**RCC Research II Allocation Request – Fall 2023**

PI: Stefano Allesina, Departments of Ecology and Evolution

September 14, 2022

**Research Goals and Impact**

The broad theme of the Allesina lab is the development of new methods new mathematical, statistical and computational methods for theoretical ecology. The main areas we cover are ecological networks, stability and feasibility of large ecological systems, game theoretical models in ecology, Lotka-Volterra dynamics, and science of science. Since we often work in the regime of high ecological diversity, the methods we develop often require integration of many coupled and non-linear Ordinary Differential Equations, which is computationally intensive. Often, we use mathematics as a tool to simplify the complexity of our computations, sidestepping the computational limitations of desktop machines. Additionally, coding in compiled languages such as Stan, or Julia, have also helped us to substantially increase the performance of our code. However, as the field of Theoretical Ecology advances, some tasks are mathematically intractable, and even numerical approaches on compiled languages become insufficient on regular machines. In particular, we are currently facing computational bottlenecks for three important questions; First, the probability of existence of one equilibrium containing all the species in Lotka-Volterra systems with Higher-Order interactions (HOIs). Second, is the inference of total abundances from relative abundances in microbial time series. Third, is the development of statistical tests for the effect of phylogeny on competition among plants. These three problems, described in more detail below, require solving many coupled, non-linear Ordinary Differential Equations (ODEs), and fitting sophisticated statistical models. As such, we resort to the computer cluster to tackle these problems numerically.

**Feasibility in Lotka-Volterra models with HOIs.**

Ecological communities are often abstractly represented as a collection of species interacting randomly, in pairs. Models of this class provide a null expectation regarding the coexistence of species in the natural world. These random ecosystems are unlikely to exhibit equilibria containing all the species at positive abundances (feasible equilibria) when the size of the community is large. Here, we relax the pairwise interactions assumption, considering models with Higher-Order Interactions (HOIs). In doing so, we uncover a novel ecological regime displaying widespread feasibility at high diversity. This work uncovers ecological models reflecting the invariably high diversity observed in natural ecosystems without imposing particular ecological mechanisms or parameter fine tuning, thus aligning theoretical and experimental null expectations. We argue that this reconciliation is a crucial step in order to move beyond the classical coexistence paradox, towards new avenues of ecological inquiry. We seek support to complete our calculations; a publication acknowledging this support and targeting the leading journal Ecology Letters is in preparation.

**Inferring microbial dynamics from compositional time-series**

High-throughput sequencing has provided a reliable way to determine the makeup of microbial

communities; however, only the relative abundances of microbial strains can be measured

accurately---as of today, microbial data is thus eminently compositional. In many applications, researchers need to determine how the absolute abundance of a given strain has changed in time (e.g., can *P. aeruginosa* establish in this synthetic community? Is an antibiotic targeting specific strains? Does changing the diet of mice favor certain bacterial families?) This requires supplementing sequencing with other experimental techniques; as counting colony-forming units; or add known amount of extraneous microbial DNA prior to sequencing ("spike-in"); or run the samples through a flow- cytometer to count cells; etc. All these methods are laborious and

costly. The goal of this project is to develop a purely statistical way to determine total microbial load having observed exclusively the fluctuations of the community composition in time. To this end, we need to fit statistical models written in the Bayesian programming language *stan* to ecological data. Bayesian fitting of time-series is notoriously computationally-intensive, and the added complication that we can only rely on proportions, rather than abundances, makes the calculation even more complex. If successful, the project will lead to a publication in a top interdisciplinary journal.

**Effect of phylogeny on ecosystem functioning**

Since the work of Darwin, the notion of phylogeny and common descent has been a powerful tool to organize our thinking about ecological communities. Despite the widespread use of phylogenetic data across all branches of biology, attempts to relate phylogenetic relatedness to the strength of competition between plants have been so far unfruitful. We argue that this is due to the limitations of the statistical models developed so far. We have developed a new statistical framework, which yields a simple test for the effect of phylogeny on the growth of plants that are co-cultured experimentally. Results are very strong, and a publication is currently in review in PNAS. The calculations needed for the publication were the focus of our previous allocation. We seek support for repeating the calculations in light of the reviewers’ comments, which we should receive in the next week or so.

We have found the RCC technical staff incredibly supportive and helpful, and we look forward to working with them this year, and beyond. For questions, please contact Stefano Allesina (sallesina@uchicago.edu).

**Results and Publications from Previous Allocation**

We have used 100,000 SUs to run part of the computations required for these two projects, with promising results. We are currently writing the two manuscripts, and about to revise a third publication. We need more SUs to strengthen our results, and produce new results.

**Resources Requested**

For 2023–2024, we hope to start using the RCC in our lab these three projects. To support these projects, for the upcoming service period we have requested **860,000 SUs**; the breakdown of this SU allocation request by project is given in the table toward the bottom of this document (“Justification for request”).

**Justification for Request**

A specific accounting of each project to be enabled by the requested allowance is provided in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Project** | **Description** | **Prior Results** | **Request** |
| Feasibility in Lotka-Volterra models with HOIs | Compute numerically the roots of systems of n multivariate polynomials of degree d.  This task requires solving 2\*d^n ODEs. To illustrate our claims, we approximate that we need to solve 10,000 systems with n ranging from 1 to 8, and d ranging from 1 to 6. After writing the code in an efficient language (Julia) and spending a week optimizing it, we have brought the time to solve the hardest system (n=8, d=6) down to several hours. We want to sweep the entire parameter range, that is, 48 parameter combinations. Performing one of these sweeps can take up to 32 h. Thus, we request for 10,000 \* 32 = 320,000 SUs. | NA | 320,000 SUs |
| Inferring microbial dynamics from compositional time-series | Fit ODEs to time series tracking community composition stemming from experiments with algae, bacteria and protozoans.  We have developed a maximum-likelihood approach to determine good priors for our model. Our ambition is to analyze 5 data sets of different complexity. The maximum-likelihood calculation requires about 24 hours for the largest system, and our goal is to repeat the search 250 times with different initial random seeds, to ensure that we have found the mle for all parameters. This step thus requires about 25 \* 5 \* 250 = 31,250 SU. We then will use stan to fit the model in a Bayesian framework. Each run will take about 500 hours, and we are thus estimating the needed SUs as 500 \* 5 \* 100 = 250,000. We also account for about 9,000 SU for testing. | NA | 290,000 SUs |

|  |  |  |  |
| --- | --- | --- | --- |
| Effect of phylogeny on ecosystem functioning | We need to perform 2,000 randomizations, each requiring fitting a model, for 19 data sets. The preliminary results presented in our manuscript required about 150,000 SUs, of which 100,000 were provided by RCC. We request 250,000 SUs for improving and repeating the calculations according to the comments provided by the reviewers. | A preprint is available here:  https://www.biorxiv.org/content/10.1101/2023.09.04.556236v1 | 250,000 SUs |