Variational inference of latent hierarchical dynamical systems in neuroscience: an application to calcium imaging data

Anonymous authors
Paper under double-blind review

Abstract

A key problem in neuroscience, and life sciences more generally, is that data is generated by a hierarchy of dynamical systems. One example of this is in *in-vivo* calcium imaging data, where data is generated by a lower-order dynamical system governing calcium flux in neurons, which itself is driven by a higher-order dynamical system of neural computation. Ideally, life scientists would be able to infer the dynamics of both the lower-order systems and the higher-order systems, but this is difficult in high-dimensional regimes. A recent approach using sequential variational auto-encoders demonstrated it was possible to learn the latent dynamics of a single dynamical system for computations during reaching behaviour in the brain, using spiking data modelled as a Poisson process. Here we extend this approach using a ladder method to infer a hierarchy of dynamical systems, allowing us to capture calcium dynamics as well as neural computation. In this approach, spiking events drive lower-order calcium dynamics, and are themselves controlled by a higher-order latent dynamical system. We generate synthetic data by generating firing rates, sampling spike trains, and converting spike trains to fluorescence transients, from two dynamical systems that have been used as key benchmarks in recent literature: a Lorenz attractor, and a chaotic recurrent neural network. We show that our model is better able to reconstruct Lorenz dynamics from fluorescence data than competing methods. However, though our model can reconstruct underlying spike rates and calcium transients from the chaotic neural network well, it does not perform as well at reconstructing firing rates as basic techniques for inferring spikes from calcium data. These results demonstrate that VLAEs are a promising approach for modelling hierarchical dynamical systems data in the life sciences, but that inferring the dynamics of lower-order systems can potentially be better achieved with simpler methods.

1 Introduction

Many datasets in the life sciences are generated by a hierarchy of dynamical systems, wherein lower-order dynamical systems that directly generate the data are driven by higher-order dynamical systems that are not observable. This problem is outlined in figure 1A, in which noisy observations x depend on the state z_1 of a low-order dynamical system that is perturbed by inputs u_1. The state of this dynamical system is also coupled to the state z_2 of a higher-order dynamical system, which can be perturbed independently by inputs u_2.

One example of such a system in *in-vivo* two-photon calcium imaging from neuroscience. Calcium imaging provides systems neuroscientists with the ability to observe the activity of hundreds of neurons simultaneously during behavioural experiments. Such experiments have allowed neuroscientists to ask questions about the underlying computations and algorithms that neural circuits are implementing in perception, decision-making, memory, and many other processes. Such experiments can be characterized as observing a hierarchical dynamical system (Fig 1B) in which measurable calcium fluorescence is primarily determined by dynamics based on voltage-gated calcium channels and calcium binding to fluorescence dyes, and the rate of fluorescence transients controlled by the underlying computation.
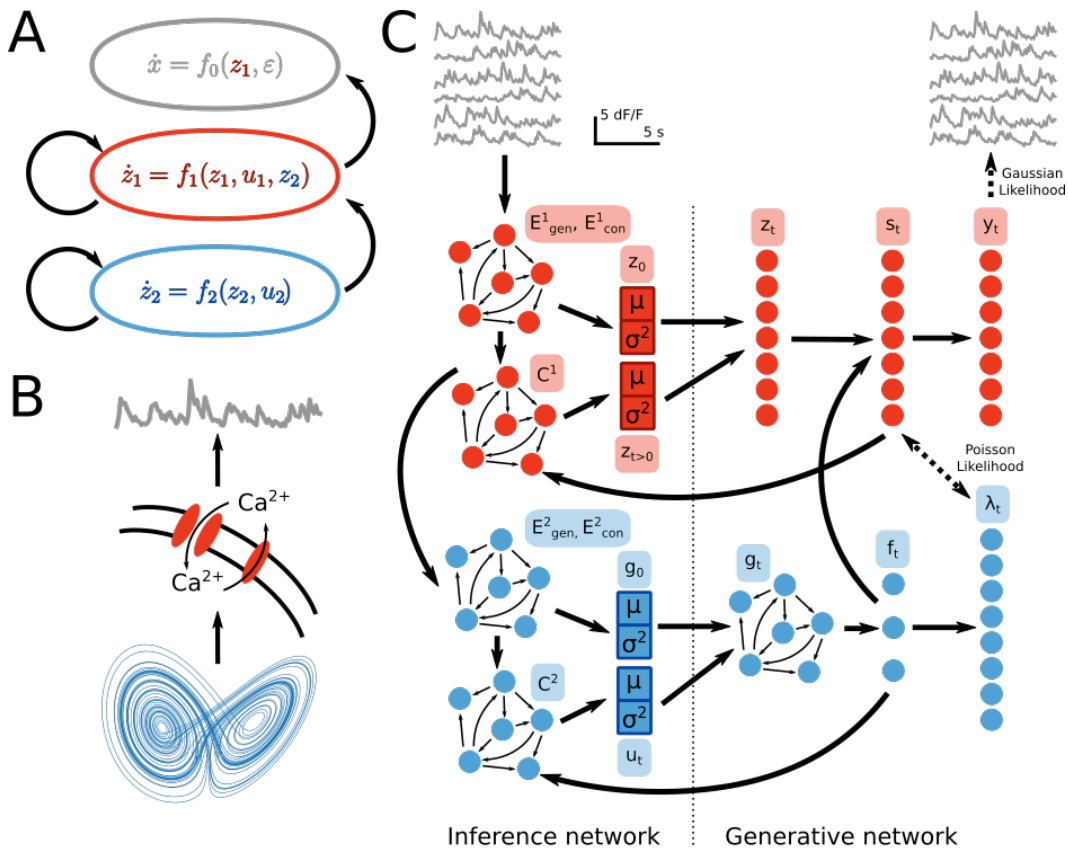
Figure 1: A) Hierarchy of dynamical systems, B) Schema of calcium and Lorenz dynamics, C) Schema of our hierarchical model. Latent dynamics model in blue, spike inference and calcium dynamics model in red.

Recent applications of sequential variational autoencoders to neural data analysis has seen great success in inferring underlying computations in populations of cells in macaque and human motor cortex Pandarinath et al. (2018). By characterizing neural computation as low-dimensional dynamic factors in a non-hierarchical dynamical systems, Pandarinath et al. (2018) showed that these dynamic factors trained to generate the inhomogeneous intensity functions explaining the rate of spikes assumed to follow a Poisson process. Crucially, these low-dimensional factors could also decode reaching behaviour of macaques and humans with much higher fidelity than any other dimensionality reduction method.

Although this is a significant advance in our ability to analyze neural data in the form of spikes trains, two-photon calcium imaging poses the additional problem of identifying latent spike trains in fluorescence traces. This problem has been independently addressed in a number of different ways, including deconvolution Friedrich et al. (2017) and variational inference Speiser et al. (2017).

If we continue to model the frequency of events as being generated by a Poisson process, this can be seen as hierarchy of dynamical systems (Fig 1A), in which low dimensional dynamics generate spike probabilities that in turn drive fluctuations in biophysical dynamics of calcium activity (Fig 1B. Here we propose a method that extends LFADS to accommodate calcium activity using this hierarchical dynamical systems approach, in which we can infer both the latent dynamics and the latent spike trains from the observed calcium fluorescence signal.

2 MODEL

The model is a variational ladder autoencoder (VLAE) (Zhao et al., 2017) with recurrent neural networks (RNNs) that supports uncovering latent dynamical systems (Fig 1C, full directed acyclic graph in Fig A1). It can be seen as a unification of two recent applications of variational autoencoders (VAEs) in neuroscience: 1) Latent Factor Analysis for Dynamical Systems (LFADS) (Pandarinath et al., 2018) and 2) DeepSpike, a VAE approach to inferring spike counts from calcium imaging data (Speiser et al., 2017). We choose the VLAE approach since it has been shown to avoid the problem of trying to reconstruct the data solely from the lower-order features by separating latent variables via divergent deterministic pathways in the encoder, and convergent deterministic pathways in the generator (Sønderby et al., 2016; Zhao et al., 2017). Model hyperparameters are shown in table A1.

The inferred dynamical system underlying the frequency of calcium events in the data is identical to that of LFADS (Fig 1C, blue modules). The prior distribution of initial conditions g_0 and external inputs u_t are modelled as Gaussian distributions $P(g_0) = \mathcal{N}(\mu_{g_0}, \sigma_{g_0}^2)$, and $P(u_t) = \mathcal{N}(\mu_{u_t}, \sigma_{u_t}^2)$. The underlying dynamical system $\dot{g} = G(g_t, u_t)$ is modelled by a Gated Recurrent Unit (GRU) taking the initial hidden state g_0 and inputs u_t. Low dimensional factors f_t are calculated as a linear transformation of the generator hidden state $f_t = W^{fac} g_t$. These factors are used to reconstruct the Poisson process intensity function with a fully connected layer and exponential non-linearity $\lambda_t = \exp(0.5(W_\lambda f_t + b_\lambda))$

Inferred spike counts s_t are generated by sampling z_t from Gaussian distributions $P(z_t) = \mathcal{N}(\mu_{z_t}, \sigma_{z_t}^2)$ and projecting these through an affine transformation and non-linearity along with the factors from the deeper layer, i.e., $s_t = \Phi(W_s[f_t, z_t] + b_s)$, where $\Phi(x) = ReLU(\exp(x) - 1)$ (Figure 1C blue modules). We assume a simple model of calcium dynamics: $\dot{y} = -y_t/\tau_y + \alpha_y s_t + \beta_y$ where the parameters $\tau_y, \alpha_y, \beta_y$ are measured from the data, however it is a topic for future research to fit the calcium dynamics simultaneously. In our synthetic data, these are valued at 0.4 s, 1, and 0 respectively. The value of τ_y is chosen as it is the known decay time constant of GCamP6, a commonly used calcium fluorescence indicator used in calcium imaging experiments.

2.1 ENCODING MODEL

The calcium fluorescence signal x_t is mapped onto the parameters of the variational posterior distributions $Q(z_t|x)$, $Q(g_0|x)$, $Q(u_t|x)$. These distributions are all modelled as Gaussians, with the mean and standard deviations parameterized by a stack of bidirectional GRUs, E_{gen}^1, E_{gen}^2, E_{con}^1, E_{con}^2. The final hidden states of E_{gen}^1 and E_{gen}^2 are mapped onto the parameters of $Q(z_0|x)$ and $Q(g_0|x)$ respectively with fully connected layers. The hidden states E_{con}^1 and E_{con}^2 are concatenated and passed as inputs to single direction GRUs C^1 and C^2. The hidden states of C^1 and C^2 are concatenated at each time step t with s_{t-1} and f_{t-1}. Subsequently these concatenated activities are mapped onto the parameters of $Q(z_t|x)$ and $Q(u_t|x)$ with fully connected layers.

2.2 LOSS FUNCTION AND TRAINING

One of the advantages of using VLAEs is that the evidence lower bound (ELBO) formulation is the same as for VAEs despite the hierarchical latent space (Zhao et al., 2017). As such, our cost function remains very similar to that of LFADS.

$$\begin{aligned}
\mathcal{L} = &-\mathbb{E}_{z \sim Q(z|x)}[\log(x|y)] \\
&- \mathbb{E}_{z,g_0,u \sim Q(z,g_0,u|x)}[\log(s|\lambda)] \\
&+ D_{\text{KL}}(Q(z, g_0, u|x) | P(z, g_0, u)) \\
&+ ||W^{gg}||_2 + ||s||_1
\end{aligned} \tag{1}$$

Where W^{gg} is the weight matrix in the GRU that take g as input. The likelihood function $P(x_t|y_t)$ is modelled as a Gaussian distribution $x_t \sim \mathcal{N}(y_t, \sigma_y^2)$, where σ_y^2 is learned. Although s_t is not discrete, $P(s_t|\lambda_t)$ is treated as an approximate Poisson process $s_t \sim Poisson(\lambda_t) = s_t^{\lambda_t} \exp(-\lambda_t)/\Gamma(s_t + 1)$. This is an inductive bias to ensure that the inferred spike trains in the

lower-order dynamic system is constrained to follow the dynamics of the higher-order dynamic system. The latent variables are treated as independent, with hierarchy imposed in the deterministic paths, i.e., $Q(z, g_0, u|x) = Q(z|x)Q(g_0|x)Q(u|x)$

Parameters of our model were optimized with ADAM, with an initial learning rate of 0.01, which decreased by a factor of 0.95 whenever plateaus in training error were detected. As in LFADS training, KL and L2 terms in the cost function were 'warmed up', i.e., had a scaling factor being 0 and 1 applied that gradually increased. Warm-up for the deeper parameters (blue modules in Figure 1) was delayed until warm-up for shallower parameters was completed (red modules in Figure 1).
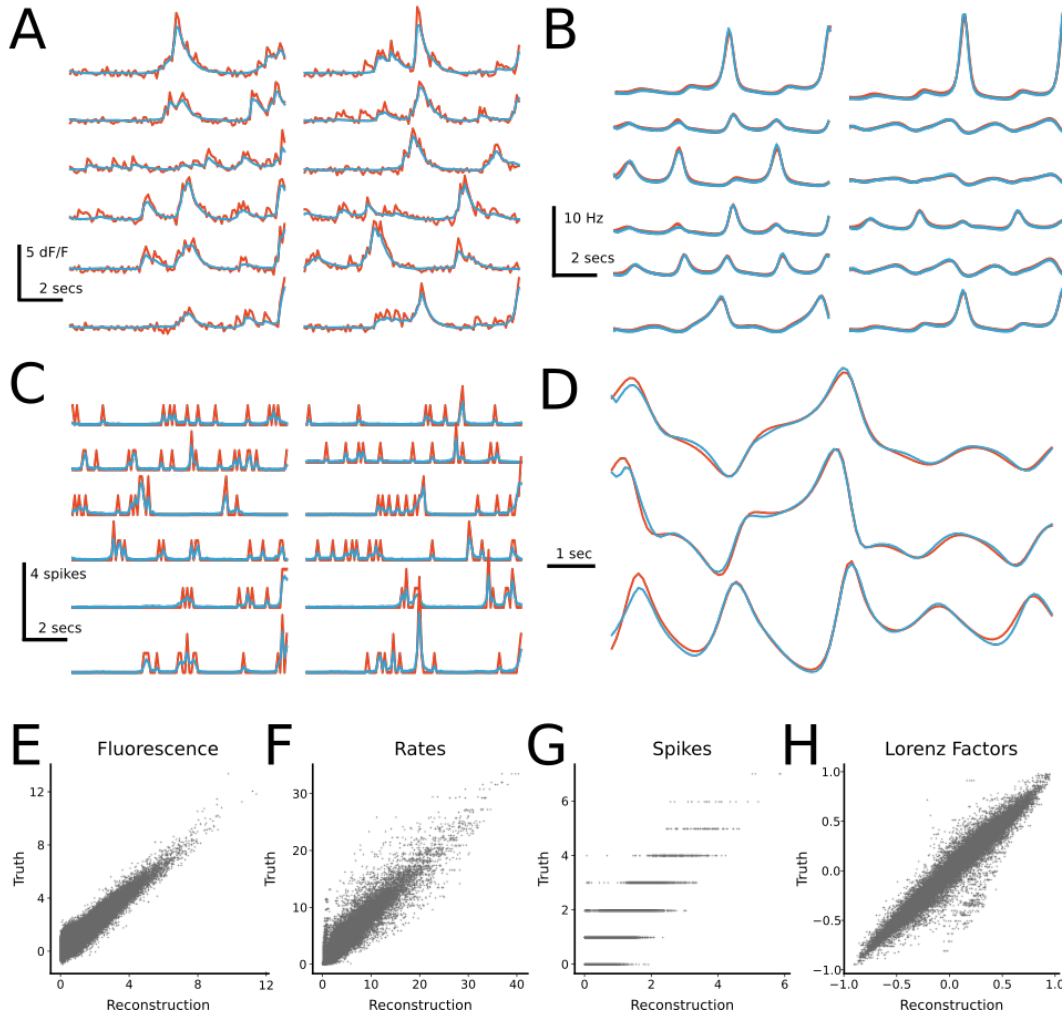
3 RESULTS

Figure 2: A) Example calcium fluorescence traces, B) Example Poisson intensity functions, C) Example spike trains, D) Inferred dynamics. Red: Ground-truth, Blue: Reconstructed, E-H Scatter plots of reconstructions and ground truth

3.1 LORENZ SYSTEM

The model was tested on synthetic data with Lorenz dynamics embedded in the frequency of calcium fluorescence transients, as described by Zhao & Park (2017), where generated spikes were convolved with an exponential kernel with a time constant of 0.4 ms, and white noise added to the resulting traces. We measure the performance of the model in three ways: 1) uncovering the

Table 1: Comparison of model performance on synthetic Lorenz dataset. A hyphen indicates it is not possible to compare, as the model does not infer this variable. The top row is italicized as it is considered the upper limit on performance in this task with LFADS since there is no additional observation noise from fluorescence.

Data	Model	End-to-end	Goodness-of-fit (R^2)			
			Lorenz	Rates	Spikes	Fluorescence
Spikes	*LFADS*	-	*.978*	*.970*	-	-
Fluorescence	Oasis + LFADS	No	.924	**.946**	**.950**	-
Fluorescence	Gaussian LFADS	Yes	.898	.771	-	.001
Fluorescence	Edgeworth LFADS	Yes	.614	.097	-	.001
Fluorescence	Ladder LFADS	Yes	**.962**	**.943**	.697	**.850**

underlying Lorenz dynamics, 2) reconstructing the rates of calcium transients an inhomogeneous Poisson intensity functions, 3) reconstructing the spike counts contributing to increases in the calcium fluorescence signal. The model was compared against a ground-truth where the spike counts are known, and LFADS is used to reconstruct the latent dynamics and intensity function, and against a model where spike counts are extracted using a deconvolution algorithm (Friedrich et al., 2017) before using LFADS to reconstruct the rates and intensity function (OASIS + LFADS). It was also tested against a model that used a 1-D convolution of the intensity function to reconstruct either the first two (Gaussian-LFADS) or four (Edgeworth-LFADS) time-varying moments of fluorescence, as used previously in estimating the intensity functions of filtered Poisson processes in neuroscience (Brigham & Destexhe, 2015).

Figure 2 shows examples of performance of our model in reconstructing the fluorescence traces (Fig 2A), Poisson intensity functions (Fig 2B), spikes (Fig 2C) and Lorenz dynamics (Fig 2D). Visually, the model provides very close fit to the fluorescence traces, intensity functions, and Lorenz dynamics. The model also captures spike-timing, although these spike trains appear smoothed. Table 1 compares the R^2 goodness-of-fit on reconstructing held-out validation data with ground-truth latent dynamic structure. Of all approaches, our model easily performs best at reconstructing fluorescence traces. It should be noted that the reason the Gaussian and Edgeworth LFADS models perform so poorly at reconstructing fluorescence is that it only attempts to reconstruct the time-varying moments of fluorescence as a 1-D convolution of the intensity function, which stems from the approximation of exponential-kernel shot noise processes (Brigham & Destexhe, 2015). Additionally, our model almost performs as well as LFADS in reconstructing the Lorenz dynamics. It is to be expected that LFADS performs better than our method, since there is an additional source of observation noise in our synthetic dataset generating fluorescence transients from spikes. Notably, our model does not perform as well as the deconvolution method OASIS in reconstructing spike trains, however this does not impact the ability of our model to reconstruct the latent dynamics. In fact, constraining the reconstructed spikes by the latent dynamics may mitigate any errors in spike train reconstruction that occur by deconvolution, since the deconvolution algorithm may erroneously drop spikes during high rates, whereas our model should be less likely to do so. It will be necessary to assess this possibility further.

It should be noted that the deconvolution algorithm performs much better at reconstructing spike trains in our synthetic dataset than in real datasets where ground-truth spiking is known (Pachitariu et al., 2018). To our knowledge, there are no known dual recordings of population 2-photon calcium imaging with ground-truth electrophysiology in a subpopulation of neurons in-vivo during behaviour driven by hypothesized low-dimensional dynamics that we would be able to validate this with. Nevertheless, since the relationship between calcium dynamics and somatic spiking is highly non-linear, especially in dendrites, it remains to be seen how useful it is to faithfully reproduce unseen spike trains in calcium fluorescence activity.

3.2 CHAOTIC RECURRENT NETWORK

The model was then tested on synthetic data generated from a 50 cell recurrent neural network with chaotic dynamics, and an external perturbation at a random time, as described in Pandarinath et al. (2018), with parameters adjusted to make the data more representative of firing rates and time scales

observed in calcium imaging experiments. Spike trains were transformed into fluorescence signals using the same procedure as with the Lorenz system dataset.
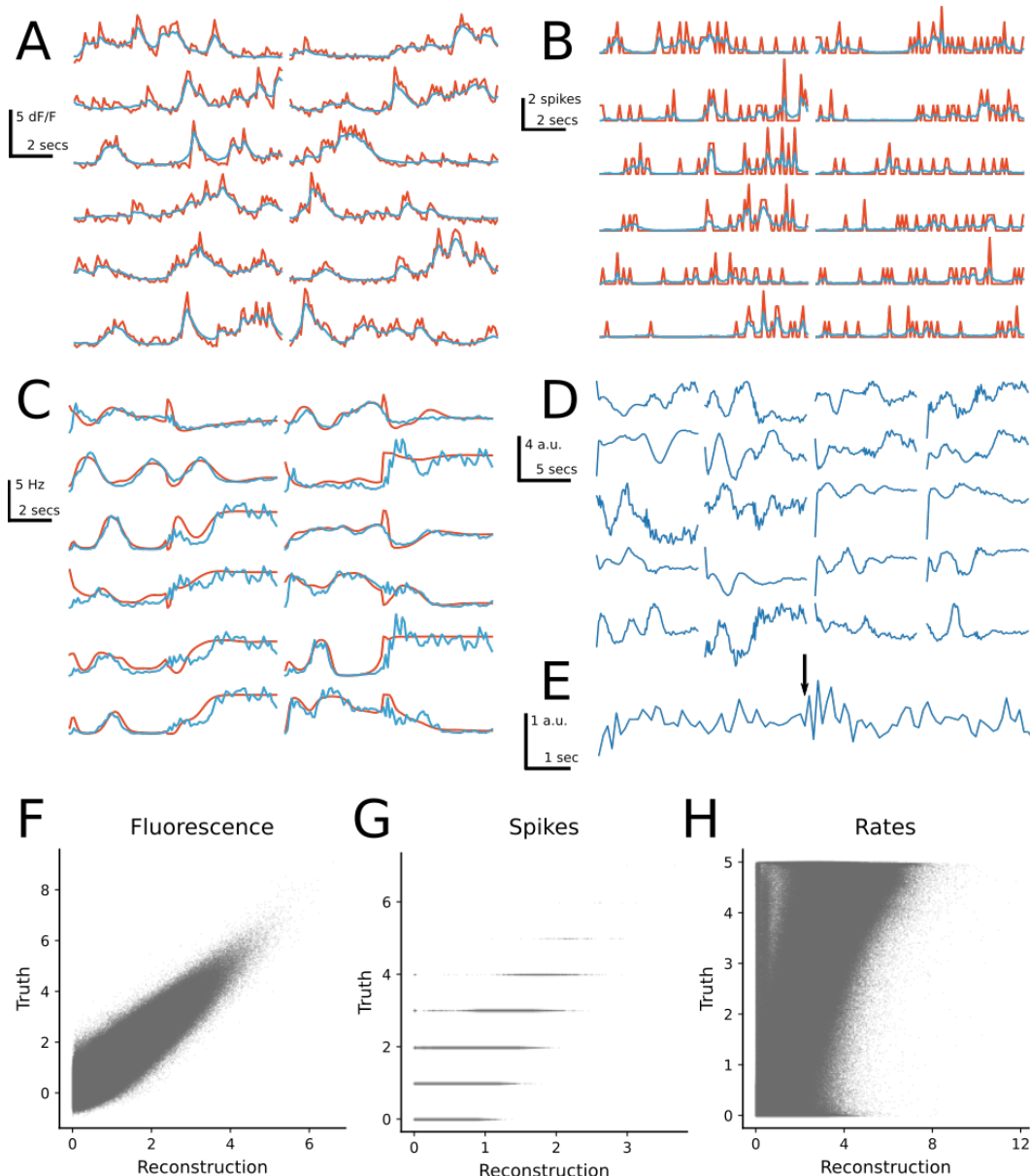

Figure 3: A) Example calcium fluorescence traces, B) Example spike trains, C) Example Poisson intensity functions, D) Inferred latent dynamics, E) Inferred input. The black arrow indicates the true timing of external perturbation. Red: Ground-truth, Blue: Reconstructed

Figure 3A-C shows examples that illustrates the performance of our model in reconstructing fluorescence traces (Fig 3A), spike trains (Fig 3B), and intensity functions (Fig 3C). Visually, the fluorescence traces and spike trains have been reconstructed reasonably well, whereas reconstructions of the the intensity functions are highly noisy. The latent dynamics represented by the factors (Fig 3D) and external inputs (Fig 3E) also show a lot of noise. This appears to be due to difficulty in inferring external inputs, in which it is not clear whether the timing of external perturbation has been accurately inferred, despite a slight transient at roughly the time of the perturbation in the example. Table 2 compares the R^2 goodness-of-fit on reconstructing held-out validation data, which demonstrates that our model performs very well in reconstructing the fluorescence signal and does reasonably well at reconstructing spikes. However, in this more complex benchmark where the deeper dynamic

system is perturbed by external input, the pre-processing of fluorescence transients with the OASIS deconvolution algorithm reconstructs firing rates far better than our method.

Table 2: Comparison of model performance on synthetic chaotic recurrent network dataset. A hyphen indicates it is not possible to compare, as the model does not infer this variable. The top row is italicized as it is considered the upper limit on performance in this task with LFADS since there is no additional observation noise from fluorescence.

			Goodness-of-fit (R^2)		
Data	Model	End-to-end	Rates	Spikes	Fluorescence
Spikes	*LFADS*	-	*0.839*	-	-
Fluorescence	Oasis + LFADS	No	**0.808**	**0.956**	-
Fluorescence	Gaussian LFADS	Yes	0.656	-	6.27×10^{-6}
Fluorescence	Edgeworth LFADS	Yes	0.706	-	6.55×10^{-6}
Fluorescence	Ladder LFADS	Yes	0.719	0.520	**0.811**

4 DISCUSSION

We present a hierarchical recurrent variational autoencoder model capable of reconstructing latent dynamics, latent spike trains, and calcium fluorescence traces in a benchmark synthetic dataset. Of the four methods tested, our model is the only one capable of reconstructing all three. Furthermore, our model performed best in reconstructing latent dynamics in our synthetic dataset We will need to assess our model on further synthetic benchmark data to assess the validity of our approach.

Since our model is trained end-to-end, it should be possible to extend to reconstructing raw 2-photon imaging videos, which could enable us to train models to uncover latent dynamics from arbitrarily shaped neuronal structures. This would of great use to neuroscientists who are largely restricted to techniques that extract fluorescence traces from regions of interest with somatic shapes, whereas the morphological diversity of dendrites is much greater.

An additional advantage of using our hierarchical model is that we can obtain measures of the uncertainty in both the latent dynamics, and the latent spike trains. The correlation in uncertainty between layers of this hierarchy may be what allows superior inference of the latent dynamics, despite less accurate reconstructions of the spike trains than OASIS, which provides no measure of uncertainty. We hope to improve our model to better capture the relationships between layers of this hierarchy in future.

We describe a use-case in neuroscience (2-photon calcium imaging data) for which this model may be very useful. However, we are keen to investigate the general case of hierarchical dynamical systems and their utility in uncovering structure in datasets outside this domain.

REFERENCES

Marco Brigham and Alain Destexhe. Nonstationary filtered shot-noise processes and applications to neuronal membranes. *Physical Review E*, 91(6):062102, June 2015.

Johannes Friedrich, Pengcheng Zhou, and Liam Paninski. Fast online deconvolution of calcium imaging data. *PLOS Computational Biology*, 13(3):e1005423, March 2017.

Marius Pachitariu, Carsen Stringer, and Kenneth D. Harris. Robustness of Spike Deconvolution for Neuronal Calcium Imaging. *The Journal of Neuroscience*, 38(37):7976–7985, September 2018.

Chethan Pandarinath, Daniel J. OShea, Jasmine Collins, Rafal Jozefowicz, Sergey D. Stavisky, Jonathan C. Kao, Eric M. Trautmann, Matthew T. Kaufman, Stephen I. Ryu, Leigh R. Hochberg, Jaimie M. Henderson, Krishna V. Shenoy, L. F. Abbott, and David Sussillo. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815, October 2018.

Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. Ladder Variational Autoencoders. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 29*, pp. 3738–3746. Curran Associates, Inc., 2016.

Artur Speiser, Jinyao Yan, Evan W Archer, Lars Buesing, Srinivas C Turaga, and Jakob H Macke. Fast amortized inference of neural activity from calcium imaging data with variational autoencoders. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 30*, pp. 4024–4034. Curran Associates, Inc., 2017.

Shengjia Zhao, Jiaming Song, and Stefano Ermon. Learning Hierarchical Features from Deep Generative Models. In *International Conference on Machine Learning*, pp. 4091–4099, July 2017.

Yuan Zhao and Il Memming Park. Variational Latent Gaussian Process for Recovering Single-Trial Dynamics from Population Spike Trains. *Neural Computation*, 29(5):1293–1316, March 2017.

A   APPENDIX

Table A1: Glossary of variables, with descriptions and dimensionality across two synthetic datasets

| Variable	Description	Dimensions	
		Lorenz	Chaotic RNN
x_t	Calcium fluorescence signal	30	50
$e_t^{g_1}$	Latent spikes embedding initial condition encoder state	128	400
$e_t^{g_2}$	Latent dynamics embedding initial condition encoder state	64	200
$e_t^{c_1}$	Latent spike embedding input encoder state	128	256
$e_t^{c_2}$	Latent dynamics embedding input encoder state	-	128
c_t^1	Latent spike embedding controller state	128	256
c_t^2	Latent dynamics embedding controller state	-	128
z_t	Latent spike embedding	100	250
g_t	Latent dynamics embedding	64	200
u_t	Latent inputs embedding	0	1
f_t	Latent dynamic factors	3	20
s_t	Latent spikes	30	50
y_t	Mean fluorescence signal	30	50
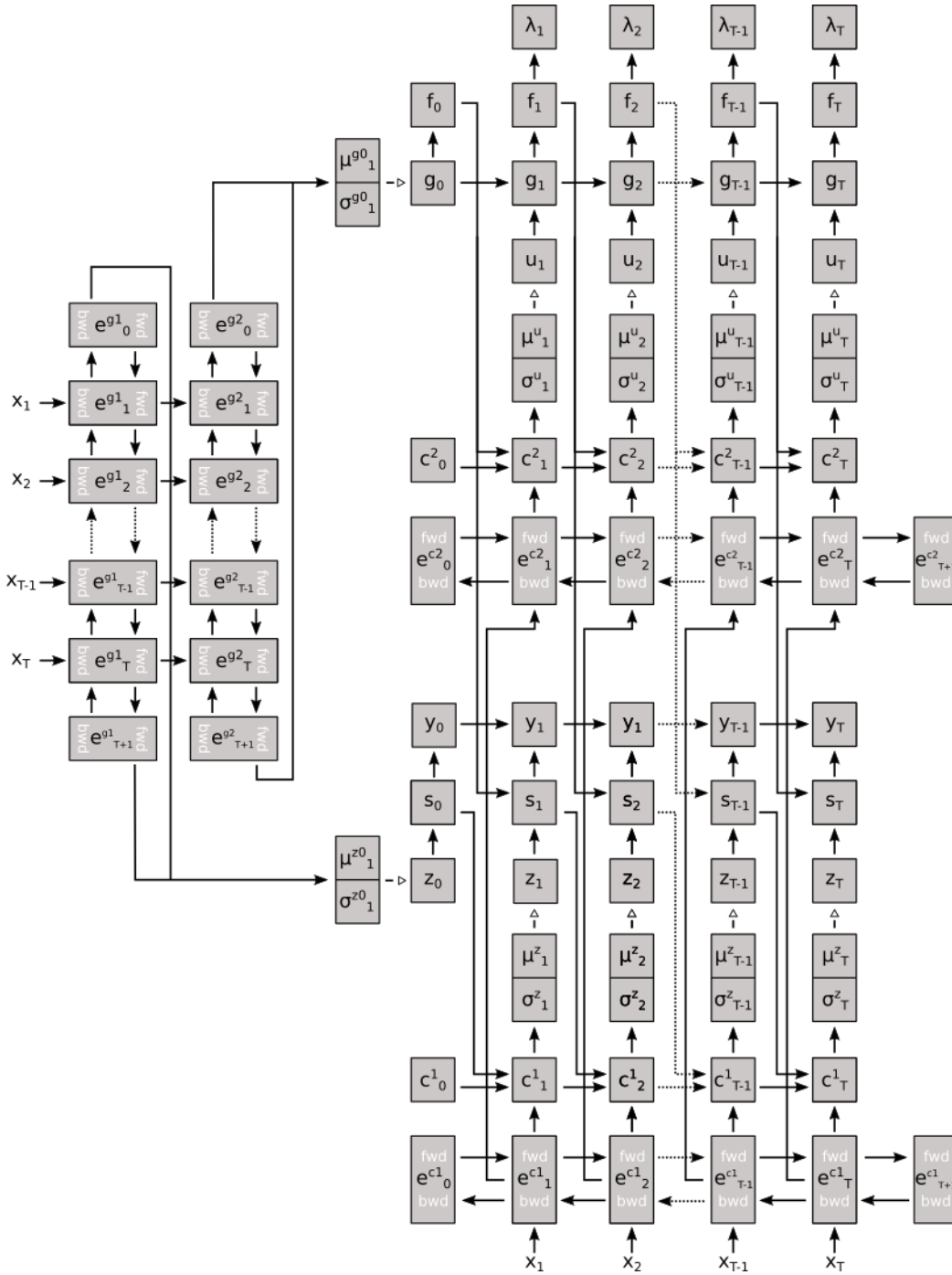
Figure A1: Directed acyclic graph for hierarchical model. Solid arrows denote deterministic mappings, open arrows denote sampling steps.