17th International Conference on Environmental Science and Technology

Athens, Greece, 1 to 4 September 2021

Artificial Neural Network (ANNs) for predicting petroleum hydrocarbons from heavy metals contaminated soils around fuel stations

BONELLI M.G.[1,2*], MANNI A.[3,4] and SAVIANO G.[4]

[1]Programming and Grant Office Unit (UPGO), Italian National Research Council (CNR), Piazzale Aldo Moro 7, 00185 Rome, Italy
[2]InterUniversity Consortium Georesources Engineering (CINIGeo), Corso Vittorio Emanuele II 244, 00186 Rome, Italy
[3]Chemical Research 2000 S.r.l., Via Santa Margherita di Belice 16, 00133 Rome, Italy
[4]Department of Chemical, Materials and Environmental Engineering (DICMA), "La Sapienza" University of Rome, Via Eudossiana 18, 00184 Rome, Italy

*corresponding author:
e-mail: mariagrazia.bonelli@cnr.it

Abstract Petrol stations are classified as a dangerous source of pollution for the human population due to the toxicity of emissions from evaporated vehicle fuels and fuel spillages. The contaminants released in the environment are mainly complex mixtures of petroleum hydrocarbon compounds (PHCs) and heavy metals, especially lead. Lead phased out as a fuel additive by the dawn of the 21st century, but some soils near old or long-standing gas stations have been contaminated.
Contamination found at these sites, affecting groundwater, drinking water, and the soil, can run deep and spread over an area that extends well beyond the site's border.
The correlation between heavy metals and heavier petroleum hydrocarbons (C>12) in soils from an urban area near petrol stations has been studied, looking to predict the organic concentration through inorganic contaminants concentration values. Metals were analyzed by ICP-OES and FP-XRF (Field Portable XRF), while PHCs were analyzed by GC/FID. No linear statistical correlation has been proved between Pb, Cu, Mn, V, Zn, Sn, Fe, and PHCs. The ANNs model, instead, has been demonstrated to have the capability to determine the relationships between organic and inorganic contaminants, allowing an accurate prediction of PHCs (C>12) (R^2=0,86).
Keywords: Artificial Neural Network predictions, heavy metals, field portable XRF, petroleum hydrocarbon compounds

1. Introduction

Hundreds of hydrocarbon compounds constitute PHCs. They are often released in city petrol stations' surroundings and could determine serious health risks for fuel station operators. This environmental threat is also joint with heavy metal dispersion. It can be relevant, especially for Lead (Pb), which was used as a fuel additive for a long time. Also, due to the spillage from underground storage tanks, by the actions of volatilization, biodegradation, photo-decomposition, chemical oxidation, bioaccumulation, dispersion, diffusion, binding to the soil, and leaching groundwater, both contaminations can be easily spread out of the fuel station site creating a potential risk for the population living in their surroundings (Onutu et al., 2018; Konwuruk et al., 2021). If used for irrigation purposes, contaminated waters may provoke topsoil contamination (Balseiro-Romero et al., 2016). Several analytical techniques can be used to determine both classes of contaminants. Heavy metals can be analyzed by legally accepted methods such as ICP (OES or MS) or AA (FL or GF). FP-XRF has received particular attention as a low-cost and fast turnaround analytical technique suitable for screening heavy metals in soils, minerals, and wastes (EPA 5200). Results are not always comparable due to a large number of sample preparation differences (Laperce, 2021). PHCs can be analyzed by GC/FID (EPA 8015D), IR (EPA 418.1), or Gravimetry (EPA 413.2).

When large areas are contaminated, it may be really useful to have a screening method available to establish the contamination borders and determine the presence of hot spots. Immunoassay has been used for this purpose (EPA SW846 Test Method 4030).

Artificial neural networks (ANNs) are statistical algorithms that perform statistical modeling used in alternative to traditional predictive methodology, such as multilinear regression, for approximating complex relationships between the input and output variables with a non-linearity optimization (Haykin, 1994). The net's architecture is a directed graph with multiple layers of nodes, fully connected. The most used algorithm is Multi-Layer Perceptron (MLP), with three or more layers – one for input, one or more hidden layers, and one layer for output. The concept is sharing the information from the input to hidden layers to forecast output variables through an activation function in a training process that analyzes

the unknown relationships between predictors and the target variable. The MLP accuracy is estimated by comparing the value of the Root Mean Square Error of the training set and test set. The model performance is assessed by calculating the coefficient of determination (R^2) between the actual values and the values provided by the model (Twomey et al., 1995).

The ANNs are often used in environmental analysis, especially in screening procedures, to analyze pollution sources in large areas, estimating difficult and expensive to detect contaminants from other easy measurable pollutants (Tao et al., 2019). The present study has aimed to propose an alternative PHCs screening method by coupling FP-XRF's information on metals, and machine learning by Artificial Neural Networks (ANNs).

2. Materials and Methods

2.1. Sampling and analytical procedure

In this study, samples were collected in a strongly urbanized area of Palermo, Sicily, in the surrounding of 3 different petrol stations, at 0,1m depth. They were analyzed by GC/FID for organic compounds and FP-XRF for heavy metals. Pb was also analyzed by ICP-OES as a confirmatory method. All standards used were purchased from an ISO17034 producer (O2Si, Charleston, NC, USA).

The FP-XRF intrusive preparation procedure was adopted (11.4 EPA 6200).

The statistical analysis has been performed by the package IBM-SPSS v. 25

2.2. The Field Portable – XRF tool

XRF measurements were performed using a Skyray Genius 9000 Portable XRF Spectrometer (Jangshu Instrument Co. Ltd., Yushan, Kushan, China). A Silver (Ag) thin window X-ray tube was the excitation source (X-ray generator 5–40 kV with 1–100 μA). Samples were analyzed for 90 sec with a software, combinating Fundamental Parameters (FP) and External Calibration (EC) modes. During intrusive measurements, the pXRF gun was mounted upside down on a stand. Soil samples, contained in polyethylene bags, were directly placed over the sensing window and gently pressed by the weight of a Teflon bar; the same bar also constitutes the XRF blank sample.

2.3. Artificial Neural Networks

A Multi-Layer Perceptron model (§. 1) was trained to predict the PHCs concentration (output) through FP-XRF heavy metal values (input).

ANN architecture was built using the optimal design obtained through Taguchi's factorial analysis to identify the best parameters' set (number of hidden layers, number of hidden neurons, choice of input factors, training algorithm parameters, etc.) (Manni et al. 2021). The MLP model performed is a 7-4-1 net, based on three layers with 7 inputs, 4 hidden variables and 1 output. The original dataset has been normalized using the formula:

$$p_i = \frac{x_i - min\,(x)}{max\,(x) - min\,(x)}$$

A hyperbolic tangent was proved to be the best form of the activation function for both the hidden layer and the output node (figure 1).

3. Results and Discussion

4. To minimize the variability of the XRF measurements, each soil sample has been analyzed multiple (n) times until the difference between the geometric mean of the n analysis and the geometric mean of the n-1 analysis is below a certain level, typically 10%, for all the elements analyzed. After each analysis, the sample has been slightly moved over the XRF sensing window to maximize the surface size of the analyzed sample. Once the stability of the measurements has been assessed, the best elemental concentration to be used between median, geometric, and arithmetic means in these data sets was also verified. Typically, geometric mean is used to minimize the effect of outliers (Das & Imon, 2013). We have found that the stability of the measurements does not determine any particular difference in the results provided from the three means. Table 1 shows the results for one sample.

5. 10 soils sample were analyzed for Pb, Cu, Mn, V, Zn, Sn, Fe, and PHCs (C>12). Table 2 shows the descriptive statistics for each variable, including minimum, maximum, mean, and standard deviation of the distribution. 7 soil samples were also analyzed by ICP-OES for Pb to compare with the FP-XRF data. The results were found in good agreement, as shown in Figure 2. All samples were analyzed for PHCs (C>12) by GC/FID. The correlation matrix, reported in Table 3, demonstrated no linear relations between metals and PHCs. There is, instead, multicollinearity between the metals. Principal Component Analysis (PCA) also indicates that the multicollinearity phenomena are related to different contamination sources, as reported in Figure 3. Metals are spread into two major clusters, which are coincident with those showing multicollinearity. PC1 is positively correlated with Pb (XRF values), Zn, and Cu. Therefore, this component could be identified as emissions from tire and brake wears. PC2 is positively correlated with Fe and Mn. Thus, the source identified for PC2 is vehicular exhausted fuel emissions. PHCs do not belong to either cluster, and this situation does not allow any prediction of its concentration from metals' data.

6. ANNs were applied to forecast PHCs from metal concentrations. The MLP model provides reliable

predictions with an R^2=0,86, and a comparison between predicted and observed PHCs values is shown in figure 4.

This method can be applied after a sufficient number of samples have been analyzed for the dependent variable, in this case PHCs, according to the "legally accepted methods" to be able to train the ANNs. After this preliminary step, metals' data from successive samples can be used for the screening of PHCs. Whenever hotspots are determined, a confirmatory analysis will be necessary. Confirmatory analysis can also be used to increase the dataset extension and improve the ANNs' results' quality in a recursive procedure, shown in Figure 5 (Bonelli MG et al., 2017).

4. Conclusions

Artificial Neural Networks have shown an excellent predictive capacity for PHCs from metals' concentrations determined by FP-XRF. Their joint use may constitute an easy, inexpensive, and fast turnaround screening analytical method, alternative to the immunoassay. The method, however, could only be used after a consistent number of "legally accepted PHCs analysis" has been already performed, and will be improved after "confirmatory analysis" are taken on "hot spot screened samples".

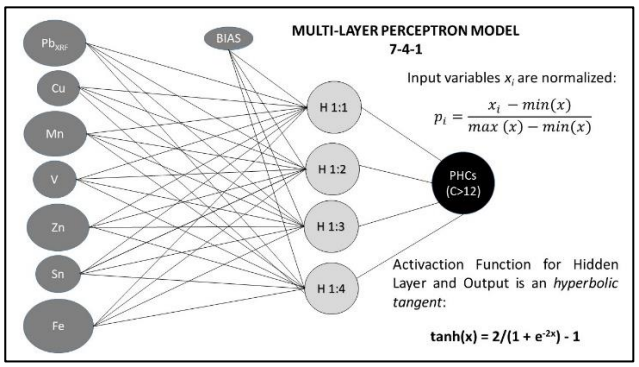

Figure 1 Three-layer MLP, with 7 input nodes, 4 hidden nodes, and 1 output nodes

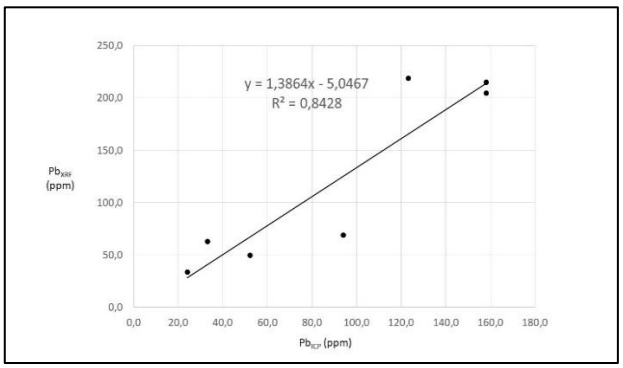

Figure 2. Comparison between ICP-OES for Pb and FP-XRF data

Table 1. Example of the maximization of variability in a sample soil

	XRF values (ppm)			Geometric mean (ppm)			Difference of geometric mean (%)		
	Cu	Zn	Pb	Cu	Zn	Pb	Cu	Zn	Pb
	85	149	240	85,00	149,00	240,00	0,00	0,00	0,00
	84	155	226	84,50	151,97	232,89	0,59	-1,99	2,96
	86	151	238	85,00	151,65	234,58	-0,59	0,21	-0,73
	86	158	270	85,25	153,21	242,98	-0,29	-1,03	-3,58
	102	147	241	88,36	151,95	242,58	-3,65	0,82	0,16
Geometric mean	88	152	243						
Median	86	151	240						
Arithmetic mean	89	152	243						

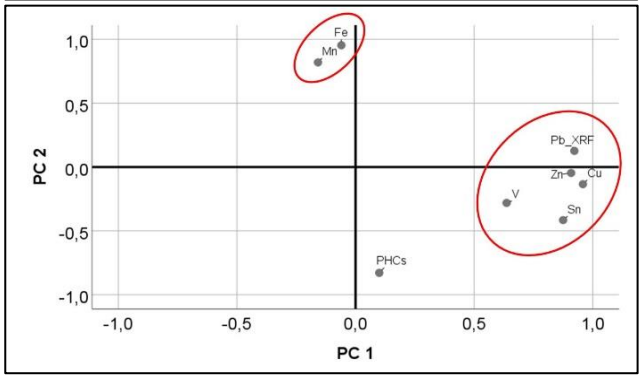

Figure 3. PCA results

Table 2. Descriptive statistics

VARIABLES	MIN	MAX	MEAN	STANDARD DEVIATION
Pb XRF	36,4	242,6	128,4	78,6
Cu	12,2	160,1	61,6	51,6
Mn	3,7	582,3	333,2	215,8
V	77,3	94,0	88,5	5,7
Zn	59,6	268,0	148,6	72,1
Sn	24,2	68,4	41,2	16,8
Fe	8.510,3	36.660,4	20.575,7	9.837,9
C>12	5,0	48,8	20,1	17,0

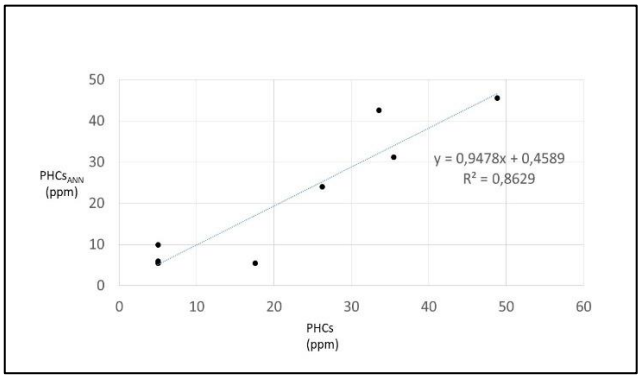

Figure 4. Comparison between predicted and observed PHCs values

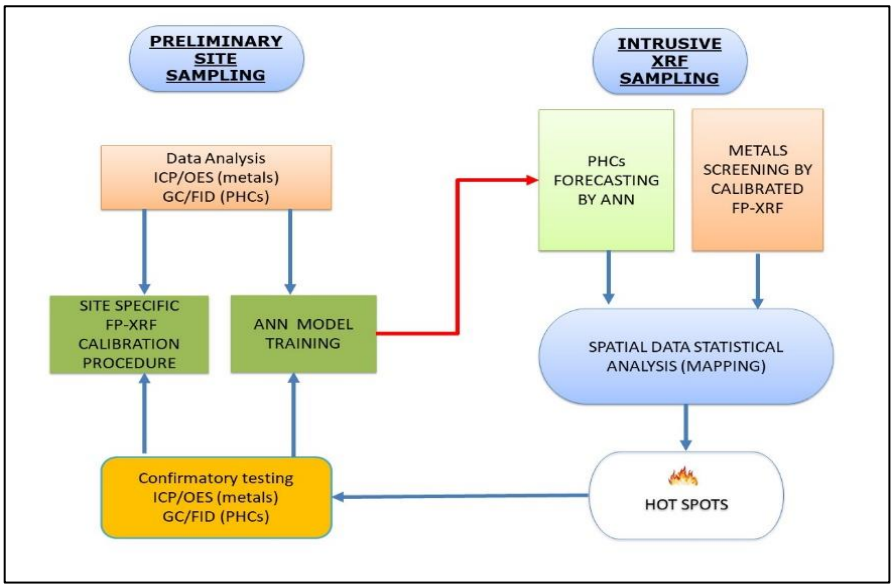

Figure 5. Recursive procedure to screen the "hot spot" screening by FPXRF metals determination coupled to ANN forecasting.

References

Onutu I., Tita M. (2018), Soil contamination with petroleum compounds and heavy metals – case study, *Scientific Papers. Series E. Land Reclamation, Earth Observation & Surveying, Environmental Engineering. Vol. VII*, 140-145

Balseiro-Romero M., Macías F. and Monterroso C. & (2016), Characterization and fingerprinting of soil and groundwater contamination sources around a fuel distribution station in Galicia (NW Spain), *Environ Monit Assess* 188:292.

Konwuruk N., Sheringham Borquaye L., Darko G., Dodd M. (2021), distribution, bioaccessibility and human health risks of toxic metals in peri-urban topsoils of the Kumasi Metropolis, *Scientific African* ,**11**, *e00701*

Laperche V. and Lemière B. (2021), Possible Pitfalls in the Analysis of Minerals and Loose Materials by Portable XRF, and How to Overcome Them, *Minerals*, **11**, 33.

EPA 6200 Method: Field Portable X-RAY Fluorescence Spectrometry for the determination of elemental concentration in soil and sediment.

Haykin, S. (1994) Neural Networks a Comprehensive Foundation, IEEE Computer Society Press.

Twomey J. M. and Smith A. E. 1995 Performance Measures, Consistency and Power for Artificial Neural Network Models Mathl. Comput. Modelling 21, pp. 243-258.

Tao H., Liao X., Zhao D., Xuegang G. and Cassidy D.P. 2019, Delineation of soil contaminant plumes at a co-contaminated site using BP neural networks and geostatistics, Geoderma Volume 354, 113878, ISSN 0016-7061

Das K.R., Imon AHRM, 2013, Geometric median and its application in the identification of multiple outliers, *Journal of Applied Statistics*, **41**(4), 817-831

Manni, A., Saviano, G., Bonelli, M.G. 2021, Optimization of the ANNs Predictive Capability Using the Taguchi Approach: A Case Study, Mathematics 9, 766-781

Bonelli, M.G., Ferrini M., Manni, A., 2017 Artificial neural networks to evaluate organic and inorganic contamination in agricultural soils, Chemosphere 186 124 - 131