ANÁLISE DAS CHUVAS MENSAIS POR MODELOS DE BOX & JENKINS: UM ESTUDO PRELIMINAR

João Baptista da Silva[1], Ivan Saraiva[2], Márcio Porto Basgalupp[3]

ABSTRACT – The study of the time series has as basis that the past it is the best element to estimate the future. So, the past and present values of the series will be project to the future. The time series data presents a serial dependence among them so, the study of the series implies on analyzing and modeling this kind of dependence. There are two distinct methods to analyze time series: the frequency domain and the time domain. The Box & Jenkins models belong to the last one and are based upon linear filters operations. ARIMA models are adjusted to the observed data. This paper reports the results to a study to adjust Box & Jenkins models to the monthly rainfall totals in Pelotas, RS, during the 1900/1999 time interval. The multiplicative seasonal ARIMA model was adjusted: ARIMA $(0,1,1)x(0,1,1)_{12}$, as suggested by the autocorrelation functions analysis. The estimated values for the unconditional least squares (UCL) method lead to the model: $V_t = (1-0{,}97204\ B)\ (1-0{,}99999\ B^{12})\ a_t$, where $V_t = (\bar{Z}_t - \bar{Z}_{t-1}) - (\bar{Z}_{t-12} - \bar{Z}_{t-13})$. The residual analysis denote that this model is adequate and by the means of the t test the statistical non-significance of the forecast for the 2000' year can be proved.

INTRODUÇÃO

De acordo com Morettin e Toloi (1987), o estudo de séries temporais tem por filosofia a idéia básica de que o passado é o melhor elemento para se estimar o futuro; assim, os valores passados e presentes da série serão projetados para o futuro. Uma série temporal é um conjunto de observações ordenadas em intervalos eqüidistantes no tempo; entretanto, tempo pode ser substituído por qualquer variável como espaço, profundidade, etc. As observações apresentam uma dependência serial entre elas e o estudo de uma série temporal consiste em analisar e modelar esta dependência.

Para análise de séries temporais existem dois métodos distintos, mas não necessariamente exclusivos: no domínio da freqüência ou no domínio do tempo.

No domínio da freqüência supõe-se que a série temporal é melhor considerada como uma soma ou superposição linear de ondas periódicas de senos e cossenos de diferentes períodos ou freqüências.

Por outro lado, pela a abordagem no domínio do tempo afirma-se que as observações presentes podem ser preditas como a soma de uma combinação linear de valores passados de uma série de ruídos e um componente determinístico ortogonal àquela combinação linear (Wold, 1938).

Os trabalhos de Box & Jenkins, alicerçados nos resultados de Wold (1938), propuseram uma classe geral de modelos lineares conhecidos como modelos de Box & Jenkins, baseados em operações de filtros lineares. Tal método consiste em ajustar modelos auto-regressivos integrados de médias móveis, ARIMA, a um conjunto de dados.

O objetivo do presente trabalho é modelar os dados das chuvas mensais de Pelotas, RS, do intervalo temporal de 1900 a 1999, no domínio do tempo, usando o modelo de Box & Jenkins.

MATERIAL E MÉTODOS

Os dados utilizados foram os totais mensais das precipitações pluviais de Pelotas, RS, no período de 1900 a 1999, coletados pela Estação Agroclimatológica de Pelotas, situada no Campus da Universidade Federal de Pelotas (UFPel), a 15 quilômetros do centro da cidade de Pelotas (latitude: 31°52' S, longitude: 52°21' W, altitude: 13,2m).

A metodologia consistiu da modelagem dos dados, pelos modelos auto-regressivos integrados médias móveis, ARIMA, de Box & Jenkins (Box at al.,1994).

Considere um modelo auto-regressivo médias móveis, ARMA (p,q), da forma: $\bar{\mathbf{Z}}_t = \phi_1 \bar{\mathbf{Z}}_{t-1} + \ldots + \phi_p \bar{\mathbf{Z}}_{t-p} + \mathbf{a}_t - \theta_1 \mathbf{a}_{t-1} - \ldots - \theta_q \mathbf{a}_{t-p}$, onde $\bar{Z}_t = Z_t - \mu$, no qual os valores correntes da série temporal dependem dos valores passados da série e dos erros correntes e passados. Pode-se representar o modelo ARMA na forma compacta: $\phi(\mathbf{B})\bar{\mathbf{Z}}_t = \theta(\mathbf{B})\mathbf{a}_t$, $\bar{\mathbf{Z}}_t = \phi^{-1}(\mathbf{B})\theta(\mathbf{B})\mathbf{a}_t$, sendo $B\bar{Z}_t = \bar{Z}_{t-1}$. Os modelos ARMA são apropriados para descrever séries estacionárias, que se desenvolvem no tempo aleatoriamente ao redor de uma média constante. Quando as séries não são estacionárias, pode-se obter a estacionariedade da série pela aplicação do operador diferença Δ^d entre valores consecutivos, isto é, calculando-se $W_t = \Delta^d \bar{Z}_t$. Ajustando-se um modelo ARMA a nova série estacionária W_t, tem-se: $\phi(\mathbf{B})\mathbf{W}_t = \theta(\mathbf{B})\mathbf{a}_t$.

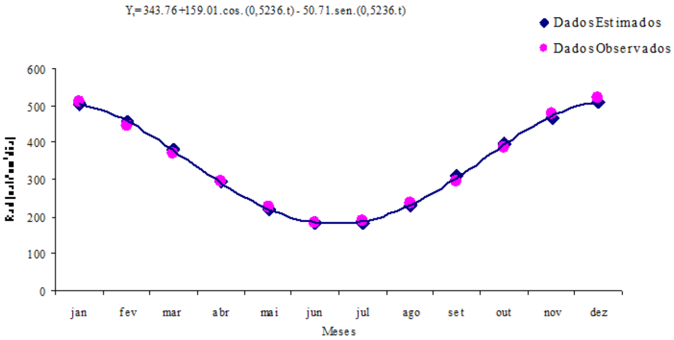
Uma vez que $W_t = \Delta^d \bar{Z}_t = (1-B)^d \bar{Z}_t$, obtém-se $\phi(B)\Delta^d \bar{Z}_t = \theta(B) a_t$, que é denominado modelo auto-regressivo integrado médias móveis ou ARIMA(p,d,q).

A construção dos modelos é baseada num ciclo iterativo, no qual a estrutura do modelo emana dos próprios dados. O ciclo iterativo envolve as fases de identificação (identify), estimação (estimate), verificação e obtenção das previsões (forecast) (Brocklebank & Dickey, 2003).

A fase de identificação tem por objetivo identificar e determinar os valores de p, d e q do modelo ARIMA (p,d,q) e estimar preliminarmente os parâmetros, a serem utilizados na fase de estimação.
Na estimação propriamente dita dos parâmetros, será necessário utilizar um procedimento iterativo de estimação de mínimos quadrados não-linear e as estimativas preliminares encontradas na fase de identificação serão usadas como valores iniciais.

A fase de verificação pode ser feita pela análise dos resíduos. Os erros são, em geral, assumidos normais e independentes, ou seja, constituem um ruído branco. Assim, se o modelo for adequado, os resíduos deverão possuir estas propriedades.

[1] Engenheiro Agrônomo, Livre Docente, Doutor em Ciências, Bolsista do CNPq, Professor Titular (Aposentado) do Instituto de Física e Matemática (UFPel). jbsilva@ufpel.edu.br
[2] Estudante de Meteorologia da UFPel. Bolsista de Iniciação Científica do CNPQ. ivansaraiva@hotmail.com
[3] Bacharel em Ciência da Computação pela UFPel. Ex-bolsista de Iniciação Científica do CNPq basgalupp@inf.pucrs.br



seja, tem-se que tomar a diferença simples e D diferenças sazonais da série $\overline{Z}_t$, afim de produzir estacionaridade, obtém–se então: $W_t = \Delta^d \Delta_s^D \overline{Z}_t = (1-B)^d (1-B^s)^D \overline{Z}_t$, denominado modelo SARIMA (p,d,q)x(P,D,Q).

O ajuste do modelo passa pelos processos de identificação, estimação e previsão.

Quanto à previsão os valores foram testados estatisticamente pelo teste **t**, onde $\mathbf{t} = \left\{ \left[(\mathbf{n}-1)\mathbf{MBE}^2 \right] / \left(\mathbf{RMSE}^2 - \mathbf{MBE}^2 \right) \right\}^{1/2}$, sendo $\mathbf{MBE} = \frac{1}{\mathbf{n}} \Sigma (\mathbf{x_o} - \mathbf{x_e})$ e $\mathbf{RMSE} = \left[\frac{1}{\mathbf{n}} \Sigma (\mathbf{x_o} - \mathbf{x_e})^2 \right]^{1/2}$, entre os valores observados X_o e os estimados X_e pelo modelo. Este valor de t é comparado com o valor da tabela de t, para o nível de significância α, com **(n-1)** graus de liberdade. A hipótese de nulidade expressa a não diferença estatística entre os valores de X_o e X_e (Togrul & Togrul, 2002).

RESULTADOS E DISCUSSÃO

Inicialmente, foi calculada a média dos dados, μ = 108,2 e os desvios da média, $\overline{Z}_t = Z_t - \mu = Z_t - 108,2$, que foram submetidos à modelagem. Pode-se constatar pela queda suave das funções de autocorrelação, autocorrelação inversa e autocorrelação parcial, que a série não é estacionária. Para adquirir esta propriedade utiliza-se o processo de diferenciação, calculando-se uma primeira diferença simples na fase de identificação do modelo.

Quando a série é não estacionária e com padrão sazonal, como costuma ocorrer em registros mensais de variáveis meteorológicas no conjunto dos anos, pode-se usar a identificação: Identify var = X (1,12), que indica uma diferença de segunda ordem de X, tal que a série analisada é a diferença entre as atuais mudanças, período a período, em X e a troca de 12 períodos passados. Ou seja, $V_t = (\overline{Z}_t - \overline{Z}_{t-1}) - (\overline{Z}_{t-12} - \overline{Z}_{t-13})$.

O modelo é identificado desde as autocorrelações. A identificação depende do reconhecimento do padrão no gráfico da função de autocorrelação. As autocorrelações teóricas das séries $\mathbf{V_t} = (1-\theta_1\mathbf{B})(1-\theta_2\mathbf{B}^{12})\mathbf{e_t}$, onde $\theta_1 > 0$ e $\theta_2 > 0$, deveriam ter: a) um pico (negativo) no lag 1; b) um pico (negativo) no lag 12; c) iguais (e positivos) picos nos lags 11 e 13 chamados de lóbulos laterais do pico do lag 12; d) correlações zeros para todos os outros lags.
Na Figura 1, este padrão foi comparado com a função de autocorrelação da variável X(1,12) e foi encontrado um acordo razoável. O pico e os lóbulos laterais no lag sazonal são características dos modelos multiplicativos sazonais (Broocklebank & Dickey, 2003).

Na estimação dos parâmetros usou-se os métodos de estimação ULS (unconditional least squares), CLS (conditional least squares) e ML(maximum likelihood). As variâncias estimadas pelos três métodos foram 4829,49; 5284,46 e 4827,29, respectivamente. Optou-se pelo método ULS, com as estimativas a seguir:

Tabela 1. Estimativas (significativas) dos parâmetros do modelo ARIMA (p,d,q)x(P,D,Q)12.

Parâmetro	Estimativa	Erro Padrão	t	lag
MA1,1	0,97204	0,0080378	120,93	1
MA2,1	0,99999	0,05815	17,20	12

O modelo sazonal multiplicativo ARIMA(0,1,1)x(0,1,1)12, ou seja, $V_t = (1-\theta_1 B)(1-\theta_2 B^{12})\,a_t$ foi estimado como $V_t = (1-0{,}97204B)(1-0{,}99999B^{12})\,a_t$. As análises dos resíduos, sob a hipótese de ruído branco, foram também realizadas pelos testes Kappa de Fisher e Kolmogorov-Smirnov de Bartlett. O valor do teste de Kappa foi de 5,2699, não significativo a 5% ($K_{0,05;593}$=9,313) e o KS foi de 0,0303 também não significativo a 5% ($KS_{0,05;592}$=0,0558). Por estes testes o modelo estabelecido é considerado adequado na apresentação dos dados. O teste de Qui-Quadrado das autocorrelações dos resíduos, nos lags 6,12,18...48, mostrou-se não significativo, o que foi confirmado pelos resultados dos dois testes anteriores.

As previsões foram obtidas para doze passos à frente, com intervalo de confiança de 95%. Aos valores previstos deve ser acrescida a média de 108,2. O valor de t = 0,174 foi não significativo a 5% ($t_{0,05;11}$=2,20), o que indica que o modelo é um previsor adequado das chuvas mensais de Pelotas.

Lag	Covariance	Correlation	Std
0	18407.479	1.00000	0
1	-8356.332	-0.45396	0.029025
2	-780.839	-0.04242	0.034492
3	-1099.632	-0.05974	0.034536
4	1548.282	0.08411	0.034623
5	-627.715	-0.03410	0.034795
6	452.674	0.02459	0.034823
7	227.123	0.01234	0.034837
8	-1582.005	-0.08594	0.034841
9	1084.761	0.05893	0.035019
10	351.455	0.01909	0.035103
11	4598.998	0.24984	0.035111
12	-9210.635	-0.50037	0.036578
13	3276.768	0.17801	0.041950
14	531.316	0.02886	0.042582
15	1393.691	0.07571	0.042599
16	-1230.411	-0.06684	0.042712
17	291.523	0.01584	0.042800
18	-524.019	-0.02847	0.042805
19	-196.167	-0.01066	0.042821
20	1617.790	0.08789	0.042823
21	-894.836	-0.04861	0.042975
22	-256.788	-0.01395	0.043021
23	-315.238	-0.01713	0.043025
24	323.284	0.01756	0.043030
25	466.082	0.02532	0.043037
26	352.101	0.01913	0.043049
27	287.540	-0.06995	0.043056
28	430.889	0.02341	0.043152
29	541.665	0.02943	0.043163
30	-551.206	-0.02994	0.043179
31	1226.317	0.06662	0.043197
32	-1379.073	-0.07492	0.043283
33	-48.584405	-0.00264	0.043393
34	586.396	0.03186	0.043393
35	-106.290	-0.00577	0.043412
36	-15.726261	-0.00085	0.043413

Figura 1. Autocorrelações dos desvios da média das chuvas mensais em Pelotas, RS (1900 – 1999).

As análises permitiram o ajuste de um modelo ARIMA sazonal aos dados observados. O modelo adequado foi o ARIMA (0,1,1)x(0,1,1)12, sendo estimado como $V_t = (1-0{,}97204\;B)(1-0{,}99999\;B^{12})\,a_t$. O modelo foi considerado estatisticamente como um bom previsor dos valores futuros da série de precipitações mensais de Pelotas, RS.

REFERÊNCIAS

Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C. Time series analysis. Forecasting and control. 3rd ed. , Prentice Hall, Inc., New Jersey, 1994, 598p..

Brocklebank, J.C. ; Dickey, D.A. SAS for forecasting time series. 2nd ed. , John Wiley & Sons Inc. , 2003. 398p.

Morettin, P.A.; Toloi C.M.C. Previsão de séries temporais. 2 ed. , São Paulo: Atual, 1987. 439p.

Togrul, I. T. ; Togrul, H. Global solar radiation over Turkey: comparison of predicted and measured data. Renewable Energy, 25, 2002. 55-67p.

Wold, H. A large sample test for MA Processes. Journal of The Royal Statistics Societ., B, 11,237-305.1938.