19th WORLD CONFERENCE OF THE ASSOCIATED RESEARCH CENTRES FOR THE URBAN UNDERGROUND SPACE
Belgrade, Serbia, November 4-7, 2025.

ACUUS BELGRADE 2025

Paper No: 5.1.136
DOI: 10.46793/ACUUS2025.5.1.136

GEOLOGICAL ASPECTS OF URBAN PERSONAS: A COMPUTATIONAL AI PIPELINE FOR THE MULTIDIMENSIONAL CHARACTERIZATION OF CITY STREETS IN THE PROVINCE OF QUÉBEC

Michael R. Doyle[1]

Abstract: City streets tend to be conceptualized from a viewpoint of a particular or general observer. Smaller-scale studies and interventions focus on the perspectival and photographic views of particular streets, providing rich but highly nuanced information. On a larger scale, streets are elements within orthographic and axonometric views from a general observer placed at infinity, situating them in a standardized way within a single coordinate space. These representations require a selection of what will be shown. As research into the urban subsurface has observed, disciplines whose work is directly impacted by the geological conditions of a site are rarely confronted by geology in the early phases of the design process. The urban underground tends to remain hidden. Long-term holistic and multidimensional planning, which includes the subsurface as one among many characterizations of the location of a project, is complicated by the increasing number of dimensions that compete for priority in the politics of territorial transformation. Geographical information systems (GIS) have helped centralize and standardize heterogeneous data sources. These information management platforms have been accompanied by work conducted on how to synthesize that data and present it on a large scale. What has yet to be explored extensively is how geology can be looked at through a high-dimensional data model that harnesses the pattern-seeking capabilities of machine-learning techniques. From the standpoint of our contemporary information technology, artificial intelligence should be able to provide an impersonal viewpoint from which to look at all the available data for a street on a planetary scale.

The project presented as part of this communication will explore this computational potential, taking as a case the Canadian province of Quebec. It recognizes that working with the components of artificial intelligence challenges our inherited modes of thinking about the world. The underlying conceptual shift is one from a personal to an impersonal subjectivity. This can be summarized somewhat figuratively by saying that the streets 'know' without knowing what they know. In conceptual terms, it is a shift away from theory-driven modelling to data-driven modelling. The design of the research project becomes less about what a street 'is' in preparation for a statistical model than about how to invent the keys to revealing what the streets 'know' without knowing they know it. As such, a street is treated as an individual whose identity is simply the fact that it is not every other street. This conception preserves the aspects of identity that are not only rational or real, but also imaginary. Mathematically speaking, we could say identity is complex and, like the complex numbers, identity comprises both the real and the imaginary. Geological conditions become part of the street's character, along with other characteristics, such as the amount of vegetation in its vicinity, the moisture content, its topography and geometry, the degree of centrality of the street in the urban network in terms of different types of activities and land uses, the qualities and value of available property, as well as photos taken by people on and around the street.

This communication will focus mainly on the method for data acquisition and processing, conducted principally in Python from freely accessible sources. The street objects were scraped from OpenStreetMap at the level of the province of Quebec using the osmnx library. The satellite data was downloaded from the Sentinel-2 satellite data using the Google Earth engine API. The 2022 georeferenced geological maps of unconsolidated surface sediments and of the underlying geological formations of the province of Quebec (an area of 1.7 million km2) were *downloaded from the web site of the Ministère des ressources naturelles et des forêts.* In order to characterize each street according to its underlying geology, the algorithm takes 50 by 50-meter samples of the two geological maps at 50-meter intervals along each street object, recording the formations for each sample area and the percentage of coverage of each. Each street segment then has n samples with each formation and the percentage of the sample it occupies. Borrowing from natural language processing, the formations are treated as words and the percentages

[1] Michael R. Doyle, PhD, Associate professor, École d'architecture, Édifice du Vieux-Séminaire, Canada, email: michael.doyle@arc.ulaval.ca

as counts, producing a matrix where each row is a street and each column is a geological formation with the corresponding percentage of each.

In order to render the streets comparable to each other, the computational setup processes the matrices of formation proportions with a one-dimensional Self-Organizing Map (SOM). The algorithm is applied as a nonparametric encoder that fits an artificial neural network to the high-dimensional space, producing an ordered set of indexes whose relative distances preserve the topological structure of the original data. Each street is assigned two metrical values, one for each near subsurface and subsurface formations, which situates it on a line of positive real numbers. As such, the values are measures of relative geological similarity to all other streets. Although the SOM is not a new algorithm, this application is novel and takes advantage of the building blocks of contemporary AI, unsupervised machine learning and largely heterogeneous datasets. Mapping these values shows unsurprisingly that the characterizations strongly resemble the geological maps because we are basically visually summarizing trends that are already visible. Running the SOM on the geological formations from both unconsolidated near subsurface and subsurface bedrock conditions tells a more nuanced story, which will be presented in this communication.

An additional exercise was performed using ChatGPT and the resulting indexes of the one-dimensional SOM. A significant challenge for architects and urban planners provided with geological information is knowing how to interpret it. Although collaboration with geologists, hydrogeologists and geotechnical engineers cannot be replaced by AI, professionals outside those expert fields might not know where to begin when confronted with a list or map of geological formations. The premise of this exercise was that, with the right prompt, an initial assessment can be provided by ChatGPT (or another pretrained transformer) that can guide future discussions with geological specialists. The approach is inspired by recent attempts to us large language models as automated proxies for expert judgment. In the resulting SOM model, each neuron indexes streets having a similar but not necessarily identical proportional distribution of geological formations in its sampled vicinity. The SOM model stores, for each neuron, a prototype vector that represents the probable distribution of the streets indexed by that neuron. This provides a single 'typical' vector of values that can be used instead of providing each individual street to ChatGPT. This option was preferred as proof-of-concept. Using the OpenAI API, each neuron's characteristic vector was provided to ChatGPT-4o with a prompt asking it to provide a brief assessment for the combination of geological formations of the presence of groundwater or aquifers, for that installation of geothermal systems for heating or cooling, for shallow to deep foundation construction and for reuse of excavated geomaterials. The call to the API specifies that ChatGPT should respond as an expert speaking to architects and urban planners, providing justifications for its assessment if necessary. Although it cannot replace the expertise of local geologists and geotechnical engineers, working with LLMs can help architects and urbanists to formulated informed questions to ask of human experts.

The presentation will conclude with new questions and next steps in the ongoing research. Because of its exploratory nature, the author will not make any claims as to the utility of the approach in solving any pressing practical problems. Nor will it attempt to say how it is better or worse than existing approaches already being tested in planning departments or proposed by other researchers. This will need to be part of future work when the prototype has been tested further. The conclusions drawn will be limited to the overall pipeline, which, in its genericness, has the advantage but also the burden of a potentially general applicability.

Keywords: artificial intelligence, geospatial data, geology, urban streets, computational methodology