International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 7, Issue 1, Dec - Jan 2018; 1 - 12

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

A PRINCIPAL COMPONENT REGRESSION MODEL, FOR FORECASTING DAILY PEAK AMBIENT GROUND LEVEL OZONE CONCENTRATIONS, IN THE PRESENCE OF MULTICOLLINEARITY AMONGST PRECURSOR AIR POLLUTANTS AND LOCAL METEOROLOGICAL CONDITIONS: A CASE STUDY OF MAUN

W. M. Thupeng[1], T. Mothupi[2], B. Mokgweetsi[3], B. Mashabe[4] & T. Sediadie[5]

[1,2,3,5]Department of Statistics, University of Botswana, Gaborone, Botswana

[4]Department of Mathematics and Statistical Sciences, Botswana International University of Science and Technology, Palapye, Botswana

ABSTRACT

The increasing public awareness of the negative health effects of exposure to peak air pollution levels, particularly to the most sensitive population sub-groups like children and the elderly, has made short-term forecasts of episodes of peak concentrations of air pollutants at a local level, increasingly necessary. The main objective of the present study is to develop a statistical model for predicting a day in advance the daily maximum 1- hour average ambient ground level ozone concentration for Maun town, using principal component regression (PCR) technique. The predictor variables are the precursor air pollutants of ground level ozone, namely, nitrogen oxides, nitrogen dioxide and the previous day's ground level ozone concentration, on the one hand, and meteorological variables that include wind speed, wind direction, relative humidity, surface temperature, atmospheric pressure and solar radiation. The data consist of maximum 1-hour interval concentrations every day, on the response and each of these predictor variables collected from 1 May 2014 to 30 September 2015. A biased regression method of PCR is applied to try and minimise the problem of multicollinearity, usually associated with multiple regression models. The detection of multicollinearity is performed by using the Pearson partial correlation matrix, and variance inflation factor (VIF). Model assessment tools include the tests for significance of individual regression coefficients in the PCR model, the coefficient of determination and F test to test for the validity of the overall model. It is found that the estimated PCR model is based on principal components that are highly correlated with maxima of the ozone concentration the day before, nitrogen oxides concentrations and surface temperature. Furthermore, wind speed, wind direction, relative humidity and nitrogen dioxide are identified as possible causes of multicollinearity, in the available data.

***KEYWORDS**: Ambient Air Pollution, Meteorological Conditions, Multicollinearity, Peak Ambient Ground level Ozone, Precursor Air Pollutants, Principal Component Regression*

Article History

***Received:** 22 Nov 2017 | **Revised:** 12 Dec 2017 | **Accepted:** 19 Dec 2017*