

CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ
ISSN impresso 1413-9022 / ISSN on-line 2317-4536 - v. 43, p.39 - 59, 2017
DOI: 10.12957/cadest.2017.31211

PREVISÃO DA UMIDADE RELATIVA DO AR DIÁRIA DE PORTO ALEGRE

Liane Werner

Universidade Federal do Rio Grande do Sul
liane.werner@ufrgs.br

Cleber Bisognin

Universidade Federal do Rio Grande do Sul
cbisognin@ufrgs.br

Letícia Menegotto

Universidade Federal do Rio Grande do Sul
leticia.menegotto@gmail.com

Resumo

A umidade relativa interfere em vários aspectos na vida do ser humano, e devido as muitas consequências que um baixo ou um alto percentual podem acarretar, o controle de seu nível é de suma importância. Desta forma, realizar previsões acuradas da umidade relativa do ar com vistas a subsidiar as decisões necessárias nos mais variados segmentos, torna-se fundamental. Posto isto, este artigo tem como objetivo comparar vários modelos e métodos de previsão, buscando encontrar o mais acurado para prever adequadamente a umidade relativa do ar diária de Porto Alegre. Para tanto foram obtidos cinco modelos (suavização exponencial, SARIMA e ARFIMA), além das combinações dois a dois de cada um dos cinco modelos para três métodos de combinação: variância mínima, média e regressão. Por meio das medidas de acurácia RSME, MAPE e coeficiente U de Theil verificamos que a combinação de previsões por regressão entre o Modelo 2 (Modelo de Suavização Exponencial Multiplicativo) e Modelo 3 (Modelo ARFIMA (2,d,0)) obteve as menores medidas de acurácia em todos os caso, sendo utilizada para realizar as previsões desejadas.

Palavras-chave: Umidade Relativa do Ar, Combinação de Previsões, Modelagem.

1. Introdução

Para o adequado desenvolvimento de diversos setores da atividade humana - tal como o setor agrícola ou econômico - é preciso entender os comportamentos climáticos de uma determinada região. O estudo de variáveis meteorológicas é chave para o entendimento do clima, pois com elas é possível identificar melhores épocas de plantio ou mapear os riscos de eventos climáticos extremos (CARVALHO *et al.*, 2016).

Uma característica meteorológica importante é a umidade relativa do ar (ou simplesmente umidade relativa), seja na saúde, no planejamento de irrigação ou em controles ambientais. Quando a umidade relativa está alta, além de causar problemas respiratórios, pode ser responsável pelo aumento das precipitações, e estas em excesso podem causar inundações, deslizamentos de terra em áreas urbanas ou danos à agricultura.

Outro incidente que tem vínculo com a umidade relativa do ar é aumento potencial de incêndios em pastagens e florestas. Incêndios florestais no Brasil e no mundo resultam em mudanças drásticas no ambiente e danos econômicos. Segundo Batista (2000), a umidade é um dos fatores mais importantes na propagação de incêndios florestais, pois existe uma troca entre o vapor d'água e o combustível (em geral, material morto e seco depositado no piso da floresta).

A umidade relativa do ar é expressa como a razão entre a quantidade efetiva de vapor d'água no ar e a quantidade máxima de vapor d'água que a mesma quantidade de ar poderia conter, se estivesse saturada desta substância, em determinada temperatura (VIRGENS *et al.*, 2009).

Umidade relativa do ar elevada deixa janelas, pisos e paredes molhados. Pisos molhados se tornam escorregadios podendo ocasionar quedas. Além disto, muita umidade gera mofo permitindo aos fungos se proliferarem causando crises de rinite e outras doenças respiratórias. Já a umidade relativa baixa pode causar problemas de saúde, tais como: alergias, ataques de asma, desidratação, sangramento nasal.

Adicionado a estes fatos, tem ainda que os seres humanos são muito sensíveis à umidade, pois a pele depende do ar no processo de transpiração. O processo de transpiração é a tentativa do corpo de manter a temperatura atual. Se o ar estiver com 100 por cento de umidade relativa, o suor não se evaporará no ar e o corpo sente mais quente

do que a real temperatura. Se a umidade relativa for baixa, o corpo sente mais frio do que a temperatura real, porque o nosso suor evapora facilmente.

Segundo o Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI, 2017), existem alguns cuidados com relação aos diferentes níveis de umidade relativa do ar: (i) Estado de Atenção, quando a umidade relativa está entre 20% e 30%, nesse caso, recomenda-se evitar exercícios e caminhadas nas horas mais quentes do dia e com maior insolação, geralmente entre as 11h e 15h no horário convencional; (ii) Estado de Alerta quando a umidade relativa está entre 12 e 20%, nesta situação, o tempo de proibição dos exercícios ao ar livre aumenta, indo das 10h às 16h. Também é preciso buscar sempre a hidratação e procurar umidificar o ar. (iii) Estado de emergência, quando ela se encontra abaixo de 12% e além de todas as medidas anteriores, é necessário evitar exercícios físicos durante o dia e procurar sempre se manter hidratado com água, umidificadores e panos úmidos sobre o rosto e as narinas, evitando a irritação das vias nasais.

Como exposto, a umidade relativa interfere em vários aspectos na vida do ser humano, e em vista das muitas consequências que um baixo ou um alto percentual podem acarretar, o controle de seu nível é de suma importância, principalmente para a saúde. Desta forma, realizar previsões acuradas da umidade relativa do ar com vistas a subsidiar as decisões necessárias nos mais variados segmentos, torna-se fundamental. Posto isto, este artigo tem como objetivo comparar vários modelos e métodos de previsão, buscando encontrar o mais acurado para prever adequadamente a umidade relativa do ar diária de Porto Alegre.

O presente artigo é composto por cinco seções. A primeira seção consiste nesta introdução, após é apresentado um referencial teórico sobre os modelos e métodos de previsões, na sequência o procedimento metodológico utilizado e resultados e discussões são detalhados na quarta seção e para finalizar são expostas as considerações finais.

2. Técnicas de Previsão

Nesta seção são apresentados os modelos usados para realizar a análise e previsão da série temporal da Umidade Relativa do Ar diária de Porto Alegre, a saber, os modelos $SARIMA(p,d,q) \times (P,D,Q)_s$, $ARFIMA(p,d,q)$ e Suavização Exponencial, além de três

métodos para realizar combinações de previsões: variância mínima, por regressão e média aritmética.

A seguir são definidos os processos SARIMA(p, d, q) \times (P, D, Q) $_s$ proposto por Box; Jenkins (1976).

Definição 1: Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação (1)

$$\Phi(B^s)\phi(B)(1 - B^s)^D(1 - B)^d(X_t - \mu) = \theta(B)\Theta(B^s)\varepsilon_t, \quad (1)$$

onde: μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é o processo ruído branco, $s \in \mathbb{N}$ é a sazonalidade, B é o operador defasagem ou de retardo, isto é, $B^j(X_t) = X_{t-j}$ e $B^{sj}(X_t) = X_{t-sj}$, para $j \in \mathbb{N}$, $\phi(\cdot)$, $\theta(\cdot)$, $\Phi(\cdot)$ e $\Theta(\cdot)$ são os polinômios de ordem p , q , P e Q , respectivamente, definidos pelas equações (2) e (3).

$$\phi(z) = \sum_{\ell=1}^p (-\phi_\ell)z^\ell, \quad \theta(z) = \sum_{m=1}^q (-\theta_m)z^m, \quad (2)$$

$$\Phi(z) = \sum_{r=1}^P (-\Phi_r)z^r, \quad \Theta(z) = \sum_{l=1}^Q (-\Theta_l)z^l, \quad (3)$$

com $\phi_\ell, 1 \leq \ell \leq p$, $\theta_m, 1 \leq m \leq q$, $\Phi_r, 1 \leq r \leq P$ e $\Theta_l, 1 \leq l \leq Q$, constantes reais e $\phi_0 = \Phi_0 = -1 = \theta_0 = \Theta_0$. Então $\{X_t\}_{t \in \mathbb{Z}}$ é um processo sazonal autorregressivo integrado de média móvel de ordem $(p, d, q) \times (P, D, Q)_s$ com sazonalidade s , denotado por SARIMA(p, d, q) \times (P, D, Q) $_s$, onde d e D são, respectivamente, o grau de diferenciação e o grau de diferenciação sazonal.

Durante as últimas décadas, houve muito interesse em estudar séries temporais com a propriedade de longa dependência. O estudo de séries temporais com esta característica foi apresentado, inicialmente por Hurst em 1951, enquanto investigava a série temporal das vazões do rio Nilo.

Utilizando a definição de longa dependência, Granger; Joyeux (1980), Hosking (1981), Hosking (1984) e Geweke; Porter-Hudak (1983) apresentam os processos autorregressivos fracionariamente integrados de média móvel (ARFIMA(p, d, q)) como um exemplo de processos com a característica de longa dependência. A seguir os processos ARFIMA(p, d, q) são definidos.

Definição 2: Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação (4)

$$\phi(B)(1 - B)^d(X_t - \mu) = \theta(B)\Theta\varepsilon_t, \quad (4)$$

onde: μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é o processo ruído branco, B é o operador defasagem ou de retardo, isto é, $B^j(X_t) = X_{t-j}$, para $j \in \mathbb{N}$, $\phi(\cdot)$ e $\theta(\cdot)$, são os polinômios de ordem p e q , respectivamente, definidos na equação (2), com ϕ_ℓ , $1 \leq \ell \leq p$ e θ_m , $1 \leq m \leq q$, constantes reais e $\phi_0 = -1 = \theta_0$. Então $\{X_t\}_{t \in \mathbb{Z}}$ é um processo autorregressivo fracionariamente integrado de média móvel de ordem (p, d, q) , denotado por ARFIMA(p, d, q), onde d é o grau de diferenciação.

Hosking (1981), demonstra que os processos ARFIMA (p, d, q) são estacionários se $d < \frac{1}{2}$ e as raízes da equação $\phi(z) = 0$ estão fora do círculo unitário; e é invertível se $d > -\frac{1}{2}$ e as raízes da equação $\theta(z) = 0$ estão fora do círculo unitário.

Os modelos de suavização exponencial são os mais utilizados frente a outras técnicas de previsão devido a sua simplicidade e facilidade de ajustes além da boa acurácia (PELLEGRINI, 2000). Assumindo que os valores extremos da série são flutuações aleatórias, o propósito destes modelos é identificar um padrão básico (MORETTIN; TOLOI, 2006). Estes modelos valorizam mais os últimos valores observados na série temporal através da ponderação exponencial dos mesmos, de acordo com a proximidade ao período da previsão h . Os métodos mais tradicionais de suavização exponencial são: (i) a suavização exponencial simples, para séries que apresentam apenas variações em torno de um nível; (ii) o modelo linear de Holt, para as séries que apresentam a componente de tendência e (iii) os modelos de Holt-Winters, quando a série apresenta tanto o componente de tendência quanto o componente sazonal (MAKRIDAKIS et al., 1998).

A seguir são definidos os modelos de Holt-Winters para efeitos sazonais aditivos e multiplicativos. Maiores detalhes podem ser encontrados em Makridakis *et al.* (1998) e em Morettin; Toloi (2006).

Considere uma série temporal $\{X_t\}_{t=1}^n$ com período s .

Definição 3: Para os Modelos de Holt-Winters para efeitos sazonais aditivos, considera-se que tal série é formada pela soma do nível, tendência, um fator sazonal e um erro aleatório, como apresentado pela equação (5).

$$X_t = L_t + T_t + S_t + \varepsilon_t, \text{ para } t = 1, \dots, n. \quad (5)$$

As estimativas do nível da série no tempo t , denotado por L_t , da tendência, denotada por T_t e do fator sazonal, denotado por S_t , são dadas, respectivamente pelas equações de (6) a (8).

$$L_t = \alpha(X_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (6)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (7)$$

$$S_t = \gamma(X_t - L_t) + (1 - \gamma)S_{t-s}, \quad (8)$$

onde: α é o coeficiente de ponderação exponencial do nível ($0 \leq \alpha \leq 1$), β é o coeficiente de ponderação exponencial da tendência ($0 \leq \beta \leq 1$) e γ é o coeficiente de ponderação exponencial do fator sazonal ($0 \leq \gamma \leq 1$).

As previsões h passos à frente, são dadas pela expressão dada pela equação (9)

$$\hat{X}_t(h) = L_t + hT_t + S_{t-s+h} \quad (9)$$

A notação $\hat{X}_t(h)$ indica a previsão de origem t e horizonte $h \geq 1$.

Definição 4: Para os Modelos de Holt-Winters para efeitos sazonais multiplicativos, considera-se que tal série é formada por uma tendência aditiva, por um fator sazonal multiplicativo e um erro aleatório, como apresentado pela equação (10).

$$X_t = L_t \times S_t + T_t + \varepsilon_t, \text{ para } t = 1, \dots, n. \quad (10)$$

As estimativas do nível da série no tempo t , denotado por L_t , da tendência, denotada por T_t e do fator sazonal, denotado por S_t , são dadas, respectivamente pelas equações de (11) a (13).

$$L_t = \alpha \frac{X_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (11)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (12)$$

$$S_t = \gamma \frac{X_t}{L_t} + (1 - \gamma)S_{t-s}, \quad (13)$$

onde: α é o coeficiente de ponderação exponencial do nível ($0 \leq \alpha \leq 1$), β é o coeficiente de ponderação exponencial da tendência ($0 \leq \beta \leq 1$) e γ é o coeficiente de ponderação exponencial do fator sazonal ($0 \leq \gamma \leq 1$).

As previsões h passos à frente, são dadas conforme a equação (14)

$$\hat{X}_t(h) = (L_t + hT_t) \times S_{t-s+h}. \quad (14)$$

A notação $\hat{X}_t(h)$ indica a previsão de origem t e horizonte $h \geq 1$.

Um método comumente utilizado para melhorar a acurácia das previsões é a combinação de previsões, segundo Costantini; Pappalardo (2010), este método consiste

em utilizar um mecanismo para captar os diversos fatores que afetam cada técnica de previsão individual usada como base de obtenção da previsão combinada.

O método da variância mínima, proposto por Bates; Granger (1969) consiste em realizar a combinação linear de duas previsões com diferentes pesos. Neste método a combinação das previsões é obtida atribuindo-se um peso para cada uma das previsões individuais que serão combinadas. Sua estrutura é apresentada conforme equação (15).

$$F_c = wF_1 + (1 - w)F_2 \quad (15)$$

onde: w é o peso atribuído à previsão de menor variância dos erros e F_1 e F_2 são as previsões individuais (modelos de séries temporais) a serem combinadas.

Para a obtenção dos pesos descritos na equação (15) é interessante atribuir menor peso às previsões de maior variabilidade nos erros e considerar a correlação existente entre os erros (absolutos) das duas previsões individuais realizadas. Os pesos são obtidos conforme equação (16).

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (16)$$

onde: ρ é o valor da correlação linear entre os erros das previsões obtidas em F_1 e F_2 , σ_1^2 e σ_2^2 são as variâncias dos erros (absolutos) de previsão de F_1 e F_2 , respectivamente.

Outra forma de realizar previsão por combinação é o método considerado o mais popular de combinação de previsões individuais, a média aritmética. Na equação (17) é apresentada a expressão para a combinação de duas previsões individuais. Além de ser um dos métodos mais conhecidos e utilizados é fácil de calcular. Segundo Menezes *et al.* (2000), uma possível resposta para o sucesso da média pode estar associada à instabilidade dos pesos ao longo do tempo na matriz de covariância dos erros das previsões individuais.

$$F_c = \frac{F_1 + F_2}{2} \quad (17)$$

onde: F_1 e F_2 são as previsões individuais a serem combinadas.

Na concepção de Granger; Ramanathan (1984) a combinação de previsões poderia ser uma forma estruturada de regressão, conforme consta na equação (18).

$$F_c = \beta_0 + \beta_1 F_1 + \beta_2 F_2 \quad (18)$$

onde: β_0 é uma constante, β_1 é o peso da previsão 1, β_2 é o peso da previsão 2 e F_1 e F_2 são as previsões individuais (modelos de séries temporais) a serem combinadas.

A proposta dos autores para obter os pesos de cada previsão utilizada na combinação consiste em estimar os coeficientes da regressão pelo Método dos Mínimos Quadrados Ordinários. Para realizar a regressão a variável dependente consiste dos valores observados dentro da amostra e as previsões individuais como variáveis explicativas.

Para optar por qual técnica de previsão - individual ou combinação - é a mais adequada, faz-se necessário obter medidas de acurácia (qualidade do ajuste). De acordo com Kim; Kim (2016) a média dos erros percentuais absolutos (MAPE) é uma das medidas de acurácia mais utilizada, devido as vantagens de interpretação e independência na escala. Essa medida está expressa na equação (19).

$$MAPE = \frac{\sum_{t=1}^n \left| (Y_t - \hat{Y}_t) / Y_t * 100 \right|}{n} \quad (19)$$

onde: Y_t é o valor observado no tempo t , \hat{Y}_t é o valor estimado no tempo t , n é o número de observações, $t = 1, 2, \dots, n$.

Se a acurácia de técnicas de previsão é avaliada em séries temporais que apresentam a mesma escala, conforme Shcherbakov *et al.* (2013) uma das medidas sugeridas é a raiz quadrada da média dos erros quadráticos (RMSE), dada pela equação (20). Além disto, esta medida tem sido usada como métrica padrão para medir o desempenho de técnicas de previsão em estudos de meteorologia, qualidade do ar e pesquisa climática (CHAI; DRAXLER, 2014).

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}} \quad (20)$$

onde: Y_t é o valor observado no tempo t , \hat{Y}_t é o valor estimado no tempo t , n é o número de observações, $t = 1, 2, \dots, n$.

O coeficiente de desigualdade de Theil é uma medida de acurácia frequentemente citada na literatura. No entanto, segundo Bliemel (1973), existe confusão sobre esse coeficiente, que pode ser resultado do fato de que o próprio Theil propôs duas fórmulas diferenciadas em momentos distintos, porém com o mesmo nome. Ambos são chamados de "coeficiente de desigualdade" e rotulados com o símbolo "U." A primeira proposta do U de Theil é delimitada entre 0 e 1. Já na segunda proposta o limite superior é infinito.

Neste estudo será utilizada a primeira proposta do U de Theil e sua expressão é dada conforme a equação (21).

$$U \text{ de Theil} = \sqrt{\frac{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\left[\sqrt{\frac{1}{n} \sum_{t=1}^n Y_t^2} + \sqrt{\frac{1}{n} \sum_{t=1}^n \hat{Y}_t^2} \right]}} \quad (21)$$

onde: Y_t é o valor observado no tempo t , \hat{Y}_t é o valor estimado no tempo t , n é o número de observações, $t = 1, 2, \dots, n$.

O coeficiente U de Theil analisa a qualidade das previsões, sendo que, quanto mais próximo de zero significa que o erro de previsão, gerada por determinado modelo ou combinação, é menor que da previsão ingênua, ou seja, que uma previsão é melhor que a previsão trivial (MAKRIDAKIS *et al.*, 1998).

3. Procedimentos Metodológicos

A pesquisa em questão é de natureza aplicada, pois busca a solução de problemas específicos e foi realizada para gerar conhecimentos sobre a umidade relativa do ar diária de Porto Alegre. A abordagem da pesquisa foi quantitativa, uma vez que os dados coletados são analisados por meio de modelagem estatística. Neste trabalho é realizado um estudo de caso, que de acordo com Gil (2017), consiste no profundo e exaustivo estudo de um ou poucos objetos, de maneira que permita seu amplo e detalhado conhecimento.

Na etapa de obtenção de dados, os dados históricos da umidade relativa do ar diária de Porto Alegre serão obtidos do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP), mantido pelo INMET - Instituto Nacional de Meteorologia (<http://www.inmet.gov.br>). O Banco de Dados Meteorológicos para Ensino e Pesquisa é um banco de dados para apoiar as atividades de ensino e pesquisa e outras aplicações em meteorologia, hidrologia, recursos hídricos, saúde pública, meio ambiente, etc. Serão utilizadas as 2706 observações diárias disponíveis, de 01/01/2010 a 29/05/2017. Os dados foram acessados em 04/08/2017. O uso deste período se deve ao fato de o período de dados mais completo disponíveis na internet e que foram encontrados pelos autores.

Na etapa de análise preliminar, os dados históricos serão representados graficamente, auxiliando assim na identificação do comportamento da série temporal. Nesta etapa também se busca identificar padrões, tendências na série temporal estudada,

para assim fornecer informações mais confiáveis para a escolha do modelo quantitativo a ser adotado (MAKRIDAKIS *et al.*, 1998). O ajuste dos modelos SARIMA e ARFIMA foi realizado utilizando, respectivamente, as rotinas *auto.arima* e *arfima* do pacote *forecast* do software R 3.4.0. Para os modelos de suavização exponencial foi utilizado a rotina *HoltWinters*, do mesmo software. A função busca encontrar valores ótimos para α , e/ou β , e/ou γ minimizando o erro quadrado de previsão de um passo à frente quando nenhum dos parâmetros de suavização é informado pelo usuário.

Para a estimação dos parâmetros dos processos SARIMA $(p,d,q) \times (P,D,Q)_s$, a rotina *auto.arima* calcula a verossimilhança exata via representação de Estado de Espaço do modelo enquanto as inovações são encontradas via Filtro de Kalmann. A estimação dos coeficientes dos polinômios utiliza a estacionariedade do processo e é baseada em Gardner *et al.* (1980).

No caso dos processos ARFIMA (p,d,q) , utilizamos a rotina *arfima*, do pacote *forecast* que seleciona automaticamente o melhor modelo baseado na log-verossimilhança e na análise de resíduos.

Uma vez que os modelos a serem utilizados foram definidos e os seus parâmetros foram estimados, serão aplicadas as técnicas de combinação de variância mínima, média aritmética e por regressão, implementadas no software R. Tais combinações serão obtidas somente utilizando dois modelos individuais, a combinação de três modelos não será analisada neste estudo.

A verificação da acurácia da técnica ou modelo a ser escolhido para realizar as predições para o período de 30/05/2017 a 13/06/2017, se dará por meio do RMSE (raiz quadrado da média dos erros quadráticos), pelo MAPE (média dos erros percentuais absolutos) e pelo U de Theil. As medidas foram calculadas utilizando-se a rotina *accuracy* do pacote *forecast*.

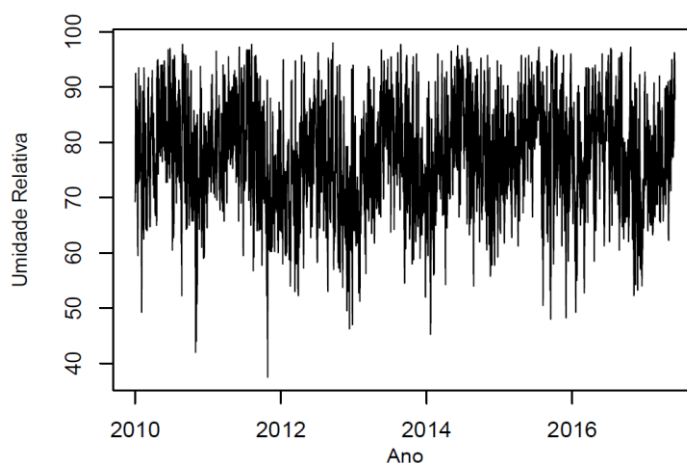
4. Resultados e Discussões

Uma vez que se busca uma técnica adequada prever a umidade relativa do ar diária de Porto Alegre, capital do Estado do Rio Grande do Sul, obtivemos as previsões utilizando os modelos de Suavização Exponencial, ARFIMA (p,d,q) e SARIMA $(p,d,q) \times (P,D,Q)_s$ e suas respectivas combinações.

Como análise preliminar construímos um gráfico temporal. A Figura 1 apresenta o gráfico da série temporal da umidade relativa do ar diária de Porto Alegre, Brasil, de 01/01/2010 a 29/05/2017, de onde se observa que a série parece apresentar sazonalidade, porém não apresenta tendência.

Foram aplicados os testes de raiz unitária de Dickey-Fuller e de Phillips-Perron. Os testes foram realizados utilizando, respectivamente, as rotinas *adf.test* e *pp.test*, do pacote *tseries* do R. Para os dois testes, as hipóteses avaliadas foram: H_0 : série temporal é não estacionária versus H_1 : série temporal é estacionária. Ambos os testes apresentaram p-valor $< 0,01$ e os resultados indicam que a série temporal é estacionária.

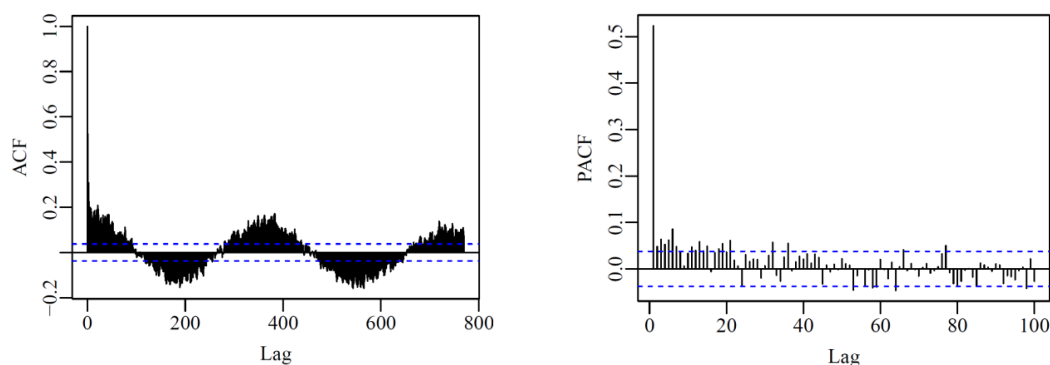
Figura 1 – Série temporal da umidade relativa do ar diária de Porto Alegre, de 01/01/2010 a 29/05/2017.



Fonte: Os autores.

A Figura 2 apresenta as funções de autocorrelação amostral e autocorrelação parcial amostral da série temporal da umidade relativa do ar diária de Porto Alegre. Podemos perceber, pelo gráfico da série temporal e pela sua função de autocorrelação amostral, que a série apresenta as características de longa dependência e sazonalidade mensal. Ambas as características também podem ser observadas através da Figura 3 (sazonalidade) e das Tabelas 1 e 2 (longa dependência).

Figura 2 – Funções de Autocorrelação e Autocorrelação Parcial amostrais da Série Temporal da Umidade Relativa do Ar Diária de Porto Alegre, de 01/01/2010 a 29/05/2017.



Fonte: Os Autores.

Uma vez que se têm indícios dos componentes presentes na série, neste caso apenas sazonalidade. Passamos a estimar os modelos indicados. A seguir são apresentados os modelos que foram ajustadas à Série Temporal da Umidade Relativa do Ar de Porto Alegre.

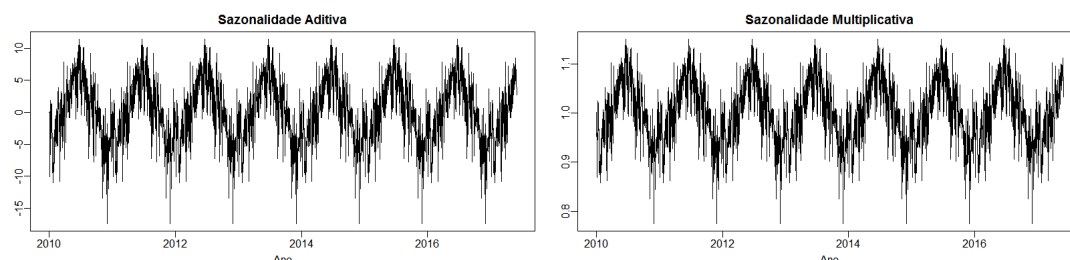
Modelo 1 - Modelo de Suavização Exponencial (Aditivo): as estimativas para os parâmetros do modelo são: $\hat{\alpha} = 0,1448$, $\hat{\beta} = 0$ e $\hat{\gamma} = 0,3598$. Vale ressaltar que no modelo aditivo, a amplitude da variação sazonal da série é constante ao longo do tempo. Para este modelo, obtivemos a soma do quadrado dos erros igual à $SSE=250480,6$.

Modelo 2 - Modelo de Suavização Exponencial (Multiplicativo): as estimativas para os parâmetros do modelo são: $\hat{\alpha} = 0,0767$, $\hat{\beta} = 0$ e $\hat{\gamma} = 0,3328$. Vale ressaltar que no modelo multiplicativo, a amplitude da variação sazonal aumenta com o acréscimo no nível médio da série. Para este modelo, obtivemos a soma do quadrado dos erros igual à $SSE=253574,3$.

A Figura 3 apresenta as sazonalidades (aditiva e multiplicativa) da Série Temporal da Umidade Relativa do Ar Diária de Porto Alegre, de 01/01/2010 a 29/05/2017. Ambas as sazonalidade foram calculadas utilizando a rotina *Holt-Winters*, do R. Analisando os resíduos de ambos os modelos, obtemos que a soma dos quadrados das autocorrelações amostrais, para 100 lags, é 1,2159, para o modelo de *Holt-Winters* para efeitos sazonais aditivos, e 1,2363, para o modelo com efeitos sazonais multiplicativos. Desta forma percebemos uma pequena vantagem em utilizarmos o modelo aditivo. Mesmo assim, neste trabalho, optamos em analisar a predição utilizando os modelos de *Holt-Winters* com efeitos sazonais aditivos e multiplicativos. Como os modelos $ARFIMA(p,d,q)$ não levam em consideração a sazonalidade da série, iremos retirar a sazonalidade utilizando

a rotina *Decompose*, considerando efeitos sazonais aditivos e multiplicativo. Ver Modelos 3 e 4, na sequência.

Figura 3 – Sazonalidades da Série Temporal da Umidade Relativa do Ar Diária de Porto Alegre, de 01/01/2010 a 29/05/2017.



Fonte: Os Autores.

Modelo 3 - Modelo ARFIMA (p,d,q) , depois de retirada a sazonalidade utilizando a rotina *Decompose* (Aditivo). Foi ajustado, como mostra a Tabela 1, o modelo ARFIMA $(2,d,0)$, com $\hat{d} = 0,1038$, $\hat{\phi}_1 = 0,3439$ e $\hat{\phi}_2 = -0,0383$. Para este modelo, obtivemos uma variância estimada igual a $\hat{\sigma}^2 = 59,28$, log-verossimilhança: -9363 e Critério de Informação de Akaike AIC=18733,76. Para a análise de resíduos utilizamos o teste de Box-Pierce (H_0 : Os resíduos não são correlacionados (ruído branco)). O p-valor encontrado foi de 0,9711, não há evidências para rejeitar H_0 , ou seja, os resíduos do modelo são não correlacionados, sendo assim, o modelo proposto é adequado.

Tabela 1 - Modelo ARFIMA(2,d,0) ajustado, após retirada a sazonalidade utilizando a rotina *Decompose* (Aditivo).

Parâmetros	Estimativa	Erro Padrão	Estat. z	P(> z)
d	0,1038	0,02452	4,233	2,31e-05
ϕ_1	0,3439	0,03100	11,094	<2e-16
ϕ_2	-0,03825	0,01997	-1,915	<0,05

Fonte: Os Autores.

Modelo 4 - Modelo ARFIMA (p,d,q) , após retirada a sazonalidade utilizando a rotina *Decompose* (Multiplicativo). Foi ajustado, como demonstra a Tabela 2, o modelo ARFIMA $(0,d,1)$, com $\hat{d} = 0,2509$ e $\hat{\theta}_1 = -0,2135$. Para este modelo, obtivemos uma variância estimada igual a $\hat{\sigma}^2 = 67,66$, log-verossimilhança: -9542 e Critério de Informação de Akaike AIC=19089,78. Para a análise de resíduos utilizamos o teste de Box-Pierce, cujo p-valor encontrado foi de 0,8949, ou seja, não há evidências para rejeitar

H_0 , indicando que os resíduos do modelo são não correlacionados e que o modelo proposto é adequado.

Tabela 2 - Modelo ARFIMA(0,d,1) ajustado, após retirada a sazonalidade utilizando a rotina *Decompose* (Multiplicativo).

Parâmetros	Estimativa	Erro Padrão	Estat. z	P(> z)
d	0,2509	0,0183	13,664	<2e-16
θ_1	-0,2135	0,0243	-8,789	<2e-16

Fonte: Os Autores.

Analisando as Tabelas 1 e 2, percebemos que o valor estimado do parâmetro d , dos processos ARFIMA(p,d,q), para os Modelos 3 e 4, pertence ao intervalo (0; 0,5), além disso é significativo. Ou seja, a série temporal da Umidade Relativa do Ar Diária de Porto Alegre, de 01/01/2010 a 29/05/2017 apresenta longa dependência, o que corrobora nossa análise da função de autocorrelação amostral, apresentada na Figura 2.

Modelo 5 - Modelo SARIMA (p,d,q)(P,D,Q) s . Foi ajustado, como apresentado na Tabela 3, o modelo Modelo SARIMA ($1,0,0$) \times ($3,1,0$) $_{12}$, com $\hat{\phi}_1 = 0,4525$, $\hat{\Phi}_1 = -0,7483$, $\hat{\Phi}_2 = -0,5389$ e $\hat{\Phi}_3 = -0,2391$. Para este modelo, obtivemos uma variância estimada igual a $\hat{\sigma}^2 = 82,86$, log-verossimilhança: -9777,12 e Critério de Informação de Akaike AIC=19564,25. Para a análise de resíduos utilizamos o teste de Box-Pierce, cujo p-valor encontrado foi de 0,8421. Assim, ambos indicam que os resíduos do modelo são não correlacionados.

Tabela 3 - Modelo SARIMA ($1,0,0$) \times ($3,1,0$) $_{12}$ ajustado.

Parâmetros	Estimativa	Erro Padrão	Estat. z	P(> z)
ϕ_1	0,4525	0,017213	26,286	<2,2e-16
Φ_1	-0,748330	0,018773	-39,862	<2,2e-16
Φ_2	-0,538926	0,021320	-25,278	<2,2e-16
Φ_3	-0,239126	0,018843	-12,691	<2,2e-16

Fonte: Os Autores.

Para todos os modelos, também analisamos a função de autocorrelação amostral dos resíduos dos modelos. No caso dos modelos 3, 4 e 5, a função de autocorrelação amostral dos resíduos corrobora o resultado do teste de Box-Pierce. Os Modelos 1 e 2 (Suavização Exponencial Aditivo e Multiplicativo, respectivamente), conforme o teste de resíduos de Box-Pierce, apresentam resíduos não significativos (p-valor < 0,05). No

entanto, pela análise da função de autocorrelação amostral, os resíduos apresentam um comportamento quase errático. Por este motivo optamos por incluí-los nos passos seguintes da análise.

Após estabelecer os modelos e estimar seus parâmetros, passamos a obtenção dos pesos w pelo método de variância mínima e estimar os coeficientes da combinação pelo método de regressão. Na Tabela 4 encontram-se os referidos pesos para todas as combinações de dois modelos individuais.

Tabela 4 - Modelos de combinação de previsão: \hat{w} é o peso atribuído à previsão de menor variância, no método de variância mínima; $\hat{\beta}_1$ e $\hat{\beta}_2$ são os coeficientes pelo método de regressão, utilizando o método dos Mínimos Quadrados Ordinários.

Combinação de Previsão	\hat{w}	$\hat{\beta}_1$	$\hat{\beta}_2$
Modelos 1 e 2	0,5032	0,6312	0,3614
Modelos 1 e 3	0,9221	-0,15107	1,15027
Modelos 1 e 4	0,8059	0,06349	0,93682
Modelos 1 e 5	0,6635	0,29277	0,7063
Modelos 2 e 3	0,9140	-0,14941	1,14924
Modelos 2 e 4	0,7938	0,06995	0,93007
Modelos 2 e 5	0,6613	0,2853	0,71253
Modelos 3 e 4	0,7148	0,94936	0,050231
Modelos 3 e 5	0,7729	0,96068	0,038907
Modelos 4 e 5	0,7537	0,93783	0,06245

Fonte: Os Autores.

De posse dos parâmetros dos modelos individuais e os pesos das combinações foram calculadas todas as predições (que consistem nas estimativas obtidas um passo à frente pelo método utilizado, dentro do período observado), que possibilitou a avaliação da acurácia, feita por meio do RMSE, MAPE e U de Theil. A Tabelas 5 apresenta as medidas de acurácia utilizando os Modelos 1 a 5 ajustados, de onde observamos que o Modelo 3 apresentou as menores medidas de acurácia.

Tabela 5 - Medidas de acurácia dos Modelos 1 a 5.

Modelos Ajustados	RMSE	MAPE	U de Theil
Modelo 1	10,344	10,406	0,1337
Modelo 2	10,407	10,573	0,1339
Modelo 3	7,6933	8,0132	0,0994
Modelo 4	8,2299	8,5979	0,1064
Modelo 5	9,0825	9,3887	0,1339

Fonte: Os autores.

Na Tabela 6 são apresentadas as combinações de previsões, sendo que estas são resultados da combinação de dois modelos de previsões individuais e para cada método de combinação encontram-se em negrito as melhores combinações. A combinação de previsões por regressão dos Modelos 2 e 3 apresentam os menores RMSE, MAPE e U de Theil, sendo considerada assim o melhor método para realizar as previsões. Com medidas de acurácia muito próximas tem-se a combinação de previsões por regressão dos Modelos 1 e 3 (realçadas em cinza na referida tabela), que mesmo sendo dois modelos aditivos, a forma como o modelo de suavização exponencial capta informações da série, deve levar a um aumento da acurácia. Para os outros dois métodos de combinação, por variância mínima e média aritmética, as menores medidas de acurácia em estudo foram encontradas para os Modelos 3 e 5 e, para os Modelos 3 e 4, respectivamente. Destacamos que as melhores medidas de acurácia obtidas nestas combinações - menores RMSE, MAPE e U de Theil, todas envolvem o Modelo 3, que foi o modelo individual com as melhores medidas de acurácia (Tabela 5). Isso deve-se ao fato que o Modelo 3 leva em consideração o efeito sazonal aditivo e a característica de longa dependência da série temporal, ambas as características encontradas na série temporal, conforme descrição realizada nas Figuras 2 e 3.

Tabela 6 - Medidas de acurácia para as Combinações de Previsão.

Combinação de Previsão	Média Aritmética			Regressão Linear			Variância Mínima		
Medidas de Acurácia	RMSE	MAPE	U de Theil	RMSE	MAPE	U de Theil	RMSE	MAPE	U de Theil
Modelos 1 e 2	10,3338	10,446	0,1331	10,3124	10,4001	0,1328	10,334	10,4466	0,133
Modelos 1 e 3	8,5489	8,8463	0,1101	7,5587	7,8681	0,0974	7,6882	8,0232	0,1417
Modelos 1 e 4	8,7157	9,0335	0,1123	8,2066	8,5819	0,1057	8,2535	8,6004	0,1299
Modelos 1 e 5	8,9221	9,2186	0,1149	8,7768	9,091	0,1131	8,7835	9,105	0,1204
Modelos 2 e 3	8,5715	8,9194	0,1104	7,5584	7,8503	0,0974	7,6991	8,0464	0,1426
Modelos 2 e 4	8,7217	9,0968	0,1124	8,2036	8,5823	0,1057	8,2571	8,6362	0,1303
Modelos 2 e 5	8,9474	9,2989	0,1153	8,7874	9,1197	0,1132	8,799	9,1571	0,1212
Modelos 3 e 4	7,7981	8,1177	0,0999	7,6868	8,0073	0,0981	7,915	8,2594	0,0981
Modelos 3 e 5	8,0191	8,2873	0,1027	7,6883	7,9981	0,0981	7,6734	7,9681	0,0981
Modelos 4 e 5	8,4186	8,7145	0,1083	8,2255	8,5863	0,1058	8,2481	8,5918	0,1058

Fonte: Os Autores.

Complementando a análise foram calculados os coeficientes de determinação para as combinações pelo método de regressão. Tais coeficientes são apresentados na Tabela 7, de onde podemos verificar que a combinação considerada como a melhor combinação (Modelos 2 e 3) possui maior R^2 ajustado, junto com a combinação entre os Modelos 1 e 3, sendo que essas combinações explicam 99,05% da variabilidade dos dados.

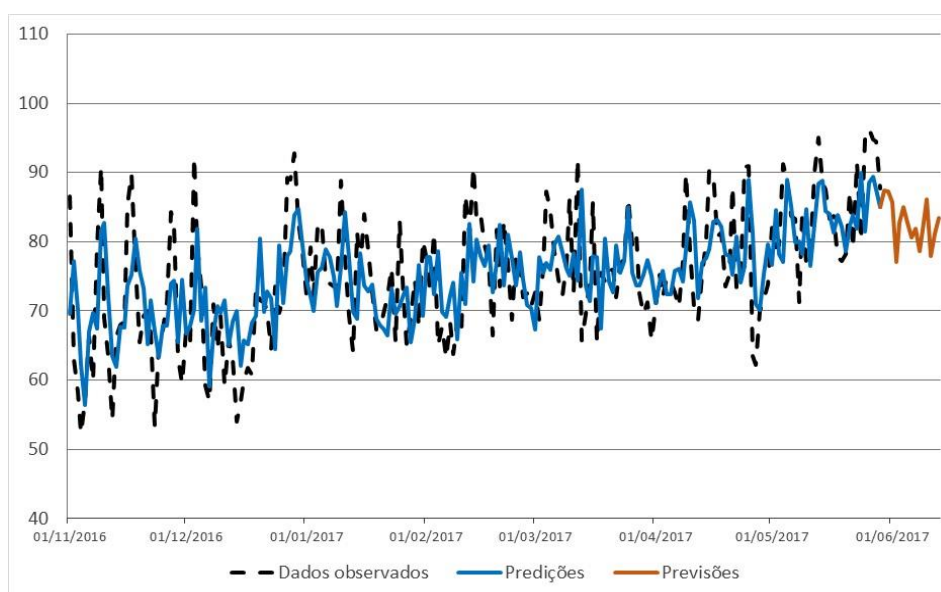
Tabela 7 - Combinação de previsão pelo método de regressão e os valores de R^2 ajustado.

Combinação de Previsão dos Modelos	1 e 2	1 e 3	1 e 4	1 e 5	2 e 3	2 e 4	2 e 5	3 e 4	3 e 5	4 e 5
R^2 ajustado	0,9823	0,9905	0,9888	0,9872	0,9905	0,9888	0,9872	0,9904	0,9904	0,9888

Fonte: Os Autores.

Apesar do estudo ter encontrado duas boas combinações (via regressão), para fins de obtenção das previsões, consideramos aquela que apresenta as menores medidas de acurácia para a realizar as previsões, a saber a combinação por regressão para os Modelos 2 e 3. A Figura 4 apresenta as predições (dentro da amostra) e as previsões (fora da amostra), obtidas com base no último período observado, da série temporal da umidade relativa do ar de Porto Alegre utilizando a combinação de previsões por regressão dos Modelos 2 e 3. Podemos observar ainda que esta combinação capta bem o comportamento dos dados da série.

Figura 4 – Predições e Previsões da Série Temporal da Umidade Relativa do Ar Diária de Porto Alegre, Brasil, para o período de 01/11/2016 a 13/06/2017.



Fonte: Os Autores.

5. Considerações Finais

Ao concluirmos nossas análises, com relação aos resultados dos modelos individualmente, os modelos de suavização exponencial foram os que obtiveram maiores medidas de acurácia. No entanto, ao combinarmos usando o método de regressão com o modelo no qual foi ajustado um processo ARFIMA (p,d,q) após a retirada da sazonalidade utilizando a rotina *Decompose* (aditivo), as previsões se tornam as melhores dentro todas as realizadas no estudo. Desta forma, é possível que mesmo não apresentando bons resultados na acurácia, os modelos de suavização exponencial captam algo do modelo que melhora o ajuste quando combinado com o ARFIMA (p,d,q) .

Verificamos também, que embora todas as medidas de acurácia apresentem valores próximos, a combinação de previsões por regressão entre os Modelos 2 (Modelo de Suavização Exponencial Multiplicativo) e 3 (Modelo ARFIMA (p,d,q)) obteve as menores medidas de acurácia para as três medidas usadas no estudo e um R^2 ajustado que explica 99,05% da variabilidade dos dados, sendo por estas as razões utilizada para obter as previsões. Contudo cabe salientar, que a combinação de previsões por regressão entre os Modelos 1 (Modelo de Suavização Exponencial Aditivo) e 3 (Modelo ARFIMA (p,d,q)) também obteve medidas de acurácia baixas e um R^2 ajustado no mesmo valor de 99,05%. Embora os modelos de Suavização Exponencial (Modelos 1 e 2) não estejam bem especificados conforme o teste de *Box-Pierce*, estes modelos estão presentes nas combinações que obtiveram a melhor capacidade preditiva dentre todos os modelos.

Outro aspecto a ser ressaltado, consiste de que o Modelo 3, que apresentou as melhores medidas de acurácia dentre os cinco modelos individuais estudados, e se encontra em todas as combinações que apresentam as melhores medidas de acurácia em cada um dos três métodos de combinação avaliados.

Este estudo contemplou a análise da contribuição das combinações de previsões e como elas exerceram um papel otimizador dos resultados de acurácia, sendo uma boa alternativa para obter valores próximos dos reais na série de umidade relativa do ar de Porto Alegre.

Uma vez que o presente artigo proporcionou encontrar resultados para prever a umidade relativa do ar em Porto Alegre – RS com base na análise de modelos de análise de séries temporais e suas combinações. A utilização de outros tipos de modelos específicos para analisar o conjunto de dados meteorológicos, como por exemplo o

modelo ETA (INPTE/CPTEC), não foram contemplados ficando como sugestão para outros trabalhos a serem realizados futuramente.

Além deste aspecto, podem ser ainda realizadas combinações de previsões utilizando como base três modelos, que não foi realizado neste trabalho por não ser o objetivo. O estudo da magnitude na acurácia comparando diversas combinações é uma oportunidade de pesquisa no qual os autores têm atuado, uma vez que a implementação e avaliação de combinações com três ou mais modelos é facilmente implementável e certamente trará melhorias de acurácia.

Referências

- BATES, J. M.; GRANGER, C. W. J. The Combining of Forecasts. **Operational Research Quarterly**, v.20, n.4, 1969, p. 451-468.
- BATISTA, A. C. **Mapas de risco**: uma alternativa para o planejamento de controle de incêndios florestais. *Floresta*, v.30, n.1/2, p.45-54, 2000.
- BLIEMEL, F. Theil's Forecast Accuracy Coefficient: A Clarification. **Journal of Marketing Research**, v. 10, P.444 – 446, November, 1973.
- BOX, G. E.; JENKINS, G. M. **Time series analysis, control, and forecasting**. San Francisco, CA: Holden Day, 1976.
- CARVALHO, R. L. S.; NASCIMENTO, B. I. S.; QUERINO, C. A. S.; SILVA, M. J. G.; DELGADO, A. R. S. Comportamento das séries temporais de temperatura do ar, umidade e precipitação pluviométrica no município de Ariquemes (Rondônia-Brasil). **Revista Brasileira de Climatologia**, ano 12 v.18, p.123-1442, jan-jun 2016.
- CEPAGRI. **Escala psicrométrica unicamp para indicação de níveis de umidade relativa do ar prejudiciais à saúde humana**. Disponível em: <http://www.cpa.unicamp.br/artigos-especiais/umidade-do-ar-saude-no-inverno.html>. Acessado em: 12 de setembro de 2017.
- CHAI, T.; DRAXLER, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, v. 7, n.3, p. 1247–1250, 2014.
- COSTANTINI, M.; PAPPALARDO, C. A hierarchical procedure for the combination of forecasts. **International Journal of Forecasting**, v. 26 n.4, p.725–743, 2010.
- GARDNER, G.; HARVEY, A. C.; PHILLIPS, G. D. Algorithm as 154: An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v. 9, n.3, p.311–322, 1980.
- GEWEKE, J.; PORTER-HUDAK, S. The estimation and application of long memory time series models. **Journal of Time Series Analysis**, v.4, n.4, p.221–238, 1983.
- GIL, A. C. **Como elaborar projetos de pesquisa**. Ed. Atlas, 2017.
- GRANGER, C. W.; JOYEUX, R. An introduction to long-memory time series models and fractional differencing. **Journal of Time Series Analysis**, v.1, n.1, p.15–29, 1980.

GRANGER, C. W.; RAMANATHAN, R. Improved methods of combining forecasts. **Journal of Forecasting**, v.3, n.2, p.97–204, 1984.

HOSKING, J. R. Fractional differencing. **Biometrika**, v.68, n.1, p.165–176, 1981.

HOSKING, J. R. Modeling persistence in hydrological time series using fractional differencing. **Water resources research**, v.20, n.12, p.1898–1908, 1984.

KIM, S.; KIM, H. A new metric of absolute percentage error for intermittent demand forecasts **International Journal of Forecasting**, v. 32, n.3, p. 669-679, July–September 2016.

MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: methods and applications**. John Wiley & Sons, 3rd edition, 1998.

MENEZES, L. M.; BUNN, D. W.; Taylor, J. W. Review of guidelines for the use of combined forecasts. **European Journal of Operational Research**, v.120, n.1, p.190–204, 2000.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. Ed. Blücher. 2006.

PELLEGRINI, F. R. **Metodologia para implementação de sistemas de previsão de demanda**. Dissertação (Mestrado em Engenharia de Produção). Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal do Rio Grande do Sul. Porto Alegre, 2000.

SHCHERBAKOV, M.V.; BREBELS, A.; SHCHERBAKOVA, N. L.; TYUKOV, A. P.; JANOVSKY, T. A.; KAMAEV, V. A. A Survey of Forecast Error Measures. **World Applied Sciences Journal**, v.24, n.24 (Information Technologies in Modern Industry, Education & Society), p.171-176, 2013.

VIRGENS F^o, J. S.; LEITE, M. L.; FRANCO, J. R.; KORELO, M. Modelo computacional estocástico para simulação de séries climáticas diárias de umidade relativa do ar, baseado na parametrização dinâmica das distribuições de probabilidade decorrente da retroalimentação de dados. **Revista Brasileira de Climatologia**. v. 5, p. 133 -151, setembro 2009.

FORECAST OF THE DAILY RELATIVE HUMIDITY OF THE AIR OF PORTO ALEGRE

Abstract

Relative humidity interferes with many aspects of human life, and due to the many consequences that a low or high enough percentage of it can bring, the control of its level is of the utmost importance. Thus, making accurate forecasts of the relative humidity of the air, with the objective of subsidizing the necessary decisions in various segments of society, becomes fundamental. Given this context, this article aims to compare various models and methods of forecasting, seeking the most accurate one in order to forecast the daily relative humidity of the air of Porto Alegre, Rio Grande do Sul state, Brazil. To this intent five models were obtained (exponential smoothing, SARIMA, and ARFIMA), in addition to two-by-two combinations of each of the five models among three combination methods: minimum variance, mean, and regression. Through the RSME and MAPE accuracy measures we verified that the combination of forecasts by regression among Model 2 (Multiplicative Exponential Smoothing Model), with Model 3 (ARFIMA Model (2,d,0)) obtaining the least accuracy measures in all cases, being then used to create the desired forecasts.

Key-words: *Relative Humidity, Combination of Forecasts, Modeling*