Fitting Models to Data in Vector-Borne Disease Systems

Samraat Pawar

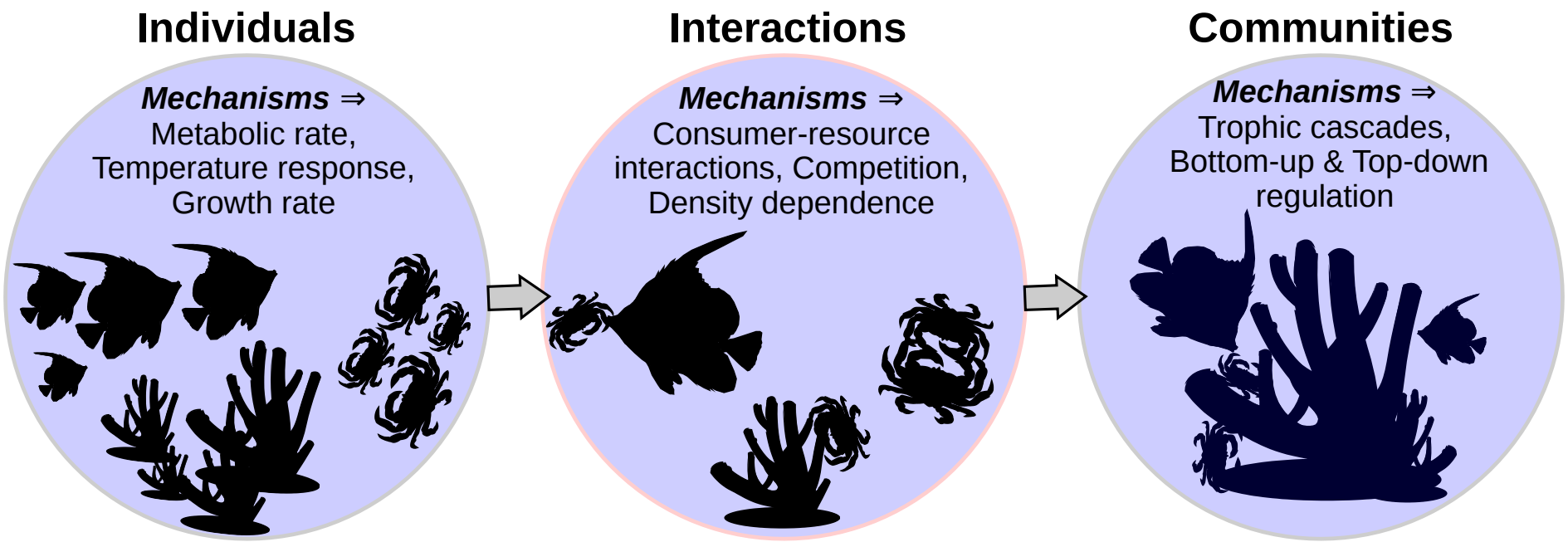
June 10, 2018

MECHANISTIC VS. PHENOMENOLOGICAL MODELS

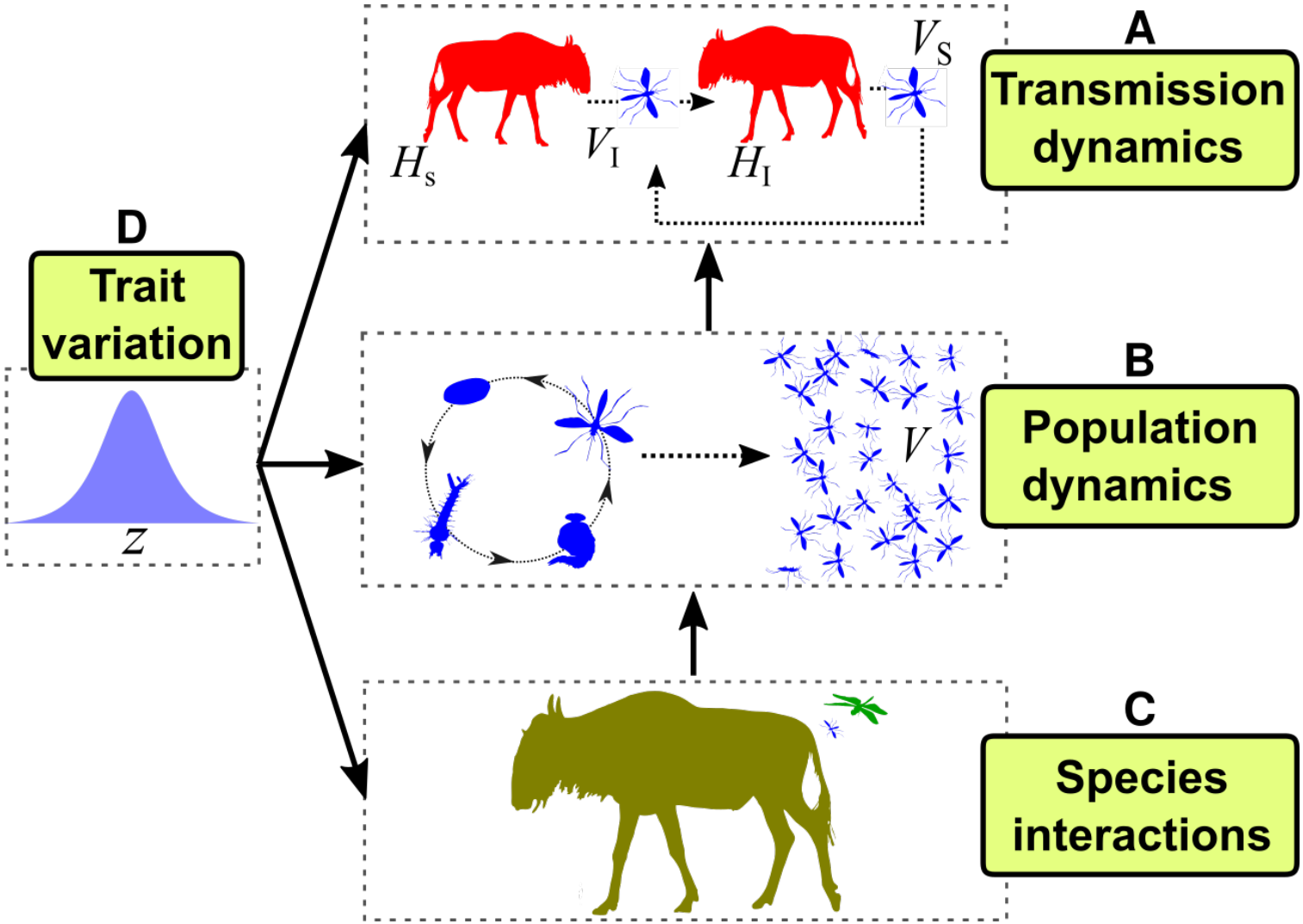
- Mechanistic models aim to explain the PROCESSES underlying observed patterns
- Empirical or phenomenological models show relationships between observed data (e.g. vector population size or disease incidence as a function of temperature or rainfall), but provide no insights into the processes/mechanisms that inter-link these data.

MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- Mechanistic models aim to explain the PROCESSES underlying observed patterns
- Empirical or phenomenological models show relationships between observed data (e.g. vector population size or disease incidence as a function of temperature or rainfall), but provide no insights into the processes/mechanisms that inter-link these data.



MECHANISMS IN VECTOR-BORNE DISEASE DYNAMICS?



MECHANISMS IN VECTOR-BORNE DISEASE DYNAMICS?

- Ecological studies often focus on explaining phenomena using somewhat phenomenological models.

MECHANISMS IN VECTOR-BORNE DISEASE DYNAMICS?

- Ecological studies often focus on explaining phenomena using somewhat phenomenological models.
- LINK

MECHANISMS IN VECTOR-BORNE DISEASE DYNAMICS?

- Ecological studies often focus on explaining phenomena using somewhat phenomenological models.
- LINK
- For example, disease invasions, outbreaks and spread: Why cycles? Why traveling waves?

MECHANISMS IN VECTOR-BORNE DISEASE DYNAMICS?

- Ecological studies often focus on explaining phenomena using somewhat phenomenological models.
- LINK
- For example, disease invasions, outbreaks and spread: Why cycles? Why traveling waves?

MECHANISMS IN VECTOR-BORNE DISEASE DYNAMICS?

- Ecological studies often focus on explaining phenomena using somewhat phenomenological models.
- LINK
- For example, disease invasions, outbreaks and spread: Why cycles? Why traveling waves? — What mechanisms operate? (Vector-parasite interaction? Vector-environment interaction?

What are Mechanisms?

- Somewhat subjective!

What are Mechanisms?

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

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

What are Mechanisms?

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

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

- What is the mechanism? — Density dependence through scramble competition (Brannstrom & Sumpter 2005)

What are Mechanisms?

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

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

- 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:

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

- 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?

What are Mechanisms?

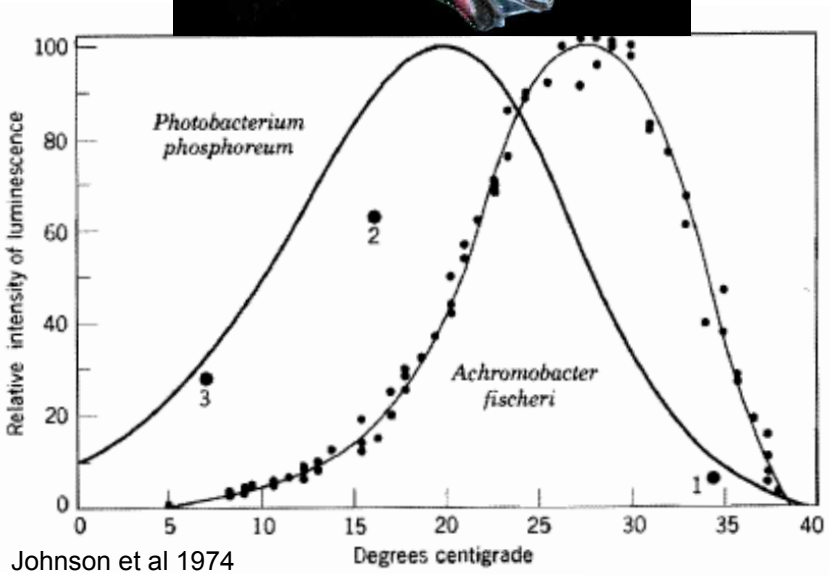
- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

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

- 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?
- Many (including yours truly!) now argue that we have not progressed far enough because the first level has been ignored!

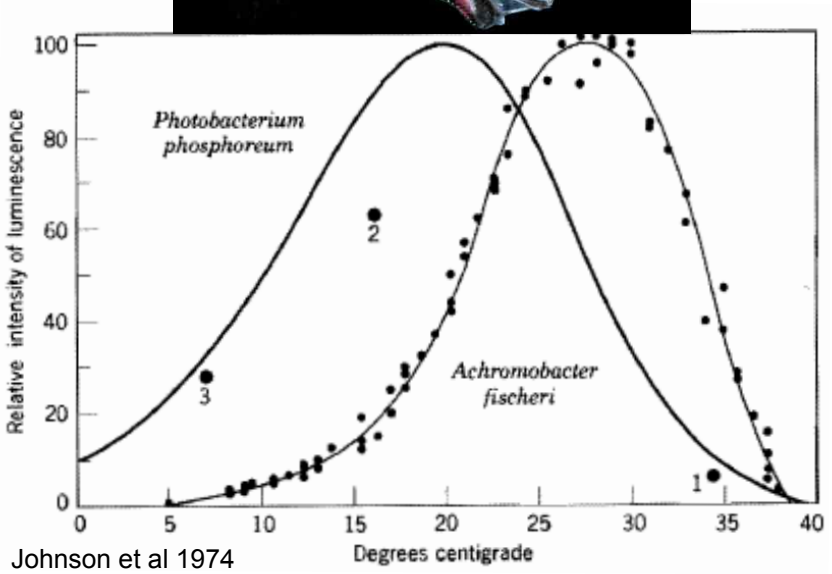
An example of a fundamental mechanism: metabolism

An example of a fundamental mechanism: metabolism



Johnson et al 1974

An example of a fundamental mechanism: metabolism

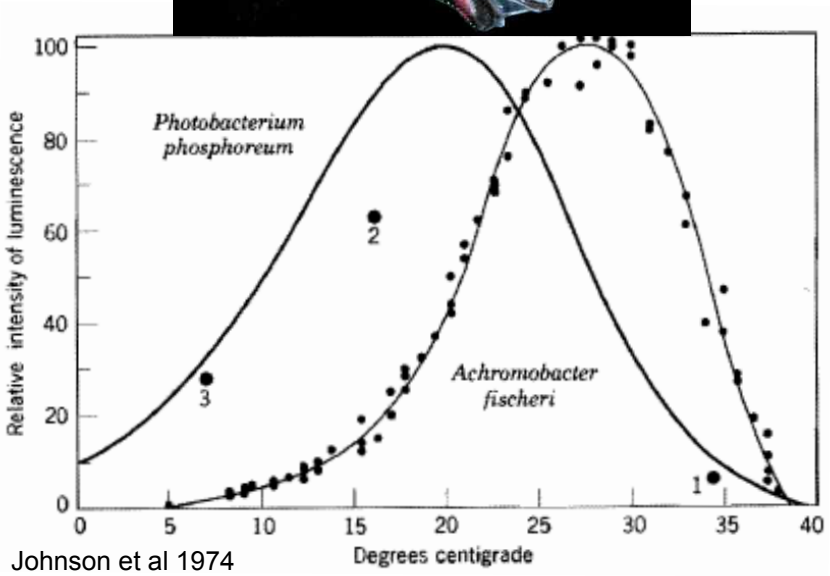


Johnson et al 1974

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

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)

An example of a fundamental mechanism: metabolism



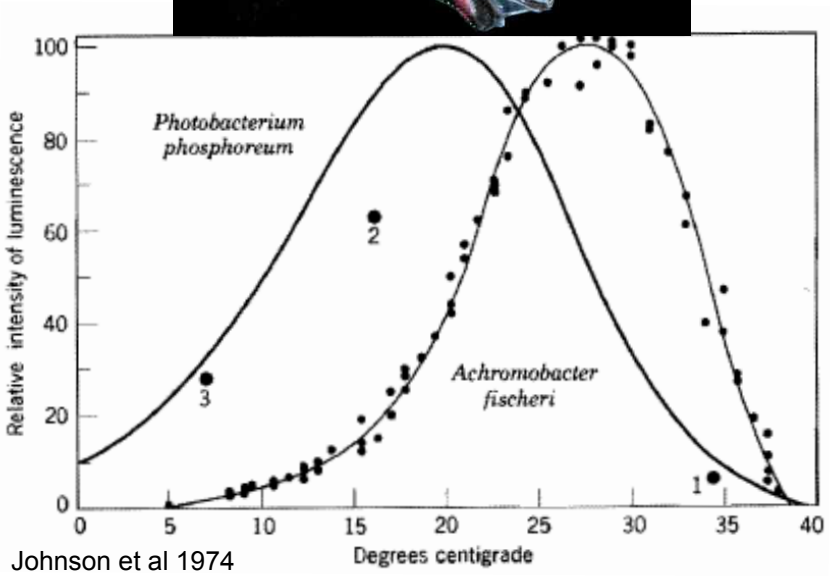
Johnson et al 1974

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

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?
  - Oxygen limitation
  - Complexity of metabolic network
  - Hormonal regulation

An example of a fundamental mechanism: metabolism



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?
  - Oxygen limitation
  - Complexity of metabolic network
  - Hormonal regulation
- *What about alternative 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
- 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 models often perform better than mechanistic ones

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)

What is model selection?

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

What is model selection?

- 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

What is model selection?

- 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.

WHAT IS MODEL SELECTION?

- 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 (read, VBD) patterns and dynamics

FITTING MODELS TO DATA

Two main ways to do it:

- One-step forecasting and machine learning (appropriate for discrete models) and time series data — focus in on maximizing ability to predict at the cost of mechanistic insights
- Ensemble fitting (appropriate for full time series or responses)
  - Least Squares methods
    - Linear
    - Non-linear
  - Likelihood based methods
    - Maximum Likelihood Estimation (MLE)
    - Bayesian

ENSEMBLE FITTING

- These include MLE, Bayesian methods, and least squares optimization or 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 are more robust if you are able to calculate the likelihood function.

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 – you should report it)

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 at the end of Miniproject Chapter of your CMEE Notes