HELIOS

Hessian-based **E**nvironment **I**dentifier for **L**arge-scale **O**bservational **S**urvey-data

Suman Sarkar

February 24, 2026

Abstract

HELIOS is a parallelised scientific Python package designed to classify large-scale galaxy environments in spectroscopic redshift surveys. The methodology is physically motivated by gravitational instability theory and tidal field analysis. The pipeline constructs volume-limited samples, estimates smoothed density fields, solves Poisson's equation for the gravitational potential, computes the Hessian (deformation tensor), and classifies cosmic web environments based on eigenvalue collapse criteria. This document presents the physical foundations, mathematical formalism, and software implementation strategy.

1 Physical Background and Motivation

The large-scale structure of the Universe emerges from the gravitational amplification of primordial density fluctuations. In the linear regime, overdense regions grow according to gravitational instability:

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{1}$$

where $\bar{\rho}$ is the mean cosmic matter density. As structure evolves non-linearly, anisotropic gravitational collapse leads to the formation of the **cosmic web**, consisting of:

- Voids (expanding regions),
- Sheets (collapse along one axis),
- Filaments (collapse along two axes),
- Nodes/Clusters (collapse along three axes).

This anisotropic collapse picture naturally arises from the Zel'dovich approximation and tidal field theory. HELIOS operationalises this physical picture for observational galaxy surveys.

2 HELIOS Pipeline

1. Input Galaxy Catalogue

Spectroscopic redshift survey data

↓

2. Volume-Limited Sample Construction

Specific absolute magnitude cut

↓

3. Cloud-In-Cell Density Estimation

Construct overdensity field $\delta(\mathbf{x})$ from 3D distribution of galaxies

↓

4. Gaussian Smoothing

Apply Gaussian smoothing kernel $W_G(r)$

↓

5. Compute gravitational potential Φ from δ

Solve the Poisson Equation in Fourier space

↓

6. Deformation Tensor (Hessian)

Compute tidal tensor: $T_{ij} = \frac{\partial^2 \Phi}{\partial x_i \partial x_j}$

↓

7. Environment Classification

Count the number of eigenvalues Evaluate eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ above threshold λ_{th} to find Void / Sheet / Filament / Node

↓

8. Environment of Galaxies

Interpolate eigenvalues to galaxy positions using Trilinear Interpolation to find the environment of galaxies

3 Volume-Limited Sample Construction

Flux-limited surveys are susceptible to redshift-dependent selection bias. To obtain a uniform tracer density, HELIOS constructs a volume-limited sample before performing the environment classification. The distribution of galaxies mapped by redshift surveys is inhomogeneous in nature due to Malmquist bias, i.e. intrinsically faint galaxies missing at higher redshifts. Instead of using a selection function for counterbalancing this selection bias, we use a volume-limited sample to characterise the cosmic web environment. A sample with an upper bound of absolute magnitude ensures the galaxies with similar intrinsic brightness are accessible across the entire redshift range probed. This is to ensure that the variation in number density occurs only due to the fluctuations in the actual matter distribution of the Universe. The spectroscopic redshifts are used to obtain the exact comoving coordinates of the galaxies in 3D. An upper redshift cut z_c is obtained by assigning the upper limit of the absolute magnitude cut M_c. Galaxies below these two upper limits are chosen for further analysis. This step ensures unbiased environmental statistics.

4 Environment Classification Methodology

4.1 Construction of the Outer parallelepiped with random mock population

Spectroscopic surveys do not occupy a regular cubic geometry. Instead, they cover irregular angular masks with finite redshift depth. However, solving Poisson's equation via Fast Fourier Transform (FFT) requires a rectangular periodic domain. To enable a consistent gravitational potential reconstruction, HELIOS embeds the survey region inside a larger `outer cube` (throughout this document, the parallelepiped will be referred to as a cube) that encompasses the entire survey region.

4.1.1 Outer Cube Construction

Let the comoving coordinates of galaxies be $\mathbf{x} = (x, y, z)$. The minimal bounding cube is defined such that

$$L = \{L_x, L_y, L_z\} = \{\,(x_{\max} - x_{\min}),\, (y_{\max} - y_{\min}),\, (z_{\max} - z_{\min})\,\} \tag{2}$$

The cube volume is therefore

$$V_{\text{cube}} = L_x \times L_y \times L_z \tag{3}$$

The survey region occupies only a sub-volume $V_{\text{survey}} \subset V_{\text{cube}}$. This embedding ensures

- Regular grid construction.
- Compatibility with FFT-based solvers.
- Well-defined boundary conditions.

4.1.2 Need for Random Mock Population

If the outer cube outside the survey region were left empty, the density contrast would artificially drop to $\delta = -1$ outside the footprint. This would generate spurious large-scale tidal forces near the survey boundary. To avoid this, HELIOS populates the cube

outside the survey mask with uniformly distributed mock galaxies having the same mean number density as the volume-limited sample.
Let the observed mean number density be

$$\bar{n} = \frac{N_{\rm gal}}{V_{\rm survey}}. \tag{4}$$

Then the expected number of mock galaxies required is

$$N_{\rm mock} = \bar{n}\,(V_{\rm cube} - V_{\rm survey}) \tag{5}$$

Hence, the region outside the survey mask is populated with mock galaxies drawn from a uniform Poisson distribution. This procedure enforces $\langle\delta\rangle = 0$ over the entire computational domain.

Physically, the random population represents an *unclustered homogeneous background field* and the combined density field becomes

$$\rho_{\rm total}(\mathbf{x}) = \rho_{\rm survey}(\mathbf{x}) + \rho_{\rm mock}(\mathbf{x}) \tag{6}$$

Since the mock galaxies are uniformly distributed, they contribute only to the mean density and not to tidal anisotropy. This crucial step

- Prevents artificial underdensities outside the survey mask.
- Suppresses boundary-induced tidal artefacts.
- Ensures stable Poisson inversion.
- Preserves physical interpretation of deformation tensor.

Without this correction, eigenvalues near survey edges would be biased, leading to systematic misclassification of filaments and nodes.

4.2 Cloud-In-Cell Density Estimation

Galaxies are discretely sampled tracers of the underlying matter field. To estimate the continuous density field, HELIOS uses the Cloud-In-Cell (CIC) scheme. Each galaxy contributes to 8 surrounding grid vertices with weights

$$W = (1 - \Delta x)(1 - \Delta y)(1 - \Delta z) \tag{7}$$

where Δx, Δy, and Δz are the fractional distances (in grid units) from the adjacent grid point to either side of the galaxy. The resulting density field at a given grid is calculated by summing the contributions from all galaxies around the grid within 1 grid unit. i.e.

$$\rho(\mathbf{x}_g) = \sum_p W_p \tag{8}$$

The overdensity is calculated as

$$\delta(\mathbf{x_g}) = \frac{\rho(\mathbf{x_g})}{\bar{\rho}} - 1 \tag{9}$$

Where $\bar{\rho} = N_p/N_g$ with N_p as the number of particles and N_g as the total number of grids. CIC conserves mass and minimises aliasing.

4.3 Gaussian Smoothing

The raw density field contains shot noise. We smooth using a Gaussian kernel,

$$W_G(r) = \frac{1}{(2\pi R_s^2)^{3/2}} \exp\left(-\frac{r^2}{2R_s^2}\right) \tag{10}$$

In Fourier space

$$\tilde{\delta}_s(\mathbf{k}) = \tilde{\delta}(\mathbf{k}) \exp\left(-\frac{k^2 R_s^2}{2}\right) \tag{11}$$

The smoothing scale R_s determines the physical scale of environment classification. Large R_s probes supercluster scales; small R_s probes local finer structure like small filaments.

4.4 Solution of Poisson's Equation on a Discrete Periodic Grid

In comoving coordinates, the gravitational potential Φ is related to the density contrast δ through Poisson's equation:

$$\nabla^2 \Phi(\mathbf{x}) = \delta(\mathbf{x}) \tag{12}$$

where Φ is the peculiar gravitational potential (in suitable normalization).

HELIOS solves Poisson's equation on a regular three-dimensional periodic grid. We employ the **finite-difference discrete Laplacian operator**. For a grid with spacings $(\Delta x, \Delta y, \Delta z)$, the discrete Laplacian in Fourier space is

$$\tilde{\nabla}^2(\mathbf{k}) = -4\left[\frac{\sin^2\left(\frac{k_x \Delta x}{2}\right)}{\Delta x^2} + \frac{\sin^2\left(\frac{k_y \Delta y}{2}\right)}{\Delta y^2} + \frac{\sin^2\left(\frac{k_z \Delta z}{2}\right)}{\Delta z^2}\right] \tag{13}$$

This expression corresponds to the Fourier transform of the second-order central finite-difference approximation

$$\nabla^2 \Phi_{i,j,k} = \frac{\Phi_{i+1,j,k} - 2\Phi_{i,j,k} + \Phi_{i-1,j,k}}{\Delta x^2} \tag{14}$$

Taking the discrete Fourier transform of the density field

$$\tilde{\delta}(\mathbf{k}) = \mathcal{F}[\delta(\mathbf{x})] \tag{15}$$

Poisson's equation becomes

$$\tilde{\Phi}(\mathbf{k}) = \frac{\tilde{\delta}(\mathbf{k})}{\tilde{\nabla}^2(\mathbf{k})} \tag{16}$$

The $k = 0$ mode corresponds to the spatial mean of the density field. Since the potential is defined only up to an additive constant, we set

$$\tilde{\Phi}(\mathbf{k} = 0) = 0. \tag{17}$$

Finally, the gravitational potential in real space is obtained by inverse Fourier transform

$$\Phi(\mathbf{x}) = \mathcal{F}^{-1}[\tilde{\Phi}(\mathbf{k})]. \tag{18}$$

4.5 Deformation Tensor (Tidal Tensor)

The deformation (tidal) tensor is defined as the Hessian of the gravitational potential,

$$T_{ij}(\mathbf{x}) = \frac{\partial^2 \Phi}{\partial x_i \partial x_j}, \tag{19}$$

and quantifies anisotropic gravitational compression or expansion along principal axes.

In HELIOS, spatial derivatives are computed in Fourier space for numerical stability and efficiency. However, rather than using the continuum operator ik_i, we employ the spectral representation of the second-order central finite-difference derivative. For a grid with spacing Δx_i, the discrete first-derivative operator is

$$D_i(k_i) = \frac{\sin(k_i \Delta x_i)}{\Delta x_i}. \tag{20}$$

Taking the Fourier transform of the potential, $\tilde{\Phi}(\mathbf{k})$, the tensor in Fourier space becomes

$$\widetilde{T}_{ij}(\mathbf{k}) = -\tilde{\Phi}(\mathbf{k})\, D_i(\mathbf{k}) D_j(\mathbf{k}), \tag{21}$$

and the real-space field is obtained via inverse Fourier transform.

This formulation is consistent with the discrete Laplacian used in solving Poisson's equation. In the long-wavelength limit ($k\Delta x \ll 1$), $D_i \to k_i$, recovering the continuum expression $T_{ij} \to -k_i k_j \tilde{\Phi}$.

The resulting symmetric tensor field provides the eigenvalues used for cosmic web classification. The deformation tensor describes anisotropic gravitational collapse with the following physical meaning

- Positive eigenvalue $\to$ local compression.
- Negative eigenvalue $\to$ expansion.

4.6 Eigenvalues of the Deformation tensor and the Physical Interpretation

At each grid point, the deformation tensor T_{ij} is a real symmetric 3×3 matrix and therefore admits three real eigenvalues. We denote the ordered eigenvalues as

$$\lambda_1 \geq \lambda_2 \geq \lambda_3. \tag{22}$$

The ordering ensures a consistent identification of the principal axes of collapse and expansion across the volume. Since the trace of the deformation tensor equals the Laplacian of the potential,

$$\sum_{i=1}^{3} T_{ii} = \nabla^2 \Phi = \delta, \tag{23}$$

It follows that the eigenvalues satisfy the important consistency relation

$$\lambda_1 + \lambda_2 + \lambda_3 = \delta. \tag{24}$$

Environment	index	Classification criteria		
Void	0	$\lambda_1 \le \lambda_{\rm th}$	$\lambda_2 \le \lambda_{\rm th}$	$\lambda_3 < \lambda_{\rm th}$
Sheet	1	$\lambda_1 > \lambda_{\rm th}$	$\lambda_2 \le \lambda_{\rm th}$	$\lambda_3 \le \lambda_{\rm th}$
Filament	2	$\lambda_1 \ge \lambda_{\rm th}$	$\lambda_2 > \lambda_{\rm th}$	$\lambda_3 \le \lambda_{\rm th}$
Node	3	$\lambda_1 > \lambda_{\rm th}$	$\lambda_2 \ge \lambda_{\rm th}$	$\lambda_3 \ge \lambda_{\rm th}$

Table 1: Criteria for environment classification

Thus, the local density contrast is equal to the sum of the principal tidal compressions. A positive eigenvalue corresponds to gravitational compression along the associated principal axis, while a negative eigenvalue corresponds to expansion. The number of eigenvalues exceeding a threshold $\lambda_{\rm th}$ determines the environment. The threshold parameter $\lambda_{\rm th}$ accounts for non-linear effects and controls the effective collapse criterion. Usual choices of λ_{th} ranges between $0 \le \lambda_{th} \le 0.6$

4.7 Environment Assignment at Galaxy Positions

The deformation tensor T_{ij} is computed on a regular periodic grid. To assign a cosmic-web environment to galaxies located at arbitrary positions $\mathbf{x}_p$, the tensor field is interpolated from the grid to particle positions using trilinear interpolation.

4.7.1 Trilinear Interpolation

Let a galaxy lie inside a grid cell whose lower vertex has an integer index (i, j, k). Writing the position in grid units, we define the fractional coordinates inside the unit cube as

$$(x_d, y_d, z_d) = (x - i,\ y - j,\ z - k)\,, \qquad 0 \le x_d, y_d, z_d < 1. \tag{25}$$

If f_{abc} denotes the value of a scalar field at the eight cube vertices ($a, b, c \in \{0, 1\}$), the trilinear interpolated value is

$$f(x_d, y_d, z_d) = \sum_{a,b,c=0}^{1} f_{abc}(1 - a + (-1)^a x_d)(1 - b + (-1)^b y_d)(1 - c + (-1)^c z_d). \tag{26}$$

This procedure is applied independently to each component of the tidal tensor T_{ij}, yielding an interpolated tensor $T_{ij}(\mathbf{x}_p)$ at the galaxy position. At each galaxy position, the interpolated tensor is diagonalised to obtain real eigenvalues,

$$\lambda_1 \ge \lambda_2 \ge \lambda_3. \tag{27}$$

The environment is then determined by counting the number of eigenvalues exceeding a threshold $\lambda_{\rm th}$, as shown in Equation 4.6

5 Installation and Execution

5.1 Requirements

- Python 3.9+
- pip
- git

5.2 Installation

```
git clone https://github.com/ss1ts2code/HELIOS.git
cd HELIOS
pip install -e .
```

5.3 Run

```
galenv-find
```

If a Windows PATH warning appears:

```
C:\Users\your-username\AppData\Roaming\Python\Python312\Scripts\galenv-find.exe
```

Input catalogue must be placed at:

```
data/raw/galaxy_catalog.csv
```

Configuration parameters are stored in:

- config/prep_config.toml
- config/classification_config.toml

---

6 Scientific Applications

- Cosmic web mapping
- Environmental quenching studies
- Galaxy morphology-density relation
- Large-scale anisotropic collapse studies
- Redshift-space distortion modelling

Author

Suman Sarkar [https://sites.google.com/view/suman-sarkar/home]
