Smearing Method for generator level data

WG6 meeting

CERN

5/05/2011

Outline

- The need and what already exists
- A new solution envisionned
- Description of the principle
- Results on ttbar events
- Results on WW events
- Conclusion
- Outlook

The need and what already exists

- Why smearing ?
  - Huge amount of background events
  - Long time to fully simulate every single event
  - Generation is fast
  - Smearing can give a realistic approximation the full simulation

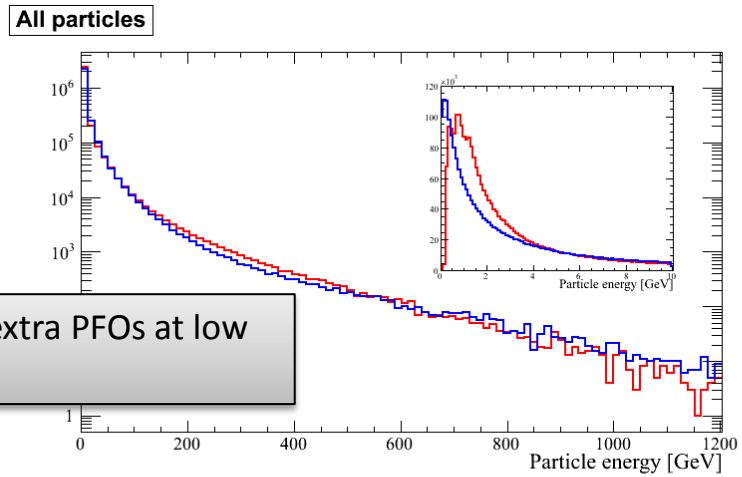
The need and what already exists

- Available yet: Lars Weuste Marlin driver
  - Developped for a specific channel (ee -> squark squark -> qq XX)
  - Based on the smearing of jets (Durham dijet)
  - Rely on 2009 PFA paper results, not updated for later improvement
- I needed more:
  - Optimized for ttbar events
  - Smearing of non-jet-related variables
  - Various jet deffinitions (kt, hemispheres)
  - Mimic of the DST PFA performance

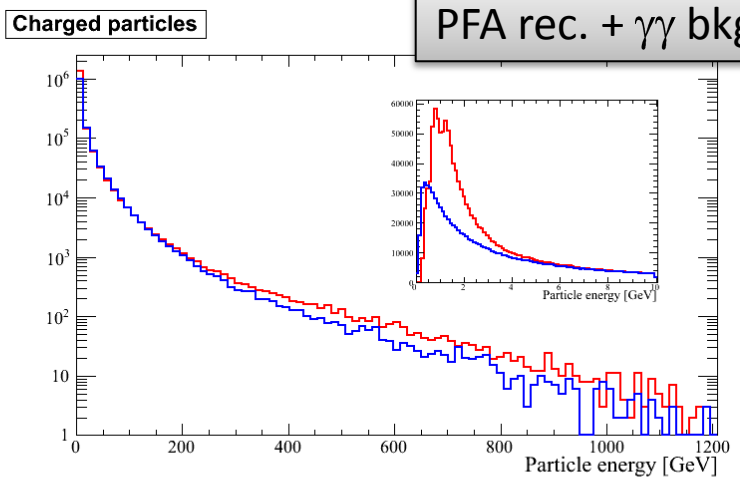
A new solution envisionned

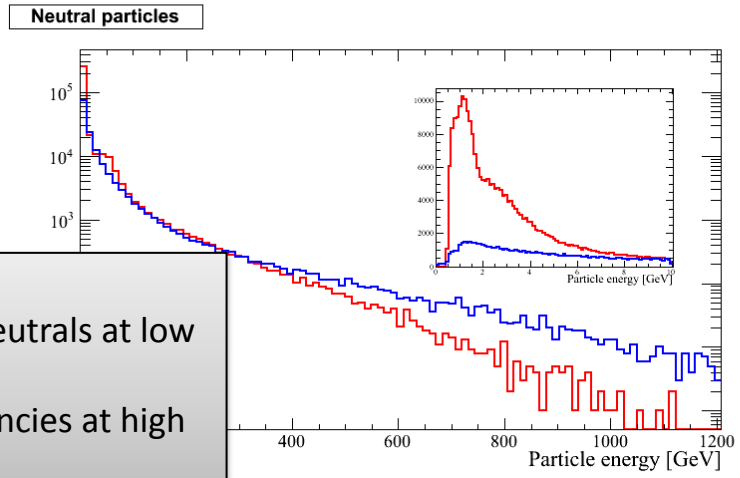
- Smearing at PFO level ?

PFA rec. + $\gamma\gamma$ bkg

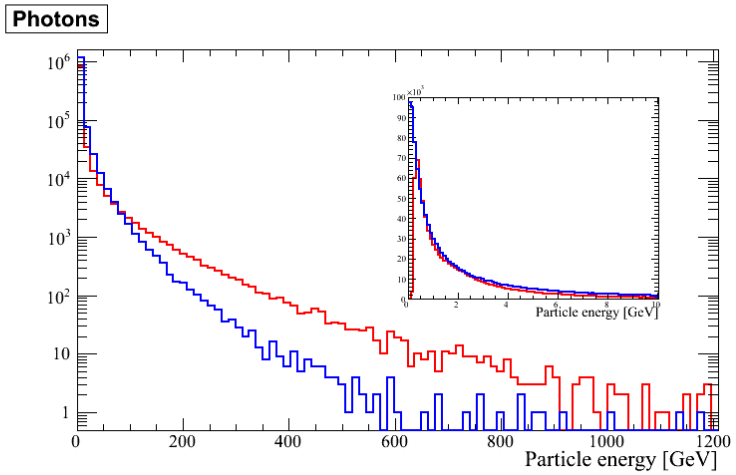


$\gamma\gamma$ bkg -> extra PFOs at low energy



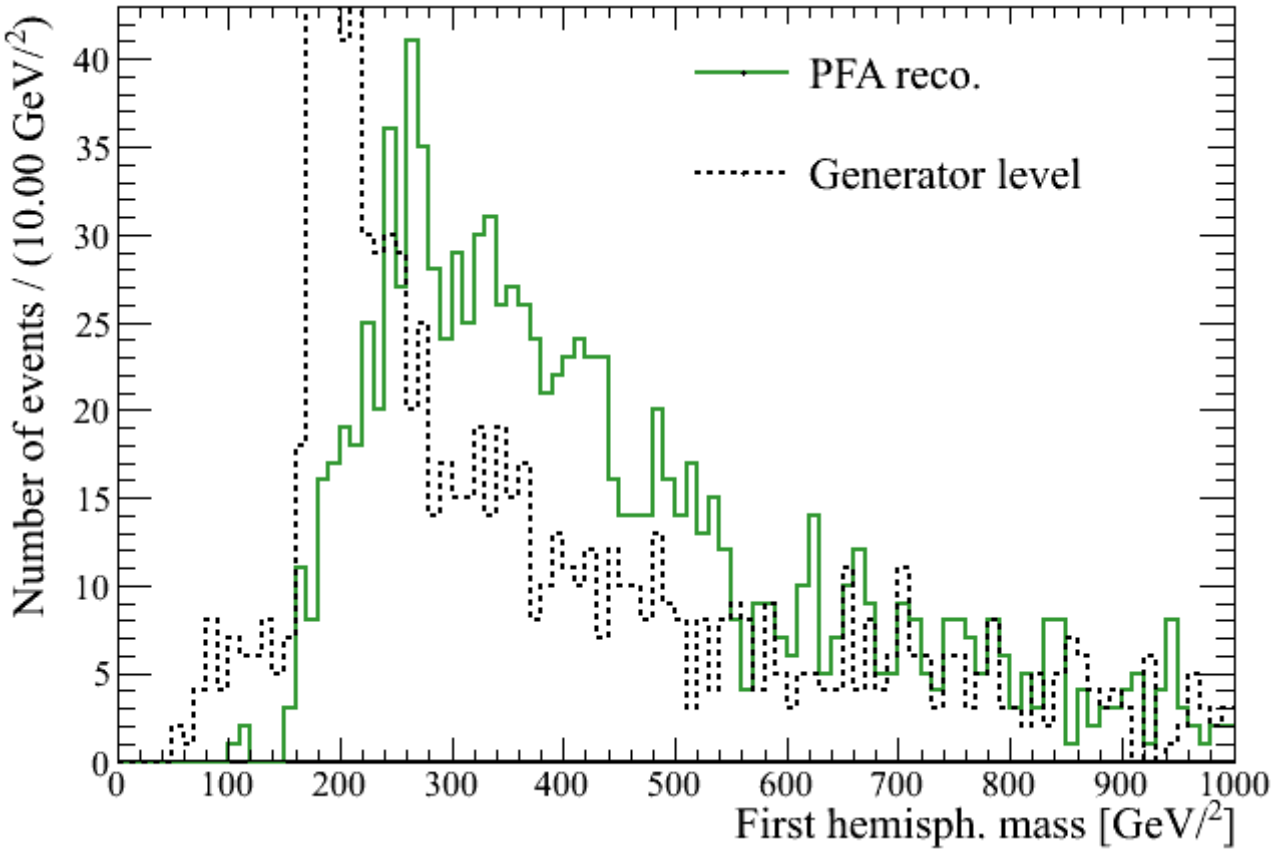


PFA rec.
-> extra neutrals at low energy
-> dicrepancies at high energies



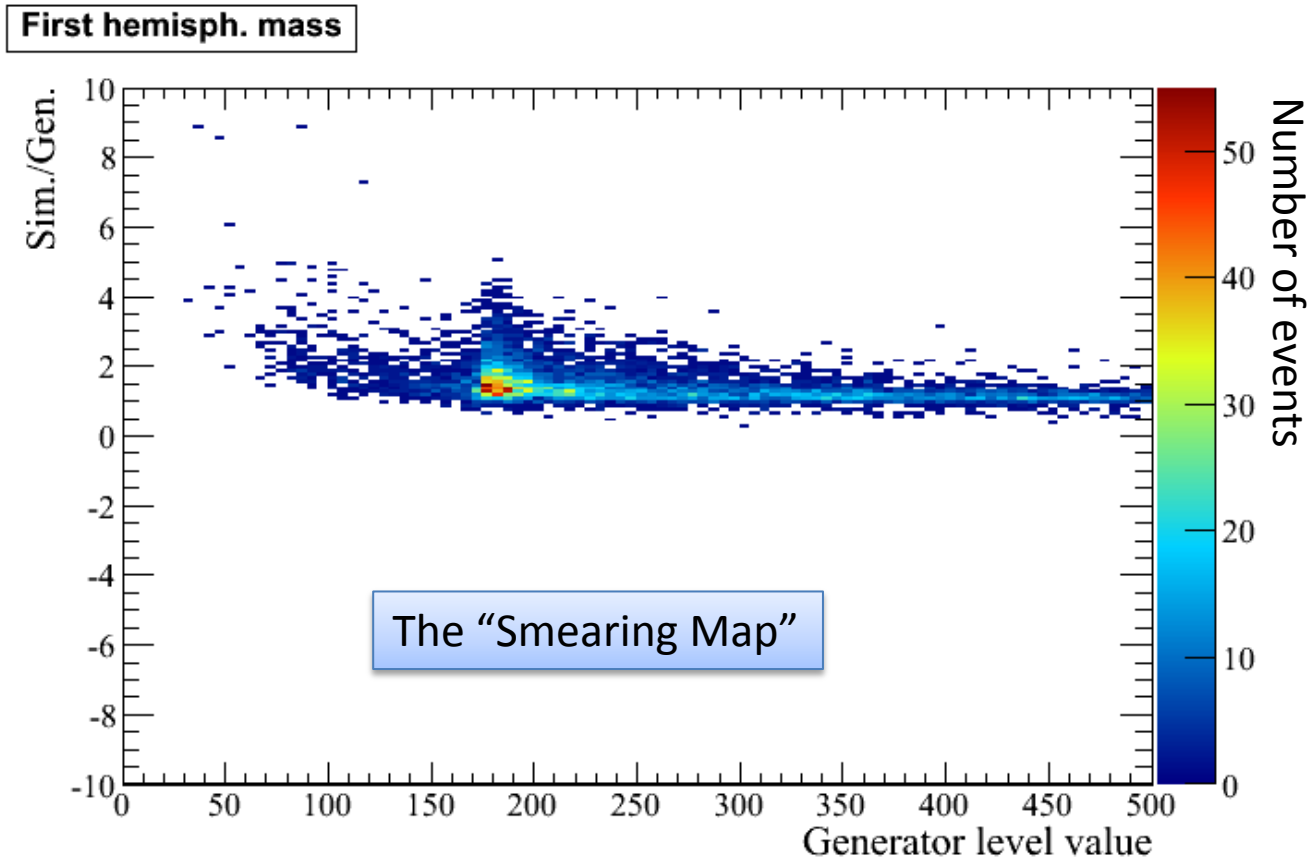
A new solution envisionned

- Smearing at the level of observables or finalstate analysis variables (PFA + $\gamma\gamma$ + TightCut)



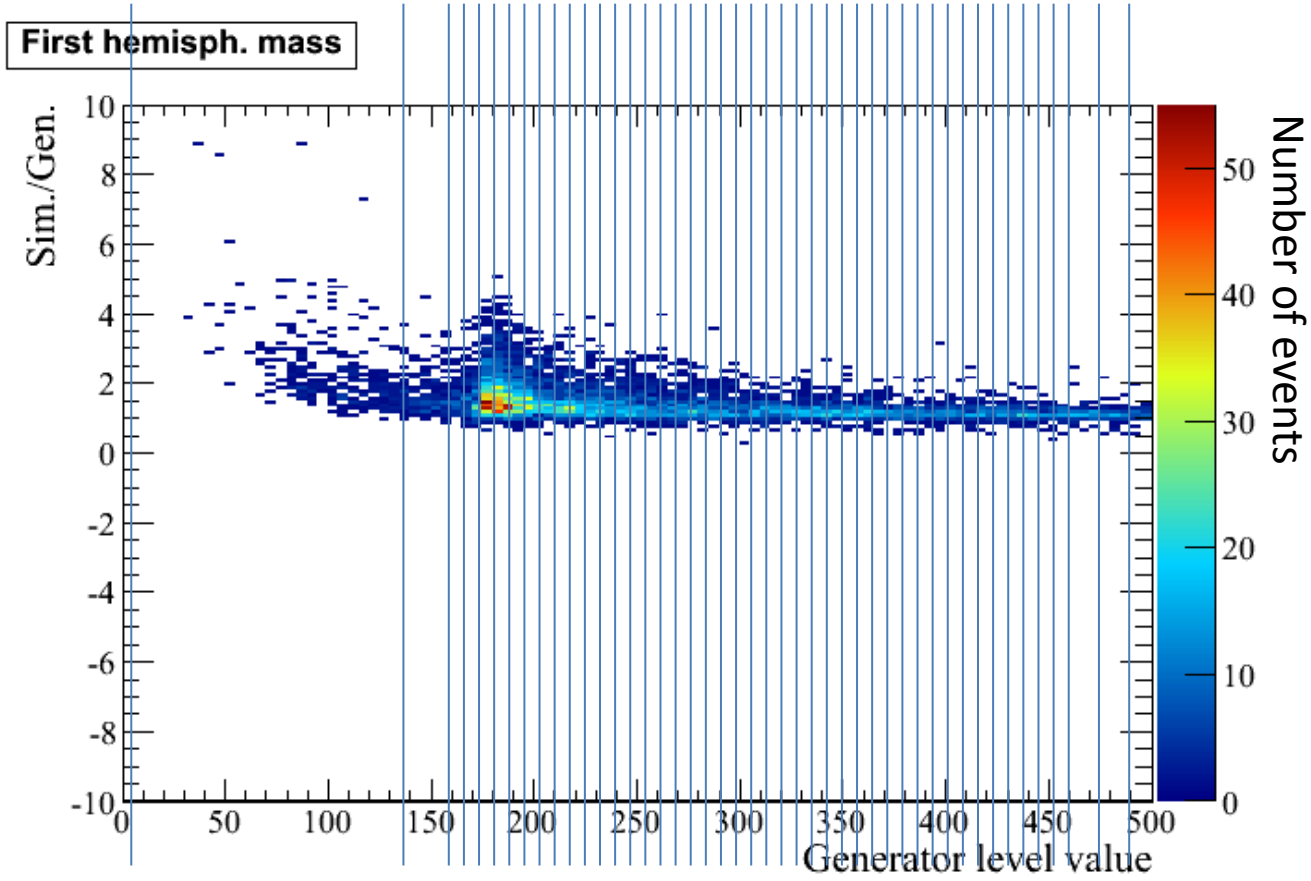
Description of the Principle

- For each variable to smear plot:
  Reco. value/ Gen. value VS. Gen Value



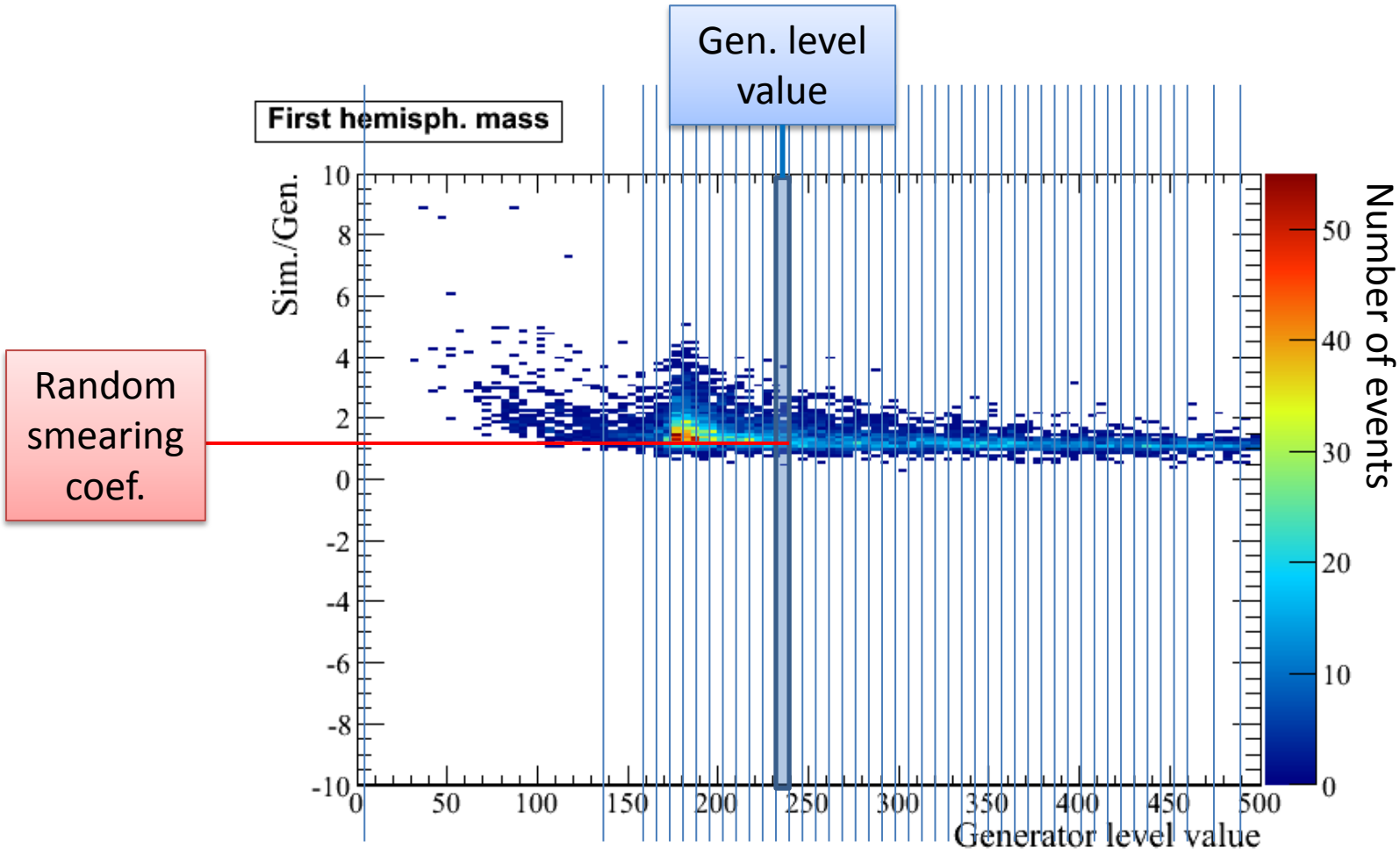
Description of the Principle

- Cut vertical slices (mind the stat, min=250)



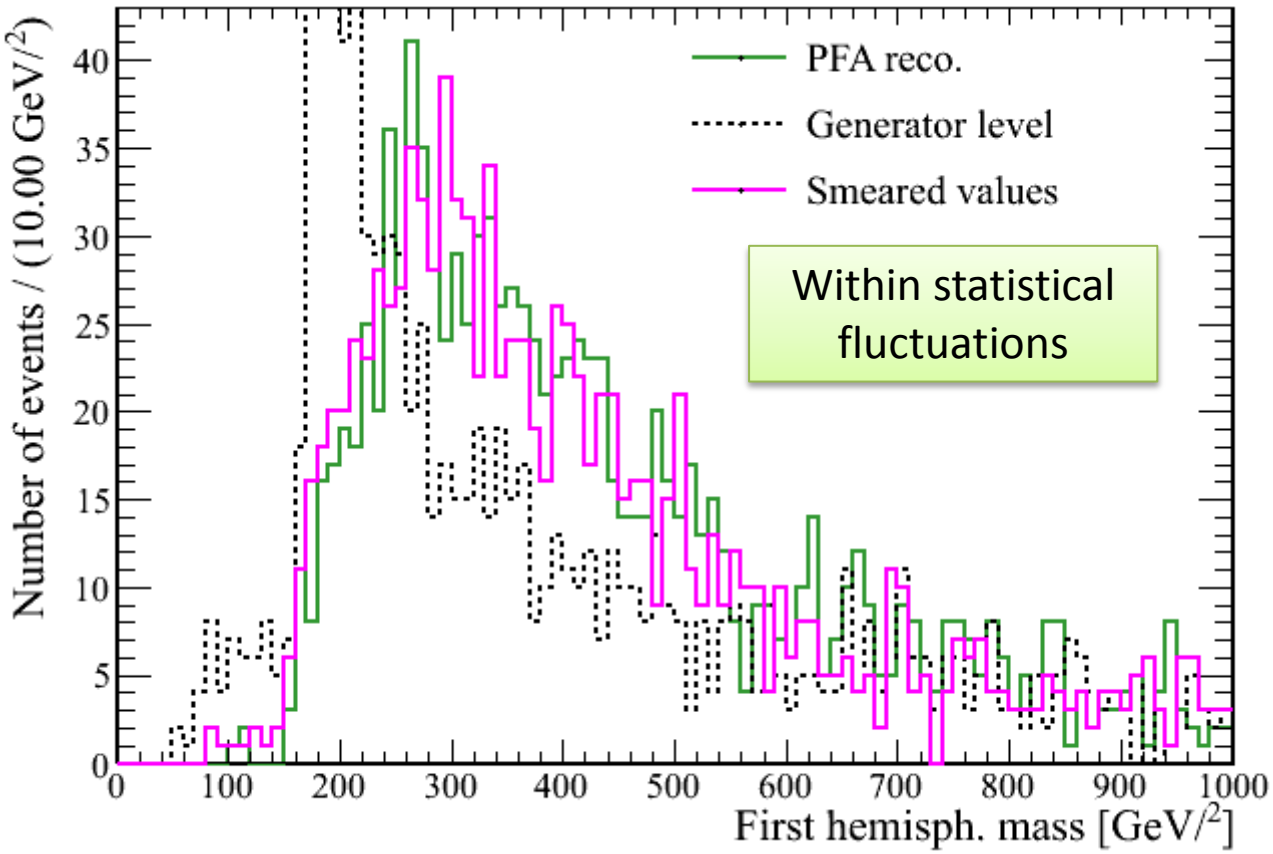
Description of the Principle

- Take generator level value to smear:
  - Select the corresponding slice
  - Generate a random coef. with the corresponding distribution.



Description of the Principle

- Multiply the generator level value by the smearing coef.



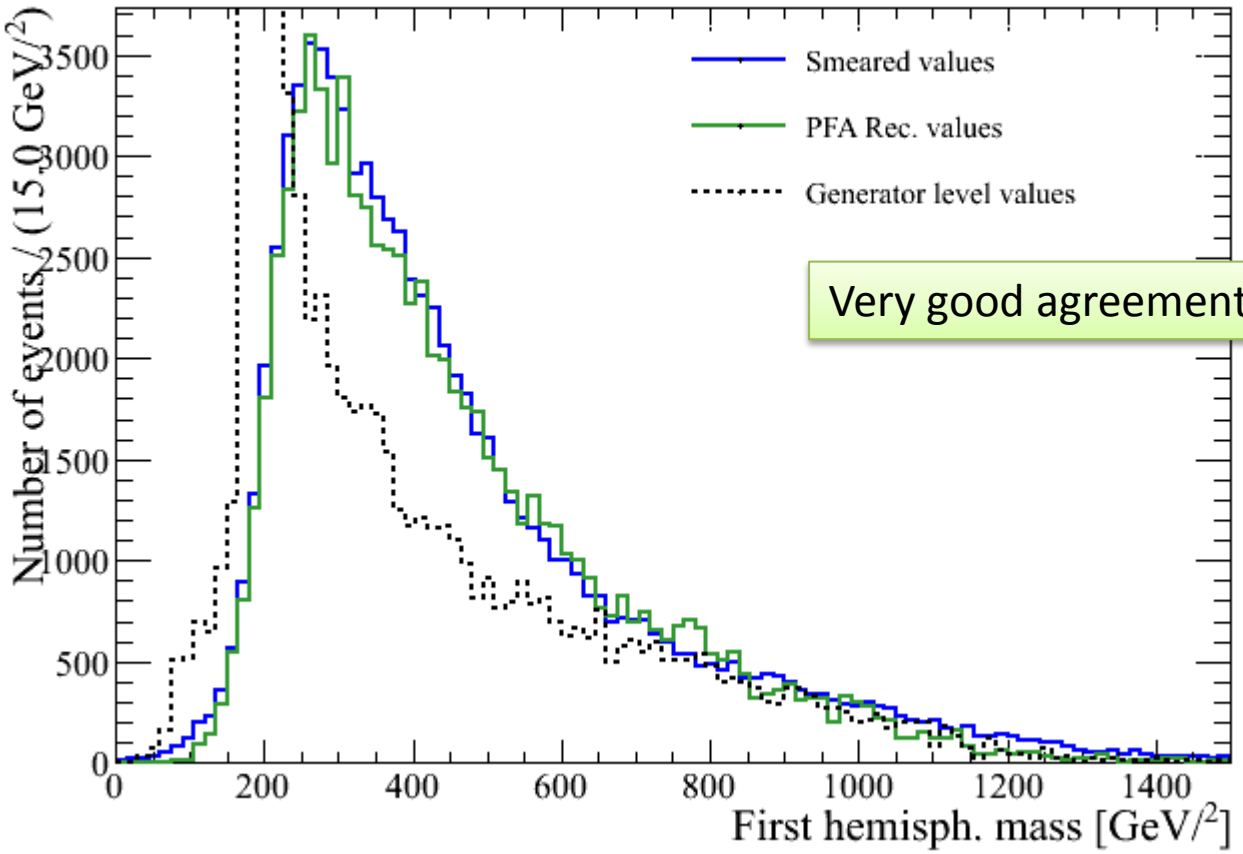
Results on ttbar events

- Smeared data obtained by looping on generator level data sample:

  -> Each gen. lev. event give many independent smeared events.
  -> The full desired statistics is obtained (e.g. 180k smeared ttbar out of 19k gen.)

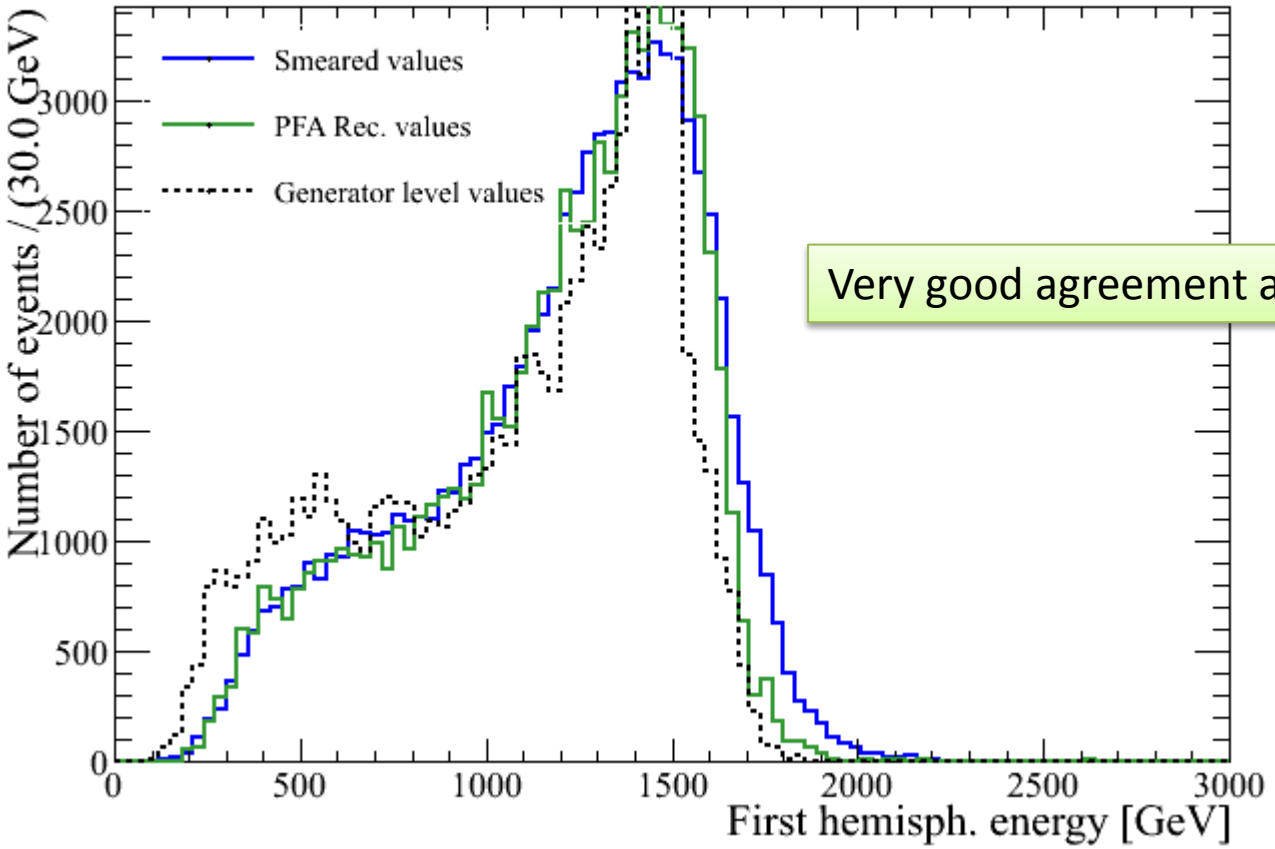
- Generator level and Full. Sim. data are scaled to match area.

Results on ttbar events



Very good agreement achivable

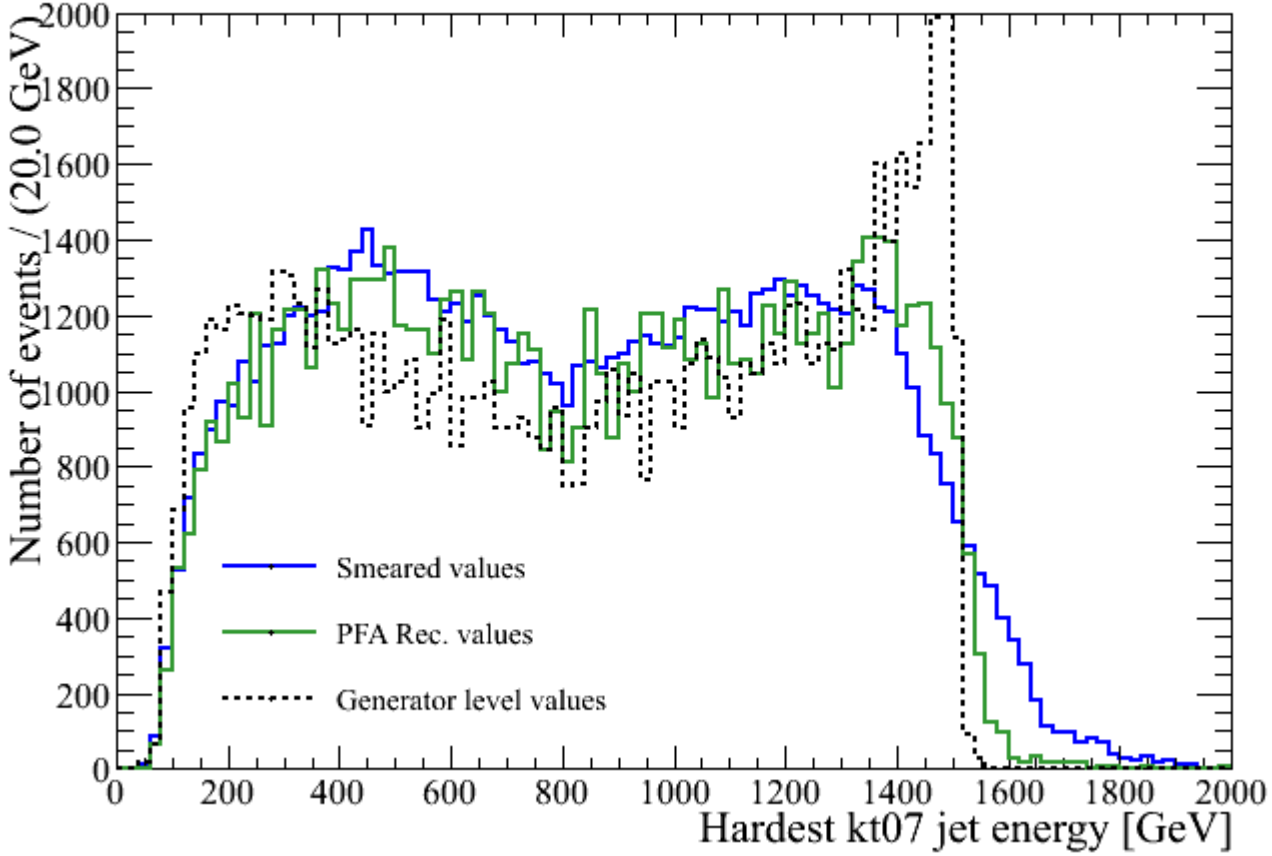
Results on ttbar events



Very good agreement achivable

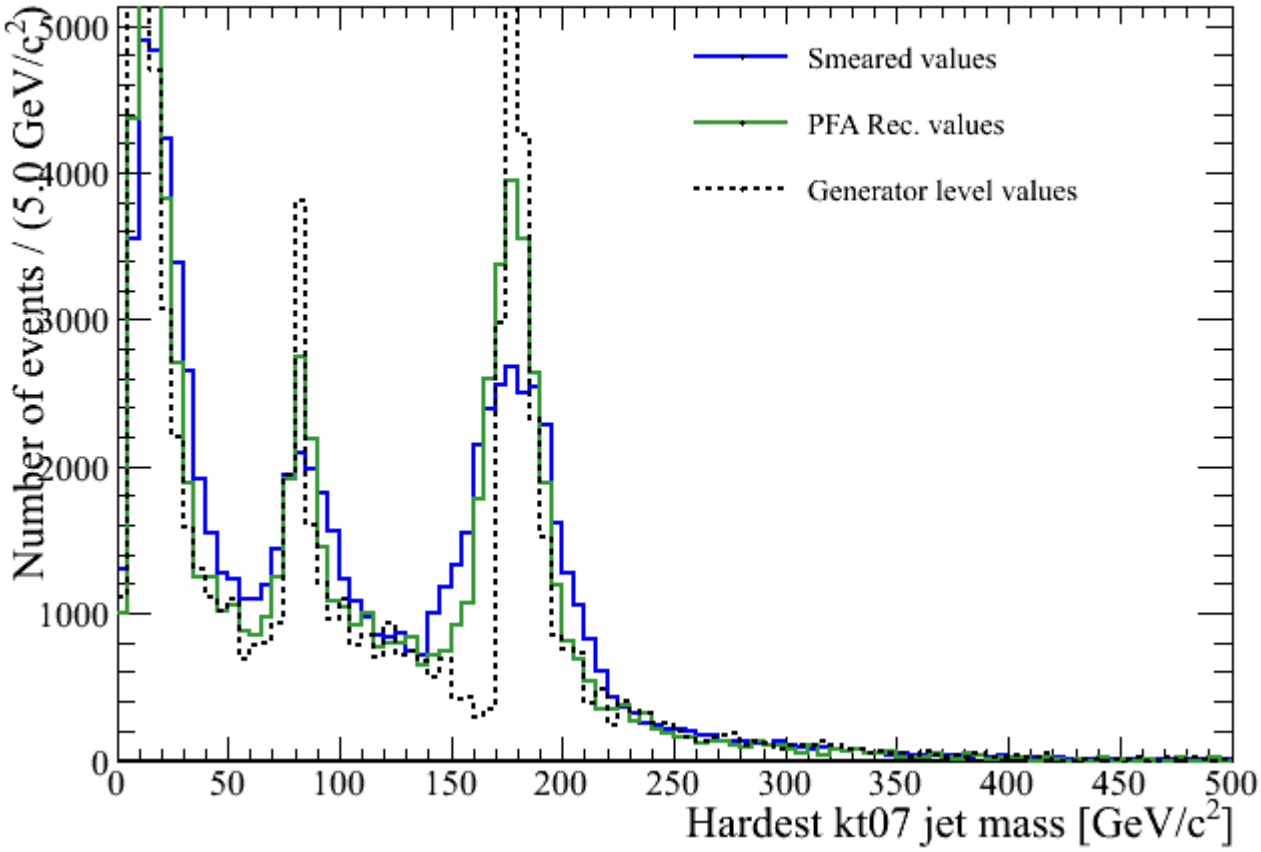
Results on ttbar events

Discrepancies are still conservative



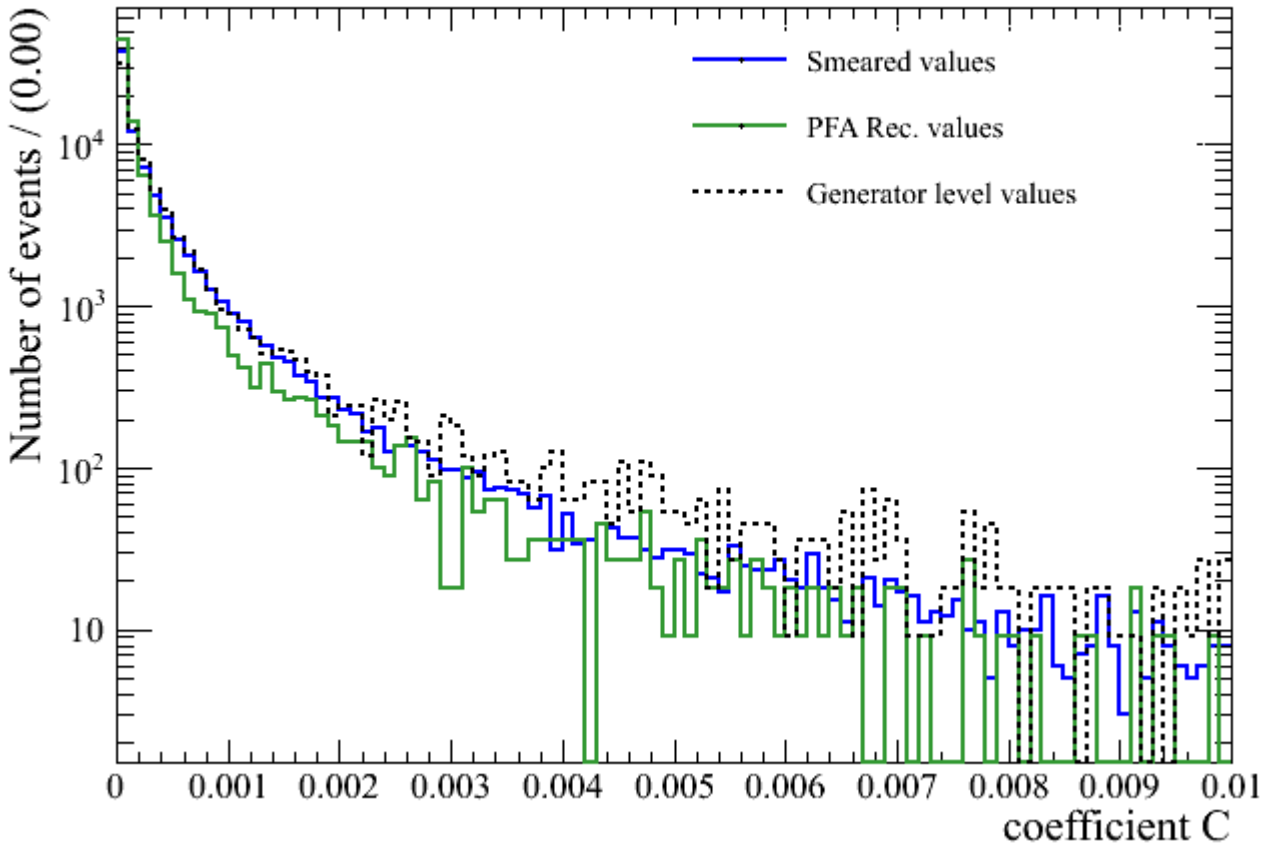
Results on ttbar events

Discrepancies are still conservative



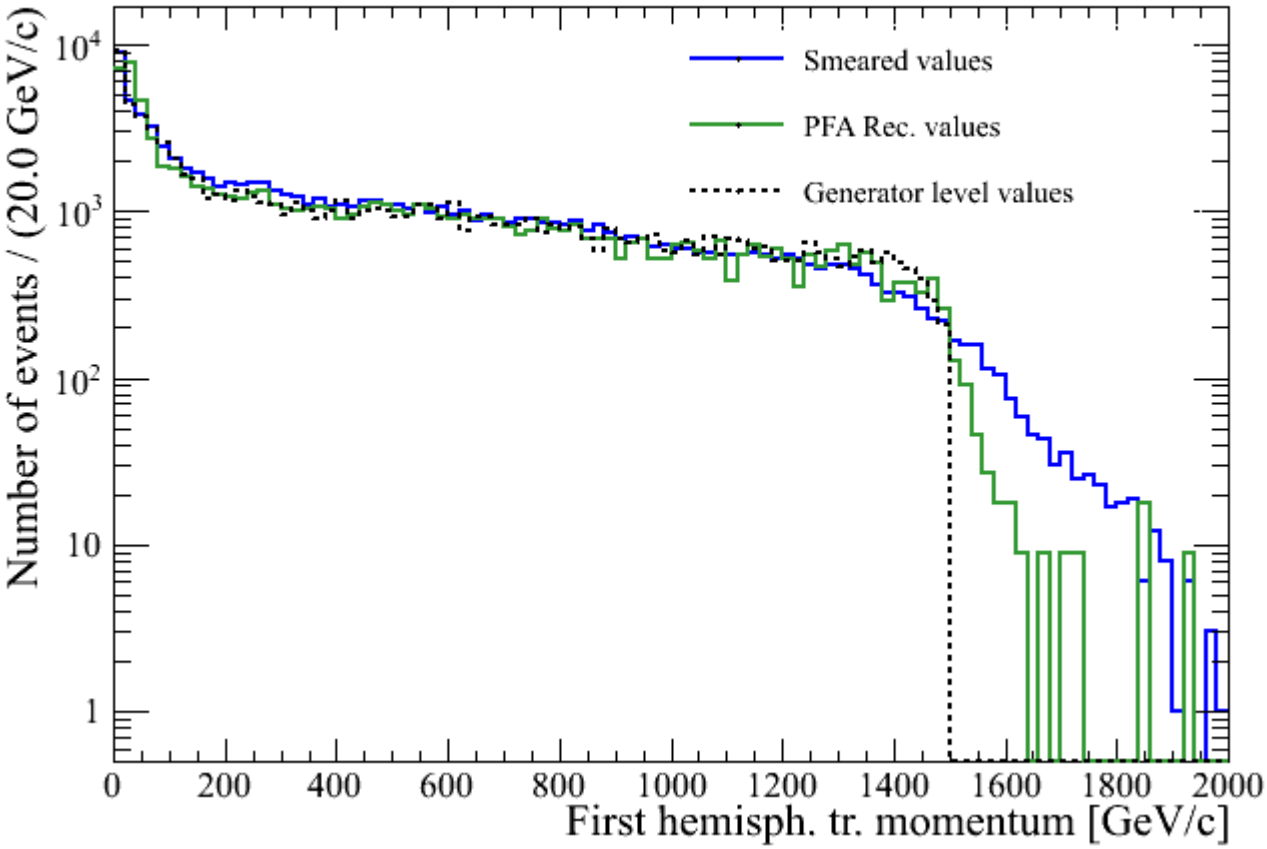
Results on ttbar events

Tails are populated (rather realistically)



Results on ttbar events

Tails are populated (rather realistically)



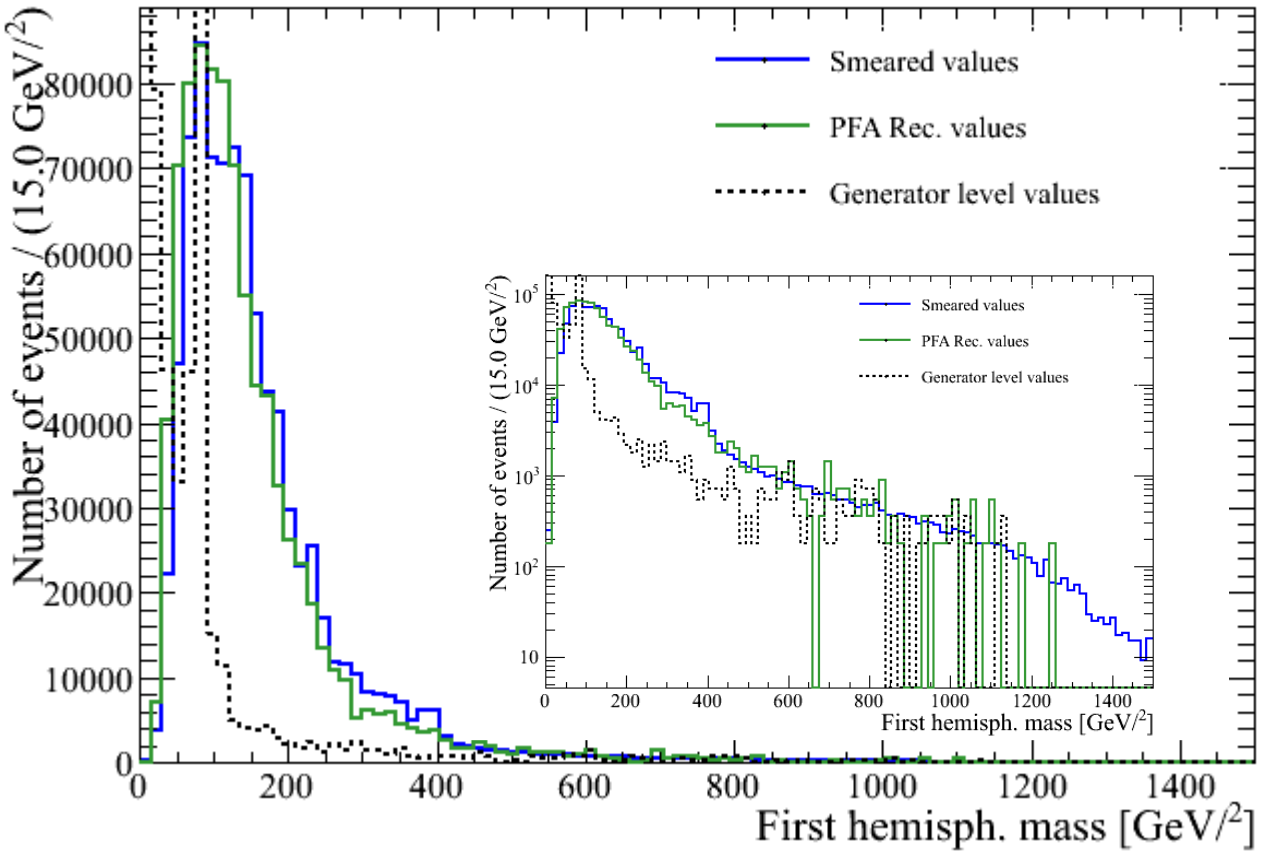
Results on WW events

- Same framework used
  - "Smearing Maps" may then not have optimized size or binning
  - Expect higher discrepancy

- 2M Smeared WW evts out of 10k generated !

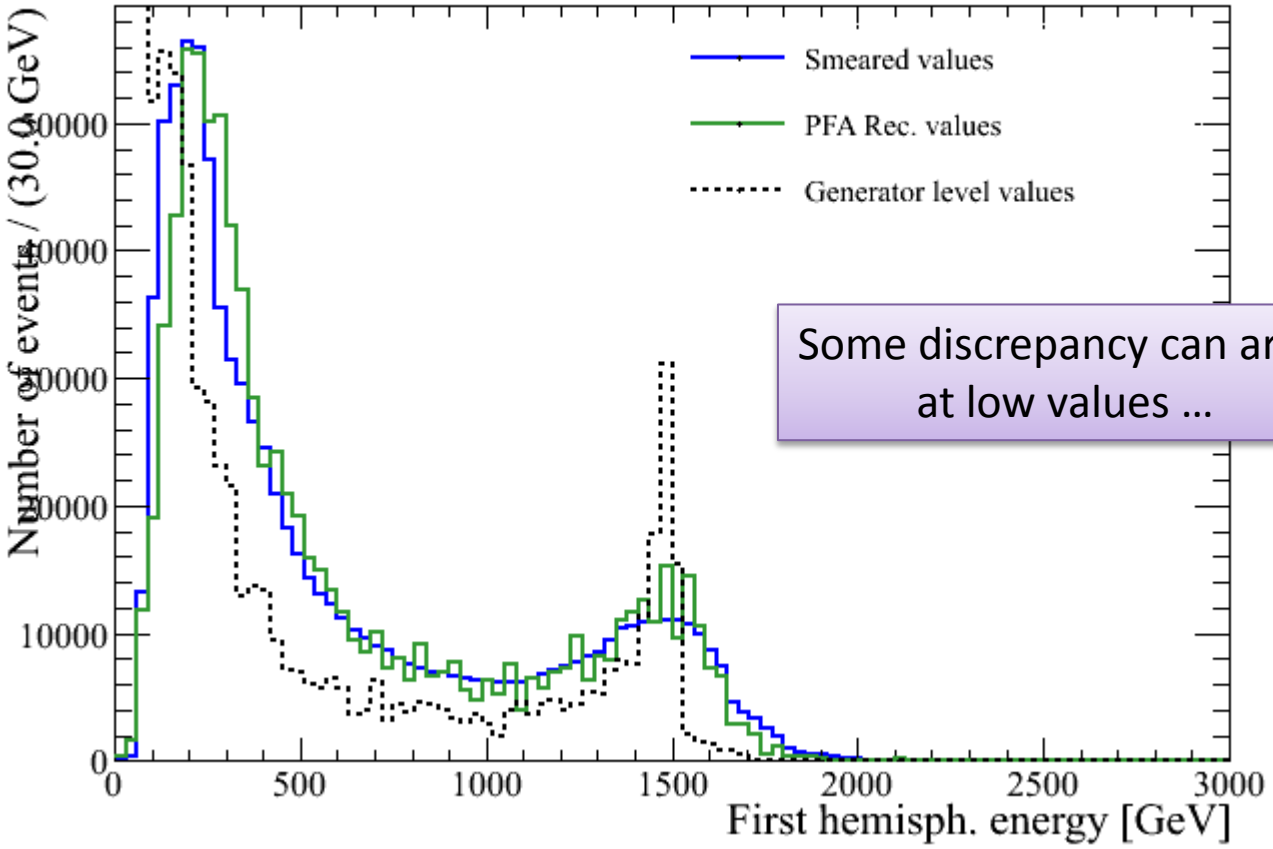
Results on WW events

Non-optimized smearing works already quite well !



Results on WW events

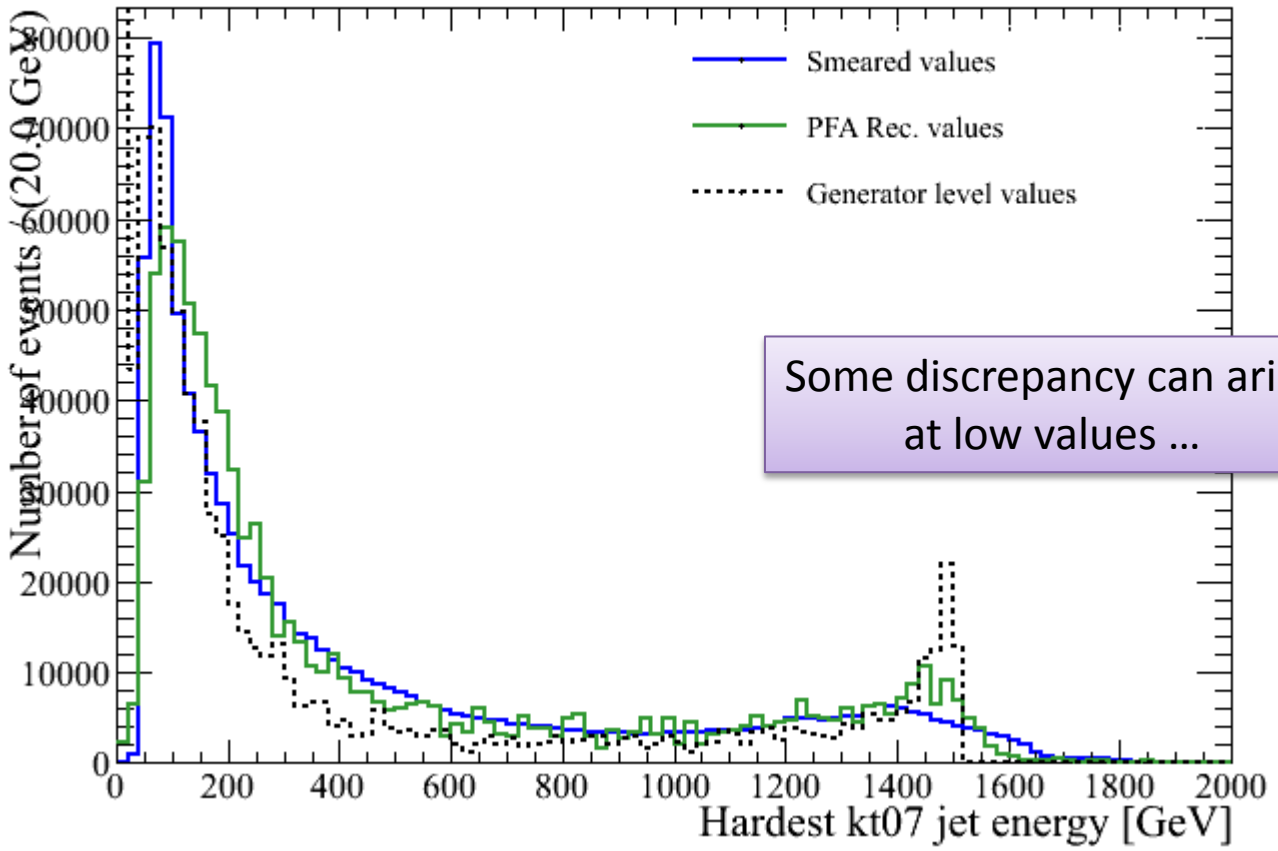
Non-optimized smearing works already quite well !



Some discrepancy can arise at low values …

Results on WW events

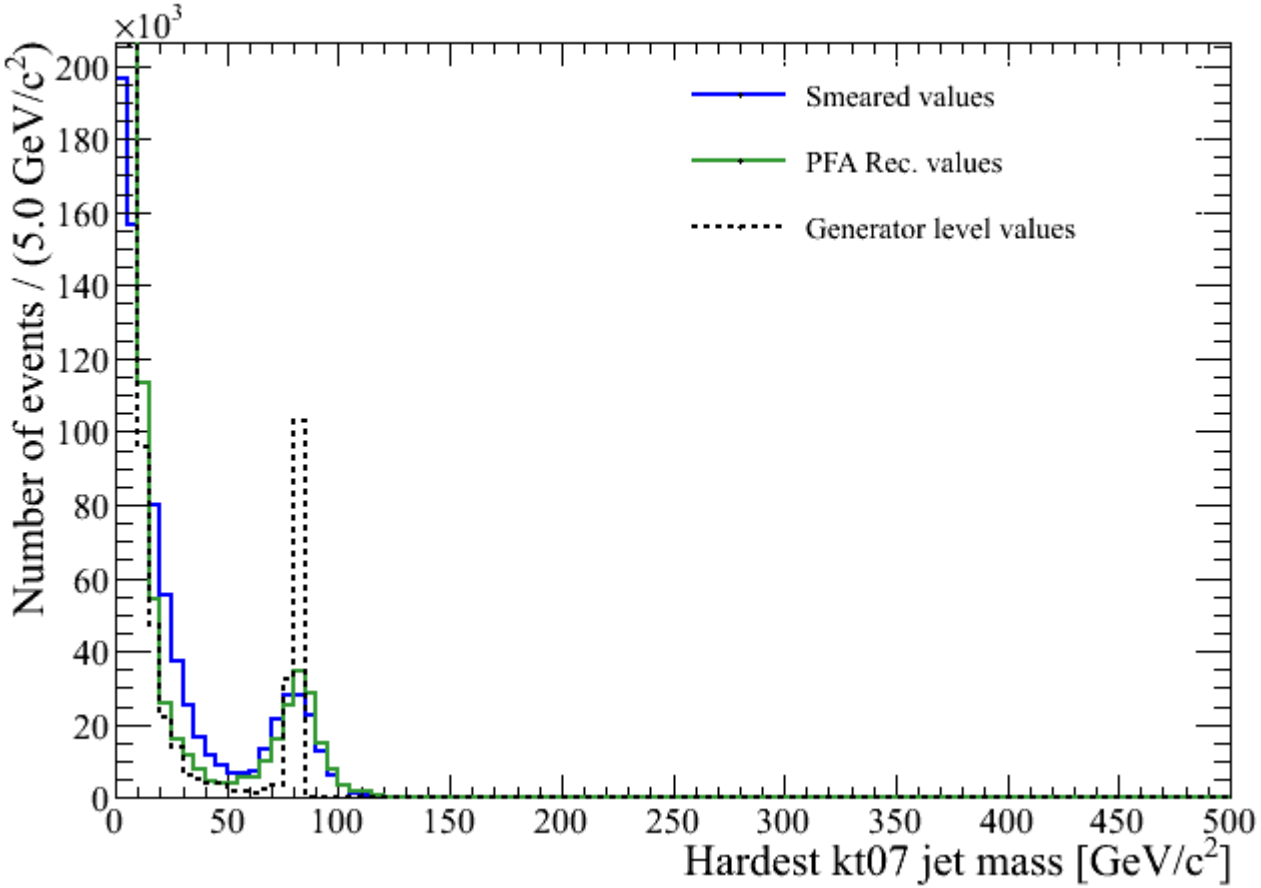
Non-optimized smearing works already quite well !



Some discrepancy can arise at low values …

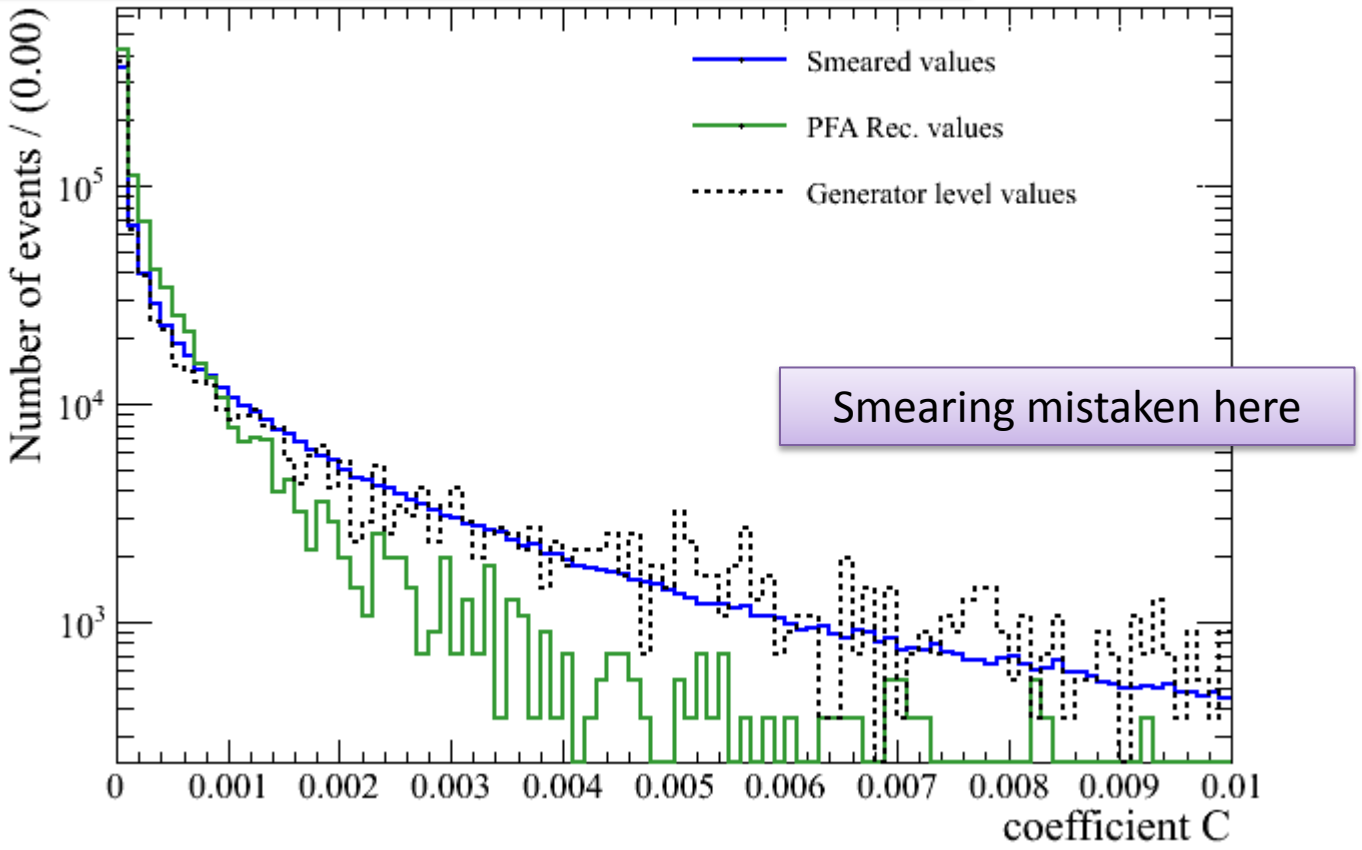
Results on WW events

Non-optimized smearing works already quite well !



Results on WW events

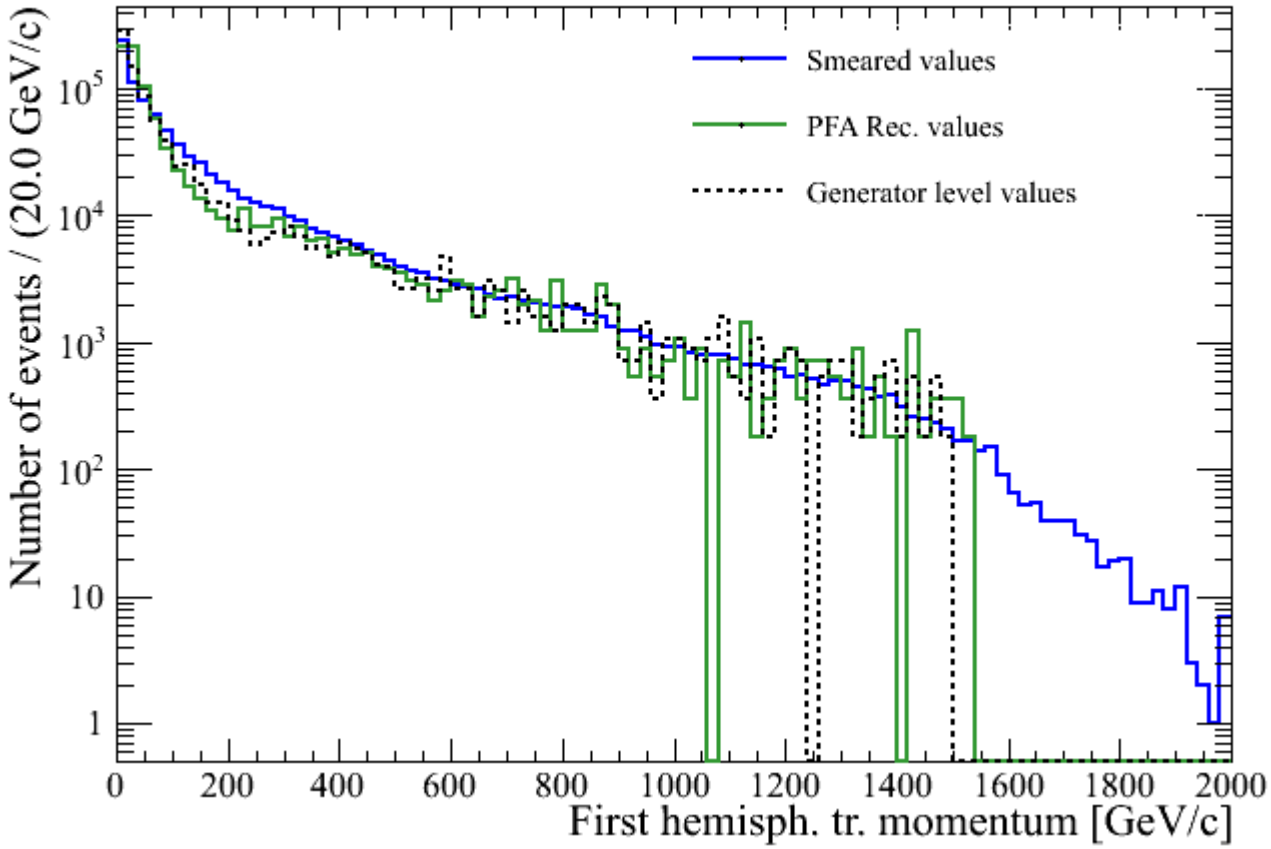
Tail very nicely and smoothly populated at high statistics



Smearing mistaken here

Results on WW events

Tail very nicely and smoothly populated at high statistics



Conclusion

- Advantages
  - Technique very efficient
  - Can produce any amount of realistic data
  - Process mainly automatic
  - Optimisation for every channel possible

- Problems
  - Low values must be looked at and sometimes tricked
  - Each variable must be treated individually
  - Variable correlations have to be checked
  - Optimisation for each channel needed ?

Outlook

- Full Sim. Signal vs Smeared Background not checked yet
  - Lower separation should be expected OK
  - Higher separation may occur ! THEN ?

- If this technique is to be adopted for CLIC CDR Background data:
  - Precise requirements ?
  - Time scale ?
  - Can involve 1 person full time …