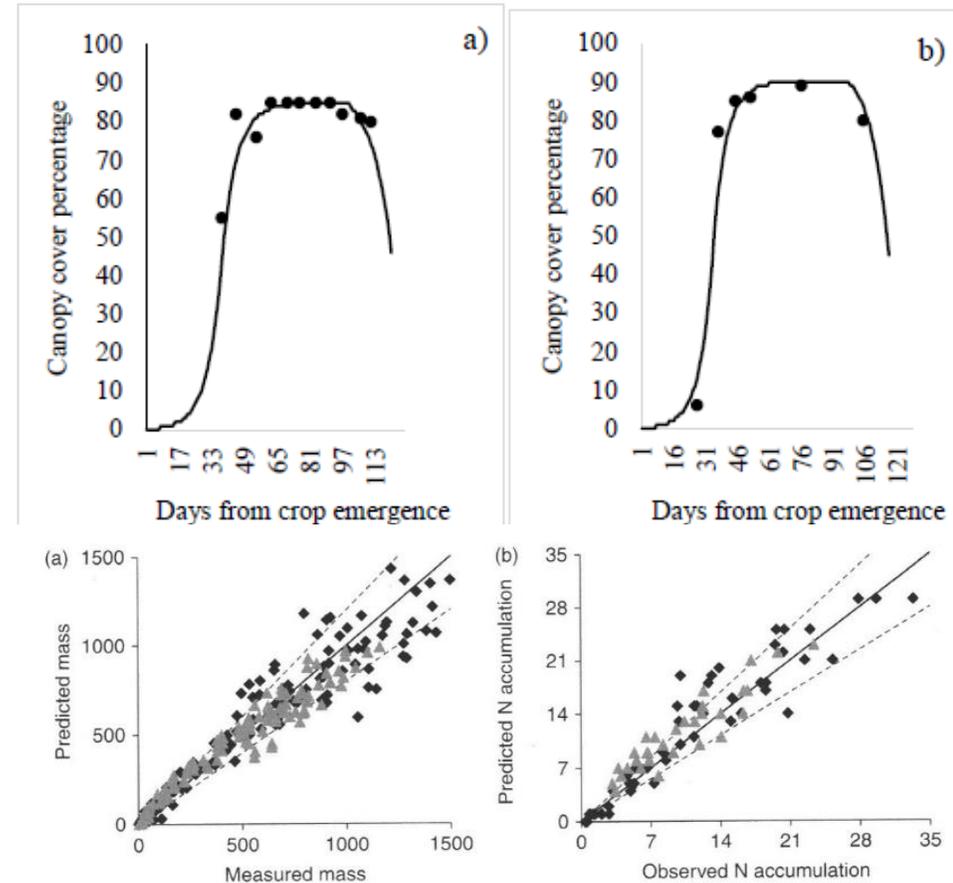


CALIBRAÇÃO E AVALIAÇÃO DE MODELOS



Calibração e avaliação de modelos

- Calibrar um modelo -> ajustar os parâmetros do modelo respeitando seu significado biofísico, buscando aproximar as simulações dos dados experimentais (medições). O ajuste é feito objetivando a minimização do erro estatístico de predição do modelo, observando as variáveis de saída.
- Nota: a estrutura do modelo não é alterada pela calibração.
- Com o modelo calibrado, segue a etapa da avaliação, também chamada de teste ou validação. O termo validação tem a conotação de que o modelo seja “validado”, implicando na sua validade em qualquer circunstância. Neste sentido, há uma célebre frase: “*pode-se apenas invalidar, mas nunca definitivamente validar um modelo*”.
- O que se faz, na verdade, é testar o desempenho do modelo para determinados cenários, avaliando seu desempenho com base em indicadores estatísticos e utilizando um conjunto de dados independentes daqueles usados na calibração.

Calibração e avaliação de modelos

- O processo de avaliação de modelos consiste na verdade em quantificar a habilidade do modelo em simular a realidade. O erro total de um modelo é dado pela soma de dois tipos de erros: a) erros estruturais e erros de estimativa de parâmetros; conforme os modelos tornam-se mais complexos, os erros estruturais diminuem porque o modelo se aproxima do sistema real; os erros estruturais podem ser reduzidos até um certo limite (ver figura abaixo).
- Ainda, conforme aumenta-se a complexidade do modelo, os erros de parâmetros aumentam. Isto ocorre por que aumenta-se o número de equações e parâmetros e isso eleva a dificuldade de ajuste desses parâmetros.

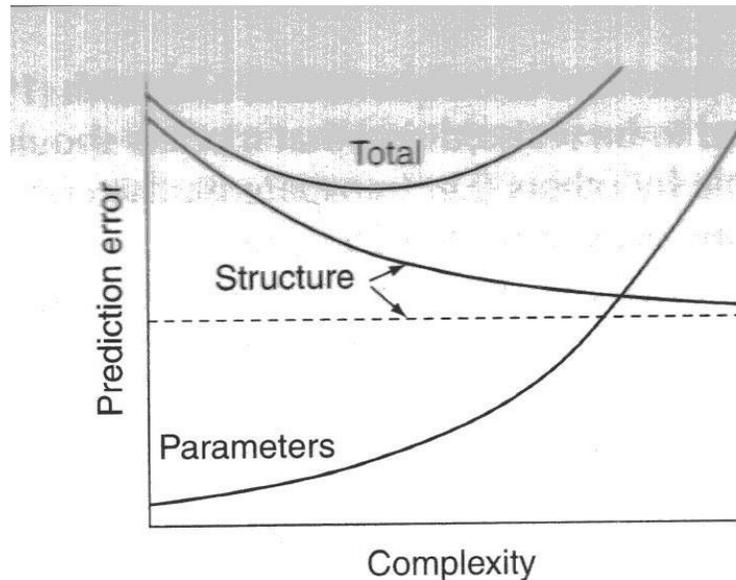


Fig. 3.1. Relationship between complexity and prediction error in biological (crop) models. Horizontal line represents irreducible structural error due to unknown structure of the system (Reynolds and Acock, 1985, as cited in Passioura, 1996).

Exemplo de procedimento de avaliação de um modelo

Example 1

Suppose that our model for predicting response Y for individual i is

$$\hat{Y}_i = f(X_i; \theta) = \hat{\theta}^{(0)} + \hat{\theta}^{(1)}x_i^{(1)} + \hat{\theta}^{(2)}x_i^{(2)} + \hat{\theta}^{(3)}x_i^{(3)} + \hat{\theta}^{(4)}x_i^{(4)} + \hat{\theta}^{(5)}x_i^{(5)} \quad (1)$$

The explanatory variables are $X = (x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)})^T$. The parameter vector is

$$\hat{\theta} = (\hat{\theta}^{(0)}, \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \hat{\theta}^{(3)}, \hat{\theta}^{(4)}, \hat{\theta}^{(5)})^T = (1.9, 7.8, 2.5, -0.2, 0.1, 0.7)^T \quad (2)$$

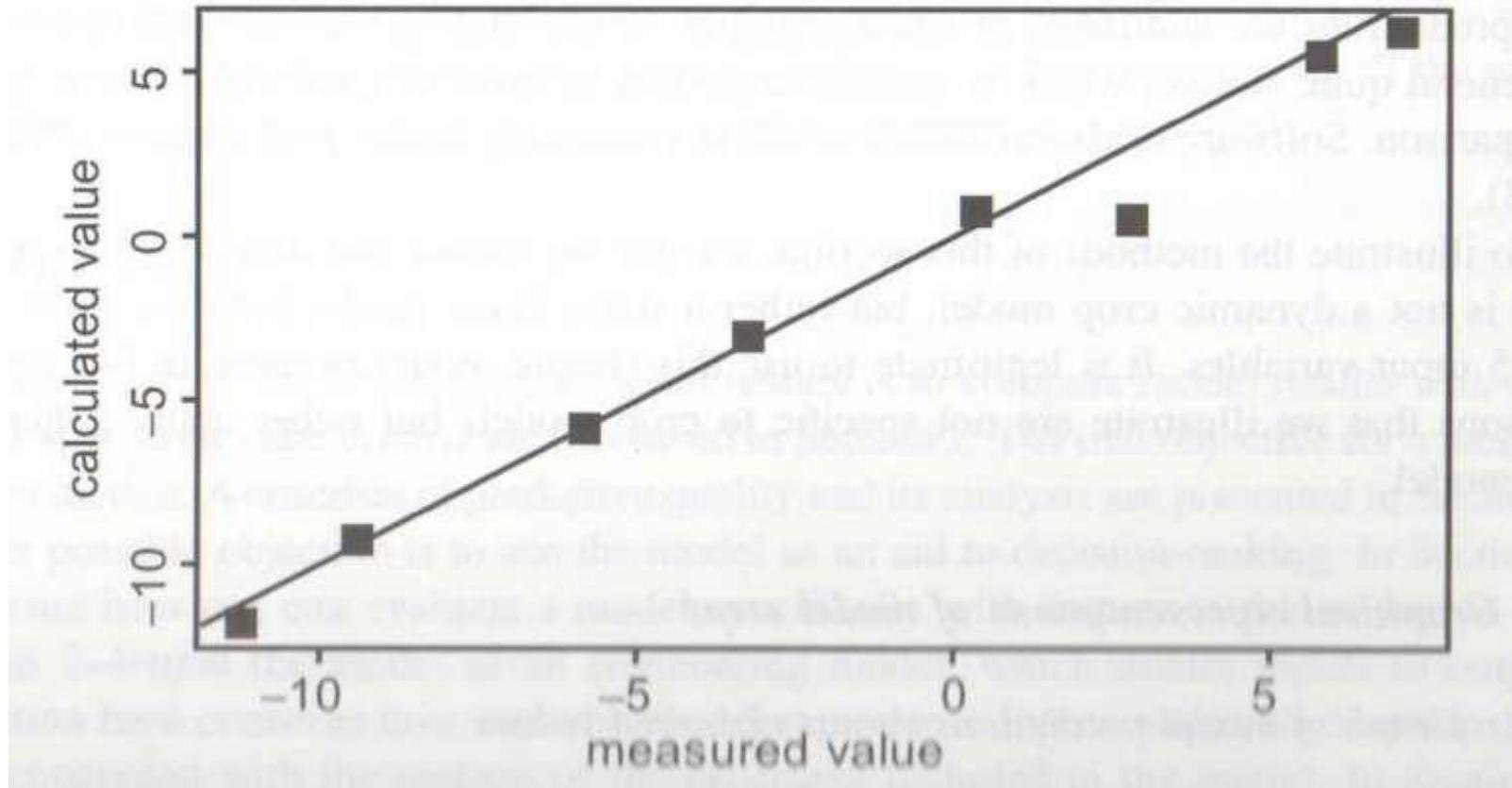
The hat notation is used to indicate that the parameter values are estimates. We do not need to bother here with the origin of these estimates. The data set for evaluating the model is given in Table 1.

Table 1. Measured values (Y_i), 5 explanatory variables, calculated values (\hat{Y}_i) and model errors (D_i) for 8 situations.

i	Y_i	$x_i^{(1)}$	$x_i^{(2)}$	$x_i^{(3)}$	$x_i^{(4)}$	$x_i^{(5)}$	\hat{Y}_i	$D_i = Y_i - \hat{Y}_i$
1	-9.39	-1.63	0.80	0.44	-0.45	-0.47	-9.31	-0.08
2	-3.23	-0.95	1.07	0.50	0.53	-0.33	-3.10	-0.13
3	0.37	-0.25	0.20	0.51	0.82	0.45	0.72	-0.35
4	7.10	0.38	1.02	-0.22	0.82	-1.93	6.18	0.92
5	5.82	0.15	1.14	1.07	1.63	-0.55	5.45	0.38
6	-11.21	-1.20	-1.79	0.35	-0.26	0.26	-11.84	0.63
7	-5.81	-0.97	-0.15	0.11	1.13	0.10	-5.90	0.09
8	2.82	0.39	-1.20	2.01	-0.79	-1.44	0.47	2.35

Fonte: Wallach et al. Working with Dynamic Crop Models. 2017.

Representação Gráfica



Normalmente se adiciona aqui a equação, o r^2 e o valor-p do ajuste.

Fonte: Wallach et al. *Working with Dynamic Crop Models*. 2017.

1100222 - Modelagem de Crescimento de Culturas Agrícolas
LEB5048 - Modelagem de Culturas Agrícolas I



Table 2. Measures of agreement between a model and measured data.

Name	Equation
Bias	$\text{Bias} = \frac{1}{N} \sum_{i=1}^N D_i$
Mean squared error	$\text{MSE} = (1/N) \sum_{i=1}^N (D_i)^2$
Root mean squared error	$\text{RMSE} = \sqrt{\text{MSE}}$
Mean absolute error	$\text{MAE} = \frac{1}{N} \sum_{i=1}^N D_i $
Relative root mean squared error	$\text{RRMSE} = \frac{\text{RMSE}}{\bar{Y}}$
Relative mean absolute error	$\text{RMAE} = \frac{1}{N} \sum_{i=1}^N \frac{ Y_i - \hat{Y}_i }{ Y_i }$
Modeling efficiency	$\text{EF} = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$
Correlation coefficient	$r = \frac{\sum_{i=1}^N [(Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})]}{\sqrt{\sum_{i=1}^N [(Y_i - \bar{Y})^2] \sum_{i=1}^N [(\hat{Y}_i - \bar{\hat{Y}})^2]}}$
Agreement index	$\text{index} = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (\hat{Y}_i - \bar{Y} + Y_i - \bar{Y})^2}$
Concordance correlation coefficient	$\rho_c = \frac{2\sigma_{Y\hat{Y}}}{\sigma_{\hat{Y}}^2 + \sigma_Y^2 + (\mu_{\hat{Y}} - \mu_Y)^2}$
Total deviation index	$\text{TDI}(p) = \text{minimal value of } d \text{ such that } D_i \leq d \text{ for at least } p\% \text{ of the observed situations.}$
Coverage probability	$\text{CP}(d) = \text{the smallest value of } p \text{ such that, for a percentage } p \text{ of the observed situations, } D_i \leq d $

Indicadores estatísticos

Formas de representação gráfica

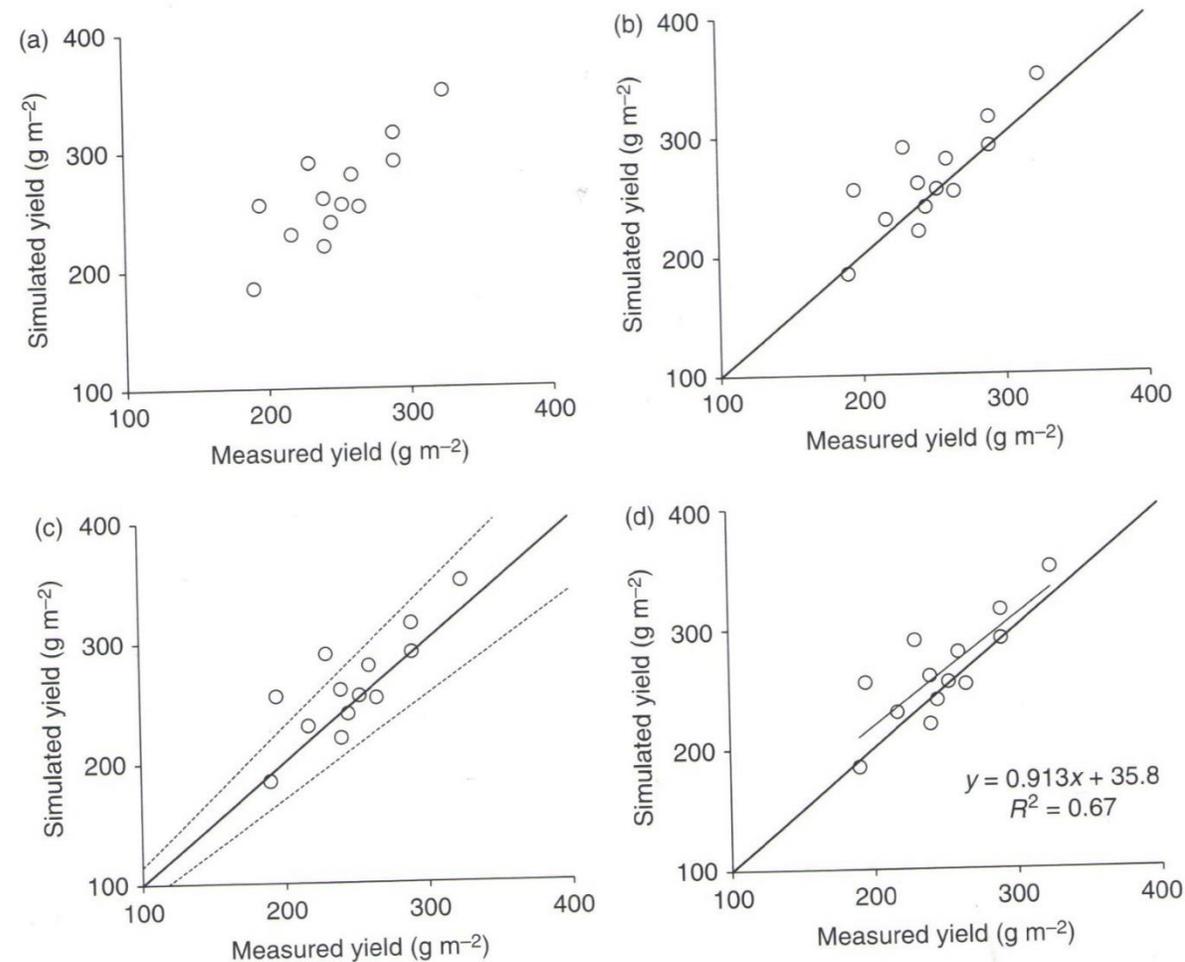


Fig. 3.2. Evaluation of model robustness using simulated versus measured graphs: (a) a simple graph of simulated versus measured variable, (b) simulated versus measured variable plus 1:1 line, (c) simulated versus measured variable plus 1:1 line plus $\pm 15\%$ discrepancy lines, and (d) simulated versus measured variable plus 1:1 line plus regression line.

Formas de representação gráfica

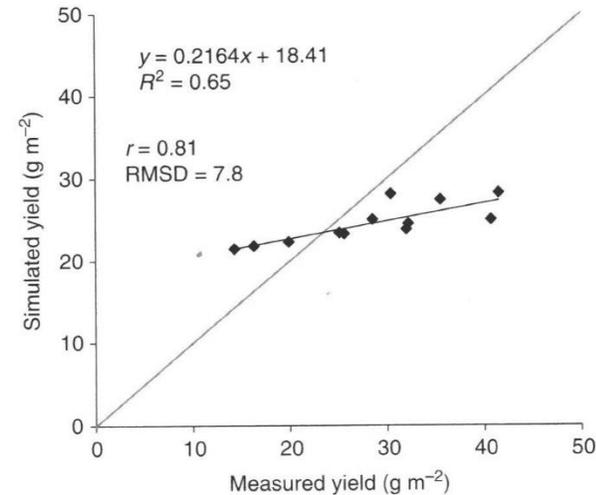


Fig. 3.3. A sample graph of simulated versus measured variable in which the linear regression between simulated and measured variable resulted in high R^2 and r , indicating that the model obviously is not a good simulator of the system.

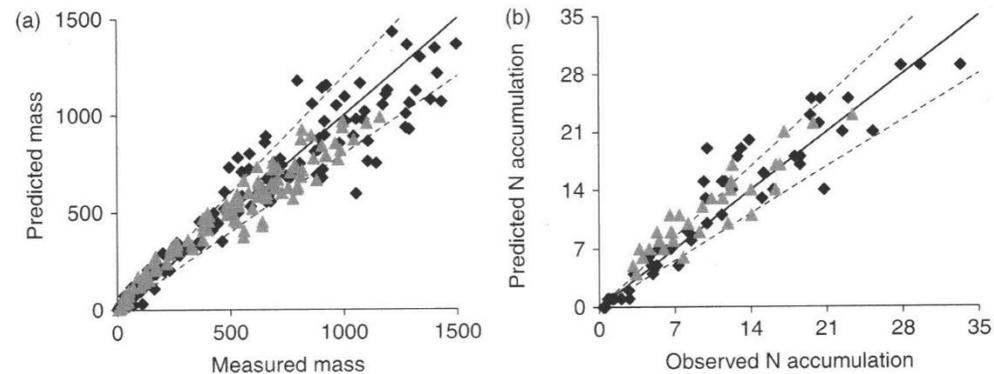


Fig. 3.4. Sample graphs of (a) predicted crop mass and (b) nitrogen (N) accumulation, both in g m⁻², during growing season in chickpea crops grown under a wide range of sowing date and density. Different symbols indicate two different experiments. The 18% ranges of discrepancy between simulated and measured values are indicated by dashed lines. Solid line is 1:1 line (Soltani and Sinclair, 2011).

Formas de representação gráfica

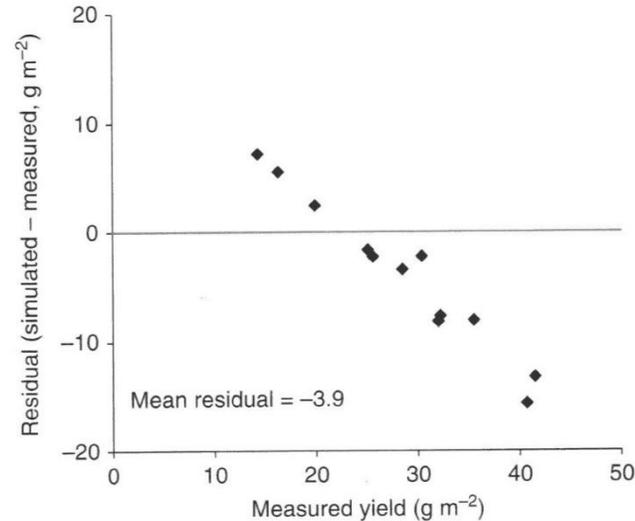


Fig. 3.5. A sample graph of residual (simulated – measured) versus measured variable derived from Fig. 3.3. This example indicates an obvious bias in model prediction. Mean bias (residual mean) is -3.9 g m^{-2} .

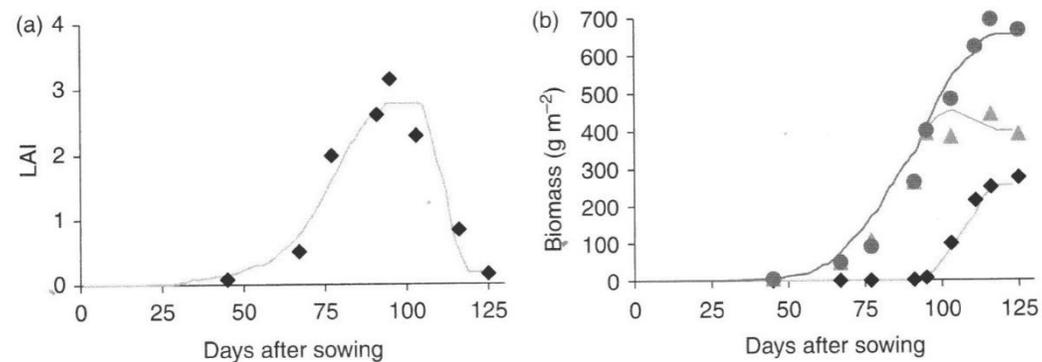
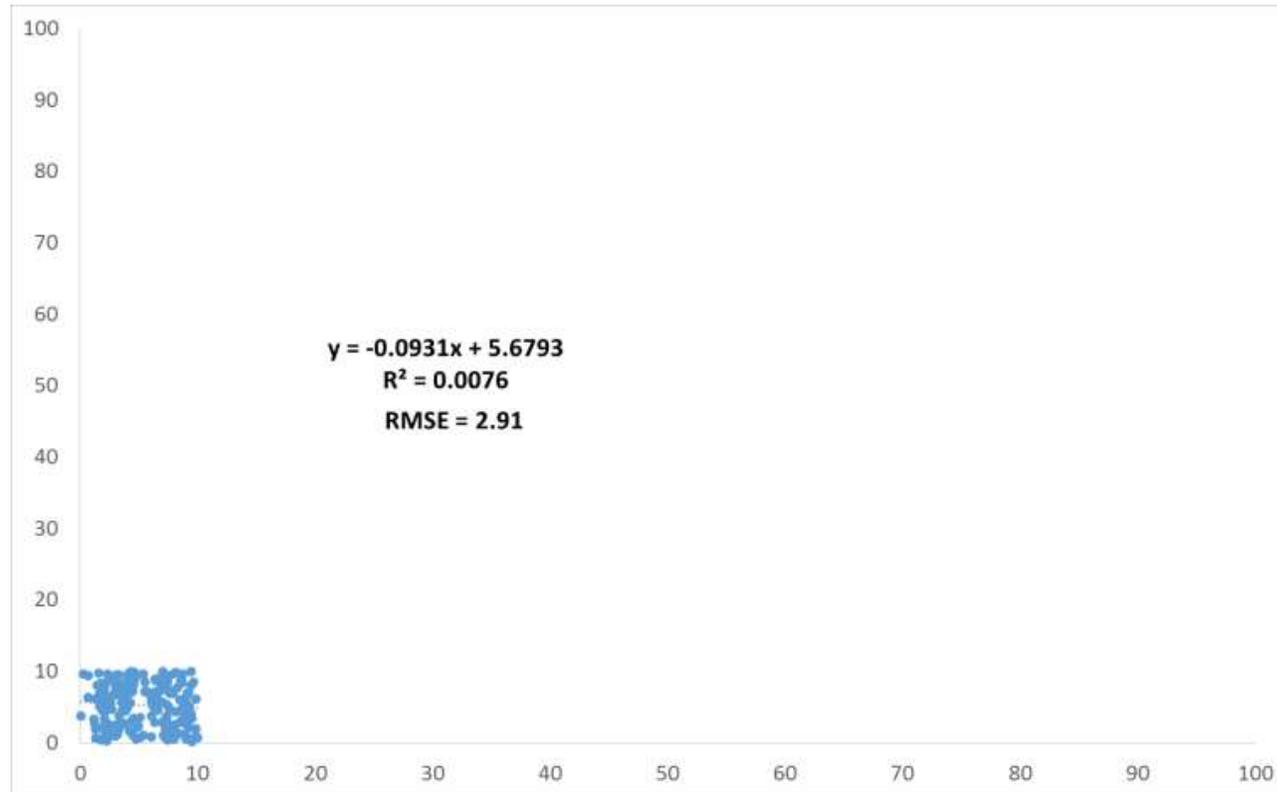


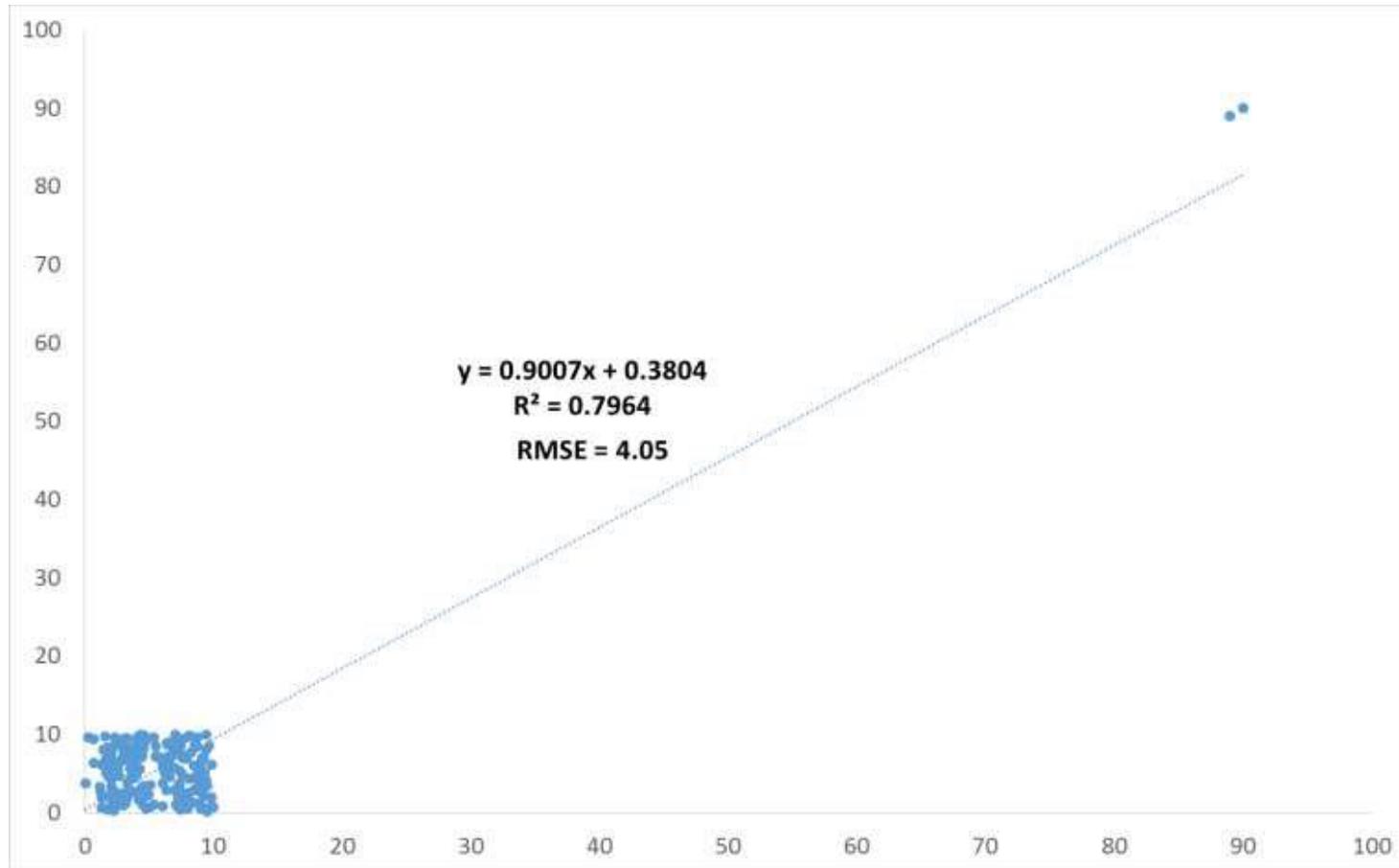
Fig. 3.6. Simulated and measured values of (a) leaf area index (LAI) and (b) vegetative, seed, and total dry matter by a chickpea crop (Soltani and Sinclair, 2011).

Observação sobre o R^2



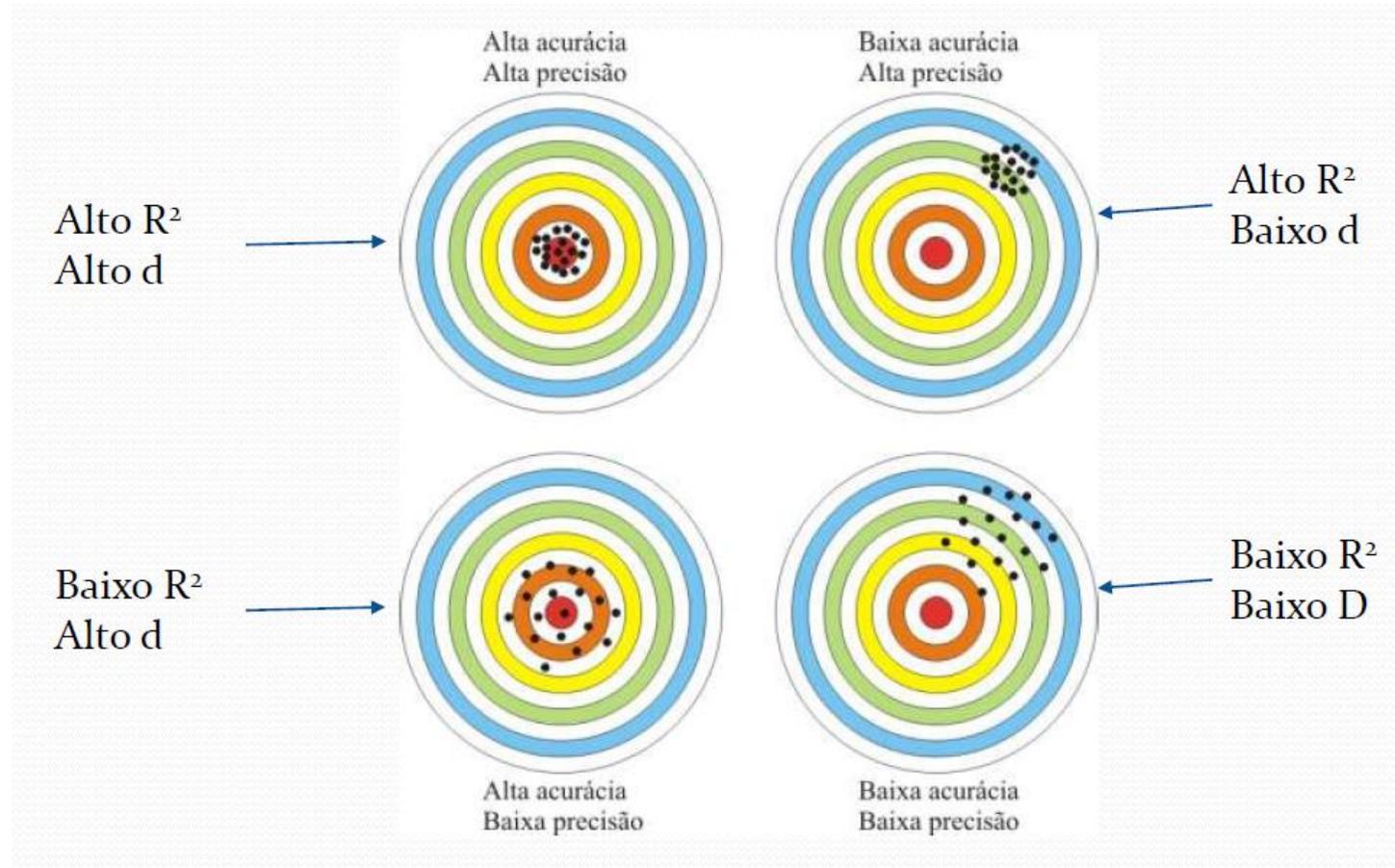
O R^2 não é um bom indicador para a qualidade preditiva de um modelo mecânico ...

Observação sobre o R^2



Note que adicionando-se apenas 2 pontos, o R^2 aumenta expressivamente. Observe abaixo e compare com a figura do slide anterior.

Acurácia e precisão



Roteiro sugerido num processo de calibração eye-fitting

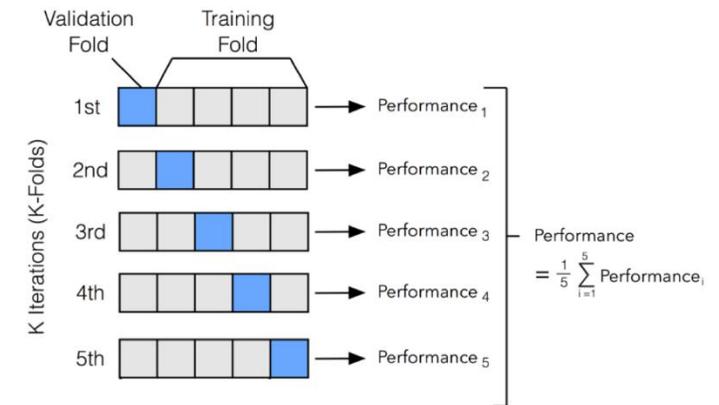
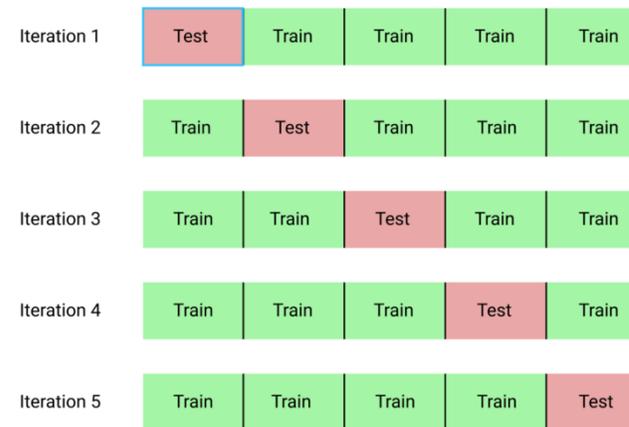
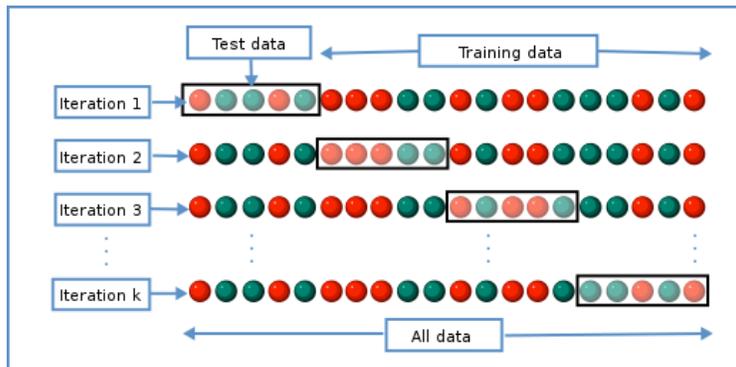
- 1) atribuir um valor para cada parâmetro considerando o significado biofísico do parâmetro, a análise de sensibilidade e os valores reportados na literatura;
- 2) utiliza-se um critério estatístico para quantificar a incerteza associado a esta primeira calibração
- 3) ajusta-se os valores do parâmetros (sempre considerando o significado biofísico do parâmetro e a análise de sensibilidade) e observa-se a variação no indicador estatístico;
- 4) o conjunto final de parâmetros é selecionado quando se atinge o valor mínimo do indicador estatístico (mínima incerteza na calibração)

Validação Cruzada – Cross-Validation

- Assumindo que o conjunto de dados seja uma amostra aleatória da população alvo, normalmente divide-se os dados em duas partes: uma para calibração e outra para avaliação.
- Como o segundo conjunto de dados não foi utilizado na calibração, diz-se que ele é um estimador não-parcial da incerteza do modelo
- Quando temos, contudo, poucos dados, a validação cruzada é uma opção interessante. Além disso, esta abordagem acima descrita tem algumas limitações:
 - Em primeiro lugar, normalmente se usa apenas metade dos dados na calibração e para estimar a incerteza do modelo, tornando nossas estimativas menos precisas.
 - Em segundo lugar, este procedimento de separação dos dois grupos é arbitrário, bem como a definição das proporções a serem utilizadas na calibração e na validação.

Validação Cruzada – Cross-Validation

- Observe os seguintes esquemas explicativo sobre a validação cruzada:



Como mostram as figuras acima, a validação cruzada consiste em dividir o conjunto de dados em vários grupos e utilizar no processo de calibração todos eles, com exceção de um conjunto. A validação nessa primeira rodada de calibração seria então feita com este conjunto que ficou fora. Na sequência, o processo é repetido, mas agora deixando de fora um segundo (diferente) conjunto de dados. O procedimento continua até que todos os conjuntos tenham sido usados tanto na calibração como na avaliação do modelo. Ao final do procedimento, teremos N estimativas de incerteza do vetor de parâmetros do modelo, mas sugere-se usar como conjunto final de parâmetros calibrados aquele construído utilizando-se todos os dados.

Tarefa

- Com base nos dados apresentados na Tabela do slide 4, representar graficamente o ajuste entre os dados estimados e observados e calcular os indicadores estatísticos que constam no slide 6.