1 The Challenge

Scientists develop models to explain diverse empirical observations. Rigorously validating a 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. Analogizing between scientific model validation and software testing, we argue that precise, automated validation criteria are essential for answering a central question in science: "Is this model's output consistent with the full and current set of data available about phenomena within its scope?"

2 Intellectual Merit

The most common form of software testing is unit testing. A unit test 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". We propose developing cyberinfrastructure for scientific model validation around analogous validation tests – executable functions validating models against an empirical observation by producing a score indicating model/data agreement. This enables collaborative construction and logical grouping of tests by scientific communities, and updating of tests and their results continuously as new data and new 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. Here we describe a framework, *SciUnit*, meeting this vision, and discuss design objectives encouraging adoption of this scientific workflow among modelers. **The primary output of this project** will be **(1)** *SciUnit*, software that implements the testing interface, **(2)** *SciDash*, a web portal providing access to test results, and **(3)** *NeuroUnit*, a library of tests for neuroscience. The biggest challenge is designing a framework that applies to a wide range of models and experimental data. We solve this 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 takes only those tests appropriate to its scope. With computer scientists, computational modelers, and experimentalists working together, we have the expertise to achieve these goals.

3 Broader Impacts

While initial development of *SciUnit* is targeted towards the neuroscience community, this does not limit its 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 the adoption of validation testing by other scientific communities. Code re-use across research areas will 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 biology. *SciUnit* also serves the larger public good by providing 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 in section 5. 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.