TIEE

Teaching Issues and Experiments in Ecology - Volume 14, August 2018

EXPERIMENTS

Exploring how climate will impact plant-insect distributions and interactions using open data and informatics

Wendy L. Clement[1], Kathleen L. Prudic[2], and Jeffrey C. Oliver[3]

[1]Department of Biology, The College of New Jersey, 2000 Pennington Road, Ewing, NJ 08638, clementw@tcnj.edu

[2]Department of Entomology, University of Arizona, PO Box 210036, Tucson, AZ 85721, klprudic@email.arizona.edu

[3]Office of Digital Innovation & Stewardship, University of Arizona Libraries, 1510 E. University Blvd. Tucson, AZ 85721, jcoliver@email.arizona.edu

ABSTRACT
Open data repositories, including those from citizen science efforts, are rich sources of research grade data that are becoming key to asking and answering questions in ecology. Simultaneously, informatics tools are becoming increasingly accessible to the non-specialist and are more commonly integrated into the college curriculum of biology students. This series of 3 classes (~360 minutes of in class activity time) guides students on how to collect, curate, and analyze citizen science data using common research computing tools: R, RStudio, Git, and GitHub. These are *in silica* experiments examining (1) the species distributions of butterflies and their host plants based on observations submitted to the web platform iNaturalist and (2) how those distributions may change in the future due to global climate change. Students will download and install software, retrieve and curate citizen science data, model the occurrence data to produce a species distribution of butterfly and host plant, and develop hypotheses on how climate change may or may not affect the future distribution of butterfly and host plant. Students then test these hypotheses using estimates of future climate variables, evaluate the strength of their results, and present a summary of these explorations to their peers using additional class time if desired. This series of experiments will result in 4 group products and 1 individual product for evaluation.

KEYWORD DESCRIPTORS

- **Ecological Topic Keywords:** Biodiversity, climate change, citizen science, ecoinformatics, open data, R programming, species distribution modeling, species interactions

- **Science Methodological Skills Keywords:** Collaborative web platforms, data analysis, geospatial mapping skills, hypothesis generation and testing, office productivity software, open data, oral presentation, R programming, version control, visualizing data

- **Pedagogical Methods Keywords:** Computing, cooperative learning, critical thinking, guided inquiry, metacognition, misconceptions, open-ended inquiry

CLASS TIME

This module requires three 2-hour lab periods plus time for in-class presentations (allotted time varies depending on group and class size and whether presentations are face to face or online). In the initial 2-hour lab period, class time is used to introduce students to the topics of ecoinformatics, open data repositories, citizen science, and climate change. Student groups then identify a butterfly-host plant interaction and learn about its natural history. In the second 2-hour period, students download the relevant data from iNaturalist and then are given instructions on how to create a species distribution map. Students develop a hypothesis concerning how climate change may impact this specific butterfly-host plant pair in the next 50 years. In the third 2-hour lab period, students evaluate their hypothesis by creating another series of species distribution maps using predicted climate climate variables for the next 50 years. Students then synthesize the information into a final group oral presentation and individual written assignment.

OUTSIDE OF CLASS TIME

Instructors' time preparing for class will depend on familiarity with research computing resources including R, RStudio, and GitHub. Students may spend several hours gathering natural history data on their butterfly-host plant interaction, (re-)analyzing their data, interpreting results, creating figures, and developing presentations. We estimate 6-8 hours total of out of class time depending on prior experiences and skills. Collaborative platforms such as Google Docs, Open Science Framework, Slack etc., can help facilitate group work outside of class.

STUDENT PRODUCTS

Students will have a variety of written, visual, and oral products for evaluation. Student Product 1: Natural History of Butterfly-Host Plant Interaction; Student Product 2: Species distribution maps (SDMs) and hypothesis; Student Product 3: Future Species Distribution Models and Hypothesis Evaluation; Student Product 4: Presentation of project and results; Student Product 5: Synthesis and reflection of group projects. Student Products 1-4 are assessed by group and Student Product 5 is assessed individually.

SETTING

Students will complete this exercise using a computer with access to the internet. Possible implementations include: 1) a shared computer lab with the ability to install software, or 2) a wi-fi enabled classroom in which students can use individual laptops with software installed prior to class. Guides for downloading relevant software have been included as supplementary documents.

COURSE CONTEXT

This exercise was first implemented in an upper level undergraduate course, *Ecology and Evolution of Plant Insect Interactions* at The College of New Jersey in Fall 2017. This course meets twice a week for two hours per meeting (4 hours/week total) and is capped at 24 students. The class was divided among groups of three or four students during this module. This activity could be scaled up to hundreds of students depending on the instructor's capabilities with the research computing skills and class support. The module would be particularly amenable to large enrollment courses with a laboratory or recitation component.

INSTITUTION

This activity was performed at a public, 4-year, primarily undergraduate institution with upper level biology students.

TRANSFERABILITY

This exercise is designed to scale in a face-to-face or online environment given the students have access to download the freely distributed software (R [https://www.r-project.org/], RStudio [https://www.rstudio.com/products/rstudio/download/#download], Git [https://git-scm.com/downloads]), and source code (GitHub [https://github.com]). Species pairs can be exchanged for any symbiotic species interactions based on course content and the learning goals. This content is designed to be flexible; any species on iNaturalist can be analyzed with minor modifications to the code.

These activities could be scaled to larger classes; however, such a venture would depend greatly on instructor knowledge, student experience, and the human capital resources necessary to properly implement the activities at scale. Of particular importance to a large classroom setting are skilled assistants (e.g., teaching assistants) necessary to assist and troubleshoot various software challenges during instruction. Archiving the materials on Open Science Framework allows the authors to update the code and work with instructors as needed. As programming and data science skills become more prevalent in pre-college instruction and curricula, we anticipate this activity becoming more applicable and accessible to introductory college courses.

ACKNOWLEDGEMENTS

This activity would not have been possible without the amazing open science and data sources specifically iNaturalist (Ken-ichi Ueda and Scott Loarie) and National Oceanic and Atmospheric Administration (NOAA) climate predictions and the willingness of the students enrolled in the Fall 2017 Plant-Insect Interactions course at The College of New Jersey to participate in this project.

AUTHOR CONTRIBUTIONS

All authors contributed to this activity. WLC's course on Plant-Insect Interactions was the catalyst for developing this group project. WLC worked with KLP and JCO to design an exercise and test it in the classroom. WLC designed and implemented the learning evaluation assessments. JCO designed and packaged the code for the students. KLP worked with available data on iNaturalist to assess feasibility of the group project and wrote a first draft of the manuscript that the other authors edited.

SYNOPSIS OF THE EXPERIMENT

Principal Ecological Question Addressed

How will climate change affect the distributions of butterflies and their larval hosts across a continent over time?

What Happens

Students identify a butterfly-host plant interaction, download and analyze distribution data, and develop a hypothesis regarding the change in distributions in 50 years based on overlapping species distribution maps they generate. Data are obtained from the citizen science project iNaturalist, uploaded into RStudio, and analyzed with R code available on GitHub. To evaluate their hypotheses, students develop a predictive distribution of butterfly and host plants for the year 2070 using forecast climatic variables. Groups give an oral presentation on their work and individuals write an essay to synthesize the class findings and reflect on the experience.

Experiment Objectives

1. Describe how biodiversity science data initiatives, such as ecoinformatics, can make use of citizen science and museum digitization efforts to ask and inform questions in ecology.
2. Use research computing tools (Citizen science crowd sourced data, R programming language, GitHub collaborative web platform, data visualization) to study a butterfly-host plant interaction.
3. Communicate findings in the form of an oral presentation
4. Synthesize potential outcomes of the effects of climate change on plant-insect interactions

Equipment/ Logistics Required

- Computers
- Access to internet
- Access and support for R, RStudio, Git, and GitHub (all freeware)
- Access to Microsoft® PowerPoint or other software to prepare slides for an oral presentation

Summary of What is Due

Students will have five products, and each is submitted for evaluation (examples and rubrics are provided in supplementary data for Student Products 2, 3, and 4).

Student Product 1: Natural History of Butterfly-Host Plant Interaction
Student Product 2: Species distribution maps (SDMs) and hypothesis
Student Product 3: Future Species Distribution Models and Hypothesis Evaluation
Student Product 4: Presentation of project and results
Student Product 5: Synthesis and reflection of group projects

DETAILED DESCRIPTION OF THE EXPERIMENT

Introduction

Plant-insect interactions do not occur in a vacuum, but rather in complex ecosystems involving interactions with other biotic and abiotic components of their communities. Global climate change, and specifically global warming, have been shown to affect the distribution, phenology, and persistence of plants and insects (Bale et al., 2002; Biesmeijer et al. 2006; Kelly and Goulden, 2008; Chung et al., 2011) as well as other organisms (Moritz et al., 2008; Tingley et al., 2009; Rubidge et al., 2011; Rowe et al, 2015). Studying these complex interactions often requires a vast amount of data from the past and present difficult for any one scientist to collect in a reasonable period of time (Hochachka et al, 2012; Hurlbert and Liang 2012; Todd et al., 2016). The creation of massive online citizen science web platforms and smartphone applications has resulted in real-time biodiversity data collection across continental and global scales (e.g. iNaturalist, eBird, eButterfly, Project BudBurst, Mushroom Observer; see Silvertown 2009 for a more extensive list). This large scale data acquisition of the world's biodiversity can help us address questions regarding the future of plant-insect interactions and ecosystem conservation at large. With these data, continental-scale spatial analyses allow us to quantify the status of each of Earth's inhabitants, assess the effectiveness of our conservation efforts, and determine when additional intervention is warranted.

Climate change is happening fast and impacting the abundance and distribution of many plants and animals - not just plant-insect interactions. Sometimes these interactions are positively impacted, other times negatively impacted. Examples of changing interactions over the past century have already shown major changes and identified likely future threats to conservation (Kerr et al. 2015, Matthew et al. 2017). We expect these types of trends to continue in the near future, particularly at the level of the ecosystem. Predictive models using information on future climatic variables are an important tool to estimate where and when a species may occur in the next 25, 50, or 100 years. Many scientists have been developing forecast models, such as those at the Geophysical Fluid Dynamics Laboratory who have developed an Earth System Model (ESM) which models a variety of atmospheric variables and cycles, as well as plant ecology and land use (https://www.gfdl.noaa.gov/earth-system-model/).

The goal of this project is to study the changing distributions of plant-insect interactions, specifically a butterfly and host plant, over the next 50 years through the use of ecoinformatics and citizen science data. First, you will work collaboratively to identify a butterfly-host plant interaction and learn about its natural history. Then, you will work together to generate species distribution models (SDM) for this butterfly and its primary larval host plant using iNaturalist open source data and common research computing tools such as R, RStudio, and Git (Figure 1). Your group will use these maps to generate a testable hypothesis about the effects of climate change on the distributions and interactions between the butterfly and its larval host. Next, your group will estimate future SDM for the year 2070 using open data provided by citizen scientists and government funded sensors, and use the results to evaluate your hypothesis about the effects of climate change on species' distributions. Finally, your group will present a summary of findings to the class, and each individual will be responsible for addressing essay questions about common themes amongst the presentations. By completing this project, you will be introduced to analyzing biodiversity data with common research computing tools and provided experience in communicating the results and relevance of original scientific research.

Materials and Methods

Overview of Data Collection and Analysis Methods:

<u>Class 1</u>: Introduction to Ecoinformatics, Citizen Science, and Climate Change

Here we introduce the concepts and importance of ecoinformatics, citizen science, and climate change. While many of us know something about climate change and conservation, less often we have heard of ecoinformatics and citizen science efforts. We begin with a discussion of a chapter from Elizabeth Kolbert's book, *Field Notes on a Catastrophe*, to offer a different perspective on the wide-ranging effects of climate change on biodiversity. Then, we move on to talk about the study of biodiversity science in the 21st century and the enormous advances made possible by digitization, computing resources, and the significant data contributions that are being made by citizen science. This discussion leads into introducing ecoinformatics with a specific focus on species distribution modeling (SDM). Finally, we introduce how we can use ecoinformatics combined with citizen science to study the potential effects of climate change on butterfly-host plant interactions.

Prior to class, you will read Silvertown (2009) which introduces citizen science efforts and their impacts on current studies of ecology and evolution. You are then tasked to navigate iNaturalist to learn about the types of data available from this citizen science effort and gain familiarity navigating the website (see Pre-class Assignment 1 in Supplementary Materials). This work helps facilitate the

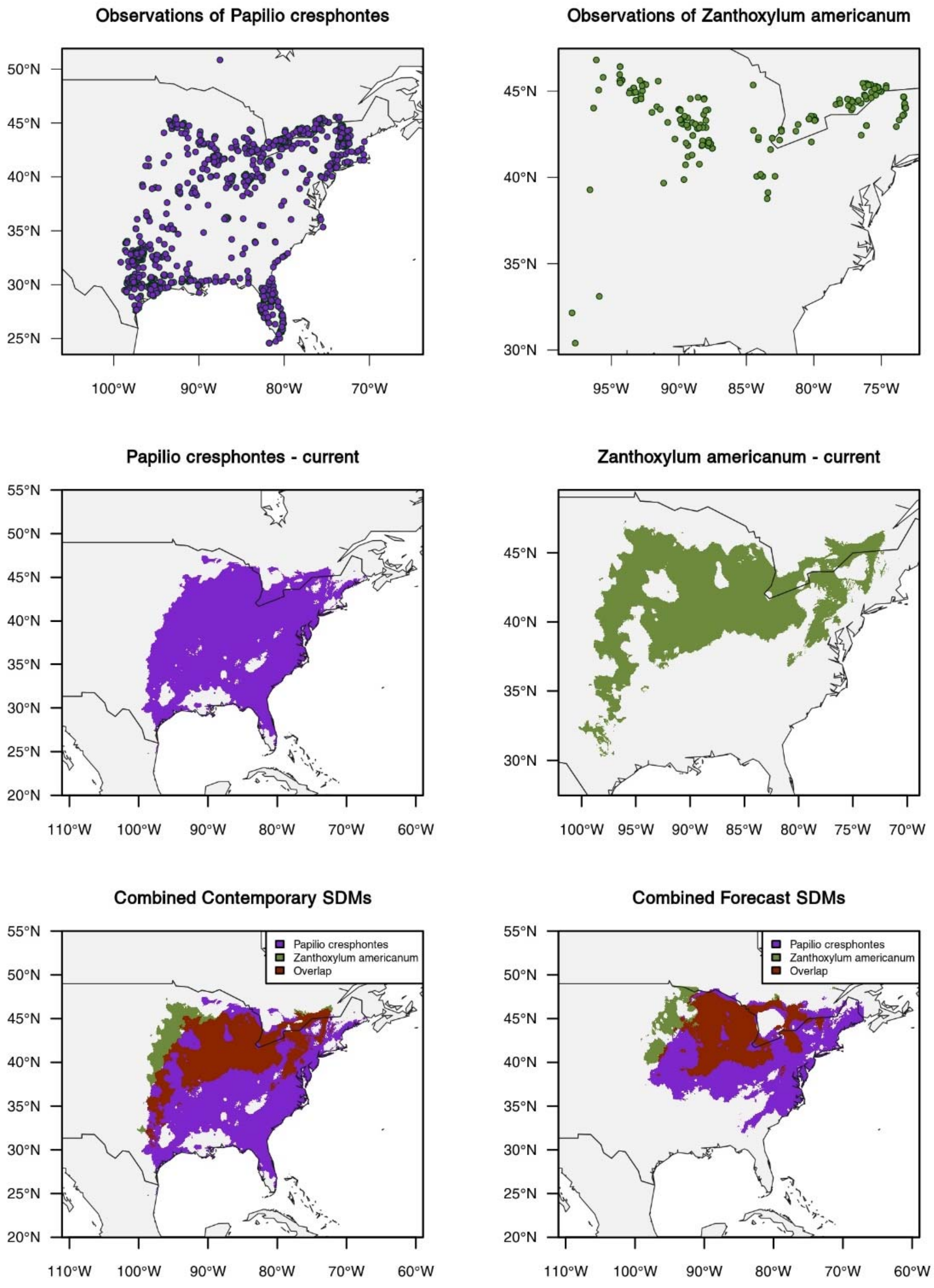

Figure 1. Examples of maps produced by scripts in this project.

first part of the project in which your group selects a butterfly-host plant interaction as the focus of their project.

Butterflies are often associated spatially with their larval host plants. The number of larval host plants and their corresponding distributions vary widely across butterfly species. Some butterflies are specialists feeding on a small number of host plants as caterpillars, others are generalists feeding on a much broader range of host plants. Some host plants have much broader ranges than others encompassing the entire range of the butterfly, while other host plants are geographically limited compared to the butterfly in question. Other ecological interactions not related to the butterfly host plant interaction, such as competitors, physiological limits, or natural enemies, may limit one or both species distributions. When selecting species pairs, be sure to select one where the larval host plant can be identified to the species level. You should verify that this is the most common host plant using BAMONA, eButterfly, or another reputable resource (websites provided in student handout).

Class 1 Learning Goals and Objectives:

1. Describe how biodiversity science data initiatives, such as ecoinformatics, can make use of citizen science and museum digitization efforts to ask and inform questions in ecology.
    a. Identify a butterfly-host plant interaction using online resources including citizen science projects
    b. Report on the natural history of chosen butterfly-host plant interaction

Class 1 Resources:

1. PreClass1-Assignment.docx
2. Class1-Slides.pptx
3. SP1-InstructionsandAssignment-NaturalHistoryofButterfly-HostPlantInteraction.docx
4. SpeciesPairsSuggestions.xlsx

Class 2: Generating Species Distribution Models and Hypothesis Development

We will generate species observation maps and species distribution models (SDMs) for their butterfly-host plant interaction. Prior to class, you should complete SP1 (if you did not do so by the end of Class 1), download the software needed for class, and read Biesmeijer et al. (2006) that examines changing distributions of plant-pollinator interactions in Britain and the Netherlands. Download instructions for software as well as the pre-class assignment are provided in the Class 2 Resources section below. Additionally, a succinct set of slides are provided to help you think about the nature of SDM and ways in which

SDM can be generated. The slides end with a set of tips for success as you engage with research computing resources (e.g., RStudio) potentially for the first time. Further information about the what the analyses are doing is available in SP2-AnalysisScriptInformation.docx.

In class, you can use SP2-Instructions to help navigate iNaturalist to generate an observation map and R to generate species distribution models. We have also included a list of common errors and help documents to assist you especially if you are new to these data analysis programs. For those well-versed in R, we have also provided more simplified instructions (SP2-SimplifiedInstructions).

SP2-Assignment provides a number of questions to help you interpret and discuss the data your group is collecting for this activity. You will be guided in hypothesis development using data from both SP1 and SP2 that you will test in the next class.

Class 2 Learning Goals and Objectives:

2. Use research computing tools (Citizen science crowd sourced data, R programming language, GitHub collaborative web platform, data visualization) to study a butterfly-host plant interaction.
    a. Apply common research computing tools including RStudio and GitHub to visualize butterfly-host plant distributions
    b. Search, filter and download butterfly and host-plant distribution data from biodiversity citizen science web platforms, specifically iNaturalist
    c. Differentiate descriptive and predictive species distribution models
    d. Create maps for species distribution models from open source biodiversity data for hypothesis generation
    e. Develop a hypothesis regarding the effects of climate change on a plant-insect interaction 50 years from now

Class 2 Resources:

1. PreClass2-Assignment.docx
2. Class2-Slides.pptx
3. SP2-Instructions-SpeciesDistributionMapsandHypothesis.docx
4. SP2-SimplifiedInstructions-SpeciesDistributionMapsandHypothesis.docx
5. SP2-Assignment-SpeciesDistributionMapsandHypothesis.docx
6. SP2-Rubric-SpeciesDistributionMapandHypothesis.docx
7. SP2-AnalysisScriptInformation.docx
8. Program Download Instructions (for prior to class)

   a. DownloadInstructions-Git.docx
   b. DownloadInstructions-R.docx
   c. DownloadInstructions-RStudio.docx
9. Computer code and instructions: https://github.com/jcoliver/biodiversity-sdm-lesson
10. HelpDocumentForCommonErrorsAndHelpfulWebsites.docx

Class 3: Forecast Models and Evaluating Hypotheses

Prior to class 3, you should have completed SP2. In some cases, the work may need to be finished outside of class. Referring to the list of common errors (see help document in Class 2 resources) can help you self-diagnose errors that you encounter outside of class. During class, you should use the SP3-Instructions to generate future SDMs which can be used to evaluate your hypothesis. In the SP3-Assignment, you will record your data, be led to make appropriate comparisons, and evaluate your hypothesis. With any remaining class time, your groups can begin to prepare your presentations (see Class 4).

Class 3 Learning Goals and Objectives:

2. Use research computing tools (Citizen science crowd sourced data, R programming language, GitHub collaborative web platform, data visualization) to study a butterfly-host plant interaction.
   a. Generate forecast maps for species distribution models in 2070 from open source biodiversity data and forecast climate models
   b. Evaluate the hypothesis using present and forecast SDMs

Class 3 Resources:

1. SP3-Instructions-FutureSpeciesDistributionModels.docx
2. SP3-Assignment-FutureSpeciesDistributionModels.docx
3. SP3-Rubric-FutureSpeciesDistributionModels.docx

Class 4: Group Project Presentations

Your oral presentations provide the opportunity to disseminate your research findings to an informed audience and to learn about the different ways in which climate change may impact future distributions of plants and butterflies. Thus your learning objectives for this class are about visual and auditory communication, critical thinking, and professional behavior. Instructions for developing the presentation are in SP4-InstructionsandAssignment.

The final written assignment will be a series of essay questions related to all research findings from the class, thus you are encouraged to take good notes on all presentations. The objective of this assignment is to have you synthesize broader trends in changes to plant-insect interactions due to climate change over the next 50 years, and provide you as an individual an opportunity to reflect on their achievements during this project.

Class 4 Learning Goals and Objectives:

3. Communicate findings in the form of an oral presentation
    a. Describe the natural history of a butterfly-host plant interaction
    b. Infer present and future species distributions using SDMs
    c. Evaluate the hypothesis using data generated during the project
    d. Interpret findings in the context of the natural history of butterfly-host plant interaction and climate change
4. Synthesize potential outcomes of the effects of climate change on plant-insect interactions
    a. Formulate hypotheses on the general effects of climate change on plant-insect interactions by comparing and contrasting results from group projects in class

Class 4 Resources:

1. SP4-InstructionsandAssignment-PresentationofProjectandResults.docx
2. SP4-Rubric-PresentationofProjectandResults.docx
3. SP5-InstructionsandAssignment-SynthesisandReflectiononGroupProjects.docx

Questions for Further Thought and Discussion:

1. Identify and describe three ecological questions that have been addressed using an ecoinformatics approach. Provide a citation from the primary literature to support each of your questions.
2. All organisms face a set of abiotic constraints (e.g., temperature, salinity, aridity, etc.) that ultimately affect their distribution. What role do abiotic constraints play in understanding the effect climate change has on changing distributions of organisms? Related to this question, how might migratory organisms (e.g., monarch butterflies) be affected differently than non-migratory organisms in the face of climate change?
3. The species distribution modeling approach applied in this exercise relied solely on using abiotic components of the environment, primarily temperature and precipitation. How might this affect our interpretation of models for species like butterflies and plants, which are also highly dependent on *biotic* factors, such as predators and disease?
4. Citizen science is a growing source of big data for scientists and conservation managers. This type of data collection and verification is

often controlled through a specific web interface usually on a smartphone or tablet. What biases might be present in these data and how can we account for these biases in our data analyses? How might these biases be the same or different than those collected by professionals or through different interfaces such as museum collections?

5. Our studies of a butterfly and host plant generated SDMs for all currently available data in iNaturalist throughout the year. How might you refine this project to take into account stages of the life cycles or seasonal changes of butterflies and their host plants to more accurately study the overlap in distributions of these organisms? Why might it be important to account for life stages in the butterfly or growth patterns in the plant? How might you use data from other citizen science efforts (such as Project BudBurst http://budburst.org/) to supplement your research?
6. In this project, we examined the distributions of one butterfly and the most common host plant of that butterfly when quite often butterflies have many larval host plants (essentially treating all butterflies as specialists). How might the findings of this research be altered when multiple larval host plants are accounted for?
7. Butterflies and host plants are only one example of organisms that interact in the same space and time. Identify two other groups of organisms that depend on each other in some way (predator-prey relationship, mutualism, etc.) and predict how climate change will impact this relationship. Support your argument with citations from the primary literature.
8. The predictions of ranges in the year 2070 rely on a single model of what the Earth's climate will be at that time, but there are over twenty different other models for the climate in 2070. These other models vary in the predicted changes in temperature and precipitation; i.e. some predict a greater rise in temperature, some predict a lower rise in temperature. How could we use these other models to better understand potential changes in species' ranges? Additionally, these future temperature and precipitation data are all based on simulations of how the climate is *expected* to change. How might this affect our interpretations of the anticipated changes in species ranges?

References

Barbet-Massin, M., F. Jiguiet, C.H. Albert, W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3:327-338.

Bale, J.S., G.J. Masters, I.D. Hodkinson, C. Awmack, T.M. Bezemer, V.K. Brown, J. Butterfield, A. Buse, J.C. Coulson, J. Farrar, J.E.G. Good, R. Harrington, S. Hartley, T.H. Jones, R.L. Londroth, M.C. Press, I. Symrioudis, A.D. Watt, and J.B. Whittaker. 2002. Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. Global Change Biology 8:1-16.

Biesmeijer J.C., S.P.M. Roberts, M. Reemer, R. Ohlemüller, M. Edwards, T. Peeters, A.P. Schaffers, S.G. Potts, R. Kleukers, C.D. Thomas, J. Settele, and W.E. Kunin. 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science 313:351-354.

Cech, R. and G. Tudor. 2005. *Butterflies of the East Coast: An Observer's Guide*. 1st edition. Princeton University Press, Princeton, NJ.

Chung, U., L. Mack, J.I. Yun, and S-H. Kim. 2011. Predicting the timing of cherry blossoms in Washington, DC and Mid-Atlantic states in response to climate change. PLoS ONE 6:e27439.

Cook, J.A., S.V. Edwards, E.A. Lacey, R.P. Guralnick, P.S. Soltis, D.E. Soltis, C.K. Welch, K.C. Bell, K.E. Galbreath, C. Himes, J.M. Allen, T. A. Heath, A.C. Carnaval, K.L. Cooper, M. Liu, J. Hanken, and S. Ickert-Bond. 2014. Natural History Collections as Emerging Resources for Innovative Education. Bioscience 64:725-734.

Dickinson, J.L., B. Zuckerberg, and D.N. Bonter. 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. Annual Review of Ecology, Evolution and Systematics 41:149-172.

Dickinson, J.L., J. Shirk, D. Bonter, R. Bonney, R.L. Rain, J. Martin, T. Phillips, and K. Purcell. 2012. The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment 10:291-297.

Elith, J. and J.R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677-697.

Hochachka, W.M., D. Fink, R.A. Hutchinson, D. Sheldon, W-K. Wong, and S. Kelling. 2012. Data-intensive science applied to broad-scale citizen science. Trends in Ecology and Evolution 27:130-137.

Hurlbert, A.H. and Z. Liang. 2012. Spatiotemporal variation in avian migration phenology: citizen science reveals effects of climate change. PLoS One 7:E31662.

Kelly, A.E. and M.L. Goulden. 2008. Rapid shifts in plant distribution with recent climate change. Proceedings of the National Academy of Sciences of the United States of America 105:11823-11826.

Kerr, J.T., Pindar, A., Galpern, P., Packer, L., Roberts, S.M., Rasmont, P., Schweiger, O., Colla, S.R., Richardson, L.L., Wagner, D.L., Gall, L.F., Sikes, D.S., Pantoja, A. 2015. Climate change impacts on bumblebees converge across continents. Science 349:177-180.

Kolbert, E. 2006. Ch. 4: The butterfly and the toad. Pages 67-87 *in Field Notes from a Catastrophe: Man, Nature, and Climate Change.* Bloomsbury Publishing USA.

Lacey, E.A., T.T. Hammond, R.E. Walsh, K.C. Bell, S.V. Edwards, E.R. Ellwood, R. Guralnick, S.M. Ickert-Bond, A.R. Mast, J.E. McCormack, A.K. Monfils, P.S. Soltis, D.E. Soltis, and J.A. Cook. 2017. Climate change, collections and the classroom: using big data to tackle big problems. Evolution Education Outreach 10:2.

Lepczyk, C.A. 2005. Integrating published data and citizen science to describe bird diversity across a landscape. Journal of Applied Ecology 42:672-677.

McSweeney, C.F., R.G. Jones, R.W. Lee, and D.P. Rowell. 2015. Selecting CMIP5 GCMs for downscaling over multiple regions. Climate Dynamics 44:3237-3260.

Matthew, J., Lorrae, E., Carina, A., Georgina, M., Wendy, B., Colin, I. 2017. Transforming conservation science and practice for a postnormal world. Conservation Biology 31:1008-1017.

Monfils, A.K., K.E. Powers, C.J. Marshall, C.T. Martine, J.F. Smith, and L.A. Prather. 2017. Natural history collections: teaching about biodiversity across time, space, and digital platforms. Southeastern Naturalist 16:47-57.

Moritz, C., J.L. Patton, C.J. Conroy, J.L. Parra, G.C. White, and S.R. Beissinger. 2008. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. Science 32:261-264.

Pearson, R.G. 2010. Species' distribution modeling for conservation educators and practitioners. Lessons in Conservation 3:54-89.

Rowe, K.C., K.M.C. Rowe, M.W. Tingley, M.S. Koo, J.L. Patton, C.J. Conroy, J.D. Perrine, S.R. Beissinger, and C. Moritz. 2015. Spatially heterogeneous impact of climate change on small mammals of montane California. Proceedings of the Royal Society B, Biological Sciences 282:20141857.

Rubidge, E.M., W.B. Monahan, J.L. Parra, S.E. Cameron, and J.S. Brashares. 2011. The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. Global Change Biology 17:696-708.

Scott, J.A. 1992. *Butterflies of North America: A Natural History and Field Guide*. Stanford University Press, Stanford, California, USA.

Silvertown, J. 2009. A new dawn for citizen science. Trends in Ecology and Evolution 24:467-471.

Soltis, D.E. and P.S. Soltis. 2016. Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. Plant Diversity 38:264-270.

Tingley, M.W., W.B. Monahan, S.R. Beissinger, and C. Moritz. 2009. Birds track their Grinnellian niche through a century of climate change. Proceedings of the National Academy of Sciences of the United States of America 106:19637-19643.

Todd, B.D., J.P. Rose, S.J. Price, and M.E. Dorcas. 2016. Using citizen science data to identify the sensitivity of species to human land use. Conservation Biology 30:1266-1276.

Useful Websites:

iNaturalist
    www.inaturalist.org/taxa

BAMONA: Butterflies and Moths of North America
    www.butterfliesandmoths.org

eButterfly
    www.e-butterfly.org

The R Project for Statistical Computing
    https://www.r-project.org/

RStudio
https://www.rstudio.com/
Example of species distribution modeling in R
http://www.molecularecologist.com/2013/04/species-distribution-models-in-r/
Git
https://git-scm.com/
GitHub
https://github.com/
Open Science Framework (OSF)
https://osf.io/
eBird
https://ebird.org/home
Project Budburst
http://budburst.org/
Mushroom Observer
http://mushroomobserver.org/

Tools for Assessment of Student Learning Outcomes:
The total grade for this group project is based on five components; four components are evaluated per group and one component is evaluated per individual. Each Student Product (SP) contains a series of questions that are graded. Examples for each Student Products 1-3 have been made available in supplementary documents. Rubrics are provided for SP2-4. SP1 is worth 5 points and is graded based on completion and the overall quality of responses. For SP5, each question is worth 10 points (20 pts total) and essays are graded based on quality of the essay responses and each student's ability to demonstrate a deep understanding of the learning goals of the project.

NOTES TO FACULTY

Challenges to Anticipate and Solve

1. **Operating system idiosyncrasies:** With most exercises that rely heavily on computation, there are opportunities for differences between operating systems to present challenges due to differences in file management and software compatibilities. In this lesson, installing Git and connecting it with the RStudio interface presented the most immediate challenge, as most

Macintosh OS X systems do not show hidden files by default. This made navigating to Git resources challenging for some students. Solutions for these challenges include (a) having resource experts accessible, especially at the start of the lesson, to help troubleshoot these problems; (b) having resources (print or digital) on hand for self-directed troubleshooting; and (c) minimizing operating system heterogeneity by conducting lesson in a computer lab. The latter solution has the drawback that it may limit students' access to computing resources (based on availability of computer lab resources) and miss the valuable exercise of installing software on their own computers.

2. **The fear of coding:** Students may be wary of typing commands into a computer, for fear that they may damage the computer. Additionally, scripting programs like R and python have syntax rules (e.g. case sensitivity) that may seem overly strict to novices. Overcoming this challenge largely requires allowing time for students to get comfortable with the process of typing commands, making mistakes, and editing commands to run correctly. Developing a supportive environment, where students are not rushed to finish in a short amount of time, is critical to overcome students' aversion to coding. Giving students access to the suggestions in the provided HelpDocumentforCommonErrorsandHelpfulWebsites.docx may also be useful to empower students to begin to troubleshoot errors they encounter.
3. **Expertise on-hand to explain relevance and troubleshoot errors:** The ecological concepts covered in this lesson, including species distribution modeling and climate forecasting, require domain-specific knowledge to be able to establish value of the exercise. Having such an expert, even if they are not present for the entire exercise, is important to make the process relevant. Also important is to have someone available who has a level of familiarity with the computational tools of this lesson (Git and R) to help troubleshoot errors as they arise. It is important to note that this expertise need not be local - when we ran this lesson, the instruction occurred in New Jersey, while domain and computational expertise was afforded, in real time via instant messaging and video chats, from Arizona.

4. **Pre-class testing:** Echoing point 1 above, there are often unanticipated challenges when dealing with software and data products from third parties. We cannot recommend strongly enough that instructors attempt all analyses before the class in which they should be run. Also useful would be tests downloading data from iNaturalist before the class session. Although we encountered no extremely long download times, this can vary based on local internet speeds and current traffic on the iNaturalist web server.

Comments on Introducing the Experiment to Your Students:

As this is an interdisciplinary project drawing from ecology, biodiversity science, data science, and the rapidly growing field of citizen science, we have worked to ensure our introduction helps students (1) understand the major objectives of each discipline and (2) see the power of connecting these disciplines to address questions in plant-insect interactions from an academic and applied perspective. Engaging students as citizen scientists has been an excellent way to bring an active approach to many traditional 'ology courses' classrooms (e.g., mammalogy, herpetology, botany, etc). Here we move beyond engaging students as citizen scientists and help them develop questions that can be addressed by these data. As such, this helps students better understand the nature of 'big data' that defines their generation and the importance of gaining skills in data science to be able to navigate and analyze these data to answer fundamental questions in biology.

We implemented this project in an upper level biology course, where a major goal of the course is for students to research a novel question in plant-insect interactions. For many students who do not immediately see the relevance of plant-insect interactions, this group project helps them understand the potential impacts of climate change on organisms that live nearby (e.g., eastern North America) and begin to impress upon them the changes global warming is likely to bring to our planet.

Before beginning any of the work required by the project itself, the major themes of the project (ecoinformatics, biodiversity science, citizen science, data science, and climate change) are introduced through an interactive lecture during Class 1. Students are required to complete a pre-class assignment prior to Class 1 that provides a foundation for discussion climate change and biodiversity science, introducing citizen science efforts, and engaging in the first part of the project - identifying a butterfly-host plant and learning about its natural history. We provide a suggested host plant-butterfly pair list to recommend to the students (SpeciesPairsSuggestions.xlsx). For our class on plant-insect interactions, we had yet to discuss many of these topics so we found it helpful to dedicate an entire class to the theoretical background of ecology and biodiversity studies, but also provide time for students to engage with the first part of the project (SP1). Toward the end of class, we provide an overview of the various parts of the project (Table 1), to inform students of the guidance they will receive as they begin their work on this project as most students had never engaged with research computing resources such as R.

Table 1. Overview of Student Products along with suggested point values that are assessed for this module.

Assignment	Description	Points
SP1	Natural history of butterfly-host plant	5
SP2	Species distribution maps (SDMs) and hypothesis	25
SP3	Future species distribution models and hypothesis evaluation	20
SP4	Presentation of project and results	50
SP5	Synthesis and reflection of group projects	20
Group work assessment	Survey assessment by peers	5
	Total	*125*

Comments on the Data Collection and Analysis Methods Used in the Experiment:

Students should be informed about the origin of the data they download from Citizen Science webportals. Citizen scientists collected and verified these data based on the protocol established by the biodiversity web platform, iNaturalist (https://www.inaturalist.org/). The data can be uploaded via smartphone, tablet, laptop, or desktop. If participants are using a smartphone or tablet to collect data and they are out of range when they are collecting data, the data will be transferred to iNaturalist automatically once they are back in range. The data are constantly changing with the addition of new observations and the verification of older ones. For example, one student group downloaded data two days after their first download and noticed that more observations appeared during that 48-hour period.

Given the public nature of the Citizen Science data, research grade observations can be downloaded directly from the iNaturalist website as either .csv or .kml formats. Web interfaces can change over time so the data downloading may change through time. Data in iNaturalist is designated as research grade when it has a photo, is georeferenced, and has been confirmed by two other iNaturalist users. Research grade observations are then shared with the Global Biodiversity Information Facility (GBIF https://www.gbif.org/).

Comments on the Data Analyses:

In an attempt to consolidate analytical tools with documentation for their use, the GitHub repository includes worked examples, full instructions, troubleshooting common errors, and links to additional resources. See https://github.com/jcoliver/biodiversity-sdm-lesson.

Comments on Questions for Further Thought:

1. Identify and describe three ecological questions that have been addressed using an ecoinformatics approach. Provide a citation from the primary literature to support each of your questions.

For many students, ecoinformatics and research computing might be quite new. To help students construct their own understanding of the driving questions that underlie ecoinformatics and better understand the power of research computing tools available to assist researchers, students might be best served by finding examples from the primary literature. Having each student in the course search the literature independently offers the opportunity to collect a number of different articles covering a wider array of active areas of investigation in ecoinformatics and problems addressed via research computing tools. Using course management software to facilitate sharing these articles and summaries (e.g., through a discussion board), is one way to have students read and learn about research found by their peers.

2. All organisms face a set of abiotic constraints (e.g., temperature, salinity, aridity, etc.) that ultimately affect their distribution. What role do abiotic constraints play in understanding the effect climate change has on changing *distributions of organisms? Related to this question, how might migratory organisms (e.g., monarch butterflies) be affected differently than non-migratory organisms in the face of climate change?*

As students in our plant-insect interactions course worked on this project, we wanted to ensure there were ample opportunities to connect to the project to relevant course material. By doing so, students could see the relevance of what may otherwise appear to be arbitrary details of the natural history of plants and insects. For instance, prior to the start of the project students learned about the abiotic needs of insects - from water availability to temperature effects on an insect's life cycle. In a changing climate, the limits imposed by these abiotic factors may be met in the environments of the plants and insects. Further, to be able to discuss and interpret the forecast maps generated during this project, students must consider the current abiotic constraints of their chosen organisms. By having students discuss some of these general types of traits beforehand, such as how migration patterns may change with global warming, they will be better prepared to consider the connection between the abiotic constraints of their organisms and the change in distribution they detect over the next 50 years.

3. The species distribution modeling approach applied in this exercise relied solely on using abiotic components of the environment, primarily temperature and precipitation. How might this affect our interpretation of models for species like butterflies and plants, which are also highly dependent on biotic factors, such as predators and disease?

An important part of understanding species' distribution is consideration of both abiotic and biotic components of the environment that influence the suitability of an ecosystem for a particular species. Some species are significantly affected by certain biotic components of their environment, but the climate variables used in these species distribution models only account for abiotic variation in environments. Thus if a species, such as a butterfly, is reliant on certain biotic factors, such as the presence of a larval host plant or predator, species distribution models based on abiotic components alone may overestimate the actual species' distribution, or the generated map is larger than it should be.

4. Citizen science is a growing source of big data for scientists and conservation managers. This type of data collection and verification is often controlled through a specific web interface usually on a smartphone or tablet. What biases might be present in these data and how can we account for these biases in our data *analyses? How might these biases be the same or different than those collected by professionals or through different interfaces such as museum collections?*

Discussion of these questions can help elucidate misconceptions students have regarding citizen science and simultaneously push them to think broadly about biases present in data. Students may have the impression that citizen science data cannot be used for scientific investigations because the data were collected by non-scientists. For many citizen science projects, there are mechanisms to ensure data quality - for instance on iNaturalist there are different filters that can be placed on data from 'research grade' to 'unverified.' This also allows discussion of data curation as part of the citizen science data collection pipeline. Ultimately, no data are perfect and have biases. In citizen science data, common biases include over-representation of observations from densely populated areas such as cities, more observations in warmer weather, and fewer records of 'less-appealing' organisms such as insects. The class can engage in a discussion of how to account for these biases in the assumptions of the study or to supplement the data where necessary with other types of museum records. From citizen science efforts to published genomes, our students are being trained in an era of big data. As larger data sources become available and are incorporated into scientific investigations, we must train our students to think carefully about the quality and biases in these massive data sets.

5. Our studies of a butterfly and host plant generated SDMs for all currently available data in iNaturalist throughout the year. How might you refine this project to take into account stages of the life cycles or seasonal changes of butterflies and their host plants to more accurately study the overlap in distributions of these organisms? Why might it be important to account for life stages in the butterfly or growth patterns in the plant? How might you use data from other citizen science efforts (such as Project BudBurst http://budburst.org/) to supplement your research?

This question, similar to question 2, allows connection to course content. In our plant-insect course, we discuss the stages of insects' life cycles, different nutritional requirements, etc. As written, this project restricts students to thinking about the larval stage of butterfly development and how it relates to plants, but they may not have restricted the data they collected to larval life stages. Filtering the data to contain only particular stages of the life cycles would refine the SDMs and allow one to move beyond whether the organism will be impacted by climate change, but rather if there is a particular stage in the life cycle that makes that organism that much more susceptible to a changing climate. For the butterflies, images provided by users of iNaturalist would potentially allow students to score their data by life stage, but this may be more difficult from the perspective of plants. However, other citizen science efforts, such as Project BudBurst, could provide more information about when plants are flowering and leafing out to be suitable hosts for caterpillars. This would acquaint students with additional citizen science efforts.

6. In this project, we examined the distributions of one butterfly and the most common host plant of that butterfly when quite often butterflies have many larval host plants (essentially treating all butterflies as specialists). How might the findings of this research be altered when multiple larval host plants are accounted for?

This is one example of a life history trait (see Student Product 1) that students describe after identifying a butterfly-host plant to study. The forecast SDMs may show little to no overlap between butterfly and host plant leaving students to explain how these butterflies can persist in areas with no host plants. In some cases, where the butterflies are restricted to one or a few larval host plants, this leaves few options - essentially, the butterfly species can switch host plants or go extinct. For butterflies with multiple host plants, a future host-shift is more likely and can have profound impacts on the ecology of the new host plant species. Students could supplement their projects by looking at the distribution of additional host plants of the butterfly.

7. Butterflies and host plants are only one example of organisms that interact in the same space and time. Identify two other groups of organisms that depend on each other in some way (predator-prey relationship, mutualism, etc.) and predict how climate change will impact this relationship. Support your argument with citations from the primary literature.

To help students connect what they have learned in class with other biological systems and prevent them from confining this learning experience to plant-insect interactions, this question can facilitate a wide-ranging discussion of the impacts of climate change on symbioses. Alternatively, one could restrict the question to specific kinds of interactions. For instance, in our plant-insect interaction course we might ask if climate change will have a greater impact on plant-insect mutualisms or antagonisms. Although there is unlikely to be a clear answer to this particular question, it forces the students to think about both biotic and abiotic components of these interactions and how that might be affected by climate change.

8. The predictions of ranges in the year 2070 rely on a single model of what the Earth's climate will be at that time, but there are over twenty different other models for the climate in 2070. These other models vary in the predicted changes in temperature and precipitation; i.e. some predict a greater rise in temperature, some predict a lower rise in temperature. How could we use these other models to better understand potential changes in species' ranges? Additionally, these future temperature and precipitation data are all based on simulations of how the climate is expected to change. How might this affect our interpretations of the anticipated changes in species ranges?

Forecasting climate change is an area of active development, with considerable uncertainty around just how much change will really occur. Given the variation in what might occur, relying on a single model for forecasting species distributions may ultimately be inaccurate. For more accurate forecasting, separate SDMs based a variety of different scenarios could be run to generate a "consensus" distribution. This should also be an opportunity to discuss the caution to use when drawing conclusions based on simulated data, which the climate data for 2070 are.

Comments on the Assessment of Student Learning Outcomes:

Prior iterations of this module asked students to compare the distribution of present-day butterfly-host plant ranges to 50-100 years in the past. Students downloaded data from GBIF and did not use tools such as R. In its current form, the project provides the opportunity for hands-on experience using data-science tools, while engaging students in an inquiry project for which the outcome is unknown. With these changes, students seemed more motivated to consider how biological interactions will change in the future as compared to the past. All students struggled with R and many were initially intimidated by it. However, students are aware that these data science tools are becoming increasingly important to scientists, and they felt a great sense of accomplishment when they successfully ran their code. Students exhibited much ownership over their projects. While participating in conversation with students when groups were discussing the implications of their findings, we observed how much more the students were engaging with the material as compared to a traditional lecture.

Finally, this project allows students the opportunity to become familiar with and engage in citizen science efforts in a different way - some students had engaged with citizen science initiatives by collecting and contributing data, but none had used the data to address their own question before.

A student's work on this project is assessed through the five student products (SP; refer to section 2). The first four assignments (SP1-4) are completed and graded by group (note: there is an individual component in SP4 - see provided rubric for details) and the last assignment (SP5) is graded per individual. Each of these assignments keeps students moving through the material and allows the opportunity for the instructor to provide feedback and assessment on student achievement of the learning goals and objectives of the project (see below). Additionally, the SPs offer opportunities for students to respond to feedback and improve their work. For instance, students first develop a hypothesis in SP2 and are asked to restate their hypothesis in SP3 including any modifications they may have made based on feedback from SP2.

The rubrics provided (refer to **Tools for Assessment of Student Learning Outcomes** for rubrics) for the Student Products reflect the various components of the assignments necessary for students to achieve the Learning Goals and Objectives of the assignment. However, some other aspects of the rubrics, such as writing a figure legend, tie into broader learning goals of this specific course, such as "written communication of science." These rubrics can and should be modified to suit the emphasis of the course in which the module is being implemented. For reference, here are the major Learning Goals and Objectives assessed by this group project:

1. Describe how biodiversity science data initiatives, such as ecoinformatics, can make use of citizen science and museum digitization efforts to ask and inform questions in ecology.
    a. Identify a butterfly-host plant interaction using online resources including citizen science projects
    b. Report on the natural history of a butterfly-host plant interaction
2. Use research computing tools (Citizen science crowd sourced data, R programming language, GitHub collaborative web platform, data visualization) to study a butterfly-host plant interaction.
    a. Apply common research computing tools including RStudio and GitHub to visualize butterfly-host plant distributions
    b. Search, filter and download butterfly and host-plant distribution data from biodiversity citizen science web platforms, specifically iNaturalist
    c. Differentiate descriptive and predictive species distribution models

  d. Create maps for species distribution models from open source biodiversity data for hypothesis generation
  e. Develop a hypothesis regarding the effects of climate change on a plant-insect interaction 50 years from now
  f. Generate forecast maps for species distribution models in 2070 from open source biodiversity data and forecast climate models
  g. Evaluate the hypothesis using present and forecast SDMs
3. Communicate findings in the form of an oral presentation
   a. Describe the natural history of a butterfly-host plant interaction
   b. Infer present and future species distributions using SDMs
   c. Evaluate the hypothesis using data generated during the project
   d. Interpret findings in the context of the natural history of a butterfly-host plant interaction and climate change
4. Synthesize potential outcomes of the effects of climate change on plant-insect interactions
   a. Formulate hypotheses on the general effects of climate change on plant-insect interactions by comparing and contrasting results from group projects in class

Comments on Formative Evaluation of this Experiment:

This experiment contains considerable hands-on work, where students interface with web-based citizen science portals and computer programming tools. Given these exposures to novel resources, there are several opportunities for evaluating progress. In-class work includes the development of information literacy skills, when students are tasked with finding information about butterfly species and respective host plant species, as well as how to extract data from citizen science programs. When students get to the point of making changes to R code and attempting to run analyses, there will inevitably be problems. Most issues will arise due to typos; seeing error messages during class time provides an assessment of students' understanding of the importance of relevant syntax such as case sensitivity and quoting strings in R. Such in-class assessments rely on instructors' attentiveness and availability during exercises; examples and troubleshooting for common mistakes are provided in supplementary materials and on the GitHub repository. In addition to the hands-on work, the **Questions for Further Thought and Discussion** section provides additional material for formative assessment of students' learning progress.

Comments on Translating the Activity to Other Institutional Scales or Locations:

1. This module could be scaled up to a larger class size with the right amount of instructor expertise, confidence, and assistance. Of particular concern is the necessary human capital to identify and troubleshoot software problems during instruction. As outlined in the challenges section, managing a large course will depend on the number of knowledgeable assistants available to the instructor.
2. This module could also be implemented during a laboratory or recitation period connected to large enrollment courses. Typically, laboratory and recitation sections break the class into smaller groups where implementation of this module would be more feasible.
3. This module is derived from open data accessible from a computer with an internet connection. This experiment could be modified to different study species or interactions (e.g. predator-prey) with minimal modifications in the code.
4. This module can easily accommodate students with mobility challenges. However, it will require support from the campus disability center to adapt this experiment for other disabilities.

STUDENT COLLECTED DATA FROM THIS EXPERIMENT

Examples of Student Products 1-3 have been provided in supplemental materials to help instructors see some of the various types of responses that should be expected from students. These files include:

1. SP1-Example-NaturalHistoryofButterfly-HostPlantInteraction
2. SP2-Example-SpeciesDistributionMapandHypothesis
3. SP3-Example-FutureSpeciesDistributionModels

COPYRIGHT STATEMENT

Lastly, we request that you return your students' and your comments on this activity to the TIEE Managing Editor (tieesubmissions@esa.org) for posting at this site.

GENERIC DISCLAIMER

Adult supervision is recommended when performing this lab activity. We also recommend that common sense and proper safety precautions be followed by all participants. No responsibility is implied or taken by the contributing author, the editors of this Volume, nor anyone associated with maintaining the TIEE web site, nor by their academic employers, nor by the Ecological Society of America for anyone who sustains injuries as a result of using the materials or ideas, or performing the procedures put forth at the TIEE web site, or in any printed materials that derive therefrom.