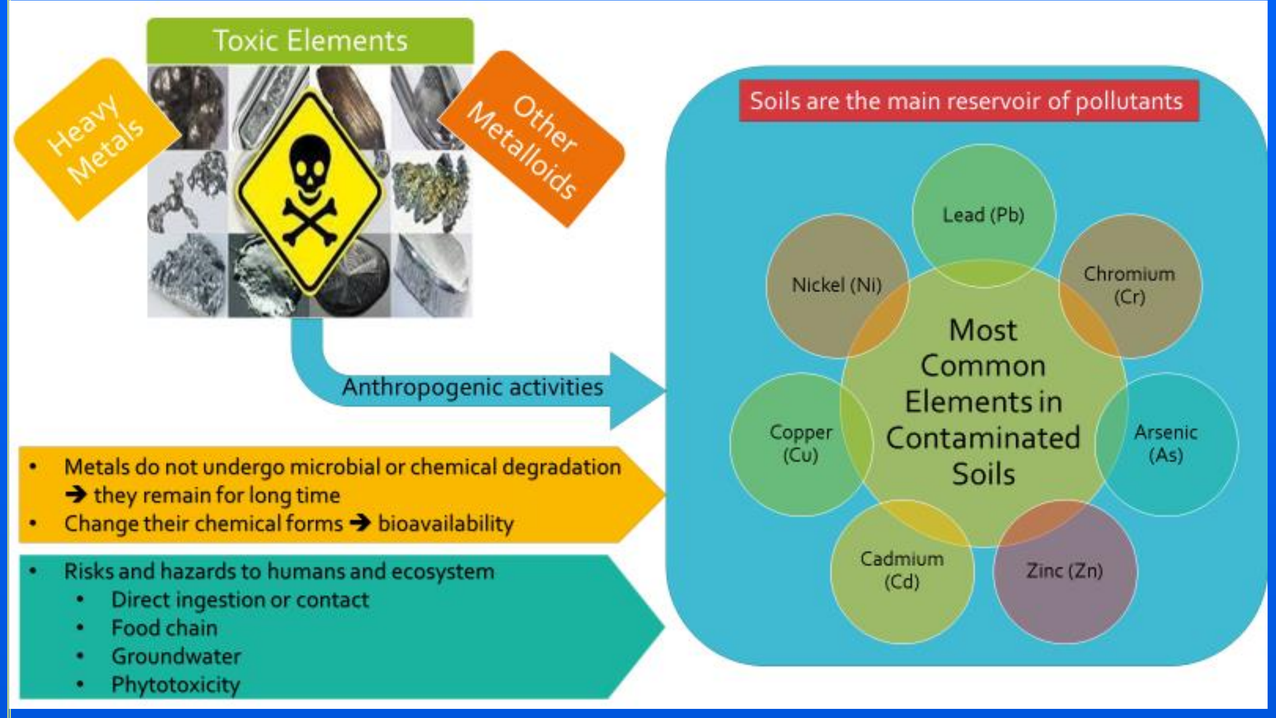
Use of machine learning methods to study the availability of nutrients and the distribution of toxic metals in agricultural Mediterranean soils

E.E. Golia[1], G. Benardos[2] and V. Diakoloukas[3]

[1] School of Agriculture, Aristotle University of Thessaloniki, Greece (email: egolia@auth.gr)
[2] AiGROW Effective Agriculture S.M.P.C, Greece (email: gbenardos@aigrow.gr)
[3] School of Electrical & Computer Engineering, Technical University of Crete, Greece (vdiakoloukas@tuc.gr)

Introduction



Elements

One of the 16 essential elements for plant growth and reproduction

- Most abundant form is Fe_2O_3 (Hematite)
- Highly insoluble (red color-rust)
- Oxide form usually hydrated

Under **oxidizing conditions**: the oxide, hydroxide, and phosphate forms control the concentration of Fe in solution

Under **reducing conditions** (addition of H+ or other reducing substances): the solubility increases

- Fe can be bound to the soil as an exchangeable ion.
- In certain soil conditions, carbonate or sulphide compounds can be formed with Fe.
- If sulphates are abundant in the soil, they become a source of oxygen for bacteria and black iron sulphide is formed.

Many organic compounds and organic acids (aliphatic acids or amino acids) and complex polymers (humic and fulvic acids) can form soluble complexes with Fe or act as chelating agents and thus increase the availability of Fe to plants

If percentage of organic matter in soil is high, Fe may be in a reduced state as Fe++ in the soil solution or adsorbed on the surfaces of soil particles

- Organic matter in soils plays an important role in the availability of Fe to plants.

48 Cd Cadmium

One of the most toxic pollutants

- Ranked seventh based on its toxicity
- It is highly toxic even at lower concentrations
- Persistent in soil for thousands of years

Cadmium as a threat

- Inhibits plant growth and metabolism
- Enters food chain
- Potential human carcinogen, causes kidney damage and cardiovascular diseases

Cd circulation in soils dependents on factors such as pH, soil type, organic content, presence of ligands and other ions in soil

Cadmium in the environment

- It exists in the form of complexes of sulphides, oxides and carbonates along with other metals instead of its pure form
- Non-biodegradable ➔ intricate to remove
- Soluble in water and highly mobile ➔ bioaccumulation and bioavailability
- Anthropogenic sources are mainly responsible for Cd contamination

Models, Data and Methodology

Multiple Linear Regression

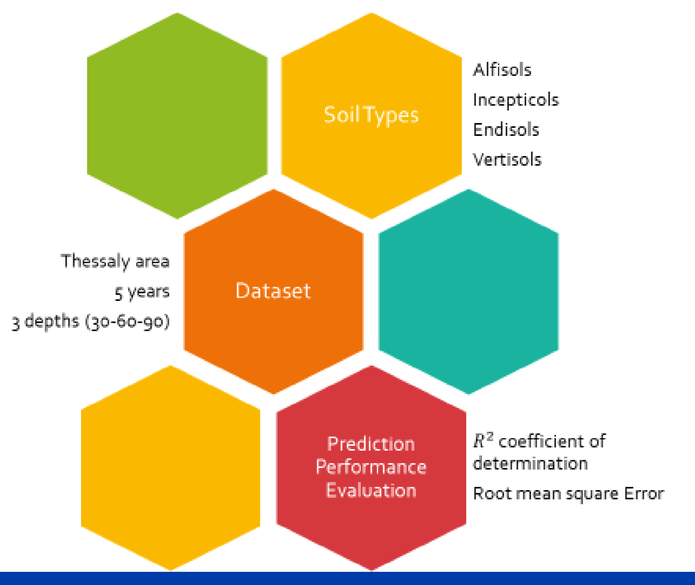
$$\hat{y} = b_0 + \sum_{i=1}^{K} a_i x_i + \sum_{i=1}^{K} b_i x_i^2 + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} c_{ij} x_i x_j$$

Robust Quadratic Regression

Goal: Formulate correlations of the target metal contaminant to other soil factors

Quadratic ➔ non-linear relations

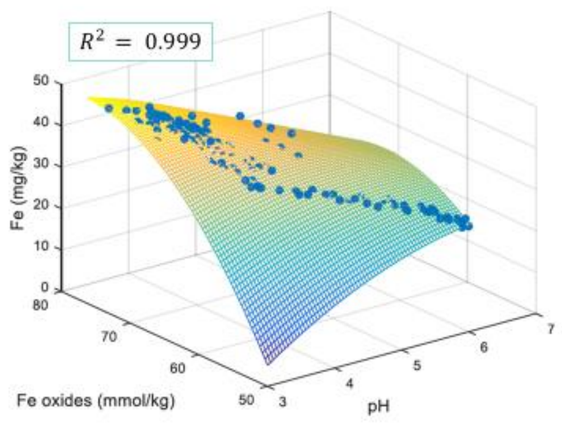
Robust (Iteratively Reweighted Least Squares parameter estimation) ➔ eliminate **outliers**



Results & Discussion

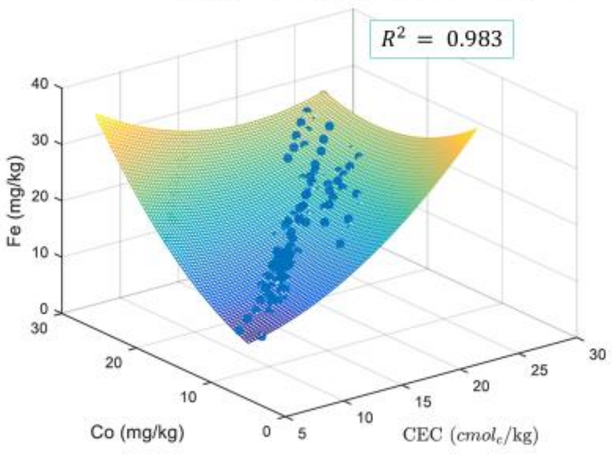
Alfisols

$$Fe = -35.62 + 2.1pH + 1.05 Fe\ oxides - 0.01pH \cdot Fe\ oxides - 0.14pH^2$$



Vertisols

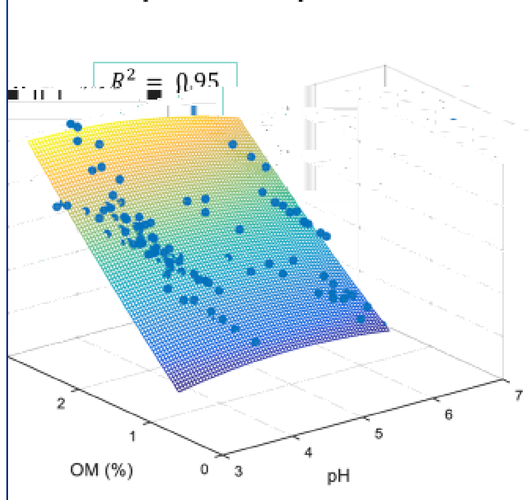
$$Fe = 1.55 + 0.33\ CEC + 0.47Co - 0.09CEC \cdot Co + 0.05CEC^2 + 0.04Co^2$$



3D plots of the regression curves that relate the Fe concentrations with two soil parameters (predictors) in each soil order.

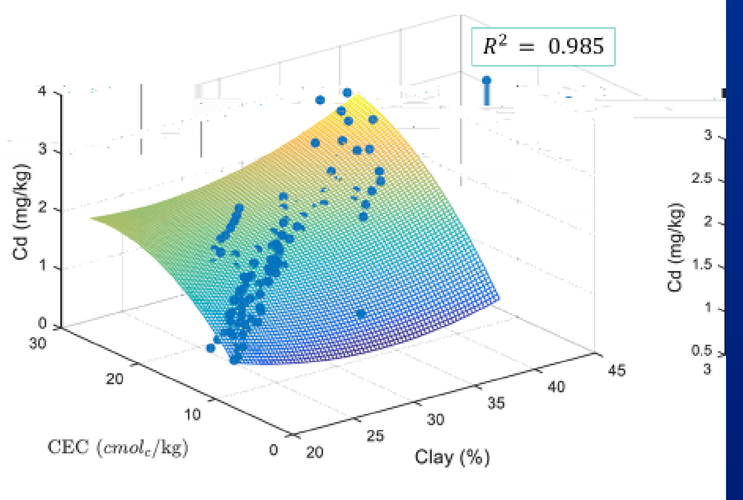
Alfisols

$$Cd = 2.19 - 0.5pH - 0.02OM + 0.08pH \cdot OM + 0.04pH^2 + 0.16\ OM^2$$



Vertisols

$$Cd = 0.25 - 0.06Clay + 0.17\ CEC - 0.01ClayCEC$$



3D plots of the regression curves that relate the Cd concentrations with two soil parameters (predictors) in each soil order.

Conclusions

- Multiple Quadratic Robust Regression can be successfully used to predict Iron and Cadmium concentrations from other explanatory soil parameters
- The proposed methodology can be extended to other regions and soil classes with similar climatic conditions
- Future Work:
  - Collect additional data ➔ Verify and optimize regression models
  - Evaluate new machine learning algorithms