Forecasting particulate matter concentration using nonlinear autoregression with exogenous input model

ABSTRACT

Air quality in some developing countries is dominated by particulate matter, especially those with size 10 micrometers and smaller or PM10. They can be inhaled and sometimes can get deep into lungs; some may even get into bloodstream and cause serious health problems. Therefore, future PM10 concentration forecasting is important for early prevention and in urban development planning, which is crucial for developing cities. This paper presents the development of PM10 forecasting model using nonlinear autoregressive with exogenous input model. To improve performance of nonlinear autoregressive with exogenous input model, principal component analysis is used prior to the model for variable selection. The first stage of principal component analysis involves Scree plot, which determines the number of principal components based on explained variance. This is then followed by selecting variables using a rotated component matrix, based on their strength of contribution towards variation of PM10 concentration. To test the model, PM10 data in Kota Kinabalu from 2003 – 2010 was used. Neural network models are developed using this data by varying number of input variables with the inclusion of temporal variables. The developed forecasting models are evaluated using data PM10 in the city from 2011 to 2012. Four performance indicators, namely root mean square error, mean absolute error, index of agreement and fractional bias are reported. Results from principal component analysis show that five variables including wind direction index, relative humidity, ambient temperature, concentration of nitrogen dioxide and concentration of ozone strongly contribute to the variation of PM10 concentration. By using these variables together with temporal variables as input in the nonlinear autoregressive with exogenous input models, the resultant model shows good forecasting performance, with root mean square error of 7.086±0.873 μg/m3. The selection of significant variables helps in reducing input variables inside the forecast model without degrading its forecast performance. This model shows very promising performance in forecasting PM10 concentration in Kota Kinabalu as it requires fewer input variables and does not require variable transformation.