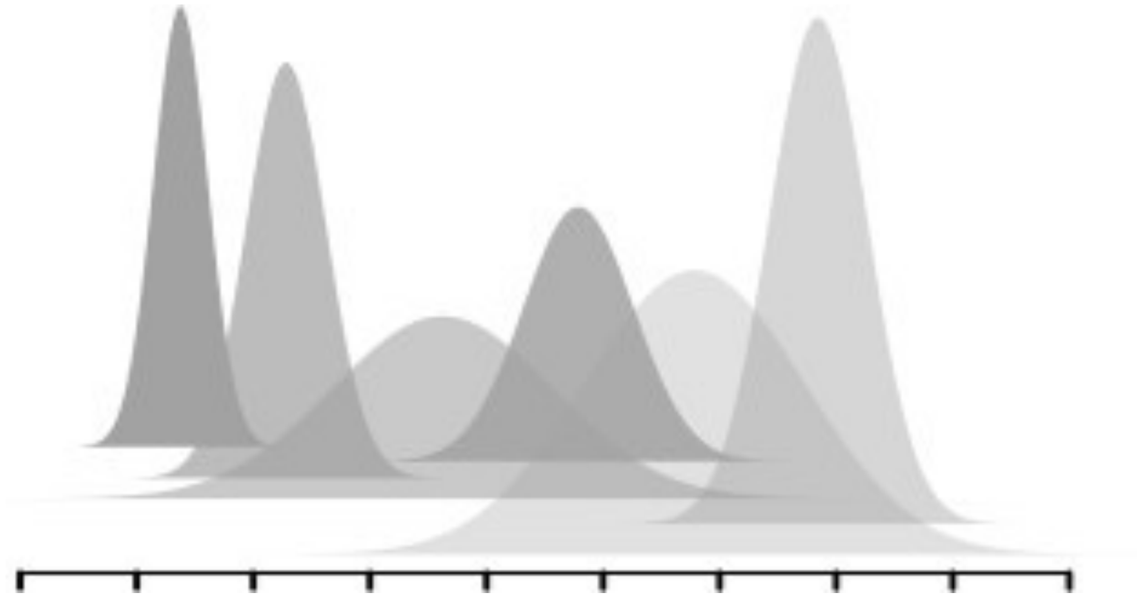
4.3 Fitting empirical data to models



Jelena H. Pantel

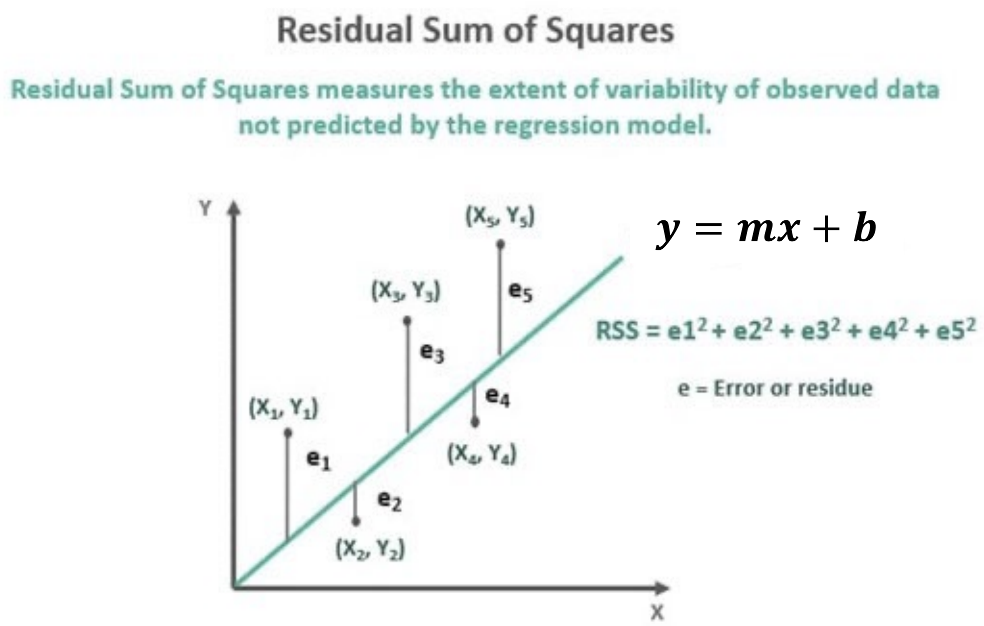
Faculty of Biology

University of Duisburg-Essen

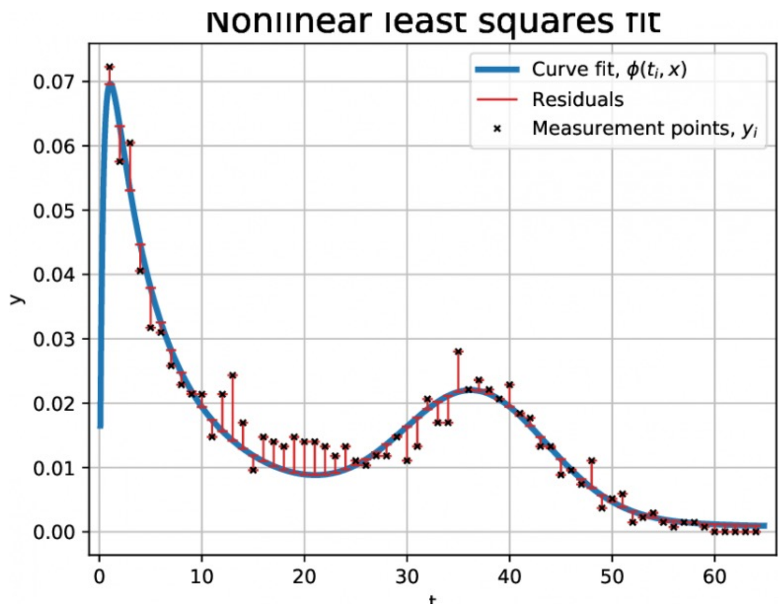
jelena.pantel@uni-due.de

Approaches to fit data to models

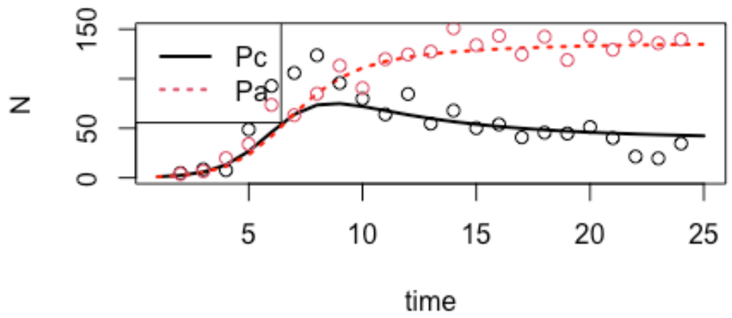
Approach 1. Least squares and non-linear least squares



$$y = mx + b$$



$$\frac{dn_i}{dt} = rn_i\left(1 - \frac{n_i + \alpha_{ij}n_j}{K_i}\right)$$



Goal: estimate parameters in linear or non-linear models by minimizing the residual sum of squares (RSS)

The smaller the RSS, the smaller the deviation between model-predicted line and observed data → the better the model fits the data

Approaches to fit data to models

Other approaches: Bayesian parameter estimation

What is the goal of these methods? To estimate the *most likely value* of the model parameters, given the data

We preresent this as – we want to know the *likelihood of the model parameter values given the data*

Pr (Φ | y)

What is the <u>probability</u>

of the model parameter value Φ

Given the observed data *y*

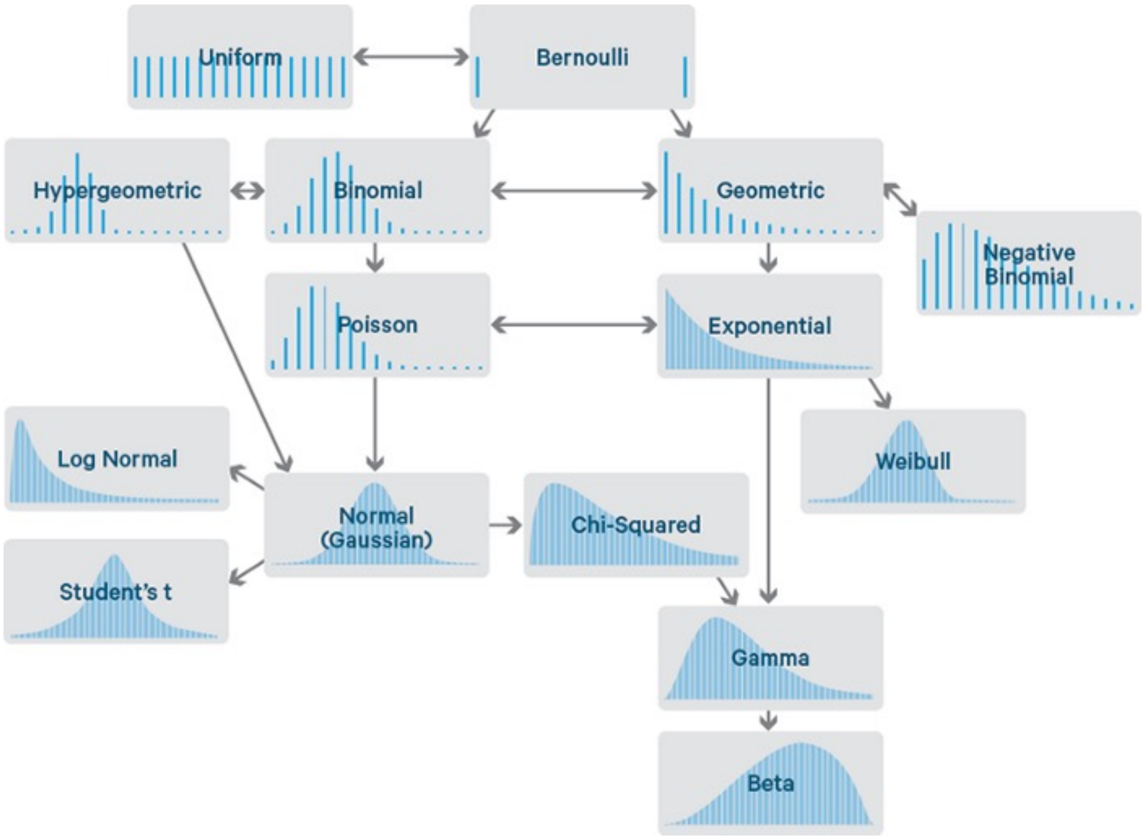
In a Bayesian framework, we (1) propose a *<u>prior distribution</u>* for each model parameter Φ

$$\frac{dn_i}{dt} = rn_i\left(1 - \frac{n_i + \alpha_{ij}n_j}{K_i}\right)$$

Φ – r, α, K

- What is the possible ranges of the parameter? – positive, negative, minimum, maximum?
- Continuous, discrete?

Possible probability distributions to choose from for model parameters



Approaches to fit data to models

Other approaches: Bayesian parameter estimation

(1) In a Bayesian framework, we propose a *prior distribution* for each model parameter Φ

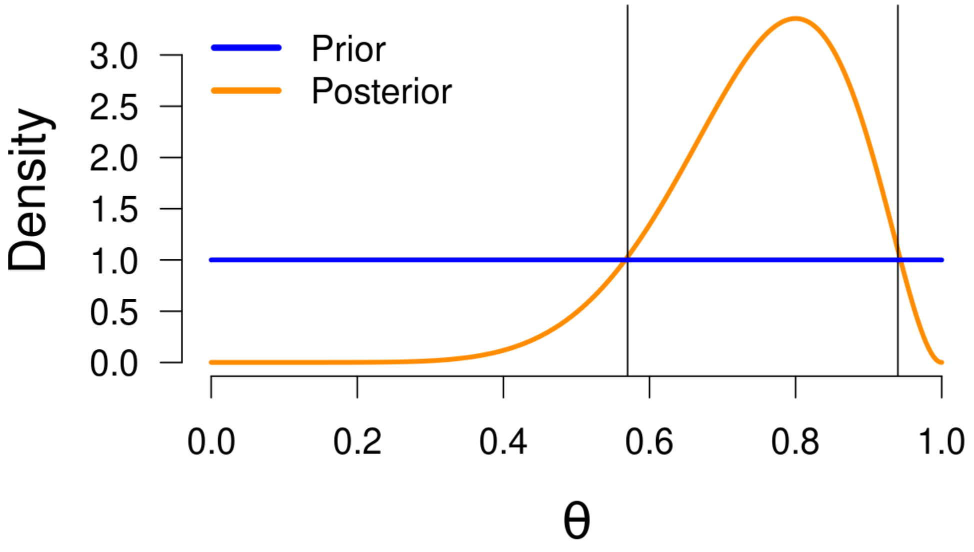
Pr (Φ | y)

What is the probability

of the model parameter value Φ

Given the observed data *y*

In other words, we use some kind of *likelihood* value to go from a prior distribution to a posterior distribution of model parameters



(2) Our goal is to:

- **propose different possible values of the parameters Φ (using the *prior distribution*)**
- **calculate the likelihood those being the 'true' parameter values given the data Pr (Φ | y)**
- **Find the *posterior distribution* of parameters, based on the likelihood calculations above**

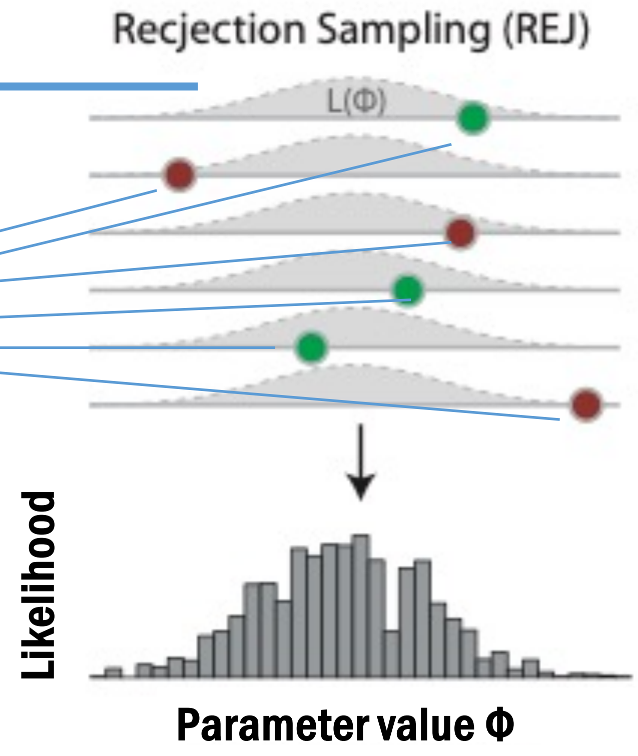
Approaches to fit data to models

Other approaches: Bayesian parameter estimation

Bayesian approach 1. Rejection sampling

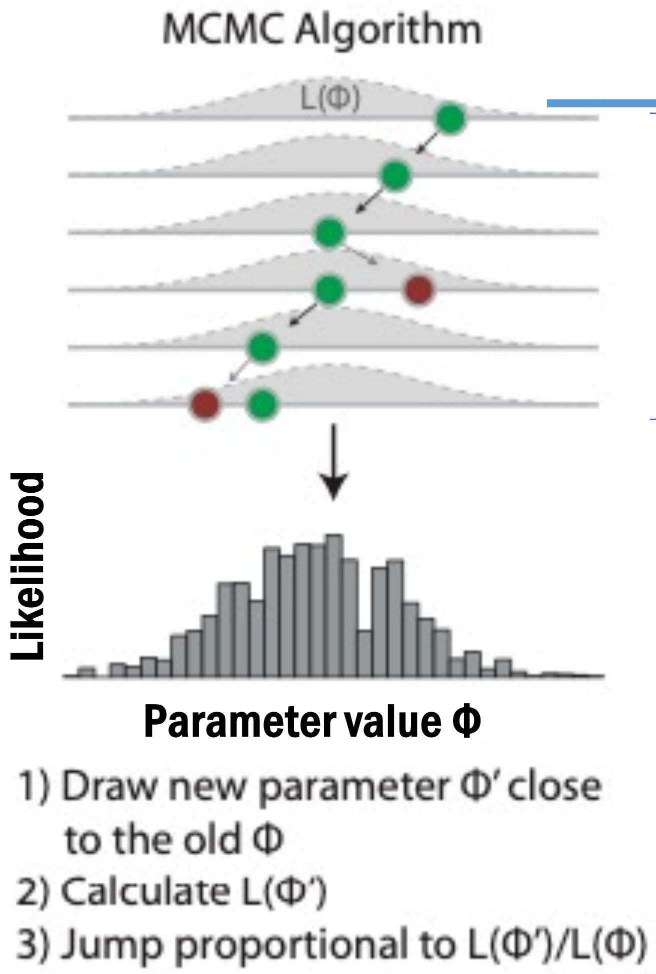
Prior distribution

Random samples from prior distribution



Note: all samples are independent

Bayesian approach 2. MCMC



Prior distribution

- Samles from prior are not independent
- Markov chain Monte Carlo (MCMC)
- Value of next step's sample from the prior depends on Likelihood of new value / Likelihood of last steps' value
- Jump is rejected (red) if Likelihood ratio is too far from a threshold

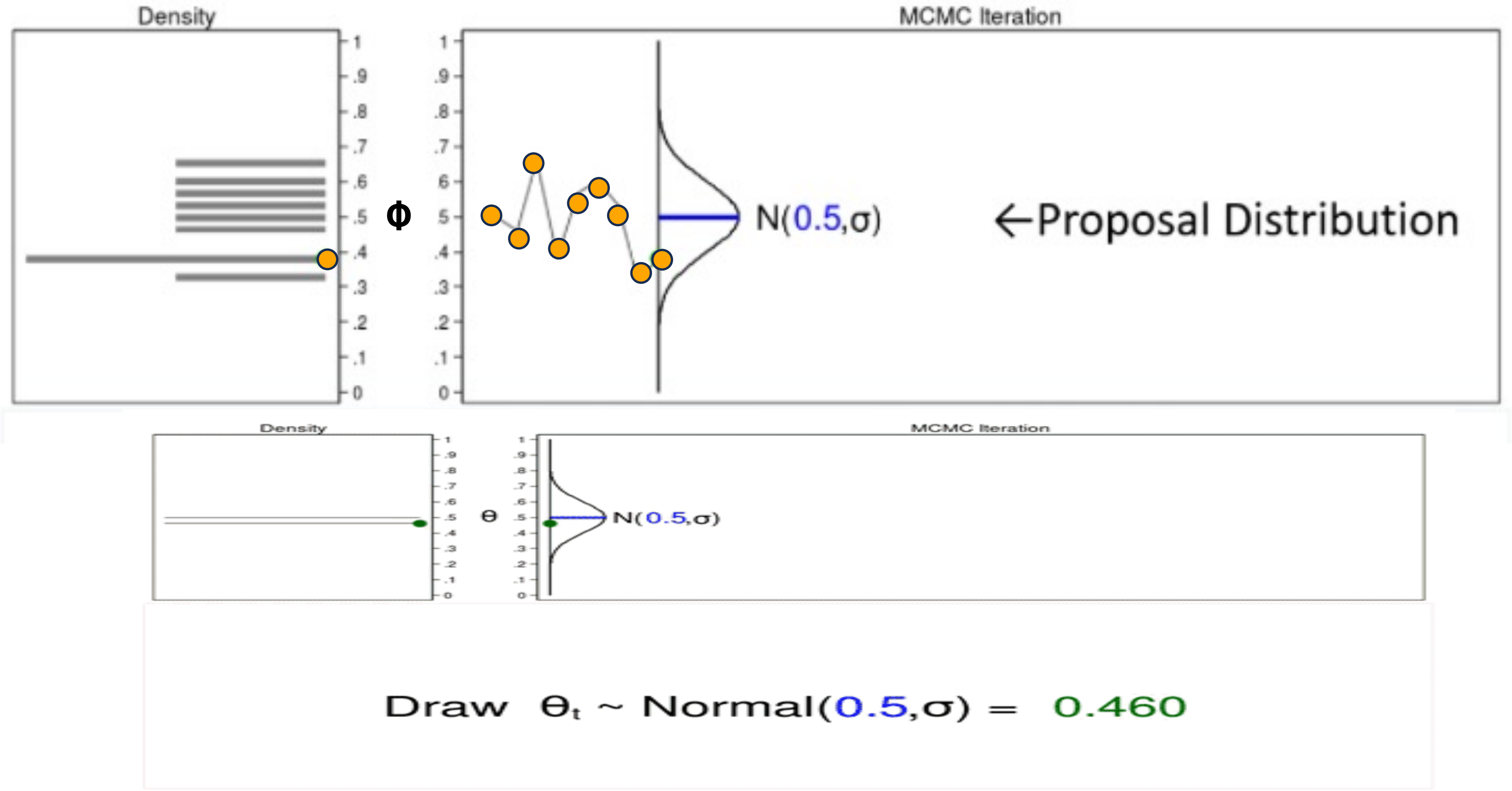
Approaches to fit data to models

Approach 2. Bayesian MCMC

Example: Estimate the mean of a normal distribution using MCMC

In this example, the true value Φ = 0.5

Figure 1: Proposal distributions, trace plots, and density plots

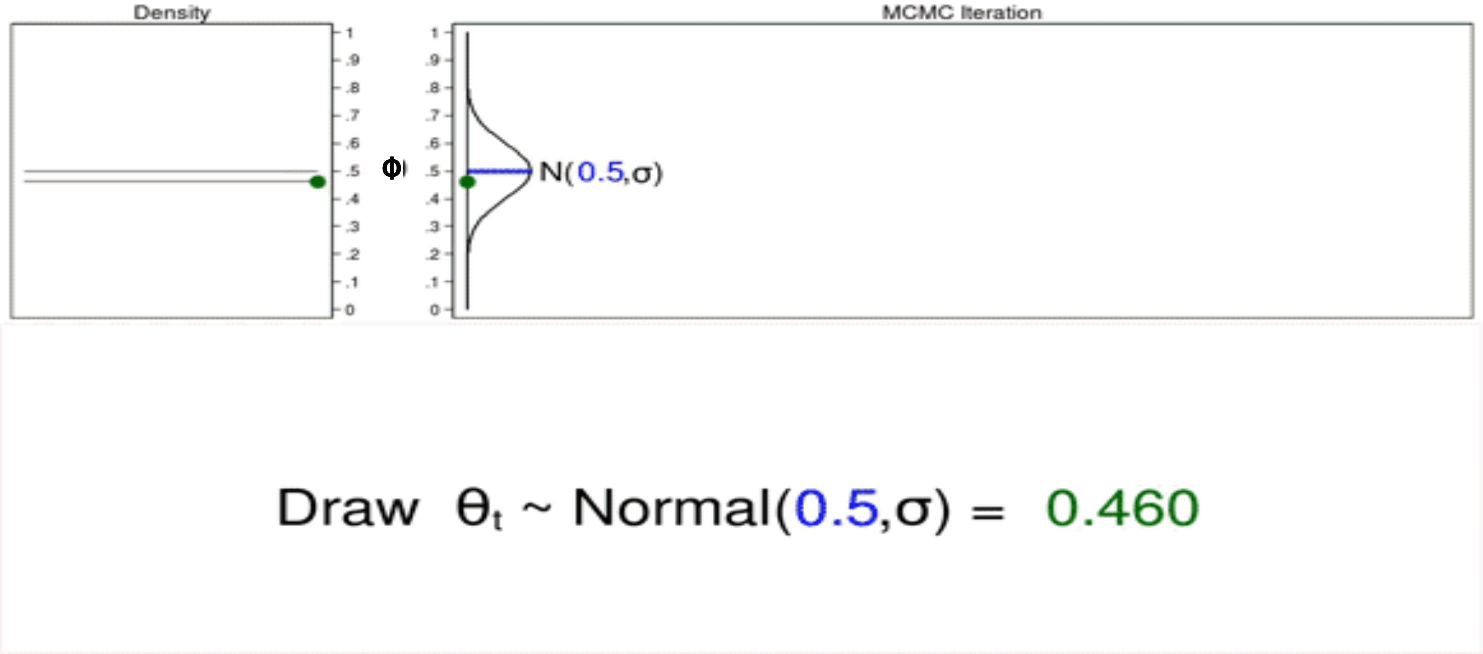


Approaches to fit data to models

Approach 2. Bayesian MCMC

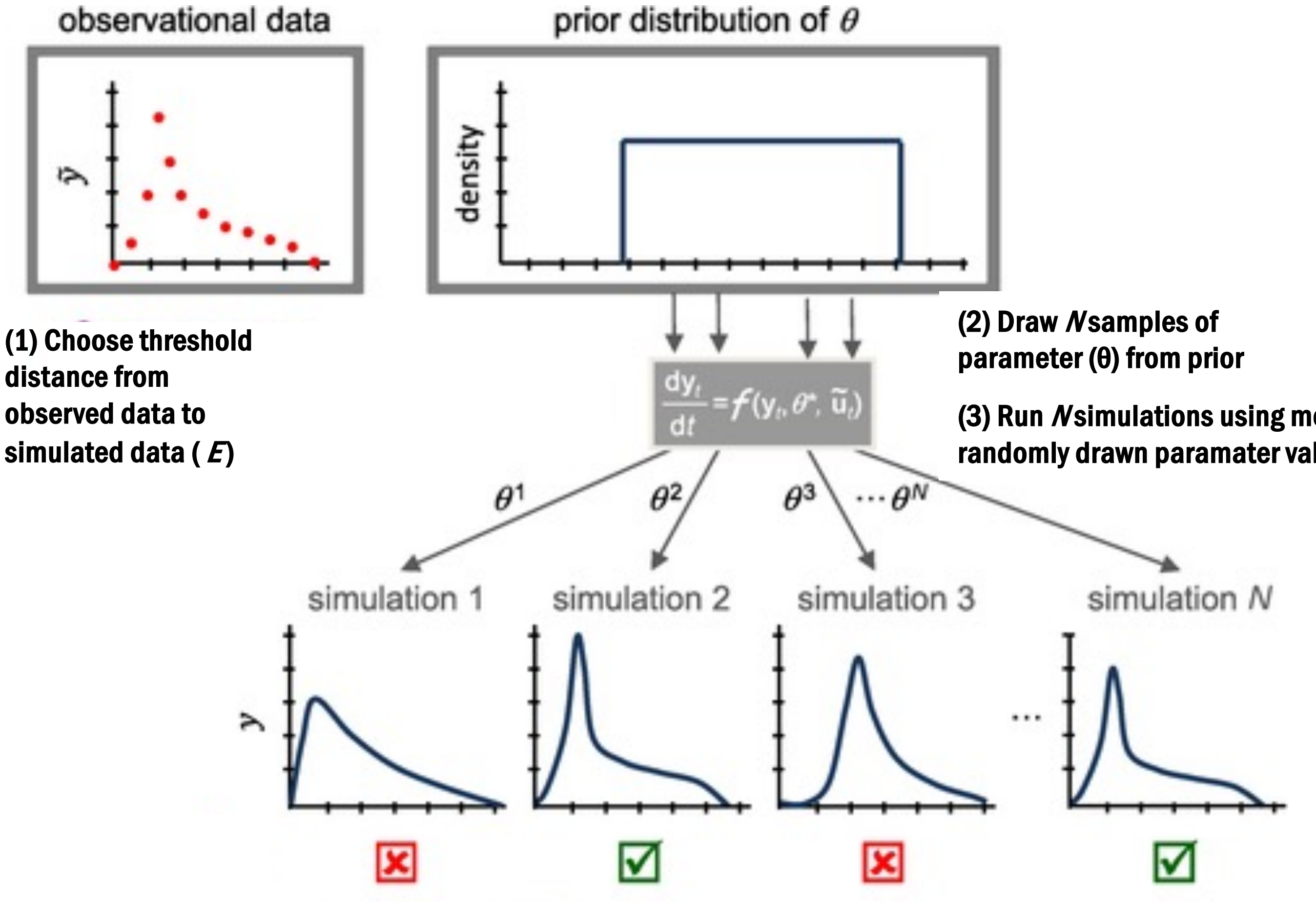
Example: Estimate the mean of a normal distribution using MCMC

In this example, the true value Φ = 0.5



Approaches to fit data to models

Approach 3. Approximate Bayesian computation (ABC)



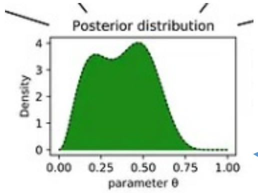
(1) Choose threshold distance from observed data to simulated data (E)

(2) Draw N samples of parameter (θ) from prior

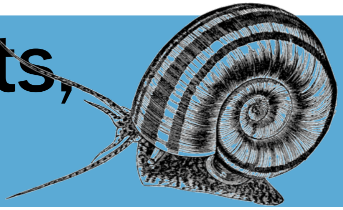
(3) Run N simulations using model and randomly drawn paramater values

(4) Accept or reject simulations based on threshold distance to observed data

(5) The accumulated accepted parameter values are used to approximate the posterior distribution for the parameter

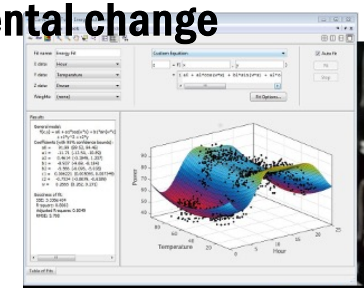


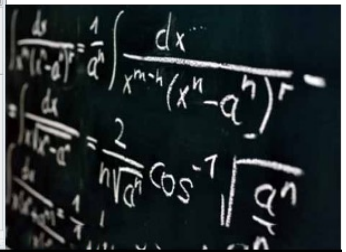
Research approach – combine surveys, experiments, and models

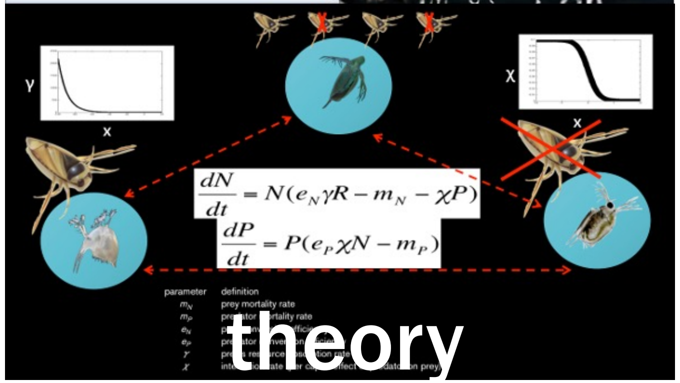


Research goals

- **Develop accurate models to explain and predict biodiversity**
- **Understand the consequences of evolution for communities, metacommunities, and ecosystems**
- **Determine how eco-evolutionary dynamics influence the response of biodiversity to environmental change**







theory

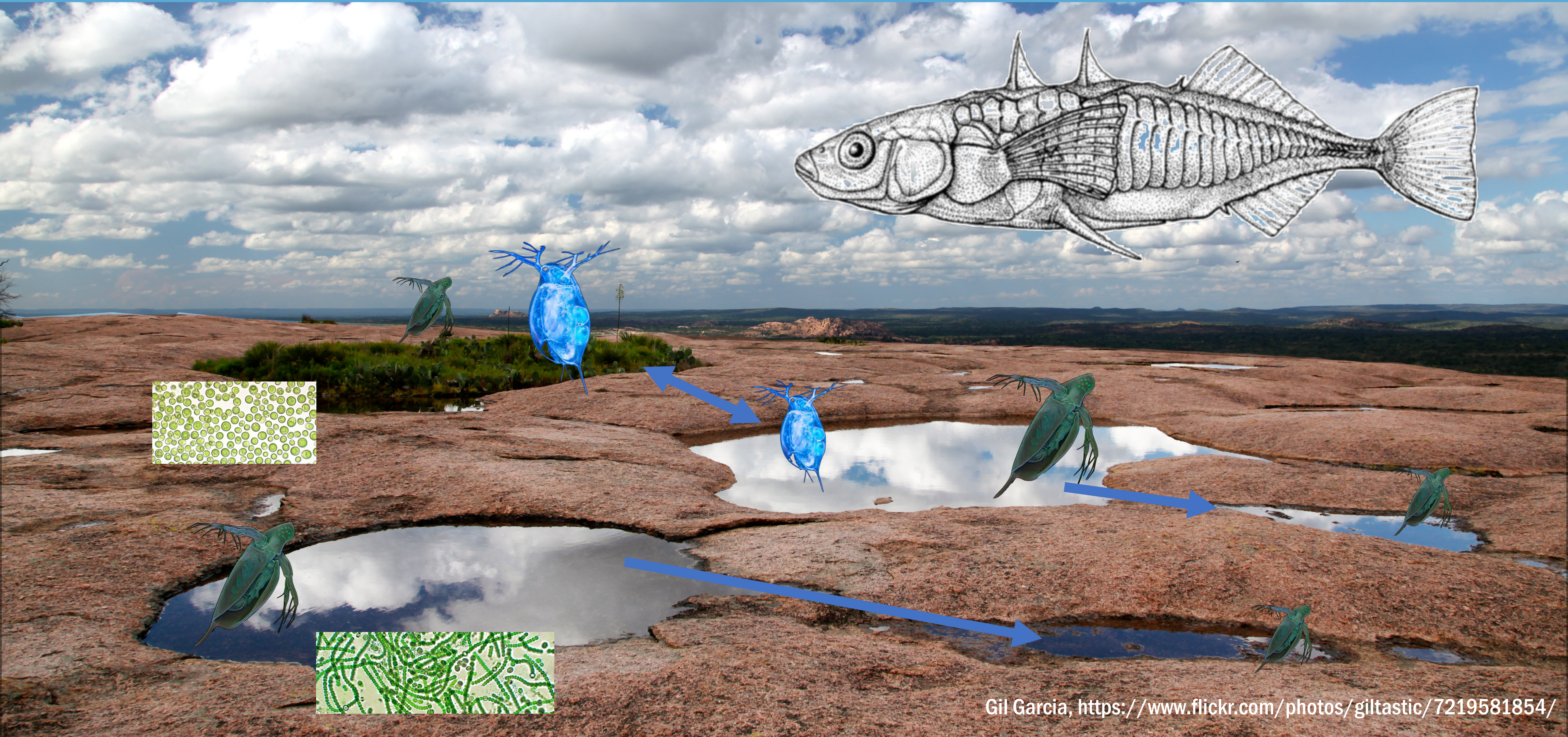
experiment

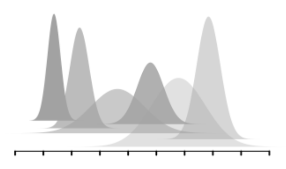
observation

Eco-evolutionary dynamics – overlapping ecological and evolutionary processes that are not independent of one another

Gil Garcia, https://www.flickr.com/photos/giltastic/7219581854/

Eco-evolutionary dynamics – overlapping ecological and evolutionary processes that are not independent of one another





Research approach –eco-evolutionary hypothesis testing to fit data to models

Candidate model

1. Identify candidate models for processes that structure observed data

Ecoevo H_0,H_A

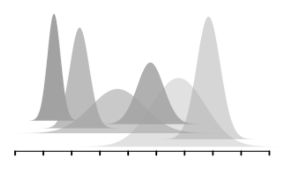
2. Generate alternative eco-evolutionary hypotheses

Simulations

3. Simulate possible observed data under each hypothesis

Compare to observed data

4. Compare observed data to alternative hypothesis simulations

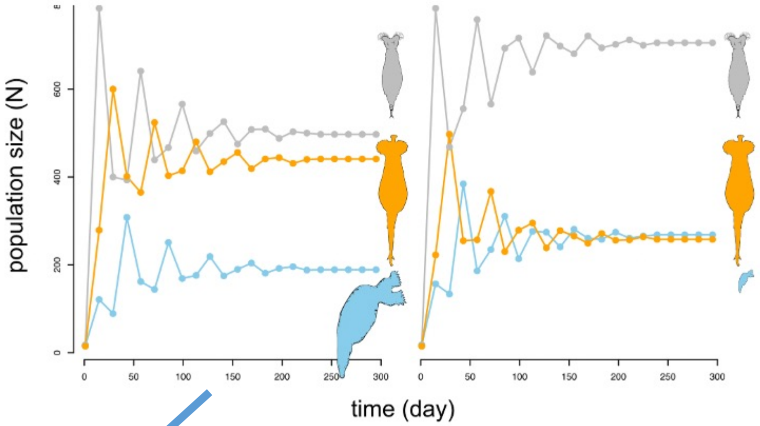


Research approach –eco-evolutionary hypothesis testing to fit data to models

Candidate model

1. Identify candidate models for processes that structure observed data

Ex 1) Coexistence / competition for a few species



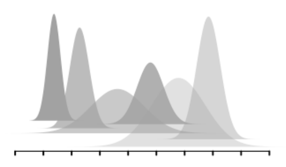
growth rate (as a function of a trait x)

$$\frac{N_{i,t+1}}{N_{i,t}} = \frac{\lambda_i(x)}{1 + \alpha_{ii} N_{i,t} + \alpha_{ij} N_{j,t}}$$

Intraspecific competition rate

Interspecific competition rate

Pantel & Becks 2023, *Trends in Ecology & Evolution*



Research approach –eco-evolutionary hypothesis testing to fit data to models

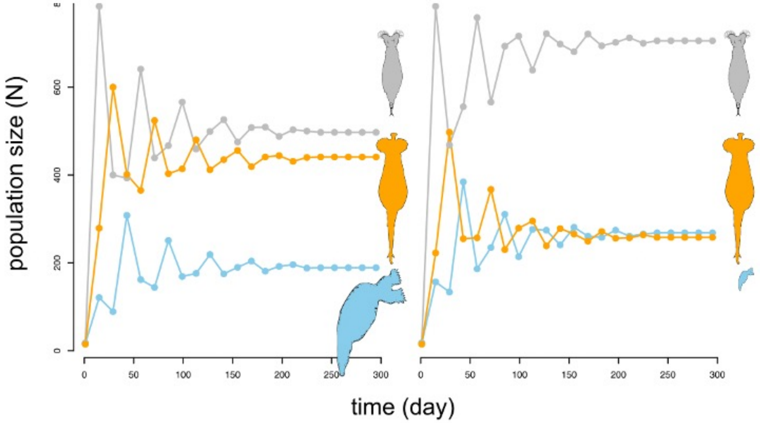
Ecoevo H_0,H_A

2. Generate alternative eco-evolutionary hypotheses

Simulations

3. Simulate possible observed data under each hypothesis

Ex 1) Coexistence / competition for a few species



H_{eco}

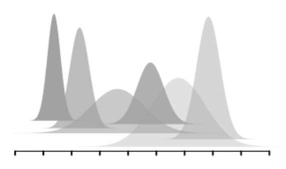
Trait x does not evolve

$H_{eco\text{-}evo}$

Trait x can evolve

$$\frac{N_{i,t+1}}{N_{i,t}} = \frac{\lambda_i(x)}{1 + \alpha_{ii}N_{i,t} + \alpha_{ij}N_{j,t}}$$

Pantel & Becks 2023, *Trends in Ecology & Evolution*



Research approach –eco-evolutionary hypothesis testing to fit data to models

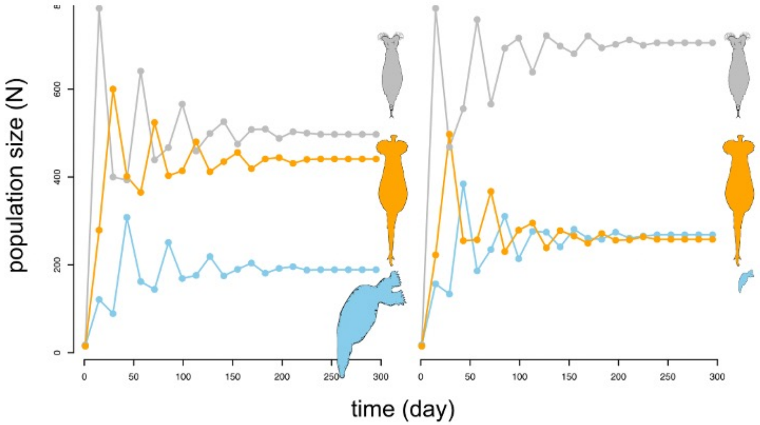
Ecoevo H_0, H_A

2. Generate alternative eco-evolutionary hypotheses

Simulations

3. Simulate possible observed data under each hypothesis

Ex 1) Coexistence / competition for a few species



H_{eco}

Trait *x* does not evolve

$H_{eco\text{-}evo}$

Trait *x* can evolve

$$\frac{N_{i,t+1}}{N_{i,t}} = \frac{\lambda_i(x)}{1 + \alpha_{ii} N_{i,t} + \alpha_{ij} N_{j,t}}$$

Growth rate → fitness

$$\lambda_t = \overline{W}_t = \widehat{W} e^{\frac{-\left[\left(\frac{w+(1-h^2)P}{P+w}\right)(E-x_t)\right]^2}{2(P+w)}}$$

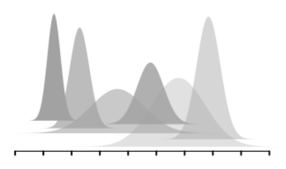
strength of selection

width of phenotypic distribution

heritability of trait *x*

Distance between trait *x* and optimum trait value E

Pantel & Becks 2023, *Trends in Ecology & Evolution*



Research approach – eco-evolutionary hypothesis testing to fit data to models

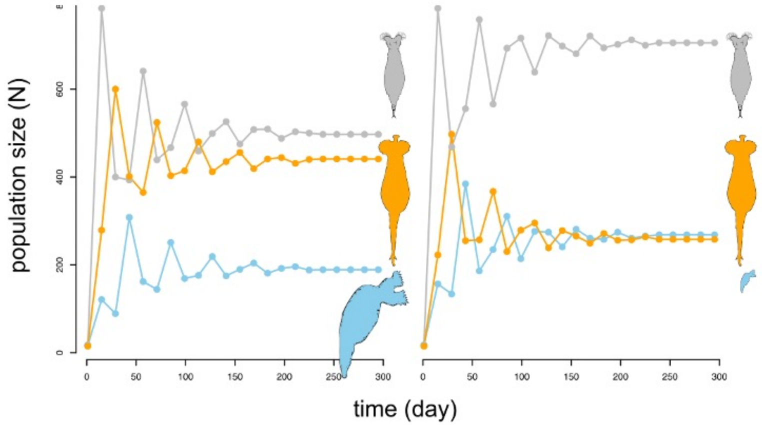
Ecoevo H_0, H_A

2. Generate alternative eco-evolutionary hypotheses

Simulations

3. Simulate possible observed data under each hypothesis

Ex 1) Coexistence / competition for a few species



H_{eco}

Trait *x* does not evolve

$H_{eco\text{-}evo}$

Trait *x* can evolve

$$\frac{N_{i,t+1}}{N_{i,t}} = \frac{\lambda_i(x)}{1 + \alpha_{ii} N_{i,t} + \alpha_{ij} N_{j,t}}$$

Growth rate → fitness

strength of selection

width of phenotypic distribution

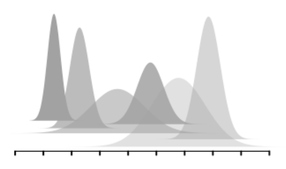
heritability of trait *x*

Distance between trait *x* and optimum trait value E

$$\lambda_t = \overline{W}_t = \widehat{W} e^{\frac{-\left[\left(\frac{w+(1-h^2)P}{P+w}\right)(E-x_t)\right]^2}{2(P+w)}}$$

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-[(\frac{w+(1-h^2)P}{P+w})(E-x_t)]^2}{2(P+w)}} N_{i,t}}{1 + \alpha_{ii} N_{i,t} + \sum_j \alpha_{ij} N_{j,t}}$$

Pantel & Becks 2023, *Trends in Ecology & Evolution*

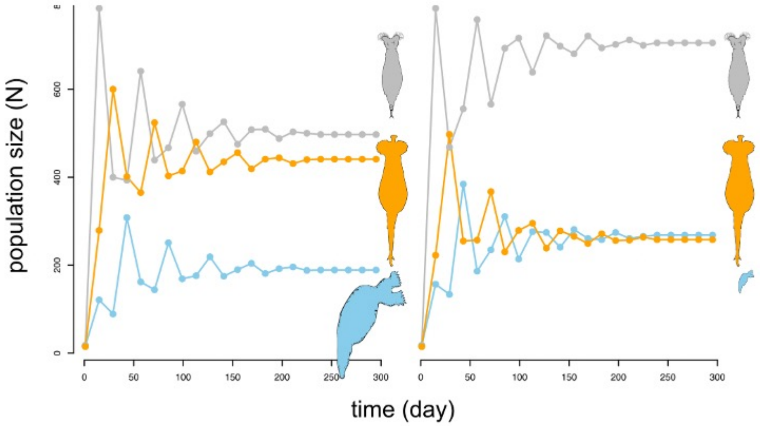


Research approach –eco-evolutionary hypothesis testing to fit data to models

Ecoevo H_0,H_A — 2. Generate alternative eco-evolutionary hypotheses

Simulations — 3. Simulate possible observed data under each hypothesis

Ex 1) Coexistence / competition for a few species



H_{eco}

Trait x does not evolve

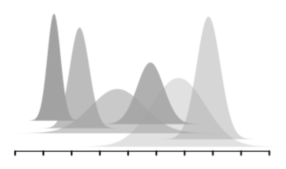
$$\frac{N_{i,t+1}}{N_{i,t}} = \frac{\lambda_i(x)}{1 + \alpha_{ii}N_{i,t} + \alpha_{ij}N_{j,t}}$$

$H_{eco\text{-}evo}$

Trait x can evolve

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-[(\frac{w+(1-h^2)P}{P+w})(E-x_t)]^2}{2(P+w)}} N_{i,t}}{1 + \alpha_{ii}N_{i,t} + \sum_j \alpha_{ij}N_{j,t}}$$

Pantel & Becks 2023, *Trends in Ecology & Evolution*



Research approach –eco-evolutionary hypothesis testing to fit data to models

Simulations

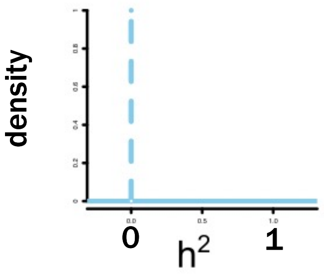
3. Simulate possible observed data under each hypothesis

Compare to observed data

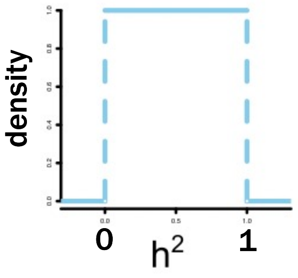
4. Compare observed data to alternative hypothesis simulations

sample a candidate parameter vector θ* from a proposed prior distribution, 100,000 simulations per model

H_1: No evolution (h^2=0)

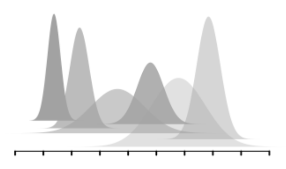


H_2: Evolution (h^2>0)



Approximate Bayesian Computing (eg Beaumont 2010)

Model parameters: h^2, α, w, $\widehat{W}$, P



Research approach –eco-evolutionary hypothesis testing to fit data to models

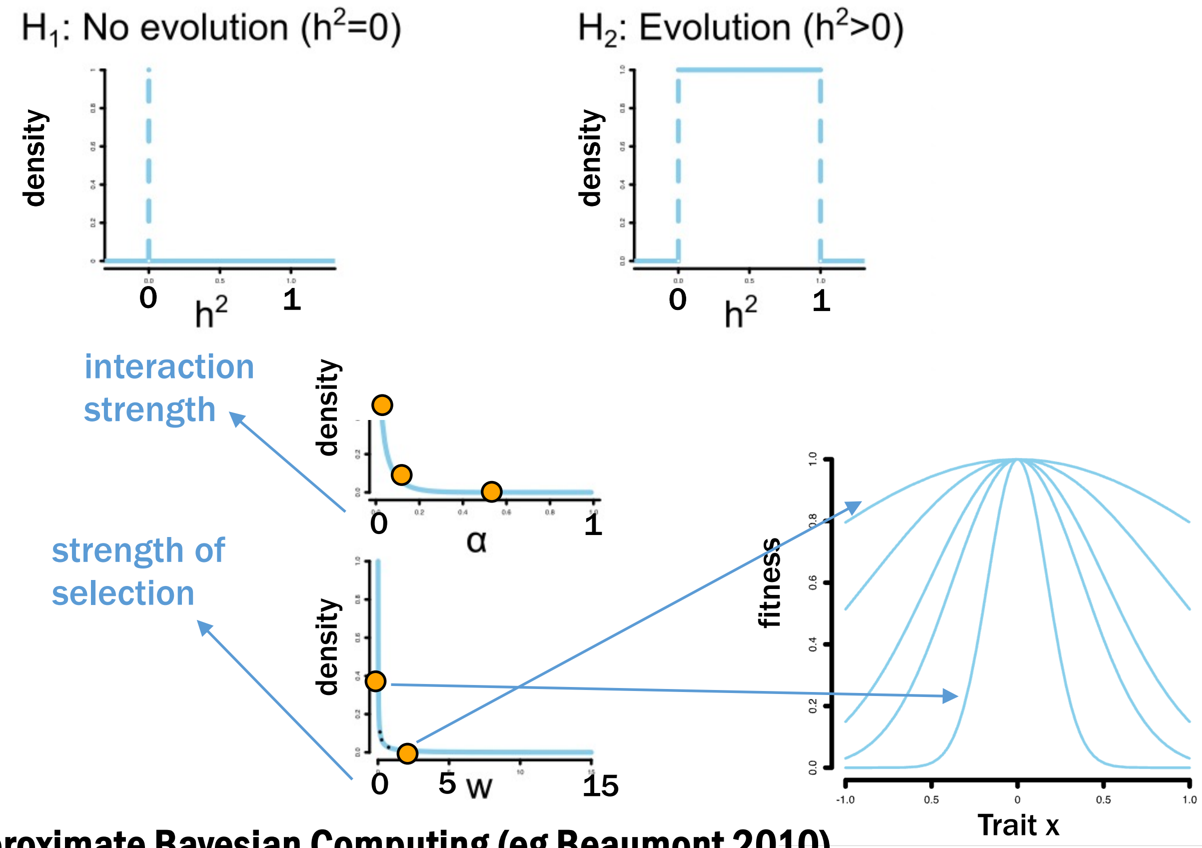
Simulations

3. Simulate possible observed data under each hypothesis

Compare to observed data

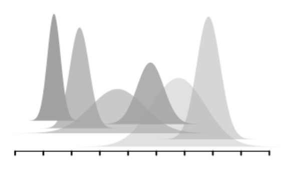
4. Compare observed data to alternative hypothesis simulations

sample a candidate parameter vector θ* from a proposed prior distribution, 100,000 simulations per model



Approximate Bayesian Computing (eg Beaumont 2010)

Model parameters: h^2, α, w, $\widehat{W}$, P



Research approach –eco-evolutionary hypothesis testing to fit data to models

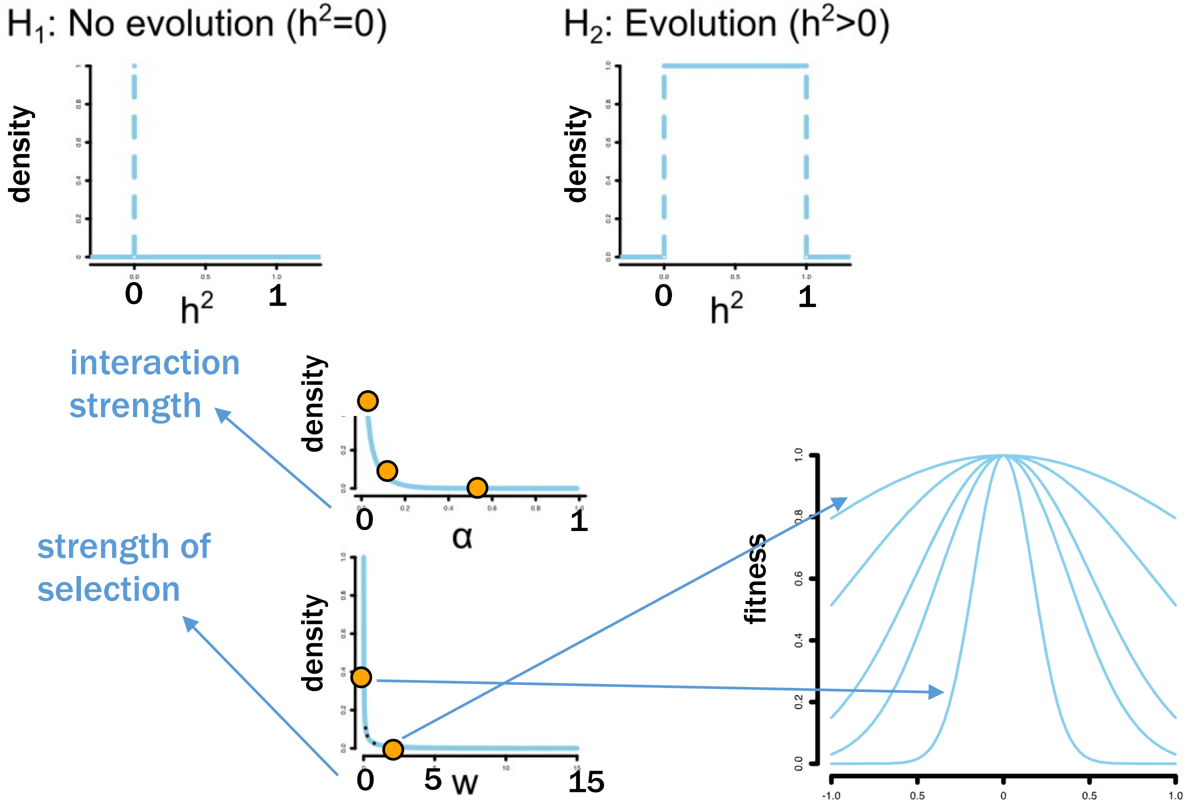
Simulations

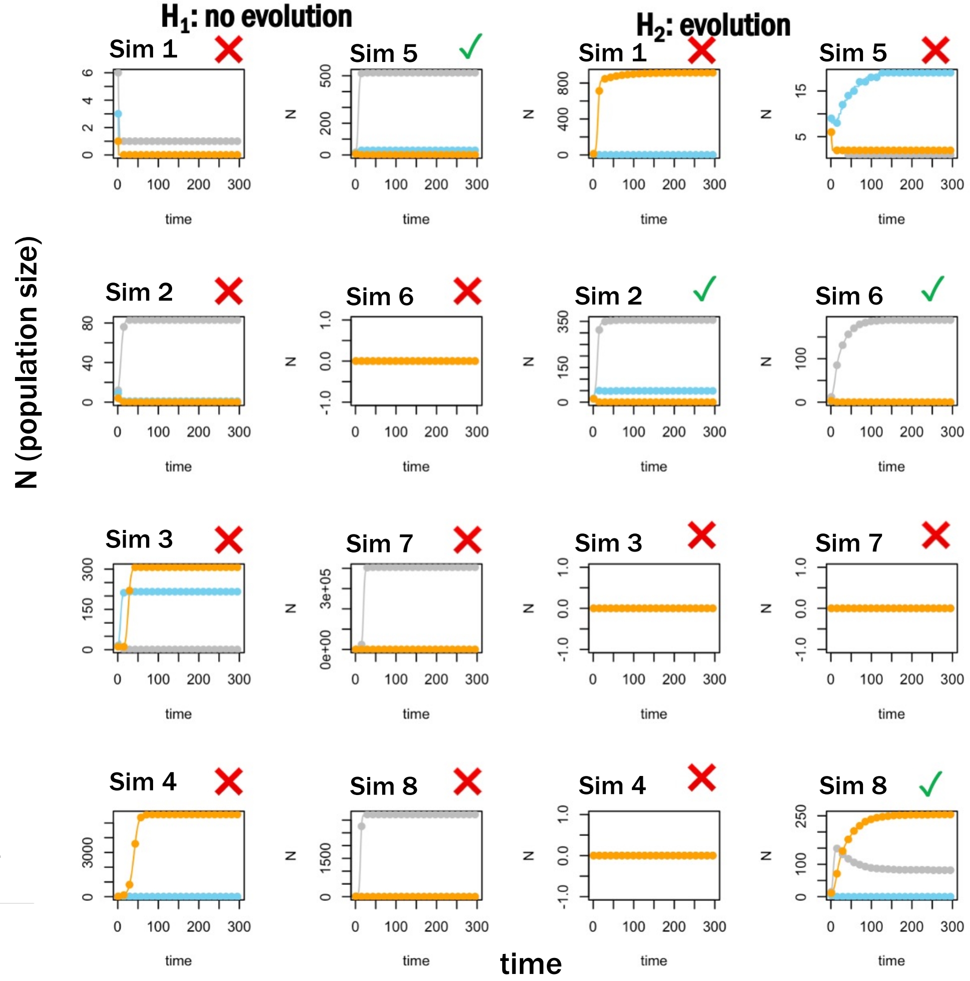
3. Simulate possible observed data under each hypothesis

Compare to observed data

4. Compare observed data to alternative hypothesis simulations

sample a candidate parameter vector θ* from a proposed prior distribution, 100,000 simulations per model

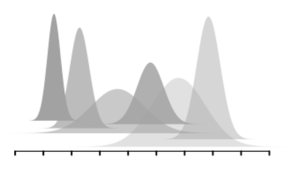




compare the simulated dataset x* to the observed data x_0, using a distance function d and a tolerance ε: if $d(x_0, x^*) \leq \varepsilon$, accept θ*

Approximate Bayesian Computing (eg Beaumont 2010)

Model parameters: h^2, α, w, $\widehat{W}$, P



Research approach –eco-evolutionary hypothesis testing to fit data to models

Compare to observed data

4. Compare observed data to alternative hypothesis simulations

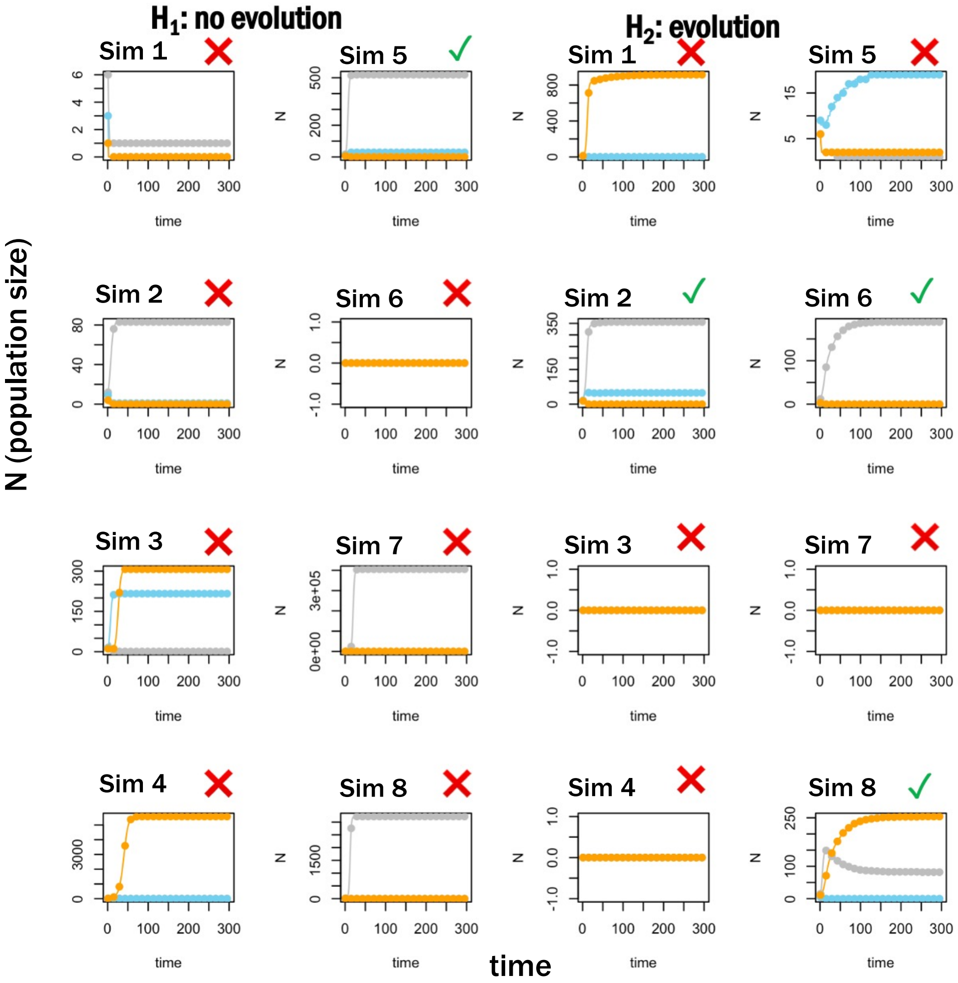
(i) Estimate posterior probability of alternative model hypotheses

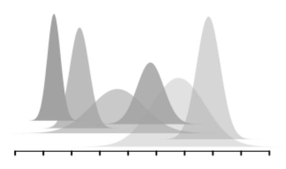
H_{eco}

$$\frac{n_{i,t+1}}{n_{i,t}} = \frac{\lambda_i}{1 + \alpha_{ii} N_{i,t} + \alpha_{ij} N_{j,t}}$$

$H_{eco\text{-}evo}$

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-\left[\left(\frac{w+(1-h^2)P}{P+w}\right)(E-x_t)\right]^2}{2(P+w)}} N_{i,t}}{1 + \alpha_{ii} N_{i,t} + \sum_j \alpha_{ij} N_{j,t}}$$





Research approach –eco-evolutionary hypothesis testing to fit data to models

Compare to observed data

4. Compare observed data to alternative hypothesis simulations

(i) Estimate posterior probability of alternative model hypotheses

H_{eco}

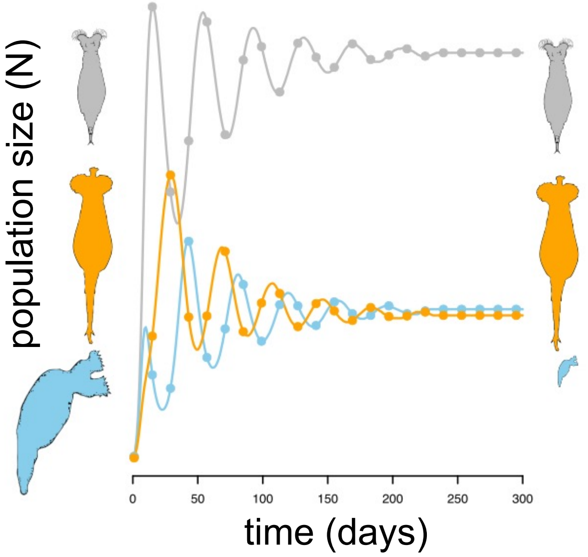
$$\frac{n_{i,t+1}}{n_{i,t}} = \frac{\lambda_i}{1 + \alpha_{ii} N_{i,t} + \alpha_{ij} N_{j,t}}$$

$H_{eco\text{-}evo}$

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-[(\frac{w+(1-h^2)P}{P+w})(E-x_t)]^2}{2(P+w)}} N_{i,t}}{1 + \alpha_{ii} N_{i,t} + \sum_j \alpha_{ij} N_{j,t}}$$

Case 1. With simulated model + data

With x_0 and x_{300} trait values



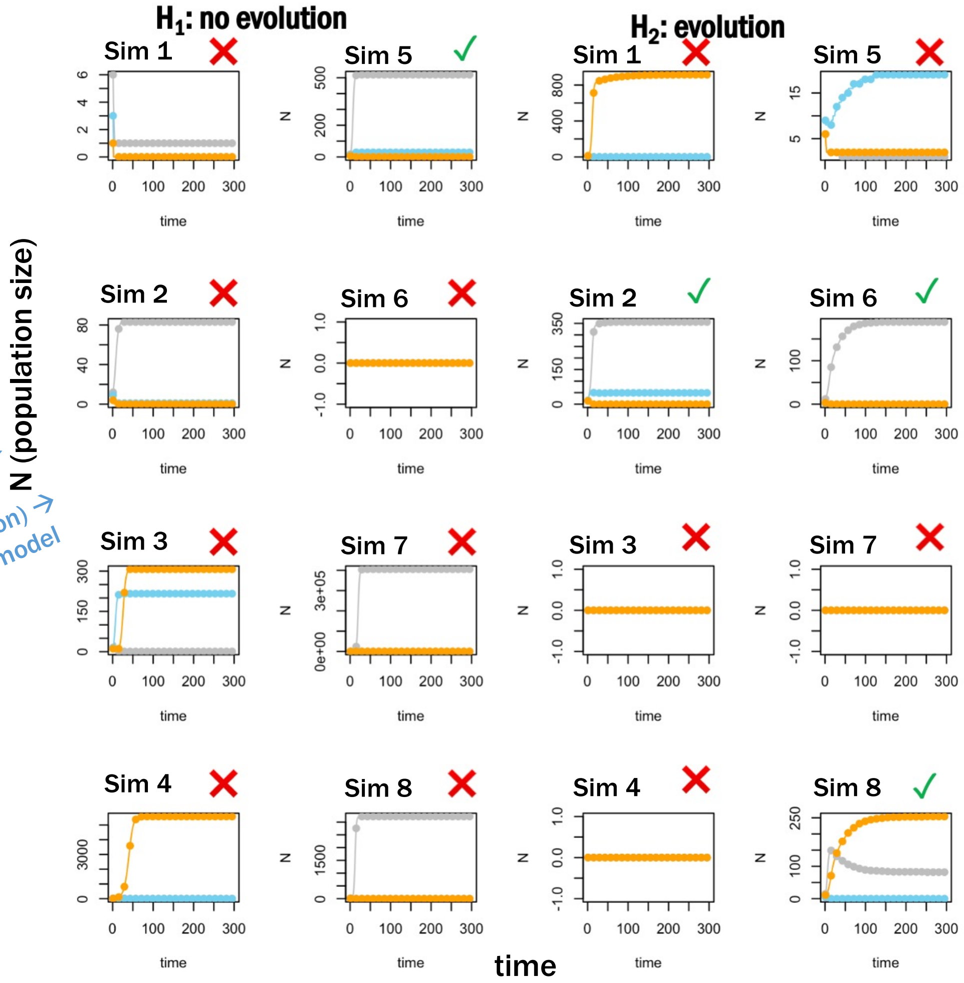
Posterior model probabilities for Observed Data

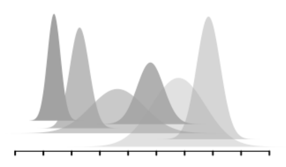
H_{eco} = 0.0001

$H_{eco\text{-}evo}$ = 0.9999

Based on % Accepted vs. Rejected per H_A model

Approximate Bayesian Computing + Neural network (for model classification) → trained on 100,000 simulations per model





Research approach –eco-evolutionary hypothesis testing to fit data to models

Compare to observed data

4. Compare observed data to alternative hypothesis simulations

(i) Estimate posterior probability of alternative model hypotheses

H_{eco}

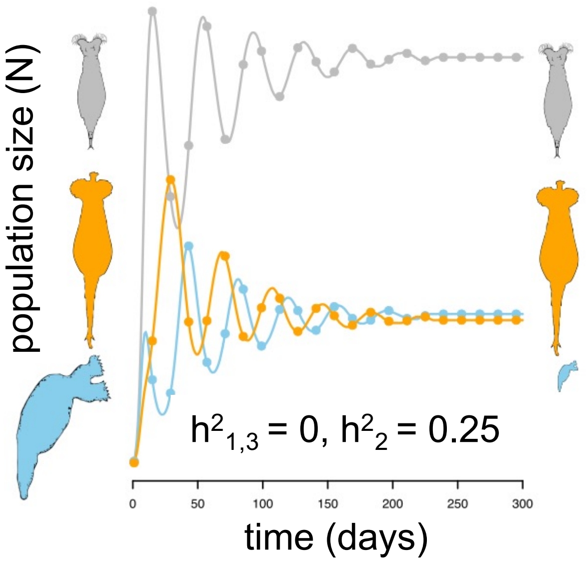
$$\frac{n_{i,t+1}}{n_{i,t}} = \frac{\lambda_i}{1+\alpha_{ii}N_{i,t}+\alpha_{ij}N_{j,t}}$$

$H_{eco\text{-}evo}$

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-\left[\left(\frac{w+(1-h^2)P}{P+w}\right)(E-x_t)\right]^2}{2(P+w)}} N_{i,t}}{1+\alpha_{ii}N_{i,t}+\sum_j \alpha_{ij}N_{j,t}}$$

Case 1. With simulated model + data

With x_0 and x_{300} trait values



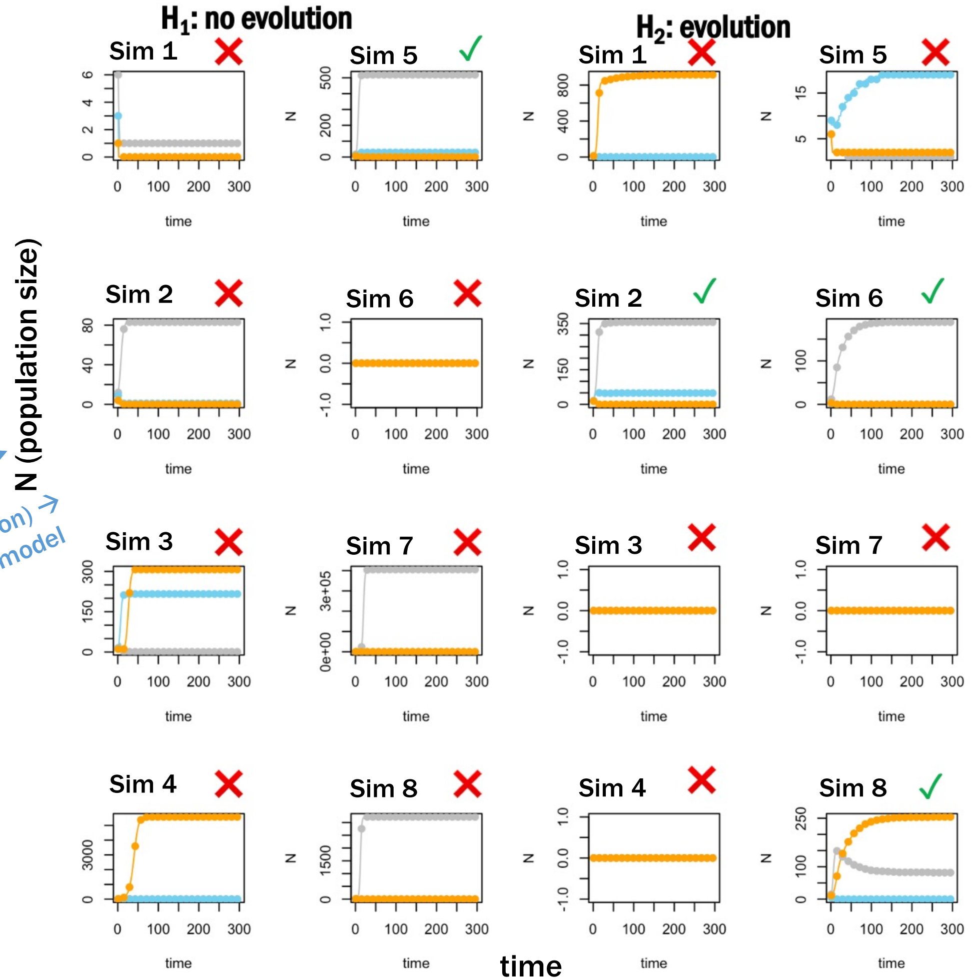
Posterior model probabilities for Observed Data

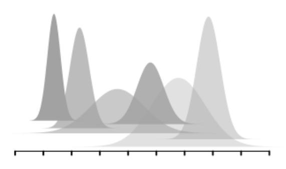
H_{eco} = 0.0001 ✗

$H_{eco\text{-}evo}$ = 0.9999 ✓

Based on % Accepted vs. Rejected per H_A model

Approximate Bayesian Computing + Neural network (for model classification) → trained on 100,000 simulations per model





Research approach – eco-evolutionary hypothesis testing to fit data to models

Compare to observed data

4. Compare observed data to alternative hypothesis simulations

(i) Estimate posterior probability of alternative model hypotheses

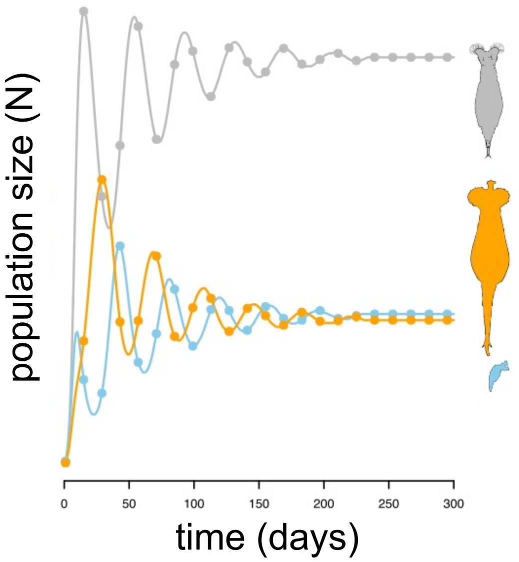
H_{eco}

$$\frac{n_{i,t+1}}{n_{i,t}} = \frac{\lambda_i}{1 + \alpha_{ii} N_{i,t} + \alpha_{ij} N_{j,t}}$$

$H_{eco\text{-}evo}$

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-[(\frac{w+(1-h^2)P}{P+w})(E-x_t)]^2}{2(P+w)}} N_{i,t}}{1 + \alpha_{ii} N_{i,t} + \sum_j \alpha_{ij} N_{j,t}}$$

Case 2. With simulated model + data
With x_{300} trait values only (!!)



Posterior model probabilities for Observed Data

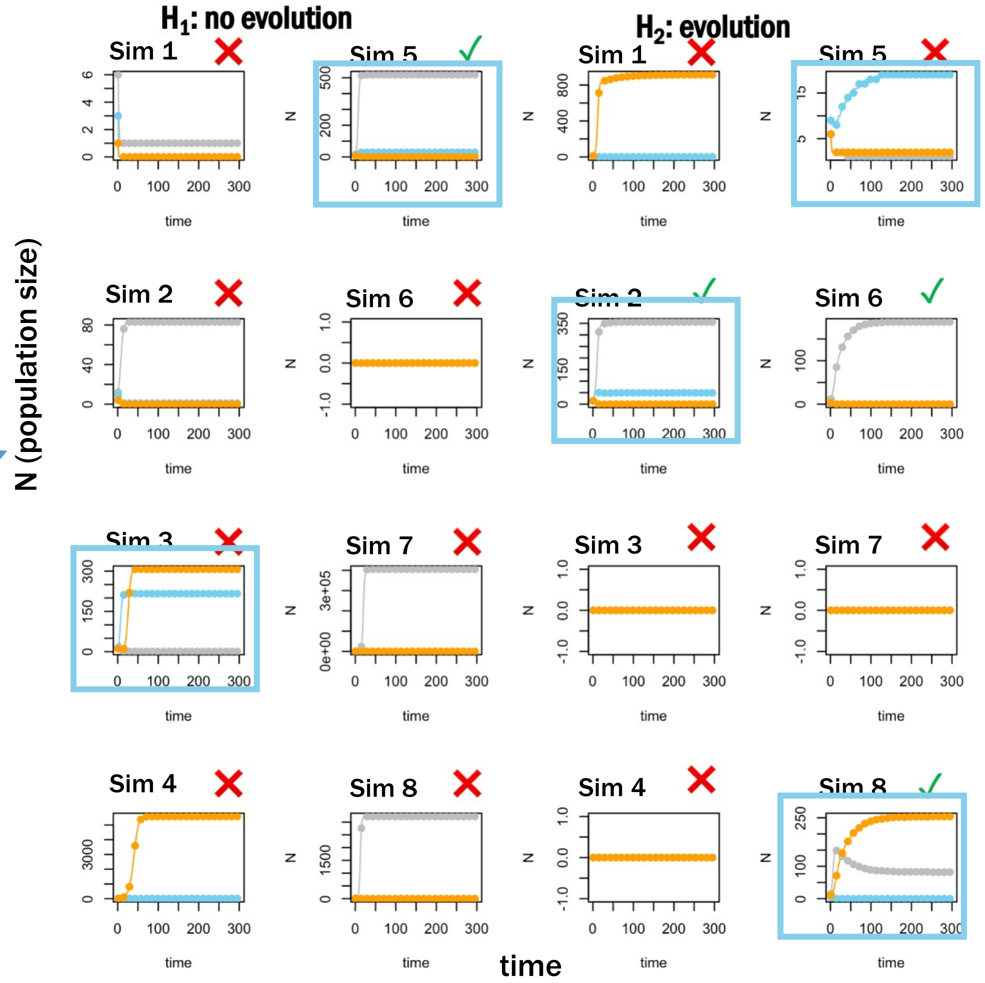
H_{eco} = 0.4676
$H_{eco\text{-}evo}$ = 0.5324

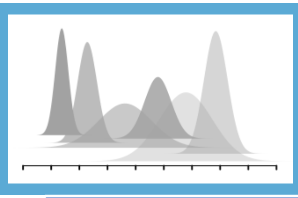
Based on % Accepted vs. Rejected per H_A model

Case 3. With simulated model + data
With x_{300} trait values only (!!)
Restricted to 3-species persisting simulations only

Posterior model probabilities for Observed Data

H_{eco} = 0.0783
$H_{eco\text{-}evo}$ = 0.9217





Research approach –eco-evolutionary hypothesis testing to fit data to models

Compare to observed data

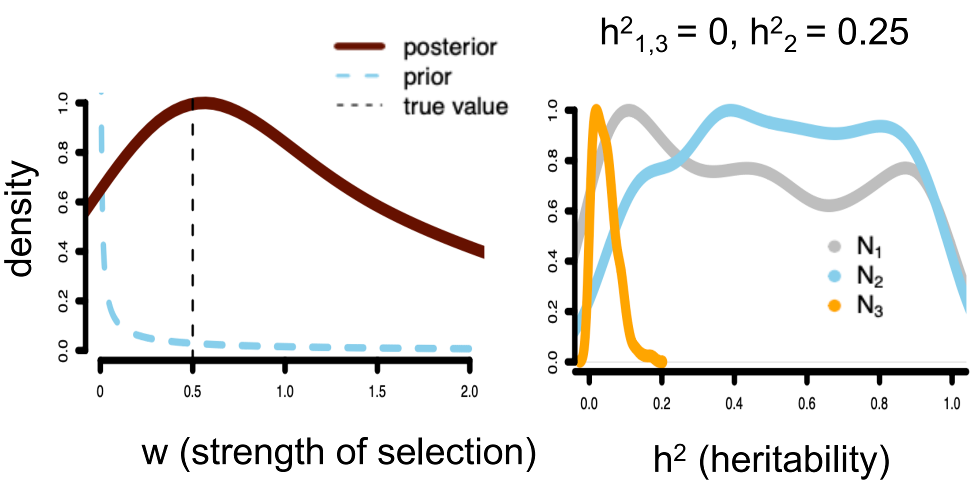
4. Compare observed data to alternative hypothesis simulations

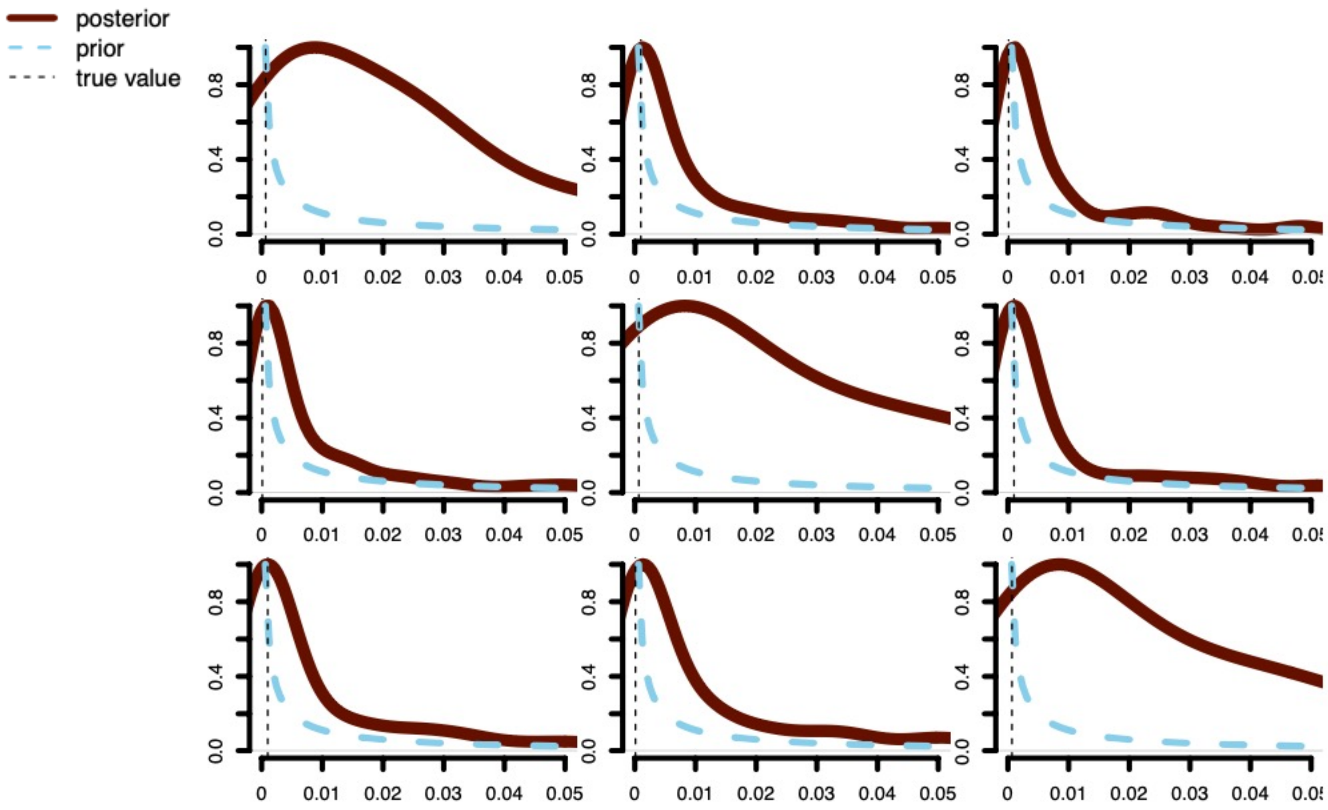
(i) Estimate posterior probability of alternative model hypotheses

(ii) Use best-fit (or consensus) models to estimate parameters for observed data, make predictions for future cases

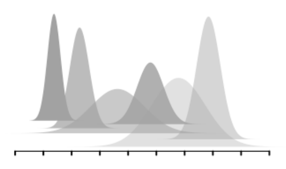
$$\begin{bmatrix}\alpha_{11} & \alpha_{12} & \alpha_{13}\\ \alpha_{21} & \alpha_{22} & \alpha_{23}\\ \alpha_{31} & \alpha_{32} & \alpha_{33}\end{bmatrix} = \begin{bmatrix}0.000666 & 0.001 & 0.000111\\ 0.000111 & 0.000666 & 0.001\\ 0.001 & 0.000111 & 0.000666\end{bmatrix}$$

Posterior parameter values (unobserved)





Pantel & Becks 2023, *Trends in Ecology & Evolution*



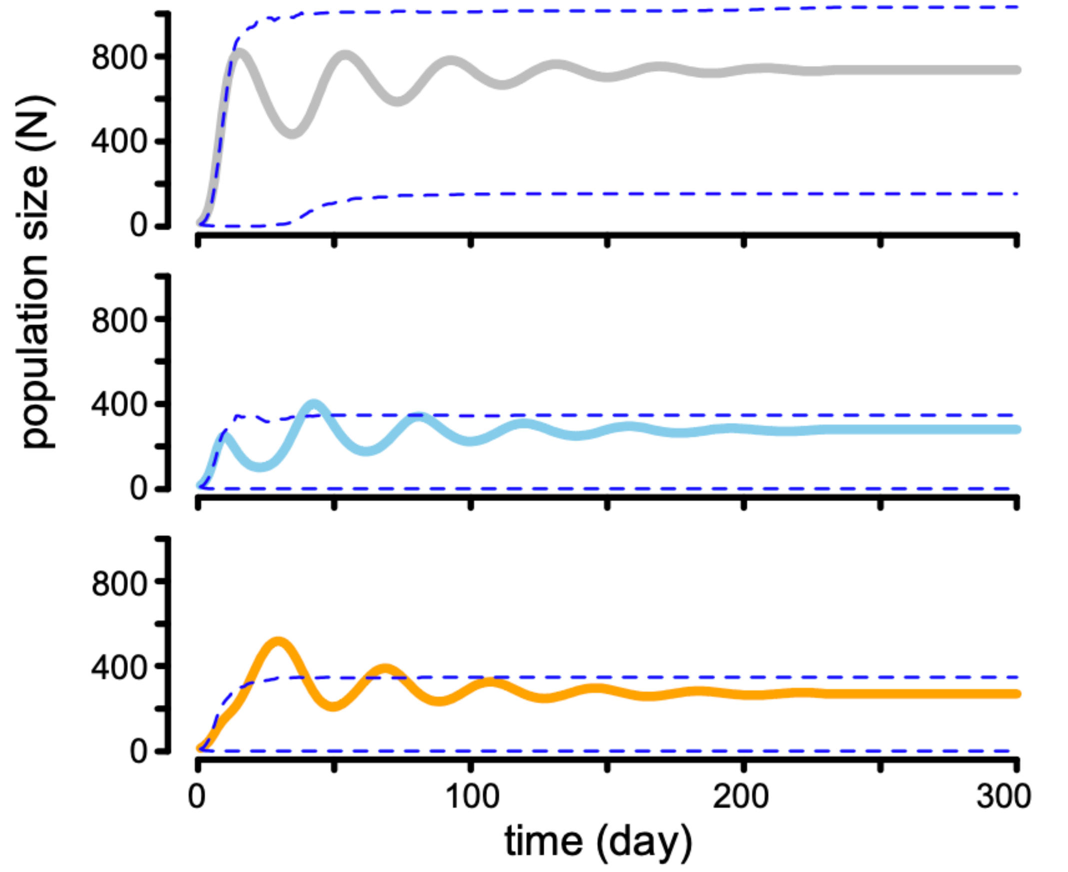
Research approach –eco-evolutionary hypothesis testing to fit data to models

Compare to observed data

4. Compare observed data to alternative hypothesis simulations

(i) Estimate posterior probability of alternative model hypotheses
(ii) Use best-fit (or consensus) models to estimate parameters for observed data, make predictions for future cases

Posterior predictive checks with 95% CI (- -) generated using acceptedsimulations (✓) with lowest Euclidean distance to observed summary statistics ( )



Pantel & Becks 2023, *Trends in Ecology & Evolution*