Imperial College
London

Curving the manifold: a recurrent neural network approach to understanding neural population dynamics in the motor cortex

Author:

Jorge Jacobo Bennasar Vázquez

B.Eng in Industrial Engineering (Electronics)

Comillas Pontifical University, 2019

Supervised by:

Juan Álvaro Gallego, PhD

A thesis submitted in partial fulfilment of the requirements for the award of MSc in Biomedical Engineering from Imperial College London

Department of Bioengineering

Imperial College London

September 2020

Word count: 5807

Abstract

During laboratory tasks, the dynamics of population activity in the sensorimotor cortices remain confined to lower-dimensional neural manifolds. These manifolds are typically identified using linear dimensionality reduction methods, implicitly assuming them as hyperplanes. However, since the cortex is made of networks of recurrently connected neurons with non-linear properties, population activity could be better captured by curved surfaces.

We investigated the geometry of neural manifolds by analysing activity in a rhesus macaque's primary motor cortex (M1) during a reaching behaviour. As predicted, curved manifolds consistently needed fewer dimensions to explain the neural activity. These results were most evident when manifolds were computed from large numbers of neurons. To verify that these non-linearities result from neural interconnectivity, we trained LSTMs with varying degrees of recurrent connectivity to perform the task. Only networks with high connectivities recapitulated all aspects of the data, illustrating how the recurrence in neural circuitry shapes the geometry of the manifold. Finally, we studied the relationship between behavioural complexity and manifold curvature, obtaining results that hint that both might be strongly correlated.

Our results indicate that neural population dynamics are better captured by curved manifolds, with such curvature arising from network connectivity and computational demands. Adopting non-linear dimensionality reduction methods may be especially important when studying a wider variety of more complex behaviours, or when analysing areas involved in more complicated computations.

Acknowledgements

Over the course of writing this thesis, I have received the support and assistance of many people around me, only to some of whom I can directly thank here.

First and foremost, I would like to express my deep gratitude to my supervisor Dr Juan Álvaro Gallego, whose invaluable insights have helped me to push the work presented here to a higher level. Your continuous support and advice have made me a better scientist, and I can proudly say that you will always be a role model as a researcher, advisor, and team leader.

My sincere thanks also go to all the members of the Gallego Lab. Your thoughtful suggestions and inquisitiveness have undoubtedly raised the quality of this project.

I would also like to acknowledge Dr Claudia Clopath and Dr A. Aldo Faisal, among other professors, for their inspiring lectures throughout my master's studies, which have fostered my love for Machine Learning and Neuroscience.

I wholeheartedly thank my parents, Jorge and Marta, and my partner, Sarah, for their unconditional support. You have helped me to be a happier person during these challenging times.

Last but not least, I would like to express my gratitude to the Endesa Foundation for funding my studies at Imperial College London.

Introduction

Since the work of Edward Evarts in the 1960s[1], neuroscientists have debated on the relationship between cortical activity and movement. For decades, approaches to understanding M1 computations relied on the assumption that neural activity encodes specific behavioural covariates[1,2,3,4,5,6,7,8]. The sought representations varied from force[1] to muscle dynamics[2,3] or movement direction[4,5]. These studies relied on models of non-interacting neurons and, barring exceptions, were initially focused on single-neuron activity. Later works shifted the focus onto population-level dynamics, nevertheless still presupposed that the motor cortex ought to represent movement, while its actual role is to generate it[7].

More recent studies conducted by Churchland, Shenoy and colleagues[9,10,11,12,13] suggest that, while it is true that some neurons exhibit high correlation with movement-related covariates, explicit representations may not be essential for movement generation. In their work, they suggest that cortical activity needs to be analysed from a different angle: the dynamical systems perspective. While traditional research aimed to determine what behavioural variables were controlled by M1, this new approach encourages to pay less attention to outputs and focus on the underlying neural dynamics that lead to movement.

This new outlook led to a fundamental change on how we understand the brain, where behaviour is no longer thought as the product of the action of many independently modulated neurons, but as the result of their interactions[14]. Under this standpoint, a novel theoretical framework for analysing population activity emerged: the neural manifold hypothesis[15]. This framework suggests that the underlying dynamics that govern movement ought to be confined within lower-dimensional manifolds spanned by 'neural modes', covariance patterns which arise from the coordinated activity of groups of neurons (Figure 1a). Hence, according to this framework, motor behaviours can be successfully explained by a much smaller number of signals, the 'latent variables', which describe the dynamics of the neural modes[16,17] (Figure 1b).

Gallego and colleagues supported this theory in ref. 18, where they demonstrated that the underlying dynamics associated with a particular task are stable through time. Interestingly, this stability was not observed as consistently in single neurons[19,20,21], implying that neural activity lies within reasonably steady manifolds[22]. In 2014, Sadtler et al.[23] went further by showing that neural manifolds might be

involved in shaping learning. In their study, monkeys had an easier time when learning tasks by using neural activity patterns comprised within a manifold, in comparison to patterns outside of it. This research hints that learning itself could consist in the re-shaping of already built neural manifolds (e.g. by synaptic plasticity[24,25,26]).

Inspired by this new philosophy, multiple studies have shown that motor behaviours are easily explained by a relatively small number of signals – commonly linear combinations of the activity of multiple neurons – obtained with a variety of dimensionality reduction methods[16,17,18,23,27,28,29,30].

While the neural manifold hypothesis has led to remarkable advances in the understanding of motor behaviour generation, it has also raised a variety of questions that must be tackled. Until now, most studies have applied linear dimensionality reduction methods, such as principal component analysis (PCA) or factor analysis (FA), to obtain the neural modes and its corresponding task-specific manifolds[18,23,27,28] (Figure 1c). These methods have undoubtedly provided satisfactory results with activity derived from standard laboratory tasks. However, it is still an open question whether neural manifolds are essentially flat or appear to be because we are 'sampling' a small section of a geometrically complex (curved) manifold (Figure 1d). It is reasonable to think that, when analysing simple task-specific neural dynamics, a local linear approximation to a curved manifold would work appropriately. Nevertheless, for more complex behaviours generated by dynamics embedded in higher-dimensional manifolds, the divergence between the performance of linear and non-linear methods might increase. Studies have shown that other cortical areas, such as V1, require high-dimensional hyperplanes to be explained[31]. Their dynamics could perhaps be better captured by lower-dimensional non-flat manifolds.

To address these questions, we first studied if neural manifolds present significant non-linearities. We began by comparing the performance of linear (PCA) and non-linear (ISOMAP[32]) dimensionality reduction methods at capturing the latent dynamics of neural activity (see Methods). Curved manifolds learned by ISOMAP provided consistently better results than PCA, needing fewer dimensions to explain the dynamics. This trend was particularly evident when analysing larger groups of neurons. We subsequently investigated whether these non-linearities were a by-product of overfitting the activity of some neurons or, on the other hand, are inherent to neural manifolds. To do this, we studied the similarity between the

underlying dynamics of different subsets of simultaneously recorded neurons for both ISOMAP and PCA. Non-linear dynamics showed higher consistency, suggesting that manifolds are essentially curved.

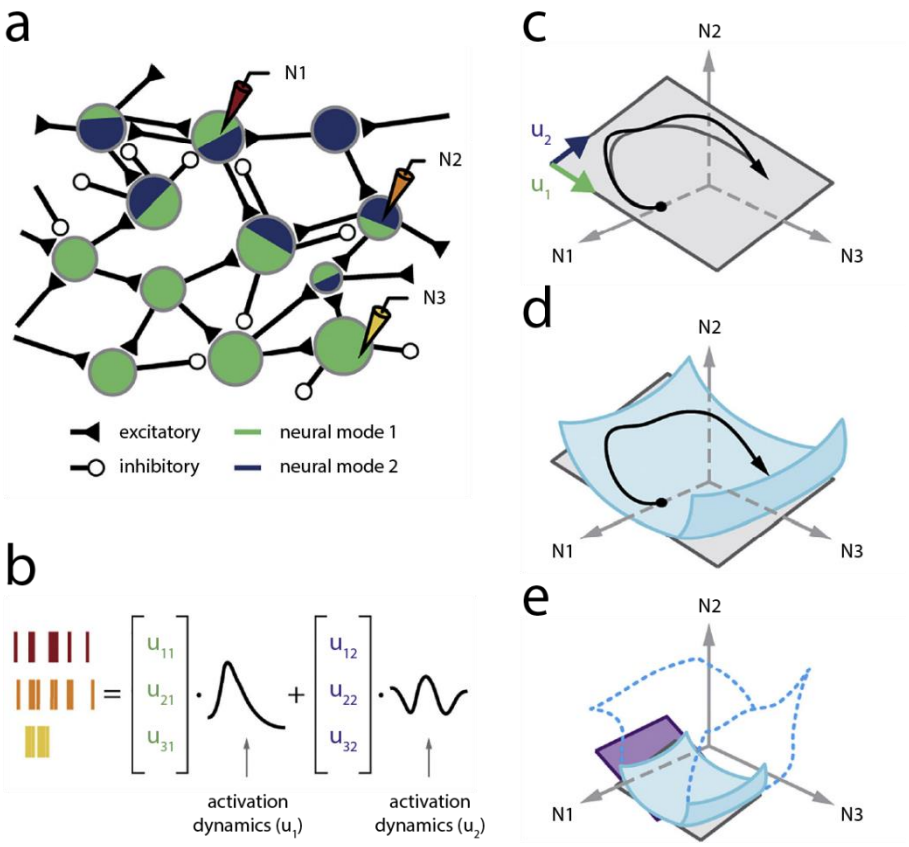


Figure 1. The neural manifold hypothesis.

Figure in ref. 15. Used here with the permission of its authors.

(**a**) Participation of each individual neuron on the underlying dynamics of the neural modes. The coloured regions in each neuron illustrate the contribution of each latent variable to the neuron's activity. (**b**) Activity of three distinct neurons as linear combinations of two latent variables. (**c**) Time-varying trajectory of the activity of a population of three recorded neurons. A two-dimensional linear plane spanned by neural modes u_1 and u_2 almost succeeds at enclosing the dynamics. (**d**) A curved manifold successfully encloses the neural dynamics. A plane (or hyperplane) provides a linear approximation to a specific region of the manifold. (**e**) When analysing task-specific activity, we might be capturing a small section of a broader, highly non-linear manifold that confines dynamics involved in the generation of multiple motor behaviours.

We then examined why these non-linearities arise in the neural dynamics. We hypothesised that manifold curvature could be associated with the recurrent interconnectivity between neurons[26,33,34,35]. To test this intuition, we proceeded to build LSTM models (long short-term memory networks[36,37]) with varying degrees of connectivity to replicate the behavioural task. We built these models inspired by previous research on the application of recurrent neural networks (RNNs) to simulate neural population activity[38,39,40,41,42,43,44,45]. Only high recurrent connectivities emulated all trends observed in the recorded data, exemplifying how the connectivity of neural circuitry shapes the curvature of manifolds.

Lastly, we considered if behavioural complexity is associated with non-linear neural dynamics. With this in mind, we analysed the geometry of manifolds corresponding to datasets consisting of single and multiple reaching targets. The latter consistently presented stronger non-linearities, suggesting that the analysis of activity derived from a wider variety of tasks may lead to the detection of more curved manifolds that span a larger range of neural dynamics (Figure 1e).

Hence, we established that neural manifolds most likely present inherent non-linearities, which seem to increase when analysing large neural populations and complex behaviours. These non-linear dynamics appear to be caused – at least partially – by the recurrent connectivity characteristic of neural circuitry.

Results

Behavioural task

We recorded neural activity in the primary motor cortex (M1) of a rhesus macaque (Monkey C) while it performed an eight-target centre-out reaching task (Figure 2a) (for recording specifications and neural data processing, see Methods). For the analysis, we employed data from eight different datasets of neural activity in the right (C_R) and left (C_L) hemispheres of M1. Each trial began with the monkey holding at a central position. An outer target was then presented, yet the monkey needed to withhold action until an auditory 'go cue' took place. After this delay period (which was only present in C_L), the monkey moved a manipulandum to the required target and held the position for half a second to receive a liquid reward. Figure 2b shows the neural activity and hand velocity of the monkey while it performed the task.

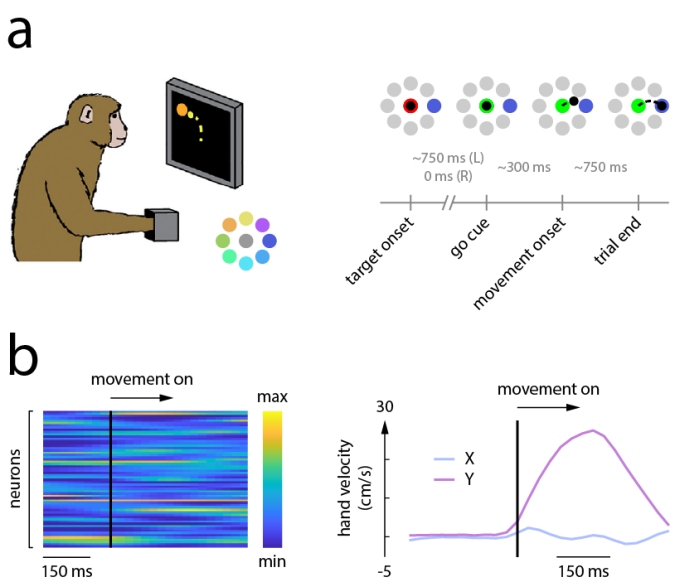


Figure 2. Behavioural task.

(**a**) Monkey C used a planar manipulandum to perform an eight-target centre-out reaching task. The schematic illustrates the mean duration of each of the task's stages for both C_L and C_R. (**b**) Neural activity and hand velocity of the monkey for a given trial.

Neural manifolds present non-linear dynamics

To investigate the geometry of neural manifolds, we systematically compared the performance of both PCA and ISOMAP at explaining the data. To evaluate the algorithms fairly, we measured the efficacy of both methods by analysing their explained variance distributions (as in ref. 46, see Methods). Figure 3a shows how ISOMAP needed fewer dimensions than PCA to explain the neural dynamics, implying that manifolds are highly curved. These results were consistent throughout all datasets, as illustrated in Supplementary Figure 1a.

We then examined if the size of the analysed neural populations correlated with the non-linearities found in the latent dynamics. To do this, we first determined the estimated dimensionality of the computed manifolds (see Methods). We then calculated the ratio between the dimensionalities provided by ISOMAP and PCA (what we call the 'non-linearity index' – μ –) for groups of different population sizes. The results revealed that the performance dissimilarity increases – in favour of ISOMAP – with the number of neurons of the analysed sample, suggesting that non-linear methods might perform notably better for larger populations. (Figure 3b, Supplementary Figure 1b).

To prove that manifold curvature is not a consequence of overfitting the data, we proceeded to study if ISOMAP provided more robust underlying dynamics (see Methods). We did this by comparing the r-squared similarity between the latent variables acquired from different simultaneously recorded neural populations – of ten neurons – for both dimensionality reduction methods. If ISOMAP is truly capturing better the manifold, then its similarity between subgroups should be higher than that of PCA (Supplementary Figure 2a). Figure 3c shows how this is the case for both C_L and C_R (for the individual results of each dataset, see Supplementary Figure 2b). Additionally, we did the same analysis with groups of twenty neurons. We observed that the divergence between the performances of ISOMAP and PCA increased, while the latent variables learned by both methods also presented notably higher similarities (Figure 3d). These results further prove that larger populations exhibit higher manifold curvatures while also yielding better approximations of the underlying dynamics. Naturally, analysing the activity of small sets of neurons would provide acceptable linear manifold estimates for simple tasks[47]. However, this might not be the case for complex behaviours, that would most likely require more extensive recordings to decode their dynamics robustly.

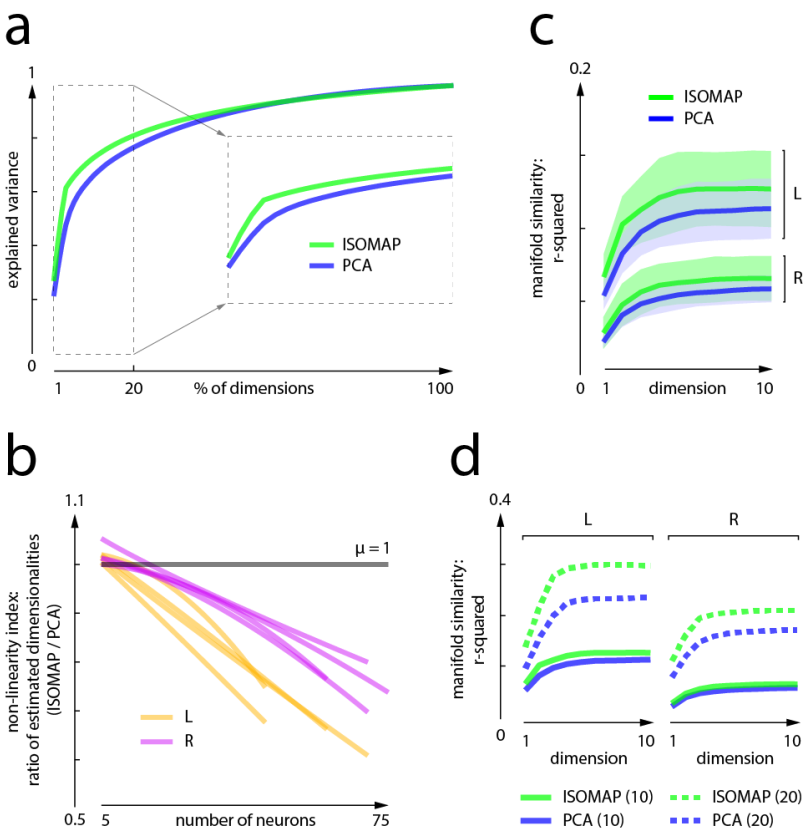


Figure 3. Analysing neural manifold geometry.

(**a**) Explained variance obtained by both ISOMAP and PCA for a C_L dataset. ISOMAP needs fewer dimensions to explain most of the neural dynamics' variance. (**b**) Relationship between the non-linearity index and the number of neurons employed to calculate the manifold. The analysis of larger populations translates into higher non-linearities. (**c**) Similarity (r-squared) between latent dynamics learned from different simultaneously recorded groups of ten neurons. As visible, ISOMAP yields better results in both C_L and C_R. (**d**) Increasing the number of neurons that belong to each subgroup (to twenty) results in performance improvements for both ISOMAP and PCA. The breach between both also widens.

An LSTM-based approach to modelling neural activity

After observing that neural manifolds present significant non-linearities, we proceeded to study why this is the case. Inspired by the work of Sussillo et al.[40], we built an LSTM model (Figure 4a) to replicate the behavioural task and investigate how various parameters affected the simulated dynamics. By doing this, we

considered that we could gain new intuitions on what shapes manifold geometry. We chose LSTMs over other RNN models because they tend to outperform most neural network architectures in neural decoding tasks[48].

The network, given an input, had to predict the velocity of the monkey's hand in the X and Y directions. To provide some variability between trials to the same target, we did not make the inputs to the LSTM condition-specific – or, in this case, target-specific –, as it was previously done in ref. 40. Instead, we constructed them from the preparatory activity of each individual trial (see Methods, Supplementary Figure 3). To illustrate this, Figure 4b shows how the first two principal components of the neural activity employed to design the inputs, while granting some variability, also provide useful information about the target to be reached. These inputs allowed us to simulate single trials instead of trial-averaged dynamics.

After assuring that the prediction performance was satisfactory – this is, it had a variance-accounted-for greater than 70% for the testing dataset –, we rectified the network activations (the activity of the LSTM units) and processed them as we previously did with the recorded data (see Methods). These processed activations were the signals that we used as simulations of the neural activity.

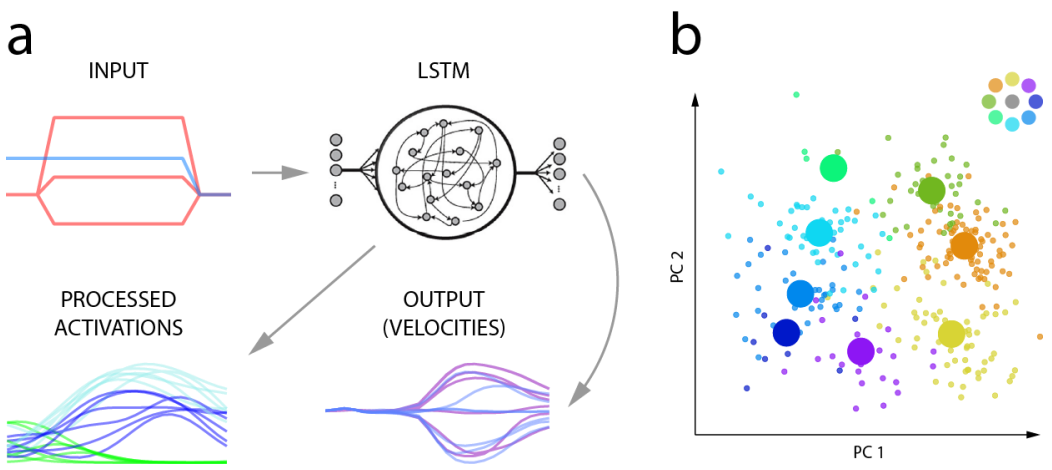


Figure 4. The LSTM model.

(**a**) Simple schematic of the LSTM model. (**b**) Sum over time of the first two PCs of preparatory activity – of the neural recordings – for each trial. The colours are associated with the eight different targets, while the bigger dots represent the mean across trials with the same objective.

Neural recordings share significant similarities with the model

As in ref. 40, the LSTM was not trained to imitate neural responses. In the model, the only constraint the network had was to predict the monkey's hand velocity successfully. Due to this, we could make comparisons between the network activations and the neural activity without worrying about possible biases.

To analyse the resemblance between the simulated units and real neurons, we first qualitatively compared their trial-averaged (Figure 5a) and single-trial activity (Figure 5b). In both cases, we observed considerable similarities. For instance, M1 and the model show a relatively constant preparatory activity, with large dynamical variations around movement onset. In Figure 5b, where we show neural and network activity for several trials with different targets, it is visible how model units – and neurons – are target-tuned, yet also present inter-trial variability. As the recorded data is not entirely target-specific, this makes the model better at replicating the neural activity. The latent dynamics of the trial-averaged activity also shared a strong resemblance (Figure 5c), although we observed a higher smoothness on the model data. This result reflects the success of the model at yielding highly similar dynamics across the first – and most important – dimensions.

We then quantitatively compared the model data with the recorded neural activity by using Procrustes analysis (see Methods). In short, Procrustes analysis is a statistical method used to compare the shapes of two or more objects. It does so by optimally translating, rotating, and scaling the data. In this case, we used it to quantify the dissimilarity between different datasets. We first compared the neural dynamics with trained and untrained models (Figure 5d). The trained networks shared significantly more similarities with the recorded activity, proving that the model performs better than random (baseline) networks.

We also compared the performance of trained networks with surrogate distributions obtained by shuffling the data in different ways (through time, targets, and both time and targets). When necessary, we also smoothed the data to match the Fourier spectrum of the model's unshuffled activity. In Figure 5e, we show how the surrogate (shuffled) datasets have higher dissimilarities with the neural activity. We confirmed this by using canonical correlation analysis (CCA), another frequently used method[40] that finds common patterns between datasets (see Methods). The results were consistent, with the non-shuffled activity yielding the highest canonical correlations (Figure 5f).

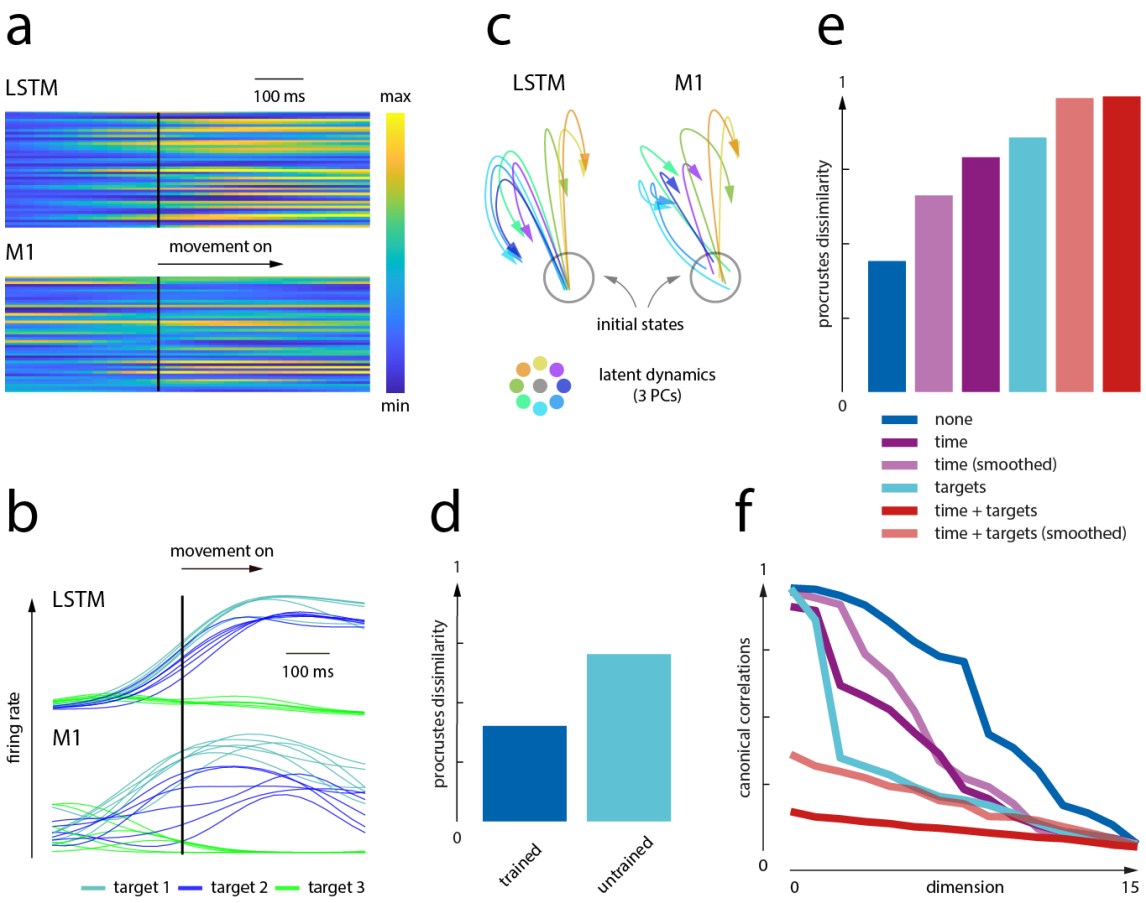


Figure 5. Model and neural data comparisons.

(**a**) Comparison of the trial-averaged activity of the LSTM model and the neural activity for a specific target. A significant shift in the dynamics takes place around the movement onset in both datasets. (**b**) Illustration of the activity of a single neuron and an LSTM unit for different trials and targets. (**c**) Latent dynamics (first three PCs) for the trial-averaged activity of the recorded data and the model. (**d**) Procrustes dissimilarity between the neural activity and trained and random (untrained) models. (**e**) Effects of shuffling the network activity on the Procrustes similarity between the model and the recorded data. (**f**) Same analysis as in (e), yet this time performed with canonical correlation analysis (CCA).

In view of the results, we deemed that the simulated activity shared significant similarities with the recorded neural data. Subsequently, we proceeded to use the model to study if network connectivity is a direct cause of manifold non-linearity.

Network connectivity shapes manifold geometry

To test the hypothesis that neural interconnectivity shapes manifold geometry, we added a connectivity parameter – φ – in the model (see Methods). With this parameter, we could modulate the percentage of propagating connections – or 'synapses' – between the network units, and therefore see how this altered the resultant dynamics.

First, we analysed how parameter φ affected the performance of ISOMAP and PCA at explaining the model. We did this by simulating networks with 100%, 50%, 25% and 12.5% connectivity. As previously hypothesised, we discovered that the connectivity of the model determined the non-linearity of its associated manifold, with only high connectivities yielding similar results to those obtained with the neural activity (Figure 6a). For low connectivities, PCA explained the simulated dynamics better than ISOMAP. We also observed these trends when investigating the relationship between the number of sampled units and the non-linearity index (Figure 6b). Non-linearities correlated positively with the population size, as they did with the neural data, solely when network units had abundant connections.

We then studied whether the connectivity of the model had an impact on how well ISOMAP and PCA computed consistent underlying dynamics from subsets of network activity. To do that, we compared the similarity between latent activity learned from different groups of ten units. To determine what models matched the neural data, we then calculated the ratio between the results obtained by ISOMAP and PCA. Again, only high connectivities provided results that resembled those observed in the primary motor cortex (Figure 6c).

Interestingly, the model compensated for low connectivity incrementing the 'synaptic' strength – or, more specifically, the propagation weights – of the available connections drastically (see Figure 6d, where we show the LSTM output weights – W_o – for high (100%) and low (1%) connectivities). This increment in 'synaptic' strength leads to a high homogeneity in the network, with a strongly correlated activity that favours planar manifolds. Correlations increase because, when parameter φ is low, units only have a few, really strong, connections, and are therefore prone to share high dynamical similarities with their neighbours. Figure 7a illustrates the impact of connectivity in the correlations between network units.

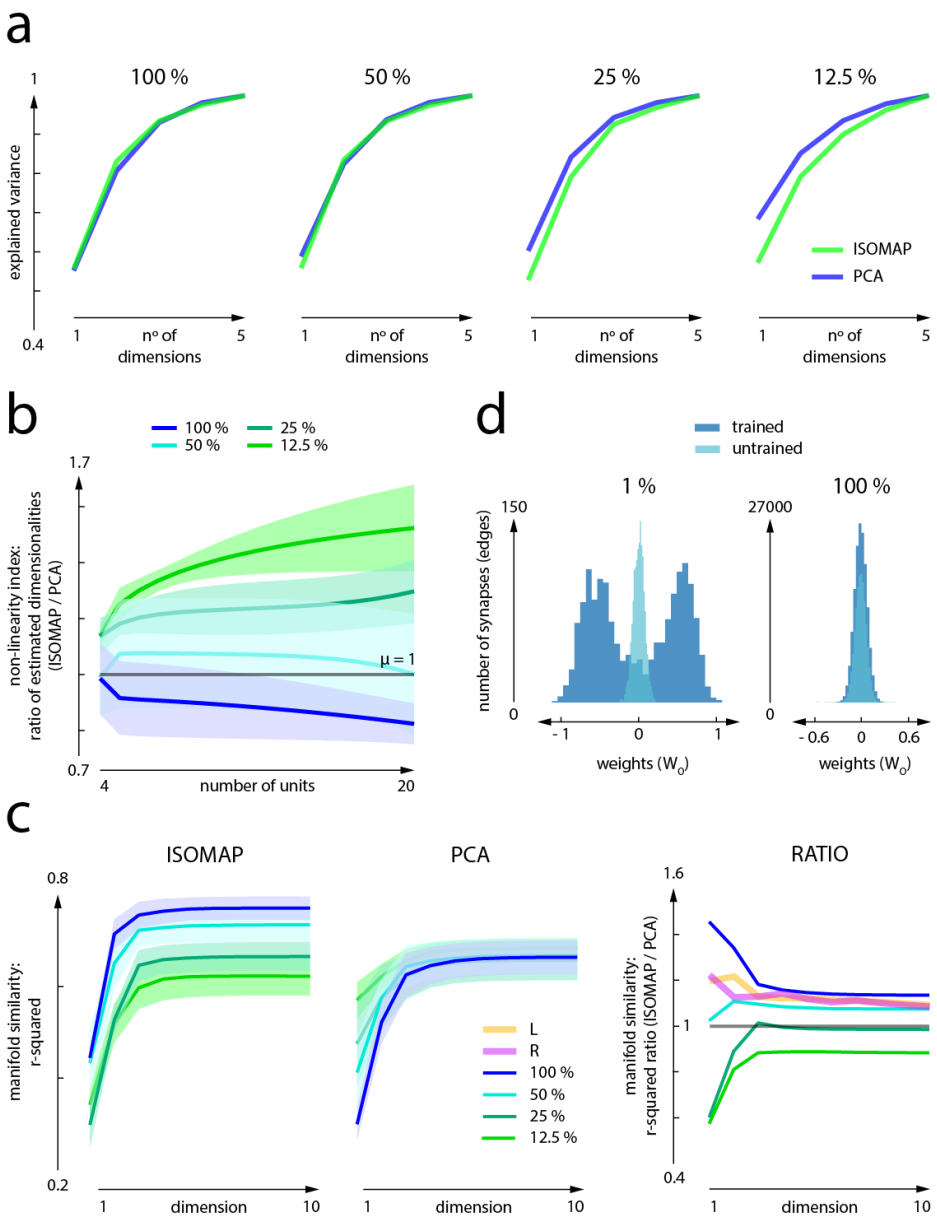


Figure 6. Network connectivity shapes manifold geometry.

(**a**) Explained variance for various connectivities (ISOMAP and PCA). (**b**) Relationship between the number of units and the ratio of estimated dimensionalities (non-linearity index) for different connectivities. (**c**) Similarity (r-squared) between latent dynamics derived from different groups of ten units for several connectivities and both ISOMAP and PCA. (**d**) Output weights for networks with 100% and 1% connectivity before and after training.

Additionally, we compared the correlation maps of the different models (Figure 7a) to those of the recorded M1 data (Figure 7b, Supplementary Figure 4), noticing that

high connectivities best replicated the correlation structure of the actual neural activity (Figure 7c). Figure 7d shows how a reduction of the correlation between units is heavily associated – $r^2 \approx 0.9$ – with higher manifold non-linearities.

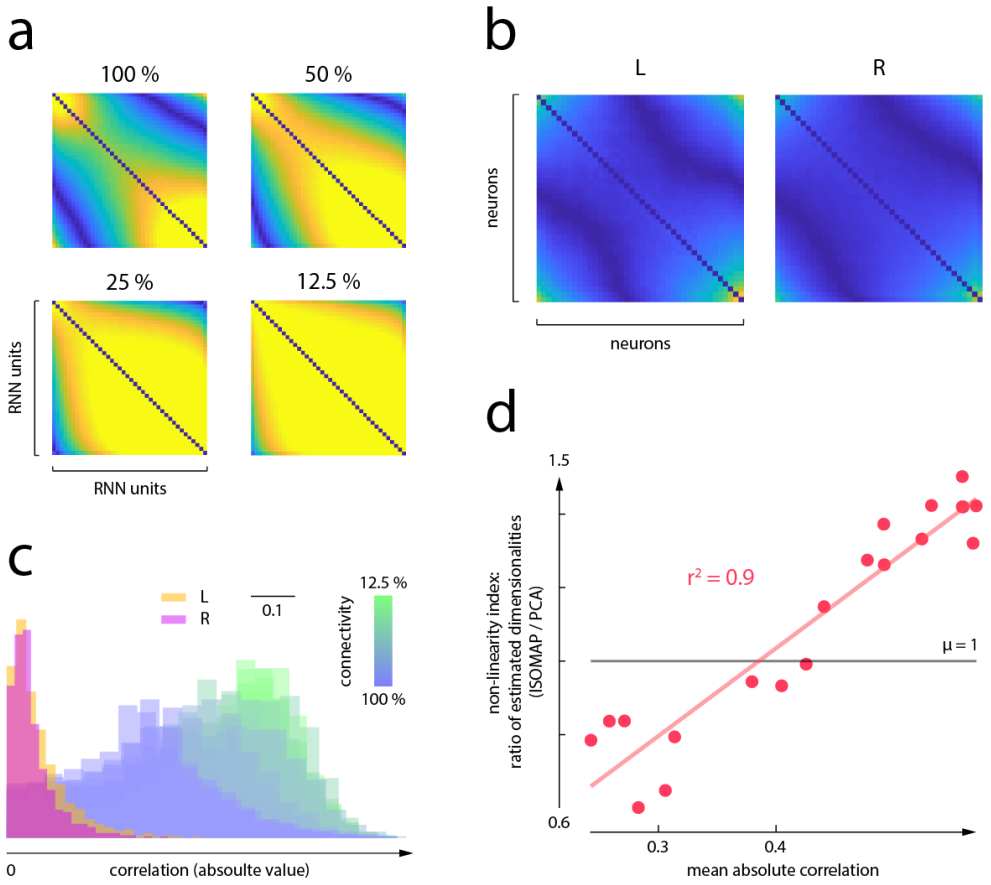


Figure 7. Correlation analysis.

(**a**) Correlation maps for 40 units (mean of multiple iterations) of models with various connectivities. (**b**) Same as in (a) but with neural activity. Each correlation map shows the mean over the four datasets of both C_L and C_R. (**c**) Histogram of the absolute value of the correlations for the recorded data and several models with varying connectivities. (**d**) Relationship between the non-linearity index and the mean absolute correlation between units of several LSTM models. As observed in the data, low correlations yield manifolds with stronger non-linearities.

To further demonstrate that correlations play a significant role in shaping manifold geometry, we also added a decorrelation parameter – ξ – to the LSTM model (see Methods). By incrementing parameter ξ, we managed to reduce the correlations between the network units, and thus increase the intrinsic non-linearities of the

manifold (Supplementary Figure 5). These results are consistent with Figure 7c: lowly correlated models show a higher resemblance with the neural dynamics.

Behavioural complexity and manifold non-linearities

Finally, we studied if behavioural complexity is associated with manifold curvature. Our intuition was that manifolds which span a wider variety of behaviours should present more prominent non-linearities than those associated with single actions. If this is the case, linear approximations should fail to perform as well when analysing more complex tasks, needing substantially more dimensions than curved manifolds to explain the neural activity (Figure 8a). We tested this hypothesis by comparing manifolds learned from the eight-target reaching task with others calculated from datasets comprised of trials in which the monkey always moved towards the same target. As we anticipated, multiple-target dynamics consistently showed sharper non-linearities, especially when increasing the number of neurons (Figure 8b).

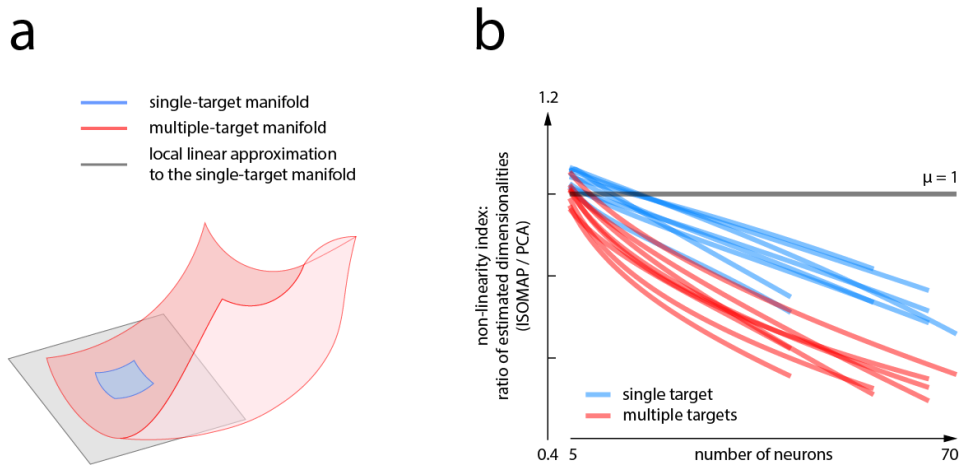


Figure 8. Behavioural complexity and manifold non-linearity.

(**a**) Analysing neural activity with higher task complexity might lead to the detection of broader, more curved manifolds that span wider dynamical ranges. (**b**) Relationship between the non-linearity index and the neural population size for multiple-target and single-target tasks. Across all datasets, multiple-target dynamics showed stronger non-linearities.

These results suggest that when investigating the dynamics underlying more complex behaviours – which may require recordings of more neurons –, or when

analysing brain areas with more complex computations, using non-linear dimensionality reduction methods could be necessary.

Discussion

Throughout the last decade, a large variety of studies have suggested that the neural manifold hypothesis is a valid framework to analyse motor behaviours. In this light, we proceeded to examine the geometries that span the underlying dynamics in a rhesus macaque's motor cortex (M1). We found that manifolds present significant non-linearities, with ISOMAP needing notably less neural modes than PCA to explain the recorded activity. Furthermore, these non-linear dynamics proved to be more consistent when examining different groups of neurons separately, and analysing larger neural populations yielded more curved manifolds. Thus, non-linearities are unlikely to be a result of overfitting or undersampling.

We also investigated the possible physiological causes of the inherent curvature of neural manifolds. Along with the non-linear properties of neurons, we found that neuronal interconnectivity could be a factor to take into consideration. Since RNNs have proven to be a valuable tool for simulating neural dynamics, we evaluated the role of connectivity by building a long short-term memory network (LSTM) model that replicated the behavioural task. The results were enlightening: network connectivity strongly determined manifold geometry, and only highly connected models mirrored all the trends observed in the recorded data.

Besides, we discovered that the curvature of neural manifolds is dependent on the complexity of the performed motor task. Hence, we might need to account for non-linearities when studying complex behaviours that require more extensive neural recordings. Currently, brain-computer interfaces[49,50,51,52] (BCIs) are capable of successfully decoding simple actions typical of standard laboratory tasks. However, factoring in the non-linearities of neural dynamics may be necessary when building naturalistic BCIs, capable of encompassing more complex and varied movements characteristic of unconstrained behaviour.

Regarding other brain areas, research on the visual cortex (V1) has revealed that hyperplanes may fall short at approximating its latent dynamics[31]. Interestingly, V1 has been discovered to be the most neuron-dense cortical area in several primate species[53], alluding to a complex circuitry that could require the utilisation of curved manifolds. Non-linear dimensionality reduction methods might also be necessary when analysing higher brain areas with more complex computations. For instance, activity in the prefrontal cortex (PFC), a region involved in decision making, behaviour planning and personality, seems to be best explained by the joint

dynamics of large neural populations[54]. We think that non-linear algorithms might better explain these complicated dynamics observed in PFC.

Likewise, it would be desirable to study how the coordination between brain areas, their inputs, and the hierarchies among them influence the geometry of their respective neural manifolds. In this regard, it would be interesting to analyse activity in the dorsal premotor cortex (PMd) – which communicates towards M1 – and compare its neural dynamics to those exhibited in this study. Furthermore, building two-layered RNN models with varying degrees of inter and intra-layer connectivity might be a useful framework to investigate how the complexity of the communications between areas[55] is related to manifold geometry.

Recording larger populations of neurons during naturalistic conditions, while subjects perform more complex and varied behaviours, may shed more light on the results presented here. This study should also be replicated with data recorded from other monkeys, alternative recurrent neural network architectures, and when analysing different behavioural tasks. Regarding the latter, we know that manifolds derived from different motor behaviours have similar properties[27], and therefore matching conclusions should be reached. Presumptively, other RNN models would also show trends alike to the ones exhibited here. Further investigating manifold geometry and dimensionality would most probably help to deepen our knowledge on how neural populations generate movement and other behaviours.

Methods

Recordings of physiological data

All procedures involving animals undertaken to record the data employed in this study complied with the ethical standards of Northwestern University's Institutional Animal Care and Use Committee.

After training the subject to perform the centre-out reaching task, we surgically implanted 96-channel electrode arrays in the primary motor cortex (M1) of a male rhesus macaque (Monkey C, *M. mulatta*) using standard surgical procedures. The monkey received two implants, one in each hemisphere of M1 (right – C_R – and left – C_L –). We recorded the data with a Blackrock Cerebus system (Blackrock Microsystems, Salt Lake, UT).

The raw data was then digitalised and band-pass filtered (250 – 5000 Hz). Subsequently, the recordings were processed using a spike sorting software (Offline Sorter v3, Plexon, Dallas, TX) to detect single-neuronal activity. To identify the spike times, we used a root-mean square based threshold ($5.5 \times \boldsymbol{RMS}$).

Neural data processing

Preceding the analysis, we processed the neural activity in several ways. First, we selected the trials in which the monkey was successful at reaching the correct target within the specified time, and thus received the corresponding reward. We then removed neurons with a low firing rate (< 0.1 Hz) and, lastly, we computed the smoothed firing rates of the neural activity by applying a Gaussian kernel (standard deviation of 100 ms) to the square-root-transformed binned (30 ms) data[18,28].

We employed recordings from eight different datasets – four of C_L and four of C_R – consisting of a varying number of neurons (49 – 77) and trials (136 – 386). The data was comprised of 300 ms of preparatory (pre-reaching) activity followed by another 450 ms recorded after the movement started, which was enough time for the monkey to reach the target.

For most analyses, we averaged the data acquired from multiple subsets of fifty randomly selected trials. Regarding the behavioural complexity analysis, however, we examined reduced groups of ten attempts. We did this due to a lack of more recorded trials towards specific targets in some datasets. Additionally, we averaged

the results obtained separately for the eight different targets to calculate the single-target data.

Principal component analysis (PCA)

Principal component analysis (PCA) is one of the most commonly used linear dimensionality reduction methods. PCA identifies uncorrelated – orthogonal – linear combinations of the original variables that successively maximise variance. It does so by computing the eigenvalues and eigenvectors of the covariance matrix of the data.

ISOMAP

ISOMAP is a non-linear low-dimensional embedding method described at length in ref. 32. In brief, ISOMAP extends multidimensional scaling (MDS) by incorporating geodesic distances induced by a neighbourhood graph. The n highest eigenvectors of the calculated geodesic distance matrix define the coordinates of the new n-dimensional Euclidean space. In our case, we used the twenty nearest neighbours (k = 20) to construct the neighbourhood graph, yet we obtained similar results with other values (12 – 30). We implemented ISOMAP using the MATLAB package from Tenenbaum et al.[32].

Explained variance and dimensionality estimation

To analyse the results yielded by ISOMAP and PCA, we first determined the variance explained by each of their predicted dimensions. We did this by dividing the eigenvalue of a given dimension by the total sum of eigenvalues. To estimate the dimensionality of each manifold, we then applied the participation ratio[47] – PR – to the eigenvalues. The participation ratio was calculated as follows:

$$PR = \frac{(\sum \lambda_i)^2}{\sum {\lambda_i}^2}$$

Where λ_i are the eigenvalues associated with the calculated dimensions. The participation ratio helped us to approximate the number of dimensions needed by each method to describe the neural dynamics. This difference in performance between ISOMAP and PCA was quantified with the ratio of estimated dimensionalities – μ – (or 'non-linearity index'):

$$\mu = \frac{PR_{ISOMAP}}{PR_{PCA}}$$

Therefore, we considered that ISOMAP performed better than PCA at explaining the data when μ < 1 (Supplementary Figure 6). We observed that with a small number of neurons – N < 10 – μ was roughly 1, suggesting linear manifolds. However, when increasing the sample size – N = 70 –, ISOMAP presented a significantly better performance ($\mu \approx 0.6$ for C_L and $\mu \approx 0.7$ for C_R). This tendency hints towards higher non-linearities with larger datasets.

Overfitting analysis

To determine if manifold non-linearities were a result of overfitting the data, we compared, for both ISOMAP and PCA, the consistency of the underlying dynamics calculated from different simultaneously recorded groups of neurons (Supplementary Figure 2). To do this, we generated multiple training and testing sets of ten (and twenty) neurons and studied the r-squared similarity between their Euclidean distance vectors – u – (L_2 norm matrices – U – of time bins – T – across all trials – m – reshaped into one dimension).

$$u \in \mathbb{R}^{1\times(T\times m)^2} \qquad U \in \mathbb{R}^{(T\times m)\times(T\times m)}$$

The analysis consisted of the following steps:

1. To begin with, we used the first ten neural modes, derived from either PCA or ISOMAP, to build the testing vector of distances – u^{test} –.
2. We then calculated the r-squared similarity between u^{test} and several distance vectors – u_n^{train} – constructed from the first n neural modes of the training dataset (from one to ten, computed with the same method as above). By doing this, we could also assess the contribution of each neural mode at providing similar dynamics between datasets.

Long short-term memory (LSTM) model

We used a long short-term memory (LSTM) architecture to build the model. LSTMs are a particular type of artificial recurrent neural networks (RNNs) characterised for having units which are composed of a cell and three different gates (input, output and forget). The cell is capable of remembering values over time – solving the vanishing gradient problem typically seen in conventional RNNs –, while the gates regulate the cell's flow of information. For more information on LSTMs, see ref. 36

and 37. We chose an LSTM over other types of recurrent networks because it tends to perform better at movement decoding tasks[48].

To train the LSTM we employed ADAM optimisation[56]. As for the network's non-linearity, we decided to use the hyperbolic tangent ($tanh$). The hyperparameters of the model and their values can be checked at Supplementary Table 1, Supplementary Table 2 and Supplementary Table 3.

We trained the network until the error – E –, which was measured by equation [1], reached a certain threshold. The model was then considered valid if, for the testing data, it reached a variance-accounted-for of at least 70% (Supplementary Figure 7).

$$E = \frac{1}{2 \times T \times M} \times \sum \left(\hat{Y} - Y\right)^2 \qquad [1]$$

With $\hat{Y}$ and Y being the predicted and measured velocities (respectively), T the number of time bins (T = 25) and M the mini-batch size (M = 8 trials).

Before comparing the neural activity to the model data (which were the activations observed when forward propagating the testing dataset after training the network), we performed several transformations to the latter. This additional processing started by rectifying the activations, which were later square-root-transformed and smoothed (as with the recordings).

Inputs to the LSTM

To generate some variability in the network for trials with the same target, we constructed the inputs to the model from individual attempts (as opposed to the standard trial-averaged-based inputs seen in other studies[40]). To build these inputs, we first applied PCA to several bins of preparatory activity (from 210 to 30 ms before movement) of each trial. After that, we selected the first D principal components ($D \in$ [3,6], depending on the dataset) and summed them over time. The resulting vector of values v (of size 1 × D) was then used to form the network inputs, such that:

$$X_i = 0 \text{ for } t < t_1 \text{ and } t \geq t_2 \qquad X_i = v_i \text{ for } t_2 > t \geq t_1$$

Where X is the input to the LSTM, t_1 the time in which the target is presented (t_1 = 4), and t_2 the bin in which movement starts (t_2 = 11). Both t_1 and t_2 are simulation parameters of the model. In this case, t_2 was designed to match the time in which the monkey commenced to move.

Furthermore, we also added a trial-independent hold cue defined as follows:

$$X_{D+1} = 0.5 \text{ for } t < t_2 \qquad X_{D+1} = 0 \text{ for } t \geq t_2$$

Therefore, X can be defined as $X \in \mathbb{R}^{I \times T}$, with $I = D + 1$.

We show a schematic of the input generation process in Supplementary Figure 3.

Varying network connectivity

To analyse the role of network interconnectivity in shaping manifolds, we introduced a connectivity parameter – φ – in the model. This parameter regulated the percentage of connections between units by tuning matrix G, which altered weight initialisation and backward propagation in the LSTM. Connectivity matrix G was formulated as follows:

$$G = [G_1 \mid G_2] \in \mathbb{R}^{H \times (H+I)}$$

$$G_1 = (g_{ij}) \in \mathbb{R}^{H \times H}$$

$$G_2 = (1) \in \mathbb{R}^{H \times I}$$

With H and I being the number of hidden units and inputs, respectively.

G_1 is a symmetrical (Erdös–Rényi) and binary matrix which multiplies the activations, being therefore responsible for the 'synaptic' propagation throughout the network. Each edge g_{ij} forms a connection with probability $\varphi \in [0,1]$, such that:

$$\varphi \approx \frac{\sum g_{ij}}{H^2}$$

G_2, conversely, multiplies the inputs. Thus, it is a matrix of ones.

Matrix G multiplies the propagation weights (W_f, W_i, W_c and W_o) in the initialisation [2] and their derivatives (dW_f, dW_i, dW_c and dW_o) during backward propagation [3]:

$$W_{f,i,c,o} = N_{f,i,c,o} \odot G \qquad [2]$$

$$dW_{f,i,c,o} = dW_{f,i,c,o} \odot G \qquad [3]$$

Where $N_{f,i,c,o}$ (N_f, N_i, N_c and N_o) are He-initialised random matrices[57].

Decorrelating the model

To reduce the correlations between the activity of different units of the model, we introduced a decorrelation parameter – ξ – in the LSTM. This parameter regulates

the effect of matrix R at reducing the weights of highly correlated (positively or negatively) units. Matrix R was defined as follows:

$$R = [R_1 \mid R_2] \in \mathbb{R}^{H\times(H+I)}$$

$$R_1 \in \mathbb{R}^{H\times H}$$

$$R_2 = (0) \in \mathbb{R}^{H\times I}$$

Where R_1 is the correlation matrix of the activations and R_2 a null matrix. With decorrelation, the derivatives of the propagation weights are updated (during backward propagation) as it is shown in [4]:

$$dW_{f,i,c,o} = \left(dW_{f,i,c,o} + \xi \times R \odot dW_{f,i,c,o}\right) \odot G \quad [4]$$

Note that the higher the correlation – both positive and negative – and ξ, the stronger the weight penalisation.

Procrustes and canonical correlation analyses

We used Procrustes analysis and CCA to quantify the similarity between the model and the recorded data. For both methods, the inputs were the first fifteen principal components of the two datasets. By doing this, we could compare their underlying dynamics instead of individual activity. In canonical correlation analysis, applying PCA to the data also allowed us to assess the similarity for each dimension considering its respective importance at explaining the dynamics.

References

1. Evarts, E. V. Relation of pyramidal tract activity to force exerted during voluntary movement. *J. Neurophysiol.* 31, 14–27 (1968).

2. Todorov, E. Direct cortical control of muscle activation in voluntary arm movements: A model. *Nat. Neurosci.* 3, 391–398 (2000).

3. Mussa-Ivaldi, F. A. Do neurons in the motor cortex encode movement direction? An alternative hypothesis. *Neurosci. Lett.* 91, 106–111 (1988).

4. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* 2, 1527–1537 (1982).

5. Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. Neuronal population coding of movement direction. *Science (80-.).* 233, 1416–1419 (1986).

6. Aflalo, T. N. & Graziano, M. S. A. Relationship between unconstrained arm movements and single-neuron firing in the macaque motor cortex. *J. Neurosci.* 27, 2760–2780 (2007).

7. Scott, S. H. Inconvenient Truths about neural processing in primary motor cortex. *J. Physiol.* 586, 1217–1224 (2008).

8. Schwartz, A. B. Direct cortical representation of drawing. *Science (80-.).* 265, 540–542 (1994).

9. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* 487, 51–56 (2012).

10. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical Control of Arm Movements: A Dynamical Systems Perspective. *Annu. Rev. Neurosci.* 36, 337–359 (2013).

11. Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I. & Shenoy, K. V. Cortical Preparatory Activity: Representation of Movement or First Cog in a Dynamical Machine? *Neuron* 68, 387–400 (2010).

12. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: Permitting preparation without movement. *Nat. Neurosci.* 17, 440–448 (2014).

13. Rokni, U. & Sompolinsky, H. How the brain generates movement. *Neural Comput.* 24, 289–331 (2012).

14. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. Computation Through Neural Population Dynamics. *Annu. Rev. Neurosci.* 43, 249–275 (2020).

15. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural Manifolds for the Control of Movement. *Neuron* vol. 94 978–984 (2017).

16. Pandarinath, C. *et al.* Latent factors and dynamics in motor cortex and their application to brain–machine interfaces. *J. Neurosci.* 38, 9390–9401 (2018).

17. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience* vol. 17 1500–1509 (2014).

18. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* 23, 260–270 (2020).

19. Rokni, U., Richardson, A. G., Bizzi, E. & Seung, H. S. Motor Learning with Unstable Neural Representations. *Neuron* 54, 653–666 (2007).

20. Stevenson, I. H. *et al.* Statistical assessment of the stability of neural movement representations. *J. Neurophysiol.* 106, 764–774 (2011).

21. Chestek, C. A. *et al.* Single-neuron stability during repeated reaching in macaque premotor cortex. *J. Neurosci.* 27, 10742–10750 (2007).

22. Degenhart, A. D. *et al.* Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity. *Nat. Biomed. Eng.* 4, 672–685 (2020).

23. Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* 512, 423–426 (2014).

24. Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: From synapses to maps. *Annu. Rev. Neurosci.* 21, 149–186 (1998).

25. Kaas, J. H. Plasticity of Sensory and Motor Maps in Adult Mammals. *Annu. Rev. Neurosci.* 14, 137–167 (1991).

26. Stepanyants, A., Hof, P. R. & Chklovskii, D. B. Geometry and structural plasticity of synaptic connectivity. *Neuron* 34, 275–288 (2002).

27. Gallego, J. A. *et al.* Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat. Commun.* 9, 1–13 (2018).

28. Yu, B. M. *et al.* Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* 102, 614–635 (2009).

29. Pandarinath, C. *et al.* Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 15, 805–815 (2018).

30. Farshchian, A. *et al.* Adversarial Domain Adaptation for Stable Brain-Machine Interfaces. *7th Int. Conf. Learn. Represent. ICLR 2019* (2018).

31. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* 571, 361–365 (2019).

32. Tenenbaum, J. B., de Silva, V. & Langford, J. C. *A Global Geometric Framework for Nonlinear Dimensionality Reduction*. *Philos. Trans. R. Soc. London Ser. B* vol. 67 www.sciencemag.org (1995).

33. Chklovskii, D. B. Synaptic connectivity and neuronal morphology: Two sides of the same coin. *Neuron* 43, 609–617 (2004).

34. Rgy Buzsá Ki, G., Geisler, C., Henze, D. A. & Wang, X.-J. Interneuron Diversity series: Circuit complexity and axon wiring economy of cortical interneurons. doi:10.1016/j.tins.2004.02.007.

35. Ringo, J. L. Neuronal interconnection as a function of brain size. *Brain, behavior and evolution* vol. 38 1–6 (1991).

36. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* 9, 1735–1780 (1997).

37. Gers, F. A., Schmidhuber, J. & Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* 12, 2451–2471 (2000).

38. Richards, B. A. *et al.* FOCUS | PersPective A deep learning framework for neuroscience. *Nat. Neurosci.* 16, 42.

39. Sussillo, D. & Abbott, L. F. Generating Coherent Patterns of Activity from Chaotic Neural Networks. *Neuron* 63, 544–557 (2009).

40. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* 18, 1025–1033 (2015).

41. Sussillo, D. *et al.* A recurrent neural network for closed-loop intracortical brain-machine interface decoders. *J. Neural Eng.* 9, 026027 (2012).

42. Sussillo, D. & Barak, O. Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.* 25, 626–649 (2013).

43. Feulner, B. & Clopath, C. Neural manifold under plasticity in a goal driven learning behaviour. 1–22 (2020) doi:10.1101/2020.02.21.959163.

44. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* 16, 925–933 (2013).

45. Depasquale, B., Cueva, C. J., Rajan, K., Escola, G. S. & Abbott, L. F. full-FORCE : A Target-Based Method for Training Recurrent Networks arXiv : 1710 . 03070v1 [cs . NE] 9 Oct 2017. 1–20 (2018) doi:10.5281/zenodo.1154965.Funding.

46. Yan, Y., Goodman, J. M., Moore, D. D., Solla, S. A. & Bensmaia, S. J. Unexpected complexity of everyday manual behaviors. *Nat. Commun.* 11, 1–8 (2020).

47. Gao, P. *et al.* A theory of multineuronal dimensionality, dynamics and measurement. 214262 (2017) doi:10.1101/214262.

48. Glaser, J. I. *et al.* Machine learning for neural decoding. (2017).

49. Golub, M. D., Chase, S. M., Batista, A. P. & Yu, B. M. Brain-computer interfaces for dissecting cognitive processes underlying sensorimotor control. *Current Opinion in Neurobiology* vol. 37 53–58 (2016).

50. Oby, E. R. *et al.* New neural activity patterns emerge with long-term learning. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15210–15215 (2019).

51. Ganguly, K. & Carmena, J. M. Emergence of a Stable Cortical Map for Neuroprosthetic Control. *PLoS Biol.* 7, e1000153 (2009).

52. Bensmaia, S. J. & Miller, L. E. Restoring sensorimotor function through intracortical interfaces: Progress and looming challenges. *Nature Reviews Neuroscience* vol. 15 313–325 (2014).

53. Collins, C. E., Airey, D. C., Young, N. A., Leitch, D. B. & Kaas, J. H. Neuron densities vary across and within cortical areas in primates. *Proc. Natl. Acad. Sci. U. S. A.* 107, 15927–15932 (2010).

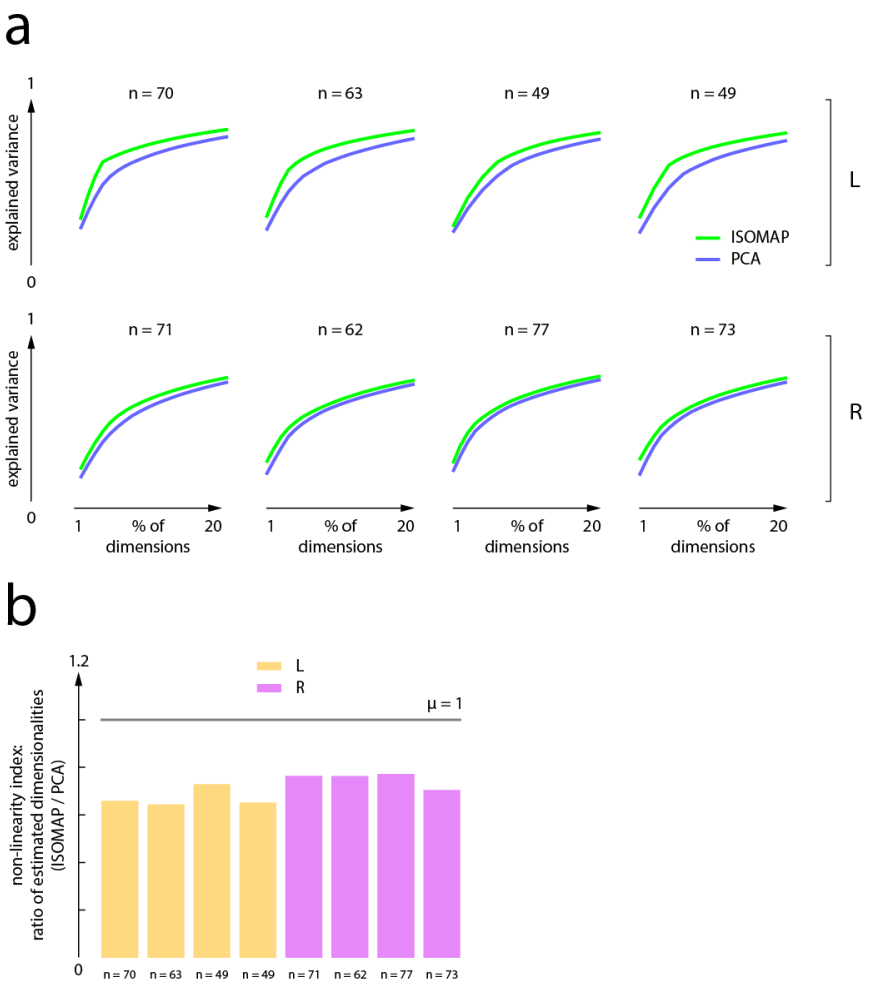
54. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84 (2013).

55. Joã Semedo, A. D. *et al.* Cortical Areas Interact through a Communication Subspace Article Cortical Areas Interact through a Communication Subspace. *Neuron* 102, 249-259.e4 (2019).

56. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic optimization. *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.* 1–15 (2015).

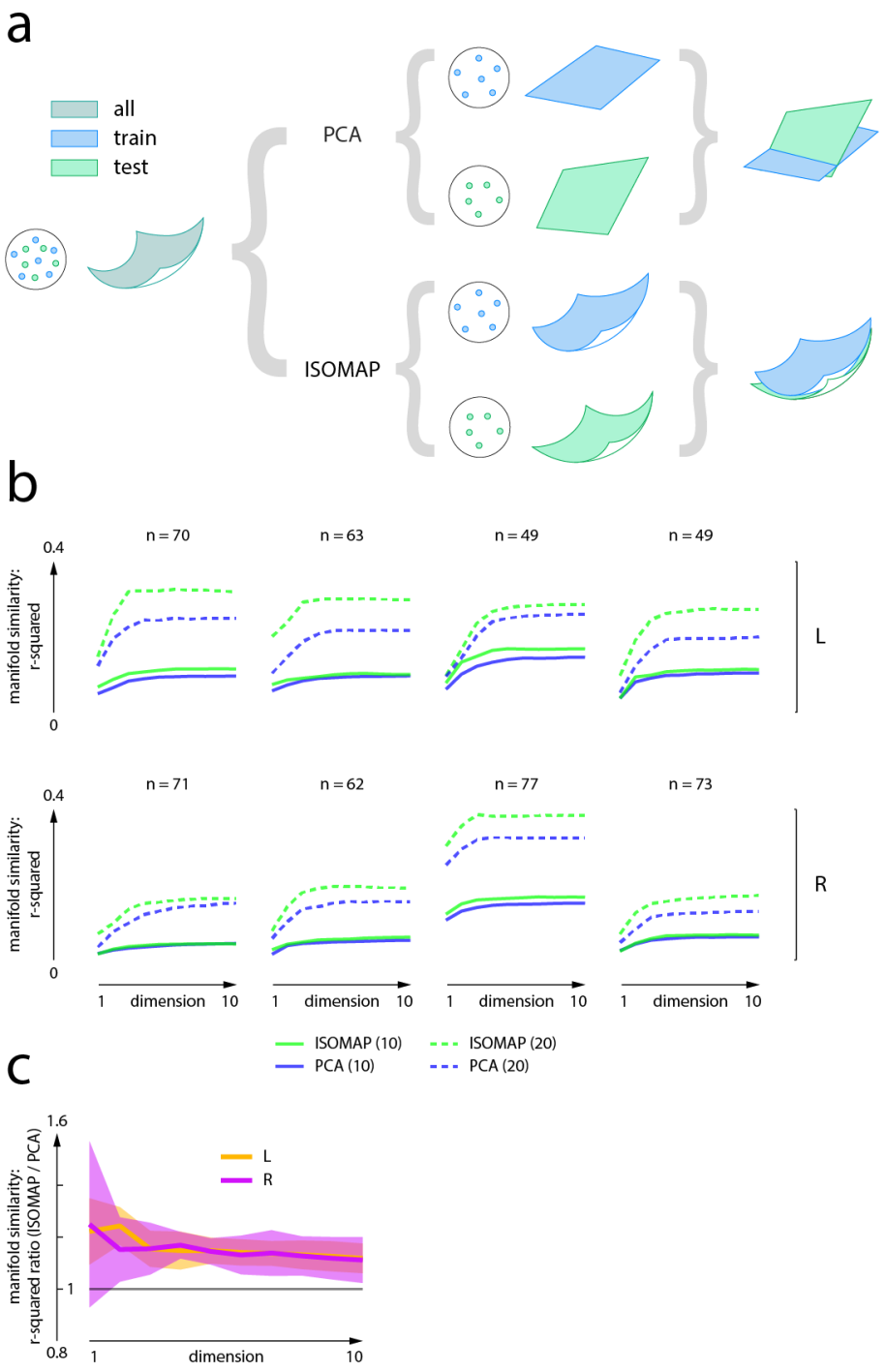
57. He, K., Zhang, X., Ren, S. & Sun, J. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*.

Supplementary information



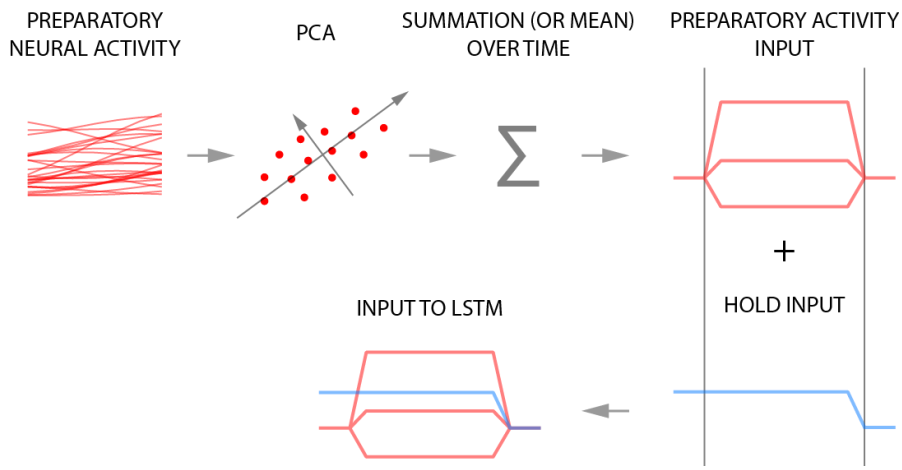
Supplementary Figure 1. Explained variance analysis.

(**a**) Explained variance (ISOMAP and PCA) for all datasets. (**b**) Non-linearity index for all neural activity in each dataset.



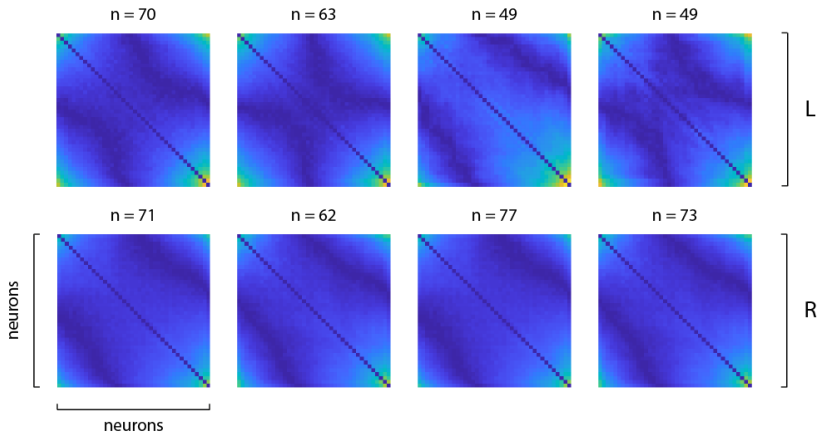
Supplementary Figure 2. ISOMAP outperforms PCA at finding consistent dynamics.

(**a**) Schematic of the overfitting analysis. (**b**) Results of the analysis for all datasets. (**c**) Ratio between the performance of ISOMAP and PCA for both C_L and C_R.



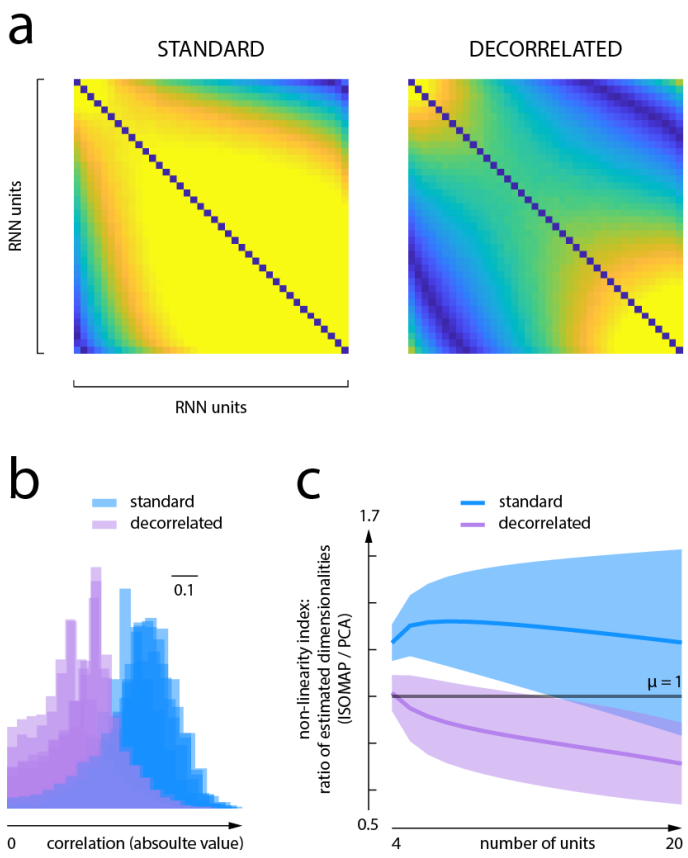
Supplementary Figure 3. LSTM input generation.

Schematic of the generation of the model's input.



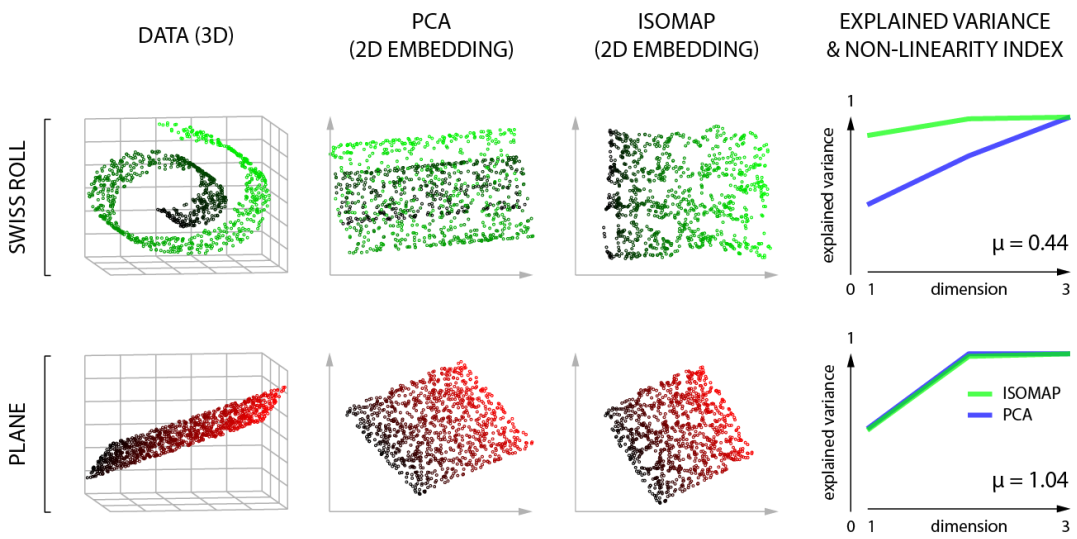
Supplementary Figure 4. Neural activity correlation maps.

Correlation maps between neurons for all datasets.



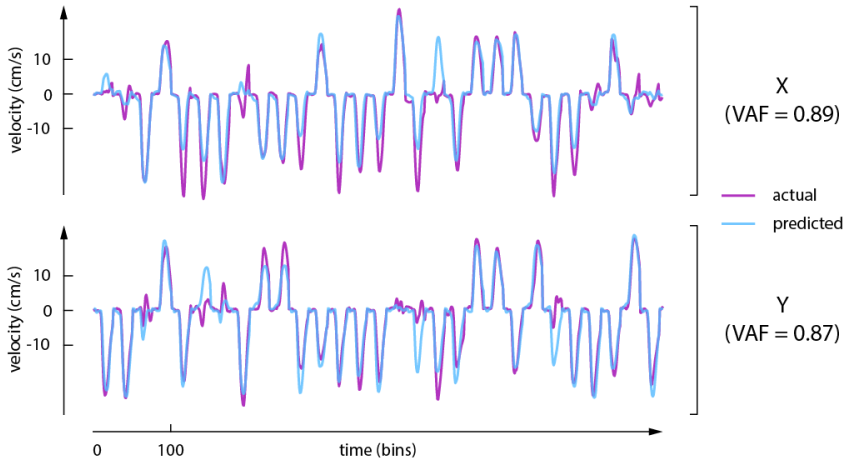
Supplementary Figure 5. Decorrelation analysis.

(**a**) Correlation maps for a decorrelated and a standard – non-decorrelated – model. (**b**) Effect of the decorrelation parameter on the distribution of correlations of the network. (**c**) Decorrelating the network's activity increases the non-linearity of the model, especially when analysing larger groups of units.



Supplementary Figure 6. Testing the non-linearity index.

We tested the non-linearity index – μ – for 'swiss roll' and 'plane' datasets to evaluate how μ reflects manifold curvature. As expected, 'swiss roll' manifolds yielded stronger non-linearities – $\mu \approx 0.44$ – than planes – $\mu \approx 1.04$ –, which were better explained by PCA. For this analysis, we used the seven nearest neighbours ($k = 7$) to construct the ISOMAP neighbourhood graph and one thousand data points for each set, as in ref. 32.



Supplementary Figure 7. Velocity prediction with the LSTM.

Velocity prediction – of the testing dataset – in the X (89% of variance-accounted-for) and Y (87% of variance-accounted-for) directions with the LSTM model.

φ	ξ	H	α	λ	β_1	β_2	ε
1 0.5 0.25 0.125	0	400	0.005	0	0.9	0.999	10^{-8}

Supplementary Table 1. Parameters for the connectivity analysis.

φ	ξ	H	α	λ	β_1	β_2	ε
0.25	0 60	400	0.005	0	0.9	0.999	10^{-8}

Supplementary Table 2. Parameters for the decorrelation analysis.

φ	Connectivity parameter
ξ	Decorrelation parameter
H	Number of hidden units
α	Learning rate
λ	L_2 regularisation parameter
β_1, β_2, ε	Other ADAM optimisation parameters

Supplementary Table 3. General parameters of the model.

Author information

Main author: **Jorge Jacobo Bennasar Vázquez**

MSc in Biomedical Engineering (Neurotechnology), Department of Bioengineering, Imperial College London.

For e-mail correspondence: jorgebvm97@gmail.com.

Supervisor: **Juan Álvaro Gallego**

Lecturer, Department of Bioengineering, Imperial College London.

Data availability

The neural data analysed in this study will be shared upon reasonable demand.

Code availability

Most of the MATLAB code used has been deposited in the following GitHub repository: https://github.com/gallegolab/manifold_geometry.