Welcome to the Webinar

Math, Computing, Undergraduate Ecology Education and Large Datasets:

an Example from a Citizen Science Program

We will begin at 1:00PM EDT

NIMBioS
National Institute for Mathematical and Biological Synthesis

Ecological Society of America
esa

http://nimbios.org/press/ecoedwebinar_resources.html


Mathematical
Ecology

Math, Computing, Undergraduate Ecology Education and Large Datasets:

an Example from a Citizen Science Program

Will Godsoe and Louis J. Gross
National Institute for Mathematical and Biological Synthesis
Departments of Ecology and Evolutionary Biology and Mathematics
University of Tennessee, Knoxville

Supported by NSF Award EF-0832858 to the University of Tennessee and co-sponsored by the Ecological Society of America

NIMBioS
National Institute for Mathematical and Biological Synthesis

NIMBioS
National Institute for Mathematical and Biological Synthesis

Vision

Foster new collaborative efforts to investigate fundamental and applied questions arising in biology using appropriate mathematical and computational methods

Enhance the essential human capacity to analyze complex biological questions and develop necessary new mathematics

Encourage broader public appreciation of the unity of science and mathematics.

NSF

USDA

U.S. Department of Homeland Security

Specific Methods

- Focused research projects (**Working Groups**) to build collaboration among diverse communities.
- Building collaborations through more open-ended general problems, addressed through multiple approaches (**Investigative Workshops**).
- Skill and methods-based programs (**Tutorials**) that foster a broader understanding of potential applications of modern math and computational science in biology.
- An expansive set of **education-linked-to-research** endeavors from elementary through **post-doctoral** level that provide diverse opportunities at the math/biology interface.

NIMBioS is "community-driven" - requested activities are evaluated by an external advisory board, not chosen by the directors

Overview

- Science, models and math
- Some recent reports on biology education
- Our REU program linking biology and math undergrads
- Data derived from the ATBI citizen science project in Great Smoky Mountains National Park
- How undergrads dealt with deciding what to analyze and how to do so from a component of the ATBI dataset

NIMBioS
National Institute for Mathematical and Biological Synthesis

What is science?

Science is thought to be a process of pure reductionism, taking the meaning out of mystery, explaining everything away, concentrating all our attention on measuring things and counting them up. It is not like this at all. The scientific method is guesswork, the making up of stories. The difference between this and other imaginative works of the human mind is that science is then obliged to find out whether the guesses are correct, the stories true. Curiosity drives the enterprise, and the open acknowledgement of ignorance.

Lewis Thomas - Sierra Club Bulletin, March/April 1982, P. 52

What is science?

Science is thought to be a process of pure reductionism, taking the meaning out of mystery, explaining everything away, concentrating all our attention on measuring things and counting them up. It is not like this at all. The scientific method is guesswork, the making up of stories. The differ between this and other imaginative works of the human mind is that science is then obliged to find out whether the guesses are correct, the stories true. Curiosity drives the enterprise, and the open acknowledgement of ignorance.

Lewis Thomas - Sierra Club Bulletin, March/April 1982, P. 52

Models

NIMBioS
National Institute for Mathematical and Biological Synthesis

What is science?

Science is thought to be a process of pure reductionism, taking the meaning out of mystery, explaining everything away, concentrating all our attention on measuring things and counting them up. It is not like this at all. The scientific method is guesswork, the making up of stories. The difference between this and other imaginative works of the human mind is that science is then obliged to find out whether the guesses are correct, the stories Curiosity drives the enterprise, and the open acknowledgement of ignorance.

Data

Lewis Thomas - Sierra Club Bulletin, March/April 1982, P. 52

NIMBioS
National Institute for Mathematical and Biological Synthesis

Key Point

Student comprehension of math and appreciation for its importance in everyday experience would be enhanced at every level of the educational experience if we encourage connections between the math and the models (internal to their experience, as well as those used to understand scientific processes) students use, and the data they collect from their own observations of the world around them.

Tie a math and computational world view to everyday experience.

NIMBioS
National Institute for Mathematical and Biological Synthesis

Training Biologists: Quantitative Concepts for all our Students

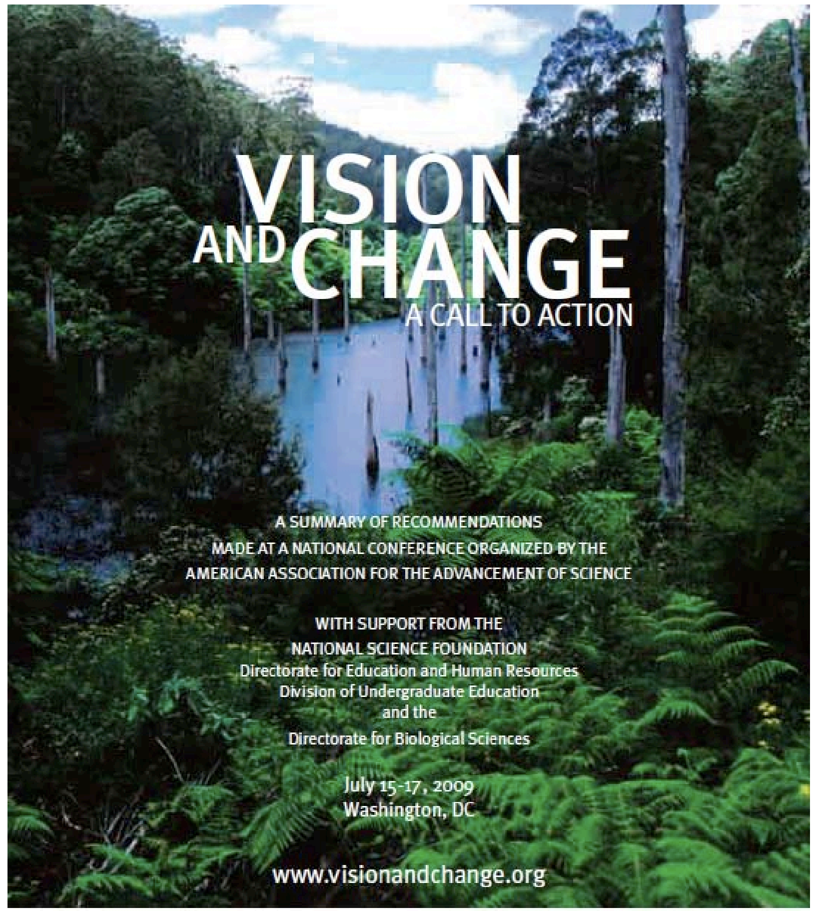
1. Rate of change
2. ***Modeling***
3. Equilibria and stability
4. Structure
5. Interactions
6. ***Data and measurement***
7. Stochasticity
8. ***Visualizing***
9. ***Algorithms***

See Bio2010 report for more details

NIMBioS
National Institute for Mathematical and Biological Synthesis

Vision & Change in Undergraduate Biology Education

A VIEW FOR THE 21st CENTURY



NIMBioS National Institute for Mathematical and Biological Synthesis

www.visionandchange.org

Overarching and unifying key competencies

- Process and nature of science
- Interdisciplinarity
- Communication and Collaboration
- Quantitative competency
- Science and Society

All the above competencies are reinforced through the research experience program we will discuss.

NIMBioS National Institute for Mathematical and Biological Synthesis

Entering Medical Student Expectations

Premedical students, no matter what their primary fields of study, should learn the major concepts and skills of science and mathematics, leaving to medical schools the task of building on this scientific foundation the further scientific competencies that provide them the ability to practice science-based medicine.

Overarching Competency at the Time of Entry into Medical School:

Demonstrate both knowledge of and ability to use basic principles of mathematics and statistics, physics, chemistry, biochemistry, and biology needed for the application of the sciences to human health and disease; demonstrate observational and analytical skills and the ability to apply those skills and principles to biological situations.

Competency E1
Apply quantitative reasoning and appropriate mathematics to describe or explain phenomena in the natural world.

Competency E8
Demonstrate an understanding of how the organizing principle of evolution by natural selection explains the diversity of life on earth.


Mathematical
Ecology

Specific lessons for undergraduates from Large datasets

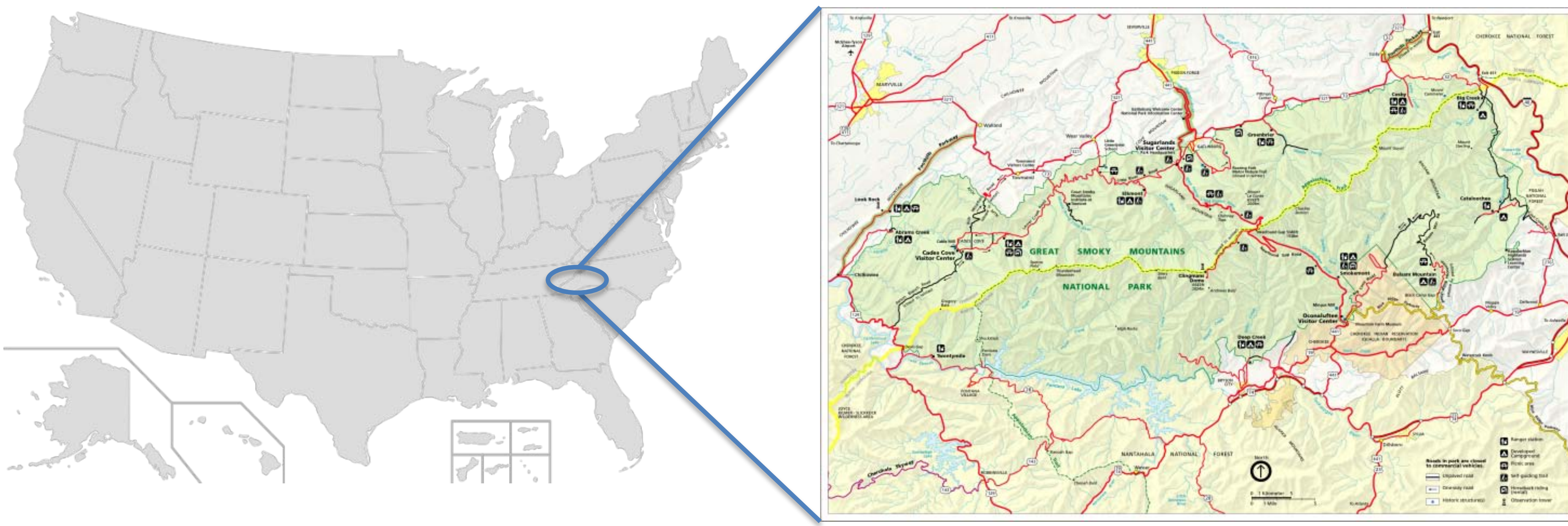
- Connecting ecological concepts to field data is complex. (Project overview)
- Math makes biological ambiguities more concrete.
- Even simple mathematical models can strengthen hypothesis tests.

Research Experience Undergraduate (REU) program at NIMBioS

- 8 week project
- 13 undergrads, 2 high school teachers, 3 vet students
- Large Datasets project: 4 students, myself and another advisor- Prof. Paul Armsworth


William and Mary
NIMBioS
NIMBioS

Great Smoky Mountains National Park



Biodiversity hotspot

All Taxa Biodiversity Inventory (ATBI)

- Used a variety of insect traps

Dataset includes:

- 23 Orders
- 226 Families
- 1,736 Species
- 189,562 Individuals

Pilot study sites

Research problem: summarize a large set of observations into ecological insights

Students did (limited) field biology

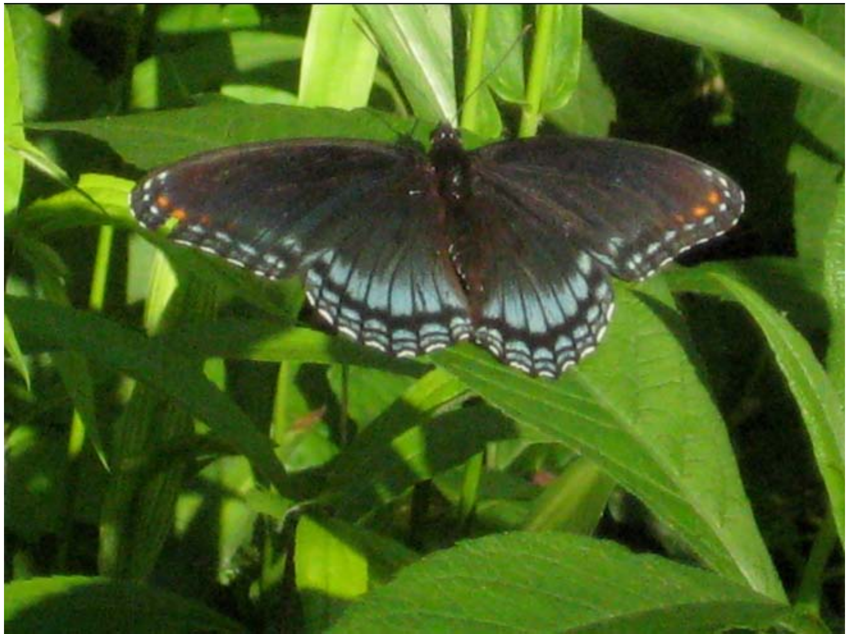
Insect Identification

Find field sites

Find insect taxa under study

Pre-formatted data: Brush footed butterflies (*Nymphalidae*)

	Andrews Bald	Snakeden Ridge	Twin Creeks
Polygonia_comma	0	0	4
Speyeria_diana	0	0	0
Speyeria_cybele	1	0	0
Phyciodes_tharos	1	0	0
Vanessa_virginiensis	4	0	0
Vanessa_atalanta	1	0	0



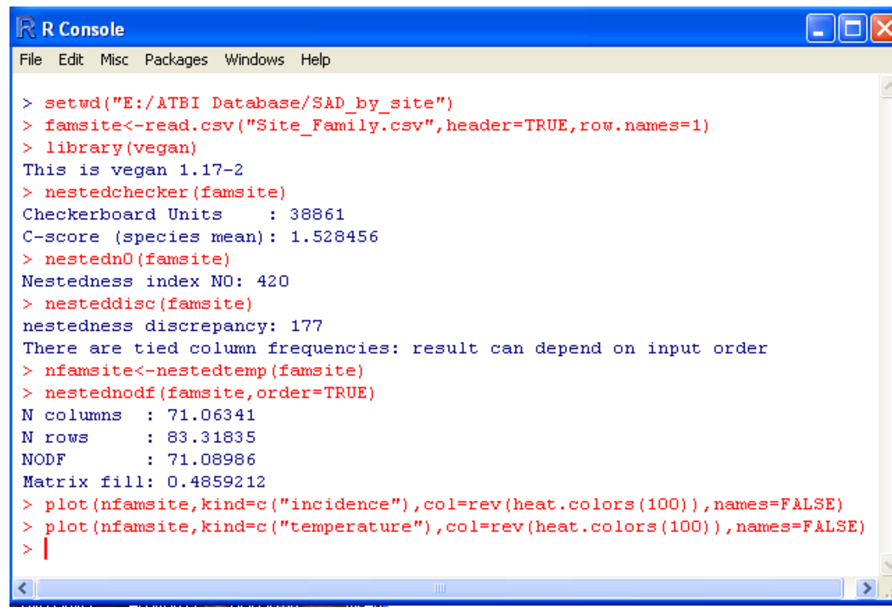
Exploration in R:

Two hour lecture, two hour walk-through using the vegan package for biodiversity analysis

Multivariate Analysis of Ecological Communities in R: vegan tutorial

Jari Oksanen

March 11, 2010



R tutorials:

http://cc.oulu.fi/~jarioksa/softhelp/vegan.html
http://www.nimbios.org/products/RforBiologistv1.1.pdf

Complexities arise naturally out of analyses

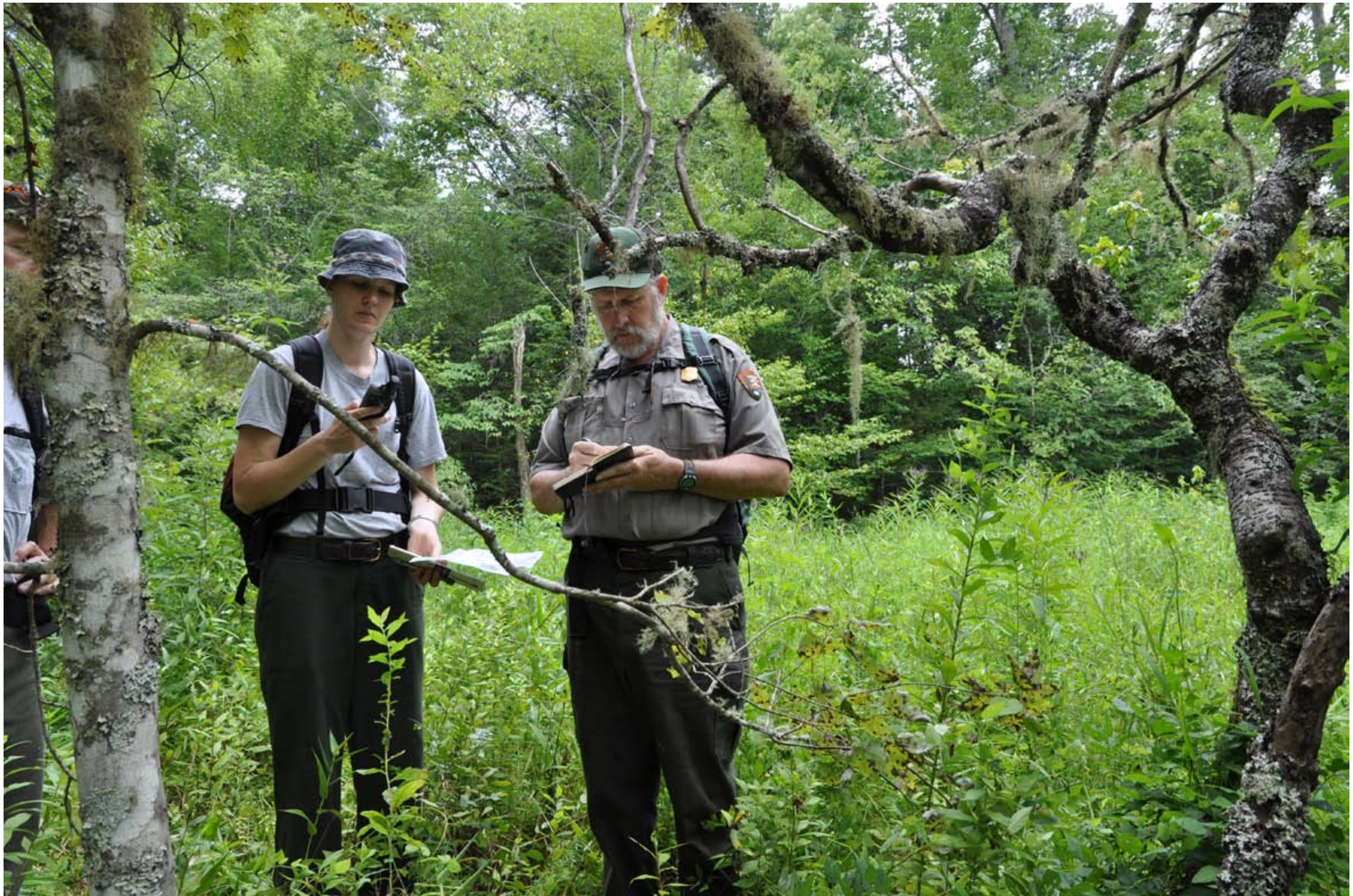
- Hundreds of analyses are possible.
- It is easier to run analyses than it is to understand them.
- In some cases it is easier to automate analyses.

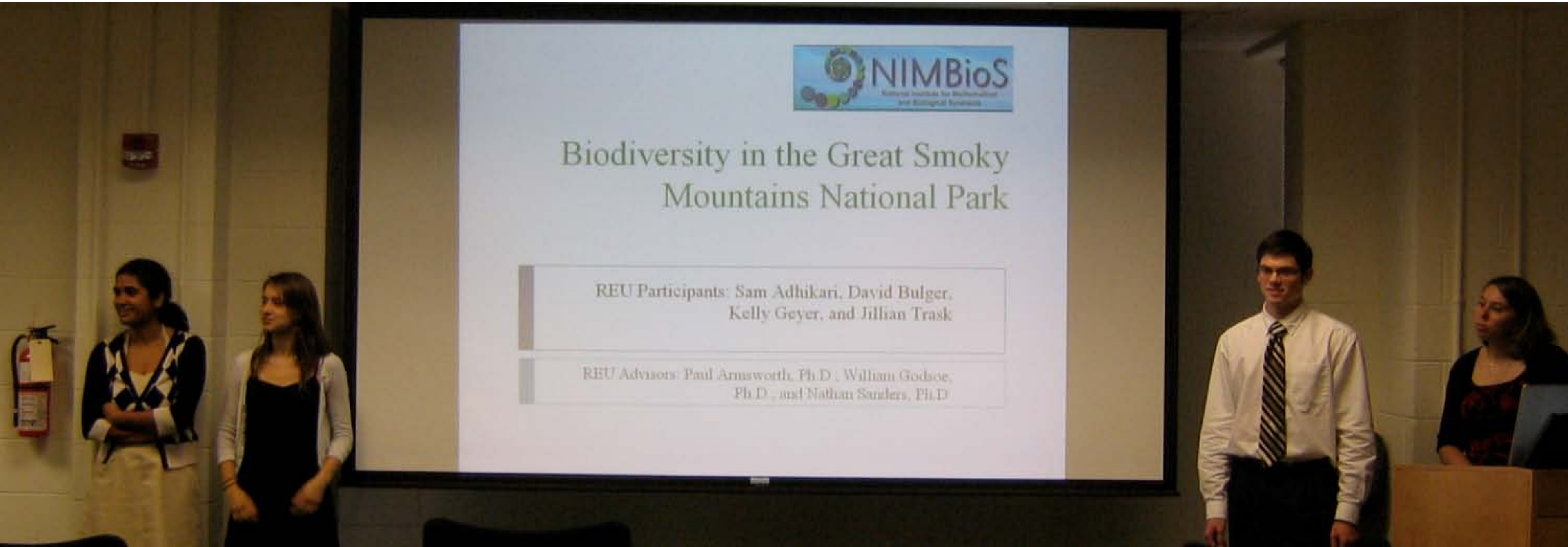
Communication forces students to simplify problems

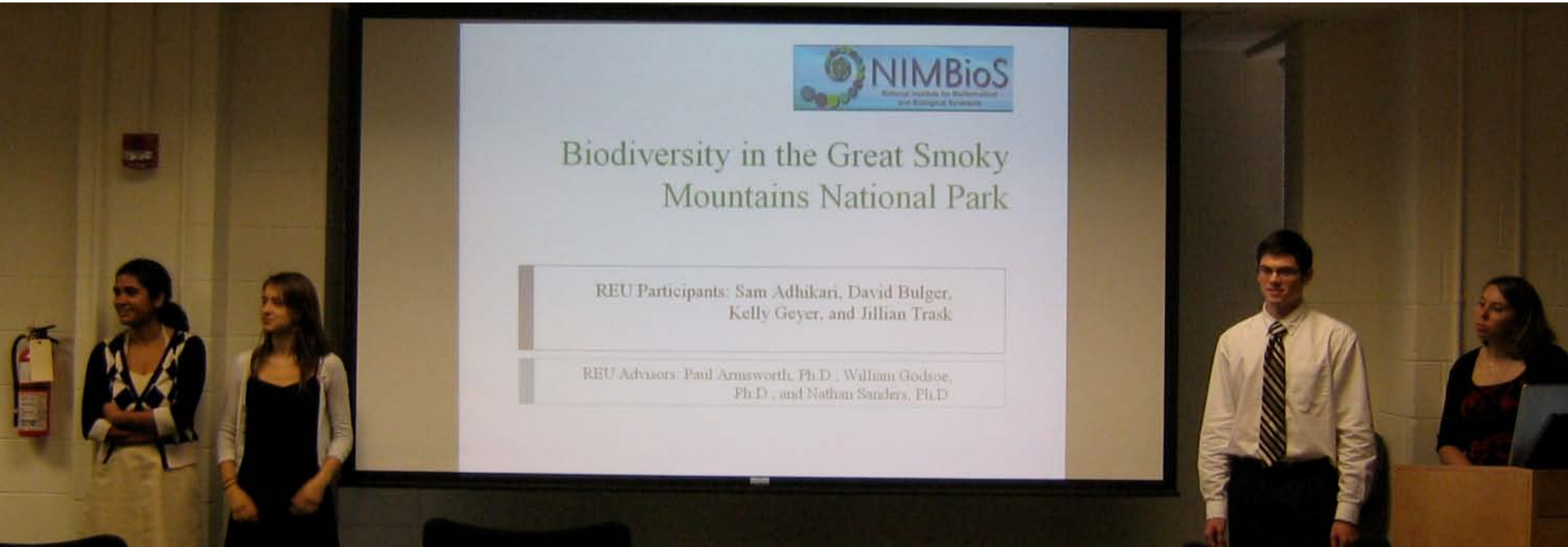
- A lot of the meetings were spent getting students to explain problems or models

- Ambiguities showed up very quickly

Fostered interactions with the National Park Service




NIMBioS
Biodiversity in the Great Smoky Mountains National Park
REU Participants: Sam Adhikari, David Bulger, Kelly Geyer, and Jillian Trask
REU Advisors: Paul Armsworth, Ph.D., William Godsoe, Ph.D., and Nathan Sanders, Ph.D


NIMBioS
Biodiversity in the Great Smoky Mountains National Park
REU Participants: Sam Adhikari, David Bulger, Kelly Geyer, and Jillian Trask
REU Advisors: Paul Armsworth, Ph.D., William Godsoe, Ph.D., and Nathan Sanders, Ph.D

Students identified two research questions:

Question 1

- What are the major components of diversity within the ATBI dataset?
  - How different are the insect communities in each site?

Question 2

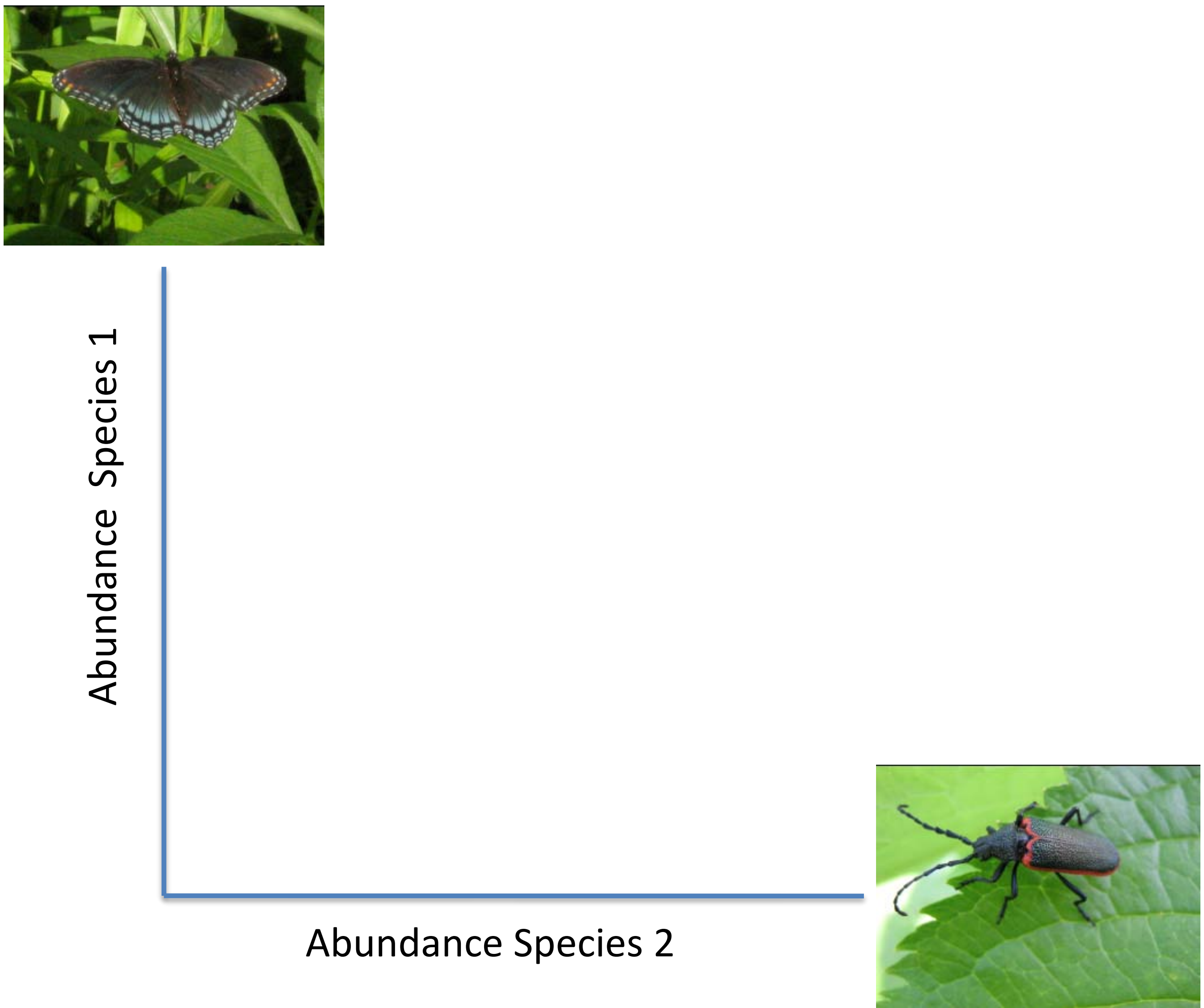
- Are our data consistent with previous observations?
  - Are particular insect groups more abundant in wet sites?

Question 1: How different are the insect communities in each site?

- Read: Legendre, P. & Gallagher, E. D. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia*.

- Sarachana developed an intuitive explanation of why there are different ways to measure diversity based on Orloci (1967).

Distance

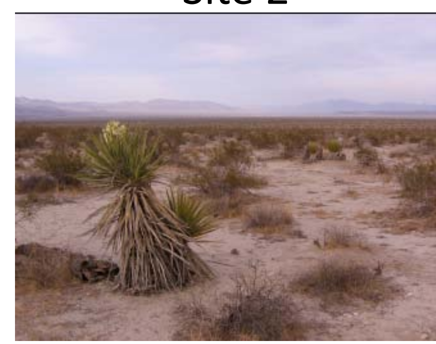


Distance

Abundance Species 1

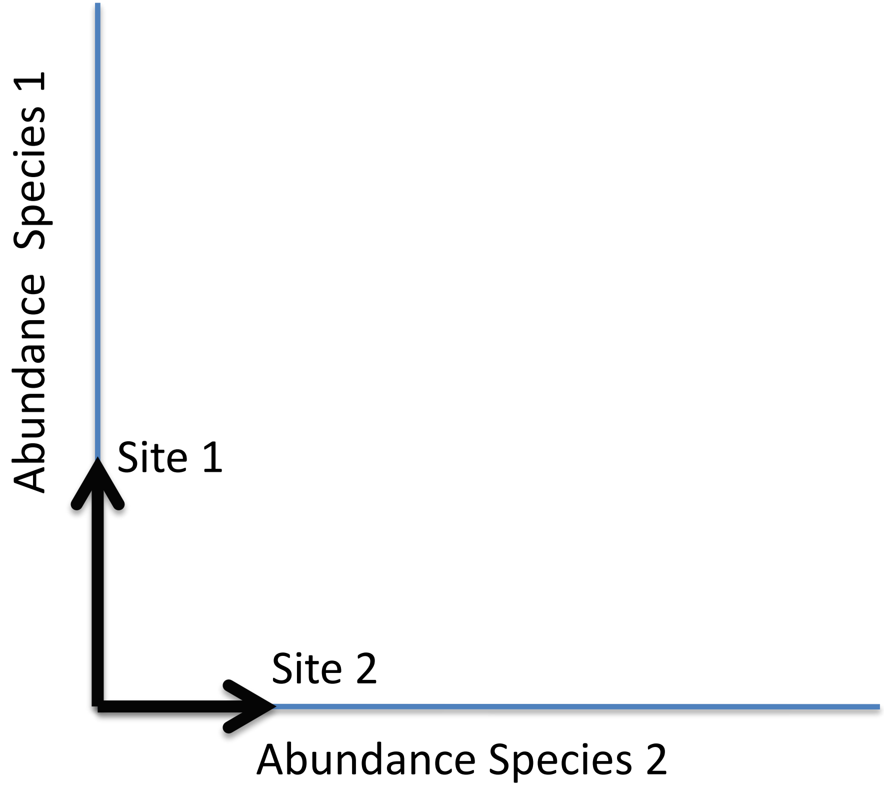
Site 1

Site 2



Abundance Species 2

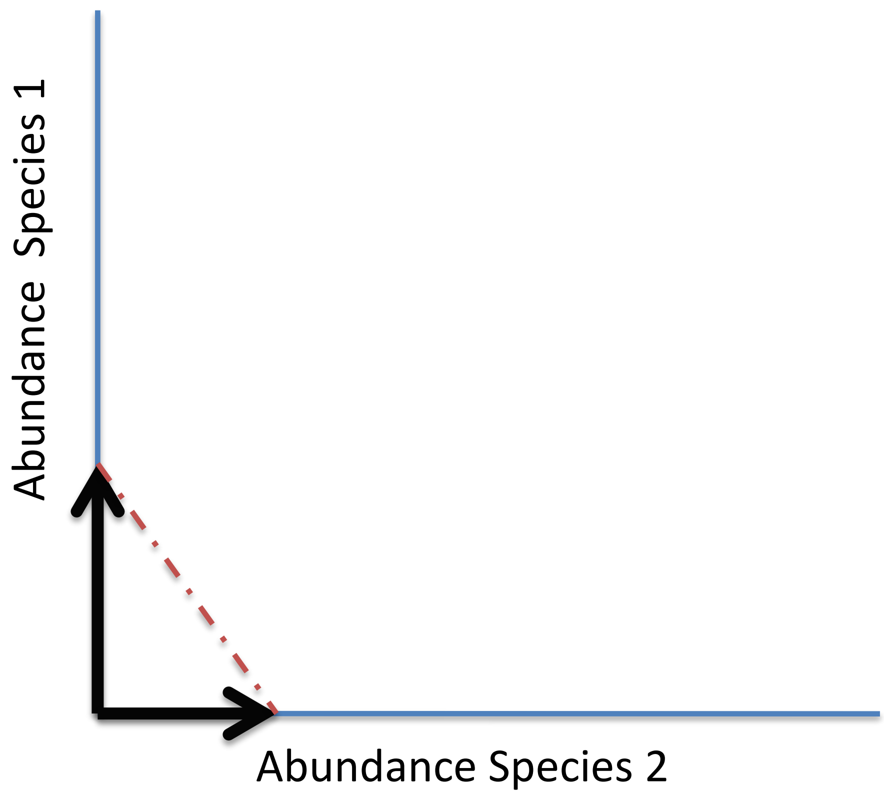
Distance



Site 1

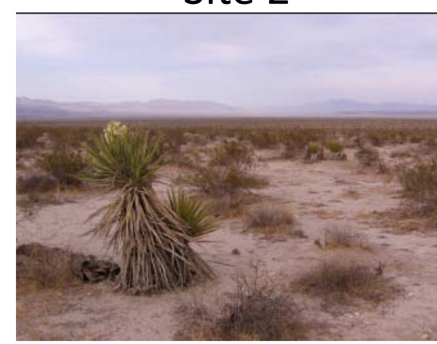
Site 2

Distance

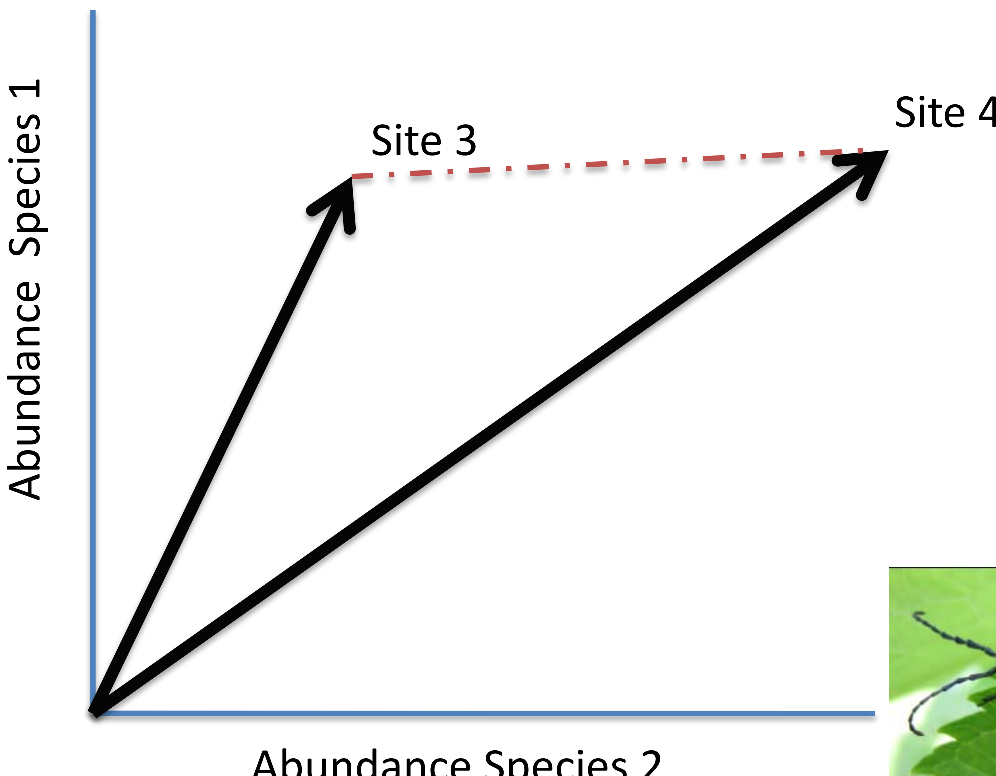


Site 1

Site 2



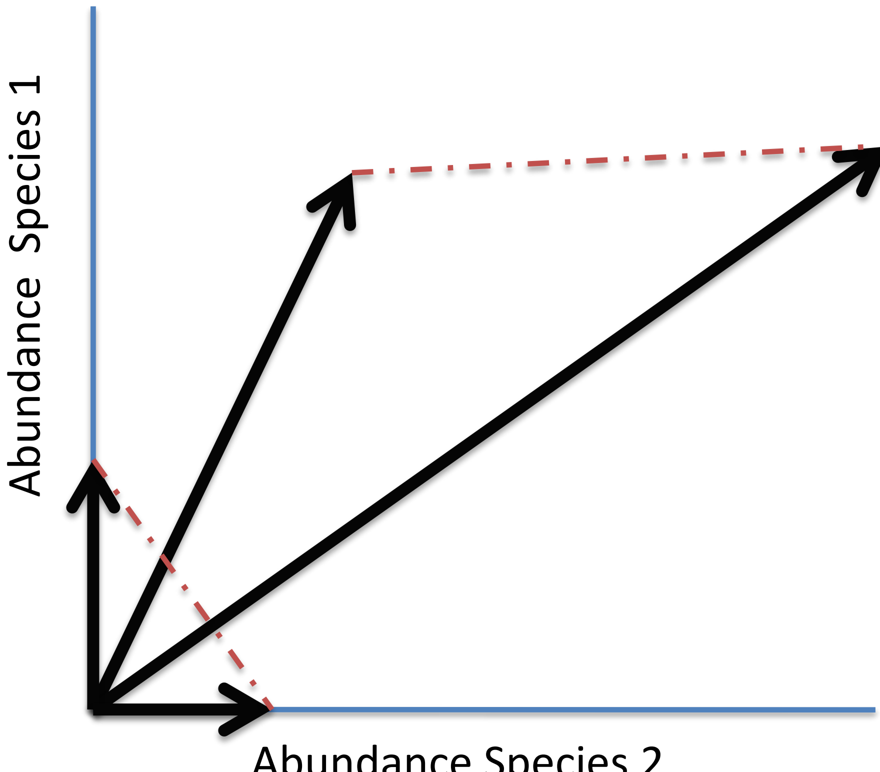
Distance



Site 3

Site 4

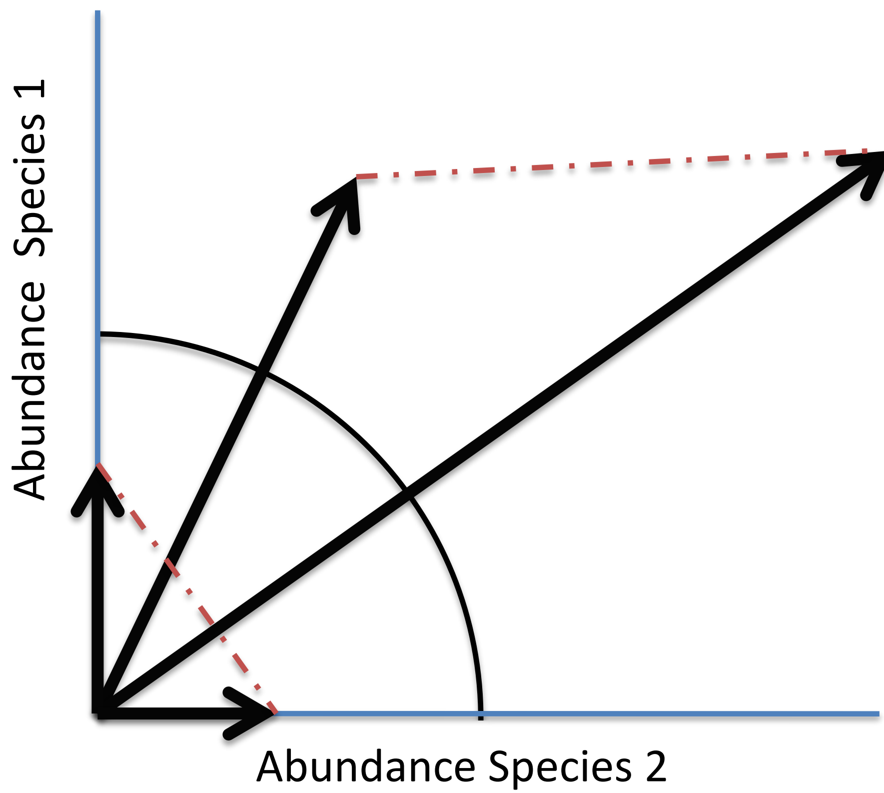
Distance



Site 3

Site 4

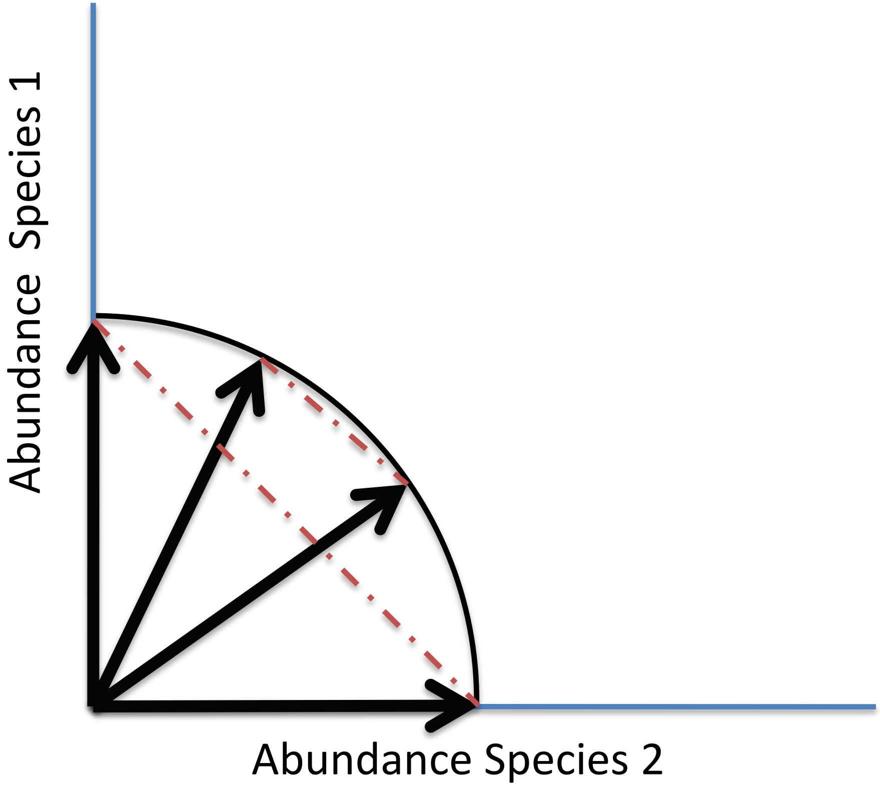
Distance



Site 3

Site 4

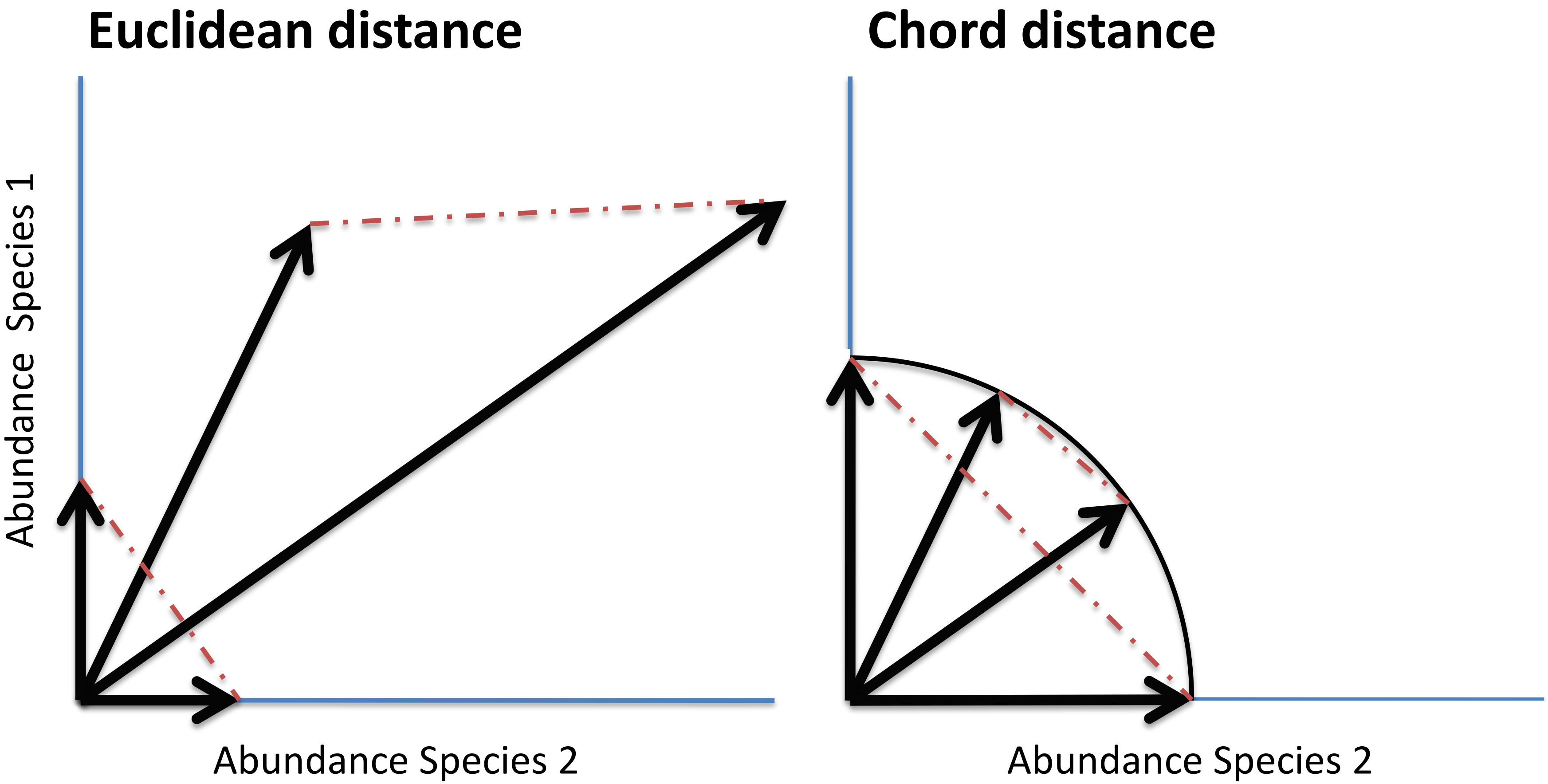
Distance



Site 3

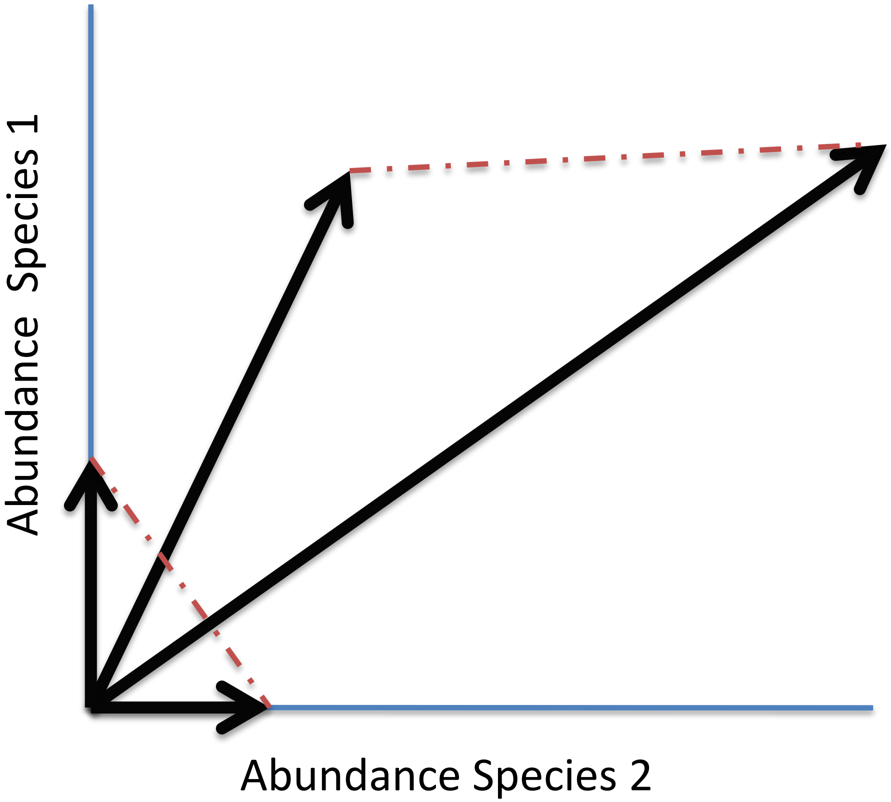
Site 4

Two (of many) ways to calculate dissimilarity, each emphasizes different components of the data

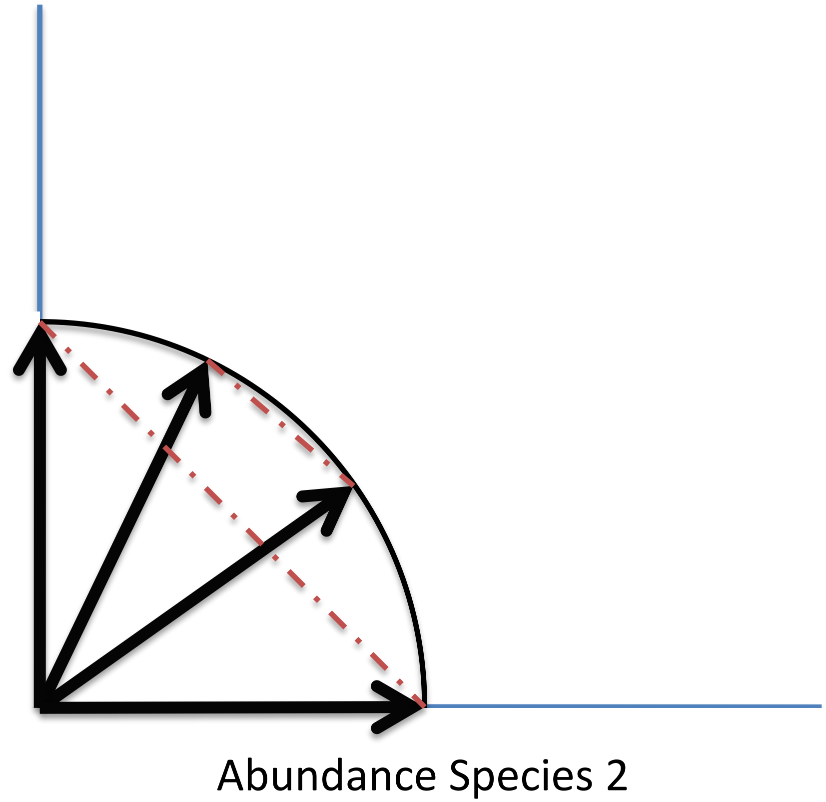


The math makes ambiguities concrete

Euclidean distance



Chord distance



Math improves hypothesis testing

- Many potential explanations exist for patterns in a large database




Only 11 replicates!

- Not enough Information to reliably test competing hypotheses
- You may need to pool some observations to test hypotheses

Statistical theory provides limits on the hypotheses you can test with a given dataset

- Not enough Information to reliably test competing hypotheses
- You may need to pool some observations to test hypotheses

Question 2: Are our data consistent with previous observations?




Focus on using existing literature to develop hypotheses

- Whittaker, R. H. 1952. A study of summer foliage insect communities in Great Smoky Mountains National Park. *Ecological Monographs* 22: 2-44.

- Stated that some families are more abundant in wet locations

ATBI dataset characterizes some sites as “wet” and some as “dry”

- n- total number of insects collected in one family in all sites
- K- number of insects in one family collected in wet sites

Model: Assume that a collected insect is equally likely to be from any site










The probability that this insect will be at a wet site is 6/11










We now model the number of insects in wet sites

Binomial model

- n- total number of insects collected in one family in all sites
- K- number of insects in one family collected in wet sites
- p- the probability that an insect will be in a wet site

$$p(K=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

You can generate hypothesis tests

- Replicates are now individual insects, not sites
- As a result the test is quite powerful.
- Many of the observations made by Whitaker are inconsistent with the observations in the ATBI database.

Importance of math to biologists

- Linking models with data is a fundamental scientific skill.

- Life science curricula will increasingly emphasize quantitative approaches.

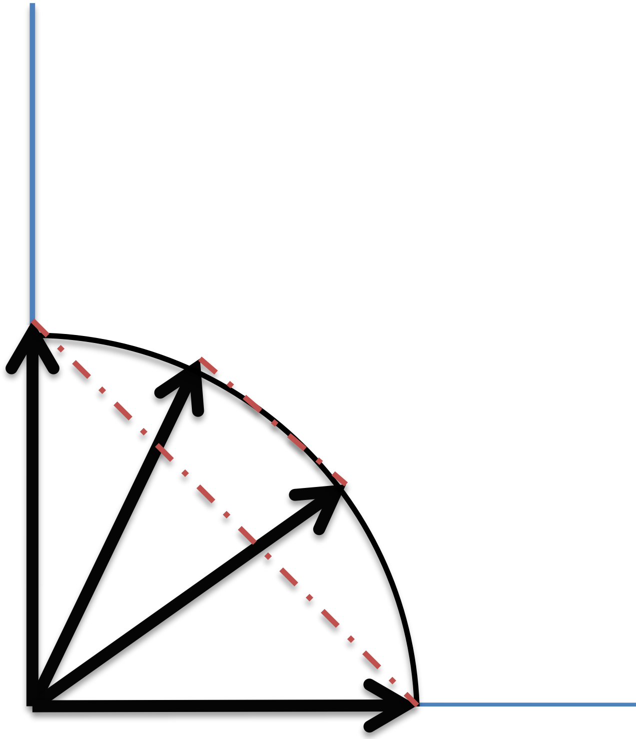
Large ecological datasets can teach important quantitative lessons

- Field work and analyses force students to experience complex biological problems

Large ecological datasets can teach important quantitative lessons

- Field work and analyses force students to experience complex biological problems

- Mathematical calculations can make ambiguities more concrete



Large ecological datasets can teach important quantitative lessons

- Field work and analyses force students to experience complex biological problems
- Mathematical calculations can make ambiguities more concrete
- Even simple mathematical models can strengthen hypothesis tests

$$p(K=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

Thanks to

Samrachana Adhikari
David Bulger
Jillian Trask
Kelly Geyer

Paul Armsworth
Nate Sanders
Suzanne Lenhart