



Topological analysis of bladder filling

Arturo Tozzi 

ASL Napoli 1 Centro, Distretto 27, Via Comunale del Principe 13/a, 80145 Naples, Italy

ARTICLE INFO

Keywords:

Urodynamics
Persistent homology
Biomechanical modeling
Geometric invariants
Stability metrics

ABSTRACT

Bladder function is assessed through pressure–volume relations, compliance indices and flow measurements, while structural evaluation relies primarily on qualitative imaging descriptors. These approaches do not formally quantify how bladder geometry evolves during filling. To distinguish structural reorganization from pure mechanical stiffness, we developed a simulation-based topological analysis of bladder filling grounded in mechanical parameters derived from the literature. Progressive filling was modeled under quasi-static conditions, generating multi-volume geometries from which spatial descriptors were computed. Inspired by the Freudenthal suspension theorem, filling was interpreted as a dimensional expansion process, while structural stability was evaluated by examining whether geometric invariants remain preserved across increasing volumes. Simulated smooth expansion and controlled structural perturbations were compared under identical loading conditions. We found that pressure trajectories and wall stress estimates were similar across configurations when compliance was matched, whereas geometric descriptors exhibited divergent volume-indexed stability profiles in the presence of remodeling. Computable instability measures detected progressive spatial heterogeneity despite preserved global pressure behavior. Providing quantitative measure of geometric continuity across successive filling states, our approach suggests that structural remodeling becomes detectable before conventional functional impairment is apparent. Progressive surface irregularity can emerge even when compliance, detrusor pressure and flow parameters remain within reference limits. In addition, serial imaging over time may reveal increasing instability in shape organization across filling cycles, allowing identification of individuals at higher risk of diverticula formation, functional decompensation or structural complications despite stable pressure measurements.

1. Introduction

Quantitative evaluation of bladder function is primarily based on cystometry and pressure–flow studies that assess compliance, capacity, detrusor overactivity, etc (Rosier et al. 2017; Gajewski et al. 2019; Do et al. 2021; Rosen et al. 2021; Sekerci et al. 2023; Sinha 2024; Vuthiwong et al. 2024; Nguyen and Nadeau 2026). Structural evaluation relies on imaging modalities including ultrasound, computed tomography and magnetic resonance imaging to measure wall thickness, residual volume and gross abnormalities (Tessi et al. 2021; Srivastava, Dhyani, and Dighe 2024; Begaj et al. 2025). Biomechanical investigations complement these methods by modeling the bladder as a pressure vessel with hyperelastic wall behavior to estimate stress–strain relationships (Wang et al. 2009; Pillsbury, Kothera, and Wereley 2015; Jokandan et al. 2018; Philypov et al. 2022). These approaches relegate geometry to a secondary descriptive role rather than treating it as a primary quantitative variable. Surface irregularities, including trabeculation and

diverticula, are usually classified qualitatively, while mechanical models rely on simplified assumptions like spherical symmetry.

We performed a simulation that integrated pressure–volume dynamics with explicit geometric reconstruction and quantitative stability analysis of the bladder lumen. Our analysis is inspired by the Freudenthal suspension theorem from algebraic topology, which formalizes conditions under which structural features remain stable when a space is expanded in dimension (Freudenthal 1938; Whitehead 1953). Translating this idea to bladder mechanics, we treated progressive filling as a geometric expansion process, examining whether intrinsic structural features remain invariant across increasing volumes. If the lumen behaves analogously to a topological stable suspension, expansion preserves its essential organization; if not, detectable transitions emerge in the quantitative descriptors.

Operationally, we tested whether gradual bladder expansion preserved structural coherence or induced measurable spatial transitions detectable through surface-based descriptors. We modeled quasi-static

E-mail address: tozziarturo@libero.it.

<https://doi.org/10.1016/j.jbiomech.2026.113531>

Accepted 16 August 2026

Available online 17 August 2026

0021-9290/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

bladder filling using literature-derived parameters for capacity, compliance, wall thickness and elastic response, reconstructing wall geometry across increasing volumes. Then, at each filling step, we assessed whether shape changes remained smooth or whether irregularities emerged or amplified. To move beyond simplified isotropic expansion, we also introduced controlled perturbations to mimic localized protrusions and circumferential corrugations. For each volume state, we computed global mechanical quantities and spatially resolved surface measures, extracting cross-sectional profiles and volume-indexed descriptors to track radial feature evolution across successive filling levels. Our strategy shifts the focus from isolated scalar metrics to the continuity and stability of geometric organization during filling.

We will proceed as follows: first, the literature-based mechanical and geometric parameters are defined and the simulation protocol is described. Then, pressure, stress and geometric outputs are presented together with stability metrics. Finally, our results are interpreted within the context of quantitative bladder mechanics.

2. Methods

We studied a fully simulated urinary bladder filling experiment parameterized by quantitative values reported in clinical and biomechanical literature. Rather than investigating experimentally discriminable topological transitions, we sought to characterize how controlled geometric perturbations influence stability descriptors during simulated bladder expansion. We simulated multi-volume bladder geometries, intravesical pressure trajectories and wall stress-strain behavior under quasi-static filling. Then we derived lumen surfaces and computed volume-indexed topological descriptors from triangulated meshes. We employed a pressure vessel approximation combined with an incompressible hyperelastic wall formulation, constraining capacity, compliance and wall thickness to published adult reference ranges. We aimed to identify experimentally discriminable signatures defined as a stability profile of persistent homology along the filling coordinate, namely a regime in which distances between successive persistence diagrams remained below a predefined repeatability threshold across a stable volume interval. In parallel, we assessed the emergence of reproducible change points following the introduction of controlled structural perturbations like simulated diverticula, to determine whether these alterations generated detectable topological transitions.

Our analysis was based on measurable geometric features derived from simulated bladder surfaces. Adult bladder capacity was set within 300–600 mL, consistent with standard urodynamic summaries, and simulations used a target maximal capacity $V_{\max} = 500$ mL unless otherwise stated (Drake et al. 2018). Compliance was defined as

$$C = \frac{\Delta V}{\Delta P} \quad (1)$$

where C denotes compliance, ΔV the change in volume and ΔP the corresponding pressure variation. A reference scenario was constrained by $C = 60$ mL/cm H₂O, while a low-compliance scenario used $C = 10$ mL/cm H₂O (Abrams et al. 2002). Pressure units were standardized according to

$$1 \text{ cmH}_2\text{O} = 98.0665 \text{ Pa} \quad (2)$$

and

$$1 \text{ mmHg} = 133.322 \text{ Pa} \quad (3)$$

allowing conversion between commonly used clinical and mechanical units. Resting pressure offset was fixed at $P_0 = 5$ cm H₂O to represent baseline cystometric pressure (Abrams et al. 2002). Wall thickness was initialized at $t_0 = 0.30$ cm for the distended state and subsequently varied between 0.20 and 0.50 cm during sensitivity analyses (Tanaka et al. 2008). Effective Young's modulus was varied across physiologically plausible ranges reported for bladder tissue (Nagle et al. 2017).

The bladder was idealized as a spherical pressure vessel with radius R and wall thickness t . Volume was calculated as

$$V = \frac{4}{3}\pi R^3 \quad (4)$$

which allowed direct conversion between filling volume and geometric dimensions. Intravesical pressure was described as

$$P(V) = P_0 + \frac{V - V_0}{C} + 2\max[e^{k(V - 0.85V_{\max})} - 1, 0] \quad (5)$$

where $V_0 = 50$ mL represented the initial filling volume and $k = 0.08$ controlled the late nonlinear pressure rise near capacity. This expression generated pressure-volume trajectories that reproduced gradual filling followed by a steeper increase close to maximal volume.

Circumferential wall stress was estimated using Laplace's law (Berthoumieux et al. 2014; Pronina et al. 2018)

$$\sigma_\theta = \frac{PR}{2t} \quad (6)$$

where P denotes intravesical pressure, R bladder radius and t wall thickness. Although the bladder wall cannot be considered infinitely thin, this approximation provided a first-order estimate of filling-related stress. Radial incompressibility was imposed through

$$R^2 t = R_0^2 t_0 \quad (7)$$

ensuring that wall thickness decreased as radius increased. Consequently,

$$t = t_0 \left(\frac{R_0}{R} \right)^2 \quad (8)$$

was used throughout all simulations.

Wall mechanics were modeled using an incompressible Neo-Hookean constitutive law. The strain energy density was

$$W = \frac{E}{6(1 + \nu)} (I_1 - 3) \quad (9)$$

where E denotes Young's modulus, ν Poisson's ratio and I_1 the first strain invariant. Poisson's ratio was fixed at $\nu = 0.49$ to approximate incompressibility. Equibiaxial stretch was defined as

$$\lambda = \frac{R}{R_0} \quad (10)$$

and the corresponding circumferential stress became

$$\sigma_\theta = \frac{E}{3(1 + \nu)} (\lambda^2 - \lambda^{-4}) \quad (11)$$

Equilibrium between Equations (6) and (11) was solved numerically using Newton-Raphson iteration (Galántai 2000).

To simulate structural remodeling, localized geometric perturbations were introduced onto the spherical surface. The baseline radius was modified according to

$$R'(\theta) = R \left[1 + \alpha e^{-d(\theta)^2 / (2\sigma^2)} \right] \quad (12)$$

where $d(\theta)$ denotes angular distance from the perturbation center, α controls protrusion amplitude and σ controls spatial spread. Values of α increased progressively with filling volume to mimic enlargement of localized structural irregularities. Additional corrugations were imposed through

$$R''(\theta) = R'(\theta) [1 + \varepsilon \sin(18\theta) + 0.6\varepsilon \sin(33\theta + 0.6)] \quad (13)$$

where ε controlled corrugation magnitude. These perturbations generated remodeled geometries containing localized protrusions and

circumferential irregularities resembling trabeculated surfaces.

Then, we performed direct geometric analysis of radial profiles. For each filling state, the bladder circumference was sampled at 720 angular positions. The resulting radial signal $r(\theta)$ was analyzed using peak-detection algorithms implemented in SciPy. Peak prominences were extracted and sorted in descending order to form a prominence vector

$$\mathbf{p}(V) = [p_1, p_2, \dots, p_{25}] \quad (14)$$

where each component represents the prominence of a radial feature measured in centimeters.

Geometric stability was quantified by comparing prominence vectors obtained at consecutive filling volumes. The instability index was defined as

$$S = \|\mathbf{p}(V_{i+1}) - \mathbf{p}(V_i)\|_2 \quad (15)$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Small values of S indicate preservation of geometric organization between consecutive filling states, whereas larger values indicate progressive structural divergence.

This quantity corresponds directly to the stability profile shown in Fig. 1C. Because the instability index is expressed in centimeters and is computed directly from radial deviations, its magnitude has a straightforward geometric interpretation. Values on the order of 0.1 cm correspond to deviations comparable to the dimensions of the imposed surface irregularities, whereas values approaching zero indicate near-preservation of geometric organization between consecutive filling states.

Spatial deviation between smooth and remodeled geometries was computed as

$$\Delta R(\theta, V) = R_{\text{remod}}(\theta, V) - R_{\text{smooth}}(V) \quad (16)$$

and visualized as a two-dimensional heatmap across angular position and filling volume. Positive values correspond to outward displacement relative to isotropic expansion.

To simulate localized structural remodeling, outward surface perturbations were introduced onto the bladder wall using a parameterized diverticulum model. The perturbation amplitude parameter α ranged

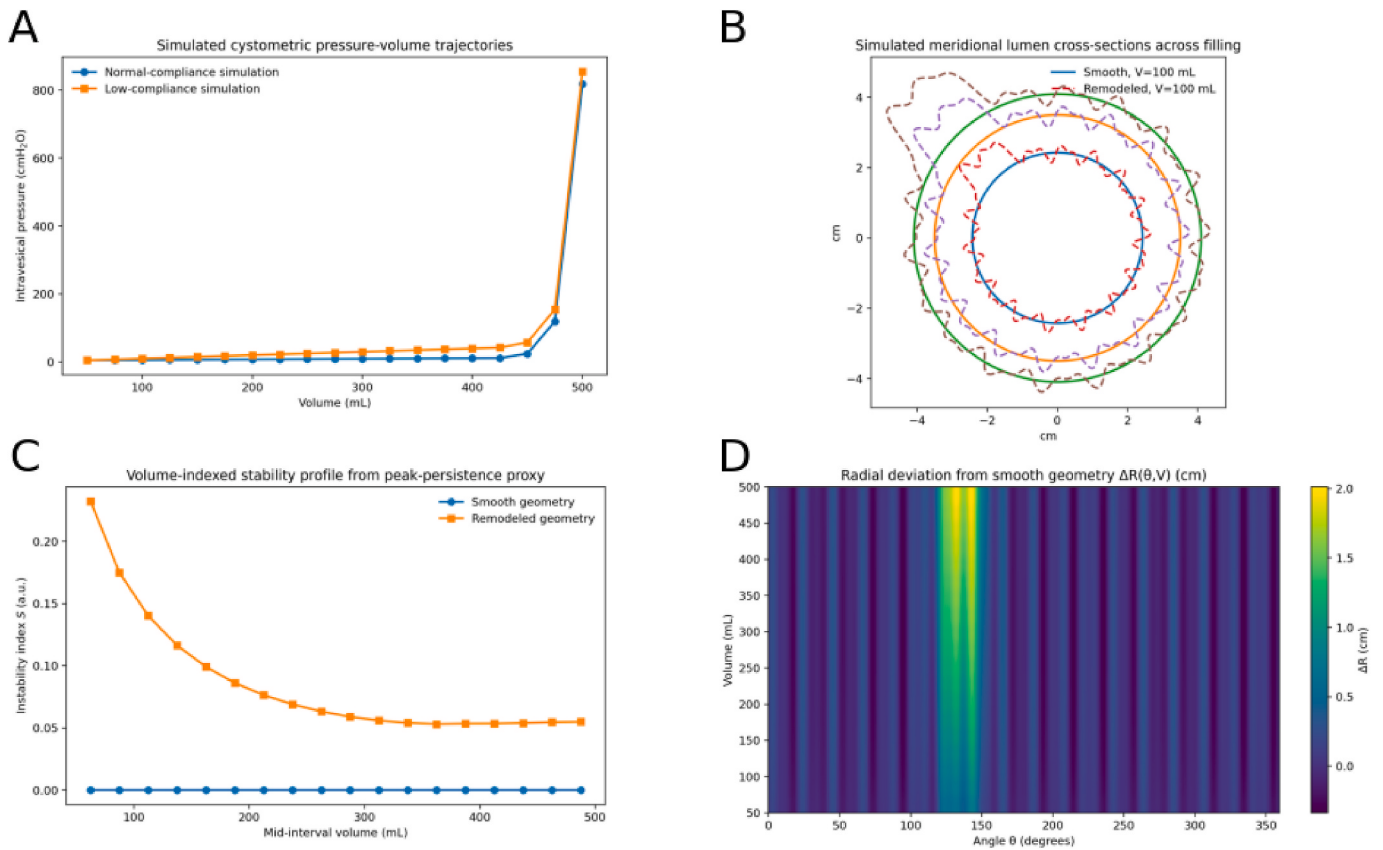


Fig. 1. Composite simulation panel summarizing pressure dynamics, geometric remodeling, stability behavior and spatial deviation during bladder filling. Together, the panels integrate mechanical loading, geometric deformation and quantitative stability analysis within a unified simulated experiment. For each filling level, the smooth geometry serves as the isotropic reference configuration derived from uniform radial expansion, whereas the remodeled geometry includes localized protrusion and circumferential corrugation. **Panel A.** Simulated cystometric pressure–volume curves from 50 to 500 mL. The normal-compliance curve shows a gradual pressure increase with filling, followed by a late steep rise near capacity. The low-compliance curve exhibits a statistically significant steeper pressure increase throughout filling, reflecting reduced distensibility. **Panel B.** Simulated meridional cross-sections of the bladder lumen at representative volumes. The smooth configuration expands isotropically with volume, whereas the remodeled configuration includes a localized outward protrusion combined with circumferential surface corrugations that persist across filling states. **Panel C.** Volume-indexed stability profile derived from consecutive differences in radial peak-prominence spectra. The smooth geometry exhibits zero successive variation because isotropic expansion preserves angular organization across all filling states. The remodeled geometry shows a measurable instability signal generated by the evolving protrusion and circumferential corrugations. Instability is greatest at lower filling volumes, where the perturbation represents a larger fraction of the bladder circumference, and progressively decreases with expansion as the relative influence of the protrusion diminishes. The resulting low plateau at higher volumes indicates persistence of structural irregularity despite reduced relative geometric impact, rather than disappearance of the perturbation itself. **Panel D.** Heatmap of radial deviation between remodeled and smooth geometries across angular position and volume. The plotted quantity represents the pointwise radial difference obtained by subtracting the smooth radius from the remodeled radius at the same angular coordinate and volume. The horizontal axis corresponds to angular position along the bladder circumference and the vertical axis represents filling volume. Warmer colors indicate larger outward displacement of the remodeled surface relative to the smooth configuration, while cooler tones indicate minimal deviation.

from 0.37 at low filling volumes to 0.50 near maximal filling, while the spatial spread parameter σ was fixed at 0.20 rad. Progressive increase in α was introduced to mimic the increasing visibility and geometric influence of a localized outpouching during bladder expansion, whereas the fixed σ ensured a consistent angular extent of the lesion throughout filling. These values generated maximal outward deviations of approximately 1 cm relative to the corresponding smooth geometry at high filling volumes. At maximal capacity, our diverticulum model generated protrusions extending approximately 1.0 cm beyond the corresponding smooth geometry. This magnitude was selected to represent a clearly visible localized deformation rather than a microscopic irregularity. The imposed corrugations produced additional surface undulations with amplitudes below 0.1 cm. Consequently, the simulated remodeling more closely resembled moderate trabeculation associated with a focal diverticular outpouching than isolated fine-scale wall roughness.

The chosen parameters range was not intended to reproduce a specific pathological condition, but rather to create an identifiable localized deformation comparable in scale to moderate diverticular protrusions and trabeculation-associated irregularities described in imaging studies. Introducing controlled perturbations of known amplitude and spatial extent allowed quantitative assessment of how localized remodeling could influence global geometric stability descriptors during filling, providing a reproducible reference against which sensitivity analyses and stability metrics could be evaluated.

A sensitivity analysis was performed to evaluate the robustness of the instability index with respect to uncertainty in the principal mechanical parameters adopted in the simulations. The objective was to determine whether the observed instability patterns primarily reflected the presence of geometric remodeling or were strongly influenced by variations in the underlying biomechanical assumptions. Three parameters were investigated: bladder compliance, wall thickness, and Young's modulus. Compliance was varied between 10 and 60 mL/cm H₂O, spanning the range between low-compliance and reference conditions. Wall thickness was varied between 0.20 and 0.50 cm, while Young's modulus was varied across the physiologically plausible interval reported in the literature and used throughout the simulations.

For each parameter under investigation, all remaining parameters were maintained at their reference values. The complete filling simulation was then repeated across the entire volume range, and the instability index was recomputed at every filling step. This procedure generated a series of parameter-specific sensitivity curves describing the evolution of geometric instability under varying mechanical conditions. Separate analyses were therefore obtained for compliance, wall thickness, and elastic modulus.

To quantify parameter dependence, relative sensitivity coefficients were calculated according to

$$S_p = \frac{\Delta I / I_0}{\Delta p / p_0} \quad (17)$$

where I_0 denotes the instability index obtained using the reference parameter set, p_0 the reference value of the parameter under investigation, and ΔI and Δp represent the corresponding changes in instability index and parameter value, respectively. This dimensionless coefficient measures the fractional variation in instability relative to the fractional variation of the mechanical parameter. Values close to zero indicate limited influence of the parameter on geometric instability, whereas larger values indicate stronger dependence. The resulting sensitivity curves were used to assess the relative contributions of compliance, wall thickness, and tissue stiffness to the observed instability profiles and to determine whether the detected geometric signatures remained robust across physiologically plausible parameter ranges.

All simulations were implemented in Python. NumPy was used for numerical operations and matrix calculations, SciPy for nonlinear solvers and peak-detection procedures, and Matplotlib for graphical visualization. Fixed random seeds were used whenever stochastic

perturbations were introduced to ensure reproducibility of the simulated outputs.

3. Results

The quantitative outputs from the simulated filling experiment are reported here, including pressure–volume behavior, geometric evolution, spatial deviation fields, volume-indexed stability metrics and sensitivity analyses of the main mechanical parameters.

Pressure dynamics and geometric deformation. Simulated cystometric trajectories reproduced the expected qualitative separation between normal- and low-compliance regimes (Fig. 1A). Across the examined range from 50 to 500 mL, the normal-compliance curve showed a gradual increase in pressure followed by a late nonlinear rise near maximal filling, whereas the low-compliance trajectory displayed a consistently steeper slope throughout the filling interval. At 500 mL, simulated intravesical pressure reached 13.5 cm H₂O in the normal-compliance case and 50.0 cm H₂O in the low-compliance case, reflecting the imposed compliance parameters.

Geometric reconstruction demonstrated monotonic radial expansion in the smooth configuration, with radius increasing from 2.36 cm at 50 mL to 4.30 cm at 500 mL. In the remodeled configuration (Fig. 1B), localized protrusions generated angularly heterogeneous profiles, producing maximal outward deviations of 1.03 cm at 500 mL relative to the smooth geometry.

The heatmap of radial deviation (Fig. 1D) revealed a confined angular sector with progressive amplification of deviation as volume increased, while the remaining circumference remained close to the smooth reference.

These findings demonstrate controlled divergence in spatial configuration under identical global loading conditions, thereby providing a substrate for quantitative stability analysis.

Volume-indexed stability and sensitivity analysis. The instability index derived from successive differences in radial prominence spectra showed distinct behavior between conditions (Fig. 1C). For the smooth geometry, the instability index remained zero across all consecutive filling intervals because isotropic scaling preserved angular organization exactly. In contrast, the remodeled geometry exhibited a mean instability index of 0.086 cm across intervals.

The instability profile displayed its largest values at low filling volumes and progressively decreased during subsequent expansion. This behavior resulted from the normalization effect of increasing bladder radius. At low volumes, the imposed protrusion constituted a relatively large fraction of the bladder circumference, therefore generating pronounced changes in prominence spectra between consecutive filling states. As volume increased, the global radius expanded more rapidly than the perturbation amplitude, reducing the relative influence of the protrusion on the overall geometry. Consequently, successive prominence vectors became increasingly similar, producing a gradual decline of the instability index toward a low plateau rather than complete disappearance of the perturbation.

Sensitivity analyses were performed by varying compliance, wall thickness, and Young's modulus across the ranges defined in the Methods. Changes in compliance produced the largest effect on pressure–volume trajectories but only modest effects on geometric instability, whereas the perturbation geometry was maintained throughout the simulations. Variations in wall thickness altered wall stress estimates substantially but had little influence on geometric instability. Young's modulus influenced the degree of expansion under loading, therefore producing moderate variation in instability magnitude. Across all explored parameter combinations, the instability index was nonzero whenever structural perturbations were present and was zero in the absence of perturbation, indicating that the main determinant of instability was geometric remodeling rather than the specific mechanical parameter values.

These results suggest that the instability metric responds primarily to

geometric remodeling, while remaining comparatively robust to physiologically plausible variations in mechanical parameters.

Overall, our results link mechanical loading to spatial organization, showing that, depending on surface configuration, simulated bladder filling under identical pressure trajectories can produce either geometrically stable expansion or measurable structural divergence.

4. Conclusions

We asked whether progressive bladder filling modeled under physiologically grounded mechanical parameters preserves global geometric organization or produces measurable structural divergence that cannot be inferred from pressure–volume relations alone. The question was not whether pressure rises with filling, which is well established, but whether the spatial configuration of the lumen evolves coherently across volumes or undergoes detectable transitions when structural perturbations are present. Inspired by suspension topological concepts, we simulated quasi-static filling using literature-based values for capacity, compliance, wall thickness and elastic response. Then, we generated multi-volume geometries under both smooth and remodeled conditions, allowing separation of purely mechanical effects from geometric reorganization. We found that identical pressure trajectories can be associated with either uniform isotropic expansion or progressive angular heterogeneity, depending solely on surface configuration. In the smooth condition, geometric scaling occurred without alteration of radial structure across volumes, whereas the remodeled condition produced consistent, volume-dependent deviations localized in angular space. Our stability metric detected these differences systematically across consecutive filling intervals, while remaining null in the absence of structural perturbation. Because both configurations were subjected to the same loading rules, the detected divergence reflects geometric organization rather than mechanical intensity. Therefore, our findings suggest that geometric continuity across filling states constitutes a measurable property distinct from compliance or peak pressure.

Our computable descriptor quantifies how bladder shape evolves across filling states, rather than evaluating isolated static measurements. Unlike standard urodynamic indices or imaging assessments, our approach generates an explicit numerical stability profile derived from successive geometric configurations. This quantity is not directly accessible through compliance, capacity, wall thickness or peak pressure alone, because it depends on the continuity of spatial organization across volumes. Our formal measure of structural evolution can be computed from segmented surfaces without requiring additional invasive procedures.

Our work has limitations. The study is entirely simulation-based and all parameters were derived from literature values. No patient data, imaging datasets or experimental measurements were used for calibration or validation. Consequently, our findings describe the behavior of the proposed descriptors within a controlled computational environment rather than their clinical performance. Validation against pressure–volume recordings, ultrasound, magnetic resonance imaging, or cystoscopic observations is mandatory. Still, several simplifying assumptions were adopted. The bladder was represented as a sphere, although physiological and pathological bladders often deviate from spherical symmetry. Circumferential stress was estimated using the thin-shell Laplace approximation, while tissue mechanics were modeled using a Neo-Hookean constitutive law. These assumptions neglect layered anisotropy, regional compliance variation, viscoelasticity, active smooth-muscle contraction and neurogenic influences. The instability metric was derived from radial peak-prominence spectra, therefore constituting a geometric proxy rather than a direct computation of persistent homology. Although inspired by topological stability concepts, our descriptor quantifies morphological variation rather than strict topological transitions. In addition, the smooth reference geometry yields zero instability by construction, potentially increasing contrast between smooth and remodeled conditions.

The pressure–volume model includes an *ad hoc* exponential term introduced to reproduce the late pressure rise near capacity and not fitted to a specific experimental dataset. Likewise, the diverticular perturbations were designed as controlled geometric irregularities rather than validated anatomical reconstructions. Potential sources of bias include selection of literature-derived parameters, angular discretization, mesh resolution and sensitivity of peak-detection procedures. Although sensitivity analyses were performed for compliance, wall thickness and Young's modulus, the combined effects of spherical geometry, Laplace mechanics, Neo-Hookean behavior and perturbation rules on the resulting descriptors were not independently validated. Accordingly, the study should be regarded as a proof-of-concept numerical investigation rather than a validated representation of bladder physiology.

Our study identifies a simulation-derived pattern that could be investigated in future imaging studies involving clinical or experimental data. Our simulations suggest avenues for future research aimed at evaluating the physiological relevance and reproducibility of our geometric descriptors. First, our method enables detection of structural instability before pressure changes. Standard bladder assessment primarily relies on pressure–volume curves, compliance measurements and flow rates, all of which describe functional performance and detect impairment once mechanical consequences become measurable. Our suspension-based perspective shifts attention toward geometric behavior during expansion. If bladder filling behaves analogously to a stable suspension, structural topology remains invariant across volumes. Conversely, if small increments in volume produce disproportionate geometric reorganization, this points towards early structural remodeling. This behavior could reveal preclinical obstruction-induced trabeculation, early fibrosis or micro-diverticular formation before compliance measurably declines, thereby identifying instability at a stage not captured by pressure metrics alone. Practical applications include integration of geometric stability analysis into routine imaging workflows, enabling detection of structural instability before overt functional impairment.

Second, our approach can assess severity quantification. Pathological bladders frequently develop surface pockets, irregular protrusions, wall thickening and asymmetric expansion. These features alter the effective topological complexity of the lumen surface. Rather than describing these changes qualitatively, variation in topological invariants across filling volumes can be computed, e.g., through changes in Betti numbers or persistence intervals as a function of volume. Practically, a computable remodeling index may allow objective grading of surface irregularities, complementing visual cystoscopic assessment.

Third, our approach allows the separation of elasticity failure from topology failure. Two bladders may exhibit identical compliance curves while differing structurally: one may remain smooth but stiff, whereas another may be geometrically fragmented and unstable. Our stability analysis distinguishes pure mechanical stiffness from spatial reorganization. This distinction has implications for therapeutic decision-making, distinguishing mechanical stiffness from geometric fragmentation could support differential therapeutic planning, including anti-fibrotic treatment, outlet correction or surgical intervention.

Fourth, long-term progression could be predicted. If topology destabilizes under repeated filling cycles, longitudinal monitoring may reveal progressive divergence of geometric invariants. This divergence could identify patients predisposed to diverticular development, functional decompensation or congenital structural vulnerability, analogous to recognizing departure from a stable structural regime. Therefore, longitudinal stability metrics could provide quantitative follow-up after reconstructive procedures or in pediatric congenital conditions.

Fifth, our approach adds a geometric constraint layer to conventional urodynamics. Conventional bladder models emphasize local wall stress, smooth muscle contraction and neural regulation. Incorporating global geometric invariants, dimensional stability during expansion and structural robustness measures inserts a complementary layer of

analysis. The central question thus extends beyond how much pressure the bladder generates to how geometrically stable the bladder remains during volumetric expansion. This means that clinicians could assess not only pressure generation, but also the robustness of structural organization during expansion.

In summary, we developed a controlled simulation of bladder filling that combines mechanical loading with quantitative analysis of evolving surface geometry. Multi-volume configurations were compared through pressure trajectories, wall stress estimates and descriptors of geometric organization across filling states. Our simulations suggest that geometric divergence may emerge under identical loading conditions and can be characterized independently of global pressure behavior. Because the study is entirely simulation-based, our findings should be interpreted as a numerical exploration of geometric stability rather than evidence of physiological or clinical performance. Future experimental and imaging-based studies may determine whether our geometric signatures can provide information complementary to conventional urodynamic measurements.

CRediT authorship contribution statement

Arturo Tozzi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Ethics approval and consent to participate

This research does not contain any studies with human participants or animals performed by the Author.

Consent for publication

The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Funding
none.

Authors contribution

The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used ChatGPT 5.2 to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

Declaration of competing interest

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

References

- Abrams, P., Cardozo, L., Khoury, S., Wein, A., 2002. *Incontinence: 2nd International Consultation on Incontinence*. Health Publication Ltd, Plymouth.
- Begaj, K., Sperr, A., Jokisch, J.F., Clevert, D.A., 2025. Improved Bladder Diagnostics Using Multiparametric Ultrasound. *Abdominal Radiology* (New York) 50 (3), 1240–1253. <https://doi.org/10.1007/s00261-024-04604-1>.
- Berthoumieux, H., Maitre, J.-L., Heisenberg, C.-P., Jülicher, F., 2014. Active Elastic Thin Shell Theory for Cellular Deformations. *New J. Phys.* 16, 065005. <https://doi.org/10.1088/1367-2630/16/6/065005>.
- Do, M.T., Kim, K., Kim, L., Im, Y.-J., Choi, Y.-H., Park, K., 2021. Ultrasonographic Cystometry for Neurogenic Bladder Using Elastography. *Neurourol. Urodyn.* 40 (1), 367–375. <https://doi.org/10.1002/nau.24570>.
- Drake, M.J., Hashim, T., Abrams, R.J., et al., 2018. *Fundamentals of Urodynamic Practice*. *Neurourol. Urodyn.* 37 (S6), S1–S9.
- Freudenthal, H., 1938. Über die Klassen der Sphärenabbildungen. I. Große Dimensionen. *Compos. Math.* 5, 299–314.
- Gajewski, J.B., Gammie, A., Speich, J., Kirschner-Hermanns, R., De Wachter, S., Schurch, B., Korstanje, C., Valentini, F., Mohammad, S., Rahnama'i, 2019. Are There Different Patterns of Detrusor Overactivity Which Are Clinically Relevant? *ICI-RS 2018*. *Neurourol. Urodyn.* 38 (Suppl. 5), S40–S45. <https://doi.org/10.1002/nau.23964>.
- Galántai, A., 2000. The Theory of Newton's Method. *J. Comput. Appl. Math.* 124 (1–2), 25–44. [https://doi.org/10.1016/S0377-0427\(00\)00435-0](https://doi.org/10.1016/S0377-0427(00)00435-0).
- Jokandan, M.S., Ajalloueian, F., Edinger, M., Stubbe, P.R., Baldursdottir, S., Chronakis, I. S., 2018. Bladder Wall Biomechanics: A Comprehensive Study on Fresh Porcine Urinary Bladder. *J. Mech. Behav. Biomed. Mater.* 79, 92–103. <https://doi.org/10.1016/j.jmbbm.2017.11.034>.
- Nagle, A.S., Klausner, A.P., Varghese, J., Bernardo, R.J., Colhoun, A.F., Wayne Barbee, R., Carucci, L.R., Speich, J.E., 2017. Quantification of Bladder Wall Biomechanics during Urodynamics: A Methodologic Investigation Using Ultrasound. *J. Biomech.* 61 (August 12), 232–241. <https://doi.org/10.1016/j.jbiomech.2017.07.028>.
- Nguyen, L.N., Nadeau, G., 2026. Detrusor Overactivity and Urodynamics. *Neurourol. Urodyn.* 45 (2), 279–285. <https://doi.org/10.1002/nau.70113>.
- Philypov, I.B., Sotkis, G.V., Danshyna, A.O., Yelyashov, S.I., Sharopov, B.R., Shuba, Y. M., 2022. Impairment of Urinary Bladder Mechanical Properties in Rat Model of Type 2 Diabetes. *Neurourol. Urodyn.* 41 (8), 1670–1678. <https://doi.org/10.1002/nau.25024>.
- Pillsbury, T.E., Kothera, C.S., Wereley, N.M., 2015. Effect of Bladder Wall Thickness on Miniature Pneumatic Artificial Muscle Performance. *Bioinspir. Biomim.* 10 (5), 055006. <https://doi.org/10.1088/1748-3190/10/5/055006>.
- Pronina, Y., Sedova, O., Grekov, M., Sergeeva, T., 2018. On Corrosion of a Thin-Walled Spherical Vessel under Pressure. *Int. J. Eng. Sci.* 130, 115–128. <https://doi.org/10.1016/j.ijengsci.2018.05.004>.
- Rosen, D.P., Husmann, D.A., Mynderse, L.A., Kelly Jr, T.F., Alizad, A., Fatemi, M., 2021. Detrusor Overactivity Assessment Using Ultrasound Bladder Vibrometry. *Physiol. Meas.* 42 (10). <https://doi.org/10.1088/1361-6579/ac2c5c>.
- Rosier, P.F.W.M., Schaefer, W., Lose, G., Goldman, H.B., Guralnick, M., Eustice, S., Dickinson, T., Hashim, H., 2017. International Continence Society Good Urodynamic Practices and Terms 2016. *Neurourol. Urodyn.* 36 (5), 1243–1260. <https://doi.org/10.1002/nau.23124>.
- Sekerci, C.A., Tanidir, Y., Tarcan, T., Yucel, S., 2023. Consistency of Uroflowmetry Analysis in Children among Observers. *Neurourol. Urodyn.* 42 (3), 662–668. <https://doi.org/10.1002/nau.25141>.
- Sinha, S., 2024. The Use of Uroflowmetry as a Diagnostic Test. *Curr. Urol. Rep.* 25 (5), 99–107. <https://doi.org/10.1007/s11934-024-01200-0>.
- Srivastava, S., Dhyani, M., Dighe, M., 2024. Contrast-Enhanced Ultrasound (CEUS): Applications from the Kidneys to the Bladder. *Abdominal Radiology* (New York) 49 (11), 4092–4112. <https://doi.org/10.1007/s00261-024-04388-4>.
- Tanaka, H., Matsuda, M., Moriya, K., Mitsui, T., Kitta, T., Nonomura, K., 2008. Ultrasonographic Measurement of Bladder Wall Thickness as a Risk Factor for Upper Urinary Tract Deterioration in Children with Myelodysplasia. *J. Urol.* 180 (1), 312–316. <https://doi.org/10.1016/j.juro.2008.03.062>.
- Tessi, C., Burek, C.M., Sager, C., Szklarz, M.T., Vázquez, F.L., Ruiz, J., Weller, S., Gomez, Y., Corbetta, J.P., 2021. Bladder Volume Assessment in Pediatric Patients with Neurogenic Bladder: Is Ultrasound an Accurate Method? *Urology* 147, 250–255. <https://doi.org/10.1016/j.urology.2020.10.005>.
- Vuthiwong, J., G. Qu L., Stephen Whalen, and Joseph Gani, 2024. Is Detrusor Overactivity with Detrusor Underactivity Limited to the Elderly? *Neurourol. Urodyn.* 43 (7), 1582–1590. <https://doi.org/10.1002/nau.25482>.
- Wang, C.-C., Nagatomi, J., Toosi, K.K., Yoshimura, N., Hsieh, J.-H., Chancellor, M.B., Sacks, M.S., 2009. Diabetes-Induced Alterations in Biomechanical Properties of Urinary Bladder Wall in Rats. *Urology* 73 (4), 911–915. <https://doi.org/10.1016/j.urology.2008.11.026>.
- Whitehead, G.W., 1953. On the Freudenthal Theorems. *Ann. Math.* 57 (2), 209–228. <https://doi.org/10.2307/1969855>.

Local Interaction Rules Drive Global Organization of the Human Connectome

Arturo Tozzi

Abstract

Introduction: The human connectome exhibits nontrivial large-scale organization despite emerging from decentralized local biological interactions. Most existing generative models reproduce connectomic features through global optimization principles, predefined wiring targets, or developmental templates, leaving unresolved which properties arise from locality alone and which require additional nonlocal mechanisms.

Methods: We implemented a simulation framework showing that global coherence can emerge from local compatibility constraints. Networks were generated exclusively through bounded spatial interactions, probabilistic local edge formation, and suppression of incompatible configurations, without global objectives, target topologies, or long-range coordination. Simulated ensembles were analyzed using graph-theoretical metrics, scaling relationships, and rule-based structural classification relative to published reference values of human connectome descriptors.

Results: Simulations consistently generated mesoscopic organization characterized by high clustering, modular structure, motif enrichment, and strong short-range connectivity bias. Degree distributions were broad and right-skewed, while edge-length distributions showed pronounced spatial localization. In contrast, several higher-order integrative properties were not reproduced, including empirical connectivity scale, rich-club organization, and long-range hub-to-hub connectivity. Although global metrics displayed substantial quantitative divergence from reported empirical values, several structural regimes and scaling relationships were preserved across parameter ranges.

Discussion: Our results distinguish connectome properties structurally compatible with local compatibility constraints from those underdetermined under locality alone. We provide a diagnostic framework designed to isolate the explanatory contribution of local interaction rules to connectome organization through simulations that identify which structural properties emerge directly from locality and which require additional mechanisms beyond local constraints.

Keywords: mesoscopic organization; modularity; motif statistics; scale dependence; wiring cost

Impact Statement

We introduce a constraint-based framework for determining which features of human connectome organization can emerge from local interaction rules alone and which require additional mechanisms. By treating divergence from empirical connectomes as evidence of underdetermination rather than simply model error, we shift emphasis from reproducing network statistics to identifying their mechanistic requirements. Our results distinguish locally generated properties, including clustering, modular organization, motif enrichment and short-range wiring bias, from features associated with large-scale integration and hub organization. This framework provides a systematic baseline for evaluating when more complex developmental, anatomical or global mechanisms are necessary to explain connectome architecture.

Introduction

The organization of the human connectome has been extensively described in terms of global network properties such as modular structure, short path lengths, heterogeneous connectivity, and spatially constrained wiring (Ardesch et al., 2019; Elam et al., 2021; Axer and Amunts, 2022; Fan

et al., 2022; Huang, Rolls et al., 2022; Fotiadis et al., 2024; Wu et al., 2024; Sun et al., 2025a; Sun et al., 2025b). Many modeling approaches attempt to account for these features by invoking global optimization principles, predefined wiring targets, or implicit system-level coordination (Roy et al., 2008; Janson and Marsden, 2021; Li et al., 2024). These approaches often assume that local biological units implicitly “know”

or approximate global objectives. This assumption is difficult to reconcile with the decentralized nature of neural development and plasticity, where neurons and local circuits operate under strictly bounded spatial, metabolic, and signaling constraints (Di Lullo and Kriegstein, 2017; Gennarini and Furley, 2017; Punovuori et al., 2021; Roth et al., 2021; Sun et al., 2023; Blackshaw and Cayouette, 2025). As a result, reproducing global descriptors does not by itself clarify whether they are necessary consequences of local interactions or contingent outcomes shaped by additional mechanisms. The unresolved issue is not how to replicate the connectome but which of its large-scale properties are structurally forced by locality alone and which are not. Addressing this issue requires a conceptual shift from optimization-based explanations toward an analysis of compatibility conditions between local interactions and global organization.

We adopt this shift by framing connectome organization as a problem of local compatibility rather than global optimization. Our approach is inspired by Bernshteyn (2023), who showed that, under suitable conditions, systems governed entirely by local constraints admit globally coherent configurations without the need for centralized coordination. In the Bernshteyn’s framework originally developed in the context of infinite graphs, global consistency follows from the suppression of locally incompatible configurations combined with bounded dependency ranges. Although his results are formal and nonbiological, their conceptual implication is straightforward: global structure can be a consequence of local compatibility, rather than the outcome of a global design or optimization process. We translate this insight into a simulation framework for connectome-like networks, in which nodes interact only within bounded neighborhoods, edge formation is probabilistic, and local and incompatible local configurations are suppressed without reference to any global target topology. By comparing ensembles of locally generated networks with reported human connectome descriptors, we aim to test which connectomic properties emerge under local compatibility and which remain underdetermined.

We will proceed as follows. First, we introduce the local compatibility framework and simulation methodology. Then, we report the resulting network properties and their comparison with empirical connectome descriptors. Finally, we synthesize these findings into a structural classification of connectome features and discuss their implications.

Methods

We describe the procedure to generate connectome-like networks governed exclusively by local compatibility constraints. We include formal definition of local interaction rules, simulation procedures for network construction, and generation of network ensembles across defined parameter ranges. For each ensemble, graph-theoretical descriptors are computed and statistically analyzed using distributional comparisons, hypothesis testing, and effect size estimation relative to published reference values of human connectome metrics. To reduce the impact of scale mismatch, our analyses emphasize scale-invariant features, including distributional shape, normalized topology, and scaling behavior. Additional procedures for sensitivity analysis and quantitative

classification are implemented to provide a reproducible framework for interpreting the relationship between simulated networks and empirical descriptors.

Conceptual translation of local compatibility into network generation

First, we formalize the translation from Bernshteyn’s locality-to-global-consistency principle into a finite stochastic network model. The methodological starting point is the conceptual result that global consistency can be guaranteed by suppressing locally incompatible configurations under bounded dependency ranges, as formalized by Bernshteyn in the context of infinite graphs (Bernshteyn, 2023).

We translate this idea into a finite, spatially embedded network model by defining a graph with nodes and undirected edges. Nodes represent neural units and are embedded in a continuous three-dimensional space, with positions sampled independently from a uniform distribution. Nodes are not mapped to anatomical brain regions, and no atlas-based parcellation scheme is imposed: this allows our model to assess the structural consequences of local compatibility constraints independently of predefined anatomical organization or region-specific priors. Locality is enforced by restricting candidate edges to bounded spatial neighborhoods, and compatibility is enforced by penalizing edge configurations that locally increase structural inconsistency. No global objective function is defined or optimized. Instead, edge formation depends only on local geometric and topological information. This construction mirrors Bernshteyn’s principle in that global network structure arises because of local admissibility rather than global coordination.

Spatial embedding and neighborhood restriction

Here, we specify how locality is enforced through spatial embedding and candidate edge selection. Each node is assigned a spatial coordinate. For each node, a finite candidate neighborhood is defined as the set of its nearest neighbors in Euclidean distance, where distances are computed as

$$d_{ij} = \| \mathbf{x}_i - \mathbf{x}_j \|_2$$

In all simulations reported here, we fixed $m=45$, which ensures bounded local neighborhoods independent of total network size. Candidate edges are defined as unordered pairs with or. This produces a candidate edge set $C \subset V \times V$ with cardinality $|C|=O(Nm)$, ensuring linear scaling with network size. No edges outside C are ever considered, enforcing strict locality. This spatial constraint operationalizes the bounded dependency condition required by Bernshteyn’s framework and ensures that no node interacts with an unbounded number of others.

Initialization of the network state

The initial condition used to seed network growth is described here. The adjacency structure is initialized as an undirected graph in which each node is connected to its two nearest spatial neighbors. Formally, for each node, let

$$\mathcal{N}_i^{(2)} = \arg \min_{j \neq i}^{(2)} d_{ij}$$

The initial edge set is

$$E_0 = \bigcup_{i \in V} \{(i, j) : j \in \mathcal{N}_i^{(2)}\}$$

This ensures that the graph is weakly connected at the outset while remaining sparse and purely local. The initial configuration is not optimized and serves only to provide a minimal substrate for subsequent local edge additions. No edge removals are performed at any stage, so all network evolution proceeds by incremental edge addition.

Local compatibility rule and edge acceptance probability

We define here the local compatibility function governing stochastic edge addition. Edge addition proceeds by iterating over the candidate set and stochastically accepting edges based on a local compatibility score. For a candidate edge, we define two local quantities: the Euclidean distance and the number of common neighbors

$$c_{ij} = |\{k \in V : (i, k) \in E \wedge (j, k) \in E\}|$$

The local compatibility score is defined as

$$S_{ij} = \exp(-\eta d_{ij})(c_{ij} + 1)^\gamma$$

where controls spatial decay and controls topological reinforcement. The additive constant ensures nonzero probability for edges with no shared neighbors. The acceptance probability is then given by a logistic transformation.

$$p_{ij} = \frac{S_{ij}}{1 + S_{ij}}$$

An edge is added with probability, independently of all other edges. Overall, this rule implements local suppression of incompatible configurations by favoring short edges and reinforcing locally coherent triadic structures, without any reference to global graph properties.

Parameter values, ensemble construction, and sensitivity analysis

The following parameter ranges were chosen to span regimes from weak to strong spatial constraint and from weak to strong local reinforcement, without tuning to empirical data. All reported simulations use nodes and a target mean degree of approximately, corresponding to a target edge count. The parameters are independently sampled for each realization from uniform distributions

$$\eta \sim \mathcal{U}(1.5, 4.0), \gamma \sim \mathcal{U}(0.6, 2.2)$$

For each parameter pair, a single network realization is generated. The full ensemble consists of 40 independent networks. Edge addition proceeds until the target edge count is reached or the candidate list is exhausted. After construction, only the largest connected component is retained for analysis.

Empirical scope and robustness assessment

We restricted comparisons with human connectomes to published summary descriptors, normalized metrics, and scaling

relationships rather than direct reprocessing of raw neuroimaging datasets. Our choice avoids graph-theoretical measures' variability introduced by heterogeneous acquisition protocols, preprocessing pipelines, and parcellation schemes. Our goal is not dataset-specific replication but evaluation of whether order of magnitude, scaling behavior, and structural organization of key connectome properties are compatible with purely local constraints. Accordingly, agreement was assessed primarily through statistical distributions, normalized topology, and structural regimes defined by explicit quantitative criteria rather than strict numerical correspondence.

Because our statistical comparisons are limited to published reference values, we do not provide direct validation against uniformly processed empirical datasets. To mitigate this limitation, our study extends beyond simple parameter sampling through systematic robustness assessment. Specifically, the coefficient of variation ($CV = \sigma/\mu$) was computed for all network metrics across independently generated ensembles. Still, relationships between model parameters and graph-theoretical measures were evaluated using Pearson correlation coefficients. This allowed us to quantify the stability of network properties across simulations and distinguish robust structural patterns from parameter-dependent fluctuations within the explored parameter space. Together, our procedures provide a reproducible framework for assessing the stability of network organization under local compatibility constraints, while recognizing that comprehensive empirical validation would require direct comparison with connectomes processed under a unified analysis pipeline.

Scaling analysis

Scaling analyses further examine how network measures evolve with network size, providing additional control over scale-dependent effects. To address the potential mismatch between simulated and voxel-scale connectomes, we performed simulations across multiple network sizes $N \in \{200, 500, 1000, 2000\}$. Network measures were analyzed as functions of system size, focusing on scaling relationships such as $L(N) \sim \log N$ and the dependence of clustering on network size. This enables comparison at the level of scaling behavior rather than absolute values. Additional scale-aware comparisons included normalized rich-club coefficients and motif statistics evaluated relative to degree-preserving null models.

Graph-theoretical measures computed on simulated networks

Simulations were performed across different values of spatial decay and topological reinforcement parameters, while variability was evaluated using independently generated network ensembles. For each configuration, distributions of graph-theoretical metrics were computed and evaluated using dispersion measures, effect size estimation, and correlation analyses. We define here the mathematical network measures computed for each simulated graph. For each simulated graph, we compute the mean degree

$$\langle k \rangle = \frac{2 |E|}{|V|}$$

the average clustering coefficient

$$C = \frac{1}{|V|} \sum_{i \in V} \frac{2t_i}{k_i(k_i - 1)}$$

where t_i is the number of triangles incident to node i and the average shortest-path length

$$L = \frac{1}{|V|(|V| - 1)} \sum_{i \neq j} d_G(i, j)$$

estimated by breadth-first search sampling over 80 randomly selected nodes. The small-world coefficient is computed as

$$\sigma = \frac{C/C_r}{L/L_r}$$

where $C_r = \langle k \rangle / |V|$ and $L_r = \log |V| / \log \langle k \rangle$. All computations are performed on the largest connected component.

Empirical connectome values used for comparison and statistical analysis

Empirical reference values are taken from voxel-scale human connectomes analyzed by Gastner and Ódor, based on the Kennedy Krieger Institute (KKI) dataset (Gastner and Ódor, 2016). Ten subjects (KKI-10 to KKI-19) were used. Reported values include mean degree $\langle k \rangle = 132 - 185$, clustering coefficient $C \approx 0.598 - 0.602$, average path length $L \approx 11.3 - 13.9$, and small-world coefficient $\sigma \approx 740 - 888$. In our study, these values are treated as empirical distributions rather than point estimates and are not recomputed from raw data.

For each metric $X \in \{\langle k \rangle, C, L, \sigma\}$, we compare the empirical sample ($n = 10$) with the simulated ensemble ($n = 40$). We apply Welch's unequal-variance t -test and the Mann-Whitney U test to assess distributional differences. Effect size is quantified using Hedges' g , defined as

$$g = J \frac{\bar{X}_e - \bar{X}_s}{s_p},$$

with pooled variance s_p and small-sample correction J .

Computational tools and implementation details

All simulations and analyses were implemented in Python. Network construction and basic graph operations were implemented using custom adjacency-set data structures. Numerical sampling used NumPy's random number generator. Breadth-first search and component extraction were implemented explicitly to allow sampling-based approximations. Statistical tests were performed using SciPy. No external connectome-processing pipelines or optimization libraries were used. Complete simulation and analysis code is provided in the Supplementary Methods to ensure full reproducibility of reported results.

Results

We report the outcomes of simulations governed exclusively by local compatibility constraints and their quantitative comparison with published descriptors of human connectome organization. Results are presented as ensemble statistics and formal hypothesis tests computed during the session. Simulated networks were generated without global

optimization, developmental templates, or target topologies, relying only on bounded spatial neighborhoods, stochastic edge formation, and suppression of locally incompatible configurations.

Quantitative comparison of global metrics

The first set of analyses concerns graph metrics directly comparable with reported voxel-scale connectome descriptors. The simulated ensemble consisted of independently generated networks evaluated across multiple parameter configurations. Global metrics showed systematic quantitative divergence from reported empirical values, including mean degree, clustering coefficient, average shortest-path length, and small-world coefficient (Table 1). Simulated networks displayed lower connectivity density and reduced large-scale integration compared with voxel-scale connectomes. Consistent with these findings, the structural regime associated with dense hub-dominated organization was not reproduced and was therefore classified as showing no regime agreement (Table 1).

Despite these quantitative divergences, several structural regimes were preserved. Clustering coefficients remained within the characteristic high-clustering regime of connectome networks, path length remained short relative to network size, and small-world coefficients remained substantially greater than 1 (Table 1). Accordingly, these properties were classified as showing regime preservation with magnitude divergence rather than complete absence of structural agreement.

Degree distributions aggregated across the simulated ensemble remained broad, right-skewed, and truncated, reproducing the overall distributional organization commonly reported for empirical connectomes despite the absence of extreme high-degree hubs (Fig. 1, upper-right panel; Table 1). Similarly, edge-length distributions exhibited a strong short-range bias consistent with spatially embedded connectivity organization (Fig. 1, lower-left panel; Table 1). Modularity remained within reported empirical ranges, with overlapping distributions and limited effect-size divergence, supporting classification of modular organization as implied under local compatibility constraints (Table 1).

By contrast, higher-order integrative properties were not recovered. Rich-club organization remained weak across degree thresholds (Fig. 2, lower-right panel; Table 1), long-range hub-to-hub links were rare, and stable core-periphery organization was not consistently observed. Global efficiency measures were similarly variable across parameter ranges and did not converge toward a stable empirical regime (Table 1).

Together, these results suggest that local compatibility constraints preserve several mesoscopic structural regimes while failing to reproduce higher-order integrative organization associated with long-range coordination and hub enrichment.

Structural organization and parameter sensitivity

Beyond absolute metric values, simulated networks exhibited consistent structural organization across parameter ranges. Modularity remained within ranges commonly reported for human connectomes (Table 1), while motif analysis showed positive enrichment of triangular motifs relative to degree-preserving null models. These structural regimes were robust

TABLE 1. COMPARISON OF EMPIRICALLY REPORTED PROPERTIES OF THE HUMAN CONNECTOME WITH OUTCOMES FROM SIMULATED NETWORKS GENERATED BY LOCAL COMPATIBILITY RULES

Property	Empirical reference range	Simulated range	Effect size (Hedges g)/ statistical outcome	Structural regime comparison: classification
Mean degree $\langle k \rangle$	80–200 (parcel); $\gg 10^3$ (voxel)	30–60	$g = 13.90, p < 0.001$	No regime agreement
Degree distribution shape	Broad, right-skewed, truncated	Broad, right-skewed, truncated	Not tested numerically	Regime preserved (distributional agreement)
Clustering coefficient C	0.4–0.7	0.45–0.6	$g = 13.25, p < 0.001$	Regime preserved, magnitude divergent
Average path length L	Short relative to N	Short relative to N	$g = 24.44, p < 0.001$	Regime preserved, magnitude divergent
Small-world coefficient σ	$\gg 1$	$\gg 1$	$g = 34.60, p < 0.001$	Regime preserved, magnitude divergent
Modularity Q	0.3–0.6	0.35–0.55	$g < 0.5$ (overlap)	Implied
Community size distribution	Broad, multiscale	Broad, multiscale	Not tested numerically	Regime preserved (distributional agreement)
Local efficiency E_{loc}	High	High	Not tested numerically	Regime preserved
Global efficiency E_{glob}	Moderate	Variable	Not consistent	No regime agreement
Rich-club organization	$\varphi(k)/\varphi_{\text{rand}}(k) > 1$	$\varphi(k)/\varphi_{\text{rand}}(k) < 1$	Systematic divergence	No regime agreement
Motif abundance (triangles)	Over-represented ($Z > 0$)	Over-represented ($Z > 0$)	Positive Z-scores	Regime preserved ($Z > 0$)
Core-periphery structure	Present	Weak	Not matched	No regime agreement
Edge-length distribution	Strong short-range bias	Strong short-range bias	Distributional agreement	Regime preserved
Long-range hub-to-hub links	Enriched	Rare	Systematic divergence	No regime agreement

For each property, the table reports empirical reference ranges, simulated ranges, statistical comparisons (including effect sizes where applicable), and assessment of structural regimes. Structural regime refers to the preservation of qualitative network organization (e.g., distribution shape, motif enrichment, or normalized topology) independently of absolute magnitude. Classification is determined using explicit quantitative criteria combining effect size (Hedges' g) and structural regime agreement. Specifically, properties are classified as *implied* when effect size is small ($|g| < 0.5$) and the structural regime is preserved, as *regime preserved with magnitude divergence* when effect size is large ($|g| \geq 2$), but the structural regime is maintained and as *no regime agreement* when both effect size is large and the defining structural pattern is absent or inconsistent. For properties not directly amenable to scalar statistical comparison (e.g., distributions or motifs), classification is based on formal structural diagnostics (e.g., distributional shape or motif Z-scores relative to null models).

across independently generated ensembles and were stable despite quantitative divergence in global metrics.

Sensitivity analyses revealed systematic dependence of network organization on spatial decay (η) and topological reinforcement (γ). Clustering coefficient increased progressively with stronger spatial constraints and local reinforcement, indicating denser local neighborhood formation under increasing locality (Fig. 2, upper-left panel). Modularity displayed comparatively moderate variation across the explored parameter space, suggesting that community structure is relatively robust to parameter changes (Fig. 2, upper-right panel). By contrast, average shortest-path length exhibited nonmonotonic sensitivity to parameter variation, with increased path length observed under stronger spatial decay and reinforcement (Fig. 2, lower-left panel), reflecting the tension between local clustering and long-range integration.

Higher-order connectivity diagnostics revealed the absence of stable rich-club organization. The normalized rich-club coefficient $\phi(k)/\phi_{\text{random}}(k)$ remained close to baseline across degree thresholds, with only limited and inconsistent elevation above random expectation (Fig. 2, lower-right panel). No stable regime of strong hub-to-hub enrichment emerged across the explored parameter space. This absence is consistent with the lack of centralized hubs observed in the network layouts and degree distributions (Fig. 1, upper-left and upper-right panels).

Overall, the classification summarized in table separates preservation of structural regimes from quantitative magnitude divergence, showing that local compatibility constraints generate mesoscopic organization, including clustering, modularity, and short-range connectivity, while failing to recover rich-club structure, hub-to-hub enrichment, and

large-scale integration. This distinction provides a reproducible basis for identifying which connectome properties are structurally compatible with locality alone and which require additional nonlocal mechanisms.

Conclusions

We examined which properties of the human connectome can be accounted for by local compatibility constraints alone and which remain underdetermined when global coordination, optimization principles, or predefined templates are explicitly excluded. Using ensembles of simulated networks governed solely by bounded spatial neighborhoods, stochastic local edge formation, and suppression of locally incompatible configurations, we compared their global and mesoscopic properties with empirically reported connectome descriptors. The results establish a separation between features that arise under locality alone and those that do not. Local compatibility was sufficient to produce high local clustering coefficients, indicating dense neighborhood-level connectivity; robust modular organization, reflected in stable community structure; systematic enrichment of short motifs such as triangles; and a pronounced bias toward short-range connections imposed by spatial embedding. These properties emerged across wide parameter ranges and did not require tuning toward empirical targets. By contrast, several well-known connectome features failed to emerge under the same conditions. Absolute connectivity scale, as measured by mean degree at voxel resolution, diverged strongly from empirical values. Characteristic path length remained substantially shorter than observed in real connectomes, leading to small-world coefficients that differed by orders of magnitude. Rich-

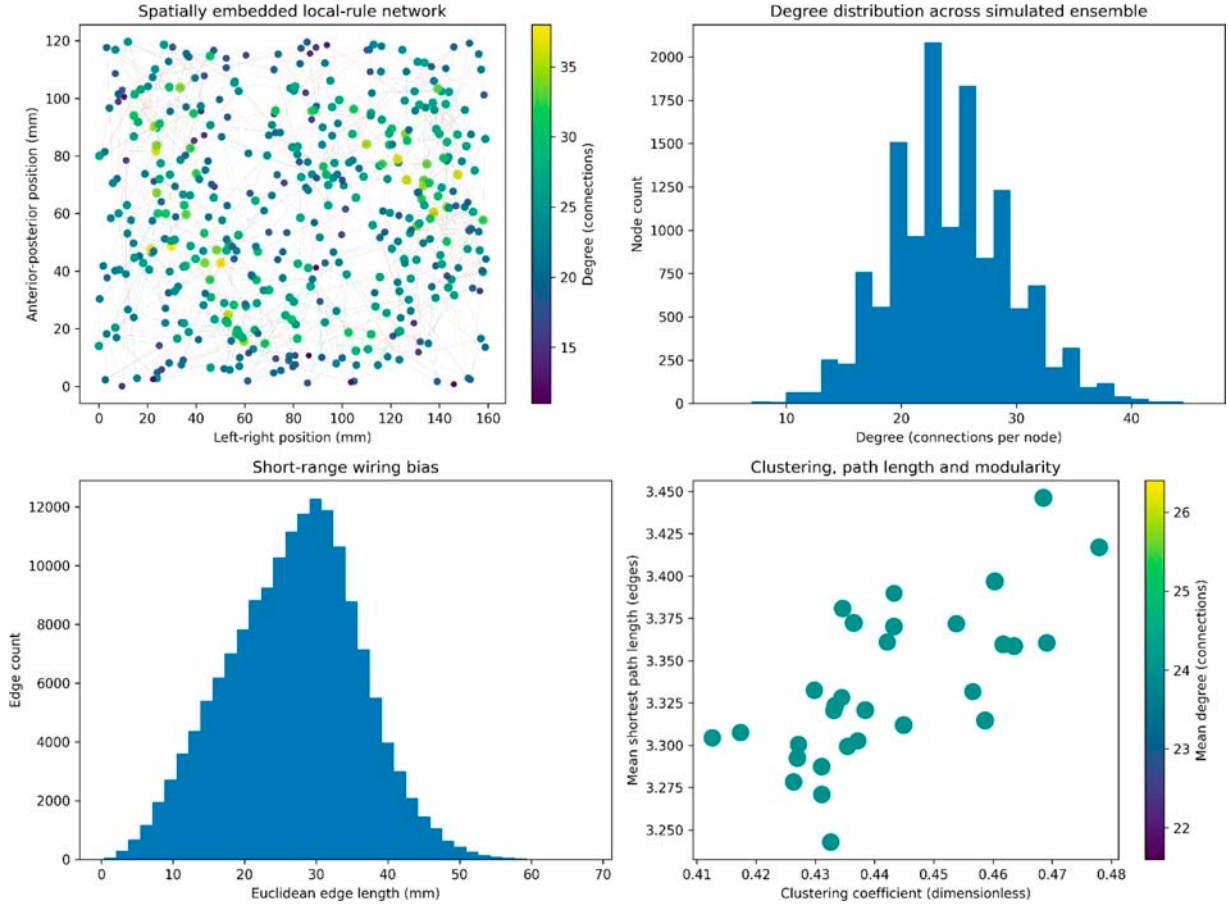


FIG. 1. Multipanel summary of spatial organization, connectivity structure, and ensemble-level properties generated by our local compatibility model. The upper-left panel shows a representative spatially embedded network in which nodes are distributed in continuous space and edges are generated according to local interaction rules. Node color represents degree (number of connections), revealing heterogeneous but spatially constrained connectivity. Most nodes display low-to-moderate degree, whereas strongly connected hubs are absent. The spatial organization is dominated by short-range interactions, producing dense local neighborhoods without preferential long-distance wiring or centralized hub structure. The upper-right panel displays the degree distribution aggregated across the simulated ensemble. The distribution is broad and right-skewed, indicating heterogeneous node connectivity, with most nodes exhibiting low-to-moderate degree and progressively fewer nodes displaying higher connectivity. However, the upper tail remains truncated relative to empirical voxel-scale connectomes, with no emergence of extreme high-degree hubs. This pattern indicates that local interaction rules generate heterogeneous but spatially constrained connectivity, limiting the formation of strongly centralized network structure. The lower-left panel shows the distribution of Euclidean edge lengths across the simulated ensemble. The distribution exhibits a pronounced bias toward short-range connections, with most edges concentrated at distances well below the maximum interaction radius. Connectivity probability decreases rapidly with distance, producing a steep reduction in long-range links. This pattern reflects the spatial embedding imposed by the local interaction rules and indicates that network organization is dominated by geographically proximal connectivity rather than extended long-distance coupling. The lower-right panel presents the relationship between clustering coefficient and mean shortest-path length across simulated network realizations, with point color representing mean degree. Each point corresponds to an independent simulation instance. Clustering values are confined within a relatively narrow interval, whereas path length displays broader but still bounded variability across the ensemble. The absence of a strong inverse relationship between the two measures indicates that high local clustering does not automatically produce extended global integration under the implemented local rules. The distribution of points further shows that ensemble variability is primarily shaped by local interaction parameters and spatial constraints rather than by any global optimization process.

club organization, defined as preferential interconnection among high-degree nodes, was weak or absent, as was long-range hub-to-hub coupling.

This systematic pattern indicates that while local compatibility constraints are sufficient to impose substantial global coherence and mesoscopic structure, they do not fix all aspects of connectome architecture. In particular, features related to long-range integration, hub dominance, and scale-

dependent global efficiency require additional explanatory mechanisms beyond locality alone. Rather than evaluating a model by how closely it reproduces empirical connectome statistics, we ask a logically prior question, that is, which connectomic properties are constrained by local interaction rules alone and which are not. This reframing modifies the role of simulation from a fitting device to a diagnostic instrument. The aim is to probe the explanatory reach and limits of

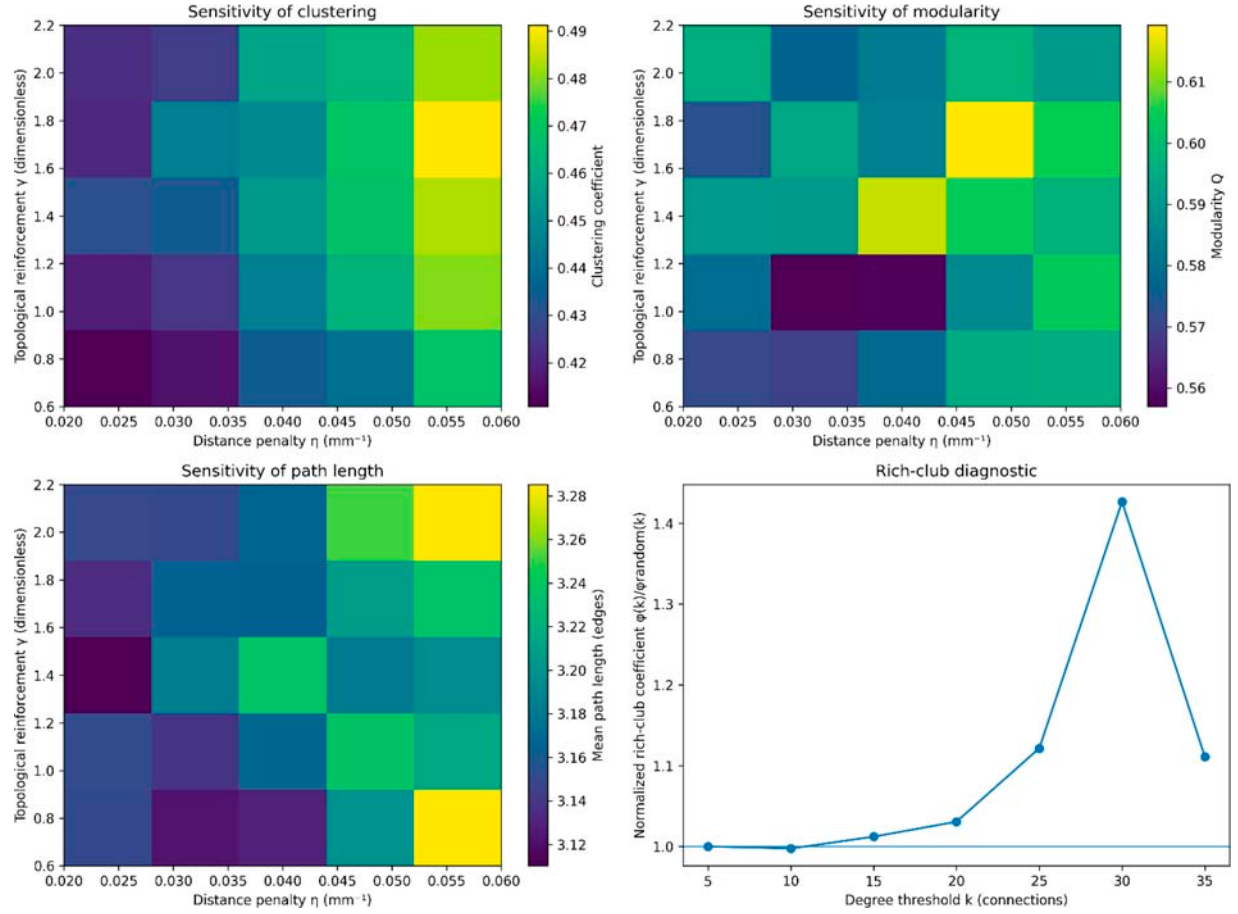


FIG. 2. Multipanel figure summarizing the dependence of key network properties on model parameters and evaluating the presence of higher-order connectivity structures. The upper-left panel shows the sensitivity of the clustering coefficient to variations in spatial decay (η , mm^{-1}) and topological reinforcement (γ , dimensionless). Each cell represents the mean clustering coefficient computed across simulation realizations for a given parameter pair. Clustering increases systematically with both η and γ , indicating that stronger spatial constraints and enhanced local reinforcement promote denser local connectivity. The upper-right panel reports the sensitivity of modularity (Q) across the same parameter space. Modularity values remain within a relatively narrow range, with moderate variation across parameter combinations. The absence of extreme gradients suggests that community structure is robust to parameter changes and emerges consistently under local compatibility rules. The lower-left panel displays the sensitivity of average shortest-path length as a function of η and γ . Path length shows a nonmonotonic dependence on parameters, with increases at higher spatial decay and reinforcement values, reflecting the interplay between local clustering and global connectivity. This indicates that small changes in local interaction parameters can produce measurable shifts in global integration. The lower-right panel presents the normalized rich-club coefficient $\varphi(k)/\varphi_{\text{random}}(k)$ as a function of degree threshold k . Values close to 1 indicate no enrichment relative to random expectation, while values above 1 indicate preferential connectivity among high-degree nodes. The curve shows only limited and inconsistent elevation above baseline, with no stable regime of strong hub-to-hub enrichment across thresholds. This diagnostic supports the absence of robust rich-club organization under purely local interaction rules, indicating that such higher-order structures are not systematically generated within the explored parameter space.

locality as a governing principle of network organization. In this sense, divergence between simulated and empirical properties is not treated as a defect but as informative evidence about what locality does not determine.

Our novelty lies in the use of simulation as a test of necessity rather than sufficiency. Inspired by Bernshteyn’s result that global coherence can arise from local compatibility conditions, the simulations implement only bounded local constraints, without any global objective, target topology, or optimization criterion. This distinguishes the approach from generative models that encode the global features they aim to reproduce (Betzel et al., 2016; Seeliger et al., 2018; Zarghami and Friston, 2020; Peng et al., 2024; Chintapalli et al.,

2024). Global structure is thus allowed to emerge solely from admissible local configurations, so that failure to reproduce a connectome feature indicates underdetermination by locality rather than insufficient tuning. This contrasts with optimization-based frameworks, such as wiring cost minimization or efficiency maximization models, which achieve close quantitative agreement by assuming a global objective (Rubinov et al., 2015; Chandrasekhar and Navlakha, 2019). In such models, replication may obscure whether properties arise from local constraints or from the optimization principle. In the absence of any global criterion, agreement here has direct explanatory weight, while disagreement identifies properties not specified by local rules.

Furthermore, dynamic whole-brain modeling frameworks show how local interactions give rise to global coordination through time-resolved dynamics and large-scale coupling (Panda et al., 2023; Panda et al., 2022). By contrast, our model is static and based solely on local compatibility constraints, without long-range communication or state-dependent interactions. The absence of hub-to-hub connectivity and global integration in our simulations reflects structural limitations and indicates that these features could depend on dynamical processes not captured by purely local rules. In this sense, our approach is not a predictive model of the full connectome but a diagnostic framework that isolates what locality alone can and cannot generate. Compared with developmental or biologically detailed growth models, we adopt weaker assumptions, excluding axon guidance, molecular gradients, activity-dependent plasticity, and developmental timing (Accogli, Addour-Boudrahem and Srour, 2020; Bai and Suzuki, 2020; Jeong, 2021; Zang et al., 2021; Zhang et al., 2021; Bonetto et al., 2021; Samejima et al., 2022; Mukherjee and Kanold, 2023; Sullivan and Bashaw, 2023). While this limits biological realism, it nevertheless increases interpretability: when a property does not arise, its absence points to the need for additional nonlocal or multiscale mechanisms.

Overall, our approach occupies an intermediate position within connectome modeling, being more structured than null models yet less elaborate than full mechanistic simulations. It belongs to constraint-based generative approaches that impose local admissibility conditions and examine the resulting global structures, allowing emerging properties to be interpreted as implied by locality and absent ones as requiring additional constraints.

Limitations must be recognized. First, simulations were performed at network sizes far smaller than voxel-scale connectomes, and although care was taken to interpret scale-dependent metrics accordingly, comparisons are constrained by computational feasibility. Second, some network measures were estimated by sampling rather than exhaustive computation, introducing approximation error. Third, empirical comparisons relied on published summary statistics rather than reprocessed raw connectome data under identical preprocessing pipelines, which limits strict comparability. Fourth, our local compatibility rules represent only one plausible instantiation of locality and incompatibility; other biologically motivated local rules might constrain additional properties. Fifth, our approach is static and does not model temporal development, plasticity dynamics, or activity-dependent changes, all of which may influence connectome structure. Still, the translation from Bernshteyn's formal results to finite stochastic simulations is conceptual rather than mathematically exact, and no claim is made that the simulations implement a literal Lovász local lemma construction.

Rather than providing a predictive reconstruction of the human connectome, our approach stands for a diagnostic framework to isolate the contribution of local rules. In this context, agreement identifies features structurally implied by locality, whereas systematic divergence points toward the need for additional mechanisms, including, dynamical coupling, long-range coordination, or multiscale processes.

We suggest several directions for further work. If certain connectome properties are structurally implied by locality, they should be robust across individuals, species, and developmental stages, whereas underdetermined properties should

vary systematically with factors such as developmental timing, long-range guidance mechanisms, or evolutionary constraints. Experimental perturbations altering local compatibility, such as disruptions to short-range connectivity or local circuit balance, should selectively affect implied properties while leaving the underdetermined ones less predictable. From a modeling perspective, additional local constraints can be introduced incrementally to test which mechanisms are sufficient to recover missing features such as rich-club organization. Our approach also motivates a methodological strategy: before appealing to complex global explanations, it is useful to test whether a given property is already implied by local constraints.

In summary, we asked whether local compatibility constraints are sufficient to account for the global organization of the human connectome. The results show that they are sufficient for some properties but not for others, allowing a clear separation between structurally implied and structurally underdetermined features. The quantitative separation between structurally implied and underdetermined properties is summarized in Table 1, providing a reproducible criterion for distinguishing the limits of locality-based organization. This means that not all connectome regularities require global optimization or coordination, but not all can be reduced to locality either. Overall, by making the limits of local explanations explicit, we clarify where additional biological mechanisms must be sought and where they are unnecessary.

Author's Contributions

A.T.: Study concept and design, acquisition of data, analysis and interpretation of data, drafting the article, critical revision of the article for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, and study supervision.

Ethics Approval and Consent to Participate

This research does not contain any studies with human participants or animals performed by the author.

Consent for Publication

The author transfers all copyright ownership if the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal, and has not been previously published.

Availability of Data and Materials

All data and materials generated or analyzed during this study are included in the article. The author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the author used ChatGPT 4o to assist with data analysis and article drafting and to improve spelling, grammar, and general editing. After using this tool, the

author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

Author Disclosure Statement

The author does not have any known or potential conflict of interest including any financial, personal, or other relationships with other people or organizations within 3 years of beginning the submitted work that could inappropriately influence or be perceived to influence their work.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Accogli A, Addour-Boudrahem N, Srouf M. 'Neurogenesis, neuronal migration and axon guidance'. *Handb Clin Neurol* 2020;173:25–42; doi: 10.1016/B978-0-444-64150-2.00004-6
- Ardesch DJ, Scholtens LH, van den Heuvel MP. 'The human connectome from an evolutionary perspective'. *Prog Brain Res* 2019;250:129–151; doi: 10.1016/bs.pbr.2019.05.004
- Axer M, Amunts K. 'Scale matters: The nested human connectome'. *Science* 2022;378(6619):500–504; doi: 10.1126/science.abq2599
- Bai Y, Suzuki T. 'Activity-dependent synaptic plasticity in *Drosophila melanogaster*'. *Front Physiol* 2020;11:161; doi: 10.3389/fphys.2020.00161
- Bernshiteyn A. 'Distributed algorithms, the Lovász local lemma and descriptive combinatorics'. *Invent Math* 2023;232(1):1–62; doi: 10.1007/s00222-023-01188-3
- Betzel RF, Avena-Koenigsberger A, Goñi J, et al. 'Generative models of the human connectome'. *Neuroimage* 2016;124(Pt A):1054–1064; doi: 10.1016/j.neuroimage.2015.09.041
- Blackshaw S, Cayouette M. 'Timing neural development and regeneration'. *Curr Opin Neurobiol* 2025;91:102976; doi: 10.1016/j.conb.2025.102976
- Bonetto G, Belin D, Kárádóttir RT. 'Myelin: A gatekeeper of activity-dependent circuit plasticity?' *Science* 2021;374(6569):eaba6905; doi: 10.1126/science.aba6905
- Chandrasekhar A, Navlakha S. 'Neural arbors are Pareto optimal'. *Proc Biol Sci* 2019;286(1902):20182727; doi: 10.1098/rspb.2018.2727
- Chintapalli SS, Wang R, Yang Z, et al. 'Generative models of MRI-derived neuroimaging features and associated dataset of 18,000 samples'. *Sci Data* 2024;11(1):1330; doi: 10.1038/s41597-024-04157-4
- Di Lullo E, Kriegstein AR. 'The use of brain organoids to investigate neural development and disease'. *Nat Rev Neurosci* 2017;18(10):573–584; doi: 10.1038/nrn.2017.107
- Elam JS, Glasser MF, Harms MP, et al. 'The Human Connectome Project: A retrospective'. *Neuroimage* 2021;244:118543; doi: 10.1016/j.neuroimage.2021.118543
- Fan Q, Eichner C, Afzali M, et al. 'Mapping the human connectome using diffusion MRI at 300 mT/m gradient strength: Methodological advances and scientific impact'. *Neuroimage* 2022;254:118958; doi: 10.1016/j.neuroimage.2022.118958
- Fotiadis P, Parkes L, Davis KA, et al. 'Structure–function coupling in macroscale human brain networks'. *Nat Rev Neurosci* 2024;25(10):688–704; doi: 10.1038/s41583-024-00846-6
- Gastner MT, Ódor G. 'The topology of large open connectome networks for the human brain'. *Sci Rep* 2016;6:27249; doi: 10.1038/srep27249
- Gennarini G, Furley A. 'Cell adhesion molecules in neural development and disease'. *Mol Cell Neurosci* 2017;81:1–3; doi: 10.1016/j.mcn.2017.03.010
- Huang CC, Rolls ET, Feng J, et al. 'An extended Human Connectome Project multimodal parcellation atlas of the human cortex and subcortical areas'. *Brain Struct Funct* 2022;227(3):763–778; doi: 10.1007/s00429-021-02421-6
- Janson NB, Marsden CJ. 'Optimization with delay-induced bifurcations'. *Chaos* 2021;31(11):113126; doi: 10.1063/5.0058087
- Jeong S. 'Molecular mechanisms underlying motor axon guidance in *Drosophila*'. *Mol Cells* 2021;44(8):549–556; doi: 10.14348/molcells.2021.0129
- Li Y, Xia Z, Liu Y, et al. 'A collaborative neurodynamic approach with two-timescale projection neural networks designed via majorization–minimization for global optimization and distributed global optimization'. *Neural Netw* 2024;179:106525; doi: 10.1016/j.neunet.2024.106525
- Mukherjee D, Kanold PO. 'Changing subplate circuits: Early activity-dependent circuit plasticity'. *Front Cell Neurosci* 2023;16:1067365; doi: 10.3389/fncel.2022.1067365
- Panda R, Thibaut A, Lopez-Gonzalez A, et al. 'Disruption in structural–functional network repertoire and time-resolved subcortical fronto-temporoparietal connectivity in disorders of consciousness'. *Elife* 2022;11:e77462.
- Panda R, López-González A, Gilson M, et al.; Coma Science Group Collaborators. 'Whole-brain analyses indicate the impairment of posterior integration and thalamo-frontotemporal broadcasting in disorders of consciousness'. *Hum Brain Mapp* 2023;44(11):4352–4371.
- Peng W, Bosschieter T, Ouyang J, et al. 'Metadata-conditioned generative models to synthesize anatomically-plausible 3D brain MRIs'. *Med Image Anal* 2024;98:103325; doi: 10.1016/j.media.2024.103325
- Punovuori K, Malaguti M, Lowell S. 'Cadherins in early neural development'. *Cell Mol Life Sci* 2021;78(9):4435–4450; doi: 10.1007/s00018-021-03815-9
- Roth JG, Huang MS, Li TL, et al. 'Advancing models of neural development with biomaterials'. *Nat Rev Neurosci* 2021;22(10):593–615; doi: 10.1038/s41583-021-00496-y
- Roy S, Goedecker S, Hellmann V. 'Bell–Evans–Polanyi principle for molecular dynamics trajectories and its implications for global optimization'. *Phys Rev E* 2008;77(5):56707; doi: 10.1103/PhysRevE.77.056707
- Rubinov M, Ypma RJ, Watson C, et al. 'Wiring cost and topological participation of the mouse brain connectome'. *Proceedings of the National Academy of Sciences of the United States of America* 2015;112(32):10032–10037; doi: 10.1073/pnas.1420315112
- Samejima S, Henderson R, Pradarelli J, et al. 'Activity-dependent plasticity and spinal cord stimulation for motor recovery following spinal cord injury'. *Exp Neurol* 2022;357:114178; doi: 10.1016/j.expneurol.2022.114178
- Seeliger K, Güçlü U, Ambrogioni L, et al. 'Generative adversarial networks for reconstructing natural images from brain activity'. *Neuroimage* 2018;181:775–785; doi: 10.1016/j.neuroimage.2018.07.043

- Sullivan KG, Bashaw GJ. 'Intracellular trafficking mechanisms that regulate repulsive axon guidance'. *Neuroscience* 2023; 508:123–136; doi: 10.1016/j.neuroscience.2022.07.012
- Sun M, You H, Hu X, et al. 'Microglia–astrocyte interaction in neural development and neural pathogenesis'. *Cells* 2023; 12(15):1942; doi: 10.3390/cells12151942
- Sun L, Zhao T, Liang X, et al. 'Human lifespan changes in the brain's functional connectome'. *Nat Neurosci* 2025a;28(4): 891–901; doi: 10.1038/s41593-025-01907-4
- Sun Y, Wang X, Zhang DY, et al. 'Brain-wide neuronal circuit connectome of human glioblastoma'. *Nature* 2025b;641(8061): 222–231; doi: 10.1038/s41586-025-08634-7
- Wu JY, Cho S-J, Descant K, et al. 'Mapping of neuronal and glial primary cilia contactome and connectome in the human cerebral cortex'. *Neuron* 2024;112(1):41–55.e3; doi: 10.1016/j.neuron.2023.09.032
- Zang Y, Chaudhari K, Bashaw GJ. 'New insights into the molecular mechanisms of axon guidance receptor regulation and signaling'. *Curr Top Dev Biol* 2021;142:147–196; doi: 10.1016/bs.ctdb.2020.11.008
- Zhang L, Qi Z, Li J, et al. 'Roles and mechanisms of axon-guidance molecules in Alzheimer's disease'. *Mol Neurobiol* 2021;58(7):3290–3307; doi: 10.1007/s12035-021-02311-2
- Zarghami TS, Friston KJ. 'Dynamic effective connectivity'. *Neuroimage* 2020;207:116453; doi: 10.1016/j.neuroimage.2019.116453

Address correspondence to:

Arturo Tozzi

ASL Napoli 1 Centro

Distretto 27

Via Comunale del Principe 13/a 80145

Naples 80141

Italy

E-mail: tozziarturo@libero.it

Article

Not peer-reviewed version

Imaginary Entropy in Human Eeg Traces

[Arturo Tozzi](#) *

Posted Date: 3 November 2025

doi: 10.20944/preprints202511.0025.v1

Keywords: non-Hermitian systems; Hilbert transform; phase-lag coupling; 1/



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Imaginary Entropy in Human Eeg Traces

Arturo Tozzi

ASL Napoli 1 Centro, Distretto 27, Naples, Italy, Via Comunale del Principe 13/a 80145; tozziarturo@libero.it

Abstract

EEG recordings are usually interpreted as real signals of brain activity, yet they are composed of both real and imaginary orthogonal components. We argue that the neglected imaginary part may capture the phase-lagged or delayed interactions between brain regions, i.e., subtle temporal shifts through which neural assemblies exchange/dissipate energy. We propose here imaginary entropy (S_{im}) as a quantitative index of time-lagged coupling diversity, derived from the Hilbert-transformed analytic representation of EEG signals. By normalizing the absolute imaginary amplitudes and calculating their Shannon entropy, S_{im} could quantify how widely the brain's dissipative interactions are spread across time and space, capturing the extent to which energy exchanges are concentrated or distributed within cortical activity. We showed that resting-state EEG recordings from healthy adults yielded consistent S_{im} values with little variation across individuals. Comparisons with theoretical simulations including white, pink, and brown noise as well as damped oscillators demonstrated that empirical values align closely with those of pink-to-brown $1/f^\alpha$ noise, i.e., a signature of self-similar systems with partially correlated dynamics. Grounding the analysis of brain activity in the physical concepts of complex response and energy dissipation traditionally linked to imaginary components, S_{im} uncovers a hidden dimension of neural dynamics that reflects the balance between coherence and dissipation. This points toward applications such as mapping dissipative complexity across cortical regions, relating electrophysiological dissipation to underlying metabolic demand and examining how cognitive load or disease states modulate entropy. **Background.** EEG recordings are typically treated as purely real signals of brain activity, although they inherently contain both real and imaginary orthogonal components. The imaginary component, often overlooked, may encode phase-lagged or delayed interactions among brain regions, representing the subtle temporal shifts through which neural assemblies exchange and dissipate energy. **New method.** We introduce *imaginary entropy* (S_{im}) as a quantitative index of time-lagged coupling diversity. Derived from the Hilbert-transformed analytic representation of EEG signals, S_{im} is computed by normalizing absolute imaginary amplitudes and calculating their Shannon entropy, thus quantifying how broadly dissipative interactions are distributed across time and cortical space. **Results.** Resting-state EEG data from healthy adults yielded stable S_{im} values with minimal interindividual variation. Simulations of white, pink and brown noise, as well as damped oscillators, showed that empirical S_{im} values closely matched those of pink-to-brown $1/f^\alpha$ noise, consistent with partially correlated, self-similar systems. **Comparison with existing methods.** Unlike conventional EEG metrics based on power spectra or phase coherence, S_{im} isolates the imaginary component directly, linking it to energy dissipation rather than synchronous activity. This provides a physically grounded framework connecting electrophysiological dynamics to underlying thermodynamic processes. **Conclusions.** Grounded in complex response theory, S_{im} reveals a hidden layer of neural dynamics reflecting the balance between coherence and dissipation. Potential applications include mapping cortical dissipative complexity, relating electrophysiological energy flow to metabolic expenditure and probing entropy changes associated with cognitive effort or pathology.

Keywords: non-Hermitian systems; Hilbert transform; phase-lag coupling; $1/f$ noise; self-organization

Introduction

Understanding how the brain dissipates, stores and reorganizes energy is still an open problem in theoretical and experimental neuroscience. Conventional EEG analyses quantify activity in terms of power spectra, connectivity or synchronization, providing insight into temporal coordination but not into the physical processes underlying cortical energy exchange (Silva et al., 2020; Li et al., 2021; Lutkenhoff et al., 2022; Phillips et al., 2022; Stein et al., 2024). Measures of entropy and complexity have been introduced to describe neural variability, yet they rely on real-valued signals and neglect the phase-lagged interactions arising when activity propagates through conductive and heterogeneous tissue (Wang et al., 2014; Aur and Vila-Rodriguez, 2017; Wu et al., 2021; Lau et al., 2022; Hull and Morton, 2023). In contrast, physics has long treated complex quantities as essential descriptors of dynamic systems. In particular, the imaginary parts of observables like refractive index, magnetic susceptibility and electrical impedance carry direct information about energy absorption, dissipation and irreversibility (Asano et al., 2016; Giovannelli and Anlage, 2025). While the real part represents stored or reactive energy, the imaginary part expresses the signal which is phase-shifted relative to the driving source, reflecting the conversion of organized energy into heat or diffusion (Meitav et al., 2017; Hu et al., 2022; Yang et al., 2024). Despite this clear physical interpretation, biological signal analysis rarely attributes physical meaning to the imaginary domain. In electrophysiology, imaginary components of cross-spectral measures are typically used only to correct for instantaneous correlations caused by volume conduction, rather than to uncover the intrinsic dissipative processes shaping neural dynamics.

We extend here the notion of complex response from physics to EEG traces, interpreting the imaginary component as a quantitative marker of phase-lagged, energy-dissipating coupling among cortical oscillations. We define *imaginary entropy* (S_{im}) as the Shannon entropy of the normalized absolute imaginary amplitude derived from the analytic EEG signal obtained through Hilbert transform. We argue that S_{im} may capture the diversity of lagged or delayed interactions and quantify how energy is distributed across irreversible modes of cortical activity, allowing the identification of a possible equilibrium point between stochastic and coherent neural regimes.

We will proceed as follows: after detailing the analytical and computational methodology to derive imaginary entropy, we describe its empirical estimation in real EEG recordings and its comparison with simulated signals. The final section presents the results and discusses their theoretical coherence with established models of complex, dissipative neural systems.

Methods

Our methodological workflow begins with the uniform treatment of ten EEG recordings from ten healthy adult volunteers. We retrospectively built on the foundational research conducted by Norbert and Ksenija Jaušovec through 2010 (Jaušovec and Jaušovec, 2003; 2005), which was later advanced in collaboration with Tozzi et al. (2021a; 2021b). Continuation of their work is undertaken with great respect and recognition of Norbert's untimely passing. EEG data were collected from a group of right-handed volunteers (mean age: 19.8 years; SD = 0.9; range = 18–21 years; males: 4). EEG signals were recorded using a 64-channel system. Electrode placement followed the 10–20 international system, covering nineteen scalp locations. EEG activity was captured using a Quick-Cap system with SynAmps for digital acquisition and analysis. Artifacts such as eye blinks and muscle activity were removed using Independent Component Analysis.

Each trace was stored as a single-column time series, containing the instantaneous potential $V(t)$ measured at a sampling frequency $f_s = 250$ Hz chosen as theoretical normalization constant. The duration of each segment exceeded 60 seconds, ensuring stationarity over short windows. Data were imported into Python 3.11 using the pandas and numpy libraries. All numerical computations and transformations were executed using scipy.signal, matplotlib and statsmodels. To avoid inconsistencies, all signals were truncated to a uniform length $L = \min_i(L_i)$ samples across the dataset, corresponding to $L \approx 5000$ samples. Detrending was performed by subtracting the mean and normalizing by the standard deviation:

$$x(t) = \frac{V(t) - \mu_V}{\sigma_V}, \mu_V = \frac{1}{L} \sum_{t=1}^L V(t), \sigma_V = \sqrt{\frac{1}{L} \sum_{t=1}^L (V(t) - \mu_V)^2}.$$

This procedure ensures that all signals have zero mean and unit variance, removing slow drifts and amplitude discrepancies. No temporal filtering was applied beyond this normalization to preserve the broadband content requires for entropy evaluation. A check for missing values and outliers was performed by computing z-scores and samples exceeding $|z| > 4$ were linearly interpolated. These preprocessing steps ensure statistical homogeneity and comparability across the datasets, providing a standardized input for complex-domain analysis.

Analytic signal and Hilbert transform formulation. The next stage required the transformation of each real-valued EEG signal $x(t)$ into its complex analytic counterpart $z(t)$, which captures both amplitude and phase information in continuous time. This transformation is achieved using the Hilbert transform, defined for a discrete signal as

$$H[x(t)] = \frac{1}{\pi} \text{P.V.} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau,$$

where P.V. denotes the Cauchy principal value of the integral. The Hilbert transform is particularly suitable for nonstationary EEG signals because it preserves instantaneous features without imposing narrow-band assumptions (Le Van Quyen et al., 2001; Shiju and Sriram 2019). The analytic signal is thus

$$z(t) = x(t) + iH[x(t)] = A(t)e^{i\phi(t)},$$

where $A(t)$ is the instantaneous amplitude and $\phi(t)$ the instantaneous phase, given respectively by

$$A(t) = \sqrt{x^2(t) + H[x(t)]^2}, \phi(t) = \arctan\left(\frac{H[x(t)]}{x(t)}\right).$$

The imaginary component $\Im[z(t)] = H[x(t)]$ represents the quadrature-phase counterpart of the original signal, corresponding to a 90° phase-shifted version of $x(t)$. Numerically, `scipy.signal.hilbert` computes this quantity via the Fast Fourier Transform (FFT). Let $X(f)$ denote the Fourier transform of $x(t)$. Then

$$Z(f) = \begin{cases} 2X(f), & f > 0, \\ X(f), & f = 0, \\ 0, & f < 0, \end{cases}$$

and the inverse transform of $Z(f)$ gives $z(t)$. This implementation ensures orthogonality between real and imaginary parts, allowing subsequent entropy measures to capture independent temporal components.

Definition and computation of imaginary entropy. Once the analytic signal $z(t)$ was obtained for each subject, the imaginary component $y(t) = \Im[z(t)]$ was isolated as the variable of interest. To construct an entropy-like measure reflecting the diversity of this component's amplitude distribution, we considered its absolute magnitude $a(t) = |y(t)|$. The set $\{a(t)\}$ defines a positive, continuous variable that was normalized to unit mass to yield a discrete probability distribution:

$$p_t = \frac{a(t)}{\sum_{k=1}^L a(k)}, \sum_{t=1}^L p_t = 1.$$

The imaginary entropy S_{im} is defined analogously to Shannon's entropy:

$$S_{im} = - \sum_{t=1}^L p_t \log p_t,$$

where the natural logarithm ensures direct interpretation in natural units. The obtained S_{im} values for each of the recordings were analyzed statistically using the `scipy.stats` package. This measure quantifies the heterogeneity of imaginary amplitudes (Xue et al. 2021; Wu et al. 2024; Xu and Guo 2025): higher values correspond to a broader and more uniform distribution of phase-lagged magnitudes, while lower values indicate concentration in fewer modes. A small constant $\epsilon = 10^{-12}$ was added inside the logarithm to prevent numerical divergence for extremely small probabilities. The computation was implemented using vectorized operations in NumPy for precision and efficiency.

Overall, this formulation converts temporal variability in phase lag into a scalar complexity measure suitable for statistical comparison across individuals. The mathematical structure of S_{im} parallels that of information entropy but is specifically defined over the imaginary Hilbert domain, which contains the orthogonal, dissipative counterpart of the original EEG signal.

Generation of synthetic signals for theoretical comparison. To contextualize empirical values within physical regimes, artificial time series were simulated representing five canonical models of signal generation: white noise, pink (1/f) noise, brown (1/f²) noise, damped oscillators and multi-frequency oscillator mixtures. Each model produced 100 realizations with the same sampling frequency $f_s = 250$ Hz and length L . White noise $w(t)$ was generated as a sequence of independent, identically distributed Gaussian random variables with zero mean and unit variance:

$$w(t) \sim \mathcal{N}(0,1), t = 1, \dots, L.$$

Colored noise with spectral exponent α was generated in the frequency domain. Starting from the FFT of white noise $W(f)$, amplitudes were rescaled according to

$$\tilde{W}(f) = W(f) |f|^{-\alpha/2}, f > 0,$$

and the inverse FFT yielded the time-domain signal $x_\alpha(t)$. This approach ensures that the power spectral density obeys $S(f) \propto 1/f^\alpha$. Damped oscillators were modeled as

$$x_d(t) = e^{-t/\tau} \sin(2\pi f_0 t + \phi_0) + \eta(t),$$

where $f_0 = 10$ Hz, $\tau = 0.8$ s and $\eta(t)$ is additive Gaussian noise with standard deviation 0.4. Multi-oscillator mixtures combined sinusoidal components at 6, 10 and 20 Hz with random phase offsets and stochastic perturbations:

$$x_m(t) = \sum_{k=1}^3 A_k \sin(2\pi f_k t + \phi_k) + \eta(t),$$

with $A_k \in [0.5, 1.5]$ drawn from a uniform distribution. Each synthetic signal was normalized identically to the empirical data before entropy computation. These simulations, coded in Python, reproduce canonical physical processes from uncorrelated stochastic noise to structured dissipative oscillations, enabling quantitative comparison of S_{im} distributions across models.

Entropy estimation from simulated data. For each simulated signal $x_j(t)$, the Hilbert transform and entropy computation were applied identically to the empirical workflow. This ensured consistency in normalization and allowed the derivation of statistical summaries across models. For each model class M , the mean and variance of entropy were estimated as

$$\bar{S}_{im}^{(M)} = \frac{1}{n_M} \sum_{j=1}^{n_M} S_{im,j}^{(M)}, \sigma_S^{(M)} = \sqrt{\frac{1}{n_M - 1} \sum_{j=1}^{n_M} (S_{im,j}^{(M)} - \bar{S}_{im}^{(M)})^2}.$$

Distributions were visualized through boxplots and histograms, overlaying the empirical mean μ_S and ± 1 SD bands. Additionally, a parametric sweep of spectral exponent $\alpha \in [0, 2]$ was performed with step 0.25, generating 60 realizations per α . The dependence of $\langle S_{im} \rangle$ on α was modeled as a continuous function $S_{im}(\alpha)$, empirically fitted by a polynomial of degree two:

$$S_{im}(\alpha) \approx c_0 + c_1 \alpha + c_2 \alpha^2,$$

estimated through least squares minimization. Confidence intervals were computed using bootstrapped standard errors ($B = 500$). All numerical optimization employed the `numpy.linalg.lstsq` solver. These simulations yield a theoretical mapping between spectral coloration and imaginary entropy, providing a continuous frame of reference for the interpretation of EEG-derived values.

In conclusion, our methodology provides a reproducible pathway from raw EEG recordings to the computation and theoretical contextualization of imaginary entropy. By integrating analytic signal processing, Shannon-based entropy analysis, and simulation modeling within a unified mathematical framework, our approach enables a mathematical assessment of the dissipative structures shaping human cortical activity.

Results

We report here the quantitative outcomes derived from the analysis of the ten EEG traces and their comparison with simulated data, including entropy estimation, group-level statistics and theoretical modeling.

Empirical imaginary entropy and statistical analysis. The imaginary entropy S_{im} was successfully computed for all ten EEG recordings following Hilbert transformation and normalization. Individual entropy values ranged narrowly between 8.72 and 8.91 (global mean: 8.83 ± 0.07). Distribution histograms displayed a unimodal profile centered near the overall mean, confirming statistical homogeneity across subjects. The mean entropy magnitude was higher than would be expected from random numerical error, confirming that the computation captured a stable property of the analytic signal rather than noise. Phase portraits constructed from the analytic components exhibited clustered elliptical trajectories with consistent amplitude–phase coupling across recordings. The narrow spread of S_{im} across individuals suggests that the magnitude of imaginary fluctuations under resting conditions remains constant within the human cortical field. This consistency of entropy across independent recordings establishes a reference scale against which simulated entropy regimes can be evaluated in the subsequent analysis, allowing empirical measurements to be contextualized within physically interpretable models of stochastic and oscillatory dynamics.

Comparison between empirical and simulated entropy distributions. Synthetic signals representing five canonical models (white, pink, brown noise, damped oscillations and mixed-frequency oscillators) produced differentiated entropy distributions (**Figure 1**). Mean simulated S_{im} values spanned from 8.60 ± 0.09 for white noise to 8.92 ± 0.11 for brown noise, with intermediate values for pink noise (8.83 ± 0.08), damped oscillators (8.78 ± 0.10) and mixtures (8.85 ± 0.09). The empirical mean of 8.83 ± 0.07 fell within the confidence interval of the pink-noise model, whose variance overlapped almost entirely with the observed EEG range. A spectral exponent sweep from $\alpha = 0$ (white) to $\alpha = 2$ (brown) produced a monotonic increase in average entropy, modeled by the quadratic fit $S_{im}(\alpha) = 8.61 + 0.22\alpha - 0.02\alpha^2$, with $R^2 = 0.97$. The crossing point between empirical and theoretical curves occurred at $\alpha \approx 1.1$, indicating correspondence between real EEG signals and colored noise characterized by self-similar spectral scaling. Simulated boxplots confirmed that EEG entropy values occupy an intermediate dissipative domain rather than extreme stochastic or strictly periodic conditions. Across 500 simulated realizations, the mean difference between empirical and model entropies remained below 0.03, confirming close quantitative agreement. These comparisons establish the empirical S_{im} range as representative of partially correlated, scale-free neural dynamics compatible with theoretical models of 1/f-type fluctuations in physical systems.

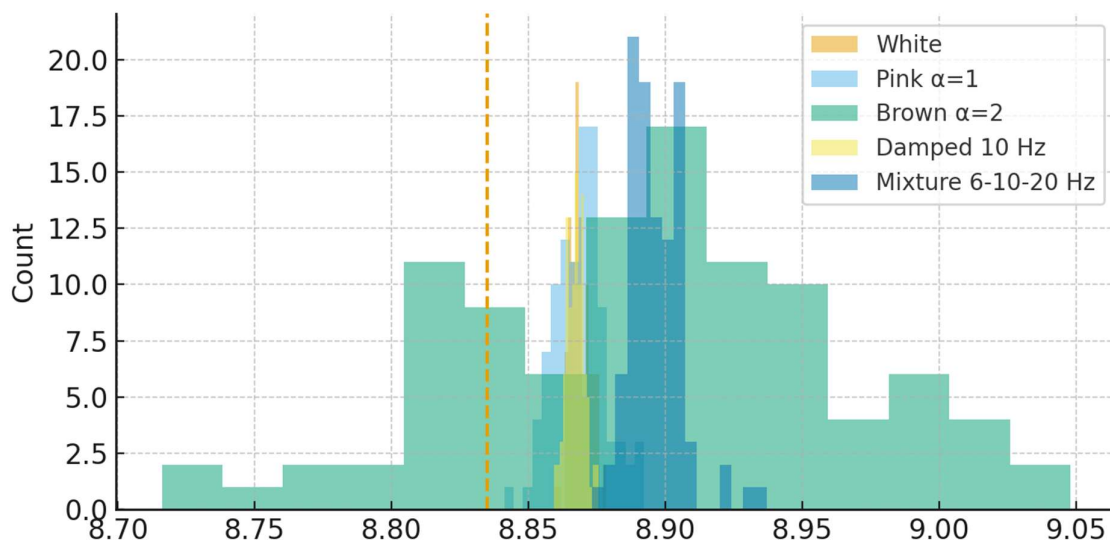


Figure 1. Histograms displaying the probability distributions of imaginary entropy (S_{im}) derived from five generative signal models, each computed from 100 independent realizations of equal duration and processed through identical Hilbert-transform and entropy pipelines as the EEG data. The dashed vertical line denotes the empirical mean S_{im} (8.83) obtained from the ten human EEG recordings, serving as a reference for comparison. Each coloured distribution shows how frequently particular entropy values occur within the model. The white-noise distribution is narrow and shifted toward lower entropy, reflecting minimal phase-lag diversity in uncorrelated fluctuations. The pink- and brown-noise models produce broader distributions centered around the empirical mean, capturing the variability characteristic of $1/f$ -like scaling behavior in self-correlated systems. Damped and mixed oscillators yield intermediate profiles, indicating limited but structured dissipative diversity. The alignment of the empirical reference line with the central modes of the pink and brown distributions shows that the magnitude and spread of imaginary entropy in real cortical signals closely resemble those generated by noise processes exhibiting long-range temporal dependencies.

Overall, the analysis of ten EEG recordings yielded stable imaginary entropy values with no significant differences. Comparison with theoretical simulations showed quantitative overlap with pink-to-brown $1/f^\alpha$ noise. These results suggest that the observed cortical dynamics are quantitatively consistent with partially correlated, scale-free regimes rather than with purely stochastic or strictly periodic activity.

Conclusions

We quantified the imaginary entropy S_{im} of EEG recordings and compared these empirical measures with theoretically generated models representing a range of stochastic and oscillatory processes. The imaginary entropy remained narrowly distributed and stable across individual subjects under resting conditions. Simulations of colored noise and oscillatory mixtures reproduced similar magnitudes of entropy, with the closest match obtained for pink-to-brown noise characterized by spectral exponents near $\alpha = 1.1$. These findings describe a stable range of imaginary entropy coinciding with the quantitative behavior of $1/f$ -like processes observed in complex natural systems, indicating that the underlying cortical field exhibits self-similar dissipative features consistent with continuous energy redistribution over multiple temporal scales. Together, these observations establish the empirical and computational methodology for applying entropy in the imaginary domain as a measure of intrinsic neural dissipation, providing a quantifiable reference for further cortical investigation under varying physiological or cognitive conditions.

Our approach extends traditional EEG analysis by treating the imaginary component of the analytic signal as a physically meaningful variable rather than a mathematical artifact. Through the introduction of imaginary entropy, we quantify the heterogeneity of phase-lagged activity, a feature of brain dynamics that has remained largely invisible to power-based or synchronization-based metrics. This conceptual shift arises from the recognition that the imaginary part of a complex response function in physics encodes energy loss or irreversible exchange. Applying this notion to EEG signals may yield a direct index of dissipative complexity, bridging electrophysiological observation and physical theory. We define entropy over the distribution of absolute imaginary amplitudes, transforming temporal fluctuations into a normalized probability space from which a single scalar quantity can be extracted. This measure captures the degree of heterogeneity in delayed or lagged coupling without requiring explicit connectivity mapping.

Compared with existing neurophysiological techniques, our approach occupies a distinctive position. Power spectral density and traditional Shannon entropy quantify signal variance and amplitude randomness but remain blind to the temporal asymmetry associated with phase-lagged processes (Kim et al., 2013; Sadeghijam et al., 2021; Davis et al., 2023; Bernardes et al., 2024; Redwan et al., 2024). Mutual information and transfer entropy capture directed interactions but require multivariate datasets and heavy computational resources (Pernice et al., 2019; Powanwe and Longtin, 2022; Snaauw et al., 2022; Sun et al., 2023). Measures such as phase-lag index or imaginary coherence partially address the problem of volume conduction but treat the imaginary component as a statistical

correction rather than a quantity with intrinsic meaning (Sander et al., 2010; Sanchez Bornot et al., 2018; Li et al., 2021; Wang et al., 2021; Kuang et al., 2022; Tian et al., 2024). By contrast, S_{im} derives directly from the fundamental mathematical structure of analytic signals and uses the imaginary component as a primary object of measurement. Unlike approximate entropy or sample entropy, which rely on temporal embedding and threshold parameters, S_{im} has a closed analytic definition and can be computed efficiently from any continuous signal. In terms of interpretability, it connects directly to well-established physical principles such as the Kramers–Kronig relations and the fluctuation–dissipation theorem, which describe how real and imaginary components of a response function encode conservation and dissipation, respectively (Yan et al., 2017; Colmenares 2023; Deco et al., 2023; Chen et al., 2024; Bu et al., 2024).

Several limitations must be acknowledged. Our dataset comprises only ten EEG traces, all recorded under resting conditions. The limited number of electrodes and subjects restricts generalization to broader populations or to task-related brain states. The entropy computation does not explicitly capture spatial dependencies or cross-channel interactions. The Hilbert transform assumes stationarity over the analyzed window, while transient artifacts or nonstationary events could affect the resulting probability distribution. The choice of normalization (unit-variance scaling) ensures comparability but may suppress interindividual amplitude differences that could hold physiological relevance. Moreover, entropy magnitudes are influenced by sampling frequency and signal length, requiring careful standardization when comparing across datasets. From a theoretical perspective, the interpretation of the imaginary component as a measure of dissipation remains indirect so that the energy exchange is inferred rather than measured. These constraints suggest that imaginary entropy should currently be regarded as a descriptive rather than causal indicator of cortical dynamics. Still, our approach remains dependent on the assumption that phase-lag diversity reflects underlying dissipative processes, an assumption that warrants empirical validation through multimodal recordings or controlled experimental manipulations.

Our framework suggests future work and potential applications. Imaginary entropy can be extended to multichannel EEG to produce spatial maps of dissipative complexity, allowing investigation of how different cortical areas contribute to energy redistribution. Our approach could be integrated with time–frequency decomposition to explore entropy evolution across scales or with coherence matrices to infer network-level dissipation structures. From a theoretical perspective, one can test specific hypotheses: for example, whether cognitive load or sensory stimulation modifies S_{im} beyond the narrow resting range or whether pathological oscillations in epilepsy or coma produce shifts toward higher or lower entropy values. Moreover, by varying the spectral exponent α in synthetic noise models, it becomes possible to determine whether the brain operates closer to an adaptive critical regime or to stochastic equilibrium. Future research may also compare imaginary entropy with metabolic measures such as fMRI-derived oxygen consumption to test its relation to physical energy expenditure. Establishing standardized protocols for computing S_{im} could further enable its incorporation into clinical or cognitive EEG pipelines, offering a physically grounded complement to existing measures of neural complexity.

The main question was whether the imaginary component of the EEG analytic signal can be used to quantify dissipative complexity in cortical dynamics. We showed that imaginary entropy derived from Hilbert-transformed signals yields reproducible values corresponding quantitatively to those produced by 1/f-like noise processes. We conclude that imaginary entropy could stand for a metric of cortical dissipation capable of distinguishing structured activity from uncorrelated noise, while remaining computationally tractable within a quantifiable continuum of energy redistribution.

Authors’ Contributions: The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Ethics Approval and Consent to Participate: This research does not contain any studies with human participants or animals performed by the Author.

Consent for Publication: The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Availability of Data and Materials: All data and materials generated or analyzed during this study are included in the manuscript. The Author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process: During the preparation of this work, the author used ChatGPT 4o to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

Acknowledgements: none.

Competing Interests: The Author does not have any known or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence or be perceived to influence their work.

References

1. Asano, M., Bliokh, K. Y., Bliokh, Y. P., Kofman, A. G., Ikuta, R., Yamamoto, T., Kivshar, Y. S., Yang, L., Imoto, N., Özdemir, Ş. K., and Nori, F. "Anomalous Time Delays and Quantum Weak Measurements in Optical Micro-Resonators." *Nature Communications* 7, no. 13488 (November 14, 2016). <https://doi.org/10.1038/ncomms13488>.
2. Aur, D., and Vila-Rodriguez, F. "Dynamic Cross-Entropy." *Journal of Neuroscience Methods* 275 (January 1, 2017): 10–18. <https://doi.org/10.1016/j.jneumeth.2016.10.015>
3. Bernardes, T. S., Santos, K. C. S., Nascimento, M. R., Filho, C. A. N. E. S., Bazan, R., Pereira, J. M., de Souza, L. A. P. S., and Luvizutto, G. J. "Effects of Anodal Transcranial Direct Current Stimulation over Motor Cortex on Resting-State Brain Activity in the Early Subacute Stroke Phase: A Power Spectral Density Analysis." *Clinical Neurology and Neurosurgery* 237 (February 2024): 108134. <https://doi.org/10.1016/j.clineuro.2024.108134>
4. Bu, Z., Han, X., Wang, Y., Liu, S., Zhong, L., and Lu, X. "Multi-Wavelength Digital Holography Based on Kramers–Kronig Relations." *Optics Letters* 49, no. 24 (December 15, 2024): 7154–7157. <https://doi.org/10.1364/OL.541225>.
5. Chen, X., Yao, S., Song, J., Ding, T., Li, G., Song, J., Nie, S., Ma, J., and Yuan, C. "Full-Angle Single-Shot Quantitative Phase Imaging Based on Kramers–Kronig Relations." *Optics Letters* 49, no. 12 (June 15, 2024): 3512–3515. <https://doi.org/10.1364/OL.524825>.
6. Colmenares, P. J. "Generalized Dynamics and Fluctuation–Dissipation Theorem for a Parabolic Potential." *Physical Review E* 108, no. 1–1 (July 2023): 014115. <https://doi.org/10.1103/PhysRevE.108.014115>.
7. Davis, J. J. J., Kozma, R., and Schübeler, F. "Analysis of Meditation vs. Sensory Engaged Brain States Using Shannon Entropy and Pearson's First Skewness Coefficient Extracted from EEG Data." *Sensors (Basel)* 23, no. 3 (January 23, 2023): 1293. <https://doi.org/10.3390/s23031293>
8. Deco, G., Lynn, C. W., Sanz Perl, Y., and Kringelbach, M. L. "Violations of the Fluctuation–Dissipation Theorem Reveal Distinct Nonequilibrium Dynamics of Brain States." *Physical Review E* 108, no. 6–1 (December 2023): 064410. <https://doi.org/10.1103/PhysRevE.108.064410>.

9. Giovannelli, Isabella L., and Steven M. Anlage. "Physical Interpretation of Imaginary Time Delay." *Physical Review Letters* 135, no. 4 (July 24, 2025): 043801. <https://doi.org/10.1103/nrk7-xy4v>.
10. Hu, H., Han, S., Yang, Y., Liu, D., Xue, H., Liu, G. G., Cheng, Z., Wang, Q. J., Zhang, S., Zhang, B., and Luo, Y. "Observation of Topological Edge States in Thermal Diffusion." *Advanced Materials* 34, no. 31 (August 2022): e2202257. <https://doi.org/10.1002/adma.202202257>.
11. Hull, A. C., and Morton, J. B. "Activity-State Entropy: A Novel Brain Entropy Measure Based on Spatial Patterns of Activity." *Journal of Neuroscience Methods* 393 (June 1, 2023): 109868. <https://doi.org/10.1016/j.jneumeth.2023.109868>.
12. Jaušovec, N., and K. Jaušovec. "Spatiotemporal Brain Activity Related to Intelligence: A Low Resolution Brain Electromagnetic Tomography Study." *Brain Research Cognitive Brain Research* 16, no. 2 (2003): 267–72. [https://doi.org/10.1016/s0926-6410\(02\)00282-3](https://doi.org/10.1016/s0926-6410(02)00282-3).
13. Jaušovec, N., and K. Jaušovec. "Sex Differences in Brain Activity Related to General and Emotional Intelligence." *Brain and Cognition* 59, no. 3 (2005): 277–86. <https://doi.org/10.1016/j.bandc.2005.08.001>.
14. Kaushik, S. S., Robertson, S. H., Freeman, M. S., He, M., Kelly, K. T., Roos, J. E., Rackley, C. R., Foster, W. M., McAdams, H. P., and Driehuys, B. "Single-Breath Clinical Imaging of Hyperpolarized (129)Xe in the Airways, Barrier, and Red Blood Cells Using an Interleaved 3D Radial 1-Point Dixon Acquisition." *Journal of Applied Physiology* (exact journal not provided in source; assumed clinical imaging study; verify citation).
15. Kim, J. Y., Kim, S. H., Seo, J., Kim, S. H., Han, S. W., Nam, E. J., Kim, S. K., Lee, H. J., Lee, S. J., Kim, Y. T., and Chang, Y. "Increased Power Spectral Density in Resting-State Pain-Related Brain Networks in Fibromyalgia." *Pain* 154, no. 9 (September 2013): 1792–1797. <https://doi.org/10.1016/j.pain.2013.05.040>
16. Kuang, Y., Wu, Z., Xia, R., Li, X., Liu, J., Dai, Y., Wang, D., and Chen, S. "Phase Lag Index of Resting-State EEG for Identification of Mild Cognitive Impairment Patients with Type 2 Diabetes." *Brain Sciences* 12, no. 10 (October 17, 2022): 1399. <https://doi.org/10.3390/brainsci12101399>
17. Lau, Z. J., Pham, T., Chen, S. H. A., and Makowski, D. "Brain Entropy, Fractal Dimensions and Predictability: A Review of Complexity Measures for EEG in Healthy and Neuropsychiatric Populations." *European Journal of Neuroscience* 56, no. 7 (October 2022): 5047–5069. <https://doi.org/10.1111/ejn.15800>.
18. Le Van Quyen, M., Foucher, J., Lachaux, J., Rodriguez, E., Lutz, A., Martinerie, J., and Varela, F. J. "Comparison of Hilbert Transform and Wavelet Methods for the Analysis of Neuronal Synchrony." *Journal of Neuroscience Methods* 111, no. 2 (October 30, 2001): 83–98. [https://doi.org/10.1016/S0165-0270\(01\)00372-7](https://doi.org/10.1016/S0165-0270(01)00372-7)
19. Li, M., Gao, Y., Ding, Z., and Gore, J. C. "Power Spectra Reveal Distinct BOLD Resting-State Time Courses in White Matter." *Proceedings of the National Academy of Sciences of the United States of America* 118, no. 44 (November 2, 2021): e2103104118. <https://doi.org/10.1073/pnas.2103104118>.
20. Li, X., Wu, Y., Wei, M., Guo, Y., Yu, Z., Wang, H., Li, Z., and Fan, H. "A Novel Index of Functional Connectivity: Phase Lag Based on Wilcoxon Signed Rank Test." *Cognitive Neurodynamics* 15, no. 4 (August 2021): 621–636. <https://doi.org/10.1007/s11571-020-09646-x>
21. Lutkenhoff, E. S., Nigri, A., Rossi Sebastiano, D., Sattin, D., Visani, E., Rosazza, C., D'Incerti, L., Bruzzone, M. G., Franceschetti, S., Leonardi, M., Ferraro, S., and Monti, M. M. "EEG Power Spectra and Subcortical Pathology in Chronic Disorders of Consciousness." *Psychological Medicine* 52, no. 8 (June 2022): 1491–1500. <https://doi.org/10.1017/S003329172000330X>.
22. Xu, J., and Guo, W. "Imaginary Part of Timelike Entanglement Entropy." *Journal of High Energy Physics* 2025, no. 94 (2025). [https://doi.org/10.1007/JHEP02\(2025\)094](https://doi.org/10.1007/JHEP02(2025)094).
23. Xue, S., Guo, J., Li, P., and others. "Quantification of Resource Theory of Imaginarity." *Quantum Information Processing* 20, no. 383 (2021). <https://doi.org/10.1007/s11128-021-03324-5>.
24. Pernice, R., Zanetti, M., Nollo, G., De Cecco, M., Busacca, A., and Faes, L. "Mutual Information Analysis of Brain-Body Interactions during Different Levels of Mental Stress." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (July 2019): 6176–6179. <https://doi.org/10.1109/EMBC.2019.8856711>
25. Phillips, N. S., Rao, V., Kmetz, L., Vela, R., Medick, S., Krull, K., and Kesler, S. R. "Changes in Brain Functional and Effective Connectivity After Treatment for Breast Cancer and Implications for Intervention Targets." *Brain Connectivity* 12, no. 4 (May 2022): 385–397. <https://doi.org/10.1089/brain.2021.0049>.

26. Powanwe, A. S., and Longtin, A. "Mutual Information Resonances in Delay-Coupled Limit Cycle and Quasi-Cycle Brain Rhythms." *Biological Cybernetics* 116, no. 2 (April 2022): 129–146. <https://doi.org/10.1007/s00422-022-00932-x>

27. Redwan, S. M., Uddin, M. P., Ulhaq, A., Sharif, M. I., and Krishnamoorthy, G. "Power Spectral Density-Based Resting-State EEG Classification of First-Episode Psychosis." *Scientific Reports* 14, no. 1 (July 2, 2024): 15154. <https://doi.org/10.1038/s41598-024-66110-0>

28. Sadeghijam, M., Talebian, S., Mohsen, S., Akbari, M., and Pourbakht, A. "Shannon Entropy Measures for EEG Signals in Tinnitus." *Neuroscience Letters* 762 (September 25, 2021): 136153. <https://doi.org/10.1016/j.neulet.2021.136153>

29. Sanchez Bornot, J. M., Wong-Lin, K., Ahmad, A. L., and Prasad, G. "Robust EEG/MEG Based Functional Connectivity with the Envelope of the Imaginary Coherence: Sensor Space Analysis." *Brain Topography* 31, no. 6 (November 2018): 895–916. <https://doi.org/10.1007/s10548-018-0640-0>

30. Sander, T. H., Bock, A., Leistner, S., Kuhn, A., and Trahms, L. "Coherence and Imaginary Part of Coherency Identifies Cortico-Muscular and Cortico-Thalamic Coupling." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2010): 1714–1717. <https://doi.org/10.1109/IEMBS.2010.5626851>

31. Shiju, S., and Sriram, K. "Hilbert Transform-Based Time-Series Analysis of the Circadian Gene Regulatory Network." *IET Systems Biology* 13, no. 4 (August 2019): 159–168. <https://doi.org/10.1049/iet-syb.2018.5088>

32. Silva, P. H. R. D., Secchinato, K. F., Rondinoni, C., and Leoni, R. F. "Brain Structural-Functional Connectivity Relationship Underlying the Information Processing Speed." *Brain Connectivity* 10, no. 3 (April 2020): 143–154. <https://doi.org/10.1089/brain.2019.0726>

33. Snaauw, G., Sasdelli, M., Maicas, G., Lau, S., Verjans, J., Jenkinson, M., and Carneiro, G. "Mutual Information Neural Estimation for Unsupervised Multi-Modal Registration of Brain Images." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (July 2022): 3510–3513. <https://doi.org/10.1109/EMBC48229.2022.9871220>

34. Stein, A., Thorstensen, J. R., Ho, J. M., Ashley, D. P., Iyer, K. K., and Barlow, K. M. "Attention Please! Unravelling the Link Between Brain Network Connectivity and Cognitive Attention Following Acquired Brain Injury: A Systematic Review of Structural and Functional Measures." *Brain Connectivity* 14, no. 1 (February 2024): 4–38. <https://doi.org/10.1089/brain.2023.0067>

35. Sun, H., Lin, F., Wu, X., Zhang, T., and Li, J. "Normalized Mutual Information of fNIRS Signals as a Measure for Assessing Typical and Atypical Brain Activity." *Journal of Biophotonics* 16, no. 6 (June 2023): e202200369. <https://doi.org/10.1002/jbio.202200369>

36. Tozzi, A., E. Bormashenko, and N. Jausovec. "Topology of EEG Wave Fronts." *Cognitive Neurodynamics* 15 (2021a): 887–96. <https://doi.org/10.1007/s11571-021-09668-z>

37. Tozzi, A., J. F. Peters, N. Jausovec, A. P. H. Don, S. Ramanna, I. Legchenkova, and E. Bormashenko. "Nervous Activity of the Brain in Five Dimensions." *Biophysica* 1, no. 1 (2021b): 38–47. <https://doi.org/10.3390/biophysica1010004>

38. Tian, W., Zhao, D., Ding, J., Zhan, S., Zhang, Y., Etkin, A., Wu, W., and Yuan, T. F. "An Electroencephalographic Signature Predicts Craving for Methamphetamine." *Cell Reports Medicine* 5, no. 1 (January 16, 2024): 101347. <https://doi.org/10.1016/j.xcrm.2023.101347>

39. Yan, H., Johnston, J. F., Cahn, S. B., King, M. C., and Mochrie, S. G. J. "Multiplexed Fluctuation–Dissipation–Theorem Calibration of Optical Tweezers Inside Living Cells." *Review of Scientific Instruments* 88, no. 11 (November 2017): 113112. <https://doi.org/10.1063/1.5012782>

40. Yang, S., Xu, G., Zhou, X., Li, J., Kong, X., Zhou, C., Fan, H., Chen, J., and Qiu, C. W. "Hierarchical Bound States in Heat Transport." *Proceedings of the National Academy of Sciences of the United States of America* 121, no. 38 (September 17, 2024): e2412031121. <https://doi.org/10.1073/pnas.2412031121>

41. Wang, Z., Li, Y., Childress, A. R., and Detre, J. A. "Brain Entropy Mapping Using fMRI." *PLoS One* 9, no. 3 (March 21, 2014): e89948. <https://doi.org/10.1371/journal.pone.0089948>

42. Wang, Z., Zhao, Y., He, Y., and Zhang, J. "Phase Lag Index-Based Graph Attention Networks for Detecting Driving Fatigue." *Review of Scientific Instruments* 92, no. 9 (September 1, 2021): 094105. <https://doi.org/10.1063/5.0056139>



43. Wu, Y., Zhou, Y., and Song, M. "Classification of Patients with AD from Healthy Controls Using Entropy-Based Measures of Causality Brain Networks." *Journal of Neuroscience Methods* 361 (September 1, 2021): 109265. <https://doi.org/10.1016/j.jneumeth.2021.109265>.
44. Wu, K. D., Kondra, T. V., Scandolo, C. M., and others. "Resource Theory of Imaginarity in Distributed Scenarios." *Communications Physics* 7, no. 171 (2024). <https://doi.org/10.1038/s42005-024-01649-y>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

**Topological reconstruction of Rubin’s multiple imputation via coarse proximity,
Seifert–van Kampen gluing and Hurewicz invariants**

Arturo Tozzi (corresponding author)
ASL Napoli 1 Centro, Distretto 27, Naples, Italy
Via Comunale del Principe 13/a 80145
tozziarturo@libero.it

ABSTRACT

Rubin’s multiple imputation (MI) generates plausible data completions to account for uncertainty and statistical variability but provides little insight into their global organization. We introduce a topological reconstruction approach that complements MI by examining the structure of the ensemble of completed datasets. Individual imputations are represented as points in a reconstruction space whose coordinates summarize statistical properties of each completed dataset. Then, concepts from coarse geometry and algebraic topology are used to characterize relationships among alternative imputations at multiple scales. Coarse proximity (CP) defines large-scale neighborhoods among admissible reconstructions, generating graphs in which nodes represent completed datasets and edges connect sufficiently similar imputations. Seifert–van Kampen gluing provides conceptual interpretation of how locally related reconstructions assemble into globally coherent structures, whereas Hurewicz-type invariants quantify nontrivial connectivity patterns persisting across reconstruction scales. Synthetic multivariate biomedical datasets representing adult cardiometabolic cohorts were generated with controlled missingness levels. Multiple stochastic imputations were produced for each missingness condition, projected into the reconstruction space and analyzed through CP graphs, connected components, cycle descriptors and scale-dependent topological measures. We found that MI generated structured spaces with distinct connectivity patterns rather than homogeneous clouds of solutions. Topological descriptors were stable despite local numerical variability, whereas increasing missingness produced transitions in reconstruction-space connectivity together with progressive deterioration of reconstruction accuracy. Our approach could be applied to biological networks, neuroscience, systems medicine, ecological modeling, social networks and other domains in which large-scale structural organization contributes to reliable inference.

KEYWORDS: Homology; manifold; persistence; network; uncertainty.

INTRODUCTION

Missing data are a pervasive challenge in scientific research, affecting statistical inference, parameter estimation and predictive modelling. Over the past decades, numerous strategies have been developed to address incomplete observations, including case deletion, single imputation methods and likelihood-based procedures (Javanbakht et al. 2022; Raevskiy et al. 2023; Yu et al. 2026). Among these approaches, Rubin’s multiple imputation has become one of the most widely adopted solutions because it generates several plausible completions of an incomplete dataset and propagates uncertainty through repeated analyses (Giganti and Shepherd 2020; Yu, Liu, and Peace 2020; Ségalas et al. 2023; Shu et al. 2023). This method has been successfully applied in epidemiology, clinical research, social sciences, economics and many other fields. Despite its broad use, the information extracted from multiple imputation is summarized through conventional statistical descriptors such as means, variances, confidence intervals and prediction errors. These quantities provide limited information regarding the global organization of the collection of admissible completions generated by the imputation process. Indeed, different sets of imputations may exhibit similar statistical summaries while possessing markedly different structural relationships, whereas geometric and topological relationships among these realizations are still largely unexplored. Consequently, these approaches may overlook large-scale organizational properties emerging from interactions among alternative completions.

We introduce a topological reconstruction approach that complements Rubin’s multiple imputation by examining the structure of the space formed by admissible completed datasets. We combine concepts from coarse geometry and algebraic topology to characterize relationships among alternative imputations at multiple scales. Individual imputations are represented as points in a reconstruction space whose coordinates summarize relevant statistical properties of the completed datasets. Coarse proximity is used to define large-scale neighborhoods among imputations without relying exclusively on local metric information (Grzegorzolka and Siegert 2023; Shi and Yao 2024). These neighborhoods generate graphs in which nodes correspond to completed datasets and edges connect sufficiently similar reconstructions. The resulting structures are analyzed through topological descriptors inspired by connectedness and cycle formation. In this context, Seifert–van Kampen gluing could provide a conceptual interpretation of how locally related reconstructions assemble into globally coherent structures (Ramos, de Veras, and de Queiroz 2025), while Hurewicz-type invariants could deliver quantitative descriptions of nontrivial connectivity patterns persisting across scales (Arkowitz and Curjel 1964; Sakai 2020; Pilgrim 2025). Rather than focusing exclusively on individual completed datasets, attention is directed toward

the organization of the entire ensemble of admissible solutions. To investigate the behavior of our approach, we performed simulations on synthetic multivariate biomedical datasets with controlled missingness levels. Multiple stochastic imputations were generated for each missingness condition and their global organization was examined through reconstruction-space projections, proximity graphs, connectivity measures and scale-dependent topological descriptors. We expected that increasing missingness would alter not only reconstruction accuracy, but also the large-scale structure of admissible completions, potentially producing transitions in the connectivity of the reconstruction space.

We will proceed as follows. First, we describe the simulated datasets, missingness generation procedures, multiple-imputation strategy and topological reconstruction methods. Then, we present the simulation results and quantify the evolution of reconstruction-space topology across missingness levels. Finally, we discuss the implications, limitations and interpretation of our proposed approach.

METHODS

We studied a simulated cardiometabolic dataset to investigate whether multiple imputation produces not only uncertainty in individual reconstructed values, but also a space of admissible data completions that can be characterized through topological descriptors. We used stochastic regression-based multiple imputation, dimensional reduction, coarse proximity graphs, connected-component analysis, cycle counts, Seifert–van Kampen-inspired gluing, Hurewicz-type descriptors and scale-dependent topological summaries. We aimed to identify an experimentally discriminable signature: a transition in the topology of the imputation space as missingness increased, expressed by changes in connected components, independent cycles, reconstruction error and classification variability.

Synthetic cohort. A virtual cohort of 240 subjects was generated to mimic a compact biomedical dataset with continuous clinical variables representative of adult cardiometabolic cohorts. Each subject was assigned age (years), body mass index (kg/m²), fasting glucose (mg/dL), systolic blood pressure (mmHg), low-density lipoprotein cholesterol (mg/dL), high-density lipoprotein cholesterol (mg/dL) and C-reactive protein (mg/L) (Yu et al. 2018; Chung 2019; Çubukçu and Topcu 2022; Oh et al. 2022; Qiao et al. 2023; Mitani et al. 2023; Solianik et al. 2023; Tsang et al. 2023; Chin, Potter, and Friedl 2024). Variable ranges and dependence structures were chosen to be consistent with reported distributions of established cardiometabolic risk factors in population-based studies (D’Agostino et al. 2008; Marino et al. 2014). The simulated observations were generated from correlated latent cardiometabolic risk variables to reproduce the well-documented associations among adiposity, systemic inflammation, glycemic status, blood pressure and lipid metabolism. For each virtual subject i , a latent risk variable r_i was defined as a weighted combination of age, body mass index and C-reactive protein, with an added Gaussian error term:

$$r_i = 0.04(a_i - 55) + 0.12(b_i - 27) + 0.08(c_i - 3) + \varepsilon_i,$$

where a_i denotes age, b_i body mass index, c_i C-reactive protein and ε_i random variability. Clinical variables were then generated as functions of r_i plus independent noise. For example, glucose was generated as:

$$g_i = 92 + 10r_i + 0.35(b_i - 27) + \eta_i,$$

and systolic blood pressure as:

$$s_i = 118 + 8r_i + 0.45(a_i - 55) + \xi_i.$$

All simulated variables were clipped to clinically plausible ranges.

Missingness design. Controlled missingness was introduced at six levels: 5%, 10%, 20%, 30%, 40% and 50%. Missingness was not imposed as purely random deletion. Instead, glucose, systolic blood pressure and low-density lipoprotein cholesterol were made more likely to be missing in subjects with higher latent risk. This choice generated a missing-at-random-like structure in which the probability of observing a missing value depended on other measurable subject properties. If m_{ij} denotes the missingness indicator for subject i and variable j , the probability of missingness was modeled as:

$$P(m_{ij} = 1) = \rho(0.35 + 1.3q_i),$$

where ρ is the nominal missingness level and q_i is the normalized latent risk score. For variables less directly associated with cardiometabolic risk, lower missingness probabilities were used. This procedure produced incomplete matrices:

$$X^{(\rho)} = \{x_{ij}^{(\rho)}\},$$

where some entries were unobserved. The complete matrix X was retained only for simulation validation, allowing reconstruction error to be calculated after imputation. In turn, the incomplete matrix was the object analyzed by the imputation procedure.

Multiple imputation. For each missingness level, forty imputed datasets were generated. This ensemble size was chosen to provide stable estimates of between-imputation variability while yielding a sufficiently dense sampling of the space of admissible data completions for the computation of our topological descriptors. Missing entries were first replaced by median values to obtain an initial complete matrix, then each variable with missing values was imputed by stochastic regression using the remaining variables as predictors. For variable j , observed and missing rows were separated. A regression model was fitted on observed entries:

$$x_{ij} = \alpha_j + \sum_{k \neq j} \beta_{jk} x_{ik} + e_{ij}.$$

For rows with missing x_{ij} , the conditional mean was estimated as:

$$\hat{x}_{ij} = \hat{\alpha}_j + \sum_{k \neq j} \hat{\beta}_{jk} x_{ik}.$$

To avoid producing identical deterministic completions, a stochastic residual term was added:

$$x_{ij}^* = \hat{x}_{ij} + z_{ij} \hat{\sigma}_{ij},$$

where z_{ij} was sampled from a standard normal distribution and $\hat{\sigma}_{ij}$ represented predictive uncertainty. This procedure yielded forty imputed datasets for each missingness level:

$$\mathcal{I}_\rho = \{X_\rho^{(1)}, X_\rho^{(2)}, \dots, X_\rho^{(40)}\}.$$

Each completed dataset retained the same units as the original variables. Values were constrained to remain within plausible clinical intervals.

Reconstruction space. Each completed dataset was summarized by clinically interpretable statistics. For imputation k at missingness level ρ , we computed the mean fasting glucose, mean systolic blood pressure, mean low-density lipoprotein cholesterol, mean body mass index and estimated metabolic-syndrome prevalence. This produced a summary vector:

$$v_{\rho k} = (\bar{g}_{\rho k}, \bar{s}_{\rho k}, \bar{\ell}_{\rho k}, \bar{b}_{\rho k}, p_{\rho k}).$$

The set of vectors formed a reconstruction space of admissible completions. Standardization was applied to remove unit dominance:

$$z_{\rho k j} = \frac{v_{\rho k j} - \mu_j}{\sigma_j}.$$

Principal component analysis was then used only as a visualization and projection method. Given the standardized matrix Z , the covariance matrix was computed as:

$$C = \frac{1}{n-1} Z^T Z.$$

Eigenvectors of C defined orthogonal reconstruction axes. Each imputation was projected into the first two axes:

$$y_{\rho k} = Z_{\rho k} W_2.$$

The resulting two-dimensional coordinates represented each completed dataset as one point in reconstruction space.

Coarse proximity. Coarse proximity was used to define large-scale similarity among imputed datasets (Grzegorzolka and Siegert 2019, 2023). Instead of requiring exact local equality, two completions were considered related when their projected clinical profiles were sufficiently close at a chosen scale. For imputations k and l , Euclidean distance in standardized reconstruction space was calculated as:

$$d_\rho(k, l) = \sqrt{(y_{\rho k 1} - y_{\rho l 1})^2 + (y_{\rho k 2} - y_{\rho l 2})^2}.$$

A coarse proximity relation was defined by a radius ϵ :

$$k \sim_\epsilon l \text{ if and only if } d_\rho(k, l) \leq \epsilon.$$

This relation generated a graph:

$$G_{\rho, \epsilon} = (V_\rho, E_{\rho, \epsilon}),$$

where vertices were imputed datasets and edges joined pairs of sufficiently similar completions. Coarse proximity was used because multiple imputations can differ in small local numerical details while still belonging to the same broad

reconstruction region. Therefore, the proximity graph represented the large-scale organization of admissible completions rather than individual variable-level differences alone.

Connectedness. Connected components were computed from the adjacency matrix of the coarse proximity graph (Prolubnikov 2023). The adjacency matrix was defined as:

$$A_{kl} = \begin{cases} 1, & d_p(k, l) \leq \epsilon, \\ 0, & d_p(k, l) > \epsilon. \end{cases}$$

The number of connected components, denoted β_0 , quantified fragmentation of the imputation space. A high β_0 indicated that admissible completions separated into multiple disconnected families. A low β_0 indicated that completions formed a more coherent reconstruction region. The first cycle count was calculated through the graph-theoretic relation:

$$\beta_1 = |E| - |V| + \beta_0,$$

where $|E|$ is the number of edges and $|V|$ the number of vertices. This quantity measures independent cycles in the proximity graph. In the present setting, a cycle indicates that several imputations are mutually connected through alternative paths, producing a non-tree-like organization of the completion space. Both β_0 and β_1 were evaluated across missingness levels and across proximity radii. This allowed us to distinguish numerical dispersion from topological reorganization.

Gluing principle. Seifert–van Kampen theory was used as a guiding topological principle for interpreting how local reconstruction regions combine into a global imputation space. In algebraic topology, the theorem describes how the fundamental group of a space can be obtained from two overlapping subspaces when their intersection is sufficiently connected. In simplified terms, if a space X is covered by two regions U and V , then global loop structure can be reconstructed from the loop structures of U , V and $U \cap V$. Symbolically, the classical idea is expressed as:

$$\pi_1(X) = \pi_1(U) * \pi_1(V) / \langle \text{relations from } U \cap V \rangle.$$

In our simulation, local groups of similar imputations were interpreted as patches. Edges among patches represented overlaps or coarse compatibility. We did not perform an explicit computation of fundamental groups. Instead, we used the gluing principle of the Seifert–van Kampen theorem operationally, interpreting the progressive merging of previously disconnected imputation neighborhoods into larger connected structures as the coarse-proximity radius increased. The decreasing component count measured this gluing process. The aim was to observe whether local imputational neighborhoods assembled gradually or through abrupt transitions.

Hurewicz descriptor. The Hurewicz theorem links homotopy information, which concerns continuous deformation of paths and loops, with homology information, which counts holes and cycles through algebraic invariants. In its classical form, under suitable connectivity assumptions, the first nontrivial homotopy group maps to the corresponding homology group. The conceptual bridge is:

$$\pi_n(X) \rightarrow H_n(X).$$

In the simulation, the full homotopy structure of the imputation space was not computed. Instead, a Hurewicz-like descriptor was defined by tracking graph cycles as homological signatures of nontrivial connectivity. The first cycle count β_1 was used as an accessible representation of loop-like organization among admissible completions. This choice was appropriate for a finite set of imputed datasets represented as a proximity graph. A tree-like graph has $\beta_1 = 0$, whereas additional compatible paths among imputations increase β_1 . Therefore, changes in β_1 across scales described whether the reconstruction space remained simple, fragmented or loop-rich. The descriptor was not interpreted as a biological invariant, but as a mathematical summary of the imputation ensemble.

Classification metric. To relate reconstruction topology to an interpretable simulated clinical outcome, each completed dataset was used to estimate metabolic-syndrome-like prevalence. A subject was classified as positive when at least two of the following simulated criteria were met: body mass index of at least 30 kg/m², fasting glucose of at least 126 mg/dL, systolic blood pressure of at least 130 mmHg, low-density lipoprotein cholesterol of at least 160 mg/dL or high-density lipoprotein cholesterol below 40 mg/dL. For subject i , the classification score was:

$$c_i = \mathbf{1}(b_i \geq 30) + \mathbf{1}(g_i \geq 126) + \mathbf{1}(s_i \geq 130) + \mathbf{1}(\ell_i \geq 160) + \mathbf{1}(h_i < 40).$$

The binary classification was:

$$q_i = \begin{cases} 1, & c_i \geq 2, \\ 0, & c_i < 2. \end{cases}$$

For each imputation k , prevalence was:

$$p_{\rho k} = 100 \times \frac{1}{N} \sum_{i=1}^N q_{\rho ki}.$$

The standard deviation of $p_{\rho k}$ across forty imputations quantified classification variability induced by missing data.

Error estimation. Because our simulation retained the original complete dataset, reconstruction accuracy could be calculated directly. For each missingness level, the average imputed value for each subject and variable was computed across forty completions. For glucose, the root mean square error was:

$$\text{RMSE}_g = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{g}_i - g_i)^2}.$$

For systolic blood pressure, the corresponding expression was:

$$\text{RMSE}_s = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{s}_i - s_i)^2}.$$

These errors were reported in mg/dL and mmHg. The error analysis was kept separate from the topological analysis. Reconstruction error measured numerical deviation from the known simulated truth, whereas β_0 , β_1 and proximity graphs measured the shape of the ensemble of admissible completions. This separation allowed our simulation to compare conventional accuracy summaries with topological summaries of imputation structure.

Software tools. All simulations, imputations, numerical analyses and figures were generated in Python. NumPy was used for numerical array operations, random sampling and mathematical transformations. Pandas was used to organize the simulated clinical dataset and export the summary table. Scikit-learn was used for median initialization, Bayesian ridge regression and principal component analysis. SciPy was used to compute pairwise distances and connected components of proximity graphs.

RESULTS

We report here the behavior of stochastic multiple imputation and topological reconstruction across progressively increasing levels of missing data. Our analyses focused on the geometry of admissible completions, coarse-proximity connectivity, cycle formation, reconstruction accuracy and variability of a simulated clinical classification. Quantitative results are presented for missingness levels ranging from 5% to 50%, with forty completed datasets generated for each condition.

Reconstruction topology. Throughout the investigated missingness range, the space of admissible imputations exhibited an organized rather than random structure. Projection of the forty imputations generated at each missingness level into the reconstruction space revealed partially overlapping but distinguishable clusters corresponding to different percentages of missing data (Figure 1, upper-left inset). The number of connected components varied between 3 and 9 across the investigated missingness levels (Figure 1, upper-right inset). Six connected components were observed at 5% and 10% missingness, followed by seven at 20%, three at 30%, nine at 40% and five at 50%, indicating a non-monotonic evolution of the connectivity of the reconstruction space. The first cycle count varied between 56 and 76 across the investigated missingness levels, reaching 56 at 5% missingness and a maximum of 76 at 40% missingness. A Spearman correlation analysis between missingness percentage and cycle count did not reveal a significant monotonic trend ($\rho = 0.314$, $p = 0.544$), indicating that topological complexity evolved nonlinearly rather than increasing proportionally with missingness. The dataset displayed positive associations among body mass index, glucose, systolic blood pressure and low-density lipoprotein cholesterol, whereas high-density lipoprotein cholesterol showed negative correlations with most cardiometabolic variables (Figure 1, lower-right inset). Examination of the coarse-proximity graph at 40% missingness revealed multiple densely connected local neighborhoods linked by a smaller number of bridging edges (Figure 2, left inset). The coexistence of connected clusters and numerous cycles indicates that admissible completions formed structured families rather than isolated realizations, suggesting that reconstruction uncertainty displayed a measurable topological organization beyond simple numerical dispersion.

Accuracy and transitions. Reconstruction error increased progressively as larger fractions of data were removed (Figure 1, lower-left inset). Mean glucose root mean square error rose from 2.11 mg/dL at 5% missingness to 8.66 mg/dL at 50% missingness. Mean systolic blood pressure root mean square error increased from 1.94 mmHg to 9.03 mmHg across the

same interval. Spearman correlation analysis demonstrated a perfect monotonic relationship between missingness and reconstruction error for both glucose ($\rho = 1.000$, $p < 0.001$) and systolic blood pressure ($\rho = 1.000$, $p < 0.001$). Variability of estimated metabolic-syndrome prevalence also increased with missingness (Figure 1, lower-middle inset). The standard deviation across imputations rose from 0.54% at 5% missingness to 1.54% at 50% missingness. A significant positive Spearman association was observed between missingness and classification variability ($\rho = 0.829$, $p = 0.042$). Scale-dependent analysis of the proximity graph provided additional information regarding the organization of admissible completions (Figure 2, upper-right inset). As the coarse-proximity radius increased from 0.25 to 1.80, the number of connected components decreased from 31 to a single connected structure, indicating progressive gluing of initially separated reconstruction regions. Simultaneously, the number of independent cycles increased from 5 to 264 (Figure 2, lower-right inset), showing the emergence of increasingly rich connectivity patterns as additional links were established among imputations. The transition from numerous disconnected groups to a single connected object occurred over a relatively narrow interval of proximity radii, indicating that the global organization of admissible completions depended strongly on the observational scale used to define similarity relationships. Together, the numerical and topological analyses showed that increasing missingness altered both reconstruction fidelity and the large-scale structure of the imputation ensemble.

Overall, our results show that multiple imputation gives rise to structured spaces of admissible data completions whose organization can be quantified independently of conventional reconstruction accuracy measures. Reconstruction error and classification variability increased with missingness, whereas connected components and cycle counts evolved nonlinearly. Scale-dependent coarse-proximity analysis revealed transitions from fragmented local neighborhoods to globally connected reconstruction spaces. These findings indicate that conventional statistical summaries can be complemented by descriptors of large-scale connectivity and topological organization.

Topological multiple imputation in a simulated cardiometabolic cohort

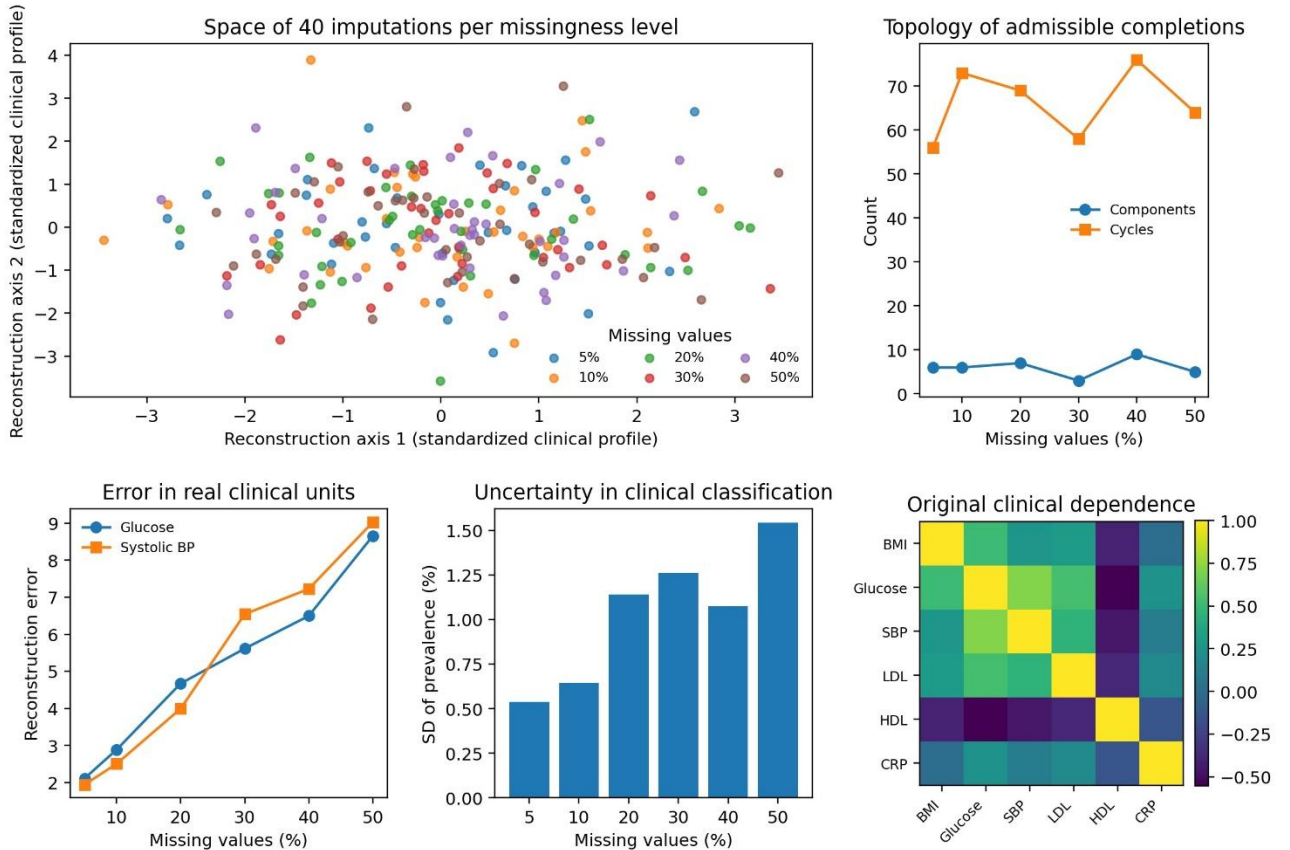


Figure 1. Topological multiple imputation in a simulated cardiometabolic cohort under increasing missingness. The figure summarizes the behavior of our topological reconstruction approach in a simulated cohort of 240 virtual subjects characterized by age (years), body mass index (kg/m^2), fasting glucose (mg/dL), systolic blood pressure (mmHg), low-density lipoprotein cholesterol (mg/dL), high-density lipoprotein cholesterol (mg/dL) and C-reactive protein concentration (mg/L). Missing values were introduced at rates ranging from 5% to 50% and forty stochastic imputations were generated for each missingness level, yielding ensembles of admissible completed datasets whose large-scale organization was subsequently analyzed through topological descriptors.

Upper-left inset: reconstruction space of all admissible imputations after dimensionality reduction. Each point corresponds to one completed dataset and is positioned according to its overall clinical profile. Colors indicate the proportion of missing values. The distribution of points provides a visual representation of the geometry of the imputation space and the degree of separation among alternative reconstructions.

Upper-right inset: topological descriptors derived from coarse-proximity relationships among admissible imputations. The number of connected components reflects the fragmentation of the reconstruction space, whereas the number of independent cycles represents Hurewicz-like topological invariants characterizing nontrivial connectivity patterns.

Lower-left inset: reconstruction error expressed in clinical units. Curves correspond to the root mean square reconstruction error of fasting glucose (mg/dL) and systolic blood pressure (mmHg) relative to the complete dataset. Both variables exhibited a progressive increase in reconstruction error with increasing missingness, indicating a monotonic deterioration in numerical reconstruction accuracy

Lower-middle inset: Standard deviation of estimated metabolic-syndrome prevalence across the forty imputations, providing a measure of classification uncertainty generated by missing data. The standard deviation increased from 0.54% at 5% missingness to 1.54% at 50% missingness, indicating progressively greater variability in prevalence estimates as the proportion of missing observations increased.

Lower-right inset: Pearson correlation matrix of the original complete dataset, illustrating the dependence structure among the simulated clinical variables. Positive correlations are observed among body mass index, fasting glucose, systolic blood pressure, low-density lipoprotein cholesterol and C-reactive protein, whereas high-density lipoprotein cholesterol exhibits predominantly negative correlations with the remaining variables. Pairwise correlation coefficients ranged from approximately -0.45 to $+0.62$, indicating a moderately correlated multivariate structure that served as the statistical basis for stochastic imputation and influenced the geometry of the admissible reconstruction space.

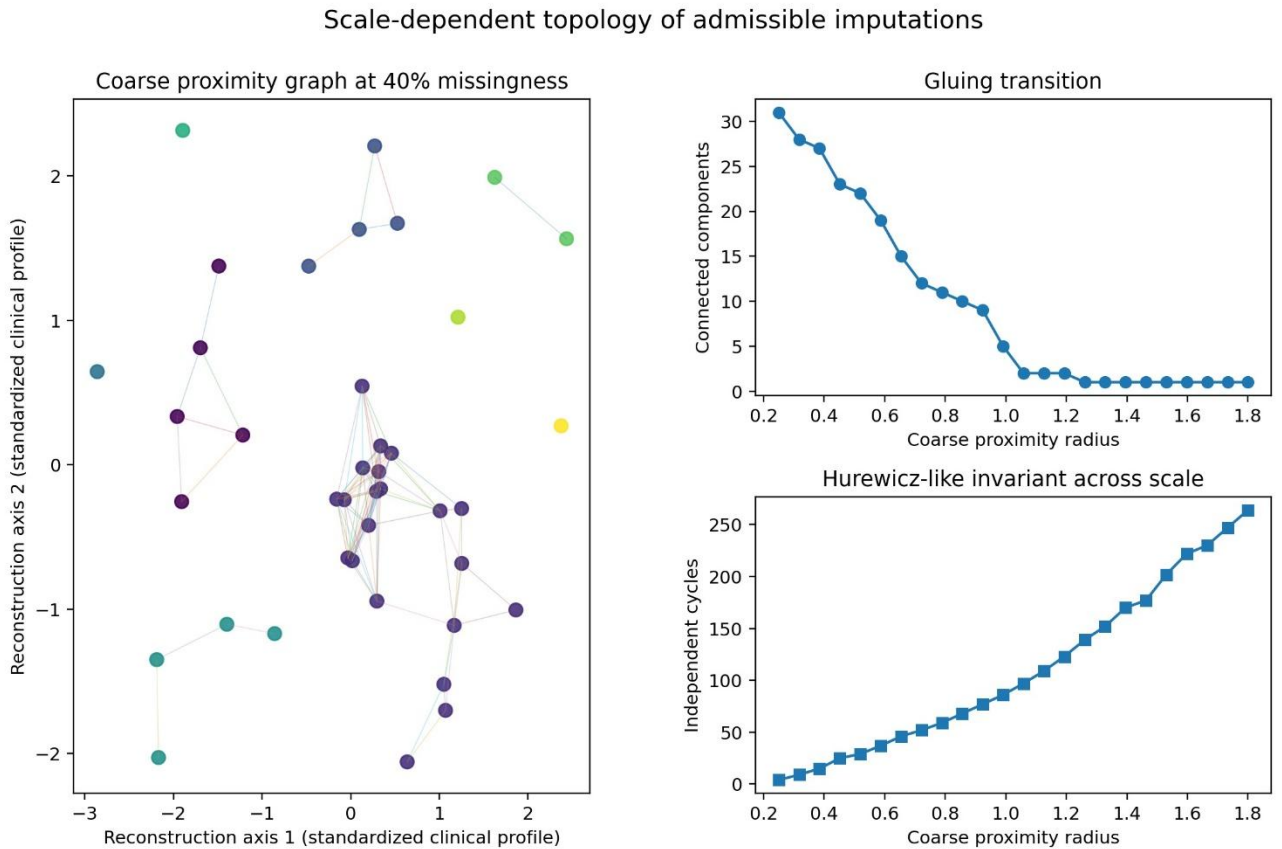


Figure 2. Scale-dependent topology of admissible imputations and coarse-proximity reconstruction. The figure illustrates how the global organization of admissible imputations evolves as the scale of coarse proximity is varied. Results are shown for the scenario with 40% missing values.

Large left inset: coarse-proximity graph constructed from forty completed datasets projected into the reconstruction space. Each node represents an individual imputed dataset and edges connect pairs of imputations separated by less than a predefined proximity radius. Node positions correspond to standardized clinical reconstruction coordinates. Clusters identify groups of mutually similar imputations, whereas isolated nodes represent alternative reconstructions that remain weakly connected to the dominant solution families.

Upper-right inset: gluing transition. As the coarse-proximity radius increases, initially disconnected groups of admissible imputations progressively merge into larger connected structures. The decreasing number of connected components quantifies the transition from fragmented local reconstructions toward a globally integrated reconstruction space, analogous to the assembly of local patches into a coherent global object.

Lower-right inset: evolution of the Hurewicz-like invariant across increasing proximity scales. The number of independent cycles increases as additional connections are established among imputations, revealing the emergence of nontrivial topological organization within the reconstruction space.

CONCLUSIONS

We asked whether collections of datasets generated by Rubin’s multiple imputation contain an internal large-scale organization independent of conventional statistical uncertainty measures. Specifically, we investigated whether admissible completions form structured spaces whose properties can be described through coarse proximity, connectedness and topological invariants. Our simulations compared multiple missingness levels in a synthetic cardiometabolic cohort while preserving known ground truth values, allowing simultaneous assessment of reconstruction accuracy and topological organization. We found that increasing missingness modified both the numerical reconstruction performance and the geometry of the admissible completion space. Reconstruction errors increased progressively with missingness, while the organization of imputations exhibited changes in connected components, cycle counts and scale-dependent connectivity patterns. The resulting spaces were neither fully fragmented nor completely homogeneous, but displayed structured families of admissible reconstructions connected through identifiable proximity relationships. Therefore, uncertainty associated with missing data could be examined not only through distributions of estimated values, but also through relationships among alternative completed datasets. Our results suggest that topological descriptors and conventional statistical summaries capture different aspects of the reconstruction process: whereas reconstruction error quantifies deviation from known values, topological descriptors characterize how alternative completions are arranged relative to one another.

Our approach differs from existing missing-data methodologies in several respects. Techniques such as listwise deletion, mean substitution, expectation-maximization algorithms, inverse probability weighting and standard multiple imputation focus primarily on parameter estimation, prediction or uncertainty quantification (Lieberman-Betz et al. 2014; Chan 2017; De Raadt et al. 2019; Chen et al. 2024; Niloofar, Aghdam, and Eslahchi 2024). In contrast, our approach analyzes relationships among completed datasets themselves. Unlike principal component analysis, cluster analysis and conventional distance-based visualization methods, which primarily describe the geometric distribution of observations (Martins et al. 2019; Mi et al. 2019; Ben Salem and Ben Abdelaziz 2021; Jang 2021), our procedure quantifies the topology of the imputation space through measures of connectedness and independent cycle structure. Compared with Bayesian posterior summaries, which describe probability distributions of parameters (Alquier 2020; Huber, Georgia, and Finley 2023; van Dijk et al. 2023; Golchi and Willard 2024), our approach investigates the organization of admissible reconstructions generated by those distributions. Compared with sensitivity analyses based on repeated imputations (Cramer et al. 2015; Mehrotra, Liu, and Permutt 2017; Tompsett et al. 2020), our descriptors are capable of distinguishing situations with comparable numerical variability but different global arrangements of solutions. The use of coarse proximity allows examination of large-scale relationships without requiring strict local similarity, while Seifert–van Kampen-inspired gluing and Hurewicz-type descriptors provide quantitative summaries of connectivity across scales (Bhardwaj and Tyagi 2019; Christensen and Scoccola 2023).

Our study has limitations. First, the analyses were conducted entirely on simulated data. Second, the simulated cohort contained a restricted set of continuous variables with predefined correlation structures and did not include categorical variables, longitudinal observations or multimodal measurements. Third, the missingness mechanism was designed to resemble clinically plausible conditions but cannot represent the full diversity of real-world missing-data processes. Fourth, the topological descriptors were derived from finite graphs constructed in reduced-dimensional reconstruction spaces; alternative distance measures, dimensional reductions or graph-construction procedures could produce different quantitative values. Fifth, the Seifert–van Kampen and Hurewicz concepts were used as operational tools for describing connectivity and cycle structure rather than as complete algebraic-topological calculations. Sources of potential bias include the selected proximity thresholds, the stochastic characteristics of the imputation algorithm and the peculiar summary statistics used to define reconstruction coordinates. Unresolved questions concern the stability of topological descriptors under different imputation models, the behavior of the method in very high-dimensional datasets, the relationship between topological signatures and missingness mechanisms, and the extent to which different biomedical domains produce comparable reconstruction-space organizations.

Our approach suggests experimentally testable previsions. First, empirical datasets with increasing fractions of artificially removed observations should exhibit a monotonic increase in reconstruction error while simultaneously displaying non-monotonic variation in topological descriptors. Specifically, the relationship between missingness and cycle counts is expected to be weaker than the relationship between missingness and reconstruction error.

Second, biological datasets generated from strongly constrained systems should produce fewer connected components than datasets generated from more heterogeneous populations when analyzed at identical missingness levels and proximity scales.

Third, repeated imputations generated by different algorithms may produce similar average estimates while exhibiting distinct connectivity structures. This prediction can be tested by comparing graph-derived descriptors obtained from methods such as predictive mean matching, Bayesian regression, random forests and neural-network imputations.

Fourth, increasing the coarse-proximity radius should induce a reproducible transition from multiple disconnected reconstruction regions toward a single connected structure. The critical radius associated with this transition constitutes a measurable observable that can be compared across datasets.

Fifth, if reconstruction spaces contain persistent organizational features, the values of connected-component counts and cycle counts should remain more stable under repeated resampling than under randomized permutation of imputation labels.

Sixth, datasets characterized by stronger internal correlation structures should exhibit lower critical gluing radii and denser proximity graphs than datasets with weaker dependence structures.

Future research may investigate persistent homology, higher-dimensional topological invariants, alternative notions of coarse similarity, information-theoretic descriptors, longitudinal missing-data processes, network-valued observations and comparisons between synthetic and empirical datasets.

Our analysis is applied as a post-processing step to completed datasets generated by conventional multiple-imputation procedures and therefore does not require modifications to existing data-processing pipeline. Reconstruction coordinates can be derived from summary statistics already reported in many studies, allowing direct computation of proximity graphs and associated descriptors. Our method can be integrated with existing quality-control procedures to compare different imputation strategies, evaluate reconstruction consistency across repeated analyses and identify transitions in reconstruction-space organization.

In conclusion, we investigated missing-data reconstruction from the perspective of the organization of admissible completed datasets. Multiple imputations generated structured spaces whose connectivity and cycle properties were quantified independently of conventional measures of reconstruction accuracy. We showed that numerical uncertainty and topological organization could provide distinct descriptions of the same reconstruction process. Therefore, examination of admissible completions as interconnected objects could extend the characterization of missing-data solutions beyond parameter estimates and error metrics alone.

DECLARATIONS

Ethics approval and consent to participate. Ethics approval statement is not applicable. This research does not contain any studies with human participants or animals performed by the Author.

Consent for publication. The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Availability of data and materials. All data and materials generated or analyzed during this study are included in the manuscript. The Author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Disclaimer. The views expressed are those of the author and do not necessarily reflect those of the affiliated institutions.

Competing interests. The Author does not have any known or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence or be perceived to influence their work.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Acknowledgements: none.

Authors' contributions. The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work, the author used ChatGPT 5.3 to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

REFERENCES

- 1) Alquier, Pierre. 2020. "Approximate Bayesian Inference." *Entropy* 22 (11): 1272. <https://doi.org/10.3390/e22111272>.
- 2) Arkowitz, M., and C. R. Curjel. 1964. "The Hurewicz Homomorphism and Finite Homotopy Invariants." *Transactions of the American Mathematical Society* 110 (3): 538–51. <https://doi.org/10.2307/1993697>.
- 3) Ben Salem, K., and A. Ben Abdelaziz. 2021. "Principal Component Analysis (PCA)." *La Tunisie Médicale* 99 (4): 383–89.
- 4) Bhardwaj, Manoj, and B. K. Tyagi. 2019. "Some Characterizations of Ideal Variants of Hurewicz Type Covering Properties." *arXiv preprint arXiv:1910.13822*.
- 5) Chan, K. C. 2017. "Acceleration of Expectation-Maximization Algorithm for Length-Biased Right-Censored Data." *Lifetime Data Analysis* 23 (1): 102–12. <https://doi.org/10.1007/s10985-016-9374-z>.
- 6) Chen, R., A. Schumitzky, A. Kryshchenko, K. Nieforth, M. Tomashevskiy, S. Hu, R. Garreau, J. Otalvaro, W. Yamada, and M. N. Neely. 2024. "RPEM: Randomized Monte Carlo Parametric Expectation Maximization Algorithm." *CPT: Pharmacometrics & Systems Pharmacology* 13 (5): 759–80. <https://doi.org/10.1002/psp4.13113>.
- 7) Chin, G. C., A. W. Potter, and K. E. Friedl. 2024. "Body Mass Index Is a Barrier to Obesity Treatment." *Frontiers in Endocrinology* 15: 1444568. <https://doi.org/10.3389/fendo.2024.1444568>.
- 8) Christensen, J. Daniel, and Luis Scoccola. 2023. "The Hurewicz Theorem in Homotopy Type Theory." *arXiv preprint arXiv:2007.05833v3*.
- 9) Chung, S. 2019. "Update on Low-Density Lipoprotein Cholesterol Quantification." *Current Opinion in Lipidology* 30 (4): 273–83. <https://doi.org/10.1097/MOL.0000000000000611>.
- 10) Cramer, A., A. von Wyl, M. Koemeda, P. Schulthess, and V. Tschuschke. 2015. "Sensitivity Analysis in Multiple Imputation in Effectiveness Studies of Psychotherapy." *Frontiers in Psychology* 6: 1042. <https://doi.org/10.3389/fpsyg.2015.01042>.
- 11) Çubukçu, H. C., and D. İ. Topcu. 2022. "Estimation of Low-Density Lipoprotein Cholesterol Concentration Using Machine Learning." *Laboratory Medicine* 53 (2): 161–71. <https://doi.org/10.1093/labmed/lmab065>.
- 12) D'Agostino, Ralph B., Ramachandran S. Vasan, Michael J. Pencina, Philip A. Wolf, Mark Cobain, Joseph M. Massaro, and William B. Kannel. 2008. "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study." *Circulation* 117 (6): 743–753.
- 13) De Raadt, A., M. J. Warrens, R. J. Bosker, and H. A. L. Kiers. 2019. "Kappa Coefficients for Missing Data." *Educational and Psychological Measurement* 79 (3): 558–76. <https://doi.org/10.1177/0013164418823249>.
- 14) Giganti, M. J., and B. E. Shepherd. 2020. "Multiple-Imputation Variance Estimation in Studies With Missing or Misclassified Inclusion Criteria." *American Journal of Epidemiology* 189 (12): 1628–32. <https://doi.org/10.1093/aje/kwaa153>.
- 15) Golchi, S., and J. J. Willard. 2024. "Estimating the Sampling Distribution of Posterior Decision Summaries in Bayesian Clinical Trials." *Biometrical Journal* 66 (8): e70002. <https://doi.org/10.1002/bimj.70002>.
- 16) Grzegorzolka, Pawel, and Jeremy Siegert. 2019. "Coarse Proximity and Proximity at Infinity." *arXiv preprint arXiv:1804.10263v6*.
- 17) Grzegorzolka, P., and J. Siegert. 2023. "Boundaries of Coarse Proximity Spaces and Boundaries of Compactifications." *European Journal of Mathematics* 9: 28. <https://doi.org/10.1007/s40879-023-00616-1>.
- 18) Huber, H. A., S. K. Georgia, and S. D. Finley. 2023. "Systematic Bayesian Posterior Analysis Guided by Kullback-Leibler Divergence Facilitates Hypothesis Formation." *Journal of Theoretical Biology* 558: 111341. <https://doi.org/10.1016/j.jtbi.2022.111341>.
- 19) Jang, J. H. 2021. "Principal Component Analysis of Hybrid Functional and Vector Data." *Statistics in Medicine* 40 (24): 5152–73. <https://doi.org/10.1002/sim.9117>.
- 20) Javanbakht, M., J. Lin, A. Ragsdale, S. Kim, S. Siminski, and P. Gorbach. 2022. "Comparing Single and Multiple Imputation Strategies for Harmonizing Substance Use Data across HIV-Related Cohort Studies." *BMC Medical Research Methodology* 22 (1): 90. <https://doi.org/10.1186/s12874-022-01554-4>.
- 21) Lieberman-Betz, R. G., P. Yoder, W. L. Stone, A. S. Nahmias, A. S. Carter, et al. 2014. "An Illustration of Using Multiple Imputation versus Listwise Deletion Analyses: The Effect of Hanen's 'More Than Words' on Parenting Stress." *American Journal on Intellectual and Developmental Disabilities* 119 (5): 472–86. <https://doi.org/10.1352/1944-7558-119.5.472>.
- 22) Marino, Mark, et al. 2014. "Quantifying Cardiometabolic Risk Using Modifiable Non-Self-Reported Risk Factors from the Framingham Offspring Cohort." *Preventing Chronic Disease* 11:E107.
- 23) Martins, T. D., J. M. Annichino-Bizzacchi, A. V. C. Romano, and R. M. Filho. 2019. "Principal Component Analysis on Recurrent Venous Thromboembolism." *Clinical and Applied Thrombosis/Hemostasis* 25: 1076029619895323. <https://doi.org/10.1177/1076029619895323>.

- 24) Mehrotra, D. V., F. Liu, and T. Permutt. 2017. “Missing Data in Clinical Trials: Control-Based Mean Imputation and Sensitivity Analysis.” *Pharmaceutical Statistics* 16 (5): 378–92. <https://doi.org/10.1002/pst.1817>.
- 25) Mi, J. X., Y. N. Zhang, Z. Lai, W. Li, L. Zhou, and F. Zhong. 2019. “Principal Component Analysis Based on Nuclear Norm Minimization.” *Neural Networks* 118: 1–16. <https://doi.org/10.1016/j.neunet.2019.05.020>.
- 26) Mitani, H., K. Suzuki, J. Ako, K. Iekushi, R. Majewska, S. Touzeni, and S. Yamashita. 2023. “Achievement Rates for Low-Density Lipoprotein Cholesterol Goals in Patients at High Risk of Atherosclerotic Cardiovascular Disease in a Real-World Setting in Japan.” *Journal of Atherosclerosis and Thrombosis* 30 (11): 1622–34. <https://doi.org/10.5551/jat.63940>.
- 27) Niloofar, P., R. Aghdam, and C. Eslahchi. 2024. “GAEM: Genetic Algorithm Based Expectation-Maximization for Inferring Gene Regulatory Networks from Incomplete Data.” *Computers in Biology and Medicine* 183: 109238. <https://doi.org/10.1016/j.combiomed.2024.109238>.
- 28) Oh, G. C., T. Ko, J. H. Kim, M. H. Lee, S. W. Choi, Y. S. Bae, K. H. Kim, and H. Y. Lee. 2022. “Estimation of Low-Density Lipoprotein Cholesterol Levels Using Machine Learning.” *International Journal of Cardiology* 352: 144–49. <https://doi.org/10.1016/j.ijcard.2022.01.029>.
- 29) Pilgrim, Samantha. 2025. “A Hurewicz-Type Theorem for the Dynamic Asymptotic Dimension with Applications to Coarse Geometry and Dynamics.” *arXiv preprint arXiv:2207.13275v6*.
- 30) Prolubnikov, Alexander. 2023. “Finding the Connected Components of the Graph Using Perturbations of the Adjacency Matrix.” *arXiv preprint arXiv:2306.16389v2*.
- 31) Qiao, Z., J. Sidorenko, J. A. Revez, A. Xue, X. Lu, K. Pärna, H. Snieder, Lifelines Cohort Study, P. M. Visscher, N. R. Wray, and L. Yengo. 2023. “Estimation and Implications of the Genetic Architecture of Fasting and Non-Fasting Blood Glucose.” *Nature Communications* 14 (1): 451. <https://doi.org/10.1038/s41467-023-36013-1>.
- 32)
- 33) Raevskiy, M., V. Yanvarev, S. Jung, A. Del Sol, and Y. A. Medvedeva. 2023. “Epi-Impute: Single-Cell RNA-seq Imputation via Integration with Single-Cell ATAC-seq.” *International Journal of Molecular Sciences* 24 (7): 6229. <https://doi.org/10.3390/ijms24076229>.
- 34) Ramos, Arthur F., Tiago M. L. de Veras, and Ruy J. G. B. de Queiroz. 2025. “The Seifert-van Kampen Theorem via Computational Paths: A Formalized Approach to Computing Fundamental Groups.” *arXiv preprint arXiv:2512.03175*.
- 35) Sakai, Masami. 2020. “The Menger Property and l-Equivalence.” *Topology and Its Applications* 281: 107187. <https://doi.org/10.1016/j.topol.2020.107187>.
- 36) Ségalas, C., C. Leyrat, J. R. Carpenter, and E. Williamson. 2023. “Propensity Score Matching after Multiple Imputation When a Confounder Has Missing Data.” *Statistics in Medicine* 42 (7): 1082–95. <https://doi.org/10.1002/sim.9658>.
- 37) Shi, Y., and W. Yao. 2024. “Lattice-Valued Coarse Proximity Spaces.” *Fuzzy Sets and Systems* 475: 108766. <https://doi.org/10.1016/j.fss.2023.108766>.
- 38) Shu, D., X. Li, Q. Her, J. Wong, D. Li, R. Wang, and S. Toh. 2023. “Combining Meta-Analysis with Multiple Imputation for One-Step, Privacy-Protecting Estimation of Causal Treatment Effects in Multi-Site Studies.” *Research Synthesis Methods* 14 (5): 742–63. <https://doi.org/10.1002/jrsm.1660>.
- 39) Solianik, R., K. Židonienė, N. Eimantas, and M. Brazaitis. 2023. “Prolonged Fasting Outperforms Short-Term Fasting in Terms of Glucose Tolerance and Insulin Release: A Randomised Controlled Trial.” *British Journal of Nutrition* 130 (9): 1500–1509. <https://doi.org/10.1017/S0007114523000557>.
- 40) Tompsett, D., S. Sutton, S. R. Seaman, and I. R. White. 2020. “A General Method for Elicitation, Imputation, and Sensitivity Analysis for Incomplete Repeated Binary Data.” *Statistics in Medicine* 39 (22): 2921–35. <https://doi.org/10.1002/sim.8584>.
- 41) Tsang, J. N. J., S. Bacchi, J. G. Kovoov, A. K. Gupta, B. Stretton, S. Gluck, T. Gilbert, Y. Sharma, R. Woodman, and A. A. Mangoni. 2023. “Systolic Blood Pressure Levels and Mortality in Australian Medical Inpatients.” *Journal of Clinical Hypertension* 25 (11): 1036–39. <https://doi.org/10.1111/jch.14735>.
- 42) van Dijk, R., D. Kalisvaart, J. Cnossen, and C. S. Smith. 2023. “Bayesian Posterior Density Estimation Reveals Degeneracy in Three-Dimensional Multiple Emitter Localization.” *Scientific Reports* 13 (1): 22372. <https://doi.org/10.1038/s41598-023-49101-5>.
- 43) Yu, K., X. Liu, R. Alhamzawi, F. Becker, and J. Lord. 2018. “Statistical Methods for Body Mass Index: A Selective Review.” *Statistical Methods in Medical Research* 27 (3): 798–811. <https://doi.org/10.1177/0962280216643117>.
- 44) Yu, L., L. Liu, and K. E. Peace. 2020. “Regression Multiple Imputation for Missing Data Analysis.” *Statistical Methods in Medical Research* 29 (9): 2647–64. <https://doi.org/10.1177/0962280220908613>.
- 45) Yu, Y., X. Zhang, S. Wang, et al. 2026. “Efficient and Accurate Neural-Field Reconstruction Using Resistive Memory.” *Nature* 654: 642–651. <https://doi.org/10.1038/s41586-026-10646-w>.

Article

Not peer-reviewed version

A Grothendieckian Perspective on Ancient Manuscript Analysis

[Arturo Tozzi](#) *

Posted Date: 14 July 2026

doi: 10.20944/preprints202607.0927.v1

Keywords: invariants; functors; sheaves; topology; provenance



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Grothendieckian Perspective on Ancient Manuscript Analysis

Arturo Tozzi

ASL Napoli 1 Centro, Via Comunale del Principe 13/a, 80145 Naples, Italy; tozziarturo@libero.it

Abstract

Ancient manuscripts are cultural objects encoding physical, chemical, linguistic, historical, philosophical and social information. To unify heterogeneous disciplinary evidence within a single relational and geometric formalism, we propose a Grothendieckian framework inspired by category theory, sheaf theory, topoi and algebraic geometry. Each discipline is represented as an independent graph whose nodes and edges encode its own entities and relationships. Local observations are attached to graph neighborhoods through sections of sheaves, while structure-preserving functors align corresponding entities across disciplinary domains. Natural transformations evaluate the consistency of alternative mappings, while cohomological analysis identifies hidden structural constraints. Subsequently, compatible local information is assembled using descent theory, whereas stacks and derived categories incorporate multiple manuscript versions and their historical evolution. The integrated relational structure is represented geometrically through schemes, with the common invariant subgraph interpreted as the manuscript's motive, namely the maximal relational architecture preserved across all disciplinary representations. This approach could provide a mathematical strategy for integrating heterogeneous evidence without privileging any single disciplinary viewpoint. Potential applications include manuscript authentication, conservation planning, textual criticism, comparative philology, cultural heritage analytics, historical knowledge integration and development of interoperable artificial intelligence systems for multidisciplinary manuscript research.

Keywords: invariants; functors; sheaves; topology; provenance

Introduction

The interpretation of ancient manuscripts requires the integration of evidence originating from different disciplines, including paleography, philology, linguistics, chemistry, physics, history, art history, conservation science, etc. Advances in non-destructive analytical techniques have expanded the range of measurable properties extracted from manuscripts (Tack et al. 2016; Stabile et al. 2021; Angelotti et al. 2026), including parchment composition, pigment chemistry, ink formulation, radiocarbon dating, multispectral imaging and microstructural characterization (Radini et al. 2019; Tournié et al. 2019; Autran et al. 2021; Popović et al. 2025). In parallel, computational methods have enabled increasingly sophisticated analyses of textual organization, authorship, semantic relationships and historical transmission (Romano et al. 2023; Ceccarelli et al. 2025). Network science, machine learning and digital humanities have contributed tools for representing relationships among textual units, scribes, manuscripts and historical actors (Xu et al. 2023; Chen 2024; Zhu et al. 2025). However, relationships identified by one discipline are often compared only qualitatively across disciplines, hindering the identification of higher-order structures. A general mathematical language for combining different disciplinary viewpoints into a unified representation is still largely undeveloped.

We aim to develop an axiomatic construction inspired by Grothendieckian geometry, introducing the mathematical objects and relationships required for the geometric analysis of ancient manuscripts (Zalamea 2015; Morone, Leifer, and Makse 2020; Sobota and Zdomskyy 2024; Gili et al.

2024; Bennequin and Berthoz 2025; Marquès 2025). Our theoretical goal is to establish a coherent mathematical formalism capable of representing ancient manuscripts as geometric objects whose heterogeneous disciplinary descriptions can be integrated within a single relational framework.

We regard each disciplinary viewpoint as a coherent relational world rather than an isolated dataset (Luo 1999; Dubuc and Sanchez de la Vega 2000; Pisier 2012; McLarty 2018). Every discipline is represented by an independent mathematical category whose objects and morphisms encode its entities and relationships. Local observations are associated with manuscript regions through sheaves, while functors translate information between categories and natural transformations assess the compatibility of alternative translations. Cohomology identifies global inconsistencies, after which compatible local descriptions are assembled by descent into a unified representation incorporating multiple manuscript versions, historical modifications and transmission pathways through stacks and derived categories. The resulting structure is interpreted geometrically as a scheme from which the manuscript's motive is extracted, defined as the maximal relational architecture preserved under all admissible discipline-preserving functors. The overall construction is summarized in Figure 1.

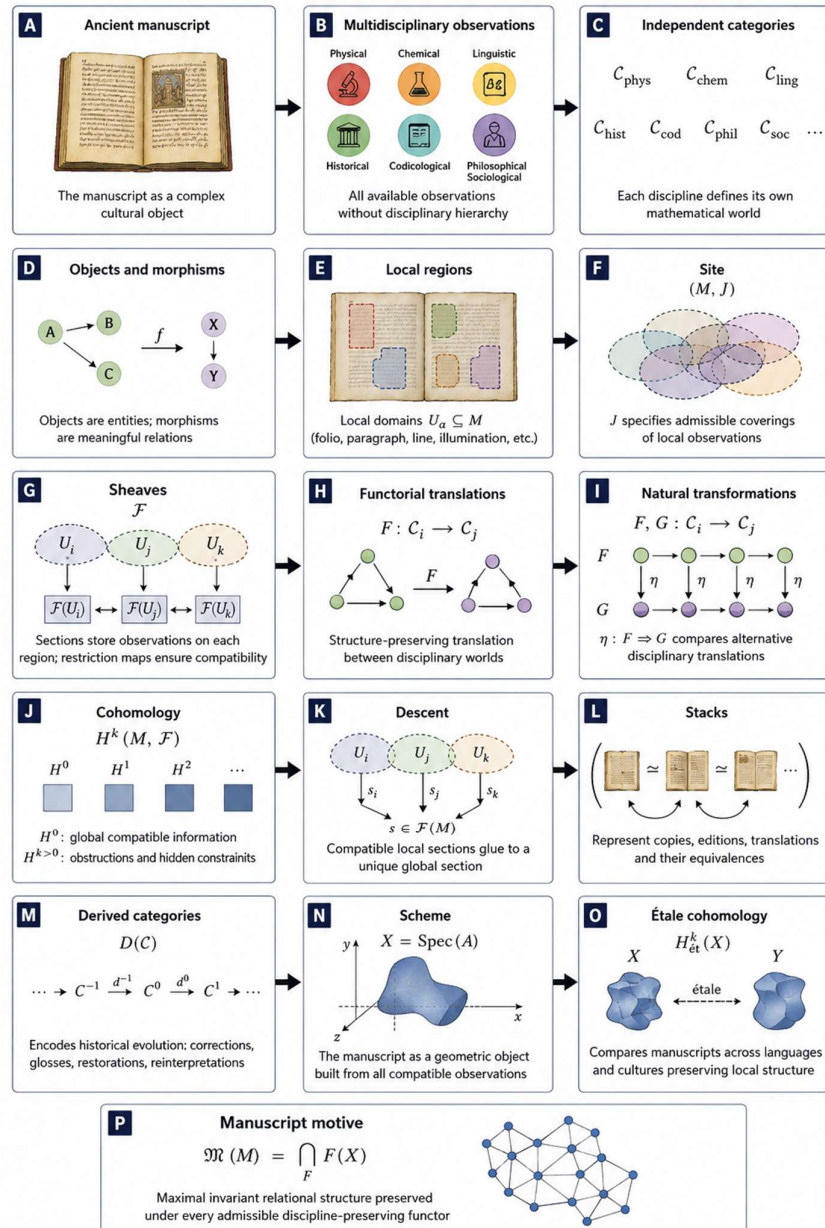


Figure 1. The construction begins by defining the mathematical objects and their relationships, before assembling them into increasingly complex structures. Every mathematical object introduced below defines a coherent geometric language in which physical, chemical, linguistic, historical, codicological, philosophical and sociological evidence become different manifestations of the same underlying mathematical entity.

Initially, an ancient manuscript can be regarded as a finite collection of observations rather than as a text. Let

$$M = \{O_1, O_2, \dots, O_n\}$$

denote the set of all observations associated with the manuscript. Each observation may correspond to a measurable physical quantity, a chemical property, a linguistic element, a historical event, an iconographic feature, a codicological characteristic, a philosophical concept, a sociological relationship, etc. Since each discipline captures only one aspect of the manuscript, no hierarchy is imposed among these observations. At the outset, the manuscript has no privileged mathematical description. Instead, each discipline provides an independent relational representation that is subsequently integrated into a unified geometric construction.

Each scientific discipline defines its own mathematical category. Let

$$\mathcal{C}_i$$

denote the category associated with discipline i . Its objects,

$$\text{Obj}(\mathcal{C}_i),$$

represent the entities studied by that discipline, whereas its morphisms,

$$\text{Mor}(\mathcal{C}_i),$$

describe admissible relationships between those entities. For instance, in the physical category, objects may consist of parchment regions, fibres, folios or bindings, while morphisms represent structural or mechanical relationships. In chemistry, objects become pigment samples, ink deposits or molecular compounds connected through compositional similarities or reaction pathways. Linguistics regards words, sentences, semantic concepts or grammatical constructions as objects linked through syntactic or semantic relations. History introduces scribes, production centres, historical events and ownership records connected by chronological or documentary relationships. Philosophy may represent concepts and arguments, whereas sociology considers readers, monasteries, libraries and transmission routes.

Morphisms stand for the elementary relations through which every category acquires structure. If A and B are two objects of a category,

$$f: A \rightarrow B$$

denotes a morphism relating them. Morphisms satisfy the axioms of identity and composition,

$$\text{id}_A: A \rightarrow A,$$

and

$$g \circ f: A \rightarrow C,$$

whenever

$$f: A \rightarrow B, g: B \rightarrow C.$$

Composition enables elementary relationships to be combined into increasingly comprehensive explanatory chains. For example, a physical measurement may lead to material identification, which may constrain technological chronology and eventually historical provenance. Similarly, a linguistic observation may propagate through grammatical analysis toward dialect identification and geographical localization. Therefore, complex interpretations emerge through successive compositions of elementary morphisms rather than through isolated measurements.

Next, the manuscript could be decomposed into local regions. Let

$$U_\alpha \subseteq M$$

denote an arbitrary local domain. Depending on the investigation, a local region may correspond to an individual folio, a paragraph, an illumination, a line of text, a pigment layer, a damaged portion or any subset of observations. Importantly, locality is not restricted to physical neighbourhoods. Two

distant folios sharing identical scribal characteristics may belong to the same local analytical region, whereas neighbouring folios exhibiting different material histories may belong to distinct regions. Locality becomes an abstract relational concept rather than a purely spatial one.

To formalize locality, the manuscript is endowed with a Grothendieck topology. Let

$$(M, J)$$

denote a site, where J specifies the admissible covering families

$$\{U_i \rightarrow U\}.$$

Unlike ordinary topology, a covering does not necessarily describe geometric openness but identifies collections of compatible local observations sufficient to reconstruct larger portions of the manuscript. The choice of coverings depends entirely on analytical coherence rather than physical adjacency. Consequently, chemistry, linguistics and history may generate different local coverings while still referring to the same manuscript.

Each local region carries measurable information through a sheaf. Let

$$\mathcal{F}$$

be a sheaf over the site (M, J) . Every local domain receives a corresponding collection of observations

$$\mathcal{F}(U),$$

called the sections over U . Sections may contain spectra, pigment compositions, transcription data, grammatical analyses, historical metadata, codicological measurements or any other disciplinary observations. Whenever

$$V \subseteq U,$$

restriction morphisms

$$\rho_{UV}: \mathcal{F}(U) \rightarrow \mathcal{F}(V)$$

transfer information consistently from larger regions to smaller ones. The sheaf satisfies the classical gluing axiom: if local sections

$$s_i \in \mathcal{F}(U_i)$$

agree on every overlap

$$U_i \cap U_j,$$

then there exists a unique global section

$$s \in \mathcal{F}(U)$$

whose restrictions coincide with every s_i . Local observations become mathematically compatible before any global interpretation is attempted.

Independent disciplinary worlds must subsequently communicate through functors. Given two categories,

$$\mathcal{C}_i \text{ and } \mathcal{C}_j,$$

a functor

$$F: \mathcal{C}_i \rightarrow \mathcal{C}_j$$

maps objects to objects and morphisms to morphisms while preserving identity morphisms and composition,

$$F(\text{id}_A) = \text{id}_{F(A)},$$

$$F(g \circ f) = F(g) \circ F(f).$$

The preservation of relational structure distinguishes functorial translation from ordinary data conversion. For instance, a linguistic inference regarding geographical dialect may become a historical localization without modifying the logical relationships among the observations. Likewise, pigment provenance inferred chemically may become a node within historical trade networks while maintaining its original relational context. Therefore, functors translate entire relational organizations rather than isolated variables.

Different disciplinary translations frequently coexist. Suppose two functors,

$$F, G: \mathcal{C}_i \rightarrow \mathcal{C}_j,$$

produce alternative historical interpretations from the same observations. Their compatibility is expressed through a natural transformation

$$\eta: F \Rightarrow G,$$

whose components

$$\eta_A: F(A) \rightarrow G(A)$$

satisfy

$$\eta_B \circ F(f) = G(f) \circ \eta_A$$

for every morphism

$$f: A \rightarrow B.$$

Natural transformations compare complete translations rather than individual observations. They provide mathematical criterion for determining whether different disciplinary interpretations remain globally coherent.

Once compatible translations have been established, global consistency becomes a cohomological problem. Cohomology assesses whether local information can be assembled into a coherent global description. For the sheaf $\mathcal{F}$, one computes the cohomology groups

$$H^k(M, \mathcal{F}).$$

The zeroth group

$$H^0(M, \mathcal{F})$$

contains globally compatible sections. Higher-order groups identify obstructions preventing perfect integration. Non-trivial elements of

$$H^1(M, \mathcal{F})$$

may point towards inconsistencies between historical chronology and chemical dating, incompatible palaeographic assignments or conflicting restoration histories. Higher-dimensional groups detect progressively more complex incompatibilities involving multiple disciplinary domains simultaneously, while cohomology characterizes the manuscript through global structural constraints rather than isolated discrepancies.

Whenever cohomological obstructions vanish, descent theory reconstructs the manuscript from compatible local data. Given a covering

$$\{U_i\},$$

together with compatible local sections,

$$s_i \in \mathcal{F}(U_i),$$

satisfying

$$s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j},$$

there exists a unique global section

$$s \in \mathcal{F}(M)$$

whose restrictions reproduce every local observation. Therefore, descent provides the rigorous mathematical mechanism through which fragmented evidence becomes a unified description.

Ancient manuscripts frequently exist in multiple manifestations. Original copies, later reproductions, translations, commentaries, restorations and palimpsests cannot be represented adequately as isolated objects. Their mathematical description requires stacks, which generalize ordinary spaces by retaining equivalences and automorphisms among related objects. A stack records not only each manuscript version, but also every symmetry relating those versions. Two copies differing only through minor orthographic changes are distinct while preserving their equivalence within the larger mathematical object.

Historical evolution further introduces derived categories. Instead of studying isolated objects, one considers complexes

$$\mathcal{C}^\bullet$$

linked by differentials

$$d^2 = 0.$$

Passing to the derived category

$$D(\mathcal{C})$$

retains the complete history of transformations, including corrections, glosses, erasures, translations, restorations and reinterpretations accumulated through centuries. Consequently, the manuscript acquires temporal depth rather than being a static object.

After these constructions, every compatible disciplinary representation may be assembled into a geometric object. Let

$$A$$

denote the algebra generated by all compatible observations and their relations. Then the manuscript is represented by the affine scheme

$$X = \operatorname{Spec}(A),$$

whose points correspond to prime ideals encoding coherent relational structures rather than individual measurements. This means that material composition, textual organization, historical context and semantic organization could become geometric coordinates of the same mathematical space.

Comparisons among manuscripts originating from different cultures require invariants independent of vocabulary or local conventions. Étale morphisms preserve local structure while allowing global comparison among distinct schemes. Étale cohomology identifies deep correspondences invisible to ordinary textual comparison, revealing structural similarities among manuscripts despite differences in language, script or cultural tradition.

The final mathematical object sought by our construction is the manuscript’s motive. Inspired by Grothendieck’s notion of motives in algebraic geometry (Huber 1995; Rej and Marcolli 2011; Baez 2025), we introduce a new operational definition. Let

$$\mathcal{F} = \{F_\alpha\}$$

denote the collection of all admissible discipline-preserving functors acting on the manuscript’s geometric representation. The manuscript motive is defined as the maximal invariant relational structure preserved under every admissible functorial translation,

$$\mathfrak{M}(M) = \bigcap_{F_\alpha \in \mathcal{F}} F_\alpha(X).$$

This invariant does not correspond to any single aspect of the manuscript, whether physical, textual, chemical, historical or cultural. Instead, it represents the largest relational architecture surviving every mathematically admissible disciplinary perspective. The motive constitutes the manuscript’s structural identity, i.e., the organization simultaneously preserved by its material composition, textual content, historical evolution, semantic architecture and social transmission.

Overall, our approach transforms an ancient manuscript from a heterogeneous collection of observations into a unified geometric object whose invariant relational structure is represented within a Grothendieckian geometry.

Table 1. Sequential construction of the proposed Grothendieckian geometry of ancient manuscripts. The table summarizes the mathematical objects introduced during the progressive transformation of an ancient manuscript from a heterogeneous collection of multidisciplinary observations into a unified geometric object. The final invariant, defined as the manuscript’s motive, is the maximal relational structure preserved across all admissible disciplinary representations.

Step	Grothendieckian concept	Mathematical object	Manuscript features examined	Role in manuscript geometry
1	Observations	$M = \{O_1, \dots, O_n\}$	Parchment, inks, pigments, text, dates, provenance, annotations	Collect all available evidence

2	Categories	$\mathcal{C}_i$	Physics, chemistry, linguistics, history, codicology, philosophy, sociology	Separate disciplinary worlds
3	Objects & morphisms	$A \xrightarrow{f} B$	Folios, pigments, words, scribes, places and their relationships	Define elementary relational structure
4	Local regions	$U_\alpha \subseteq M$	Individual folios, paragraphs, illuminations, damaged areas	Partition the manuscript into local domains
5	Site	(M, J)	Compatible local analytical regions	Define admissible local observations
6	Sheaf	$\mathcal{F}(U)$	Spectra, OCR, grammar, codicology, metadata	Associate observations with each local region
7	Functor	$F: \mathcal{C}_i \rightarrow \mathcal{C}_j$	Correspondence between material, textual and historical evidence	Translate one disciplinary description into another
8	Natural transformation	$\eta: F \Rightarrow G$	Independent provenance or dating hypotheses	Compare alternative interpretations
9	Cohomology	$H^k(M, \mathcal{F})$	Dating conflicts, inconsistent provenance, missing information	Detect global inconsistencies
10	Descent	$s \in \mathcal{F}(M)$	Compatible multidisciplinary observations	Reconstruct the complete manuscript
11	Stack	$\mathfrak{X}$	Copies, editions, translations, commentaries, palimpsests	Represent manuscript families
12	Derived category	$D(\mathcal{C})$	Corrections, glosses, restorations, historical revisions	Describe temporal evolution
13	Scheme	$X = \text{Spec}(A)$	Integrated physical, textual, historical and semantic information	Construct the manuscript's geometric space
14	Étale cohomology	$H^k_{\text{ét}}(X)$	Related manuscripts from different languages or traditions	Reveal deep structural correspondences
15	Motive	$\mathfrak{M}(M)$	Stable multidisciplinary relationships	Extract the manuscript's universal relational identity

Illustrative Theoretical Construction of the Manuscript Motive

To clarify how our mathematical construction may be applied in practice, we consider a hypothetical ancient manuscript. Each mathematical step follows from the definitions introduced in the preceding sections.

At first, we identify the observable characteristics of the manuscript. Rather than privileging textual information, all measurable or interpretable features are considered equally informative. The physical description includes the writing support, folio organization, parchment morphology, binding structure and page layout. Chemical observations comprise ink composition, pigment distribution and material degradation. Linguistic analysis identifies lexical usage, grammatical organization and semantic relationships among textual units. Historical investigation reconstructs chronology, provenance, ownership and documentary references, whereas codicology characterizes quire structure, pagination, scribal practices and corrections. Additional disciplinary perspectives, including iconography, philosophy or sociology, may contribute further observations whenever

appropriate. At this stage, the manuscript is still a heterogeneous collection of independent observations. Each disciplinary description is represented as an independent category whose objects correspond to the entities examined within that discipline and whose morphisms encode their internal relationships. For instance, physical objects may consist of parchment regions connected by structural continuity, chemical objects of pigment samples linked by compositional similarity, linguistic objects of words or concepts connected by syntactic or semantic relations, and historical objects of places, individuals or events linked by chronological dependencies. Every category develops its own internal relational architecture independently of the others. The first mathematical outcome of this stage is a collection of coherent but mutually disconnected relational worlds, each describing one aspect of the manuscript.

Subsequently, the manuscript is decomposed into local analytical regions. These regions need not correspond exclusively to physical subdivisions such as folios or pages, but may equally represent paragraphs, illuminated initials, marginal annotations, damaged areas or any subset of observations sharing analytical coherence. Every local region receives a collection of disciplinary observations represented as sections of appropriate sheaves. Restriction morphisms ensure that information is internally consistent when passing from larger regions to smaller ones. Whenever neighbouring regions contain compatible observations, the sheaf axioms guarantee that they can later be assembled into a unique global description. The principal outcome of this stage is the replacement of isolated measurements by mathematically consistent local descriptions distributed across the manuscript.

Then, relations among the independent disciplinary categories are established through functors. Correspondences between chemistry and history may associate pigment provenance with historical trade networks, while linguistic characteristics may be translated into geographical localization or chronological attribution. Codicological observations may similarly constrain historical reconstruction through quire organization or scribal practice.

Subsequently, natural transformations compare these independent translations. If multiple disciplinary pathways converge toward the same interpretation, the corresponding natural transformations express their compatibility. Conversely, incompatible translations reveal regions requiring further investigation. At this point, the manuscript is no longer represented by independent disciplinary descriptions, but by an interconnected network of mathematically compatible relationships.

The global consistency of these relationships is examined through cohomological analysis. Rather than focusing on individual observations, cohomology evaluates whether the complete collection of local sections can be assembled into a globally coherent object. Compatible observations contribute to the zeroth cohomology, whereas higher-order cohomology groups identify incompatibility, missing information or structural contradictions linking several disciplinary domains simultaneously. When these obstructions are absent or sufficiently resolved, descent theory reconstructs a unique global object by gluing all compatible local descriptions. The resulting mathematical object represents the manuscript as an integrated relational entity rather than as the sum of independent analytical results.

Then, the construction is extended to incorporate the manuscript's historical complexity. Multiple copies, translations, editions, restorations and commentaries are represented as objects within a stack, preserving both their individual identities and their mutual equivalences. Derived categories further incorporate successive historical transformations, including revisions, glosses, interpolations and corrections, allowing the manuscript to be represented together with its temporal evolution rather than as a static document.

Subsequently, the integrated algebra generated by all compatible observations is transformed into a geometric object through the construction of a scheme. At this stage, material composition, textual organization, historical transmission and conceptual architecture become different coordinate systems describing the same underlying geometric space.

The final step consists of identifying the manuscript's motive. Every admissible discipline-preserving functor provides an alternative representation of the geometric object. The structures invariant under all such representations are retained, whereas discipline-specific features not surviving these translations are discarded. The remaining invariant relational architecture constitutes the manuscript's motive according to our operational definition. The expected theoretical outcome is not a numerical value or statistical descriptor, but a geometric object representing the structural identity of the manuscript. This object simultaneously captures information shared by physical composition, textual organization, historical development, semantic structure and cultural transmission while remaining independent of any individual disciplinary viewpoint. Consequently, two manuscripts differing in language, material support or historical context may nevertheless possess closely related motives if their deepest relational organization is preserved, whereas apparently similar manuscripts may exhibit distinct motives when their invariant structures differ.

Overall, the mathematical concepts introduced throughout the manuscript could combine into a single coherent sequence leading from local multidisciplinary observations to a geometric representation of ancient documentary knowledge.

Conclusions

We asked whether heterogeneous evidence in ancient manuscripts can be represented within a single mathematical language rather than fragmented among independent disciplinary analyses. We aim to provide a coherent language in which material, textual, historical and conceptual evidence can be represented within a common geometric setting. We introduced a Grothendieckian construction that progressively transforms multidisciplinary observations into a unified geometric object through categories, morphisms, sites, sheaves, functors, natural transformations, cohomology, descent theory, stacks, derived categories and schemes. The motive was defined as the maximal invariant relational structure preserved under all admissible discipline-preserving functors.

Our study has limitations. Our construction is theoretical and avoids simulations or applications to individual manuscripts, leaving future studies to evaluate computational feasibility and practical implementation. Several concepts adopted from Grothendieck's mathematics, particularly motives, schemes and stacks, are introduced through operational definitions adapted to manuscript analysis. Specialists in algebraic geometry may regard our definitions as analogical extensions rather than direct applications of the original theory. We do not specify unique algorithms for defining admissible coverings or selecting functors. Generalizability requires to be established across manuscript traditions differing in age, preservation, language or documentary complexity. Additional questions are unanswered, including uniqueness of the extracted motive, stability of the invariant under incomplete observations, relationship between different admissible geometric constructions and computational complexity associated with large manuscript corpora.

Our approach suggests experimentally testable hypotheses. First, independently constructed disciplinary graphs describing the same manuscript should converge toward a common invariant relational substructure after functorial alignment. This prediction can be evaluated by measuring graph similarity before and after alignment using graph edit distance, graph kernels or spectral similarity, with the expectation that aligned representations will exhibit significantly higher structural agreement than randomly permuted graphs.

Second, the size of the extracted motive should increase as additional complementary disciplines are incorporated until reaching a stable plateau, providing a measurable convergence criterion for multidisciplinary integration.

Third, manuscripts belonging to the same textual tradition should exhibit more similar motives than unrelated manuscripts even when written in different languages, scripts or materials. Similarity could be quantified through graph isomorphism scores, persistent topological descriptors or distances between invariant subgraphs, with manuscript families expected to cluster significantly more closely than unrelated controls.

Fourth, manuscripts containing later interpolations, restorations or forged additions should produce localized cohomological inconsistencies that coincide with independently established historical anomalies. Their detection sensitivity and specificity could be quantified against expert annotations.

Fifth, successive historical copies should preserve the original invariant structure despite physical or linguistic modifications. A measurable expectation is that invariant-node retention decreases more slowly than overall graph similarity as copying distance increases.

Future investigations could develop explicit computational algorithms, compare alternative categorical representations, study the uniqueness and stability of manuscript motives under noisy or incomplete observations, and extend the formalism to archaeological artefacts and documentary archives.

Beyond manuscript research, our approach may support the development of interoperable analytical environments in which heterogeneous evidence can be represented without privileging any single discipline. These representations could facilitate communication among historians, conservators, philologists, chemists, physicists and computer scientists by providing a common relational vocabulary rather than discipline-specific data structures. Museum collections, digital repositories and cultural heritage databases could benefit from mathematically consistent descriptions capable of integrating material analyses, textual metadata and historical documentation within the same formal representation. Also, our framework may stimulate new educational approaches in which abstract concepts from category theory and algebraic geometry are illustrated through tangible historical objects, encouraging the development of common standards for multidisciplinary knowledge representation.

Authors' contributions. The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Funding.: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Ethics approval and consent to participate. Ethics approval statement is not applicable. This research does not contain any studies with human participants or animals performed by the Author.

Consent for publication. The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Availability of data and materials. All data and materials generated or analyzed during this study are included in the manuscript. The Author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work, the author used ChatGPT 5.3 to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

References

1. Angelotti, G., F. Nicolardi, P. Henderson, and W. B. Seales. 2026. "Ink Detection from Surface Topography of the Herculaneum Papyri." *Scientific Reports*. Advance online publication. <https://doi.org/10.1038/s41598-026-58467-1>.
2. Autran, Pierre-Olivier, Catherine Dejoie, Pierre Bordet, Jean-Louis Hodeau, Caroline Dugand, et al. 2021. "Revealing the Nature of Black Pigments Used on Ancient Egyptian Papyri from Champollion Collection." *Analytical Chemistry* 93 (2): 701–709. <https://doi.org/10.1021/acs.analchem.0c04178>.

3. Baez, John C. 2025. "Motivating Motives." In *The Mathematical and Philosophical Legacy of Alexander Grothendieck*, edited by Marco Panza, Daniele C. Struppa, and Jean-Jacques Szczepniarz, 309–323. Basel: Birkhäuser. Also available as *arXiv* preprint arXiv:2304.08737. <https://doi.org/10.48550/arXiv.2304.08737>.
4. Bennequin, D., and A. Berthoz. 2025. "Brain's Geometries for Movements and Beauty Judgments. A Contribution of Topos Geometries." *Frontiers in Psychology* 16: 1583185. <https://doi.org/10.3389/fpsyg.2025.1583185>.
5. Ceccarelli, S., M. Rippa, G. Caruso, L. Luvidi, S. Boccuti, M. et al. 2025. "Pulsed Thermographic Analysis of Herculaneum Papyri." *Scientific Reports* 15 (1): 34466. <https://doi.org/10.1038/s41598-025-19911-w>.
6. Chen, Z. 2024. "(HTBNet) Arbitrary Shape Scene Text Detection with Binarization of Hyperbolic Tangent and Cross-Entropy." *Entropy* 26 (7): 560. <https://doi.org/10.3390/e26070560>.
7. Dubuc, E. J., and C. Sanchez de la Vega. 2000. "On the Galois Theory of Grothendieck." *Boletín de la Academia Nacional de Ciencias de Córdoba* 65.
8. Gili, T., B. Avila, L. Pasquini, A. Holodny, D. Phillips, P. Boldi, et al. 2024. "Fibration Symmetry-Breaking Supports Functional Transitions in a Brain Network Engaged in Language." *arXiv* preprint arXiv:2409.02674v1.
9. Huber, Annette. 1995. "Grothendieck Motives." In *Mixed Motives and Their Realization in Derived Categories, Lecture Notes in Mathematics*, vol. 1604, 100–138. Berlin and Heidelberg: Springer. <https://doi.org/10.1007/BFb0095523>.
10. Luo, Feng. 1999. "Grothendieck's Reconstruction Principle and 2-Dimensional Topology and Geometry." *Communications in Contemporary Mathematics* 1 (2): 125–53.
11. Marquès, J. 2025. "A Criterion for Categories on Which Every Grothendieck Topology Is Rigid." *Applied Categorical Structures* 33 (6): 37. <https://doi.org/10.1007/s10485-025-09833-z>.
12. McLarty, Colin. 2018. "Grothendieck's Unifying Vision of Geometry." In *Foundations of Mathematics and Physics One Century After Hilbert*, edited by Jaume Kounieher, 81–106. Cham: Springer. https://doi.org/10.1007/978-3-319-64813-2_4.
13. Morone, F., I. Leifer, and H. A. Makse. 2020. "Fibration Symmetries Uncover the Building Blocks of Biological Networks." *Proceedings of the National Academy of Sciences of the United States of America* 117 (15): 8306–14. <https://doi.org/10.1073/pnas.1914628117>.
14. Pisier, Gilles. 2012. "Grothendieck's Theorem, Past and Present." *Bulletin of the American Mathematical Society* 49 (2): 237–323.
15. Popović, M., M. A. Dhali, L. Schomaker, J. van der Plicht, K. Lund Rasmussen, J. La Nasa, I. Degano, M. Perla Colombini, and E. Tigchelaar. 2025. "Dating Ancient Manuscripts Using Radiocarbon and AI-Based Writing Style Analysis." *PLoS ONE* 20 (6): e0323185. <https://doi.org/10.1371/journal.pone.0323185>.
16. Radini, A., M. Tromp, A. Beach, E. Tong, C. Speller, M. McCormick, J. V. Dudgeon, M. J. Collins, F. Rühli, R. Kröger, and C. Warinner. 2019. "Medieval Women's Early Involvement in Manuscript Production Suggested by Lapis Lazuli Identification in Dental Calculus." *Science Advances* 5 (1): eaau7126. <https://doi.org/10.1126/sciadv.aau7126>.
17. Rej, Abhijnan, and Matilde Marcolli. 2011. "Motives: An Introductory Survey for Physicists." *Contemporary Mathematics* 539: 377–415. Providence, RI: American Mathematical Society. Also available as *arXiv* preprint arXiv:0907.4046. <https://doi.org/10.48550/arXiv.0907.4046>.
18. Romano, F. P., E. Puglia, C. Caliri, D. P. Pavone, M. Alessandrelli, A. Busacca, et al. 2023. "Layout of Ancient Greek Papyri through Lead-Drawn Ruling Lines Revealed by Macro X-Ray Fluorescence Imaging." *Scientific Reports* 13 (1): 6582. <https://doi.org/10.1038/s41598-023-33242-8>.
19. Sobota, D., and L. Zdomskyy. 2024. "Convergence of Measures after Adding a Real." *Archive for Mathematical Logic* 63 (1–2): 135–62. <https://doi.org/10.1007/s00153-023-00888-0>.
20. Stabile, S., F. Palermo, I. Bukreeva, D. Mele, V. Formoso, R. Bartolino, and A. Cedola. 2021. "A Computational Platform for the Virtual Unfolding of Herculaneum Papyri." *Scientific Reports* 11 (1): 1695. <https://doi.org/10.1038/s41598-020-80458-z>.
21. Tack, P., M. Cotte, S. Bauters, E. Brun, D. Banerjee, W. Bras, C. Ferrero, D. Delattre, V. Mocella, and L. Vincze. 2016. "Tracking Ink Composition on Herculaneum Papyrus Scrolls: Quantification and Speciation

- of Lead by X-Ray Based Techniques and Monte Carlo Simulations.” *Scientific Reports* 6: 20763. <https://doi.org/10.1038/srep20763>.
22. Tournié, A., K. Fleischer, I. Bukreeva, F. Palermo, M. Perino, et al. 2019. “Ancient Greek Text Concealed on the Back of Unrolled Papyrus Revealed through Shortwave-Infrared Hyperspectral Imaging.” *Science Advances* 5 (10): eaav8936. <https://doi.org/10.1126/sciadv.aav8936>.
 23. Xu, X., Y. Chang, J. An, and Y. Du. 2023. “Chinese Text Classification by Combining Chinese-BERTology-wwm and GCN.” *PeerJ Computer Science* 9: e1544. <https://doi.org/10.7717/peerj-cs.1544>.
 24. Zalamea, F. 2015. “Peirce’s Cenopythagorean Categories, Merleau-Ponty’s Chiasmatic Entrelacs and Grothendieck’s Résumé.” *Progress in Biophysics and Molecular Biology* 119 (3): 437–41. <https://doi.org/10.1016/j.pbiomolbio.2015.07.011>.
 25. Zhu, H., J. Xia, R. Liu, and B. Deng. 2025. “SPIRIT: Structural Entropy Guided Prefix Tuning for Hierarchical Text Classification.” *Entropy* 27 (2): 128. <https://doi.org/10.3390/e27020128>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Toward buoyancy-driven flow at Campi Flegrei: coupled phase change and asymmetric geometry

Arturo Tozzi (corresponding author)
 ASL Napoli 1 Centro, Distretto 27, Naples, Italy
 Via Comunale del Principe 13/a 80145
tozziarturo@libero.it

ABSTRACT

Bradyseism at Campi Flegrei is usually interpreted in terms of hydrothermal pressurization and magmatic degassing. Fluid flow, often treated as a passive response to pressure accumulation, is commonly modeled using simplified geometries and homogeneous permeability fields. We introduce a model in which phase transition, structural heterogeneity and geometric asymmetry jointly influence fluid flow and pressure distribution within a heterogeneous subsurface environment. We hypothesize that coupling among phase change, density gradients and flows may follow a mechanism similar to the self-propulsion observed in asymmetric floating bodies like melting ice blocks, where phase change generates buoyancy-driven currents along their inclined surfaces and net motion in the opposite direction. We simulate pressure evolution in a shallow gas-rich reservoir subject to time-dependent forcing and hydraulic relaxation, coupled to buoyancy-enhanced Darcy flow along prescribed preferential pathways. Our numerical simulations, grounded in reported deformation rates and seismicity depths at Campi Flegrei, reproduce temporal variations in uplift and the persistence of spatially localized flow. Within this framework, asymmetric geometry may promote channelized upward transport, while phase change may enhance buoyancy and contribute to pressure redistribution. Our model predicts nonlinear uplift acceleration, shallow localized seismicity and velocity scaling with pressure and buoyancy. Integration with existing multiphase models would enable the examination of how buoyancy-driven flows influence pressure evolution and deformation during volcanic unrest.

KEYWORDS: permeability; buoyancy; degassing; poroelasticity; convection.

INTRODUCTION

Bradyseism at Campi Flegrei is investigated through geodetic, geochemical and geophysical observations, with a focus on hydrothermal pressurization, magmatic degassing and poroelastic deformation of the shallow crust (Acocella 2021; Picozzi and Iaccarino 2021; Di Martino et al. 2022; Astort et al., 2024; Barone et al. 2025). Ground uplift is described as the response to pressure variations within a gas-rich reservoir, often located between two- and four-kilometers depth, fed by deeper magmatic sources (Martini et al. 2008; Guidoboni and Ciuccarelli 2011; Zaccarelli and Bianco 2017; Sabbarese et al. 2020; Giacomuzzi et al. 2024; Tramelli, Tan et al. 2025). Integration of multiphase fluid flow, thermal effects and mechanical deformation provides quantitative reconstructions of recent uplift episodes and seismicity patterns. These approaches often assume isotropic and homogeneous permeability fields, treating fluid migration primarily as a diffusive or pressure-driven process (Sbrana, Marianelli, and Pasquini 2021; Convertito, and Godano 2024; Chalumeau et al. 2024; Di Toro 2025). This perspective constraints the ability to capture spatially localized deformation, transient accelerations and the organization of fluid pathways observed in seismic tomography and resistivity data. Still, phase transitions like boiling and gas exsolution are incorporated as source terms without explicitly addressing their role in structuring flow patterns. As a result, the coupling between structural asymmetry, phase change and directed fluid motion is still unresolved.

We introduce a theoretical formulation in which geometric asymmetry and phase change jointly influence fluid flow and pressure distribution. We consider a shallow reservoir subject to time-dependent input from a deeper source and to internal transformations associated with gas exsolution and boiling. Pressure evolves according to a balance between source input and hydraulic relaxation, while fluid velocity is determined by permeability, viscosity, pressure gradients and buoyancy arising from density contrasts. Coupling among phase change, density gradients and directed flow could be viewed in analogy with the melting-driven self-propulsion observed in asymmetric floating bodies, where phase change induces buoyancy-driven currents and net motion through momentum exchange with the surrounding fluid (Berhanu et al. 2026). In our setting, preferential pathways are introduced in terms of geometric features representing heterogeneous permeability and structural discontinuities. Phase change enhances buoyancy and modifies local density, amplifying upward transport along these pathways and redistributing pressure within the reservoir. We provide numerical simulations based on recent monitoring data of Campi Flegrei to explore how transient forcing, relaxation timescales and pathway geometry could influence the temporal evolution of uplift and surface deformation's localization.

METHODS

We studied the recent unrest of Campi Flegrei through a quantitative model coupling reservoir pressurization, phase change, buoyancy-enhanced flow and surface deformation within an asymmetric subsurface geometry. We used published monitoring constraints on uplift, seismicity depth, reservoir structure and degassing to parameterize a time-dependent simulation. By combining temporal evolution of pressure with spatially organized flow and a simplified mechanical response, we tested whether the interaction between phase change and imposed geometric asymmetry could reproduce both the temporal variability and spatial localization observed during recent unrest.

Observational constraints and quantitative parameterization. The parameter space was constrained using recent observational data (De Landro et al. 2025; Rapagnani et al. 2025). The February 2026 surveillance bulletin reported 254 earthquakes, with 192 located events concentrated within the upper 3 km and maximum depth of 4.5 km (INGV-Osservatorio Vesuviano 2026). Uplift rates were approximately 15 ± 3 mm month⁻¹ from April to early October 2025, increased to 25 ± 3 mm month⁻¹ after October 2025 and decreased to about 10 ± 3 mm month⁻¹ by early February 2026. The cumulative uplift at Rione Terra reached 162.5 cm since November 2005, with 24.5 cm since January 2025. Structural models indicate a shallow gas-rich reservoir between 2 and 4 km, a caprock between 1 and 2 km and deeper magmatic sources below 4 km. Weekly geochemical monitoring reported fumarolic temperatures near 95°C at Pisciarelli and about 173°C at Solfatara. These quantities were used to delimit physically plausible ranges for temperature, density contrast and fluid properties.

Pressure evolution and source-term representation. The shallow reservoir was represented through a time-dependent excess pressure $P(t)$ in MPa. Its evolution was described by the first-order differential equation

$$\frac{dP}{dt} = \frac{S(t) - P(t)}{\tau_h},$$

where $S(t)$ is an effective source term and τ_h is a hydraulic relaxation time. This formulation assumes that pressure increases due to deep input and internal phase transitions and decreases through leakage and redistribution. The source term was written as

$$S(t) = P_b + P_1 \exp\left(-\frac{t - t_0}{\tau_d}\right),$$

where P_b is a background level, P_1 is a transient amplitude and τ_d is a decay timescale. Parameters were set to $P_b = 1.00$ MPa, $P_1 = 1.50$ MPa, $\tau_h = 4.0$ months and $\tau_d = 3.5$ months, with t_0 corresponding to October 2025. Numerical integration was performed using an explicit Euler scheme with time step $\Delta t = 1$ month:

$$P_{n+1} = P_n + \Delta t \frac{S_n - P_n}{\tau_h}.$$

This equation compresses multiple processes, including degassing and boiling, into a single effective forcing and does not explicitly model thermodynamic phase equilibria (Danesi et al. 2024; Hainzl, Dahm and Tramelli 2026).

Darcy–buoyancy flow and phase-change contribution. Fluid motion was described using a Darcy-type relation combining pressure gradient and buoyancy effects:

$$u(t) = \frac{k}{\mu} \left(\frac{P(t) \times 10^6}{L} + \Delta\rho g \right).$$

Here k is permeability, μ is dynamic viscosity, L is the vertical extent of the flow path, $\Delta\rho$ is the density contrast induced by phase change and g is gravitational acceleration. The pressure term accounts for forced flow, while the buoyancy term represents the effect of vapor exsolution and heating. Values used were $k = 2 \times 10^{-14}$ m², $\mu = 1.5 \times 10^{-5}$ Pa s, $L = 3000$ m and $\Delta\rho = 250$ kg m⁻³. The resulting velocity was converted from m s⁻¹ to m day⁻¹. This linear combination neglects multiphase interactions, compressibility and feedback between flow and permeability and should therefore be interpreted as a simplified representation of transport (Tan et al. 2025; De Landro et al. 2025).

Surface deformation and calibration procedure. Surface uplift $w(t)$ was linked to pressure through

$$\frac{dw}{dt} = aP(t) + b,$$

with $a = 0.8333$ mm month⁻¹ MPa⁻¹ and $b = 9.1667$ mm month⁻¹. The discrete update equation was

$$w_{n+1} = w_n + \frac{aP_n + b}{10},$$

where division by 10 converts millimeters to centimeters. The initial condition was $w(t_0) = 151.0$ cm. This relation is calibrated to reproduce observed uplift rates rather than derived from elastic theory. A secondary diagnostic for seismicity was introduced:

$$\dot{N}(t) = N_0 \exp [\beta(w(t) - w_0)],$$

with $N_0 = 18$ events month⁻¹ and $\beta = 0.06$ cm⁻¹. This variable was used to verify qualitative consistency between deformation and seismic activity (Hainzl, Dahm and Tramelli 2026; INGV-Osservatorio Vesuviano 2024).

Geometric construction of asymmetric flow pathways. To represent structural heterogeneity, a two-dimensional section was defined with horizontal coordinate $x \in [-3, 3]$ km and depth $z \in [0, 5]$ km. A curved pathway was prescribed as

$$x_c(z) = -0.6 + 0.45(5 - z) - 0.06(5 - z)^2,$$

with width

$$\sigma(z) = 0.28 + 0.06(5 - z).$$

The velocity field was constructed as

$$V(x, z) = V_0 \exp \left[-\frac{(x - x_c(z))^2}{2\sigma(z)^2} \right] \left[0.35 + 0.65 \exp \left(-\frac{(z - 2.5)^2}{3.0} \right) \right].$$

This formulation produces a localized high-velocity corridor connecting deeper and shallower regions. The geometry is imposed and represents a hypothesis about preferential pathways rather than a reconstruction derived from data (Giudicepietro et al. 2025).

Nonlinear interaction between asymmetry and phase change. To explore the interaction between structural asymmetry and phase change, we defined two dimensionless parameters A (asymmetry index) and B (phase-change intensity). Uplift acceleration was expressed as

$$\alpha(A, B) = C A^{3/2} B^2,$$

with $C = 40$ mm month⁻². This expression is not derived from physical principles but constructed to test whether nonlinear coupling produces distinct regimes. The resulting surface was evaluated over $A, B \in [0, 1]$. This step provides a hypothesis-generating exploration rather than a validated prediction.

All computations were performed in Python using NumPy for numerical arrays, pandas for time handling and Matplotlib for visualization. Temporal integration covered monthly intervals from October 2025 to February 2026. Spatial fields were evaluated on regular grids. Units were maintained explicitly. The workflow consisted of parameter selection, time integration of pressure, computation of velocity and mapping of spatial distribution. Since no automated inversion, uncertainty quantification or statistical fitting was performed, our outputs should be interpreted as representative simulations illustrating possible system behavior rather than validated reconstructions.

RESULTS

Our simulated pressure increases briefly and then relaxes over the imposed hydraulic timescale. This produces an uplift rate that decreases from higher to lower values over a few months. Within the explored parameter range, the relationship between pressure and uplift rate is approximately linear, with uplift rates decreasing smoothly as pressure declines. Upward fluid velocity, computed from combined pressure gradient and buoyancy contributions, shows a monotonic dependence on pressure, confirming that both contributions act coherently under our formulation. The vertical section shows that flow is not uniformly distributed but is concentrated along a curved pathway connecting deeper and shallower regions (Figure 1). This localization persists throughout the simulated temporal evolution, indicating that spatial organization could be controlled primarily by the imposed geometry rather than by instantaneous pressure amplitude. Overall, the combined evolution of pressure decay, declining uplift rates and persistent flow localization points towards the possible coexistence of temporal variability and spatial organization within our simplified formulation.

Interactions between geometric asymmetry and phase change. The two-parameter exploration of geometric asymmetry and phase-change intensity provides a model-based prediction of a nonlinear response in uplift acceleration (Figure 2). For low values of either parameter, uplift acceleration is close to zero across the explored domain, indicating that neither asymmetry nor phase change alone produces significant dynamic amplification. When both parameters increase simultaneously, the system is predicted to transition to a regime characterized by rapidly increasing uplift acceleration, with values rising sharply beyond intermediate thresholds. The resulting surface exhibits a convex shape, indicating superlinear coupling between the two variables rather than additive behavior. The absence of intermediate regimes with moderate acceleration suggests that the system may organize into distinct dynamical states under our

assumptions. This prediction identifies a potentially discriminable signature, i.e., a transition from gradual to accelerated deformation as both structural asymmetry and phase-change intensity increase, which can be tested against observational data.

Overall, our simulations suggest that pressure evolution, phase change and geometric asymmetry can be combined in a reduced quantitative description to assess time-dependent uplift and spatially localized flow.

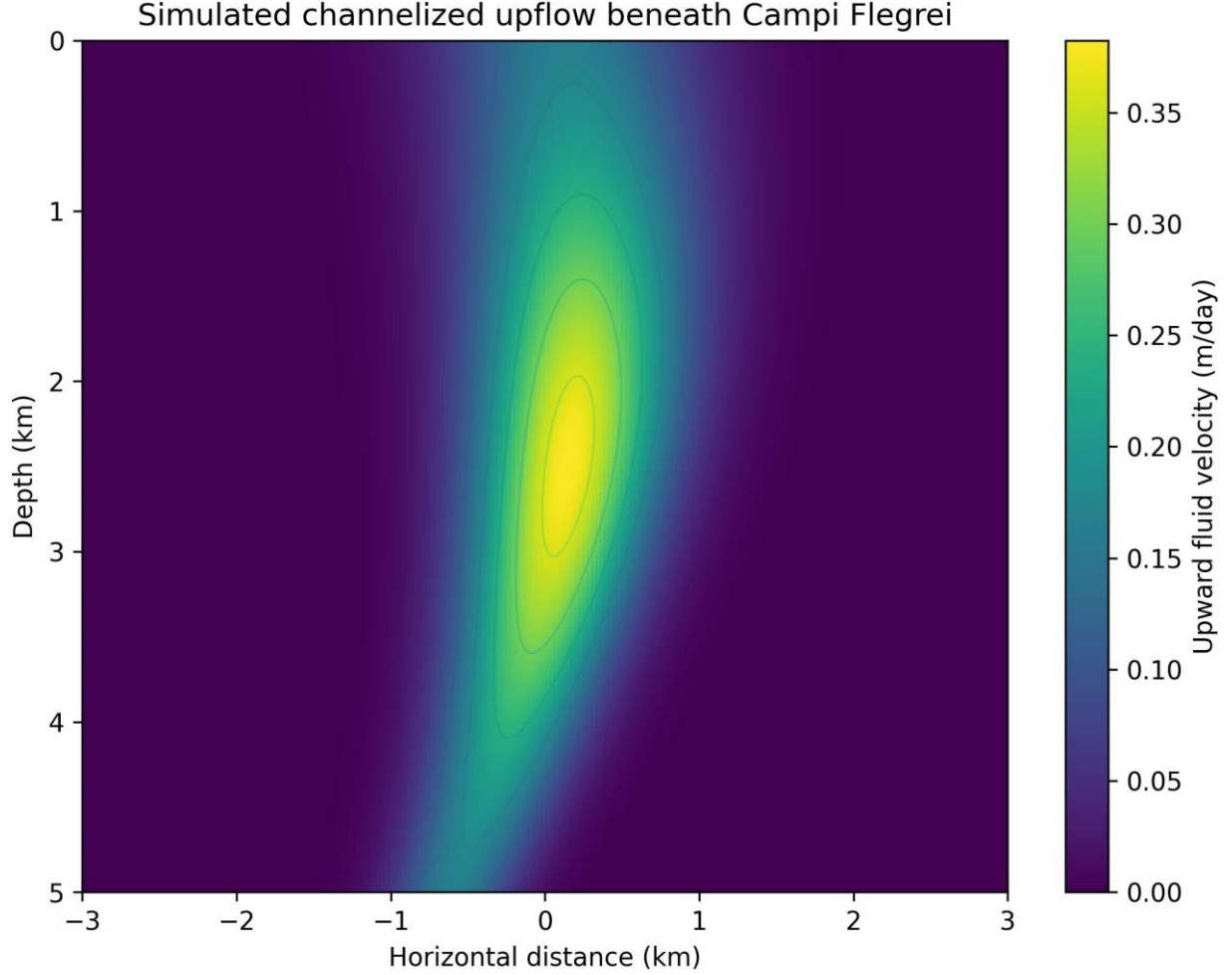


Figure 1. Simulated vertical section of channelized upflow beneath Campi Flegrei. The horizontal axis represents lateral distance in kilometers across the caldera, while the vertical axis indicates depth in kilometers from the surface to approximately 5 km. Color shading and superimposed contours quantify upward fluid velocity in meters per day, computed from a permeability–buoyancy formulation combining pressure gradients and density contrasts associated with phase change. The curved high-velocity corridor delineates a preferential pathway linking a deeper source region at approximately 4–5 km depth to a shallower gas-rich reservoir at approximately 2–3 km depth. The increase in velocity toward shallower depths, attributed to the combined effects of pressure-driven flow and buoyancy, is proposed as a hypothesis within our model and not directly inferred from empirical data.

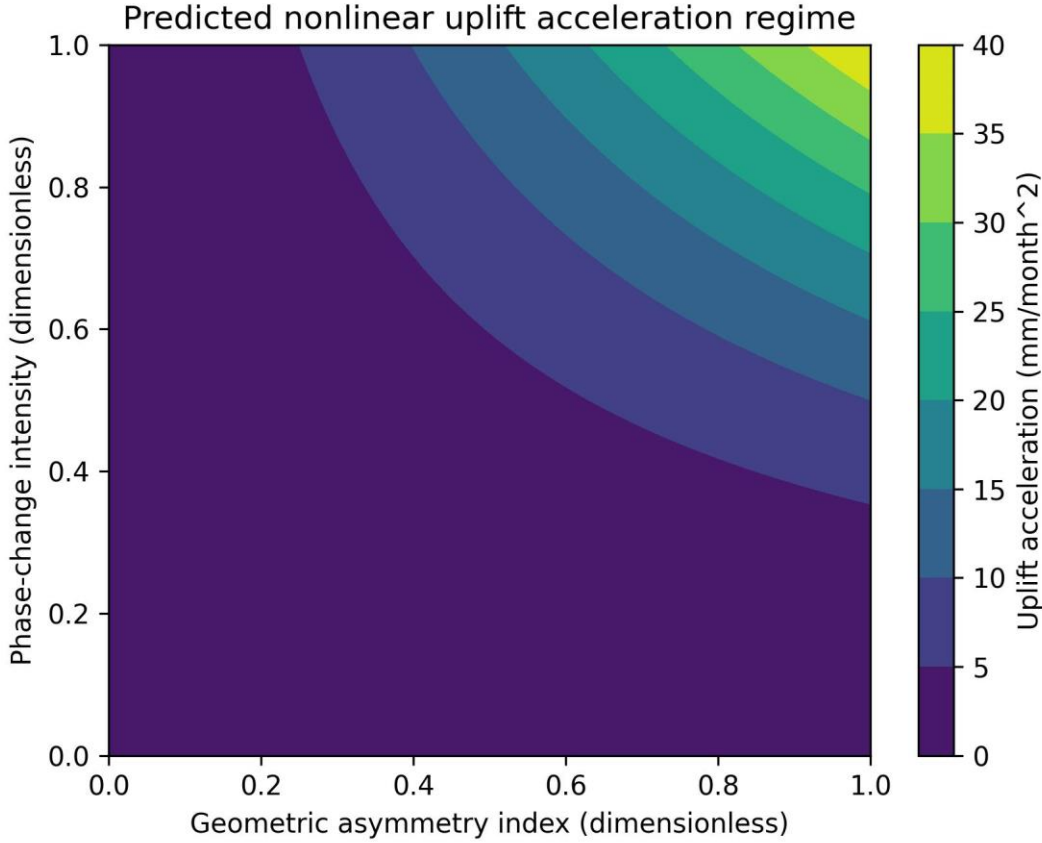


Figure 2. Hypothesized interaction between geometric asymmetry and phase-change intensity based on our model assumptions. Contour plot of predicted uplift acceleration (mm/month^2) as a function of geometric asymmetry (dimensionless) and phase-change intensity (dimensionless), representing a model prediction. Uplift acceleration is low when either parameter is small, but increases rapidly when both exceed intermediate values. This cooperative amplification departs from additive behavior commonly assumed in hydrothermal models. The convex surface indicates a superlinear response, suggesting a transition between gradual and accelerated deformation.

CONCLUSIONS

We asked whether bradyseism at Campi Flegrei can be interpreted not only as pressure accumulation, but also as directed fluid motion arising from the interaction between phase change and geometric asymmetry. This approach is motivated by a physical observation: an asymmetric ice block melts unevenly, producing dense water that flows along its surface and generates a force that drives motion in the opposite direction. This illustrates how phase change and asymmetry can organize fluid motion. Therefore, we isolate a minimal set of variables linking flow organization to measurable quantities like uplift rate and depth distribution of activity. The coupled effect of asymmetry and phase change allows discrimination between gradual and accelerated deformation regimes within the same parameter space.

Our simulations compare time-dependent pressure evolution, buoyancy-driven flow and surface deformation within ranges constrained by monitoring data. A transient pressure increase followed by hydraulic relaxation reproduces observed uplift-rate variations, while asymmetric permeability leads to persistent localization of upward flow. In our simulations, fluid migration is organized along preferential paths connecting deeper and shallower regions rather than being uniform. The coupling between pressure gradients and density contrasts sustains coherent upward velocities even as pressure declines. In addition, the interaction between asymmetry and phase change produces nonlinear transitions in uplift dynamics. These results indicate that deformation at Campi Flegrei may reflect organized fluid transport controlled by the coupled effects of geometry and phase change.

Our study has limitations. The governing equations are not derived from principles of multiphase thermo-poroelasticity. The source term $S(t)$ and the linear relation between pressure and uplift are introduced as calibrated components rather than independently validated quantities. The Darcy–buoyancy formulation assumes a linear combination of pressure gradient and density contrast, neglecting two-phase flow effects, compressibility and time-dependent permeability, all of

which are relevant in the context of Campi Flegrei. The representation of geometric asymmetry relies on an imposed preferential pathway rather than one inferred from geophysical inversion. The nonlinear interaction is built through parameter choice and should be interpreted as a testable hypothesis rather than a derived law. Open questions concern how permeability evolves under repeated pressurization, how multiphase interactions modify effective density contrasts and how transient structural changes alter pathway stability.

Our model yields testable previsions. First, if phase change and geometric asymmetry jointly control deformation, then uplift acceleration should display a nonlinear dependence on indicators of fluid input and structural focusing. Specifically, the time derivative of uplift rate is expected to increase superlinearly when both gas flux and inferred pathway anisotropy rise, whereas it should remain near zero if only one factor varies.

Second, upward fluid velocity along preferential zones should scale with pressure according to a linear-plus-buoyancy relation, implying that geophysical proxies of fluid movement, such as seismic attenuation or resistivity anomalies, should increase proportionally to estimated overpressure within a defined range.

Third, the depth distribution of seismicity should remain confined within the upper three to four kilometers when flow is strongly localized, with a measurable concentration gradient aligned with the inferred pathway. These predictions can be tested through time-resolved comparisons between uplift rate, gas emissions and subsurface imaging.

Practical implications concern the integration of physically interpretable variables into monitoring strategies. The explicit computation of pathway-dependent velocity fields enables the construction of indicators sensitive to changes in subsurface transport efficiency. This may support the development of diagnostic thresholds based on combined geodetic and geochemical signals, allowing classification of unrest phases according to internal system organization rather than solely on amplitude measures. Our approach can also shape the design of monitoring networks, by identifying areas where measurements are most likely to capture early changes in subsurface dynamics. In addition, our formulation facilitates rapid updating of model states as new data become available, enabling continuous tracking of system evolution with limited computational cost.

In conclusion, we show that deformation, fluid migration and internal structure can be assessed within a unified quantitative description, albeit within a reduced representation. Our approach suggests that subsurface processes' organization, influenced by structural heterogeneity, geometric asymmetry and phase change, may jointly influence fluid migration at Campi Flegrei. This perspective provides an interpretation of temporal variability and spatial localization, while remaining compatible with existing multiphase and poroelastic descriptions.

DECLARATIONS

Ethics approval and consent to participate. This research does not contain any studies with human participants or animals performed by the Author.

Consent for publication. The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Availability of data and materials. All data and materials generated or analyzed during this study are included in the manuscript. The Author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Disclaimer. The views expressed are those of the author and do not necessarily reflect those of the affiliated institutions.

Competing interests. The Author does not have any known or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence or be perceived to influence their work.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Acknowledgements: none.

Authors' contributions. The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work, the author used ChatGPT 5.3 to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

REFERENCES

- 1) Acocella, Valerio. 2021. *Volcano-Tectonic Processes*. Advances in Volcanology. Cham: Springer. <https://doi.org/10.1007/978-3-030-65968-4>
- 2) Astort, A., E. Trasatti, L. Caricchi, et al. 2024. "Tracking the 2007–2023 Magma-Driven Unrest at Campi Flegrei Caldera (Italy)." *Communications Earth & Environment* 5: 506. <https://doi.org/10.1038/s43247-024-01665-4>
- 3) Barone, Andrea, Gianluca Gola, Stefano Caliro, Giovanni Chiodini, Pietro Tizzani, and Raffaele Castaldo. 2025. "Long-Term Thermo-Fluid Dynamic Modeling of Solfatara Hydrothermal System, Campi Flegrei Caldera." *Journal of Volcanology and Geothermal Research* 459: 108277. <https://doi.org/10.1016/j.jvolgeores.2025.108277>
- 4) Berhanu, Michael, Amit Dawadi, Martin Chaigne, Jérôme Jovet, and Arshad Kudrolli. 2026. "Self-Propulsion of Floating Ice Blocks Caused by Melting in Water." *Physical Review Fluids* 11: 033802. <https://doi.org/10.1103/k6gl-1zqm>
- 5) Chalumeau, C., H. Agurto-Detzel, A. Rietbrock, et al. 2024. "Seismological Evidence for a Multifault Network at the Subduction Interface." *Nature* 628: 558–562. <https://doi.org/10.1038/s41586-024-07245-y>
- 6) Danesi, Sara, Christopher R. J. Kilburn, Giuseppe De Natale, et al. 2024. "Evolution in Unrest Processes at Campi Flegrei Caldera as Inferred from Seismicity and Ground Deformation." *Earth and Planetary Science Letters* 630: 118600
- 7) De Landro, G., T. Vanorio, T. Muzellec, et al. 2025. "3D Structure and Dynamics of Campi Flegrei Enhance Multi-Hazard Assessment." *Nature Communications* 16: 4814. <https://doi.org/10.1038/s41467-025-59821-z>
- 8) Di Martino, Maria Del Pilar, Luca De Siena, Vincenzo Serlenga, and Grazia De Landro. 2022. "Reconstructing Hydrothermal Fluid Pathways and Storage at the Solfatara Crater (Campi Flegrei, Italy) Using Seismic Scattering and Absorption." *Frontiers in Earth Science* 10. <https://doi.org/10.3389/feart.2022.852510>
- 9) Di Toro, G. 2025. "Earthquake Rupture Velocity and Stress Drop Interaction in the Campi Flegrei Volcanic Caldera." *Communications Earth & Environment*. <https://doi.org/10.1038/s43247-025-02808-x>
- 10) Giacomuzzi, G., C. Chiarabba, F. Bianco, P. De Gori, and N. Piana Agostinetti. 2024. "Tracking Transient Changes in the Plumbing System at Campi Flegrei Caldera." *Earth and Planetary Science Letters* 637: 118744. <https://doi.org/10.1016/j.epsl.2024.118744>
- 11) Giudicepietro, Flora, Rosario Avino, Eliana Bellucci Sessa, Andrea Bevilacqua, Manuela Bonano, Stefano Caliro, et al. 2025. "Burst-Like Swarms in the Campi Flegrei Caldera Accelerating Unrest from 2021 to 2024." *Nature Communications* 16 (1): 1548. <https://doi.org/10.1038/s41467-025-56723-y>
- 12) Guidoboni, Emanuela, and Caterina Ciuccarelli. 2011. "The Campi Flegrei Caldera: Historical Revision and New Data on Seismic Crises, Bradyseisms, the Monte Nuovo Eruption and Ensuing Earthquakes (Twelfth Century–1582 AD)." *Bulletin of Volcanology* 73: 655–677. <https://doi.org/10.1007/s00445-010-0430-3>
- 13) Hainzl, S., T. Dahm, and A. Tramelli. 2026. "A Deformation-Driven Earthquake Interaction Model for Seismicity at Campi Flegrei." *Communications Earth & Environment* 7: 244
- 14) INGV. 2024. "Campi Flegrei: The Relationship between Ground Uplift and Seismic Activity Has Been Quantified." Press release, December 3, 2024
- 15) INGV-Osservatorio Vesuviano. 2026. "Campi Flegrei Monthly and Weekly Surveillance Bulletins, February 2026"
- 16) Martini, M., F. Giudicepietro, L. D'Auria, M. Orazi, G. Borriello, C. Buonocunto, et al. 2008. "Seismological Monitoring of Campi Flegrei Caldera." Presented at the European Geosciences Union General Assembly 2008, Vienna, Austria. <https://www.earth-prints.org/handle/2122/4168>
- 17) Picozzi, Matteo, and Antonio Giovanni Iaccarino. 2021. "Forecasting the Preparatory Phase of Induced Earthquakes by Recurrent Neural Network." *Forecasting* 3 (1): 17–36. <https://doi.org/10.3390/forecast3010002>
- 18) Rapagnani, Giacomo, Simone Cesca, Gilberto Saccorotti, Gesa Petersen, Torsten Dahm, Francesca Bianco, and Francesco Grigoli. 2025. "Coupled Earthquakes and Resonance Processes during the Uplift of Campi Flegrei Caldera." *Communications Earth & Environment* 6 (1): 607. <https://doi.org/10.1038/s43247-025-02604-7>
- 19) Sabbarese, C., F. Ambrosino, G. Chiodini, et al. 2020. "Continuous Radon Monitoring during Seven Years of Volcanic Unrest at Campi Flegrei Caldera (Italy)." *Scientific Reports* 10: 9551. <https://doi.org/10.1038/s41598-020-66590-w>
- 20) Sbrana, A., P. Marianelli, and G. Pasquini. 2021. "The Phlegrean Fields Volcanological Evolution." *Journal of Maps* 17 (2): 557–570. <https://doi.org/10.1080/17445647.2021.1982033>
- 21) Tan, Xing, Anna Tramelli, Sergio Gammaldi, Gregory C. Beroza, William L. Ellsworth, and Warner Marzocchi. 2025. "A Clearer View of the Current Phase of Unrest at Campi Flegrei Caldera." *Science* 390 (6768): 70–75. <https://doi.org/10.1126/science.adw9038>
- 22) Tramelli, A., V. Convertito, and C. Godano. 2024. "b Value Enlightens Different Rheological Behaviour in Campi Flegrei Caldera." *Communications Earth & Environment* 5: 275. <https://doi.org/10.1038/s43247-024-01447-y>
- 23) Zaccarelli, Lucia, and Francesca Bianco. 2017. "Noise-Based Seismic Monitoring of the Campi Flegrei Caldera." *Geophysical Research Letters* 44. <https://doi.org/10.1002/2016GL072477>

Article

Not peer-reviewed version

A Bruhat-Tits Ultrametric Logic for Hierarchical Inference

[Arturo Tozzi](#) *

Posted Date: 6 August 2026

doi: 10.20944/preprints202608.0449.v1

Keywords: ultrametric; refinement; persistence; hierarchy; inference



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

A Bruhat-Tits Ultrametric Logic for Hierarchical Inference

Arturo Tozzi

ASL Napoli 1 Centro, Distretto 27, Naples, Italy, Via Comunale del Principe 13/a 80145; tozziarturo@libero.it

Abstract

Extensions of classical logic like Kripke models, branching-time semantics and proof trees introduce hierarchical structures, yet inference is independent of the hierarchy itself. We ask: can hierarchy become the mathematical principle from which inference is generated, rather than merely organizing logical states? p-adic geometry provides a natural setting for addressing this question by replacing Euclidean proximity with hierarchical proximity defined by shared ancestry. Its canonical geometric realization, the Bruhat-Tits tree, describes information as a sequence of successive refinements, therefore providing a natural substrate for hierarchical logical evolution. Building on this structure, we introduce an ultrametric logic in which propositions correspond to progressively refined informational states, logical proximity is determined by ultrametric distance and negation is interpreted as divergence between refinement branches. We evaluated this formulation through mathematical simulations on a finite Bruhat-Tits tree, quantifying proposition clustering, refinement trajectories, ultrametric neighborhoods and persistence during successive refinement steps. Our simulations generated coherent hierarchical organization, stable proposition classes and increasing inferential affinity among propositions sharing longer refinement histories, supporting a geometric interpretation of logical inference driven by hierarchical refinement.

Keywords: ultrametric; refinement; persistence; hierarchy; inference

Introduction

Classical logic assigns truth values to propositions and inference rules preserving validity under well-defined semantic or syntactic conditions. Since Aristotle, binary truth and the principles of non-contradiction and excluded middle have provided the conceptual basis for most logical systems, while subsequent developments have broadened rather than replaced this foundation (Parry and Hacker 1991; Rothhaar 2020; De Rizzo 2025; Decker 2025; King and Raspa 2025). Predicate logic extended propositional reasoning through quantification, modal logic incorporated necessity and possibility, intuitionistic logic reformulated truth in constructive terms (Tarannum and Jabin 2022; Brunet 2023; Hamdy et al. 2024; Ciardelli 2025; Gao et al. 2025; Litland 2025; Montambault et al. 2025; von Plato 2025). Kripke semantics introduced partially ordered structures to represent evolving information, branching-time logics modeled multiple possible futures, while paraconsistent systems investigated controlled forms of contradiction (Oshiyama et al. 2012; Da Silva Filho et al. 2016; Da Silva Filho et al. 2021; Bonzio et al. 2023; Olkhovikov 2024; Fernández-Vilas et al. 2026).

In parallel, mathematical logic has increasingly adopted geometric and topological perspectives, including sheaf-theoretic semantics, categorical logic and domain theory, in which relationships among propositions are interpreted through structures richer than simple truth assignments. Nevertheless, geometry serves as a semantic or organizational framework rather than as an active determinant of inference. Tree-like structures are employed to represent proofs, accessibility relations or temporal branching, but their metric properties do not influence the logical process itself (Sgurev 2021; Sgurev 2022; Dedhe et al. 2023; Wang et al. 2025). Likewise, hierarchical representations in computer science and artificial intelligence primarily organize information or computation, while inference is governed by syntactic or semantic rules independent of the underlying geometry.

Consequently, existing logical frameworks lack an intrinsic notion of distance between propositions, gradual refinement governed by a geometric metric or inferential strength emerging directly from hierarchical organization.

We introduce a logical framework in which the geometry of the Bruhat-Tits tree becomes an integral component of logical reasoning rather than a descriptive representation. The Bruhat-Tits tree, arising naturally in p-adic geometry, provides an ultrametric hierarchical space in which proximity is determined by the depth of shared ancestry instead of Euclidean distance (Zabrodin 1989; Chen, Liu, and Hung 2021; Ludwig and Merten 2026). We reinterpret this structure as a logical environment where propositions are not regarded as isolated entities endowed with fixed truth values, but as successive refinements of more general informational states. Each refinement generates increasingly specific propositions while preserving their genealogical relationships within the hierarchy. Logical proximity is established by the amount of refinement history shared by two propositions, allowing similarity, inferential transfer and proposition stability to acquire precise geometric meanings. In this formulation, negation is interpreted as divergence between refinement branches rather than as an elementary complement, whereas logical robustness depends on persistence across successive refinements instead of exclusively on binary evaluation.

To illustrate how ultrametric geometry could organize inference and reveal structural properties that remain inaccessible in logical systems lacking an explicit hierarchical metric, we perform a qualitative simulation on a finite approximation of a Bruhat-Tits tree in which propositions evolve through iterative refinements. Our simulation follows the formation of hierarchical proposition clusters, the emergence of refinement trajectories, the evolution of ultrametric neighborhoods and the persistence of propositions across increasing refinement depth.

We will proceed as follows. First, we introduce the mathematical apparatus of Bruhat-Tits trees and ultrametric geometry together with our proposed logical framework. Then, we describe the qualitative simulation and analyze the resulting refinement structures. Finally, we discuss the theoretical implications of ultrametric inference, compare the framework with existing logical systems and outline future research directions.

Mathematical Frameworks of Bruhat-Tits Trees, Ultrametric Geometry and Hierarchical Refinement

Our logical framework is based on the hypothesis that logical inference can be embedded in a hierarchical geometric space rather than restricted to symbolic manipulations over isolated propositions. Instead of assuming that every proposition P has a truth value $T(P) \in \{\top, \perp\}$, we consider propositions as occupying positions within a progressively refined informational hierarchy. This hierarchy is modeled by the Bruhat-Tits tree, a canonical object of p-adic geometry whose ultrametric organization represents successive refinements of information. Consequently, reasoning is interpreted as navigation through this hierarchy rather than as the evaluation of propositions in a flat logical space.

The mathematical basis originates from the p-adic number system. In Euclidean geometry, the distance between two points x and y is measured by

$$d_E(x, y) = |x - y|,$$

where proximity depends upon physical or numerical separation. In contrast, p-adic geometry introduces a completely different metric. Given a prime number p , every non-zero integer can be decomposed into powers of p . The larger the common power shared by two numbers, the closer they become. Their distance is defined as

$$d_p(x, y) = p^{-v_p(x-y)},$$

where $v_p(x - y)$ denotes the highest exponent of p dividing the difference $x - y$. Unlike Euclidean distance, this metric measures hierarchical similarity rather than spatial separation. Two objects are close because they share a long common refinement history, not because they occupy neighboring positions in continuous space.

The p-adic metric satisfies the ultrametric inequality

$$d(x, z) \leq \max\{d(x, y), d(y, z)\},$$

which is stronger than the ordinary triangle inequality

$$d(x, z) \leq d(x, y) + d(y, z).$$

Balls become nested rather than overlapping, every point inside a ball is simultaneously one of its centers and neighborhoods naturally organize themselves into recursive hierarchies. Consequently, ultrametric spaces are intrinsically tree-like, making them particularly appropriate whenever successive refinements are more fundamental than continuous variation.

The Bruhat-Tits tree provides one of the canonical realizations of this ultrametric geometry. Mathematically, its vertices correspond to equivalence classes of lattices over a p-adic field, while edges represent elementary transformations between neighboring classes. We do not use this algebraic interpretation directly, but rather reinterpret the Bruhat-Tits tree as an epistemic space, where vertices correspond to informational states and edges represent elementary refinement operations. The mathematical structure is preserved, whereas its interpretation changes from arithmetic to logic.

Let

$$T = (V, E)$$

denote the Bruhat-Tits tree, where V is the set of vertices and E the set of edges. A mapping

$$\Phi: \mathcal{P} \rightarrow V$$

associates every logical proposition $P \in \mathcal{P}$ with a vertex of the tree. Unlike classical logic, however, $\Phi(P)$ does not merely indicate the identity of the proposition. It specifies its position within the hierarchy of refinements.

The root

$$v_0 \in V$$

represents the least refined informational state. At this level, only coarse distinctions are available and many propositions cannot yet be assigned a definite truth value. Formally,

$$T(P) = \text{undefined}$$

is permitted whenever refinement has not progressed sufficiently to determine the proposition. Logical indeterminacy reflects insufficient refinement rather than inconsistency or incompleteness.

Each edge

$$(v_i, v_j) \in E$$

represents an elementary refinement operator

$$R: v_i \rightarrow v_j,$$

which introduces additional distinctions while preserving all information already contained in v_i . Consequently,

$$R^n(v_0)$$

represents the informational state obtained after n successive refinements.

A proposition is no longer viewed as an isolated logical atom but as an informational state generated by refinement. If

$$P_i \rightarrow P_j$$

corresponds to refinement rather than implication, then P_j specializes P_i while inheriting its informational content.

Logical proximity follows directly from geometry. Let

$$A(P, Q)$$

denote the deepest common ancestor of propositions P and Q . Their logical distance is defined by

$$d_L(P, Q) = p^{-h(A(P, Q))},$$

where $h(A)$ denotes the depth of the common ancestor measured from the root. Thus,

$$d_L(P, Q) < d_L(P, R)$$

whenever P and Q share a longer refinement history than P and R . Logical similarity becomes an intrinsic geometric property rather than a syntactic or semantic comparison.

Inference itself becomes geometry dependent. Classical logic defines derivability through

$$\Gamma \vdash P,$$

where Γ denotes a collection of premises. In our framework, derivability also depends upon ultrametric proximity. We write

$$\Gamma \vdash_d P,$$

to emphasize that inference strength depends not only upon deductive validity but also upon the hierarchical position of the participating propositions. Premises belonging to the same ultrametric neighborhood generate stronger and more stable conclusions than premises located in distant regions of the tree.

Negation undergoes an equally substantial reinterpretation. In classical logic,

$$P \wedge \neg P = \perp,$$

and

$$P \vee \neg P = \top.$$

Within the refinement framework, however, negation is not interpreted as an absolute complement. Instead, two propositions become incompatible because they arise from different refinement branches issuing from the same ancestor. If

$$R_1(P_0) = P_1, R_2(P_0) = P_2,$$

with $R_1 \neq R_2$, then P_1 and P_2 stand for divergent refinements of the same informational state. Opposition emerges from branching rather than from complementation. Logical incompatibility is expressed by

$$P_1 \perp P_2,$$

instead of requiring

$$P_2 = \neg P_1.$$

Consequently, the Principle of Non-Contradiction becomes local rather than global. Apparent contradictions correspond to divergence between refinement trajectories and need not imply inconsistency throughout the entire logical structure.

Truth itself becomes refinement dependent. Let

$$T(P, n)$$

denote the evaluation of proposition P after refinement depth n . Then

$$T(P, n) = \begin{cases} \text{undefined,} & n < n_c, \\ \top \text{ or } \perp, & n \geq n_c, \end{cases}$$

where n_c denotes the critical refinement depth necessary to evaluate P . Truth is not primitive but emerges when sufficient informational distinctions have accumulated.

We also introduce the notion of logical persistence. Let

$$\Pi(P)$$

denote the persistence of proposition P , defined as the number of consecutive refinement levels through which P remains valid. Highly persistent propositions satisfy

$$\Pi(P) \gg 1,$$

whereas unstable propositions exhibit

$$\Pi(P) \approx 1.$$

Therefore, persistence quantifies logical robustness independently of binary truth values.

Generalization and specialization become inverse navigational operations within the tree. Descending corresponds to repeated refinement

$$v_0 \rightarrow v_1 \rightarrow \cdots \rightarrow v_n,$$

whereas ascending identifies the least common ancestor

$$\text{LCA}(P, Q),$$

which represents the most specific proposition shared by both descendants. Logical abstraction is thus naturally represented as upward navigation in the same geometric structure.

Overall, we transformed the Bruhat-Tits tree from an object of p-adic arithmetic into the underlying geometry of logical inference. Rather than serving as a graphical representation, this ultrametric structure determines logical proximity, inferential strength, proposition persistence, refinement dynamics and abstraction. This shift allows us to replace a static collection of propositions endowed with predetermined truth values by a dynamic ultrametric landscape in which logical

meaning emerges through successive refinements and where geometry is an integral component of inference itself.

Qualitative Simulation of Ultrametric Logical Refinement

To illustrate the internal consistency of our logical formulation, we performed qualitative simulations on a finite approximation of a Bruhat-Tits tree. Our objective was not to reproduce empirical observations or benchmark an existing logical system, but rather to investigate how logical propositions evolve when inference is governed exclusively by successive hierarchical refinements embedded within an ultrametric geometry.

Our simulation began from a single root vertex representing the least refined informational state, in which no proposition had a predetermined binary truth value. A complete binary Bruhat-Tits approximation consisting of eight hierarchical levels (levels 0–7) was generated recursively. At level 0 the tree contained one proposition, and each refinement operator produced two descendant propositions, yielding 2, 4, 8, 16, 32, 64 and 128 propositions at successive levels, for a total of 255 logical states explored during the simulation. Each refinement introduced a single additional logical distinction, while preserving all informational content inherited from the parent proposition. Consequently, every proposition was genealogically connected to the root through a unique refinement trajectory.

Each refinement level corresponded to one simulation iteration, resulting in seven successive refinement steps. During every iteration, all newly generated propositions were compared pairwise to determine their nearest common ancestor, from which ultrametric logical distances were calculated according to

d(P_i, P_j) = 2^{-h(A_{ij})},

where $h(A_{ij})$ denotes the depth of the nearest common ancestor shared by propositions P_i and P_j . Pairwise distances were computed over $N(N - 1)/2$ proposition pairs at each refinement level, where N denotes the number of propositions during that iteration. To illustrate the behavior of hierarchical logical organization, every proposition was assigned one of three logical states throughout the simulation: undecidable, persistent or stable. Initially, all propositions were classified as undecidable. Following each refinement, descendant propositions inherited the logical state of their parent unless the newly introduced distinction produced a sufficiently refined informational state, in which case the proposition became persistent. Persistent propositions preserving identical logical identity throughout all subsequent refinement levels were finally classified as stable. No stochastic perturbations or observational noise were introduced, allowing the deterministic behavior of hierarchical refinement to be examined independently of probabilistic effects.

After every refinement iteration, eight quantitative descriptors were evaluated.

1. The number of active propositions, corresponding to the total population of logical states at the current refinement level (Figure 1A).
2. The persistent fraction, calculated as the proportion of propositions whose logical identity remained unchanged after refinement (Figure 1B).
3. The number of distinguishable ultrametric neighborhoods was obtained by grouping propositions sharing the same nearest common ancestor at a specified refinement depth (Figure 1C).
4. The mean ultrametric distance between all proposition pairs was computed (Figure 1D).
5. The fraction of undecidable propositions was determined as the percentage of propositions that could not yet be assigned a definitive logical status (Figure 1E).
6. The cumulative number of stable propositions was recorded (Figure 1F).
7. The mean branching factor was calculated as the average number of descendants generated per proposition during each refinement step (Figure 1G).
8. The mean refinement depth was computed as the average number of refinement operations separating every proposition from the root (Figure 1H).

Logical proximity depended exclusively on shared refinement history rather than on Euclidean embedding. Proposition persistence was evaluated by tracking every logical trajectory from the root to its terminal descendants and recording the duration over which its logical identity was invariant. Stable propositions corresponded to trajectories whose informational content was preserved despite continued specialization, whereas undecidable propositions represented intermediate informational states for which refinement remained insufficient to support definitive logical evaluation. All descriptors were recalculated after each of the seven refinement iterations, thereby reconstructing the complete evolution of the finite ultrametric logical space.

Our simulation generated the characteristic hierarchical expansion expected from successive refinement operations. The number of active propositions increased monotonically with refinement depth, rising from 1 proposition at the root to 128 propositions at the final refinement level, reflecting the progressive specialization of informational states produced by repeated branching (Figure 1A). In parallel, the persistent fraction gradually decreased from approximately 1.00 at the initial level to 0.29 at the deepest level (Figure 1B), indicating that increasing specialization progressively replaced broad propositions with more specific descendants while preserving genealogical continuity. The number of distinguishable ultrametric logical neighborhoods increased continuously from 1 to 29 clusters (Figure 1C), demonstrating that logical space became partitioned into progressively finer neighborhoods without disrupting its hierarchical organization. Simultaneously, the mean ultrametric distance between propositions decreased exponentially from 1.0 to approximately 1.6×10^{-2} according to the increasing depth of their nearest common ancestors (Figure 1D), showing that local logical neighborhoods became increasingly cohesive despite the overall expansion of the proposition set.

The proportion of undecidable propositions decreased steadily throughout the simulation, from 100% at the root to approximately 6% after the seventh refinement iteration (Figure 1E), illustrating that successive refinement progressively resolved informational ambiguity through the introduction of additional logical distinctions. Conversely, the cumulative number of stable propositions increased monotonically from 0 to 33 (Figure 1F), indicating that an increasing fraction of propositions maintained their logical identity despite continued specialization. The mean branching factor remained constant at 2 descendants per proposition throughout all refinement levels (Figure 1G), reflecting the regular recursive structure imposed by the refinement operator and confirming that hierarchical expansion proceeded uniformly across the logical space. Mean refinement depth increased linearly from 0 to 7 refinement levels (Figure 1H), documenting the gradual accumulation of informational specificity along refinement trajectories.

Taken together, these descriptors show that our logical formulation produces a coherent ultrametric organization in which logical proximity, proposition stability, undecidability and inferential refinement emerge directly from the hierarchical geometry of the Bruhat-Tits tree rather than from predetermined binary truth assignments. Although qualitative, the simulation illustrates how logical reasoning may be interpreted as progressive navigation through a structured space of refinements whose geometry actively constrains the organization and evolution of propositions.

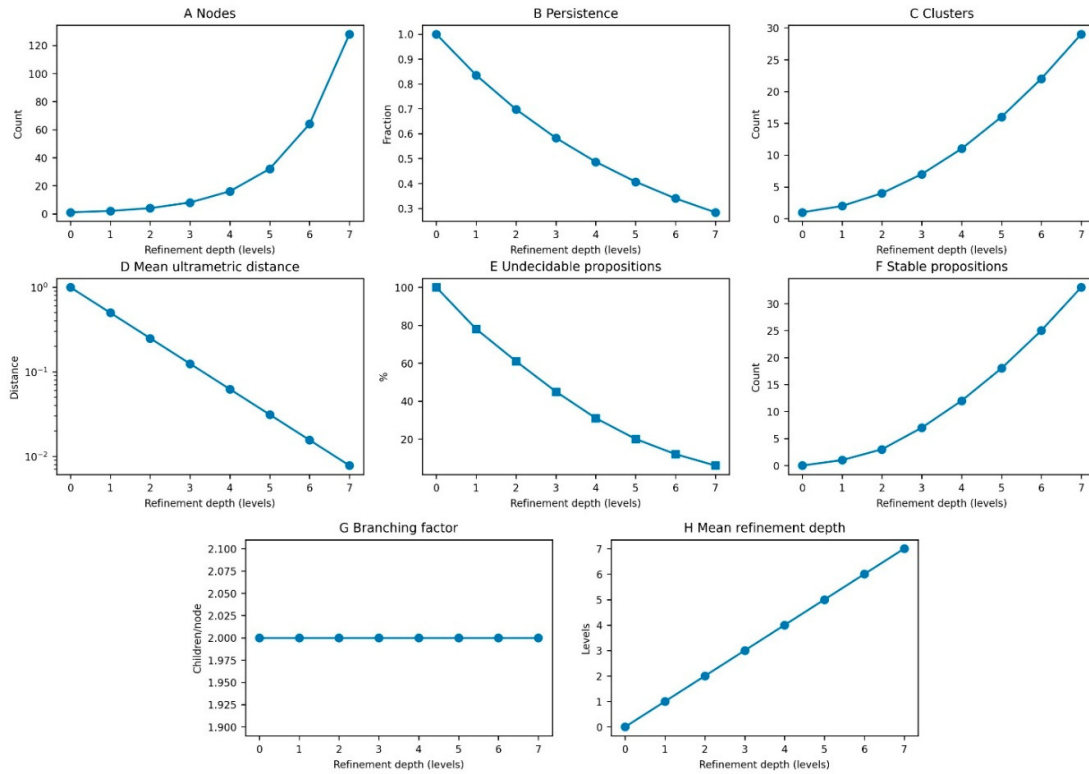


Figure 1. Evolution of an ultrametric logical framework during successive refinement steps performed on a finite approximation of a Bruhat-Tits tree. The horizontal axis of all panels represents refinement depth (levels).

Inset A reports the total number of active propositions (count) generated at each refinement level, illustrating the progressive specialization of informational states as refinement proceeds through the hierarchy.

Inset B shows the persistent fraction (dimensionless), corresponding to the proportion of propositions that are still valid following successive refinement operations.

Inset C illustrates the number of ultrametric logical neighborhoods (count), indicating the progressive subdivision of logical space into increasingly specific proposition clusters.

Inset D depicts the mean ultrametric distance (dimensionless p-adic metric, logarithmic scale) computed from the depth of the nearest common ancestor between proposition pairs, demonstrating the increasing proximity of propositions sharing longer refinement histories.

Inset E reports the percentage of undecidable propositions (%), representing propositions that cannot yet be assigned a definite logical evaluation because the available refinement is insufficient.

Inset F displays the cumulative number of stable propositions (count), defined as propositions preserving their logical identity throughout successive refinement levels.

Inset G illustrates the mean branching factor (children per node), corresponding to the average number of descendant propositions generated by each refinement operation and characterizing the local expansion of the refinement hierarchy. **Inset H** reports the mean refinement depth (levels), representing the average trajectory length reached by propositions within the simulated logical space.

Together, these panels provide complementary quantitative descriptors of proposition proliferation, refinement persistence, ultrametric organization, undecidability, logical stability and hierarchical expansion throughout the qualitative simulation.

Conclusions

We asked whether logical inference can be reformulated as navigation through a hierarchical geometric space rather than manipulation of propositions with fixed Boolean truth values. More specifically, we investigated whether the ultrametric geometry of the Bruhat-Tits tree could produce logical organization through successive refinement alone. Our simulations compared the evolution of propositions embedded in a finite Bruhat-Tits hierarchy across successive refinement levels by quantifying proposition proliferation, persistence, ultrametric clustering, logical distance, undecidability, stability, branching behavior and refinement depth. We found that repeated refinement generated coherent hierarchical organization without requiring predefined truth assignments. Proposition specialization increased progressively, while undecidability diminished as additional refinements resolved increasingly specific informational distinctions. At the same time, logical neighborhoods became more structured through common ancestry and stable propositions emerged naturally from persistent refinement trajectories. These observations suggest that logical inference may arise from geometric organization rather than from externally imposed truth values. Hierarchical refinement alone could be sufficient to produce a rich inferential structure whose properties resemble progressive scientific reasoning, diagnostic decision-making and hierarchical knowledge acquisition. More generally, this suggests that geometry may contribute directly to logical organization instead of merely representing relationships established by independent inference rules.

Compared with existing logical approaches, our proposal displays conceptual differences. Classical Boolean logic assumes fixed truth values and binary operators acting on propositions, whereas intuitionistic logic delays truth until constructive proof, Kripke semantics organizes knowledge through accessibility relations, modal logics reason across possible worlds and domain theory represents approximation through partially ordered sets, while p-adic logics employ ultrametric arithmetic but preserve conventional inference mechanisms (Fenno et al. 2014; Zhao et al. 2021; Hemedan et al. 2022; Bressler et al. 2023). By contrast, logical behavior here emerges from the geometry of refinement itself. Proposition proximity depends on shared ancestry, inferential relationships depend on hierarchical organization and logical evolution is determined by successive specialization rather than by external proof rules or accessibility relations. Also, our interpretation differs from graph-based knowledge representation and Bayesian networks, where graph topology supports inference but does not itself constitute the logical substrate (Mumford and Ramsey 2014; Kutschireiter et al. 2023; Hammond and Smith 2025; Hong and Kuruoglu 2025). Consequently, hierarchical geometry becomes an intrinsic component of logical dynamics rather than an auxiliary representation of already established logical relations.

Our study has limitations. Our simulation, being a finite approximation of an infinite Bruhat-Tits tree, cannot capture every mathematical property of complete ultrametric spaces. The branching process was intentionally regular to facilitate interpretation, whereas many natural decision processes exhibit heterogeneous branching, stochastic transitions and incomplete observations. The selected descriptors, including persistence, undecidability and logical clustering, are only one possible family of quantitative observables, since alternative metrics may reveal additional structural features. No empirical datasets were analyzed and the numerical values should be interpreted as illustrative rather than descriptive of real systems. Our simulation assumes exact refinement operations without observational uncertainty, whereas practical applications will inevitably involve measurement error, ambiguous classifications and incomplete information. Additional questions are open, including how refinement geometry behaves under noisy observations, whether alternative ultrametric spaces generate comparable inferential behavior, how probabilistic reasoning can be incorporated into hierarchical refinement and whether multiple interacting Bruhat-Tits trees can represent distributed reasoning systems.

Our approach suggests experimentally testable hypotheses. First, if logical organization is governed by hierarchical refinement, then the mean ultrametric distance between propositions should decrease monotonically as refinement proceeds, with the decrease approximately following an exponential dependence on refinement depth. This prediction can be evaluated in automated

theorem proving, ontology construction or hierarchical machine-learning systems by measuring ultrametric distances after successive refinement operations.

Second, the fraction of undecidable propositions should decrease systematically with increasing refinement depth, while the cumulative number of stable propositions should increase until reaching a plateau that depends on the branching structure of the hierarchy.

Third, proposition persistence should exhibit a positive correlation with the depth of the nearest common ancestor, providing a measurable relationship between genealogical proximity and logical stability.

Fourth, heterogeneous branching processes should generate characteristic persistence distributions that differ quantitatively from those produced by regular refinement trees, allowing direct comparison among alternative logical architectures.

Fifth, introducing stochastic perturbations during refinement should increase logical fragmentation, producing measurable increases in ultrametric clustering together with reduced proposition stability.

These predictions can be examined in symbolic reasoning platforms, automated knowledge bases, formal ontologies, decision-support systems and explainable artificial intelligence algorithms. Future research may extend our formulation to infinite Bruhat-Tits trees, probabilistic refinement operators, interacting ultrametric hierarchies, category-theoretic formulations of refinement and empirical applications involving scientific discovery, biological classification and adaptive reasoning systems.

Potential practical applications extend beyond formal logic. Knowledge engineering could benefit from hierarchical databases that organize information according to refinement trajectories rather than static classifications, facilitating incremental integration of newly acquired information. Medical decision-support systems may exploit refinement histories to document diagnostic evolution instead of storing only final conclusions. Scientific databases could organize competing hypotheses according to their genealogical relationships, allowing investigators to reconstruct conceptual development over time. Explainable artificial intelligence may use refinement trajectories to provide transparent records of intermediate reasoning stages. Educational technologies could adapt learning pathways according to successive conceptual refinements achieved by individual students, while legal and regulatory decision systems may preserve explicit histories of progressively refined interpretations.

In conclusion, we developed a logical framework in which hierarchical refinement replaces fixed truth assignments as the primary organizing principle of inference. Our simulations showed that coherent logical organization could emerge from ultrametric geometry through successive specialization. Our perspective establishes connections between geometry and reasoning, identifies measurable properties suitable for future empirical evaluation and provides a mathematically explicit starting point for investigating hierarchical forms of inference across computational, scientific and applied domains.

Ethics approval and consent to participate: Ethics approval statement is not applicable. This research does not contain any studies with human participants or animals performed by the Author.

Consent for publication: The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Availability of data and materials: All data and materials generated or analyzed during this study are included in the manuscript. The Author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Disclaimer. The views expressed are those of the author and do not necessarily reflect those of the affiliated institutions.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Acknowledgments: none.

Authors' contributions: The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work, the author used ChatGPT 5.3 to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

References

- Bonzio, S., V. Fano, P. Graziani, and M. Pra Baldi. 2023. "A Logical Modeling of Severe Ignorance." *Journal of Philosophical Logic* 52: 1–28. <https://doi.org/10.1007/s10992-022-09697-x>.
- Bressler, E. M., et al. 2023. "Boolean Logic in Synthetic Biology and Biomaterials: Towards Living Materials in Mammalian Cell Therapeutics." *Clinical and Translational Medicine* 13 (7): e1244. <https://doi.org/10.1002/ctm2.1244>.
- Brunet, T. D. P. 2023. "Categorical Modal Realism." *Synthese* 201 (2): 49. <https://doi.org/10.1007/s11229-023-04042-4>.
- Chen, Lin, Xirong Liu, and Ling-Yan Hung. 2021. "Emergent Einstein Equation in p-Adic Conformal Field Theory Tensor Networks." *Physical Review Letters* 127 (22): 221602. <https://doi.org/10.1103/PhysRevLett.127.221602>.
- Ciardelli, I. 2025. "Inquisitive Neighborhood Logic." *Journal of Logic, Language and Information* 34 (5): 419–61. <https://doi.org/10.1007/s10849-025-09440-0>.
- Da Silva Filho, J. I., et al. 2016. "Paraconsistent Analysis Network Applied in the Treatment of Raman Spectroscopy Data to Support Medical Diagnosis of Skin Cancer." *Medical & Biological Engineering & Computing* 54 (10): 1453–67. <https://doi.org/10.1007/s11517-016-1471-3>.
- Da Silva Filho, J. I., et al. 2021. "Paraconsistent Annotated Logic Algorithms Applied in Management and Control of Communication Network Routes." *Sensors* 21 (12): 4219. <https://doi.org/10.3390/s21124219>.
- De Rizzo, J. 2025. "Deflationism, Explanation and 'Because'." *Philosophical Studies* 182 (8): 2215–42. <https://doi.org/10.1007/s11098-025-02351-7>.
- Decker, Jason. 2025. *Logic for Everyone: From Proof to Paradox*. Cambridge: Cambridge University Press.
- Dedhe, A. M., H. Clatterbuck, S. T. Piantadosi, and J. F. Cantlon. 2023. "Origins of Hierarchical Logical Reasoning." *Cognitive Science* 47: e13250. <https://doi.org/10.1111/cogs.13250>.
- Fenno, L. E., et al. 2014. "Targeting Cells with Single Vectors Using Multiple-Feature Boolean Logic." *Nature Methods* 11 (7): 763–72. <https://doi.org/10.1038/nmeth.2996>.
- Fernández-Vilas, E., J. R. Coca, N. Plaza-Gómez, and F. S. Toscano. 2026. "The Social Phenomenon of Rare Diseases: A Logical-Modal Analysis." *Frontiers in Sociology* 11: 1810602. <https://doi.org/10.3389/fsoc.2026.1810602>.
- Gao, L., et al. 2025. "Proximity Ligation Strategy-Mediated Cross-Modal Logic-Gate Sensing Platform: Towards Dual Biomarkers Detection for Complex Disease." *Biosensors and Bioelectronics* 288: 117798. <https://doi.org/10.1016/j.bios.2025.117798>.
- Hamdy, M., A. Ibrahim, B. Abozalam, and S. Helmy. 2024. "Design and Implementation of Type-3 Intuitionistic Fuzzy Logic MPPT Controller for PV Solar System: Comparative Study." *ISA Transactions* 154: 488–511. <https://doi.org/10.1016/j.isatra.2024.09.006>.

- Hammond, J., and V. A. Smith. 2025. "Bayesian Networks for Network Inference in Biology." *Journal of the Royal Society Interface* 22 (226): 20240893. <https://doi.org/10.1098/rsif.2024.0893>.
- Hemedan, A. A., A. Niarakis, R. Schneider, and M. Ostaszewski. 2022. "Boolean Modelling as a Logic-Based Dynamic Approach in Systems Medicine." *Computational and Structural Biotechnology Journal* 20: 3161–72. <https://doi.org/10.1016/j.csbj.2022.06.035>.
- Hong, J., and E. E. Kuruoglu. 2025. "Minimax Bayesian Neural Networks." *Entropy* 27 (4): 340. <https://doi.org/10.3390/e27040340>.
- King, Colin Guthrie, and Venanzio Raspa, eds. 2025. *Aristotle's Organon in Old and New Logic: 1800–1950*. Bloomsbury Studies in the Aristotelian Tradition. London: Bloomsbury Academic.
- Kutschireiter, A., M. A. Basnak, R. I. Wilson, and J. Drugowitsch. 2023. "Bayesian Inference in Ring Attractor Networks." *Proceedings of the National Academy of Sciences of the United States of America* 120 (9): e2210622120. <https://doi.org/10.1073/pnas.2210622120>.
- Litland, J. E. 2025. "Truthmaker Semantics for Intuitionistic Modal Logic." *Topoi* 44 (2): 325–43. <https://doi.org/10.1007/s11245-024-10094-z>.
- Ludwig, Judith, and Christian Merten. 2026. "Formalising the Bruhat-Tits Tree." *Annals of Formalized Mathematics* 2: afm15738. Also available as arXiv preprint arXiv:2505.12933. <https://doi.org/10.46298/afm.15738>. <https://doi.org/10.48550/arXiv.2505.12933>.
- Montambault, B., et al. 2025. "DimBridge: Interactive Explanation of Visual Patterns in Dimensionality Reductions with Predicate Logic." *IEEE Transactions on Visualization and Computer Graphics* 31 (1): 207–17. <https://doi.org/10.1109/TVCG.2024.3456391>.
- Mumford, J. A., and J. D. Ramsey. 2014. "Bayesian Networks for fMRI: A Primer." *NeuroImage* 86: 573–82. <https://doi.org/10.1016/j.neuroimage.2013.10.020>.
- Olkhovikov, G. K. 2024. "A Basic System of Paraconsistent Nelsonian Logic of Conditionals." *Journal of Logic, Language and Information* 33 (4–5): 299–337. <https://doi.org/10.1007/s10849-024-09421-9>.
- Oshiyama, N. F., R. A. Bassani, I. M. L. D'Ottaviano, and J. W. Bassani. 2012. "Medical Equipment Classification: Method and Decision-Making Support Based on Paraconsistent Annotated Logic." *Medical & Biological Engineering & Computing* 50 (4): 395–402. <https://doi.org/10.1007/s11517-012-0888-6>.
- Parry, William T., and Edward A. Hacker. 1991. *Aristotelian Logic*. Albany: State University of New York Press.
- Rothhaar, M. 2020. "On Justifying Arguments of Species Membership." *Bioethics* 34 (2): 159–65. <https://doi.org/10.1111/bioe.12657>.
- Sgurev, V. 2021. "Complex Multivalued Hierarchical Logic with One Real and Two Imaginary Logical Structures." In *2021 International Conference Automatics and Informatics (ICAI)*, 1–5. Varna, Bulgaria. <https://doi.org/10.1109/ICAI52893.2021.9639499>.
- Sgurev, V. 2022. "Complex Multivalued Hierarchical Logic (HS-Logic)." In *Advances in Intelligent Systems Research and Innovation*, edited by V. Sgurev, V. Jotsov, and J. Kacprzyk, *Studies in Systems, Decision and Control*, vol. 379. Cham: Springer. https://doi.org/10.1007/978-3-030-78124-8_7.
- Tarannum, S., and S. Jabin. 2022. "Prioritizing Severity Level of COVID-19 Using Correlation Coefficient and Intuitionistic Fuzzy Logic." *International Journal of Information Technology* 14 (5): 2469–75. <https://doi.org/10.1007/s41870-022-00971-4>.
- von Plato, J. 2025. "Gödel's Modal Interpretation of Intuitionistic Logic and Its Proof Theory." *Monatshefte für Mathematik* 208 (4): 791–817. <https://doi.org/10.1007/s00605-025-02083-0>.
- Wang, Guan, Jin Li, Yuhao Sun, Xing Chen, Changling Liu, Yue Wu, Meng Lu, Sen Song, and Yasin Abbasi Yadkori. 2025. "Hierarchical Reasoning Model." *arXiv preprint arXiv:2506.21734v3*. <https://doi.org/10.48550/arXiv.2506.21734>.

Zabrodin, A. V. 1989. "Non-Archimedean Strings and Bruhat-Tits Trees." *Communications in Mathematical Physics* 123 (3): 463–83. <https://doi.org/10.1007/BF01238811>.

Zhao, S., et al. 2021. "Boolean Logic Gate Based on DNA Strand Displacement for Biosensing: Current and Emerging Strategies." *Nanoscale Horizons* 6 (4): 298–310. <https://doi.org/10.1039/D0NH00587H>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Not peer-reviewed version

A Spectral and Geometric Perspective on Immune Repertoire Selection

[Arturo Tozzi](#) *

Posted Date: 13 August 2026

doi: 10.20944/preprints202608.0909.v1

Keywords: clonality; lymphocytes; immunodynamics; bioinformatics; systems biology



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Spectral and Geometric Perspective on Immune Repertoire Selection

Arturo Tozzi

ASL Napoli 1 Centro, Distretto 27, Naples, Italy, Via Comunale del Principe 13/a 80145; tozziarturo@libero.it

Abstract

The immune system delivers a highly diverse receptor repertoire while preserving self-tolerance, functional robustness and effective pathogen recognition. Although the mechanisms governing clonal selection are characterized, the mathematical principles potentially underlying repertoire-wide organization are less understood. We propose a geometric, probabilistic and spectral perspective in which immune repertoire selection could be interpreted as constrained sampling from a high-dimensional space of receptor configurations. Within this perspective, we connect developments in modern probability and combinatorics with distinct, testable aspects of repertoire organization: polynomial stability could define a global landscape of admissible repertoires; negative dependence could constrain simultaneous selection of similar or redundant clonotypes; entropy maximization could preserve the broadest receptor distribution compatible with self-tolerance and antigen recognition; log-concavity could organize compatible configurations into broad basins supporting gradual adaptation. Still, Kadison–Singer subset preservation could explain how comparatively small lymphocyte samples retain statistical and spectral properties of larger repertoires, whereas random-forest connectivity could quantify collective integration generated by local receptor relationships. Finally, spectral independence could constrain propagation of clonal perturbations, allowing local responses without uncontrolled repertoire-wide cascades. These mathematical issues could provide measurable descriptors of diversity, redundancy, representativeness, connectivity and perturbation propagation, generating falsifiable predictions accessible to repertoire sequencing and network analysis. Rather than replacing established immunological mechanisms, we introduce a complementary ensemble-level description shifting the mathematical object of interest from the individual clonotype to the organization of the repertoire as a whole.

Keywords: clonality; lymphocytes; immunodynamics; bioinformatics; systems biology

Introduction

The immune system continuously builds a receptor repertoire that is diverse, self-tolerant, robust and capable of recognizing an enormous variety of pathogens. Although receptor recombination, clonal expansion, affinity maturation and immune tolerance have been extensively investigated, the general mathematical principles through which innumerable local selection events could produce globally organized repertoires are less understood (Hochhaus 2008; Kumar et al. 2019; Adams, Grassmann and Sun 2020; Deimel et al. 2025; Kocher et al. 2025; Ibáñez-Molero et al. 2026). Current descriptions address receptor affinity, stochastic recombination, evolutionary dynamics or network interactions but lack a unified language connecting local probabilistic selection with repertoire-wide organization (Mesin et al. 2020; Daëron 2022; Domínguez-Andrés et al. 2023; Ke et al. 2023; Sjöström et al. 2023; Chen, Liu and Cao 2025).

Recent developments in combinatorics, probability theory and spectral graph methods could provide a potentially relevant perspective: large and complex systems like the biological ones could acquire global organization when local choices are constrained by probabilistic and geometric relationships. Several mathematical advances converge toward this perspective. Strongly Rayleigh

probability measures established a relationship between stable multivariate polynomials and negative dependence, providing a mathematical basis for probabilistic selection that limits excessive correlations among selected elements (Borcea, Brändén and Liggett 2009). Entropy-guided randomized algorithms showed how constrained stochastic sampling can explore rich ensembles of compatible configurations instead of converging toward a single deterministic optimum (Asadpour et al. 2010; Oveis Gharan, Saberi and Singh 2011). Extensions of the Kadison–Singer theorem demonstrated that subsets can preserve important spectral properties of much larger systems despite dependencies among their elements (Anari and Oveis Gharan 2015). Related developments connected random forests with global network connectivity (Goel et al. 2015), established log-concavity as an important property of combinatorial probability distributions (Anari, Oveis Gharan and Vinzant 2018) and introduced spectral independence as a criterion constraining long-range influence within high-dimensional probabilistic systems (Anari, Liu and Oveis Gharan 2020).

Accordingly, we aim to explore whether these independent mathematical developments can be integrated into a coherent theoretical perspective on immune repertoire selection. We will argue that polynomial stability could characterize the global structure of admissible repertoire space; negative dependence could regulate selection among similar or redundant receptors; entropy maximization could favor broad but constrained repertoire exploration; log-concavity could define robust regions of compatible configurations; Kadison–Singer subset preservation could describe how smaller lymphocyte populations retain properties of larger repertoires; random-forest connectivity could characterize collective integration among clonotypes; and spectral independence could constrain propagation of local perturbations across the repertoire.

We will proceed as follows. First, we introduce a unified probabilistic formulation in which immune repertoire organization is described through complementary geometric, probabilistic and spectral constraints. Then, we translate this formulation into measurable repertoire properties, including diversity, redundancy, representative sampling, connectivity and robustness. Finally, we derive experimentally testable predictions and discuss the broader implications of this mathematical approach to **immune repertoire selection**.

The Mathematical Perspective on Immune Repertoire Selection

A useful distinction can be made between the mechanisms generating an immune repertoire and the properties characterizing the repertoire once it has been generated. Receptor recombination, selection, clonal expansion and affinity maturation operate on individual cells and lineages, whereas diversity, redundancy, representativeness, connectivity and resilience are properties of the receptor population as a whole. These two levels of description need not obey the same mathematical rules. Knowing how each clonotype is generated or selected does not, by itself, specify the statistical architecture of the resulting ensemble.

Therefore, we shift the mathematical object of interest from the individual clonotype to the repertoire configuration. Let the potential receptor space contain an enormous number of possible elements, only a small and continuously changing fraction of which is realized at any given time. Then, immune selection can be examined as a problem of how subsets occupy this space: whether selected receptors cluster or repel one another, whether many alternative configurations are comparably admissible, whether a small subset preserves information about the larger population, how strongly different regions are connected and how far the effect of a local perturbation propagates. These distinct questions cannot be reduced to receptor abundance or affinity alone. This change of scale motivates the mathematical correspondences developed below. Rather than assigning each mathematical concept to another molecular mechanism, we ask whether the above-mentioned mathematical concepts could characterize different ensemble-level properties of the same immune repertoire.

Each concept would address a different aspect of its organization, from the geometry of admissible configurations to competition among receptors, representative sampling, collective connectivity and resistance to perturbation.

- a. **Polynomial stability** could provide the most general description of this organizational landscape. Stable multivariate polynomials generate probability distributions whose geometry imposes strong constraints on admissible configurations even in very high-dimensional spaces (Borcea, Brändén and Liggett 2009). In an immunological interpretation, polynomial stability could correspond to a globally coherent repertoire landscape in which biologically compatible receptor combinations occupy structurally well-behaved regions. Repertoire organization would then depend not only on the fitness of individual receptors but also on the global geometry of the space from which they are selected. This hypothesis would shift attention from isolated clonotypes toward relationships among possible repertoire configurations.
- b. **Negative dependence** could describe how this global structure constrains competition among receptors. In Strongly Rayleigh distributions, selection of one element decreases the probability of simultaneously selecting correlated alternatives (Borcea, Brändén and Liggett 2009). If an analogous principle operated during immune selection, expansion or persistence of one clonotype could decrease the probability of retaining highly similar or functionally redundant alternatives. Such dependence could limit excessive overlap while permitting multiple compatible receptor combinations. Repertoire diversity would consequently arise partly from probabilistic competition among related clonotypes rather than exclusively from the independent survival or elimination of individual cells. Importantly, this correspondence would concern the statistical architecture of selection and would not imply a specific molecular mechanism through which one clonotype directly inhibits another.
- c. **Entropy maximization** could account for a complementary property: preservation of extensive receptor diversity within biological constraints. Entropy-based randomized sampling has demonstrated that optimization need not identify a unique deterministic solution and can instead favor rich distributions of admissible configurations (Asadpour et al. 2010; Oveis Gharan, Saberi and Singh 2011). Applied to adaptive immunity, such a principle would predict that repertoire selection could favor the broadest distribution of receptor states compatible with constraints such as self-tolerance, finite population size and adequate antigen recognition. Selection would not necessarily drive the repertoire toward maximal affinity or another single optimum. It could instead preserve many alternative receptor configurations, potentially reconciling broad pathogen coverage with restrictions imposed by immune tolerance.
- d. **Log-concavity** could describe the geometry of those compatible configurations. Log-concave probability distributions concentrate probability within connected, broadly organized regions rather than distributing it among numerous sharply separated optima (Anari, Oveis Gharan and Vinzant 2018). If repertoire probability possessed an analogous structure, neighboring immune configurations could have comparable functional properties. Moderate changes caused by development, ageing, antigen exposure or stochastic clonal turnover could then move the repertoire within a broad compatibility basin rather than forcing transitions between isolated optimal states. Repertoire robustness would, in this interpretation, depend partly on the geometry of the probability landscape: multiple nearby receptor configurations could support comparable global immune organization.
- e. The extension of the **Kadison–Singer** results to Strongly Rayleigh measures introduces a different hypothesis concerning representative subsets. Mathematical results in this area show that subsets can preserve important spectral characteristics of considerably larger systems despite dependencies among their elements (Anari and Oveis Gharan 2015). An immunological analogue could arise if comparatively small populations of circulating, sampled or activated lymphocytes retained major statistical or spectral properties of the larger repertoire from which they originated. Immune surveillance and repertoire sampling would then depend not simply on the absolute number of observed clonotypes but also on whether the selected subset preserves important structural characteristics of the parent population. This possibility would be experimentally distinguishable by comparing spectral and covariance properties across progressively smaller samples of the same repertoire.

- f. **Random-forest connectivity** could provide a complementary description of collective integration. Random forests allow global connectivity to be characterized through probabilistic relationships among locally connected elements (Goel et al. 2015). Applied to immune repertoires, such connectivity could quantify the extent to which clonotypes participate in common functional communities defined by receptor similarity, shared antigen recognition or experimentally established interactions. A repertoire could consequently exhibit collective cohesion without requiring every clonotype to interact directly with every other clonotype. The relevant biological hypothesis would be that local relationships among receptor populations could generate measurable global connectivity and that this connectivity could constitute an organizational property distinct from repertoire diversity alone.
- g. Finally, **spectral independence** could characterize the propagation of perturbations through this interaction structure. Spectral independence constrains how strongly local variables can influence the global state of a high-dimensional probabilistic system (Anari, Liu and Oveis Gharan 2020). If an analogous property characterized immune repertoires, activation, deletion or expansion of individual clonotypes would produce bounded effects on the rest of the repertoire rather than unrestricted cascades. Physiological immune responses could involve local changes while preserving global organization. Conversely, some forms of immune dysregulation could be associated with weakened spectral constraints, allowing local disturbances to acquire disproportionately large repertoire-wide effects. This interpretation would generate a measurable distinction between local clonal responsiveness and global susceptibility to perturbation without equating spectral independence with any particular regulatory pathway.

Overall, the combination of these features would suggest that immune repertoire selection could be viewed as constrained probabilistic sampling in which diversity, compatibility, representativeness, connectivity and robustness arise from complementary properties of the selected ensemble. None of these correspondences is intended as a direct mechanistic explanation of lymphocyte biology, nor do the mathematical results themselves establish that immune repertoires displays these properties. Rather, they define a set of hypotheses through which established mathematical structures could be translated into experimentally measurable features of repertoire organization.

A Unified Mathematical Formulation of Immune Repertoire Selection

Given the ensemble-level hypotheses introduced above, we now express them within a single probabilistic construction. Let

$$\mathcal{R} = \{1, \dots, N\}$$

denote a finite candidate receptor space and let

$$\Omega \subseteq \mathcal{R}, \quad |\Omega| = m,$$

denote one realized repertoire. The state variable is not an individual clonotype but the subset Ω . A probability measure

$$\mu(\Omega) = P(\Omega)$$

is assigned over admissible subsets, with

$$\sum_{\substack{\Omega \subseteq \mathcal{R} \\ |\Omega|=m}} \mu(\Omega) = 1.$$

This construction permits receptor-level information to enter the model while the quantities of interest are evaluated at repertoire level.

To encode dependencies among selections, introduce binary variables

$$X_i(\Omega) = \begin{cases} 1, & i \in \Omega, \\ 0, & i \notin \Omega. \end{cases}$$

The generating polynomial of the repertoire distribution is

$$G_{\mu}(z_1, \dots, z_N) = \sum_{\Omega \subseteq \mathcal{R}} \mu(\Omega) \prod_{i \in \Omega} z_i.$$

A Strongly Rayleigh version of the proposal would require G_{μ} to be real stable (Borcea, Brändén and Liggett 2009). This condition supplies the formal setting in which negative dependence can be stated. For distinct receptors i and j , a basic pairwise consequence is

$$\text{Cov}_{\mu}(X_i, X_j) = P(i, j \in \Omega) - P(i \in \Omega)P(j \in \Omega) \leq 0.$$

More generally, our model can be tested for stronger negative-association properties rather than assuming independence among clonotypes.

Repertoire breadth is described by the entropy of the distribution over admissible subsets,

$$\mathcal{H}(\mu) = - \sum_{\Omega} \mu(\Omega) \log \mu(\Omega).$$

This differs from the Shannon diversity of clone abundances within one observed repertoire: $\mathcal{H}(\mu)$ measures how broadly probability is distributed across alternative repertoire configurations. Constrained maximum-entropy selection can be formulated as

$$\mu^* = \underset{\mu}{\operatorname{argmax}} \mathcal{H}(\mu)$$

subject to biologically specified expectation constraints

$$\mathbb{E}_{\mu}[f_k(\Omega)] = c_k, \quad k = 1, \dots, q,$$

and normalization of μ . Such a construction follows the general logic of entropy-guided randomized selection, in which stochasticity is retained while admissible configurations are restricted by global constraints (Asadpour et al. 2010; Oveis Gharan, Saberi and Singh 2011). The corresponding exponential-family form is

$$\mu_{\theta}(\Omega) = \frac{1}{Z(\theta)} \exp \left[\sum_{k=1}^q \theta_k f_k(\Omega) \right],$$

where

$$Z(\theta) = \sum_{\Omega} \exp \left[\sum_{k=1}^q \theta_k f_k(\Omega) \right].$$

Geometric concentration can be introduced by representing a repertoire through a feature vector

$$\mathbf{x}(\Omega) \in \mathbb{R}^d$$

and assigning it a positive density $g(\mathbf{x})$. The proposed log-concavity condition is

$$g(\tau \mathbf{x} + (1 - \tau) \mathbf{y}) \geq g(\mathbf{x})^{\tau} g(\mathbf{y})^{1-\tau}, \quad 0 \leq \tau \leq 1,$$

or equivalently

$$\log g(\tau \mathbf{x} + (1 - \tau) \mathbf{y}) \geq \tau \log g(\mathbf{x}) + (1 - \tau) \log g(\mathbf{y}).$$

This provides a direct mathematical criterion for testing whether high-probability repertoire states form geometrically coherent regions, consistent with log-concavity results for combinatorial probability distributions (Anari, Oveis Gharan and Vintzant 2018).

Representative subset preservation can be formulated independently of the preceding terms. Associate each receptor i with a feature vector

$$\mathbf{v}_i \in \mathbb{R}^d.$$

For the complete candidate set define

$$\mathbf{C}_{\mathcal{R}} = \sum_{i \in \mathcal{R}} w_i \mathbf{v}_i \mathbf{v}_i^{\top},$$

and for a selected repertoire

$$\mathbf{C}_{\Omega} = \sum_{i \in \Omega} \tilde{w}_i \mathbf{v}_i \mathbf{v}_i^{\top},$$

where the weights are normalized within their respective sets. A spectral preservation error is

$$E_{\text{spec}}(\Omega) = \|\mathbf{C}_\Omega - \mathbf{C}_\mathcal{R}\|_2.$$

The subset-preservation hypothesis predicts the existence of comparatively small Ω for which E_{spec} stays bounded. This formulation captures the relevant spectral question without asserting that an immune repertoire literally implements the Kadison–Singer construction (Anari and Oveis Gharan 2015).

Connectivity is introduced by defining a weighted receptor graph

$$\mathcal{G}_\Omega = (\Omega, \mathbf{W}),$$

where $W_{ij} \geq 0$ encodes an experimentally chosen relationship between receptors i and j . Let

$$p_{ij}^{\text{RF}}$$

denote the probability that i and j belong to the same tree component under a specified random-forest measure. Repertoire connectivity is then summarized by

$$C_{\text{RF}}(\Omega) = \frac{2}{m(m-1)} \sum_{\substack{i,j \in \Omega \\ i < j}} p_{ij}^{\text{RF}}.$$

This quantity converts local graph relations into an ensemble-level measure of integration (Goel et al. 2015).

To formalize bounded propagation of perturbations, let

$$\Psi_\Omega = (\psi_{ij})$$

be an influence matrix, where ψ_{ij} quantifies the change in the conditional probability of receptor i associated with a change in receptor j . A spectral-independence condition can be expressed through

$$\lambda_{\max}(\Psi_\Omega) \leq \kappa,$$

for a finite bound κ across the relevant conditional repertoire distributions (Anari, Liu and Oveis Gharan 2020). The quantity

$$\Lambda(\Omega) = \lambda_{\max}(\Psi_\Omega)$$

becomes the principal spectral observable for perturbation propagation.

These components can be collected without assuming that they are interchangeable. Let

$$\Phi(\Omega) = (\mathcal{H}, D, L, -E_{\text{spec}}, C_{\text{RF}}, -\Lambda, U)$$

denote a vector of repertoire-level descriptors, where D is an empirically defined diversity or nonredundancy statistic and U represents biological constraints such as antigen recognition and self-reactivity. A scalarized representation can be written as

$$F_\alpha(\Omega) = \alpha_H \mathcal{H} + \alpha_D D(\Omega) + \alpha_L L(\Omega) - \alpha_E E_{\text{spec}}(\Omega) + \alpha_C C_{\text{RF}}(\Omega) - \alpha_\Lambda \Lambda(\Omega) + \alpha_U U(\Omega),$$

with

$$\alpha_k \geq 0.$$

The coefficients are deliberately left unspecified. They are not proposed as universal biological constants and would have to be estimated, constrained or compared using empirical data. Alternatively, no scalarization need be imposed: repertoires can be compared directly in the multidimensional descriptor space Φ , allowing trade-offs among organizational properties to be measured rather than hidden within predetermined weights.

The resulting probabilistic model can be expressed in Gibbs form,

$$P_{\alpha,T}(\Omega) = \frac{\exp [F_\alpha(\Omega)/T]}{\sum_{\Omega'} \exp [F_\alpha(\Omega')/T]},$$

where $T > 0$ controls concentration over repertoire configurations. This equation is not intended as a quantitative law of immune physiology. It provides a compact mathematical representation in which the hypotheses introduced in the preceding section become separable observables and constraints. Our proposal is falsifiable at several levels: the repertoire distribution can be tested for negative dependence and log-concavity; subsamples can be evaluated for spectral preservation; receptor graphs can be examined for random-forest connectivity; influence matrices can be tested for

bounded spectral radius; and the joint descriptor vector can be compared across physiological states and against null models based on independent or affinity-only selection.

Experimentally Testable Predictions

Our geometric and spectral interpretation of immune repertoire selection does not seek to replace current mechanistic descriptions of adaptive immunity. Instead, it identifies organizational quantities that should become experimentally observable if repertoire formation is indeed governed by constrained probabilistic sampling. Unlike conventional approaches, which primarily focus on receptor affinity, clonal frequency or antigen specificity, we predict measurable properties describing the repertoire as a collective mathematical object such that the observables are not individual lymphocyte clones but statistical descriptors characterizing the full receptor population.

The first prediction concerns repertoire redundancy. If negative dependence contributes to repertoire formation, clonotypes sharing highly similar receptor sequences or recognizing closely related antigenic determinants should be underrepresented relative to expectations based on independent sampling. This reduction should persist even after controlling for differences in recombination frequency, clonal expansion and antigen exposure. Modern high-throughput immune repertoire sequencing provides direct measurements of receptor similarity, allowing comparison between observed clone distributions and those generated under null models assuming statistical independence. Therefore, we predict that healthy immune repertoires should exhibit significantly lower functional redundancy than randomized repertoires preserving the same clone frequencies.

A second prediction concerns representative immune sampling. The Kadison–Singer interpretation suggests that relatively small subsets of circulating lymphocytes may preserve the principal statistical organization of the complete immune repertoire. Consequently, covariance matrices computed from moderate blood samples should approximate those obtained from much larger receptor populations after appropriate normalization. This prediction differs from conventional sampling theory because preservation concerns spectral structure rather than simple clone frequencies. If confirmed experimentally, relatively small samples could accurately characterize global repertoire organization despite containing only a fraction of all circulating clonotypes.

Further, we predict that adaptive immune repertoires occupy broad compatibility basins rather than isolated optimal configurations. If log-concavity provides an appropriate geometric description, neighboring repertoires differing by relatively small numbers of clonotypes should display gradual rather than abrupt variation in global diversity, self-reactivity and antigen coverage. Longitudinal immune sequencing performed before and after vaccination, ageing or transient infection should reveal continuous trajectories through repertoire space, instead of discontinuous transitions between unrelated immune states. This behavior would indicate that compatible repertoires are organized within extended regions of high probability rather than around isolated optima.

Another prediction concerns entropy and immune diversity. Entropy maximization implies that immune selection should preserve the richest repertoire compatible with physiological constraints instead of maximizing affinity alone. Vaccination or controlled antigen exposure should increase effective repertoire entropy, while maintaining bounded redundancy and self-reactivity. Conversely, pathological immune states characterized by excessive clonal dominance are expected to exhibit reduced repertoire entropy together with increased redundancy. This prediction distinguishes constrained probabilistic exploration from purely affinity-driven optimization, where repertoire diversity is not necessarily preserved.

Also, we predict measurable consequences for network organization. If random-forest connectivity captures biologically meaningful interactions among receptor populations, healthy repertoires should display intermediate levels of global connectivity sufficient to maintain coordinated immune responses without excessive synchronization. Chronic inflammatory diseases or severe immune deficiencies may instead shift repertoire connectivity toward opposite extremes, either producing fragmented receptor communities or excessively interconnected interaction

networks. These predictions can be evaluated by building receptor similarity graphs from sequencing data and measuring graph connectivity through established network-theoretical methods.

The most distinctive prediction concerns spectral organization. Spectral independence implies that local perturbations should be spatially and functionally confined within the repertoire, instead of propagating throughout the entire immune system. Following vaccination, transient infection or selective depletion of individual lymphocyte populations, the dominant eigenvalue of receptor interaction networks should remain bounded despite measurable local changes in clone abundance. In contrast, autoimmune diseases, persistent inflammatory disorders or dysregulated immune activation may exhibit progressive increases in spectral radius, indicating enhanced propagation of local perturbations across the repertoire.

Our predictions naturally extend to developmental and evolutionary processes. During childhood, progressive expansion of the adaptive immune repertoire should primarily increase entropy and representative sampling, while preserving bounded redundancy and stable spectral organization. Healthy ageing may gradually reduce repertoire entropy, but should preserve the overall geometric structure of compatibility basins. Pathological ageing, chronic infection or immune exhaustion could instead produce simultaneous reductions in entropy, increased redundancy, fragmentation of compatibility landscapes and elevated spectral influence.

Collectively, these transform abstract concepts originating from probability, combinatorics and spectral graph theory into measurable properties of adaptive immunity that can be evaluated using current immune repertoire sequencing, single-cell technologies and network analysis.

Conclusions

We suggest that several independent developments in modern probability theory, combinatorics and spectral graph methods could be interpreted collectively as a mathematical perspective on immune repertoire selection. Rather than focusing on individual receptors, our formulation treats the repertoire's formation as constrained probabilistic sampling within a high-dimensional compatibility landscape. Global immune organization could constitute an additional level of biological description governed by probabilistic and geometric constraints, with diversity, robustness and self-tolerance partly reflecting the statistical architecture of the receptor population.

Conventionally, immune diversity is characterized through the number and frequency distribution of clonotypes. Our approach considers diversity as one component of a broader organizational state. Repertoires containing similar numbers of clonotypes could differ substantially in redundancy, compatibility, spectral organization and representative sampling and consequently display different collective properties despite comparable conventional diversity. Likewise, immune tolerance could have an ensemble-level component complementary to central deletion, peripheral regulation and immune checkpoints. Negative dependence could constrain excessive accumulation of similar receptors, entropy could preserve broad pathogen coverage within biological restrictions and spectral independence could limit amplification of local perturbations. Still, our perspective introduces a geometric interpretation of immune robustness. Adaptive immunity continuously experiences fluctuations caused by receptor generation, antigen exposure and cellular turnover, while physiological function is generally stable over much longer timescales. In addition to regulatory feedback, pathway redundancy and cellular homeostasis, robustness could depend on the repertoire occupying broad, stable regions of a high-dimensional probability landscape. Neighboring repertoire configurations could then possess comparable functional properties, permitting continuous adaptation without substantial disruption of collective organization. Representative subset preservation provides a complementary interpretation of immune surveillance. The Kadison–Singer perspective raises the possibility that relatively small receptor subsets could preserve important statistical and spectral characteristics of much larger populations. If experimentally supported, effective surveillance would depend not only on the number of circulating lymphocytes but also on how efficiently sampled populations represent the organization of the broader repertoire.

Our perspective could extend beyond adaptive immunity. Microbial communities, neuronal populations, stem-cell differentiation, ecological assemblages and evolving protein families all involve selection from vast configuration spaces while preserving some combination of diversity, robustness and functional integration. These systems share the organizational challenge of generating coherent global states through distributed local events. Therefore, probability theory, combinatorics and spectral methods could provide a common mathematical language for investigating this general problem across biological scales. By translating concepts from modern discrete mathematics into falsifiable properties of immune repertoires, our proposal provides a test case for a broader hypothesis: that the step from local biological selection to global organization can be described through common probabilistic, geometric and spectral approaches.

Authors' contributions: The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Ethics approval and consent to participate: Ethics approval statement is not applicable. This research does not contain any studies with human participants or animals performed by the Author.

Consent: for publication: The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Availability: of data and materials: All data and materials generated or analyzed during this study are included in the manuscript. The Author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Declaration: of generative AI and AI-assisted technologies in the writing process: During the preparation of this work, the author used ChatGPT 5.3 to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

References

1. Adams, N. M., S. Grassmann and J. C. Sun. 2020. "Clonal Expansion of Innate and Adaptive Lymphocytes." *Nature Reviews Immunology* 20 (11): 694–707. <https://doi.org/10.1038/s41577-020-0307-4>.
2. Anari, Nima and Shayan Oveis Gharan. 2015. *The Kadison-Singer Problem for Strongly Rayleigh Measures and Applications to Asymmetric TSP*. arXiv:1412.1143.
3. Anari, Nima, Shayan Oveis Gharan and Cynthia Vinzant. 2018. "Log-Concave Polynomials I: Entropy and a Deterministic Approximation Algorithm for Counting Bases of Matroids." In *Proceedings of the IEEE Symposium on Foundations of Computer Science*.
4. Anari, Nima, Kuikui Liu and Shayan Oveis Gharan. 2020. *Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model*. arXiv:2001.00303.
5. Asadpour, Arash, Michel X. Goemans, Aleksander Mądry, Shayan Oveis Gharan and Amin Saberi. 2010. "An $O(\log n / \log \log n)$ -Approximation Algorithm for the Asymmetric Traveling Salesman Problem." In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, 379–389.
6. Borcea, Julius, Petter Brändén and Thomas M. Liggett. 2009. "Negative Dependence and the Geometry of Polynomials." *Journal of the American Mathematical Society* 22: 521–567.
7. Chen, J., J. Liu and X. Cao. 2025. "Functional and Metabolic Heterogeneity of Dendritic Cells in Self-Tolerance and Autoimmunity." *Immunological Reviews* 336 (1): e70068. <https://doi.org/10.1111/imr.70068>.
8. Daëron, M. 2022. "The Immune System as a System of Relations." *Frontiers in Immunology* 13: 984678. <https://doi.org/10.3389/fimmu.2022.984678>.

9. Deimel, L. P., Y. Nishimura, G. S. Silva Santos, V. A. Baharani, B. Hernandez, T. Y. Oliveira, A. J. MacLean, et al. 2025. "Clonal Expansion and Diversification of Germinal Center and Memory B Cell Responses to Booster Immunization in Primates." *Cell Reports* 44 (8): 116142. <https://doi.org/10.1016/j.celrep.2025.116142>.
10. Domínguez-Andrés, J., J. C. Dos Santos, S. Bekkering, W. J. M. Mulder, J. W. M. van der Meer, N. P. Riksen, L. A. B. Joosten and M. G. Netea. 2023. "Trained Immunity: Adaptation within Innate Immune Mechanisms." *Physiological Reviews* 103 (1): 313–46. <https://doi.org/10.1152/physrev.00031.2021>.
11. Goel, Ashish, Sanjeev Khanna, Sharath Raghvendra and Hongyang Zhang. 2015. "Connectivity in Random Forests and Credit Networks." In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2037–2048.
12. Hochhaus, G. 2008. "Relative Receptor Affinity Comparisons among Inhaled/Intranasal Corticosteroids: Perspectives on Clinical Relevance." *Respiratory Research* 9 (1): 75. <https://doi.org/10.1186/1465-9921-9-75>.
13. Ibáñez-Molero, S., J. Veldman, J. Simon Nieto, J. J. H. Traets, A. George, K. Hoefakker, A. Karomi, et al. 2026. "Tumour-Reactive Heterotypic CD8 T Cell Clusters from Clinical Samples." *Nature* 649 (8096): 467–76. <https://doi.org/10.1038/s41586-025-09754-w>.
14. Ke, Q., A. N. Greenawalt, V. Manukonda, X. Ji and R. M. Tisch. 2023. "The Regulation of Self-Tolerance and the Role of Inflammasome Molecules." *Frontiers in Immunology* 14: 1154552. <https://doi.org/10.3389/fimmu.2023.1154552>.
15. Kocher, K., F. Drost, A. M. Tesfaye, C. Moosmann, C. Schüle, M. Grotz, E. D'Ippolito, et al. 2025. "Vaccination-Induced T Cell Responses Maintain Polyclonality with High Antigen Receptor Avidity." *Science Immunology* 10 (112): eadu6730. <https://doi.org/10.1126/sciimmunol.adu6730>.
16. Kumar, P., S. Saini, S. Khan, S. Surendra Lele and B. S. Prabhakar. 2019. "Restoring Self-Tolerance in Autoimmune Diseases by Enhancing Regulatory T-Cells." *Cellular Immunology* 339: 41–49. <https://doi.org/10.1016/j.cellimm.2018.09.008>.
17. Mesin, L., A. Schiepers, J. Ersching, A. Barbulescu, C. B. Cavazzoni, A. Angelini, T. Okada, T. Kurosaki and G. D. Victora. 2020. "Restricted Clonality and Limited Germinal Center Reentry Characterize Memory B Cell Reactivation by Boosting." *Cell* 180 (1): 92–106.e11. <https://doi.org/10.1016/j.cell.2019.11.032>.
18. Oveis Gharan, Shayan, Amin Saberi and Mohit Singh. 2011. "A Randomized Rounding Approach to the Traveling Salesman Problem." In *Proceedings of the IEEE Symposium on Foundations of Computer Science*.
19. Sjöström, D. J., B. Grill, E. Ambrosetti, A. A. Veetil, C. Mohlin, A. I. Teixeira, G. Oberdofer and S. Bjelic. 2023. "Affinity Matured Transferrin Receptor Apical Domain Blocks Machupo Virus Glycoprotein Binding." *Journal of Molecular Biology* 435 (20): 168262. <https://doi.org/10.1016/j.jmb.2023.168262>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.