

MODELAGEM E PREVISÃO DAS CHUVAS ANUAIS PARA REGIÕES HOMOGÊNEAS DO ESTADO DO RIO GRANDE DO SUL

João Baptista da Silva¹, Ivan Saraiva², Jorge Andrioni³

^{1,3}Universidade Federal de Pelotas, Pelotas – RS, Brasil

²Universidade de São Paulo – USP, São Paulo – SP, Brasil

¹jbs39@terra.com.br

RESUMO

Neste trabalho, foram estudados os componentes das médias de chuvas anuais do estado do Rio Grande do Sul, a partir de 28 pontos de grades agrupadas em quatro regiões homogêneas definidas pela análise de agrupamentos. Os dados seguiram aproximadamente distribuição normal e apresentaram variâncias homogêneas, podendo ser analisados, diretamente sem necessidade de uma transformação prévia. O periodograma das médias anuais de cada grupo permitiram estabelecer modelos de três ondas senoidais. As ondas de períodos 5, 3,8 e 3,3 anos destacaram-se em 75% dos modelos. Os modelos representaram adequadamente o comportamento dos dados e de forma satisfatória permitiram a previsão de valores futuros.

Palavras chaves: análise espectral, periodograma, auto-correlações.

ABSTRACT

In this work, they were studied the components of the averages of annual rains of the state of the Rio Grande do Sul, starting from 28 grating points contained in four defined homogeneous areas by the analysis of groupings. The data followed normal distribution approximately and they presented homogeneous variations, could be analyzed directly without, without need of a previous transformation. The periodogram of the annual averages of each group allows

¹ Eng. Agr., Livre Docente, Doutor em Ciências, Professor Titular (Aposentado) do Instituto de Física e Matemática da UFPel, Bolsista do CNPq, jbs39@terra.com.br;

² Meteorologista, Aluno de Mestrado da Universidade de São Paulo – USP, Bolsista CAPES ivansaraiva@gmail.com;

³ Meteorologista - UFPel jorge.andrioni@gmail.com.

establishing models of 3 sinusoidal waves. The waves of periods 5, 3,8 and 3,3 years stood out in 75% of the models. The models represent the behavior of the data appropriately and in a satisfactory way they allowed the forecast of future values.

Key words: Spectral analysis, periodogram, autocorrelation.

1. INTRODUÇÃO

O Estado do Rio Grande do Sul caracteriza-se por uma intensa atividade agropecuária, com grande diversidade de culturas. O conhecimento do comportamento das variáveis climáticas é de suma importância para o planejamento das atividades agrícolas, seja para a escolha das culturas mais adequadas de uma determinada região ou para a determinação da melhor época de plantio de certa variedade. A precipitação destaca-se como um dos fatores mais importantes para o bom desenvolvimento das plantas.

De acordo com MORETTIN e TOLOI, [4], o estudo das séries temporais baseia-se no fato de que o passado é o melhor elemento para se estimar o futuro. Desta forma, os modelos são criados baseados em observações passadas na tentativa de se projetar para o futuro o comportamento de determinado fenômeno.

Destacam-se os seguintes objetivos para este trabalho:

- determinar as regiões homogêneas de chuvas anuais no Estado do Rio Grande do Sul, através da técnica de análise de agrupamento;
- identificar, pela técnica da análise espectral, os harmônicos mais significativos nas séries temporais formadas pelas médias das chuvas anuais das regiões homogêneas, de modo que se possa modelar o comportamento dessa variável;
- avaliar os modelos quanto à capacidade preditiva de novos valores.

2. MATERIAL E MÉTODOS

Foram utilizados no trabalho dados de chuvas anuais, provenientes do Climate Prediction Center /National Oceanic & Atmospheric Administration (CPC/NOAA), interpolados pelo programa GrADS, resultando em 28 pontos de grade, de 1° x 1° (latitude – longitude), cobrindo todo o Estado do Rio Grande do Sul. A figura 1 mostra a distribuição dos 28 pontos da grade sobre o Estado do RS. Os dados são referentes ao período de 1970 a 2004.

A análise da grade e o sistema de exposição (GrADS) são ferramentas interativas usadas para o acesso, a manipulação e a visualização de variáveis meteorológicas.

Na determinação das regiões homogêneas utilizou-se o método hierárquico aglomerativo de agrupamentos, o método de Ward, em virtude de suas conhecidas propriedades. O número de grupos foi determinado de forma subjetiva por um corte no dendograma, sempre respeitando a parcimônia [2].

A seguir, foram calculadas as médias das chuvas anuais correspondentes àqueles pontos de grade, pertencentes a cada uma das regiões homogêneas. Verificou-se após, a normalidade e a homogeneidade das variâncias das séries constituídas, pela aplicação dos testes de Shapiro-Wilk e de Cochran, respectivamente. Caso não se verificasse essa propriedade, buscar-se-ia uma adequada transformação dos dados.

Para cada uma das séries temporais formadas foi aplicada a análise espectral, baseada na Transformada Discreta de Fourier, de modo a identificar as possíveis periodicidades existentes, que originarão os modelos representativos do comportamento das chuvas anuais em cada região homogênea do Estado. Para a modelagem foram utilizados dados de 30 anos, de 1970 a 1999 e, os 5 anos restantes, de 2000 a 2004, para as previsões.

O modelo, baseado na representação periódica [5], apresenta a seguinte expressão:

$$\hat{X}_t = \frac{A_0}{2} + \sum_n [A_n \cos \omega_n t + B_n \text{sen} \omega_n t] + e_t \quad (1)$$

onde: \hat{X}_t : valor estimado da variável temporal.

ω_n : frequência angular da n-ésima onda.

A_n e B_n : parâmetros a serem estimados.

e_t : erros

$t = 0, 1, 2, \dots, (N-1)$; $n = 1, 2, 3, \dots, I(N/2)$; N = número de observações.

A seleção das ondas utilizadas na composição dos modelos foi baseada, conforme BAPTISTA DA SILVA et al. [1], na magnitude dos picos no periodograma e, aqui também, na observação dos valores estimados da função de densidade espectral. Foram incluídas, em cada modelo, as componentes senoidais ($\text{sen} \omega_n$ e $\text{cos} \omega_n$) que tiveram coeficientes significativos, estimados segundo a técnica da regressão linear múltipla. Cada modelo está

composto por um número de ondas adequado, de modo a garantir uma boa representatividade dos dados, sem que se torne excessivamente complexo.

O periodograma apresenta no eixo das abscissas a frequência e nas ordenadas o valor calculado pela soma dos quadrados dos coeficientes do seno e do cosseno, em cada frequência, multiplicada por $N/2$.

Estimado o modelo, a verificação do seu ajuste foi feita pela análise dos resíduos.

Se os modelos são verdadeiros, os resíduos devem ser normais e independentes, ou seja, constituir-se num ruído branco. Se os modelos são adequados, os resíduos estimados devem estar próximos dos reais e ser aproximadamente não correlacionados. Para se verificar isso, foram aplicados dois testes, descritos abaixo, baseados na análise das auto-correlações dos resíduos.

Define-se a função de auto-correlação de z com lag k como:

$$r_k = \frac{\frac{1}{n} \sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t-k} - \bar{z})}{\frac{1}{n} \sum_{t=1}^{n-k} (z_t - \bar{z})^2} \quad (2)$$

com $K = 0, 1, 2, \dots, n$.

Se os r_k indicarem as auto-correlações dos resíduos $\hat{\epsilon}_t$, então se deve ter $r_k \cong 0$. Tendo-se em vista que r_k tem aproximadamente distribuição normal, com média zero e variância $1/N$, pode-se considerar, de um modo geral, que o modelo é adequado quando r_k está dentro do intervalo $\pm 2/\sqrt{N}$, aceitando-se, no máximo, 5% dos seus valores fora deste intervalo, para $\alpha = 0,05$.

Determina-se, também, a existência de ruído branco a partir da comparação de r_k com $\pm 2\sigma(r_k)$, onde $\sigma(r_k)$, desvio padrão de r_k , é definido como:

$$\sigma(r_k) \cong \left[\frac{1}{N} \left(1 + 2 \sum_{v=1}^q \rho_v^2 \right) \right]^{1/2}, \quad k > q \quad (3)$$

onde q é o maior “lag”, além do qual as funções de auto-correlação tendem a zero [3].

Considerando-se que a série seja um processo “ruído branco”, isto é, que todas as auto-correlações são iguais a zero, o erro padrão de r_k fica simplificado para:

$$\sigma(r_k) = \sqrt{\frac{1}{N} \left[\frac{N-k}{N+2} \right]} \quad (4)$$

Novamente, se aceita como ruído branco quando, no máximo, 5% das auto-correlações ficam fora do intervalo $\pm 2\sigma(r_k)$, para $\alpha = 0,05$.

O teste t pode ser utilizado para avaliação do ajuste dos modelos [6], entretanto, neste trabalho foi usado apenas para verificar a eficiência dos modelos quanto à previsão de novos valores. Embora os indicadores estatísticos MBE (*erro médio tendencioso*) e RMSE (*raiz quadrada do erro médio*) forneçam um razoável procedimento para comparar modelos, eles não indicam objetivamente se as estimativas do modelo são estatisticamente significativas. Uma medida estatística baseada no teste t permite que os modelos sejam comparados e, ao mesmo tempo, indica se suas estimativas são estatisticamente significativas.

O valor de t, a ser calculado, é:

$$t = \sqrt{\frac{(N-1)MBE^2}{RMSE^2 - MBE^2}} \quad (5)$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (X_o - X_e) \quad , \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_o - X_e)^2} \quad (6)$$

Onde:

X_o = valor observado

X_e = valor estimado pelo modelo

Este valor deve ser comparado com o valor de t da tabela bilateral, para o nível de significância escolhido, com (N-1) graus de liberdade. Quanto menor o valor de t, comparado com o valor limite da tabela, melhor o desempenho do modelo.

3. RESULTADOS E DISCUSSÃO

A análise do dendograma, resultante do método de Ward, indicou a constituição de quatro grupos (pontos da grade): - Grupo I: 1, 2, 6, 7 e 8; - Grupo II: 3, 4, 5, 9, 10 e 11; - Grupo III: 12, 13, 14, 15, 16, 17, 18, 21, 22, 23 e 24; - Grupo IV: 19, 20, 25, 26, 27 e 28 (Figura 1). A delimitação das áreas no mapa foi feita apenas para permitir a visualização dos grupos, sem nenhum critério objetivo, não representando isolinhas.

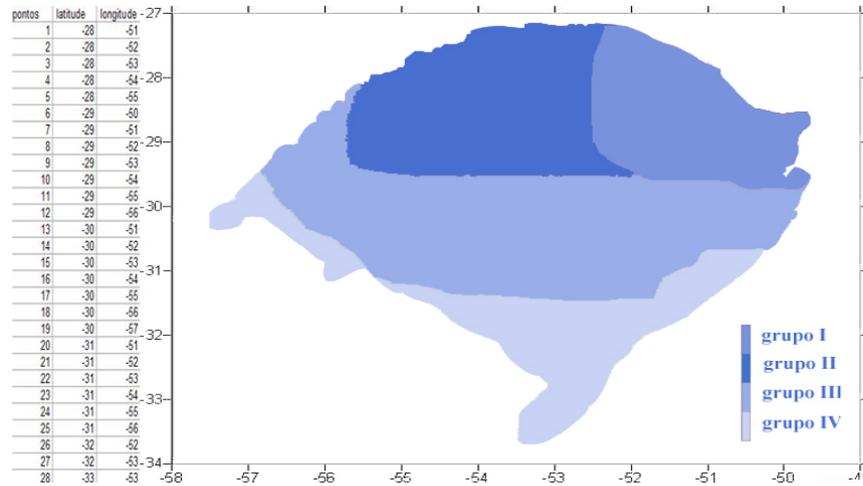


Figura 1 – Regiões homogêneas em chuvas anuais do estado do Rio Grande do Sul no período.

Pelo teste de Shapiro-Wilk obtiveram-se os seguintes resultados: - Grupo I: $W=0,9564$, $p<0,2286$; - Grupo II: $W=0,9672$, $p<0,4485$; - Grupo III: $W=0,9591$, $p<0,2719$; - Grupo IV: $W=0,9467$, $p<0,1160$. Como se pode observar todos os resultados foram não significativos, ao nível de probabilidade de 5%, indicando que os dados de cada grupo seguem aproximadamente a distribuição normal.

O teste de Cochran, para a homogeneidade de variâncias resultou para os 4 grupos, em valores de $C_I=0,2154$, $C_{II}=0,2365$, $C_{III}=0,3175$ e $C_{IV}=0,3521$, respectivamente. Em vista disto, podemos considerar que os grupos apresentam variâncias homogêneas a 5% ($C_{0,05;7;4}=0,4307$).

Os periodogramas estimados indicaram, pelos picos nos gráficos, as ondas senoidais mais destacadas, selecionadas por suas magnitudes para compor os modelos com 3 ondas, apresentados a seguir:

$$\begin{aligned} \hat{X}_{1,t} &= 1\,641,647 - 141,273 * \cos(1,256 * t) - 8,4236 * \text{sen}(1,256 * t) \\ &\quad + 133,655 * \cos(1,884 * t) - 39,3334 * \text{sen}(1,884 * t) \\ &\quad - 123,806 * \cos(1,675 * t) + 24,3008 * \text{sen}(1,675 * t) \\ \hat{X}_{2,t} &= 1\,720,1 - 176,248 * \cos(1,675 * t) + 93,4435 * \text{sen}(1,675 * t) \\ &\quad + 162,824 * \cos(1,884 * t) - 89,4319 * \text{sen}(1,884 * t) \\ &\quad + 166,541 * \cos(0,419 * t) - 25,7055 * \text{sen}(0,419 * t) \\ \hat{X}_{3,t} &= 1\,447,845 - 136,396 * \cos(1,256 * t) + 48,7638 * \text{sen}(1,256 * t) \\ &\quad + 126,748 * \cos(1,675 * t) + 36,2932 * \text{sen}(1,675 * t) \\ &\quad + 86,275 * \cos(2,512 * t) - 96,4791 * \text{sen}(2,512 * t) \\ \hat{X}_{4,t} &= 1\,163,883 - 88,3293 * \cos(1,256 * t) + 46,6312 * \text{sen}(1,256 * t) \\ &\quad + 84,0878 * \cos(1,884 * t) + 32,7878 * \text{sen}(1,884 * t) \\ &\quad + 53,1381 * \cos(0,419 * t) - 67,6776 * \text{sen}(0,419 * t) \end{aligned}$$

onde $t = 0, 1, 2, 3, \dots, 29$.

Processou-se uma análise dos resíduos, por meio de suas funções de auto-correlação, para cada um dos modelos. Os resultados dos dois testes usados ($\alpha = 0,05$), foram os seguintes:

- a) **Teste 1:** no máximo 5% das auto-correlações fora do intervalo de $\pm 2/\sqrt{N} = \pm 2/\sqrt{30} = \pm 0,3651$. Em nenhum dos quatro grupos ocorreram valores fora do intervalo.
- b) **Teste 2:** no máximo 5% das auto-correlações fora do intervalo de $\pm 2\sigma(r_k)$. Em nenhum dos quatro grupos ocorreram valores fora do intervalo.

Os resultados dos testes das auto-correlações evidenciaram que os resíduos dos quatro modelos podem ser aceitos como ruído branco, adotando-se os modelos, para representar as chuvas médias anuais das regiões homogêneas (Figura 2).

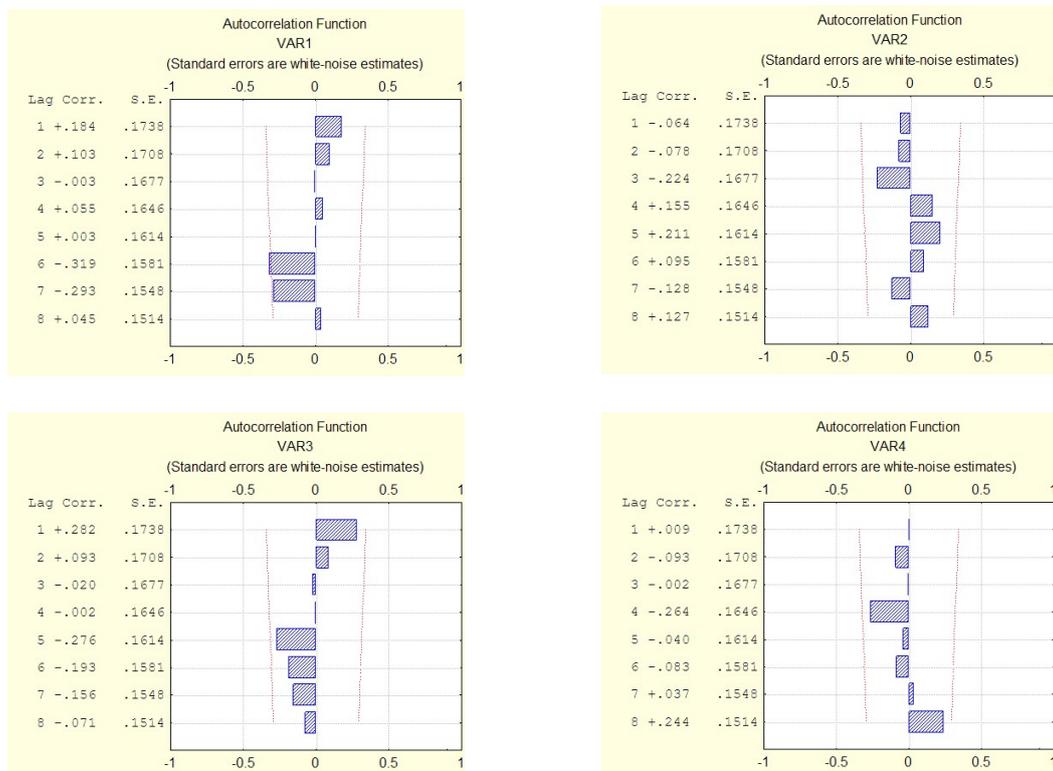


Figura 2: Função de autocorrelação dos resíduos do ajuste dos modelos espectrais das chuvas de 1970 a 1999, para: a) grupo 1; b) grupo 2; c) grupo 3; d) grupo 4.

Quanto à previsão de novos valores, os modelos foram avaliados pelo teste t, para os anos de 2000 a 2004, cujos dados de chuva não participaram no estabelecimento dos mesmos, indicando $t_I = 0,54$, $t_{II} = 0,28$, $t_{III} = 2,08$ e $t_{IV} = 2,24$, todos não significativos a 5% ($t_{0,05;4} = 2,78$). Isto nos permite aceitar os modelos como bons previsores de valores futuros (Figura 3).

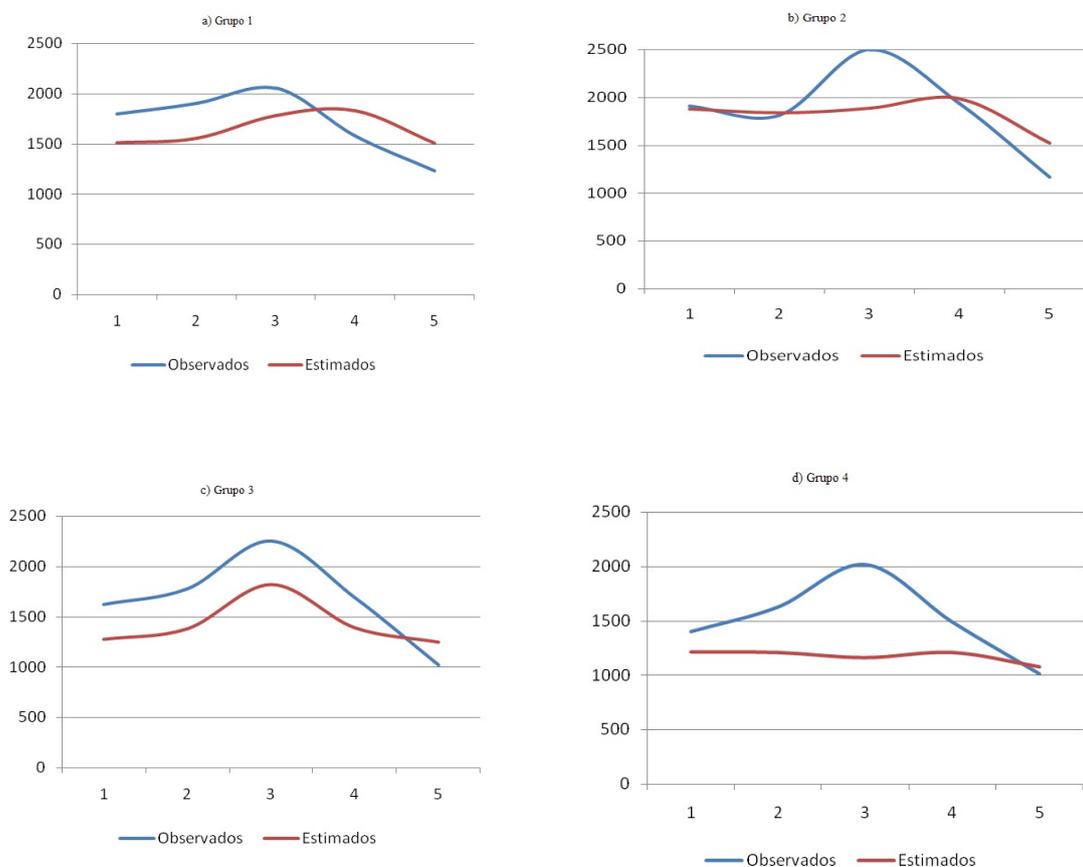


Figura 3 – Precipitações anuais observadas e previstas para o período de 2000 a 2004, para: a) grupo 1; b) grupo 2; c) grupo 3; d) grupo 4.

4. CONCLUSÕES

1) Os 28 pontos identificados sobre o Estado do Rio Grande do Sul são classificados, pela análise de agrupamentos, em quatro regiões homogêneas quanto ao regime de chuvas.

2) Os modelos propostos se constituem de três ondas senoidais, destacando-se em 75% deles, as ondas de 5 anos, 3,8 anos e 3,3 anos.

3) Os modelos representam adequadamente o comportamento das chuvas anuais nas quatro regiões homogêneas selecionadas no Estado e de forma satisfatória fazem a previsão de novos valores, na maioria destas regiões.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BAPTISTA DA SILVA, J. ; CUNHA, F.B. ; NETO, W.P. Modelagem das chuvas trimestrais por regiões homogêneas no Estado do Ceará. **Revista Brasileira de Agrometeorologia**, Santa Maria, v.9, n.2, p.317-324, 2001.
- [2] BAPTISTA DA SILVA, J. ; CUNHA, F.B. ; PAZ, S.R. ; BASGALUPP, M.P. Análise do comportamento das chuvas em algumas localidades do semi-árido do Nordeste Brasileiro. **Revista Brasileira de Agrometeorologia**, v.13, n.1, p.173-181, 2005.
- [3] BOX, G.E.P. ; JENKINS, G.M. ; REINSEL, G.C. **Time series analysis. Forecasting and control**. New Jersey: Prentice-Hall, Inc., 3nded., 1994, 598p.
- [4] MORETTIN, P.A. ; TOLOI, C.M.C. **Modelos para previsão de séries temporais**. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, Vols. 1 e 2, 1981, 623p.
- [5] PRIESTLEY, M.B. **Spectral analysis and time series**. Elsevier Academic Press, San Diego, Califórnia, USA, Vols. 1 e 2, 2004, 890p.
- [6] TOGRUL, I.T. ; TOGRUL, H. Global solar radiation over Turkey: comparison of predicted and measured data. **Renewable Energy**, Turkey, v.25, p.55-67, 2002.