Modeling Conscious–Unconscious Neural Dynamics with Operator Theory

Anonymous submission

Abstract

Awareness emerges from large-scale patterns of neural activity, yet modeling the interplay between conscious and unconscious dynamics remains elusive. Leading techniques for data-driven analysis of neural signals rely on Koopman operator theory (KOT) which lacks mechanisms to disentangle conscious and unconscious modes of awareness. We leverage Mori-Zwanzig formalism to separate unconscious and conscious modes as *resolved* and *unresolved* states, respectively.

Introduction

Artificial intelligence driven analysis of neuroscience enables new insights into neural processing. Advances in neuroimaging, deep learning, and representational similarity analysis (RSA) now enable unprecedented insight into how awareness modulates neural representations across the brain, revealing distinct patterns associated with conscious and unconscious processing(Mei and Soto 2025; Weber et al. 2024). Methods based on Koopman-operator theory (KOT), including dynamic mode decomposition (DMD), model neural activity as evolving dynamical systems, and recover coherent spatiotemporal modes underlying perception and cognition (Brunton, Proctor, and Kutz 2014; Casorso et al. 2019). However, KOT approaches assume Markovian dynamics–the evolution of the system is entirely self-contained–lacking principled mechanisms to separate the perceptively distinct modes of conscious and unconscious awareness. Recent work studying the effects of propofol anesthesia on the loss of consciousness (LOC) provides empirical access to the split between conscious and unconscious dynamics (Eisen et al. 2024; Xiong et al. 2024). These studies examined propofol-induced LOC as a source of dynamical instability in neural activity, using the Hankel alternative view of Koopman (HAVOK) framework (Brunton et al. 2017) to analyze the forcing mechanisms driving such instabilities. In this work, we aim to extend these insights by applying the Mori–Zwanzig (MZ) projection-operator formalism (Mori 1965; Zwanzig 2001), which distinguishes between resolved (conscious) and unresolved (unconscious) components of neural dynamics.

Background

Our approach is based on a hierarchy of operator theories. Closed KOT is purely Markovian and enables the analysis of coherent neural dynamics. KOT incorporates a stochastic forcing term, enabling the study of dynamic instabilities. MZ incorporates separate neural dynamics for resolved and unresolved states, resulting in forcing and dissipative terms.

Koopman Operator Theory (KOT)

The Koopman operator $\mathcal{K}$ is a linear operator describing the temporal dynamics of a neural state given by g. If the dynamics are *closed* then the evolution of g is given by

$$\frac{d}{dt}g(t) = \mathcal{K}g(t). \tag{1}$$

DMD generates a matrix approximation $\boldsymbol{K} \simeq K$ given a history of states $\{\boldsymbol{g}(t_i)\}_{i=1}^{n}$. This is achieved by generating snapshot matrices

$$\boldsymbol{X} = \begin{bmatrix}\boldsymbol{g}(t_1) & \cdots & \boldsymbol{g}(t_{n-1})\end{bmatrix}, \quad \boldsymbol{Y} = \begin{bmatrix}\boldsymbol{g}(t_2) & \cdots & \boldsymbol{g}(t_n)\end{bmatrix},$$

and finding the least squares solution $\boldsymbol{A} = \boldsymbol{Y}\boldsymbol{X}^{+}$, where $\boldsymbol{X}^{+}$ is the Moore–Penrose pseudoinverse. Coherent modes are recovered as the leading DMD eigenvectors.

General Koopman operator theory. If the dynamics are driven by an external force $\eta(t)$ the evolution equation is

$$\frac{d}{dt}g(t) = \mathcal{K}g(t) + \eta(t). \tag{2}$$

From the singular decomposition of the Hankel matrix

$$\boldsymbol{H} = \begin{bmatrix} g(t_1) & g(t_2) & \cdots & g(t_m) \\ g(t_2) & g(t_3) & \cdots & g(t_{m+1}) \\ \vdots & \vdots & \ddots & \vdots \\ g(t_\ell) & g(t_{\ell+1}) & \cdots & g(t_{\ell+m-1}) \end{bmatrix} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\top},$$

HAVOK models the forcing term as the right singular vector of $\boldsymbol{H}$. A linear model is fit on the reduced coordinates $\tilde{\boldsymbol{g}}_{n+1} = \boldsymbol{K}\tilde{\boldsymbol{g}}_n + b\boldsymbol{u}_n$ where the right singular vector $\boldsymbol{u}_n$ is a low-rank approximation of the residual.

Mori–Zwanzig Operator Theory (MZ)

MZ models the residual term explicitly using resolved and unresolved subspaces. In particular, g is decomposed by orthogonal subspaces P and Q so that $g = \hat{g} + \tilde{g}$ where $\hat{g} = Pg$

and $\tilde{g} = Qg$ are *resolved* and *unresolved* components. The evolution of a resolved observable $\hat{g}(t)$ is given by

$$\frac{\partial}{\partial t}\hat{g}(t) = \underbrace{P\mathcal{L}\hat{g}(t)}_{\text{Markov}} + \underbrace{\int_0^t P\mathcal{L}e^{(t-s)Q\mathcal{L}}Q\mathcal{L}\hat{g}(s)\,ds}_{\text{Dissipative Memory}} + \underbrace{P\mathcal{L}e^{tQ\mathcal{L}}Qg(0)}_{\text{Fluctuating Force}}. \tag{3}$$

Mori-Zwanzig Approach to Neuroscience

Our approach utilizes equation 3 to describe how conscious and unconscious signals co-evolve. In particular, the Markov term accounts for direct interactions within the unconscious baseline, the dissipative memory term captures delayed influences of conscious processes (e.g., feedback effects reflecting awareness integration), and the fluctuating force term reflects inherent noise in conscious states.

The MZ operator theory is *closed* by correlating the fluctuating force and dissipating memory operators

$$P\mathcal{L}e^{tQ\mathcal{L}}Q\mathcal{L}\langle Pg, Pg^\top\rangle = \langle P\mathcal{L}e^{tQ\mathcal{L}}Qg, P\mathcal{L}Qg\rangle.$$

This fluctuation-dissipation relation effectively couples the noise and memory terms. **The result is a practical mechanism for modeling conscious influence as signal noise.**

Recent work (Lin et al. 2021) provides data-driven methods for modeling the effects of unresolved variables from snapshots of resolved variables. The evolution of $\hat{\boldsymbol{g}}$ is

$$\hat{\mathbf{g}}(t_{n+1}) = \sum_{k=0}^{\ell} \boldsymbol{K}^{(k)} \cdot \hat{\mathbf{g}}(t_n) + \mathbf{F}\,,$$

where $\boldsymbol{K}^{(k)}$ incorporates the Markov and dissipative memory terms and $\boldsymbol{F}$ incorporates the fluctuating force. We remark on the connection between MZ and the extended DMD via $\boldsymbol{K}^{(0)} = (\boldsymbol{Y}\boldsymbol{X})(\boldsymbol{X}\boldsymbol{X})^{+}$; see (Lin et al. 2021) for details.

Experiments

Ornstein-Uhlenbeck (OU) process In anesthesia modeling, OU processes serve as interpretable drivers of the dynamics of the neural population. This is because the OU process models conscious signals as deviations from the unconscious mean, where there is a strong tendency to return to the mean (Sleigh et al. 2004). Mathematically, the OU process is a continuous-time mean-reverting SDE akin to equation 2

$$\frac{dx_t}{dt} = -\theta x_t + \sigma\eta(t)$$

where $\eta(t)$ is the derivative of the Wiener process dW_t/dt.

Figure 1 illustrates the advantages of using MZ in forecasting the OU process. In particular, we train on a trajectory with 180 equally partitioned time steps, and then perform 20 step roll-out predictions. Moreover, we perform an ablation of the MZ computation with and without the memory term $\ell = 0$. We observe that the forward evolution of the MZ with $\ell = 18$ outperforms the baseline methods. Additionally, the error term remains reduced for time steps on the order of the memory length $\ell = 18$.

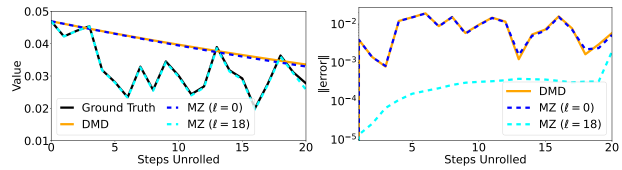

Figure 1: A 20 time step roll-out of the DMD and MZ methods for the OU process, and the respective error. We observe that MZ with memory adheres to the ground truth while the time steps are on the order of the memory length $l = 18$.

Propofol induced loss of consciousness (LOC) Propofol induces loss of consciousness (LOC) by amplifying inhibitory neurotransmission. In awake states, neural feedback loops reduce signal noise so that the system can focus on surprises (Xiong et al. 2024). Under propofol, this feedback mechanism weakens so that sensory signals become louder, but cross area coordination is reduced. MZ treats missing feedback as a learnable memory kernel plus noise, so that we can capture how past context influences new inputs.

Model	Awake	Anesthetized
DMD	3.140	0.620
MZ	0.101	0.011

Table 1: Mean squared error for the DMD and MZ methods on the propofol induced LOC. We observe that MZ achieves significantly improved results over the baseline DMD.

Table 1 reports the results for forecasting 13 unrolled time steps using DMD and MZ with $l = 10$. Training is performed on 728 time steps for the mean sensor data over each of the 96 frequencies. In both awaken and anesthetized states, MZ significantly improves over the baseline.

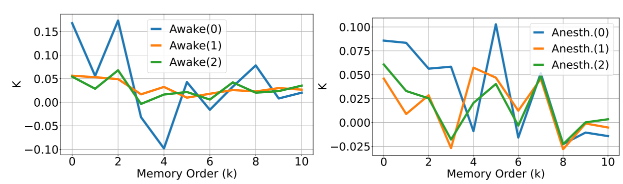

Figure 2: The top three eigen values of the memory operators at each order k. The modes all correlate in the anesthetized state. In the awake state the dominant mode is oscillatory.

Figure 2 illustrates the top three eigenvalues of the MZ memory kernel at each order k. We observe significant differences in the dominant mode of the awake state corresponding to the neural feedback loop. Meanwhile, we observe correlating modes in the anesthetized state.

Conclusion

We show that incorporating a separation of resolved and unresolved variables can boost the model's interpretability of conscious and unconscious modes. Our approach is limited to available data on propofol LOC which separates awake and anesthetic states. Future work aims at analyzing aware and unaware levels of consciousness in multi-modal deep learning approaches (Mei and Soto 2025)

References

Brunton, S. L.; Brunton, B. W.; Proctor, J. L.; Kaiser, E.; and Kutz, J. N. 2017. Chaos as an intermittently forced linear system. *Nature communications*, 8(1): 19.

Brunton, S. L.; Proctor, J. L.; and Kutz, J. N. 2014. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Frontiers in Computational Neuroscience*.

Casorso, J.; et al. 2019. Dynamic Mode Decomposition of resting-state and task fMRI. *NeuroImage*.

Eisen, A. J.; Kozachkov, L.; Bastos, A. M.; Donoghue, J. A.; Mahnke, M. K.; Brincat, S. L.; Chandra, S.; Tauber, J.; Brown, E. N.; Fiete, I. R.; et al. 2024. Propofol anesthesia destabilizes neural dynamics across cortex. *Neuron*, 112(16): 2799–2813.

Lin, Y. T.; Tian, Y.; Livescu, D.; and Anghel, M. 2021. Data-Driven Learning for the Mori–Zwanzig Formalism: A Generalization of the Koopman Learning Framework. *SIAM Journal on Applied Dynamical Systems*, 20(4): 2558–2601.

Mei, N.; and Soto, D. 2025. Brain Representation in Conscious and Unconscious Vision. *Journal of Cognition*.

Mori, H. 1965. Transport, collective motion, and Brownian motion. *Progress of Theoretical Physics*, 33(3): 423–455.

Sleigh, J. W.; Steyn-Ross, D. A.; Steyn-Ross, M. L.; Grant, C.; and Ludbrook, G. 2004. Cortical entropy changes with general anaesthesia: theory and experiment. *Physiological measurement*, 25(4): 921.

Weber, N.; et al. 2024. Correlates of implicit semantic processing revealed by representational similarity analysis. *iScience*.

Xiong, Y.; Donoghue, J. A.; Lundqvist, M.; Mahnke, M.; Major, A. J.; Brown, E. N.; Miller, E. K.; and Bastos, A. M. 2024. Propofol-mediated loss of consciousness disrupts predictive routing and local field phase modulation of neural activity. *Proceedings of the National Academy of Sciences*, 121(42): e2315160121.

Zwanzig, R. 2001. *Nonequilibrium Statistical Mechanics*. Oxford University Press.