MODELO MARKOVIANO PARA A PRECIPITAÇÃO DIÁRIA NA REGIÃO DE PATOS-Pb.

MÁRIO DE MIRANDA VILAS BOAS RAMOS LEITÃO [1] e PEDRO VIEIRA DE AZEVEDO [1]

INTRODUÇÃO

A distribuição anual média da precipitação é um parâmetro importante para a agricultura porque permite o estabelecimento da estação de cultivo e da melhor época para o plantio. No entanto, não prever a ocorrência de períodos secos (veranicos) dentro da estação chuvosa, os quais prejudicam sobre maneira o cultivo não irrigado. Para se obter tais informações, torna-se necessário conhecer a função de distribuição de sequências de dias sêcos dentro da estação chuvosa.

A precipitação diária não pode ser considerada como um evento aleatório independente que segue uma distribuição de Bernoulli. Apesar da ocorrência ou não de precipitação num determinado dia ser uma decorrência das condições atmosféricas naquele dia e local, é condicionada, também, ao tempo passado, isto é, a situação ocorrida no dia ou dias anteriores. Assim, a precipitação diária tende a apresentar uma dependência estocástica, com o surgimento de seqüências de dias secos e chuvosos intercalados. Tal persistência meteorológica pode ser descrita apropriadamente por um modelo de cadeia de Markov, cuja ordem corresponde à ordem de dependência condicional do fenômeno (Chin, 1977). A ordem do modelo corresponde ao grau de dependência sobre o passado, por exemplo, a ordem 1 significa que a probabilidade de que um certo dia seja chuvoso ou seco, depende apenas da situação ocorrida (seco ou chuvoso) no dia anterior. A ordem 2 quando a probabilidade de ocorrência de dia chuvoso ou seco depende das situações ocorridas nos 2 dias anteriores e assim por diante.

O modelo de primeira ordem tem sido empregado por alguns autores (Gabriel & Neumann, 1962; Katz, 1974). Entretanto, vários autores (Chin,

[1] Professor do Centro de Ciências e Tecnologia da UFPb, Campina Grande-Pb.

1976 e 1977; Bruhn & Fry, 1980; Tong, 1975; Momorou, 1983) têm mostrado que nem sempre o modelo de primeira ordem é o mais apropriado e estabelecem a metodologia para determinação da ordem mais adequada em cada caso.

O presente trabalho objetiva determinar a ordem do modelo de Markov mais apropriada para prever a precipitação diária para a região de Patos-Pb, visando determinar as probabilidades de ocorrência de períodos secos ao longo do trimestre mais chuvoso.

METODOLOGIA

Foi utilizada uma série de dados diários de precipitação de 27 anos (1938 a 1965) num modelo de Markov definido como uma seqüência de eventos aleatórios discretos de ordem k, onde k é o menor inteiro positivo, tal que para todo n, as probabilidades condicionais são satisfeitas pela seguinte equação:

$$p\{X_n/X_{n-1}, X_{n-2}, \ldots, X_{n-k}, X_{n-k-1} \ldots\} = p\{X_n/X_{n-1}, \ldots, X_{n-k}\} \qquad (1)$$

onde X_n/X_{n-1} representa a ocorrência do estado específico X no dia de ordem n, dado que X ocorreu no dia anterior (n-1). No presente trabalho, considerou-se um processo de Markov para a precipitação diária, com 2 estados: chuvoso (c) quando ocorre precipitação superior a 0,4mm e seco (s) quando a precipitação for igual ou menor que 0,4mm. A análise foi feita para o trimestre mais chuvoso (fevereiro-março-abril) e para a estação chuvosa (janeiro a junho). Os períodos secos dentro da estação considerada foram definidos como a seqüência de N dias secos, seguidos e antecedidos de, pelo menos, um dia chuvoso, onde $N \geq 1$ é o comprimento do período (dias).

A determinação da ordem da cadeia de Markov que melhor se ajusta à série de dados analisados foi determinada usando o Critério de Informação de Akaike (AIC), baseando-se no modelo de regressão de múltipla variância:

$$AIC = -2 \log(L = \text{máxima verossimilhança}) + 2k \qquad (2)$$

em que k é o número de parâmetros indenpendentes ajustados dentro do modelo.

Em notação simplificada, as probabilidades de transição de uma cadeia de Markov podem ser dadas por $P_{i,\ i+1,\ \ldots,\ j-1,\ j}$ e as frequências destas transições por $N_{i,\ i+1,\ \ldots\ j-1,j}$. Para grandes séries de observações a função log-verossimilhança é dada por:

$$\log(L) = \sum_{i,i+1,\ldots,j-1,j} N_{i,i+1,\ldots,j-1,j} \log(N_{i,i+1,\ldots,j-1,j}/N_{i,i+1,\ldots,j-1}) \quad (3)$$

onde $(N_{i,i+1,\ldots,j-1,j}/N_{i,i+1,\ldots,j-1})$ é o tamanho máximo da verossimilhança.

Para determinar a ordem que melhor se ajusta ao modelo, compara-se a validade relativa entre um modelo de ordem m com um modelo de ordem (m-1), através da razão de log-verossimilhança ($\log \lambda_{m-1,m}$). Hoel (1954) mostrou que $-2 \log \lambda_{m-1,m}$ para uma cadeia ergódica é assintoticamente uma variável χ^2 com $\upsilon = \nabla^2 S^{m+1} = \nabla S^{m+1} - \nabla S^{k+1}$ graus de liberdade.

Substituindo $-2 \log \lambda_{m-1,m}$ por ${}_{m-1}H_m$ na equação (3) obtem-se:

$$
\begin{aligned}
{}_{m-1}H_m = 2 \sum_{i,i+1,\ldots,j-1,j} N_{i,i+1,\ldots,j-1,j} \Big[\log \frac{N_{i,i+1,\ldots,j-1,j}}{N_{i,i+1,\ldots,j-1}} - \\
\log \frac{N_{i,i+1,\ldots,j-1,j}}{N_{i+1,\ldots,j-1}}\Big]. \quad (4)
\end{aligned}
$$

A equação (4) é uma medida de como a seqüência observada melhor suporta as hipóteses de uma cadeia de ordem m versus uma cadeia de ordem (m-1). Para $k < (m-1)$, pode-se mostrar que:

$$ {}_kH_m = {}_kH_{k+1} + \cdots + {}_{m-1}H_m \quad (5)$$

onde os termos individuais são considerados assintoticamente independentes. Então, dado que a cadeia é de ordem k, ${}_kH_m$ tem uma distribuição χ^2 com ν graus de liberdade.

A ordem da cadeia é melhor identificada através da função perda dada por (Tong, 1975):

$$ R(K) = {}_kH_m - 2\nu \quad (6)$$

onde m é a maior ordem em consideração.

Os períodos secos para cadeias de ordem 1, 2 e 3, foram obtidos pelas equações:

$$p(N = 1,2,\ldots,n) = p(c).p(s/c).p(c/s).p(s/s)^{N-1}, \text{ para } k=1 \quad (6)$$

$$p(N=1) = p(c).p(s/c).p(c/s,c)$$

e $$p(N \geq 2) = p(c).p(s/c).p(s/s,c).p(c/s,s).\ [p(s/s,s]^{N-2}, \text{ para } k = 2 \quad (7)$$

$$p(N=1) = p(c).\ p(s/c).p(c/s,c),$$

$$p(N=2) = p(c).p(s/c).p(s/s,c).p(c/s,s,c) \text{ e}$$

$$p(N \geq 3) = p(c).p(s/c).p(s/s,c).p(s/s,s,c).p(c/s,s,s).[p(s/s,s,s)]^{N-3}, \text{ para } k=3 \quad (8)$$

Os valores esperados para os períodos secos de N dias foram obtidos multiplicando-se os valores das probabilidades (eqs, 6, 7 e 8) pelo número total de dias do período considerado.

RESULTADOS E DISCUSSÕES

A TABELA I mostra os valores da função perda para o período correspondente ao trimestre mais chuvoso (fevereiro-março-abril), para Patos-Pb. Esses resultados evidenciam que a cadeia de Markov de ordem 3 se ajusta melhor (menor valor de R) aos dados considerados. Esses cálculos foram também efetuados para a estação chuvosa (janeiro a junho) e para cada mês individualmente dentro dessa estação. Para a estação como um todo houve uma tendência de aumento da dependência — tendência para um melhor ajustamento com um modelo de ordem 4 — ao passo que houve uma tendência de redução, principalmente nos meses mais chuvosos, da ordem — de 3 para 2 — no caso dos meses individuais. Sugere-se que tal redução de dependência se deva à maior concentração de dias chuvosos consecutivos durante os meses mais chuvosos. Estes resultados mostram o alto grau de dependência da precipitação diária sobre o tempo passado, na região estudada.

A TABELA II mostra os resultados da análise do comprimento e freqüência de períodos secos durante o trimestre mais chuvoso, assim como os cálculos de probabilidades de ocorrência desses períodos para os modelos de ordem 1, 2 e 3. Observa-se que, mesmo para o trimeste mais chuvoso, a freqüência de ocorrência de períodos curtos (N = 2 e N = 3) é siginificativa, apesar de reduzir substancialmente com o aumento de N. Pela TABELA II pode-se confirmar que a cadeia de ordem 3 é a que melhor se ajusta à série analisada, uma vez que as freqüências esperadas por esse modelo se aproximam mais das freqüências observadas. Observa-se também que a precisão do modelo diminui com o aumento do comprimento do período seco.

SUMMARY

A Markov chain model was applied to the time serie of daily precipitation for Patos-Pb. The analysis showed that the data are best adjusted to a third order Markov model. The within season frequence distribuition of dry periods was studied along with probabilities of occurrence. This analysis confirmed the third order model as the best one and showed that the accuracy of the model for predicting the frequence of occurrence of dry period reduces as the degree of dependence increases.

BIBLIOGRAFIA CITADA

CHIN, E.H., 1976. Modeling daily precipitation occurrence process with Markov chain. Water Res. Research, 13: 949-956.

FEYERHERN, A.M. and BARK, L.D., 1967. Goodness of fit for a Markov chain model for sequence of wet and dry days. J. Appl. Met., 6: 770-77.

GABRIEL, K.R. and NEUMANN, J., 1962. A Markov chain model for rainfall occurrence at Tel Aviv. Quart. J.R. Met. Soc., 88: 90-95

HAAN, C.T., ALLEN, D.M. and STREET, J.D., 1976. A Markov chain model for daily rainfall. Water Res. Research, 12(3): 443-449.

HOEL, R.G., 1954. A test for Markov chains. Biometrika 41: 430-433.

KATZ, R.W., 1974. Computing probabilities associated with the Markov chain model for precipitation. J. Appl. Met. 13: 953-954.

MIMIROU, M., 1983. Daily precipitation occurrence modeling with Markov chain of seasonal order. Hydrol. Sc. Journal, 28: 221-232.

TONG, H., 1975. Determination of the order of a Markov chain by Akaike's Information criterion. J. Appl. Prob., 12: 488-497.

TABELA I

FUNÇÃO PERDA R(K) PARA K IGUAL A 4.

K(ordem)	R(K) Fevereiro-Março-Abril
0	232,582
1	46,119
2	10,533
3	4,414
4	0

TABELA II

COMPARAÇÃO DA DISTRIBUIÇÃO DE PERÍODOS SECOS OBSERVADOS COM OS RESPECTIVOS VALORES ESPERADOS, CALCULADOS PARA CADEIAS DE MARKOV DE 1ª ORDEM (K = 1), 2ª ORDEM (K = 2) e 3ª ORDEM (K = 3), PARA O TRIMESTRE FEV./MR./ABRIL DE 1938 a 1965, EM PATOS-Pb.

PERÍODOS SECOS COM N DIAS				
		ESPERADO		
N	OBSERVADO	K = 3	K = 2	K = 1
1	134	136	136	91
2	53	56	50	69
3	33	37	40	53
4	33	30	32	40
5	27	24	25	31
6	21	19	20	23
7	10	16	16	18
8	6	13	13	14
9	6	10	10	10
10	10	8	8	8
11	5	7	7	6
12	3	5	5	5
13	5	4	4	3
14	0	3	3	3
≥15	14	9	9	7