



## CIÊNCIAS EXATAS E DA TERRA

## Comparação do desempenho dos modelos estatísticos aplicados a séries temporais de umidade relativa do ar

### *Comparison of the performance of statistical models applied to temporal series of relative humidity*

Cleiton Tibulo<sup>1</sup>; Roselaine Ruviano Zanini<sup>1</sup>; Vaneza De Carli Tibulo<sup>1</sup>

### RESUMO

Modelos estatísticos aplicados a séries temporais vêm sendo empregados em diversas áreas e têm surgido como necessidade atual para empresas sobreviverem em um mercado globalizado e competitivo. Nesse contexto, o presente trabalho tem por objetivo apresentar uma comparação do desempenho dos modelos estatísticos: Autorregressivo de Médias Móveis Sazonal (SARMA), Autorregressivo de Médias Móveis e entrada de variáveis exógenas (ARMAX) e modelos de Alisamento Exponencial Holt-Winters aditivo/multiplicativo, ajustados a dados de umidade relativa do ar (UR). Os dados foram coletados junto ao Instituto Nacional de Meteorologia (INMET) no período de 01 de outubro de 2001 a 22 de janeiro de 2014. Na comparação dos resultados e na seleção do melhor modelo foram utilizados os critérios Erro Absoluto Percentual Médio (MAPE), Erro Quadrático Médio (EQM), Média absoluta dos erros (MAD) e Soma dos quadrados das diferenças (SSE). No período analisado, observou-se que a média da UR foi de 76,03% com coeficiente de variação de 7,68% e a série em estudo apresentou sazonalidade, justificando o uso dos modelos propostos. Os resultados mostraram que o modelo ARMAX (3,0) com a inclusão de variáveis exógenas ajustou-se de forma adequada à UR, apresentando melhores resultados de previsão em relação aos seus concorrentes SARMA (3,0)(1,1)12 e o Holt-Winters multiplicativo/aditivo.

**Palavras-chave:** Umidade Relativa do Ar; Modelos SARIMA; Modelos de Alisamento Exponencial; Modelos ARMAX.

### ABSTRACT

*Statistical models applied to time series have been used in several areas and have emerged as a current need for companies to survive in a globalized and competitive market. In this context, the present work aims to present a comparison of the performance of the statistical models: Seasonal Moving Average Self-Correlation (SARMA), Self-Regressive Moving Average and Entry of Exogenous Variables (ARMAX), and additive / multiplicative Holt-Winters exponential smoothing models, adjusted to relative air humidity (RH) data.. Data were collected from the National Institute of Meteorology (INMET) from October 1, 2001 to January 22, 2014. In the comparison of the results and in the selection of the best model, the criteria Absolute Average Percent Error (MAPE) were used, Mean Square Error (EQM), Absolute Mean of Errors (MAD) and Sum of Squares of Differences (SSE). In the analyzed period, it was observed that the mean RH was 76.03% with a coefficient of variation of 7.68% and the series under study presented seasonality, justifying the use of the proposed models. The results showed that the ARMAX (3.0) model with the inclusion of exogenous variables was adjusted*

<sup>1</sup> UFSM – Universidade Federal de Santa Maria, Santa Maria/RS – Brasil.

*appropriately to the UR, presenting better prediction results than its competitors SARMA (3,0) (1,1) 12 and or multiplicative / additive Holt-Winters.*

**Keywords:** *Relative Humidity; SARIMA models; Exponential Smoothing Model; ARMAX models.*

## 1. INTRODUÇÃO

O aprimoramento e o desenvolvimento de modelos estatísticos aplicados a variáveis climáticas vêm ganhado espaço na sociedade atual e tem surgido como necessidade para o aprimoramento de pesquisas. Desta forma os modelos de previsão de séries temporais vêm se destacando e podem auxiliar consideravelmente nessa área do conhecimento.

Conhecer previamente um valor futuro pode representar grande diferença na tomada de decisões em inúmeras áreas. Nesse contexto, a análise de séries temporais aplicadas a dados climatológicos atrai interesse especial, considerando que o clima interfere diretamente na qualidade de vida dos habitantes de qualquer região ou ainda, que na área econômica determina o sucesso ou o fracasso de vários empreendimentos.

Para Chechi e Sanches (2013), a partir da análise de séries temporais climatológicas é possível construir-se modelos de previsão que necessitam basicamente de uma série de valores observados ao longo do tempo.

Modelos estatísticos de previsão vêm sendo utilizados de maneira satisfatória em variáveis climatológicas para diferentes regiões. Chechi e Sanches (2013), ajustaram modelos de séries temporais para a previsão da precipitação no município de Erechim-RS. Antonini et al. (2009) desenvolveram um modelo matemático para estimativa da temperatura média diária do ar para o estado de Goiás.

Para o nordeste brasileiro, Cavalcanti et al. (2006) estabeleceram um modelo de estimativa da temperatura do ar em função das coordenadas geográficas e das anomalias de temperaturas da superfície do mar. Para a cidade de Uberlândia, Minas Gérias, Silva et al. (2008) analisaram a temperatura média mensal, através do modelo de série temporal do tipo SARIMA.

As variações de temperatura da cidade de Nossa Senhora da Glória estado de Sergipe foram investigadas por Andrade et. al. (2011), por meio do modelo de séries temporais ARIMA.

Soebiyanto et al. (2010), aplicaram o modelo de série temporal ARIMA, nas regiões de Hong Kong (China) e no Condado de Maricopa, Arizona (Estados Unidos da América) para verificar efeitos diretos das principais variáveis climáticas (umidade do ar, precipitação, temperatura) no inverno como fator de transmissão da gripe em climas temperados.

Com o objetivo de prever focos de calor no Estado do Mato Grosso e diminuir as queimadas, Espinosa et al. (2010) utilizaram modelos de séries temporais ARIMA (SARIMA). Modelos de previsão de prognósticos em escala sazonal da variável precipitação são discutidos por Lúcio et al. (2010).

Apesar de um número crescente de pesquisas abordando aplicações de modelos estatísticos em inúmeras variáveis, dentre elas as climáticas, existem poucos resultados de trabalhos com o emprego de modelos de séries temporais para dados coletados em porcentagem (%).

Desta forma este trabalho tem por objetivo apresentar um estudo comparativo do ajuste e desempenho dos modelos de séries temporais ARIMA, ARMAX e Alisamento Exponencial de Holt-Winters, ajustados a dados de umidade relativa do ar.

## 2. REFERENCIAL TEÓRICO

### 2.1 Modelo SARIMA

Em estudos de variáveis ambientais é comum encontrar sazonalidade nas séries segundo Reisen et al. (2008), o que remete à necessidade de utilizar modelos que incorporem essa importante característica temporal. Nesse contexto, utilizou-se os modelos SARIMA (p,d,q)(P,D,Q) representados por:

$$\nabla^D \Delta^d Z_t = \underbrace{\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p}}_{\text{autorregressivo (p)}} + \varepsilon_t - \underbrace{\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \dots - \theta_q \varepsilon_{t-q}}_{\text{médias móveis (q)}} + \underbrace{\Phi_1 Z_{t-1s} + \Phi_2 Z_{t-2s} + \dots + \Phi_P Z_{t-P}}_{\text{autorregressivo sazonal (P)}} - \underbrace{\Theta_1 \varepsilon_{t-1s} - \Theta_2 \varepsilon_{t-2s} + \dots + \Theta_Q \varepsilon_{t-Q}}_{\text{médias móveis sazonal (Q)}}$$

em que (p,d,q) são as ordens do modelo referentes a parte ordinal ou simples, enquanto que (P,D,Q) são referentes às ordens do modelo da parte sazonal. Os parâmetros  $\phi_1, \dots, \phi_p$  são os parâmetros referentes à parte autorregressiva;  $\theta_1, \dots, \theta_q$  são os parâmetros de médias móveis referentes à parte sazonal;  $\varepsilon_t$  é o erro que não pode ser estimado do modelo; d representa número de diferenças da parte ordinal e D representa o número de diferenças da parte sazonal.

A utilização dos modelos SARIMA em estudos ambientais tem apresentado resultados consistentes na modelagem e previsão de séries temporais, mostrando assim ser uma ferramenta útil e que deve ser considerada quando se pretende realizar um ajuste de modelo, segundo Reisen et al. (2008).

### 2.2 Modelo ARMAX

O modelo ARMAX é um modelo múltiplo que possibilita a análise de mais de uma variável correlacionada à previsão de  $Z_t$ , sendo a variável endógena dependente e as variáveis explicativas exógenas independentes, de acordo com Silva (2011).

Segundo Martínez e Zamprogno (2003) e Ozaki e Admi (2012), o modelo ARMAX (autorregressivo com médias móveis e entrada de variáveis exógenas) baseia-se em um modelo ARMA, onde é possível combiná-lo com uma estrutura de regressão (variáveis explicativas) na forma:

$$Z_t = \underbrace{\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p}}_{\text{autorregressivo (p)}} + \underbrace{\beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_r Y_{t-r}}_{\text{variável independente}} + \varepsilon_t - \underbrace{\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \dots - \theta_q \varepsilon_{t-q}}_{\text{médias móveis (q)}}$$

onde  $Z_t$  é a variável dependente,  $y_t$  é a variável independente,  $\varepsilon_t$  é ruído branco,  $\phi_i$  e  $\theta_i$  são os parâmetros desconhecidos,  $\beta_i$  é o coeficiente da variável exógena e  $\varepsilon_t$  é a sequência de resíduos do

modelo. Esta estrutura pode considerar a inclusão de mais variáveis explicativas (exógenas). Quando a série é não estacionária, sendo necessário aplicar diferenças na série, o modelo ARMAX passa a se denominar ARIMAX. Para Almeida (2008), quando se inclui ao modelo ARIMA a variação sazonal e variáveis externas explicativas, o novo modelo denomina-se de SARIMAX.

Conforme Chechi e Sanches (2013) e Lizzi (2012), o modelo SARIMAX é uma extensão do modelo ARIMA/SARIMA onde podem se incluir variáveis explicativas por meio de estrutura de regressão.

Desse modelo podem derivar outros modelos como ARIX (p,d,0,r[]), IMAX (0,d,p,r[]) ARMAX (p,0,q,r[]), ARX (p,0,0,r[]), MAX (0,0,q,r[]), de acordo com Andrade (2009).

Ainda, segundo este autor, as etapas para estimação de um modelo ARIMAX/SARIMAX são as mesmas do modelo ARIMA/SARIMA, com a diferença que, além da estimação das ordens dos coeficientes dos modelos, é preciso estimar os coeficientes das variáveis exógenas.

### 2.3 Modelo Holt-Winters

O método de Alisamento Exponencial de Holt-Winters (AEHW) é um dos mais utilizados para a previsão de curto prazo, devido a sua simplicidade, baixo custo de operação, boa precisão e capacidade de ajustamento automático rápido a mudanças na série, de acordo com Lúcio et al.(2010).

Para Nascimento (2013), o método de AEHW é uma técnica determinística, que faz parte de um conjunto de técnicas de alisamentos exponenciais indicados para séries que apresentam tendências e/ou sazonalidade. O método distingue-se dos demais, pois seu objetivo não é ajustar uma reta aos dados e sim uma reta a cada instante de tempo.

Cada um dos procedimentos (aditivo ou multiplicativo) requer a determinação de três constantes de alisamento (A, C e D), associadas a cada uma das componentes para sazonalidade, nível e tendência.

Os valores das constantes (A, C e D) são obtidos de modo que se obtenha a mínima soma dos quadrados dos erros de ajustamento das constantes de suavização e uma das vantagens do método é a facilidade de manuseio.

De acordo com Veríssimo et al. (2012) e Morettin e Tolo (2004), o modelo aditivo caracteriza-se pelo padrão sazonal não se alterar a medida que o nível da série muda. Já o modelo multiplicativo se caracteriza por ter a sazonalidade da série afetada pela tendência sendo o resultado do produto das componentes individuais. As equações do modelo estão descritos em Morettin e Tolo (2004).

### 2.4 Critérios de seleção dos Modelos de Box Jenkins

Os critérios a serem utilizados para verificar o ajuste do modelo autorregressivo integrado e de médias móveis são: AIC e BIC, relatados abaixo.

O critério de AIC (*Akaike Information Criteria*) proposto por Akaike (1974):

$$AIC = \ell_n(\hat{\sigma}_e^2) + \frac{2(p+q)}{n}$$

e o critério BIC (*Bayesian Information Criterion*) desenvolvido por Akaike (1978):

$$BIC = \ell_n(\hat{\sigma}_e^2) + \frac{(p+q)\ln(n)}{n},$$

onde  $\hat{\sigma}_e^2$  é variância estimada dos erros;  $n$  é o tamanho da amostra e  $(p,q)$  é o número de parâmetros de cada componente AR e MA. Os critérios levam em conta a variância do erro, o tamanho da amostra  $n$  e os valores de  $(p,q)$ .

Nesse sentido, vários modelos concorrentes são ajustados e, a partir dos valores dos critérios citados, é possível escolher o modelo mais adequado para realizar previsões  $h$  passos à frente. Quanto menores forem estes valores, melhor é a adequação do modelo para realizar previsões.

Além disso, a condição de ruído branco dos resíduos é considerada, ou seja, faz-se necessário que os erros do modelo devem ter média zero e variância constante, além de serem descorrelacionados ao longo do tempo, indicando que o modelo está tão bem ajustado restando apenas à parte aleatória que não pode ser explicada pelo modelo.

## 2.5 Critérios de comparação dos modelos

Os critérios de avaliação da qualidade das previsões mais utilizados para os modelos de alisamento exponencial são o Erro Quadrático Médio (EQM) e Soma dos quadrados das diferenças (SSE) e o MAD Média absoluta dos erros, descritos na Tabela 1, de acordo com Tidre et al. (2013), Esquivel et al. (2012) e Barbosa e Chaves (2012).

**Tabela 1** – Critérios de avaliação de previsão modelo Holt-Winters

Critério	Equação
Erro quadrático médio	$EQM = \frac{\sum_{t=1}^n (Z_t - \hat{Z}_t)^2}{n}$
Soma dos quadrados das diferenças	$SSE = \sum_{t=1}^n (Z_t - \hat{Z}_t)^2$
Média absoluta dos erros	$MAD = \frac{\sum_{t=1}^n  Z_t - \hat{Z}_t }{n}$
Erro Absoluto Percentual Médio	$MAPE = \frac{\sum_{t=1}^n \left  \frac{Z_t - \hat{Z}_t}{Z_t} \right }{n} \cdot 100\%$

onde:  $Z_t$  é o valor da série no instante  $t$ ;  $\hat{Z}_t$  é o valor previsto para o instante  $t$ ;  $n$  é o número período de tempos.

De acordo com Feliciano (2009), o Erro Absoluto Percentual Médio (MAPE) e MAD são as medidas de seleção mais utilizadas para a comparação de modelos. A diferença dos métodos encontra-se na

origem dos dados: o MAPE não depende de escala dos dados, podendo ser utilizado para comparar modelos com diferentes conjuntos de dados, já o MAD é útil para comparar diferentes modelos aplicados ao mesmo banco de dados.

### 3. METODOLOGIA

Para realização deste estudo, foram considerados os dados da série histórica correspondente à média semanal da umidade relativa do ar, compreendidos no período de 01 de outubro de 2001 a 22 de janeiro de 2014. Os dados correspondem a Estação Climatológica de Santa Maria - RS (OMM: 83936), os quais foram obtidos no Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) mantido pelo Instituto Nacional de Meteorologia (INMET), totalizando 6657 observações coletados em três períodos diariamente (12:00h, 18:00h e 00:00h). Essas observações foram transformadas em médias diárias e posteriormente em médias semanais, com o objetivo de facilitar a comparação das previsões dos modelos propostos.

Inicialmente foram calculadas algumas medidas descritivas da variável (média, mediana, desvio-padrão, coeficiente de variação, mínimo e máximo) com o objetivo de caracterizá-la. Posteriormente foram ajustados os modelos SARMA, ARMAX e Holt-Winters à série histórica. Para fins de comparação foram reservados as últimas vinte e uma observações da série original para verificar a qualidade do ajuste do modelo e comparar os modelos propostos dentro de cada classe considerada, utilizando os critérios MAPE, EQM, MAD e SSE. Como software de apoio foi utilizado o R.

## 4. 4 Análise dos resultados

### 4.1 Características Estatísticas da Série

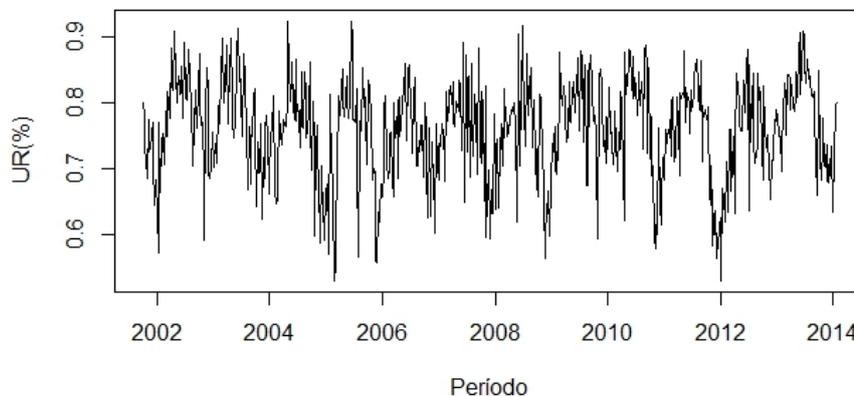
Para um melhor entendimento da variável UR, são apresentadas, na Tabela 2, algumas medidas descritivas, onde pode-se observar uma média elevada da UR, característica do clima local, com alta concentração de umidade no período de abril a agosto, considerado período de inverno e, as baixas concentrações, nos meses de dezembro e janeiro, período de verão. A UR manteve-se acima dos 50% em todos os períodos da série em estudo. A assimetria negativa indica que a curva possui uma cauda mais longa à esquerda. O valor do coeficiente de curtose demonstra uma distribuição platicúrtica, ou seja, os dados se concentram em maior número em torno da média que a distribuição normal.

**Tabela 2** - Medidas descritivas da UR de Santa Maria – RS, 2001 a 2014

Medida Descritiva	Resultado	Medida Descritiva	Resultado
Média	0,76037	Valor máximo	0,92476
Mediana	0,76952	Valor mínimo	0,53000
Desvio Padrão	0,07682	Curtose	-0,13351
Coeficiente de variação	10,103%	Assimetria	-0,51288

Inicialmente, pode-se observar, na Figura 1, o comportamento da série original buscando identificar suas componentes, como tendência e/ou sazonalidade.

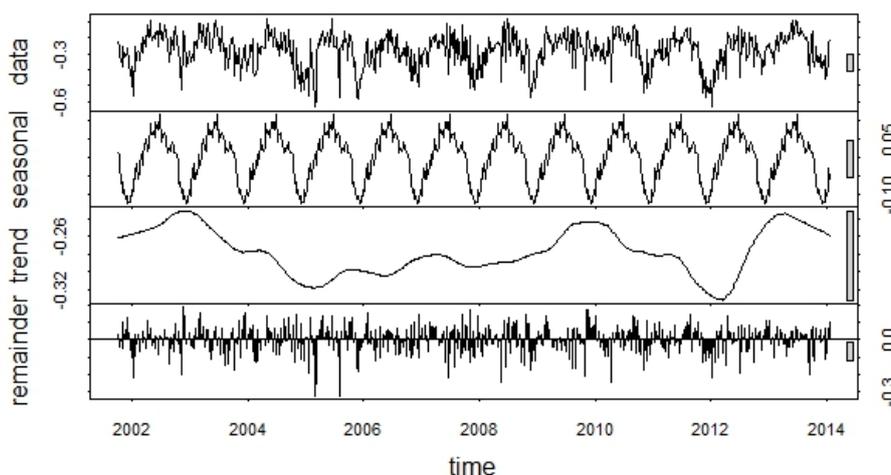
**Figura 1** - Série Original da Umidade Relativa do Ar em Santa Maria – RS



Analisando a Figura 1, verifica-se que a série possui indícios de sazonalidade, não demonstrando a presença de tendência, o que indica não ser necessário aplicar diferenciações na série original; entretanto, para confirmar a análise gráfica foram aplicados testes estatísticos para verificar a condição de estacionariedade e sazonalidade.

A decomposição da série original também permite identificar componentes importantes da série e confirmar se há presença de sazonalidade e/ou tendência. Na Figura 2, pode-se observar a decomposição da série em estudo nas suas componentes ciclo, sazonalidade, tendência e variações aleatórias.

**Figura 2** - Decomposição da série temporal em componentes de ciclo, sazonalidade, tendência e aleatoriedade



Observa-se, na Figura 2, que a série não possui tendência definida, mas há de se considerar a presença de sazonalidade. A série demonstra um comportamento definido ao longo do tempo e, portanto, pode ser ajustada num modelo de série temporal.

Após a etapa inicial de análise descritiva dos dados foram realizados testes estatísticos para confirmar as hipóteses de estacionariedade e sazonalidade levantadas na análise gráfica.

De acordo com o teste Dickey-Fuller, encontrou-se  $\hat{\tau} = -5,2621$  e  $\tau = -2,86$ , para o nível de significância  $\alpha = 5\%$ , rejeitando-se a hipótese nula  $H_0$  de que a série em estudo possui raiz unitária, aceitando a hipótese alternativa  $H_1$  de não possuir raiz unitária (estacionariedade dos dados).

Para o teste de sequências, tendo em vista  $n_1$  e  $n_2$  serem maior que 20 utilizou-se a aproximação para a normal  $T_1 \sim N(\mu, \sigma^2)$ , em que  $Z_{cal.} = 1,2522$ , e  $Z_{tab.} = 1,96$ , ao nível de significância  $\alpha = 5\%$ . De acordo com o teste, se  $Z_{cal.} < Z_{tab.}$  aceita-se a hipótese nula  $H_0$  de que não há indícios de tendências significativas na série em estudo, confirmando o teste Dickey-Fuller de estacionariedade dos dados.

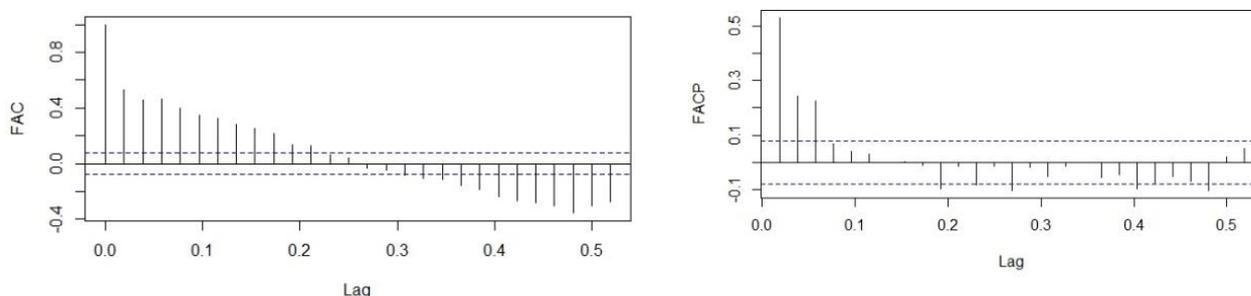
Para avaliar a sazonalidade foi realizado o teste Kruskal-Wallis sendo rejeitada a hipótese nula  $H_0$  (p-valor  $< 0,001$ ) e aceita a hipótese alternativa  $H_1$  de presença de efeito sazonal na série, ao nível de significância  $\alpha = 5\%$ .

Confirmando o resultado do teste Kruskal-Wallis, o teste de Friedman também rejeitou a hipótese nula  $H_0$  (p-valor  $< 0,001$ ), confirmando a presença de sazonalidade na série em estudo.

## 4.2 Modelagem SARMA

Os gráficos das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP), que estão expressos na Figura 3, apresentam as correlações seriais no domínio do tempo. Com base em suas frequências estimadas é possível verificar sazonalidades e ciclos na série.

**Figura 3** - Função de Autocorrelação (FAC) e Autocorrelação Parcial (FACP)



Ao se analisar os gráficos das FAC e FACP, percebe-se características de um modelo estacionário, autorregressivo de ordem 3 ou 4 e de médias móveis de ordem 1 ou 0. A FAC em nível permite identificar a presença de sazonalidade corroborando com os testes realizados de Friedman e Kruskal-Wallis, além do gráfico da decomposição (Figura 2), que indicam que um modelo do tipo SARMA deve ser considerado. Essa análise é sumária para dar início ao processo de modelagem, sendo que vários modelos de ordem diferentes vão ser considerados e testados.

Na Tabela 3, estão expostos os dois melhores modelos concorrentes ajustados para a série UR, com os respectivos valores de AIC e BIC, os parâmetros do modelo, a média e a variância dos resíduos referentes a cada modelo significativo.

**Tabela 3 - Modelos concorrentes SARMA e os respectivos critérios**

SARMA (p,q)(P,Q)	Parâmetros	Critérios de seleção		Média/Variância dos resíduos
		AIC	BIC	
(3,0)(1,2) <sub>12</sub>	$\phi_1 = 0,3389$	-1681,52	-1646,08	$\mu \cong 0,0001$ $\sigma^2 = 0,0379$
	$\phi_2 = 0,1471$			
	$\phi_3 = 0,2312$			
	$\Phi_1 = 0,0315$			
	$\theta_1 = -0,0617$			
	$\theta_2 = -0,0524$			
	$c = 0,7619$			
(3,0)(1,1) <sub>12</sub>	$\phi_1 = 0,3477$	-1682,08	-1651,07	$\mu \cong 0,0001$ $\sigma^2 = 0,0037$
	$\phi_2 = 0,1462$			
	$\phi_3 = 0,2315$			
	$\Phi_1 = 0,3413$			
	$\theta_1 = -0,3841$			
	$c = 0,7619$			

Considerando os dois melhores modelos significativos ajustados, analisou-se a condição de ruído branco (média zero, variância constante, erros não correlacionados). Na análise dos resíduos dos modelos SARMA, observou-se que a média dos modelos pode ser considerada como zero conforme Tabela 3.

Os testes de autocorrelação dos resíduos Ljung-Box (p-valor = 0,6936) e Box-Pierce (p-valor = 0,6943), do melhor modelo SARMA(3,0)(1,1)<sub>12</sub> de acordo com os critérios de seleção AIC e BIC, indicam que os erros do modelo ajustado são descorrelacionados ao longo do tempo, aceitando a hipótese nula  $H_0$ .

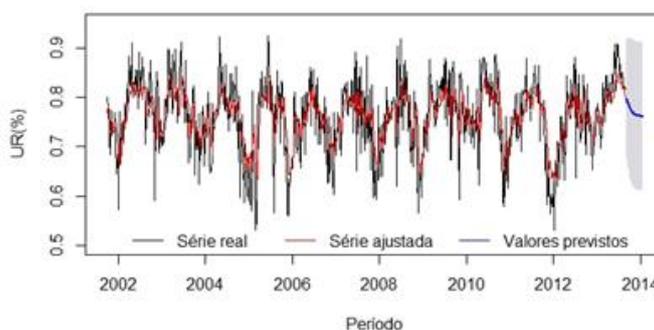
Os testes de normalidade dos resíduos Jarque-Bera (p-valor = 0,06005) e Shapiro-Wilk (p-valor = 0,07497) demonstram que os resíduos do modelo SARMA (3,0)(1,1)<sub>12</sub>, seguem distribuição normal e concluem a última etapa da escolha do modelo.

Considerando a análise residual e os critérios AIC e BIC, o modelo SARMA (3,0)(1,1)<sub>12</sub>, atendeu a todas as condições de ruído branco e normalidade dos resíduos sendo o escolhido para o ajuste do modelo e previsões.

Nessa etapa do trabalho, foram obtidas as previsões da UR para um horizonte de 21 passos à frente, ou seja, semanas que não foram consideradas no ajuste do modelo.

Na Figura 4 é demonstrada a série real, ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

**Figura 4 – Modelo SARMA (3,0)(1,1)<sub>12</sub> ajustado à série observada**



### 4.3 Ajuste do modelo ARMAX

Para o ajuste do modelo ARMAX e identificar a influência dessas variáveis na UR utilizaram-se as variáveis exógenas explicativas temperatura média (máxima e mínima) semanal.

Dentre os diversos modelos testados, na Tabela 5, estão apresentados os dois modelos com melhor desempenho nos critérios de seleção AIC e BIC.

**Tabela 5** - Modelos concorrentes ARMAX e os respectivos critérios

ARMAX (p,q,r[...])	Parâmetros	Critérios de seleção		Média/Variância dos resíduos
		AIC	BIC	
(3,0,[1,1])	$\phi_1 = 0,2185$ $\phi_2 = 0,1648$ $\phi_3 = 0,2012$ $c = 0,93334$ T. max. = -0,008 T. min. = 0,0021	-1773,64	-1742,63	$\mu \cong 0,0001$ $\sigma^2 = 0,0032$
(2,0,[1,1])	$\phi_1 = 0,2611$ $\phi_2 = 0,2117$ $c = 0,9450$ T. max. = -0,0081 T. min. = 0,0016	-1750,72	-1724,14	$\mu \cong 0,0001$ $\sigma^2 = 0,0034$

Considerando os dois melhores modelos significativos ajustados, analisou-se a condição de ruído branco (média zero, variância constante, erros não correlacionados). Na análise dos resíduos do modelo ARMAX(3,0,[1,1]), observou-se que a média do modelo pode ser considerada como zero conforme Tabela 5.

Os testes de autocorrelação dos resíduos Ljung-Box (p-valor = 0,7762) e Box-Pirce (p-valor = 0,7732), do melhor modelo ARMAX de acordo com os critérios de seleção AIC e BIC, indicam que os erros do modelo ajustado não são correlacionados ao longo do tempo.

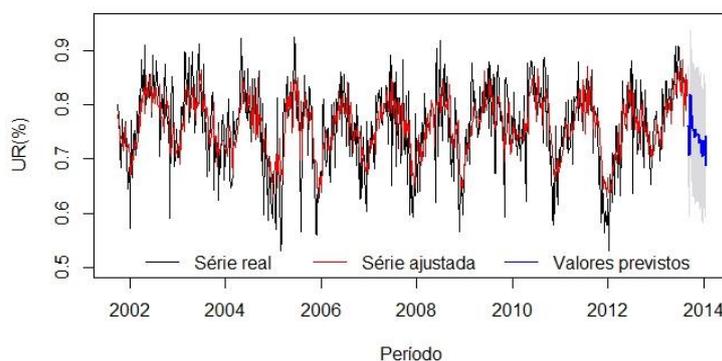
Os testes de normalidade dos resíduos Jarque-Bera (p-valor = 0,16005) e Shapiro-Wilk (p-valor = 0,1863) comprovam que os resíduos do modelo ARMAX(3,0[1,1]), seguem distribuição normal e concluem a última etapa da escolha do modelo que melhor se ajusta à série de dados para realizar previsões.

Considerando a análise residual e os critérios AIC e BIC, o modelo ARMAX(3,0,[1,1]) atendeu a todas as condições de ruído branco e normalidade dos resíduos, sendo o escolhido para o ajuste do modelo e previsões.

Nessa etapa do trabalho, foram obtidas as previsões da UR para um horizonte de 21 passos à frente, ou seja, semanas que não foram consideradas no ajuste do modelo, conforme a metodologia proposta.

Na Figura 5 é demonstrada a série ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

**Figura 5** – Modelo ARMAX(3,0,[1,1]) ajustado à série observada



#### 4.4 Alisamento Exponencial de Holt-Winters (AEHW)

Dentre os modelos de alisamento exponencial o que melhor se ajusta a dados sazonais, característica presente na série, é o modelo AEHW. Seguindo as características da série, também foi utilizado esse modelo para ajuste da previsão e comparação com os modelos da classe ARIMA.

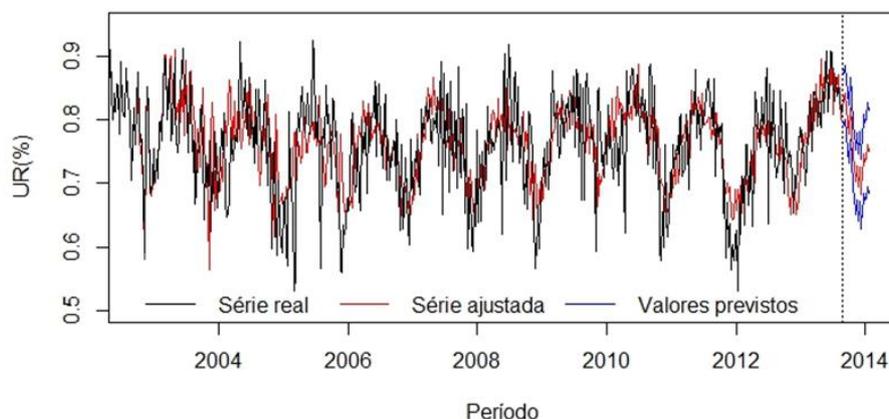
Foram ajustados dois modelos de (AEHW), o aditivo e multiplicativo e, posteriormente, foi feita a comparação utilizando os critérios de seleção.

No ajuste dos modelos, os valores das constantes de suavização que melhor se ajustaram ao modelo aditivo foram:  $A=0,2493845$  (constante de suavização do nível);  $C=0,0000$  (constante de suavização da tendência) e  $D=0,08384394$  (constante de suavização da sazonalidade) e, para o multiplicativo, foi:  $A=0,2377823$  (constante de suavização do nível);  $C=0,0000$  (constante de suavização da tendência) e  $D=0,08741191$  (constante de suavização da sazonalidade).

Utilizando os critérios de seleção MAPE, EQM, MAD e SSE, o modelo multiplicativo ajustou-se de forma mais adequada à série histórica em relação ao modelo aditivo.

Na Figura 6, podem ser observados os valores reais e previstos, além do intervalo de previsão ( $h=21$ ) do modelo Holt-Winters multiplicativo.

**Figura 6** - Holt-Winters multiplicativo



#### 4.5 Comparação dos modelos ajustados SARMA, Holt-Winters e ARMAX

Um dos objetivos de uma modelagem de séries temporais é fazer previsões de valores futuros. Para tanto, se faz necessário verificar a eficiência e o bom ajuste do modelo.

Os critérios a serem utilizados para comparação estão descritos na seção 3.5. Na Tabela 9, estão expostos os resultados encontrados para os modelos que melhor se ajustaram à série de dados.

**Tabela 9** - Comparação dos Modelos utilizando os critérios de seleção

Modelo	MAPE(%)	EQM	MAD	SSE
Holt-Winters multiplicativo	7,25413	0,00485	0,05149	0,10171
Holt-Winters aditivo	7,8564	0,00499	0,05278	0,10468
SARMA(3,0)(1,1) <sub>12</sub>	8,77223	0,00493	0,06159	0,10342
ARMAX(3,0,[1,1])	6,52960	0,00362	0,04643	0,07620

Considerando os resultados dos critérios propostos, o modelo ARMAX(3,0,[1,1]), apresentou melhor desempenho em todos os critérios utilizados em relação ao modelo SARMA(3,0)(1,1)<sub>12</sub> e o Holt-Winters multiplicativo e aditivo, demonstrando que as previsões se aproximam de forma mais adequada à série real em estudo.

#### 5. CONSIDERAÇÕES FINAIS

Ao final do presente estudo, observou-se que a série histórica de UR em Santa Maria – RS possui sazonalidade, com elevados valores de umidade no período de abril a agosto e, menores valores, nos meses de dezembro e janeiro.

As classes de modelos aqui utilizadas e comparadas para dados de Umidade Relativa do Ar demonstraram bom desempenho e ajuste. Verificando que os modelos podem ser empregados para taxas e proporções, desde que sejam seguidos todos os pressupostos exigidos pelos modelos.

Em relação à comparação do desempenho dos modelos, conforme os resultados obtidos, pode-se considerar que o modelo ARMAX com a inclusão das variáveis exógenas temperaturas média (mínima e máxima) ajustou-se de forma adequada à série de dados e apresentou melhores resultados de previsão. O modelo Holt-Winters também se aproximou de forma adequada à série de dados estudada e devido a sua fácil manipulação deve ser levando em consideração. Não se descarta o modelo SARMA ou outros modelos estatísticos possam produzir melhores resultados de previsão em outras séries de dados.

Em estudos futuros, pretende-se empregar modelos estatísticas para outras variáveis climáticas.

## 6. REFERÊNCIAS

- AKAIKE, H. (1974). A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v.19, p.716-723.
- AKAIKE, H. (1978). A Bayesian analysis of the minimum AIC procedure. **Annals of the Institute of Statistical Mathematics**, v.30, p.9-14.
- ALMEIDA, E. S. (2008). **Previsão de carga a curto prazo: uma comparação de modelos**. (Dissertação de Mestrado em Engenharia de Produção) Universidade Estadual do Norte Fluminense (UENF), Rio de Janeiro.
- ANDRADE, N. S.; CARVALHO, P. S.; RUSSO, S. L. (2011). Análise das variações de temperaturas na plataforma de Nossa Senhora da Glória via Box e Jenkins. **Revista GEINTEC**, v.1, p. 47-54.
- ANDRADE, W. L. S. (2009) **Estimação de Modelos ARIMA/ARIMAX e Aplicação em Inferência de Propano**. (Dissertação de Mestrado em Engenharia Elétrica e Computação), Universidade Federal do Rio Grande do Norte (UFRN), Natal.
- ANTONINI, J. C. A.; SILVA, E. M.; OLIVEIRA, L. F. C.; SANO, E. E. (2009). Modelo Matemático para a Temperatura média diária do ar do Estado de Goiás. **Pesquisa Agropecuária Brasileira**, Brasília, v.44, n.3, p.333-338.
- BARBOSA, C. M.; CHAVES, C. A. (2012). Um estudo sobre gerenciamento da demanda nos sistemas de planejamento e controle da energia de produção. **VIII Congresso Nacional de Excelência em Gestão**, de 08 a 09 de junho, São Paulo.
- CAVALCANTI, E. P.; SILVA, V. P. R.; SOUSA, F. A. S. (2006). Programa computacional para a estimativa da temperatura do ar para a Região Nordeste do Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.10, p.140-147.
- CHECHI, L.; SANCHES, F. O. (2013). Análise de uma série temporal de precipitação para Erechim (RS) e um possível método de previsão climática. **Ambiência Guarapuava**, v.9, p. 43-45.
- ESQUIVEL, R. M.; SENNA, V.; GOMES, G. S. S. (2012). Análise espectral singular: comparação de previsão em séries temporais. **Revista do Mestrado em Administração e Desenvolvimento Empresarial**, v.16, n.2, p. 87-101.
- ESPINOSA, M. M.; PRADO, S. M.; GHELIERRE, M. (2010). Uso de modelo SARIMA na previsão de focos de calor para os meses de junho a outubro no Estado do Mato Grosso. **Ciência em Natura**, v.32, n.2, p 7-21.
- FELICIANO, R. A. (2009). **Uma proposta de gerenciamento integrado de demanda e distribuição, utilizando sistemas de apoio à decisão (SAD) com Bussiness Intelligence (BI)**. (Dissertação de Mestrado em Engenharia) Politécnica Universidade de São Paulo, São Paulo.
- INMET, 2014. Dados da rede **Instituto Nacional de Meteorologia**. Disponível em: <http://www.inmet.gov.br/>, acesso em 02/02/2014.
- LIZZI, E. A. S. **Predição do número de casos mensal dos casos de dengue por modelos de séries temporais**. 2012, 131f. Dissertação (Mestrado em Saúde na Comunidade) Universidade de São Paulo (USP), Ribeirão Preto, 2012.

LÚCIO, P. S.; SILVA, F. D. S.; FORTES, L. T. G.; SANTOS, L. A. R.; FERREIRA, D. B.; SALVADOR, M. A. S.; BALBINO, H. T.; SARMANHO, G. F.; SANTOS, L. S. F. C.; LUCAS, E. W. M.; BARBOSA, T. F.; DIAS, P. L.S. (2010). Um modelo estocástico combinado de previsão sazonal para a precipitação no Brasil. . **Revista Brasileira de Meteorologia**, v.25, n.1, p. 70-87.

MARTÍNEZ, R. O.; ZAMPROGNO, B. (2003). Comparação de algumas técnicas de previsão em análise de séries temporais. **Revista Colombiana de Estatística**, v.26, n.2, p. 129-157.

MORETTIN, P. A.; TOLOI, C. M. C. (2004) **Análise de Séries Temporais**. São Paulo: Edgard Blucher LTDA.

NASCIMENTO, H. H. S. (2013). **Planejamento de mercado de energia elétrica no novo modelo do setor elétrico brasileiro pela aplicação de análise de componente principal como métodos de combinação de previsão**. (Dissertação de Mestre em Engenharia Elétrica) Universidade Federal do Mato Grosso do Sul (UFMS), Campo Grande.

OZAKI, V. A.; ADAMI, A. C. O. (2012). Modelagem estatística dos prêmios do seguro rural. **Revista Política Agrícola**, v.21, n.1, p. 60-75.

REISEN, V. A.; MOLINARES, F. A. A.; TEIXEIRA, E. C. (2008). Modelagem de Séries Temporais Sazonais na Presença de Outliers Estudo de Caso da Vazão Máxima Mensal do Rio Jucu, ES, Brasil. **Revista Brasileira de Recursos Hídricos**, v.13, n.2.

SILVA, M. I. S.; GUIMARÃES, E. C.; TAVARES, M. (2008). Previsão da temperatura média mensal de Uberlândia, MG, com modelos de séries temporais. **Revista Brasileira de Engenharia agrícola e Ambiental**, v.12, p.480-485.

SILVA, L. C. (2011). **Utilizando o ARMAX para estimar a influência do PIB, SELIC e Inflação no faturamento de empresas do setor siderúrgico com ações na bolsa**. (Monografia de Bacharelado em Administração), Universidade de Brasília (UNB), Brasília.

SOEBIYANTO, R. P.; ADIMI, F.; KIANG, R. K. (2010). Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. **PLoS ONE**, v.5, p.1-10.

TIDRE, P. V.; BIASE, N. G.; SILVA, M. I. S. (2013). Utilização dos modelos de séries temporais na previsão do consumo mensal de energia elétrica da região norte do Brasil. **Revista Eletrônica Matemática e Estatística em Foco**, v.1, n.1, p. 57-63.

VERÍSSIMO, A. J.; ALVES, C. C.; HENNING, E.; AMARAL, C. E.; CRUZ, A. C. (2012). Métodos estatísticos de suavização exponencial Holt-Winters para a previsão da demanda em uma empresa do setor metal mecânico. **Revista Gestão Industrial**, v.8, n.4, p. 154-171.

Submissão: 20/02/2018

Aceito: 04/06/2018