

MODELAGEM ESTATÍSTICA DA PRECIPITAÇÃO DIÁRIA NO ESTADO DO PARANÁ

Cláudio Marchand Krüger, Eloy Kaviski e Ingrid Illich Müller

Centro de Hidráulica e Hidrologia Prof. Parigot de Souza – CEHPAR
Caixa Postal 1309 – CEP 80001-970 Curitiba, PR

RESUMO

Histogramas de dados de precipitação diária são normalmente assimétricos à direita, com as maiores probabilidades concentradas na origem. Esta forma de distribuição sugere que os dados de precipitação diária sejam modelados em dois estágios: modelagem do padrão de ocorrência de dias secos e úmidos e modelagem da quantidade de precipitação em dias úmidos. No presente estudo, dois modelos lineares generalizados foram utilizados para a modelagem de dados de chuva, sendo um deles uma regressão logística. Estudos anteriores concluem que, em muitos casos, uma dependência de primeira ordem correspondente a uma cadeia de Markov fornece um modelo satisfatório para a ocorrência e não-ocorrência de chuva. Em um segundo estágio, um modelo para a quantidade de chuva é necessário, o qual deve ser definido para valores positivos, apenas, e com assimetria positiva. A distribuição Gama foi utilizada para este fim. Este método foi utilizado para a modelagem de dados de precipitação diária de estações pluviométricas do Estado do Paraná. O objetivo principal do estudo é a detecção de tendências na frequência e intensidade da chuva diária na bacia do rio Paraná.

INTRODUÇÃO

Pesquisas no assunto de modelagem de precipitações diárias (Stern e Coe, 1982, 1984; Clarke, 1994, 1998) têm proposto a aplicação de modelos de cadeia Markoviana para o estudo da frequência e distribuições Gama para a intensidade das precipitações médias diárias. Esta abordagem foi utilizada para a modelagem de dados de chuva, com o objetivo de comparar a forma dos modelos

matemáticos ajustados aos dados de precipitações médias diárias de dois sub-períodos das amostras, de forma a detectar eventuais alterações no comportamento da frequência e intensidade dos eventos pluviais. No caso da frequência, esta abordagem consiste em ajustar funções às frequências observadas na série histórica, das sequências de dias secos e úmidos. No caso do presente estudo, os modelos matemáticos ajustados consideram cadeias de Markov de primeira ordem, em que um estado (seco ou úmido) em um dia só depende do estado no dia anterior. Para a intensidade, uma variável é definida como o excesso de chuva acima de um valor mínimo. A distribuição da intensidade da chuva diária é geralmente bastante assimétrica, recomendando-se o uso de distribuições Gama. A análise de estacionariedade da frequência e intensidade da precipitação média diária foi baseada na comparação das funções ajustadas a dois sub-períodos da amostra; no caso, os períodos anteriores e posteriores a 1970. Expressões teóricas foram ajustadas separadamente para cada sub-período, e a seguir foram comparadas para verificação estatística da probabilidade de existir uma alteração nos parâmetros obtidos para cada sub-período. As análises da precipitação diária foram realizadas para postos localizados no Estado do Paraná.

DESCRIÇÃO DOS MÉTODOS UTILIZADOS

Frequência de dias secos e úmidos

Consiste em ajustar funções às frequências observadas na série histórica, das

seqüências de dias secos e úmidos. No caso do presente estudo, os modelos matemáticos ajustados consideram cadeias de Markov de primeira ordem, em que um estado (seco ou úmido) em um dia só depende do estado no dia anterior. Como normalmente existe sazonalidade na precipitação diária dentro do ano, recomenda-se que as funções sejam ajustadas levando-se em conta a variação das freqüências para cada dia do ano.

Desta forma, inicia-se com uma matriz $(N \times 365)$ onde N é o número de anos de observação. A seguir, condensa-se a matriz anterior em outra do tipo $(k \times 365)$ onde k depende da ordem da cadeia de Markov escolhida. Se for de ordem 1, $k = 4$, representando as freqüências observadas $n_{ij}(t) \leq N$ dos tipos de seqüência ij possíveis. Para uma cadeia de Markov de primeira ordem: 00, 01, 10, 11, onde “0” significa “dia seco” e “1” significa “dia úmido”. O conceito de “dia úmido” varia conforme a quantidade de chuva mínima a ser considerada. No presente estudo, este valor inferior foi adotado igual a 0,1 mm.

As estimativas das probabilidades de cada tipo de evento podem ser obtidas da seguinte forma (para uma cadeia de ordem 1):

$$p_{00}(t) = n_{00}(t) / [n_{00}(t) + n_{01}(t)] \quad (1)$$

$$p_{01}(t) = n_{01}(t) / [n_{00}(t) + n_{01}(t)] \quad (2)$$

$$p_{10}(t) = n_{10}(t) / [n_{10}(t) + n_{11}(t)] \quad (3)$$

$$p_{11}(t) = n_{11}(t) / [n_{10}(t) + n_{11}(t)] \quad (4)$$

onde (t) , $t = 1, \dots, 365$ significa que as freqüências variam para cada dia do ano, se houver sazonalidade.

É fácil observar, pelas expressões anteriores, que $p_{00}(t) + p_{01}(t) = 1$, da mesma forma que $p_{10}(t) + p_{11}(t) = 1$. Portanto, neste caso, o estudo será restrito às grandezas $p_0(t)$ e $p_1(t)$, ou seja, a freqüência de dias úmidos precedidos de um dia seco ou um dia úmido, respectivamente. Aos valores estimados das probabilidades, Stern e Coe (1982)

recomendam o ajuste de modelos lineares generalizados, com a seguinte forma:

$$\ln\left[\frac{p_0(t)}{1-p_0(t)}\right] = \alpha_0 + \alpha_1 \sin(2\pi t / 365) + \alpha_2 \cos(2\pi t / 365) \quad (5)$$

$$\ln\left[\frac{p_1(t)}{1-p_1(t)}\right] = \beta_0 + \beta_1 \sin(2\pi t / 365) + \beta_2 \cos(2\pi t / 365) \quad (6)$$

A regressão não pode ser realizada pelos métodos de regressão convencionais, pois a variável dependente não é normal (vem do cálculo de uma proporção). O ajuste deve ser feito através da maximização das respectivas funções de verossimilhança (Clarke, 1994). A função de verossimilhança pode ser estimada como o produto em t de todos os termos

$$[p_1(t)]^{n_{11}} [1-p_1(t)]^{n_{10}} [p_0(t)]^{n_{01}} [1-p_0(t)]^{n_{00}} \quad (7)$$

A decisão sobre o número de harmônicos (termos em seno e cosseno) e também sobre a ordem mais adequada da cadeia de Markov (primeira ou segunda ordem) pode ser realizada através da análise da estatística de viância (Clarke, 1998). No presente estudo, por simplicidade, o modelo adotado usa apenas dois harmônicos e cadeia de Markov de primeira ordem.

Intensidade da precipitação diária

A variável $X(t)$ é definida como o excesso de chuva acima de um valor mínimo (0,1 mm, por exemplo). A distribuição de $X(t)$ geralmente é bastante assimétrica, recomendando-se o ajuste de distribuições Gama. Para uma distribuição Gama com 2 parâmetros:

$$f(x) = [\kappa / \mu(t)]^\kappa x^{\kappa-1} \exp[-\kappa x / \mu(t)] / \Gamma(\kappa) \quad (8)$$

O parâmetro κ geralmente pode ser considerado constante ao longo do ano (determina o coeficiente de variação da

distribuição $= 1/\sqrt{\kappa}$). Se $\kappa = 1$, a expressão acima torna-se uma exponencial.

A dependência temporal do parâmetro $\mu = \mu(t)$, valor esperado da chuva no dia t , é levada em conta de forma semelhante à frequência, ou seja,

$$\ln[\mu_0(t)] = \lambda_0 + \lambda_1 \sin(2\pi t / 365) + \lambda_2 \cos(2\pi t / 365) \quad (9)$$

$$\ln[\mu_1(t)] = \delta_0 + \delta_1 \sin(2\pi t / 365) + \delta_2 \cos(2\pi t / 365) \quad (10)$$

Da mesma forma, o número de harmônicos a ser usado pode ser escolhido pela análise da deviança. No caso de uma cadeia de primeira ordem, existem dois tipos de média $\mu_k(t)$, conforme o dia anterior ao dia chuvoso foi seco ($k = 0$) ou chuvoso ($k = 1$). Assim como para a frequência, o ajuste dos modelos deve ser realizado pela maximização das respectivas funções de verossimilhança, neste caso, levando em consideração que as precipitações seriam distribuídas segundo uma distribuição Gama (Clarke, 1994).

Análise de tendências

A análise de tendências na frequência e intensidade da precipitação média diária foi baseada na comparação das equações ajustadas a dois sub-períodos da amostra. O objetivo desta análise é a detecção de modificações no padrão das seqüências de chuvas que possam ter ocorrido após 1970 na bacia do rio Paraná. O ano de 1970 pode ser considerado aproximadamente como sendo o final do período das principais alterações de uso do solo (principalmente desmatamento) que ocorreram no Estado do Paraná neste século. Para detectar as possíveis tendências na precipitação após 1970, as expressões mostradas anteriormente, ajustadas separadamente para cada sub-período da amostra, foram comparadas para verificação estatística da probabilidade de existir uma alteração nos parâmetros obtidos para cada sub-período.

A comparação é realizada em duas etapas. Na primeira etapa, as expressões ajustadas aos dados amostrais são comparadas em termos do paralelismo das curvas obtidas. Esta comparação se baseia na verificação dos coeficientes obtidos para os termos em seno e cosseno, realizando-se um novo ajuste, agora com os dados dos dois sub-períodos conjuntamente, acrescentando-se variáveis "dummy" em seno e cosseno e verificando a diferença em termos da deviança obtida com este novo ajuste, em relação aos ajustes obtidos anteriormente. A diferença de deviança é uma variável distribuída segundo uma distribuição qui-quadrado com o número de graus de liberdade igual ao número de variáveis "dummy" adicionadas. Este procedimento encontra-se descrito em Clarke (1998).

Se a hipótese de paralelismo não for rejeitada, procede-se à verificação do intercepto, ou seja, testa-se a hipótese das curvas paralelas diferirem apenas por um parâmetro de deslocamento. Da mesma forma, acrescenta-se uma variável "dummy" referente ao intercepto e realiza-se um novo ajuste com a totalidade dos dados dos dois sub-períodos, testando-se as diferenças das devianças entre o novo ajuste e os anteriores. Se a hipótese de interceptos iguais não for rejeitada, conclui-se que os ajustes realizados sobre os dois sub-períodos podem ser considerados estatisticamente indistinguíveis.

RESULTADOS

As análises anteriormente descritas foram realizadas com os dados de precipitação diária de 38 estações pluviométricas com período suficiente para as análises. A Tabela 1 mostra os dados cadastrais dos postos utilizados. Para cada estação, foram comparados os ajustes obtidos para cada sub-amostra e realizados os testes estatísticos de paralelismo e deslocamento das curvas obtidas, para dois níveis de significância e com o ponto de divisão da amostra em 1970.

Tabela 1. Dados cadastrais das estações pluviométricas utilizadas no estudo.

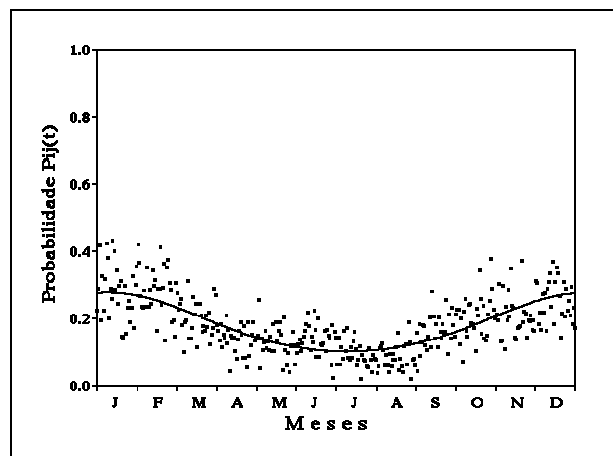
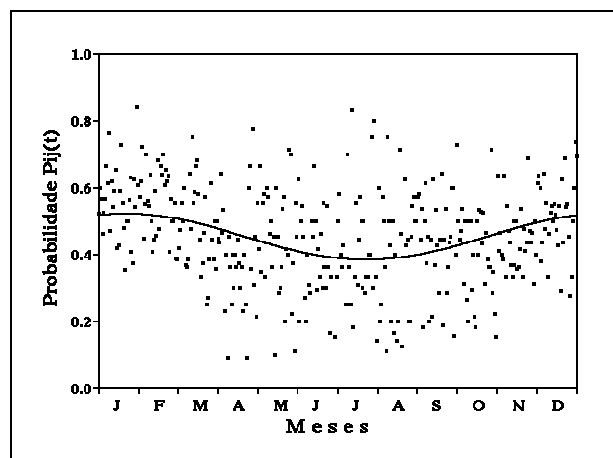
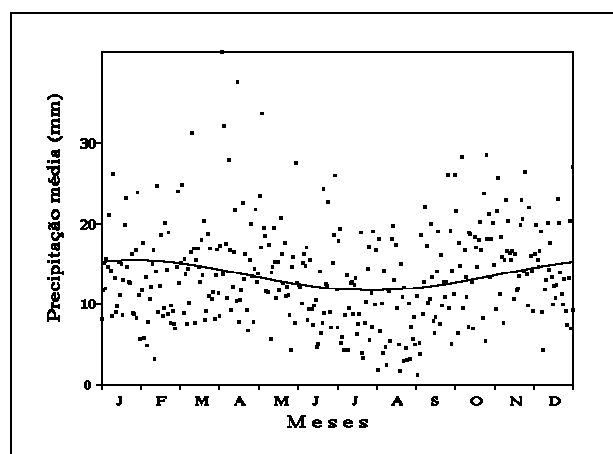
Código	Nome	Latitude	Longitude	Período
02349028	Jacarezinho	23° 10'	49° 58'	05/41 a 12/94
02349032	Siqueira Campos	23° 41'	49° 51'	01/37 a 12/80
02349033	Tomazina	23° 46'	49° 57'	07/37 a 09/96
02349034	Wenceslau Braz	23° 52'	49° 48'	01/39 a 12/82
02349066	Ponte Melo Peixoto	23° 01'	49° 54'	01/39 a 12/80
02350000	Jataizinho	23° 15'	50° 59'	01/38 a 07/96
02350001	Porto Santa Terezinha	23° 06'	50° 27'	10/45 a 08/96
02350002	Andirá	23° 05'	50° 17'	10/45 a 09/96
02350015	Usina Apucarantina	23° 44'	50° 55'	08/48 a 12/95
02350016	Usina Cachoeira - Salto Fiu	23° 44'	50° 57'	07/55 a 12/95
02351010	Londrina	23° 19'	51° 09'	01/46 a 11/92
02352000	Porto Paraíso do Norte	23° 19'	52° 40'	03/53 a 09/96
02353001	Santa Izabel do Ivaí	23° 00'	53° 11'	10/57 a 07/96
02449006	Balsa do Cerro Azul	24° 48'	49° 17'	01/56 a 06/96
02449007	Turvo - Cerro Azul	24° 45'	49° 20'	02/46 a 07/96
02449008	Cerro Azul	24° 49'	49° 16'	12/38 a 07/96
02449019	Jaguariaíva	24° 14'	49° 42'	10/17 a 08/96
02450002	Tibagi	24° 30'	50° 24'	04/38 a 12/95
02450005	Telêmaco Borba	24° 17'	50° 38'	04/49 a 09/96
02450006	Usina Mauá (Salto Mauá)	24° 03'	50° 42'	10/45 a 09/96
02450009	Bom Jardim	24° 42'	50° 29'	08/41 a 09/96
02450010	Castro	24° 47'	50° 00'	01/25 a 12/94
02451014	Ivaiporã	24° 15'	51° 39'	10/55 a 08/96
02454000	Porto Mendes Gonçalves	24° 30'	54° 20'	01/41 a 08/96
02454001	Guaira (Porto Guaira)	24° 04'	54° 15'	01/41 a 08/96
02550000	Rio dos Patos	25° 12'	50° 56'	05/37 a 12/95
02550002	Sumidouro (Usina São Jorge)	25° 02'	50° 05'	03/42 a 12/95
02550003	Santa Cruz	25° 12'	50° 09'	01/45 a 09/96
02550006	Ivaí	25° 01'	50° 51'	12/12 a 03/96
02550024	Ponta Grossa - Vila Velha	25° 13'	50° 01'	01/54 a 12/95
02551000	Guarapuava	25° 27'	51° 27'	03/53 a 09/96
02551004	Santa Clara	25° 38'	51° 58'	08/49 a 12/95
02551006	Guarapuava - 83834	25° 24'	51° 28'	01/25 a 01/85
02552000	Campo Novo	25° 28'	52° 54'	08/49 a 09/96
02552005	Usina Cavernoso	25° 28'	52° 12'	01/52 a 12/95
02554000	Foz do Iguaçu (Ponto Oficial)	25° 33'	54° 35'	01/40 a 10/90
02554001	Parque Nacional do Iguaçu	25° 37'	54° 29'	01/41 a 07/96
02554004	Foz do Iguaçu - 83826	25° 33'	54° 34'	01/38 a 04/81

As Figuras 1 a 8 mostram alguns exemplos dos resultados das análises da precipitação diária, referentes ao posto pluviométrico Jataizinho (02350000). São apresentados os gráficos dos ajustes obtidos para cada parâmetro analisado e os gráficos com a comparação dos resultados dos modelos para cada um dos sub-períodos da amostra. A Tabela 2 mostra os resultados dos testes estatísticos de paralelismo e deslocamento, para os níveis de

confiança de 1% e 5%. A convenção utilizada é a seguinte: p_0 representa a variável $p_0(t)$, p_1 representa a variável $p_1(t)$, m_0 representa a variável $\mu_0(t)$ e m_1 representa a variável $\mu_1(t)$. Os elementos da tabela iguais a “0” indicam que os ajustes obtidos para as duas sub-séries podem ser considerados equivalentes e os valores iguais a “1” indicam que os modelos ajustados para cada trecho da série não podem ser considerados equivalentes.

Tabela 2. Resultados dos testes estatísticos de igualdade dos modelos.

Código	1 %				5 %			
	p_0	p_1	m_0	m_1	p_0	p_1	m_0	m_1
02349028	1	0	1	0	1	0	1	0
02349032	1	1	1	1	1	1	1	1
02349033	1	1	1	0	1	1	1	1
02349034	1	1	1	1	1	1	1	1
02349066	1	0	1	1	1	1	1	1
02350000	1	1	0	0	1	1	0	1
02350001	1	1	0	0	1	1	1	1
02350002	1	1	0	1	1	1	0	1
02350015	1	0	1	0	1	1	1	0
02350016	0	1	0	0	0	1	0	1
02351010	1	1	1	0	1	1	1	0
02352000	1	0	1	0	1	1	1	1
02353001	1	1	1	1	1	1	1	1
02449006	1	1	1	1	1	1	1	1
02449007	1	1	0	0	1	1	0	1
02449008	0	1	1	0	0	1	1	1
02449019	1	1	1	1	1	1	1	1
02450002	1	1	0	1	1	1	0	1
02450005	0	1	0	0	1	1	0	0
02450006	0	0	0	0	0	0	0	0
02450009	1	1	1	1	1	1	1	1
02450010	1	1	1	1	1	1	1	1
02451014	0	1	1	1	1	1	1	1
02454000	1	1	1	1	1	1	1	1
02454001	1	1	1	1	1	1	1	1
02550000	0	0	1	1	0	0	1	1
02550002	1	0	0	0	1	1	0	1
02550003	1	1	0	1	1	1	1	1
02550006	1	1	1	1	1	1	1	1
02550024	0	1	1	1	1	1	1	1
02551000	1	1	1	1	1	1	1	1
02551004	1	1	0	1	1	1	1	1
02551006	1	1	0	0	1	1	0	0
02552000	1	1	1	0	1	1	1	0
02552005	0	1	1	1	1	1	1	1
02554000	1	0	1	1	1	0	1	1
02554001	1	1	1	1	1	1	1	1
02554004	1	1	1	1	1	1	1	1

**Figura 1. Ajuste do modelo.
Frequência - Evento 01 (02350000).****Figura 2. Ajuste do modelo.
Frequência - Evento 11 (02350000).****Figura 3. Ajuste do modelo.
Intensidade - Evento 01 (02350000).**

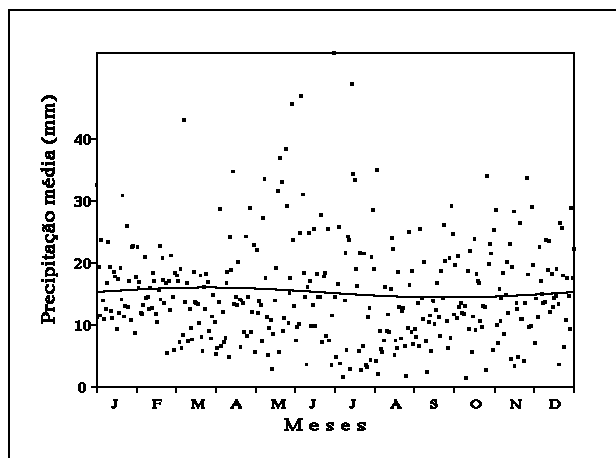


Figura 4. Ajuste do modelo.
Intensidade - Evento 11 (02350000).

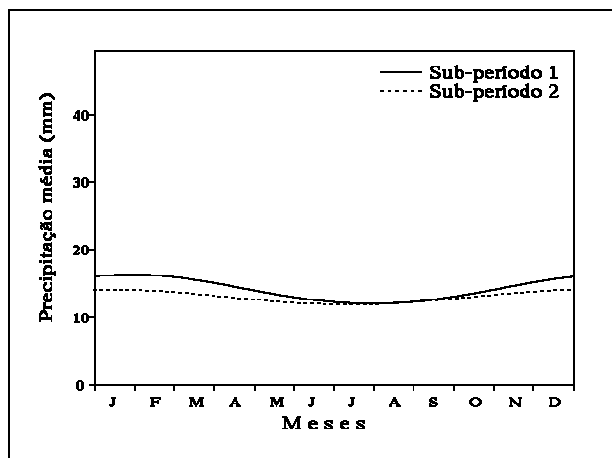


Figura 7. Comparação dos modelos.
Intensidade - Evento 01 (02350000).

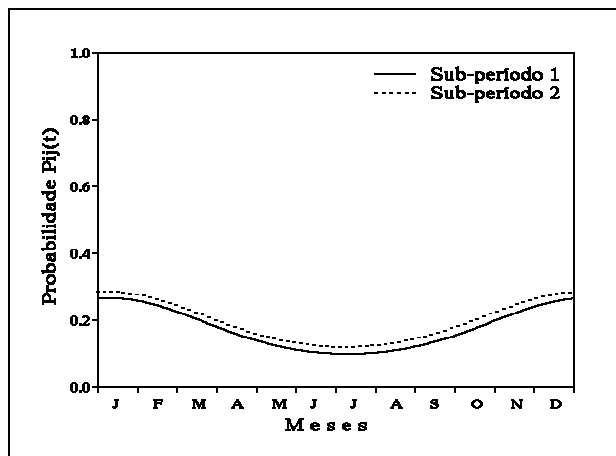


Figura 5. Comparação dos modelos.
Frequência - Evento 01 (02350000).

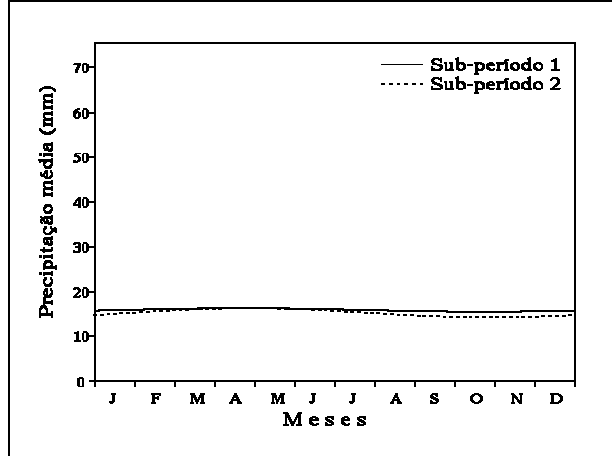


Figura 8. Comparação dos modelos.
Intensidade - Evento 11 (02350000).

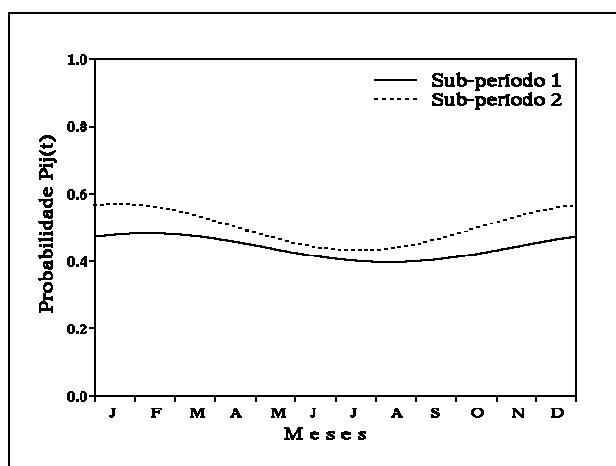


Figura 6. Comparação dos modelos.
Frequência - Evento 11 (02350000).

ANÁLISE DOS RESULTADOS

Apesar do grande número de resultados indicando alteração nas séries de precipitação média diária (Tabela 2), a análise dos gráficos não permite concluir por uma alteração consistente dos parâmetros analisados em uma dada direção, por exemplo, aumento ou diminuição da intensidade de chuva ao longo do tempo para a maioria das estações. Algumas estações apresentam diferenças de comportamento captadas pelos testes no sentido de aumento da frequência e diminuição da intensidade e outras inversamente, pela diminuição de frequência e aumento de intensidade.

O limiar de chuva utilizado nos testes (0,1 mm) é o mesmo utilizado em outros trabalhos encontrados na literatura (Stern e Coe, 1982). Alguns estudos utilizam um limiar igual a zero. Observou-se que o limite mínimo utilizado tem influência nos resultados dos testes. Um limite um pouco maior (1 mm, por exemplo) produz resultados com menor número de rejeições das hipóteses de igualdade das séries.

Cabe observar que, apesar de grande parte das estações apresentarem um período de observação razoavelmente longo, o erro amostral envolvido é bastante grande para alguns parâmetros, particularmente a probabilidade de ocorrência do evento 11 ($p_1(t)$). Isto ocorre devido à análise ser realizada através da estimativa de parâmetros separadamente para cada dia do ano. A grande dispersão observada nos gráficos do evento 11 mostra que seriam necessários períodos observados maiores para a obtenção de estatísticas mais precisas.

Tendo em vista tais dificuldades, realizou-se uma análise visual dos resultados, locando-se no mapa do Paraná os pontos das estações analisadas, juntamente com a seguinte convenção: "0" significa que não foi possível identificar visualmente um comportamento de aumento ou diminuição consistente de determinada grandeza durante todo o ano; "1" significa que foi possível identificar pelos resultados um aumento de frequência dos eventos, acompanhada por uma conseqüente diminuição da intensidade média da chuva; e "2" significa que identificou-se uma diminuição de frequência dos eventos, acompanhada por aumento de intensidade. A Figura 9 mostra estes resultados, para os casos em que as séries foram subdivididas em 1970. Segundo esta abordagem de análise, ocorreu predomínio do comportamento "1" em 53% dos postos. Apenas 8% dos postos apresentou comportamento "2" e em 39% não foi possível estabelecer nenhuma conclusão, ou seja, "0".

CONCLUSÕES

Em resumo, a análise dos resultados foi dificultada pelos seguintes fatores: (i) a grande dispersão apresentada nos gráficos dos eventos analisados, principalmente a frequência e intensidade de ocorrência do evento "11" (dia chuvoso seguido de dia chuvoso); (ii) os gráficos dos ajustes realizados para a obtenção de um padrão sazonal dentro do ano, quando comparados para os dois sub-períodos das séries, nem sempre permitem uma conclusão a respeito do aumento ou diminuição de uma determinada grandeza (frequência ou intensidade), pois este aumento ou diminuição pode ocorrer em uma determinada época do ano, para alguns casos; (iii) a distribuição espacial dos postos analisados não resultou ideal, devido à escassez de postos com dados suficientes para a divisão em duas sub-amostras. A maior parte das estações analisadas localizam-se no nordeste do Estado do Paraná.

A análise visual dos resultados indica que pode haver uma tendência no sentido do aumento de frequência e diminuição da intensidade da precipitação diária para a região. No entanto, um volume maior de dados diários seria necessário para a obtenção de uma conclusão mais segura. A grande variação amostral das estimativas demonstra a necessidade de séries mais longas para a adequada aplicação do método utilizado.

REFERÊNCIAS

- CLARKE, R. T. 1994, *Statistical Modelling in Hydrology*. New York: John Wiley, 412 p.
- CLARKE, R. T. 1998, *Stochastic Processes for Water Scientists: Developments and Applications*. John Wiley & Sons. Inglaterra.
- STERN, R. D., COE, 1982, The Use of Rainfall Models in Agricultural Planning. *Agricultural Meteorology*, v. 26, p. 35-50.
- STERN, R. D., COE, 1984, A Model Fitting Analysis of Daily Rainfall Data. *Journal of the Royal Statistical Society*, v. 147, p. 1-17.

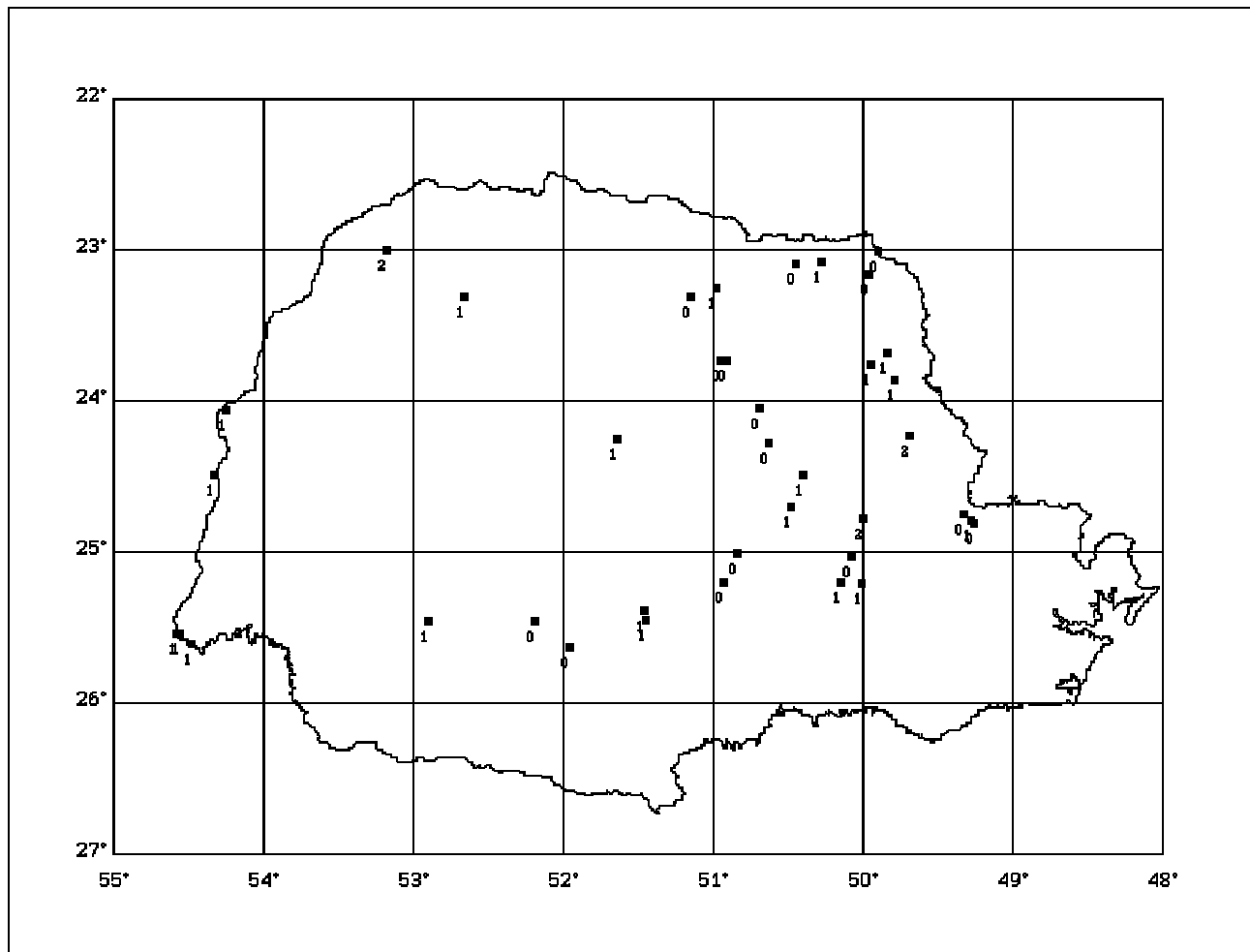


Figura 9. Distribuição espacial dos resultados.

Statistical Modeling of Daily Rainfall Data in Brazil

ABSTRACT

Histograms of daily rainfall data are usually skewed to the right, with greater probabilities concentrated at the origin. This form of distribution suggests that rainfall data might be modeled in two stages, one stage being concerned with modeling the pattern of occurrence of wet and dry days, and the second stage being concerned with the depth of rain falling on wet days. Two generalized linear models, one of which was a logistic regression, were used to model daily rainfall data. It is found that, in many cases, first-order dependence corresponding to a Markov chain gave a satis-

factory model for the occurrence and non-occurrence of rain. In the second stage, a model of rainfall depth was required to be defined for positive values only, and be positively skewed. The gamma distribution was used for this purpose. The method was used to model daily rainfall from rain gauge stations in the Paraná state. The main purpose of the present study was the detection of trends in the occurrence and depth of rainfall in the Paraná river basin.