Enabling Naturalistic, Long-Duration and Continual Experimentation

Joaquin Rapela

February 14, 2025

Contents

1 Vision

At the Sainsbury Wellcome Centre for Neural Circuts and Behavior (SWC) and at the Gatsby Computational Neuroscience Unit (GCNU) we are performing a radically new type of experimentation, that is Naturalistic, Long Duration, and Continual (NaLoDuCo). We are recording bahavioral and electrophsiological data continuosly for weeks to months, while mice forage, indidually or socially, in large naturalistic arenas.

The Allen Institute for Neural Dynamics (AIND) is creating software and hadware technology to study foraging along two parallel paths. First, they are investigating the neural basis of foraging behavior in shorter virtual reality experiments on head fixed mice. Second, they are studying how freely moving mice learn over extended periods of time, while they spontaneously explore odor sources.

The SWC/GCNU and the AIND have complementary expertise in foraging research. Since 2021 the SWC/GCNU have been creating software and hardware infrastructure to enable NaLoDuCo foraging research, and have successfully used this technology to perform simultaneous behavioral and neural recordings in mice foraging in large arenas for weeks to months. Also, the GCNU is a leader in developing advanced neural data analysis methods, that will be essential to understand data generated by NaLoDuCo experiments. The AIND is performing foraging research, in head-fixed mice using virtual reality, and in freely moving mice spontaneously exploring odour sources in long-duration experiments. In addtion, the AIND is a pioneer in generating and openly disseminating high-quality behavioral and neural datasets at scale.

Here we propose to join forces to co-develop critical NaLoDuCo experimentation technology, test it thorougly in our three experimental setups (1. freely moving foraging at the SWC/GCNU, 2. head-fixed foraging at the AIND, and 3. freely moving odour exploration at the AIND), and disseminate it openly to facilitate the adoption of NaLoDuCo experimentation by research groups around the world.

2 Data acquisition, management, quality control and alerts

2.1 Deliverables

1. hardware specifications for recordings of behaviour and neural activity used at the SWC/GCNU[1] and at the AIND for head-fixed foraging[2] and freely moving odour exploration.

2. software for managing long-duration recordings (e.g., data storage, data indexing).

3. software for online/offline quality control.

4. software for creating alerts.

5. software for online (behavioral and neural) data visualisation.

6. software for online (behavioral and neural) data analysis.

2.2 Previous work

- the SWC has performed foraging experiments
  - lasting xx weeks and recording behaviour only
  - lasting yy weeks and recording behaviour and electrophysiology
  - data is stored in files and in a MySQL database

- the AIND has performed foraging experiments in head-fixed mice (ask Josh, Saskia or David about the status of this project).

- the AIND is setting up the odour exploration experiments.

- items 1–2: above have been completed for the SWC foraging experiments

- item 3: the SWC has developed several quality control for behavior.

[1] https://sainsburywellcomecentre.github.io/aeon_docs/reference/hardware.html

[2] https://www.allenneuraldynamics.org/platforms/behavior

- item 4: alerts have been developed for behaviour.
- item 5: the SWC has developed some online behavioural data visualisation tools in Bonsai.
- item 6: funded by BBSRC, we have integrated into Bonsai tools for online data analysis:
  - estimate kinematics of mice
  - estimate kinematic states of mice using Hidden Markov Models
  - clusterless point-process decoder of mice position and replay from spikes
- disseminated documentation of hardware used at the AIND to perform VR and head-fixed foraging exeperiments[3].
- disseminated documentation on software used at the SWC to control NaLoDuCo foraging experiments (see repo)
- disseminated documentation on machine learning methods integrated into Bonsai for analyzing behavioral and neural tie series in real time (see repo)

2.3 Future work

- item 3: at the SWC and at the AIND we have developed several tools for offline quality control. We next need to build online versions of them.
- item 4: develop more software for data visualisation.
- items 5 and 6: develop more software for online data analysis.
  - online estimate of latent variables from Neuropixels recordings.
  - online estimate of RL models.
- help Dr. Carl Schoonover (AIND) use hardware and software developed at the SWC/GCNU to create his olfaction NaLoDuCo experiments.

[3] https://www.allenneuraldynamics.org/platforms/behavior

3 Data sharing

3.1 Deliverables

1. dashboard (or examples of how) to convert data collected in NaLoDuCo experiments to the Zarr format, and to upload the Zarr files to DANDI.

2. dashboard to stream data collected from NaLoDuCo experiments to DANDI

3.2 Previous work

- the AIND is experienced on sharing their recordings on DANDI. However, these recordings are not as large as those in NaLoDuCo experiments and they are not continual.

3.3 Future work

- develop dashboard to convert data collected in NaLoDuCo experiments to the Zarr format

- develop dashboard to stream data collected from NaLoDuCo experiments to DANDI

- test dashboards on data from:

    1. freely-moving foraging mice (SWC)
    2. head-fixed foraging mice (AIND)
    3. freely-moving odour exploration mice (AIND)

- DANDI is typically used to store neurophysiological datasets much smaller than those generated in NaLoDuCo experiments. Conventional methods to access data in DANDI may not be fast enough to allow performant data visualisation and/or data analysis. We may need to explore parallel computing and/or resource efficient cloud configurations (i.e., optimising cloud configurations to improve runtime performance).

4 Data visualisation

4.1 Need for cloud-based visualisation

A unique feature of the NaLoDuCo recordings collected at the SWC and the AIND, is that they are long-duration and continual. Greatest insights will come from investigating these recordings as a whole, and not by analysing separately its parts. For example, the analysis of shorter duration recordings will not be able to capture long-term temporal dependencies in neural activity, that could be critical to understand infradian modulations of behaviour. Thus, we need software infrastructure to browse and visualise week-to month-long experimental recordings on the order of hundreds of terabytes. It is not feasible to download these huge datasets in order to visualise them. Hence, **offline data visualisation needs to be done on the cloud**, as in Neurosift.

4.2 Deliverables

1. web-based dashboard for **online** visualisation of quality control measures.
2. web-based dashboard for **online** data analysis and visualisation of its results.
3. web-based dashboard for **offline** visualisation of NaLoDuCo behavioural and neural recordings on DANDI.
4. web-based dashboard for **offline** visualisation of data analysis results on DANDI.

4.3 Questions

- can I see the visualisation tools from the AIND?
- does the AIND has visualisation tools running on the cloud?
- what visualisation tools do we have at the SWC? developed by Datajoint?
- is the AIND collaborating with Jeremy Magland?

4.4 Previous work

- Neurosift (repo, paper) allows to visualise shorter-duration behavioural and neural recordings in DANDI.
- Dendro (repo) allows to perform analysis on the cloud and visualise the results of such analysis
- offline and precomputed visualisations developed at the SWC, with the help of Datajoint.
- offline and precomputed quality control visualisations developed at the AIND.
- offline and precomputed visualisations developed at the AIND for shorter-duration experiments.
- offline and precomputed visualisations from IBL for short-duration experiments.

4.5 Future work

- Neurosift has been designed to visualise relatively short duration datasets. We will extend it with data pyramids (e.g., https://github.com/carbonplan/ndpyramid) to enable it to operate on long-duration recordings.

5 Spike sorting

5.1 Deliverables

1. methods (and software implementation) for sorting spikes from Neuropixels probes in long-duration experiments

2. software for curation of results of sorting spikes from Neuropixels probes in long-duration experiments

3. quality control measures for the results of sorting spikes from Neuropixels probes in long-duration experiments

5.2 Previous work

- methods exist for sorting, curating and quality control spikes from short duration experiments

- Dr. Carl Schoonover (AIND) has developed methods to sort spikes from long-duration experiments (find out more about these methods)

- the SWC has managed to sort spikes from a small subset of channels of a Neuropixels probe (how many channels? what duration?)

6 Data analysis

6.1 Non-Stationarity

Most current data analysis methods assumme some sort of stationarity in the data used to estimate and evaluate their models. This assumption may be reasonable for experimental data from short duration experiments. However, for long-duration experiments, with subjects learning and adapting over time, this stationarity assumption is questionable (**?**).

Here we will rigorously test data stationarity and adapt the proposed methods to work in non-stationary evnvironments, when necessary.

6.2 Deliverables

1. methods to analyse behavioural and electrophysiological recordings from NaLoDuCo experiments that are **online** and **adaptive to non-stationarity** in measurements.

   For behavioural data, we will investigate methods to

   - track multiple body parts of animals (deep neural networks)
   - infer kinematics of foraging mice (linear dynamical systems)
   - segment behaviour into discrete states
   - characterize short- and long-term periodicities in behavior
   - infer the rules that govern mice behaviour from behavioural observations only] (i.e., policy inference).

   For neural data, we will investigate methods to:

   - estimate low-dimensional continual representations of neural activity (i.e., latents inference)
   - segment neural activity into discrete states
   - characterize short- and long-term periodicities in neural activity
   - decode environment variables from neural activity

2. integration of these methods into DANDI, as in Dendro, so that users can run them on NaLoDuCo datasets stored in DANDI.

6.3 Previous work

- At the Gatsby we have invented several methods for the characterisation of neural time series (e.g., Yu et al., 2009; Duncker and Sahani, 2018; Rutten et al., 2020; Yu et al., 2024; Buesing et al., 2012b,a; Macke et al., 2015; Soulat et al., 2021; Walker et al., 2023; Turner and Sahani, 2014; O'Shea et al., 2022)

- Dendro allows to perform advanced data analysis on DANDI. It allows to reuse previous analysis.

7 Intelligent experimental control

7.1 Relevance

Most experiments in Neuroscience are controled by predefined rules or by simple neural and/or behavioral observations. Here we propose to control neuroscience experiments using advanced inferences on behavioral and/or neural data.

Intelligent experimental control is relevant to long-duration naturalistic experimentation for at least two reasons. First the extremely large size of recorded dataset may forbid storing all raw data and online machine learning algorithms can help decide what data to store. For instance, we want to record the behaviour of multiple animals in large arenas with high resolution. This requires using multiple high-definition cameras to record videos of different parts of the arena. It is not feasible to store the videos by all cameras in long duration experiments. To overcome this problem, we can use probabilistic machine learning methods to track online the positions of the mice in the arena. When the tracking confidence of this methods is high, we would only save the highresolution videos of the cameras filming mice, but when their confidence is low we would save the videos of all cameras.

Second, we want to make neural interventions informed by online inferences from machine learning methods. For example, in a foraging experiment we could find that a neural latent variable peaks before the instant when a mouse starts accelerating to leave the current patch (the latent variable could be estimated using a Poisson linear dynamical system and the acceleration using a linear dynamical system). We could hypothesise that the neural population associated with the latent variable is responsible for the decision of leaving a patch. We could then test this hypothesis by optogenetically silencing this population while an animal is on a patch and checking if it leaves the patch or not.

7.2 Deliverables

1. software, hardware and papers demonstrating applications of intelligent experimental control in neuroscience experimentation.

7.3 Previous work

- Bonsai is an excellent software for close-loop neuroscience experimental control.
- funded by a BBSRC award[4] we are integrating machine learning methods into Bonsai.

7.4 Future work

- Bonsai has been used for close-loop control using direct observations. With the advanced machine learning methods that we have added to Bonsai, we can know infer subtle patterns in behavioral and neural recordings not visible to the nacked eye. Now we will find applications where these patterns can be used for experimental control. Fortunately, we work at the SWC, where such applications abound.
- it would be very useful to perform control based on patterns inferred from spiking activity recorded by Neuropixels probes. To do so, we need online spike sorting methods for Neuropixels probes. These methods are underdeveloped and, if neccessary, we will need to develop one.

[4] https://gow.bbsrc.ukri.org/grants/AwardDetails.aspx?FundingReference=BB%2FW019132%2F1

References

Buesing, L., Macke, J. H., and Sahani, M. (2012a). Learning stable, regularised latent models of neural population dynamics. *Network: Computation in Neural Systems*, 23(1-2):24–47.

Buesing, L., Macke, J. H., and Sahani, M. (2012b). Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. In *Advances in Neural Information Processing Systems 25: 26th Conference on Neural Information Processing Systems (NIPS 2012)*, pages 1691–1699.

Duncker, L. and Sahani, M. (2018). Temporal alignment and latent gaussian process factor inference in population spike trains. In *Advances in Neural Information Processing Systems*, pages 10445–10455.

Macke, J. H., Buesing, L., and Sahani, M. (2015). *Estimating state and parameters in state space models of spike trains*, page 137–159. Cambridge University Press.

O'Shea, D. J., Duncker, L., Goo, W., Sun, X., Vyas, S., Trautmann, E. M., Diester, I., Ramakrishnan, C., Deisseroth, K., Sahani, M., et al. (2022). Direct neural perturbations reveal a dynamical mechanism for robust computation. *bioRxiv*, pages 2022–12.

Rutten, V., Bernacchia, A., Sahani, M., and Hennequin, G. (2020). Non-reversible gaussian processes for identifying latent dynamical structure in neural data. *Advances in neural information processing systems*, 33:9622–9632.

Soulat, H., Keshavarzi, S., Margrie, T. W., and Sahani, M. (2021). Probabilistic tensor decomposition of neural population spiking activity. In *Advances in Neural Information Processing Systems*.

Turner, R. E. and Sahani, M. (2014). Time-frequency analysis as probabilistic inference. *IEEE Transactions on Signal Processing*, 62(23):6171–6183.

Walker, W. I., Soulat, H., Yu, C., and Sahani, M. (2023). Unsupervised representation learning with recognition-parametrised probabilistic models. In *International Conference on Artificial Intelligence and Statistics*, pages 4209–4230. PMLR.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of neurophysiology*, 102(1):614–635.

Yu, C., Sahani, M., and Lengyel, M. (2024). Discovering temporally compositional neural manifolds with switching infinite gpfa. *bioRxiv*, pages 2024–10.