# SAMSI Spring 2006 Program on Astrostatistics
# Workshop on Source and Feature Detection
# Sub-Groups and Research Topics

## Group I: Modeling High-Energy Count Data

**Participants:** Keith Arnaud, Jim Chiang, Alanna Connors, Peter Freeman, Jiashun Jin, Vinay Kashyap, Taeyoung Park, Adam Roy, Jeff Scargle, Aneta Siemiginowska, David van Dyk Alex Young, Yaming Yu

### Topic I-A. Image Reconstruction for GLAST Data

*Point Person:* Jim Chiang

*Participants:* Alanna Connors, Adam Roy, Jeff Scargle, David van Dyk,Martin Weinberg (if we try to use his BIE)

We'd like to generalize EMC2 (or similar methods) to handle GLAST data in order to provide a means of characterizing point sources and extended structures. This would involve extending the current EMC2 algorithm to include the energy dependence of the PSF, the projection effects onto the Celestial sphere for large fields-of-view, and accounting for a known background component. A related topic would be source identification in the reconstructed images.

Other topics related topics relating to GLAST might include characterizing the null distribution for LRT for point sources, computing point source upper limits, fast methods for point source identification, blind searches for periodicity (e.g., pulsars), and multi-wavelength spectral fitting (joint with data from other wavebands). Maybe we can tell from the other projects that are submitted whether there is potential for organizing such a group.

### Topic I-B: Solar Astrophysics: high energy imaging and spectroscopy with RHESSI

*Point Person:* Alex Young

*Participants:* Vinay Kashyap, David van Dyk

The current solar high energy mission is the Ramaty High Energy Solar Spectroscopic Imager (RHESSI). This experiment combines multiple high resolution germanium detectors with rotating grid collimators to obtain X-ray/gamma-ray spectra and images from solar flares.

(a) The spectroscopy aspect of RHESSI requires a method for determining photon model fits to data. The data are measured counts so an instrument response must be applied to move from count space to photon space. The known physics tells us that for the full energy range a model typically contain a continuum component composed of a power law and a thermal component along with Gaussian like lines in the X-ray and gamma-ray ranges. The currently used method for determining a model fit is the so-called iterative forward folding method with Chi-squared goodness-of-fit test. The standard minimization algorithm is Levenberg-Marquardt. This approach has many problems one of the biggest being the lack of straightforward way to include instrument systematic errors (e.g. pulse pile-up). This problem is not unlike those encountered in x-ray spectroscopy (e.g. with Chandra).

(b) Imaging with RHESSI is again not unlike that in x-ray astronomy. Though the details of the instrument are different from Chandra that data analysis needs are similar. RHESSI image analysis is in need of a robust image reconstruction method that provides a way to estimate errors for the reconstruction.

For this two part project a reasonable task would seem to be to adapt the current methods that have be developed for Chandra to RHESSI.

**Topic I-C: Source and Feature Significance in Poisson Data**

*Point Person:* Aneta Siemiginowska

*Participants:*

1. How to define a borderedge for low counts data? The features are seen in the smoothed images. Can the smoothed image be used to define the border and the source size? How do we calculate the significance of such sources?
2. How to choose the smoothing scales for unsharp-mask images and describe the significance of the structures in such data? What tells us that such structure is real?
3. Detecting changes in time binned images at low counts limits. This is related to the last example in my presentation, which I did not have time to describe in details.

## Group II: Modeling Optical and Solar Images

**Participants:** Vinay Kashyap, Hyunsook Lee, Thomas Lee, Aneta Siemiginowska, Rebecca Willett, Alex Young, Yaming Yu, Lingsong Zhang

**Topic II-A: Solar Astrophysics: Feature Detection in White Light and Extreme Ultraviolet Images**

*Point Person:* Alex Young

*Other Participants:* Vinay Kashyap

Extreme ultraviolet and white light coronagraph images of the sun show us the dynamic nature of the solar corona. These data are very rich in both slowly evolving and transient structures. Some researchers have worked to detect and classify these structures but most research in this area is very basic. There is a strong need for rigorous and robust methods for use with this data. This is important not only because of the need to have better quantitative analysis methods but it is a vital aspect of automated computer systems that will be required to handle the massive data volumes of future ground and space-based solar missions. This project will be tasked with developing such statistical image processing methods.

## Group III: Statistical Issues with Low-Count Data

**Participants:** Keith Arnaud, Jim Chiang, Peter Freeman, Woncheal Jang, Jiashun Jin, Viney Kashyap, Hyunsook Lee, Ji Meng Loh, Xiao-Lim Meng, Taeyoung Park, Aneta Siemiginowska, David van Dyk, Rebecca Willet, Alex Young, Yaming Yu, Lingsong Zhang

**Topic III-A: Upper Limits**

*Point People:* Vinay Kashyap and Peter Freeman

*Participants:* David van Dyk

The basic issue here is to develop a statistically meaningful definition of the concept that astronomers and physicists have in mind when they use the term "upper limits", and to provide a prescription of how to decide what its value should be in any given problem. And also to work it out in some specific cases such as wavdetect, or a line in a spectrum, etc.

**Topic III-B: Poisson Data in the Presence of a Poisson Background**

*Point Person:* Keith Arnaud

*Participants:*

For example: suppose we have two observations made by an X-ray detector. One is of source and background, the other just contains background. We have $N$ energy channels in each observation. The special case of $N = 1$ is the source detection problem.

We have a source model with $M_S$ parameters and one for the background with $M_B$ parameters. Following the Bayesian track we can marginalize over the $M_B$ background parameters given priors on the source and background parameters. We can then calculate the posterior distribution for the $M_S$ source parameters.

Alternatively, as frequentists we can maximize the likelihood for the $M_B$ background parameters in terms of the source parameters and the data then substitute back into the log likelihood expression to get the profile likelihood. We can then maximize this and determine confidence regions on the $M_S$ source parameters.

Questions :

1. In the Bayesian approach how should we choose priors to get the best frequentist coverage properties on the posteriors?
2. In the frequentist approach how we can calculate confidence regions when we are away from the asymptotic limit?

Possibly relevant work I came across

Stern, S., 1997, J.R.Statist.Soc. B, 59, 653

Reid, N., 2003, PHYSTAT2003

**Topic III-C: Classification of Sources—Extended v. Point**

*Point Person:* Peter Freeman

*Participants:*

A major unsolved problem of source detection is how to classify sources as either point-like or extended in a robust, automated matter. The fitting of a point-spread function (PSF) to image data can in theory allow one to perform at least a alternative- free hypothesis test ($H_0$: source is point-like; $H_1$: it has arbitrary shape). Problems with this approach include the fact that PSFs are energy-dependent (an issue usually swept under the rug as dealing with it can greatly slow analyses) and have shapes that are not analytically expressible. How can one hope to identify an almost-point-like high-z X-ray cluster using a PSF which is at best known only approximately?

So to perform robust classification, one would probably have to "leave the image" and gather evidence by some other means, such as exploring the hyperspace of source properties or doing an a posteriori classification after analyzing hundreds/thousands of fields.

Or perhaps methods could be combined, e.g., use Voronoi tesselation to examine imagelets containing sources detected with wavelets. Etc. Development of a good classifier would make one a hero amongst cluster hunters!

## Group IV: Quantification of Uncertainty In Image Reconstruction

**Participants:** Keith Arnaud Jim Chiang, Alanna Connors, Peter Freeman, Woncheal Jang, Jiashun Jin, Vinay Kashyap, Hyunsook Lee, Thomas Lee, Ji Meng Loh, Xiao-Li Meng, Adam Roy, Jeff Scarglem Aneta Siemiginowska, David van Dyk, Alex Young

**Topic IV-A:** : **Quantification of Uncertainty In Image Reconstruction**

*Point Person:* Alanna Connors

*Participants:* Vinay Kashyap, David van Dyk

## Group V: Statistical Computing

**Participants:** Keith Arnaud, Peter Freeman, Vinay Kashyap, Thomas Lee, Xiao-Li Meng, Taeyoung Park, Aneta Siemiginowska, David van Dyk, Alex Young, Lingsong Zhang, Zhengyuan Zhu

## Other Topics

*Point Person:* Jeff Scargle

*Participants:*

Some general topics to keep in mind throughout the program:

- one dimension (e.g. time series and other sequential data) [e.g., "bump" hunting is like detecting sources in images; jump and/or change point detection is of much interest for GLAST and other high-energy contexts.]
- event/point data: from the bottom up ("each photon is an image") as in my presentation at the planning meeting