

1 Overview

Rigorously validating a scientific model’s explanatory power requires comparing its predictions against empirical data – both data available during model development and data gathered after publication. However, in most fields model validation remains an informal and incomplete process. This makes it difficult to evaluate a model’s strengths and weaknesses, or to adequately compare two models.

This project aims to improve model validation by drawing inspiration from unit tests, a common form of software testing that validates a single component of a computer program against a single correctness criterion. For example, when writing a program to check string length, a unit test might pass “smith” into program input and check for the output “5”. The project is developing cyberinfrastructure for scientific model validation around analogous scientific validation tests – executable functions validating models against an empirical observation by producing a score indicating model/data agreement.

2 Intellectual Merit

The project enables collaborative construction and logical grouping of tests by scientific communities, and updating of tests and their results continuously as new data and models emerge. Visual summaries of aggregate results provide scientists with an up-to-date report of progress in their research area. Merits and deficiencies of competing models are clearly visible, benefiting ongoing modeling efforts and informing new theories and experiments.

These goals are realized through a framework, *SciUnit*, whose design meets this vision and enables adoption of this workflow among modelers. **The primary output of this project** is (1) *SciUnit*, software that implements the testing interface, (2) *SciDash*, a web portal providing access to test results, and (3) *NeuronUnit*, a library for test construction in neuroscience. The biggest challenge is designing a framework that applies to a wide range of models and data. This research addresses the challenge by providing an interface between models and tests that makes no assumptions, but enables developers to specify model capabilities and test requirements, ensuring that each model only takes tests appropriate to its scope.

3 Broader Impacts

While initial development of *SciUnit* is targeted towards neuroscience, this does not limit applicability to other fields of biology, since the design philosophy embraces generality, extensibility, and modularity. The case studies, discipline-agnostic tools and infrastructure, and examples this work generates will facilitate adoption of validation testing by other communities. Code re-use across research areas will reduce duplication of effort. This project promotes dissemination of research using a mechanism that complements the existing system of manuscript publication. Because *SciUnit* makes successful instances of modeling transparent, and highlights outstanding scientific questions without a deep literature search, it bridges interdisciplinary gaps, both within neuroscience and across biology. *SciUnit* also serves the public good by providing journalists and educators with an expert-curated resource where contributions of different scientific models are identified without expert training, as well as a window to a fundamental component of the scientific method – validation of scientific theories; This work broadly transforms science: by giving modelers a tool to develop models quickly and with clear purpose; by identifying the best models rigorously so further efforts can focus on these; and by helping experimentalists identify the impact of their work on a body of theory.