Model selection

Why model selection?

- *Nested models*: how much complexity is necessary to fit the data?

Why model selection?

- *Nested models*: how much complexity is necessary to fit the data?
- *Non-nested models*: compare fit of different models (e.g. alternative hypotheses)

Why model selection?

- *Nested models*: how much complexity is necessary to fit the data?
- *Non-nested models*: compare fit of different models (e.g. alternative hypotheses)
  - But building larger model might be better than choosing any of them!

Overfitting and balanced model complexity

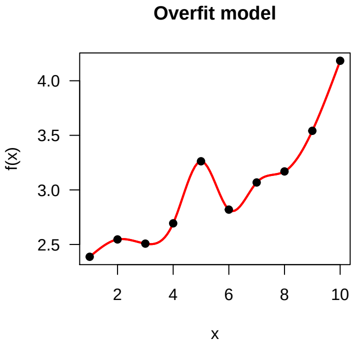


Figure 1: Overfitted model

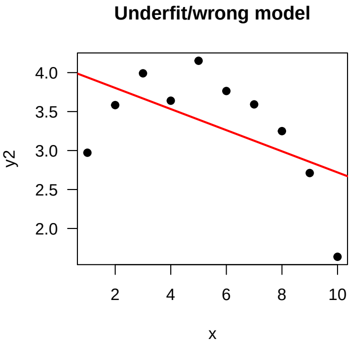


Figure 2: Wrong model

Overfitting: an example with niche modelling

Wenger & Olden (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods Ecol Evol*.

GLMM

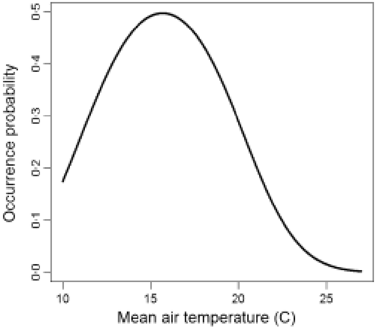


Figure 3:

Random forests (overfit)

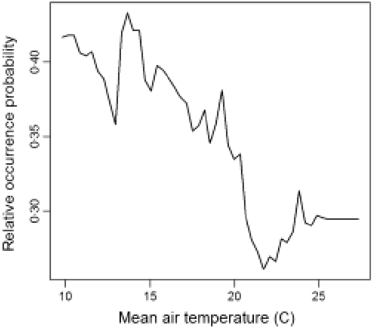


Figure 4:

So, two important aspects of model selection

- On one hand, we want to maximise fit.

So, two important aspects of model selection

- On one hand, we want to maximise fit.
- On the other hand, we want to avoid overfitting and overly complex models.

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out...)

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out. . .)
- Alternatives:

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out. . .)
- Alternatives:
  - AIC

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out...)
- Alternatives:
  - AIC
  - BIC

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out...)
- Alternatives:
  - AIC
  - BIC
  - DIC

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out. . .)
- Alternatives:
  - AIC
  - BIC
  - DIC
  - WAIC. . .

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out…)
- Alternatives:
  - AIC
  - BIC
  - DIC
  - WAIC…
- All these attempt an impossible task:

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out...)
- Alternatives:
  - AIC
  - BIC
  - DIC
  - WAIC...
- All these attempt an impossible task:
  - estimating out-of-sample prediction error without external data or further model fits!

Evaluating models' predictive accuracy

- Cross-validation (k fold, leave one out. . .)
- Alternatives:
  - AIC
  - BIC
  - DIC
  - WAIC. . .
- All these attempt an impossible task:
  - estimating out-of-sample prediction error without external data or further model fits!
- All these methods have flaws!

AIC

$$\mathrm{AIC} = -2 \log p(y|\hat{\theta}_{\mathrm{mle}}) + 2k$$

Figure 5:

- First term: model fit (deviance, log likelihood)

AIC

$$\mathrm{AIC} = -2\log p(y|\hat{\theta}_{\mathrm{mle}}) + 2k$$

Figure 5:

- First term: model fit (deviance, log likelihood)
- k: number of estimated parameters (penalisation for model complexity)

AIC

$$\mathrm{AIC} = -2 \log p(y|\hat{\theta}_{\mathrm{mle}}) + 2k$$

Figure 5:

- First term: model fit (deviance, log likelihood)
- k: number of estimated parameters (penalisation for model complexity)
- AIC biased towards complex models.

AIC

$$\mathrm{AIC} = -2\log p(y|\hat{\theta}_{\mathrm{mle}}) + 2k$$

Figure 5:

- First term: model fit (deviance, log likelihood)
- k: number of estimated parameters (penalisation for model complexity)
- AIC biased towards complex models.
- AICc recommended with 'small' sample sizes (n/p $<$ 40). But see Richards 2005 Ecology.

AIC

$$\mathrm{AIC} = -2 \log p(y|\hat{\theta}_{\mathrm{mle}}) + 2k$$

Figure 5:

- First term: model fit (deviance, log likelihood)
- k: number of estimated parameters (penalisation for model complexity)
- AIC biased towards complex models.
- AICc recommended with 'small' sample sizes (n/p $<$ 40). But see Richards 2005 Ecology.
- Doesn't work with hierarchical models or informative priors!

Problems of IC

- No information criteria is panacea: all have problems.

Problems of IC

- No information criteria is panacea: all have problems.
- They give average out-of-sample prediction error, but prediction errors can differ substantially within the same dataset (e.g. populations, species).

Problems of IC

- No information criteria is panacea: all have problems.
- They give average out-of-sample prediction error, but prediction errors can differ substantially within the same dataset (e.g. populations, species).
- Sometimes better models rank poorly (Gelman et al. 2013). So, combine with thorough model checks.

So which variables should enter my model?

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)
  - pairs() or similar

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)
  - pairs() or similar
  - If $|r| > 0.5$ - 0.7, consider leaving one variable out, but keep it in mind when interpreting model results.

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)
  - pairs() or similar
  - If $|r| > 0.5 - 0.7$, consider leaving one variable out, but keep it in mind when interpreting model results.
  - Or combine 2 or more in a synthetic variable (e.g. water deficit $\sim$ Temp + Precip).

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)
  - pairs() or similar
  - If $|r| > 0.5 - 0.7$, consider leaving one variable out, but keep it in mind when interpreting model results.
  - Or combine 2 or more in a synthetic variable (e.g. water deficit $\sim$ Temp + Precip).
  - Many methods available, e.g. sequential, ridge regression... (see Dormann et al)

Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)
  - pairs() or similar
  - If $|r| > 0.5 - 0.7$, consider leaving one variable out, but keep it in mind when interpreting model results.
  - Or combine 2 or more in a synthetic variable (e.g. water deficit $\sim$ Temp $+$ Precip).
  - Many methods available, e.g. sequential, ridge regression... (see Dormann et al)
  - Measurement error can seriously complicate things (Biggs et al 2009; Freckleton 2011)

Choosing predictors

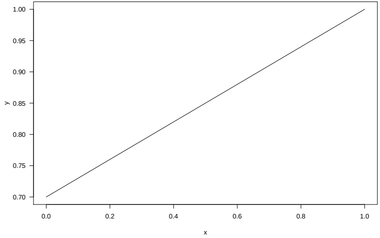
- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)
  - pairs() or similar
  - If $|r| > 0.5$ - 0.7, consider leaving one variable out, but keep it in mind when interpreting model results.
  - Or combine 2 or more in a synthetic variable (e.g. water deficit $\sim$ Temp $+$ Precip).
  - Many methods available, e.g. sequential, ridge regression... (see Dormann et al)
  - Measurement error can seriously complicate things (Biggs et al 2009; Freckleton 2011)
- For predictors with large effects, consider interactions.

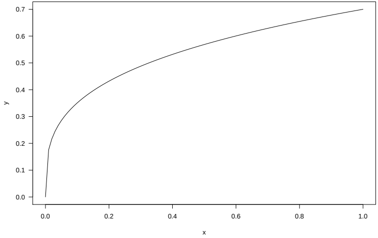
Choosing predictors

- Choose variables based on **ecological understanding**, rather than throwing plenty of them in a fishing expedition.
- Propose single global model or small set ($<$ 10 - 20) of **reasonable** candidate models.
- Number of variables balanced with sample size (at least 10 - 30 obs per param)
- Assess collinearity between predictors (Dormann et al 2013)
  - pairs() or similar
  - If $|r| > 0.5$ - 0.7, consider leaving one variable out, but keep it in mind when interpreting model results.
  - Or combine 2 or more in a synthetic variable (e.g. water deficit $\sim$ Temp $+$ Precip).
  - Many methods available, e.g. sequential, ridge regression... (see Dormann et al)
  - Measurement error can seriously complicate things (Biggs et al 2009; Freckleton 2011)
- For predictors with large effects, consider interactions.
- See also Zuur et al 2010.

Think about the shape of relationships

$y \sim x + z$
Really? Not everything has to be linear! Actually, it often is not.
Think about shape of relationship. See chapter 3 in Bolker's book.





Removing predictors

Do not use stepwise regression

- Whittingham et al. (2006) Why do we still use stepwise modelling in ecology and behaviour? J. Animal Ecology.

Do not use stepwise regression

- Whittingham et al. (2006) Why do we still use stepwise modelling in ecology and behaviour? J. Animal Ecology.
- Mundry & Nunn (2009) Stepwise Model Fitting and Statistical Inference: Turning Noise into Signal Pollution. Am Nat.

Do not use stepwise regression

- Whittingham et al. (2006) Why do we still use stepwise modelling in ecology and behaviour? J. Animal Ecology.
- Mundry & Nunn (2009) Stepwise Model Fitting and Statistical Inference: Turning Noise into Signal Pollution. Am Nat.
- This includes stepAIC (e.g. Dahlgren 2010; Burnham et al 2011; Hegyi & Garamszegi 2011).

Gelman's criteria for removing predictors

(assuming only potentially relevant predictors have been selected a priori)

- NOT significant + expected sign = let it be.

Gelman's criteria for removing predictors

(assuming only potentially relevant predictors have been selected a priori)

- NOT significant + expected sign = let it be.
- NOT significant + NOT expected sign = remove it.

Gelman's criteria for removing predictors

(assuming only potentially relevant predictors have been selected a priori)

- NOT significant + expected sign = let it be.
- NOT significant + NOT expected sign = remove it.
- Significant + NOT expected sign = check… confounding variables?

Gelman's criteria for removing predictors

(assuming only potentially relevant predictors have been selected a priori)

- NOT significant + expected sign = let it be.
- NOT significant + NOT expected sign = remove it.
- Significant + NOT expected sign = check... confounding variables?
- Significant + expected sign = keep it!

The modelling process

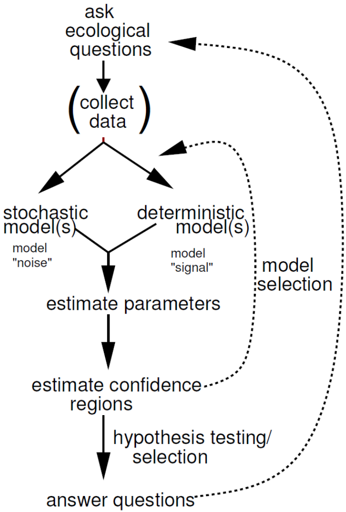


Figure 1.5 Flow of the modeling process.

Figure 6:

Summary

1. Choose meaningful variables

Summary

1. Choose meaningful variables
   - Beware collinearity

Summary

1. Choose meaningful variables
    - Beware collinearity
    - Keep good n/p ratio

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets
   - Don't assume linear effects: think about appropriate functional relationships

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets
   - Don't assume linear effects: think about appropriate functional relationships
   - Consider interactions for strong main effects

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets
   - Don't assume linear effects: think about appropriate functional relationships
   - Consider interactions for strong main effects
3. If > 1 model have similar support, consider model averaging (or blending).

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets
   - Don't assume linear effects: think about appropriate functional relationships
   - Consider interactions for strong main effects
3. If > 1 model have similar support, consider model averaging (or blending).
4. Always check thoroughly fitted models

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets
   - Don't assume linear effects: think about appropriate functional relationships
   - Consider interactions for strong main effects
3. If > 1 model have similar support, consider model averaging (or blending).
4. Always check thoroughly fitted models
   - Residuals, goodness of fit...

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets
   - Don't assume linear effects: think about appropriate functional relationships
   - Consider interactions for strong main effects
3. If > 1 model have similar support, consider model averaging (or blending).
4. Always check thoroughly fitted models
   - Residuals, goodness of fit...
   - Plot. Check models. Plot. Check assumptions. Plot. (Lavine 2014).

Summary

1. Choose meaningful variables
   - Beware collinearity
   - Keep good n/p ratio
2. Generate global model or (small) set of candidate models
   - Avoid stepwise and all-subsets
   - Don't assume linear effects: think about appropriate functional relationships
   - Consider interactions for strong main effects
3. If > 1 model have similar support, consider model averaging (or blending).
4. Always check thoroughly fitted models
   - Residuals, goodness of fit...
   - Plot. Check models. Plot. Check assumptions. Plot. (Lavine 2014).
5. Always report effect sizes