NO_2 TOTAL COLUMN FORECASTING, THROUGH ARIMA STATISTICAL MODEL FOR THE METROPOLITAN REGION OF SÃO PAULO

Bruno R.G. JATOBÁ, Rodrigo L. da ROCHA JUNIOR, Paulo A.D.P. CALADO, Diogo R.S. MONTEIRO[2], Daniel M.C. LIMA, Glauber L. MARIANO
bruno.jatoba@icat.ufal.br

Instituto de Ciências Atmosféricas (ICAT/UFAL) – Brazil

ABSTRACT

An ARIMA (Auto Regressive Integration Moving Average) model was adjusted to predict the time series of the Total Nitrogen Dioxide Column in the Metropolitan Region of São Paulo. At the end of this study, we observed that the actual and predicted values had an excellent correlation. We conclude that the ARIMA model is an efficient statistical method, which allows the application to issue alerts with a certain degree of confidence.

RESUMEN

Se ajustó un modelo ARIMA (Auto Regressive Integration Moving Average) para predecir las series de tiempo de la Columna Total de Dióxido de Nitrógeno en la Región Metropolitana de São Paulo. Al final de este estudio, observamos que los valores reales y predichos tenían una excelente correlación. Concluimos que el modelo ARIMA es un método estadístico eficiente, al que posibilita la aplicación para la emisión de alertas con un cierto grado de confianza.

Keywords: ARIMA, Nitrogen Dioxide, OMI/AURA

1. INTRODUCTION

The implementation of public policies aiming at air quality management requires methods to accurately predict the concentration of pollutants (Sharma and Chandra, 2009) aiming to establish the control of air pollution in urban areas, which is intensified by the increase in vehicular traffic (Kumar and Jain, 2010). The OMI sensor (Ozone Monitoring Instrument) of the NASA Aura satellite has a database of observations of several pollutants that began in October 2004 in the ozone layer and for NO_2 analysis for near surface studies, data are used from the OMI sensor Tropospheric Column (Krotkov et al., 2016). For the present study, an ARIMA (Auto Regressive Integration Moving Average) model will be adjusted and validated to predict NO_2 concentration in the Metropolitan Region of São Paulo (RMSP), considering that it is an area where there are daily critical problems in air quality and statistical models are good tools for generating alerts since physical models are not efficient for the same purpose.

2. METODOLOGY

Total NO_2 column concentration data, for the area comprised of the RMSP, were obtained through the OMI Tropospheric Column NO_2 sensor in the Giovanni / NASA platform (https://giovanni.gsfc.nasa.gov/giovanni/). An ARIMA model was adjusted and used to predict future values of the NO_2 time series, the model consists of extracting the predicted values based on the trend and correlation of the series from the observed data with the models of autoregression (AR) and moving average (MA) for N amount of data where they are replaced by the differentiation of the current values with their previous ones through the integration operators (I). Further details on ARIMA models can be obtained in Sharma et.al, 2009. The first 300 days of 2007 were selected and we applied a cross-validation process. We consider a need for NO_2 concentration alerts for 24h. Thus, we adjusted the model to generate forecasts for the day after the present day using observations from previous days. We adopted Pearson's correlation as the forecast quality metric. We use the R programming environment (https://www.r-project.org/) with the help of the ggplot2 and forecast packages for the treatment, analysis and visualization of the data.

3. RESULTS

From the various configurations tested to find the best model for the NO_2 forecast in the RMSP, we arrive at the ARIMA configuration (2,1,0). Image 1 shows the scatter plot between the predicted and actual data by the model. It is possible to observe that the points form a linear pattern around the blue line, indicating a considerable relationship between the data. Figure 2 corroborates Figure 1, showing that the model is predicting NO_2 satisfactorily. The estimated Pearson coefficient was 0.86 and the coefficient of determination was 0.73.

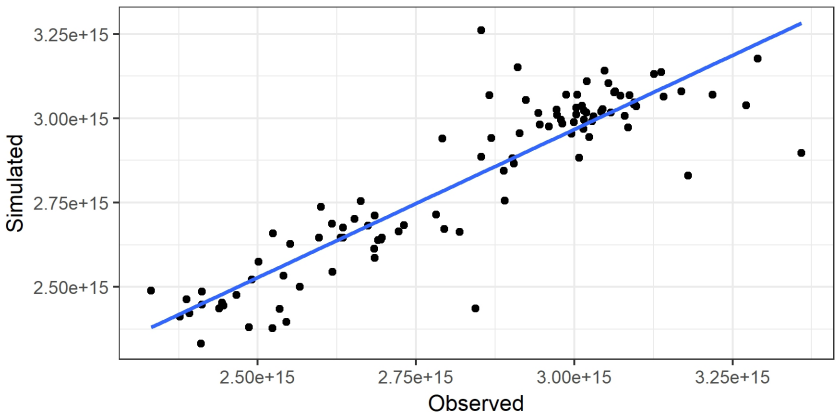

Figure 1: Scatter plot of NO_2 concentration for the RMSP. The black points are the set of the observed and predicted data (molecules/cm2) and the blue line represents the line of adjustment linear regression between the data.

4. CONCLUSIONS

At the end of the present study, we highlight the competence of the ARIMA model for application in the prediction of nitrogen dioxide with the objective of providing methods that allow the daily emission of alerts by air quality bodies, especially in large urban centers. The use of this methodology in the analysis of the concentration data is essential to carry out studies aiming to improve the quality of life of the population, minimizing the effects of extreme values, and allows to establish strategies to avoid exposure to potentially dangerous rates of human health.

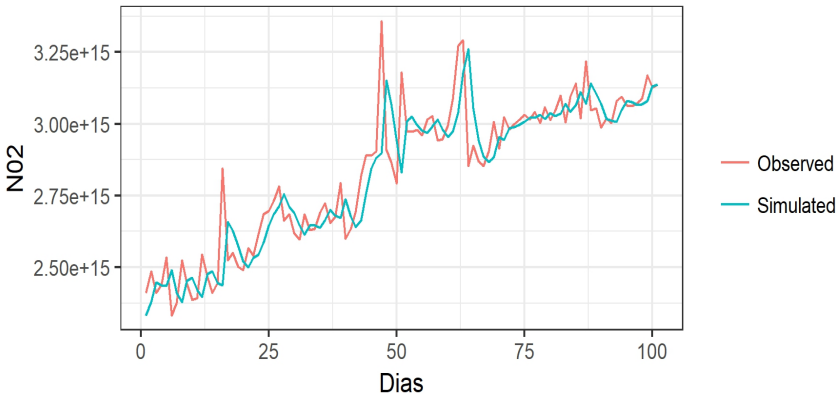

Figure 2: Temporal series of the observed and predicted data of NO2 concentration

ACKNOWLEDGE

We thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior) for the financial support.

REFERENCES

Krotkov, N.A.; McLinder, C.A. Li, C.; Lamsal, L.N.; Celarier, E.A.; Marchenko, S.V.; Swarts, W.H. and et.al., 2016: Aura OMI observations of regional SO_2 and NO_2 pollution changes from 2005 to 2015. Atmospheric Chemistry and Physics, 16, 4605-4629.

Kumar, U.; Jain, V.K., 2010: ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO). Stochastic Environmental Research and Risk Assessment, 24, 751-760.

Sharma, P.; Chandra, A.; Kaushik, S.C., 2009. Forecasting using Box-Jenkins models for the ambient air quality data of Delhi City. Environmental monitoring and assessment, 157, 105-112.