

Métodos de validação

Alguns métodos para validação de modelos ambientais

Definição

- Validação é verificar se os resultados produzidos pelos modelos correspondem aos resultados que se espera dos mesmos. Pode-se, por exemplo, simular uma situação conhecida e verificar se o modelo consegue reproduzi-la com algum grau de confiabilidade.

Métodos de validação

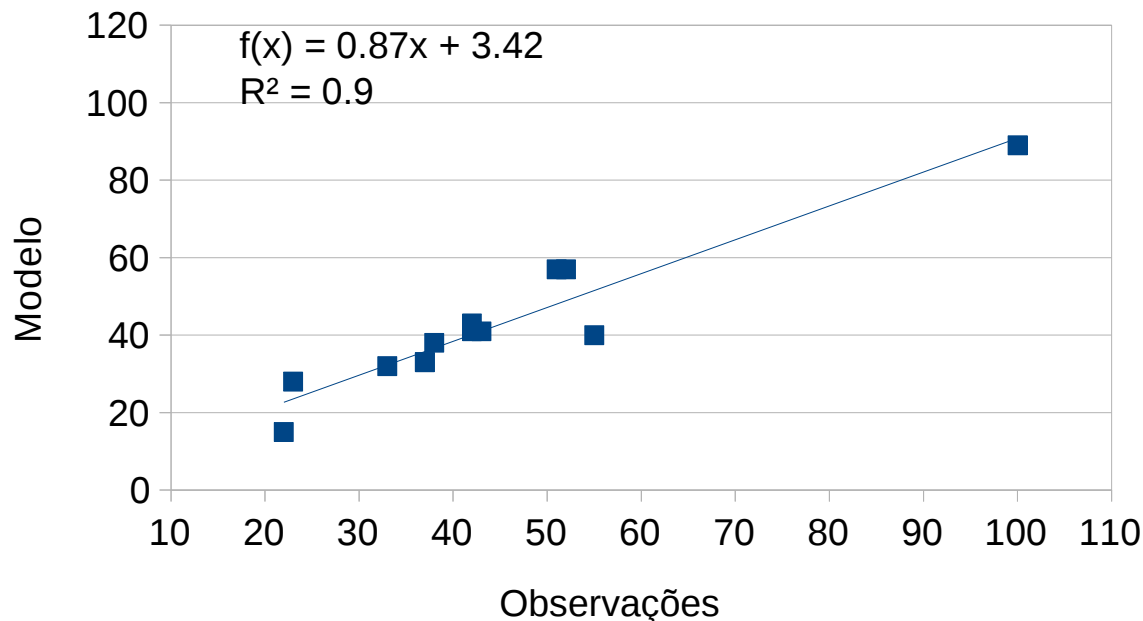
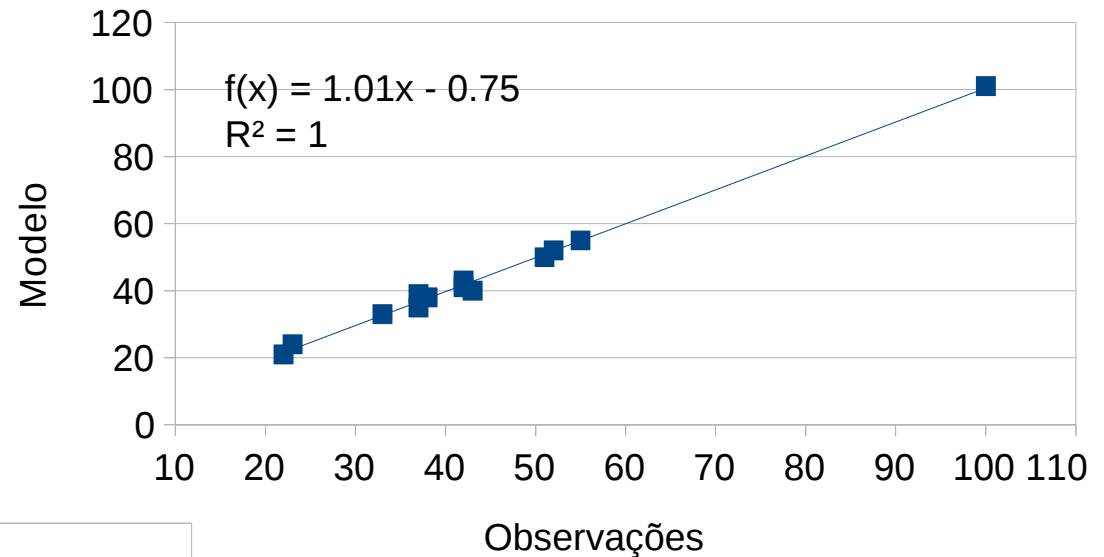
- Podemos usar alguns métodos estatísticos para fazer a validação de modelos.
 - Correlação linear
 - Erro médio quadrático / RMSE
 - Bias (erro sistemático)
 - Index of agreement (IOA) → índice de concordância
 - Hit rate (HR) → taxa de acerto

Correlação linear

- Já falamos sobre correlação linear para verificar se há uma relação linear entre duas variáveis. Aqui queremos calcular a correlação linear entre dados observados e os resultados do modelo.
- Fazemos um gráfico de dispersão, onde os dados observados estão em um dos eixos e os resultados no outro eixo.
- Cada ponto do gráfico leva o valor de um dado observado ao valor calculado pelo modelo nas mesmas condições.
- Um modelo “perfeito” calcularia sempre o mesmo valor que o observado, então o gráfico de dispersão teria todos os pontos sobre a reta $y = x$. Essa reta não tem intercepto e sua inclinação é 1.
- Assim, quando utilizamos o método dos mínimos quadrados para encontrar uma reta que relacione os dados observados com os calculados, queremos uma inclinação o mais próxima de 1 possível.

Correlação linear - exemplos

- Em qual dos exemplos o modelo demonstra melhor desempenho?



Erro Médio Quadrático

- MSE → mean square error
- É uma maneira de avaliar a média do quadrado do erro (diferença entre valor observado e modelado).
- É igual à soma da variância e a tendenciosidade do quadrado do valor modelado.
- Quanto mais próximo de zero, mais o modelo se aproxima da “realidade” (valor observado).
- Consideração: esse método coloca mais peso em erros grandes que em pequenos (quadrado).

- Fórmula:

$$EMQ = \frac{\sum_{i=1}^n (o_i - m_i)^2}{n}$$

RMSE

- Root mean square error → é a raiz quadrada do erro médio quadrático

- Fórmula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (o_i - m_i)^2}{n}}$$

BIAS (Erro sistemático)

- Quando comparamos as observações com os valores modelados, podemos ter 2 tipos de erro: aleatório e sistemático.
 - Aleatório: devido a inúmeras razões
 - Sistemático: o modelo sistematicamente superestima, ou subestima, a variável
- O erro sistemático (bias) pode ser estimado como a diferença entre a média dos valores observados e a dos valores calculados.

Fórmula:

$$BIAS = \frac{\sum_{i=1}^n m_i - \sum_{i=1}^n o_i}{n}$$

IOA – Index of agreement

- Proposto por Wilmot (1981).
- Indica a porcentagem de valores modelados que “concordam” com os valores observados.
- É calculado pela razão entre o erro médio quadrático e o erro potencial.

Fórmula:
$$IOA = 1 - \frac{\sum_{i=1}^n (o_i - m_i)^2}{\sum_{i=1}^n (|m_i - \bar{o}| + |o_i - \bar{o}|)^2}$$

HR – Hit Rate

- Representa a porcentagem de valores modelados que ficaram dentro de um determinado intervalo em torno da correspondente observação.
- É necessária a determinação do valor do intervalo (critério).
- Por exemplo: qual a porcentagem de valores modelados que ficaram apenas a 2 unidades de diferença do valor observado, tanto para mais quanto para menos?
 - Valor observado = 37.5 e valor modelado = 35.7 → está no intervalo
 - Valor observado = 35.6 e valor modelado = 33.1 → não está no intervalo

Fórmula:

$$HR = \frac{\text{número de valores modelados dentro do intervalo}}{\text{número total de valores comparados}}$$

Exercício

- Usando a tabela abaixo, calcule a inclinação da reta de linear de tendência, o erro médio quadrático, o RMSE, o BIAS, o IOA e o HR.

Observações	Modelo
33	32
45	49
53	54
70	70
62	59