
A Latent Neural ODE-VAE for Modeling Hippocampal Population Activity on Low-Dimensional Manifolds

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Neural population activity traces trajectories in a high-dimensional state space, yet
2 accumulating evidence suggests these trajectories are confined to low-dimensional
3 manifolds that encode both task variables and internal state. Existing manifold
4 inference pipelines can recover geometry and explain variability, but often rely on
5 multi-stage local models and do not impose globally smooth continuous-time dy-
6 namics. We develop a latent Neural ODE variational autoencoder (ODE-VAE) that
7 jointly learns (i) a low-dimensional stochastic initial condition, (ii) continuous-time
8 latent dynamics parameterized by a mixture-of-experts ODE, and (iii) a decoder
9 back to neural activity. To better align reconstruction with temporally structured
10 variability, our v5 implementation adds transition-consistency regularization in
11 observation space and a soft locally linear embedding (LLE) constraint in latent
12 space. On synthetic random-foraging sequences, the model achieves high recon-
13 struction accuracy ($R^2 = 0.9789$) while exhibiting seed sensitivity. On the E65
14 hippocampal calcium dataset, we observe moderate reconstruction in archived
15 runs ($R^2 = 0.4368$) and substantial dependence on preprocessing, split strategy,
16 and evaluation space. Together, these results highlight both the promise and cur-
17 rent fragility of end-to-end continuous-time manifold models for noisy biological
18 recordings.

19 1 Introduction

20 Neural activity can be described as a point in a high-dimensional space, where each coordinate
21 axis corresponds to a neuron’s activity. Despite this ambient dimensionality, population trajectories
22 frequently occupy structured, low-dimensional manifolds. Such manifolds have been implicated in
23 motor control [Gallego et al., 2017, Russo et al., 2018], spatial coding in hippocampus [O’Keefe
24 and Dostrovsky, 1971, O’Keefe and Nadel, 1978], and the representation of non-spatial variables
25 [Aronov et al., 2017, Nieh et al., 2021]. These findings motivate a geometric view of hippocampal
26 computation: learned knowledge may be organized as trajectories on manifolds that encode both
27 physical and abstract structure [Tolman, 1948, Stachenfeld et al., 2017, Bellmund et al., 2018].

28 Recent hippocampal manifold analyses estimate intrinsic dimensionalities on the order of 4–6 and
29 show that manifold geometry can capture both task variables and structured trial-to-trial variability
30 [Low et al., 2018, Chaudhuri et al., 2019, Nieh et al., 2021]. However, many approaches are
31 multi-stage: they model transitions locally, define a data-dependent distance, embed points into a low-
32 dimensional coordinate system, and then learn a separate reconstruction mapping [Low et al., 2018,
33 Tenenbaum et al., 2000, Yu et al., 2009]. While effective for geometry, these pipelines do not directly
34 couple representation learning, temporal evolution, and reconstruction in a single continuous-time
35 generative model.

We propose a latent Neural ODE variational autoencoder (ODE-VAE) for trialized population recordings. The model encodes trial onset activity into a stochastic latent initial condition, evolves it via a continuous-time latent ODE, and decodes latent trajectories back to neural activity. Our v5 implementation introduces mixture-of-experts latent dynamics and adds two regularizers inspired by manifold inference: (i) transition consistency in observation space to encourage accurate step-to-step changes, and (ii) a soft LLE constraint to encourage locally linear latent structure without forcing a globally linear embedding. We evaluate this family on synthetic and hippocampal calcium datasets and analyze the sensitivity of performance to preprocessing and evaluation choices.

Contributions.

- We formalize an ODE-VAE for trialized population sequences with mixture-of-experts latent dynamics and explicit geometric regularizers.
- We instantiate this formulation in a versioned codebase (v1-v6) and present v5 as the primary model with transition and soft-LLE regularization.
- We provide a reproducible evaluation on synthetic and hippocampal calcium datasets and identify protocol factors that strongly affect reconstruction metrics.

2 Related Work

Our approach lies at the intersection of manifold-based neuroscience and latent dynamical systems. In hippocampus, the cognitive map framework and subsequent experimental work motivate geometric organization of population codes [O’Keefe and Dostrovsky, 1971, O’Keefe and Nadel, 1978, Eichenbaum and Cohen, 2014], including abstract and non-spatial representations [Constantinescu et al., 2016, Schuck and Niv, 2019, Park et al., 2020, Aronov et al., 2017, Nieh et al., 2021]. Manifold inference from neural dynamics can recover low-dimensional structure and explain structured variability beyond measured task variables [Low et al., 2018, Chaudhuri et al., 2019].

In machine learning, variational autoencoders [Kingma and Welling, 2014] and neural ODEs [Chen et al., 2018] provide a principled framework for continuous-time latent-variable modeling. Latent ODEs extend this idea to irregularly sampled sequences [Rubanova et al., 2019]. We adopt this framework but tailor the encoder, evaluation protocol, and regularization to the neuroscience setting, emphasizing trialized sequences, explicit geometric constraints, and manifold visualizations that facilitate comparison to prior manifold studies.

3 Problem Setup and Data

We study trialized population activity sequences. Let $y_b(t_\ell) \in \mathbb{R}^N$ denote the raw activity of N simultaneously recorded units/ROIs on trial $b \in \{1, \dots, B\}$ at resampled time t_ℓ , where $\ell \in \{1, \dots, L\}$ indexes a fixed-length grid. We write $Y_b \in \mathbb{R}^{L \times N}$ for the stacked sequence.

Observation space. In v5, the model is trained on a PCA-projected representation of activity. Let $x_b(t_\ell) \in \mathbb{R}^K$ be the K -dimensional PCA coordinate at time t_ℓ , and let $X_b \in \mathbb{R}^{L \times K}$ be the corresponding trial sequence. Unless otherwise stated, all losses and reported R^2 values for v5 are computed in this PCA space. For some evaluations (e.g., the optional sweep path), reconstructions are mapped back to raw ROI space via inverse PCA and de-normalization.

Time grid. Trials are resampled to a common duration and the time vector is normalized to $[0, 1]$; we denote the resulting grid by $0 = t_1 < \dots < t_L = 1$. The latent dimension is denoted by D .

E65 dataset. We use the Schotttdorf Lab E65 dataset (E65_data.npz), containing calcium activity ($\Delta F/F$) from $N = 375$ ROIs over $T = 7434$ frames, along with trial IDs, timestamps, and aligned behavioral covariates. In the v5 preprocessing path: (i) PCA is fit to the full recording and retains 95% variance, producing $K = 129$ components; (ii) frames are grouped by trial, the first 10 trials are dropped, and each trial is linearly interpolated to a fixed length $L = 120$ (trial_len_s=12, fps=10); (iii) the time vector is normalized to $[0, 1]$; (iv) each PCA component is standardized over time (session-level z-score) and an optional per-trial baseline is removed by subtracting the mean of

the first 5 resampled bins. After filtering, 180 trials are available; default validation holds out the last 3 trials (train 177 / val 3). For efficiency, the default configuration further subsamples to 100 sequences via greedy landmark selection.

Synthetic benchmark. We additionally evaluate on `synthetic_rat_data.npz` (4000 frames, 300 neurons, 20 trials), which provides a controlled benchmark for recoverability of smooth low-dimensional dynamics.

4 Model: Latent Neural ODE-VAE

4.1 Stochastic encoder

For each trial, the encoder uses only $x_b(t_1)$ and outputs a diagonal Gaussian posterior on the latent initial state:

$$q_\phi(z_{0,b} \mid x_b(t_1)) = \mathcal{N}(\mu_b, \text{diag}(\sigma_b^2)), \quad (1)$$

with reparameterization

$$z_{0,b} = \mu_b + \sigma_b \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I). \quad (2)$$

Here $z_{0,b} \in \mathbb{R}^D$, $\mu_b \in \mathbb{R}^D$, $\sigma_b \in \mathbb{R}_{>0}^D$, and \odot denotes elementwise multiplication.

4.2 Continuous-time latent dynamics

Latent trajectories are generated by a neural ODE:

$$\frac{dz_b(t)}{dt} = f_\theta(z_b(t), t), \quad z_b(t_1) = z_{0,b}. \quad (3)$$

In v5, f_θ is a mixture of experts:

$$f_\theta(z, t) = \sum_{e=1}^E \pi_e(z) f_e(z), \quad \pi(z) = \text{softmax}(g(z)), \quad (4)$$

with $E = 4$ latent experts by default and Dormand–Prince integration (`dopri5`). Each expert $f_e : \mathbb{R}^D \rightarrow \mathbb{R}^D$ is an MLP and $\pi_e(z) \in [0, 1]$ are gating weights satisfying $\sum_e \pi_e(z) = 1$.

4.3 Decoder family

A decoder maps latent states back to observations:

$$\hat{x}_b(t_\ell) = g_\psi(z_b(t_\ell)). \quad (5)$$

The codebase supports MLP, neuron-aware, local-attention, and MoE decoders; v5 default is MoE decoder with 8 decoder experts. In all cases, $g_\psi : \mathbb{R}^D \rightarrow \mathbb{R}^K$ outputs the mean of a factorized Gaussian observation model in PCA space.

5 Training Objective and Regularization

We optimize a variational objective with auxiliary regularizers. Under a Gaussian observation model $p_\psi(x_b(t_\ell) \mid z_b(t_\ell)) = \mathcal{N}(g_\psi(z_b(t_\ell)), \sigma^2 I)$ with fixed σ^2 , maximizing the ELBO corresponds (up to constants and a scale factor) to minimizing mean-squared reconstruction error plus a KL penalty.

The base objective combines reconstruction and KL terms:

$$\mathcal{L}_{\text{base}} = \mathcal{L}_{\text{rec}} + \beta \mathcal{L}_{\text{KL}}, \quad (6)$$

where

$$\mathcal{L}_{\text{rec}} = \frac{1}{B L K} \sum_{b=1}^B \sum_{\ell=1}^L \|\hat{x}_b(t_\ell) - x_b(t_\ell)\|_2^2, \quad (7)$$

$$\mathcal{L}_{\text{KL}} = \frac{1}{B} \sum_b D_{\text{KL}}(q_\phi(z_{0,b} \mid x_b(t_1)) \parallel \mathcal{N}(0, I)). \quad (8)$$

Equivalently, the (negative) ELBO per trial is

$$\mathcal{L}_{\text{ELBO}} = -\mathbb{E}_{q_\phi(z_{0,b}|x_b(t_1))} \left[\sum_{\ell=1}^L \log p_\psi(x_b(t_\ell) | z_b(t_\ell)) \right] + \beta D_{\text{KL}}(q_\phi(z_{0,b} | x_b(t_1)) \| p(z_{0,b})), \quad (9)$$

with prior $p(z_{0,b}) = \mathcal{N}(0, I)$. In practice, the code uses a single Monte Carlo sample of $z_{0,b}$ per trial and minibatch.

Smoothness regularization.

$$\mathcal{L}_{\text{smooth}} = \frac{1}{B(L-1)D} \sum_{b,\ell} \left\| \frac{z_b(t_{\ell+1}) - z_b(t_\ell)}{t_{\ell+1} - t_\ell} \right\|_2^2. \quad (10)$$

Transition-aware regularization (v5).

$$\mathcal{L}_{\text{trans}} = \frac{1}{B(L-1)K} \sum_{b,\ell} \left\| (\hat{x}_b(t_{\ell+1}) - \hat{x}_b(t_\ell)) - (x_b(t_{\ell+1}) - x_b(t_\ell)) \right\|_2^2. \quad (11)$$

This term is linearly warmed up for the first 30 epochs.

Soft LLE latent regularization (v5). For flattened latent points $\{z_i\}_{i=1}^M \subset \mathbb{R}^D$, with k -NN set $\mathcal{N}_k(i)$, we add a soft locally linear embedding penalty [Saul and Roweis, 2003]:

$$\mathcal{L}_{\text{LLE}} = \frac{1}{M} \sum_{i=1}^M \left\| z_i - \sum_{j \in \mathcal{N}_k(i)} w_{ij} z_j \right\|_2^2, \quad w_{ij} \propto \exp\left(-\frac{\|z_i - z_j\|_2}{\tau}\right). \quad (12)$$

Default parameters: $k = 8$, $M \leq 256$, $\tau = 0.1$.

Total loss.

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \beta_t \mathcal{L}_{\text{KL}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}} + \lambda_{\text{trans},t} \mathcal{L}_{\text{trans}} + \lambda_{\text{LLE}} \mathcal{L}_{\text{LLE}}. \quad (13)$$

The KL coefficient β_t is warmed up over 30 epochs to a final value $\beta = 0.02$.

6 Experimental Protocol

6.1 Configurations

Main v5 settings from `config.txt`: latent dimension $D = 5$ (with a sweep to $D = 8$), batch size 8, 150 epochs, Adam optimizer (learning rate 0.002, weight decay 10^{-5}), $\lambda_{\text{smooth}} \in \{5 \times 10^{-4}, 2 \times 10^{-4}\}$, $\lambda_{\text{trans}} = 0.01$ with a 30-epoch warmup, $\lambda_{\text{LLE}} = 0.01$, landmark count 100, and baseline correction enabled.

Implementation details. The encoder is an MLP with hidden widths 512–256–128. The latent vector field uses $E = 4$ experts with hidden width 128 and a learned gating network; derivatives are layer-normalized for stability. For reconstruction, the default v5 decoder is a mixture-of-experts network with 8 decoder experts and hidden width 256. Latent dynamics are integrated with Dormand–Prince (dopri5) using tolerances `rtol=1e-3` and `atol=1e-4`. Gradients are clipped to max norm 1.0.

6.2 Metrics

The primary training metric is coefficient of determination,

$$R^2 = 1 - \frac{\sum_{b,\ell} \|x_b(t_\ell) - \hat{x}_b(t_\ell)\|_2^2}{\sum_{b,\ell} \|x_b(t_\ell) - \bar{x}\|_2^2}, \quad (14)$$

where $\bar{x} = \frac{1}{BL} \sum_{b,\ell} x_b(t_\ell)$ denotes the mean activity vector across all validation entries in the evaluation space. Our implementation supports two evaluation styles: (i) PCA-space R^2 and (ii) strict raw-neuron-space R^2 via inverse PCA and de-normalization. We report values as saved in the run metadata for each experiment.

Table 1: Synthetic random-foraging benchmark from `seed_sweep_results.txt`.

Seed	Final R^2	Best validation loss
1	0.9789	0.06642
42	0.6757	0.40521
1337	0.9116	0.16298
2025	0.7467	0.31145
777	0.0000 (NaN collapse)	∞
Mean (all seeds)	0.6626	–
Mean (non-collapsed seeds)	0.8282	–

Table 2: E65 reconstruction metrics extracted from saved run artifacts. Sweep runs report both best-epoch and final R^2 as logged during training.

Variant	Space	D	λ_{smooth}	Best R^2	Final R^2
archived ODE-VAE	PCA	–	–	–	0.4368
v5 sweep	PCA	5	5×10^{-4}	0.0354	-0.2391
v5 sweep	PCA	5	2×10^{-4}	0.0353	-0.1382
v5 sweep	PCA	8	5×10^{-4}	0.0569	-0.2022
v6 (no PCA)	raw	–	–	–	0.0956

7 Results

7.1 Synthetic benchmark: high ceiling with seed sensitivity

Table 1 summarizes a five-seed sweep on synthetic data. The best seed reaches $R^2 = 0.9789$. Excluding the divergent run, the mean performance is $R^2 = 0.828 \pm 0.141$ (std. dev.), indicating sensitivity to initialization even in a controlled setting.

7.2 E65 hippocampal data: mixed performance across variants

Table 2 reports available E65 run metrics for the ODE-VAE family. An archived ODE-VAE run reports $R^2 = 0.4368$. In contrast, v5 sweep artifacts are substantially lower (best-epoch $R^2 \leq 0.0569$, negative final R^2 in three runs), while a no-PCA v6 run records $R^2 = 0.0956$. These discrepancies are consistent with the codebase supporting multiple evaluation spaces and split protocols; therefore, comparisons across runs should be interpreted cautiously unless the evaluation definition is matched.

7.3 Manifold interpretability

The codebase saves latent manifold projections (MDS) and reconstruction diagnostics for each run. Figure 1 shows an example latent trajectory embedding from the trained model artifacts.

8 Discussion

The model captures the intended inductive bias: low-dimensional continuous latent trajectories with explicit geometric regularization. On synthetic data, this bias is highly effective. On real E65 recordings, however, results are sensitive to implementation and evaluation choices.

Three factors emerge from the saved run artifacts:

1. **Metric-space mismatch.** PCA-space training can look favorable while strict raw-space R^2 may degrade.
2. **Data-efficiency tradeoff.** Landmark subsampling (100 selected sequences from 180 usable trials) accelerates training but may reduce generalization.
3. **Optimization stability.** Strong regularization with small validation sets (3 trials) and stiff latent dynamics can produce unstable or negative final R^2 , despite early high points.

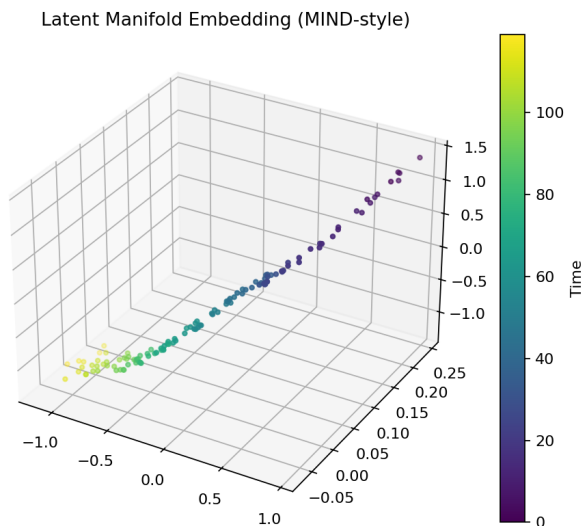


Figure 1: Latent manifold embedding produced by the ODE-VAE analysis pipeline.

These observations suggest that future gains likely require protocol-level changes in addition to architectural changes: larger and randomized holdout splits, early stopping on a stable cross-validated objective, trial-level (not frame-level) landmark selection, and direct raw-space reconstruction losses.

9 Limitations and Reproducibility

This study is bounded by the available run artifacts and inherits version-specific logging differences. In particular, some run files report “best” and “final” R^2 under different conditions, and not all checkpoints include identical metadata fields. We therefore report values exactly as saved in each artifact path. The implementation also exhibits training fragility (including occasional NaN divergence), which should be addressed before drawing definitive biological conclusions.

10 Conclusion

We presented a mathematically grounded latent Neural ODE-VAE framework for neural manifold modeling and analyzed a sequence of model variants (v1-v6), with v5 as the primary model. The method can recover smooth low-dimensional dynamics and high synthetic reconstruction quality, but real-data performance remains sensitive to preprocessing and evaluation protocol. This work provides a formal foundation and concrete directions for improving robustness of ODE-VAE manifold modeling for neuroscience.

Societal impact. This work is basic research on neural representation learning from animal neuroscience data and has no immediate direct societal deployment. Potential long-term impact is improved scientific understanding of memory and cognition.

References

- Dmitriy Aronov, Rachel Nevers, and David W. Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722, 2017.
- Jacob L. S. Bellmund, Peter Gärdenfors, Edvard I. Moser, and Christian F. Doeller. Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415):eaat6766, 2018.
- Rishidev Chaudhuri, Burak Gerçek, Bikash Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, 22:1512–1520, 2019.

190 Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary
191 differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

192 Alexandra O. Constantinescu, Jill X. O’Reilly, and Timothy E. J. Behrens. Organizing conceptual
193 knowledge in humans with a grid-like code. *Science*, 352(6292):1464–1468, 2016.

194 Howard Eichenbaum and Neal J. Cohen. Can we reconcile the declarative memory and spatial
195 navigation views on hippocampal function? *Neuron*, 83(4):764–770, 2014.

196 Juan A. Gallego, Matthew G. Perich, Lee E. Miller, and Sara A. Solla. Neural manifolds for the
197 control of movement. *Neuron*, 94(5):978–984, 2017.

198 Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference
199 on Learning Representations (ICLR)*, 2014.

200 Ryan J. Low, Sean Lewallen, Dmitriy Aronov, Rachel Nevers, and David W. Tank. Probing variability
201 in a cognitive map using manifold inference from neural dynamics. *bioRxiv*, 2018. doi: 10.1101/
202 418939.

203 Edward H. Nieh et al. Geometry of abstract learned knowledge in the hippocampus. *Nature*, 595
204 (7865):80–84, 2021.

205 John O’Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map. preliminary evidence
206 from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175, 1971.

207 John O’Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map*. Clarendon Press, 1978.

208 Sang Ah Park, David S. Miller, Hamed Nili, Charan Ranganath, and Erie D. Boorman. Map making:
209 Constructing, combining, and inferring on abstract cognitive maps. *Neuron*, 107(6):1226–1238.e8,
210 2020.

211 Yulia Rubanova, Ricky T. Q. Chen, and David Duvenaud. Latent ordinary differential equations for
212 irregularly-sampled time series. In *Advances in Neural Information Processing Systems (NeurIPS)*,
213 2019.

214 Abigail A. Russo et al. Motor cortex embeds muscle-like commands in an untangled population
215 response. *Neuron*, 97(4):953–966.e8, 2018.

216 Lawrence K. Saul and Sam T. Roweis. Think globally, fit locally: Unsupervised learning of low
217 dimensional manifolds. *Journal of Machine Learning Research*, 4:119–155, 2003.

218 Nicolas W. Schuck and Yael Niv. Sequential replay of nonspatial task states in the human hippocam-
219 pus. *Science*, 364(6447):eaaw5181, 2019.

220 Kimberly L. Stachenfeld, Matthew M. Botvinick, and Samuel J. Gershman. The hippocampus as a
221 predictive map. *Nature Neuroscience*, 20(11):1643–1653, 2017.

222 Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for
223 nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

224 Edward C. Tolman. Cognitive maps in rats and men. *Psychological Review*, 55(4):189–208, 1948.

225 Byron M. Yu, John P. Cunningham, Gopal Santhanam, Stephen I. Ryu, Krishna V. Shenoy, and
226 Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of
227 neural population activity. *Journal of Neurophysiology*, 102(1):614–635, 2009.