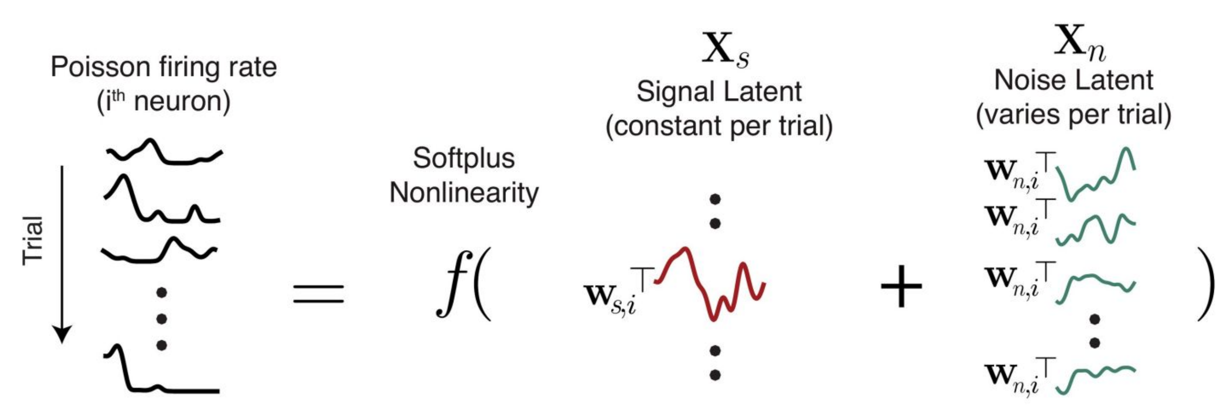
# Preprocessing Data of Varying Trial Duration with Linear Time Warping to Extend on the Applicability of SNP-GPFA

Arjan Dhesi, Arno Onken
School of Informatics, University of Edinburgh, UK

## Neuroscience data of varying trial duration is length and feature aligned with time warping

Signal-noise Poisson-spiking Gaussian Process Factor Analysis (SNP-GPFA) is a popular model for analyzing neural data [1]. However, a limitation exists, in that it cannot be applied to data of varying trial duration, limiting the range of experiments that can be performed. This work proposes data preprocessing techniques to feature align uneven length spike data, as well as findings from the application of SNP- GPFA to transformed rodent V1 data.

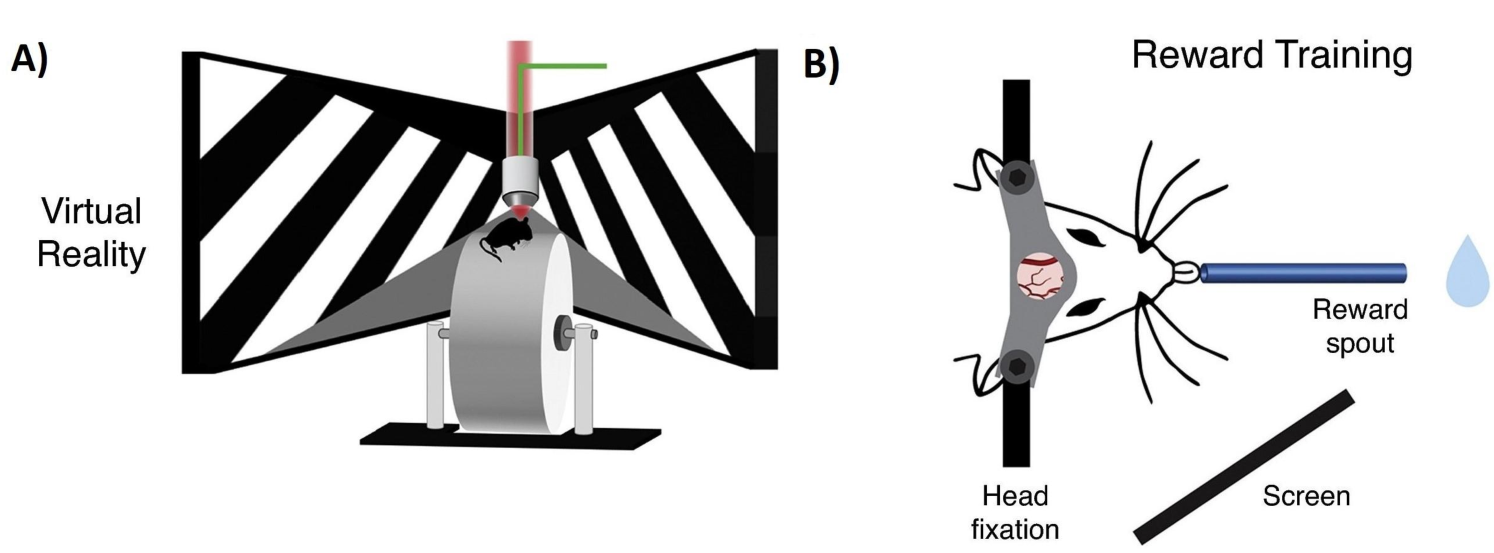
One issue with typical neuroscience recordings is that there is a notable amount of trial varying noise that can make results more difficult to clearly examine and analyze. The noise in this case is defined as the variations about the determined "true signal" that is fully dependent on the presented stimulus. The SNP-GPFA model was designed to show both the signal and noise latents. GPFA uses a Gaussian Process prior to link together a set of Factor Analysers. SNP-GPFA builds off this with the model structure below.




$$\mathbf{y}_j = \mathrm{Poiss}(f(\mathbf{W_s}^\top \mathbf{X}^s + \mathbf{W_n}^\top \mathbf{X}_j^n))$$

An issue with SNP-GPFA is that it requires data of fixed length, but the duration of trials in typical behavioral data varies.

## The dataset

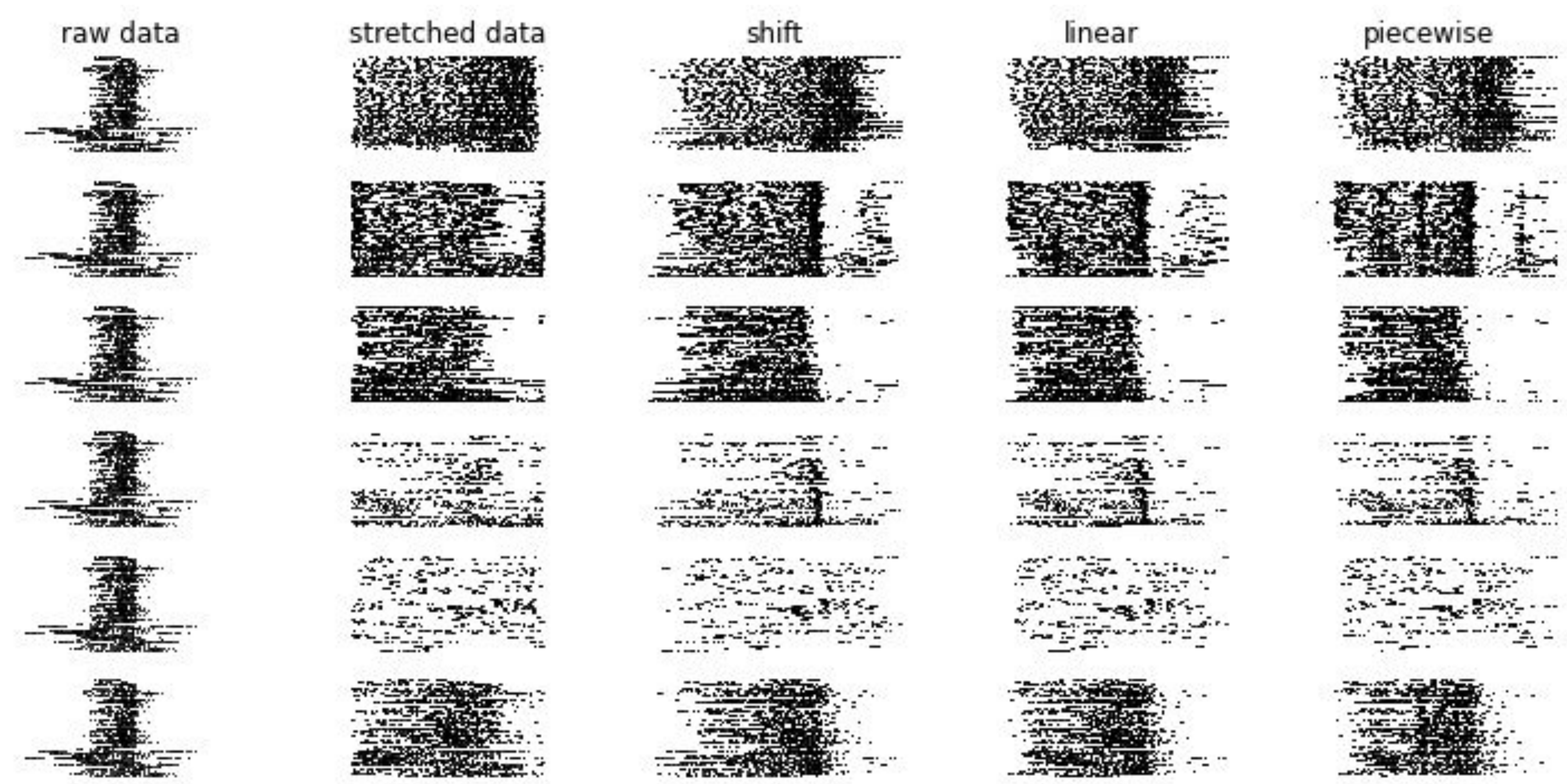



(A) The mouse on a fixed treadmill, with screens to create the appearance of a corridor in the virtual environment. (B) The head fixated mouse with cranial window to allow for the recording of its neurons in V1. The reward spout provides the mouse with water when it reaches the reward zone of the virtual corridor. (Taken from [2])

The neuroscience dataset for this project contains the activity of many neurons in mice V1 as the subjects walk down a 160 cm length corridor in a virtual reality environment towards a reward zone beginning at the 120 cm mark, defined by walls of a different visual pattern, where a water reward is supplied to the mice. We analyzed a sub-dataset that contained the recordings of 23 neurons over 69 trials. No neurons or trials were omitted.

## Time warping

The raw spike data is first stretched to match the length of the longest trial. After this, the time warping methods of shift, linear and piecewise linear [3] warping are applied to the data.




Here each large row is a different neuron, and within these rows are 69 smaller rows for each trial. The raw data is stretched and warped with shift, linear and 1 knot piecewise linear warping.

It can be seen that all the warping methods have aligned certain features in the spike data. Through cross validation it was found that linear warping proved the most effective, as piecewise linear warping starts to overfit.

### References

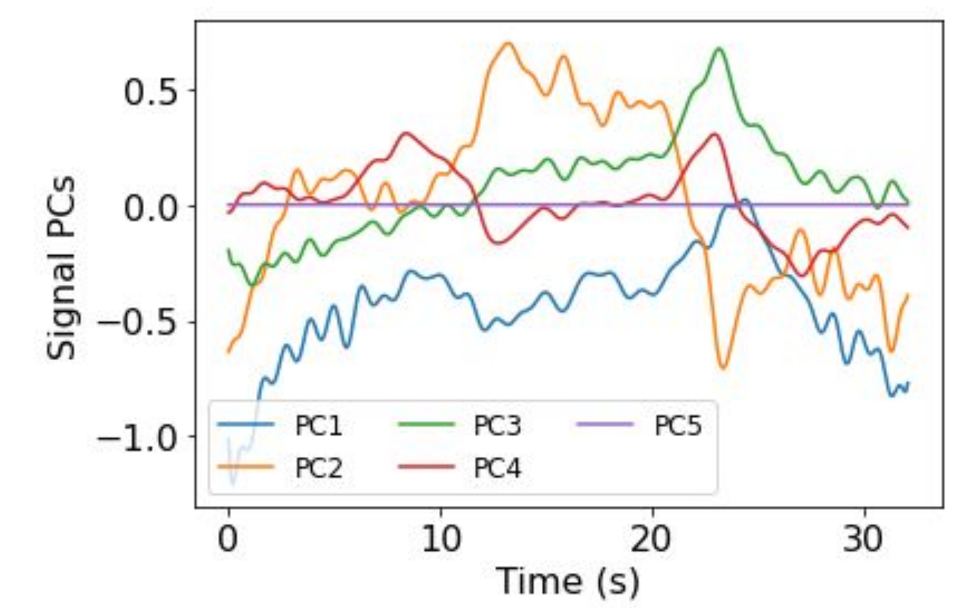
[1] Stephen Keeley, Mikio Aoi, Yiyi Yu, Spencer Smith, and Jonathan W Pillow. Identifying signal and noise structure in neural population activity with gaussian process factor models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 13795–13805. Curran Associates, Inc., 2020.

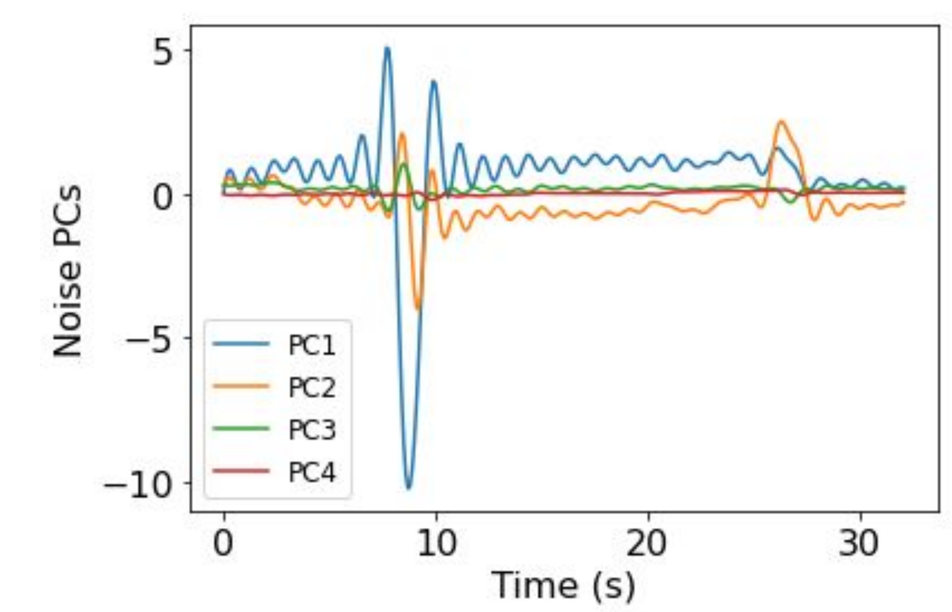
[2] Janelle MP Pakan, Valerio Francioni, and Nathalie L Rochefort. Action and learning shape the activity of neuronal circuits in the visual cortex. Current Opinion in Neurobiology, 52:88–97, 2018. Systems Neuroscience.

[3] Alex H. Williams, Ben Poole, Niru Maheswaranathan, Ashesh K. Dhawale, Tucker Fisher, Christopher D. Wilson, David H. Brann, Eric M. Trautmann, Stephen Ryu, Roman Shusterman, Dmitry Rinberg, Bence P. Ölveczky, Krishna V. Shenoy, and Surya Ganguli. Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping. Neuron, 105(2):246–259.e8, 2020.

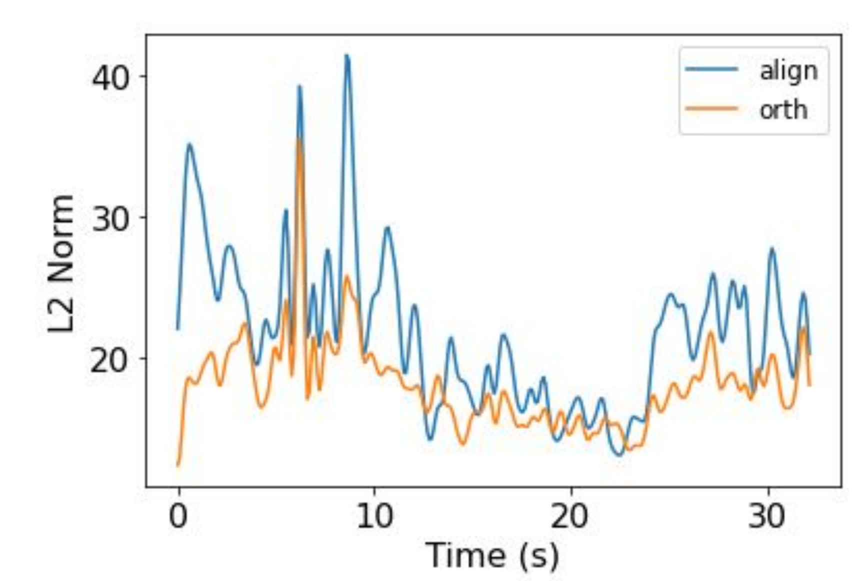
## SNP-GPFA: Signal and noise latents

The stretched, linearly warped data is fed to the SNP-GPFA model to find the underlying signal and noise principal components.




The signal PCs 1, 3 and 4 can be seen to exhibit peaking at reward time, while PC 2 exhibits sustained decrease beginning at the reward zone. PC5 is flat due to some neurons having very little to no activity. The noise PCs are mainly oscillatory, except at around the 10 second mark and towards the end of the trials. These large spikes are determined to be an artifact of specifically applying linear warping with this dataset, as altering the warping method or dataset removes these features.




When viewing how the noise subspace orientation in relation to the signal subspace orientation changes with time, it can be seen that the noise subspace is more likely to be orthogonal than aligned to the signal subspace when the overall noise variance is low. This along with that stretching followed by linear time warping proves the best preprocessing method are the main takeaways.