POTENTIAL OF RADON DEFICIT AS A MONITORING TOOL IN ORGANIC SOIL REMEDIATION: A MACHINE LEARNING-BASED PREDICTIVE APPROACH

Jaime Montalvo-Piñeiro*, Fernando Barrio-Parra, Humberto Serrano-García, Miguel Izquierdo-Díaz y Eduardo De Miguel

*jaime.montalvo@upm.es

Acknowledgements:
This study was funded through the CARESOIL–CM [TEC-2024/ECO-69] and the agreement-subvention for the encouragement and promotion of research and technology transfer at the Universidad Politécnica de Madrid, in Line A, Emerging Doctors ML-TRACER [DOCTORES-EMERGENTES-24-FL6X] funded by the Regional Government of Madrid (Comunidad de Madrid).

INTRODUCTION

- Organic contaminants act as persistent sources of contamination
- Their distribution is difficult to characterize and delineate
- Traditionally, intrusive methods have been used that do not accurately reflect the actual distribution
- Uncertainties can increase the cost of monitoring and remediation programs
- Screening methods can be very useful tools for locating sources of contamination

USEPA 2004. Site Characterization Technologies for DNAPL investigations.

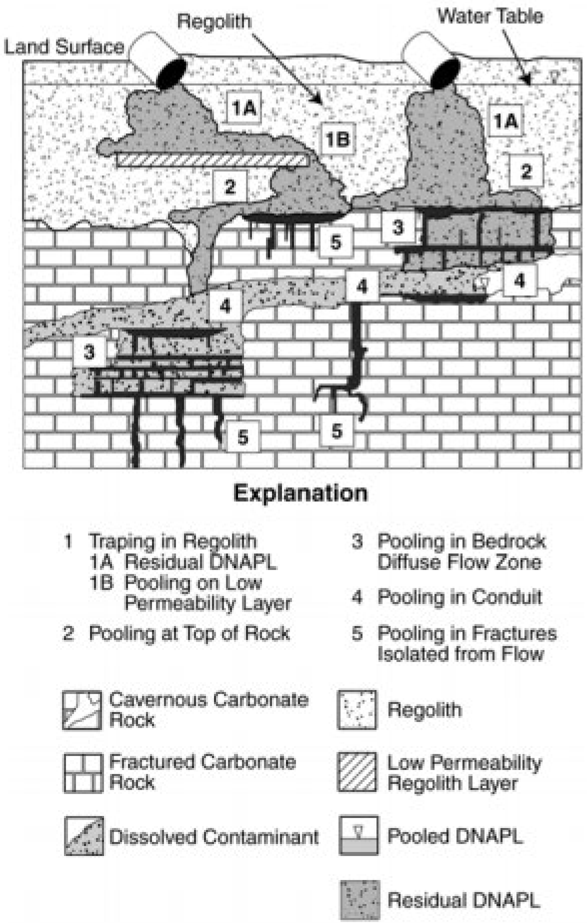
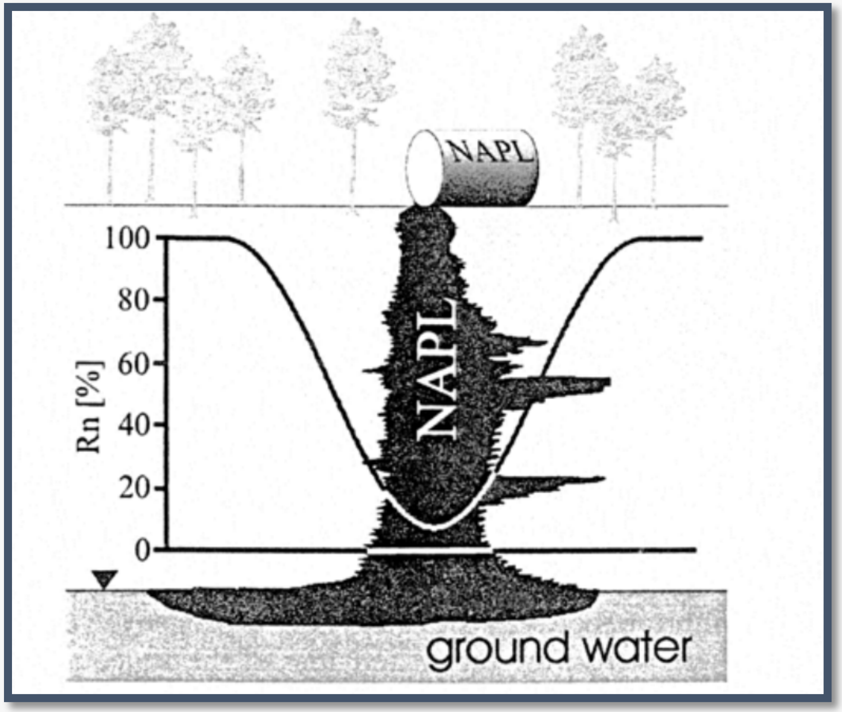


Figure 1-1. Distribution of Potential DNAPL Accumulation in a Hypothetical Karst Setting.

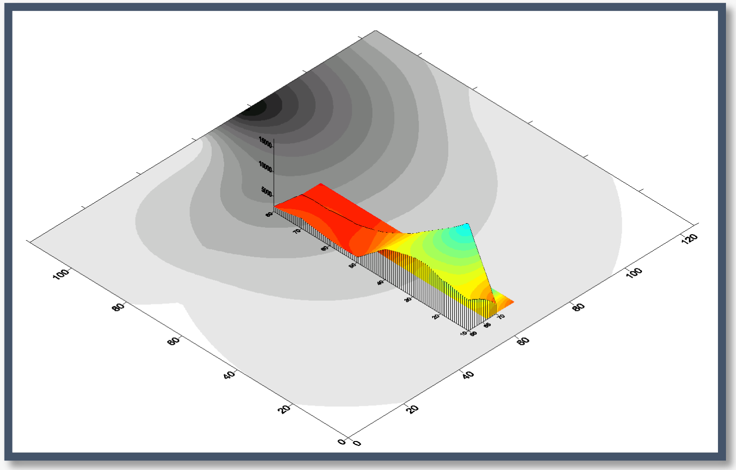
Source: Wolfe et al. 1997

^{222}Rn Deficit Technique

- ^{222}Rn is ubiquitous in soil and groundwater.
- It is preferentially distributed in organic phases.
- Radon levels vary depending on the presence of contamination.
- Limited by parameters that interfere with the signal



Schubert et al. 2001.
Journal of Soils and Sediments



García-Gonzalez et al. 2008.
Applied Geochemistry

OBJECTIVES

Objectives of the study

- **Investigate the use of machine learning models to support environmental surveying** and the interpretation of complex spatial data.
- **Analyze the potential relationships and the role of contaminants** in the variability of radon concentrations in the subsurface.
- Evaluate the potential utility of the **model's predictive capabilities** for monitoring decontamination processes.

MATERIALS AND METHODS

Data acquisition:

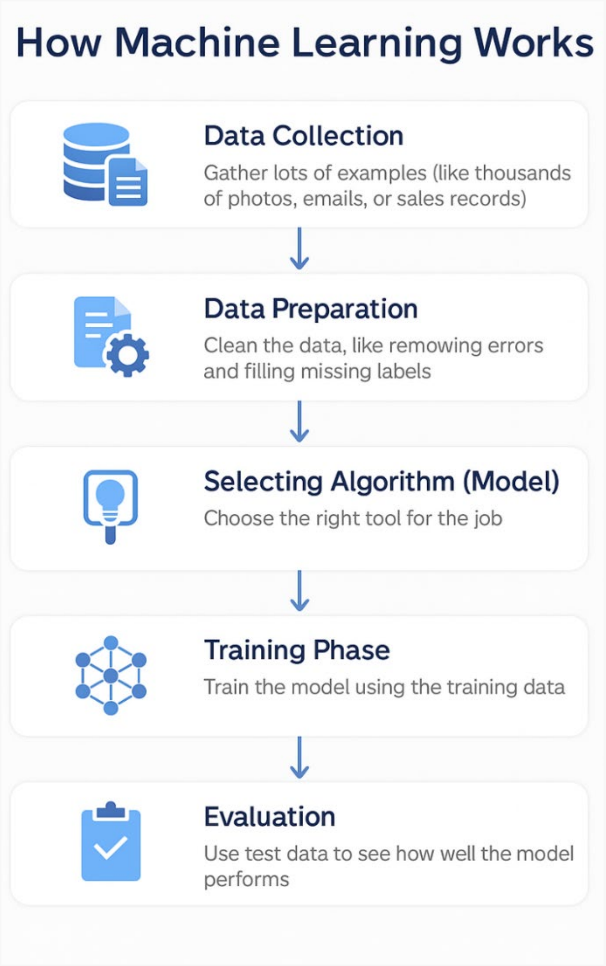
- **3 field campaigns** (February–March 2023).
- **242 valid measurements** for activity 222Rn.
- **64 sampling points** (blind sampling)
- **46 piezometers** (associated contaminants)

Sharing not permitted

MATERIALS AND METHODS

Data processing

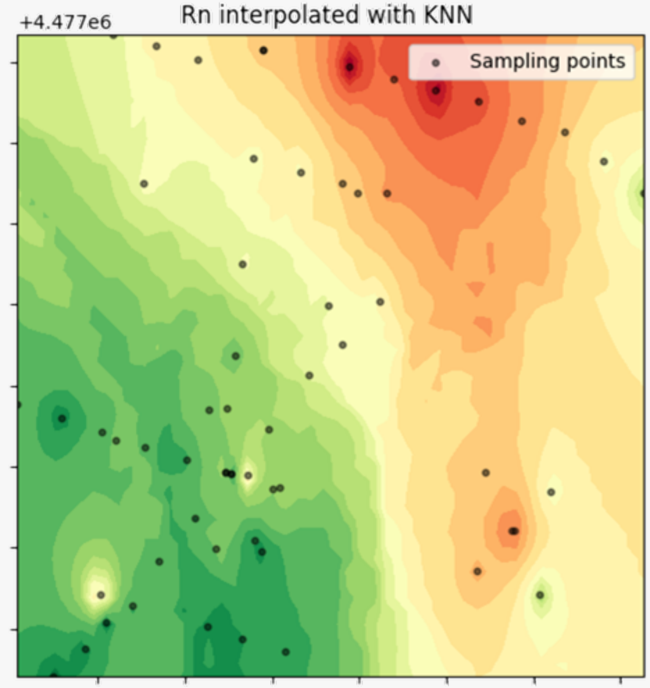
- Temporal variability→ Average ^{222}Rn activity
- Depth-profiled and preprocessed contamination data.
- Removal of outliers → Interquartile Range (IQR) method
- Spatial interpolation of predictor variables.
- ML model trained using preprocessed variables.



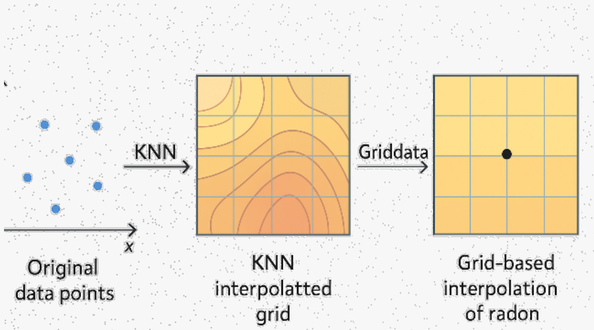
MATERIALS AND METHODS

Spatial modeling: interpolation and integration with ML

- **Spatial mismatch** between radon measurement points and piezometers.
- **Spatial harmonization** of variables using interpolation (KNN and RBF).
- Construction of a **Random Forest model** using hydrogeochemical variables and contaminants.
- **Model evaluation** using cross-validation (K-fold and GroupK-fold).



Estimating radon levels using KNN interpolation



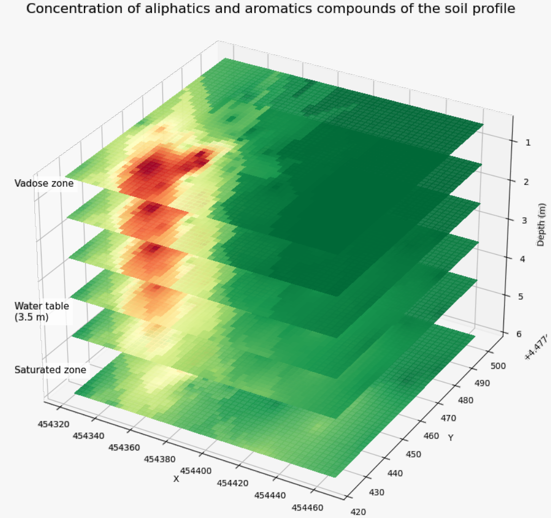
Data interpolation across the grid

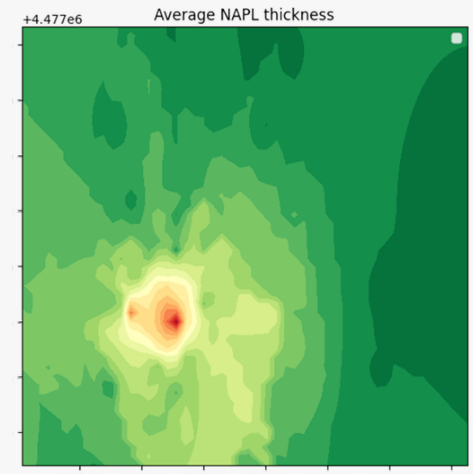
Sharing not permitted

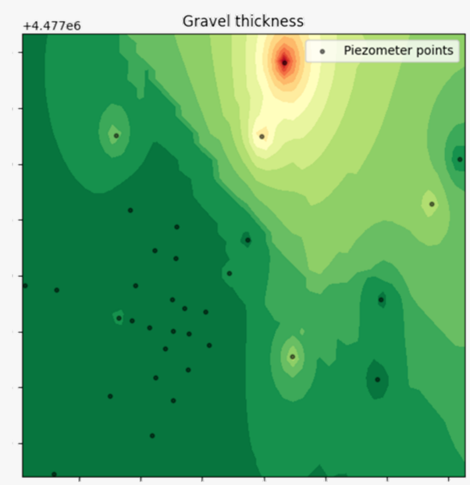
RESULTS

Spatial correlation analysis

- The radon signal exhibits a significant spatial structure.
- Bivariate analysis shows that the variables input into the model vary spatially with radon in a significant manner.
- These spatial patterns are complemented by Pearson's and Spearman's linear correlation analyses, which reveal consistent associations between variables.







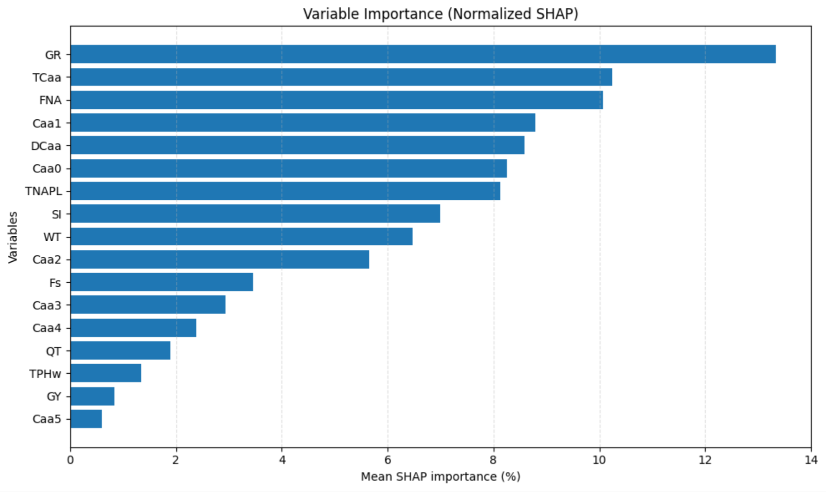
Spatial distribution of the variables across the study area

Sharing not permitted

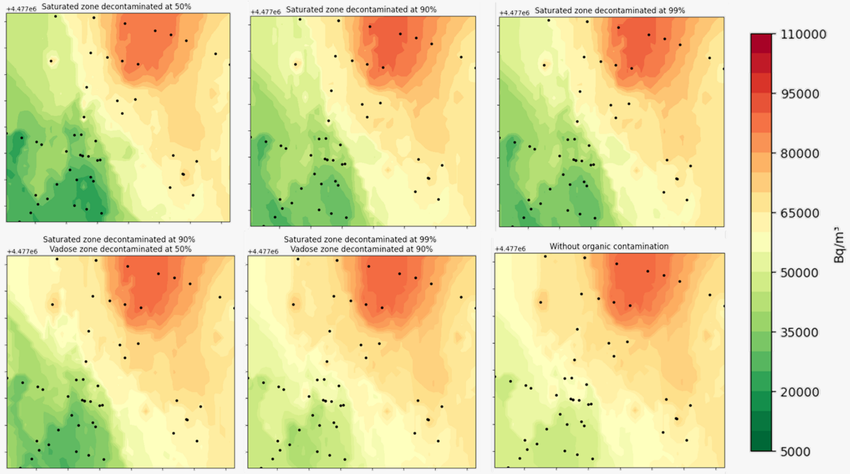
RESULTS

Performance evaluation and validation

- Differences in the influence of the contaminant are observed depending on depth, reflecting vertical effects and mixing phenomena ("smearing").
- The simulated reduction of organic compounds provides a theoretical approximation of radon evolution during remediation.
- The model allows the evaluation of hypothetical decontamination scenarios and their impact on the radon signal.



Importance of variables captured by the RF model



Radon modeling using machine learning in decontamination scenarios.

Sharing not permitted

CONCLUSIONS

- The ^{222}Rn deficit technique (RDT) has been confirmed as an effective screening method for LNAPL.
- Machine learning (ML) modeling reinforces the functional relationship between ^{222}Rn and the organic contaminants.
- The ^{222}Rn signal is influenced by the vertical continuity and geometry of the contaminant, as well as by hydrogeological variables.
- Deep NAPL may be underrepresented in shallow sampling, which limits the interpretation of RDT.
- Applicable as a supporting tool in the monitoring and tracking of remediation processes, with field validation.

Thank you!

jaime.montalvo@upm.es

Jaime Montalvo-Piñeiro*, Fernando Barrio-Parra, Humberto Serrano-García, Miguel Izquierdo-Díaz y Eduardo De Miguel

Acknowledgements:
This study was funded through the CARESOIL–CM [TEC-2024/ECO-69] and the agreement-subvention for the encouragement and promotion of research and technology transfer at the Universidad Politécnica de Madrid, in Line A, Emerging Doctors ML-TRACER [DOCTORES-EMERGENTES-24-FL6X] funded by the Regional Government of Madrid (Comunidad de Madrid).