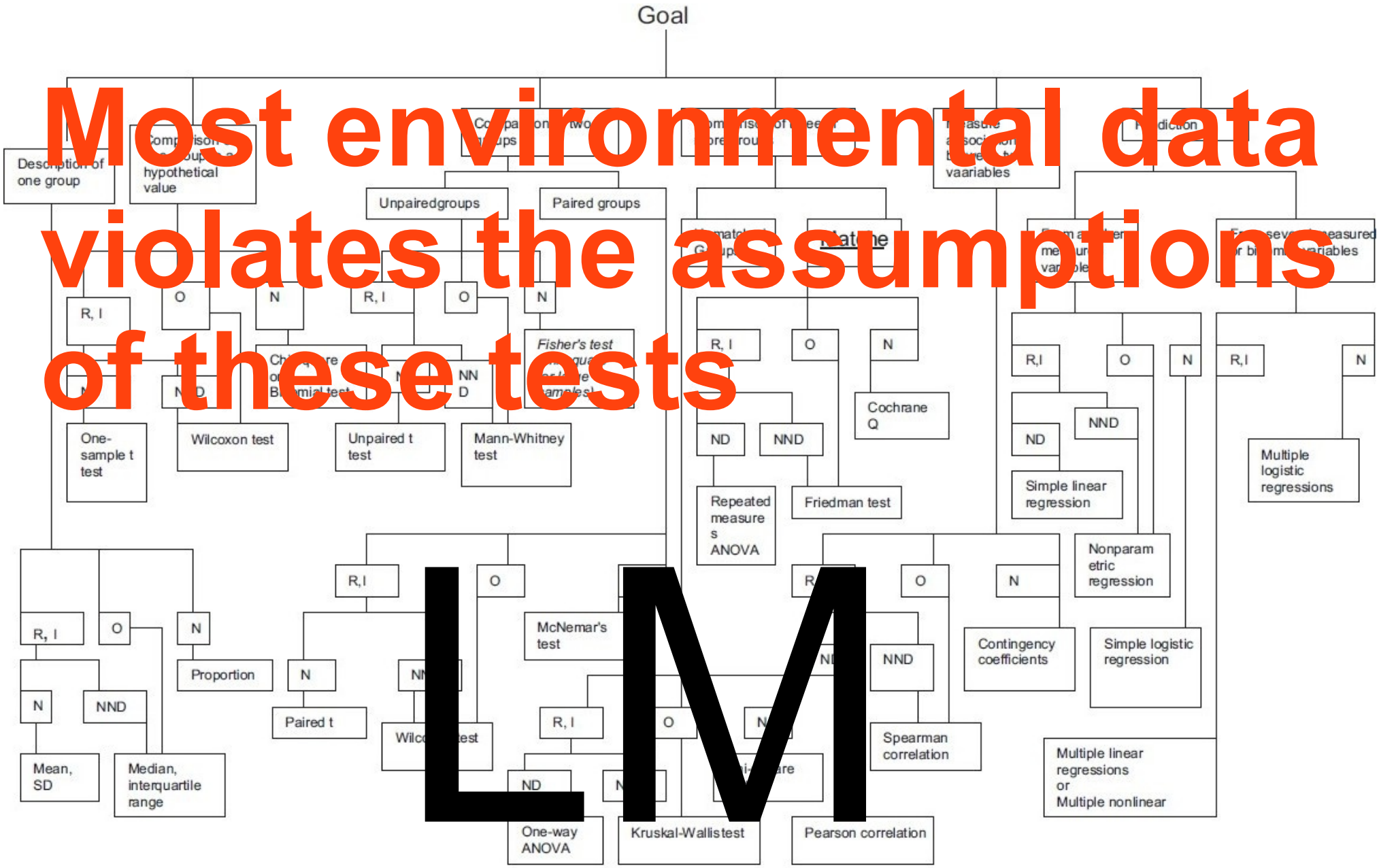
# Applied Environmental Statistics

GE 509

Instructor: Prof. Michael Dietze

# Introductions


Goal
Most environmental data violates the assumptions of these tests
LM
Description of one group
hypothetical value
Unpairedgroups
Paired groups
vaariables
R, I
O
N
One-sample t test
Wilcoxon test
Unpaired t test
Mann-Whitney test
Fisher's test
Cochrane Q
ND
NND
Repeated measures ANOVA
Friedman test
Simple linear regression
Multiple logistic regressions
Nonparametric regression
Simple logistic regression
Contingency coefficients
Proportion
Paired t
McNemar's test
Mean, SD
Median, interquartile range
One-way ANOVA
Kruskal-Wallistest
Pearson correlation
Spearman correlation
Multiple linear regressions or Multiple nonlinear
R, I = Ratio and Interval data
O= Ordinal data
N = Nominal data
N = Normal distribution
NND = Non normal distribution

# What is statistical modeling?

# What is statistical modeling?

## “Confronting models with data”

- Model fitting / parameter estimation
- Model comparison
- Estimation, partitioning, and propagation of uncertainties

# What is statistical modeling?

"Confronting models with data"

**Design the statistical analysis to fit the data rather than the data to fit the test**

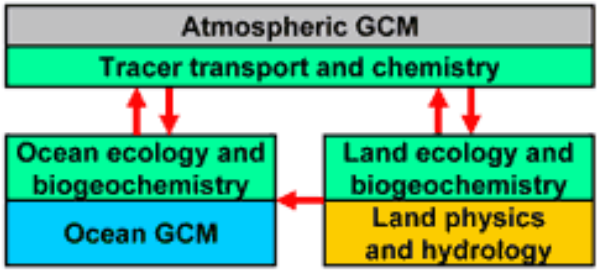
# What is a model?

# What is a model?

A conceptual, graphical, or mathematical representation / abstraction of some empirical process(es).

A mathematical function that formalizes our conceptual model / theory

$$f(x) = a \qquad f(x) =$$



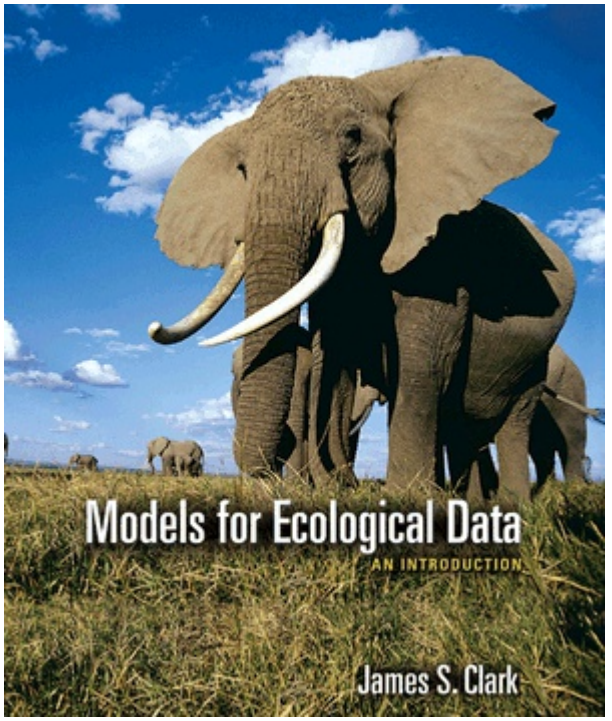
# What is a model?

Models are
HYPOTHESES

# Syllabus

# Course Materials

- Reading assignments, lecture slides, project details, etc. are all posted on the lab website http://people.bu.edu/dietze/Bayes2020/EE509.htm
- Primary Text:
  "Models for Ecological Data"
  Clark 2007 Princeton U Press
- Software:
  - R / RStudio
  - OpenBUGS / JAGS
  - Git / GitHub



# Grading

Grading will be based on lab reports, a semester-long project, and four exams.

| | | |
|---|---|---|
| Lab reports/problem sets (10 points each) | | = 150 |
| Semester project | | = 95 |
| project proposal | 2/14 | (10) |
| model description | 3/6 | (15) |
| preliminary analysis | 4/10 | (20) |
| Final report | before exam 4 | (50) |
| Exams (30, 25, 30, 30 points ) [non-cumulative] | | = 115 |
| Total | | = 360 |

# Labs

- **LAB IS MANDATORY**

- Labs will be posted in git repository https://github.com/mdietze/EE509
- Due FOLLOWING WEEK by the start of lab

- Must be turned in individually
- Can work together

# Semester Project

- Final product:
  "Journal article" on a data analysis
  - You choose topic
  - ENCOURAGED to use your own data
  - Analysis must be new, use concepts from class
  - "Methods" heavy
- Four milestones
- One lab is peer critique

# Lecture & Exams

- Four sections
  - Probability theory and Maximum Likelihood
  - Bayesian methods
  - Hierarchical/mixed models
    - Linear regression → nonlinear, non-gaussian
  - Advanced topics
    - Time series
    - Spatial

# Exams

- Multiple Choice
- Matching
- Fill in the blank
- Short Answer / Derivation
- ~15 questions

# Expectations

- You have seen basic calculus at some point
  - Primarily need to follow derivations
- Basic familiarity with statistical concepts
  - e.g. experimental design, randomization, mean, median, variance
- Open mind
- You will work hard
- You won't 'get' Bayes the first time they see it (but will need to by the 2nd exam)

# Objectives

- Literacy
  - Read and evaluate advanced stats used in papers
- Proficiency
  - underlying statistical concepts
  - Software: R, JAGS
- Exposure to advanced topics
- Paradigm shift

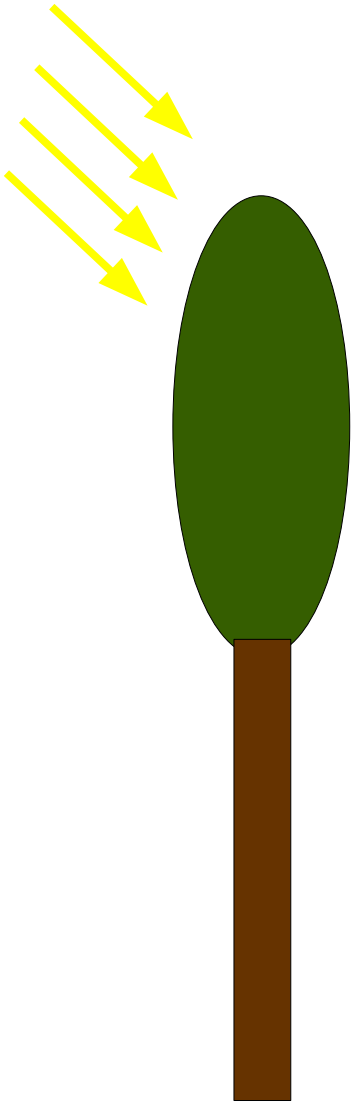
# A bit more on motivation....

Data are usually complex

Violate the assumptions of classical tests

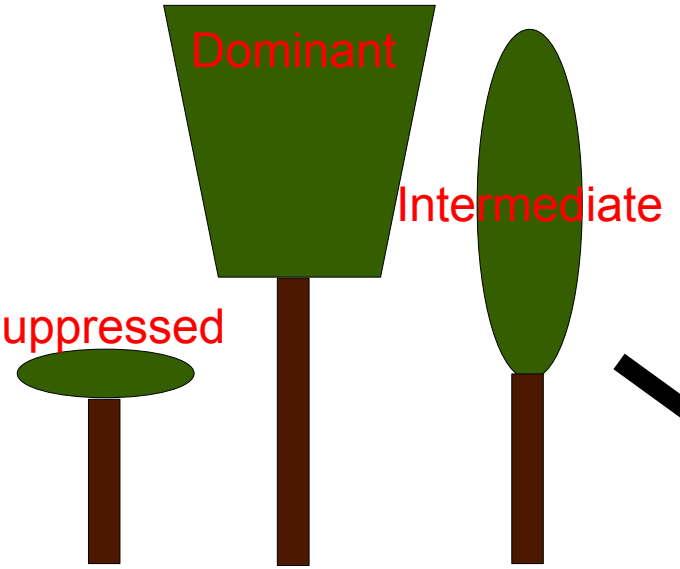
This complexity can be addressed with modern techniques

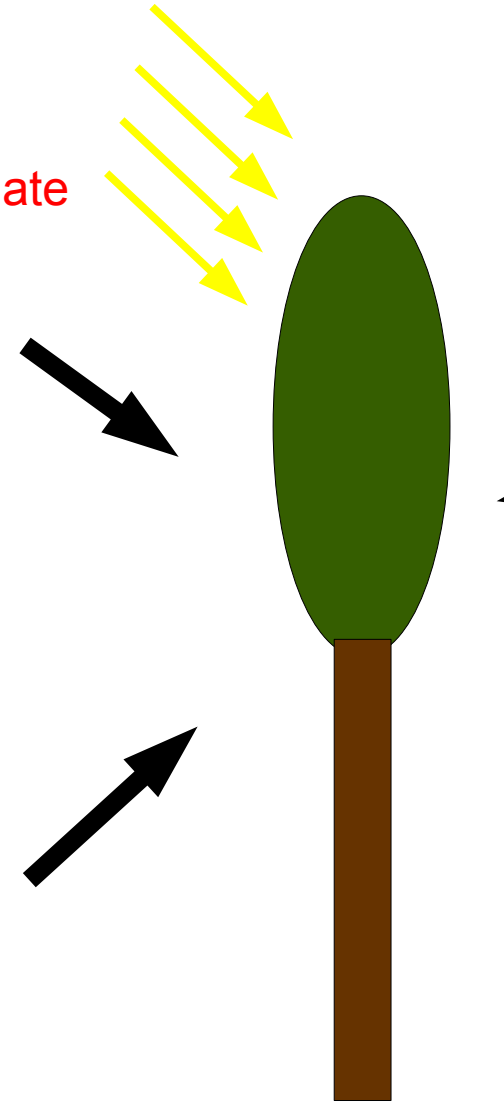
# Example:
# How much light is a tree getting?



# Example:
# How much light is a tree getting?

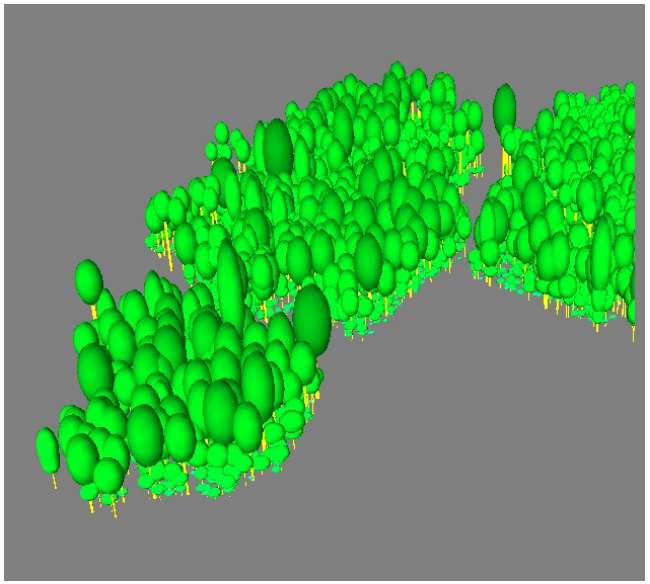
FIELD

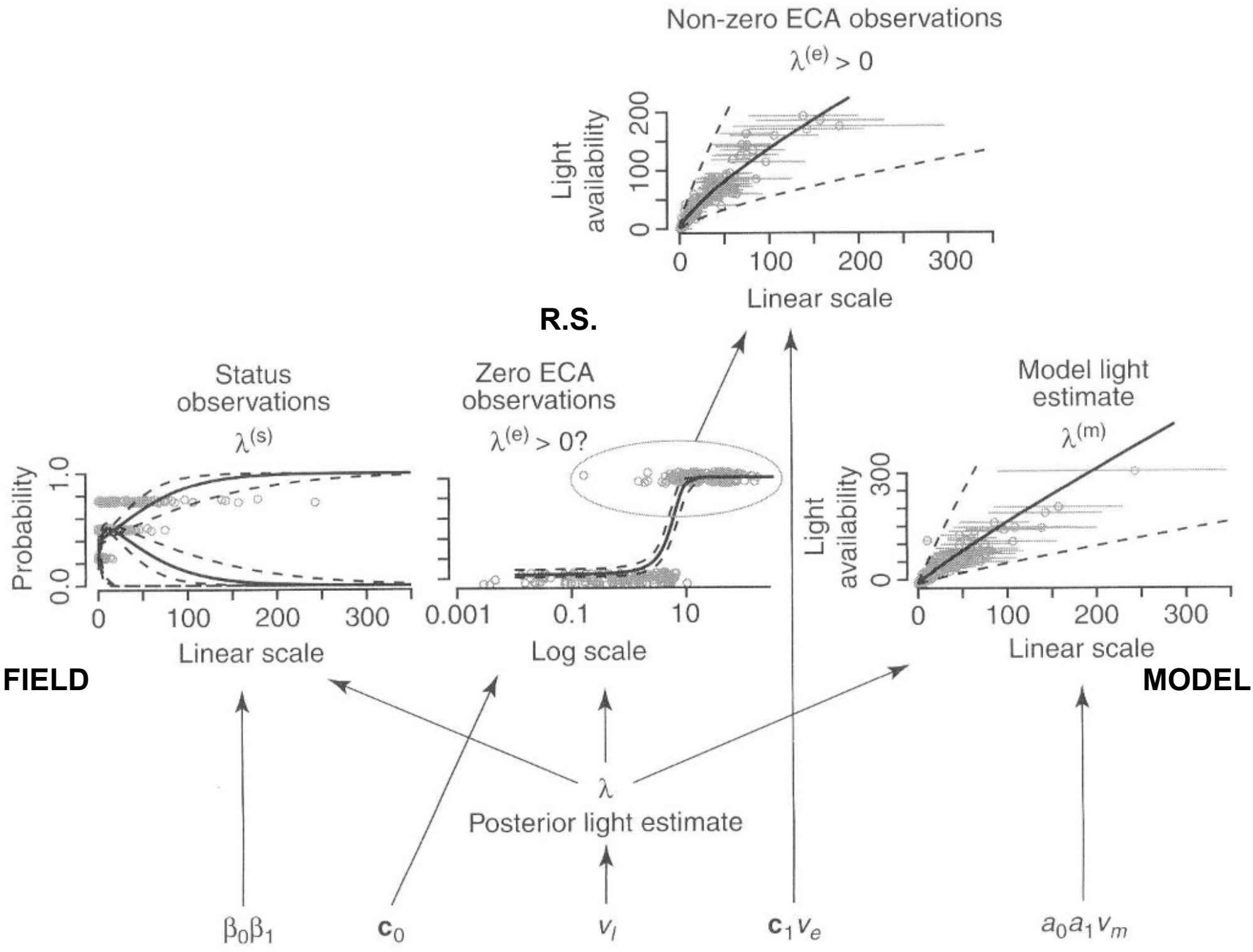


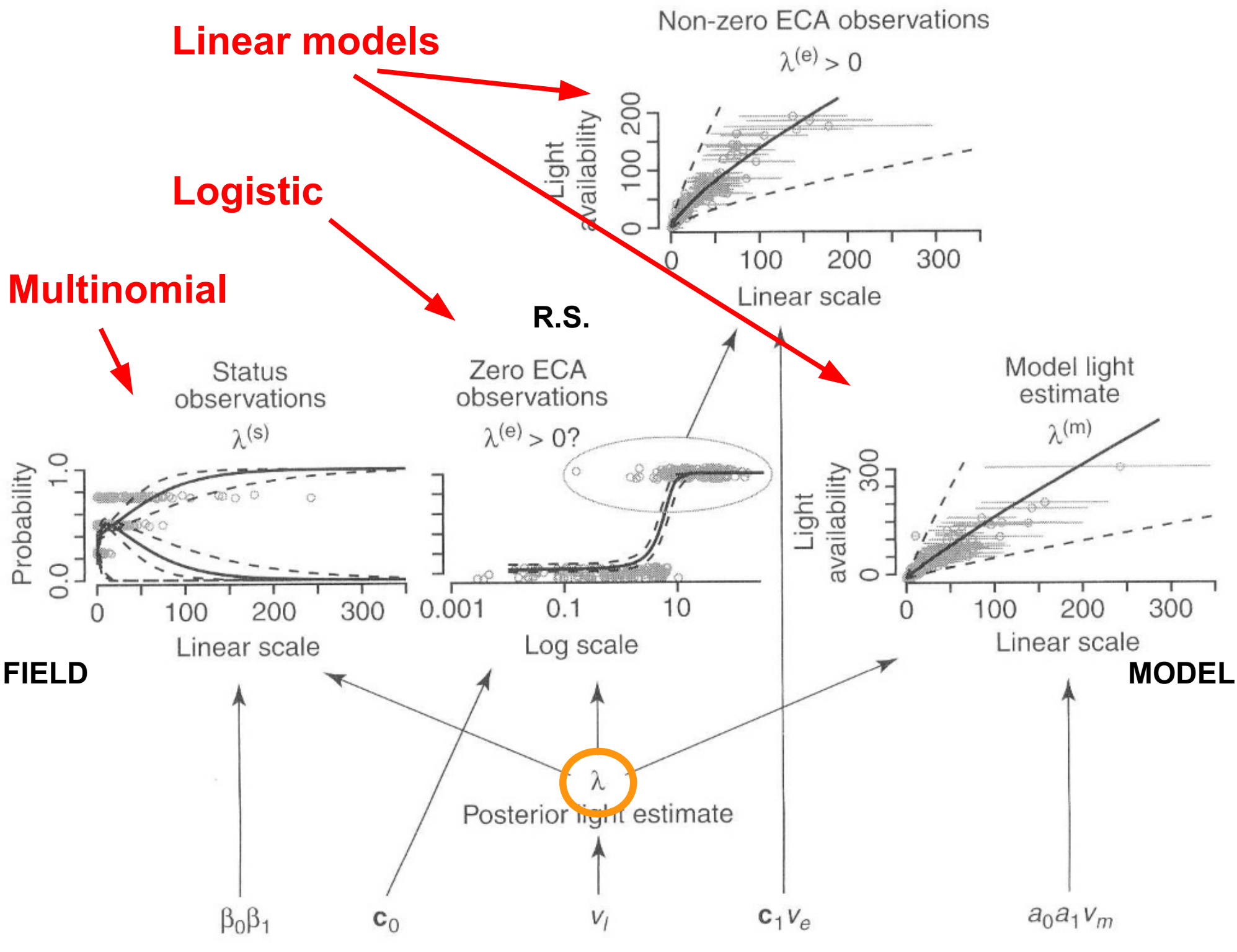



REMOTE SENSING

MODEL




Non-zero ECA observations
$\lambda^{(e)} > 0$
Light availability
0 100 200
0 100 200 300
Linear scale
R.S.
Status observations
$\lambda^{(s)}$
Probability
0.0 1.0
0 100 200 300
Linear scale
Zero ECA observations
$\lambda^{(e)} > 0?$
0.001 0.1 10
Log scale
Model light estimate
$\lambda^{(m)}$
Light availability
0 300
0 100 200 300
Linear scale
FIELD
MODEL
$\lambda$
Posterior light estimate
$\beta_0\beta_1$
$c_0$
$v_l$
$c_1 v_e$
$a_0 a_1 v_m$


Linear models
Logistic
Multinomial
R.S.
Non-zero ECA observations
$\lambda^{(e)} > 0$
Light availability
0 100 200
0 100 200 300
Linear scale
Status observations
$\lambda^{(s)}$
Probability
0.0 1.0
0 100 200 300
Linear scale
Zero ECA observations
$\lambda^{(e)} > 0?$
0.001 0.1 10
Log scale
Model light estimate
$\lambda^{(m)}$
Light availability
0 300
0 100 200 300
Linear scale
FIELD
MODEL
$\lambda$
Posterior light estimate
$\beta_0\beta_1$
$c_0$
$v_l$
$c_1 v_e$
$a_0 a_1 v_m$

# Problem Characteristics

- Multiple data constraints
- Non-linear relationships
- Non-Normal residuals
- Non-constant variance
- Latent variables (response variable not being observed directly)
- Distinction between observation error and process variability
- Missing data

# Statistical Paradigms

- Classical (e.g. sum of squares)
- Maximum Likelihood
- Bayesian

# Statistical Paradigms

| | Statistical Estimator | Method of Estimation | Output | Data Complexity | Prior Info |
|---|---|---|---|---|---|
| **Classical** | Cost Function | Analytical Solution | Point Estimate | Simple | No |
| **Maximum Likelihood** | Probability Theory | Numerical Optimization | Point Estimate | Intermediate | No |
| **Bayesian** | Probability Theory | Sampling | Probability Distribution | Complex | Yes |

# Statistical Paradigms

| | Statistical Estimator | Method of Estimation | Output | Data Complexity | Prior Info |
|---|---|---|---|---|---|
| **Classical** | Cost Function | Analytical Solution | Point Estimate | Simple | No |
| **Maximum Likelihood** | Probability Theory | Numerical Optimization | Point Estimate | Intermediate | No |
| **Bayesian** | Probability Theory | Sampling | Probability Distribution | Complex | Yes |

The unifying principal for this course is statistical estimation based on **probability**

# Next lecture

- Will cover basics of probability theory
- Read
  - Clark 2007 - Chapter 1
  - **Hilborn and Mangel p39-62** (course website)
- Optional
  - Clark 2007 – Appendix D (Probability)
  - Otto and Day – Appendix 1 (Math) and 2 (Calculus) (course website)