Designing reproducible analysis workflows for experimental and simulated activity using Elephant

Danylo Ulianich, Robin Gutzen, Julia Sprenger, Elena Pastorelli, Giulia De Boni, Peter Stanislao Paolucci, Andrew Davison, Sonja Grün, and Michael Denker

1 Inst. of Neuroscience and Medicine (INM-6), Inst. for Adv. Simulation (IAS-6), JARA Inst. Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Germany
2 National Institute for Nuclear Physics (INFN), Rome, Italy
3 Unité de Neurosciences, Information et Complexité, Neuroinformatics group, CNRS FRE 3693, Gif-sur-Yvette, France
4 Theoretical Systems Neurobiology, RWTH Aachen University, Germany

Introduction

The need for reproducible research is a topic of intense discussion in the neurosciences. In the context of data analysis, we develop the Electrophysiology Analysis Toolkit (Elephant) [1] as a central resource to provide tested and validated reference implementations of common analysis methods for activity data. However, reproducibility also requires such tools to be embedded in collaborative, holistic workflows [2] providing clear, traceable analysis steps from data acquisition to publication.

Analysis using Elephant

Elephant is a community-centered, open-source software package that provides components for the analysis of multi-scale electrophysiological data (e.g., spike trains, local field potentials) from experiments and neuronal simulations, focusing on:

- methods for the analysis of parallel recordings
- correlational features of brain dynamics
- bridging different scales of observation

Summary

The presented analysis workflow...  
- combines several public, community-centered software tools to achieve a reproducible analysis.
- provides a comprehensible data flow across scales independent of the data format using the Bio library.
- leads the way towards the implementation of future analysis workflows based on the Elephant library.

Find further resources:
http://python-elephant.org
https://github.com/NeuralEnsemble/elephant

Multi-scale analysis workflow on the Collaboratory

Here, we showcase how Elephant is integrated into an analysis workflow running on the Collaboratory, reproducing work in [3]. The workflow consists of complementary open-source tools and services [3] for metadata management, versioning, and collaboration ([5,6,7]), data query ([8]), knowledge graph ([9]), data versioning ([10]), data storage ([11]), data handling ([12]), and containers ([13]). Finally, we outline how these building blocks, combined with generic tools, can be assembled into formalized workflows to support reproducible research, e.g. the validation of network simulations with NetworkUnit.

Validation with NetworkUnit

- Validation is the process of establishing confidence in a model by quantitatively testing whether its prediction accuracy is within an acceptable agreement to its system of interest.
- Network-level validation evaluates the model simulation on the level of the network activity as opposed to the complementary approach of validating on a single-cell level.
- Model-to-model validation compares models (or their implementations) for consistency, cross-validation, simulator evaluation, or quantification of model developments.

The Python module NetworkUnit [13] is based on SciUnit [14] and Elephant, and provides a formalized framework along with a battery of standardized tests for network-level validation.

- Models are matched to appropriate tests via ‘capabilities’.
- New tests can be easily derived from a range of base tests.
- Tests can be adapted to also compare multiple models.
- Test scores are annotated with their provenance.

Standardization → Reproducibility

Modularization → Versatility

Formalization → Understandability

References