Machine learning and magnetic parameters to monitor potentially toxic elements in urban road dust of Mexico City

Aprendizaje automatizado y parámetros magnéticos para monitorear elementos potencialmente tóxicos en polvo de vialidades urbanas de la Ciudad de México

Ruben **Cejudo**[1,*], Francisco **Bautista Zuñiga**[2], Jaime **Urrutia-Fucugauchi**[3], Avto **Goguitchaichvili**[1], Vadim **Kravchinsky**[4], Patricia **Quintana Owen**[5], Daniel **Aguilar**[6]

[1] Instituto de Geofísica-UNAM campus Morelia, Antigua Carretera a Pátzcuaro 8701 Col. San José de la Huerta C.P 58190, Morelia, Mexico.

[2] Centro de Investigaciones en Geografía Ambiental-UNAM campus Morelia, Antigua Carretera a Pátzcuaro 8701, Col. San José de la Huerta C.P 58190, Morelia, Mexico.

[3] Instituto de Geofísica-UNAM, Circuito de la Investigación Científica s/n, Ciudad Universitaria, Alcaldía Coyoacán. C.P. 04510. Ciudad de México.

[4] Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2E1.

[5] CINVESTAV Mérida, Km. 6 Antigua carretera a Progreso Apdo. Postal 73, Cordemex, 97310, Mérida, Yuc., México.

* Corresponding author: (R. Cejudo)
ruben@igeofisica.unam.mx

How to cite this article:
Cejudo, R., Bautista Zuñiga, F., Urrutia-Fucugauchi, J., Goguitchaichvili A., Kravchinsky, V., Quintana Owen, P., Aguilar, D., 2024, Machine learning and magnetic parameters to monitor potentially toxic elements in urban road dust of Mexico City: Boletín de la Sociedad Geológica Mexicana, 76 (3), A290224. http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Manuscript received: October 26, 2023
Corrected manuscript received: February 20, 2024
Manuscript accepted: February 28, 2024

Peer Reviewing under the responsibility of Universidad Nacional Autónoma de México.

ABSTRACT

The high concentration of potentially toxic elements in urban road dust has been commonly associated with serious health problems. However, there is no systematic method for monitoring toxic elements in road dust of metropolitan areas. Here we develop a mathematical model based on machine learning algorithms using magnetic parameters (specific susceptibility, frequency-dependent susceptibility and saturation isothermal remanent magnetization) for proxy monitoring of toxic elements. Machine learning algorithms included artificial neural networks, classification trees and linear regression model in order to analyse up to 140 urban road dust samples collected in the metropolitan area of Mexico City. The study is intended to determine their ability to predict the concentrations of chromium, copper, lead, vanadium and zinc in urban road dust and evaluate the pollutant load index. We found that the algorithm based on artificial neural networks is a 20% better estimator of the concentration of toxic elements, compared to the decision tree and multiple linear regression models. The use of mathematical models based on neural networks and magnetic parameters is proposed to design a proxy monitoring system for potentially toxic elements in urban road dust.

Keywords: pollution load index, artificial neural network, magnetic parameters, urban road dust, Mexico City.

RESUMEN

La alta concentración de elementos potencialmente tóxicos en el polvo de carreteras urbanas se ha asociado con graves problemas de salud. Sin embargo, no existe un método sistemático de monitoreo de elementos tóxicos en polvo sedimentado de carreteras en zonas metropolitanas. Aquí desarrollamos un modelo matemático basado en algoritmos de aprendizaje automático, empleando parámetros magnéticos (susceptibilidad específica, susceptibilidad dependiente de la frecuencia y magnetización remanente isotérmica) para el monitoreo proxy de elementos tóxicos.
Se aplicaron algoritmos de aprendizaje automatizado: redes neuronales artificiales, árboles de clasificación y modelo de regresión lineal para el análisis de 140 muestras de polvo vial, recolectadas en el área metropolitana de la Ciudad de México, para determinar su capacidad de predecir las concentraciones de cromo, cobre, plomo, vanadio y zinc en polvo de carretera y evaluar el índice de carga contaminante. Encontramos que el algoritmo basado en redes neuronales artificiales es 20% mejor estimador de la concentración de elementos tóxicos, en comparación con los modelos de árbol de decisión y regresión lineal múltiple. Se propone el uso de modelos matemáticos basados en redes neuronales y parámetros magnéticos para diseñar un sistema de monitoreo proxy de elementos potencialmente tóxicos en polvo urbano.

Palabras clave: índice de carga contaminante, redes neuronales artificiales, parámetros magnéticos, polvo urbano, Ciudad de México.

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

1. Introduction

The high concentrations of potentially toxic elements (PTEs) in the urban road dust of big cities are becoming a serious health concern since it may trigger the occurrence of degenerative diseases (heart and cerebrovascular issues, among others; Tchounwou *et al.*, 2012; Briffa *et al.*, 2020). The emissions of particulate matter from natural sources (forest fire, floods, erosion of soil, volcanic eruption and others) and anthropic (particulate emissions from motor vehicles, industrial, mining and domestic activities) contain a variety of PTEs. The continued deposition of particulate matter with PTEs over the city causes an increase in the concentration of elements in urban soils and dust on roads, which can reach toxic levels (Ihl *et al.*, 2015; Qing *et al.*, 2015; Covarrubias and Peña, 2017).

Several cities around the world have reported high concentration levels of chromium (Cr), copper (Cu), lead (Pb), vanadium (V), and zinc (Zn) in soils, plants, urban sediments and street dust, so PTEs exposure in the population is almost unavoidable (Ihl *et al.*, 2015; Bautista *et al.*, 2017; Chávez-Gómez *et al.*, 2017; Delgado *et al.*, 2019). Effective monitoring systems of toxic elements in an urban zone are needed, and available monitoring methods are expensive and generate toxic waste and multiple environmental problems, such as the inductively coupled plasma mass spectrometry (ICP-MS) or atomic absorption spectroscopy (AAS). For this reason, it is necessary to develop reliable, environmentally safe alternatives (Faciu *et*

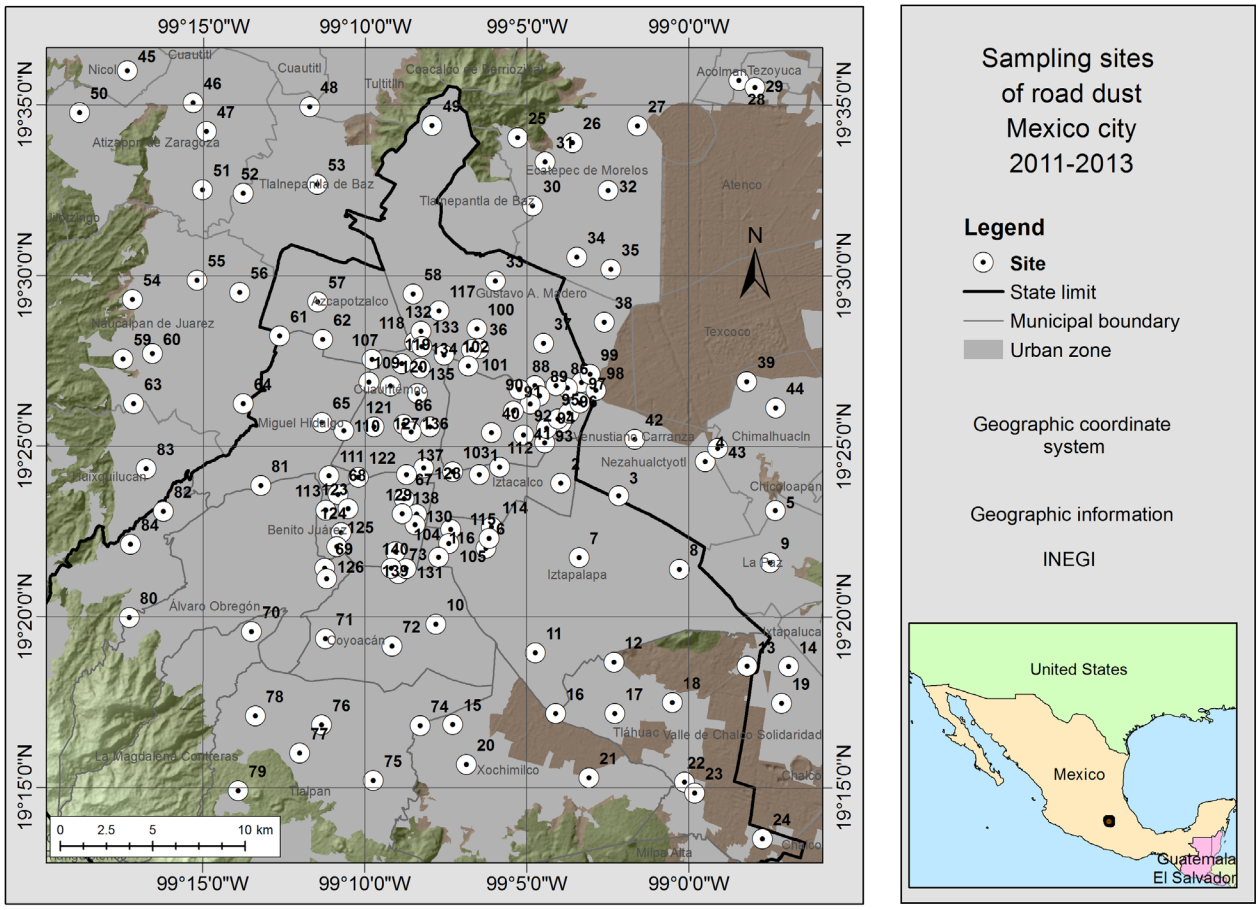


Figure 1 Sampling sites within the Mexico City metropolitan area (2011 and 2013 sample collection campaigns).

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

al., 2012; Jones *et al.*, 2014; Legarreta-Perusquia *et al.*, 2016).

Studies on magnetic parameters as proxies for estimating the concentration of PTEs became a quite routine approach. Typically, the properties of ferrimagnetic minerals of absorbing and adsorbing PTEs ions, abundant in the contamination emissions, are studied worldwide (Morton-Bermea *et al.*, 2009; Qian *et al.*, 2011; Zdoroveyshchev *et al.*, 2019; Chaparro *et al.*, 2020). Linear regression, multivariant analyses and artificial neural networks are the mathematical models usually used to estimate the relationship of magnetic parameters with the concentration of PTEs (Morton-Bermea *et al.*, 2009; Aguilar *et al.*, 2012; Zhang *et al.*, 2017, 2020).

In this study, we evaluate the predictive capacity of mathematical models, such as multiple linear regression, classification trees and artificial neural network, to classify and estimate the concentration of PTEs through the pollution load index in the studied samples of street dust.

2. Materials and Methods

The study analyzes 140 samples of urban street dust from different localities in Mexico City (Figure 1). The sampling was conducted in April 2011 (84 samples) and in 2013 (56 samples). Each sample was collected from a surface of 1 m^2 over the road and below the sidewalk. The material was obtained with a brush and plastic non-magnetic trowels and placed inside plastic bags. Eight grams of sieved material was placed inside a cubic plastic box for magnetic analyses, and 10 g of sieved material was used pellets for X-Ray Fluorescence (XRF) analysis.

The magnetic susceptibility (MS) was measured at low and high frequencies with MS3/MS2B dual sensor Bartington MS meter. Magnetic parameters of specific susceptibility (Equation 1).

$$\chi_{lf} = \kappa_{LF}/\rho \quad (1)$$

Where ρ is density in m^3/kg, and percentage of frequency-dependent susceptibility χ_{fd}% were computed (Dearing, 1999; Equation 2).

$$\chi_{fd}\% = \frac{\kappa_{LF} - \kappa_{HF}}{\kappa_{LF}} 100\% \quad (2)$$

Where k_{LF} is the low-frequency MS, and k_{HF} is the high-frequency MS (Dearing, 1999).

To obtain the saturation isothermal remanent magnetization (SIRM) value, the sample was subjected to a 1000 mT magnetic field pulse using an ASC-Scientific IM-10. The measurement of the remanent magnetization was then performed (RM_{1000}) with a JR-6 spinner magnetometer. The SIRM was obtained using Equation 3:

$$SIRM = RM_{1000}/\rho \quad (3)$$

Where, ρ is density in m^3/kg and RM_{1000} is the value of remanent magnetization to 1000 mT (Evans and Heller, 2003).

The concentrations of chromium (Cr), copper (Cu), lead (Pb), vanadium (V) and zinc (Zn) were determined in pellets with a hand-held Skyray Genius 7000 XRF spectrometer. A local threshold value for each element was applied in the absence of a threshold value in the element concentration in urban street dust (SEMARNAT, 2007). The pollution load index (PLI) for Cr, Cu, Pb, V and Zn was determined using the method of Tomlinson *et al.* (1980; Equation 4).

$$PLI = \sqrt[n]{\prod_{i=1}^{n}\left(\frac{Cm_i}{CF_i}\right)} \quad (4)$$

Where, Cm_i = mean concentration of i's element; CF_i = Background concentration of i's element, n represents the number of elements analyzed.

We proposed the value of PLI fell into three pollution levels: PLI ≤ 1 unpolluted, PLI ϵ (1, 2) mildly polluted, and PLI > 2 highly polluted (Thomlison *et al.*, 1980, Wang *et al.*, 2019).

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

2.1. MATHEMATICAL MODELS

The database obtained of magnetic parameters and PLI was divided into two subsets: (1) the so-called testing database (70 % of data) for predictive models, including multiple linear regression (MLR_{PLI}), artificial neural network (ANN) and classification tree (CT), and (2) the validation database (30 % of data) to determine the PLI value from magnetic parameters.

2.2. LINEAR REGRESSION MODEL

A multiple linear regression model was constructed with magnetic parameters to obtain the Pollution Load Index (Equation 5)

$$MRL_{PLI} = \left(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\right) + e \qquad (5)$$

Where, β_0 is the interceptor while β_1, β_2 and β_3 are the correlation coefficients of magnetic, and e is the error of the estimation (Rodriguez del Águila and Benítes-Parejo, 2011; Sabogal *et al.*, 2015).

2.3. ARTIFICIAL NEURAL NETWORK MODEL

The multi-layer perceptron-based artificial neural network was used to predict the PLI values (Y) for magnetic parameters χ_{lf}, χ_{fd}% and SIRM. All values of input were normalized with (Equation 6):

$$\chi_{lfnorm} = \frac{\chi_{lf(i\,med)} - 2.08}{7.47}, \chi_{fd\%norm} = \frac{\chi_{fd\%(i\,med)}}{4.6}, SIRM_{norm} = \frac{MRIS_{(i\,med)} - 31.03}{99.42} \qquad (6)$$

Where, $\chi_{lf\ (i\ med)}$, $\chi_{fd}\%_{\ (i\ med)}$ and $SIRM_{\ (i\ med)}$ are magnetic parameters measured in urban road dust of each site i. This data was used to activate specific network function (w) with a bias or threshold expressed as b in Equation 7. The equation describes the sigmoid activation function, which transforms the input signal so that it varies between 0 and 1 through the application of a non-linear response ($\varphi = 1/(1+e^{-t})$):

$$Y_i = \varphi\left(\sum_{i=1}^{N}\left(w_{ij}x_i + b_{ij}\right)\right) \qquad (7)$$

Where, x_i - value of the input, w_{ij} - propagation rule of each input in the hidden neuron j, b is the weight of the adaptive neuron (Trujillano *et al.*, 2004; Wickham and Grolemund, 2016; Alizamir and Sobhanardakani, 2018; Cejudo *et al.*, 2021).

2.4. CLASSIFICATION TREE

A classification tree model was constructed by a supervised learning method to create a model that predicts the value of PLI in terms of discrete variables (PLI ≤ 1 - no contaminant, and PLI >1 - contaminant is present). Magnetic parameters were used as input data.

The Classification Tree algorithm, also known as the Classification and Regression Tree (CART), is based on a binary division (R_1 and R_2) of the data through the threshold values at a cut-off value (s) of measured variables ($s(x_j)$). The division of sets is given as Equation 8.

$$R_1(j,s) = \left\{x \vee x_i < s\right\} and\ R_2 = \left\{X \vee x_j \geq s\right\} \qquad (8)$$

Where, x_i represents lower values with respect to the threshold value s (PLI ≤1), and x_j are the values higher than s (PLI>1).

The homogeneity of partition was evaluated using the Gini index (GI; Equation 9), which evaluates the probabilities considering classification categories at each node. The values of GI = 0 indicate that the data are classified. If GI ≠ 0, then the node presents mixed data requiring another division.

$$GI = 1 - \sum_{i=1}^{n}(P_i)^2 \qquad (9)$$

Where P_i is the probability of an element is of class i.

The process is iterated until a threshold value of GI = 0 or GI ≈ 0 is reached (Trujillano *et al.*, 2004; Krzywinski and Altman, 2017).

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

The identification of a reliable predictive model was made using a confusion matrix. For such purpose, we compared the PLI values computed directly from the data (real PLI) with the PLI values estimated from the different predictive models MRL, ANN, and CT. The difference between PLI estimate and PLI calculated (PLIreal) is less than 25 % for acceptable values for this proxy method. The statistical analyses and the algorithm of prediction models were performed using the R programming software with the RStudio version 1.3.1093 and the packages MASS, neuralnet, ggplot2 (ANN), caret, rparta and rpart.plot (AC).

3. Results and discussion

Table 1 presents the results of the statistical analyses of magnetic parameters and the concentration of PTEs for 140 samples of urban road dust in Mexico City. The value of the first quartile (Q1) of residential sites (samples of 2011) was used as the background value (BV) to calculate the PLI value for each sample (Table 1).

The mean values of urban road dust magnetic parameters for 2011 collection samples are χ_{lf} = 5.6±1.9 μm^3/kg; χ_{fd}% = 1.7±1.2 %, and SIRM = 73.4±20.7 mA m^2/kg. The sample collection 2013 is characterized by χ_{lf} = 4.0±1.1 μm^3/kg; χ_{fd}% = 2.9±1.6 %, and SIRM = 55.3±17.2 mA m^2/kg. The Q1 value is characterized by χ_{lf} = 4.4 μm^3/kg; χ_{fd}% = 0.9 %, and SIRM = 54.9 mA m^2/kg (Table 1). The 2011 urban road dust data shows higher average values of χ_{lf} and SIRM compared to the urban road dust values of 2013, which indicates the diminution of ferrimagnetic mineral contribution in the dust. On the other hand, the average of χ_{fd}% showed an increase in 2013 compared to 2011. This is a possible indicator of the growth of the concentration of ultrafine superparamagnetic minerals (<50 nm) in urban road dust.

Lower magnetic parameter values in urban road dust indicate a change in the ferrimagnetic mineral emission source. The increment of ultra-

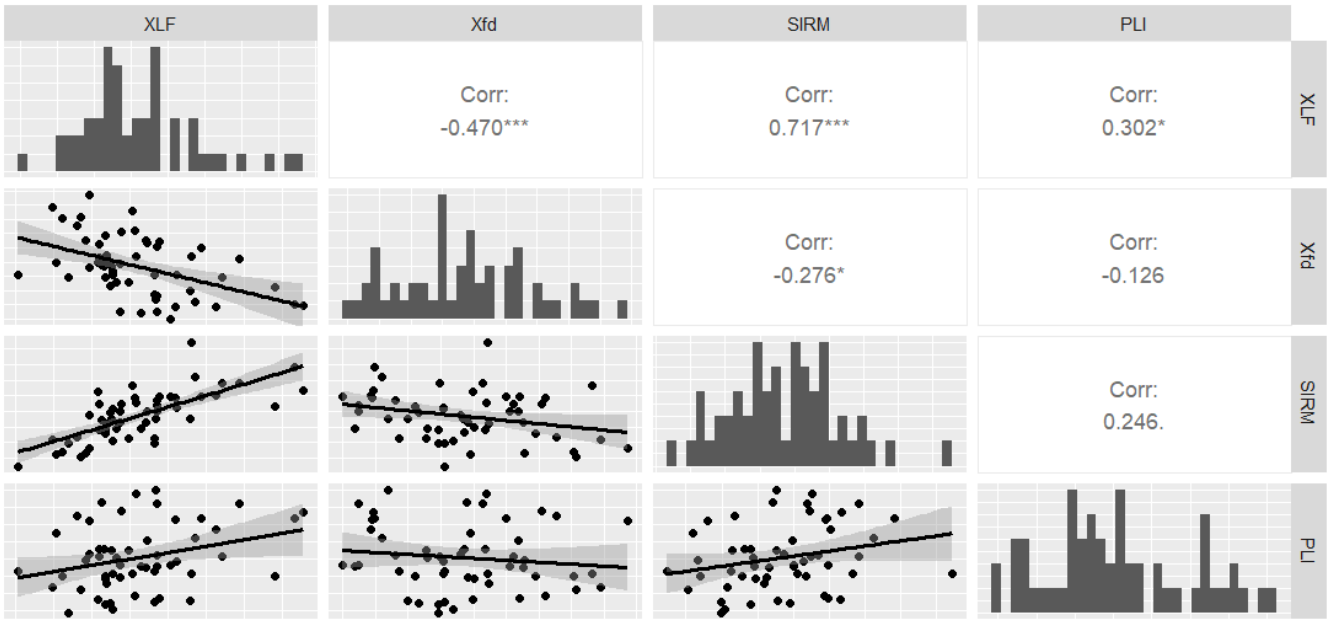
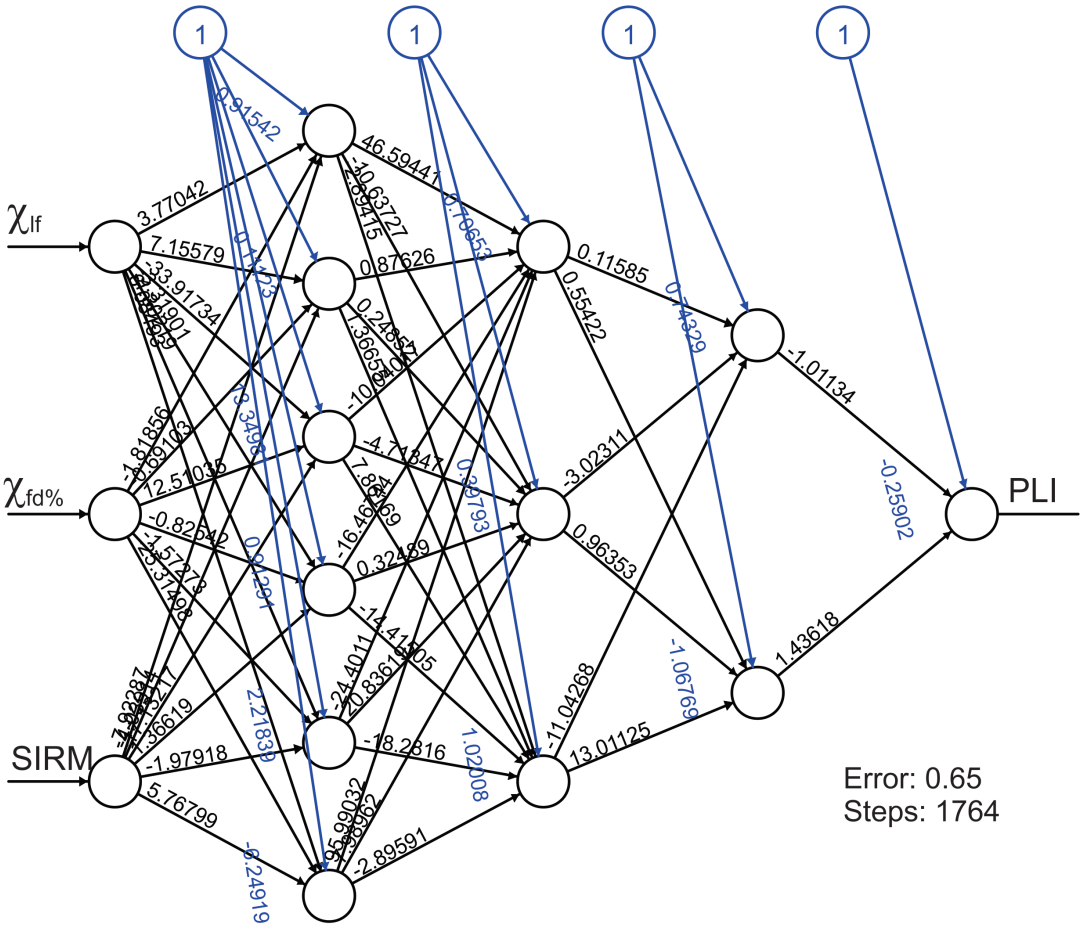


Figure 2 Linear correlation between magnetic parameters and pollution load index of urban road dust in Mexico City (test data from the 2011 sampling are in Table 2). The distribution of each variable is shown on the diagonal. On the bottom of the diagonal: the bivariate scatter plots with a fitted line are displayed. On the top of the diagonal: the value of the correlation plus the significance level as stars. Level of significance. * = 0.05, ** = 0.01, *** = 0.001

http://dx.doi.org/10.18268/BSGM2024v76n3a290224



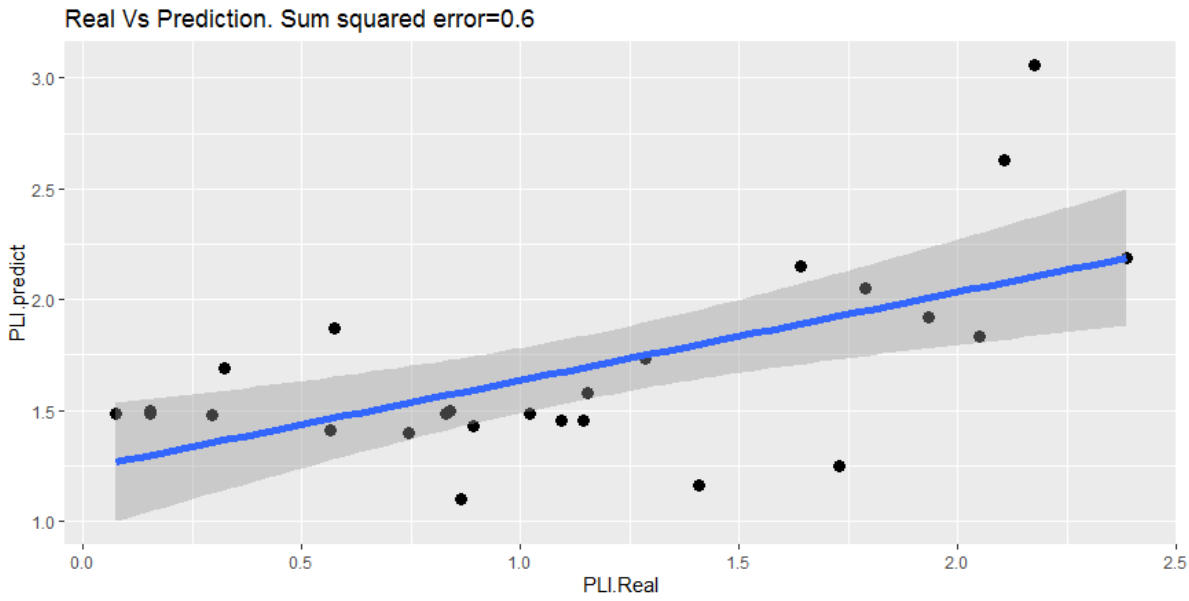


Figure 3 a) Architecture of artificial neural network for the PLI prediction (PLI_{ANN}) value of magnetic parameters in urban road dust, b) Cross-validation of PLI_{ANN} vs PLI_{real} for test data from 2011 collection.

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

fine particles is typically associated with industrial emission sources and fossil combustibles (Menshov *et al.*, 2020).

The arithmetic mean elemental concentration values and standard deviation (A.M.±S.D.) for 2011 are $A.M._{Cr}$ = 113±33 mg/kg, $A.M._{Cu}$ = 94±59 mg/kg, $A.M._{Pb}$ = 203±130 mg/kg, $A.M._{V}$ = 87±14 mg/kg, and $A.M._{Zn}$ = 309±126 mg/kg. The recorded mean concentration of the elements was in a decreasing order Zn>Pb>Cr>Cu>V. Mean concentrations for 2013 are $A.M._{Cr}$ = 60±21 mg/kg; $A.M._{Cu}$ = 144±69 mg/kg; $A.M._{Pb}$ = 225±87 mg/kg; $A.M._{V}$ = 90±15 mg/kg and $A.M._{Zn}$ = 567±163 mg/kg (Table 1). The mean concentrations are in a decreasing order Zn>Pb>Cu>V>Cr.

The urban road dust concentrations showed a decrease in the average concentration of Cr from 2011 to 2013, while the concentration of Cu and Zn exhibited an increase. The concentration of Zn in both years was abundant; Zn is related to the production of metals and the use of fossil fuels, so its increase in urban road dust may be associated with the increase in emissions from the vehicle fleet (Jones *et al.*, 2014).

The mean concentrations of all samples for Pb and V in 2011 and 2013 were similar (Table 1). In the residential area of Mexico City, the average concentration of lead significantly increased from 181 mg/kg in 2011 to 225 mg/kg in 2013, indicating a 24 % increase. Additionally, the percentage of samples of urban road dust in residential land use containing Cr, Cu, and Zn increased from 2011 to 2013. However, according to the Mexican regulation threshold (NOM-147-SEMARNAT/SSA1-2004), the values found do not indicate any contamination issues as they were within the normal range (SEMARNAT, 2007).

High concentrations of elements such as Cr, Cu, Pb, V, and Zn were found at 126 sites, representing 89 % of the samples (Tables 1 and 2).

The PLI calculated for 140 samples of urban road dust in Mexico City, revealed values ranging from 0.7 to 3.1. This suggests the presence of both contaminated and uncontaminated sites. Specifically, 16 sites exhibited a PLI ≤ 1 (uncontaminated), 89 sites had PLI values ranging from 1 to 2 (moderate contamination), and 35 sites had PLI > 2 (high contamination).

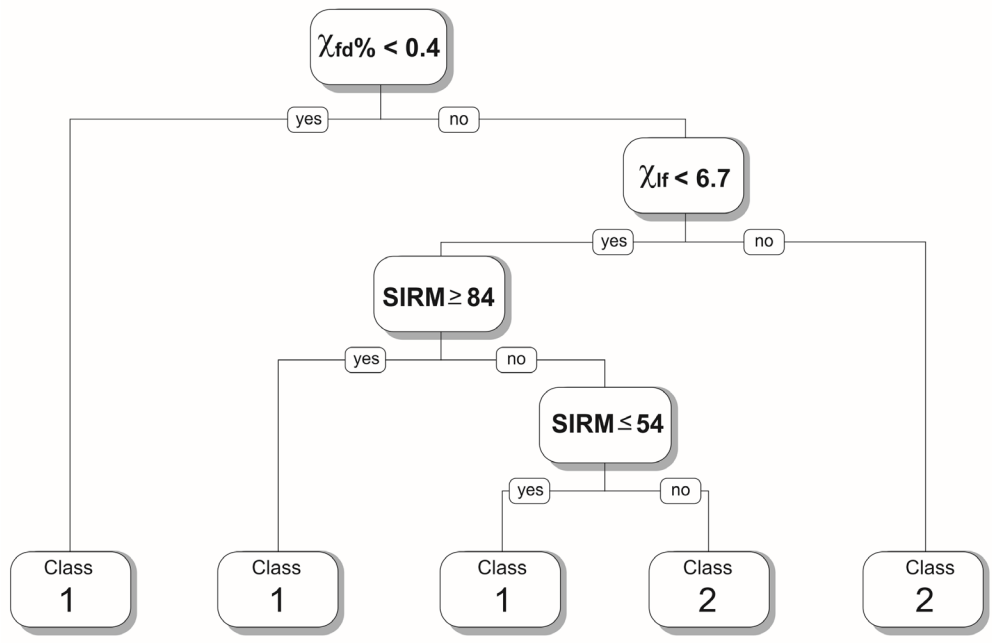


Figure 4 **Classification tree to classify contaminated and uncontaminated samples in urban road dust of Mexico City.**

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Table 1. Descriptive statistics of magnetic parameters and concentration of elements in urban road dust of Mexico City recollected in 2011 and 2013.

Year	Par		Mean	Med	SD	VC	Min	Max	Q1	SK	Kr
2011 (n=84)	χlf	(μm^3/kg)	5.6	5.2	1.9	33 %	2.1±0.3	9.8±0.3	4.3	2.4	-0.4
	χ_{fd}%	(%)	1.7	1.6	1.2	69 %	0.0±0.8	4.7±0.8	0.75	2.5	-0.3
	SIRM	(mAm^2 kg)	73.4	71.9	20.7	28 %	31.0±4.7	130.5±4.7	58.3	1.6	0.1
	Cr	(mg/kg)	113	111	33	30 %	40±7	186±7	87	1.4	-0.9
	Cu	(mg/kg)	94	83	59	63 %	10±5	340±5	53	5.8	6.9
	Pb	(mg/kg)	203	183	130	64 %	17±3	610±3	102	3.6	1.1
	V	(mg*/kg)	87	89	14	16 %	48±9	116±9	77	-1.5	-0.1
	Zn	(mg/kg)	309	291	126	41 %	114±4	651±4	225	2.8	-0.1
	PLI	DI	1.6	1.5	0.5	34 %	0.7±0.3	3.1±3	1.2	2.0	-0.4
2013 (n=56)	χlf	(μm^3/kg)	4.0	3.7	1.1	29 %	2.1±0.5	7.1±0.5	3.35	2.4	0.6
	χ_{fd}%	(%)	2.9	2.6	1.6	55 %	0.2±0.7	8.3±0.7	1.9	4.4	4.2
	SIRM	(mAm^2 kg)	55.3	51.6	17.2	31 %	29.3±4.7	111.7±4.7	42.34	3.3	2.0
	Cr	(mg/kg)	60	58	21	35 %	28±7	127±7	42	2.0	0.6
	Cu	(mg/kg)	144	123	69	48 %	40±5	370±5	101	4.9	4.0
	Pb	(mg/kg)	225	210	87	39 %	70±3	525±3	177	2.9	2.5
	V	(mg*/kg)	90	88	15	17 %	57±9	124±9	82	1.1	-0.2
	Zn	(mg/kg)	567	565	163	29 %	252±4	1109±4	456	1.9	1.7
	PLI	DI	1.8	1.8	0.4	20 %	0.9±0.3	2.9±0.3	1.6	0.6	1.2
2011 Urban road dust in residential land use (n=51)	χlf	(μm^3/kg)	5.4	5.3	1.7	32 %	2.1±0.5	8.8±0.5	**4.4**	0.9	-0.7
	χ_{fd}%	(%)	1.8	1.5	1.2	67 %	0.0±0.8	4.7±0.8	**0.9**	2.1	0.0
	SIRM	(mAm^2 kg)	71.4	70.1	20.1	28 %	31.0±4.7	130.5±4.7	**54.9**	1.8	1.5
	Cr	(mg/kg)	111	107	31	28 %	53±7	182±7	**86**	1.0	-1.0
	Cu	(mg/kg)	84	71	57	67 %	21±5	277±5	**43**	4.4	3.6
	Pb	(mg/kg)	181	120	142	78 %	17±3	610±3	**82**	4.1	2.3
	V	(mg*/kg)	87	90	14	16 %	48±9	116±9	**77**	-1.1	0.0
	Zn	(mg/kg)	283	266	111	39 %	114±4	576±4	**198**	2.0	0.0
	PLI	DI	1.5	1.4	0.5	35 %	0.7±0.3	3.0±0.3	**1.0**	2.3	0.4

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Morton-Bermea *et al.* (2009) suggested a magnetic susceptibility value of $\chi_{lf} > 4.0$ μm^3/kg to detect soils with high concentrations of potentially toxic elements (PTEs) in Mexico City. Therefore, to determine the background concentration of element values in urban road dust in Mexico City, we propose using the first quartile (Q1) values for residential land use type. The values for Cr, Cu, Pb, V, and Zn are 86 mg/kg, 43 mg/kg, 82 mg/kg, 77 mg/kg, and 198 mg/kg, respectively (Table 1).

3.1. MATHEMATICAL MODELS TO DETERMINATION VALUES OF PLI BY MAGNETIC PARAMETERS

3.1.1 LINEAR REGRESSION MODEL

Various linear regression models were tested to obtain a between magnetic parameters and PLI values. Only χ_{lf} correlated weakly with PLI values (Figure 2).

The linear regression model is expressed here (Equation 10):

$$PLI_{LR} = 1.0372 + 0.0921\ \chi_{lf} \quad (10)$$

This model shows values of correlation coefficient was 0.30, R^2=0.09, and P-value= 0.027 ($p<0.05$). Since the P-value in the variance analyse was less than 0.05, it is a statistically significant relationship between PLI and χ_{lf} with a confidence level of 95 %. The variables are weakly related. The model satisfies the statistical assumptions, the values of Shapiro-Wilk normality test (w=0.95, P-value=0.06), and Breusch-Pagan Test (BP = 43.12, P-value = 0.88). In this case, the p-value is 0.88, which is greater than 0.05. Therefore, we cannot reject the null hypothesis and we can assume homoscedasticity in the residuals of the linear regression model.

3.1.2 ARTIFICIAL NEURAL NETWORK

The artificial neural network model was constructed considering the three magnetic parameters χ_{lf}, χ_{fd}%, and SIRM as input data (Equation 5). The architecture had three input layers; 6, 3, 2 hidden layers and an output layer for calculating the normalized PLI (PLInorm). Equation 11 to 14 show the values that conform to the neural network trained by the back propagation algorithm:

$$\varphi\left[\begin{pmatrix} 3.77 & -1.82 & -7.92 \\ 7.16 & -0.69 & -4.04 \\ -33.92 & 12.51 & -17.13 \\ -3.32 & -0.83 & 1.37 \\ -4.30 & -1.57 & -1.98 \\ 6.08 & -25.31 & 5.77 \end{pmatrix}\begin{pmatrix} \chi_{lf\ norm} \\ \chi_{fd\%\ norm} \\ SIRM_{norm} \end{pmatrix} + \begin{pmatrix} 0.92 \\ 0.11 \\ 13.35 \\ 0.91 \\ 2.22 \\ -6.25 \end{pmatrix}\right] = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{pmatrix} \quad (11)$$

$$\varphi\left[\begin{pmatrix} 46.59 & 0.88 & -10.04 & -16.47 & -24.40 & 95.99 \\ -10.64 & 0.25 & -4.71 & 0.32 & -20.81 & 7.99 \\ -2.89 & 1.37 & 7.86 & -14.42 & -18.28 & -2.90 \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{pmatrix} + \begin{pmatrix} 0.71 \\ 0.40 \\ 1.02 \end{pmatrix}\right] = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \quad (12)$$

$$\varphi\left[\begin{pmatrix} 0.12 & -3.02 & -11.04 \\ 0.55 & 0.96 & -13.01 \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{pmatrix} 0.74 \\ -1.07 \end{pmatrix}\right] = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (13)$$

$$\varphi\left[(-1.01 \quad 1.44)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + (-0.26)\right] = PLI_{norm} \quad (14)$$

Where, h, z, and z are results obtained by entering the input value into the artificial neural network. It is necessary to use a data transformation to obtain the PLI value by ANN model. In this case, the associated equation was Equation 15.

$$PLI_{ANN} = \left(2.36 * PLI_{norm}\right) + 0.71 \quad (15)$$

The values of PLI_{ANN} were compared to a data of PLI_{real} of 2011, yielding a squared error 0.6 (Figure 3). The squared error is a measure of the difference between the predicted value and the actual value in regression problems. It is calculated as the square of the difference between the predicted value and the actual value and is commonly used as a loss function in machine learning algorithms to minimize the difference between predicted values and actual values. The results of predicted PLI by ANN method with magnetic parameters showed a similitud with values calculated (Figure 3).

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Table 2. Sampling site, values of magnetic parameters, and polluted load index of urban road dust of Mexico City.

n	Latitude	Longitude	Year	Type	χ_{lf} ($\mu m^3/kg$)	χ_{fd}% (%)	SIRM (mAm^2/kg)	PLI_{real}	PLI_{RL}	PLI_{ANN}	PLI_{CT}
1	19° 24' 10.21"	99° 6' 27.00"	2011	P	8.5	1.9	100.45	2.4	1.9	3.1	2
2	19° 23' 54.76"	99° 3' 56.16"	2011	E	5.5	2.7	74.54	1.8	1.5	1.4	1
3	19° 23' 32.77"	99° 2' 8.81"	2011	E	4.7	1.3	60.77	1.5	1.3	1.4	1
4	19° 24' 33.13"	98° 59' 28.31"	2011	P	4.6	0.8	65.04	1.3	1.3	1.5	1
5	19° 23' 6.67"	98° 57' 18.55"	2011	E	4.3	1.9	67.74	2.3	1.2	1.4	2
6	19° 22' 0.312"	99° 6' 15.19"	2011	E	5.2	2.2	81.53	2.4	1.4	1.5	2
7	19° 21' 44.41"	99° 3' 21.86"	2011	P	8.6	0.4	88.69	2.2	2.0	1.9	2
8	19° 21' 23.46"	99° 0' 16.62"	2011	P	6.3	1.9	96.8	1.3	1.6	1.4	2
9	19° 21' 35.22"	98° 57' 27.90"	2011	E	5.7	0.8	83.46	1.6	1.5	1.5	2
10	19° 19' 47.28"	99° 7' 47.88"	2011	E	9.3	0.5	110.36	2.1	2.1	2.2	2
11	19° 18' 57.05"	99° 4' 43.68"	2011	E	7.3	0.4	87.87	1.9	1.7	1.8	2
12	19° 18' 40.25"	99° 2' 17.75"	2011	P	5.1	2.5	82.55	1.6	1.4	1.5	2
13	19° 18' 33.29"	98° 58' 10.78"	2011	P	4.1	2.2	70.42	1.5	1.2	1.5	2
14	19° 18' 32.46"	98° 56' 53.75"	2011	E	4.6	1.7	51.1	0.8	1.3	0.9	1
15	19° 16' 50.35"	99° 7' 16.08"	2011	E	7.4	1.5	97.82	1.5	1.8	1.6	2
16	19° 17' 10.19"	99° 4' 5.77"	2011	E	4.2	2.2	60.2	1.6	1.3	1.4	2
17	19° 17' 9.97"	99° 2' 16.33"	2011	E	5.1	3.8	95.7	1.3	1.3	1.4	2
18	19° 17' 29.11"	99° 0' 29.5"	2011	E	4.2	2.6	81.11	1.5	1.2	1.5	2
19	19° 17' 28.08"	98° 57' 7.13"	2011	E	3.4	1.5	49.56	0.7	1.1	1.5	2
20	19° 15' 40.67"	99° 6' 50.75"	2011	E	5.7	2.6	75.01	1.4	1.5	1.4	1
21	19° 15' 16.86"	99° 3' 3.95"	2011	E	5.2	3.1	85.74	1	1.4	1.4	2
22	19° 15' 9.60"	99° 0' 6.57"	2011	E	4.2	2.0	91.3	0.9	1.2	1.5	2
23	19° 14' 49.92"	98° 59' 47.94"	2011	E	4.6	1.5	76.54	1.3	1.3	1.5	1
24	19° 13' 29.28"	98° 57' 42.65"	2011	P	4.7	0.4	85.03	1.5	1.3	1.5	2
25	19° 34' 2.695"	99° 5' 16.38"	2011	E	6.6	1.0	69.33	0.9	1.7	0.9	1
26	19° 33' 53.75"	99° 3' 34.75"	2011	E	4.4	1.5	60.09	0.8	1.3	1.4	1
27	19° 34' 23.52"	99° 1' 34.39"	2011	E	2.1	1.6	31.03	1.3	0.9	1.5	2
28	19° 35' 43.39"	98° 58' 26.04"	2011	E	4.5	1.2	73.85	0.9	1.3	1.5	1
29	19° 35' 31.02"	98° 57' 56.11"	2011	P	3.0	4.5	50.25	0.8	1.0	1.5	2
30	19° 32' 2.82"	99° 4' 48.76"	2011	P	6.9	2.3	81.97	1	1.7	1.7	2
31	19° 33' 20.22"	99° 4' 25.25"	2011	E	5.6	0.9	53.36	1.3	1.5	1.6	1
32	19° 32' 29.87"	99° 2' 28.31"	2011	E	4.4	2.0	61.93	1.2	1.3	1.4	1
33	19° 29' 50.56"	99° 5' 58.06"	2011	P	6.3	1.2	92.25	1.1	1.6	1.4	2
34	19° 30' 32.34"	99° 3' 26.99"	2011	P	6.2	0.9	82.47	1.6	1.6	1.6	2

n: site number, E: training data, P: test data, χ_{lf}: in $\mu m^3/kg$, χ_{fd}% in %, SIRM: Saturation isothermal remanent magnetization, PLI_{real}: Pollution load index calculated for Cr, Cu, Pb, V, and Zn concentration means, PLI_{RL}: estimated value with regression model, PLI_{ANN}: estimated value with artificial neural network model, PLI_{CT} estimated values of the classification tree model.

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Table 2. Sampling site, values of magnetic parameters, and polluted load index of urban road dust of Mexico City. (continuation)

n	Latitude	Longitude	Year	Type	χ_{lf} ($\mu m^3/kg$)	χ_{fd}% (%)	SIRM (mAm^2/kg)	PLI_{real}	PLI_{RL}	PLI_{ANN}	PLI_{CT}
35	19° 30' 11.11"	99° 2' 23.41"	2011	E	3.9	1.8	71.72	1.6	1.2	1.5	2
36	19° 27' 51.54"	99° 6' 28.48"	2011	P	9.6	0.8	118.73	2.3	2.1	2.6	2
37	19° 28' 0.84"	99° 4' 28.68"	2011	E	6.7	0.6	103.14	1.8	1.6	1.9	2
38	19° 28' 38.16"	99° 2' 35.56"	2011	E	3.6	3.3	54.32	2.1	1.2	1.5	2
39	19° 26' 53.28"	98° 58' 11.69"	2011	P	2.7	1.1	44.03	0.8	1.0	1.5	2
40	19° 25' 23.62"	99° 6' 4.93"	2011	E	2.7	1.1	44.03	2.2	1.0	1.5	2
41	19° 25' 41.87"	99° 3' 54.97"	2011	E	6.9	2.5	86.78	2.1	1.7	1.7	2
42	19° 25' 12.13"	99° 1' 39.47"	2011	E	4.5	1.8	68.89	1.6	1.3	1.4	1
43	19° 24' 56.65"	98° 59' 5.34"	2011	E	4.4	2.2	64.74	1.1	1.3	1.4	1
44	19° 26' 7.07"	98° 57' 17.68"	2011	E	3.2	3.5	42.42	1.3	1.1	1.5	2
45	19° 35' 59.93"	99° 17' 21.47"	2011	E	5.7	1.6	49.55	0.9	1.5	0.8	1
46	19° 35' 3.41"	99° 15' 19.01"	2011	P	8.8	0.5	104.01	2.0	2.0	2.2	2
47	19° 34' 12.95"	99° 14' 54.07"	2011	P	5.7	0.7	51.35	1.4	1.5	1.5	1
48	19° 34' 57.01"	99° 11' 42.60"	2011	P	5.0	0.6	65.83	1.7	1.4	1.7	1
49	19° 34' 23.74"	99° 7' 55.38"	2011	E	4.8	0.2	75.22	1.4	1.3	1.5	1
50	19° 34' 45.70"	99° 18' 49.93"	2011	E	6.1	0	87.29	1.4	1.5	1.9	2
51	19° 32' 30.53"	99° 15' 1.19"	2011	P	7.0	0.1	68.58	1.4	1.7	1.4	1
52	19° 32' 24.64"	99° 13' 45.55"	2011	E	9.6	0.5	91.63	2.2	2.1	2.1	2
53	19° 32' 40.55"	99° 11' 28.27"	2011	E	5.0	1.3	53.6	1.6	1.4	1.6	1
54	19° 29' 17.35"	99° 17' 11.22"	2011	E	8.8	1.1	79.46	1.3	2.0	1.3	1
55	19° 29' 51.01"	99° 15' 11.41"	2011	P	9.7	0	97.59	2.1	2.2	2.1	2
56	19° 29' 30.35"	99° 13' 52.44"	2011	P	5.3	0.8	54.85	2.3	1.5	1.8	1
57	19° 29' 13.44"	99° 11' 27.30"	2011	E	8.8	1.3	126.21	3.0	2.0	2.8	2
58	19° 29' 27.95"	99° 8' 30.03"	2011	E	5.7	0.2	79.63	2.3	1.5	1.9	2
59	19° 27' 32.33"	99° 17' 28.56"	2011	E	8.1	0	62.88	0.9	1.9	0.9	1
60	19° 27' 42.31"	99° 16' 33.90"	2011	P	4.2	0.3	55.29	0.8	1.3	1.5	2
61	19° 28' 13.63"	99° 12' 37.93"	2011	E	5.7	0.7	69.12	2.5	1.5	1.9	1
62	19° 28' 7.67"	99° 11' 17.76"	2011	P	6.8	0.5	74.26	2.0	1.7	1.3	1
63	19° 26' 14.22"	99° 17' 8.53"	2011	P	8.4	1.1	83.56	1.4	2.0	1.1	2
64	19° 26' 14.41"	99° 13' 45.30"	2011	E	6.2	0.4	86.02	2.1	1.6	1.9	2
65	19° 25' 41.75"	99° 11' 19.27"	2011	E	5.3	0.2	61.27	1.4	1.4	1.9	1
66	19° 25' 38.05"	99° 8' 47.45"	2011	P	9.8	0.4	115.56	2.5	2.1	2.2	2
67	19° 23' 27.01"	99° 8' 46.41"	2011	E	4.6	3.3	58.72	3.1	1.3	3.0	1
69	19° 21' 25.39"	99° 11' 13.80"	2011	E	3.9	4.4	45.13	2.1	1.2	2.0	2
70	19° 19' 33.28"	99° 13' 29.18"	2011	E	4.4	2.9	57.18	1.2	1.3	1.5	1

n: site number, E: training data, P: test data, χ_{lf}: in $\mu m^3/kg$, χ_{fd}% in %, SIRM: Saturation isothermal remanent magnetization, PLI_{real}: Pollution load index calculated for Cr, Cu, Pb, V, and Zn concentration means, PLI_{RL}: estimated value with regression model, PLI_{ANN}: estimated value with artificial neural network model, PLI_{CT} estimated values of the classification tree model.

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Table 2. Sampling site, values of magnetic parameters, and polluted load index of urban road dust of Mexico City. (continuation)

n	Latitude	Longitude	Year	Type	χ_{lf} ($\mu m^3/kg$)	χ_{fd}% (%)	SIRM (mAm^2/kg)	PLI_{real}	PLI_{RL}	PLI_{ANN}	PLI_{CT}
71	19° 19' 21.11"	99° 11' 12.39"	2011	E	3.8	2.8	41.6	1.5	1.2	1.2	2
72	19° 19' 8.40"	99° 9' 8.90"	2011	E	5.4	2.8	66.49	1.4	1.5	1.3	1
73	19° 21' 15.98"	99° 8' 57.09"	2011	E	4.7	1.9	65.66	1.2	1.3	1.4	1
74	19° 16' 48.71"	99° 8' 16.38"	2011	E	7.9	2.1	97.96	2.3	1.8	2.6	2
75	19° 15' 11.78"	99° 9' 43.21"	2011	P	3.4	4.7	58.89	1.2	1.1	1.9	2
76	19° 16' 50.01"	99° 11' 19.10"	2011	E	3.0	3.9	51.92	1.1	1.0	1.5	2
77	19° 15' 59.99"	99° 11' 59.68"	2011	E	6.6	2.2	130.45	1.3	1.6	1.2	2
78	19° 17' 5.68"	99° 13' 21.78"	2011	E	4.8	3.1	80.51	2.2	1.3	2.1	2
79	19° 14' 53.78"	99° 13' 53.70"	2011	P	3.0	4.4	70.08	0.9	1.0	1.5	2
80	19° 19' 58.08"	99° 17' 16.20"	2011	E	5.8	2.7	87.21	0.9	1.5	1.4	2
81	19° 23' 50.68"	99° 13' 12.70"	2011	P	6.6	1.7	72.09	1.8	1.7	1.2	1
82	19° 23' 4.89"	99° 16' 13.21"	2011	E	3.7	3.6	38.13	1.1	1.2	1.1	2
83	19° 24' 19.51"	99° 16' 45.29"	2011	E	6.2	1.5	78.42	1.5	1.6	1.6	1
84	19° 22' 6.89"	99° 17' 13.89"	2011	E	4.3	1.9	57.88	1.6	1.3	1.3	2
85	19° 26' 52.09"	99° 3' 17.65"	2013	P	2.8	3.9	52.52	1.9	1.0	1.5	2
86	19° 26' 42.45"	99° 3' 44.33"	2013	P	2.8	4.6	55.06	1.7	1.0	1.5	2
87	19° 26' 46.44"	99° 4' 5.42"	2013	P	3.6	3.8	69.59	1.8	1.1	1.6	2
88	19° 26' 46.46"	99° 4' 44.25"	2013	P	3.2	5.4	46.51	1.5	1.1	2.0	2
89	19° 26' 28.89"	99° 4' 35.73"	2013	P	3.5	3.2	47.76	1.8	1.1	1.4	2
90	19° 26' 14.28"	99° 4' 52.94"	2013	P	4.9	2.7	73.21	1.4	1.4	1.7	1
91	19° 26' 1.54"	99° 5' 24.52"	2013	P	2.2	8.3	33.8	1.7	0.9	3.3	2
92	19° 25' 19.71"	99° 5' 5.23"	2013	P	2.1	7.9	33.71	1.3	0.9	2.4	2
93	19° 25' 5.97"	99° 4' 25.96"	2013	P	7.1	2.1	111.72	2.1	1.7	1.5	2
94	19° 25' 31.51"	99° 4' 22.30"	2013	P	2.1	6.0	36.75	2.2	0.9	1.5	2
95	19° 25' 47.82"	99° 4' 1.15"	2013	P	5.4	3.4	98.8	0.9	1.4	1.4	2
96	19° 25' 58.04"	99° 3' 41.05"	2013	P	3.8	4.2	63.39	1.6	1.2	2.2	2
97	19° 26' 16.46"	99° 3' 20.72"	2013	P	3.8	4.3	58.6	1.5	1.2	2.3	2
98	19° 26' 38.04"	99° 2' 51.01"	2013	P	5.1	3.7	71.16	2.1	1.4	3.5	1
99	19° 27' 6.28"	99° 3' 2.08"	2013	P	3.5	5.2	63.19	1.1	1.1	2.9	2
100	19° 28' 26.67"	99° 6' 31.94"	2013	P	5.1	2.1	70.53	2.1	1.4	1.4	1
101	19° 27' 49.58"	99° 6' 42.27"	2013	P	3.6	2.6	45.08	1.8	1.2	1.4	2
102	19° 27' 21.07"	99° 6' 48.08"	2013	P	2.9	2.4	46.12	1.9	1	1.5	2
103	19° 24' 14.77"	99° 7' 16.93"	2013	P	3.4	2.3	40.75	1.6	1.1	1.5	2
104	19° 22' 33.91"	99° 7' 20.45"	2013	P	3.0	3.1	47.33	1.9	1.1	1.5	2
105	19° 22' 8.53"	99° 7' 23.55"	2013	P	3.4	3.0	42.34	1.4	1.1	1.4	2

n: site number, E: training data, P: test data, χ_{lf}: in $\mu m^3/kg$, χ_{fd}% in %, SIRM: Saturation isothermal remanent magnetization, PLI_{real}: Pollution load index calculated for Cr, Cu, Pb, V, and Zn concentration means, PLI_{RL}: estimated value with regression model, PLI_{ANN}: estimated value with artificial neural network model, PLI_{CT} estimated values of the classification tree model.

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Table 2. Sampling site, values of magnetic parameters, and polluted load index of urban road dust of Mexico City. (continuation)

n	Latitude	Longitude	Year	Type	χ_{lf} ($\mu m^3/kg$)	χ_{fd}% (%)	SIRM (mAm^2/kg)	PLI_{real}	PLI_{RL}	PLI_{ANN}	PLI_{CT}
106	19° 27' 40.21"	99° 7' 32.79"	2013	P	2.7	1.9	46.99	2.1	1.0	1.5	2
107	19° 27' 32.23"	99° 9' 46.17"	2013	P	6.2	1.6	69.26	1.4	1.6	1.5	1
108	19° 26' 39.81"	99° 5' 13.26"	2013	P	4.7	1.2	48.79	1.9	1.4	1.1	1
109	19° 26' 52.73"	99° 9' 52.44"	2013	P	6.3	1.7	44.04	2.3	1.7	0.1	1
110	19° 25' 27.41"	99° 10' 38.86"	2013	P	3.8	0.2	35.21	2.2	1.2	1.8	2
111	19° 24' 7.55"	99° 11' 5.93"	2013	P	3.2	2.6	54.31	2.3	1.1	1.5	2
112	19° 24' 23.54"	99° 5' 49.87"	2013	P	5.0	2.4	62.66	2.3	1.4	1.3	1
113	19° 23' 7.51"	99° 11' 11.82"	2013	P	3.6	2.6	38.52	1.5	1.2	1.4	2
114	19° 22' 37.60"	99° 6' 4.42"	2013	P	6.8	1.2	91.11	1.8	1.7	1.4	2
115	19° 22' 17.52"	99° 6' 9.27"	2013	P	3.6	2.9	55.89	1.8	1.2	1.5	2
116	19° 21' 44.67"	99° 7' 42.80"	2013	P	4.4	2.6	57.87	2.0	1.3	1.3	1
117	19° 28' 58.027"	99° 7' 42.18"	2013	P	4.5	1.3	44.59	1.9	1.3	1.0	1
118	19° 28' 2.91"	99° 8' 27.62"	2013	P	4.3	2.0	56.3	1.8	1.3	1.3	2
119	19° 27' 25.09"	99° 8' 51.22"	2013	P	3.9	1.6	69.34	1.7	1.2	1.5	2
120	19° 26' 45.90"	99° 9' 12.04"	2013	P	4.0	1.7	41.75	1.8	1.2	1.3	2
121	19° 25' 33.91"	99° 9' 42.76"	2013	P	3.8	2.0	43.47	1.2	1.2	1.4	2
122	19° 24' 5.35"	99° 10' 11.75"	2013	P	4.9	1.3	85	2.1	1.3	1.5	2
123	19° 23' 10"	99° 10' 30.14"	2013	P	3.5	3.1	29.32	1.4	1.2	1.3	2
124	19° 22' 28.24"	99° 10' 43.57"	2013	P	3.1	3.0	42.34	1.8	1.1	1.5	2
125	19° 22' 3.41"	99° 10' 51.77"	2013	P	3.3	3.7	35.73	1.6	1.1	1.4	2
126	19° 21' 6.39"	99° 11' 10.21"	2013	P	4.0	4.4	42.01	2.2	1.2	1.9	2
127	19° 25' 34.38"	99° 7' 58.89"	2013	P	2.5	2.0	37.07	1.7	1.0	1.5	2
128	19° 24' 22.08"	99° 8' 10.56"	2013	P	5.3	2.0	65.04	1.8	1.4	1.5	1
129	19° 23' 0.70"	99° 8' 23.14"	2013	P	4.4	3.1	50.63	1.6	1.3	1.2	1
130	19° 22' 42.09"	99° 8' 26.07"	2013	P	6.1	1.9	86.71	1.5	1.5	1.4	2
131	19° 21' 25.04"	99° 8' 42.91"	2013	P	2.4	5.3	37.07	1.2	1	1.5	2
132	19° 28' 22.31"	99° 8' 15.36"	2013	P	3.5	2.4	42.25	1.6	1.2	1.4	2
133	19° 27' 55.18"	99° 8' 14.20"	2013	P	3.8	1.6	68.1	1.8	1.2	1.5	2
134	19° 27' 17.57"	99° 8' 16.81"	2013	P	3.6	2.0	64.48	1.4	1.1	1.5	2
135	19° 26' 32.34"	99° 8' 22.40"	2013	P	4.4	0.8	68.14	1.7	1.3	1.5	1
136	19° 25' 25.44"	99° 8' 33.83"	2013	P	3.5	1.4	44.55	2.2	1.2	1.5	2
137	19° 24' 10.25"	99° 8' 42.09"	2013	P	4.2	1.6	47.09	2.0	1.3	1.2	2
138	19° 23' 1.37"	99° 8' 50.60"	2013	P	5.6	2.3	65.34	2.3	1.5	1.6	1
139	19° 21' 54.63"	99° 9' 2.54"	2013	P	3.6	2.6	60	2.9	1.1	1.5	2
140	19° 21' 25.54"	99° 9' 10.36"	2013	P	3.5	3.5	58.58	1.8	1.1	1.5	2

n: site number, E: training data, P: test data, χ_{lf}: in $\mu m^3/kg$, χ_{fd}% in %, SIRM: Saturation isothermal remanent magnetization, PLI_{real}: Pollution load index calculated for Cr, Cu, Pb, V, and Zn concentration means, PLI_{RL}: estimated value with regression model, PLI_{ANN}: estimated value with artificial neural network model, PLI_{CT} estimated values of the classification tree model.

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

3.1.3 CLASSIFICATION TREE (CT)

The decision tree structure uses the three values of parameters χ_{lf}, χ_{fd}% and SIRM of 2011 test data to establish two types of samples: uncontaminated (Class 1) and contaminated (Class 2). The constructed model showed the values into four nodes: the first node is classified for χ_{fd}% < 0.4 %, the second node with χ_{lf} < 6.7 μm^3/kg, the third node with SIRM ≥ 84 mA m^2/kg, and the fourth node with SIRM ≤ 54 mA m^2/kg (Figure 4).

In the first node, the data were classified with χ_{fd}% related to ultrafine particles (< 50 nm). In the second node, data were classified with χlf, related to magnetic mineral concentrations. In the third node, the data were classified with high concentration of mineral ferrimagnetic (SIRM ≥ 84 mA m^2/kg) and in the four node, data were classified with low concentration of ferrimagnetic concentrations (Figure 4).

In this study, we have proposed a model of CT using three magnetic parameters to classify contaminated sites for PTEs, unlike other models that use a single parameter (Morton-Bermea *et al.* 2009; Cejudo *et al.* 2015). The χ_{lf} values (6.7 μm^3/kg) and SIRM values (54 mA m^2/kg) reported in this study to identify contaminated sites (PLI>1) are comparable to those reported by Morton-Bermea *et al.* (2009) and Cejudo *et al.* (2015) for soils and urban dust in Mexico City (χ_{lf}:4.0 μm^3/kg and SIRM:46 mA m^2/kg, respectively). The urban dust of Mexico City shows that higher concentrations of ferrimagnetic minerals are associated with

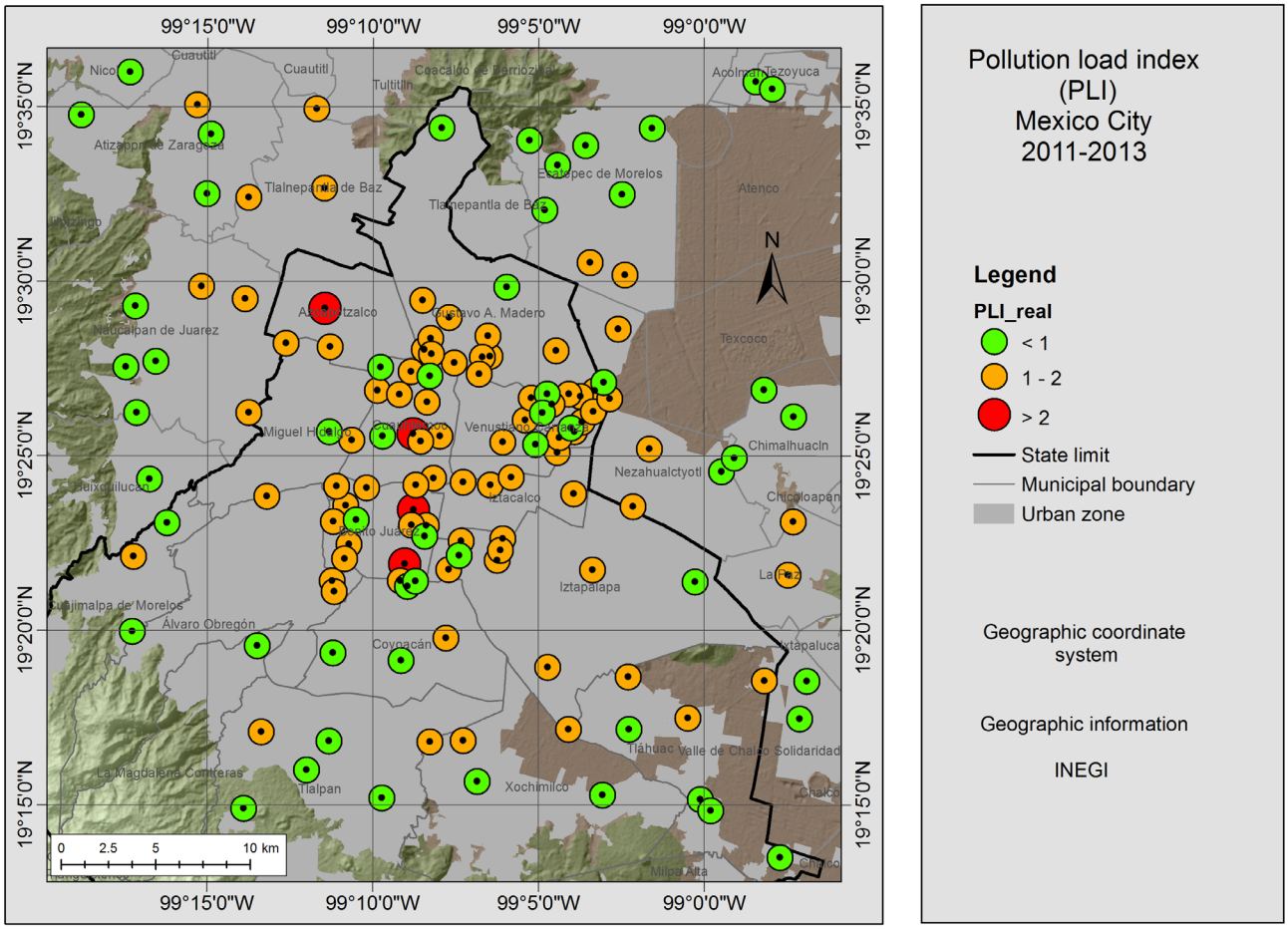


Figure 5 Map of the pollution load index for urban road dust in Mexico City, obtained by measuring the concentration of elements (PLI_{real}).

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

Table 3. Confusion matrix for predictive models of values of pollution load index obtained by magnetic parameters.

Model	Class		Estimate 1	Estimate 2	Estimate 3	Total	Precision
Linear regression	Real	1	**48**	7	0	55	61 %
		2	45	**36**	0	81	
		3	2	1	**1**	4	
						140	
	%	1	**34 %**	5 %	0 %	39 %	
		2	32 %	**26 %**	0 %	58 %	
		3	1 %	1 %	**1 %**	3 %	
						100 %	

Model	Class		Estimate 0	Estimate 1	Estimate 2	Estimate 3	Estimate 4	Total	Precision
Artificial neural network	Real	1	0	**39**	15	1	0	55	84 %
		2	1	1	**75**	3	1	81	
		3	0	0	1	**3**	0	4	
								140	
	%	1	0 %	**28 %**	11 %	1 %	0 %	39 %	
		2	1 %	1 %	**54 %**	2 %	1 %	58 %	
		3	0 %	0 %	1 %	**2 %**	0 %	3 %	
								100 %	

Model	Class		Estimate 1	Estimate 2	Estimate 3	Total	Precision
Classification tree	Real	1	**34**	21		55	60 %
		2	31	**50**		81	
		3	1	3		4	
						140	
	%	1	**24 %**	15 %		39 %	
		2	22 %	**36 %**		58 %	
		3	1 %	2 %		3 %	
						100 %	

higher concentrations of PTEs.

3.1.4 CONFUSION MATRIX

The confusion matrix for the estimated models of PLI is shown in Table 3. It was considered that the difference between the PLI estimate and the PLI calculation (PLIreal) was less than 25 %. The confusion matrix for the linear regression model showed that it is capable of estimating up to 61 % of correct values of PLI through χ_{lf} and predicting the PLI ≤ 1 (34 %) and values between 1 < PLI ≤ 2 (26 %) more accurately. The confusion matrix for the artificial neural network model is capable of estimating up to 84 % of PLI correct values through χ_{lf}, χ_{fd}% and SIRM and more accurately predict values of PLI ≤ 1 (28 %) and values between 1 < PLI ≤ 2 (54 %; Table 3).

The confusion matrix for the decision tree model showed that it is capable of classifying the contaminated and uncontaminated sites through χ_{lf}, χ_{fd}% and SIRM, with values ≥ 84 mA m^2/kg and ≤ 54 mA m^2/kg (Figure 4). The model predicts 36 % contaminated sites with PLI >1 and 28 % uncontaminated sites with PLI ≤ 1 (Table 3). Thus, the artificial neural network was the best model to estimate PLI values from the magnetic parameters. The comparison of the spatial distribution of PLI values predicted for the artificial neural network model and PLI values computed were similar in the city center but showed different outputs in the suburban areas (Figures 5 and 6).

The magnetic parameters are a tool to quickly evaluate the concentration of PTEs in urban road

http://dx.doi.org/10.18268/BSGM2024v76n3a290224

dust and soils (Chaparro *et al.*, 2006; Oudeika *et al.*, 2020). However, it is observed that when machine learning algorithms are used, their level of precision considerably increases, in comparison to the linear regression methods used in many studies of soil and dust in urban areas (Liu *et al.*, 2016; Wang *et al.*, 2019, Oudeika *et al.*, 2020).

The urban road dust from Mexico city shows very high values of magnetic parameters compared to those reported in other cities. For examples, in this study the χ_{lf} values was 5.2 $\mu m^3/kg$ in 2011; in Turkey was 0.010 $\mu m^3/kg$ (Canbay *et al.*, 2010); in Shanghai, China was 1.55 $\mu m^3/kg$ (Wang *et al.*, 2019) and Isfahan Iran was 0.74 $\mu m^3/kg$ (Karimi *et al.*, 2011). However, a relationship between magnetic parameters and the concentration of PTEs has been observed in all of these cities. It is possible that the artificial neural network model could improve the prediction of sites with high concentrations based on magnetic parameters.

4. Conclusions

The artificial neural network proved to be the best approach to predict the pollution load index of urban road dust from magnetic parameters. The accuracy was 20 % better than the regression linear model and the decision tree technique. Therefore, the artificial neural network could be

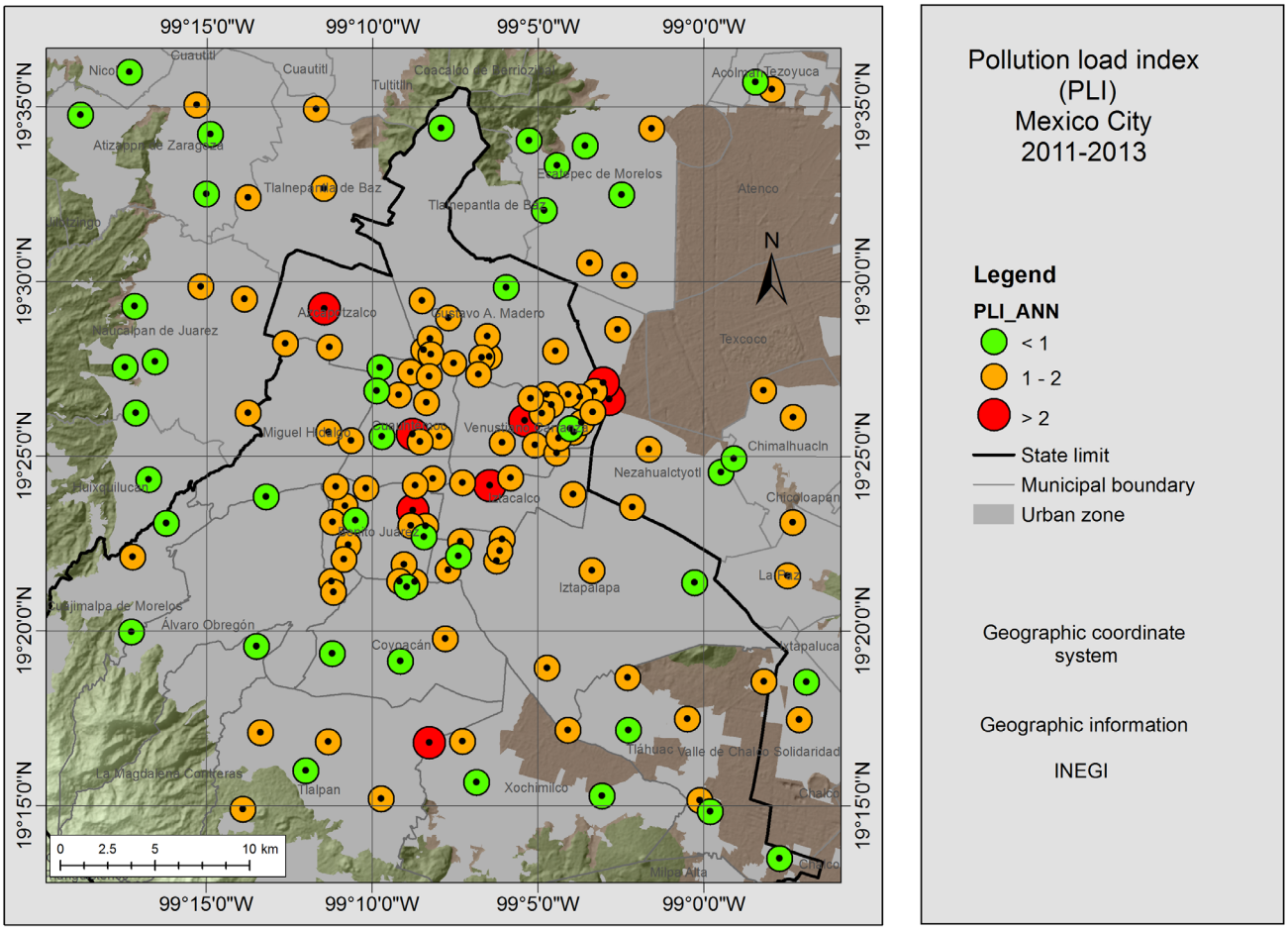


Figure 6 Map of the pollution load index of urban road dust of Mexico City, obtained by estimated with artificial neural network models (PLI_{ANN}).

suggested as a proxy for high concentrations of potentially toxic elements.

The linear regression and decision tree models are alternative methods to estimate the pollution index load in urban road dust. However, these models have less precision in samples with a very high concentration of potentially toxic elements.
The machine learning algorithms increased the prediction levels of magnetic parameters for classification and identification of problems of accumulation of potentially toxic elements in urban road dust. The method could be used in monitoring future proxy studies.

Contributions of authors

(1) Conceptualization: RC, FG, AG; (2) Analysis or data acquisition: RC, AG, FB, PQO, DA; (3) Methodologic/technical development: RC, FB, AG, PQO, DA; (4) Writing of the original manuscript: RC, AG, VK; (5) Writing of the corrected and edited manuscript: RC, AG, VK, JU; (6) Graphic design: RC, FG, AG; (7) Fieldwork: RC; (8) Interpretation: RC, FB, AG; (9) Financing: RC, AG, FB.
All authors contributed to the study conception and design. All authors read and approved the final manuscript.

Financing

To Consejo Nacional de Ciencia y Tecnología (CONACyT) by financing to proyect CB-2011-01-169915. To Secretaría de Ciencias, Tecnología e Innovación de la Ciudad de México, research project Sistema de monitoreo de la contaminación por metales pesados en polvos urbanos de la CDMX (SECITI/051/2016).

Acknowledgements

A.G. is grateful for the support provided by Universidad Nacional Autonoma de Mexico (UNAM-DGAPA) for his sabbatical fellowship. This study was partially funded by the Natural Sciences and Engineering Research Council of Canada for V.K. (NSERC grant RGPIN-2019-04780).

Conflicts of interest

The authors declare no conflicts of interest.

Handling editor

Daniele Brandt.

References

Aguilar, B., Cejudo, R., Martínez-Cruz, J., Bautista, F., Gogichaishvili, A., Carvallo, C., Morales, J., 2012, Ficus Benjamina leaves as indicator of atmospheric pollution: a Reconnaissance study: Studia Geophysica et Geodaetica, 56(3), 879–887. https://doi.org/10.1007/s11200-011-0265-1

Alizamir, M., Sobhanardakani, S., 2018, An artificial neural network - particle swarm optimization (ANN-PSO) approach to predict heavy metals contamination in groundwater resources: Jundishapur Journal of Health Sciences, 10(2), e67544, 1-8. https://doi.org/10.5812/jjhs.67544

Bautista, F., Campuzano, E., Delgado, C., Goguitchaichvil, A., 2017, Índices de adsorción de metales pesados en suelos urbanos: el caso de Morelia, Michoacán: Boletín de la Sociedad Geológica Mexicana., 69(2), 433–445. https://doi.org/10.18268/BSGM2017v69n2a8

Briffa, J., Sinagra, E., Blundell, R., 2020, Heavy metal pollution in the environment and their toxicological effects on humans: Heliyon, 6(9), e04691, 1-26. https://doi.org/10.1016/j.heliyon.2020.e04691

Canbay, M., Aydin, A., Kurtulus, C., 2010, Magnetic susceptibility and heavy-metal contamination in topsoils along the Izmit Gulf coastal area and IZAYTAS (Turkey): Journal of Applied Geophysics, 70(1), 46-57. https://

doi.org/10.1016/j.jappgeo.2009.11.002

Cejudo, R., Bautista, F., Quintana, P., Delgado, M., Aguilar, D., Goguitchaichvili, A., Morales, J., 2015, Correlación entre elementos potencialmente tóxicos y propiedades magnéticas en suelos de la ciudad de México para la identificación de sitios contaminados: definición de umbrales magnéticos: Revista Mexicana de Ciencias Geológicas, 32(1), 50–61.

Cejudo, R., Bayona, G., Goguitchaichvili, A., Cervantes, M., Bautista, F., Mendiola, F., 2021, Modelo de red neuronal para el pronóstico de la contaminación en polvos urbanos de principales vialidades de Bogotá, Colombia: Boletín de la Sociedad Geológica Mexicana, 73(1), A031020 1-18. https://doi.org/10.18268/BSGM2021v73n1a031020

Chaparro, M.A.E., Gogorza, C.S.G., Chaparro, M.A.E., Irurzun, M.A., Sinito, A.M., 2006, Review of magnetism and heavy metal pollution studies of various environments in Argentina: Earth Planets and Space, 58(suppl. 10), 1411 – 1422. https://doi.org/10.1186/BF03352637

Chaparro, M., Ramírez-Ramírez, M., Chaparro, M., Miranda-Avilés, R., Puy-Alquiza, M., Böhnel, H., Zanor, G., 2020, Magnetic parameters as proxies for anthropogenic pollution in water reservoir sediments from Mexico: an interdisciplinary approach: Science of the Total Environment, 700, 134343, 1-14. https://doi.org/10.1016/j.scitotenv.2019.134343

Chávez-Gómez, N., Cabello-López, A., Gopar-Nieto, R., Aguilar-Madrid, G., Marin-López, K., Aceves-Valdez, M., Jiménez-Ramírez, C., Cruz Angulo, M.C., Juárez-Pérez, C.A., 2017, Chronic kidney disease in Mexico and its relation with heavy metals: Revista Médica del Instituto Mexicano del Seguro Social, 55(6), 725-734.

Covarrubias, S.A., Peña Cabriales, J.C., 2017, Contaminación ambiental por metales pesados en México: problemática y estrategias de fitorremediación: Revista Internacional de Contaminación Ambiental, 33(1), 7–21. https://doi.org/10.20937/RICA.2017.33.esp01.01

Dearing J., 1999, Environmental magnetic susceptibility: Using the Bartington MS2 System: Kenilworth, UK, Chi Publishing, 43 p.

Delgado, C., Bautista, F., Gogichaishvili, A., Cortés, J.L., Quintana, P., Aguilar, D., Cejudo, R., 2019, Identificación de las zonas contaminadas con metales pesados en polvo urbano de la Ciudad de México: Revista Internacional de Contaminación Ambiental, 35(1), 81–100. https://doi.org/10.20937/RICA.2019.35.01.06

Evans, M.E., Heller, F., 2003, Environmental magnetism, principles and applications of enviromagnetics: San Diego, CA., Academic Press, 299 p. https://doi.org/10.1016/s0074-6142(03)x8309-6

Faciu, M., Ifrim, I., Lazar, G., 2012, Building and integrated environmental monitoring system for heavy metals in Romanian soils: Moldova region case study: Environmental Engineering and Management Journal, 11(12), 2185–2201. https://doi.org/10.30638/eemj.2012.272

Ihl T., Bautista, F., Cejudo, R., Delgado, M., Quintana, P., Aguilar, D., Goguitchaichvili, A., 2015, Concentration of toxic elements in topsoils of the metropolitan area of Mexico City: a spatial analysis using ordinary kriging and indicator kriging: Revista Internacional de Contaminación Ambiental, 31(1), 47–62.

Jones, F., Bankiewicz, D., Hupa, M., 2014, Occurrence and sources of zinc in fuels: Fuel, 117(Part A), 763–775. https://doi.org/10.1016/j.fuel.2013.10.005

Karimi, R., Ayoubi, S., Jalalian, A., Sheikh-Hosseini, A.R., Afyuni, M., 2011, Relationships between magnetic susceptibility and heavy metals in urban topsoils in the arid region of Isfahan, Central Iran: Journal of Applied Geophysics, 74(1), 1-7. https://doi.

org/10.1016/j.jappgeo.2011.02.009

Krzywinski, M., Altman, N., 2017, Classification and regression trees: Nature Methods, 14(8), 757–758. https://doi.org/10.1038/nmeth.4370

Legarreta-Perusquia, A., Corral Avitia, A.Y., Delgado Rios, M., Torres Pérez, J., Flores Marguez, J.P., 2016, Material particulado y metales pesados en aire en ciudades mexicanas: CULCyT , 56(1), 234-245.

Liu, D., Ma, J., Sun, Y., Li, Y., 2016, Spatial distribution of soil magnetic susceptibility and correlation with heavy metal pollution in Kaifeng City, China: Catena, 139, 53-60. https://doi.org/10.1016/j.catena.2015.11.004

Menshov, O., Spassov, S., Camps, P., Vyzhva, S., Pereira, P., Pastushenko, T., Demidov, V., 2020, Soil and dust magnetism in semi-urban area Truskavets, Ukraine: Environmental Earth Sciences, 79, 182, 1-10. https://doi.org/10.1007/s12665-020-08924-5

Morton-Bermea. O., Hernandez, E., Martinez-Pichardo, E., Soler-Arechalde, A., Lozano R., Gonzalez-Hernandez, G., Beramendi-Orosco, L., Urrutia-Fucugauchi, J., 2009, Mexico City topsoils: Heavy metals vs. magnetic susceptibility: Geoderma, 151(3–4), 121–125. https://doi.org/10.1016/j.geoderma.2009.03.019

Oudeika, M.S., Altinoglu, F.F., Akbay, F., Aydin, A., 2020, The use of magnetic susceptibility and chemical analysis data for characterizing heavy metal contamination of topsoil in Denizli city, Turkey: Journal of Applied Geophysics, 183, 104208, 1-12. https://doi.org/10.1016/j.jappgeo.2020.104208

Qian, P., Zheng, X., Zhou, L., Jiang, Q., Zhang, G., Yang, J., 2011, Magnetic properties as indicator of heavy metal contaminations in roadside soil and dust along G312 Highways: Procedia Environmental Sciences, 10(Parte B), 1370–1375. https://doi.org/10.1016/j.proenv.2011.09.219

Qing, X., Yutong, Z., Shenggao, L., 2015, Assessment of heavy metal pollution and human health risk in urban soils of steel industrial city (Anshan), Liaoning, Northeast China: Ecotoxicology and Environmental Safety, 120, 377–385. https://doi.org/10.1016/j.ecoenv.2015.06.019

Rodríguez del Águila, M.M., Benítez-Parejo, N., 2011, Simple linear and multivariate regression models: Allergologia Immunopathologia, 39(3), 159–173. https://doi.org/10.1016/j.aller.2011.02.001

Sabogal, O., Hincapié, J.D., Santa, J.J., Escobar, J.W., 2015, Modelos de regresión lineal para estimación de tiempos de viaje en sistemas de transporte masivo: Ciencia e Ingeniería Neogranadina, 25(1), 77-89. https://doi.org/10.18359/rcin.434.

Secretaria de Medio Ambiente y Recursos Naturales (SEMARNAT), 2007, Norma Oficial Mexicana que establece criterios para determinar las concentraciones de remediación de suelos contaminados por arsénico, bario, berilio, cadmio, cromo hexavalente, mercurio, níquel, plata, plomo, selenio, talio y/o vanadio (NOM-147-SEMARNAT/SSA1-2004): Ciudad de México, Diario Oficial de la Federación, 2 de marzo de 2007, 1-69. Disponible en <https://www.gob.mx/profepa/documentos/norma-oficial-mexicana-nom-147-semarnat-ssa1-2004>, consultado 27 de marzo de 2023.

Tchounwou, P., Yedjou, C., Patlolla, A., Sutton, D., 2012, Heavy Metal Toxicity and the Environment, in Luch, A., (ed.), Molecular, clinical and environmental toxicology: Volume 3 Environmental toxicology, Experientia supplementum volume 101: Berlin, Germany, Springer Basel, 133–164. https://doi.org/10.1007/978-3-7643-8340-4_6

Tomlinson, D.L., Wilson, J.G., Harris, C.R., Jeffrey, D.W., 1980, Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index: Helgoländer Meeresuntersuchungen, 33, 566–575.

https://doi.org/10.1007/BF02414780

Trujillano, J., March, J., Sorribas, A., 2004, Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina: Medicina Clínica, 122(S1), 59–67.

Wang, G., Chen, J., Zhang, W., Chen, Y., Ren, F., Fang, A., Ma, L., 2019. Relationship between magnetic properties and heavy metal contamination of street dust samples from Shanghai, China: Environmental Science and Pollution Research, 26, 8958–8970. https://doi.org/10.1007/s11356-019-04338-4

Wickham, H., Grolemund, G., 2016, R for Data Science: Ca., USA, O'Reilly Media, Inc., 518 p.

Zdoroveyshchev, A.V., Vikhrova, O.V., Demina, P.B., Dorokhin, M.V., Kudrin, A.V., Temiryazev, A.G., Temiryazeva, M.P., 2019, Micromagnetic and magnetooptical properties of ferromagnetic/heavy metal thin film structures: Physics of the Solid State, 61(9), 1577–1582. https://doi.org/10.1134/S1063783419090294

Zhang, H., Wu, P., Yin, A., Yang, X., Zhang, M., Gao, C., 2017, Prediction of soil organic carbon in an intensively managed reclamation zone of eastern china: a comparison of multiple linear regressions and the random forest model: Science of the Total Environment, 592, 704–713. https://doi.org/10.1016/j.scitotenv.2017.02.146

Zhang, H., Yin, S., Chen, Y., Shao, S., Wu, J., Fan, M., Chen, F., Gao, C., 2020, Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China: Journal of Cleaner Production, 273, 122858, 1-10. https://doi.org/10.1016/j.jclepro.2020.122858