Title: Analysis of spatial Big Data for the qualitative-quantitative evaluation of geohazards through *data mining* processes

Tutor: Dr Diego Di Martire

Co-tutor: Prof. Domenico Calcaterra

Research program

The decades of collection and archiving of geological data provide access to new application opportunities in the field of Geosciences, the analysis of the large available datasets can be considered the new scientific paradigm of the 21st century. According to the estimate of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), 2.5 quintillion bytes of data are generated every day. Much of this information is now entered and managed in GIS platforms that have now achieved excellent performance for the mathematical and statistical analysis of the different informative layers (thanks to the development of specific tools) and that are constantly evolving. Moreover, in recent years, big data and artificial intelligence have provided new methods and opportunities for many applications in Geology (creation of detailed geological models of the bedrock, precision geolocation with GPS data, artificial neural networks associated with the calculation of landslide susceptibility, 3D seismic tomography, identification and estimation of geo-resources, etc.).

However, big data and data mining operations are still in their initial stage, methods and objectives are not yet fully consolidated, lacking a unified theoretical and application framework. Indeed, on one hand, current techniques for analysing big data can be transversal, such as the creation of relational geo-databases and real-time data management, on the other, specialised techniques for spatial big data based on the real situation should be developed, such as distributed spatio-temporal indexing for geological data, the evaluation of the revisiting time of a phenomenon for the purposes of hazard evaluation and the use of filtering methods that take into account the spatio-temporal relationship and geological semantics, which are peculiarities of the geology reseacher. In fact, most of the algorithms used to date for qualitative and quantitative assessments in the various sectors of use of big data are based on types of data that do not have spatial properties and therefore do not physically interact with each other and above all with the geological and geomorphological context.

The main scientific and technological issues related to the development of Geosciences in the era of big data can be listed as follows: (i) integrated storage and management of structured, semi-structured and unstructured data, big and small data, hybrid and precision data, (ii) model and data, static exploration model and dynamic monitoring model, (iii) combination of data mining and data analysis, correlation and causality unification, data mining and (iv) big data visualization of Geological Sciences.

The proposed research project aims to deepen the themes of data mining processes starting from the analysis of physical and spatial big data in order to characterise, model and verify the spatio-

temporal occurrence of possible types of geohazards that interact with the geological context of certain urban and suburban areas.
To this, initially, the use of a large amount of data already available will be considered and then continue to collect specific data according to the needs of the research project; literature data and the knowledge acquired during the training period will allow progress in the state of knowledge, adopting methodologies that are up-to-date with the state of the art. In the perspective of the development of "smart digital cities", the final objective of the research project will be the development of decision support platforms for forecasting and quantitative assessment in geohazard management starting from the analysis of multi-source geological datasets, properly processed with appropriately modelled big data analysis algorithms.

Proposal for a PhD position

On the basis of the aforementioned scientific premises, it is proposed a PhD position in which will be exploited collaborations with national and international research leading-edge centres on the analysis of spatio-temporal big data (e.g.: University of Twente, British Geological Survey, etc). We consider such collaborations fundamental for the cultural growth of the doctoral student, as well as for the achievement of shared and validated scientific results within the international scientific community. It is proposed, therefore, that the research can be articulated, during the three years of the PhD, as described below:

- first year: training is institutionally foreseen for all doctoral students of the college and specific to deal with the research topic; collection and reasoned analysis of the specific bibliography; acquisition of skills related to main methods of analysis and data mining; individuation of the study areas; collection of data for the construction of a specific database;
- second year: integration of available knowledge with ad hoc data acquisition in the study areas; development of a multi-hazard analysis procedure related to the project objectives and the identified study areas;
- third year: verification and validation of the multi-hazard analysis procedure; writing of the thesis.

Potential study areas will fall in portions of the territory affected by different geohazards known from the literature (landslides, floods...), where the availability of monitoring data is known at the same time.

Cost for field surveys and other project-related activities will be covered by the departmental funds of the Research Group in Applied Geology and Geotechnics.

The candidate should have a solid background in natural hazard analysis, mathematics, physics, geo-statistics, computer science and GIS. Knowledge of programming languages represents an appreciate skill (Matlab, R, etc.).