Developing a Modeling Orientation to Science: Teaching and Learning Variability and Change in Ecosystems

Leigh Peake[1], Richard Lehrer[2], Alison Miller[3], Amanda Dickes[1], Panchompoo Wisittanawat[2] & Eliza Jacobs[1]

Background/Literature Review

- There is an increasing emphasis for Data Science education that is "grounded in **consequential investigations** in which learners pose questions, obtain data, and communicate findings within **meaningful disciplinary contexts**" (Wilkerson & Polman, 2020, p. 3).
- However, concepts and practices that students need to participate in consequential and meaningful investigations with data (Watson et al., 2018; Wild & Pfannkuch, 1999) require **coherent and sustained support** to develop (Lehrer et al., 2014; Makar & Rubin, 2009; Pfannkuch et al., 2018).
- Therefore, the current Data Science Education emphasis requires **considerations about trade-offs** between disciplinary contexts that are consequential and meaningful to students and those that are conducive to emergence and development of concepts and practices for analyzing and modeling data (Jones et al., 2017; Lehrer & English 2018; Makar & Ben-Zvi, 2011).
- Ecological issues are becoming one of the most pressing concerns for youth (e.g., O'Brien et al., 2018). As a context to initiate and sustain practices for visualizing, measuring and modeling variability (Lehrer & Schauble, 2017), **ecology** is
  - *Promising,* because variability is ubiquitous, ecological data are becoming more available, and students can have direct experiences in constructing and taking samples, as well as
  - *Challenging,* because sources of variability are often not readily visible to students.

Research Questions

We aim to investigate the development of middle school students' understanding and practice of modeling in the context of data-rich citizen science ecological investigations. Two questions guide our research:

(1) How do students conceive of and employ material analog models (microcosms) to support inquiry and investigation of ecosystems?

(2) How do students conceive of and employ data models, including models of random processes, to support inquiry and investigation of ecosystems?

In this poster, we focus on *how students constructed and revised models of chance as accounts for variability and uncertainty* across different ecological investigations.

Methods: Classroom Design Research

Participants & Setting

- 7th-grade science classroom (3 sections, ~15 students per section)
- Two ecological investigations each year: intertidal ecosystems (9 weeks, fall semester), vernal pool ecosystems (5 weeks, spring semester)

Instructional Sequences

Typical sequences included 1) introduction to citizen-science questions ("What is the distribution of species? How will their distribution change as the climate warms?"), 2) initial visit to field site, 3) microcosm investigations, 4) field sampling, 5) data analysis and modeling, 6) communicating results. These were intertwined during instruction to support relations across models and forms of investigations.

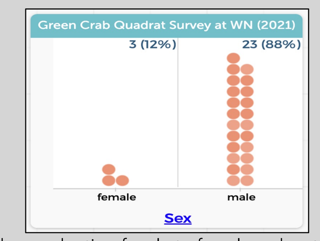
Semester	Microcosm	Field Sampling	Data & Models
Fall Year 1 (Intertidal)	One classroom microcosm (Where and how much crabs hide?)	What is the sex ratio of green crabs at the intertidal?	• Construct chance models of field outcomes (sex ratio) • Analyze data to compare sex ratio at our site in relation to other sites
Spring Year 1 (Vernal pools)	Multiple microcosms (e.g., What leaves do caddisfly larvae prefer?)	How many samples should we take to be confident about species absence?	• Analyze microcosm data • Analyze field data to compare species absence at one pool to detection probability across sites • Construct chance models of field outcomes (sample size)
Fall Year 2 (Intertidal)	Multiple microcosms (e.g., Which substrates do crabs prefer? How aggressive are green crabs?)	n/a	• Analyze microcosm data • Construct chance models of microcosm outcomes (e.g., food preference)
Fall Year 2 (Vernal pools)	Multiple microcosms (e.g., Does water influence egg masses and tadpole growth?)	How many samples should we take to be confident about species absence?	• Analyze microcosm data • Analyze field data to compare species absence at one pool to detection probability across sites • Construct chance models of field outcomes (sample size)

Data Collection

- Videotaped whole-class and small-group conversations
- Videotaped post-instruction interviews with a stratified random sample of students (n = 14 in Year 1, and n = 28 in Year 2).

Modeling Chance in Field Samplings

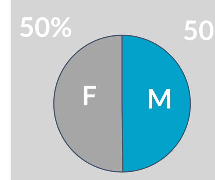
Project Intertidal Crabs

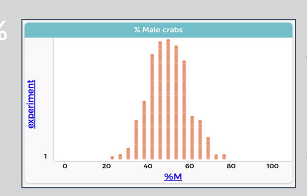


The observed ratio of male to female crabs at field site

A 50-m transect was placed along the water line and 26 quadrats randomly placed along the transect. The number, sex, and species of intertidal crabs was recorded. Exploration of data collected at the field site revealed a "wild" ratio of 88% male, 12% female.

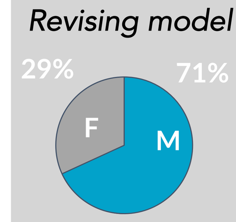
How likely is it our observed sex ratio occurred just by chance?

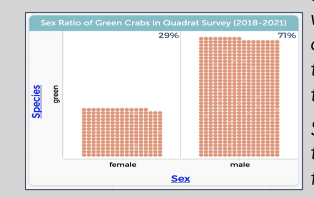




S: I don't have anything over 80 or under 20 … it makes me think that's it's even less likely. Like, I was thinking before that it probably was just by chance when we were [modeling] with 10, but with 26 it shows that it's even less likely because I didn't even get one point that was above 80 [percent male].

Revising model assumption





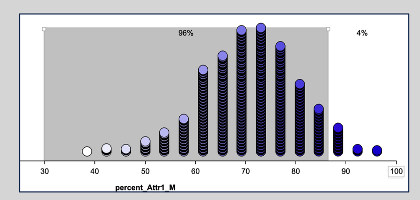
S1: I agree with [classmate], like, when we did the 50/50 spinner, we almost didn't have it up towards 90%, so I think that there's just going to be more males than females out in the ocean.

S2: I agree, but I don't [know] if it's like the whole ocean has more male than female crabs, it might just be at [our field site]. I definitely think the number of male crabs to female crabs is more than 50/50.

T: What is actually happening when we find 88 in this [revised] model, but not in the 50/50 model. What does that mean… for our understanding of the sex ratio of male to females in the field, and our data collection?

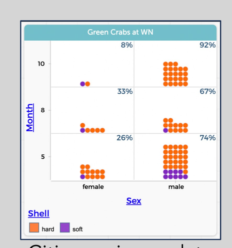
S1: Because most likely there are more males on the coast of Maine.

S2: I feel like this data is more accurate because our data can fit into this one, and it couldn't fit in the other one.



Search for possible mechanism

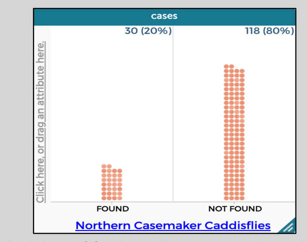
S: That it's mostly because of their molting that we don't see many females because we checked only in the Fall and that's when females molt and will hide, and the males are hard shells, and they don't typically hide.



Citizen science data

Professional data

Project Vernal Pools

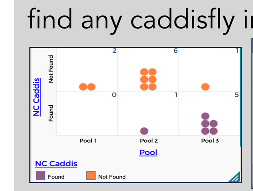


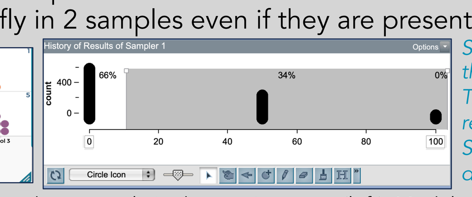
Probability of finding NC in a sample in Maine

One scoop of pond detritus equals one sample. Sampled detritus is searched for the presence or absence of indicator species. Combined data collected from pools across Maine indicated the probability of finding a Northern Casemaker caddisfly in one sample was 20%.

Modeling Detectability: How many samples should we collect?

Students considered data collected by seventh graders the previous year. Only 2 samples were collected from Pool 1 and no indicator species found in either sample. The students considered how likely it is to not find any caddisfly in 2 samples even if they are present in the pool.





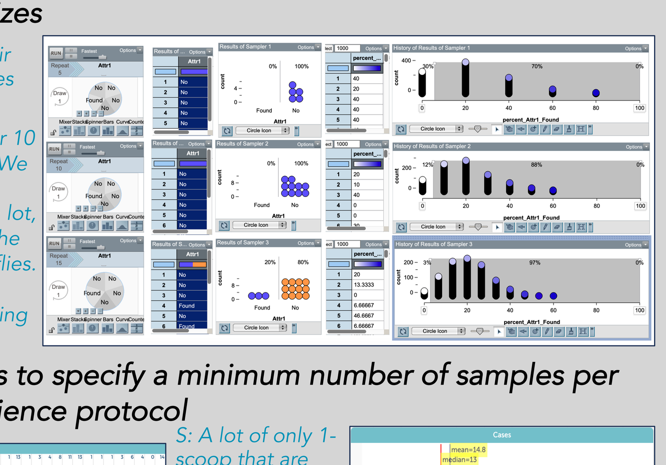
S: There's a lot more zeroes than I thought.
T: And what do those zeros represent?
S: The ones [trips] when you didn't find anything.

NC Caddisfly detected per sample in the previous year (left). Modeling 2-samples at 20% detection probability (right).

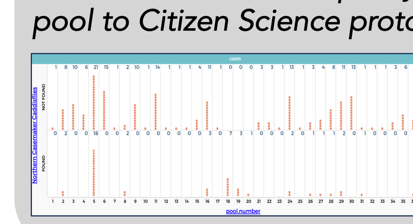
Considering the probability of finding no caddisfly in any sample at different sample sizes

T: Has anyone revised their idea for how many samples we should take?

S: I originally had said 5 or 10 [samples], and I think 15. We don't want to take 20 [samples] because that's a lot, but 15 we found a lot of the time we're finding caddisflies. There's not as many [caddisflies] that aren't being found.

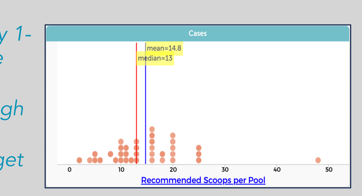


Recommendations to specify a minimum number of samples per pool to Citizen Science protocol



Data collected from 38 pools in Maine

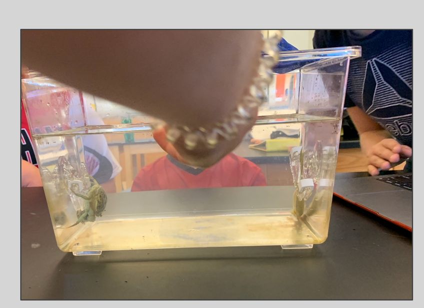
S: A lot of only 1-scoop that are getting not found. It's a high chance you're not going to get it [a caddisfly] from one scoop.

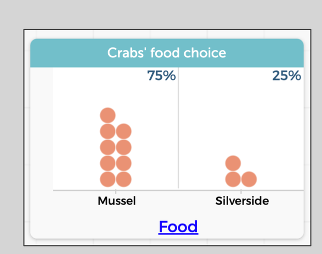


Recommended # of samples/pool

Modeling Chance in Microcosm Investigations

Crab Food Preference

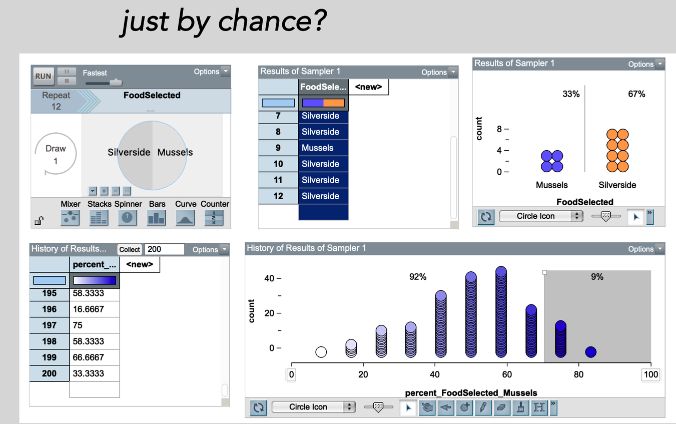




A student team designed a microcosm investigation to study green crabs' food preference. They attached two food sources (mussel and silverside) on opposite sides of the tank, placed a hungry crab in the middle, and observed which food the crab ate first. Over 12 observations, they found that a crab ate mussel first 9 times (75%). Therefore, crabs prefer mussel.

How likely is it a crab chooses mussel first 7 out of 12 times just by chance?

Students built a model to simulate outcomes of the number of times crab selecting mussel first in 12 trials even when crabs had no preference (i.e., equal chance to select either food).



S: If [food selection] is random, only 9% of the time we would get the data that we got, and 9% is low.

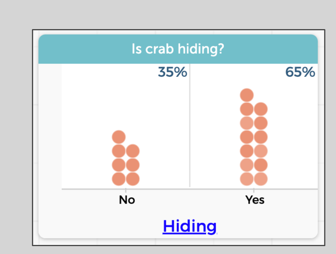
Challenges in grasping model assumption

Students ascribed intentionality to animal behaviors, e.g., "We think that crabs have a preference just like every other animal on the planet. Why would crabs not have one too?" Therefore, some students were skeptical about using a random device to model crab behaviors:

S1: It's not like the crab is the spinner that you're spinning. … the spinner is not the crab's head.

S2: From your perspective, yes, this data may look like it's not even related to the crab's brain or anything. But we're saying that the crab's brain is preferring something. This [model] is saying that the crab is completely random 50/50. We're not saying that this [model] is the crab's brain, we're saying if it was the crab's brain, it would be like this.

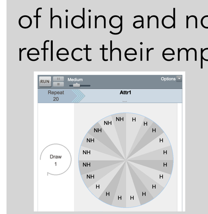
Crab Hiding & Substrate Preference



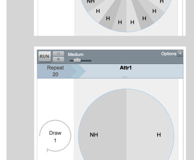
A student team designed a microcosm investigation to study green crabs' hiding (refuge-seeking) behavior. The tank included multiple crabs and potential hiding spots—first using natural materials such as seaweed, but later using artificial objects to aid observations. They observed the tank once a day and recorded whether a crab was hiding or not hiding. Overall, they found a crab hiding 7 times and a crab not hiding 13 times. Therefore, students concluded that crabs spend more time hiding than not hiding.

How likely is it we observe a crab was hiding 13 out of 20 times just by chance?

Students disagreed about how to construct a chance model to estimate probability of finding crabs hiding more times (13 out of 20 times) just by chance. Some students thought their model should represent equal chance of hiding and not hiding, but other students thought their model should reflect their empirical data to estimate the probability of crab hiding.



S1: I think [the spinner] should be 7 [not hiding] and 13 [hiding] because it's supposed to be representing the data that we've already collected and try to figure out if that data is random or not random so I think it should be data that you already collected.



S2: I think it should be evenly split … I feel like it makes more sense because if it's really even, you could get more a more of an even chance of getting hiding or not hiding.

A similar issue arose in another team's microcosm investigation.

T: So, why did you choose to build your spinner like that [equipartitioned in 3 splits to represent 3 substrates]?

S: So, one because we only have three different substrates, seaweed, rock and sand. So, it would give each one an even chance. But at first, I thought that seaweed should have had a bigger one [consistent with tank data: finding crabs most often in seaweed]. But then once my group really explained it to me, which I think was really helpful, it helped me see that if it's all, if it's like that, how would we be testing if it's random, it has a higher percent chance landing on that one every time.

Student Interviews

Year 1: Interview protocol focused on modeling chance in field data (sex ratio)

Our post-instruction flexible interview (N=14) concluded with model simulations of a random sample and of sampling variability, with the goal of probing students' conceptions of the purposes and interpretations of models of chance.

Concepts	% correct	Examples
Probability structure of the spinner	93	"Well, it makes sense that it has 30% female and 70% male, because that's what the probability supposedly is."
Sample size	93	"Oh, repeat 20 because she found 20 crabs."
A single case in a sample	100	S: So, like it (a dot in the plot of frequencies of male, female) stands for one spin around the circle. I: Okay. And so, on that spin, what did the spinner land on? S: On this in particular one, it landed on the male.
A single case in a sampling distribution	85	S: So, that's for one of the rounds that there were 40% female. I: And so how many crabs does that dot represent? S: 20.
Using a sampling distribution to express uncertainty about claims	92	"Well, it shows that it's not very likely because at the 50% spot, there's only like four or five and it's out of a hundred, so it shows that's not very likely."

Year 2: Revised interview protocol to include a sharper focus on the counterfactual nature of statistical reasoning

Let's imagine we're having a conversation with an 8th grader who visited [gym teacher's] pools last year when they were in seventh grade.

Seventh grader: We found NC caddisflies in Pool 1!
Eighth grader: That's so great! I'm happy the NC caddisflies have been discovered in Pool 1 this year. Last year we took 2 scoops from Pool 1 and didn't find any caddisflies, so we know for a fact they were not there last year.

1. What do we think about the 8th grader's statement? [Students likely say they can't be sure, especially because they only took two samples.]
2. How would you modify this model to estimate how likely the eighth graders missed the caddisflies in their 2 scoops even though the caddisflies were there?
3. We took 16 scoops this year, and we found caddisflies in Pool 1. What if next year's 7th graders take 16 scoops in this same pool and do not find anything? What is the probability that they don't find any caddisflies in all 16 scoops?
4. Next year's 7th grader says, "We have taken 16 scoops from Pool 1, and we don't find any caddisflies, so we now know for a fact the caddisflies are no longer here this year." What do you think about that statement?

Discussion & Next Steps

- Our findings contribute to the ongoing efforts in Data Science Education to support consequential investigations in meaningful disciplinary contexts.
- Ecological investigations can serve as contexts for introducing and developing classroom practices for modeling variability. We have identified fruitful candidate tasks (Lehrer & Schauble, 2021) that integrate modeling random variability in students' field sampling and microcosm investigations.
- Our findings also capture tensions in designing to integrate practices for modeling variability across different forms of ecological investigations. We are currently exploring these tensions in our analyses. Clarifying these tensions can inform future design work.

Context	Affordances	Challenges
Sex ratio (field sampling)	• The equiprobable model is **consistent with students' initial expectation** about equal sex ratio. • Models of chance provide accounts that **field outcomes are indeed unlikely**, though consistent with expected sampling variability if the probability structure is altered to reflect a non-equiprobable estimate of probability. • This motivates a **search for mechanisms.**	• Observations in citizen science data—largely contributed by school groups—are dense during field-trip seasons but sparse or absent otherwise. Potential **mechanisms** involving seasonality are **thus uncoverable in citizen science data.** We've begun to complement student data investigations with analogous professional data.
Number of samples (field sampling)	• Using random devices to represent species detection is **consistent with students' first-hand experiences** that detection varies across samples. • Modeling can **inform field sampling** protocols.	• For some students, the **equiprobable assumption persists** in model construction (Biehler et al., 2017; Noll & Kirin, 2017) despite the evidently low relative frequencies of detecting any caddisfly in a sample in the larger, citizen science data corpus.
Behavioral studies (microcosms)	• Behavioral studies are **recurring and generative foci** of students' microcosm investigations.	• Students ascribe intentionality to sentient behaviors. **Using a chance model to represent behaviors is thus counter-intuitive for students.** This reflects historical and ongoing contests over statistical modeling of behaviors (Fisher, 1935; Rosen et al., 2006).

This project is funded by the National Science Foundation, grant # DRL-2010207, 2010119, and 2010155. Any opinions, findings, and conclusions or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[1] Gulf of Maine Research Institute lpeake@gmri.org; adickes@gmri.org; ejacobs@gmri.org
[2] Vanderbilt University rich.lehrer@vanderbilt.edu; panchompoo.wisittanawat@vanderbilt.edu
[3] Bowdoin College amiller2@bowdoin.edu