Model evaluation and model comparision

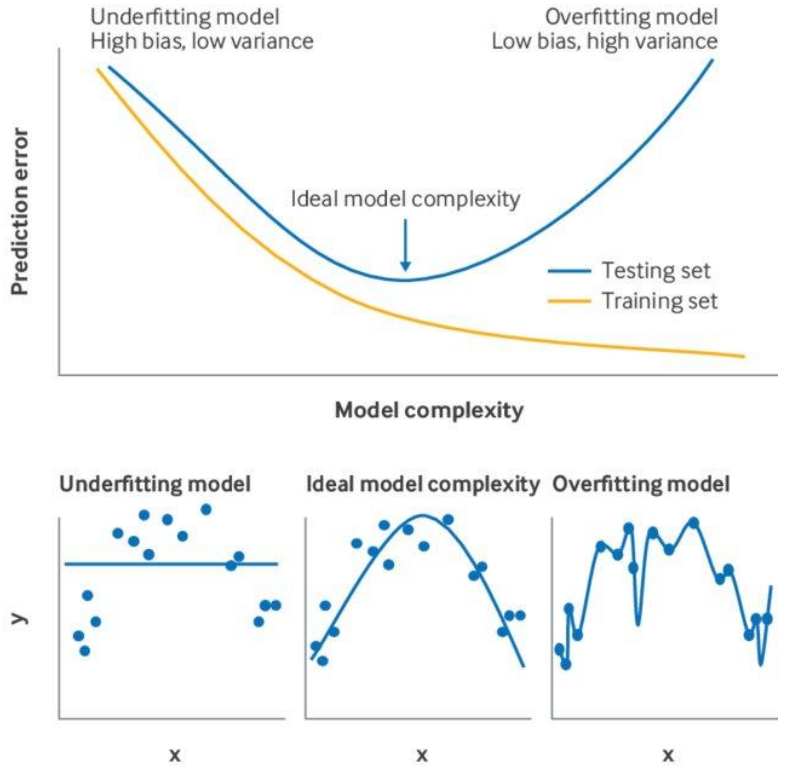
Joana Amaral Paulo

Information included in this presentation is available in Fenix!

According to **Vanclay and Skovsgaard (1997),** a summary on the key aspects on model evaluation include:

1. **Examine the model and its components in terms of logic structure and from theoretical and biological views** to see if they are:

- Parsimonious (more complex is not always better);



According to **Vanclay and Skovsgaard (1997)**, a summary on the key aspects on model evaluation include:

1. **Examine the model and its components in terms of logic structure and from theoretical and biological views** to see if they are:

- Parsimonious (more complex is not always better);
- biologically realistic (sigmoidal curve, allometric relationship…);
- consistent with existing theories of forest growth;
- predict sensible responses to management actions (ex: thinning => N and G), stand growth characteristics (site index…), etc…

2. **Characterize errors in terms of:**

- bias and precision;
- model efficiency;

model evaluation

Selection of data for model evaluation

- **Independent data set** – was not used during model fitting and is used solely to assess how well the model generalizes to unseen data (model evaluation). Note: this data set must include dependent and independent variables available!
  - Not Used in fitting: It must be separate from the fitting data to prevent information leakage.
  - Representative of “real data”: It should be similar in distribution to the actual data the model will encounter in deployment.
  - Sufficiently Large: It should have enough samples to provide a reliable estimate of model performance.

Selection of data for model evaluation

- **Cross validation -** splitting or partition of data into subsets - one for fitting and another for evaluation.
  - Trees? Plots?
  - How many? 10% of the data? X%? N plots?
  - The size of the dataset for cross-validation depends on several factors, including the total dataset size, model complexity, and computational resources.

CV Method	Description	When to Use
Leave-One-Out Cross-Validation (LOOCV or press)	Each observation is left out once, and the model is trained on the rest.	When data is **very small** but computational cost is high.
k-Fold Cross-Validation	Splits data into **k** equal parts, trains on **k-1** folds, tests on the remaining fold. Repeats **k** times.	Standard choice when data is large enough.

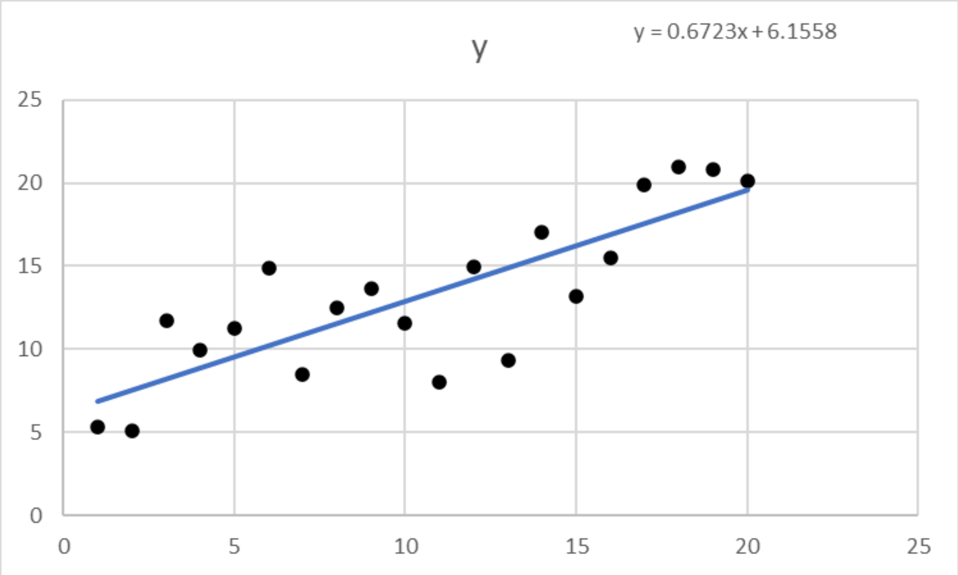
Measuring bias

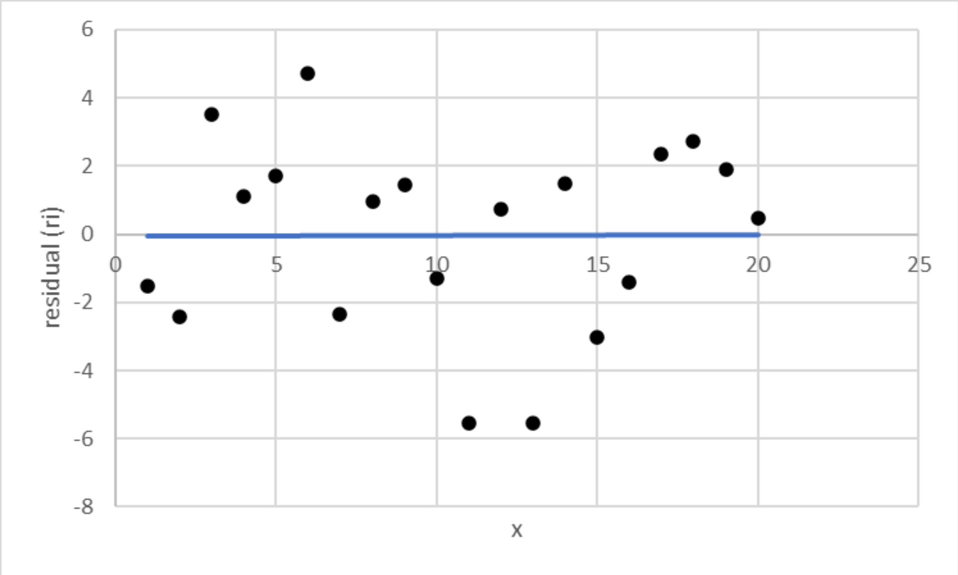
- **Compute the residual i (r_i) of each observation from the evaluation data set (r_i = $yobs_i$ - $yest_i$ = $yobs_i$ - $\hat{y}_i$)**
- **Bias:** refers to a systematic error in a model estimate, that causes it to deviate from the true value or underlying pattern.
  - It may be positive or negative.
  - As close to zero the better.
  - Evaluated by r_i values distribution (percentiles, mean…)

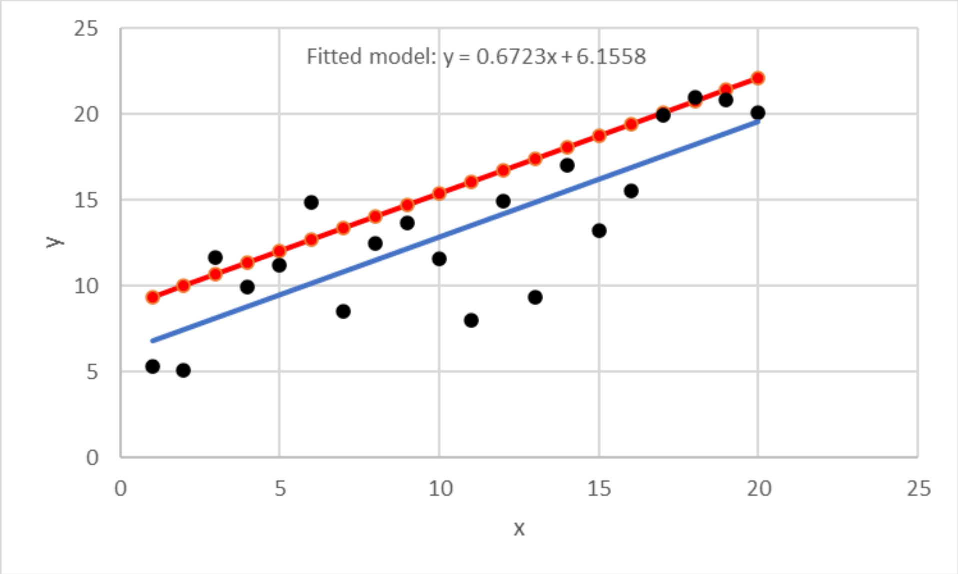
Mean value of the residuals (M_r) to evaluate bias:

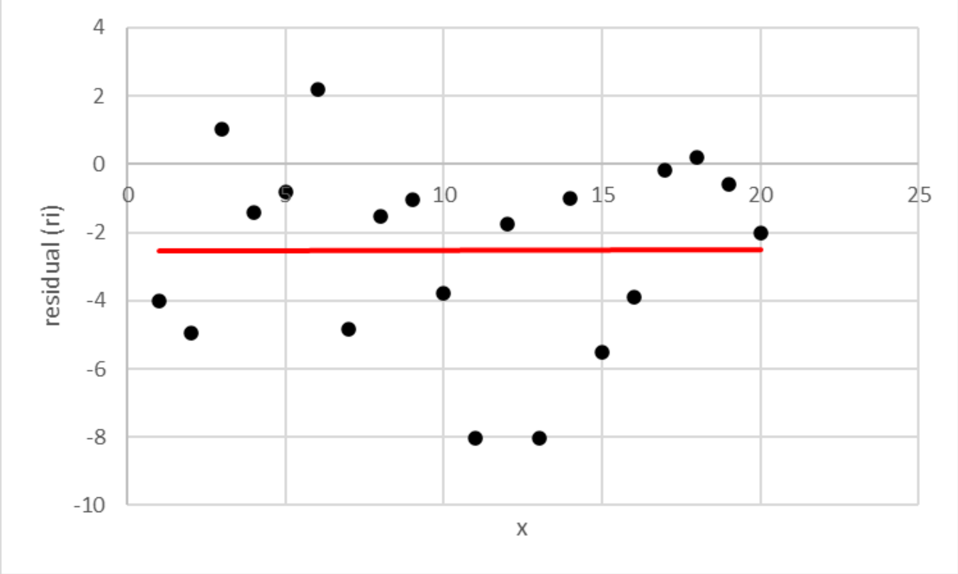
$$M_r = \frac{\sum_{i=1}^{n} r_i}{n}.$$

Measuring bias

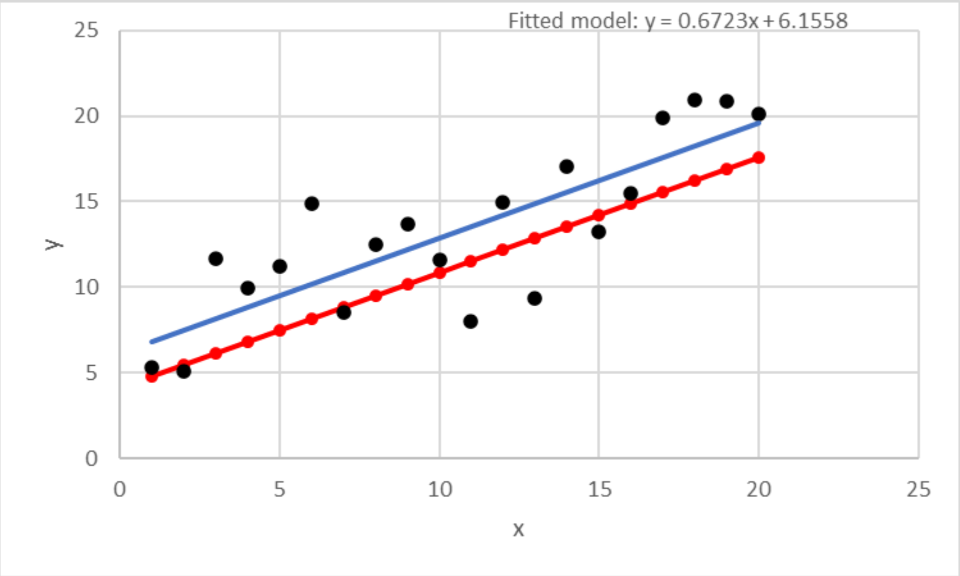


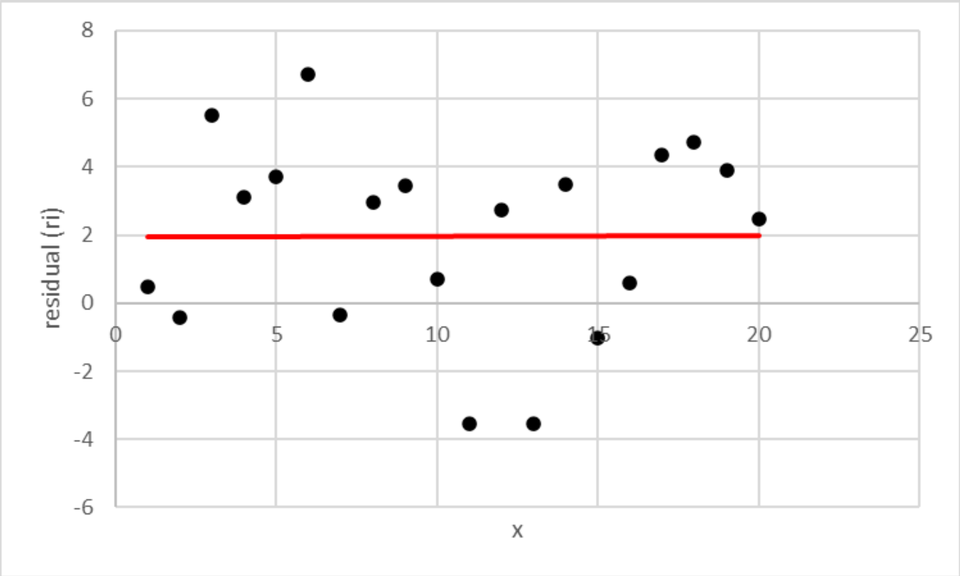


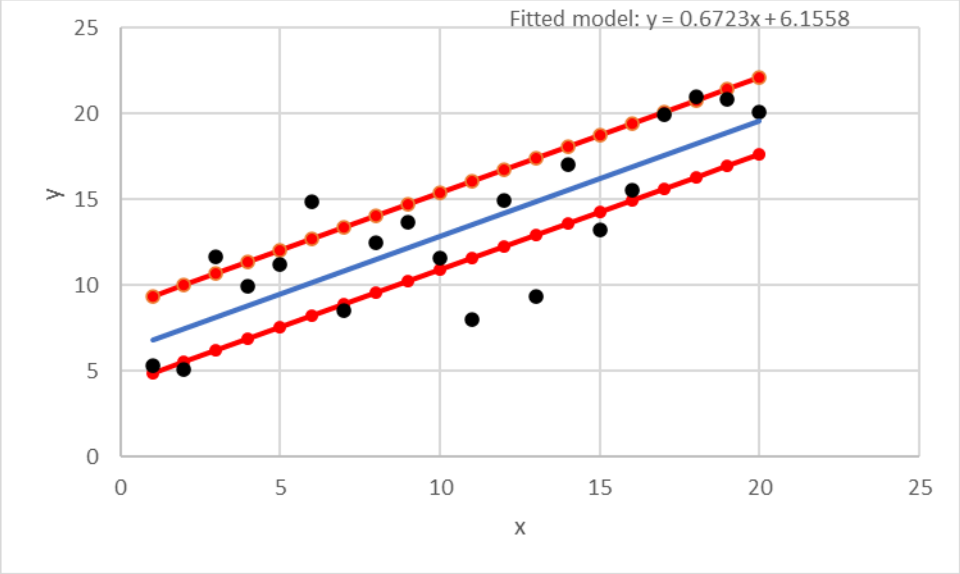




Measuring bias







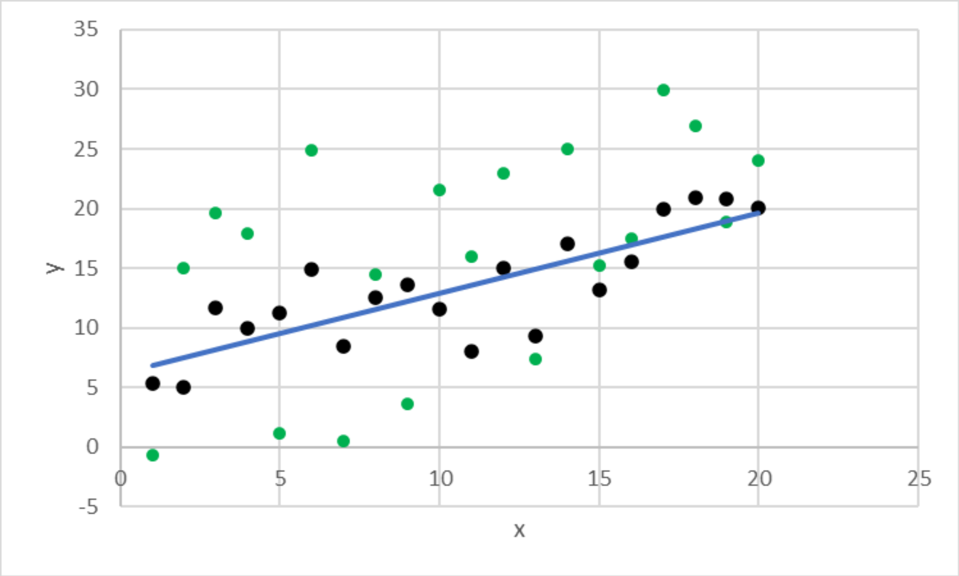
Measuring precision

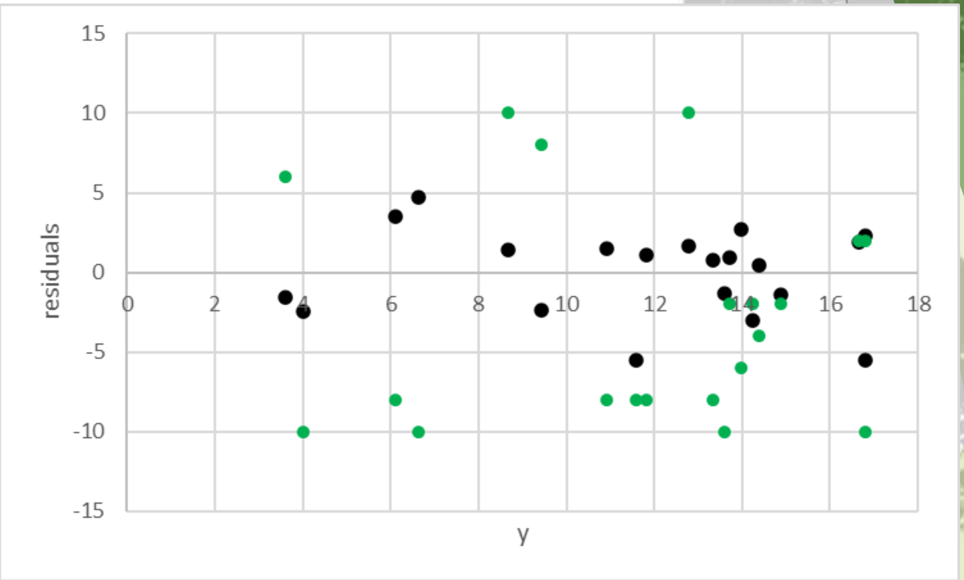
- **Precision:** refers to how accurately a model predicts the dependent (y) variable.
  - As close to zero the better.
  - Evaluated by the absolute value of the residuals $|r_i|$ distribution (percentiles, mean…)

Mean of the absolute value of the residuals ($M_{|r|}$) to evaluate precision:

$$M_{|r|} = \frac{\sum_{i=1}^{n} |r_i|}{n},$$

Measuring precision





Model efficiency

- **Model efficiency (ef)** – the proportion of variation explained by the model.

Model efficiency or the proportion of variation explained by the model (ef):

$$\mathrm{ef} = 1 - \frac{\sum_{i=1}^{n} r_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

where r_i is the residual for observation i, y_i is observation or observed data i, $\bar{y}$ is the mean value of all observed values from the evaluation data set and n is the number of observations from the evaluation data set.

Comparing models

Table 8
Validation statistics. M_r, mean value of the residuals (kg); $M_{|r|}$, mean of absolute value of the residuals (kg); P_5, percentile 5 of residuals (kg); P_{95}, percentile 95 of residuals (kg); ef, model efficiency.

Model	M_r	$M_{\|r\|}$	P_5	P_{95}	ef
I	−1.17	7.81	−14.20	24.59	0.825
II	0.71	7.24	−13.67	21.13	0.843
III	−0.21	5.19	−13.07	12.43	0.931
IV	0.21	4.38	−9.18	9.84	0.957

Paulo and Tomé (2010) http://dx.doi.org/10.1016/j.foreco.2010.02.010

Quadro 7 *- Estatísticas do ajustamento e da capacidade preditiva dos modelos candidatos para avaliação do peso seco de cortiça virgem*

Designação	R^2	M_{press}	M_{apress}	P_{99}	P_{95}	P_5	P_1	R^2press
Mod1	0,620	0,000	0,99	3,66	2,21	-1,85	-3,45	0,611
Mod3	0,645	0,000	0,97	3,67	2,27	-2,00	-3,00	0,633
Mod5	0,627	0,120	0,97	3,73	2,29	-1,76	-3,00	0,613

Paulo and Tomé (2014). http://www.scielo.gpeari.mctes.pt/pdf/slu/v22n1/v22n1a02.pdf

Reading for this topic

- Vanclay, J. K., Skovsgaard, J. P. 1997. Evaluating forest growth models. Ecological Modelling 98 (1997) 1-12. Available: https://fenix.isa.ulisboa.pt/downloadFile/563022967901630/vanclay%201995%20Modelacao.pdf
- Burkhart, H. E., Tomé, M. 2012. Modeling Forest Trees and Stands. Springer. Chapter 18.2