Overview

It is difficult to describe, predict, and preserve biodiversity in the face of rapid rates of environmental change, yet these are critical goals for conservation. A macroecological approach to understanding diversity uses patterns that exhibit regular behavior across taxonomic groups, ecosystems and continents. These patterns are increasingly being used as tools to predict diversity under global change scenarios, even though pattern response to stressors remains largely unknown. My overarching research goal is to address three limiting factors in developing a *dynamic* framework for macroecology, **1)** data aggregation and access, **2)** mechanistic complex system modeling, and **3)** cross-scale analysis. The research will meet these objectives using informatics, complex system models, and experimental data.

Intellectual Merit

Body size patterns integrate taxonomic, functional, and trait diversity. Thus, changes in body size structure may be vital to understanding changes in ecosystem function in a way that taxonomic-level patterns cannot. Size patterns are expected to be particularly sensitive to perturbation since hypothesized mechanisms that influence individual size and size evolution include species interactions and metabolic constraints, which can be altered on ecological timescales. My specific research goals are to describe, predict, and quantify body size pattern response (*e.g.,* species size distribution, individual size distribution) under a suite of environmental change scenarios and to determine if pattern response is scale-dependent. Products of this research will be of broad interest to a diverse group of scientists, including macroecologists, conservation biologists, computer scientists, and statisticians.

Broader Impacts

The proposed project will formalize a database schema for future macroecology queries, publish a new open-access dataset for research, and an interactive model for predicting pattern response to global change scenarios. I will disseminate the results of this research to three main audiences, **1)** the scientific community (*i.e.*, peer reviewed papers, dataset, scientific meetings), **2)** undergraduate students (*e.g.*, course development, student-centered learning), and **3)** the public (*i.e.*, open-access data and code, Twitter, blogging, popular science talks, outreach with local female programmers). I will integrate my research tools (*e.g.,* database, interactive model) as key elements of active learning in core biology classes and in a programming class that I will develop at the University of Maine. I will also play an active role in promoting the role of women in STEM careers via the NSF ADVANCE program at The University of Maine and via "Girl Develop It", a local coding group in Burlington, VT with the goal of getting teen girls and women excited about programming and science. The Sustainability Solutions Initiative will provide opportunities to participate in K-12 educational outreach events. Code and data will be open-access (*i.e.,* GitHub, figshare), and I will use social media tools such as Twitter and blogs to communicate with the public about my research and teaching experiences. I will assess my broader impacts yearly by tracking interactions on my blog, twitter, GitHub, and figshare accounts, collecting anonymous feedback from guest lectures and course instruction, and quantifying the number of new collaborations and mentorships gained from participating in the women in science and technology groups.

Intersections of Biology and Math

My research will combine advanced computational and statistical toolkits to address complex questions unifying macroecology with global change science that have not been effectively resolved using current data and tools. My mathematical/computational mentor, Brian McGill (University of Maine) is an ideal sponsoring scientist because of his experience with advanced statistical and informatics tools, including 10 years of experience in industry developing a leading data warehousing technology. My biology mentor, Nicholas Gotelli (University of Vermont) is primarily an experimental field biologist, but has published papers and developed software applications null and stochastic ecological models. Both universities provide strong collaborative environments that will facilitate my ability to learn cutting edge computational and statistical modeling techniques from interactions with faculty across Biology, Mathematics and Statistics, Computer Science, Engineering, and Economics departments.

Introduction

Dynamic macroecology: Globally assessing body size diversity response to environmental change

It is difficult to describe, predict, and preserve biodiversity in the face of rapid rates of human-induced environmental change, yet these remain critical goals for conservation. One solution to understanding diversity uses macroecology patterns that exhibit regular behavior across taxonomic groups, ecosystems and continents. For example, taxonomic-level patterns (*i.e.*, species area relationship, species abundance distribution) have been used to estimate the number of species that occur in areas difficult to sample, and to predict extinction (Myers *et al* 2000, Thomas *et al.* 2004). These patterns are increasingly being used as tools to predict and monitor biodiversity (*e.g.,* Brummit & Lughadha 2003, Diniz-Filho & Bini 2008).

Macroecology is a powerful way to estimate and predict diversity, but the field is currently doing a poor job of linking macroecology to global change and conservation research. There are several reasons for this: **1)** Biodiversity data remain broadly distributed throughout the literature and there are no formal guidelines for researchers to effectively publish and share data (Scholes *et al* 2008). **2)** Patterns have largely been regarded as static, and pattern response to perturbations remains unknown (Fisher *et al* 2010, Beck *et al.* 2012, Keith *et al* 2012). **3)** Patterns are typically generated at large spatial scales, and pattern form across scales has received little attention (*but see* Palmer & White 1994, Bakker & Kelt 2000).

This proposal aims to address three major limiting factors in developing a *dynamic* framework for macroecology. I will focus on body size rather than taxonomic diversity because emerging evidence suggests that taxonomic diversity is expected to be *insensitive* sensitive to many perturbations (Supp *et al* 2012, Supp 2013). Size patterns have been studied regarding manipulations of specific taxonomic groups (*e.g.*, fishing pressure; Bianchi *et al* 2005, Duplisea & Castonguay 2006), but have received almost no attention in the context of global change. Bridging the gap between conservation and macroecology requires understanding pattern *dynamics* in altered systems. This project represents an important step in developing dynamic macroecology by combining computational tools, statistical modeling, and empirical data to describe body size patterns and to predict response to change (Fig 1).

Why body size?

Most biodiversity research uses a strictly taxonomic approach, but changes to ecosystem function are a conservation concern that taxonomic approaches alone cannot address (Supp *et al* 2012). Taxonomic-level diversity metrics are often weak measures of community response, and as such, are poor indicators of important change (Hooper *et al.* 2005, Supp 2013). Body size patterns integrate taxonomic, functional, and trait diversity in a way that is vital to understanding ecosystem function (Dornelas *et al* 2011). Body size is a fundamental property related to key ecological traits (*e.g.* lifespan, dispersal) and processes (*e.g.* resource use, nutrient cycling; Jennings *et al* 2001, White *et al* 2007). Thus, size distributions are an important link between individual- and population-level processes and community- and ecosystem-level dynamics (Woodward *et al* 2005, White *et al* 2007). Patterns include the species size distribution (SSD; frequency distribution of species average body sizes) and the individual size distribution (ISD; frequency distribution of individual body sizes), and are ideally positioned to gain insight into the influence of environmental change on a form of diversity likely to have a major impact on ecosystem function or services (Folke *et al* 2004).

Figure 1. Theoretical individual size distribution (ISD) response based on body size sensitivity to global change scenarios. **A)** small-bodied individuals experience higher fitness under change, **B)** large-bodied individuals experience higher fitness. Alternatively, ISD peak(s) could flatten or the ISD could fail to respond.

There are several mechanistic hypotheses for how body size will respond to perturbation. **1)** Species interactions may constrain the evolution of body size on ecological timescales (Grant & Grant 2002, Woodward *et al* 2005). **2)** Phylogenetic models
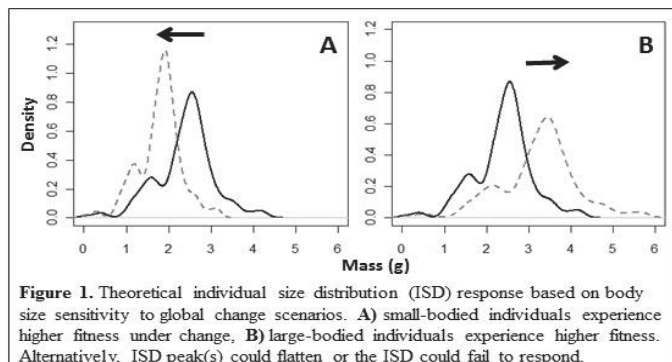
suggest that body size evolution is constrained among taxonomic clades (Smith *et al* 2004). A null hypothesis is that size distributions reflect evolution via Brownian motion (Diniz-Filho & Torres 2002). **3)** Metabolic theories suggest that environmental changes that shift the ratio of energy allocation among survival, reproduction, and growth or that impact total energy availability will lead to shifts in community body size structure (Brown *et al* 1993, Yom-Tov & Yom-Tov 2005, Anderson-Teixeira *et al* 2012).

Research Plan

Research Objectives
The primary objective of this study is to create an integrated research and education program to mechanistically describe, quantify, and predict body size pattern (SSDs, ISDs) response to environmental change and to determine if pattern response is scale-dependent. This project will address the questions:

1. Can we enable better research via data aggregation, database management & open-access tools?

2. Do global change scenarios influence body size distributions at local scales? Do model simulations support hypotheses for how body size distributions will respond to environmental change scenarios?

3. Do regional and local body size distributions have fundamentally different dynamic responses to environmental change? Is scale-dependence altered under global change scenarios?

Methods

I propose an integrated research and education program to address three major limitations to developing a mechanistic framework for dynamic macroecology, connect broad-scale diversity research to the types of change that conservation scientists are interested in, and provide a set of toolkits, including a course that will be shared with undergraduate and graduate students, peer scientists, and the public.

Empirical Data and Statistical Pattern Analysis

I will use community-level data from **1)** Breeding Bird Survey (http://www.pwrc.usgs.gov/bbs; BBS), **2)** Forest Inventory Analysis (http://fiatools.fs.fed.us; FIA), and **3)** a dataset of manipulated sites that I will compile from the literature. Aggregating valuable field data from published papers and shared online datasets is crucial for testing mechanistic hypotheses for macroecological patterns. BBS and FIA datasets include abundance, richness, and, for FIA, individual body size, for thousands of communities in the USA and Canada. In all datasets, sites undergoing change (*e.g.* land use change, regeneration post clear-cut), paired with nearby unmanipulated sites will be used as natural experiments to assess avian and tree size distribution response. My new dataset will include abundance, richness, and body size data across a broad range of taxa and ecosystems that have undergone experimental or temporal manipulation.
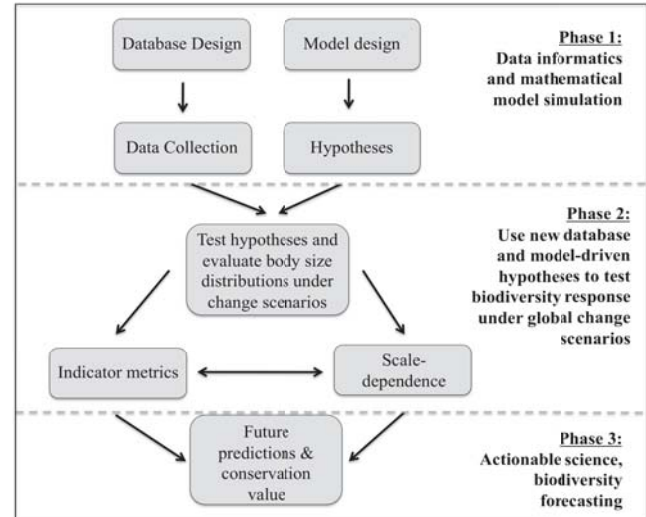
Where individual size information is not recorded in the BBS or my dataset, I will estimate individual size distributions using mean and variance data for species from published literature (*e.g.,* BBS, Dunning 2008; mammals, Smith *et al* 2003). Continuous ISDs can be estimated by randomly sampling individual sizes for each species from a normal distribution with the mean mass and estimated variance from published body size. This approach has been rigorously tested (Thibault *et al* 2011) and allows large-scale datasets, previously unavailable, to be used for evaluating body size patterns. Testing my hypotheses

using large amounts of data from a broad array of taxa and habitat types should allow for strong general conclusions about the dynamic behavior of body size patterns and what mechanisms drive response.

Body size distributions may take several different forms (*i.e.*, monotonically decreasing, unimodal, or multimodal; White *et al* 2007), and can be difficult to describe or compare using standard statistical approaches (*e.g.,* histograms and kernel density estimation are sensitive to bin width choice; Silverman 1986). I will perform statistical curve fitting on the data using Gaussian mixture modeling and likelihood ratio tests (McLachlan & Peel 2000, Thibault *et al* 2011, Legendre & Legendre 2012). Model fit to the data and pattern comparison will be evaluated using AIC (Burnham & Anderson 2002, Dornelas & Connolly 2008). I will also calculate the cumulative abundance profile (CAP) for paired manipulated and control communities (for modeled and empirical data), a metric that considers dissimilarity in both species identity and size structure (De Caceres *et al* 2010, 2013), thus linking trait and taxonomic diversity. These methods will allow me to accurately characterize size distributions and to evaluate their response to each change scenario. Deviations from model predictions and significant differences among paired empirical sites will indicate size sensitivity to change and provide clues for important mechanisms.

Objective 1: Develop data tools: Can we enable better research through data aggregation, database management & open-access tools?

Diversity data from manipulated sites that include information on abundance and body size of organisms at multiple spatial scales is rare and broadly distributed throughout the literature (Scholes *et al* 2008). Gaining access to detailed datasets is a key part of conducting macroecological research, and studies are often limited by the quantity and quality of available data. To investigate new research questions, it is imperative that data be collected and updated in a centralized database. Databases should be designed to maintain efficiency and reduce redundancy in the data and display high levels of interoperability, the ability to seamlessly merge information from different datasets (White *et al* 2013). Interoperability can be achieved using carefully defined ontologies, formal definitions of data terms and their relationships using mathematical logic (Kattge *et al* 2011). McGill has extensive knowledge of data informatics, and we will collaborate to standardize database methods designed for research and to make the data and methods accessible to researchers who are currently unfamiliar with working with "big data".

Objective 2: Describe and predict body size pattern behavior in modeled and empirical manipulated communities: Does environmental change influence size distributions at local scales?

Body size patterns have largely been regarded as static, and the relationship between patterns and change scenarios on ecological timescales remains unknown. I will test hypotheses of body size pattern response (Fig 1) and mechanistic drivers (ecological, phylogenetic, metabolic) to a suite environmental change scenarios of crucial conservation concern using statistical model simulation and empirical data (Table 1).

Environmental change scenarios	Hypothesized body size shift	Detailed description of expected body size diversity response to change	Taxonomic biodiversity metrics
Habitat fragmentation	Decrease	Large organisms have larger home range size, will be more susceptible to habitat size manipulation (Crooks 2002)	Decrease in total abundance, decrease in species richness
Predator removal	Decrease	Predators limited by gape size favor prey small enough to eat. Removal of predators may decrease mortality of small individuals (Scharf *et al* 2000, Woodward *et al* 2005)	Increased total abundance, variable impact on species richness
Fertilization (nutrient addition)	Increase	Adding energy to a system may lead to an increase in body size assuming larger organisms are more efficient at taking up resources quickly (Woodward *et al* 2005)	Increase total abundance, variable impacts on species richness
Temperature increase	Decrease	Small-bodied individuals are more commonly found in warm climates. Large bodied individuals may be more susceptible to stressful thermal environments (Freckleton *et al* 2003, Anderson-Teixiera *et al* 2012)	Increase total abundance, decrease species richness, or no change at all (compensation)

Table 1. Body size response hypotheses relevant to each of 4 major environmental change scenarios. Since body size is related to key ecological traits, such as dispersal, resource use, and nutrient cycling, I expect size sensitivity and direction of pattern response to be predictable based on the category of ecological change.

I will simulate body size pattern response under each change scenario using the free open-source NetLogo agent-based modeling (ABM) program (Wilensky 1999, 2005). ABMs use a series of simulation experiments to capture the complex effects of interacting adaptable individuals on patterns that emerge at community and ecosystem levels. Thus, it is possible to experiment with the connection between pattern emergence and individual behavior in a way that is impossible in reality (Railsback & Grimm 2011). ABMs are robust research tools that incorporate spatially explicit evolutionary processes and can easily translate into undergraduate classroom participatory learning (Wilensky & Stroup 1999, Wilensky 2005).

I will generate ABMs to evaluate and predict the behavior of body size distributions under each of the four environmental change scenarios (Table 1), by decomposing individual behavior into simple heuristic rules based on ecological interactions, phylogenetic relationships, and metabolic constraints. I will use existing scaling relationships to set metabolic constraints, and I will program individual "agents" with realistic body sizes (using mean & variation of existing data, Portal Project, Ernest *et al* 2009), life history (*e.g.,* lifespan, reproductive rate) and metabolic traits (*e.g.,* amount consumed daily). In a series of simulation experiments of increasing complexity, change scenarios will be modeled by altering parameters that simulate the affected abiotic and biotic features of the system (*e.g.*, habitat fragmentation, predator removal), which in turn, will influence individual behavior based on the provided constraints. Simulated control and manipulated site community-level results will be compared using conventional statistics (*see above*) to describe and characterize differences in size patterns, and sensitivity analyses will be used to analyze model response to variation in key parameters (Grimm & Railsback 2005). I will empirically test ABM predictions to each environmental change scenario using the datasets described above. Empirical sites will be categorized by change scenario and manipulated sites paired with nearby control sites to describe each size distribution, assess response to perturbation, and the fit of model predictions to empirical data. Testing model predictions with data is key for refining and retesting ABMs.

I expect my results to be directly important to peer scientists interested in using macroecology to describe and predict trait and taxonomic diversity in systems undergoing change. If size distributions are generally sensitive to ecological change, this research also has the potential to demonstrate how a general theory of ecology can be linked to conservation science. General predictions for *how* diversity will change and how changes influence ecosystem function are especially important for areas of high conservation concern, since logistics often limit researchers' ability to adequately monitor and respond to observed changes.

Question 3: Quantify scale-dependence of body size distribution form, behavior, and response: Do regional and local body size distributions have different responses to environmental change?

Body size patterns have typically been generated over large geographic scales, and the form of patterns across spatial scales is poorly understood (*but see* Brown & Nicoletto 1991, Bakker & Kelt 2000, Kelt *et al* 2009). Several mechanisms suggested to influence scale-dependence include resource competition, degree of specialization, range size (Brown & Nicoletto 1991), habitat heterogeneity (Hutchinson & MacArthur 1959, Morse *et al* 1985, Gardezi & daSilva 1999), and metabolic constraints (Brown *et al* 1993). Mechanisms influencing pattern behavior should differ depending on spatial scale of the pattern. I expect that regional-scale body size distributions should respond primarily to climate and habitat change, while local-scale distributions may be more sensitive to changes that influence population demographic processes such as competition, predation, and dispersal limitation.

I will modify my ABMs to provide output at three spatial scales (representing local patch, region, and continent) and assess the scale-dependence of size diversity response to each change scenario. I will also assess pattern response at two disturbance scales: small scale (local patch) and global (all patches experience change) and statistically compare simulated patterns across scales and between control and manipulated communities. Body size response emerging from model simulations will inform my model hypotheses for the scale-dependent response of empirical communities. Using data from BBS, FIA, and my new dataset, I will empirically test ABM predictions across at least three spatial scales, describe scale-dependence of body size pattern form, and evaluate cross-scale response under each change scenario.

Supp

Research Significance and Broader Impacts

Describing and predicting biodiversity in ecological systems undergoing change remains an important challenge for basic and conservation science. My results will advance the understanding of mechanisms influencing the form and scale-dependence of body size patterns, and develop a statistical modeling framework that can be used to evaluate other pattern dynamics (*e.g.,* taxonomic, functional). This project does not seek to suggest specific management regimes, but it does aim to enable scientists to determine what conservation questions *can* be addressed using macroecology and to reduce complexity in a way that enables rapid prediction of important ecosystem changes.

This project will result in a new multi-scale database that will facilitate sifting through large amounts of data and several open-access interactive models evaluating size structure under a suite of change scenarios. My overarching goal is to generate a framework for dynamic macroecology that can be used to evaluate other patterns. All research products will be open-access on websites (*i.e.*, GitHub, figshare) and publications will be hosted as preprints arXiv or PeerJ. I practice open science; my research ideas, products, and publications are publicly available and I follow a strong template for reproducible research.

One of my major motivations for this project is to engage with the scientific community and the public to disseminate ideas and to connect basic research with broader topics of general interest and conservation goals. I will share my research through traditional venues, via talks at scientific meetings (*e.g.*, Ecological Society of America, Gordon Research Conference), popular science talks, and via teaching and mentoring undergraduate and graduate students. I will integrate the computational and statistical toolkits I develop (databases, code, ABMs) with the classroom to help students gain skills that are becoming essential for scientists. My experience as an instructor for Software Carpentry (http://software-carpentry.org) has prepared me to be successful in this role, where I can connect computational skills with data and research. I will also play an active role in promoting the role of women in STEM careers via the NSF ADVANCE program at The University of Maine, "Girl Develop It", a local group in Burlington, VT, and Software Carpentry. "Girl Develop It" is open to all regardless of gender or skill-level, and the group hosts monthly informal "Code & Coffee" events, workshops to learn new skills, and meetings to make career connections, with a primary goal of getting teen girls and women excited about science and technology.

I will also share my research findings through social media. Tools such as Twitter, Mendeley, and GitHub allow rapid dissemination of information, broad discussion of questions and approaches, and researchers can connect with a wide audience, including non-scientists. On Twitter, I am followed by a diverse group of scientists across disciplines, science writers, educators, and non-scientists. My blog will be an educational resource that will explain my science in plain language, share results, discuss papers, and initiate conversations in perspectives on scientific education, mentoring, and outreach. McGill contributes to a widely read science blog, *Dynamic Ecology*, and will provide mentorship. I will assess my broader impacts yearly by tracking interactions on my blog, twitter, GitHub, and figshare accounts, collecting anonymous feedback from workshop and course instruction, and quantifying the number of new mentorships and collaborations gained from participating in the women in science and tech groups.

Training Objectives and Career Development

Training in data informatics, ABMs and advanced statistical techniques will complement my current strengths and areas of expertise. My grad research focused on fieldwork, data analysis, and macroecology, and my postdoc work uses remote sensing data and physiology to understand climate change impact on hummingbird diversity in the western hemisphere. I have limited experience with simulation-based approaches for studying trait-based forms of diversity, such as body size patterns. This project would give me strong training in statistical pattern analysis, model evaluation, and build new connections across macroecology, physiology, conservation, and statistics. I will receive training in ABMs using textbooks (*e.g.*, Railsback & Grimm 2011), attending courses (Univ. of Maine - Math & Statistics Dept), and interacting with the Agent-Based Computing research cluster (Engineering & Math, Univ. of Vermont).

Educational training will culminate with teaching basic undergraduate biology and an independently developed course on computational biology. I will use the Center for Excellence in Teaching and Assessment at University of Maine to learn research-based techniques that increase student engagement, empowerment, and concept mastery. Guest lecturing in introductory biology courses will help me explore learning challenges at these levels and gain familiarity with different teaching approaches. I will mentor at least one undergraduate and graduate student in the McGill lab. I will participate in K-12 outreach events via the Sustainability Solutions Initiative (SSI) to gain public communication training.

I am committed to pursing a career in ecology as a faculty member at a university where I can integrate research and education. These training objectives are designed to prepare me to work with students in and out of the classroom to conduct cutting-edge interdisciplinary research. The statistical and computational skills I will gain will complement my field ecology background and provide a solid foundation for this career path. This will enable me to act as a bridge between computational scientists and field ecologists and to develop a strong teaching philosophy geared towards empowering students in math and biology.

Choice of Sponsors and Host Institution

Brian McGill (Univ. of Maine)-Math-Stats/Biology & Nicholas Gotelli (Univ. of Vermont)-Biology/Modeling

McGill's and Gotelli's research interests and skill sets make them ideal co-sponsors for this project's intellectual development and practical implementation. **McGill** is an Associate Professor in the School of Biology and Ecology, the Sustainability Solutions Initiative (SSI), and holds a joint appointment with the Climate Change Institute. He has a B.S. in math, worked for 10 years in the software industry, and was an NSF Bioinformatics postdoc. He is known for his work synthesizing biodiversity, informatics and global change research and has extensive experience with advanced spatio-temporal statistical methods and informatics approaches. **Gotelli** is a Professor in the Dept. of Biology. He is an experimental ecologist, but has also published null and stochastic ecological models. He has made significant contributions in the study of complex biological systems and climate change, ranging from individual to ecosystem-level focus, using advanced modeling methods. His experimental and modeling background will be valuable in navigating the statistical complexities of this project.

My primary location will be the **University of Maine**, an ideal place for this research because of the SSI, whose mission statement shares a conceptual goal with my project: to connect basic and applied sciences with a goal to improve and preserve ecosystem health. I will also spend a significant amount of time at the **University of Vermont** for modeling mentorship and collaboration opportunities in the Complex Systems Center and Bioinformatics Center. My interests fit within the research scope of current grad students and postdocs in McGill and Gotelli's labs, which will provide excellent platforms for mentorship and sharing skills. Both universities have strong collaborative environments that will allow me to learn cutting-edge computational/statistical modeling techniques from faculty across the Biology, Math and Statistics, Computer Science, and Economics departments.

Timeline

Dates	Intersections of Math and Biology Training	Education & Communication Training	Dissemination of Results
Year 1	**Phase 1,** Database design, data collection from literature, parameterize agent-based model	Classroom observation, receive blogging mentorship & advice	Attend & present results at annual scientific meetings. Participate in K-12 outreach via SSI. Use social media. Submit first publications
Year 2	**Phase 2,** Continue development & expand agent-based model, analyze empirical data	Guest lectures, assessment from mentor & students, blog about research & computation methods	Present results at scientific meetings. Publish data, models, & code in repositories (GitHub, figshare, NetLogo). Submit publications
Year 3	**Phase 3,** Finalize analyses. Develop simple agent-based model for undergrad classroom	Teach independent course, mentor undergrad researcher. Blog about research & teaching methods	Organize a seminar on teaching methods, present results at annual scientific meetings. Submit final publications

References

Anderson-Teixeira, K.J., F.A. Smith and S.K.M. Ernest. 2012. Climate Change. *In* Metabolic Ecology: A scaling approach. Eds. R.M. Sibly, J.H. Brown, and A. Kodric-Brown. Wiley & Sons, Ltd. Oxford.

Bakker, V.J. and D.A. Kelt. 2000. Scale-dependent patterns in body size distributions of neotropical mammals. Ecology. 81: 3530-3547.

Beck, J., L. Ballesteros-Mejia, C. M. Buchmann, J. Dengler, S. A. Fritz, B. Gruber, C. Hof, F. Jansen, S. Knapp, H. Kreft, A. Schneider, M. Winter, and C. F. Dormann. 2012. What's on the horizon for macroecology? Ecography. 35: 673-683.

Bianchi, G., H. Gislason, K. Graham, L. Hill, X. Jin, K. Koranteng, S. Manickchand-Heileman, I. Payá, K. Sainsbury, F. Sanchez, and K. Zwanenburg. 2000. Impact of fishing on size composition and diversity of demersal fish communities. ICES Journal of Marine Science. 57: 558-571.

Brown J. H., P. A. Marquet, and M. L. Taper. 1993. Evolution of body size: consequences of an energetic definition of fitness. The American Naturalist. 142: 573-584.

Brown, J.H. and P.F. Nicoletto. 1991. Spatial scaling of species composition: body masses of North American land mammals. The American Naturalist. 138: 1478-1512.

Brummit, N. and E.N. Lughadha. 2003. Biodiversity: Where's hot and where's not. Conservation Biology. 17: 1442-1448.

Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodal inference: a practical information-theoretical approach. 2nd Ed. Springer-Verlag, New York.

Crooks, K.R. 2002. Relative sinsitivities of mammalian carnivores to habitat fragmentation. Conservation Biology. 16: 488-502.

De Caceres, M., X,. Font, and F. Oliva. 2010. The management of vegetation classifications with fuzzy clustering. Journal of Vegetation Science, 21, 1138–1151.

De Cáceres, P. Legendre, and F. He. 2013. Dissimilarity measurements and the size structure of ecological communities. Methods in Ecology and Evolution. doi: 10.1111/2041-210X.12116

Diniz-Filho, J.A.F. and L.M. Bini. 2008. Macroecology, global change, and the shadow of forgotten ancestors. Global Ecology and Biogeography. 17: 11-1

Diniz-Filho, J.A. F. and N. M. Torres. 2002. Phylogenetic methods and the geographic range size – body size relationship in new world terrestrial carnivora. Evolutionary Ecology. 16: 351-367.

Dornelas, M., C.U. Soykan, and K.I. Ugland. 2011. Biodiversity and disturbance. *In* Biological Diversity. Eds. A.E. Magurran and B.J. McGill. Oxford University Press, Oxford.

Dornelas, M. and Connolly, S. R. 2008. Multiple modes in a coral reef species abundance distribution. Ecology Letters. 11: 1008-1016.

Dunning, J.B.J. 2008. CRC handbook of avian body masses. CRC Press, Boca Raton, FL.

Duplisea, D.E. and M. Castonguay. 2006. Comparison and utility of different size-based metrics of fish communities for detecting fishery impacts. Canadian Journal of Fisheries and Aquatic Sciences. 63: 810-820.

Ernest, S. K. M., T. J. Valone, and J. H. Brown. 2009. Long-term monitoring and experimental manipulation of a Chihuahuan Desert ecosystem near Portal, Arizona, USA. Ecology. 90: 1708.

Fisher, J.A.D., K.T. Frank, and W.C. Leggett. 2010. Dynamic macroecology on ecological time-scales. Global Ecology and Biogeography. 19: 1-15.

Folke, C., S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, L. Gunderson, and C.S. Holling. 2004. Regime shifts, resilience, and biodiversity in ecosystem management. Annual Review of Ecology, Evolution, and Systematics. 35: 557-581.

Freckleton, R.P., P.H. Harvey, and M. Pagel. 2003. Bergmann's Rule and body size in mammals. American Naturalist. 161: 821-825.

Gardezi, T. and J. da Silva. 1999. Diversity in relation to body size in mammals: a comparative study. The American Naturalist. 153: 110-123.

Grant, P.R. and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. Science. 296: 707-711.

Grimm, V. and S.F. Railsback. 2005. Individual-based modeling and ecology. Princeton University Press.

Hooper, D.U., F.S. Chapin III, J.J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J.H. Lawton, D.M. Lodge, M. Loreau, S. Naeem, B. Schmid, H. Setälä, A.J. Symstad, J. Vandermeer, and D.A. Wardle. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecological Monographs. 75: 3-35.

Hutchinson, G. E. and R. H. MacArthur. 1959. A theoretical ecological model of size distributions among species of animals. The American Naturalist. 43: 117-125.

Jennings, S., J.K. Pinnegar, N.V.C. Polunin, and T.W. Boon. 2001. Weak cross-species relationships between body size and trophic level belie powerful size-based trophic structuring in fish communities. Journal of Animal Ecology. 70: 934-944.

Kattge, J., K. Ogle, G. Bonisch, S. Diaz, S. Lavorel, J. Madin, K. Nadrowski, S. Nollert, K. Sartor, C. Wirth. 2011. A generic structure for plant trait databases. Methods in Ecology and Evolution. 2: 202-213.

Keith, S.A., T.J. Webb, K. Böhning-Gaese, S.R. Connolly, N.K. Dulvy, F. Eigenbrod, K.E. Jones, T. Price, D.W. Redding, I.P.F. Owens, and N.J.B. Isaac. 2012. What is macroecology? Biology Letters. 8: 904-906.

Kelt, D. A. and M. D. Meyer. 2009. Body size frequency distributions in African mammals are bimodal at all spatial scales. Global Ecology and Biogeography. 18: 19-29.

Legendre and Legendre. 2012. Numerical Ecology, 3rd Edition. Elsevier. Amsterdam.

McLachlan, G. and D. Peel. 2000. Finite mixture models. John Wiley and Sons, Inc., New York.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature. 403: 853-858.

Morse, D. R., J. H. Lawton, M. M. Dodson, and M. H. Williamson. 1985. Fractal dimension of vegetation and the distribution of arthropod body lengths. Nature. 314: 731-733.

Palmer, M.W. and P.S. White. 1994. Scale dependence and the species-area relationship. The American Naturalist. 144: 717-740.

Railsback, S.F. and V. Grimm. 2011. Agent-based and individual-based modeling: A practical introduction. Princeton University Press.

Scharf, F. S., F. Juanes, R. A. Rountree. 2000. Predator size-prey relationships of marine fish predators: interspecific variation and effects of ontogeny and body size on trophic-niche breadth. Marine Ecology Progress Series. 208: 229-248.

Scholes, R.J., G.M. Mace, W. Turner, G.N. Geller, N. Jurgens, A. Larigauderie, D. Muchoney, B.A. Walther, and H.A. Mooney. 2008. Toward a global biodiversity observing system. Science. 321: 1044-1045.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. Chapman and Hall, New York.

Smith, F. A., S. Kathleen Lyons, S. K. Morgan Ernest, Kate E. Jones, Dawn M. Kaufman, Tamar Dayan, Pablo A. Marquet, James H. Brown, and John P. Haskell. 2003. Body mass of late Quaternary mammals. Ecology. 84:3403.

Smith, F.A., J. H. Brown, J. P. Haskell, S. K. Lyons, J. Alroy, E. L. Charnov, T. Dayan, B. J. Enquist, S. K. M. Ernest, E. A. Hadly, K. E. Jones, D. M. Kaufman, P. A. Marquet, B. A. Maurer, K. J. Niklas, W. P. Porter, B. Tiffney, and M. R. Willig. 2004. Similarity of mammalian body size across the taxonomic hierarchy and across space and time. The American Naturalist. 163: 672-691.

Supp, S.R., X. Xiao, S.K.M. Ernest, and E.P. White. 2012. An experimental test of the response of macroecological patterns to altered species interactions. Ecology. 93: 2505-2511.

Supp, S.R. 2013. Local and regional drivers of biodiversity: from life-history traits to system-level properties. Graduate Dissertation. Utah State University. http://digitalcommons.usu.edu/etd/1503

Thibault, K.M., E.P. White, A.H. Hurlbert, and S.K.M. Ernest. 2011. Multimodality in the individual size distributions of bird communities. Global Ecology and Biogeography. 20: 145-153.

Thomas, C.D., A. Cameron, R.E. Green, M. Bakkenes, L.J. Beaumont, Y.C. Collingham, B.F.N. Erasmus, M. Ferreira de Siqiera, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A.S. van Jaarsveld, G.F. Midgley, L. Miles, M.A. Ortega-Huerta, A.T. Peterson, O.L. Phillips, and S.E. Williams. 2004. Extinction risk from climate change. Nature. 427: 145-148.

USDA Forest Service. 2010. Forest inventory and analysis national core field guide (Phase 2 and 3). Version 4.0. USDA Forest Service. Forest Inventory and Analysis. Washington, D.C., USA

USGS Patuxent Wildlife Research Center. 2012. North American Breeding Bird Survey, ftp dataset, version 2012.0 (ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/datafiles/).

White, E.P., S.K.M. Ernest, A.J. Kerkhoff, and B.J. Enquist. 2007. Relationships between body size and abundance in ecology. Trends in Ecology and Evolution. 22: 323-330.

White, E.P., E.M. Baldridge, Z.T. Brym, K.J. Locey, D.J. McGlinn, and S.R. Supp. 2013. Nine simple ways to make it easier to (re)use your data. Ideas in Ecology and Evolution. 6: 1-10

Wilensky, U. 2005. NetLogo Echo model. http://ccl.northwestern.edu/netlogo/models/Echo. Center for Connected Learnning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. and W. Stroup. 1999. HubNet. http://ccl.northwestern.edu/netlogo/hubnet.html. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL

Woodward, G., B. Ebenman, M. Emmerson, J.M. Montoya, J.M. Olesen, A.Valido, and P.H. Warren. 2005. Body size in ecological networks. Trends in Ecology and Evolution. 20: 402-409.

Yom-Tov, Y. and J. Yom-Tov. 2005. Global warming, Bergmann's rule and body size in the masked shrew *Sorex cinereus*. Journal of Animal Ecology. 74: 803-808.

Data Management Plan

Research Products

The project will develop two main products, a new dataset to be used for macroecological inquiries and several publicly available agent-based simulation models using the NetLogo software platform. The data will be compiled from experimentally manipulated communities in the literature, in collaboration with McGill and the models and model testing will be developed in collaboration with McGill and Gotelli.

Data Format and Content

I will maintain an open lab notebook on my website. The open lab notebook will contain details on the development and design of my projects, including the new database and the agent-based model simulations. I will develop all code using freely available software, R, Python and NetLogo. Code will be stored in public GitHub repositories (www.github.com/sarahsupp), with detailed README files outlining the purpose of the code, and how to use the files contained in the repository.

All data collected for the project and all accompanying metadata will be stored and managed using an SQL database. All data tables will be stored as .csv files, which are considered standard current best practice. The database will be compiled to be optimized for my (and other) macroecological inquiries. We will use Level V ecological metadata language (EML) documents to optimally describe the datasets and structures including provenance information detailing all reference datasets used in the compilation.

I will target journals for the results of this research project that provide open-access downloadable PDF copies of the manuscript on the web. I will link to these publications from my website, blog, and twitter.

Data Access and Archiving

The new dataset will be published as a Data Paper on *Ecological Archives,* which is widely accessible and can be cited and used by other scientists. Data Papers are permanently housed on ESA's servers providing long-term stable access to the datasets. An online abstract published in *Ecology* actively promotes and advertises the existence of the new data, increasing the potential that the scientific community will utilize it in other studies. Figures or results generated during the course of the project that do not make it into published documents will be shared on figshare (www.figshare.com), an open-access site that provides a permanent and citable doi for all uploaded products and expands the scientific reach of previously unpublished data or results. All code will be open-access on GitHub (www.github.com/sarahsupp) and will enable my research results to be fully replicated, including figures, tables, and manuscripts.

Software Storage and Provision

All code for analysis of the data will be developed in R or Python, and stored and served from a centralized public GitHub software repository. All collaborators involved in writing code will commit their changes to this central source. In this manner, we will facilitate real-time peer review of each addition and modification to the code by the participants in the project. We will implement standard testing procedures to prevent errors. The simulation model will be constructed in collaboration with McGill and Gotelli and will be housed along with metadata and instructions on the NetLogo website (http://ccl.northwestern.edu/netlogo/) as an open-access tool. A simpler model designed as an in-class participatory simulation will be housed on HubNet (http://ccl.northwestern.edu/netlogo/hubnet.html).

Reuse and Redistribution

All research products will be clearly licensed following the guidelines in the Panton Principles (http://pantonprinciples.org) to communicate the way the data can be used. I will use the Creative Commons Zero (CC0) public domain license or the Berkeley Software Distribution (BSD) 2-clause

license for my research products (*i.e.*, code, models, and presentation material). These are liberal licenses that allow software to be accessed and used by others, provided that the author(s) are cited and retain the copyright to the material. Manuscripts will be published in preprint servers, including arXiv (arXiv.org) and PeerJ (https://peerj.com), to open them up for wider review, audience and usage by the scientific community. Manuscripts will be published in peer-reviewed journals, with preference towards journals that have rapid review times and/or offer open-access options for publishing (*e.g.*, Ecology, American Naturalist, Ecology and Evolution, PLoS).

Roles and Responsibilities

Sarah Supp will be primarily responsible for data collection, data management, and data dissemination. Brian McGill will provide guidance on database development and management and on interacting with the community in scientific blogs. McGill and Gotelli will provide support for developing, managing, and sharing code for the agent-based models and in communicating with the NetLogo user community. The data archiving for this project will provide strong provenance and long-term open-access to data, code, and computational tools into the future, after the grant ends.

Teaching Plan

One of my major motivations in pursing this fellowship is the opportunity to gain training and experience in undergraduate education. Most undergraduates will take a biology course during their education, regardless of their major discipline or career goals, but many students leave the classroom frustrated and demotivated. In particular, many introductory biology courses poorly connect core concepts to the larger framework of scientific knowledge or other educational disciplines. A focus on fact memorization and a failure to root facts in the context of real-world problems leads to disconnect between the classroom, the scientific process, and what students perceive as relevant to their lives. In general, this represents a major pedagogical challenge for science teachers. I have included a teaching plan as part of my training goals because I feel that developing a strong student-centered focus towards teaching is a significant challenge and an important skill to develop.

My teaching plan is a tiered approach that includes training in the formal classroom as an undergraduate mentor and in public communication (Table 1). Brian McGill at the University of Maine (UMaine) will be my primary teaching mentor, but I will also receive teaching mentorship and advice from Nicholas Gotelli at the University of Vermont (UVM). Both have received awards for outstanding undergraduate and graduate teaching. My training will occur throughout the entire postdoc, with a focus on instructing an independently developed course and mentoring undergraduate and graduate researchers in the third year. Gaining experience in multiple teaching venues (*e.g.*, classroom, mentorship, and outreach), will give me invaluable training for pursuing an academic career and for becoming an effective teacher and communicator.

Classroom training

The main focus of my teaching plan will be in teaching and mentoring undergraduates. During years 1-2, I will gain teaching experience by observing introductory biology courses (*e.g.,* BIO100 - Basic Biology, BIO200 - Biology of Organisms) and conducting guest lectures in interdisciplinary classes (*e.g.*, BIO319 - General Ecology, ECO590 - Mathematical Models of Social Evolution, CS365 - Introduction to Scientific Modeling). During year 3, I will develop and teach my own course.

Core Course Experience

I will gain teaching experience during the first year by observing teaching methods used in introductory biology courses primarily at UMaine, but also during my visits to UVM. During the second year, I will guest lecture in one or more of these core courses, as opportunities become available. I have had the opportunity to be a teaching assistant (*e.g.,* BIOL5560 –*Ornithology*, BIOL2420 - *Human Physiology*), and to guest lecture in several courses during my graduate work (*e.g.*, BIOL4750 – *Macroecology*, BIOL5250 - *Evolutionary Biology*) and I independently developed and instructed a course (BIOL5250 - *Evolutionary Biology*) at Utah State University. These opportunities gave me important teaching experience, but nearly all were upper-level courses, which present significantly different challenges from introductory courses. For these reasons, opportunities to observe, teach, and receive assessment on effective methods and approaches to teaching introductory level biology courses will be extremely important to me. In addition, I will gain interdisciplinary teaching experience by guest lecturing in a mid-level computer science course at UVM (CS302 – *Modeling Complex Systems*) where I can use my research, data, and agent-based model as the basis for a "case study" for the class to grapple hands-on with mathematical/computational approaches to solve scientific problems and analyze results.

Independent Course Details

At UMaine, I will independently develop and teach an undergraduate course*, Programming for Biologists*. This course include, but not be limited to, basic skills in R, SQL, and the Unix Shell, would be a hands-on interactive class modeled on successful courses taught by Ethan White (Utah State University) and short-courses that I have already instructed through Software Carpentry. The course will introduce

simple, practical programming tools to biology undergraduates. Acquiring even a basic knowledge of computational approaches will allow students to be better prepared for graduate school and for careers in science. These skills will enhance student research experiences by enabling students to conduct in-depth data analysis, generate better figures, and to connect with the growing community of scientist-programmers.

In the course, students will learn basic programming in R, a language that is widely used among ecologists and is easily translatable into other common scientific programming languages (*e.g.,* Python, MatLab). Students will also learn how to use shell scripts, basic database management (SQL) and agent-based modeling approaches (using HubNet software). All of these computational tools are freely available, and will require no extra monetary investment from the university or students. A 'flipped classroom' model is ideal for this course: students follow lecture materials (*e.g.*, reading assignments, online tutorials) outside of class, and engage in hands-on activities and concept practice during class time. This model enables students to interact directly with their instructor and with each other to problem-solve and discuss core concepts. After learning core concepts, students will conduct a small biology research project of their choice using the tools learned in class (*i.e.*, coding, modeling and/or dataset analysis). Gotelli has taught courses in R geared toward graduate students, and can provide insight via his personal experiences at UVM. The UMaine School of Biology and Ecology does not currently have a course focused on computational skills for scientists as part of its curriculum, but McGill has been approached regarding a departmental interest in developing such a course for future students.

Assessment

The UMaine Center for Excellence in Teaching and Assessment (CETA) is dedicated to improving teaching and learning quality of university instructors. I will attend CETA organized events for graduate students, postdocs, and faculty geared towards assessing and improving teaching skills. CETA provides informational meetings about student services (*e.g.*, Disability Support Services), opportunities to connect with a community of other new teachers at UMaine, and collaborative assessment through the Peer Consulting program, in which I would participate. Peer Consultants engage in a multi-part assessment program that includes reviewing teaching materials, classroom observation, and student evaluations. In addition to formal assessment and training via CETA, I will receive direct feedback from my teaching mentor, McGill, and I will use weekly interactive discussions, short quizzes and anonymous polls to assess student learning and to give me feedback on what material I covered effectively and what material I needed to revisit. This mixed format will provide frequent feedback and help me tweak lesson plans on-the-fly to better address difficult concepts.

Teaching-Focused Learning Objectives	Year 1	Year 2	Year 3
Student learning styles	Observe intro BIO, connect with online community, read educational literature	Observe upper-level BIO courses, blog about education styles	Experiment with research-based teaching methods, follow-up with evaluations
Effective teaching	Guest lecture, participate in a teaching workshop	Guest lecture an entire module in intro BIO and CS courses	Instruct independent course, guide small research projects
Teaching philosophy	Read examples	Revise based on observations and experiences	Revise based on experiences, share strategies in a seminar
Assessment	Mentor-based	Mentor and student-based, CETA	Student-based, ongoing throughout course, CETA
Mentoring	Participate in weekly lab meetings	Weekly lab meetings, undergrad researchers	Mentor undergraduate & graduate researchers

Table 1. *Timeline for achieving training objectives and receiving assessment from mentors, students, and the Center for Excellence in Teaching and Assessment (CETA)*

Integrating Research and Teaching

The student-centered, active learning classroom I will develop is ideal for integrating my own research goals with teaching and student learning objectives. My research plan involves generating a new large

dataset and several interactive agent-based models. I can bring both of these products into the classroom for students to use and explore. For example, the dataset is ideal to use for class examples and small projects where students can practice using databases, analyzing data, and plotting results. In addition to practice problems sets, I will encourage students to create their own research experiences by asking their own questions using the dataset. The agent-based model can be used in a similar way: to demonstrate how computational concepts are used to evaluate biological questions, and to allow students to re-parameterize and explore the model to ask their own questions. In my more advanced course, students will learn to build their own agent-based models to test hypotheses, whereas in a core biology course, students will use an existing model with sliding-scale parameters to experiment with simulation outcomes as they violate assumptions and change parameters (*e.g.*, Hardy-Weinberg models). One of my main goals in teaching is to empower students with the skills and toolkits necessary to conduct their own scientific research.

Research mentorship training

In addition to teaching in the classroom, helping undergraduates gain concrete research experience is essential to training the next generation of scientists. I will participate in weekly meetings with both the McGill and Gotelli labs (depending on my current location). During this time, I will build collaborative working relationships with undergraduate and graduate students in both labs, and share knowledge and skill sets across the two groups. I will also mentor at least one undergrad in the McGill lab developing an independent research project. At UMaine, students in the School of Biology and Ecology are encouraged to seek out independent research experiences. The Center for Undergraduate Research (CUGR) at UMaine is a member of the national Council on Undergraduate Research (CUR) and provides a venue for advertising research opportunities to undergraduates. I will encourage undergraduate researchers to present their work at one of the CUR biennial conferences, which are specifically aimed at undergraduate research and scholarship. CUR also provides learning opportunities to teachers in promoting and improving undergraduate research experiences, in which I would participate.

Science communication training

Effective communication with the public is becoming more important for the scientific community. Using social media outlets such as blogging and twitter represents an important way to reach a broad audience. McGill is a member of the Dynamic Ecology blog, a multi-contributor science blog that currently includes Jeremy Fox (Associate Professor of Population Ecology, University of Calgary), Meghan Duffy (Assistant Professor of Ecology and Evolutionary Biology, University of Michigan), and Christopher Klausmeier (Associate Professor of Plant Biology, University of Minnesota). The blog reaches ~1000 readers/week and McGill is well qualified to give blogging mentorship and advice. I will blog and tweet (I currently have > 330 twitter followers) regularly about my research progress, teaching experiences and methods, and general thoughts on the intersection of ecology, programming, and statistics. My blog will be aimed at a scientific audience, but will be basic enough in scope to be read by non-scientists or by undergraduates.

Sharing Experiences

At the end of my teaching fellowship, I will organize and lead seminars for grad students, postdocs, and faculty at UMaine and at UVM on student-centered teaching strategies. Seminar discussion will focus on bringing toolkits (*e.g.,* computational methods, data analysis, bioinformatics) and interdisciplinary concepts into the classroom using research-based teaching methods, and on strategies for translating computational concepts to a broad scientific audience. I will share my results and new insights from the discussion at the seminars at UMaine and at UVM on my blog and on the Teaching Software Carpentry blog (http://teaching.software-carpentry.org) in order to reach a wider audience of scientists and teachers.