



Estatística de Mallows na seleção de modelos de predição da precipitação média mensal e anual no Rio Grande do Sul¹

Mallows statistic in the selection of models to predict the monthly and annual average rainfall in Rio Grande do Sul, Brazil

**Claudia Fernanda Almeida Teixeira^{2*}, Rita de Cássia Fraga Damé², Pamela Bilhafan Disconzi³,
Marília Alves Brito Pinto⁴, Antonyony Severo Winkler³, Jacira Porto dos Santos³**

Resumo - A estatística Cp de Mallows pode ser utilizada na seleção de melhores subconjuntos na modelagem hidrológica, principalmente nos casos em que são utilizadas muitas variáveis. Com base no fato de que há, em muitos casos, o interesse em estimar a precipitação média mensal e anual baseada nas coordenadas geográficas latitude e longitude, e altitude, objetivou-se com este trabalho verificar o ganho de informação quando é aplicada a estatística Cp de Mallows na seleção dos melhores subconjuntos da regressão linear múltipla, para a predição da precipitação de alguns municípios localizados no estado do Rio Grande do Sul. Para tanto foram utilizados dados de precipitação diária de 26 estações climatológicas, além de outras sete, utilizadas para a validação dos modelos lineares propostos, pertencentes a sete mesorregiões do Rio Grande do Sul. Após a constituição das séries, os valores de precipitação foram ajustados a partir de modelos lineares, utilizando a regressão linear múltipla, em que a variável dependente foi a precipitação e as variáveis independentes, as coordenadas geográficas latitude e longitude, e a altitude. A estatística Cp foi aplicada na seleção dos conjuntos e, na sequência, aplicados os índices estatísticos erro quadrático médio, erro padrão de predição, fator de viés e fator de precisão para comparação entre os valores observados de precipitação versus os preditos. A partir dos resultados obtidos pode-se concluir que, sob o ponto de vista da parcimônia, a estatística proposta por Mallows mostrou-se adequada na seleção de modelos de predição da precipitação mensal e anual das estações analisadas.

Palavras-chave - Modelagem hidrológica. Regressão linear múltipla. Seleção dos subconjuntos.

Abstract - The Mallows Cp statistic can be used in the selection of the best subsets in hydrological modeling, especially in cases where many variables are used. Because there are, in many cases, the interest in estimating the monthly and annual average rainfall based on geographic coordinates of latitude and longitude, and altitude. Consequently, the aim of this study was to verify the information gain when applied to statistical Cp Mallows in the selection of the best subsets of multiple linear regression to predict the precipitation of some municipalities in the state of Rio Grande do Sul. Daily precipitation data from 26 meteorological stations, in addition to seven others, used to validation of the proposed linear models, belonging to seven mesoregions of Rio Grande do Sul were collected and analyzed. After the formation of the series, precipitation values were adjusted from linear models, using multiple linear regression in which the dependent variable was the precipitation and independent variables, the geographic coordinates of latitude and longitude, and altitude. The Cp statistic was used in the selection of sets and, subsequently applied statistical indexes mean square error, standard error of prediction bias factor were used to obtain the accuracy factor for comparison between observed versus predicted precipitation. From the results obtained it can be concluded that, from the point of view of parsimony, the statistic proposed by Mallows proved adequate in the selection of models for prediction of monthly and annual rainfall of the stations analyzed.

Key words - Hydrologic modeling. Multiple linear regression. Selection of subsets.

*Autor para correspondência

¹Recebido para publicação em 18/09/2013 e aprovado em 10/07/2013

²Professor do Centro de Engenharias, Universidade Federal de Pelotas - UFPel, Pelotas - RS, cfeixei@ig.com.br; rita2o@hotmail.com

³Mestrando em Manejo e Conservação do Solo e da Água/MACSA, Universidade Federal de Pelotas - UFPel, Pelotas - RS, pamelabilhafan@yahoo.com.br; antonionysw@hotmail.com, portojacira@gmail.com

⁴Mestrando em Manejo e Conservação do Solo e da Água/MACSA, Universidade Federal de Pelotas - UFPel, Pelotas - RS.

Introdução

A precipitação pluvial é considerada um dos elementos que mais interfere na produção agrícola (SILVA *et al.*, 2009; PRELA-PANTANO, 2011), ocasionando, devido às suas características de alta variabilidade geográfica, temporal e sazonal, riscos na estimativa errônea daquela variável. A regionalização da precipitação exige uma análise abrangente, tanto no tempo quanto no espaço, cujo tratamento dos dados, para a grande maioria dos problemas hidrológicos, é estatístico.

Para Montgomery e Runger (2012), a análise de regressão é uma técnica estatística capaz de modelar e investigar a relação entre duas ou mais variáveis, sendo um modelo com mais de uma variável regressora, chamado de regressão múltipla. Mello e Silva (2009) ajustaram equações lineares de regressão múltipla para a estimativa dos totais médios precipitados em função das coordenadas geográficas e da altitude para o estado de Minas Gerais. Os autores utilizaram o coeficiente de determinação ajustado para a seleção do melhor modelo encontrado, ou seja, consideraram o referido coeficiente para medida da qualidade do ajuste do modelo proposto.

Segundo Pimenta *et al.* (2008) em estudos com modelos lineares, é necessário explorar as variáveis independentes para realizar estudos sobre a variável dependente, sendo algumas removidas seletivamente, com objetivo de reduzir seu número na aplicação do modelo. Isto porque, um modelo de regressão com muitas variáveis independentes possui elevados custos computacional e de atualização do modelo, dificultando seu uso e análise. Há, portanto, que se considerar que o número de variáveis utilizadas na construção de um modelo deva possibilitar a descrição, o controle e a predição adequada. Neste contexto, Levine *et al.* (2011) consideram, ainda, que deve ser utilizado o princípio da parcimônia, ou seja, deve-se selecionar o modelo mais simples e que cumpra sua função adequadamente.

Para a seleção adequada do modelo, existem na literatura (MONTGOMERY; RUNGER, 2012; LEVINE *et al.*, 2011) alguns critérios quantitativos, tais como o coeficiente de determinação (R^2), o coeficiente de determinação ajustado (R^2_{ajustado}), o quadrado médio dos resíduos (QMR) e a estatística de Mallows (C_p). Floriano *et al.* (2006) com o objetivo de ajustar e selecionar modelos tradicionais para série temporal de dados de altura de árvores utilizaram o coeficiente de determinação (R^2), a estatística C_p de Mallows, o critério de informação de Akaike (Akaike's information criterion – AIC), o quadrado médio dos resíduos (QMR) e a análise gráfica de resíduos para comparação entre os modelos.

A estatística de Mallows (C_p) (MALLOWS, 1975; 1995) é utilizada (IAFFE *et al.*, 2000; FLORIANO *et*

al., 2006; PIMENTA *et al.*, 2008; OLIVEIRA, 2010) em modelos onde o valor da mesma seja pequeno e próximo de p (número de variáveis regressoras no modelo). Um valor baixo de C_p indica que o modelo é relativamente preciso e possui variância pequena na estimativa verdadeira dos coeficientes da regressão e na previsão de respostas futuras. Assim como, modelos com baixa capacidade preditiva e tendência, têm valor de C_p maior do que p (MONTGOMERY; RUNGER, 2012). Segundo Hair Júnior *et al.* (1998), a estatística C_p é delineada contra o número de parâmetros (p), ou seja, quanto mais próximo for C_p de p , menos tendenciosas são as estimativas dos parâmetros e melhor é o modelo.

Com base no fato de que as variáveis, altitude, latitude e longitude influenciam as características climáticas de uma região, objetivou-se com o presente trabalho aplicar a estatística C_p de Mallows com o intuito de selecionar os melhores subconjuntos, quando da modelagem dos valores de precipitação média mensal e anual para 26 municípios localizados no estado do Rio Grande do Sul.

Material e métodos

O presente trabalho foi realizado a partir dos dados de precipitação diária de 26 estações climatológicas obtidos do banco de dados da Agência Nacional de Águas (ANA, 2012), pertencentes a sete mesorregiões do estado do Rio Grande do Sul. Além destas foram consideradas outras sete estações que foram utilizadas para a validação dos modelos lineares propostos. Na Figura 1 encontra-se a localização de todas as estações pluviométricas utilizadas no presente trabalho e nas Tabelas 1 e 2, a caracterização das mesmas.

A partir do banco de dados de precipitação, expressos em altura de lâmina d'água (mm), foram constituídas as séries média mensal e anual, para o período de 1913 a 2009 (N = 97 anos). As coordenadas geográficas latitude e longitude, em graus, e altitude, em m, de cada estação climatológica também foram obtidas diretamente da base de dados da ANA. Após a constituição das séries, os valores de precipitação foram ajustados a partir de modelos lineares, utilizando a regressão múltipla, em que a variável dependente foi a precipitação e as variáveis independentes, as coordenadas geográficas latitude e longitude, e a altitude. Para tanto foi utilizado o software “Statistical Package for Social Science for Windows” (SPSS), considerando, inicialmente, todas as três variáveis independentes e posteriormente, mediante a abordagem dos melhores subconjuntos de modelos para um determinado número de variáveis independentes (NETER *et al.*, 1996), utilizando o software PHStat (LEVINE *et al.*, 1999).



Figura 1 – Localização das 33 estações pluviométricas nas sete mesorregiões do estado do Rio Grande do Sul ANA (2012)

A estatística C_p (MALLOWS, 1973; 1995) definida na Equação (1) mede as diferenças entre os valores obtidos mediante um modelo de regressão ajustado e um observado. Quando um modelo de regressão com k variáveis independentes contém somente diferenças aleatórias em relação a um modelo real, o valor médio de C_p é $p + 1$, o número de parâmetros. Por conseguinte, ao avaliar muitos modelos de regressão alternativos, o objetivo é encontrar modelos cuja C_p esteja próxima ou inferior a $(p + 1)$.

$$C_p = \frac{(1 - r_p^2)(n - T)}{1 - r_T^2} - (n - 2(p + 1)) \quad (1)$$

em que p é o número de variáveis independentes em um modelo de regressão; T é o número total de parâmetros (incluindo a interseção) a serem estimados no modelo de regressão completo; r_p^2 é o coeficiente de determinação múltipla para um modelo de regressão que possua p variáveis independentes; r_T^2 é o coeficiente de determinação múltipla para um modelo de regressão completo que contenha todos os T parâmetros estimados.

Para comparar os dados estimados pelos modelos obtidos por regressão linear múltipla e pela estatística C_p com os observados, a partir das sete estações consideradas na validação, foram utilizados os seguintes índices de

Tabela 1 - Caracterização das 26 estações de precipitação localizadas nas sete mesorregiões do estado do Rio Grande do Sul, utilizadas para a modelagem da precipitação

| Estação | Nome | Latitude | Longitude | Altitude (m) |
|---------|----------------------------|------------|------------|--------------|
| 3154001 | Bagé | 31°18'17'' | 54°07'11'' | 230 |
| 2850001 | Bom Jesus | 28°40'07'' | 50°26'31'' | 1048 |
| 3052002 | Cachoeira do Sul | 30°02'00'' | 52°53'00'' | 73 |
| 2951008 | Caxias do Sul | 29°11'39'' | 51°11'09'' | 760 |
| 2835005 | Cruz Alta | 28°37'28'' | 53°36'12'' | 432 |
| 3052005 | Encruzilhada do Sul | 30°32'37'' | 52°31'25'' | 428 |
| 2853025 | Ijuí | 28°23'17'' | 53°54'50'' | 448 |
| 2753003 | Iraí | 27°11'20'' | 53°15'12'' | 247 |
| 2953011 | Júlio de Castilhos | 29°13'00'' | 53°40'00'' | 516 |
| 2851014 | Lagoa Vermelha | 28°13'19'' | 51°30'45'' | 842 |
| 2852020 | Passo Fundo | 28°13'33'' | 52°24'12'' | 684 |
| 3152014 | Pelotas | 31°45'00'' | 52°21'00'' | 13 |
| 3051011 | Porto Alegre | 30°03'13'' | 51°10'24'' | 47 |
| 3252020 | Rio Grande | 32°04'44'' | 52°10'00'' | 3 |
| 2953017 | Santa Maria | 29°43'27'' | 53°43'12'' | 95 |
| 2754005 | Santa Rosa | 27°51'00'' | 54°25'00'' | 360 |
| 3353007 | Santa Vitória do Palmar | 33°31'57'' | 53°20'58'' | 24 |
| 3055001 | Santana do Livramento | 30°53'00'' | 55°32'00'' | 328 |
| 2856005 | São Borja | 28°39'44'' | 56°00'44'' | 83 |
| 3054011 | São Gabriel | 30°20'00'' | 54°19'00'' | 124 |
| 2854011 | São Luiz Gonzaga | 28°24'00'' | 54°58'00'' | 245 |
| 2951030 | Taquari | 29°48'00'' | 51°49'00'' | 76 |
| 2949002 | Torres | 29°20'41'' | 49°43'41'' | 31 |
| 2957001 | Uruguaiana | 29°45'00'' | 57°05'00'' | 62 |
| 2850011 | Vacaria | 28°31'00'' | 50°57'00'' | 1050 |
| 2851042 | Veranópolis | 28°56'14'' | 51°33'11'' | 705 |

Tabela 2 - Caracterização das estações de precipitação utilizadas para validação das equações geradas, as quais representam as sete mesorregiões do estado do Rio Grande do Sul

| Mesorregião | Estação | Nome | Latitude | Longitude | Altitude (m) |
|------------------|---------|-------------------------|------------|------------|--------------|
| Centro Ocidental | 2954001 | Cacequi | 29°52'43'' | 54°49'31'' | 100 |
| Centro Oriental | 2955012 | Santa Cruz do Sul | 29°43'00'' | 56°26'00'' | 53 |
| Metropolitana | 2950021 | Osório | 29°53'00'' | 56°16'00'' | 10 |
| Nordeste | 2951003 | Bento Gonçalves | 29°09'00'' | 51°31'00'' | 640 |
| Noroeste | 2753005 | Palmeira das Missões | 27°53'00'' | 53°26'00'' | 634 |
| Sudeste | 3053002 | Caçapava do Sul | 30°30'00'' | 53°29'00'' | 450 |
| Sudoeste | 3053002 | Alegrete | 29°47'00'' | 55°46'00'' | 121 |

desempenho estatístico: erro quadrático médio (EQM), erro padrão de predição (EPP), fator de viés (FV) e fator de precisão (FP) (ROSS, 1996).

Resultados e discussão

As estatísticas descritivas dos valores médios de precipitação mensal e anual para as 26 estações localizadas no estado do Rio Grande do Sul, considerando 97 anos de informação (1913 a 2009), são apresentadas na Tabela 3. Os valores médios de precipitação mensal revelam que não é possível definir com clareza os meses do ano correspondentes às estações seca e chuvosa, evidenciando que na região sul, a distribuição da precipitação mostra-se regular, oscilando entre 115,28 e 153,36 mm, para os meses de novembro e setembro, respectivamente. Blain *et al.* (2009) analisaram a variabilidade das séries mensais de precipitação pluvial, nas regiões de Campinas-SP e Pelotas-RS e verificaram que ao contrário do observado na região de Campinas-SP, na região de Pelotas-RS, as chuvas são distribuídas de forma semelhante ao longo do ano.

Os coeficientes de variação (CV) da precipitação média mensal e anual mostram que a variabilidade desse atributo é considerada média ($10\% < C.V. < 20\%$) (PIMENTEL-GOMES, 1995), com valores que variam de 11,33 a 18,99%, para o mês de fevereiro e dezembro,

respectivamente (Tabela 3). Silva *et al.* (2003), com o objetivo de estudar o comportamento temporal das chuvas de Uberaba/MG, utilizando a série das precipitações pluviais mensais e anuais de 1914 a 2000, encontraram valores de coeficiente de variação relativamente altos, principalmente na estação seca, com valores de 61,37 e 161,08%. Os autores associaram os altos valores à ausência total de chuvas em alguns anos da série, no período da seca, o que indica que estimativas feitas com a média aritmética podem apresentar precisão e confiabilidade duvidosas, pelo fato de essa medida de posição não ser a mais adequada para representar a variável, pois é altamente influenciada por valores extremos.

Com o objetivo de verificar a possibilidade de utilização das variáveis latitude, longitude e altitude, na predição dos valores das precipitações média mensal e anual, são apresentados na Tabela 4, os coeficientes de correlação de Pearson para as 26 estações estudadas. Verifica-se que para a variável latitude, os valores de coeficiente de correlação apresentaram-se significativos e altamente significativos para todos os períodos analisados, à exceção do mês de julho, evidenciando que para este período a altitude não influenciou na predição da precipitação mensal. Para os meses de fevereiro, março, outubro e novembro, os coeficientes mostraram-se não significativos para as variáveis longitude e altitude, representando que os valores de precipitação média mensal deverão ser estimados considerando apenas a latitude. Das três variáveis analisadas, a longitude e a altitude parecem

Tabela 3 - Estatísticas básicas dos valores médios de precipitação mensal, sazonal e anual para as 26 estações localizadas no estado do Rio Grande do Sul, para o período de 1913 a 2009

| Período | Média (mm) | D.P. | C.V. (%) | L.I. (95%) | L.S. (95%) | Mediana (mm) | Mínimo (mm) | Máximo (mm) |
|---------|------------|--------|----------|------------|------------|--------------|-------------|-------------|
| Jan | 131,41 | 18,82 | 14,32 | 124,18 | 138,65 | 127,11 | 94,86 | 163,09 |
| Fev | 128,26 | 14,53 | 11,33 | 122,67 | 133,84 | 124,94 | 101,41 | 159,77 |
| Mar | 123,65 | 15,04 | 12,16 | 117,87 | 129,43 | 120,16 | 95,21 | 155,36 |
| Abr | 129,46 | 24,25 | 18,73 | 120,14 | 138,78 | 125,65 | 94,22 | 180,73 |
| Mai | 127,24 | 17,94 | 14,10 | 120,34 | 134,13 | 140,19 | 97,96 | 157,40 |
| Jun | 137,24 | 21,50 | 15,66 | 128,98 | 145,50 | 129,86 | 90,12 | 167,12 |
| Jul | 127,73 | 18,92 | 14,81 | 120,46 | 135,00 | 126,45 | 75,95 | 152,76 |
| Ago | 127,55 | 20,01 | 15,69 | 119,86 | 135,24 | 126,49 | 70,27 | 159,81 |
| Set | 153,36 | 22,50 | 14,67 | 144,72 | 162,01 | 157,49 | 105,81 | 192,51 |
| Out | 151,02 | 28,35 | 18,77 | 140,13 | 161,92 | 153,14 | 92,89 | 197,50 |
| Nov | 115,28 | 20,23 | 17,55 | 107,50 | 123,05 | 114,35 | 76,80 | 144,20 |
| Dez | 120,86 | 22,95 | 18,99 | 112,04 | 129,69 | 124,31 | 68,57 | 148,62 |
| Anual | 1573,07 | 179,53 | 11,41 | 1504,06 | 1642,07 | 1583,06 | 1218,94 | 1818,02 |

D.P.: desvio padrão; C.V.: coeficiente de variação; LI (95%): limite inferior do intervalo de confiança para média com 95% de confiança; LS (95%): limite superior do intervalo de confiança para média com 95% de confiança; Md: Mediana; Mínimo: valor mínimo; Máximo: valor máximo.

ser as que menos influenciam na estimativa dos valores de precipitação. Estes resultados podem ser devido à ausência de chuvas orográficas na região estudada. Esta suposição baseia-se nos resultados de Mello e Silva (2009), que

encontraram, em contrapartida, que a latitude e a altitude proporcionaram os maiores valores de coeficientes, refletindo as características climáticas e a influência orográfica no comportamento das precipitações da região.

Tabela 4 - Valores de coeficiente de correlação de Pearson dos valores de precipitação média mensal e anual em função das coordenadas geográficas latitude e longitude, e altitude, para as 26 estações localizadas no Rio Grande do Sul, no período analisado

| Período | Coeficiente de Correlação de Pearson | | |
|---------|--------------------------------------|----------------------|----------------------|
| | Latitude | Longitude | Altitude |
| Jan | 0,853** | 0,121 ^{ns} | 0,571** |
| Fev | 0,492* | 0,121 ^{ns} | 0,273 ^{ns} |
| Mar | 0,411* | -0,358 ^{ns} | -0,007 ^{ns} |
| Abr | 0,473* | -0,694** | -0,015 ^{ns} |
| Mai | 0,660** | -0,460* | 0,183 ^{ns} |
| Jun | 0,512** | 0,158 ^{ns} | 0,428* |
| Jul | 0,244 ^{ns} | 0,574** | 0,481* |
| Ago | 0,411* | 0,729** | 0,692** |
| Set | 0,710** | 0,417* | 0,715** |
| Out | 0,873** | -0,297 ^{ns} | 0,386 ^{ns} |
| Nov | 0,722** | -0,274 ^{ns} | 0,309 ^{ns} |
| Dez | 0,858** | -0,098 ^{ns} | 0,583** |
| Anual | 0,844** | -0,025 ^{ns} | 0,537** |

**Significativo a 1% de probabilidade de erro; *significativo a 5% de probabilidade de erro; ns: não significativo.

As equações de regressão linear múltipla e o coeficiente de determinação ajustado ($R^2_{ajustado}$) para a estimativa dos valores de precipitação média mensal e anual de 26 estações localizadas no Rio Grande do Sul são apresentados na Tabela 5.

Com o objetivo de representar a variável dependente de interesse, tem-se buscado desenvolver modelos parcimoniosos, ou seja, aqueles em que há o menor número de parâmetros (MORETTIN; TOLOI, 2006). Como forma de selecionar qual o modelo mais adequado na estimativa da variável dependente, pode-se, dentre outros (MONTGOMERY; RUNGER, 2012), utilizar como critério, àquele que leva em conta a estatística C_p de Mallows ou os valores de $R^2_{ajustado}$. Considerando a abordagem dos melhores subconjuntos para a construção de modelos, aplicada para a estimativa dos valores de precipitação, são apresentados na Tabela 6, os valores de C_p , p , $R^2_{ajustado}$ e desvio padrão. Esta é uma estatística de grande utilidade, e alguns estudos (FLORIANO *et al.*, 2006; COLET *et al.*, 2007; PIMENTA *et al.*, 2008) a indicam como de maior sensibilidade para perceber mudanças na adequabilidade do modelo para predição em relação ao coeficiente de determinação e à estatística F (HOCKING, 1976).

Observa-se que para o mês de janeiro, o $R^2_{ajustado}$ atingiu valor máximo de 0,734 (Tabela 6), quando a

Tabela 5 - Equações de regressão linear múltipla e coeficiente de determinação ajustado ($R^2_{ajustado}$) para estimativa dos valores de precipitação média mensal e anual de 26 estações localizadas no Rio Grande do Sul

| Período | Equação | $R^2_{ajustado}$ |
|---------|---|------------------|
| Jan | $P = 426,549 - 9,898LAT - 0,113LON + 0,011ALT$ | 0,722 |
| Fev | $P = 321,407 - 4,989LAT - 0,854LON + 0,0001ALT$ | 0,150 |
| Mar | $P = 125,057 - 5,066LAT + 2,836LON - 0,005ALT$ | 0,227 |
| Abr | $P = -154,797 - 8,3LAT + 9,979LON + 0,002ALT$ | 0,704 |
| Mai | $P = 89,777 - 8,130LAT + 5,222LON + 0,003ALT$ | 0,636 |
| Jun | $P = 351,122 - 6,104LAT - 0,713LON + 0,013ALT$ | 0,210 |
| Jul | $P = 428,868 - 1,309LAT - 5,041LON + 0,014ALT$ | 0,332 |
| Ago | $P = 548,054 - 2,816LAT - 6,503LON + 0,022ALT$ | 0,720 |
| Set | $P = 554,59 - 8,123LAT - 3,193LON + 0,025ALT$ | 0,692 |
| Out | $P = 301,380 - 16,293LAT + 6,186LON + 0,011ALT$ | 0,871 |
| Nov | $P = 186,808 - 9,629LAT + 3,983LON + 0,007ALT$ | 0,572 |
| Dez | $P = 260,468 - 11,159LAT + 3,439LON + 0,024ALT$ | 0,800 |
| Anual | $P = 3439,253 - 91,814LAT + 15,229LON + 0,127ALT$ | 0,714 |

*P: precipitação (mm); LAT: latitude (graus decimais); LON: longitude (graus decimais); ALT: altitude (m).

Tabela 6 - Resultados da estatística C_p , valores de p , equações de regressão, coeficiente de determinação ajustado e desvio padrão considerando as coordenadas geográficas latitude e longitude, e altitude para a estimativa da precipitação média mensal e anual, para as 26 estações localizadas no Rio Grande do Sul

| Período | C_p | $(p + 1)^*$ | Equação | $R^2_{ajustado}$ | D.P. |
|-----------|-------|-------------|---|------------------|-------|
| Janeiro | 2,01 | 3 | $P = 48,1317 + 1,3363LON + 0,036ALT$ | 0,734 | 9,70 |
| Fevereiro | 0,29 | 2 | $P = 274,91 - 4,9554LAT$ | 0,210 | 12,91 |
| Março | 2,27 | 3 | $P = 85,81 - 4,4741LAT + 3,2105LON$ | 0,252 | 13,01 |
| Abril | 2,04 | 3 | $P = -139,7294 - 8,5271LAT + 9,8350LON$ | 0,716 | 12,93 |
| Mai | 2,15 | 3 | $P = 113,9725 - 8,4948LAT + 4,9908LON$ | 0,650 | 10,62 |
| Junho | 1,35 | 2 | $P = 363,2624 - 7,6374LAT$ | 0,232 | 18,84 |
| Julho | 2,27 | 3 | $P = 378,8402 - 4,8492LON + 0,0176ALT$ | 0,353 | 15,21 |
| Agosto | 4,00 | 4 | $P = 548,0541 - 2,8159LAT - 6,5031LON + 0,0224ALT$ | 0,720 | 10,59 |
| Setembro | 4,00 | 4 | $P = 554,5896 - 8,1231LAT - 3,1928LON + 0,0248ALT$ | 0,692 | 12,49 |
| Outubro | 4,00 | 4 | $P = 301,3799 - 16,2929LAT + 6,1862LON + 0,0110ALT$ | 0,871 | 10,16 |
| Novembro | 2,39 | 3 | $P = 233,7367 - 10,3373LAT + 3,5351LON$ | 0,584 | 13,06 |
| Dezembro | 4,00 | 4 | $P = 260,4684 - 11,1588LAT + 3,4394LON + 0,024ALT$ | 0,800 | 10,25 |
| Anual | 4,00 | 4 | $P = 3439,2722 - 91,8147LAT + 15,2293LON + 0,1267ALT$ | 0,714 | 95,97 |

* p variáveis independentes

variável latitude foi excluída, e quando são consideradas todas as variáveis independentes, seu valor é alterado para 0,722 (Tabela 5). Observa-se que embora tenha ocorrido um acréscimo no valor de $R^2_{ajustado}$, este é pequeno. O mesmo comportamento de $R^2_{ajustado}$ ocorreu nos meses de abril e novembro, quando foi excluída a variável altitude. Mesmo assim, o princípio da parcimônia (MORETTIN; TOLOI, 2006) permite que sejam selecionadas as equações apresentadas na Tabela 6, sob o critério do valor de $R^2_{ajustado}$.

Considerando a abordagem dos melhores subconjuntos para a construção de modelos, utilizando a estatística desenvolvida por Mallows (1973, 1995), são apresentados na Tabela 6, os valores de C_p e $(p+1)$. Verifica-se que nos meses de agosto, setembro, outubro e dezembro, e na escala anual, os valores de C_p são iguais aos de $(p + 1)$, indicando que quando são consideradas as três variáveis independentes e mais o termo relativo ao intercepto, estes são os modelos que devem ser selecionados. À medida que são excluídas uma das variáveis independentes, os valores de C_p são menores do

que p , por exemplo, quando da exclusão da altitude, para os meses de março, abril, maio e novembro, observando-se também diminuição dos valores do $R^2_{ajustado}$.

Para os meses de janeiro e julho, quando é excluída a latitude, os valores de $R^2_{ajustado}$ e C_p também diminuiram. Quando o valor de C_p é pequeno, significa que a predição do modelo apresenta pequenas variações para estimar os verdadeiros coeficientes de regressão e prever futuras respostas. Apesar de nos meses de fevereiro e junho, em que é mantida apenas a variável latitude, com baixíssimos valores de $R^2_{ajustado}$, há que se considerar: (i) a utilização de modelos alternativos parcimoniosos; (ii) seleção do modelo que melhor representa o comportamento da variável de interesse e; (iii) obtenção dos valores simulados próximos aos observados.

Nesse sentido, quanto aos índices estatísticos, a Tabela 7 mostra de maneira geral, o bom desempenho dos modelos considerando os valores estimados e os observados, para as escalas mensal e anual da precipitação. Entretanto, verifica-se que os índices estatísticos (ROSS,

Tabela 7 - Índices de desempenho estatístico utilizados para comparar os dados de precipitação observados (Obs.) e os estimados pelos modelos de regressão linear múltipla (Reg. Múlt.) e pela estatística C_p (Est. C_p), considerando as coordenadas geográficas latitude e longitude, e altitude para as estações de precipitação utilizadas para validação das equações geradas

| Período | Obs. versus Reg. Múlt. | | | | Obs. versus Est. C_p | | | |
|-----------|------------------------|-------|------|------|------------------------|-------|------|------|
| | EQM* | EPP | FV | FP | EQM | EPP | FV | FP |
| Janeiro | 25,80 | 17,78 | 0,91 | 2,07 | 29,32 | 20,21 | 0,90 | 2,16 |
| Fevereiro | 18,93 | 14,43 | 0,99 | 1,95 | 20,19 | 15,39 | 0,99 | 2,03 |
| Março | 22,15 | 16,48 | 0,95 | 2,59 | 22,68 | 16,88 | 0,94 | 2,59 |
| Abril | 20,79 | 15,76 | 1,03 | 1,99 | 20,90 | 15,84 | 1,03 | 2,03 |
| Mai | 13,24 | 10,00 | 1,00 | 1,87 | 13,78 | 10,41 | 1,00 | 1,92 |
| Junho | 19,18 | 13,45 | 0,97 | 2,20 | 21,70 | 15,23 | 0,98 | 2,63 |
| Julho | 18,64 | 14,14 | 0,94 | 2,47 | 18,27 | 13,86 | 0,94 | 2,43 |
| Agosto | 20,45 | 15,56 | 0,93 | 2,38 | 20,40 | 15,52 | 0,94 | 2,37 |
| Setembro | 18,83 | 11,40 | 0,91 | 1,92 | 18,87 | 11,43 | 0,91 | 1,93 |
| Outubro | 14,96 | 9,62 | 1,00 | 1,60 | 14,96 | 9,62 | 1,00 | 1,60 |
| Novembro | 6,52 | 5,55 | 1,00 | 1,33 | 5,71 | 4,86 | 1,00 | 1,27 |
| Dezembro | 19,97 | 15,65 | 0,96 | 1,85 | 19,97 | 15,65 | 0,96 | 1,85 |
| Anual | 104,98 | 6,38 | 0,96 | 1,45 | 105,05 | 6,38 | 0,96 | 1,45 |

*EQM: Erro Quadrático Médio; EPP: Erro Padrão de Predição; FV: Fator de viés; FP: Fator de Precisão.

1996) mostraram-se mais adequados dentro dos intervalos ($< EQM$; $< EPP$; $FV = 1$ e $FP = 1$), quando foram considerados os valores observados e os estimados por meio da regressão múltipla, comparativamente aos obtidos via estatística C_p de Mallows.

Conclusões

A estatística C_p de Mallows é uma ferramenta capaz de selecionar de forma robusta o modelo de predição de precipitação mensal e anual para as 26 estações climatológicas localizadas no Rio Grande do Sul.

Sob o ponto de vista da parcimônia para a estimativa dos valores de precipitação mensal e anual, a estatística de proposta por Mallows mostrou-se adequada.

Literatura científica citada

AGÊNCIA NACIONAL DE ÁGUAS (Brasil) (ANA). HidroWeb: sistemas de informações hidrológicas. Disponível em: <http://hidroweb.ana.gov.br/HidroWeb>. Acesso em: 05 janeiro 2012.

BLAIN, G. C.; KAYANO, M. T.; CAMARGO, M. B. P.; LULU, J. Variabilidade amostral das séries mensais de precipitação pluvial em duas regiões do Brasil: Pelotas-RS e Campinas-SP. *Revista Brasileira de Meteorologia*, v.24, n.1, p.1-11, 2009.

COLET, M. J.; WEIRICH NETO, P. H.; GARBUIO, W.; SCHIMANDEIRO, A. Processo de semeadura e o rendimento do milho na região dos Campos Gerais do Paraná. *Ciência Rural*, v.37, n.4, p.994-999, 2007.

FLORIANO, E. P.; MULLER, I.; FINGER, C. A. G.; SCHNEIDER, P. R. Ajuste e seleção de modelos tradicionais para série temporal de dados de altura de árvores. *Ciência Florestal*, v.16, n.2, p.177-199, 2006.

HAIR JUNIOR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. *Multivariate data analysis: with readings*. 5. ed. New Jersey: Prentice-Hall, 1998. 730 p.

HOCKING, R. R. The analysis and selection of variables in linear regression. *Biometrics*, v.32, p.1-49, 1976.

IAFFE, A.; ARRUDA, F. B.; SAKAI, E.; PIRES, R. C. M.; CALHEIROS, R. O. Quantificação do efeito dos elementos agrometeorológicos na produção de um ensaio de café em Ribeirão Preto, SP. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 1., Poços de Caldas. Anais... Minas Gerais: 2000. p.794-797.

LEVINE, D. M.; BERENSON, M. L.; STEPHAN, D. *Statistics for Managers Using Microsoft Excel*. 2nd Ed., Prentice Hall, New Jersey, EUA, 1999.

LEVINE, D. M.; STEPHAN, D. F.; KREHBIEL, T. C.; BERENSON, M. L. *Estatística: Teoria e aplicações*. 5. ed. Rio de Janeiro: LTC, 2011. 752 p.

MALLOWS, C. L. Some comments on C_p . *Technometrics*, v. 15, p.661-675, 1973.

- MALLOWS, C. L. More comments on Cp. **Technometrics**, v. 37, p. 362–372, 1995.
- MELLO, R.; SILVA, A. M. Modelagem estatística da precipitação mensal e anual e no período seco para o estado de Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.13, n.1, p.68–74, 2009.
- MONTGOMERY, D. C.; RUNGER, G. C. Estatística aplicada e probabilidade para engenheiros. 5. ed. Rio de Janeiro: LTC, 2012. 521p.
- MORETTIN, P. L.; TOLOI, C. M. C. Análises de séries temporais. 2. ed. São Paulo: Edgard Blucher, 2006. 531p.
- NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J.; WASSERMAN, W. Applied linear statistical models. 4th ed. Homewood, IL: Irwin, 1996. 1408 p.
- OLIVEIRA, A. B. Melhoria do processo Batschek, por meio de tratamento de dados históricos, para fabricação de telhas onduladas. 2010. 49 f. Dissertação (Mestrado em Engenharia) - Universidade de São Paulo, São Paulo.
- PIMENTA, C. D.; SILVA, M. B.; RIBEIRO, R. B.; RAMOS, A. W. Planejamento de experimentos (DOE) aplicado no processo de têmpera e revenimento de arames de aço SAE 9254. **Revista Janus**, lorena, v.5, n.8, p.119-139, 2008.
- PRELA-PANTANO, A.; DUARTE, A. P.; SILVA, D. F.; ROLIM, G. S.; CASER, D. V. Produtividade de milho, precipitação e ocorrência de enos na região do médio Paranapanema, SP, BRASIL. **Revista Brasileira de Milho e Sorgo**, v.10, n.2, p. 146-157, 2011.
- ROSS, T. Indices for performance evaluation of predictive models in food microbiology. **Journal Applied Bacteriology**, v. 81, p. 501–508, 1996.
- SILVA, L. L.; COSTA, R. F.; CAMPOS, J. H. B.; DANTAS, R. T. Influência das precipitações na produtividade agrícola no Estado da Paraíba. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n. 4, p. 454–461, 2009.
- SILVA, J. W.; GUIMARÃES, E. C.; TAVARES, M. Variabilidade temporal da precipitação mensal e anual na estação climatológica de Uberaba-MG. **Ciência e Agrotecnologia**, v.27, n.3, p. 665-674, 2003.