Detecting Objects in Astronomical Images with Statistical Error Control

David Friedenberg, Christopher Genovese and Arthur Kosowsky*

Carnegie Mellon University and *University of Pittsburgh

International Computational Astrostatistics Group
http://incagroup.org

Problem Statement

•Detecting objects like stars, galaxies and galaxy clusters against a background is a common problem in Astronomy.

•A good detection algorithm should balance **power**, the ability to detect objects, against making too many **false detections**, which leads to low purity.

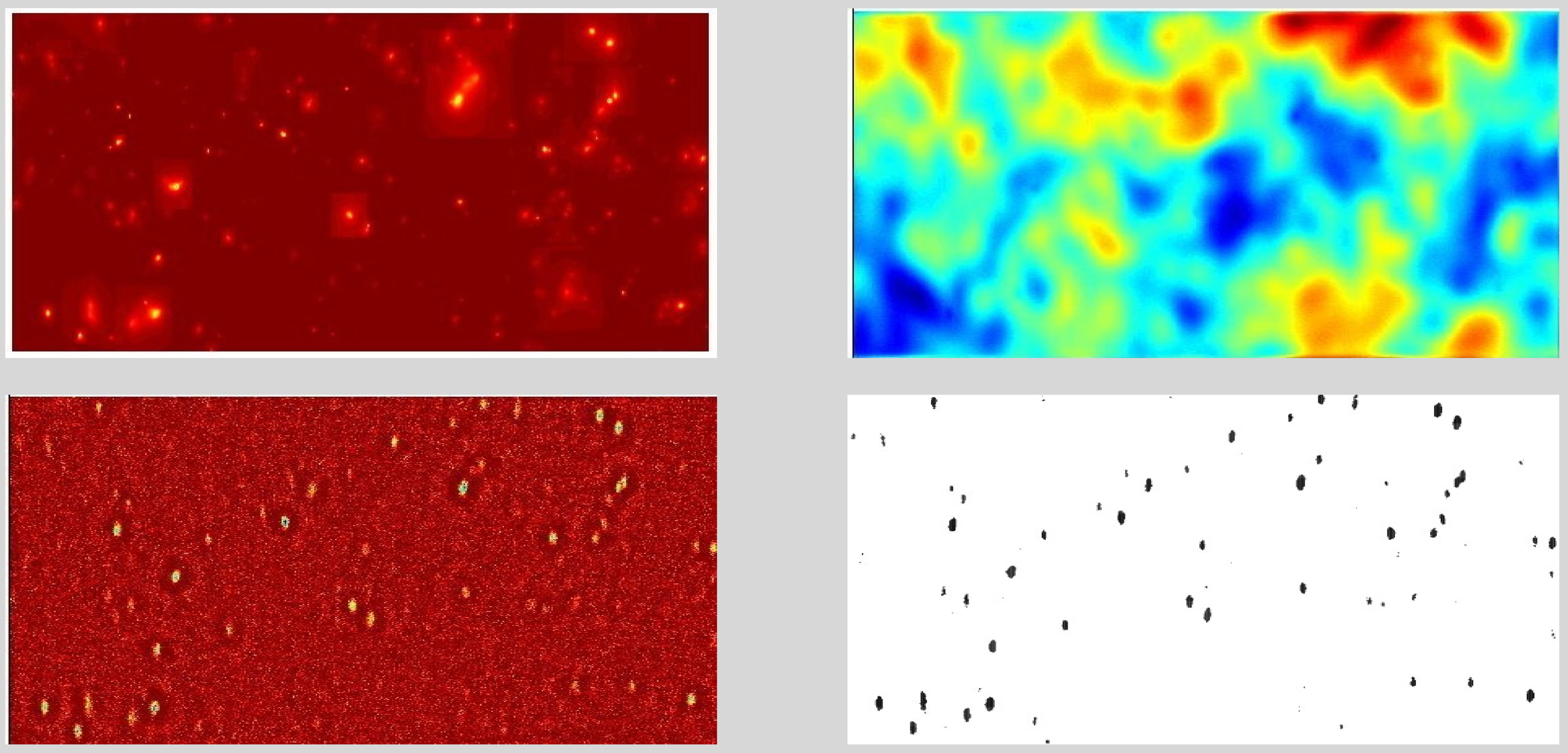
•A standard approach is to use simple heuristic peak finding algorithms.
- •These can have reasonable power but provide **no error control**
- •Purity is typically estimated using simulations but these estimates assume that real data will have identical properties to the simulated data.
- •Thus when the peak finding algorithm is applied to actual data the true purity is **unbounded** and **unknown**

•We introduce the **False Cluster Proportion** (FCP) Algorithm [1]
- •This statistical procedure makes a probabilistic guarantee that the **purity is bounded** below a bound specified by the user
- •We demonstrate this procedure on simulated galaxy cluster data from the Atacama Cosmology Telescope team
- •We show that the False Cluster Proportion technique can **effectively make detections** while **maintaining a pre-specified purity level**

FCP applied to simulated image from the Atacama Cosmology Telescope

We want to **detect galaxy clusters** via their Sunyaev-Zeldovich (SZ) signature. These data are simulated to look like a patch of sky, approximately 12 degrees square, we would expect to see with the Atacama Cosmology Telescope [5]. The Atacama Cosmology Telescope [3] as well as the South Pole Telescope[4] will make microwave sky maps with an angular resolution of approximately 1 arcminute, with the typical amplitude of the fluctuations being up to 100muK around the 2.7K mean temperature. Both telescopes are designed for detecting clusters via their SZ signature

The top left image is a simulated patch of sky showing only the 202 clusters from the simulation. The top right image is what we would observe with the telescope and includes the clusters but also CMB, dust, and point sources. There is also smoothing from the beam and white noise that also need to be filtered out.



The bottom left image comes from applying an appropriate filter to the top right image. We get a noisy reconstruction of the location of the clusters. Finally, in the bottom right panel,we apply the FCP procedure to the filtered image. We **detect virtually all the clusters** we can see in the top left image. Additionally, we have **bounded the proportion of falsely clusters detected to be less than .1** with probability.95

FCP Method

We treat detection as a **statistical multiple testing** problem where the image is a realization of a random field

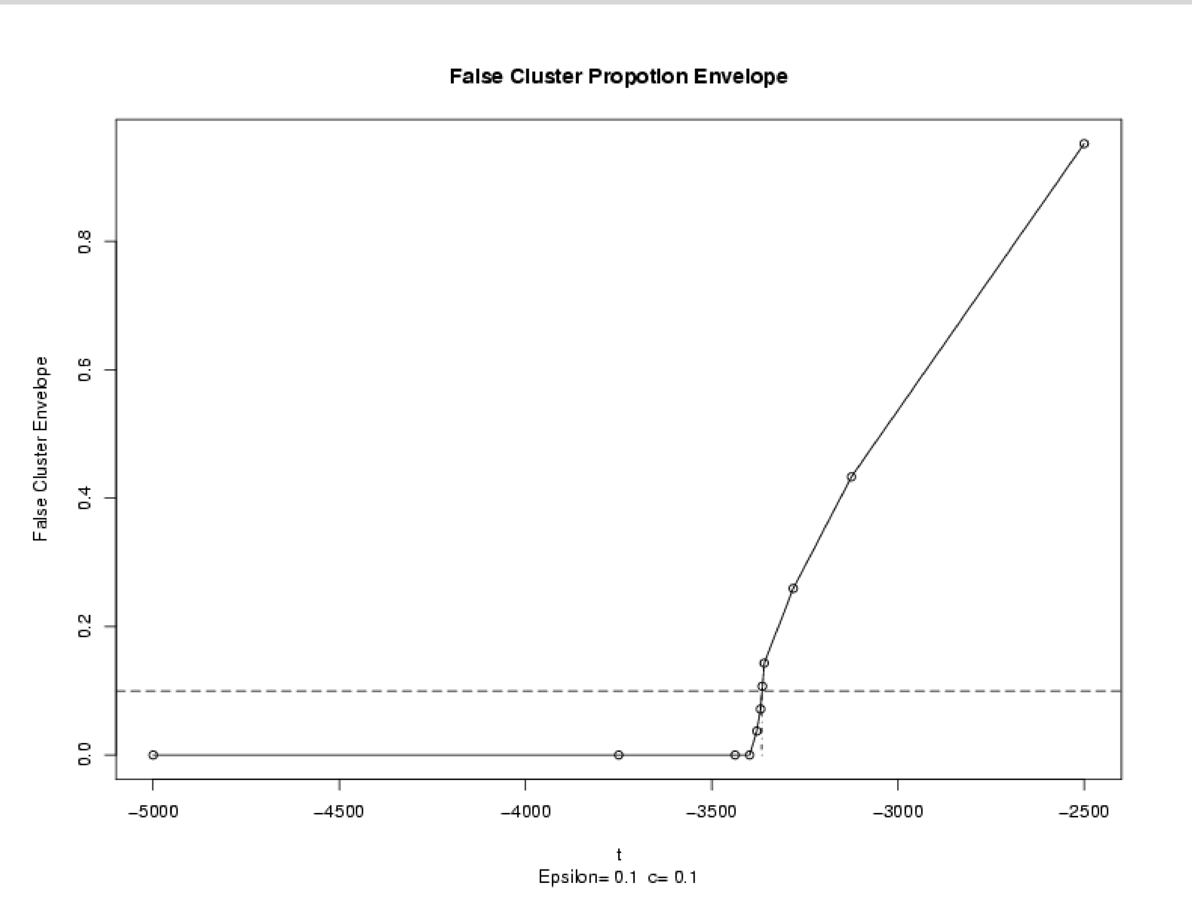
1.) We first derive a set containing the background with at least the specified probability

2.) Next we decompose the image into clusters with all pixels above an intensity threshold and touching being considered a cluster

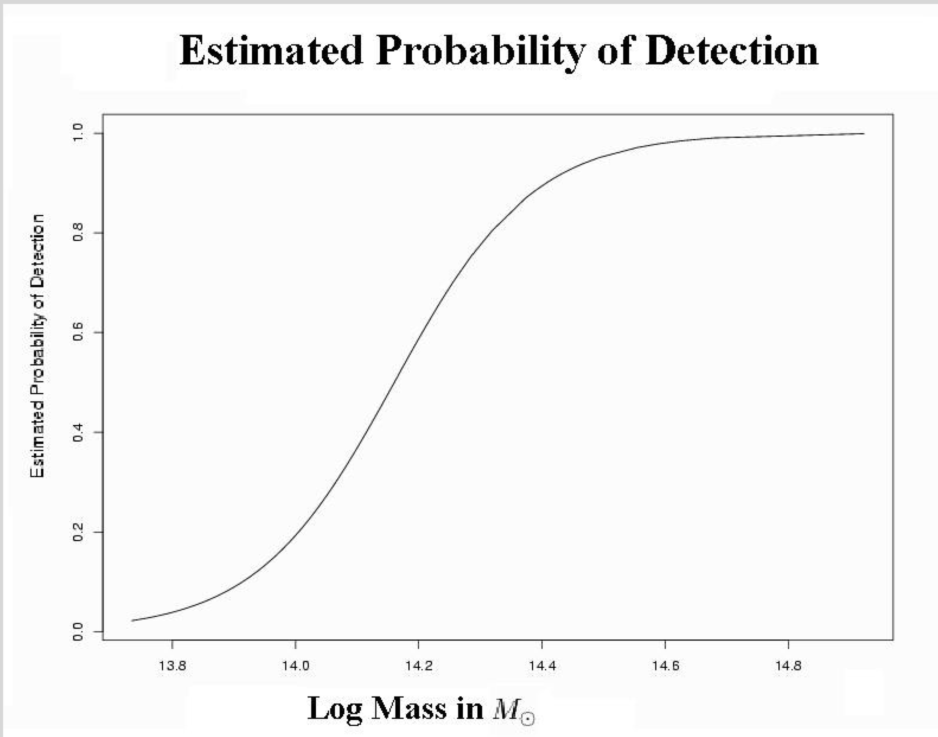
3.) We use our superset to classify the clusters as real or false.

4.) We vary the intensity threshold to match the proportion of false clusters with our desired threshold. This will tell us the threshold value for which clusters to keep

Below is the plot of a typical search for the proper cutoff value, we see a cutoff intensity around -3400 will bound the proportion of false clusters to be less than .1 (dashed line)



Modeling Detections and Future Work



We know that more massive clusters are easier to detect. In fact, we can **model our ability to detect clusters** as a function of mass. As expected there is a high probability of detecting massive clusters and as the mass decreases, so does the **probability of detection.**

•We can use the estimated probability of detection to estimate how many clusters we miss at a given mass

•Future work will focus on generalizing the procedure to **include mass**

•We hope to analyze a map and claim that with high probability we **controlled the proportion of false clusters** and have **detected everything above a specified mass threshold**.

•This would insure that we are getting an accurate sample of clusters of a certain on mass

References and Acknowledgments

We are always looking for new and interesting datasets, please contact David Friedenberg dfrieden@stat.cmu.edu

Thanks to Arthur Kosowsky, Neelima Sehgal, Paul Bode, Hy Trac, Kevin Huffenberger and the ACT team for the maps. Thanks to Chris Genovese, the Carnegie Mellon University Department of Statistics, and Arthur Kosowsky for financial assistance.

References:

[1]Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal Statistic Society, B(1995)

[2]A. Kosowsky, The atacama cosmology telescope, New Astronomy Reviews, 47 (2003)

[3]M.P. Pacifico, C. Genovese, I. Verdinelli, and L. Wasserman,false discovery rates for random fields, Tech Report 771, Carnegie Mellon University Department of Statistics

[4]J.Ruhl et al., "The South Pole Telescope", Proc. SPIE, Vol. 5498, p 11-29, 2004.

[5]N. Sehgal, H. Trac, K. Huffenberger, and P. Bode, microwave sky simulations and projections for galaxy cluster detection with the atacama cosmology telescope, The Astrophysical Journal, 664(2007)