Geophysical Research Abstracts
Vol. 20, EGU2018-11497, 2018
EGU General Assembly 2018

Development of a Land Use Regression model using Support Vector Regression for estimating ambient air pollution exposure to $PM_{2.5}$ in Dublin, Ireland

Bidroha Basu, Md. Saniul Alam, Bidisha Ghosh, Laurence Gill, and Aonghus McNabola
Trinity College Dublin, Civil, Structural & Environmental Engineering, Dublin, Ireland (bbasu@tcd.ie)

Assessment for epidemiological studies of long-term exposure to ambient air pollution is a challenging task. Land Use Regression (LUR) has been considered to be an effective and economical means of estimating air pollution exposures for epidemiological studies due to its simplicity and ability to explain small-scale variation in air pollution concentrations.

The development of LUR models is often based on air pollution concentration data obtained from fixed monitoring stations. In most major cities, the number of monitoring stations are limited and the data are inadequate to model the spatial variability. To overcome this limitation, additional mobile station data can be obtained at several locations covering a significant portion of the city. However, since obtaining additional data is costly, those data are usually collected for a short period of time. The uncertainty associated with short term, mobile station data are considerably higher than that of the fixed monitoring stations data, which needs to be accounted for while developing an LUR model. A weighted function was considered here to account for differences in record duration of data in this study. Another important factor in developing an LUR model is the identification of proper set of predictors that influence the ambient air pollution concentration. Predictors considered for LUR modelling include topography, land use, traffic volume and meteorological variables. The appropriate distance of influence (buffer) area corresponding to each predictors are not known a priori. A buffer distance decay curves-based approach was considered in this study to identify appropriate predictors and their associated buffer distance.

The analysis was performed in Dublin city for fine particulate matter ($PM_{2.5}$) where 8 fixed monitoring stations data were available, with average daily concentration data records of 2 to 7 years in length. 102 mobile station data were also collected at several locations in the city for a duration of 12 hours to 2 days. The performance of a developed multiple linear regression based LUR model was evaluated by using the leave-one-out-cross-validation procedure. The model explained variances (R^2) of the LUR model were found to be 0.44. One of the primary reason for poor performance of the LUR model was that the relationship between the predictors and the mean $PM_{2.5}$ concentration were found to be non-linear which cannot be modelled by using multiple linear regression approach. To address this, a weighted support vector regression approach was also considered which can account for non-linearity while developing the LUR model. The model performance in terms of R^2 was found to improve to 0.56 using this method. Further research is going on to account for temporal variability along with spatial variation in the developed LUR model.