

SSA 200: Strategic Use of Data

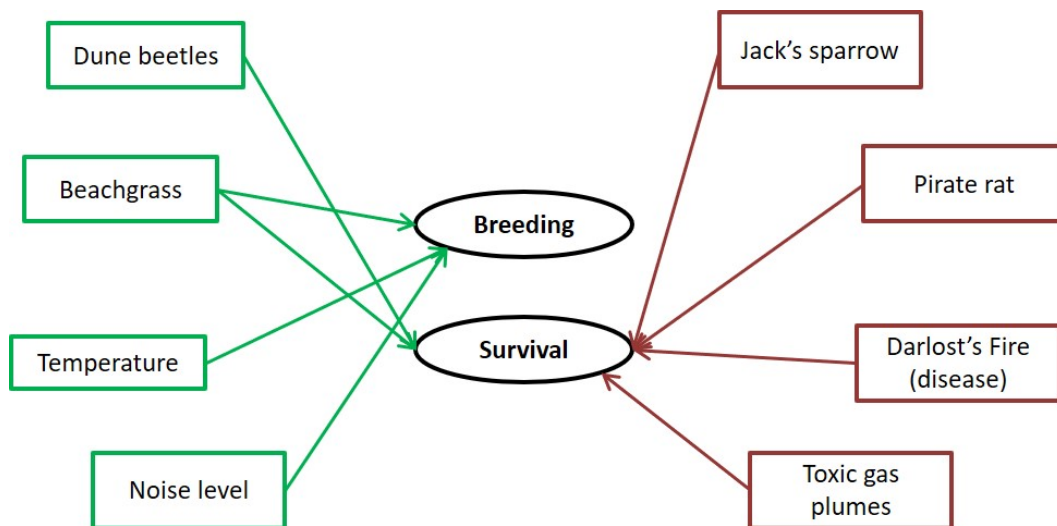
Activity 1 – Linking Conceptual Diagrams to Statistical Models

Objectives:

- Understand the link between conceptual diagrams and statistical models
- Develop ecological hypotheses about species needs and translate hypotheses into biologically-meaningful models
- Choose appropriate ecological variables from available data sources to include in models

Background Information:

The Island Mouse (*Zapus islandsonious*) is a species that lives on Darlost's Island. There are 10 current populations, although historically they occurred at 11 sites across the island. As you covered in SSA 100, the mice are adapted to the unique ecology of Darlost's Island. Their primary food source is dune beetles, which are found across the small island. They also rely on beach grass that grows along the coasts to build their nests and create shelter from the sun and cold winters. They require warm spring temperatures for successful breeding, and in years with exceptionally cold trade winds there is increased winter mortality. They are also very sensitive to noise, and loud noises can lead to depressed breeding activity and, in extreme cases, death from cardiac arrest. There are also several stressors on the island that influence their population dynamics. The two main predators of Island mice are Jack's sparrow, which are found across the island, and the pirate rat, which are only found along the coasts. Darlost's fire is a disease that causes mange and can be fatal, and toxic gas plumes from the island's volcanoes can also be deadly. This conceptual diagram summarizes the key species needs (green boxes) and threats (red boxes) and how they influence breeding and survival.



Using this conceptual model and what you know about the ecology of the mice, you are tasked with developing models that will inform how these factors influence Island mouse occurrence, abundance, and breeding success across their range.

The goal of model selection is to find the model that best explains the data using the fewest number of parameters, which will hopefully identify the most important ecological drivers of species success. It is standard practice to always include a **null model** in your model set (this may sometimes also be referred

to as the “dot model”), which is a model that does not include any covariates. Sometimes none of the data we have are good predictors of the response variable of interest. Does that mean the models are useless? This result still tells us about the system, in that it can tell us that the things we thought were important drivers of species success maybe aren’t as important, and there could be other factors that we should consider.

When lots of different data sources are available, it can be tempting to fit “all subsets”, or all possible combinations of available covariates. This kitchen sink approach is not good practice for a few reasons. First, it can be time consuming and computationally intensive to fit all possible combinations. Second, and more importantly, it can result in the selection of models that might not make a lot of sense ecologically or be hard to use in future predictions. Finally, it’s just not good science because it usually results in trying to come up with theory to explain observations *post hoc*, instead of conducting meaningful tests of biological hypotheses.

It’s also important to carefully consider the combinations of covariates to include. We want to avoid **collinearity** in model sets, which happens when one or more covariates are correlated with each other. This can be an issue for estimation of the effects of each variable alone because they are not independent of each other. For example, if we are interested in predicting grizzly bear occurrence, we may want to test a model that includes the presence of a water source or the presence of fish at a given site. Since the presence of fish is highly correlated with the presence of water, those covariates could be collinear, and we may not want to include both in the same model. The presence of fish implies the presence of water, and may be closer to the real ecological reason why bears occur near water. However, if data about fish presence isn’t available, including water in the model can be a good proxy.

For each of the following scenarios, don’t worry about the mechanics or details of fitting the models, but focus instead on the combinations of covariates that can best explain the response variable of interest. Write your models both as a formula that lists all covariates (feel free to use your own abbreviations) and as a sentence that describes the ecological hypothesis that the model represents. We’ve included space for six models, but feel free to include fewer or more candidate models in your list. For example, consider the following set of models to predict grizzly bear occurrence:

	Model (formula)	Hypothesis (sentence)
1	water + trees + rocks	Bear occurrence is dependent on the presence of water source, forest cover, and rocky substrate.
2	water + trees	Bear occurrence is dependent on the presence of a water source and forest cover.
3	water	Bear occurrence depends primarily on the presence of a water source.
4	null model	Bear occurrence is not strongly associated with water source, forest cover, or rocky substrate.

Part 1: Occupancy

The Island Mouse Recovery Team is interested in mapping the drivers of occurrence across Darlost's island in order to assess the suitability of other nearby islands for possible translocations. They have collected occurrence data from a variety of sources and want to use species distribution modeling to find the environmental covariates that best predict occurrence. Using the following data sources, develop a set of candidate models to predict occurrence of Island mice.

Response variable = Probability of island mouse presence

Source	Data Type	Variables Measured
National Land Cover Database (NLCD)	Land cover/land use in 30km grid	<i>Each cell is classified as one of:</i> Open water Perennial Ice/Snow Developed, Open Developed, Low Developed, Medium Developed, High Rock/sand/clay Deciduous forest Evergreen forest Mixed forest Dwarf scrub Shrub/scrub Grassland Sedge Lichens Moss Pasture/Hay Cultivated crops Woody wetlands Emergent wetlands
WorldClim Bioclimatic variables	Climate data in 30km grid	Annual Mean Temperature Max Temperature of Warmest Month Min Temperature of Coldest Month Temperature Annual Range (Max – Min) Mean Temperature of Wettest Quarter Mean Temperature of Driest Quarter Mean Temperature of Warmest Quarter Mean Temperature of Coldest Quarter Annual Precipitation Precipitation of Wettest Month Precipitation of Driest Month Precipitation of Wettest Quarter Precipitation of Driest Quarter Precipitation of Warmest Quarter Precipitation of Coldest Quarter
NOAA Weather Station	Daily local weather data	Maximum air temperature

	from 3 weather stations across island and 1 weather buoy offshore	Minimum air temperature Average air temperature Maximum wind speed Minimum wind speed Average wind speed Total precipitation Maximum water temperature Minimum water temperature Average water temperature
--	---	--

Occupancy candidate model set

	Model (formula)	Hypothesis (sentence)
1		
2		
3		
4		
5		
6		

Part 2: Abundance

A graduate student has developed a project to estimate Island mouse abundance and map the drivers of abundance across its range. His field crew conducted transect surveys at randomly-selected points across Darlost's island to estimate the abundance of Island mice while accounting for detection probability. He and his crew conducted vegetation surveys at each transect point and also recorded the presence or absence of Jack's sparrows and pirate rats each survey. Use the following data to construct a set of models to predict mouse abundance.

Response variable = Site abundance

Source	Data Type	Variables Measured
Vegetation surveys	Vegetation data collected at each transect	Percent canopy cover Percent shrub cover Percent herbaceous cover Substrate: rock, soil, sand, other Soil type: clay, silt, loam, NA Beachgrass density Average DBH Ambient noise level (decibels) Toxic gas level
Predator surveys	Presence/absence of predators during each survey	Jack's sparrow presence/absence Pirate rat presence/absence
Other transect info	Measured via GIS	Distance to nearest coast Distance to nearest volcano Ecotype: coastal, palms, mountains
NOAA Weather Station	Daily local weather data from 3 weather stations across island and 1 weather buoy offshore	Maximum air temperature Minimum air temperature Average air temperature Maximum wind speed Minimum wind speed Average wind speed Total precipitation Maximum water temperature Minimum water temperature Average water temperature
NLCD	Land cover/land use in 30km grid	<i>Each cell is classified as one of:</i> Open water Perennial Ice/Snow Developed, Open Developed, Low Developed, Medium Developed, High Rock/sand/clay Deciduous forest Evergreen forest Mixed forest

		Dwarf scrub Shrub/scrub Grassland Sedge Lichens Moss Pasture/Hay Cultivated crops Woody wetlands Emergent wetlands
--	--	---

Abundance candidate model set

	Model (formula)	Hypothesis (sentence)
1		
2		
3		
4		
5		
6		

Part 3: Breeding success

A study is conducted to measure breeding success in the Beach Bums, Dead Man's Dunes, and Misty Mountain populations. Over the course of five years, mouse nests are surveyed throughout breeding to estimate the number of offspring produced per female. Pitfall traps are also placed at each study population to estimate invertebrate abundance. Use the following data sources to develop models to determine factors associated with the number of offspring per female.

Response variable = Number off offspring per female

Source	Data Type	Variables Measured
Nest surveys	Data collected at each nest	Ecotype: coastal, palms, mountains Toxic gas level at nest Nest substrate: beachgrass, sand, shrub, other Female age Female disease status (test positive/negative)
Pitfall traps	Index of invertebrate diversity and abundance collected for each site in each year	Dune beetle abundance Total insect abundance Number of species detected
WorldClim Bioclimatic variables	Climate data in 30km grid	Annual Mean Temperature Max Temperature of Warmest Month Min Temperature of Coldest Month Temperature Annual Range (BIO5-BIO6) Mean Temperature of Wettest Quarter Mean Temperature of Driest Quarter Mean Temperature of Warmest Quarter Mean Temperature of Coldest Quarter Annual Precipitation Precipitation of Wettest Month Precipitation of Driest Month Precipitation of Wettest Quarter Precipitation of Driest Quarter Precipitation of Warmest Quarter Precipitation of Coldest Quarter
NOAA Weather Station	Daily local weather data from 3 weather stations across island and 1 weather buoy offshore	Maximum air temperature Minimum air temperature Average air temperature Maximum wind speed Minimum wind speed Average wind speed Total precipitation Maximum water temperature Minimum water temperature Average water temperature
NLCD	Land cover/land use in 30km grid	<i>Each cell is classified as one of:</i> Open water

		Perennial Ice/Snow Developed, Open Developed, Low Developed, Medium Developed, High Rock/sand/clay Deciduous forest Evergreen forest Mixed forest Dwarf scrub Shrub/scrub Grassland Sedge Lichens Moss Pasture/Hay Cultivated crops Woody wetlands Emergent wetlands
--	--	--

Breeding success candidate model set

	Model (formula)	Hypothesis (sentence)
1		
2		
3		
4		
5		
6		