Point Source Search with 2007 & 2008 data:

Study of data modeling...

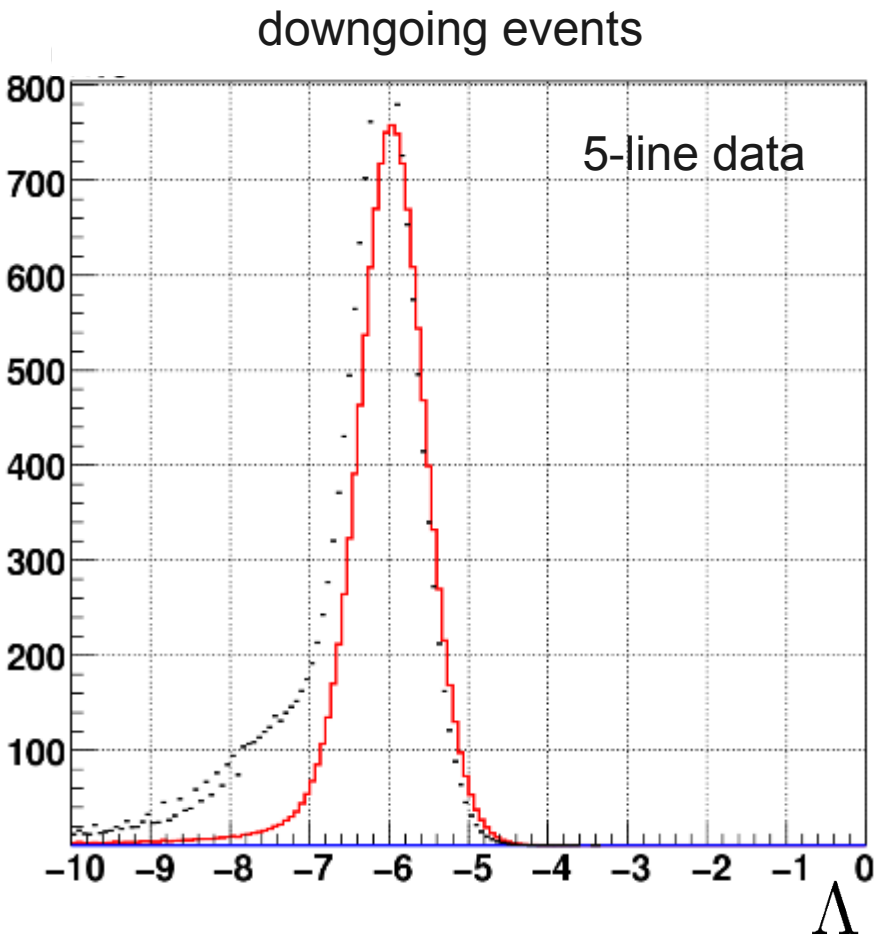
Aart Heijboer, Nikhef.

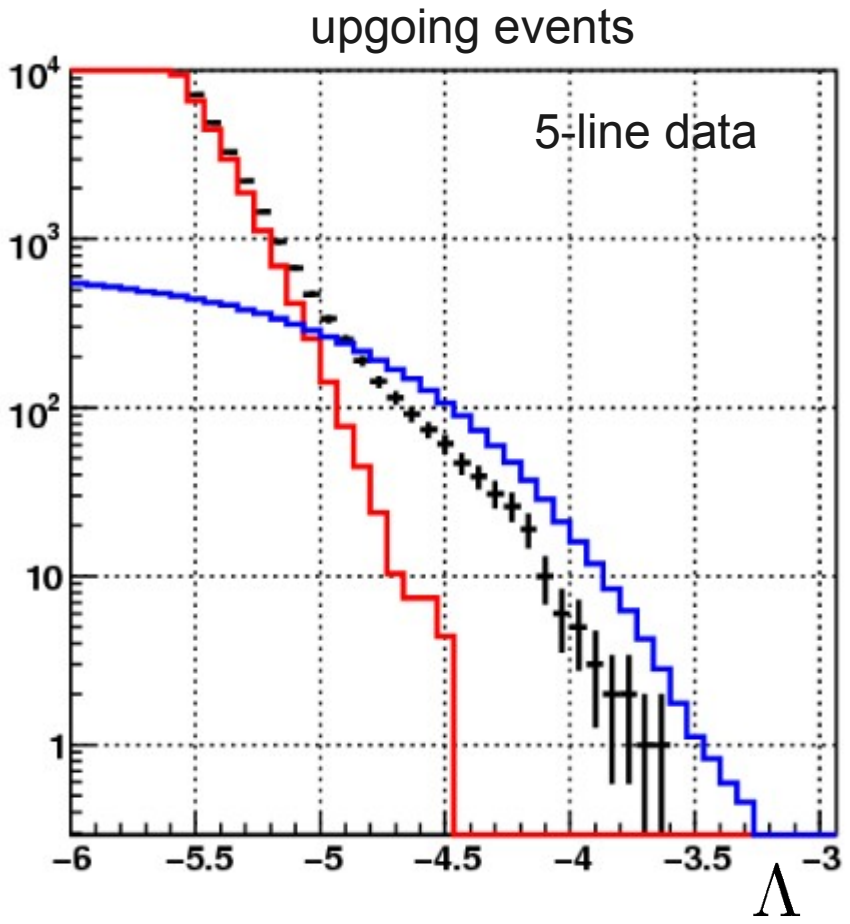
- 'low' level studies:
  towards understanding the data (this talk)
- high level stuff:
  limits, systematics, discovery potential (next talk)
- other bug-fixes/improvements (see lots of talks, before)

soon: let dust settle, finalize, rerun and combine:

- all plots will change, but
- many tools/methods are in place.

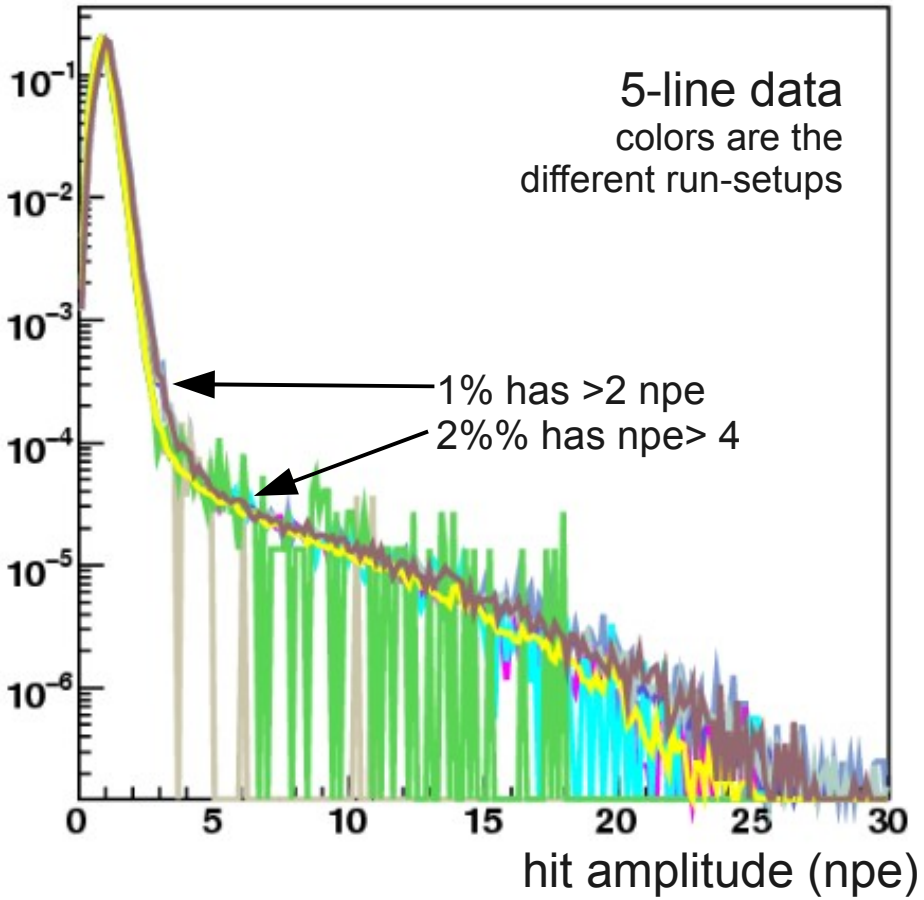
The problem

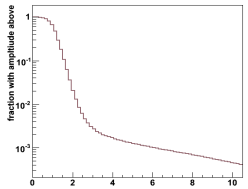




- MonteCarlo consistently over-predicts quality value Λ (by few tenths of a unit)
- clue: strong dependence on background rate (see my Erlangen talk and Simone's talk)

Hit amplitudes





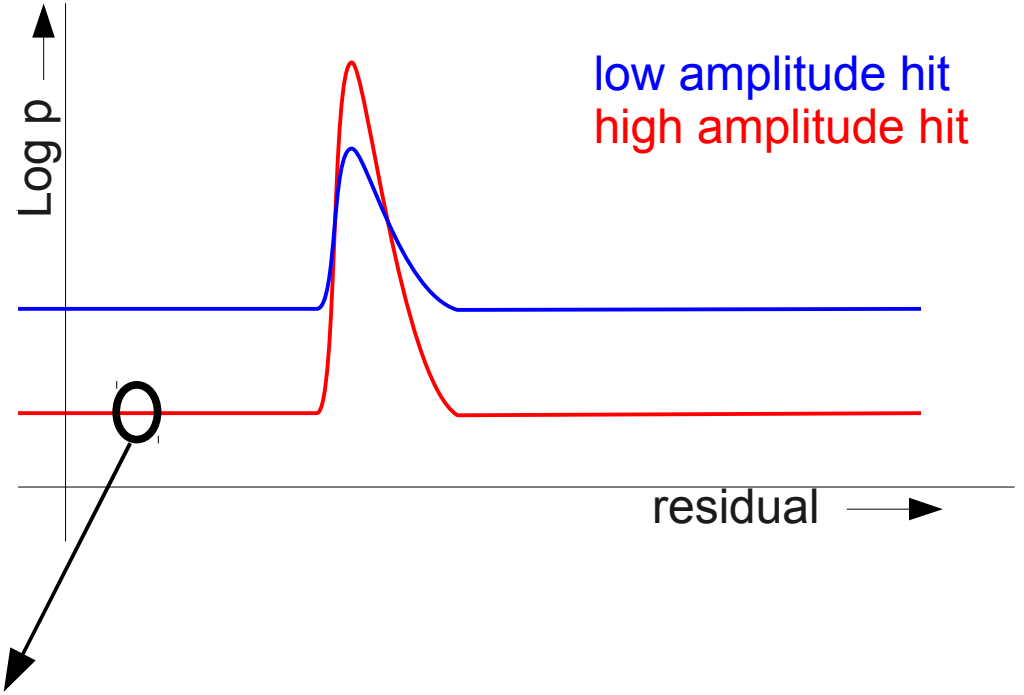
- Amplitude distribution of *background* hits
  - actually: earliest 10 hits of the PhysicsEvents (could also use L0 data or minbias)
- Large tail present at sub-% level (origin of these hits : ^{40}K decay in sphere ?)

- Monte-carlo (TriggerEfficiency):
  - assumes single-photo-electron, smeared by 30%, cutoff at 0.5 pe.
  - this is wrong for sure
    - already proposed to fix this in Erlangen
    - but does it matter?
      → there are reasons to suspect: yes

How do background hits contribute to Λ?

$$\Lambda = \log(\mathcal{L})/\mathrm{ndof} + 0.1 \times N_{\mathrm{comp}}$$

$$\log \mathcal{L} = \sum_i \log P(t_i|\mathrm{npe}_i)$$



- Background hits with residuals incompatible with the track contribute a ~constant number to the likelihood.
- If the amplitude is large, log(P) is a very negative number → significant contribution to numerical value of Λ.
- nb: the fit may be mostly unaffected, since such a hit contributes to the likelihood as a constant (as long as residual is incompatible with the track)

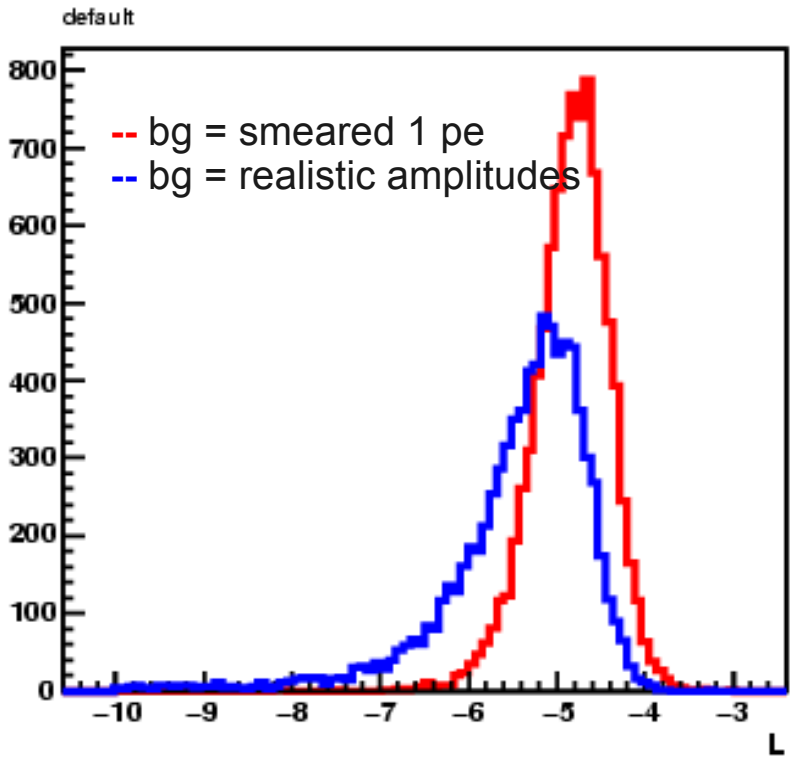
Simple simulation setup

- read MCEW or .evt geasim file
- simulate addition of random background (usually 80 kHz)
  - either npe = Gauss (1, 0.3)
  - or npe = random sample from histogram
- simulate 40 ns ARS integration time
- *not* simulating ars dead time or trigger

- Feed simulated hits directly into stand-alone version of aartstrategy.
  - run full reconstruction
  - investigate contributions to likelihood on hit-by-hit basis (*while having full hit-level MC truth info at my disposal*)

Likelihood study

- ν-MC with 100 GeV < E_μ < 10 TeV, with at least 10 signal hits.
- take **MC-truth muon**, based on this track:
  - apply the same hit selection that is done for the final likelihood fit
  - compute the likelihood



aartstrategy final hit selection

for final fit, select hits that:

- have a time residual -250 < r < 250 ns and a distance-to-track < 100 m.

OR

- are part of a coincidence

OR

- have an amplitude > 2.5 p.e.

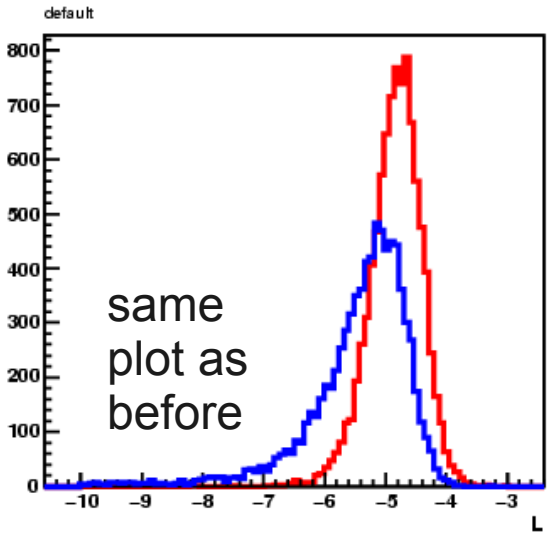
$$L = \log(\mathcal{L})/N_{\text{hits}}$$

simplified version of Λ

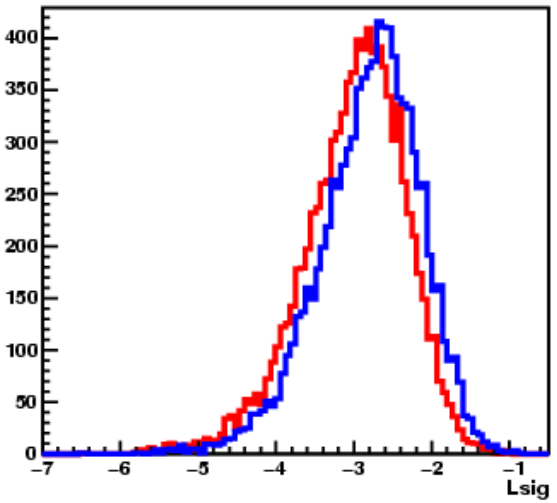
modeling of background amplitudes has huge impact on lambda-distribution!

Likelihood study

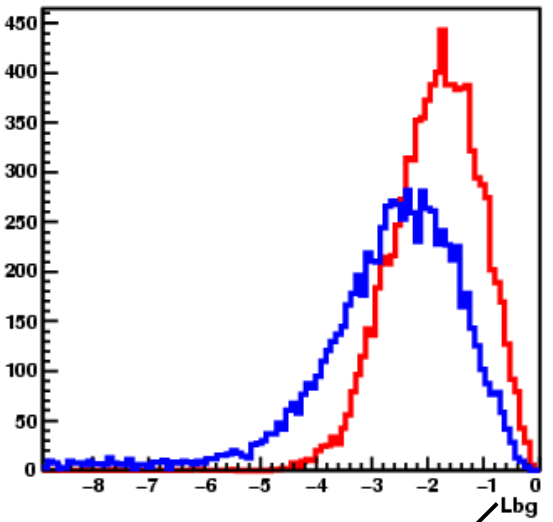
-- bg = smeared 1 pe
-- bg = realistic amplitudes



=

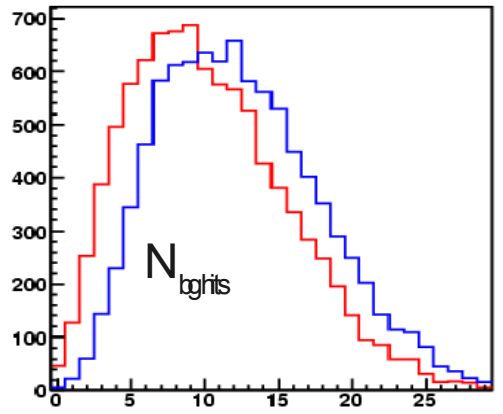


+



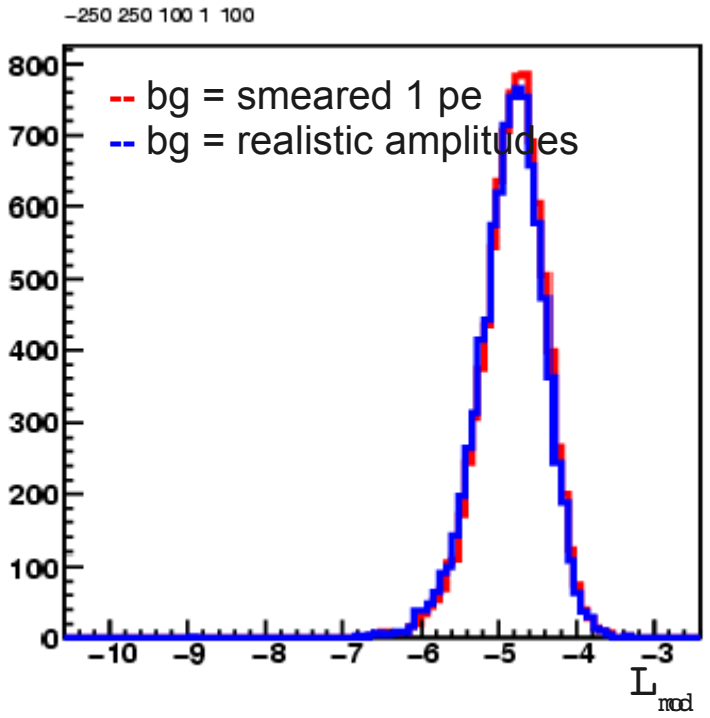
$$\sum_{\text{allhits}} \log P(r_i)/N_{\text{hits}} = \sum_{\text{sig-hits}} \log P(r_i)/N_{\text{hits}} + \sum_{\text{bg-hits}} \log P(r_i)/N_{\text{hits}}$$

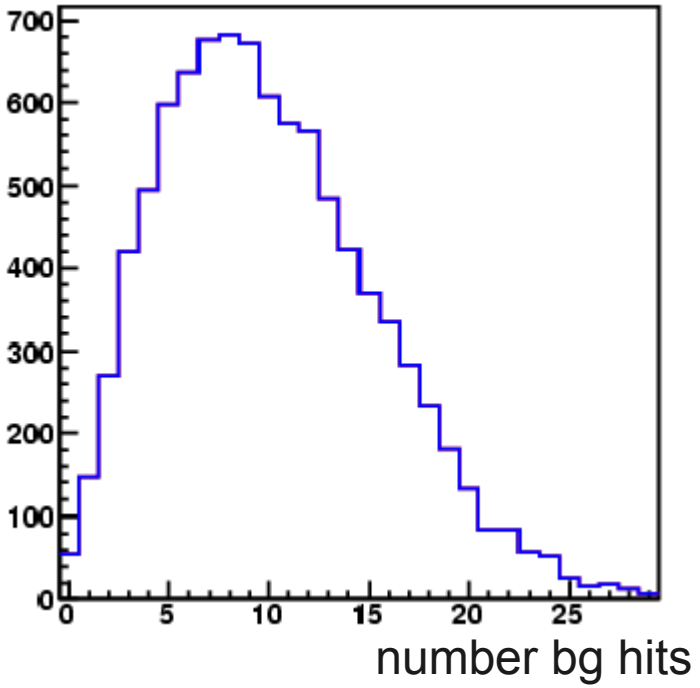
- difference in L-shape due to contribution from background hits (slight difference in L_{sig} is due to differing N_{hits})
- only a **few** more bg hits in the blue case, but they have a **huge** effect on the likelihood.
- it all makes sense.



Likelihood study: Robuster version of Λ

Drop the inclusion of hits with amplitude > 2.5 npe
when computing L_{mod}



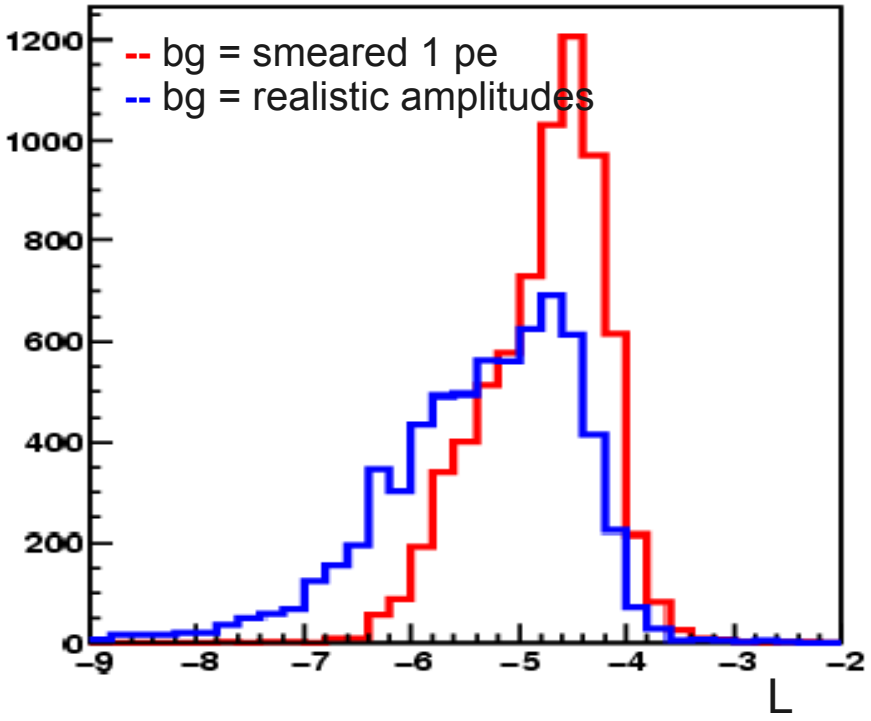


much more robust against mismodeling of background amplitudes.

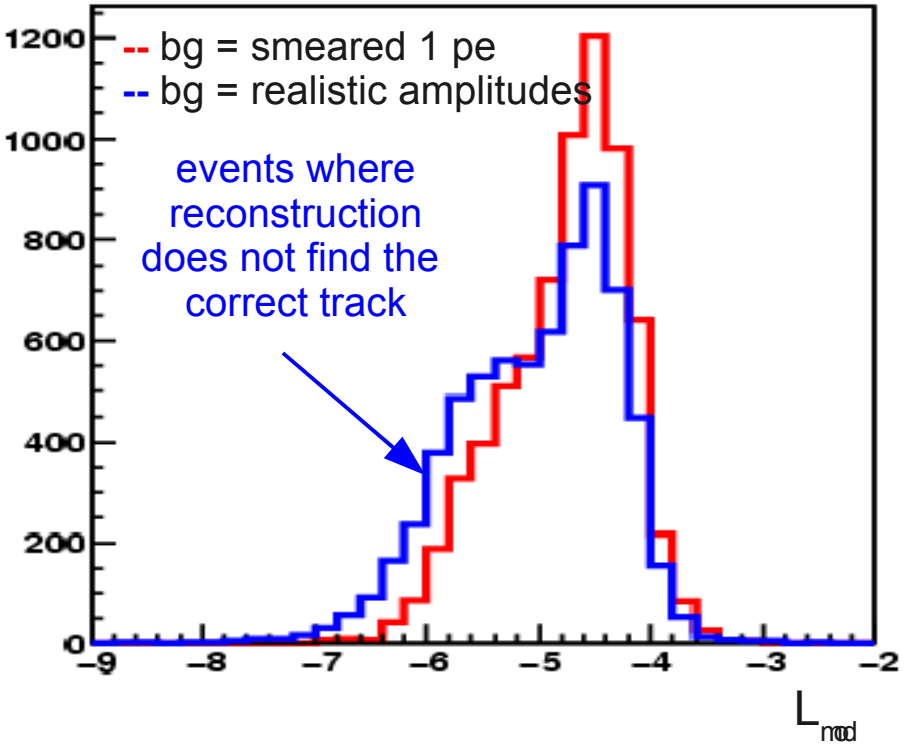
reminder: this is the likelihood of the true track, not of the reconstructed track

Reconstruction

- now run full reconstruction algorithm -> no mc truth information used here
- reconstruction algorithm and PDF itself unmodified (compute L and L_{mod} afterwards)

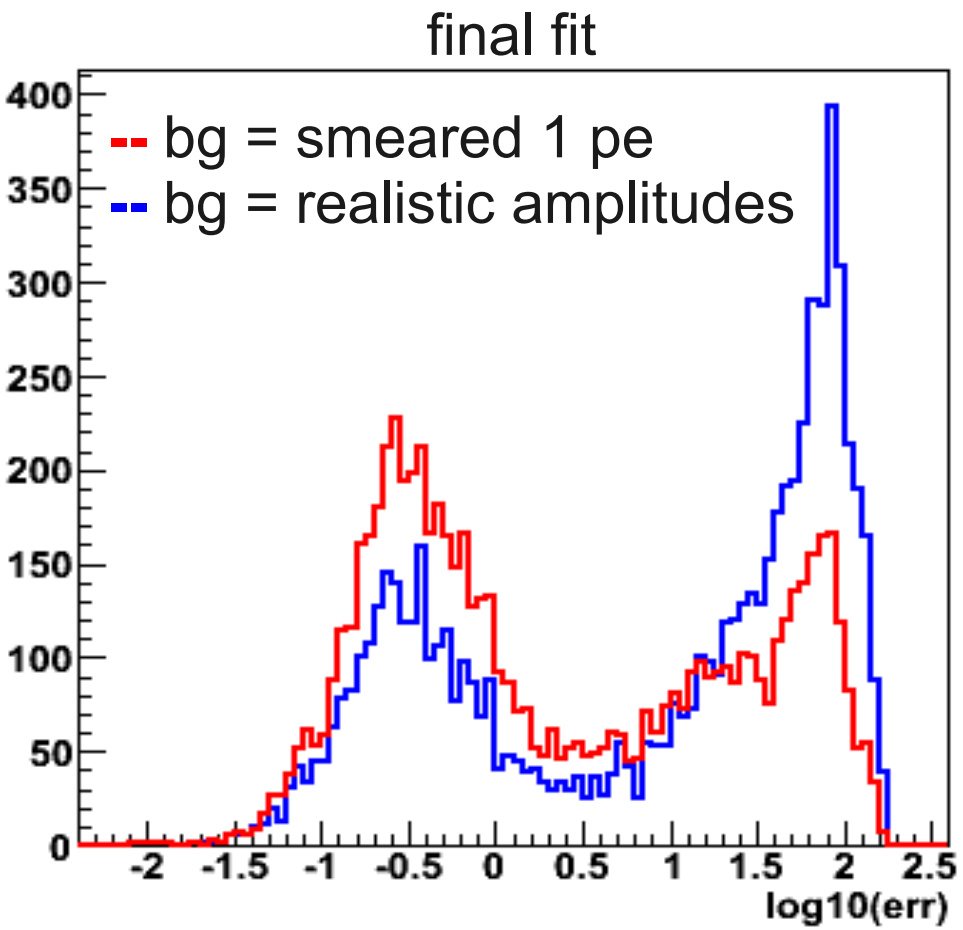


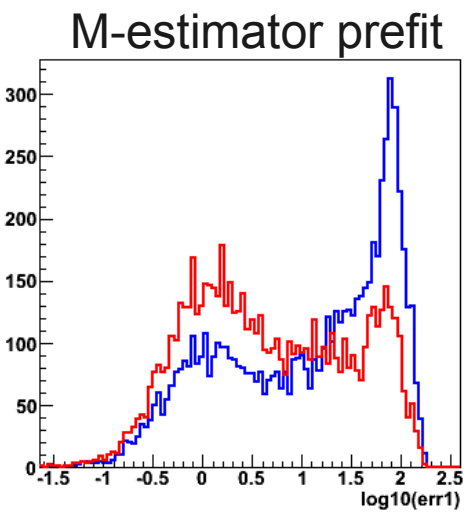
as expected: L distribution very sensitive to bg-modeling

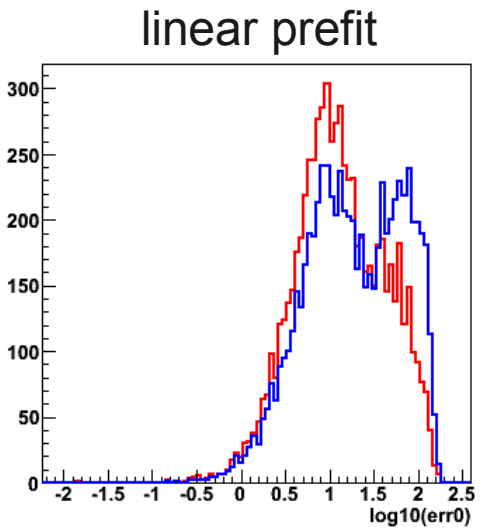


when using more robust variant of L, still some dependence left

Reconstruction: performance





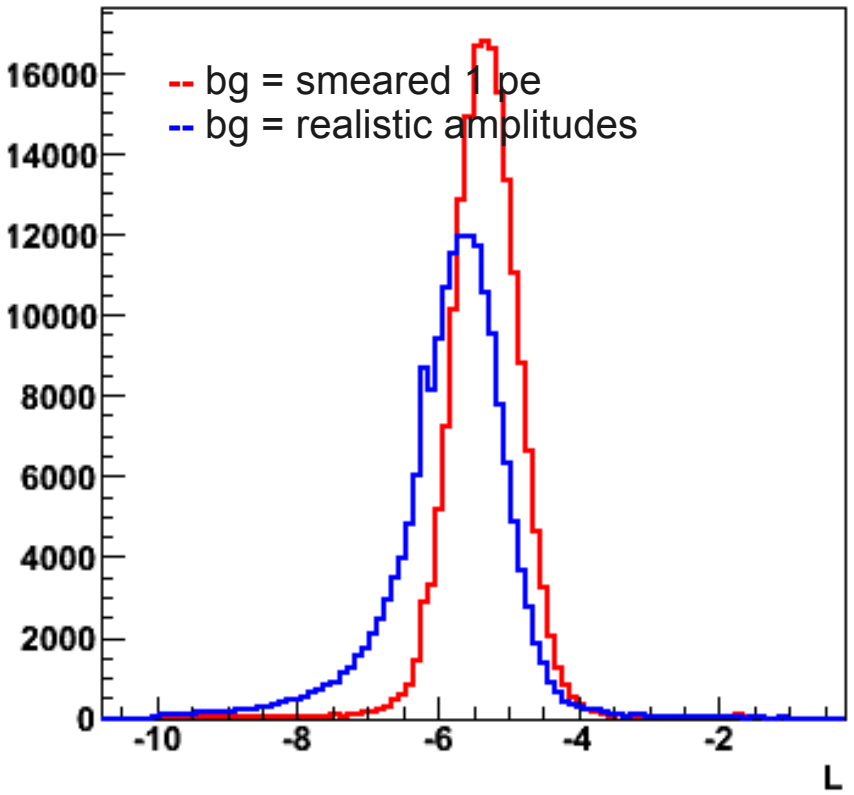


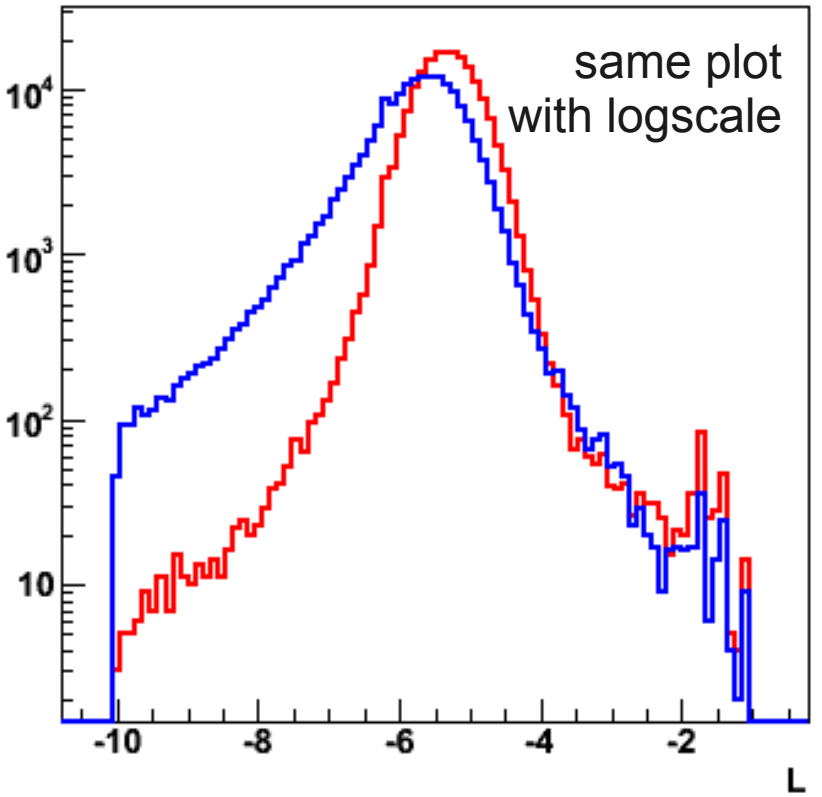
- performance of reconstruction severely affected by large amplitude bg hits
- already in first stages of algorithm (not a likelihood issue, but hit-selection)
- want to recuperate, should be fairly easy

nb: the blue curves should agree with what we have in data, but we would like the reconstruction on realistic MC to be as performant as it is on events with single-pe background.

Effect in downgoing muon mc.

- mupage mc
- mean background rates and dead-om-mask from run 37128
- no trigger

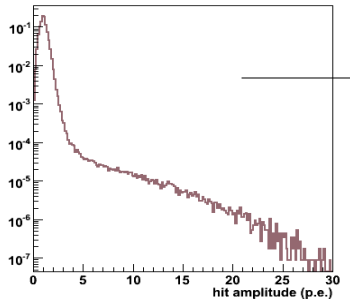




“TriggerEfficiency -C2”

- new feature to throw amplitudes of background hits from a histogram
  see analysis elog entry 345
- but simple consistency check fails:

TriggerEfficiency



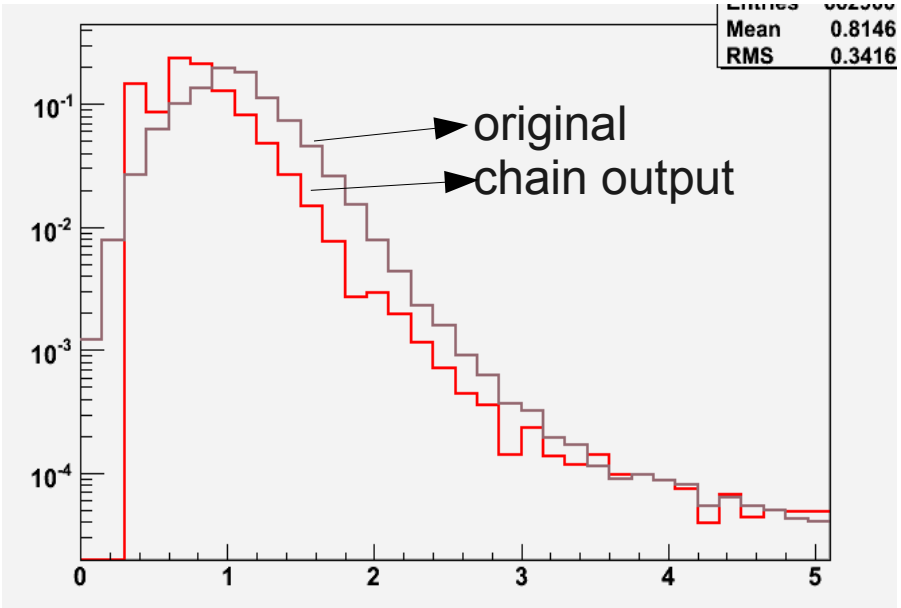
- throw amplitudes
- **convert** to AVC values
- run trigger
- write PhysicsEvent

CalReal

- read PhysicsEvent
- **convert** AVC values to amplitudes
- reconstruct
- write FullEvent (with amplitudes)

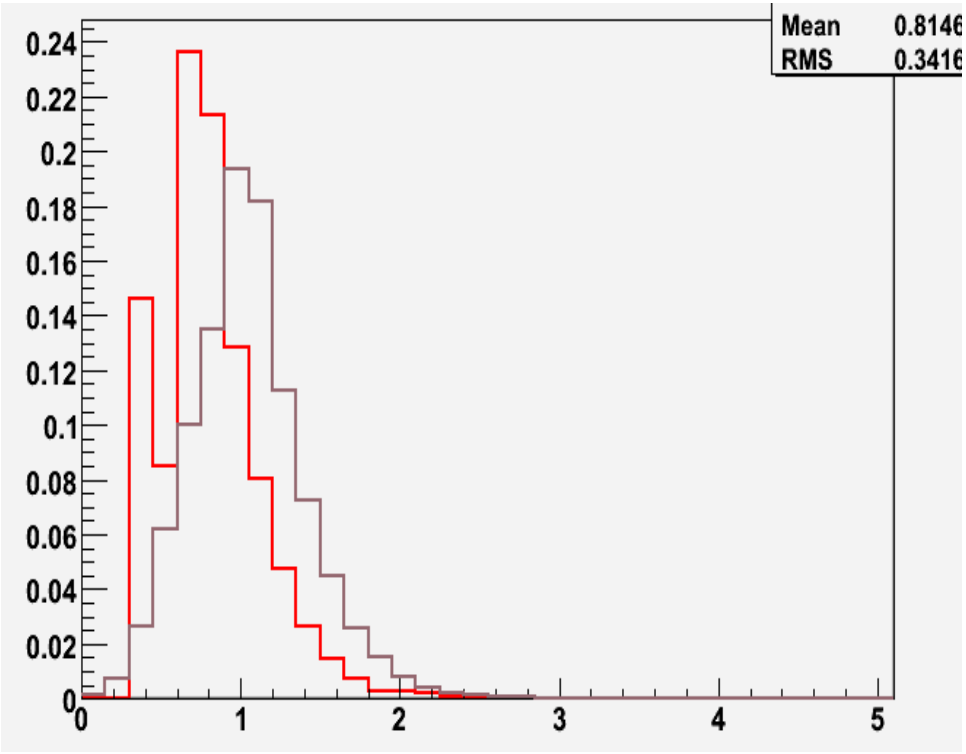
me

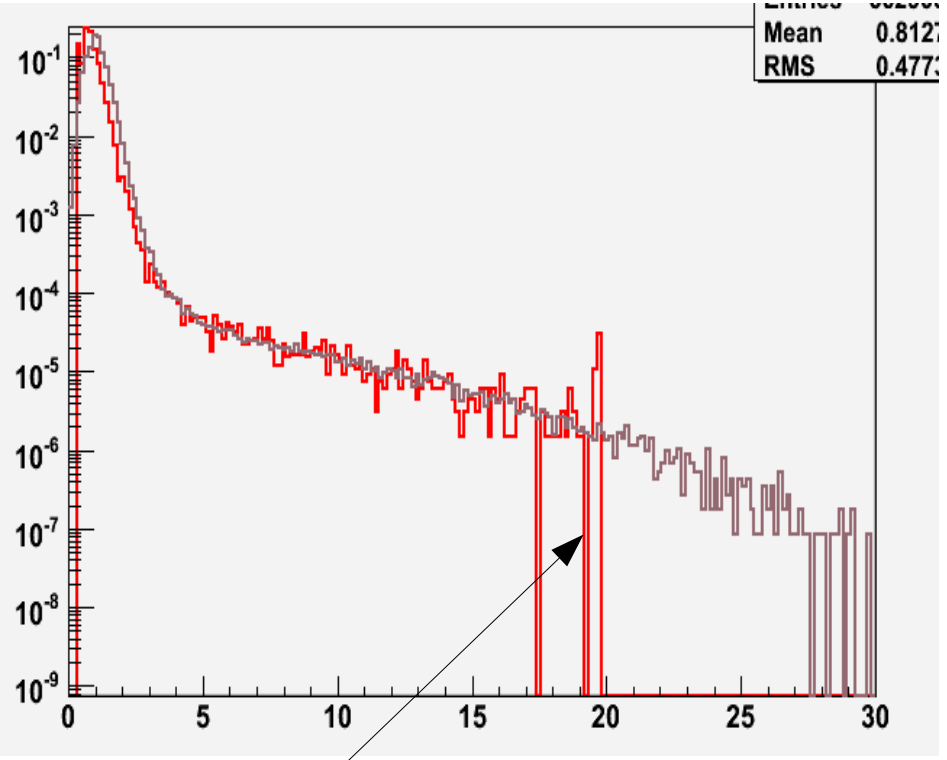
- plot amplitudes of earliest 10 hits in each FullEvent
- compare with the start-distribution



- **something is wrong! problem could be in any of the boxes...**
- **whatever it is : may well influence data/mc agreement some more**

“TriggerEfficiency -C2”





same plot as before, on different scales

- tail is actually nicely modeled
  - peak understood: saturation at 20 p.e. simulated (not in current cvs version)
- but one/few spe peak completely off!

Conclusions

general:

- Modeling of the optical background hit amplitudes is currently quite wrong.
- Long tail up-to >30 p.e. not modeled
- fixing MC should be easy and will cause a shift in the right direction.
- New option in TriggerEfficiency, but consistency check fails →
  - still some problem somewhere in this chain
    (either in TriggerEfficiency or CalReal or in my stuff/command-line options).

reconstruction:

- Quality parameter Λ happens to be very very sensitive to this mismodeling.
  - Very easy to come up with a more robust variable, but
  - performance also affected, already at prefit-stages
    → will hopefully recover soon by revising hit-selection algorithms.

summary:

- A large part of data/mc discrepancy is probably understood.
- next steps
  - fix bug and run full-chain mc
  - update reconstruction to recover performance.