Principal Components Analysis and spatial analysis integration for enhanced assessment of pollution emission sources

Maria Grazia Bonelli\textsuperscript{1,2,4} and Andrea Manni\textsuperscript{3}

\textsuperscript{1} Dept. ICMA, University of Rome “La Sapienza”, Via Eudossiana 18, Rome, Italy, 00198; \\
\textsuperscript{2} CINIGeo, Corso Vittorio Emanuele II, 244, Rome, Italy, 00186; \\
\textsuperscript{3} Chemical Research 2000 Srl, Via Santa Margherita di Belice 16, Rome, Italy, 00133. \\
\textsuperscript{4} Email: mariagrazia.bonelli@uniroma1.it

Abstract. Principal Component Analysis (PCA) is a statistical technique to reduce the dimension of the original dataset. This method is often used in environmental analysis to detect the pollution sources in ambient air and in soil. However, PCA results can be difficult to read, especially when some variables seem to correlate to any principal component. In this case, it should be helpful to combine PCA with another method to better explain the uncertainty. In this study, the use of PCA coupled to Kriging algorithm has been investigated to better identify the emission sources responsible for soil pollution in an agricultural field near an industrial site in the surroundings of the city of Brescia, in Northern Italy, where organic and inorganic micropollutants concentrations in soils are beyond the legal limits. Three clusters of variables have determined as many as the emission sources in the investigated area. PCs condensed the information from most heavy metals and organic compounds. It has been demonstrated that the spatial assessment can be useful to find emission sources for those elements where it has not been possible by PCA analysis only.

1. Introduction
Agricultural soils close to industrial sites are characterized by distinct groups of organic and inorganic contaminants, which depend on different emission sources and anthropogenic activities. Among organic micropollutants, dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs) and polychlorinated biphenyls (PCBs) are highly toxic and considered as persistent organic pollutants (POPs) [1]. Based on the number of chlorine atoms in the chemical structure and their positions, PCDD/Fs consist of 210 congeners. Among them, 17 congeners have chlorine atoms at least in the positions 2, 3, 7, and 8 of the parent molecules: they show similar toxicological responses with the most toxic being the 2,3,7,8-tetrachlorodibenzo-p-dioxin (2,3,7,8-TCDD) and the 2,3,7,8-tetrachlorodibenzo-p-furan (2,3,7,8-TCDF) [2].

Twelve of the 209 congeners of PCBs have the same toxicological properties of dioxins and furans, and are internationally recognized by UNEP (United Nations Environment Programme) to be dangerous for the environment and for public health. For this reason, they are called “dioxin like PCBs” (dl-PCBs) [3].

In polluted soils, POPs are often associated with the presence of inorganic compounds, such as heavy metals [4]. In the present study, soil samples have been collected in a large agricultural land in
Brescia, an Italian city in the region of Lombardy shown in Figure 1, whose large scale industrial development has resulted in an impact on the surrounding environment [5] [6].

Figure 1. Agricultural land close to industrial area, city of Brescia (Lombardy region).

As illustrated in Figure 2, the investigated area includes an important industrial centre with high population density. The detected fields were chosen in the proximity of Caffaro’s chemical factory, a PCBs producer in a chloralkaly process plant, operating from 1930 to 1984, indicated as the principal responsible for PCDD/Fs, PCBs and Hg pollution [7] [8].

Figure 2. “Caffaro” chemical factory (red area). ESPG:3003 1593185,5043968 1:8864.

The aim of the present study has been to identify other emission sources, different than “Caffaro”, in charge for soil pollution, using Principal Component Analysis (PCA) coupled to Kriging algorithm to perform the spatial distribution of contamination. Geostatistical approach has been used to confirm the relationships between original variables (heavy metals and POPs) and Principal Components (PCs).
2. Details experimental

2.1. Materials and procedures
In the considered site, 73 soil samples were collected by a grid sampling of the surface soil (10 cm. of depth) and analysed for some heavy metals (Cu, As, Zn, Pb, Ni, Hg, Cr and Mn), PCDD/Fs and dl-PCBs. Figure 3 shows the investigated areas and the sampling points [9].

![Figure 3](image_url). Investigated fields and grid sampling. ESPG:3003 1592259,5043401 1:8864.

Heavy metals have been analysed by ICP-MS (Agilent, USA) according to the EPA 6020A 1998 method [10], after their mineralization (Ethos Touch Control, Milestone, Italy) according to EPA 3051A 2007 method [11]. ISO 17034 analytical standards were from O2Si, Charleston, USA.

Organic compounds have been evaluated by HRGC/HRMS (Trace GC Ultra/ DFS, Thermo, USA) according to EPA1668C method [12]. ISO 17034 native and labelled standards were from Wellington Laboratories, Canada.

2.2. Principal Component Analysis
Principal Component Analysis (PCA) technique is a classical statistical approach to reduce the dimension of the original dataset by transforming it to a new set of variables (called “principal components”) to summarize the features of the above mentioned original variables. Principal components are uncorrelated and ordered such that the kth PC has the kth largest variance among all PCs [13]. So, the first component has the maximum variance. Successive components progressively explain smaller portions of the variance and are all uncorrelated with each other.

In this work PCA was used to identify the different contaminant sources in the considered site.

2.3. Kriging algorithm
Kriging is a geostatistical interpolation method that has proven useful and popular in many fields. In general, by the Kriging algorithm it is possible to locate the values of a spatial variable (i.e. organic or inorganic compounds) in areas where this variable has not been directly measured [14].

This method produces visually appealing contour and surface plots from irregularly spaced data. It can be custom fit to a data set by specifying the appropriate variogram model. It can be either an exact interpolator or a smoothing interpolator depending on the parameters specified by the user. It incorporates anisotropy and underlying trends in an efficient and natural manner.

In this work, Kriging interpolation technique has been used to analyse the spatial distribution of the contaminants in the investigated area.
3. Results and discussion

3.1. Principal component analysis results

PCA has been applied to 73 surface soil samples taken from the above described site. Results are described in Table 1.

| Elements  | PC1    | PC2    | PC3    |
|-----------|--------|--------|--------|
| As        | 0.745  | -0.045 | 0.464  |
| Mn        | 0.315  | 0.226  | 0.772  |
| Pb        | 0.303  | 0.060  | -0.628 |
| Cu        | 0.945  | 0.161  | 0.074  |
| Zn        | 0.852  | 0.125  | -0.095 |
| Hg        | 0.953  | 0.006  | 0.067  |
| Ni        | 0.694  | 0.355  | -0.052 |
| dl-PCBs   | 0.170  | 0.929  | 0.060  |
| PCDD/Fs   | 0.879  | 0.193  | -0.119 |

% variance 49 17 12
Cumulative variance % 49 66 78

The extraction of principal components amounts to a variance maximizing (varimax) rotation of the original variable space. This type of rotation is called variance maximizing because the goal of the rotation is to maximize the variance of the "new" variable (PC) [15]. Three PCs have been extracted by Kaiser criterion, retaining only factors with eigenvalues greater than 1 [16]. In soil samples, they accounted for 78% of the cumulative variance. PC1 shows high positive correlation with Ni, Cu, Zn, As and PCDD/Fs; PC2 presents high positive correlation with Hg and PCBs; PC3 is positively correlated with Mn and negatively correlated with Pb.

Municipal solid waste incinerator (MSWI), chemical, metallurgical and steel plants, cement factories, thermoelectric power station and vehicular traffic are the main emission sources in the study area. They have direct relationships with the analysed contaminants [6] [17] [18] [19] [20], as reported in Table 2.

Relationships shown in Table 2 and in Figure 4 suggest three principal types of pollutant’s emission, corresponding to three clusters of the original variables:

PC1: Municipal solid waste incinerators, characterized by the cluster As, Zn, Cu, Ni and PCDD/Fs;
PC2: Chemical farm, PCBs producer, characterized by Hg and PCB;
PC3: Metallurgical and steel plants, characterized by Mn.

However, neither of the three components perfectly correlate with Pb in surface soil: in other words lead concentration cannot be perfectly explained from the above mentioned source typology.

PCA results have been proved that “Caffaro” chemical farm is not the only responsible of the soil pollution in investigated area.
Table 2. Relationships between emission sources and soil contaminants.

| Emission source                        | Typical emissions                                      |
|----------------------------------------|--------------------------------------------------------|
| Thermoelectric power stations           | Ni                                                     |
| Municipal solid waste incinerators      | Zn, Cu, Ni, Mn, Pb, PCB, PCDD/Fs                       |
| Vehicular traffic                       | Pb, Cu, Cr, Ni                                        |
| Metallurgical and steel plants          | Fe, Mn                                                 |
| Carbon combustion plants                | As, Cr, PCDD/Fs                                        |
| Pigment plants                          | Pb                                                     |
| Chemical plants                         | Hg, PCB                                                |

* * * * *

3.2. Contour maps of pollution

A contour map with cross-validations was performed to describe the spatial distribution of each contaminant (organic and inorganic) [21], using ordinary kriging interpolation technique (Golden Software Surfer 8.0). As an example, Figures 4-9 illustrate the contour maps created for Cu and PCDD/Fs (as an example of micropollutants correlated with PC1), Hg and PCBs (correlated to PC2), Mn (correlated to PC3) and Pb (not clearly correlated to any of the above mentioned PCs).

![Figure 4](image1.png) Contour map for Cu (PC1).

![Figure 5](image2.png) Contour map for PCDD/Fs (PC1).

![Figure 6](image3.png) Contour map for Hg (PC2).

![Figure 7](image4.png) Contour map for PCBs (PC2).
It should be noted that the pollution sources identified by each PCs show off the same spatial distribution of pollutants. Noteworthy Pb spatial distribution looks quite similar to those identified from PC1 and it should therefore be accounted to the MSW incinerator. However, as showed Table 2, the presence of lead in the surface soil could also depend on vehicular traffic or pigment plants. While MSWI and vehicular traffic produce diffuse contamination [22], the “hot-spot” on the left side is due to a point source pollution. As lead is very poorly mobilized in the environment it is quite easy to assess an old pigments manufacturer in close proximity of the investigated area as its pollution source. This assumption has been confirmed by PCA analysis, introducing a fourth principal component. PC4 shows a positive correlation with Pb only, as represented in Table 3.

Table 3. Principal Components Analysis: rotated component matrix and variance explained (4PCs).

| Elements   | PC1    | PC2    | PC3    | PC4    |
|------------|--------|--------|--------|--------|
| As         | 0.757  | -0.063 | 0.520  | 0.111  |
| Mn         | 0.085  | 0.232  | 0.872  | -0.141 |
| Pb         | 0.140  | 0.041  | -0.057 | 0.966  |
| Cu         | 0.863  | 0.124  | 0.368  | 0.168  |
| Zn         | 0.906  | 0.085  | 0.047  | 0.004  |
| Hg         | 0.082  | 0.948  | 0.123  | 0.087  |
| Ni         | 0.885  | -0.032 | 0.336  | 0.125  |
| Cr         | 0.770  | 0.322  | 0.021  | -0.062 |
| dl-PCBs    | 0.869  | 0.153  | 0.137  | 0.182  |
| PCDD/Fs    | 0.210  | 0.922  | 0.065  | -0.032 |
| Cumulative % variance | 49 | 66 | 78 | 87 |

Therefore, the use of the spatial assessment coupled to PCA analysis has proved a multisource lead pollution, due to the effects of a diffusion contamination, such as the MSWI, and a point emission source, such as the old pigment factory.

4. Conclusions
The present work has investigated the integration between Principal Components Analysis, often used in pollution source determination, and the pollution spatial assessment obtained by the means of geostatistical data interpolation. It has been demonstrated that different clusters of organic and
inorganic micropollutants, determined by PCA, show a well-defined and “unique shape” of the pollutant’s spatial distribution. Moreover, the spatial assessment can be useful to find emission sources for those elements where it has not been possible by PCA analysis only.

References
[1] UNEP 2000 Proceedings of the Subregional Workshop on Identification and Management of PCBs and Dioxins/Furans, Cavtat, Croatia, 29 May–June 2000
[2] Fiedler H. Dioxins and Furans (PCDD/PCDF) in The Handbook of Environmental Chemistry Vol. 3, Part O Persistent Organic Pollutants - (2003) © Springer-Verlag Berlin
[3] UNEP Chemicals: http://www.chem.unep.ch/pos
[4] Orta-García S T, Ochoa-Martínez A C, Carrizalez-Yañez L, Varela-Silva J A, Pérez-Vázquez F J, Pruneda-Álvarez L G, Torres-Dosal A, Guzmán-Mar J L, Pérez-Maldonado I N 2016 Persistent Organic Pollutants and Heavy Metal Concentrations in Soil from the Metropolitan Area of Monterrey, Nuevo Leon, Mexico Archives of Environmental Contamination and Toxicology 70 452-463
[5] 2007 Regione Lombardia - ERSAF, Analisi del contenuto in rame ed altri metalli nei suoli agricoli lombardi, Quaderni della ricerca n. 61 Milano
[6] Facchinelli A, Sacchi E, Mallen L, 2001 Environmental Pollution 114 313-324
[7] ERSAF Regione Lombardia, Convegno Fitoretentione e sperimentazione di tecnologie di bioremediation nei terreni agricoli, 26 giugno 2017
[8] Asl Brescia 2002 Il caso Caffaro. Un’analisi oggettiva, Asl Brescia, 3 maggio 2002
[9] USEPA U.S. Environmental Protection Agency, Method 6020A (SW-846): Inductively Coupled Plasma-Mass Spectrometry, Washington, DC. 1998
[10] USEPA U.S. Environmental Protection Agency, SW-846 Test Method 3051A: Microwave Assisted Acid Digestion of Sediments, Sludges, Soils, and Oils, Washington, DC. 2007
[11] USEPA U.S. Environmental Protection Agency, Method 1668C Chlorinated Biphenyl Congeners in Water, Soil, Sediment, Biosolids, and Tissue by HRGC/HRMS, Washington, DC. 2010
[12] Jolliffe I T 2002 Principal Component Analysis, New York, Springer Verlag
[13] D’Or, D Demouget-Renard, H Garcia M 2009 An integrated geostatistical approach for contaminated site and soil characterization Mathematical Geosciences 41 307-322
[14] McLean J E, Bledsoe 1992 Behaviour of metals in soils (EPA/540/S- 92/018) Washington, DC – U.S. Environmental Protection Agency 1-25
[15] Abdi H 2015 Factor rotation in Factor Analysis in Encyclopedia of Social Sciences Research Methods, SAGE Publications, Inc.
[16] Costello A B, Osborne J W 2005 Best practices in exploratory factor Analysis: four recommendations for getting the most from your analysis, Pratical assessment, research evaluation 10(7)
[17] Croce G. Collana Ambiente ARPA Toscana, Scheda Informativa n.15, 2013
[18] Rutherford J Gettens, Hermann Kühn, W T Chase 1967 Identification of the Materials of Paintings: Lead White Studies in Conservation 12(4) 125–139
[19] Ferri R, Donna F, Smith D R, Guazzetti S, Zacco A, Rizzo L, Bontempi E, Zimmerman NJ, Lucchini R G 2012 Journal of Environmental Protection 3 374-385
[20] Ndikwere C L 1984 A study of heavy metal pollution from motor vehicle emissions and its effect on roadside soil, vegetation and crops in Nigeria, Environmental Pollution Series B, Chemical and Physical 7(1) 35-42
[21] Dao L, Morrison L, Kiely G and Zhang C 2012 Spatial distribution of potentially bioavailable metals in surface soils of a contaminated sports ground in Galway, Ireland Environmental Geochemistry and Health 1–12
[22] Ellis J B and Revitt D M 2008 Water Air Soil Pollution: Focus 8 577