Persistent Homology of the Intergalactic Medium via the Lyman-alpha Forest

Jessi Cisewski
Carnegie Mellon University

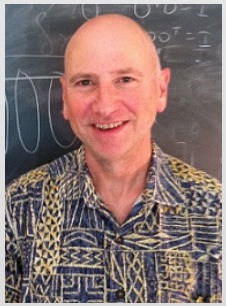
Topological Data Analysis Workshop
SAMSI

February 7, 2014

Collaborators

Peter Freeman

Christopher R. Genovese



Larry Wasserman

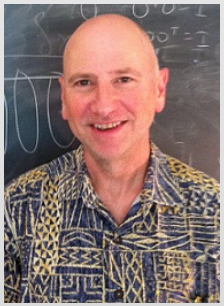
Rupert A. C. Croft

Melih Özbek

CMU TopStat: http://www.stat.cmu.edu/topstat/

Fabrizio Lecci

Alessandro Rinaldo



Larry Wasserman

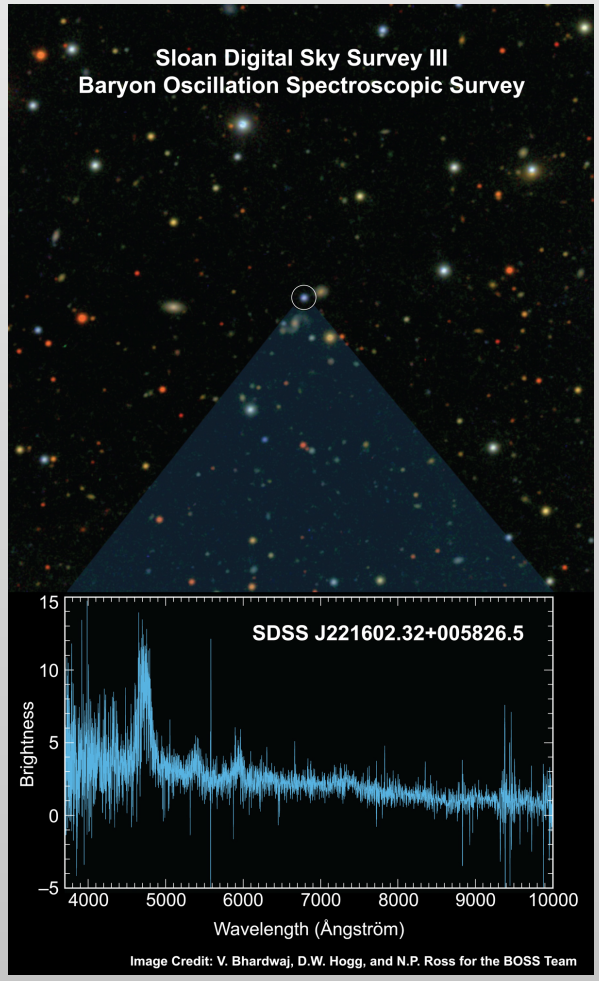
Brittany Fasy

Jisu Kim

And others...

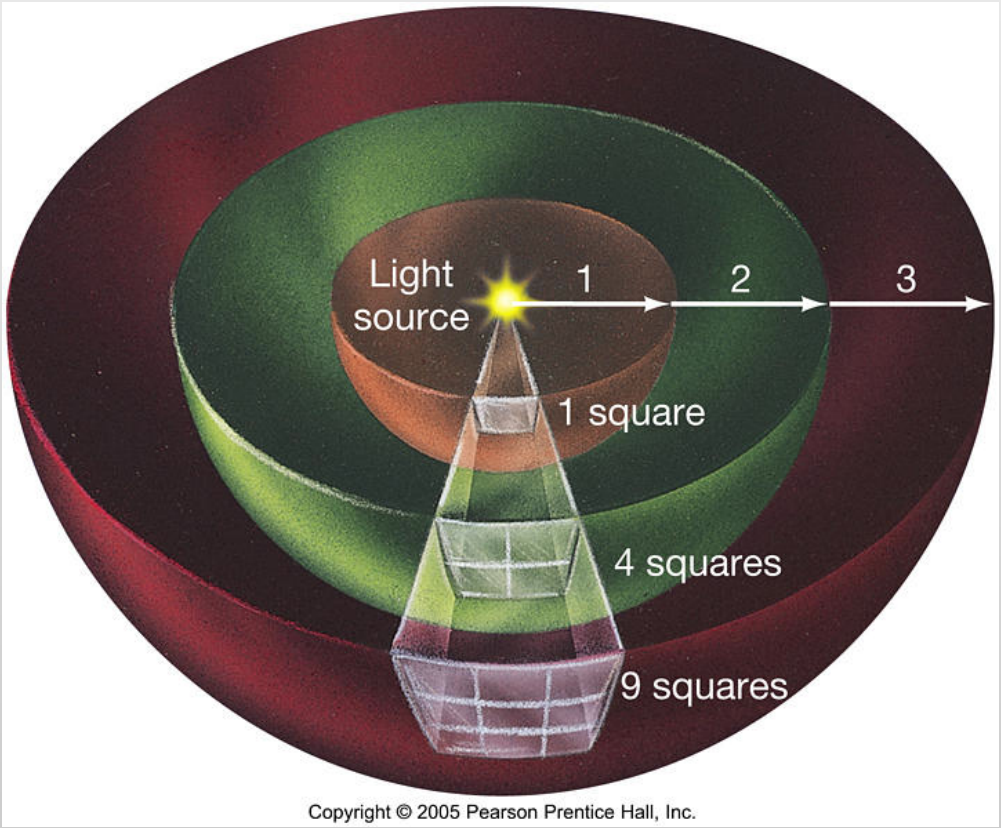
Outline

1. **Background**
   - Lyman-alpha forest allows us to "see" parts of the Universe that were previously not accessible
2. **Methodology**
3. **Persistent homology for cosmology**



Source: http://newscenter.lbl.gov

Light: our source of information about the Universe



Source: Astronomy Today (Chaisson & McMillian 2011)

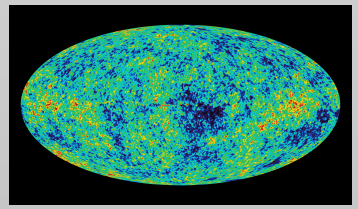
$$\text{Inverse-square Law: Apparent brightness} \propto \frac{\text{luminosity}}{\text{distance}^2}$$

Distance measurements

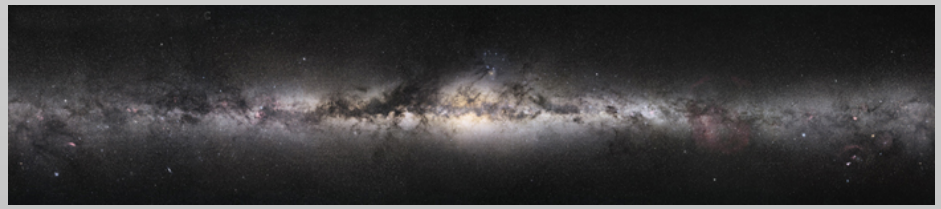
Source: http://www.universetoday.com

Source: http://www.spacetelescope.org

1 Megaparsec $\approx 3.26 \times 10^6$ light years

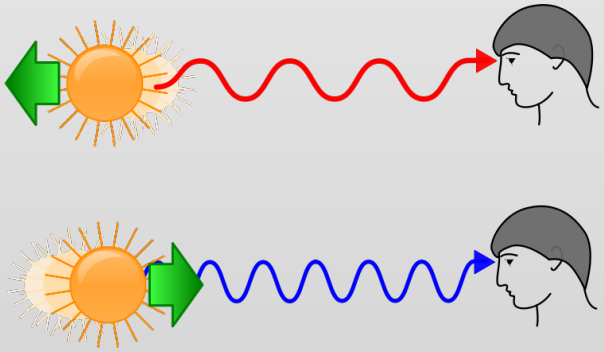


Source: http://astrohow.org



Source: http://www.digitalskyllc.com

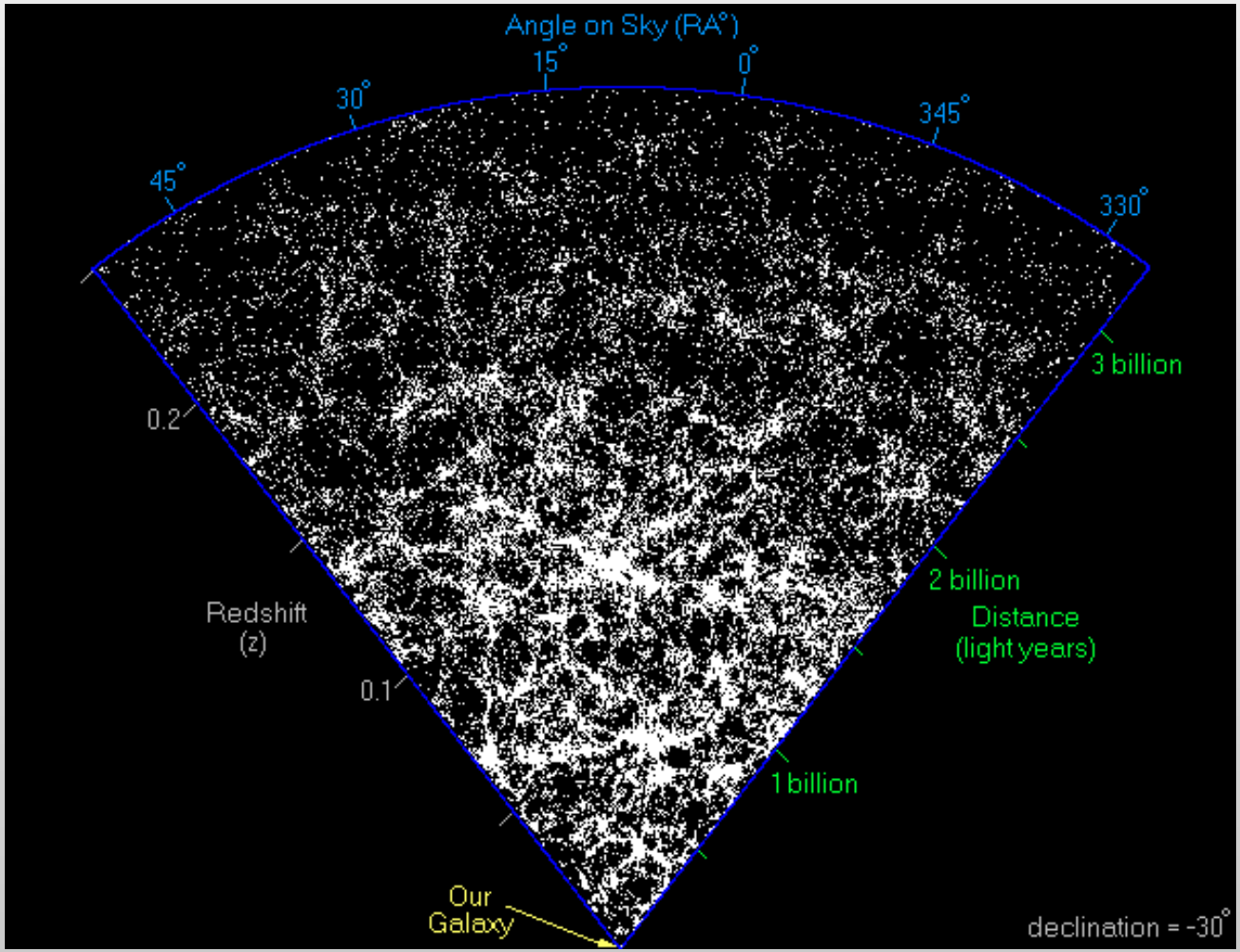
Redshift



Source: http://en.wikipedia.org

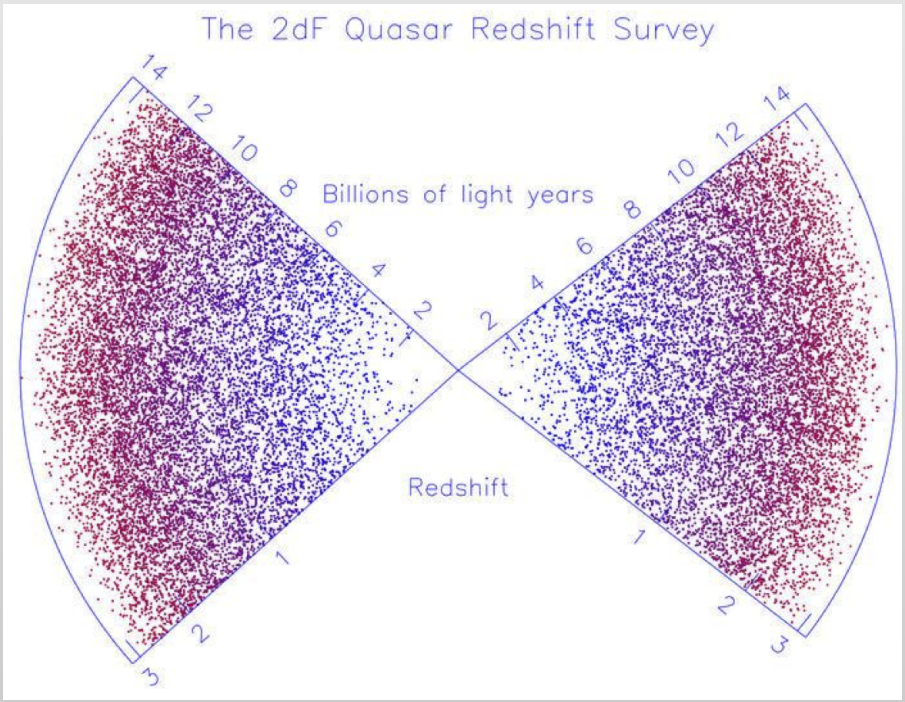
Source: http://coolcosmos.ipac.caltech.edu

$$\textbf{Redshift (z)} = \frac{\lambda_{\textbf{observed}} - \lambda_{\textbf{emitted}}}{\lambda_{\textbf{emitted}}} = \frac{\textbf{recessional velocity}}{\textbf{speed of light}}$$

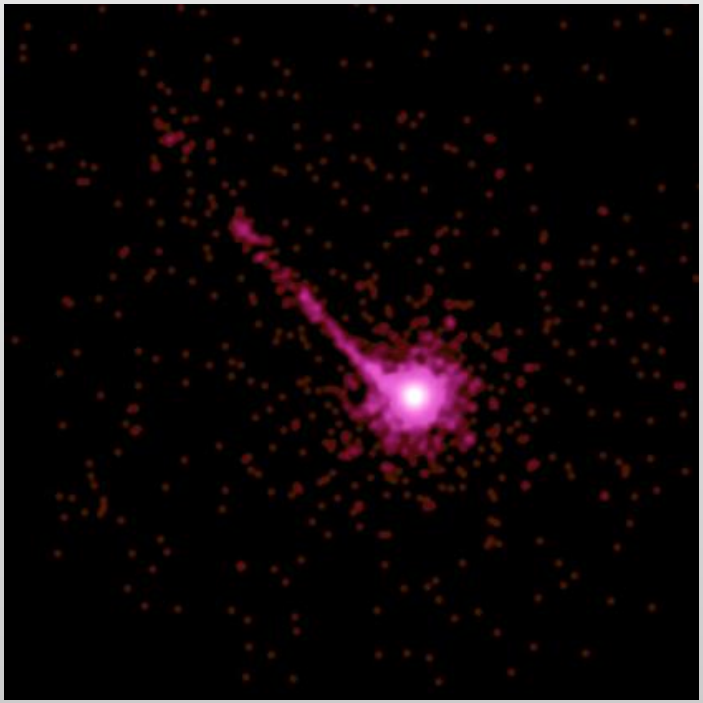


Source: `http://www.atlasoftheuniverse.com`

Quasars



Source: `http://www.2dfquasar.org`



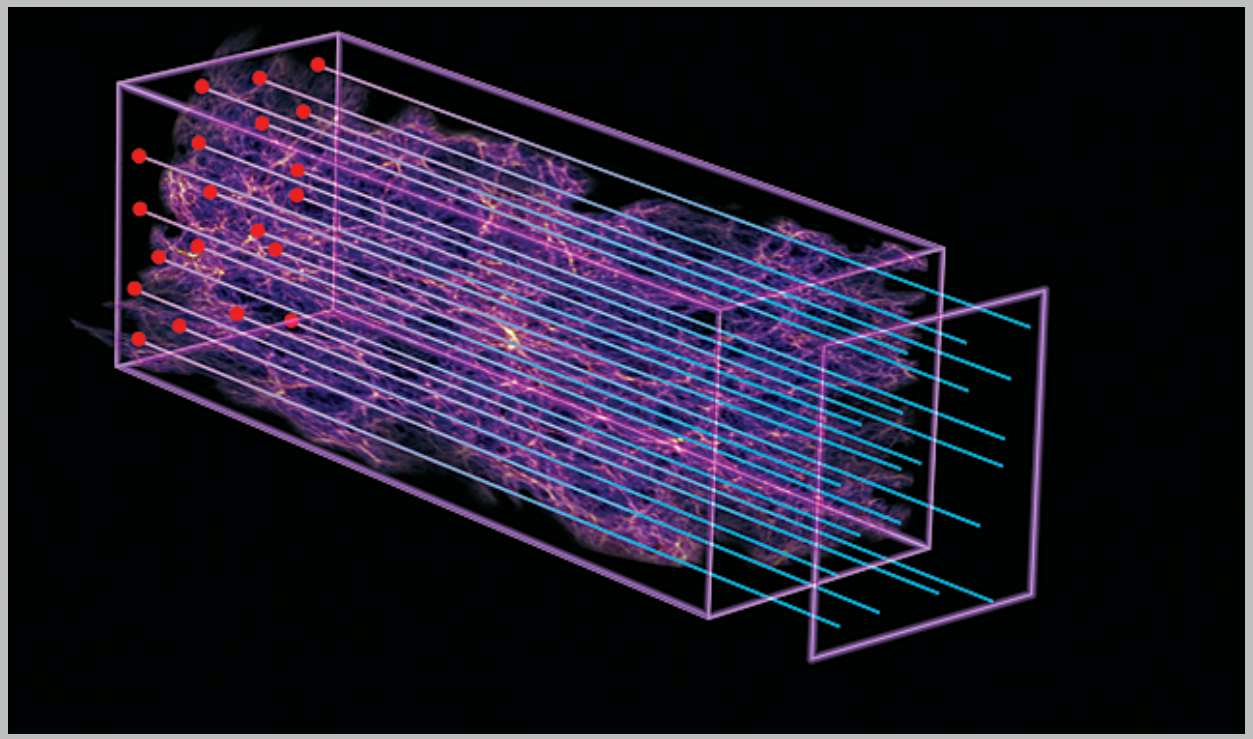
The X-ray image of the quasar PKS 1127-145

Source: `http://chandra.harvard.edu`

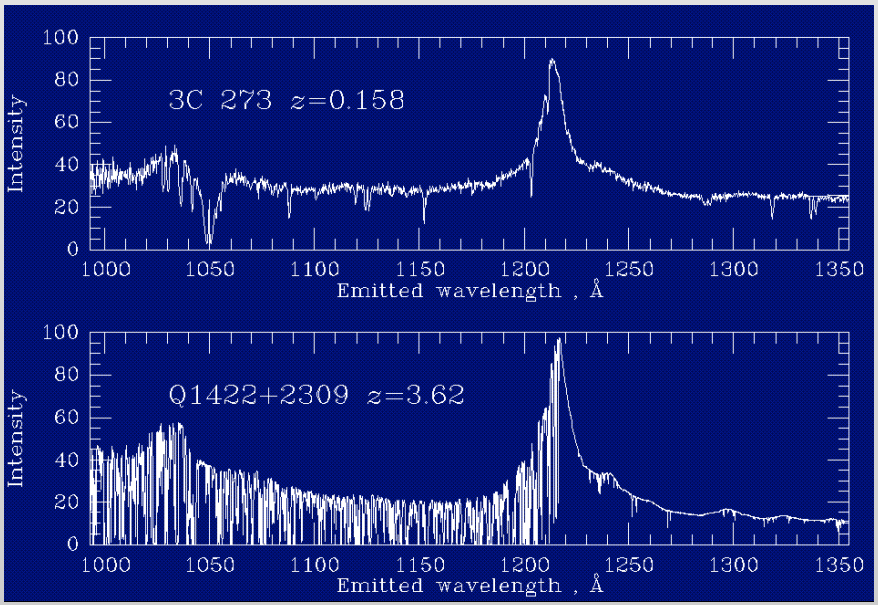
SDSS-III
Baryon Oscillation Spectroscopic Survey (BOSS)

- 54,468 quasar spectra with redshifts $2.0 < z < 5.7$

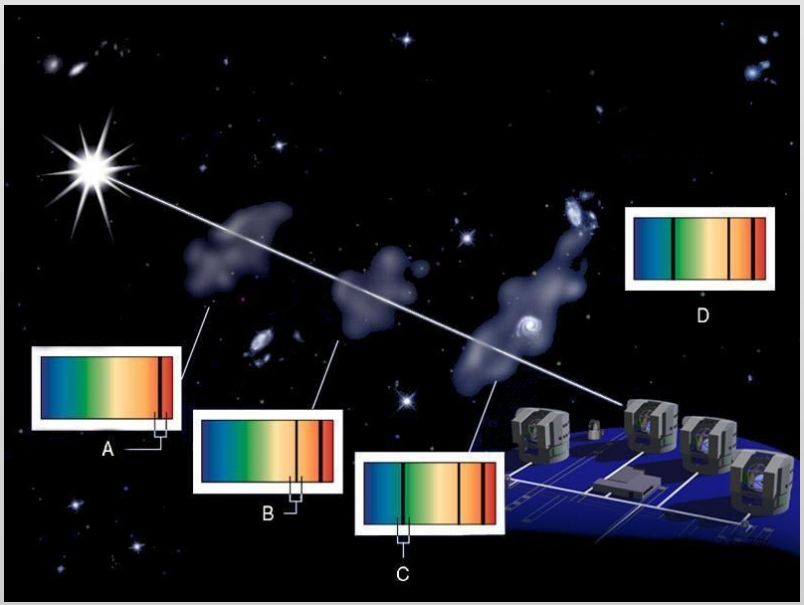
Lines of sight from distant quasars



Lyman-alpha forest

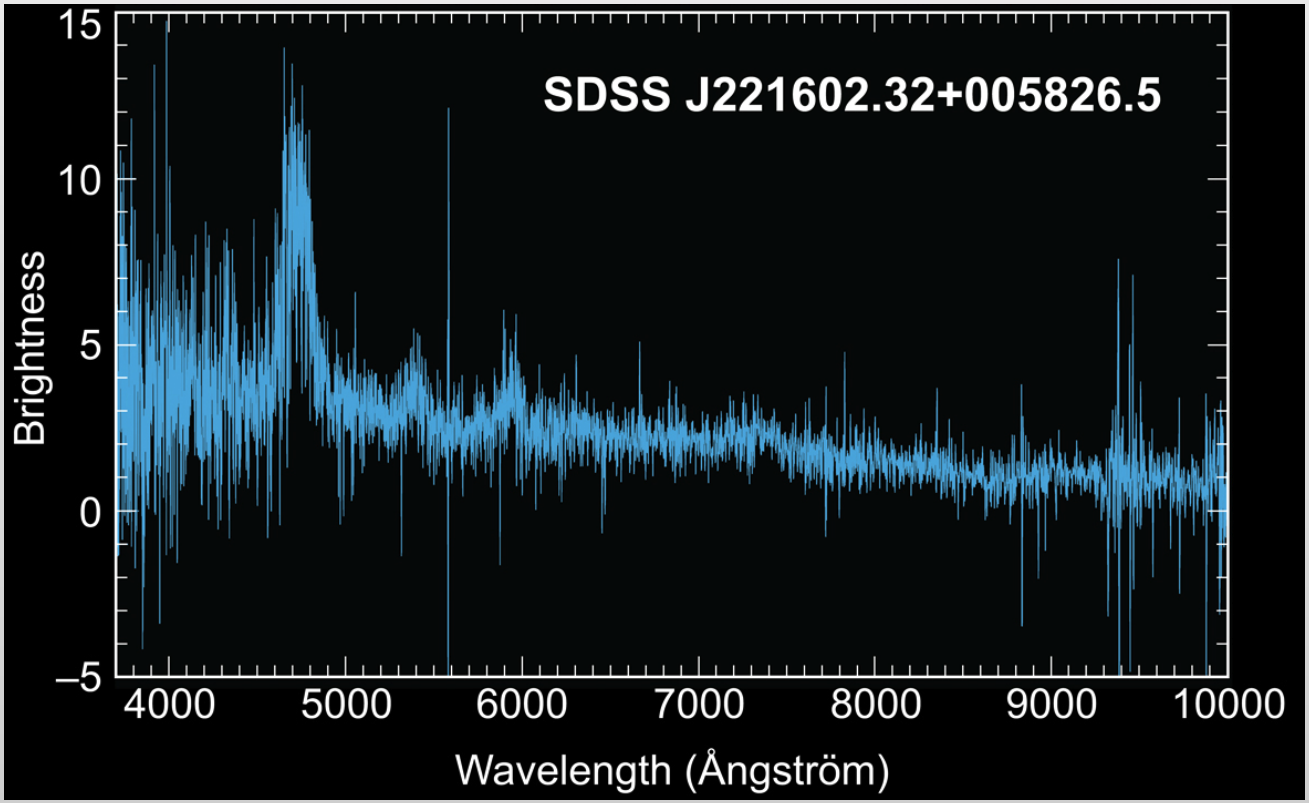


Source: William C. Keel (`www.astr.ua.edu/keel/`)



Source: `http://www.hs.uni-hamburg.de`

Quasar spectrum



Mattia Ciollaro
(CMU Stat Ph.D. student)

What does the Lyman-alpha forest reveal?

- 3D map of high redshift Intergalactic Medium at a large scale
- Constraints on initial conditions (and subsequent evolution of the Universe)
- Location of early galaxies
- Tracer of Dark Matter; possibly constrain theories on dark matter
- A means for measuring BAO at redshifts > 2

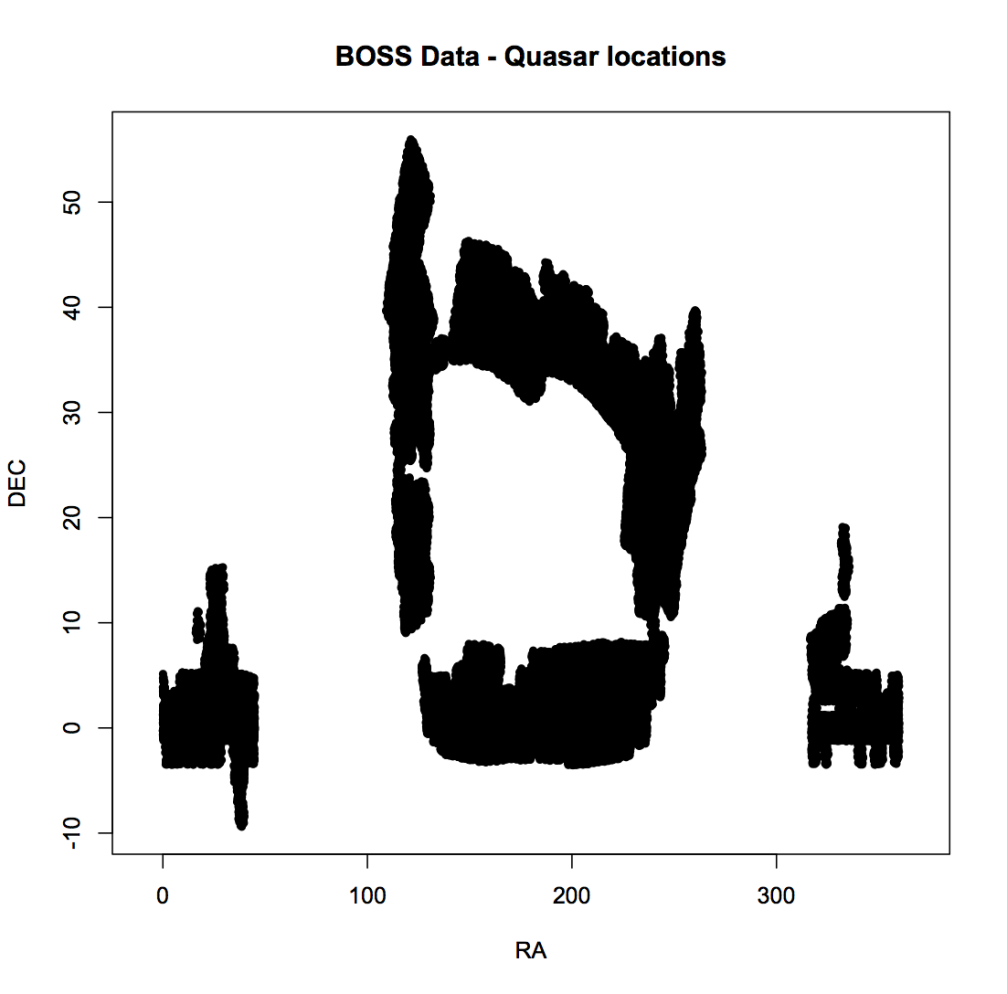
BOSS survey

- BOSS = Baryon Oscillation Spectroscopic Survey
- Part of SDSS - III Data Release 9
- 54,468 quasar spectra with redshift > 2.15
- Absorption redshifts $2.0 < z < 5.7$ over 3275 square degrees, comoving volume of $20h^{-3}Gpc^3$

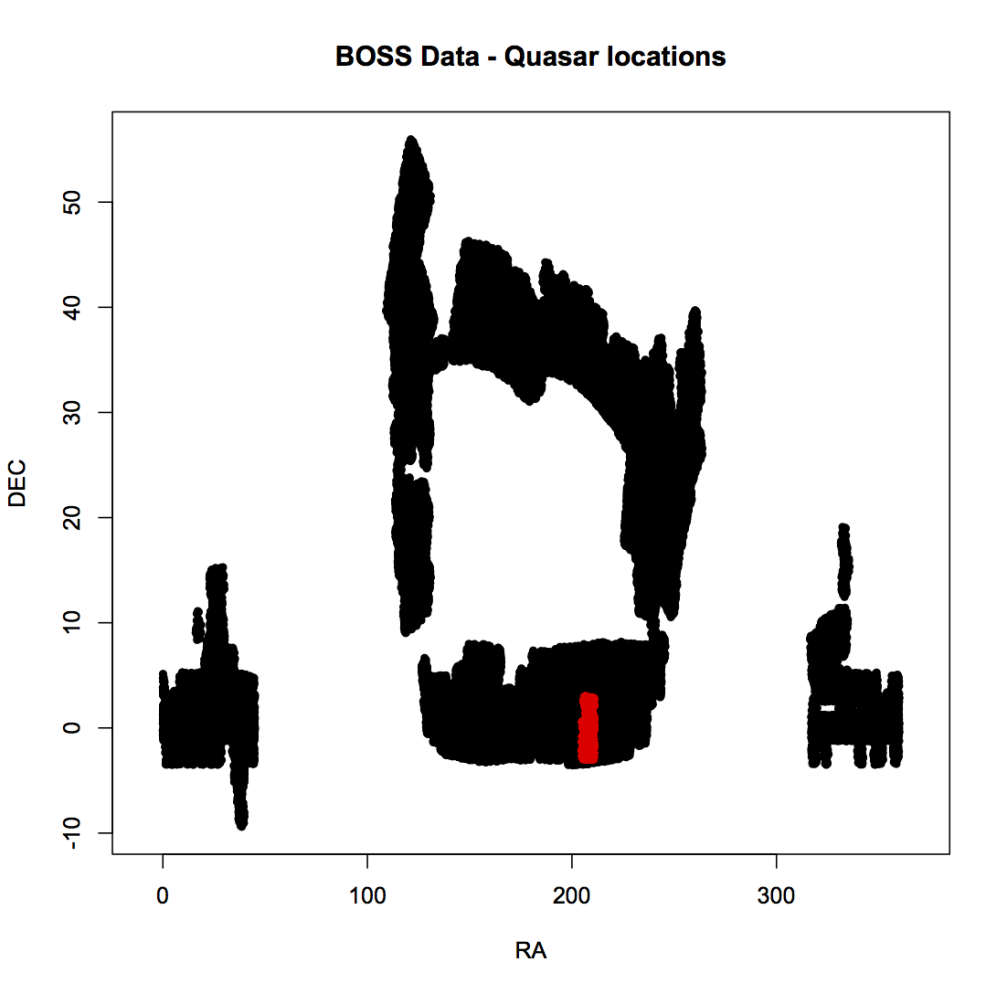
Data available to the public at `http://www.sdss3.org/dr9/`

Source: `http://www.apo.nmsu.edu`

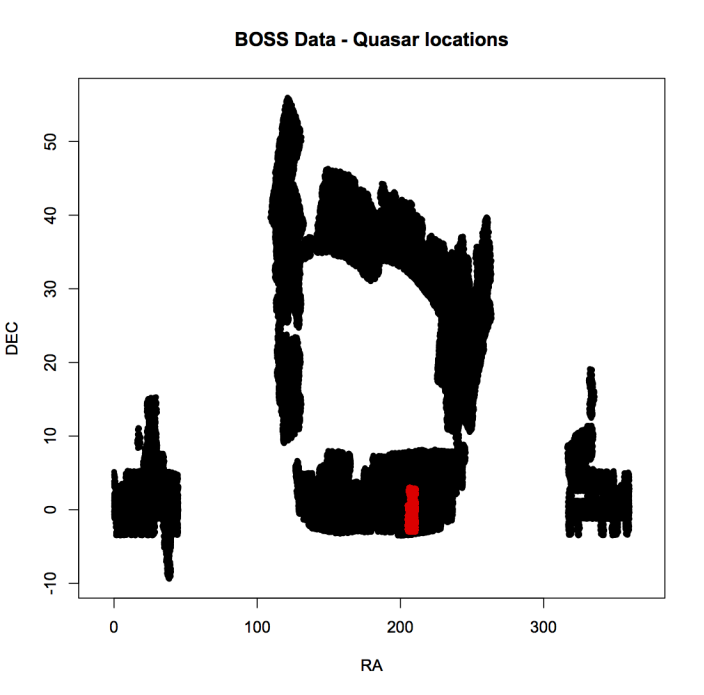
Locations of BOSS quasars

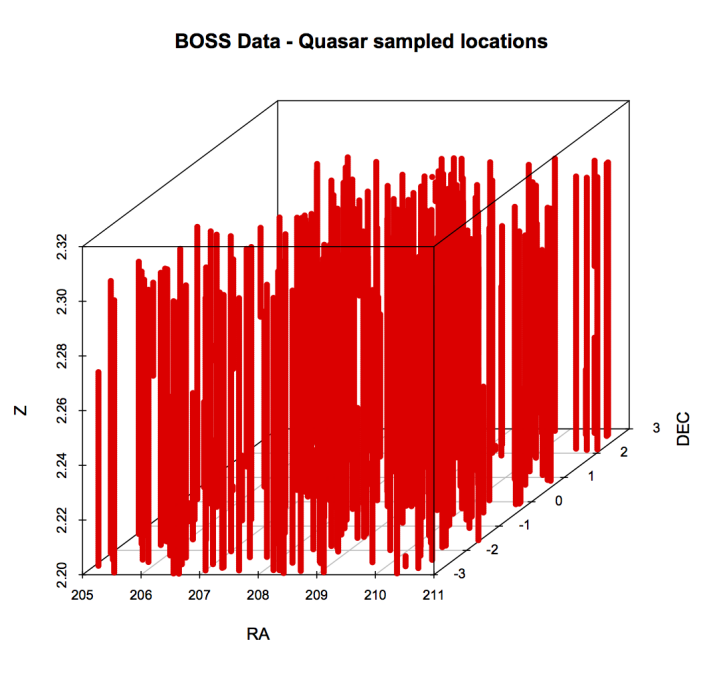


Locations of BOSS quasars



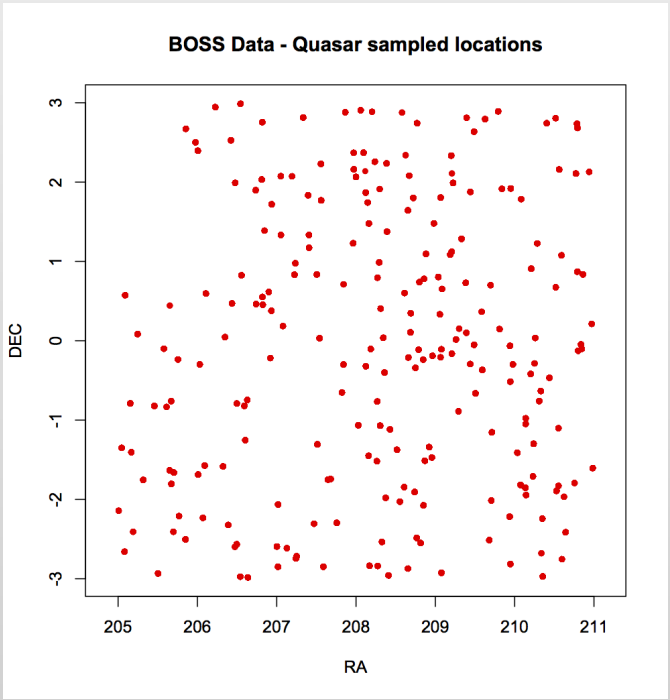
Locations of BOSS quasars

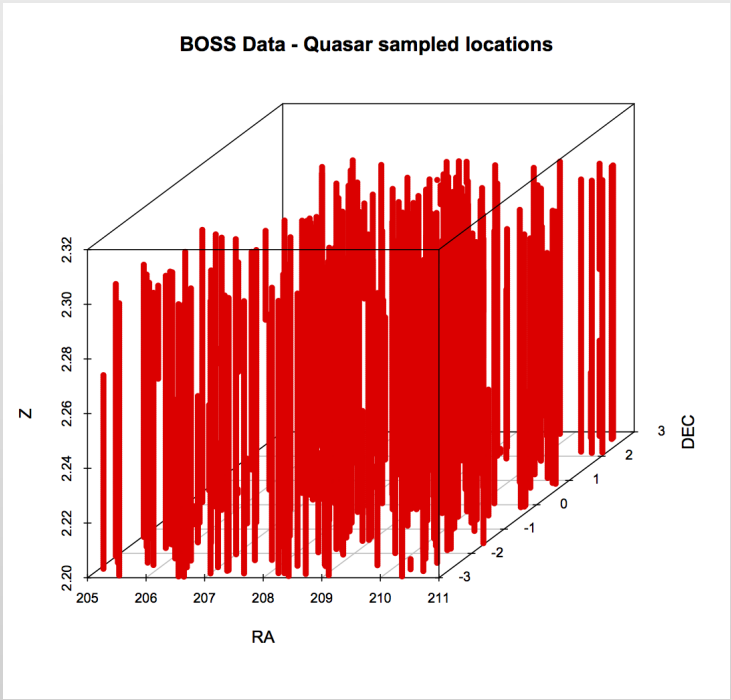




- 400 Mpc^3
- RA: 205 to 211, DEC: -3 to 3, Redshift: 2.2 to 2.3
- 24,596 measurements, 234 quasars
- 1 - 134 measurements/LOS (median = 134/LOS)

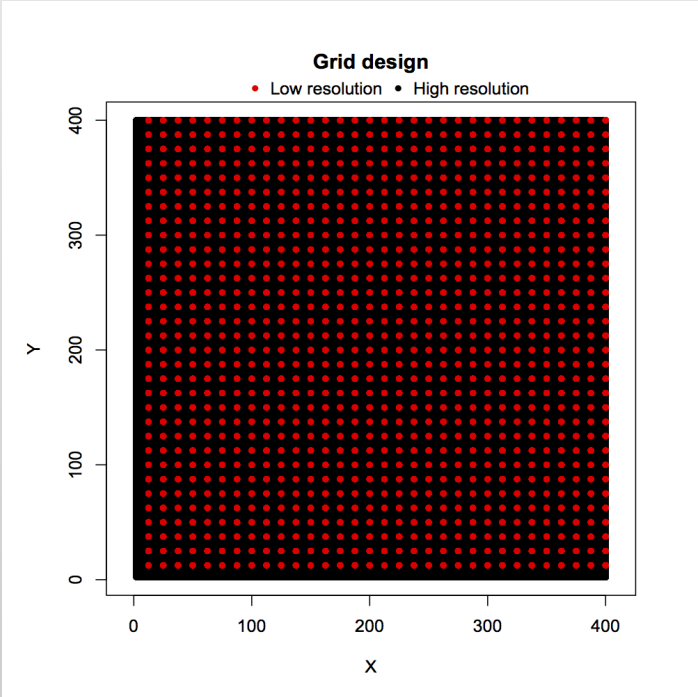
Locations of BOSS quasars

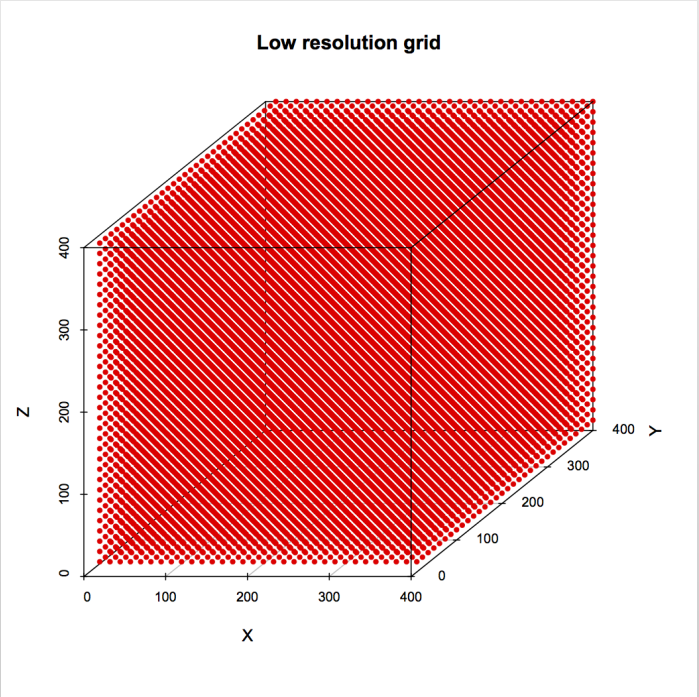




- 400 Mpc^3
- RA: 205 to 211, DEC: -3 to 3, Redshift: 2.2 to 2.3
- 24,596 measurements, 234 quasars
- 1 - 134 measurements/LOS (median = 134/LOS)

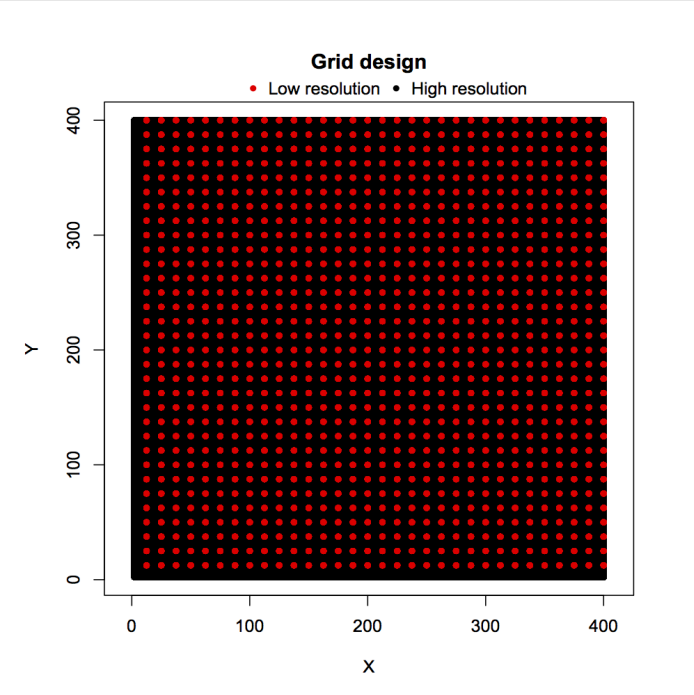
Cosmological simulations

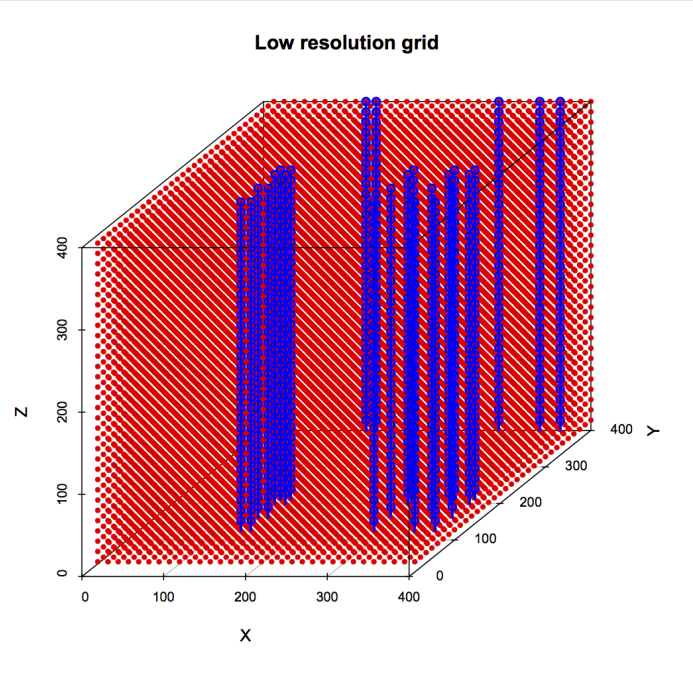




- 400 Mpc^3 at a redshift centered around 2.25

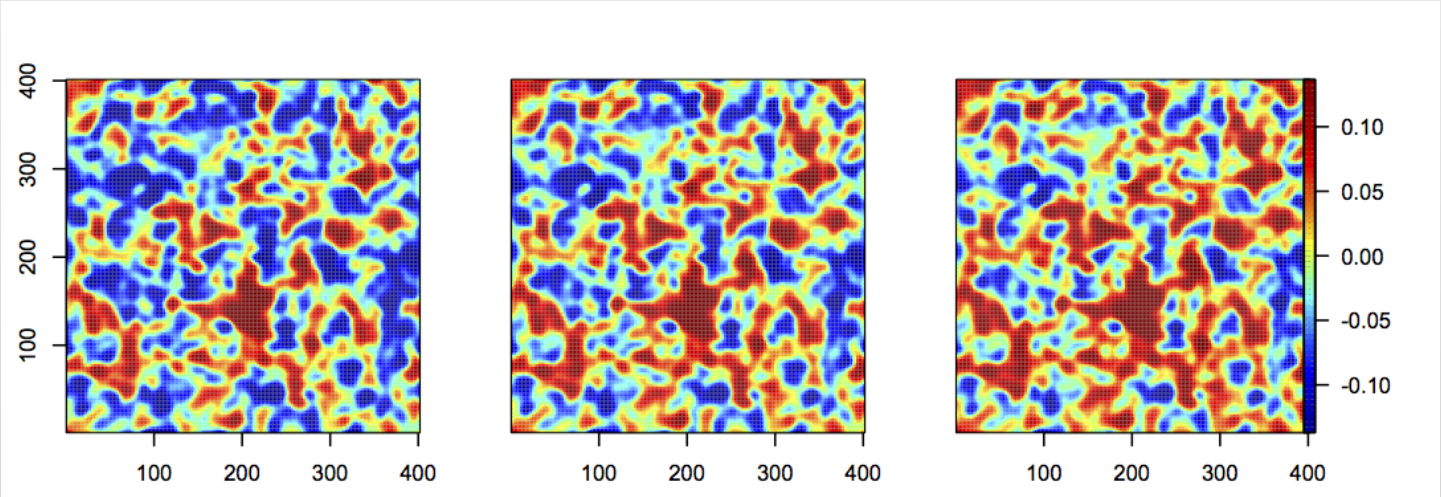
Cosmological simulations

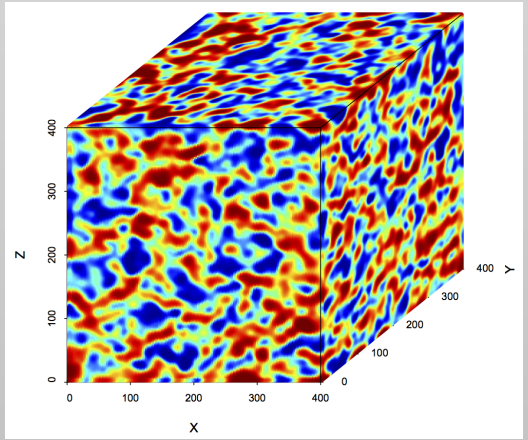




- 400 Mpc^3 at a redshift centered around 2.25
- Lines of sight of quasars

Local polynomial smoothing of simulation

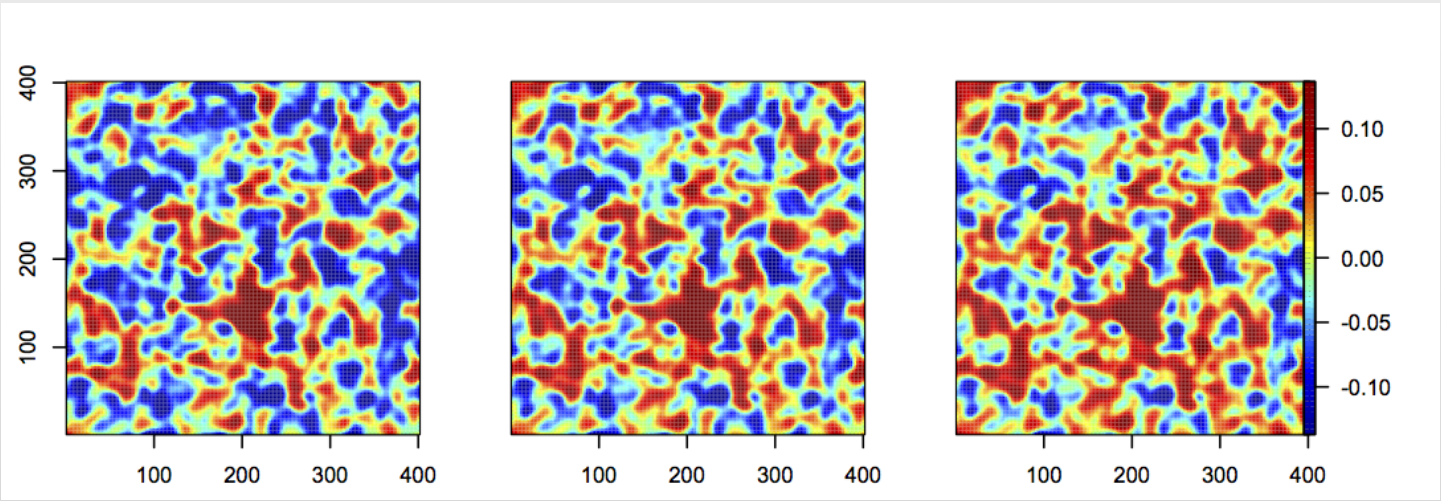




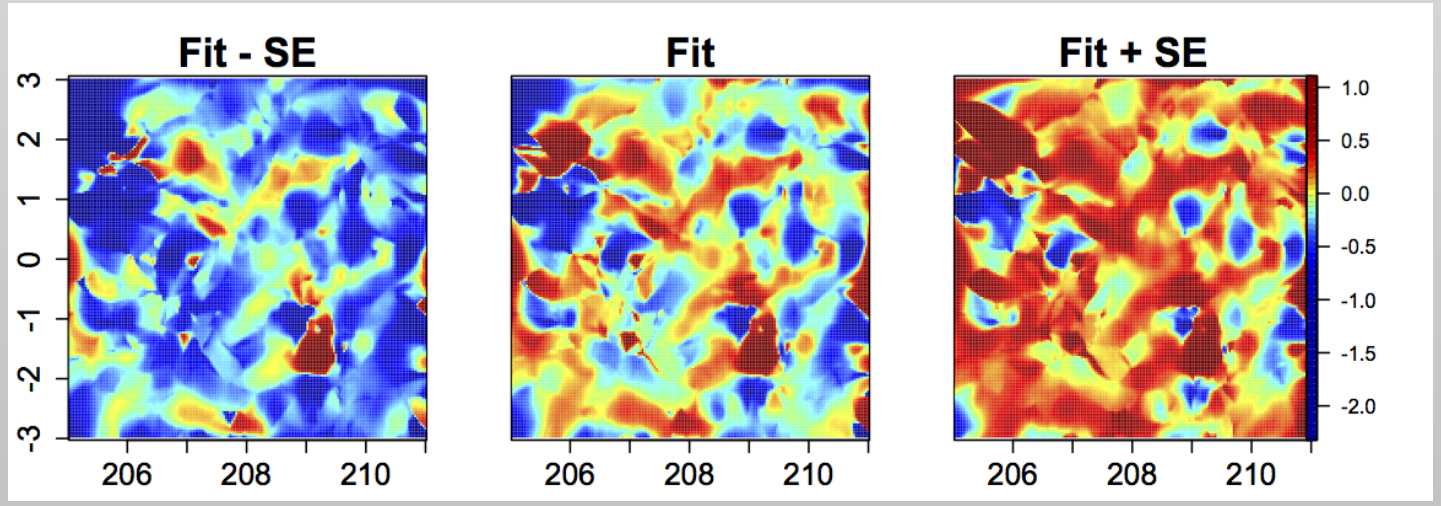
→ Redder = more neutral Hydrogen

→ Bluer = less neutral Hydrogen

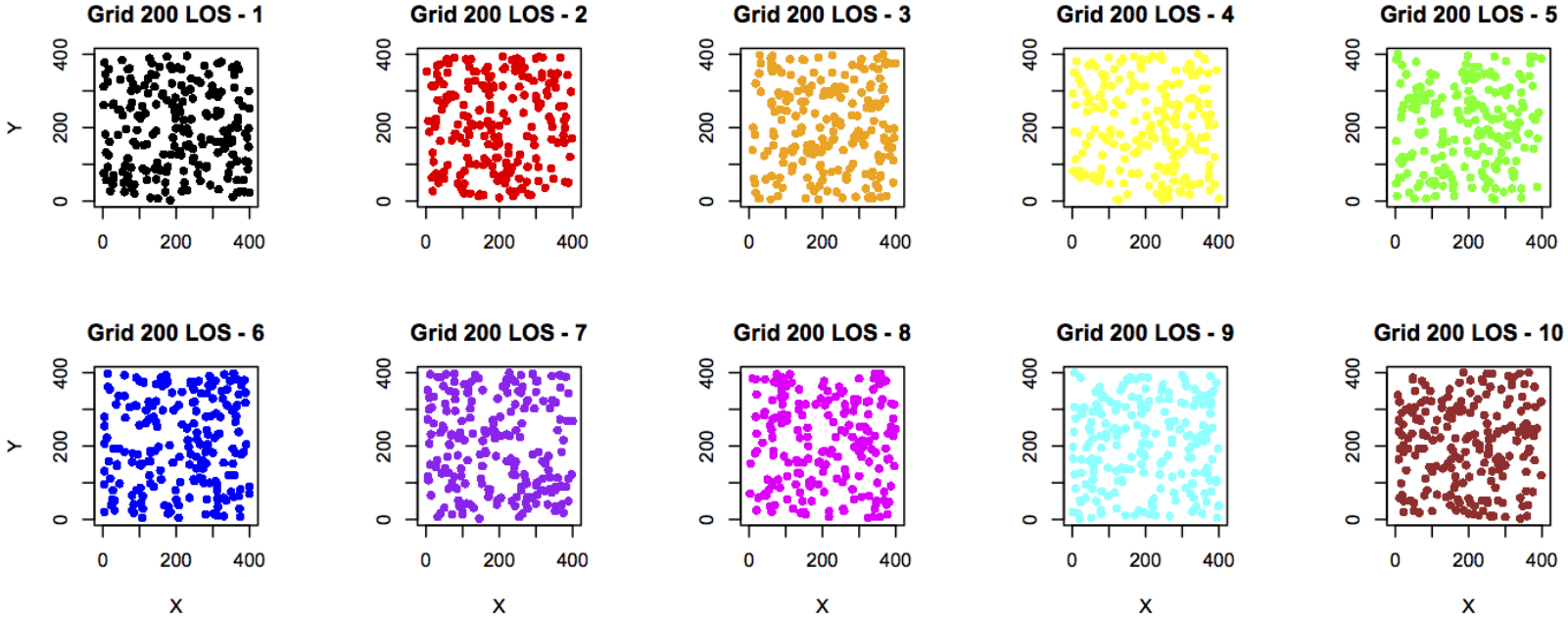
Local polynomial smoothing of simulation



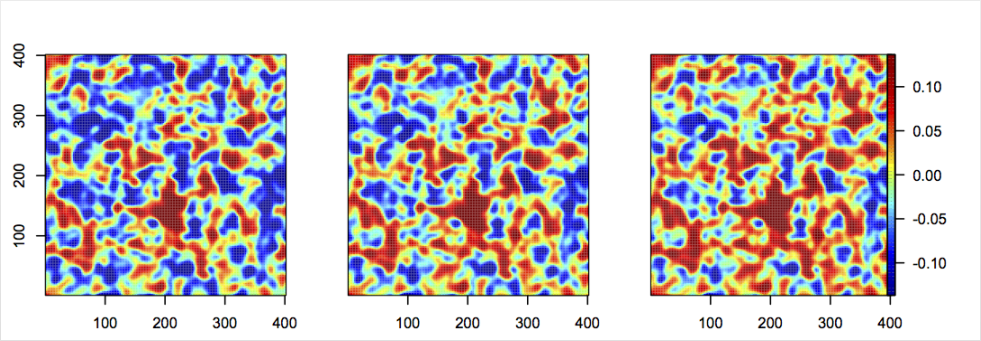
And real data (from SDSS-III BOSS)



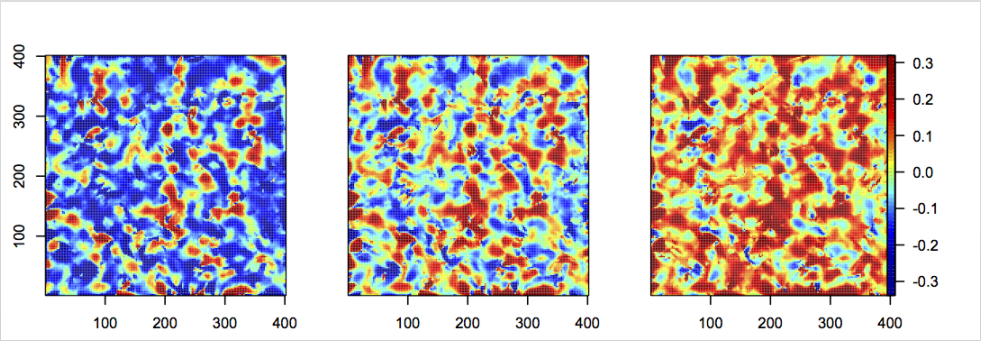
Sampled grids with 200 lines-of-sight



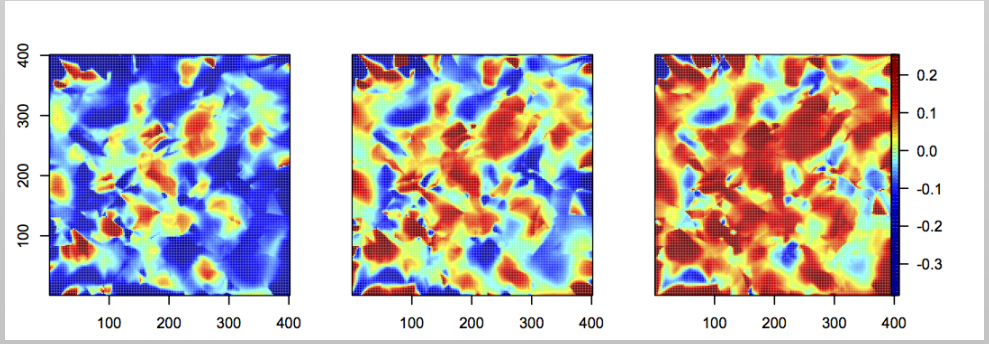
Full (30,976 LOS)



1000 LOS



200 LOS



- **Question: How do we evaluate the fit? Are the cosmological simulations are "matching" the real data?**

- **Question: How do we evaluate the fit? Are the cosmological simulations are "matching" the real data?**

Common measures:

1. Visual comparison (compare slices)
   - Any obvious difference between images/figures?
2. PDFs by redshift
   - Symmetric/skewed, number/location of modes, etc.
3. Correlation function
4. Local modes
5. Euler characteristic
6. And others...

- **Question: How do we evaluate the fit? Are the cosmological simulations are "matching" the real data?**

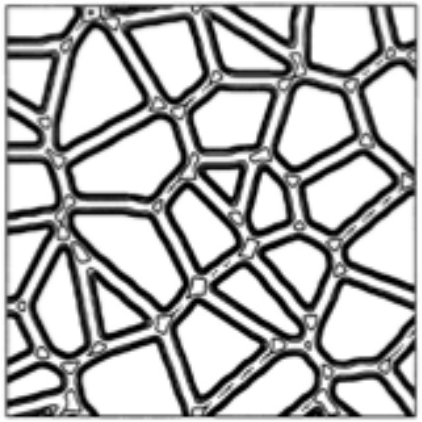
  Common measures:

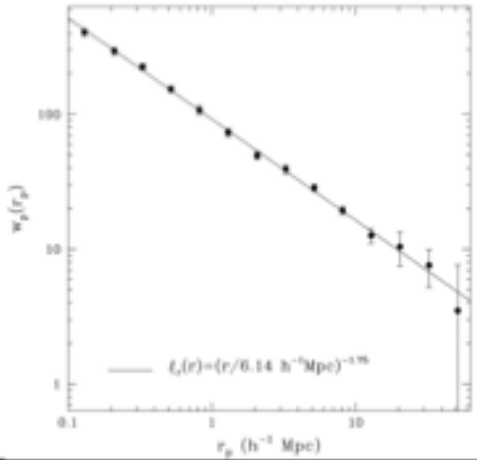
  1. Visual comparison (compare slices)
     - Any obvious difference between images/figures?
  2. PDFs by redshift
     - Symmetric/skewed, number/location of modes, etc.
  3. Correlation function
  4. Local modes
  5. Euler characteristic
  6. And others...

- **Are these measures capturing all the important features of the data?**

Correlation functions: Structural Insensitivity

Voronoi foam, R=1.6, smoothed original



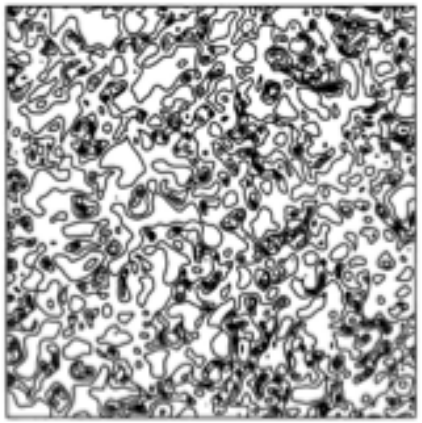


$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$$

$$\gamma = 1.8$$

$$r_0 = 5h^{-1}Mpc$$

Voronoi foam, R=1.6, random phases



2-pt correlation function is highly insensitive to the geometry & morphology of weblike patterns:

compare 2 distributions with same ξ(r), cq. P(k), but totally different phase distribution

Figure from Pratyush Pranav (Rijksuniversiteit te Groningen)

Persistent homology

What is persistent homology?

A tool of topological data analysis for understanding, describing and quantifying certain topological features of data.

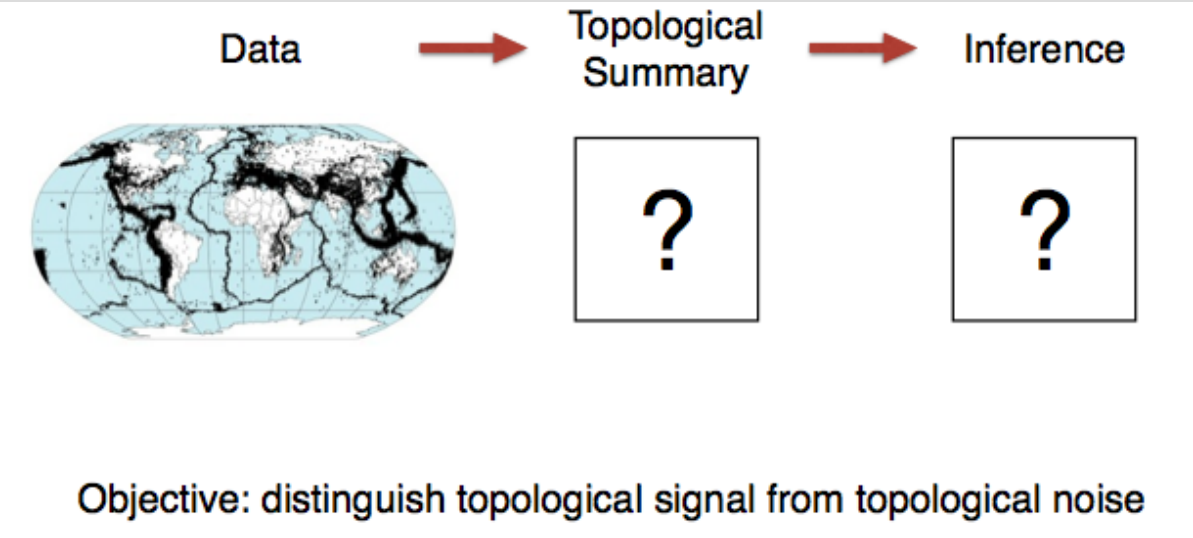


Figure from Fabrizio Lecci (CMU)

Upper level sets

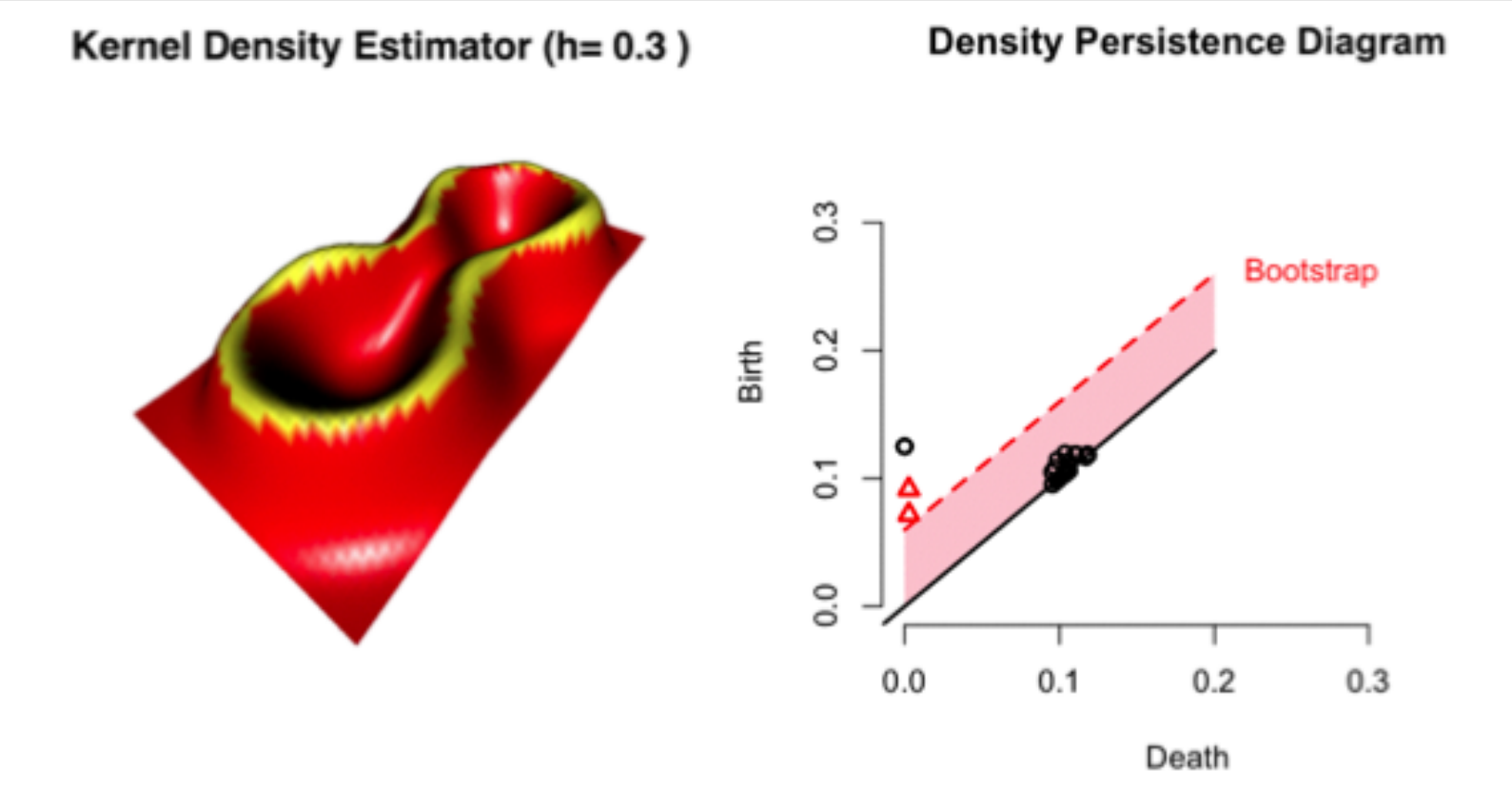
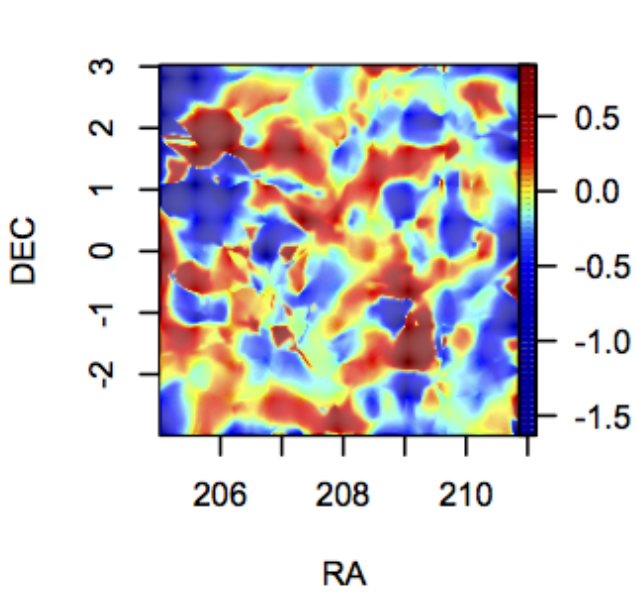
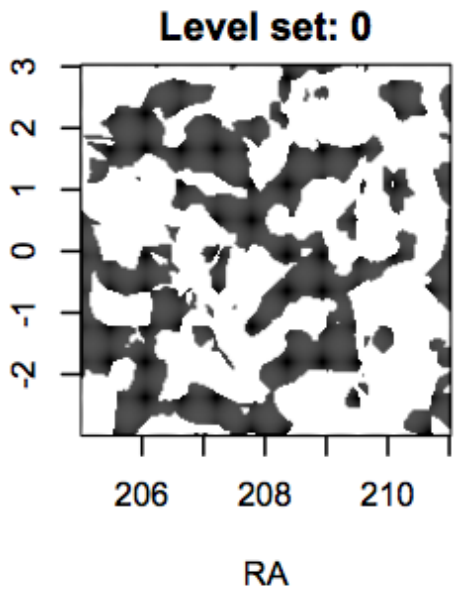


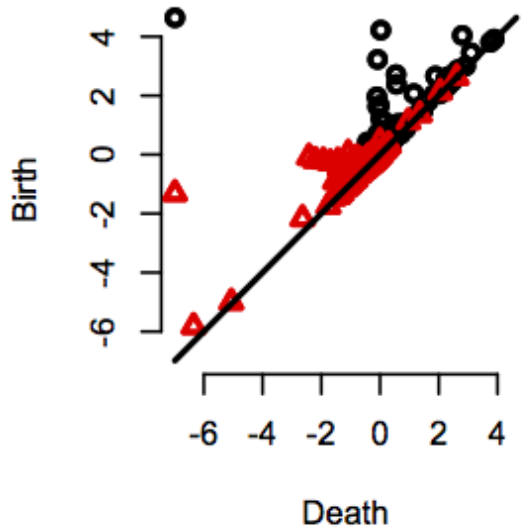
Figure from Fabrizio Lecci (CMU)

BOSS data slice at z ≈ 2.21

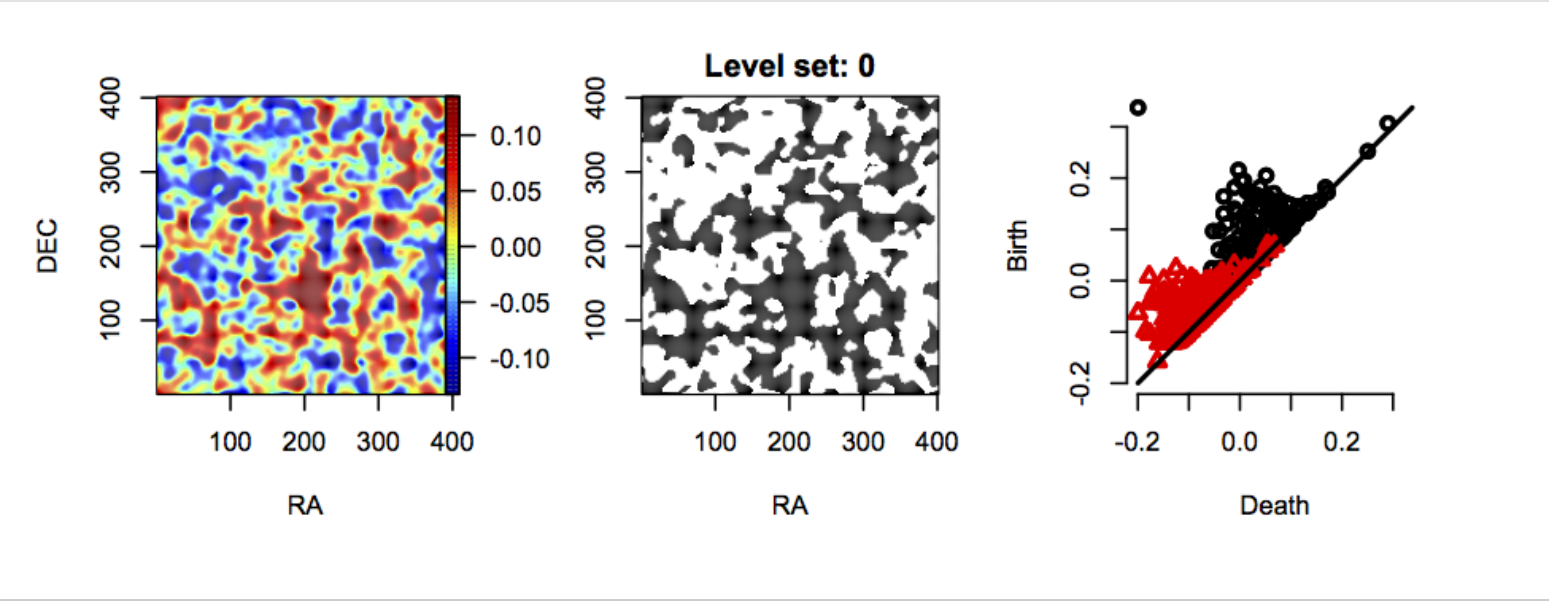
Level set at 0



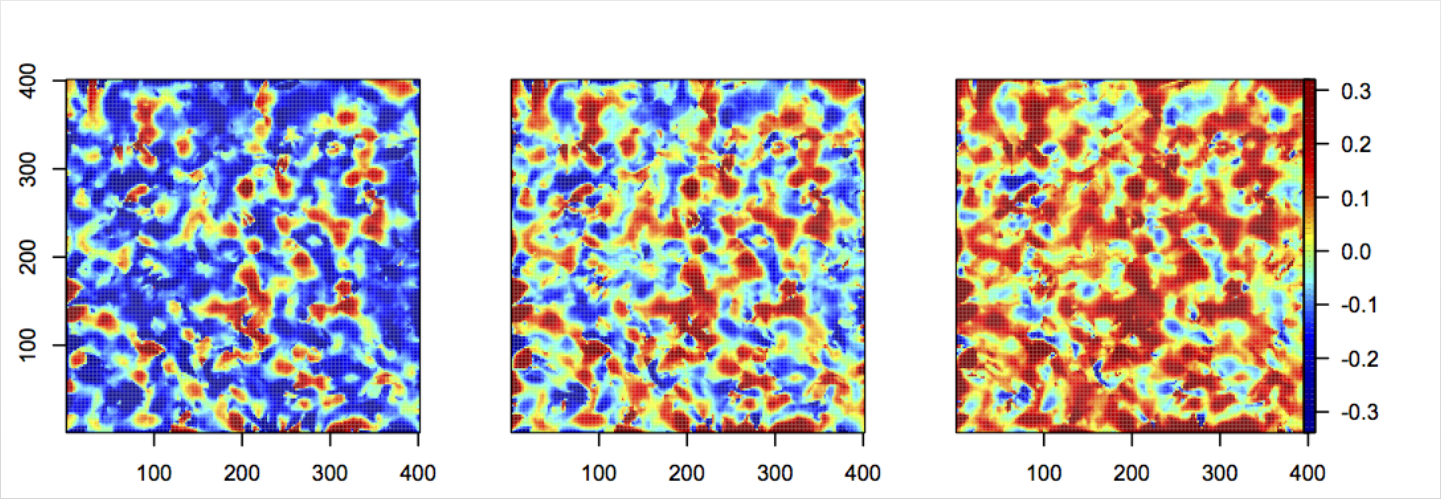


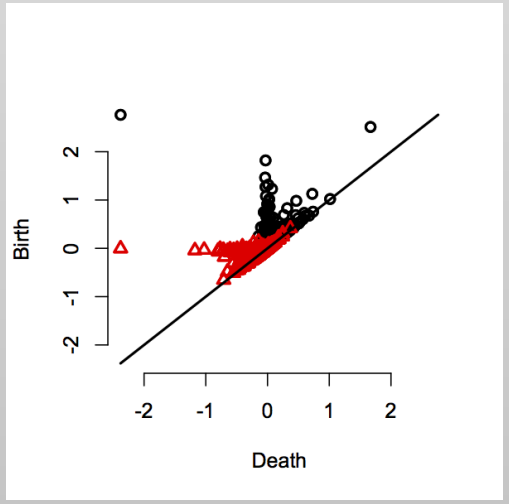


Persistent homology: Full (30,976 LOS)

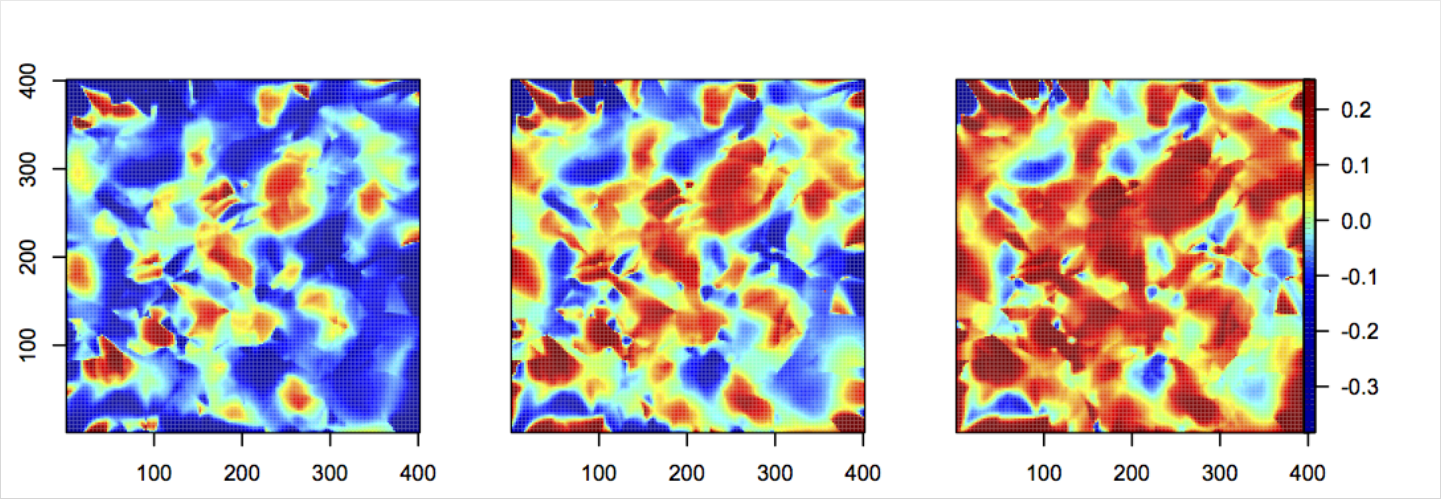


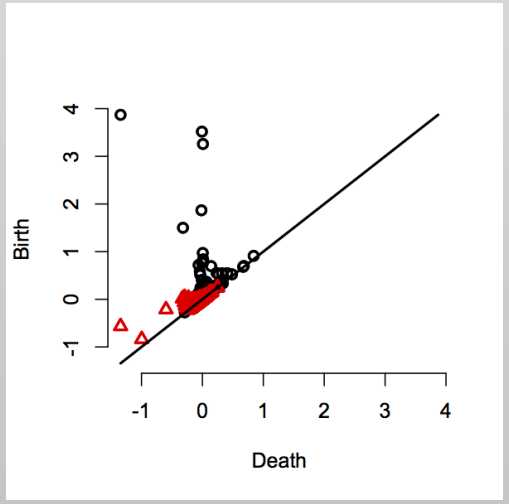
Persistent homology: 1000 LOS



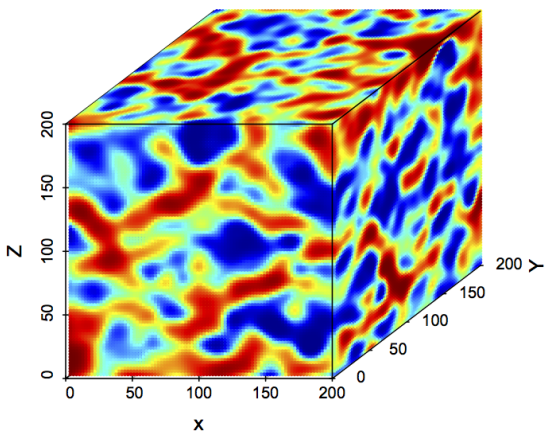


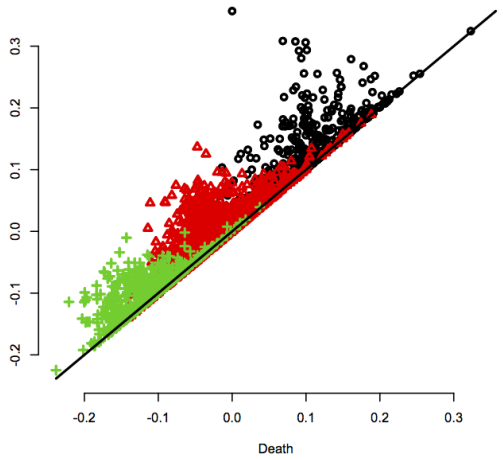
Persistent homology: 200 LOS

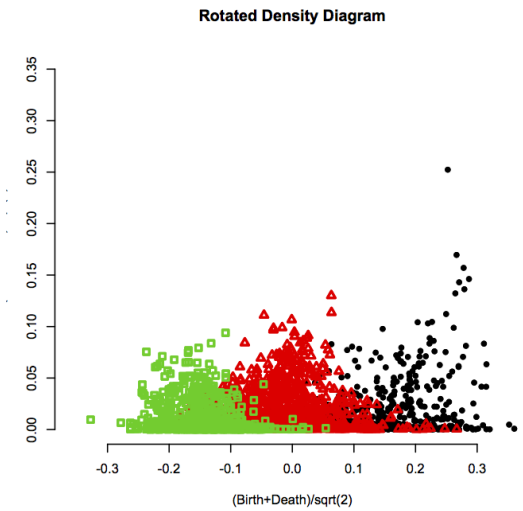




Persistent homology: Full (30,976 LOS)



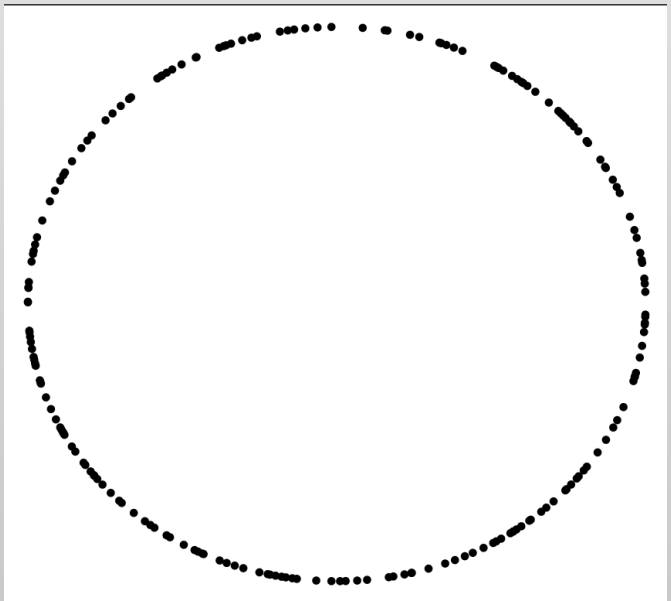


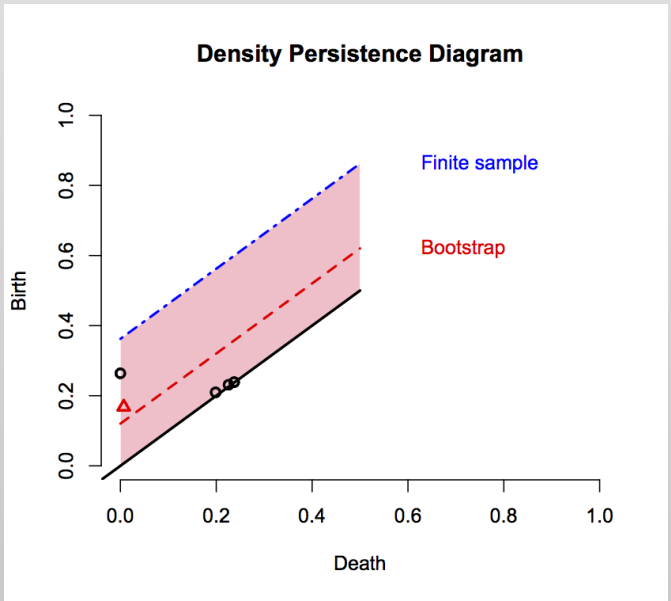


Inference for persistent homology

Confidence intervals for persistent homology

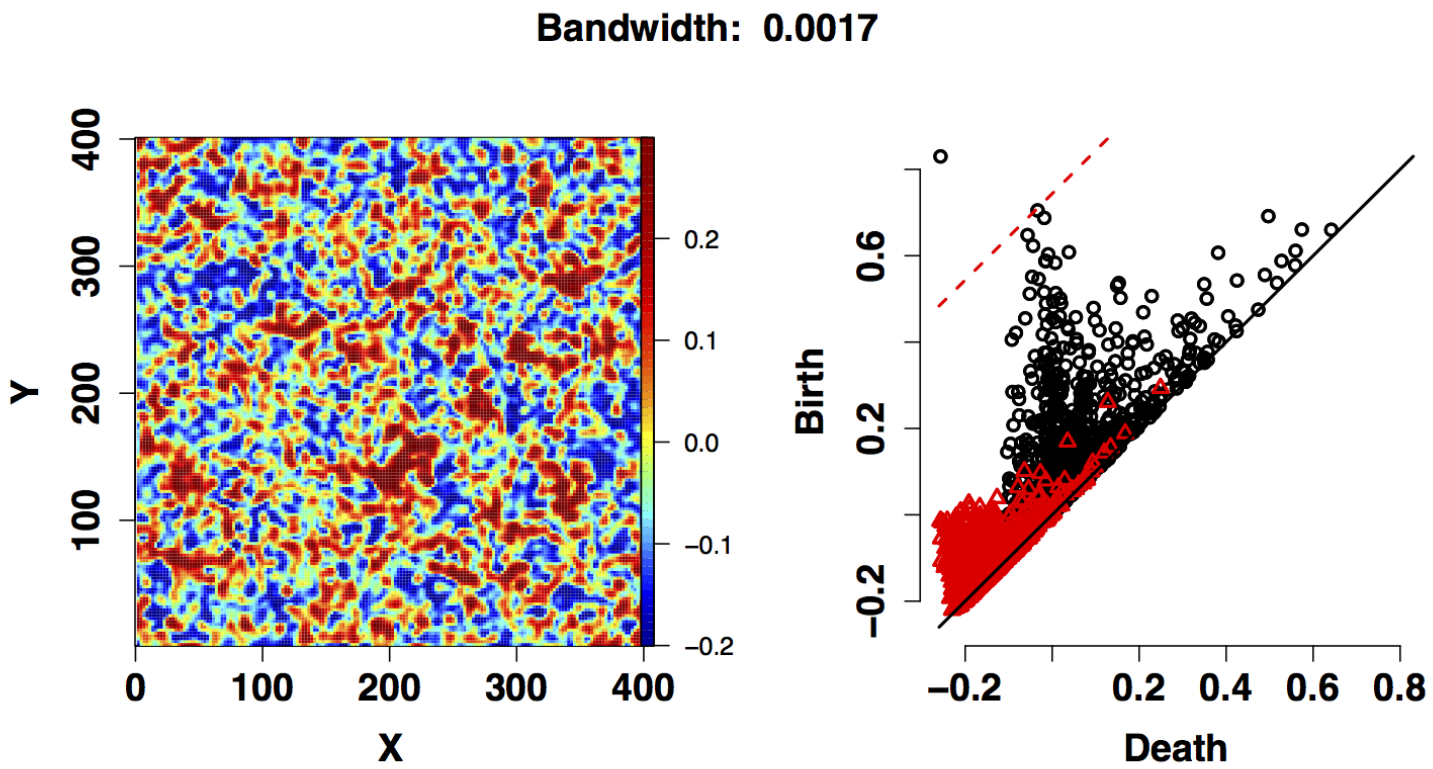
- Sivaraman Balakrishnan, Brittany Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, Larry Wasserman (2013)



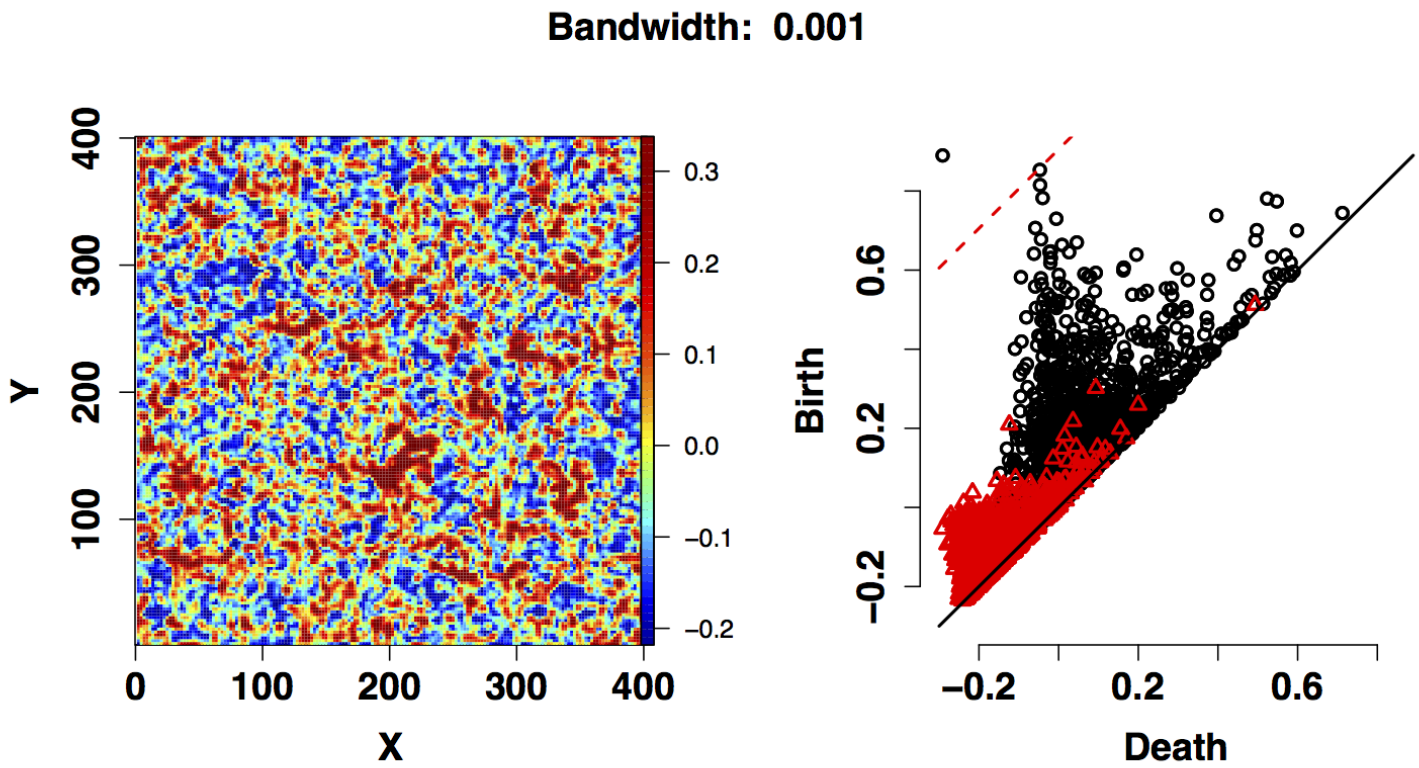


Code: `http://www.stat.cmu.edu/~flecci/research/index.html`

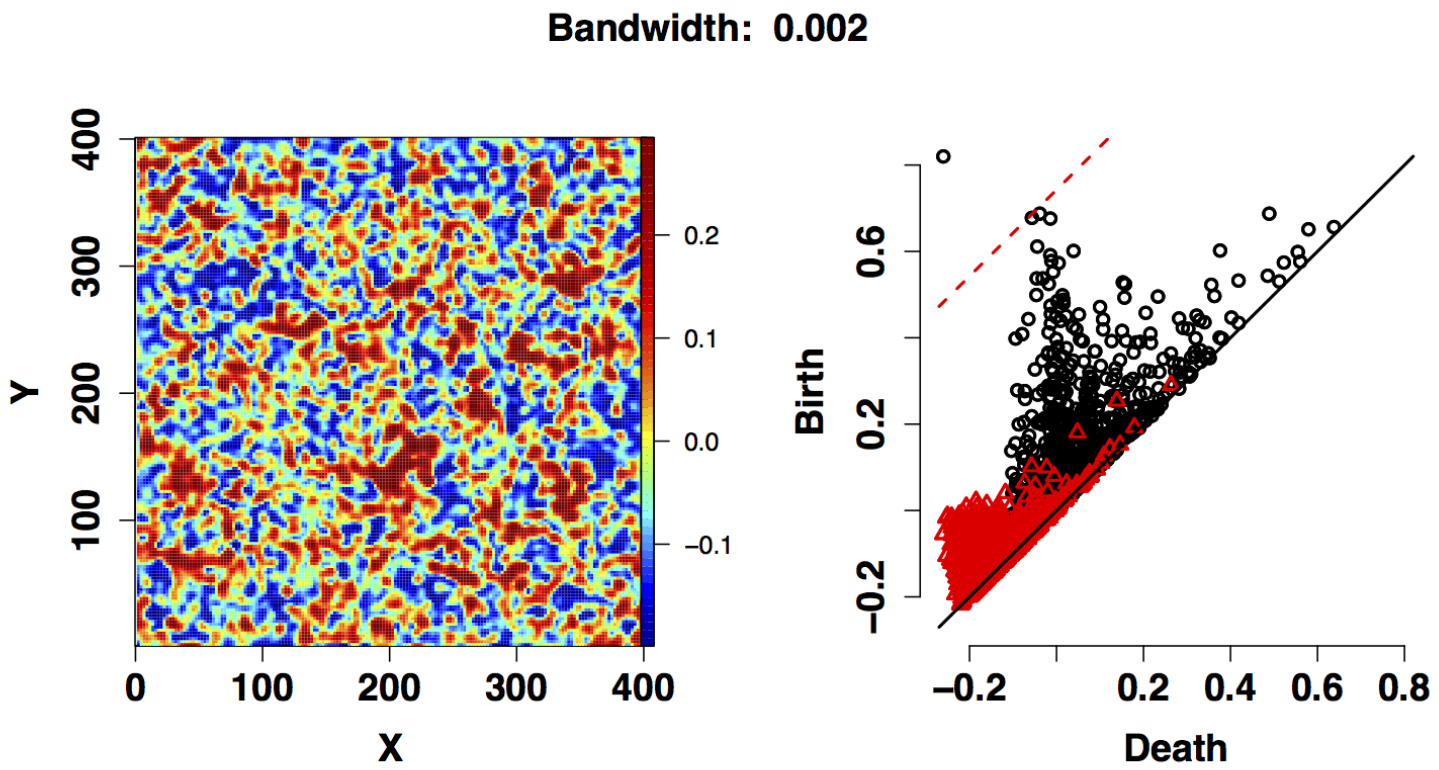
Bootstrap confidence interval



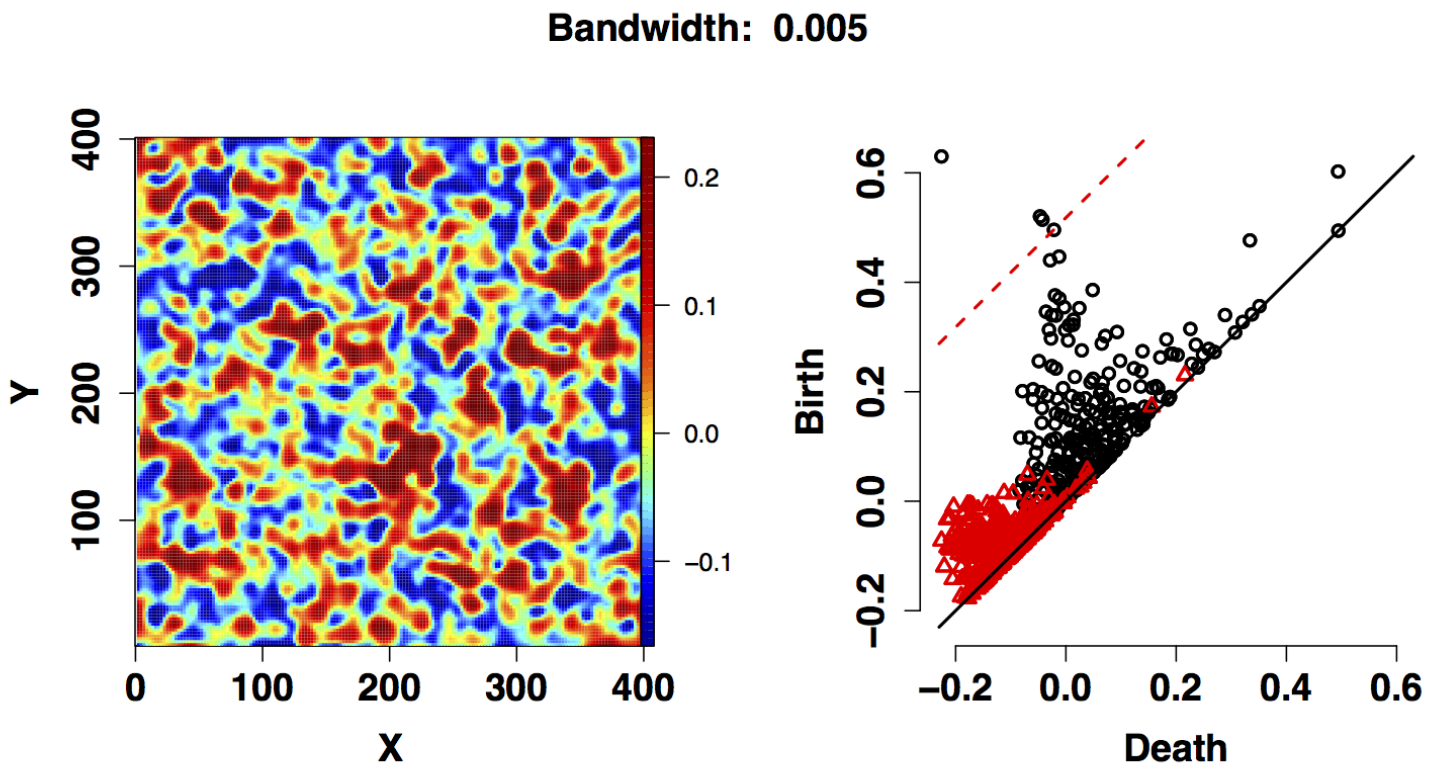
Bootstrap confidence interval



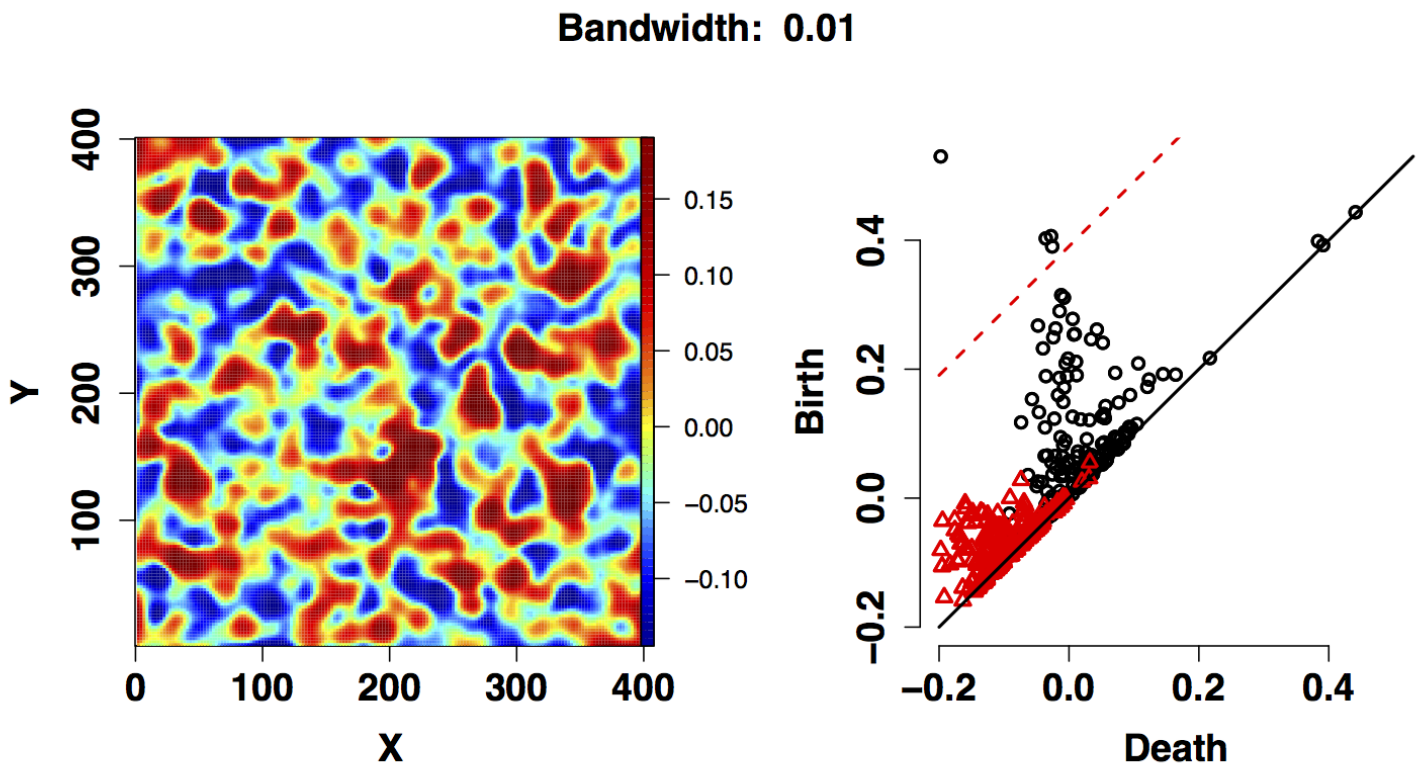
Bootstrap confidence interval



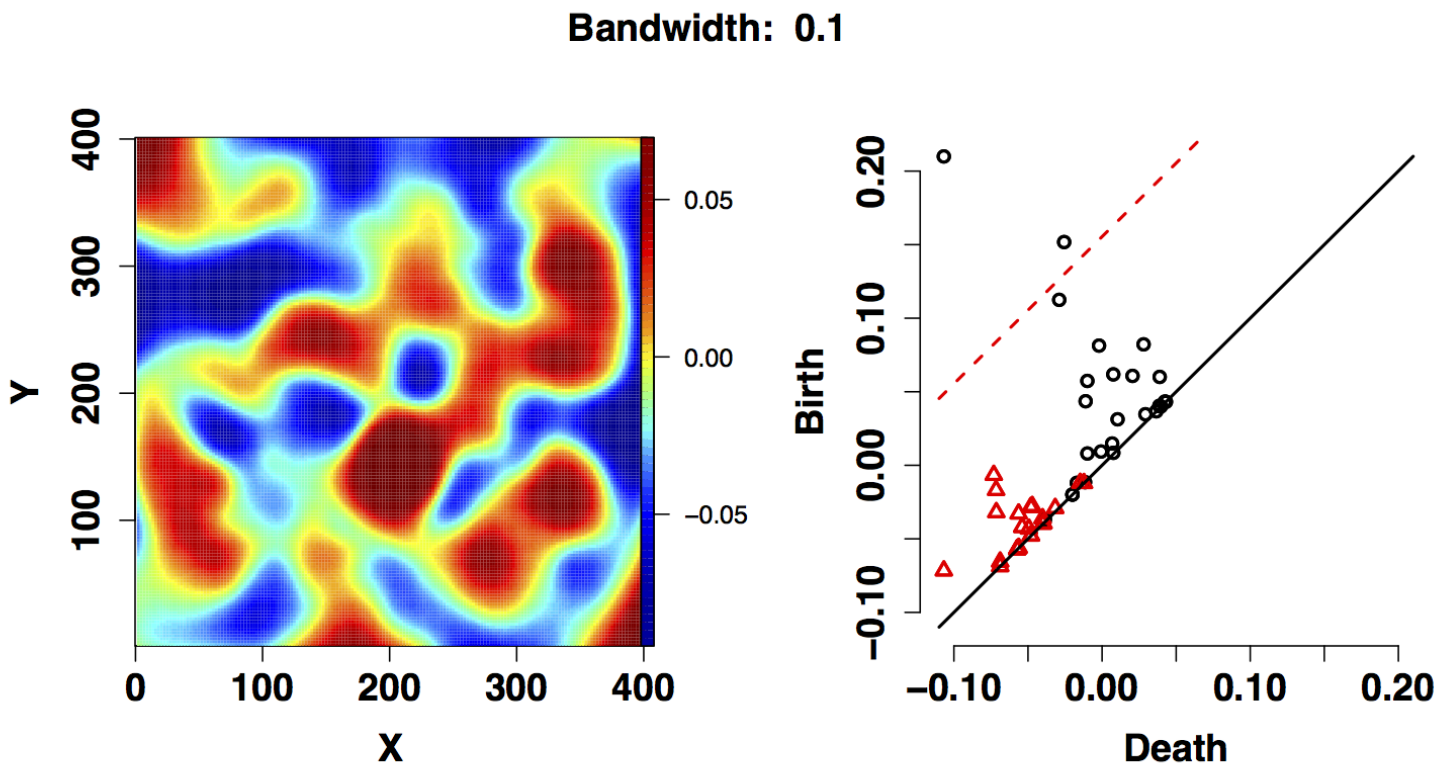
Bootstrap confidence interval



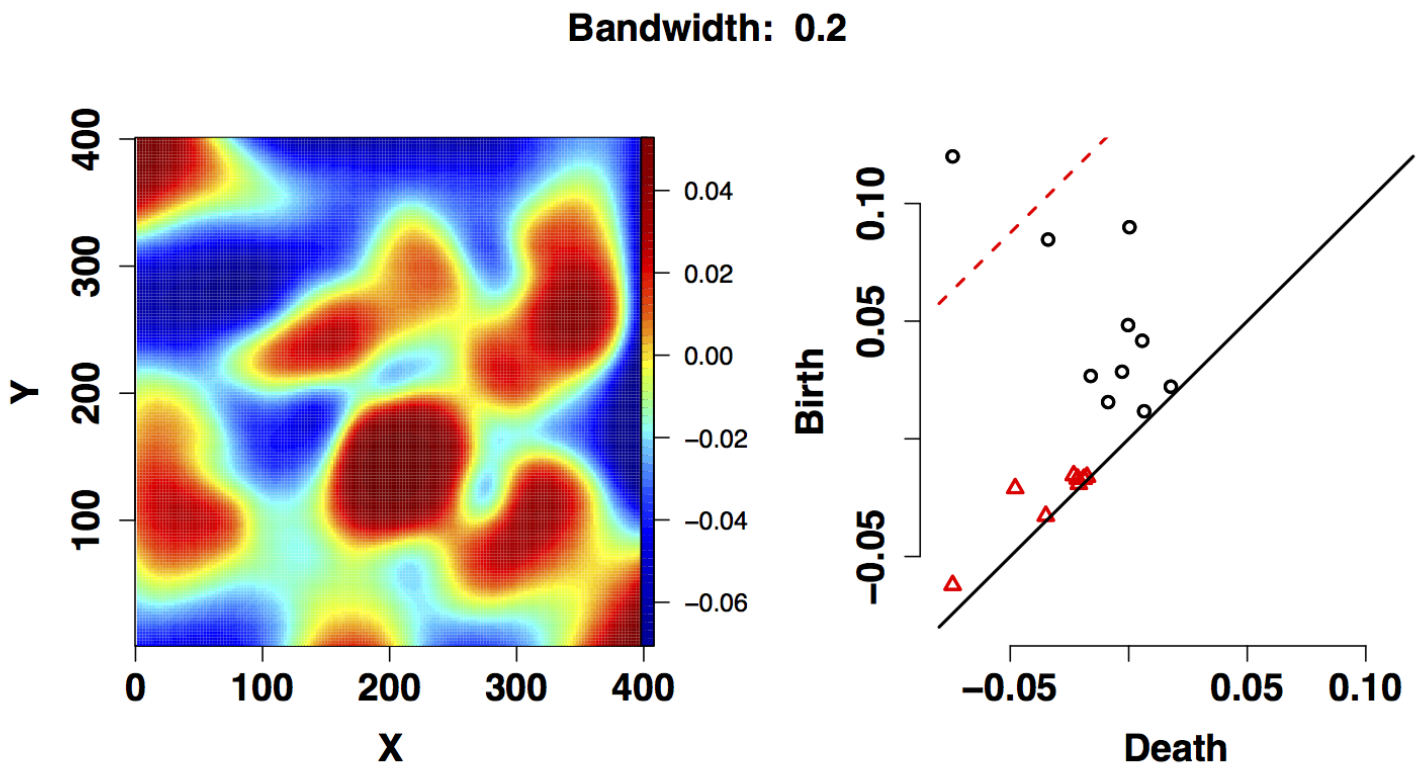
Bootstrap confidence interval



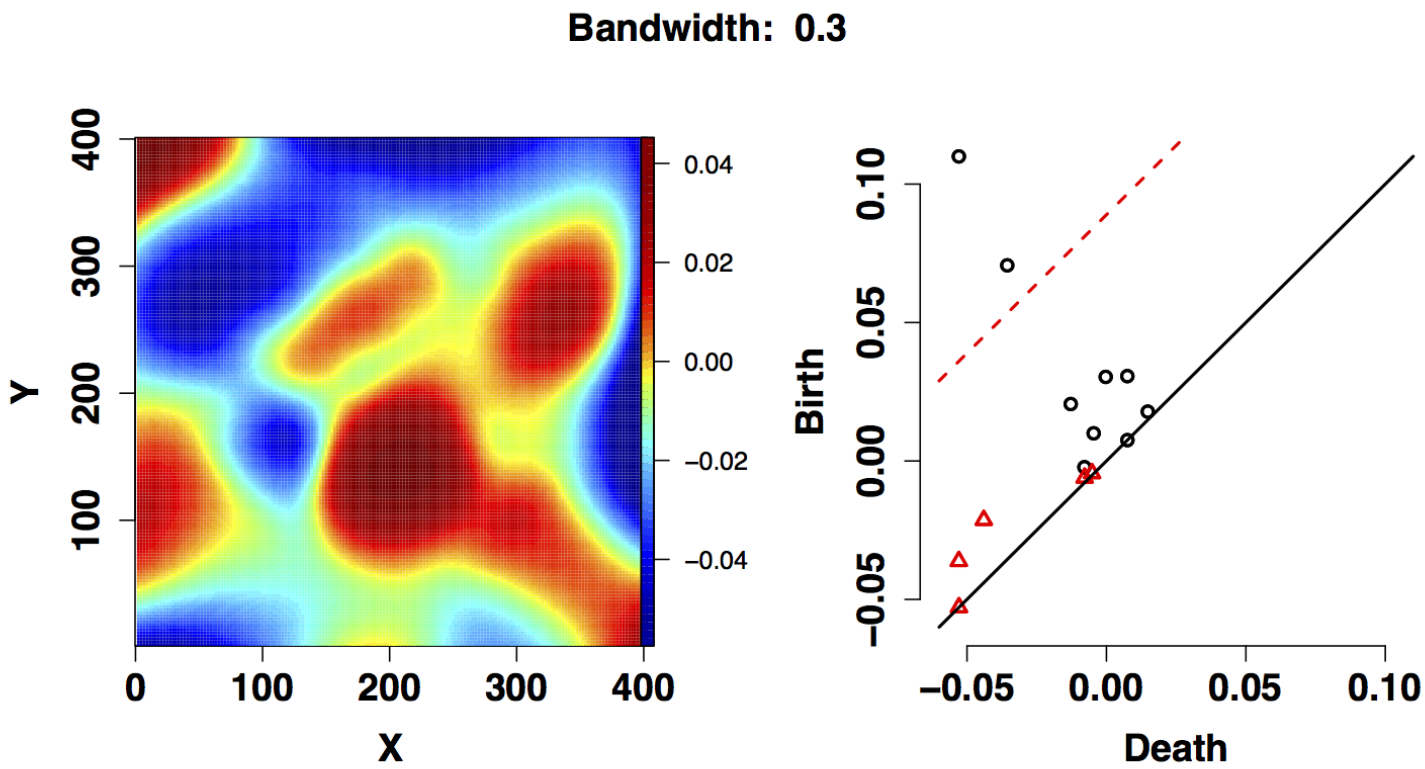
Bootstrap confidence interval



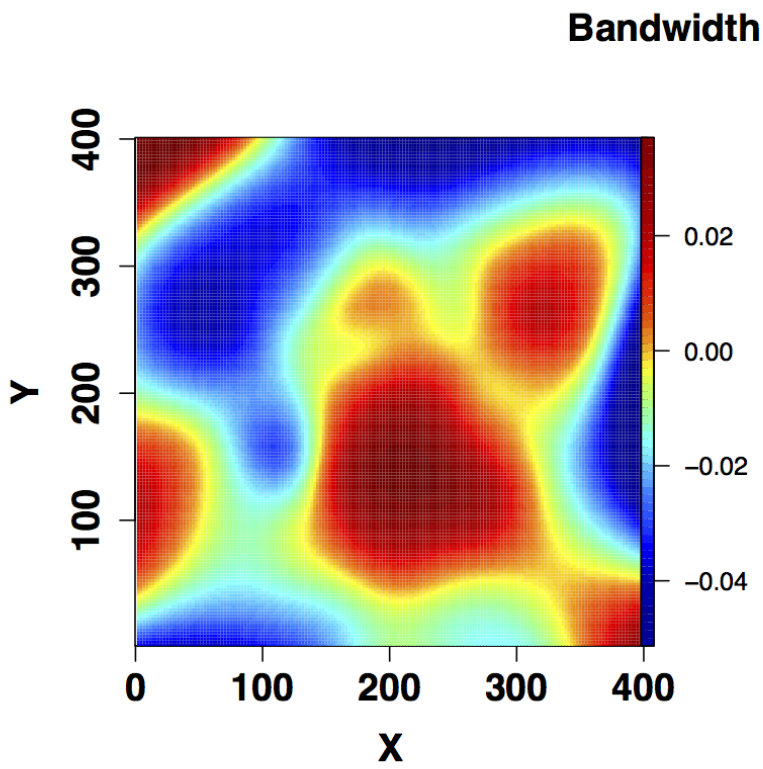
Bootstrap confidence interval

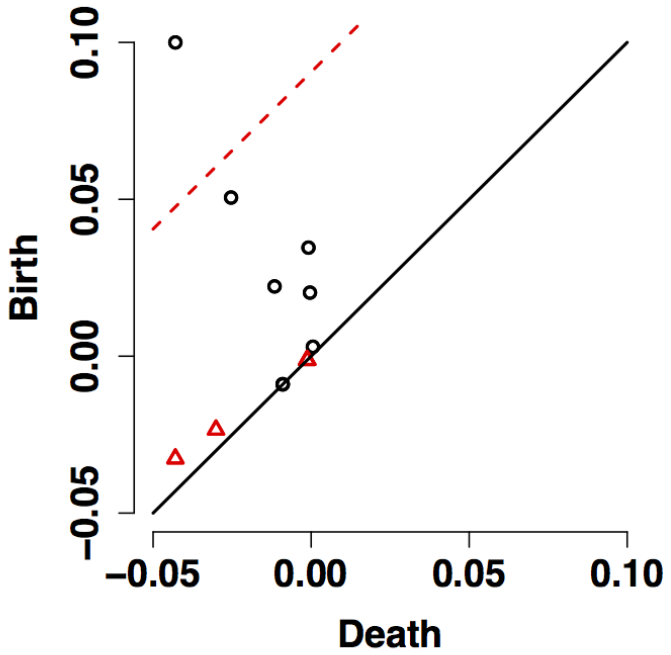


Bootstrap confidence interval



Bootstrap confidence interval





Persistence Landscapes

- Persistence diagrams can be summarized using landscapes

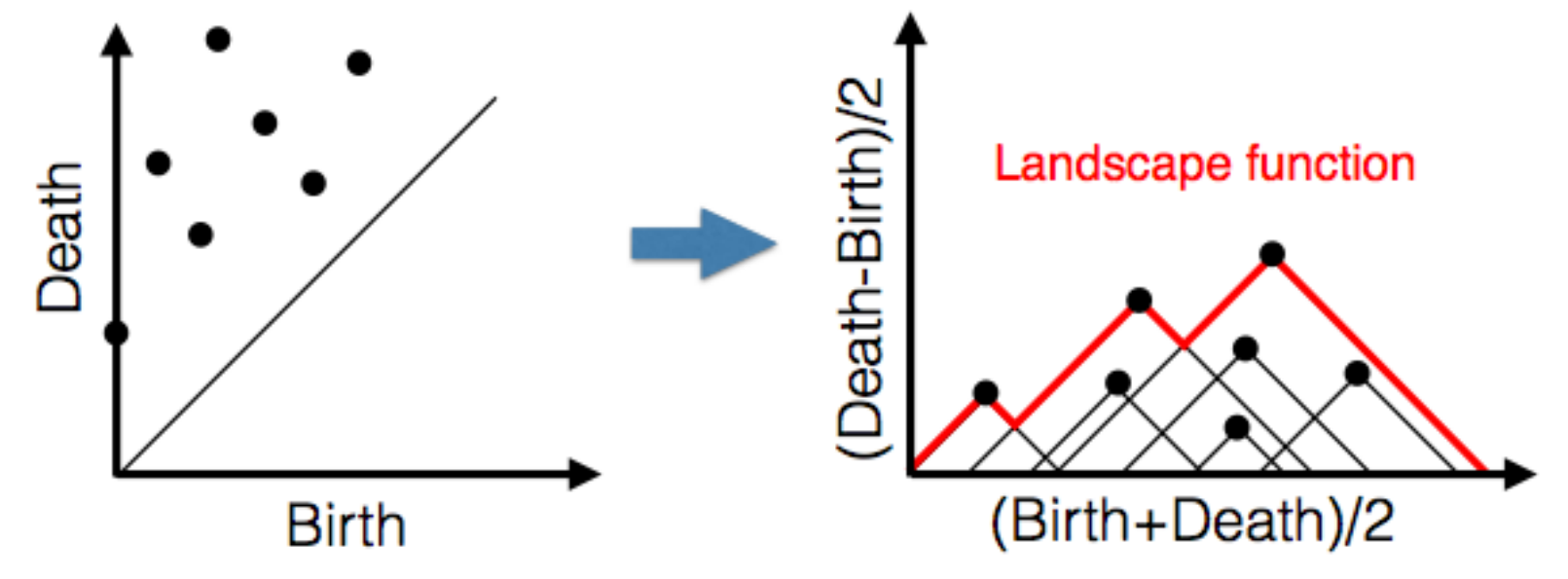


Figure from Fabrizio Lecci (CMU)

Confidence band for persistence landscapes

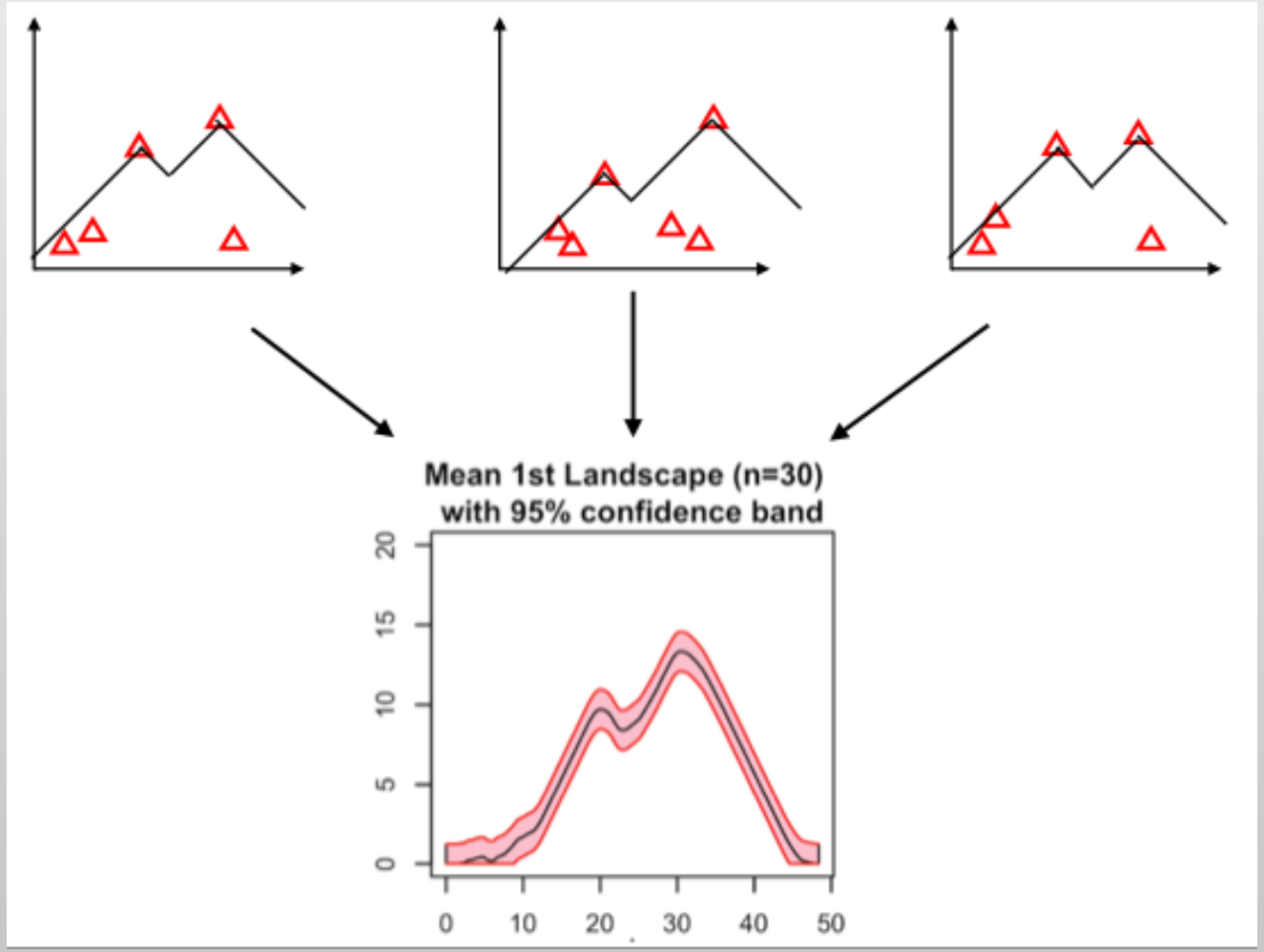
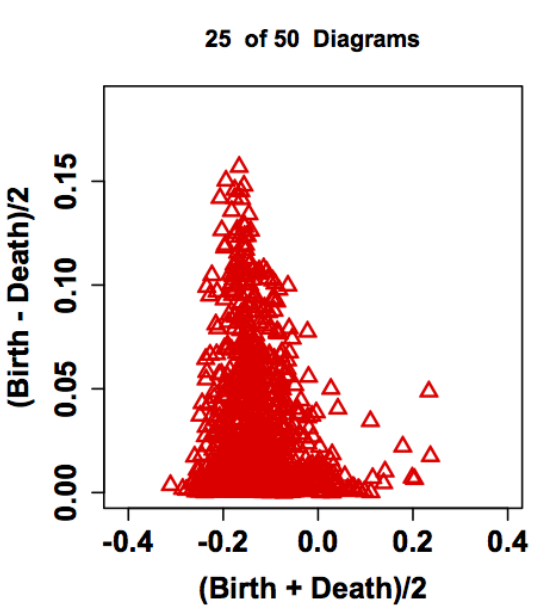
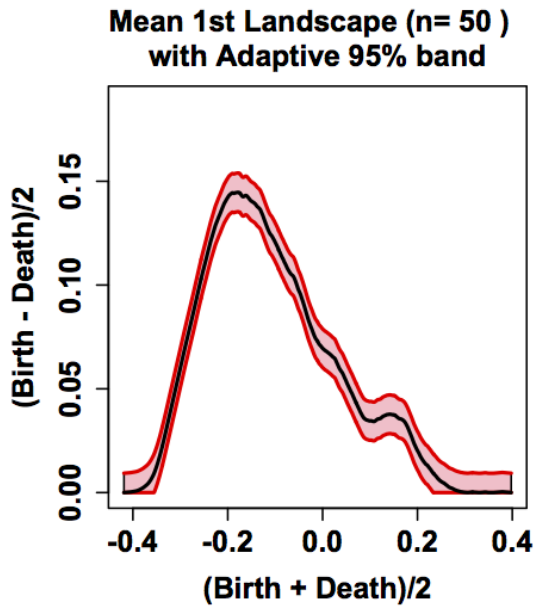
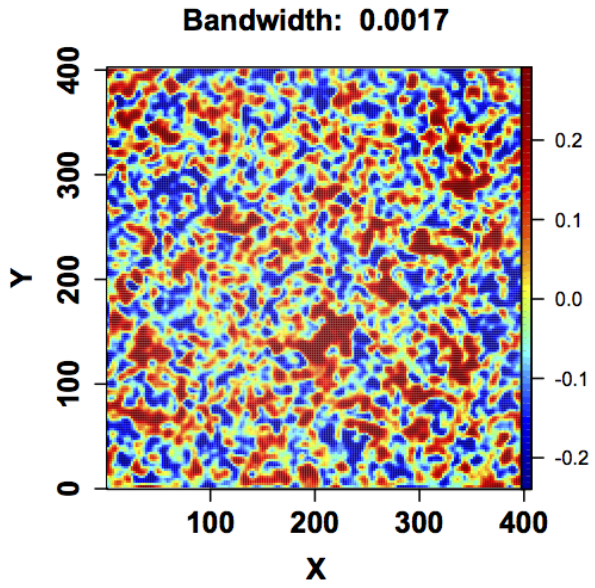


Figure from Fabrizio Lecci (CMU)

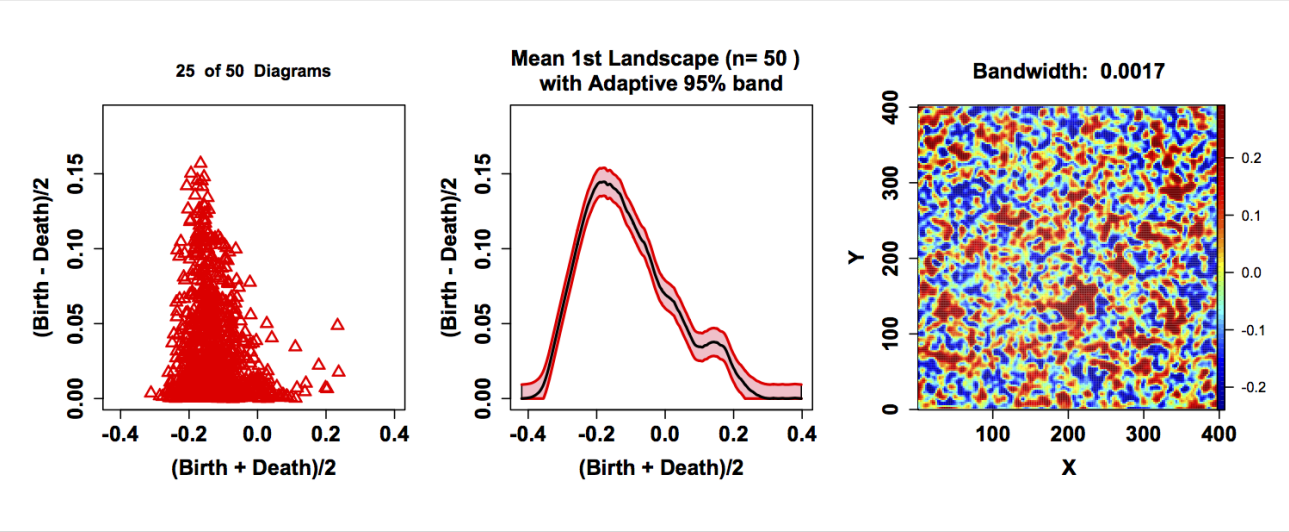
Persistent Landscapes: high-resolution simulation slice

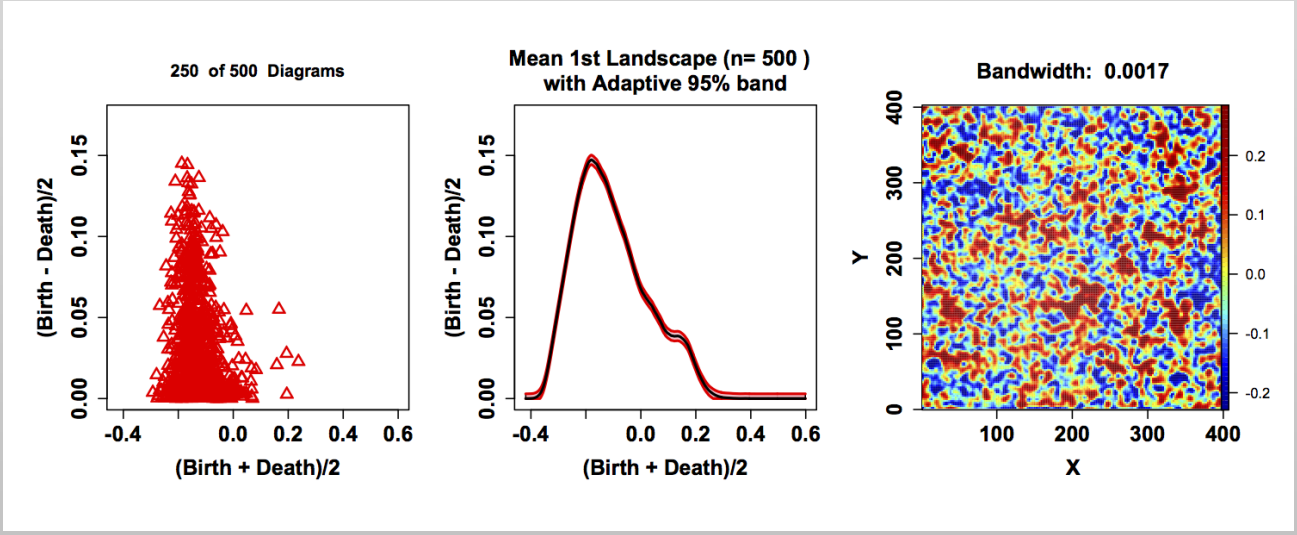




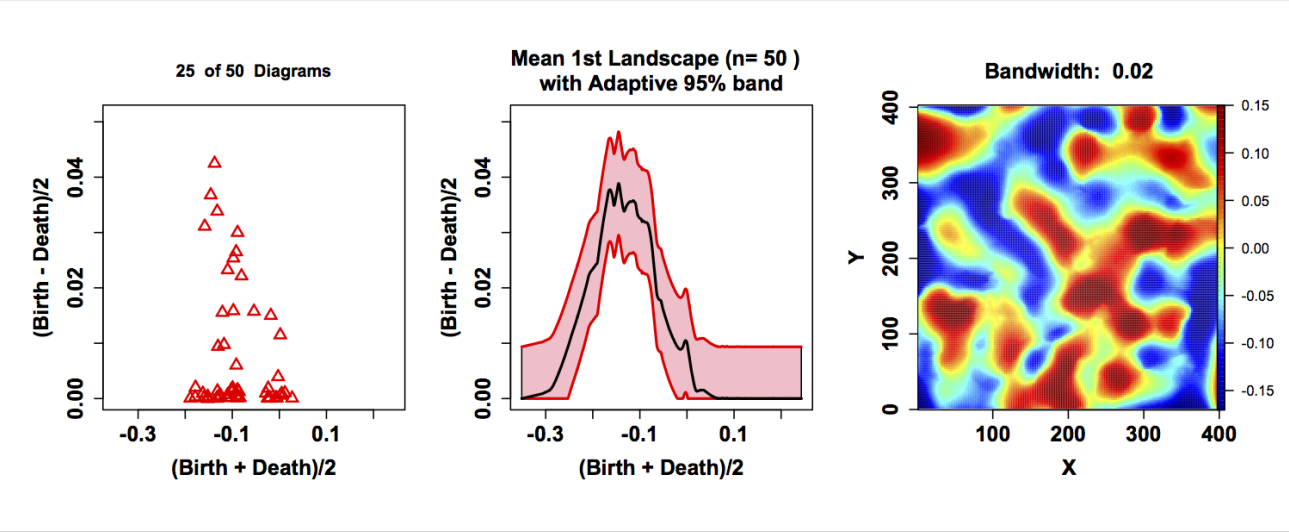


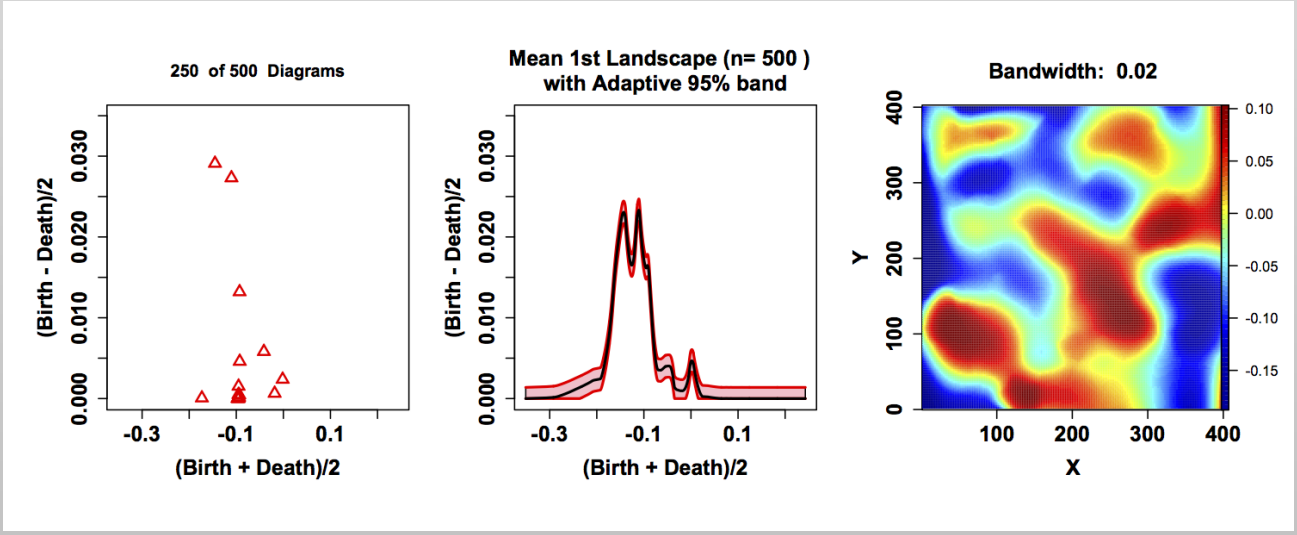
Persistent Landscapes: high-resolution simulation slice



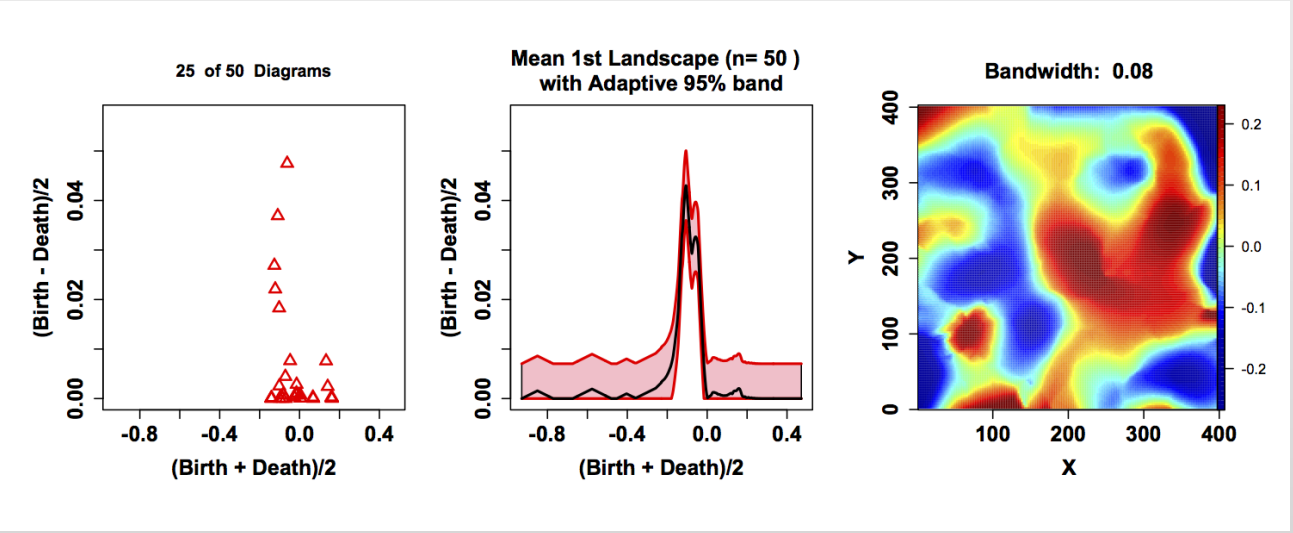


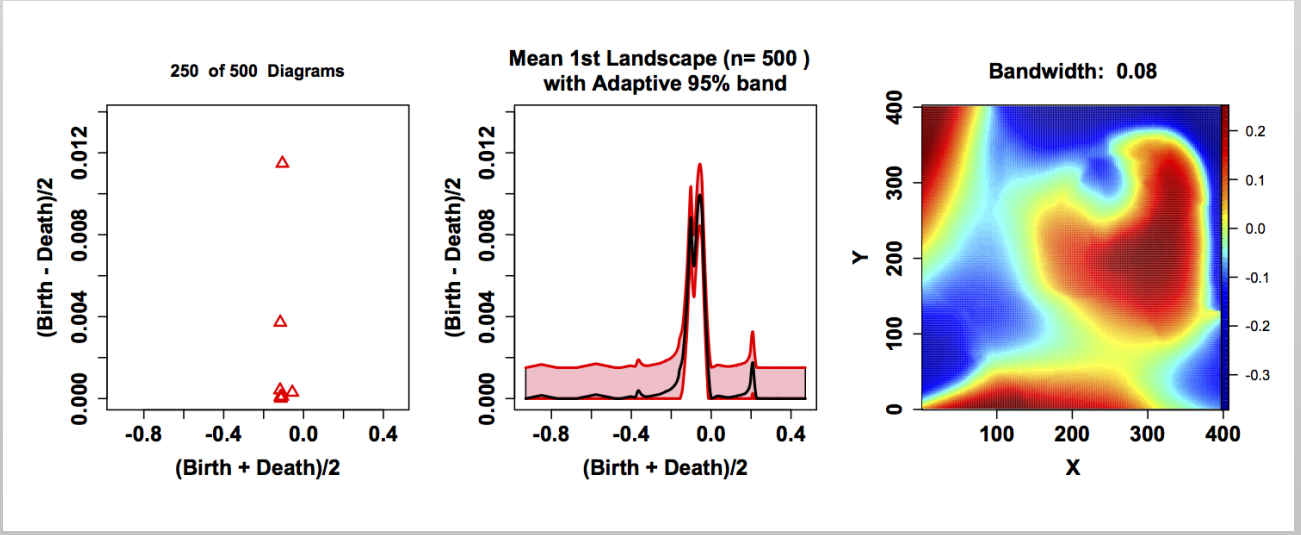
Persistent Landscapes: 1000 LOS simulation slice



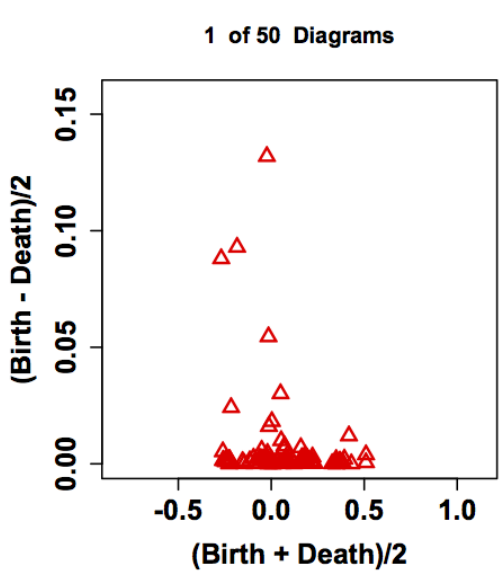


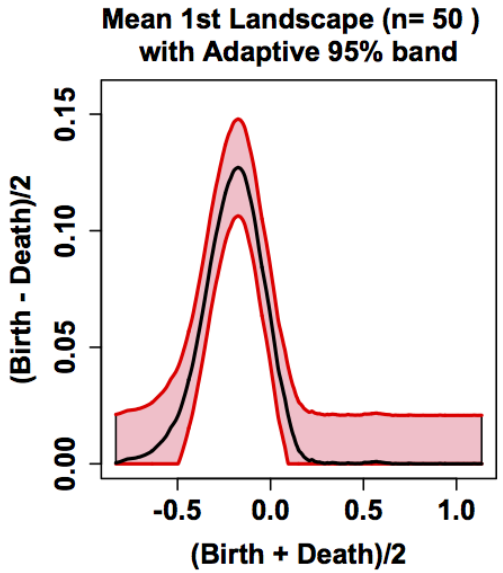
Persistent Landscapes: 200 LOS simulation slice

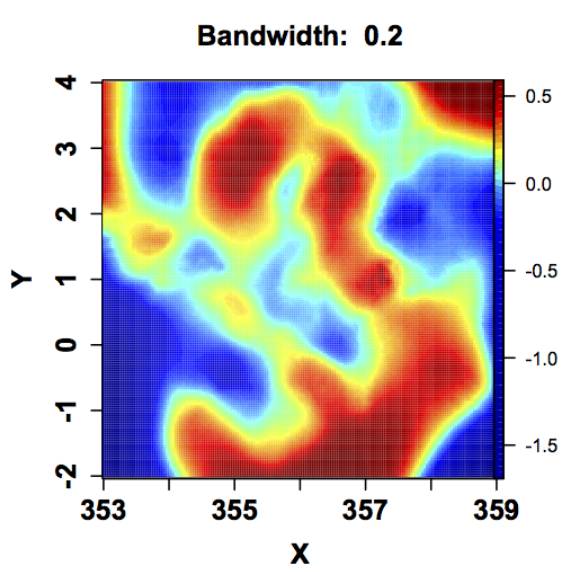




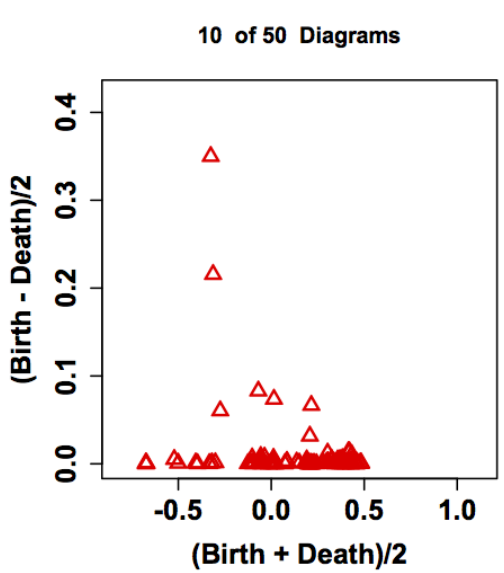
SDSS BOSS data slices - landscape bands

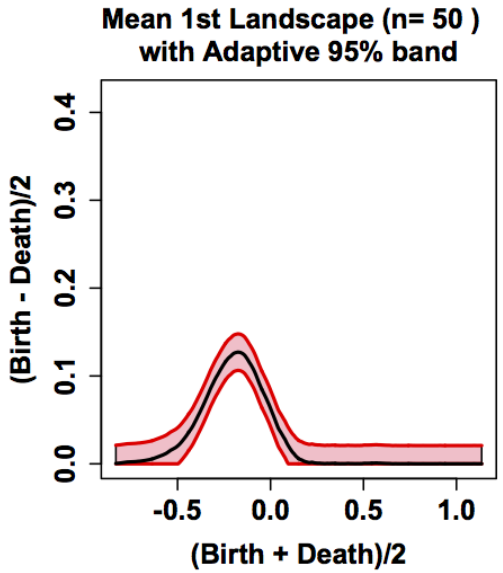


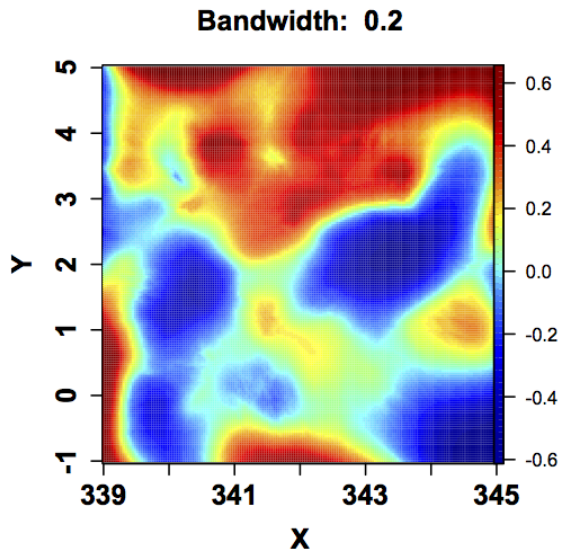




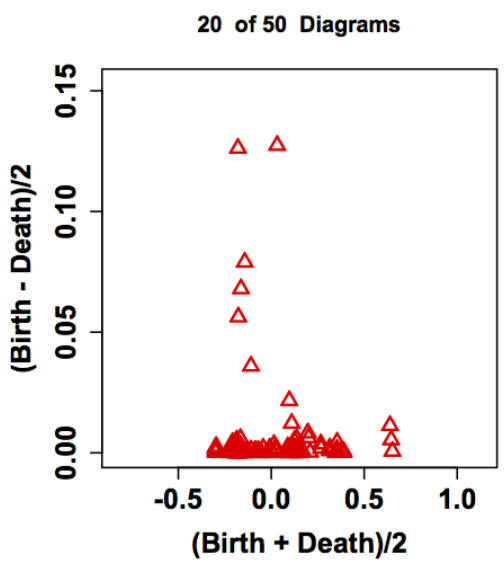
SDSS BOSS data slices - landscape bands

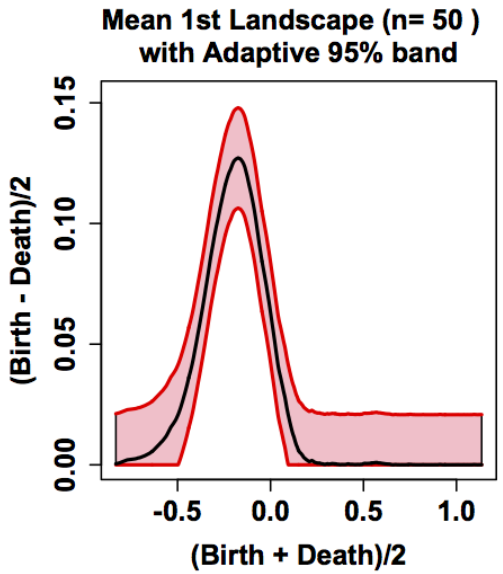


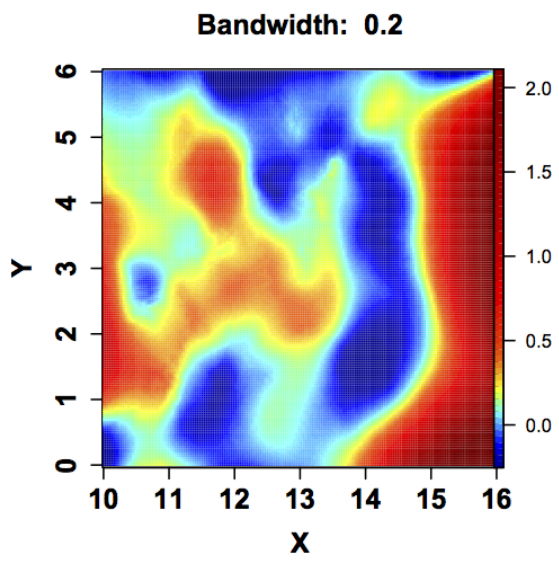




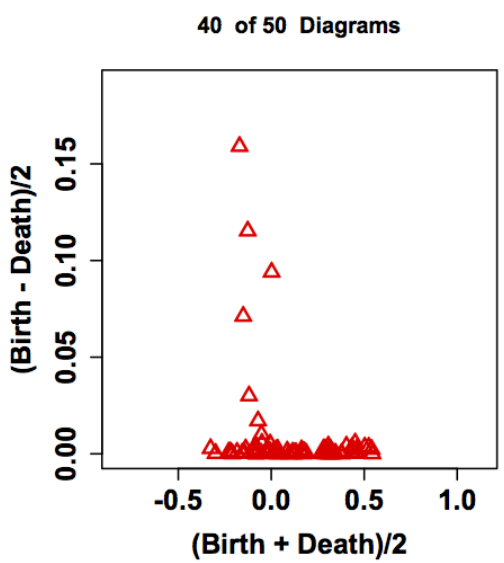
SDSS BOSS data slices - landscape bands

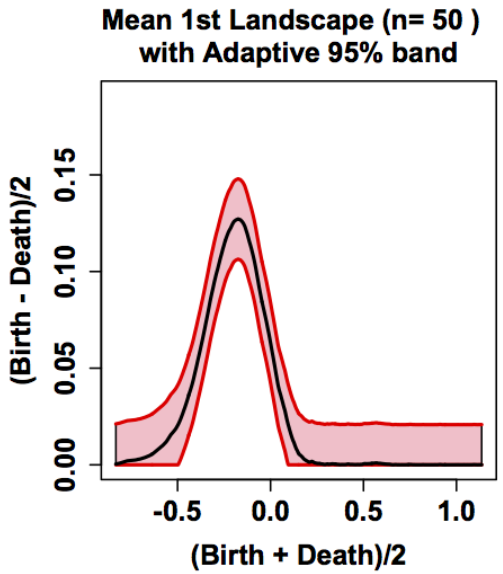


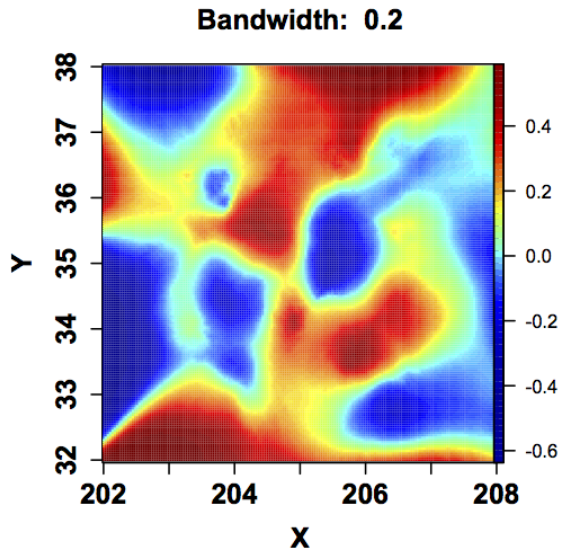




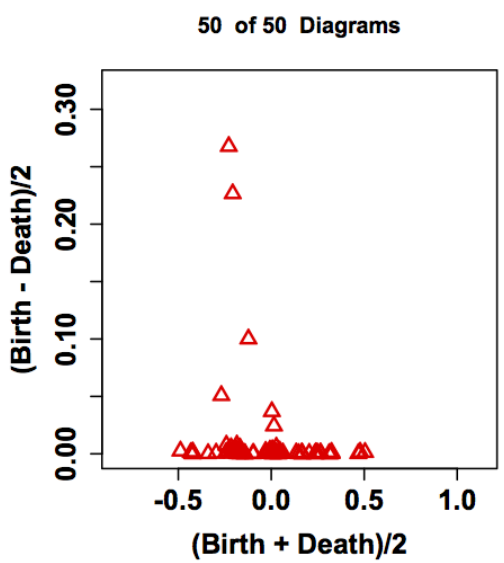
SDSS BOSS data slices - landscape bands

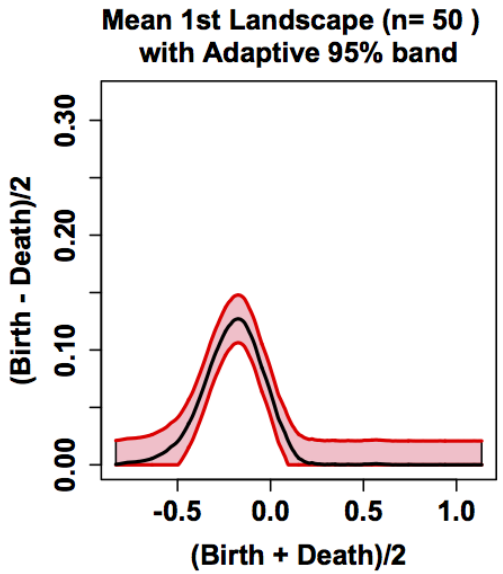


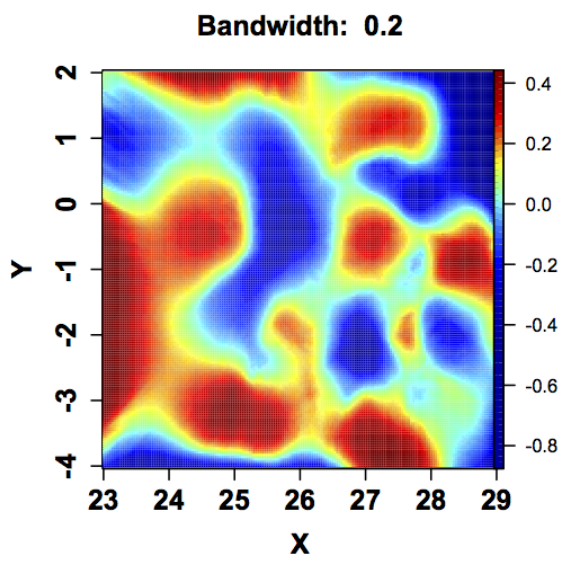




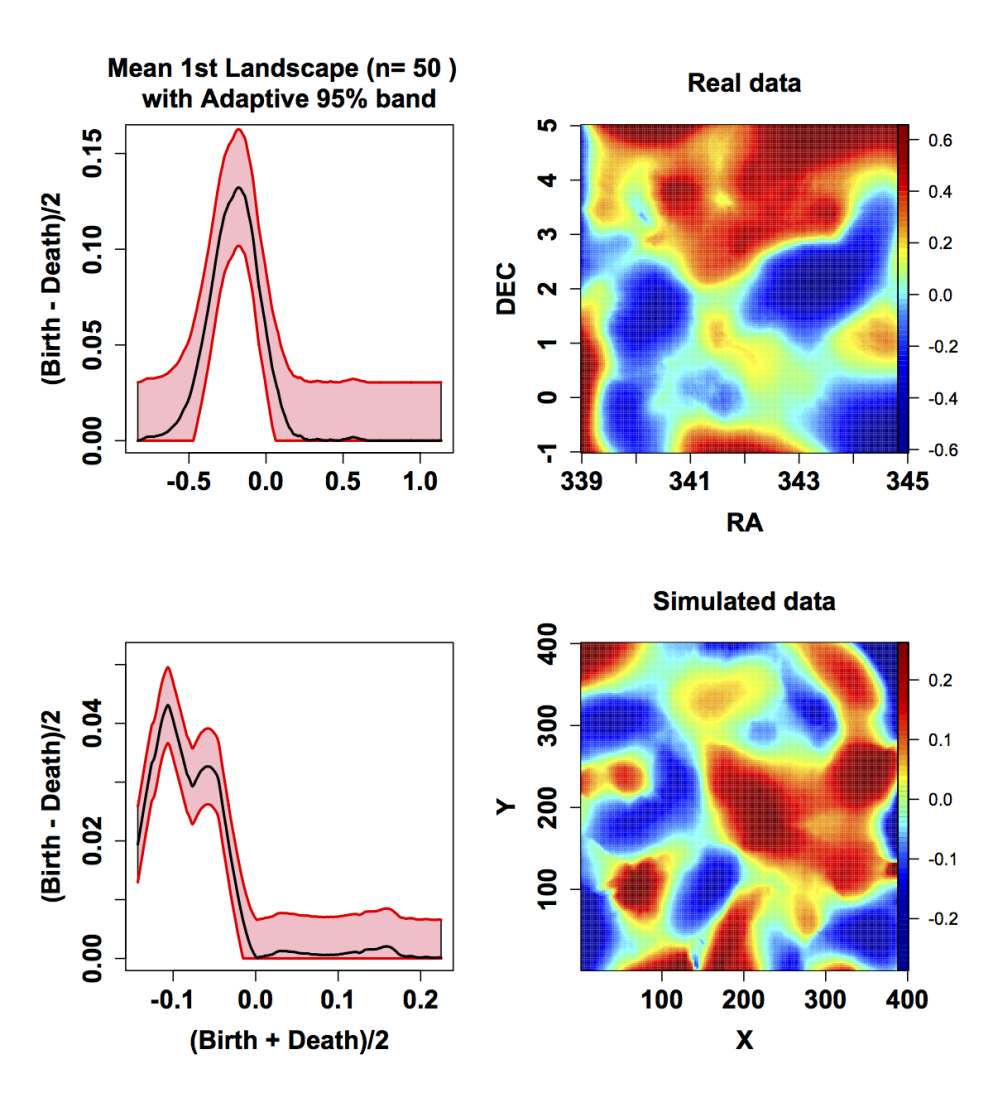
SDSS BOSS data slices - landscape bands



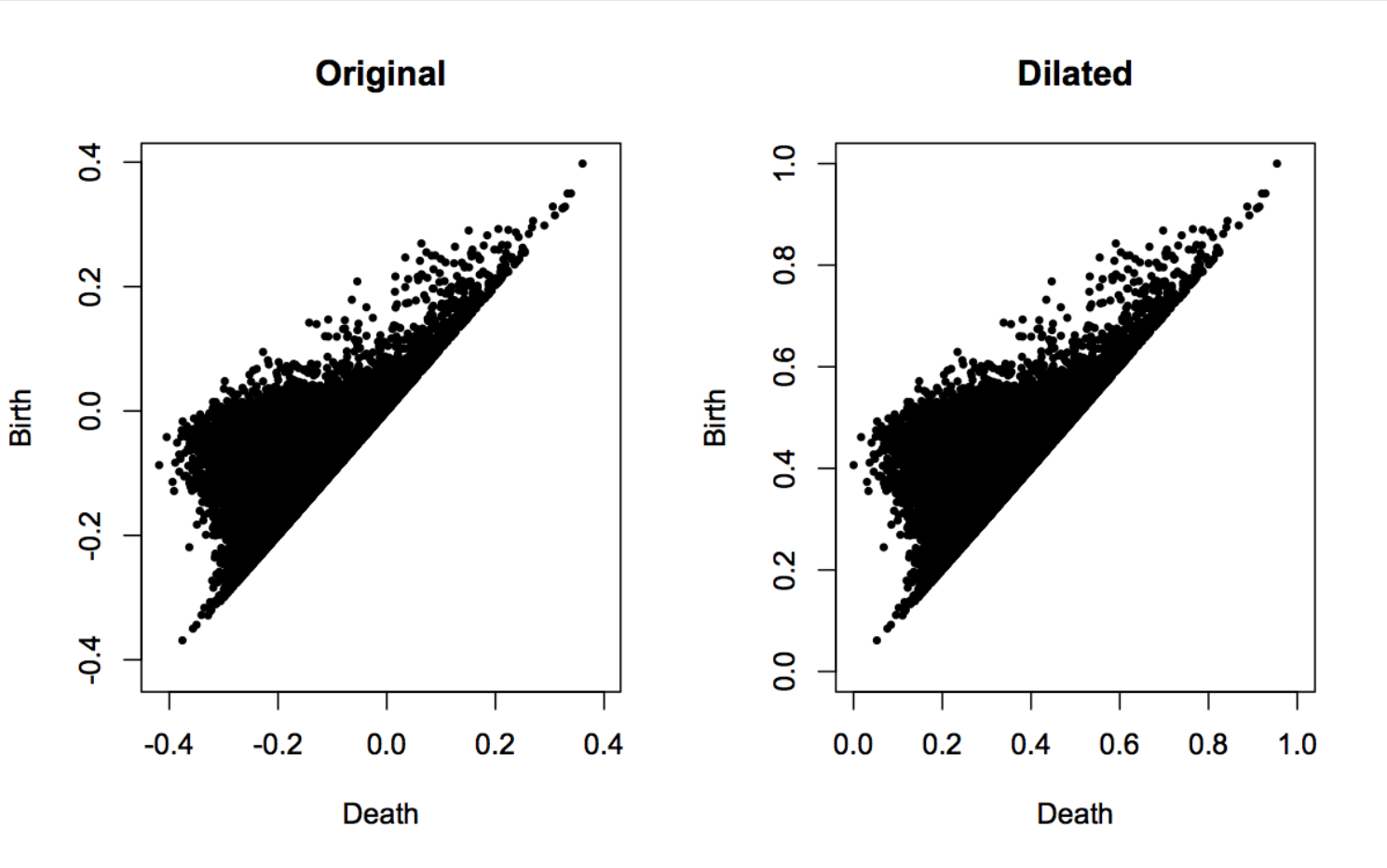




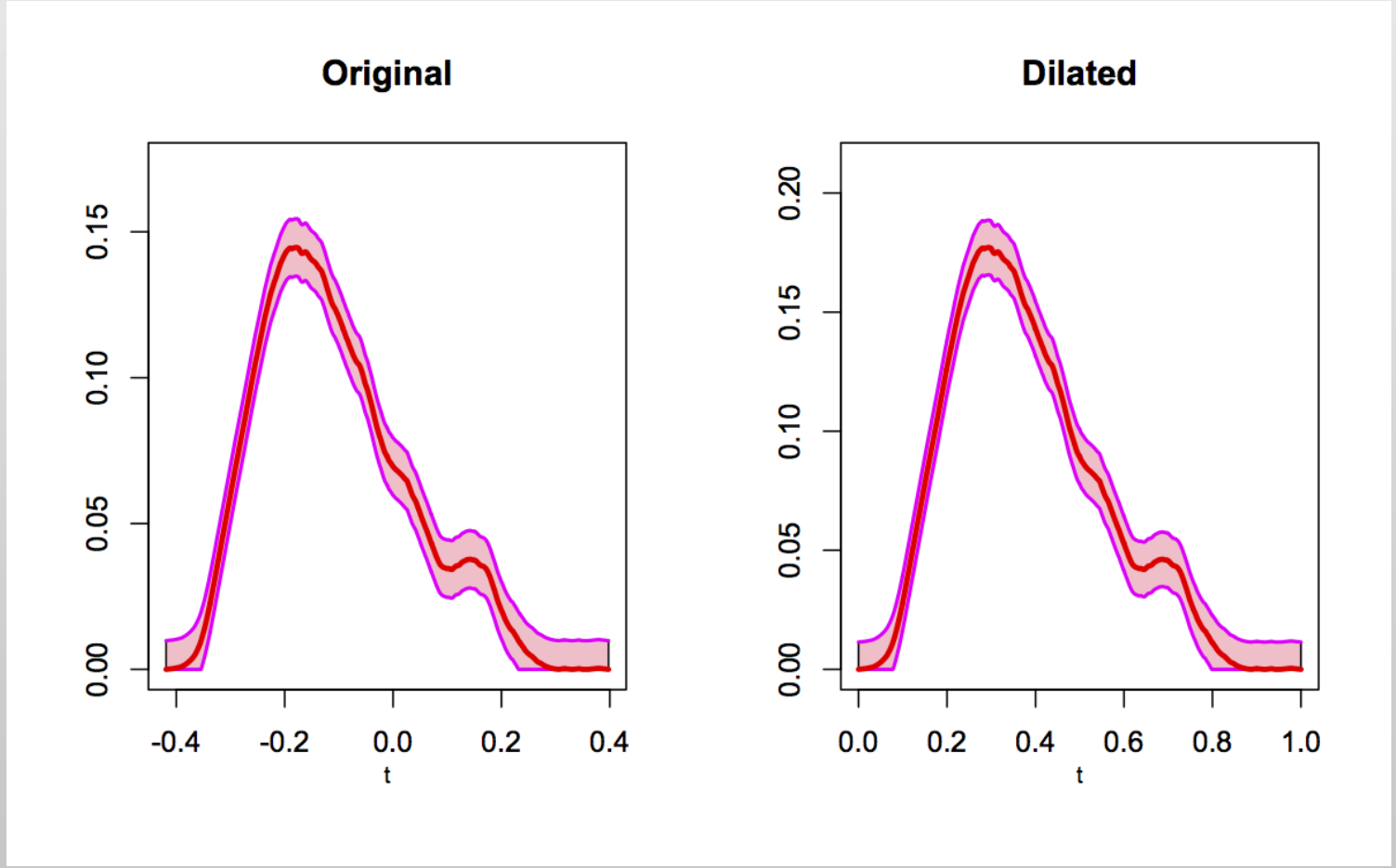
Compare simulated and real data



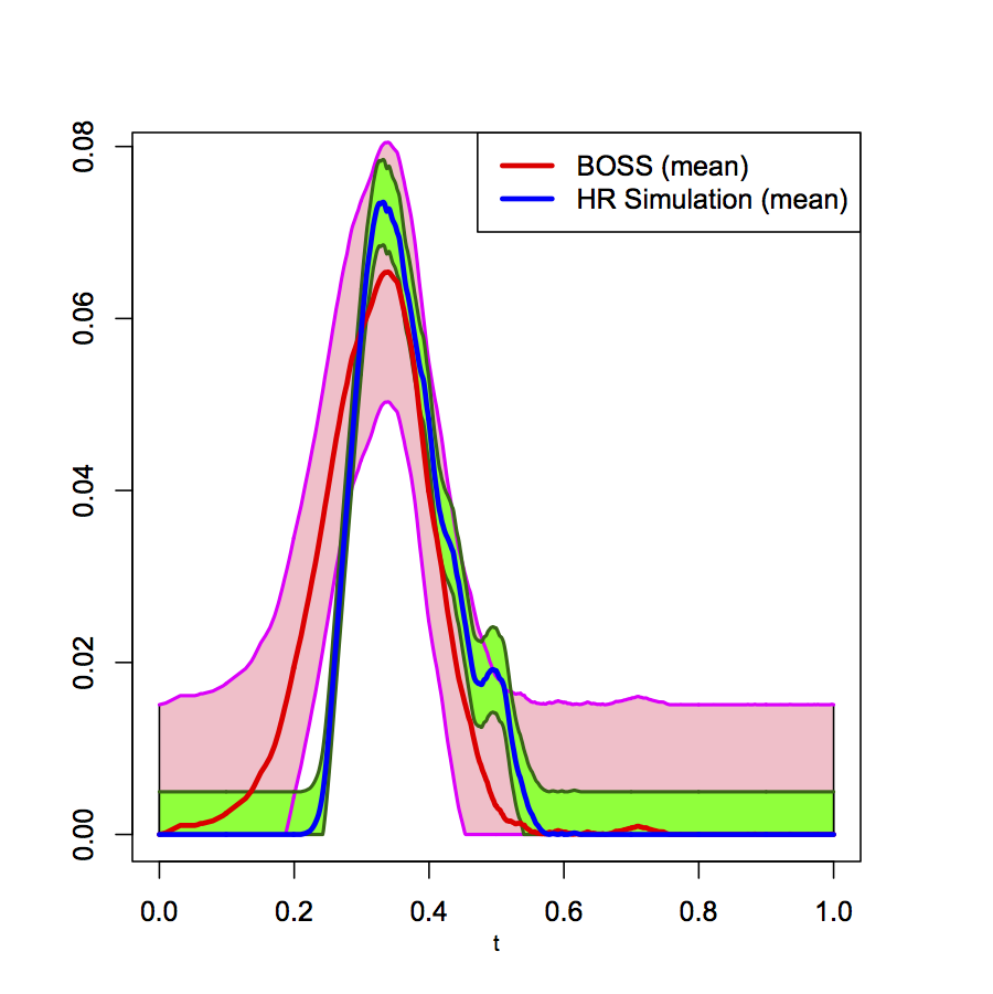
Rescaled persistence



Rescaled persistence



Compare simulated and real data



Concluding remarks

- Quasars provide a means for probing the intergalactic medium
- Persistent homology may be a useful tool in cosmology
  - $\longrightarrow$ Compare modeling fits: is the methodology capturing "important" features?
  - $\longrightarrow$ Do the topological properties of cosmological simulations match the topological properties of the observations?
  - $\longrightarrow$ Example: compare model fit of BOSS Lyman-alpha data with model fit of comological simulation output
- More investigation is needed to understand the statistical properties of persistent homology

Concluding remarks

- Quasars provide a means for probing the intergalactic medium
- Persistent homology may be a useful tool in cosmology
  - $\longrightarrow$ Compare modeling fits: is the methodology capturing "important" features?
  - $\longrightarrow$ Do the topological properties of cosmological simulations match the topological properties of the observations?
  - $\longrightarrow$ Example: compare model fit of BOSS Lyman-alpha data with model fit of comological simulation output
- More investigation is needed to understand the statistical properties of persistent homology
- Lots of really (scientifically and statistically) interesting problems in astronomy

Concluding remarks

- Quasars provide a means for probing the intergalactic medium
- Persistent homology may be a useful tool in cosmology
  - $\longrightarrow$ Compare modeling fits: is the methodology capturing "important" features?
  - $\longrightarrow$ Do the topological properties of cosmological simulations match the topological properties of the observations?
  - $\longrightarrow$ Example: compare model fit of BOSS Lyman-alpha data with model fit of comological simulation output
- More investigation is needed to understand the statistical properties of persistent homology
- Lots of really (scientifically and statistically) interesting problems in astronomy

THANK YOU!!!