Multimodel inference (model selection and model averaging)

SEEC Toolbox Seminar

Res Altwegg

27/05/2021

Overview

- Ranking a set of model
- Multiple working hypotheses
- Model averaging

SEEC - Statistics in Ecology, Environment and Conservation

http://www.seec.uct.ac.za/

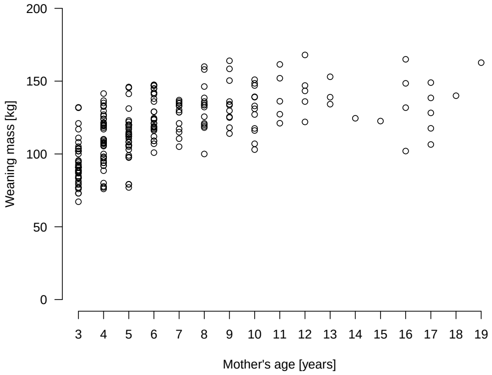
@SEEC_UCT

Good statistical models

Should fit the structure in the data but not the noise.

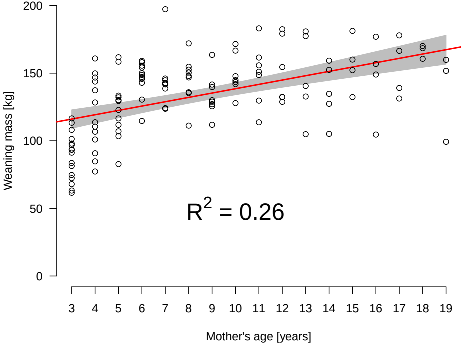
- Underfitting: failure to fit structure in the data $\rightarrow$ prediction bias
- Overfitting: fits to noise $\rightarrow$ loss of precision

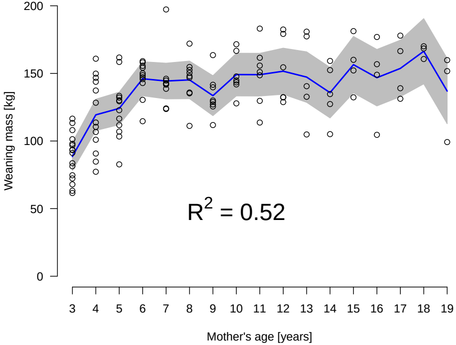
Elephant seals on Marion Island



Thanks to Chris Oosthuizen and the Marion Island Marine Mammal Program, University of Pretoria (www.marionseals.com)

Elephant seals

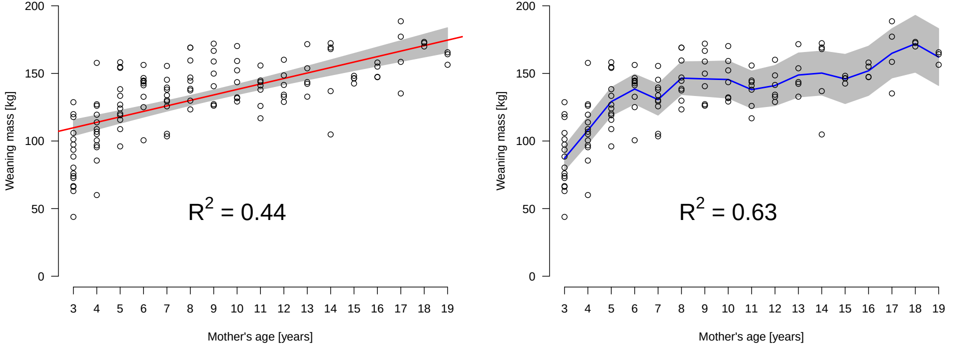




$$Y_j = \beta_0 + \beta_1 \times X_j + \epsilon_j$$

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

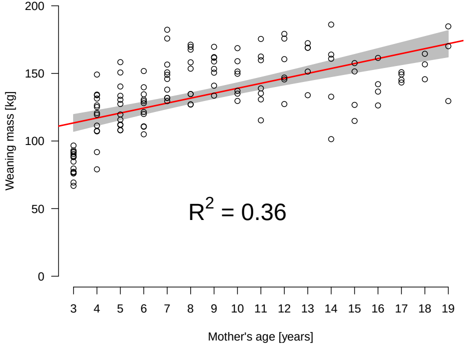
Elephant seals

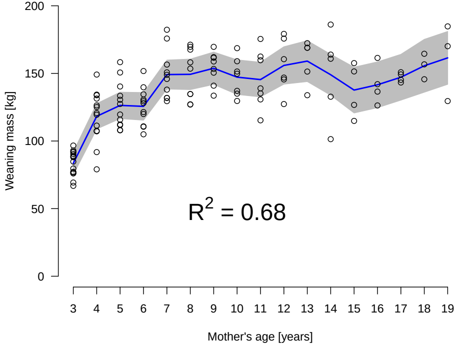


$$Y_j = \beta_0 + \beta_1 \times X_j + \epsilon_j$$

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

Elephant seals

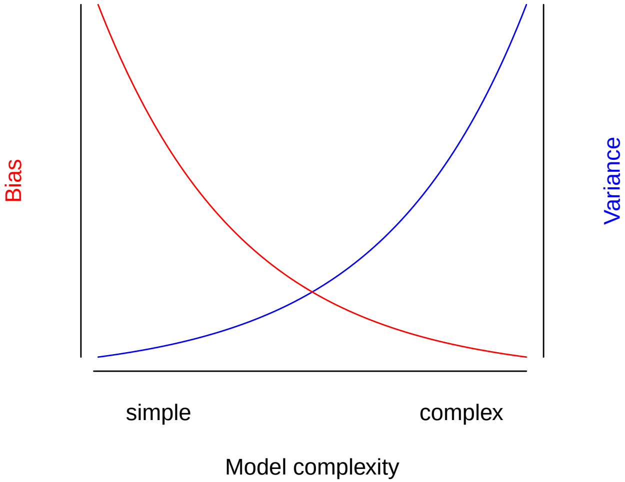




$$Y_j = \beta_0 + \beta_1 \times X_j + \epsilon_j$$

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

Bias-variance trade-off

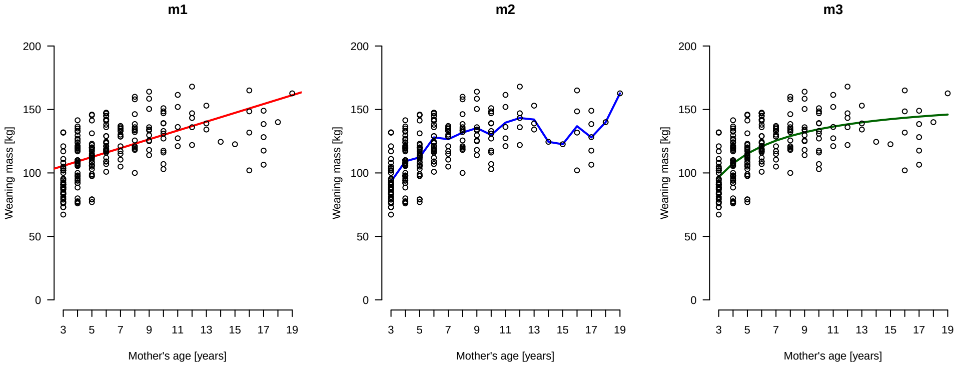


Akaike's Information Criterion

$$AIC = -2\log(\mathcal{L}(\hat{\theta}|\text{Data})) + 2K$$

- Best balance between bias and variance
- Smaller value is better

Model selection analysis

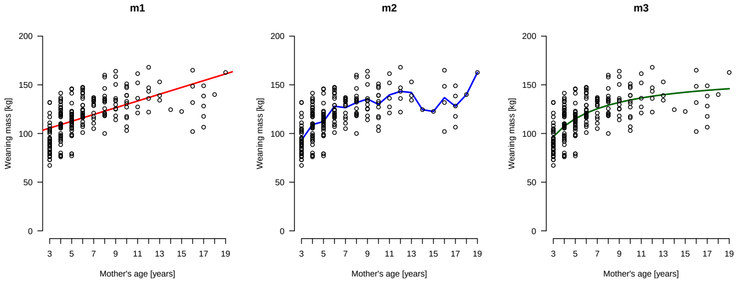


$$Y_i = \beta_0 + \beta_1 \times X_i + \epsilon_i$$

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

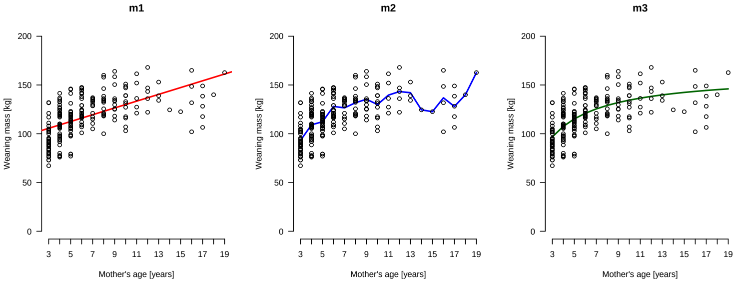
$$Y_i = \frac{\beta_0 \times X_i}{1 + \beta_0 \times \beta_1 \times X_i} + \epsilon_i$$

Model selection analysis



	-2 × loglik	K	AIC
m1	1940.50	3	1946.50
m2	1868.14	18	1904.14
m3	1891.49	3	1897.49

Model selection analysis



	-2 × loglik	K	AIC	Δ AIC
m1	1940.50	3	1946.50	49.00
m2	1868.14	18	1904.14	6.65
m3	1891.49	3	1897.49	0

Akaike weights

The weight of evidence in favor of model i being the best in the set:

$$w_i = \frac{exp(-\frac{1}{2}\Delta_i)}{\sum_{r=1}^{R} exp(-\frac{1}{2}\Delta_r)}$$

	-2 $\times$ loglik	K	AIC	Δ AIC	w
m1	1940.50	3	1946.50	49.00	0.00
m2	1868.14	18	1904.14	6.65	0.03
m3	1891.49	3	1897.49	0	0.97

Model m3 had 97% of the support relative to the other models.

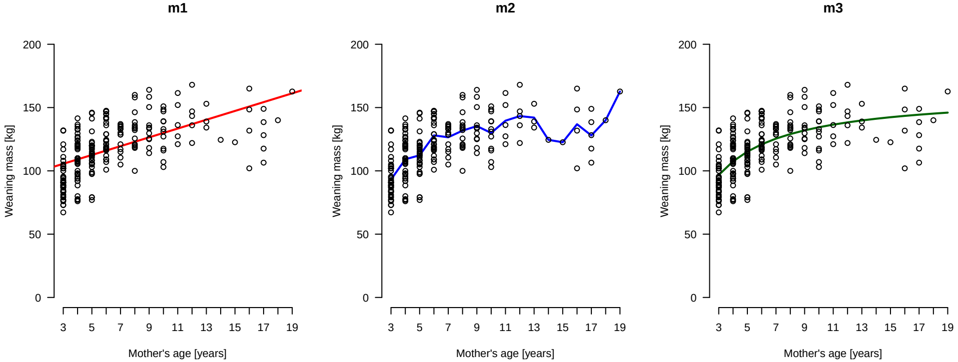
Evidence ratios

$$\text{Evidence ratio} = \frac{w_i}{w_j}$$

	-2 × loglik	K	AIC	Δ AIC	w
m1	1940.50	3	1946.50	49.00	0.00
m2	1868.14	18	1904.14	6.65	0.03
m3	1891.49	3	1897.49	0	0.97

$$\frac{w_3}{w_2} = \frac{0.97}{0.03} = 27.8$$

Model m3 was 28 times more likely than m2 to be the best in the set.



$$Y_i = \beta_0 + \beta_1 \times X_i + \epsilon_i$$

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

$$Y_i = \frac{\beta_0 \times X_i}{1 + \beta_0 \times \beta_1 \times X_i} + \epsilon_i$$

Model selection analysis

Two different scientific goals need fundamentally different approaches.

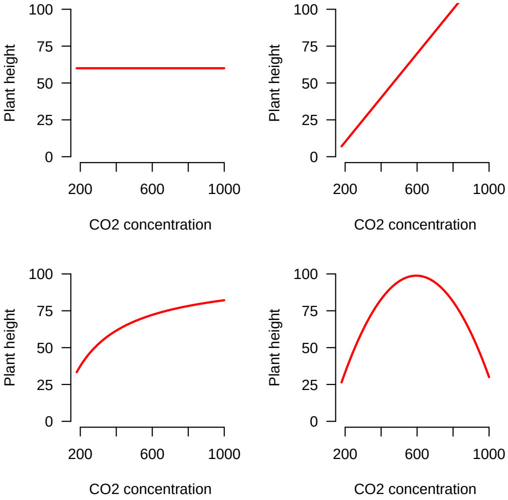
- Hypothesis-based research
- Data mining / hypothesis generation

Statistical Modelling / Scientific Research: The steps

1. come up with a set of biological hypotheses
2. translate the hypotheses into statistical models
3. data collection: field, experiment, observations
4. fit the models to these data
5. evaluate the relative support each model (hypothesis) gets from the data
6. answer the biological question

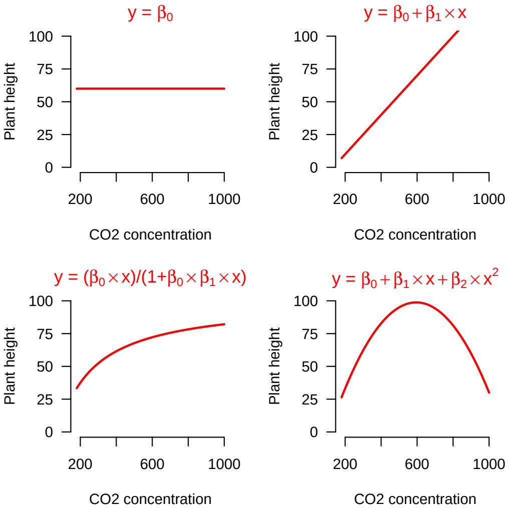
Step 1: Formulate biological hypotheses

How does $[CO_2]$ affect growth of *Acacia karroo*?



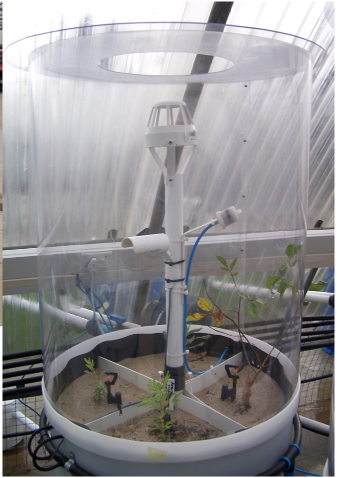
"Acacia karroo, bloeityd, Roodeplaat NR" by JMK - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons

Step 2: Translate hypotheses into models



"Acacia karroo, bloeityd, Roodeplaat NR" by JMK - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons

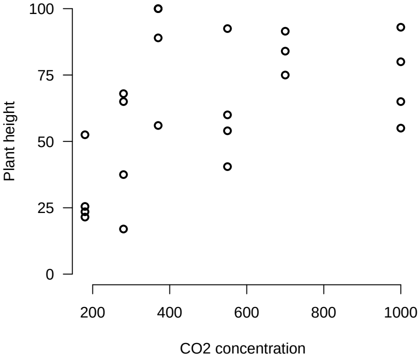
Step 3: Data collection



Kgope et al 2009. Austral Ecology 35:451–463.

Fotos: Guy F Midgley

Step 3: Data collection

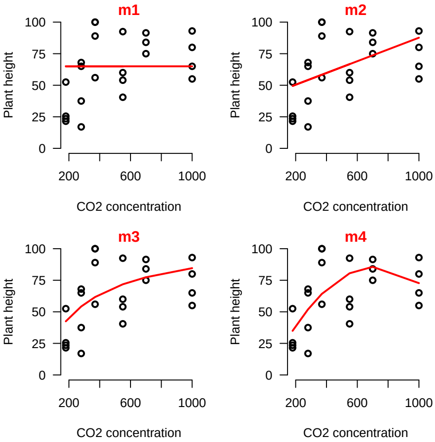


Kgope et al 2009. Austral Ecology 35:451–463.

Step 4: Fit models to data

```
m1<-lm(Total.stem.length~1)
m2<-lm(Total.stem.length~CO2)
m3<-nls(Total.stem.length~(b0*CO2)/(1+b0*b1*CO2),
        start=list(b0=1,b1=0.01))
m4<-lm(Total.stem.length~CO2+I(CO2^2))
```

Step 4: Fit models to data



Step 5: Evaluate relative support

```
aics <- AIC(m1,m2,m3,m4)
delta.aics <- aics$AIC - min(aics$AIC)
wi <- exp(-0.5*delta.aics)/sum(exp(-0.5*delta.aics))

logliks <- c(logLik(m1),logLik(m2),logLik(m3),logLik(m4))

models <- c("m1","m2","m3","m4")
mstable <- data.frame(models, -2*logliks, aics$df,
                      aics$AIC, delta.aics,wi)
```

Step 5: Evaluate relative support

	-2 $\times$ loglik	K	AIC	Δ AIC	w
m1	226.56	2	230.56	8.58	0.01
m2	220.44	3	226.45	4.46	0.06
m3	216.60	3	222.60	0.61	0.40
m4	213.99	4	221.99	0.00	0.54

Step 6: Answer biological question

	-2 × loglik	K	AIC	Δ AIC	*w*
m1	226.56	2	230.56	8.58	0.01
m2	220.44	3	226.45	4.46	0.06
m3	216.60	3	222.60	0.61	0.40
m4	213.99	4	221.99	0.00	0.54

- Model 4 was best supported by the data suggesting that plant height increased with increasing [CO_2] up to a maximum above which it started to decline.
- However, model 3 was nearly as well supported as model 4 (evidence ratio $= \frac{0.54}{0.40} = 1.35$).
- Model 1 was poorly supported showing that it is highly unlikely that [CO_2] had no effect on plant height.

Careful when interpreting P values after model selection

- Don't mix model selection and P values

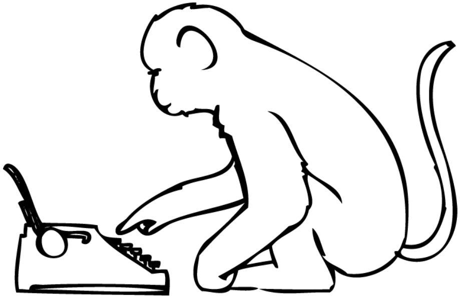
```
##
## Call:
## lm(formula = Total.stem.length ~ CO2 + I(CO2^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40.040 -13.688  -4.692  16.462  35.776
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.198e+00  2.022e+01  -0.208  0.83754
## CO2          2.483e-01  8.096e-02   3.068  0.00584 **
## I(CO2^2)    -1.714e-04  6.733e-05  -2.546  0.01881 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Why not fit "all possible" models and see which one comes out best?

- leads to fitting of lots of models

e.g. with 10 covariates there are $2^{10} = 1024$ possible regression models (ignoring interactions and polynomial effects)

- overfitting guaranteed



Wikimedia Commons

Why not use step-wise model selection?

- leads to fitting of lots of models
- overfitting guaranteed
- spurious results guaranteed
- different procedures don't lead to the same result
- misleading if explanatory variables are correlated

$\rightarrow$ don't do it!

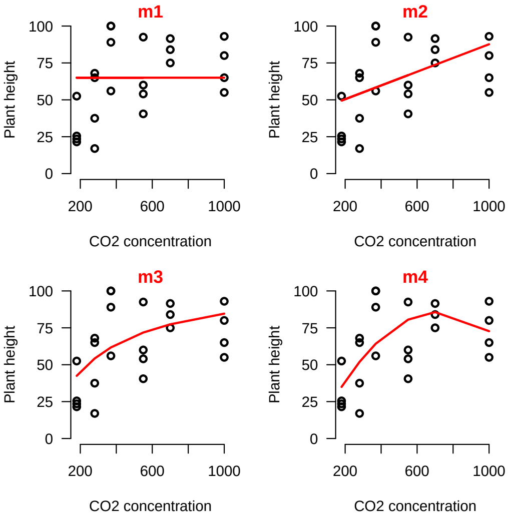
Types of uncertainty

	-2 × loglik	K	AIC	Δ AIC	*w*
m1	226.56	2	230.56	8.58	0.01
m2	220.44	3	226.45	4.46	0.06
m3	216.60	3	222.60	0.61	0.40
m4	213.99	4	221.99	0.00	0.54

- Structural uncertainty (which model is correct) → Akaike weights
- Uncertainty conditional on model structure → standard errors

"Acacia karroo, bloeityd, Roodeplaat NR" by JMK - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons

Model-averaged predictions



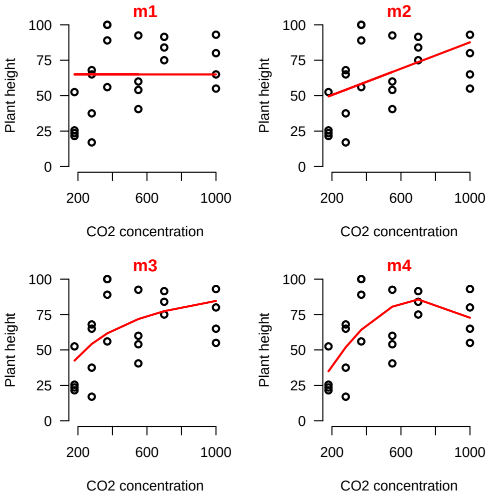
Average?

$$\hat{\bar{\theta}} = \frac{\sum_{i=1}^{R} \hat{\theta}_i}{R} = \sum_{i=1}^{R} \frac{1}{R} \hat{\theta}_i$$

Weighted average using Akaike weights:

$$\hat{\bar{\theta}} = \sum_{i=1}^{R} w_i \hat{\theta}_i$$

Model-averaged predictions



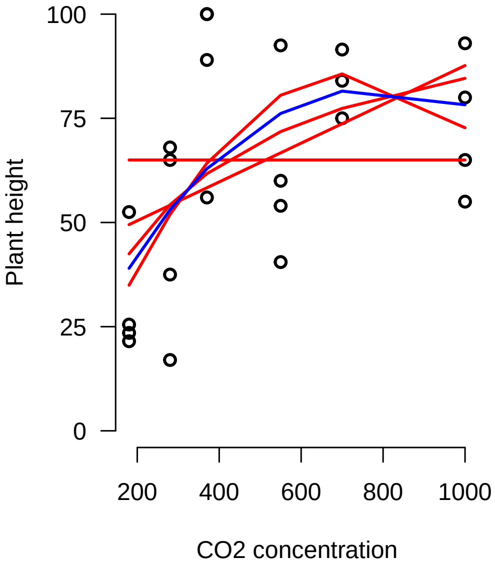

```r
cbind(pred200, wi, wi * pred200)
```

```
##      pred200      wi
## [1,]    65.0 0.00739  0.481
## [2,]    50.4 0.05798  2.923
## [3,]    45.2 0.39641 17.925
## [4,]    38.6 0.53822 20.783
```

```r
sum(wi * pred200)
```

```
## [1] 42.1
```

Model-averaged predictions



Predictions from individual models
Model-averaged predictions

Unconditional standard error

Averaged point estimate:

$$\hat{\hat{\theta}} = \sum_{i=1}^{R} w_i \hat{\theta}_i$$

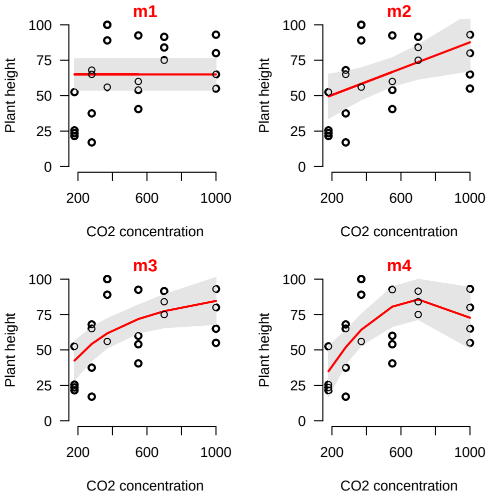
Unconditional measure of uncertainty:

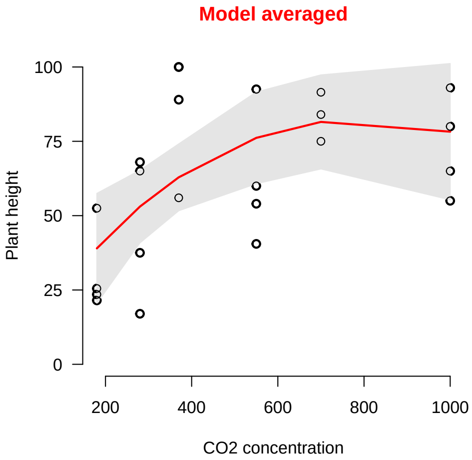
$$\hat{se}(\hat{\hat{\theta}}) = \sum_{i=1}^{R} w_i [\hat{var}(\hat{\theta}|g_i) + (\hat{\theta}_i - \hat{\hat{\theta}})^2]^{\frac{1}{2}}$$

$\hat{var}(\hat{\theta}|g_i)$: variance of the model-specific estimate

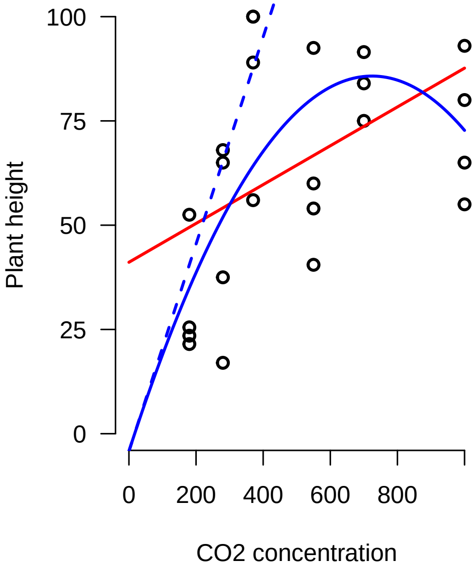
$(\hat{\theta}_i - \hat{\hat{\theta}})^2$: variance among model-specific point estimates

Unconditional measures of uncertainty





Averaging parameters? – Careful!



Predictions from model m2

$$y = \beta_0 + \beta_1 \times x$$

Predictions from model m4

$$y = \beta_0 + \beta_1 \times x + \beta_2 \times x^2$$

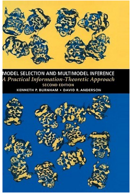
β_0 and β_0 do not have the same meaning

β_1 and β_1 do not have the same meaning

Literature

Standard textbook on multi-model inference:

- *Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd edition. Springer, New York.*



Model averaging:

- *Cade, B. S. 2015. Model averaging and muddled multimodal inferences. Ecology 96:2370–2382.* **Why you shouldn't average parameters.**
- *Dormann, C. F., et al 2018. Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. Ecological Monographs 88:485–504.*