Quantitative Analysis, Spatial Modelling, and Predictive Urban Grammar of Roman Cities

Daniel Tejerina
DIIn
Università degli Studi di Salerno
Fisciano, Italy
dtejerina@unisa.it

Jaime Molina Vidal
Departamento de Prehistoria, Arqueología, Historia Antigua, Filología Griega y Filología Latina
Universidade de Alicante
Spain
jaime.molina@ua.es

Francesco Colace
DIIn
Università degli Studi di Salerno
Fisciano, Italy
fcolace@unisa.it

Angelo Lorusso
Facoltà di Ingegneria e Informatica
UniPegaso
Italy
angelo.lorusso@unipegaso.it

Michele Pellegrino
DIIn
Università degli Studi di Salerno
Fisciano, Italy
mipellegrino@unisa.it

Domenico Santaniello
DIIn
Università degli Studi di Salerno
Fisciano, Italy
dsantaniello@unisa.it

***Abstract*— Understanding how Roman cities were organized has traditionally relied on qualitative archaeological interpretation, while computational approaches have often remained limited to descriptive spatial analyses or black-box predictive models. This paper proposes an interpretable computational framework for identifying, quantifying, and comparing recurrent spatial, architectural, infrastructural, and socio-economic patterns within Roman urban environments. Rather than treating archaeological sites as isolated case studies, the proposed methodology investigates whether a transferable *Roman Urban Grammar* can be inferred from measurable relationships among urban morphology, accessibility, building typology, infrastructure, land use, and demographic proxies. The research adopts Pompeii and Ostia Antica as complementary urban laboratories representing two distinct yet comparable expressions of Roman urbanism. Pompeii provides a well-preserved, predominantly low-rise urban fabric suitable for establishing the baseline analytical model, whereas Ostia Antica introduces higher density, multi-storey architecture, complex commercial functions, and port-related infrastructures that challenge the robustness and generalizability of the proposed framework. The comparative analysis enables the identification of both site-specific characteristics and recurrent spatial configurations that may extend across other Roman settlements. The proposed framework integrates relational spatial databases, Geographic Information Systems, archaeological knowledge representation, explainable machine learning, and statistical modelling to analyse interactions among accessibility, architectural typology, infrastructure, functional zoning, preservation conditions, and excavation bias. Particular attention is devoted to uncertainty-aware modelling in order to distinguish genuine archaeological patterns from biases introduced by differential preservation and documentation. Beyond functional classification, the research investigates how urban morphology correlates with residential capacity, population-density scenarios, socio-economic differentiation, and the spatial organisation of productive and commercial activities. The objective is not to automate archaeological interpretation but to provide transparent analytical tools capable of supporting hypothesis generation, quantitative comparison, and evidence-based reasoning. The resulting framework introduces a scalable methodology for comparative computational archaeology, demonstrating how interpretable Artificial Intelligence can transform exceptionally documented Roman cities into reference laboratories for modelling urban organization while preserving archaeological transparency, historical contextualization, and scholarly interpretability**

Keywords*— *Roman Urban Grammar; Computational Archaeology; Spatial Modelling; Explainable Artificial Intelligence; Urban Morphology; Cultural Heritage Analytics

I. INTRODUCTION

Roman cities represent one of the most extensively documented urban systems of the ancient world. Over more than a century of archaeological research, sites such as Pompeii and Ostia Antica have provided an extraordinary amount of information concerning architecture, infrastructure, urban planning, commercial organization, and social life. Despite this wealth of evidence, the interpretation of Roman urban organization has remained largely qualitative, relying on archaeological expertise, historical reconstruction, architectural typologies, and descriptive spatial analyses. While these approaches have produced a profound understanding of individual sites, they often make it difficult to compare urban contexts quantitatively or to evaluate whether recurring spatial configurations represent local adaptations or more general characteristics of Roman urbanism.

Recent advances in Geographic Information Systems, spatial statistics, network analysis, and Artificial Intelligence have introduced new opportunities for computational archaeology. Machine learning techniques have demonstrated considerable potential for archaeological prediction, feature extraction, and pattern recognition, while graph-based methods and spatial analytics have improved the quantitative investigation of complex urban systems. Nevertheless, many existing computational approaches remain either strongly data-driven or highly site-specific. As a consequence, they frequently provide limited interpretability, insufficient treatment of archaeological uncertainty, and only partial integration between quantitative modelling and historical reasoning.

A further limitation concerns the representation of archaeological evidence itself. Roman cities are characterized by heterogeneous preservation conditions, uneven excavation histories, varying documentation quality, and multiple levels of interpretative certainty. Computational models that neglect these factors risk confusing archaeological absence with documentary absence, producing statistically accurate but historically misleading interpretations. Consequently, uncertainty should not be regarded as a secondary issue but as an intrinsic component of archaeological knowledge that must be explicitly represented throughout the analytical process.

This paper addresses these challenges by proposing a computational framework centred on the concept of a Roman Urban Grammar. Roman Urban Grammar is defined as the set of measurable spatial, architectural, infrastructural and functional relationships that characterize Roman urban organization and can be inferred through interpretable computational modelling. Rather than assuming that Roman cities follow rigid or deterministic planning rules, the proposed framework investigates whether recurrent relationships among urban morphology, accessibility, infrastructure, architectural typology, functional organization, and demographic proxies can be identified, quantified, and compared through transparent computational methods. Within this perspective, the notion of urban grammar does not describe a fixed planning model, but an operational representation of recurring spatial relationships that can be tested quantitatively across different archaeological contexts.

To investigate this hypothesis, Pompeii and Ostia Antica are adopted as complementary urban laboratories. Pompeii provides a remarkably preserved example of a predominantly low-rise, street-oriented urban fabric, whereas Ostia represents a denser, vertically articulated, and functionally more complex urban environment closely connected to imperial commercial and port activities. Their comparison offers a controlled yet sufficiently heterogeneous experimental setting for evaluating both local spatial regularities and the transferability of computationally identified patterns.

The proposed methodology integrates GIS-based spatial databases, graph analysis, feature engineering, uncertainty-aware modelling, explainable machine learning, and comparative validation within a unified analytical pipeline. Particular emphasis is placed on model transparency and archaeological interpretability, ensuring that quantitative results remain understandable, reproducible, and directly comparable with historical and archaeological evidence.

The main contribution of this paper is therefore methodological rather than purely technological. Specifically, the paper:

- introduces the concept of Roman Urban Grammar as a quantitative and operational framework for analysing Roman urban organization;
- proposes a bias-aware computational methodology that explicitly incorporates preservation state, excavation history, documentation quality, and interpretative uncertainty into the modelling process;
- demonstrates how comparative analysis between Pompeii and Ostia Antica can distinguish site-specific characteristics from potentially transferable spatial patterns;
- provides an interpretable analytical workflow capable of supporting archaeological reasoning through reproducible quantitative evidence rather than replacing expert interpretation.

By combining archaeological knowledge with explainable computational modelling, this work contributes to the emerging field of comparative computational archaeology and provides a scalable methodological framework for investigating Roman urban systems beyond individual case studies.

II. Comparative Urban Laboratories: Pompeii and Ostia Antica

Pompeii will serve as the first modelling environment. Its exceptional preservation, extensive documentation, and relatively coherent early imperial urban fabric make it suitable for building the initial database schema, testing the variable architecture, and evaluating relationships between street access, building type, commercial concentration, domestic architecture, public spaces, and infrastructural features.

Ostia Antica will provide the second and more complex modelling environment. Its denser urban fabric, multi-storey architecture, port-related functions, horrea, insulae, and strong infrastructural articulation make it especially relevant for testing whether the model can deal with verticality, functional complexity, and a different urban and chronological context.

In terms of Comparative rationale:

- Pompeii allows the construction of the baseline data model under relatively controlled spatial conditions.
- Ostia stress-tests the model under greater density, verticality, and functional specialization.
- The comparison avoids a single-site case study while remaining feasible for the remaining doctoral timeframe.
- The two sites allow the project to move from description to validation: not only asking what each city looks like, but which rules are transferable and which are not.

III. Research Questions, Hypotheses, and Predictive Targets

The central research question can be summarized in the following way: can a transparent computational model identify recurrent spatial and functional patterns in Roman urban fabrics, reveal latent socio-economic and demographic tendencies within Pompeii and Ostia, and evaluate the transferability of selected features across other Roman urban contexts? In particular, the research questions can be the following:

- Which combinations of spatial accessibility, architectural morphology, infrastructure, and proximity to key urban features best explain the distribution of functions inside Pompeii and Ostia?

- Can computational models distinguish between different types of urban space - such as commercial, residential, storage, production, public, service, or circulation-related areas - without reducing archaeological complexity to a single deterministic label?
- How do excavation history, preservation state, documentation density, and functional uncertainty affect the reliability of the model outputs?
- Which latent socio-economic, functional, and socio-spatial patterns can be quantified within Pompeii and within Ostia independently of broader cross-site transferability?
- How do urban morphology, architectural typology, infrastructure, and accessibility correlate with demographic density, residential capacity, and socio-economic differentiation within Pompeii and Ostia?
- Which patterns are site-specific, and which may be transferable across Roman urban contexts?
- How can the results be communicated through an interactive platform that supports expert interpretation rather than black-box prediction?

In order to obtain a robust model, we need to fix the following core hypotheses:

- Roman urban fabrics contain recurrent spatial relationships between access, function, infrastructure, and density that can be measured and compared computationally.
- Pompeii and Ostia represent complementary but comparable urban laboratories: one mainly horizontal and early imperial, the other denser, more vertical, and strongly connected to metropolitan and port-related dynamics.
- Within each site, latent socio-spatial and socio-economic patterns can be detected quantitatively, even where they have often been described only qualitatively in the archaeological literature.
- Urban morphology, residential typology, verticality, accessibility, and infrastructure are expected to correlate with population-density scenarios and with forms of socio-economic differentiation.
- Bias-aware modelling can help distinguish archaeological absence from documentary absence, reducing the risk of overinterpreting uneven excavation and publication records.
- Interpretable machine learning can reveal non-obvious patterns while remaining accountable to archaeological evidence, Roman topography, epigraphy, and historical interpretation.

To avoid an overly vague use of the term prediction, the project will define a limited set of concrete predictive targets. The predictive targets summarized in Table I guide the design of the feature space and the evaluation strategy adopted throughout the computational workflow.

Predictive target	Operational meaning	Expected output
Functional probability of urban units	Estimate the probability that mapped units or areas correspond to broad functional classes such as commercial, residential, storage, production/workshop, public/service, or circulation interface.	Interpretable probability maps and feature-importance summaries, always accompanied by uncertainty indicators.
Urban accessibility and density signatures	Measure how street integration, node centrality, permeability, proximity, building dimensions, and floor-level information relate to observed functional and socioeconomic patterns.	Comparative indicators for Pompeii and Ostia, including accessibility profiles and density-related clusters.
Infrastructure and socioeconomic clustering	Identify spatial associations between infrastructure, routes, commercial fronts, storage areas, port-related functions, water/drainage indicators, and concentrations of activity.	Cluster maps, latent urban zones, and hypotheses on socioeconomic organization.
Cross-site transferability	Test whether models or variables that explain patterns in one urban laboratory retain predictive value in the other, and in selected external comparative contexts.	Validation metrics, transferability scores, and a clear distinction between local and generalizable rules.
Intra-site socio-spatial pattern discovery	Detect latent clusters and correlations within each city, such as relationships among street hierarchy, commercial frontage, residential typology, water access, storage, or vertical density.	Site-specific pattern maps and quantitatively testable hypotheses for Pompeii and Ostia.
Demographic density and residential-capacity scenarios	Model population-density proxies through built surface, housing typology, floor levels, access patterns, service catchments, and mixed-use intensity.	Scenario-based estimates and comparative indicators of demographic pressure and socio-economic differentiation.
Interactive Archaeological Intelligence output	Translate variables, model assumptions, uncertainty levels, demographic scenarios, and spatial outputs into an exploratory dashboard for historians and archaeologists.	A transparent interface for interpretation, communication, and reproducible data visualization.

Table 1 The Design of the Feature Space

The objective is not to replace archaeologists with intelligent systems but to create environments where collective archaeological reasoning can emerge and evolve.

IV. Methodological Framework

The proposed framework is designed to identify, quantify, and compare recurrent spatial and functional relationships within Roman urban environments through an interpretable computational pipeline. Rather than relying on a single predictive algorithm, the methodology integrates archaeological knowledge, spatial analysis, graph theory, statistical modelling, and explainable machine learning into a unified analytical workflow.

A. Data Collection and Spatial Representation

The analysis begins with the construction of a harmonized geospatial database integrating archaeological, architectural, and topographical information derived from Pompeii and Ostia Antica. Spatial entities are represented at multiple levels of granularity, including streets, intersections, insulae, buildings, rooms, public spaces, commercial units, storage facilities, and infrastructural elements.

Each spatial object is associated with a set of descriptive attributes including chronology, architectural typology, preservation state, functional interpretation, accessibility, construction techniques, excavation history, documentation quality, and confidence level. For Ostia, additional variables describing vertical development (e.g., number of floors and estimated elevation) are introduced to preserve the three-dimensional nature of the urban fabric.

The resulting dataset is implemented within a GIS-enabled relational database, allowing both geometric and semantic relationships to be queried simultaneously.

B. Spatial Feature Engineering

Once the urban database has been constructed, a feature engineering process transforms archaeological observations into quantitative variables suitable for computational analysis.

Morphological descriptors include building dimensions, parcel geometry, floor area, frontage length, courtyard presence, and volumetric indicators. Spatial descriptors include street hierarchy, connectivity, centrality measures, intersection density, accessibility indices, and proximity to strategic urban elements such as forums, baths, markets, temples, gates, and harbour infrastructures.

Functional variables describe residential, commercial, productive, storage, administrative, and public activities, while contextual variables model preservation quality, excavation intensity, documentation completeness, and interpretative uncertainty.

These variables collectively define the quantitative representation of the proposed Roman Urban Grammar.

C. Bias-Aware Modelling

Archaeological evidence is intrinsically incomplete and heterogeneous. Consequently, uncertainty is treated as an explicit analytical dimension rather than as residual noise.

Each observation is associated with confidence indicators derived from preservation state, excavation chronology, source reliability, and documentation density. Instead of assigning deterministic functional labels, uncertain interpretations are represented through weighted probabilities whenever appropriate.

This strategy reduces the influence of excavation bias and prevents computational models from confusing the absence of evidence with evidence of absence.

D. Pattern Discovery and Predictive Modelling

The analytical workflow combines exploratory spatial analysis with machine learning techniques to investigate both intra-site organization and cross-site transferability.

Initially, descriptive statistics, spatial autocorrelation indices, density estimation, and graph-based network analysis are employed to characterize the structural organization of each city.

Subsequently, unsupervised learning algorithms identify latent spatial clusters corresponding to functional districts, socio-economic zones, commercial corridors, and infrastructural systems.

Interpretable supervised learning models—including Logistic Regression, Random Forests, Gradient Boosting, and Elastic Net classifiers—are then employed to estimate the probability that specific urban units belong to different functional categories and to evaluate the relative contribution of each explanatory variable.

Rather than maximizing predictive accuracy alone, model selection prioritizes interpretability, robustness, and archaeological plausibility.

E. Comparative Validation

Validation is performed at three complementary levels.

First, each city is analysed independently to identify local spatial regularities and latent socio-economic structures.

Second, models trained on one urban laboratory are evaluated on the other to assess the transferability of learned spatial relationships and to distinguish site-specific behaviours from potentially general Roman urban patterns.

Finally, all computational outputs are systematically compared with archaeological literature, expert interpretations, epigraphic evidence, and historical topography to ensure that statistical findings remain historically meaningful.

F. Explainable Archaeological Intelligence

The final stage of the framework consists of transforming analytical outputs into an interactive decision-support environment.

Probability maps, feature importance rankings, accessibility indicators, uncertainty layers, demographic scenarios, and socio-spatial clusters are integrated within an exploratory visualization platform that allows archaeologists to inspect, validate, and interpret model results.

The proposed system therefore does not aim to replace archaeological reasoning, but to augment it by providing transparent, reproducible, and quantitatively testable evidence capable of supporting comparative studies across Roman urban contexts.

.

CONCLUSIONS

This paper has presented a computational framework for the quantitative analysis of Roman urban environments through the concept of a Roman Urban Grammar. Rather than viewing ancient cities as isolated archaeological case studies, the proposed approach investigates whether recurrent spatial, architectural, infrastructural, and socio-economic relationships can be identified, quantified, and compared using transparent and interpretable computational methods.

The comparative adoption of Pompeii and Ostia Antica as complementary urban laboratories represents the core methodological contribution of this research. The two cities provide contrasting yet compatible urban configurations that allow the proposed framework to distinguish between site-specific characteristics and potentially transferable patterns of Roman urban organization. This comparative perspective moves beyond descriptive urban analysis by introducing a reproducible methodology capable of evaluating how accessibility, morphology, infrastructure, functional zoning, and demographic proxies interact within complex archaeological landscapes.

A central aspect of the proposed framework is the explicit treatment of archaeological uncertainty. Preservation state, excavation history, documentation quality, and interpretative confidence are incorporated as analytical variables rather than being considered external limitations. In this way, the computational models are designed not only to identify latent spatial patterns but also to provide results that remain consistent with archaeological reasoning and historical interpretation.

The proposed methodology does not aim to replace the expertise of archaeologists or historians. Instead, it provides a transparent analytical environment in which quantitative evidence can support hypothesis generation, comparative reasoning, and the evaluation of alternative interpretations. The adoption of explainable machine learning, graph-based spatial analysis, and reproducible data pipelines ensures that computational results remain understandable, verifiable, and scientifically accountable.

Beyond the specific case studies considered in this work, the framework offers a scalable methodological foundation for comparative computational archaeology. Future developments will focus on extending the proposed Roman Urban Grammar to additional urban contexts across the Roman world, refining predictive models through larger archaeological datasets, and integrating advanced spatial reasoning techniques to investigate broader questions concerning urban evolution, socio-economic organization, and demographic dynamics.

Ultimately, this work argues that the next generation of computational archaeology should not be measured solely by increasingly sophisticated algorithms, but by its ability to transform archaeological evidence into transparent, interpretable, and reproducible scientific knowledge. The proposed Roman Urban Grammar represents a first step toward this objective by establishing a general methodological framework through which ancient urban systems can be analysed, compared, and interpreted using quantitative yet historically grounded approaches.

References

[1] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. New York: Springer.

[2] Carriero, V. A., Gangemi, A., Mancinelli, M. L., Marinucci, L., Nuzzolese, A. G., Presutti, V., & Veninata, C. (2019). ArCo: The Italian Cultural Heritage Knowledge Graph. In The Semantic Web - ISWC 2019 Satellite Events. Cham: Springer.

[3] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). An Introduction to Statistical Learning with Applications in Python. Cham: Springer.

[4] Rey, S. J., & Anselin, L. (2010). PySAL: A Python library of spatial analytical methods. In M. M. Fischer & A. Getis (Eds.), Handbook of Applied Spatial Analysis (pp. 175-193). Berlin: Springer.

[5] Zhang, L., & Chiaradia, A. J. F. (2021). Urban volumetrics: From space syntax to three-dimensional spatial modelling and wayfinding analysis. Environment and Planning B: Urban Analytics and City Science.

[6] Bevan, A. (2012). Spatial methods for analysing large-scale artefact inventories. Antiquity, 86(332), 492-506. Daniel Tejerina - Revised PhD Proposal - Page 10

[7] Clarke, D. L. (1968). Analytical Archaeology. London: Methuen.

[8] Conolly, J., & Lake, M. (2006). Geographical Information Systems in Archaeology. Cambridge: Cambridge University Press.

[9] Hillier, B., & Hanson, J. (1984). The Social Logic of Space. Cambridge: Cambridge University Press.

[10] Verhagen, P., & Whitley, T. G. (2012). Integrating archaeological theory and predictive modeling: A live report from the scene. Journal of Archaeological Method and Theory, 19, 49-100.

[11] Wheatley, D., & Gillings, M. (2002). Spatial Technology and Archaeology: The Archaeological Applications of GIS. London: Taylor & Francis.