Cross-correlating GW Events and Galaxy Catalogs

Federico Berlfein

Topics

Idea

- Method of **cross-correlating** GW events and galaxy catalogs

- **Cosmology** comes into play:
  - Galaxies have **redshifts**
  - GW events have **luminosity distances**
  - We need a **common coordinate system**
  - **Transform** the redshifts and luminosity distances into **comoving distances** using a given **cosmological model**

- **Infer H_0** from cosmology that produces **maximum cross-correlation**!

Comparison of Distance Measures 0 < z < 10,000
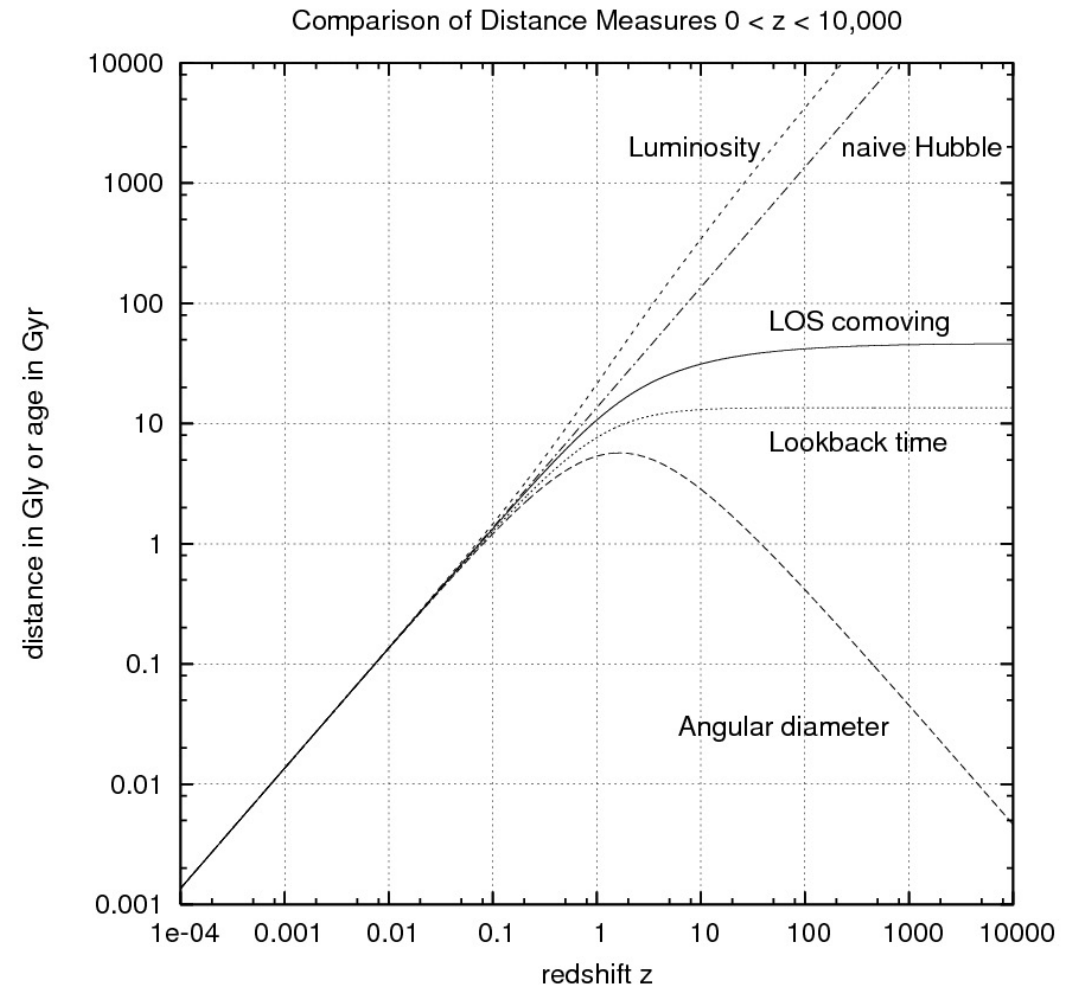
Computing Cross-Correlations using TreeCorr

- Galaxy weight: w

- GW pixel weight: ω, and probability: k

- Look at different **separation bins**: r

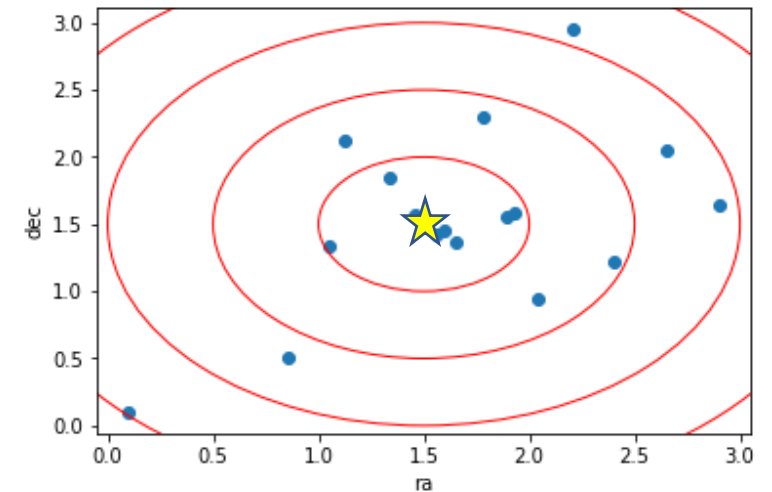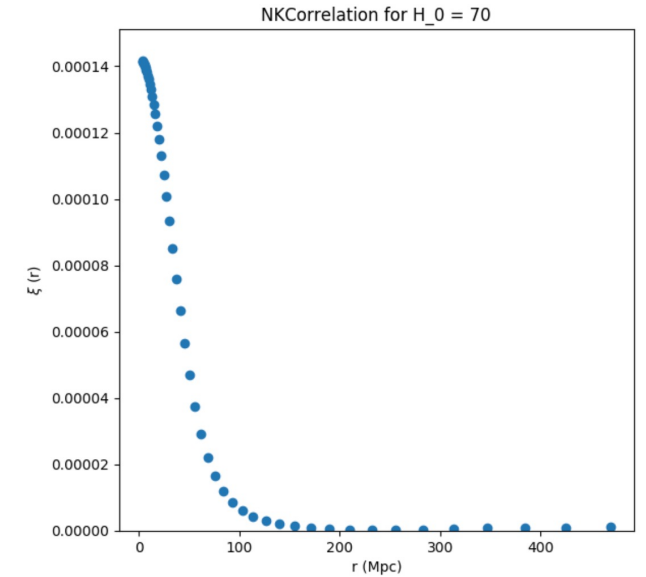- For **N galaxies** and **M pixels**

$$\chi(r) = \sum_{i=1}^{N}\sum_{j=1}^{M} \delta_{ij}^{r} w_i \omega_j k_j, \qquad W(r) = \sum_{i=1}^{N}\sum_{j=1}^{M} \delta_{ij}^{r} w_i \omega_j$$

$\xi(r) = \frac{\chi(r)}{W(r)}$ **("finished"** correlation)

$$\delta_{ij}^{r} = \begin{cases} 1, & r_{ij} < r \\ 0, & else \end{cases}$$

- Compute cross-correlation for **random** galaxy catalog and have a **corrected** $\xi(r)$:

$$\xi(r)_{corrected} = \xi(r) - \xi(r)_{Randoms}$$

Understanding the Inputs

GW Events

distance: 241±26 Mpc

- **Pixelated sky map** where each pixel contains:
  - **Probability** of hosting event
  - **Luminosity distance** measurement and **uncertainty** (can be up to 20-30%)
- TreeCorr **does not allow** pixels or uncertainties as inputs
- How do we represent the GW events in TreeCorr?
  - Pixels: RA and DEC of pixel center
  - Luminosity distances are not that simple...

50% area: 4 deg²
90% area: 19 deg²
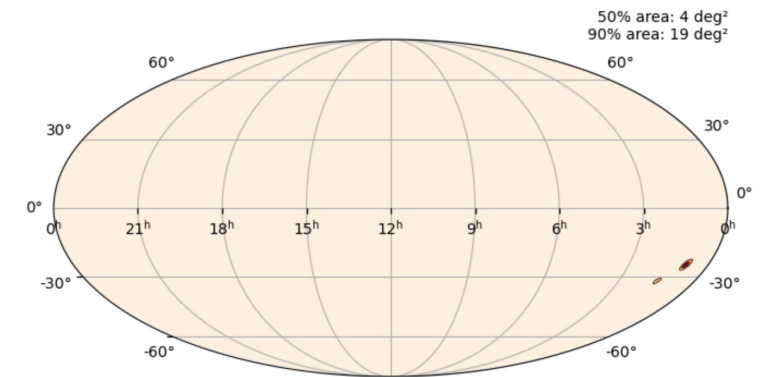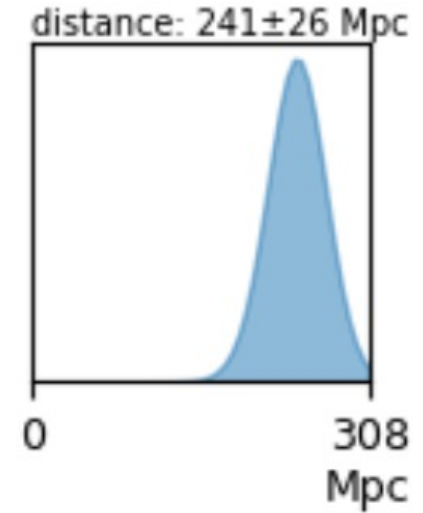
Luminosity Distance Uncertainty

- Use **weighted instances** of each pixel

- 100 instances of the same pixel with **different luminosity distances**. Each instance is **weighted** using its respective **pdf value**.

- **Cross correlate** 100 weighted instances of each event with the galaxy catalog

GW Luminosity Distance Posterior

Mean Value = 539.2 Mpc
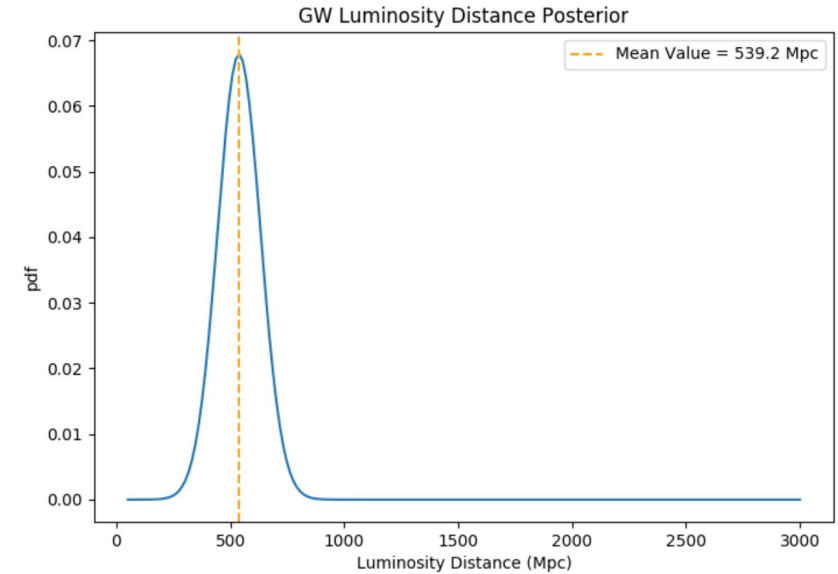
Histogram blob of galaxies and GW-event distances (blue lines): H_0 =70.0 km / (Mpc s)
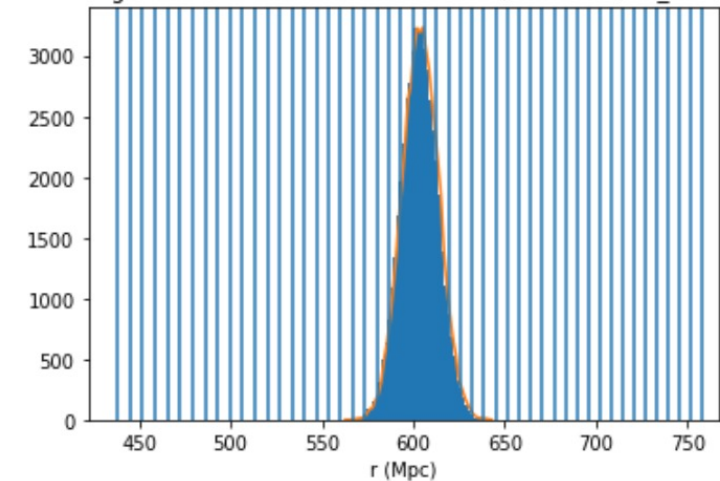
Galaxy Catalogs

- **Straight forward**

- Already provide galaxy sky localizations (RA and DEC) with **high precision**

- Redshifts: **smaller uncertainties**

- **Not considered** in this analysis, although they should eventually

- Luminosity distance errors dominate
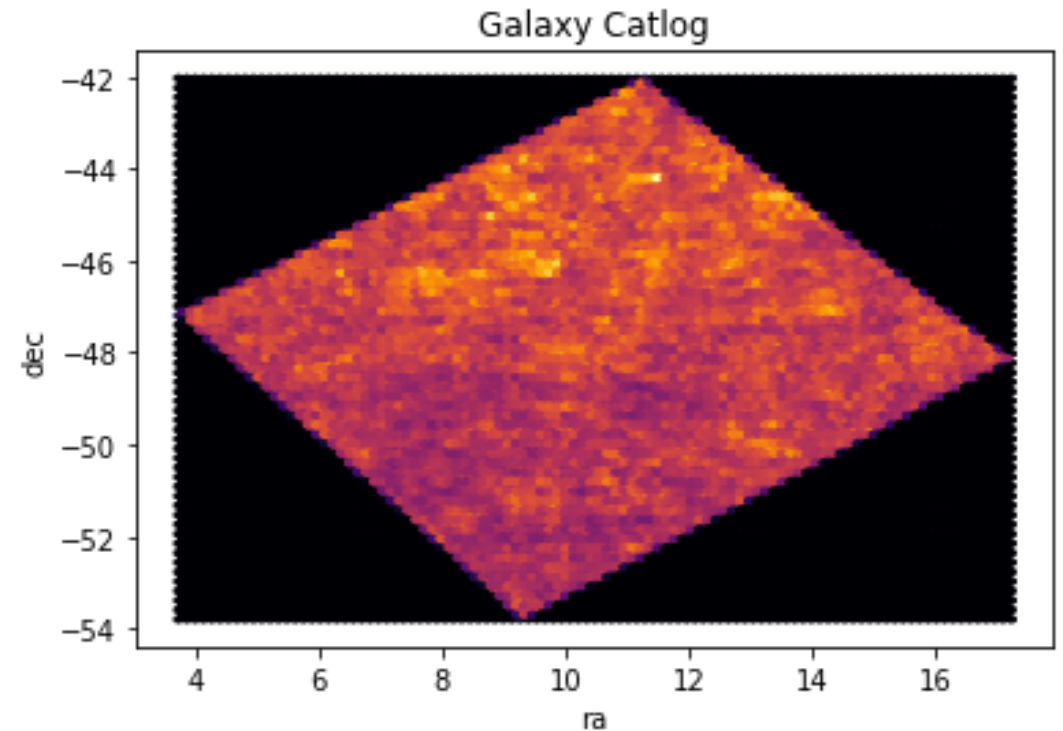
Galaxy Catlog

Cross-Correlation Pipeline

Settings

1 Paths to files or directories to GW events, galaxy catalogs, and output directory

2 Galaxy catalog, GW column names, and ordering of GW pixels

3 Range and step of H_0 values to analyze

4 Minimum percentage of pixels containing galaxies

5 Whether or not to calculate randoms or skip already analyze events

6 TreeCorr Settings

7 Parallelization Settings

```
# gw_dir is the directory of all GW events to analyze. This can also be a single file
gw_dir : 'des40a/mockTest/25GW_mockEvents/'
# cat_dir is the directory where correspodning        catalogs are. This can be a single file
cat_dir : 'des40a/mockTest/Galaxy_Catalogs/'
#directory for all output folders, files, and plots
outdir : 'des40a/mockTest/Final_Mock_NKObjects/'

# column names for galaxy catalog
ra_col_name : 'RA'
dec_col_name : 'DEC'
redshift_col_name : 'Z'

# column names for GW event file
prob_col_name : 'PROB'
distmu_col_name : 'DISTMU'
distsigma_col_name : 'DISTSIGMA'
distnorm_col_name : 'DISTNORM'

# nest for GW events (True or False)
nest : False


# set H0 values to analyze. h0_step represents size of step between H0 values
h0_min : 40
h0_max : 100
h0_step : 2

# min_completeness for GW event to be cross-correlated
min_completeness : 0.8
# do_randoms sets if the random cross-correlation for each event are done. If skip_done = True, analysis will be skipped
# for events in set output directory that have already been analyzed
do_randoms : True
skip_done : False

# TreeCorr settings
min_sep : 3
max_sep : 500
nbins : 50

#Parralelization settings. n_jobs sets the number of parallelized jobs. n_threads sets number of CPU's to use per job
n_jobs: 8
n_threads : 3
```

Creating Directories

Pre-Processing

- GW Event

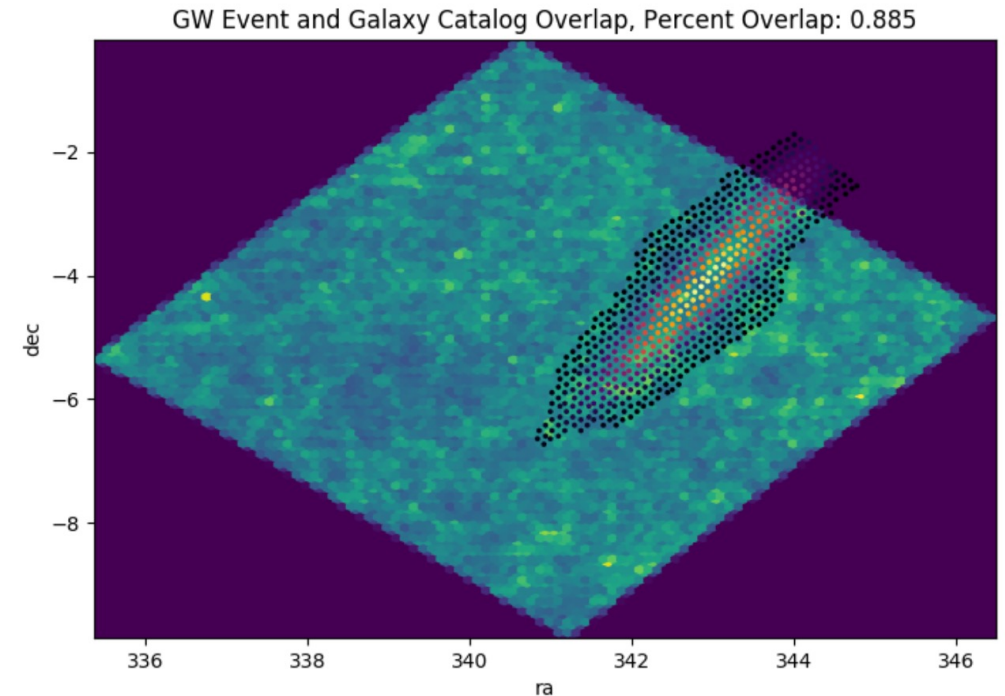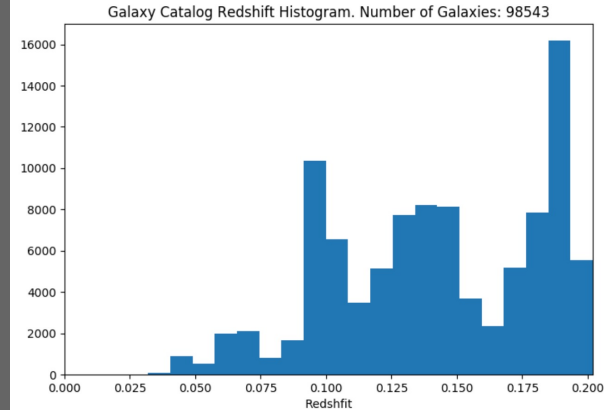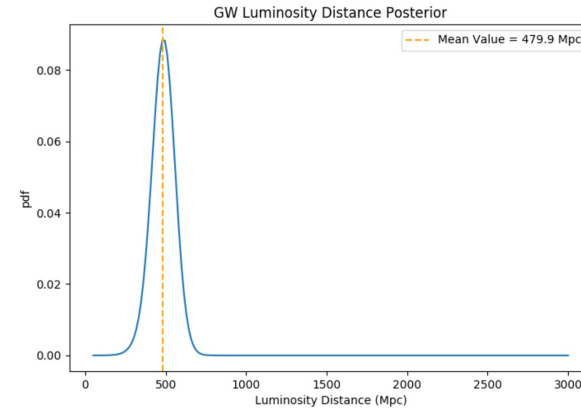 - **Eliminates pixels** :
    - Zero probability
    - Luminosity distance uncertainties higher than 50%
  - **Weighted instances** of event:
    - **100 equally spaced distances** in each pixels' 95% luminosity distance range
    - **Weighting them by the pdf value** of that distance squared
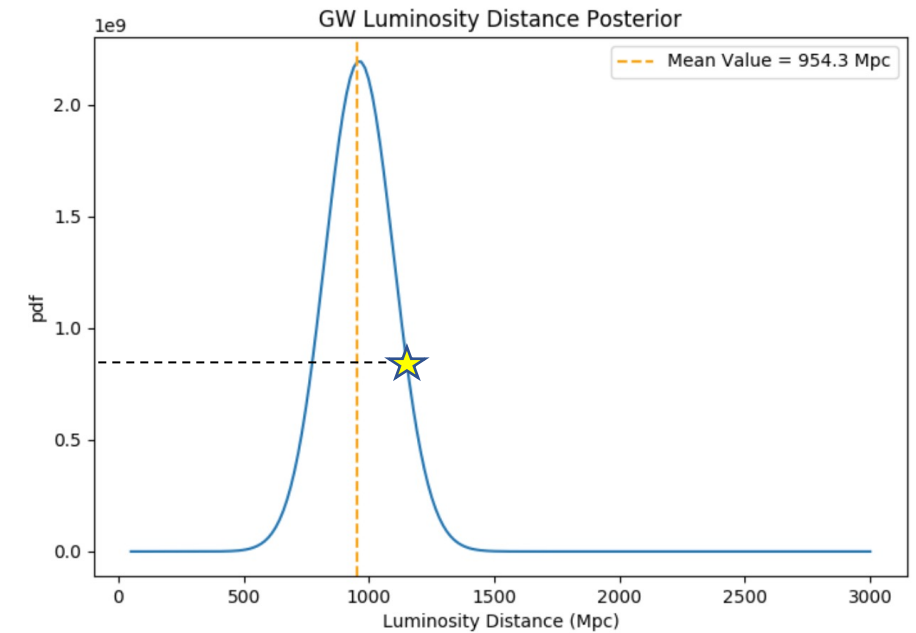   - Using squared value to **amplify** effect of distances closer to mean

- Galaxy Catalog

  - **Redshift range**
  - Only considers galaxies inside GW event **95% confidence region**
  - Calculates completeness: percentage overlap between GW event and Galaxy Catalog

GW Luminosity Distance Posterior

Galaxy Catalog Redshift Histogram. Number of Galaxies: 98543

GW Event and Galaxy Catalog Overlap, Percent Overlap: 0.885

Running Cross-Correlations

- **Parallelization** over H_0 values
- For every cosmology:
  - Calculate **weights for galaxies**
  - Transform to common coordinate system
  - Calculates $\chi(r)$ and $W(r)$ for **all instances** of GW event and **adds them** together
    - **Does not finish** correlation
- Randoms galaxies are **uniformly distributed in comoving volume**
- **Saves** unfinished correlations
- **Updates** logs

GW Luminosity Distance Posterior

--- Mean Value = 954.3 Mpc

📄 log_40.0
📄 log_42.0
📄 log_44.0
📄 log_46.0
📄 log_48.0
📄 log_50.0
📄 log_52.0
📄 log_54.0
📄 log_56.0
📄 log_58.0
📄 log_60.0
📄 log_62.0
📄 log_64.0
📄 log_66.0
📄 log_68.0
📄 log_70.0
📄 log_72.0
📄 log_74.0
📄 log_76.0
📄 log_78.0
📄 log_80.0

```
Complete! 100/100
Time taken: 00:07:56
```

Plots and Outputs w/o Randoms
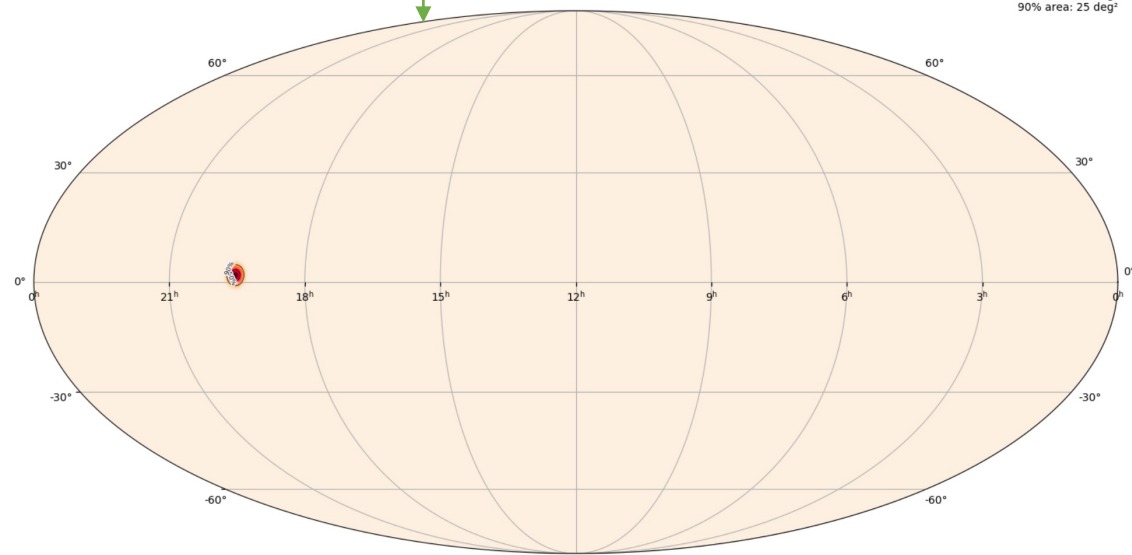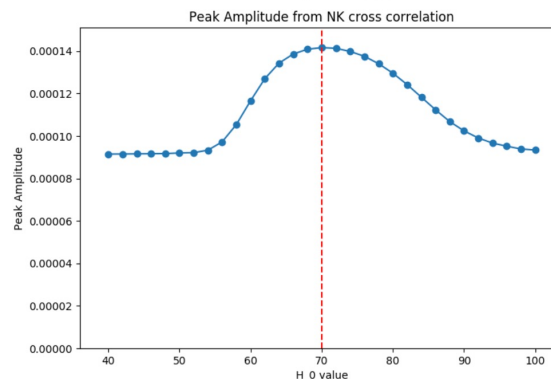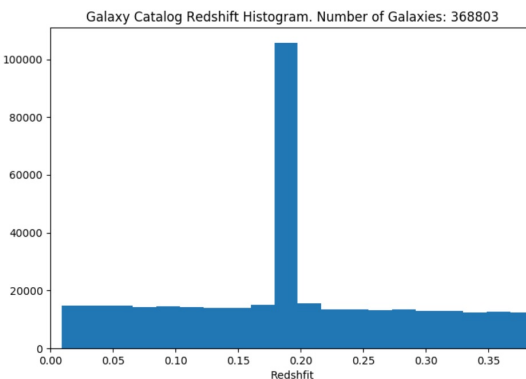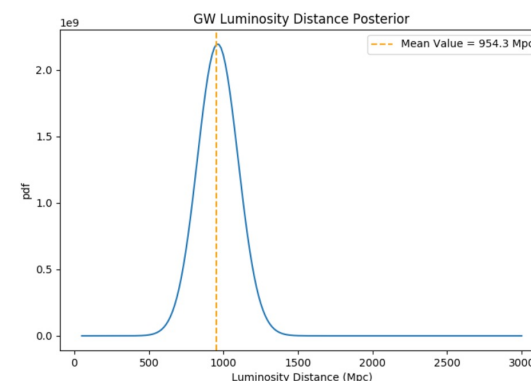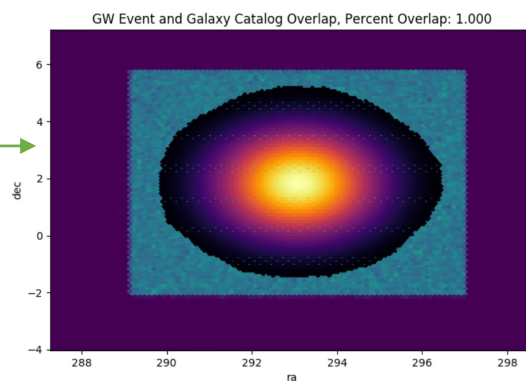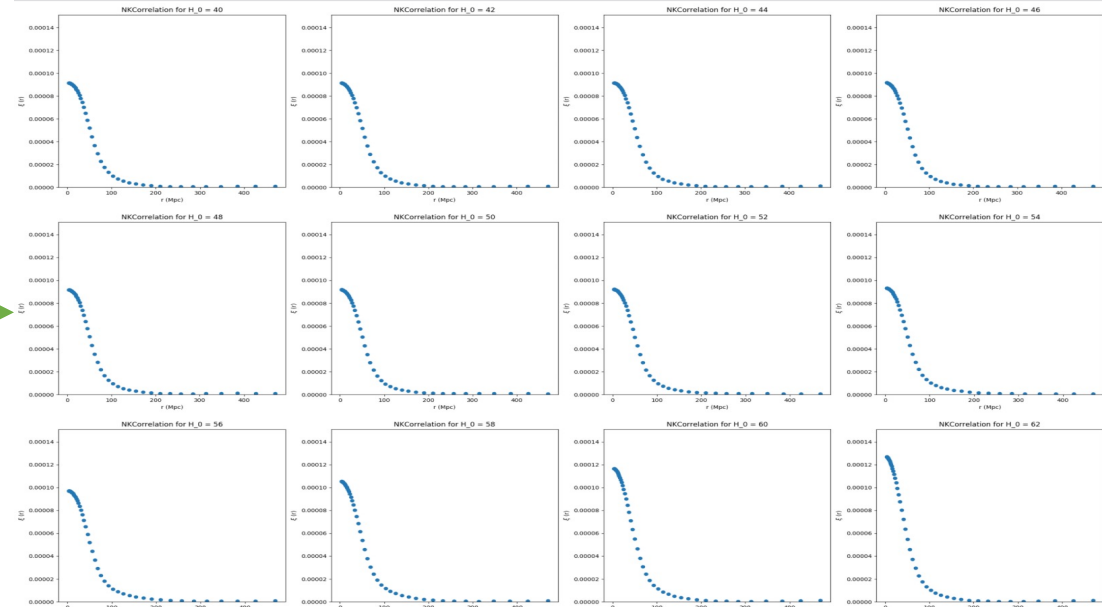
Plots and Outputs with Randoms
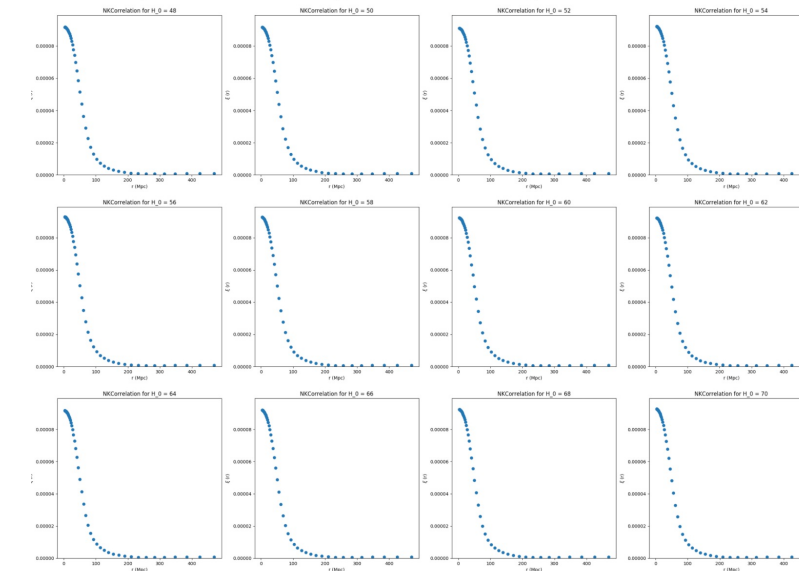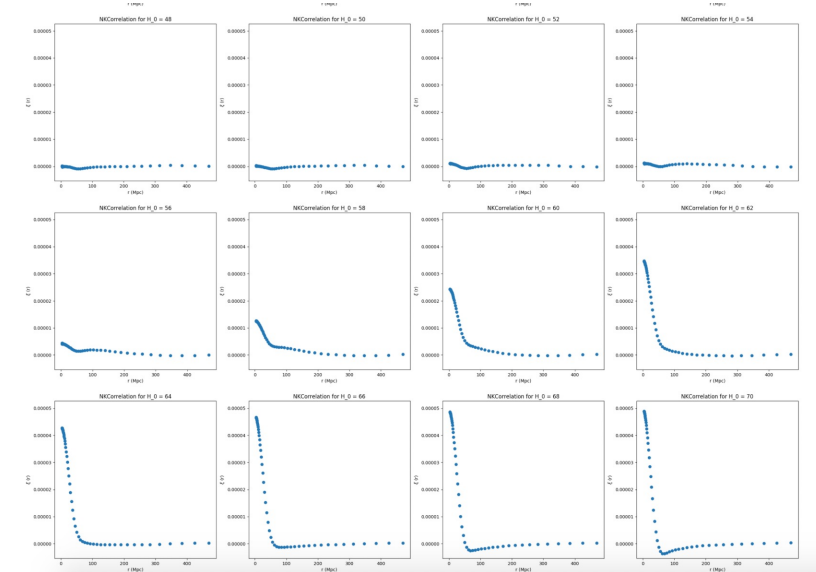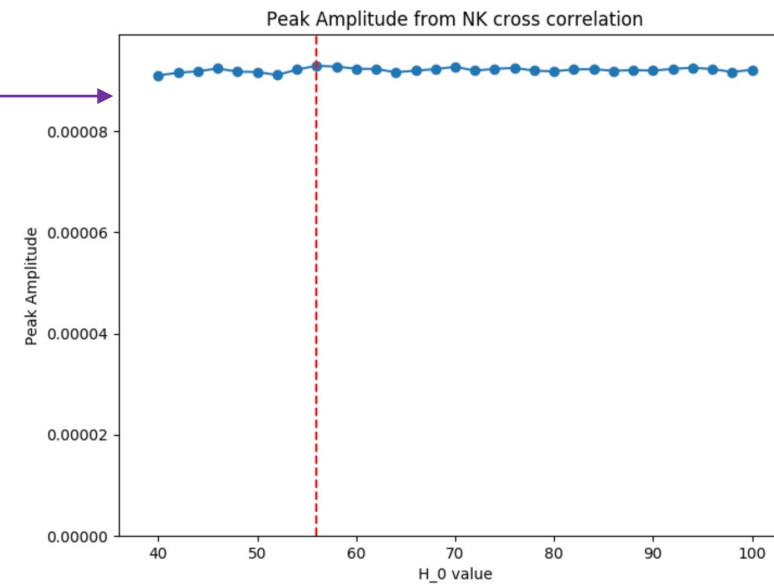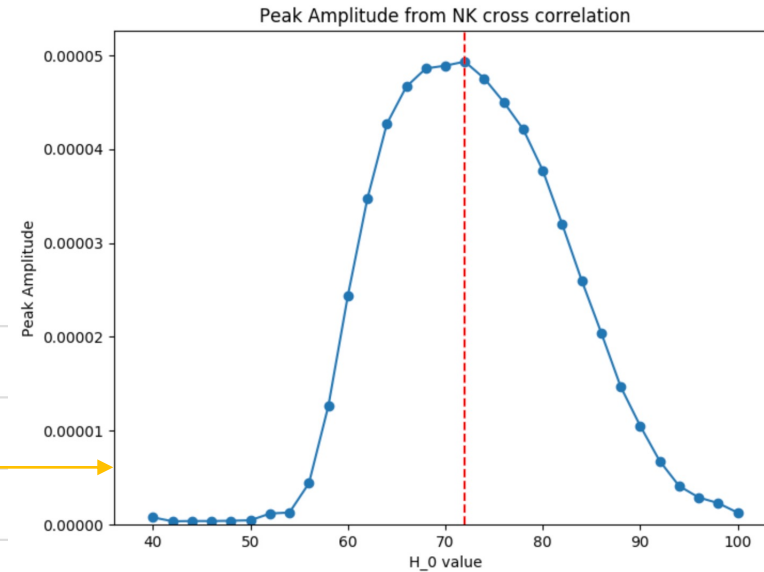
Plots and Outputs All Events
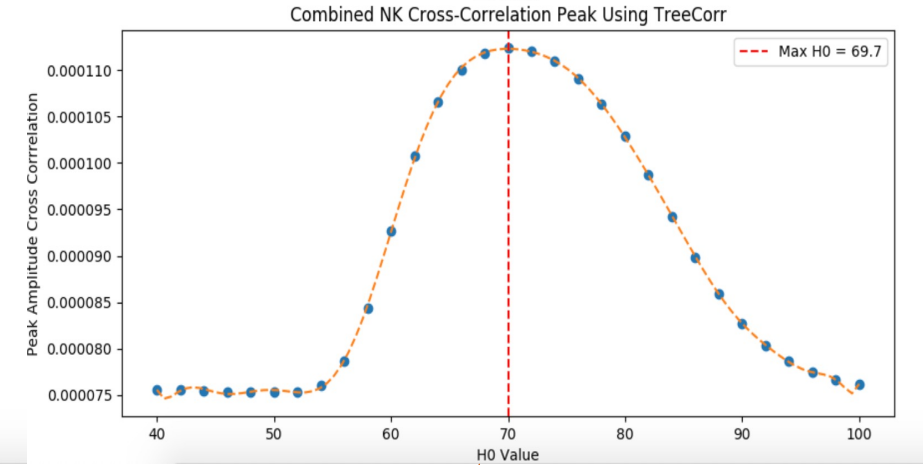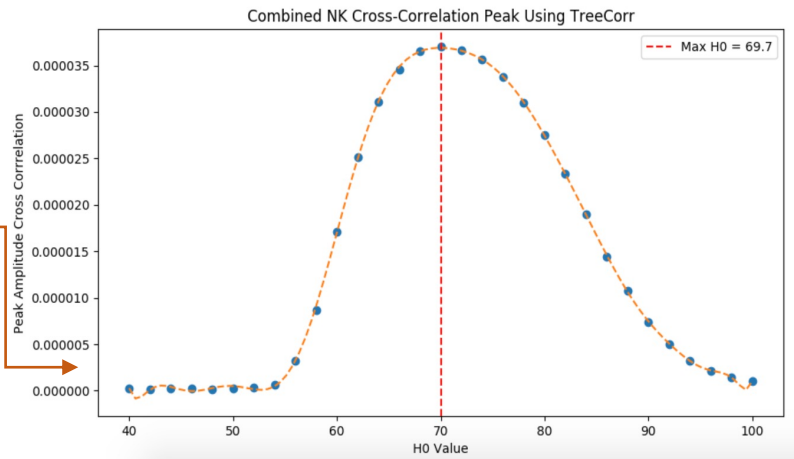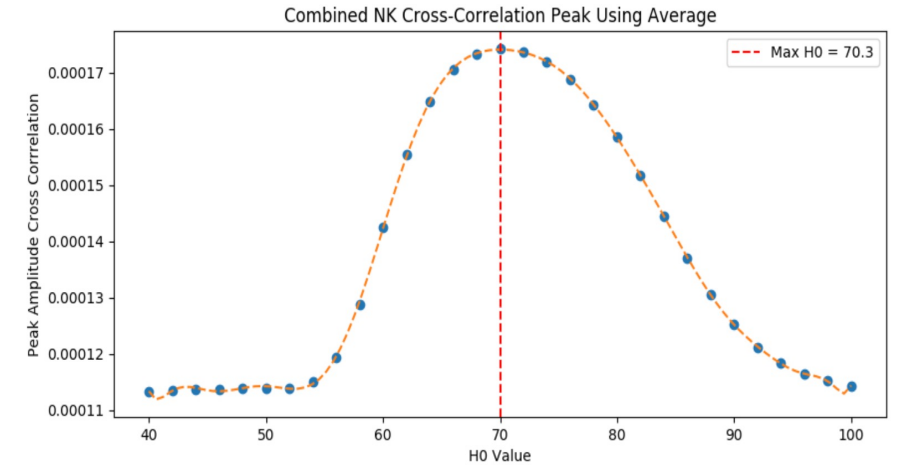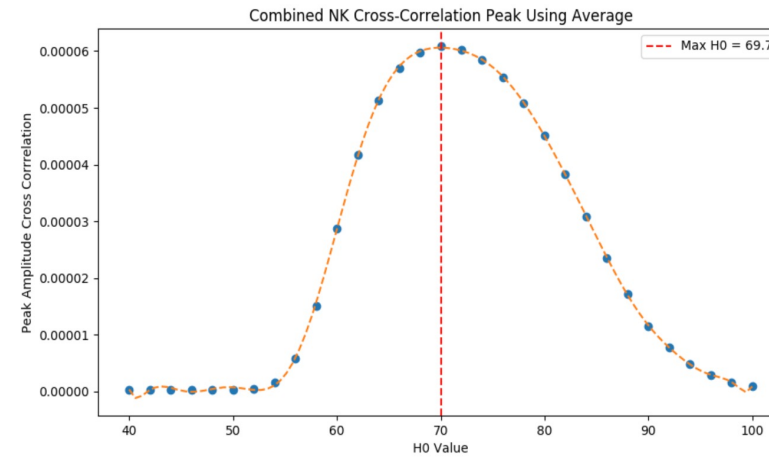
Results

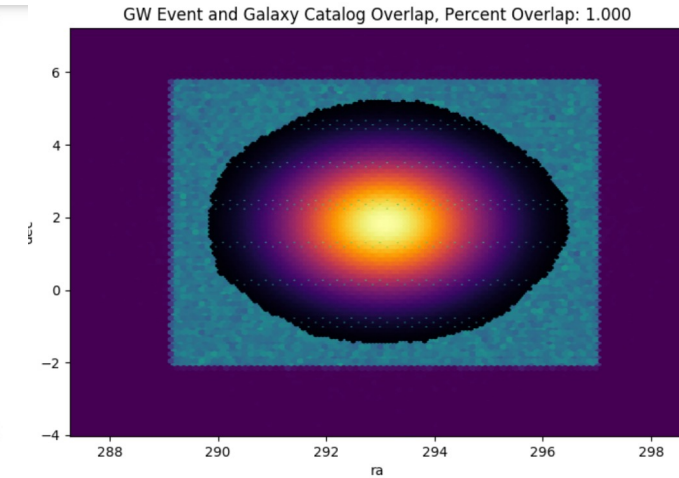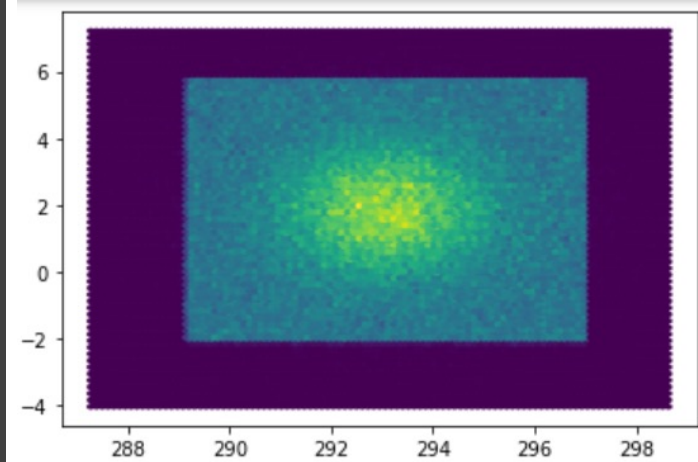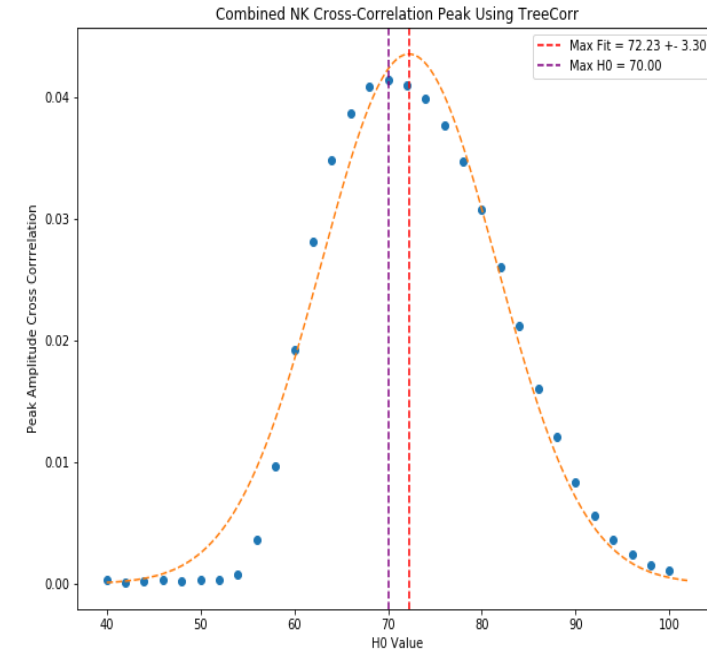# Gaussian Blobs

- 25 Gaussian Blobs

- Luminosity Distance: 200-1000 Mpc

- 15% uncertainty

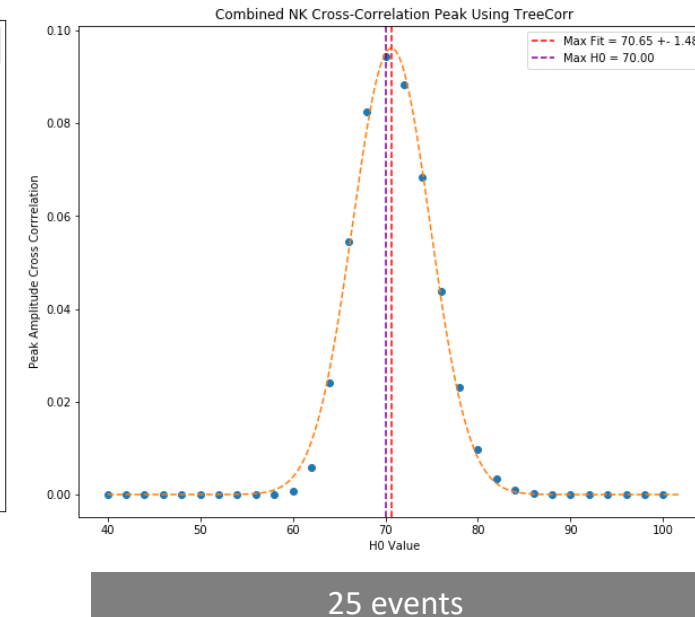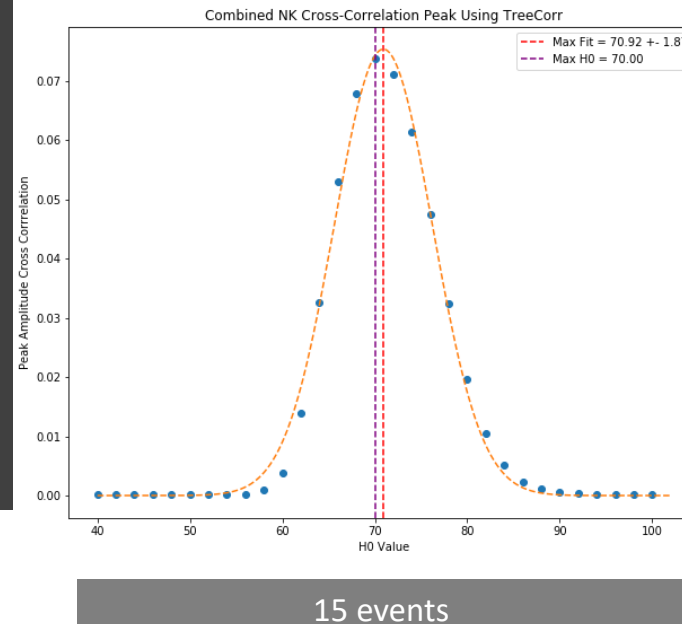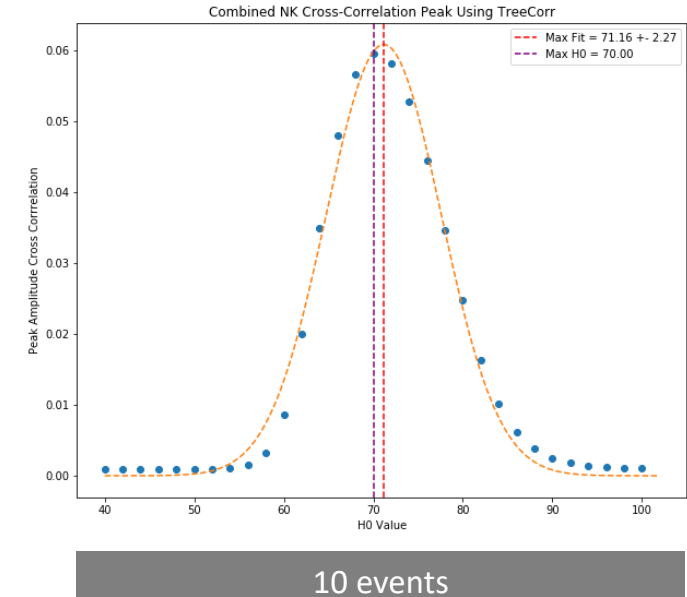- TreeCorr Method **does not produce narrower** distribution with multiple similar events
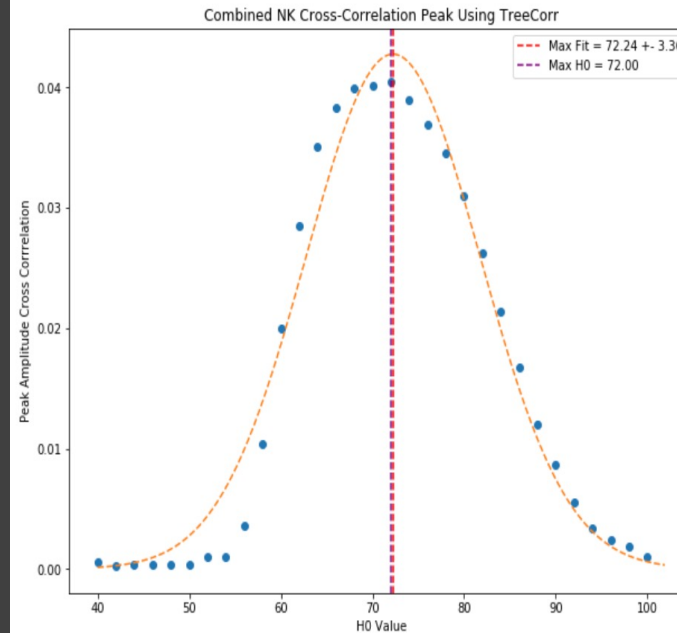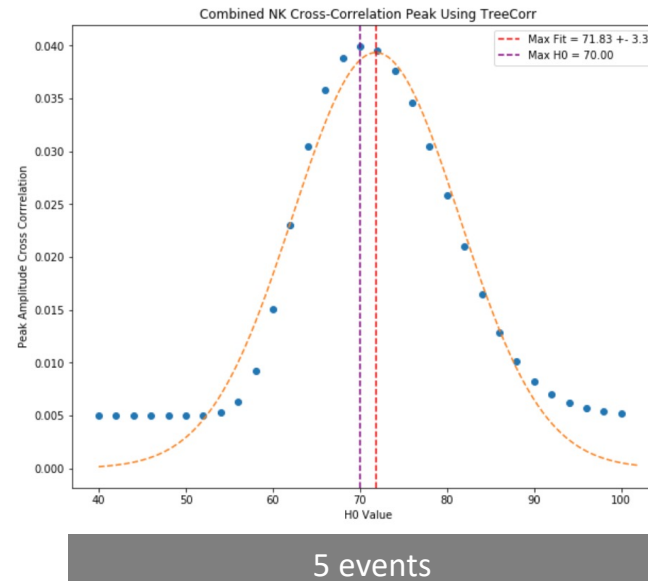

1 event

25 events

Different Combination Method

- **Multiplying finalized**

- Distribution **narrows** with more events

- **No randoms** included in this type of combination

  - Randoms correction can produce negative cross-correlations

5 events

10 events

15 events

25 events

Antonella's BCC Sims

- 200 GW Events
- Minimum completeness: **80%**
- **38 events used**

Fractions overlap between 95% GW conf Region and BCC Catalogs

Combined NK Cross-Correlation Peak Using TreeCorr

Max H0 = 70.19 +- 5.67
Max H0 = 70.00

Combined NK Cross-Correlation Peak Using Multiplication

Max Fit = 71.07 +- 2.39
Max H0 = 70.00
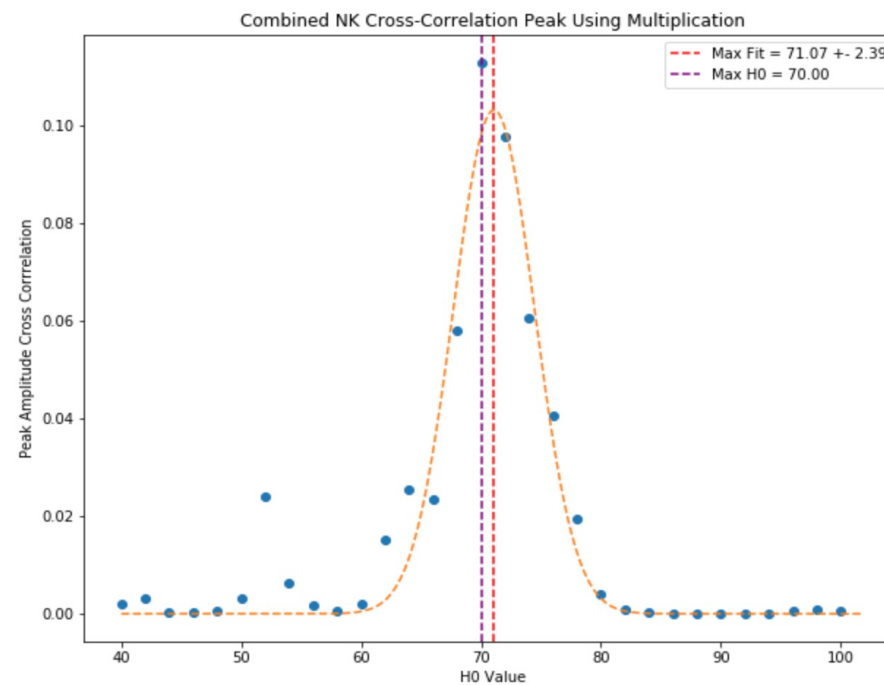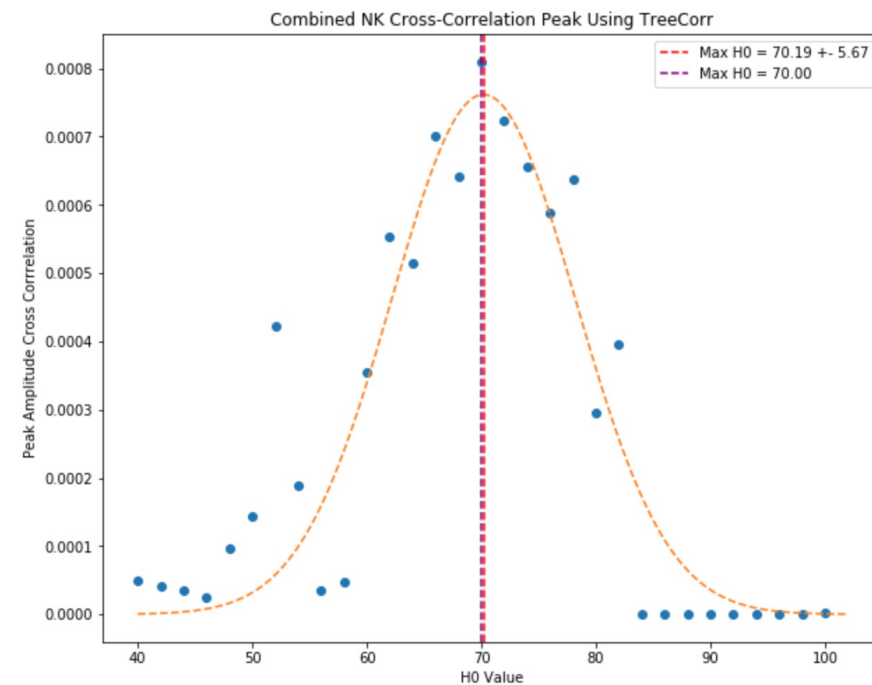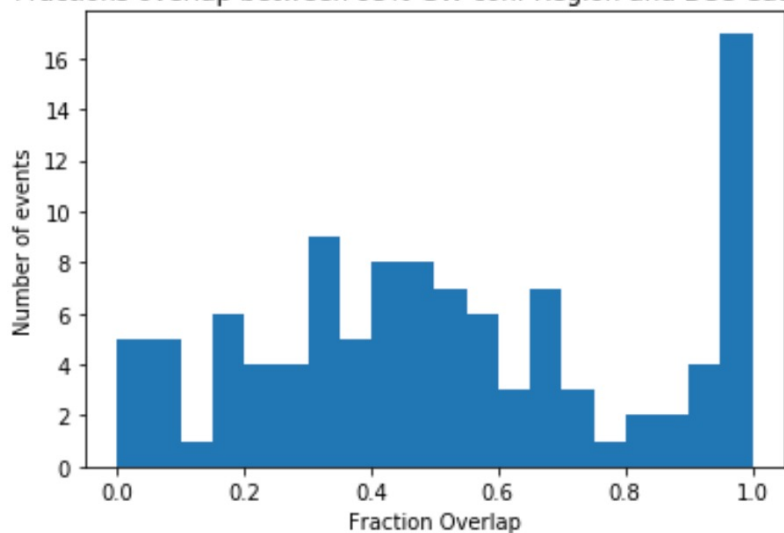
Combined NK Cross-Correlation Peak Using Multiplication

Conclusion

- Maximum cross-correlation analysis seems like a viable method to **infer H_0**

- More work needed in **understanding combination methods**
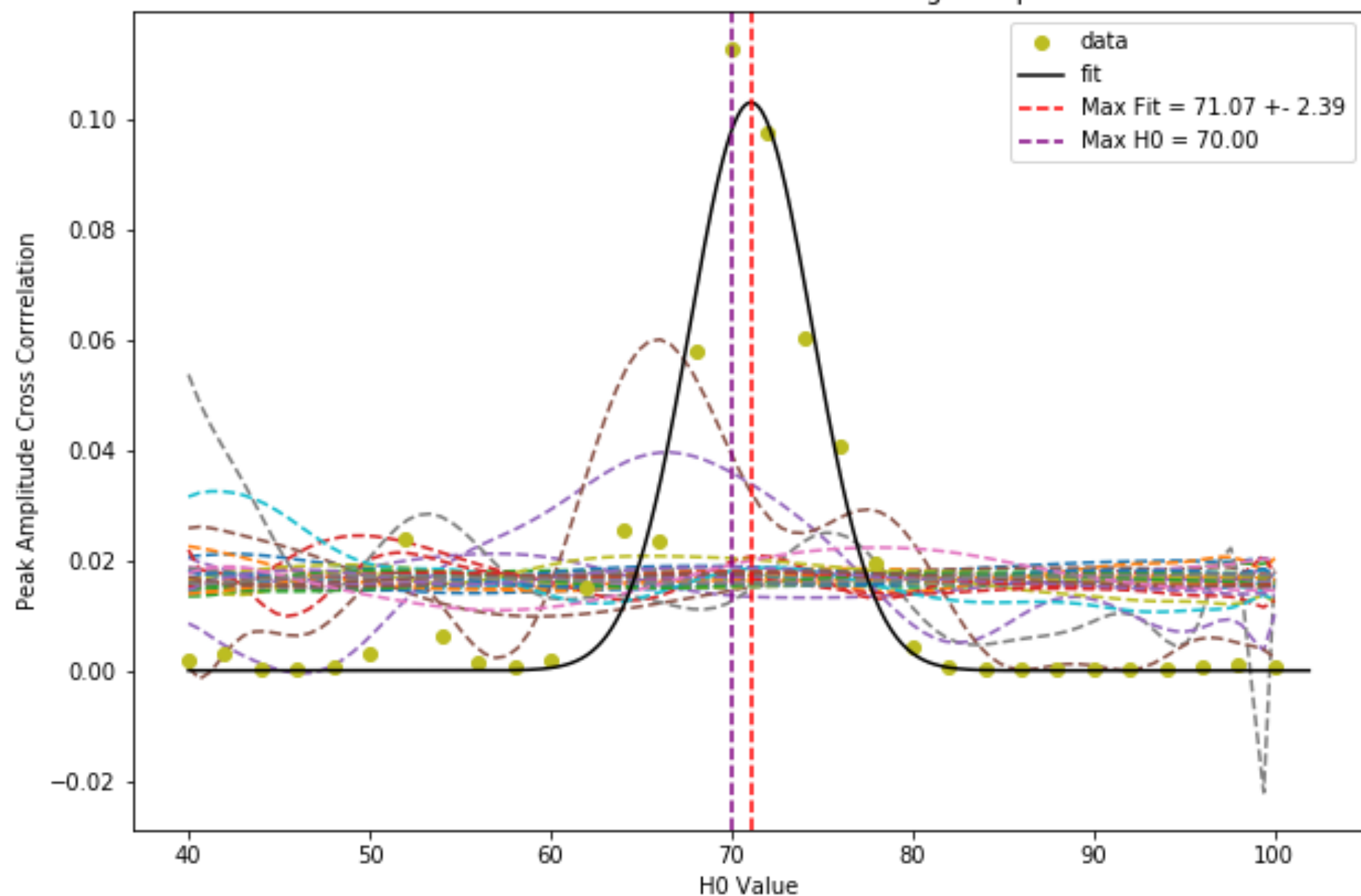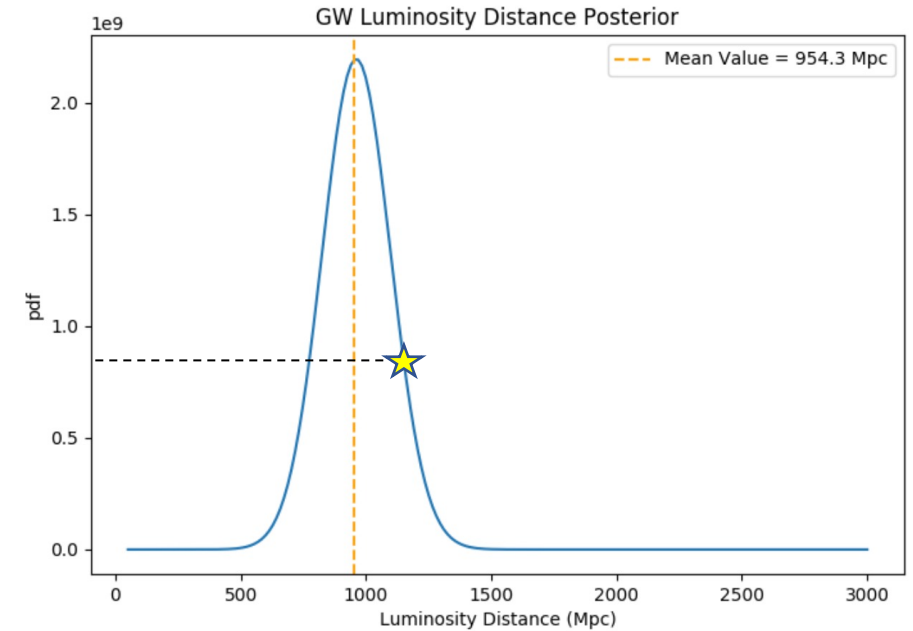
- Other questions:
  - When creating **randoms, how many do we use?** Same as galaxy catalog? Fixed density?
  - Are **redshift uncertainties** statistically significant in this analysis? If so how much?

Running Cross-Correlations

- **Parallelization** over H_0 values

- For every cosmology:
  - Calculate **weights for galaxies**
  - Transform to common coordinate system
  - Calculates $\chi(r)$ and $W(r)$ for **all instances** of GW event and **adds them** together
  - **Does not finish** correlation

- Randoms galaxies are **uniformly distributed in comoving volume**

- **Saves** unfinished correlations

- **Updates** logs

GW Luminosity Distance Posterior

--- Mean Value = 954.3 Mpc

Luminosity Distance (Mpc)

- log_40.0
- log_42.0
- log_44.0
- log_46.0
- log_48.0
- log_50.0
- log_52.0
- log_54.0
- log_56.0
- log_58.0
- log_60.0
- log_62.0
- log_64.0
- log_66.0
- log_68.0
- log_70.0
- log_72.0
- log_74.0
- log_76.0
- log_78.0
- log_80.0

Complete! 100/100
Time taken: 00:07:56