

**Università degli Studi di Padova**

Dipartimento di Fisica e Astronomia

*“Galileo Galilei”*

**Predicting brain activity from arousal measurements**  
**Predire l'attività cerebrale tramite misure di eccitazione**

**Corso di Laurea in Fisica**

**Anno Accademico 2024/2025**

**Autore: Federico Sbarbati**

Matricola: 2008448

**Relatore: Dr. Michele Allegra**

**Padova, March 24, 2025**

# Contents

<b>Abstract</b>	<b>3</b>
<b>1 Introduction</b>	<b>5</b>
<b>2 Methods</b>	<b>6</b>
2.1 Delay Coordinate Embedding . . . . .	6
2.1.1 Takens' Embedding Theorem . . . . .	6
2.1.2 Choice of Delay and Embedding Dimension . . . . .	7
2.2 Dimensionality reduction . . . . .	8
2.2.1 Discrete Legendre polynomials . . . . .	8
2.2.2 Variational Autoencoders . . . . .	9
2.3 Outline of the Methods . . . . .	10
<b>3 Results</b>	<b>12</b>
3.1 Lorenz Toy . . . . .	12
3.1.1 Preprocessing . . . . .	12
3.1.2 Encoder training and performance . . . . .	13
3.1.3 Decoder training and performance . . . . .	14
3.2 Zenodo brain data . . . . .	18
3.2.1 Dataset analysis . . . . .	18
3.2.2 Preprocessing . . . . .	20
3.2.3 Network performance . . . . .	20
<b>4 Conclusions</b>	<b>25</b>
<b>Bibliography</b>	<b>25</b>

# Abstract

Several imaging techniques, such as functional magnetic resonance imaging (fMRI) or calcium imaging, allow spatially fine-grained measurements of brain activity in time. By combining an established technique in dynamical systems analysis (time-delayed embedding) with recently developed artificial-neural-network based methods (variational autoencoders), one can retrieve an explicit description of a low-dimensional dynamical system underlying the observed time series. How this low-dimensional dynamics relates to the overall level of arousal, a key physiological parameter, was until recently not known. Adding measurements of arousal (which is non-invasively captured by the pupil diameter) to the autoencoder, a recent publication (Raut et al., 2023 [6]) showed that arousal can predict a large part of the observed dynamics. In this thesis, we will review the methodology of Raut et al., apply it to simplified artificial scenarios, and try to reproduce a few results of Raut et al.

# Abstract

Diverse tecniche di imaging, come la risonanza magnetica funzionale (fMRI) e l'imaging del calcio, permettono di ottenere precise misure dell'attività cerebrale nel tempo. Combinando una tecnica consolidata nell'analisi dei sistemi dinamici (l'embedding a ritardo temporale) con metodi basati su reti neurali artificiali recentemente sviluppati (variational autoencoders), è possibile ricavare una descrizione esplicita di un sistema dinamico a bassa dimensionalità che sottende le serie temporali osservate. Fino a poco tempo fa, non era noto come questa dinamica a bassa dimensionalità fosse correlata al livello generale di eccitazione (arousal), un importante parametro fisiologico. Tuttavia, tramite misure di eccitazione (che può essere catturata in modo non invasivo tramite il diametro pupillare) l'autoencoder è capace di predire gran parte della dinamica osservata, come dimostrato nella recente pubblicazione di Raut et al., 2023 [6]. In questa tesi, esamineremo la metodologia di Raut et al., la applicheremo a scenari artificiali semplificati e cercheremo di riprodurre alcuni dei risultati presentati nello studio.

# Chapter 1

## Introduction

In the past decade, several studies have demonstrated a strong link between behavioral and brain activity patterns. In particular, the time course of pupil diameter reflects changes in brain states and cognitive processes. This is why decoding the relationship between pupil dynamics and resting-state functional magnetic resonance imaging (fMRI) has become a fertile research topic, with a growing body of literature. Sobczak et al. [7] proved that pupil dynamics are tightly coupled with different neuromodulatory centers, showing that they are correlated with the leading principal components (PCs) of resting-state fMRI. These components exhibit patterns related to arousal fluctuations and autonomous regulation. Accordingly, Raut et al. [6] hypothesized that “an organism-wide regulatory process constitutes the primary mechanism underlying spatially structured patterns of spontaneous activity observed on the timescale of seconds,” which they referred to as *arousal*. They proposed a methodology that combines classical techniques in state-space reconstruction—such as delay coordinate embedding and Principal Component Analysis (PCA)—with modern artificial neural network-based approaches to model arousal. Their framework clearly shows the benefits of integrating these methods. Their results suggest that arousal may account for even more neural variance than what is conventionally attributed to it. This improvement in predictive performance was achieved by employing a combination of linear and nonlinear techniques, featuring a specific neural architecture: the Variational Autoencoder (VAE). In their experiments, Raut et al. performed simultaneous multimodal widefield optical imaging and face videography in awake mice. They trained a neural network to reconstruct widefield calcium dynamics from pupil diameter measurements, demonstrating the effectiveness of their approach in capturing brain-wide physiological fluctuations. In this thesis, we will review the methodology applied by Raut et al. and evaluate its performance first in an artificial explanatory scenario and then on real data. Firstly, we will test the methodology on a well-known dynamical system (Stochastic Lorenz attractor) with different levels of observation and dynamical noise, showing that it can reconstruct the original 3-dimensional phase space from a one-dimensional coordinate. Then we will use the publicly available data from Reference [7] to predict the resting-state fMRI of anesthetized rats from pupil diameter measurements.

# Chapter 2

## Methods

In this section, we present the methodology and procedures adopted in our analysis. Our objective is to develop a method for accurately predicting resting-state fMRI (rs-fMRI) dynamics using only the temporal evolution of the pupil diameter, under the assumption that both signals are governed by a shared latent process associated with arousal fluctuations [6] [7]. Our goal is to develop a framework capable of reconstructing the dynamics of a set of observables driven by a common latent process, using only one of these as observable as input. To achieve optimal results, we will employ a combination of Delay Coordinate Embedding, dimensionality reduction techniques - such as PCA and discrete Legendre polynomials [2] - and a Variational Autoencoder [5] to learn a latent representation of the unknown phase space.

### 2.1 Delay Coordinate Embedding

#### 2.1.1 Takens' Embedding Theorem

Takens' Embedding Theorem (Floris Takens, 1981) is a fundamental result in the reconstruction of dynamical systems and time series. The theorem states that, given a deterministic dynamical system with a  $d$ -dimensional phase space, it is possible to reconstruct an equivalent representation of the system in a higher-dimensional space using a technique known as Delay Coordinate Embedding.

Formally, for a discrete-time dynamical system described by a function  $f : M \rightarrow M$ , where  $M$  is a  $d$ -dimensional smooth manifold, and an observable in the system  $h : M \rightarrow \mathbb{R}$ , it is possible to create an embedding map:

$$\Phi : M \rightarrow \mathbb{R}^m \tag{2.1}$$

where  $m \geq 2d + 1$ . The map  $\Phi$  is built using a time delay sequence of the observable  $h$ :

$$\Phi(x) = (h(x), h(F(x)), h(F^2(x)), \dots, h(F^{m-1}(x))) \tag{2.2}$$

where  $F(x)$  is the flow of the system. This means that there exists a diffeomorphism  $\Phi$  from  $M$  to  $\mathbb{R}^m$  that preserves the topological properties of the system in this higher-dimensional space.

As long as a proper time delay  $\tau$  and embedding dimension  $m$  are chosen, this theorem can considerably simplify the analysis of complex systems when we do not have access to the whole phase space but only to scalar measurements of the system.

### 2.1.2 Choice of Delay and Embedding Dimension

The choice of parameters  $\tau$  (time delay) and  $m$  (embedding dimension) is crucial to ensure a correct and meaningful reconstruction of the phase space through delay coordinate embedding. To achieve optimal results, different methods were employed such as **mutual information** for determining the time delay and the **false nearest neighbors algorithm** for selecting the embedding dimension.

However, while these methods provide a valuable guideline for selecting the embedding parameters, they may suggest a suboptimal parameter choice when delay embedding is followed by dimensionality reduction steps.

#### Time Delay through Mutual Information

The time delay  $\tau$  is the parameter that determines the temporal distance between consecutive observations in the delay sequence. Its selection can be guided by the **mutual information** ( $I(X; Y)$ ) between successive observations of the system. Mutual information is a measure of statistical dependence between two random variables  $X$  and  $Y$ , and its key advantage is its ability to capture not only linear but also non-linear correlations between variables, which are typical in the systems under study.

For a time series, mutual information is evaluated between pairs of points  $x(t)$  and  $x(t + \tau)$ , where  $x(t)$  represents the observable at time  $t$ . This evaluation is repeated for different values of  $\tau$ . The goal is to identify values of  $\tau$  that minimize the mutual information between delayed coordinates. This ensures that the offsets in the delay vector are as independent as possible, enabling the acquisition of complementary information from each delayed observation. To evaluate the mutual information we used equation:

$$I(\tau) = \sum_{h=1}^{n_{\text{bins}}} \sum_{k=1}^{n_{\text{bins}}} p_{XY}^{(\tau)}(h, k) \ln \left( \frac{p_{XY}^{(\tau)}(h, k)}{p_X(h) p_Y(k)} \right) \quad (2.3)$$

where:

- $X = x(t)$  and  $Y = x(t + \tau)$ .
- $p_X(h)$  is the probability that  $x(t)$  falls in bin  $h$ .
- $p_Y(k)$  is the probability that  $x(t + \tau)$  falls in bin  $k$ .
- $p_{XY}^{(\tau)}(h, k)$  is the joint probability that  $x(t)$  falls in bin  $h$  and  $x(t + \tau)$  falls in bin  $k$ .

#### Embedding Dimension and False Nearest Neighbors

The embedding dimension is a crucial parameter in the reconstruction of the system under consideration. According to Takens' theorem, it is possible to faithfully reconstruct the system's dynamics as long as the embedding dimension  $m$  is sufficiently large to preserve the system's topological properties. If the embedding dimension is underestimated, the reconstructed trajectories become "flattened", leading to a loss of critical information about the system. In such cases, some points in the reconstructed space appear as **false neighbors**—pairs of points that seem close due to the insufficient dimensionality but are actually distant in the true phase space. To address this issue, we employ the **False Nearest Neighbors (FNN)** method discussed in Reference [3] [4], which quantifies the fraction of false neighbors as the embedding dimension increases for a fixed time delay. By analyzing

how the fraction of false neighbors decreases with increasing embedding dimension, we can determine the embedding dimension that keeps these artifacts minimal, ensuring an accurate reconstruction. To evaluate the fraction of false neighbors we used the equation:

$$F = \frac{1}{N} \sum_{i=1}^N \mathbf{1} \left( \frac{|x_{i+m\tau} - x_{j(i)+m\tau}|}{\|\mathbf{X}_i - \mathbf{X}_{j(i)}\|} > R_{\text{th}} \right) \quad (2.4)$$

where:

- $\mathbf{X}_i$  is the  $m$ -dimensional delay-embedded vector at time  $i$ .
- $j(i)$  is the index of the nearest neighbor of  $\mathbf{X}_i$  in embedding space.
- $R_{\text{th}}$  is the threshold ratio.
- $\mathbf{1}(\cdot)$  is the indicator function (1 if its argument is true, 0 otherwise).

We have to note that the FNN algorithm and Takens' theorem provide us just with a lower limit for the embedding dimension, and not actually a value corresponding to the optimal choice. This point will become clear when discussing the effect of dimensionality reduction on the embedded data.

## 2.2 Dimensionality reduction

The delay coordinates embedding provides a representation of our manifold of dimension at least  $2d+1$ . However, in practice, the embedding dimensionality is often significantly higher than this lower bound. While this is not a theoretical limitation, practically handling high-dimensional data can lead to computational challenges requiring excessive processing power and time. Therefore, even if the optimal embedding parameters can be determined, we will apply dimensionality reduction techniques to compress our data and show that they can indeed bring benefit to our analysis.

### 2.2.1 Discrete Legendre polynomials

The first step of dimensionality reduction is performed using discrete Legendre polynomials. In the *regime of short delay windows* [2], this approach provides an approximation of *Principal Component Analysis (PCA)* but it avoids the need to diagonalize a high-dimensional matrix. Moreover, it provides a good approximation of the eigenfunctions of the Koopman operator  $K$ , yielding a universal analytic basis to represent the dynamics across different datasets [2]. To effectively leverage this approach, it is crucial to work within the limit of short delay windows. In practical terms, this requires that the time span covered by an embedding vector  $\tau_w = \tau \cdot (m - 1)$  remains smaller than a critical threshold given by:

$$\tau_w^* = 2\sqrt{\frac{3K_0}{K_1}}$$

where  $K_0 = \langle x^2(t) \rangle$ ,  $K_1 = \langle (x'(t))^2 \rangle$  with  $x(t)$  representing the observed time series. The condition  $\tau_w < \tau_w^*$  imposes a constraint on the choice of  $\tau$  and  $m$ . Specifically, for a given embedding dimension satisfying the FNN and Takens' Theorem conditions it is necessary to verify whether the delay suggested by Mutual information remains within the short delay windows limit. It is important to note that an agreement between these methods is not guaranteed, making the selection of an appropriate

delay non-trivial for certain systems. If the short delay condition is not met, then there is no real convenience in using this method of dimensionality reduction instead of classical PCA as the eigenvectors of the covariance matrix will not be well approximated by the discrete Legendre polynomials.

## 2.2.2 Variational Autoencoders

The last step in our framework is dimensionality reduction through neural networks. A **Variational autoencoder** (VAE) [5] is a deep generative model that combines unsupervised learning with a probabilistic representation of the data. Its goal is to learn a latent distribution of the input data that enables the generation of new coherent observations from the same distribution. VAE's are composed of two paired networks: *encoder* (or *recognition model*) and *decoder* (or *generative model*). The *encoder* learns stochastic mappings from the observation space to a low-dimensional latent space, where the latent variables  $z$  are assumed to follow a predefined *prior distribution*  $p_\theta(z)$  in the *generative model*. This is achieved by an *approximate posterior distribution*  $q_\phi(z|x)$  optimized to approximate the *true posterior*  $p_\theta(z|x)$ . On the other hand, the decoder learns to map the latent space back to the observation space by modeling the *likelihood*  $p_\theta(x|z)$  allowing the generation of new coherent observations from the latent distribution. We refer to  $\phi$  and  $\theta$  respectively as the encoder model parameters and the decoder model parameters. A key problem in VAEs training is the intractability of the *posterior distribution*  $p_\theta(z|x)$  while maximizing the marginal likelihood. To avoid this intractability the optimization objective is the *evidence lower bound* (**ELBO**):

$$\mathcal{L}_{\theta,\phi}(x) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x, z) - \log q_\phi(z|x)]$$

which satisfies the inequality:

$$\mathcal{L}_{\theta,\phi}(x) = \log p_\theta(x) - D_{KL}(q_\phi(z|x) \parallel p_\theta(z|x)) \leq \log p_\theta(x)$$

where the term  $D_{KL}(q_\phi(z|x) \parallel p_\theta(z|x))$  is the Kullback-Leibler divergence between  $q_\phi(z|x)$  and  $p_\theta(z|x)$  and is non-negative quantity. Maximizing the **ELBO** will approximately maximize the marginal likelihood  $p_\theta(x)$  and minimize the distance of  $q_\phi(z|x)$  from the true *posterior*  $p_\theta(z|x)$ . However, even though **ELBO** allows joint optimization with respect to all parameters through *stochastic gradient descent*, it is difficult to compute unbiased gradients for the *encoder model* parameters  $\phi$  since the gradient does not commute with the expectation in this case:

$$\nabla_\phi \mathcal{L}_{\theta,\phi}(x) = \nabla_\phi \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x, z) - \log q_\phi(z|x)] \neq \mathbb{E}_{q_\phi(z|x)} [\nabla_\phi (\log p_\theta(x, z) - \log q_\phi(z|x))] \quad (2.5)$$

In case of continuous latent variables, it is possible to use the **reparameterization trick** that consists of a change in variables from  $z \sim q_\phi(z|x)$  to  $z = g(\epsilon, \phi, x)$  where  $g$  is a differentiable transformation of another random variable  $\epsilon$  sampled from a certain distribution  $p(\epsilon)$ . As a result, expectations can be rewritten as:

$$\mathbb{E}_{q_\phi(z|x)} [f(z)] = \mathbb{E}_{p(\epsilon)} [f(z)]$$

and the gradients using the approximation  $\epsilon \sim p(\epsilon)$ :

$$\nabla_\phi \mathbb{E}_{q_\phi(z|x)} [f(z)] \simeq \nabla_\phi f(z)$$

The reparameterization trick allows efficient gradient computation by rewriting the latent variables as:

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

$$(\mu, \log \sigma) = \text{Encoder}_\phi(x)$$

In the end, the state space mapping of a sample  $h$  from the observation space through the network can be resumed as:

$$h \xrightarrow{\text{Encoder}} z \xrightarrow{\text{Decoder}} y$$

where  $y$  is another sample from the observable space. Thus, the VAE in this framework serves not only as a dimensionality reduction tool but also as a model for learning a mapping between pupil dynamics and brain dynamics.

## 2.3 Outline of the Methods

This framework was applied to both the stochastic Lorenz system and the Zenodo database. The input consists of a time series  $y_1(t)$  sampled at  $p$  discrete time points and our outputs corresponds to another vector-valued time series  $y_2(t)$ , which is an observable from the same phase space.

First we preprocess data to determine the optimal embedding parameters following the guideline in Sections 2.1.2 and 2.2.1. Once the embedding parameters are selected, we construct the embedding matrix  $H_i$  for all the observables in the dataset:

$$H_i^T(t) = \begin{bmatrix} y_i(t_1) & y_i(t_1 + \tau) & y_i(t_1 + 2\tau) & \cdots & y_i(t_1 + (d-1)\tau) \\ y_i(t_2) & y_i(t_2 + \tau) & y_i(t_2 + 2\tau) & \cdots & y_i(t_2 + (d-1)\tau) \\ y_i(t_3) & y_i(t_3 + \tau) & y_i(t_3 + 2\tau) & \cdots & y_i(t_3 + (d-1)\tau) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_i(t_{p-d+1}) & y_i(t_{p-d+1} + \tau) & y_i(t_{p-d+1} + 2\tau) & \cdots & y_i(t_{p-d+1} + (d-1)\tau) \end{bmatrix} \in \mathbb{R}^{d \times (p-d+1)}$$

where each row of  $H_i^T$  represent a delay vector  $h_i(t)$  associated with the  $i$ -th observable. Next, we select a dimension  $r$  for the Legendre basis and construct the corresponding Legendre basis matrix:

$$P^{(r)} = \begin{bmatrix} P_0(x_0) & P_1(x_0) & \cdots & P_{r-1}(x_0) \\ P_0(x_1) & P_1(x_1) & \cdots & P_{r-1}(x_1) \\ P_0(x_2) & P_1(x_2) & \cdots & P_{r-1}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_{d-1}) & P_1(x_{d-1}) & \cdots & P_{r-1}(x_{d-1}) \end{bmatrix} \in \mathbb{R}^{d \times r}$$

where each column corresponds to a Legendre polynomial evaluated on a equally spaced grid of  $d$  points, with  $x_i \in [-1, 1]$ . The embedding matrix  $H_i$  is then projected onto the Legendre basis using the transformation:

$$Y_i^T = H_i \cdot P^{(r)} \in \mathbb{R}^{(p-d+1) \times r}$$

The rows of the projected matrix  $Y_i^T$  serve as inputs to the neural network. The input vectors  $\hat{y}_i^T = h_i P^{(r)}$  are normalized prior to dimensionality reduction and their  $r$ -dimensional shape must match the dimension of the first layer of the *encoder model*, while the targets are  $r \cdot n$ -dimensional,

where  $n$  is the dimensionality of the vector-valued observable  $y_2$ , and must match the dimension of the last layer of the *decoder model*. An initial **VAE** is used to reconstruct the input using a *dummy decoder*, enabling the model to learn an early representation of the latent space. This step prevents the final model from excessively optimizing the decoder at expense of the latent representation. Then the pre-trained encoder is used to build a VAE designed to reconstruct the desired observable, leveraging the previous learned latent space representation. To train the model we minimized the negative **ELBO**:

$$\mathcal{L}_{\theta,\phi}(y) = -\mathbb{E}_{q_{\phi}(z|y)}[\log p_{\theta}(y|z)] + \gamma \cdot D_{KL}(q_{\phi}(z|y) \parallel p_{\theta}(z))$$

where the  $D_{KL}$  is weighted by the hyperparameter  $\gamma$ , which is gradually annealed to 0.1.

# Chapter 3

## Results

Here we will describe the analysis results on the two systems previously mentioned in the Introduction.

### 3.1 Lorenz Toy

We simulated the three-dimensional *stochastic Lorenz system* using the following stochastic differential equations:

$$\begin{aligned} dz_1 &= 10(z_2 - z_1) dt + \alpha dW_1, \\ dz_2 &= [z_1(28 - z_3) - z_2] dt + \alpha dW_2, \\ dz_3 &= [z_1 z_2 - \frac{8}{3} z_3] dt + \alpha dW_3. \end{aligned}$$

where the intrinsic noise is modeled as independent **Wiener processes**, scaled by the noise parameter  $\alpha$ . The system observables are then defined as:

$$\begin{aligned} y_1 &= z_2 + \beta \sigma_1, & \sigma_1 &\sim \mathcal{N}(0, 1). \\ y_2 &= [z_1, z_3] + \beta \sigma_2, & \sigma_2 &\sim \mathcal{N}(0, 1). \end{aligned}$$

These correspond to a scalar and a vector-valued observables derived from the simulated system, with added observation noise modeled as a standard Gaussian and scaled by the parameter  $\beta$ . We aim to reconstruct the vector-valued observable  $y_2$  from the scalar observable  $y_1$  using the proposed framework.

#### 3.1.1 Preprocessing

Using the **FNN** algorithm and the **Mutual Information**, we obtained a threshold of 4 as embedding dimension and a suggested delay of 15 time steps, corresponding to the first local minimum of the graph, as can be seen in Figure 3.3. We then computed the critical time window  $\tau_w^* = 0.47s$  and selected an appropriate ratio between  $d$  and  $\tau$  to respect the short delay window constraint. Following a *posteriori* analysis of reconstruction performance, we decided to set the embedding dimension to  $d = 30$ , as it yielded the best results, and trained models for two different delays:  $\tau = 15$ , as suggested

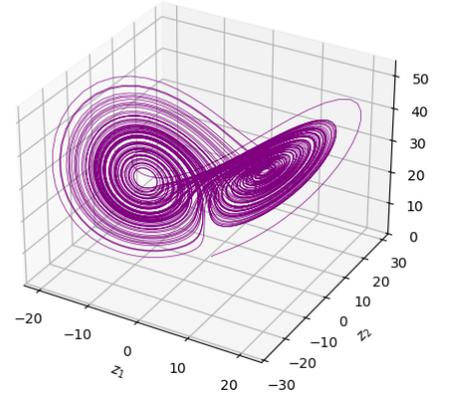


Figure 3.1: Simulated trajectories for noise parameters:  $(\alpha = 0, \beta = 0)$

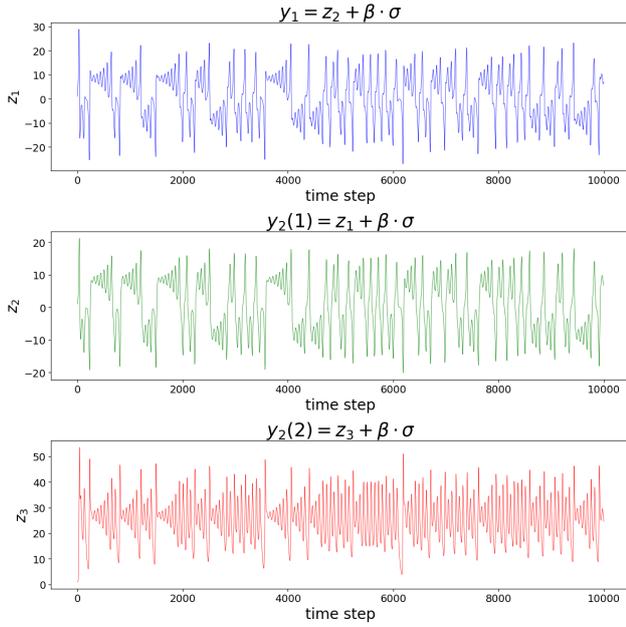


Figure 3.2: Observables  $y_1(t)$  and  $y_2(t)$  time series obtained from simulated trajectories

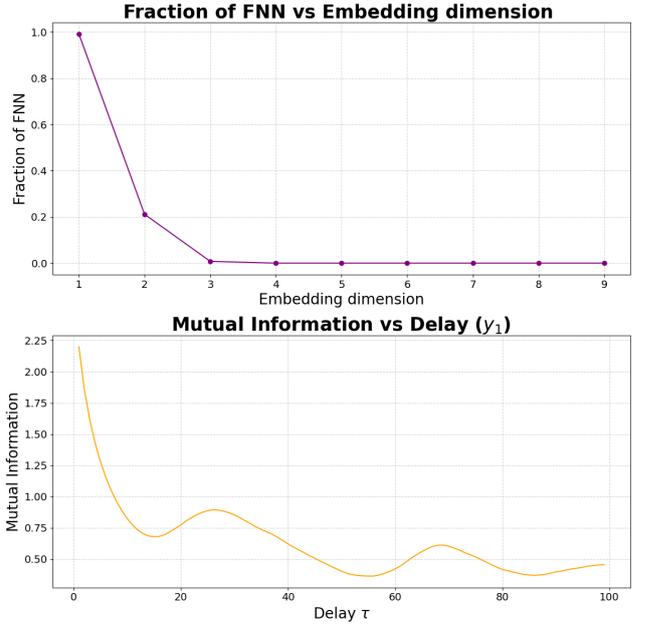


Figure 3.3: Graph corresponding to the fraction of **FNN** in function of the embedding dimension  $d$  and the Mutual information in function of delay  $\tau$  for the observable  $y_1$

Figure 3.4: Data properties for  $(\alpha = 0, \beta = 0)$

by **MI** and  $\tau = 3$  ( $\tau_w \simeq 0.80s$ ), which was chosen based on the **short delay window constraint** to ensure operation in the *moderate regime*. Finally, we set the Legendre basis dimension to  $r = 20$  and fed the transformed data into the networks.

### 3.1.2 Encoder training and performance

We trained the encoder to reconstruct its input for both selected values of  $\tau$ . The VAE was implemented as a feed-forward network with a fixed latent dimension of  $m = 5$  and two hidden linear layers in both the encoder and decoder, using **ReLU** activation function everywhere except for the last layer of the *decoder model* where a **sigmoid** activation was applied to match data normalization. For training, we used the Adam Optimizer with initial *learning rate* of 0.02, a **ReduceLRonPlateau** scheduler, and a batch size of 32. During training, the parameter  $\gamma$  followed a sigmoid schedule, gradually annealing to 0.01 over the first 250 epochs out of the total 300. To evaluate the network's performance, we tested its ability to reconstruct the principal components (PCs) of the original data embeddings, as well as the original time series, using the inverse transformation:

$$H_i^T = Y_i^T \cdot P^\dagger \quad (3.1)$$

where  $P^\dagger$  denotes the pseudo-inverse of  $P$ . For a delay of  $\tau = 3$ , the results showed an excellent ability to reconstruct the first PCs of the original data, as well as the original time series, even for very high noise level as shown in Figure 3.5.

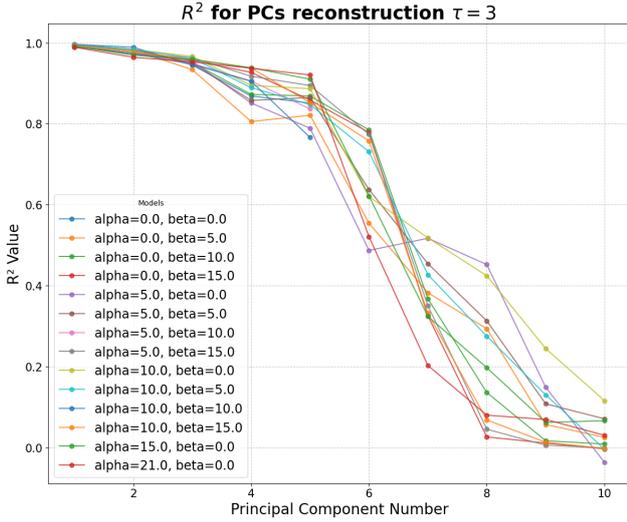


Figure 3.5:  $y_1$  PCs reconstruction via neural network.  $R^2$  is evaluated between the PCs of the original embeddings and the projection of network outputs onto their directions

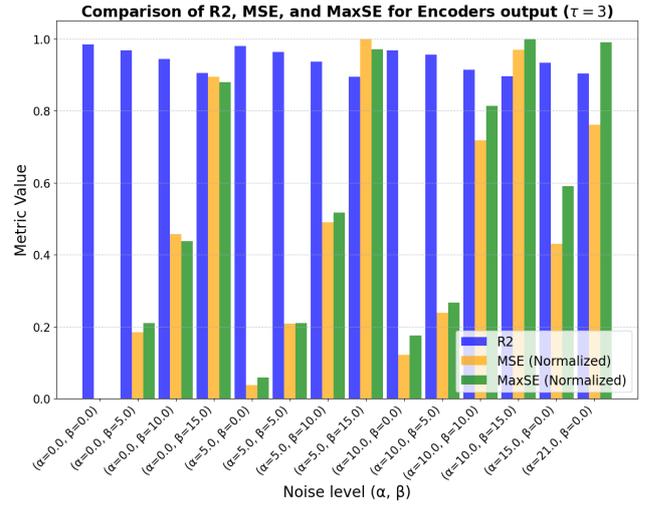


Figure 3.6: Reconstruction metrics evaluated between network outputs and targets. **MSE** and **maxSE** are normalized in range  $[0, 1]$

Figure 3.7: Encoder performance with increasing level of noise for  $\tau = 3$

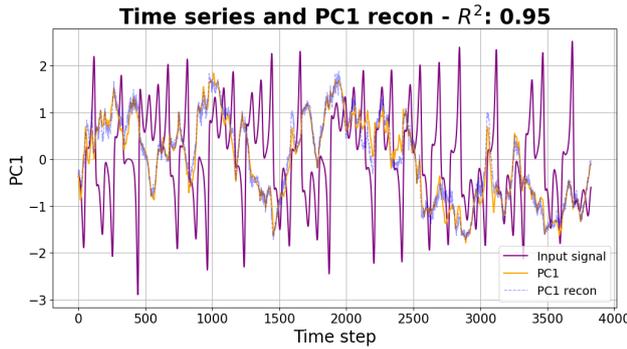


Figure 3.8:  $y_1$ 's PC1 and PC1 reconstruction via neural network compared

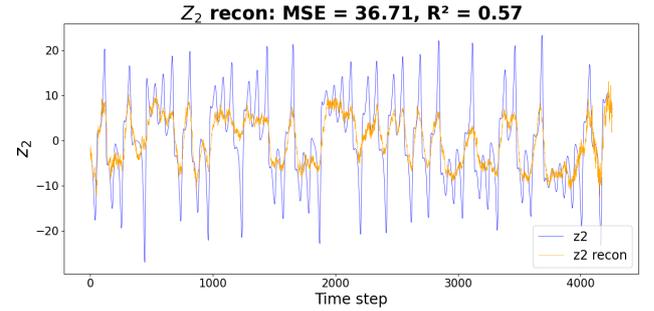


Figure 3.9:  $y_1$  time series reconstruction using inverse transformation in Equation 3.1

Figure 3.10: Encoder performance for  $\tau = 15$

The encoder training for  $\tau = 15$  exhibited a completely different behavior. While the reconstruction of the first PCs did not show any anomalies, the time series displayed a pathological reconstruction of the original data, likely due to excessive information loss during the projection onto the Legendre basis. This loss made the inverse transformation unable to fully recover the original data. In this case we did not investigate the performance under increasing noise level and will report just the best result achieved, shown in Figure 3.10.

### 3.1.3 Decoder training and performance

We trained models to take as input a dimensionally reduced delay vector of the observable  $y_1$ , corresponding to a column of the matrix  $Y_1$ , and as target the corresponding vector obtained by applying the same transformation to the components of the observable  $y_2$ , which are the columns of the matrices  $Y_2$  and  $Y_3$ . A shared decoder was used to simultaneously reconstruct both observables. In this case, due to the previous regularization of the latent space, the contribution of  $D_{KL}$  to the total training

loss was negligible. Consequently, we decided to freeze the encoder and latent space parameters during training and instead use a **MSE loss** as the objective function. To evaluate the network's performance, we assessed its ability to reconstruct the principal components of the original embeddings and computed the  $R^2$  score between the reconstructed and original time series. For models trained with a delay of  $\tau = 3$ , we observed a slight systematic effect, where the reconstruction of  $z_1$  dynamics was marginally better than that of  $z_3$ .

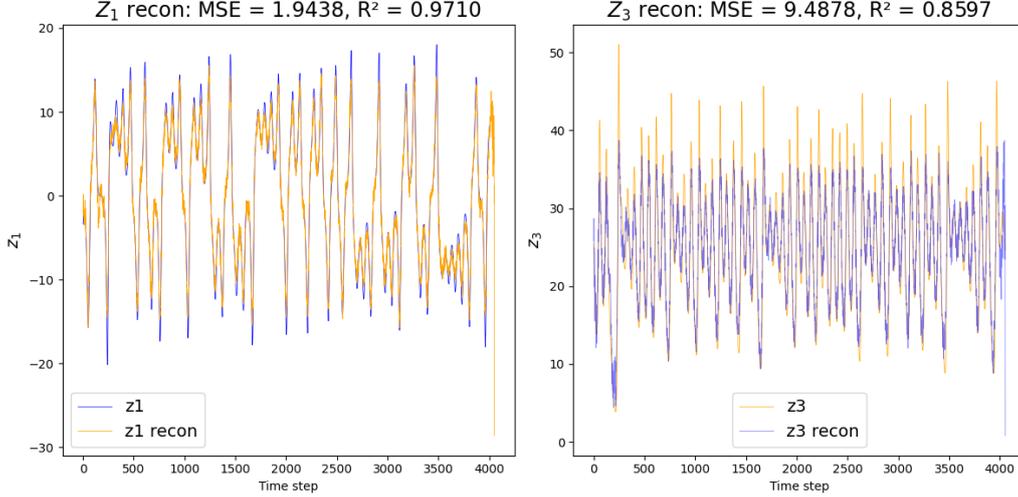


Figure 3.11: Example of  $y_2(t)$  time series reconstruction for  $\tau = 3$  and noise parameters ( $\alpha = 0$ ,  $\beta = 0$ )

Delay vector reconstruction yields very good results for both  $z_1$  and  $z_3$  at low noise levels, with performance degrading as noise increases. While the reconstruction quality of  $z_1$  delay vectors remained acceptable even at high noise levels, the reconstruction of  $z_3$  delay vectors deteriorated rapidly (Figure 3.12), in some cases showing no correlation with the original data.

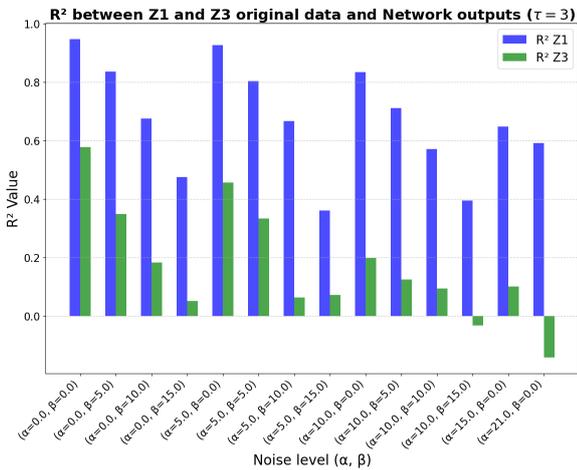


Figure 3.12:  $R^2$  metric evaluated between network's outputs and targets for  $y_2(t)$  reconstruction

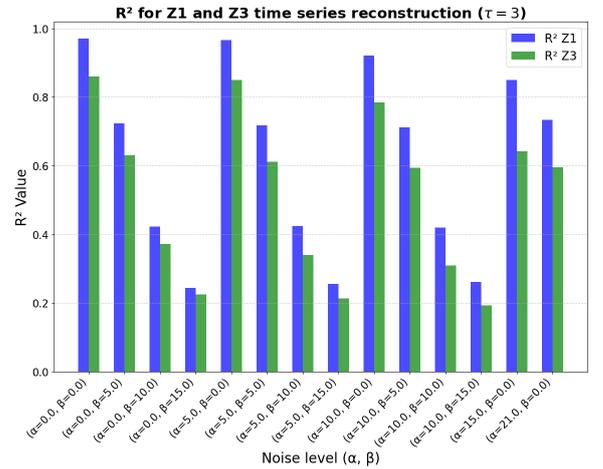


Figure 3.13:  $R^2$  metric evaluated between original time series and reverse transformed network outputs for  $y_2(t)$  time series reconstruction

Figure 3.14: Decoder performance with increasing level of noise for  $\tau = 3$

However, the applying of inverse transformation from equation 3.1 to unfold the embedding and reconstruct the time series for comparison with the original data, demonstrated that the network is capable of recovering the dynamics both for  $z_1$  and  $z_3$ , ultimately yielding acceptable results. A more detailed analysis was conducted on the reconstruction of the PCs of the original time series.

The first six principal components were computed from original embeddings, and the reconstructed embeddings were projected onto these directions using the `fit.transform` function from `sklearn`. Subsequently, the  $R^2$  indices between original and reconstructed PCs were evaluated. Results indicate that the first PCs of  $z_1$  are consistently well reconstructed, whereas for  $z_3$ , the PCs reconstruction exhibits anomalous behaviour, as portrayed in figure 3.16.

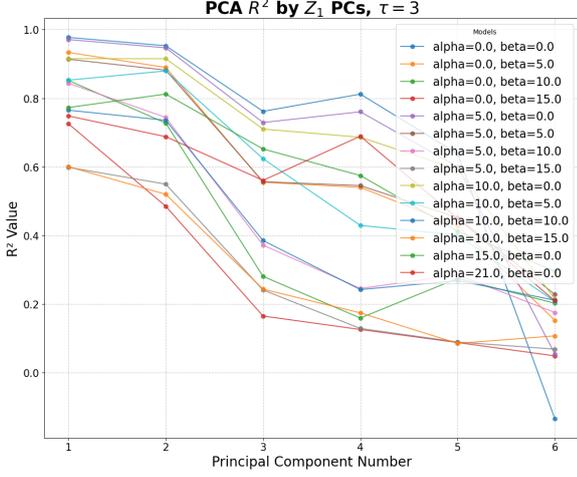


Figure 3.15:  $z_1$  PCs reconstruction via neural network.  $R^2$  is evaluated between the PCs of the original embeddings and the projection of network outputs onto their directions

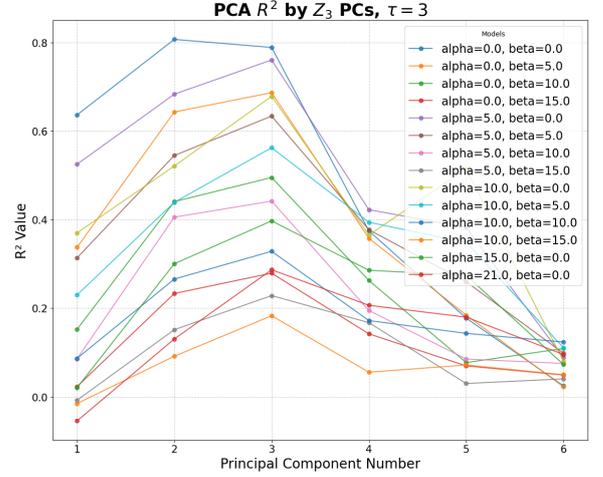


Figure 3.16:  $z_3$  PCs reconstruction performance

Figure 3.17: Decoder ability to reconstruct PCs of original  $y_2$  data for  $\tau = 3$

The first notable observation from figure 3.16 is that the first PC is systematically reconstructed worse than the subsequent two. Additionally, the overall reconstruction quality is lower for  $z_3$  compared to  $z_1$ . To possibly mitigate this effect, we adjusted the loss function by assigning a higher weight to  $z_3$  compared to  $z_1$ . However, this method had no effect on the reconstruction quality. We also attempted to unfreeze the encoder and latent space parameters to refine the latent representation of the system but this approach yielded no improvement, as the  $D_{KL}$  term remained stuck on the previously reached plateau. Our hypothesis is that, although the embedding parameters correspond to a time window of the same order as the critical one, we are operating in the **transition regime**, where the eigenvalues of the covariance matrix deviate from the *small-windows solution* but still decrease exponentially with increasing orders. In other words, while Legendre polynomials provide a useful universal basis to study the dynamics across different datasets, they are less effective than **PCA** as a dimensionality reduction tool. This limitation restricts the network's ability to learn a low-dimensional latent representation of the system, ultimately reducing the reconstruction accuracy of one of its components.

Regarding the models trained with a delay of  $\tau = 15$ , we were unable to find any network parameters that allowed the model to learn a sufficiently meaningful latent representation of the entire phase space. Results indicate that while a delay of 15 is sufficient to achieve a good reconstruction of  $z_1$  time series, the network fails to capture any correlation for  $z_3$ , resulting in a complete lack of reconstruction ability (Figure 3.20). Moreover, the network's outputs for  $z_3$  are unstable under the pseudo-inverse matrix transformation, preventing the recovery of the time series from the reconstructed embeddings. These results are likely due to a strong dependence of  $z_3$  reconstruction on short-term correlations between consecutive points. While a delay of 3 time step can capture these dependencies, a 15 time step delay effectively ignores them. It is also important to note that for a delay of  $\tau = 15$  the corresponding

time window is  $\tau_w = 4.35s$ , which is significantly larger than the critical value  $\tau_w^* = 0.47s$ . This implies that dimensionality reduction through Legendre polynomials no longer corresponds to PCA. These results suggest that, for optimal performance within this framework, the selection of embedding parameters should prioritize adherence to the *small window constraint* rather than focusing solely on selecting a delay that ensures independence between successive coordinates of a delay vector.

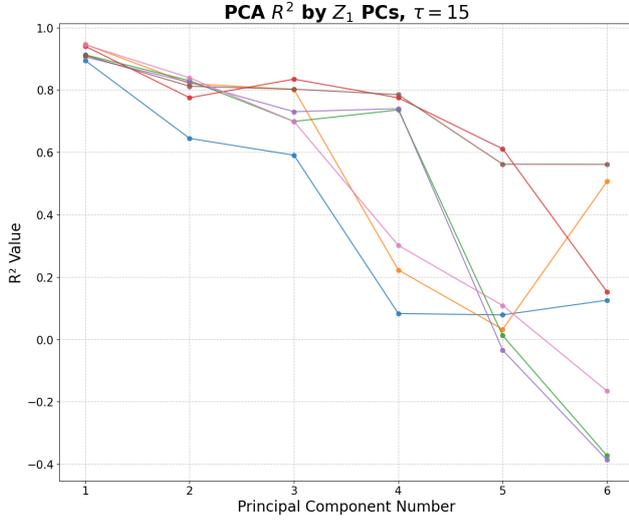


Figure 3.18:  $z_1$  PCs reconstruction via neural network.  $R^2$  is evaluated between the PCs of the original embeddings and the projection of network outputs onto their directions.

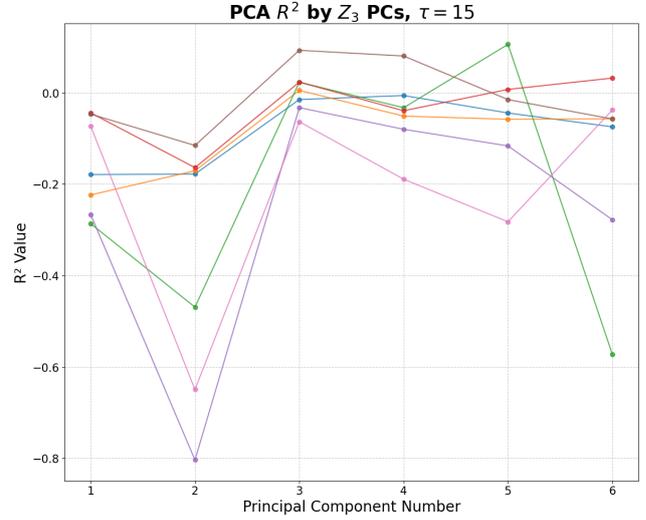


Figure 3.19:  $z_3$  PCs reconstruction.

Figure 3.20: Decoder performance for  $\tau = 15$  in absence of noise ( $\alpha = 0, \beta = 0$ ). Different lines in the plot represent neural networks trained with different configurations.

## 3.2 Zenodo brain data

This methodology has been tested on the publicly available data, accessible at link [zenodo.org], discussed in Reference [7]. The dataset consists of simultaneous rs-fMRI and pupillometry recordings, acquired using the Echo Planar Imaging (**EPI**) technique with a repetition time (**TR**) of 1 s for a total duration of 925 s. The dataset includes 74 trials conducted on 10 anesthetized rats using alpha-chloralose. The rs-fMRI file contains volumetric brain images with a voxel resolution of  $(56 \times 48 \times 32)$ , temporally aligned with the pupillometry measurements. Since the voxel-wise time evolution of each trial requires excessive computational power to train a network, we chose instead to reconstruct the time series of the first  $n$  spatial principal components from selected rs-fMRI data using the proposed framework, in analogy with the approach presented for the Lorenz Toy model.

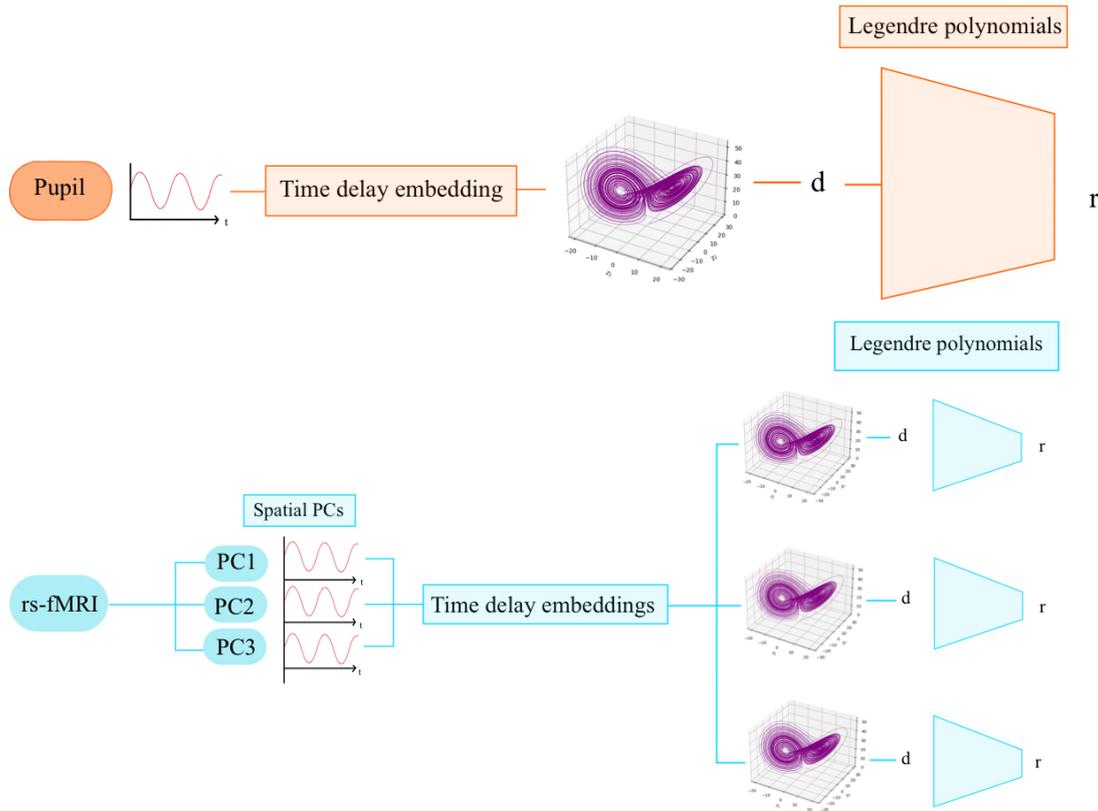


Figure 3.21: Dataset preparation workflow

### 3.2.1 Dataset analysis

For analysis purposes it is necessary to distinguish between different rats. Unfortunately, the dataset did not provide an identifier associating each trial with a specific rat. To overcome this limitation, we hypothesized that individuals would be recognizable from their unique brain shape. By using a simple threshold, we were able to discriminate between brain and non-brain voxels in the images, so that we could associate a brain volume estimate (number of brain voxels) to each image. Brain volume estimates obtained in different trials perfectly clustered in seven clusters (3.23), allowing us to identify seven animals. At first glance, most of the trials exhibit little or no correlation between fMRI and pupillometry recordings, but this may be simply due to a subject-specific physiological lag between the two signals. We attempted to shift the **PC1** time series to determine the temporal delay that would

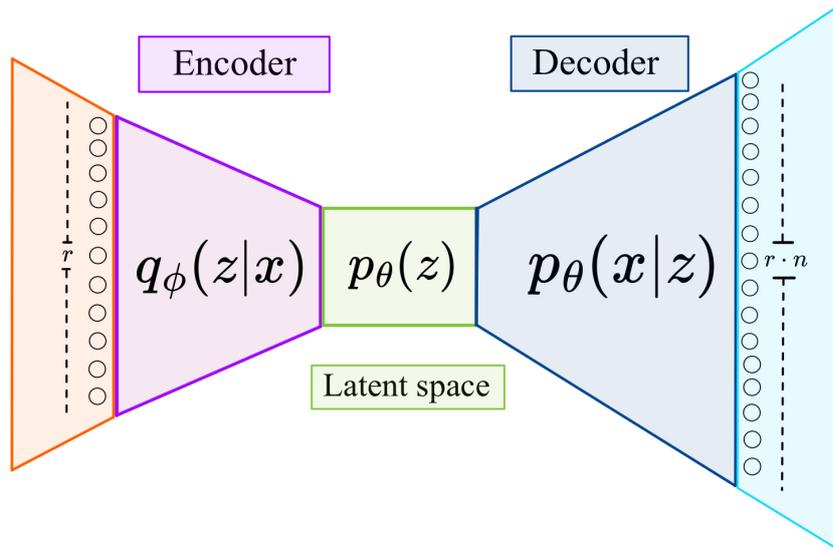


Figure 3.22: VAE's configuration

maximise the correlation. This approach led to significantly higher values in most datasets, with some correlations being very strong. However, in the majority of cases, the optimal delay was too long to be associated with arousal dynamics. In awake mice, the expected delay is on the scale of seconds, whereas in our dataset, the best correlation was often found at delays of approximately 20s or more, which is excessive even when considering the effects of sedation on rats. To account for the differences between the conditions in the provided dataset and those studied in Reference [6], we selected just the trials exhibiting the strongest correlation between fMRI PC1 and pupil dynamics, with a maximum delay of 15s. Among the 74 available trials, the selected ones are (2,3,4), which correspond to rat 1, and (16,17), which correspond to rat 2. Since only a limited subset of trials exhibits such strong correlation, identifying the animals corresponding to these specific trials is sufficient for our purposes, without requiring full identification of all ten subjects.

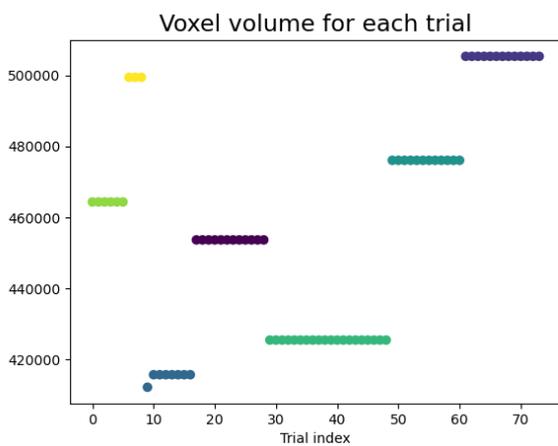


Figure 3.23: Brain volume for each trial

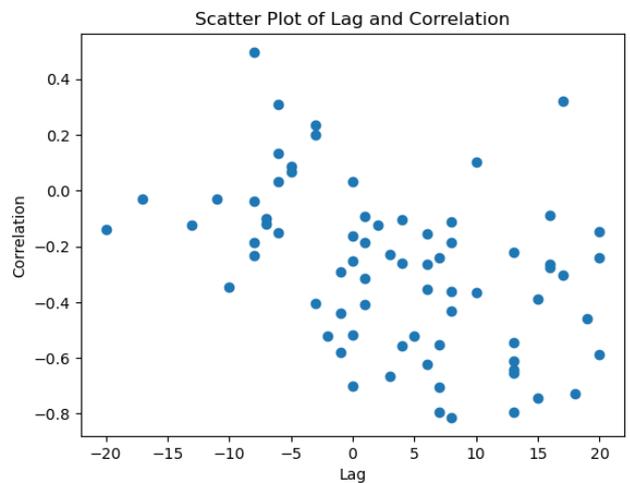


Figure 3.24: Physiological lag and correlation between rs-fMRI (PC1) and pupil

### 3.2.2 Preprocessing

Following the procedure outlined in Reference [6], the rs-fMRI data from the selected trials were detrended and spatially smoothed using Gaussian smoothing with  $\sigma = 0.7$ . To distinguish the brain from the background, we generated a voxel-wise brain mask by excluding from analysis all voxels with a temporal standard deviation below a fixed threshold. The indices associated with each selected voxel and its corresponding time series were stored to compress the data, reducing its shape from (56,48,32,925) to (21680,925). Here 21680 represents the number of active voxels per trial, as the fMRI volumes were affine-transformed to be aligned with the Paxinos atlas. In this work, we will discuss only the results obtained for the first rat. Results of the second rat were analogous, with a similar network performance. For each trial, we computed the first three principal components (PCs) of the fMRI time series and applied a Butterworth filter from `scipy.signal` to isolate the frequency bands associated with arousal, both in the fMRI PCs and pupil signal. Since the dataset consists of anesthetized rats we selected lower frequency ranges than those considered in Reference [6], as the arousal pattern unfolds over longer time windows, as evidenced by the systematic increase of the physiological delay. We choose the range from 0.01 Hz to 0.05 Hz instead of the range 0.01 Hz-0.2 Hz studied in Reference [6]. The data was then normalized using a z-score normalization, and the pupil and fMRI PCs time series were aligned based on the previously computed optimal delay. Although the dataset refers to the same rat, different trials exhibit slightly different physiological delay. To construct a symmetric dataset with an equal number of time samples across trials, we removed a number of samples from each dataset corresponding to the maximum delay. For each time series, we computed the critical time window  $\tau_w^*$  and used it to determine the optimal embedding parameters.

### 3.2.3 Network performance

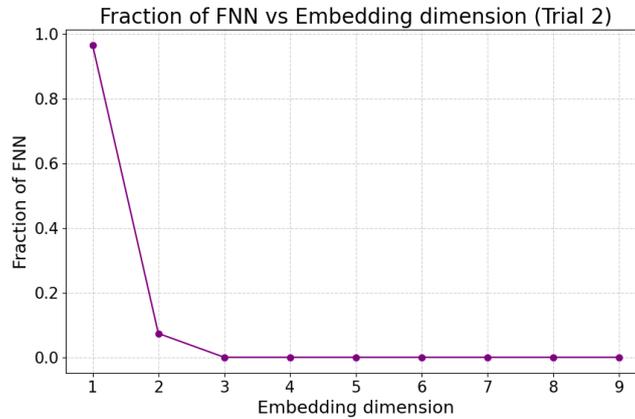


Figure 3.25: **FNN algorithm** performed on Trial 2 for a delay  $\tau = 3$ . All selected trials had the same minimum embedding dimension.

Following the proposed framework we trained a neural network to reconstruct the dimensionally reduced delay vector, obtained from the first  $n$  PCs of the fMRI time series, using as input the dimensionally reduced delay vector obtained from the pupil time series. The number of selected PCs was determined *a posteriori* through an analysis of the network’s performance, while the embedding parameters and delay were chosen according to the provided guidelines. The critical time window  $\tau_w^*$  of the selected time series was approximately 50s for each trial. Based on this, we initially selected an embedding dimension of  $d = 20$  and a delay of  $\tau = 2$ , resulting in a time window of 38s which is significantly smaller than the critical value. For the Legendre basis dimension, we selected  $r = 15$

and set the latent space dimension of the VAE to  $m = 10$ , taking into account that **FNN** algorithm suggests an embedding dimension higher than 2 (Figure 3.25). The dimension of Legendre basis and latent space were determined *a posteriori* based on the values that yielded optimal performance. The Network was a feedforward Variational Autoencoder with a layer structure  $[r, m, r \cdot n]$  with **tanh** activation function for each layer, except for the last layer of the decoder, which had no activation function.

### Dataset [2,3,4]

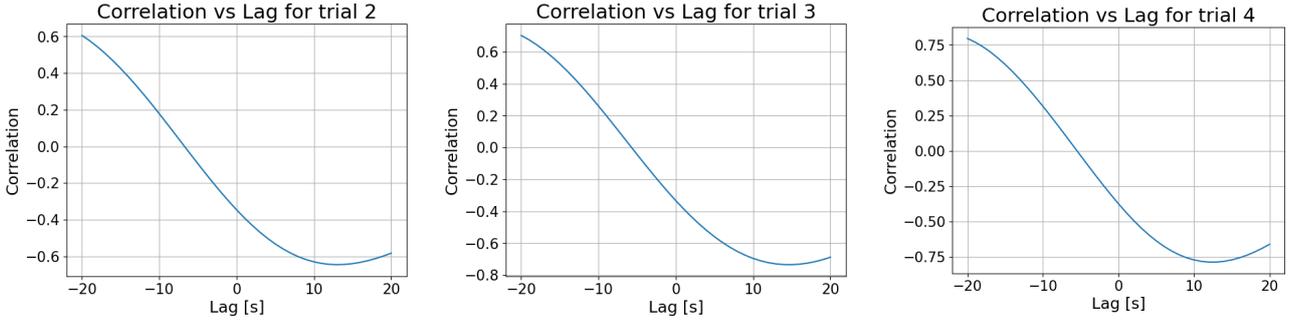


Figure 3.26: Cross-correlation between the first principal component of rs-fMRI and pupil size as a function of lag for trials referring to the first rat

The following trials exhibited a strong anti-correlation between PC1 and the pupil signal, with correlation values of  $-0.65$ ,  $-0.74$  and  $-0.79$ , and corresponding lags of 13s, 15s and 12s, respectively (Figures 3.26, 3.27). These values are still consistent with arousal-related dynamics.

We selected  $n = 3$  spatial principal components to be reconstructed by the Network and trained a VAE with a layer configuration of  $[15,10,45]$ . A sigmoid annealing schedule was applied to gradually set the  $\gamma$  parameter to 0.1 over 500 epochs to mitigate posterior collapse [6]. The model was trained using a constant learning rate of  $10^{-4}$  for 1000 epochs with a batch size of 32. After lag correction, each trial consisted of 910 time steps, allowing us to construct 870 delay vectors. For each trial, the first 70% of the data was used for training, while the remaining 30% was equally split into validation and test sets. Practically, the first 647s of the original time series were used for training, and the model performance was evaluated on the last 168s.

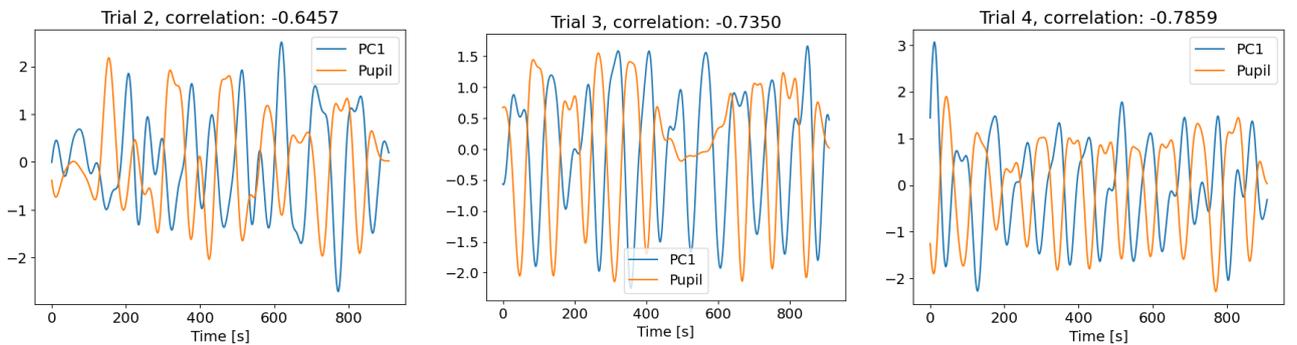


Figure 3.27: Time courses of pupil size and PC1 for each trial after re-alignment to remove the physiological lag

As shown in Figure 3.29, the loss function is successfully optimized through backpropagation, even

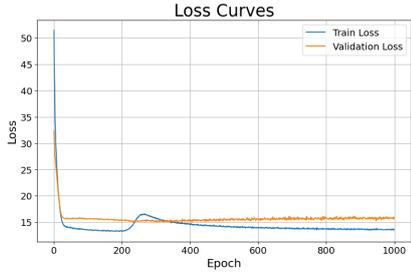


Figure 3.28: Training and Validation loss through Epochs

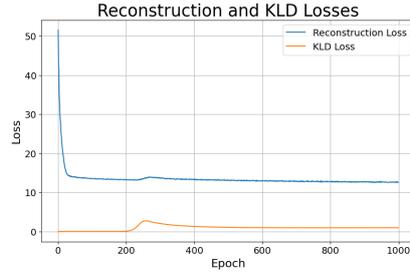
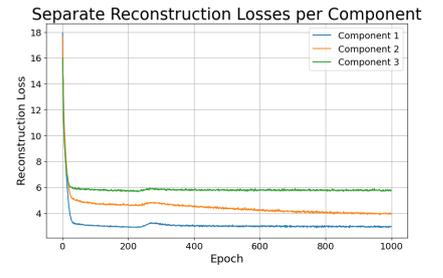
Figure 3.29: Reconstruction Loss and  $D_{KL}$  loss during training

Figure 3.30: Reconstruction loss for the first three PCs of rs-fMRI

Figure 3.31: Training behavior of VAE model

when the weight of the  $D_{KL}$  reaches its maximum value following the sigmoid annealing schedule for the parameter  $\gamma$  in epoch 250. Moreover, performance in the validation dataset during training did not show signs of overfitting 3.28.

It is important to note that performance degrades as the number of PCs increases 3.30. However, this does not pose a significant issue, as higher-order PCs explain only minor fraction of the total variance (Table 3.1). Including up to PC3 proved beneficial for capturing secondary temporal and spatial patterns that reinforce the primary dynamics (PC1). Consequently the **VAE** effectively captures the dominant signal subspace, although relying on only three spatial PCs inherently leaves out part of the variance not represented in those components. As a result, fine-scale or localized structures beyond the main patterns cannot be fully recovered by the PCA inverse transform. This partial reconstruction nonetheless suffices for our focus on the main spatiotemporal features, as higher-order PCs account for negligible variance. To quantify model performance, we evaluated the  $R^2$  score between the original time series and the VAE predictions, using the inverse transformation as described in Equation 3.1. The results are reported in Table 3.1, and the reconstruction of **PC1** is shown in Figure 3.32.

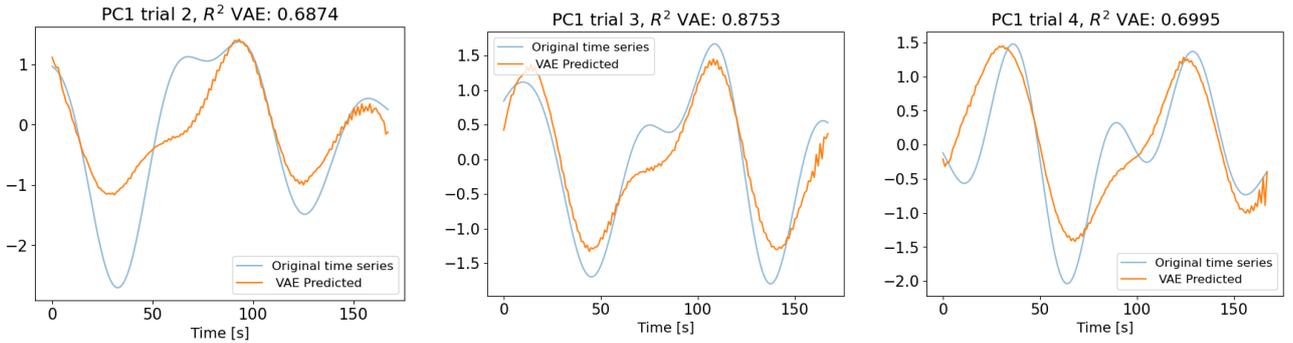


Figure 3.32: PC1 reconstruction via neural network

Table 3.1: Cumulative explained variance by PCs and  $R^2$  scores for PCs reconstruction across different trials.

Trial	Explained Variance			$R^2$ Score		
	PC1	PC2	PC3	PC1	PC2	PC3
2	0.51	0.58	0.62	0.69	0.31	0.17
3	0.51	0.57	0.62	0.88	0.44	-0.67
4	0.47	0.53	0.58	0.70	0.68	0.57

The results indicate that the VAE is capable of learning a latent representation that explains between 60% and 80% of the variance in **PC1**. However, since **PC1** itself accounts for only 47% to 51% of

the variance in the original data, the overall fraction of variance explained in the original dataset is inevitably lower. This limitation is intrinsic to our approach, as we opted to reconstruct only the first few PCs to mitigate the computational challenges posed by the dataset’s high dimensionality. Nonetheless, one could apply the same framework directly to all active voxel time series, thereby leveraging the dataset’s complete information for potentially improved performance. Despite this constraint, our findings highlight the flexibility of this approach and its potential for making predictions across a broader range of observables. In the specific case analyzed, it is important to highlight a key difference from Reference [6], arising from the different sampling frequencies: 20 Hz versus 1 Hz. The large difference in sampling rates directly influenced the time window associated with each delay vector, making it impossible to use higher embedding dimensions or delays without violating the constraint imposed by the critical window value. Moreover, we applied this framework on a dataset consisting of only 1830 samples for training and 390 samples each for validation and testing. By slowing down dynamics, sedation further aggravated the latter problem.

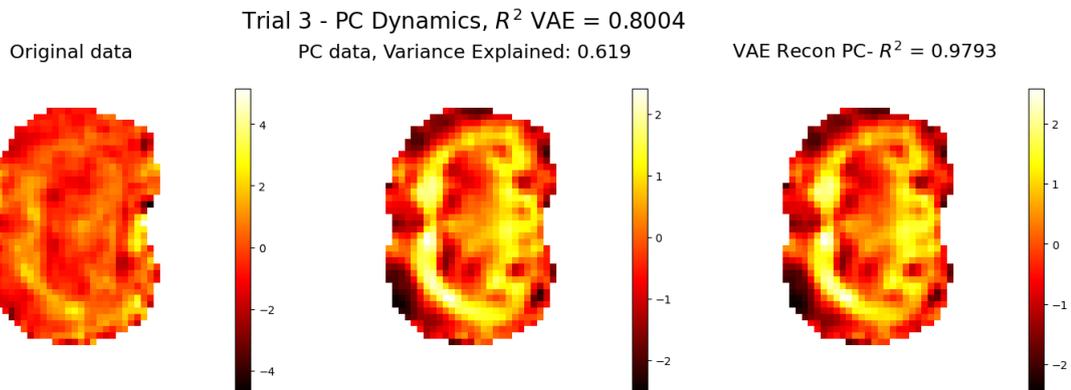


Figure 3.33: Visual comparison between 16th axial slice of original data for trial number 3, PCs inverted original data and VAE reconstruction of PCs. Every voxel contain the mean activity over the time span covered in the test dataset. The reported  $R^2$  VAE score is computed between the full time series of active voxels from PCA-inverted data and VAE’s reconstruction, and thus reflects the reconstruction accuracy of the entire brain dynamic and not just the single slice shown. The  $R^2$  score shown in the title of the rightmost plot instead refers only to the voxels within the selected axial slice.

In addition to the quantitative  $R^2$  metrics, we provide a volumetric visualization to illustrate the main spatiotemporal features captured by the first three PCs and our VAE reconstructions. Specifically Figures 3.33 and 3.34 compares the average activity map (over time) for a representative axial slice in the original rs-fMRI data, the data reconstructed by only the first three spatial principal components, and the final **VAE** output (also limited to the three PCs). In Figure 3.35, we present the mean activity over five consecutive 8-second intervals, comparing the original data to our reconstructions for each temporal segment. This visualization highlights how, in addition to preserving the broader spatiotemporal patterns, the model also retain local dynamic features throughout the time window. This allows us to visually assess how faithfully the dominant spatial patterns are preserved once we reduce the dimensionality and then re-invert the signal. Despite omitting higher-order PCs, the main structure and large-scale activity distribution remain well captured, as anticipated by our  $R^2$  scores. Hence, these time-averaged slices illustrate that retaining just a few PCs suffices to preserve the essential arousal-related spatiotemporal features we aim to study.

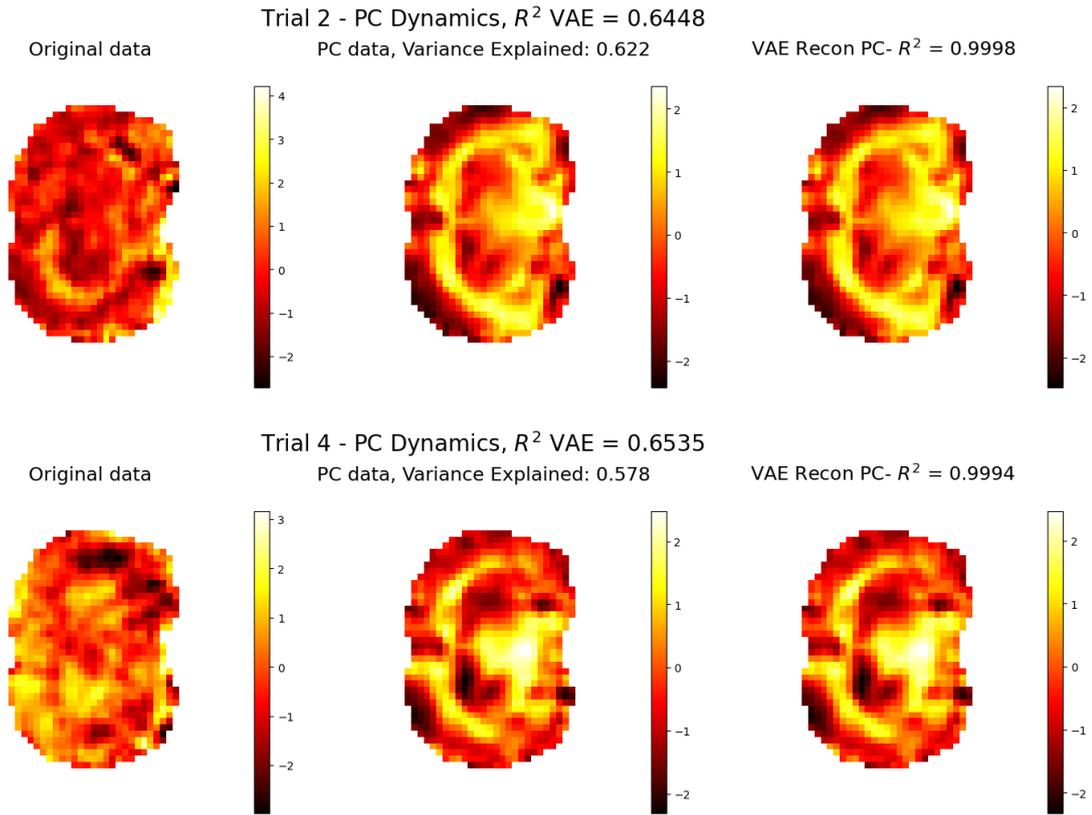


Figure 3.34: Visual comparison for trial 2,4 as described in caption of Figure 3.33

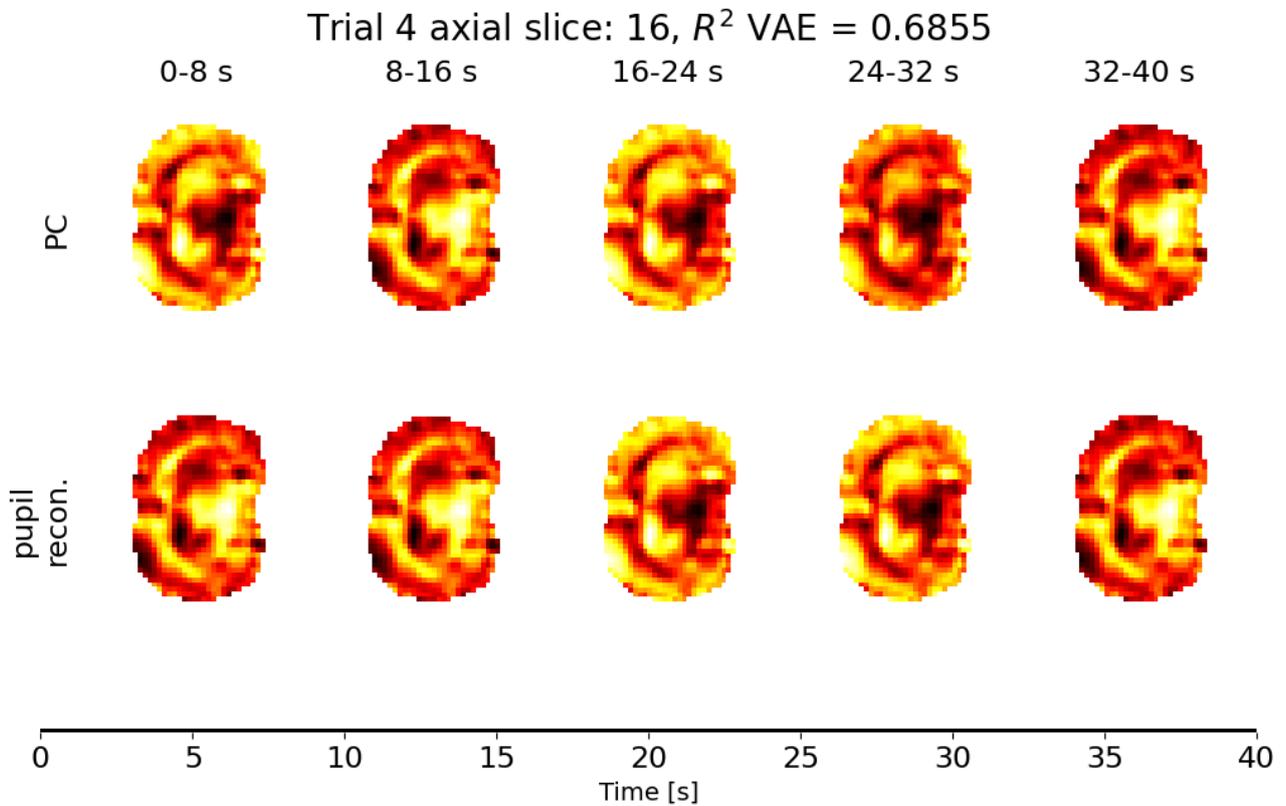


Figure 3.35: Mean activity in five 8-second interval over the first 40s of the time window covered by the test dataset

## Chapter 4

# Conclusions

In this thesis, we reviewed and tested the framework proposed by Raut et al. [6] that combines delay-coordinate embedding, dimensionality reduction (via *discrete Legendre polynomials*), and variational autoencoders to reconstruct high-dimensional brain-like dynamics from low-dimensional signals (e.g., pupil diameter). First, we tested it on a Stochastic Lorenz system and investigated the best method to choose the optimal embedding parameters, finding that while Mutual Information would suggest certain embedding parameters, the *small-window constraint* [2] yields better reconstructions within our pipeline. This underlines the importance of balancing the independence of delayed coordinates with practical limits on embedding size. We then applied the framework to real rs-fMRI data from the study "*Decoding the brain state-dependent relationship between pupil dynamics and resting state fMRI signal fluctuation*" [7], showing that these methods are able to capture a significant portion of variance in principal components associated with arousal, despite the physiological lags introduced by anesthesia. Overall, these results highlight the flexibility and robustness of the method, but also call for further developments both in the modeling methodology (e.g., using more sophisticated network architectures) and in the experimental techniques - such as acquiring data at a higher sampling rate - to further improve the accuracy and interpretability of arousal-related brain dynamics reconstruction.

# Bibliography

- [1] Toy example: Stochastic lorenz system. Supplementary Appendix of *Arousal as a universal embedding for spatiotemporal brain dynamics*, 2023. Extracted from Supplementary Appendix.
- [2] John F. Gibson, J. Dooyne Farmer, Martin Casdagli, and Stephen Eubank. An analytic approach to practical state space reconstruction. Technical Report, Los Alamos National Laboratory and Santa Fe Institute, April 1992. Report dated April 23, 1992.
- [3] Matthew B. Kennel and Henry D. I. Abarbanel. False neighbors and false strands: A reliable minimum embedding dimension algorithm. *Physical Review E*, 66(2):026209, 2002.
- [4] Matthew B. Kennel, Reggie Brown, and Henry D. I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45(6):3403–3411, Mar 1992.
- [5] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, xx(xx):1–18, 2019.
- [6] Ryan V. Raut, Zachary P. Rosenthal, Xiaodan Wang, Hanyang Miao, Zhanqi Zhang, Jin-Moo Lee, Marcus E. Raichle, Adam Q. Bauer, Steven L. Brunton, Bingni W. Brunton, and J. Nathan Kutz. Arousal as a universal embedding for spatiotemporal brain dynamics. *bioRxiv*, 2023. Preprint.
- [7] Filip Sobczak, Patricia Pais-Roldán, Kengo Takahashi, and Xin Yu. Decoding the brain state-dependent relationship between pupil dynamics and resting state fmri signal fluctuation. *eLife*, 10:e68980, 2021.