A GLAST Analysis

Agenda

- Overarching Approach & Strategy
- Flattening Analysis Variables
- Classification Tree Primer
- Sorting out Energies
- PSF Analysis
- Background Rejection
- Assessment

Strategy

Terminology and GLAST Phase space:

Light Gathering Power: $A_{eff} \times \Delta\Omega$

GLAST S.R.: 8000 cm^2 x 2.0 str = 16000 cm^2-str
Goal: 10000 cm^2 x 2.4 str = 24000 cm^2-str
Triggerable: 19630 x .65 x 2.4str = 30600 cm^2-str
EGRET: ~ 1000 cm^2 x .6 str = 600 cm^2-str

Input Data:

All Gamma: 18 MeV – 18 GeV into 6 m^2 x 2π str (= 37.7 m^2-str)
Energy Spectrum: 1/E (Flat in Log(E))
"Pre-Cuts": AcdActiveDist < -20 mm & TkrNumTracks > 0
Background: Generic On-Orbit Mix - same $A_{eff} \times \Delta\Omega$

Variables:

To cover GLAST Phase space – make variables independent of Energy and $\cos(\theta)$
Alternative: make analysis "cuts" energy and angle dependent

Key Methodology: Classification Trees

Strategy 2

Game Plan:

1) Flatten important variables used in the analysis

2) Use CT technology to determine events with "well measured" energies

3) Use CT technology to determine events with "well measured" directions

4) Filter background events (BGE's) and γ's through the above CT scripts and form training and testing samples for background rejection

5) Use CT technology to separate γ's from BGE's

Flattening the Variables

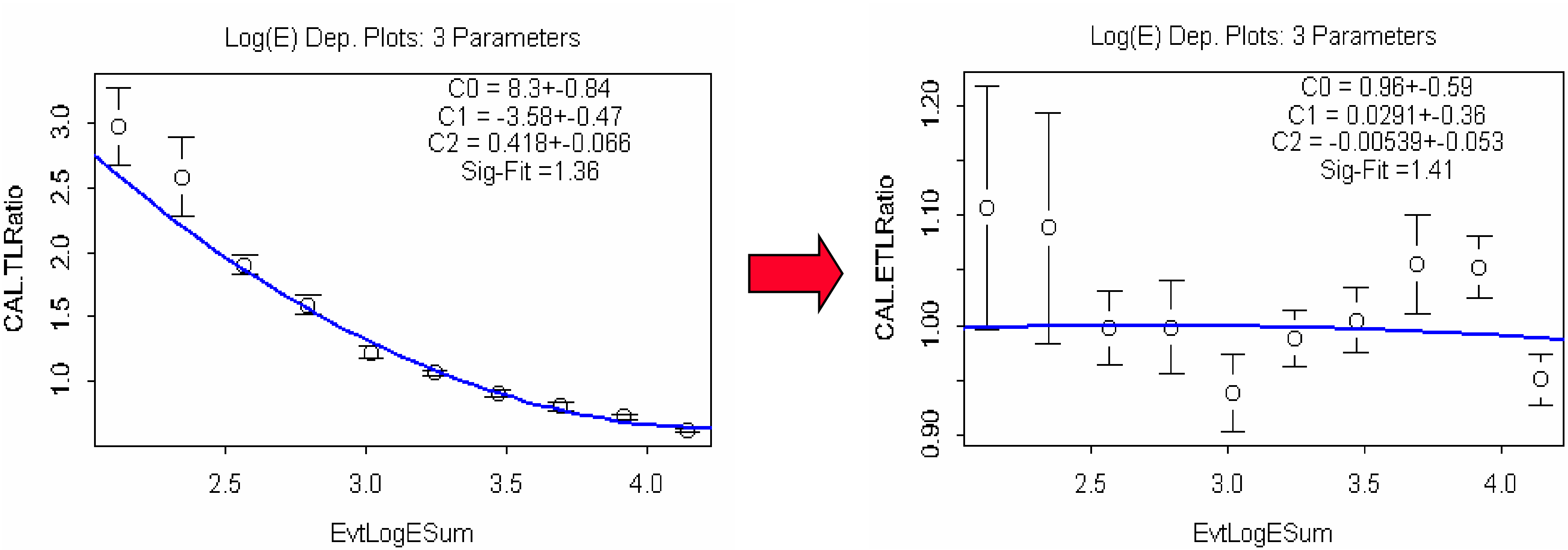
Many analysis variables vary (albeit) slowly with energy and cos(θ).

Assume averages can be modeled by $\langle v_i \rangle = f\left(\log(E)\right) \cdot g(\cos(\theta))$

Least Squares Fit to 2nd order.

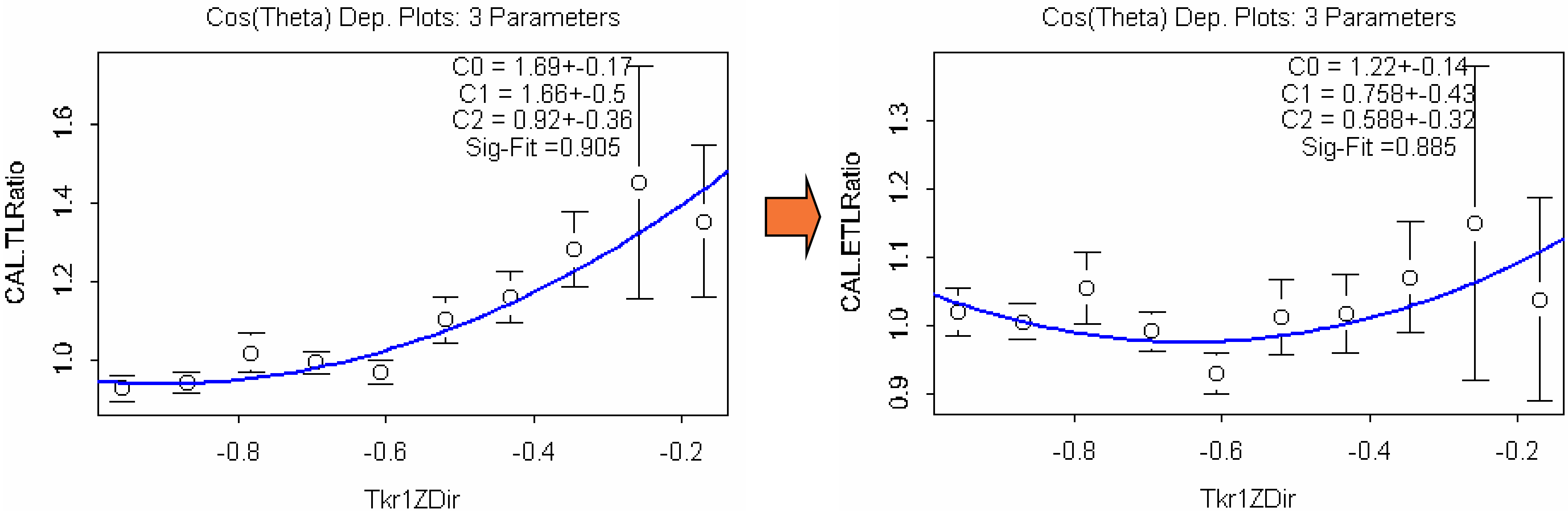
First do *log(E)* dependence:

$$f(\log(E)) = c_0 + c_1 \cdot \log(E) + c_2 \cdot \log(E)^2$$



Flatten 2

Next do cos(θ): $g(\cos(\theta)) = c_0 + c_1 \cdot \cos(\theta) + c_2 \cos(\theta)^2$



Variables which have been "flattened" include:

Tkr1Chisq	Tkr2Chisq	EvtTkrComptonRatio	EvtCalXtalTrunc
Tkr1FirstChisq	Tkr2FirstChisq	EvtCalTLRatio	EvtCalTrackDoca
Tkr1Qual	Tkr2Qual	EvtCalXtalRatio	EvtCalTrackSep
EvtVtxEAngle	EvtVtxDoca	EvtVtxHeadSep	

Classification Tree Primer

Origin: Social Sciences - 1963

How a CT works is simple:
A series of "cuts" parse the data into a "tree" like structure where final nodes (leaves) are "pure"

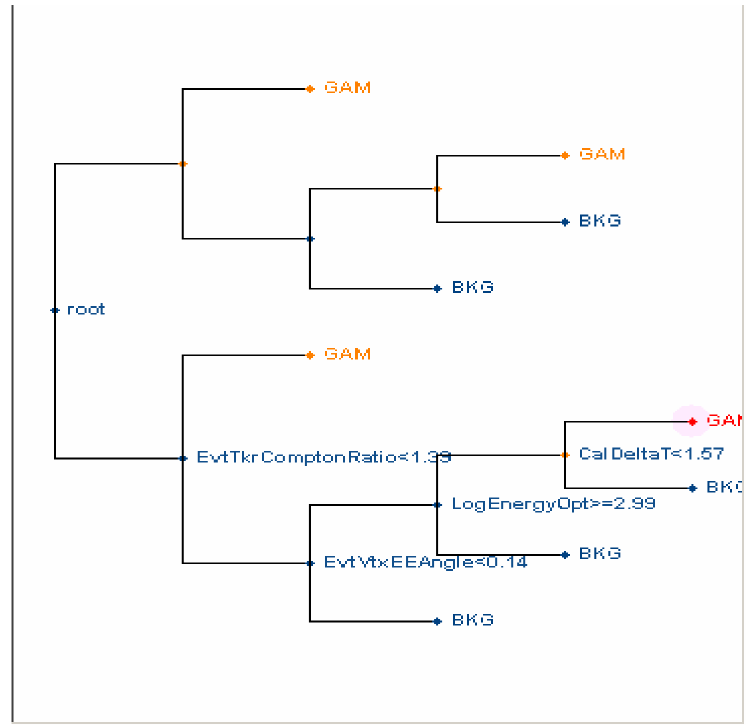
How the Cuts are determined is harder:
(Called Partitioning)
Total Likelihood for a tree is:

$$L = \prod_{i-leaves} \prod_{k-classes} p_{ik}^{n_{ik}}$$

where p_{ik} are the probabilities and n_{ik} are the number of events. For each node define a deviance

$$D_i = -2 \sum_{k-classes} n_{ik} \log(p_{ik})$$

Splitting node *i* into two smaller nodes *s* & *t* Results in a reduction in deviances given by



$$D_s - D_t - D_u = 2 \sum_{k-classes} \left[n_{tk} \log(\frac{p_{tk}}{p_{sk}}) + n_{uk} \log(\frac{p_{uk}}{p_{sk}}) \right]$$

where $n_{sk} = n_{tk} + n_{uk}$

Tree Primer (2)

The probabilities are not know *a priori* so the event counts in the training sample are used.

Example: $\tilde{p}_{tk} = \dfrac{n_{tk}}{n_t}$

From this the value of a split can be determined by

$$D_s - D_t - D_u = 2\left[\sum_{k-classes} \left(n_{tk} \log n_{tk} + n_{uk} \log n_{uk} - n_{sk} \log n_{sk} \right) + n_s \log n_s - n_t \log n_t - n_u \log n_u \right]$$

Note that splitting nodes with large numbers of events is favored.

Splitting of each node continues until change in deviance is too small or the number of events in the node has fallen below a minimum.

Tree construction is a “look one step ahead” process – it does not necessarily find the ultimate optimal tree.

Trees readily adapt to the “training” data if the event count in the leaves or the deviance reduction at each split is allowed to be too small.

Sorting Out the Energies

Energy Types: *Percentages:*

1) No CAL Events: < 5 MeV OR < 2 r.l. in CsI 46%
2) Low CAL Events: < 100 MeV 13%
3) High CAL Events: > 100 MeV 41%

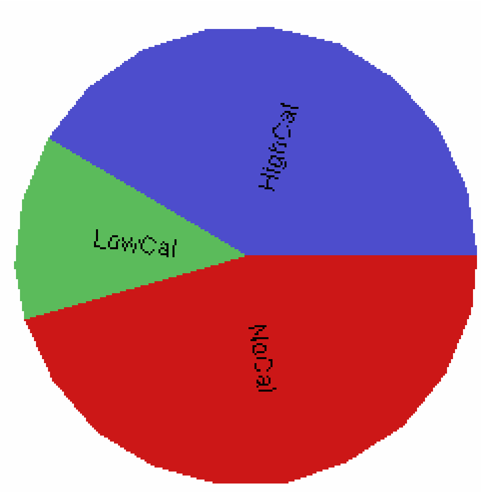
Good Energy Definition

Model: $$\frac{\Delta E}{E} = \frac{E_{Obs} - E_{MC}}{.8E_{MC} + 40MeV}$$

(Maps energy errors onto a common scale.
Example: for σ_{Energy} = .1 (GLAST Nominal)
$\Delta E_{100\ MeV}$ = 12 MeV & $\Delta E_{1000\ MeV}$ = 84 MeV)

No CAL: -.4 < $\Delta E/E$ < 1.5 (-60% + 150%)
Low/Hi CAL: -.5 < $\Delta E/E$ < .5 (+- 50%)

Break Down of Energy Classes

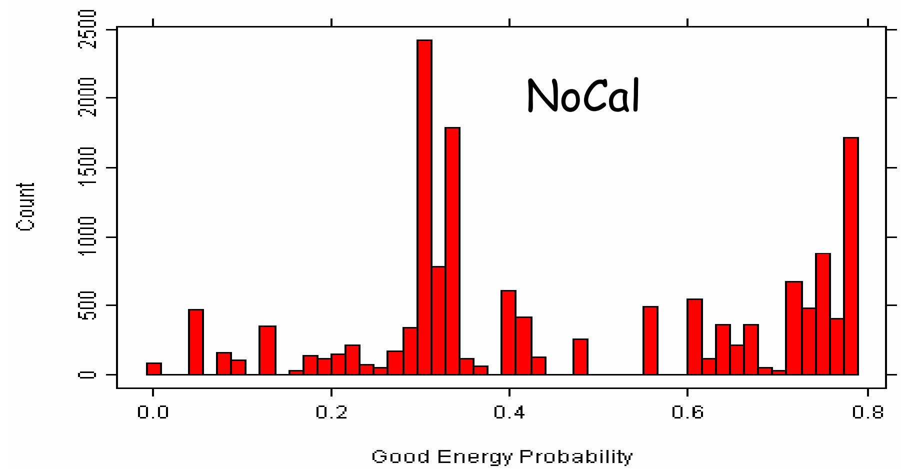


Energy Classes

Input Node - NO CAL (422)

		Predicted		Totals
		GoodEnergy	*BadEnergy*	
Observed	*GoodEnergy*	4496	2609	7105
	BadEnergy	1841	6410	8251
Totals		6337	9019	15356

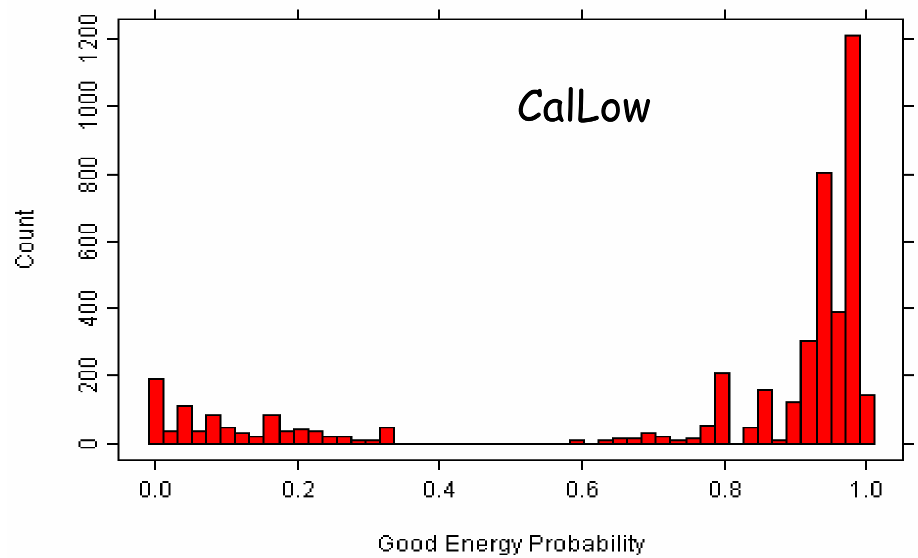
	Observed		Overall
	GoodEnergy	*BadEnergy*	
% Agree	*63.3%*	*77.7%*	*71.0%*



Input Node - CAL Low (419)

		Predicted		Totals
		GoodEnergy	*BadEnergy*	
Observed	*GoodEnergy*	3307	94	3401
	BadEnergy	266	758	1024
Totals		3573	852	4425

	Observed		Overall
	GoodEnergy	*BadEnergy*	
% Agree	*97.2%*	*74.0%*	*91.9%*

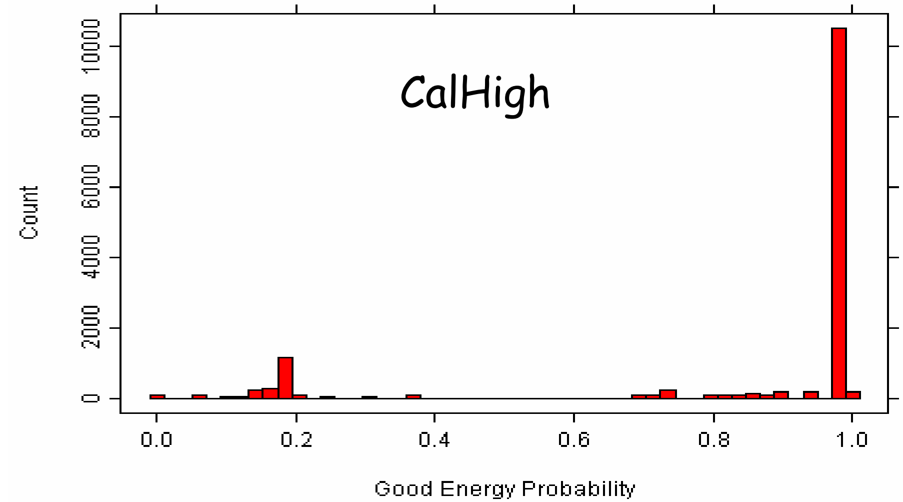


Energy Classes

Input Node - CAL High (361)

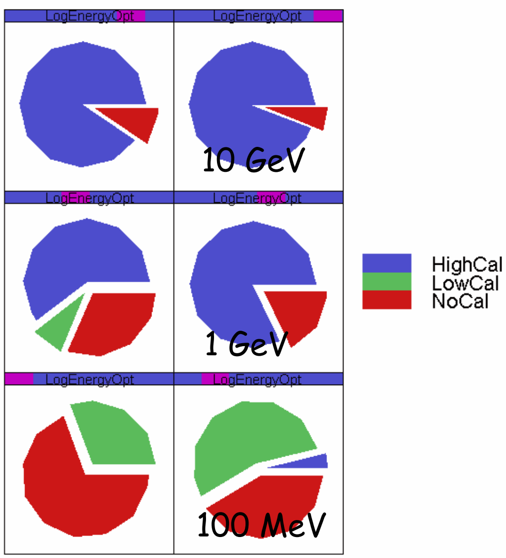
		Predicted		Totals
		BadEnergy	*GoodEnergy*	
Observed	*BadEnergy*	1731	387	2118
	GoodEnergy	355	11508	11863
Totals		2086	11895	13981

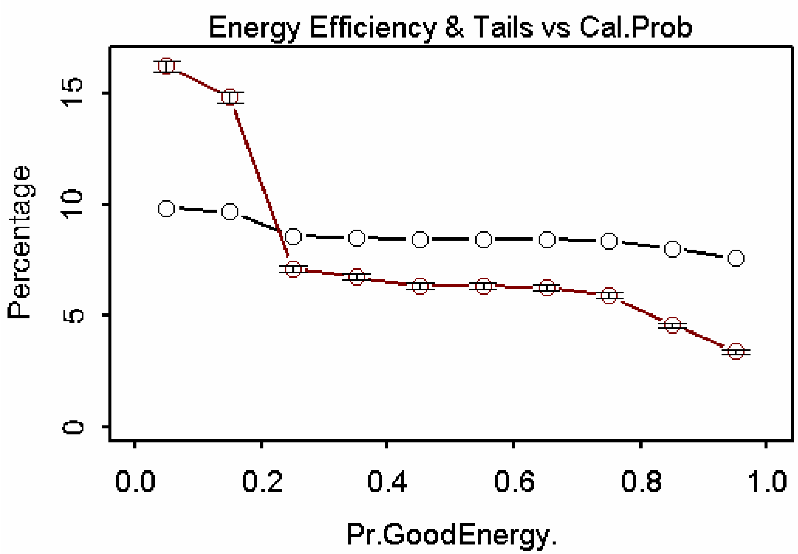
	Observed		Overall
	BadEnergy	*GoodEnergy*	
% Agree	*81.7%*	*97.0%*	*94.7%*



Energy Class Break Down

Prob. > 50%



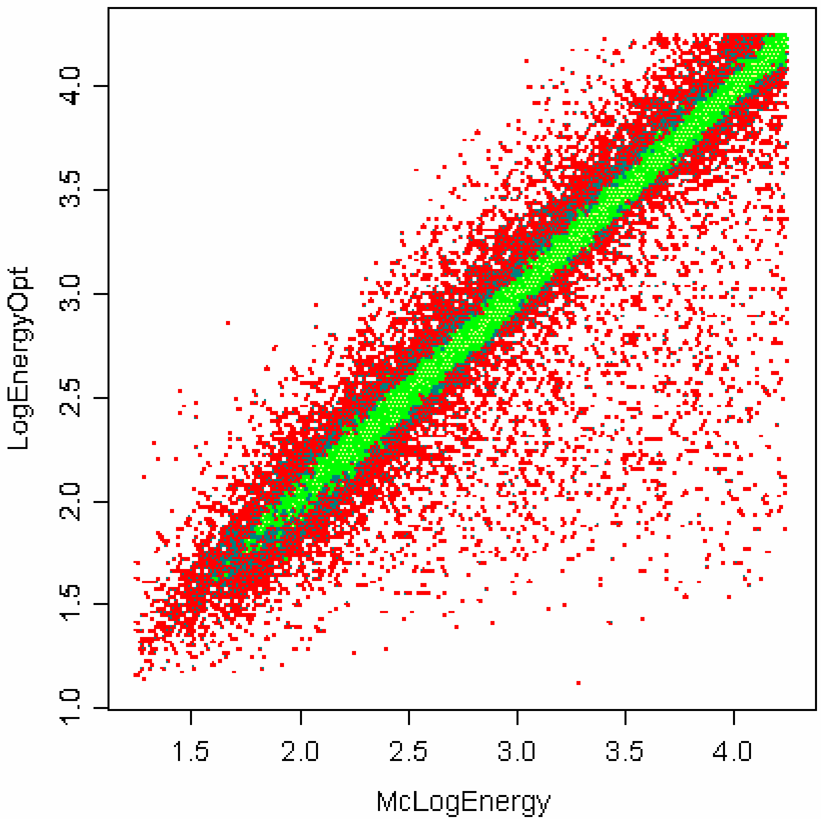


Energy Summary

Low & High CAL Classes

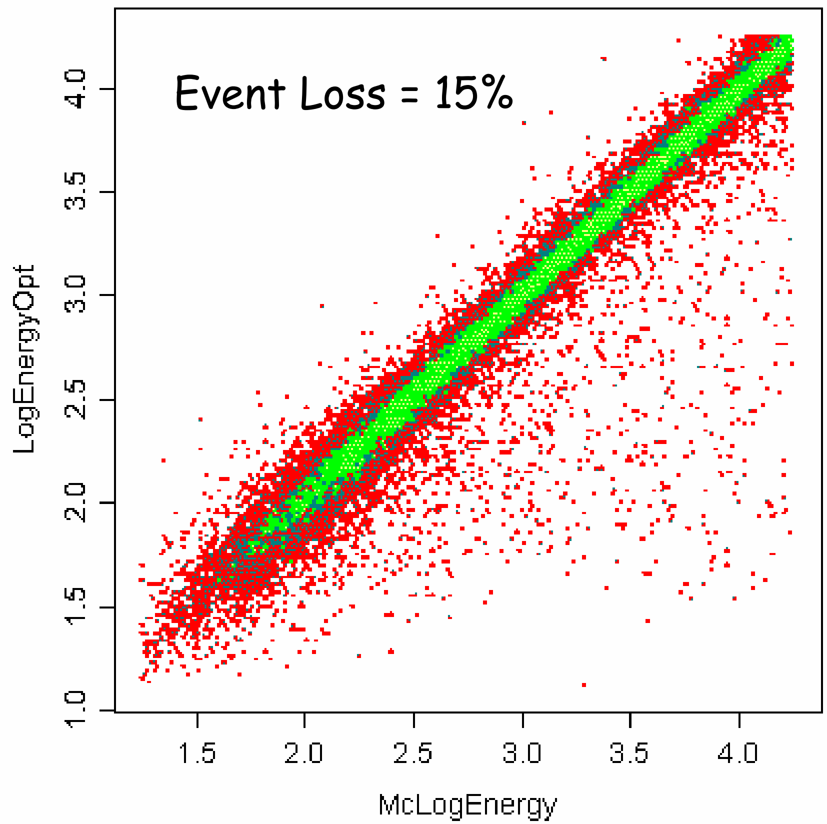
No Probability Cut

$A_{eff} \times \Delta\Omega = 2.74$ m^2-str

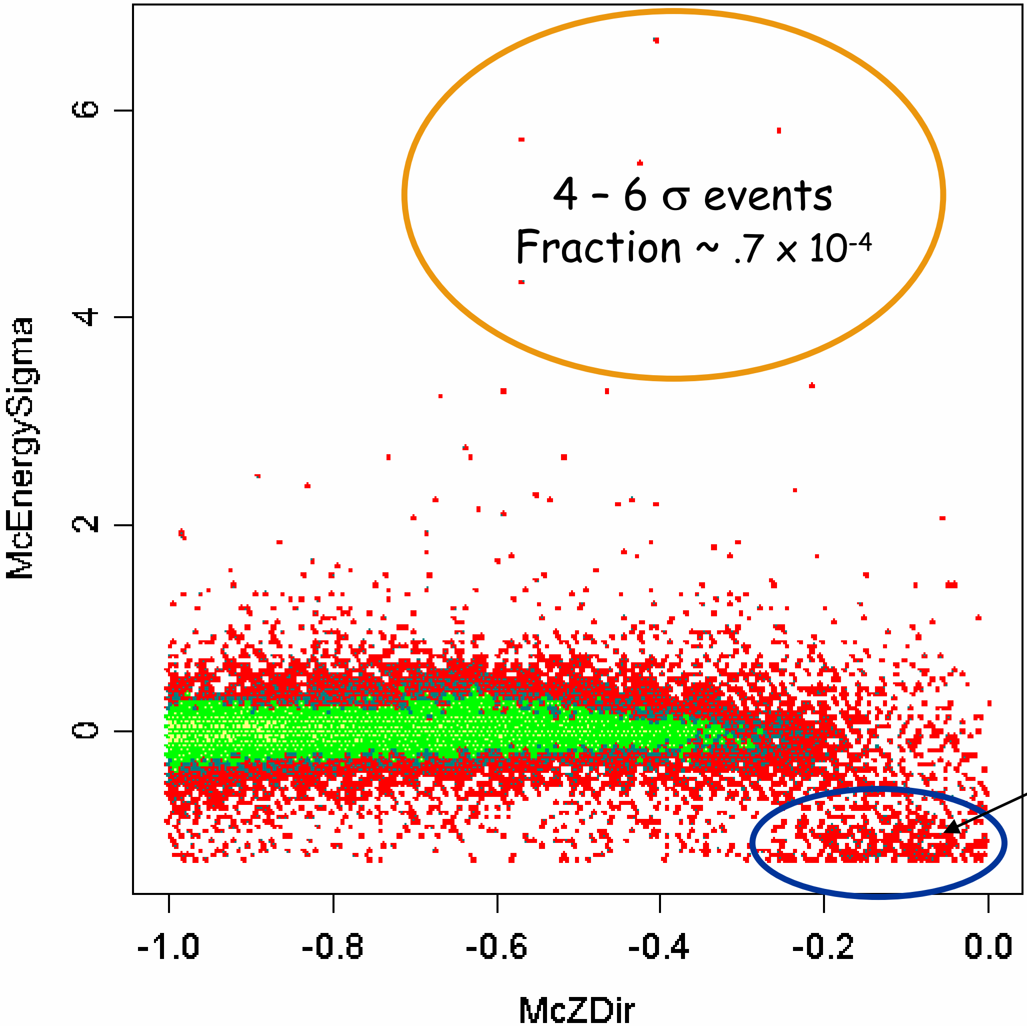


Prob. > .50

$A_{eff} \times \Delta\Omega = 2.33$ m^2-str



Energy Summary (cont')



Energy is "FLAT" in dimensions of θ and E.

Remaining $A_{eff} \times \Delta\Omega = 2.33$ m^2-str

Remaining "Bad Energy": 6.3%

Remaining "Good Energy": 84%

Horizontal Events - Not so easy to remove at this stage.
Note: This is where they are generated – NOT where they are reconstructed

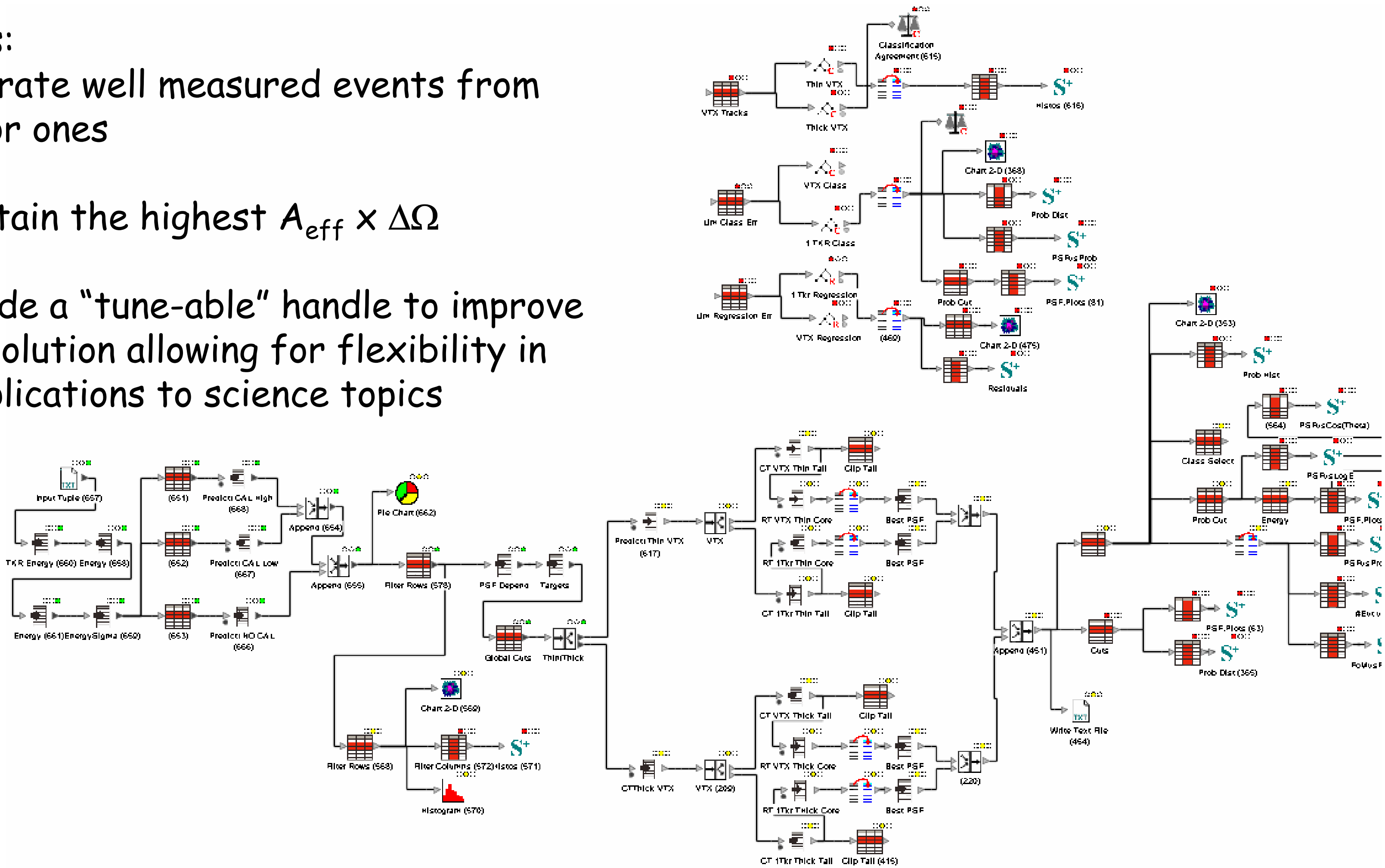
PSF Analysis

Goals:

Separate well measured events from poor ones

Maintain the highest $A_{eff} \times \Delta\Omega$

Provide a "tune-able" handle to improve resolution allowing for flexibility in applications to science topics



GLAST

PSF Classes

<u>Conversion Location</u>: Thick & Thin

First hit occurs in Thin radiator section → Thin

First hit occurs in Thick radiator section → Thick

<u>Analysis Type</u>: VTX & 1Trk

> 50% of Events have a "VTX" solution

(VTX solution → 2 tracks combined to give γ direction)

VTX Solution not always better than the "Best Track Solution"

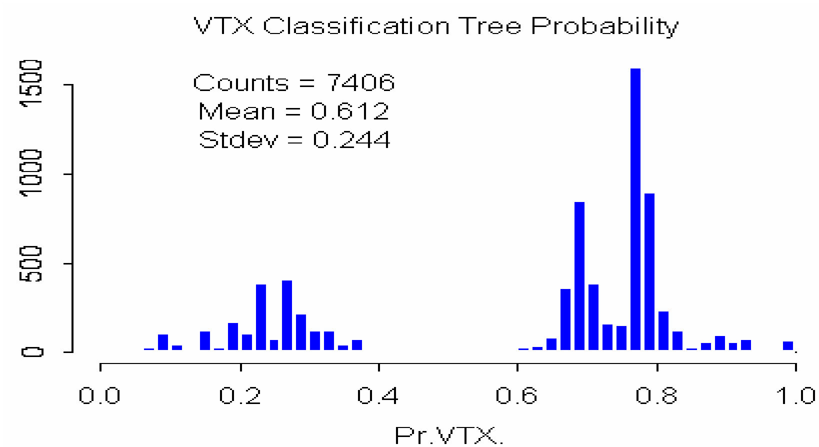
Types sorted out via a Classification Tree

4 PSF Classes x 3 Energy Classes

The VTX Decision

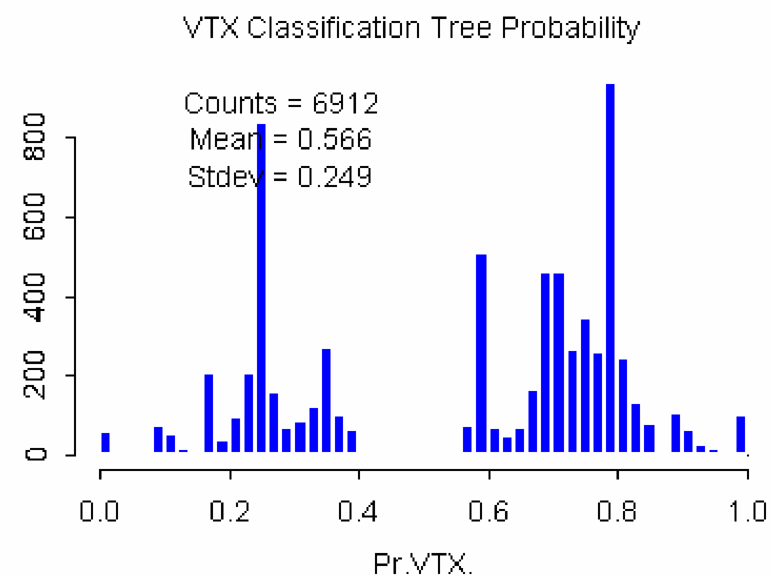
Input Node - Thin VTX (259)		Predicted		Totals
		1TKR	VTX	
Observed	1TKR	1595	1277	2872
	VTX	503	4031	4534
Totals		2098	5308	7406

	Observed		Overall
	1TKR	VTX	
% Agree	55.5%	88.9%	76.0%



Input Node - Thick VTX (259)		Predicted		Totals
		1TKR	VTX	
Observed	1TKR	1843	1159	3002
	VTX	622	3288	3910
Totals		2465	4447	6912

	Observed		Overall
	1TKR	VTX	
% Agree	61.4%	84.1%	74.2%

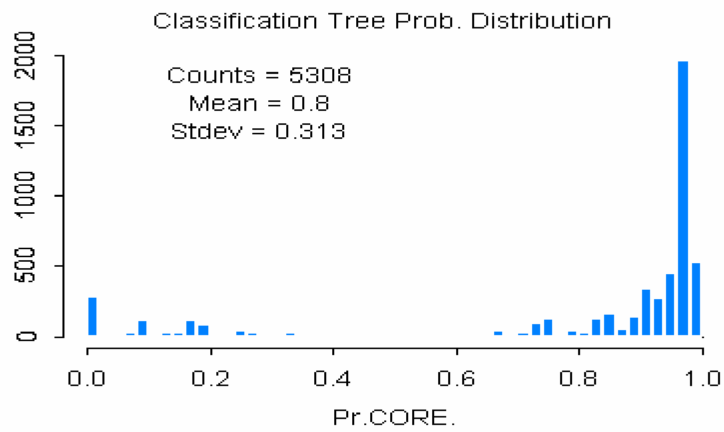


VTX Thin

Input Node - (231)

		Predicted		Totals
		CORE	TAIL	
Observed	CORE	4569	81	4650
	TAIL	154	485	639
Totals		4723	566	5289

	Observed		Overall
	CORE	TAIL	
% Agree	98.3%	75.9%	95.6%

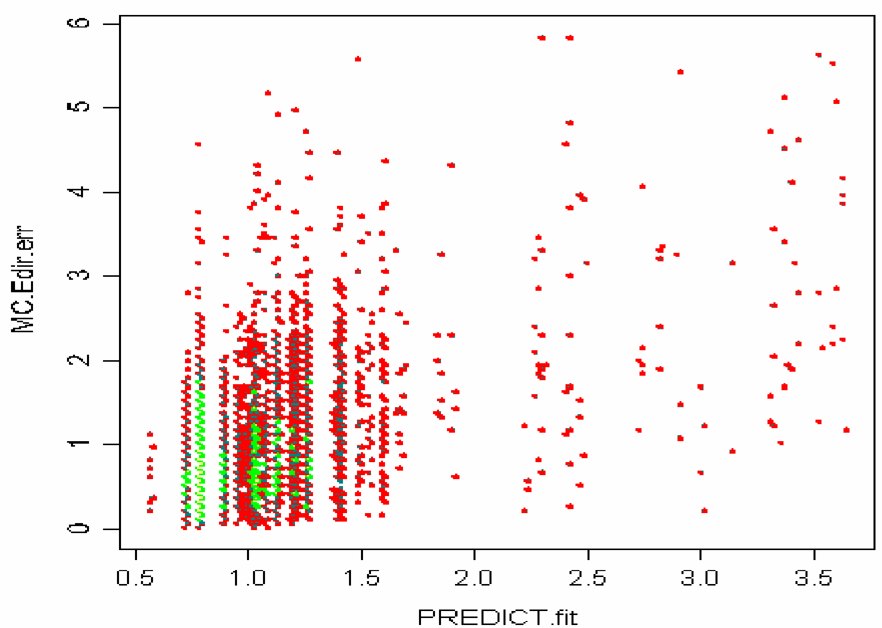


Clip Bad Events using a CT

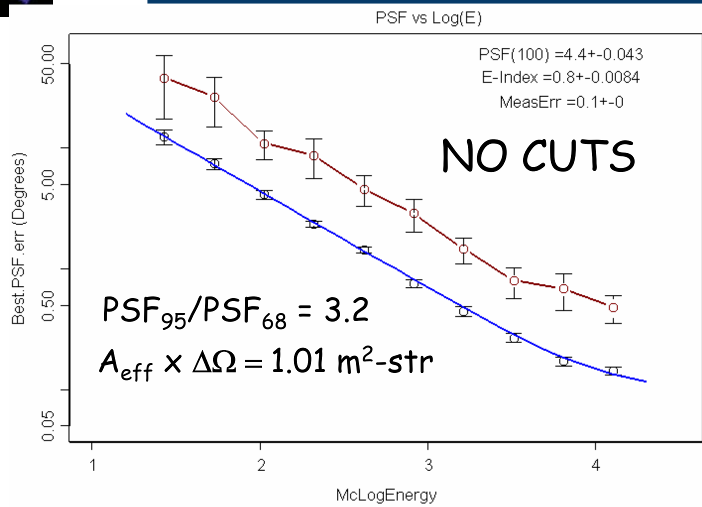
Predict how "good" using a

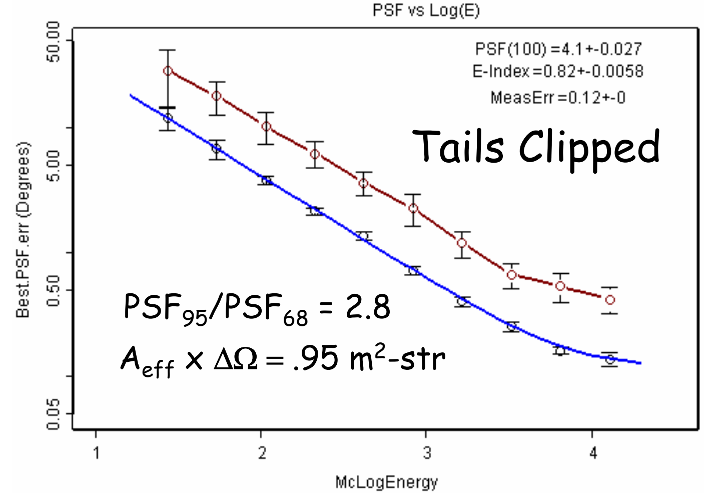
Regression Tree

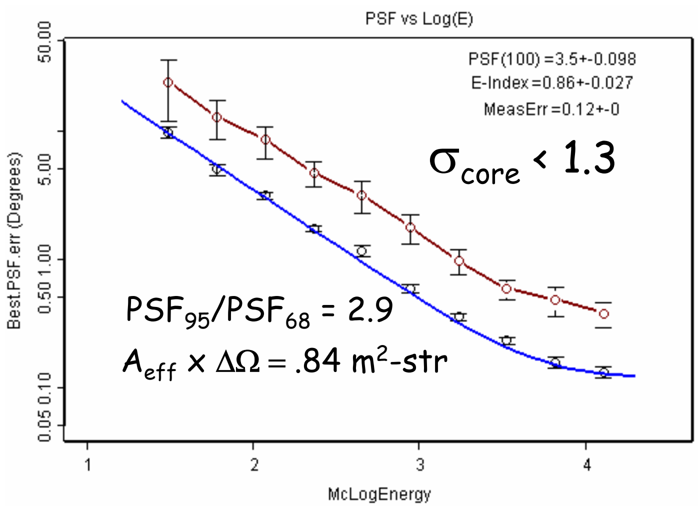
This process is repeated for the 4 Tracking Event Classes

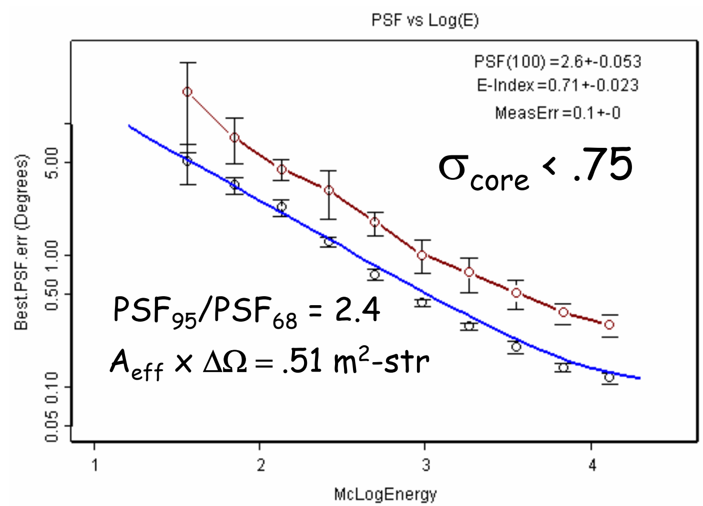


PSF Results – Thin Radiator







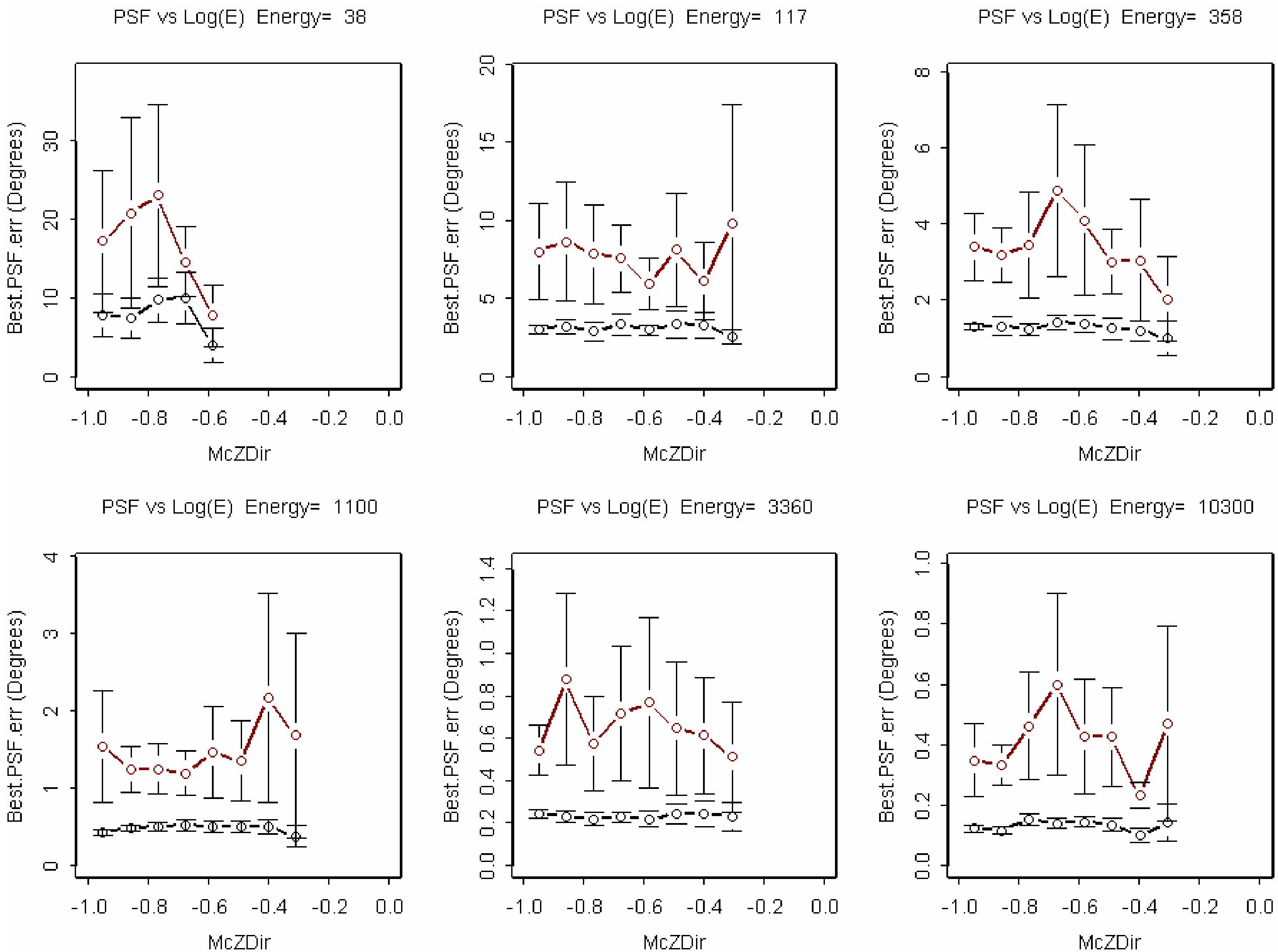


Thin Radiator PSF 2

Cos(θ) Dependence

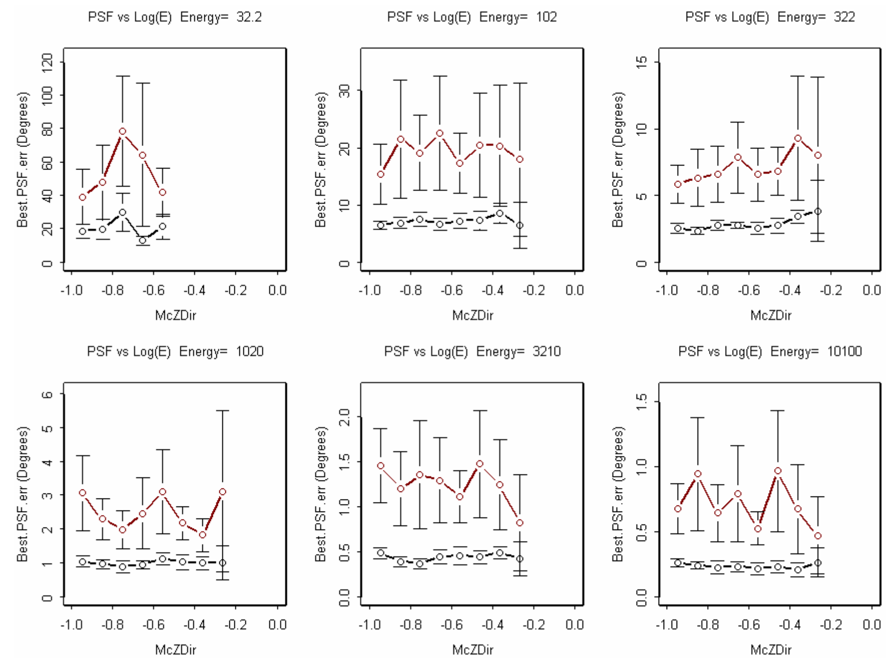
Cuts:

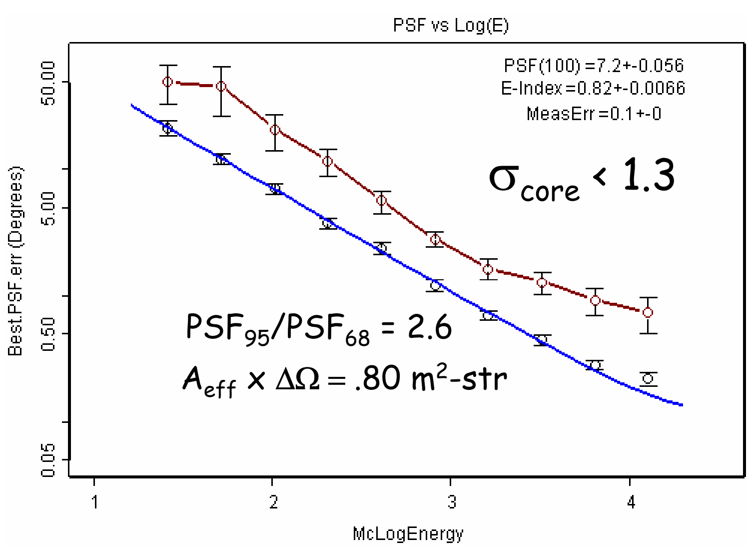
1) Tails Clipped
2) σ_{core} < 1.3



PSF Results – Thick Radiator

Thick Radiator Events:		Expect
1)	Similar to Thin	yes
2)	ΔA_{eff} ~95% A_{eff}(Thin)	76%
3)	~2 x worse PSF	2.1 x

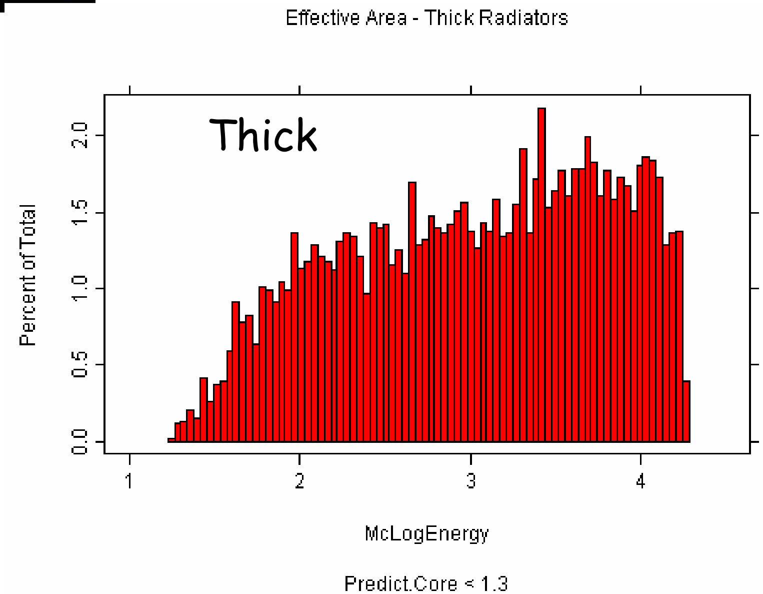


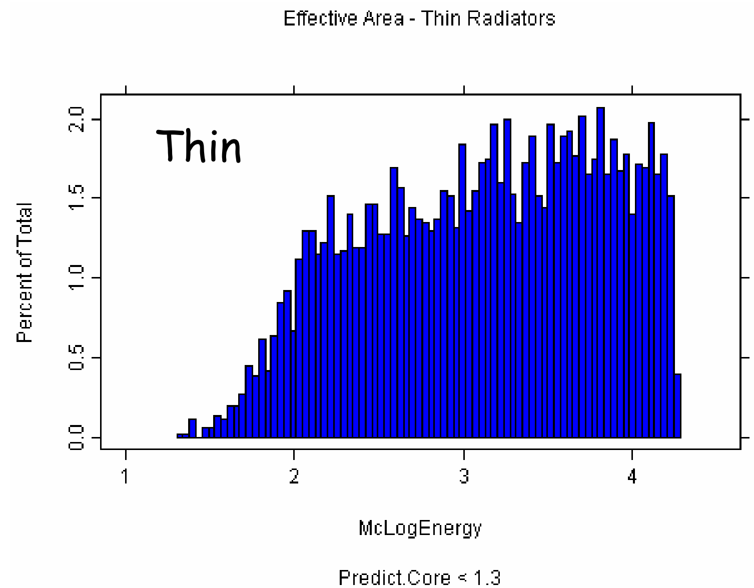


At high energy PSF_{thick} → PSF_{thin}

Multiple Scattering becomes less important then measurement errors.

PSF Results – What Remains





$A_{eff} \times \Delta\Omega$ distributions approximately the same

SR case (σ_{core} < 1.3): $A_{eff} \times \Delta\Omega = .80\ m^2$-str + $.84\ m^2$-str = $1.64\ m^2$-str

Ratio of Integral log(E) plots to flat (as generated) distribution: ~ 1.8

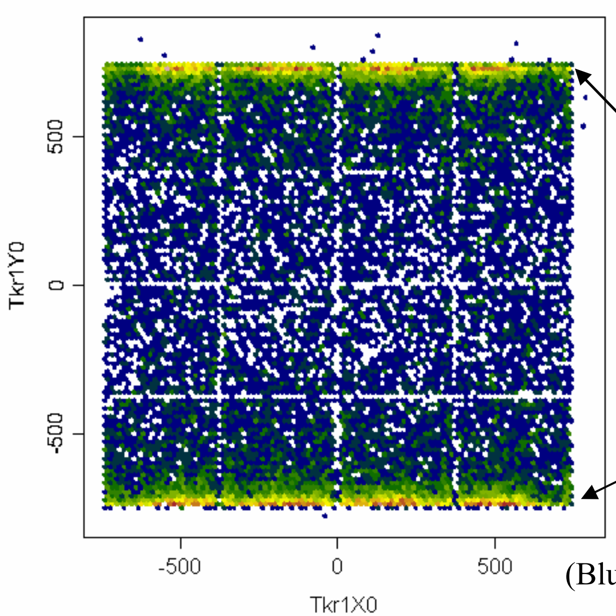
Hence Asymptotic $A_{eff} \times \Delta\Omega = 2.94\ m^2$-str (lots of light gathering power left)

Background Rejection

Goal: remove most of the BGE's while preserving the γ signal

Problem: Large imbalance between #BGE's and # γ's.
CT's need sufficient #'s of events to establish unbiased model trees.

Show Stopper: 11th hour discovery of problems in ACD Sim & Analysis



BGE Data Set

AcdTileCount = 0

No Side Tiles Fired!
The events pour in!

Also there's trouble with Top Tiles as well!

(Blue ~ 1/pixel, Brown ~ 50/pixel)

Forge Ahead! *(DTFSA)*

The Formal portion of the talk is now ended!

What lies ahead is presented to show the direction which is being pursued.

All the quantitative results are given as illustrative only!

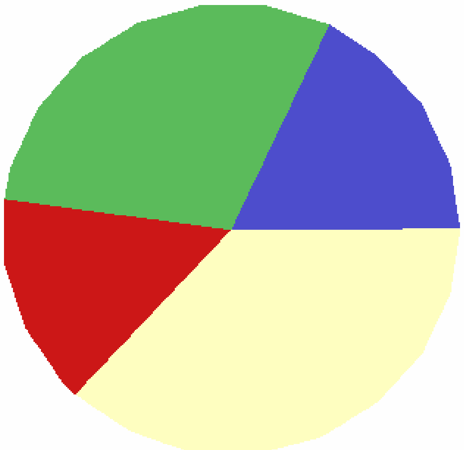
IN SHORT: QUOTE NOTHING FROM THIS!

Step 1:

Events are first processed in the PSF Analysis script.

a) Good Energy Prob. > .50
b) Determine Event Classes
c) Compute CT's for PSF Analysis
d) No cuts on PSF - goodness

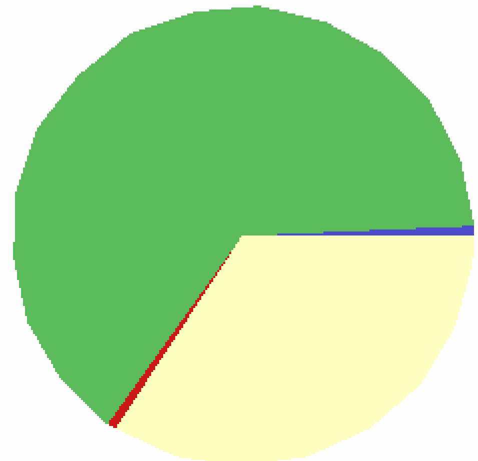
Breakdown after PSF Processing



γ's

Thin-VTX .179
Thin-1Tkr .300
Thick-VTX .151
Thick-1Tkr .370

Total: 1.000



BGE's

Thin-VTX .007
Thin-1Tkr .644
Thick-VTX .007
Thick-1Tkr .342

Total: 1.000

Note: the disparity among the Event Classes

BGE Accounting:

1) 2×10^6 generated
2) 8.5×10^5 Triggered
3) 8.9×10^4 Post-Prunning
4) 12.4×10^3 Post Energy Selection

Survival Factors*

Event Class	Factor
Thin-VTX	6×10^{-4}
Thick-VTX	8×10^{-4}
Thin-1Tkr	3×10^{-2}
Thick-1Tkr	2×10^{-2}

*Factors relative to Triggered and are corrected for relative Signal Fractions

(For SR case – factors 2x smaller)

At this point γ events have lost 4% due to ACD cuts & 15% due to energy cut

Losses: minimal and the VTX Event Classes already have S/N ~ 1 : 1 !

BGE Rejection CT's

Step 2:

Mixes of BGE's and γ's are formed

a) Training Sample – 50:50 BGE:γ
(Split the BGE sample 50:50 Training/Testing)
- Leaves only ~ 6500 of each type
- Statistics allow for only shallow CT's
- For demonstration – Lump Thick & Thin Event Classes Together

b) Test Sample - 80:1 BGE:γ (relative to "as-generated" totals)

The available statistics don't even allow for this!
- Leaves only ~ 500 γ's (after SR Case PSF Cuts ~ 400 γ's)

Caveat: What ratio of events should the train sample have?
- Need sufficient numbers of both classes to establish patterns
- At "real" analysis ratios – the CT splitting mechanism work poorly. Deviance per split will be too small.
- Trial & Error shows that ratio needs to be within a factor of 2.

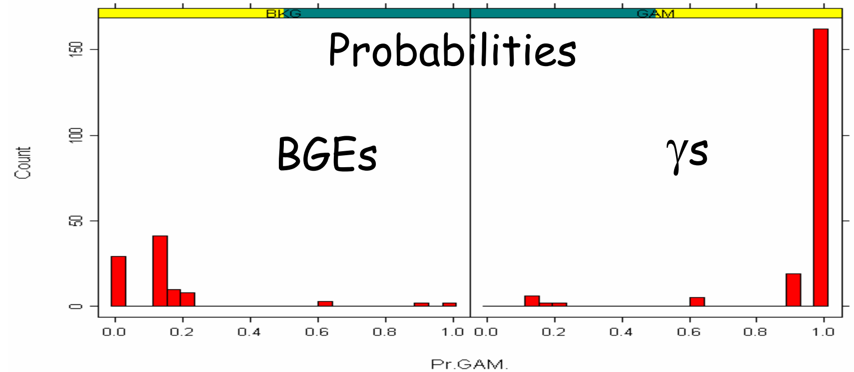
BGE Rejection CT's: VTX Events

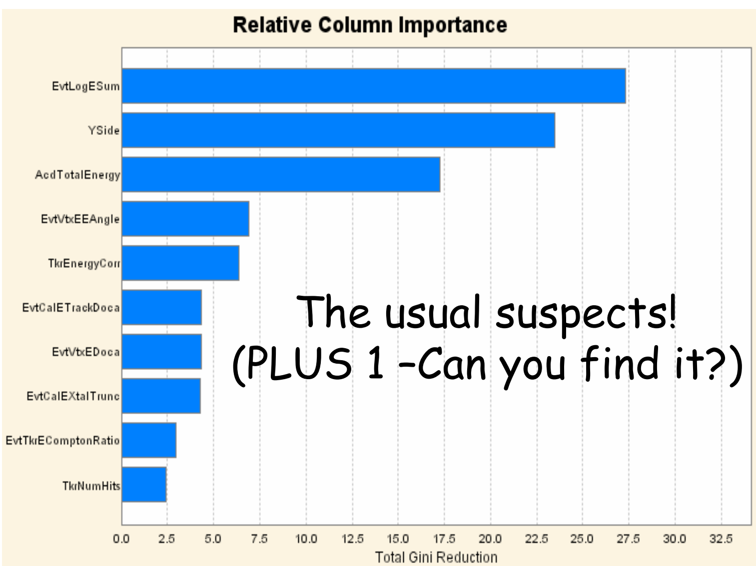
Note the sparse stats

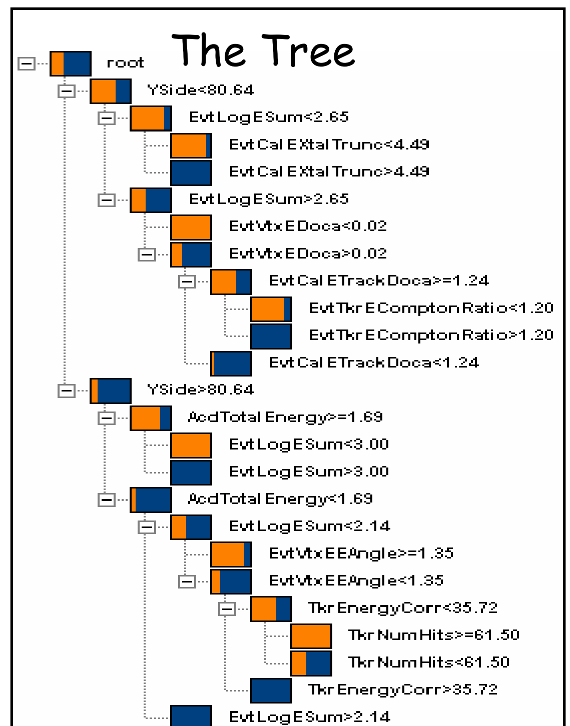
Input Node – Classification Tree (9)

		Predicted BKG	Predicted GAM	Totals
Observed	BKG	88	7	95
Observed	GAM	10	186	196
Totals		98	193	291

	Observed BKG	Observed GAM	Overall
% Agree	92.6%	94.9%	94.2%







For VTX Events
The CT gives
> 10x more
Rejection

Limited Rejection
due to low stats

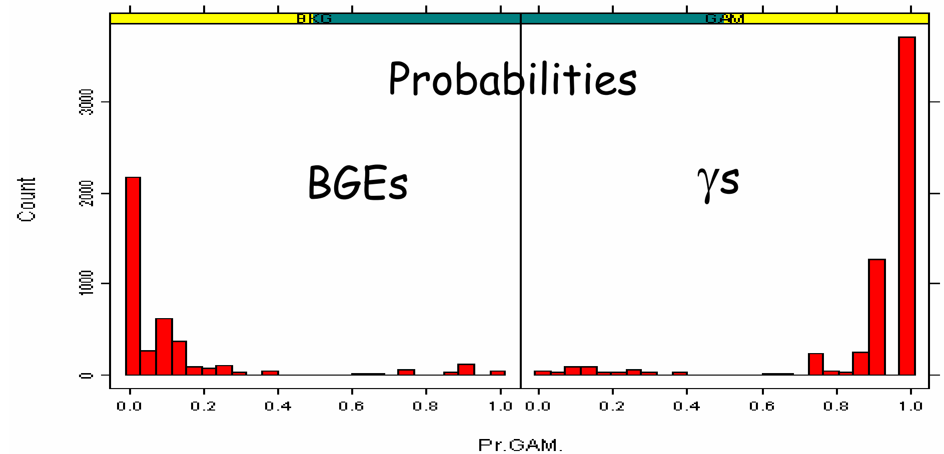
BGE Rejection CT's: 1Trk Events

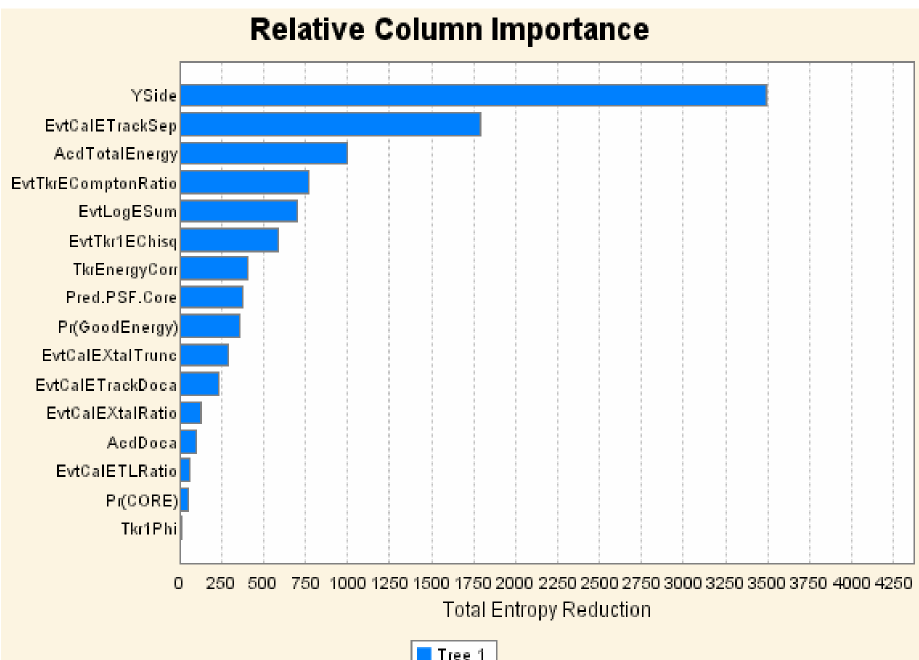
Stats large enough to grow a moderate size tree

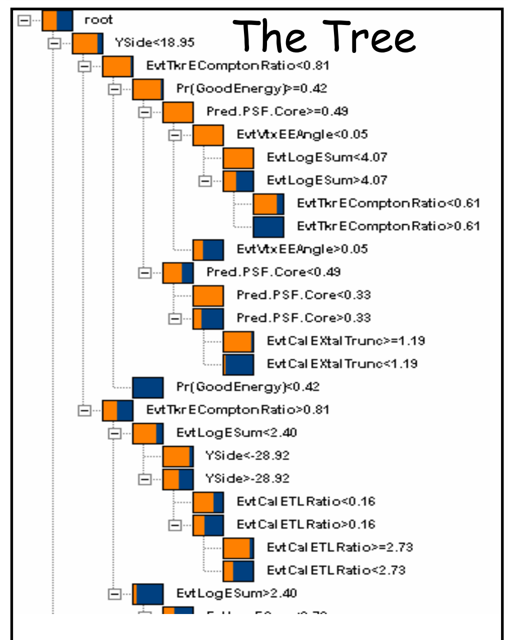
Input Node - 1Tkr Case (9)

		Predicted BKG	Predicted GAM	Totals
Observed	*BKG*	5940	434	6374
	GAM	505	7114	7619
Totals		6445	7548	13993

	Observed BKG	Observed GAM	Overall
% Agree	*93.2%*	*93.4%*	*93.3%*



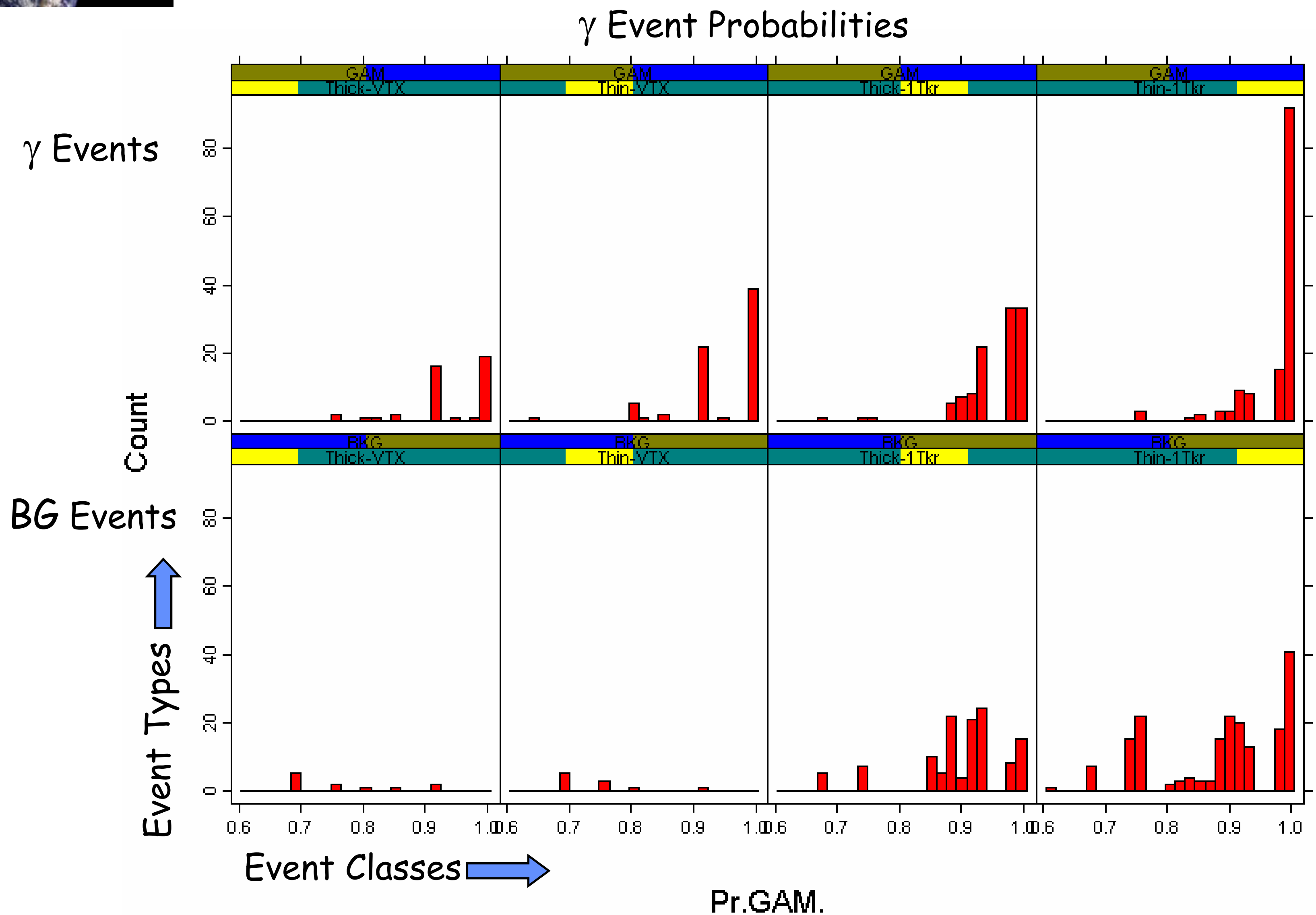




For 1Tkr Events The CT gives > 10x more Rejection

Would do better if Thick and Thin were done separately

Background Rejection Summary



BGE Rejection Summary 2

VTX Events (undifferentiated w.r.t. Thin/Thick)

1) Remaining background: 3% (But recall test sample is only 80:1)
2) Good Event Loss: 17%
3) BGE Reduction Factor: 16x (post SR Case selection)
4) Further progress stop for lack of statistics
   (there were 3 BGE's events left)

1Tkr Events (undifferentiated w.r.t. Thin/Thick)

1) Remaining background: 32% (No there yet!)
2) Good Event Loss: 3%
3) BGE Reduction factor: 60x (post SR case selection
4) Further progress limited by state of present software

This exercise is an example of what will happen to the science if we lose two sides of the ACD and put a big hole in the top of it as well!