Astronomy and Computing 56 (2026) 101123

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full Length Paper

Local intrinsic dimensionality as an estimator of random generator uniformity

Ashot Chilingarian

Yerevan Physics Institute, Alikhanyan Brothers st. 2, Yerevan 36, Armenia

ARTICLE INFO

Keywords:
Random number generators
Intrinsic dimensionality
Monte Carlo simulation
Large-scale structure of the universe

ABSTRACT

The uniformity of random point distributions is vital for all types of simulations, numerical integration, and data mining in high-dimensional spaces. In this study, we suggest using the parameters of the local intrinsic dimensionality (LID) estimates as a sensitive, scalable metric to evaluate the spatial uniformity of random number generators. We estimate LID at 10,000 fixed reference points within the unit N-dimensional hypercube to compare the robustness against border effects of both a deterministic generator and a pseudorandom generator. Based on the bias and standard deviation of LID distributions, we develop a metric indicating the generator's uniformity as a function of the size of the random sample. Our findings show that the deterministic generator performs notably better, delivering accurate spatial coverage of the feature space. Unlike traditional benchmarking methods—such as spectral tests and discrepancy analysis—that measure global uniformity, our approach evaluates local geometric regularity by analyzing the variance and bias of LID estimates. This provides a highly sensitive method for assessing the spatial uniformity of random number generators in high-dimensional domains. The proposed approach is especially valuable for applications in simulation, optimization, and dimensionality analysis, where spatial coherence is crucial.

1. Introduction

Random number generators (RNGs) support a wide range of numerical methods in computational science, including Monte Carlo simulations, numerical integration, and stochastic optimization. While traditional measures of generator quality focus on randomness, independence, or period length, many applications require not only statistical randomness but also high-quality spatial uniformity of multivariate point distributions, particularly in astrophysical contexts.

The large-scale structure of the Universe—the cosmic web—is a complex network of high-density clusters, elongated filaments, broad sheet-like walls, and extensive underdense voids. Identifying and measuring these structural elements in galaxy surveys is a key challenge in cosmology. Traditional statistical tools (such as two-point correlations and power spectra) quantify clustering strength but do not directly classify local shapes. A promising method is to use intrinsic dimensionality as a structural indicator. In a 3D galaxy distribution, a dense cluster should appear locally three-dimensional (isotropic), a thin filament one-dimensional (linear), and a wall two-dimensional (planar). Recent studies have demonstrated that the local intrinsic dimension (LID) can vary across a dataset, and detecting such changes can uncover meaningful substructures. This approach suggests that estimating LID could become a powerful way to separate cosmic web components in an unsupervised way.

However, achieving this vision depends on accurate and stable estimation of LID initially. Estimating the intrinsic dimension locally is challenging: it requires determining the dimensionality of the distribution near a point, often using only the distances to a limited set of nearest neighbors. Modern cosmology typically uses Monte Carlo random catalogs—synthetic point distributions. Variations in how well these random points fill space uniformly can influence scientific results, such as clustering measurements or dimension estimates. Therefore, the intrinsic dimensionality should be estimated not from galaxy points but from a sequence of points uniformly covering the cosmic web. In this context, the uniformity of these sequences is the top priority, and methods for comparing the uniformity of test points in N-dimensional manifolds—where the intrinsic dimensionality is examined to uncover complex structures—are of critical importance.

In two recent studies, we highlighted the impact of random generator quality in astrophysical applications. In Chilingarian (2025), a low-discrepancy deterministic sequence was used to sample sparse 3D galaxy point clouds for intrinsic dimension analysis, emphasizing that

E-mail address: chili@aragats.am.

https://doi.org/10.1016/j.ascom.2026.101123
Received 6 January 2026; Accepted 23 April 2026
Available online 23 April 2026

the choice of sampling scheme can affect the stability of dimension inferences. In Chilingarian (2026), we introduced a multiple random search approach for analyzing globular cluster data, illustrating how robust random sampling procedures aid high-dimensional feature discovery. A poorly distributed set of "random" points may introduce artificial anisotropies or clustering, skewing metrics like LID. Thus, the goal of this work is to quantitatively evaluate random generators via LID-based diagnostics. If LID estimates are sensitive to spatial non-uniformity, we can use them as a probe of generator quality in high dimensions, complementing classical tests (such as n-dimensional discrepancy or spectral tests).

1.1. Related work

Low-discrepancy deterministic sequences, such as those proposed by Halton (1960) and Sobol (1967), are designed for this purpose and offer superior coverage compared to purely pseudorandom generators. Both Halton and Sobol sequences are crafted to fill the unit hypercube uniformly; however, they differ fundamentally in their construction and distribution behavior. Halton sequences rely on radical inverses in mutually prime bases (e.g., 2, 3, 5), leading to a unidirectional expansion that can reveal correlation patterns in higher dimensions. In contrast, Sobol sequences are generated using base-2 direction numbers optimized to distribute points evenly along all coordinate axes, especially when sample sizes are powers of two. While both sequences achieve low global discrepancy, their local structures and uniformities vary, influencing nearest-neighbor statistics and metrics such as intrinsic dimensionality. Modern high-quality pseudorandom generators, such as PCG64 (O'Neill, 2014), implemented on platforms such as NumPy (Lam et al., 2015), are now popular due to their computational efficiency and excellent statistical properties. To compare generator performance from a geometric perspective, we adopt a new metric based on local intrinsic dimensionality (LID), as calculated by the local dimensionality algorithm.

Braams (1974) introduced this algorithm, which estimates local dimensionality by analyzing the relationship between the neighborhood radius and the number of points within it, thereby enabling accurate mapping of nonlinear manifolds. In 1989, A. Chilingarian refined and implemented this approach to improve local dimensionality estimation in empirical datasets (Chilingarian and Harutunyan, 1989; Chilingarian, 1992). His updated version (the TIDIM algorithm) removed the assumption of strict homogeneity, enabling the method to be applied to real-world data such as high-energy cosmic rays or multiparticle production at colliders.

Random number generators are usually evaluated using well-known criteria, such as spectral tests, discrepancy measures, statistical uniformity tests, and lattice-structure analysis. For example, the spectral test examines the spacing between hyperplanes containing generated points in multidimensional space (Knuth, 1997). Discrepancy theory, applied to low-discrepancy sequences such as Sobol and Halton, measures how much the sequence deviates from uniformity via coverage gaps (Niederreiter, 1992). Other tests include the chi-square, Kolmogorov–Smirnov, and serial-correlation tests (Marsaglia, 1968), which assess the statistical properties of generator outputs rather than their geometric distribution.

Unlike traditional benchmarks such as discrepancy measures or spectral tests, which assess uniformity across the entire space, our approach is local, scalable, and geometry-aware. It measures how evenly a generator fills neighborhoods, using LID variance and root-mean-square error (RMSE) as quality metrics. To our knowledge, this is the first time intrinsic dimensionality has been used as a quantitative benchmark for generator uniformity.

We selected two uniform point generators for analysis: the Sobol sequence, a widely used low-discrepancy deterministic generator, and NUMBA's PCG64-based generator, a high-quality modern pseudorandom generator included in Python's numerical computing ecosystem. These choices span the spectrum from globally structured (Sobol) to statistically uniform (PCG64), enabling a meaningful comparison of the two approaches. Reference points were generated using another deterministic uniform reference set (Halton, 1960) at varying sample sizes, and the resulting variances and biases in the LID distribution were compared.

Recent research on intrinsic dimension and local-ID has grown quickly in machine learning and statistics. For reproducible benchmarking, several community toolkits now implement and compare many estimators across synthetic and real datasets (e.g., Bac et al., 2021). Beyond traditional kNN log-ratio estimators, recent likelihood-based local-ID methods provide explicit uncertainty estimates and greater robustness across scales (Denti, 2023; Tempczyk et al., 2022). A large body of recent work summarizes the advantages, limitations, and hyperparameter sensitivity of various estimators (Binnie et al., 2025). Our contribution complements these efforts: we do not introduce a new estimator; instead, we investigate how the sampling method (pseudorandom versus low-discrepancy) affects the variance and bias of local uniformity diagnostics on manifolds.

Meanwhile, Monte Carlo methods have advanced toward randomized, application-focused designs that improve robustness in practice, with modern approaches emphasizing effective dimension, scrambling, and projection properties (Keller, 2022; Paulin et al., 2021). These developments motivate our focus on variance-based local-uniformity diagnostics: even when two generators display similar overall discrepancy, their local neighborhood statistics can differ, which is critical when analyzing sparse structures. In cosmology, extracting large-scale structure relies on local statistics sensitive to sampling irregularities and the topology of voids and filaments. Recent research explicitly evaluates cosmic-web topology through clustering-based techniques (Kelesis et al., 2022), emphasizing the importance of understanding how the choice of point generator affects local uniformity measures used in structure detection.

2. Estimation of the LID uniformly filling a unit N-dimensional hypercube

2.1. Theory and estimator definition

LID measures the effective dimension of data near a point by analyzing distances to its nearest neighbors. Its variance across the dataset indicates the uniformity of the underlying point distribution. Understanding how LID variance depends on the random generator's properties provides a valuable metric for assessing the generator's quality.

To evaluate the spatial uniformity of random point distributions, we computed the parameters of the LID distributions at a fixed set of 10,000 reference points generated by the Halton generator (the reference set). To examine the edge effect, we simulated the tested sequences (clouds of points) in two versions: within the same hypercube and in a slightly smaller region, where each side of the N-dimensional cube was reduced from 0–1 to 0.25–0.975 (to analyze border effects).

The LID at each reference point was computed using its k-nearest neighbors (KNN) among cloud points generated by a deterministic uniform generator based on the Sobol sequence (labeled SOBOL) and a pseudorandom generator PCG64 (NumPy) with Numba acceleration (labeled NUMBA). This separation between reference and evaluation sets ensures independence between query locations and surrounding densities, preventing geometric bias in LID variance estimates.

We define the local intrinsic dimensionality at a reference point q via the k-nearest-neighbor distance ratios. Let $r_1(q) \le r_2(q) \le \ldots \le r_{k(q)}$ be the distances from q to its k nearest cloud points. The local dimension estimate at scale *k* is computed as:

$$\hat{d}(q,\ k) = (k-1) \Big/ \sum_{i=1}^{k-1} ln\left(\frac{r_{k(g)}}{r_i(g)}\right).$$

This kNN form is equivalent to the local log–log slope used in the TIDIM algorithm (Chilingarian, 1992) when the neighborhood is parameterized by k rather than by radius.

2.2. Experimental protocol and results

Unless stated otherwise, we evaluate $\hat{d}$ at M = 10,000 fixed reference points generated by a Halton sequence. Using a generator-independent reference set makes the comparison between the tested cloud generators fair, and M = 10,000 provides stable estimates of distribution width (variance/RMSE) while keeping runtime manageable.

Reference points near the cube's edges miss neighbors in several directions. As a result, their neighborhoods are highly asymmetric and may produce biased LID estimates, increasing variance. In Fig. 1, we illustrate the simulation scheme for 2- and 3-dimensional distributions, emphasizing the edge effects. The reference points fill the entire hypercube (blue points), whereas the simulated cloud points used for LID estimation cover only part of it (0.25–0.975 in each dimension).

In Fig. 2, we demonstrate the LID distribution estimates. On the boundaries of the cloud area (0.25–0.975 in each dimension), the dimensionality is significantly larger than 2, as indicated by the green colors at the borders.

We used the SOBOL and NUMBA generators to produce 100,000 points for each cloud. LID estimates with $K = 300$ nearest neighbors were employed to compute LID at 10,000 points generated by the HALTON sequence. The result of each simulation was a distribution of LID values, from which we calculated the mean and standard deviation. See an example of the local dimensionality histogram in Fig. 3, where it is clear that the distribution of LID estimates for the SOBOL generator (a) is much more compact than that of NUMBA (b). The standard deviation (σ) of the LID estimates for the SOBOL cloud is 0.0227, significantly lower than that for the NUMBA cloud (0.1182). Thus, this demonstrates that the uniformity of the SOBOL sequence over the unit square is considerably better than that of the pseudorandom generator.

For $N = 100{,}000$ cloud points, $\sqrt{N} \approx 316$, so k ≈ $\sqrt{N}$ serves as a standard, scale-adaptive heuristic that balances variance (when k is too small) and bias (when k is too large) in kNN-based dimension estimation. Additionally, Fig. 5 shows that the RMSE curve has a broad minimum or plateau around $K \approx 250$–350 for both generators, indicating that the qualitative conclusions (compact SOBOL histogram versus broad NUMBA histogram) are not affected by small variations in k within this stable range. Importantly, all generator comparisons in the manuscript use the median over a k-sequence, rather than a single fixed k, specifically to avoid manually choosing a neighborhood size when the optimal k shifts with N and local density.

In Fig. 4, we illustrate how biases in LID estimates vary with cloud size (n) for both generators in 2D and 3D and for the restricted and nonrestricted cases. In the restricted variant, cloud points are drawn only from the interior (excluding the boundary layer). In the unrestricted variant, the cloud points cover the full unit domain, so a non-negligible fraction lies near the boundaries. For small and moderate N, this produces a pronounced negative bias (underestimation) because neighborhoods are truncated. As N increases, the bias rises toward zero, then reaches a small positive plateau, shifting from underestimation to overestimation in both restricted and unrestricted scenarios, across both generators. Increasing the number of cloud points makes neighborhoods smaller and more localized. This initially enhances the estimator's accuracy, reducing bias near zero. Near boundaries, the distribution becomes anisotropic, resulting in fewer neighbors in one direction. This distorts the local topology and inflates larger LID values. Even in well-distributed point clouds, density effectively decreases near boundaries, causing the estimator to interpret the local volume as expanding, leading to higher-dimensional estimates. In practical terms, for revealing galaxy structure, the restricted setup is best interpreted as a “local uniformity” probe of the quality of interior sampling (thereby reducing boundary-driven artifacts). In contrast, the unrestricted setup better reflects what happens when the estimator must operate across the entire domain, including edge and low-density regions.

Interestingly, the bias is consistently smaller in 3D than in 2D across a wide range of sample sizes, in both restricted and unrestricted setups. Several factors explain this: in 3D, the proportion of points near the boundary is lower because of the additional degree of freedom in volume. The volume growth in 3D ($\propto r^3$) enhances the localization of neighborhoods for a fixed number of neighbors, k, which better supports the assumptions of local uniformity. Distance ratios used in LID estimation become more consistent and predictable in higher dimensions, leading to lower variance and less bias. This pattern emphasizes the connection between sample size, cloud geometry, and ambient dimension in determining the bias and reliability of LID estimators.

3. Selecting neighborhood size k: fixed-k and k-sequence median

To enhance the stability of LID estimation, we used a median-based approach over a set of k values. As shown in Fig. 5, while the variance of LID estimates initially decreases with larger k and then starts to rise, the median LID estimate across the k range consistently results in a lower root mean squared error (RMSE) than any single fixed k value. This method reduces sensitivity to the choice of k and captures stable geometric structure, which is especially useful for irregular or sparse point clouds Fig. 6.

In both cases, the LID was estimated using the TIDIM method at 10,000 fixed reference points generated by the Halton generator. For each value of the number of nearest neighbors, RMSE was calculated as the square root of the sum of squared bias and MSD relative to intrinsic dimensionality. The dashed red line indicates the RMSE for the median ensemble estimate across all tested k values.

$$\text{RMSE(k)} = \sqrt{1} \Big/ \text{M} \sum_{m=1}^{M} \left(\hat{\text{d}}\text{m(k)} - D_{\text{intinsic}}\right)^2.$$

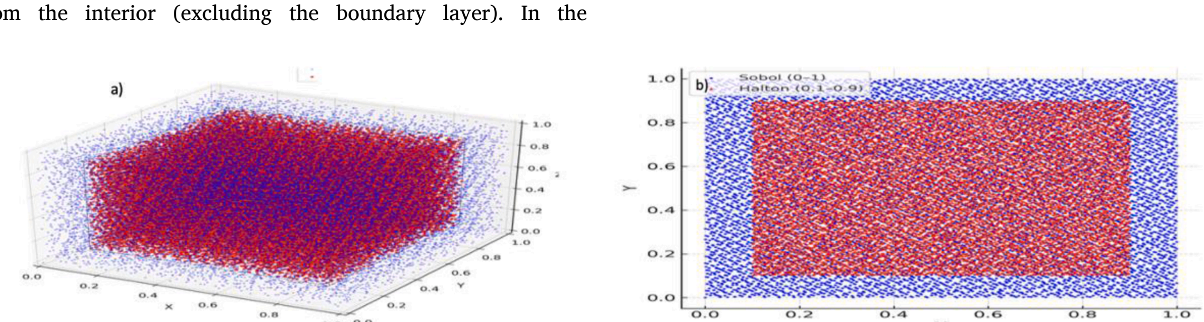

Fig. 1. 3- and 2-dimensional simulation schemes. Blue are reference points (HALTON) where local dimensionality is estimated, including the boundaries of unit hypercubes. However, the tested clouds (red points) are distant from the hypercube's boundaries (SOBOL).

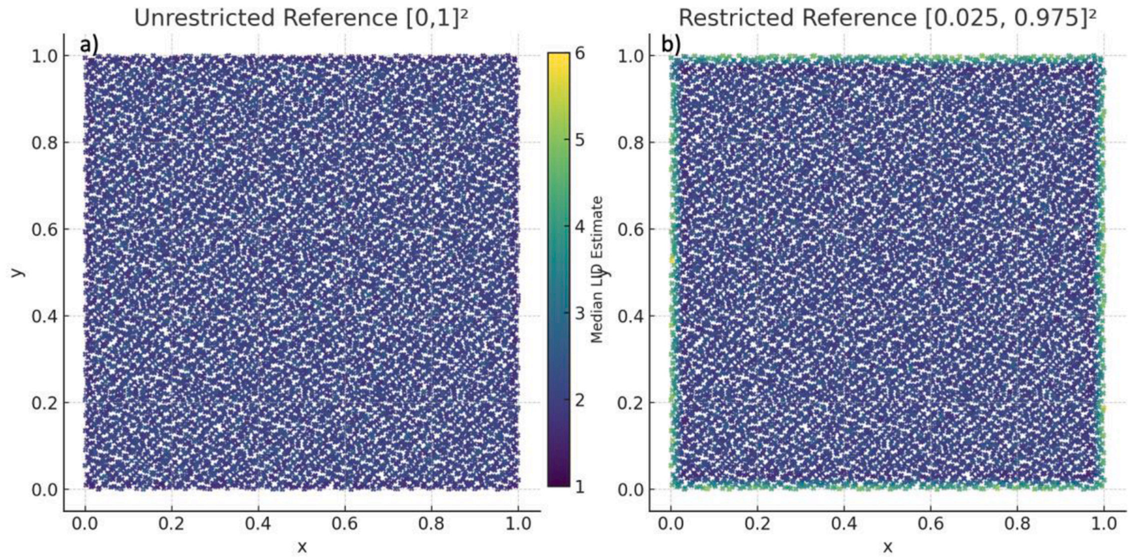

Fig. 2. LID estimates using the kNN method with unrestricted (a) and restricted (b) cloud points. In the right frame, we see higher estimated dimensionality values along all borders (green dots).

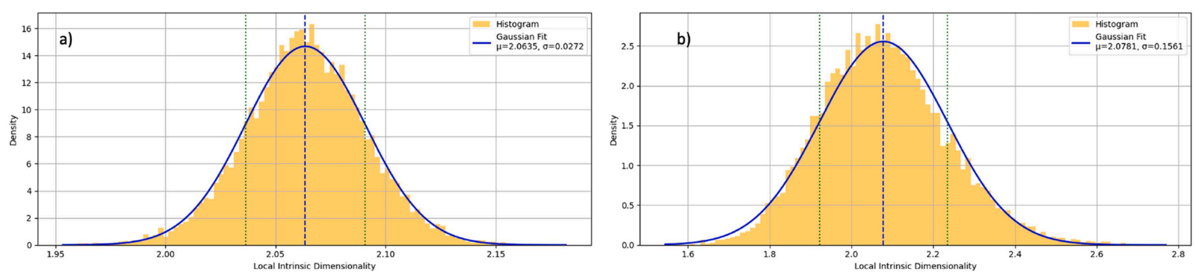

Fig. 3. Gaussian fit of LID distribution. Estimates of the distribution parameters for SOBOL and NUMBA point clouds.

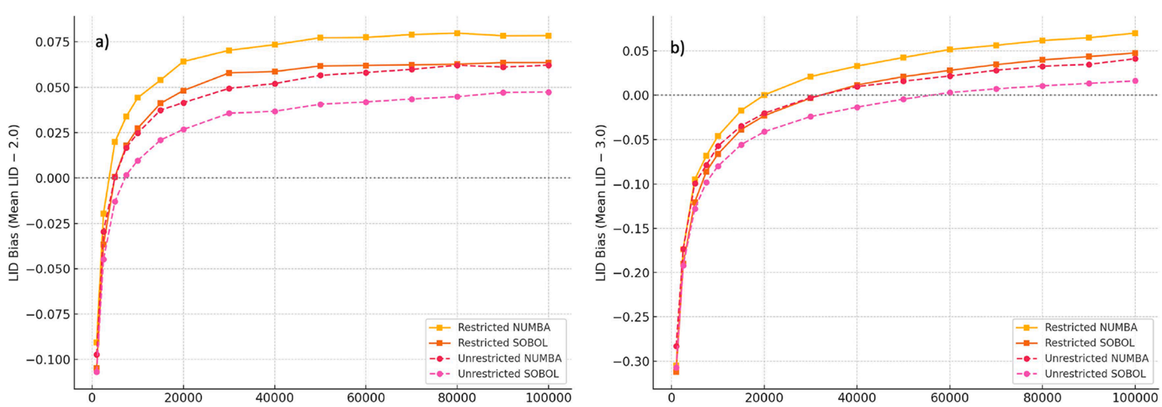

Fig. 4. Bias of LID estimates vs sample size for restricted and unrestricted cloud points in 2D and 3D (a) 2D case: LID bias is defined as (mean LID − 2.0). (b) 3D case: LID bias is defined as (mean LID − 3.0).

The RMSE of the median estimate is lower than that of any individual k, demonstrating the robustness and accuracy of the median-based ensemble. For SOBOL and NUMBA, the RMSE curve flattens near $K = 30$, and the NUMBA generator never reaches the low values observed with SOBOL. SOBOL, a deterministic low-discrepancy sequence, maintains structured coverage of the space; by contrast, the distribution of the LID estimates for NUMBA, a pseudorandom generator, is much more spread out, indicating nonuniformity across the whole domain.

4. Multidimensional tests

In this section, we pushed our tests to 10 dimensions, an order of magnitude beyond the 3 spatial dimensions of the Universe, to challenge the method in severely sparse regimes. In practice, cosmological structure analyses often operate in higher-dimensional feature spaces (e.g., position combined with redshift, velocities, or photometric/derived attributes), so 10D provides a realistic stress test beyond purely 2D–3D geometry. In Fig. 6, we show the RMSE dependance on the sample size.

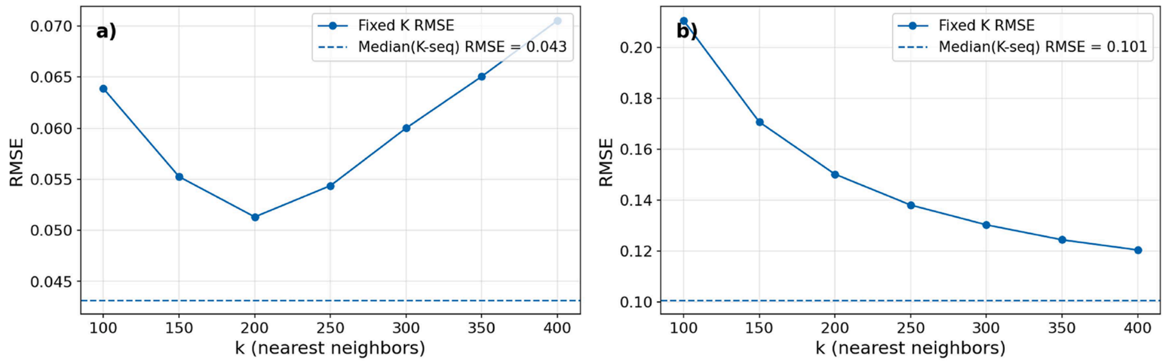

Fig. 5. Comparison of RMSE of LID Estimate vs. k for Two Random Generators. a) RMSE curves for the Sobol sequence with cloud size $N = 100{,}000$; b) RMSE curves for NUMBA (PCG-based pseudorandom generator) with $N = 100{,}000$ cloud points.

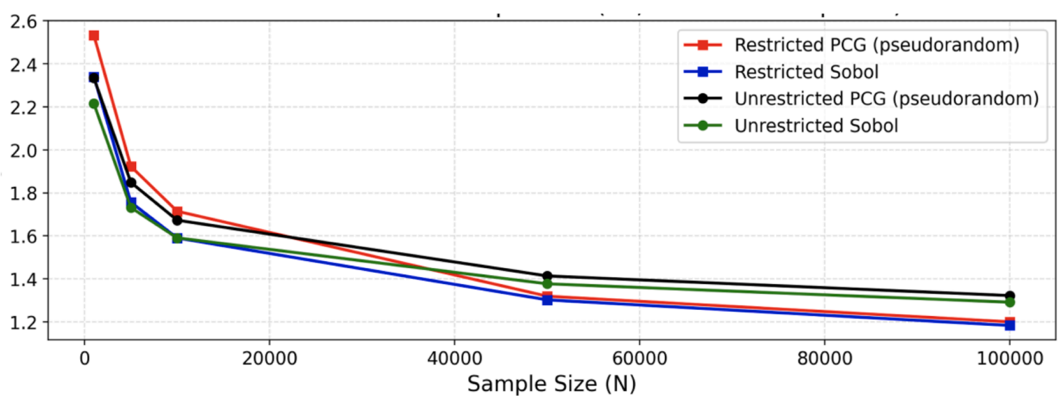

Fig. 6. 10D TiDIM RMSE versus N (10,000 reference points). Restricted: red/blue; Unrestricted: black/green. All curves use the median over the adaptive K-sequence.

We generate a cloud of N points uniformly distributed in the 10D unit hypercube. Two generators are compared for the cloud: (i) a deterministic low-discrepancy Sobol sequence (no scrambling) and (ii) a high-quality pseudorandom generator (PCG64; corresponding to the NUMBA/PCG variant used in Section 3). Intrinsic dimension is evaluated at independent reference points (not at the cloud points), which are generated from a Halton sequence to provide a reproducible and generator-independent probing set. Separating cloud and reference points avoids mixing the sampling non-uniformity of the cloud with the evaluation grid used to diagnose local geometry. In the unrestricted case, cloud points are drawn from the full hypercube. In the restricted case, cloud points are drawn from the interior region to emphasis the boundary effects (near the boundary, kNN balls are truncated and TiDIM becomes biased). Only the cloud-point domain changes; the reference itself always lies within the [0, 1]10 hypercube. For each reference point and each k in k-sequence, TiDIM finds distances $r1 \le r2 \le \ldots \le rk$ to the k nearest cloud points and computes a local dimension estimate. Instead of fixing k a priori, we evaluate the estimator on an adaptive k-sequence and take the median across that sequence as the final, robust estimate at q. The k-sequence is defined by $k_{min} = 10$ and $k_{max} = \mathrm{sqrt}(N)$, with 11 logarithmically spaced values between k_{min} and k_{max} (rounded to integers). The best fixed k that minimizes RMSE drifts with N and with local density, so the k-sequence captures this variability, and the median suppresses outliers from too-small or too-large k.

In both restricted and unrestricted variants, the Sobol-generated cloud produces lower RMSE than the pseudorandom cloud for all tested N, with improvements of approximately 2–9% in this 10D setting. This RMSE-based conclusion aligns with the earlier 2D and 3D RMSE comparisons: low-discrepancy sampling provides more stable local neighborhoods and thus more reproducible kNN-based dimension estimates. The restricted points cloud domain reduces boundary-driven bias and generally improves RMSE compared to the unrestricted cloud domain.

5. Discussion and conclusions

In this study, we introduced a novel method to benchmark the spatial uniformity of random point generators by examining the variance and bias of local estimates of intrinsic dimensionality. Our experiments, conducted in 2D, 3D, and 10D with both restricted and unrestricted point clouds, uncovered several key trends and mechanisms that affect the performance of random generators. The results indicate that low-discrepancy deterministic sequences, such as SOBOL, outperform pseudorandom generators like NUMBA (PCG64-based) in terms of local uniformity. This is shown by lower RMSE in LID estimates and more stable, symmetric, and consistent spatial patterns. In contrast, pseudorandom generators tend to produce higher variance and more prominent edge effects, especially at smaller sample sizes, which lead to distorted LID distributions and unstable estimates near the boundaries.

In low-dimensional spaces, a low-discrepancy sequence (Sobol) distributes points more uniformly than pseudorandom points, causing k-neighborhoods to vary less in size and shape. This directly reduces the variability of kNN distances and improves local dimension estimates, particularly when boundary effects and “holes” are geometrically significant. In 2D–3D, these geometric irregularities are straightforward to generate and easily detected by kNN-based estimators, making Sobol’s advantage apparent.

Interestingly, our findings reveal a non-monotonic pattern in LID bias: at small sample sizes, the LID estimator underestimates the true dimension due to its coarse spatial resolution. As the number of points increases, this bias shifts toward overestimation, especially near domain boundaries, where neighborhood asymmetry increases local volumes. To mitigate the instability caused by k-dependence in nearest-neighbor methods, we use a median k-ensemble estimator for LID. The median k-ensemble consistently yields lower RMSE than any fixed k. This is because the median combines stable information across scales, effectively reducing the influence of local geometric fluctuations and noisy nearest-neighbor distances. Unlike tuning a fixed k, the median does not require prior knowledge of the optimal neighborhood size and automatically adjusts to changes in generator and density.

The methodologies presented here lay the groundwork for new scientific analyses. For instance, applying the median LID estimator to galaxy survey data to create 3D maps of the cosmic web classified by the intrinsic dimension. This could complement existing structure-finding algorithms by providing a continuous "dimension field" identifying filaments, voids, and sheets. Similarly, in star cluster or galaxy evolution studies, LID could help identify dimensionality transitions in phase-space distributions (e.g., when a system moves from a thin disk to a puffed thick disk). Additionally, the idea of using LID variance as a metric of spatial uniformity can be extended to design better sampling strategies for simulations (e.g., importance sampling schemes that maintain uniform local coverage in transformed variables).

In conclusion, as astronomical datasets grow larger and more complex, understanding their structure in an unsupervised, data-driven way becomes increasingly essential. Modern wide-field surveys and multi-wavelength observatories now produce extremely high-dimensional datasets describing galaxy clusters and large-scale structure, such as those obtained by the Vera C. Rubin Observatory survey program, the eROSITA telescope X-ray mission, and the Atacama Cosmology Telescope, which together measure optical, X-ray, and Sunyaev–Zeldovich properties of galaxy clusters (Hilton et al., 2021; Ivezić et al., 2019; Predehl et al., 2021). Local intrinsic dimensionality provides insight into the form of data in high-dimensional spaces, and the improvements discussed in this paper make it a more reliable and practical tool. By integrating ideas from statistical theory, computer science, and domain expertise in cosmology, we developed a solution that meets the rigorous standards of both mathematicians and astronomers. We anticipate this approach will enable discoveries by uncovering hidden low-dimensional manifolds in survey data and ensuring the accuracy of computational experiments, thus supporting the broader goal of understanding the Universe's structure through data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Bac, J., Mirkes, E.M., Gorban, A.N., Tyukin, I., Zinovyev, A., 2021. Scikit-Dimension: a Python package for intrinsic dimension estimation. Entropy 23 (10), 1368. https://doi.org/10.3390/e23101368.

Binnie, A., Dłotko, P., Harvey, A., Malinowski, C., Yim, J., 2025. A Survey of Dimension Estimation Methods arXiv:2507.13887.

Braams, B.J., 1974. Intrinsic dimensionality estimation. Internal report. KVI, Groningen.

Chilingarian, A., Harutunyan, S., 1989. On the possibility of a multidimensional kinematic information analysis using nearest-neighbour estimation of dimensionality. Nucl. Instrum. Method. Phys. Res. A 281, 388.

Chilingarian, A., 1992. Dimensionality analysis of multiparticle production at high energies. Comput. Phys. Commun. 69, 347–359. https://doi.org/10.1016/0010-4655(92)90173-V.

Chilingarian, A., 2025. Intrinsic dimensionality estimation for galaxy distribution structure analysis. Astron. Comput. 53, 100989. https://doi.org/10.1016/j.ascom.2025.100989.

Chilingarian, A., 2026. Search for the best diagnostics in globular clusters (MRSES approach). Astron. Comput. 55, 101047. https://doi.org/10.1016/j.ascom.2025.101047.

Denti, F., 2023. IntRinsic: an R package for model-based estimation of the intrinsic dimension of a dataset. J. Stat. Softw. 106 (9), 1–45. https://doi.org/10.18637/jss.v106.i09.

Ivezić, Ž., Kahn, S.M., Tyson, J.A., 2019. LSST: from science drivers to reference design and anticipated data products. Astrophys. J. 873 (2), 111. https://doi.org/10.3847/1538-4357/ab042c.

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2 (1), 84–90. https://doi.org/10.1007/BF01386213.

Keller, A. (Ed.), 2022. Monte Carlo and Quasi-Monte Carlo methods 2020. Springer. https://doi.org/10.1007/978-3-030-98319-2.

Kelesis, D., Basilakos, S., Papadopoulou Lesta, V., Fotakis, D., Efstathiou, A., 2022. Detecting and analysing the topology of the cosmic web with spatial clustering algorithms I: methods. Mon. Not. R. Astron. Soc. 516 (4), 5110–5124. https://doi.org/10.1093/mnras/stac2444.

Hilton, M., Sifón, C., Naess, S., 2021. The atacama cosmology telescope: DR5 cluster catalog. Astrophys. J. Suppl. Ser. 253 (1), 3. https://doi.org/10.3847/1538-4365/abd023.

Knuth, D.E., 1997. The art of computer programming In: Seminumerical Algorithms, 3rd ed., 2. Addison-Wesley.

Lam, S.K., Pitrou, A., Seibert, S., 2015. Numba: a LLVM-based Python JIT compiler. In: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, pp. 1–6. https://doi.org/10.1145/2833157.2833162.

Marsaglia, G., 1968. Random numbers fall mainly in the planes. Proc. Natl. Acad. Sci. 61 (1), 25–28. https://doi.org/10.1073/pnas.61.1.25.

Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods. SIAM.

O'Neill, M., 2014. PCG: A Family of Simple Fast Space-Efficient Statistically Good Algorithms For Random Number Generation. Technical report HMC-CS-2014-0905, Harvey Mudd College.

Paulin, L., Coeurjolly, D., Iehl, J.C., Bonneel, N., Keller, A., Ostromoukhov, V., 2021. Cascaded Sobol sampling. ACM Trans. Graph. 40 (6), 274. https://doi.org/10.1145/3478513.3480547 article.

Predehl, P., Andritschke, R., Böhringer, H., 2021. The eROSITA X-ray telescope on SRG. Astron. Astrophys. 647 (A1). https://doi.org/10.1051/0004-6361/202039313.

Sobol, I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. Math. Phys. 7 (4), 86–112. https://doi.org/10.1016/0041-5553(67)90144-9.

Tempczyk, P., Teisseyre, P., Liu, D., Alessie, D., Jewell, S., 2022. LIDL: local intrinsic dimension estimator. In: Proceedings of the 39th International Conference on Machine Learning (ICML 2022), 162. PMLR, pp. 21205–21231. https://doi.org/10.48550/arXiv.2206.14882.