



Information theory for data-driven model reduction in biology

Matthew Schmitt, Vincenzo Vitelli, Michael Rust, Margaret Gardel

Our work addresses a central challenge in biology: how to find minimal mathematical models that capture the essential dynamics of a system. This is especially difficult for complex non-equilibrium or living systems where we don't have much physical intuition to guide us. Here, we present a data-driven pipeline that is grounded on a mathematical mapping between model reduction and optimal signal compression.

This information-theoretic formulation allows us to construct neural networks that identify the relevant variables based only on their information content. For example, in the figure on the left, our networks automatically extract the dominant slow collective variables from uncured videos of atmospheric flows (figure, top) and discover an emergent synchronization order parameter from experimental videos of cyanobacteria colonies (figure, middle). We are currently applying our approach to epithelia to better understand spatial heterogeneity in tissue dynamics (figure, bottom).

We stress that the key feature of these neural networks is that they are interpretable by construction. By unveiling a theoretical link between optimal compression and the operator-theoretic formalism of dynamical systems, we can (i) show analytically that the learned latent variables are given by eigenfunctions of the transfer operator and (ii) systematically answer the question of when to stop increasing the complexity of a minimal model.

Reference:

Matthew S. Schmitt, Maciej Koch-Janusz, Michel Fruchart, Daniel S. Seara, Michael Rust, Vincenzo Vitelli, arXiv:2312.06608, submitted to PNAS.

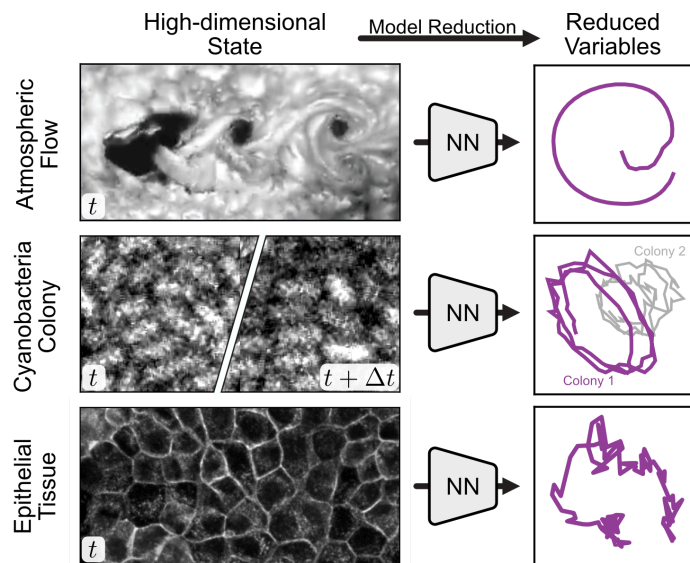


Figure: By viewing model reduction as a compression problem, we design neural networks which extract physically relevant reduced variables directly from data. From top to bottom, we apply our method to videos of atmospheric flows, bacterial colonies (Rust lab), and evolving epithelial tissues (Gardel lab).