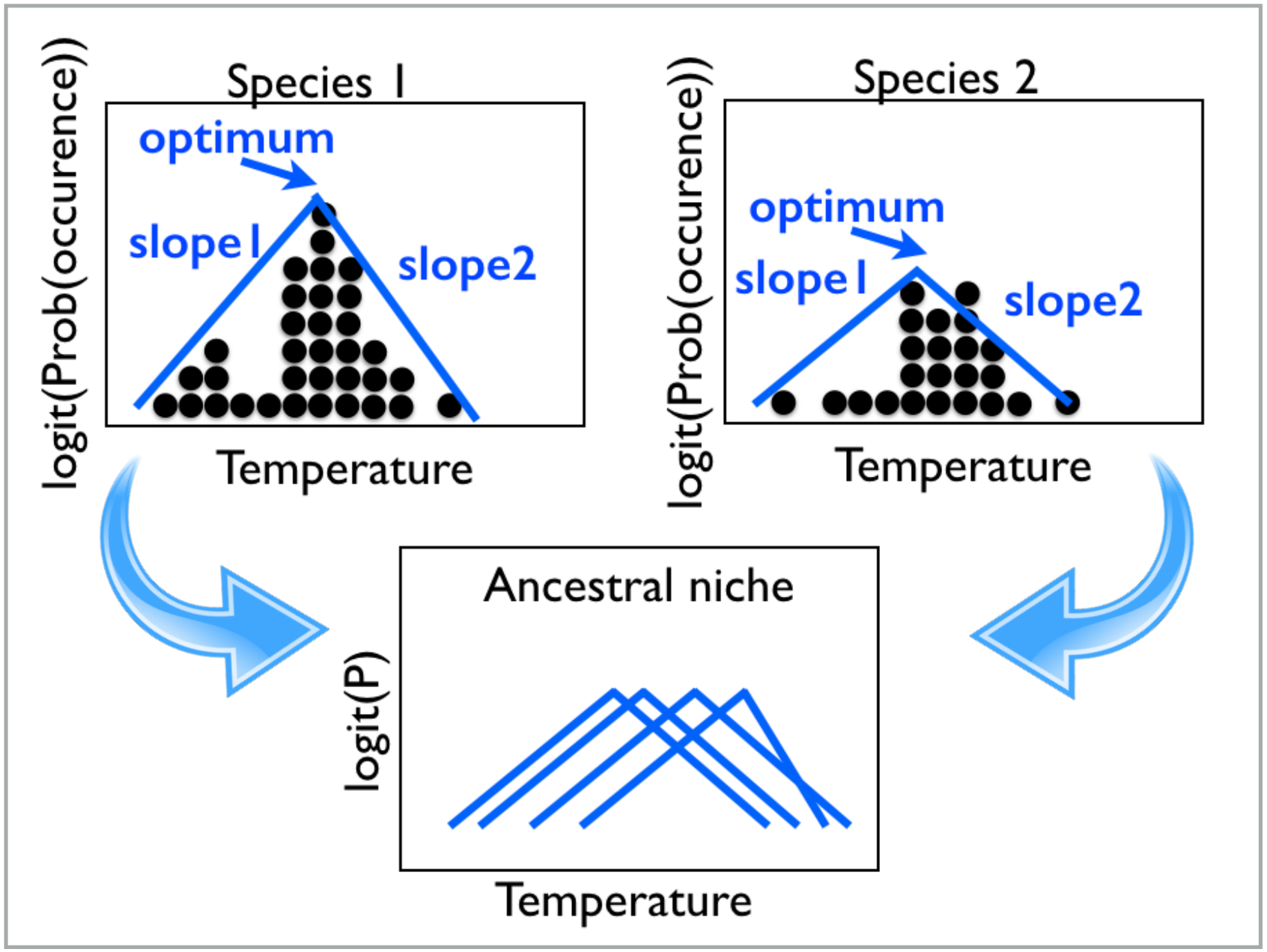
Putting Evolution Into Ecological Niche Modelling: Background and Paths Forward



Nicholas J. Matzke
DECRA Fellow
Division of Ecology, Evolution, and Genetics
Research School of Biology
The Australian National University
phylo.wikidot.com/nicholas-j-matzke

Evolution 2016
Sunday, June 19, 8:15 - 11:45 am, Ballroom C

Thanks for support from:

SSB Symposium: Putting Evolution Into Ecological Niche Modelling

1. **Nick Matzke —** Introduction to Symposium; and phyloSDM with joint Bayesian estimation
2. **Dan Warren —** Incorporating evolutionary information into niche and distribution models: learning to love terrible SDMs
3. **Erin Saupe —** A Tale Of Two Sisters: Ecological & Evolutionary Divergence In Geminate Species Pairs
4. **Kaitlin Maguire —** The fossil record and projecting species distribution models through time
5. **Jenny McGuire —** Evolution and epigenetics: unraveling error in niche models
6. **COFFEE BREAK**
7. **Renee Catullo —** Extending spatial modelling of climate change responses beyond the realized niche: estimating, and accommodating, physiological limits and adaptive evolution
8. **Matt Van Dam —** Environmental Niche Trackers and Nice Adapters Revealed Through Fine Scale Phenological Niche Modeling
9. **Dan Rosauer —** Distribution models for intraspecific lineages – which method is best?
10. **Laura Pollock —** Protecting evolutionary history into the future
11. **All presenters —** Panel discussion and further questions

<u>Outline</u>:

1. **Ecological Niche Modelling/ Species Distribution Modelling methods**
   - and common critiques

2. **Phylogenetic Comparative Methods**
   - and common critiques

3. **Phylogenetic Niche Conservatism methods**
   - and less common critiques

4. **Joint Bayesian modelling: a way forward?**

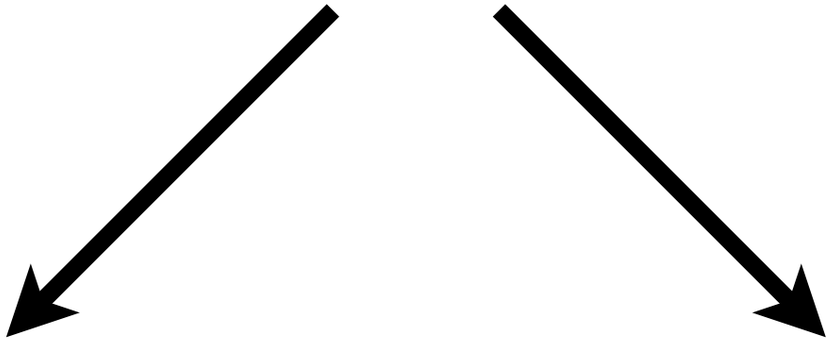
<u>Definitions</u>:

ENM: Ecological Niche Modelling

SDM: Species Distribution Modelling

PCMs: Phylogenetic Comparative Methods

Biogeography:
The geography of life



Ecological biogeography

Historical biogeography

Biogeography:
The geography of life

Ecological biogeography

- Focus: Environmental controls on distribution
- Ecological interactions
- Timescale: the Recent

Historical biogeography

Biogeography:
The geography of life

Ecological biogeography

- Focus: Environmental controls on distribution
- Ecological interactions
- Timescale: the Recent
- **Methods revolution: SDM/ENM**

Historical biogeography

Biogeography:
The geography of life

Ecological biogeography

- Focus: Environmental controls on distribution
- Ecological interactions
- Timescale: the Recent
- **Methods revolution: SDM/ENM**

Historical biogeography

- Focus: Evolutionary history of distributions
- Dispersal/vicariance events
- Timescale: Millions of years

Biogeography:
The geography of life

Ecological biogeography

- Focus: Environmental controls on distribution
- Ecological interactions
- Timescale: the Recent
- **Methods revolution: SDM/ENM**

Historical biogeography

- Focus: Evolutionary history of distributions
- Dispersal/vicariance events
- Timescale: Millions of years
- **Methods revolution: Phylogenetics, PCMs**

Phylogenetics in the scientific literature

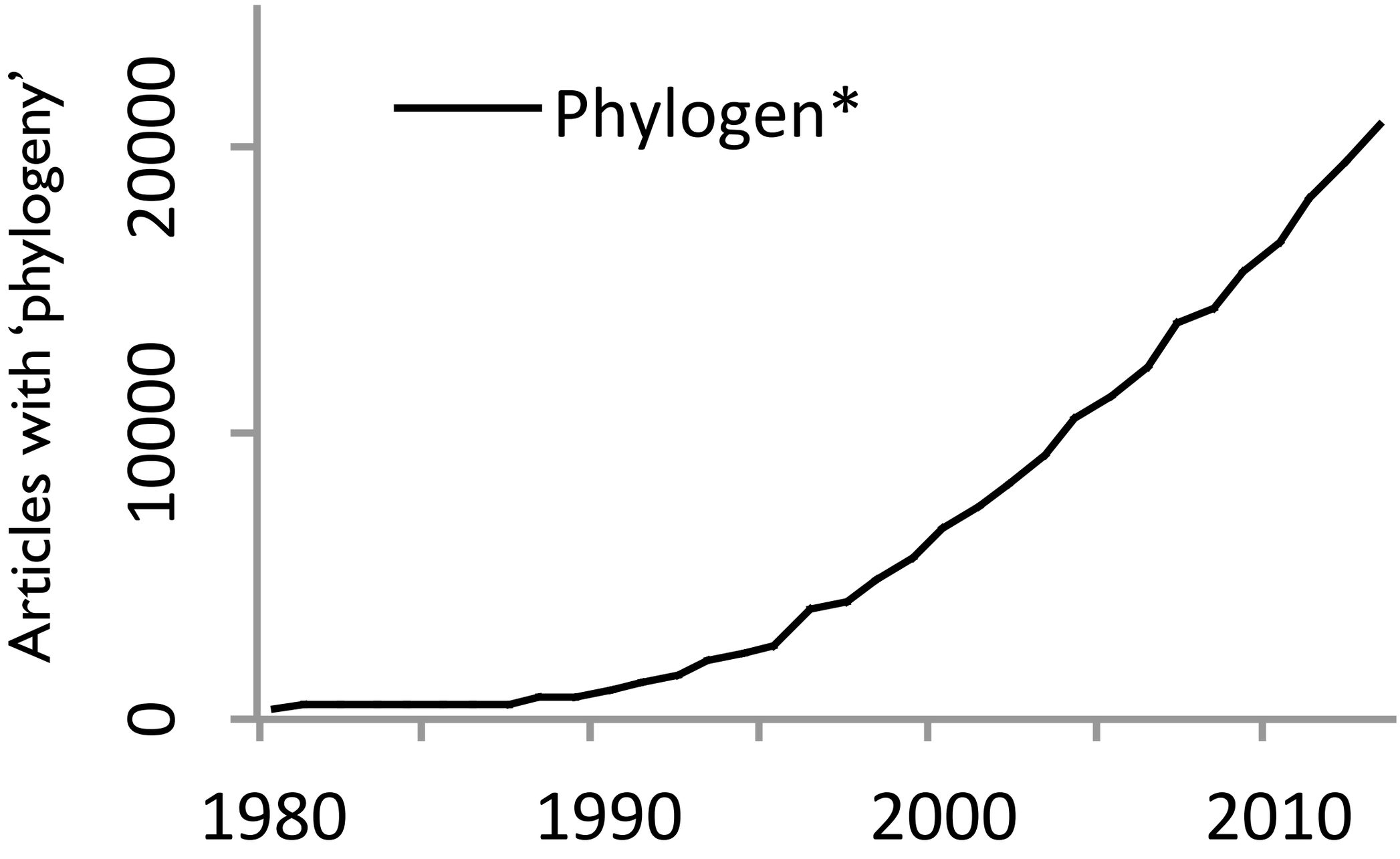
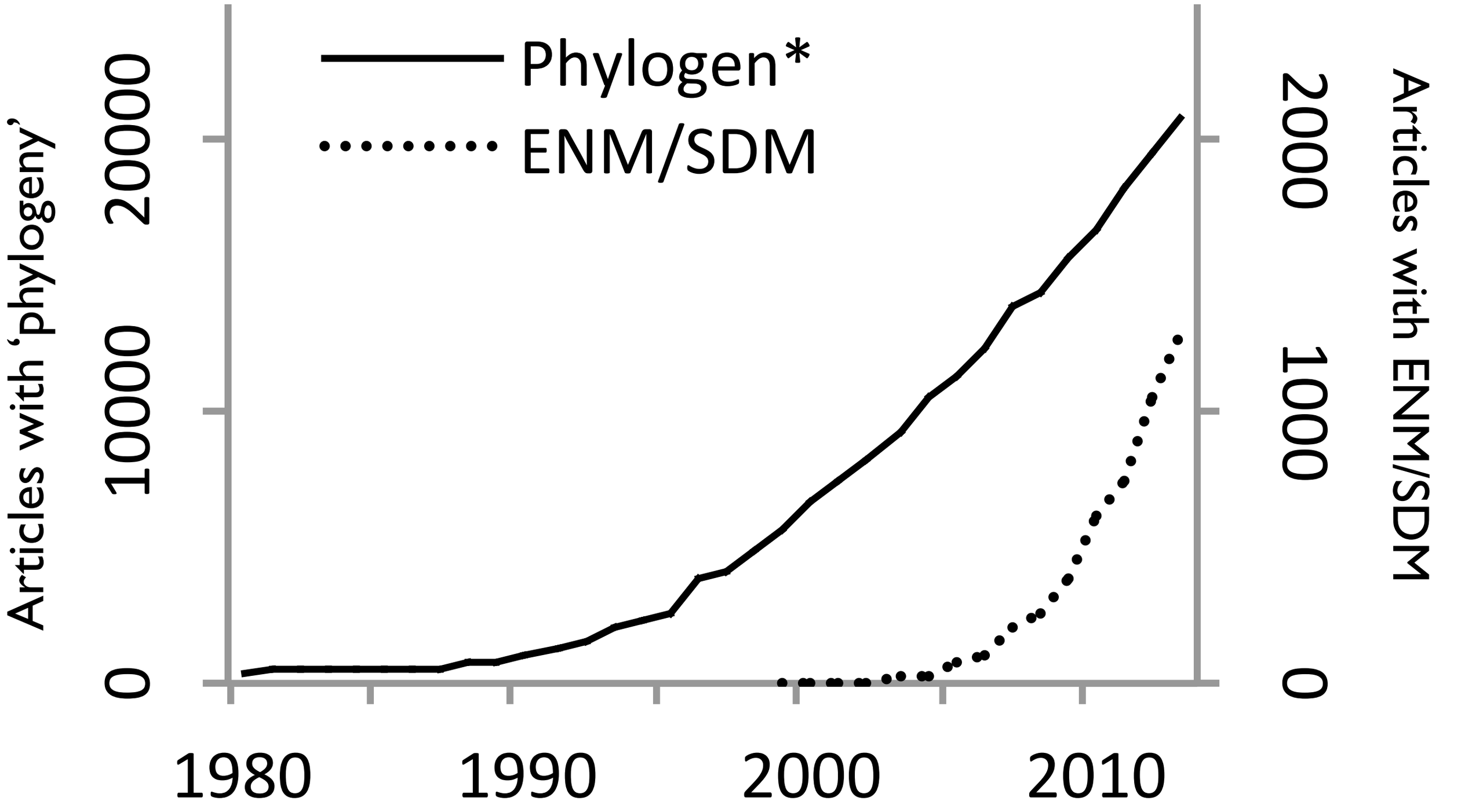


Figure. Growth in phylogenetics studies. By 2013, studies involving phylogenies were ~1% of all published science (David Hillis, pers. comm.).

Data: Scopus search, 2014

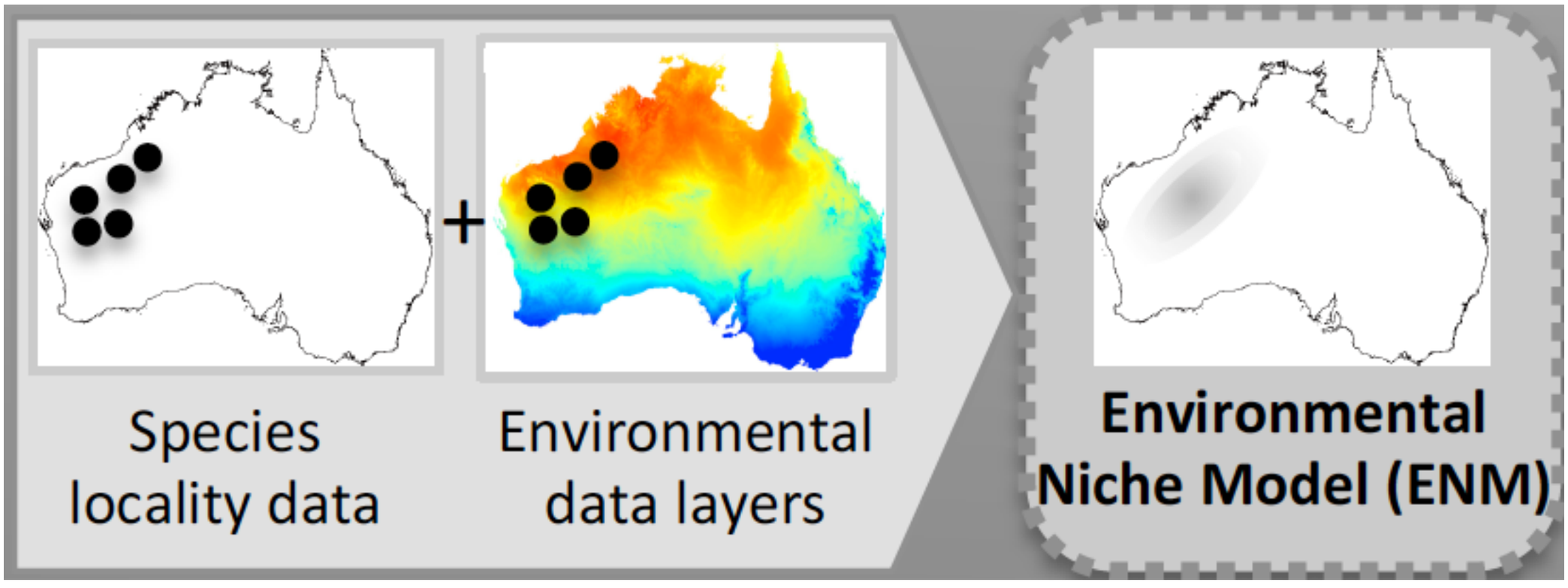
ENMs in the scientific literature



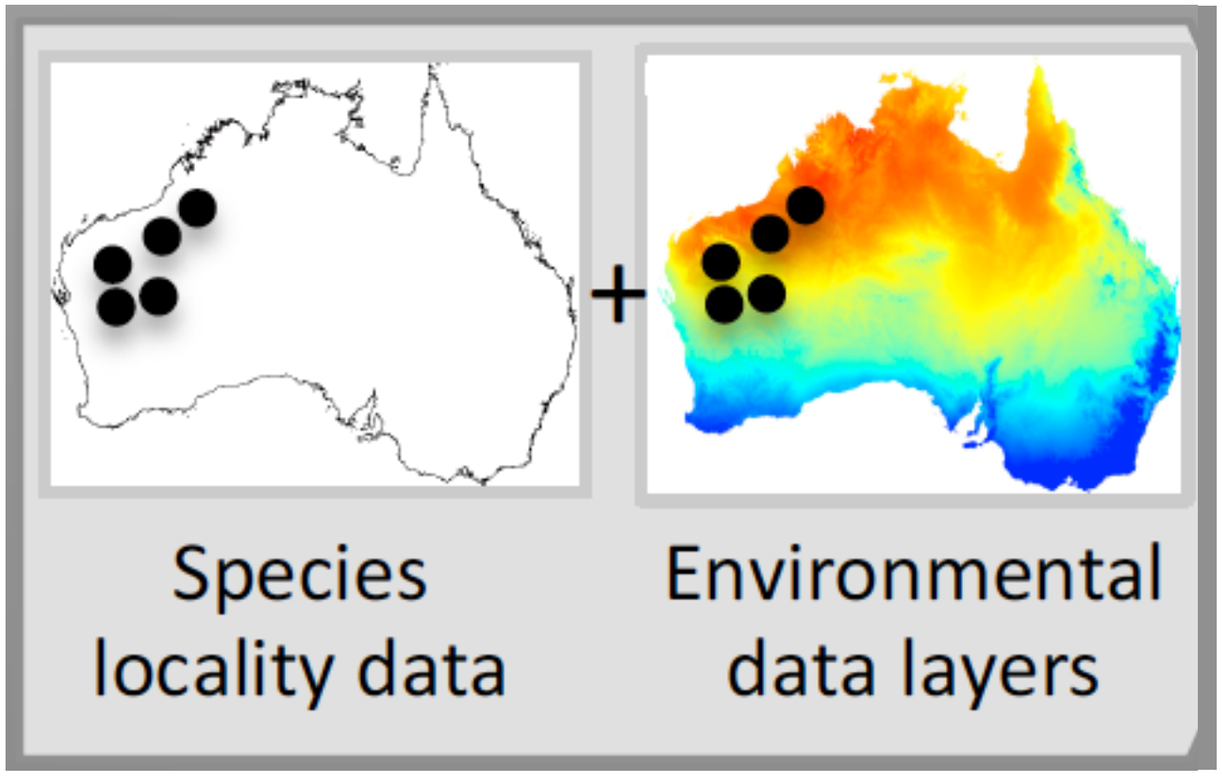
Data: Scopus search, 2014

Ecological Niche Modelling (ENM)/ Species Distribution Modelling (SDM):

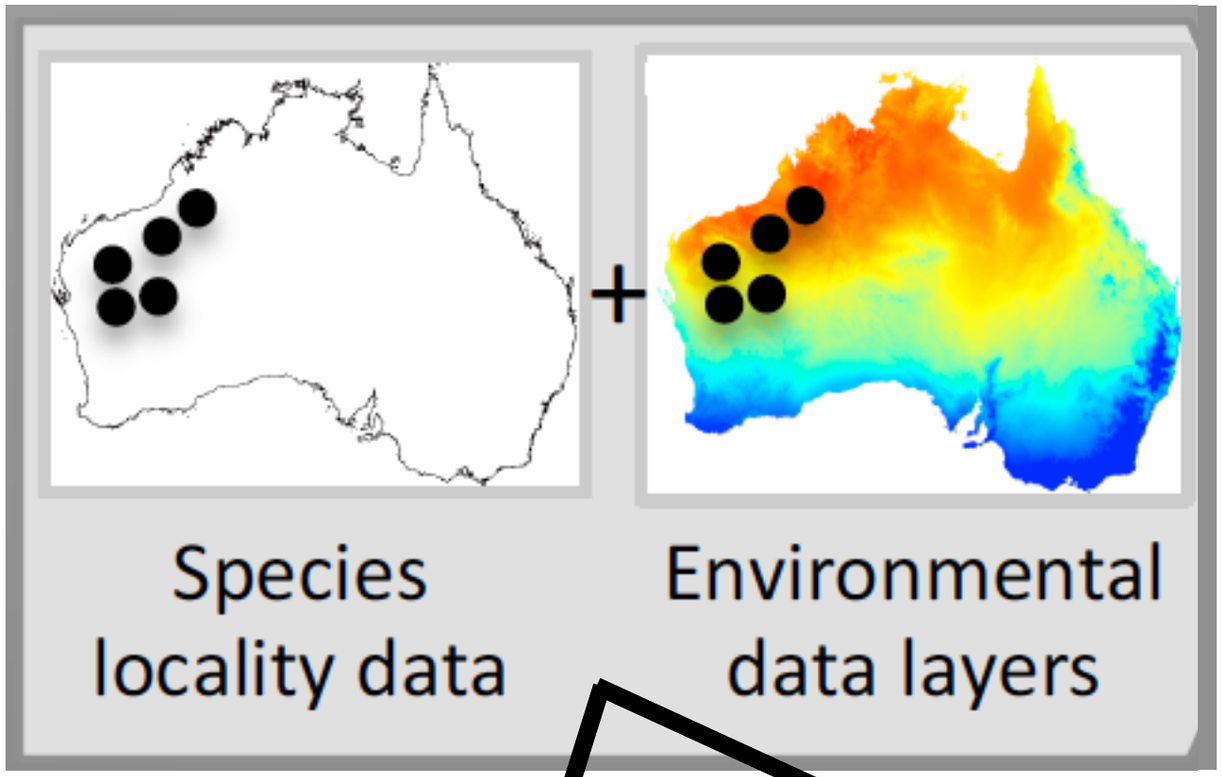
A brief introduction



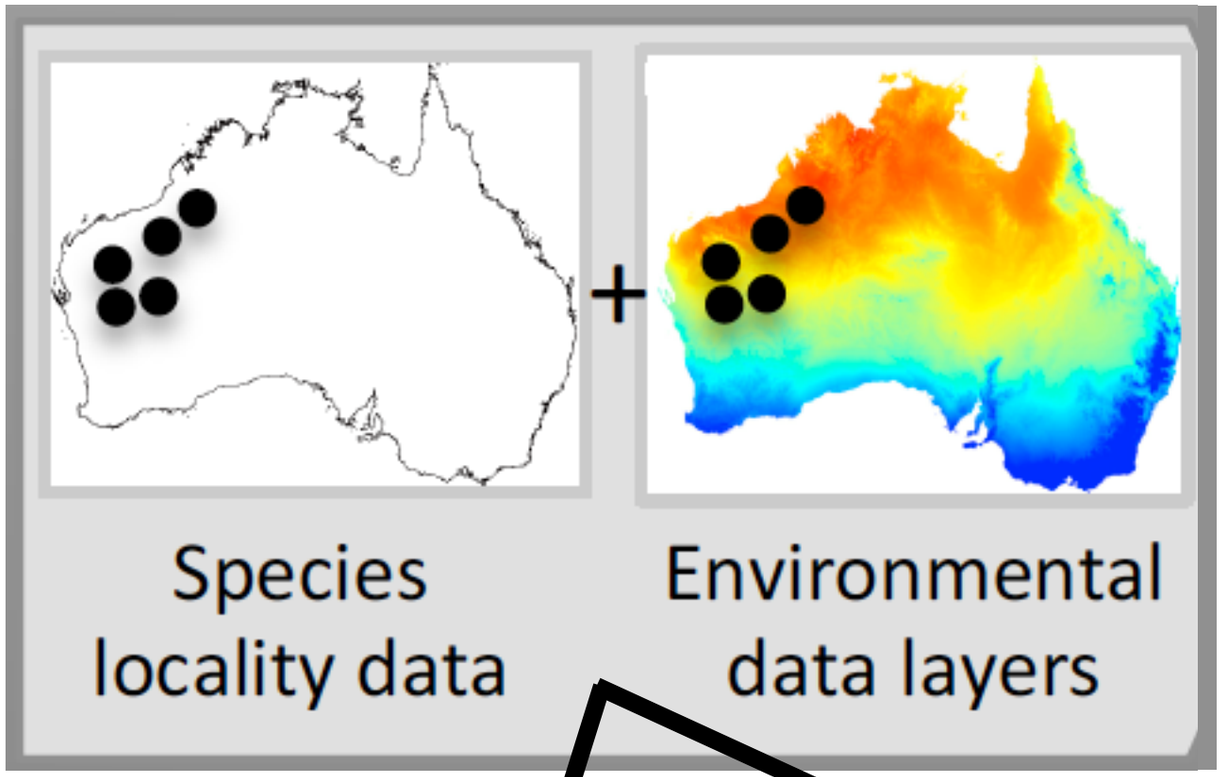
Niche modelling: basic procedure



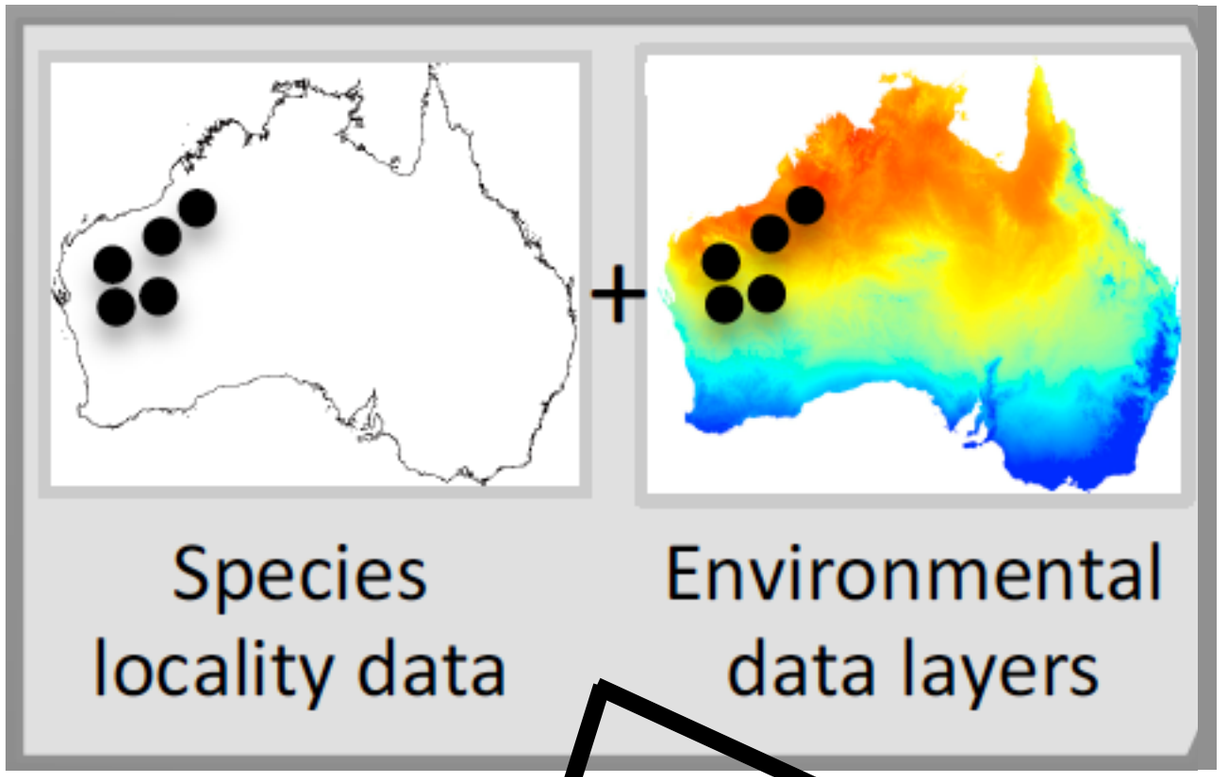
Niche modelling: basic procedure

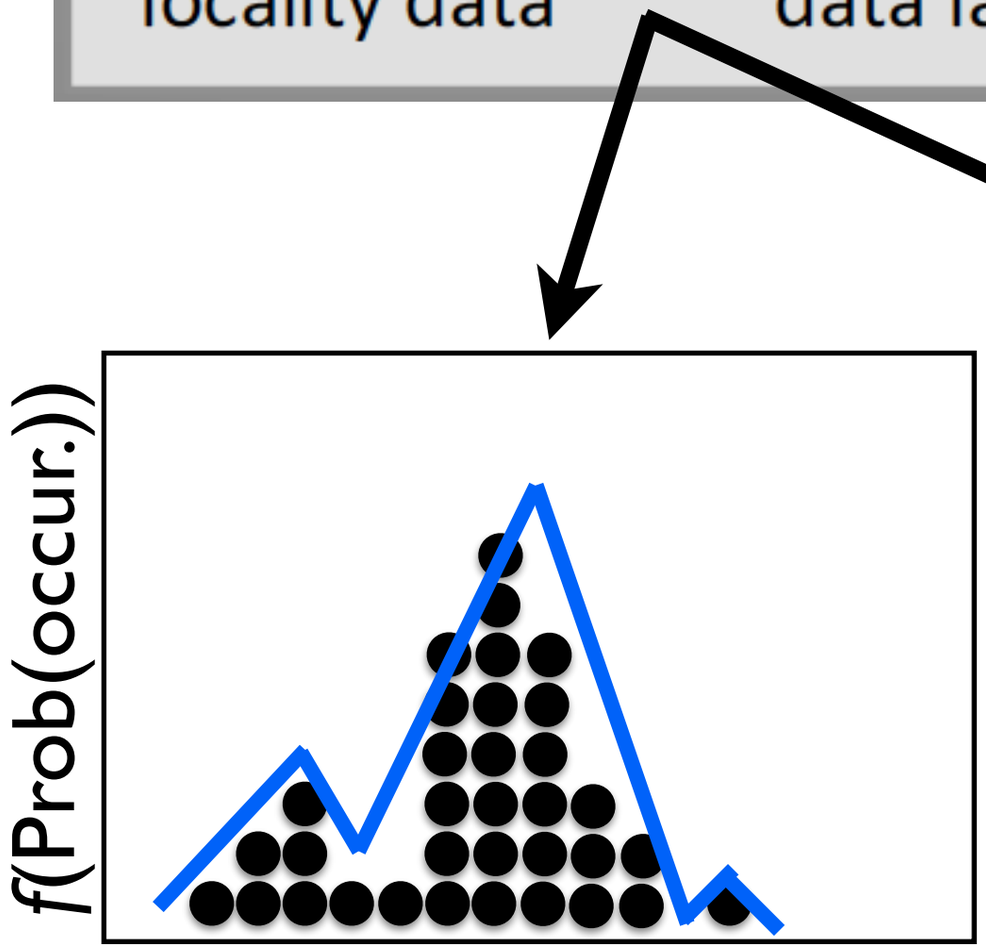


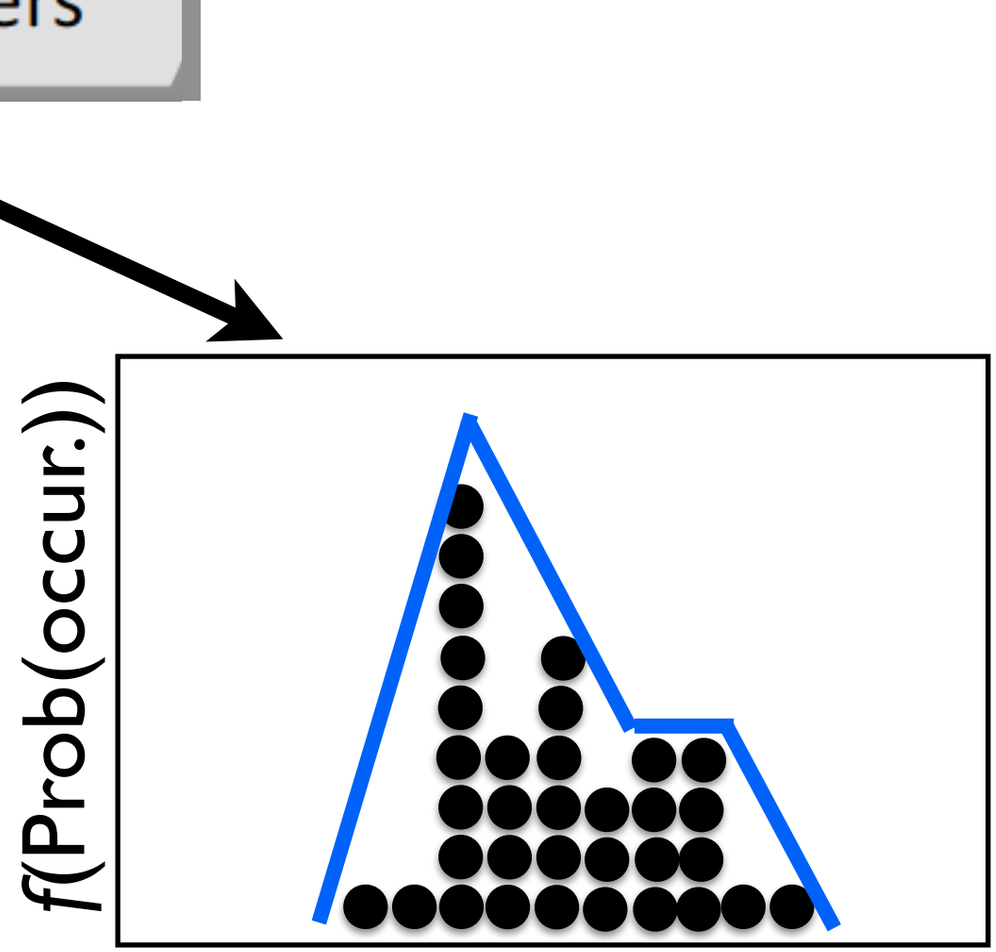
Niche modelling: basic procedure



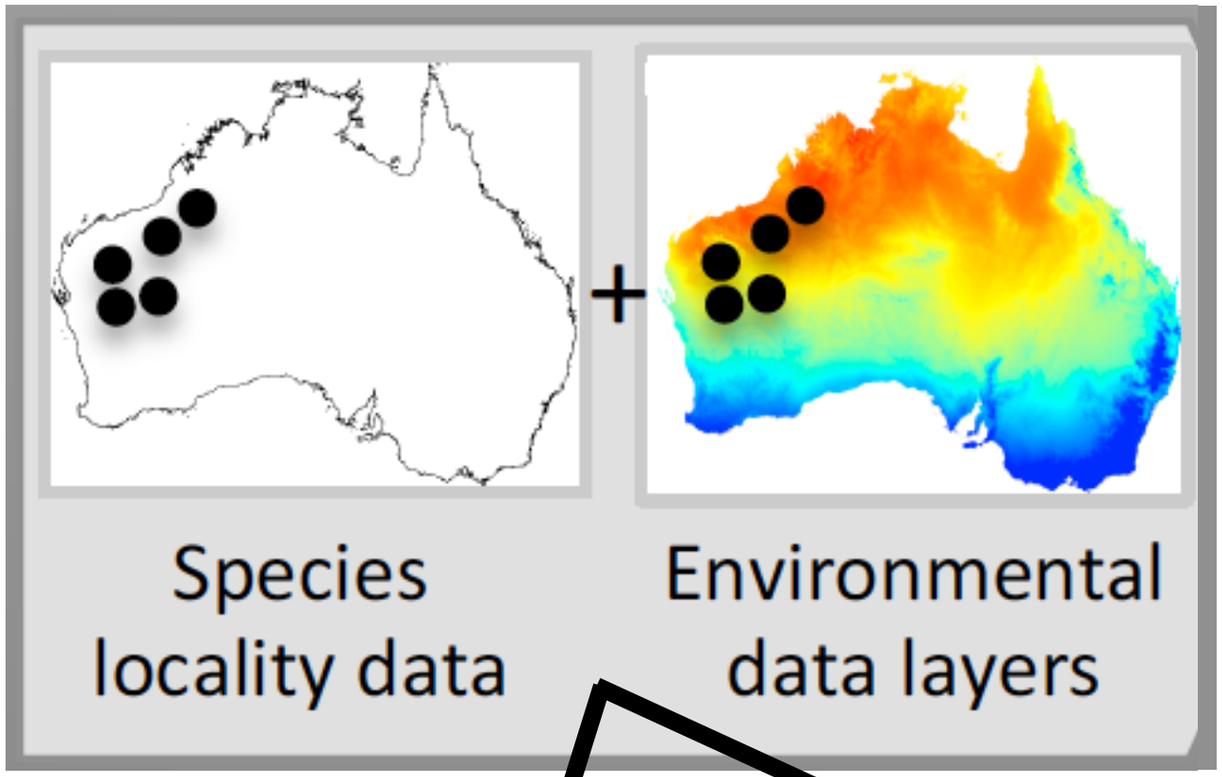
Niche modelling: basic procedure

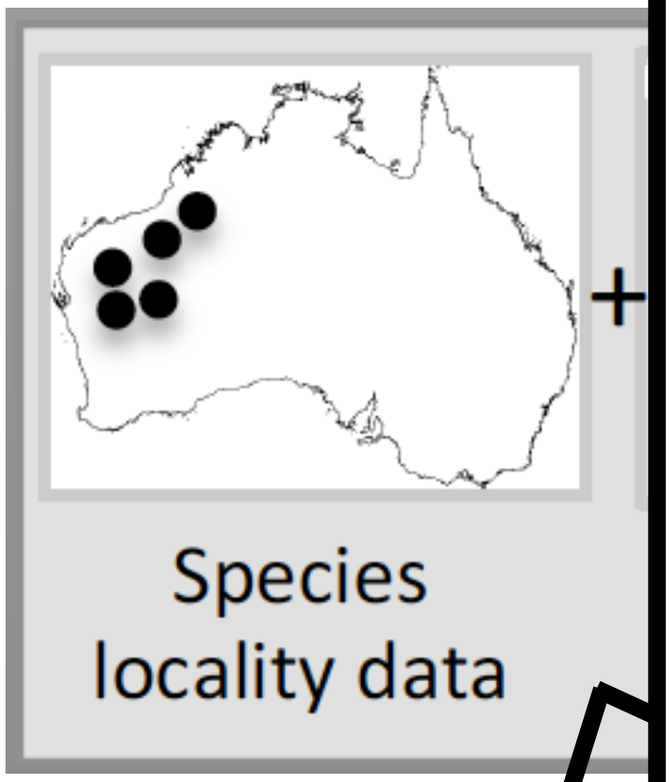






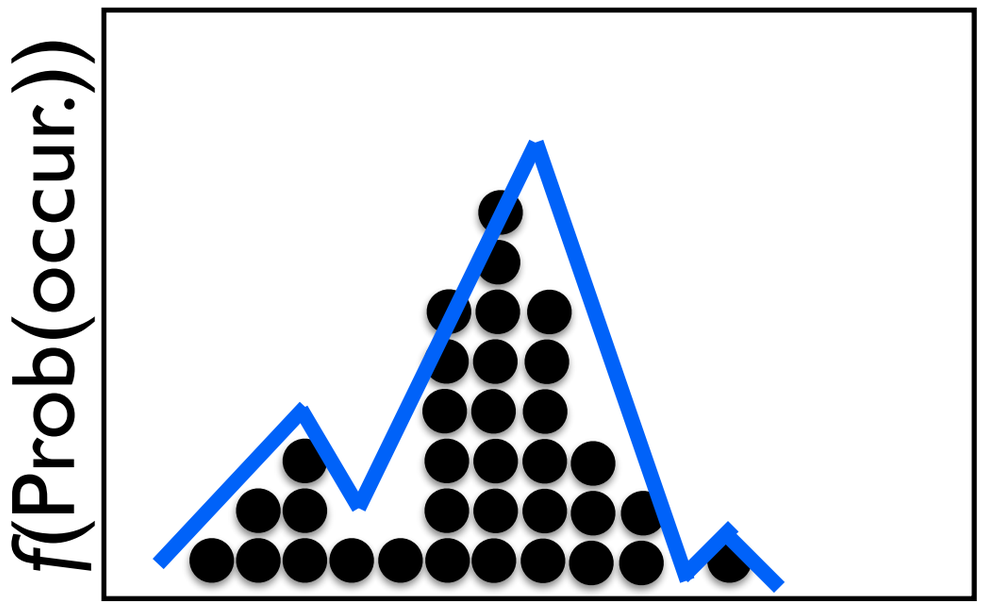
Niche modelling: basic procedure

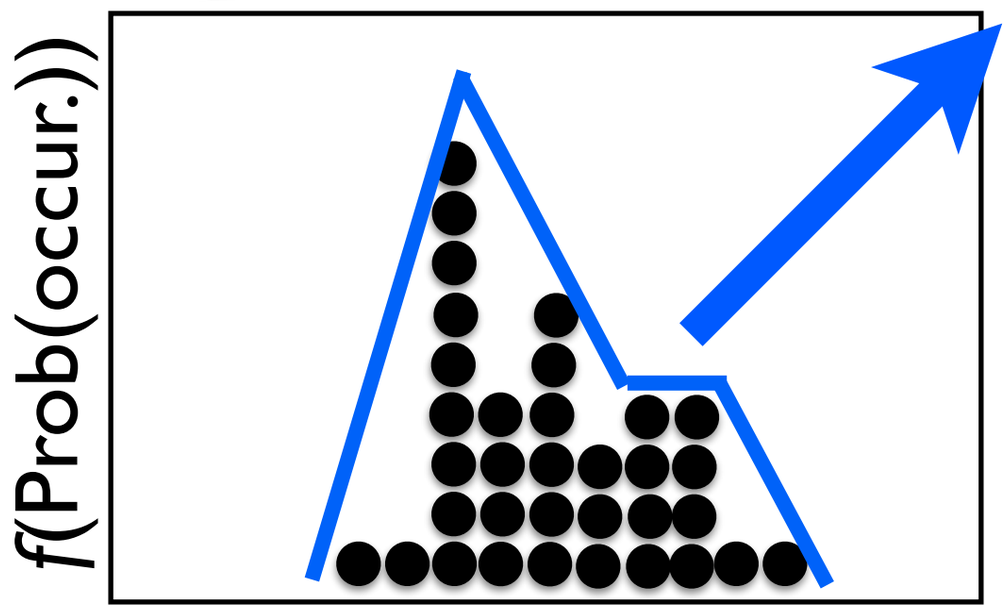




Common ENM tools

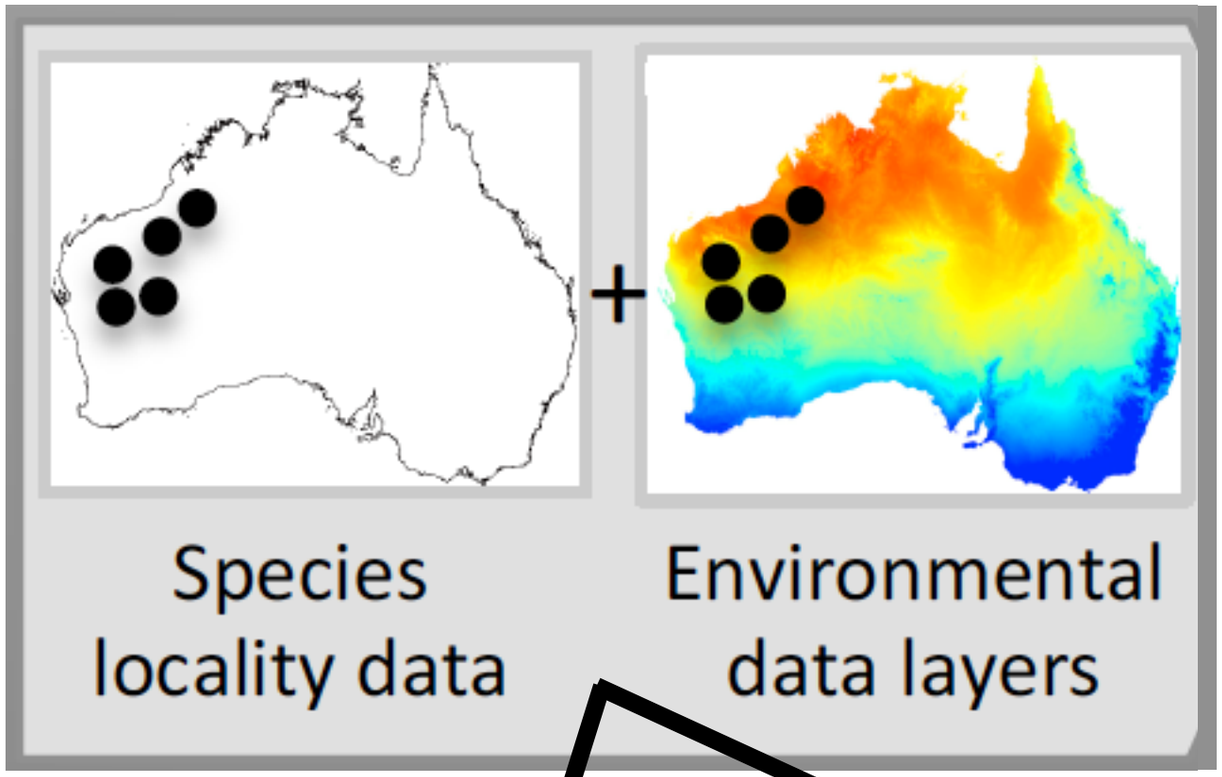
- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest



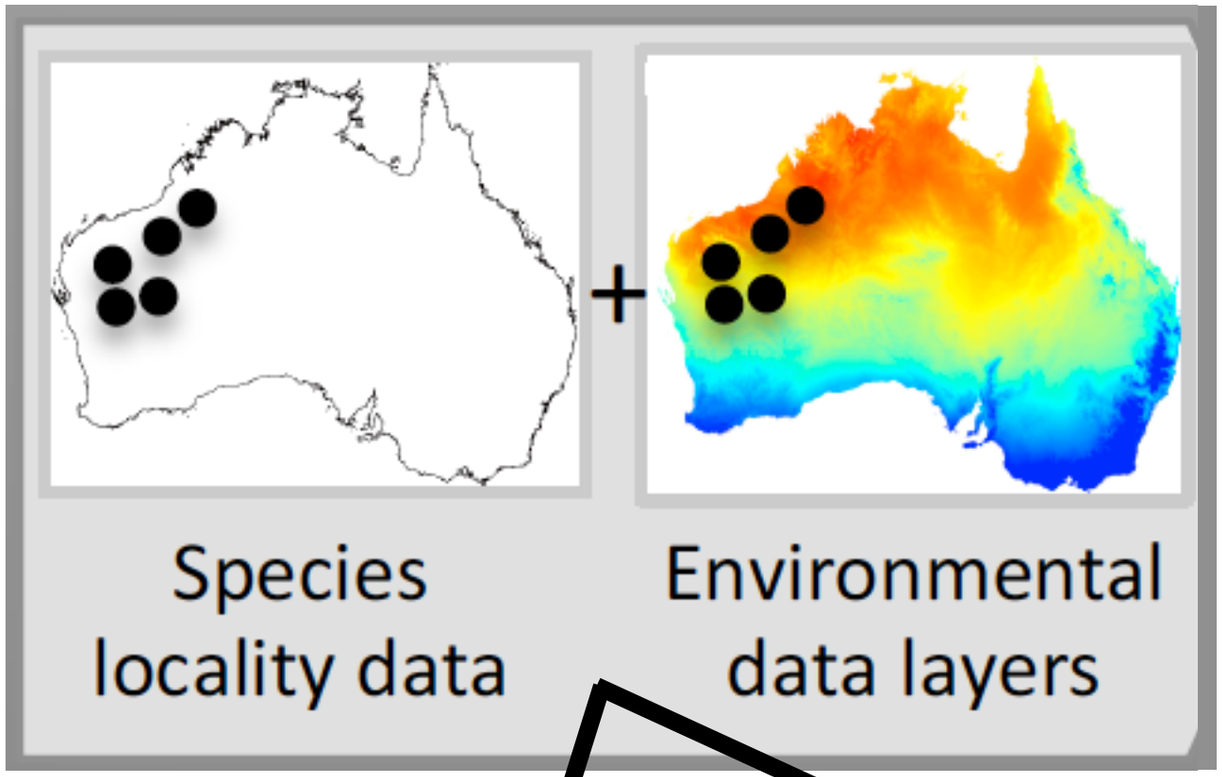


Ecological niche model

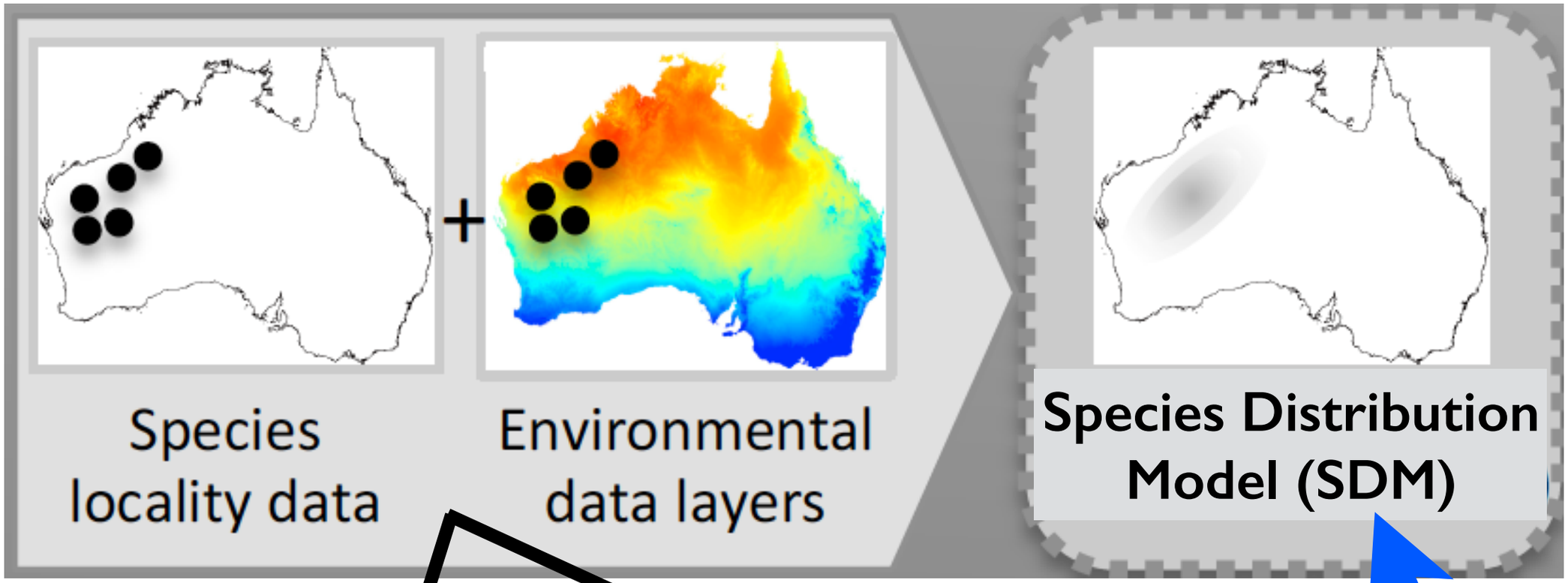
Niche modelling: basic procedure



Niche modelling: basic procedure



Niche modelling: basic procedure



ENM procedure

Occurrence data

+

Environmental Data

=

ENM procedure

Occurrence data

+

Environmental Data

=

Ecological niche model

ENM procedure

Occurrence data

+

Environmental Data

=

Ecological niche model

ENM procedure

Occurrence data

+

Environmental Data

=

Ecological niche model

Projection onto map

ENM procedure

Occurrence data

+

Environmental Data

=

Ecological niche model

Projection onto map

Species Distribution Model

ENM procedure

Occurrence data

+

Environmental Data

=

Ecological niche model

Projection onto map

Species Distribution Model

Common critiques

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ niche model

Projection onto map

Species Distribution Model

Common critiques

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ Environmental niche model

Projection onto map

Species Distribution Model

Common critiques

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ Environmental niche model

Projection onto map

Species Distribution Model

Common critiques

* **Presence-only data**
* **Sampling issues**

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ Environmental niche model

Projection onto map

Species Distribution Model

Common critiques

* **Presence-only data**
* **Sampling issues**

Correlational modelling

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ Environmental niche model

Projection onto map

Species Distribution Model

Common critiques

* **Presence-only data**
* **Sampling issues**

Correlational modelling

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ Environmental niche model

Projection onto map

Species Distribution Model

Common critiques

* **Presence-only data**
* **Sampling issues**

Correlational modelling

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest

***correlation vs. causation**

***overfitting**

***poor extrapolation ability**

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ Environmental niche model

Projection onto map

Species Distribution Model

Common critiques

* **Presence-only data**
* **Sampling issues**

Correlational modelling

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest

***correlation vs. causation**

***overfitting**

***poor extrapolation ability**

Projection into space assumes:

ENM procedure

Occurrence data

+

Environmental Data

=

~~Ecological~~ Environmental niche model

Projection onto map

Species Distribution Model

Common critiques

* **Presence-only data**
* **Sampling issues**

Correlational modelling

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest

***correlation vs. causation**

***overfitting**

***poor extrapolation ability**

Projection into space assumes:

* **Niche model equals true niche**
* **Niche is wholly environmental**
* **Equilibrium**
* **No dispersal limitation**
* **No competition, etc.**

Nevertheless:

All models are wrong, but some models are useful

ENMs/SDMs are clearly useful, even if imperfect

Let's take them more seriously, as statistical *models*

PCM procedure

Tip data

+

Phylogeny

=

PCM procedure

Tip data

+

Phylogeny

=

Fit continuous trait model(s)

PCM procedure

Tip data

+

Phylogeny

=

Fit continuous trait model(s)

PCM procedure

Tip data

+

Phylogeny

=

Fit continuous trait model(s)

1. **Estimate ancestral states**
2. **Test hypotheses about processes, rates**

PCM procedure

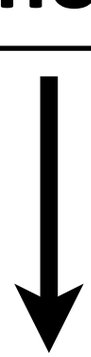
Tip data

+

Phylogeny

=

Fit continuous trait model(s)



1. **Estimate ancestral states**
2. **Test hypotheses about processes, rates**

Common critiques

PCM procedure

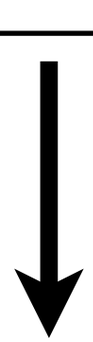
Tip data

+

Phylogeny

=

Fit continuous trait model(s)



1. **Estimate ancestral states**
2. **Test hypotheses about processes, rates**

Common critiques

* **Tip data have error**

PCM procedure

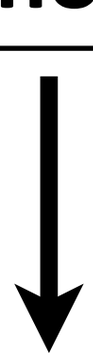
Tip data

+

Phylogeny

=

Fit continuous trait model(s)



1. **Estimate ancestral states**
2. **Test hypotheses about processes, rates**

Common critiques

* **Tip data have error**
* **Phylogeny, branch lengths have error**

PCM procedure

Tip data

+

Phylogeny

=

Fit continuous trait model(s)

Common PCM models

- Brownian motion
- OU process
- Early Burst
- Trend
- Multiple regimes, shifting means, etc.

(fit with ML or Bayes)

1. **Estimate ancestral states**
2. **Test hypotheses about processes, rates**

Common critiques

* **Tip data have error**
* **Phylogeny, branch lengths have error**

PCM procedure

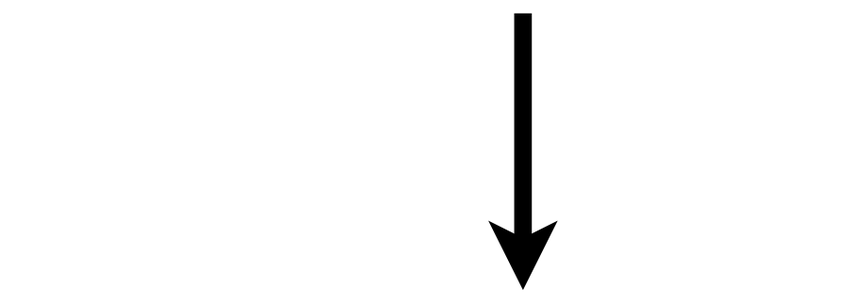
Tip data

+

Phylogeny

=

Fit continuous trait model(s)



1. Estimate ancestral states
2. Test hypotheses about processes, rates

Common PCM models

- Brownian motion
- OU process
- Early Burst
- Trend
- Multiple regimes, shifting means, etc.

(fit with ML or Bayes)

Common critiques

* **Tip data have error**
* **Phylogeny, branch lengths have error**
* **Identifiability**
* **Without paleo: weighted averaging**
* **Interpretation**
* **Fitting multiregime models, overfitting issues**

Nevertheless:

All models are wrong, but some models are useful

Models of continuous trait evolution are useful and can be improved

1

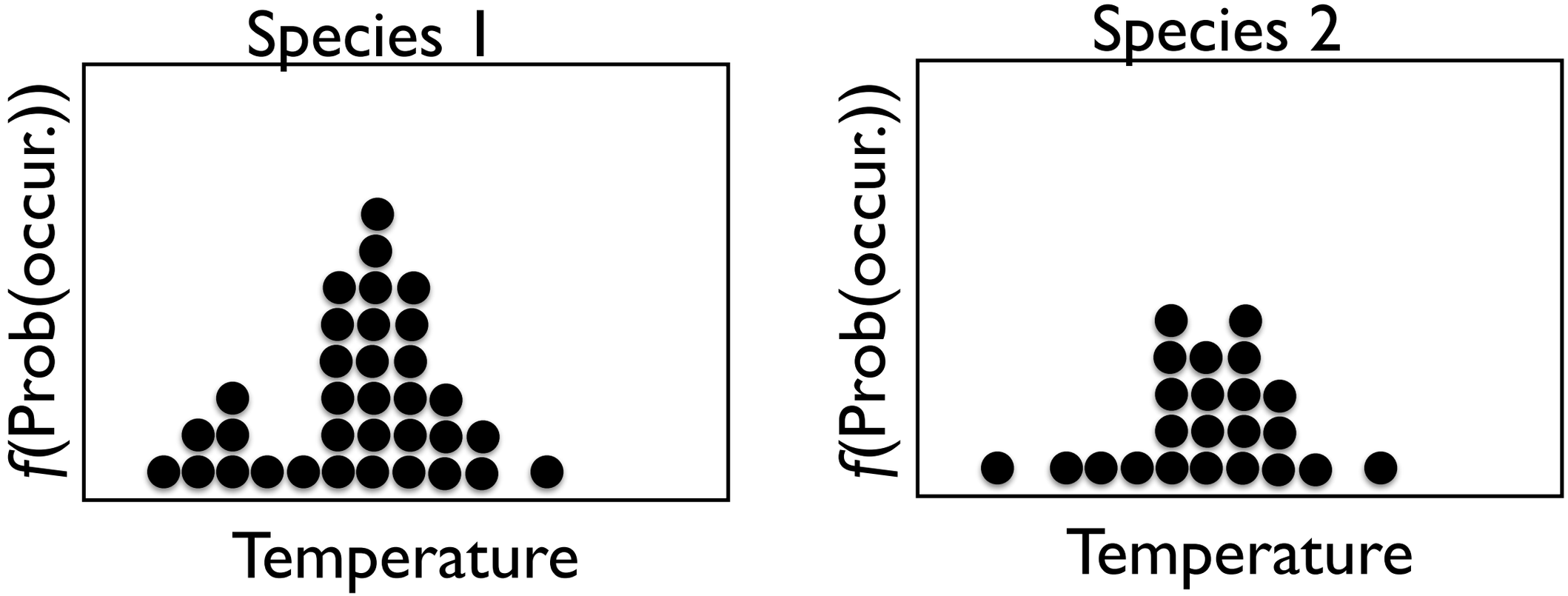
2

Combining ENMs and PCMs

ENM+PCM: use the mean?

1. Fit a niche model

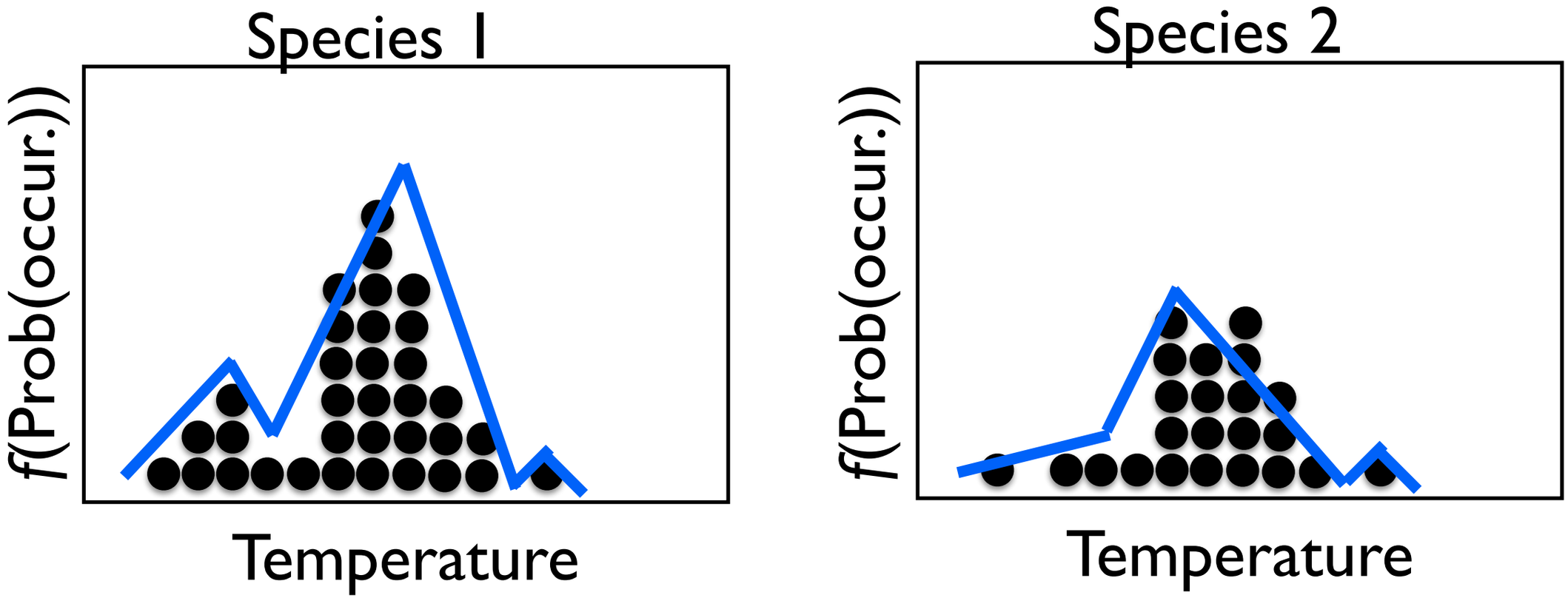
2a. Sample from curve:



ENM+PCM: use the mean?

1. Fit a niche model

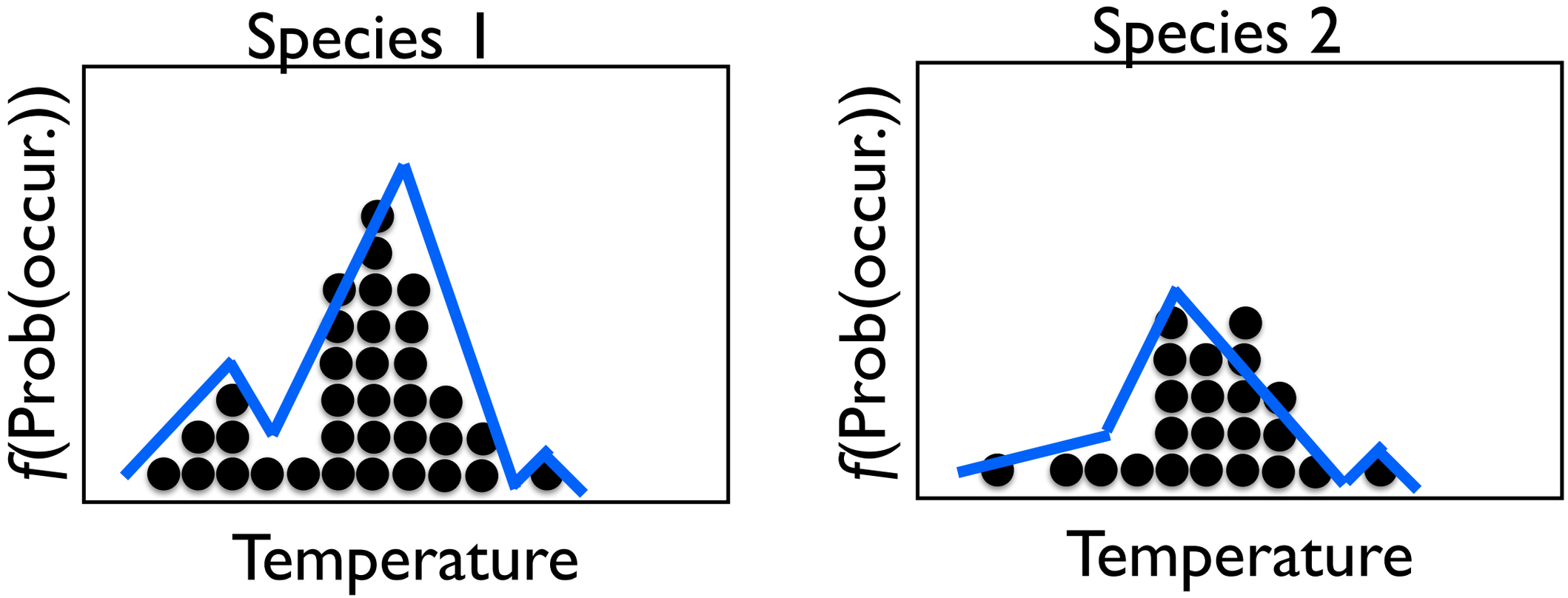
2a. Sample from curve:



ENM+PCM: use the mean?

1. Fit a niche model

2a. Sample from curve:

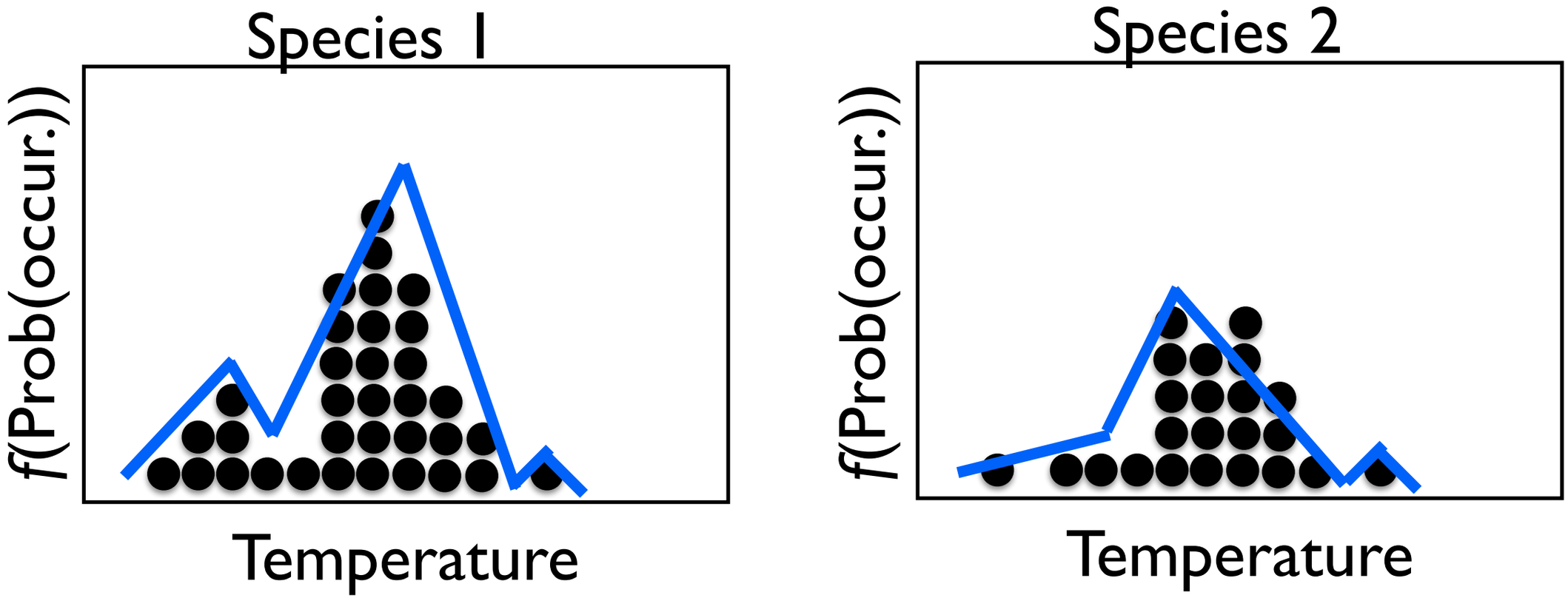


2b. Ancestral states w. Brownian:

ENM+PCM: use the mean?

1. Fit a niche model

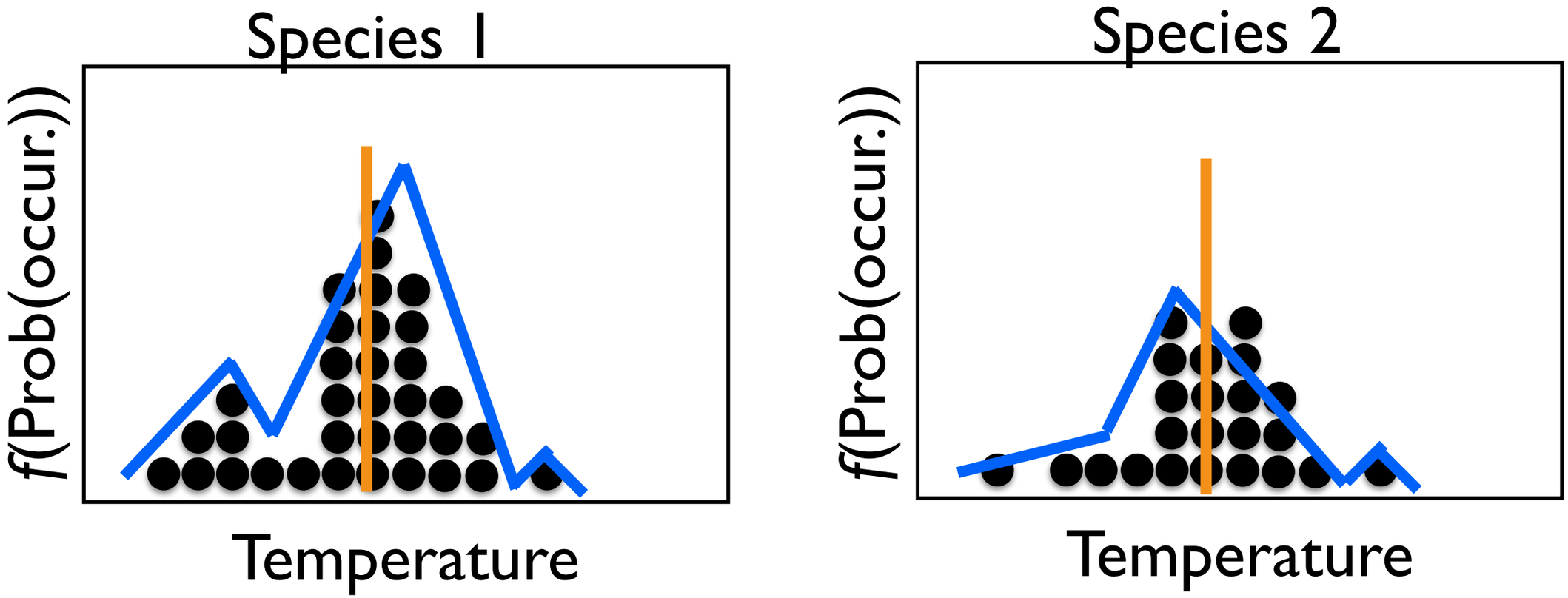
2a. Sample from curve:



ENM+PCM: use the mean?

1. Fit a niche model

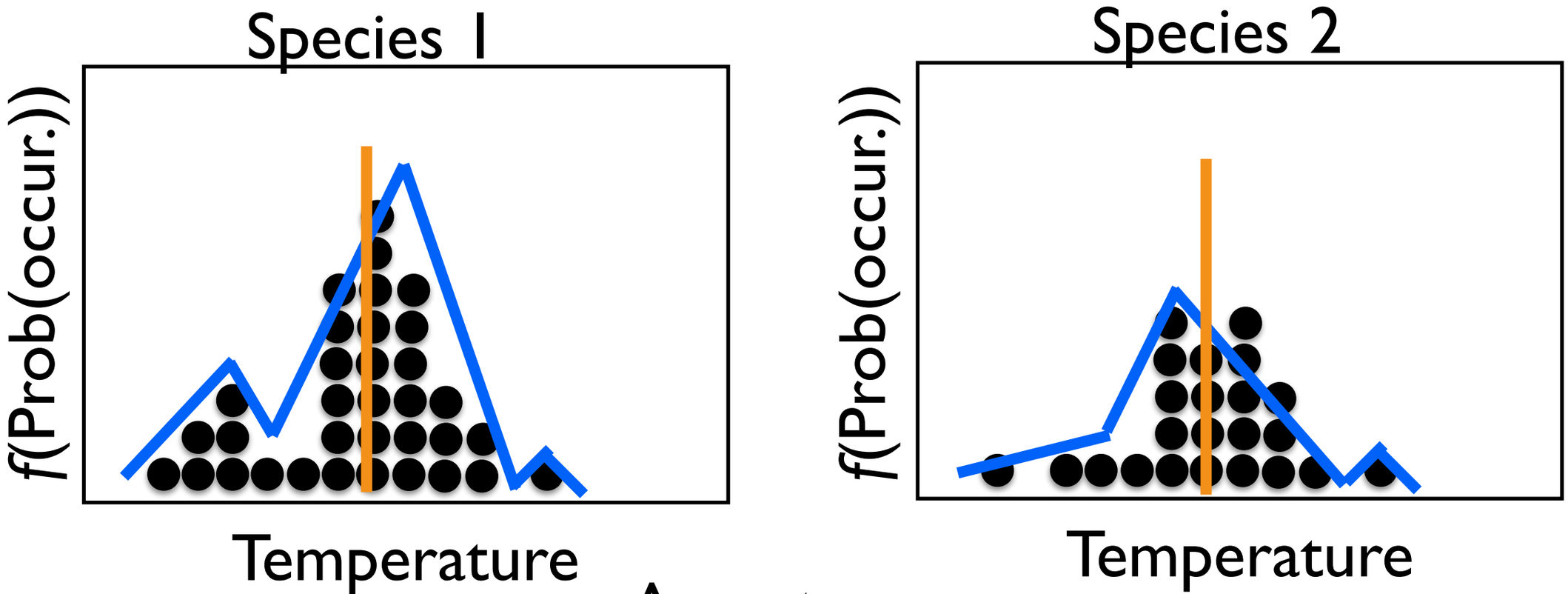
2a. Sample from curve:



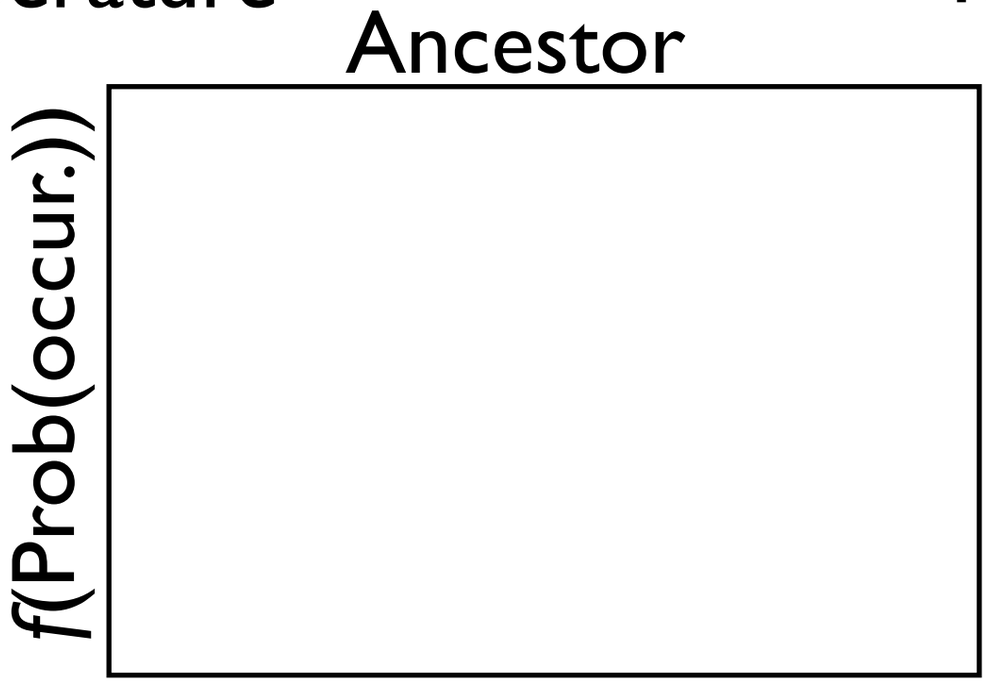
ENM+PCM: use the mean?

1. Fit a niche model

2a. Sample from curve:



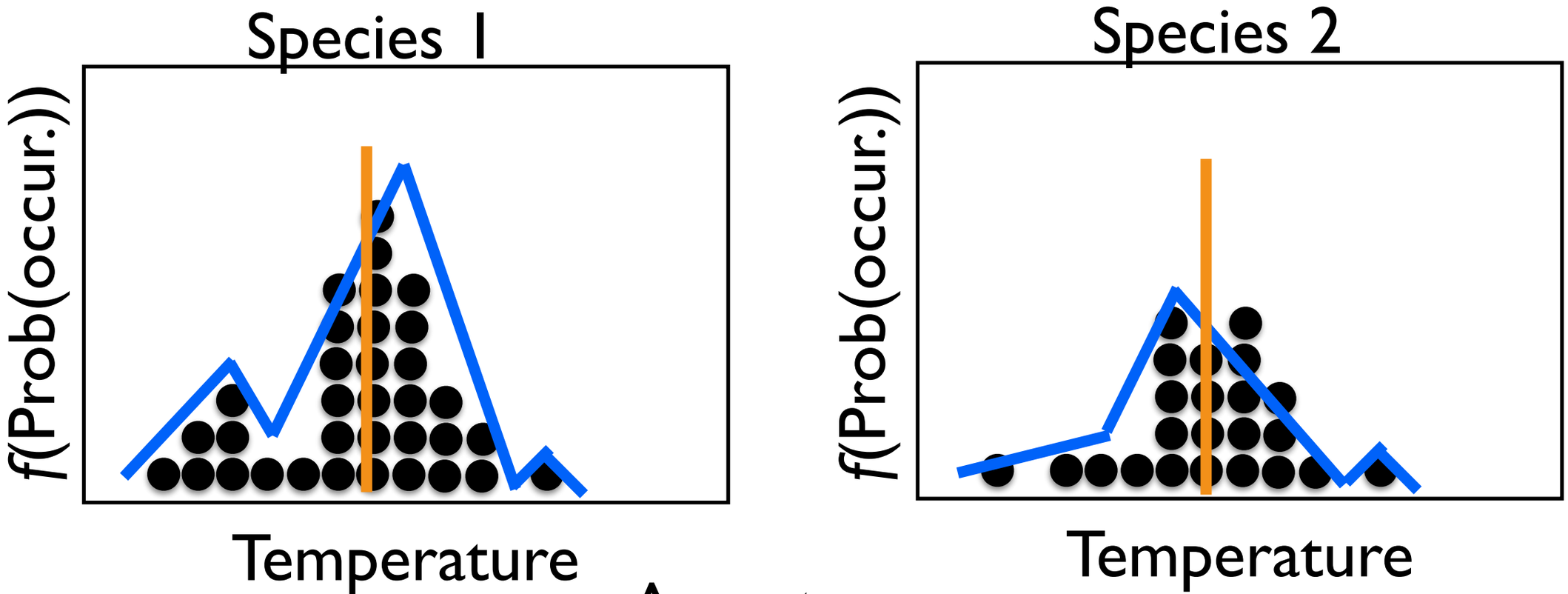
2b. Ancestral states w. Brownian:



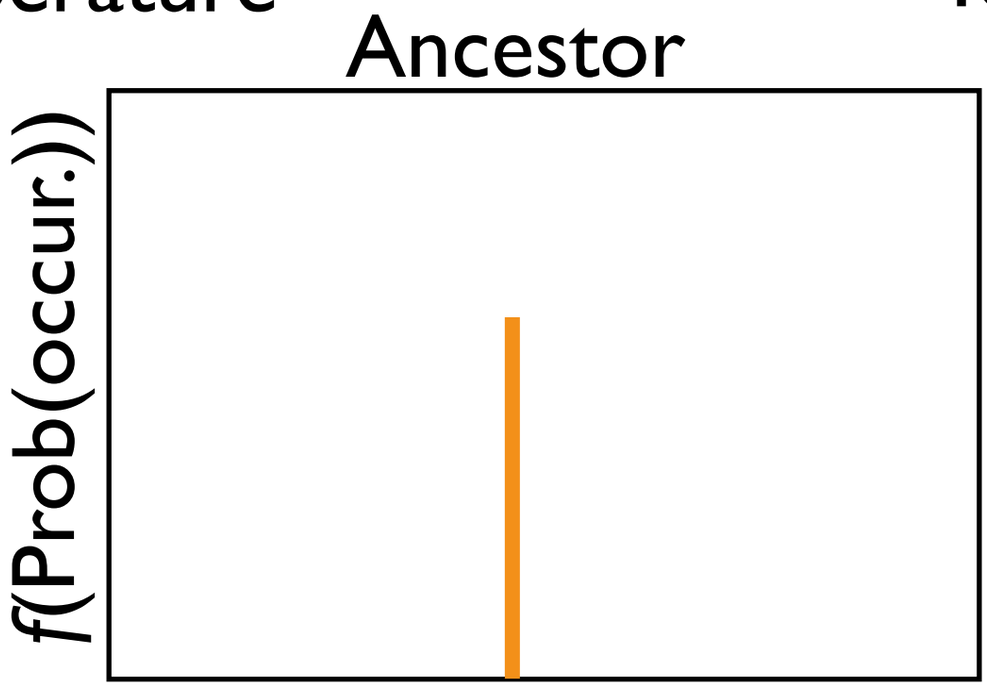
ENM+PCM: use the mean?

1. Fit a niche model

2a. Sample from curve:



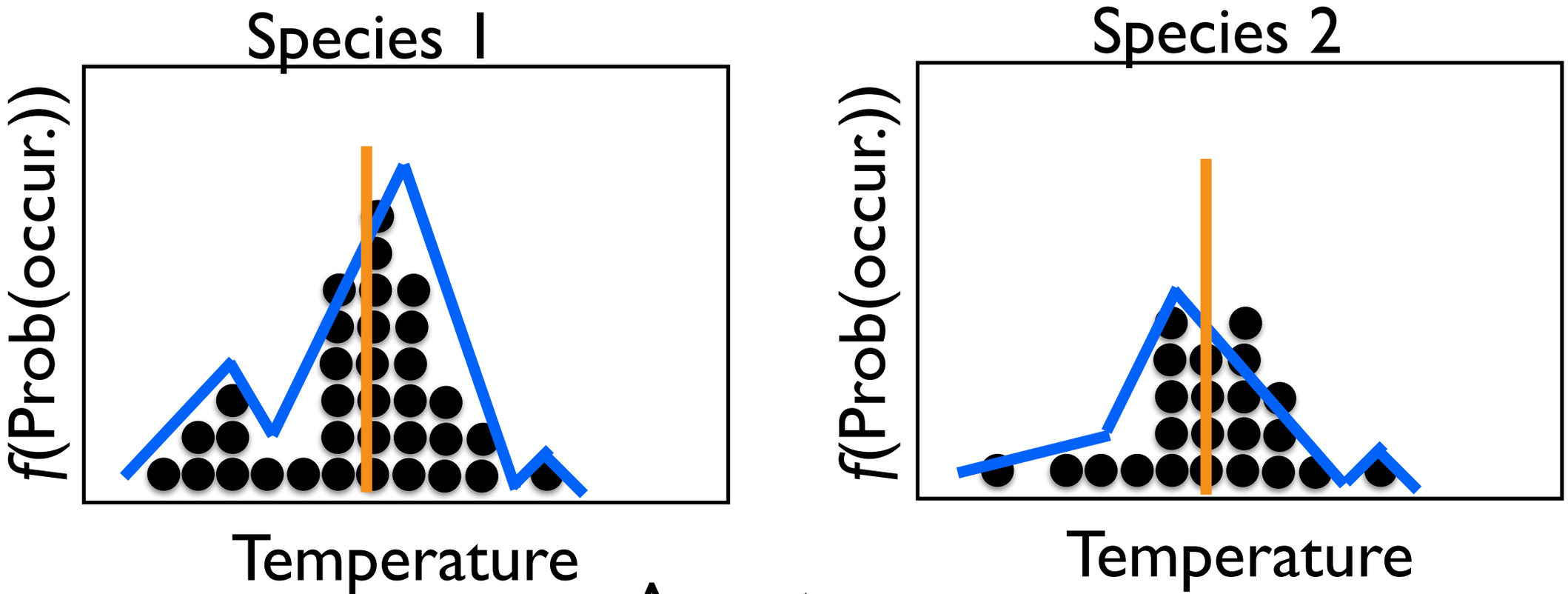
2b. Ancestral states w. Brownian:



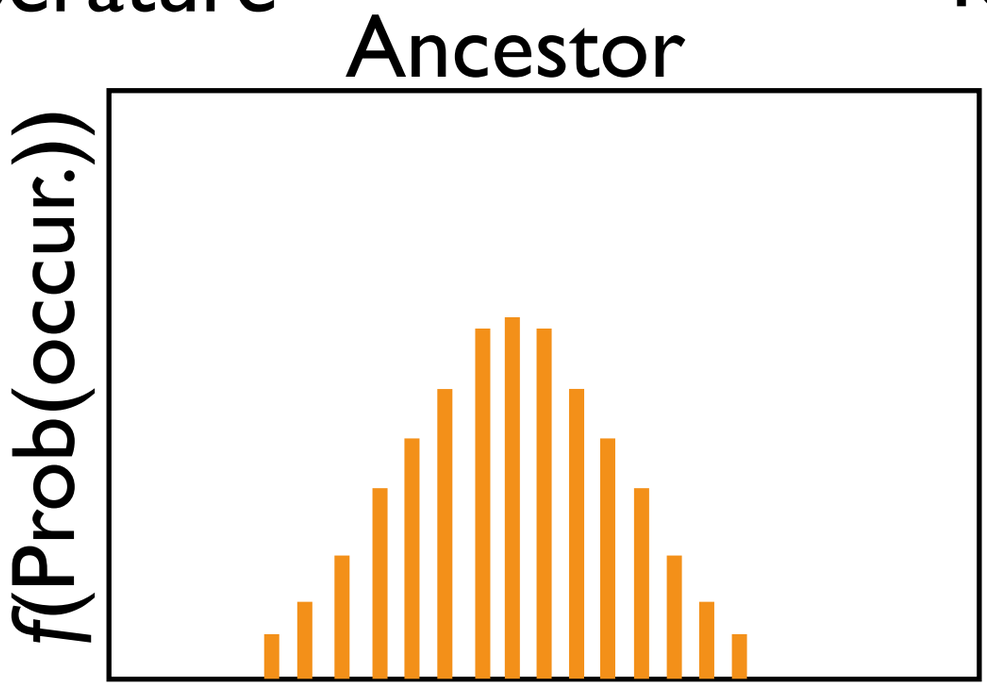
ENM+PCM: use the mean?

1. Fit a niche model

2a. Sample from curve:



2b. Ancestral states w. Brownian:

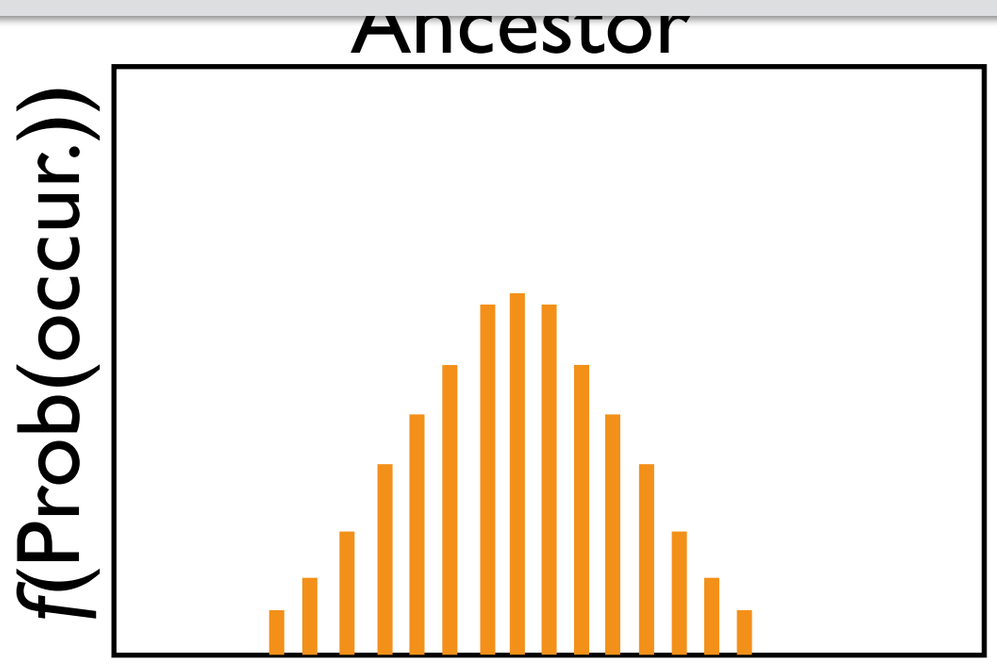


ENM+PCM: use the mean?

Critiques of the mean approach:

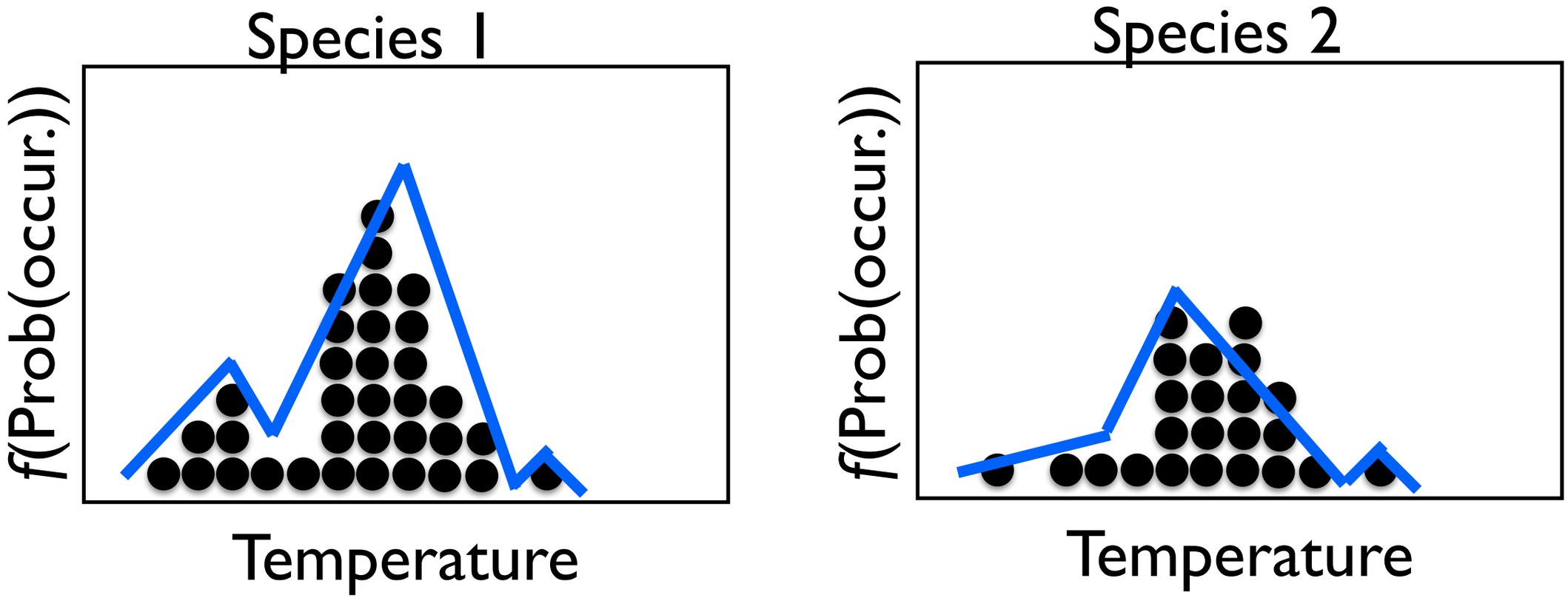
1. Mean is *inferred* from an *inferred* model (2 steps from data)
2. Is distribution gaussian?
3. Limits might be crucial

2b. Ancestral states w. Brownian:



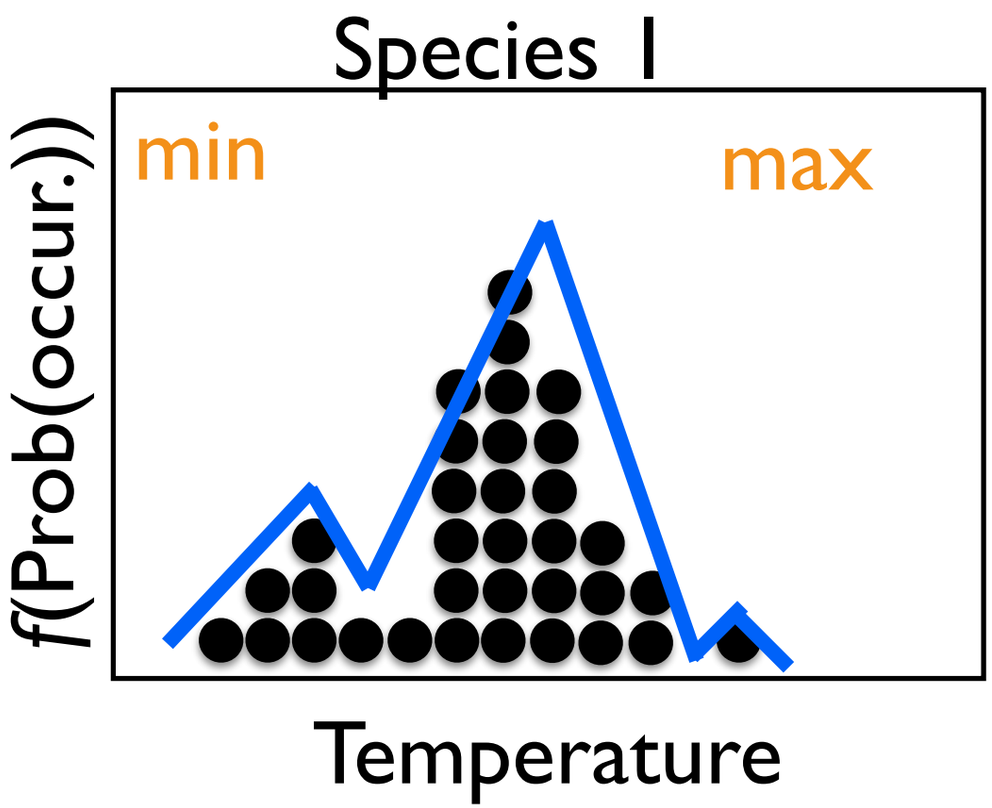
ENM+PCMs: Min-max approach

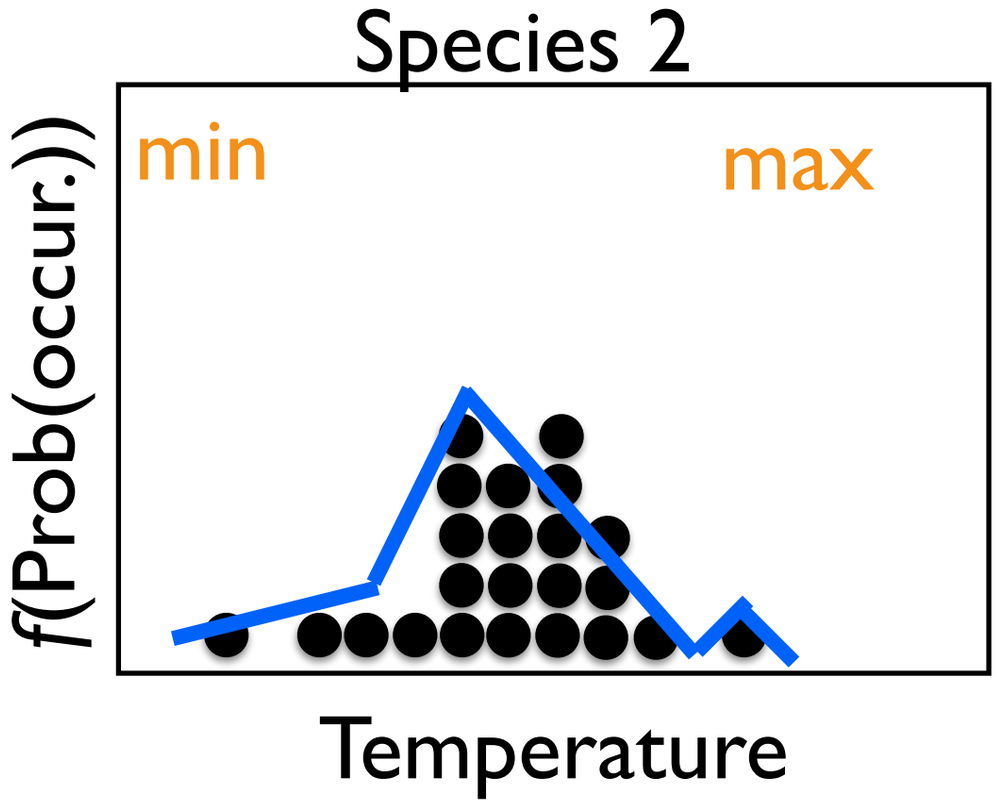
1. Fit a niche model



ENM+PCMs: Min-max approach

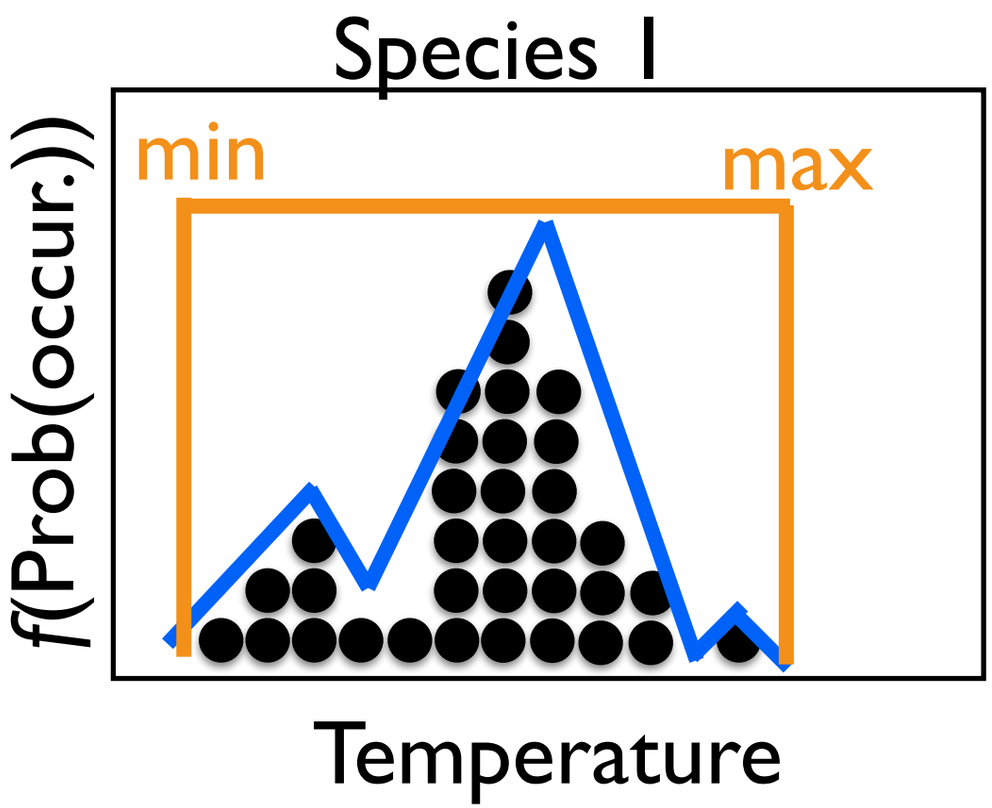
1. Fit a niche model

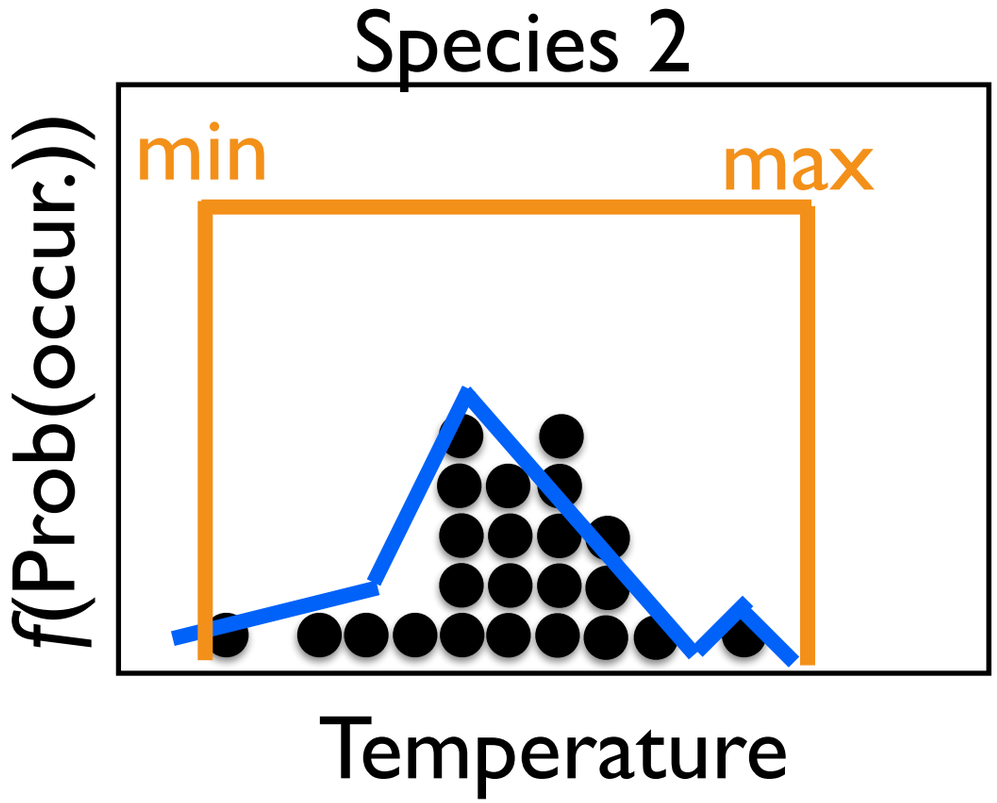




ENM+PCMs: Min-max approach

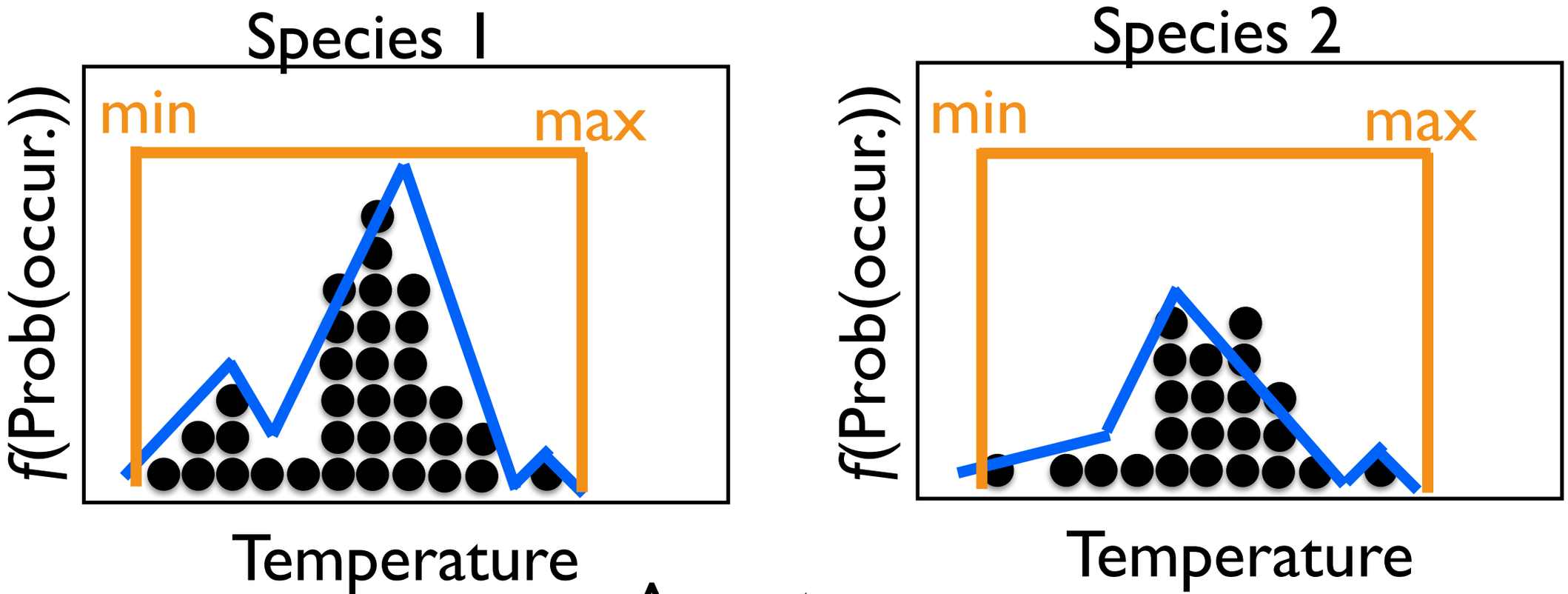
1. Fit a niche model



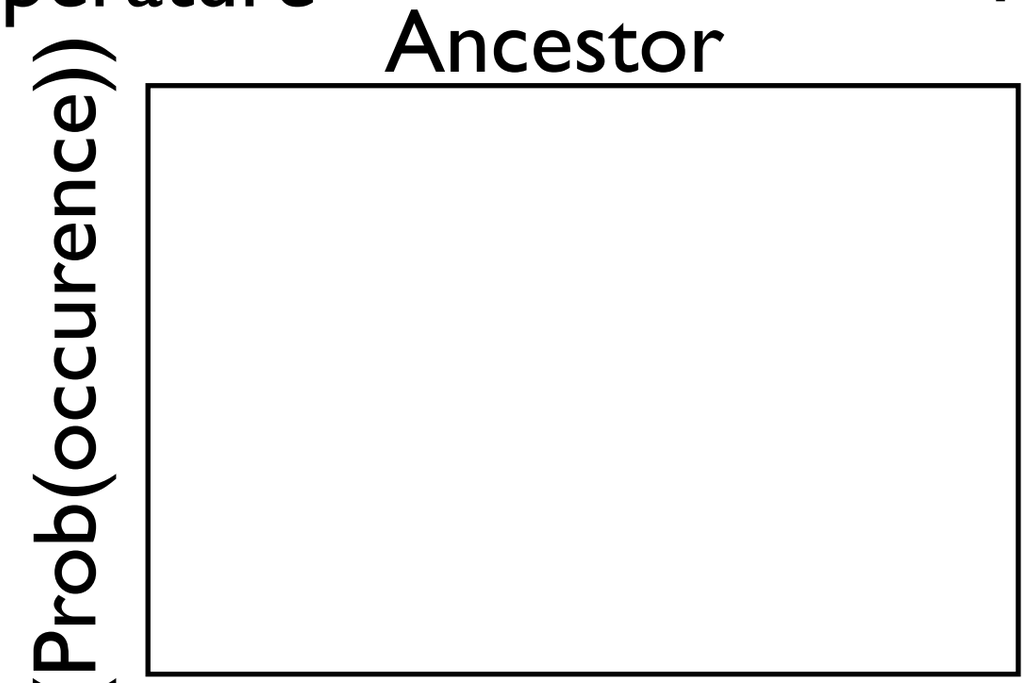


ENM+PCMs: Min-max approach

1. Fit a niche model

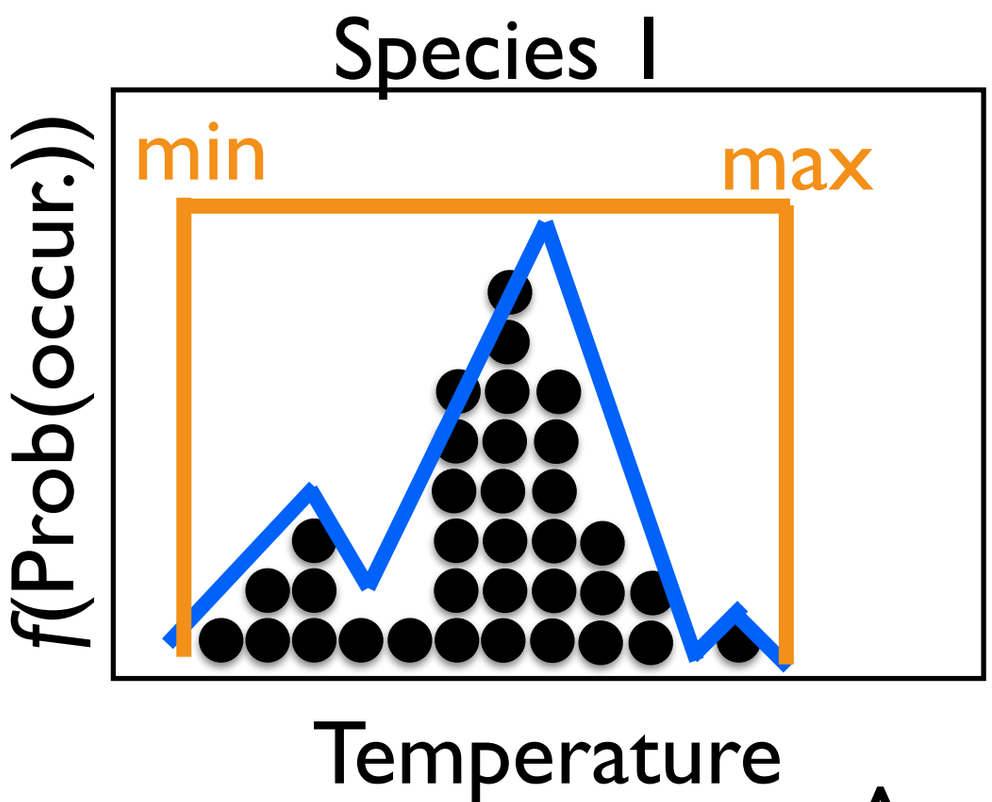


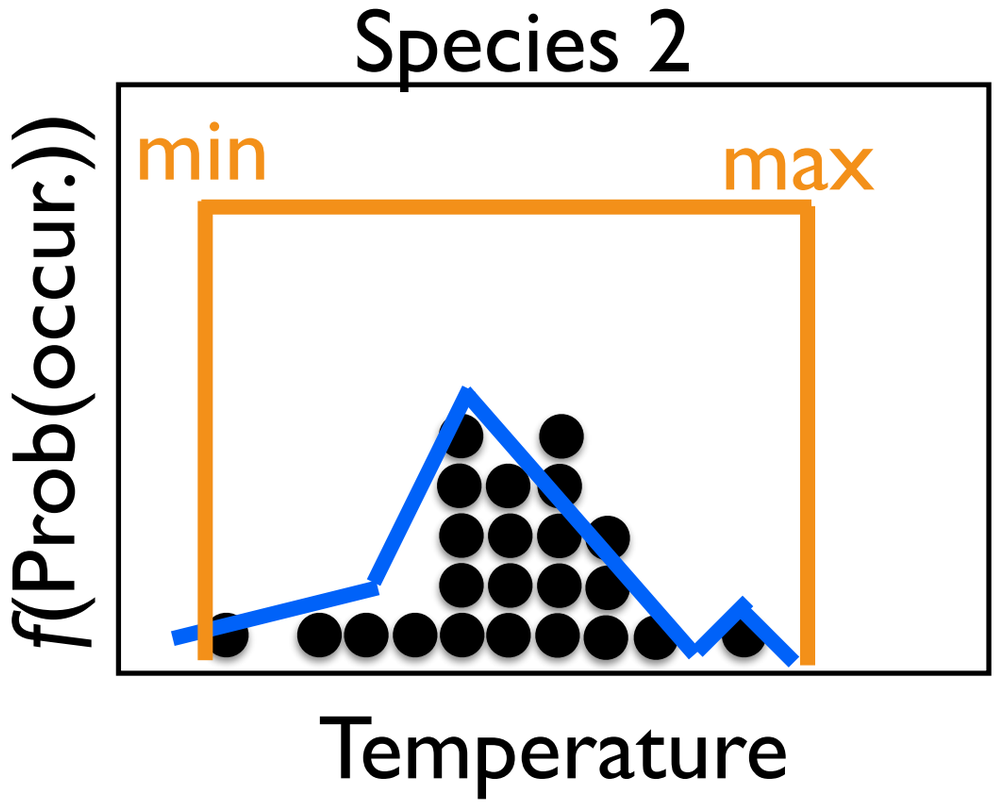
2. Ancestral states w. Brownian:



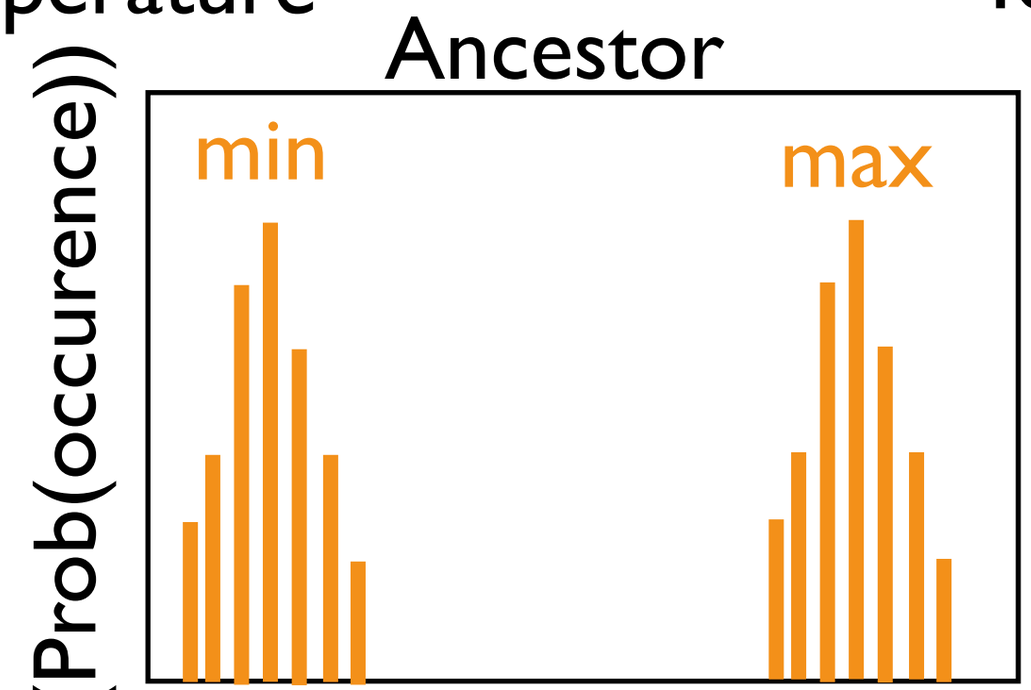
ENM+PCMs: Min-max approach

1. Fit a niche model





2. Ancestral states w. Brownian:

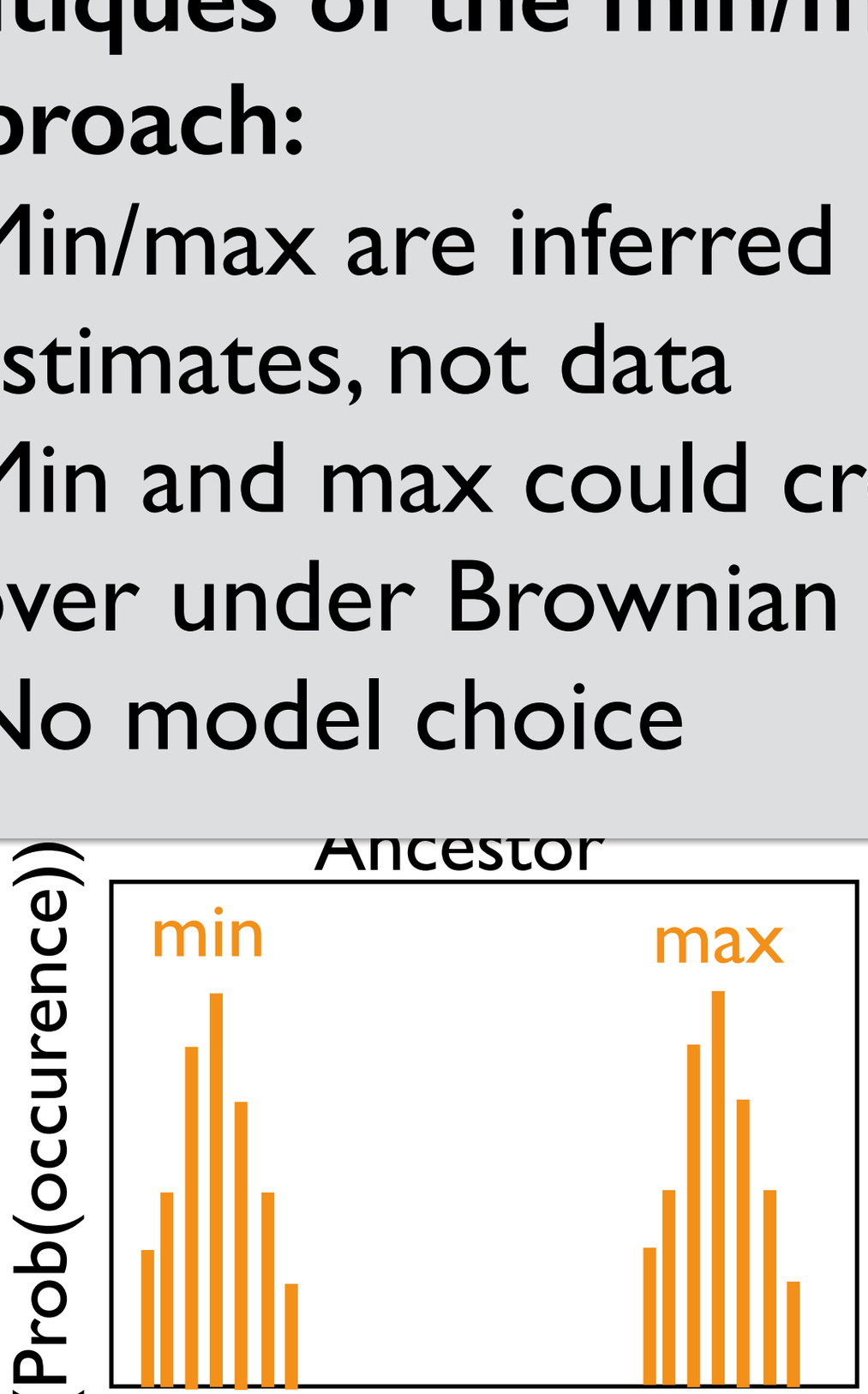


ENM+PCMs: Min-max approach

Critiques of the min/max approach:

1. Min/max are inferred estimates, not data
2. Min and max could cross over under Brownian motion
3. No model choice

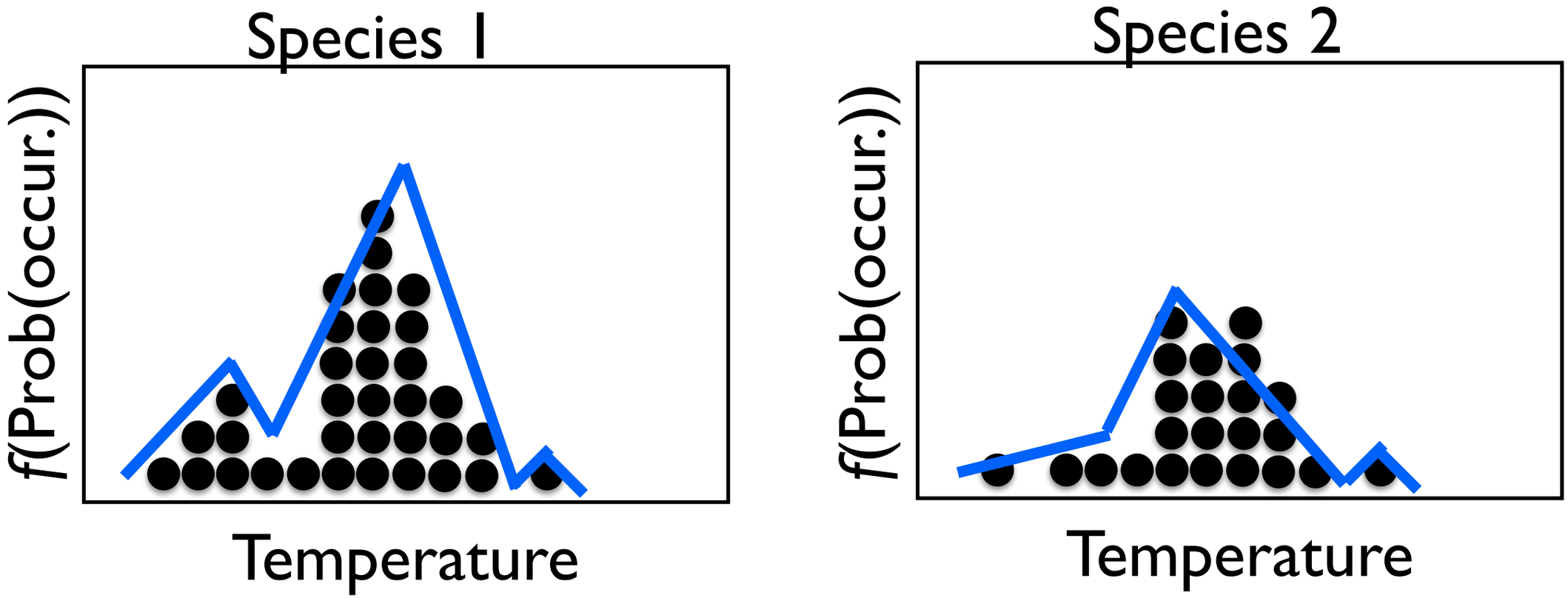
2. Ancestral states w. Brownian:



ENM+PCMs: Sampling approach

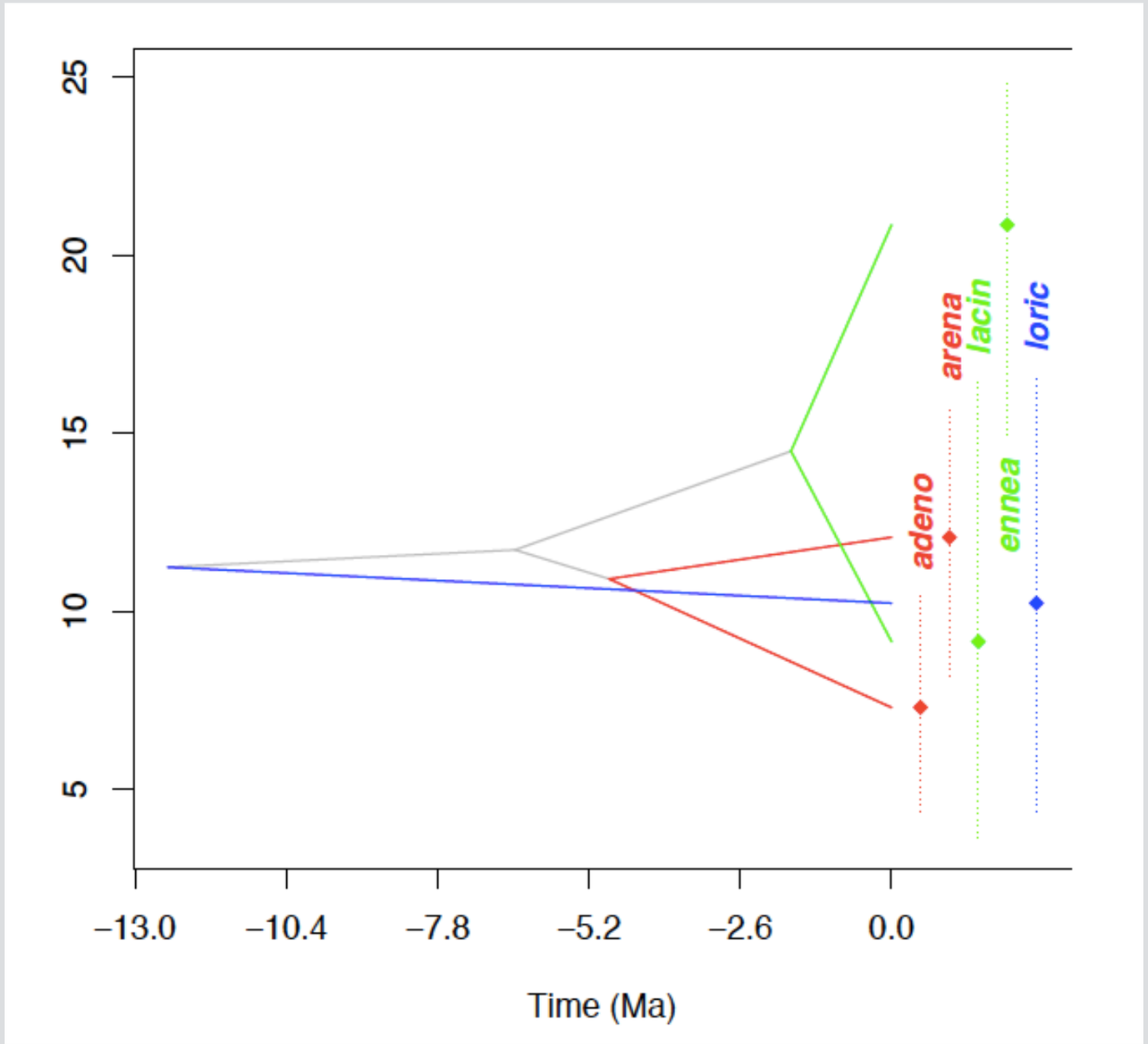
1. Fit a niche model

2a. Sample from curve:



ENM+PCMs: Sampling approach

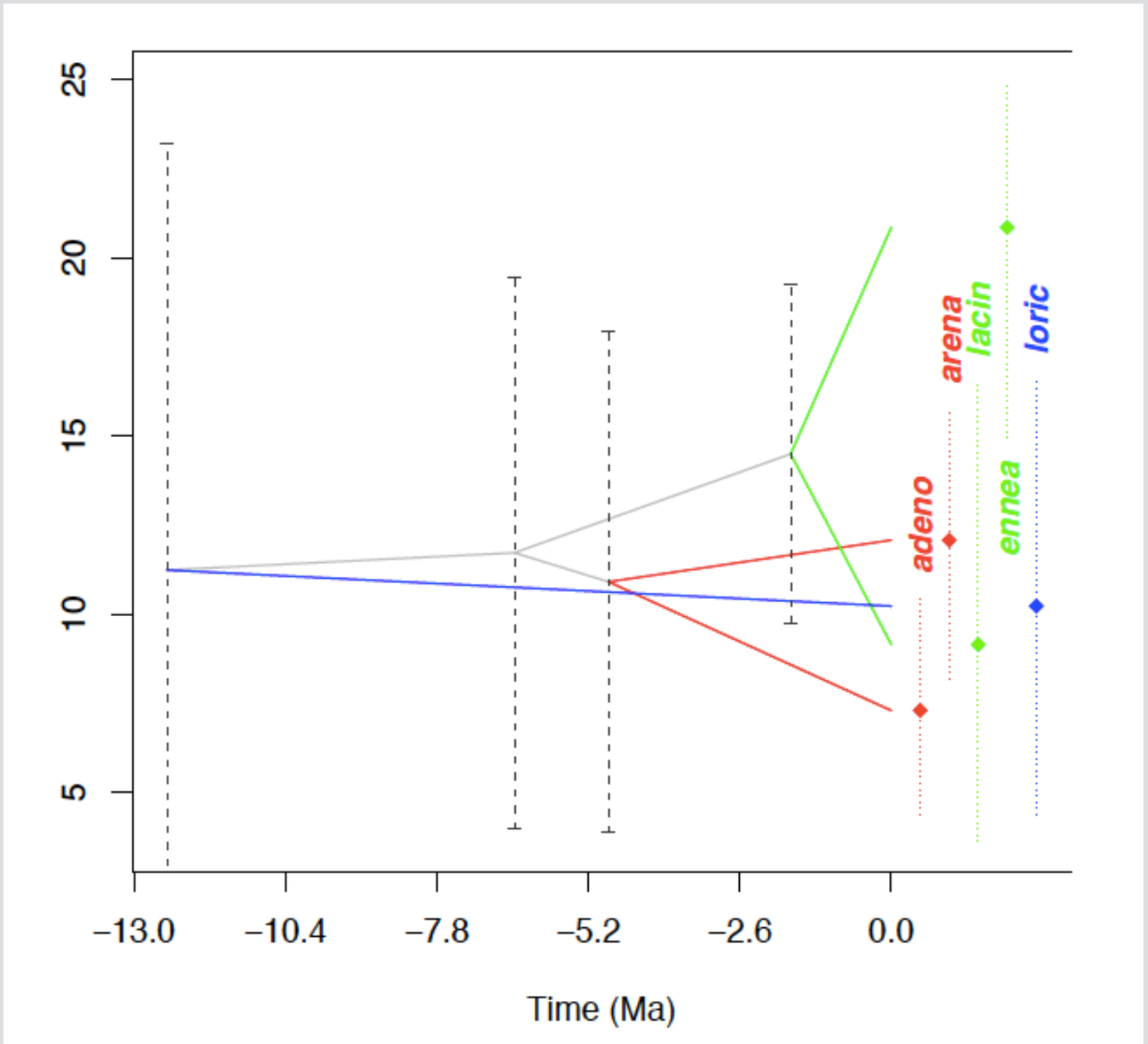
Pioneered by Evans et al. (2009); Available in phyloclim R package



(example dataset from Heibl's phyloclim R package)

ENM+PCMs: Sampling approach

Pioneered by Evans et al. (2009); Available in phyloclim R package



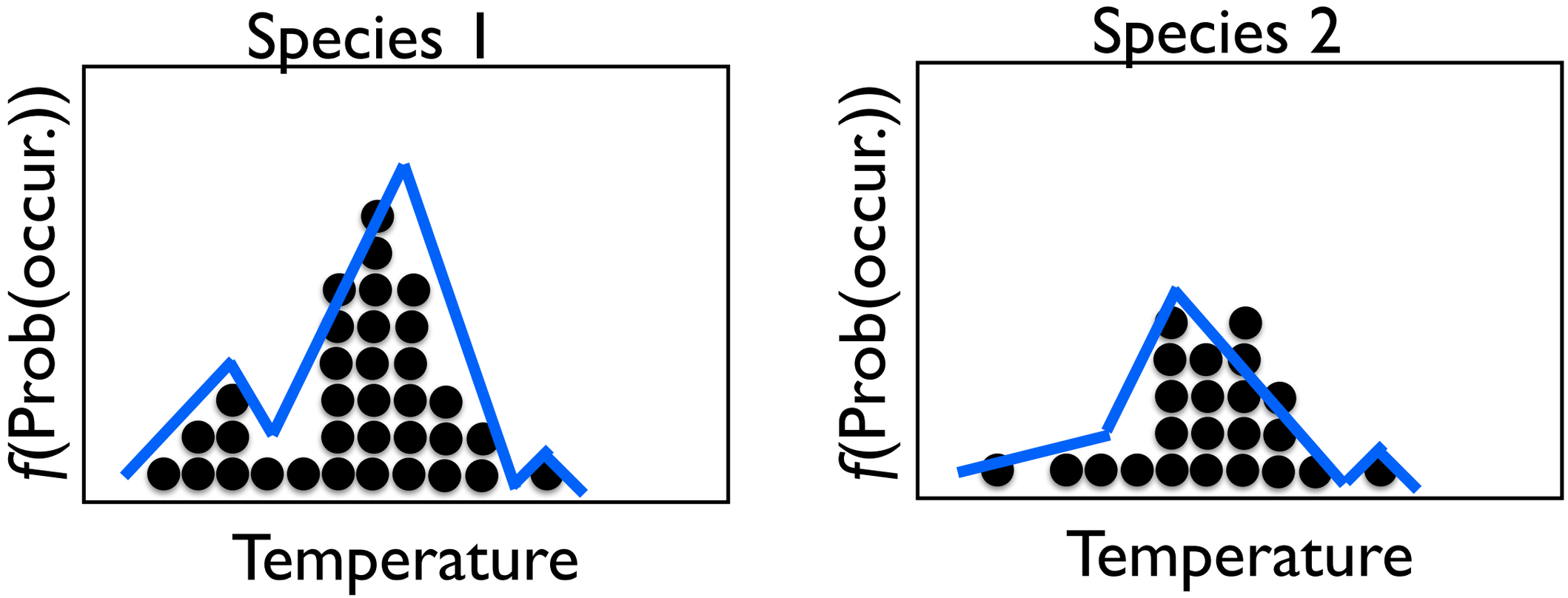
(example dataset from Heibl's phyloclim R package)

(adding the 95% CIs on ancestral means)

ENM+PCMs: Sampling approach

1. Fit a niche model

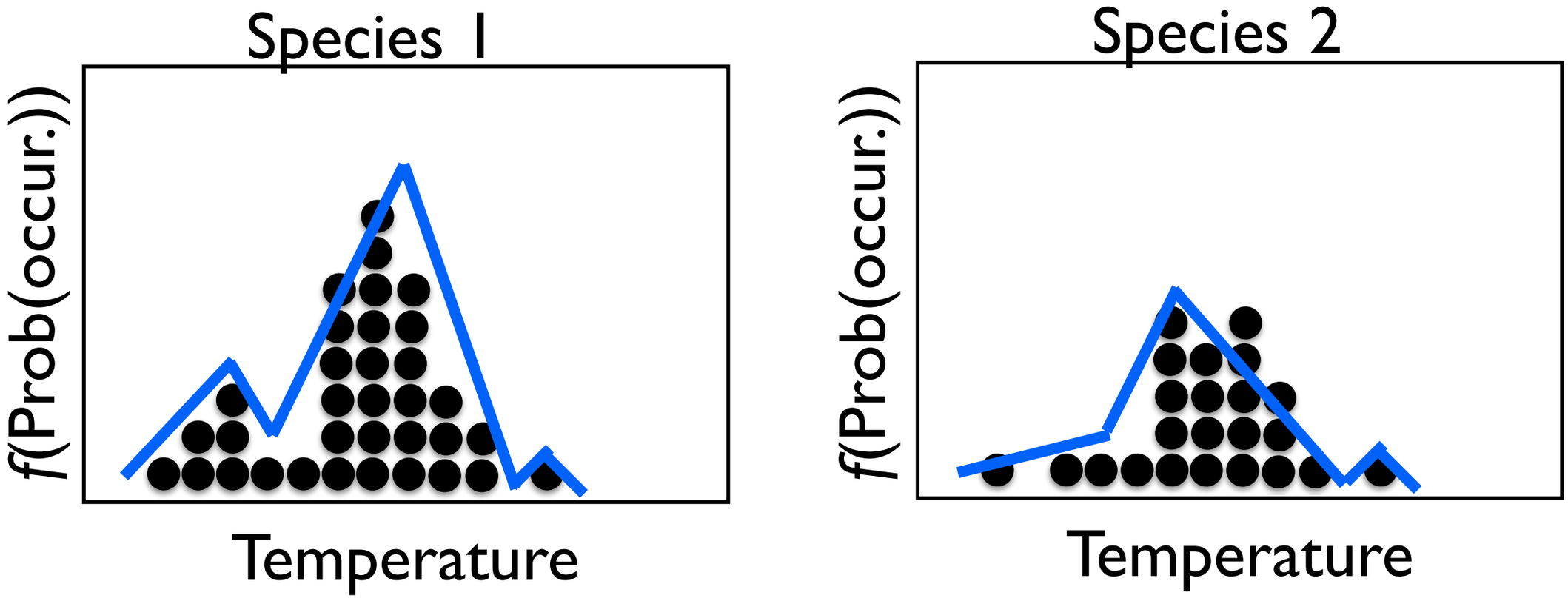
2a. Sample from curve:



ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:

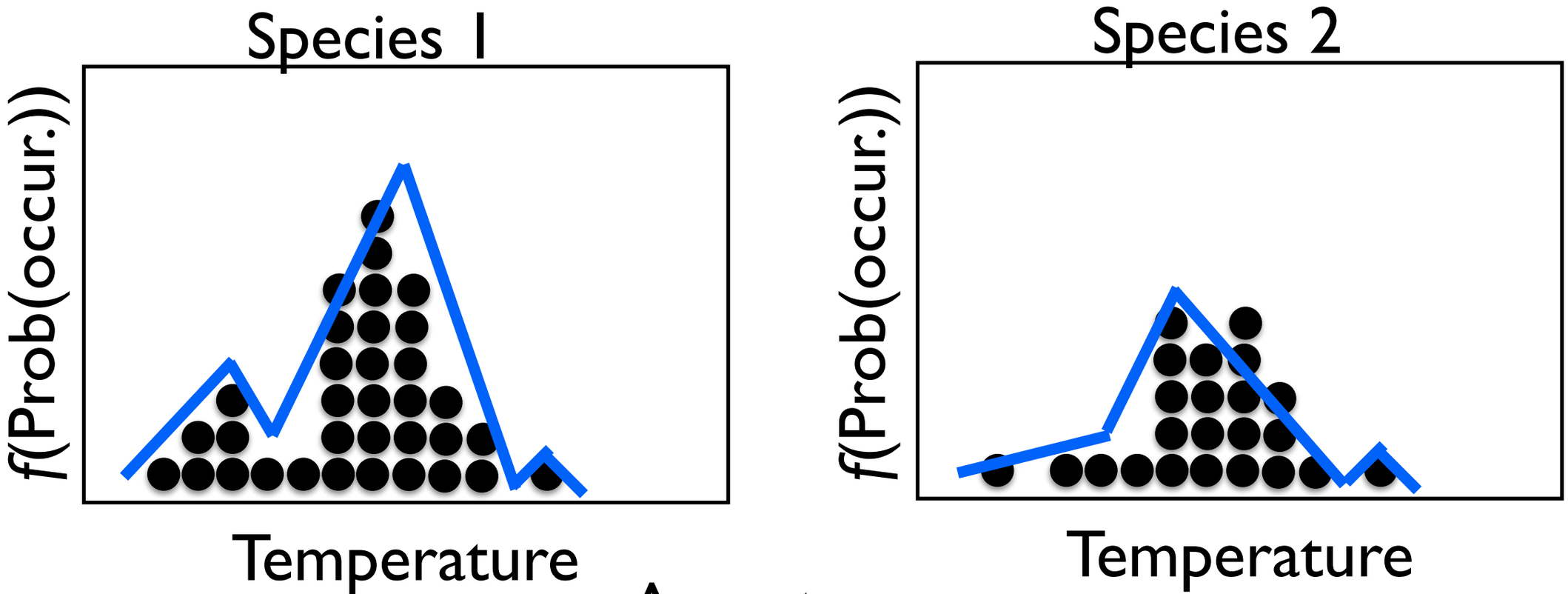


2b. Ancestral states w. Brownian:

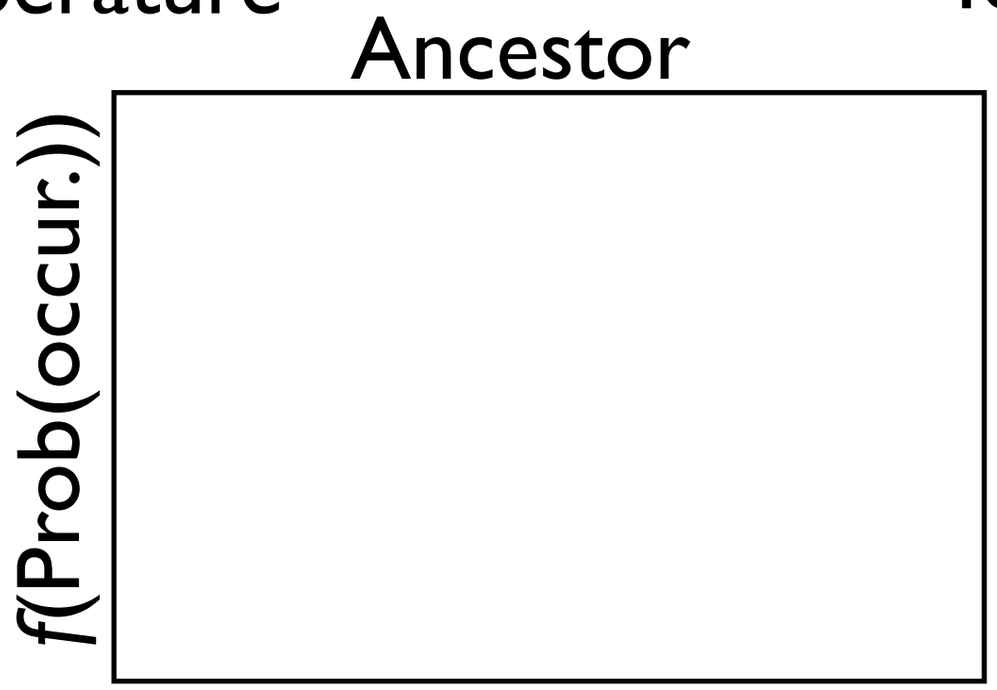
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



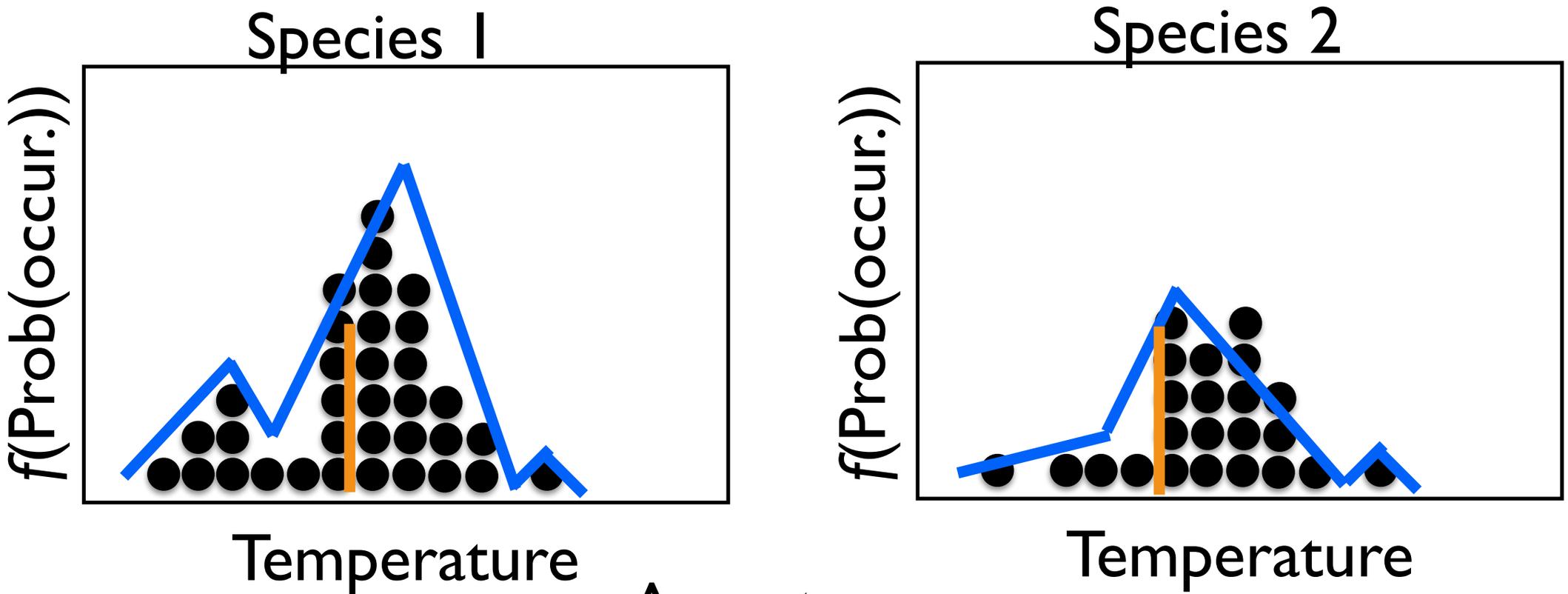
2b. Ancestral states w. Brownian:



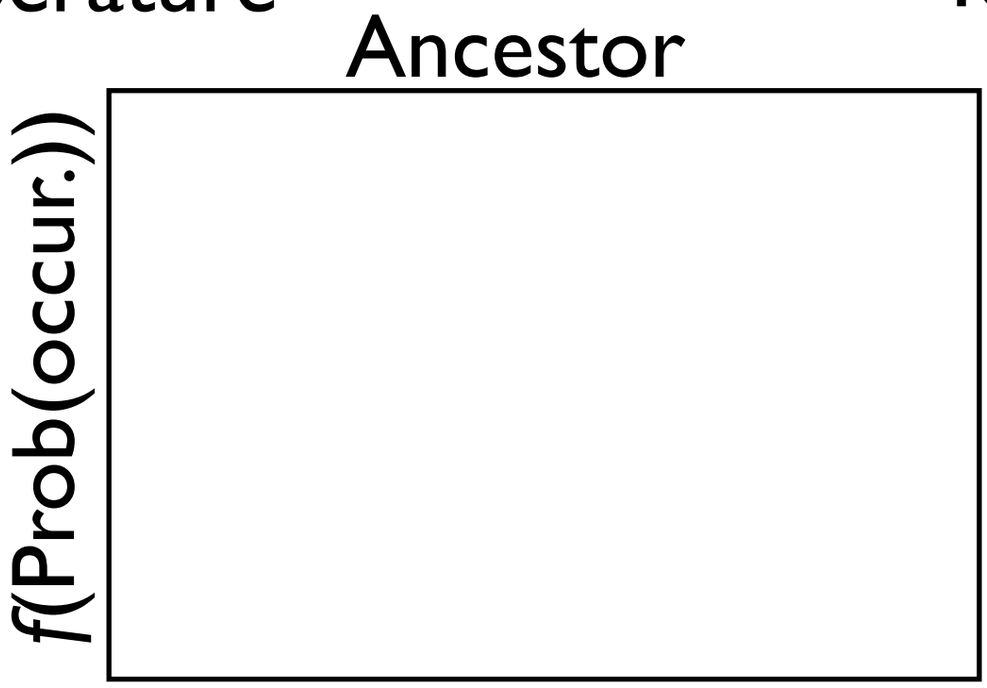
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



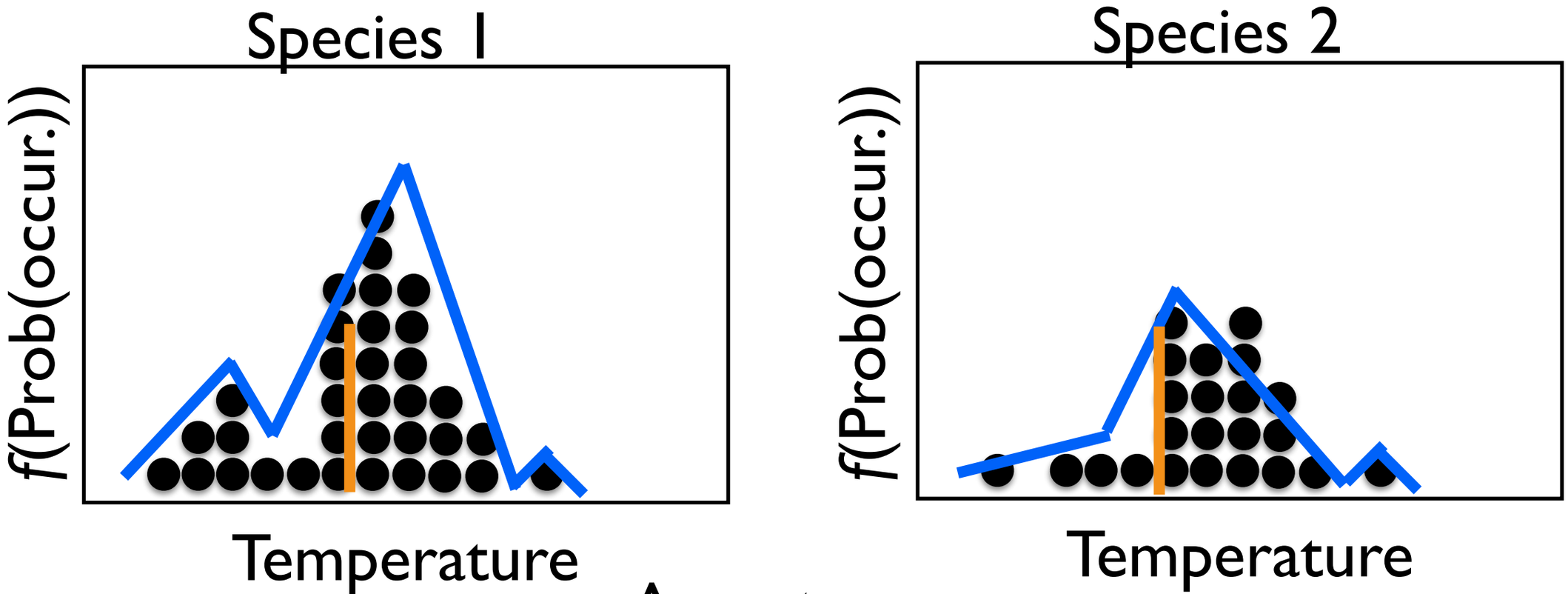
2b. Ancestral states w. Brownian:



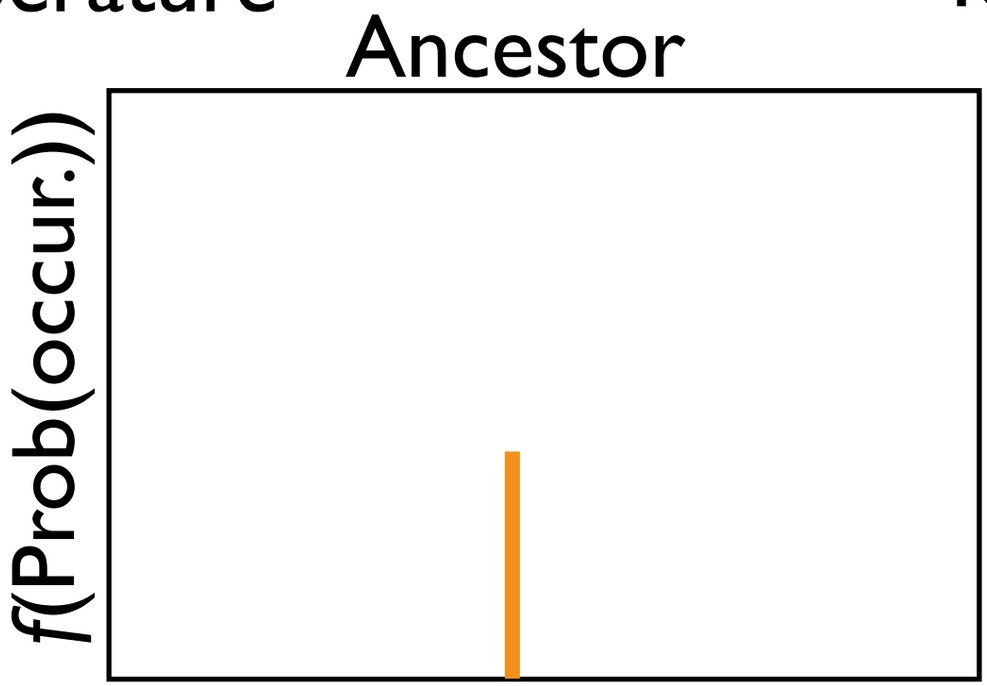
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



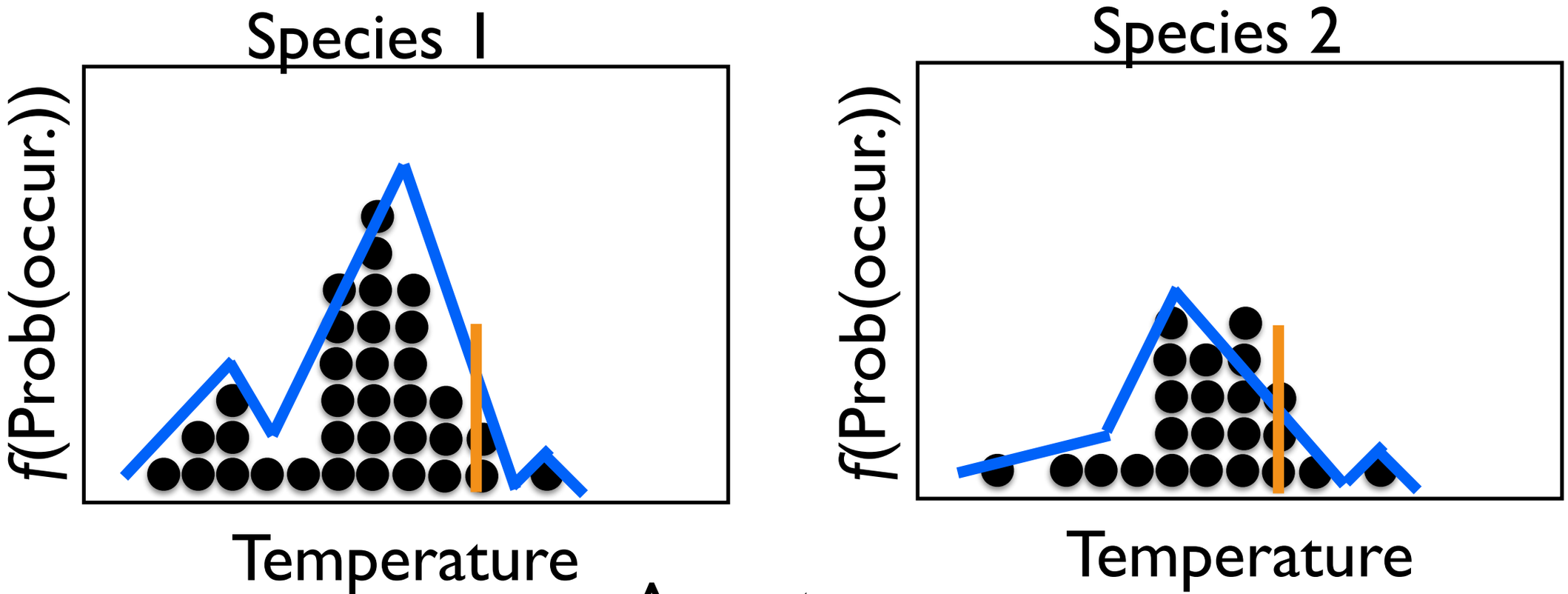
2b. Ancestral states w. Brownian:



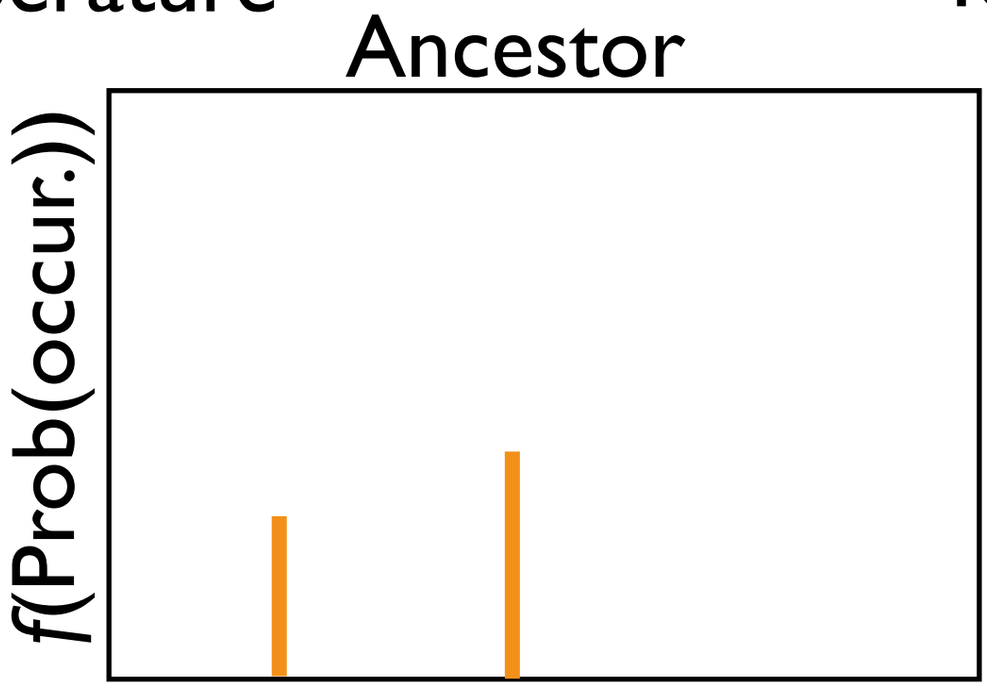
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



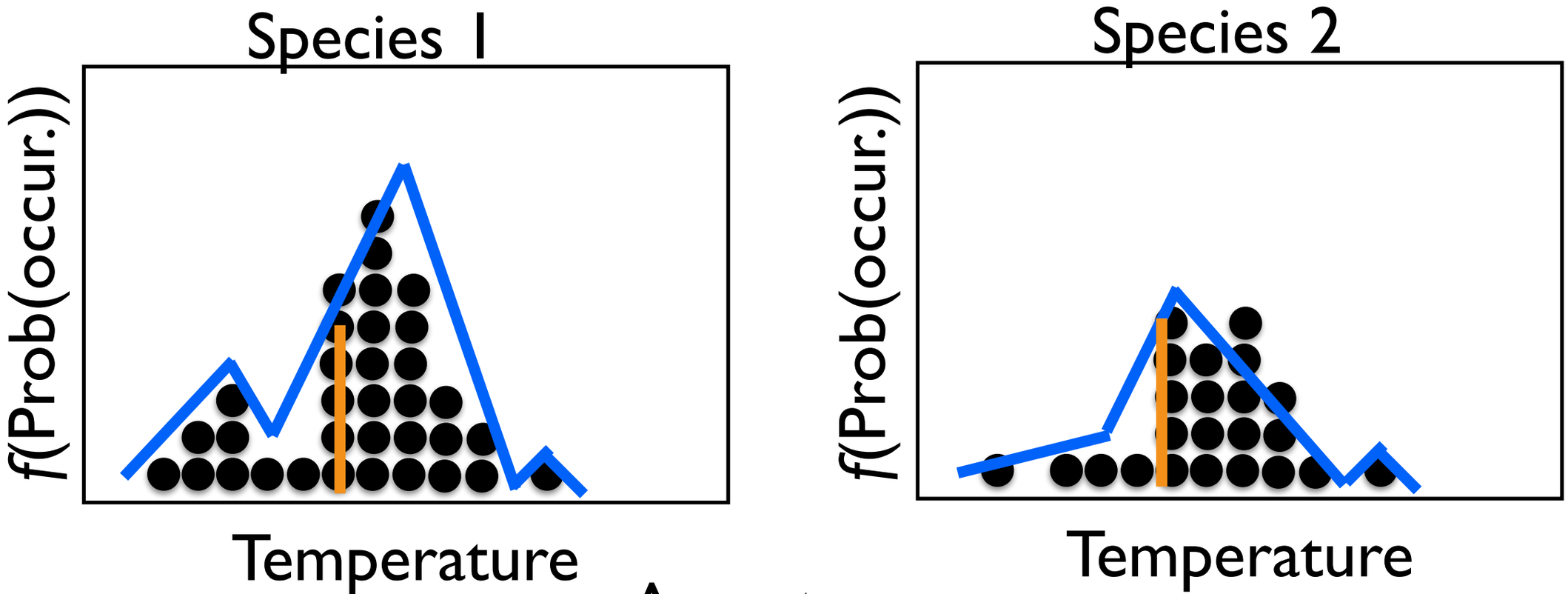
2b. Ancestral states w. Brownian:



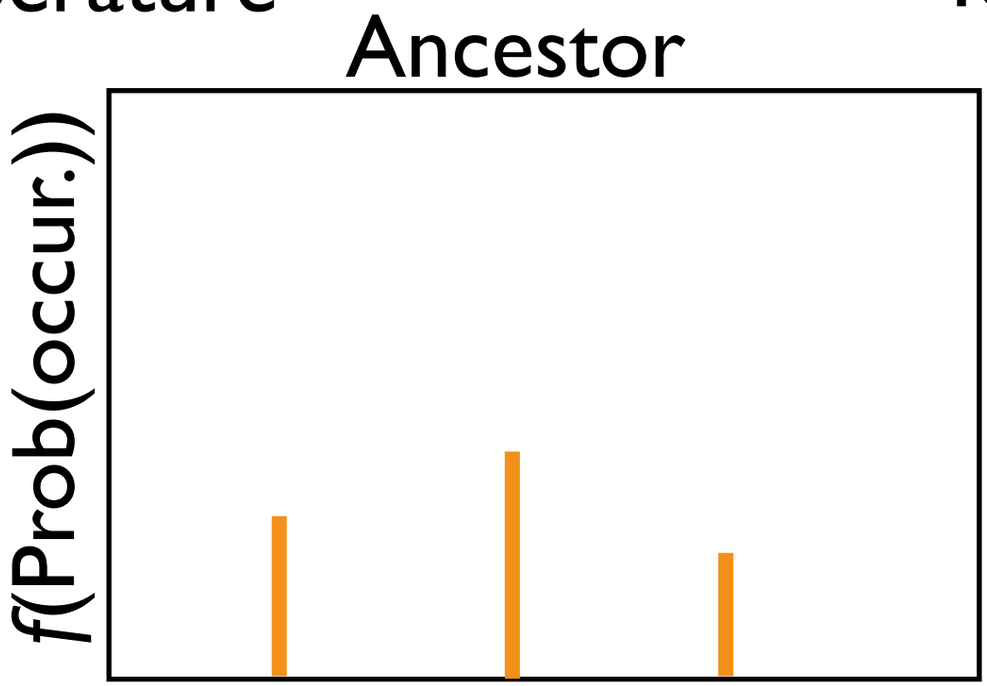
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



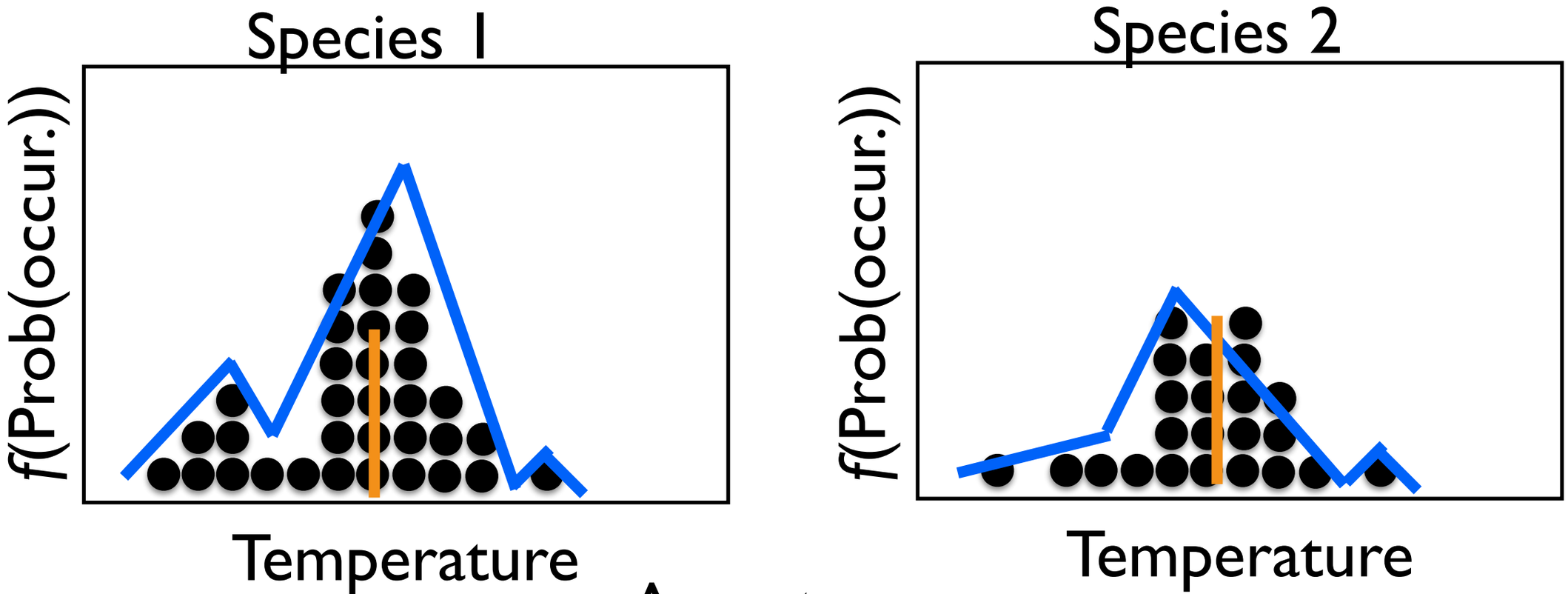
2b. Ancestral states w. Brownian:



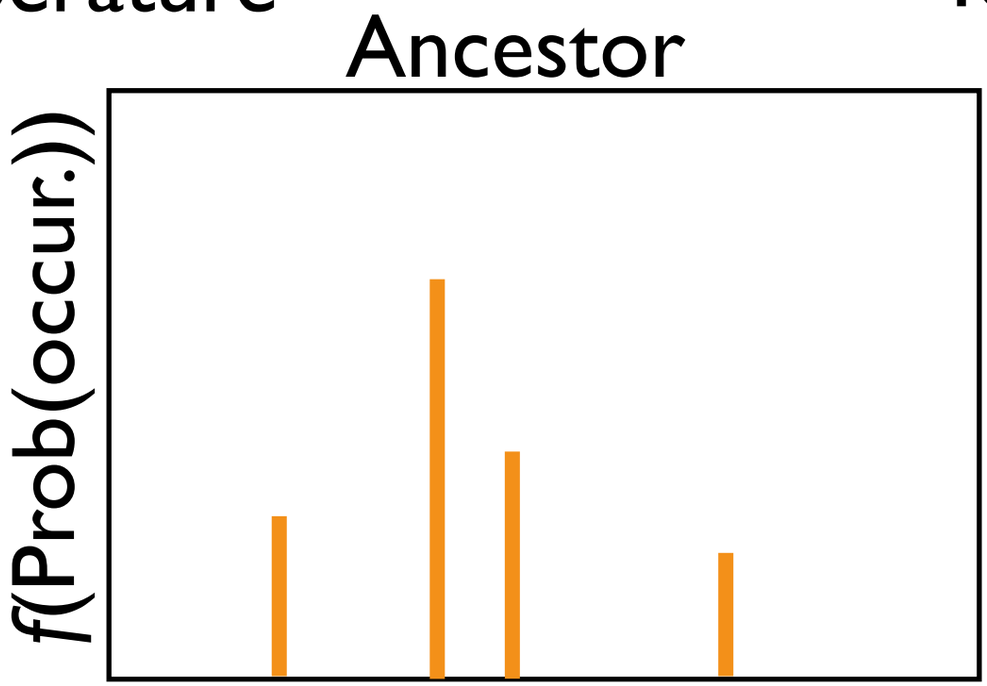
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



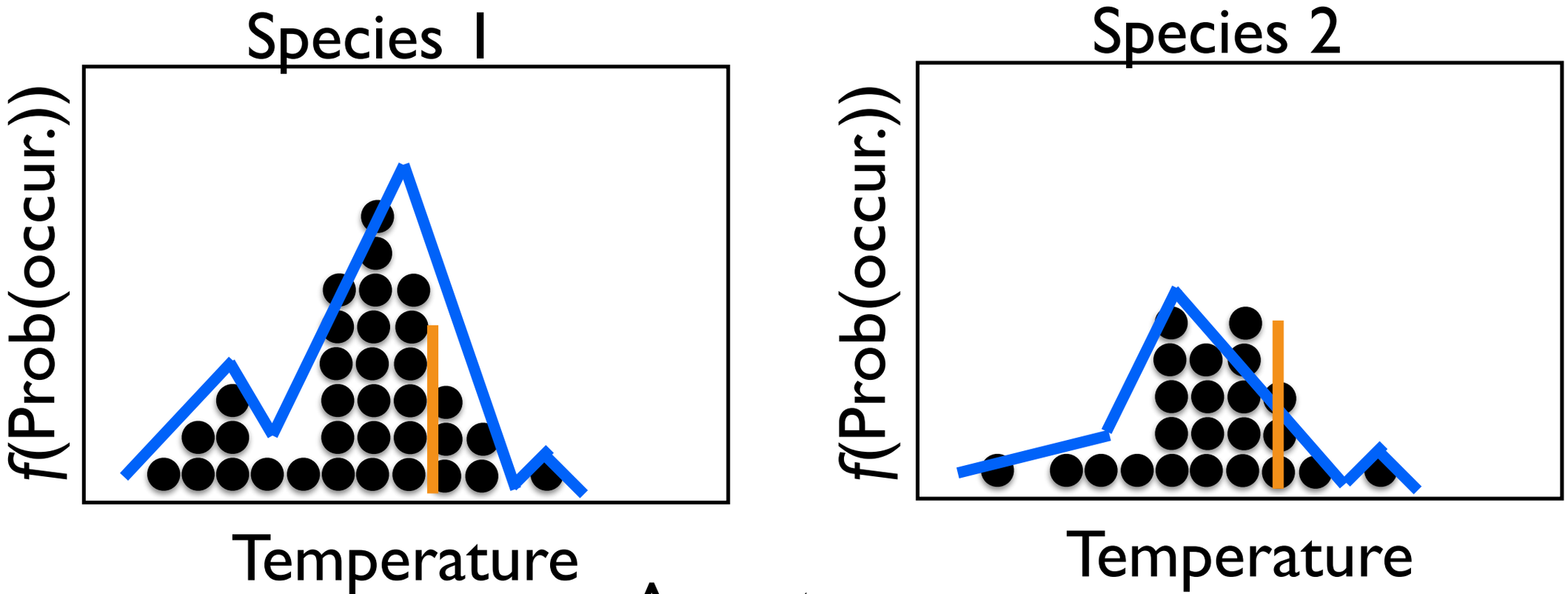
2b. Ancestral states w. Brownian:



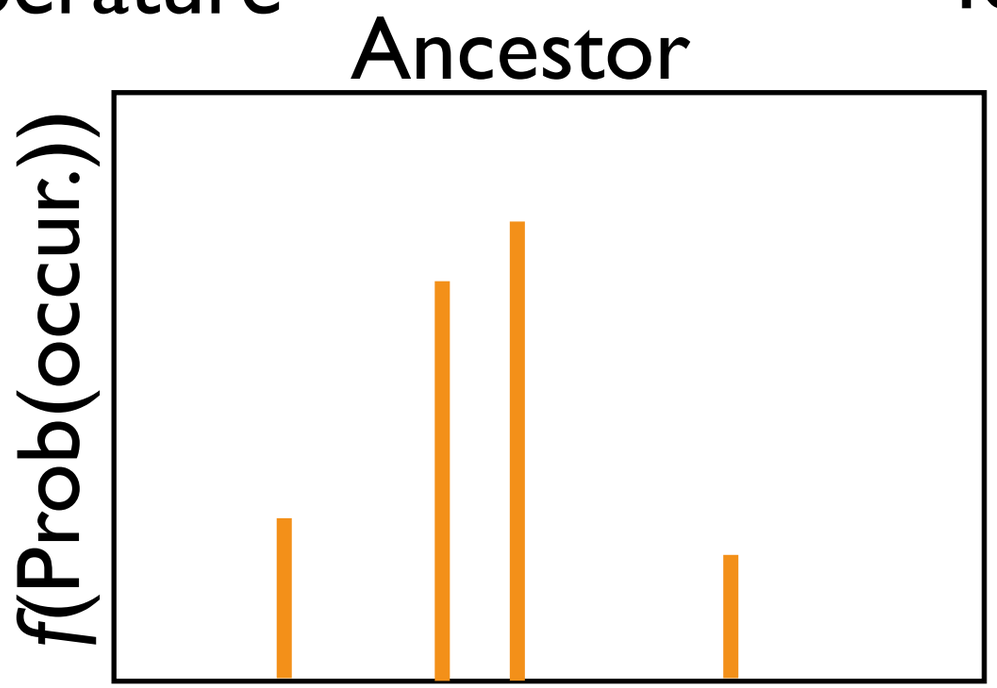
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



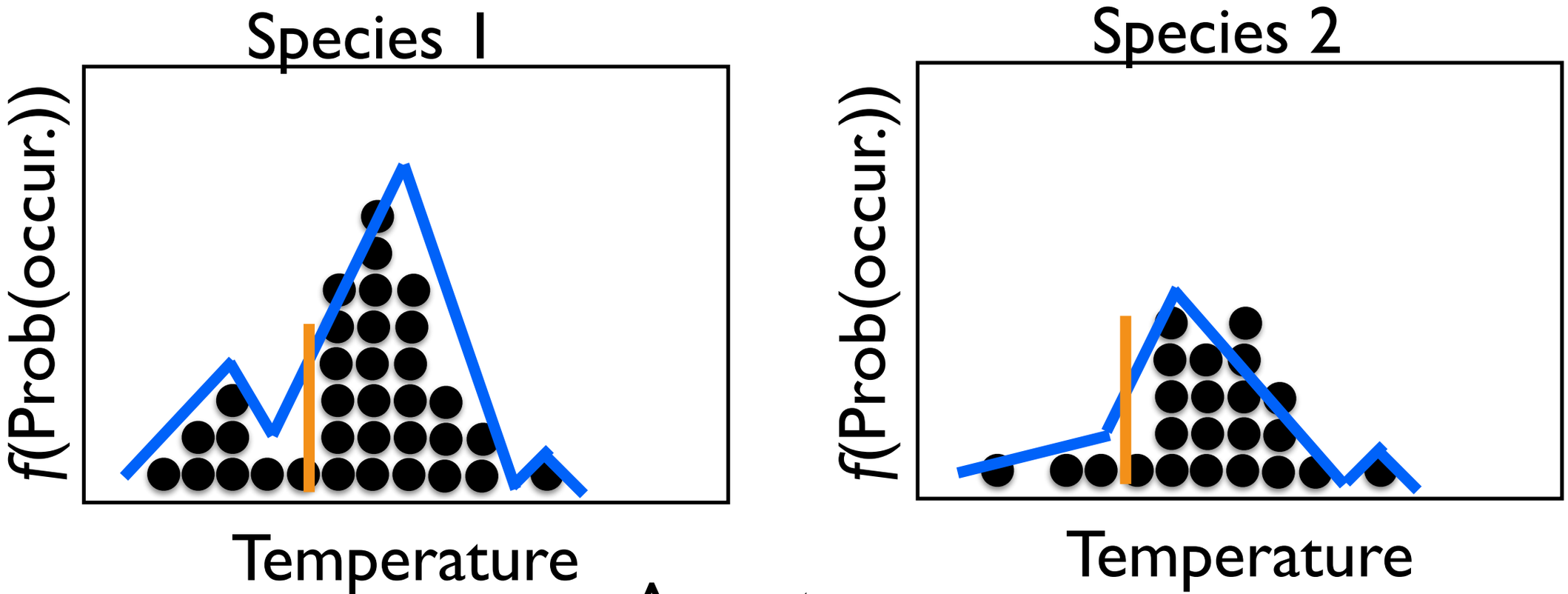
2b. Ancestral states w. Brownian:



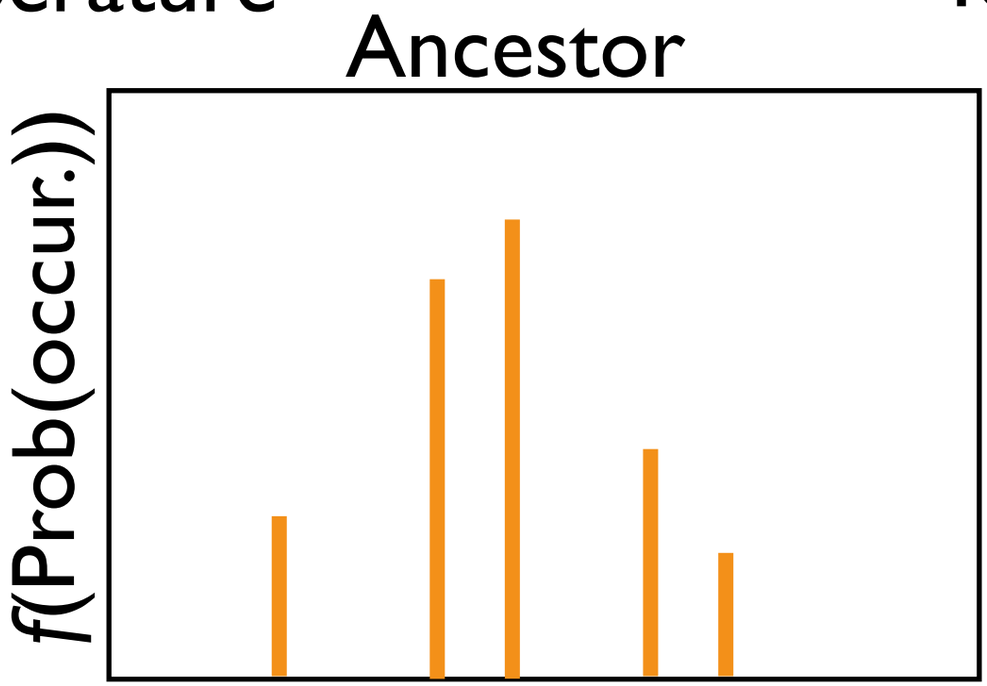
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



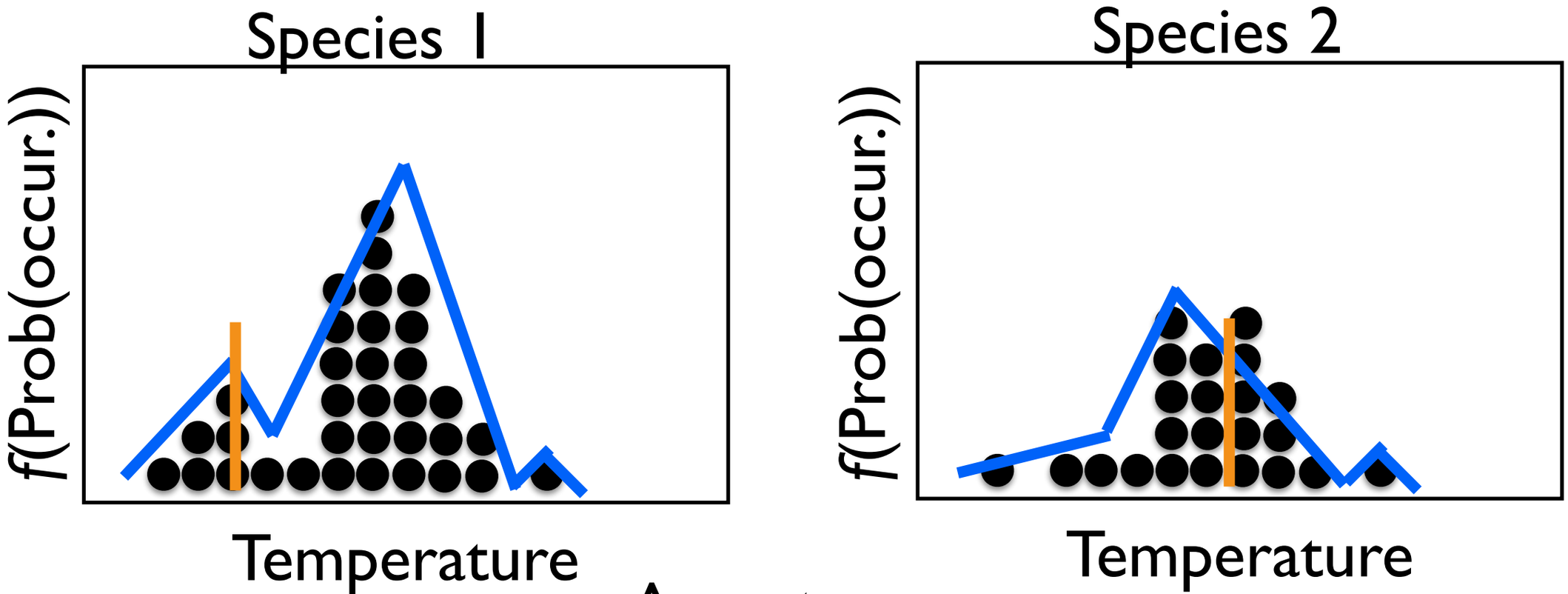
2b. Ancestral states w. Brownian:



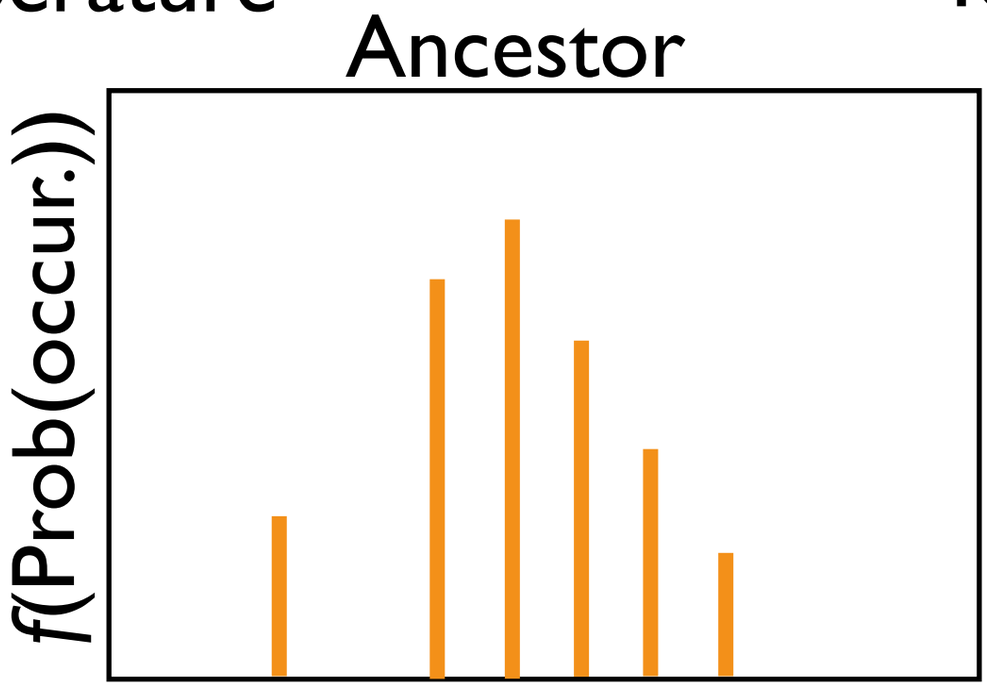
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



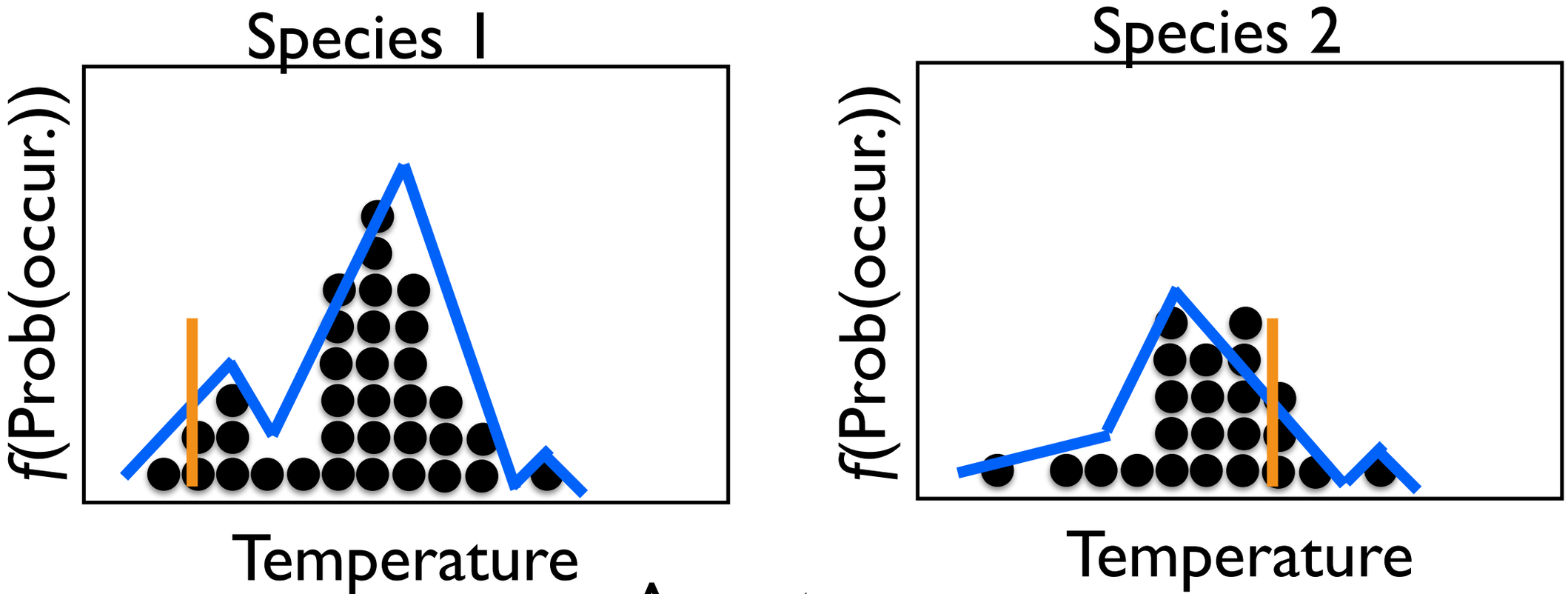
2b. Ancestral states w. Brownian:



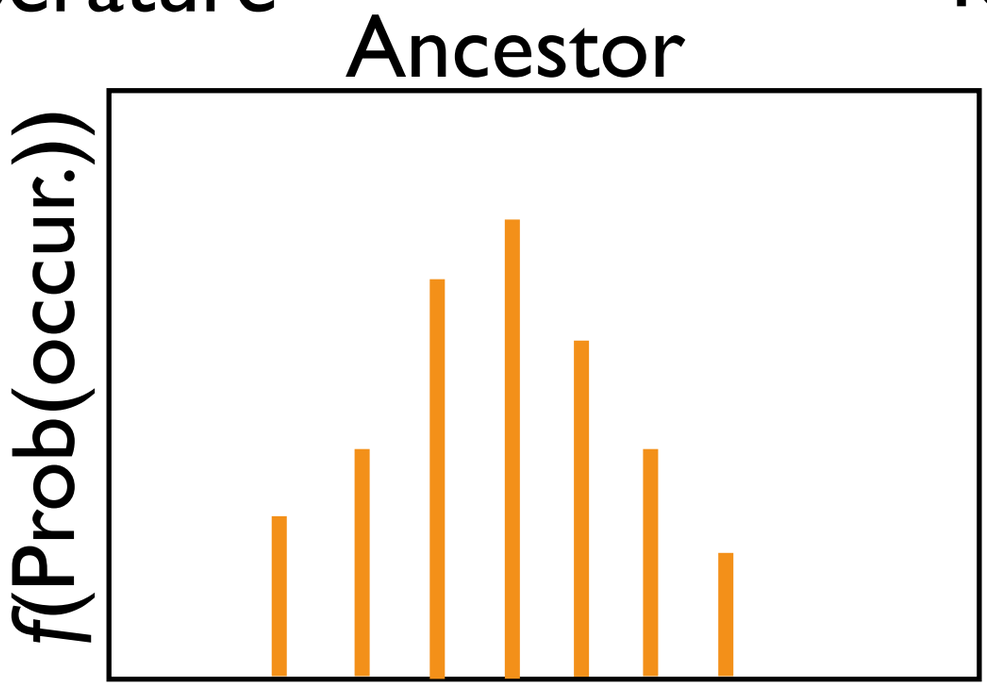
ENM+PCMs: Sampling approach

1. Fit a niche model

2a. Sample from curve:



2b. Ancestral states w. Brownian:

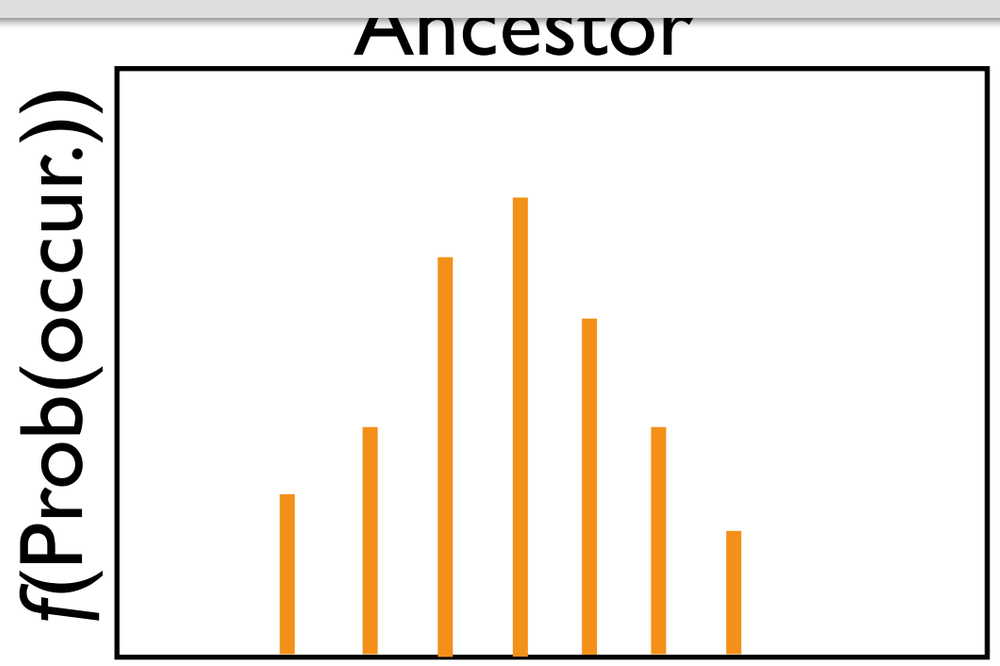


ENM+PCMs: Sampling approach

Critiques:

1. Niche curves are inferred estimates, not data
2. Sampling from the curves adds variance
3. No model choice

2b. Ancestral states w. Brownian:



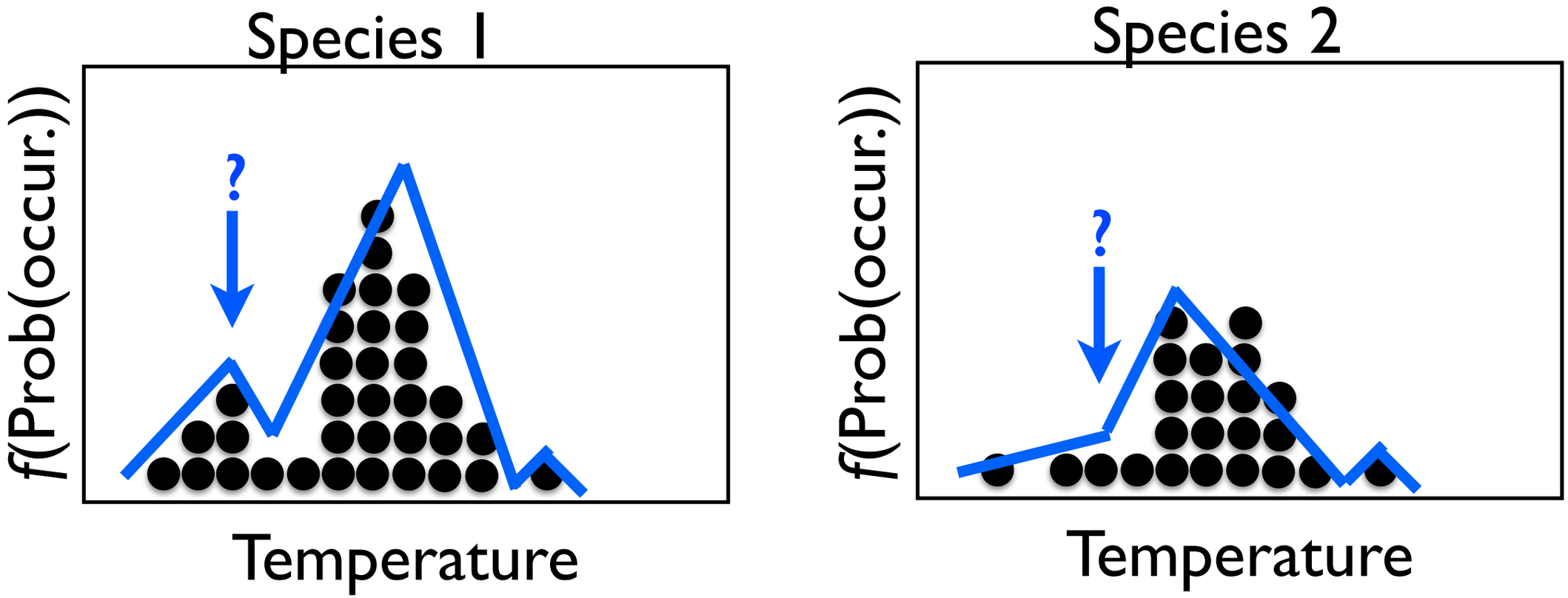
Combining ENMs and PCMs

Combining ENMs and PCMs

How *should* we link ENMs/SDMs and PCMs?

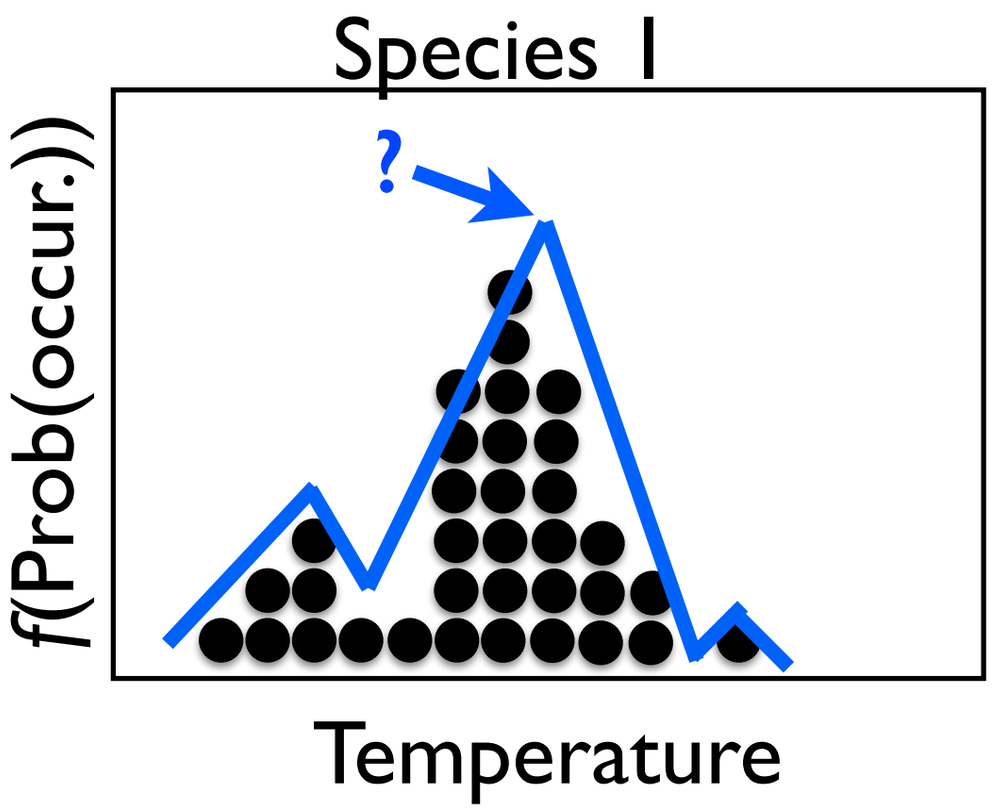
How *should* we link ENMs/SDMs and PCMs?

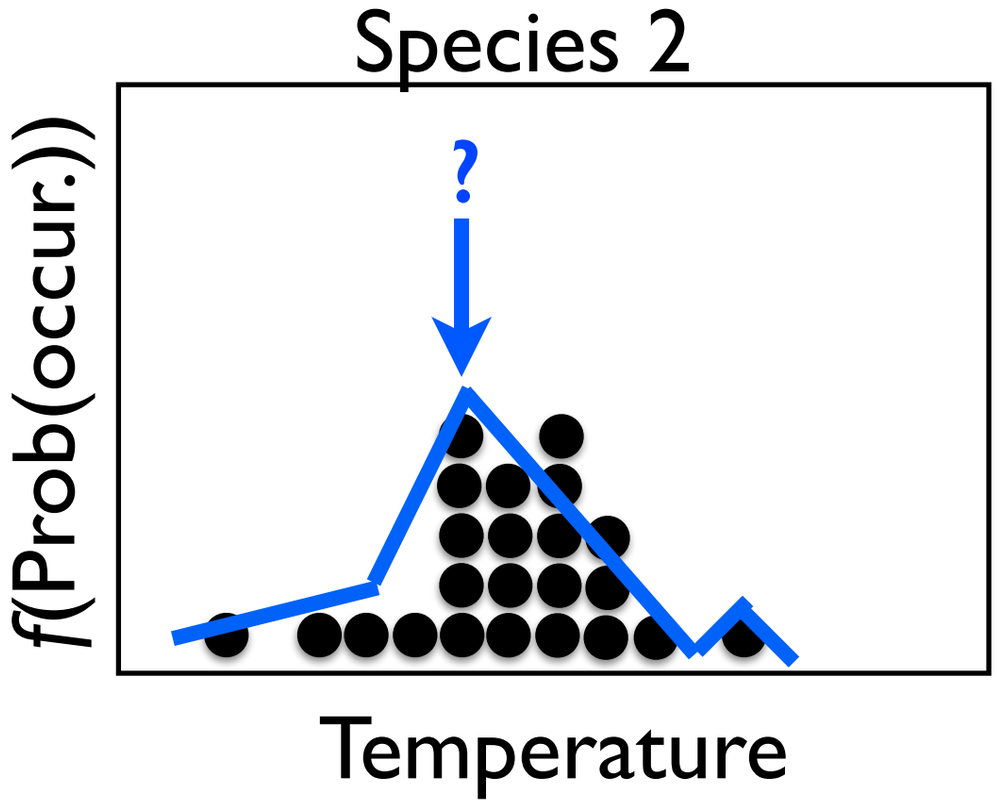
Unsolved problem: how are niches *homologous?*



How *should* we link ENMs/SDMs and PCMs?

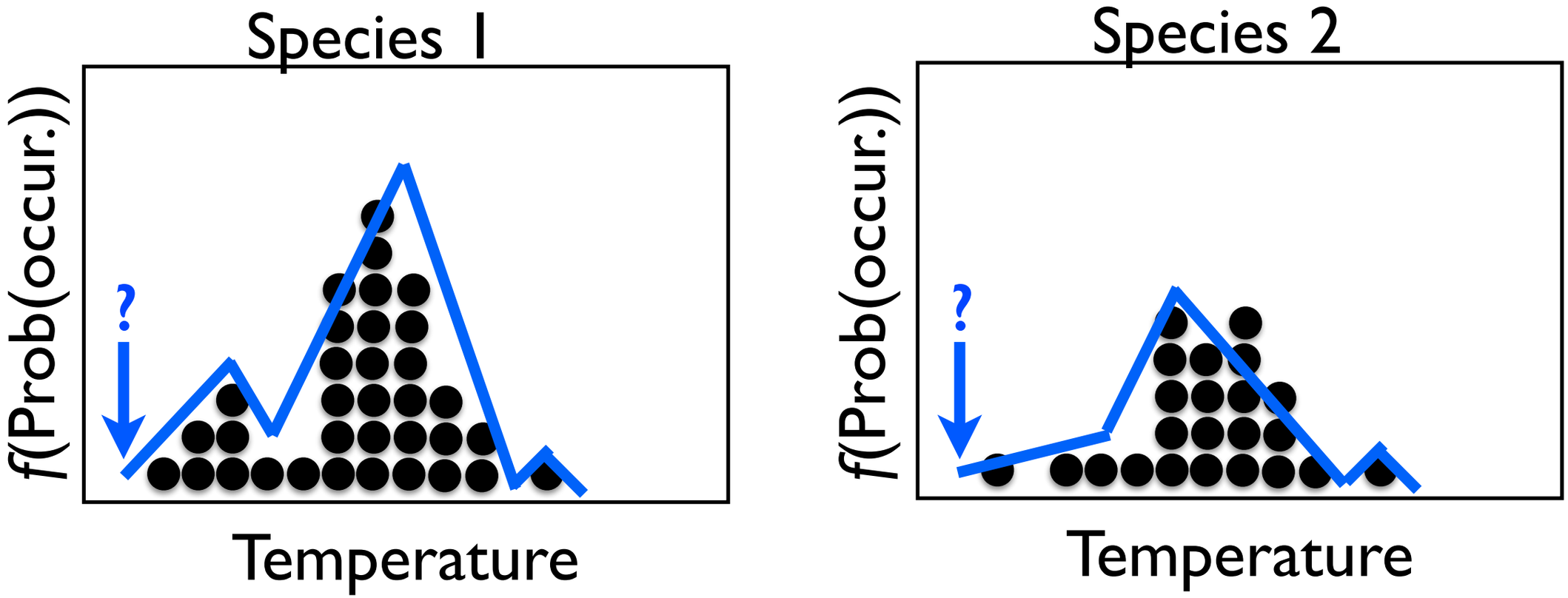
Unsolved problem: how are niches *homologous?*





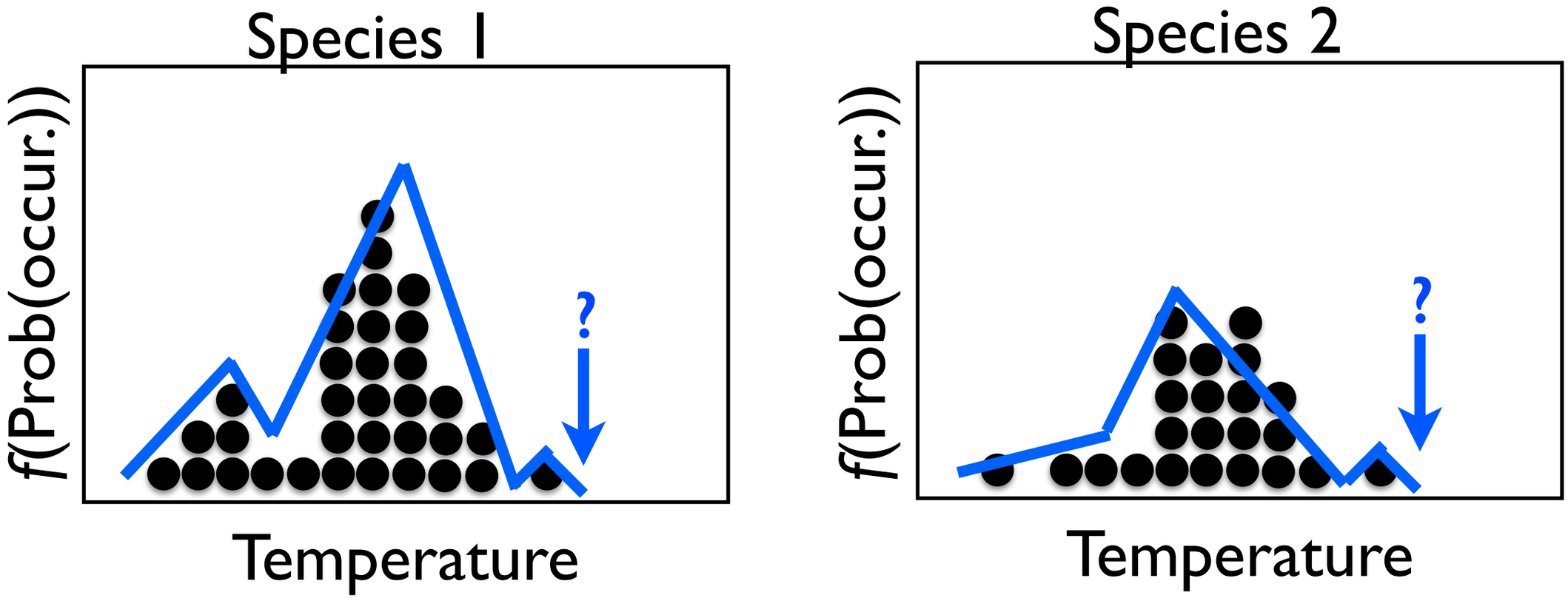
How *should* we link ENMs/SDMs and PCMs?

Unsolved problem: how are niches *homologous?*



How *should* we link ENMs/SDMs and PCMs?

Unsolved problem: how are niches *homologous?*



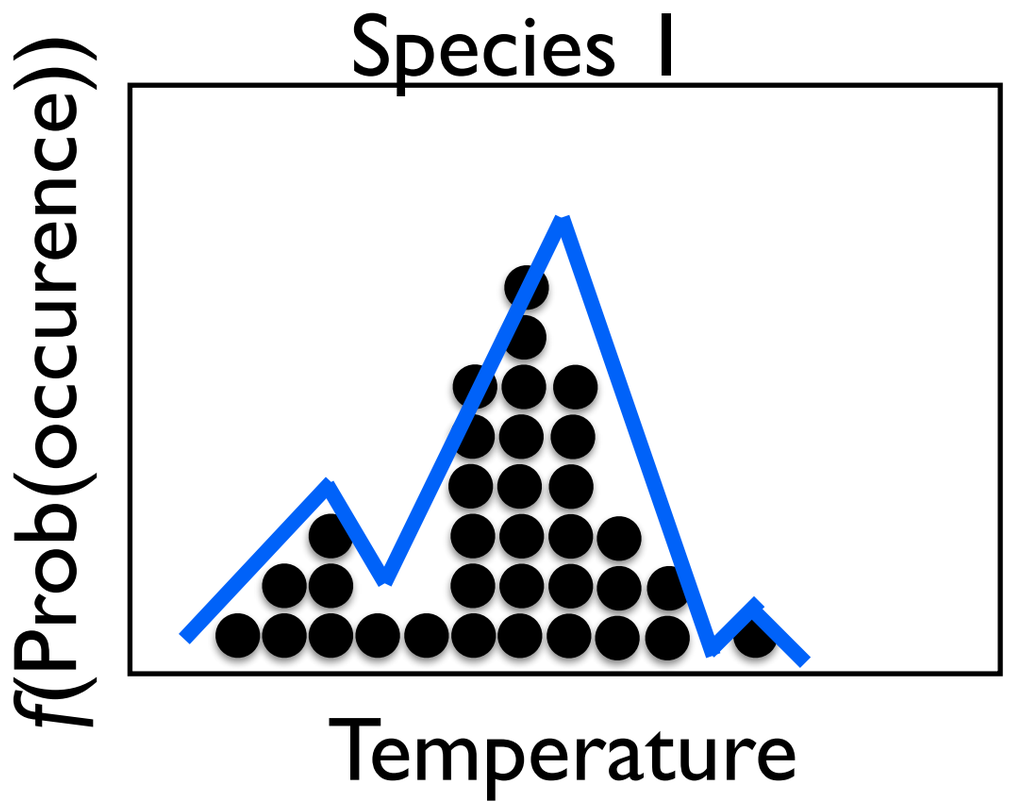
How *should* we link ENMs/SDMs and PCMs?

Homology problem: standard Niche Modelling tools won't give identical niche models between methods

…or even between random number seeds

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest



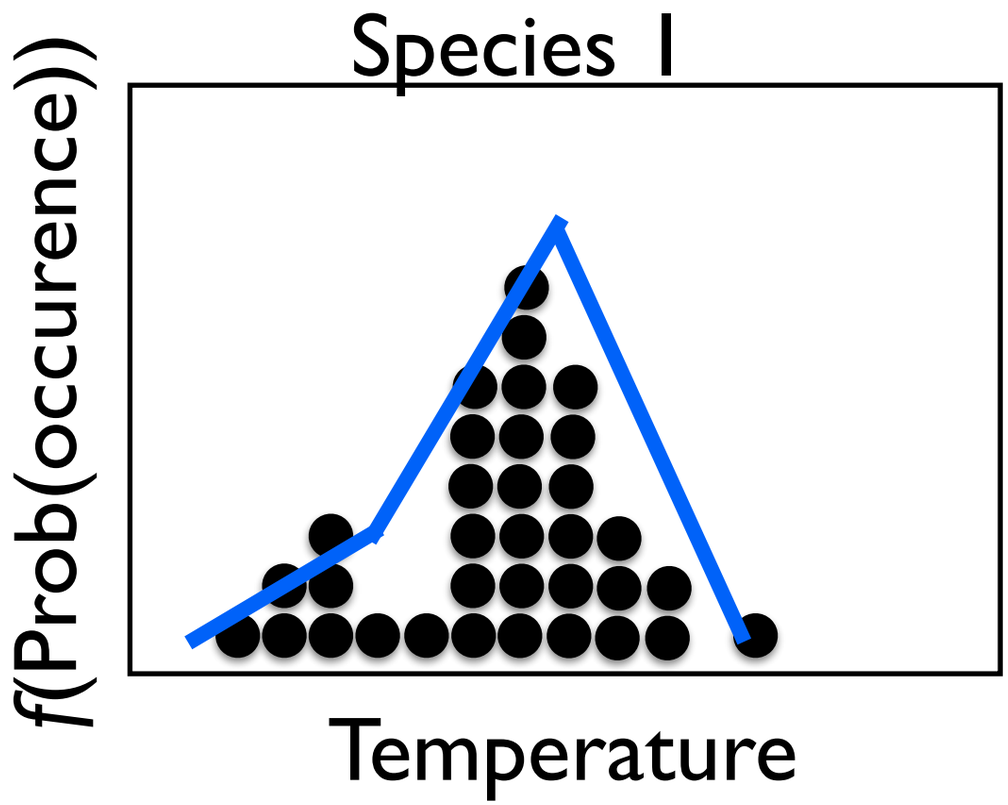
How *should* we link ENMs/SDMs and PCMs?

Homology problem: standard Niche Modelling tools won't give identical niche models between methods

…or even between random number seeds

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest



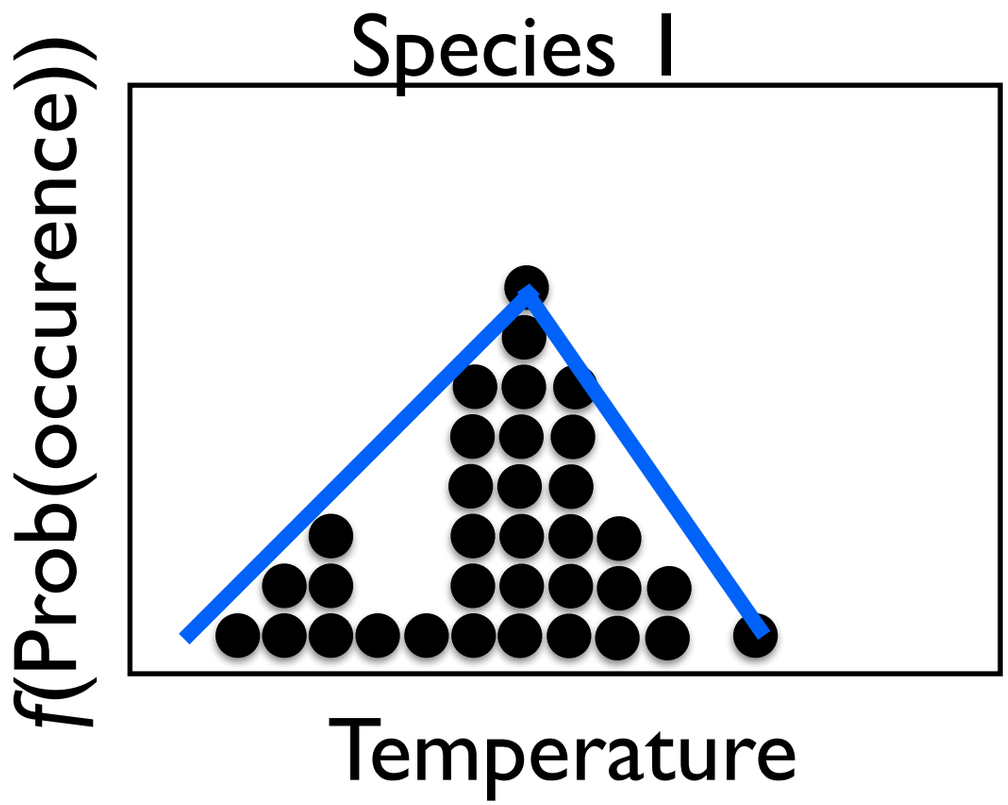
How *should* we link ENMs/SDMs and PCMs?

Homology problem: standard Niche Modelling tools won't give identical niche models between methods

…or even between random number seeds

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest



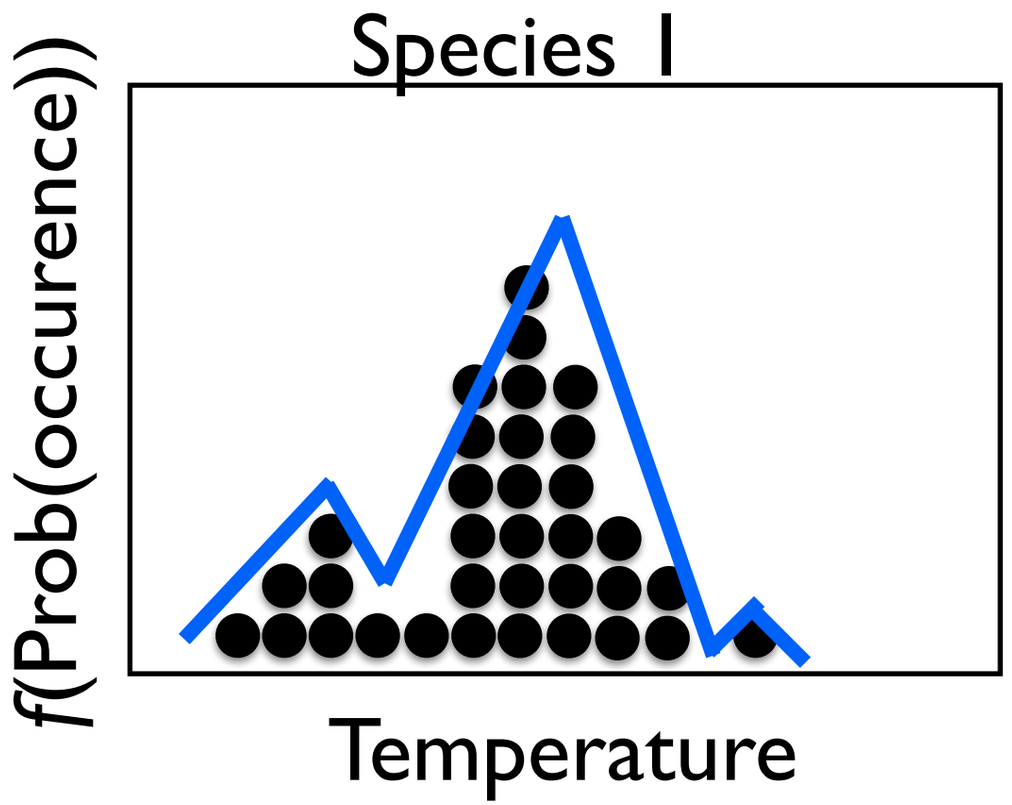
How *should* we link ENMs/SDMs and PCMs?

Homology problem: standard Niche Modelling tools won't give identical niche models between methods

…or even between random number seeds

Common ENM tools

- BioClim (min/max)
- GLMs
- GAMs
- MaxEnt
- MARS
- RandomForest



Ho ls?

Ho lling

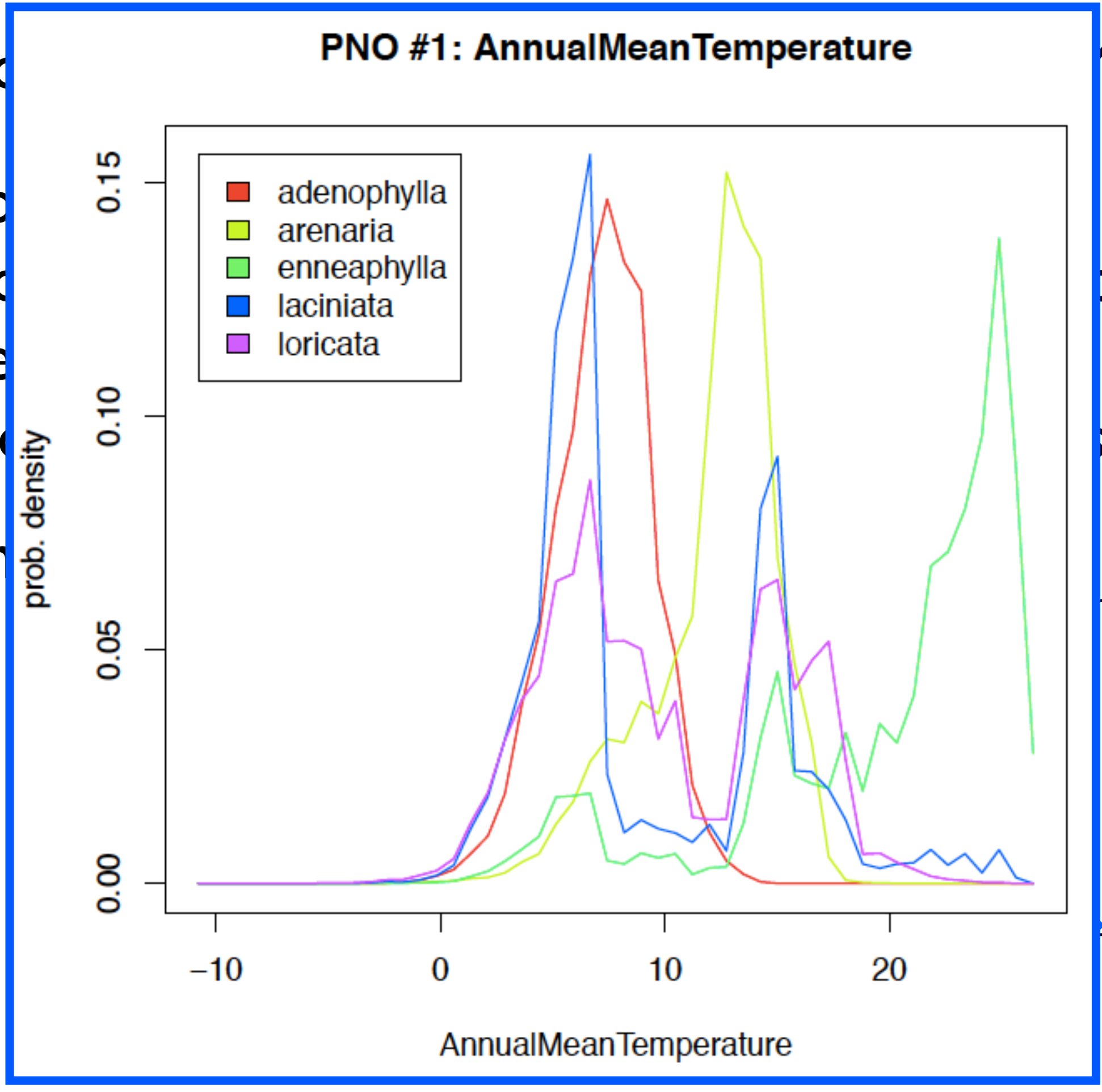
too tween

me

…

Cor

-
-
-
-
-
-



(example dataset from Heibl's phyloclim R package)

Ho ls?

Ho lling

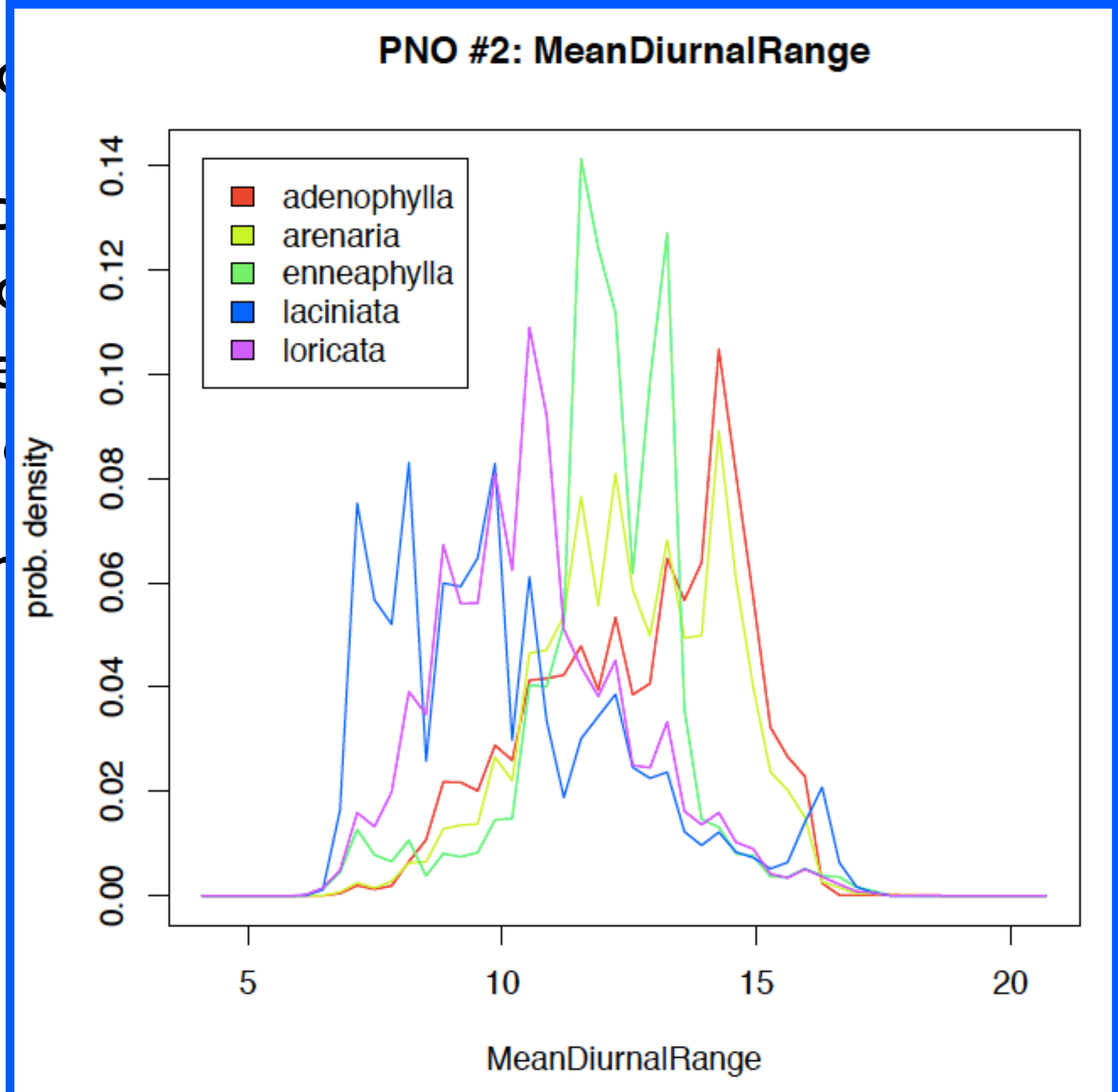
too tween

me

…

Cor

-
-
-
-
-
-



(example dataset from Heibl's phyloclim R package)

Ho ls?

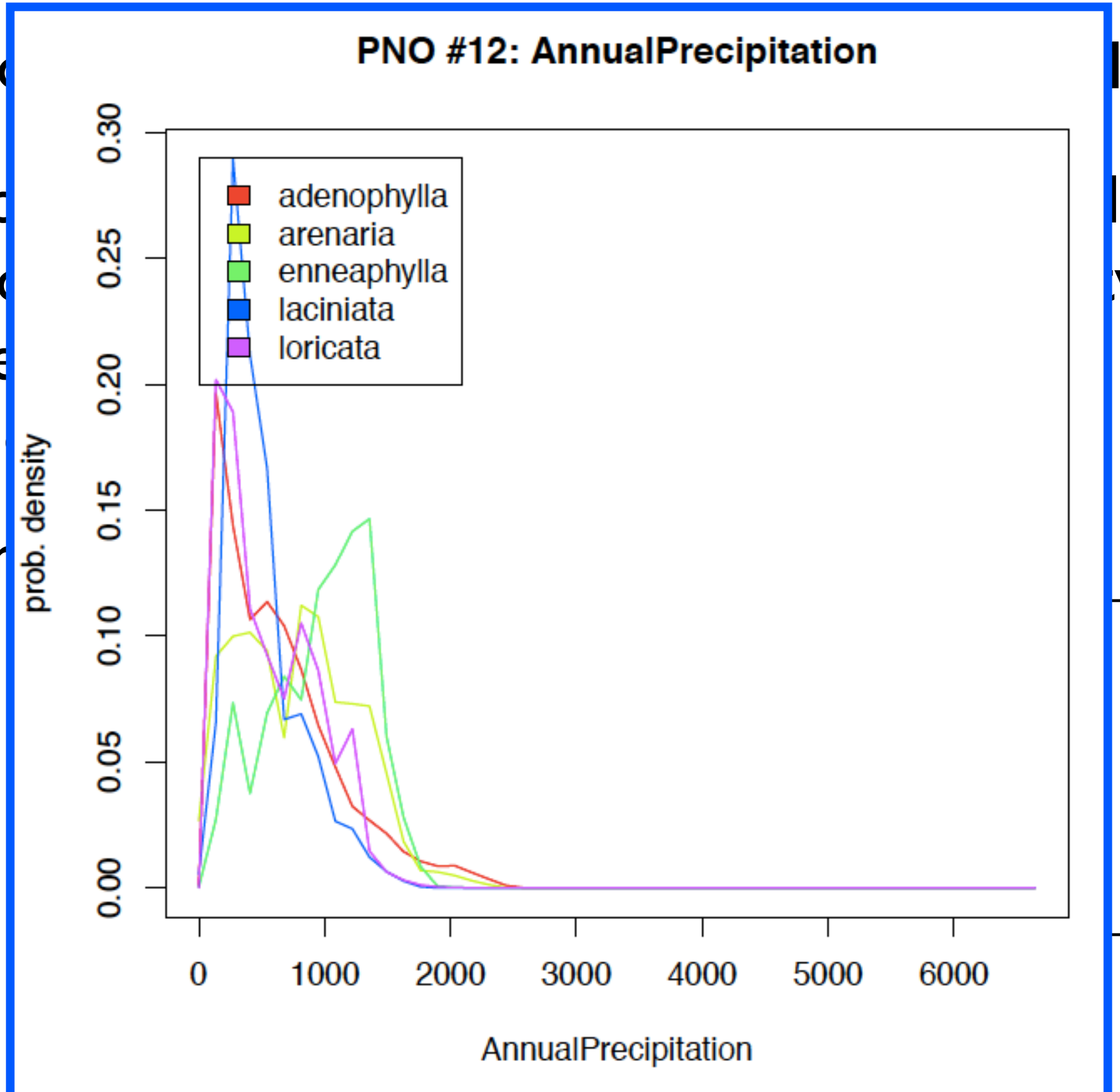
Ho lling

too tween

me

…

Cor



(example dataset from Heibl's phyloclim R package)

Ho ls?

Ho lling

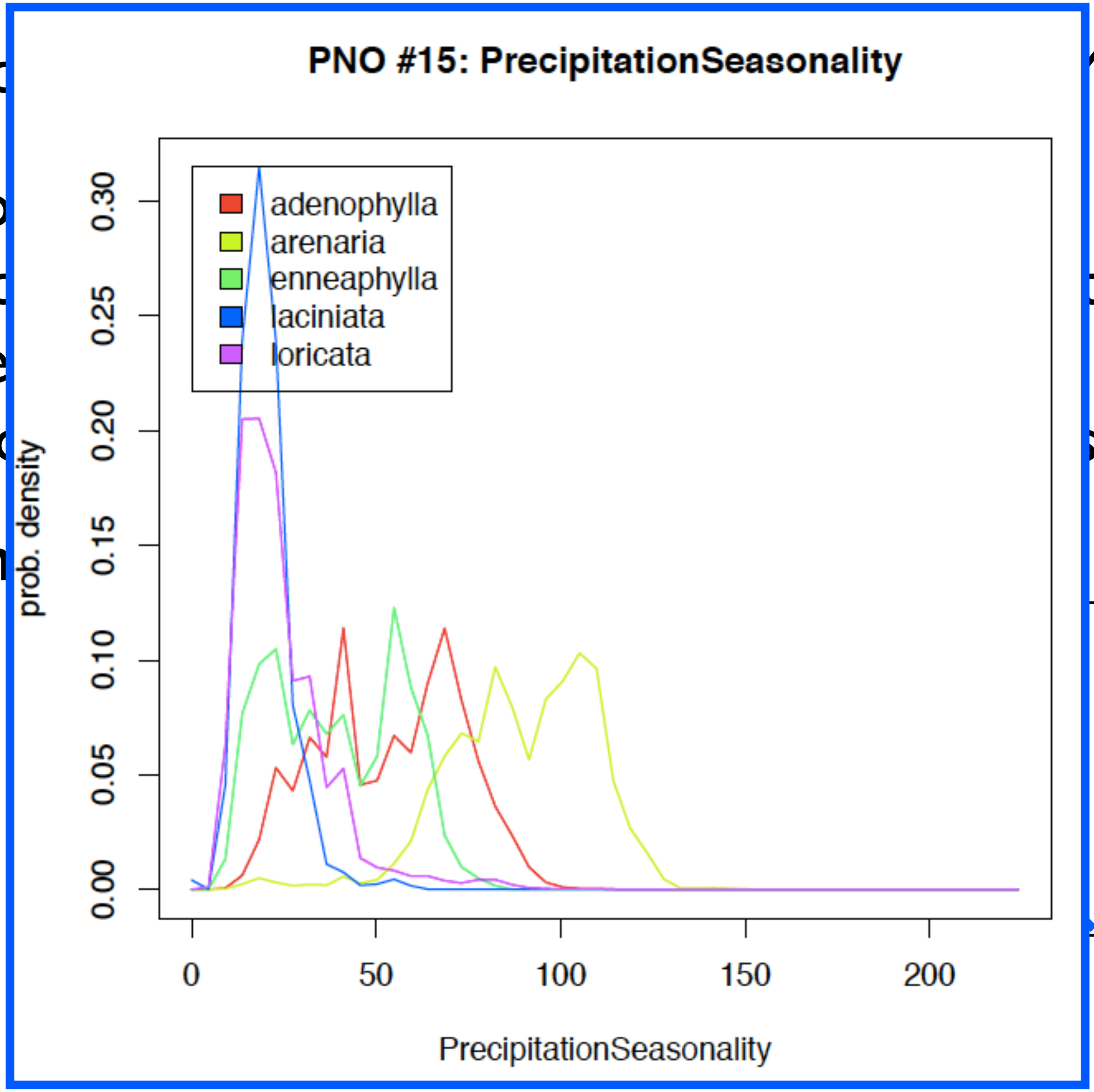
too tween

me

…c s

Con

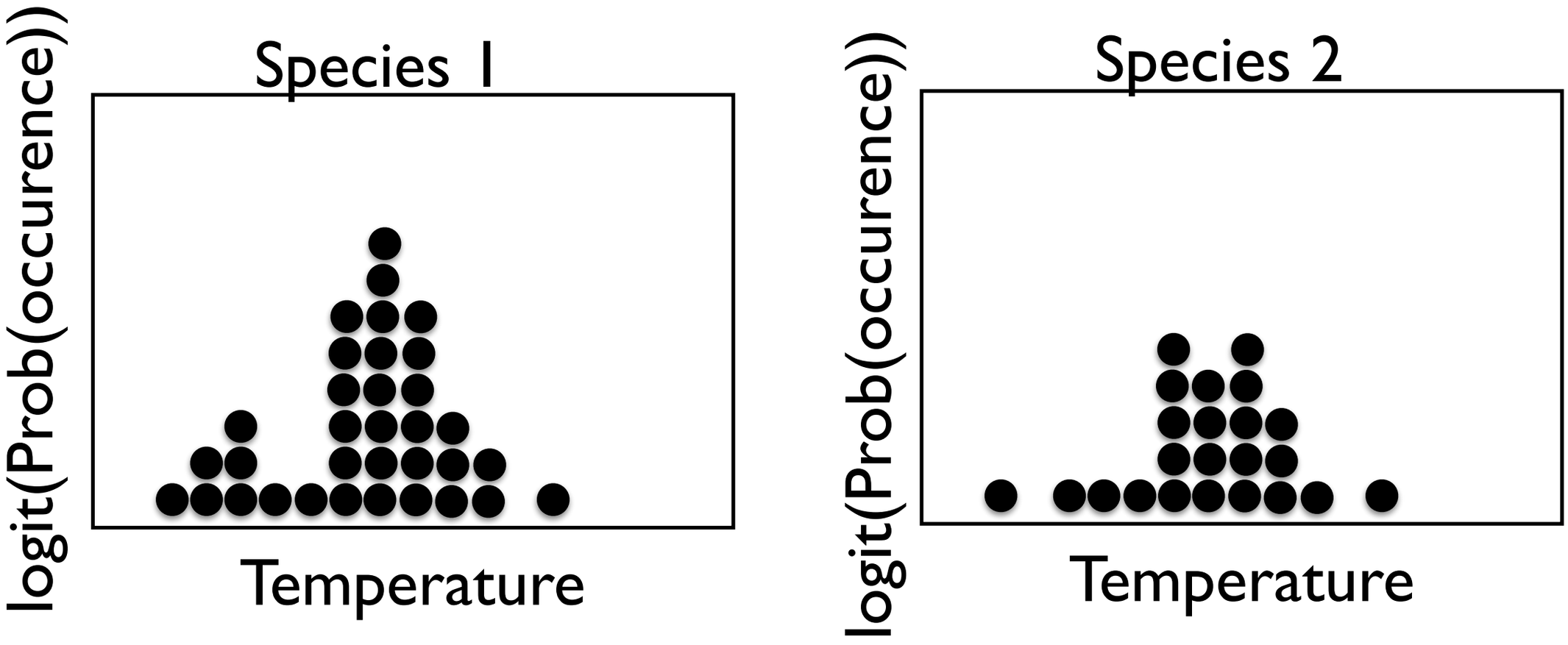
-
-
-
-
-
-



(example dataset from Heibl's phyloclim R package)

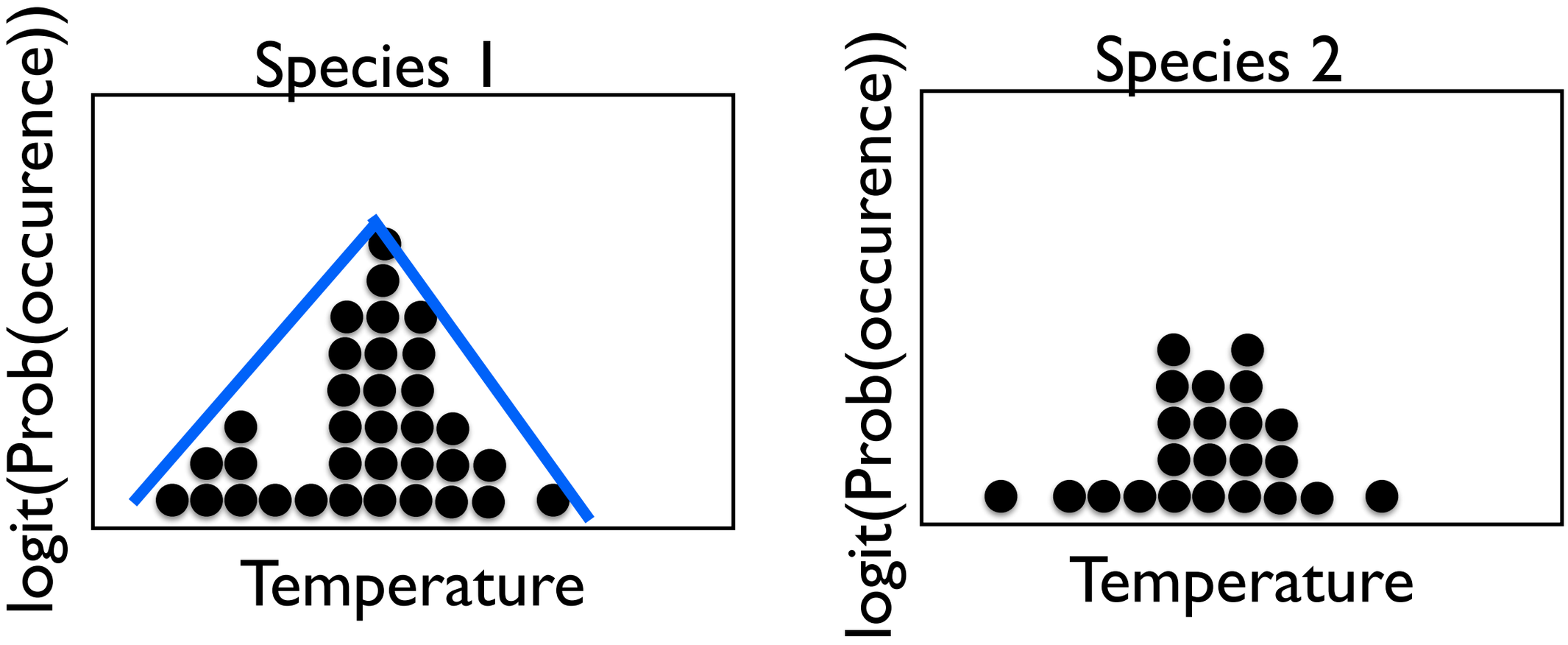
Proposal: PhyloSDM

1. Niche model: simple spline, 3 parameters/species/trait



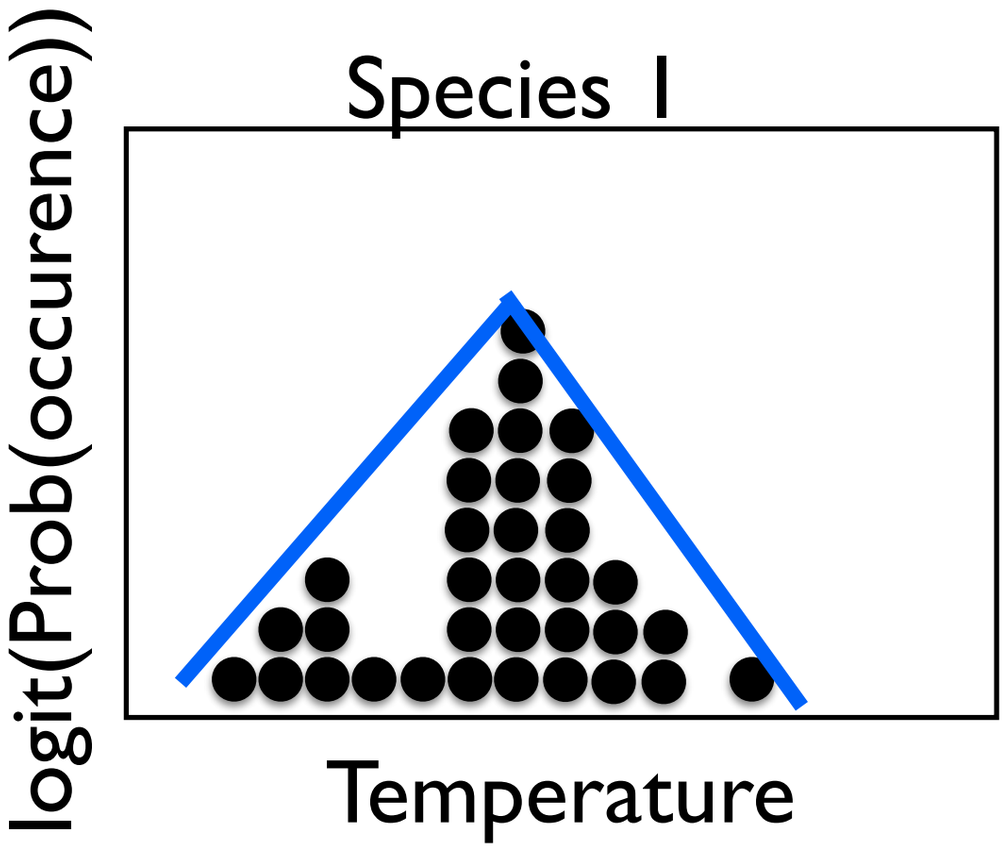
Proposal: PhyloSDM

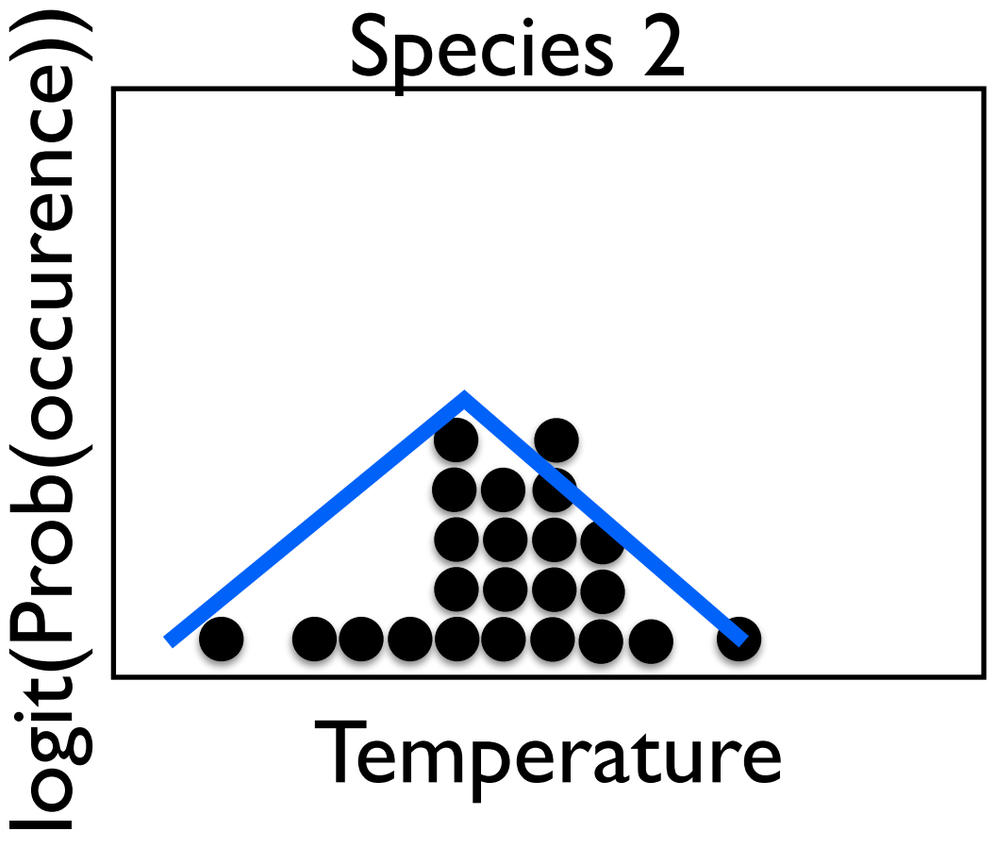
1. Niche model: simple spline, 3 parameters/species/trait



Proposal: PhyloSDM

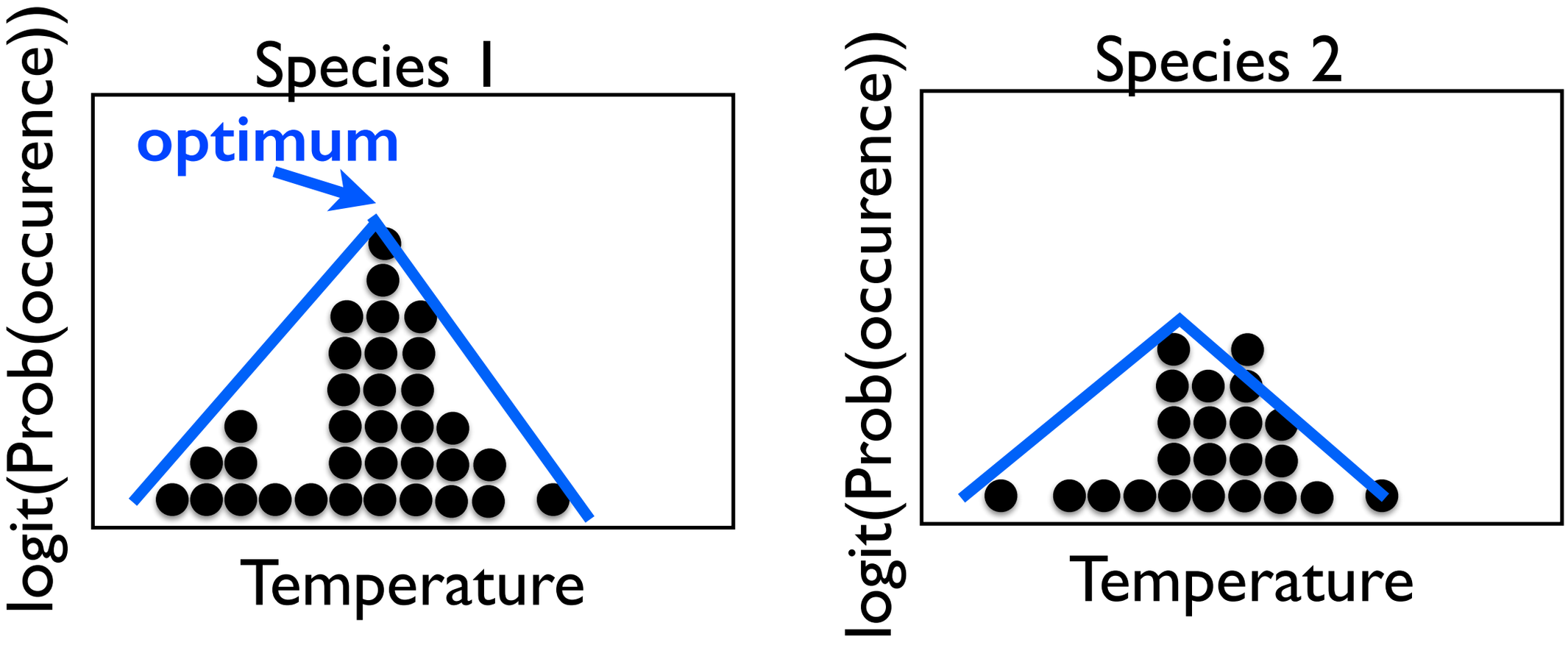
1. Niche model: simple spline, 3 parameters/species/trait





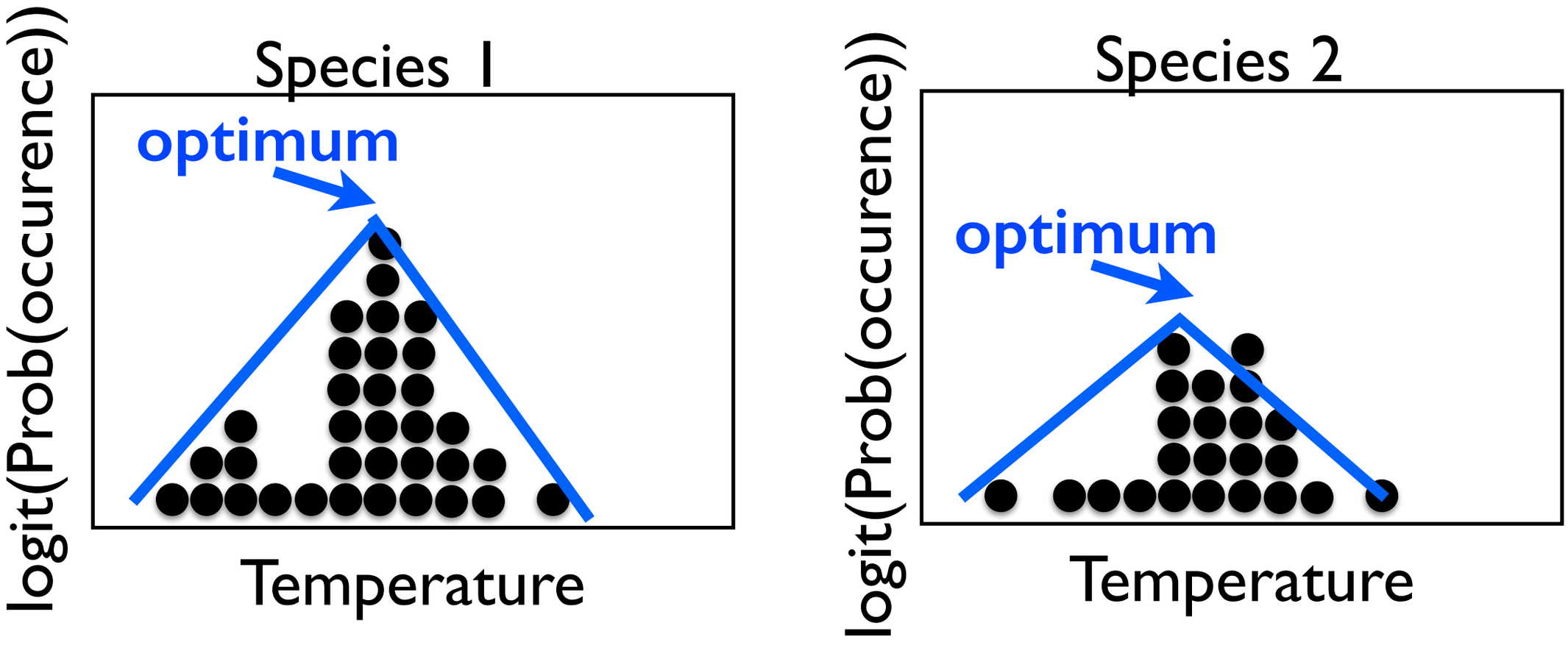
Proposal: PhyloSDM

1. Niche model: simple spline, 3 parameters/species/trait



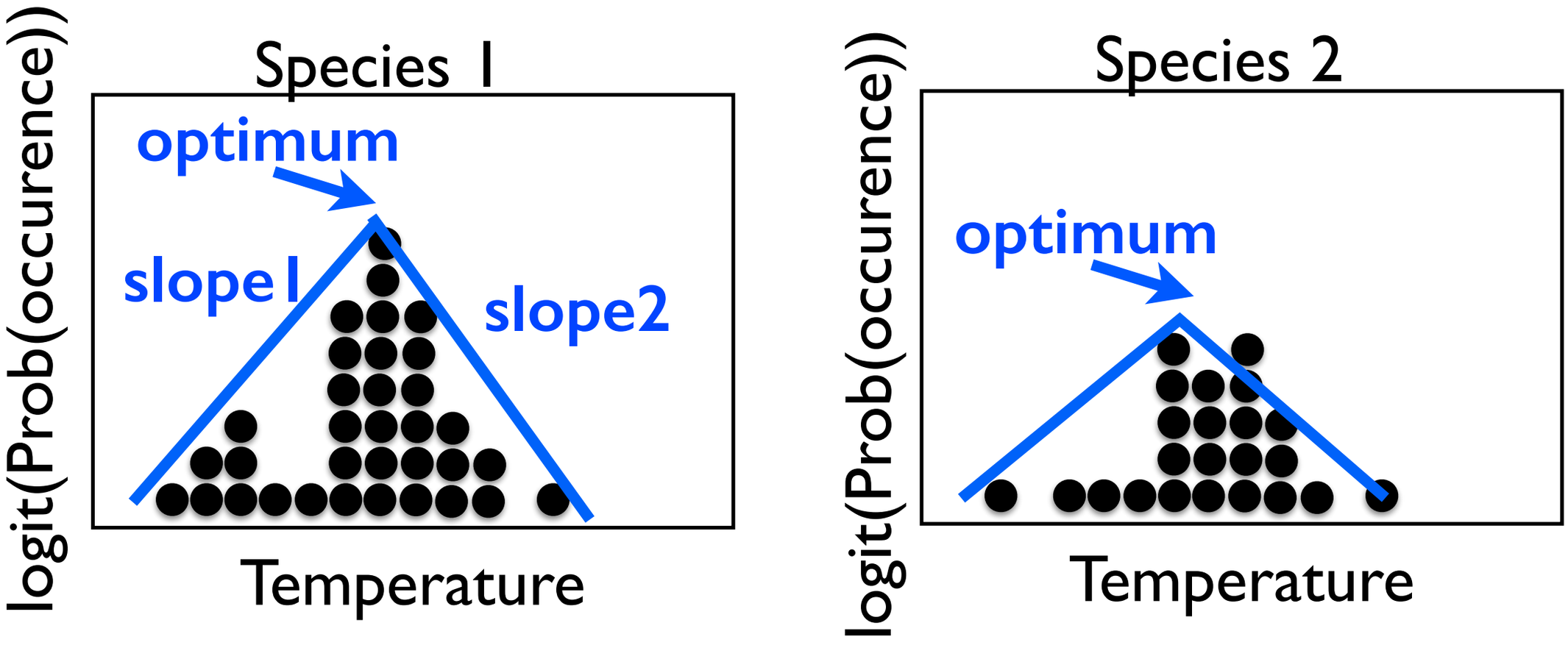
Proposal: PhyloSDM

1. Niche model: simple spline, 3 parameters/species/trait



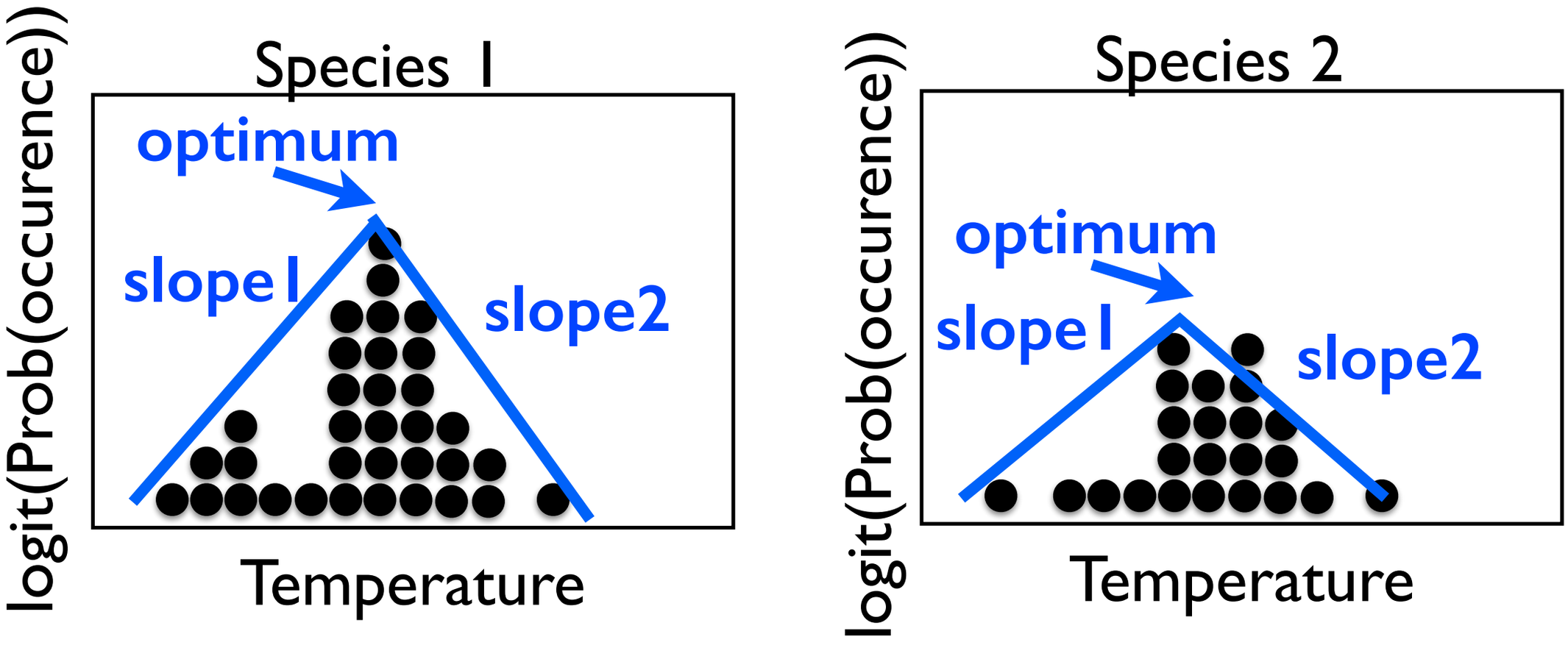
Proposal: PhyloSDM

1. Niche model: simple spline, 3 parameters/species/trait



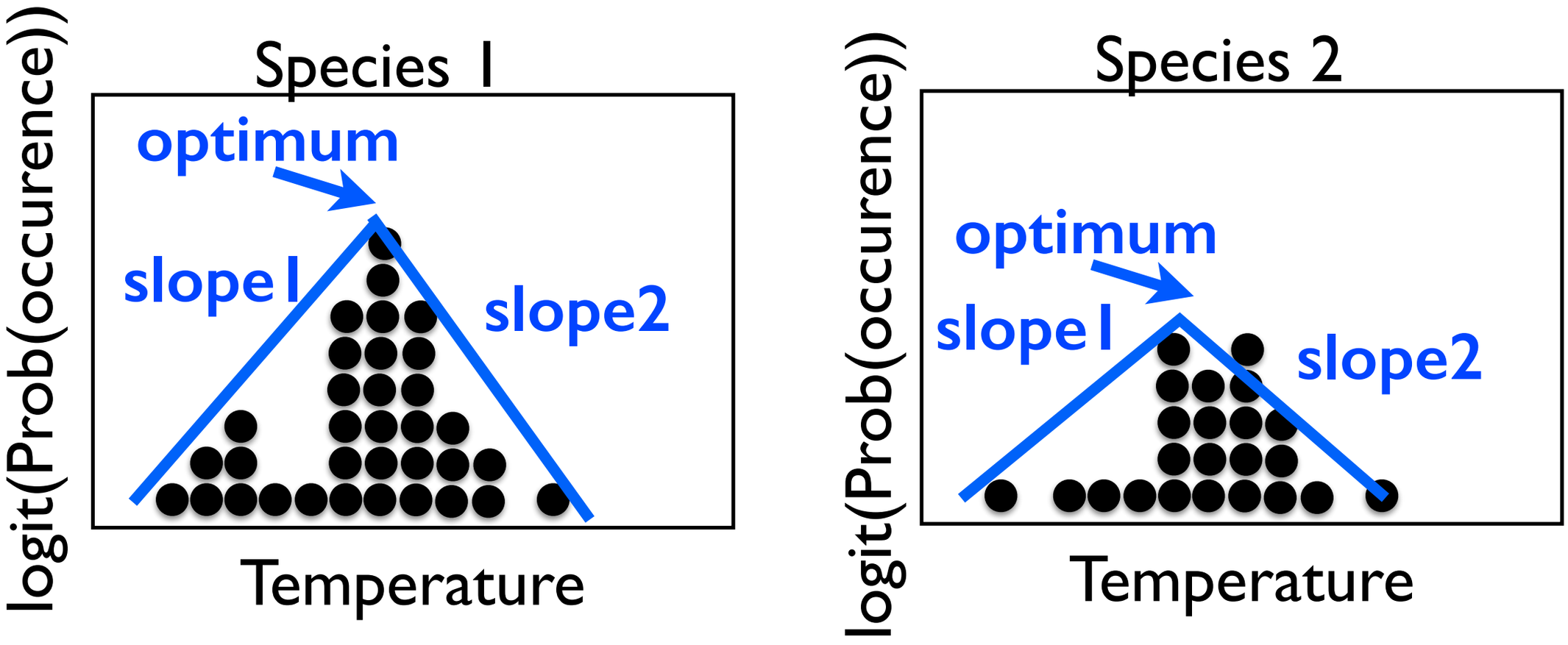
Proposal: PhyloSDM

1. Niche model: simple spline, 3 parameters/species/trait



Proposal: PhyloSDM

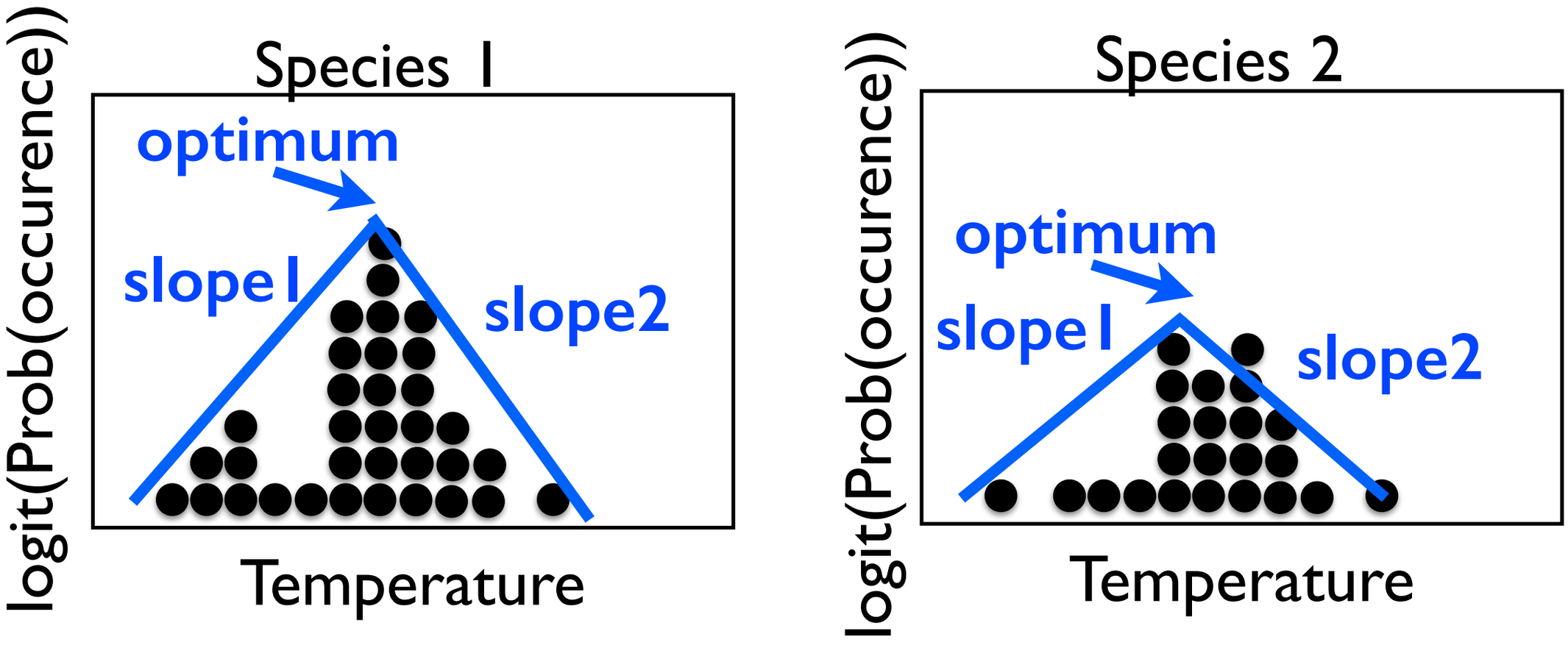
1. Niche model: simple spline, 3 parameters/species/trait



2. Evolve niche parameter(s) with Brownian motion, Ornstein-Uhlenbeck, shifting means, etc.

Proposal: PhyloSDM

1. Niche model: simple spline, 3 parameters/species/trait

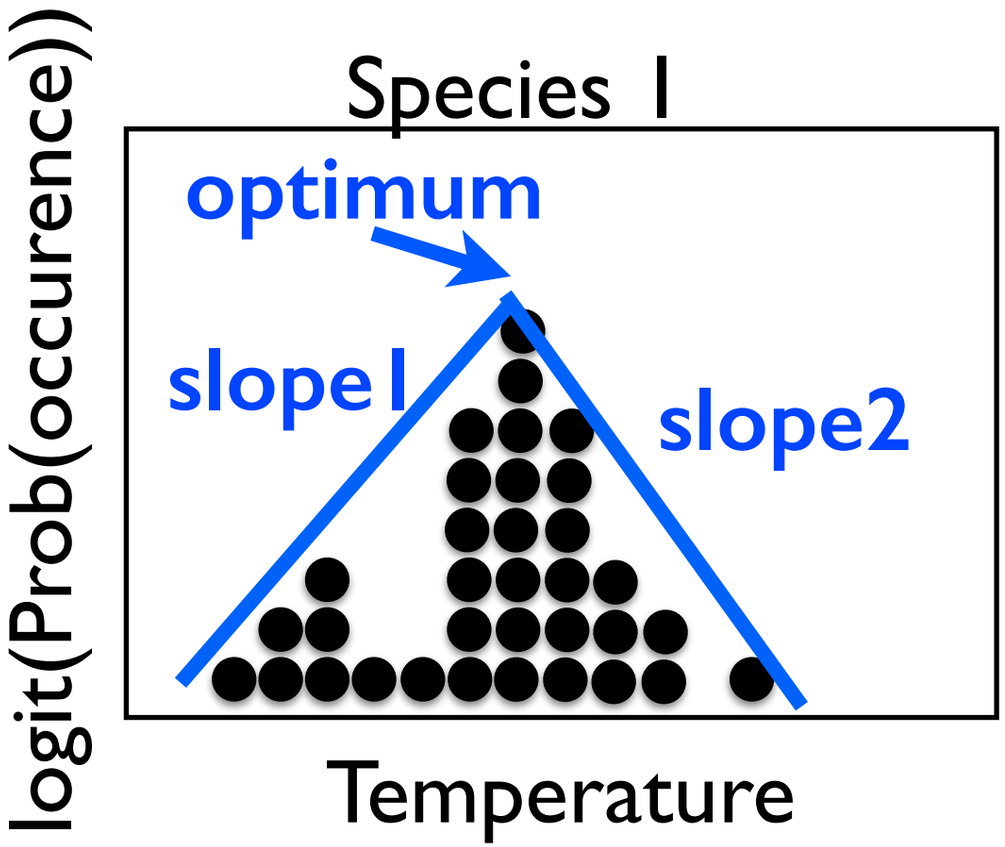


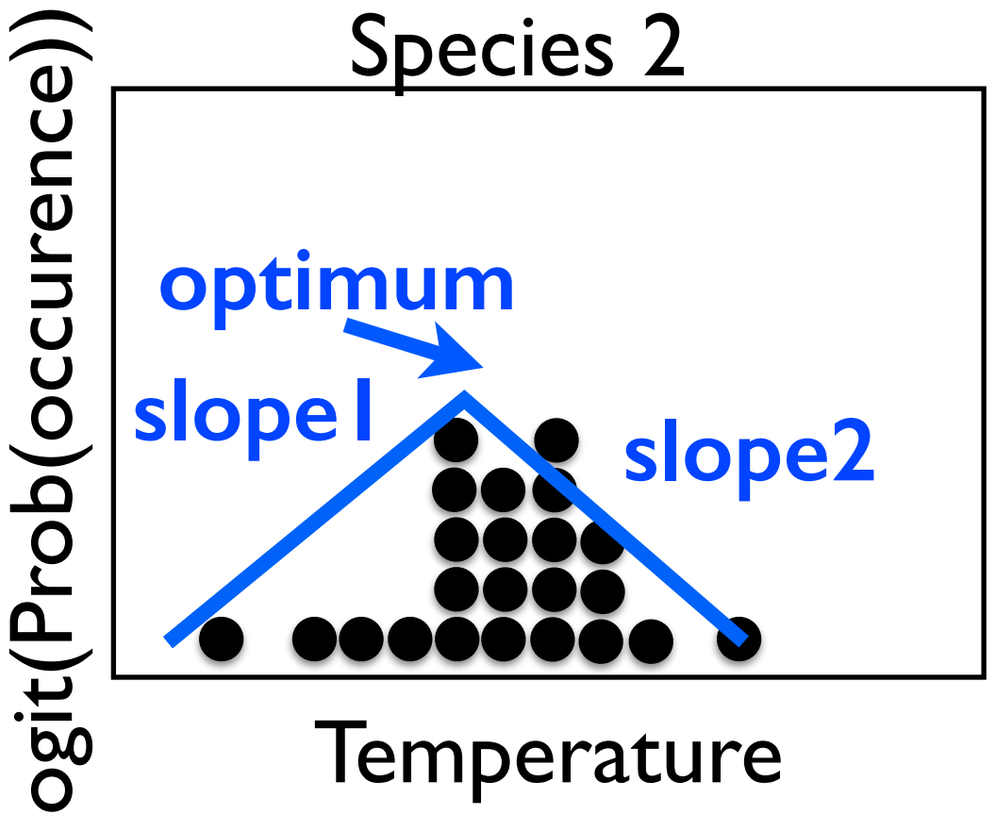
2. Evolve niche parameter(s) with Brownian motion, Ornstein-Uhlenbeck, shifting means, etc.

3. MCMC

Proposal: PhyloSDM

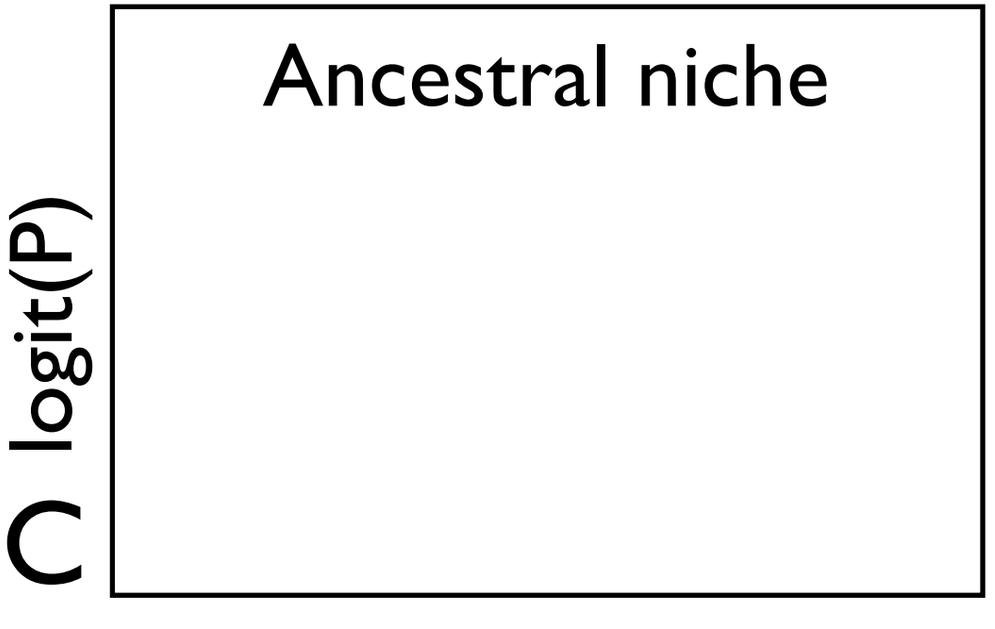
1. Niche model: simple spline, 3 parameters/species/trait





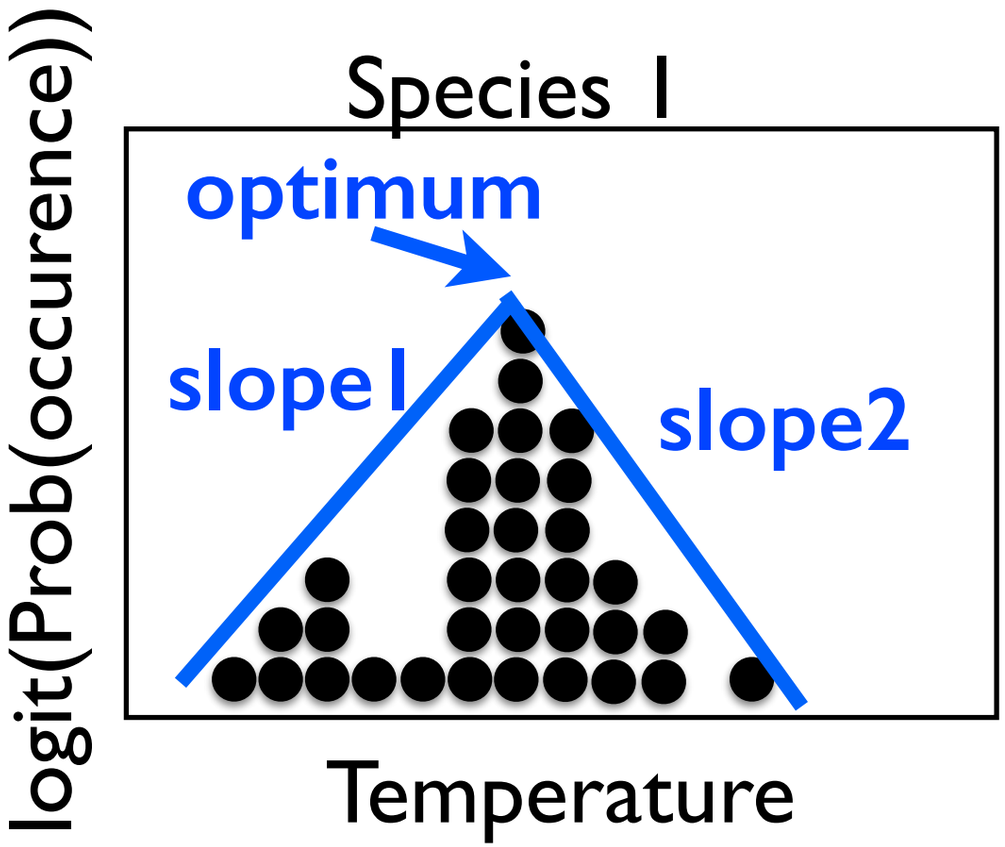
2. Evolve niche parameter(s) with Brownian motion, Ornstein-Uhlenbeck, shifting means, etc.

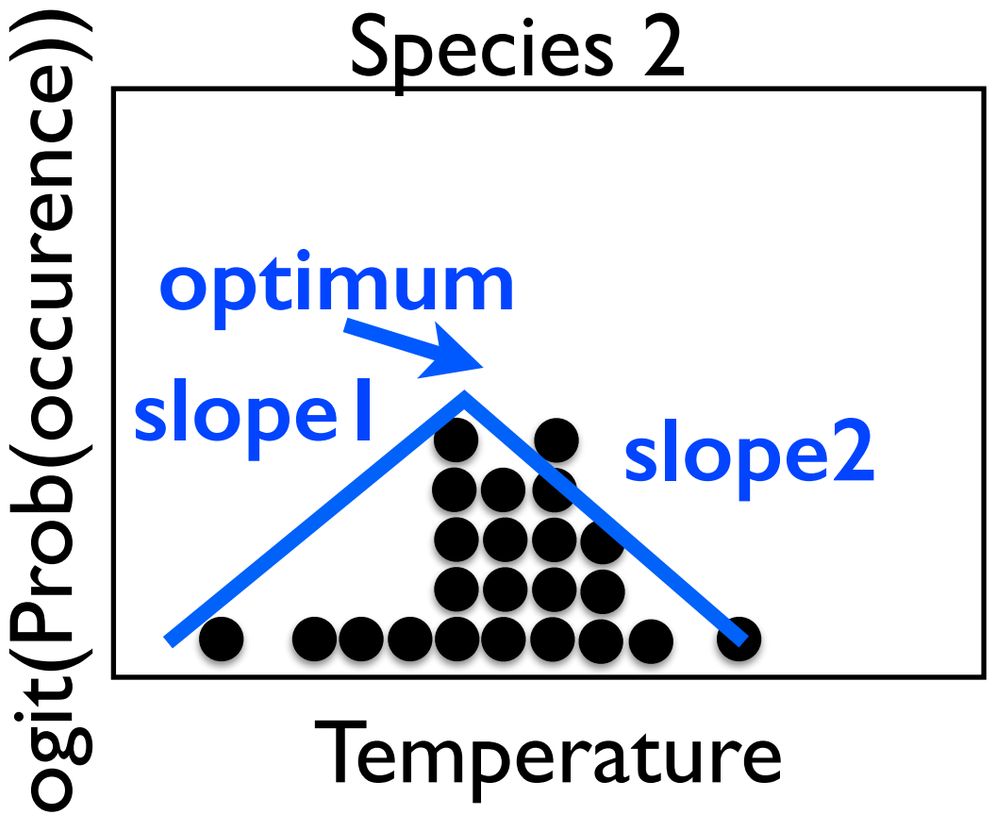
3. MCMC



Proposal: PhyloSDM

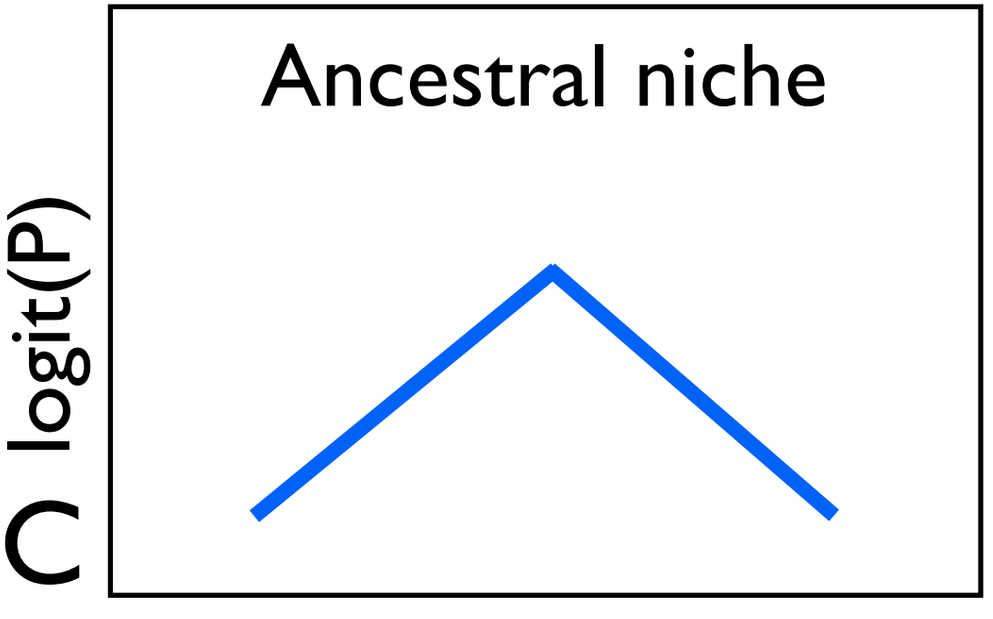
1. Niche model: simple spline, 3 parameters/species/trait





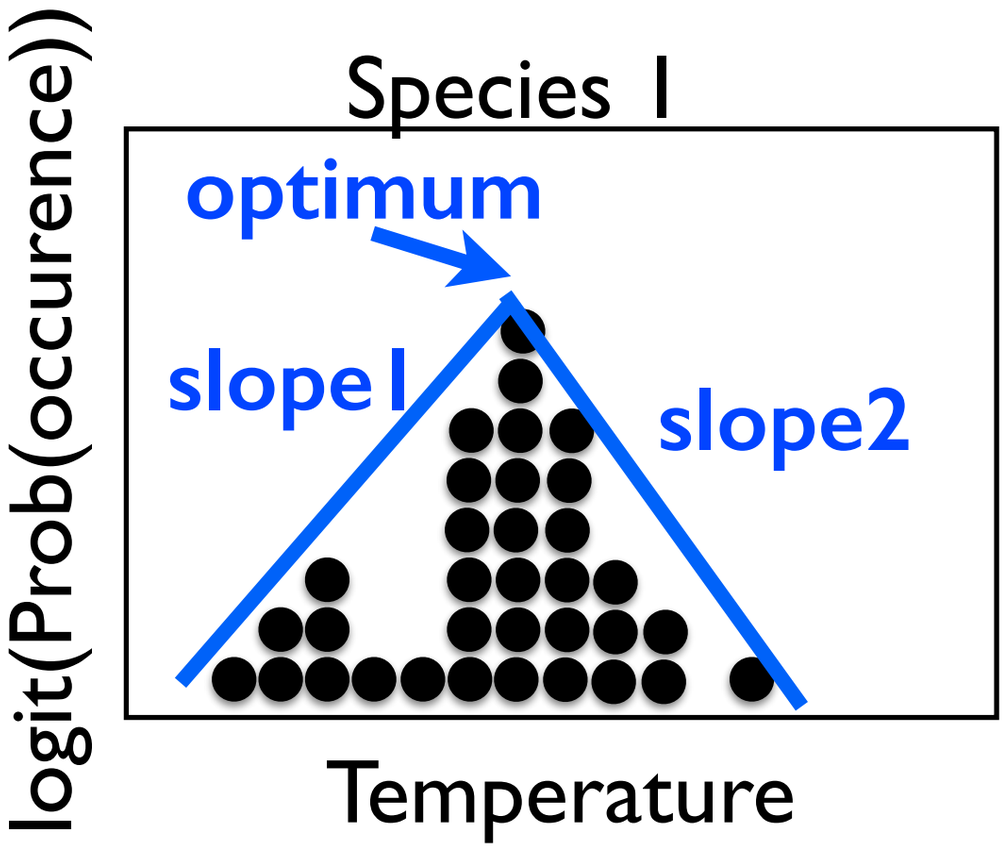
2. Evolve niche parameter(s) with Brownian motion, Ornstein-Uhlenbeck, shifting means, etc.

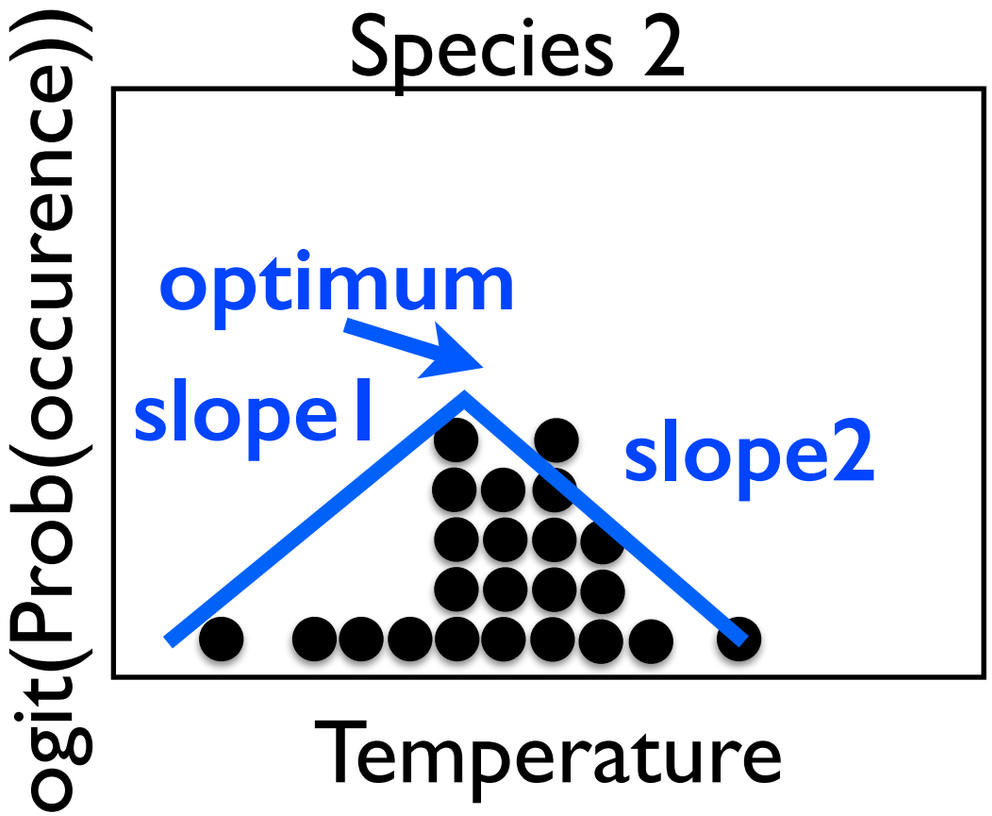
3. MCMC



Proposal: PhyloSDM

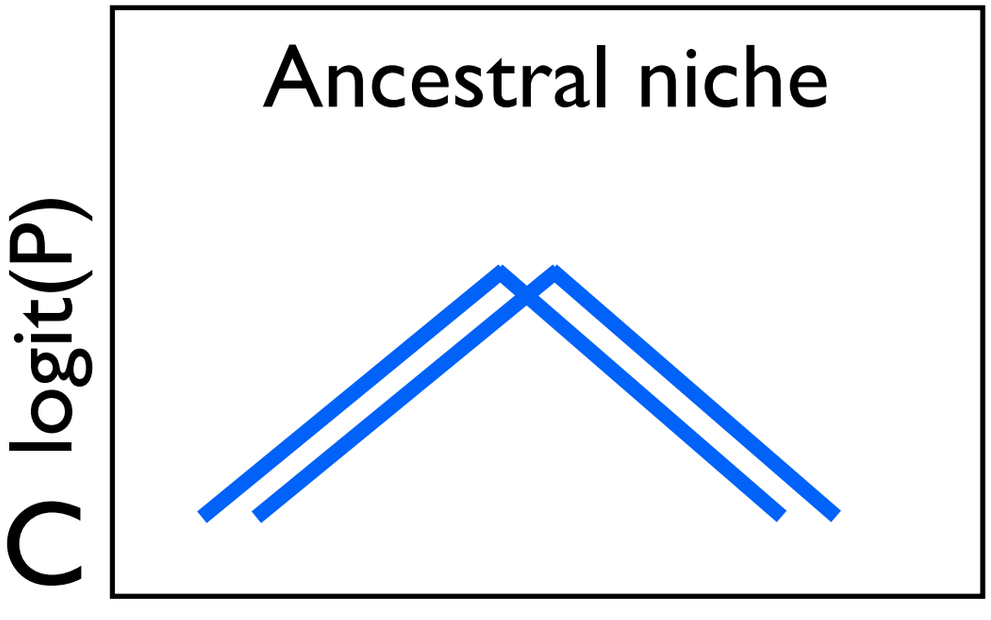
1. Niche model: simple spline, 3 parameters/species/trait





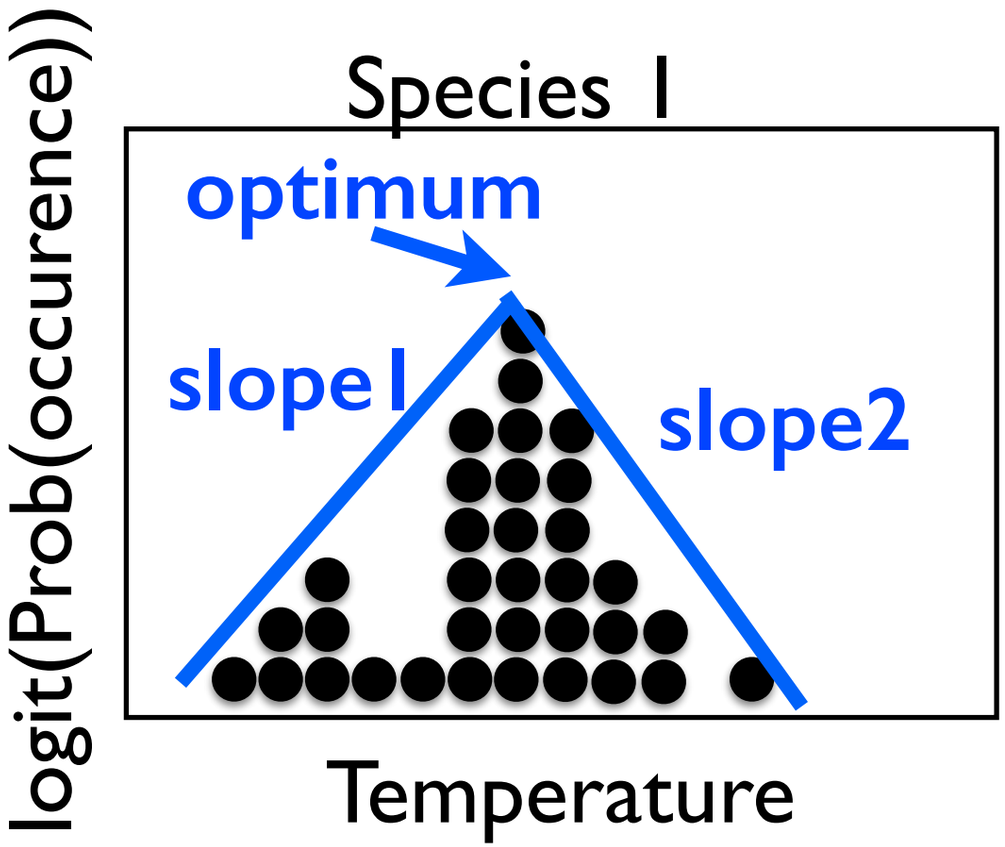
2. Evolve niche parameter(s) with Brownian motion, Ornstein-Uhlenbeck, shifting means, etc.

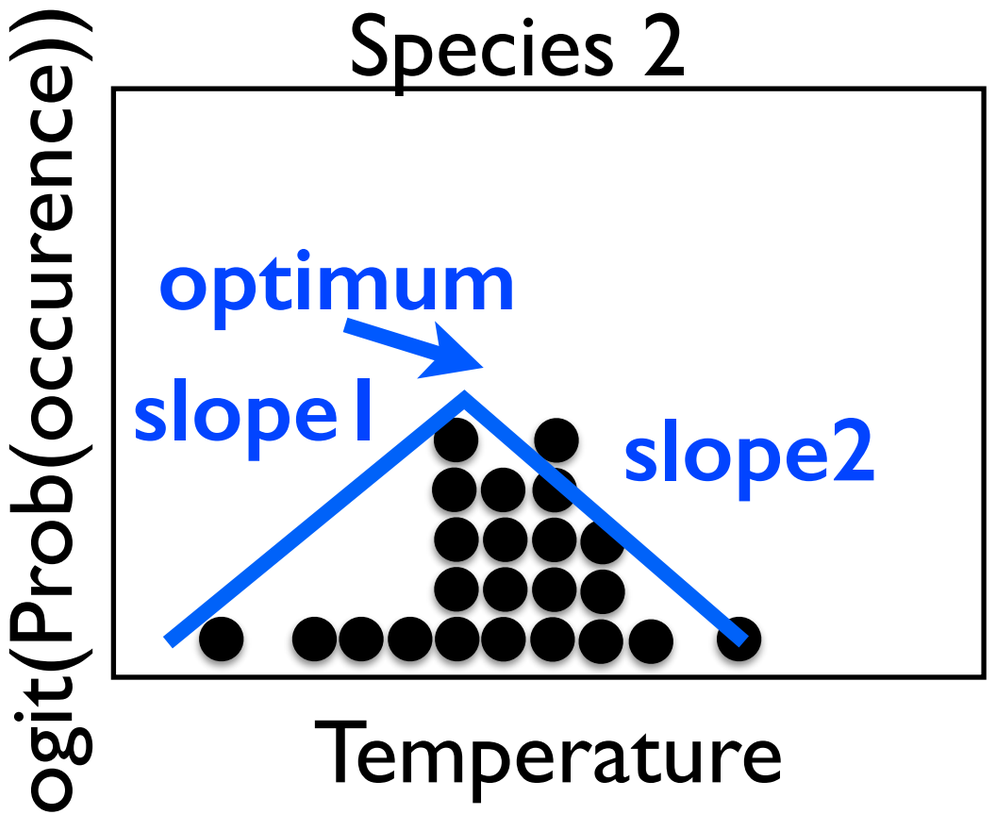
3. MCMC



Proposal: PhyloSDM

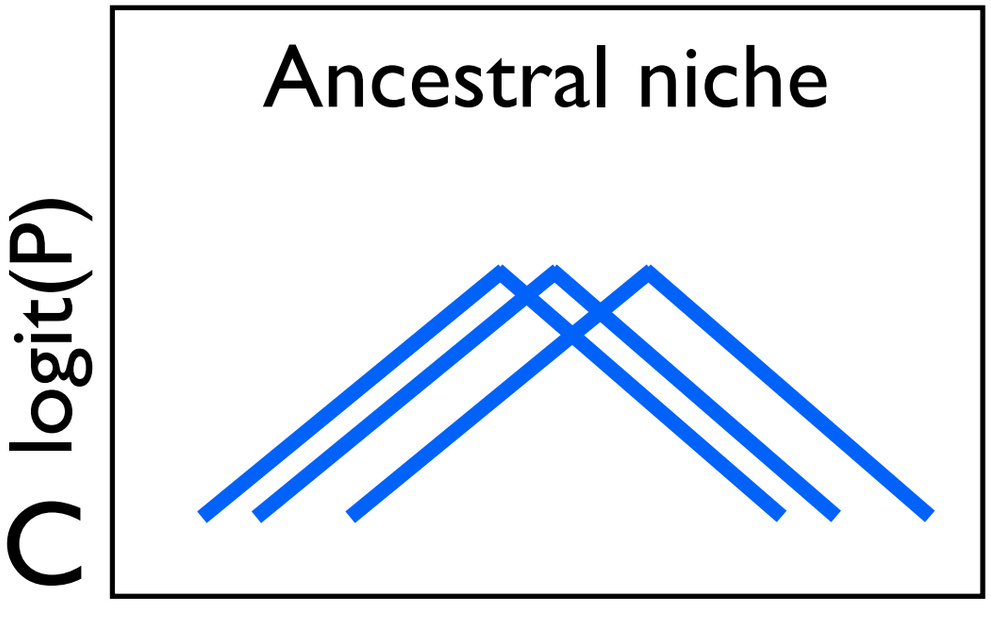
1. Niche model: simple spline, 3 parameters/species/trait





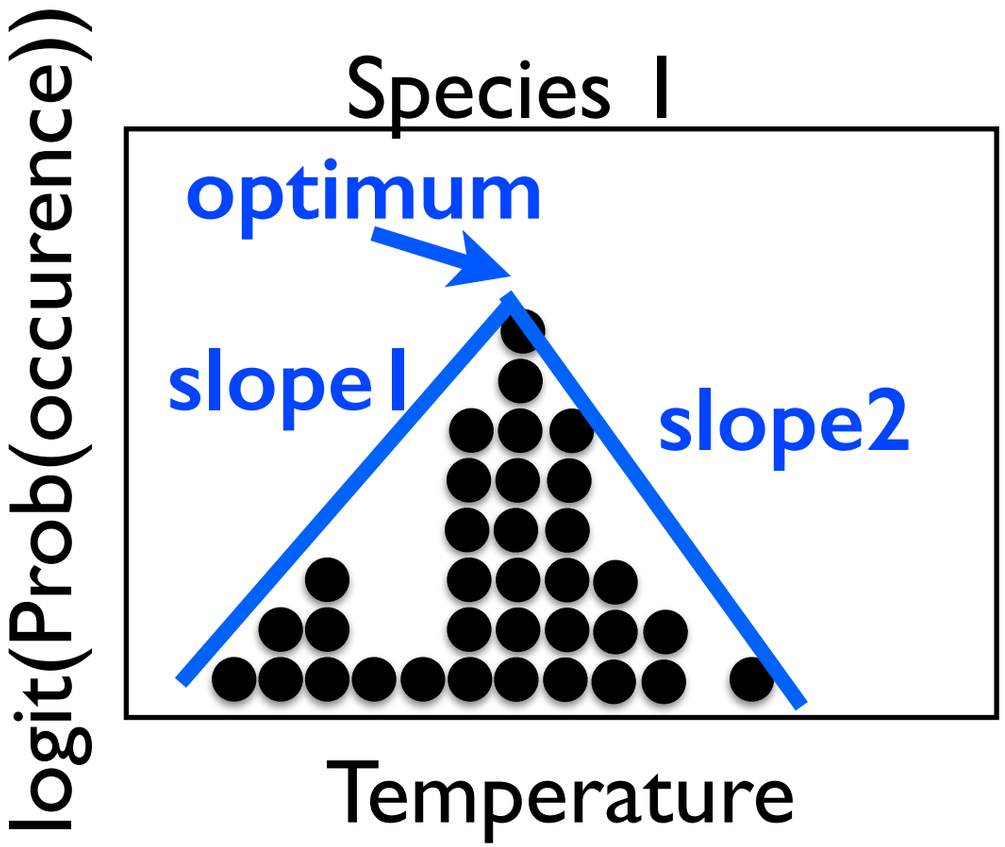
2. Evolve niche parameter(s) with Brownian motion, Ornstein-Uhlenbeck, shifting means, etc.

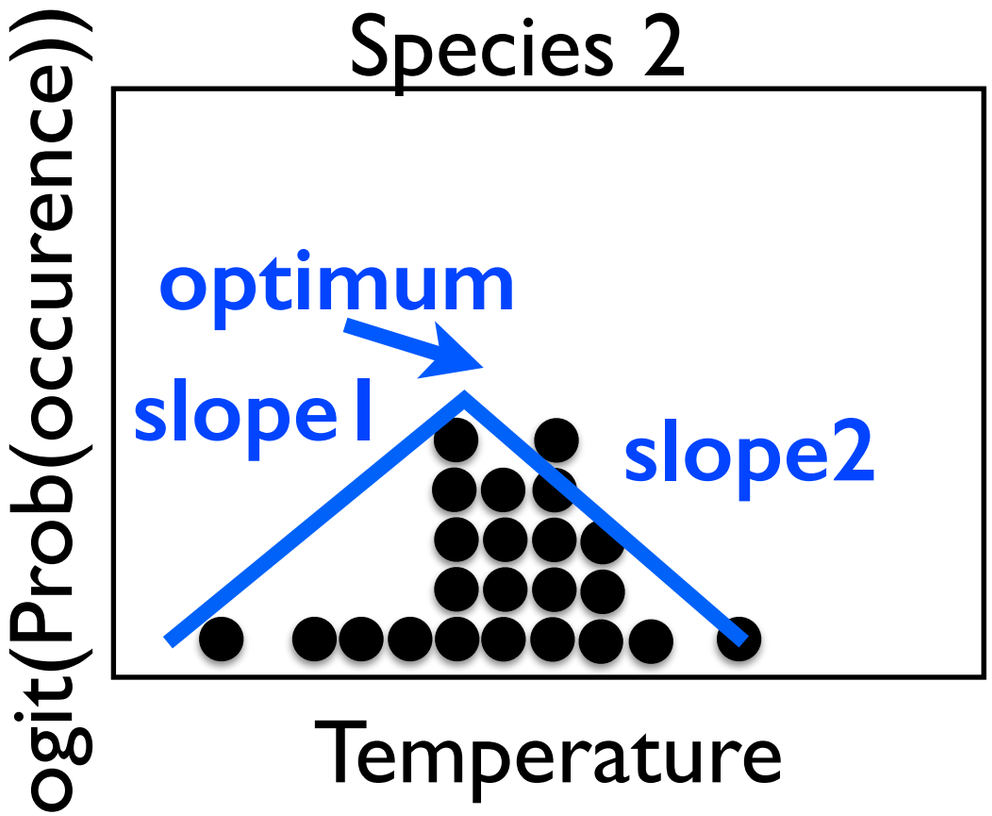
3. MCMC



Proposal: PhyloSDM

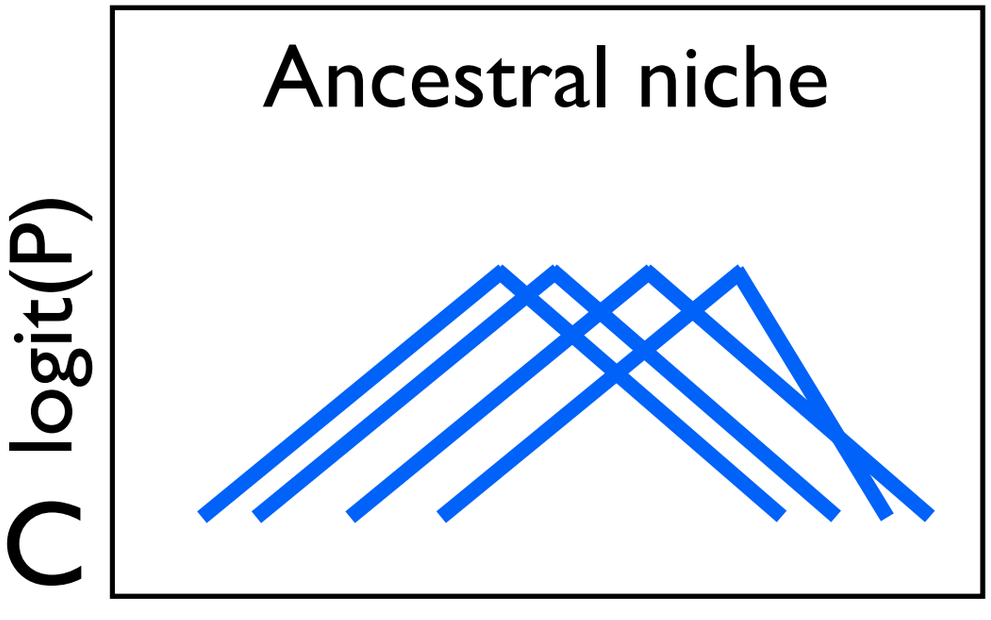
1. Niche model: simple spline, 3 parameters/species/trait





2. Evolve niche parameter(s) with Brownian motion, Ornstein-Uhlenbeck, shifting means, etc.

3. MCMC



Proposal: PhyloSDM

Advantages:

1. Niche models and evolutionary models are simultaneously inferred

Proposal: PhyloSDM

Advantages:

1. Niche models and evolutionary models are simultaneously inferred
2. Information sharing between close relatives

Proposal: PhyloSDM

Advantages:

1. Niche models and evolutionary models are simultaneously inferred
2. Information sharing between close relatives
3. Error, historical biogeography etc. could be modelled also

Proposal: PhyloSDM

Advantages:

1. Niche models and evolutionary models are simultaneously inferred
2. Information sharing between close relatives
3. Error, historical biogeography etc. could be modelled also

Difficulties:

Proposal: PhyloSDM

Advantages:

1. Niche models and evolutionary models are simultaneously inferred
2. Information sharing between close relatives
3. Error, historical biogeography etc. could be modelled also

Difficulties:

1. # of parameters:
   3(nspecies)(nvars)+
   3*3(nvars)
   (bare minimum!)

Proposal: PhyloSDM

Advantages:

1. Niche models and evolutionary models are simultaneously inferred
2. Information sharing between close relatives
3. Error, historical biogeography etc. could be modelled also

Difficulties:

1. # of parameters:
   3(nspecies)(nvars)+
   3*3(nvars)
   (bare minimum!)
2. Speed vs. traditional heuristic searches

Proposal: PhyloSDM

Advantages:

1. Niche models and evolutionary models are simultaneously inferred
2. Information sharing between close relatives
3. Error, historical biogeography etc. could be modelled also

Difficulties:

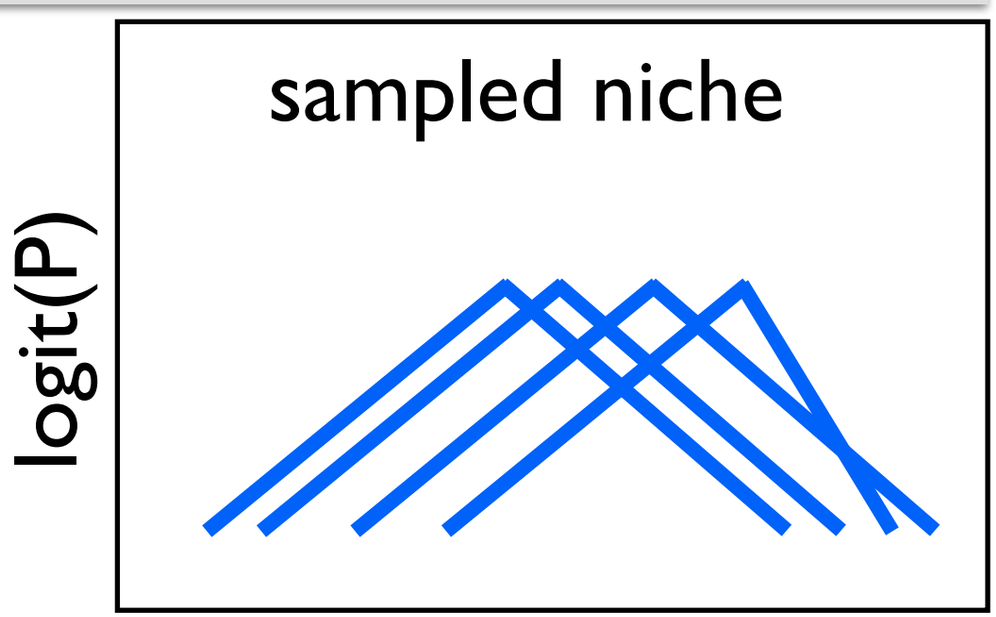
1. # of parameters:
   3(nspecies)(nvars)+
   3*3(nvars)
   (bare minimum!)
2. Speed vs. traditional heuristic searches
3. Evaluation metrics - unexplored

Current status of phyloSDM

Programming in JAGS / R2jags:

- Detection model: P(detect|presence)
- Niche model: P(presence|environ, params)
- Evolutionary model: P(params|evo. model)
- E.g. OU process with shifting means (using a stochastically mapped trait)
- Sample from distrib. of trees, maps

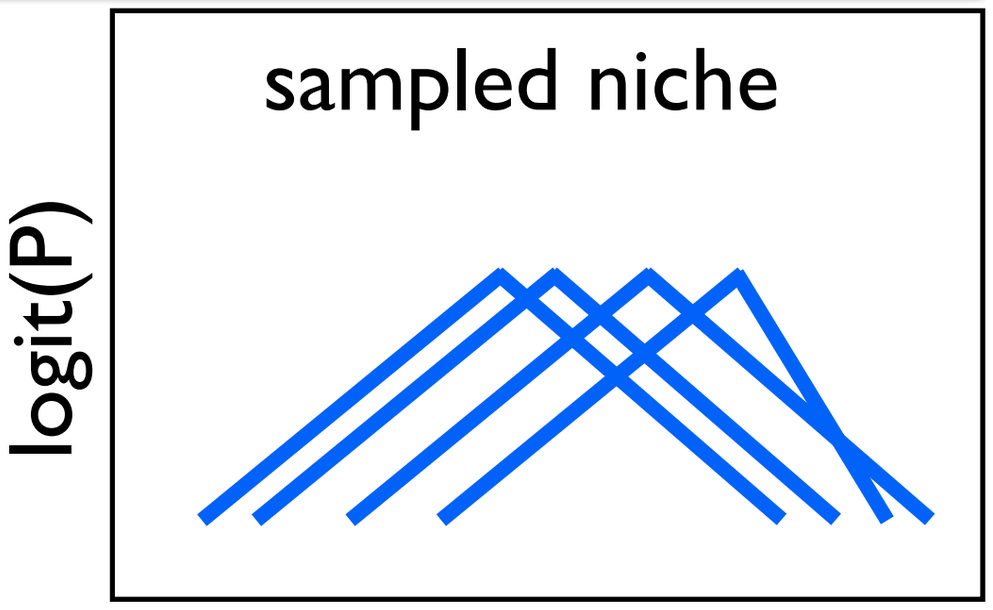
MCMC sampling of possible niche models



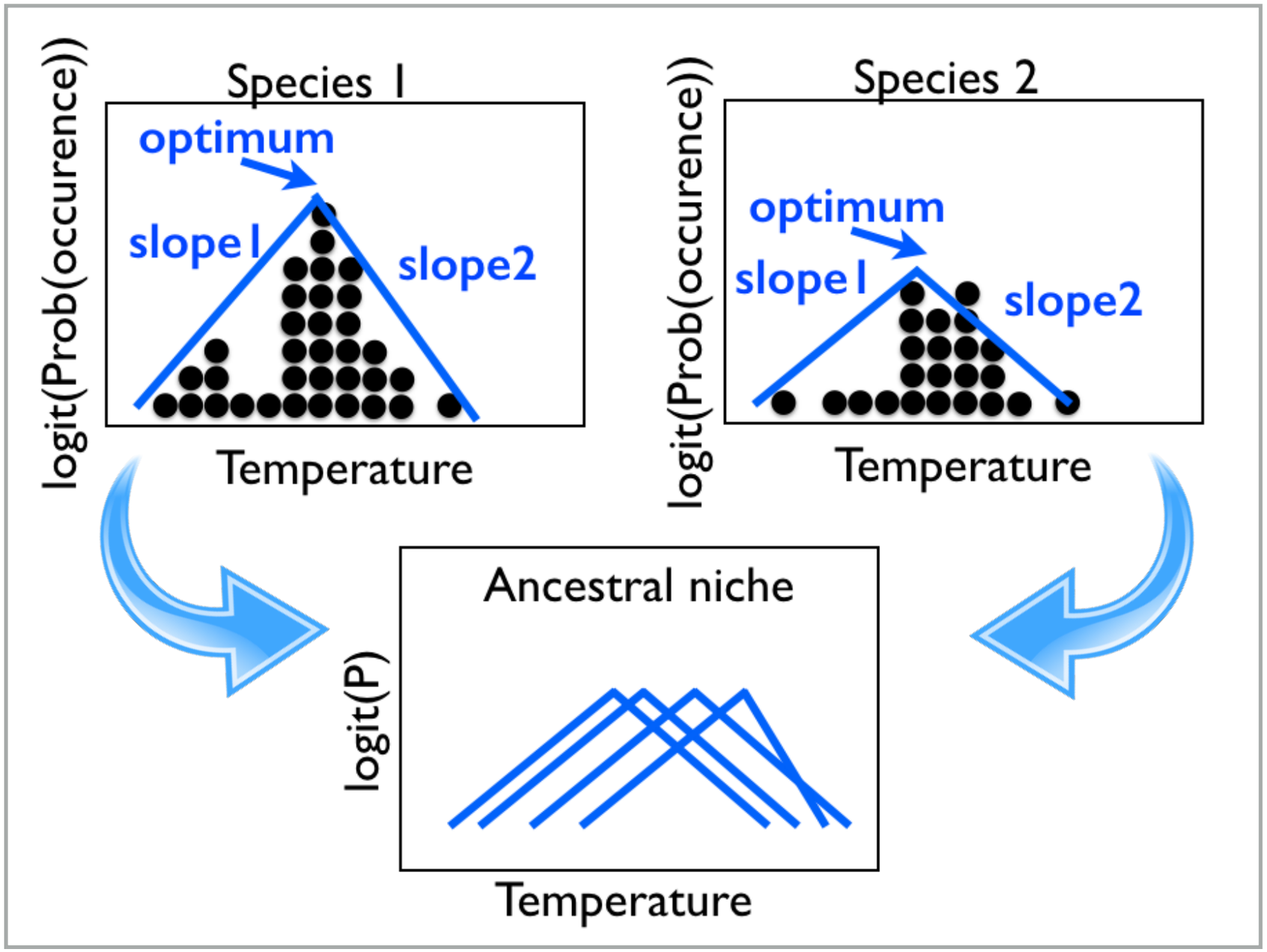
Open questions for symposium

- What do we miss in ENM if we leave out evolution/phylogeny?
- How can we include paleontological information?
- How can we combine models of ecological niche, and niche evolution?

MCMC sampling of possible niche models



Putting Evolution Into Ecological Niche Modelling: Background and Paths Forward



Nicholas J. Matzke
DECRA Fellow
Division of Ecology, Evolution, and Genetics
Research School of Biology
The Australian National University
phylo.wikidot.com/nicholas-j-matzke

Evolution 2016
Sunday, June 19, 8:15 - 11:45 am, Ballroom C

Thanks for support from: