



# **Neural network and Monte Carlo simulation approach to investigate variability of copper concentration in phytoremediated contaminated soils**

Nour Hattab, Ridha Hambli, Mikael Motelica-Heino, Michel Mench

## **► To cite this version:**

Nour Hattab, Ridha Hambli, Mikael Motelica-Heino, Michel Mench. Neural network and Monte Carlo simulation approach to investigate variability of copper concentration in phytoremediated contaminated soils. *Journal of Environmental Management*, 2013, 129, pp.134-142. <10.1016/j.jenvman.2013.07.003>. <insu-00852439>

**HAL Id: insu-00852439**

**<https://insu.hal.science/insu-00852439v1>**

Submitted on 12 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# **Neural network and Monte Carlo simulation approach to investigate variability of Copper concentration in phytoremediated contaminated soils**

Nour HATTAB<sup>a\*</sup>, Ridha HAMBLI<sup>b</sup>, Mikael MOTELICA-HEINO<sup>a</sup>, Michel MENCH<sup>c</sup>

<sup>a</sup> ISTO

UMR 7327 - CNRS/Université d'Orléans

Campus Géosciences, 1A, rue de la Férollerie, 45071 Orleans cedex 2, France

mail : [nour.hattab@univ-orleans.fr](mailto:nour.hattab@univ-orleans.fr)

<sup>b</sup> Prisme Institute – MMH

8, Rue Léonard de Vinci, 45072 Orléans cedex 2, France

mail : [ridha.hambli@univ-orleans.fr](mailto:ridha.hambli@univ-orleans.fr)

<sup>c</sup> UMR BIOGECO INRA 1202, Ecologie des communautés, Université Bordeaux 1, Avenue des facultés, F-33405 Talence, France

## **Abstract**

Statistical variation of soil properties and their stochastic combinations may affect the extent of soil contamination by metals. This paper describes a method for the stochastic analysis of the effects of the variation in some selected soil factors (pH, DOC and EC) on the concentration of Copper in dwarf bean leaves (Phytodisponibility) grown in the laboratory on contaminated soils treated with different amendments. The method is based on a hybrid modeling technique: artificial neural network (ANN) and Monte Carlo Simulations (MCS). Because the repeated analyses required by MCS are time-consuming, the ANN is employed to predict the Copper concentration in dwarf bean leaves in response to stochastic (random) combinations of soil inputs. The input data for the ANN are a set of selected soil parameters generated randomly according to a Gaussian distribution to represent the parameter variabilities. The output is the Copper concentration in bean leaves. The results obtained by the stochastic (hybrid) ANN-MCS method show that the proposed approach may be applied (i) to perform a sensitivity analysis of soil factors in order to quantify the most important soil

parameters including soils properties and amendments on a given metal concentration, (ii) to contribute toward the development of decision-making processes at a large field scale such as the delineation of contaminated sites.

**Keywords:** Artificial neural network, Monte Carlo simulation, Soil contamination, Copper, Bean leaves, Soil factors variability.

## **1. Introduction**

Statistical variation related to the variabilities of soil properties (soil inputs) combined with various soil amendments, may affect trace metal mobility (Goovaerts, 2001; Broos et al., 1999, Schnabel et al., 2004). Many factors can influence the variability of soil parameter measurements, ranging from field sampling technique and soil location, to sample preparation and quality control in the laboratory. The ability to predict the concentration of a given metal in a given soil depends on the accuracy with which the soil inputs can be measured (Minasny and McBratney, 2002).

Laboratory soil tests to investigate the effect of the variability of soil properties are usually time-consuming and laborious. A large number of factors (inputs) involving random fluctuations in time and space are adequately described by stochastic processes. Furthermore, soils that are regularly ploughed tend to present a greater spatial dependency as depth increases, because management leads to surface homogeneity (Souza et al., 2006). Camacho-Tamayo et al., (2008) investigated the spatial variability of some chemical properties in different agricultural fields in Columbia. Their results showed that spatial variability of the soil chemical properties depends upon the use of amendments with a greater influence in the upper 100 mm of soil. Hence, a rapid modeling tool is needed to perform stochastic analysis

(Goovaerts, 2001). Such stochastic modeling approach in soil problems are still lacking, for several reasons: (i) the complexity and length of time needed to perform the experiments and (ii) the very high number of experiments to be conducted to obtain statistically significant results. Ramsey et al. (2002) developed an analytical approach called the duplicate method based on balanced experimental design, where a small proportion of sampling targets are sampled in duplicate. The data produced from the implementation of the duplicate method was then analyzed using robust analysis of variance (ANOVA) to separate and quantify the individual contributions to uncertainty as standard deviations. The values for the total measurement uncertainty, and the separate contributions from the sampling was estimated from the resultant measurements using robust analysis of variance (ANOVA) to separate and quantify the individual contributions to uncertainty as standard deviations (Ramsey et al., 1992, Ramsey and Argyraki, 1997).

The application of stochastic simulation algorithms in the field of soil analysis is a common tool for decision-making processes such as the delineation of contaminated sites (Goovaerts, 2001; McKenna, 1998; Broos et al., 1999, Betrie et al., 2013). Several methodologies have been used to account for uncertainty such as Kalman filtering (Peter, 1979; Ahsam and O'Connor, 1994), first-order analysis (FOA) (Chaubey et al., 1999; Haan and Skaggs, 2003a, 2003b), Monte Carlo Simulations (MCS) (Haan and Skaggs, 2003a, 2003b; Ogle et al., 2003; Wang et al., 2005), Latin hypercube sampling (LHS) (Pebesma and Heuvelink, 1999) and generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992; Beven, 1993). Recently, Oporto et al. (2012) applied MCS to identify the cause of soil cadmium contamination. The parameter uncertainty was taken into account with the MC analysis. Different algorithms have been applied by several authors to perform stochastic modeling (Deutsch, 1994; Gotway and Rutherford, 1994; Srivastava, 1996). These studies indicated that (i) no simulation algorithm is valid for all cases, and (ii) most of them require

the mathematical description of the relationships between soil inputs (properties) and soil outputs which have to be stated a priori in the regression models (Goovaerts, 2001).

Rapid and accurate analysis of the variabilities involved in the assessment of soil factors by stochastic methods is still lacking compared to deterministic procedures.

In the current work, a hybrid stochastic analysis approach Neural network (ANN) and Monte Carlo simulation (MCS) was developed allowing for the prediction of soil output variability based on soil input parameters variabilities and to examine the predictive variability as a function of model inputs across the full range of parameter space. In the present study we consider the specific case of the prediction of variability of Copper (Cu) toxicity in contaminated soils that have undergone phytoremediation versus the variabilities of some selected soil parameters such as: pH, the electrical conductivity (EC) and the dissolved organic carbon (DOC). Cu toxicity was evaluated by the Cu concentration in bean leaves (BL) grown in the laboratory on contaminated soils treated with four different amendments.

The ANN-MCS procedure generates random statistical combinations of soil inputs and performs the corresponding stochastic predictions compatible with prescribed statistical distributions, which are found to be Gaussian, related to the soil factors. The use of the normal distribution is often defended by invoking the central limit theorem, which states that any distribution will tend to behave like a normal distribution as the values are averaged together mainly for independent variables.

## **2. Material and methods**

The following flowchart describes the different steps and their interdependencies to perform combined ANN-MCS (Fig.1).

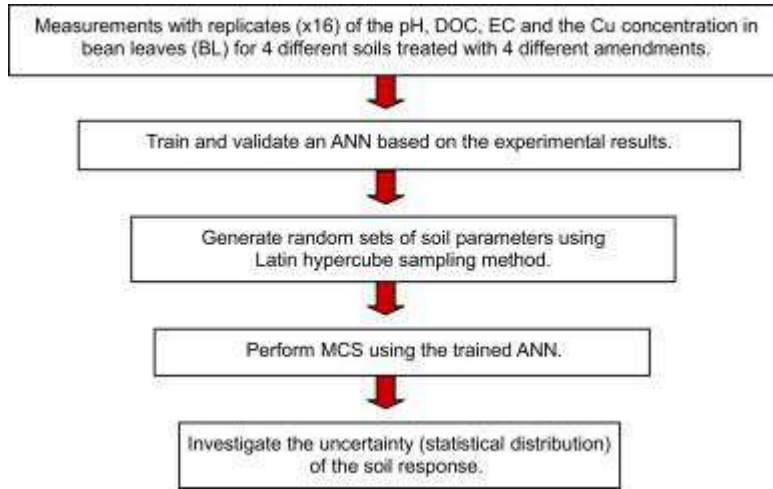


Fig. 1. : Flowchart of the hybrid stochastic computation approach: ANN-MC.

For an individual case of Cu prediction, the procedure for obtaining the stochastic solution that incorporates soil parameter variabilities is as follows:

- (1) *The values of each input variable (pH, DOC and EC) are generated randomly based on their statistical distribution (normal distribution in our case with a mean value and a standard deviation for each parameter);*
- (2) *Deterministic ANN prediction of Cu concentration in BL is performed in response to the random data set generated in step 1;*
- (3) *The above two steps are repeated ten thousand times as part of the MCS;*
- (4) *Finally, the predicted Cu concentrations in BL obtained in step 3 are used to determine the Cu concentration statistical distribution function related to the variability of each input factor (pH, DOC and EC).*

## 2.1. Experiment

### 2.1.1. Soil sampling and preparation

Sixteen soil samples with 16 replicates of each one were collected in 2011 from 16 plots (1x3m) to a depth of 0.25 m from the BIOGECO phytostabilization platform installed on

a former wood preservation site located in south-western France, Gironde County (44°43'N; 0°30'O). This site has been contaminated with high concentrations of Cu. The history of the site and its characteristics are detailed in (Mench and Bes, 2009; Bes et al., 2010). Long-term aided phytostabilization experiments are established at the site. The plant communities cultivated in the zone of the field trial were *Agrostis capillaris*, *Elytrigia repens*, *Rumex acetosella*, *Portulaca oleracea*, *Hypericum perforatum*, *Hypochaeris radicata*, *Euphorbia chamaesyce*, *Echium vulgare*, *Agrostis stolonifera*, *Lotus corniculatus*, *Cerastium glomeratum*, and *Populus nigra* (Bes et al., 2010). Weber et al. (2007) reported that the use of organic amendments in the soil requires testing of the long-term changes in available Cu. Therefore, on these 16 plots four different amendments were applied in May 2006, one per plot (Latin square design) and carefully mixed in the top soil (0-0.30 m) with a stainless steel spade with sixteen replicates: untreated soil (UNT), 0.2% of dolomite limestone (DL), 5% of compost of poultry manure and pine bark (CPM) and a mixture of 0.2% DL along with 5% CPM (DLX CPM). One kilo of each soil collected in 2011 from each plot was placed in a pot after sieving (2 mm). Four seeds of dwarf beans (*Phaseolus vulgaris*) were sown in each pot and cultivated for 18 days in controlled conditions (16 h light/8 h darkness regime). The soil moisture was maintained at around 50% of the field water capacity with additions of distilled water after weighing. Then the soil moisture was raised to 80% (11-13% of air-dried soil mass) at the beginning of the germination stage of the seeds. At the end of the growing period the plants were harvested and the dry weight of BL was determined after drying at 70°C.

The BL were weighed (35-150mg) directly into Savillex Polytetrafluoroethylene PTFE 50mL vessels, 2mL H<sub>2</sub>O and 2mL supra-pure 14 M HNO<sub>3</sub> were added and the vessels were heated open at 65°C for 2 hours. Then the caps were closed and the containers were left overnight at 65°C (12-14h). After that they were opened, 0.5mL of H<sub>2</sub>O<sub>2</sub> (30%) was added to each sample and left open at 75°C for 3 hours. Then 1.5±0.5mL of hydrofluoric acid HF

(48%) was added to each sample, caps were closed and left at 100°C overnight. The containers were opened and kept at 120°C for 4-5 hours evaporating to dryness, then taken off the heat; 1mL HNO<sub>3</sub> + 5mL H<sub>2</sub>O + 0.1mL H<sub>2</sub>O<sub>2</sub> were added to each, gently warmed up and after cooling down made up to 50mL. Finally, trace element concentrations in digests were determined by ICP-MS (Varian 810-MS) using standard solution of measured trace elements, in our case, Reference Standard Solution of Copper (1000ppm ±1%/Certified) was used. The accuracy of the metals determination was checked by performing calibrations with standard reference solution. Strong correlation was found between the measured and the reference results ( $R^2=0.9992$ ) indicating that the measurement are accurate. After the calibration phase, 16 repeated measurements were performed for each digest generating 256 testing results (4 repeated × 4 amendments × 16 digests). The precision of the metal content measurement was assessed by the standard deviation (SD) (Table 1).

### **2.1.2. Characterization of soil solution**

The soils were watered with distilled water after harvesting the dwarf beans and daily maintained at 80% of field capacity (11-13% of air-dried soil mass) for 2 weeks. Three Rhizon soil-moisture samplers (SMS) from Rhizosphere Research Products (Wageningen, Holland) were inserted after 2 weeks for 24h with a 45° angle into each potted soil (4x16 soils) to collect soil pore water (30 mL) from each pot. Then DOC was analyzed in the soil solution by a Shimadzu<sup>®</sup> TOC 5000A analyzer. Soil pH and EC were determined in the same soil solution by pH meter and the EC readings.

### **2.1.3. Selection of soil factors**

It has been reported that the relationships between the concentrations of metals in plants and a given soil are mainly influenced by the soil properties related to the ion charges (Sauvé



et al., 2000, Fisher and Binkley, 2000; Weng et al., 2002). Therefore, in the current preliminary study, the soil inputs for the ANN model were limited to the three measurable factors (pH, DOC and EC) considered to be the most influential on the mobility and availability of metals in the soil.

#### **2.1.4. Estimation of the parameters' statistical distribution**

Four soil samples were collected from 4 plots treated with 4 different amendments (UNT, CPM, DL and DLX CPM). 16 replications were performed for each measurement, generating 256 (4 soils x 4 amendments x 16 replicates) measured data used to estimate the statistical characteristics of the factors and to train the ANN. The total soil solution concentrations of Cu, pH, DOC and EC were statistically analyzed by (Statistica) to evaluate the Probability Density Function (PDF) of each measured variable and its corresponding characteristics. The Anderson-Darling normality test was performed. It was found that the retained soil factors distributions are Gaussian.

The calculated P-Value was greater than 0.05 and the data points follow an almost straight line. The estimation of the different parameters' mean values, their standard deviations and coefficient of variation (COV) is reported in table 1. It can be seen in table 1 that the average concentration and standard deviation vary with the treatment applied.

Table 1.

Statistical properties of the selected soil factors for different soil treatments.

<b>Treatment</b>	<b>Factor</b>	<b>Mean</b>	<b>Standard deviation</b>	<b>COV (%)</b>	<b>PDF</b>
UNT	pH	6.87	0.09	1.34	Normal
	EC (mS cm <sup>-1</sup> )	159.5	15.40	9.65	Normal
	DOC (mg L <sup>-1</sup> )	347.4	25.35	7.30	Normal
CPM	pH	6.98	0.055	0.79	Normal
	EC (mS cm <sup>-1</sup> )	161	14.30	8.88	Normal
	DOC (mg L <sup>-1</sup> )	352.4	23.77	6.75	Normal
DL	pH	7.05	0.055	0.78	Normal
	EC (mS cm <sup>-1</sup> )	162.4	13.35	8.22	Normal
	DOC (mg L <sup>-1</sup> )	362.5	22.07	6.09	Normal
DLX CPM	pH	7.15	0.04	0.57	Normal
	EC (mS cm <sup>-1</sup> )	163.1	12.55	7.70	Normal
	DOC (mg L <sup>-1</sup> )	356.7	21.03	5.90	Normal

(UNT: untreated; CPM: compost of poultry manure and pine bark; DL: dolomite limestone; DLX CPM: mixture of 0.2% DL along with 5% CPM; PDF: Probability Density Function; COV, coefficient of variation).

## 2.2. Artificial Neural networks

The ANN architecture is composed of an input layer, a certain number of hidden layers and an output layer in forward connections. Each neuron in the input layer represents a single input parameter. These values are directly transmitted to the subsequent neurons of the hidden layers. The neurons of the last layer represent the ANN outputs (Fig. 2).

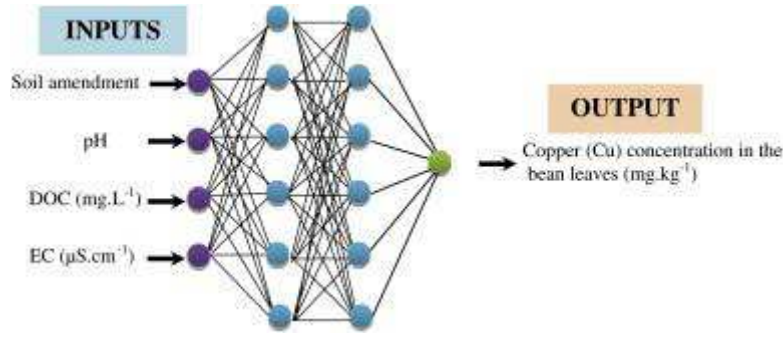


Fig. 2. : Artificial neural network architecture composed of 4 inputs, two hidden layers and one output layer.

The output  $y_i^m$  of neuron  $i$  in a layer  $m$  is calculated by (Johansson et al., 1992; Hagan et al., 1996; Haykin, 1999; Hambli et al., 2006; Hambli, 2010, 2011; Hambli et al., 2011).

$$y_i^m = f(v_i^m) \quad (1)$$

$$v_i^m = \sum_{j=1}^L w_{ji}^{m-1} y_j^{m-1} + b_i^m \quad (2)$$

Where  $y_i^0$  are the model inputs,  $v_i^m$  are the outputs of the layer  $m$ ,  $f$  is the activation function,  $L$  is the number of connections to the previous layer,  $w_{ji}^{m-1}$  corresponds to the weights of each connection and  $b_i^m$  is the bias, which represents the constant part in the activation function.

Among the activation functions, the sigmoid (logistic) function is the one most usually employed in ANN applications. It is given by (Haykin, 1999; Hambli, 2011):

$$f(v_i^m) = \frac{1}{1 + \exp(-\theta v_i^m)} \quad (3)$$

where  $\theta$  is a parameter defining the slope of the function ( $\theta = 0.9$ ).

### 2.2.1. Neural network training

Sixty four measurements were performed; 40 were used for training, 16 samples for testing and 8 samples covering a wide range of the experiments for validation. The testing data were not used for training. The testing data provided cross validation during the ANN

training for verification of the network prediction accuracy. The validation data were used to measure the performance of the predictive capability of the ANN after complete training.

In the present work, an in-house ANN program called Neuromod written in FORTRAN (Hambli, 2010, 2011; Hambli et al., 2011) was applied. The basic ANN configuration employed in this study has a double hidden layer with six neurons in each layer with a learning rate factor  $\eta = 0.1$  and momentum coefficient  $\alpha = 0.1$ .

The learning rate coefficient  $\eta$  and the momentum term  $\alpha$  are two user-defined ANN algorithm training parameters that affect the learning procedure of the ANN. The training is sensitive to the choice of these net parameters. The learning rate coefficient, employed during the adjustment of weights ( $w_{ji}^{m-1}$ ), was used to speed up or slow down the learning process. A larger learning coefficient increases the weight changes, hence large steps are taken towards the global minimum of error level, while smaller learning coefficients increase the number of steps taken to reach the desired error level. Tests performed for more than two hidden layers and different  $\eta$  and  $\alpha$  parameters showed no significant improvement in the obtained results.

### **2.3. Monte Carlo simulation**

For stochastic analysis using MCS,  $N$  samples of the vector of random soil inputs were generated randomly according to a statistical distribution function. The implementation of the method consisted in the numerical simulation of these samples with an ANN. As shown in Figure 3, the stochastic modeling concept based on MC sampling can be considered as the direct mapping of the input space onto the output space. The input space represents the random combinations of input factors, each of which follows a probability density function (PDF) representing the statistical variability.

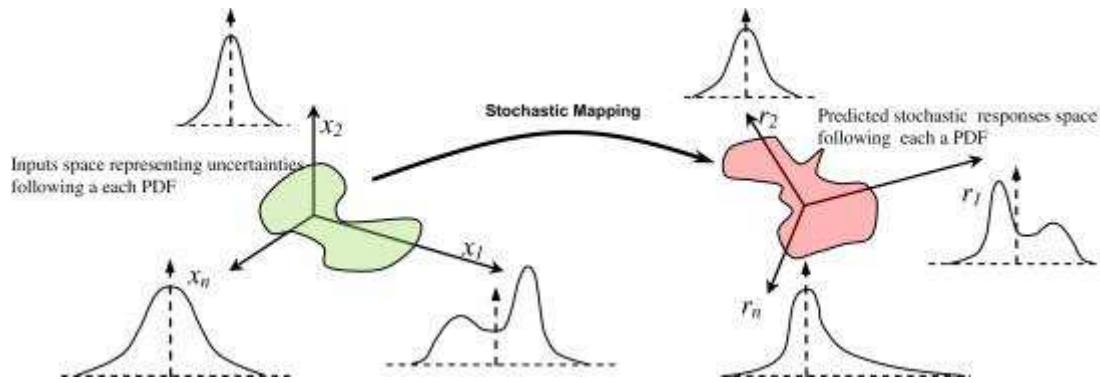


Fig. 3. : Stochastic modeling concept based on MC sampling considering factor variability consisting of mapping input space (due to variability following probability density functions) onto output space

Statistical applications for the study of contaminated sites commonly make use of theoretical distributions such as the normal, log-normal and exponential distributions. The most commonly used distribution model in soil statistics is the normal distribution (Al-Omran et al., 2004).

### 2.3.1. Sampling method for Monte Carlo simulations

Latin hypercube sampling (LHS) was used for the MCS to select random points from the uniformly distributed parameter space. One of the advantages of the LHS method which makes it appropriate for this study is that LHS ensures full coverage over the range of each variable so that all areas of the sample space are represented by the selected input values (Pebesma and Heuvelink, 1999). The more points that are selected from the parameter space, the more densely the space will be covered and the more reliable the results will be. In the current work, 10,000 random combinations were used for the ANN-MCS analysis.

## 3. Results

Figure 4 shows the accuracy of predicting Cu concentration with the trained ANN for the four different soil treatments. It can be seen that (i) the ANN prediction is in good

agreement with the experimental results, which means that the ANN has been sufficiently trained and (ii) the accuracy of the four soils with the four amendments is quite comparable. The  $R^2$  value of the regression is greater than 0.97 and the slope of the regression is close to 1.

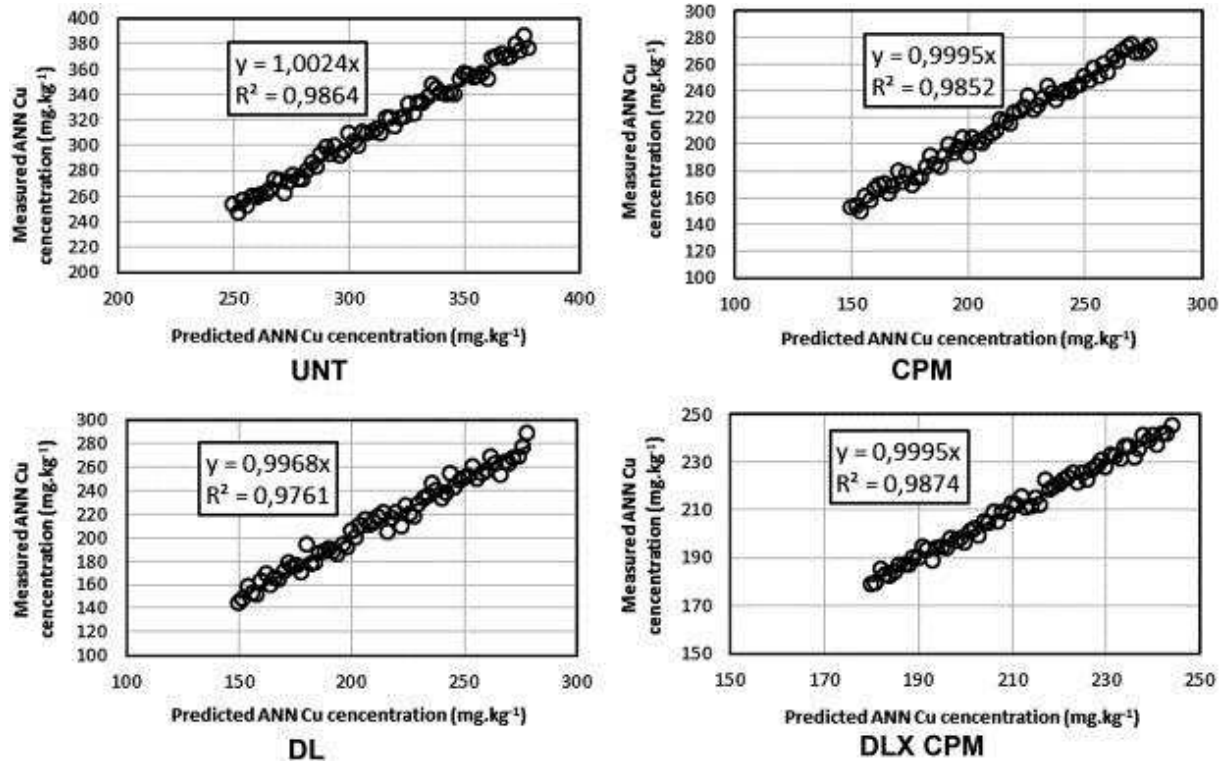


Fig. 4. : Predicted (ANN) versus Experimental results of Cu concentration in BL Diagrams showing the accuracy of the ANN predicting for different soil treatments. (UNT: untreated; CPM: compost of poultry manure and pine bark; DL: dolomite limestone; DLX CPM: mixture of 0.2% DL along with 5% CPM and BL: bean leaves).

First, the ANN-MCS model was tested using different number of random combinations ranging from 1000 to 100,000. We found that the calculation converges (insensitivity to the combination numbers) for a number of combinations greater than 5000. Therefore, to investigate the role of soil parameters' variability, 10,000 random combinations of the variability soil factors generated by MCS were applied. Each soil factors (inputs) was generated according to the normal distribution (Table 1). These results were then processed to determine the statistical characteristics of Cu response regarding input variability.

Figure 5 shows the effects of the three soil factors variabilities on the Cu concentration in BL. For an explicit comparison between the different soil treatments and the corresponding

data scatter, the mean value and standard deviation were normalized (unit normal distribution:  $\mu = 0$  and  $\sigma = 1$ ).

In order to study the influence of each soil parameter, only one parameter was generated randomly for each calculation. For the other factors were fixed, the mean values were taken (Table 1).

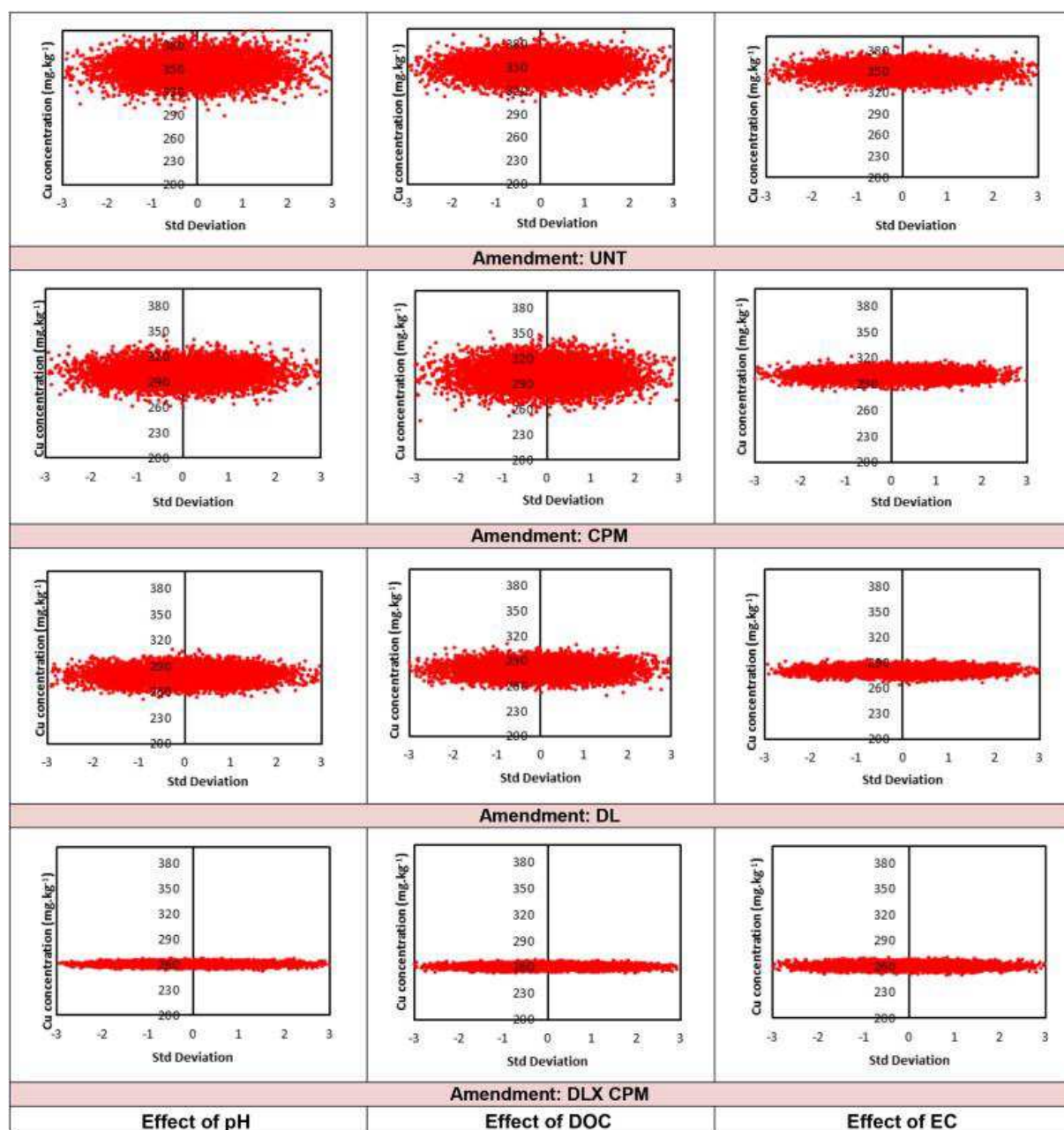
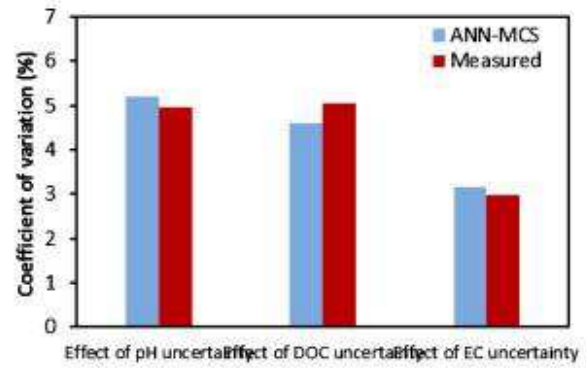
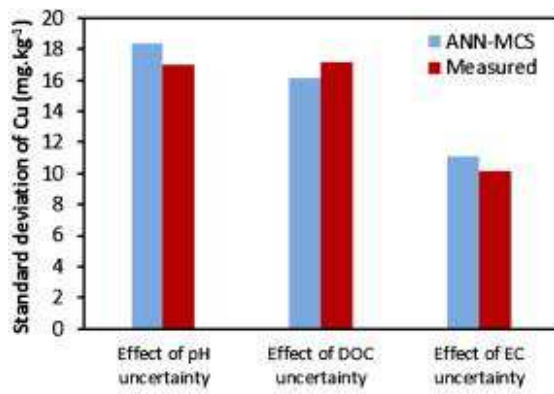


Fig. 5. : Copper concentration variation versus the statistical variation of soil factors for different amendments. (UNT: untreated; CPM: compost of poultry manure and pine bark; DL: dolomite limestone; DLX CPM: mixture of 0.2% DL along with 5% CPM).

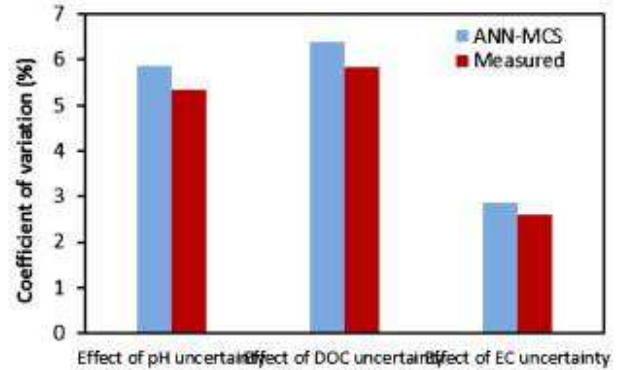
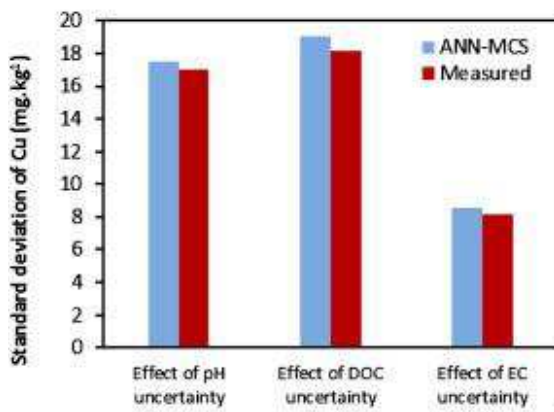
It can be observed that the scatter is significant due to the variability related to the soil parameters. The figure shows the stochastic nature of the predicted Cu concentration in BL generated by the statistical variation of the soil inputs. The measured variability is clearly statistically significant. The scatter plots in Figure 5 indicate that the predicted Cu concentration in BL for the different treatments follows the normal distribution with varying mean values, standard deviations and coefficients of variation (COV). In addition, the results indicated that the variability of the Cu concentration in BL was reduced by the application of soil amendments.

The effect of the pH, DOC and EC variabilities on the variability of Cu concentration in BL under the influence of the different amendments are as follows, from the greatest impact to the least impact: effect of pH > effect of DOC > effect of EC (low value of standard deviation and COV) (Fig. 6).

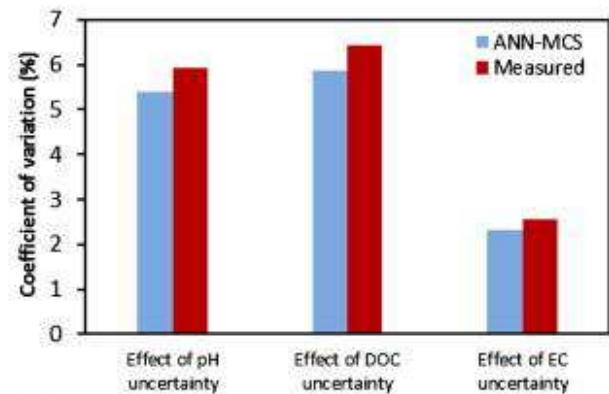
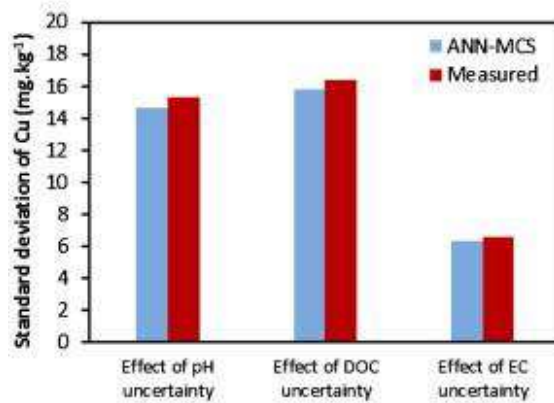




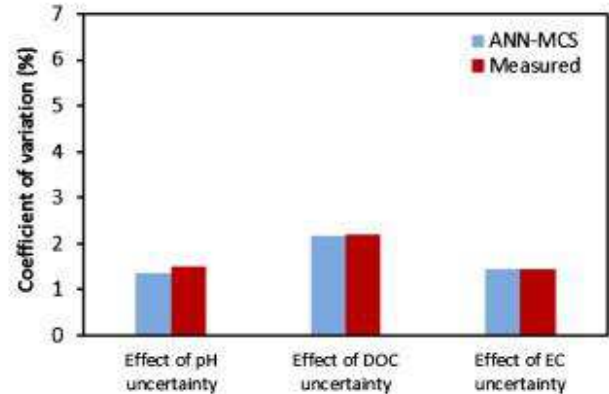
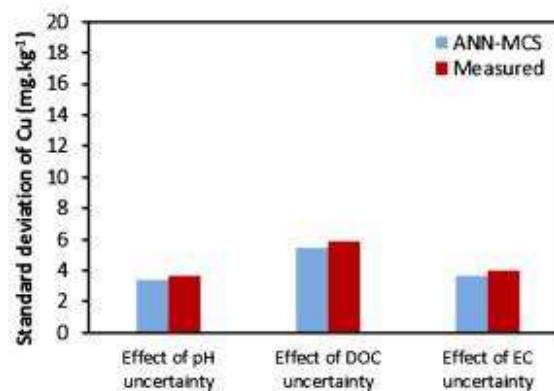
(a) UNT



(b) CPM



(c) DL



(d) DLX CPM

Fig. 6. : Predicted and measured standard deviations and coefficients of variation related to the uncertainties of pH, DOC and EC for four amendments (UNT, CPM, DL and DLX CPM). (UNT: untreated, CPM: compost of poultry manure and pine bark, DL: dolomite limestone and DLX CPM: mixture of 0.2% DL along with 5% CPM).

#### **4. Discussion**

The focus of the proposed ANN-MCS stochastic modeling approach was not on achieving an in-depth investigation of the variation of soil contamination by Cu and corresponding statistical analysis of the data but rather on (i) presenting the concept of combined rapid ANN-MCS modeling and (ii) presenting the potential of the method in performing and exploiting the stochastic analysis in a large field scale.

To assess the prediction accuracy, a comparison between the measured Cu concentration variability and the predicted ones shown in Figure 5 is summarized in table 2.

Table 2.

Comparison between predicted and measured Cu concentration (mean value, standard deviation and coefficient of variation) related to the selected soil factors variability (pH, DOC and EC) for four different soil treatments (UNT, CPM, DL and DLX CPM).

Treatment	Factor effect	Mean value (mg kg <sup>-1</sup> )			Standard Deviation (mg kg <sup>-1</sup> )			COV (%)		
		ANN-MC	Measured	Error (%)	ANN-MC	Measured	Error (%)	ANN-MC	Measured	Error (%)
UNT	pH	352.6	343.4	-2.7	18.3	17.0	-7.7	5.2	4.9	-4.9
	DOC (mg L <sup>-1</sup> )	350.7	340.1	-3.1	16.1	17.1	6.1	4.6	5.0	8.9
	EC (mS cm <sup>-1</sup> )	351.9	338.4	-3.9	11.1	10.1	-9.9	3.1	2.9	-5.7
CPM	pH	299.2	317.7	5.8	17.5	16.9	-3.0	5.8	5.3	-9.3
	DOC (mg L <sup>-1</sup> )	298.1	311.5	4.3	19.0	18.2	-4.7	6.3	5.8	-9.5
	EC (mS cm <sup>-1</sup> )	299.2	314.4	4.8	8.5	8.1	-4.5	2.8	2.6	-9.8
DL	pH	271.1	256.9	-5.5	14.7	15.2	3.9	5.4	5.9	8.9
	DOC (mg L <sup>-1</sup> )	269.3	254.2	-5.9	15.7	16.3	3.6	5.8	6.4	9.0
	EC (mS cm <sup>-1</sup> )	271.2	257.8	-5.2	6.3	6.5	4.1	2.3	2.5	8.8
DLX CPM	pH	252.8	243.1	-3.9	3.4	3.6	6.0	1.3	1.4	9.6
	DOC (mg L <sup>-1</sup> )	250.9	267.5	6.2	5.4	5.8	7.5	2.1	2.1	1.3
	EC (mS cm <sup>-1</sup> )	252.8	278.4	9.2	3.6	4.1	9.4	1.4	1.4	0.2

(UNT: untreated; CPM: compost of poultry manure and pine bark; DL: dolomite limestone; DLX CPM: mixture of 0.2% DL along with 5% CPM; PDF: Probability Density Function; COV, coefficient of variation).

It can be seen that the predicted and measured results are in good agreement, with an error ranging from -9.9 % to 9.7 %. This indicates that the proposed ANN-MCS model is acceptable for predicting variabilities (statistical characteristics and distribution) as a consequence of the variabilities of the soil inputs.

The ANN-MCS results indicated that pH variability generated the highest variability of Cu concentration suggesting that pH play a main role in stabilization of the

phytoremediation process. Considering the pH variability effect, the predicted Cu concentration varied between 304.3 mg.kg<sup>-1</sup> and 400.6 mg.kg<sup>-1</sup> in the case of UNT soil, between 265.7 mg.kg<sup>-1</sup> and 334.4 mg.kg<sup>-1</sup> in the case of CPM, between 258.1 mg.kg<sup>-1</sup> and 316.2 mg.kg<sup>-1</sup> in the case of DL and between 255.3 mg.kg<sup>-1</sup> and 271.4 mg.kg<sup>-1</sup> in the case of DLX CPM. The computed mean values and standard deviation for every applied amendment are respectively (UNT: Cu=352.6 mg.kg<sup>-1</sup>, SD=18.3 mg.kg<sup>-1</sup>), (CPM: Cu=299.2 mg.kg<sup>-1</sup>, SD=17.5 mg.kg<sup>-1</sup>), (DL: Cu=271.1 mg.kg<sup>-1</sup>, SD=14.7 mg.kg<sup>-1</sup>) and (DLX CPM: Cu=252.8 mg.kg<sup>-1</sup>, SD=3.4 mg.kg<sup>-1</sup>).

The second influencing factor was the DOC. ANN-MCS prediction showed that Cu concentration varied between 309.6 mg.kg<sup>-1</sup> and 395.7 mg.kg<sup>-1</sup> in the case of UNT soil, between 257.2 mg.kg<sup>-1</sup> and 349.5 mg.kg<sup>-1</sup> in the case of CPM, between 255.8 mg.kg<sup>-1</sup> and 310.2 mg.kg<sup>-1</sup> in the case of DL and between 252.3 mg.kg<sup>-1</sup> and 273.1 mg.kg<sup>-1</sup> in the case of DLX CPM. The computed mean values and standard deviation for every applied amendment are respectively (UNT: Cu=35.7 mg.kg<sup>-1</sup>, SD=16.1 mg.kg<sup>-1</sup>), (CPM: Cu=298.1mg.kg<sup>-1</sup>, SD=19.0 mg.kg<sup>-1</sup>), (DL: Cu=285.3 mg.kg<sup>-1</sup>, SD=15.7 mg.kg<sup>-1</sup>) and (DLX CPM/ Cu=250.9 mg.kg<sup>-1</sup>, SD=5.4mg.kg<sup>-1</sup>).

We found that EC was the less influencing factor compared to pH and DOC. ANN-MCS prediction showed that Cu concentration varied between 319.9 mg.kg<sup>-1</sup> and 385.3 mg.kg<sup>-1</sup> in the case of UNT soil, between 287.2 mg.kg<sup>-1</sup> and 321.2 mg.kg<sup>-1</sup> in the case of CPM, between 260.1 mg.kg<sup>-1</sup> and 291.2 mg.kg<sup>-1</sup> in the case of DL and between 248.4 mg.kg<sup>-1</sup> and 277.1 mg.kg<sup>-1</sup> in the case of DLX CPM. The computed mean values and standard deviation for every applied amendment are respectively (UNT: Cu=351.9 mg.kg<sup>-1</sup>, SD=11.1 mg.kg<sup>-1</sup>), (CPM: Cu=299.2 mg.kg<sup>-1</sup>, SD=8.5 mg.kg<sup>-1</sup>), (DL: Cu=271.2mg.kg<sup>-1</sup>, SD=6.3 mg.kg<sup>-1</sup>) and (DLX CPM: Cu=252.8 mg.kg<sup>-1</sup>, SD=3.8 mg.kg<sup>-1</sup>).

Interestingly, the simulation indicated that the soil amendment plays an important role in mediating the Cu concentration variability. The influence of the different amendments in generating scattered Cu response, from the greatest to the least impact, is as follows: effect related to UNT > Effect related to CPM > Effect related to DL soil > Effect related to DLX CPM soil.

The predicted results highlighted the role of the interactions between soil factors variability and applied amendments in controlling the level and the scatter of Cu concentration and suggested that the Cu variability is not driven by a particular soil parameter but by interactions among them. In all cases, DLX CPM amendment lead to the stabilization (reduced variability versus the variability of pH, DOC and EC) of Cu concentration (lower value of standard deviation) compared to UNT, CPM and DL amendments.

In all cases, DLX CPM amendment lead to reduced variability of Cu (Stabilization) versus the variability of pH, DOC and EC compared to UNT, CPM and DL amendments.

In contrast, the average value of Cu in the soil amended with DLX CPM was decreased. This result has important implication related to the selection of appropriate amendment for Cu phytoremediation in large field scale.

The study demonstrated the potential of ANN-MCS modeling procedure for prediction of Cu concentration variability (means and standard deviation) versus the variability of some selected soils properties (pH, DOC and EC) subjected to different amendments. Therefore, ANN-MCS approach may be applied to derive guidelines for specific phytoremediation at large field scale in regard to variability (spatial, temporal and environmental) of different soil properties as well as their interactions with the applied amendment.

In large field scale, many factors can influence variability of soil parameter measurements, ranging from field sampling technique and soil location, to sample preparation

and quality control in the laboratory. Hence, modeling soil processes must take into account the lack of large data covering the whole field and uncertainty in model parameters. Model predictions should incorporate, when possible, analyses of their uncertainty and sensitivity. This issue was investigated previously by several authors to deal with different causes of uncertainty such as heterogeneity of contaminant distribution in space and time (Ramsey et al., 1992, 2002; Ramsey, 2009 and Boon and Ramsey, 2010), measurement error in both the sampling and the chemical analysis (Ramsey and Boon, 2010) and contaminant heterogeneity and plant growth/uptake (Millis et al., 2004).

According to the current obtained results, one may suggest that the developed ANN-MCS model may be applicable at a large field scale for decision-making processes of Cu phytoremediation by controlling the sensitivity of the amount of a given metal uptake by plants in response to the soils factor variability (Deutsch, 1994; Gotway and Rutherford, 1994; Srivastava, 1996; Goovaerts, 1999). For example, Mench and Bes (2009) reported 1460 mg.kg<sup>-1</sup> of Cu in the topsoil of a large wood preservation site at different locations and between 65–2600 mg.kg<sup>-1</sup> on other site plots. Scholz and Schnabel (2006) reported that selection of specific soil remediation technique can be a challenging task due to among others to the uncertainty in assessment of level of contamination. Goovaerts (2001) addressed the importance of assessing the uncertainty about the value of soil properties and of the need to incorporate this assessment in subsequent decision-making processes, such as delineation of contaminated areas or identification of zones that are suitable for crop growth.

This needs solutions based on the representation of the soil factors inputs as probability density functions representing data variability rather than considering deterministic values (Scholz and Schnabel, 2006). Also, data from pot experiments are often not validated in field trials (Song et al. 2004; Hu et al. 2007). However, simple, reliable and rapid approaches to select remedial solutions are needed to optimize the solution and to

evaluate the potential of selected phytoremediation options (Ramsey and Boon, 2010, Moreno-Jimenez et al. 2011). Ramsey et al. (2002) developed a low cost analytical approach based a small proportion of sampling targets to optimized contaminated land related to the factors variability.

Present approach is based on computational one. From a practical point of view, if the tests performed on large field comprise a limited measurements data record, ANN-MCS procedures may provide a suitable tool for (i) modeling the inputs-outputs relationship with ANN and (ii) performing stochastic analysis using MCS to control the soil response. For example, based on maps of pH, DOC and EC spatial distribution, with the ANN-MCS, (i) one can draw the map of Cu variability distribution which can be useful to start any management methods in any field area and investigate the role of a given amendment on the Cu uptake stabilization, (ii) the method may be used to perform statistical analysis to obtain output probability distribution functions and confidence limits (Beven and Binley, 1992; Beven, 1993) and (iii) Optimization procedures based on sensitivity analysis may be developed to control the level and variability of the Cu concentration by the optimal combination of soil factors and applied amendments.

Considering other large fields, it should be speculated that the stochastic solution that incorporates the same soil properties and same uncertainties may be employed to predict the Cu variability of other similar soils in similar conditions.

In a more general case, ANN-MCS algorithm can contribute toward the development of software and/or implemented into in-situ measurements devices to control the Cu uptake process. For example, Ferreira et al., (2008) developed an ANN as a calibration strategy for Cu determination in soil samples using a portable Laser Induced Breakdown Spectroscopy system. Also, ANN-MCS procedure is suitable in many other soil problems where mapping soil outputs variability to soil factors variability is needed.

Despite its good performance, the proposed hybrid ANN-MCS method suffers from a number of limitations.

First, ANN-MCS is not based on a physical description of the problem, which limits the physical understanding of these uncertainties. Moreover there are various types of uncertainties related to other soil/plant factors, and the entire spectrum of uncertainties is not yet known. In this first work, only the uncertainty of three selected soil parameters was studied to demonstrate the potential of the method. Second, the available statistical information is mainly restricted to the mean value, the standard deviation, upper and lower fractal values or upper and lower bounds which characterize the normal distribution. Other probability density functions such as log-normal and exponential distributions may be applied instead of the normal distribution. However, with the ANN-MCS approach, whatever the statistical information available, it can be used in relation with a specific sampling procedure for MCS. Third, there are no general guidelines which can help in the design of ANN-MCS modeling by a non-specialist. There is a need to develop a user-friendly interface that can be used by a non-specialist.

Finally, it is worthy to notice that the total time required to develop and run the ANN-MCS is dependent upon: the time required to generate the training and test data used; the time taken to train a number of ANN models in order to select the appropriate complexity; and the time required to run the ANN for the required number of MC samples. The majority of time is spent on generating the training data. In this study, the time required for the ANN prediction of 10,000 input combinations was approximately 25 sec.

In conclusion, the proposed ANN-MCS approach can be applied:

- To perform sensitivity analysis of soil factors to quantify the most important parameters in a soil design problem that could then serve to formulate a



mechanistic model and to determine where future research efforts should be targeted.

- To quantify and classify the most important soil parameters including soils properties and amendments on a given metal concentration
- For decision-making processes at a large field scale such as the delineation of contaminated sites.

## References

- Ahsam, M., O'Connor, K. M., 1994. A reappraisal of the Kalman filtering technique, as applied in river flow forecasting. *J. Hydrology* 161: 197-226.
- Al-Omran, A.M., Abdel-Nasser, G., Choudhary, I., AL-Otuibi, J., 2004. Spatial variability of soil pH and salinity under date palm cultivation, *Res Bult*, (128), Agric Res Center, King Saud Univ, 5-36.
- Bes C M, Mench M, Aulen M, Gaste H and Taberly J 2010 Spatial variation of plant communities and shoot Cu concentrations of plant species at a timber treatment site. *Plant and Soil* 330, 267-280.
- Betrie G.D., Sadiq R., Morin K.A., Tesfamariam S., 2013, Selection of remedial alternatives for mine sites: A multicriteria decision analysis approach, *J Environ Manage.* 15; 119:36-46.
- Beven, K. J., Binley, A. M., 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* 6: 279-298.
- Beven, K. J., 1993. Prophecy, reality, and uncertainty in distributed hydrological modeling. *Advances in Water Resources* 16(1): 41-51.
- Boon K.A. and Ramsey M.H., 2010, Uncertainty on the measurements or on the mean for the reliable classification of contaminated land . *Sci Total Environ.* 409, 423–429.
- Broos, M.J., Aarts, L., van Tooren, C.F., Stein, A., 1999. Quantification of the effects of spatially varying environmental contaminants into a cost model for soil remediation. *J. Environ. Manage.* 56, 133–145.

- Bucher, C.G., 1988. Adaptive sampling- an iterative fast Monte Carlo procedure, *Struct Safety* 5, 119–126.
- Camacho-Tamayo J.H., Luengas C.A. and Leiva F.R., 2008, Effect of Agricultural Intervention on the Spatial Variability of some Soils Chemical Properties in the Eastern Plains of Colombia, *Chilean journal of agricultural research*, (1):42-55.
- Chaubey, I., Haan, C. T., Grunwald, S., Salisbury, J. M., 1999. Uncertainty in model parameters due to spatial variation of rainfall. *J. Hydrology* 220(1-2): 48-61.
- Deutsch, C.V., 1994. Algorithmically-defined random function models. In: Dimitrakopoulos, R. Ž. Ed. , *Geostatistics for the Next Century*. Kluwer Academic Publishing, Dordrecht, pp. 422–435.
- Ferreira E.; Milori D.; Ferreira E.; Dasilva R. and Martinneto L., 2008, Artificial neural network for Cu quantitative determination in soil using a portable Laser Induced Breakdown Spectroscopy system, *Spectrochimica Acta Part B: Atomic Spectroscopy*; 63, 10, 1216-1220
- Fisher, R.F., Binkley, D., 2000. *Ecology and Management of Forest Soils* (3<sup>rd</sup> Edition) Wiley, New York. 489 p.
- Goovaerts, P., 1999. Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties. *Stochastic Environ. Res. Risk Assess.* 13, 161–182.
- Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science, *Geoderma* 103. 3–26
- Gotway, C.A., Rutherford, B.M., 1994. Stochastic simulation for imaging spatial uncertainty: comparison and evaluation of available algorithms. In: Armstrong, M., Dowd, P.A. Eds., *Geostatistical Simulations*. Kluwer Academic Publishing, Dordrecht, pp. 1–21.
- Haan, P. K., Skaggs, R. W., 2003a. Effect of parameter uncertainty on DRAINMOD predictions: I. Hydrology and yield. *Trans. ASAE* 46(4): 1061-1067.
- Haan, P. K., Skaggs, R. W., 2003b. Effect of parameter uncertainty on DRAINMOD predictions: II. Nitrogen loss. *Trans. ASAE* 46(4): 1069-1075.
- Hagan, M., Demuth, H., Beale, M., 1996. *Neural network design*. PWS Publishing Company;.
- Hambli, R., 2010. Application of Neural Networks and Finite Element Computation for Multiscale Simulation of Bone Remodeling, 2010, in *Journal of Biomechanical Engineering*, (132):11, 114502

- Hambli, R., 2011. Numerical procedure for multiscale bone adaptation prediction based on neural networks and finite element simulation, *Finite elements in analysis and design*, (47):7, 835-842
- Hambli, R., Chamekh, A., Bel Hadj Salah, H., 2006. Real-time deformation of structure using finite element and neural networks in virtual reality applications, *Finite Elements in Analysis and Design*, (42):11, 985-991.
- Hambli, R., Katerchi, K., Benhamou, C.L., 2011. Multiscale methodology for bone remodelling simulation using coupled finite element and neural network computation, *Biomech Model Mechanobiol.*, 10(1):133-45.
- Haykin, S., 1999. *Neural networks, a comprehensive foundation*. 2nd ed. Prentice- Hall;. p. 181–2 [Chapter 4.6].
- Hu NJ, Luo YM, Wu LH, Song J. 2007. A field lysimeter study of heavy metal movement down the profile of soils with multiple metal pollution during chelate-enhanced phytoremediation. *Int J Phytorem.* 9:257–268.
- Johansson, E.M., Dowla, F.U., Goodman ,D.M., 1992. Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method. *Int J Neural Syst*;2(4):291–301.
- McKenna, S.A., 1998. Geostatistical approach for managing uncertainty in environmental remediation of contaminated soils: case study. *Environ. Eng. Geosci.* 4, 175–184.
- Mench M, Bes C. 2009. Assessment of the ecotoxicity of topsoils from a wood treatment site. *Pedosphere.* 19:143–155.
- Millis P.R., Ramsey M.H. and John E.A., 2004, Heterogeneity of cadmium concentration in soil has implications for human health risk assessment, *Sci Total Environ.* 326, 1-3, 29,49-53.
- Minasny, B., McBratney A.B., 2002. The neuro-m methods for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal*, 66: 352-361.
- Moreno-Jimenez E, Beesley L, Lepp N W, Dickinson N M, Hartley W and Clemente R 2011 Field sampling of soil pore water to evaluate trace element mobility and associated environmental risk. *Environmental Pollution* 159, 3078-3085.
- Ogle, S. M., Breidt, F. J., Eve, M. D., Paustian ,K., 2003. Uncertainty in estimating land use and management impacts on soil organic carbon storage for U.S. agricultural lands between 1982 and 1997. *Global Change Biology* 9(11): 1521-1542.

- Oporto, C., Smolders, E., Vandecasteele, C., 2012. Identifying the cause of soil cadmium contamination with Monte Carlo mass balance modelling: a case study from Potosi, Bolivia. *Environ Technol.* 33(4-6):555-61.
- Pebesma, E. J., Heuvelink, G. B. M., 1999. Latin hypercube sampling of Gaussian random fields. *Technometrics* 41(4): 303-312.
- Peter, S.M., 1979. *Stochastic Models, Estimation, and Control*. Vol. 1. New York, N.Y.: Academic Press.
- Ramsey M.H., Thompson M., Hale M., 1992, Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance. *J Geochem Explor*;44: 23–36.
- Ramsey M.H. and Argyraki A., 1997, Estimation of measurement uncertainty from field sampling: implications for the classification of contaminated land. *Sci Total Environ*; 198:243–57.
- Ramsey M.H., Taylor P. D. and Lee J.C., 2002, Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose, **J. of Env. Monit.**, 4, 5, 809 - 814.
- Ramsey M.H., 2009, Uncertainty in the assessment of hazard, exposure and risk. *Environmental Geochemistry and Health*, 31, 2, 205-217.
- Ramsey M.H. and Boon K.A., 2010, New Approach to Geochemical Measurement: Estimation of measurement uncertainty from sampling, rather than an assumption of representative sampling. *Geostandards and Geoanalytical Research*, 34, 293-304.
- Sauvé, S., Hendershot, W., Allen, H. E., 2000. Solid–solution partitioning of metals in contaminated soils: dependence on pH, total metal burden, and organic matter. *Environ. Sci. Technol.* 34, 1125–1131.
- Scholz, R.W., Schnabel, U., 2006. Decision making under uncertainty in case of soil remediation. *Journal of Environmental Management* 80, 132-147.
- Schnabel U., Tietje O., Scholz R.W., 2004, Uncertainty assessment for management of soil contaminants with sparse data. *Environmental Management* 33 (6), 911–925.
- Souza, Z.M. de, J. Marques Júnior, G.T. Pereira, and D.M. Barbieri. 2006. Small relief shape variations influence spatial variability of soil chemical attributes. *Sci. Agric. (Piracicaba, Braz.)* 63:161-168.
- Song J, Zhao FJ, Luo YM, McGrath SP, Zhang H. 2004. Copper uptake by *Elsholtzia splendens* and *Silene vulgaris* and assessment of copper phytoavailability in contaminated soils. *Environ Pollut.* 128:307–315.

- Srivastava, M.R., 1996. An overview of stochastic spatial simulation. In: Mowrer, H.T., Czaplowski, R.L., Hamre, R.H. Eds. , Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium, General Technical Report RM-GTR-277. US Department of Agriculture, Forest Service, Fort Collins, pp. 13–22.
- Wang, X., He, X., Williams, J.R., Izaurralde, R.C., Atwood, J.D., 2005. Sensitivity and uncertainty analysis of crop yields and soil organic carbon simulated with EPIC, American Society of Agricultural Engineers 48(3): 1041–1054.
- Weber, J., Karczewska, A., Drozd, J., Licznar, M., Licznar, S., Jamroz, E., Kocowicz, A., 2007. Agricultural and ecological aspects of a sandy soil as affected by the application of municipal solid waste composts. *Soil Biol. Biochem.* 39, 1294–1302.
- Weng, L.P., Temminghoff, E.J. M., Lofts, S., Tipping, E., Van Riemsdijk, W.H., 2002. Complexation with dissolved organic matter and solubility control of heavy metals in a sandy soil. *Environ. Sci. Technol.* 36, 4804–4810.