

QUALIDADE DO AR

UTILIZAÇÃO DE SÉRIES TEMPORAIS PARA PREVISÃO DO MATERIAL PARTICULADO (MP₁₀) NA CIDADE DE PAULÍNIA (SP)

Nathália Morgana Rissi– rissi.nathalia@gmail.com
Universidade Estadual de Campinas (Unicamp).

Marla Rubia Garcia de Lima – marla_rubia@hotmail.com
Universidade Estadual de Campinas (Unicamp)

Danilo Nogarotto– nogarotto.danilo@gmail.com
Universidade Estadual de Campinas (Unicamp)

Simone Andréa Pozza – simone.pozza@gmail.com
Universidade Estadual de Campinas (Unicamp)

Resumo: A poluição atmosférica traz prejuízos à saúde humana, tornando-se de vital importância seu estudo e o desenvolvimento de técnicas que visam seu monitoramento e controle eficazes. O estudo na ênfase em material particulado inalável (MP₁₀) é necessário dado o dano causado à saúde da população, além de outros. Foram obtidas concentrações diárias de MP₁₀ no Sistema QUALAR da CETESB, referente ao município de Paulínia (SP). Tais dados foram tratados com o intuito de prever concentrações futuras de MP₁₀. Como método estatístico para manipulação destes dados optou-se por modelos de séries temporais utilizando-se de modelos ARIMA e SARIMA. Os resultados mais satisfatórios foram obtidos pelo modelo SARIMA, pois apresentou os melhores ajustes em relação aos parâmetros estabelecidos.

Palavras-chave: Séries temporais, MP₁₀, Poluição atmosférica.

1. INTRODUÇÃO E OBJETIVOS

A poluição do ar é caracterizada pela presença de uma ou mais substâncias químicas com concentrações suficientes para causarem danos aos seres humanos, animais, vegetais, ou em materiais. Para garantir a qualidade de vida da população, o estudo e o monitoramento de poluentes atmosféricos têm sido cada vez mais necessários. Um dos poluentes de estudo é o MP₁₀, no qual são partículas inaláveis que possuem como característica diâmetro menor ou igual a 10 µm, podendo ficar retidas na parte superior do sistema respiratório ou até alcançar os alvéolos pulmonares (CETESB, 2015).

Paulínia foi selecionada como objeto de estudo pelo contexto ambiental que contempla a influência da grande região urbanizada de Campinas e entorno, bem como regiões altamente

industrializadas como as de Americana e Limeira, além da concentração importantes empreendimentos nos setores químicos, petroquímicos, de industrialização e distribuição de petróleo, fertilizantes, agroquímicos e fibras sintéticas na própria cidade (CLEMENTE, 2000; AMORIM, 2004).

A presença de dados faltantes é um problema muito frequente em muitos campos científicos, agravando-se em pesquisa ambiental (XIA *et al.*, 1999). O método sugerido por Plaia e Bondi (2006) denominado como *Site-Dependent Effect Method* (SDEM), foi aplicado em seus estudos com o objetivo de propor um novo meio de preenchimento de dados faltantes. De modo geral, o método SDEM considera um efeito principal (cujas características são similares entre si) e efeitos secundários em relação a esse efeito principal. Em seu trabalho foram consideradas oito estações de monitoramento como o efeito principal (fator de espaço), pois suas características são mais próximas uma das outras, e os efeitos secundários são a semana, dia e hora, que são considerados em relação às estações (efeitos de tempo, em relação ao espaço).

Para a análise de dados de MP10, contamos com ferramentas matemáticas e estatísticas, que auxiliam nas análises e estudo de seu comportamento. Atualmente, tem sido utilizada por diversos autores a análise de séries temporais no estudo do comportamento das variáveis poluentes, especificamente o MP, ao longo do tempo. Ehlers (2009) descreveu que uma série temporal é a coleção de observações feitas sequencialmente ao longo do tempo. Modelos de Séries Temporais são técnicas que visam explicar a influência do tempo nos dados, tanto nas observações do passado, quanto nas do futuro. Eles auxiliam na compreensão e na previsão dos dados temporais (MORETTIN e TOLOI, 2006).

Bell, Samat e Dominici (2004) utilizaram as séries temporais para estimar os riscos à saúde associados com a exposição a material particulado (MP) em curto prazo, colocando enfoque analítico no contexto mais amplo dos estudos epidemiológicos da poluição do ar em geral.

Lima *et al.* (2009) também aplicaram séries temporais como ferramenta para prever as concentrações de MP10, encontrando resultados positivos para o uso da técnica. Neste caso, as metodologias de ajuste foram via modelo ARIMA sazonais e método de suavização exponencial de Holt-Winters.

Este trabalho tem como objetivo de utilizar séries temporais e para fazer previsões de valores futuros da concentração de MP10 da cidade de Paulínia, visando encontrar padrões de tendência, bem como a existência de variação sazonal ou cíclica e alterações estruturais.

2. METODOLOGIA

2.1 Dados

Os dados da concentração de MP10 da cidade de Paulínia (SP) foram obtidos pela base de dados do QUALAR (Qualidade do Ar em Tempo Real), que consiste num banco de informações sobre a qualidade do ar (ANDRADE, 2011). Foram utilizadas as médias diárias da concentração de MP10 de um período de cinco anos (de 2010 a 2014), e os dados do mês de janeiro de 2015 foram utilizados para a comparação da previsão obtida pelo modelo.

2.2 Tratamentos dos dados

Como no banco de dados disponibilizado pelo QUALAR possuía 1,48% de dados faltantes, houve a necessidade de realizar um tratamento de preenchimento de dados para realizar a análise da série.

Para o preenchimento dos dados faltantes, o método SDEM desenvolvido por Plaia e Bondi (2006) foi adaptado a fim de se adequar as estruturas dos dados da cidade de Paulínia. Dessa maneira consideramos o mês como efeito principal, enquanto ano, semana e dia da semana foram os efeitos secundários.

2.3 Séries temporais

A ferramenta estatística utilizada para fazer o tratamento dos dados são os modelos de séries temporais. Estas séries são coleções de dados feitos sequencialmente ao longo do tempo. Sua característica mais importante é que as observações vizinhas são dependentes, e os modelos têm como seu principal objetivo analisar essas dependências. Neste caso o objetivo de utilizar esse método é para fazer previsões de valores futuros da concentração de MP10, visando assim encontrar padrões de tendência, bem como a existência de variação sazonal.

Foi utilizada a abordagem de Box Jenkins, que consiste em ajustar modelos autor-regressivos integrados de médias móveis, ARIMA(p,d,q). Além disso, para estudar a componente sazonal, foram utilizados os modelos ARIMA sazonal multiplicativa (SARIMA) de ordem (p,d,q) x (P,D,Q)_S (MORETTIN e TOLOI, 2006).

Para fazer essa análise dos dados foi utilizada a plataforma computacional R (R CORE TEAM, 2015), de acesso livre, para análises estatísticas. Esta plataforma oferece um excelente ambiente para análises estatísticas com recursos gráficos com alta qualidade (RIBEIRO JUNIOR, 2011).

3. RESULTADOS E DISCUSSÕES

Foram identificados diversos modelos ARIMA e SARIMA a fim de comparar qual obteve melhores resultados. Para isso utilizamos os valores de AIC (Critério de Informação Akaike), LOG (Log-verossimilhança) e EQM (erro quadrático médio de previsão). Este último foi calculado baseado na previsão dos próximos 30 dias (janeiro de 2015). As medidas do AIC e EQM são melhores quando menores. Já o LOG, quanto maior, melhor.

Para identificar os modelos foram analisados as funções de auto correlação (fac) e as funções de auto correlação parcial (facp) (MORETTIN e TOLOI, 2006). Nos modelos SARIMA considerou-se a sazonalidade com um período de setedias, isto supondo que, concentrações de MP10 podem ter comportamentos similares no mesmo dia da semana.

De acordo com a Tabela 1, o ajuste que se destacou entre os demais foi o Ajuste9, pois apresentou os melhores AIC e LOG. Em geral, os modelos SARIMA foram mais bem ajustados que os modelos ARIMA, analisando os parâmetros. Somente no EQM foi melhor no modelo ARIMA Ajuste6.

Tabela 1: Resultados dos modelos

Ajuste	Modelo	p	d	q	P	D	Q	S	AIC	LOG	EQM
Ajuste1	ARIMA	1	0	2	0	0	0	0	13.990,52	-6.990,26	8,8088
Ajuste2	ARIMA	1	0	1	0	0	0	0	14.028,10	-7.010,05	9,7378
Ajuste3	SARIMA	1	0	2	1	0	2	7	13.920,60	-6.952,30	8,0794
Ajuste4	SARIMA	1	0	1	1	0	1	7	13.928,42	-6.958,21	8,1772
Ajuste5	SARIMA	2	0	2	2	0	2	7	13.897,76	-6.938,88	7,7707
Ajuste6	ARIMA	2	0	2	0	0	0	0	13.956,49	-6.972,25	7,5784
Ajuste7	SARIMA	1	1	2	1	0	2	7	13.896,84	-6.941,42	8,6326
Ajuste8	SARIMA	1	1	1	1	0	1	7	13.894,85	-6.942,43	8,7352
Ajuste9	SARIMA	1	1	1	1	1	1	7	13.863,39	-6.926,69	8,5345

O gráfico (Figura 1) representa a série obtida pelo Ajuste9 em vermelho, pela série temporal utilizando o modelo SARIMA, e em preto os dados obtidos pela estação de monitoramento localizada na cidade.

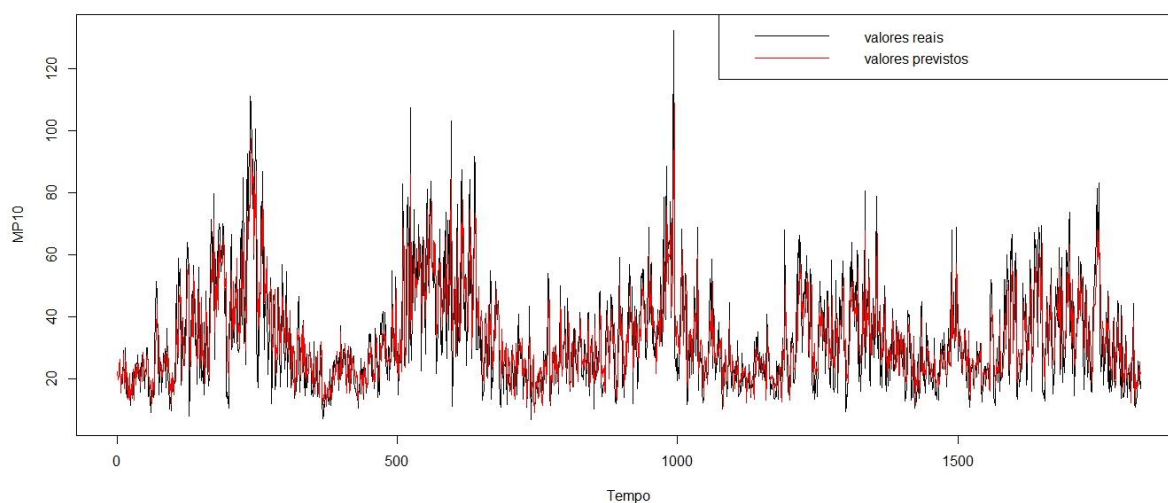


Figura 1: Série ajustada no período de 2010 a 2014

O gráfico apresentado na Figura 2 representa a série obtida pelo Ajuste9, nos seis primeiros meses de janeiro de 2010, assim observa-se que a previsão está se baseando no dia anterior, com isso a série está ficando defasada em um dia em sua previsão.

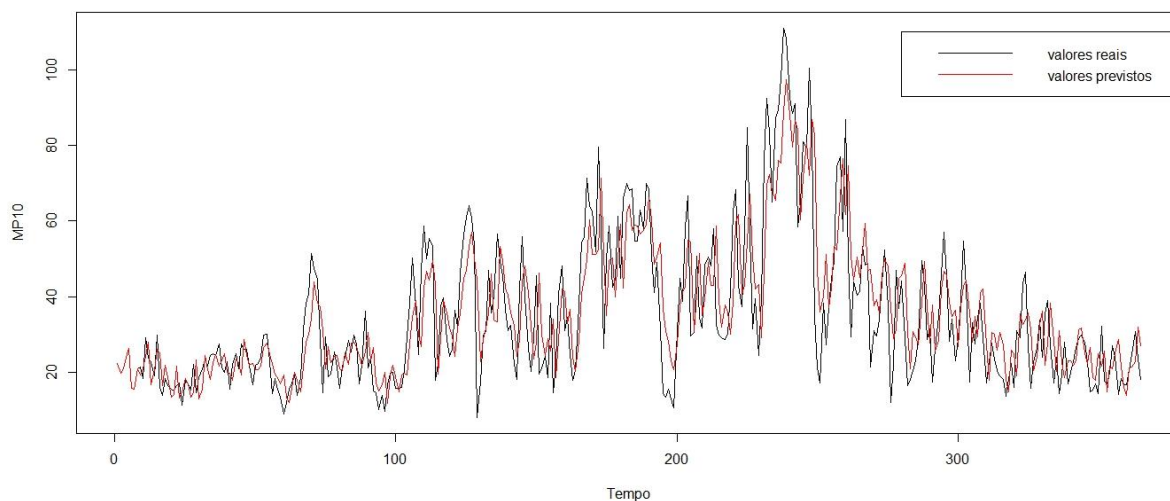


Figura 2: Série ajustada nos primeiros seis meses de 2010

A Figura 3 representa a previsão realizada foi do mês de Janeiro de 2015 e observa-se que a série obtida através do modelo SARIMA não obteve mesmo comportamento que os dados da estação de monitoramento, tendo uma previsão das concentrações com valores mais baixos que os reais, sobressaindo apenas no início do mês.

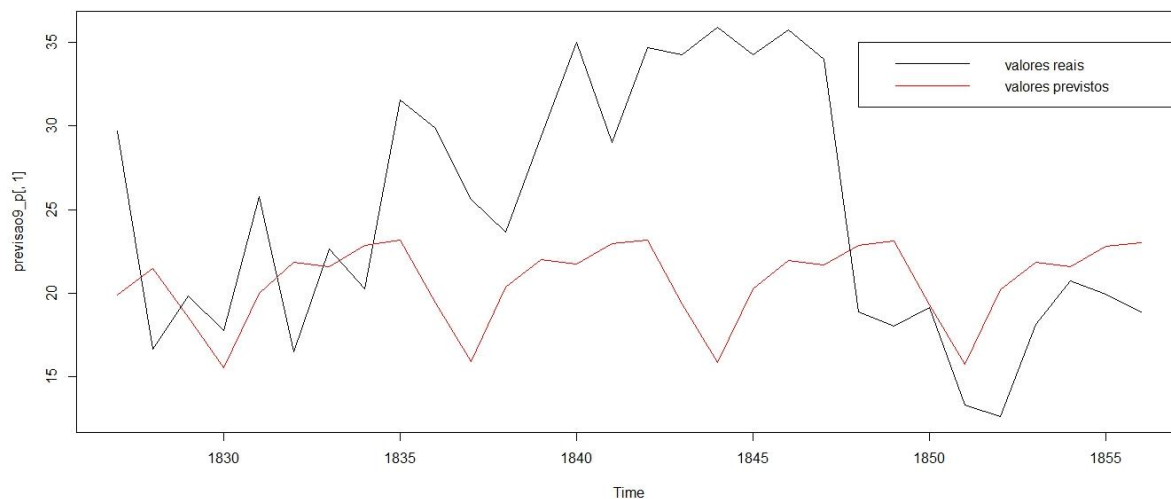


Figura 3: Previsão do melhor modelo Ajuste9.

4. CONSIDERAÇÕES FINAIS

O modelo que apresentou melhor resultado foi SARIMA de ordem (1,1,1) x (1,1,1) em relação aos parâmetros AIC e LOG. Em geral os modelos SARIMA apresentaram resultados mais satisfatórios quando comparados aos ARIMAS.

Há evidências que a série temporal do MP10 tem um comportamento de memória longa, por isso a defasagem no ajuste, e a previsão não ser tão precisa. Como sugestão de

trabalho futuro talvez o melhor modelo que se encaixe nessa situação seja os modelos ARFIMA (Modelo Auto Regressivo Fracionário Integrado de Médias Móveis).

Agradecimentos

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), pela concessão da bolsa de estudos.

REFERÊNCIAS

ANDRADE, J. M. F. **A atmosfera, o Ar e o QUALAR**. 2011. Disponível em: <<http://www.cetesb.sp.gov.br/ponto-de-vista/52,a-atmosfera--o-ar-e-o-qualar>> Acesso em 10 Abril 2015.

AMORIM, W. B. **Monitoramento da concentração e caracterização do material particulado suspenso na atmosfera**. 2004, Tese (Doutorado) – Faculdade de Engenharia Química da Unicamp, Campinas, 2004

BELL, M. L.; SAMAT, J. M.; DOMINICE, F. 2004. **Time-Series studies of particulate matter**. Annu. Rev. Public Health 25 pp.247–80.

CETESB (Companhia Ambiental do Estado de São Paulo) 2015. **Qualidade do ar - Poluentes**. São Paulo. Disponível em: <<http://www.cetesb.sp.gov.br/ar/Informa??es-B?sicas/21-Poluentess>>. Acesso em 15 de abril de 2015.

CLEMENTE, D. A. **Estudo do Impacto Ambiental das Fontes Industriais de Poluição do Ar no Município de Paulínia- S.P: empregando o Modelo ISCST3**. Dissertação (Mestrado) - Curso de Engenharia Química, Unicamp, Campinas, 2000, 179p.

EHLERS, R. S. 2009. **Análises de séries temporais**. 5^a ed., 114p.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. Saraiva, São Paulo, 7. ed. 2002, 526p.

LIMA, E. P.; POZZA, S. A.; GIMENES, M. L.; COURRY, J. R. 2009. **Uso de modelos arima sazonais no estudo da série temporal de MP10 da cidade de São Carlos**. Anais do XXXIV ENEMP – Congresso Brasileiro de Sistemas Particulados. Campinas-SP.

PLAIA, A.; BONDI A. L. 2006. **Single imputation method of missing values in environmental pollution data sets**. Atmospheric Environment 40, 2006; 7316-7330.

QUALAR CETESB: **banco de dados**. Disponível em: <http://ar.cetesb.sp.gov.br/qualar/>. Acesso em: 15 abril, 2016.



R Core Team (2015). **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. Acesso em 8 de abril de 2016.

RIBEIRO JUNIOR, P. J. 2011. **Introdução ao Ambiente Estatístico R.** Disponível em: <<http://www.leg.ufpr.br/~paulojus/embrapa/Rembrapa/>>. Acesso em 16 Abril de 2015.

XIA, Y.; FABIAN, P.; STOHL, A.; WINTERHALTER, M.. 1999. **Forest climatology: estimation of missing values for Bavaria.** Agricultural and Forest Meteorology 96, 131-144.