

Comparison of Dimensionality Reduction Methods for Large Scale Neuronal Recordings

Michael Jambor

Master of Science by Research
University of Edinburgh
2022

Abstract

The activity patterns of neurons are determined by their physiological properties and mutual interactions. These patterns in turn influence the behaviour and other internal workings of neural circuits. It is of great interest to neuroscience to understand how information, in the form of action potentials across neural populations, is encoded in these internal processes. However, this approach requires interpreting the activity of many simultaneously spiking neurons over time, which can pose a significant challenge. One of the emerging methods used to study the neural populations activity is dimensionality reduction of neuronal data. Dimensionality reduction methods are a set of algorithms that can interpret large multivariate neural data sets in a comprehensible way. They do so by discovering neural manifolds; low-dimensional objects in high-dimensional spaces that capture important aspects of neural activity. However, these methods are not fully established and require further implementation and basic testing. We tested three of these methods: Principal Component Analysis (PCA), Laplacian Eigenmaps (LEM) and Uniform Manifold Approximation and Projection (UMAP). Using a simulated neural data sets, we were able to reverse the manifold discovery process. We created a simple and easy-to-understand model of the place cells, generated a multitude of varying data sets, and applied the methods to them. Since we knew the intrinsic dimensionality of the simulated data, we were able to predict what lower dimensional manifolds the methods should discover. We demonstrated that PCA and LEM are unable to capture the inherently two-dimensional activity of the place cell populations. On the other hand, when provided enough data, UMAP was able to clearly represent the intrinsic dimensionality of neural populations and generate neural manifolds that, when visualized, provided an intuitive understanding of the data. We also showed how the performance of the methods varies for place cell populations of different sizes.

Acknowledgements

First of all, I would like to express my infinite gratitude to my research supervisor, Dr. Matthias H. Hennig, for giving me the opportunity me to conduct this research under his guidance. For introducing me to the world of theoretical neuroscience, scientific programming, and machine learning. And for his ever-present kind attitude that helped me and encouraged me to persevere in difficult circumstances.

I would also like to express my gratitude to my personal tutor Dr. Kim Picozzi and BMTO Student Support Officer Jenny Blair for helping me to navigate my studies in times of adversity.

Last but not least, I want to thank my friend and former neuroscientist Matej Macak for his moral support and discussions about my work, code, neuroscience, and machine learning in general.

Table of Contents

1	Introduction	1
1.1	Modern study of neural activity	1
1.2	Dimensionality reduction overview	2
1.3	Project motivation and background	4
2	Hypothesis & Aims	6
3	Methods	8
3.1	Simulation setup	8
3.1.1	Place cell model	8
3.1.2	Biological plausibility variables	10
3.1.3	Population activity model	11
3.2	Dimensionality reduction	12
3.3	Experimental script	12
3.4	Data analysis	13
3.4.1	Manifold visualization	13
3.4.2	Correlation of distances	13
4	Results	14
4.1	Single manifold analysis	14
4.1.1	Manifold visualization	14
4.1.2	Quantitative assessment	16
4.2	Populations analysis	18
5	Discussion	22
5.1	Outcomes	22
5.2	Limitations	23
5.3	Future Directions	24

Bibliography	26
A Appendix	32

Chapter 1

Introduction

1.1 Modern study of neural activity

The common denominator of all brain activity, whether in humans, mosquitoes or cephalopods, is the presence of action potentials. This type of rapid information transfer, if present in sufficient quantities, can be considered the basic building block of all intelligent biological systems. Yet it is highly unlikely that random and uncoordinated brain cell spiking activity would generate any kind of intelligence on its own, let alone high-level cognition. So what gives neural systems, such as the brain or spinal cord, their emergent properties? Another element necessary for the emergence of intelligent systems of this kind is coordination; sets of rules, modulations, and constraints that determine the patterns of neural activity. The existence of these is evident. The activity of individual neurons is limited by their physiological properties and the spiking of neuronal ensembles and larger circuits is governed by their highly complex mutual connectivity and other physical influences (Yuste, 2015; Shepherd, 2004). But how can these rules governing neural populations activity be captured and further investigated?

This question can generally be addressed in two ways. The first approach is to focus on (electro)physiological properties of neurons and their interconnectivity (Shepherd, 2004). This approach seeks to derive the functionalities and features of neural circuits through their physical structure. While this practice clearly serves neuroscientific discovery to this day, it may not be as compelling as previously assumed (Hennig, 2022). The second approach to studying the characteristics of neural populations activity is to record and examine the spiking of the populations as such (Stevenson and Kording, 2011; Yuste, 2015; Whiteway and Butts, 2019).

Neuroscience faces two major challenges in this regard. Namely, data acquisition and data interpretation. In the last decades, neuroscience has seen great advances in the methods of

multielectrode recordings (Steinmetz et al., 2021; Kipke et al., 2008) and optical recordings (Ahrens et al., 2013; Kerr and Denk, 2008). These techniques make it possible to record spiking activity of hundreds to thousands of neurons simultaneously in different brain regions under various types of experimental setups. This is a significant difference compared to more traditional methods such as single-electrode or tetrode recordings, which can capture the spiking of only a few neurons at a time (Buzsáki, 2004). Technological innovations of this magnitude enable researchers to reevaluate the types of scientific questions that are being proposed. This is also true for well-established experimental paradigms and thoroughly researched brain areas that have been studied for decades now (Cunningham and Yu, 2014). The question at this point is, how can we relate this large amount of hard-to-interpret spiking activity of potentially thousands of neurons to animal behavior and/or other experimental variables? The more recent challenge that neuroscientists currently face is the development of methods capable of finding meaningful interpretation of these highly multivariate data sets (Brown et al., 2004; Stevenson and Kording, 2011; Paninski and Cunningham, 2018).

1.2 Dimensionality reduction overview

In recent years, a set of statistical and machine learning techniques called *dimensionality reduction methods* have shown promise as a tool for analyzing and interpreting large neural data sets (Whiteway and Butts, 2019; Cunningham and Yu, 2014). Dimensionality reduction methods leverage the fact that sample points of neural population activity can be represented as multi-dimensional vectors, where each component of the vector represents the activity of a particular neuron (Whiteway and Butts, 2019). The activity of neurons in this case is represented as the number of spikes per time interval. Further, the number of neurons determines the number of dimensions of the vector. Inherently, these vectors occupy multidimensional space where they depict the activity of the recorded neural population.

This is where dimensionality reduction methods are applicable for neural recordings. The rationale for their application is the assumption that although the activity of all recorded neurons takes place in a high-dimensional space, its underlying dynamics occupy a considerably lower-dimensional space (Cunningham and Yu, 2014; Whiteway and Butts, 2019; Churchland et al., 2007; Mitchell-Heggs et al., 2022). Essentially, these methods scale down the number of dimensions of the original data set, while retaining particular aspects of the neural activity and discarding others. They turn N dimensional data set, where N is the number of neurons, to its K

dimensional interpretation, where $K < N$. These K dimensions are termed *latent variables* and they reflect complex patterns of neural activity in a smaller set of summarizing features. The type of aspects that the methods capture depends solely on the design of each method. There is a multitude of dimensionality reduction methods applicable to neural recordings, each focusing on different aspects of the data set. Dimensionality reduction methods have already proven successful in uncovering the diverse neural mechanisms underlying decision-making (Mante et al., 2013; Harvey et al., 2012; Stokes et al., 2013; Briggman et al., 2005), motor system (Pandarinath et al., 2018; Kaufman et al., 2014; Churchland et al., 2012; Yu et al., 2008; Churchland et al., 2010b), olfactory system (Stopfer et al., 2003; Broome et al., 2006; Saha et al., 2013), auditory system (Luczak et al., 2009), spatial representation (Gardner et al., 2022; Chaudhuri et al., 2019; Rubin et al., 2019), visual attention (Cohen and Maunsell, 2010), working memory (Rigotti et al., 2013), and many others. But how exactly are the outputs of these methods used to better understand the functions of neural populations?

The lower-dimensional representations of the neural activity can be treated as mathematical objects called manifolds, or in this case, *neural manifolds*. Manifold is a concept originally derived from topological algebra and its definition can be highly complicated and abstract. For the purposes of this work, we adopted the intuitive definition of Mitchell-Heggs et al.: "The neural manifold is, in brief, a low-dimensional surface within the higher-dimensional space of neural activity which explains the majority of the variance of the neural dynamics." (Mitchell-Heggs et al., 2022).

As mathematical objects of a specific kind, neural manifolds can be analyzed by means of topology and geometry (Mitchell-Heggs et al., 2022; Khona and Fiete, 2021). The most common approach of analyzing a neuronal manifold is by demonstrating a strong relationship between its dimensions, i.e., latent variables, and shape with respect to animal behavior or some other experimental variables. A prime example is a recent publication showing how head direction information is represented in thalamic neuronal populations in mice. After reducing the dimensionality of neuronal activity, Chaudhuri et al. retrieved one-dimensional ring in which the head direction was represented by a single variable (Chaudhuri et al., 2019). In another recent publication, Gardner et al. have shown that grid cells activity in the medial entorhinal cortex can be effectively reduced to two latent variables that lie on a geometrical object known as a torus (a hollow donut) (Gardner et al., 2022). Furthermore, dimensionality reduction methods may reveal new experimental variables, the existence of which was not previously registered or assumed (Rubin et al., 2019).

Although the use of dimensionality reduction methods for neural recordings is very promising and yields significant results, it is still relatively new and not devoid of its own challenges. Some of the complications facing this approach are experimental design issues, e.g., what methods to use for different experimental setups and recorded brain regions, how many neurons to record, what should be the duration of a trial, whether it is always better to have more data, etc.? Other, equally important, are the interpretations of the resulting neural manifolds. While some of these challenges can be addressed theoretically, others, especially the former, require empirical investigation.

1.3 Project motivation and background

Dimensionality reduction methods can be classified into several overlapping categories. Among the most important of these categories are linear and non-linear methods, both of which are used to reduce the dimensionality of neural recordings (Cunningham and Yu, 2014; Mitchell-Heggs et al., 2022). As their names suggest, linear methods seek to find linear relationships between the neural activity and latent variables, while non-linear methods seek non-linear relationships. In general, the application of linear methods to neural recordings is preferable due to their better intelligibility and easier interpretability of the results. However, it is reported that non-linear methods are more successful at generating lower-dimensional neural manifolds at the expense of potentially more complex interpretability of the resulting manifolds (Mitchell-Heggs et al., 2022).

The general objective of this work was to compare the performance of one linear and two non-linear dimensionality reduction methods on simulated neural activity. The linear method of interest was Principal Component Analysis (PCA) (Jolliffe, 2002). The non-linear methods were Laplacian Eigenmaps (LEM) (Belkin and Niyogi, 2003), and Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018).

PCA, invented as early as 1901 (Pearson, 1901), is one of the most standard and well-established dimensionality reduction methods used for far more than just neural data. To briefly describe the method, PCA performs a series of linear transformations on the data set to reduce its dimensionality while preserving as much uncorrelated variance as possible. The other two methods, LEM and UMAP, are considerably more novel than PCA. LEM is a geometry-based non-linear method that utilizes the insights of graph theory to generate lower-dimensional manifolds. UMAP is non-linear method built on algebraic topology and Riemannian geometry. The method

is also partially based on mathematical principles of LEM. PCA (Ahrens et al., 2012; Gao and Ganguli, 2015; Mazor and Laurent, 2005; Churchland et al., 2010a), LEM (Rubin et al., 2019; Sun et al., 2019), and UMAP (Tombaz et al., 2020; Lee et al., 2021), have all successfully produced new findings about neural activity in the past.

Rather than applying the dimensionality reduction methods to unexplored neural data sets to yield new insights, as is usually done, we reversed the process in this project. The main contribution of this work was the application of the methods on simulated, *ground truth*, neural activity, i.e., data whose underlying dynamics were well understood prior to analysis. Hence, we knew what the general properties of the lower-dimensional neural manifolds should be. The methods were applied to data sets consisting of simulated spiking activity of place cell populations. We simulated one of the basic and ever-present properties of place cells, namely the response of the cell to the current position of the animal in its local environment (O’Keefe and Dostrovsky, 1971; O’Keefe, 1976). We set up the model so that the entire multi-dimensional neural data set was the output of only two variables; the x and y coordinates in the experimental environment. This helped us to predict what type of neural manifold and number of latent variables the dimensionality reduction methods should reveal. We applied PCA, LEM, and UMAP methods to this neural activity to generate neural manifolds. We then compared the geometric properties of these manifolds to determine which method best captured the underlying mechanisms of the simulated place cell populations activity and what conditions were necessary for this.

Chapter 2

Hypothesis & Aims

Hypothesis

Dimensionality reduction methods can produce a lower-dimensional representation of simulated activity of place cell populations. Furthermore, non-linear methods are more likely to successfully construct a two-dimensional neural manifold whose shape and latent variables can be strongly linked to the simulated position of the animal in the experimental environment. We further expect that the quality of the neural manifolds will be positively affected by the number of place cells recorded and the duration of the trial.

The hypothesis was tested by meeting the following aims.

Aim 1: Model

We constructed a simplistic but biologically plausible computational model of place cell activity as a function of the position of the simulated animal in the experimental environment. We then recorded the activity of variety of different populations of place cells, while varying the number of neurons and the duration of trials, to obtain different data sets.

Aim 2: Dimensionality reduction

We systematically applied each PCA, LEM and UMAP dimensionality reduction methods to all generated data sets. All the methods were applied to each data set. The resulting two-dimensional neural manifolds were stored for subsequent analysis.

Aim 3: Geometrical analysis

We compared the geometrical properties of the latent variables of the manifolds to the experimental variables. We developed color map encoding to visually compare the neural manifolds with the experimental environment. We also compared the distances of randomly selected locations within the experimental environment and within the manifolds. Finally, we applied this analysis to all the populations of neural manifolds and generalized the results.

Chapter 3

Methods

The methodology of this research project was solely computational. The code was written in Python 3 programming language and its corresponding libraries.

Contributions

All code used in this project was written by myself under the supervision of Dr. Matthias H. Hennig. With the exception of open source Python libraries and modules, no previously written code was implemented in this project. The following libraries and modules were used: NumPy for handling arrays, linear algebra operations, and to retrieve mathematical functions (Harris et al., 2020), Matplotlib for visualizing figures (Hunter, 2007), SciPy for statistical operations (Virtanen et al., 2020), Scikit-learn to retrieve PCA and LEM dimensionality reduction methods (Pedregosa et al., 2011), and github.com/lmcinnes/umap repository to retrieve UMAP dimensionality reduction method (McInnes et al., 2018).

3.1 Simulation setup

3.1.1 Place cell model

The place cell activity model we developed and used to simulate each individual neuron consisted of two mathematical functions. The first function, which we used to define the general area and strength of the neuron's place field, was un-normalized bivariate Gaussian function (**Function 3.1**).

$$f(x) = A \times \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \quad (3.1)$$

The parameter μ , the center of the function, consisted of a two-dimensional vector which established the location of the place field. The two dimensions of the vector μ were defined by arbitrary x and y coordinates from the simulated environment. The coefficient A established the maximum firing rate of the place cell. The diagonal covariance matrix Σ defined the size of the place field. The intrinsic property of the Gaussian function established decreasing strength of the place field in the locations further away from the field's center. An example of the 2D Gaussian plot is shown in **Figure 3.1**.

To add stochasticity for biological plausibility of neural spiking and to generate integer values of spike counts per time interval, we used the Poisson random process (Tolhurst et al., 1983) (**Function 3.2**) as the second function of the model.

$$P(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (3.2)$$

The variable λ was the expected number of spike counts of a place cell for given location, i.e., λ was the output of the **Function 3.1**. The variable k was the number of potential spikes ($k = 0, 1, 2, \dots$). The Poisson random process calculates the probabilities of each k given λ and then draws samples from the distribution based on those probabilities.

The spike counts of an individual place cell were then randomly sampled by **Function 3.3**, i.e., Function 3.2 of Function 3.1.

$$y \sim P(k; f(x)) \quad (3.3)$$

After establishing the place cell model, the spike count of an individual place cell per arbitrary x and y coordinates inside the simulated environment was obtained by using the coordinates as a 2D vector input for **Function 3.3**. The result was a scalar integer number of action potentials over a specific time interval (time bin specified below) at the location of the x and y coordinates. An example of a place cell activity plot is shown in **Figure 3.1**.

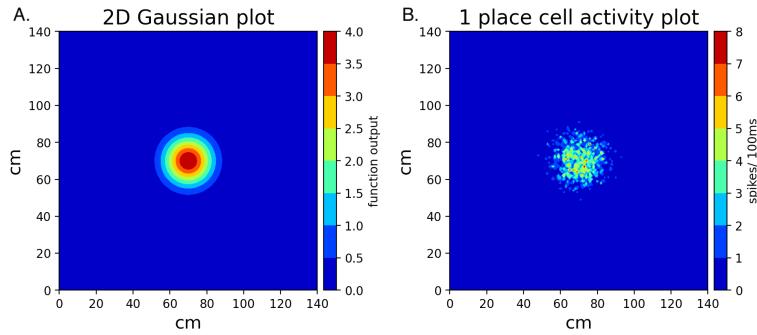


Figure 3.1: **(A)** Spatial plot of Function 3.1; the precursor of the place field simulation. Gaussian function centered in the middle of a square experimental environment. The function generates a scalar value for each coordinate within the environment. The value at each coordinate is represented by a heat map. **(B)** Spatial plot of Function 3.3; a place field. Spike counts of one simulated place cell per each coordinate within the environment. The Poisson random process (Function 3.2) adds stochasticity to the activity. The number of spike counts at each coordinate is represented by a heat map. The simulated place cell has the highest activity when the simulated animal is in the middle of the environment. The number of spike counts decreases as the animal moves further away from the place field center. The graph shows the number of spikes at an arbitrary time t . The general shape would be preserved at time $t + 1$ but the values would be slightly different due to the stochastic element of the model.

3.1.2 Biological plausibility variables

To further cultivate the biologically plausibility of the place cell model to fit the experimental setting, we set the un-normalized Gaussian function (**Function 3.1**) covariance matrix Σ as:

$$\Sigma = \begin{bmatrix} 80 & 0 \\ 0 & 80 \end{bmatrix}$$

This established that each place field covered approximately the same area size of the simulated environment ($140\text{cm} \times 140\text{cm}$) as was observed for rat place cells in a real experimental setup of similar dimensions (O’Keefe and Burgess, 1996). To emulate the spiking rate of real place cells and to collect an appropriate number of spikes per time bins of 100ms we set the coefficient A of **Function 3.1** as $A = 4$. In combination with the Poisson random process (**Function 3.2**), this established the mean maximum firing rate of the place cell as 4 spikes per 100ms (McClain et al., 2019). At this point, the place cell activity model was fully built for our experimental setup.

The variables Σ and A could also be adjusted accordingly to match and simulate different experimental conditions. This gives our model a modest degree of versatility for different experimental conditions.

3.1.3 Population activity model

To simulate the activity of a place cell population, we initialized N place cell models and recorded their activity for T time bins per one trial. Each place field location within the simulated environment was selected randomly. All simulations were run in a $140\text{cm} \times 140\text{cm}$ square environment. Centers of the place fields did not extend beyond the environmental barriers and all sample points of the population activity were recorded inside the environment. **Figure 3.2** displays examples of the spatial activity of different place cell populations.

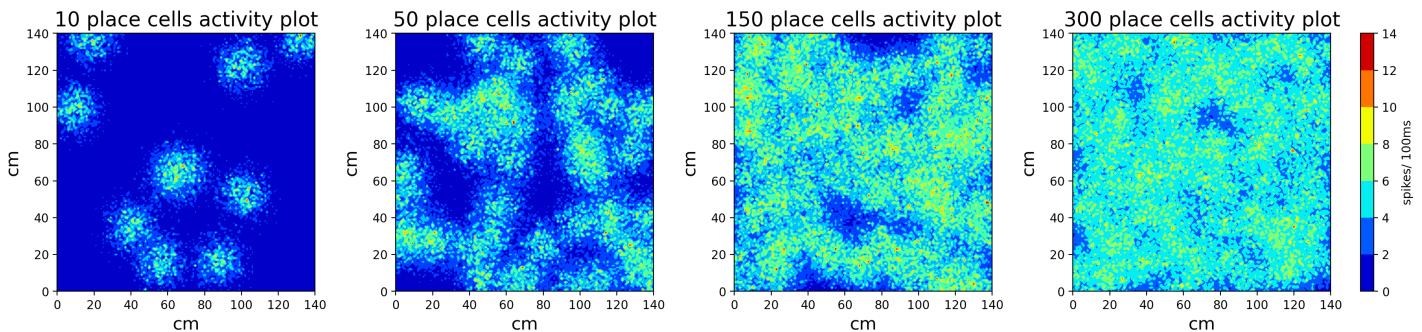


Figure 3.2: Examples of activity of four place cell populations inside a $140\text{cm} \times 140\text{cm}$ environment. This is an extension of the concept shown in Figure 3.1(B), but for multiple place cells. The number of spikes per 100ms of each neuron at each location is indicated by the heat map. The plots show superimposed activity of these neurons. The plots illustrate activity and range of four different populations of 10, 50, 150, and 300 place cells respectively. These populations are examples drawn from much larger data files.

Population activity sample points were collected in the following way. Two-dimensional vector containing randomly selected x and y coordinates of the simulated environment was set for each sample point. The spiking activity of all N neurons was calculated as the function **Function 3.3** of the coordinates vector. The neural spike counts were collected in a row vector where each column represented a different place cell. This process was repeated T times to collect T sample points. All the row vectors were stacked horizontally. The final data set of each trial consisted of T by N matrix where each column represented activity of one place cell per T time bins and each row represented the activity of all place cells per one 100ms time bin. Each sample point vector had N dimensions. Thus, the number of neurons N determined the number of dimensions

of the data set. The simulation of the animal movement described here is clearly unrealistic, but the non-dynamic nature of the dimensionality reduction methods we used allowed for this (more in the discussion).

3.2 Dimensionality reduction

To generate two-dimensional neural manifolds contained in the population activity, we used three dimensionality reduction methods. These methods were PCA, LEM, and UMAP. We used the PCA and LEM methods in their default setting as provided by the developers of Scikit-learn Python library (version 1.1.1) (Pedregosa et al., 2011) and the UMAP method in its default setting as provided by the developers of its Python module¹ (version 0.5.3) (McInnes et al., 2018).

We applied all three methods to each trial data set to obtain its two-dimensional neural manifolds for each dimensionality reduction method. We thus obtained a set of two-dimensional vectors that were projections of the original N-dimensional sample points. We then compared the resulting two-dimensional manifolds with the experimental ground truth environment using different analysis methods to evaluate the performance of each method.

3.3 Experimental script

To systematically explore the influence of the number of recorded place cells and the duration of the trial on the quality of the neural manifolds, we created an experimental script. The objective of the script was to systematically increase the number of recorded neurons, the duration of the trial, and to apply all three dimensionality reduction methods on the resulting data sets. The script generated tests with 10 up to 300 recorded place cells with an increment of 10 cells, that is $N = (10, 20, \dots, 300)$. Per each test, 5 different place cell populations of N neurons were generated. Neural activity of every population was recorded for 5, 10, 15, 20, 25, and 30 minutes using 100ms time bins. Each recording was considered as one trial. PCA, LEM and UMAP were applied on each trial data set. The script generated 150 different place cell populations, 900 trials, and 2700 two-dimensional neural manifolds.

¹github.com/lmcinnes/umap

3.4 Data analysis

3.4.1 Manifold visualization

As the first method to examine the embeddings, we used color coded visualization. Each coordinate of the experimental environment was assigned color based on its location. We used the 'binary' color map of the Python Matplotlib library for the x coordinates and the 'spring' color map for the y coordinates. These were combined to create a two-dimensional continuous color map that produced unique color code for each coordinate (see **Figure 4.1**). We used this coding to label the sample points. When plotting the two-dimensional manifolds, we used the color codes of the sample points to display how the original experimental coordinates are represented in the manifolds. We compared the visualizations of the neural manifolds to the original experimental environment.

3.4.2 Correlation of distances

To quantitatively analyze the quality of the manifolds, we randomly selected 5000 pairs of coordinates in the experimental environment and calculated their Euclidean distances. We then selected the projections of these pairs inside the neural manifolds and calculated the Euclidean distances of those. As a quantitative measure of the manifold quality we calculated the Pearson correlation coefficient of the experimental environment distances and their projections inside the manifolds. The resulting correlation coefficients quantitatively indicated how well the manifolds preserve relative distances between the transformed coordinates and how well do the latent variables relate to the ground truth variables.

To compare the performance of the dimensionality reduction methods over neural population size and trial duration, we calculated the averages of the Pearson correlation coefficients for all populations of N neurons of trial duration T . Because correlation coefficients are values from probabilistic space, calculating their arithmetic mean would be deficient. To compute the average, we first transformed all correlations into the Fisher Z-space. There, we calculated their arithmetic mean and standard deviation. We then applied the inverse of the Fisher Z-transformation to the results and transformed them back to the probabilistic space.

Chapter 4

Results

4.1 Single manifold analysis

4.1.1 Manifold visualization

Due to the extensive numbers of place cell populations, trials, and even larger number of their manifolds that we generated (150, 900, and 2700 respectively), we visually examined only a smaller subset of these. The subset consisted of 36 manifolds of 12 trials of 4 different populations. We picked these populations to reflect the impact of sufficiently large increments of the population size on manifold quality. The example populations consisted of one of 10, 50, 100 and 200 simulated neurons. For each population we examined 3 manifolds of each dimensionality reduction method. The manifolds inspected were those of 5, 15, and 30 minutes of the trial duration. Again, we did so to examine the effect of sufficiently large increments in trial duration on the quality of the manifolds. We visualized the manifolds using the color map coding (see **Figure 4.1**).

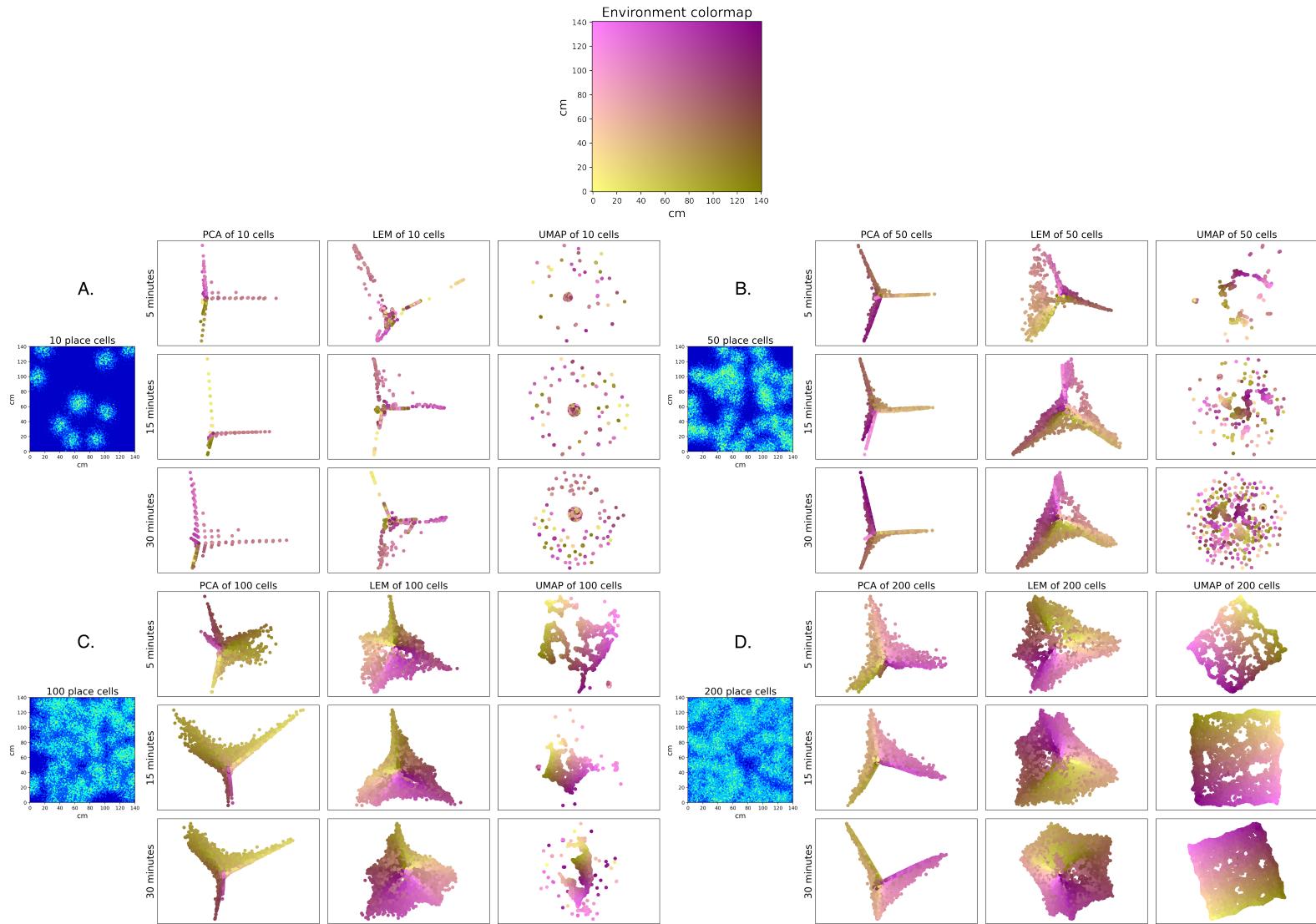


Figure 4.1: Color map of the original experimental environment (top), each 2D coordinate has unique color based on its location. Color map visualization of 36 neural manifolds of 4 different place cell populations of 10, 50, 100, and 200 neurons (below). The plots are divided into 4 sections A, B, C, and D, each rendering 9 manifolds of one place cell population. Each section displays 3 manifolds of PCA, LEM and UMAP dimensionality reduction methods with respect to the duration of the trial. The durations were 5, 15, and 30 minutes. The experimental color map (top) serves for universal comparison of all manifolds to the experimental environment coordinates. The manifolds are plotted in arbitrary units. The axes scales of the embeddings are omitted in these plots as it is solely the shape and color code that is of interest here.

The color code plotting of the manifolds visually demonstrates how well the different dimensionality reduction methods capture the original two-dimensional environment with respect to the number of recorded neurons and trial duration. The two-dimensinal latent variables were plotted in Cartesian space. The color coded layout of the latent variables shows the shape of the neural manifold and provides direct visual comparison to the experimental coordinates.

With regards to the examples shown in **Figure 4.1**. For the population A with 10 place cells, no method was able to successfully capture the experimental environment, even with longer trial duration. This is not surprising, as the place field activity covered only a small fraction of the experimental environment. For the population B of 50 neurons, all three methods began to successfully cluster some coordinates of the environment. It appears that especially LEM managed to reconstruct some local environments better than the other methods. Nevertheless, still none of the methods captured the ground truth environment as a whole. For the figure C of 100 neurons LEM and UMAP improved in reconstructing some local environments. PCA manifolds did not improve significantly compared to the previous population. Again, however, none of the methods managed to capture the whole ground truth environment.

For population D of 200 cells, PCA, again, did not achieve significantly better results than in the previous case. LEM captured some local environments well, but could not reconstruct the entire environment. With a sufficient number of recorded neurons, the visualization of UMAP manifolds suggests that this method is capable of effectively capturing and reconstructing the entire environment. It also appears that UMAP manifolds become progressively more accurate with increasing experimental duration as the number of holes in the geometrical representation of the manifold was decreasing.

4.1.2 Quantitative assessment

Although visualization can reveal important information about the manifolds, such as the shape and direct location of the transformed coordinates, it does not express how well the relative distances between the transformed coordinates are preserved. For that reason, we chose the Pearson correlation coefficient as quantitative measure of the quality of the manifolds. We calculated the correlation coefficient between 5000 randomly selected pairs of distances in the experimental environment and their projections in the neural manifolds. **Figure 4.2** shows examples of the quantitative quality of six manifolds of two different populations of place cells, three per each.

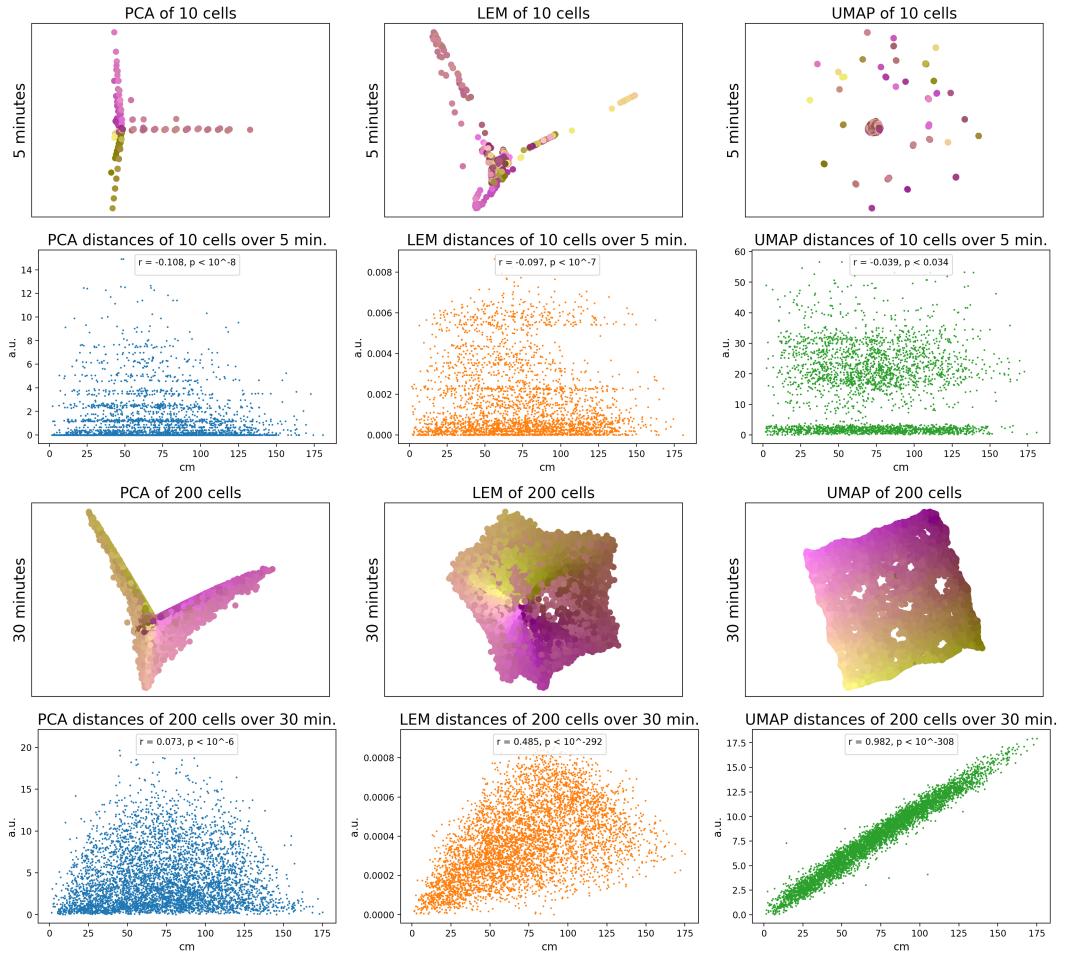


Figure 4.2: Neural manifolds and associated scatter plots of distances in the experimental environment (x-axis) relative to the distances of their neural manifold projections (y-axis). The plots show two examples of neural populations with low (top) and high (bottom) number of place cells and trial durations. The scatter plots also contain Pearson correlation coefficients and their associated p values. Scatter plots and Pearson r inform on how well the different neural manifold coordinates correlate with the experimental environment.

Figure 4.2 shows the quantitative quality of PCA, LEM, and UMAP manifolds for a population of 10 place cells recorded for 5 minutes and a population of 200 place cells recorded for 30 minutes. For manifolds created from the activity of 10 cells, the analysis confirms that the quality of any method is low and fails to capture the relative distances between the coordinates. PCA, LEM, and UMAP perform almost equally poorly with correlation coefficients of -0.108 ($p < 10^{-8}$), -0.097 ($p < 10^{-7}$), and -0.039 ($p < 0.034$) respectively. As the second example in **Figure 4.2**, we selected a population where some of the neural manifolds were significantly more accurate. For the activity of 200 neurons for 30 minutes, PCA did not achieve better results

than in the previous example with a correlation coefficient equal to 0.073 ($p < 10^{-6}$). LEM produced much better manifold compared to the previous example with correlation coefficient of 0.485 ($p < 10^{-292}$) and so did UMAP with correlation coefficient of 0.982 ($p < 10^{-308}$). Of all the examples above, UMAP neural manifold best captured the experimental variables when provided sufficient data.

4.2 Populations analysis

To perform broader quantitative analysis of all the 150 different place cell populations, 900 trials and their corresponding 2700 neural manifolds, we averaged the Person correlation coefficients of all neural manifolds across population sizes and trial durations. In doing so, we also revealed how the quality of the manifolds varies with the number of place cells recorded and the duration of the trial. We calculated the mean values of the Pearson correlation coefficients for each type of manifold of all 5 populations for population size N and trial duration T. The values were calculated using the Fisher Z-transformation (motivation explained in the methods).

First, we plotted the average correlation values of the PCA, LEM, and UMAP manifold quality over time per number of neurons. Examples of these plots can be seen in **Figure 4.3** (the rest in the appendix).

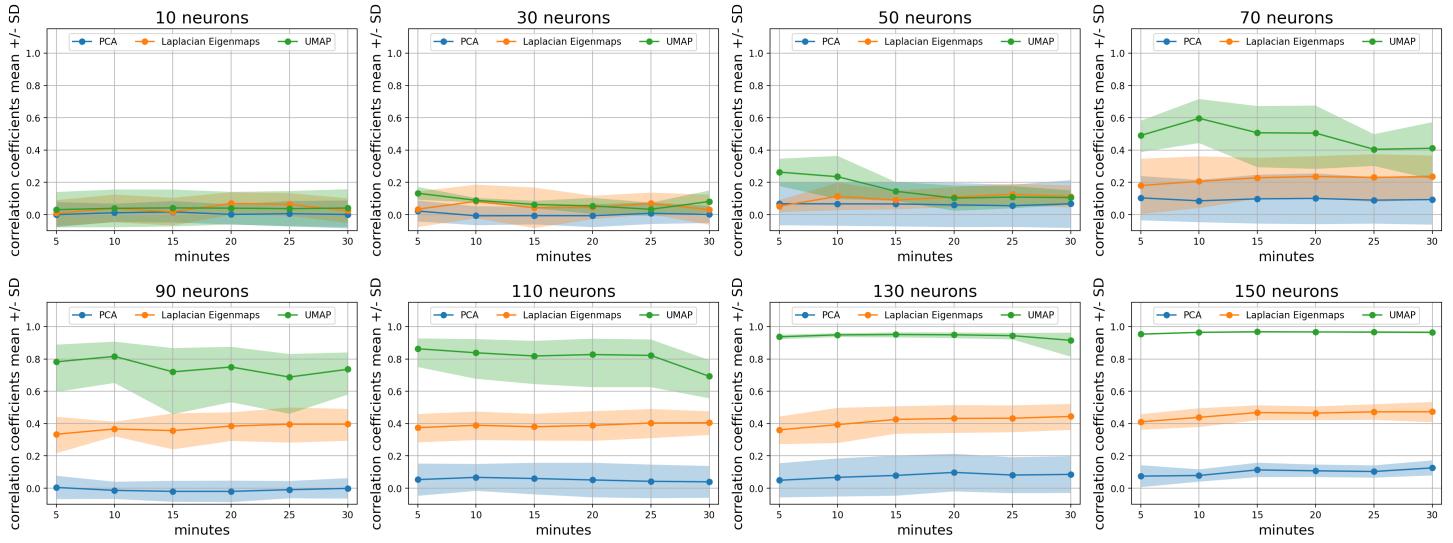


Figure 4.3: The plots show the quality of the neural manifolds produced by PCA, LEM, and UMAP expressed by averaging the Pearson correlation coefficients over time per recorded number of neurons. The x-axis shows the duration of the trial. The y-axis shows the average quality of the neural manifolds. Each point on the graphs represents the average quality (+/- one standard deviation) of 5 manifolds for each dimensionality reduction method. Each graph shows the quality per number of neurons.

As the number of recorded place cells increased, the quality of the PCA manifolds did not significantly improve. The values of the correlation coefficient mean oscillated approximately between 0.0 and 0.2 for each number of recorded place cells. The average quality of LEM and UMAP manifolds improved with the number of recorded place cells. Nevertheless, the same cannot be said with regards to the duration of the trials that we measured. The duration of the trials had either minimal or no positive impact, and in some cases even a negative impact on the manifolds quality. It is clearly visible from the plots (**Figure 4.3**) that number of neurons had much more significant impact on the quality of neural manifolds produced by LEM and UMAP.

We therefore next focused on how the quality of manifolds changes with respect to the number of recorded neurons per trial duration. We plotted the data in a graph to better illustrate the dependence of Pearson r means on the number of recorded place cells. Examples of these plots can be seen in **Figure 4.4** (the rest in the appendix).

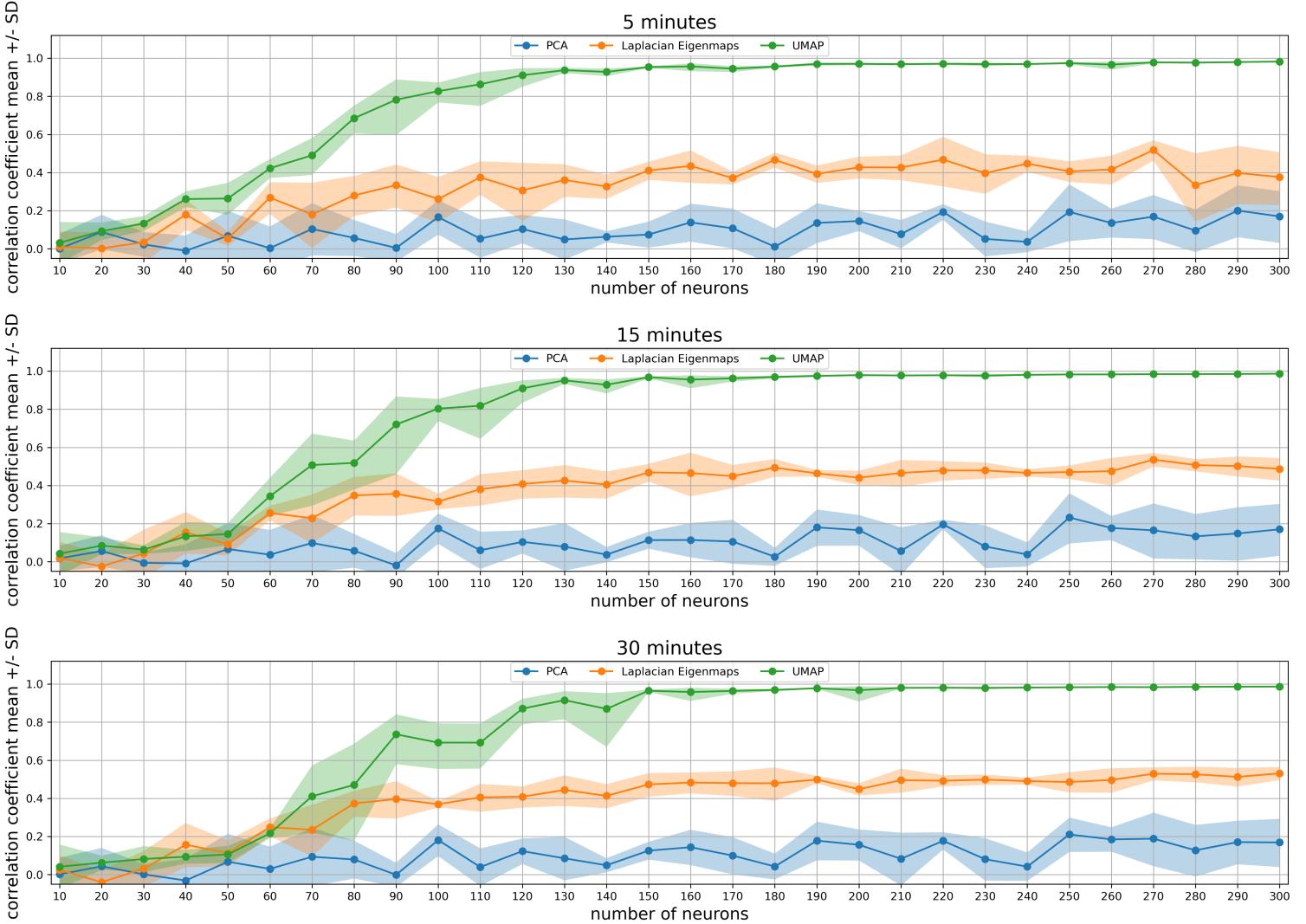


Figure 4.4: The plots show the quality of the neural manifolds produced by PCA, LEM, and UMAP expressed by averaging the Pearson correlation coefficients over recorded number of neurons per trial duration. The x-axis shows the number of recorded place cells. The y-axis shows the average quality of the neural manifolds. Each point on the graphs represents the average quality (+/- one standard deviation) of 5 manifolds for each dimensionality reduction method. Each graph shows the quality per duration of a trial.

The graphs clearly show that as the number of neurons increases, the quality of LEM and UMAP manifolds increases as well. The quality of PCA manifolds does not change significantly with the rising number of recorded neurons. Mean Pearson r of the PCA manifolds oscillates between 0 and 0.2 (with smaller outliers) for all populations between 10 to 300 neurons. This also does not change with duration of the trial. The quality of LEM improves with respect to number of neurons but never exceeds above the value of mean Pearson r of 0.6. However, this method performs better than PCA and is capable of clustering a few local areas and preserving some relative distances between the coordinates.

The only method able to generate neural manifolds with quality of mean Pearson r \geq 0.9 was UMAP. Quality of UMAP manifolds saturated in populations of 150 neurons or larger in all tested trial durations. UMAP manifolds achieved the mean correlation coefficients of 0.95, 0.96, 0.97, 0.97, and 0.97 of 150 place cells activity during all trial durations of 5, 10, 15, 20, 25, and 30 minutes respectively. For the other UMAP neural manifolds generated with recordings using larger number of neurons, the quality of the manifolds did not change significantly.

Chapter 5

Discussion

5.1 Outcomes

In this research project, we tested the properties of neural manifolds generated by three dimensionality reduction methods; PCA, LEM, and UMAP. We focused primarily on their reliability, i.e., how well can each method capture the intrinsic dimensionality of neural data. The key asset of the project was that we applied the methods to data of simulated place cell population activity where the intrinsic dimensionality was known. Therefore, we had informed expectations about what latent variables the resulting manifolds should capture.

As expected, PCA failed to reproduce two-dimensional neural manifolds with latent variables that could be easily related to the actual simulated two dimensions of the original data set. We expected that this method would not be successful in the task of reducing the data sets to only two latent variables that would clearly inform about the experimental variables. As mentioned above, PCA is a linear dimensionality reduction method. However, the simulated place cell activity was calculated using two non-linear functions. Therefore, we did not expect PCA to capture the non-linear features in the data set. All analysis methods concluded that PCA could not capture the original experimental variables in two dimensional manifolds under any conditions that we varied. Yet it would be false to claim that the first two latent variables do not carry any meaningful information about the data sets. Further analysis of the first two, as well as other latent variables, may reveal what information these variables carry (Recanatesi et al., 2021).

As the number of simulated place cells increased, LEM performed significantly better than PCA in reducing the data sets to two latent variables. LEM is a non-linear method, so we expected it to perform significantly better than PCA when reducing the data sets to two dimensions since

the data were generated by non-linear methods. However, LEM also failed to clearly capture the original intrinsic variables of the simulated data sets. The method was unable to generate neuronal manifolds whose quantitative quality exceeded Pearson r of 0.6 , even as the number of recorded neurons increased. Visualization of the color coding also showed that the LEM does not clearly capture the original experimental variables, and suggested the reason why this is the case. Visualization of the LEM neural manifolds revealed that the LEM captured some local areas well, but failed to capture the global environment. Yet again, it would be incorrect to claim that the two latent variables produced by LEM do not carry any meaningful information about the data. It may be the case that these manifolds carry information about variables that arose in our experimental design but that we are not aware of (Rubin et al., 2019).

Of the methods tested, UMAP was the most successful in generating two latent variables corresponding to the original experimental variables. UMAP outperformed both PCA and LEM in this regard. If given data from a sufficient number of simulated place cells, 150 or more, UMAP was able to generate neural manifolds that faithfully captured the ground truth experimental variables. This was evident in the color coded visualizations, where the neural manifolds had the shape and spatial arrangement of coordinates that closely resembled the color coding of the experimental environment. Furthermore, quantitative analysis of the quality of the manifolds demonstrated that UMAP also preserved the relative distances between the coordinates projections remarkably well. This is a very useful finding as it strengthens the position of UMAP as a dimensionality reduction method that was designed to preserve relative distances within its manifolds (McInnes et al., 2018). UMAP was the only method that, when provided with sufficient data, captured the original experimental variables.

Based on the trials performed, the quantitative quality of UMAP manifolds saturated at 150 recorded place cells. Any number above this value yields neural manifolds of virtually the same level of quality (mean Pearson $r \geq 0.95$). The duration of the trials had no significant effect on the manifolds. UMAP was able to generate high-quality neural manifolds using the shortest trials we tested, i.e., 5 minutes.

5.2 Limitations

Most of the limitations of this work stem from our computational model of place cell population activity and the method of data collection. However, some of the limitations are intended for the experimental design. For example, creating only a simple model of place cells and ignoring

other aspects of place cell activity that are well documented. This helped us to only focus on handful of variables that we wanted to rediscover with dimensionality reduction instead of creating complex combinations of place cell activity features that would make the model much harder to understand. Therefore, we omitted properties of place cells activity such as their spiking outside of their place fields (Foster and Wilson, 2006; Ferbinteanu and Shapiro, 2003; Johnson and Redish, 2007) and others. These properties of place cells could potentially be added to the model in future experiments studied with dimensionality reduction. It would also probably be easier to study them in real animals.

Another limitation of this project was the method of movement of the simulated animal. As stated in the methods, the movement of the animal in the experimental environment during which neural activity was recorded was completely random. That is, each location was unrelated to the animal's previous location and the animal did not move in any kind of natural or even realistic pattern. This experimental design was largely justified because we did not use dynamic dimensionality reduction methods. All of the methods we used are non-dynamic, meaning that they do not take into account time and the order of the sample points. Still, it could potentially be useful to create some more biologically plausible movement pattern of the simulated animal. This could, for example, affect the amount of environment explored during different experimental durations or allow for the use of dynamic dimensionality reduction methods in future studies.

In the results section, we stated that the duration of the trial did not have a considerable impact on the resulting neural manifolds. However, this is only true for the durations we tested, with 5 min being the minimum and 30 min being the maximum. To further investigate the effect of this factor, it would be necessary to compare different durations of less than 5 minutes with different increments of change.

5.3 Future Directions

This research can be used and/or elaborated in two general directions. The first is to use the results to inform and facilitate the design of experimental setups aimed at studying real place cell populations. Obviously, the outcomes of this work cannot be directly linked to the activity of actual hippocampal place cells and therefore no real conclusions can be drawn about them based on this work. However, the results of this research may provide useful information to neuroscientists planning to perform dimensionality reduction experiments on place cells in real

animals. For example, our findings can inform researchers about what potential results can be expected from the methods we employed. Our results can also inform experimentalists about how many place cells, and in what kind of environment, need to be recorded in order to obtain UMAP neural manifolds that can capture the experimental variables well. Otherwise, not capturing similar manifolds in real world experiments may also yield valuable insights that could potentially inform us about how populations of place cells do or do not function, e.g., what are the relevant experimental variables.

Another way of utilizing this work is a further elaboration of the model and employing other dimensionality reduction methods. First, the model would need to be refined, especially the movement patterns of the simulated animal. This would allow for the application of state-of-the-art non-linear dynamic dimensionality reduction methods designed exclusively for neural activity. These methods are Latent Factor Analysis via Dynamical Systems (LFADS) (Sussillo et al., 2016) and Targeted Neural Dynamical Modeling (TNDM) (Hurwitz et al., 2021). The great advantage of these methods is that they are dynamic, that is, they take into account the time and order of the sample points. These methods could potentially indicate what role time as a variable plays in the activity of place cell populations and what effect it may have on the role of other variables.

Bibliography

- Ahrens, M. B., Li, J. M., Orger, M. B., Robson, D. N., Schier, A. F., Engert, F., and Portugues, R. (2012). Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*, 485(7399):471–477.
- Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M., and Keller, P. J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods*, 10(5):413–420.
- Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.
- Briggman, K. L., Abarbanel, H. D., and Kristan Jr, W. (2005). Optical imaging of neuronal populations during decision-making. *Science*, 307(5711):896–901.
- Broome, B. M., Jayaraman, V., and Laurent, G. (2006). Encoding and decoding of overlapping odor sequences. *Neuron*, 51(4):467–482.
- Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature neuroscience*, 7(5):456–461.
- Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature neuroscience*, 7(5):446–451.
- Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., and Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520.
- Churchland, M. M., Byron, M. Y., Sahani, M., and Shenoy, K. V. (2007). Techniques for extracting single-trial activity patterns from large-scale neural recordings. *Current opinion in neurobiology*, 17(5):609–618.

- Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, 487(7405):51–56.
- Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., and Shenoy, K. V. (2010a). Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron*, 68(3):387–400.
- Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., et al. (2010b). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378.
- Cohen, M. R. and Maunsell, J. H. (2010). A neuronal population measure of attention predicts behavioral performance on individual trials. *Journal of Neuroscience*, 30(45):15241–15253.
- Cunningham, J. P. and Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509.
- Ferbinteanu, J. and Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron*, 40(6):1227–1239.
- Foster, D. J. and Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683.
- Gao, P. and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155.
- Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., Moser, M.-B., and Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.
- Harvey, C. D., Coen, P., and Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62–68.

- Hennig, M. H. (2022). The sloppy relationship between neural circuit structure and function. *The Journal of Physiology*.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.
- Hurwitz, C., Srivastava, A., Xu, K., Jude, J., Perich, M., Miller, L., and Hennig, M. (2021). Targeted neural dynamical modeling. *Advances in Neural Information Processing Systems*, 34:29379–29392.
- Johnson, A. and Redish, A. D. (2007). Neural ensembles in ca3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, 27(45):12176–12189.
- Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.
- Kaufman, M. T., Churchland, M. M., Ryu, S. I., and Shenoy, K. V. (2014). Cortical activity in the null space: permitting preparation without movement. *Nature neuroscience*, 17(3):440–448.
- Kerr, J. N. and Denk, W. (2008). Imaging in vivo: watching the brain in action. *Nature Reviews Neuroscience*, 9(3):195–205.
- Khona, M. and Fiete, I. R. (2021). Attractor and integrator networks in the brain. *arXiv preprint arXiv:2112.03978*.
- Kipke, D. R., Shain, W., Buzsáki, G., Fetz, E., Henderson, J. M., Hetke, J. F., and Schalk, G. (2008). Advanced neurotechnologies for chronic neural interfaces: new horizons and clinical opportunities. *Journal of Neuroscience*, 28(46):11830–11838.
- Lee, E. K., Balasubramanian, H., Tsolias, A., Anakwe, S. U., Medalla, M., Shenoy, K. V., and Chandrasekaran, C. (2021). Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. *Elife*, 10:e67490.
- Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62(3):413–425.
- Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474):78–84.
- Mazor, O. and Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*, 48(4):661–673.

- McClain, K., Tingley, D., Heeger, D. J., and Buzsáki, G. (2019). Position–theta-phase model of hippocampal place cell activity applied to quantification of running speed modulation of firing rate. *Proceedings of the National Academy of Sciences*, 116(52):27035–27042.
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- Mitchell-Heggs, R., Prado, S., Gava, G. P., Go, M. A., and Schultz, S. R. (2022). Neural manifold analysis of brain circuit dynamics in health and disease. *arXiv preprint arXiv:2203.11874*.
- O’Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1):78–109.
- O’Keefe, J. and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, 381(6581):425–428.
- O’Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*.
- Pandarinath, C., O’Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., Trautmann, E. M., Kaufman, M. T., Ryu, S. I., Hochberg, L. R., et al. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815.
- Paninski, L. and Cunningham, J. P. (2018). Neural data science: accelerating the experiment–analysis–theory cycle in large-scale neuroscience. *Current opinion in neurobiology*, 50:232–241.
- Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Recanatesi, S., Farrell, M., Lajoie, G., Deneve, S., Rigotti, M., and Shea-Brown, E. (2021). Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature communications*, 12(1):1–13.

- Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590.
- Rubin, A., Sheintuch, L., Brande-Eilat, N., Pinchasof, O., Rechavi, Y., Geva, N., and Ziv, Y. (2019). Revealing neural correlates of behavior without behavioral measurements. *Nature communications*, 10(1):1–14.
- Saha, D., Leong, K., Li, C., Peterson, S., Siegel, G., and Raman, B. (2013). A spatiotemporal coding mechanism for background-invariant odor recognition. *Nature neuroscience*, 16(12):1830–1839.
- Shepherd, G. M. (2004). *The synaptic organization of the brain*. Oxford university press.
- Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, 372(6539):eabf4588.
- Stevenson, I. H. and Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139–142.
- Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2):364–375.
- Stopfer, M., Jayaraman, V., and Laurent, G. (2003). Intensity versus identity coding in an olfactory system. *Neuron*, 39(6):991–1004.
- Sun, G., Zhang, S., Zhang, Y., Xu, K., Zhang, Q., Zhao, T., and Zheng, X. (2019). Effective dimensionality reduction for visualizing neural dynamics by laplacian eigenmaps. *Neural Computation*, 31(7):1356–1379.
- Sussillo, D., Jozefowicz, R., Abbott, L., and Pandarinath, C. (2016). Lfads-latent factor analysis via dynamical systems. *arXiv preprint arXiv:1608.06315*.
- Tolhurst, D. J., Movshon, J. A., and Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, 23(8):775–785.
- Tombaz, T., Dunn, B. A., Hovde, K., Cubero, R. J., Mimica, B., Mamidanna, P., Roudi, Y., and Whitlock, J. R. (2020). Action representation in the mouse parieto-frontal network. *Scientific reports*, 10(1):1–14.

- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- Whiteway, M. R. and Butts, D. A. (2019). The quest for interpretable models of neural population activity. *Current opinion in neurobiology*, 58:86–93.
- Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S., Shenoy, K. V., and Sahani, M. (2008). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Advances in neural information processing systems*, 21.
- Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature reviews neuroscience*, 16(8):487–497.

Appendix A

Appendix

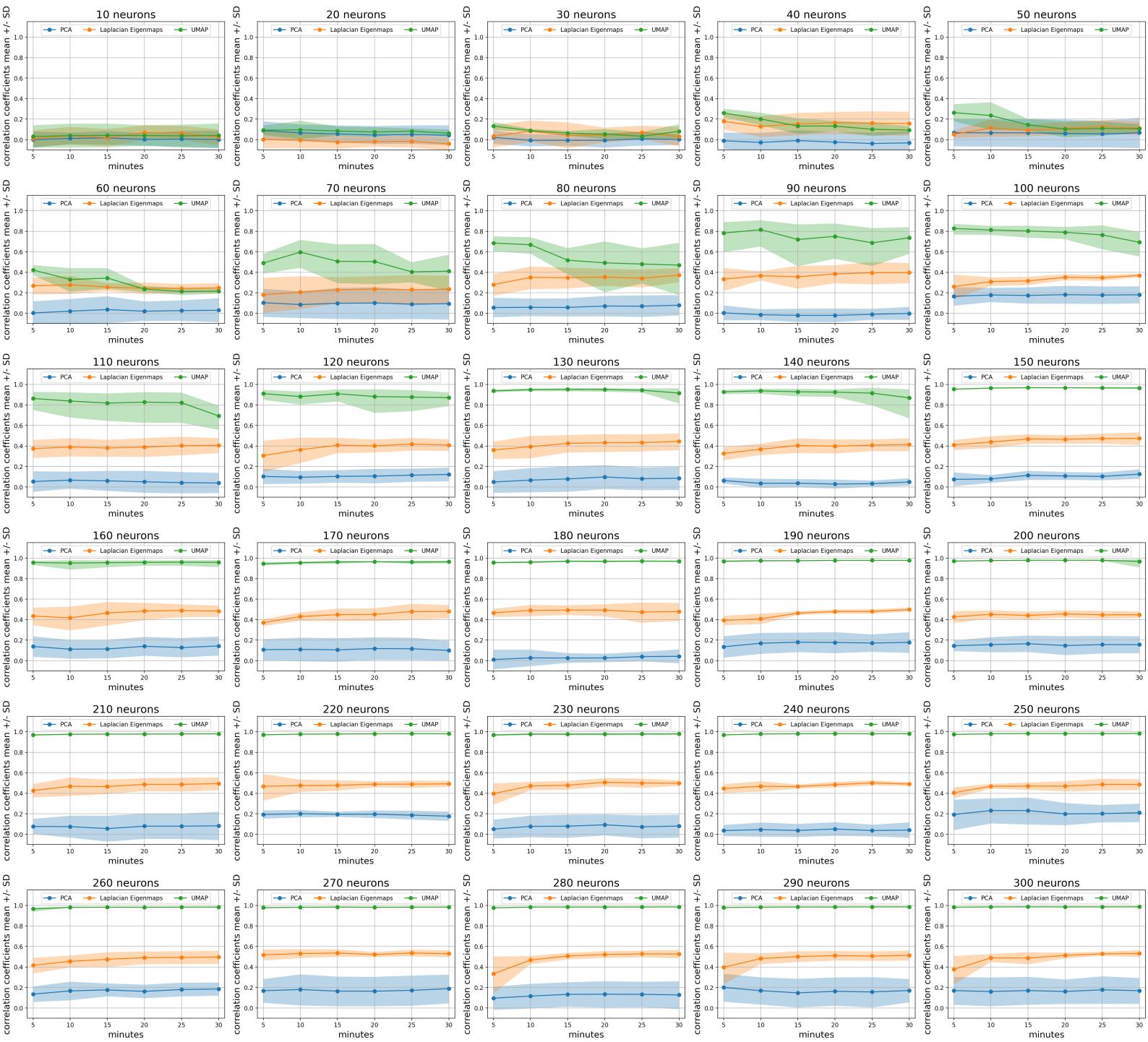


Figure A.1: The plots show the quality of the neural manifolds produced by PCA, LEM, and UMAP expressed by averaging the Pearson correlation coefficients over time per recorded number of neurons. The x-axis shows the duration of the trial. The y-axis shows the average quality of the neural manifolds. Each point on the graphs represents the average quality (+/- one standard deviation) of 5 manifolds for each dimensionality reduction method. Each graph shows the quality per number of neurons. These graphs contain all experimental data

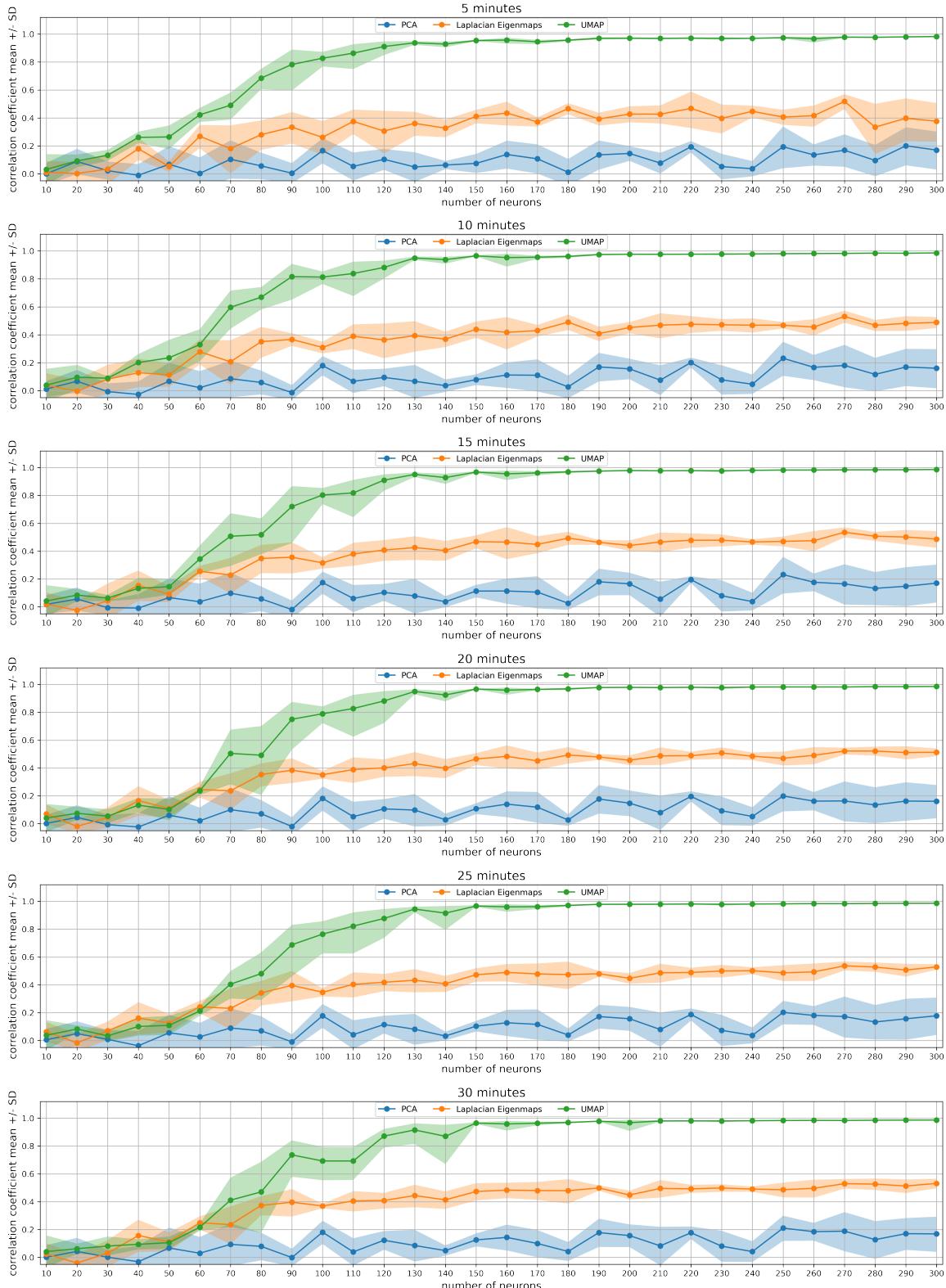


Figure A.2: The plots show the quality of the neural manifolds produced by PCA, LEM, and UMAP expressed by averaging the Pearson correlation coefficients over recorded number of neurons per trial duration. The x-axis shows the number of recorded place cells. The y-axis shows the average quality (+/- one standard deviation) of 5 manifolds for each dimensionality reduction method. Each graph shows the quality per duration of a trial. These graphs contain all experimental data