

1 The Challenge

Scientists develop theoretical models to provide a coherent explanation for diverse empirical observations. Rigorously validating a model’s explanatory power requires comparing its predictions against experimental data – both data available when the model is developed as well as data gathered after it is published. In most fields, however, model validation remains a highly informal and incomplete process. Drawing an analogy between scientific model validation and software testing, we argue that precise, automated validation criteria are essential for answering a central question in science: “Is the output of this model consistent with the full and current set of data available about the phenomena within its scope?”

2 Intellectual Merit

The most common form of software testing is unit testing. A unit test is a procedure that validates a single component of a computer program against a single correctness criterion. For example, when writing a program designed to compute exponentials, one could write a unit test which passes 10^2 into the program input and checks for the correct output 100. We propose the development of cyberinfrastructure for scientific model validation structured around analogous validation tests – executable functions that validate models against a single empirical observation to produce a score indicating agreement between model and data. Our work enables the collaborative construction and logical grouping of tests by members of a scientific community, and the updating of tests and their results continuously as new data is gathered and new models are developed. 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 both ongoing modeling efforts and informing new theoretical and experimental directions. Here we describe a framework, *SciUnit*, that meets this vision, and discuss practical design objectives and activities that encourage adoption of this scientific workflow in the larger modeling community. **The primary output of this project will be (1) *SciUnit*, software that implements the testing interface, (2) *SciDash*, a web portal to provide access to test results for models, and (3) *NeuroUnit*, a library of tests for neuroscience.** The greatest challenge is designing a testing framework that can apply to a wide range of models and a wide range of experimental data. We solve this by providing an interface between models and tests that makes no assumptions, but enables developers to easily specify the capabilities of models and the requirements of tests, ensuring that each model takes the precise set of tests which are appropriate to its scope. With computer scientists, computational modelers, and experimentalists working together, our team has the expertise to achieve these goals.

3 Broader Impacts

While initial development of *SciUnit* is targeted towards the neuroscience community, this in no way limits applicability of *SciUnit* to other fields of biology, since the design philosophy embraces generality, extensibility, and modularity. Developing tests and integrating models into this framework requires subfield domain expertise, and our team provides this for experimental and computational neuroscience. However, the case studies, discipline-agnostic tools and infrastructure, and examples this work generates will facilitate the adoption of validation testing by other scientific communities. Code re-use across research areas will help to reduce duplication of effort. This project promotes the 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 the broader field of biology. *SciUnit* also serves the larger public good by providing scientific journalists and educators with an expert-curated resource where contributions of different scientific models are easily identified without expert training, as well as a window to a fundamental component of the scientific method – validation of scientific theories. We provide an illustration of such educational utility in section 5. This work broadly transforms science: by giving modelers a tool to develop models more quickly and with clearer purpose; by identifying the best models more rigorously so further efforts can focus on these; and by helping experimentalists identify the impact of their work on a body of theory.

Keywords: computational model; validation; unit test; Python; scientific method