

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu in fulfilment of the requirements for the degree of Master of Science

**MULTIVARIATE STATISTICAL ANALYSIS FOR NEXT HOUR PM<sub>10</sub>  
CONCENTRATION FORECASTING DURING EPISODIC TRANS-  
BOUNDARY HAZE EVENTS IN MALAYSIA**

**NUR NAZMI LIYANA MOHD NAPI**

**APRIL 2021**

**Main Supervisor : Samsuri Abdullah, Ph.D**

**Co- Supervisor : Professor Marzuki Ismail, Ph.D  
Ali Najah Ahmed, Ph.D**

**Faculty : Faculty of Ocean Engineering Technology and  
Informatics**

Episodic trans-boundary haze events had happens caused higher particulate matter (PM<sub>10</sub>) concentration in Malaysia's atmosphere. The high concentration of PM<sub>10</sub> had exceeded the New Malaysia Ambient Air Quality Standard (NMAAQS) and caused a higher Air Pollutant Index (API) reading. Hence, this situation prompts an adverse impact on human health and also the country's economic sector. All the meteorological factors (ambient temperature, relative humidity, and wind speed) data and other gaseous pollutants (carbon monoxide, sulphur dioxide, nitrogen oxide, nitrogen dioxide, and ozone) that involved in this study were acquired from the Department of Environment Malaysia from the year 2005 until 2014 based on the haze chronology information. Thus, the correlation among the influenced factors led to multicollinearity problems, especially in model development. The aim of this study is to develop Multiple Linear Regression (MLR) and Principal Component Regression (PCR) models. The multicollinearity problem has occurred among the independent variables, and the Principal Component Analysis (PCA) has been introducing in the PCR model. The PCA can minimize the multicollinearity problem among the independent variables by grouped it to the respective principal component, PCs. The generated PCs are used as an input in the PCR model. In the MLR model, all the

parameters involved have been using as input in that model development. As a result, the MLR model was performed well compared to the PCR model in forecasting the concentration of the next hour of  $PM_{10}$  during the trans-boundary haze event. In conclusion, the MLR model had selected as the best-fitted model in this study. It has lower error measures of Root Mean Square Error (RMSE, 126.728), Normalized Absolute Error (NAE, 0.325), and Mean Absolute Error (MAE, 89.280). Meanwhile, it also has a higher accuracy measure in term of the correlation coefficient ( $R^2$ , 0.445) and Prediction Accuracy (PA,0.668) as compared to the PCR model with a value of RMSE (261.471), NAE (0.945), MAE (227.013), MAE (227.013),  $R^2$  (0.445) and PA (0.667). Therefore, the MLR model forecasting is very useful for local authorities and policymakers in managing air quality during trans-boundary haze events in Malaysia.

Abstrak tesis yang dikemukakan kepada Senat Universiti Malaysia Terengganu sebagai memenuhi keperluan untuk Ijazah Sarjana Sains

**ANALISIS MULTIVARIAT BAGI PERAMALAN KEPEKATAN PM<sub>10</sub> PADA JAM YANG BERIKUTNYA SEMASA EPISOD JEREBU RENTAS SEMPADAN DI MALAYSIA**

**NUR NAZMI LIYANA MOHD NAPI**

**APRIL 2021**

**Penyelia Utama : Samsuri Abdullah, Ph.D**

**Penyelia Bersama : Professor Marzuki Ismail, Ph.D  
Ali Najah Ahmed, Ph.D**

**Fakulti : Faculty of Ocean Engineering Technology and Informatics**

Episod kejadian jerebu rentas sempadan yang berlaku telah menyebabkan kepekatan bahan partikulat (PM<sub>10</sub>) yang lebih tinggi di lapisan atmosfera Malaysia. Kepekatan PM<sub>10</sub> telah melebihi Standard Baru Kualiti Udara Ambien Malaysia (NMAAQS) dan menyebabkan bacaan Indeks Pencemaran Udara (API) menjadi lebih tinggi. Oleh itu, keadaan ini telah memberi kesan buruk kepada kesihatan manusia dan juga sektor ekonomi negara. Semua data faktor meteorologi (suhu persekitaran, kelembapan relatif, dan kelajuan angin) dan pencemaran gas lain (karbon monoksida, sulfur dioksida, nitrogen oksida, nitrogen dioksida dan ozon) yang terlibat dalam kajian ini telah diperolehi daripada Jabatan Alam Sekitar Malaysia dari tahun 2005 sehingga 2014 berdasarkan maklumat dalam kronologi jerebu. Oleh itu, korelasi di antara faktor-faktor yang dipengaruhi telah menyebabkan masalah multikolineariti, terutama dalam pengembangan model. Tujuan utama kajian ini adalah untuk mengembangkan model Regresi Linear Pelbagai (MLR) dan Regresi Komponen Utama (PCR). Masalah multikolineariti telah berlaku di antara pemboleh ubah bebas, dan Analisis Komponen Utama (PCA) telah diperkenalkan dalam model PCR. PCA dapat meminimumkan masalah multikolineariti di antara pemboleh ubah bebas dengan mengelompokkannya di dalam Komponen Utama (PC) yang

berkenaan. PC yang dihasilkan digunakan sebagai input di dalam model PCR. Di dalam model MLR, semua parameter yang terlibat di dalam kajian ini telah digunakan sebagai input di dalam pengembangan model tersebut. Hasilnya, model MLR telah menunjukkan prestasi yang baik berbanding model PCR dalam meramalkan kepekatan  $PM_{10}$  pada jam yang berikutnya semasa kejadian jerebu rentas sempadan berlaku. Kesimpulannya, model MLR telah dipilih sebagai model yang paling sesuai dalam kajian ini. Ia mempunyai ukuran ralat yang lebih rendah dalam ralat punca min kuasa dua (RMSE, 126.728), ralat mutlak normal (NAE, 0.325), dan ralat mutlak purata (MAE, 89.280). Sementara itu, ia juga mempunyai ukuran ketepatan lebih tinggi dari segi perkali korelasi ( $R^2$ , 0.445) dan ketepatan ramalan (PA, 0.668) berbanding dengan model PCR dengan nilai RMSE (261.471), NAE (0.945), MAE (227.013), MAE (227.013),  $R^2$  (0.445) dan PA (0.667). Oleh yang demikian, model peramalan MLR sangat berguna kepada kerajaan tempatan dan juga penggubal polisi dalam menguruskan kualiti udara semasa kejadian jerebu rentas sempadan di Malaysia.