DIOXIN AND INORGANIC MICROPOLLUTANTS CONTAMINATION ANALYSIS: A STATISTICAL MULTI-METHOD APPROACH

Bonelli MG [1]*, Ferrini M[1], Guerriero E[2], Manni A[3]

[1]Dept. ICMA, University of Rome "La Sapienza", Via Eudossiana 18, 00198 Rome, Italy; [2] CNR – Institute for Atmospheric Pollution, Via Salaria, Km 29.300, 00015 Monterotondo Scalo, (Rome), Italy; [3]Spectra 2000 Srl, Via Santa Margherita di Belice 16, 00133 Rome, Italy;

Introduction

Current development of the society is characterized by a continuous and progressive transformation of the agricultural and industrial production. These processes are accomplished through the use of organic and inorganic contaminants and, often, their release in the environment may generate pollution and toxicity for plants, animals as well as humans when overcome certain concentration thresholds.

Among the organic contaminants, dioxins are one of the twelve classes of stable and persistent organic pollutants (Persistent Organic Pollutants - POPs), internationally recognized by UNEP (United Nations Environment Programme) to be toxic for both the environment and humans[1]. However, over all the species of dioxins only 7 are highly toxic. Their chemical-physical characteristics are according to their chlorination's degree. Among all the 75 congeners, the most toxic isomers are the ones with chlorine in positions 2, 3, 7, and 8, and, in particular, the 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)[2], which owns a lesser chlorination degree while, on the contrary, the octachlorodibenzo-p-dioxins (OCDD), have a higher chlorination degree and a lower toxicity.

There are different types of inorganic pollutants; between them heavy metals, as well as the above mentioned organic contaminants, can present problems of toxicity in addition to long persistence time. Among them Al, Fe, Ba, Be, Cd, Co, Cr, Hg, Mo, Ni, Pb, Cu, Sn, Ti, Ta, V, Zn together with to certain metalloids with similar properties to those of heavy metals such as As, Bi and Se.

The present study takes into consideration topsoil samples from an site in the surroundings of a former chemical industry.

Some authors investigated concentrations of As, Cd, Cu and Zn in agricultural soil originating from applications of pesticides, animal manures and fertilizers[3,4]. Other studies considered the heavy metals and organic pollutant combination in agricultural soils near emission sources (e.g. solid waste incinerators, urban sites or chemical industries)[5].

The aim of this study was to use a statistical multi-method approach to analyse relationship existing between organic and inorganic micropollutants detected in the above-mentioned soil samples, in order to formulate a statistically valid model for the mathematical representation of the results that could also be used for predictive purposes.

Materials and methods

75 surface soils samples has been taken from an agricultural site characterized by high levels of pollution. They have been analyzed for heavy metals (Ca, Fe, Mg, K, As, Cd, Cr, Mn, Hg, Ni, Pb, Cu, Sn, Zn) as well as PCDDs. Results of the latter one have been consolidated in 5 families of congeners, namely TCDD, PeCDD, HxCDD, HpCDD and OCDD. The smallest and the largest concentration of each element, mean and standard deviation are summarized in table 1. OCDD concentration is much higher than the other families of congeners: if the sources are herbicides pentachlorophenol (PCP) and chloronitrophen (CNP) contain PCDD/Fs impurities, with the dominant congener in PCP being OCDD[6], moreover , higher degree of chlorination as well as its lesser bioavailability for animal and vegetable organisms. Furthermore, OCDD are "heavier" and more persistent than others, and so they are less subject to dissipation and photochemical degradation[7]. For the above reason, in the present work, OCDD were analyzed separately from the other families of congeners, mentioned above. Their concentrations have been summed and taken into consideration as a distinct variable, called PCDD_tot_no_OCDD. Therefore, within this study, separate relations between heavy metals and PCDD_tot_no_OCDD or OCDD have been assessed.

All statistical calculations and elaborations were performed using the software package SPSS v. 22.

More information about the relation model between variables have been obtained using artificial neural networks technique (ANN) combined with multiple linear regression (MLR). Specifically, ANN approach has been used

for the selection of explicative variables to consider in MLR, to generate a correct mathematical model explaining how independent variables influence dependent variable.

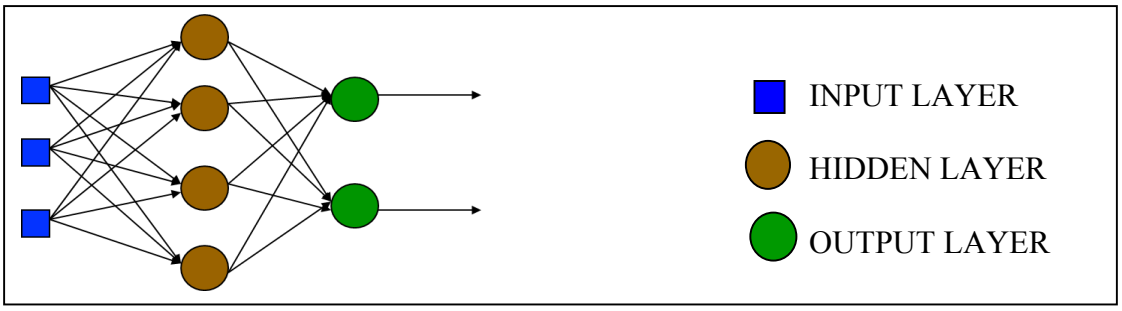
Table 1 Descriptive Statistics: PCDD isomers and heavy metals

Variable	Min	Max	Mean	Std. Dev.	Variable	Min	Max	Mean	Std. Dev.
Ca	11400	52800	28467,5	12292,9	**Sn**	1,9	14,4	4,6	2,0
Fe	16400	32720	24581,2	4032,6	**Cr**	21,8	53,2	31,6	6,4
Mg	9040	26880	17740	4052,5	**Hg**	0,4	12,2	2,5	1,9
K	4240	11520	7145	1878,2	**Cd**	0,30	1,0	0,5	0,1
Mn	563,2	2016	1080,1	386,2	**TCDD**	-	445,60	21,13	58,35
Ni	19,7	114,4	32,6	12,2	**PeCDD**	0,10	159,20	11,34	26,53
Zn	61,8	302,4	138,2	34,6	**HxCDD**	0,10	280,80	20,09	45,19
Pb	44,5	564,8	116,8	71,5	**HpCDD**	2,90	184,80	22,91	26,33
Cu	35,0	341,6	71,4	38,0	**OCDD**	15,10	1864,00	207,54	274,20
As	20,8	175,2	47,4	23,3					

Metal concentrations in ppm (mg/kg); PCDDs concentration in ng/kg

Artificial neural networks is a statistical methodology being able to model complicated and non-linear relationships, but it works as a black box: it can also provide accurate results from a series of very different input data but it cannot explain why and how it got these results. A neural network can approximate a wide range of statistical models without requiring to hypothesize in advance any relationships between the dependent and independent variables. Instead, the form of the relationships is determined during the learning process. If a linear relationship between the dependent and independent variables is appropriate, the results of the neural network should closely approximate those of the linear regression model. If a nonlinear relationship is more appropriate, the neural network will automatically approximate the "correct" model structure.
Neuron (node) is the basic processing unit in neural networks. Neural networks impose minimal demands on model structure and assumptions, but it is necessary to correctly choose the general network architecture, that consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. In this study Multi Layer Perceptron (MLP) model, consisting of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes[8], has been preferred: it is a function of predictors (also called inputs or independent variables) that minimize the prediction error of target variables (also called outputs).
Fig. 1 it shows an example of MLP's architecture, with 3 input nodes, 4 hidden nodes and 2 output nodes.

Fig. 1 MLP's architecture 3-4-2



The network's learning process occurs in the following consecutive phases:

a) Training phase, where original sample set was divided into three subsets: training set, test set and holdout set. Model is trained by pairing the input with expected output.
b) Validation phase, where the model was taught based on the training set and further validated using test set.
c) Test phase, where the accuracy of the model was estimated to be based on the value of root mean square error (RMSE) of prediction calculated for the holdout set.

Results and discussion

The first step was to measure how variables relate to each other, determining Pearson's correlation coefficients of PCDD_tot_no_OCDD or OCDD and heavy metals. The significance was established at $p<0,05$ (table 2).

High and significative correlations have been determined between PCDD_tot_no_OCDD and As, Hg, Ni, Cu, Zn. Moreover, the highly observed correlation among several independent variables (metals) indicates multicollinearity. Instead, the results of OCDD suggest that the relationship with heavy metals is non-linear, but it does not necessarily mean that this connection is non-existent.

Table 2 Correlations matrix

		Ca	Fe	Mg	K	As	Cd	Cr	Mn	Hg	Ni	Pb	Cu	Sn	Zn	OCDD	PCDD_tot_no_OCDD
Ca		1	-,324**	,669**	-,468**	,509**	,214	,138	,821**	,083	,308*	-,132	,346**	,023	,162	,067	,296*
	Sig. (2-tailed)		,009	,000	,000	,000	,090	,276	,000	,514	,013	,300	,005	,859	,201	,597	,017
Fe		-,324**	1	,355**	,836**	-,337**	,241	,415**	-,251*	-,121	,082	,133	-,101	,354**	,479**	-,031	-,134
	Sig. (2-tailed)	,009		,004	,000	,007	,056	,001	,045	,342	,521	,293	,429	,004	,000	,808	,293
Mg		,669**	,355**	1	,178	,162	,377**	,428**	,532**	-,058	,342**	-,110	,215	,301*	,454**	,068	,102
	Sig. (2-tailed)	,000	,004		,160	,202	,002	,000	,000	,651	,006	,388	,088	,016	,000	,596	,423
K		-,468**	,836**	,178	1	-,465**	,148	,466**	-,479**	-,070	,048	,148	-,136	,356**	,384**	-,034	-,181
	Sig. (2-tailed)	,000	,000	,160		,000	,242	,000	,000	,582	,704	,243	,285	,004	,002	,789	,153
As		,509**	-,337**	,162	-,465**	1	,495**	,291*	,528**	,676**	,727**	,073	,803**	,150	,474**	,150	,608**
	Sig. (2-tailed)	,000	,007	,202	,000		,000	,020	,000	,000	,000	,567	,000	,236	,000	,238	,000
Cd		,214	,241	,377**	,148	,495**	1	,791**	,429**	,427**	,782**	,069	,747**	,520**	,734**	,101	,440**
	Sig. (2-tailed)	,090	,056	,002	,242	,000		,000	,000	,000	,000	,587	,000	,000	,000	,426	,000
Cr		,138	,415**	,428**	,466**	,291*	,791**	1	,323**	,400**	,695**	,119	,611**	,622**	,773**	,129	,378**
	Sig. (2-tailed)	,276	,001	,000	,000	,020	,000		,009	,001	,000	,350	,000	,000	,000	,308	,002
Mn		,821**	-,251*	,532**	-,479**	,528**	,429**	,323**	1	,059	,333**	-,111	,374**	,139	,192	,061	,249*
	Sig. (2-tailed)	,000	,045	,000	,000	,000	,000	,009		,643	,007	,383	,002	,273	,128	,632	,047
Hg		,083	-,121	-,058	-,070	,676**	,427**	,400**	,059	1	,690**	,246*	,789**	,308*	,669**	,205	,687**
	Sig. (2-tailed)	,514	,342	,651	,582	,000	,000	,001	,643		,000	,050	,000	,013	,000	,104	,000
Ni		,308*	,082	,342**	,048	,727**	,782**	,695**	,333**	,690**	1	,186	,938**	,487**	,805**	,120	,592**
	Sig. (2-tailed)	,013	,521	,006	,704	,000	,000	,000	,007	,000		,142	,000	,000	,000	,346	,000
Pb		-,132	,133	-,110	,148	,073	,069	,119	-,111	,246*	,186	1	,212	,438**	,170	-,061	,094
	Sig. (2-tailed)	,300	,293	,388	,243	,567	,587	,350	,383	,050	,142		,093	,000	,178	,630	,458
Cu		,346**	-,101	,215	-,136	,803**	,747**	,611**	,374**	,789**	,938**	,212	1	,427**	,747**	,149	,733**
	Sig. (2-tailed)	,005	,429	,088	,285	,000	,000	,000	,002	,000	,000	,093		,000	,000	,240	,000
Sn		,023	,354**	,301*	,356**	,150	,520**	,622**	,139	,308*	,487**	,438**	,427**	1	,558**	,091	,264*
	Sig. (2-tailed)	,859	,004	,016	,004	,236	,000	,000	,273	,013	,000	,000	,000		,000	,472	,035
Zn		,162	,479**	,454**	,384**	,474**	,734**	,773**	,192	,669**	,805**	,170	,747**	,558**	1	,175	,562**
	Sig. (2-tailed)	,201	,000	,000	,002	,000	,000	,000	,128	,000	,000	,178	,000	,000		,167	,000
OCDD		,067	-,031	,068	-,034	,150	,101	,129	,061	,205	,120	-,061	,149	,091	,175	1	,117
	Sig. (2-tailed)	,597	,808	,596	,789	,238	,426	,308	,632	,104	,346	,630	,240	,472	,167		,356
PCDD_tot_no_OCD		,296*	-,134	,102	-,181	,608**	,440**	,378**	,249*	,687**	,592**	,094	,733**	,264*	,562**	,117	1
	Sig. (2-tailed)	,017	,293	,423	,153	,000	,000	,002	,047	,000	,000	,458	,000	,035	,000	,356	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Two models were identified using MLP, one for each dependent variable, PCDD_tot_no_OCDD (Model A) and OCDD (Model B). Results are summarized in fig. 2.

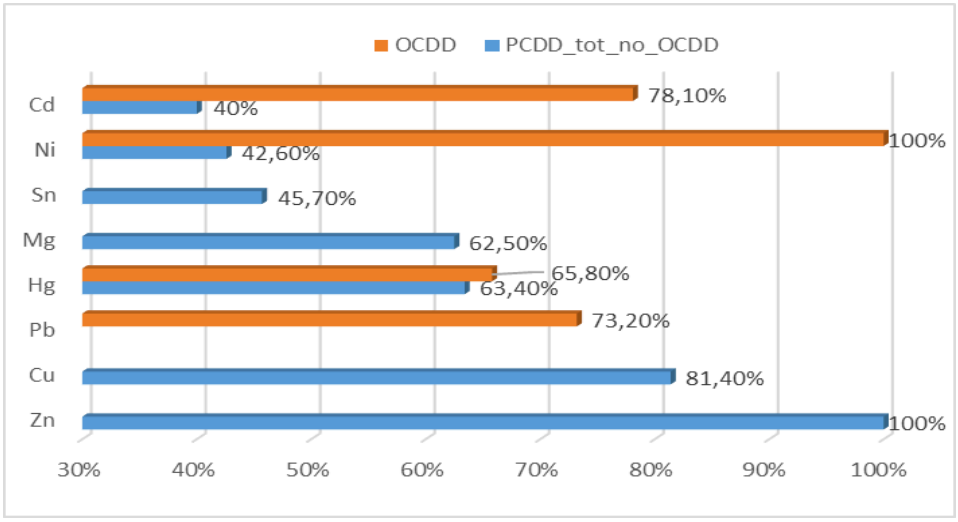


Fig. 2 Normalized importance matrix value (in percent)

Model A is based on 14-1-1 perceptron architecture, with 53,1% sample's units in training set, 15,6% in test set and 31,3% in holdout set, and Model B is based on 14-2-1 perceptron architecture, with 45,3% sample's units in training set, 25% in test set and 29,7% in holdout set. The models accuracy, based on RMSE's holdout set A is 89,4% and on RMSE's holdout set B is 94,2%. Normalized importance matrix (fig. 2) shows that Zn, Cu, Hg, Mg, Sn, Ni, and Cd are the independent variables that more than others have a relationship with PCDD_tot_no_OCDD, and Ni, Cd, Pb and Hg are the independent variables that more than others have a relationship with OCDD.

Thereafter, MLR has been applied between dioxins and the set of variables selected by the two above mentioned networks. Moreover, considering the multicollinearity between the independent variables (as mentioned above), a stepwise linear regression has been conducted, in order to select the significant contributors to estimated dioxin concentrations.

Inasmuch as not a single metal has a significative correlation with OCDD, a new variable, called log_OCDD (OCDD's logarithmic transformed), was used to conduct linear regression, in agreement with another reference[9] showing that the model with logarithm of dioxins concentrations fit better than that one with untransformed concentrations data. For each of the two dependent variables, two models have been determined. Their regression coefficients (R^2) have been reported in table 3.

Table 3 Linear regression models

Dependent variable	*Independent Variables*	*R^2 value*
PCDD_tot_no_OCDD		
Linear regression	Zn, Cu, Hg, Mg, Sn, Ni, Cd	0,657
Stepwise regression	Cu, Ni, Zn	0,641
log_OCDD		
Linear regression	Ni, Cd, Pb, Hg	0,265
Stepwise regression	Pb, Hg	0,219

R^2 values are significant only for regression models fitted to PCDD_tot_no_OCDD. Linear relationship between "light" dioxins and the set of metals selected by the neural network can be assumed; it "explains" the 66% of PCDD_tot_no_OCDD's variability. Stepwise regression improves the result, selecting only Cu, Ni and Zn as significant contributors for predictive purposes.

Instead, OCDD's regression model confirm that there is not a significant influence of the metals selected by the neural network; Pb and Hg alone explain the 22% of "heavy" dioxin's variability. It demonstrates that the relationship between OCDD and the above mentioned metals is not linear. To find a good mathematical representation of this connection will be the aim of future studies.

The findings of the present work show that the determination of few metals, in example by the mean of a portable XRF, could lead to a very fast and inexpensive on-site screening procedure for the determination of tetra to hepta chlorodibenzodioxins.

References:

1. APAT - Agenzia per la protezione dell'ambiente e per i servizi tecnici, (2006) workpaper
2. Poland A., KnutsonDenison J.C., (1982) *Annual Review of Pharmacology and Toxicology vol 22, 517-554*
3. Chen T.B., Wong J.W, Zhou H.Y. and Wong M.H. (1997) *Enviromental Pollution, 96, n.1, 61-68*
4. Huang S.S., Liao Q.L., Hua M., Wu X.M., Bi K.S., Yan C.Y., Chen B., Zhang X.Y. (2007) *Chemosphere, 67, 2148–2155*
5. Vives I, Müller A., Umlaf G., Christoph E.H., Mariani G., Skejo H. Cenci, R.M., Sena F. and Beone G.M. (2008), *Enviromental International 34, 994-1000*
6. Masunaga, S., Takasuga, T., Nakanishi, J., (2001) .. *Chemosphere, 44, 873–885*
7. Magdalena Urbaniak (2013).http://www.intechopen.com/books/biodegradation-engineering-and-technology/biodegradation-of-pcdds-pcdfs-and-pcbs
8. IBM SPSS, Neural networks 22.0 (2013), http://www-03.ibm.com/software/products/it/spss
9. Hattemer H.A. and Travis C. *Health effects of municipal waste incineration* (1991), *CRC Press, Inc*.