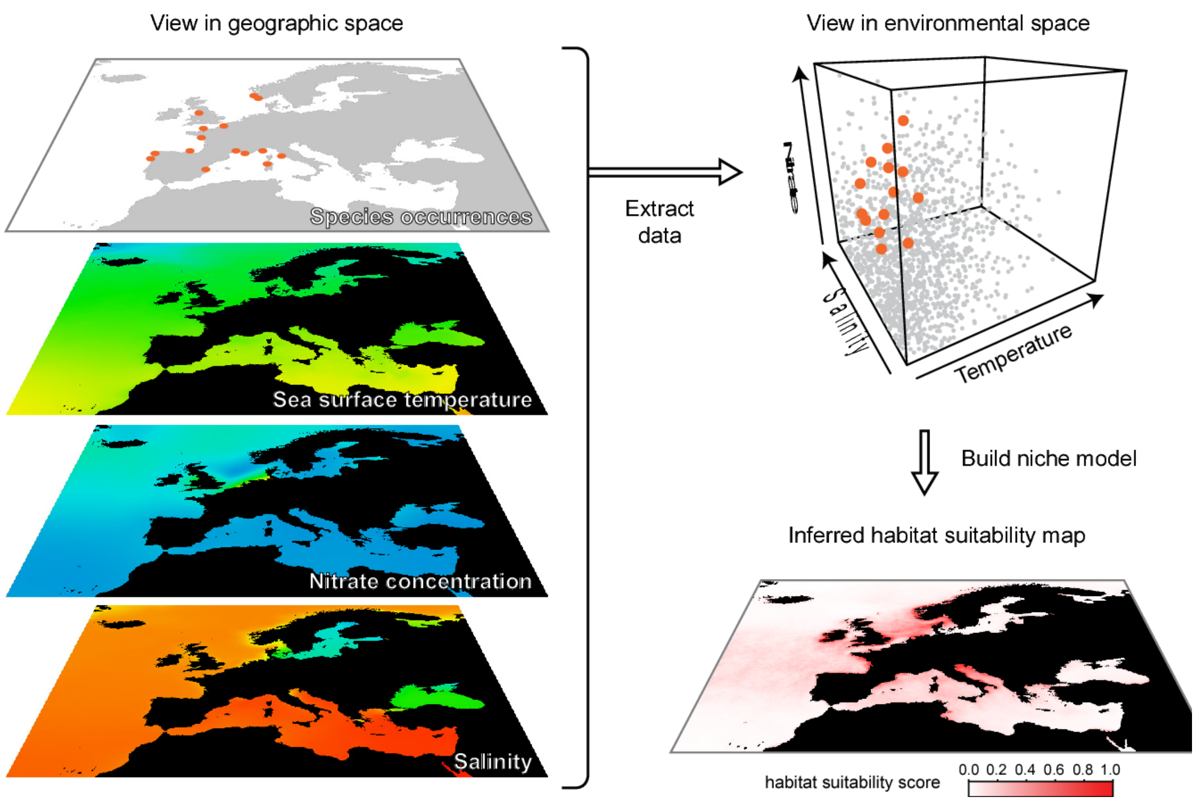
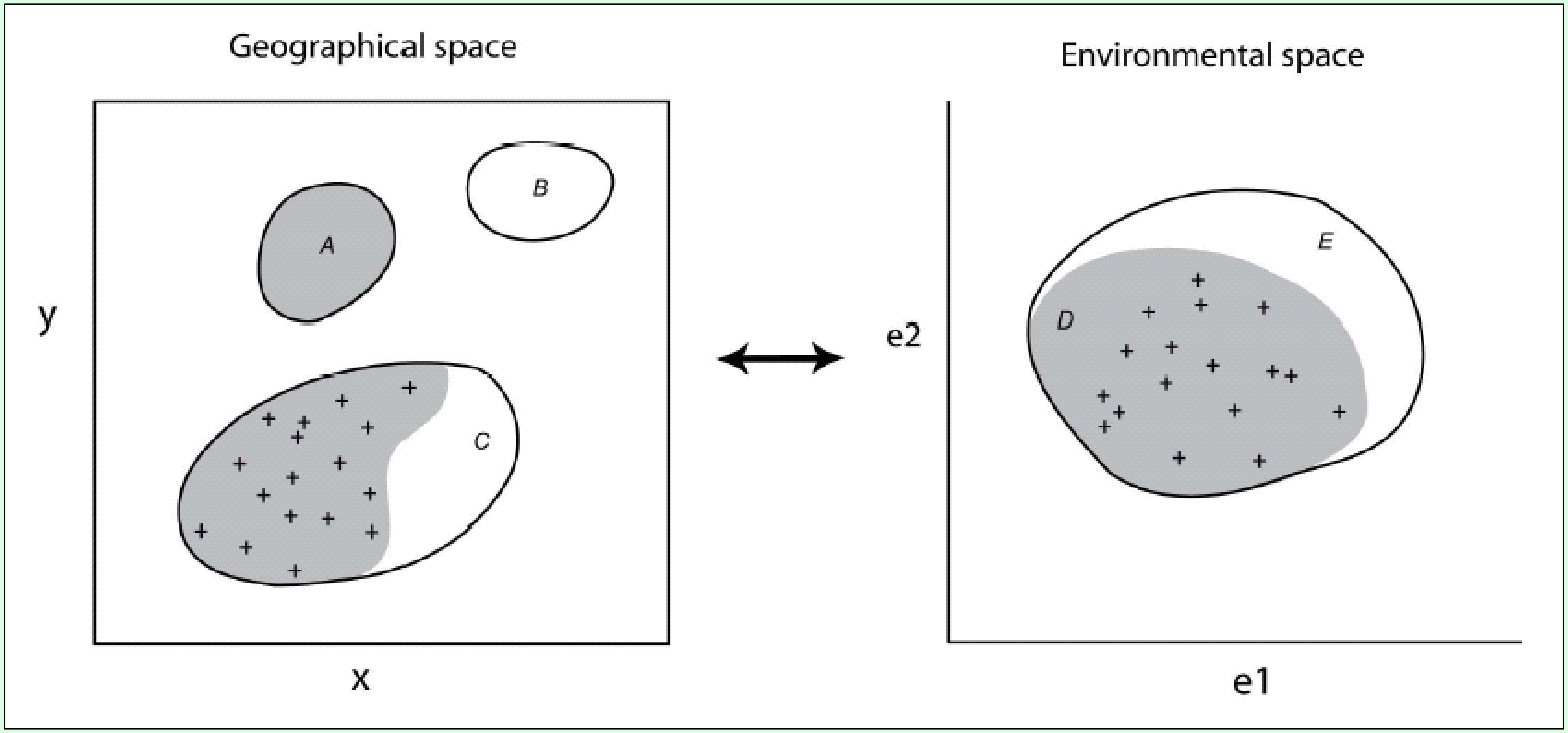
Intro to Ecological Niche Models with MaxEnt

Eugenio Valderrama ev243@cornell.edu



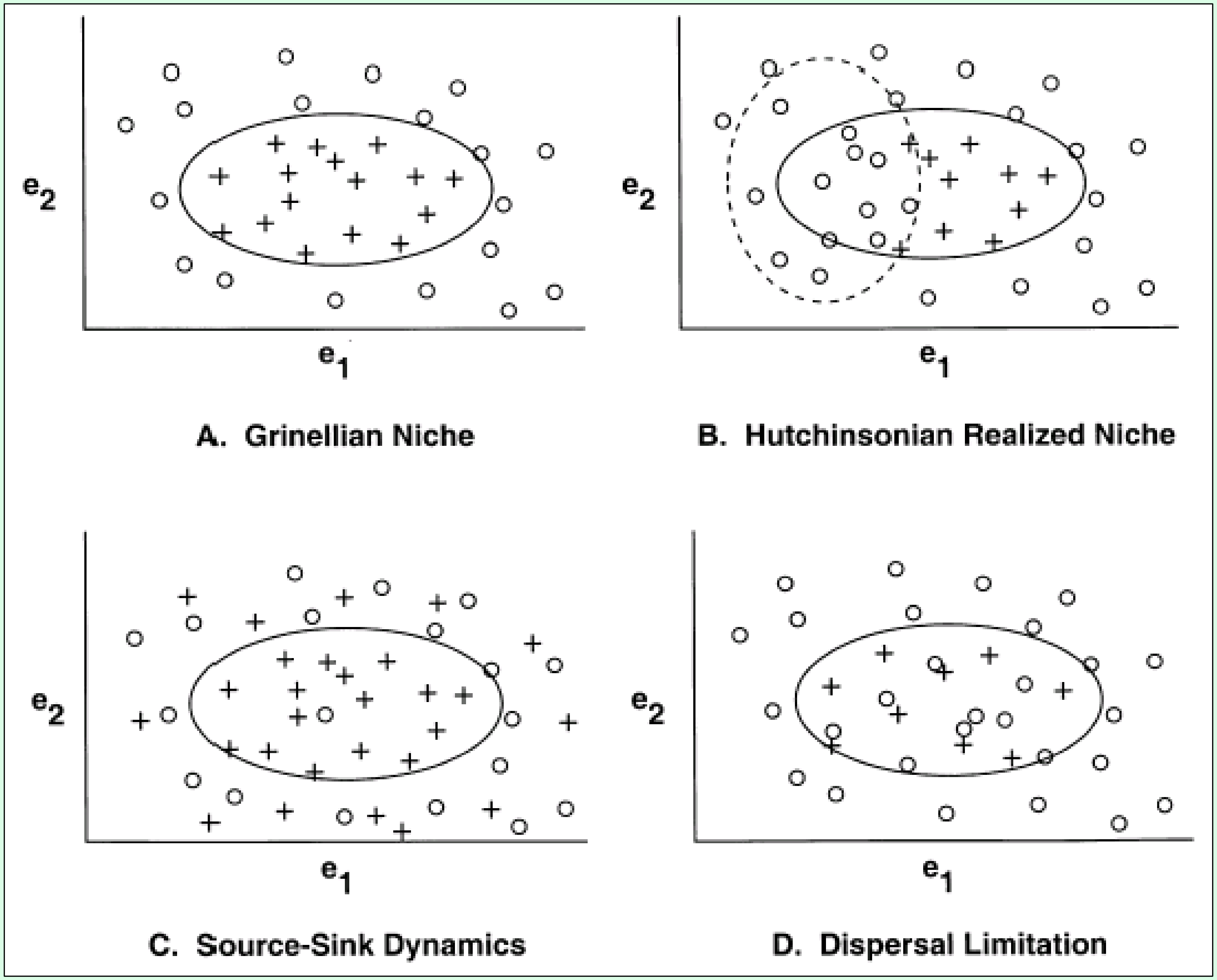
Rossetto Marcelino & Verbruggen 2015

G-space and E-space



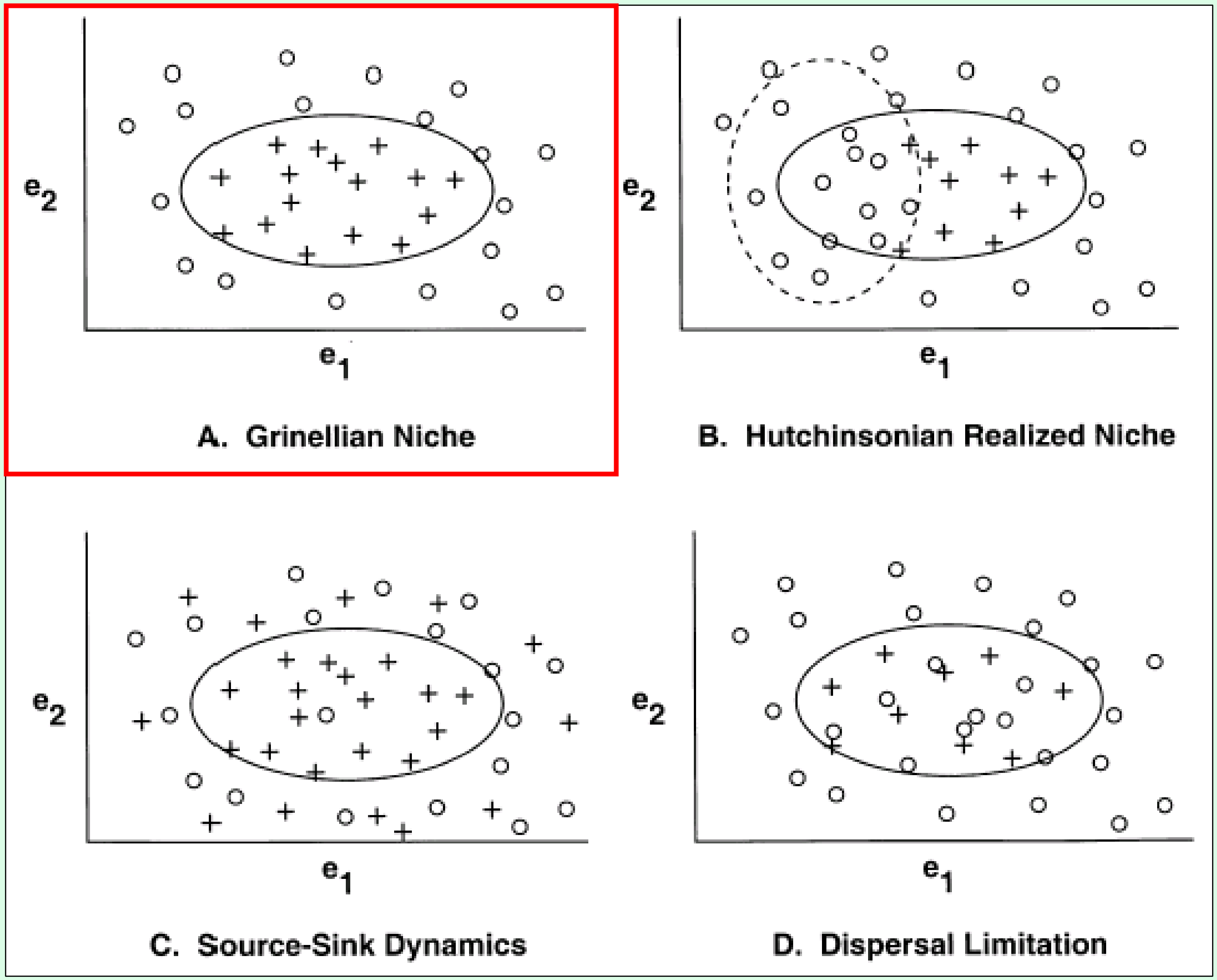
(Pearson, 2008)

Niche theory



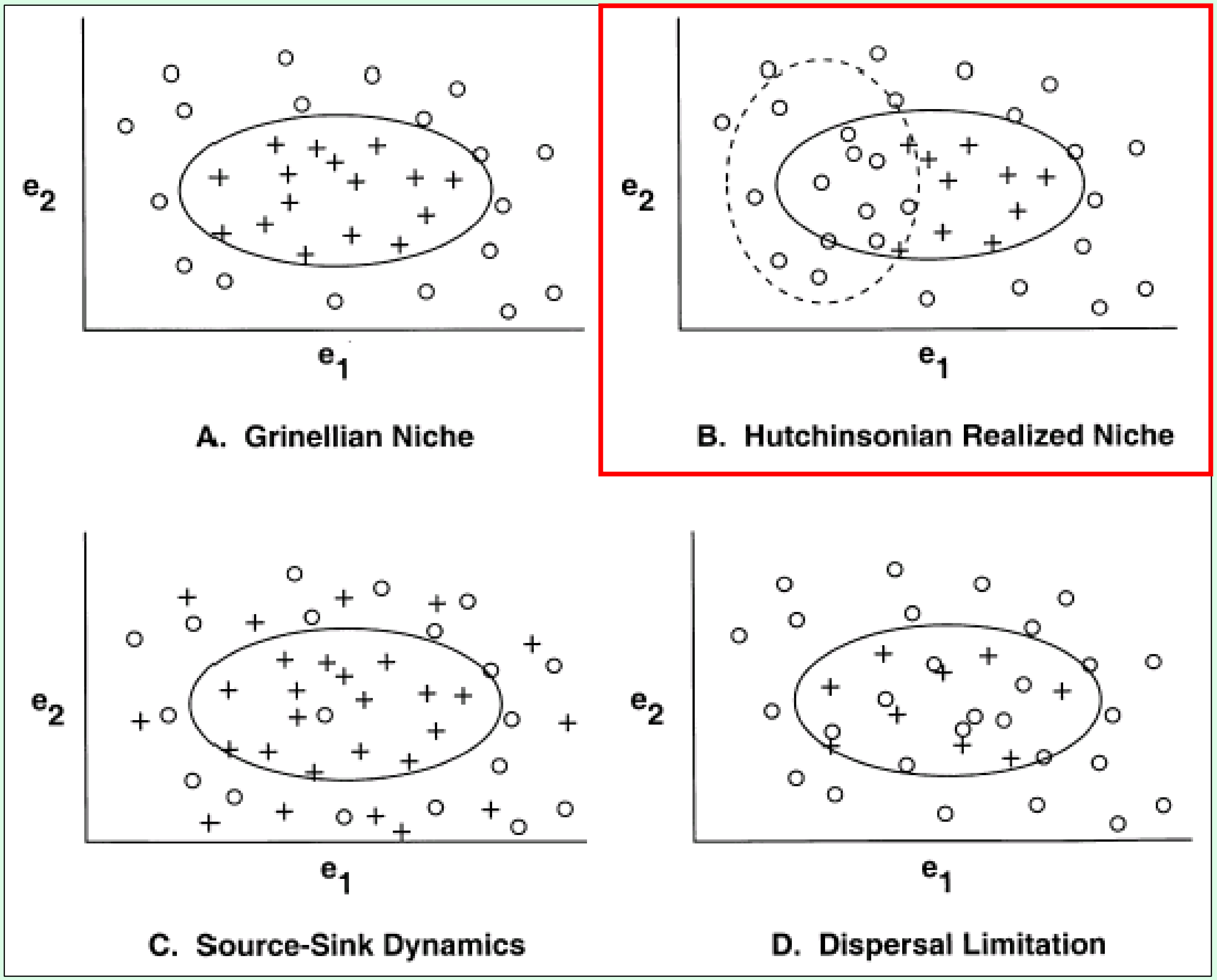
(Pulliam, 2000)

Niche theory



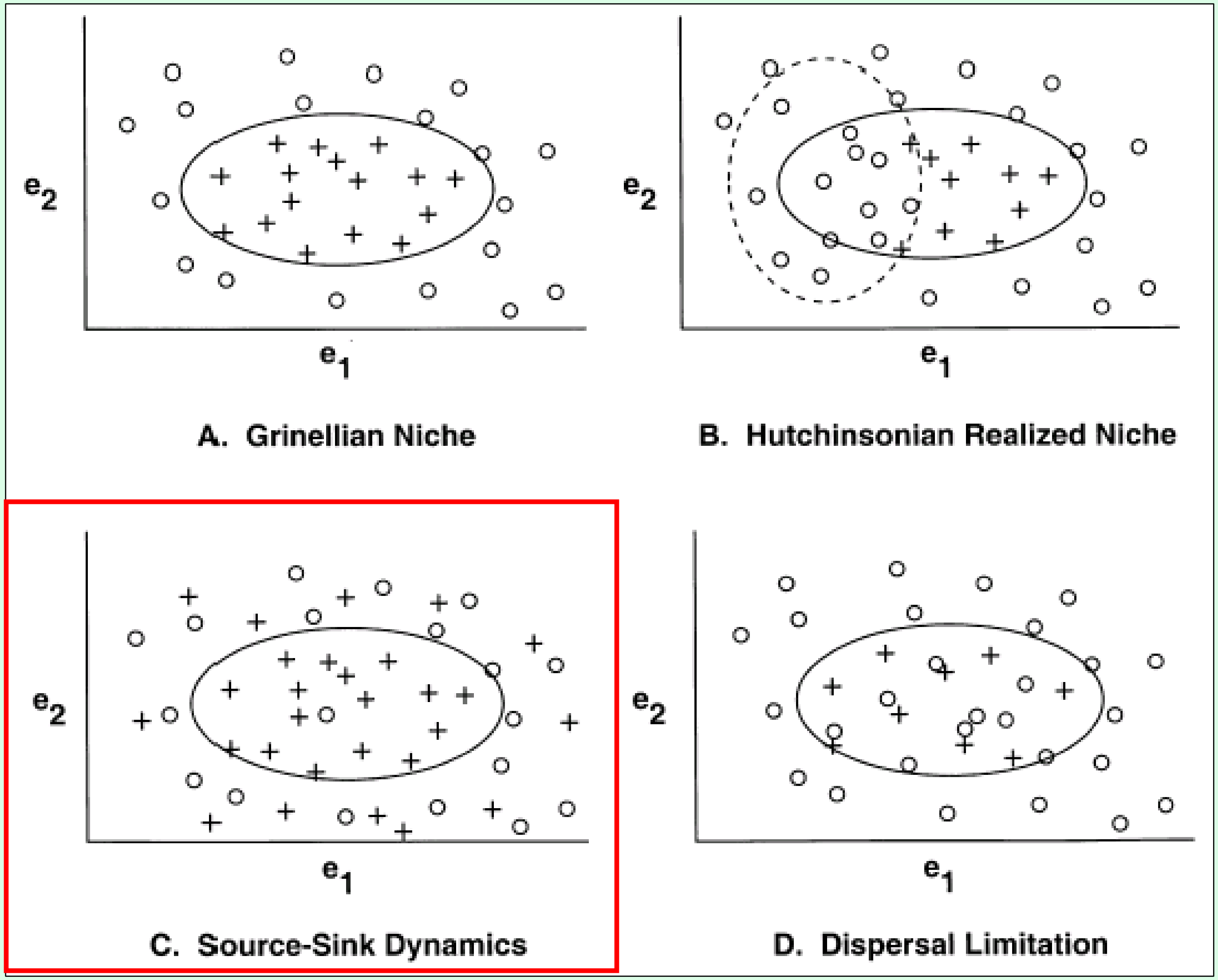
(Pulliam, 2000)

Niche theory



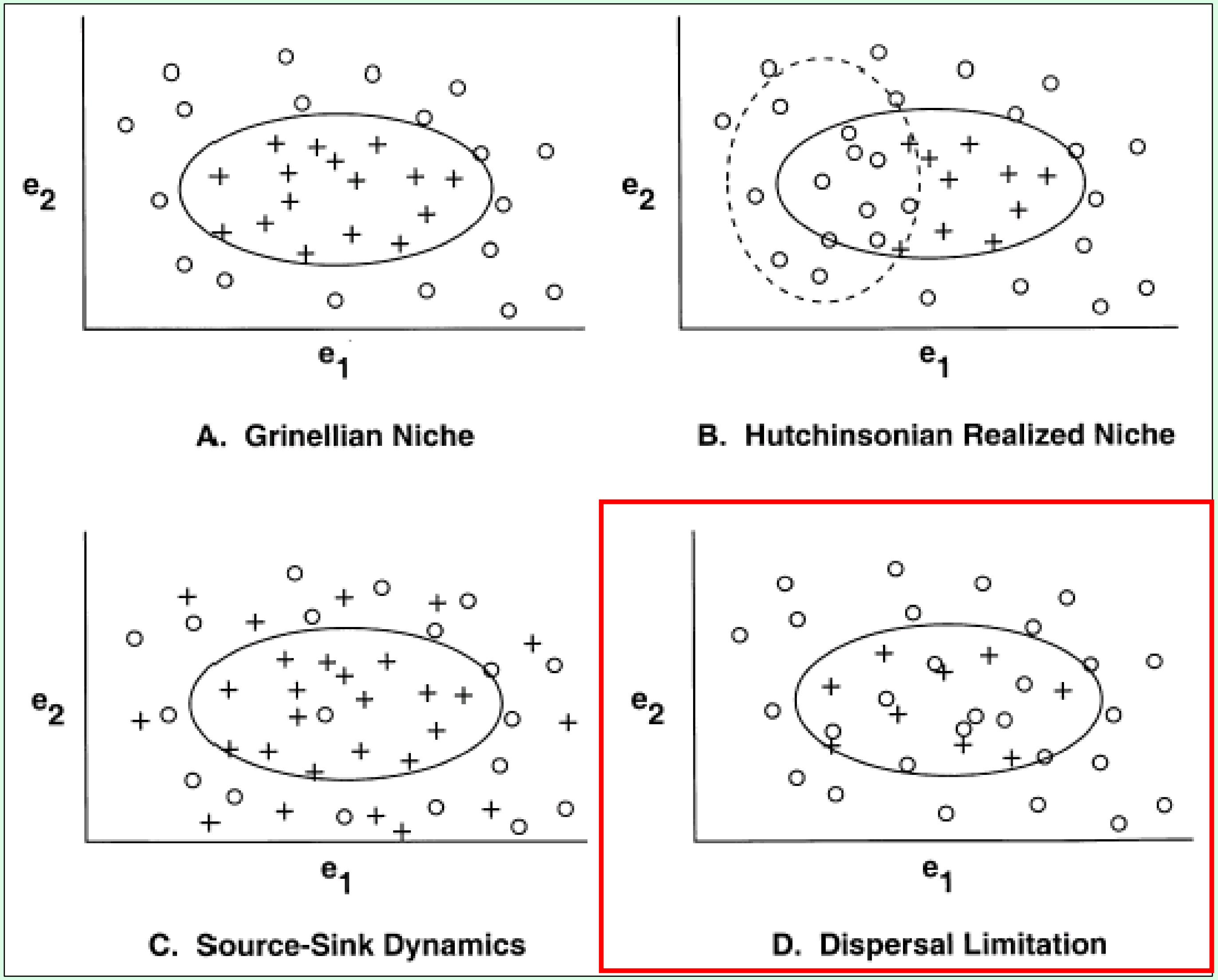
(Pulliam, 2000)

Niche theory



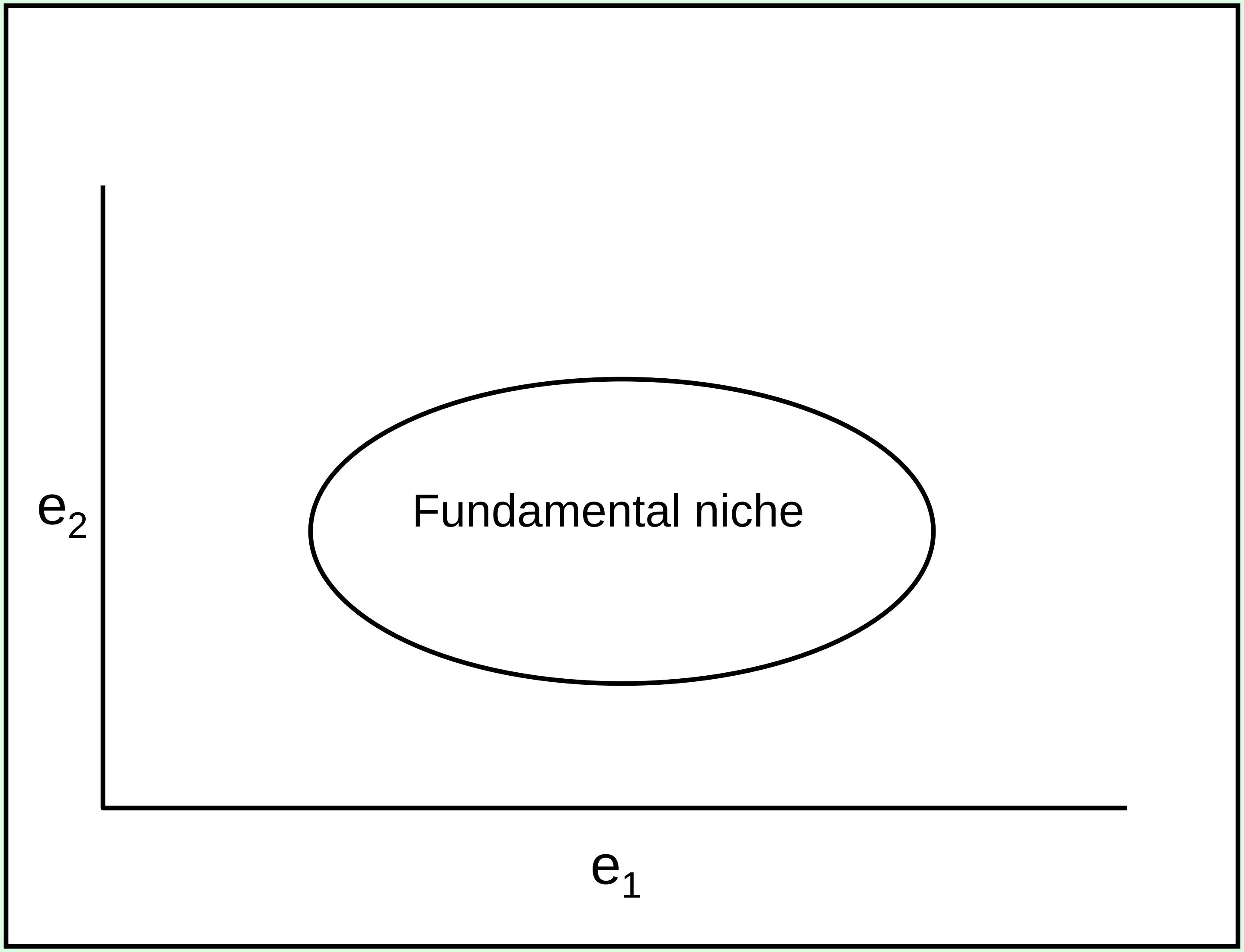
(Pulliam, 2000)

Niche theory



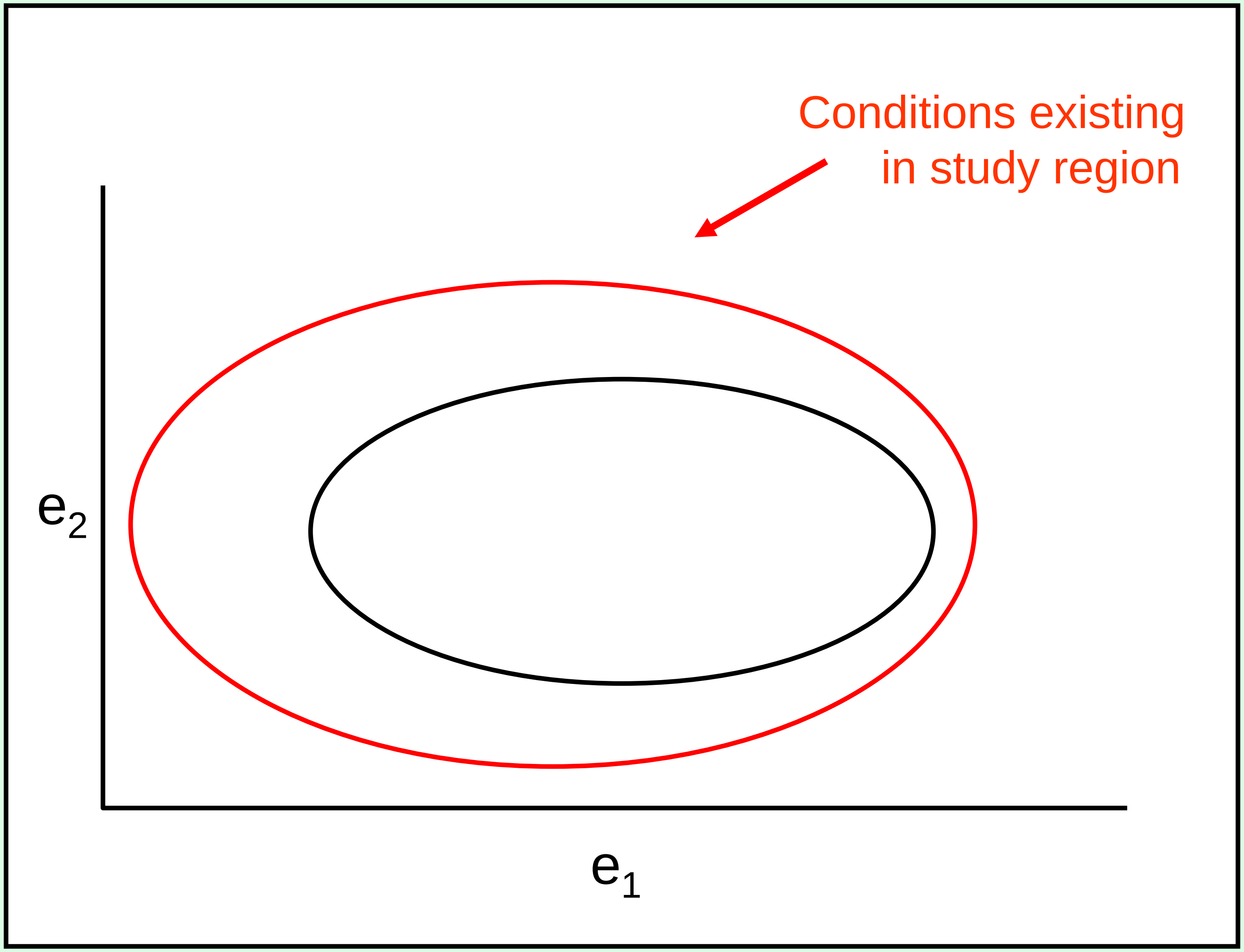
(Pulliam, 2000)

Niche theory



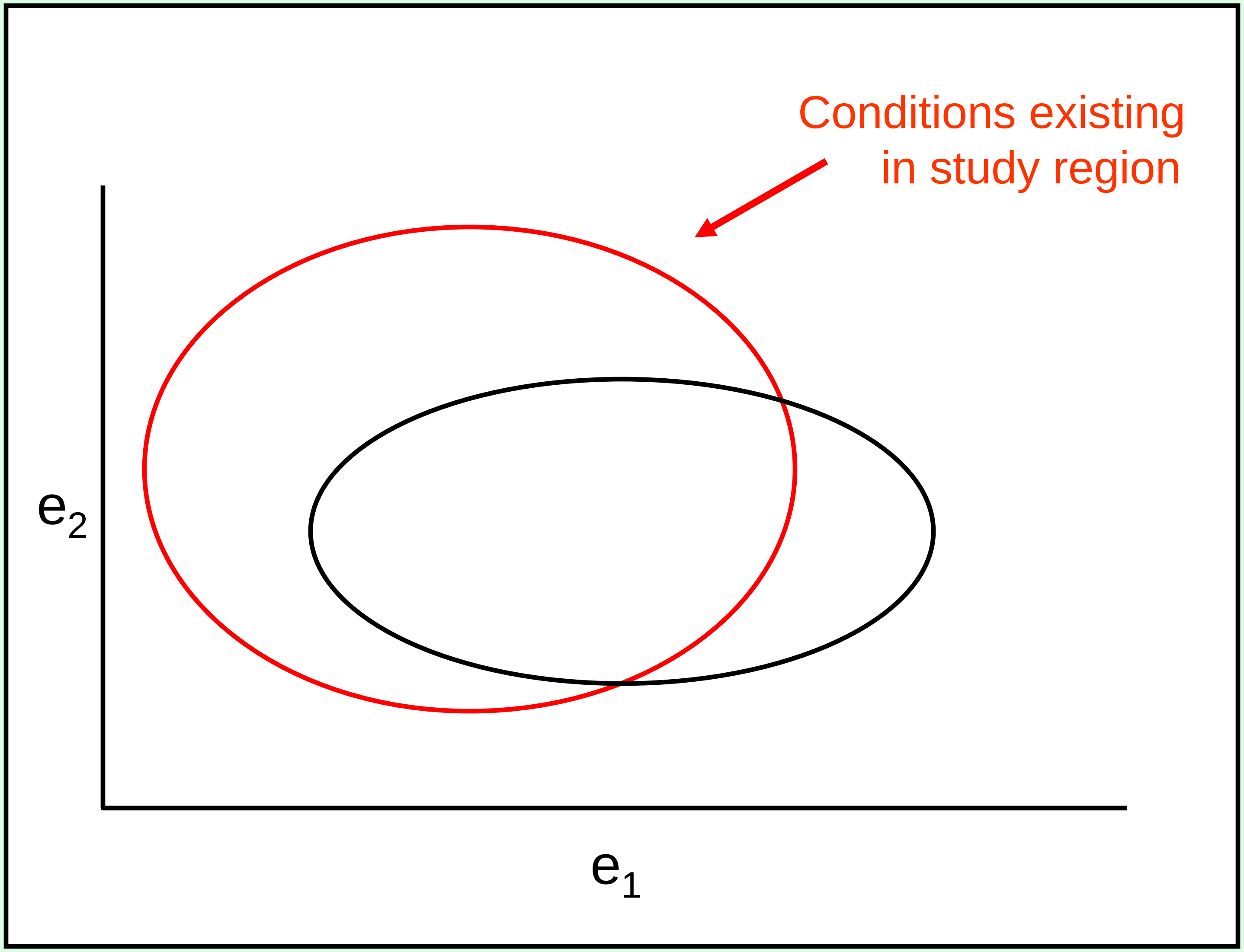
(Pulliam, 2000)

Niche theory



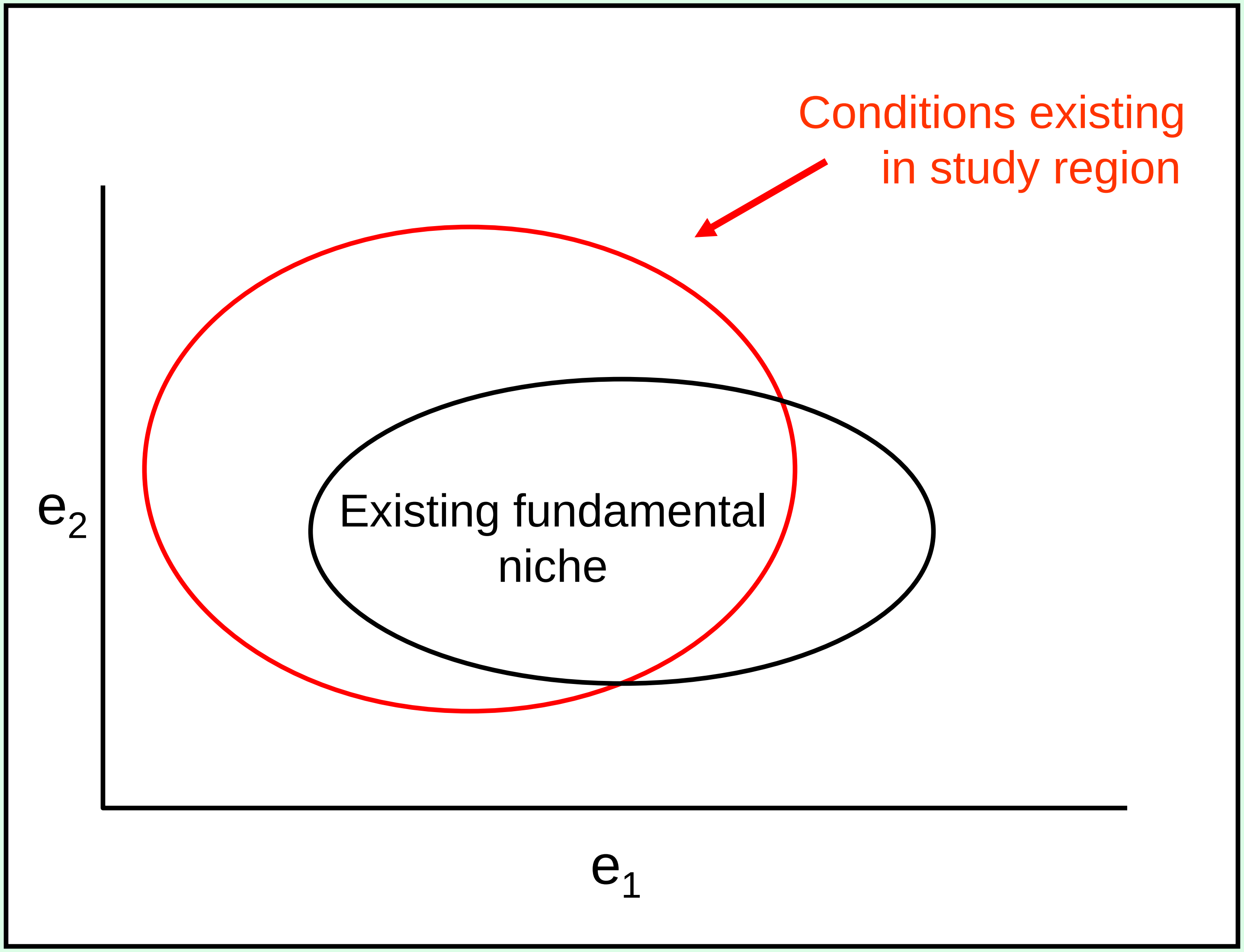
(Jackson and Overpeck, 2000)

Niche theory



(Jackson and Overpeck, 2000)

Niche theory



(Jackson and Overpeck, 2000; their “potential niche”)

Non-equilibrium distributions

Occupied distributions may be smaller than *abiotically suitable distributions* due to:

1. Contingent factors

   a. lack of dispersal

   b. local extinction

Non-equilibrium distributions

Occupied distributions may be smaller than *abiotically suitable distributions* due to:

1. Contingent factors

   a. lack of dispersal

   b. local extinction

2. Biotic interactions (e.g., competition)

Non-equilibrium distributions

Occupied distributions may be smaller than *abiotically suitable distributions* due to:

1. Contingent factors

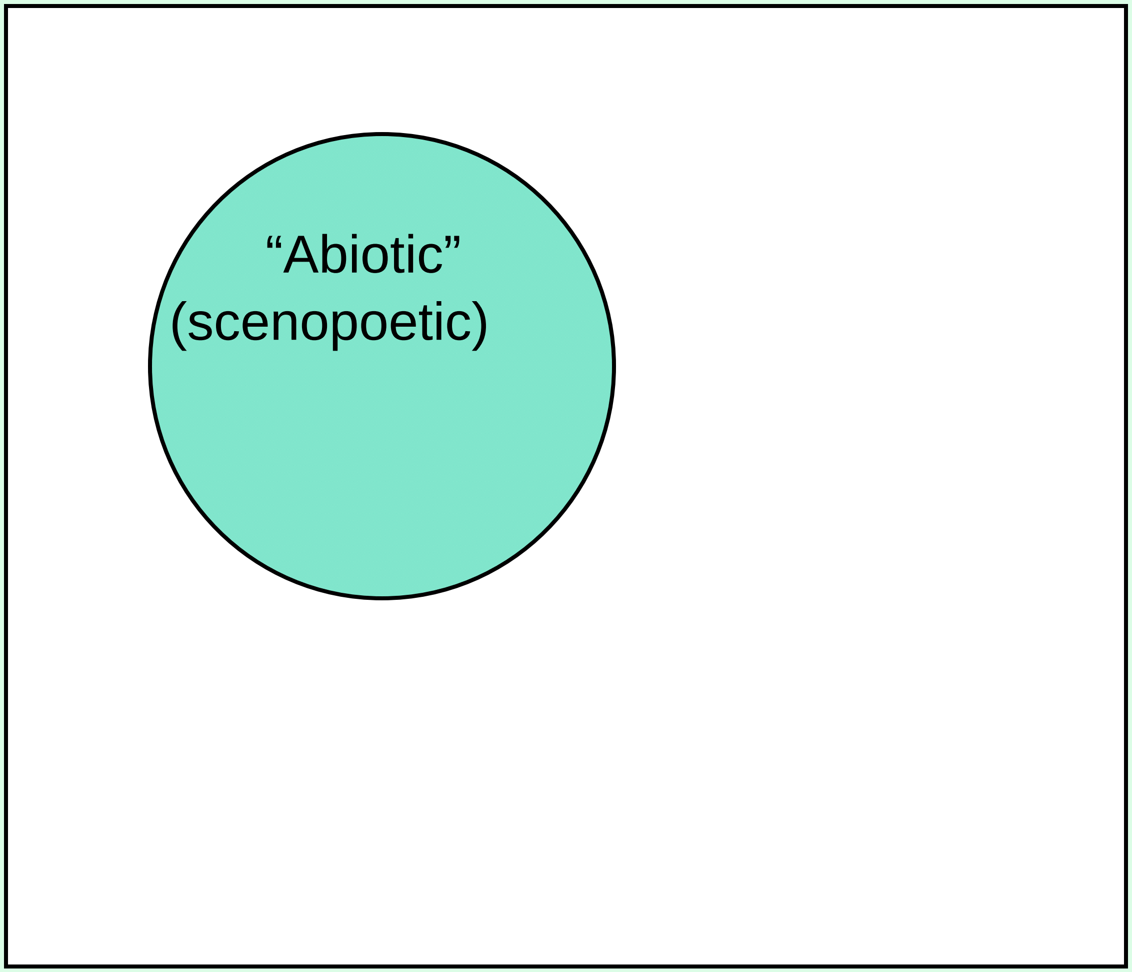
    a. lack of dispersal

    b. local extinction

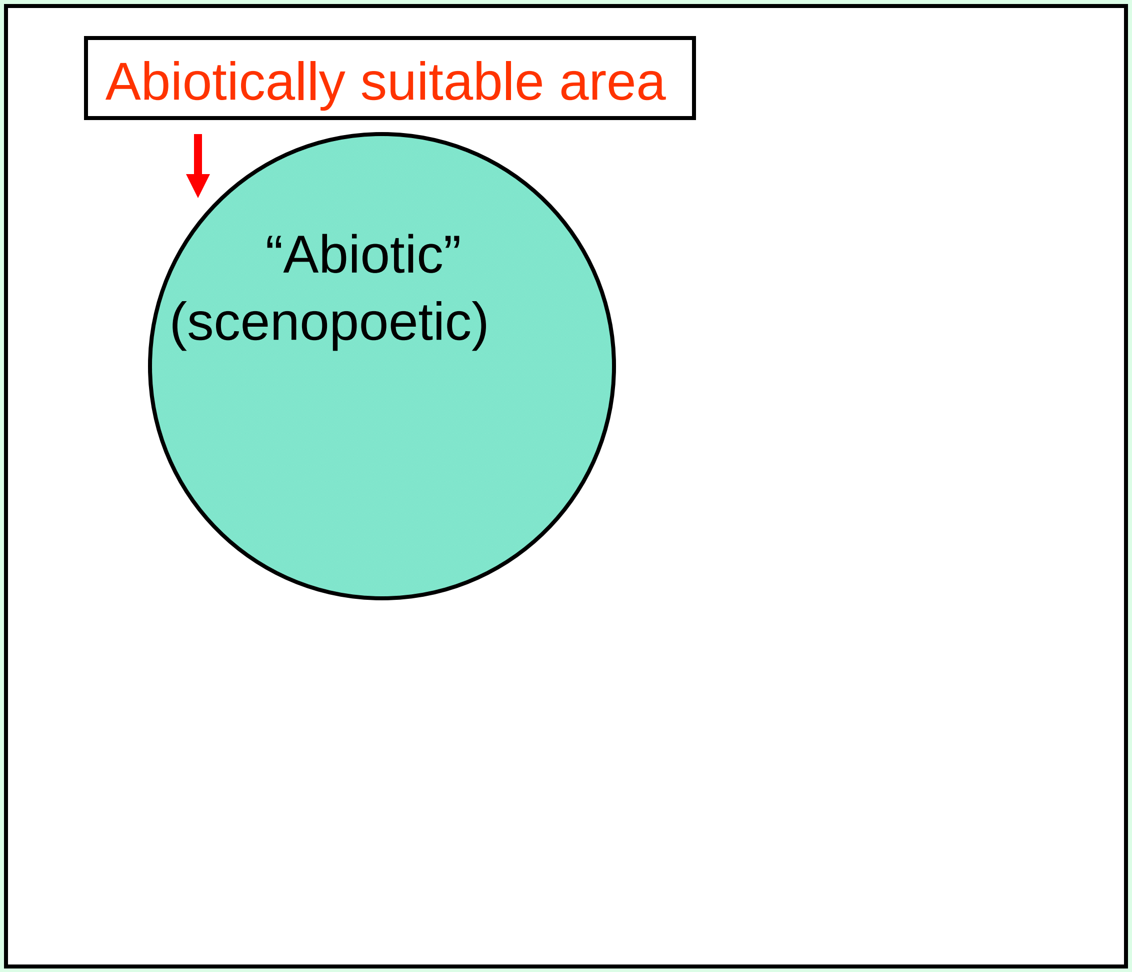
2. Biotic interactions (e.g., competition)

3. Human modifications of the landscape

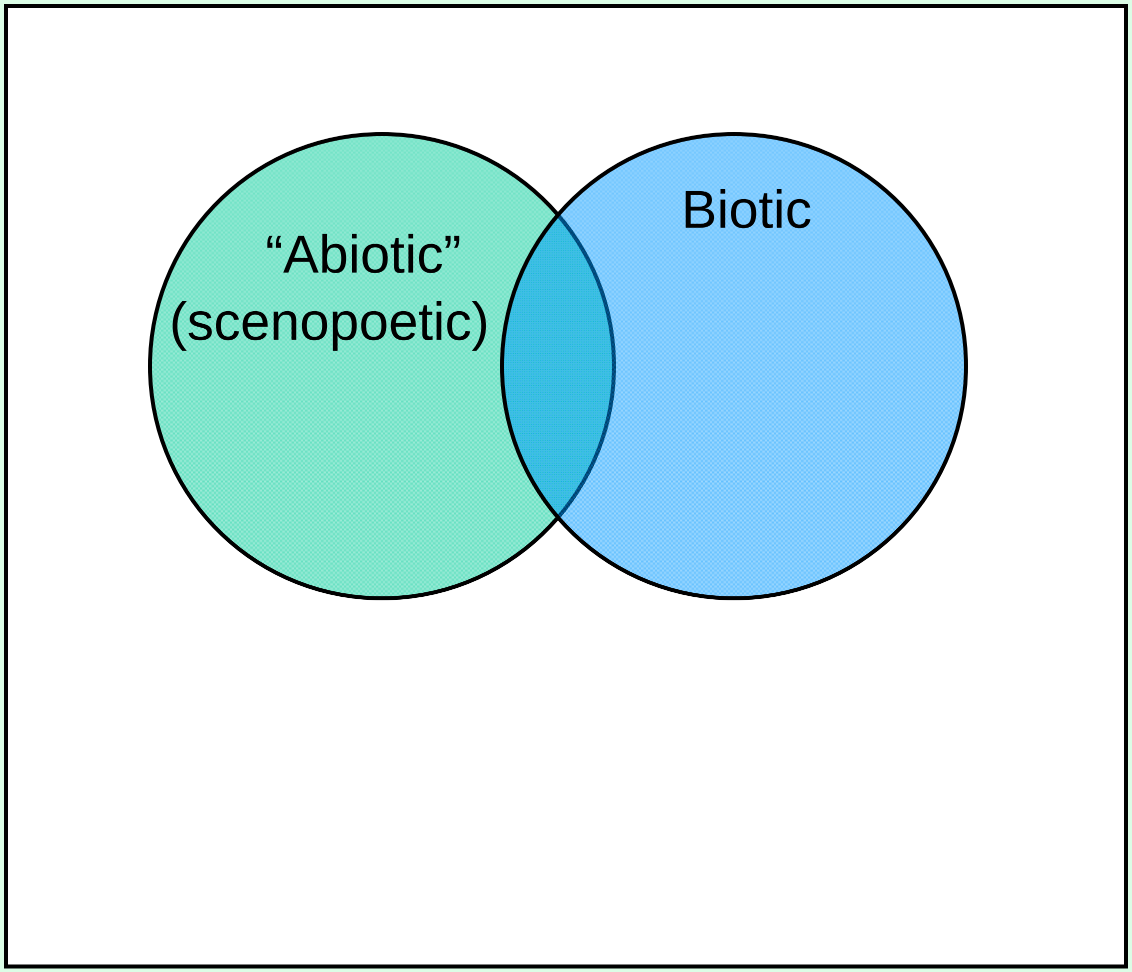
Non-equilibrium distributions



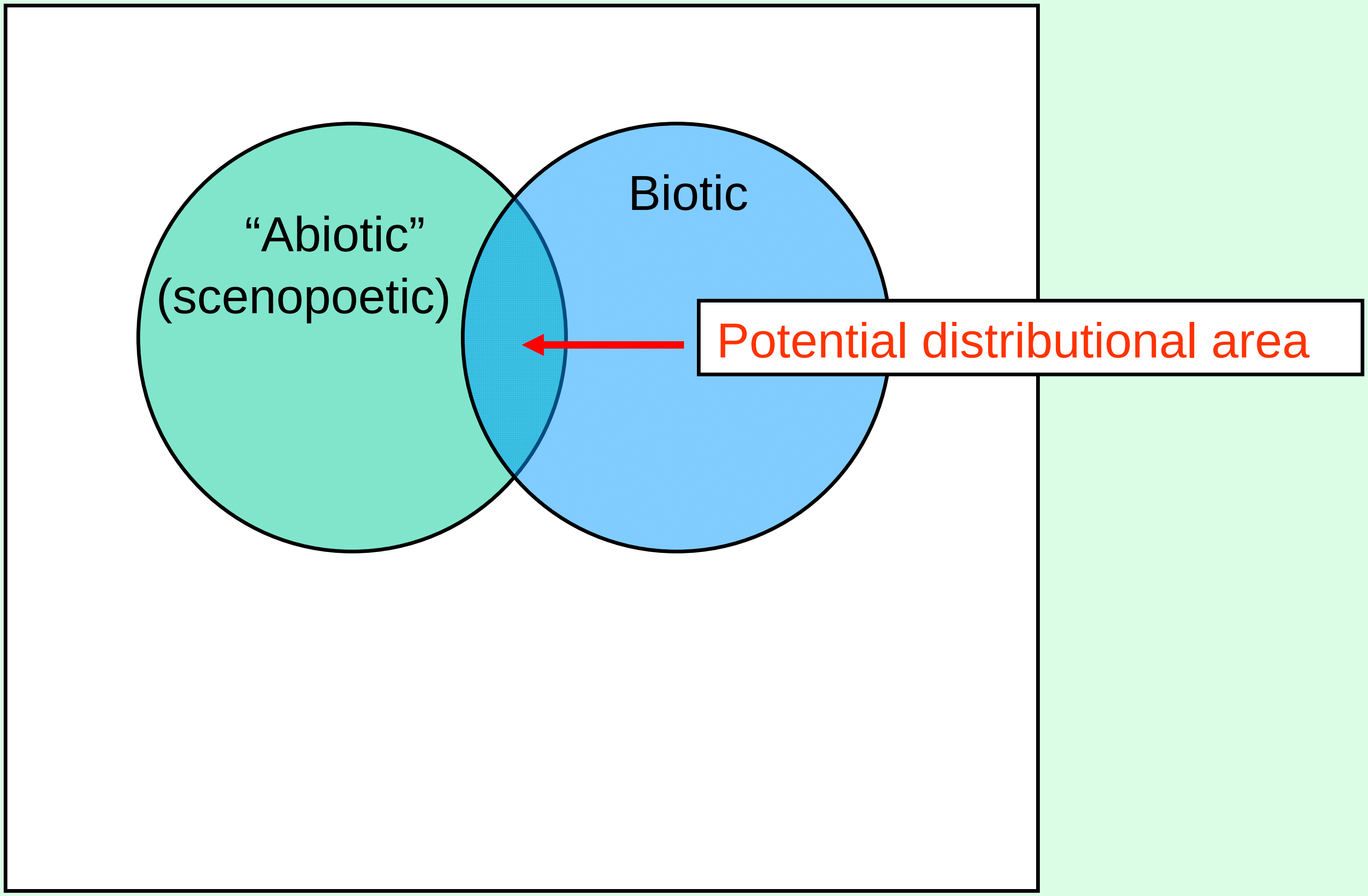
Non-equilibrium distributions



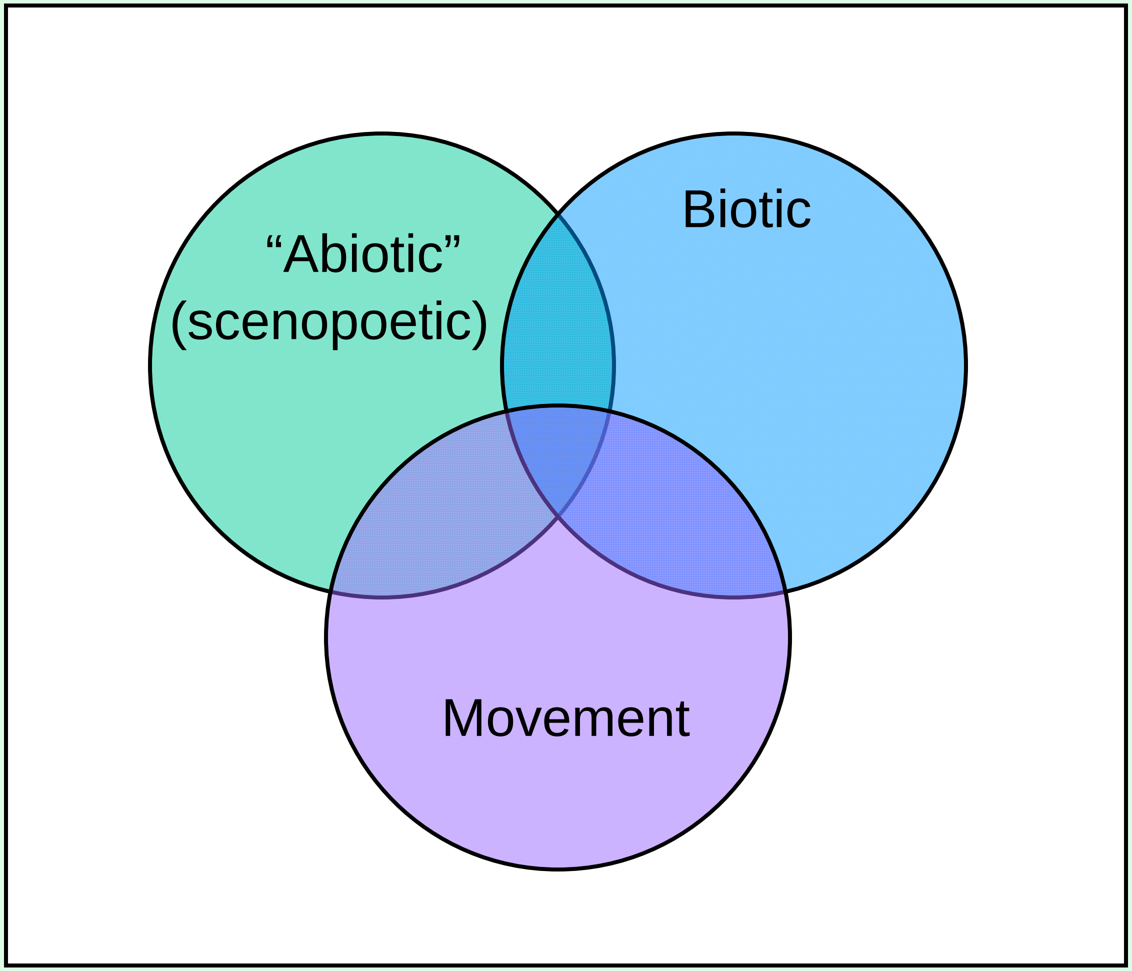
Non-equilibrium distributions



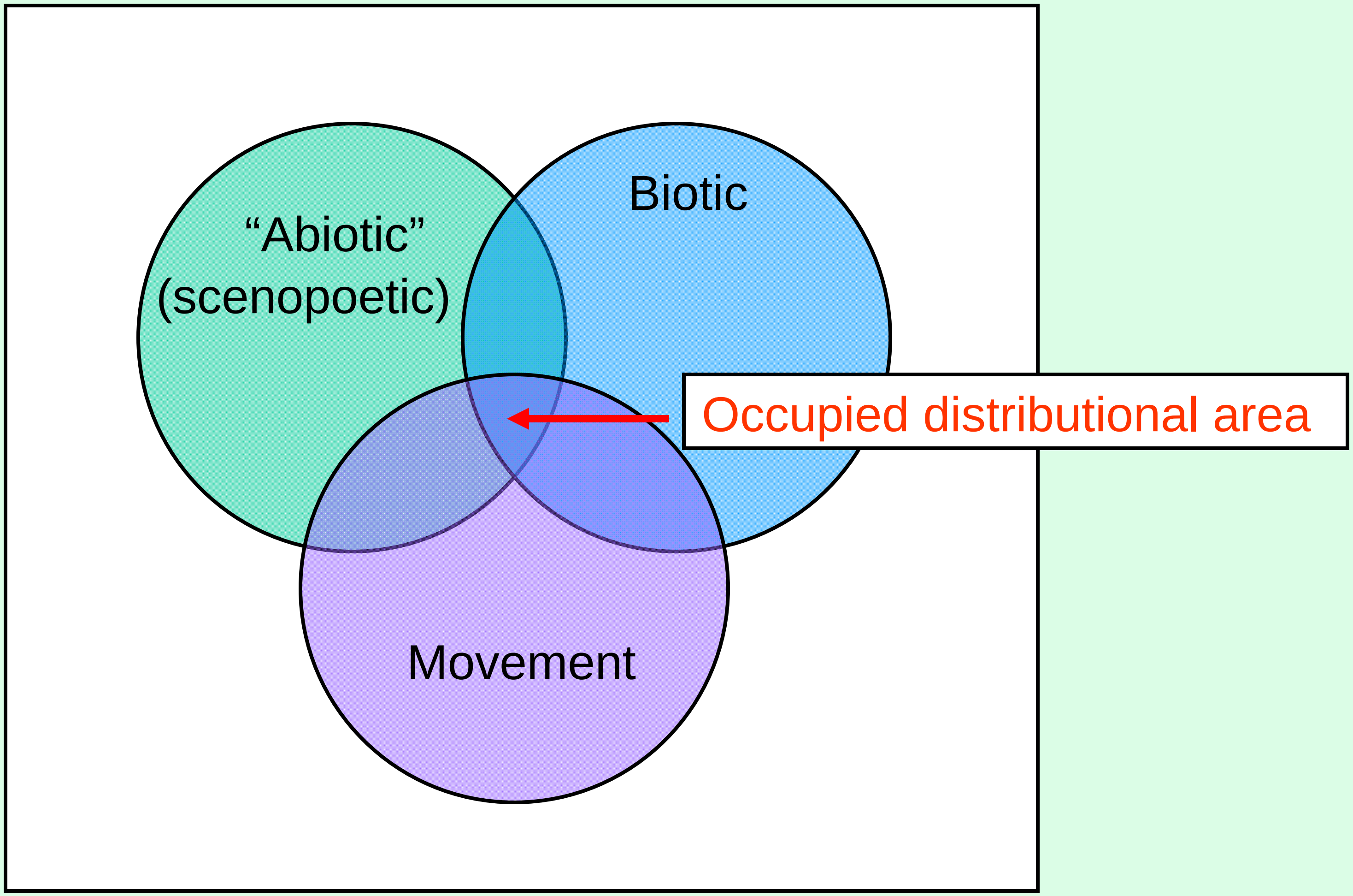
Non-equilibrium distributions



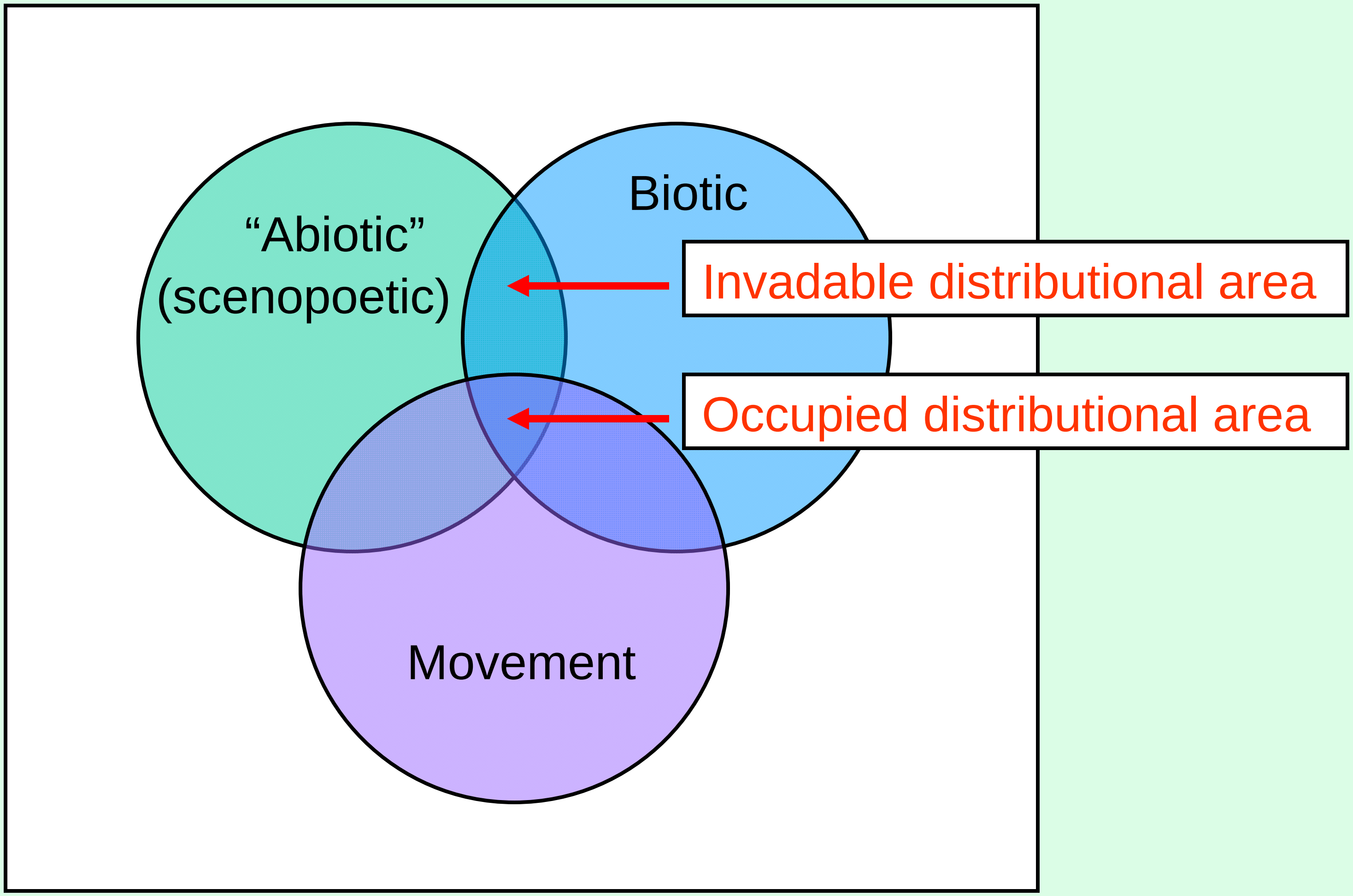
Non-equilibrium distributions



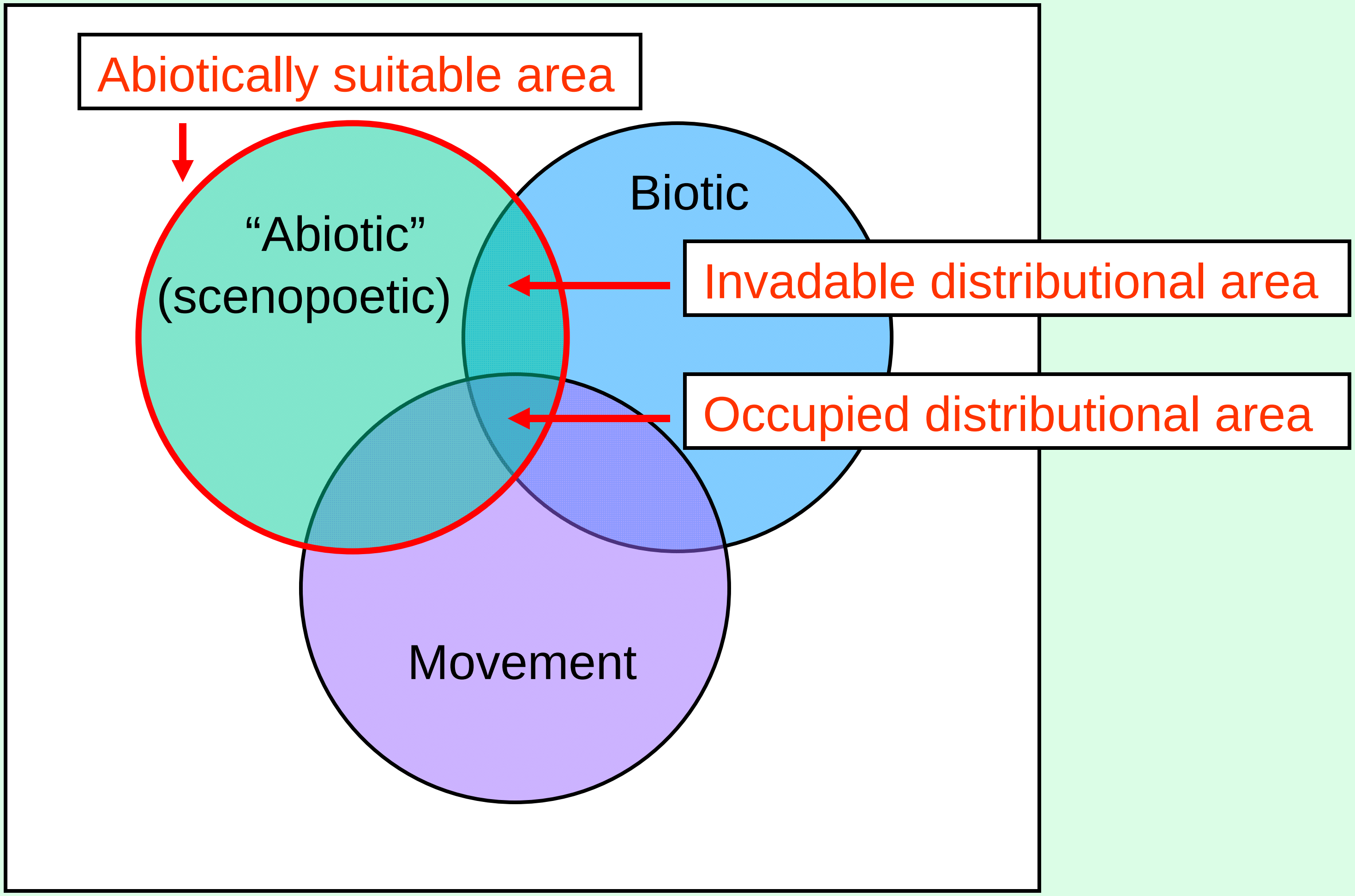
Non-equilibrium distributions



Non-equilibrium distributions



Non-equilibrium distributions



The Problem: Species Habitat Modeling

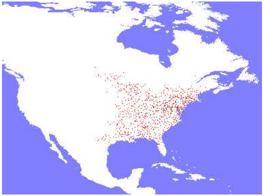
- goal: model distribution of plant or animal species

The Problem: Species Habitat Modeling

- goal: model distribution of plant or animal species

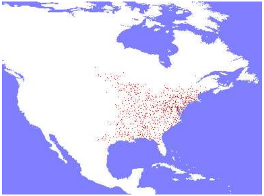
The Problem: Species Habitat Modeling

- goal: model distribution of plant or animal species
- given: presence records

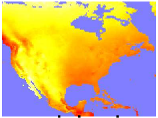


The Problem: Species Habitat Modeling

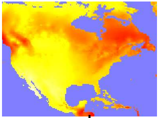
- goal: model distribution of plant or animal species
- given: presence records



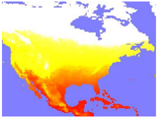
- given: environmental variables



precipitation



wet days

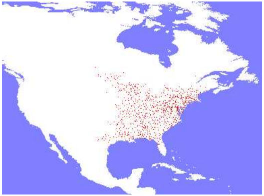


avg. temp.

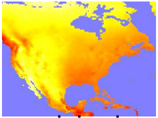
...

The Problem: Species Habitat Modeling

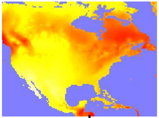
- goal: model distribution of plant or animal species
- given: presence records



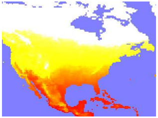
- given: environmental variables



precipitation



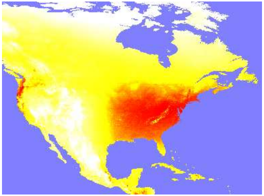
wet days



avg. temp.

. . .

- desired output: map of range



Biological Importance

- fundamental question: what are survival requirements (niche) of given species?
- core problem for conservation of species
- first step for many applications:
  - reserve design
  - impact of climate change
  - discovery of new species
  - clarification of taxonomic boundaries

A Challenge for Machine Learning

- no negative examples
- very limited data
  - often, only 20-100 presence records
  - usually, not systematically collected
  - may be museum records years or decades old
- sample bias
  - (toward most accessible locations)

Our Approach

- assume presence records come from probability distribution π
- try to estimate π
- apply maximum entropy approach

Maximum Entropy (Maxent) method

Estimates *target probability distribution*

by finding probability distribution (statistical model) of *maximum entropy* (i.e., most spread out, closest to uniform)

subject to *constraints*

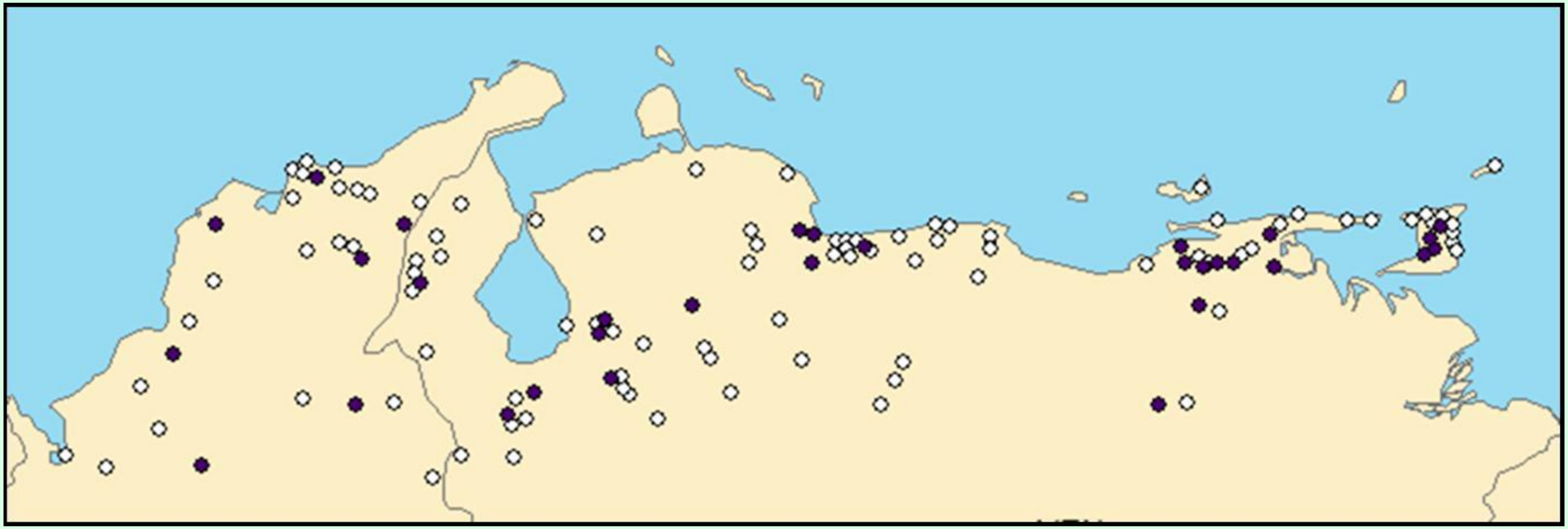
Maximum Entropy (Maxent) method

Constraints: what we know about the features

Data from the sample points (the known occurrence localities, in our case)

Random subsets

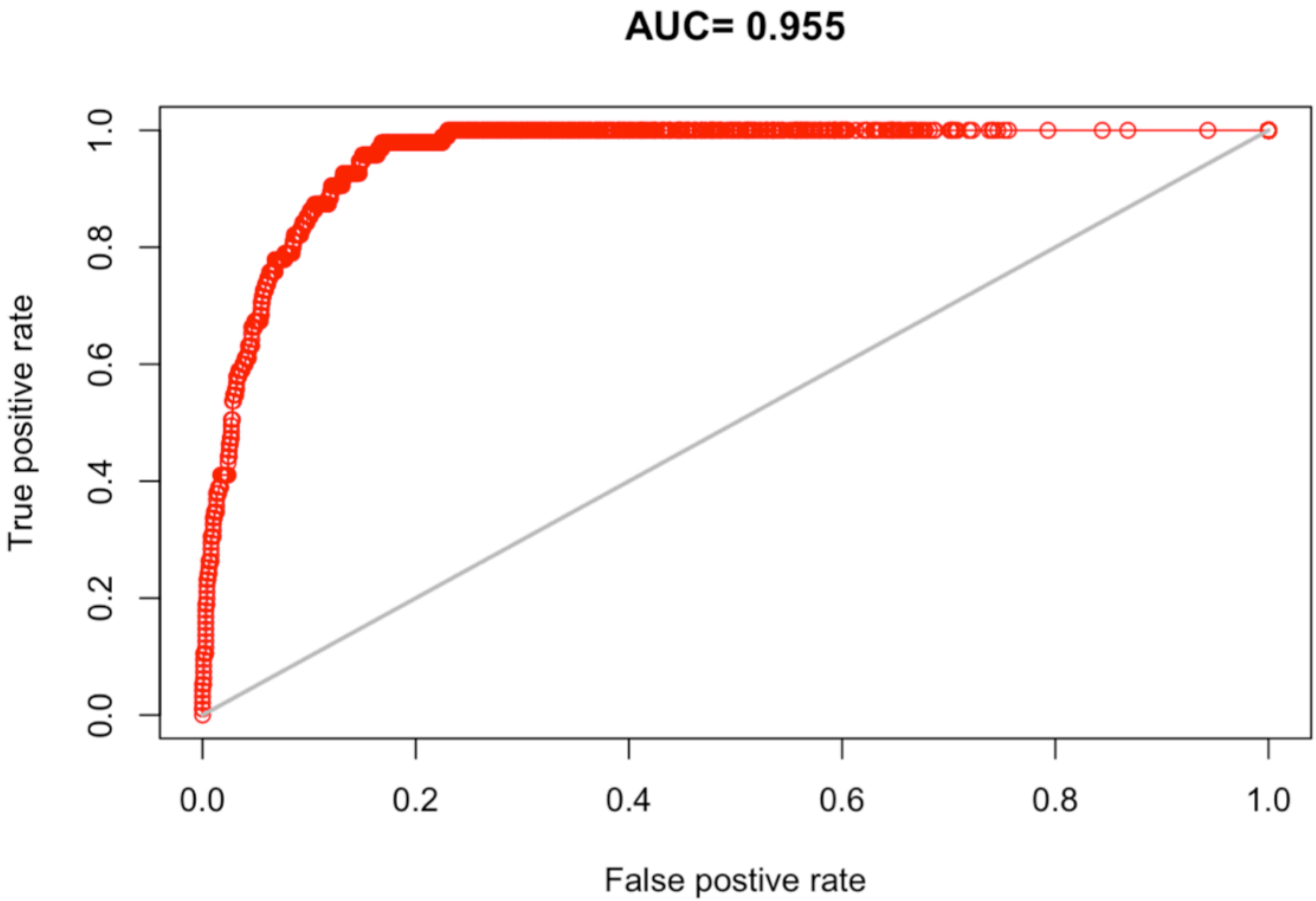
random split-sample approach: easy test, cannot detect overfitting to bias



White: calibrate the model

Black: evaluate the model

Receiver operating characteristic curve



Environmental variables

Effect of species on predictor variable (Hutchinson 1978)

Scenopoetic	Not affected by presence of focal species
Non-scenopoetic	Affected by presence of focal species

Environmental variables

Relative degree of causality (Austin 2002)

Proximal	Determines the organism’s response
Distal	Linked to proximal variable that determines response

Environmental variables

Physiological effect on species (Austin 2002)

Indirect	Does not affect focal species physiologically; correlated with distribution via correlations with other factors
Direct	Affects focal species physiologically *but is not* consumed by it
Resource	Affects focal species physiologically *and is* consumed by it

Environmental variables: examples

Effect of species on predictor variable (Hutchinson 1978)

Scenopoetic	Temperature or precipitation
Non-scenopoetic	Water or nutrients consumed by a plant

Environmental variables: examples

Relative degree of causality (Austin 2002)

Proximal	Available soluble [P] at root hair;
Distal	Total soil phosphate;

Environmental variables: examples

Relative degree of causality (Austin 2002)

Proximal	Available soluble [P] at root hair;
	Freeze durations that affect survival of cacti along poleward range margin
Distal	Total soil phosphate;
	Mean temperature of coldest month, or annual mean temperature (relatively more distal than the former)

Environmental variables: examples

Physiological effect on species (Austin 2002)

Indirect	Elevation; latitude or longitude
Direct	Temperature; pH
Resource	Water or nutrients in soil

Environmental variables: guidelines

Effect of species on predictor variable (Hutchinson 1978)

Scenopoetic	Not affected by presence of focal species USE!
Non-scenopoetic	Affected by presence of focal species AVOID!

Environmental variables: guidelines

Relative degree of causality (Austin 2002)

Proximal	Determines the organism’s response
	USE!
Distal	

Environmental variables: guidelines

Relative degree of causality (Austin 2002)

Proximal	Determines the organism's response
	USE!
Distal	Linked to proximal variable that determines response
	MAYBE: use for transfer only if the correlation with the driving variable is likely to hold across space/time
	(see also indirect and direct)

Environmental variables: guidelines

Physiological effect on species (Austin 2002)

Indirect	
Direct	Affects focal species physiologically but not consumed by it: USE!
Resource	Affects focal species physiologically and is consumed by it: MAYBE: Use if scenopoetic

Environmental variables: guidelines

Physiological effect on species (Austin 2002)

Indirect	Does not affect focal species physiologically; correlated with distribution via correlations with other factors: MAYBE (depends on correl.)
Direct	Affects focal species physiologically but not consumed by it: USE!
Resource	Affects focal species physiologically and is consumed by it: MAYBE: Use if scenopoetic

Environmental variables: guidelines

Physiological effect on species (Austin 2002)

Indirect

AVOID if correlated with distribution because of associations with factors related to dispersal/demography or with the distributions of important biotic interactors;

Environmental variables: guidelines

Physiological effect on species (Austin 2002)

Indirect

AVOID if correlated with distribution because of associations with factors related to dispersal/demography or with the distributions of important biotic interactors;

USE if correlated with distribution because of correlations with driving abiotic variables; use for transfer only if the correlation with the driving variable is likely to hold across space/time

Many of the slides were taken from presentations by Robert P. Anderson & Robert Schapire