Meta-analysis

Combining Data

IB/NRES 509 Statistical Modeling

What is a Meta-analysis?

- A quantitative synthesis of previous research
- Studies as individual observations, weighted by n, σ^2, quality, etc.
- Can combine heterogeneous data from heterogeneous studies
- Evaluate means, responses, sources of heterogeneity

When is Meta-analysis used?

- To synthesize results once a large body of work has been conducted
- To investigate large-scale patterns beyond the scope of individual experiments
- Today we will look at two examples:
  - evaluating the hypotheses that N limits NPP
  - synthesizing trait data to parameterize a DGVM

Meta-analysis is increasing in Ecology and Evoloution

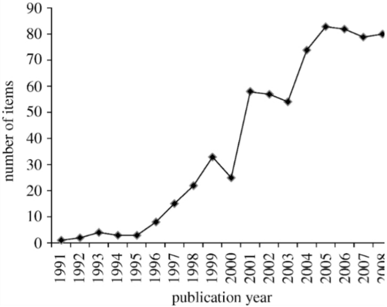


Figure: Number of meta-analyses articles published in ISI journals in Ecology and Evolution through 2008. Stewart G Biol. Lett. 2010;6:78-81

Meta-analysis Overview

- Develop a research question
- Study Literature
- Define Predictor and Response Variables
- Define Model
- Collect Data
- Calculate Response Metric and Weights
- Assess the effect of unpublished data
- Interpret, present at conference, publish...

Literature Search

- Search methodology:
  - search terms / phrases
  - follow references
  - contact researchers
- Study Selection
  - apriori criteria for inclusion:
    - study design
    - years of publication
    - language
    - sample size
    - availability of relevant information
  - block results by quality, conduct analyses in groups
  - quality based weights

Chosing Predictor Variables

- Depends on Question, available data
- Experimental vs. Observational Data
- Continuous vs. Categorical Data

Chosing Response Variables

1. Mean
$$\bar{X}$$
2. Differences in means
$$d = \bar{X}_{trt} - \bar{X}_{Ctl}$$
3. Cohen's g
$$g = \frac{\bar{X}_{trt} - \bar{X}_{Ctl}}{s_p}$$
where $s_p$is the pooled sample standard deviation, small n can bias estimate of $\sigma$and thus$g$
4. Response Ratio proportionate change
$$lnR = log(\bar{X}_T / \bar{X}_C)$$
can be used in competition intensity, relative yield, relative crowding

Fixed Effects Model

- Fixed Effects Model:

$$T_i \sim N(\Theta_i, \sigma_i^2)$$

$$\Theta_i = \mu$$

  - where Θ_i is the study mean and σ_i^2 is the within study variance
  - assumes no between-study variability; all studies have same μ
    - can be used if RE model shows negligable among-study variance ($\tau^2 \rightarrow 0$)
    - otherwise, inference limited to studies in meta-analysis and studies with identical characteristics
  - Difficult to justify in ecological context.

Random Effects Model

- Random Effects Model:

$$T_i \sim N(\Theta_i, \sigma_i^2)$$

$$\Theta_i \sim N(\mu, \tau^2)$$

  - whereΘ_i is the study mean, μ is the grand mean,$\sigma_i^2$is the within study variance and$\tau^2$is the among-study variance
    - acknowledges study-specific effects due to locatin, experimental conditions, etc.
    - models both within and among study variance
  - As a regression model:

$$\mu = \mu_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_n X_{ni}$$

    - X's can be continuous (linear regression), categorical (e.g. ANOVA) or a combination (mixed effects)

Study Weights

- Give greater weight to experiments with smaller standard error:
  - fixed effects model

$$w_i = 1/(\hat{\sigma}_i{}^2)$$

  - random effects model

$$w_i = 1/(\hat{\sigma}_i{}^2 + \hat{\tau}^2)$$

- Weighted Mean Effect

$$\bar{T} = \frac{\sum w_i T_i}{\sum w_i}$$

Sources of Bias

- Publication Bias
  - studies that show no effect are less likely to be published
  - “file drawer problem”
  - Create a funnel plot with effect size versus sample size
- Unconducted Studies
  - researchers are less likely to ask a question if they think that the answer is known
  - e.g. few N fertilization studies have been conducted in the tropics
- Study Heterogeneity
  - need to ensure that effect sizes are standardized and comparable across studies
  - e.g. a large variety of methods may be used, sites and organisms studied may vary

Assessing Study Bias

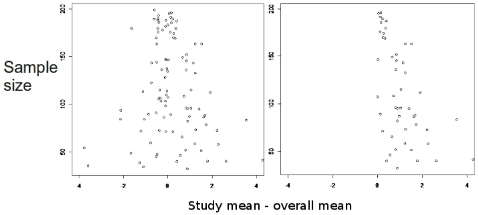


Figure: Use of a funnel plot to evaluate the file-drawer problem; x axis is deviance from mean effect size ($T_i - \hat{\mu} = \varepsilon; \varepsilon \sim N(0, \tau^2)$) as sample size increases, the variance around the mean effect size (here set to 0 by subtracting) in effect size should decrease. left: no evidence for missing studies, right: missing studies

Approaches to Inference in Meta-analysis

- Moment Matching
- Maximum Likelihood
- Bayesian

Moment Matching

- Moment Matching
  - given data, calculate moments ($\hat{\mu} = T_i = n\sum x_i$, $\hat{\sigma}^2 = S_i^2$)
  - estimate global group means
  - partition among vs. within study variance
  - adequate for simple models, but not, e.g. multiple regression

Analysis of N-fertilization Studies

- Question: Does Nitrogen limit Terrestrial Net Primary Productivity?
  - Widely accepted to be true, except for tropical forests
  - Many previous studies, no global synthesis
- Approach
  - Meta-analysis of N-fertilization studies
  - Regression of N on lnR

Analysis of N-fertilization Studies

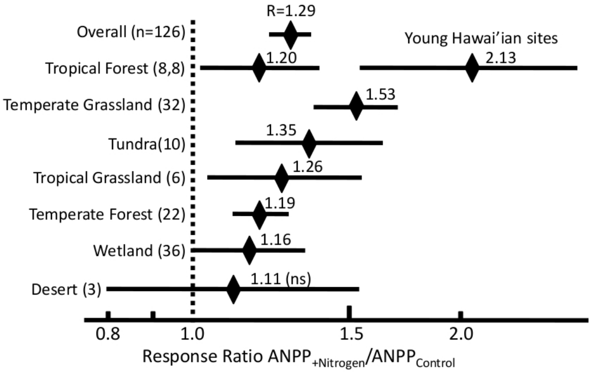
- Response Metric

$$lnR = \frac{ANPP_{+N}}{ANPP_C}$$

- Measures of ANPP
  - Grasslands & Wetlands: Peak Biomass, Repeated Harvests
  - Forests: ΔVolume, ΔDiameter, ΔBasal Area, Litterfall

Analysis of N-fertilization Studies

Biome Level Responses to N



Maximum Likelihood

- Maximum Likelihood
  - choose parameters that maximize p(data|model)
  - allows more complex models, non-normal distributions, parameter uncertainty
  - simultaneous estimation of all parameters, more inference options

Maximum Likelihood Example

- Given data, what is the pdf of different parameter values?
- equivalent to MM for simple models, but more parameters can be estimated, e.g. each Θ_i
- Random Effects Model:

$$T_i \sim N(\Theta_i, \sigma_i^2)$$

$$\Theta_i \sim N(\mu, \tau^2)$$

Maximum Likelihood Example

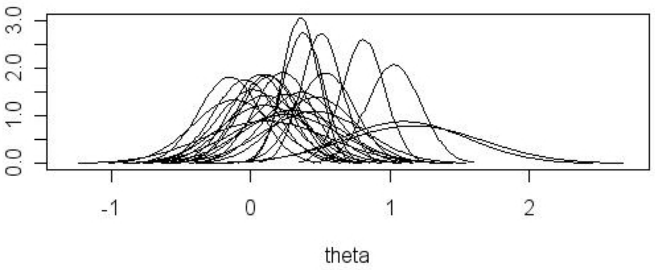


Figure: Likelihood of observing the data (T_i and $S_i^2(\hat{\sigma^2})$) from each of 25 studies, given different values of Θ

Bayesian Meta-analysis

- posterior is the Likelihood of observing parameters given the data
- like MLE, can estimate all parameters simultaneously
- Much more flexible

What Values Might A Plant Trait Hold?

A Bayesian Approach to DGVM Model Parameterization

- Overview
  - Plant traits are used to parameterize models of vegetation, ie: ED2
  - Std. Practice - select single values from the literature
  - Our goal - parameterize model with meta-analyisis posterior distributions
- Response Metric:
  - Trait Means of 24 (and counting) traits

Plant Trait Meta-analysis Model

$$T_i \sim N(\Theta_i, \sigma_i^2)$$

$$\Theta_i \sim N(\mu, \tau^2)$$

Informed Priors

$$\mu \sim distn(a, b)$$

$$s_i^2 \sim IG(\frac{n_i}{2}, \frac{n_i}{2}\sigma_i^2)$$

Priors

$$\tau^2 \sim IG(0.1, 0.1)$$

$$\sigma^2 \sim IG(0.1, 0.1)$$

Within vs. Among Study Variance

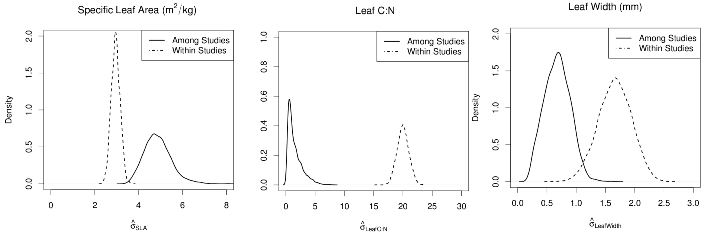


Figure: Distribution of among (τ^2)and within (σ^2)study variance of three plant traits.

Distribution of Priors and Posterior Means

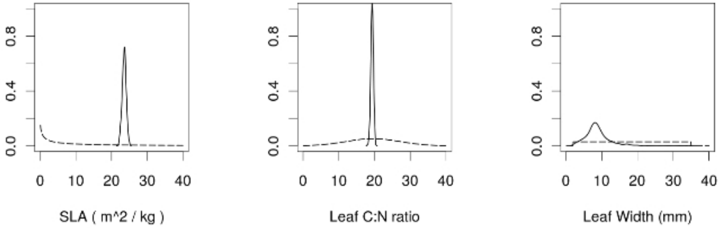


Figure: Distribution of prior (dashed line) and posterior (solid line) means of three plant traits.

Example 3: Synthesis of Plant Traits

Model Forecast With Uncertainty

A Bayesian Approach to Model Parameterization

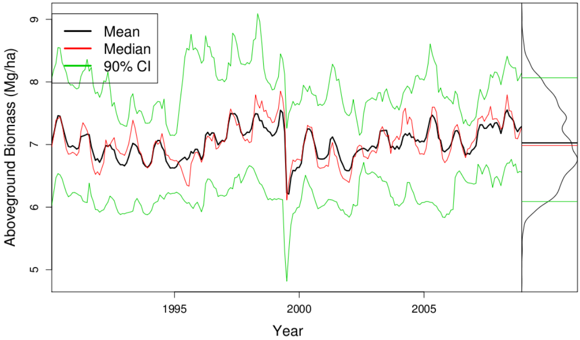


Figure: Model prediction with uncertainty estimate.