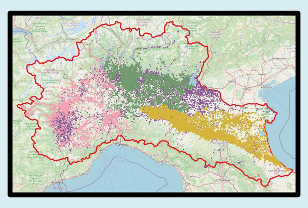
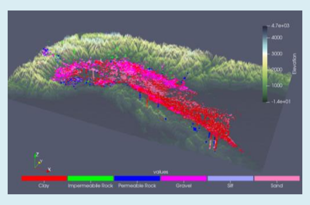
ANN Applied to Modeling the Spatial Distribution of Geomaterial at the Watershed Scale

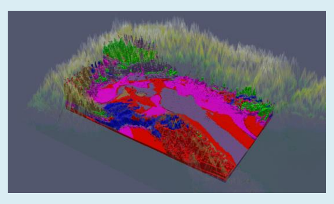
Andrea Manzoni*, Monica Riva, Giovanni M. Porta and Alberto Guadagnini
Politecnico di Milano, Dipartimento di Ingegneria Civile e Ambientale, Milano, Italy
*Contact: andrea2.manzoni@polimi.it

1. INTRODUCTION

- In a system management prospective, groundwater resources information, based on available atmospheric, surface and subsurface environmental covariates, at watershed-scale level, is required. Available data are needed to support assessment of groundwater availability and quality and provide input for flow and contaminant transport modeling. Water flow dynamics and contamination patterns in the subsurface environment are closely dependent on the spatial distribution of subsurface geomaterials.


Data Collection

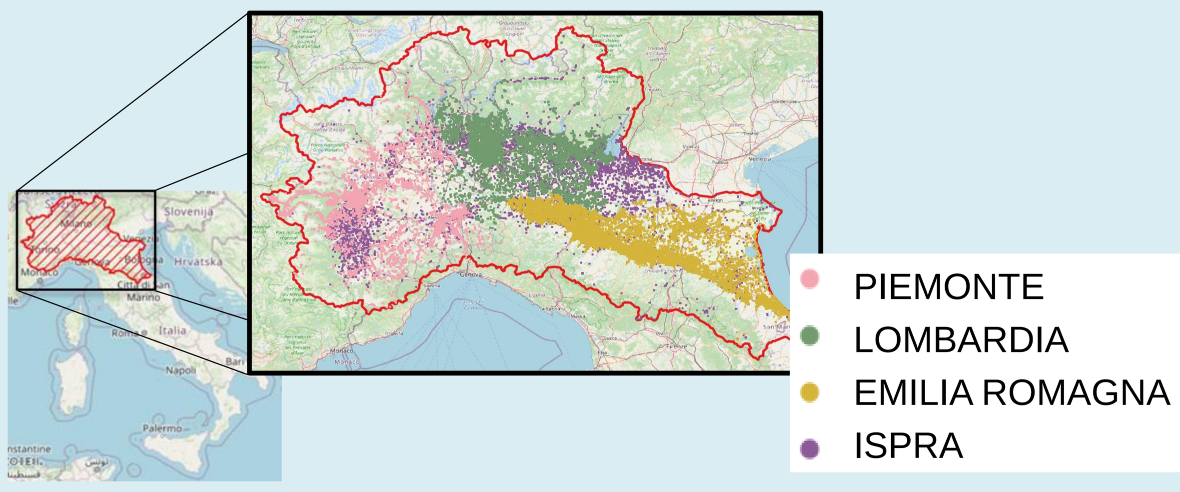

Category selection


Prediction

- Here, we set a three-dimensional (3D) georeferenced information system considering all available stratigraphic data sources, considering local, regional, and national database, available in the Po River basin scale.
- Then, we evaluate the spatial distribution of subsurface geomaterials using an approach based on Artificial Neural Networks (ANNs) trained on available data.

2. DATA COLLECTION

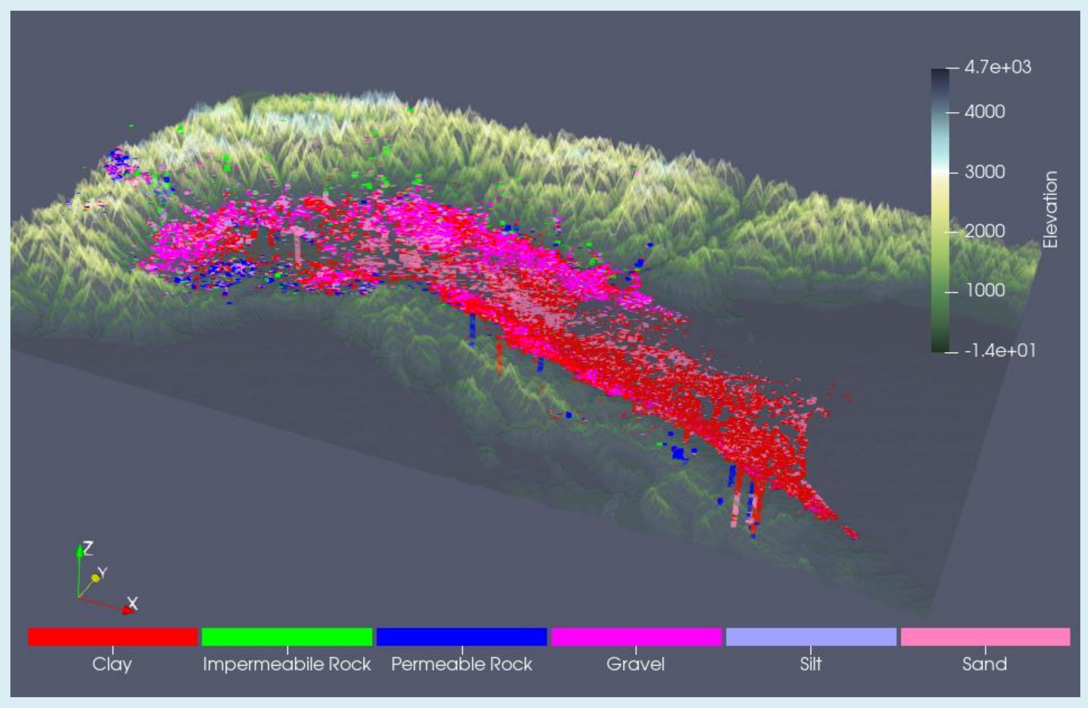
- The study area coincides with the hydrographic basin of the Po River in Northern Italy.



- Characterization has been performed on the basis of the stratigraphy obtained from the drilling of 51557 wells, for a total of 456437 stratigraphic information (Δ_i $with\ i = [1, \ldots, 456437]$)
- Available stratigraphic data here considered and collected in the study are stored in 4 databases. Data are included in three Regional database (Geoportale Lombardia, Banca dati geognostica Emilia Romagna and Banca Dati Geologica ARPA Piemonte) and one database provided by the Italian National Institute for Environmental Protection (Banca Dati Geofisici – ISPRA).

3. CATEGORIES SELECTION

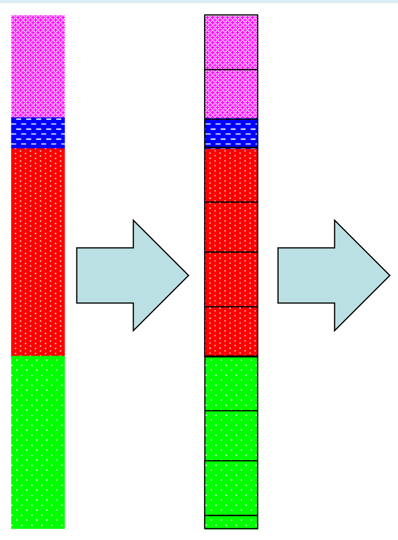
- To be usable, Δ_i must contain the geographical coordinates (latitude (X_i) and longitude (Y_i)) , upper ($Z_{u,i}$) and lower ($Z_{l,i}$) vertical coordinates, and the geomaterial description.



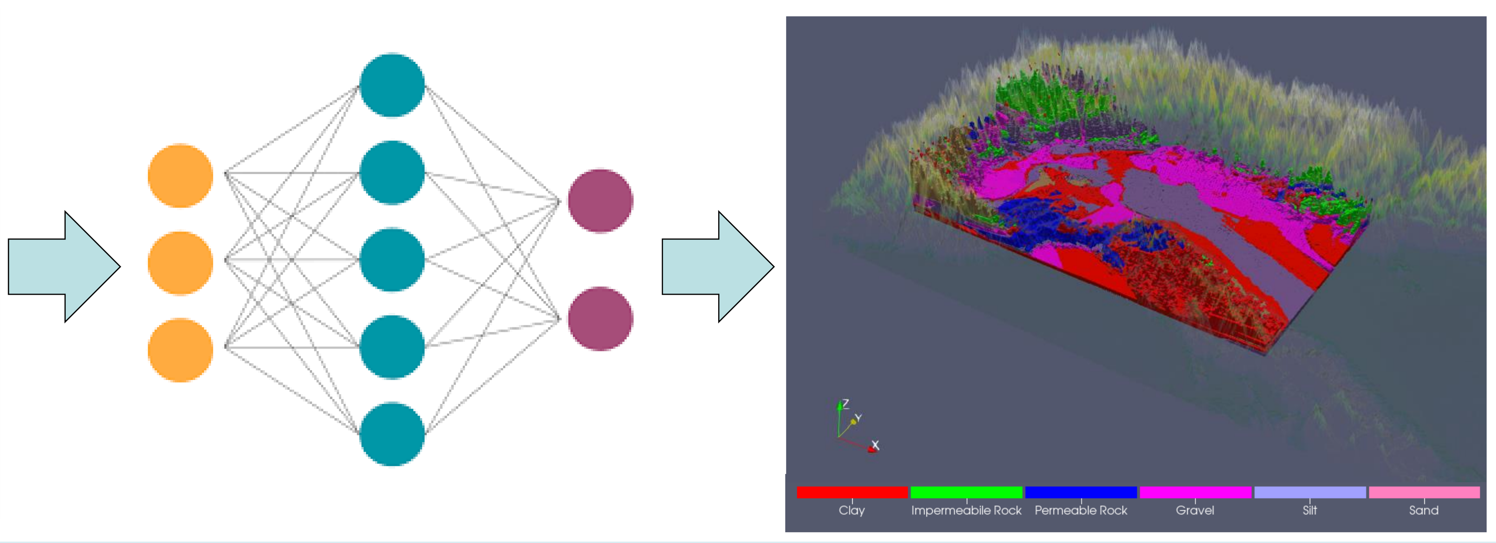
- All available data have been merged into a single 3D georeferenced lithostratigraphic data frame , where overlapping and/or inconsistent information has been automatically eliminated.
- The available descriptions were used to assign each Δ_i to one of the following categories: Gravel, sand, silt, clay, permeable rock, and impermeable rock.

4. TRAINING AND PREDICTION

- We employ a categorical Artificial Neural Network (ANN) to assess the spatial distribution of subsurface geomaterials according to the following procedure:
  - The 3D georeferenced lithostratigraphic data frame is employed to generate the training dataset. Each Δ_i with a thickness ($T_i = Z_{u,i} - Z_{l,i}$) larger than 1 m is discretized to obtain stratigraphic information Δ_j with $T_j \leq 1$ m. In this way each stratigraphic data has comparable information content during the ANN training process.
  - The training dataset is created by collecting form each Δ_j: $X_j,\ Y_j\ ,\ Z_{a,j} = (Z_{u,j} + Z_{l,j})/2,\ D_j = Z_{ground\ level}(X_j, Y_j) - Z_{a,j}$ and category (C).
- ANN's features are; (i) four nodes for the input layer (three spatial coordinates and depth with respect to the ground level), (ii) a single node associated to the lithological category for the output layer, and (iii) two hidden layers of 15 nodes.
- The data associated with the input layers are normalized to equally distribute the importance of each input.
- The trained ANN is then employed to evaluate spatial distribution of geomaterials. Our evaluation is performed on a georeferenced 3D grid with a spatial resolution of $250 \times 250 \times 4$ m and up to a depth of -200 m above sea level.



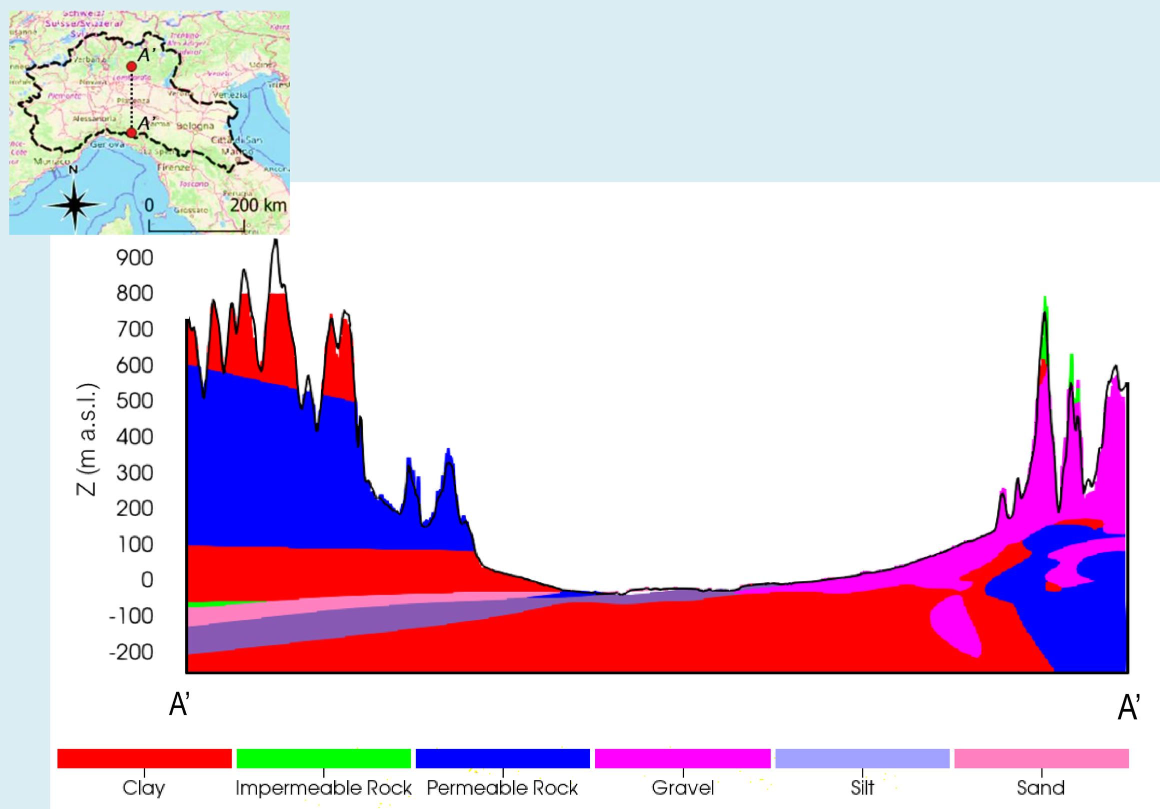
	X_j	Y_j	$Z_{a,j}$	D_j	C
Δ_1					
Δ_2					
Δ_3					
Δ_4					
…					
Δ_N					



Data Preprocessing

Training

Prediction

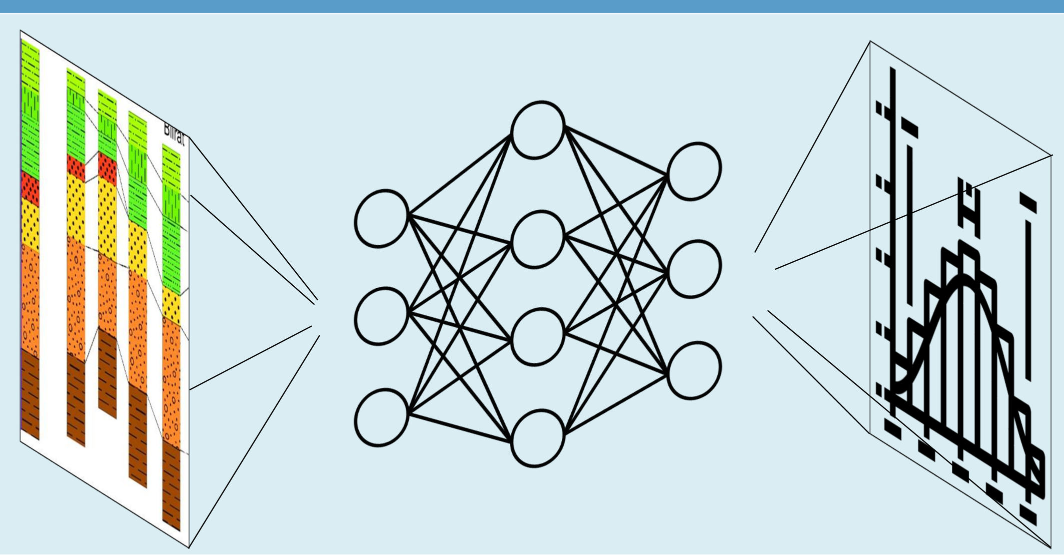


- Our ANN-based results document that deeper areas are usually characterized by low permeability geomaterials (e.g., clay), regions closer to the ground level (typically corresponding to free-surface aquifer bodies) are generally characterized by medium-high permeable geomaterials such as gravel and sand, consistent with, e.g., regional scale geological sections available in the Éupolis Lombardia report [2015].

5. ONGOING ACTIVITIES

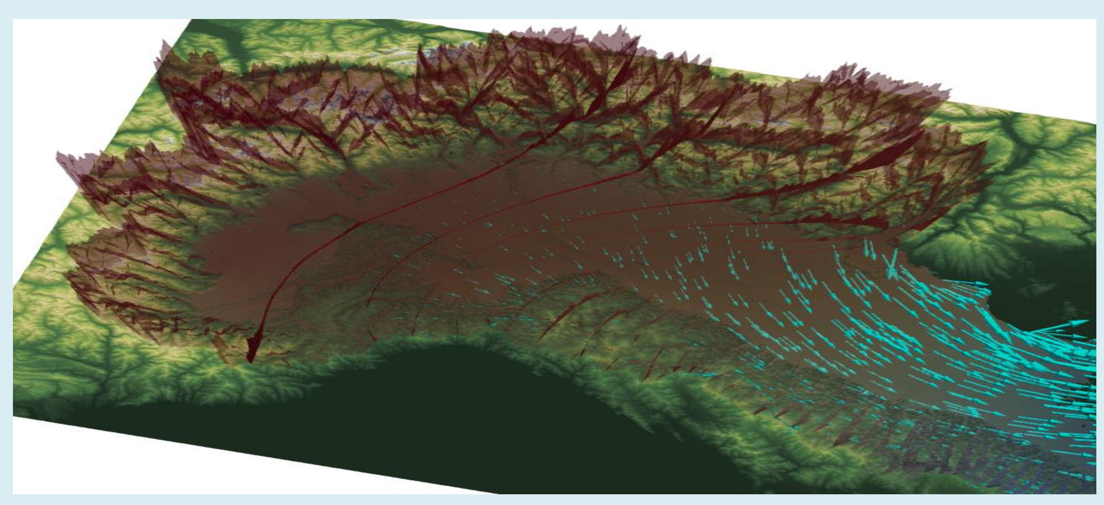
- Spatially stratified k-fold cross-validation.
  Due to the high spatial variation in observation density, the observation are spatially stratified in the domain to ensure a balanced spatial distribution within each cross-validation fold.

	Fold 1	Fold 2	Fold 3	Fold 4		Fold k
Iter. 1	Validation	Training	Training	Training	…	Training
Iter. 1	Training	Validation	Training	Training	…	Training
Iter. 1	Training	Training	Validation	Training	…	Training
	⋮	⋮	⋮	⋮		⋮
Iter. 1	Training	Training	Training	Training	…	Validation



- Uncertainty quantification.
  We are developing a probabilistic approach to infer geomaterial spatial distribution based on ANN.

- Groundwater flow modelling.
  We are employing the obtained geomaterial distribution to define the geometry, discretize the domain, define parameters, and select boundary conditions for a large-scale groundwater flow assessment within the Po River watershed.



5. REFERENCES

- Éupolis Lombardia (2015). Attività di progettazione monitoraggio e studio relative ai corpi idrici sot- terranei della Lombardia. Progetto di accompagnamento a supporto del processo di revisione del piano di tutela delle acque. Cod. Éupolis Lombardia Ter13016/001.
- Montoya, S. (2020, March 27). 3D Geological Models using Neural Networks with Python Scikit Learn and Vtk - Tutorial. Hatari Labs. https://hatarilabs.com/ih-en/3d-geological-models-using-neural-networks-with-python-scikit-learn-and-vtk-tutorial
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn. GetMobile: Mobile Computing and Communications, 19(1), 29–33. https://doi.org/10.1145/2786984.2786995