Verificação e Comparação de Resultado de Modelos

Quão bom é o meu modelo?

Objetivo:

- Comparação entre séries de dados obtidos através de modelos numéricos.

- Referências:

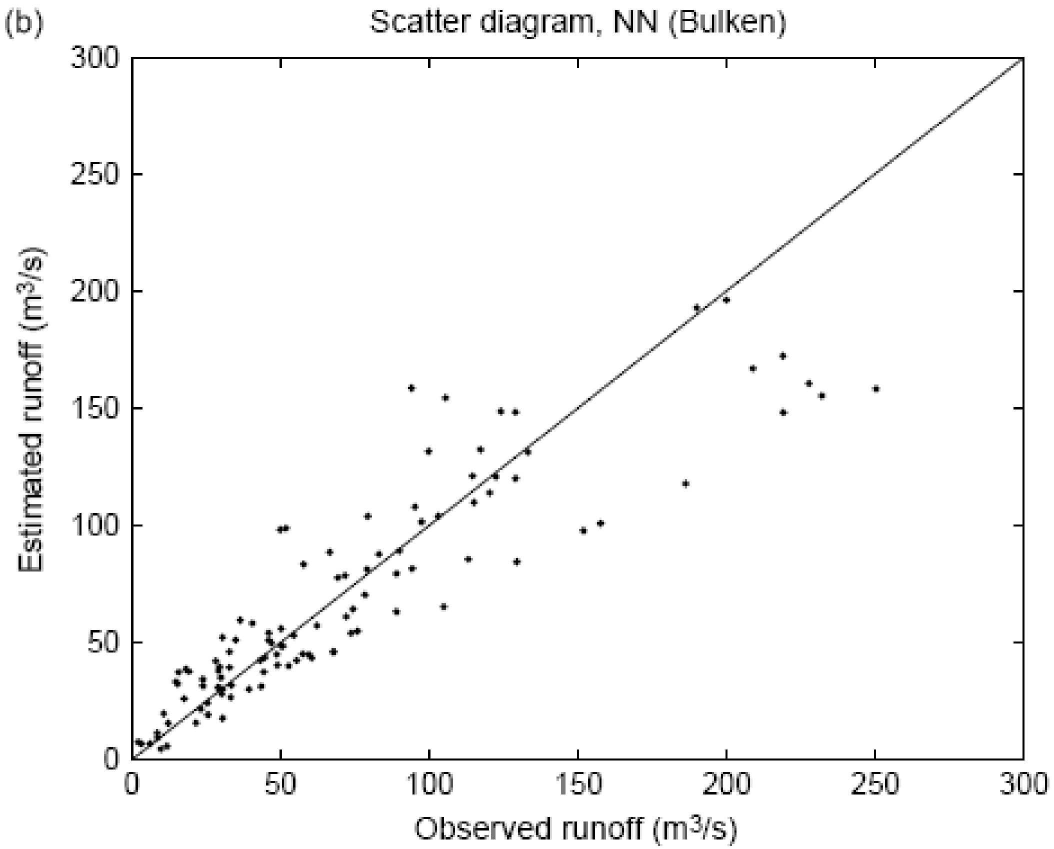
  - Willmott, CJ (1981) On the validation of models. *Physical Geography,* 2, 2, 184-194.
  - Willmott, CJ (1982) *BAMS*
  - Harmel, RD and Smith PK (2007) Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *J. Hydrology*, 337, 326 – 336.
  - **Clarke, RT (2008) Issues of experimental design for comparing the performance of hydrologic models. *Water Res. Research*.** Doi:10.1029/2007WR005927,2008
  - **Clarke, RT (2008) A Critique of Present Procedures Used to Compare Performance of Rainfall-Runoff Models. *J. Hydrology.***

Exemplos:

- Regressão Linear

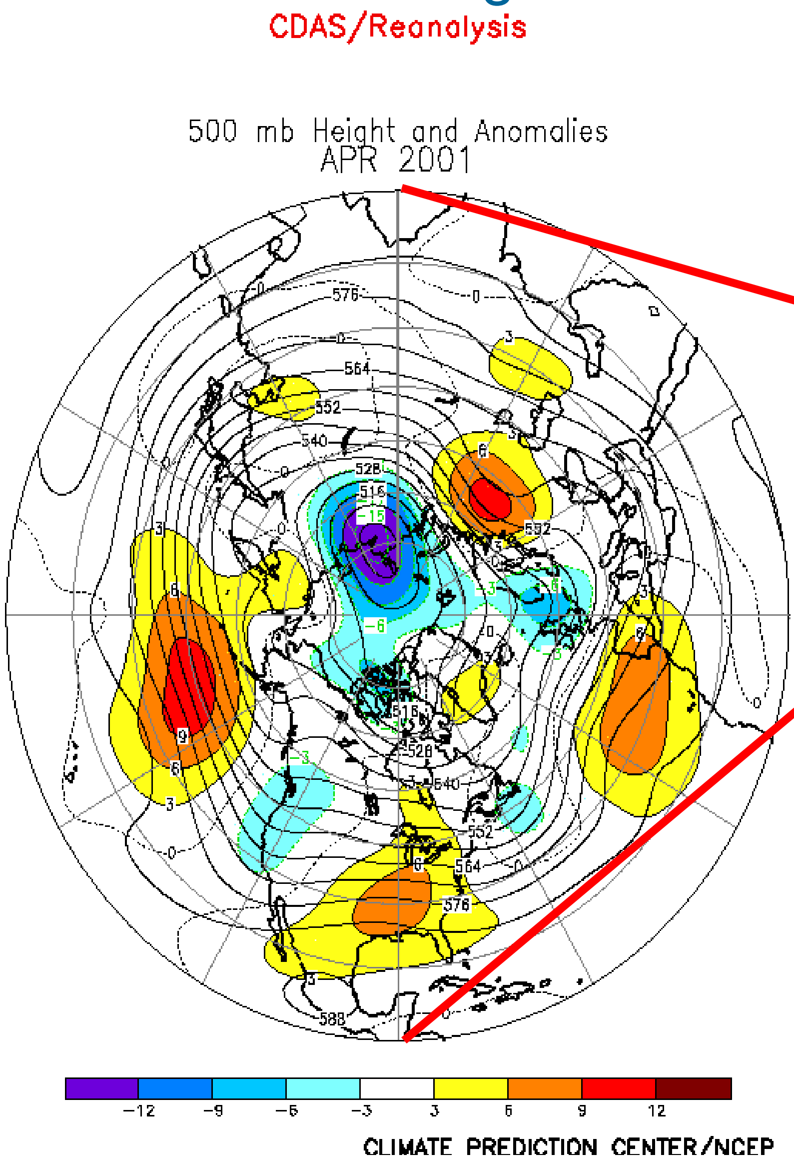
$$\hat{u} = a_1X_1+a_2X_2+...a_nX_n$$

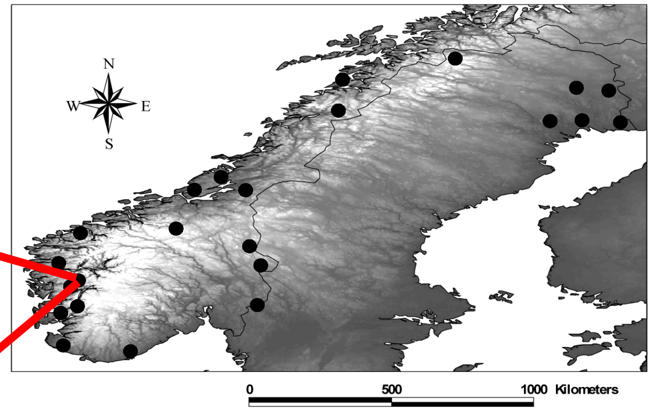
- Compara-se û com u para se obter uma idéia da qualidade da regressão.

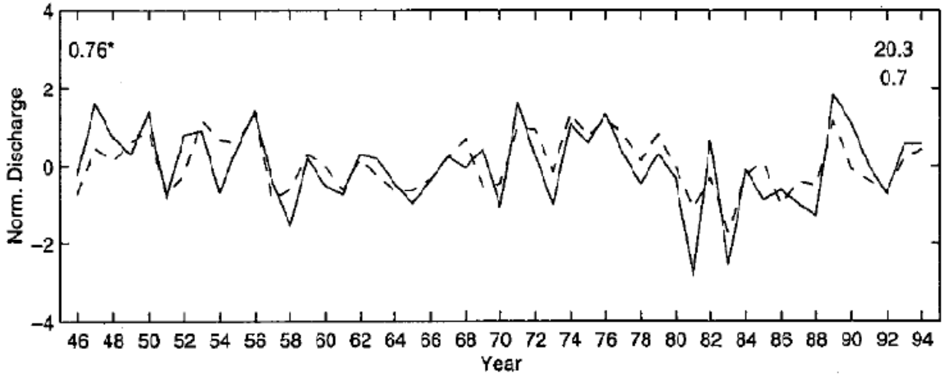


Exemplos

- Downscaling

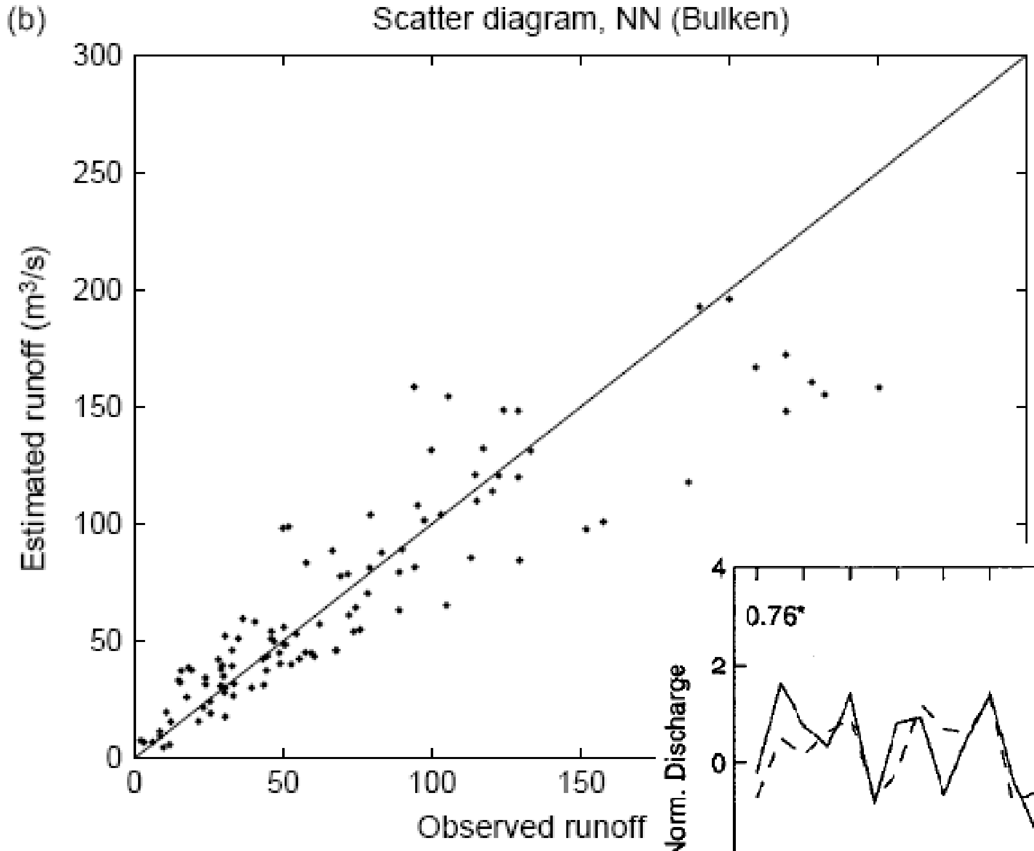


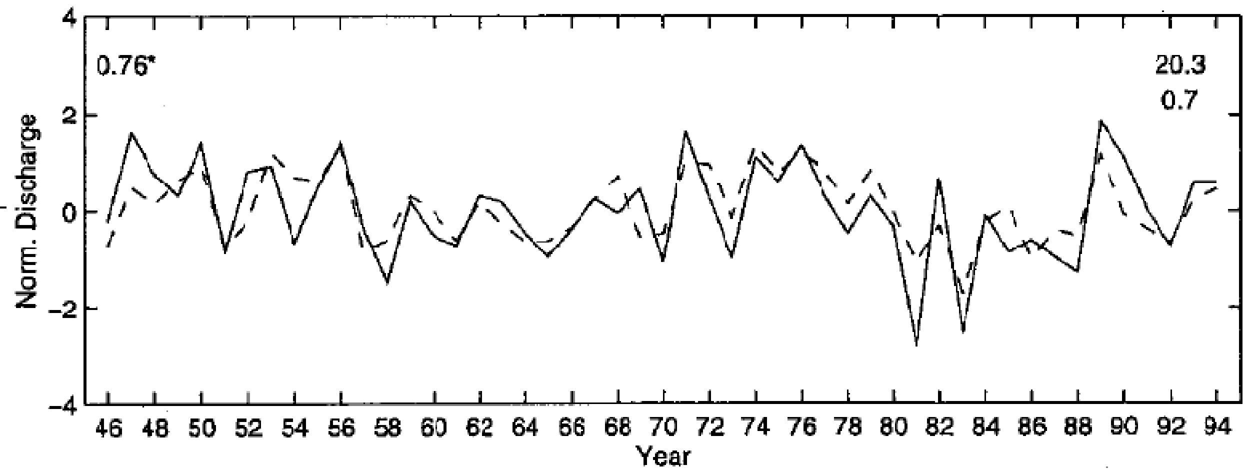




O que comparar?

- Série observada e **reservada** para verificação
- Série estimada pelo modelo





Como comparar ou definir a qualidade da série estimada?

Comece por comparar:

- Média

  $Mo = 1/n * \Sigma P_o$

  $Me = 1/n * \Sigma P_e$

- Desvio padrão

  $dp_o = (1/n * \Sigma(P_o - P)^2)^{1/2}$

  $dp_e = (1/n * \Sigma(P_e - P)^2)^{1/2}$

Como comparar ou definir a qualidade da série estimada?

- **Coeficiente de correlação** linear de produto – momento de Pearson.
- **Definição**: razão entre a covariância e o produto dos dois desvios padrões de duas variáveis X e Y.

$$r_{XY} = \frac{\text{cov}(X,Y)}{s_X s_Y} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}$$

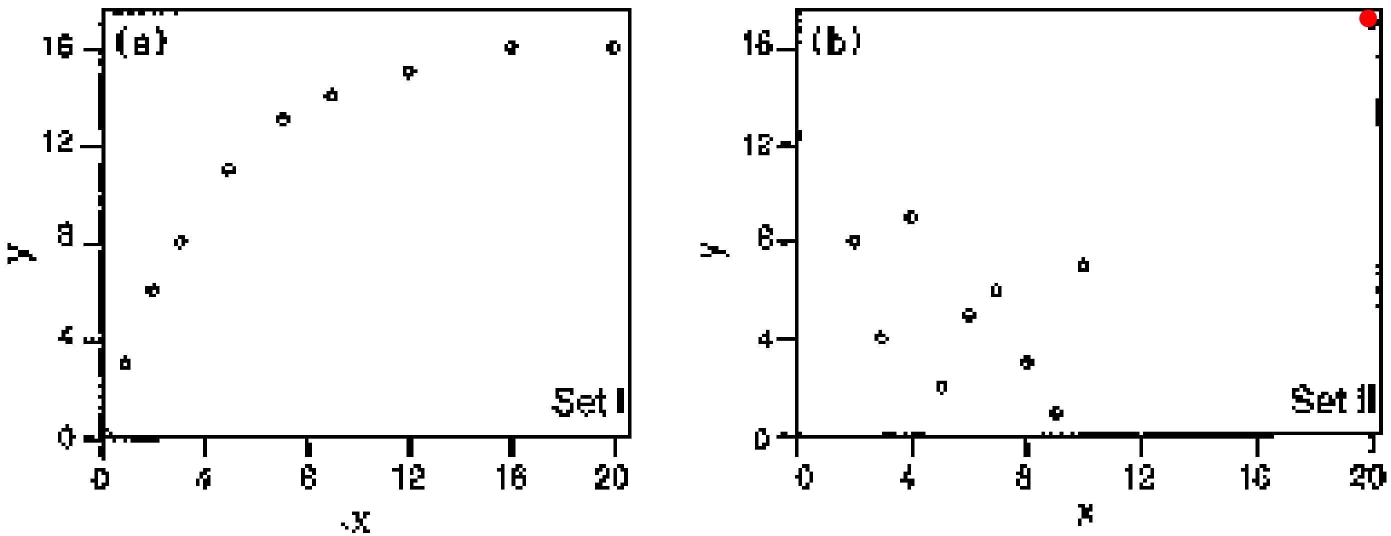
- Vantagens:

  $-1 >= r_{XY} >= 1$

  Não tem unidade

Coeficiente de Correlação r

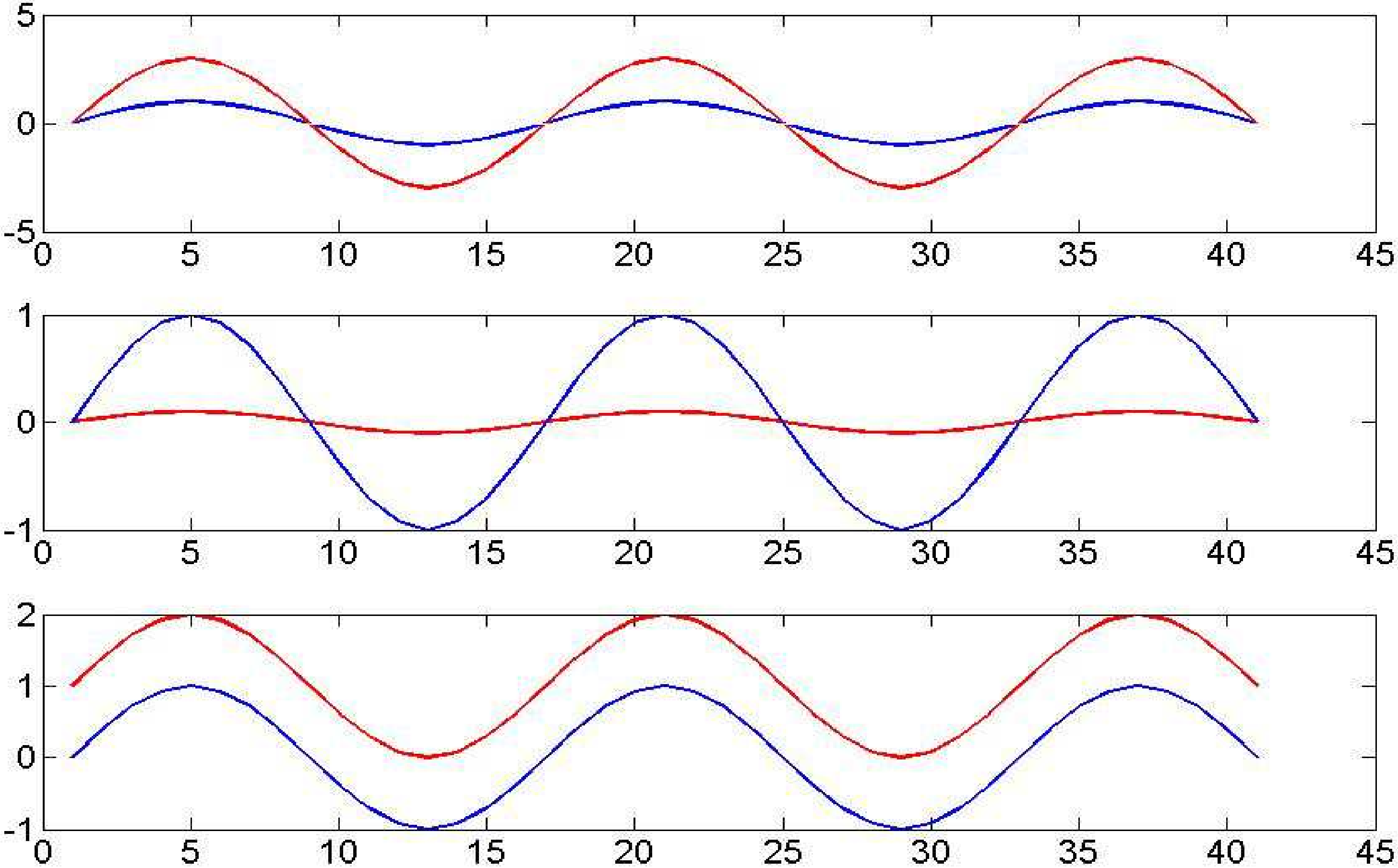
- Desvantagem:
  NÃO é robusto
- Extremamente sensível a um ou mais pares de dados fora do padrão.



Coeficientes de correlação: 0.88 e 0.61

Coeficiente de Correlação r

Insensível a diferenças proporcionais ou aditivas entre a série observada e a série estimada.



Coeficientes de correlação: 1 --obs -- est

Coeficiente de Determinação R^2

- É o quadrado do coeficiente de correlação se a variável independente tiver distribuição normal (gaussiana).

- R^2 expressa a magnitude da variabilidade da variável independente, linearmente descrita pelo modelo.

- Vantagens:

  $0 <= R^2 <= 1$

- Desvantagens:

  As mesmas que de r

Como incrementar a validação do modelo

- Volume total

  $V_o = \Sigma\ P_o$ (volume total observado),

  $V_e = \Sigma\ P_e$ (volume total estimado)

- Erro Médio Quadrático

  $EMQ = 1/n^*\Sigma(P_e - P_o)^2$ onde n é o número de eventos

  Penaliza quadraticamente os erros.

- Raiz Quadrada do Erro Médio Quadrático

  $REMQ = (1/n^*\Sigma(P_e - P_o)^2)^{1/2}$

Como incrementar a validação do modelo

- Erro Médio Absoluto

  $EMA = 1/n*\Sigma |P_e - P_o|$

  Mostra o erro médio sem penalizarão, porém não fornece qualquer informação sobre a variabilidade dos erros.

- Erro Padrão Absoluto

  $EPA = (1/n*\Sigma|P_e - P_o|)^{1/2}$

É o equivalente a medir o desvio padrão dos erros. Dá uma indicação de quanto o erro varia em torno do EMA.

Use r ou R^2 juntamente com alguma outra medida

Índices estatísticos para dados com chuva prevista e observada não nula.

	1° dia	2° dia	3° dia
Σ Volume Previsto (mm)	2050,86	2092,85	2291,99
Σ Volume Observado (mm)	2537,51	2562,05	2501,72
Σ Vol,(Previsto - Observado) (mm)	-486,65	-469,20	-209,73
EMQ (mm)	187,59	212,64	312,02
EMA (mm)	8,83	9,61	11,38
EPA (mm)	13,70	14,58	17,66
R	0,45	0,38	0,25

Outros coeficientes muito interessantes e úteis

- Coeficiente de Eficiência ou Nash-Sutcliffe E

$$E = 1 - \frac{\sum (Po_i - P\mathbf{e}_i)^2}{\sum \left(Po_i - \overline{Po}\right)^2}$$

$\overline{P_o}$ é a média dos valores observados

$-\infty < E < 1$

Fisicamente: E= 1 – EMQ / variância obs

Sensível a valores extremos (como R^2)

E = 0 se EMQ é igual à variância dos dados observados

E < 0 se EMQ é maior que a var. dos dados observados (modelo pior que estimar pela média)

E > 0 se EMQ é menor que a var. dos dados observados (modelo MELHOR que estimar pela média)

Variações de E

- Coeficiente de log Eficiência ou log Nash-Sutcliffe Elog

$$E\log = 1 - \frac{\sum(\log P_o - \log P_e)^2}{\sum(\log P_o - \log \overline{P_o})^2}$$

- ∞< E < 1

Elog dá mais peso a valores baixos. (Por que?)

E = 0 se EMQ é igual à variância dos dados observados

E < 0 se EMQ é maior que a var. dos dados observados (modelo pior que estimar pela média)

E > 0 se EMQ é menor que a var. dos dados observados (modelo MELHOR que estimar pela média)

Exemplo: E e Elog

Table 1. Nash-Sutcliffe efficiency in gaps, before and after the adjustment of model-generated inserts.

Series No.	*R2* Before	*R2* After	*R2log* Before	*R2log* After
45	0.67	0.46	0.74	0.64
83	0.43	0.22	0.58	0.55
109	0.78	0.63	0.84	0.78
135	0.62	0.58	0.74	0.75
148	0.56	0.50	0.57	0.61
164	0.84	0.77	0.91	0.92
205	0.76	0.56	0.83	0.56
237	0.58	0.72	0.65	0.80
265	0.67	0.53	0.69	0.48
365	0.74	0.27	0.72	0.43

Variações de E

- Coeficiente de Persistência

$$CP = 1 - \frac{\sum (Po_i - P\boldsymbol{e}_i)^2}{\sum (Po_i - Po_{i-1})^2}$$

- Compara a resultado do modelo com um modelo que sempre fornece o último valor ocorrido como previsão (persistência)

E = 0 se o modelo é tão bom quanto usar o último valor ocorrido como previsão (persistência)

E < 0 se o modelo é pior que persistência

E > 0 se o modelo é MELHOR que persistência

Variações de E

- Coeficiente de eficiência E'

$$E' = 1 - \frac{\sum \left| P_o - P_e \right|}{\sum \left| P_o - \overline{P_o'} \right|}$$

- Onde $\overline{P_o'}$ é qualquer parâmetro com o qual você queira comparar o seu modelo.

e.g. Média sazonal, resultado de outro modelo, etc...

E = 0 se o modelo é tão bom quanto usar o objeto de comparação

E < 0 se o modelo é pior que o objeto de comparação

E > 0 se o modelo é MELHOR que o objeto de comparação

Exemplo

Forecast from	Equation	E	R2
Oct	$L_{JAN}^{t}=0.55\ L_{OUT}^{t-1}+25.3\ SAM_{AUG}^{t-1}+48.7$	0.05	0.71
Sept	$L_{JAN}^{t}=0.48\ L_{SET}^{t-1}+22.6\ SAM_{AUG}^{t-1}+32.8$	0.29	0.72
Oct	$L_{JUN}^{t}= -0.03L_{OUT}^{t-1}+12.7SAM_{Aug}^{t-1} + 0.81L_{JUN}^{t-1}+86{,}673$	0.10	0.74
Set - Aug	$L_{JUN}^{t}= -0.79\ L_{JUN}^{t-1}+13.03SAM_{Aug}^{t-1} +88.1$	0.12	0.74

Índice de Concordância d

$$d = 1 - \frac{\sum (P_o - P_e)^2}{\sum \left(\left| P_e - \overline{P_o} \right| + \left| P_o - \overline{P_o} \right| \right)^2}$$

- 0 <= d <= 1
- d= 1- (REMQ/erro potencial)*n

  erro potencial é o maior valor que $(P_o\text{-}P_e)^2$ pode ter para cada par estimado-observado

Sensível a valores extremos devido ao quadrado da diferença

Para se estabelecer a qualidade do seu modelo

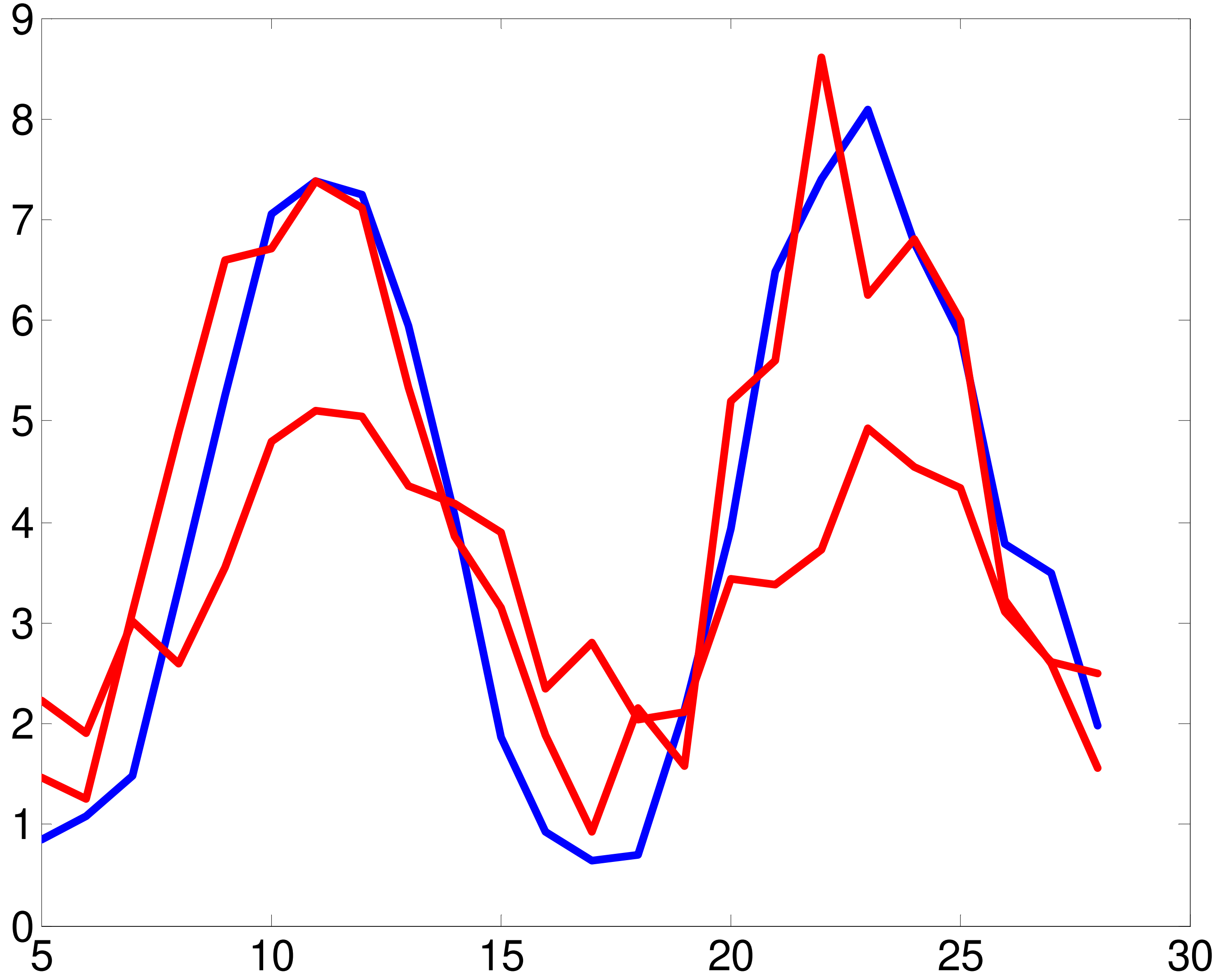
- Necessário observar pelo menos uma medida de erro relativo (E, d, E´, R2, r, etc) e uma medida de erro absoluto EMQ, EMA, EPA, etc.

- Além disso é necessário verificar média e desvio padrão.

Exemplo: Precipitação média sazonal no SE do Brasil

trimes	ncdc	dp	ncep	dp	cola	dp	regcm	dp
5	1.06	0.70	0.86	0.84	1.47	1.00	2.24	1.46
6	1.76	0.97	1.09	0.69	1.26	0.82	1.91	1.38
7	2.85	1.44	1.48	0.61	3.12	1.22	3.02	1.54
8	4.13	0.91	3.35	0.87	4.90	1.32	2.59	2.13
9	4.93	0.65	5.26	2.05	6.59	0.99	3.56	2.71
10	5.15	1.07	7.05	2.99	6.72	0.98	4.80	3.75
11	5.87	1.27	7.38	3.34	7.39	2.03	5.10	3.68
12	5.67	1.25	7.24	2.64	7.11	1.87	5.04	3.44
13	4.68	0.92	5.95	1.33	5.34	1.28	4.36	2.66
14	2.69	0.58	4.06	0.72	3.85	1.99	4.18	4.83
15	1.13	0.32	1.87	0.82	3.15	2.30	3.89	4.09
16	0.76	0.30	0.92	0.77	1.89	1.36	2.34	2.11
17	0.80	0.42	0.65	0.56	0.92	0.77	2.80	2.74
18	1.57	0.68	0.69	0.45	2.15	1.06	2.04	1.17
19	2.65	0.52	2.14	0.62	1.58	0.53	2.12	1.71
20	4.29	0.66	3.93	1.34	5.20	1.32	3.43	1.95
21	5.93	1.11	6.49	2.61	5.60	1.06	3.38	3.75
22	8.07	1.57	7.41	2.93	8.60	2.67	3.72	3.77
23	7.91	1.38	8.10	3.34	6.25	2.07	4.94	3.68
24	6.25	1.11	6.76	2.74	6.80	1.71	4.54	4.31
25	3.43	0.96	5.85	1.80	6.01	1.47	4.33	3.29
26	2.64	0.91	3.79	0.77	3.22	1.98	3.11	2.43
27	2.38	0.98	3.50	0.53	2.60	1.41	2.62	3.24
28	2.14	1.11	1.99	0.84	1.57	1.23	2.5	3.28

Observado estimado



Para se comparar modelos muito diferentes

- Se a correlação entre valores estimados e observados é 0.9 para um modelos com 6 parâmetros e também para um modelo com 60 parâmetros, não se pode concluir que esses modelos têm performance semelhante. O primeiro modelo explica muito mais que o segundo.
- Então se deve-se corrigir o valor do coeficiente de Nash-Sutcliffe pelo número de parâmetros do modelo

$$E = 1 - \frac{\sum \frac{(Po_i - Pc_i)^2}{N - p}}{\sum \frac{(Po_i - \overline{Po})^2}{N - 1}}$$

- N = número de observações
- p = número de parâmetros