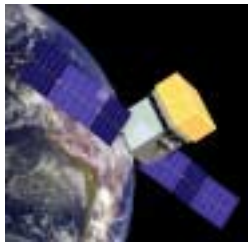
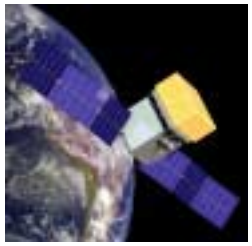


Classification PSF Analysis

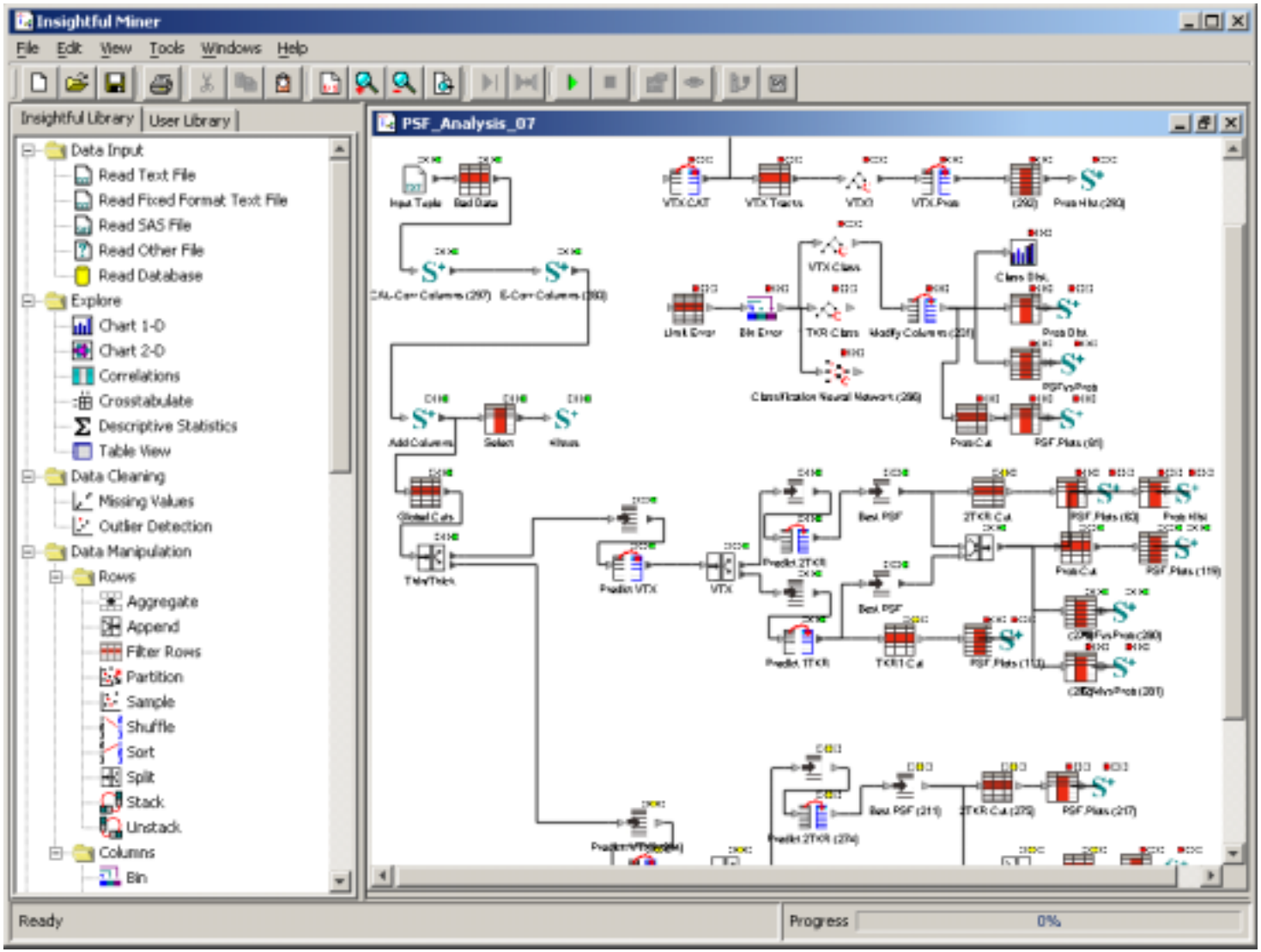
CPA

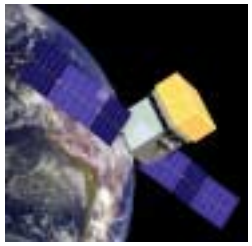
- A New Analysis Tool: Insightful Miner
- Classification Trees
- From Cuts ⇨ Classification Trees:
  Recasting of the GLAST PSF Analysis
- Energy Dependencies
- Present status of GLAST PSFs



A Data Mining Tool

An Miner Analysis Program!





Miner Details

What is a Data Miner?

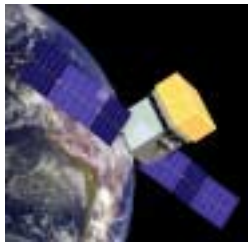
- o A graphical user programming environment
- o An ensemble of Data Manipulation Tools
- o A Set of Data Modelling Tools
- o A "widget" scripting language
- o An interface to data bases

Why use a Data Miner?

- o Fast and Easy prototyping of Analysis
- o Encourages "exploration"
- o Allows a more "Global" View of Analysis

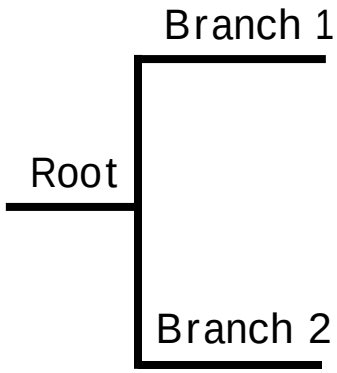
A Traditional "CUT"

A Properties Browser to set parameters



Classification Trees

Given a "catagorical varible" split the data into two pieces using "best" independent continuous varible



Example: VTX.Type =
- 1 if "vertex" direction is best
- 2 if "best-track" direction is best

Continue process – treating each branch as a new "root." Terminate according to statistics in last node and/or change in Entropy

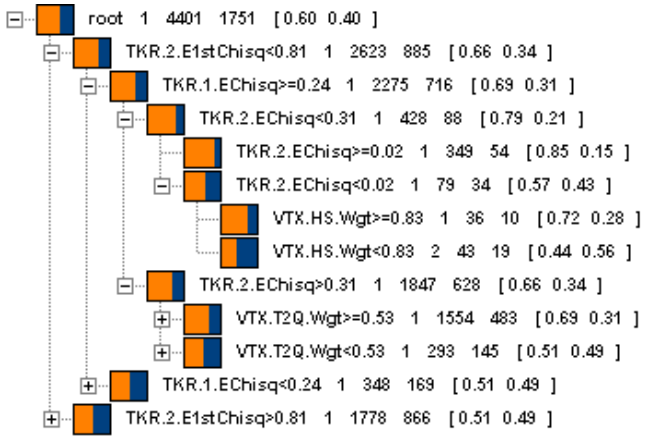
Use "Entropy" to deside which Independent varible to use:

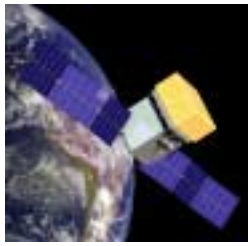
$$\text{Entropy} = \sum_k p_{ik} \log(p_{ik})$$

Where k is over catagories and i is the i^{th} Node

(There are other criteria)

Example: Classification Tree from Miner





Classification Trees

Why use Classification Trees?

1. Simplicity of method – recursive application of a decision making rule

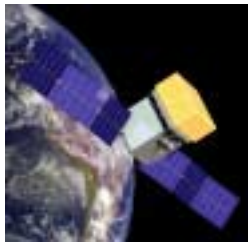
2. Easily captures non-linear behavior in predictors as well As interactions amoung them

3. Not limited to just 2 catagories

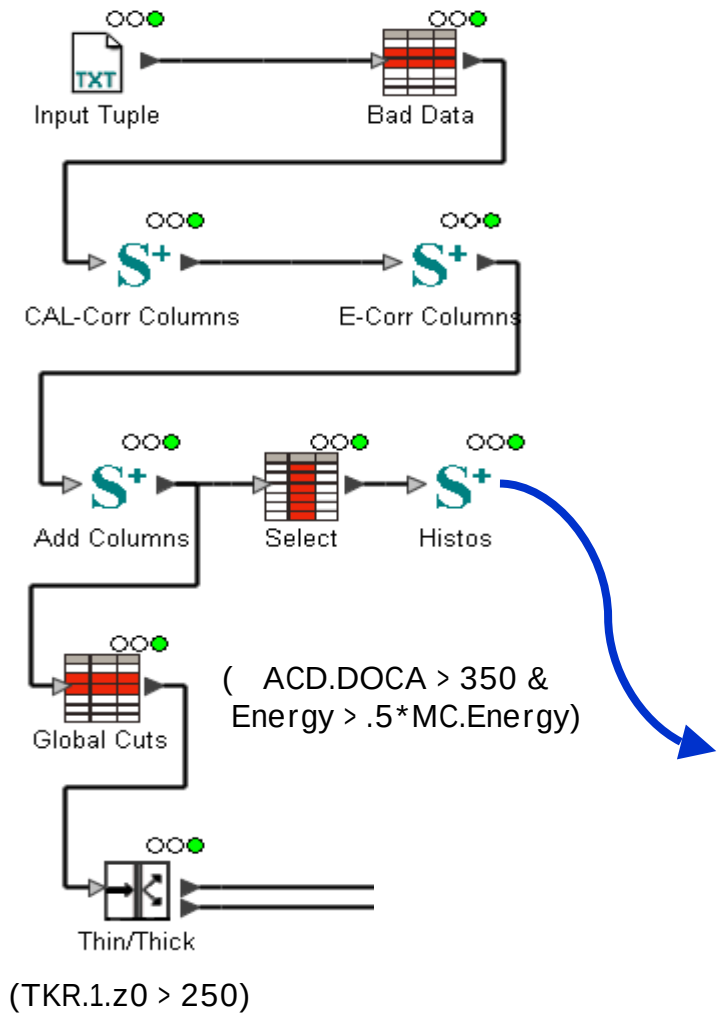
There are numerous text on this subject……

In the following analysis Classification Trees will be used to:

- Separate out the good “vertex” events
- Predict how “good” and event really is

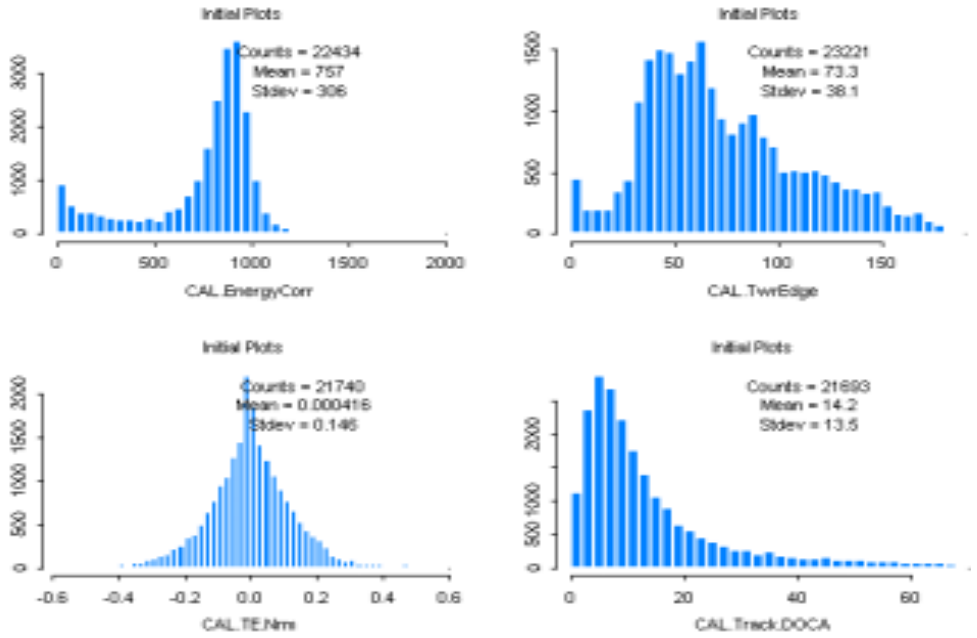


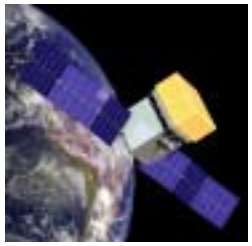
GLAST PSF Analysis



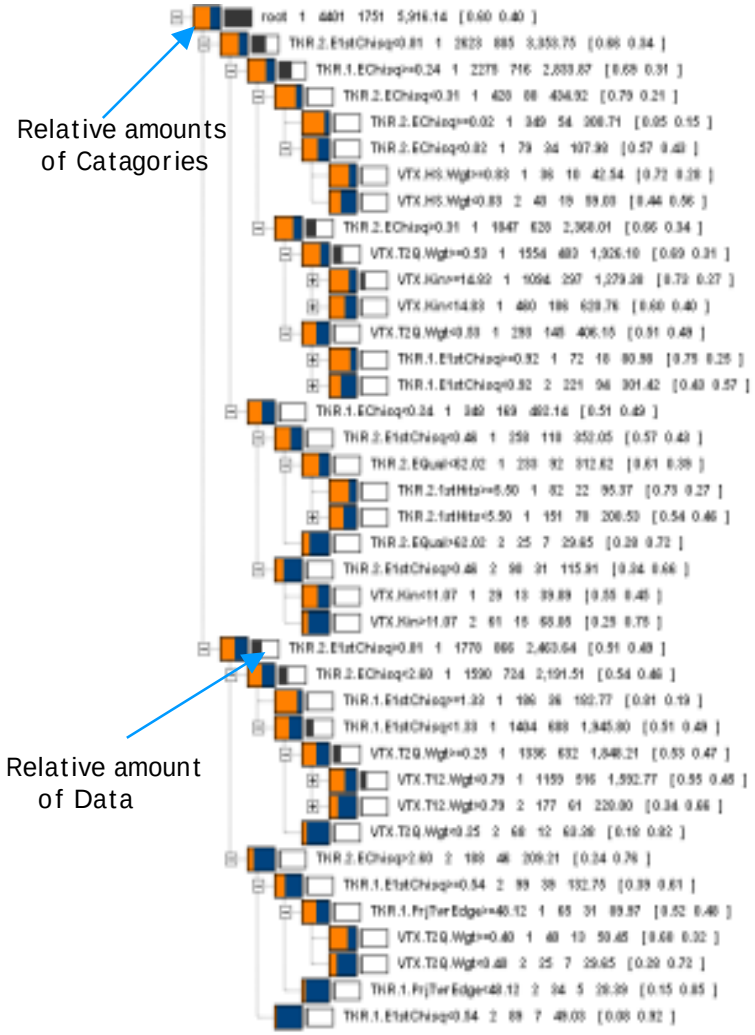
This portion of the code

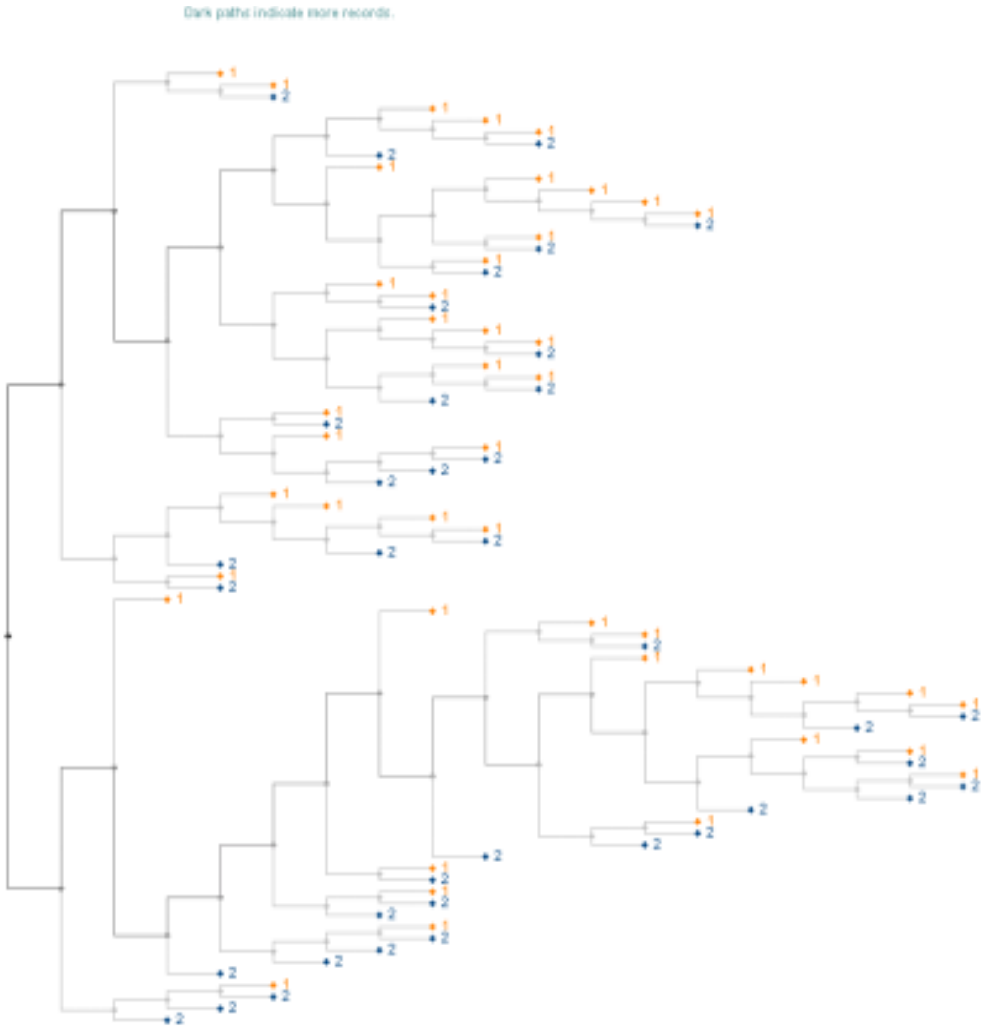
- Reads in the data
- Culls out bad data
- Adds new columns for analysis
- Makes Global Cuts
- Splits the data into 2 pieces
- Thin Radiators
- Thick Radiators

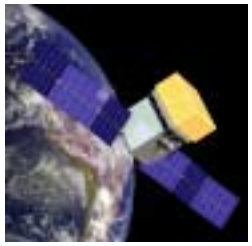




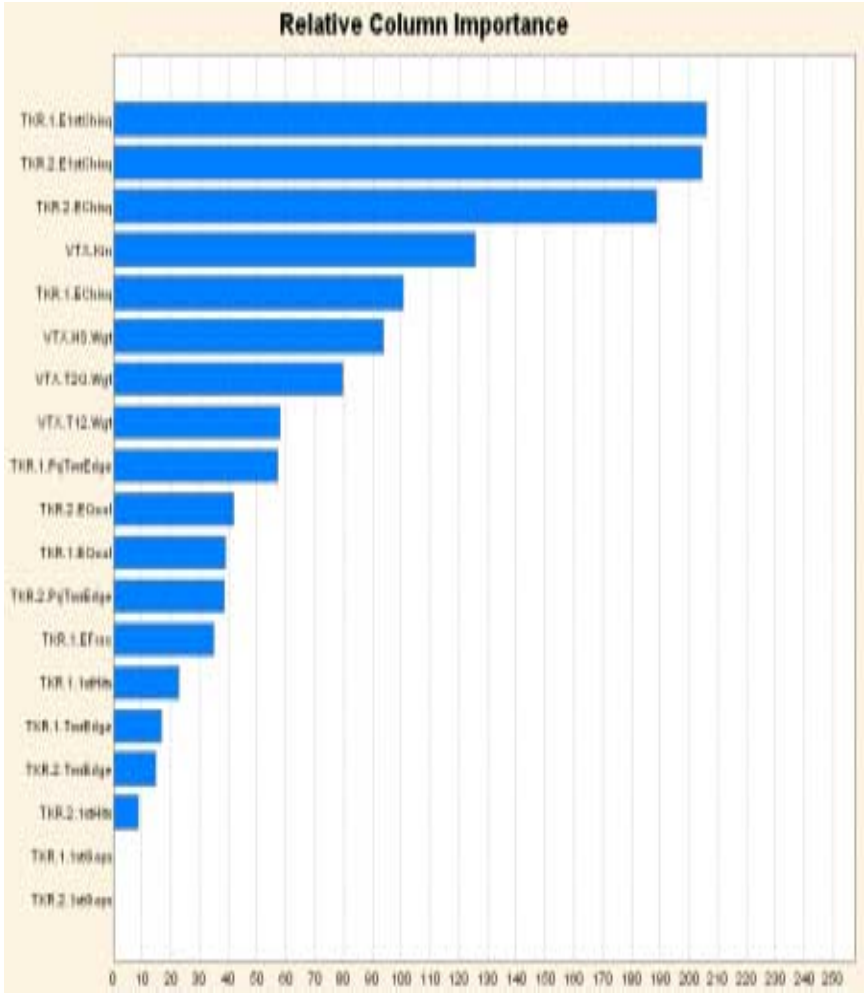
The VTX Classification Tree

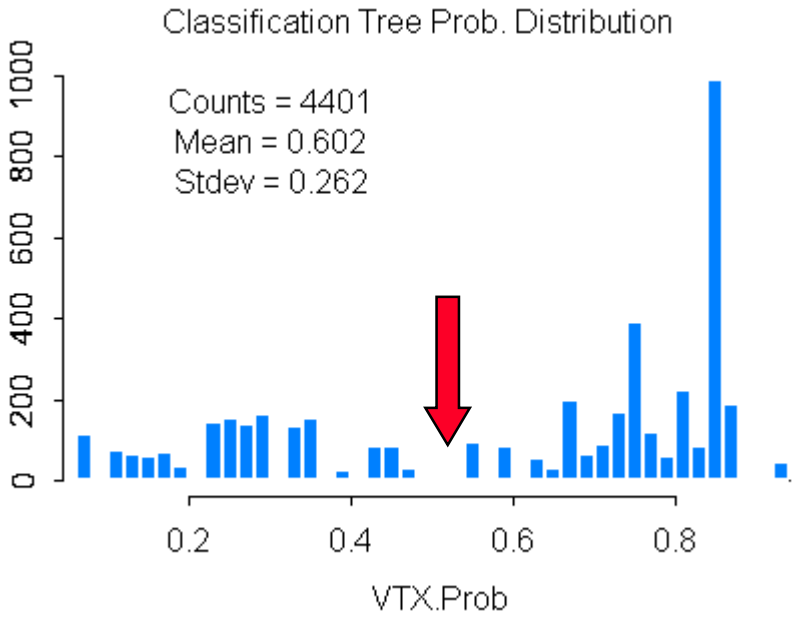






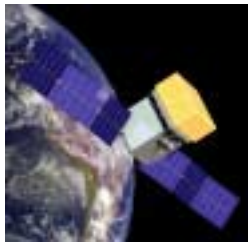
CPA: To Vertex or not to Vertex?



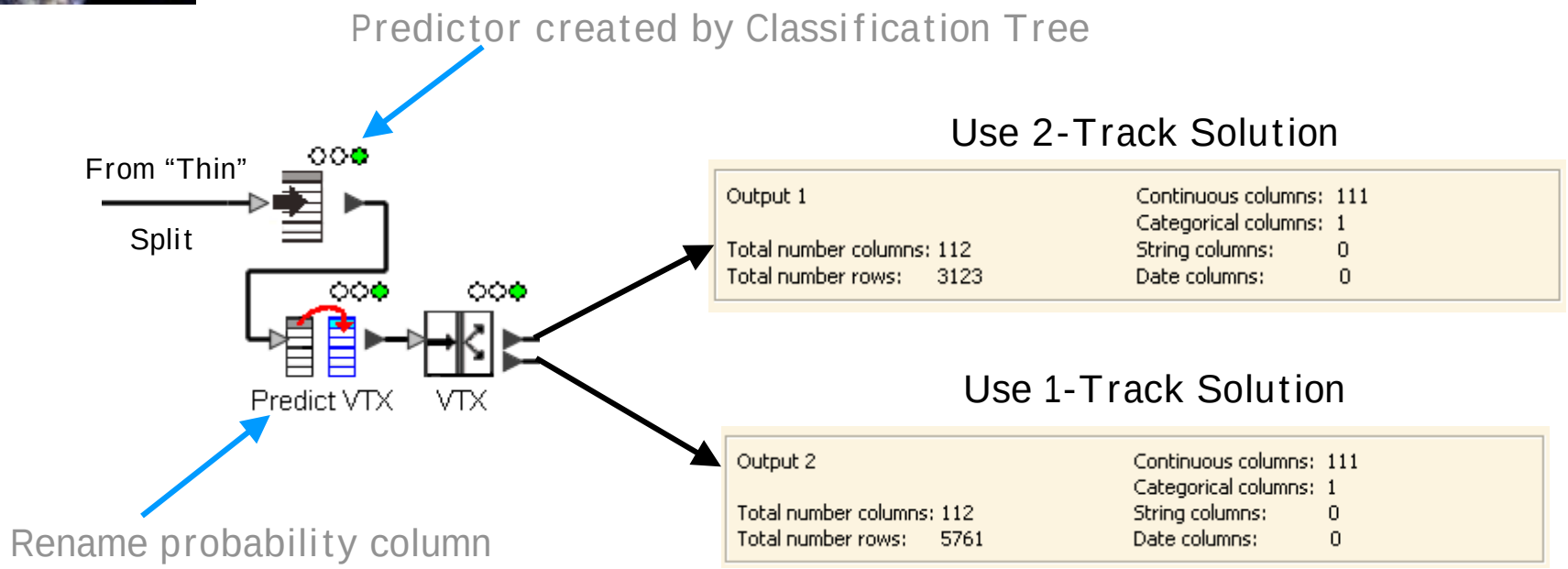


Probability is not continuous – its essentially binned by the finite number of leaves (ending nodes)

There is a “gap” at .5 - Use that to determine which solution to use

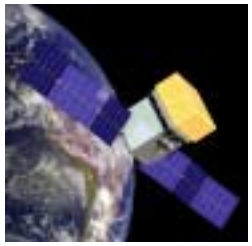


Do the Vertex Split!



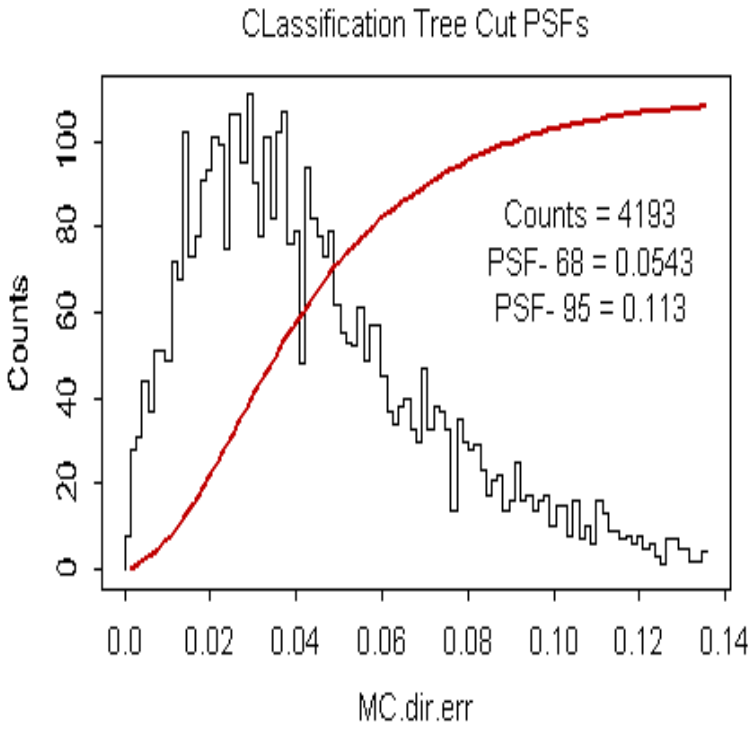
The data are now divided into 2 subsets according to the Probability that the 2-Track ("vertex") solution is best.

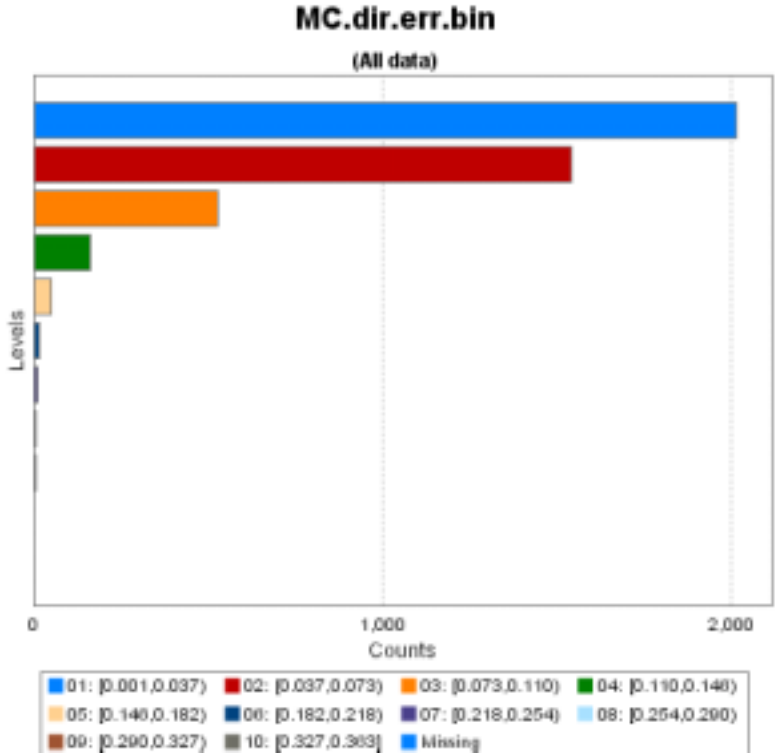
No data have been eliminated – Failed Vertexed solutions Are tried again as 1-Track events



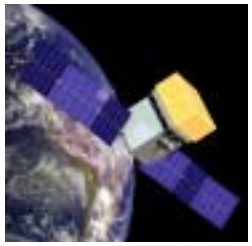
Bin the PSF

Continuous Variable → Catagroical Variable

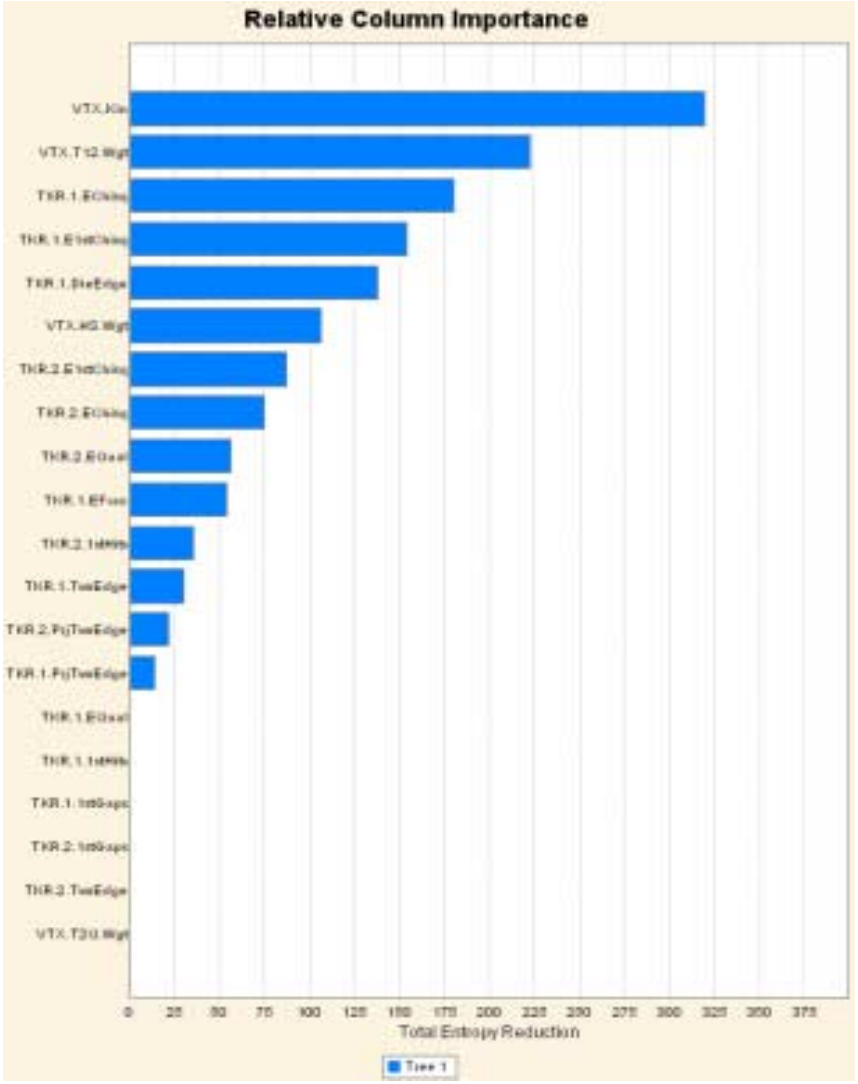


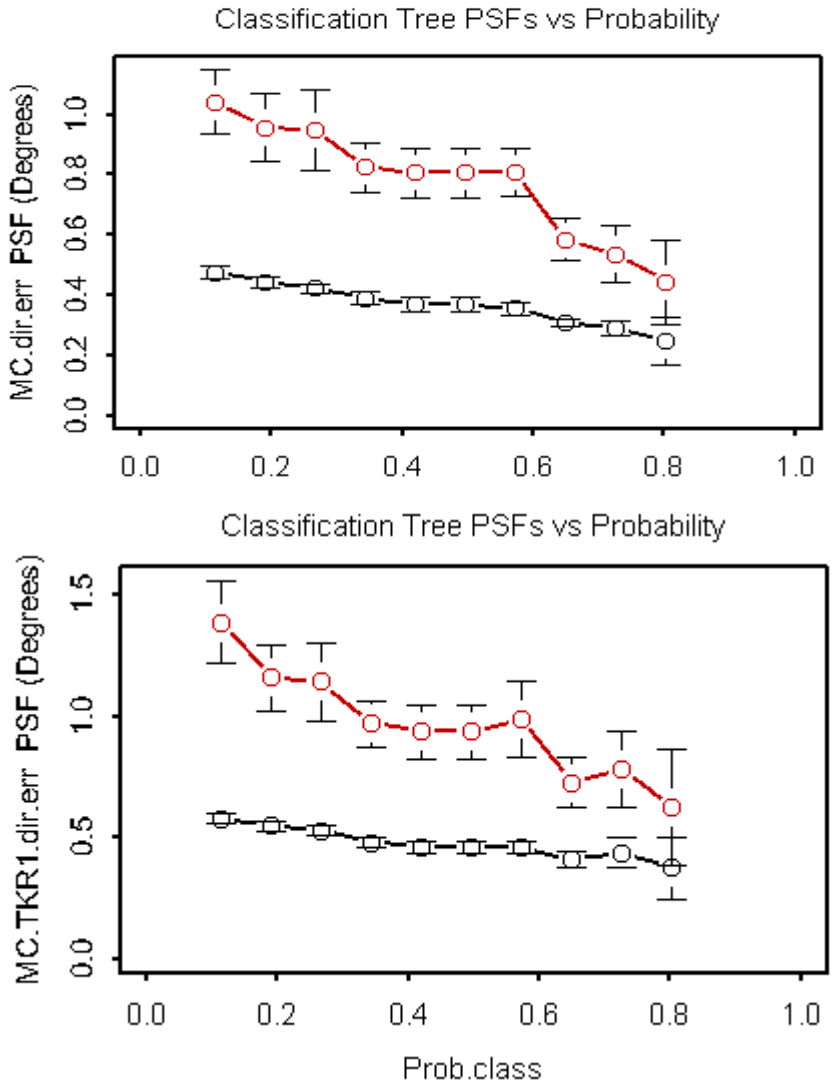


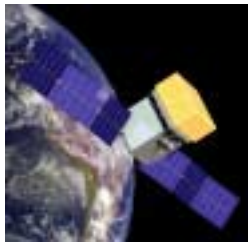
Target Class: Class #1 – MS PSF Limited Bin



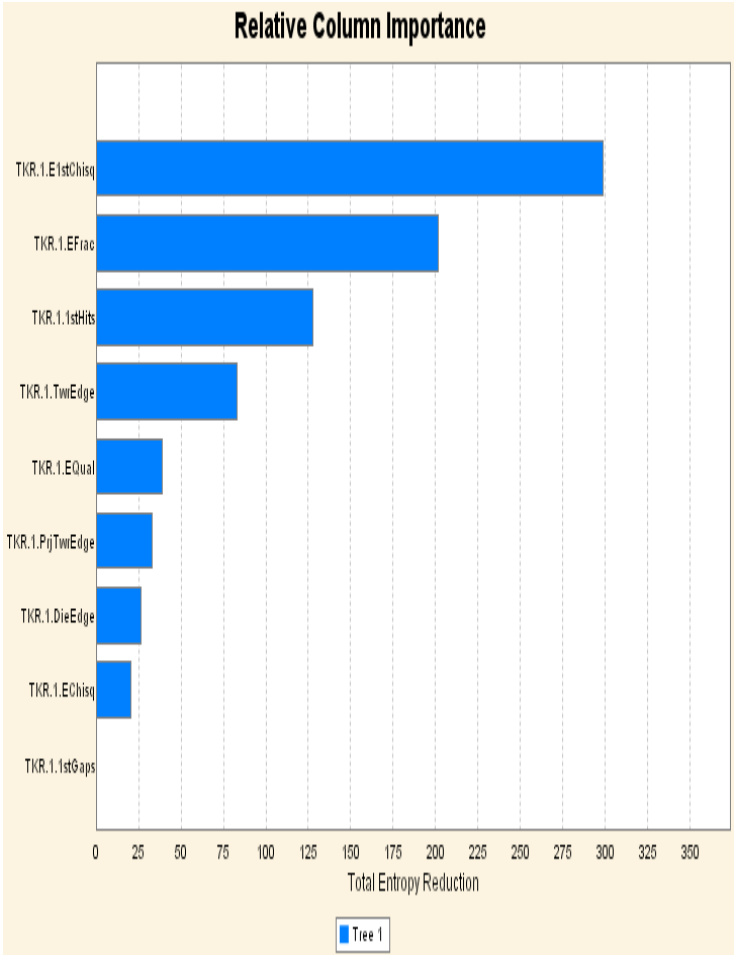
2 Track Classification Tree

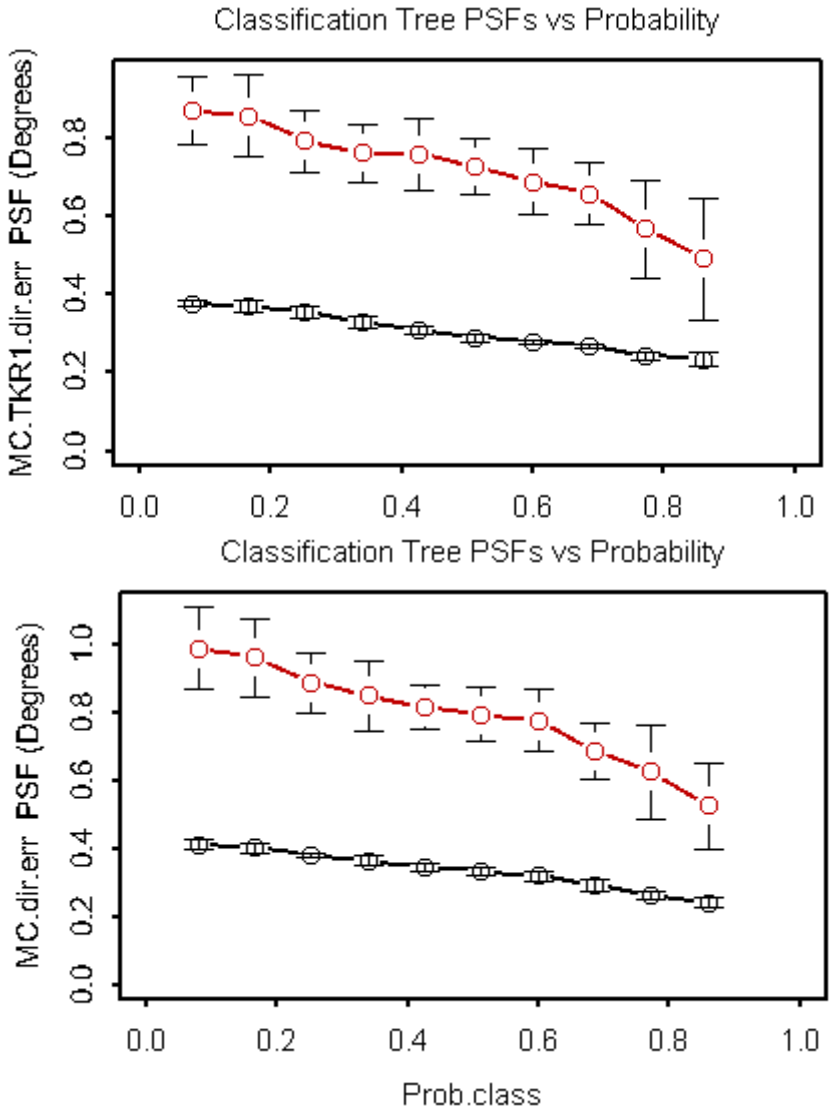


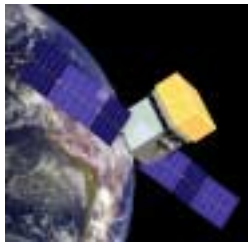




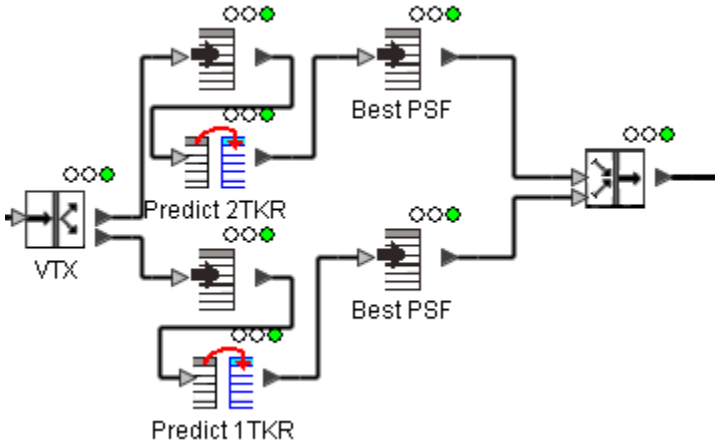
1 Track Classification Tree

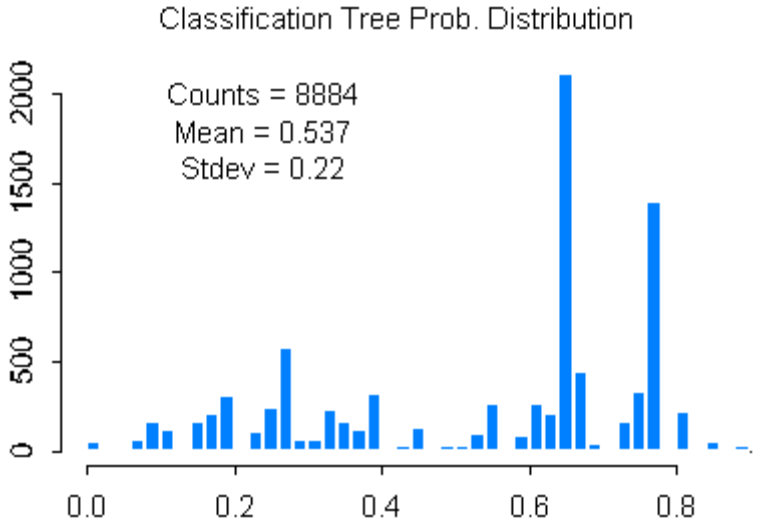


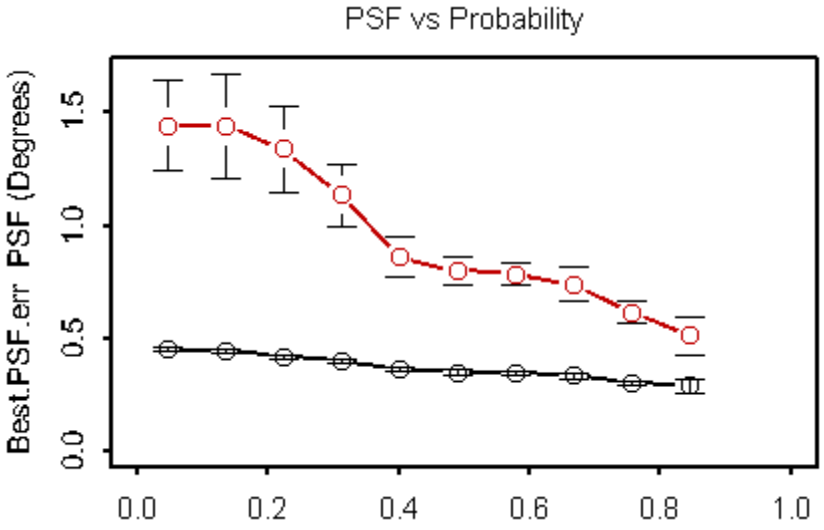


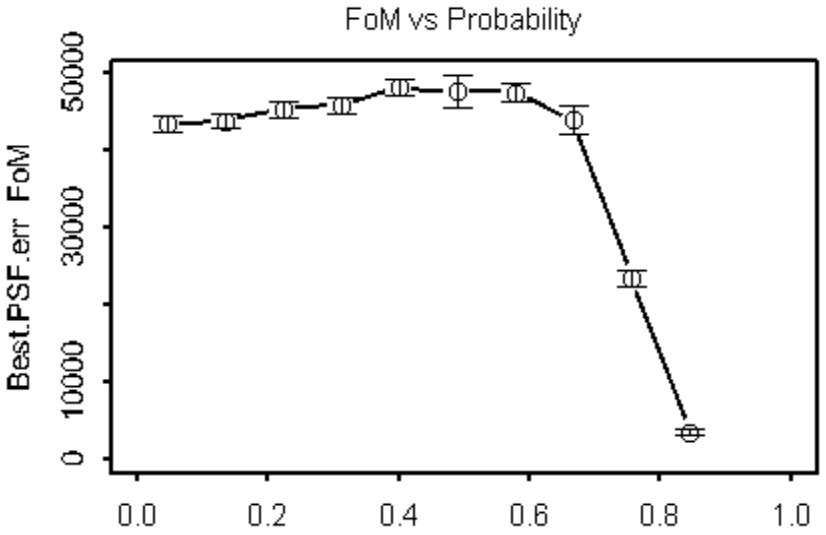


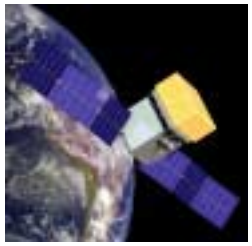
Combining Results





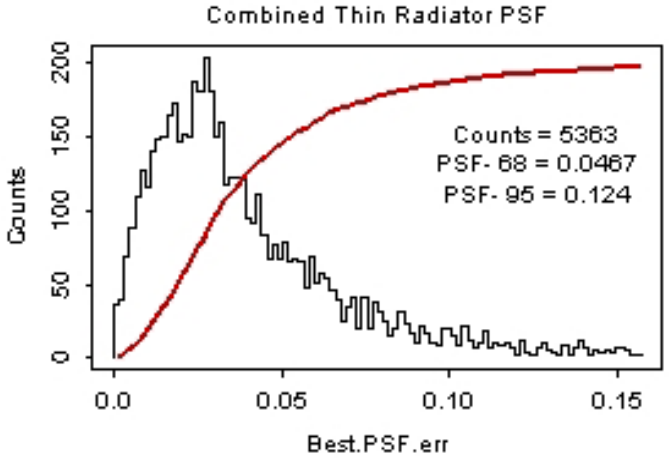


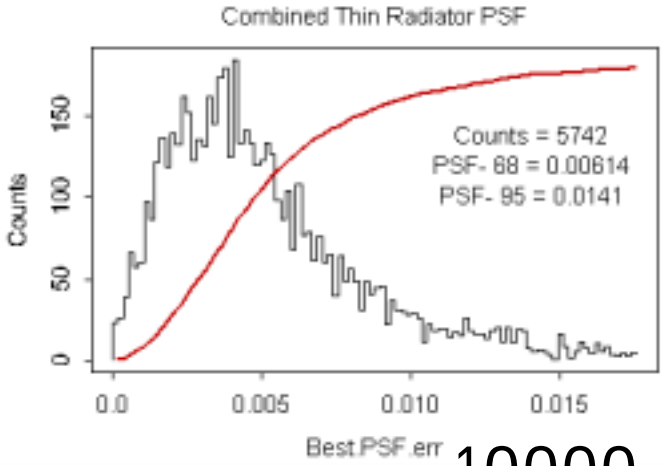




Example PSF's At FoM Max

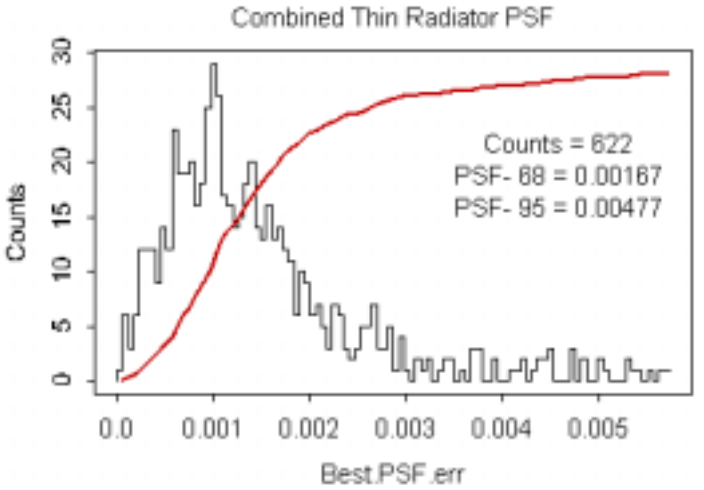
100 MeV
PSF-68 =2.7°
95/68 = 2.65

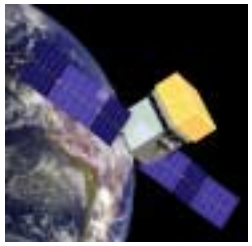




1000 MeV: PSF-68 = .35°
95/68 = 2.3

10000 MeV :
PSF-68 = .1°
95/68 = 2.9





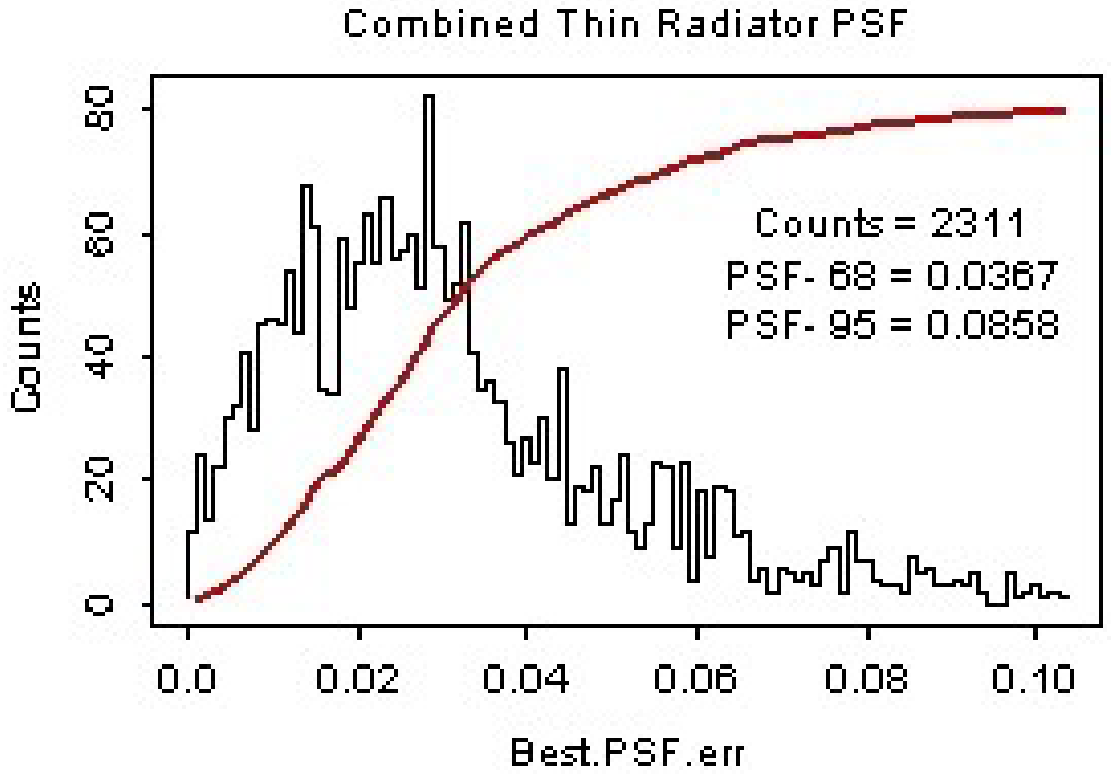
Before and After Trees

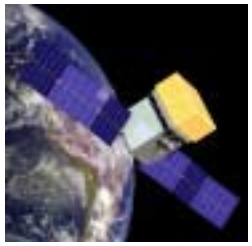
Using Classification Trees

PSF: 2.1°

95%/68% :2.34

A_{eff}: 1387 cm²





Before and After Trees

Best results o Using Classification Trees