Washington University in St. Louis

Washington University Open Scholarship

Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Winter 1-15-2021

Relating Spontaneous Activity and Cognitive States via NeuroDynamic Modeling

Matthew Singh
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Cognitive Psychology Commons, Engineering Commons, and the Neuroscience and Neurobiology Commons

Recommended Citation

Singh, Matthew, "Relating Spontaneous Activity and Cognitive States via NeuroDynamic Modeling" (2021). *Arts & Sciences Electronic Theses and Dissertations*. 2346.
https://openscholarship.wustl.edu/art_sci_etds/2346

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST.LOUIS

Division of Biolgy & Biomedical Sciences
Neurosciences

Dissertation Examination Committee:
Camillo Padoa-Schioppa, Chair
Todd Braver
ShiNung Ching
Nico Dosenbach
Jr-Shin Li
Jeffrey Zacks

Relating Spontaneous Activity and Cognitive States via NeuroDynamic Modeling

by
Matthew F. Singh

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

January 2021
St. Louis, Missouri

© 2021, Matthew F. Singh

Table of Contents

List of Figures vii

List of Tables xix

Acknowledgments xxii

Abstract xxiv

Chapter 1: Introduction 1

1.1 Cognitive and Neural Variability 1

1.2 Current Frameworks for Studying Brain Activity with fMRI 3

1.2.1 Functional Connectivity and Resting-State Dynamics 3

1.2.2 Previous Linkages Between Resting-State, Dynamics and Task fMRI . 5

1.2.3 Trial-to-Trial Variability in Task fMRI 6

1.2.4 Previous Dynamic fMRI Models 7

1.3 Motivation and Research Aims 8

Chapter 2: Estimation and Validation of Individualized Dynamic Brain Models with Resting State fMRI 12

2.1 Introduction 12

2.2 Methods 16

2.2.1 Nature of Interpretations from the Model 16

2.2.2 Robust Estimation of Individualized Neural Model Parameters 18

2.2.3 Study Design 21

2.2.4 Empirical Dataset 21

2.2.5 MINDy Fitting Procedure 25

2.2.6 Ground-Truth Simulations 27

2.2.7 Simulations for DFC analysis 32

2.2.8 DFC Analyses 33

2.2.9 Sensitivity Analyses 34

2.2.10 Statistical Analyses 35

2.3 Results 35

2.3.1 Overview of Results/Approach 35

2.3.2 MINDy Retrieves Individual Differences 36

2.3.3 MINDy Parameters are Reliable 40

2.3.4 MINDy Parameters are Personalized 41

2.3.5 Novel MINDy Parameters show reliable individual and anatomic variation 43

2.3.6 MINDy Predicts Individual Brain Dynamics 48

2.3.7 MINDy is Robust 52

2.3.8 Comparing MINDy with Dynamic Causal Modeling 58

2.3.9 MINDy Requires 15-20 minutes of Data 60

2.4 Discussion 61

2.4.1 Relationship with Functional Connectivity 61

2.4.2 Relationship with other Models 62

2.4.3 Comparison with Diffusion Imaging Seeded Neural Mass Models 65

2.4.4 Limitations 66

2.4.5 Future Applications and Directions 67

Chapter 3: Scalable Surrogate Deconvolution for Identification of Partially-Observable Systems and Brain Modeling 70

3.1 Introduction 70

3.1.1 Relevance to Neuroscience and Neural Engineering 72

3.1.2 Previous Work 73

3.2 Approach 75

3.2.1 Contributions 78

3.3 Surrogate Deconvolution 79

3.3.1 Building Surrogate Representations 80

3.3.2 Deploying Surrogate Models 82

3.4 Data-driven Model Identification 83

3.4.1 Modeling and Isolating Local Activity from the LFP 85

3.4.2 Reconstructing Firing-Rate from LFP 89
3.4.3 Computational Efficiency 90
3.4.4 Reconstructing Connectivity and Hemodynamics in Simulated fMRI 92
3.4.5 Empirical Dual Estimation with the Human Connectome 94
3.5 Discussion and Conclusion 95
3.5.1 Generalizability of the Problem Framework 95
3.5.2 The Role of Priors in Deconvolution 96
3.5.3 Conclusion 97
Chapter 4: Enhancing Task fMRI Preprocessing via Individualized Model-Based Filtering of Intrinsic Activity Dynamics 99
4.1 Introduction 99
4.1.1 Introducing Dynamics to Activity Flow 100
4.1.2 Previous Approaches using DCM 101
4.1.3 Filtering Instead of Parameterizing 102
4.2 Approach 102
4.2.1 Model Derivation 102
4.2.2 MINDy-based Filtering 104
4.3 Methods 105
4.3.1 Subjects 105
4.3.2 Scanning Protocol 105
4.3.3 Task Descriptions 106
4.3.4 Cognitive Control Demand 107
4.3.5 Task Manipulations 108
4.3.6 Pre-processing and Parcellation 108
4.3.7 Task GLM Analyses 109
4.3.8 MINDy Modeling 109
4.3.9 Control Pipelines 110
4.4 Validation and Comparison Criteria 112
4.4.1 Benchmarking Sustained Effects 112
4.4.2 Benchmarking Event-Related Effects 113

4.4.3 Testing Selective vs. Global Improvements 113
4.4.4 Sensitivity to Cognitive States 114
4.4.5 Sensitivity to Individual Differences 115
4.4.6 Robustness to Motion 115
4.4.7 Significance Testing for Construct Identification 116
4.5 Results 116
4.5.1 Structure and Presentation of Results 116
4.5.2 Resting-state Model Predictions Generalize to Task 117
4.5.3 MINDy-based Filtering Accounts for intra and inter-subject Variability 118
4.5.4 Improved Group-Level Detection Power 120
4.5.5 Identifying Individual Differences in a Latent Cognitive Construct 124
4.5.6 MINDy-based Filtering Selectively Enhances Task-Related Neural Signals 128
4.5.7 Sensitivity to Motion 133
4.6 Discussion 133
4.6.1 Relationship with Frequency-Based Filtering 134
4.6.2 Relationship with other approaches 135
4.6.3 Limitations 137
4.6.4 Task Dynamics Could Potentially Influence Statistical Improvements . 139
4.6.5 Conclusion 140
Chapter 5: Conclusion 141
5.1 Relationship to Existing Literature 141
5.2 Resting-State MINDy Modeling 142
5.2.1 Validation of the MINDy Framework 143
5.2.2 Disentangling Neural Dynamics and Hemodynamics 144
5.3 Implications for Individual Differences in Spontaneous Brain Activity 144
5.4 Implications for The Origin and Analysis of Task-Evoked Activity 145
5.5 Limitations and Directions for Future Work 146
5.5.1 Developing Multimodal MINDy 147
5.5.2 Explicating Task Mechanisms 147

5.5.3 Linking MINDy and Behavior 148
5.6 Conclusion 149
Chapter 6: References 150
References 150
6.1 References 150
Appendix A: SI for Resting-State MINDy Modeling .... [162]
A.1 Interpreting Model Parameters .... [162]
A.1.1 Interpreting Model Weights .... [163]
A.1.2 Interpreting Model Curvature .... [164]
A.1.3 Interpreting Model Decay .... [167]
A.2 Derivation of the MINDy Transfer Function .... [169]
A.3 Accelerated Stochastic Gradients through NADAM .... [172]
A.4 Hyperparameters in Model Fitting .... [174]
A.5 Interpretations of the Weight-Decomposition and Well-Posedness of the Problem [175]
A.6 Reliability and Individual Differences in the Weight Decomposition .... [177]
A.7 The Influence of Hyperparameter Choices on Sparsity .... [179]
A.8 Comparing MINDy and Spectral Dynamic Causal Modeling .... [182]
A.8.1 Benchmarking with Unbiased Ground-truths .... [182]
A.8.2 MINDy Performs Competitively with DCM .... [184]
A.9 Anatomical Distribution of Individual Differences .... [190]
A.10 Directed Connectivity Identified by MINDy .... [193]
A.11 Nonlinear Dynamics in MINDy .... [193]
A.12 MINDy Optimization: Under the Hood .... [202]
Appendix B: SI for Surrogate Deconvolution .... [205]
B.1 “Local” Field Potential Simulations .... [205]
B.2 Randomized Networks and MINDy Hyperparameters for simulated fMRI .... [206]
B.3 Empirical Selection of MINDy Hyperparameters .... [208]
B.4 HCP Data for Surrogate-HRF MINDy .... [208]
B.5 Derivation of Kernels from Partially-Observable State-Space Models .... [209]

List of Figures

Figure 1.1: Theoretical framework underlying the proposal. A) Cognition is manifest in internal states (neural activity) which evolves according to the individual's brain and afferents inputs to the brain. B) These relationships are formalized through dynamical systems in which the state variable (neural activity) evolves according to the brain's dynamics and exogeneous inputs. 2

Figure 2.1: Overview of Methods Employed. A) The MINDy model consists of coupled 1-dimensional neural-mass models (Hopfield form [1]). 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 ([2]) 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 Wiener-deconvolved ([3]) with a canonical HRF function (see Methods; [4]) before being analyzed with either MINDy or rsFC 16

Figure 2.2: Ground-truth validation of MINDy and rsFC at the level of single-subject 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 offdiagonal 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 simulated single-subject level (correlation with ground-truth [GT]) for MINDy (blue) and rsFC (red). Bottom row: same as top but for for predicting the differences in matched-pairs of simulation subjects who differed only in ground-truth connectivity. Simulation subjects were generated by sampling from the distribution of empirical (HCP) MINDy parameters (see Sec. 2.2.6). 38

Figure 2.3: 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. A.9. Blue line indicates mean....... 40

Figure 2.4: Local 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 in red and Limbic-(Temporal Pole) in blue) to illustrate anatomical gradients in this parameter. B) Same as A but for the decay parameter (D). C) Correlation between the first principal component of MINDy decay and "hierarchical heterogeneity" provided by Demirtas and colleagues ([5]) based upon erf transform of the T1/T2 ratio (MMP parcelation). This measure has been theorized to reflect a hierarchy of cognitive abstraction from sensory to associative cortices. 44

Figure 2.5: MINDy weights are structured, sparse, and directed. A) Left-side: Mean connection matrix W averaged across subjects and scanning session. Parcels are grouped according to the Schaeffer ([6]) 17-network parcellation (hemispheres combined) plus the free-surfer subcorticals. Right-side: thresholded anatomical projection (positive connections$\geq$ 20% max non-recurrent magnitude and negative connections with magnitude$\geq$ 8%). B) The MINDy weight distribution demonstrates sparser connectivity than rsFC. C) Parcel 187 ([6] 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 (e.g. [7]) see SI Sec. A.10 for additional, preliminary directed-connectivity results. Blue highlights chosen source-parcel. 47

Figure 2.6: MINDy models predict individual variation and central tendency of pairwise 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 atlas from [6] and collapsing the 17-network grouping down to 8). B) Scatterplot and Pearson correlation of group-average σ-DFC for data vs. model. C) Evidence of non-trivial dynamics in MINDy models. Example phase portrait of one subject projected onto the first 3 principal components. Complex orbits link neighborhoods of attractors 1 and 2 (orbits starting from each neighborhood are colored red, blue respectively). Inset figures show these attractors projected onto the brain. D) Example deterministic time-series for a limit cycle in a different representative subject averaged across five networks (Visual [blue], SomMot. [red], Dorsal Attn. [yellow], Salience [purple], and Control [green]). These deterministic dynamics demonstrate significant nonlinearity but are qualitatively different from the simulated model dynamics (e.g. for computing DFC) which include process noise. 51

Figure 2.7: 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 (see SI Fig. A.13H). C) MINDy performance in retrieving the ground-truth weight matrix (original matrix: red; asymmetric part ($W_{i,j} - W_{j,i}$): 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. 54

Figure 3.1: General Approach. A) The brain is an example of a convolutional system when viewed through BOLD fMRI. Dynamics among brain regions are highly nonlinear and usually cannot be directly observed. B) Measurements made using fMRI reflect latent brain activity passing through a hemodynamic response function (HRF). C) Surrogate Deconvolution workflow: a deconvolution surface is estimated by sampling the time-series deconvolution across a variety of kernel parameters (left). A separate surrogate model is formed for each time-point using basis regression to approximate this surface (middle). The combined surrogate models then represent the deconvolution process during parameter estimation. 74

Figure 3.2: Surrogate surfaces for performing deconvolution. A) Example fMRI time-series deconvolved by hemodynamic models with different values for the kernel parameters (α, β). The inset equation is the hemodynamic response function (kernel). Subplots show the surrogate surfaces for four sequential time-points (e.g. the deconvolved signal amplitude at that time as a function of kernel parameters). Note that the surrogate surfaces at a fixed time-point are smooth, whereas the variation in deconvolved signal across time is much less regular. B) Performance in reconstructing the iterative Richardson-Lucy deconvolution of the same signal across an 81x81 grid (same range as A) using third-order bivariate polynomials (Appendix Sec. B.4). Inset shows a representative stretch of 200 points (144s)......................... 77

Figure 3.3: Surrogate deconvolution's performance in inverting a neural-mass model of Local Field Potential. A) Model schematic: output signals from each population arrive at post-synaptic terminals. Electrode measurements primarily reflect the post-synaptic potentials generated from synaptic activity. B) Bench-marking total error in identifying the synaptic time-constants and network connectivity. Surrogate deconvolution is compared to the current gold-standard: joint-Kalman Filters (Unscented and Extended). C) Same as (B), but displaying performance in terms of correlation rather than mean-square-error. D) Performance in reconstructing local spiking activity from electrode measurements using the identified system models for a variety of state-estimation techniques. E) State estimation performance for a representative case (the simulation with median jUKF+UKF performance). F) Computational complexity of system identification methods in terms of the number of brain regions considered. The top inset shows the run-time and least-squares fit on rescaled x-axes (n_{pop}^4) to demonstrate the $\mathcal{O}(n_{pop}^4)$ complexity of Kalman-Filtering in terms of the number of brain regions (n_{pop}). The bottom inset is the same, but for surrogate deconvolution and $\mathcal{O}(n_{pop}^2)$. Calculations were performed single-core on an Intel Xeon E5-2630v3 CPU.................. 84

Figure 3.4: Incorporating HRF surrogate-deconvolution into MINDy. A) Without HRF modeling, connectivity estimates degrade with spatial variability in the neurovascular coupling. Fitting the HRF through surrogate deconvolution preserves performance. B) Same as (A) but for the asymmetric component of connectivity. C) HRF parameter estimates from HCP data are reliable across scanning days and subject-specific. D) Spatial map of the mean α parameter estimate across subjects. E) Same as (D), but for the second HRF parameter (β). F) Spatial map of the mean time-to-peak in the fitted HRF's............................... 91

Figure 4.1: Resting-state MINDy models generalize to task. A) MINDy models trained on resting-state data produce similarly accurate predictions for task data. Goodness-of-fit is quantified in terms of the mean R^2 value across all brain parcels ([6]) and scanning runs (n=6) in predicting the difference time series: $X_{t+1} - X_t$ ([8]). B) Individual-differences in model accuracy are highly correlated between resting-state and task data... 117

Figure 4.2: MINDy-based Filtering reduces variability within and between subjects. A) MINDy-based Filtering accounts for a significant portion of unique variability within each subject's data. This effect holds across tasks (results averaged over all parcels, subjects). Variance partitioning was performed after removing variation due to nuisance factors (motion and drift). B) MINDy-based Filtering reduces the between-subject variability of task-evoked signals. Example shown is the mean signal over the DMCC34 parcels for the Cued-TS high control-demand condition (incongruent trials). C) Variability also decreases for contrasts between conditions. Example shown is for the AX-CPT (BX-BY contrast). D) Spatial-map for the proportion of variability in cognitive control signals attributable to group (collapsed across tasks) for MINDy-filtered data. E) Difference in the relative group-explained variability between MINDy and the original data. Note that MINDy-based filtering actually decreases the proportion of group variance in some regions, but increases for task-implicated regions (lPFC, anterior insula, etc.). F) MINDy-based Filtering increases the group-level effect size for each task. This can be seen in the number of parcels exhibiting higher t-values after MINDy-based Filtering relative to conventional analyses (i.e., above the identity line). Yellow dots indicate significant parcels (in terms of the control-demand effect) which also had increased effect sizes from MINDy-based Filtering, while blue dots denote significant parcels whose effect sizes were larger with conventional analyses. Teal dots denote parcels which did not exhibit a significant control-demand effect for either method. 119

Figure 4.3: MINDy-based Filtering increases power in detecting event-related and sustained task effects. A) Spatial distribution of high-vs.-low conflict effects averaged across tasks using conventional analyses and B) the proposed technique. Thresholds correspond to $p < .05$ for a single task. C) Paired comparisons of effect size between control methods and the proposed approach in identifying high-vs.-low conflict effects for significant brain parcels. Values less than zero indicate that the propose technique improves upon controls. D) Comparisons for the absolute sustained effect over significant parcels (positive or negative). Parcels were deemed significant if they passed $p < .05$ for either method within a pair. Comparison techniques are denoted:"Orig"=original analysis (no filter), "Pre-Reg"=motion and drift pre-regressed before GLM fitting, "Loc"=AR model filtering with local/heterogeneous parameters, "Glob"=AR model filtering with a single (anatomically global/homogeneous) parameter, and "MINDy"=MINDy-based Filtering. 121

Figure 4.4: Improved detection power over a set of 34 parcels with consistent event-related effects A) 34 parcels were identified as exhibiting a consistent cognitive control effect across tasks during study piloting. B) Paired comparisons of the cognitive control effect over the DMC34 parcels: MINDy-based Filtering minus control. C) Comparisons for the absolute sustained effect the DMCC34 parcels: MINDy-based Filtering minus control. Values greater than zero indicate improvement. Comparison techniques are denoted:"Orig"=original analysis (no filter), "Pre-Reg"=motion and drift pre-regressed before GLM fitting, "Loc"=AR model filtering with local/heterogeneous parameters, "Glob"=AR model filtering with a single (anatomically global/homogeneous) parameter, and "MINDy"=MINDy-based Filtering. 123

Figure 4.5: MINDy-based Filtering improves identification of cognitive control demand. A) Cross-task generalizability (ICC) for group means in the low-conflict condition (low construct purity across tasks) and the high vs. low conflict contrast (high purity) by method. B) Generalizability of individual differences (z-scores) for the "DMCC34" parcels. Boxplots display distributions across parcels. C) Anatomical distribution of generalizability for the proposed technique and (D) conventional analyses for the high vs. low conflict contrast. E) The difference in cross-task generalizability between MINDy-based Filtering and conventional analyses. Comparison techniques are denoted:"Orig"=original analysis (no filter), "Pre-Reg"=motion and drift pre-regressed before GLM fitting, "Loc"=AR model filtering with local/heterogeneous parameters, "Glob"=AR model filtering with a single (anatomically global/homogeneous) parameter, and "MINDy"=MINDy-based Filtering. 125

Figure 4.6: MINDy-based Filtering magnifies task-related signals relative to all controls. A) Scatter plots compare methods in terms of group-level T-values for the cognitive control effect collapsed across tasks (all tasks plotted together). Unlike Fig. 4.2F, each plot corresponds to one of the four control methods combined across tasks (instead of one plot per task). In particular, the upper left corner plots all of the points from Fig. 4.2F combined. Improvements can be seen in the number of parcels exhibiting higher t-values after MINDy-based Filtering relative to conventional analyses (i.e., above the identity line). Yellow dots indicate significant parcels (in terms of the control-demand effect) which also had increased effect sizes from MINDy-based Filtering, while blue dots denote significant parcels whose effect sizes were larger with conventional analyses. Teal dots denote parcels which did not exhibit a significant control-demand effect for either method. B) Same as (A) but for sustained effects. 129

Figure A.1: Interpreting Phenomenological Model Parameters. A) The weight matrix largely captures monotone causal relationships. However, the sign of the causal relationship depends upon the sign of the actual inter-regional connection and which neurons are involved. Excitatory connections/cells are depicted with green arrows and inhibitory in red dots. B) When the sign of connections between regions is mixed, it is possible for indirect relationships to appear stronger than direct connections. Local network structure could influence transfer function and decay parameters. C) Networks with greater reciprocal inhibition (red lines) have a faster time-scale, hence greater decay than those with reciprocal excitation (green). D) A toy example of a network with near binary output due to reciprocal excitation in the output cells (triangles). E) A toy example with a more graded output rule due to inhomogeneities in the excitation of output cells. .... [165]

Figure A.2: Schematics of data-processing and data-generation pipelines. A) Secondary preprocessing consisted of the pipeline proposed by Siegel and colleagues ([2]) followed by Weiner deconvolution ([3]) with a canonical HRF function. B) Simulated BOLD signals were generated by integrating MINDy models as stochastic differential equations, downsampling results to the scanner TR, convolving with a (potentially random) HRF and adding measurement noise. The resultant signal was then deconvolved with the canonical HRF. C) MINDy parameter estimation consists of iteratively updating estimates using current and past error gradients according to NADAM ([9]).......... [168]

Figure A.3: Schematic of NADAM benefits with illustrative error surfaces (y-axis) for fitting a parameter (x-axis values) on the first scanning session (left) and the second (right). The NADAM algorithm uses momentum to avoid shallow local minima (green). This feature also prevents convergence to overly sharp minima (even if they are global) because such error surfaces can often correspond to overfitting (blue) and hence do not generalize across sessions. Rather, NADAM emphasizes solutions to deep basins (purple) which may prove the most robust. . . [174]

Figure A.4: Local MINDy parameters (curvature and decay) exhibit consist anatomical structure within and between networks. A) Distribution of curvature parameter values for each brain parcel grouped according to network (17-network [6]). 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 ([10]). D) Profile for the gwMRF ([6]) parcellation. E-H) same as A-D but for the decay parameter. I) Hierarchical Heterogeneity map by Demirtas and colleagues ([5]) using group T1/T2 ratio. J,K) same as G,H but for the first principal component of decay across subjects. . . . . [179]

Figure A.5: Comparison of accuracy and run-time for MINDy and spDCM. A) Accuracy in estimating ground-truth connectivity from Hopfield-network simulations by network size. "Reduced MINDy" indicates that all regularization terms were removed from MINDy to avoid bias (analogous results for the full MINDy model are in Sec. 2.3.7). Lines indicate mean and bars indicate first/third. quartiles. B) Accuracy in estimating ground-truth connectivity from Neural Mass simulations by network size. Results for the full MINDy model are in Sec. 2.3.7. C) Model performance as a function of HRF spatial variability for two network sizes: 6 and 8 nodes. Note that MINDy performance decreases with HRF spatial variability, whereas the effect for spDCM is minor. D) Full run times for MINDy and spDCM for each simulation as a function of simulation type/size. E) Run time per EM iteration (spDCM) and for 10,000 mini-batches in MINDy. We chose to compare with 10,000 mini-batches so that the run-times would be comparable for the smallest network size (n_{Pop}=6). F) Model performance as a function of HRF spatial variability for two network sizes: 6 and 8 nodes. Note that MINDy performance decreases with HRF spatial variability, whereas the effect for spDCM is minor. . . . . [181]

Figure A.6: Anatomical distribution of inter-individual variation in MINDy. A) QCD of the curvature parameter with top-20% threshold. B) Same as A) but for the decay parameter. C) QCD for MINDy connection weights. Weights in which the sign was inconsistent across subjects (<75% agreement) or low reliability (Fisher ICC<.5) were censored (grey). D) Mean weight QCD within each network combinations. If over 95% of parcel-wise connections were censored, the network-level connection was also censored (white). .... [189]

Figure A.7: Connection asymmetries identified by MINDy 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 ([6]) .... [192]

Figure A.8: Eigenvalue analyses indicate that empirical MINDy models do not possess a global, Lyapunov stable attractor. A) Distribution of eigenvalues (real-part) for local-linearization about the origin. Shading indicates ±SD and black lines give the data's maximum/minimum. Note the presence of positive eigenvalues which indicate nontrivial dynamics (Proposition 2). B) Scatterplot of these eigenvalues (all subjects) in the complex plane suggests that the greatest "spin" (complex components) occurs along the unstable subspaces (corresponding to eigenvalues of positive real-part). .... [195]

Figure A.9: MINDy strongly 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. The identity line (perfect match) is indicated in red G) Same as F) but with a slightly decreased noise term (85% of original). .... [196]

Figure A.10: MINDy differentiates model parameters between individuals and 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. .... [197]

Figure A.11: MINDy reproduces reliable, accurate estimates of dynamic functional connectivity (DFC). We use σ-DFC to disambiguate the standard-deviation measure for DFC (e.g. [11],[12]) 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. 2.6) .... [198]

Figure A.12: After pre-processing, 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). .... [200]

Figure A.13: Sensitivity to various forms of noise in data (Same as Fig. 2.7 A,B, but with 17 levels of noise). A-C) MINDy estimates all parameters of ground-truth models accurately even in the presence of additive measurement noise. D) Examples of how increasing the variability of hemodynamic parameters changes the shape of randomly drawn hemodynamic response functions (HRF). E-G) Hemodynamic variability does not alter the mean performance of MINDy estimates. H) Hemodynamic variability decreases the consistency of MINDy performance (more variable correlations with ground truth). .... [201]

Figure A.14: MINDy estimates are robust to secondary preprocessing choices: Motion-only (Red), Motion+CompCor (Green), Motion+CompCor+GSR (Red). A) Test-retest reliability of rsFC and the MINDy weights (W), curvature (α) and Decay (D) parameters by secondary preprocessing pipeline. B) Correlation between parameter estimates for different preprocessing pipelines applied to the same scanning data (single session). Color of the bar indicates which pipelines are being compared: R+G indicates similarity between Motion-only and CompCor estimates, R+B indicates Motion-only vs. GSR, G+B indicates CompCor vs. GSR (results are displayed in this order left to right within each column) .... [201]

List of Tables

Table 3.1: Common neuroimaging measures subject to convolution 73

Table A.1: Types of assumptions ("Difference") made by MINDy vs. spDCM and our controls to mitigate these differences in simulated comparisons (SI Sec.A.8). .... [187]

Table A.2: Accuracy (r) for each method (Regr.=regression) for the neural-mass simulation (SI Sec. A.8) with variable population sizes (left side). Test-statistics (paired t-test; 2-tailed) for MINDy-spDCM are provided on the right side. The number of nodes per simulation is listed under "nodes" whereas the number of simulation instances is listed under "N". The paired difference in accuracy between MINDy and spDCM is denoted Δr. We denote $* = p < .01$ and $** = p < .001$. .... [187]

Table A.3: Accuracy (r) for each method (Regr.=regression) for the Hopfield-network simulation (SI Sec. A.8) with variable population sizes (left side) and a random global HRF ($\sigma = .5$). For this simulation we added a new population size (9 Nodes) post-hoc to see whether the unexpected decrease in spDCM accuracy between 8 and 10 nodes was continuous (it was). Since this special case contained an odd number of nodes the hyperdistribution parameter q in Sec. 2.2.6 was always equal to one (instead of one and two with equal probability). Test-statistics (paired t-test; 2-tailed) for MINDy-spDCM are provided on the right side. The number of nodes per simulation is listed under "nodes" whereas the number of simulation instances is listed under "N". The paired difference in accuracy between MINDy and spDCM is denoted Δr. We denote $* = p < .01$ and $** = p < .001$. .... [188]

Table A.4: Accuracy (r) for each method (Regr.=regression) for the Hopfield-network simulation (SI Sec. A.8) with spatially variable HRF and 6 nodes (left side). Test-statistics (paired t-test; 2-tailed) for MINDy-spDCM are provided on the right side. The standard-deviation of the HRF parameters is listed under "σ" whereas the number of simulation instances is listed under "N". The paired difference in accuracy between MINDy and spDCM is denoted Δr. We denote $* = p < .01$ and $** = p < .001$ .... [188]

Table A.5: Ground-truth validation performance of MINDy and rsFC in recovering the weight matrix of a single subject and the arithmetic difference of weight matrices between subjects. We denote significance with $^{**} = p < .001$, 2-tailed for the contrast Weights minus rsFC. .... [195]

Table A.6: Sensitivity of MINDy weights and rsFC to changes in non-connectivity parameters in a ground-truth simulation Performance is measured in terms of false positives=the percentage of connections that change (thresholded by $p < .05$) due to a change in the ground truth model's curvature/decay. Thus, lower values indicate less false positives (less sensitivity) due to non-connectivity variables. We denote significance with $^{**} = p < .001$, 2-tailed for the contrast Weights minus rsFC. ..... [196]

Table A.7: Test-retest correlation across scanning sessions for MINDy parameters when sessions were drawn from the same subject or from different subjects. Group-level statistics are present in mean correlation (SD) form. Group level permutation testing (100,000 each) produced $p's \approx 0$ 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 .... [197]

Table A.8: Comparing the test-retest reliability and pre-processing sensitivity of the MINDy connectivity parameter 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. 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. .... [199]

Table A.9: MINDy performance in inverting the weight matrix and its asymmetries in cases of model mismatch. Ground truth models were either a tanh rate-model downsampled to the fMRI TR, a rate model with spatially heterogeneous hemodynamic response functions, or a neural mass model using the nonlinear Balloon-Windkessel model of hemodynamics. [199]

Table A.10: Definition of variables ("Name") used in MINDy gradient calculations and their interpretation ("Meaning"). The term n_{Batch} denotes the number of samples (time-points) included in each minibatch (the training data for a given NADAM iteration). .... [203]

Table A.11: Equations used to efficiently calculate MINDy parameter gradients. These equations leverage the additional variables defined in SI Tab. A.10. Note that the decay parameter is updated in terms of its square-root (D_2) and the curvature parameter is updated in terms of the linearized form ξ .... [203]

Table A.12: NADAM hyperparametes for each MINDy parameter and the distributions used to initialize each parameter. NADAM hyperparameters consist of the update rate ("Rate"), decay rate of gradients (μ), decay rate of squared-gradients (ν), and regularization term ε. .... [203]

Table A.13: Chosen values for other hyperparameters used in MINDy. These (non-NADAM) hyperparameters consist of the four regularization terms (λ_i) in the cost function (Eq. 2.4), the minimum allowable value for D (D_{min}), and the scaling factor of the transfer-function (b) .... [204]

Acknowledgments

This work was supported by the US National Science Foundation: DGE-1143954 and NCS-FO 1835209, and National Institute of Mental Health: MH066078-15S1.

I am profoundly indebted to my mentors Todd Braver and ShiNung Ching (ordered alphabetically) for shouldering the risk and commitment inherent in my research. Prior to our collaboration, ShiNung and Todd had never met nor operated in a common research space. By agreeing my co-mentorship, they assumed signifnicant risk for which I am very grateful. They have always provided candid and thoughtful advice and also promoted the self-discipline that has enabled me to mature both academically and personally.

I thank CCP lab members Anxu Wang, Mitch Jeffers, Jo Etzel, and Michael Freund for their discussions and assistance. I thank Michael W. Cole for discussions regarding MINDy.

I thank my children, Bodhi and Dominc for their exuberance and my spouse, Nanoy, for her love. I thank my parents for their unconditional support throughout all stages of my career.

Matthew F. Singh

Washington University in Saint Louis

January 2021

Dedicated to my loving spouse Nanoy, and my two frenetic dinosaurs: Bodhi and Dominic.

ABSTRACT OF THE DISSERTATION

Relating Spontaneous Activity and Cognitive States via NeuroDynamic Modeling

by

Matthew F. Singh

Doctor of Philosophy in Biology and Biomedical Sciences

Neurosciences

Washington University in St. Louis, 2021

Professor Todd Braver

Professor ShiNung Ching

Stimulus-free brain dynamics form the basis of current knowledge concerning functional integration and segregation within the human brain. These relationships are typically described in terms of resting-state brain networks—regions which spontaneously coactivate. However, despite the interest in the anatomical mechanisms and biobehavioral correlates of stimulus-free brain dynamics, little is known regarding the relation between spontaneous brain dynamics and task-evoked activity. In particular, no computational framework has been previously proposed to unite spontaneous and task dynamics under a single, data-driven model. Model development in this domain will provide new insight regarding the mechanisms by which exogeneous stimuli and intrinsic neural circuitry interact to shape human cognition. The current work bridges this gap by deriving and validating a new technique, termed Mesoscale Individualized NeuroDynamic (MINDy) modeling, to estimate large-scale neural population models for individual human subjects using resting-state fMRI. A combination of ground-truth simulations and test-retest data are used to demonstrate that the approach is robust to various forms of noise, motion, and data processing choices. The MINDy formalism is then extended to simultaneously estimating neural population models and the neurovascular coupling which gives rise to BOLD fMRI. In doing so, I develop and validate a new optimization framework for simultaneously estimating system states and parameters. Lastly, MINDy models derived from resting-state data are used to predict task-based activity and remove the effects of intrinsic dynamics. Removing the MINDy model predictions from task fMRI, enables separation of exogenously-driven components of activity from their indirect consequences (the model predictions). Results demonstrate that removing the predicted intrinsic dynamics improves detection of event-triggered and sustained responses across four cognitive tasks. Together, these findings validate the MINDy framework and demonstrate that MINDy models predict brain dynamics across contexts. These dynamics contribute to the variance of task-evoked brain activity between subjects. Removing the influence of intrinsic dynamics improves the estimation of task effects.

Chapter 1

Introduction

1.1 Cognitive and Neural Variability

Human behavior and cognition are remarkably variable across both situational (task) contexts and individuals ([13], [14]). Traditionally, cognitive science paradigms have sought to isolate features of these complex relationships through experimental control of task contexts, as well as through comparisons between individuals/groups. Although complex interactions likely subserve many human behaviors, much of the cognitive science and neuroscience literature has studied cognition/behavior as interactions among three main variables ([15]): an individual (e.g. their brain), the individual's internal state (neural activity) and an external context (e.g. a task). In neural terms, this relationship may be stated as follows: at any moment in time, brain activity evolves according to constraints imposed by brain structure, the current state of endogenous (or intrinsic) brain activity, and modulations due to stimulation arriving via peripheral (afferent) inputs (Fig. 1.1A). This framework has thus far granted numerous insights into individual differences, environmental influences and neural correlates of specific

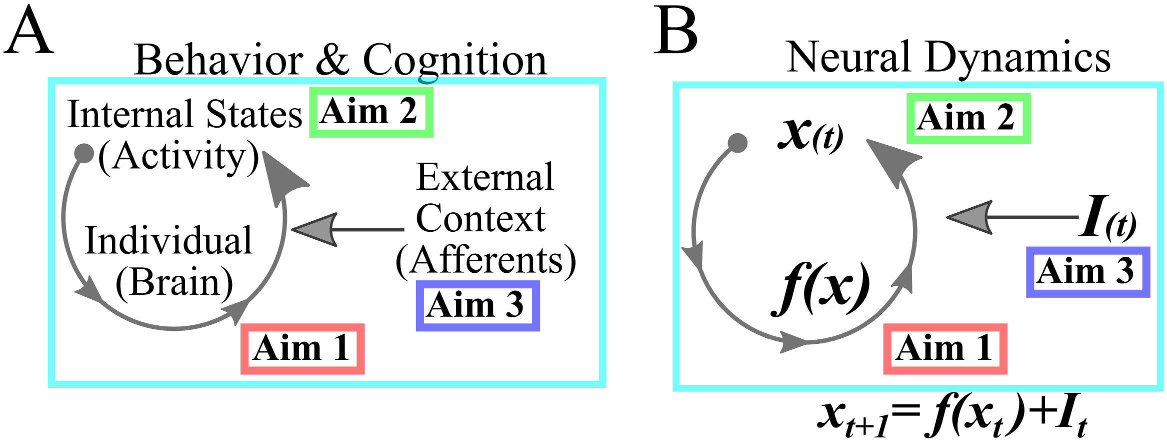

Figure 1.1: Theoretical framework underlying the proposal. A) Cognition is manifest in internal states (neural activity) which evolves according to the individual's brain and afferents inputs to the brain. B) These relationships are formalized through dynamical systems in which the state variable (neural activity) evolves according to the brain's dynamics and exogeneous inputs.

human behaviors. These interactions can be formalized through the language of dynamical systems. In this interpretation, the nervous system is treated as a dynamical system with states (neural activity) that evolve in time according to rules and constraints imposed by physiology. External forcing (context) modulates this system and is typically modeled as input from exogenous sources (e.g. [16]; Fig. 1.1B). These inputs alter neural activity as opposed to modifying the system itself (e.g. changing architecture).

By leveraging generative models, the dynamical systems approach offers concrete predictions by expressing future neural activity or behavior as a direct function of the system architecture (brain), its current state (activity), and the input to the system (context). These approaches have proven highly influential in specific cognitive domains such as perceptual decision making (e.g. [17]) and may prove vital for personalized neuroscience/brain medicine ([18],[19],[20]). Given a personalized brain model, dynamical systems analyses predict how various inputs (treatments) will affect the system state (individual brain activity). Conversely, such models can also be used to predict how individual differences in brain structure might lead distinct brain responses to the same context; as such, this formalism has the potential to provide a linking proposition between aberrant physiology and psychopathology. Dynamical systems techniques will also benefit basic cognitive neuroscience by providing a parsimonious, quantitative mapping from individual differences in brain architecture to individual differences in brain activity and functioning (and possibly to observable behavioral correlates) across different situational or task contexts.

1.2 Current Frameworks for Studying Brain Activity with fMRI

1.2.1 Functional Connectivity and Resting-State Dynamics

The nervous system has long been studied as a dynamical system (e.g. [21]) and dynamical-systems principles are increasingly informing high-level theories of brain function(e.g. [22], [23]). Moreover, dynamical signatures within fMRI data are being increasingly used to characterize cognitive states and have offered new insight into the intrinsic interactions between brain networks ([24],[25]). However, fMRI brain dynamics exist at multiple spatial and temporal scales ([26]) so there remains substantial controversy regarding which dynamics in the fMRI signal are meaningful. At the simple-end of dynamics, the resting-state functional connectivity (rsFC) describes the coevolution of brain regions through the correlation coefficient statistic ([27]). Functional connectivity has been well-studied in terms of its relation to the underlying anatomy ([26],[28]), and this literature has provided new insight into brain organization (e.g. [29]) and its relation to coordinating function ([30]). Functional connectivity has also proved to be a reliable measure ([31],[32]) and has thus become widely employed within the systems neuroscience community.

In contrast, the study of fMRI dynamics, using approaches that extend basic functional connectivity approaches, such as dynamic functional connectivity (dFC), has proven to be quite controversial. Specifically, typical dFC approaches considers the changing nature of the correlational relationship between brain regions as they evolve in time. Unlike conventional rsFC, dynamic FC does not consider resting state brain activity to be a stationary process and instead estimates time-varying functional connectivity: how brain areas become more or less correlated with each other using a temporally local sliding-window approach ([33],[34], [35]). Some methods of computing dynamic functional connectivity (dFC) have been found to be less reliable than conventional functional connectivity ([11]) and the debate continues over whether the dynamic FC is in fact, an artifact of sampling variability, physiological states such as sleep, or noise within a stationary process ([36],[37]), rather than the identification of a phenomenon that is qualitatively distinct from static FC. More recent studies have integrated these criticisms to build stronger null-models for dFC analyses ([34], [35]) and studies employing these methods have continued to find some evidence for resting state non-stationarity ([38],[34]).

Both static and dynamic functional connectivity have been used to investigate the organization of brain function (e.g. [29],[39], [25]) and a new thrust in systems neuroscience is to study how dynamic patterns (whether simple correlations or higher-order forms) relate to cognition ([40]). Functional connectivity methods have demonstrated the ability to identify individual variation ([41],[42]) in brain organization, and the ability to predict patterns of behavior and neural activation during task ([30], [43], [44]). Thus, statistical approaches to resting-state fMRI can identify individual traits and predict static patterns of brain activity across contexts. However, the mechanisms underlying the transition from resting-state to task-driven activity remains unknown. Transitions are, by definition, dynamic phenomena and are therefore ill-suited to static analyses. In this dissertation, I aim to fill this gap, by leveraging dynamic brain models (as opposed to static descriptions) as a primary investigative tool.

1.2.2 Previous Linkages Between Resting-State, Dynamics and Task fMRI

Until recently, separate methods were employed to study resting-state and task fMRI. In part, this separation may reflect the distinct theoretical accounts of their origins. Early task-fMRI analyses assumed that neural activity was largely context-driven and thus largely modeled the BOLD signal as the sustained, selective recruitment of functionally-relevant areas ([4]) rather than seeking to describe regional interactions at large-scale (e.g. whole-brain). By contrast, resting-state BOLD activity ([27]) was thought to reflect spontaneous coactivation of brain regions due to the underlying propagation and inter-regional interaction of signals that could not be temporally resolved using fMRI. Thus, traditional resting-state analyses treated the BOLD signal as structured in space but approximately random in time. Over the past decade, however, the distinction between spontaneous and task-evoked neural activity has diminished. Two major findings have illuminated the relationship between task and resting-state fMRI signals.

First, a finding that was observed across multiple recording modalities, is a reduction in neural variability following stimulus presentation that correlates with the evoked response magnitude ([45], [46], [16]). Subsequent findings suggested that these effects reflect a reduction in brain entropy during task states (i.e. task activity is more structured), and that this constriction results from the propagation of exogenous inputs through a fixed network structure ([16]). Secondly, a growing body of work now demonstrates that the spatial structure of spontaneous brain coactivation is largely preserved during task states ([30],[43]). After task evoked effects are estimated and removed, the covariance of the residual BOLD signal contains similar spatial structure as that observed during rest states ([30]). Moreover, the spatial patterns of task-evoked brain activity in fMRI largely reflect the main components of resting-state BOLD network connectivity ([43]); likewise, individual variation in resting-state network structure predicts individual variation in task-evoked activity ([43],[47]). Together, these findings indicate that during task states, brain activity patterns are more temporally restricted (i.e., in dynamical systems terms, they can characterized as inhabiting a smaller region of activity space; [16]), whereas the spatial patterns of coactivation are highly similar (if not identical) during task contexts and resting-state. These findings suggest that the coupling strength of brain regions during task is similar to rest (after removing the evoked-signal) so models based upon resting-state (e.g. MINDy) may be able to also explain brain function during task.

1.2.3 Trial-to-Trial Variability in Task fMRI

Task-related analyses in fMRI typically involve the use of statistical general linear models (GLMs), to estimate the mean time-course of evoked-responses after removing nuisance covariates. These approaches have proven statistically powerful for simple tasks, and as such, characterize much of the current literature regarding task-induced activation. However, over the past decade, improvements in the accuracy and speed of fMRI acquisitions has given birth to a new literature concerning within-subject trial-to-trial variation in brain activity. These analyses studies have generated two key findings relevant to the current study: 1) trial-to-trial variation in BOLD responses predict within-subject behavioral variation ([48],

[49], [50], [46], [51]) and 2) the BOLD signal elicited by a stimulus is dependent upon the previous pattern of brain activity ([46]). We use the term “brain activity” in the latter case to indicate that this history dependence is thought to be neural, rather than solely reflecting potential nonlinearity in the hemodynamic coupling. These results indicate that trial-to-trial variation in the BOLD response is due, in part, to variation in the underlying neural activity ([46],[16]). Moreover, this variability is behaviorally relevant and is history dependent, i.e. reflects underlying dynamics ([46]). Thus, the neural activity associated with BOLD is increasingly considered as a nonlinear dynamical system—one in which the spatiotemporal responses to an input depend upon its current state and evolve according to a fixed set of rules. This framework contrasts with both current statistical (i.e., GLM) models and dynamical causal modeling (DCM; [52]) approaches. These approaches either treat the neural activity as a noisy autoregressive signal (most GLMs) or as a linear dynamical system (popular DCMs). This distinction is critical because, in a linear system. the evolution of internal states is mathematically separable from the downstream-effects of input to the system. Thus, it is not possible for a linear dynamical system (i.e. DCM [52]) to explain trial-to-trial variation in brain activity in response to the same “input” (e.g. trial type).

1.2.4 Previous Dynamic fMRI Models

Dynamical systems modeling for fMRI has historically taken two forms: either individualized models that only deal with restricted, small networks (e.g. [52]), or instead large whole-brain models, in which parameters are either assumed or derived from group-averaged structural data ([53]). For individualized fMRI the dominant technique and current state-of-the-art is dynamic causal modeling (DCM, [52]), which attempts to model both intrinsic dynamics and task-evoked changes. Rather than specifying a biophysical form, the DCM approach to modeling involves the use of a bilinear form, as a first-order approximation for an unknown system of equations. Thus, the neural model underlying DCM consists of a simple approximation (local linearization) rather than a biophysically motivated equation. However, despite their name (”Dynamic” Causal Modeling) these individualized models are not intended to be generative or truly dynamic, as they approach a single equilibrium in the absence of input (or with constant input) and are limited to a relatively small number of brain regions. The limitation in terms of the number of regions is in part due to the Bayesian estimation technique used for parameterization, which does not scale well with dimensionality. Recently, DCM methods have also been applied to fMRI data in the absence of task (”resting state”) to produce a purely linear model ([54]). Rather than attempting to model task effects, individualized resting-state DCMs are designed for network discovery ([55], [56]) and offer improved spatial resolution by modeling up to 36 brain regions ([57]). However, modern brain-parcellation schemes have demonstrated the importance of including hundreds of regions (e.g. [58], [10]) so the currently available high-resolution and generative fMRI models still preclude individual-subject, whole-brain modeling. In contrast, current generic (non-individualized) models (e.g. [53], [59]) offer high spatial resolution and generative power ([60], [61]). Large-scale brain modeling has already proven useful for studying brain pathology ([62]) and explaining brain activity patterns. However, the vast number of parameters required for large-scale models has thus-far prevented detailed brain modeling, at the level of individual subjects. The goal of the current proposal is to remedy this situation, using a novel framework, termed MINDy (for Mesoscale Individualized NeuroDynamic Models).

1.3 Motivation and Research Aims

Over the past few decades, the ability to develop whole-brain neural models and use these to predict experimental, stimulation, or treatment effects has steadily matured (e.g. [63],[64][5]) although the role of individualized, functional data is still limited (e.g. [5]). By contrast, previous precision medicine initiatives (e.g., Human Connectome Project; HCP [65]) have generated increasingly sophisticated means to identify and characterize human brain networks as well as variation in these networks among individuals (e.g. [47], [66]). These methods can also predict the spatial structure of mean task activity (e.g. [47]). However, as static, statistical models, these existing approaches have limited ability to characterize the within-subject variation in neural/behavioral responses that arises as a function of different task or contextual states (neural activity preceding task, [49], [50], [67]). Conversely, the studies that have focused on within-subject variation provide a description of the statistical interaction of internal states and external context but do not possess a formal, mechanistic framework from which to characterize this interaction. Thus, although the previous literature has quantitatively linked pairs of factors in the person-state-context interaction, empirical models which seek to link all three are currently lacking.

The aim of the current work is to provide the first steps towards the establishment and validation of such a model. Our approach consists of first developing individualized neural and neurovascular models to describe the evolution of each subject's latent brain activity and its relation to measured BOLD. Using the neurovascular models, we estimate latent neural states from BOLD data. Lastly, we use task fMRI data from the same subjects, to isolate the influence of task-contexts (stimuli, trial-type/task-set etc.) on brain activity by filtering-out the influence of intrinsic dynamics (as identified by resting-state models). Together, these three components form a generative model which predicts how brain activity will evolve for each subject × task-condition × initial state (brain activity at trial onset). The current work is organized around these components with one chapter/aim devoted to the development of each.

Aim 1-Develop and validate MINDy models. In this aim, I capitalize on a novel numerical optimization framework ([8]) to fit large-scale neural models to resting-state fMRI data in the Human Connectome Project (HCP; [65]). These models are then validated using ground-truth simulations and evaluated in terms of robustness to physiological confounds. Critically, the final phase of validation tests MINDy model dynamics in terms of the accuracy of cross-validated model predictions over both short and long timescales (predicting BOLD and predicting dynamic functional connectivity [DFC]; [35], respectively). This work is currently published as: Singh, Braver, Cole, & Ching (2020). Estimation and validation of individualized dynamic brain models with resting state fMRI, *NeuroImage*, 221:117046.

Aim 2-Estimate hemodynamics and latent neural activity from BOLD signals. In this aim, I extend the MINDy estimation approach by incorporating an additional novel estimation technique (Surrogate Deconvolution), that provides the means to model region-specific neurovascular coupling underlying resting-state fMRI time-series. Surrogate Deconvolution enables rapid simultaneous estimation of each region's hemodynamic response function, latent neural activity, and neural model. The Surrogate Deconvolution technique is validated via ground-truth simulations. I demonstrate its relevance in estimating hemodynamic kernels for each brain region using resting-state data, and then identifying systematic and individual-specific spatial variation in the neurovascular coupling from resting-state fMRI. Improvements in model predictions relative to the base MINDy estimations are then demonstrated. This work is currently published as: Singh, Wang, Braver, & Ching (2020). Scalable surrogate deconvolution for identification of partially-observable systems and brain modeling. *Journal of Neural Engineering.*

Aim 3-Identify exogeneous and intrinsic dynamics underlying task fMRI. In this aim, I apply the dual hemodynamic and neural estimation approach to identify, at the individual subject level, latent neural activity present in the task BOLD signal. Specifically, applying MINDy models derived from each individual's resting-state fMRI data, I demonstrate how exogenously driven activity can be identified in task-related brain dynamics in relation to its complement (i.e., intrinsic activity dynamics). Through this separation, I quantify how task events alter neural activity over fast time scales (1s) and how these changes modulate subsequent brain activity for prolonged cognitive states (the extracted "exogenous" and "intrinsic" dynamics, respectively).

Chapter 2

Estimation and Validation of Individualized Dynamic Brain Models with Resting State fMRI[1]

2.1 Introduction

To understand human brain function, it is necessary to understand the spatial and temporal computations that govern how its components interact. This understanding can take multiple levels, 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. However, 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 ([53], [68], [69]). 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 ([53],[68]).

---

[1] Chapter reprinted verbatim from previously published work: [8]

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, [65]). 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; [27]). As a result, it is sometimes difficult to make mechanistic inferences based upon functional connectivity correlations ([70]). 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 direct-parameterization approaches, with methods being developed to identify directed, causal influences between brain regions (e.g. [57]). Conversely, neural mass modeling approaches have also been extended to study human brain activity in a generative fashion ([61]), and these have provided new insights into the computational mechanisms underlying fMRI and MEG/EEG activity dynamics ([59], [63], [64], [5], [71]). However, unlike (linear) data-driven approaches (e.g. Dynamic Causal Modeling; [52],[57]), neural mass models have been limited to replicating higher-level statistical summaries, such as functional connectivity, rather than predicting the actual time-series. This fact may not be relevant for some applications in which statistical descriptions will suffice. However, there remain many applications in basic neuroscience, neural medicine, and neural engineering for which more precise descriptions could be profitably leveraged.

Unfortunately, current approaches of both types have important limitations. In particular, the existing approaches to directly parameterize models (e.g. DCM) are subject to potential misinferences due to assumptions of linearity ([72]), and, in some cases, limitation to a relatively small number of brain regions ([73],[74],[75]). This number has increased dramatically in recent years by assuming a fixed hemodynamic response function ([76]), but remains well below modern brain parcellations, which feature several hundred regions (e.g. [10], [6], See Discussion). Likewise, with current neural mass modeling approaches, their ability to quantitatively recreate key features of individual-level functional connectivity has also been limited ([28],[5],[71]). 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 ([77]). Efforts have been made to personalize these models by using individualized diffusion imaging data rather than group-average and/or tuning a small number of free-parameters to better approximate each subject's summary statistics (e.g. [63],[64][5]). However, again, these models are not directly inferred from the brain activity time-series, which could limit their ability to accurately simulate the dynamical features of these time-series. Indeed, up until this point, it has not been shown that individual-level brain models can be directly parameterized and fit from fMRI while retaining sufficient complexity to capture – and predict – whole-brain activity. 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 ([18],[19],[20]) – individualized whole-brain models are required.

In the current work, we aim to fill this gap, by advancing high-resolution characterization of the human connectome through the parameterization of nonlinear dynamical systems models that go beyond statistical correlation matrices. The models consist of hundreds of interacting neural populations, each of which is modeled as an abstracted neural mass model evolving over time-scales commensurate with fMRI. 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 rapidly fit these models directly from human resting-state fMRI. The algorithm extends data-driven techniques towards the estimation of biologically interpretable models, and conversely enables the parameterization of dynamical neural models in a data-driven, individualized fashion with relatively few priors on the dynamics within and between brain regions. Our approach represents a significant departure and alternative approach to that of previous modeling efforts, in that every parameter in our model is individually estimated without consideration of prior anatomical constraints or long-term summary statistics.

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 powerful alternative to rsFC, and even to more closely related modeling approaches, such as DCM. 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

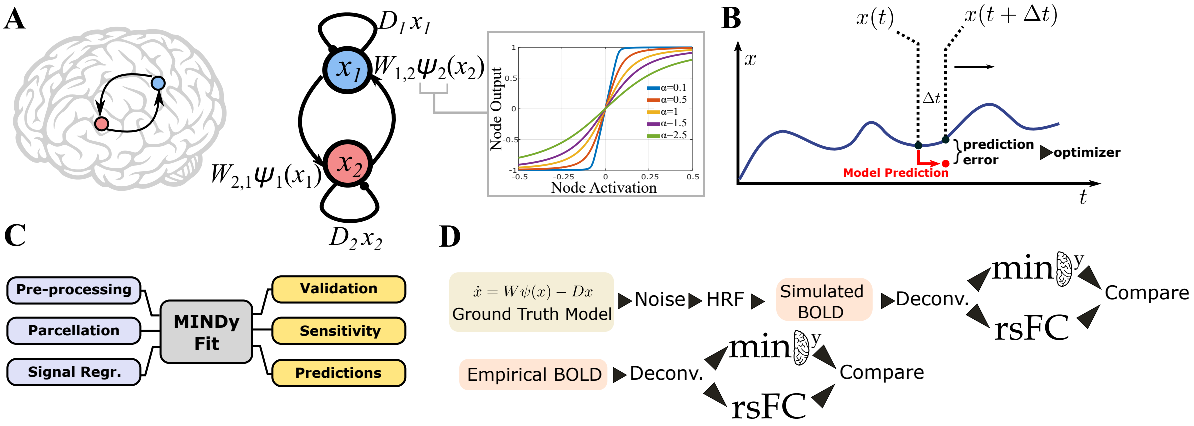
A
$D_1 x_1$
x_1
$W_{1,2}\psi_2(x_2)$
x_2
$W_{2,1}\psi_1(x_1)$
$D_2 x_2$
Node Output
Node Activation
1
0.5
0
-0.5
-1
-0.5
-0.25
0
0.25
0.5
α=0.1
α=0.5
α=1
α=1.5
α=2.5
B
$x(t)$
$x(t+\Delta t)$
x
Δt
prediction error
optimizer
Model Prediction
t
C
Pre-processing
Parcellation
Signal Regr.
MINDy Fit
Validation
Sensitivity
Predictions
D
$\dot{x} = W\psi(x) - Dx$
Ground Truth Model
Noise
HRF
Simulated BOLD
Deconv.
min y
Compare
rsFC
Empirical BOLD
Deconv.
min y
Compare
rsFC

Figure 2.1: Overview of Methods Employed. A) The MINDy model consists of coupled 1-dimensional neural-mass models (Hopfield form [1]). 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 ([2]) 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 Wiener-deconvolved ([3]) with a canonical HRF function (see Methods; [4]) before being analyzed with either MINDy or rsFC.

results that validate its utility as an analytic tool for investigating the neural mechanisms and individual differences present in fMRI data.

2.2 Methods

2.2.1 Nature of Interpretations from the Model

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

$$\dot{x} = W\psi_\alpha(x_t) - Dx_t + \epsilon_t. \tag{2.1}$$

This model, which resembles a neural mass model ([78], [1], [24], [61]) 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. 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. 2.3, Fig. 2.1 A, [79]), 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. 2.1 A). Process noise is denoted ϵ_t and is assumed to be uncorrelated between parcels. The additional parameters (α and D) reflect regional variation in intrinsic dynamics (D) and efferent signaling (α); critically, as described below, these parameters also show consistent anatomical distributions. These properties vary with brain network and are consistent even at the finer within-network scale (Fig. 2.4 A,B). Thus, our model, like a neural mass model, parameterizes both the interactions between brain regions and the processes that are local to each brain region that make it distinct.

It is important to recognize that this model is a phenomenological model in the sense that the state variables are more abstract than encountered in traditional mean-field models which combine biophysical first-principles and phenomenological approximations (e.g. the sigmoidal nonlinearity). Thus, inferences gained from the model are bounded by the inherent limitations of fMRI data (e.g. low temporal resolution and the indirectness of BOLD). The parametric form that we have chosen leads itself to interpretability. However, we stress that interpretability should not be confused with biophysical equivalence. As described in SI, there are likely many biophysical processes (including non-neuronal) contributing to each estimated parameter (A.1).

2.2.2 Robust Estimation of Individualized Neural Model Parameters

While theoretical neural mass models operate in continuous-time, fMRI experiments have limiting temporal sampling rates. Therefore, we approximate the continuous time neural model by fitting a discrete-time analogue for temporal resolution Δt (e.g. the sampling TR; Fig. 2.1 B):

$$x_{t+\Delta t} - x_t = (W\psi_\alpha(x_t) - Dx_t + \epsilon_t)\Delta t. \tag{2.2}$$

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:

$$\psi_\alpha(x_t) := \sqrt{\alpha^2 + (bx_t + .5)^2} - \sqrt{\alpha^2 + (bx_t - .5)^2}. \tag{2.3}$$

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. This form of transfer function is general enough to capture conventional choices (see SI Sec. A.2 for a derivation of the function and its relation to conventional transfer functions). Secondly, we make use of recent advances in optimization to ensure that the fitting procedure (SI Fig. A.2 C) is robust.

By using Nesterov-Accelerated Adaptive Moment Estimation (NADAM, [9]) 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. A.3 for discussion). This approach leads to a very reasonable time duration for estimation (approximately one minute on a standard laptop; see SI Sec. A.8 for a comparison with spDCM).

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 masses (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 ([80]) with the resultant fitting objective being to minimize:

$$\begin{aligned} J = \frac{1}{2} E_T[\|(X_{T+\Delta t} - X_T) - [(W_S + W_L)\psi_\alpha(X_T) - DX_T]\|_2^2] \\ + \lambda_1 \|W_S\|_1 + \lambda_2 Tr(|W_S|) + \lambda_3(\|W_1\|_1 + \|W_2\|_1) + \frac{\lambda_4}{2}\|W_L\|_2^2 \quad (2.4) \end{aligned}$$

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.

We employ this two-component weight formulation as a heuristic that facilitates high-dimensional model fitting. In most analyses we only analyze the composite weight matrix than its components. However, preliminary results indicate that properties of this decomposition, namely the ratio of sparse vs. low-rank components, may be a marker of individual differences (see SI Sec. A.6). Interestingly, recent work by Mastrogiuseppe and Ostojic ([81]) has also considered models in which connectivity is the sum of two terms: one low-rank and one random. The authors found that these structures produced low-dimensional dynamics which could be predicted based upon network structure and exogeneous (task) input. Such analyses may be relevent for understanding the role of connectivity in MINDy. Bayesian and algebraic interpretations of this penalty function are presented in SI Sec. A.5. We also discuss the well-posedness of this problem (SI Sec. A.5).

Throughout, we use the term "weights" to refer to the matrix W in estimated dynamic 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.

2.2.3 Study 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 both numerical simulations to validate the accuracy of models with respect to a known ground-truth, as well as 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).

2.2.4 Empirical Dataset

HCP Resting-State Scans

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 [82]; WU-Minn Consortium). Each subject underwent two scanning sessions on separate days. Each scan session included two 15-minute resting-state runs (two scans × two days) 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 clean segments of scanning sessions.

Preprocessing

Data were preprocessed through the rsFC pipeline proposed by Siegel and colleagues ([2]; SI Fig. A.2 A). The first stage of this pipeline is the HCP minimal pre-processing pipeline (see [82]) with FSL's ICA-FIX correction ([83],[84]). We then applied one of three second-stage pipelines developed by Siegel and colleagues ([2]; Sec. 2.3.7), 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 ([85]). Respiratory artifact was mitigated with a 40th-order .06-.14 Hz band-stop filter applied to FD and DVARS for all pipelines ([2]).

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 ([86]). 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 ([87]). Lastly, the third pipeline variant also added global signal regression (GSR; [88]), 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. For most analyses we used the full (third) pipeline, but we also compared the effects of pipeline choice (Sec. 2.3.7).

After the second-stage preprocessing pipelines, we deconvolved the parcellated data (see below) with the generic SPM hemodynamic kernel ([4]) using the Wiener deconvolution ([3]). 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. As described further below, we also tested the effect of mismatches between "true" and canonical HRF models (Sec. 2.3.7, 2.3.7).

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: $dX(t) = (X(t+2) - X(t))/2$.

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 ([89])). For cortical parcels, we generally employed the gradient-weighted Markov Random Field (gwMRF) parcellation with 200 parcels per hemisphere ([6]) and organized according to the 17 cortical networks described in [90]. 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. A.4 C,G) to ensure generality : the MMP atlas

([10]) 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.

2.2.5 MINDy Fitting Procedure

MINDy models were fit by applying the iterative NADAM algorithm ([9]) to optimize the MINDy cost-function (Eq. 2.4; see SI Sec. A.12). This algorithm belongs to the family of stochastic gradient-descent techniques and we provide further detail/discussion regarding NADAM in SI Sec. A.3. To ensure algorithmic stability, we used two transformations (one each for the curvature and decay parameters) which are detailed in SI Sec. A.12. The gradient equations for each parameter in detailed in SI Tab. A.11.

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 ([91]) as implemented by MATLAB2018a. Specifically, we fit two scalar parameters: p_W, p_D in regressing $dX(t) = p_W W\psi(X(t)) - p_D Dx$ 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. Although we performed this step using robust regression, we obtained identical results using conventional linear regression. The choice of robust regression was made as a safeguard for high leverage points as might occur due to motion artifact. However, results indicate that conventional regression may suffice for sufficiently clean data.

Selecting 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 ([9]). 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 selection criteria were to maximize cross-validated goodness-of-fit under the constraint that test-retest correlations were greater than .7 for all parameters. Regularization values were sampled with resolution .005. The chosen set of hyperparameters was then constant for all test subjects. Hyperparameter values and discussion are included in SI (Tables A.12, A.13). The initialization distributions for the algorithm were similarly selected using the same subjects and are included in the SI (Table A.12). We explored the effect of hyperparameter choices on the sparsity of MINDy relative to rsFC and found that for any choice of regularization hyperparameter (even 0), the group-average MINDy weights are sparser than rsFC (SI Sec. A.7).

2.2.6 Ground-Truth Simulations

Realistic Whole-Brain Simulations

For the analyses of sensitivity and individual differences we generated new, synthetic individuals by randomly sampling neural mass 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 $(dX = f(X)dt + \sigma_W dW_t)$ with $f(X)$ the deterministic neural mass model and units time measured in terms of the fMRI TR. Models were Euler-Maruyama integrated with $dt = 1/4$ and $\sigma_W = .45$ in units TR $(720ms)$ 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 SPM-style HRF kernel ([4]; SI Fig. A.13 C):

$$h(t; \alpha_{\{1,2\}}, \beta_{\{1,2\}}, c) := \frac{t^{\alpha_1 - 1} e^{-\beta_1 t} \beta_1^{\alpha_1}}{\Gamma(\alpha_1)} - \frac{t^{\alpha_2} e^{-\beta_2 t} \beta_2^{\alpha_2}}{c\Gamma(\alpha_2)} \tag{2.5}$$

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 double-gamma distribution. Parameters were set to their default values $(\alpha_1 = 6, \alpha_2 = 16, \beta_1 = 1, \beta_2 = 1, c = 1/6)$ except for the simulation featuring HRF variability. 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. 2.1 D).

Randomized Network Simulations

Although some ground-truth simulations leveraged the empirical MINDy distributions to maximize realism (Sec. 2.3.2, 2.3.2,2.3.7,2.3.7), others used randomly generated networks of Hopfield or neural mass models (Sec. 2.3.7, 2.3.8, SI Sec. A.8). The latter ground-truth simulations prevent circularity (i.e. using MINDy distributions to test MINDy) by drawing parameters from random hyperdistributions 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 standard-deviation parameters was randomly sampled for each ground-truth model from normal distributions: $\sigma_1, \sigma_a \sim N(4, .05^2)$ and $\sigma_2 \sim N(3, .05^2)$. Connectivity matrices were then randomly parameterized as follows:

$$
\begin{aligned}
&M_1 \sim [N(0, 1/\sigma_1^2) + N(0, 1/\sigma_1^2)^3]_{n/q \times n/q} \\
&M_2 \sim [N(0, 1/\sigma_2^2)^3]_{n \times n} \\
&M_3 \sim [N(0, 1/\sigma_1^2) + N(0, 1/\sigma_1^2)^3]_{n \times k} \times [N(0, 1/\sigma_1^2) + N(0, 1/\sigma_1^2)^3]_{k \times n}
\end{aligned}
\tag{2.6}
$$

Here, the bracket 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:$\hat{M}_1 := 1_{q\times q} \otimes M_1$. The three component matrices are then combined as follows:

$$Q = \hat{M}_1 + M_2 + M_3, \qquad \hat{Q} = (Q + (Q - Q^T)/\sigma_a) \tag{2.7}$$

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 networks of Hopfield models with homogeneous, heterogeneous, or nonlinear hemodynamic effects and realistically-paramaterized neural mass models with nonlinear hemodynamics.

Hopfield Network Simulations

We employed two cases of non-MINDy ground truths: Hopfield networks and neural-mass models (Sec. 2.3.7,2.3.8). Continuous, asymmetric Hopfield models are similar in form to the MINDy model, but use a tanh transfer function:

$$dx = \big(W tanh(b_0 \circ x) - Dx\big)dt + \sigma_W dW. \tag{2.8}$$

Here, the slope vector $b_0 \in \mathbb{R}^n \sim N(6, (.5)^2)$ and diagonal elements of the decay matrix D drawn from $N(.4, (.1)^2)$ (non-diagonal elements are zero). As elsewhere, the symbol $\circ$ 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 $N(6, (.25)^2)$ and the β_1 parameter drawn from $N(1, (.25/6)^2)$. 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 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.

Neural Mass and Windkessel-Balloon Model Simulations

For our neural mass ground-truth simulations (Sec. 2.3.7, 2.3.8), we largely followed the approach of Wang and colleagues ([71]) in using single-population neural mass models (20 masses/simulation in Sec. 2.3.7 and 6 to 16 in Sec. A.8) with Windkessel-Balloon model hemodynamics ([92])). Similar to the MINDy model, the neural mass model ([93]) contains a monotone nonlinearity ($\hat{\psi}$) and linear decay $1/\tau_S$:

$$\begin{aligned}\dot{S}_i &= -S_i/\tau_S + r(1-S_i)H(x_i) + \sigma_W dW \\ H(x_i) &= \hat{\psi}(-d(ax_i - b)) \\ x_i &= \sum_j C_{i,j} S_j + I_{Sub} \end{aligned} \tag{2.9}$$

The variable S describes the average synaptic gating, while H describes the population firing-rate. We used the default parameter settings: $\tau_S = .1s$, $a = 270n/C$, $b = 108Hz$, $d = .154s$, $r = .641$. Unlike Wang and colleagues ([71]), 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 ([92]):

$$\begin{aligned}\dot{z}_i &= S_i - \kappa z_i - \gamma(f_i - 1) \\ \dot{f}_i &= z_i \\ \tau\dot{v}_i &= f_i - v_i^{\alpha_G^{-1}} \\ \tau\dot{q}_i &= \frac{f_i}{\rho}\left[1-(1-\rho)^{1/f_i}\right] - q_i v_i^{\alpha_G^{-1}-1}\end{aligned} \tag{2.10}$$

The variables z,f,v, and q model vasodilation, inflow, blood volume, and deoxyhemoglobin content, respectively. Parameters were: $\rho = .34, \kappa = .65s^{-1}, \gamma = .41s^{-1}, \tau = .98s, \alpha_G = .32$.

The simulated BOLD signal at each TR is then modeled as:

$$BOLD(v_i, q_i) = V_0 \Big[k_1(1 - q_i) + k_2\big(1 - \frac{q_i}{v_i}\big) + k_3(1 - v_i) \Big] \tag{2.11}$$

Resting blood volume fraction is denoted $V_0 = .02$. Scanning parameters $k1, k2, k3$ were set to 3T values according to Demirtas and colleagues ([5]): $k_1 = 3.72, k_2 = .53 k_3 = .53$. Simulations were run with dt=25ms and $\sigma_W = .005$ 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 $|N(0, 1)|$. Initial conditions for the neural variable (S) were generated by first sampling $S_0 \sim |N(0, 1)|$ and then performing the transformation $S_0/(1 + S_0)$.

2.2.7 Simulations for DFC analysis

For analyses of dynamic functional connectivity, models were estimated for each subject (one per session) using the full HCP temporal resolution $dX(t) = X(t + 1) - X(t)$. These models were then used to generate simulated resting-state fMRI data, but with additional process noise added as would be expected in observed fMRI timeseries data. We used the same time-scale for simulation as in the validation models (dt=.5 TR). However, whereas the validation simulations employed process noise containing constant variance across parcels, we used a naive 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 ([6]) and removed subcortical ROIs in order to reduce computational complexity of subsequent DFC analyses.

2.2.8 DFC Analyses

DFC analyses consisted of the standard deviation and excursion ([12]) of the time-varying correlation between brain regions. To calculate time-varying correlations we used Dynamic Conditional Correlation (DCC; [94]). To avoid confusion with other references to "standard-deviation" we refer to this measure as "σ-DFC" as it pertains to time-varying correlations. Formally, σ-DFC is calculated by first estimating the time-varying covariance using DCC. Under this approach, the data, (y_t) is modeled as a zero-mean stochastic process with auto-regressive covariance:

$$y_t \sim \mathcal{N}(0, \Sigma_t) \tag{2.12}$$

with time-varying covariance matrix Σ evolving according to the first-order autoregressive model:

$$\Sigma_t = \Omega + A \circ y_{t-1}y_{t-1}^T + B \circ \Sigma_{t-1}. \tag{2.13}$$

The matrices Ω, A, B are estimated in DCC using maximum-likelihood. We define the σ-DFC matrix as the standard deviation (over time) of the time-varying correlation matrix Q_t:

$$\sigma - DFC := SD(Q_t) = \sqrt{\frac{\sum_t (Q_t - E_t[Q_t])^2}{T-1}} \tag{2.14}$$

with $E_t[Q_t]$ denoting the sample mean over time. 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 ([12]). Reliability was computed for each pair of region's DFC statistics using Fisher's ICC of group-demeaned DFC metrics between scanning session (ICC(2,1) in the Shrout and Fleiss convention [95]). Overall reliabilities collapsed across all regions were calculated using Image Intraclass Correlation ([96]).

2.2.9 Sensitivity Analyses

We conducted sensitivity analyses in Sec. 2.3.2 to test how the different mechanisms of ground-truth models (e.g. connections vs. decay) influence the estimates of "connectivity" in MINDy and rsFC. We were particularly interested in how each method responded to local heterogeneity (i.e. are MINDy/rsFC estimates of connection strength sensitive to local model parameters: decay and curvature). For each batch of 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.

2.2.10 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.

2.3 Results

2.3.1 Overview of Results/Approach

The Results of the paper are structured as follows. The first section serves to relate MINDy parameter estimates to resting-state Functional Connectivity (and related partial correlation approaches) in terms of differentiating/identifying sources of individual variation. The "ground-truth" models for validation in this first set of analyses are drawn from the empirical distribution of MINDy parameters to ensure that the resultant simulated data is realistic. The second section directly addresses the potential for overfitting by testing whether MINDy models cross-validate and whether parameters are reliable. The third section demonstrates that MINDy parameters have distinct anatomical gradients consistent with previous, theoretical results ([5],[71]), and highly conserved individual variation (a feature not present in over-fit models). The fourth section demonstrates models' predictive validity by reproducing individual differences in resting-state dynamics using the empirical models. In the fifth section, we demonstrate that the approach is robust to measurement noise, preprocessing pipelines, and hemodynamic confounds. This section uses three forms of "ground-truth" models. For initially testing robustness to noise and global hemodynamic variability, we again use parameters drawn from the empirical distribution to ensure maximum realism. In subsequent analyses, however, "ground-truth" parameter values are drawn from random hyper-distributions independent of the data and combined with more nuanced hemodynamics. This step tests model performance with more exotic "ground-truths" and prevents circularity. We also consider an additional case in which the simulated fMRI data is generated from randomly-parameterized neural-mass models (operating at the millisecond-scale) to provide insight into the relationship/limitations of MINDy parameter estimates from fMRI and the underlying synaptic connectivity. In the sixth section (Sec. 2.3.8), we summarize comparisons with Dynamic Causal Modeling which receive fuller treatment in the SI (Sec. A.8). The final results section directly assesses data-requirements of MINDy and provides a minimum data quantity (>15 minutes) to prevent over-fitting.

2.3.2 MINDy Retrieves Individual Differences

MINDy Retrieves Individualized 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 ([97], [47]). 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. 2.1 C; SI Fig. A.2 B). 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. 2.2 B) and the sensitivity of each procedure to different model components (SI Fig. A.10 A). 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 ground-truth weight matrix for each individual (Fig. 2.2 A,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 and partial correlation measures (computed on the simulated timeseries data) in their ability to accurately recover both the ground-truth connectivity matrix of simulated individual subjects, as well as the differences between individuals (Fig. 2.2 B; SI Table A.5). 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, rsFC may be disadvantaged in this comparison as it does not typically permit sparseness commensurate

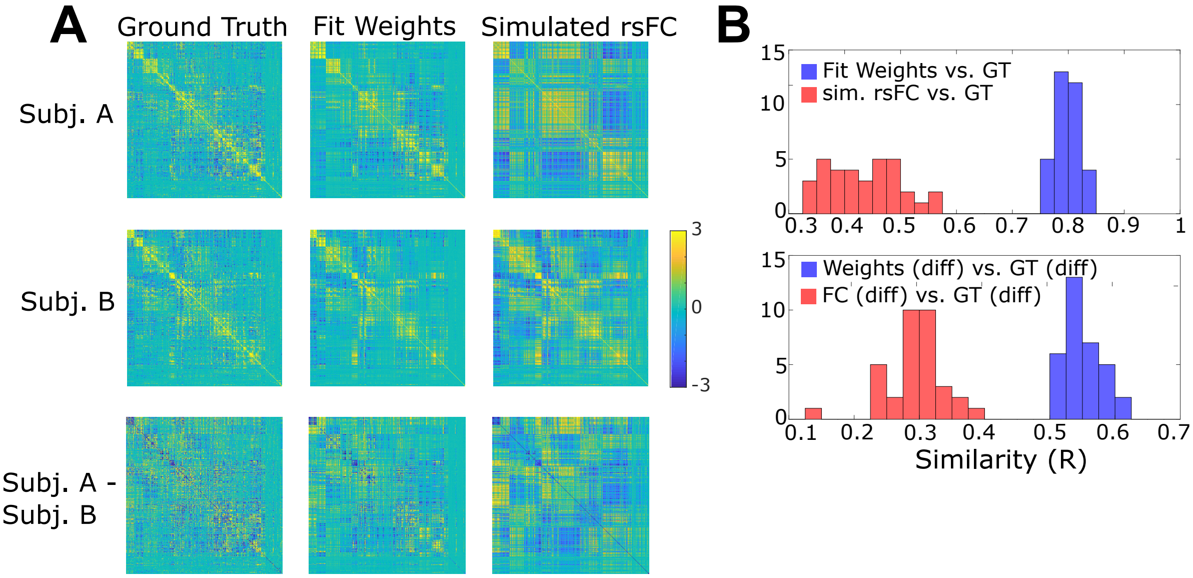

Figure 2.2: Ground-truth validation of MINDy and rsFC at the level of single-subject and inter-subject variation. A) First column: Example ground-truth weight matrices for two simulated subjects (top two rows) and the difference between ground-truth weights (bottom-row). Second column: Recovered weight 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 offdiagonal 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 simulated single-subject level (correlation with ground-truth [GT]) for MINDy (blue) and rsFC (red). Bottom row: same as top but for for predicting the differences in matched-pairs of simulation subjects who differed only in ground-truth connectivity. Simulation subjects were generated by sampling from the distribution of empirical (HCP) MINDy parameters (see Sec. 2.2.6).

with empirical MINDy weights (Fig. 2.5 A,B). Therefore, we used partial correlations as an additional benchmark. While partial correlations quantitatively improved upon rsFC estimates (single-subject: $R = .537 \pm .032$, inter-subject: $R = .392 \pm .027$), performance remained significantly lower than MINDy (single-subject: $paired - t(33) = 40.51, p \approx 0$, inter-subject: $paired - t(33) = 23.62, p \approx 0$).

The above analyses were designed to illustrate the additional utility of MINDy in empiricial contexts over the most common current approaches (rsFC and partial correlation). For this reason, we generated ground-truths from the empirical distributions to ensure maximal realism. In later analyses (Sec. 2.3.8), we compare MINDy to a much closer modeling approach (Spectral DCM; [57]). We reserve these comparisons for later as they employ a very different approach to generating ground-truth models: seeking to minimize bias and sample over a wide range of potential ground-truth scenarios. The anatomically-detailed models used in the current section are also too large for Spectral DCM to estimate using available computational resources (Sec. 2.3.8).

MINDy Disentangles Sources of Individual Differences

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 changed in strength between simulated models, when in reality only the curvature or decay terms were altered (SI Fig. A.10 A). Results indicate that MINDy correctly detects the sources of individual variation when due to local changes such as decay rate and transfer function shape, 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 ($27.5 \pm 12\%$ of connections changed vs. $7.6 \pm .6\%$ for MINDy, with 5% expected by chance), indicating that some individual differences in FC may be reflective of purely local brain differences as opposed to connectivity between brain regions (SI Fig. A.10 A; SI

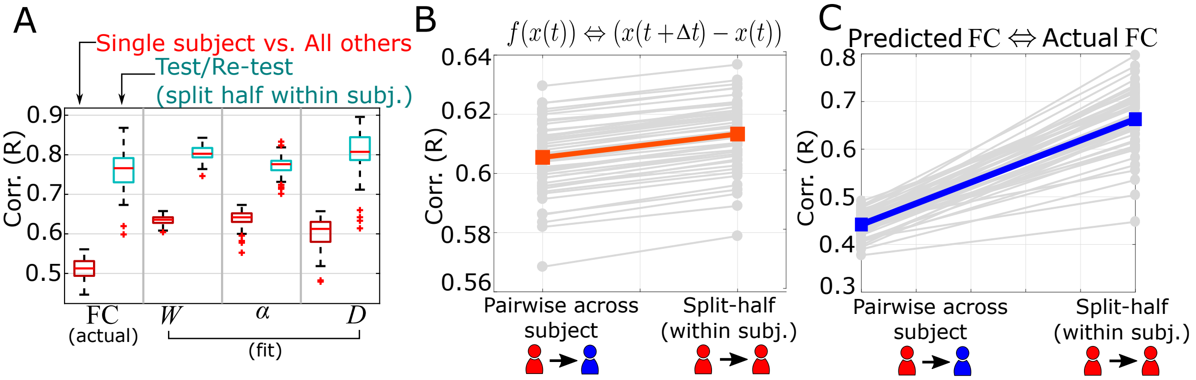

Figure 2.3: MINDy parameters and predictions are personalized and reliable. A) Comparison of the test-retest similarity between subjects (red) and the test-retest reliability (blue) for rsFC and the MINDy parameters. B) Goodness-of-fit for a single time-step prediction is uniformly (but minutely) greater 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. A.9. Blue line indicates mean.

Table A.6). 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. However, it is still the case that resting-state fMRI exhibits *generalized* sensitivity to individual differences in neurobiology, which may suffice for some applications, such as biomarker discovery (see Sec. 2.3.4).

2.3.3 MINDy Parameters are Reliable

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 ($> .75$; Fig. 2.3 A). The reliability of weight estimates was significantly higher than rsFC reliability, although the mean difference was modest ($\Delta R \approx .045$, SI Table A.7, SI Table A.8). 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 A.8).

2.3.4 MINDy Parameters are Personalized

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 (Fig. 2.5A,B) 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 ([42]), 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. 2.3 A). 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. 2.3 B) and for predictions of each subject's empirical rsFC, both cross-validated across sessions (Fig. 2.3 C). In all analyses, we found that the best predicting model for every subject was almost always their previously fit model (Table A.7). In particular, we achieved 100% accuracy when conducting connectome fingerprinting based on MINDy weight parameters (SI Fig. A.10 B), and when computing cross-validated goodness of fit/cross-validated predicted rsFC (Fig. 2.3 B,C). For pairwise analyses of subjects, see SI Fig. (A.10 F).

Similar patterns emerged but also some important differences, when conducting parallel analyses using rsFC. Replicating prior findings ([42]), 100% accuracy was also achieved in connectome fingerprinting (SI Fig. A.10E). 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 A.7). These results suggest that rsFC may actually generate an exaggerated picture of the idiosyncratic nature of connectivity, since 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 $[\alpha]$ 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).

2.3.5 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 (Fig. 2.3). 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 (Fig. 2.3 A; Table A.7). Pair-wise, between-subject, comparisons of similarity in these parameters are reported in SI Fig. (A.10 B-E).

We followed-up on the identification of reliable individual differences through MINDy, by conducting exploratory analyses to examine which brain regions/connections exhibited the greatest inter-individual variability (SI Sec. A.9). We found that the curvature parameter had greatest relative variability in prefrontal cortex, particularly inferior frontal gyrus (SI Fig. A.6A), while the decay parameter had high variability in visual regions, the “hand” portion of post-central gyrus, and medial prefrontal cortex (SI Fig. A.6B). Connections within the visual networks had the lowest individual variability while connections to/from

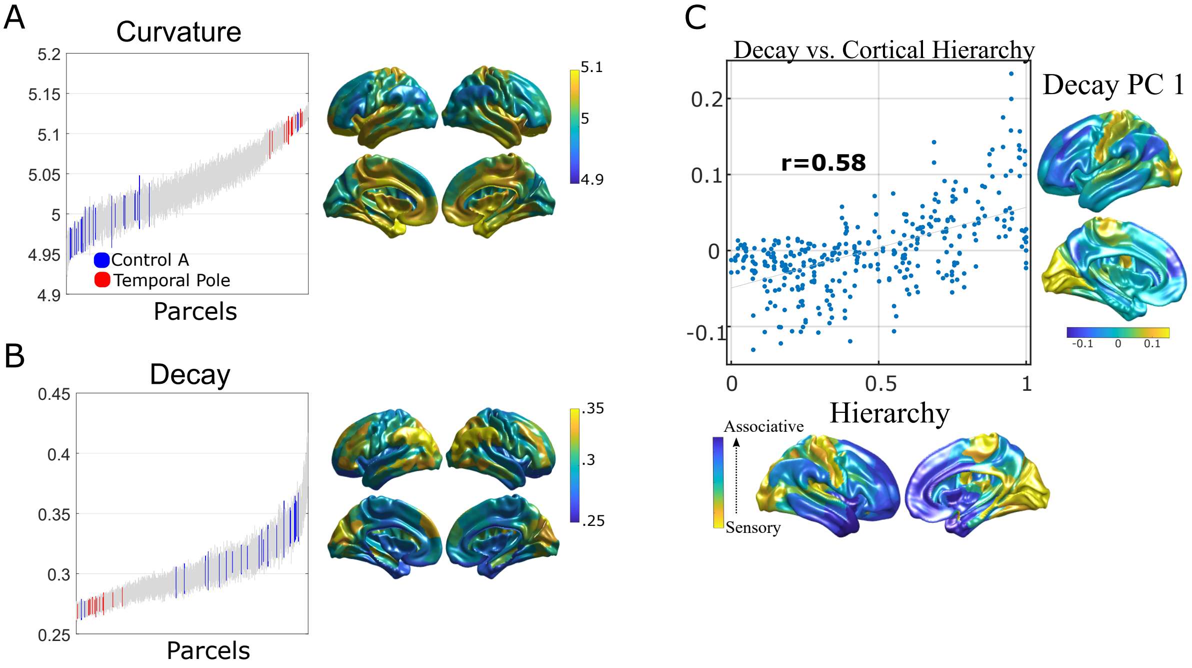

Figure 2.4: Local MINDy parameters display consistent anatomical distributions. A) The curvature-parameter displays network structure and is consistent across subjects at the finer parcel level. Parcels are ordered from least to greatest value for the curvature parameter (α) averaged across subjects and scanning sessions. Surface plots are for mean value. Two representative brain networks are highlighted (Control-A in red and Limbic-(Temporal Pole) in blue) to illustrate anatomical gradients in this parameter. B) Same as A but for the decay parameter (D). C) Correlation between the first principal component of MINDy decay and "hierarchical heterogeneity" provided by Demirtas and colleagues ([5]) based upon erf transform of the T1/T2 ratio (MMP parcelation). This measure has been theorized to reflect a hierarchy of cognitive abstraction from sensory to associative cortices.

the Temporal-Parietal network had the greatest (SI Fig. A.6C, D). Although these initial findings are intriguing, due to sample size/bias considerations and the exploratory nature of these analyses, we view them as a launching pad for future insights rather than basic neuroscientific results per se (see SI Sec. A.9).

Although the above analyses focused on individual differences in the unique MINDy parameters, these parameters also exhibited common patterns across individuals (SI Fig. A.4 B,F) that revealed interesting anatomical structure and gradients (Fig. 2.4 A-C; SI Fig. A.4 B,F). These may reflect regional variation in intrinsic dynamics (D) and efferent signaling (α) that vary across brain networks (SI Fig. A.4 A,E), but also exhibit consistency even at the finer within-network scale (Fig. 2.4 A,B; SI Fig. A.4 B,F). For example, most nodes within the Temporal-Parietal network showed high curvature, but also low decay parameters; in contrast, in nodes of the Control (A) network, the curvature parameter tended to be low, whereas the decay parameter was high. Group-mean values show the same anatomical gradient across the gwMRF ([6]; SI Fig. A.4 D,H) and MMP ([10]; SI Fig. A.4 C,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, 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; $r(377) = .911, p \approx 0$) the principal dimensions of individual variation (Fig. 2.4 C, SI Fig. A.4 I,J) recreate the hierarchical organization of primate cortex as derived from the T1/T2 ratio map ($r(358) = .583, p \approx 0$; using the MMP Hierarchy map by Demirtas and colleagues [5]). As a caveat, it is worth noting that these statistics do not take into account spatial autocorrelation (which is challenging to model, given the large and irregular shape of parcels), which could have contributed in part to the anatomical gradients we observed. This hierarchy has been the subject of recent studies into its relationship with local excitation/inhibition ([71], [5]) which is one physiological mechanism we suspect underlies the decay construct (see SI A.1). This hierarchy also predicts the time-scales of local microcircuits, patterns of gene-expression, mylein density, and function (sensory-processing hierarchy; see [71], [5]).

In addition to the curvature and decay parameters, MINDy also differentiates from rsFC in the structure of the weight matrix (W) / connectivity matrix, both in terms of asymmetry (Fig. 2.5A,C) and sparseness (Fig. 2.5A,B). The former is a direct consequence of the dynamical systems model that underlies MINDy, which provides an estimate of effective connectivity. Although regularization generally favors sparse solutions, we found that, even without any regularization, the group average Weight matrix was much sparser than rsFC (SI Sec. A.7). We provide a simple proof-of-concept to illustrate the potential insights that can be gained from investigating such asymmetries. Specifically, 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. 2.5 C). Excitatory connections of the left IFG with temporal cortex are essential features of the "language network" (e.g. [7]). Additional results revealing other brain regions showing directionality biases in connectivity are reported in SI (Sec. A.10). In a later section, we explicitly test

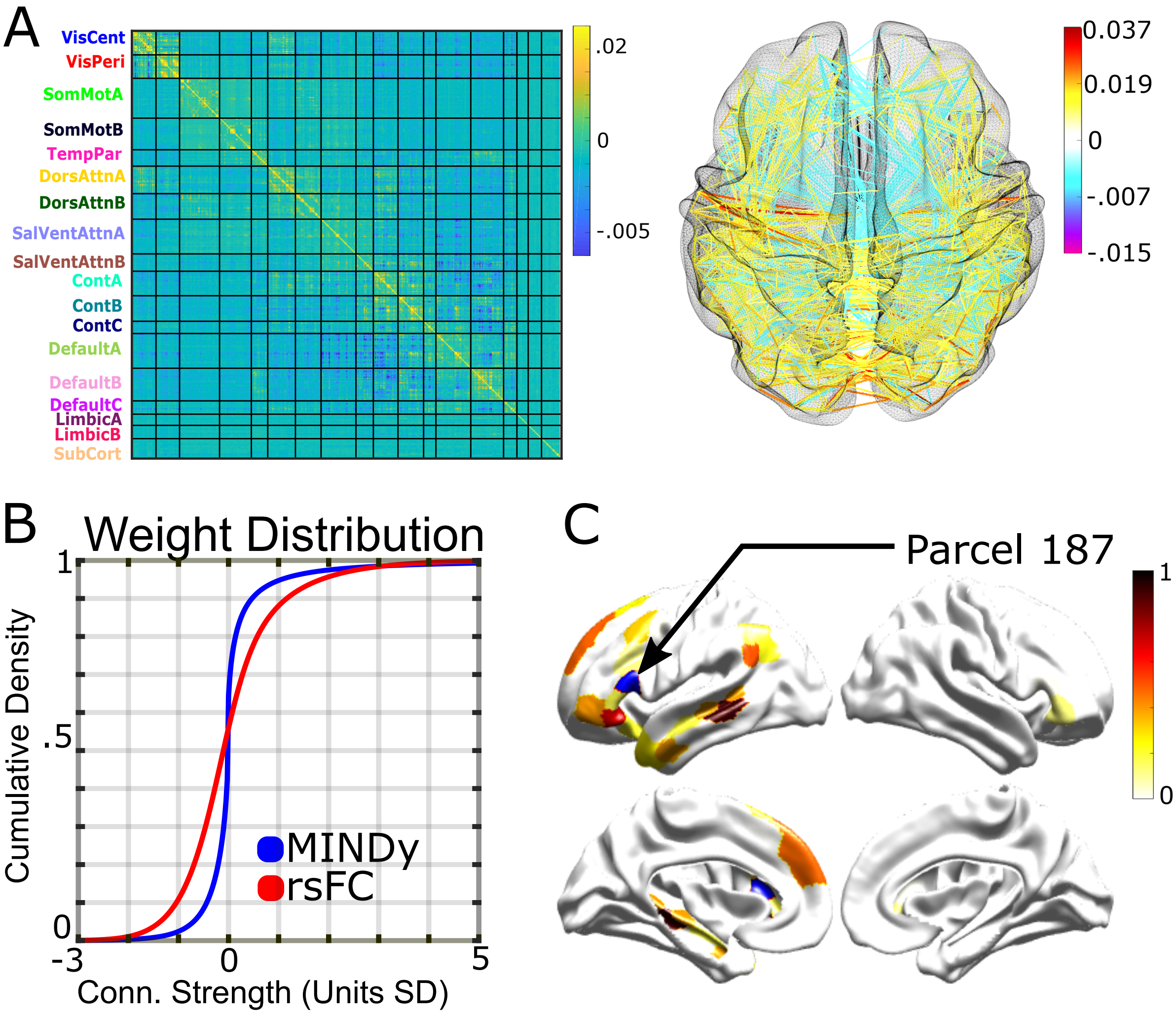

Figure 2.5: MINDy weights are structured, sparse, and directed. A) Left-side: Mean connection matrix W averaged across subjects and scanning session. Parcels are grouped according to the Schaeffer ([6]) 17-network parcellation (hemispheres combined) plus the freesurfer subcorticals. Right-side: thresholded anatomical projection (positive connections$\geq$ 20% max non-recurrent magnitude and negative connections with magnitude$\geq$ 8%). B) The MINDy weight distribution demonstrates sparser connectivity than rsFC. C) Parcel 187 ([6] 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 (e.g. [7]) see SI Sec. A.10 for additional, preliminary directed-connectivity results. Blue highlights chosen source-parcel.

the robustness of asymmetry estimates and how they are affected by assumptions regarding hemodynamics and model mismatch.

2.3.6 MINDy Predicts Individual Brain Dynamics

MINDy Predicts Individual Differences in Dynamic Functional Connectivity

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 ([33],[39],[25], [35]). 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 and C in Fig. 2.3, which reflect short and long-term predictions, respectively). 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; [94]), a method which has been recently shown to improve reliability in the HCP data-set as compared to sliding-window estimates ([11]). 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 ([11]) and excursion ([12]) are more reliable than state-based descriptions for the HCP resting-state data ([11]). Therefore, we used these measures (see Sec. 2.2.8 for equations) 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 of 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 ([11]). MINDy performed slightly better on recreating another reliable DFC statistic, 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 (Fig. A.11).

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. A.11 A,B). The image intraclass correlation (I2C2, [96]) for the model was .555 for σ-DFC and .481 for excursion. In the original experimental data, I2C2 reliabilities 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. 2.6 A). Lastly, we analyzed whether the simulated data recreates the central tendency of observed data. In general, the group-mean σ-DFC (SI Fig. A.11D) and excursion (SI Fig. A.11 E) estimates were highly similar between the simulated and observed data for both the σ-DFC (Fig. 2.6 B; $r(4948) = .761$) and excursion metrics (SI Fig. A.11 C; $r(4948) = .836$). Thus, MINDy models recreate measures of DFC at the level of both individual differences and the group-level. Moreover, in some cases (e.g. the excursion metric), MINDy models generate more reliable estimates than those of the original data (SI Fig. A.11B). A main advantage of the model in this regard is likely due to the ability to simulate an arbitrarily large amount of data with the model that is also free from nuisance signals/motion.

MINDy Models Generate Non-Trivial Dynamics

In the previous section we demonstrated that MINDy predicts individual differences in nonstationary dynamics. This finding suggests that the nonlinearities in MINDy are able to account for some features of the data (nonstationarity) that are mathematically absent from linear models. From a dynamics perspective, non-pathological (Schur-stable) linear models predict that spontaneous brain activity consists of noise-driven fluctuations about a single equilibrium. The model parameters for a linear system (e.g. "effective connectivity" in DCM) shape the spatiotemporal statistics of these fluctuations and in the case of white-noise excitation result in a unimodal distribution about the equilibrium in question. Although many nonlinear systems exhibit exotic behavior (e.g. chaos), some systems are dominated by a single equilibrium and may thus possess dynamics that are similar to a linear system. Therefore, we tested whether empirical MINDy models exhibit nontrivial dynamics in the absence of noise (see SI Sec. A.11). We found that all subjects' models were dominated by nontrivial dynamics (multistability, homo/heteroclinic cycles, limit cycles, etc.). Example nonlinear dynamics for two representative subjects are provided (Fig. 2.6C,D), although a thorough characterization of each model's full phase space is beyond our current scope (see SI Sec. A.11). Nonetheless, we were able to formally demonstrate that no subject exhibits trivial dynamics (SI Fig. A.8A,B; Proposition 2). We conclude that the nonlinearity of MINDy models is not superficial, but rather generates topologically significant dynamics which shape model behavior.

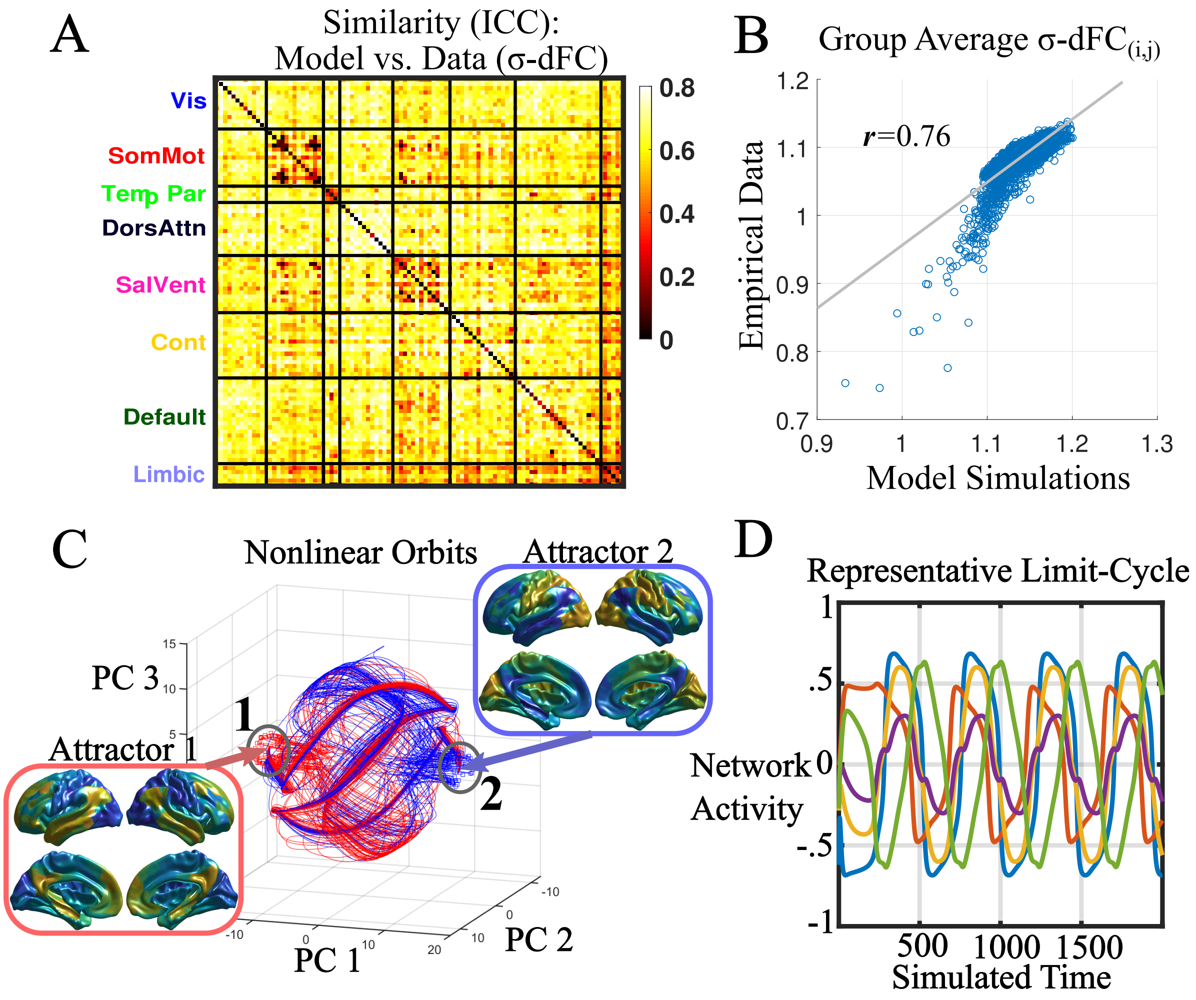

Figure 2.6: MINDy models predict individual variation and central tendency of pairwise dynamic functional connectivity (DFC) measures. A) Similarity between model and data for predicting each subject's σ-DFC for each pair of brain regions (using the 100-parcel atlas from [6] and collapsing the 17-network grouping down to 8). B) Scatterplot and Pearson correlation of group-average σ-DFC for data vs. model. C) Evidence of non-trivial dynamics in MINDy models. Example phase portrait of one subject projected onto the first 3 principal components. Complex orbits link neighborhoods of attractors 1 and 2 (orbits starting from each neighborhood are colored red, blue respectively). Inset figures show these attractors projected onto the brain. D) Example deterministic time-series for a limit cycle in a different representative subject averaged across five networks (Visual [blue], SomMot. [red], Dorsal Attn. [yellow], Salience [purple], and Control [green]). These deterministic dynamics demonstrate significant nonlinearity but are qualitatively different from the simulated model dynamics (e.g. for computing DFC) which include process noise.

2.3.7 MINDy is Robust

MINDy is Robust to Measurement Noise

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 systematically varying the amount of measurement noise added at each time-step. This approach allowed us determine how strongly these various sources of noise impacted the ability of MINDy to recover the ground-truth parameters. Results indicated that although the performance of MINDy decreased with the amount of noise added (Fig. 2.7 A), similarity to the ground-truth values generally remained high. Additional levels of noise are plotted in SI Fig. A.13. 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 empirical data exhibiting such a high level of noise would (hopefully) fail quality control.

MINDy is Robust to Individual Differences in Motion

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 absolute framewise-displacement of the subjects head across scanning sessions, and 3) the spatial standard deviation of temporal difference images (DVARS) ([86]). 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. A.12 B-D, I). A parallel analysis used cross-validated fit, in which MINDy parameters were estimated from one session, and then used to predict 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 motion 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 in the held-out session). In fact, the cross-validated fits were relatively similar in each group (SI Fig. A.12 E,J). Together, these results suggest that participant motion (within a reasonable range) may not be strong factor in determining how well MINDy model parameters can be estimated from observed fMRI data timeseries.

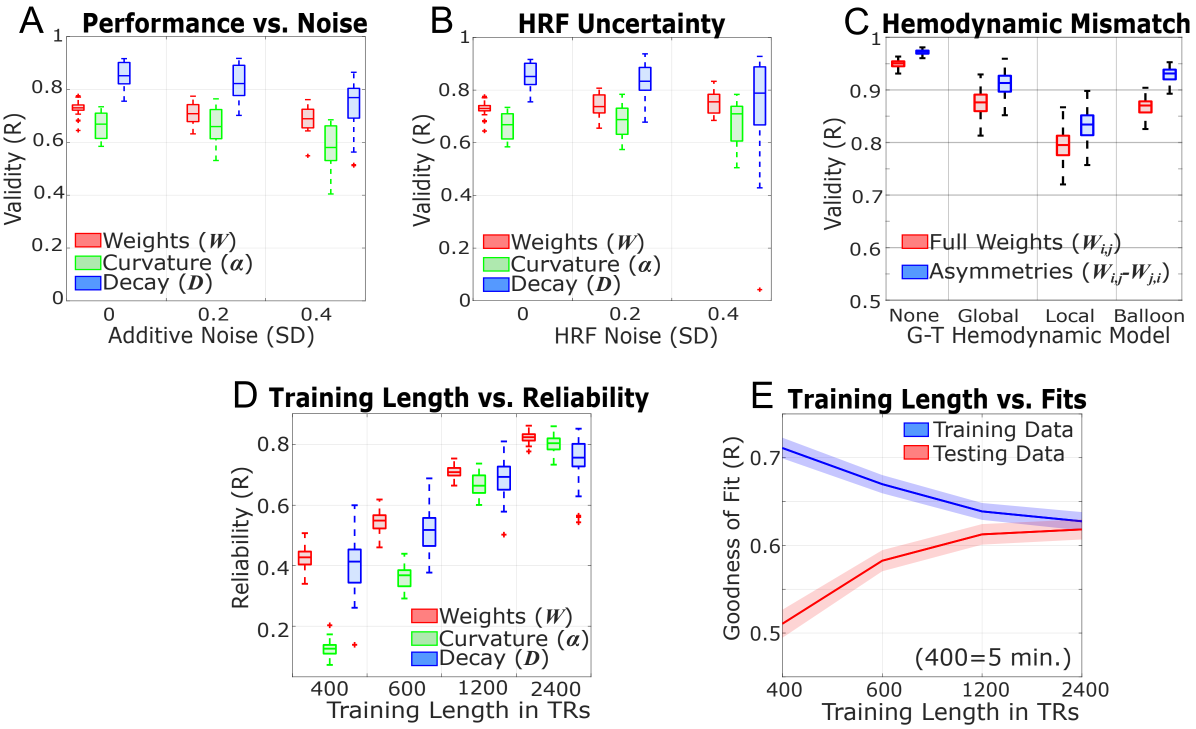

Figure 2.7: A) Increasing the amount of additive measurement noise slightly decreases MINDy performance in recovering ground-truth parameters. B) Mean performance is unaffected by the uncertainty in the ground-truth HRF, although performance does become more variable (see SI Fig. A.13H). C) MINDy performance in retrieving the ground-truth weight matrix (original matrix: red; asymmetric part $(W_{i,j} - W_{j,i})$: 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.

MINDy is Robust to Pre-Processing Pipelines

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 ([2]), one with motion-correction only, one which adds to this CompCor (a standard method that removes noise components associated with white matter and CSF; [87]), and a final, full variant that additionally includes global signal regression (GSR; [98]). We compared test-retest reliability for data-processed with each pipeline (SI Fig. A.14 A) and the similarity of parameter estimates obtained when the same data were processed using different pipelines (SI Fig. A.14 B). Results indicated that MINDy parameters had high test-retest reliability regardless of preprocessing choices (all $R > .7$, SI Fig. A.14 A) and that similar parameter estimates are obtained regardless of preprocessing choices (all $r > .85$, SI Fig. A.14 B). 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 pipelines (all $p's < .001$; Table A.8; SI Fig. A.14 A), and were less impacted by changing preprocessing pipelines (all $p's < .001$; SI Fig. A.14 B). Together, this set of analyses indicate that the choice of preprocessing pipeline will not have a large effect on estimated MINDy parameters.

MINDy is Robust to Global Hemodynamics

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 ((although we have recently begun to explore the effect of relaxing this assumption, and estimating a different HRF for each parcel and individual; [99]). Other fMRI models also assume a canonical HRF (e.g. regression-DCM; [76]). However, existing literature suggests that this assumption is likely to be incorrect ([88], [100]). To examine the impact of mis-fitting the HRF, we modeled a variety of ground-truth scenarios. The first set of ground-truth simulations were randomly parameterized according to the empirical MINDy distribution and activity timeseries were convolved with spatially homogeneous, but randomly parameterized HRFs with incrementally greater variability (SI Fig. A.13 D). We then attempted to recover MINDy parameters while again assuming the fixed, canonical HRF model ([4]). 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-value$\approx$0.75, Fig. 2.7 B). However, the variability of the fits increased across simulations, as the HRF became more variable across regions and individuals (SI Fig. A.13 H).

MINDy Parameters are Robust to Model Mismatch

We also considered the effect of violations of the MINDy model in terms of the underlying neural models (MINDy vs. Hopfield, neural mass) and neurovasculature (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 the next set of simlations, we generated complex networks from a non-empirical hyperdistribution whose characteristic parameters were randomly sampled at each run. This approach allowed us to sample over a wide range of qualitatively different network structures (Sec. 2.2.6) and these simulations did not depend upon previous empirically-fit MINDy models. We tested the ability of MINDy to recover the weight parameter (Fig. 2.7 C) from a simple rate model (tanh transfer function) 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 ([92], [5], Sec. 2.2.6). 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. 2.7 C; SI Table A.9). 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 ($r = .910.023$, $n = 1600$). Spatial heterogeneity of the HRF decreased MINDy performance in recovering overall ground-truth connectivity ($r = .793 \pm .029$; $t(3071.8) = -86.72$, $p \approx 0$; unequal-variance), but did not differentially impair the estimation of asymmetries ($r = .832 \pm .028; t(3057.7) = 11.74$, $p \approx 1$, 1-tailed, unequal variance).

We also found 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 was used

($t(3073.3) = 23.03$, $p \approx 0$; unequal variance). Thus, violations of spatial homogeneity in the hemodynamic response appear more relevant to MINDy than violations of hemodynamic linearity. However, performance was still strong in all cases considered (median $r \leq .80$). We also conducted preliminary tests of MINDy's ability to recover synaptic conductances (weights) from the simulated BOLD signal (Balloon-Windkessel) of a biophysically parameterized neural mass model ([93]) 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 ($paired - t(1399) = -109.172$, $p \approx 0$). 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.

2.3.8 Comparing MINDy with Dynamic Causal Modeling

The earlier analyses, in which we compared MINDy and rsFC (Sec. 2.3.2), serve to demonstrate the potential linkages between methods and how MINDy can resolve some ambiguities inherent in rsFC (e.g., directionality). However, these analyses should not be interpreted as stating that MINDy is unambiguously "better" than rsFC as the two approaches represent fundamentally different constructs. The correlation matrix (rsFC) is a statistical quantification whereas MINDy is an approach for estimating a dynamical-systems model and they may have complementary roles for exploring individual differences/biomarker discovery. In order to benchmark MINDy as a model-fitting technique we compared performance with spectral DCM

([57]) in recovering connectivity weights for a variety of simulated ground-truth scenarios. Spectral DCM (spDCM) is a recently developed Dynamic Causal Modeling (DCM) approach for simultaneously estimating linear dynamical systems and (region-specific) hemodynamic kernels from resting-state fMRI ([57]). To be clear, we view the primary contributions of MINDy relative to modeling approaches such as spDCM in terms of its scalability, biological interpretability, and the ability to predict nonstationary resting-state dynamics. However, the question remains whether these advantages come at the expense of accuracy—i.e. whether MINDy is inferior to DCM within the latter's scope.

We compared performance of MINDy and spDCM across a variety of ground-truth scenarios (see SI Sec. A.8) to test whether MINDy performs at least as well as spDCM in the lower-dimensional scenarios in which the latter is applicable (i.e., estimating parameters for a small number of nodes or neural masses). These simulations were specifically designed to reduce bias based upon either model's assumptions (see SI Tab. A.1) and considered ground-truths based upon mesoscale Hopfield-style models (SI Fig. A.5A) and biophysical neural mass models (SI Fig. A.5B). In the former case, we manipulated the degree of spatial variability in the hemodynamic response (SI Fig. A.5C). When arbitrary choices were necessary, we chose the option than empirically favored spDCM. Results support that MINDy's advantages do not come at a cost to accuracy. In all settings considered, MINDy was at least as accurate, on average, as spDCM and significantly (orders of magnitude) faster. We observed that spDCM was more robust than MINDy to spatial variability in the ground-truth HRF (although see extensions in [99]), but even under the most extreme cases considered, MINDy was at least as accurate as spDCM (SI Fig. A.5). The empirical examination of run-time overwhelmingly favored MINDy (SI Fig. A.5D-F). For example, the largest network we tested contained 16 neural masses (SI Fig. A.5D) for which MINDy estimation took 3.5s on average vs. 2.7 hours for spDCM. We estimate that fitting spDCM models using our chosen parcellation, involving 419 brain regions/nodes (400 of which are cortical [6]) would take a minimum of 44 years each (and likely much longer; see SI Sec. A.8). We conclude that MINDy's advantages (scalability, dynamics etc.) do not come at the expense of accuracy relative contemporary approaches.

2.3.9 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. 2.7 D). 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. 2.7 E). 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 ([97],[47]). 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).

2.4 Discussion

2.4.1 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. 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 ([61]). 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 ([64]).

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. A.10 A). Conversely, we have demonstrated that both the transfer function and decay rate of brain regions can also serve reliable markers of individual differences and anatomical structure. By using MINDy, researchers can identify which neural mechanisms (i.e. which of MINDy's parameters) give rise to individual differences of interest.

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. 2.5 C, SI Fig. A.7). Further, we demonstrated that MINDy may prove a more valid measure of brain connectivity and individual differences in connectivity than rsFC (Fig. 2.2 E).

2.4.2 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 [59], [64]) and directly-parameterized linear models (e.g. Dynamic Causal Modeling [52], [57]). Although nonlinear variants of DCM have been proposed for task-fMRI ([101]), the techniques used for resting-state fMRI (e.g. [57]) are fundamentally linked with the statistics of linear systems. These two approaches (proxy-parameterized neural-masses and linear models) 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, since 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, in the manner by which these models are parameterized. Even in state-of-the-art techniques (e.g. [5],[71]) 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, typically 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 FC, [28], [5], [71]).

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 ([102]) despite using a simplified 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 their short-term predictions (by virtue of fitting parameters), the linear form guarantees that they will be unable to capture the extended pattern of brain spatio-temporal 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 ([36], [103]). 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 takes a reduced (linear) 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. [57]). 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. A.10 A).

MINDy attempts to leverage the advantages from both approaches. Like current neural-mass models, MINDy employs a nonlinear dynamical systems model 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 functional 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 ([5], [71]) to fitting every parameter in biologically-plausible individualized 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 ([57],[76]) 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 ([71],[57]; see SI Sec. A.8) 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. 2.3.2).

2.4.3 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 ([28],[5],[71]). Two recent studies fit free parameters with the explicit optimization objective of predicting the group-average rsFC matrix ([5],[71]). 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. 2.3 C). 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. A.9). As such, the group-average MINDy fit compares very favorably with the analogous measures for diffusion-parameterized models which typically don't surpass $r = .6$ ([26],[5],[71]).

2.4.4 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. Thus, MINDy cannot describe relationships which, depending upon local activity, change sign from net excitatory to net inhibitory. 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 ([61]), MINDy employs a single population rather than two or more local subpopulations of excitatory and inhibitory neurons. 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 A.1).

2.4.5 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.[93]), 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.

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 ([46]) and intrinsic network structure ([43]). 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. In forthcoming work, we have been using MINDy to estimate task-related activity in this manner, and the initial results strongly indicate that this approach improves the statistical power and temporal specificity of estimated neural events ([104]). Thus, MINDy has the potential to improve estimates of task-evoked activity from fMRI data, although future validation is needed.

Illuminating Dynamics

Present results indicate that MINDy is able to replicate some features of infraslow brain-dynamics observed in the data (see Sec. 2.3.6). 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 ([105],[106]) and daydreaming ([107]). 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 ([49],[50],[67]), 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.

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: https://github.com/singhmf/MINDy-Beta.

Chapter 3

Scalable Surrogate Deconvolution for Identification of Partially-Observable Systems and Brain Modeling [2]

3.1 Introduction

A key challenge in neural engineering pertains to estimating neural model parameters from indirect observations that are temporally convolved from source measurements. For example, many imaging modalities reflect convolution of neural activity with temporal kernels associated with slower physiological processes such as blood flow (Fig. 3.1A), or molecular concentrations (Tab. 3.1). Often, these kernels are not known, necessitating so-called ‘dual estimation’ of both the latent neural activity and the neural model (including convolutional kernels) at the same time. Our paper presents a computational framework for addressing this problem. Specifically, we assume that the system can be described in the following form (or its discrete-time equivalent):

---

[2] Chapter reprinted verbatim from previously published work: [99]

$$\dot{x} = f(\theta, x, \hat{z}) + \varepsilon(t) \tag{3.1}$$

$$\hat{z}_i(t) = [h_i(\eta_i) * x_i](t) \tag{3.2}$$

$$z_i(t) = \hat{z}_i + \nu_i(t) = [h_i(\eta_i) * x_i](t) + \nu_i(t) \tag{3.3}$$

Here, $x \in \mathbb{R}^n$ are the hidden non-convolutional state variables and $\hat{z}$ are the physiological variables generated by convolution. We denote unknown parameters for the non-convolutional plant as $\theta \in \mathbb{R}^q$ and for the convolutional plant as $\eta_i \in \mathbb{R}^{v_i}$. Each parameterized kernel (h_i) represents the process generating the corresponding measurable variable z_i via convolution (denoted $*$). This formulation requires the assumption that these processes may be well-approximated by a finite-impulse response function and that structural priors may be placed on each kernel (i.e. h_i is known up to a small number of parameters: η_i; see Sec. 3.5.2 for discussion). We denote process noise in the hidden state variables by $\varepsilon(t) \in \mathbb{R}^n$) and denote measurement noise $\nu_i(t)$, both of which we assume to be drawn from stationary distributions, independently realized at each time step (noise is not auto-correlated). In the current context, x represents latent neural state-variables. The measurements z_i are multi-dimensional recordings of neural data and $\hat{z}$ are the corresponding physiological sources. These sources can either feed back into the latent system (e.g. Ca^{2+} concentration) or be modeled as purely downstream (as is typical for BOLD). Formally, we seek to estimate the convolutional kernel parameters $\{\eta\}$ and the neural model parameters θ using the measurements z (i.e., the 'dual' estimation). This problem formulation is highly relevant to neuroscience and neural engineering since it would enable inferences regarding brain activity via indirect and uncertain dynamical transformations.

3.1.1 Relevance to Neuroscience and Neural Engineering

Whereas many neural models emphasize membrane potentials, channel conductances, and/or firing rate as state variables, high-coverage measurements often consist of the extracellular ("local") field potential, concentration of signaling molecules (e.g. Ca^{2+}), blood-oxygenation (and the derived BOLD-fMRI contrast) or radionuclide concentrations (e.g. PET). In all of these cases, the measurements reflect downstream, temporally extended consequences of neuronal firing (Tab. 3.1). Thus, in the context of neuroscience, the dual-problem consists of simultaneously estimating the parameters of neural systems, while inverting measured signals into their unmeasured neural generators (the state-variables specified by a given model framework). Often this linkage (from generator to measurement) is assumed to be a linear time-invariant (LTI) system so that the relationship can be described via convolution with parameterized kernels. For, example, post-synaptic currents are often modeled via synaptic "kernels" (e.g. "alpha-synapse", [108]), kernels for molecular concentrations (e.g. Ca^{2+}, [109],[110]) are derived from Markovian kinetic-schemes ([111],[109]), and the neurovascular coupling kernel (linking BOLD-fMRI and neural "activity") is described by a Hemodynamic-Response Function (HRF, [4]; Fig. 3.1A,B,). If these functions are assumed fixed, it may be possible to estimate the neural state-variables via deconvolution, in which case, conventional modeling approaches are feasible. However, in many cases only the general form of the kernel function is known (e.g. up to a small number of unknown parameters). This underspecification results in1 computationally difficult dual estimation problems (estimating the neural states and the model parameters). The current work aims to treat such dual problems in a computationally efficient, and highly scalable manner.

Table 3.1: Common neuroimaging measures subject to convolution

Modality	Physiology	Popular Kernels	Interpretation
fMRI	BOLD Signal	Di-Gamma [4]	Hemodynamic Response
PET	Radionucleotide Concentration	Multi-Exponential [112],[113]	Exchange, Radio decay
Ca^{2+} Imaging	Ca^{2+} Concentration	Multi-Exponential [109],[110]	Diffusion +Kinetics
LFP (low freq.)	Membrane Potential	Multi-Exponential/ Alpha [114]	K^+ Leak +Kinetics
Dendritic Recording	Post-Synaptic Potential	Multi-Exponential/ Alpha [21],[108]	K^+ Leak +Kinetics

3.1.2 Previous Work

Currently, there are several methods to deal with dual-identification for small systems and these approaches may be grouped into black-box and grey box models. However, whereas black-box modeling encompasses diverse approaches such as neural networks ([115]), Volterra Expansion ([116]), and Nonlinear Autoregressive Moving Average (NARMA) models ([117]); grey-box identification (model parameterization) has largely centered upon the dual/joint Kalman-Filters (linear, extended, unscented etc. [118, 119, 120, 121]) and related Bayesian methods. Under these approaches, the convolutional component is converted into the equivalent linear time-invariant (LTI) system format and free-parameters are modeled as additional state-variables. Thus, joint state-space techniques re-frame the dual-estimation problem as conventional state-estimation with a fully-determined model.

However, none of these methods are well situated to perform dual-identification for large (grey-box) systems due to the high computational complexity and data-intensive nature of Bayesian dual-estimation. These features are particularly limiting to neuroscience applications which typically feature a large number of connectivity parameters and potentially few sampling times (e.g. fMRI). These approaches also increase in complexity with the number of additional

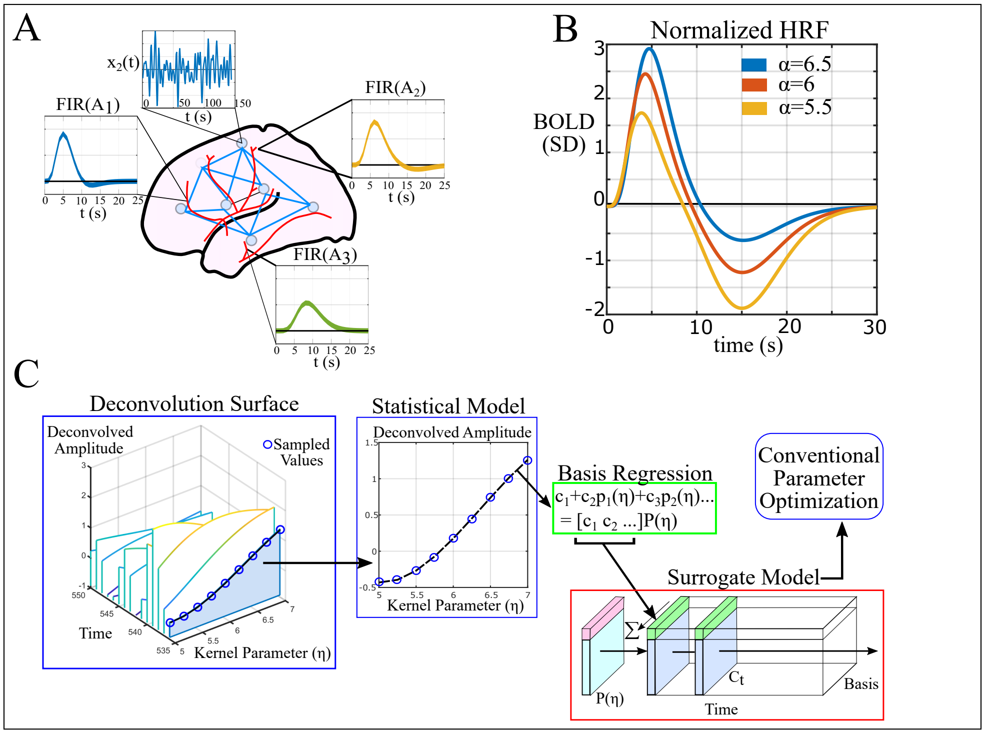
A
x2(t)
t (s)
FIR(A1)
FIR(A2)
FIR(A3)
t (s)
B
Normalized HRF
α=6.5
α=6
α=5.5
BOLD (SD)
time (s)
C
Deconvolution Surface
Deconvolved Amplitude
Sampled Values
Time
Kernel Parameter (η)
Statistical Model
Deconvolved Amplitude
Kernel Parameter (η)
Basis Regression
c1+c2p1(η)+c3p2(η)...
= [c1 c2 ...]P(η)
Conventional Parameter Optimization
Surrogate Model
Σ
P(η)
Ct
Time
Basis

Figure 3.1: General Approach. A) The brain is an example of a convolutional system when viewed through BOLD fMRI. Dynamics among brain regions are highly nonlinear and usually cannot be directly observed. B) Measurements made using fMRI reflect latent brain activity passing through a hemodynamic response function (HRF). C) Surrogate Deconvolution workflow: a deconvolution surface is estimated by sampling the time-series deconvolution across a variety of kernel parameters (left). A separate surrogate model is formed for each time-point using basis regression to approximate this surface (middle). The combined surrogate models then represent the deconvolution process during parameter estimation.

state variables necessary to represent complex kernels such as the hemodynamic response function (Fig. 3.1B).

Neural systems present two challenges to the current status quo: the dimensionality of the neural system/parameters and the complexity of the convolutional kernel. Neurobiological recordings are often high-dimensional, containing dozens or hundreds of neurons/neural populations. Moreover, the number of free parameters often scales nonlinearly with the number of populations (e.g. quadratically for the number of connectivity parameters). Current dual-estimation approaches such as joint-Kalman-Filtering are computationally limited in these settings due to their high computational complexity in terms of both the number of state variables and the number of parameters estimated.

Previous approaches are also limited in terms of kernel complexity. Since joint-Kalman-Filtering employs a state-space representation, convolutional variables are implicitly generated via linear time-invariant (LTI) systems. This issue is not inherently problematic, as many neural models contain simple exponential kernels which are easily converted to an additional LTI variable (e.g. Tab. 3.1). However, specific domains feature higher-order kernels such as the Hemodynamic Response Function (HRF) that relates latent neural activity and observed BOLD signal in fMRI. Approximating the HRF through a linear time-invariant (LTI) system requires multiple additional layers of state-variables which greatly increases the difficulty of estimating neural activity and also increases the overall computational burden.

3.2 Approach

We propose to treat this problem by directly performing optimization within the latent state-space using Surrogate Models to replace the state-estimation step (Fig. 3.1C). Surrogate functions comprise a means to approximate computationally intensive functions, typically through a linear combination of *a priori* specified nonlinear bases (e.g. polynomial families, radial-basis functions etc.). In the current case, we propose using surrogate models to explicitly estimate latent variables by deconvolving the measured time-series by the current estimate of the convolutional kernel at each iteration. Deconvolution is typically performed either by iterative algorithms such as the Richardson-Lucy algorithm ([122],[123]), Alternating Direction Method of Multipliers (e.g. [124]; ADMM) or explicit transformations in the Fourier domain. The proposed surrogate techniques are compatible with any combination of deconvolution algorithm and additional signal processing that are smooth with respect to the kernel parameters. In a later example with empirical fMRI data, we use the Wiener-deconvolution ([3]) coupled with variance normalization in the time-domain:

$$x_i(t) \approx \frac{w(z_i(t), h_i(\eta_i), K_i)}{\sigma(w(z_i(t), h_i(\eta_i), K_i))} \tag{3.4}$$

$$w(z_i(t), h_i(\eta_i), K_i) := \mathcal{F}^{-1}\left[\frac{\mathcal{F}^*[h_i(\eta_i)]\mathcal{F}[z_i(t)]}{|\mathcal{F}[h_i(\eta_i)]|^2 + K_i}\right] \tag{3.5}$$

with $w(z_i, h_i(\eta_i), K_i)$ denoting the Wiener deconvolution of z_i with respect to kernel $h_i(\eta_i)$ and noise-factor K_i equal the mean power-spectral density of the measurement noise $\nu_i(t)$ divided by the mean power spectral density of z_i. We denote standard deviation by σ and $\mathcal{F}, \mathcal{F}^*$ denote the Fourier transform and its complex conjugate, respectively. Through deconvolution, we reduce the dual-estimation problem to conventional system identification with the convolutional kernel as an additional free parameter. Using surrogate models we reduce deconvolution-algorithms into simple, differentiable functions of the kernel parameters (Fig. 3.2A). Thus rather than solving the dual estimation problem:

$$\arg\min_{\hat{\theta}, \hat{\eta}, \hat{x}_t} \left(J(\hat{\theta}, \hat{\eta}, \hat{x}_t, z_t) \right) \tag{3.6}$$

for some loss function J, we solve the parameter-estimation problem:

$$\arg\min_{\hat{\theta},\hat{\eta}} \left(J(\hat{\theta}, \hat{\eta}, S(t, \hat{\eta}), z_t) \right); \quad S(t, \hat{\eta}) \approx h(\hat{\eta}) *^{-1} z_t \tag{3.7}$$

for which S denotes the Surrogate Deconvolution model and $*^{-1}$ is a user-defined deconvolution algorithm, potentially incorporating priors on the distribution of $\nu_i(t)$ and further signal processing (e.g. normalization or additional filtering). In later experiments, we set J as the mean-squared error of 1-step predictions, e.g.

$$J = E_{t \in T} \left[\| z_{t+1} - f(\theta, S_t(\{\eta\}), z_t) \|_2^2 \right] \tag{3.8}$$

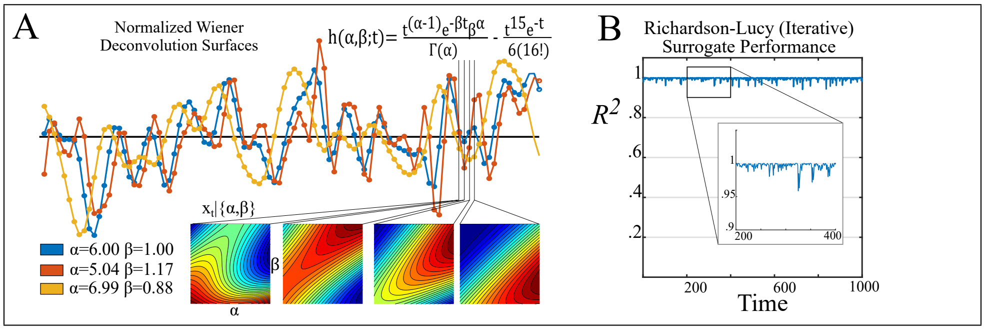

Figure 3.2: Surrogate surfaces for performing deconvolution. A) Example fMRI time-series deconvolved by hemodynamic models with different values for the kernel parameters (α, β). The inset equation is the hemodynamic response function (kernel). Subplots show the surrogate surfaces for four sequential time-points (e.g. the deconvolved signal amplitude at that time as a function of kernel parameters). Note that the surrogate surfaces at a fixed time-point are smooth, whereas the variation in deconvolved signal across time is much less regular. B) Performance in reconstructing the iterative Richardson-Lucy deconvolution of the same signal across an 81x81 grid (same range as A) using third-order bivariate polynomials (Appendix Sec. B.4). Inset shows a representative stretch of 200 points (144s).

with T denoting the set of initialization times during training. The surrogate model S is a linear combination of smooth, nonlinear bases and is therefore smooth for both iterative deconvolution algorithms, such as Richardson-Lucy, and for explicit transformations. Thus, algorithms which are natively nonsmooth due to randomization or stopping criteria (e.g. Richardson-Lucy) are converted to an accurate, but differentiable form via the Surrogate representation (e.g. Fig. 3.2B). The remaining, (surrogate-assisted) fitting problem is thus amenable to highly scalable techniques such as gradient-based optimization.

3.2.1 Contributions

Our contribution in this regard is generating surrogate models to explicitly approximate the deconvolution process in a computationally-efficient closed form. To our knowledge, previous approaches have not sought to estimate nonlinear models using parameterized deconvolution. We do so in a two-step process. First, we build a surrogate model of the deconvolution process (deconvolving $z_i(t)$ by $h_i(\eta_i)$ as a direct function of the kernel parameters η_i). We fit one surrogate function per measurement in the deconvolved space: the value of a deconvolved channel evaluated at a specific time. For a fixed basis, this representation may be fit rapidly at scale. For example, the empirical brain data treated later requires nearly two million surrogate functions per subject (419 brain regions $\times$ 4444 time points), all of which can be fit in seconds as the only computation of nonlinear complexity is shared across time points (Eq. 3.9). In the second step, we directly integrate the surrogate model into optimization algorithms. By doing so, the time-course of each latent state-variable is expressed as a direct, easily differentiable, function of the kernel parameters (η).

We present these results as follows. First we introduce surrogate methods and the proposed technique, Surrogate Deconvolution, in which surrogate models of the latent variable are directly integrated into the optimization procedure. In the next section we consider the special case of gradient-based optimization and demonstrate how error-gradients are efficiently back-propagated through the Surrogate Model. We then test Surrogate Deconvolution in two sets of experiments. First, we consider a low-dimensional case (a small LFP model) in which existing techniques for dual-estimation remain tractable. This simplified setting allows us to benchmark Surrogate Deconvolution's accuracy in parameter/state estimation relative the joint-Extended Kalman Filter and the joint-Unscented Kalman Filter. Results demonstrate that Surrogate Deconvolution is competitive even within the Kalman Filter's operating domain. We then consider more complicated fMRI models in which current dual-estimation techniques are not applicable due to high-dimensionality and kernel complexity. We demonstrate that Surrogate Deconvolution is robust to spatial variation in the HRF kernel in contrast to state-of-the-art non-dual approaches. Lastly, we demonstrate the approach's feasibility to empirical fMRI data. Thus, Surrogate Deconvolution performs competitively within the scope of current dual-estimation approaches and enables robust dual-estimation for a much larger set of problems than previously considered.

3.3 Surrogate Deconvolution

Our procedure contains two parts. First, we construct a surrogate function for each channel and time-point, a process which can be massively parallelized, if necessary. We use the surrogate construction to express the estimation of unobserved state variables as a direct function of η. The surrogate function then replaces unobserved variables in a user-chosen identification-algorithm for fully observable systems. This process is advantageous as it enables direct calculation of how changing parameters η influence the final estimate of unobserved state variables (for the current set of parameters) as opposed to techniques such as the dual Kalman Filter which do not "look-ahead" to see how changing downstream parameters will affect state estimates (since $\nabla_{\eta} f = 0$ without a surrogate model).

The key insight underlying surrogate deconvolution regards the effect of varying a kernel parameter. As demonstrated in Fig. 3.2, changing a kernel parameter produces intricate effects upon deconvolved estimates when viewed from the time-domain. Even when these effects can be expressed analytically (as in the Wiener deconvolution) they are not readily reduced to a temporally-local calculation using first-principles. However, when the kernel is lower frequency than the signal (as usually happens in biology), the effect of kernel variation on a single estimate is often quite smooth with respect to the kernel parameter. Thus, the effect of kernel variation on a single deconvolved estimate is very well-approximated by simple functions of the kernel parameter. Together, these functions comprise the surrogate model.

3.3.1 Building Surrogate Representations

We efficiently define and evaluate surrogate models by storing coefficients in tensor format. For a vector of m_i stacked basis functions $P_i(\eta_i) : \mathbb{R}^{v_i} \rightarrow \mathbb{R}^{m_i}$ we define the 3-tensor C defined for each channel ("i") and a prior distribution on η:

$$C_{i,t,:} = E_{\eta_i}[w_i(t)P_i^T]E_{\eta_i}[P_iP_i^T]^{-1}. \tag{3.9}$$

Thus, C stores the coefficients of regressing the basis functions P_i on the deconvolved time series w_i (one of P_i's bases should be $[P_i]_j = 1, \forall \eta_i$ to provide the intercept). For clarity of presentation, we have reduced the input arguments of w_i to time alone. By E_{η_i} we denote the expectation taken over some prior distribution on η_i. In practice, the choice of prior is not usually impactful, as an arbitrarily fine sampling of the response surface can be quickly computed in parallel and the surrogate goodness-of-fit can be similarly increased by adding additional (linearly independent) basis functions. In all later examples, we simply assume a uniform distribution over reasonable bounds. The tensor C holds coefficients of each time-point's surrogate model with $C_{i,t,j}$ denoting the coefficient of basis j in predicting the deconvolution of channel i at time t in the deconvolved time-domain (which is shifted from the measurement times). We evaluate the surrogate functions in parallel by defining the following product between 3-tensor C and a 2-tensor-valued function $[P(\{\eta\})]_{i,j} := [P_i(\eta_i)]_j$:

$$[P(\{\eta\}) \star C]_{i,t} = \sum_j [P_i(\eta_i)]_j C_{i,j,t} \approx x_i(t)|\eta_i \tag{3.10}$$

with the right-hand side denoting the optimal estimate of $x_i(t)$ (e.g. in the least-squares sense for Wiener deconvolution) given $\eta_i, z_i(t)$ and any fixed priors used to define the chosen deconvolution. In principle, this technique could be used for system identification objectives in which errors are defined in terms of predicting x_t or z_t or both. In practice, however, we have found that including x_t predictions within the objective function leads to a moving-target problem in which identification algorithms enter periods of attempting to maximize auto-covariance (by changing η). Therefore, we assume that objectives are given of the form:

$$J = \sum_{k \in \hat{\mathbf{k}}} \big(J_k \big([z_{t+k}]_{Actual}, [\bar{\mathbf{z}}_{t+k} | \theta, z_t, \{\eta\}] \big). \tag{3.11}$$

The final cost function J is a sum of the sub-costs J_k evaluated at the time-steps $k \in \hat{\mathbf{k}}$. Here, $\hat{\mathbf{k}}$ denotes the user-determined time steps at which to evaluate the cost function J which potentially varies by time-step (e.g. choosing to weight temporally distant predictions less). The right-hand side denotes the current estimate ($\bar{\mathbf{z}}$) of z_{t+k} given $\theta, \{\eta\}$, and z_t. Thus, the new cost function incorporates the actual measurements and their prediction. However, unlike conventional dual approaches, the predictions are a direct, explicit function of previous measurements, rather than in terms of both measurements and an iteratively estimated latent variable.

3.3.2 Deploying Surrogate Models

To evaluate the cost function, we use make forward predictions in the latent-variable (deconvolved) domain and then convolve those predictions forward in time to evaluate error in terms of observations. For k-step predictions and kernel length τ, this corresponds to:

$$\bar{\mathbf{z}}_{t_0+k|t_0} := h * [f^k(x_{t_0-\tau}, z|\eta) \;\; f^k(x_{t_0-\tau+1}, z|\eta)...] \tag{3.12}$$

$$= \sum_{k=1}^{\tau} \left(h_{1+\tau-k} \circ f^{t-t_0}\big(P(\{\eta\}) \star C_{t_0+k-\tau}, z\big) \right) \tag{3.13}$$

We use $\bar{z}_{t_0+k|t_0}$ to denote the estimate of z_{t+k} using initial conditions for the convolutional variable (z) and latent variable (x) prior to t_0. The operator $\circ$ denotes the Hadamard product (element-wise multiplication). In the latter equation, we have condensed notation for the effect of z on f as follows:

$$f^{k+1}(x_t, z) := f(f^k(x_t, z), z_{t+k}) \tag{3.14}$$

with $f(x_t, z) := f(x_t, z_t)$. Thus, f^k is not a proper iterated composition when it accepts both x_t and z_t as arguments, since only one variable (x_{t+1}) is output. However, we abuse this notation for clarity of presentation. For brevity, we also use $\hat{*}$ to indicate convolution over initial conditions as indicated in the variable indices. Hence the earlier equation (Eq. 3.13) condenses to:

$$\bar{\mathbf{z}}_{t|t_0} := h \hat{*} f^{t-t_0}(P(\{\eta\}) \star C_{[t_0-\tau,t_0]}, z) \tag{3.15}$$

As a general technique for re-representing dual estimation problems, Surrogate Deconvolution is compatible with most estimation techniques. However, the approach is particularly advantageous for gradient-based estimation as the deconvolution process is replaced with an easily differentiable surrogate form. For single-step prediction, the resulting error gradients for the nonlinear plant's parameters (θ) and the convolution kernel parameters ($\{\eta\}$) are as follows:

$$\frac{\partial J}{\partial \theta} = \frac{\partial J}{\partial \bar{\mathbf{z}}} \left(h(\{\eta\}) \hat{*} \frac{\partial f(\theta, P \star C_t, z_t)}{\partial \theta} \right) \tag{3.16}$$

$$\frac{\partial J}{\partial \{\eta\}} = \frac{\partial J}{\partial \bar{\mathbf{z}}} \left[\frac{\partial h}{\partial \{\eta\}} \hat{*} f + h \hat{*} \left(\frac{\partial f}{\partial x} \left[\frac{\partial P}{\partial \{\eta\}} \star C_t \right] \right) \right]. \tag{3.17}$$

Thus, surrogate deconvolution re-frames dual-estimation problems into conventional parameter-estimation problems which are amenable to gradient-based approaches. The analogous gradients for multi-step prediction are derived by augmenting the one-step prediction gradients with back-propagation through time. We demonstrate the power of surrogate deconvolution by reconstructing large brain network models from either simulated data or empirical recordings.

3.4 Data-driven Model Identification

We present two applications to brain discovery to illustrate the advantage of surrogate deconvolution-enhanced methods for both state-estimation (Kalman-Filtering) and grey-box parameter identification. Both examples are dual-estimation problems (state and parameter), but we assess their performance in the state and parameter components separately to make comparisons with existing work which may be particularly designed for either domain. For instance, dual-estimation using the joint unscented Kalman Filter has been particularly successful in parameterizing black-box models for filtering (e.g. [120]), but requires further modification in some more complicated grey-box models for which the conventional Kalman gain (1-step prediction) is provably rank-deficient. To demonstrate the adaptability of surrogate deconvolution we consider two different simulated system identification /estimation problems and one empirical application.

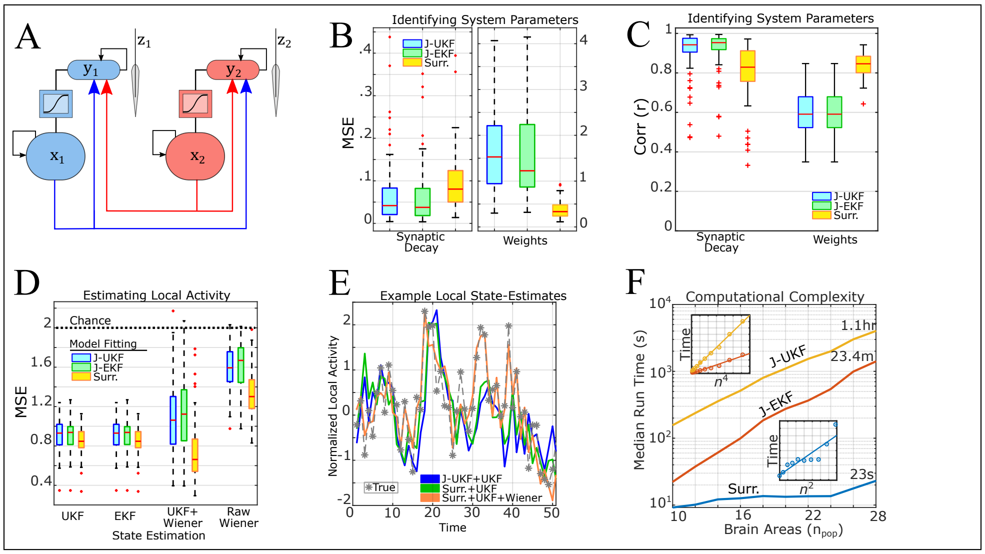

Figure 3.3: Surrogate deconvolution's performance in inverting a neural-mass model of Local Field Potential. A) Model schematic: output signals from each population arrive at post-synaptic terminals. Electrode measurements primarily reflect the post-synaptic potentials generated from synaptic activity. B) Bench-marking total error in identifying the synaptic time-constants and network connectivity. Surrogate deconvolution is compared to the current gold-standard: joint-Kalman Filters (Unscented and Extended). C) Same as (B), but displaying performance in terms of correlation rather than mean-square-error. D) Performance in reconstructing local spiking activity from electrode measurements using the identified system models for a variety of state-estimation techniques. E) State estimation performance for a representative case (the simulation with median jUKF+UKF performance). F) Computational complexity of system identification methods in terms of the number of brain regions considered. The top inset shows the run-time and least-squares fit on rescaled x-axes (n_{pop}^4) to demonstrate the $\mathcal{O}(n_{pop}^4)$ complexity of Kalman-Filtering in terms of the number of brain regions (n_{pop}). The bottom inset is the same, but for surrogate deconvolution and $\mathcal{O}(n_{pop}^2)$. Calculations were performed single-core on an Intel Xeon E5-2630v3 CPU.

3.4.1 Modeling and Isolating Local Activity from the LFP

Our first example compares performance across methodologies designed for state-estimation. This simulated problem consists of identifying the wiring of a neural system and subsequently reconstructing cellular activity from simulated extracellular recording of the "local" field potential (LFP). This signal is primarily generated by the combined currents entering into the local population of nerve cells as opposed to the currents directly generated by neural firing or the trans-membrane potential. Thus, the measured LFP reflects the temporally extended effects of input into a brain region rather than the current population activity (Fig. 3.3A). To describe this process, we use a three-level discretized-model combining 10 coupled neural-mass models ($n_{pop} = 10$) with passive integration of post-synaptic currents:

$$x_{t+1} = a\zeta(by_t) + \frac{x_t}{\tau} + c + \epsilon_t \tag{3.18}$$

$$y_{t+1} = S \circ y_t + Wx_t + \omega_t \tag{3.19}$$

$$z_t = y_t + \nu_t \tag{3.20}$$

Here, x is the synaptic-gating variable which describes neural activity. The sigmoidal transfer-function is denoted $\zeta(x) := (1 + exp[-x/5])^{-1}$ with scaling coefficient $a = 3$. The time constant of x is denoted τ and baseline drive to x is denoted c. The parameters $a, \tau \in \mathbb{R}$ and $c \in \mathbb{R}^{n_{pop}}$ are assumed known as are the covariances of process noise ϵ_t, ω_t and measurement noise ν_t (see Appendix). Thus, the unknown parameters are the connections between neural populations (W) and the synaptic time-constants S. We considered two general approaches to system identification: either using the current gold-standard (joint Kalman estimation) or using surrogate deconvolution for least-squares optimization. The joint Kalman filter and associated variants operate identically to the original Kalman filter, except that the state-space model is augmented with unknown parameters being treated as additional state-variables with trivial dynamics (e.g. $W_{t+1} = W_t + \hat{\epsilon}_t$ and similarly for S). The "noise" terms $\hat{\epsilon}_t$ for parameter state-variables are assumed iid. and with a user-defined variance that determines the learning rate. Based upon numerical exploration, we found that the best performance for both EKF and UKF was with an initial prior on parameter variance $var[\hat{\epsilon}] = .001$. Every 50 time-steps we decreased the variance prior by 5% of its current value.

For comparison with existing techniques we used both the joint-Extended Kalman Filter (jEKF) which linearizes the nonlinear portion of dynamics and the joint-Unscented Kalman Filter (jUKF) which directly propagates noise distributions through nonlinearities using the Unscented Transformation ([119]). We compare these traditional methods with system identification through surrogate deconvolution. The benefit of surrogate deconvolution is the ability to apply a wide variety of optimization techniques to partially-observable identification problems which can decrease computation time over conventional techniques (Fig. 3.3F) and expand the scope of problems which may be tackled. For this first example, we have chosen a relatively simple case (low-dimensional, single-exponential kernels) so that conventional methods (Kalman Filtering). Therefore, the goal of this test is not to demonstrate an overwhelming advantage of surrogate deconvolution over Kalman Filtering, but to show that the proposed technique can perform competitively in cases for which Kalman-Filtering is applicable, but imperfect. Subsequent examples will consider cases in which Kalman Filtering is not tenable.

To implement Surrogate Deconvolution, we first reformulate this problem as a convolutional equation through the change of variable $r_t := Wx_t$:

$$r_{t+1} = aW\zeta(y_t) + \frac{r_t}{\tau} + Wc + W\epsilon_t \tag{3.21}$$

$$y_{t+1} = S \circ y_t + r_t + \omega_t \tag{3.22}$$

or, equivalently,

$$r_{t+1} = \frac{r_t}{\tau} + W[a\zeta([(r+\omega) * \mathcal{P}(S)]_t) + c + \epsilon_t] \tag{3.23}$$

$$z_t = [(r+\omega) * \mathcal{P}(S)]_t + \nu_t \tag{3.24}$$

with $\mathcal{P}(S)$ denoting the discrete-time kernel formed from polynomials of S to a suitably long length $[0 \;\; 1 \;\; S \;\; S^2 \;\; S^3 ...]$ analogous to exponential decay for continuous-time systems. In this form, the parameters can be estimated using traditional least-squares methods, optimizing over W and S. However, by leveraging the tensor representation of surrogate models, this equation can be reduced into a single equation in S by representing the optimal choice of W for a given S as a direct function of S. To do so we define the matrix

$$F_t := a\zeta(z_t) + c \tag{3.25}$$

and the associated 3-tensor

$$M_{i,j,p} = (E_t[(C_{t+1,p} - C_{t,p}\tau^{-1})F_t^T]E_t[F_tF_t^T]^{-1})_{i,j}. \tag{3.26}$$

Each $n \times n$ page of this tensor (e.g. the matrix formed by holding p constant) stores the coefficients of the least squares solution for W in predicting $C_{t+1,p} - C_{t,p}\tau^{-1}$ from F_t for the pth basis function. Since $r_t^*(S) = C_t \star P(S)$, for a given synaptic decay term S we use the notation $r^*(S)$ to denote the estimate of r produced through Surrogate Deconvolution of measurements z with the kernel $\mathcal{P}(S)$. This produces the least-squares estimate for W as a direct function of S:

$$\arg\min_W ||r_{t+1}^*(S) - (WF_t + r_t^*(S)\tau^{-1})||_F^2 = \mathcal{P}(S) \star M, \tag{3.27}$$

$$z_{t+2} \approx S z_{t+1} + [\mathcal{P}(S) \star M] F_t + [\mathcal{P}(S) \star C_t] \tau^{-1} \tag{3.28}$$

Thus, in this case, surrogate deconvolution enables the approximation of $2n_{pop}$ difference equations containing $n_{pop}(n_{pop}+1)$ unknown parameters (W and S) to n_{pop} difference equations with n_{pop} unknown parameters (only S). The resultant model (from Eq. 3.28) is also compatible with a wide variety of optimization techniques. For simplicity, we fit the parameters S through ordinary least-squares optimization in terms of predicting z_{t+2} as in Eq. (3.28). Optimization was performed using Nesterov-Accelerated Adaptive Moment Estimation (NADAM; [9]) with both NADAM memory parameters set equal to .95, and the NADAM regularization parameter set to .001. Training consisted of 15,000 iterations with each minibatch containing 1000 time points. The step size (learning rate) of updates was .0001.

All methods were able to retrieve accurate estimates of the synaptic decay term S (Fig. 3.3B,C). The best-performing method varied by simulation (e.g. for different true values of W, S), but the mean error was greater for surrogate deconvolution than Kalman Filtering methods (Unscented and Extended) which performed near-identically. By contrast, surrogate deconvolution always provided a more accurate estimate of the connectivity weight parameter (W) and the advantage relative Kalman-Filtering was substantial (Fig. 3.3B,C). The poor performance of the Kalman Filter for identification in this case is not surprising as the Kalman Filter is known to be non-robust for large systems ([125]) and the W parameter adds 100 additional latent state-variables to the joint Kalman model as opposed to the 10 state-variables added by S.

3.4.2 Reconstructing Firing-Rate from LFP

We next examined the ability of each method to recover the time series of neural activity x_t using the previously generated state-space models. During this stage, models produced during the previous identification step were used to estimate the latent state variable x_t (Fig. 3.3D,E). It is important to distinguish between state-estimation techniques (e.g. UKF) which we used to estimate x_t from previously-fit models and the techniques used to fit those initial models (e.g. jUKF) as these steps need not "match" (e.g. UKF-based state-estimation from a surrogate-identified model). Measurements consisted of simulated extracellular voltages z_t generated by resimulating the same ground-truth model (i.e. the same parameters, but new initial conditions and noise realizations). As in the identification stage, we considered two general approaches to recovering the latent variable x_t: either through deconvolution or Kalman Filtering (unscented and extended). Kalman filtering in this setting produces direct estimates of x_t and y_t. By contrast, deconvolving $z_t \approx y_t$ produces an estimate of r_t, so deconvolution estimates of x_t were produced by premultiplying the deconvolved time series with W_{est}^{-1} (the inverse estimated connectivity parameter). We considered deconvolution applied either directly to the raw measurements (z_t) or to the estimates of y_t produced by Kalman filtering z_t with the estimated models (both unscented and extended Kalman filters were considered). Noise covariance estimates for Kalman filtering at this stage were the same as those assumed in the initial stage: a value close to the mean tendency, rather than the true values which were randomly selected for each simulation.

We found that the type of Kalman Filter used for state-estimation had no appreciable effect upon accuracy (Fig. 3.3D). Likewise, the technique used for system identification (surrogate deconvolution vs. EKF/UKF) had little effect, although surrogate deconvolution was slightly more accurate on average. However, model performance differed greatly for deconvolution-based state-estimation (using $x_t \approx W^{-1}[\mathcal{P}(S) *^{-1} y_{est}]_t)$. Models estimated using joint-Kalman Filtering (jEKF/jUKF) performed worse using deconvolution-based estimation of x_t than Kalman-based estimation (Fig. 3.3D). This result is unsurprising as the deconvolution-based estimate additionally requires the inverse weight parameter W^{-1} and both jUKF and jEKF poorly estimated W. Interestingly, however, estimation accuracy for surrogate-identified models decreased when using deconvolution of the raw, unfiltered measurements, but increased for the UKF+decconvolution hybrid. The former result is not surprising as pure deconvolution is clearly suboptimal in not considering the noise covariance. This result was unexpected and it suggests the possible benefit of using a two-stage estimation procedure in which Kalman-Filtering first dampens measurement noise and improves estimates of measurable state-variables. Then, subsequent deconvolution might improve the estimate of latent state-variables by considering the impact of estimates across time, rather than just the directly subsequent measurement. However, these benefits are likely situation-dependent and therefore require more study. In any case, results indicate that state-estimates from models produced by surrogate-deconvolution are at least as accurate as those produced by Kalman-Filtering and potentially more so depending upon the state-estimation procedure (Fig. 3.3D,E).

3.4.3 Computational Efficiency

Surrogate deconvolution is also computationally efficient as it scales linearly with the number of measurement channels ($\mathcal{O}(n)$) in both forming and evaluating surrogate functions which is also parallelizable across channels. However, since Surrogate deconvolution is not a system identification procedure in and of itself, time-savings depend upon how the technique is used (e.g. which optimization scheme it is coupled to). The advantage of surrogate deconvolution is that it can be combined with a wide-variety of optimization techniques which are otherwise

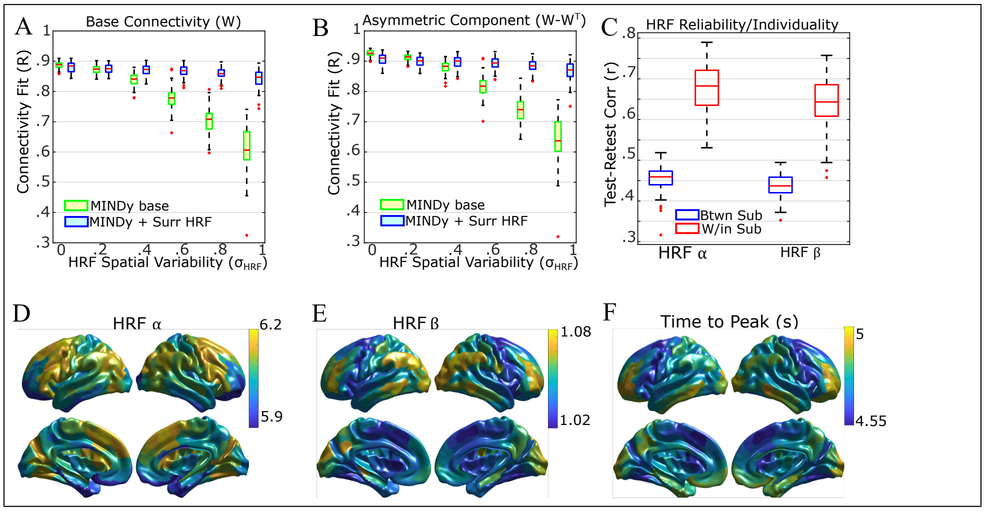

Figure 3.4: Incorporating HRF surrogate-deconvolution into MINDy. A) Without HRF modeling, connectivity estimates degrade with spatial variability in the neurovascular coupling. Fitting the HRF through surrogate deconvolution preserves performance. B) Same as (A) but for the asymmetric component of connectivity. C) HRF parameter estimates from HCP data are reliable across scanning days and subject-specific. D) Spatial map of the mean α parameter estimate across subjects. E) Same as (D), but for the second HRF parameter (β). F) Spatial map of the mean time-to-peak in the fitted HRF's.

ill-suited to partially-observable problems. In this first simulation, for instance, the number of parameters scale with $n_{pop}(n_{pop}+1)$ and the number of state variables (in the native space) scale with $2n_{pop}$. Thus, the dominant complexity of joint-UKF and joint-EKF is greater than $\mathcal{O}(n_{pop}^4)$ as joint-UKF/EKF are $\mathcal{O}(n^2)$ in the number of parameters and at least $\mathcal{O}(n^2)$ in the number of native (non-parameter) state-variables. By contrast, the gradient approaches applied with surrogate deconvolution have approximately $\mathcal{O}(n_{pop}^2)$ complexity (Fig. 3.3F). However, surrogate deconvolution is not limited to gradient-based approaches. The main effect is to simplify error functions to a direct equation in the measurable variables so surrogate deconvolution is compatible with a wide variety of non-gradient techniques, as well (e.g. heuristic-based or Bayesian). As such, surrogate deconvolution presents the opportunity to identify significantly larger partially-observable systems than previously considered.

3.4.4 Reconstructing Connectivity and Hemodynamics in Simulated fMRI

For our second example, we considered the ability to correctly parameterize large-scale brain models from simulated fMRI data. Brain regions were modeled through the continuous-valued asymmetric Hopfield model ([1]) and simulated fMRI signals were produced by convolving the simulated brain activity with randomly parameterized Hemodynamic Response Function (HRF) kernels ([4]):

$$x_{t+\Delta t} = W[tanh(b_0 \circ x_t)]\Delta t + (1-\Delta t)Dx + \epsilon_t \tag{3.29}$$

$$z_t = [x * h(\alpha, \beta)]_t \tag{3.30}$$

$$h_i(\alpha, \beta, t) = \frac{t^{\alpha_i - 1} e^{-\beta_i t} \beta_i^{\alpha_i}}{\Gamma(\alpha_i)} - \frac{t^{15} e^{-t}}{6(16!)} \tag{3.31}$$

Parameter distributions for simulation are detailed in the Appendix. Simulations were integrated at Δt=100ms and sampled every 700ms (mirroring the Human Connectome Project's scanning TR of 720ms [65]). Simulated HRF's (h_i) were independently parameterized for each brain region according to the distributions $\alpha_i \sim \mathcal{N}(6, \sigma^2)$ and $\beta_i \sim \mathcal{N}(1, (\sigma/6)^2)$ in which the variability term σ was systematically varied. Each HRF can be well approximated by a finite-length kernel and therefore can be represented as a discrete-time linear plant. However, doing so, in this case, requires multiple hidden state-variables per region which impairs Kalman-based dual-estimation procedures. Instead, most current procedures to deal with fMRI-based systems identification at scale ignore inter-regional variability in h_i and instead seek to retrieve x_t by fixing HRF parameters (e.g. [126], [8]) to the so-called "canonical HRF" (e.g. $\alpha = 6, \beta = 1$). In this example, we demonstrate the potential pitfalls of this assumption and the benefits accrued by efficiently fitting hemodynamics through Surrogate

Deconvolution. To do so, we attempted to reconstruct W using Mesoscale Individualized NeuroDynamics (MINDy) in either its base form (which assumes a canonical HRF) or in an augmented form in which the predictions are calculated as in Eq. 3.13. MINDy model fitting consists of using NADAM-enhanced gradient updates ([9]) to minimize the following cost function:

$$\begin{aligned} J = \frac{1}{2}E_T[\|(x_{T+\Delta t} - x_T) - [(W_S + W_L)\psi_\gamma(x_t) - Dx_T]\|_2^2] \\ + \lambda_1\|W_S\|_1 + \lambda_2 Tr(|W_S|) + \lambda_3(\|W_1\|_1 + \|W_2\|_1) + \frac{\lambda_4}{2}\|W_L\|_2^2 \end{aligned} \quad (3.32)$$

in which the estimated weight matrix $\hat{W}$ is decomposed into the sum of estimated sparse (W_S) and low-rank (W_L) components satisfying:

$$\hat{W} = W_S + W_L = W_S + W_1 W_2^T \quad (3.33)$$

for some $W_S \in \mathbb{M}^{n\times n}$ and $W_1, W_2 \in \mathbb{M}^{n\times k}$. The hyperparameter $k < n$ determines the rank of the low-rank component W_L and the regularization hyperparameters $\{\lambda_i\}$ define statistical priors on each of the weight matrix components (Laplace prior for W_S, W_1, W_2 and a normal prior for $W_L := W_1 W_2^T$). This decomposition has been shown useful to estimating large brain networks ([8]). The nonlinear function ψ is parameterized by the parameter vector $\gamma \in \mathbb{R}^n$ with

$$\psi_\gamma(x) = \sqrt{\gamma^2 + (bx_t + .5)^2} - \sqrt{\gamma^2 + (bx_t - .5)^2} \quad (3.34)$$

For the Surrogate-Deconvolution, however, these analyses are performed in the original space to prevent the afore-mentioned moving target problem. Hyper-parameter determination and simulation parameters are detailed in the appendix.

$$J = \frac{1}{2}E_T[\|(z_{T+\Delta t} - z_T) - [h * (W_S + W_L)\psi_\gamma(P \star C)]_T + Dz_T\|_2^2]$$
$$+ \hat{\lambda}_1\|W_S\|_1 + \hat{\lambda}_2 Tr(|W_S|) + \hat{\lambda}_3(\|W_1\|_1 + \|W_2\|_1) + \frac{\hat{\lambda}_4}{2}\|W_L\|_2^2 \quad (3.35)$$

Results demonstrated a clear benefit for additional modeling of the local hemodynamic response (Fig. 3.4A,B). When hemodynamics differed only slightly between simulated brain regions both methods produced highly accurate estimates of the connectivity parameter W. However, past $\sigma = .4$ (the SD of spatial variation in one of the HRF parameters), the accuracy of estimated connectivity rapidly decreased for conventional methods, while only slightly decreasing for surrogate deconvolution. In addition, the hemodynamic parameter estimates also became increasingly accurate. Thus, surrogate deconvolution enables accurate system (brain) identification when the downstream plants (hemodynamics) are variable and unknown.

3.4.5 Empirical Dual Estimation with the Human Connectome

Lastly, we tested the effects of using Surrogate Deconvolution in fitting MINDy models to data from the Human Connectome Project ([65]). By using empirical data, this analysis demonstrates that human hemodynamics are spatially variable and that accounting for this variability produces more nuanced and reliable brain models. Data consists of one hour of resting-state fMRI per subject spread across two days (30 minutes each). Data were processed according to the recommendations of Siegel and colleagues ([2]) and divided into 419 brain regions ([6]). We then fit MINDy models either with or without surrogate deconvolution to this data using the same fitting procedure and hyperparameters ([8]) as before. Results indicated the the parameters which describe the hemodynamic response function are reliable across scanning days and reliably differ between individuals (Fig. 3.4C). Each of the two HRF parameters had a stereotypical spatial distribution (Fig. 3.4D,E) as did the time-to-peak of the recovered HRF kernels (generated by substituting in the recovered kernel parameters). Time to peak was slowest for anterior prefrontal cortex, particularly in the right hemisphere (Fig. 3.4F). Because current knowledge of the “true” hemodynamic response is limited, future study establishing ground-truths for HRF variation across human cortex is needed to facilitate more rigorous empirical validation.

3.5 Discussion and Conclusion

3.5.1 Generalizability of the Problem Framework

The methods that we propose are dependent upon the problem satisfying two criteria: 1) unmeasured variables can be related to measured ones via convolution and 2) the form of the convolutional kernels are known up to a small number of parameters per kernel. These assumptions are satisfied in many areas of neuroscience (see Tab. 3.1) in which measurements have high spatial precision relative to the underlying models. These kernels can also be derived by analytically reducing large state-space models to a smaller convolutional form (see Sec.B.5). In state-space formulation, these problems all contain more state-variables than recording channels, but they all still contain one channel per anatomical unit (region, population, cell etc. depending upon the model). In other words, the inverse-problem from these scenarios results from mechanistic undersampling (i.e. only measuring one type of variable) rather than spatial undersampling.

The relationship between the measured and unmeasured variables can be either unidirectional (e.g. neural activity influencing BOLD but not vice-versa) or bidirectional (e.g. neural activity and synaptic currents influencing each other) and this formulation covers a large number of empirically relevant scenarios. However, the assumption of full spatial precision (relative the model) also limits our approach to specific modalities (see Tab. 3.1). As presented, our technique is not applicable to sensor-level EEG or MEG recordings since each channel's signal reflects a (weighted) summation of activity at many anatomical locations. By contrast, other techniques such as the joint Kalman Filters (with which we compare our method) are applicable to these scenarios and simultaneously perform spatial-inversion and model parameterization. Thus, the proposed techniques only cover specific classes of modeling scenarios which are but a subset of problems in which the joint Kalman Filters are applicable. However, as we have demonstrated in the results, our approach scales much better with model size. Thus, our technique is generalizable in terms of model scale, whereas the Kalman Filter is more general with respect to model type.

3.5.2 The Role of Priors in Deconvolution

A second assumption of our technique is that the convolutional kernels are known up to a relatively small number of parameters each, thus constituting semi-blind deconvolution. This assumption holds in a wide variety of scenarios in which prior empirical evidence suggests an approximate functional form (e.g. the double-gamma HRF [4]). However, there remain cases in which the general form of the kernel is unknown, or the form contains many unknown parameters (e.g. an unknown kinetic scheme with many conformations). Fortunately, several statistical approaches to blind deconvolution exist, many of which require few assumptions regarding the kernel's form (e.g. [127], [128]). The Richardson-Lucy algorithm ([122],[123],[127]) is one popular example for both blind and semi-blind deconvolution when the noise statistics are Poisson. These approaches can also be applied to unknown kernels which span both time and space, whereas our technique only considers convolution in the temporal domain. For these reasons, blind-deconvolution algorithms have been previously applied to a variety of neuroscience domains (e.g. [129], [130]). The primary drawback of statistical blind-deconvolution algorithms, however, is that solutions are at most unique up to an unknown lag for each channel so it may not be possible to discern the order of latent events between channels. By contrast, the proposed method considers the dynamical relationship between channels. As a result, solutions identify the relative timing of latent neural events across channels.

3.5.3 Conclusion

Data-driven modeling remains one of the key challenges to neuroengineering and computational neuroscience. Although a wealth of theoretical model forms have been produced, the state-variables of these models (e.g. neuronal firing rate) are often difficult to directly measure $in-vivo$ which complicates system-identification (model-parameterization). Instead, many clinical and experimental contexts record proxy variables which reflect the physiologically downstream effects of neuronal activity (e.g. on blood oxygenation, signaling molecules, and synaptic currents). In the current work, we aimed to parameterize conventional neural models using indirect measurements of neural activity. This problem involved simultaneously estimating the generative neural model as well as the latent neural activity thus comprising a dual-estimation problem. Through surrogate models, we approximated the state-estimation step as a parameterized deconvolution,, thus reducing computationally challenging dual-estimation problems to a closed-form, conventional identification problem. The primary advantage of this approach is speed/scalability.

Current approaches to model-based dual-estimation emphasize the joint/dual Kalman Filters and related Bayesian approaches (e.g. [121]). These approaches suffer, however, in terms of scalability and data quantity. As demonstrated in numerical simulations, the computational complexity of Kalman-Filtering limits application to relatively small models (Fig. 3.3F), whereas Surrogate Deconvolution enables optimization techniques that scale well with the number of parameters (Fig. 3.3F). However, despite requiring orders of magnitude fewer computations, Surrogate Deconvolution performs competitively with Kalman Filtering in estimating system parameters (Fig. 3.3B,C) as well as estimating states (latent neural activity; Fig. 3.3D,E). Thus, the computational advantages of Surrogate Deconvolution do not compromise accuracy.

Scalability is particularly salient in empirical neuroimaging, as several recent approaches have eschewed detailed modeling of physiological measurements (e.g. [8], [126]) in order to increase the spatial coverage of models. However, ground-truth simulations indicate that these reductions potentially compromise accuracy (Fig. 3.4 A,B). By contrast, methods augmented with Surrogate Deconvolution maintained high levels of performance (accuracy) even for extreme spatial variation in physiological signals. Interestingly, this variation appeared to be a reliable feature in empirical data with consistent differences across individuals (Fig. 3.4 C) and brain regions (Fig. 3.4 D-F) which can potentially lead to systematic biases (as opposed to random error) when these features are not modeled. Thus, for neuroimaging in particular, it may be critical to parameterize both the generative neural model and the measurement models to account for these biases. Surrogate Deconvolution provides a means to parameterize such models without compromising the detail of either component.

Chapter 4

Enhancing Task fMRI Preprocessing via Individualized Model-Based Filtering of Intrinsic Activity Dynamics

4.1 Introduction

Task-related analyses in fMRI typically involve statistical general linear models (GLMs) which seek to identify the amplitude and/or mean timecourse of (BOLD) evoked-responses after removing nuisance covariates. These approaches have proven statistically powerful for simple tasks and characterize much of the current literature regarding task-induced activation. However, over the past decade, improvements in the accuracy and speed of fMRI acquisitions has given birth to a new literature concerning within-subject trial-to-trial variation in brain activity. These studies have generated two key findings relevant to the current study: 1) trial-to-trial variation in BOLD responses predict within-subject behavioral variation ([51]) and 2) the BOLD signal elicited by a stimulus is dependent upon the previous pattern of brain activity ([48], [46]). We use the term “brain activity” in the latter case to indicate that this history dependence is thought to be neural, rather than solely reflecting potential nonlinearity in the hemodynamic coupling. These results indicate that trial-to-trial variation in the BOLD response is due, in part, to variation in the underlying neural activity. Moreover, this variability is behaviorally relevant ([49]) and is history dependent (i.e. reflects underlying dynamics). Thus, the neural activity associated with BOLD is increasingly considered as a nonlinear dynamical system—one in which the spatiotemporal response to an input depends upon its current state and is determined by a set of rules that dictates its temporal evolution ([16]).

This framework contrasts both with current statistical approaches, which treat the neural activity as a noisy autoregressive signal (most GLMs), and with Dynamic Causal Modeling (DCM) approaches which treat the brain as a linear system (although see [101]). In the current work, we propose a new technique for modeling intrinsic brain dynamics and their contribution to task-evoked activation patterns. This approach, which we term MINDy-based Filtering, more accurately identifies individual differences, and enhances the temporal precision and statistical power in identifying task effects.

4.1.1 Introducing Dynamics to Activity Flow

Brain activity dynamics are thought to consist of two factors: 1) the signaling between brain regions; and 2) the local dynamics that describe how each brain region integrates afferent signals over time. Several recent efforts have focused solely upon the former factor to study how evoked activity propagates through the brain. The Activity Flow ([43]) framework, for example, approximates task-evoked activity in each brain area as the FC-weighted sum of activity in all other areas. Previous studies have indicated that Activity Flow accurately describes spatial patterns of task-evoked activity. The primary limitation of the Activity Flow approach, however, has been the lack of dynamics. Since inferences are based only upon static relationships, these models reduce to mathematical statements about the admissible spatial patterns of task activity. In particular, the Activity Flow model is mathematically equivalent to stating that task activity patterns occur along the principal components (or principal subspaces) of resting-state activity. Thus, the inferences yielded by these approaches are inherently static and correlative, whereas the target construct (activity propagation) is inherently dynamic and causal (i.e., often referred to as “effective connectivity”; [131]).

4.1.2 Previous Approaches using DCM

Dynamic Causal Modeling (DCM), by contrast, incorporates the temporal evolution of brain activity and thus can consider the propagation of neural activity through brain networks. Each DCM contains an effective connectivity matrix and a set of extrinsic inputs that describe how task events impinge upon each node of the network ([52]). Many implementations also contain region-specific hemodynamic models and/or an interaction between task events and effective connectivity (i.e. the effective connectivity is parameterized by task events). Although the original DCM models were strongly limited in size, modern implementations ([57], [76]) can consider a much larger number of brain regions (although the computation cost remains considerable; [57], [8]). However, the DCM methodology also presents several constraints which limit its application. Estimating a DCM model requires pre-specifying the time-series of task effects. This assumption precludes analyses which explore the temporal dynamics of task effects such as Finite Impulse Response (FIR) modeling or nuanced task

GLMs, such as those featuring nuisance regressors (e.g. motion). In addition, all DCM implementations that support whole-brain models (i.e. more than a few regions; [57]) are strongly ([8]) dependent upon the assumption of stationary linear dynamics.

4.1.3 Filtering Instead of Parameterizing

In the current work, we aim to strike a balance between the mechanistic inferences made by DCM and the flexibility of standard analysis techniques. To do so, we generate dynamical systems models of the brain and neurovasculature (as is done in DCM). However, our approach differs substantially from DCM in how we build and utilize these models. Instead of fitting models of the brain and tasks, we propose to fit dynamic models to independent resting-state data for each subject. We then use these models to generate a mathematical filter for each subject that removes, or “partials out”, the effects of intrinsic dynamics from BOLD timeseries. This approach uses no information regarding task events and thus functions as a preprocessing step, as opposed to explicitly modeling task events. This feature is advantageous, as the proposed techniques can be inserted into any data preprocessing pipeline with minimal effort, provided that sufficient amount of resting state data (e.g. >15 minutes [8]) has been collected to build MINDy models.

4.2 Approach

4.2.1 Model Derivation

Our approach leverages individualized resting-state models in order to estimate task-evoked brain effects while making minimal modeling assumptions about the underlying mechanisms. We model brain activity (x_t) as a dynamical system containing two components: an intrinsic dynamics component $f(x)$ which is estimated from resting-state models and an exogenous input component I_t.

$$x_{t+1} = f(x_t) + I_t. \tag{4.1}$$

The latter component is exogenous with respect to the resting-state model and should not be interpreted as "exogenous to the brain". Rather, I_t represents additional input to each brain region beyond that which is created through intrinsic (resting state) dynamics embedded in $f(x)$. In principle, this technique is compatible with any resting-state model ($f(x_t)$). For the current work, we chose MINDy ([8], [99]) as it is highly scalable, nonlinear, and robust to many nuisance factors. The aim of the current work is to estimate the input (I_t) for task data and to investigate exogenous input as a marker for cognitive states. We do not assume a specific mechanism underlying this input (e.g. recurrent input, inter-regional signaling, neuronal "noise", or sensory afferents) or any spatial/temporal properties of I_t. Thus, we treat I_t as a latent signal to be estimated (i.e. filtering I_t from BOLD). By contrast, other methods, such as DCM ([52],[102]) assume a time course of I_t (the temporal aspects of I_t) based upon task design and only estimate its relative contribution to each brain area. For this reason, we term our objective MINDy-based Filtering. Although the mechanisms of interest (I_t) are modeled as neural, fMRI measures the hemodynamic BOLD contrast. For this reason, we simultaneously model neural dynamics and the hemodynamics which link neural events to fMRI measurements. We assume that BOLD signal reflects the convolution (denoted "*") of latent neural activity (x_t) with a region-specific Hemodynamic Response Function (HRF; denoted $h_i(t)$) and we estimate the HRF kernels from resting state data ([99]). Thus, for each brain region (parcel "i") our model is:

$$BOLD_t^{(i)} = [h_i * (x_\tau^{(i)} + \eta_\tau^{(i)}))]_t + \nu_t \tag{4.2}$$

We consider noise at the level of the neurovascular coupling η_t and at the level of BOLD measurements (ν_t). These terms are modeled as normal random variables which are independently and identically distributed (iid) between brain regions and time points. Process noise (physiological stochasticity) is not explicitly modeled at the neural level Eq. 4.1, as it is absorbed in the unknown inputs I_t. Substituting for x_t (from Eq. 4.1) and rearranging yields:

$$BOLD^{(i)}_{t+1} - [h_i * f^{(i)}(x)]_t = [h_i * I^{(i)}_\tau]_t + [h_i * \eta^{(i)}_\tau]_t + \nu^{(i)}_t. \tag{4.3}$$

Thus, the HRF-convolved input $[h * I]_t$ is equal to the difference between measured and predicted BOLD plus additional autocorrelated noise terms. For all current analyses we consider brain states estimated with HRF-convolved estimates of input ($[h * I]_t$) as opposed to the estimates of I_t alone. This step enables the same statistical pipelines (i.e. GLM structure) to analyze original fMRI BOLD data and the HRF-convolved input. As a result, the estimation of $[h * I]_t$ serves as an additional "preprocessing" (filtering) step that can be added to any fMRI pipeline with minimal effort. No information regarding task events is used in estimating I_t, so the same statistical frameworks are applied to model-filtered and original data.

4.2.2 MINDy-based Filtering

In the current approach, we do not explicitly model different forms of noise. The only noise factor we consider is the measurement noise power in inverting BOLD onto neural activity. Since neurovasculature noise is removed (η_t=0), Wiener deconvolution generates the least-mean-square estimate for x_t. The resultant approximation for BOLD-convolved input ($[h * I]_t$) is:

$$[h * I_\tau]_t \approx BOLD_{t+1} - [h * f(h \hat{*}^{-1} BOLD]_\tau]_t \tag{4.4}$$

With $h\hat{*}^{-1}BOLD$ denoting the Weiner deconvolution of each region's BOLD signal with respect to the corresponding HRF model. Thus, we estimate neural activity by deconvolving BOLD with the region-specific HRF's identified at rest. Predictions are made in terms of neural activity and re-convolved to produce predictions in terms of BOLD. The difference between measured and predicted BOLD approximates the HRF-convolved input.

4.3 Methods

4.3.1 Subjects

Data consisted of fMRI task and resting-state scans for 50 healthy young-adult subjects collected as part of the Dual Mechanisms of Cognitive Control (DMCC) study. The DMCC participant pool contains a large number of monozygotic and dizygotic twin pairs. However, for these analyses, these characteristics are ignored.

4.3.2 Scanning Protocol

Each participant took part in three separate scanning sessions which occurred on different days, but all had the same general procedure. Each day, participants provided two resting-state scans of 5 minutes each as well as two scans each for four cognitive tasks: the AX-Continuous Performance Task (AX-CPT), Sternberg Task, Stroop Task, and Cued Task-Switching (Cued-TS). The two scans per task were performed sequentially for each task whereas the two resting-state scans were separated in time (one at the session start and one at end). Each of the task scans (2 per task per day) contained three task-blocks separated by inter-block intervals and lasted approximately 12 minutes. For resting state and task, the two scans per day were split between anterior-posterior and posterior-anterior phase-encoding directions. Scans were performed at 3T with 1.2s TR (multi-band ×4).

4.3.3 Task Descriptions

We briefly describe the general structure of each of the four cognitive tasks in the "baseline" format which was administered on the first scanning day. Subtle changes to task structure were made on the two following days (subsequent section) but were not relevant to our analyses. The **AX-CPT** task ([132]) involves repeated sequences of cue-probe pairs, in which the response to the probe item is constrained by the preceding contextual cue. Thus, the A-X cue-probe pairing requires a target response and is frequent, leading to strong associations between the cue and probe. However, both the B-X pairing (where "B" refers to any non-X cue) and A-Y pairing (where "Y" refers to any non-X probe) require non target responses. In the **Sternberg task** ([133]), participants are sequentially presented with short list of words to memorize for that trial (called the memory set; appearing across two encoding screens). After a short retention delay, they are presented with a probe word and must determine if the probe was present in that trial's memory set. On some trials, the probe item is termed a "recent negative", in that was not present in the current trial memory set but was present in the memory set from the preceding trial. In the current implementation of the **Stroop task**, subjects are asked to verbally report the font color in which probes are displayed ([134]). Each probe is itself a color-word, and can either be congruent (font color is the same as the color word, e.g., BLUE in blue font) or incongruent (font color is different from the color-word name; e.g., BLUE in red font). Lastly, during **Cued Task-Switching** (Cued-TS, [135]) participants are pre-cued to attend to either the number or letter component of a subsequent probe (combined letter + digit). In "attend-number" trials, participants indicate whether the digital component of a probe is even vs. odd. In "attend-letter" trials, participants indicate whether the letter component is a consonant vs. vowel. The probe can be either congruent (both letter and digit are associated with the same response) or incongruent (the letter and digit are associated with different responses). With the exception of the Stroop task, participants report responses using button presses.

4.3.4 Cognitive Control Demand

The current set of trial-based analyses center upon the ability to identify neural signatures of cognitive control. Although cognitive control is a heterogeneous construct, we specifically studied the conflict resolution aspects of cognitive control, so we use the terms control-demand and conflict interchangeably when referring to these tasks, and contrasts between trial types. In particular, we operationally identify cognitive control demand as the difference in neural activity measures during high and low-conflict trials for each task. In the AX-CPT, we contrast BX trials (high conflict) vs. BY (low conflict). The BX trials are high conflict because of the target-association with the X-probe, which require contextual cue information to override. For the Sternberg task, we contrast trials with recent negative probes (high conflict) and trials containing novel negative probes (low-conflict). Thus, recent negative trials are high conflict because the familiarity of the probe, requires information actively maintained in memory to override. In the Stroop task, we contrast incongruent (high conflict) and congruent (low conflict) trials. The incongruent trials are high conflict because the task goals (name the font color) are required to override the dominant tendency to read the color-name. Lastly, in the Cued-TS we also contrast incongruent (high conflict) and congruent (low conflict) trials. The incongruent trials are high conflict because it is critical to process the task cue, in order to know what response to make (for congruent trials, the same response would be made regardless of the task being performed).

4.3.5 Task Manipulations

The four tasks (AX-CPT, Sternberg, Stroop, and Cued-TS) were chosen to measure/engage cognitive control. On the first scanning day, participants performed a “baseline” version of each task. On the subsequent days, however, participants performed modified version of each task, meant to promote either proactive or reactive cognitive control strategies. On the two subsequent scans participants performed all the reactive-mode conditions of the tasks on one day and all the proactive-mode conditions of the tasks on another, with the order of proactive vs. reactive days counter-balanced across subjects. In the current work we do not consider the influence of cognitive-control mode and combine data for each task across scanning sessions, to increase statistical power.

4.3.6 Pre-processing and Parcellation

Raw resting-state and task data were preprocessed using the same pipeline, implemented with fMRI-prep software ([136],[137]). The whole-brain surface data were then parcellated into 400 cortical parcels defined by the 400 parcel Schaeffer atlas (Schaefer [6]; 7-network version). Subcortical volumetric data was divided into 19 regions derived from FreeSurfer ([89]). Motion time-series consisted of the 3-dimensional coordinate changes for rigid-body (brain) rotation and translation (6 total). Motion and linear drift were regressed out of pre-processed resting-state data before MINDy model fitting and from task data prior to filtering. Since motion time-series are also covariates within our task GLMs (as is common), this step does not bias results as motion is implicitly removed from the unmodeled data during GLM estimation (see below). However, we also implemented controls (see Sec. 4.3.9) which used this same data (i.e. motion pre-regressed) with conventional analyses.

4.3.7 Task GLM Analyses

Statistical models of task fMRI were estimated using general linear models (GLM) as implemented in AFNI. The same analyses were performed for both the original task data and the model-subtracted data. The GLM design consisted of a mixed block-task design in which trial-type effects were modeled using a modified Finite-Impulse-Response (FIR,[138], [139], [140]) framework (AFNI TENT; [141]), whereas block effects (task vs. inter-block interval) were modeled using a canonical HRF convolved with the block regressors. The TENT bases consisted of overlapping linear-interpolation splines spanning two TRs each. The FIR models were generated by projecting TENT coefficients by the mean TENT basis-set for each trial-type (within-subject). The GLM design also included block onset/offset (modeled with a canonical HRF) and the six motion regressors corresponding to rigid body translation and rotation (3 each). Timepoints containing excessive motion (Framewise Displacement > 0.9mm) were censored from task GLMs. Estimation was performed using the built-in AFNI function "3dREMLfit".

4.3.8 MINDy Modeling

Mesoscale Individualized NeuroDynamic (MINDy, [8][99]) models were generated from each subject using the parcellated, pre-processed resting-state data for each subject, combined across scanning sessions. Thus, a single MINDy model was estimated for each subject and was used in analyzing task-data across scanning sessions. We simultaneously estimated the neurovascular coupling/HRF and latent brain networks by combining the original MINDy model with Surrogate Deconvolution as in [99]. This combination simultaneously estimates HRF kernel parameters for each brain region as well as the connectivity matrix, region-specific transfer function shape, and local decay parameter (time-constant). Previous work indicates that the inclusion of Surrogate Deconvolution renders MINDy estimates robust to spatial variation in the HRF. Moreover, the spatial distribution of estimated HRF properties such as time-to-peak are consistent with empirical literature at the group level and are also reliable at the level of individual differences ([99]). Hyperparameters used in MINDy model fitting were identical to previous studies ([8]).

4.3.9 Control Pipelines

In addition to comparing the proposed pipeline with conventional analyses, we also repeated all task analyses for several control pipelines. These control pipelines considered two factors that might explain results: 1) pre-processing and 2) mechanistic components of the model. The MINDy modeling framework assumes that nuisance covariates such as motion and drift have already been removed from time-series prior to model fitting. Therefore, to address #1, we implemented a control in which standard GLM analyses were computed directly upon the fMRI BOLD task timeseries, with motion covariates already regressed out first. The same regressors also appear in the task GLM model (which is shared across all pipelines), but regressing these factors out first will rescale estimated beta-coefficients due to the input normalization performed by many fMRI processing packages (e.g. AFNI). This control ensured that improvements in group-level sensitivity were due to increased similarity of estimated spatiotemporal patterns rather than theoretically uninteresting factors due to pre-processing pipelines. We refer to this control as “pre-regressed” (preReg).

To address #2, we considered the influence of anatomically local dynamics vs. interactions between brain regions. This contrast is significant for three reasons. First, it is theoretically significant to distinguish between purely local neural dynamics and inter-regional brain dynamics. Secondly, long distance interactions between brain regions cannot be explained solely in terms of neurovasculature since the regions involved may share anatomically distinct blood supply (i.e. different cerebral arteries). As a result, improvements identified in whole-brain models, but not purely local models, cannot be explained solely as a benefit of hemodynamic modeling (although other contaminants such as motion could still be a factor). Lastly, analyses using the purely local models are equivalent to region-specific frequency-domain filtering. Although this equivalence does not imply that neural dynamics are insignificant, the signal-processing interpretation is simpler and could render the proposed neural modeling framework unnecessary (i.e. less parsimonious). Thus, the local dynamics control serves to ensure that our guiding neural modeling framework provides additional value above its (partial) relationship to existing signal-processing techniques. This control was implemented in two distinct variants: either heterogeneous (region-specific) or homogeneous (region-invariant) autoregressive models fit to each subject.

The homogeneous model consists of an autoregressive model that is specific to subject, but not parcel:

$$BOLD_{t+1} = \beta BOLD_t + \nu_t \tag{4.5}$$

We assumed that the noise-component was independent and identically distributed between parcels and solved for β using linear regression (collapsing BOLD across parcels). The "input" estimates from this model consist of the residuals (ν_t). We fit the heterogenous model analogously to the heterogeneous model, but with region-specific autoregressive terms:

$$BOLD_{t+1}^{(i)} = \beta_i BOLD_t^{(i)} + \nu_t^{(i)} \tag{4.6}$$

for parcel "i". We use these two cases to determine whether regional heterogeneity is a significant factor in any improvements due to local modeling. We refer to the homogeneous and heterogeneous models as global ("glob") and local ("loc") autoregressive (AR) models, respectively.

4.4 Validation and Comparison Criteria

In order to assess potential advantages of the proposed technique, we considered two types of comparisons: benchmarking (is method "a" better than "b"?), and sensitivity/robustness (how does factor "x" influence method "a" vs. "b"?). The first case establishes whether the proposed method offers additional statistical power in detecting task effects.

4.4.1 Benchmarking Sustained Effects

In the present work, we consider both sustained task-effects (block-related changes) and trial-type effects across four cognitive tasks commonly used to index cognitive control (AX-CPT [132], Cued Task Switching [135], Sternberg [133], and Stroop [134] tasks; see Sec. 4.3.3). Sustained effects refer to "background" activity that is present regardless of whether participants are performing a task. Since we used FIR models to span each trial type, sustained effects in our analysis *only* refer to activity during inter-trial periods (non-trial periods of task-blocks) since effects during other periods are absorbed in the trial vs. rest-block contrasts ([142], [143]). We compared the group-level effect size of each technique (the proposed method and several controls) in detecting sustained effects. Methods were compared pairwise, and analyses were only conducted on parcels which had a significant effect for either method in a pair. Sustained analyses considered both signal increases and decreases, so the target metric was absolute t-value (1-sample group test) for the GLM sustained betas (see Sec. 4.3.7).

4.4.2 Benchmarking Event-Related Effects

Trial-types were defined by high cognitive control demand vs. low cognitive control demand across the four tasks (see Sec. 4.3.4). Trial-specific activity was modeled using a Finite Impulse Response (FIR) model with 1TR resolution (1.2s) and task-specific length (see Sec. 4.3.7). Group-level statistics were compared for the peak effect (parcel × method specific) over a task-specific 2TR interval. This interval was chosen during study piloting using the peak times in conventional analyses. Thus, the analysis targets are statistically biased *against* the proposed technique since they were chosen to maximize conventional analyses. These times qualitatively correspond with a typical HRF time-to-peak after the probe-events which define high vs. low control trials (see Sec. 4.3.4). Previous literature and present results suggest that these effects are primarily one-sided, with activity increased in the high-conflict (control demand) trials relative to low-conflict (low control demand) in relevant brain regions. Conversely, task-negative effects (significant decreases) have largely been associated with sustained signals as opposed to high vs. low control events. For these reasons, we only considered significant increases in activity for trial-type analyses. Group-level t-tests (within parcel) were compared for all parcels with significant increases (either method), or for a set of 34 parcels (pre-defined from independent conventional analyses which showed consistent control-demand effects across all tasks, [144]). Since these parcels were pre-selected based upon conventional analyses, they are statistically biased *against* the proposed method (i.e. in favor of conventional methods).

4.4.3 Testing Selective vs. Global Improvements

We further analyzed benchmarking results by testing how MINDy-based Filtering changes the distribution across parcels. The primary question was whether the MINDy-based Filtering: a) uniformly changes statistical power across the brain (by shift or scale); b) primarily identifies previously insignificant regions or c) primarily alters previously significant regions. This analysis is important for determining whether the technique globally improves statistical power or differentiates task-relevant regions from the rest of the brain. We test these effects using multilevel linear models to compare MINDy-based Filtering to the different control models. These multilevel models (presented in more detail later) contain task main effects (anatomically global) and additional terms for task-implicated (statistically significant parcels). We use these models to test the significance of model improvements (increased effect sizes) after discounting anatomically global changes.

4.4.4 Sensitivity to Cognitive States

Sensitivity analyses were performed to assess the impacts of cognitive states, individual differences, and motion. In the current case, cognitive states differ between tasks and trials. Although, each of the four tasks are commonly used to index cognitive control, cognitive tasks are not construct-pure. For instance, tasks featuring delays (AX-CPT, Cued Task Switching, and Sternberg) are thought to be more dependent upon working memory than those without delays (i.e. the Stroop task). However, many task-specific factors are the same between high and low control trials of the same task (i.e. all events prior to the probe). Thus, we control for cognitive similarity across tasks by comparing results across increasing levels of cognitive similarity: low-control trials, high-control trials, and the contrast high vs. low control trials. These levels increasingly isolate the cognitive control construct by increasing control demand (high-control trials) and controlling for other task events (high vs. low contrast). Methods which are sensitive to cognitive states will produce more similar results between task contexts when the cognitive states measured are more similar. Thus, we tested whether the proposed technique increased similarity between tasks for the high vs. low contrast relative to low-control trials. We measured this using Intraclass Correlation (ICC; [95]). Tasks differed in effect magnitude and there was no theoretical basis for assuming this factor should be identical between tasks (i.e. we don't assume each task equally taxes cognitive control), so we normalized the group-average data (divided by root-sum-of-squares) for each task × method before applying ICC. Significance was established using boot-strapping with new distributions generated for each comparison.

4.4.5 Sensitivity to Individual Differences

We also analyzed the degree to which each method was sensitive to individual differences. This analysis also used ICC across the different levels/contrasts of control demand. However, the data of interest consisted of individual differences, which were defined by z-scores relative the group (i.e. normalized deviations from the group tendency). We computed ICCs separately for each DMCC34 parcel and determined significance using parametric statistics on the distribution over parcels.

4.4.6 Robustness to Motion

Lastly, we compared methods in their robustness to motion confound. While previous work has established that the model-fitting technique (MINDy) is robust to motion ([8]) it remains unknown whether the proposed MINDy-based Filtering technique also exhibits similar motion robustness. Therefore, we compared methods in terms of sensitivity to motion artifact. We considered three motion metrics for task data including the number of frames censored based upon framewise-displacement (FD) criteria ($< 0.9mm$), the median framewise displacement, median-absolute-deviation (MAD) of DVARS ([86]). We analyzed sensitivity by comparing the similarity (ICC) of results between high-motion and low-motion groups of subjects (median split for each motion measure).

4.4.7 Significance Testing for Construct Identification

We used permutation statistics to compare the significance of generalizability tests between methods. When testing the generalizability of group-level patterns, we treated brain regions as the object of measurement in intraclass correlations (ICC, [95]) over task classes and estimated confidence intervals with bootstrap sampling over the set of brain parcels. We defined individual differences in terms of z-scored data relative the group (computed task × parcel × method). We tested significance via permutation testing z-scored data was randomly permuted between methods and then z-scored again before ICC computation (thereby maintaining the within-task distribution). Confidence intervals were computed with bootstrapped sampling over subjects.

4.5 Results

4.5.1 Structure and Presentation of Results

We designed analyses to answer four questions: 1) do resting-state MINDy models (partially) generalize to task? 2) does the proposed technique improve power in answering cognitive-neuroscience questions? 3) can these methods test hypotheses which were previously impractical? and 4) do improvements reflect theoretically interesting concepts (e.g. signal propagation) or do they stem from signal-processing/filtering side-effects? The first question resolves whether the intrinsic dynamics modeled at rest meaningfully generalizes to task (although not perfectly, as we are interested in the task versus rest differences). The second and third questions identify methodological contributions, whereas the latter determines whether techniques also offer additional theoretical insight (i.e. their success reflects some principle of brain function). This question is important for determining whether results reflect

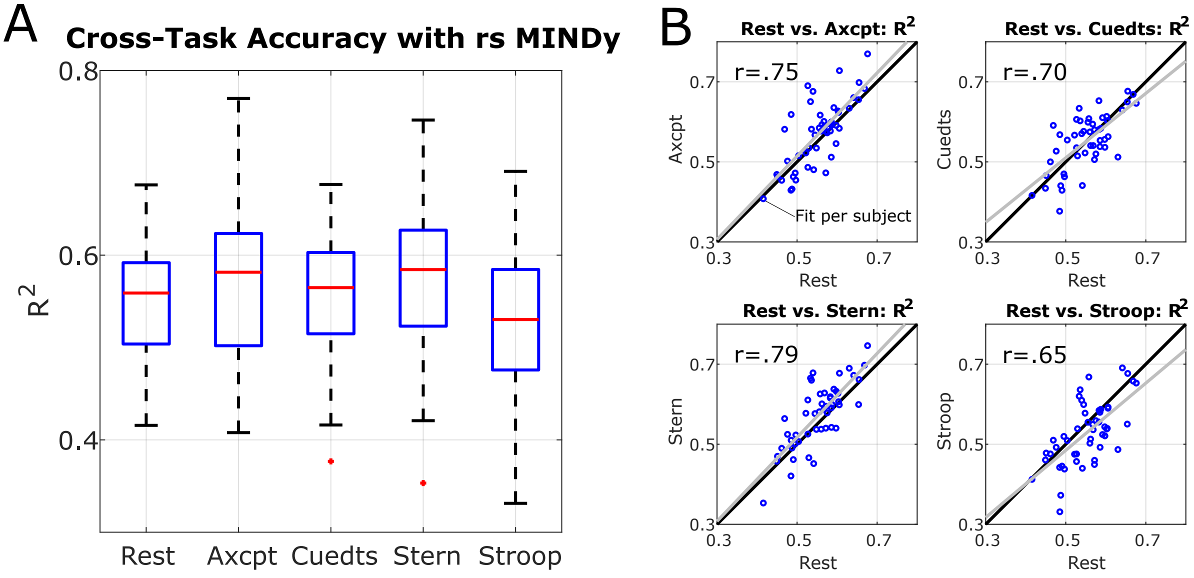

Figure 4.1: Resting-state MINDy models generalize to task. A) MINDy models trained on resting-state data produce similarly accurate predictions for task data. Goodness-of-fit is quantified in terms of the mean R^2 value across all brain parcels ([6]) and scanning runs (n=6) in predicting the difference time series: $X_{t+1} - X_t$ ([8]). B) Individual-differences in model accuracy are highly correlated between resting-state and task data.

the activity-flow framework or can be more parsimoniously explained in terms of (non-neural) signal processing effects.

4.5.2 Resting-state Model Predictions Generalize to Task

We first test whether the proposed technique actually serves as a conceptual "filter" in removing intrinsic-dynamics from task, rather than making this information more salient (as would occur when there is little overlap between task and rest). Our framework assumes systematic discrepancies between task and resting-state (i.e. we are interested in the difference between contexts), but we assume that there is some overlap of task and resting-state dynamics for us to "remove". In statistical terms, we first ensure that the approach removes variation from task data (associated with covariance of task-rest dynamics) rather than adding additional variation as would occur when subtracting independent factors. To test this possibility, we first quantified the goodness of fit for resting-state models (MINDy; [8], [99]) in predicting both task and resting-state data within subject. Goodness-of-fit was calculated using both the final model and cross-validated models which are less biased in favor of resting-state fits. Paired (within-subject) analyses did not indicate greater goodness-of-fit for resting-state scans ($R^2 = .55 \pm .06$) relative to task scans (mean R^2 across tasks=$.56 \pm .07$, paired-$t(49) = -.50, p = .62$; Fig.4.1A). Moreover, individual differences in goodness-of-fit were consistent across tasks (Fig. 4.1B), which indicates that model accuracy is also highly preserved within-subject. We conclude that the short-term evolution of brain activity is similar (but not identical) in resting-state and various task contexts. By leveraging large-scale resting-state models (MINDy) the proposed technique filters out intrinsic dynamics common to resting and task state.

4.5.3 MINDy-based Filtering Accounts for intra and inter-subject Variability

We also tested whether these intrinsic dynamics explain unique variability above the task GLM. This test is important for determining whether MINDy serves to predict the mean brain-response for each trial-type or whether it also predicts trial-to-trial variability. We quantified these properties through sum-of-squares partitioning (ANOVA). Across all tasks, we found that the proportion of unique variance explained by MINDy was significant (42.3% on average, Fig. 4.2A). However, MINDy predictions and the task effects do have some overlap (a non-zero MINDy×task sum-of-squares, Fig. 4.2A), thus MINDy predictions account for some of the variation in both the trial-to-trial variability (variation unique to MINDy) and the typical response across trials (MINDy × task interaction). We also tested how MINDy-based Filtering impacts variability in the evoked-response between subjects. We restricted these analyses to the pre-defined set of regions (the DMCC34 parcels, [144]) which were previously

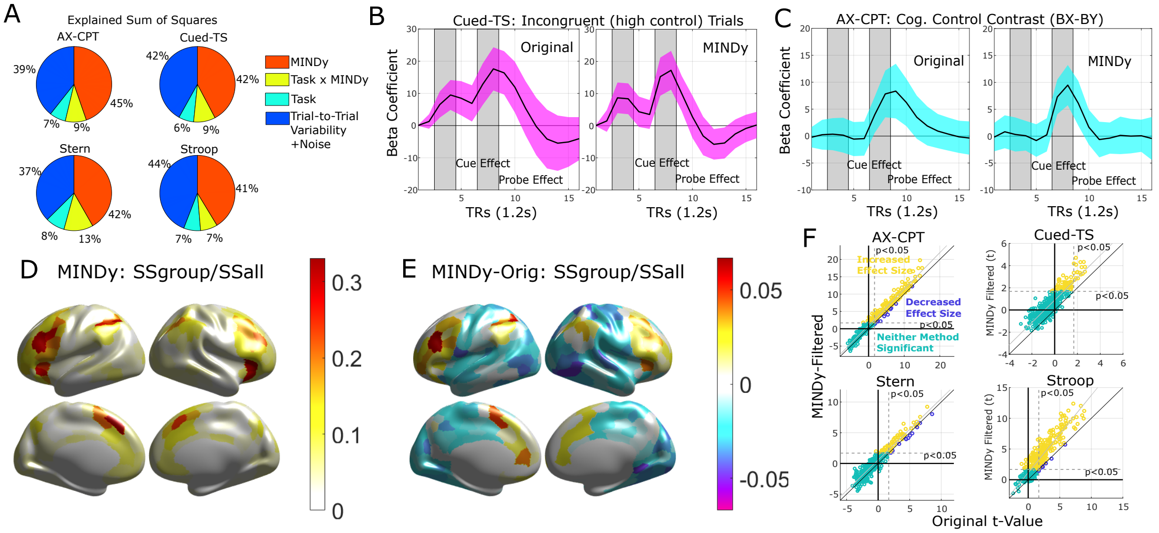

Figure 4.2: MINDy-based Filtering reduces variability within and between subjects. A) MINDy-based Filtering accounts for a significant portion of unique variability within each subject's data. This effect holds across tasks (results averaged over all parcels, subjects). Variance partitioning was performed after removing variation due to nuisance factors (motion and drift). B) MINDy-based Filtering reduces the between-subject variability of task-evoked signals. Example shown is the mean signal over the DMCC34 parcels for the Cued-TS high control-demand condition (incongruent trials). C) Variability also decreases for contrasts between conditions. Example shown is for the AX-CPT (BX-BY contrast). D) Spatial-map for the proportion of variability in cognitive control signals attributable to group (collapsed across tasks) for MINDy-filtered data. E) Difference in the relative group-explained variability between MINDy and the original data. Note that MINDy-based filtering actually decreases the proportion of group variance in some regions, but increases for task-implicated regions (lPFC, anterior insula, etc.). F) MINDy-based Filtering increases the group-level effect size for each task. This can be seen in the number of parcels exhibiting higher t-values after MINDy-based Filtering relative to conventional analyses (i.e., above the identity line). Yellow dots indicate significant parcels (in terms of the control-demand effect) which also had increased effect sizes from MINDy-based Filtering, while blue dots denote significant parcels whose effect sizes were larger with conventional analyses. Teal dots denote parcels which did not exhibit a significant control-demand effect for either method.

identified as having a significant control-demand effect across tasks. Results demonstrated that MINDy filtering decreased inter-subject variability in both main effects of trial-type (e.g. Fig. 4.2 B) and the contrast between trial-types (e.g. Fig. 4.2C). In particular, these analyses and associated event-related timecourse visualizations reveal that the peak task-related effects become sharper (more well-defined) as well as more temporally-precise after MINDy-based filtering. We used ANOVA to partition variance in the cognitive control effect into group-level variance and individual variance over the relevant trial periods.

We then tested whether MINDy increased the proportion of cognitive control effects attributed to a common group factor (sum-of-squares for the group effect divided by the total effect). As expected, regions implicated in cognitive control, such as the lateral and medial prefrontal cortex, anterior insulae, and posterior parietal cortex, had larger proportions of variability explained by the common group factor (Fig. 4.2D). Interestingly, although MINDy-based Filtering increased group variability (decreased inter-subject variability) in many of these same task implicated regions it decreased the common group factor for regions not implicated in cognitive control (Fig. 4.2E). In particular, the proportion variance explained by a common group effect increased across the DMCC34 parcels (paired-$t(33) = 6.03, p < 9E-7$). Thus, by removing individual-differences in intrinsic brain dynamics, MINDy-based Filtering reveals more similar task-effects between subjects. As a result, MINDy increased cognitive control detection power in all four cognitive tasks considered. Figure 4.2F visualizes this effect for all parcels across the whole brain.

4.5.4 Improved Group-Level Detection Power

We tested whether MINDy-based Filtering improves statistical power in detecting group-level neural effects. We considered two effects of interest: an event-related contrast between

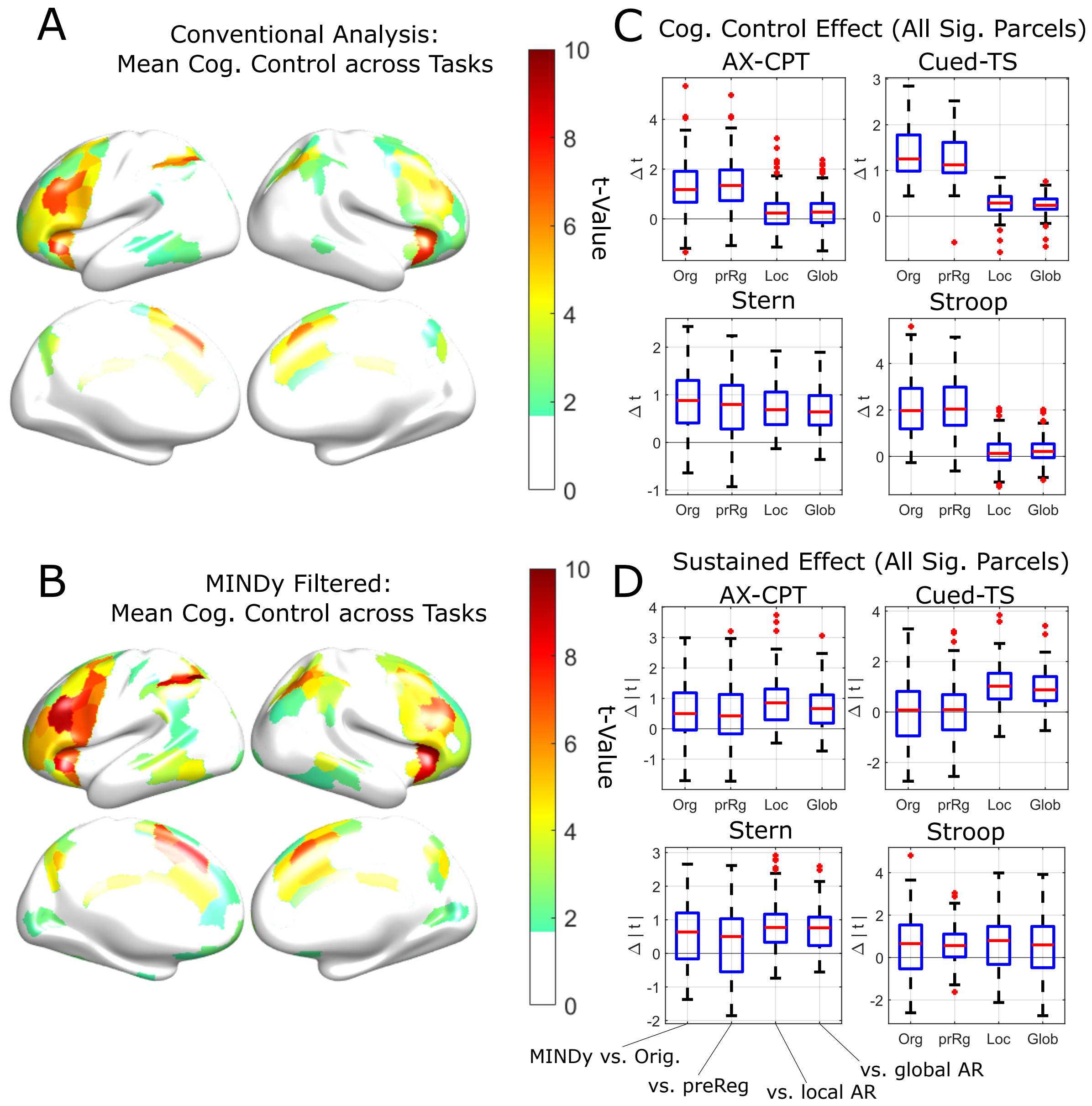

Figure 4.3: MINDy-based Filtering increases power in detecting event-related and sustained task effects. A) Spatial distribution of high-vs.-low conflict effects averaged across tasks using conventional analyses and B) the proposed technique. Thresholds correspond to $p < .05$ for a single task. C) Paired comparisons of effect size between control methods and the proposed approach in identifying high-vs.-low conflict effects for significant brain parcels. Values less than zero indicate that the propose technique improves upon controls. D) Comparisons for the absolute sustained effect over significant parcels (positive or negative). Parcels were deemed significant if they passed $p < .05$ for either method within a pair. Comparison techniques are denoted:“Orig”=original analysis (no filter), “Pre-Reg”=motion and drift pre-regressed before GLM fitting, “Loc”=AR model filtering with local/heterogeneous parameters, “Glob”=AR model filtering with a single (anatomically global/homogeneous) parameter, and “MINDy”=MINDy-based Filtering.

trial-types (high vs. low cognitive control demand; see Sec. 4.3.4) and sustained neural activity during task blocks. In both cases we employed Finite Impulse Response (FIR) designs to model task effects (trial-type specific for GLMs in event-related analyses but trial non-specific for sustained analysis GLMs). Additional regressors included motion, drift, and the start/end of task blocks (see Sec. 4.3.7). By contrast, sustained effects consist of "background" activity that is present during inter-trial periods (i.e. non-trial periods of task-blocks) since the effect of task vs. rest-block is already absorbed in the FIR trial-type regressors ([142], [143]). We compared results (group-level one-sample t-tests) between methods (MINDy-based Filtering, conventional analyses, and controls) for each task and an omnibus test across tasks (Fig. 4.3A,B). For each event-related pairwise comparison of methods, we tested the change in effect-size (group t-value) for parcels demonstrating a significant increase ($p < .05$, uncorrected) for either method within a pair. Results indicate that MINDy-based Filtering significantly increased statistical detection power on all tasks (four of four) for the event-related contrast relative to all other controls (all p‘s≤E-7; Fig. 4.2F, Fig. 4.3 C). Sustained effects were also magnified for each task relative the two auto-regressive controls (all p‘s<E-10; Fig. 4.3 D). However, sustained effects during Cued-TS did not significantly improve relative conventional analyses (paired-$t(147) = .32, p = .75$) or the pre-regressed control (paired-$t(141) = .60, p = .55$). Sustained effects for all other tasks did significantly increase (all $p‘s < .004$). Thus, the proposed technique increased statistical power in 4 of 4 event-related analyses and 3 of 4 sustained analyses relative to conventional approaches. These increases were due, in part, to considering inter-regional signaling as reflected in the improvements over models which did not consider connectivity (the auto-regressive models). Statistical power increased in 4 of 4 event-related analyses and 4 of 4 sustained analyses relative to both homogeneous and heterogeneous autoregressive models.

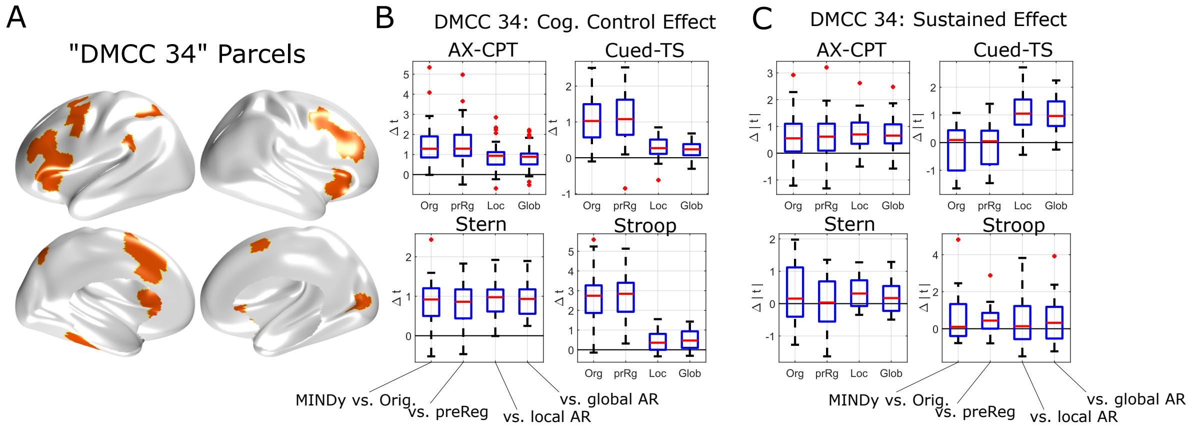

Figure 4.4: Improved detection power over a set of 34 parcels with consistent event-related effects A) 34 parcels were identified as exhibiting a consistent cognitive control effect across tasks during study piloting. B) Paired comparisons of the cognitive control effect over the DMC34 parcels: MINDy-based Filtering minus control. C) Comparisons for the absolute sustained effect the DMCC34 parcels: MINDy-based Filtering minus control. Values greater than zero indicate improvement. Comparison techniques are denoted:“Orig”=original analysis (no filter), “Pre-Reg”=motion and drift pre-regressed before GLM fitting, “Loc”=AR model filtering with local/heterogeneous parameters, “Glob”=AR model filtering with a single (anatomically global/homogeneous) parameter, and “MINDy”=MINDy-based Filtering.

For omnibus analyses, we collapsed observations across tasks (Fig. 4.3A,B). Results indicated that the proposed technique generally increases detection power for event-related analyses relative to all controls (vs. original: paired-$t(635) = 35.9$, vs. pre-regressed:$p \approx 0, t(638) = 36.8, p \approx 0$, vs. local AR: $t(662) = 12.8, p \approx 0$, global AR: $t(654) = 14.9, p \approx 0$). Detection power for sustained effects also increased relative to controls ($paired - t(680) = 9.7, p \approx 0; t(672) = 10.4, p \approx 0; t(623) = 21.5, p \approx 0; t(630) = 18.5, p \approx 0$). We conclude that the proposed techniques improve group-level detection of event-related and sustained task effects and these improvements are due, in part, to modeling the interactions between brain regions (i.e. results are not accounted for by univariate auto-regressive models).

One limitation of the previous tests, however, concerns the determination of which parcels are included in analysis: we compared effect sizes in parcels that were statistically significant (i.e.

large effect sizes). This approach is anatomically parsimonious in that the comparison regions are informed by data rather than prior assumptions. However, this dependency could produce biases due to differences in higher-order features (e.g. overdispersion) between methods. Therefore, we repeated the previous analyses over a fixed set of 34 pre-specified brain parcels (Fig. 4.4A) that demonstrated significant increases due to cognitive conflict (event-related contrast) across all four tasks during independent and pre-specified analyses (see Methods, [144]). The implicated parcels agree with previous studies mapping the neuroanatomy of cognitive control and are largely located along lateral prefrontal cortex and anterior insula (Salience/Ventral Attention and Control networks; [90][6]). Analyses over this restricted, pre-specified group of parcels agreed with the previous results: the omnibus (all task) statistical detection power improved relative controls for events (maximum $p = 4.1E-5$; Fig. 4.4B) and sustained effects (maximum $p = .0015$; Fig. 4.4C). Thus, results indicated that MINDy-based Filtering improved statistical detection even when analyses were restricted to this group of 34 pre-specified parcels.

4.5.5 Identifying Individual Differences in a Latent Cognitive Construct

The previous analyses indicate that the proposed techniques enhance the identification of neural activity associated with a set of contrasts between trial-types (theoretical high control-demand trials minus low control-demand trials). However, many cognitive neuroscience studies seek to understand cognitive constructs, as opposed to unitary tasks. In the current section, we explore how well each method identifies the neural correlates of one such construct: cognitive control. The four tasks we studied have all been previously used to index cognitive control (typically via the difference between high-conflict and low-conflict trials). However,

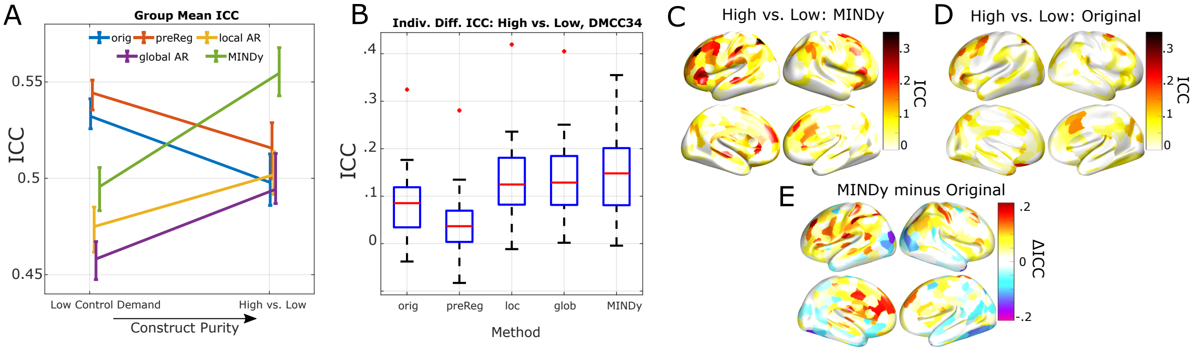

Figure 4.5: MINDy-based Filtering improves identification of cognitive control demand. A) Cross-task generalizability (ICC) for group means in the low-conflict condition (low construct purity across tasks) and the high vs. low conflict contrast (high purity) by method. B) Generalizability of individual differences (z-scores) for the "DMCC34" parcels. Boxplots display distributions across parcels. C) Anatomical distribution of generalizability for the proposed technique and (D) conventional analyses for the high vs. low conflict contrast. E) The difference in cross-task generalizability between MINDy-based Filtering and conventional analyses. Comparison techniques are denoted:"Orig"=original analysis (no filter), "Pre-Reg"=motion and drift pre-regressed before GLM fitting, "Loc"=AR model filtering with local/heterogeneous parameters, "Glob"=AR model filtering with a single (anatomically global/homogeneous) parameter, and "MINDy"=MINDy-based Filtering.

because the tasks themselves are not construct-pure (i.e. they tap multiple cognitive constructs) the neural activity associated with tasks is also expected to be non-identical. To control for this fact, we used the different trial types to generate levels of "construct-purity" in terms of cognitive control: low conflict trials (low purity) and the high-vs.-low contrast (high purity). We consider the high-vs.-low contrast to be more "construct-pure" in terms of cognitive control since it controls for many of the other cognitive processes that differentiate tasks. For instance, speech production (unique to the Stroop task), is identical between high and low-conflict trials (the same set of words are produced). Likewise, working memory maintenance during delays (Sternberg, AX-CPT, and Cued-Task Switching) does not differ between high and low control-demand trials since these trial-types are identical through the delay period (up until the probe).

We tested how sensitive each approach was to the cognitive control construct via the relationship between “construct-purity” and cross-task similarity of neural effects. For this test, we indicate that a measure is “sensitive” to a factor (cognitive constructs) if the similarity in measurements reflects the similarity in that factor. We therefore consider a measure “sensitive” to cognitive constructs if it reports higher similarity between tasks for the high “construct-purity” condition (high-vs.-low control demand contrast) than for the low “construct-purity” condition (low demand trials).

We quantified “similarity” across the four tasks using Intra Class Correlation (ICC, [95]) and performed analyses in terms of both the group-average and individual-differences. For group-average analyses, ICC “units of observation” consisted of the mean beta for each brain parcel (all 419 brain regions) and “classes” consisted of the different tasks. Results indicated that the proposed technique was sensitive to the cognitive control construct at group level (Fig. 4.5A). In the “low-purity” condition, the proposed technique identified significantly lower similarity between tasks ($ICC = .50 \pm .012$) than conventional approaches (both p‘s $<$ .002, 5,000 bootstraps), although more than the autoregressive models (both p‘s $<$ .002, 5000 paired bootstraps). Thus, the proposed technique does not generically increase the similarity of task results irrespective of cognitive construct. By contrast, for the “high-purity” condition, the proposed technique generated significantly more similar results across tasks ($ICC = .555 \pm .013$) than either conventional analyses or local modeling (all p‘s $<$ .002, 5000 paired bootstraps). We conclude that the proposed technique improves sensitivity to the cognitive control construct at group-level. Based on the nature of how these ICCs were calculated, the finding can also be interpreted as indicating that the anatomical profile of effects (i.e. the gradient of effect sizes across the brain) becomes more similar or consistent across tasks after MINDy-based filtering, relative to conventional analyses, and the controls.

We next tested the consistency of individual differences. For these analyses we normalized data between subjects for each method × task × parcel before computing ICC separately for each parcel (using subjects as the "units of observation"). In general, ICC values were highest for task-implicated regions (e.g. lateral prefrontal cortex and anterior insulae; Fig. 4.5C,D) and the proposed technique demonstrated particularly high generalizability at these locations relative methodological controls (FIg. 4.5E). We compared methods in terms of ICC for the DMCC34 set of parcels (Fig. 4.5B). Results indicated that the proposed technique improves identification of individual differences over conventional techniques for the high vs. low control contrast (vs. original: $paired-t(33)=5.5, p=4.4E-6$; vs. pre-regressed: $t=10.0, p=1.8E-11$) but did not significantly differ from filtering with autoregressive models (vs. local AR: $paired-t(33)=1.4, p=.18$; vs. global AR: $t=1.0, p=.32$). As with the group-level analysis, generalizability did not increase over conventional analysis for the low construct-purity (low demand) condition: (vs. original: $paired-t(33)=.220, p=.828$, vs. pre-regressed: $t=-2.17, p=.037$). Thus, the proposed technique improves the estimation of neural individual differences related to a cognitive construct, but do not artificial increase generalizability across cognitively dissimilar task conditions. MINDy-based filtering increased the generalizability of the group-level activity profile across tasks relative to both conventional techniques and the reduced, autoregressive models. Individual differences also became more generalizable after model-based filtering but did not depend upon the choice of model (MINDy vs. autoregressive).

These results also offer theoretical (exploratory) interpretations. The inter-task variability of conventional techniques and MINDy-based Filtering are linked since "input" is defined as the difference between the observed brain activity and the propagation along intrinsic dynamics ($x_{t+1} = f(x_t)+I_t$). Consequently, this section's results suggest that cognitive control signatures are most similar between tasks when they first impact the brain (as "inputs") but lead to more task-specific patterns of activity (x_t) as they evolve according to intrinsic brain dynamics. Group-average signatures of cognitive-control generalized better for the full MINDy model than for models without connectivity, indicating that this evolution occurs, in part, due to signaling between brain regions (i.e. the 'post-synaptic' activity is more task-specific than the 'pre-synaptic' component). By contrast, individual differences are similarly generalizable regardless of model-choice. This pattern of results suggest that individual differences become task-specific as they propagate through local dynamics (i.e. are integrated over time) as opposed to propagating through connectivity. Of course, these interpretations are post-hoc and mainly serve to demonstrate the potentially utility of MINDy-based Filtering. Future studies should explore these possibilities in more detail.

4.5.6 MINDy-based Filtering Selectively Enhances Task-Related Neural Signals

The previous analyses indicate that the proposed technique can potentially increase the spatial power and temporal resolution of neural-cognitive linkages. These properties demonstrate that the approach may improve neuroscientific analyses (i.e. the approach is useful as a *tool*). However, a method's success does not require that the motivating framework be useful (or even correct). For instance, the homogeneous autoregressive control is a special case of the MINDy model which neglects regional heterogeneity, nonlinearity, and interregional signaling. The proposed technique, when applied as a homogeneous autoregressive model (instead of MINDy) reduces to a spatially-univariate frequency-domain filter. Although the general approach is motivated by a neural framework (activity propagation), results for this

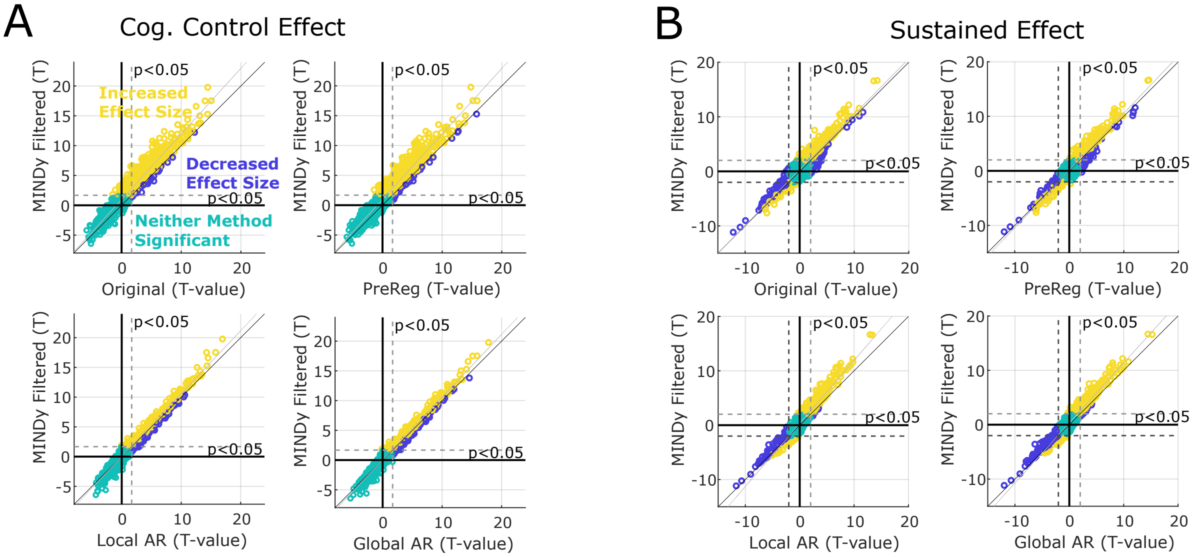

Figure 4.6: MINDy-based Filtering magnifies task-related signals relative to all controls. A) Scatter plots compare methods in terms of group-level T-values for the cognitive control effect collapsed across tasks (all tasks plotted together). Unlike Fig. 4.2F, each plot corresponds to one of the four control methods combined across tasks (instead of one plot per task). In particular, the upper left corner plots all of the points from Fig. 4.2F combined. Improvements can be seen in the number of parcels exhibiting higher t-values after MINDy-based Filtering relative to conventional analyses (i.e., above the identity line). Yellow dots indicate significant parcels (in terms of the control-demand effect) which also had increased effect sizes from MINDy-based Filtering, while blue dots denote significant parcels whose effect sizes were larger with conventional analyses. Teal dots denote parcels which did not exhibit a significant control-demand effect for either method. B) Same as (A) but for sustained effects.

reduced model are more parsimoniously explained in terms of signal-processing effects. We tested whether the proposed technique can be similarly reduced by comparing results using a full brain model (MINDy; [8][99]) compared with the result of applying our technique to local, linear brain models. We considered local models that were either homogeneous (one autoregressive coefficient per subject) or heterogeneous (one per region). The improvement of using a full neural model over these controls indicates the benefit of using a detailed dynamical systems model (i.e. including connectivity).

Earlier results also indicate that the proposed technique increases the statistical detection power of task effects (Fig. 4.2F). Statistical power and effect sizes are useful benchmarking criteria as they are easy to interpret and relate to potential applications. However, these markers are also limited in that they indicate the ability to reject a generic null hypothesis of no task effects, but do not generate scientific insight as to the source of the effects. For instance, approaches which magnify anatomically global effects may provide little benefit to functional "brain-mapping" studies which derive insight from the differentiation between brain regions. Therefore, we tested whether improvements under the proposed technique are anatomically global or serve to further differentiate regions.

We consider two sorts of global effects: additive "shifts" in the global signal and global "scaling" of task effects. In statistical modeling terminology, the former reflects a main-effect (intercept) of method, whereas the latter reflects the method-specific slope. We modeled the differentiation between brain regions as either a main effect of regional significance (i.e. whether a region has a significant effect) or as an interaction with regional significance reflecting either a shift or rescaling of effect sizes of significant regions due to MINDy-based filtering, relative to the control models. We use the logical-valued variable $Sig_{task,Parc}$ to denote whether a parcel exhibits a significant effect for either method in a given second-level task analysis.

$$X_{task,Parc} = \beta_{task} + \beta_0 Y_{task,Parc} + Sig_{task,Parc}(\beta_1 + \beta_2 Y_{task,Parc}) + \epsilon_{task,Parc} \tag{4.7}$$

We assume that ϵ is identically distributed across tasks and parcels. The coefficient β_1 represents the main effect of parcel significance as a binary factor, while β_2 represents the interaction with parcel effect size in control methods. Conceptually, these two components represent the degree to which the proposed technique further separates task-implicated and non-implicated parcels and the degree to which task-implicated are further magnified, respectively.

Results indicate that the proposed technique demonstrates differential sensitivity, in that improvements are greater in task-implicated regions (Fig. fig:VarPart F, 4.6A, B). The main effect of event-related regional significance was significant in all cases but larger for the original and pre-regressed models (t(1669)=12.73, p ≈0, t=12.94, p ≈0, respectively) than for the local and global AR models (t=2.16, p=.03, t=2.45, p=.014; Fig. 4.2 F, 4.6A, B). This result indicates that the proposed technique further separates event-implicated and non-implicated regions rather than simply increasing global statistical features. Moreover, the proposed technique also differentially magnified effect sizes relative to the original and pre-regressed cases (t(1669)=6.25, p=5.2E-10, t=4.87, p=1.2E-6, respectively). However, the proposed method did not significantly magnify effect sizes over AR control models ($t = .42, 1.47$ for local and global, respectively). Thus, differences among significant regions are magnified by filtering out local-dynamics (MINDy and AR models), while filtering-out connectivity driven dynamics (MINDy but not AR) further differentiates task-implicated regions from insignificant ones.

Similar results held for analyses of sustained effects. As before, we considered bidirectional effects for sustained analyses (i.e. parcels with significant increases or decreases in sustained activity). For this reason, we slightly modified Eq. 4.7 to model improvements in terms of magnitude rather than a linear main effect:

$$X_{task,Parc} = \beta_{task}+\beta_0 Y_{task,Parc}+Sig_{task,Parc}(\beta_1 sign(Y_{task,Parc})+\beta_2 Y_{task,Parc})+\epsilon_{task,Parc}. \quad (4.8)$$

Note that the coefficient β_1 is now multiplied $sign(Y_{task,Parc})$. Results for sustained analysis mirrored those of the event-related analysis. The proposed technique differentially increased effect sizes over task-implicated parcels when compared to both conventional controls (t(1669)=5.48, p=5.0E-8 vs. original and t=6.24, p=4.8E-10 vs. pre-regressed) and these increases were even greater relative AR controls (t(1669)=10.87, $p \approx 0$ vs. local and t=12.93, $p \approx 0$ vs. global). Sustained effects were also differentially magnified effect sizes relative to conventional analysis (t=7.60, p=5.1E-14), pre-regressed controls (t=6.00, p=2.46E-9), and local AR models (t=2.26, p=.023), but not global AR models. Together, this pattern of results suggest that the proposed technique significantly differentiates between task-implicated and non-implicated regions for both sustained and event-related effects. The proposed technique distinguishes between task-implicated and non-implicated regions (the main effect of significance) and further differentiates among highly implicated and weakly implicated regions (the interaction-term). However, the latter effects do not improve above auto-regressive filters. Improvements over conventional techniques were greatest in the identification of event-related effects, whereas improvements over autoregressive controls were greatest for identifying sustained effects.

4.5.7 Sensitivity to Motion

Lastly, we compared the sensitivity of approaches to motion artifact. For each task and scanning session we computed three motion statistics: the number of frames censored due to passing a critical value of framewise displacement, the median framewise displacement and the median DVARS statistic ([86]) for each task run and averaged over runs. We then used resampling to test the relationship between each motion variable and the group effect-size of the high-vs.-low conflict contrast and sustained effect for each task. In brief, we randomly drew 5,000 samples of 30 subjects each without replacement. We computed group-level statistics for motion and the cognitive control contrast and then tested whether the average motion or variability of motion (inter-subject) of a sample predicted the sample's group-effect (one-sample t-scores averaged over the 34 parcels). We also used the same technique for predicting the difference between methods (i.e. do improvements under our approach require low motion?). Results did not indicate a significant effect of motion for the current dataset and subject pool. The relationship between motion and the difference between methods (MINDy versus original averaged over tasks) was insignificant for event-related analyses and did not display a consistent sign (proportion of frames censored: $r = .112$, FD: $r = .056$, DVARS: $r = -.028$). Likewise, we did not observe differential sensitivity to motion in the sustained effects (frames censored: $r = .0168$, FD: $r = .016$, DVARS: $r = .013$). Thus, the degree to which MINDy-based Filtering improves upon conventional methods is not influenced by motion within reasonable bounds.

4.6 Discussion

We demonstrated that MINDy-based Filtering increases the ability to detect both event-related (cognitive control-demand) and sustained brain responses in task fMRI (Sec. 4.5.4).

These effects are strongest in task-implicated brain regions (Sec. 4.5.6) and generate higher temporal precision than the original BOLD timeseries. By accounting for intrinsic dynamics, MINDy-based Filtering accounts for trial-to-trial variability within subject and variability between subjects (Sec. 4.5.3). However, while the absolute magnitude of subject-to-subject variability decreased, individual differences (and group—level activity) in a latent cognitive construct (control-demand) generalized better between tasks after MINDy-based Filtering (Sec. 4.5.5). These results suggest that MINDy-based Filtering can enhance the detection of task-evoked brain activity.

4.6.1 Relationship with Frequency-Based Filtering

Frequency-based (spectral) filtering has been applied to fMRI signals in many previous studies. High-pass filtering is commonly applied to both resting-state and task data to remove signal drift which is thought to largely reflect changes in non-neuronal variables. Low-pass filtering is also sometimes applied, primarily for resting-state data. Although these approaches were common in early fMRI experiments ([145],[146]), the changing nature of fMRI acquisitions (e.g. TR length) and analyses (e.g. functional connectivity) has led to renewed debate over these techniques ([147]) and the development of more sophisticated methodologies (e.g. [148],[149]). In the current work, we did not perform spectral filtering (instead using AFNI's "polort" function for polynomial basis de-drifting) and the proposed technique is not a direct replacement for spectral filtering which can be applied before our technique, afterwards or not at all. However, as previously mentioned, when the connectivity parameter of our model is zero, the proposed technique reduces to a form of spectral filtering based upon autoregressive models. Empirically we have demonstrated that MINDy-based filtering outperforms filters based upon autoregressive models (Sec. 4.5.4, Fig. 4.3 C,D) so effects cannot be attributed solely to removal of particular frequency components within each region.

Notably, the proposed technique improves detection in both sustained and event-related analyses over both conventional methods and autoregressive filters. By contrast, filters based upon autoregressive models are expected to underperform in the identification of (low-frequency) sustained effects as we confirmed in our analyses (Sec. 4.5.4, Fig. 4.3D). At a statistical-level, dynamical systems models (including MINDy) capture the multivariate partial autocovariance between successive time-points (i.e. how x_{t+1} is related to x_t). As a result, removing these predictions from the training data (Rest) inherently yields a timeseries with lower autocovariance. The improved detection of sustained effects is therefore significant as it indicates that the proposed technique reveals systematic differences between the resting-state and task dynamics rather than simply acting as a high-pass filter. These effects are also more pronounced in task-implicated parcels (Sec. 4.5.6, Fig. 4.6A,B) indicating that these features are also context-related.

4.6.2 Relationship with other approaches

The current approach is conceptually related to several current initiatives for linking resting-state and task-state brain activity. Our approach uses resting-state brain dynamics to extrapolate patterns of intrinsic dynamics that also factor into brain activity during task states. Frameworks such as Activity Flow ([43]) have demonstrated similarity between the spatial aspects of evoked responses and resting-state network structure. Likewise, functional connectivity patterns have been found to be roughly similar between resting-state and task ([30]). However, whereas these frameworks are largely employed to discover similarities between spontaneous and evoked activity, we analyze the manner in which task-state deviates from resting-state activity over short time-scales.

Other approaches have also investigated the difference between brain dynamics in task-state and resting-state. Previous work ([48], [46]) has demonstrated that intrinsic dynamics shape task-evoked activity on a trial-by-trial basis and modeling studies have reproduced the statistical differences between task and resting-state activity ([16]). Our approach furthers these efforts by leveraging these underlying concepts into an empirical modeling/analysis framework.

Dynamic Causal Modeling (DCM, [52]) frameworks have also used empirical dynamical systems models to improve estimates of task effects. As previously mentioned (Sec. 4.1.2), DCM techniques allow task effects to manifest changes in the exogeneous drive to brain regions and (for small-scale DCMs) the effective coupling between brain regions. By contrast, the current technique only models a single factor: changes in the input to each brain region which collapses both of these mechanisms into a single term as is common in larger-scale DCM models (e.g. [76]). Our approach differs from all DCMs, however, in two fundamental ways.

First, whereas DCMs fit all data simultaneously, we parameterize our dynamic (MINDy) models solely from resting-state data. As a result, our model parameters are not impacted by any preconceived models of task effects (i.e., that they follow a certain temporal pattern). Secondly, we do not explicitly model task effects. Whereas DCMs directly fit parameters to task conditions, MINDy-based Filtering produces a full timeseries of estimated effects based solely upon fMRI scans (i.e. no task information is used). Thus, our approach estimates the evolution of latent variables (task-related "input" to each region) rather than estimating coefficients for a pre-specified temporal model of task effects. As such the proposed technique is much more flexible than DCM as it functions as a processing step rather than a full analysis pipeline in and of itself. In the current work, we used statistical GLMs to analyze the MINDy-filtered data with Finite Impulse Response models fit for each trial type and additional components to model task blocks (mixed block-event design). However, the end-product of our technique (a timeseries) could, in principle, be analyzed with a wide variety of methods including parcel-level multivariate techniques (e.g., multivariate pattern analysis; MVPA). Thus, although DMCC and the current technique both use empirical dynamical systems models with similar assumptions, the approaches differ radically in how these models are leveraged.

4.6.3 Limitations

The proposed work rests upon two related claims: 1) that by subtracting intrinsic dynamics we identify changes in "input" to each brain area and 2) that the signal generated by this calculation is a better marker of task effects (ostensibly task-related cognition).

Methodological Considerations

The bulk of our results concern the latter claim (improved detection power) and demonstrating that these statistical improvements are related to task-specific neural processes. We performed these tests using several controlled comparisons and lines of inquiry. However, our efforts in this domain are limited by using a specific subset of cognitive tasks: those used to index cognitive control. As the set of potential cognitive constructs remains vast, further testing in other cognitive domains may be useful.

Another limitation concerns how MINDy models are parameterized. Since we parameterize models based upon resting-state data, we require the collection of both resting-state and task data for each subject which increases data requirements. Moreover, this dependency could prove problematic for low-quality resting-state data as mis-specified resting-state models could corrupt task estimates. We found that individual differences in goodness-of-fit were consistent across tasks (Fig. 4.1B) so this possibility cannot be ruled out. However, analyses with MINDy modeling indicated that the goodness-of-fit is not related to individual differences in motion ([8]). Results also do not support model overfitting as goodness-of-fit did not uniformly decrease when applied to novel (task) data relative to training (rest) data (Fig. 4.1A). Further study may therefore be beneficial in determining which factors (neural or nuisance) influence individual differences in goodness of fit as these factors could influence estimated individual differences in task variables.

Mechanistic Considerations

Future study is necessary is necessary to disambiguate which biological mechanisms contribute to the calculated "input" signal. For decades, computational neuroscience models have largely formalized task context as an exogeneous forcing ("input" or "bias") term in neural networks and connectionist models (e.g. [150], [151], [152], [153], [154]). This formulation is appealing for its simplicity, however external contexts are known to only act as "inputs" during sensory transduction since brain activity is known to modulate sensory neurons (e.g. [155], [156]). Even when these effects are neglected, many modeling studies consider task "inputs" to regions that are not directly enervated by sensory nerves (e.g. [152]). As a result, these "inputs" should not be interpreted as literal inputs to the brain (i.e. signals from sensory nerves). Rather, these "inputs" include the initial propagation of such signals over the fMRI sampling rate (1 TR), so our approach is limited by the temporal resolution of fMRI BOLD.

The nature of these "inputs" is also somewhat underspecified. In the current approach, we use MINDy to model the propagation of brain signals during resting-state. The model predicts task-fMRI activation based upon the effective connectivity parameters estimated from resting-state. However, these parameters are limited to describing the relationship of bulk activity between brain regions. Many brain regions contain diffuse sets of neurons with heterogeneous axonal connectivity profiles. Several lines of evidence suggest that task-contexts can modulate the effective connectivity between brain regions via selective recruitment of neurons in synchronous ensembles ([157], [158], [159]). Our approach is therefore limited in that it does not explicate how changes in "input" relate to changes in the effective coupling between brain regions. Future studies may improve upon the current approach by further modeling how task events modulate effective connectivity between brain regions. Such studies could either directly parameterize connectivity × task interactions (as in DCM) or extend the filtering approach to estimate time-varying connectivity.

4.6.4 Task Dynamics Could Potentially Influence Statistical Improvements

The current approach serves to estimate latent changes in input to each brain area. In the present study we found that MINDy-based Filtering consistently improved statistical detection power across tasks. However, there may be contexts in which brain activity ($x(t)$) is a more consistent marker of task context than input ($I(t)$). Such cases occur when different input patterns (i.e. inter-trial variability in input) lead to the similar outcomes in terms of activity. In these cases, MINDy-based Filtering might actually decrease detection power, since the “input” on each trial is less consistent than its long-term consequences. Future studies might identify such cases using a wider variety of tasks.

4.6.5 Conclusion

In the current work, we proposed a new technique to estimate the influence of external contexts (task conditions) on brain activity (in our case fMRI). This technique forms a mathematical filter and therefore functions as a preprocessing step rather than as a direct tool for hypothesis testing. This property is advantageous as it allows this approach to be used in conjunction with a variety of existing methods. We have demonstrated that using this technique improves statistical power (Fig. 4.3,4.4), increases sensitivity to task-implicated regions (Sec. 4.5.6), and better identifies the neural signatures of a latent cognitive construct (cognitive conflict) in both individuals and group-level (Fig. 4.5A,B). These improvements are not sensitive to motion within a reasonable range (Fig. 4.5C). Our technique can be easily inserted into most fMRI processing pipelines and we have made code available via the primary author’s GitHub to facilitate this process.

Chapter 5

Conclusion

5.1 Relationship to Existing Literature

Brain activity varies remarkably across contexts and individuals. This variability is reflected across time scales that range from trial-to-trial variability present within-individuals all the way to life-outcomes varying between-individuals. For this reason, significant literature concerns the neural substrates of individual differences in behavior and, to a lesser extent, brain activity. Recent studies in this domain have found that the intrinsic correlation of activity between brain regions at rest (e.g. resting-state Functional Connectivity; rsFC) is a powerful way to characterize individuals' brain activity across contexts. For instance, previous studies ([43], [47]) have found that individual differences in patterns of task-free brain activity (rsFC networks) map onto anatomically-isomorphic patterns of task-evoked activity.

This literature has been constrained, however, by a reliance on static statistical descriptions. Although previous approaches can predict which sets of regions (i.e. a network) coactivate across contexts, they do not predict the temporal dynamics of how networks will be activated or under which contexts this activation will occur. These properties are significant for forming mechanistically explanatory rather than descriptive theories of neural processes. For these reasons, dynamical systems frameworks have been sought to model the generative mechanisms that describe the spatiotemporal evolution of brain activity ([57]). However, these approaches have been limited by a different set of constraints. In contrast, to the previously mentioned resting-state approaches, task dynamic models ("Dynamic Causal Modeling"; DCM [52]), have had limited scope as they do not scale well to whole-brain analyses or to a large number of task conditions due to a reliance on Bayesian estimation/model comparison. Moreover, it remains unknown whether these dynamic models generalize across contexts (i.e. whether a resting-state brain model can predict task-related activity).

5.2 Resting-State MINDy Modeling

In the current work, we aimed to bridge the methodological gaps present in previous modeling approaches by directly testing whether dynamic brain models generated from resting-state brain data predict individualized evoked activity. To do so, we developed a new technique for generating individualized brain and neurovascular models using fMRI data. These Mesoscale Individualized NeuroDynamic (MINDy) models take the form of nonlinear dynamical systems in which activity is nonlinearly propagated through a network of interconnected brain regions, where each region integrates this activity at different rates (time-constants). We hypothesized that dynamic brain models would be more tightly linked to underlying brain processes and (effective) connectivity than statistical methods. We approached this question through a combination of empirical and simulation tests.

5.2.1 Validation of the MINDy Framework

Empirical tests established that the “connectivity” parameter of MINDy models is a more stable measure of individualized connectivity as it is less sensitive to non-neural variables such as sampling variability (indexed via test-retest reliability). Simulation tests also indicated that our modeling approach provides a more veridical estimate of effective brain connectivity and connectivity-related individual differences than either statistical methods or other dynamical modeling methods (Sec. 2.3.2, 2.3.8). Of course, simulation analyses are only valid to the extent that the underlying models reflect neurophysiology, so we repeated analyses over a variety of ground-truth modeling scenarios. Results held over all ground-truth models of hemodynamics (e.g. nonlinear, spatially heterogeneous, etc.), neural dynamics (different functional forms and timescales), and means of generating simulation parameters (random or empirical).

We also demonstrated that our dynamic models form a more fundamental description of neural processes. We first demonstrated that MINDy models strongly predict rsFC at both the group and individual (Sec. 2.3.4) levels. Thus, MINDy contains sufficient information to reconstruct statistical properties, but the converse does not hold; it is impossible to reconstruct the directed connectivity of MINDy from the undirected rsFC matrix. For this reason, we say that MINDy is a more fundamental descriptor. Moreover, we demonstrated that MINDy models predict nonstationarity in statistical descriptors (rsFC) at both the individual and group level. This finding suggests that MINDy models may also be able to explain patterns of brain activity that vary in time.

5.2.2 Disentangling Neural Dynamics and Hemodynamics

One of the primary limitations to neuroimaging-based modeling of brain network dynamics, however, has been the indirectness of fMRI measurements. Current human fMRI techniques largely utilize the BOLD contrast which presents several ambiguities. One of these is the influence of non-neural variables, such as breathing rate (which can covary with task conditions), along with vascular tone and cerebral blood flow. Latent-variable effects such as these, present a general limitation of fMRI data for neural modeling, that also impacts the current work. However, methodologies introduced in the current work could potentially reduce ambiguities in the temporal linkage between neural events and BOLD. In particular, we demonstrated that the novel Surrogate Deconvolution approach enables accurate estimates of Hemodynamic Response Functions (HRF) from resting-state data (Sec. 3.4.4). In simulations, we demonstrated that the same technique improved estimation of latent neural state-variables (Sec. 3.4.2). Since our HRF estimates are independent of task effects (due to the use of resting-state data), these approaches may eventually be used to estimate the time-courses of (low frequency) neural events from fMRI

5.3 Implications for Individual Differences in Spontaneous Brain Activity

During the validation of MINDy, we examined the sensitivity of MINDy and rsFC to individual differences by comparing the similarity of repeated measures within-individuals to those computed between individualis (Sec. 2.3.4). Interestingly, we found that while MINDy parameters reliably separated indiivduals, they were actually more similar to each other in terms of the MINDy connectivity parameter, relative to rsFC parameters. Likewise, although MINDy models predicted left-out (cross-validated) data from the same individual better than for any other individual, the absolute difference in accuracy was relatively small for short-time predictions (predicting the change in activity over two time-steps [1.44s]) However, when MINDy models were simulated over large periods of time, we found that the rsFC statistics of these simulations matched those of the data including strong, reliable differences between individuals. This pattern of results suggests that individual differences in the effective connectivity between regions may be significantly smaller than estimates based upon rsFC. Rather, results suggest that the iterated accumulation of these small differences leads to markedly different patterns of rsFC.

We also demonstrated that rsFC was very sensitive to the time-constant (decay parameter) of brain regions. The decay parameter varied systematically between brain regions (correlated with the "global connectivity" statistic) and individual differences (identified via PCA) were organized along the so-called cortical hierarchy. When we permuted the decay parameters of a model between subjects, the model's simulated rsFC changed significantly. This finding suggests that the timescales of local brain integration are a significant factor mediating individual differences in rsFC. Future studies should test whether individual differences in the estimated time-constants reflect corresponding variation in other variables organized along the cortical hierarchy, such as white matter density.

5.4 Implications for The Origin and Analysis of Task-Evoked Activity

This work also suggests several implications for the study of task-evoked brain activity. As referenced earlier, previous literature (e.g. [30], [66]) indicates that Functional Connectivity is roughly preserved between resting-state fMRI and task fMRI (after removing the task main effects). Our work suggests that a similar continuity holds over small timescales since we found that MINDy models fit to spontaneous brain activity were similarly accurate in predicting brain activity during task contexts. This result suggests that the combined variance associated with task effects and 'noise' during task have similar variance to the 'noise' component during resting-state. This implication agrees with other contemporary studies, which have suggested a decrease in 'noise' variability during task-state (e,g, [45], [16])

We also found that removing resting-state model predictions (MINDy-based filtering) decreased inter-individual variability of estimated task-effects. However, individual differences in these estimates generalized better between cognitive-control tasks after MINDy-based filtering. This finding suggests that the observed post-MINDy homogenization does not reflect insensitivity to individual factors. Instead, these results suggest that individual differences are small for direct task effects ("input") but are magnified as "inputs" propagate along individualized networks and local dynamics.

5.5 Limitations and Directions for Future Work

There are two primary limitations to the current work: low temporal resolution and limited mechanistic inferences. Both constraints are due, in part, to the use of fMRI data. We chose fMRI for our analyses due to its high spatial coverage and the spatially local nature of measurements (relative the scale of cortical gradients and areal differentiation). Since the current approach assumes that all brain areas in a model are measured simultaneously, high coverage measurements are necessary. Other neuroimaging methods such as EEG and MEG also have relatively high coverage over superficial cortex, but are limited in their ability to spatially localize neural events due to the spatial mixing and orientation-dependence of electromagnetic fields (mostly MEG) and the volume conduction of signals (mostly EEG). Much of the current work emphasized validating our new technique (MINDy), so the ability to compare estimated model with known neuroanatomy proved critical. For this reason, we chose fMRI for the current work.

5.5.1 Developing Multimodal MINDy

However, the low temporal resolution of fMRI is one limitation of this work. Current knowledge concerning neural dynamics emphasizes much faster timescales than can be observed with fMRI. As a result, we expect that the current models, while accurate for fMRI data, will fail to capture many neural processes which cannot be measured in fMRI data. Future implementations of MINDy may benefit from developing multimodal modeling approaches which leverage the high spatial resolution of fMRI with the temporal resolution of MEG/EEG. Such advances will provide further insight by modeling the propagation of neural activity over time-scales commensurate with neuronal computation. This, in turn, may provide insight into the sequential aspects of neural computations (e.g., perceptual) which are thought to occur faster than the fMRI sampling rate which defines the model time scale in our current work.

5.5.2 Explicating Task Mechanisms

The current work is also limited, in part, by our model formulation for task-evoked brain activity. Rather than estimating changes (relative to the resting-state) in the effective-coupling between brain-regions during task, we estimate changes in the net input to each brain region, without specifying its source. Such descriptions may be limited in their ability to identify the neural circuitry underlying task computations. By determining the circuitry underlying task “inputs”, future work will further explicate the neural mechanisms underlying cognitive operations as manifest in task-evoked brain activity. One way to model these mechanisms is via an interaction between effective connectivity and task conditions ([52]). Current approaches for modeling such effects in DCM require first specifying the time-course of task events and then incorporating these assumptions to estimate effective connectivity. Such approaches run counter to the “Filtering” approached that we present in the current work. Instead, an alternative path may be treating effective connectivity as a latent, time-varying state variable rather than a fixed parameter using the joint Kalman filters (Sec. 3.1.2, e.g. [121]). This approach would maintain all the advantages of MINDy-based Filtering while enabling parallel analyses for effective connectivity (i.e. the timeseries of effective connectivity estimates would be analyzed using conventional GLM techniques). However, current iterations of joint Kalman filtering do not generally scale-well (e.g. 3.4.1) so progress may depend upon future development in joint estimation.

5.5.3 Linking MINDy and Behavior

Lastly, the current study has been limited in scope by only interrogating our models in their prediction of neural data. Using MINDy, we predicted individual differences in spontaneous neural activity (Dynamic Functional Connectivity; Sec. 2.3.6) and identified individual differences in activity related to cognitive control (Sec. 4.5.5). Ostensibly, individual differences in behavior possess some neural substrate. However, just because the current work identified robust individual differences in neural activity does not guarantee that such effects are behaviorally-relevant. For instance, such effects can result anatomical variability in the location (coordinates) of functional brain areas. Future studies which attempt to demonstrate that model predictions can be directly linked to individuals’ behavior will provide further validation of the MINDy approach and may generate new insight into the neural mechanisms of individualized behavior.

5.6 Conclusion

In conclusion, the present work provides a framework for individualized brain modeling and state-estimation across contexts. These models take the form of large nonlinear dynamical systems (hundreds of brain regions) which we fit to individualized fMRI data. The present work serves to validate this approach within the resting-state domain and demonstrates its utility in identifying exogenous (task) influences upon neural activity. These advances will support new insight into the relationship between intrinsic brain relationships (at rest) and goal-oriented neural computations (during task) at both the individual and group level.

Chapter 6

References

6.1 References

[1] J J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, 1984.

[2] Joshua S. Siegel, Anish Mitra, Timothy O. Laumann, Benjamin A. Seitzman, Marcus Raichle, Maurizio Corbetta, and Abraham Z. Snyder. Data quality influences observed links between functional connectivity and behavior. *Cerebral Cortex*, 27(9):4492–4502, 2017.

[3] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT Press, New York City, 1949.

[4] K.J. Friston, P. Fletcher, O. Josephs, A. Holmes, M.D. Rugg, and R. Turner. Event-related fmri: Characterizing differential responses. *NeuroImage*, 7(1):30 – 40, 1998.

[5] Murat Demirtaş, Joshua B. Burt, Markus Helmer, Jie Lisa Ji, Brendan D. Adkinson, Matthew F. Glasser, David C. Van Essen, Stamatios N. Sotiropoulos, Alan Anticevic, and John D. Murray. Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron*, 101(6):1181–1194, 2019.

[6] Alexander Schaefer, Ru Kong, Evan M. Gordon, Timothy O. Laumann, Xi-Nian Zuo, Avram J. Holmes, Simon B. Eickhoff, and B.T. Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral Cortex*, pages 1–20, 2017.

[7] Angela D Friederici and Sarah ME Gierhan. The language network. *Current Opinion in Neurobiology*, 23(2):250 – 254, 2013.

[8] Matthew F. Singh, Todd S. Braver, Michael W. Cole, and ShiNung Ching. Estimation and validation of individualized dynamic brain models with resting state fmri. *NeuroImage*, 221:117046, 2020.

[9] Timothy Dozat. Incorporating nesterov momentum into adam. *Proceedings of 4th International Conference on Learning Representations, Workshop Track, 2016*, 2016.

[10] Matthew F. Glasser, Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, Ess Harwell, John andYacoub, Kamil Ugurbil, Jesper Andersson, Christian F. Beckmann, Mark Jenkinson, Stephen M. Smith, and David C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536:171–178, 2016.

[11] Ann S. Choe, Mary Beth Nebel, Anita D. Barber, Jessica R. Cohen, Yuting Xu, James J. Pekar, Brian Caffo, and Martin A. Lindquist. Comparing test-retest reliability of dynamic functional connectivity methods. *NeuroImage*, 158:155 – 175, 2017.

[12] Chao Zhang, Stefi A. Baum, Viraj R. Adduru, Bharat B. Biswal, and Andrew M. Michael. Test-retest reliability of dynamic functional connectivity in resting state fmri. *NeuroImage*, 183:907 – 918, 2018.

[13] R. P. Tett and H. A. Guterman. Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, 34(4):397–423, 2000.

[14] Todd S Braver, Michael W Cole, and Tal Yarkoni. Vive les differences! individual variation in neural mechanisms of executive control. *Current Opinion in Neurobiology*, 20(2):242 – 250, 2010.

[15] Katharina Geukes, Steffen Nestler, Roos Hutteman, Albrecht C.P. Küfner, and Mitja D. Back. Trait personality and state variability: Predicting individual differences in within- and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, 69:124 – 138, 2017.

[16] Adrián Ponce-Alvarez, Biyu J He, Patric Hagmann, and Gustavo Deco. Task-driven activity reduces the cortical activity space of the brain: experiment and whole-brain modeling. *PLoS computational biology*, 11(8):e1004445, 2015.

[17] S Kühn, F Schmiedek, B Schott, R Ratcliff, H.J. Heinze, E Düzel, U Lindenberger, and M Lövden. Brain areas consistently linked to individual differences in perceptual decision-making in younger as well as older adults before and after training. *Journal of Cognitive Neuroscience*, 23(9):2147–58, 2011.

[18] Euan A. Ashley. The Precision Medicine Initiative: A New National Effort. *JAMA*, 313(21):2119–2120, 06 2015.

[19] Bruce M. Psaty, Olaf M. Dekkers, and Richard S. Cooper. Comparison of 2 Treatment Models: Precision Medicine and Preventive Medicine. *JAMA*, 320(8):751–752, 2018.

[20] Theodore D. Satterthwaite, Cedric H. Xia, and Danielle S. Bassett. Personalized neuroscience: Common and individual-specific features in functional brain networks. *Neuron*, 98(2):243 – 245, 2018.

[21] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117:500–544, 1952.

[22] Randall D. Beer. Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3):91 – 99, 2000.

[23] Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11:127–138, 2010.

[24] Gustavo Deco, Viktor K. Jirsa, and Anthony R. McIntosh. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12:43–56, 2011.

[25] Richard F. Betzel, Makoto Fukushima, Ye He, Xi-Nian Zuo, and Olaf Sporns. Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *NeuroImage*, 127:287–297, 2016.

[26] Christopher J. Honey, Rolf Kötter, Michael Breakspear, and Olaf Sporns. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*, 104(24):10240–10245, 2007.

[27] Bharat Biswal, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. Functional connectivity in the motor cortex of resting human brain using echoplanar mri. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995.

[28] 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*, 106(6):2035–2040, 2009.

[29] Jonathan D. Power, Alexander L. Cohen, Steven M. Nelson, Gagan S. Wig, Kelly Anne Barnes, Jessica A. Church, Alecia C. Vogel, Timothy O. Laumann, Fran M. Miezin, Bradley L. Schlaggar, and Steven E. Petersen. Functional network organization of the human brain. *Neuron*, 72(4):665 – 678, 2011.

[30] Michael W. Cole, Danielle S. Bassett, Jonathan D. Power, Todd S. Braver, and Steven E. Petersen. Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83(1):238 – 251, 2014.

[31] Urs Braun, Michael M. Plichta, Christine Esslinger, Carina Sauer, Leila Haddad, Oliver Grimm, Daniela Mier, Sebastian Mohnke, Andreas Heinz, Susanne Erk, Henrik Walter, Nina Seiferth, Peter Kirsch, and Andreas Meyer-Lindenberg. Test–retest reliability of resting-state connectivity network characteristics using fmri and graph theoretical measures. *NeuroImage*, 59(2):1404 – 1412, 2012.

[32] Xi-Nian Zuo, Clare Kelly, Jonathan S. Adelstein, Donald F. Klein, F. Xavier Castellanos, and Michael P. Milham. Reliable intrinsic connectivity networks: Test–retest evaluation using ica and dual regression approach. *NeuroImage*, 49(3):2163 – 2177, 2010.

[33] R. Matthew Hutchison, Thilo Womelsdorf, Elena A. Allen, Peter A. Bandettini, Vince D. Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H. Duyn, Gary H. Glover, Javier Gonzalez-Castillo, Daniel A. Handwerker, Shella Keilholz, Vesa Kiviniemi, David A. Leopold, Francesco de Pasquale, Olaf Sporns, Martin Walter, and Catie Chang. Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, 80:360 – 378, 2013.

[34] R. Hindriks, M.H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N.K. Logothetis, and G. Deco. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fmri? *NeuroImage*, 127:242 – 256, 2016.

[35] Maria Giulia Preti, Thomas AW Bolton, and Dimitri Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, 160:41 – 54, 2017.

[36] Timothy O. Laumann, Abraham Z. Snyder, Anish Mitra, Evan M. Gordon, Caterina Gratton, Babatunde Adeyemo, Adrian W. Gilmore, Steven M. Nelson, Jeff J. Berg, Deanna J. Greene, John E. McCarthy, Enzo Tagliazucchi, Helmut Laufs, Bradley L. Schlaggar, Nico U. F. Dosenbach, and Steven E. Petersen. On the stability of bold fmri correlations. *Cerebral Cortex*, 27(10):4719–4732, 2017.

[37] Raphaël Liégeois, Timothy O. Laumann, Abraham Z. Snyder, Juan Zhou, and B.T. Thomas Yeo. Interpreting temporal fluctuations in resting-state functional connectivity mri. *NeuroImage*, 163:437 – 455, 2017.

[38] Nora Leonardi and Dimitri Van De Ville. On spurious and real fluctuations of dynamic functional connectivity during rest. *NeuroImage*, 104:430 – 436, 2015.

[39] Elena A. Allen, Eswar Damaraju, Sergey M. Plis, Erik B. Erhardt, Tom Eichele, and Vince D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 24(3):663–676, 2014.

[40] Nancy J. Kopell, Howard J. Gritton, Miles A. Whittington, and Mark A. Kramer. Beyond the connectome: The dynome. *Neuron*, 83(6):1319 – 1328, 2014.

[41] David V. Smith, Amanda V. Utevsky, Amy R. Bland, Nathan Clement, John A. Clithero, Anne E.W. Harsch, R. McKell Carter, and Scott A. Huettel. Characterizing individual differences in functional connectivity using dual-regression and seed-based approaches. *NeuroImage*, 95:1 – 12, 2014.

[42] Emily S. Finn, Xilin Shen, Dustin Scheinost, Monica D. Rosenberg, Jessica Huang, Marvin M. Chun, Xenophon Papademetris, and R. Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neureoscience*, 18:1664–1671, 2015.

[43] Michael W Cole, Takuya Ito, Danielle S. Bassett, and Douglas H. Schultz. Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, 19(12):1718–1726, 2016.

[44] I. Tavor, O. Parker Jones, R. B. Mars, S. M. Smith, T. E. Behrens, and S. Jbabdi. Task-free mri predicts individual differences in brain activity during task performance. *Science*, 352(6282):216–220, 2016.

[45] M. Churchland, B. Yu, and J. et al. Cunningham. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13:369–378, 2010.

[46] Biyu J. He. Spontaneous and task-evoked brain activity negatively interact. *Journal of Neuroscience*, 33(11):4672–4682, 2013.

[47] Evan M. Gordon, Timothy O. Laumann, Adrian W. Gilmore, Dillan J. Newbold, Deanna J. Greene, Jeffrey J. Berg, Mario Ortega, Catherine Hoyt-Drazen, Caterina Gratton, Haoxin Sun, Jacqueline M. Hampton, Rebecca S. Coalson, Annie L. Nguyen, Kathleen B. McDermott, Joshua S. Shimony, Abraham Z. Snyder, Bradley L. Schlaggar, Steven E. Petersen, Steven M. Nelson, and Nico U.F. Dosenbach. Precision functional mapping of individual human brains. *Neuron*, 95(4):791 – 807.e7, 2017.

[48] Michael D Fox, Abraham Z Snyder, Jeffrey M Zacks, and Marcus E Raichle. Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nature Neuroscience*, 9:23–25, 2006.

[49] Michael D. Fox, Abraham Z. Snyder, Justin L. Vincent, and Marcus E. Raichle. Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, 56(1):171 – 184, 2007.

[50] M.L. Schölvinck, K.J. Friston, and G. Rees. The influence of spontaneous activity on stimulus processing in primary visual cortex. *NeuroImage*, 59(3):2700 – 2708, 2012.

[51] Biyu J. He and John M. Zempel. Average is optimal: An inverted-u relationship between trial-to-trial brain activity and behavioral performance. *PLOS Computational Biology*, 9(11):1–14, 2013.

[52] K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273 – 1302, 2003.

[53] Henry Markram. The blue brain project. *Nature Reviews Neuroscience*, 7:153–160, 2006.

[54] Karl J. Friston, Joshua Kahan, Bharat Biswal, and Adeel Razi. A dcm for resting state fmri. *NeuroImage*, 94:396–407, 2014.

[55] Karl J. Friston, Baojuan Li, Jean Daunizeau, and Klaas E. Stephan. Network discovery with dcm. *NeuroImage*, 56(3):1202 – 1221, 2011.

[56] Mohamed L. Seghier and Karl J. Friston. Network discovery with large dcms. *NeuroImage*, 68:181 – 191, 2013.

[57] Adeel Razi, Mohamed L. Seghier, Yuan Zhou, and Peter McColgan. Large-scale dcms for resting-state fmri. *Network Neuroscience*, 1(3), 2017.

[58] EM Gordon, TO Laumann, B Adeyemo, JF Huckins, WM Kelley, and SE Petersen. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex*, 26(1):288–303, 2016.

[59] Paula Sanz Leon, Stuart Knock, M. Woodman, Lia Domide, Jochen Mersmann, Anthony McIntosh, and Viktor Jirsa. The virtual brain: a simulator of primate brain network dynamics. *Frontiers in Neuroinformatics*, 7:10, 2013.

[60] Gustavo Deco, Giulio Tononi, Melanie Boly, and Morten L. Kringelbach. Rethinking segregation and integration: contributions of whole-brain modelling. *Nature Reviews: Neuroscience*, 16:430–439, 2015.

[61] Michael Breakspear. Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20:340–352, 2017.

[62] Murat Demirtaş, Carles Falcon, Alan Tucholka, Juan Domingo Gispert, José Luis Molinuevo, and Gustavo Deco. A whole-brain computational modeling approach to explain the alterations in resting-state functional connectivity during progression of alzheimer's disease. *NeuroImage: Clinical*, 16:343 – 354, 2017.

[63] Michael Schirner, Simon Rothmeier, Viktor K. Jirsa, Anthony Randal McIntosh, and Petra Ritter. An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data. *NeuroImage*, 117:343 – 357, 2015.

[64] 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*, 145:377 – 388, 2017.

[65] David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The wu-minn human connectome project: An overview. *NeuroImage*, 80:62 – 79, 2013.

[66] Caterina Gratton, Timothy O. Laumann, Ashley N. Nielsen, Deanna J. Greene, Evan M. Gordon, Adrian W. Gilmore, Steven M. Nelson, Rebecca S. Coalson, Abraham Z.

Snyder, Bradley L. Schlaggar, Nico U.F. Dosenbach, and Steven E. Petersen. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, 98(2):439 – 452, 2018.

[67] Sepideh Sadaghiani, Jean-Baptiste Poline, Andreas Kleinschmidt, and Mark D'Esposito. Ongoing dynamics in large-scale functional connectivity predict perception. *Proceedings of the National Academy of Sciences*, 112(27):8463–8468, 2015.

[68] Henry Markram, Karlheinz Meier, Thomas Lippert, Sten Grillner, Richard Frackowiak, Stanislas Dehaene, Alois Knoll, Haim Sompolinsky, Kris Verstreken, Javier DeFelipe, Seth Grant, Jean-Pierre Changeux, and Alois Saria. Introducing the human brain project. *Procedia Computer Science*, 7:39 – 42, 2011.

[69] Hideyuki Okano, Erika Sasaki, Tetsuo Yamamori, Atsushi Iriki, Tomomi Shimogori, Yoko Yamaguchi, Kiyoto Kasai, and Atsushi Miyawaki. Brain/minds: A japanese national brain project for marmoset neuroscience. *Neuron*, 92(3):582 – 590, 2016.

[70] Randy L Buckner, Fenna M Krienen, and B T Thomas Yeo. Opportunities and limitations of intrinsic functional connectivity mri. *Nature Neuroscience*, 16:832–837, 2013.

[71] Peng Wang, Ru Kong, Xiaolu Kong, Raphaël Liégeois, Csaba Orban, Gustavo Deco, Martijn P. van den 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*, 5(1), 2019.

[72] Chengyi Tu, Rodrigo P. Rocha, Maurizio Corbetta, Sandro Zampieri, Marco Zorzi, and S. Suweis. Warnings and caveats in brain controllability. *NeuroImage*, 176:83 – 91, 2018.

[73] Srikanth Ryali, Kaustubh Supekar, Tianwen Chen, and Vinod Menon. Multivariate dynamical systems models for estimating causal interactions in fmri. *NeuroImage*, 54(2):807 – 823, 2011.

[74] Alard Roebroeck, Elia Formisano, and Rainer Goebel. The identification of interacting networks in the brain using fmri: Model selection, causality and deconvolution. *NeuroImage*, 58(2):296 – 302, 2011.

[75] Gabriele Lohmann, Kerstin Erfurth, Karsten Müller, and Robert Turner. Critical comments on dynamic causal modelling. *NeuroImage*, 59(3):2322 – 2329, 2012.

[76] Stefan Frässle, Ekaterina I Lomakina, Adeel Razi, Karl J Friston, Joachim M Buhmann, and Klaas E Stephan. Regression dcm for fmri. *Neuroimage*, 155:406–421, 2017.

[77] S. Knock, A. McIntosh, O. Sporns, R. K'otter, 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*, 3(1):86–94, 2009.

[78] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1–24, 1972.

[79] André C. Marreiros, Jean Daunizeau, Stefan J. Kiebel, and Karl J. Friston. Population dynamics: Variance and the sigmoid activation function. *NeuroImage*, 42(1):147 – 157, 2008.

[80] David L. Donoho. For most large underdetermined systems of linear equations the minimal l1 norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.

[81] Francesca Mastrogiuseppe and Srdjan Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99:609–623, 2018.

[82] Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124, 2013.

[83] Ludovica Griffanti, Gholamreza Salimi-Khorshidi, Christian F. Beckmann, Edward J. Auerbach, Gwenaëlle Douaud, Claire E. Sexton, Enikő Zsoldos, Klaus P. Ebmeier, Nicola Filippini, Clare E. Mackay, Steen Moeller, Junqian Xu, Essa Yacoub, Giuseppe Baselli, Kamil Ugurbil, Karla L. Miller, and Stephen M. Smith. Ica-based artefact removal and accelerated fmri acquisition for improved resting state network imaging. *NeuroImage*, 95:232 – 247, 2014.

[84] Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F. Beckmann, Matthew F. Glasser, Ludovica Griffanti, and Stephen M. Smith. Automatic denoising of functional mri data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449 – 468, 2014.

[85] Jonathan D. Power, Bradley L. Schlaggar, and Steven E. Petersen. Recent progress and outstanding issues in motion correction in resting state fmri. *NeuroImage*, 105:536 – 551, 2015.

[86] Jonathan D. Power, Kelly A. Barnes, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *NeuroImage*, 59(3):2142 – 2154, 2012.

[87] Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T. Liu. A component based noise correction method (compcor) for bold and perfusion based fmri. *NeuroImage*, 37(1):90 – 101, 2007.

[88] G.K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human, bold hemodynamic responses. *NeuroImage*, 8(4):360 – 369, 1998.

[89] Bruce Fischl. Freesurfer. *NeuroImage*, 62(2):774 – 781, 2012.

[90] B. T. Thomas Yeo, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, Jordan W. Smoller, Lilla Zöllei, Jonathan R. Polimeni, Bruce Fischl, Hesheng Liu, and Randy L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165, 2011.

[91] Paul W. Holland and Roy E. Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6(9):813–827, 1977.

[92] K.J. Friston, A. Mechelli, R. Turner, and C.J. Price. Nonlinear responses in fmri: The balloon model, volterra kernels, and other hemodynamics. *NeuroImage*, 12(4):466 – 477, 2000.

[93] Enrique C.A. Hansen, Demian Battaglia, Andreas Spiegler, Gustavo Deco, and Viktor K. Jirsa. Functional connectivity dynamics: Modeling the switching behavior of the resting state. *NeuroImage*, 105:525 – 535, 2015.

[94] Robert Engle. Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

[95] PE Shrout and JL Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2):420–428, 1979.

[96] 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*, 13(4):714–724, 2013.

[97] Timothy O. Laumann, Evan M. Gordon, Babatunde Adeyemo, Abraham Z. Snyder, Sung Jun Joo, Mei-Yen Chen, Adrian W. Gilmore, Kathleen B. McDermott, Steven M. Nelson, Nico U.F. Dosenbach, Bradley L. Schlaggar, Jeanette A. Mumford, Russell A. Poldrack, and Steven E. Petersen. Functional system and areal organization of a highly sampled individual human brain. *Neuron*, 87(3):657 – 670, 2015.

[98] G.K. Aguirre, E. Zarahn, and M. D'Esposito. The inferential impact of global signal covariates in functional neuroimaging analyses. *NeuroImage*, 8(3):302 – 306, 1998.

[99] Matthew Singh, Anxu Wang, Todd Braver, and ShiNung Ching. Scalable surrogate deconvolution for identification of partially-observable systems and brain modeling. *Journal of Neural Engineering*, 2020.

[100] Fa-Hsuan Lin, Jonathan R. Polimeni, Jo-Fu Lotus Lin, Kevin W.-K. Tsai, Ying-Hua Chu, Pu-Yeh Wu, Yi-Tien Li, Yi-Cheng Hsu, Shang-Yueh Tsai, and Wen-Jui Kuo. Relative latency and temporal variability of hemodynamic responses at the human primary visual cortex. *NeuroImage*, 164:194 – 201, 2018.

[101] Klaas Enno Stephan, Lars Kasper, Lee M. Harrison, Jean Daunizeau, Hanneke E.M. [den Ouden], Michael Breakspear, and Karl J. Friston. Nonlinear dynamic causal models for fmri. *NeuroImage*, 42(2):649 – 662, 2008.

[102] K.J. Friston, Katrin H. Preller, Chris Mathys, Hayriye Cagnan, Jakob Heinzle, Adeel Razi, and Peter Zeidman. Dynamic causal modelling revisited. *NeuroImage*, 199:730 – 744, 2019.

[103] MohammadMehdi Kafashan, Ben Julian A. Palanca, and ShiNung Ching. Dimensionality reduction impedes the extraction of dynamic functional connectivity states from fmri recordings of resting wakefulness. *Journal of Neuroscience Methods*, 293:151 – 161, 2018.

[104] Anxu Wang, Matthew F. Singh, Joset Etzel, and Todd Braver. Enhancing task fmri preprocessing via whole-brain neural modeling of intrinsic activity dynamics. In *Organization for Human Brain Mapping (Virtual), 2020*, 2020.

[105] Anish Mitra, Abraham Z Snyder, Enzo Tagliazucchi, Helmut Laufs, and Marcus E Raichle. Propagated infra-slow intrinsic brain activity reorganizes across wake and slow wave sleep. *eLife*, 4:e10781, 2015.

[106] Anish Mitra, Andrew Kraft, Patrick Wright, Benjamin Acland, Abraham Z. Snyder, Zachary Rosenthal, Leah Czerniewski, Adam Bauer, Lawrence Snyder, Joseph Culver, Jin-Moo Lee, and Marcus E. Raichle. Spontaneous infra-slow brain activity has unique spatiotemporal dynamics and laminar structure. *Neuron*, 98(2):297 – 305.e6, 2018.

[107] Aaron Kucyi and Karen D. Davis. Dynamic functional connectivity of the default mode network tracks daydreaming. *NeuroImage*, 100:471 – 480, 2014.

[108] W. Rall. Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *Journal of Neurophysiology*, 30:1138–1168, 1967.

[109] J. T. Vogelstein, A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104:3691–3704, 2010.

[110] L. Theis, P. Berens, E. Froudarakis, J. Reimer, M. Román Rosón, T. Bader, T. Euler, A. S. Tolias, and M. Bethge. Benchmarking spike rate inference in population calcium imaging. *Neuron*, 90:471–482, 2016.

[111] A.Destexhe, Z. Mainen, and T. J. Sejnowski. An efficient method for computing synaptic conductances based on a kinetic model of receptor binding. *Neural Computation*, 6:14–18, 1994.

[112] A. A. Lammertsma and S. P. Hume. Simplified reference tissue model for pet receptor studies. *Neuroimage*, 4:153–158, 1996.

[113] H. Watabe, Y. Kimura Y. Ikoma, M. Naganawa, and M. Shidahara. Pet kinetic analysis–compartmental model. *Annals of Nuclear Medicine*, 20(9):583–588, 2006.

[114] G. T. Einevoll, C. Kayser, N. K. Logothetis, and S. Panzeri. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 13:770–785, 2013.

[115] K. S. Narendra and K. Partbsarathy. Identification and control of dynamical systems using neural networks. *IEEE Trans. Neural Networks*, 1(1):4–27, 1990.

[116] Tokunbo Ogunfunmi. *Adaptive nonlinear system identification: The Volterra and Wiener model approaches*. Springer Science & Business Media, 2007.

[117] S. Chen and S. A. Billings. Representation of nonlinear systems: The narmax model. *International Journal of Control*, 49(3):1013–1032, 1989.

[118] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.*, 82:34–45, 1960.

[119] S. J. Julier and J. K. Uhlmann. New extension of the kalman filter to nonlinear systems. *Proc. SPIE 3068, Signal Processing, Sensor Fusion, and Target Recognition VI*, pages 182–193, 1997.

[120] R. Van der Merwe and E. Wan. The square-root unscented kalman filter for state and parameter-estimation. *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 6:3461–3464, 2001.

[121] M. Xiong X. Sun, L. Jin. Extended kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks. *PLoS One*, 3(11):e3758, 2008.

[122] W. H. Richardson. Bayesian-based iterative method of image restoration. *JOSA*, 62(1):55–59, 1972.

[123] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745–755, 1974.

[124] M. Afonso, J. Bioucas-Dias, and M. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process*, 20(3):681–695, 2011.

[125] U. A. Khan and J. M. F. Moura. Distributing the kalman filter for large-scale systems. *IEEE Trans. Signal Processing*, 56(10):4919–4935, 2008.

[126] S. Frässle, E. I. Lomakina, L. Kasper, Z. M. Manjaly, A. Leff, K. P. Pruessmann, J. M. Buhmann, and K. E. Stephan. A generative model of whole-brain effective connectivity. *NeuroImage*, 179:505–529, 2018.

[127] D. A. Fish, A. M. Brinicombe, E. R. Pike, and J. G. Walker. Blind deconvolution by means of the richardson–lucy algorithm. *Journal of the Optical Society of America: A*, 12(1):58–65, 1995.

[128] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 2008.

[129] C. Ekanadham, D. Tranchina, and E. P. Simoncelli. A blind sparse deconvolution method for neural spike identification. *Advances in Neural Information Processing Systems*, 24:1440–1448, 2011.

[130] I. J. Park, Y. V. Bobkov, B. W. Ache, and J. C. Principe. Quantifying bursting neuron activity from calcium signals using blind deconvolution. *Journal of Neuroscience Methods*, 218(2):196–205, 2013.

[131] Karl J. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011.

[132] J.D. Cohen, D.M. Barch, and D. Carter, C.S.and Servan-Schreiber. Schizophrenic deficits in the processing of context: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology*, 108:120–133, 1999.

[133] S. Sternberg. High-speed scanning in human memory. *Science*, 153:652–654, 1966.

[134] J.R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18:643–662, 1935.

[135] J.M. Bugg and T.S. Braver. Proactive control of irrelevant task rules during cued task switching. *Psychological Research*, 80:860–876, 2016.

[136] O. Esteban, C.J. Markiewicz, and R.W. et al. Blair. fmriprep: a robust preprocessing pipeline for functional mri. *Nature Methods*, 16:111–116, 2019.

[137] O. Esteban, R. Ciric, and K. et al. Finc. Analysis of task-based functional mri data preprocessed with fmriprep. *Nature Protocols*, 15:2186–2202, 2020.

[138] Gary H. Glover. Deconvolution of impulse response in event-related bold fmri1. *NeuroImage*, 9(4):416 – 429, 1999.

[139] C. Goutte, F. A. Nielsen, and K. H. Hansen. Modeling the hemodynamic response in fmri using smooth fir filters. *IEEE Transactions on Medical Imaging*, 19(12):1188–1201, 2000.

[140] J.M. Ollinger, G.L. Shulman, and M. Corbetta. Separating processes within a trial in event-related functional mri: I. the method. *NeuroImage*, 13(1):210 – 217, 2001.

[141] Robert W Cox. Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29:162–173, 1996.

[142] Kristina M Visscher, Francis M Miezin, James E Kelly, Randy L Buckner, David I Donaldson, Mark P McAvoy, Vidya M Bhalodia, and Steven E Petersen. Mixed blocked/event-related designs separate transient and sustained activity in fmri. *Neuroimage*, 19(4):1694–1708, 2003.

[143] Joseph W Dubis Steven E Petersen 1. The mixed block/event-related design. *Neuroimage*, 62(2):1177–1184, 2012.

[144] Todd S. Braver, Alexander Kizhner, Rongxiang Tang, Michael Freund, Anxu Wang, Matthew Singh, and Joset A. Etzel. The dmcc project: A neuroimaging study of individual variation in cognitive control function. In *Organization for Human Brain Mapping (Virtual), 2020*, 2020.

[145] Bharat Biswal, Edgar A. Deyoe, and James S. Hyde. Reduction of physiological fluctuations in fmri using digital filters. *Magnetic Resonance in Medicine*, 35(1):107–113, 1996.

[146] Ola Friman, Magnus Borga, Peter Lundberg, and Hans Knutsson. Detection and detrending in fmri data analysis. *NeuroImage*, 22(2):645 – 655, 2004.

[147] Catherine E. Davey, David B. Grayden, Gary F. Egan, and Leigh A. Johnston. Filtering induces correlation in fmri resting state data. *NeuroImage*, 64:728 – 740, 2013.

[148] Simo Särkkä, Arno Solin, Aapo Nummenmaa, Aki Vehtari, Toni Auranen, Simo Vanni, and Fa-Hsuan Lin. Dynamic retrospective filtering of physiological noise in bold fmri: Drifter. *NeuroImage*, 60(2):1517 – 1527, 2012.

[149] Theodore D. Satterthwaite, Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, Hakon Hakonarson, Ruben C. Gur, Raquel E. Gur, and Daniel H. Wolf. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64:240 – 256, 2013.

[150] G. D. Logan and W. B. Cowan. On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91(3):295–327, 1984.

[151] M. Usher and J. L. McClelland. The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3):550–592, 2001.

[152] Nicolas P. Rougier, David C. Noelle, Todd S. Braver, Jonathan D. Cohen, and Randall C. O'Reilly. Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences*, 102(20):7338–7343, 2005.

[153] Frederick Verbruggen and Gordon D. Logan. Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience and Biobehavioral Reviews*, 33(5):647 – 661, 2009.

[154] Timothy T. Rogers and James L. McClelland. Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognitive Science*, 38(6):1024–1077, 2014.

[155] HL Fields and SD Anderson. Evidence that raphe-spinal neurons mediate opiate and midbrain stimulation-produced analgesias. *Pain*, 5(4):333–349, 1978.

[156] Anat London, Inbal Benhar, and Michal Schwartz. The retina as a window to the brain—from eye research to cns disorders. *Nature Reviews Neurology*, 9:44–53, 2013.

[157] Timothy J. Buschman, Eric L. Denovellis, Cinira Diogo, Daniel Bullock, and Earl K. Miller. Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron*, 76(4):838 – 846, 2012.

[158] T. Akam and D. Kullmann. Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nature Reviews Neuroscience*, 15:111–122, 2014.

[159] E.H. Smith, G. Horga, and M.J. et al. Yates. Widespread temporal coding of cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 22:1883–1891, 2019.

[160] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, 2000.

[161] Aileen Schroeter, Felix Schlegel, Aline Seuwen, Joanes Grandjean, and Markus 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*, 94:372 – 384, 2014.

[162] Felix Schlegel, Aileen Schroeter, and Markus Rudin. The hemodynamic response to somatosensory stimulation in mice depends on the anesthetic used: Implications on analysis of mouse fmri data. *NeuroImage*, 116:40 – 49, 2015.

[163] Nick Todd, Steen Moeller, Edward J. Auerbach, Essa Yacoub, Guillaume Flandin, and Nikolaus Weiskopf. Evaluation of 2d multiband epi imaging for high-resolution, whole-brain, task-based fmri studies at 3t: Sensitivity and slice leakage artifacts. *NeuroImage*, 124:32 – 42, 2016.

[164] Gustavo Deco, Viktor K. Jirsa, Peter A. Robinson, Michael Breakspear, and Karl Friston. The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLOS Computational Biology*, 4(8):1–35, 08 2008.

[165] Matthew Singh and David Zald. A simple transfer function for nonlinear dendritic integration. *Frontiers in Computational Neuroscience*, 9:98, 2015.

[166] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[167] Adam R. Aron, Trevor W. Robbins, and Russell A. Poldrack. Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4):170 – 177, 2004.

[168] Diane Swick, Victoria Ashley, and A. U. Turken. Left inferior frontal gyrus is critical for response inhibition. *BMC Neuroscience*, 9(1):102, 2008.

[169] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 3:455–492, 1998.

Appendix A

Supplemental Information for Resting-State MINDy Modeling[3]

A.1 Interpreting Model Parameters

As discussed in the main text, the neural modeling framework of MINDy is inherently phenomenological in that it is not directly derived from biophysical first-principals. The weight parameter (W), for instance, serves to measure effective connectivity and should not be confused with synaptic efficacy (or any other directly measurable anatomical metric, such as white matter integrity). The phenomenological nature of these equations gives tractability to the fitting problem. However, this fact does not preclude the model's interpretable and predictive nature. The parametric form that we have chosen leads itself to interpretability by separating the dynamics into three distinct components, interregional-signaling, local decay, and a nonlinear mapping between local excitation and output, which parallel the components used in conventional neural mass models. In the following subsections, we present fuller descriptions of the potential relations between model parameters and underlying biological processes.

---

[3] Appendix chapter reprinted verbatim from previously published work: [8] SI

A.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 (Fig. A.1 A, 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 ([160]) 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; [161],[162]) lead to some uncertainty in timing. In addition, there are methodological limitations inherent in rapid acquisition methods such as multiband imaging ([163]), 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 (Fig. A.1B). 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.

A.1.2 Interpreting Model Curvature

In the original theory of neural mass models ([78], [164]), 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

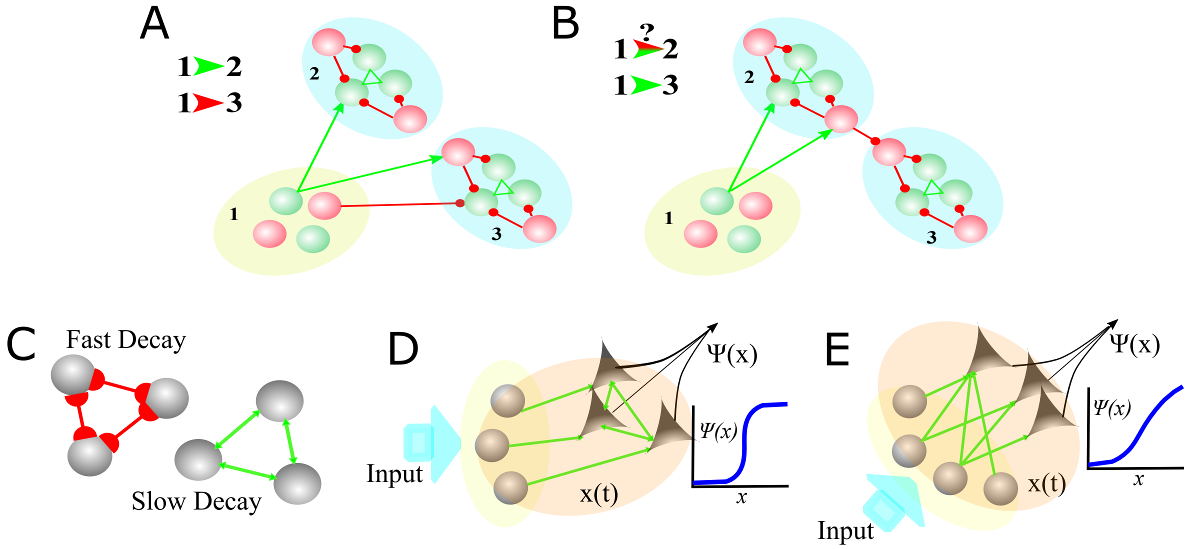

Figure A.1: Interpreting Phenomenological Model Parameters. A) The weight matrix largely captures monotone causal relationships. However, the sign of the causal relationship depends upon the sign of the actual inter-regional connection and which neurons are involved. Excitatory connections/cells are depicted with green arrows and inhibitory in red dots. B) When the sign of connections between regions is mixed, it is possible for indirect relationships to appear stronger than direct connections. Local network structure could influence transfer function and decay parameters. C) Networks with greater reciprocal inhibition (red lines) have a faster time-scale, hence greater decay than those with reciprocal excitation (green). D) A toy example of a network with near binary output due to reciprocal excitation in the output cells (triangles). E) A toy example with a more graded output rule due to inhomogeneities in the excitation of output cells.

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 (Fig. A.1 D, E). 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 (Fig. A.1 E). 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.

A.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 $-Dx$. 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

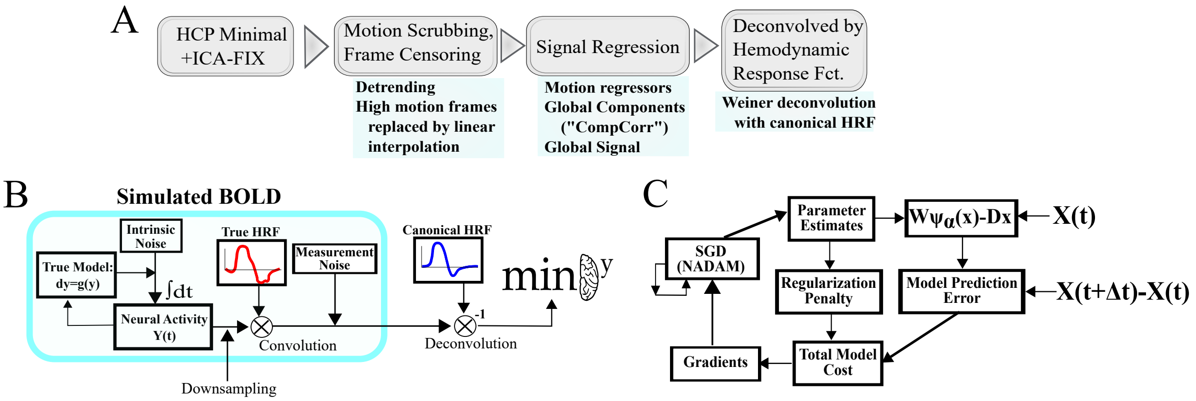
A
HCP Minimal +ICA-FIX
Motion Scrubbing, Frame Censoring
Signal Regression
Deconvolved by Hemodynamic Response Fct.
Detrending High motion frames replaced by linear interpolation
Motion regressors Global Components ("CompCorr") Global Signal
Weiner deconvolution with canonical HRF
B
Simulated BOLD
Intrinsic Noise
True HRF
Measurement Noise
True Model: dy=g(y)
∫dt
Neural Activity Y(t)
Convolution
Downsampling
Canonical HRF
-1
Deconvolution
min y
C
Parameter Estimates
Wψα(x)-Dx
X(t)
SGD (NADAM)
Regularization Penalty
Model Prediction Error
X(t+Δt)-X(t)
Gradients
Total Model Cost

Figure A.2: Schematics of data-processing and data-generation pipelines. A) Secondary preprocessing consisted of the pipeline proposed by Siegel and colleagues ([2]) followed by Weiner deconvolution ([3]) with a canonical HRF function. B) Simulated BOLD signals were generated by integrating MINDy models as stochastic differential equations, downsampling results to the scanner TR, convolving with a (potentially random) HRF and adding measurement noise. The resultant signal was then deconvolved with the canonical HRF. C) MINDy parameter estimation consists of iteratively updating estimates using current and past error gradients according to NADAM ([9]).

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 (Fig. A.1 C). 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 equilibria 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 (see

Supp. for precise conditions). In general, we expect that the decay parameter is related to the relative proportions of local excitation/inhibition within each parcel (Fig. A.1 C). The anatomical distribution of decay terms across parcels was largely consistent across subjects (Fig. 2.4B).

A.2 Derivation of the MINDy Transfer Function

The use of a new transfer function is motivated by the desire to unify the three main classes of transfer function employed in both artificial neural networks and biological neural models: the rectified linear unit (ReLu), softplus, and the logistic sigmoidal function. These functions differ in their curvature (ReLu is piecewise linear, while the others are smooth) and their boundedness (ReLu and softplus are unbounded, linear in the positive limit). Rather than specifying one of these functions explicitly, as is usually done, we chose to create a more general functional form and let the data select the function's shape on a person x region basis. This form consists of a generalized class of sigmoidal functions which can be varied from smooth to piecewise-linear. We'll later show that this property enables approximation of the other two classes (ReLu/softplus) over bounded domains. Our function is generated by integrating the difference of two shifted sigmoidal functions (denoted $\sigma(y)$).

$$\hat{\psi}_\alpha(x) := \int_{-\infty}^{x} \sigma(\alpha(y+.5)) - \sigma(\alpha(y-.5))dy \tag{A.1}$$

By Proposition 1 (below), the shift by ±.5 guarantees that $\hat{\psi}$ will have the same limits as the original function $\sigma(x)$ and also retain any of the original function's reflection symmetries about $x = 0$. Moreover, $\hat{\psi}$ reduces to the definite integral (Proposition 1):

$$\hat{\psi}_\alpha(x) = \int_{x-.5}^{x+.5} \sigma(\alpha y)dy \tag{A.2}$$

When α is small, this formulation generates smooth sigmoidal functions with curvature dependent upon the choice of σ. However, in the limiting case of large α, the function approaches a shifted ReLu function for sigmoids (σ) that have a lower-limit of zero and appropriate rescaling (b):

$$x \leq b \implies \lim_{\alpha \to \infty} b\hat{\psi}_\alpha(b^{-1}x - .5) = ReLu(x) := max(0, x). \tag{A.3}$$

For finite values of α, b the function is smoothed and can be rescaled/shifted to behave like a soft-plus function over desired intervals. An analogous means of generating functions was previously used for modeling dendritic saturation ([165]) starting from the logistic sigmoid function:

$$\sigma(x) := \frac{1}{1 + e^{-x}} \implies \hat{\psi}_\alpha(x) = ln\left(\frac{1 + e^{\alpha x + .5}}{1 + e^{\alpha x - .5}}\right) \tag{A.4}$$

We chose to use the sigmoidal function:

$$\sigma(x) := \frac{x}{\sqrt{1 + x^2}} \implies \psi_\alpha = \sqrt{\alpha^{-2} + (x + .5)^2} - \sqrt{\alpha^{-2} + (x - .5)^2} \tag{A.5}$$

which takes values on $(-1, 1)$ similar to the hyperbolic-tangent (tanh). We favored the chosen sigmoidal basis over tanh/logstic (as was used in [165]), because the resultant transfer function ψ involved slightly faster and more stable operations (i.e. avoided using log). Additional terms to further customize the slope/intercept of the transfer function were initially considered, e.g.:

$$\hat{f}(x) = W\psi_\alpha(b \circ (x + s)) - Dx + c \tag{A.6}$$

However, we observed in early tests that the scaling term could be reduced to a scalar constant ($b = 20/3$) and fitted values for s were effectively zero (for z-scored data). We also observed that when transfer functions were bounded over $[-1, 1]$ the c term became effectively zero which was not the case when we tested transfer functions bounded over $[0, 1]$. Thus, we chose to use functions bounded over $[-1,1]$ so that the s and c terms could be removed. Of course, MINDy models can always be rewritten in an equivalent form featuring a non-negative transfer function and constant drive

$$c_i := -\sum_j W_{i,j}, \tag{A.7}$$

since $\psi(x)+1$ is non-negative and

$$f(x) = W\psi(x) - Dx = W(\psi(x)+1) - Dx + c. \tag{A.8}$$

Proposition 1. *Define the operator* $\Phi : C^0 \to C^1$*:*

$$[\Phi \circ \sigma](x) := \int_{-\infty}^{x} \sigma(y+.5) - \sigma(y-.5)dy. \tag{A.9}$$

Suppose that σ *is non-decreasing and bounded. Then* $\liminf[\sigma] = \liminf[\Phi\circ\sigma]$ *and* $\limsup[\sigma] = \limsup[\Phi \circ \sigma]$*. Moreover,*

$$[\Phi \circ \sigma](x) = \int_{x-.5}^{x+.5} \sigma(y)dy. \tag{A.10}$$

Proof. By Lebesgue's Theorem for Monotone Functions, σ is differentiable almost everywhere and we can write a function $D\sigma$ equal to the derivative of σ at differentiable points and zero otherwise which satisfies:

$$\sigma(x+.5) - \sigma(x-.5) = \int_{x-.5}^{x+.5} D\sigma(y)dy. \tag{A.11}$$

By rearranging limits of integration we produce:

$$[\Phi \circ \sigma](x) = \int_{x-.5}^{x+.5} \int_{-\infty}^{z} D\sigma(y)dydz = \int_{x-.5}^{x+.5} \sigma(y)dy. \tag{A.12}$$

By monotonicity we have:

$$\sigma(x - .5) \leq \int_{x-.5}^{x+.5} \sigma(y)dy \leq \sigma(x + .5). \tag{A.13}$$

By the monotone convergence theorem, σ converges to its infimum/supremum. Taking the negative limits for x:

$$\liminf[\sigma] \leq \liminf[\Phi \circ \sigma] \leq \liminf[\sigma], \tag{A.14}$$

and similarly for the positive limit ($\limsup$). Applying the squeeze theorem completes the proof. □

A.3 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 [9]) which builds upon the earlier ADAM algorithm ([166]). 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:

$$\theta_{k+1} = \theta_k - \eta \frac{\partial E}{\partial \theta_k} \tag{A.15}$$

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 (A.3). 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 ([9]). 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:

$$m_{k+1} = \mu\, m_k + (1-\mu)\frac{\partial E}{\partial \theta_k} \tag{A.16}$$

$$n_{k+1} = \nu\, n_k + (1-\nu)\frac{\partial E}{\partial \theta_k}^2 \tag{A.17}$$

$$\theta_{k+1} = \theta_k - \eta \frac{\frac{1-\mu}{1-\mu^{k+1}}\frac{\partial E}{\partial \theta_k} + \frac{\mu}{1-\mu^{k+2}} m_{k+1}}{\sqrt{\frac{n_{k+1}}{1-\nu^{k+1}}} + \varepsilon} \tag{A.18}$$

Like its predecessor, the ADAM algorithm ([166]), NADAM makes use of momentum to avoid converging to shallow minima and also incorporates estimates of the error surface curvature ([9]). 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 (A.3). 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

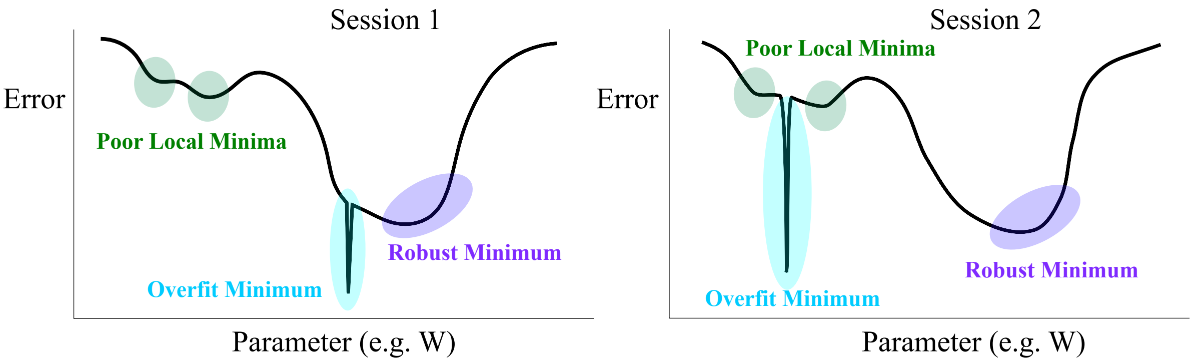
Session 1
Error
Poor Local Minima
Overfit Minimum
Robust Minimum
Parameter (e.g. W)
Session 2
Error
Poor Local Minima
Overfit Minimum
Robust Minimum
Parameter (e.g. W)

Figure A.3: Schematic of NADAM benefits with illustrative error surfaces (y-axis) for fitting a parameter (x-axis values) on the first scanning session (left) and the second (right). The NADAM algorithm uses momentum to avoid shallow local minima (green). This feature also prevents convergence to overly sharp minima (even if they are global) because such error surfaces can often correspond to overfitting (blue) and hence do not generalize across sessions. Rather, NADAM emphasizes solutions to deep basins (purple) which may prove the most robust.

.

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.

A.4 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:

$$J = \frac{1}{2}E_T[\|(X_{T+1} - X_T) - [(W_S + W_L)\psi_A(X_T) - DX_T]\|_2^2] \\ + \lambda_1\|W_S\|_1 + \lambda_2 Tr(|W_S|) + \lambda_3(\|W_1\|_1 + \|W_2\|_1) + \frac{\lambda_4}{2}\|W_L\|_2^2 \quad \text{(A.19)}$$

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 decay parameters for computing moving averages, and one "regularization" parameter ([9]). 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.

A.5 Interpretations of the Weight-Decomposition and Well-Posedness of the Problem

From a Bayesian perspective, this penalty function is equivalent to maximum a posteriori (MAP) estimation with fixed Laplace distribution (symmetric exponential) priors for each of the individual weight matrices and a normal prior on the combined low-rank component. The Laplace (distribution) prior is unrelated to Laplace approximation as used in Bayesian estimation. Since we assume that process noise is iid. between parcels, its influence (scaling the prediction error term) gets absorbed in the regularization coefficients (by multiplying all terms of the log-likelihood by the noise variance). 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 quite flexible. The values of each λ_i and their determination is detailed in SI Table A.13 and SI Sec. A.4.

The primary function of this decomposition is to prevent over-fitting. In all contexts, the potential for over-fitting is related to the difference of model and data degrees of freedom (parameters vs. measurements). While the scale of the current MINDy model might induce initial skepticism, the current problem is similarly well-posed as several recent attempts at fitting a smaller number of parameters. For instance, a recent approach by Wang and colleagues ([71]) to fit just local parameters (2 per node) in modeling the functional connectivity matrix results in 138 parameters estimated from 2,278 data points collapsed across the whole brain (approximately 16.5 measurements per parameter). Although the current approach estimates far more parameters (n+2 per node) it also utilizes many more data points: the whole multivariate time series is used for estimation rather than just the functional connectivity matrix. This results in a ratio of roughly 12 measurements per parameter using the full HCP resting-state data for subjects with MMP cortical and Freesurfer subcortical parcellation (11 for the gwMRF-400). The HCP dataset also contains double this quantity for a subset of subjects who also participated in a later retest session. Of course, these “back-of-the-envelope” calculations assume the worst-case scenario of no parameter covariation. In reality, we expect the set of underlying effective connectivity matrices to be much more constrained—a fact that we exploit via our weight decomposition. Moreover, the sparse regularization priors result in many weights becoming negligible (i.e. very near zero), so even fewer non-trivial estimates are made (Fig. 2.5A,B). The fitting process is also tractable due to the use of the NADAM algorithm ([9]) which is optimized for simultaneously fitting very large numbers of parameters in a highly efficient manner (approximately one minute per model on a latop).

A.6 Reliability and Individual Differences in the Weight Decomposition

In the main text (Sec. 2.2.2), we introduced a linear decomposition of the weight matrix into sparse and low rank components:

$$W := W_S + W_L = W_S + W_1 W_2^T. \tag{A.20}$$

This decomposition was motivated by the dual influences of sparse, long-distance connections between "hub" regions (which motivates W_S) and the propagation of these signals along subnetworks (which motivates W_L). This formulation was developed as a fitting heuristic by which we could approach the high-dimensional model-estimation inherent in MINDy (Eq. 2.4). All previous analyses have focused upon the final weight matrix W, since the dynamical systems models in MINDy do not require explicit consideration of the components (W_S and W_L). In this section, we present preliminary analyses which suggests reliable individual differences in this decomposition. We do not separately analyze the two rectangular matrices W_1 and W_2 which define W_L since they are not unique (e.g. their column indices are arbitrary). As with the full weight-matrix, we measured the degree to which estimates were similar within-subject (different sessions) vs. between-subject. We only consider non-diagonal elements of the matrices since the recurrent elements (diagonals) are distinguished by a separate regularization term in the cost function (Eq. 2.4). The results for recurrent connections in isolation were: within-subject: $r = .56 \pm .06$, between-subject: $r = .42 \pm .07$. We found that both components had greater similarity within-subject ($W_S : r = .53 \pm .06$ and $W_L : r = .67 \pm .04$) than between-subjects ($W_S : r = .35 \pm .03$ and $W_L : r = .39 \pm .04$). Thus, the component matrices had significant reliability and exhibited individual differences although the reliabilities were lower than that of the full weight matrix (see Sec. 2.3.3).

We also considered the degree to which sparsity vs. low-dimensionality of a subject's weight matrix was a reliable trait. We quantified this value by comparing the (log) relative magnitude of the sparse and low-rank matrices (U):

$$U := ln\left(\frac{\|W_S\|_F^2}{\|W_L\|_F^2}\right) \tag{A.21}$$

With $\|W\|_F^2$ denoting the squared Frobenius-norm (sum of squared matrix elements). Each model produces a single (scalar) value for U. We tested whether individual differences in this quantity were reliable either with or without including recurrent connections. We found that results were reliable in either case (with: $ICC = .763$, without: $ICC = .710$) and that individual subject's values were highly correlated for the two cases ($r(51) = .919$ for the mean across sessions). When recurrent connections are included, the sparse component is larger (i.e. $U > 0$) on average: $U = .265 \pm .261$ while the low-rank component otherwise dominates: $U = -.528 \pm .215$. Thus, results indicate that the relative magnitude of sparse vs. low-rank components is a reliable marker of individual differences in MINDy models. However, the degree to which the overall weight matrix is dominated by either component will depend upon whether the recurrent (sparse) connections are considered. Moreover, we expect that the average value of this ratio will depend upon the particular choice of regularization parameters which will favor either component. Therefore, while results are promising in

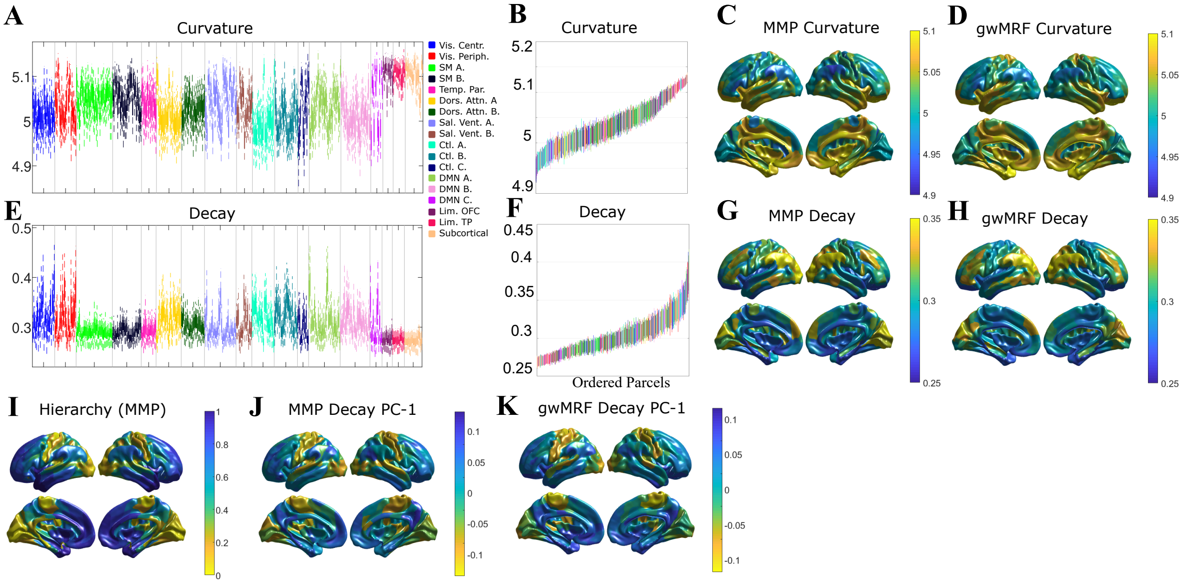

Figure A.4: Local MINDy parameters (curvature and decay) exhibit consist anatomical structure within and between networks. A) Distribution of curvature parameter values for each brain parcel grouped according to network (17-network [6]). B) Curvature parameters reordered according to mean demonstrate that within-network variability is also consistent. C) Anatomical profile of group-mean curvature for the MMP atlas ([10]). D) Profile for the gwMRF ([6]) parcellation. E-H) same as A-D but for the decay parameter. I) Hierarchical Heterogeneity map by Demirtas and colleagues ([5]) using group T1/T2 ratio. J,K) same as G,H but for the first principal component of decay across subjects.

terms of individual differences, we do not recommend using the weight decomposition to quantify the general sparseness/dimensionality of brain networks without considering the influence of regularization hyperparameters.

A.7 The Influence of Hyperparameter Choices on Sparsity

The relative value of the hyperparameters λ_1 vs. λ_2 can influence the sparsity of the MINDy connection matrix. Overly small values of λ_1 will not generate sparsity in the "sparse" matrix (W_S) while overly large values of λ_1 relative λ_2 will also decrease sparsity by biasing solution coefficients toward the low-rank component (W_L). However, in practice, even the low-rank matrix is substantially sparser than the rsFC which generates overdispersion ("heavy-tailedness"). We quantify this property via kurtosis:

$$Kurt[X] := E\left[\left(\frac{X - E[X]}{\sigma[X]}\right)^4\right] \tag{A.22}$$

which, for a normal distribution, is 3. The kurtosis for each subject's low-rank component (35.6 ± 8.1) substantially exceeds that of the rsFC matrix (7.2 ± 0.9) although both are dwarfed by the sparse component (188.7 ± 27.8). Thus, both components of the weight matrix are more sparse than $rsFC$ so hyperparameter choices which bias towards either term will still result in sparser solutions than rsFC. Lastly we considered the case in which all regularization terms are equal zero which forms a lower-bound case for sparsity (i.e. all other regularization values should produce more sparse estimates). Individual model estimates without regularization are extremely noisy (hence the need for regularization) and less sparse ($Kurt = 5.9 \pm 0.7$) than rsFC. However, the group-mean of these noisy estimates ($Kurt = 24.4$) is also more sparse than the group-mean for rsFC ($Kurt = 9.0$). Moreover, the group-mean without regularization was highly correlated ($r(175559) = .917$) with the group-mean for the full MINDy model. We conclude that individual weight estimates are sparse for a range of hyperparameter values and the group mean of estimated weights for our dataset is more sparse than rsFC, irrespective of hyperparameter choices.

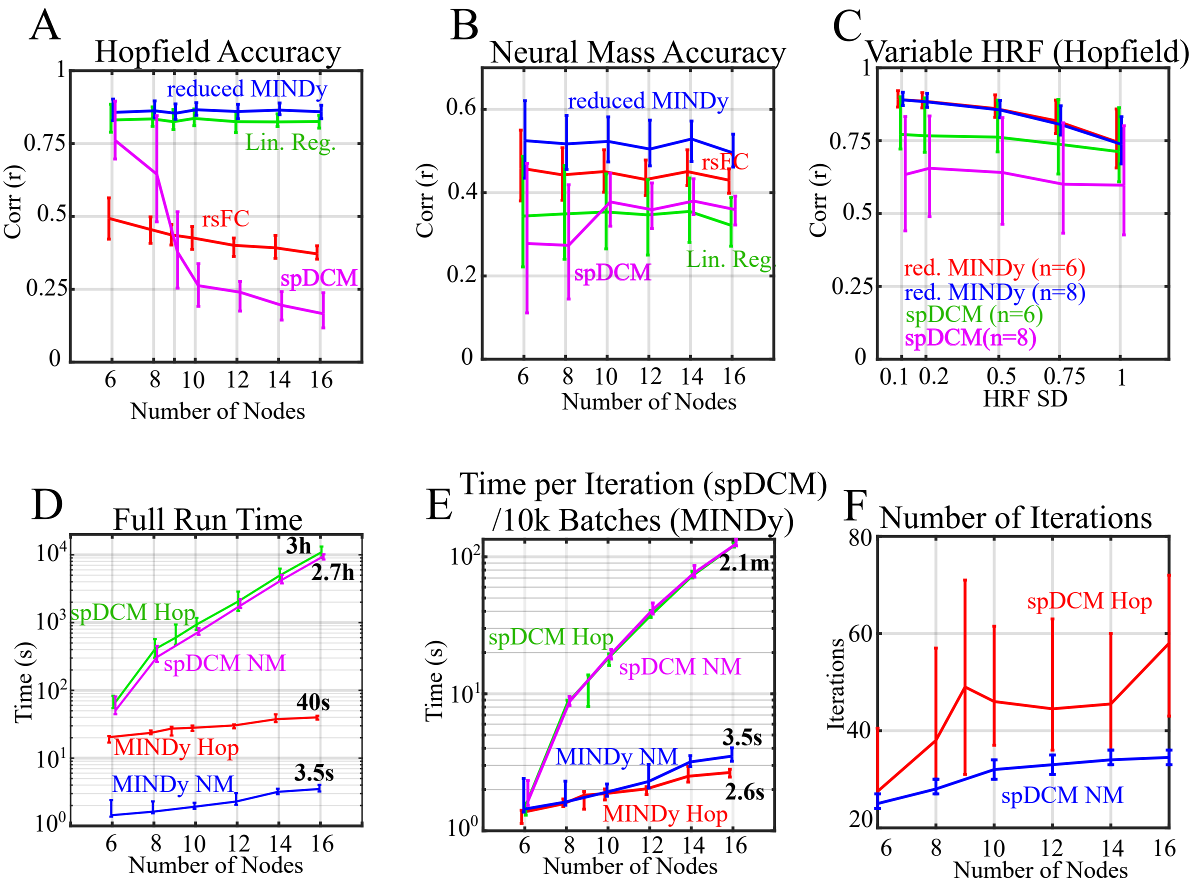

Figure A.5: Comparison of accuracy and run-time for MINDy and spDCM. A) Accuracy in estimating ground-truth connectivity from Hopfield-network simulations by network size. “Reduced MINDy” indicates that all regularization terms were removed from MINDy to avoid bias (analogous results for the full MINDy model are in Sec. 2.3.7). Lines indicate mean and bars indicate first/third. quartiles. B) Accuracy in estimating ground-truth connectivity from Neural Mass simulations by network size. Results for the full MINDy model are in Sec. 2.3.7. C) Model performance as a function of HRF spatial variability for two network sizes: 6 and 8 nodes. Note that MINDy performance decreases with HRF spatial variability, whereas the effect for spDCM is minor. D) Full run times for MINDy and spDCM for each simulation as a function of simulation type/size. E) Run time per EM iteration (spDCM) and for 10,000 mini-batches in MINDy. We chose to compare with 10,000 mini-batches so that the run-times would be comparable for the smallest network size (n_{Pop}=6). F) Model performance as a function of HRF spatial variability for two network sizes: 6 and 8 nodes. Note that MINDy performance decreases with HRF spatial variability, whereas the effect for spDCM is minor.

A.8 Comparing MINDy and Spectral Dynamic Causal Modeling

We view the primary contributions of MINDy in its scalability, biological interpretability, and the ability to predict nonstationary resting-state dynamics. However, one non-unique benefit of MINDy is the data-driven characterization of effective connectivity via the weight parameter. Other methods, such as stochastic DCM and spectral DCM have also used (linear) dynamical systems models to estimate effective connectivity. By converting problems into the frequency domain, spectral DCM (spDCM; [57]) has been applied to brain models consisting of 36 regions, but still has significantly higher computational cost than MINDy. One question, therefore, is whether the scalability of MINDy comes at the cost of accuracy. We tested this question in a series of ground-truth simulations. To be clear, we are not seeking to demonstrate that MINDy is necessarily a better estimator of low-dimensional effective connectivity, but rather that the scalability of MINDy does not significantly impair accuracy (i.e. MINDy is at least as good as DCM). One inherent advantage of spectral DCM is the ability to estimate region-specific hemodynamic kernels which is not part of the currently proposed MINDy model (although extensions for HRF estimation are being developed [99]). Thus, we consider two features when comparing MINDy and spectral DCM (spDCM): scalability and robustness to spatial variation in the HRF.

A.8.1 Benchmarking with Unbiased Ground-truths

The main difficulty in comparing MINDy and DCM is the different underlying assumptions—spDCM, for instance, has only been validated using a linear ground-truth ([57]). We took a number of steps to prevent bias based upon differing assumptions (SI Tab.A.1) and when bias was inevitable, we made choices that favored spDCM. First, we used two ground-truths which were not based upon either technique: a neural mass + balloon-Windkessel ground-truth and a continuous asymmetric Hopfield-model ground truth with either a random global HRF or spatially variable hemodynamics (the same models as for model-mismatch analyses in Sec. 2.3.7). Random HRFs were generated by sampling $\alpha_1 \sim \mathcal{N}(6, \sigma^2)$ and $\beta_1 \sim \mathcal{N}(1, (\sigma/6)^2)$ in Eq. 2.5. For the global HRF simulations, $\sigma = .5$ and the same values of α_1 and β_1 were used for each node (for a given simulation). In the spatially-variable HRF simulations, values were independently drawn for each node. We implemented spectral-DCM using the MATLAB code provided with SPM-12 (function name: "spm_dcm_fmri_csd") to compute the expected value for each connection weight based upon cross-spectral density (spectral DCM). Simulations and model-fitting were performed single-core on Intel Xeon E5-2630v3 CPUs. Since linear models like DCM do not separate recurrent connections and decay, we only compared accuracy for non-recurrent (off-diagonal) elements of the connectivity matrix for each technique. Both ground-truth simulations were integrated at time-scales faster than the sampling rate (dt=.025 vs. 725ms TR for the neural mass and dt=100ms vs. 700ms TR for Hopfield to mirror HCP acquisitions). This feature ensures that results are not biased against spDCM due to simulation time-scale, since MINDy discretizes the model in terms of TR, while spDCM maintains a continuous-time estimation framework.

We used the same hyper-distributions to parameterize neural-mass simulations and Hopfield networks as in Section 2.3.7, but for smaller network sizes. We made three further adjustments to reduce bias: first, we set all regularization terms from MINDy equal to zero, so there would be no inherent advantage to MINDy based upon the hyper-distributions of network structure. The regularization-based calculations were still performed (preserved run time), but they had no effect on the solution (were always zero). We also increased the simulation length of the neural-mass simulation from 3000 to 5000 TRs and removed the nearest-neighbor smoothing from the Hopfield Network simulations which had been included to mirror the empirical processing-pipeline. These adjustments did not affect MINDy but were found to increase the performance of spDCM.

In addition to MINDy and spDCM, we tested the accuracy of functional connectivity and one-step prediction by multiple-regression (solving $\Delta x = Mx_t$). The latter case provides an additional control by providing an alternative method to parameterize linear dynamical systems. This control is important because it can indicate that cases in which spDCM underperforms are due to the estimation technique rather than linearity per se. (i.e. cases in which multiple-regression is accurate but spDCM is not). When all regularization terms are set equal to zero and the transfer function is linear, the MINDy and multiple-regression models are equivalent. The regression approach differs from spDCM, however, in the strength of the linearity assumption. Whereas spDCM seeks a linear model that best explains statistics drawn from the full time-course (i.e. assumes global linear dynamics), the regression approach (like MINDy) considers the changes at each TR (i.e. the collection of local dynamics).

A.8.2 MINDy Performs Competitively with DCM

Results indicated that MINDy scaled-well in terms of performance and run-time. Moreover, this scalability did not generally come at a cost to performance relative contemporary higher-complexity techniques (i.e. spectral DCM). In all situations tested (model x size x HRF), the reduced MINDy model (no regularization) performed at least as well as all competitors in retrieving ground-truth connection weights (SI Fig. A.5 A-C). All methods performed poorer for the neural-mass ground-truth (SI Fig. A.5B; SI Tab. A.2) than for the Hopfield network (SI Fig. A.5A; SI Tab. A.3; SI Tab. A.4) which was expected due to the greater difficulty of the problem (much faster time-scales and more complicated models). We note that the full MINDy model performs substantially better in neural mass simulations than the reduced version (see Sec. 2.3.7). However, even when regularization was removed to prevent bias due to assumptions on network structure ("reduced" MINDy), performance remained competitive with spDCM. We also observed that spDCM performance decreased with network size for the Hopfield simulation (SI Fig. A.5B; SI Tab. A.3), but the regression-based model did not. Thus, this feature cannot be explained solely in terms of functional form (linearity) although the reduced MINDy did consistently outperform regression.

Moreover, the computational complexity of MINDy is substantially lower than that of spDCM. The theoretical limiting computational complexity of MINDy is a second-degree polynomial in the number of nodes since the highest-complexity operation in terms of n_{Pop} is analytically calculating the error-gradient with respect to the weight-matrix ($\mathcal{O}(n_{Pop}^2)$) although the empirical complexity was substantially lower for these simulations (the quadratic term only dominates for much larger n_{Pop}; SI Fig. A.5D). By contrast, the spectral-DCM code packaged with SPM has at least fourth-order complexity in terms of the population size for a fully-connected model (SI Fig. A.5D). Each Expectation-Maximization (EM) iteration is dominated by $\mathcal{O}(n_{Pop}^4)$ (SI Fig. A.5E), but the total complexity can be even greater if the number of iterations until convergence also increases with n_{Pop} (SI Fig. A.5F). For instance, the median total runtime for the neural mass and Hopfield network simulations scaled with $\mathcal{O}(n^{5.66\pm.44})$ and $\mathcal{O}(n^{5.85\pm.38})$, respectively (95% confidence estimated using the "fit" function in MATLAB 2020a for linear power functions: "power1"). The empirical complexity of each EM iteration was roughly the theoretical limit of 4: $\mathcal{O}(n^{3.84\pm.20})$ and $\mathcal{O}(n^{3.97\pm.16})$ for the neural mass and Hopfield simulations, respectively.

However, even the minimal case of fourth-order complexity can severely limit scalability. For context, a recent "large-scale" spectral DCM paper ([57]) employed 36 brain ROIs with a run-time of between 1,280 and 2,560 minutes per model (between 64 and 128 iterations at 20 minutes each). Increasing resolution from 36 ROIs to the 419-node parcellation we employed (19 subcortical + 400 cortical [6]) would increase CPU time by a factor of over 18,000 (roughly 44 to 89 years per model for the same data and hardware as [57]). By contrast, the current MINDy models for HCP data were locally fit on a laptop in less than one minute each (Intel i7-8750H CPU, 2.2GHz, 6 cores).

One area in which spectral DCM proved advantageous, however, was in robustness to spatial variability in the hemodynamic response (SI Fig. A.5.C; SI Tab. A.4). As with the analogous simulations in Section 2.3.7, MINDy performance decreased with the underlying HRF's spatial variability (see also [99]). This pattern is expected since the currently proposed MINDy assumes a fixed HRF which these simulations violate. The performance benefit of MINDy over spDCM, likewise decreased with HRF spatial variability. For the smallest network (6 nodes) and highest level of HRF variability considered, the difference between models' accuracy was negligible (although statistically significant; SI Tab. A.4). Thus, there may be cases of extreme HRF spatial variability in which spectral DCM outperforms MINDy for sufficiently small networks. Although MINDy scales significantly better than competing approaches like spectral DCM, the current version is less robust to spatial variability in the HRF (although see [99] for upcoming extensions). Nonetheless, for all simulations considered, MINDy performed competitively with spDCM within the latter's scope. We conclude that MINDy performs at least as well as spDCM, while scaling far better.

Table A.1: Types of assumptions ("Difference") made by MINDy vs. spDCM and our controls to mitigate these differences in simulated comparisons (SI Sec.A.8).

Difference	MINDy	spDCM	Adjustment/Control
Model Form	Nonlinear Network	Linear Time-Invariant	Model Mismatch; Regression Control
Regularization	Yes	No	Set MINDy regularization equal 0
Time-Scale	Discrete (TR)	Continuous	Simulation dt<<TR
Local Components	Recurrent + Decay	Decay	Only tested non-recurrent connections
HRF	Fixed	Local Estimates	Parametrically varied spatial HRF

Table A.2: Accuracy (r) for each method (Regr.=regression) for the neural-mass simulation (SI Sec. A.8) with variable population sizes (left side). Test-statistics (paired t-test; 2-tailed) for MINDy-spDCM are provided on the right side. The number of nodes per simulation is listed under "nodes" whereas the number of simulation instances is listed under "N". The paired difference in accuracy between MINDy and spDCM is denoted Δr. We denote $* = p < .01$ and $** = p < .001$.

Nodes	rsFC	Regr.	MINDy	spDCM		N	Δr	t
6	.46(.14)	.34(.20)	.52(.15)	.28(.24)		240	.25(.26)	14.8**
8	.44(.10)	.35(.16)	.52(.11)	.27(.20)		320	.24(.21)	20.4**
10	.45(.08)	.35(.12)	.52(.09)	.38(.10)		130	.15(.10)	17.3**
12	.43(.07)	.35(.11)	.50(.09)	.36(.09)		80	.15(.08)	16.6**
14	.45(.06)	.35(.09)	.53(.07)	.38(.07)		40	.15(.06)	14.5**
16	.43(.05)	.32(.09)	.50(.08)	.36(.06)		20	.14(.03)	22.8**

Table A.3: Accuracy (r) for each method (Regr.=regression) for the Hopfield-network simulation (SI Sec. A.8) with variable population sizes (left side) and a random global HRF ($\sigma = .5$). For this simulation we added a new population size (9 Nodes) post-hoc to see whether the unexpected decrease in spDCM accuracy between 8 and 10 nodes was continuous (it was). Since this special case contained an odd number of nodes the hyperdistribution parameter q in Sec. 2.2.6 was always equal to one (instead of one and two with equal probability). Test-statistics (paired t-test; 2-tailed) for MINDy-spDCM are provided on the right side. The number of nodes per simulation is listed under "nodes" whereas the number of simulation instances is listed under "N". The paired difference in accuracy between MINDy and spDCM is denoted Δr. We denote $* = p < .01$ and $** = p < .001$.

Nodes	rsFC	Regr.	MINDy	spDCM		N	Δr	t
6	.49(.11)	.83(.08)	.86(.07)	.76(.20)		200	.10(.19)	7.0**
8	.46(.08)	.83(.06)	.86(.05)	.65(.23)		200	.22(.22)	14.3**
9	.44(.06)	.83(.06)	.85(.05)	.38(.19)		105	.48(.19)	25.5**
10	.43(.06)	.84(.05)	.87(.04)	.26(.12)		80	.60(.12)	45.1**
12	.40(.06)	.83(.05)	.86(.04)	.24(.10)		50	.62(.11)	39.7**
14	.39(.05)	.82(.04)	.86(.03)	.20(.07)		38	.67(.08)	48.9**
16	.37(.04)	.83(.04)	.86(.03)	.17(.08)		13	.69(.08)	30.0**

Table A.4: Accuracy (r) for each method (Regr.=regression) for the Hopfield-network simulation (SI Sec. A.8) with spatially variable HRF and 6 nodes (left side). Test-statistics (paired t-test; 2-tailed) for MINDy-spDCM are provided on the right side. The standard-deviation of the HRF parameters is listed under "σ" whereas the number of simulation instances is listed under "N". The paired difference in accuracy between MINDy and spDCM is denoted Δr. We denote $* = p < .01$ and $** = p < .001$.

σ-HRF	rsFC	Regr.	MINDy	spDCM		N	Δr	t
.1	.48(.10)	.87(.06)	.89(.04)	.77(.18)		280	.12(.17)	11.8**
.2	.48(.11)	.86(.06)	.89(.05)	.77(.19)		280	.12(.18)	11.1**
.5	.49(.12)	.83(.08)	.86(.07)	.76(.18)		280	.10(.18)	9.3**
.75	.47(.12)	.80(.10)	.82(.10)	.74(.20)		280	.08(.21)	6.5**
1.0	.47(.11)	.72(.16)	.75(.15)	.71(.21)		280	.03(.21)	2.7*

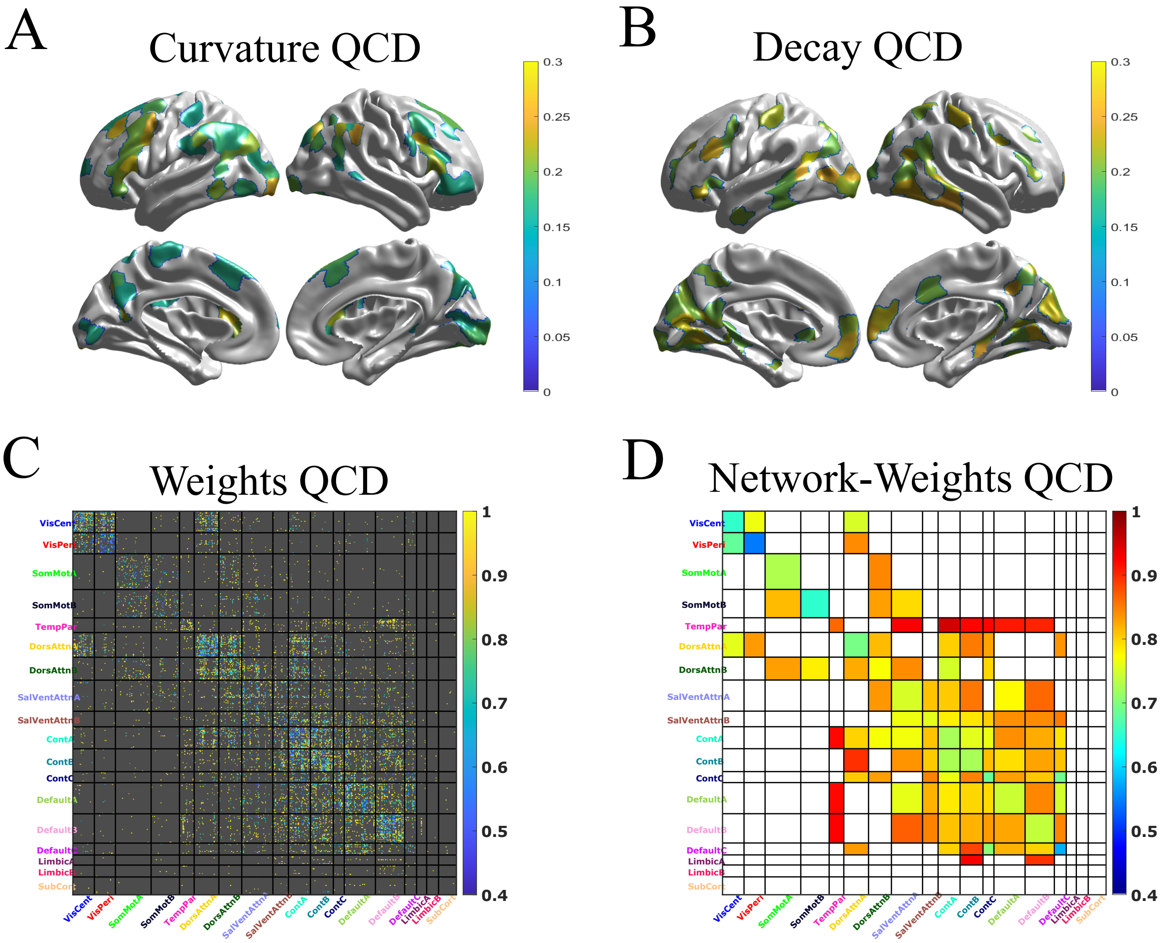

Figure A.6: Anatomical distribution of inter-individual variation in MINDy. A) QCD of the curvature parameter with top-20% threshold. B) Same as A) but for the decay parameter. C) QCD for MINDy connection weights. Weights in which the sign was inconsistent across subjects (<75% agreement) or low reliability (Fisher ICC<.5) were censored (grey). D) Mean weight QCD within each network combinations. If over 95% of parcel-wise connections were censored, the network-level connection was also censored (white).

A.9 Anatomical Distribution of Individual Differences

In concert with the previous analyses of individual differences (Sec. 2.3.5) across parameter-types (weights, curvature, decay) we also investigated the anatomical distribution of individual differences within each parameter-type. These analyses are post-hoc (exploratory) so we report results as a potential launching pad for future investigations and as a means to understand how MINDy models encode individual differences. We do not perform hypothesis-testing and we caution against interpreting these analyses as stand-alone findings due to their exploratory nature and relatively low sample-size. We also note that these analyses are performed upon a biased sample of subjects—those that had no high-motion scans (>1/3 frames censored) so these results may also fail to describe variability in the full HCP subject pool (see [2] for cognitive covariates of motion) and its target population (American young adults). These caveats aside, we considered the degree of individual variation for each parcel/connection. We used the quartile coefficient of dispersion (QCD) to quantify the degree of variability within parcels/connections. Conceptually, the QCD is a robust analogue of the more commonly used Coefficient of Variation and is defined in terms of the first (Q_1) and third (Q_3) data quartiles as:

$$QCD := \frac{Q_3 - Q_1}{Q_3 + Q_1} \tag{A.23}$$

The reason that we apply QCD instead of Coefficient of Variation is that, while both assume the true population follows a ratio-scale (e.g. are one-sided) the QCD is robust to extreme values which, due to measurement error violate the ratio-scale assumption. We censored connections which were unreliable (Fisher's ICC$\leq$.5) or incompatible with QCD (Q_1 and Q_3 differed in sign). We then transformed the other variables onto admissible distributions (non-negative ratio scales) by shifting the curvature and decay parameters so that they had a minimum value of zero. Interestingly, the anatomical distribution of QCD appeared to differ between the curvature and decay parameters. The curvature had the highest QCD in parcels of inferior frontal gyrus, early visual cortex, and a large posterior section of frontal cortex (SI Fig. A.6A). For the decay parameter, QCD was highest in visual regions and the bilateral portion of somatosensory cortex traditionally associated with hands (SI Fig. A.6B). By contrast, connection weights had the lowest QCD for connections within the visual networks (especially peripheral visual; SI Fig. A.6C,D). Some of the highest QCD connections involved the temporal-parietal network (inter and intra-network) and connections to the limbic system. Of course, as previously mentioned, these analyses are purely exploratory and should only be interpreted as an example of how MINDy separates the sources of individual differences (weights, curvature, decay) rather than as a basic neuroscientific result.

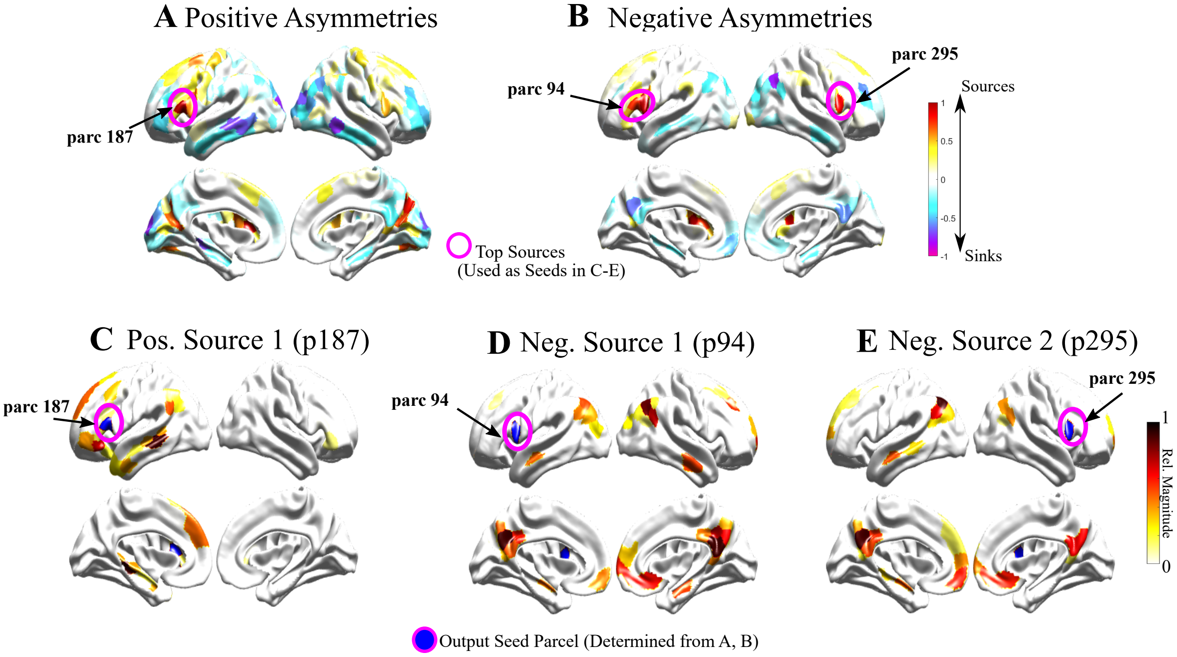
A Positive Asymmetries
B Negative Asymmetries
parc 187
parc 94
parc 295
Sources
Sinks
1
0.5
0
-0.5
-1
Top Sources
(Used as Seeds in C-E)
C Pos. Source 1 (p187)
D Neg. Source 1 (p94)
E Neg. Source 2 (p295)
parc 187
parc 94
parc 295
1
Rel. Magnitude
0
Output Seed Parcel (Determined from A, B)

Figure A.7: Connection asymmetries identified by MINDy 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 ([6])

A.10 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. A.7 A) and negative connections (SI Fig. A.7 B). For positive sources MINDy most strongly identifies inferior frontal gyrus (IFG), bilateral parieto-occipetal sulcus, and dorsal prefrontal cortex (SI Fig. A.7 A). MINDy identifies the strongest excitatory source as a region of left IFG (see main text; SI Fig. A.7 C) 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. A.7 D,E). The role of right IFG in inhibition is well-documented within neuroimaging (e.g. [167]) and lesion studies suggest an inhibitory role for left IFG as well ([168]). These results indicate that the asymmetries within MINDy weights are functionally interpretable. However, these initial findings only scratched the surface of possible analyses.

A.11 Nonlinear Dynamics in MINDy

As discussed in the main-text (Sec. 2.3.6), the fact that MINDy is nonlinear does not inherently imply that the model behaves qualitatively different from models relying upon a linear approximation (e.g. DCM). This distinction is critical to understanding how MINDy models resting-state dynamics: as random fluctuations about a single equilibrium (like DCM) or as topologically significant (nontrivial) dynamics. In general, proving the existence of global behavior in MINDy models constitutes a nontrivial endeavor given their high-dimensionality. However, we can easily rule out trivial dynamics (a Lyapunov-stable global attractor) by examining eigenvalues of the Jacobian at zero. Since all subjects contain at least one positive (real part) eigenvalue (SI Fig. A.8), we conclude (by Proposition 2) that no empirically-parameterized MINDy model is globally Lyapunov-stable.

Proposition 2. *Consider a continuous-time dynamical system evolving according to* $\dot{x} = f(x)$*, with* $f : \mathbb{R}^n \to \mathbb{R}^n = W\psi(x) - Dx$ *with* $\psi : \mathbb{R}^n \to \mathbb{R}^n$ *an odd function* ($\psi(-x) = -\psi(x)$) *and* $n \times n$ *matrices* W, D*. Suppose that the Jacobian at the origin* ($F(0)$) *has at least one eigenvalue with positive real part and none with zero real part. Then for any fixed point* $x_s : f(x_s) = 0$*, at least one of the following hold:*

1. *There exists a set* U *of non-zero measure, whose positive limit-set* $\omega^+(U)$ *does not contain* x_s*.*
2. x_s *is not Lyapunov stable—i.e. there exists* $\epsilon > 0$ *for which there is no* $\delta > 0$ *satisfying* $\|x(0) - x_s\| < \delta \implies \forall t > 0, \|x(t) - x_s\| < \epsilon$*.*

Proof. We consider two cases depending upon whether $x_s = 0$. If $x_s \neq 0$ is a fixed point then so is its reflection $-x_s \neq x_s$ since $f(-x) = -f(x)$. Suppose that x_s is locally Lyapunov stable (violating implication 2) and thus possesses an attractive basin V of non-zero measure. Then $-V$ is an attractive basin of $-x_s$ hence $x_s \notin \omega^+(-V)$ which confirms the first implication. Thus, consider the alternative case: $x_s = 0$. By the hypothesis, x_s is a hyperbolic fixed-point so there exists an open neighborhood N containing x_s for which the dynamics on N are topologically conjugate those of the linearization $\dot{x} \approx F'(0)x$ (Hartman-Grobman Theorem). The linearization $F'(0)$ is unstable (at least one eigenvalue with positive real part) which implies that any sufficiently small ball about $x_s = 0$ will also be an unstable set. This contradicts Lyapunov stability (confirming implication 2). $\square$

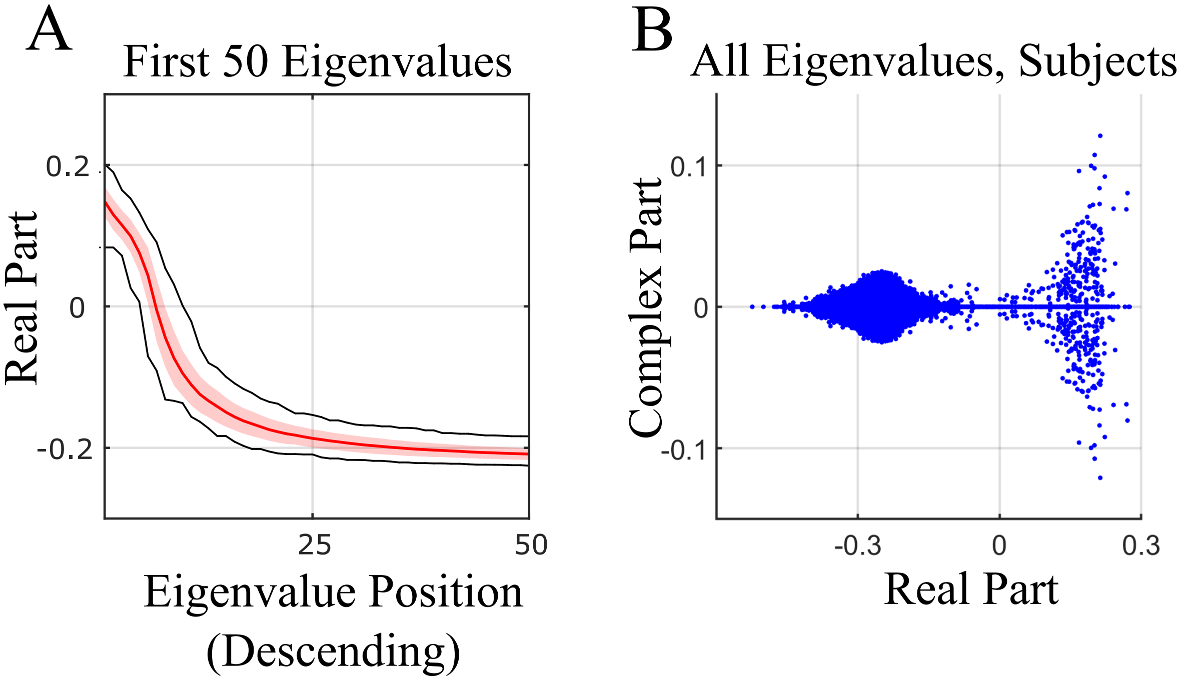

Figure A.8: Eigenvalue analyses indicate that empirical MINDy models do not possess a global, Lyapunov stable attractor. A) Distribution of eigenvalues (real-part) for local-linearization about the origin. Shading indicates ±SD and black lines give the data's maximum/minimum. Note the presence of positive eigenvalues which indicate nontrivial dynamics (Proposition 2). B) Scatterplot of these eigenvalues (all subjects) in the complex plane suggests that the greatest "spin" (complex components) occurs along the unstable subspaces (corresponding to eigenvalues of positive real-part).

Table A.5: Ground-truth validation performance of MINDy and rsFC in recovering the weight matrix of a single subject and the arithmetic difference of weight matrices between subjects. We denote significance with $^{**} = p < .001$, 2-tailed for the contrast Weights minus rsFC.

	Weights	FC	paired-t (df=33)
Single Subject	.800 (.025)	.436 (.065)	45.629**
Indiv. Differences	.544 (.049)	.293 (.046)	58.618**

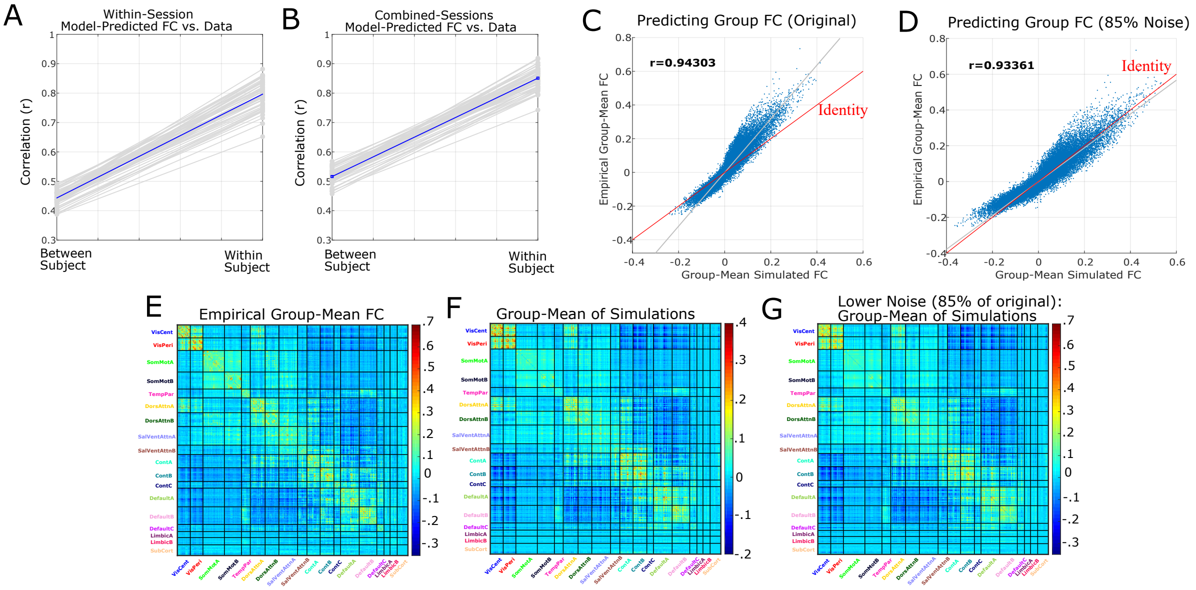

Figure A.9: MINDy strongly predicts individual and group FC. A) Correlation between empirical FC and simulated FC from the same scanning day for either the same subject or a different subject (mean across other subjects). Blue line indicates group mean. Correlations are averaged across scanning sessions. 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. The identity line (perfect match) is indicated in red G) Same as F) but with a slightly decreased noise term (85% of original).

Table A.6: Sensitivity of MINDy weights and rsFC to changes in non-connectivity parameters in a ground-truth simulation Performance is measured in terms of false positives=the percentage of connections that change (thresholded by $p < .05$) due to a change in the ground truth model's curvature/decay. Thus, lower values indicate less false positives (less sensitivity) due to non-connectivity variables. We denote significance with $^{**} = p < .001$, 2-tailed for the contrast Weights minus rsFC.

	Weights	FC	paired-t (df=339)
Changing Curvature	.063 (.002)	.079 (.027)	-10.639**
Changing Decay	.076 (.006)	.2749 (.120)	-31.753**

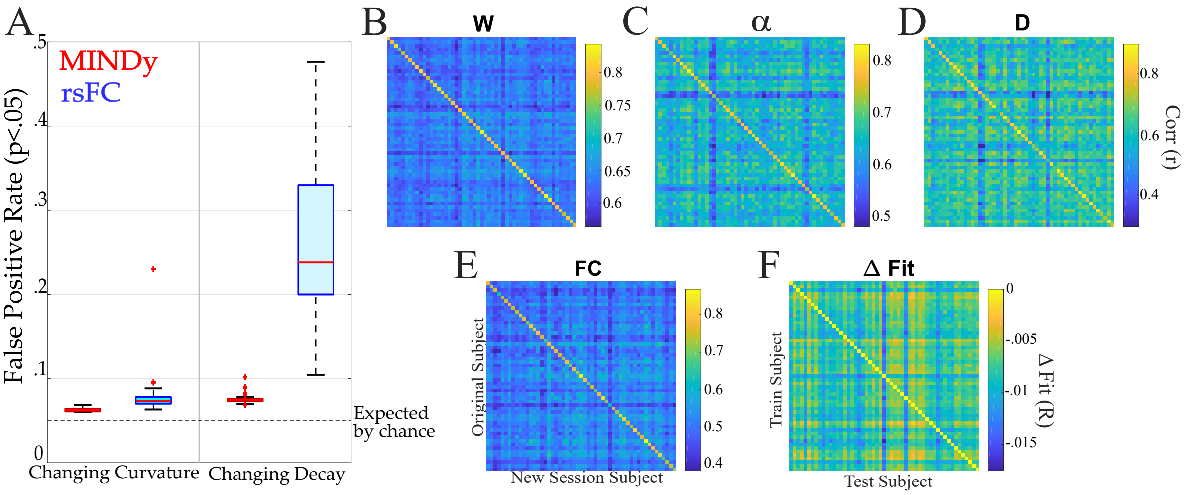

Figure A.10: MINDy differentiates model parameters between individuals and identifies the source of individual differences. A) Changing only the curvature (left) or decay (right) parameters of a simulated subject has more impact on the simulated rsFC (blue) then on MINDy weight estimates refit to the 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.

Table A.7: Test-retest correlation across scanning sessions for MINDy parameters when sessions were drawn from the same subject or from different subjects. Group-level statistics are present in mean correlation (SD) form. Group level permutation testing (100,000 each) produced $p's \approx 0$ 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

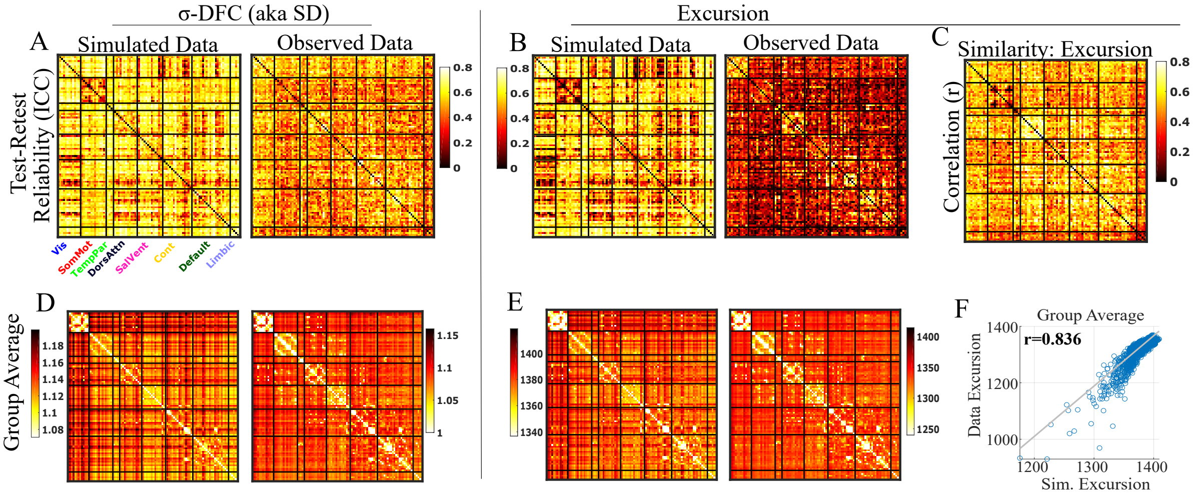

Figure A.11: MINDy reproduces reliable, accurate estimates of dynamic functional connectivity (DFC). We use σ-DFC to disambiguate the standard-deviation measure for DFC (e.g. [11],[12]) from other uses of standard-deviation. A) Data simulated from test-retest models (models fit to separate sessions) 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. 2.6)

Table A.8: Comparing the test-retest reliability and pre-processing sensitivity of the MINDy connectivity parameter 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. 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)
Motion	.8096 (.0268)	.7909 (.0619)	3.1507*	5.3219**
CompCor	.8039 (.0185)	.7481 (.0504)	10.5096**	7.378**
GSR	.8021 (.0180)	.7571 (.0523)	7.8834**	8.4200**
M vs. C	.9323 (.0178)	.8512 (.0582)	13.6332**	10.7247**
M vs. G	.9116 (.0191)	.7910 (.0776)	14.1496**	16.4823**
C vs. G	.9736 (.0037)	.9669 (.0143)	4.2162**	14.6271**

Table A.9: MINDy performance in inverting the weight matrix and its asymmetries in cases of model mismatch. Ground truth models were either a tanh rate-model downsampled to the fMRI TR, a rate model with spatially heterogeneous hemodynamic response functions, or a neural mass model using the nonlinear Balloon-Windkessel model of hemodynamics.

Model	Full W	$W - W^T$	N
Rate-Model	0.949 (0.009)	0.971 (0.007)	1700
Rate+spatial HRF	0.754 (0.026)	0.770 (0.027)	1680
Neural Mass + B-W	0.642 (0.032)	0.567 (0.036)	1480

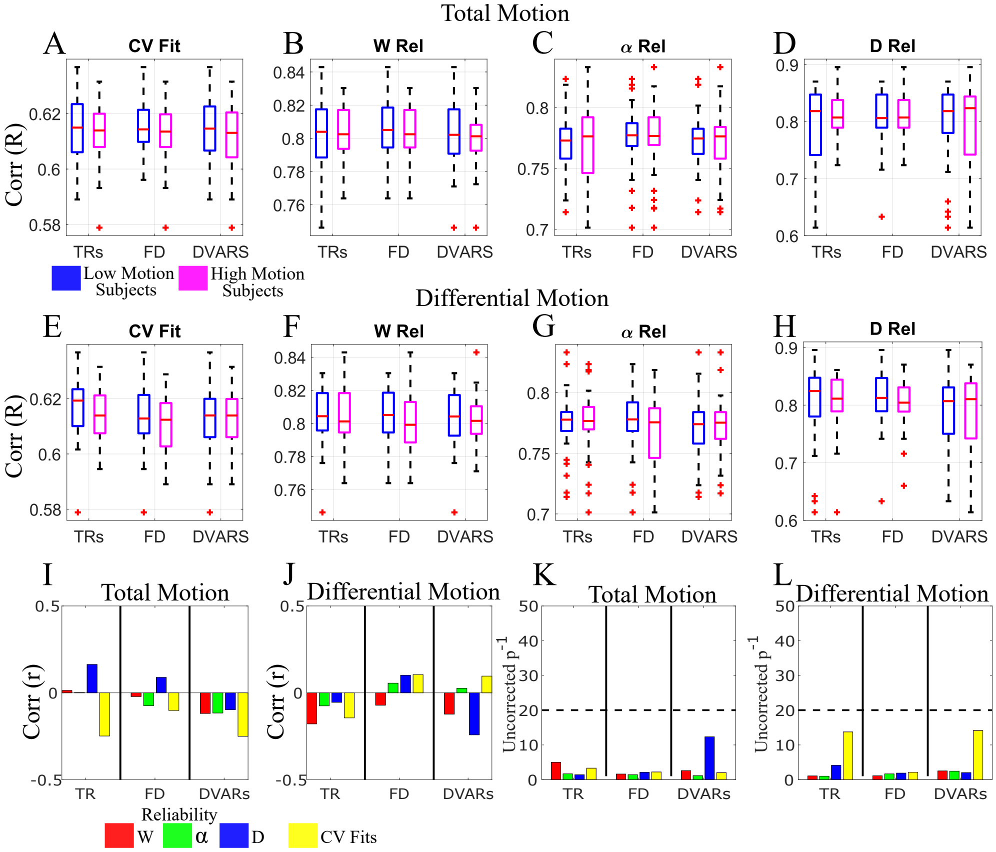

Figure A.12: After pre-processing, MINDy fits are robust to motion. Fitting performance was measured by the cross-validated goodness of fit (A,E) and the reliability for each parameter (B-D,F-H). Individual differences in motion were quantified by either membership in median-split high vs. low motion 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).

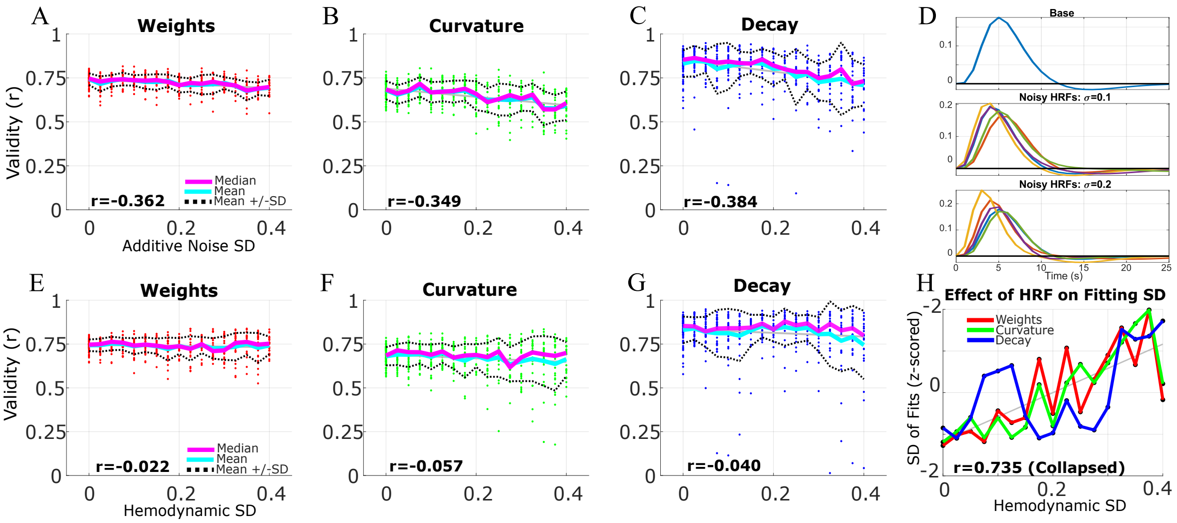

Figure A.13: Sensitivity to various forms of noise in data (Same as Fig. 2.7 A,B, but with 17 levels of noise). A-C) MINDy estimates all parameters of ground-truth models accurately even in the presence of additive measurement noise. D) Examples of how increasing the variability of hemodynamic parameters changes the shape of randomly drawn hemodynamic response functions (HRF). E-G) Hemodynamic variability does not alter the mean performance of MINDy estimates. H) Hemodynamic variability decreases the consistency of MINDy performance (more variable correlations with ground truth).

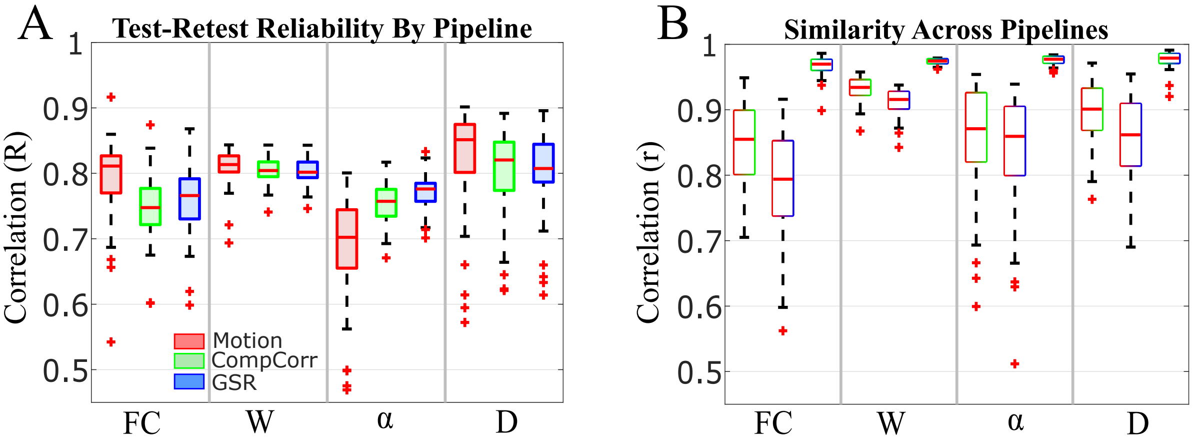

Figure A.14: MINDy estimates are robust to secondary preprocessing choices: Motion-only (Red), Motion+CompCor (Green), Motion+CompCor+GSR (Red). A) Test-retest reliability of rsFC and the MINDy weights (W), curvature (α) and Decay (D) parameters by secondary preprocessing pipeline. B) Correlation between parameter estimates for different preprocessing pipelines applied to the same scanning data (single session). Color of the bar indicates which pipelines are being compared: R+G indicates similarity between Motion-only and CompCor estimates, R+B indicates Motion-only vs. GSR, G+B indicates CompCor vs. GSR (results are displayed in this order left to right within each column)

A.12 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 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 $\xi = max_x(\psi_\alpha(x))$. 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 Table A.10 and then calculating gradients as in Table A.11. In all cases MINDy was run for 5000 iterations with batch size 300 (300 time-points used in each iteration).

Table A.10: Definition of variables ("Name") used in MINDy gradient calculations and their interpretation ("Meaning"). The term n_{Batch} denotes the number of samples (time-points) included in each minibatch (the training data for a given NADAM iteration).

Name	Equation	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 Function (ψ)
D	$D_2^2 + D_{min}$	Full Decay
R	$dX - W_K \psi(x) + Dx$	Residual Error
Q	$W_K^T R$	$2\frac{\partial R}{\partial \psi}$
Z	$R\psi(x)^T / n_{Batch}$	$E_T[\frac{\partial R}{\partial W}]$
Y	$Z - \lambda_4 W_L$	$\frac{\partial J}{\partial W_L}$

Table A.11: Equations used to efficiently calculate MINDy parameter gradients. These equations leverage the additional variables defined in SI Tab. A.10. Note that the decay parameter is updated in terms of its square-root (D_2) and the curvature parameter is updated in terms of the linearized form ξ.

Parameter	Negative Error Gradient ($-\frac{\partial J}{\partial \omega}$)
W_S	$Z - \lambda_1 sgn(W_S) - \lambda_4 diag(sgn(W_S))$
W_1	$Y W_2^T - \lambda_3 sgn(W_1)$
W_2	$W_1 Y - \lambda_3 sgn(W_2)$
ξ	$-\xi^{-3} b E_T[Q \circ (1/P_1 - 1/P_2)]$
D_2	$-2D_2 \circ E_T[R \circ X]$

Table A.12: NADAM hyperparametes for each MINDy parameter and the distributions used to initialize each parameter. NADAM hyperparameters consist of the update rate ("Rate"), decay rate of gradients (μ), decay rate of squared-gradients (ν), and regularization term ε.

	Rate ($\eta \times 10^5$)	μ	ν	ε	Initialization
W_S	2.5	.9	.95	.15	N(0, .01)
$W_{1,2}$	6.25	.9	.95	.15	N(0, .01)
α	12.5	.9	.95	.2	$.1 + \sqrt{.25 + \lvert N(0,\ .25)\rvert + b^2/4}$
D_2	1750	.9	.95	200	$1.75 + \sqrt{\lvert N(0,\ .25)\rvert}$

Table A.13: Chosen values for other hyperparameters used in MINDy. These (non-NADAM) hyperparameters consist of the four regularization terms (λ_i) in the cost function (Eq. 2.4), the minimum allowable value for D (D_{min}), and the scaling factor of the transfer-function (b)

Variable	Value
λ_1	.075
λ_2	.2
λ_3	.05
λ_4	.05
D_{min}	.1
b	20/3

Appendix B

Supplemental Information for Surrogate Deconvolution [4]

B.1 "Local" Field Potential Simulations

The "local" field potential recordings from Sec. 3.4.1 were simulated using the discrete-time neural mass model:

$$x_{t+1} = a\zeta(by_t) + \frac{x_t}{\tau} + c + \epsilon_t \tag{B.1}$$

$$y_{t+1} = S \circ y_t + Wx_t + \omega_t \tag{B.2}$$

$$z_t = y_t + \nu_t \tag{B.3}$$

The paramaters $a = 3$, $b = 1/5$, and $\tau = 2$ were fixed. For each simulation, the remaining neural-mass model parameters were redrawn from fixed, independent distributions: $c \in \mathbb{R}^{n_{pop}} \sim \mathcal{N}(-1, .25^2)$ and $S \sim \mathcal{N}(.5, .2^2) \cap [.2, .8]$. The connectivity parameter W was sampled using a two-step procedure:

---

[4] Appendix chapter reprinted verbatim from previously published work: [99]

$$W_0 \sim \mathcal{N}(0, .1^2) \quad W = W_0 + 2(W_0 - W_0^T). \tag{B.4}$$

This exaggerated asymmetry serves to ensure solutions have nontrivial dynamics in the absence of noise. The noise processes ϵ_t, ω_t, and ν_t were all independent, white Gaussian processes with the same variance for each population. For each simulation the standard deviations of ϵ_t, ω_t, and ν_t were drawn from $.25 + .5|\mathcal{N}(0,1)|$, $.05 + .1|\mathcal{N}(0,1)|$, and $.1 + .2|\mathcal{N}(0,1)|$, respectively. The variances assumed by Kalman Filtering were .5, .1, and .2 for ϵ_t, ω_t, and ν_t, respectively.

B.2 Randomized Networks and MINDy Hyperparameters for simulated fMRI

Ground-truth simulations for BOLD fMRI (Sec. 3.4.4) were produced by a 40 node Hopfield-type ([1]) recurrent neural network with asymmetric connectivity:

$$x_{t+\Delta t} = W[tanh(b_0 \circ x_t)]\Delta t + (1 - \Delta t)Dx + \epsilon_t. \tag{B.5}$$

Here, the timescale of integration was Δt=.1s and measurement occurred every $700ms$. The process noise ϵ_t was Gaussian ($\sigma^2 = .625$) and independent between channels. The simulation parameters and generic MINDy fitting hyperparameters were generally identical to those in the original 40-network MINDy simulations ([8]). Ground-truth connectivity parameters (W) for the simulations were generated by a hyperdistribution characterized by four hyperparameters which scale the reduced-rank magnitude (σ_1), sparseness (σ_2), degree of asymmetry (σ_a), and degree of population clustering ($\hat{p}$). These hyperparameters are distributed $\sigma_1, \sigma_a \sim \mathcal{N}(4, .1^2)$ and $\sigma_2 \sim \mathcal{N}(3, .1^2)$. The hyperparameter $\hat{p}$ is either 1 or 2 with equal probability. These parameters were used to generate three matrices (M_1, M_2, M_3) distributed as follows:

$$M_1 \sim [\mathcal{N}(0, 1/\sigma_1^2) + \mathcal{N}(0, 1/\sigma_1^2)^3]_{40/\hat{p} \times 40/\hat{p}} \tag{B.6}$$

$$M_2 \sim [\mathcal{N}(0, 1/\sigma_2^2)^3]_{40\times 40} \tag{B.7}$$

$$\begin{split} M_3 \sim [\mathcal{N}(0, 1/\sigma_1^2) + \mathcal{N}(0, 1/\sigma_1^2)^3]_{40\times 5} \times ... \\ [\mathcal{N}(0, 1/\sigma_1^2) + \mathcal{N}(0, 1/\sigma_1^2)^3]_{40\times 5} \end{split} \tag{B.8}$$

To generate population clustering we use the ones matrix $1_{\hat{p}\times\hat{p}}$ and define $\hat{M}_1 := 1_{\hat{p}\times\hat{p}} \otimes M_1$ in which $\otimes$ denotes the Kronecker product. The final connectivity matrix (W) for each simulation is formed as follows:

$$Q := \hat{M}_1 + M_2 + M_3 \qquad W = Q + (Q - Q^T)/\sigma_1. \tag{B.9}$$

The slope vector $b_0 \in \mathbb{R}^{40}$ is distributed $b_0 \sim \mathcal{N}(6, (.5)^2)$ and the diagonal decay matrix D has (diagonal) elements iid. distributed $D_{i,i} \sim \mathcal{N}(.4, .1^2) \cap [.2, \infty]$. Deconvolved time series were z-scored. Base MINDy regularization parameters for the 40-node simulation were generated by rescaling the empirical fMRI regularization parameters $(\hat{\lambda}_1 = .075, \hat{\lambda}_2 = .2, \hat{\lambda}_3 = .05, \hat{\lambda}_4 = .05)$ by $1/r_n, 1/r_n, 1/\sqrt{r_n}$, and $1/r_n^2$, respectively with $r_n = 10$ is the approximate ratio between the number of empirical brain regions (419) and those used in the simulation (40) ([8]) which used the method described below (Sec. B.3). The maximum-rank of the low rank component W_L was 15. Initial values for MINDy parameters were distributed as in ([8]). The NADAM update rates for the HRF parameters α and β were 5×10^{-4} and 2.5×10^{-4}, respectively for the

40-node simulation. Surrogate deconvolution used the third-order bivariate polynomial basis $\{1, \alpha, \beta, \alpha^2, \alpha\beta, \beta^2, \alpha^3, \alpha^2\beta, \alpha\beta^2, \beta^3\}$ which was fit to the z-scored deconvolution surfaces.

B.3 Empirical Selection of MINDy Hyperparameters

The MINDy hyperparameters we used were previously determined ([8]) by pseudo-optimization of empirical Human Connectome Project ([65]) fMRI data. In the former study, values were chosen to maximize cross-validated goodness of fit while retaining a test-retest correlation (reliability) of at least .7 for each type of estimated parameter (W, α, D). In brief, values were sampled from a grid over the 4-dimensional space and used to fit models to a set of 10 left-out subjects with test-retest data (none of these subjects were reused in our analyses). The gridded fits determined the likely vicinity of local minima and the final values were chosen based upon iterated coordinate-descent with a fixed resolution (.005). More sophisticated approaches for hyperparameter selection also exist ([169]) and may be more efficient in future applications.

B.4 HCP Data for Surrogate-HRF MINDy

For the empirical data, MINDy used the original regularization parameters $(\hat{\lambda}_i)$. NADAM update rates were 2.5×10^{-4} for α and 2.5×10^{-5} for β. Resting-state fMRI from the Human Connectome Project (HCP; [65]) was preprocessed according to Siegel and colleagues ([2]) and smoothed via nearest-neighbor. Deconvolution was performed using Wiener's method with noise-signal-ratio =.002. On each minibatch, next-step predictions were made for 250 sequential frames using an HRF kernel length of 30 TRs (21.6s) and parameter updates were performed using NADAM for 6000 minibatches. As before, surrogate deconvolution used the third-order bivariate polynomial basis $\{1, \alpha, \beta, \alpha^2, \alpha\beta, \beta^2, \alpha^3, \alpha^2\beta, \alpha\beta^2, \beta^3\}$ to approximate the z-scored deconvolved time-series. For fitting surrogate coefficients, α was assumed uniform on $[5, 7]$ and β was assumed uniform on $[.5, 1.5]$. Expected values were taken by sampling this two-dimensional space along an evenly-spaced 10×10 rectangular grid.

B.5 Derivation of Kernels from Partially-Observable State-Space Models

Convolutional representation can reduce differential/difference equation models of large, hierarchical systems into much smaller (integro-differential) forms. These systems are hierarchical in the sense that they contain a small set of nonlocal (potentially nonlinear) state-variables ($x_t \in \mathbb{R}^n$) with an equal number of recording channels ($z_t \in \mathbb{R}^n$). Each of these interconnected state-variables ($x_t^{(i)} \in \mathbb{R}$), however, can have several coupled local state-variables which produce linear intrinsic dynamics ($y^{(i)} \in \mathbb{R}^{k_i}$). In neuroscience applications, this scenario typically corresponds to one channel per brain area with each area defined by multiple state-variables (e.g. physiological mechanisms):

$$x_{t+1} = f(x_t, y_t) + \eta_t, \tag{B.10}$$

$$y_{t+1}^{(i)} = A_i y_{t+1}^{(i)} + b_i x_t^{(i)}. \tag{B.11}$$

Thus, the state-variables $y^{(i)}$ evolve according to the matrix $A_{k_i \times k_i}$. We assume that the A matrix is stable in the discrete-time sense (eigenvalues have absolute values strictly less than one) which prevents "exploding" solutions and guarantees the existence of an equivalent convolutional form. We note that the local state-variables ($y^{(i)}$) do not need to be the same size for each $x^{(i)}$ (e.g. brain area) and they are only defined to be local in terms of input: $y^{(i)}$ can directly influence $x^{(j \neq i)}$, but not vice-versa. The measurement from each channel $z_t^{(i)}$ is a noisy linear summation of $k_i + 1$ state-variables: $x_t^{(i)} \in \mathbb{R}$ and $y_t^{(i)} \in \mathbb{R}^{k_i+1}$. Thus, at each instance n channels measure a system with $n + \sum k_i$ (partially) coupled state-variables.

$$z_{t+1}^{(i)} = c_i^T y_{t+1}^{(i)} + a_i x_t^{(i)} + \nu_{t+1}. \tag{B.12}$$

Due to the linear intrinsic dynamics of $y_t^{(i)}$, measurements can be re-written in convolutional form:

$$z_t^{(i)} = [h_i * x^{(i)}]_t + \nu_t \tag{B.13}$$

$$h_i = [a_i \quad c_i^T b_i \quad c_i^T A_i b_i \quad c_i^T A_i^2 b_i ...]. \tag{B.14}$$

The unknown kernel parameters can factor into any of the local terms (b_i, A_i, c_i, or a_i). When little is known regarding these parameters a-priori, the mapping from state-space parameters onto the kernel (h_i) is not always invertible, so there are cases in which the parameterization problem is well-posed in convolutional form but not in state-space form (e.g. if A_i is symmetric and both c_i, b_i are unknown). Analogous results hold for the continuous-time case:

$$\dot{y}_i(t) = A y_i(t) + b_i x_i(t) \tag{B.15}$$

$$h_i(\tau) = a_i \delta(\tau) + c_i^T e^{A_i \tau} b_i \tag{B.16}$$

with A now required to be Hurwitz-stable (all eigenvalues have negative real-part), δ denoting the Dirac function, and $e^{A_i \tau}$ denoting the matrix-exponential of A_i multiplied by the lag τ.