

HFSP Fellowship

On model misspecification in neuroscience

1. Introduction

Neural simulation-based inference (SBI) is a recently developed Bayesian framework for posterior estimation, which has already found successful applications across various scientific disciplines [refs from Jan Blog]. A persistent challenge affecting this method is *model misspecification* [Spanos, Frazier]. While formal definitions may vary, this issue essentially arises when the simulated joint distribution fails to capture the essential statistical features of the true, yet inaccessible, data-generating process. As a result, neural estimators are forced to operate in an *out-of-distribution* regime, often producing unreliable estimates. Two main causes contribute to this:

1. The functional form of the simulation model is incorrectly specified (*This might not align with the statistical community's definitions, so a clarifying reference may be needed.*)
2. The noise model governing the deterministic dynamics is incorrect [Spanos, Frazier].

With the growing adoption of likelihood-free inference methods [SBI], model misspecification has become a central issue—particularly in computational neuroscience. There is a strong need for a quantitative understanding of this problem. This need is especially pressing in neuroscience, where numerous mathematical models have been developed to address complex, multi-scale data, ranging from low- to high-throughput recordings [LIF, QIF, FN, WC, HH, Gerstner, Clopath, Shirin, Lorenzo, ...]. To what extent these models are functionally equivalent—or whether they can be unified under a shared set of principles—remains an open question.

We aim to address this question using a multidisciplinary approach, beginning at the level of neuronal circuits.

2.1 Computational

Our computational work begins with a recently developed internal tool, Jaxley [I mean its public now but should we say the GoncaLab developed it?, JXL]. This package offers several key advantages for our goals: (i) it supports highly flexible network instantiations, including arbitrary connectivity, morphology, and plasticity (ii) it is implemented in [JAX], enabling compiled, parallelized execution on both GPU and CPU (iii) JAX's native automatic differentiation allows for efficient gradient computation and parameter optimization without relying on third-party tools.

Using Jaxley, we plan to explore the behavior of various neuroscientific models, analyzing both spiking activity and derived summary statistics. Specifically, we will investigate:

- **Q1:** Can a neural network distinguish between spiking datasets generated by different models? How does this depend on the models used?
- **Q2:** How robust is SBI when applied to data from functionally perturbed models?
- **Q3:** Can we use a known joint distribution (from a "ground truth" model) to infer priors for a simplified model?

To better address **Q2**, particularly in the context of neuronal dynamics, we will analyze the system's reduced dynamics. Dimensionality reduction has become central in neuroscience for making sense of complex, high-dimensional data [Mante (PCA), Alvaro-Gallego (TCA), Albert (CCA), dPCA, UMAP, t-SNE, ...]. These techniques, however, often have to choose between the linearity of the decomposition and the interpretability of the embedding.

To overcome this, we propose to explore an alternative method called **SSM (Spectral SubManifold decomposition)** [Haller], which has already been applied with success to the mechanical engineering field. This approach allows, under the presence of fixed points in the dynamics, a non-linear, analytical decomposition, potentially recovering the slow (observable) neuronal dynamics in an explicit way. Importantly, this method allows for data-driven approaches, allowing to recover reduced order dynamics agnostically from experimental observations. Importantly, this technique can also be applied in a data-driven fashion, allowing reduced-order dynamics to be extracted agnostically from empirical observations.

[IMPORTANT: THERE IS NO PYTHON IMPLEMENTATION OF SSM, SO WE WOULD NEED TO DO THAT]

2.2 Experimental

Once the model validation in §2.1 has been conducted, we plan on understanding how inference on synthetic data generalized to inference on real-world observation. This is a crucial step in addressing our scientific question, as it represents the factual test scenario of the theoretical model selection hypotheses derived in §1.

We plan on collaborating with thh experimental lab that can grow specific neurons and make them develop specific things, finally producing what can be somewhat defined as a controlled experiment. (I need to find some citations for already present datasets, and some better description of what the collabolators would to - maybe we can talk to them? see if it even makes sense?)

In particular, we are interested in exploring three questions, closely related to the misspecification issue:

- **Q1:** see if a NN can distinguish between synthetic data and natural data
- **Q2:** see if we can train a NN to reproduce true neuronal data so that we can use it as evidence generator
- **Q3:** can we use this new dataset to improve SBI training? In particular, can we use the differences between synthetic and real observation distributions to derive priors for model parameters? (given p_{θ} one optimizes q_{ϕ} to give posteriors; can we do the opposite? q_{ϕ} on data, p_{θ} on q_{ϕ} ? I still have to study)

Answering these questions would have immediate biological implications, providing distinctive neuronal and dynamical features distinguishing healthy from pathological networks - if we are collaborating on the Parkinson's disease project.

2.3 Neuromorphic

If we can characterize neuronal networks, understand how and when they work, we can propose new inspiration into neuromorphic computing.

In particular, if our models can distinguish healthy, pathological and synthetic networks, they can also shed light into the differences between natural and silico circuitry (at some reasonable level, no need to compare the spike shapes I guess). This can help design better neuromorphic hardware.

In plain words:

At the end of the day we are looking for some properties (invariances? optima of some energy measure?) that make natural networks work, and other networks work not so well. We hope to get this Graal thing by comparing different models and data (leveraging the fact that modelers have so far produced a dataset of models, so these models have hopefully something valid in common that can unify them all into a truish "neural circuit model" template that works) (basically a MonteCarlo but with models) (can we use SBI on it?). If this valid thing is what makes a neuronal network "special" then maybe we find it - unlikely though.