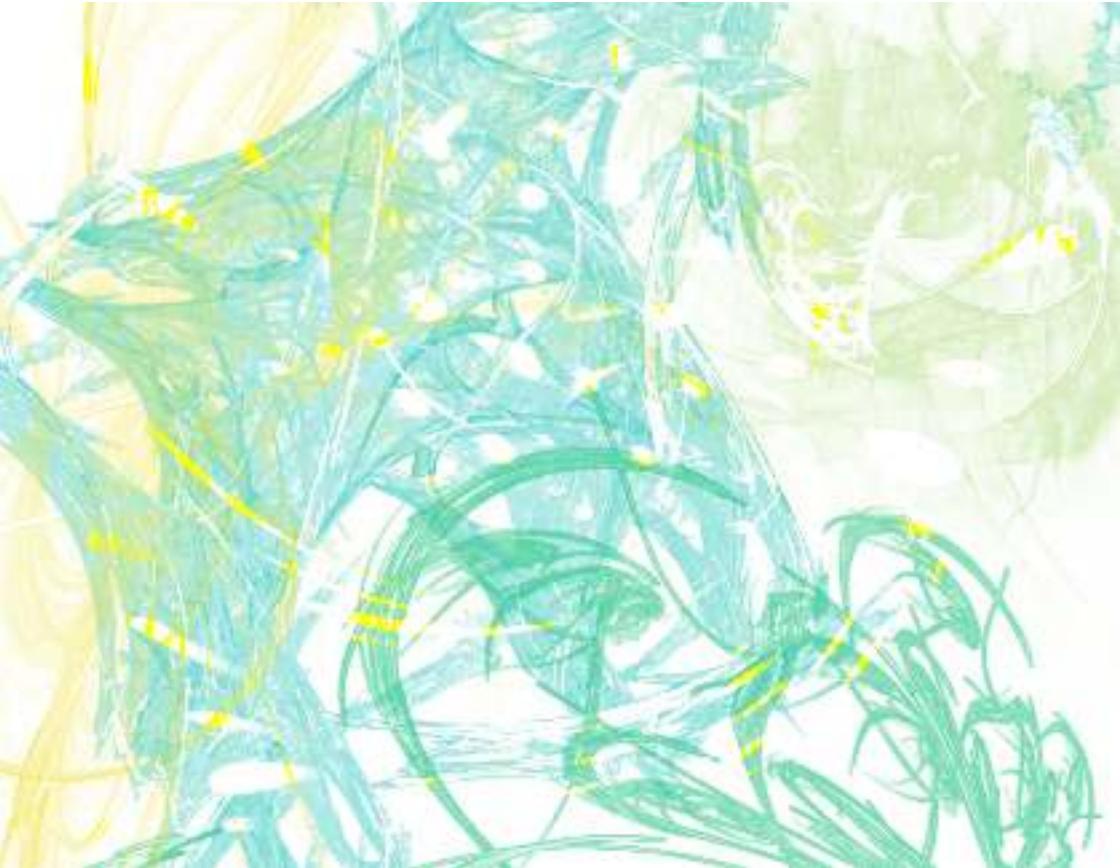


ISSN 1677-9274

Distribuição de Poisson e Sistema de Informações Geográficas: analisando secas mensais





*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

*ISSN 1677-9274
Dezembro, 2008*

Documentos 77

Distribuição de Poisson e Sistema de Informações Geográficas: analisando secas mensais

José Iguelmar Miranda

Embrapa Informática Agropecuária
Campinas, SP
2008

Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)

Av. André Tosello, 209

Cidade Universitária "Zeferino Vaz" – Barão Geraldo

Caixa Postal 6041

13083-970 – Campinas, SP

Telefone (19) 3211-5700 – Fax (19) 3211-5754

URL: <http://www.cnptia.embrapa.br>

e-mail: sac@cnptia.embrapa.br

Comitê de Publicações

Kleber Xavier Sampaio de Souza (presidente)

Marcia Izabel Fugisawa Souza

Martha Delphino Bambini

Sílvia Maria Fonseca Silveira Massruhá

Stanley Robson de Medeiros Oliveira

Suzilei Almeida Carneiro (secretária)

Supervisão editorial: *Suzilei Almeida Carneiro*

Normalização bibliográfica: *Marcia Izabel Fugisawa Souza*

Revisão de texto: *Adriana Farah Gonzalez*

Editoração eletrônica: *Área de Comunicação e Negócios (ACN)*

Suplentes

Goran Neshich

Leandro Henrique Mendonça de Oliveira

Maria Goretti Gurgel Praxedes

1ª edição on-line - 2008

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Miranda, Jose Iguelmar.

Distribuição de Poisson e sistema de informações geográficas : analisando secas mensais / José Iguelmar Miranda. – Campinas : Embrapa Informática Agropecuária, 2008.

24 p. : il. – (Documentos / Embrapa Informática Agropecuária ; 77)

ISSN 1677-9274

1. Modelo estatístico. 2. Distribuição de Poisson. 3. Sistema de informação geográfica. I. Título. II. Série.

CDD – 21st ed. 519.2

© Embrapa 2009

Autor

José Iguelmar Miranda

Ph.D. em Geoprocessamento,

Analista da Embrapa Informática Agropecuária

Av. André Tosello, 209, Barão Geraldo

Caixa Postal 6041 - 13083-970 - Campinas, SP

Telefone: 19-3211-5736

e-mail: miranda@cnptia.embrapa.br

Apresentação

O objetivo deste documento é mostrar o potencial da integração de um sistema de informações geográficas (SIG) com um modelo de probabilidade, usando a distribuição de Poisson, para espacializar variáveis discretas. Modelos estatísticos são ferramentas importantes no estudo de variáveis ambientais, principalmente com a crescente importância da valoração do capital ambiental. A distribuição do Poisson é um bom modelo estatístico para manejo de variáveis discretas, pois mostra seu comportamento. Um passo posterior seria saber como essas variáveis se comportam no espaço, mostrando sua distribuição espacial. Nesse caso, os sistemas de informações geográficas (SIG) são bastante eficientes (Miranda, 2005). Para testar o uso de ambas as ferramentas e mostrar sua eficiência, este trabalho traz uma implementação específica usando uma variável ambiental discreta, secas mensais. A hipótese deste trabalho tinha por base a afirmação de que o déficit hídrico mensal no Estado de São Paulo, definido como um mês com precipitação menor que 10 mm, obedeceria a uma distribuição de Poisson. Para testar a hipótese, a pesquisa usou um conjunto de dados contendo trinta anos de precipitação diária, de 1970 a 1999, obtidos de cento e quarenta e quatro estações climatológicas, distribuídas pelo Estado. Um teste de melhor ajuste com base na distribuição estatística do qui-quadrado, com 95% de confiança (um desvio, $\alpha = 0,05$), foi realizado em todos os resultados, e estes confirmaram a hipótese inicial. O domínio temporal foi a única preocupação nesta primeira parte do experimento. Na segunda parte, buscou-se a interpolação dos valores individuais de probabilidades encontrados para cada estação no domínio espacial, usando um SIG para produzir uma superfície de probabilidade com a distribuição espacial de Poisson. Os resultados se mostraram valiosos e poderiam ser usados, no mínimo, como entrada para um zoneamento agroclimático. A integração do SIG com o modelo de Poisson pode ser estendido a qualquer outra variável ambiental discreta.

Eduardo Delgado Assad
Chefe-Geral

Sumário

Introdução	8
Metodologia	10
Resultados e Discussão	14
Conclusão	20
Referências	21

Distribuição de Poisson e Sistema de Informações Geográficas: analisando secas mensais

José Iguelmar Miranda

Introdução

Dados coletados ao longo dos anos no Estado de São Paulo, antes considerados de uso privado pelas instituições proprietárias das estações, agora se tornaram públicos. A situação era de muitos dados disponíveis, mas sem muita informação. Na realidade, esta é uma característica em muitas áreas da ciência, devido a lacunas existentes entre a geração dos dados e seu entendimento, em que a taxa de produção dos dados é bem maior do que a de seu entendimento. Esforços para entender comportamentos “escondidos” dos dados são bem-vindos, transformando esses **dados** em algo mais útil, **informação**. A matemática e suas ramificações, como a estatística, desempenham um importante papel nesse contexto.

Modelos estatísticos desenvolvidos há muito tempo podem ser úteis no estudo do comportamento dos dados. Atualmente, pode-se contar com pelo menos duas vantagens: a existência de séries temporais, muitas delas disponíveis em endereços eletrônicos, e computadores de alto desempenho para fazerem os cálculos, permitindo o uso dos modelos. Sob essas perspectivas, este trabalho usou uma distribuição de Poisson para estudar o déficit hídrico mensal (DHM) em São Paulo, definido como um período de trinta e um dias com precipitação total menor que 10 mm.

Durante o inverno brasileiro (junho-julho-agosto), é comum a ocorrência de longas séries de dias consecutivos sem precipitação significativa. Essa é uma informação particularmente importante quando as autoridades estaduais estão tratando de planejamento agrícola ou técnicos que precisam de zoneamento agroclimático. Por exemplo, é uma prática comum plantar feijão na região sudoeste do Estado, e uma questão poderia ser mencionada: qual é a probabilidade de se ter, digamos, dois DHM durante esse período? Procedimentos para realizar essa tarefa assumem que: (1) os dados das estações usados representam o universo de amostras disponível a sere tratado, não sendo objetivo deste trabalho coletar dados novos; (2) os dados das estações usados, foram assumidos como confiáveis e não foram sujeitos à verificação.

É importante enfatizar que este trabalho considerou a ocorrência de uma variável discreta aleatória, a saber, a quantidade de DHM dentro de um ano. Existem trabalhos considerando a precipitação como variável contínua aleatória, por exemplo, Wilks & Eggleston (1992), assim como existem trabalhos considerando a precipitação como uma variável discreta aleatória, por exemplo, Schmidt et al. (1990).

A distribuição de probabilidade de Poisson frequentemente pode ser usada como um bom modelo para a distribuição de probabilidade de um número X de eventos raros que ocorrem sem frequência dentro de um *escopo de oportunidades* especificado, como tempo, espaço, volume ou outra dimensão qualquer, considerando apenas o valor médio daqueles eventos. Esse escopo de oportunidades depende do tipo de evento a ser contabilizado. Na maioria dos casos, ele pode ser o *escopo espacial* representando uma região física, por exemplo, a superfície da porta de um carro testada para defeitos de pintura. Ou pode ser um *escopo temporal*, por exemplo, a contagem semanal das ferramentas do estoque para manutenção ou a quantidade de pessoas se encaminhando para uma loja no período de uma hora, etc.

Aplicações da probabilidade de Poisson se estendem por muitas áreas de interesse, à medida que ela fornece um bom modelo para a distribuição de probabilidade dos eventos. Exemplos são transportes (acidentes automobilísticos), indústria (acidentes industriais), física (a quantidade de partículas radioativas que decaem em um período de tempo particular), biologia (probabilidade de mutação em um segmento de DNA, distribuição de espécies ou células) (Mendenhall et al., 1981). Bartlett (1970) aplicou um modelo de Poisson para estudar a distribuição de frequências ecológicas. No escopo espacial, existem muitas aplicações da distribuição de Poisson para estudar o padrão espacial pontual ou dependência espacial (Ripley, 1981; Diggle, 1983; Upton & Fingleton, 1989; Kaiser & Cressie, 1997).

Relativamente à questão do clima, Berlato et al. (2000) usaram uma distribuição de Poisson para calcular a probabilidade sazonal da ocorrência de geadas no Estado do Rio Grande do Sul. De acordo com os autores, as geadas seguiam uma distribuição de Poisson em 80% dos casos estudados. Na sua abordagem, devem-se considerar dois pontos: (1) a distribuição de geadas é extremamente dependente da altitude e isso restringia sua ocorrência a regiões montanhosas específicas ou de relevo ondulado; e (2) não foi possível espacializar a distribuição de Poisson com as linhas de contorno para todo o Estado, principalmente devido à natureza localizada do evento.

Diferentemente, o DHM não está limitado a regiões específicas, ocorrendo em todo o Estado, e como tal, foi possível desenhar as linhas de contorno representando a distribuição espacial das probabilidades de Poisson usando um SIG. Lana & Burgueno (1998) estudaram longos episódios de secas na

região espanhola da Catalunha, com base na distribuição de Poisson. Comparado a outras metodologias, como as cadeias de Markov ou as formulações de Gumbell e Jenkinson, os autores quantificaram as probabilidades com respeito a longos episódios de dias secos. Eles concluíram que os resultados concordavam com estudos de precipitação prévios e com a diversidade espacial da região, dada sua complexidade orográfica.

Metodologia

A presente pesquisa usou uma série temporal de trinta anos de precipitação diária com dados do Departamento de Água e Energia Elétrica, Secretaria de Energia, Recursos Hídricos e Saneamento, Governo do Estado de São Paulo,??? no período entre janeiro de 1970 e dezembro de 1999, coletados em cento e quarenta e quatro estações. Os dados disponíveis estão hospedados no endereço <www.sigrh.sp.gov.br>. Os dados brutos, disponíveis no formato de texto, necessitaram de filtragem e edição. Nem todas as séries temporais estavam completas para o período considerado, e em algumas estações ocorreram casos de ausências, devidamente considerados no cálculo das médias. A distribuição de chuvas ao longo do ano no Estado segue um padrão, com os meses mais chuvosos entre outubro e março, quando normalmente chove acima da média anual. A Fig. 1 mostra um gráfico de linha com a média mensal de chuva durante o período dos trinta anos para Aguai, onde a média geral foi de 128,8 mm.

A **Figura 2** mostra a moda do DHM, por mês, para todas as cento e quarenta e

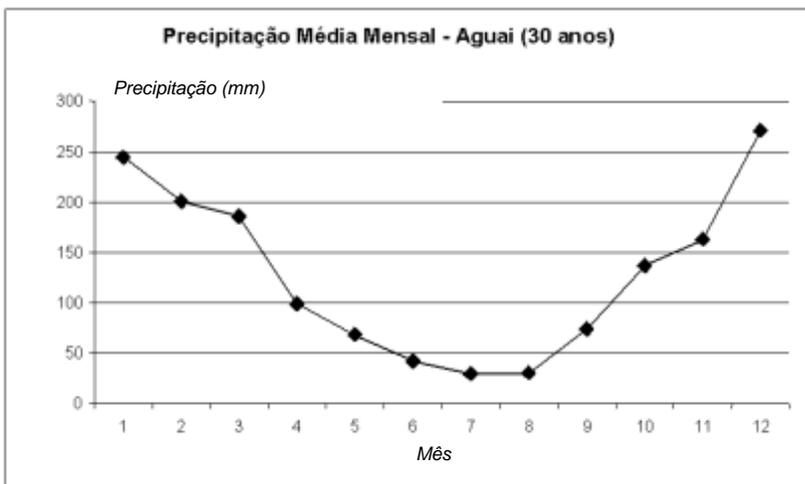


Figura 1. Precipitação média mensal para Aguai; período de trinta anos.

quatro estações no período considerado. O período seco cobre os meses entre junho e agosto, sendo outros valores não significativos.

A **Tabela 1** mostra algumas estatísticas básicas. Valores médios estão bem

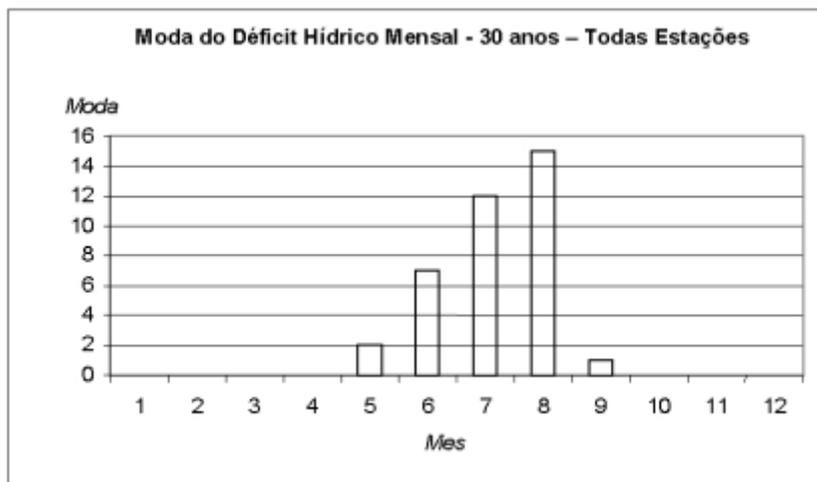


Figura 2. Moda para as cento e quarenta e quatro estações durante trinta anos.

próximos, mas os valores de máxima e moda mostram que julho e agosto concentram os períodos secos mensais mais altos.

Começando com a amostra, um modelo busca a estimativa de como um

Tabela 1. Número de déficit hídrico mensal (período de trinta anos).

	Junho	Julho	Agosto
Máximo	13	18	21
Moda	7	12	15
Média	7,08	8,77	11,35

fenômeno se comporta em locais não enumerados pela amostra. Ele permite alguém estimar valores não disponíveis, descrevendo como os dados da amostra foram criados. Um modelo *determinístico* fornece uma figura razoavelmente completa de como os dados foram criados. Um modelo *probabilístico* ou *estocástico* é usado quando não se sabe quais processos criaram os dados da amostra, ou se ele é conhecido, ou quando não se sabe o suficiente para quantificá-los.

Esse tipo de modelo permite ao analista tratar os dados da amostra como resultado de um processo aleatório, não implicando que o processo seja aleatório, mas que não se sabe ou se ignora acerca de sua natureza verdadeira. A natureza cria mecanismos, físicos ou biológicos, que geram observações não previsíveis com certeza. Tais eventos não podem ser previstos, mas a frequência relativa com a qual eles ocorrem em uma longa série de tentativas é normalmente estável. Eventos que possuem essa característica são chamados de *eventos aleatórios* ou *estocásticos*. A frequência estável relativa providencia uma medida intuitiva para um evento aleatório ocorrer em uma observação futura. É impossível prever com certeza onde um raio vai cair, mas pode-se dizer com uma razoável medida de confiança sobre a probabilidade desse evento (Ripley, 1981).

Uma *variável aleatória*, X , é uma variável cujos valores são gerados de maneira aleatória de acordo com um mecanismo estocástico. A variável aleatória pode assumir valores *discretos* ou *contínuos*. Uma variável aleatória X é dita discreta se assume apenas um número contável de valores e contínua se assume um número incontável de valores. A letra maiúscula X denota uma variável aleatória e a minúscula, x , um valor particular que a variável pode assumir, sendo sua probabilidade $p(X = x)$ ou simplesmente, $p(x)$. Cada saída de uma variável aleatória tem uma probabilidade associada, p , de maneira que:

$$p(x) = u \mid u \quad [0, 1]$$

O conjunto de saída de uma variável aleatória é denominado sua *lei de probabilidade* ou *distribuição de probabilidade*. Uma distribuição de probabilidade pode ser representada por uma fórmula, uma tabela ou um gráfico, que fornecem a probabilidade $p(x)$ correspondente a cada um dos valores de x . Suponha que a variável aleatória X assuma valores inteiros, $X = \{0, 1, 2, \dots\}$, então, a função massa de probabilidade fornece a probabilidade que uma variável aleatória X terá um valor particular x : $p(x) = Pr\{X = x\}$.

Existem poucas distribuições universais que ocorrem numa variedade surpreendente de problemas. As três principais distribuições, com ramificações na teoria das probabilidades, são a binomial, a normal e a de Poisson (Feller, 1968). O interesse aqui é estudar os eventos de DHM usando a distribuição de Poisson. A função massa de probabilidade de uma variável aleatória de Poisson com média μ é (Mendenhall et al., 1981):

$$p(x; \mu) = e^{-\mu} \frac{\mu^x}{x!}$$

A distribuição de Poisson é uma distribuição discreta de um parâmetro que assume valores inteiros não negativos. O parâmetro, μ , é tanto a média como a variância da distribuição. Portanto, à medida que os números de uma

amostra aleatória particular de Poisson crescem, da mesma maneira cresce a variabilidade dos números. Como Poisson mostrou em 1837, sua distribuição é o caso limite de uma distribuição binomial. As distribuições de Poisson e exponencial são relacionadas. Se a quantidade contabilizada segue uma distribuição de Poisson, então o intervalo entre as contagens individuais segue a distribuição exponencial. Diferentemente da binomial, o modelo de Poisson nunca se torna simétrico. Na realidade, ele é sempre deslocado para a direita (Shapiro & Gross, 1981). Com uma amostra de dados é possível ajustar uma distribuição de Poisson encontrando o valor do parâmetro μ que faz a equação acima se comportar similarmente para a mesma amostra. A estimativa do parâmetro μ é feita pelo método dos momentos.

Existem dois pontos importantes acerca da equação de Poisson que coincidem com as necessidades do presente trabalho (Shapiro & Gross, 1981): (1) a equação descreve as probabilidades de ocorrências aleatórias; e (2) a equação se aplica a “intervalos” nos eixos espaciais ou temporais. No caso presente, a sequência de eventos aleatórios são os DHM, ocorrendo no tempo, isto é, no ano. Cada evento é representado por um ponto no eixo do tempo, e o objetivo é encontrar a chance de distribuição dos pontos. A suposição física mais simples considera $p(x;\mu)$ como a probabilidade de encontrar exatamente x pontos (eventos = DHM) dentro de um intervalo fixo de tempo (um ano).

Uma suposição acerca dos eventos individuais é que eles sejam independentes ou que suas probabilidades de ocorrência em um dado intervalo dependam apenas do comprimento do intervalo onde eles serão contados. Essa afirmação de independência é difícil de demonstrar, mas a distribuição de Poisson pode produzir uma representação útil, considerando que o grau de dependência não seja muito forte (Wilks, 1995). A teoria da probabilidade em um contínuo torna possível expressar essa suposição diretamente, mas restrita a probabilidades discretas, sendo usado um modelo finito aproximado (Feller, 1968). Os passos para se obter os índices de Poisson são:

1. Totalizar, por mês, os dados diários brutos para cada estação, gerando uma matriz onde as colunas representam a precipitação total do mês e as linhas representam os anos.
2. Calcular novos valores binários para as células da matriz, de acordo com a seguinte regra:

Se (valor_da_celula < 10,0) então

 Novo_valor_da_celula = 1

Senão

 Novo_valor_da_celula = 0

3. Contar o número de ocorrências de DHM, percorrendo as linhas na matriz binária.
4. Calcular os valores médios das ocorrências de DHM. Esse valor será usado na fórmula da distribuição de Poisson.
5. Calcular os valores de probabilidade de Poisson para cada classe de DHM.

Resultados e Discussão

Devido ao fato de não ser possível mostrar todos os cento e quarenta e quatro resultados das estações, a Tabela 2 é um exemplo de ocorrência de DHM para Aguai.

A taxa média desses eventos foi 1,2 DHM por ano. A Fig. 3 mostra um gráfico com a frequência relativa da distribuição de DHM e a correspondente distribuição de Poisson. As probabilidades de Poisson aparecem como uma

Tabela 2. Ocorrência de DHM em Aguai (1970 – 1999).

Classe	Observado
0	6
1	16
2	5
3	2
4	1

linha, mas na realidade, elas são valores discretos. Com a distribuição ajustada, é possível agora calcular a probabilidade com que um número particular de eventos de DHM ocorra em Aguai anualmente.

A distribuição de Poisson permite a definição de probabilidades suavizadas, embora os dados brutos sejam discretos. Ninguém perguntará pela probabilidade de ter um e meio déficit hídrico, mas pode-se perguntar pela probabilidade de ter mais de dois desses eventos, por exemplo.

Como mostrado na Fig. 3, os dois gráficos são bastante aproximados. A distribuição ajustada aos dados brutos desempenha um importante papel no fato de que agora se pode ter idéia da transição das variações. O coeficiente de correlação entre os dois conjuntos de dados atingiu 0,87. Com a realização desses passos, a primeira parte do experimento foi finalizada.

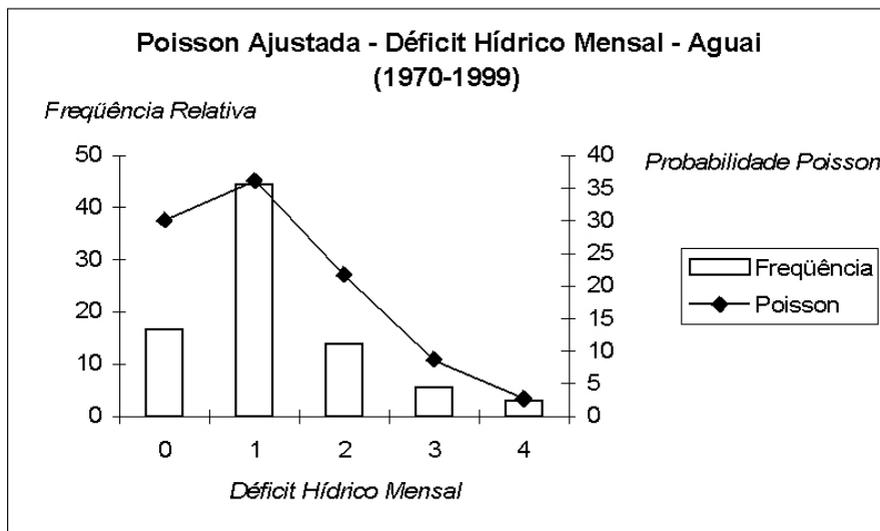


Figura 3. Déficit hídrico mensal e a distribuição de Poisson para Aguaí.

Como dito, além de trabalhar com a dimensão tempo, a experiência tinha um segundo objetivo, interpolar os valores de probabilidades individuais encontrados para cada estação na dimensão espacial, usando um SIG para produzir a superfície com a distribuição espacial de Poisson.

A Fig. 4 mostra o mapa interpolado com o SIG IDRISI, onde a distribuição de Poisson enfatiza as áreas com probabilidades de ocorrer dois ou mais DHM anualmente.

Basicamente, o Estado apresenta três regiões de DHM: norte, centro e sul. A região sul do Estado é a menos provável de ocorrerem dois ou mais tais eventos. Os municípios nessa região apresentam uma probabilidade de Poisson menor que 12% de ocorrerem dois ou mais DHM. Mesmo nos meses mais secos, os planejadores podem confiar que as culturas cultivadas nesse período apresentam baixos riscos de seca. A região central é representada pelos municípios com probabilidade de terem dois ou mais DHM entre 12% e 22%. De maneira oposta, a região norte está mais sujeita a tais eventos, com probabilidade superior a 22%, mas limitado a um máximo de 27%. Mesmo assim, não são valores elevados de probabilidade.

Pode-se perguntar, então, quão representativa, pensando espacialmente, é a distribuição das cento e quarenta e quatro estações? Para responder essa questão, existe uma função no SIG que divide o espaço em polígonos, conhecidos como polígonos ou diagramas de Voronoi (Miranda,



Figura 4. Padrão espacial da distribuição de Poisson.

2005). Trata-se de estruturas de dados da geometria computacional e são importantes em muitas áreas, como matemática e ciências naturais. Dado um conjunto de pontos, o espaço é particionado em seu entorno em células, representando a área do espaço que está mais próxima ao ponto associado, do que a qualquer outro ponto.

Polígonos de Voronoi para um conjunto de pontos são criados com as bisseções perpendiculares dos segmentos, ligando um ponto àqueles mais próximos dele. Os polígonos de Voronoi são usados principalmente para problemas de definição de adjacências espaciais e de vizinho mais próximo (Alani et al., 2001). Quando se usa um conjunto discreto de pontos para modelar a vizinhança espacial, como nesse caso, o diagrama obtido é conhecido como diagrama *ordinário* de Voronoi. Esses diagramas modelam de maneira única a proximidade espacial dos pontos. Diagramas ordinários de Voronoi são informativos e significativos apenas quando os pontos possuem pesos iguais ou quando o vizinho mais próximo é a única relação a ser considerada.

Quando se necessita modelar a proximidade espacial que dependa dos k -ésimos vizinhos mais próximos, com pontos contendo pesos diferentes, é preciso usar os diagramas *generalizados* de Voronoi (Lee & Gahegan, 2000).

O módulo do IDRISI que calcula os diagramas ordinários de Voronoi é o THIESSEN (Fig. 5). Esse módulo constrói os diagramas ao redor de um conjunto de pontos, neste caso, as estações climatológicas. Os polígonos de Voronoi ou Thiessen dividem o espaço de tal maneira que cada localização é atribuída ao ponto de controle mais próximo. Os polígonos definem regiões que são dominadas por cada ponto. Maiores informações de como esses polígonos são construídos podem ser encontradas em Miranda (2005).

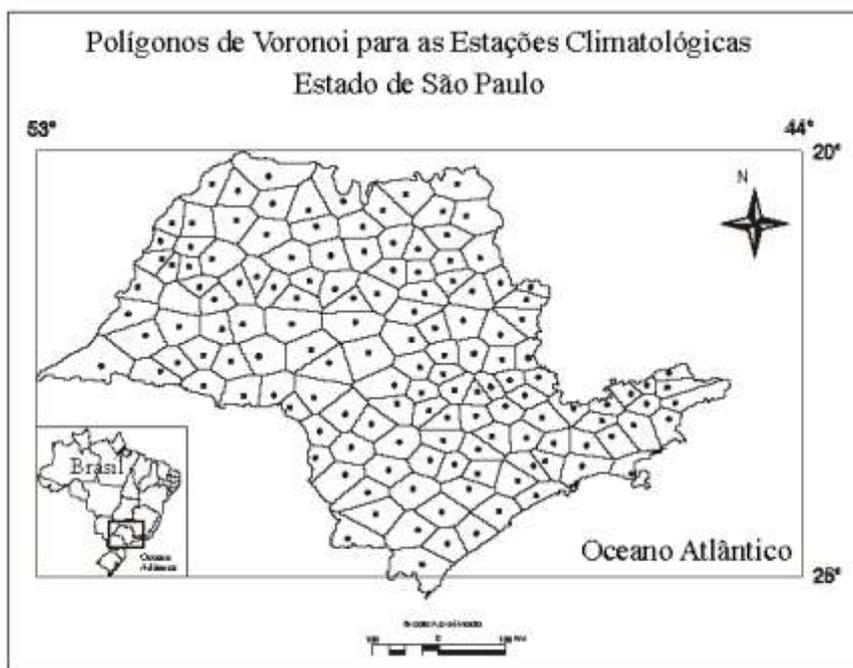


Figura 5. Padrão espacial mostrando os 144 polígonos de Voronoi.

A Tabela 3 mostra alguns dados acerca desses polígonos. O tamanho de área máxima encontrada foi de 476.708,10 ha, correspondendo a uma área na região oeste do Estado, Teodoro Sampaio e outros municípios. O menor polígono contabilizou 57.019,69 ha. O tamanho médio é de 186.152,90 ha, com desvio padrão de 64.684,82 ha. Naturalmente, aumentando a rede de estações, melhores resultados seriam obtidos. Existe um sério problema com a quantidade de valores perdidos para grande parte das estações, séries temporais incompletas e valores não auditados, tornado as estatísticas

menos confiáveis.

Uma preocupação adicional quando se ajusta distribuições a dados observados é avaliar o melhor ajuste ou teste de hipótese, da distribuição aos dados usados. Desta maneira, pode-se afirmar que o modelo fornece uma descrição adequada do fenômeno. Como mostrado na Fig. 3, a distribuição ajustada foi sobreposta ao histograma dos dados e, visualmente, pode-se verificar que não existe diferença acentuada dos dados. Embora isso não seja uma hipótese formal para testar o modelo, pode ser aceito.

Tabela 3. Estatísticas dos polígonos de Voronoi.

Tamanho da área	Estatísticas (ha)
Média	186.152,90
Mínimo	57.019,69
Máximo	476.708,10
Desvio Padrão	64.684,82

Para ser mais formal no teste do modelo, um teste de melhor ajuste, com base na distribuição de probabilidade do qui-quadrado, testou a hipótese do trabalho, de que o déficit hídrico mensal no Estado de São Paulo obedecia a uma distribuição de Poisson. Esse teste essencialmente compara o histograma dos dados com a distribuição de probabilidade, uma vez que as variáveis aleatórias são discretas. O teste qui-quadrado funciona melhor para variáveis aleatórias discretas, pois, para implementá-lo, o intervalo dos dados deve ser dividido em classes discretas (Wilks, 1995). O teste envolve a contagem dos valores em cada classe com relação às probabilidades calculadas:

O numerador indica que, em cada classe, o número de valores dos dados esperados, $E(n_i)$, de acordo com a distribuição ajustada, é a probabilidade de ocorrência naquela classe multiplicado pelo tamanho da amostra, np_i . Se a distribuição ajustada estiver próxima dos dados, os valores esperados e observados estarão próximos para cada classe, e o valor da diferença quadrática no numerador será pequeno. Se o ajuste não for bom, a discrepância será enfatizada pelo quadrado no numerador, produzindo grandes valores do qui-quadrado.

$$\chi^2 = \sum_{i=1}^k \frac{[n_i - E(n_i)]^2}{E(n_i)} = \sum_{i=1}^k \frac{[n_i - np_i]^2}{np_i}$$

Para usar esse teste, é recomendado um mínimo de cinco eventos por classe. Como os números para os déficits foram poucos, em muitas situações os números de classes foram apenas três, devido ao agrupamento dos poucos números. Originalmente, o número de classes com déficit hídrico variava entre 0 e 4. Apenas duas das estações registraram mais de quatro déficits hídricos por ano, totalizando cinco. Para ser consistente, todos os valores foram agrupados com apenas três classes. Sob essa suposição, o número de graus de liberdade foi 1, uma vez que apenas um parâmetro, a média, foi ajustado. A hipótese nula assumida, H_0 , afirmava que a variável aleatória X (ocorrência de déficit hídrico) obedecia a uma distribuição de Poisson em cada uma das cento e quarenta e quatro estações, supondo os eventos independentes, com $\alpha = 0,05$ ou 95% de confiança. O valor máximo do qui-quadrado foi 0,977, para a estação de Taciba, uma vez que:

$$\chi^2_{1,0,0,5} = 3.841$$

A hipótese nula não foi rejeitada, significando que os dados não apresentam evidência suficiente para contradizer a hipótese. As vantagens de se usar um modelo podem ser descritas como síntese e interpolação (Wilks, 1995). Síntese ajuda a reduzir o número de quantidades requeridas para a estatística de ordem; dessa maneira, não é necessário usar uma grande base de dados para antecipar o comportamento dos dados. A interpolação ajuda a *recuperar* falhas encontradas no processo de amostragem dos dados reais.

É importante enfatizar a disponibilidade crescente de dados climatológicos no Brasil através da Internet. Esse é um passo importante para a democratização dos dados coletados por vários centros de pesquisa. Embora a facilidade de obter dados esteja crescendo, o desafio ainda continua sendo de produzir informações a partir do seu uso, um dos propósitos deste trabalho.

Conclusão

O uso da distribuição de Poisson para ajustar dados do déficit hídrico mensal no Estado de São Paulo se mostrou plausível, confirmando a hipótese do trabalho. Com isso, é possível o uso desse modelo para calcular a probabilidade da ocorrência de um número particular de eventos de DHM que possam ocorrer anualmente em qualquer região do Estado. Como os dados de cada estação foram testados individualmente pelo melhor ajuste, não é possível produzir mapas automáticos com os contornos de probabilidades de Poisson. É necessário um dispendioso trabalho para se obter as saídas.

O SIG se mostrou uma tecnologia útil para a espacialização da distribuição de Poisson, permitindo ao analista ter uma visão geral acerca da regionalização dos dados. A integração de um SIG com a distribuição de Poisson pode ser estendida para outros tipos de variáveis ambientais discretas.

Referências Bibliográficas

ALANI, H.; JONES, C. B.; TUDHOPE, D. Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science*, Leicester, v. 15, n. 12, p. 287-306, 2001.

BARTLETT, M. S. *Stochastic population models*. London: Methuen, 1970. 90 p.

BERLATO, M. A. de; MELO, R. W.; FONTANA, D. C. Risco de ocorrência de granizo no Estado do Rio Grande do Sul. *Revista Brasileira de Agrometeorologia*, Santa Maria, v.8, n.1, p. 121-132, 2000.

DIGGLE, P. J. *Statistical analysis of spatial point patterns*. London: Academic Press, 1983. 148 p.

FELLER, W. *An introduction to probability theory and its application*. New York, NY: John Wiley, 1968. 509 p.

KAISER, M. S.; CRESSIE, N. Modeling Poisson variables with positive spatial dependence. *Statistics & Probabilities Letters*, Madison, n. 35, p. 423-432, 1997.

LANA, X.; BURGUENO, A. Probabilities of repeated long dry episodes based on the Poisson distribution: an example for Catalonia (NE Spain). *Theoretical and Applied Climatology*, Berlin, n. 60, p.111-120, 1998.

LEE, I.; GAHEGAN, M. What-if analysis for point data sets using generalised Voronoi diagrams. In: INTERNATINAL CONFERENCE ON GEOCOMPUTATION, 5., 2000, Chattam Maritime, UK. *Conference proceedings*. [Chattam Maritime: University of Greenwich, 2000]. Disponível em : <<http://www.geocomputation.org/2000/GC049/Gc049.htm>>. Acesso em: 10 out. 2008.

MENDENHALL, W.; SCHEAFFER, R. L.; WACKERLY, D. D. *Mathematical statistics with applications*. 2nd ed. Boston: Duxbury Press, 1981. 686 p.

MIRANDA, J. *Fundamentos de sistemas de informações geográficas*. Brasília, DF: Embrapa Informação Tecnológica, 2005. 425 p.

RIPLEY, B. D. *Spatial statistics*. New York, NY: John Wiley, 1981. 252 p.

SCHMIDT, G. M.; SMAJSTRLA, A. G.; ZAZUETA, F. S. A *nonhomogeneous parameter estimation methodology for stochastic rainfall models*. St. Joseph, MI: ASAE, [1990]. 13 p. (ASAE. Paper, 90-2019). Paper presented at The 1989 Winter Meeting American Society of Agricultural Engineers, Columbus, Ohio, 1990.

SHAPIRO, S. S.; GROSS, A. J. *Statistical modeling techniques*. New York, NY: Marcel Dekker, 1981. 367 p.

UPTON, G. J. G.; FINGLETON, B. *Spatial data analysis by example: point pattern and quantitative data*. Chichester, NY: John Wiley, 1989. 409 p.

WILKS, D. S. *Statistical methods in the atmospheric sciences: an introduction*. San Diego, CA: Academic Press, 1995. 467 p.

WILKS, D. S.; EGGLESTON, K. L. Estimating monthly and seasonal precipitation distributions using the 30- and 90-day outlooks. *Journal of Climate*, Fort Collins, n. 5, p. 252-259, 1992.



Informática Agropecuária

**Ministério da
Agricultura, Pecuária
e Abastecimento**

