



Article

# Kriging-Based Land-Use Regression Models That Use Machine Learning Algorithms to Estimate the Monthly BTEX Concentration

Chin-Yu Hsu <sup>1,2</sup>, Yu-Ting Zeng <sup>3</sup>, Yu-Cheng Chen <sup>4</sup> , Mu-Jean Chen <sup>4</sup>,  
Shih-Chun Candice Lung <sup>5,6,7</sup> and Chih-Da Wu <sup>3,4,\*</sup>

<sup>1</sup> Department of Safety, Health and Environmental Engineering, Ming Chi University of Technology, New Taipei 243303, Taiwan; gracecyhsu@mail.mcut.edu.tw

<sup>2</sup> Center for Environmental Sustainability and Human Health, Ming Chi University of Technology, New Taipei 243303, Taiwan

<sup>3</sup> Department of Geomatics, National Cheng Kung University, Tainan 70101, Taiwan; z10708105@ncku.edu.tw

<sup>4</sup> National Institute of Environmental Health Sciences, National Health Research Institutes, Miaoli 35053, Taiwan; yucheng@nhri.org.tw (Y.-C.C.); zeromagi@nhri.org.tw (M.-J.C.)

<sup>5</sup> Research Center for Environmental Changes, Academia Sinica, Taipei 11529, Taiwan; sclung@rcec.sinica.edu.tw

<sup>6</sup> Department of Atmospheric Sciences, National Taiwan University, Taipei 10617, Taiwan

<sup>7</sup> Institute of Environmental Health, National Taiwan University, Taipei 10055, Taiwan

\* Correspondence: chidawu@mail.ncku.edu.tw; Tel.: +886-6-275-7575 (ext. 63841)

Received: 11 August 2020; Accepted: 20 September 2020; Published: 23 September 2020



**Abstract:** This paper uses machine learning to refine a Land-use Regression (LUR) model and to estimate the spatial–temporal variation in BTEX concentrations in Kaohsiung, Taiwan. Using the Taiwanese Environmental Protection Agency (EPA) data of BTEX (benzene, toluene, ethylbenzene, and xylenes) concentrations from 2015 to 2018, which includes local emission sources as a result of Asian cultural characteristics, a new LUR model is developed. The 2019 data was then used as external data to verify the reliability of the model. We used hybrid Kriging-land-use regression (Hybrid Kriging-LUR) models, geographically weighted regression (GWR), and two machine learning algorithms—random forest (RF) and extreme gradient boosting (XGBoost)—for model development. Initially, the proposed Hybrid Kriging-LUR models explained each variation in BTEX from 37% to 52%. Using machine learning algorithms (XGBoost) increased the explanatory power of the models for each BTEX, between 61% and 79%. This study compared each combination of the Hybrid Kriging-LUR model and (i) GWR, (ii) RF, and (iii) XGBoost algorithm to estimate the spatiotemporal variation in BTEX concentration. It is shown that a combination of Hybrid Kriging-LUR and the XGBoost algorithm gives better performance than other integrated methods.

**Keywords:** nitrogen dioxide (NO<sub>2</sub>); hybrid Kriging-LUR model; culture-specific sources; spatiotemporal variations

## 1. Introduction

Chemical and petroleum facilities are major emitters of volatile organic compounds (VOCs) into the environment [1,2]. These industrial emissions include benzene, toluene, ethylbenzene, and xylenes, which are also known as BTEX [3,4]. On the other hand, ambient BTEX might also result from various substances and sources, including traffic, gas stations, combustion processes, and households [5–7]. BTEX emissions have a significant effect on human health. For example, the International Agency for Research on Cancer classifies benzene as carcinogenic for humans [8]. Benzene also affects blood

production, the lymphatic system, and the central nervous system [5,9]. Even at low concentrations, BTEX has an adverse effect on reproductive processes, cardiovascular disease, respiratory dysfunction, asthma, and sensitivity to common antigens [10]. Several studies report that residents near sources of emission from chemical/petroleum facilities are exposed to relatively high levels of BTEX [11,12]. Other studies also show a positive correlation between cancer risk (leukemia and brain tumor) and exposure to benzene or surrogates for residents who live close to petrochemical facilities [12,13]. This highlights the importance of predicting BTEX concentrations to help policy makers to assess the prevention measures. However, few studies have addressed within-city variability in the level of BTEX.

Kaohsiung City, which is a heavily industrialized harbor city located in southern Taiwan, has a population of 2.8 million and an area of 153.6 km<sup>2</sup>. There are numerous industrial parks, petrochemical facilities, and more than 1.5 million motor vehicles in Kaohsiung. These emission sources have a negative effect on air quality in Kaohsiung, and partially on the high level of VOCs. The air quality in autumn and winter is worst because the atmosphere is stable, winds are slow, and there is diurnal land–sea breeze circulations in the cold and dry seasons. Although the Taiwanese Environmental Protection Agency (EPA) has imposed emission standards to control ambient air quality and minimize the risk to health, residents in Kaohsiung still have a long-standing concern for pollutant exposure, partly because it is still unknown how long-term VOCs exposure affects human health. To find out, the spatial variability of pollution concentration is essential.

Kriging and land-use regression (LUR) are used to predict air pollution gradients if there is limited sampled data. Kriging is a method of spatial interpolation that assumes the distance or direction between sampling points to reflect a spatial correlation. This spatial correlation is used to explain variation in the surface. The LUR model is widely used to estimate spatial variation in VOC concentrations to determine a population's exposure to pollution [14–16]. In addition, using adjacent monitoring sites and local emission points as well as estimated air pollution concentrations from Kriging interpolation as a variable for LUR (Hybrid Kriging-LUR) allows a more accurate prediction of air pollution levels [17]. However, these models can ignore the dynamic spatial and temporal relationship between VOCs and predictor variables because LUR and Hybrid Kriging-LUR models only produce a single regression equation to summarize global relationships between independent and dependent variables. These regression models can also underestimate the pollution levels and fail to identify nonlinear relationships between predictions and observations that may not be linearly correlated.

A geographically weighted regression (GWR) model considers the spatial variation in relationships and creates maps to determine nonstationary spatial relationships [18]. Machine learning algorithms such as Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) are also widely used to determine air pollution concentrations because they identify nonlinear relationships between observations and predictors [19–21].

This study used Hybrid Kriging-LUR alone and then with GWR and machine learning algorithms (RF and XGBoost) to estimate the spatiotemporal variation in BTEX in Kaohsiung. To increase the accuracy of each model, the distribution of local temples was used as a predictor to represent the unique emissions from sources that are unique to Asian culture. In terms of residents' health and indicators of the health effect, this study shows that air epidemiological studies of ambient BTEX are important for the future.

## 2. Materials and Methods

### 2.1. Study Area

Kaohsiung is an industrial city located in southern Taiwan, with a population density of 3957 (people per km<sup>2</sup>), three petrochemical industrial parks, a large iron ore and steel factory, and many factories that use oil/coal combustion. There are about 3 million registered motor vehicles (including motorbikes, cars, and other vehicles) in this city. There are 72 vehicles per hundred people so traffic

emissions are a significant factor for air pollution in Kaohsiung. On average there are 10 factories per square kilometer and many of these are located near commercial districts and residential areas. Local culture also plays a role in this study because Taiwan features unique emission sources of BTEX, such as the frequent burning of joss paper and incense in thousands of temples [22,23]. The present study area crossed six districts in Kaohsiung and one district in Pingung (Figure S1). It also covered two large petrochemical industrial parks (Linhai and Linyuan). The study districts contributed respective ~50% of sulfur oxides (SO<sub>x</sub>), ~60% of nitrogen oxides (NO<sub>x</sub>), and ~60% of VOCs to ambient pollutants in Kaohsiung in 2018. The registered vehicles in the study districts were about 40% of total in Kaohsiung. In addition, the largest iron ore and steel factory in Taiwan is located in Linhai petrochemical industrial park.

## 2.2. Air Pollutant Database

The Taiwanese EPA requires that air pollution monitoring stations must be established in all townships in close proximity to petrochemical industrial parks. This study uses five years of BTEX data (from May 2015 to June 2019) collected from 17 monitoring stations that are close to two petrochemical industrial parks in south Kaohsiung (shown in Figure S1). The data from 2015 and 2018 were used to develop models and observations from 2019 and are also used for external data verification to assess the reliability of the models.

This study uses 10,660 hourly measurements of BTEX, which are aggregated into 939 monthly averages for the model. The concentrations of monitored pollutants are obtained from the EPA database as explanatory variables. Previous studies confirm monitored pollutants' association with BTEX concentration (e.g., NO<sub>x</sub> and O<sub>3</sub>) [24,25]. Table S1 lists the potential predictor variables used in this study.

## 2.3. Geospatial Database

To develop Hybrid Kriging-LUR, information about land-use or land-cover from several GIS layers and spatial databases is required. The land-use inventory, the landmark database, the digital road network map, the Digital Terrain Model (DTM), Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) database, and the thermal power plant distribution dataset are used. Further details of land-use or land-cover related information for potential prediction variables (Table S1) can be found in a previous study by the authors [17,26,27]. In this study, LUR models at monthly resolution were developed based on air pollutants measurements from 2015 to 2019. It is difficult to obtain clear Landsat or SPOT images for every month in Taiwan due to the humid and cloudy weather [28]. In this case, we just followed previous studies to obtain NDVI information from MODIS [29].

## 2.4. LUR Model Development and Validation

This study uses a Hybrid Kriging-LUR to identify important prediction variables. The variables that are selected by Hybrid Kriging-LUR are used for the geographically weighted regression (GWR) and two machine learning algorithms (random forest (RF) and extreme gradient boosting (XGBoost)) to develop the prediction models called RF-Hybrid LUR or XGBoost-Hybrid LUR.

This study uses the same Hybrid Kriging-LUR as that proposed by Wu et al. [17]. Kriging-based estimations of the BTEX level are used as an explanatory variable for a stepwise selection during the conventional LUR procedure, using a leave-one-out Ordinary Kriging interpolation algorithm for “n-1” observations. The Hybrid Kriging-LUR approach uses Spearman correlation coefficients to define the bivariate association between each BTEX compound and all of the potential prediction variables. A supervised stepwise procedure is used to maximize the percentage of explained variability (R<sup>2</sup>). For all potential predictor variables, an a-priori direction of effect is used for each BTEX (e.g., positive for road length and residential area to benzene). The variable initially has the highest explained variance in a univariate analysis and a regression slope with the expected direction. All other variables

are then added to this model separately by assessing whether the  $p$ -value is  $<0.1$  and the variance inflation factor (VIF) is  $<3$ . This procedure continues until none of the variables fit the specified criteria. Finally, The  $R^2$  adjusted  $R^2$  values and the Root Mean Square Error (RMSE) are used to determine the model's performance.

GWR is used to solve the model spatially. The equation (Equation (1)) for the GWR model is defined as follows:

$$Y_i = \beta_{(U_i, v_i)} + \sum_k \beta_{k(U_i, v_i)} X_{ik}, \quad (1)$$

where  $(U_i, v_i)$  denotes the coordinates of the point in a location;  $Y_i$  is BTEX concentration;  $\beta_{(U_i, v_i)}$  represents the intercept value;  $\beta_{k(U_i, v_i)}$  is a set of values of parameters at point  $i$ ; and  $X_{ik}$  are prediction variables that are obtained using Hybrid Kriging-LUR approaches.

The RF grows multiple decision trees and forces a randomly selected subset of candidate predictors into each tree [30]. RF-based Hybrid Kriging-LUR approaches produce 200 regression trees, which are extracted from randomly bootstrapped features from the training data. The extent to which a tree grows also affects the model's performance. This study uses depths of 30 for Hybrid Kriging-LUR models.

XGBoost is a common machine learning algorithm that was first proposed by Chen and Guestrin [31]. It has been proved very successful in many machine learning competitions. XGBoost is similar to a random forest approach in that it features multiple regression trees. The tree ensemble model trains weak learners to optimize the model using the bias for the loss function by boosting a scalable gradient tree. If XGBoost learners with a different feature importance score are generated across all trees, the prediction is accumulated in terms of the weight of each learner. The Hybrid Kriging-LUR approaches use 130 trees, and the maximum depths to which the trees are grown was 8. The parameter values for each method are listed in Tables S2 and S3.

Land-use/land-cover information is extracted using ArcGIS 10.5 (Esri, Redlands, CA, USA). LUR and all statistical analyses are conducted using SPSS 22.0 (IBM, New York, NY, USA) and R 3.5.2. (The R Foundation for Statistical Computing, Vienna, Austria) These machine learning models are programmed in Python 3.7, using a Jupyter Notebook platform. The computer hardware is a laptop (ASUS, Taipei City, Taiwan) with a CPU i5-8265U and 8 GB of RAM.

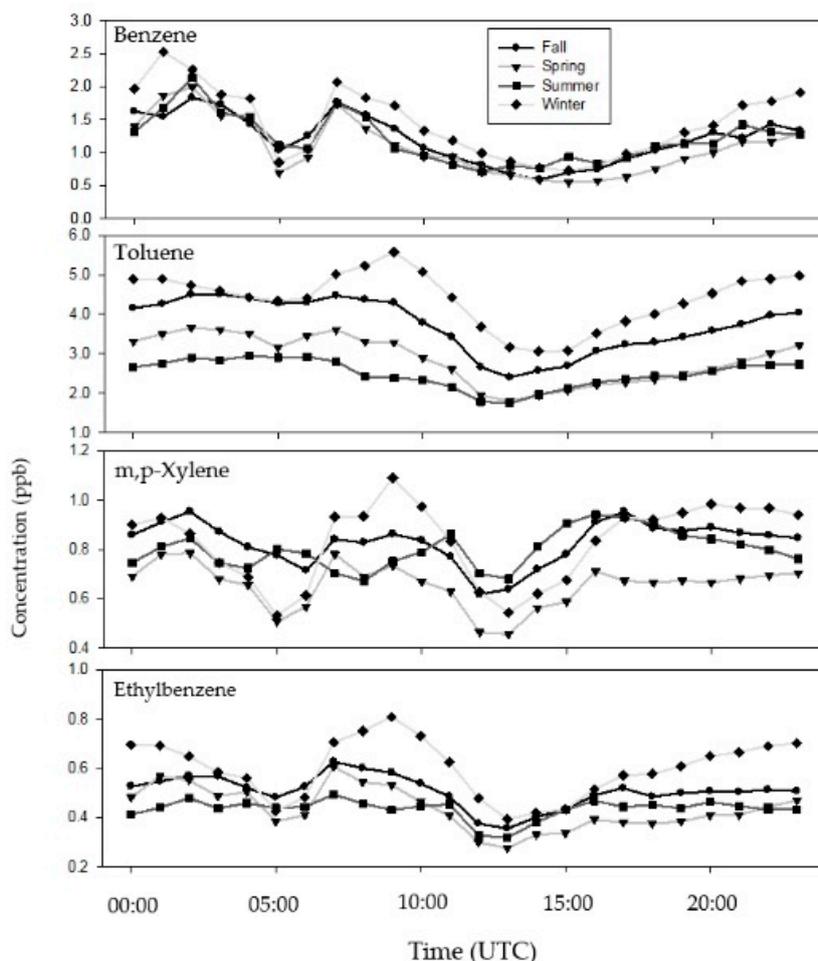
### 3. Results

#### 3.1. Descriptive Statistics for BTEX Concentrations

Figure S1 shows annual wind rose for 2015–2018. The site experiences a predominantly Westerly wind flow in the spring and winter, in all directions in the summer, and a Westerly to Northern wind flow in the fall. The winds generally blew at 0.2 to 5.37 m/s and at an average  $2.3 \pm 1.35$  m/s. In terms of annual BTEX concentration, toluene is the dominant BTEX compound in the study area ( $3.31 \pm 4.01$  ppb), followed by benzene ( $1.22 \pm 5.57$  ppb), *m,p*-xylene ( $0.78 \pm 1.46$  ppb), and ethylbenzene ( $0.49 \pm 1.16$  ppb). The highest BTEX concentrations in Kaohsiung were greater than those in Beijing (1.44 ppb for toluene, 0.54 ppb for benzene, 0.48 ppb for *m,p*-xylene, and 0.27 ppb for ethylbenzene) [32] and Tianjin (0.50 ppb for toluene, 1.22 ppb for benzene, 0.57 ppb for *m,p*-xylene, and 0.51 ppb for ethylbenzene) [33] in China. Kaohsiung also has higher BTEX figures than areas near the largest petrochemical industrial parks in Taiwan (2.56 ppb for toluene, 0.22 ppb for benzene, 0.14 ppb for *m,p*-xylene, and 0.07 ppb for ethylbenzene) [34].

Figure 1 shows BTEX average diurnal variations in each season during the study period. For example, the concentration of BTEX compounds became relatively lower in the daytime (from 10:00 to 15:00) with the lowest concentration observed at ~13:00. This is consistent with what has been reported in Shanghai for a similarly-situated large industrial estate [35]. Such a diurnal trend is likely caused by strong solar radiation and intense air convection in the daytime, both of which can photochemically react with and/or dilute VOCs [35]. In contrast, as shown in Figure 1, we can see higher BTEX concentrations during rush hours both in the morning (7:00 to 9:00) and late afternoon (~18:00), similar to the findings in previous studies [36,37]. Such a high concentration

also suggests that automobile exhaust was an important source for atmospheric BTEX in the study area. Figure 1 further shows similar diurnal variations of BTEX across four seasons, indicating that the BTEX concentration is contributed from similar sources and dispersion mechanisms in each season. In addition, some variances of BTEX concentrations in Figure 1 are likely influenced by many factors such as emission sources (mainly from vehicular exhaust, gasoline, and solvent evaporation), meteorological conditions, and their sinks, given that the study area is located in the industrial area of Kaohsiung. Indeed, benzene, toluene, and xylenes are typical tracers of vehicular exhaust, industrial production, and solvent usage, respectively [38,39]. This also explains why Figure 1 shows distinct diurnal variations of BTEX concentration in each season.



**Figure 1.** The diurnal variation in the BTEX (benzene, toluene, ethylbenzene, and xylenes) concentrations for each season, averaged over 17 sampling sites.

### 3.2. Development and Validation of The LUR and Machine Learning Models

Table 1 shows the selected prediction variables, the estimated coefficient, the partial  $R^2$  value and the VIF for the proposed Hybrid Kriging-LUR model. The variables, Benzene<sub>Kriging-based</sub>, UV, rice farm within a 150-m buffer, and harbor and industrial area within a 500-m buffer, all have a significant effect on the explanatory power of the model for benzene. In terms of toluene, the significant variables are toluene<sub>Kriging-based</sub>,  $NO_x$ , water body, purely residential area within a 250-m buffer, sandstone field within a 150-m buffer, sandstone field within a 2500-m buffer, industrial area within a 150-m buffer, all types of road within a 50-m buffer, and temple within a 250-m buffer. In terms of ethylbenzene, the significant variables are ethylbenzene<sub>Kriging-based</sub>,  $SO_2$ , winter, industrial area within a 250-m buffer, temple within a 250-m buffer, fruit orchard within a 50-m buffer, and fruit orchard within a 1500-m

buffer. In terms of m,p-xylene, the factors that have the most significant effect on the explanatory power of the model are m,p-xylene<sub>Kriging-based</sub>, sandstone field within a 150-m buffer, funeral services within a 1250-m buffer, industrial area within a 50-m buffer, local road within a 250-m buffer, and temple within a 250-m buffer. All of these variables discussed are used to develop hybrid models, such as GWR with the Hybrid Kriging-LUR (GWR-Hybrid LUR), RF with the Hybrid Kriging-LUR (RF-Hybrid LUR), and XGBoost with the Hybrid Kriging-LUR (XGBoost-Hybrid LUR). Most variables have a positive effect on BTEX, except for UV and harbor for benzene, and sandstone field for toluene, ethylbenzene, and m,p-xylene.

**Table 1.** Prediction variables for the Hybrid Kriging-LUR model. LUR—land-use regression; VIF—variance inflation factor.

BTEX	Variable	Coefficient	p-Value	Partial R <sup>2</sup>	VIF
Benzene	Intercept	1.964	<0.05	-	-
	Benzene <sub>Kriging-based</sub>	0.223	<0.05	0.006	1.395
	Ultraviolet	-0.163	<0.05	0.045	1.394
	Rice farm <sub>150m</sub>	0.002	<0.05	0.068	1.272
	Harbor <sub>Nearest distance</sub>	-1.113 × 10 <sup>-4</sup>	<0.05	0.070	1.163
	Industry <sub>500m</sub>	0.002	<0.05	0.240	1.185
Toluene	Intercept	-1.229	<0.05	-	-
	Toluene <sub>Kriging-based</sub>	0.581	<0.05	0.061	2.366
	Nitrogen Oxides	0.068	<0.05	0.246	2.311
	Water body <sub>Nearest distance</sub>	5.966 × 10 <sup>-4</sup>	<0.05	0.001	1.412
	Purely residential area <sub>250m</sub>	0.002	<0.05	0.048	1.649
	Sandstone field <sub>150m</sub>	-0.005	<0.05	0.058	1.102
	Sandstone field <sub>250m</sub>	0.002	<0.05	0.002	1.257
	Industry <sub>150m</sub>	6.208 × 10 <sup>-4</sup>	<0.05	0.025	1.406
	All types of road(width) <sub>50m</sub>	3.241 × 10 <sup>-4</sup>	0.153	0.005	1.359
	Temple <sub>250m</sub>	0.515	<0.05	0.071	1.403
Ethylbenzene	Intercept	-0.105	0.442	-	-
	Ethylbenze <sub>Kriging-based</sub>	0.072	0.239	0.007	1.097
	SO <sub>2</sub>	0.094	<0.05	0.032	1.342
	winter	0.114	<0.05	0.011	1.299
	Industry <sub>250m</sub>	3.737 × 10 <sup>-4</sup>	<0.05	0.160	1.072
	Temple <sub>250m</sub>	0.105	<0.05	0.096	1.056
	Sandstone field <sub>500m</sub>	-3.224 × 10 <sup>-4</sup>	<0.05	0.010	1.928
	Fruit orchard <sub>50m</sub>	6.428 × 10 <sup>-4</sup>	<0.05	0.038	1.635
	Fruit orchard <sub>1500m</sub>	5.927 × 10 <sup>-4</sup>	0.161	0.003	2.434
m,p-Xylene	Intercept	-0.045	0.778	-	-
	m,p-Xylene <sub>Kriging-based</sub>	0.432	<0.05	0.041	1.062
	Sandstone field <sub>150m</sub>	-8.339 × 10 <sup>-4</sup>	0.079	0.040	1.169
	Funerary services <sub>1250m</sub>	0.003	<0.05	0.011	1.041
	Industry <sub>50m</sub>	6.963 × 10 <sup>-4</sup>	<0.05	0.075	1.516
	Local road <sub>250m</sub>	16.121	<0.05	0.010	1.518
	Temple <sub>250m</sub>	0.364	<0.05	0.248	1.042

BTEX = benzene, toluene, ethylbenzene, and xylenes.

Table 2 shows the performance of the Hybrid Kriging-LUR, GWR-Hybrid LUR, RF-Hybrid LUR, and XGBoost-Hybrid LUR models. The XGBoost-Hybrid LUR better predicts the variation in all BTEXs, with a  $R^2$  value from 0.61 to 0.79. The Hybrid Kriging-LUR has the worst  $R^2$  value (from 0.37 to 0.52). Similar results as to  $R^2$  were obtained in adjusted  $R^2$  values (from 60 to 79 for the XGBoost-Hybrid LUR, which performs best; and 0.37 to 0.52 for the Hybrid Kriging-LUR, which performs worst).

**Table 2.** Performance of the Hybrid Kriging-LUR, GWR-Hybrid LUR, RF-Hybrid LUR and XGBoost-Hybrid LUR models. GWR—geographically weighted regression; LUR—Land-use regression; RF—random forest; XGBoost—extreme gradient boosting; RMSE—root mean square error.

BTEX	Statistic	Hybrid Kriging-LUR	GWR-Hybrid LUR	RF-Hybrid LUR	XGBoost-Hybrid LUR
Benzene	$R^2$ (training, testing)	0.45 (0.43, 0.55)	0.47 (0.46, 0.45)	0.57 (0.59, 0.42)	0.63 (0.65, 0.53)
	Adjusted $R^2$ (training, testing)	0.45 (0.42, 0.54)	0.47 (0.46, 0.44)	0.56 (0.59, 0.38)	0.63 (0.64, 0.50)
	RMSE (training, testing)	1.24 (1.29, 1.06)	1.22 (1.23, 0.44)	1.10 (1.11, 1.04)	1.02 (1.01, 1.03)
Toluene	$R^2$ (training, testing)	0.52 (0.52, 0.56)	0.54 (0.52, 0.60)	0.69 (0.70, 0.63)	0.72 (0.74, 0.60)
	Adjusted $R^2$ (training, testing)	0.52 (0.51, 0.56)	0.54 (0.52, 0.59)	0.68 (0.69, 0.59)	0.71 (0.73, 0.56)
	RMSE (training, testing)	1.35 (1.42, 1.10)	1.33 (1.32, 1.36)	1.09 (1.07, 1.16)	1.03 (1.03, 1.16)
Ethylbenzene	$R^2$ (training, testing)	0.37 (0.36, 0.49)	0.38 (0.31, 0.23)	0.50 (0.50, 0.45)	0.61 (0.62, 0.54)
	Adjusted $R^2$ (training, testing)	0.37 (0.34, 0.49)	0.38 (0.31, 0.22)	0.49 (0.49, 0.40)	0.61 (0.61, 0.50)
	RMSE (training, testing)	0.31 (0.33, 0.23)	0.31 (0.32, 0.17)	0.28 (0.29, 0.22)	0.60 (0.25, 0.22)
m,p-Xylene	$R^2$ (training, testing)	0.42 (0.42, 0.43)	0.44 (0.40, 0.29)	0.77 (0.77, 0.77)	0.79 (0.79, 0.79)
	Adjusted $R^2$ (training, testing)	0.42 (0.41, 0.42)	0.44 (0.40, 0.29)	0.77 (0.77, 0.77)	0.79 (0.79, 0.77)
	RMSE (training, testing)	0.70 (0.72, 0.67)	0.69 (0.72, 0.27)	0.44 (0.41, 0.44)	0.42 (0.36, 0.61)

The XGBoost-Hybrid LUR has the lowest RMSE (from 0.24 ppb to 1.03 ppb) and the Hybrid Kriging-LUR has the highest RMSE (from 0.31 ppb to 1.135 ppb). The adjusted  $R^2$  values are similar for the overfitting tests (Table 2). The respective adjusted  $R^2$  values for testing for the Hybrid Kriging-LUR, the GWR-Hybrid LUR, RF-Hybrid LUR, and the XGBoost-Hybrid LUR models are 0.34–0.56, 0.22–0.59, 0.38–0.77, and 0.50–0.79.

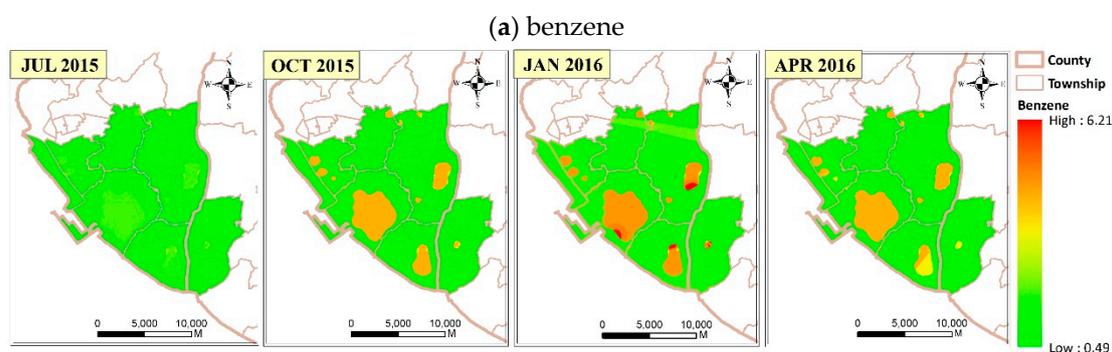
Observations from January to June in 2019 were used as external data to verify the robustness of the model (Table 3). The respective adjusted  $R^2$  values for the Hybrid Kriging-LUR, GWR-Hybrid LUR, the RF-Hybrid LUR, and the XGBoost-Hybrid LUR models are 0.34–0.65, 0.28–0.58, 0.42–0.56, and 0.41–0.55. This shows that even if the  $R^2$  value is reduced, these models still have a medium level of prediction performance. To validate the exposure estimates, we further conducted a 10-fold cross-validation to verify the model performance of the XGBoost-Hybrid LUR. 90% of the sites' data were randomly selected for model development, while the remaining 10% were used as out-of-data for model evaluation. This procedure was repeated ten times; thus, each monitoring site was used as a test data set for spatial verification. Similar  $R^2$  values with the main model (0.53 for benzene, 0.56 for toluene, 0.48 for ethylbenzene, and 0.59 m,p-xylene in Table S4) were obtained again to confirm the reliability of the developed model (0.41 for benzene, 0.55 for toluene, 0.45 for ethylbenzene, and 0.52 m,p-xylene in Table 3).

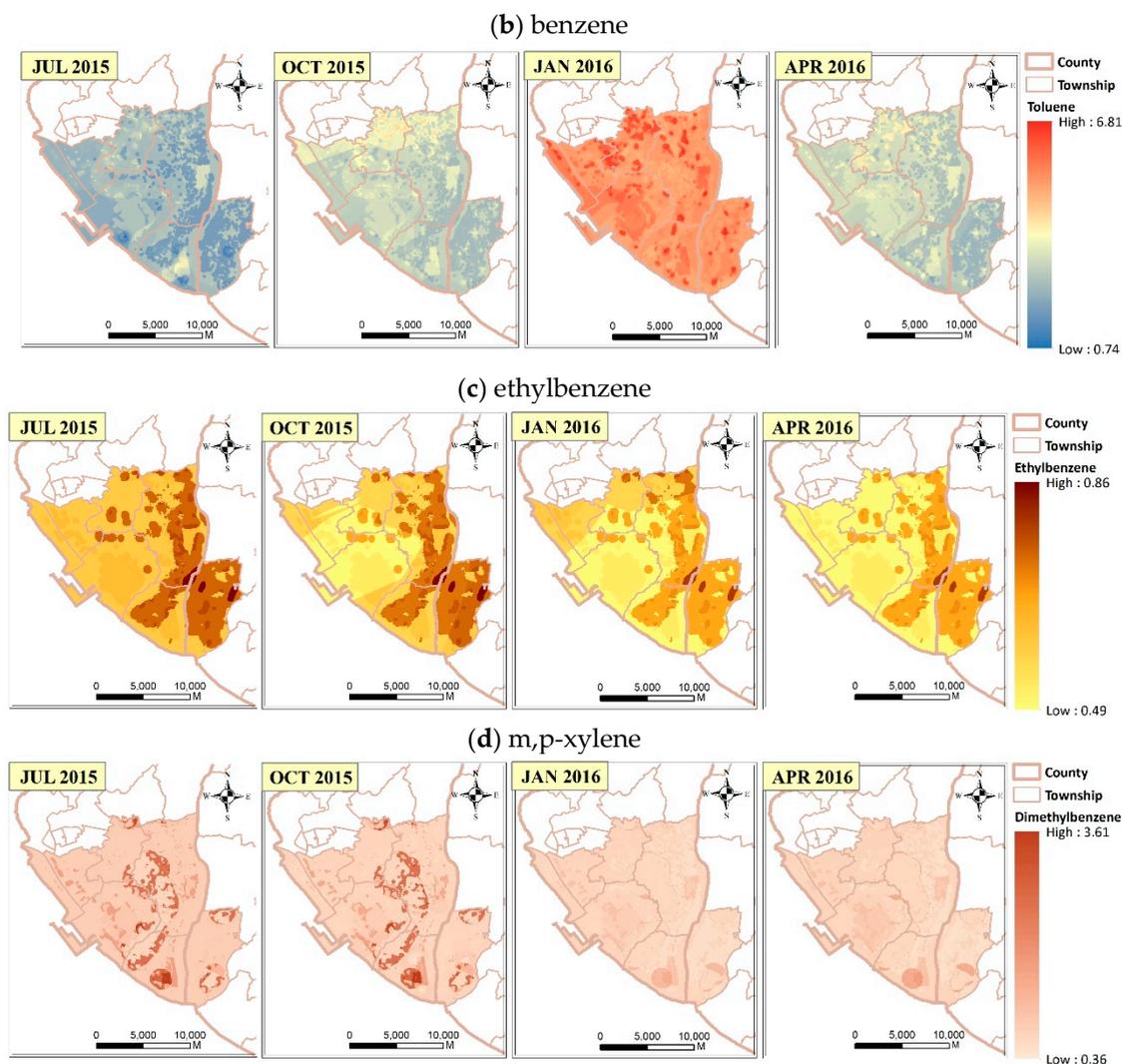
**Table 3.** External data validation for the proposed models.

BTEX	Statistic	Hybrid Kriging-LUR	GWR-Hybrid LUR	RF-Hybrid LUR	XGBoost-Hybrid LUR
Benzene	R <sup>2</sup>	0.52	0.52	0.44	0.41
	Adjusted R <sup>2</sup>	0.52	0.52	0.43	0.40
	RMSE	0.29	0.29	0.31	0.80
Toluene	R <sup>2</sup>	0.65	0.58	0.56	0.55
	Adjusted R <sup>2</sup>	0.64	0.58	0.55	0.54
	RMSE	0.81	0.88	0.90	0.91
Ethylbenzene	R <sup>2</sup>	0.47	0.43	0.42	0.45
	Adjusted R <sup>2</sup>	0.47	0.42	0.41	0.44
	RMSE	0.15	0.16	0.16	0.16
m,p-Xylene	R <sup>2</sup>	0.34	0.28	0.51	0.52
	Adjusted R <sup>2</sup>	0.34	0.27	0.51	0.52
	RMSE	0.24	0.25	0.23	0.19

### 3.3. Spatiotemporal Distribution of BTEX

By using the XGBoost-Hybrid LUR model for representative months from 2015 to 2016 (July, October, January, and April), Figure 2 shows the monthly average BTEX concentration through the study period. To begin with, the spatial variation in each season was relatively consistent, probably because the season and temperature factors were too insignificant to be selected into the models (Table 1). Second, there are higher benzene concentrations (light yellow to red color in Figure 2a) near industrial parks because of the higher partial R<sup>2</sup> for the factor of industry in the benzene model (Table 1). Third, as shown in Figure 2b, we see higher toluene levels scattering in places closer to the city or roads (red color in Jan 2010 in Figure 2b) because NO<sub>x</sub>, which is the major pollutant of traffic, had higher partial R<sup>2</sup> in the toluene model (Table 1). Fourth, higher ethylbenzene concentrations (dark brown color in Figure 2c) were shown in certain residential areas with many temples and near industrial parks because of the higher partial R<sup>2</sup> for the factor of both industry and temples in the ethylbenzene model. Fifth, higher m,p-xylene concentrations (dark brown color in Figure 2d) were seen also in certain residential areas with many temples because of the higher partial R<sup>2</sup> for the factor of temples in the m,p-xylene model.

**Figure 2.** Cont.



**Figure 2.** Monthly average concentration of BTEX: (a) benzene, (b) toluene, (c) ethylbenzene, and (d) m,p-xylene.

#### 4. Discussion

Most studies of exposure to ambient BTEX and health outcomes rely on daily monitoring of air pollution [40]. Few studies determine individual exposure levels using spatial analysis techniques [14,41,42]. These studies extrapolate actual measurements to individual exposures, so they do not reliably reflect the effect of air pollutants on health outcomes. This study proposes a method that is more economical than daily monitoring and more accurate than extrapolation to determine the effect of BTEX on health.

Four spatiotemporal models are used to predict monthly average BTEX concentration from 2015 to 2018 at a resolution of  $50 \times 50$  m. The models that combine Hybrid Kriging-LUR with machine learning (RF-Hybrid LUR and XGBoost-Hybrid LUR) have a greater predictive ability than the two regression models (Hybrid Kriging-LUR and GWR-Hybrid LUR). Specifically, the use of machine learning models in conjunction with land-use information increases the predictive power by 16% to 25% over that of the regression models. This increase is attributable to the fact that both the RF and XGBoost methods identify potential nonlinear associations between candidate predictors and ambient BTEX. To the authors' best knowledge, this is the first study to compare machine learning and standard linear regression models to predict spatial differences in ambient BTEX. It is shown that both machine learning models (RF and XGBoost method) have a greater predictive power than standard approaches.

XGBoost-Hybrid LUR is demonstrated to be the best model in this study and better explains the spatial variation in ambient BTEX in south Kaohsiung. The model also performs acceptably when verified using an external dataset. To address the problem of overfitting, 80% of data was used to train the XGBoost-Hybrid LUR model and 20% to test it. The adjusted  $R^2$  values for training and testing are similar to those for the original model so there is no overfitting problem.

The variables were selected using the Hybrid Kriging-LUR model and then used for the other three models developed by this study. Most of the significant variables are similar for the prediction of BTEX using the Hybrid Kriging-LUR model, but the individual contribution of each variable to the models is different. For example, while industrial area is a significant variable in predicting benzene, toluene, ethylbenzene, and m,p-xylenes, the significance of this variable (24%, 3%, 16%, and 8%, respectively) is different for each model (Table 1). This difference is probably caused by different levels of emissions of various compounds in each industrial area. It also highlights the need to consider different air pollutants when developing a prediction model [43].

When compared to other LUR models developed earlier in Sarnia, Ontario, Canada [43]; Toronto, Canada [44]; Dallas, Texas, USA [42]; Tehran, Iran [14]; Detroit/Dearborn, USA [45]; and New York City, USA [46]; this study selected both similar and different significant variables. For instance, Atari and Luginaah [43] reported that industrial area was the most significant factor for BTEX levels, while Smith et al. [42], Su et al. [44], and Amini et al. [14] suggested that traffic was the dominant factor for BTEX concentration. Mukerjee et al. [45] noted that both traffic and emission sources caused higher concentrations of BTEX. For this study, industrial area is the most significant factor for benzene and ethylbenzene. Table 1 also shows that traffic is the dominant factor for toluene concentration because 50% of  $\text{NO}_x$  across Taiwan and 85% of  $\text{NO}_x$  in cities is emitted from vehicles [47]. Sources of emissions that are specific to Asian culture, such as temples, are the dominant factor for m,p-xylene level because incense combustion significantly increases the concentration of m,p-xylene [48]. As strong solar radiation removes VOCs through the photochemical reactions [35], UV is a significant variable for the prediction of benzene. It is noteworthy that some greenness, such as rice farms and fruit orchards, can also increase BTEX levels.

While industry and traffic are often the dominant factors in the prediction of BTEX, some BTEX sources are specific to Asia. Going to a temple to pray and burn incense and joss paper is an important religious activity for many Asian households [49], and several studies have shown that this activity contributes to air pollution [50–52]. However, none of these used culturally specific variables to develop an LUR model to predict BTEX. This study uses the number of temples to reflect local emissions caused by the burning of joss paper and incense, which is a significant predictor for the proposed model. Future studies should also consider this unique local cultural source as a predictor of BTEX for developing LUR models in other Asian regions.

There are limitations to the selected predictors for this study. Traffic intensity is used by other studies to improve model performance [45,53] but is not used for this study because data are not generally available in Taiwan. In stand, we used  $\text{NO}_x$  as proxies for traffic because a great portion (50% to 80%) of  $\text{NO}_x$  is emitted from vehicles in Taiwan [47]. In contrast to data for a single year or an even shorter period, which is used by other studies, this study uses a much longer period (from 2015 to 2019) to represent spatial and temporal variations in compound concentrations. Using long-term pollutant data to establish an LUR model that is refined by machine learning and considering culturally specific predictors, this model has good prediction performance, which can be used to better depict the variation of BTEX in Asian cities.

## 5. Conclusions

Using machine learning algorithms to estimate individual levels of ambient air pollution is common practice. Combining a traditional LUR model and machine learning, this study develops Hybrid Kriging-LUR, GWR-Hybrid LUR, RF-Hybrid LUR, and XGBoost-Hybrid LUR models to predict BTEX concentrations. The study site is in Kaohsiung, Taiwan, where traffic, industrial area,

and temple are the main variables. Using data from seventeen measurement stations, this study shows that the machine learning LUR models (such as RF-Hybrid LUR and XGBoost-Hybrid LUR models) can better estimate fine spatial variability in long-term BTEX concentrations. This approach should be used in future studies to develop hybrid LUR models for other pollutants in Taiwan. In terms of residents' health or health effect indicators, the results of this study support the need for future air epidemiological studies of ambient BTEX.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1660-4601/17/19/6956/s1>, Figure S1: Overview of the sampling sites (1–17) and wind rose diagrams for the study periods of spring, summer, fall and winter, Table S1: Parameters proposed in hybrid-kriging LUR coupled with RF models, Table S2: Parameters proposed in hybrid-kriging LUR coupled with XGBoost models, Table S3: Parameters proposed in hybrid-kriging LUR coupled with XGBoost models, Table S4: Results of 10-fold cross-validation proposed in the XGBoost-Hybrid LUR model.

**Author Contributions:** Conceptualization, Y.-C.C., S.-C.C.L., and C.-D.W.; methodology, Y.-T.Z. and C.-D.W.; formal analysis, C.Y.H. and Y.-T.Z.; data curation, M.-J.C. and Y.-C.C.; writing—original draft preparation, C.-Y.H. and C.-D.W.; writing—review and editing, C.-Y.H., Y.-T.Z., Y.-C.C., M.-J.C., S.-C.C.L., and C.-D.W.; funding acquisition, C.-D.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the National Health Research Institutes, Taiwan (NHRI-109A1-EMCO-02202212).

**Acknowledgments:** We gratefully acknowledge the funding received from the National Health Research Institutes (NHRI-109A1-EMCO-02202212), Academia Sinica (AS-SS-107-03), and Ministry of Science and Technologies (MOST 109WFA0910475). This research was also supported in part by Higher Education Sprout Project, Ministry of Education to the Headquarters of University Advancement at National Cheng Kung University (NCKU). We also appreciate data supported from the Environmental Protection Administration, Ministry of Executive Yuan, Taiwan.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Heibati, B.; Pollitt, K.J.G.; Karimi, A.; Charati, J.Y.; Ducatman, A.; Shokrzadeh, M.; Mohammadyan, M. BTEX exposure assessment and quantitative risk assessment among petroleum product distributors. *Ecotoxicol. Environ. Saf.* **2017**, *144*, 445–449. [[CrossRef](#)]
2. Liu, J.P.; Yang, Y.; Xu, S.H.; Zhao, Y.Y.; Wang, Y.; Zhang, F.H. A Geographically Temporal Weighted Regression Approach with Travel Distance for House Price Estimation. *Entropy* **2016**, *18*, 303. [[CrossRef](#)]
3. Kennes, C.; Veiga, M.C. Fungal biocatalysts in the biofiltration of VOC-polluted air. *J. Biotechnol.* **2004**, *113*, 305–319. [[CrossRef](#)] [[PubMed](#)]
4. Yassaa, N.; Brancaleoni, E.; Frattoni, M.; Ciccioli, P. Isomeric analysis of BTEXs in the atmosphere using beta-cyclodextrin capillary chromatography coupled with thermal desorption and mass spectrometry. *Chemosphere* **2006**, *63*, 502–508. [[PubMed](#)]
5. Correa, S.M.; Arbilla, G.; Marques, M.R.C.; Oliveira, K.M.P.G. The impact of BTEX emissions from gas stations into the atmosphere. *Atmos. Pollut. Res.* **2012**, *3*, 163–169. [[CrossRef](#)]
6. Jiang, Z.H.; Gosselin, B.; Daele, V.; Mellouki, A.; Mu, Y.J. Seasonal and diurnal variations of BTEX compounds in the semi-urban environment of Orleans, France. *Sci. Total Environ.* **2017**, *574*, 1659–1664. [[CrossRef](#)] [[PubMed](#)]
7. Kountouriotis, A.; Aleiferis, P.G.; Charalambides, A.G. Numerical investigation of VOC levels in the area of petrol stations. *Sci. Total Environ.* **2014**, *470*, 1205–1224. [[CrossRef](#)]
8. IARC. Chemical agents and related occupations. *IARC Monogr. Eval. Carcinog Risks Hum.* **2012**, *100 Pt F*, 9–562.
9. Moolla, R.; Curtis, C.J.; Knight, J. Occupational Exposure of Diesel Station Workers to BTEX Compounds at a Bus Depot. *Int. J. Environ. Res. Public Health* **2015**, *12*, 4101–4115. [[CrossRef](#)]
10. Bolden, A.L.; Kwiatkowski, C.F.; Colborn, T. New Look at BTEX: Are Ambient Levels a Problem? *Environ. Sci. Technol.* **2015**, *49*, 5261–5276.
11. Mo, Z.W.; Shao, M.; Lu, S.H.; Qu, H.; Zhou, M.Y.; Sun, J.; Gou, B. Process-specific emission characteristics of volatile organic compounds (VOCs) from petrochemical facilities in the Yangtze River Delta, China. *Sci. Total Environ.* **2015**, *533*, 422–431. [[CrossRef](#)] [[PubMed](#)]

12. Yu, C.L.; Wang, S.F.; Pan, P.C.; Wu, M.T.; Ho, C.K.; Smith, T.J.; Li, Y.; Pothier, L.; Christiani, D.C. Residential exposure to petrochemicals and the risk of leukemia: Using Geographic Information System tools to estimate individual-level residential exposure. *Am. J. Epidemiol.* **2006**, *164*, 200–207. [[CrossRef](#)]
13. Liu, Y.; Shao, M.; Fu, L.; Lu, S.; Zeng, L.; Tang, D. Source profiles of volatile organic compounds (VOCs) measured in China: Part I. *Atmos. Environ.* **2008**, *42*, 6247–6260. [[CrossRef](#)]
14. Amini, H.; Schindler, C.; Hossein, V.; Yunesian, M.; Kunzli, N. Land Use Regression Models for Alkylbenzenes in a Middle Eastern Megacity: Tehran Study of Exposure Prediction for Environmental Health Research (Tehran SEPEHR). *Environ. Sci. Technol.* **2017**, *51*, 8481–8490. [[CrossRef](#)]
15. Gaeta, A.; Cattani, G.; di Bucchianico, A.D.; De Santis, A.; Cesaroni, G.; Badaloni, C.; Ancona, C.; Forastiere, F.; Sozzi, R.; Bolignano, A.; et al. Development of nitrogen dioxide and volatile organic compounds land use regression models to estimate air pollution exposure near an Italian airport. *Atmos. Environ.* **2016**, *131*, 254–262. [[CrossRef](#)]
16. Poirier, A.; Dodds, L.; Dummer, T.; Rainham, D.; Maguire, B.; Johnson, M. Maternal Exposure to Air Pollution and Adverse Birth Outcomes in Halifax, Nova Scotia. *J. Occup. Environ. Med.* **2015**, *57*, 1291–1298. [[CrossRef](#)] [[PubMed](#)]
17. Wu, C.D.; Zeng, Y.T.; Lung, S.C.C. A hybrid kriging/land-use regression model to assess PM<sub>2.5</sub> spatial-temporal variability. *Sci. Total Environ.* **2018**, *645*, 1456–1464. [[CrossRef](#)] [[PubMed](#)]
18. McMillen, D.P. Geographically weighted regression: The analysis of spatially varying relationships. *Am. J. Agric. Econ.* **2004**, *86*, 554–556. [[CrossRef](#)]
19. Joharestani, M.Z.; Cao, C.X.; Ni, X.L.; Bashir, B.; Talebiesfandarani, S. PM<sub>2.5</sub> Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere* **2019**, *10*, 373. [[CrossRef](#)]
20. Ma, J.H.; Yu, Z.Q.; Qu, Y.H.; Xu, J.M.; Cao, Y. Application of the XGBoost Machine Learning Method in PM<sub>2.5</sub> Prediction: A Case Study of Shanghai. *Aerosol Air Qual. Res.* **2020**, *20*, 128–138. [[CrossRef](#)]
21. Araki, S.; Shima, M.; Yamamoto, K. Spatiotemporal land use random forest model for estimating metropolitan NO<sub>2</sub> exposure in Japan. *Sci. Total Environ.* **2018**, *634*, 1269–1277. [[CrossRef](#)]
22. Löfroth, G.; Stensman, C.; Brandhorstsatzkorn, M. Indoor Sources of Mutagenic Aerosol Particulate Matter—Smoking, Cooking and Incense Burning. *Mutat. Res.* **1991**, *261*, 21–28. [[CrossRef](#)]
23. Tranfo, G.; Pignini, D.; Paci, E.; Bauleo, L.; Forastiere, F.; Ancona, C. Biomonitoring of Urinary Benzene Metabolite SPMA in the General Population in Central Italy. *Toxics* **2018**, *6*, 37. [[CrossRef](#)] [[PubMed](#)]
24. Alghamdi, M.A.; Khoder, M.; Abdelmaksoud, A.S.; Harrison, R.M.; Hussein, T.; Lihavainen, H.; Al-Jeelani, H.; Goknil, M.H.; Shabbaj, I.I.; Almeahmadi, F.M.; et al. Seasonal and diurnal variations of BTEX and their potential for ozone formation in the urban background atmosphere of the coastal city Jeddah, Saudi Arabia. *Air Qual. Atmos. Health* **2014**, *7*, 467–480. [[CrossRef](#)]
25. Wang, M.; Zhu, T.; Zheng, J.; Zhang, R.Y.; Zhang, S.Q.; Xie, X.X.; Han, Y.Q.; Li, Y. Use of a mobile laboratory to evaluate changes in on-road air pollutants during the Beijing 2008 Summer Olympics. *Atmos. Chem. Phys.* **2009**, *9*, 8247–8263. [[CrossRef](#)]
26. Hsu, C.Y.; Wu, C.D.; Hsiao, Y.P.; Chen, Y.C.; Chen, M.J.; Lung, S.C.C. Developing Land-Use Regression Models to Estimate PM<sub>2.5</sub>-Bound Compound Concentrations. *Remote Sens.* **2018**, *10*, 1971. [[CrossRef](#)]
27. Hsu, C.Y.; Wu, J.Y.; Chen, Y.C.; Chen, N.T.; Chen, M.J.; Pan, W.C.; Lung, S.C.C.; Guo, Y.L.; Wu, C.D. Asian Culturally Specific Predictors in a Large-Scale Land Use Regression Model to Predict Spatial-Temporal Variability of Ozone Concentration. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1300. [[CrossRef](#)]
28. Wu, C.D.; Cheng, C.C.; Lo, H.C.; Chen, Y.K. Study on estimating the evapotranspiration cover coefficient for stream flow simulation through remote sensing techniques. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 225–232. [[CrossRef](#)]
29. Wu, C.D.; Chen, Y.C.; Pan, W.C.; Zeng, Y.T.; Chen, M.J.; Guo, Y.L.; Lung, S.C.C. Land Use Regression with Long-term Satellite based Greenness Index and Culture-Specific Sources to Model PM<sub>2.5</sub> Spatial-Temporal Variability. *Environ. Pollut.* **2017**, *224*, 148–157.
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Chen, T.; Guestrin, C. *Xgboost: A Scalable Tree Boosting System*; Cornell Univ.: Ithaca, NY, USA, 2016; pp. 785–794.

32. Li, L.; Li, H.; Zhang, X.M.; Wang, L.; Xu, L.H.; Wang, X.Z.; Yu, Y.T.; Zhang, Y.J.; Cao, G. Pollution characteristics and health risk assessment of benzene homologues in ambient air in the northeastern urban area of Beijing, China. *J. Environ. Sci. (China)* **2014**, *26*, 214–223. [[CrossRef](#)]
33. Zhou, J.A.; You, Y.; Bai, Z.P.; Hu, Y.D.; Zhang, J.F.; Zhang, N. Health risk assessment of personal inhalation exposure to volatile organic compounds in Tianjin, China. *Sci. Total Environ.* **2011**, *409*, 452–459. [[CrossRef](#)] [[PubMed](#)]
34. Hsu, C.Y.; Chiang, H.C.; Shie, R.H.; Ku, C.H.; Lin, T.Y.; Chen, M.J.; Chen, N.T.; Chen, Y.C. Ambient VOCs in residential areas near a large-scale petrochemical complex: Spatiotemporal variation, source apportionment and health risk. *Environ. Pollut.* **2018**, *240*, 95–104. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, Y.C.; Li, R.; Fu, H.B.; Zhou, D.; Chen, J.M. Observation and analysis of atmospheric volatile organic compounds in a typical petrochemical area in Yangtze River Delta, China. *J. Environ. Sci. (China)* **2018**, *71*, 233–248. [[CrossRef](#)]
36. Zhang, Y.J.; Mu, Y.J.; Liu, J.F.; Mellouki, A. Levels, sources and health risks of carbonyls and BTEX in the ambient air of Beijing, China. *J. Environ. Sci. (China)* **2012**, *24*, 124–130. [[CrossRef](#)]
37. Na, K.; Kim, Y.P.; Moon, K.C. Diurnal characteristics of volatile organic compounds in the Seoul atmosphere. *Atmos. Environ.* **2003**, *37*, 733–742. [[CrossRef](#)]
38. Borbon, A.; Locoge, N.; Veillerot, M.; Galloo, J.C.; Guillermo, R. Characterisation of NMHCs in a French urban atmosphere: Overview of the main sources. *Sci. Total Environ.* **2002**, *292*, 177–191. [[CrossRef](#)]
39. Tang, J.H.; Chan, L.Y.; Chan, C.Y.; Li, Y.S.; Chang, C.C.; Liu, S.C.; Wu, D.; Li, Y.D. Characteristics and diurnal variations of NMHCs at urban, suburban, and rural sites in the Pearl River Delta and a remote site in South China. *Atmos. Environ.* **2007**, *41*, 8620–8632. [[CrossRef](#)]
40. Aguilera, I.; Sunyer, J.; Fernandez-Patier, R.; Hoek, G.; Aguirre-Alfaro, A.; Meliefste, K.; Bomboi-Mingarro, M.T.; Nieuwenhuijsen, M.J.; Herce-Garraleta, D.; Brunekreef, B. Estimation of outdoor NO<sub>x</sub>, NO<sub>2</sub>, and BTEX exposure in a cohort of pregnant women using land use regression modeling. *Environ. Sci. Technol.* **2008**, *42*, 815–821. [[CrossRef](#)]
41. Mukerjee, S.; Smith, L.; Neas, L.; Norris, G. Evaluation of Land Use Regression Models for Nitrogen Dioxide and Benzene in Four US Cities. *Sci. World J.* **2012**, *2012*, 1–8. [[CrossRef](#)]
42. Smith, L.A.; Mukerjee, S.; Chung, K.C.; Afghani, J. Spatial analysis and land use regression of VOCs and NO<sub>2</sub> in Dallas, Texas during two seasons. *J. Environ. Monit.* **2011**, *13*, 999–1007. [[PubMed](#)]
43. Atari, D.O.; Luginaah, I.N. Assessing the distribution of volatile organic compounds using land use regression in Sarnia, “Chemical Valley”, Ontario, Canada. *Environ. Health* **2009**, *8*, 16. [[PubMed](#)]
44. Su, J.G.; Jerrett, M.; Beckerman, B.; Verma, D.; Arain, M.A.; Kanaroglou, P.; Stieb, D.; Finkelstein, M.; Brook, J. A land use regression model for predicting ambient volatile organic compound concentrations in Toronto, Canada. *Atmos. Environ.* **2010**, *44*, 3529–3537.
45. Mukerjee, S.; Smith, L.A.; Johnson, M.M.; Neas, L.M.; Stallings, C.A. Spatial analysis and land use regression of VOCs and NO<sub>2</sub> from school-based urban air monitoring in Detroit/Dearborn, USA. *Sci. Total Environ.* **2009**, *407*, 4642–4651. [[PubMed](#)]
46. Kheirbek, I.; Johnson, S.; Ross, Z.; Pezeshki, G.; Ito, K.; Eisl, H.; Matte, T. Spatial variability in levels of benzene, formaldehyde, and total benzene, toluene, ethylbenzene and xylenes in New York City: A land-use regression study. *Environ. Health* **2012**, *11*, 51. [[PubMed](#)]
47. TWEPA. Taiwan Emission Data System (TEDS), Version 8.1. 2015. Available online: [http://teds.epa.gov.tw/new\\_main2-0-1.htm](http://teds.epa.gov.tw/new_main2-0-1.htm) (accessed on 15 June 2020).
48. Lee, S.C.; Wang, B. Characteristics of emissions of air pollutants from burning of incense in a large environmental chamber. *Atmos. Environ.* **2004**, *38*, 941–951.
49. Lui, K.H.; Bandowe, B.A.M.; Ho, S.S.H.; Chuang, H.C.; Cao, J.J.; Chuang, K.J.; Lee, S.C.; Hu, D.; Ho, K.F. Characterization of chemical components and bioreactivity of fine particulate matter (PM<sub>2.5</sub>) during incense burning. *Environ. Pollut.* **2016**, *213*, 524–532.
50. Lung, S.C.C.; Hsiao, P.K.; Wen, T.Y.; Liu, C.H.; Fu, C.B.; Cheng, Y.T. Variability of intra-urban exposure to particulate matter and CO from Asian-type community pollution sources. *Atmos. Environ.* **2014**, *83*, 6–13.
51. Wang, B.; Lee, S.C.; Ho, K.F.; Kang, Y.M. Characteristics of emissions of air pollutants from burning of incense in temples, Hong Kong. *Sci. Total Environ.* **2007**, *377*, 52–60.

52. Navasumrit, P.; Arayasiri, M.; Hiang, O.M.T.; Leechawengwongs, M.; Promvijit, J.; Choonvisase, S.; Chantchaemsai, S.; Nakngam, N.; Mahidol, C.; Ruchirawat, M. Potential health effects of exposure to carcinogenic compounds in incense smoke in temple workers. *Chem. Biol. Interact.* **2008**, *173*, 19–31.
53. Johnson, M.; Isakov, V.; Touma, J.S.; Mukerjee, S.; Ozkaynak, H. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmos. Environ.* **2010**, *44*, 3660–3668. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).