

APLICAÇÃO DE REGRESSÃO LINEAR MÚLTIPLA PARA ESTIMATIVA DA PRECIPITAÇÃO MÉDIA ANUAL CONSIDERANDO A VARIABILIDADE ESPACIAL NO ESTADO DE SERGIPE

Marcus Aurélio Soares Cruz^{1*} & Luciano Alves de Souza² & Ricardo de Aragão³ & Rafael Rodrigues de Souza Silva⁴ & Paulo Vinicius Melo da Mota⁵

Resumo – A precipitação é um fenômeno atmosférico caracterizado pela sua aleatoriedade e variabilidade espaço-temporal. O entendimento da variabilidade espacial da precipitação em uma região ou bacia hidrográfica tem grande importância para as ações ligadas ao planejamento e uso dos recursos hídricos. Este estudo visou obter uma equação por meio de regressão linear múltipla que correlacione valores anuais de precipitação e fatores de localização espacial para 58 postos espalhados pelo estado de Sergipe. A equação obtida apresentou coeficiente de determinação de 0,844, que pode ser considerado muito bom, com erros médios percentuais situados entre -7,5% a 5,5%, variando segundo as diferentes regiões climáticas presentes no estado.

Palavras-Chave – análise estatística, séries históricas, pluviometria.

APPLICATION OF MULTIPLE LINEAR REGRESSION FOR ESTIMATE AVERAGE ANNUAL RAINFALL INCLUDING SPATIAL VARIABILITY IN THE STATE OF SERGIPE

Abstract – Rainfall is a atmospheric phenomenon characterized by its randomness and spatio-temporal variability. Understanding the spatial variability of rainfall in a region or basin is very important for the actions related to the planning and use of water resources. This study aimed at obtaining an equation using multiple linear regression to correlate annual rainfall data and spatial factors to 58 stations throughout the state of Sergipe. The equation obtained showed a correlation coefficient of 0.844, which can be considered very good, with average percentage errors located between -7.5% and 5.5%, varying according to the different climatic regions in the state.

Keywords – statistical analysis, time series, rainfall.

INTRODUÇÃO

Os estudos hidrológicos pressupõem a detenção de informações mínimas sobre as variáveis envolvidas no ciclo hidrológico que permitam o seu desenvolvimento de forma confiável. A precipitação é sem dúvida umas das variáveis de maior demanda para tais estudos, tendo em vista a sua importância para as atividades relacionadas ao desenvolvimento de uma região, como agricultura, indústria e serviços urbanos diversos. Segundo Carvalho e Assad (2005), o conhecimento da distribuição espacial da precipitação é relevante no planejamento agrícola,

¹ Embrapa Tabuleiros Costeiros, Avenida Beira Mar 3250, Jardins, CEP: 49024-040, Aracaju-SE, marcus.cruz@embrapa.br

² Embrapa Tabuleiros Costeiros, Avenida Beira Mar 3250, Jardins, CEP: 49024-040, Aracaju-SE, luciano.alves@embrapa.br

³ Universidade Federal de Sergipe, São Cristóvão-SE, ricardoaragao@yahoo.com

⁴ Embrapa Tabuleiros Costeiros, Avenida Beira Mar 3250, Jardins, CEP: 49024-040, Aracaju-SE, maximusrrs@hotmail.com

⁵ Embrapa Tabuleiros Costeiros, Avenida Beira Mar 3250, Jardins, CEP: 49024-040, Aracaju-SE, paulovin@live.com

principalmente com relação à instalação de culturas, considerando ainda a influência nos níveis d'água dos mananciais, conservação do solo e adequado dimensionamento de obras hidráulicas.

A precipitação é um fenômeno de natureza aleatória, apresentando grande variabilidade espacial, em virtude de diversos fatores, tais como localização espacial, altitude do local, microclimas, movimentações de massas de ar, dentre outros. (Goovaerts (2000), Valeriano (2003), Fuckner et al. (2007)). Tal característica realça a importância de conhecer o comportamento médio das chuvas em uma região ou bacia de interesse.

Além da variabilidade espacial, muitos problemas estão relacionados ao desenvolvimento de estudos que considerem diferentes escalas temporais, como chuvas mensais e anuais, por exemplo, uma vez que muitas regiões no país não apresentam monitoramento contínuo, seja pela ausência de postos pluviográficos, ou ainda, pela pouca quantidade de anos de registro em virtude de falhas (Bertoni e Tucci (2007), Oliveira et al. (2010)). Tal situação demanda então a aplicação de técnicas de transferência espacial da informação conhecida (Bailey e Gatrell, 1996). As principais metodologias para esta transferência estão associadas aos interpoladores determinísticos ou estocásticos ou ainda a aplicação de análises de regressão múltipla que considere parâmetros intervenientes espacialmente (Boni et al., 2008).

Interpoladores determinísticos exatos, como Inverso do Quadrado da Distância (IDW) ou Splines, ou ainda estocásticos, que considerem a dependência espacial, como Kriging, estão disponíveis em diversos sistemas de informações geográficas e permitem a confecção de mapas ou grades de pontos com a informação espacializada, possibilitando em alguns casos, como em Kriging, a avaliação do ajuste obtido (Philips et al., 1992). No entanto, alguns destes métodos mostram ainda um elevado grau de complexidade em suas rotinas, não possibilitando a produção de uma equação que possa ser disponibilizada ao usuário não especialista para estimar com segurança o valor da precipitação média em locais desprovidos de monitoramento, a partir da entrada de informações de fácil obtenção.

Neste sentido a aplicação de regressão múltipla correlacionando variáveis intervenientes nos volumes médios anuais e mensais precipitados em grandes áreas, pode fornecer equações que permitam ao interessado obter rapidamente uma estimativa para qualquer ponto conhecendo a grandeza do erro que pode estar cometendo. Assim, este estudo apresenta a aplicação de regressão múltipla sobre dados de precipitações médias anuais, buscando correlacioná-las com variáveis de caracterização espacial e fornecer uma equação aplicável para o estado de Sergipe.

MATERIAIS E MÉTODO

Caracterização da área de estudo

O Estado de Sergipe é o menor estado em extensão territorial do país, apresentando cerca de 21.910 km², situado na região Nordeste com três regiões climáticas bem caracterizadas: Litoral úmido, Agreste e Semiárido. As precipitações têm seus maiores volumes concentrados no período de abril a agosto, com alturas pluviométricas anuais médias variando de cerca de 1600 mm junto ao litoral, 1000 mm em sua porção intermediária (Agreste) e 600 mm no semiárido sergipano (SEMARH, 2012). Na Figura 1 são apresentadas a localização do estado de Sergipe e algumas características citadas. A topografia do território sergipano é relativamente plana, com a presença de áreas de relevo levemente ondulado e altitudes inferiores a 300 metros, com planaltos situados a

noroeste. O estado de Sergipe apresenta uma boa cobertura de postos pluviográficos, no entanto a instalação da maioria destes é recente, com séries inferiores a 5 anos de dados. Considerando os postos com séries mais longas (superior a 10 anos), foram identificados 63, com registros históricos em períodos diversos, todos sob coordenação da ANA (ANA, 2013), destes foram descartados cinco por problemas nos registros. A Figura 1 apresenta também a localização dos postos identificados no estado e na Tabela 1 são apresentadas suas características básicas.

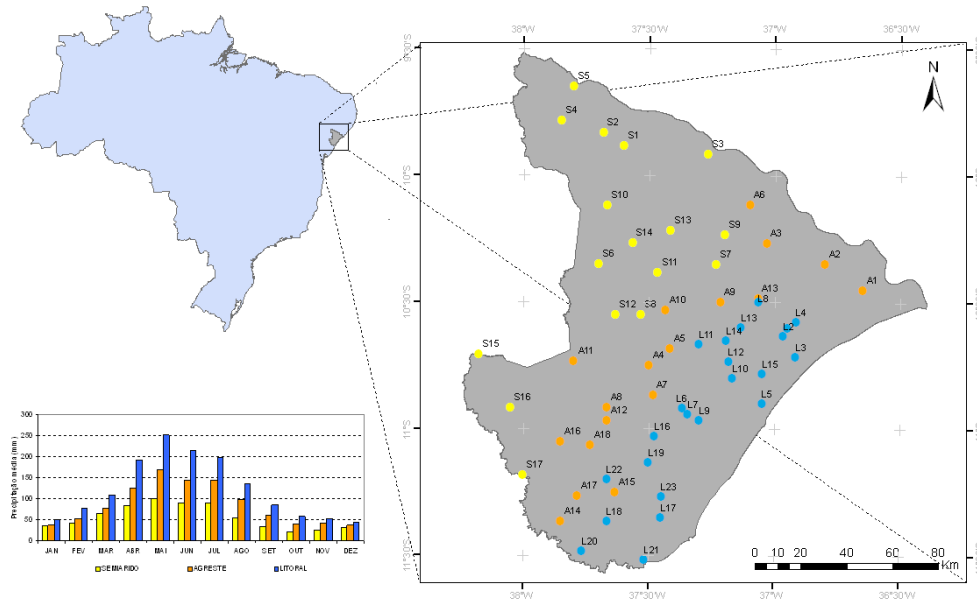


Figura 1 – Localização do estado de Sergipe e postos de registros históricos de precipitação segundo as regiões climáticas: Litoral Úmido (azul), Agreste (laranja) e Semiárido (amarelo).

A regressão linear múltipla

A Regressão Linear Múltipla é um modelo estatístico referente ao tratamento de séries temporais de dados, advindo da Regressão Linear, esta que se restringe a duas variáveis e a apenas uma equação funcional do primeiro grau de ajustamento. A análise de Regressão Múltipla é uma metodologia estatística de previsão de valores de uma ou mais variáveis de resposta (Dependentes) através de um conjunto de variáveis explicativas (Independentes) (Naghetini e Pinto, 2007). Quanto mais significativo for o peso de uma variável isolada, ou de um conjunto de variáveis explicativas, tanto mais se poderá afirmar que alguns fatores afetam mais o comportamento de uma variável de resposta especificamente procurada do que outros (Kaznar e Gonçalves, 2007). A equação que representa a regressão linear múltipla é:

$$Y = a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3 + \dots + a_p \cdot X_p \quad (1)$$

onde Y é a variável dependente ou prevista; $X_1, X_2, X_3, \dots, X_p$ são as variáveis independentes ou explicativas e $a_1, a_2, a_3, \dots, a_p$ são os coeficientes de regressão. Assim, a partir de um conjunto de n valores da variável Y , associados às n observações correspondentes das p variáveis independentes pode-se definir os coeficientes da regressão pelo método dos mínimos quadrados.

A verificação da qualidade de uma equação ajustada pode ser realizada por meio do coeficiente de determinação (R^2) da regressão múltipla:

Tabela 1 – Postos pluviométricos no estado de Sergipe com séries históricas superiores a dez anos

Posto ANA	ID	Latit.	Longit.	Altitude (m)	Prec. Média (mm/ano)	Posto ANA	ID	Latit.	Longit.	Altitude (m)	Prec. Média (mm/ano)
937024	S1	-9,883	-37,600	200,0	517,5	1037078	A13	-10,483	-37,067	139,0	1229,9
937025	S2	-9,833	-37,683	171,0	497,8	1137003	A14	-11,367	-37,850	120,0	1134,5
937026	S3	-9,917	-37,267	45,0	552,7	1137010	A15	-11,250	-37,633	109,0	952,4
937027	S4	-9,783	-37,850	200,0	531,2	1137011	A16	-11,050	-37,850	230,0	959,1
937028	S5	-9,650	-37,800	130,0	479,3	1137020	A17	-11,267	-37,783	225,0	987,6
1037010	S6	-10,350	-37,700	351,0	795,6	1137025	A18	-11,067	-37,733	167,0	1080,9
1037013	S7	-10,350	-37,233	180,0	797,8	1036014	L1	-10,600	-36,950	79,0	1270,9
1037014	S8	-10,550	-37,533	272,0	825,1	1036020	L2	-10,633	-36,967	80,0	1572,7
1037016	S9	-10,233	-37,200	200,0	790,5	1036033	L3	-10,717	-36,917	30,0	1567,4
1037027	S10	-10,117	-37,667	80,0	632,1	1036063	L4	-10,578	-36,916	10,0	1404,0
1037031	S11	-10,383	-37,467	250,0	714,6	1037002	L5	-10,900	-37,050	3,0	1632,1
1037032	S12	-10,550	-37,633	204,0	678,1	1037006	L6	-10,917	-37,367	100,0	1475,9
1037034	S13	-10,217	-37,417	290,0	698,0	1037007	L7	-10,942	-37,345	19,0	1507,6
1037038	S14	-10,267	-37,567	180,0	736,5	1037009	L8	-10,500	-37,067	148,0	1300,6
1038000	S15	-10,709	-38,178	300,0	808,8	1037022	L9	-10,967	-37,300	10,0	1502,7
1038001	S16	-10,917	-38,050	250,0	680,9	1037028	L10	-10,800	-37,167	9,0	1288,6
1138000	S17	-11,183	-38,000	157,0	759,0	1037030	L11	-10,667	-37,300	224,0	1394,4
1036021	A1	-10,450	-36,650	20,0	1188,7	1037042	L12	-10,733	-37,183	30,0	1627,9
1036054	A2	-10,350	-36,800	89,0	934,7	1037047	L13	-10,600	-37,133	90,0	1255,8
1037003	A3	-10,267	-37,033	217,0	919,8	1037049	L14	-10,653	-37,193	47,0	1345,5
1037008	A4	-10,750	-37,500	180,0	1106,7	1037050	L15	-10,783	-37,050	9,0	1350,4
1037019	A5	-10,683	-37,417	186,0	876,7	1137001	L16	-11,030	-37,478	92,0	1349,7
1037021	A6	-10,117	-37,100	187,0	926,0	1137002	L17	-11,350	-37,450	50,0	1992,6
1037024	A7	-10,867	-37,483	100,0	1206,7	1137004	L18	-11,367	-37,667	109,0	1366,8
1037025	A8	-10,917	-37,667	183,0	1043,3	1137012	L19	-11,133	-37,500	150,0	1600,9
1037036	A9	-10,500	-37,217	200,0	1050,5	1137014	L20	-11,483	-37,767	173,0	1360,6
1037044	A10	-10,533	-37,433	350,0	950,1	1137018	L21	-11,517	-37,517	21,0	1576,8
1037045	A11	-10,733	-37,800	283,0	887,4	1137023	L22	-11,200	-37,667	170,0	1418,8
1037051	A12	-10,967	-37,667	400,0	1093,0	1137015	L23	-11,267	-37,450	53,0	1491,9

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

onde R^2 é o coeficiente de determinação ($0 \leq R^2 \leq 1$), y_i é o valor observado da variável dependente, \hat{y}_i é o valor estimado da variável dependente e \bar{y} é a média da variável dependente,

No presente estudo, foi proposta a realização de uma análise por regressão múltipla considerando a precipitação anual média para os postos disponíveis, como a variável dependente e sugeridas três variáveis independentes que tenham por características principais a possibilidade de considerar a localização espacial do registro bem como a influência da altimetria nos valores de chuvas e que, além disso, sejam de obtenção simples por qualquer usuário potencial. Assim, foram consideradas as latitudes e longitudes decimais dos postos de medição, no sistema WGS1984 e a

altitude dos mesmos. Tais valores estão apresentados também na Tabela 1. Desta forma a equação para estimativa da precipitação anual média para o estado de Sergipe teria o seguinte formato:

$$\bar{P}_{anual} = a_1 \cdot LATD + a_2 \cdot LONG + a_3 \cdot ALTI + K \quad (3)$$

onde \bar{P}_{anual} é a precipitação média anual nos postos de registro, LATD é a latitude do ponto em graus decimais, LONG é a longitude do ponto em graus decimais, ALTI é a cota altimétrica do ponto em metros e K é a constante resultante do ajuste.

RESULTADOS E DISCUSSÃO

A partir das séries históricas dos 58 postos distribuídos de forma uniforme pela superfície do estado, foram calculadas as alturas precipitadas médias anuais para cada posto, e agrupados segundo a região climática do estado, conforme é apresentado na Tabela 2. Observa-se que há uma quantidade maior de postos localizados na região do Litoral Úmido (23), contrapondo com 18 postos na porção do Agreste e 17 no Semiárido. A precipitação média geral para o estado, a partir das séries disponíveis foi de cerca de 1098 mm, com desvio padrão de 356 mm. A precipitação média anual na região do Litoral Úmido, totalizou aproximadamente, 1460 mm, cerca de 42% maior que a média anual para o Agreste (1029 mm) e 116% com relação ao semiárido sergipano (676 mm). O C.V. para a região Semiárida foi maior que para as duas outras, evidenciando uma maior variabilidade entre os valores da amostra para esta região com relação à sua média.

Tabela 2 – Análise exploratória dos valores de precipitação anual por região climática

REGIÕES CLIMÁTICAS					
LITORAL ÚMIDO		AGRESTE		SEMIÁRIDO	
Parâmetros	Precipitação média anual	Parâmetros	Precipitação média anual	Parâmetros	Precipitação média anual
Amostra	23	Amostra	18	Amostra	17
Média	1463,3	Média	1029,3	Média	676,2
Desv. Pad.	165,4	Desv. Pad.	112,6	Desv. Pad.	119,5
C.V.	0,11	C.V.	0,11	C.V.	0,18
Média geral da amostra					1097,9
Desvio Padrão					356,0
Coeficiente de variação					0,32

Realizada a análise exploratória dos dados amostrais, partiu-se então para verificação da correlação entre a variável dependente (precipitação) e as variáveis independentes (latitude, longitude e altitude), por meio da realização de procedimentos de regressão linear múltipla e cálculo do coeficiente de determinação. Para fins comparativos, optou-se por realizar a análise com os dados originais em com a aplicação de logaritmos, buscando melhorar os resultados finais do ajuste. Para tanto foi utilizada a ferramenta *Análise de Dados - Regressão* do Excel. A Tabela 3 apresenta o resultado para os dados originais e logaritmos.

Observa-se a partir dos resultados apresentados que, considerando a regressão realizada com os dados originais, há uma boa correlação entre a precipitação observada e as variáveis independentes, resultando em um coeficiente de determinação igual a 0,778. Pode-se verificar também que houve

uma melhora no coeficiente em virtude da aplicação de logaritmos aos dados originais, resultando em 0,844, que pode ser considerado uma indicação de muito boa correlação entre a precipitação observada e os valores estimados a partir da regressão.

Tabela 3 – Resultados da análise por regressão múltipla dos dados originais e em logaritmos

Parâmetros	Originais	Logaritmos
Observações	58	58
R ²	0,778	0,844
Erro padrão	172,267	0,142
Coeficientes		
Interseção	14357,52	70,142
LATD	-585,733	6,752
LONG	518,964	-21,829
ALTI	-0,642	-0,025

Assim, considerando o ajuste para os dados em logaritmos, tem-se a seguinte equação para estimativa da precipitação média anual no estado de Sergipe:

$$\bar{P}_{anual} = e^{(6,752 \cdot \ln(|LATD|) - 21,829 \cdot \ln(|LONG|) - 0,025 \cdot \ln(ALTI) + 70,142)} \quad (4)$$

onde LATD e LONG estão em graus decimais no sistema WGS1984 e a ALTI em metros.

A Figura 2 representa graficamente a qualidade da estimativa realizada pela equação (4) por meio da plotagem das precipitações observadas versus estimadas. Observa-se que há uma aproximação no ângulo da reta do ajuste linear aos 45 graus, o que caracteriza a região de coincidência da estimativa.

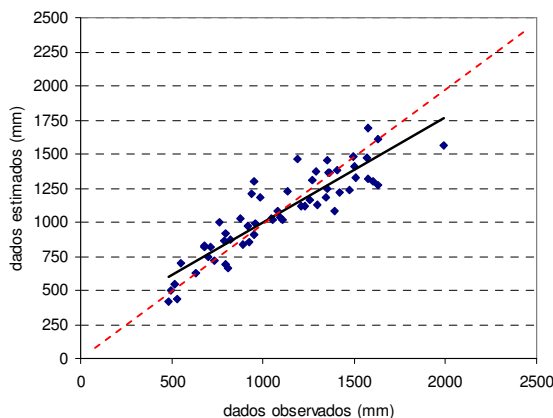


Figura 2 – Plotagem dados observados *versus* dados estimados pela equação de regressão

Para avaliar a dimensão dos erros de estimativa decorrentes da aplicação da equação obtida para qualquer região do estado, foi realizada uma análise considerando as diferentes regiões climáticas. Para tanto foi calculado o erro médio percentual de cada estimativa com relação ao valor anual registrado no posto (Tabela 4). Como pode ser observado, o maior erro médio percentual foi verificado nos postos do Litoral Úmido, com valor da ordem de - 7,5%, indicando que em média, a equação subestima o valor das precipitações anuais nesta região climática. Para Agreste e Semiárido, os erros médios foram da mesma ordem (5,1% e 5,5% respectivamente), menores que

para o Litoral Úmido e positivos, indicando que os valores estimados para estas regiões serão em média maiores que os observados. Ao observar-se o desvio padrão dos erros percentuais, verifica-se que este foi maior para a região do Semiárido sergipano, sendo o menor registrado no Litoral Úmido, evidenciando que a amplitude de variação do erro médio cometido nesta região, que possui menores alturas pluviométricas, é mais significativa. A partir dos valores de erros, pode-se inferir que a equação obtida pode ser utilizada com bom grau de confiabilidade para estimar a altura pluviométrica anual em qualquer ponto do estado de Sergipe.

Tabela 4 – Análise dos erros médios percentuais segundo as regiões climáticas

REGIÕES CLIMÁTICAS					
LITORAL ÚMIDO		AGRESTE		SEMIÁRIDO	
Parâmetros	Erro médio (%)	Parâmetros	Erro médio (%)	Parâmetros	Erro médio (%)
Média	-7,5	Média	5,1	Média	5,5
Desv. Pad.	9,6	Desv. Pad.	14,1	Desv. Pad.	15,2
C.V.	-1,28	C.V.	2,76	C.V.	2,77

Elaborou-se também mapa de confiabilidade da informação, onde são mostrados valores dos erros percentuais de forma espacializada (Figura 3). Este mapa possibilita ao usuário o conhecimento prévio da dimensão percentual da subestimativa ou sobrestimativa que a equação está produzindo para aquele local específico. Para tanto foi realizada uma interpolação a partir dos valores de erro percentual de cada posto utilizando um algoritmo simples do tipo Inverso do Quadrado da Distância (IDW). Observa-se que, conforme já mencionado, as regiões com predominância de sobrestimativa do valor da precipitação localizam-se na porção Agreste e Semiárido do estado, com exceção do extremo norte onde há registros em dois postos com subestimação dos valores. Por outro lado, na região mais próxima ao litoral, há predominância de áreas de subestimativa da precipitação. Observa-se ainda pelo mapa que a maior parte da área do estado está classificada com erros percentuais inferiores a 10% em ambos os casos.

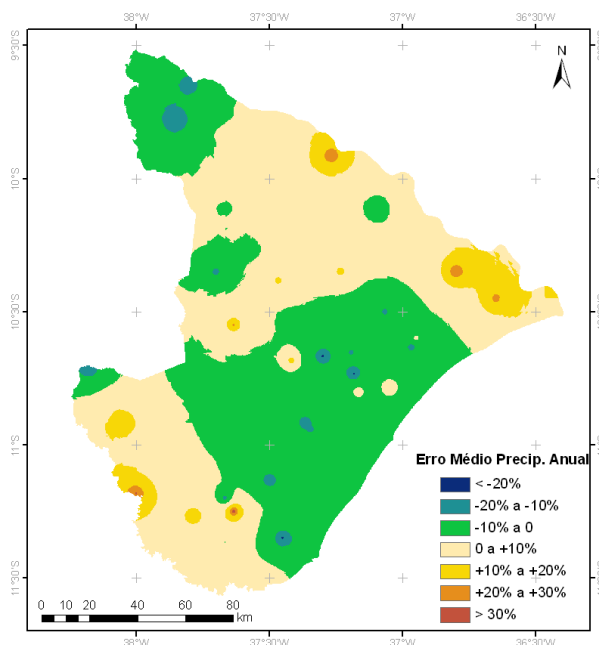


Figura 3 – Mapa de distribuição espacial do erro médio percentual da estimativa

CONCLUSÕES

A partir da aplicação da análise de regressão múltipla foi possível obter uma equação para a estimativa da precipitação anual média em função de variáveis independentes relacionadas a variabilidade espacial da chuva, que foram as coordenadas geográficas e a altitude dos postos de medição de alturas pluviométricas com séries históricas superiores a dez anos. Esta equação obteve um coeficiente de determinação de 0,844 e suas estimativas tiveram erros médios percentuais da ordem de -7,5% a +5,5%, de acordo com as regiões climáticas onde estão localizados, de forma que observou-se uma subestimação dos valores na região de maiores alturas pluviométricas (Litoral úmido) e sobrestimação no Agreste e Semiárido. O mapa de erros percentuais mostrou que na maior parte da superfície do estado estaria sendo cometido um erro da ordem de $\pm 10\%$ pela aplicação da equação gerada.

REFERÊNCIAS

- ANA. Agência Nacional de Águas. *Portal HIDROWEB*. In: www.ana.gov.br. Acesso: Abr. 2013.
- BAILEY, T., GATRELL, T. *Interactive Spatial Data Analysis*. Prentice Hall, England, 1996.
- BERTONI, J.C., TUCCI, C.E.M. Precipitação. In: TUCCI, C.E.M. *Hidrologia: Ciência e Aplicação*. Porto Alegre: UFRGS, 2007.
- BONI, G., PARODI, A., SICCARDI, F. A New Parsimonious Methodology of Mapping the Spatial Variability of Annual Maximum Rainfall in Mountainous Environments. *Journal of Hydrometeorology*, v. 9, pp. 492-506, 2008.
- CARVALHO, J. R. P.; ASSAD, E. D. Análise espacial da precipitação pluviométrica no Estado de São Paulo: Comparação de métodos de interpolação. *Engenharia Agrícola*, v. 25, n. 2, p. 377-384. 2005.
- FUCKNER, M.A., MAIA, A.L.S., IWASHITA, F., BISPO, P.C. Dados topográficos e sua relação com a distribuição espacial da precipitação: estudo em área teste no estado de São Paulo – Brasil. In: *Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto*, Florianópolis, Abr. 2007, pp. 3929-3936.
- GOOVAERTS, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, v. 228, pp. 113-129, 2000.
- KASZMAR, I. K., & GONÇALVES, B. M. L. *Regressão múltipla: uma digressão sobre seus usos*. Rio de Janeiro: IBICI. 2007.
- NAGHETTINI, M., ANDRADE PINTO, E.J. *Hidrologia estatística*. Serviço Geológico do Brasil. Belo Horizonte, 2007. 552p.
- OLIVEIRA, L.F.C., FIOREZE, A.P., MEDEIROS, A.M.M., SILVA, M.A. Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v14, n. 11, pp. 1186-1192, 2010.
- PHILIPS, D., DOLPH, J., MARKS, D. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology*, v.58, pp.119-141, 1992.
- SEMARH. Secretaria de Estado do Meio Ambiente e dos Recursos Hídricos. *Atlas Digital dos Recursos Hídricos do Estado de Sergipe*. Superintendência de Recursos Hídricos, Aracaju, 2012.
- VALERIANO, M. M. Curvatura vertical de vertentes em microbacias pela análise de modelos digitais de elevação. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 7, n. 3, pp. 539-546, 2003.