Fitting Models to Data in Ecology and Evolution

Samraat Pawar

Department of Life Sciences (Silwood Park)

Imperial College
London

October 25, 2018

MECHANISTIC VS. PHENOMENOLOGICAL MODELS

What does "modelling data" mean to you?

Mechanistic vs. Phenomenological models

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data

MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS

MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS
- *Empirical/Phenomenological models* establish the existence of STATISTICALLY SIGNIFICANT, NON-RANDOM PATTERNS or PHENOMENA in empirical data

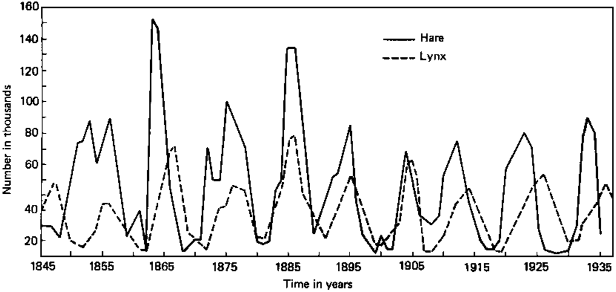
MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS
- *Empirical/Phenomenological models* establish the existence of STATISTICALLY SIGNIFICANT, NON-RANDOM PATTERNS or PHENOMENA in empirical data
  - They make no assumptions about the processes or mechanisms that generate the patterns

Mechanistic vs. Phenomenological models

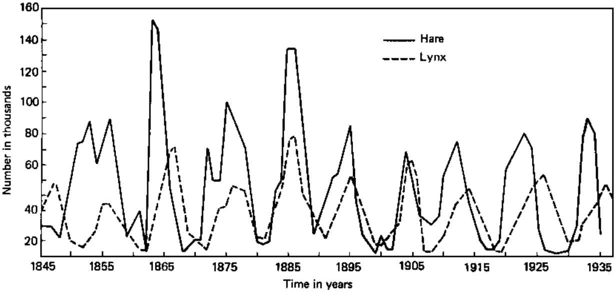
- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS
- *Empirical/Phenomenological models* establish the existence of STATISTICALLY SIGNIFICANT, NON-RANDOM PATTERNS or PHENOMENA in empirical data
  - They make no assumptions about the processes or mechanisms that generate the patterns
  - That is, these models lack a THEORETICAL BASIS

Mechanistic vs. Phenomenological model fitting



source: `https://www.cds.caltech.edu/~murray/amwiki/images/8/8f/LHgraph.gif`

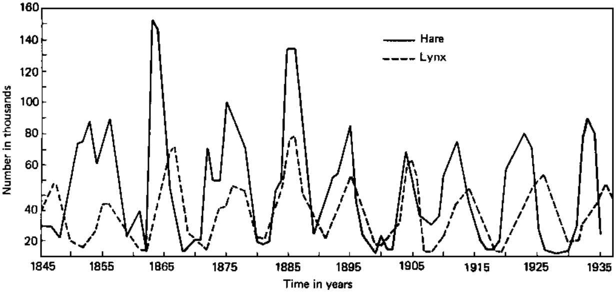
Mechanistic vs. Phenomenological model fitting



source: *https://www.cds.caltech.edu/~murray/amwiki/images/8/8f/LHgraph.gif*

- **Mechanistic model**: *The Lynx-Hare Cycle is driven by density-dependent population growth in hares*

Mechanistic vs. Phenomenological model fitting



source: `https://www.cds.caltech.edu/~murray/amwiki/images/8/8f/LHgraph.gif`

- **Mechanistic model**: *The Lynx-Hare Cycle is driven by density-dependent population growth in hares*
- **Phenomenological model**: The Lynx and Hare Cycles have a significant asynchrony (period shift) of xx years

Mechanistic vs. Phenomenological model fitting

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)

MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES

Mechanistic vs. Phenomenological model fitting

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting

MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)

MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)
- Mechanistic model-fitting *tries* to validate a mechanistic model that can explain the observed phenomenological pattern and generate MORE ACCURATE, MECHANISTIC HYPOTHESES

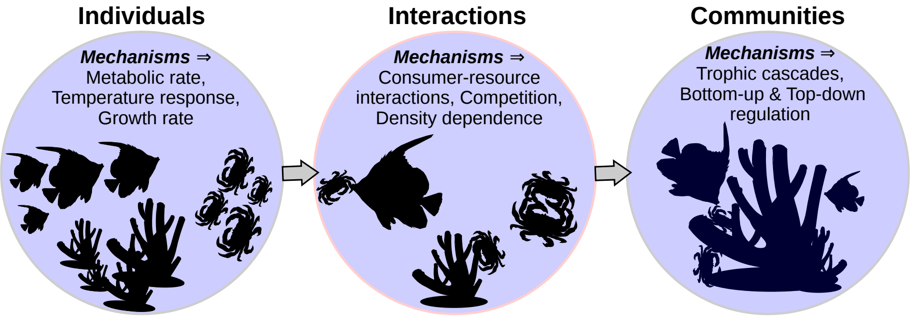
Mechanistic vs. Phenomenological model fitting

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)
- Mechanistic model-fitting *tries* to validate a mechanistic model that can explain the observed phenomenological pattern and generate MORE ACCURATE, MECHANISTIC HYPOTHESES
  - Example: *How* climatic temperature *drives* the Lynx-Hare cycle

MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)
- Mechanistic model-fitting *tries* to validate a mechanistic model that can explain the observed phenomenological pattern and generate MORE ACCURATE, MECHANISTIC HYPOTHESES
  - Example: *How* climatic temperature *drives* the Lynx-Hare cycle
- *Ultimately, successful, EMPIRICALLY-GROUNDED mechanistic models are the best path towards a THEORY in any scientific discipline (including ecology and evolution)*

Mechanistic vs. Phenomenological model fitting



Mechanistic models in Ecology and Evolution?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*

MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological

Mechanistic models in Ecology and Evolution?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological

Mechanistic models in Ecology and Evolution?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*

MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,

Mechanistic models in Ecology and Evolution?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,
  - ... and partly because we are (or are forced to be) interested in FORECASTING rather than EXPLAINING.

MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,
  - ... and partly because we are (or are forced to be) interested in FORECASTING rather than EXPLAINING.
- *So the big question is, can we FORECAST WITHOUT EXPLAINING?*

Mechanistic models in Ecology and Evolution?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,
  - ... and partly because we are (or are forced to be) interested in FORECASTING rather than EXPLAINING.
- *So the big question is, can we FORECAST WITHOUT EXPLAINING?*
  - For example, disease outbreaks: Do we really need to care about the underlying mechanisms if we can predict a future event using phenomenological modelling (e.g., Machine-learning of time series patterns)?

What are Mechanisms?

- Somewhat subjective!

What are Mechanisms?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1-\frac{N_t}{k}\right)}$$

What are Mechanisms?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1-\frac{N_t}{k}\right)}$$

- What is the mechanism?

What are Mechanisms?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1-\frac{N_t}{k}\right)}$$

- What is the mechanism?

What are Mechanisms?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1-\frac{N_t}{k}\right)}$$

- What is the mechanism? — Density dependence through scramble competition (Brannstrom & Sumpter 2005)
- If the Ricker model and another model with contest competition were compared with data — some would call it mechanistic modelling because one is trying to get at the underlying mechanism, scramble or contest competition

What are Mechanisms?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1 - \frac{N_t}{k}\right)}$$

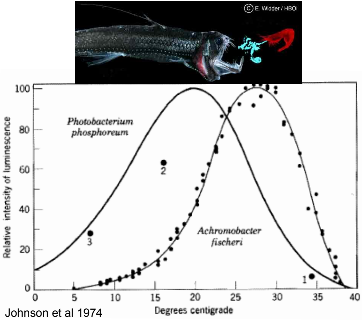
- What is the mechanism? — Density dependence through scramble competition (Brannstrom & Sumpter 2005)
- If the Ricker model and another model with contest competition were compared with data — some would call it mechanistic modelling because one is trying to get at the underlying mechanism, scramble or contest competition
- But is this REALLY mechanistic? What are *r* and *k* really?

EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

- Proponents of *Metabolic Theory of Ecology* argue that we have not progressed far enough towards mechanistic modelling because metabolism has been ignored

EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

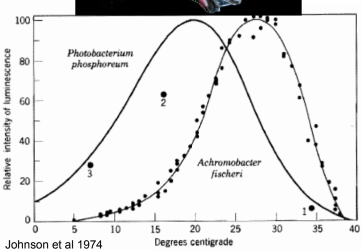
- Proponents of *Metabolic Theory of Ecology* argue that we have not progressed far enough towards mechanistic modelling because metabolism has been ignored



Johnson et al 1974

Example of a fundamental mechanism: metabolic rate

- Proponents of *Metabolic Theory of Ecology* argue that we have not progressed far enough towards mechanistic modelling because metabolism has been ignored

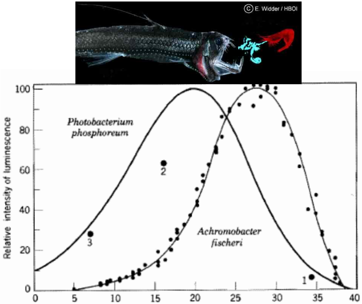


$$B = B_0 \boxed{e^{-\frac{E}{kT}}} f(T, T_{pk}, E_D)$$

T = temperature (K)
k = Boltzmann constant (eV K^{-1})
E = Activation energy (eV)
T_{pk} = Temperature of peak performance
E_D = Deactivation energy (eV)
(J H van't Hoff 1884, S Arrhenius 1889)

EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

- Proponents of *Metabolic Theory of Ecology* argue that we have not progressed far enough towards mechanistic modelling because metabolism has been ignored



Johnson et al 1974

$$B = B_0 \boxed{e^{-\frac{E}{kT}}} f(T, T_{pk}, E_D)$$

T = temperature (K)
k = Boltzmann constant (eV K^{-1})
E = Activation energy (eV)
T_{pk} = Temperature of peak performance
E_D = Deactivation energy (eV)
(J H van't Hoff 1884, S Arrhenius 1889)

- Surely there is more to thermal responses?
- *What about alternative models?*

MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models

Modelling, and fitting models to data: What's the big idea?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data

Modelling, and fitting models to data: What's the big idea?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data
- Whichever model "agrees best" is most likely to have the right mechanisms

MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data
- Whichever model "agrees best" is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc

MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data
- Whichever model "agrees best" is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!

Modelling, and fitting models to data: What's the big idea?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data
- Whichever model "agrees best" is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones.

Modelling, and fitting models to data: What's the big idea?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data
- Whichever model "agrees best" is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones.

MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data
- Whichever model "agrees best" is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones. *Why?*

MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models "agree well" with data
- Whichever model "agrees best" is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones. *Why? — because they have lest restrictive assumptions*

Models: How to build them?

- It's an art, takes practice
- Build models one mechanism at a time — in biology, it means start at the right level of organization!
- Always consider an alternative that is more parsimonious, even if it is phenomenological (the thermal performance curves example: Sharpe-Schoolfield, Briere, or Polynomial?)!
- For example, the Boltzmann-Arrhenius model is a good first try describe and uncover mechanisms underlying individual level rates (e.g., vector fecundity or development rate)
- The next step would be to include species interactions with temperature dependence of individuals (or go in an evolutionary direction)

FITTING MODELS TO DATA

Multiple ways to do it:

- Least Squares methods
  - Linear
  - Non-linear
- Likelihood-based methods
  - Maximum Likelihood Estimation (MLE)
  - Bayesian
- Artificial intelligence and Machine learning
  - Focus in on maximizing ability to discover pattern and predict at the cost of mechanistic insights

Methods you can use in the Miniproject

- Least squares: along with Linear Model fitting, Non-linear Least Squares (NLLS) fitting is a particularly versatile and powerful approach because many mechanisms in biology and inherently non-linear

- MLE/Bayesian methods: more robust if you are able to calculate the likelihood function — you will learn this in Term 2, so not recommended

- AI/machine Learning: most versatile for large amounts of noisy data — you will be introduced to these at the end of term 2, so definitely not recommended

SUMMARY: MODEL SELECTION IS THE KEY

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory

Summary: Model selection is the key

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory
- This is an advance over the traditional “null hypothesis” approach in Biology

SUMMARY: MODEL SELECTION IS THE KEY

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory
- This is an advance over the traditional “null hypothesis” approach in Biology
- Necessary for developing the advancement of Biology from from an observational and axiomatic discipline to one with general theories.

SUMMARY: MODEL SELECTION IS THE KEY

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory
- This is an advance over the traditional “null hypothesis” approach in Biology
- Necessary for developing the advancement of Biology from from an observational and axiomatic discipline to one with general theories.
- Necessary for understanding the mechanisms underlying biological patterns/phenomena

Comparing and Selecting Models

- It's all about the "Likelihood" of a model:
  the set of parameter values of the model (θ) given outcomes (x), equals the probability of those observed outcomes given those parameter values, that is,

$$\mathcal{L}(\theta|x) = P(x|\theta)$$

- The easiest thing to do for you is to use information theory (including AIC and BIC) to compare models.
- Both AIC and BIC use the *estimated likelihoods of a model*:
  AIC: $-2\ln[\mathcal{L}(\theta|x)] + 2p$
  Small sample AIC (AICc): $-2\ln[\mathcal{L}(\theta|x)] + 2p$
  BIC (Schwartz criterion): $-2ln[\mathcal{L}(\theta|x)] + p\ln(n)$
  (where $n =$ sample size, p number of free parameters)
- The lower the AIC or BIC, the better.

COMPARING AND SELECTING MODELS

This is how you calculate AIC and BIC (using python syntax):

Comparing and selecting models

This is how you calculate AIC and BIC (using python syntax):

- residuals = Observations - Predictions
- rss = sum(residuals ** 2)
- Then, AIC is n * log((2 * pi) / n) + n + 2 + n * log(rss) + 2 * p (*note n and p!*)
- And BIC is n + n * log(2 * pi) + n * log(rss / n) + (log(n)) * (p + 1)
- That is, $\mathcal{L}(\theta|x) = -\frac{n}{2/ln(RSS/n)}$
- For both AIC and BIC, If model **A** has AIC lower by 2-3 or more than model **B**, it's better — Differences of less than 2-3 don't really matter

COMPARING AND SELECTING MODELS

This is how you calculate AIC and BIC (using python syntax):

- residuals = Observations - Predictions
- rss = sum(residuals ** 2)
- Then, AIC is n * log((2 * pi) / n) + n + 2 + n * log(rss) + 2 * p (*note n and p!*)
- And BIC is n + n * log(2 * pi) + n * log(rss / n) + (log(n)) * (p + 1)
- That is, $\mathcal{L}(\theta|x) = -\frac{n}{2/ln(RSS/n)}$
- For both AIC and BIC, If model **A** has AIC lower by 2-3 or more than model **B**, it's better — Differences of less than 2-3 don't really matter

Also note that:

- R^2 = 1 - (rss/tss), where tss is total sum of squares:
  tss = sum((Observations - mean(Predictions)) ** 2)
  (a useful measure of goodness of fit)

COMPARING AND SELECTING MODELS: MORE STUFF

- You can also calculate Akaike Weights, which is very useful/important when comparing $>$ 2 models. These weights can then be used to perform *model averaging*.
- Model selection using the Likelihood-Ratio test (LRT) is another option when you are comparing 2 models.
- Adjusted R^2 can be used to get rigorous "idea" about how alternative models are performing.
- Very often, you will end up doing model simplification, especially in *for linear least squares model fitting* — starting with a complex model and then dropping terms till you have found a the most parsimonious version of the original model. There are functions in `R` to do this (of course!).

Readings

- Levins, R. (1966) The strategy of model building in population biology. Am. Sci. 54, 421–431.
- Johnson, J. B. & Omland, K. S. (2004) Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108.
- Bolker, B. M. et al. (2013) Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. Methods Ecol. Evol. 4, 501–512 .
- Some illustrative examples of (non-linear) model fitting to ecological/evolutionary data `https://groups.nceas.ucsb.edu/non-linear-modeling/projects`
- Additional readings in the git repository