Application of algebraic topology to datasets

Pratyush Pranav

CRAL, ENS de Lyon

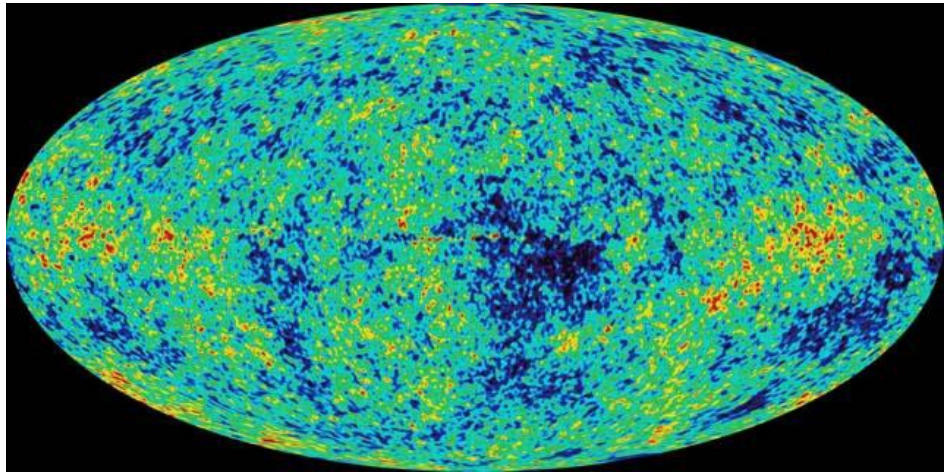
Inhomogeneous cosmologies IV
17th July, 2019

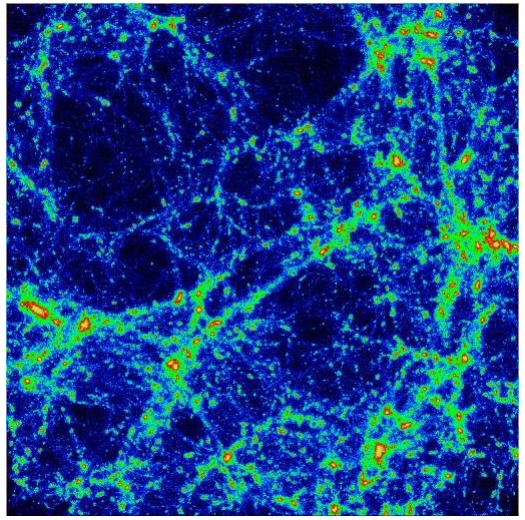
Outline

- Topological Background
  - Topological cycles and holes
  - Genus, Euler characteristic and Minkowski functionals
- Cosmic topology: Results
  - Topology of the CMB (2D scalar field)
  - Topology of simulated data sets (3D Gaussian fields)
- Conclusion

Functions on Manifolds

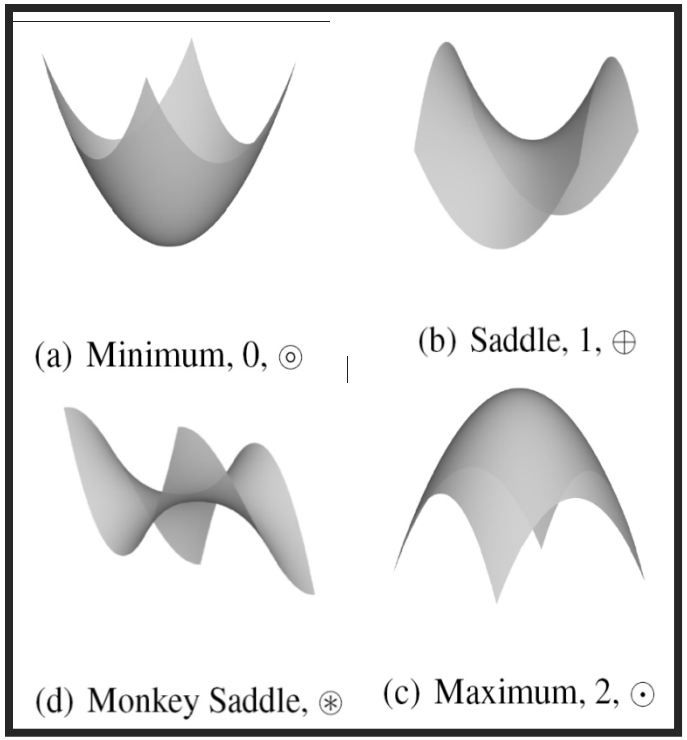
- Primoridal fluctuation field, CMB
- Grid-densities of N-body simulations
- Study the change in topology of a manifold as induced by the excursion sets of the function f





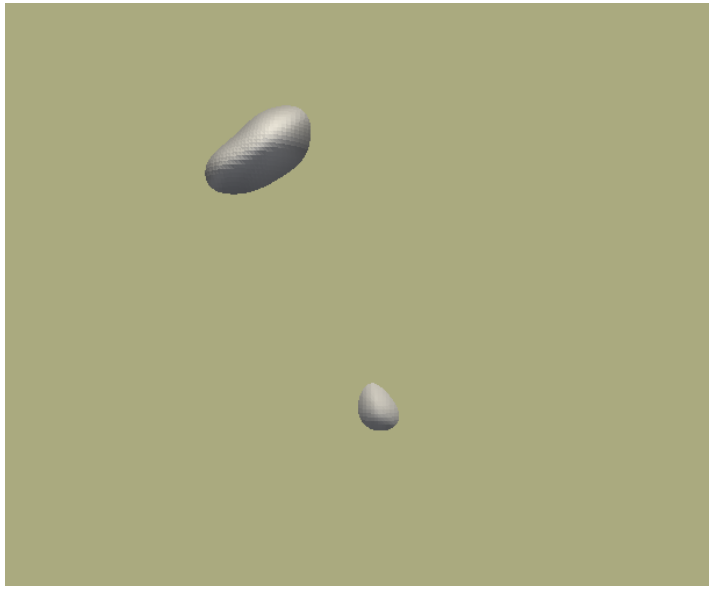
Morse functions

- Let f: $R^n \rightarrow R$
- The function is *nice*
  - » Domain R, R^2, R^3
  - » Gradient ∇f
  - » Critical points $\nabla f = 0$
  - » Smooth functions $\nabla^2 f$

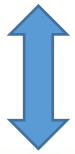


Topological cycles/holes

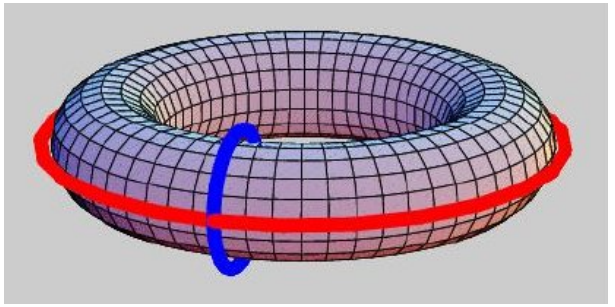
- **intuitive interpretation**



0 dimensional holes :
gaps between
connected objects



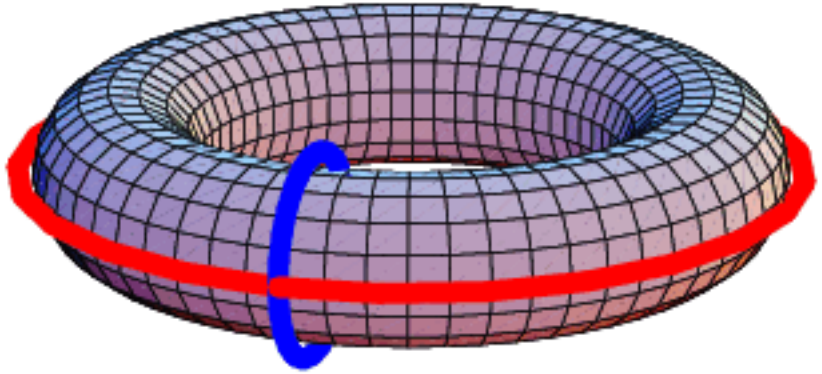
1 dimensional holes :
loops/tunnels



2 dimensional holes :
voids

Genus

For a connected, orientable surface, the Genus has a linear relationship with the maximal number of independent simple closed curves that can be drawn on the surface without rendering it disconnected



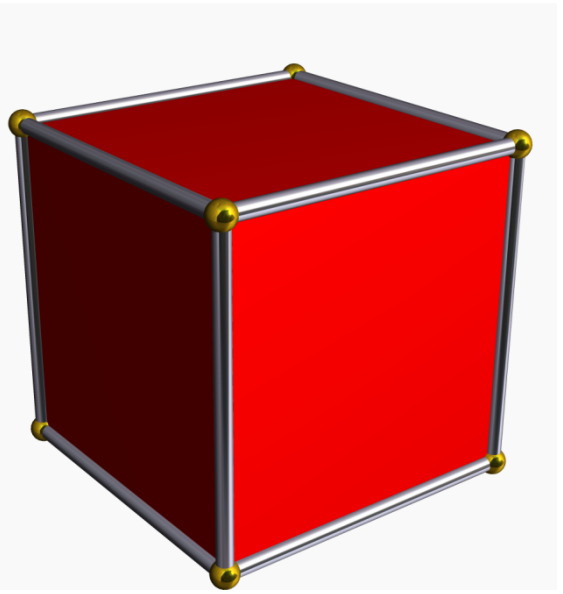
Pranav et. al. MNRAS, 485 (3), 4167-4208 (2019)

Euler characteristic

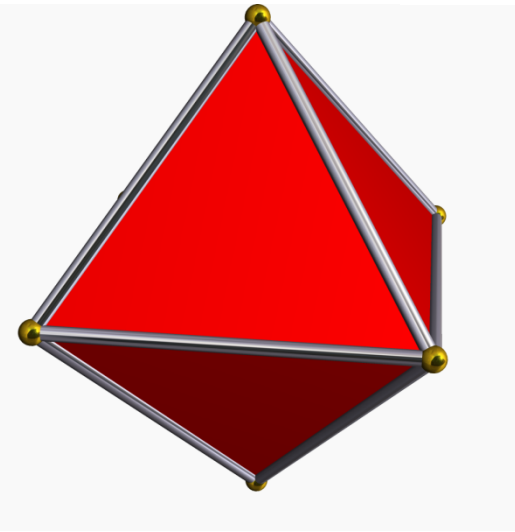
- **Originally defined for polyhedra**

$$\chi = V - E + F$$

Tetrahedron
V = 4, E = 6, F = 4



Cube
V = 8, E = 12, F = 6

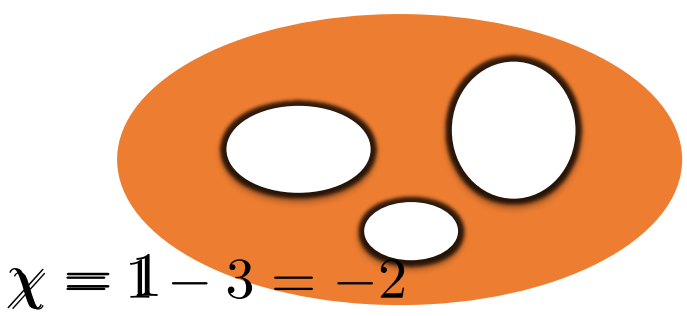


octahedron
V = 6, E = 12, F = 8

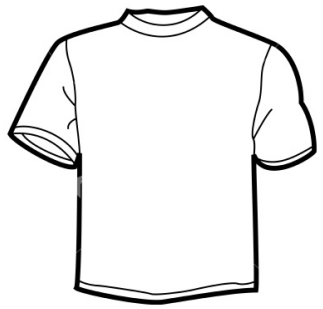
= 2

- **Modern definition through algebraic topology, specifically *Homology***

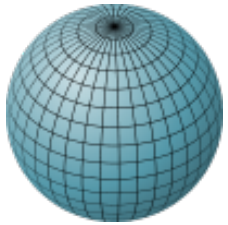
Euler characteristic and/or genus



=



Euler $\chi(A) = 2 - g(A)$ genus



Genus 0
Euler 2

genus 1
Euler 0

genus 2
Euler -2

genus 3
Euler -4

Minkowski Functionals

- For a d-dimensional manifold, *M,* there are (d + 1) Minkowski functionals

$$Q_k (k = 0,\ldots,d)$$

- Predominantly geometric in nature

- While in R^3,

Q_0 : *volume functional*

Q_1 : *area functional*

Q_2 : *integrated mean curvature*

Q_3 : *Gaussian curvature*

Genus, Euler & Betti

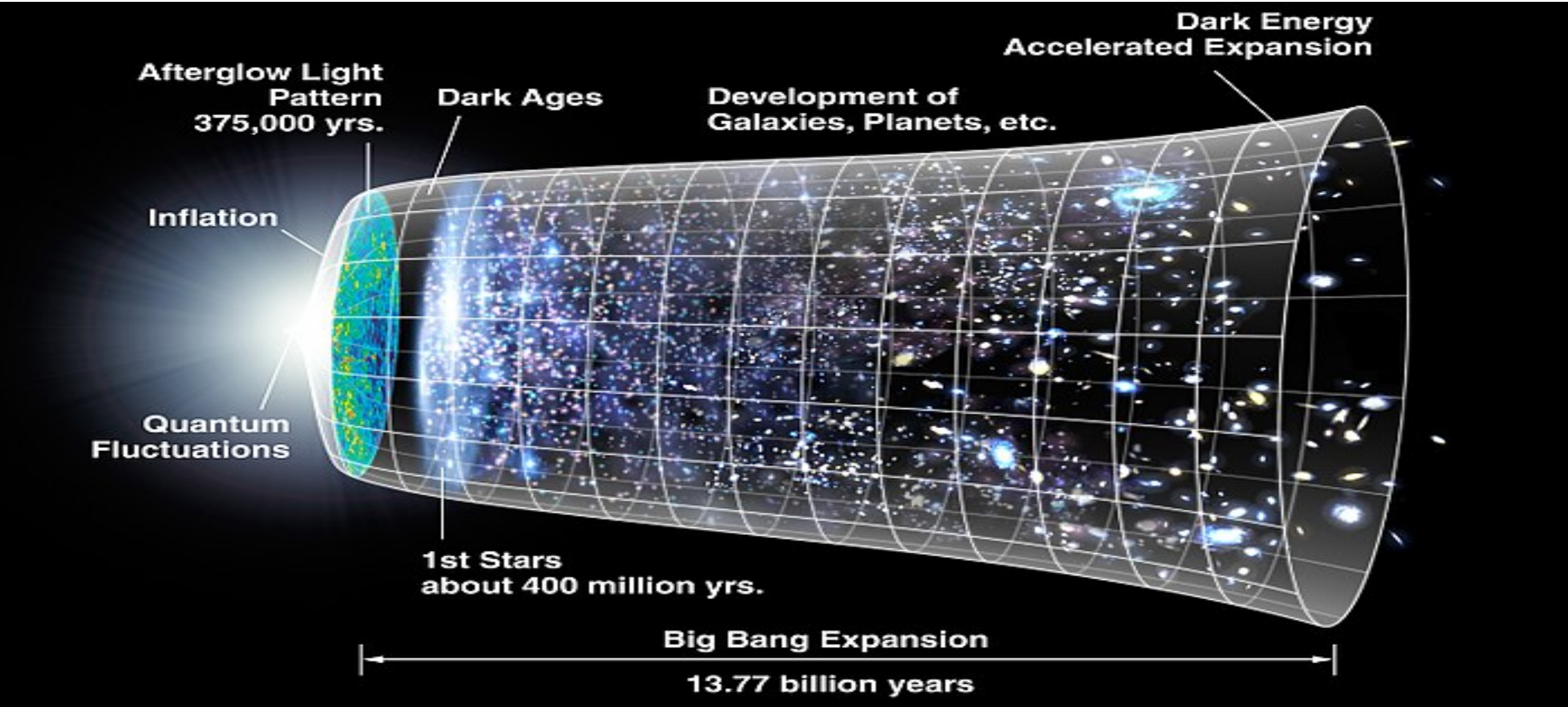
- Euler – Poincare formula

  Relationship between Betti Numbers & Euler Characteristic :

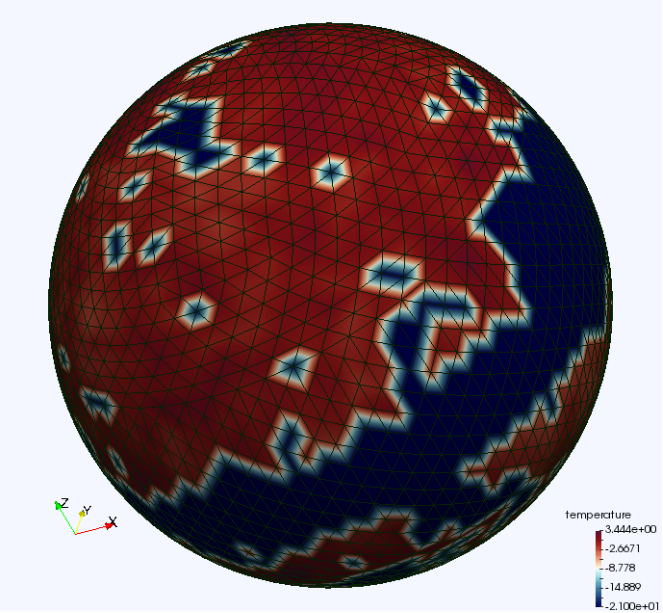
$$\chi = \sum_{k=0}^{d} (-1)^k \beta_k$$

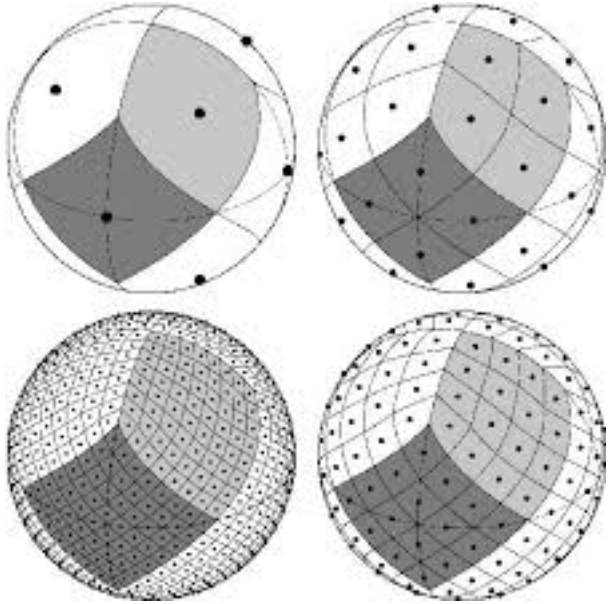
(Unexpected) Topology of the CMB

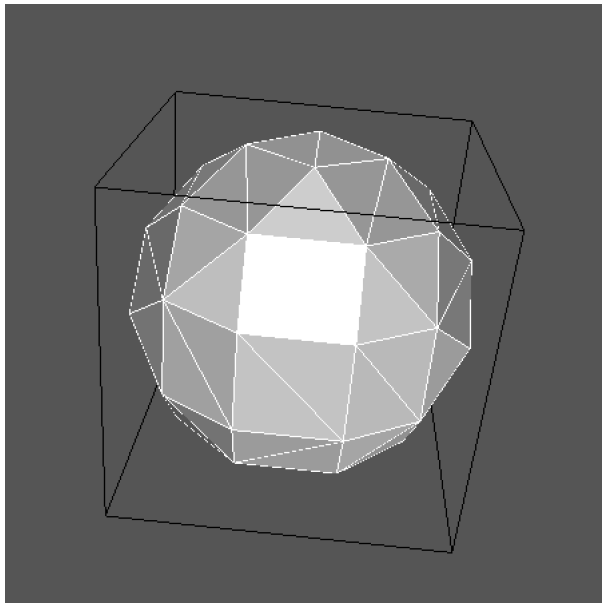
Cosmic timeline



Planck Data

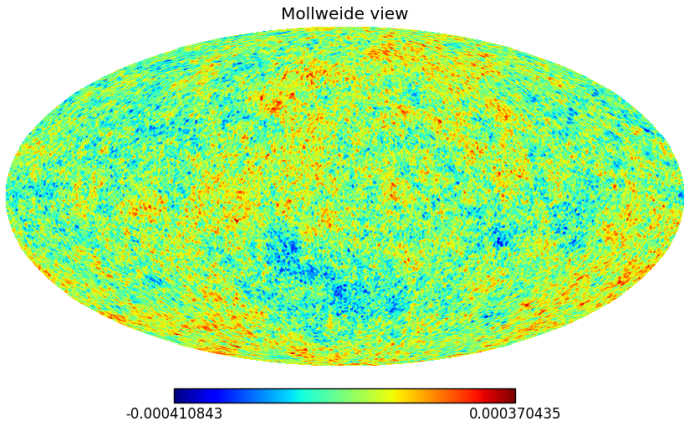


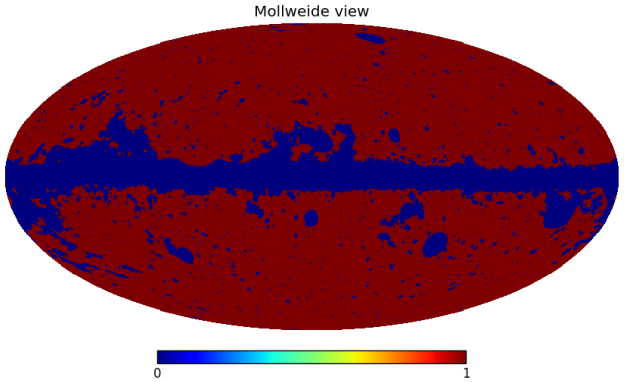




Specified on S2, as the deviation from the background average.

!!! Zone of avoidance - obscured by foregrounds





Planck data

- Sky observed in 8 distinct frequency: 30, 45, 70 (LFI), and 100, 140, 220, 350, 550 and 850 GHz (HFI).
- Final observational CMB maps synthesized from the frequency maps using four different component separation methods: C-R, NILC, SEVEM & SMICA
- Observed maps not reliable in certain zones due to obfuscation by foregrounds
- FFP8 simulations used test the Gaussian hypothesis – initial input for the simulations is a Gaussian random field
- Realistic maps that model the effects for known foregrounds e.g., gravitational lensing, Reyleigh scattering and more
- 1000 maps employed to compute the statistics
- More than 3-sigma deviations at N = 32, to 8 with respect to Gaussian ffp8 simulations, for the components and holes

Relative Homology

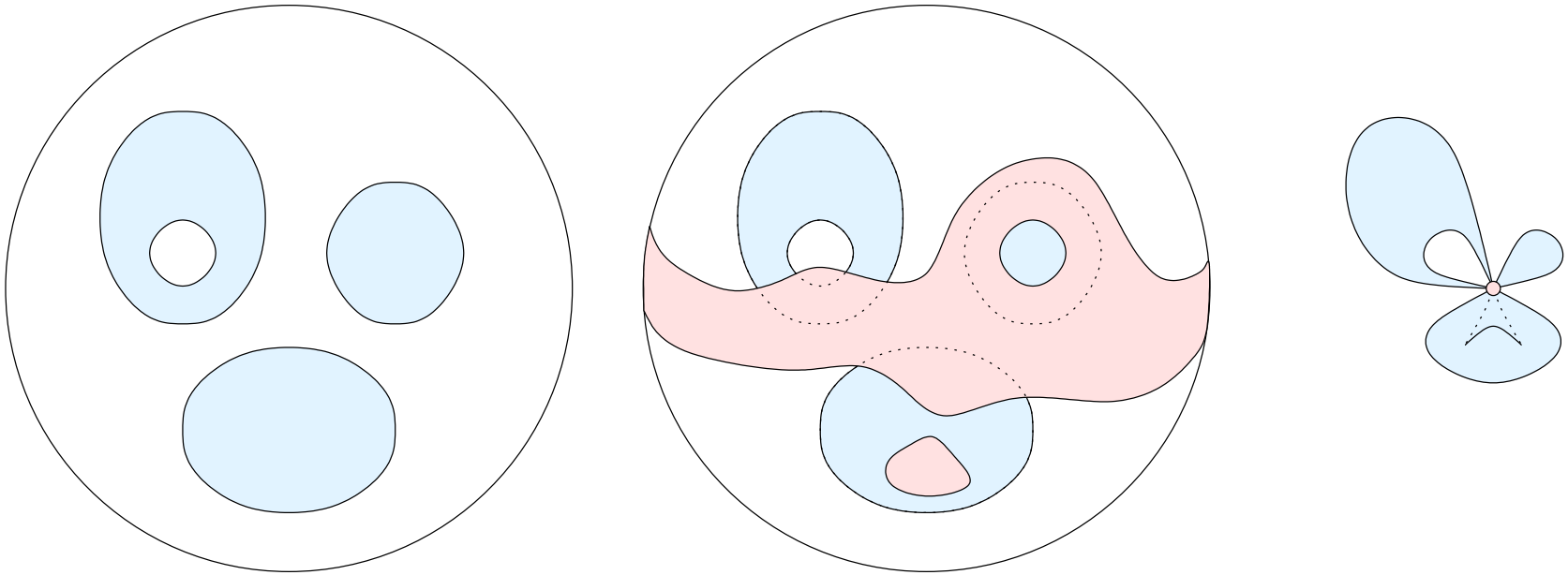


Figure 1: *Left*: A blue excursion set on the sphere consisting of an upper left component with a hole, an upper right component, and a lower component. Its Betti numbers are $\beta_0 = 3, \beta_1 = 1, \beta_2 = 0$, and its Euler characteristic is $\mathsf{EC} = 3 - 1 + 0 = 2$. *Middle*: A pink mask in which the data is not reliable. It covers part of the upper left component and hole, its hole is fully contained in the upper right component, and it overlaps the lower component in two disconnected pieces. *Right*: A visualization of the relative homology groups obtained by shrinking the mask to a point and pulling the excursion set with it. We have $b_0 = 0$ because all three components connect to the shrunken mask, $b_1 = 2$ because the loop in the upper left component is preserved and a new loop in the lower component is formed, and $b_2 = 1$ because the upper right component takes on the shape of sphere. The (relative) Euler characteristic is therefore $\mathsf{EC}_{\mathrm{rel}} = 0 - 2 + 1 = -1$.

Pranav et. al. A&A, submitted

Relative Loops in the CMB

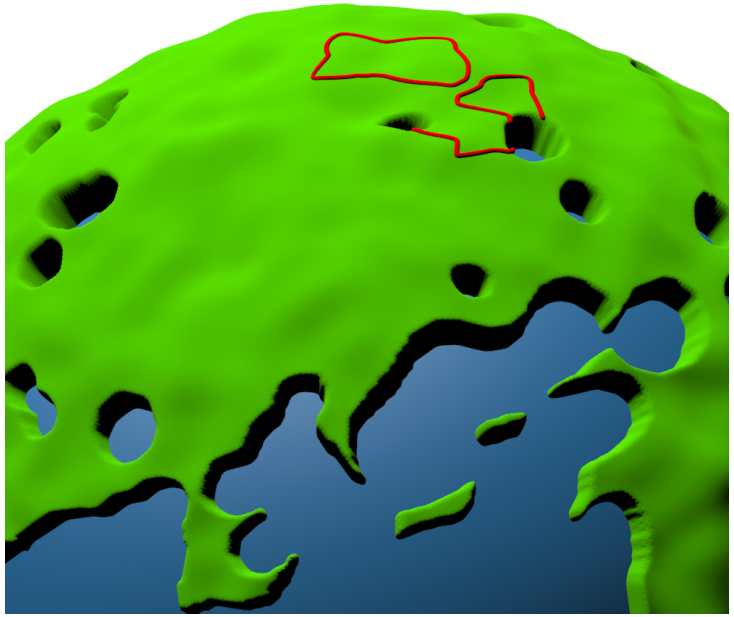
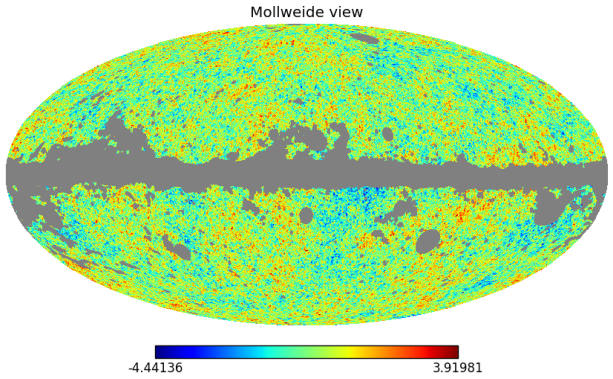
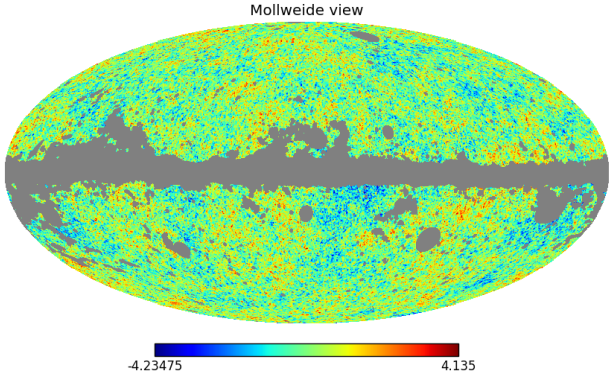
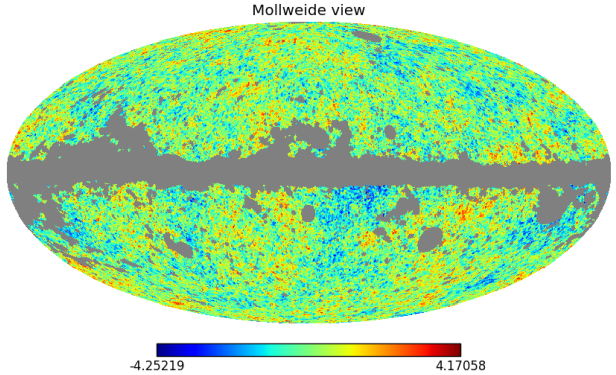


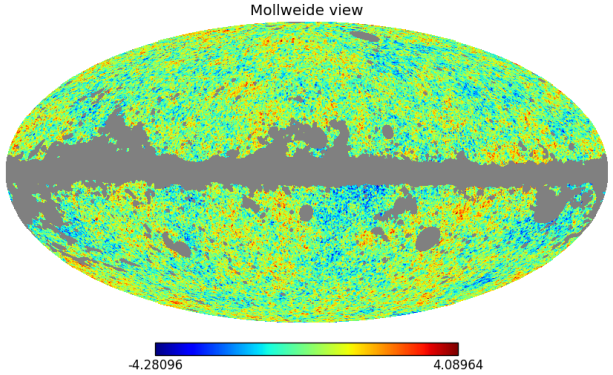
Figure 2: A small section of the sphere of directions, with the temperature field visualized by the green landscape that complements the blue mask drawn at lower altitude. We see one closed loop, surrounding a relative depression of the temperature field, and two open loops, connecting points in the mask along locally highest paths. The visualization is based on the observed CMB maps cleaned using the NILC technique, and smoothed at 4 degrees.

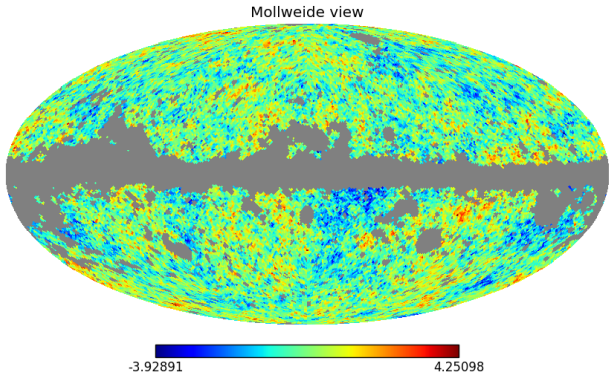
Masked degraded maps

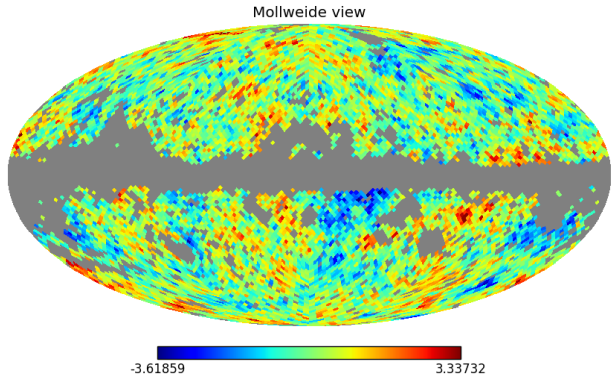


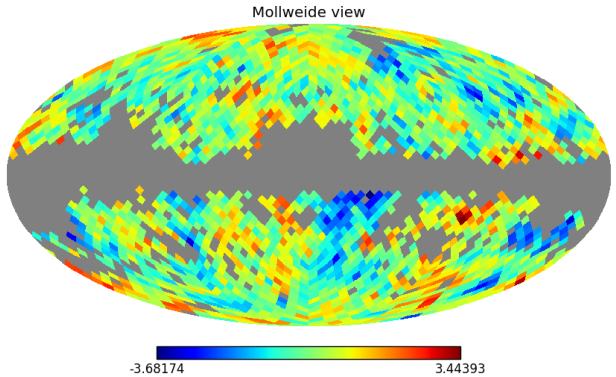




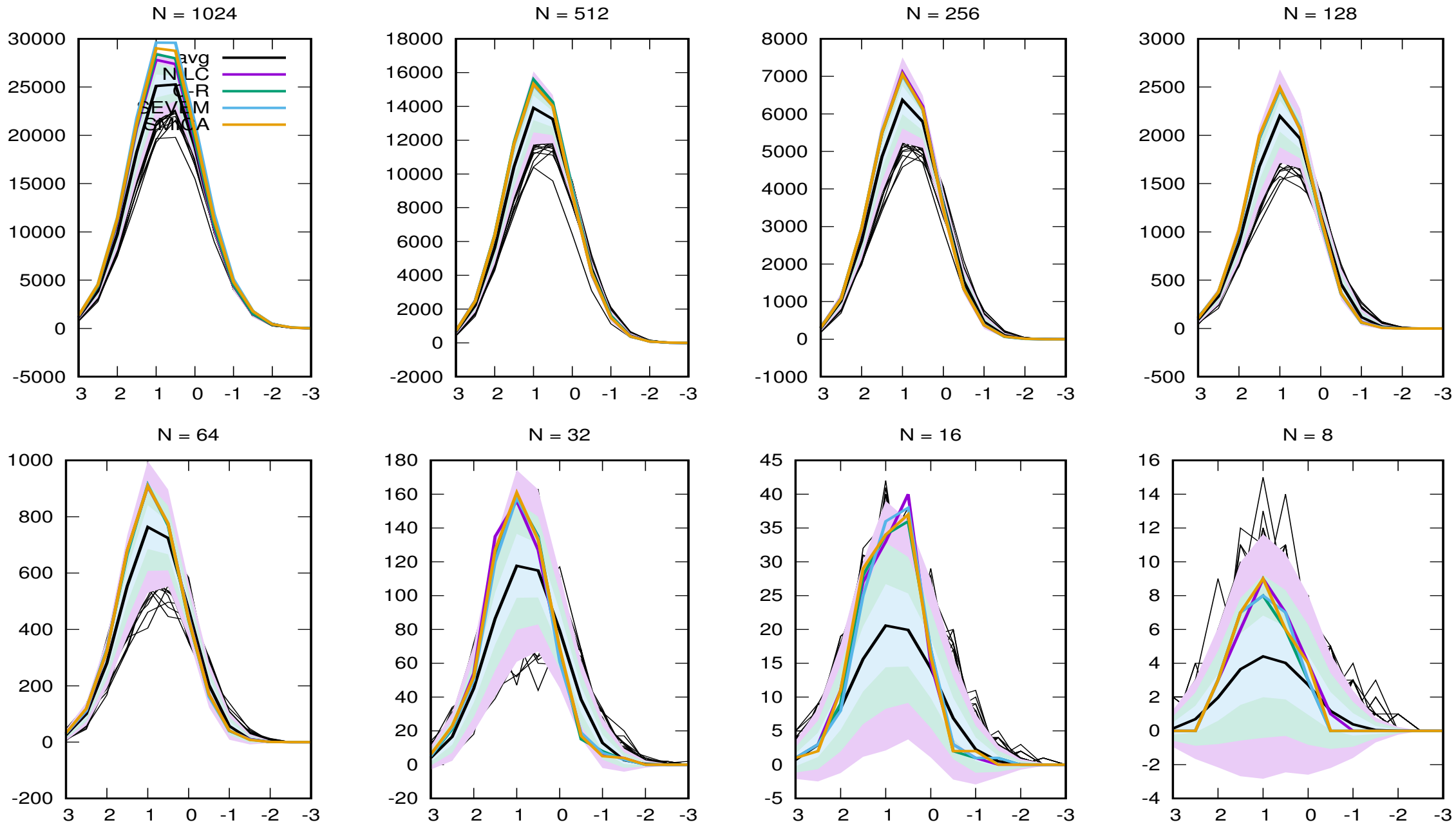






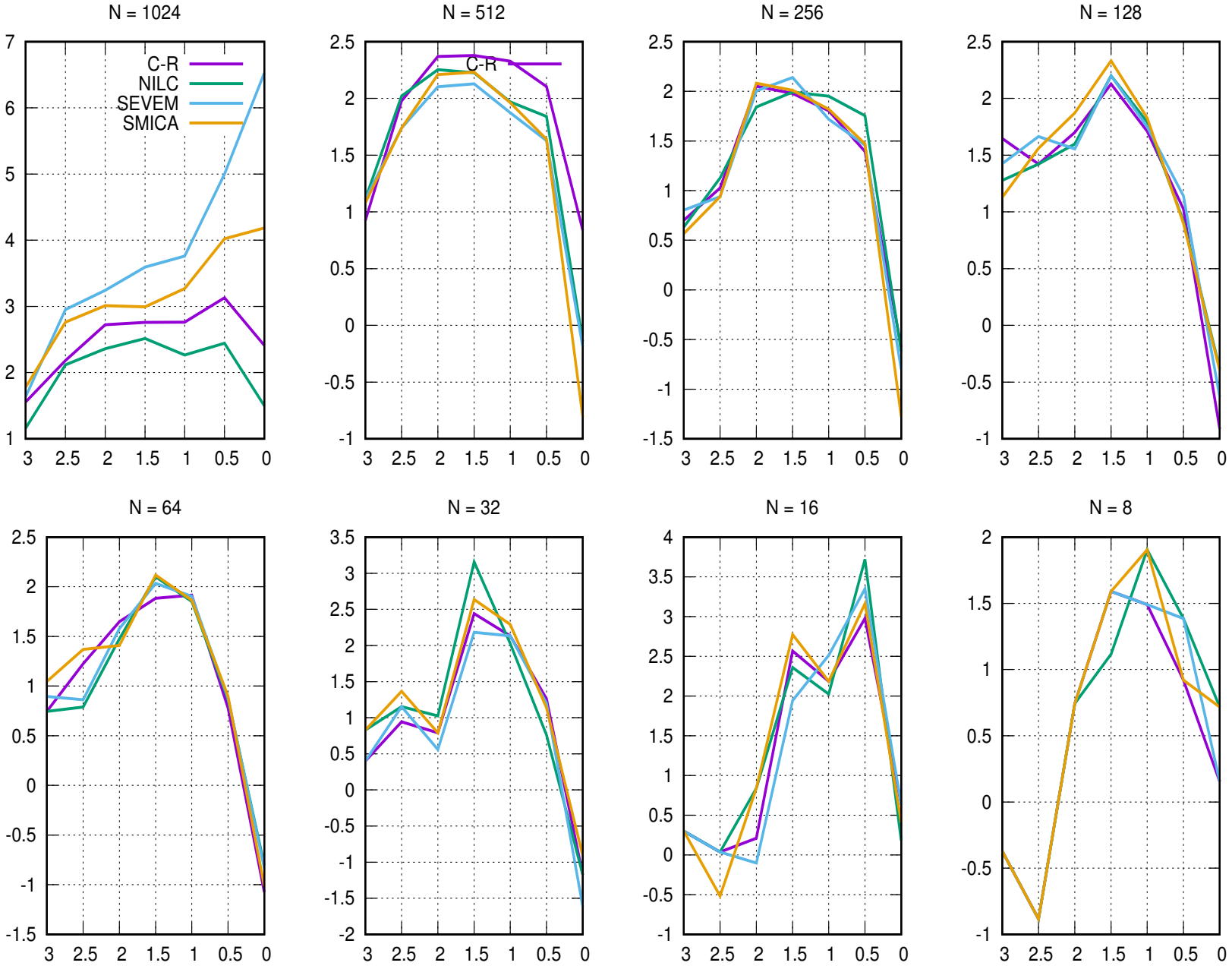


- Maps degraded to N_side = 1024, 512, 256, 128, 64, 32, 16, and 8 (not shown)
- Binary Mask degraded similarly (converts it to non-binary) – reconverted to binary by setting the threshold 0.9 (as done by Planck coll.) (additionally more thresholds: 0.7, 0.8 and 0.95)



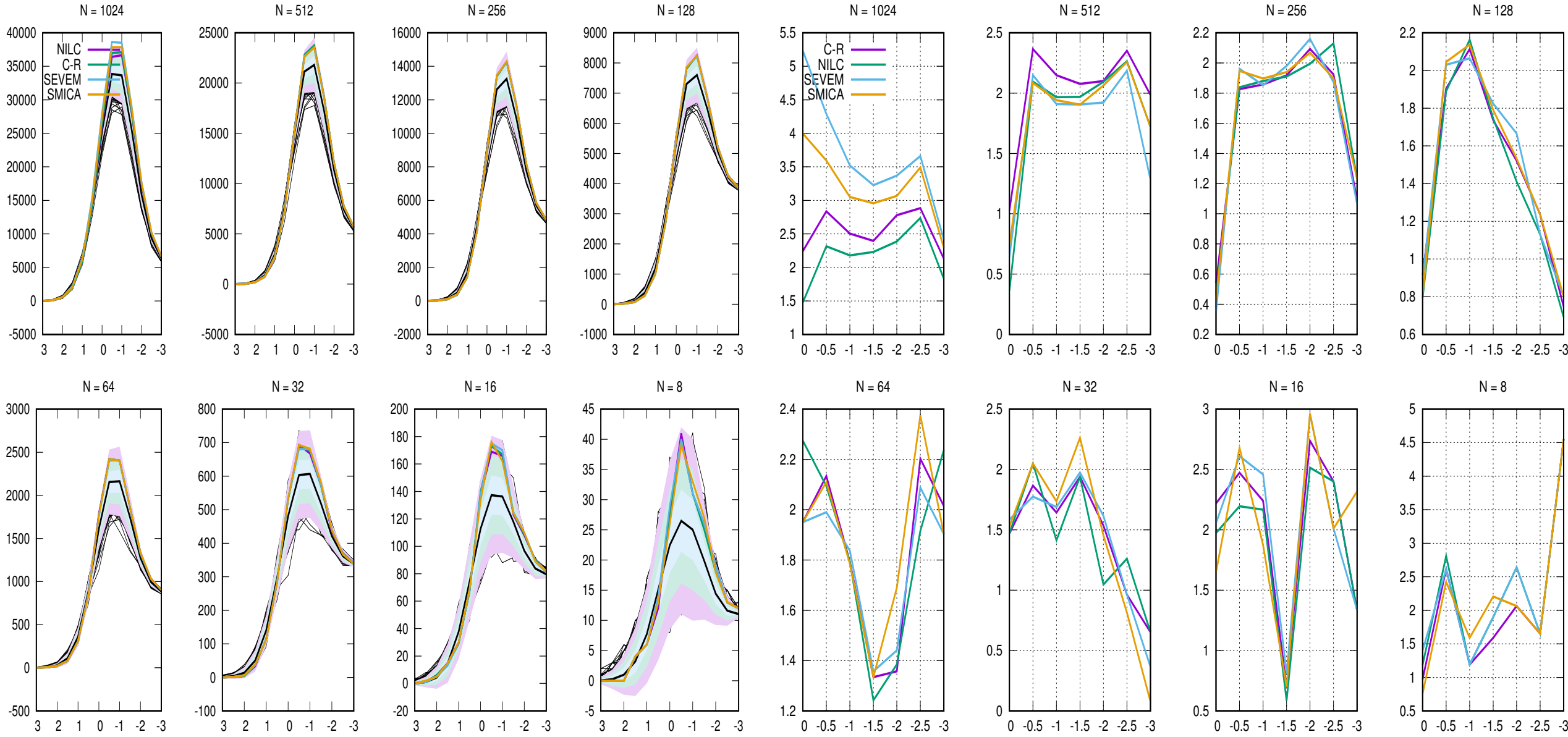
Isolated objects (betti 0)

Pranav et. al. A&A, accepted

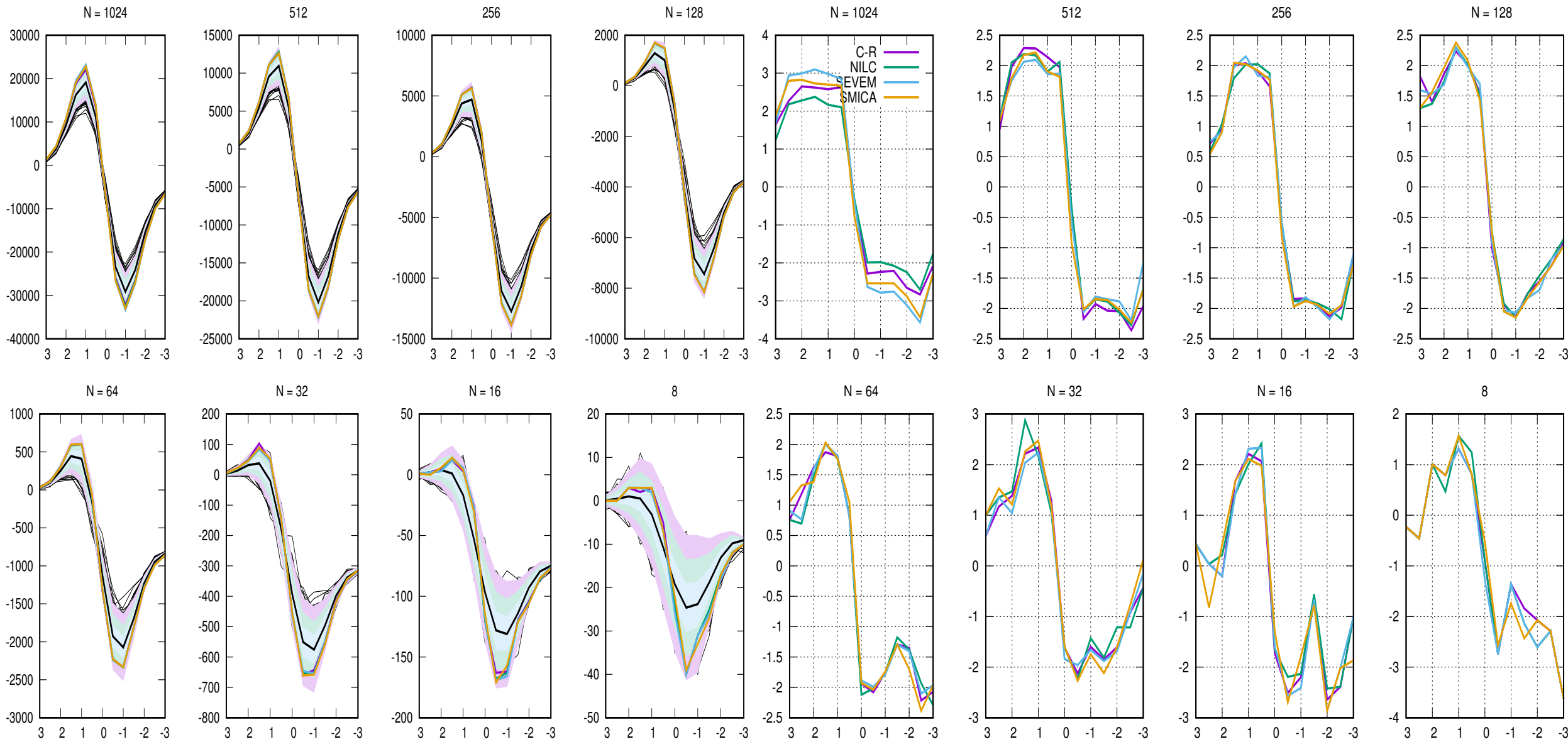


$$(f_{obs} - f_{avg,sims}) / sd_{sims}$$

Pranav et. al. A&A, accepted



Loops/holes (betti 1)



Euler Characteristic (betti0 – betti1)

Statistical tests

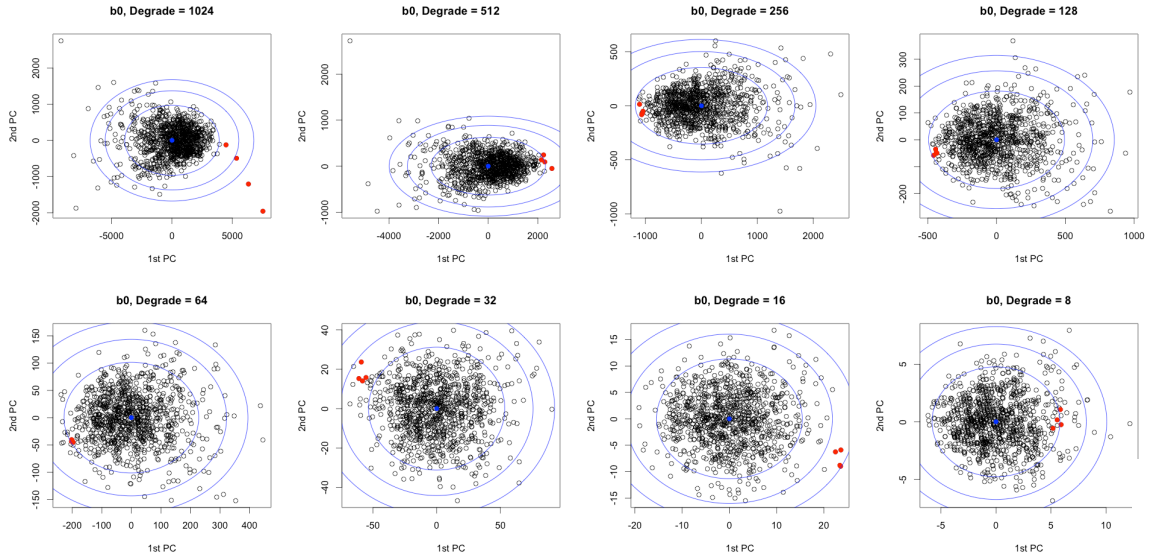
- The data consists of topological summaries (b0, b1, EC) obtained from 1000 simulations, and observed CMB field
- Goal: estimate the probability that the physical model that produced the simulations would produce quantities consistent with those from the observed CMB field
- Let $\mathbf{x}_i \in \mathbb{R}^m, i = 1, \ldots, n,$|, be a sample of i.i.d. m-dimensional vectors, drawn from a distribution F. Let $\mathbf{y} \in \mathbb{R}^m$ be another sample point, assumed to be drawn from a distribution G.
- Test the (null) hypothesis that F=G
- p-values compute the probability that **y** is `consistent' with this hypothesis.
- Two methods :
  - Mahalanobis Distance or chi^2 test : parametric (Prasanta Chandra Mahalanobis, 1936)
  - Tukey depth : non-parametric (John Tukey, 1974)

resolution	Method	Mahalanobis b_0	Mahalanobis b_1	Mahalanobis EC_{rel}	Tukey Depth b_0	Tukey Depth b_1	Tukey Depth EC_{rel}
threshold = 0.70							
1024	NILC	0.236	0.244	0.472	< **0.001**	< **0.001**	0.302
	C-R	0.048	0.170	0.130	< **0.001**	< **0.001**	< **0.001**
	SEVEM	< **0.001**	< **0.001**	0.124	< **0.001**	< **0.001**	< **0.001**
	SMICA	< **0.001**	**0.001**	0.208	< **0.001**	< **0.001**	< **0.001**
512	NILC	0.492	0.325	0.666	0.134	< **0.001**	0.685
	C-R	0.276	0.487	0.661	< **0.001**	< **0.001**	0.530
	SEVEM	0.660	0.303	0.751	0.389	< **0.001**	0.586
	SMICA	0.478	0.472	0.908	0.134	< **0.001**	0.870
256	NILC	0.602	0.513	0.760	0.201	0.211	0.579
	C-R	0.518	0.635	0.750	0.259	0.353	0.579
	SEVEM	0.390	0.490	0.496	0.136	0.211	0.313
	SMICA	0.480	0.571	0.695	0.136	0.211	0.313
128	NILC	0.260	0.441	0.627	< **0.001**	0.171	0.327
	C-R	0.331	0.547	0.705	0.152	0.171	0.451
	SEVEM	0.399	0.543	0.807	0.152	0.222	0.630
	SMICA	0.383	0.564	0.763	0.232	0.171	0.524
64	NILC	0.335	0.278	0.528	0.171	< **0.001**	< **0.001**
	C-R	0.319	0.366	0.528	0.237	< **0.001**	0.314
	SEVEM	0.211	0.352	0.488	< **0.001**	< **0.001**	< **0.001**
	SMICA	0.259	0.306	0.448	< **0.001**	< **0.001**	< **0.001**
32	NILC	0.082	0.302	0.442	< **0.001**	< **0.001**	< **0.001**
	C-R	0.166	0.292	0.509	< **0.001**	0.252	< **0.001**
	SEVEM	0.160	0.444	0.704	< **0.001**	0.252	0.351
	SMICA	0.155	0.294	0.472	< **0.001**	< **0.001**	0.351
16	NILC	**0.018**	**0.030**	0.120	< **0.001**	< **0.001**	< **0.001**
	C-R	**0.032**	**0.016**	0.102	< **0.001**	< **0.001**	< **0.001**
	SEVEM	**0.037**	**0.016**	0.178	< **0.001**	< **0.001**	< **0.001**
	SMICA	**0.017**	**0.001**	**0.021**	< **0.001**	< **0.001**	< **0.001**
8	NILC	0.373	< **0.001**	**0.012**	0.430	< **0.001**	< **0.001**
	C-R	0.706	< **0.001**	**0.022**	0.693	0.108	< **0.001**
	SEVEM	0.546	< **0.001**	**0.009**	0.563	< **0.001**	< **0.001**
	SMICA	0.401	< **0.001**	**0.004**	0.380	< **0.001**	< **0.001**
summary	NILC	**0.002**	**0.001**	**0.002**	< **0.001**	< **0.001**	< **0.001**
	C-R	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**
	SEVEM	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**
	SMICA	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**

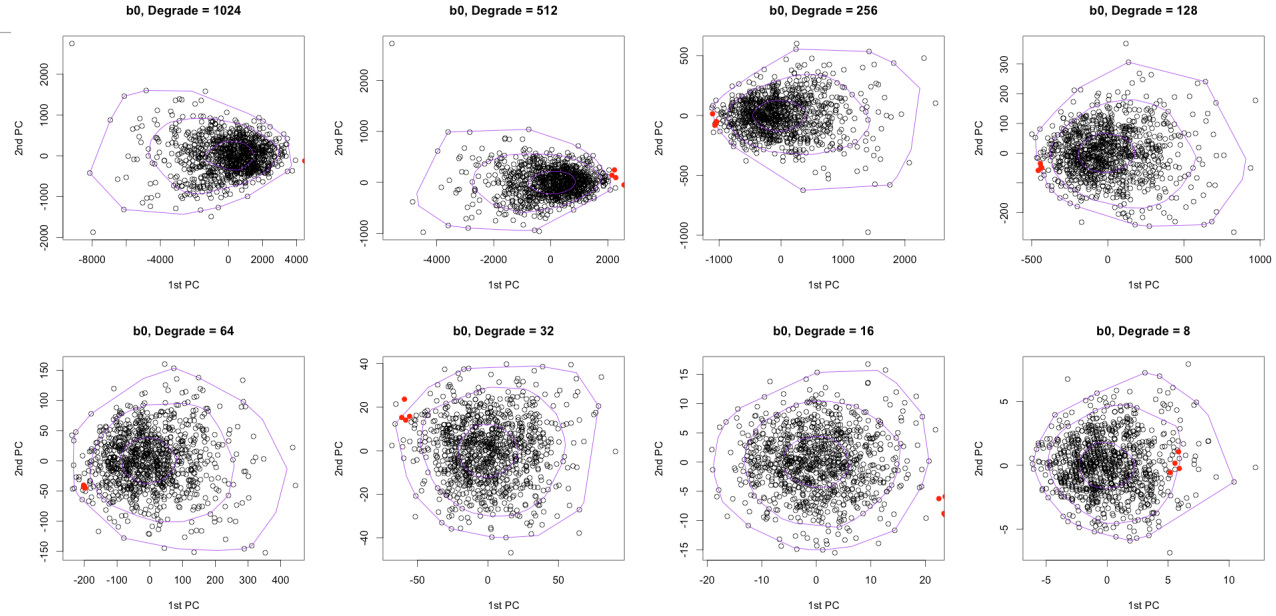
resolution	Method	Mahalanobis b_0	Mahalanobis b_1	Mahalanobis EC_{rel}	Tukey Depth b_0	Tukey Depth b_1	Tukey Depth EC_{rel}
threshold = 0.80							
1024	NILC	0.225	0.278	0.472	< **0.001**	< **0.001**	0.410
	C-R	**0.048**	0.169	0.130	< **0.001**	< **0.001**	< **0.001**
	SEVEM	< **0.001**	< **0.001**	0.095	< **0.001**	< **0.001**	< **0.001**
	SMICA	< **0.001**	0.002	0.217	< **0.001**	< **0.001**	< **0.001**
512	NILC	0.499	0.348	0.686	0.130	< **0.001**	0.649
	C-R	0.289	0.491	0.690	< **0.001**	< **0.001**	0.537
	SEVEM	0.657	0.362	0.813	0.268	< **0.001**	0.649
	SMICA	0.463	0.522	0.919	0.130	< **0.001**	0.784
256	NILC	0.559	0.481	0.679	0.139	0.218	0.538
	C-R	0.541	0.631	0.752	0.139	0.380	0.604
	SEVEM	0.377	0.512	0.503	0.139	0.149	0.334
	SMICA	0.459	0.562	0.723	0.139	0.218	0.604
128	NILC	0.295	0.484	0.633	< **0.001**	< **0.001**	0.331
	C-R	0.363	0.609	0.695	< **0.001**	0.149	0.434
	SEVEM	0.318	0.640	0.755	< **0.001**	0.149	0.517
	SMICA	0.370	0.637	0.735	< **0.001**	< **0.001**	0.517
64	NILC	0.250	0.269	0.382	< **0.001**	< **0.001**	< **0.001**
	C-R	0.192	0.363	0.438	< **0.001**	< **0.001**	0.311
	SEVEM	0.166	0.336	0.408	< **0.001**	< **0.001**	0.311
	SMICA	0.172	0.339	0.351	< **0.001**	< **0.001**	0.000
32	NILC	0.082	0.406	0.538	< **0.001**	< **0.001**	< **0.001**
	C-R	0.149	0.452	0.707	< **0.001**	0.345	0.652
	SEVEM	0.175	0.515	0.774	< **0.001**	0.292	0.810
	SMICA	0.133	0.384	0.578	< **0.001**	0.292	0.607
16	NILC	**0.024**	**0.043**	0.082	< **0.001**	< **0.001**	< **0.001**
	C-R	**0.028**	**0.042**	0.119	< **0.001**	< **0.001**	< **0.001**
	SEVEM	0.064	**0.024**	0.119	< **0.001**	< **0.001**	< **0.001**
	SMICA	**0.039**	**0.007**	**0.046**	< **0.001**	< **0.001**	< **0.001**
8	NILC	0.202	< **0.001**	**0.013**	0.142	< **0.001**	0.220
	C-R	0.573	< **0.001**	**0.013**	0.599	< **0.001**	< **0.001**
	SEVEM	0.352	< **0.001**	**0.012**	0.358	< **0.001**	< **0.001**
	SMICA	0.331	< **0.001**	**0.012**	0.323	< **0.001**	< **0.001**
summary	NILC	**0.001**	**0.001**	**0.002**	< **0.001**	**0.032**	< **0.001**
	C-R	**0.001**	**0.001**	**0.001**	**0.001**	**0.032**	**0.001**
	SEVEM	**0.001**	**0.001**	**0.001**	**0.001**	**0.032**	**0.001**
	SMICA	**0.001**	**0.001**	**0.001**	**0.001**	**0.032**	**0.001**

Pranav et. al. A&A, accepted

PC graphs

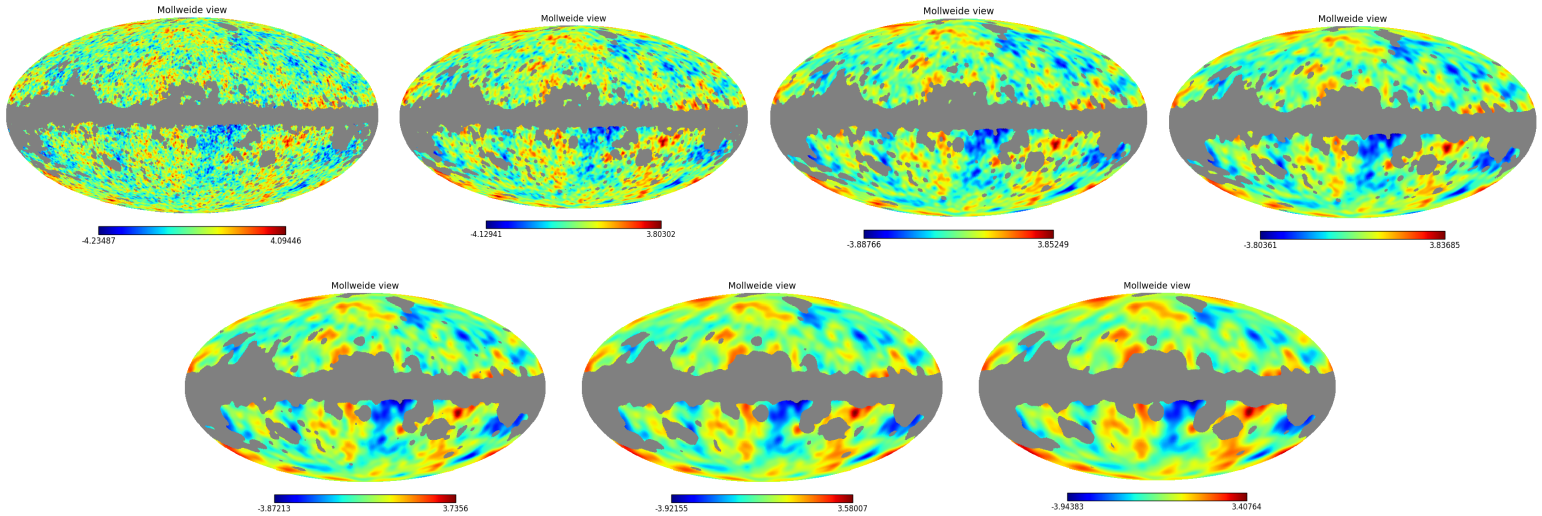


Mahalanobis/ chi^2



Tukey depth

More tests: smoothed maps



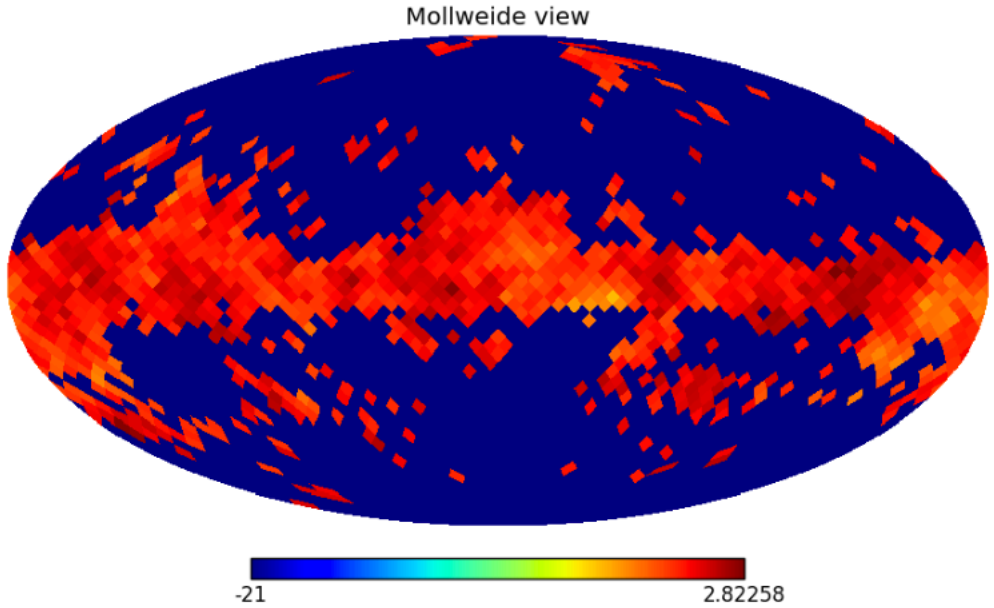
fwhm	Method	β_0	β_1	χ (E.C.)
1	NILC	0.2928	**0.0346**	0.1769
	C-R	0.2168	**0.0374**	0.3003
	SEVEM	0.1964	0.1772	0.4406
	SMICA	0.1370	0.0869	0.2130
2	NILC	0.0923	**0.0097**	0.0856
	C-R	0.1206	**0.0265**	0.1396
	SEVEM	**0.0196**	**0.0258**	**0.0241**
	SMICA	0.0803	**0.0261**	0.0992
3	NILC	0.1511	**0.0111**	0.0567
	C-R	0.1100	**0.0018**	**0.0145**
	SEVEM	0.2052	**0.0081**	0.0798
	SMICA	0.2704	**0.0106**	0.0776
4	NILC	0.1373	0.0961	0.2968
	C-R	0.0865	0.0839	0.2243
	SEVEM	**0.0304**	**0.0144**	**0.0499**
	SMICA	**0.0289**	**0.0016**	**0.0195**
5	NILC	**0.0353**	0.1261	0.1999
	C-R	0.0561	0.1297	0.1018
	SEVEM	0.1695	0.2120	0.5552
	SMICA	0.0515	**0.0084**	0.0956
6	NILC	**0.0384**	**0.0478**	0.1839
	C-R	**0.0182**	0.0561	0.1123
	SEVEM	0.1968	**0.0091**	0.2069
	SMICA	0.0793	**0.0407**	0.2600

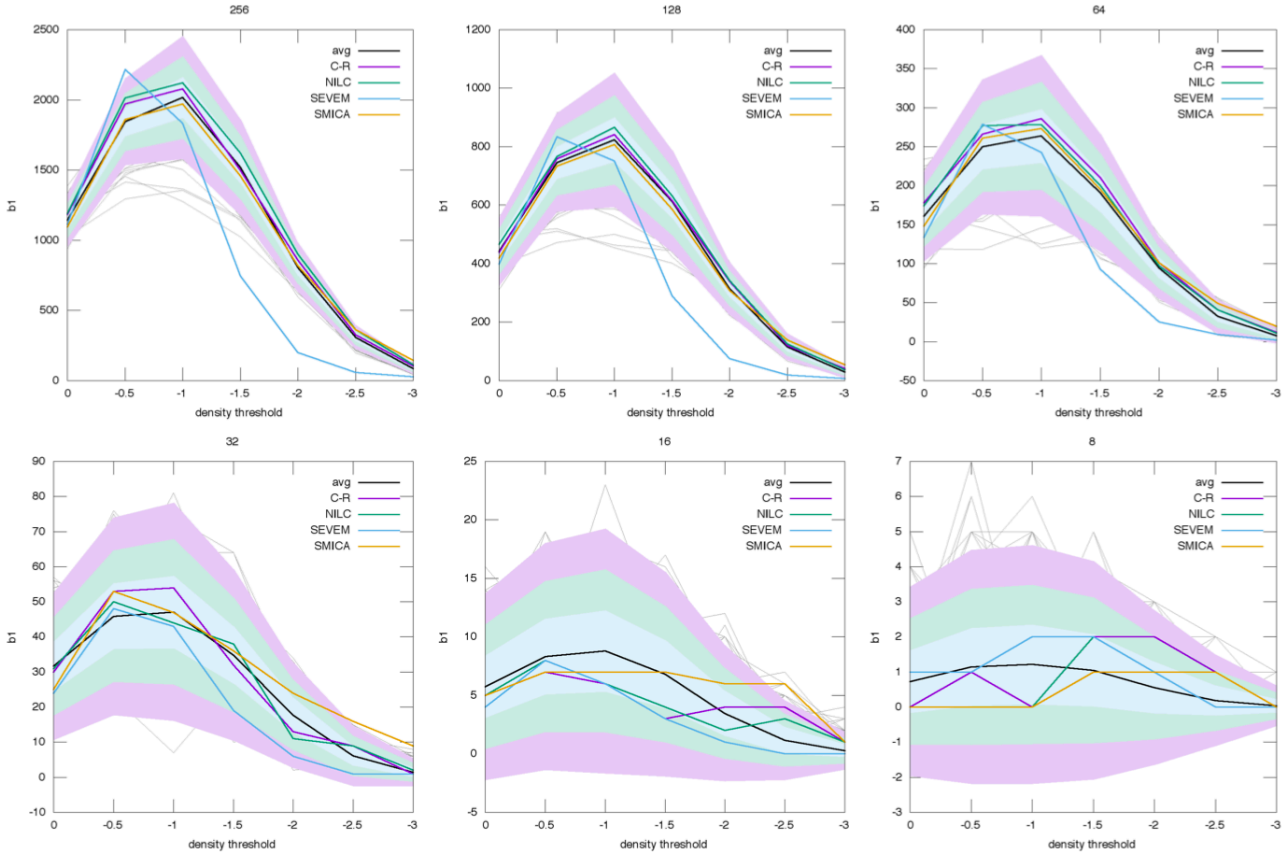
Table 1: p-value table for the significance of components, holes and the Euler characteristic. The values are shown for various resolutions ranging from 1 to 8 (fwhm in degrees), for the different component-separated maps. We set our level of significance at $p = 0.05$.

- Maps smoothed to fwhm = 1, 2, 3, 4, 5, and 6 degrees
- Binary Mask degraded similarly (converts it to non binary) – reconverted to binary by setting the threshold 0.9 (as done by planck coll.)

Pranav et. al. A&A, accepted

More tests: inverted mask



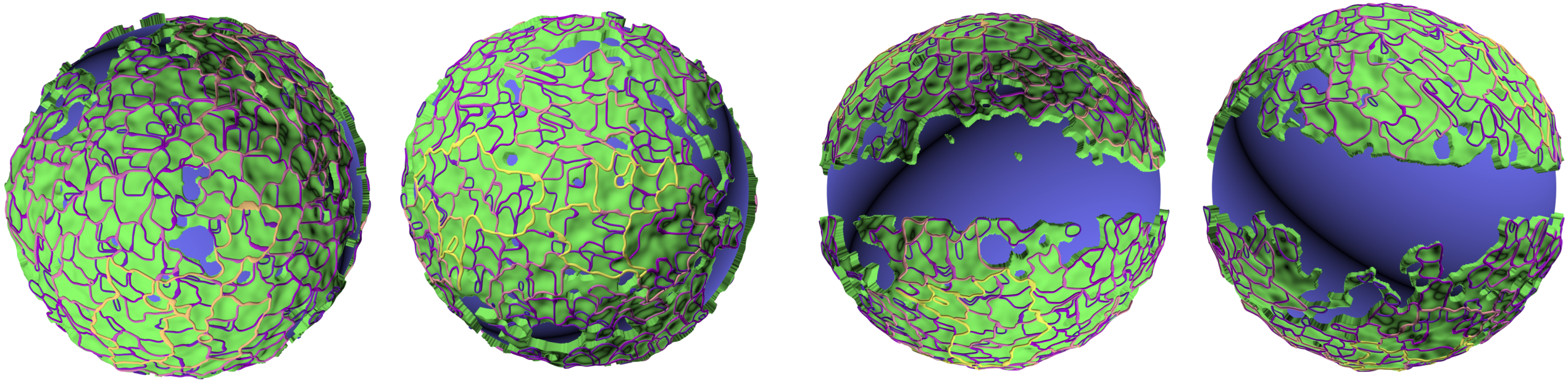


- Mask inverted and zone of reconstruction analyzed

- No significant difference between observations and simulations

Pranav et. al, in prep

CMB Loops



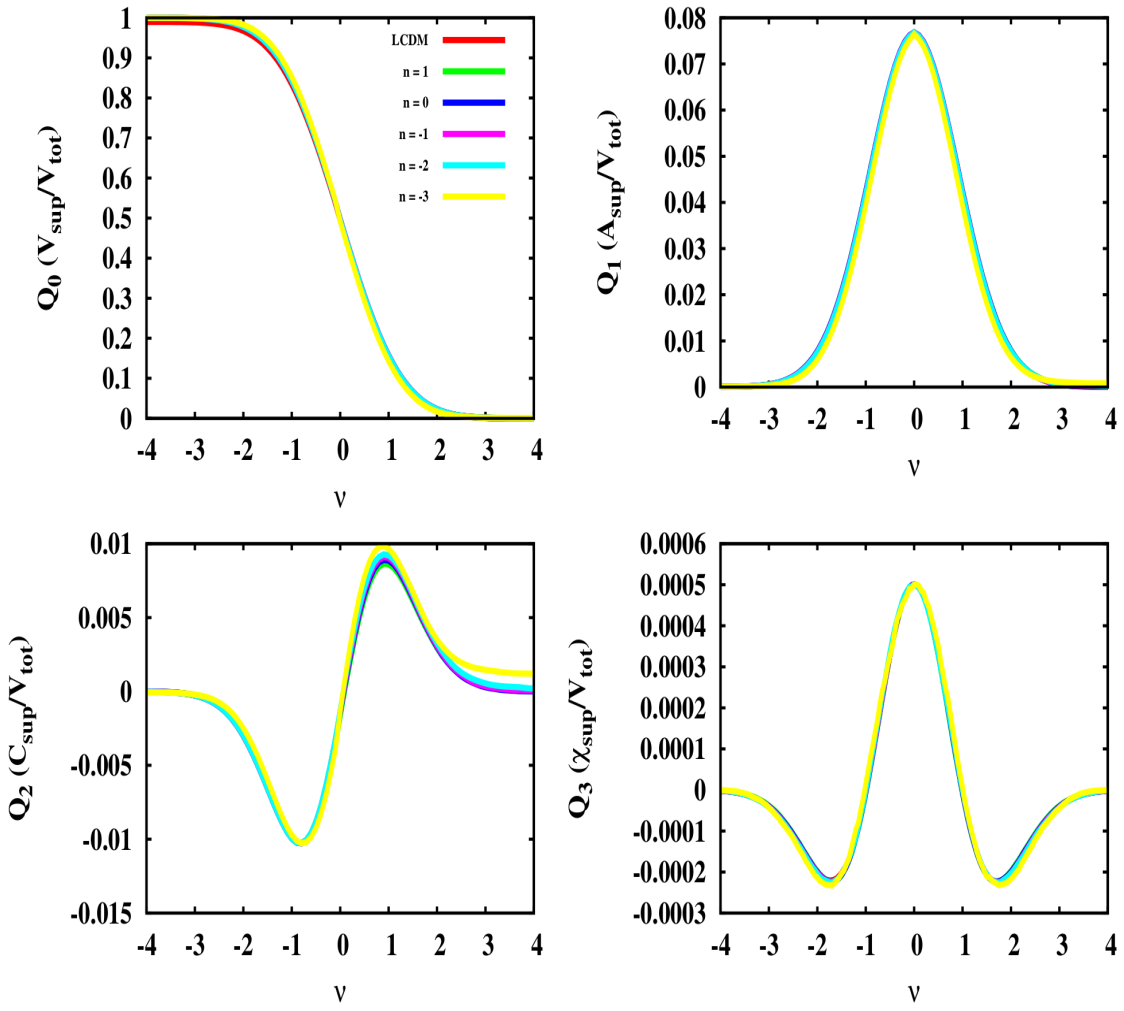
Gaussian fields: Euler and Minkowski

$$Q_0(\upsilon) = \frac{1}{2} - \frac{1}{\sqrt{2}}\Phi(\upsilon)$$

$$Q_1(\upsilon) = \frac{2}{3}\frac{\lambda}{\sqrt{2\pi}}\exp(-\frac{1}{2}\upsilon^2)$$

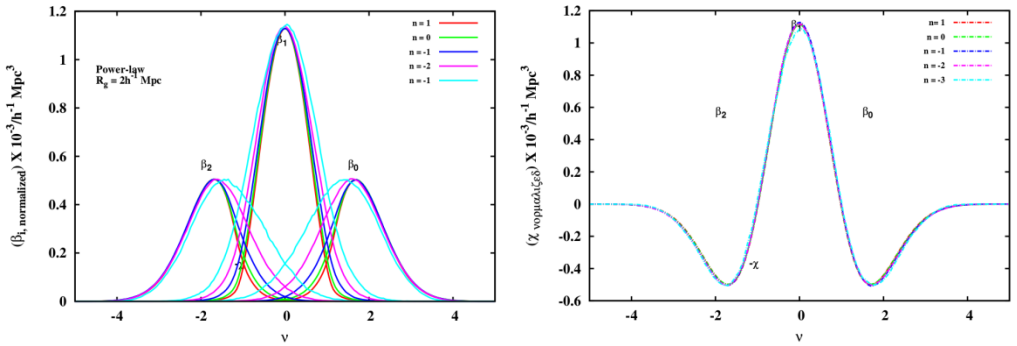
$$Q_2(\upsilon) = \frac{2}{3}\frac{\lambda^2}{\sqrt{2\pi}}\upsilon\exp(-\frac{1}{2}\upsilon^2)$$

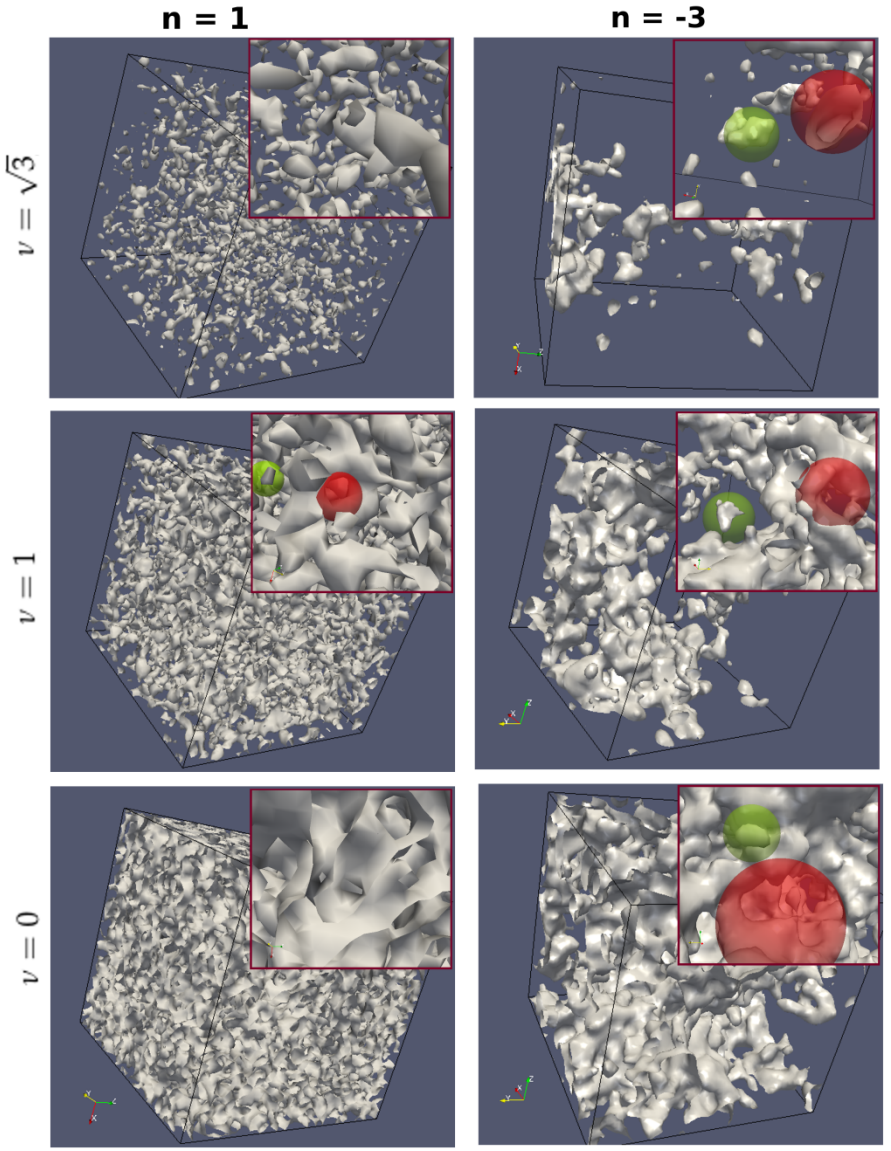
$$Q_3(\upsilon) = \frac{\lambda^3}{\sqrt{2\pi}}(\upsilon^2 - 1)\exp(-\frac{1}{2}\upsilon^2)$$



Pranav et. al, in prep

Gaussian fields: Betti Topology





- Shape of Betti numbers dependent on power spectrum; EC and MF are not
- Gott et. Al. (1986) claim the topology of LSS at median density to be *Sponge-like* based on EC
- Betti numbers reveal it is not so even for the simplest case of GRF

Pranav et. al, MNRAS 485 (3), 4167 - 4208

Conclusions

- Topology and geometry ideal tool for studying connectivity and the nature of complex spatial patterns manifested in the universe
- The topology of the CMB temperature fluctuations deviant from realistic simulations based on Gaussian prescriptions
- Earlier measurement of CMB by WMAP also shows mildly significant Euler characteristic (Eriksen 04).
- Not due to cold spot or other directional anomalies, as the loops and isolated components cover all sky.
- Betti numbers of Gaussian fields show a dependence on power spectrum, unlike MFs.

Homology & Persistence
Voronoi Models

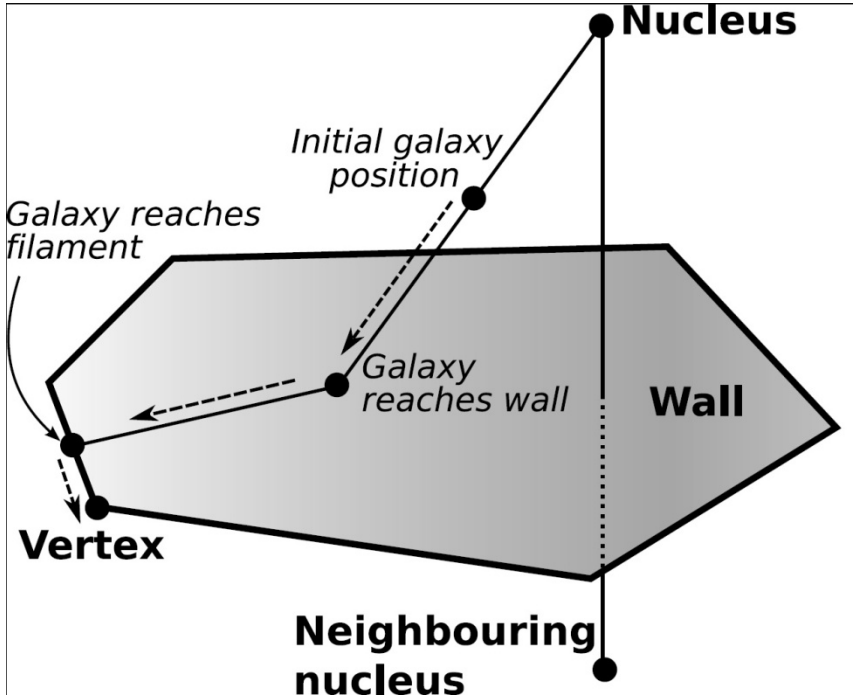
Voronoi Clustering Models

Template/Skeleton:

- Galaxy distributions in weblike networks
- Testing statistical/clustering characteristics of weblike galaxy distributions

Advantages:

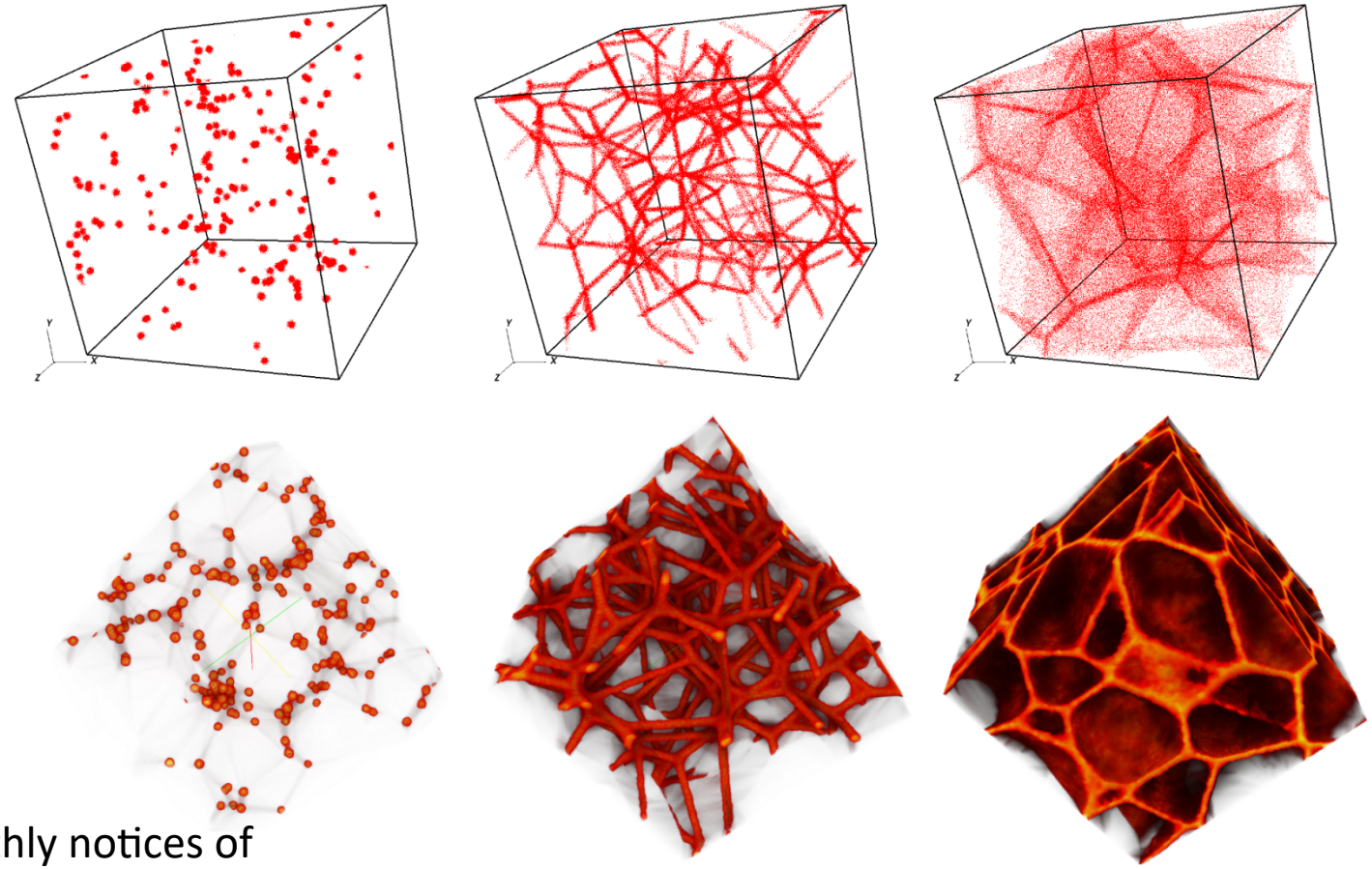
- Flexible
  low computational cost
- Versatile
  exploring widely different cosmological models



Web Elements:

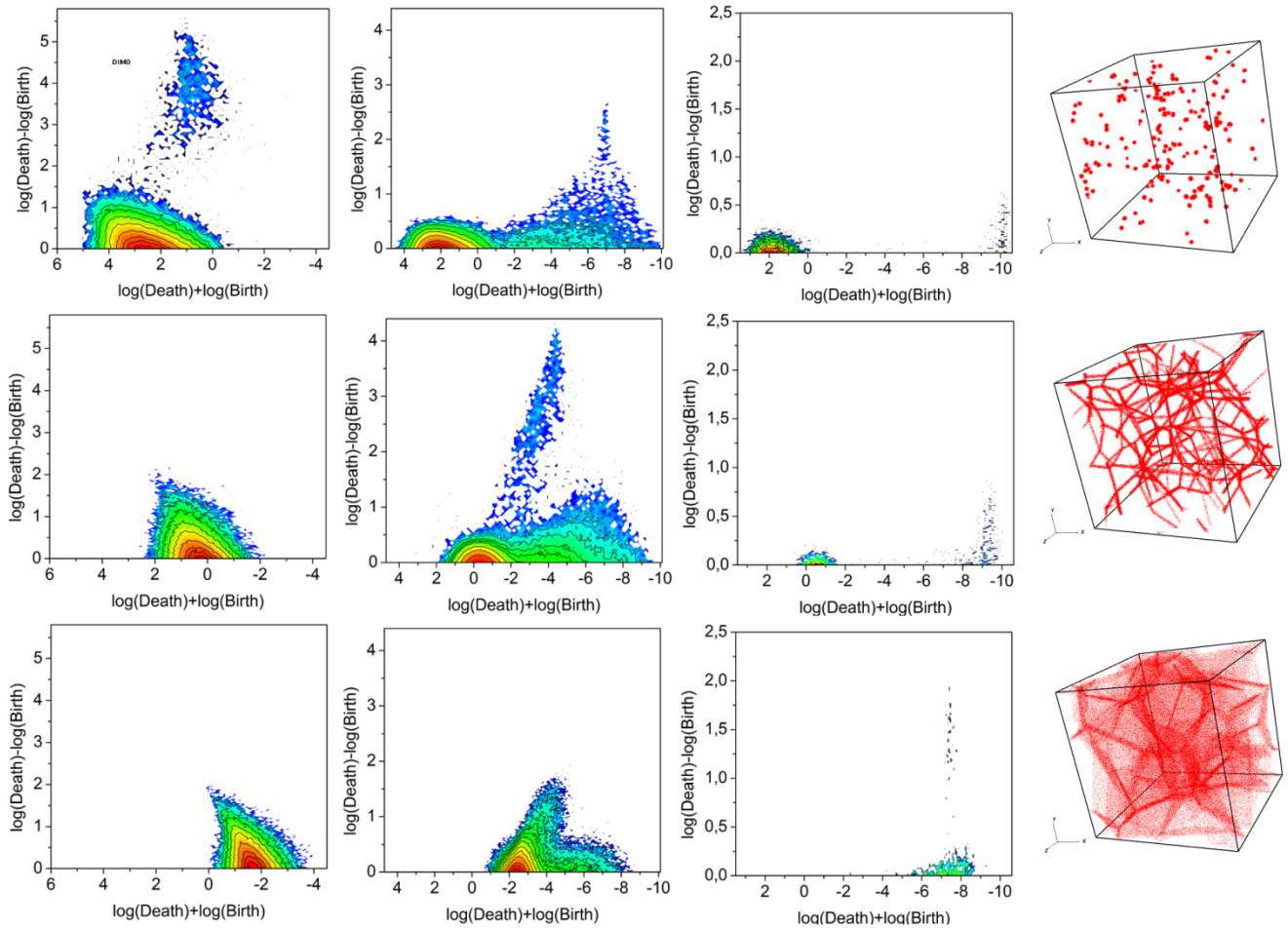
Wall:	projection onto Voronoi wall
Filament:	projection onto Voronoi edge
Cluster:	projection towards Voronoi vertex

Single Component Models



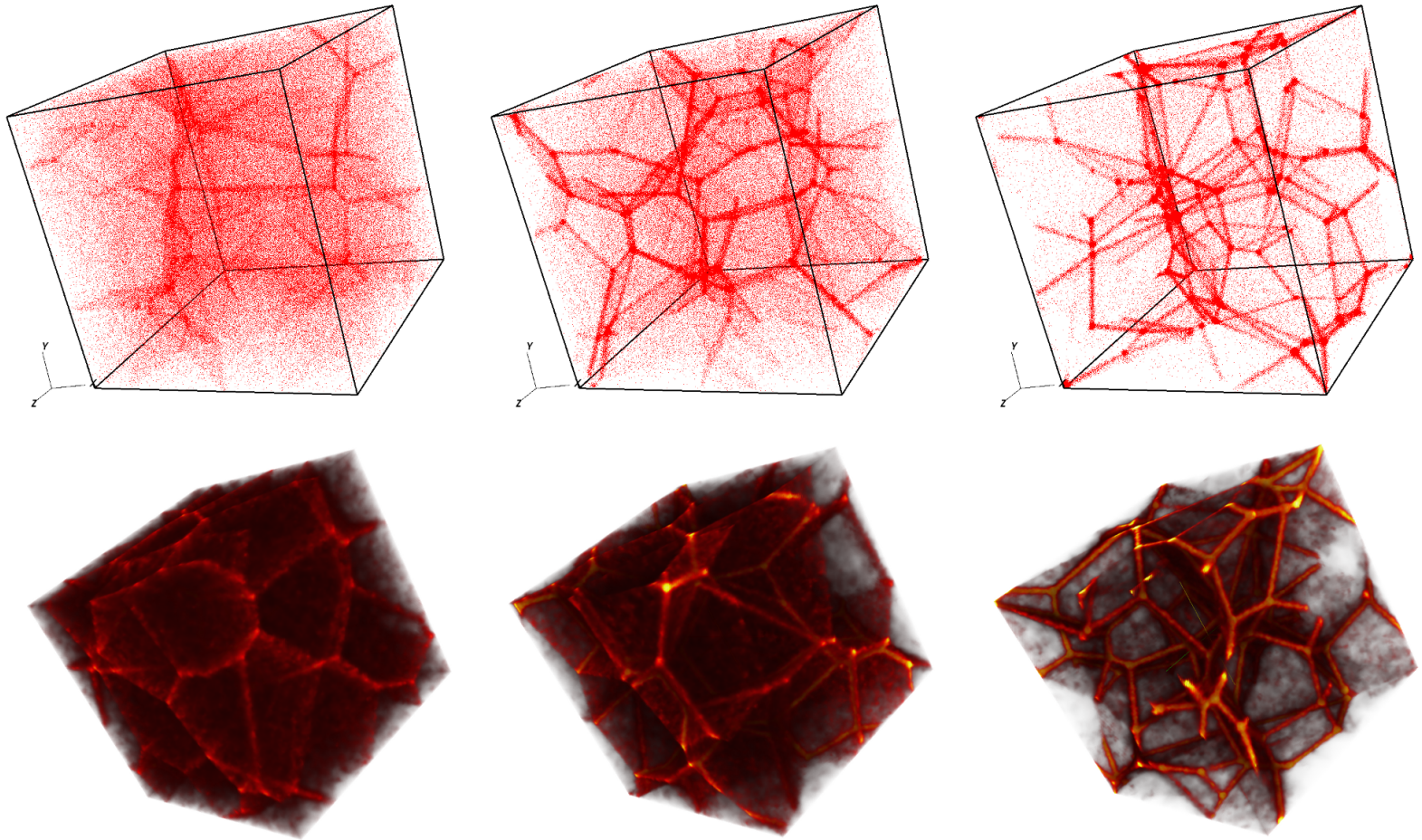
Pranav et. al 2016, Monthly notices of Royal astronomical Society, 465

Persistence : Single component models



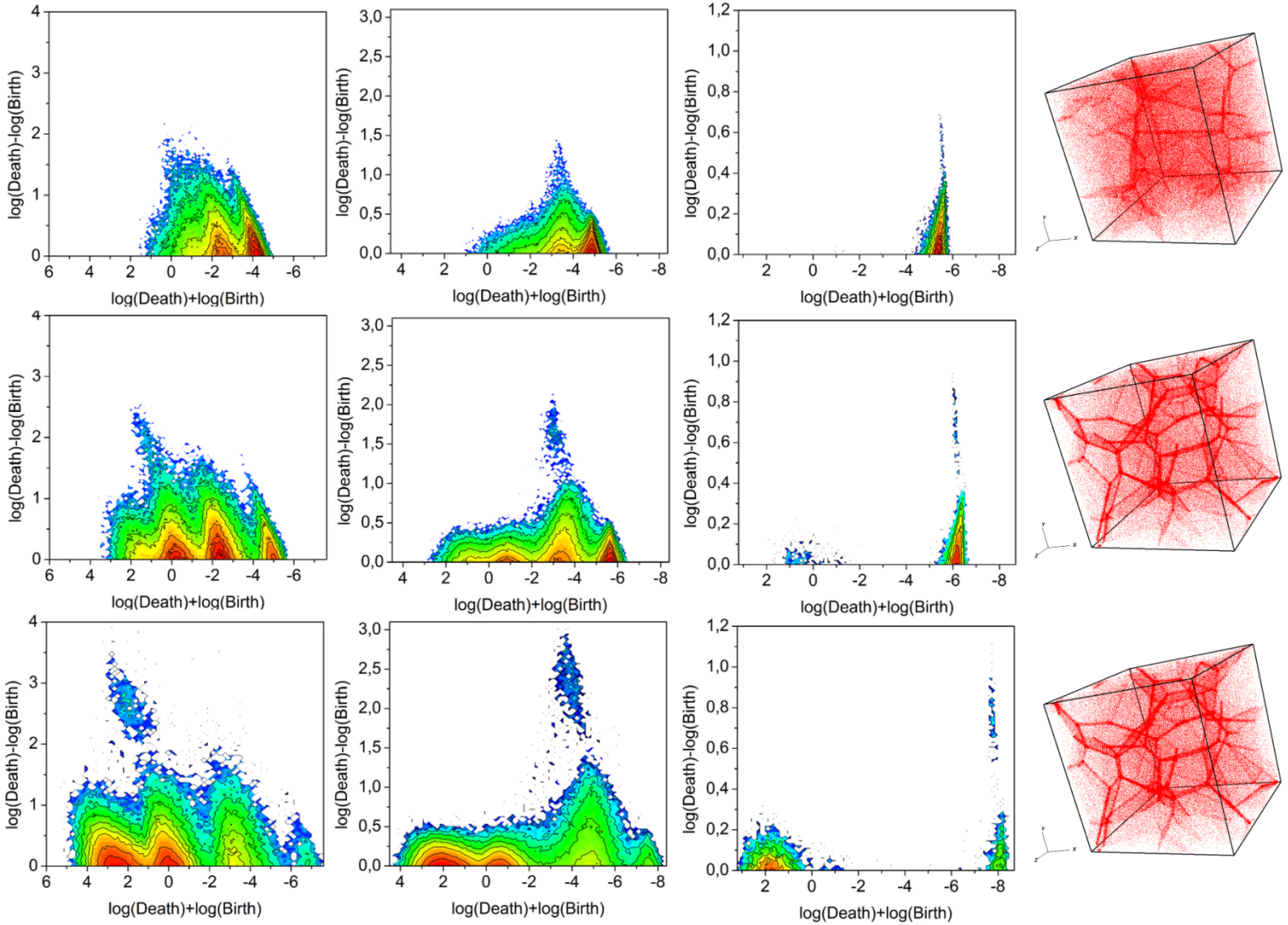
Pranav et. al 2016, Monthly notices of Royal astronomical Society, 465

Evolution Models



Pranav et. al 2016, Monthly notices of Royal astronomical Society, 465

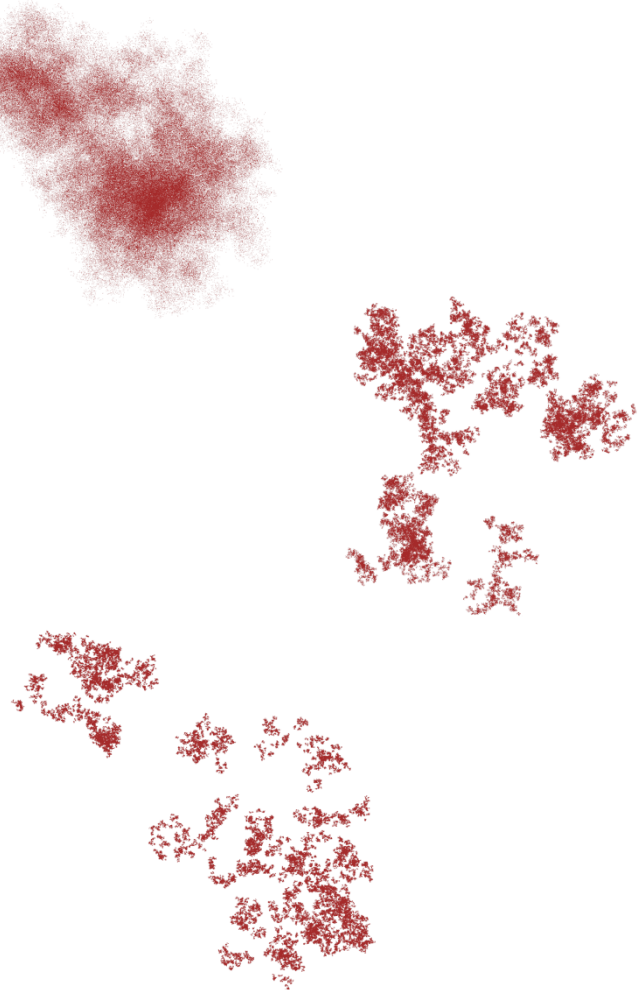
Persistence : Evolution Models



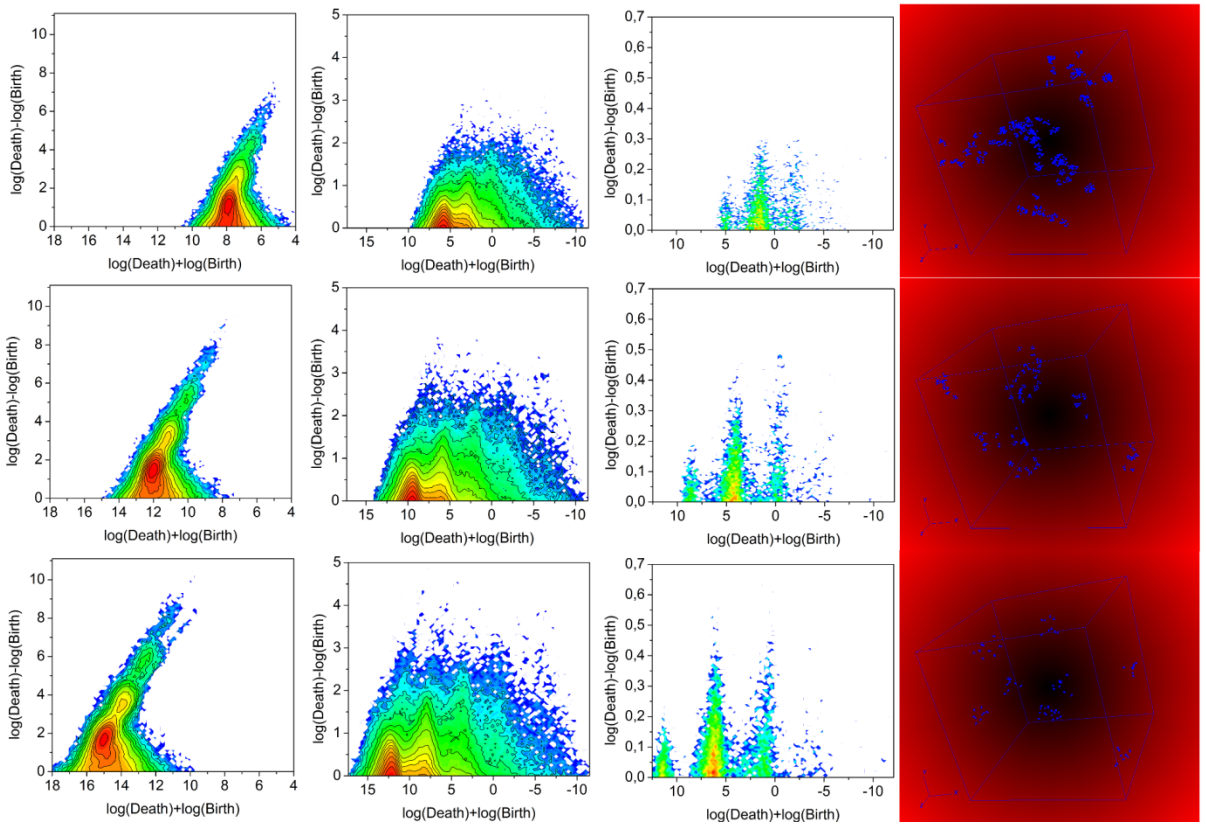
Pranav et. al 2016, Monthly notices of Royal astronomical Society, 465

Homology & Persistence
Hierarchical models

Soneira-Peebles Models:
Heuristic description of hierarchical clustering



Pranav et. al 2016, Monthly notices of Royal astronomical Society, 465



Parameters of the model :

- Number of levels (n)
- Number of children (η)
- Ratio between the radius of parent and children spheres(λ)

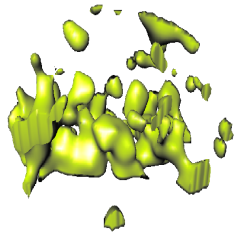
Randomly place η spheres inside the top-level and continue for all levels

And now with galactic densities

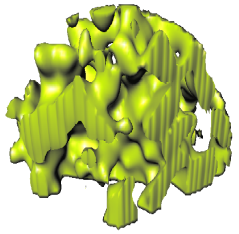
Center for Astrophysics (CfA) survey

SEPARATION OF HOMOLOGIES

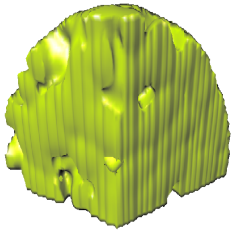
10,506 galaxies in the cone-shaped survey region, which extends out to 135 megaparsecs in the northern hemisphere, with the earth at the apex of the cone.



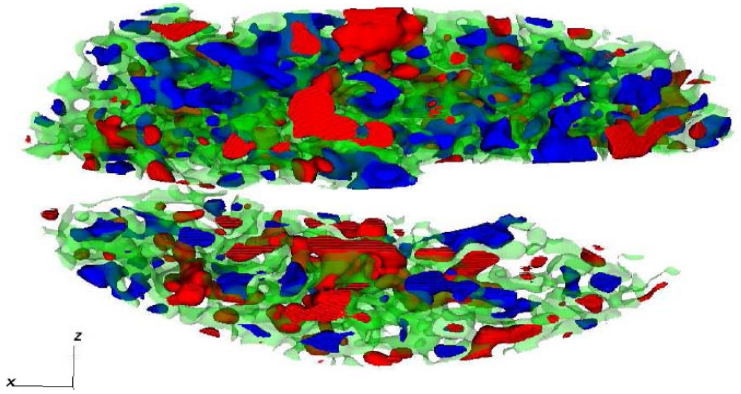
Meatball

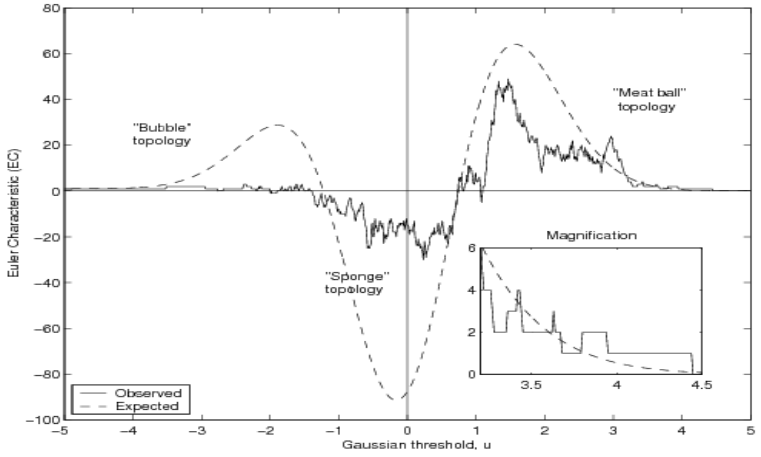


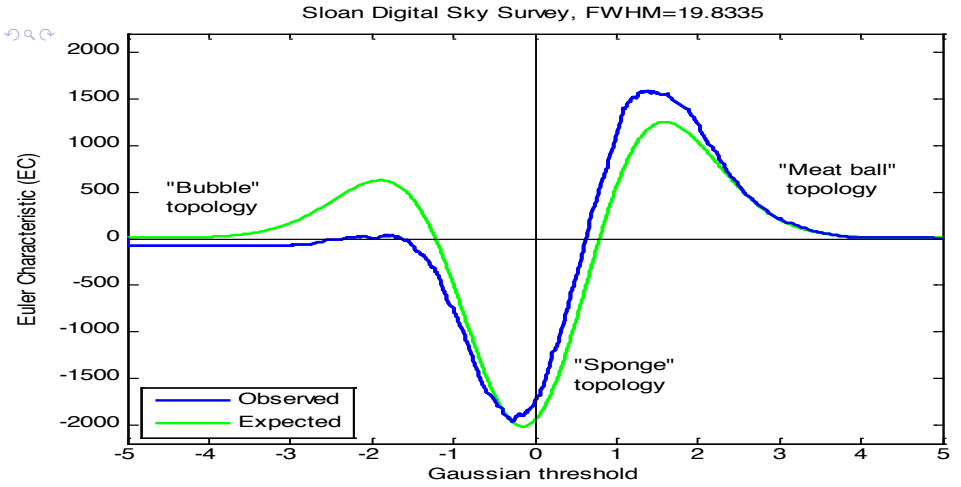
Sponge



Swiss cheese, bubble





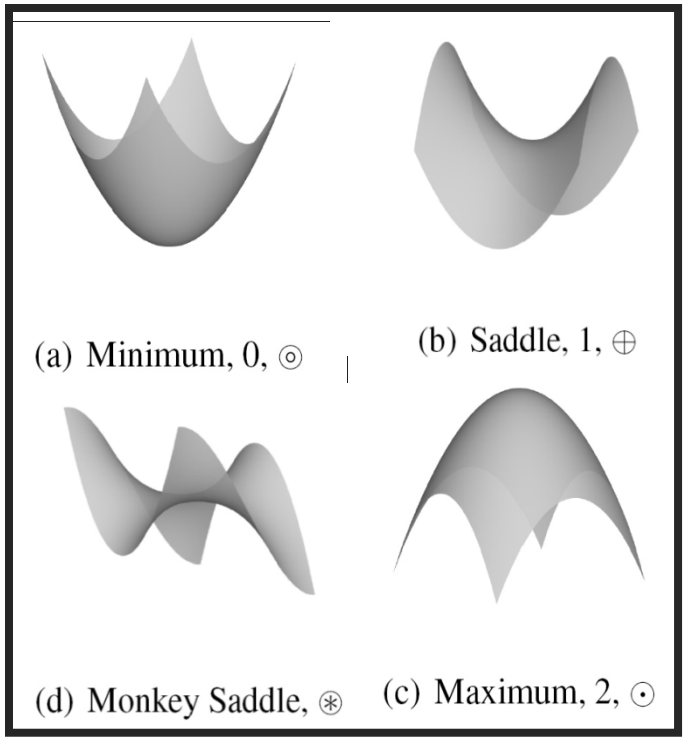


Topology and Geometry of 3D Gaussian fields

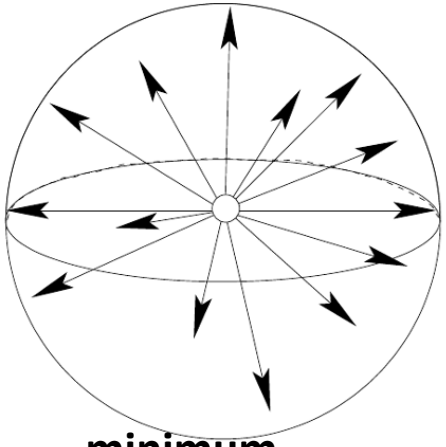
Felix:
*Fi*lament *Ex*plorer

Morse functions

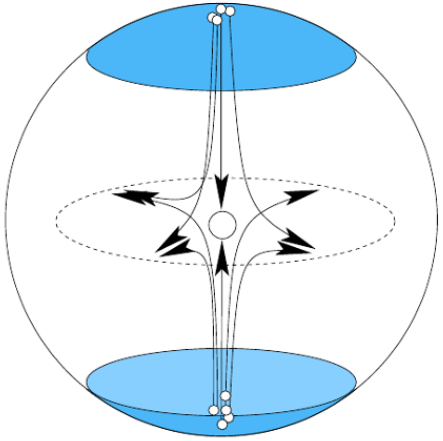
- Let f: $R^n \rightarrow R$
- The function is *nice*
  - » Domain R, R^2, R^3
  - » Gradient ∇f
  - » Critical points $\nabla f = 0$
  - » Smooth functions $\nabla^2 f$



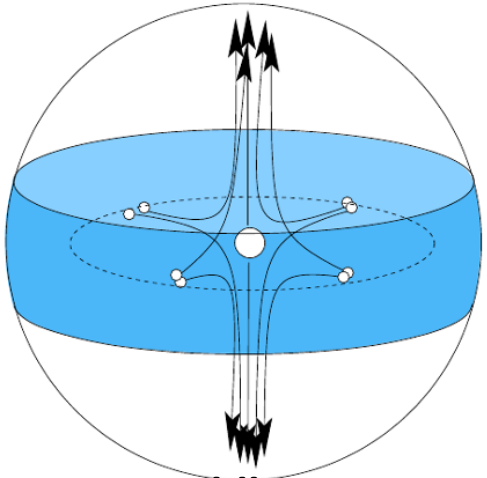
Morse geometry



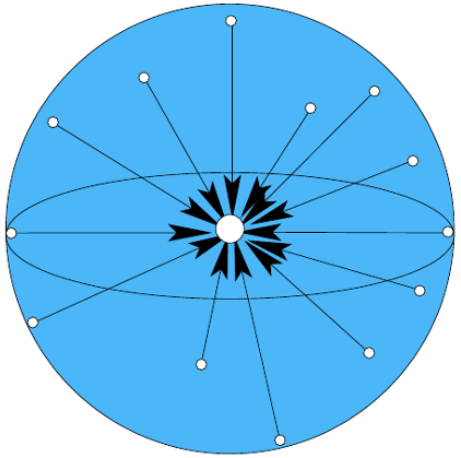
minimum



1-saddle

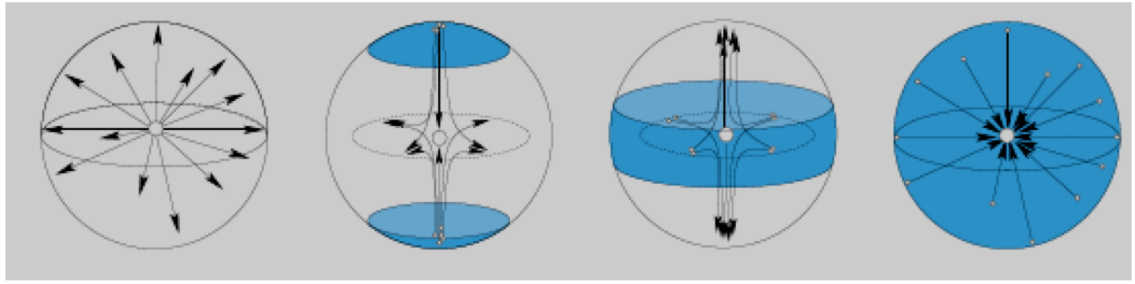


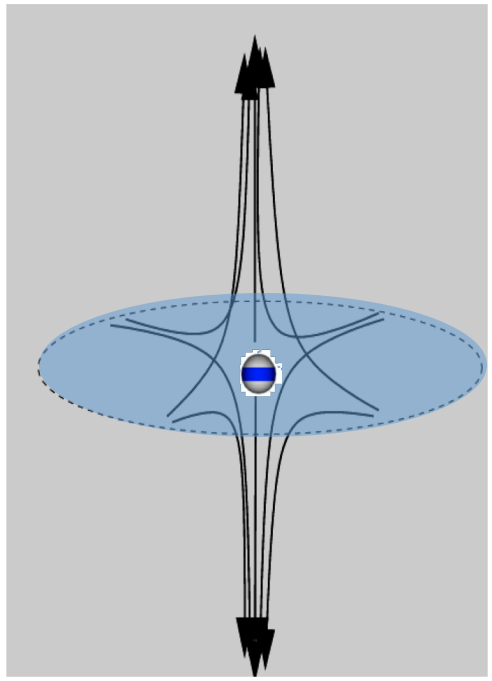
2-saddle



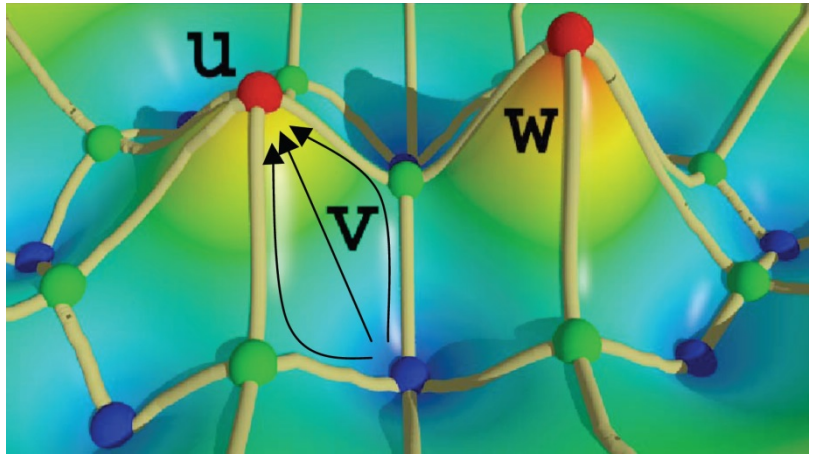
maximum

Morse geometry :
Filaments as ascending manifolds of 2-saddles





Integral lines : maximal curves in the domain of *f* that align with the gradient



Ascending manifold of a critical point *p :* set of all integral lines that originate at p, along with p

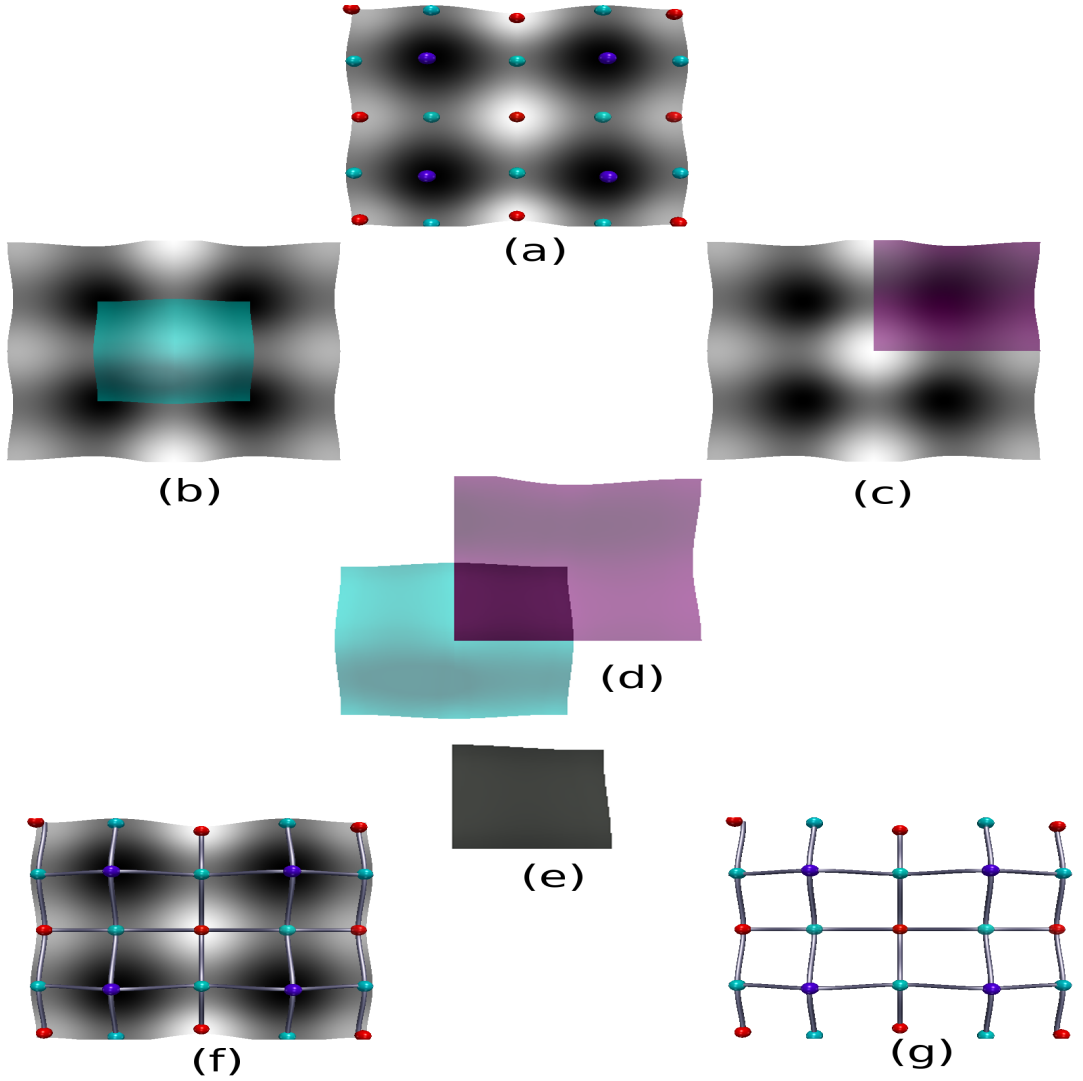
Descending manifold of a critical point *p :* set of all integral lines that terminate at p, along with p

Morse-Smale Complex

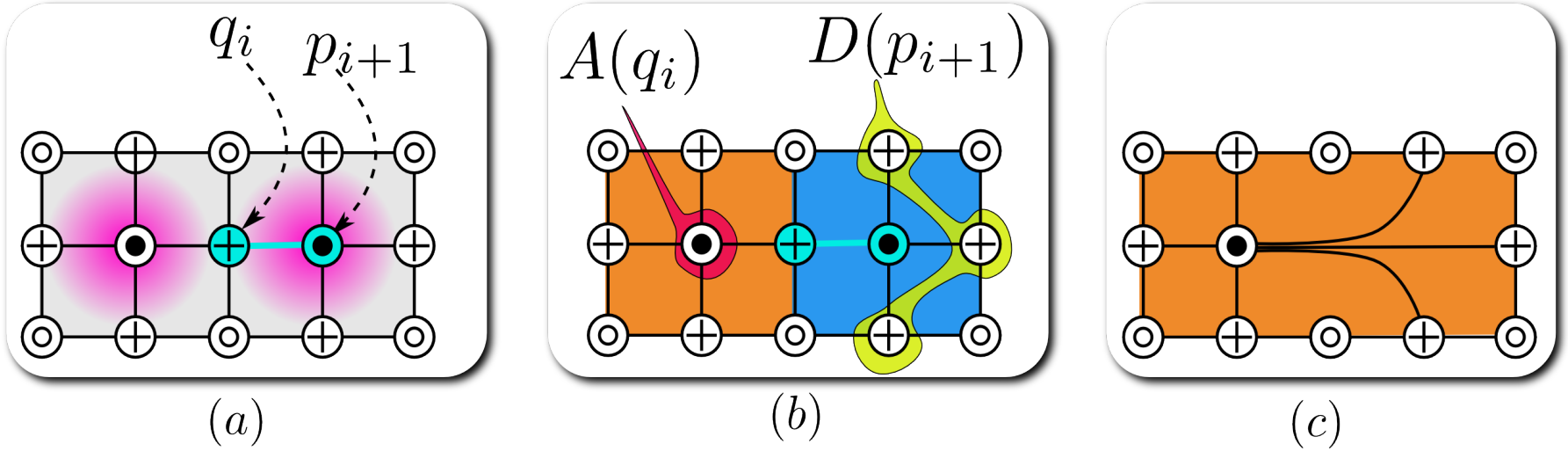
- **Morse-Smale complex : partition of the domain into cells formed by the collection of integral lines that share a common source and a common destination**

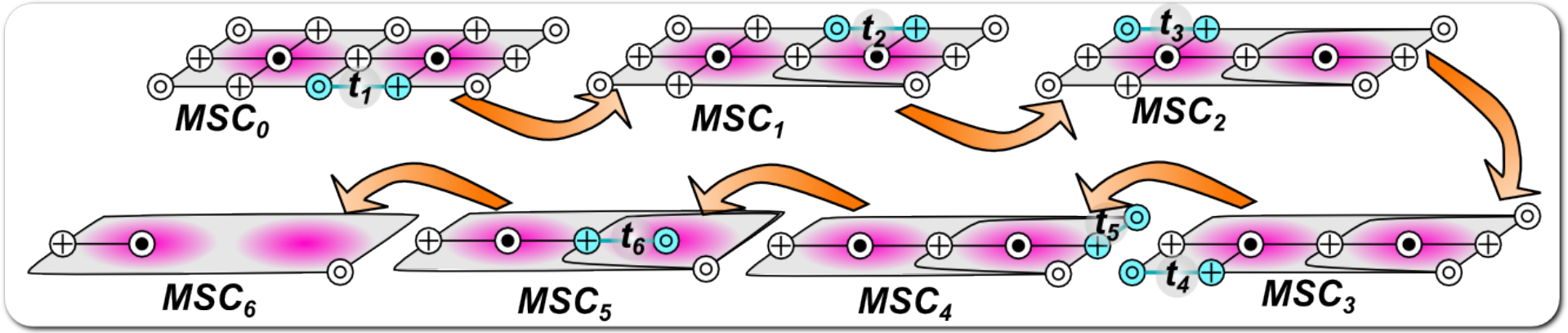
- **The function f is called a Morse-Smale function if the ascending and descending manifolds of all pairs of critical points intersect only transversally**

- **Combinatorial representation : nodes along with the 1-D arcs that connect them**



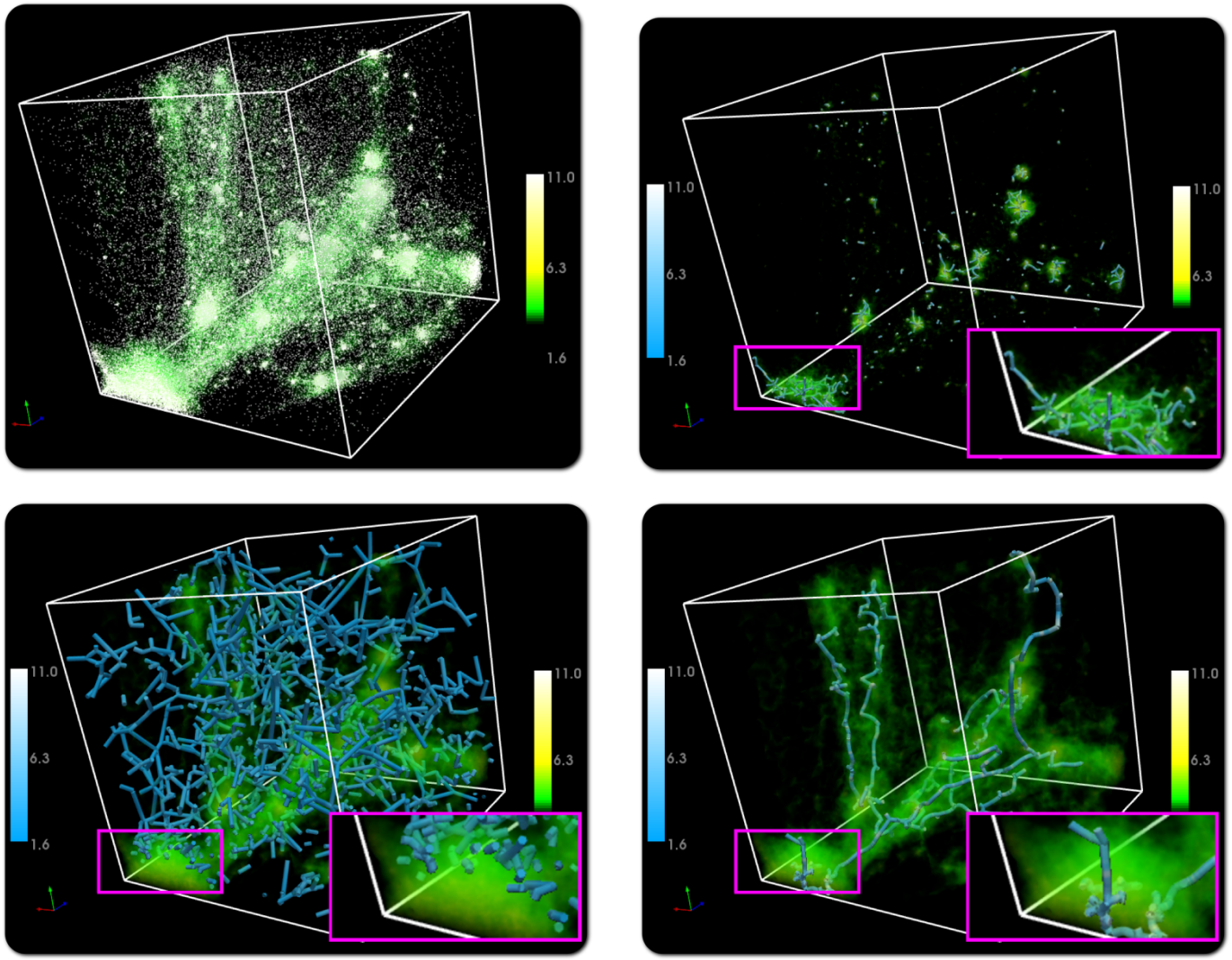
Morse-Smale Complex: simplification & hierarchy





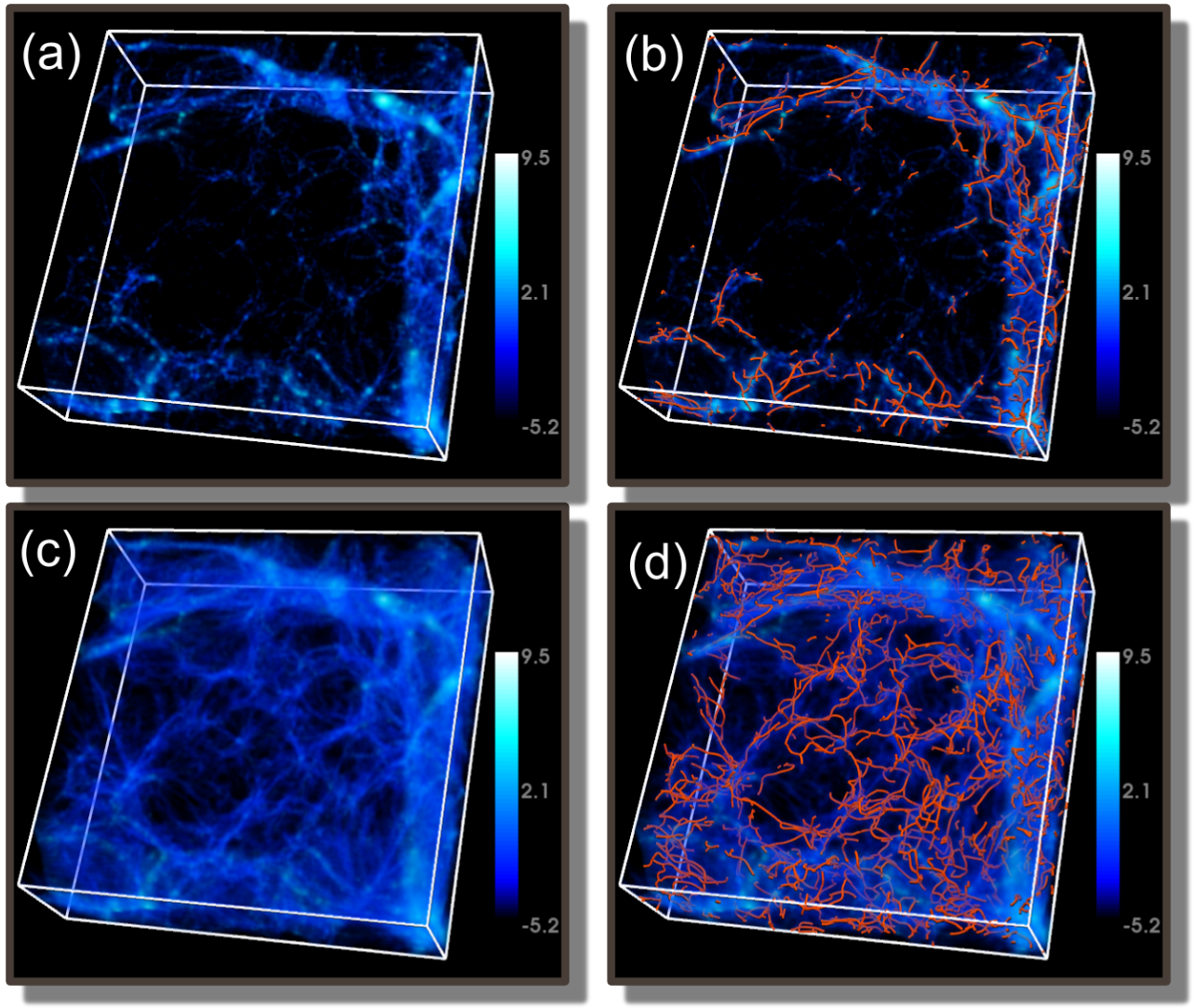
Density range based filament estimation

- filaments identified by the density range of maxima and 2-saddles



Density range based filament estimation

•filaments identified by the density range of maxima and 2-saddles



Conclusions

- Cosmic mass distribution forms and evolves hierarchically : structures ubiquitous at all density and spatial scale ranges
- Topology ideal tool for studying complex spatial patterns manifested in the universe
- Homology and persistence resolve differences between models where EC and MF are inadequate
- Persistence provides rich language for description of multi-scale (hierarchical) topology of cosmic structure
- Homology and persistence capture the different morphologies and hierarchies of the cosmic mass distribution
- Hints of violation of Gaussianity and isotropy captured by homology and persistence