

USO DO PROGRAMA LIVRE “R” NA ESTRATIFICAÇÃO DOS MUNICÍPIOS PRODUTORES DE PÊSSEGO NO RIO GRANDE DO SUL

Fernanda P. Marion¹, Alfredo J. B. Luiz², Aline H. N. Maia³, Marcos C. Neves⁴

Resumo

Informações precisas e atualizadas sobre a produção agrícola brasileira são importantes e estratégicas, tanto do ponto de vista econômico quanto em relação à segurança alimentar. A previsão oficial de safras no Brasil ainda é realizada de forma subjetiva, mas existe um esforço da Companhia Nacional de Abastecimento (Conab) para o aperfeiçoamento metodológico do atual sistema de previsão. Este trabalho pretende contribuir para esse propósito, apresentando um método de estratificação eficiente para fins de estimação de área plantada, informação fundamental para a previsão de safras. Foi utilizando o pacote computacional estatístico R, software livre e de código aberto. Como exemplo de aplicação, foi elaborado um algoritmo para estratificação dos municípios do Rio Grande do Sul, para estimação da área plantada com o pessegueiro no estado. Foi aplicado um método hierárquico de agrupamento, utilizando a função HCLUST do programa R, que resultou na formação de oito estratos sendo um deles o ‘estrato certo’, onde todos os municípios serão amostrados.

Palavras-chave: Estratificação, programa R, previsão de safra.

USING THE FREE SOFTWARE R FOR STRATIFYING PEACH- PRODUCING COUNTIES IN THE RIO GRANDE DO SUL STATE, BRAZIL

Abstract

We developed an algorithm for counties stratification using the R function HCLUST, useful for designing efficient sampling plans for estimating planted area, key information for yield prediction. The algorithm was applied for stratifying counties at Rio Grande do Sul, Brazil, with the purpose of estimating peach planted area.

Key words: Stratification, R project, yield forecast.

¹ Graduanda, Faculdade de Estatística, Unicamp, Campinas, SP. fernanda.marion@gmail.com

² Eng. Agrônomo, Doutor, Embrapa Meio Ambiente, Jaguariúna, SP. alfredo@cnpma.embrapa.br

³ Eng. Agrônoma, Doutora, Embrapa Meio Ambiente, Jaguariúna, SP. ahmaia@cnpma.embrapa.br

⁴ Eng. Eletricista, Embrapa Meio Ambiente, Jaguariúna, SP. marcos@cnpma.embrapa.br

Introdução

A coleta de dados para previsão de safra pode ser feita de forma objetiva, através da mensuração de variáveis de interesse ou de forma subjetiva, através de informações fornecidas por produtores, técnicos ou outros agentes que detêm conhecimento sobre a produção agrícola nas regiões investigadas. Atualmente, a previsão de safras no Brasil é feita de forma subjetiva, gerando assim estimativas sem quantificação de incertezas associadas. Quando são utilizados métodos de amostragem probabilística, é possível avaliar a qualidade das estimativas, via medidas de incerteza tais com variância, erro padrão e intervalos de confiança.

O GeoSafras – "Projeto de Aperfeiçoamento Metodológico do Sistema de Previsão de Safras no Brasil", iniciado em 2004, pretende desenvolver novos métodos para previsão de safras agrícolas, de modo que seja possível avaliar a qualidade das estimativas geradas. Esses novos métodos estão fundamentados na utilização de conjunta de geotecnologias e métodos de amostragem probabilística. Utilizando essa abordagem, Luiz (2003), propôs um método para estimação da área plantada de uma determinada cultura, onde inicialmente são selecionados municípios a serem amostrados e posteriormente, pontos (pixels) dentro dos municípios selecionados que serão utilizados para o cálculo da estimativa de interesse. O presente trabalho baseou-se em LUIZ et al (2006), onde foi proposto um delineamento amostral para estimação de área plantada para café em São Paulo e para grãos em Minas Gerais, utilizado a função FASTCLUS do programa proprietário SAS® (SAS, 1999). Contudo, para que faça parte do escopo do GeoSafras, é preciso que os métodos propostos sejam implementados utilizando linguagem de programação e softwares de acesso livre (software livre), como é o caso do R. Como exemplo de aplicação, foi elaborada uma rotina para estratificar os municípios gaúchos com o objetivo de estimar a área plantada com pessegueiro no estado, utilizando um método de classificação hierárquica.

Material e Métodos

A estratificação dos municípios foi feita utilizando como variável auxiliar a área plantada com pessegueiro em cada município, no ano de 2004. Para um melhor entendimento da variabilidade e padrões de agregação dos dados de área plantada por

município, foi realizada uma análise exploratória dos dados. Calculou-se o índice de Gini, medida de concentração, que varia entre 0 e 1. O valor nulo significaria que todos os municípios têm a mesma área com pessegueiro; o valor unitário significaria que apenas um município é responsável por toda a área de pêssego do estado.

Para definição dos estratos de municípios, foi utilizando um método de classificação hierárquica (EVERITT, 1980), que consiste em uma série de sucessivas divisões de elementos em grupos, até que a estratificação seja otimizada, de acordo com critérios pré-estabelecidos. A função de estratificação hierárquica do programa R utilizada foi a HCLUST. Esta função equivale à FASTCLUS do programa SAS, utilizada em Luiz et al. (2006).

Inicialmente, foram excluídos os municípios com área plantada inferior a 5 ha. A seguir, foram feitas simulações de agrupamento dos municípios restantes em 4, 5, 6, 7 ou 8 estratos, considerando o critério de máxima variabilidade entre estratos. Foram calculadas as variâncias dentro de cada estrato e a razão (R) entre a máxima e a mínima variância para cada cenário simulado. O número de estratos foi escolhido com base no critério de mínimo R. Após a seleção do número de estratos, os municípios em estratos com menos de quatro elementos foram re-allocados num novo estrato denominado 'estrato certo' (HEDLIN, 2000) e o algoritmo de estratificação executado novamente. O processo de estratificação foi finalizado quando todos os estratos tinham mais que quatro municípios.

Resultados e Discussões

O Índice de Gini obtido para a área plantada com pêssego no Rio Grande do Sul em 2004 foi 0,73, o que aponta uma grande concentração, ou seja, poucos municípios são responsáveis por grande parte da área plantada de todo o estado. A alta variabilidade da área plantada indicou a necessidade de estratificação, para posterior seleção de municípios a serem amostrados na segunda fase. A razão entre a máxima e a mínima variâncias (R) para os diferentes cenários simulados (quatro a oito estratos) variou entre 8,72 (sete estratos) e 37,70 (seis estratos). Foram necessárias três iterações para re-alocar os municípios isolados no estrato certo e obter a estratificação final (oito estratos). O 'estrato certo' constituiu-se dos municípios com maiores áreas plantadas, onde todos serão amostrados na segunda fase do processo de amostragem (seleção de pontos dentro dos

municípios). Os estratos resultantes são tais que a variância da área plantada é a menor possível dentro de cada estrato e a maior possível entre estratos, considerando a restrição de o número de municípios por estrato ser maior que quatro. As variâncias dentro de estratos na estratificação ótima variaram de 5 a 83 mil ha² para os estratos de 1, a 7; no ‘estrato certo’, que inclui os municípios com maior área plantada, a variância foi da ordem de 33 milhões de ha². A estratificação resultante é apresentada na Figura 1.

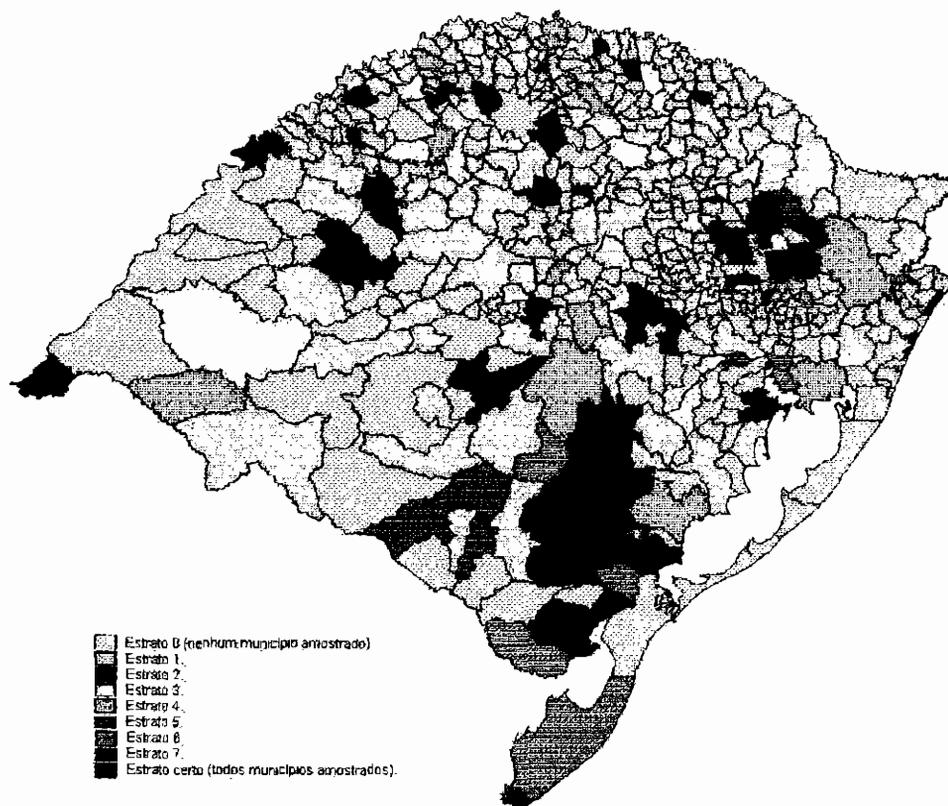


Figura 1. Estratificação dos municípios do Rio Grande do Sul de acordo com a área plantada com pêssigo no ano de 2004.

Conclusões

A estratificação ótima para estimação de área plantada com pêssigo, segundo os critérios utilizados consiste de oito estratos, sendo um deles o ‘estrato certo’, composto pelos municípios com maiores áreas plantadas.

O algoritmo desenvolvido, utilizando a função HCLUST do software livre R, é útil para estratificação de municípios visando reduzir os custos e melhorar a precisão das estimativas em estados com alta variabilidade entre municípios, da área plantada da cultura objeto de investigação.

Bibliografia

AGRITEMPO: Sistema de Monitoramento Agrometeorológico. Disponível em <<http://www.agritempo.gov.br/>>. Acesso em: 20 set. 2006.

EVERITT, B.S. **Cluster Analysis**, Second Edition, London: Heineman Educational Books Ltd, 1980.

HEDLIN, D. A procedure for stratification by an extended Ekman rule. **Journal of Official Statistics**, v.16, n.1, p.15-29, 2000.

IBGE. **Sistema IBGE de Recuperação Automática - SIDRA**. Disponível em: <<http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?z=t&c=1613>>. Acesso em: 30 mar. 2006.

LUIZ, A.J.B. **Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto**. 2003. 116p. Tese (Doutorado) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003.

LUIZ, A. J. B.; NEVES, M. C.; EVANGELISTA, S. R. M.; MAIA, A. H. N. **Sistema para estimação de área plantada por amostragem (SEARA)**: proposta de delineamento amostral para café em São Paulo e para grãos em Minas Gerais. Jaguariúna: Embrapa, 2005. 27 p. (Embrapa Meio Ambiente. Documentos, 49). Disponível em: <http://www.cnpma.embrapa.br/download/documentos_49.pdf>. Acesso em: 30 mar. 2006.

R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2006. Disponível em: <<http://www.R-project.org>>. Acesso em: 15 abr. 2006.

SAS INSTITUTE INC., **SAS/STAT® User's Guide**, Version 8, Cary, NC: SAS Institute Inc., 1999.

MÉTODOS NÃO PARAMÉTRICOS PARA QUANTIFICAÇÃO DE EFEITOS DO FENOMENO EL NINO/OSCILAÇÃO SUL (ENOS) SOBRE RISCOS CLIMÁTICOS

Aline de Holanda N. Maia¹, Holger Meinke² e Sarah Lennox³

RESUMO: Sistemas probabilísticos de previsão climática sazonal (SPCS), acoplados a modelos de simulação de sistemas agrícolas, são utilizados para processo de quantificação de riscos climáticos associados ao negócio agrícola. Uma forma simples de conectar SPCS e modelos de simulação é abordagem de anos análogos (AA) onde séries históricas de variáveis climáticas (Y) são segregadas em classes derivadas do fenômeno El Niño/ Oscilação Sul (ENOS). Nessa abordagem, a informação é usualmente sumarizada em funções de distribuição acumuladas ($F_Y(y)=P(Y\leq y)$), específicas para cada classe do ENOS. Uma das vantagens da abordagem de AA é a incorporação de informações sobre mudanças dos padrões históricos decorrentes do aquecimento global, capturadas pelos índices derivados do ENOS. Desde que haja informação disponível, riscos climáticos específicos para determinadas atividades agrícolas (ex. déficit hídrico), podem ser estimados considerando diferentes classes do ENOS. Neste trabalho, propomos o uso de medidas descritivas e métodos não paramétricos para quantificação da influência do ENOS sobre as F_Y e riscos climáticos e elas associados. Essa abordagem é útil para investigar a variabilidade espaço-temporal do sinal do ENOS, onde o uso de métodos paramétricos exigiria a verificação de pressupostos para um grande número de locais e períodos. Ilustramos a abordagem proposta com exemplos onde foi quantificada a influência do ENOS sobre chuvas trimestrais no Brasil, Argentina e Austrália.

Palavras chave: previsão climática sazonal, riscos climáticos, mudanças climáticas.

NONPARAMETRIC METHODS FOR QUANTIFYING ENSO EFFECTS ON CLIMATIC RISKS

ABSTRACT: We propose a general approach, based on descriptive measures and non-parametric tests, to quantify and communicate influence of ENSO on climatic risks in agriculture. We illustrate this approach by quantifying ENSO influence on 3-monthly rainfall over Australia and Southeast of South America.

Key words: seasonal climate forecasts, climatic risks, climate change

¹ Embrapa Meio Ambiente, Jaguariúna, São Paulo, Brasil. ahmaia@cnpma.embrapa.br.

² Wageningen University, Department of Plant Sciences, Wageningen, The Netherlands. holger.meinke@wur.nl

³ Department of Primary Industries and Fisheries, Toowoomba, Australia. sarah.lennox@dpi.qld.gov.au

1. INTRODUÇÃO

Sistemas probabilísticos para previsão climática sazonal (SPCS), acoplados a modelos de simulação de sistemas agrícolas, são frequentemente utilizados para estimação de riscos climáticos na agricultura, informação fundamental para tomada de decisão pelos produtores e estabelecimento de políticas de crédito agrícola. Riscos climáticos são usualmente estimados a partir de séries históricas das variáveis de interesse, considerando que os padrões temporais permanecem constantes no horizonte de tempo das previsões. No entanto, evidências cada vez mais consistentes de mudanças nos padrões climáticos apontam para a necessidade do uso de métodos de estimação de risco que incorporem indicadores das mudanças climáticas globais. Uma forma simples e intuitiva de considerar tais indicadores é abordagem de anos análogos onde séries históricas de variáveis climáticas Y (ex. chuva, temperatura) são segregadas em sub-séries ou classes derivadas de indicadores climáticos tais como o Índice de Oscilação Sul (IOS), El Niño/ Oscilação Sul (ENSO) ou Temperaturas da Superfície do Mar (TSM) (MEINKE & STONE, 2004). Como essas séries apresentam fraca ou nula autocorrelação serial, a representação dos dados observados (sub-séries) em cada classe pelas respectivas CDF ou PEX não implica em perda de informação sobre padrões de autocorrelação das séries. A abordagem de anos análogos tem sido utilizada em muitos países, fornecendo informações valiosas para o planejamento de atividades agrícolas (STONE ET AL., 1996, MESSINA ET AL. 1999). Neste trabalho, propomos o uso de estatísticas descritivas e níveis de significância nominais (valores p) associados a testes estatísticos não paramétricos para quantificação do efeito do ENOS sobre as distribuições acumuladas das variáveis e interesses e, conseqüentemente, sobre riscos climáticos derivados de tais distribuições. Ilustramos a abordagem proposta em dois exemplos onde foi quantificada a influência do ENOS sobre riscos climáticos associados a chuvas trimestrais no Brasil, Uruguai e Austrália.

2. METODOLOGIA

Foram utilizados dados de chuva de estações meteorológicas do Brasil, Uruguai e Austrália para demonstrar a adequação das medidas descritivas e testes não paramétricos para quantificar a influência do ENOS sobre riscos climáticos. Os dados de cada sub-série são representados por funções empíricas que estimam a probabilidade condicional de a variável Y exceder um valor y , $PEX_Y(y)$, dado que y pertence à classe $K=j$. A função probabilidade de

exceder condicional á classe K é dada por $PEX_V(y)=1-P(Y\leq y|K=j)$. Para uma descrição detalhada dos métodos inferenciais, ver MAIA et al. (2007).

Exemplo I. Variabilidade espacial da influência do ENOS sobre chuvas trimestrais na Austrália. Neste exemplo foram utilizados dados de chuva dos trimestres JJA e JAS provenientes de 590 estações meteorológicas da Austrália. Em cada estação, as séries foram segregadas em cinco subséries classes de acordo com a classificação IOS-5 com “lag” zero, resultando em 2950 subséries (PEX condicionais) de comprimentos variáveis. Evidências de sinal do ENOS em cada estação foram quantificadas pelo teste Kruskal-Wallis (KW) para comparação de medianas das distribuições condicionais. Mapas das referidas evidências foram construídos para os dois trimestres analisados.

Exemplo II. Análise descritiva e inferencial da influência do ENOS sobre chuvas trimestrais no sudeste da América do Sul. Foram calculadas distâncias máximas entre as medianas do total de chuvas do trimestre OND, em de cada classe do ENOS, e a respectiva mediana histórica para as estações de São Luiz (Brazil) e Bella Union (Uruguai). Máximas distâncias verticais entre distribuições também foram mensuradas para as mesmas localidades. Como complemento às referidas análises descritivas, foram utilizados os testes Log-Rank e KW para quantificar a influência do ENOS sobre PEX condicionais e suas correspondentes medianas, respectivamente.

3. RESULTADOS E DISCUSSÃO

Representações das séries históricas dos totais trimestrais (JJA) de chuva (Y, mm) em Dalby e Yalgoo são apresentadas na Figura 1. Os gráficos à direita mostram as funções $PEX_V(y)$ em cada local, para cada classe do sistema IOS-5 (padrões do IOS “positivo”, “negativo”, “crescente”, “decrésciente” ou “neuro”). Os valores p associados ao teste KW foram 0,1229 para Yalgoo e 0,0049 para Dalby, sinal mais intenso do ENOS para este local. Na Figura 2, observa-se a utilidade da abordagem proposta para investigação de variabilidade espacial do sinal do ENOS numa região onde há disponibilidade de séries históricas para uma densa rede de estações meteorológicas. A abordagem também é adequada para comparação dos padrões de variabilidade em diferentes períodos de interesse, por exemplo, trimestres JJA e JAS. Neste exemplo, observa-se que o sinal do ENOS, quantificada em termos de divergências entre medianas das PEX (valor p do teste KW), é mais pronunciada no Leste do país. Nota-se também que o sinal do ENOS para o período JAS é mais intenso que para o período JJA.

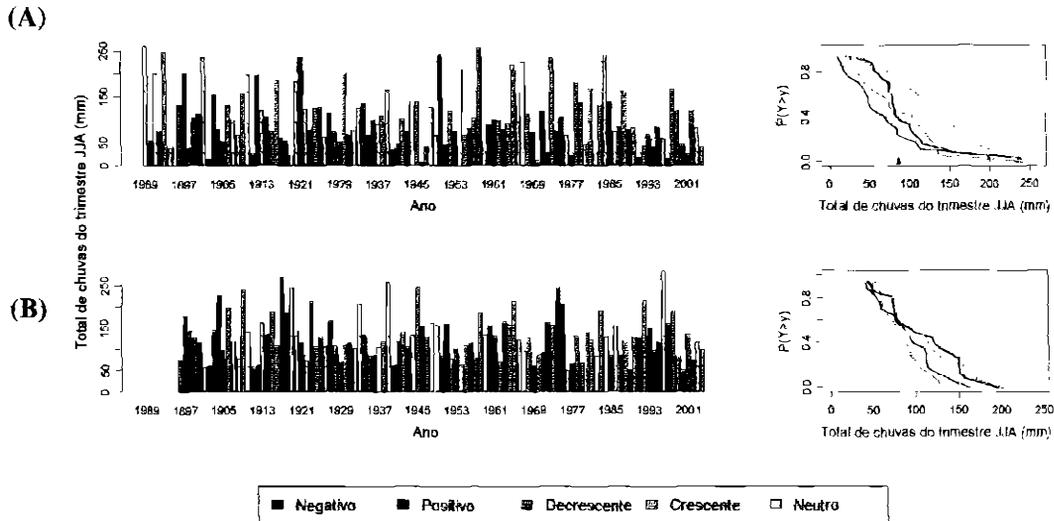


Figura 1. Séries históricas dos totais trimestrais (JJA) de chuva (Y) em Dalby (A) e Yaloo (B) e respectivas estimativas não-paramétricas das funções $PEX_Y(y)$.

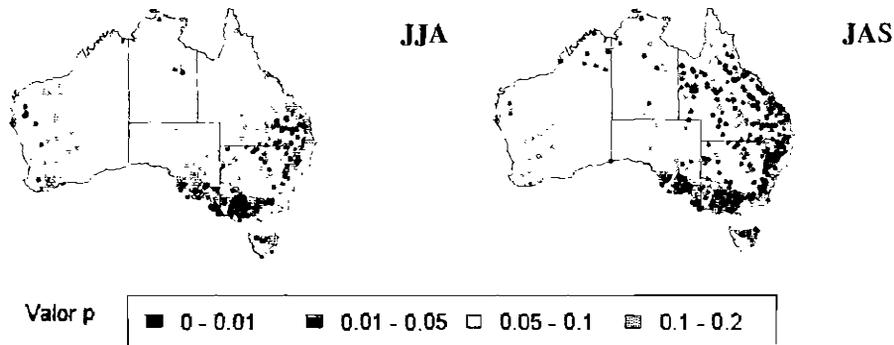


Figura 2. Variabilidade espacial do sinal do ENOS nos trimestres JJA e JAS na Australia. Evidências quantificadas pelo teste Kruskal-Wallis.

Na Figura 3, distâncias horizontais representam divergências entre medianas enquanto distâncias verticais representam a influência do ENOS sobre riscos, nesse caso, risco de o total de chuvas do trimestre exceder a mediana histórica. Os valores p dos testes KW para São Luiz e Bella Union foram 0,0084 e 0,0014, respectivamente. Para o teste Log-rank, os valores p foram inferiores a 0,0001 em ambos os locais.

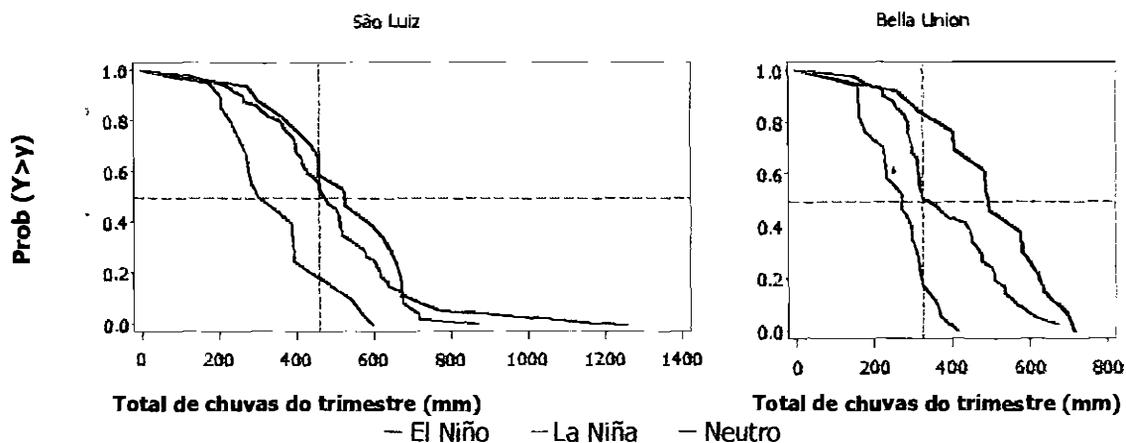


Figura 3. Estimativas não-paramétricas das funções $PEX_Y(y)$ para os totais de chuva do trimestre OND em cada classe do ENSO em São Luiz (Brasil) e Bella Union (Uruguai).

O uso de testes não paramétricos em estudos de variabilidade espacial de medidas de qualidade de SPCS tem a vantagem de não exigir avaliações de qualidade de ajuste para as distribuições propostas numa grande quantidade de locais.

4. CONCLUSÕES

A abordagem proposta para é adequada para investigação sobre variabilidade espaço temporal da influência do ENOS sobre riscos climáticos. Tais informações, associadas a estatísticas descritivas, são úteis para auxílio à tomada de decisão em sistemas de produção agrícolas.

A incorporação de indicadores do fenômeno El Niño/Oscilação Sul é útil para o aperfeiçoamento dos zoneamentos de risco climático de culturas anuais, principalmente em regiões onde a influência do ENSO é intensa e consistente.

REFERÊNCIAS BIBLIOGRÁFICAS

- MEINKE, H. AND STONE, R.C. Seasonal and inter-annual climate forecasting: the new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic Change*, 70, 221-253, 2005.
- MAIA, A.H.N., MEINKE, H., LENNOX, S., STONE, R.C. Inferential, non-parametric statistics to assess quality of probabilistic forecast systems. *Monthly Weather Review*, 135, 351-362, 2007.
- MESSINA C D, HANSSEN, J W, HALL, A. J. Land allocation conditioned on ENSO phases in the Pampas of Argentina, *Agricultural Systems*, 60:197-212, 1999.
- STONE, R. C., HAMMER G. L., MARCUSSEN, T. (1996). Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. *Nature*, 384, 252-55.