Data-driven Modeling of Multi-Region Communication Dynamics using Gated Neural ODEs

Alexander Ladd
Department of Neurobiology & Biophysics
University of Washington
aladd38@uw.edu

Abstract

Advances in multi-region electrophysiology have enabled neuroscientists to collect state-of-the-art datasets comprised of observations of neural dynamics across brain regions. These resources have the potential to help us better understand the dynamics of how populations of neurons communicate with each other during visual decision making. To this end, we introduce Multi-Region Gated Neural ODEs (MR-gnODE), extending gated neural ODEs to model communication between brain regions through separate gated ODE modules. MR-gnODE aims to leverage new multi-region data to accurately identify communication dynamics while maintaining interpretability. We validate this model by recovering communication dynamics between coupled recurrent neural nets (RNNs) and then demonstrate its utility in modeling multi-region communication in the landmark IBL dataset. This work emphasizes the importance of data-driven discovery of brain-wide communication dynamics from emerging large-scale neural datasets.

1 Introduction

Large-scale neural recording technologies now enable simultaneous observation of neural population activity across multiple brain regions, revealing highly distributed computations underlying sensory, cognitive, and motor processes [20, 15]. While computational modeling has driven significant progress in interpreting the dynamics of single populations of neurons [2, 14, 17, 16], identifying how multiple populations coordinate their dynamics remains a substantial challenge in modeling the brain-wide dynamics of behavior. Fortunately, new datasets, such as the International Brain Lab (IBL) dataset, comprising over 547 neuropixels probe insertions spanning more than 240 mouse brain regions [12], introduce new opportunities to better understand the dynamics of multi-region communication.

Despite unprecedented access to multi-region datasets, identifying communication dynamics between populations of neurons remains an outstanding challenge. Communication signals are not directly observable, models must account for unrecorded inputs and complex local dynamics, and accurate data reconstruction does not guarantee correct inference of the underlying communication patterns [13]. Several modeling approaches described in Section 1.1 have made notable progress in accurately identifying communication dynamics. But, just as neural populations display rich dynamical structure, interactions between populations may also display complex dynamical coupling. Models of communication dynamics should not only accurately infer communication dynamics, but also do so in an expressive and interpretable manner.

To this end, neural ordinary differential equations (nODEs) [1] are well-suited for data-driven discovery of communication dynamics. Recent work has demonstrated that gated neural ODEs (gnODEs) offer adaptive timescales and expressive dynamical structure [10]. To date, nODEs have only been applied to single neural populations [8]. We introduce **Multi-Region Gated Neural ODEs (MR-gnODE)**, a novel framework that extends gated neural ODEs to model communication dynamics

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Data on the Brain & Mind.

between multiple brain regions. Our approach explicitly models both within-region dynamics and between-region communication channels using separate gated ODE modules with region-specific and communication-specific timescales. Our main contributions are: (1) We formulate MR-gnODE for data-driven modeling of multi-region neural dynamics with explicit communication channels; (2) we validate that MR-gnODE can accurately capture both regional dynamics of evidence accumulation and motor control in recurrent neural networks (RNNs) as well as communication between such RNNs. (3) We apply MR-gnODE to identify communication dynamics between primary and secondary motor cortex using the IBL dataset. This approach is well-suited for identifying and interpreting rich low-dimensional dynamical structure in neural communication.

1.1 Background and Related Work

Neural computation models for inter-regional communication face fundamental tradeoffs between capturing temporal dynamics and maintaining interpretability. Static methods like reduced-rank regression [19] identify instantaneous communication subspaces through low-rank linear functions but cannot capture temporal dependencies or nonlinear relationships. Dynamical system approaches express temporal dependencies between populations as either linear [4] or nonlinear [7] switching dynamical systems. We will use the multi-population recurrent switching linear dynamical system (mp-rSLDS) described in [4] as a benchmark in our experiments. Other methods model communication dynamics as impulse response functions (IRF) [3], Gaussian processes [5], or RNNs [18]. MR-LFADS [13] uses a sequential VAE with coupled GRUs to jointly infer communication, unobserved inputs, and local dynamics, improving identifiability by constraining communication to reconstructed firing rates rather than flexible latents.

Neural ODEs model dynamics as $\dot{\mathbf{z}}_t = F_\theta(\mathbf{z}_t, \mathbf{x}_t)$ [1], where F_θ is a neural network, $\mathbf{z}_t$ are latent factors and $\mathbf{x}_t$ are observations. $\dot{\mathbf{z}}_t$ is a velocity field for latent variables $\mathbf{z}$ at time t. The "gated" neural ODE extension (Fig. 1a) provides adaptive timescales through $\tau\dot{\mathbf{z}}_t = G_\theta(\mathbf{z}_t, \mathbf{x}_t) \odot [-\mathbf{z}_t + F_\theta(\mathbf{z}_t, \mathbf{x}_t)]$ [10]. The added gating function G_θ can modulate the rate of change in latent dynamics based on both $\mathbf{x}_t$ and $\mathbf{z}_t$. This allows the multi-region extension of the gnODE (MR-gnODE) described in Section 2 to learn smooth, time-varying, nonlinear dynamics rather than linear [3] or discrete nonlinear dynamics (rSDLS) [4, 7]. Unlike all previous approaches described above, which identify dynamics through discrete changes in latent state ($\mathbf{z_{t-1}} \rightarrow \mathbf{z_t}$), MR-gnODE identifies dynamics by inferring a velocity field (flow field) conditioned on $\mathbf{z}_t$ and $\mathbf{x}_t$. This benefits interpretability, as F_θ directly expresses the attractor structure [9] rather than requiring the sampling of trajectories to identify fixed points in communication dynamics, as is necessary with RNNs [13]. Though many studies identify low-rank structure in RNNs, it's Kim et al. [10] showed that high-dimensional GRUs did not always favor low-dimensional solutions to low-dimensional tasks while gnODEs did - making them a more ideal choice for accurately inferring low-dimensional dynamics.

2 Multi-Region Gated Neural ODE (MR-gnODE)

We now present our Multi-Region Gated Neural ODE (MR-gnODE) framework, which extends gated neural ODEs (gnODE) to explicitly model communication dynamics between brain regions. Our approach decomposes multi-region systems into three key components: within-region gnODEs, between-region communication gnODEs, and region-specific readouts.

2.1 Model formulation

Consider a system with R regions, where each region r has latent state $\mathbf{z}_t^r \in \mathbb{R}^{d_r}$ and communication channels $\mathbf{m}_t^{(s\rightarrow r)} \in \mathbb{R}^{d_c}$ capture information flow from region s to region r. The MR-gnODE **region dynamics** shown in Figure 1b are given by:

$$\tau_r \dot{\mathbf{z}}_t^r = G_\theta^r \left(\mathbf{z}_t^r + \mathbf{m}_t^{(\rightarrow r)}, \mathbf{x}_t^r\right) \odot \left[-\mathbf{z}_t^r + F_\theta^r \left(\mathbf{z}_t^r + \mathbf{m}_t^{(\rightarrow r)}, \mathbf{x}_t^r\right)\right] \tag{1}$$

where aggregated messages are defined as:

$$\mathbf{m}_t^{(\rightarrow r)} = \sum_{s \neq r} w^{(s\rightarrow r)} \cdot \psi_\theta(\mathbf{m}_t^{(s\rightarrow r)}) \tag{2}$$

and **communication dynamics** are formulated as:

$$\tau_m \dot{\mathbf{m}}_t^{(s\rightarrow r)} = G_\theta^m \left(\mathbf{z}_t^s, \mathbf{m}_t^{(s\rightarrow r)}\right) \odot \left[-\mathbf{m}_t^{(s\rightarrow r)} + F_\theta^m \left(\mathbf{z}_t^s, \mathbf{m}_t^{(s\rightarrow r)}\right)\right] \tag{3}$$

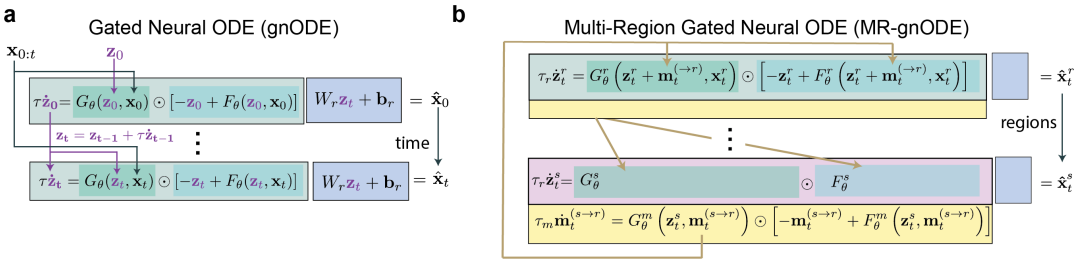

Figure 1: **Gated Neural ODE architecture. (a)** Single-region gated Neural ODE (gnODE) showing the gating mechanism $G_\theta \odot [-z + F_\theta]$, temporal dynamics, and linear readout $W_r \mathbf{z}_t + \mathbf{b}_r$ that maps latent states to observations. **(b)** Multi-Region gated Neural ODE (MR-gnODE) with multiple interacting regions connected through communication channels. Each region maintains its own latent dynamics while receiving aggregated messages from other regions. Note that panel A shows temporal dynamics of gnODE while panel B shows communication dynamics at a single timestep.

F_θ^r and F_θ^m are multi-layer perceptrons (MLPs) defining nODEs for region dynamics $\dot{\mathbf{z}}$ and communication dynamics $\dot{\mathbf{m}}$, respectively. G_θ^r and G_θ^m are MLPs that gate regional and communication dynamics, respectively, through element-wise scaling rate of change in the latent dynamics produced by F_θ^r and F_θ^m. ψ_θ is a neural network that transforms communication channel states into region-dimensional space, and $w^{(s \to r)}$ are learnable scalar weights that enable asymmetric communication effects between regions. The aggregated message $\mathbf{m}_t^{(\to r)}$ is computed by summing all transformed incoming communication channels before adding to the region state. To model **activity readouts**, we reconstruct the observable data for each region using:

$$\hat{\mathbf{x}}_t^r = W_r \mathbf{z}_t^r + \mathbf{b}_r \tag{4}$$

which is a linear readout that maps the latent state of region r to reconstructed neural activity $\hat{\mathbf{x}}_t^r$. This allows the model to be trained by minimizing the reconstruction error between the model's predictions, $\hat{\mathbf{x}}_t^r$, and the observed neural data, $\mathbf{x}_t^r$. As represented in Figure 1b, the latent state of a single region only evolves according to inputs $\mathbf{x}_t$, its own latent state $\mathbf{z_t}^r$, and the aggregated incoming messages from other regions $\mathbf{m}_t^{(\to r)}$.

Training and implementation: We train MR-gnODE using Euler integration for the forward pass with time step $dt = 0.01$, updating states as $\mathbf{z}_{t+dt} = \mathbf{z}_t + dt \cdot \dot{\mathbf{z}}_t$ and $\mathbf{m}_{t+dt} = \mathbf{m}_t + dt \cdot \dot{\mathbf{m}}_t$ (Fig 1b). The total loss combines reconstruction accuracy with communication regularization:

$$\mathcal{L} = \sum_{r=1}^{R} \sum_{t=1}^{T} \|\mathbf{x}_t^r - \hat{\mathbf{x}}_t^r\|^2 + \lambda_{\text{comm}} \sum_{t=1}^{T} \|\mathbf{m}_t\| \tag{5}$$

where $\hat{\mathbf{x}}_t^r = W^r \mathbf{z}_t^r$ are region-specific linear readouts, $\mathbf{x}_t^r$ are target outputs, and λ_{comm} penalizes excessive communication to prioritize learning internal dynamics. Model parameters include velocity field networks F^r and F^m, gating networks G^r and G^m, communication transformation ψ, asymmetric weights $w^{(s \to r)}$, and readout matrices W^r. Parameters are updated using backpropagation through time (BPTT).

3 Experiments

3.1 Decision-Motor Task

We first evaluated MR-gnODE on a synthetic decision-motor task designed to test inter-regional communication. The task consists of a multi-region RNN - an evidence integration RNN, and a motor control region RNN. During training, the evidence region receives noisy sensory input for 15 time steps and must integrate this information to guide the motor region's output velocity for a leftward or rightward reach over the remaining 15 time steps of the trials (Fig. 2a). The evidence region is trained to produce the correct integration of the noisy signal. The motor region is trained to produce the correct velocities while only receiving communication input from the evidence region. The resulting evidence RNN learns low-dimensional decision integration dynamics as represented in

Figure 2b, where the PCA transformation of the evidence RNN hidden state diverges for left evidence and right evidence trials. Figure 2b also demonstrates that motor RNN learns rotational dynamics [2] to produce sinusoidal velocity profiles for left/right reaches. The specific objective is described in Appendix 6.2.

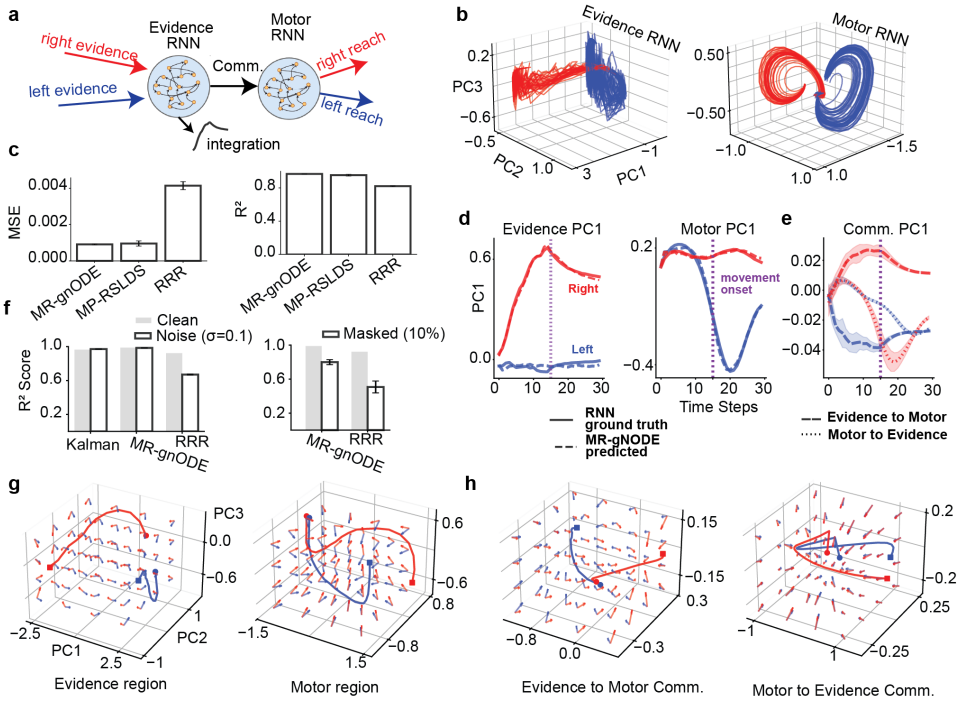

Figure 2: **Decision-motor task validation and model performance.** **(a)** Task architecture with evidence integration and motor control regions connected by communication channels. **(b)** Ground truth RNN dynamics showing diverging trajectories for left/right evidence trials in evidence region and rotational dynamics in motor region. **(c)** Model reconstruction performance comparing MR-gnODE against baselines. **(d)** MR-gnODE predictions vs ground truth trajectories for left/right trials. **(e)** Communication analysis showing evidence-motor communication diverges during accumulation while motor-evidence communication diverges near execution. **(f)** Robustness under noisy observations (left) and partial observations (right). **(g-h)** MR-gnODE generated flow fields for regional and communication dynamics and mean trajectories for held-out left/right trials. MR-gnODE produces interpretable integration and rotational region dynamics (g) and integration communication dynamics (h)

Next, we trained several multi-region models to reconstruct the observed hidden state data from the Evidence-Motor RNN data-generating model. Figure 2c demonstrates that MR-gnODE and mp-rSDLS [4] both learn to nearly perfectly reconstruct the neural population dynamics of both regions, achieving 0.97 and 0.95 R^2 scores on test data. This is verified in Figure 2d, where the first PC of MR-gnODE mean predictions for left and right trials match the first PC of ground truth RNN data for these trials. For left/right conditions, MR-gnODE evidence accumulation (evidence PC1) diverges during the integration period, while PC1 of the motor region does not diverge until just before execution. The communication analysis in (Fig. 2)e demonstrates that MR-gnODE successfully recovers both the directionality of evidence to motor communication but also the informational content as the first PC of evidence-motor communication diverges as evidence is accumulated for left and right trials. The first PC of motor-evidence communication does not diverge until just before execution.

Visualizing the first PC or evaluating models based on R^2 and mean squared error (MSE) in reconstruction tasks can be misleading, as a model that learns the identity transform could achieve perfect reconstruction on training and testing data. To investigate whether MR-gnODE is truly recovering neural dynamics, we create two challenging settings for evaluation: noisy observations and partial observations. In noisy observations, the input neural data $\mathbf{x}_t$ is corrupted as $\mathbf{x}_t + \mathcal{N}(0, 0.1)$ but the

model must still predict $\mathbf{x}_t$. Though we could not evaluate mp-rSLDS under this setting due to library constraints (Appendix 6.1), we added a Kalman filter baseline which explicitly models Gaussian noise, making it a competitive baseline. Figure 2 demonstrates that while Reduced Rank Regression (RRR) performance is significantly hindered in this setting, both Kalman Filter and MR-gnODE maintain near perfect reconstruction of uncorrupted neural data. The right panel of Figure 2f demonstrates model performance where the model must predict the full $\mathbf{x}_t$ while being trained on inputs with 10% of neurons removed. Among 5 cross-validated folds with different held-out neurons and different train/test data, MR-gnODE successfully recovers the intrinsic dynamics of both regions, while RRR fails to do so. The loss function and weight updates for MR-gnODE did not account for estimation error in the held-out neurons, so a mapping from held-in to held-out neurons could not be learned during training.

A unique feature of MR-gnODE is that it models region and communication dynamics as ODEs, so we can also test the model recovered the true Evidence-Motor RNN dynamics by investigating gradients ($\dot{\mathbf{z}}$) of the latent state $\mathbf{z}$ during trials. Figure 2g-h shows PCA-projected flow fields for the evidence region of the trained MR-gnODE model, the motor region, and bidirectional communication channels. For each 3D visualization, we compute $\dot{\mathbf{z}}$ at grid points using example trial inputs, with non-visualized dimensions fixed to mean values from left (blue) versus right (red) decision trials (see Section 6.4 for implementation details). The evidence region exhibits integrator-like flow toward decision-specific attractors. The motor region exhibits partial rotational dynamics, with opposing flow fields for leftward versus rightward reach conditions. Communication channels, $\mathbf{m}$ and $\dot{\mathbf{m}}$, show context-dependent flow patterns: evidence-to-motor channels exhibit divergent flows based on decision context, while motor-to-evidence channels remain largely quiescent except near movement states. These conditional flow projections demonstrate that MR-gnODE recovers the dynamical structure of evidence integration, communication of evidence to the motor region, and the motor control dynamics. Notably, this result differs from traditional flow-field analysis, which is not normally dependent on inputs $\mathbf{x_t}$ [10, 11]. We emphasize the importance of inferring communication flow without requiring inputs or dimensions fixed to mean values for future work.

3.2 IBL Motor Control Task

To demonstrate the utility of MR-gnODE in inferring neural dynamics from data, we analyzed multi-region mouse motor cortex data from the International Brain Laboratory (IBL) using MR-gnODE . We used a session with simultaneously recorded neural populations from primary motor cortex (MOp, 162 neurons) and secondary motor cortex (MOs, 46 neurons) during the IBL visual decision task [12]. Neural activity was aligned to movement onset (-200ms to $+300$ms) and binned at 50ms resolution. The MR-gnODE model was configured with region-specific hidden dimensions (MOs -32 units, MOp - 64 units) and trained to reconstruct the population dynamics while inferring inter-regional communication through the learned coupling parameters.

Figure 3a shows that MR-gnODE achieves accurate reconstruction of both MOp and MOs population activity. The observed MOp responses (Fig. 3a,c) display characteristic movement-related modulation patterns that are moderately captured by the model predictions (Fig. 3a,c right), achieving $R^2 = 0.5$ on held-out trials. Characteristic of the known functions of the secondary motor cortex in movement planning and the primary motor cortex in movement initiation and execution, Figure 3b shows that the distance between neural population activity projected into PC space for left and right wheel turn trials diverges sooner in MOs than MOp. The trained MR-gnODE model demonstrates this directional flow of information in Figure 3d, where the norm of average communication channel activity from MOs to MOp peaks during movement preparation and early execution, while MOp feedback to MOs gradually increases as movements become more prominent. Figure 3e shows some, albeit non-significant, divergence in MOs to MOp left turn and right communication dynamics leading up the movement. This divergence does not appear in MOp to MOs feedback until the wheel turn is executed. Despite only a minimal difference in average communication channel activity for left and rightward movements, analysis of the movement-conditioned Jacobian of communication channel states (Fig. 3f) provides a clearer picture of diverging neural states leading up to left/right movements. For each communication channel, we compute $\mathbf{J}^{(1\to2)} = \frac{\partial \dot{\mathbf{m}}^{(1\to2)}}{\partial \mathbf{m}^{(1\to2)}}$ and $\mathbf{J}^{(2\to1)} = \frac{\partial \dot{\mathbf{m}}^{(2\to1)}}{\partial \mathbf{m}^{(2\to1)}}$, where $\mathbf{m}^{(\cdot\to\cdot)}$ represents the communication channel states between regions (with MOp = 1, MOs = 2). By computing separate Jacobians for left and right choice trials (e.g., $\mathbf{J}^{(2\to1)}_{\text{left}}$ and $\mathbf{J}^{(2\to1)}_{\text{right}}$), we measure the choice-dependent angular divergence $\theta = \arccos\left(\frac{\langle \mathbf{J}_{\text{left}}, \mathbf{J}_{\text{right}} \rangle}{|\mathbf{J}_{\text{left}}||\mathbf{J}_{\text{right}}|}\right)$, quantifying how communication

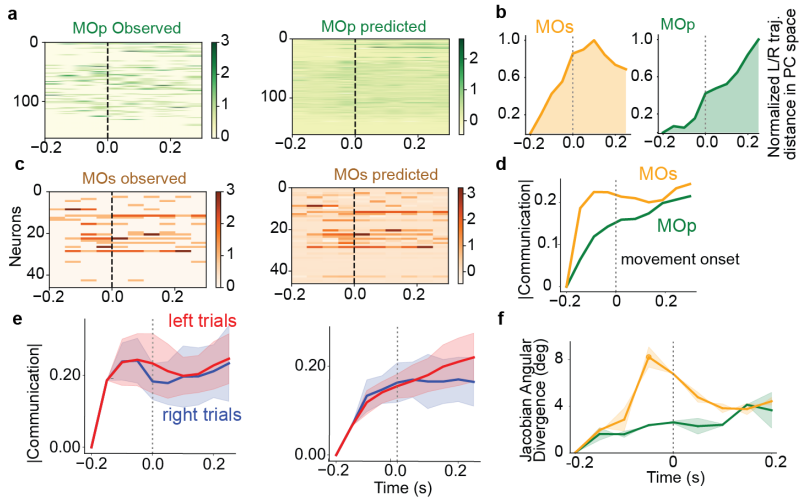

Figure 3: **IBL motor cortex multi-region communication dynamics.** **(a)** Primary motor cortex (MOp) observed population activity aligned to movement onset. **(b)** Distance between PCA transform of neural population activity for left vs right wheel turns. **(c)** Secondary motor cortex (MOs) neural responses and MR-gnODE reconstruction. **(d)** Norm of MR-gnODE MOp/MOs mean communication channel activity. **(e)** Same as d. but mean taken over left and right trials seperately. MOs on left and MOp on right. **(f)** Jacobian angular divergence analysis computing gradients of communication dynamics for left vs right trials and measuring angular separation between condition-specific messaging.

dynamics differentiate based on behavioral choice. Here, MOs to MOp communication channel choice-dependent divergence significantly increases during movement preparation (-100ms to 0ms) while MOp to MOs communication demonstrates little choice-dependent divergence. Thus, we conclude the MR-gnODE model learned communication patterns consistent with the known hierarchical organization of the motor cortex. MOs contribute higher-level control signals that appear during planning and propagates this information to MOp, which participates in initiation and execution. Together, these results demonstrate that MR-gnODE not only accurately models multi-region neural dynamics but also reveals interpretable communication principles from neural data.

4 Discussion

Understanding how distinct brain regions coordinate the neural population dynamics to generate behavior remains a fundamental challenge in neuroscience. Most previous approaches for modeling population dynamics have focused on a single region, but as more multi-region data becomes publicly available, we can start to approach modeling communication dynamics. Here, we introduced MR-gnODE, a multi-region neural ODE framework that simultaneously learns within-region dynamics and inter-regional communication patterns from neural recordings. By directly modeling region and communication velocity fields ($\dot{\mathbf{z}}$, $\dot{\mathbf{m}}$), we used MR-gnODE to visualize inferred phase portraits (Fig. 2g-h) and identify changes in the movement-conditioned Jacobian of communication dynamics in the motor cortex ($\mathbf{J}^{MOs \to MOp}$, Fig. 3f). Previously proposed multi-region approaches like MR-LFADS or mp-rSLDS would require sampling trajectories [6] to visualize attractor structure.

Our experiments demonstrate that MR-gnODE accurately captures both synthetic and biological multi-region dynamics while revealing interpretable communication principles. Yet, much future work is needed to complete and validate the MR-gnODE framework. In future work, we aim to enable the model to receive or infer exogenous inputs u_t and add variational inference for $\mathbf{z}$ and $\mathbf{m}$. Further ablation experiments and improved baselines, such as MR-LFADS [13], are required to validate/contextualize the model's capacity for identifying neural dynamics. We hope MR-gnODE can leverage emerging large datasets of multi-region neural activity for data-driven discovery of rich brain-wide communication dynamics. Regardless, this work aims to emphasize the importance of interpretable modeling of multi-region communication flow dynamics.

5 Acknowledgments

We would like to thank REDACTED for insightful discussion on best practices for gated neural ODEs and flow field modeling.

References

[1] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[2] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012.

[3] Matthew Dowling and Cristina Savin. Nonlinear multiregion neural dynamics with parametric impulse response communication channels. In *The Thirteenth International Conference on Learning Representations*.

[4] Joshua Glaser, Matthew Whiteway, John P Cunningham, Liam Paninski, and Scott Linderman. Recurrent switching dynamical systems models for multiple interacting neural populations. *Advances in neural information processing systems*, 33:14867–14878, 2020.

[5] Evren Gokcen, Anna I Jasper, João D Semedo, Amin Zandvakili, Adam Kohn, Christian K Machens, and Byron M Yu. Disentangling the flow of signals between populations of neurons. *Nature Computational Science*, 2(8):512–525, 2022.

[6] Matthew D Golub and David Sussillo. Fixedpointfinder: A tensorflow toolbox for identifying and characterizing fixed points in recurrent neural networks. *Journal of open source software*, 3(31):1003, 2018.

[7] Orren Karniol-Tambour, David M Zoltowski, E Mika Diamanti, Lucas Pinto, Carlos D Brody, David W Tank, and Jonathan W Pillow. Modeling state-dependent communication between brain regions with switching nonlinear dynamical systems. In *The Twelfth International Conference on Learning Representations*, 2024.

[8] Timothy D Kim, Thomas Z Luo, Jonathan W Pillow, and Carlos D Brody. Inferring latent dynamics underlying neural population activity via neural differential equations. In *International Conference on Machine Learning*, pages 5551–5561. PMLR, 2021.

[9] Timothy Doyeon Kim, Tankut Can, and Kamesh Krishnamurthy. Learning and shaping manifold attractors for computation in gated neural odes. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022.

[10] Timothy Doyeon Kim, Tankut Can, and Kamesh Krishnamurthy. Trainability, expressivity and interpretability in gated neural odes. *arXiv preprint arXiv:2307.06398*, 2023.

[11] Timothy Doyeon Kim, Thomas Zhihao Luo, Tankut Can, Kamesh Krishnamurthy, Jonathan W Pillow, and Carlos D Brody. Flow-field inference from neural data using deep recurrent networks. *bioRxiv*, 2023.

[12] International Brain Laboratory, Brandon Benson, Julius Benson, Daniel Birman, Niccolò Bonacchi, Kcénia Bougrova, Sebastian A Bruijns, Matteo Carandini, Joana A Catarino, Gaelle A Chapuis, et al. A brain-wide map of neural activity during complex behaviour. *Biorxiv*, pages 2023–07, 2023.

[13] Belle Liu, Jacob Sacks, and Matthew D Golub. Accurate identification of communication between multiple interacting neural populations. *arXiv preprint arXiv:2506.19094*, 2025.

[14] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474):78–84, 2013.

[15] Simon Musall, Matthew T Kaufman, Ashley L Juavinett, Steven Gluf, and Anne K Churchland. Single-trial neural dynamics are dominated by richly varied movements. *Nature neuroscience*, 22(10):1677–1686, 2019.

[16] Aditya Nair, Tomomi Karigo, Bin Yang, Surya Ganguli, Mark J Schnitzer, Scott W Linderman, David J Anderson, and Ann Kennedy. An approximate line attractor in the hypothalamus encodes an aggressive state. *Cell*, 186(1):178–193, 2023.

[17] Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815, 2018.

[18] Matthew G Perich, Charlotte Arlt, Sofia Soares, Megan E Young, Clayton P Mosher, Juri Minxha, Eugene Carter, Ueli Rutishauser, Peter H Rudebeck, Christopher D Harvey, et al. Inferring brain-wide interactions using data-constrained recurrent neural network models. *BioRxiv*, pages 2020–12, 2020.

[19] João D Semedo, Amin Zandvakili, Christian K Machens, Byron M Yu, and Adam Kohn. Cortical areas interact through a communication subspace. *Neuron*, 102(1):249–259, 2019.

[20] Nicholas A Steinmetz, Peter Zatka-Haas, Matteo Carandini, and Kenneth D Harris. Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 576(7786):266–273, 2019.

6 Technical Appendices and Supplementary Material

6.1 Code Availability

All code is available at : REDACTED FOR DOUBLE BLIND REVIEW. For mp-rSLDS, we used the SSM library: https://github.com/lindermanlab/ssm.

6.2 Motor Evidence RNN

Model architecture: The multi-region RNN consists of two functionally specialized modules connected via a communication pathway:

Evidence integration RNN: The evidence region is implemented as an LSTM network that processes noisy sensory inputs: $h_t, c_t = \text{LSTM}(x_t, h_{t-1}, c_{t-1})$. $x_t \in \mathbb{R}^2$ represents the bivariate evidence input at time t, and $h_t, c_t \in \mathbb{R}^{64}$ are the hidden and cell states respectively. The evidence integration output is computed as: $y_t = Wh_t + b$. where W and b are learned parameters.

Motor control RNN: The motor region is implemented as a separate LSTM that receives communication signals from the evidence region: $h_t, c_t = \text{LSTM}([s_t^{\text{comm}}; p_{t-1}], h_{t-1}, c_{t-1})$ where the input concatenates the communication signal $s_t^{\text{comm}} \in \mathbb{R}$ with the previous motor position $p_{t-1} \in \mathbb{R}^2$. The motor velocity output is a 2-d vector: $v_t = Wh_t + b$.

Evidence-motor communication: The communication signal from evidence to motor region is computed as:$s_t^{\text{comm}} = W^{\text{comm}} h_t + b^{\text{comm}}$ where W^{comm} transforms the evidence hidden state into a scalar communication signal.

Task data generation: Training data is generated for a decision-motor task with the following structure: During the evidence phase ($t \in [1, T_e]$ where $T_e = 15$ time steps), noisy sensory inputs are generated:

$$x_t^{(i)} = \begin{bmatrix} (1 - d_i) \cdot \kappa + \epsilon_t^{(0)} \\ d_i \cdot \kappa + \epsilon_t^{(1)} \end{bmatrix}$$

where $d_i \in \{0, 1\}$ indicates the target direction (left/right) for trial i, $\kappa = 0.3$ is the coherence level, and $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.7$ represents Gaussian noise.

The ground truth evidence integration is computed as: $I_t = \tanh\left(\sum_{k=1}^{t} (x_k^{(1)} - x_k^{(0)})\right)$

In the motor phase ($t \in [T_e, T]$) Target velocities are generated with confidence modulation: $v_t = d_i \cdot \alpha_i \cdot f(t) + \eta_t$ where:

- $\alpha_i = \text{clip}(|I_{T_e}|/20, 0, 1)$ is the confidence derived from evidence strength
- $f(t) = 0.5\sin(\pi \cdot \frac{t-T_e}{T-T_e})$ defines the velocity profile
- η_t is smooth noise generated from low-frequency sinusoids with random phase

Training procedure: The model is trained using two mean squared error losses:

Evidence integration loss:

$$\mathcal{L}_e = \frac{1}{T}\sum_{t=1}^{T}(y_t - I_t)^2 \tag{6}$$

Motor velocity loss:

$$\mathcal{L}_m = \frac{1}{T}\sum_{t=15}^{T}\|\hat{v}_t - v_t\|_2^2 \tag{7}$$

The total loss is:

$$\mathcal{L} = \mathcal{L}_e + \mathcal{L}_m \tag{8}$$

6.3 Model parameters

The parameters used in all experiments can be found in: GIT REPO FOLDER RETRACTED FOR DOUBLE BLIND REVIEW

6.4 Input Conditioned Flow Field Visualization

To visualize the dynamics of MR-gnODE, we project the high-dimensional flow fields onto three-dimensional subspaces using PCA. We construct separate visualizations for each region $\mathbf{z}^{(r)}$ and communication channel $\mathbf{m}^{(r\to s)}$.

Given trajectories $\{\mathbf{s}_t^{(i)}\}_{i=1}^{N_{\text{trials}}}$ from experimental trials where $\mathbf{s}_t = [\mathbf{z}_t^{(0)}, \mathbf{z}_t^{(1)}, \mathbf{m}_t^{(0\to1)}, \mathbf{m}_t^{(1\to0)}]$, we visualize the dynamics of each region $\mathbf{z}^{(r)}$ as follows. We compute PCA on the concatenated trial data for $\mathbf{z}^{(r)}$ to obtain projection matrix $V_r \in \mathbb{R}^{d_r\times3}$ containing the first three principal components.

To compute the flow field, we create a grid $\mathcal{G}$ in the 3D PCA space spanning the data distribution. At each grid point $\mathbf{p} \in \mathcal{G}$, we:

1. Inverse transform to original space: $\mathbf{z}^{(r)*} = V_r\mathbf{p} + \bar{\mathbf{z}}^{(r)}$
2. Construct full state using mean state $\bar{\mathbf{s}}^{(c)}$ from decision type c with $\mathbf{z}^{(r)}$ replaced by $\mathbf{z}^{(r)*}$
3. Compute MR-gnODE dynamics: $\tau\dot{\mathbf{z}} = G_\theta(\mathbf{z}, \mathbf{x}^{(c)}) \odot [-\mathbf{z} + F_\theta(\mathbf{z}, \mathbf{x}^{(c)})]$
4. Project flow back to PCA space: $\dot{\mathbf{p}} = V_r^T\dot{\mathbf{z}}^{(r)}$

where $\mathbf{x}^{(c)}$ represents example trial inputs for decision type c. Mean trajectories for each decision type are projected as $\mathbf{p}_t^{(c)} = V_r^T(\bar{\mathbf{z}}_t^{(r),c} - \bar{\mathbf{z}}^{(r)})$. The same procedure is applied to visualize communication channels $\mathbf{m}^{(r\to s)}$.