Background MC study for MVA

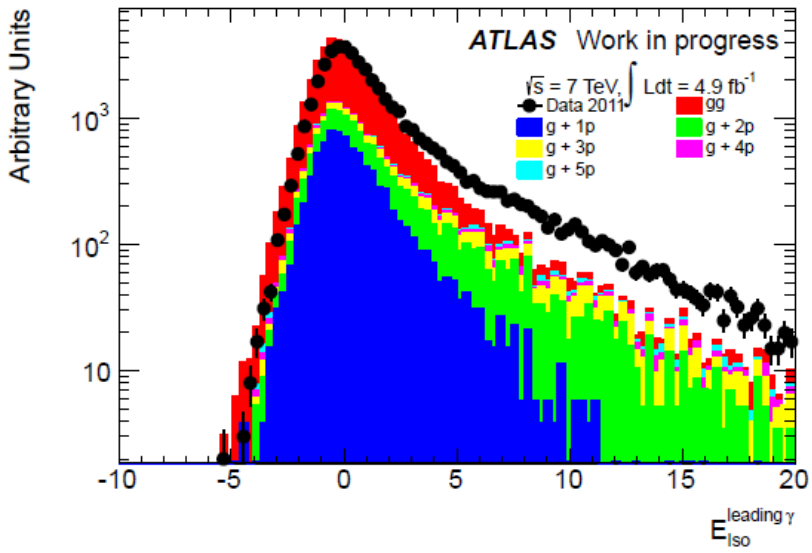
with Olivier's tools

Yuji

Describe Jet-Jet component by data

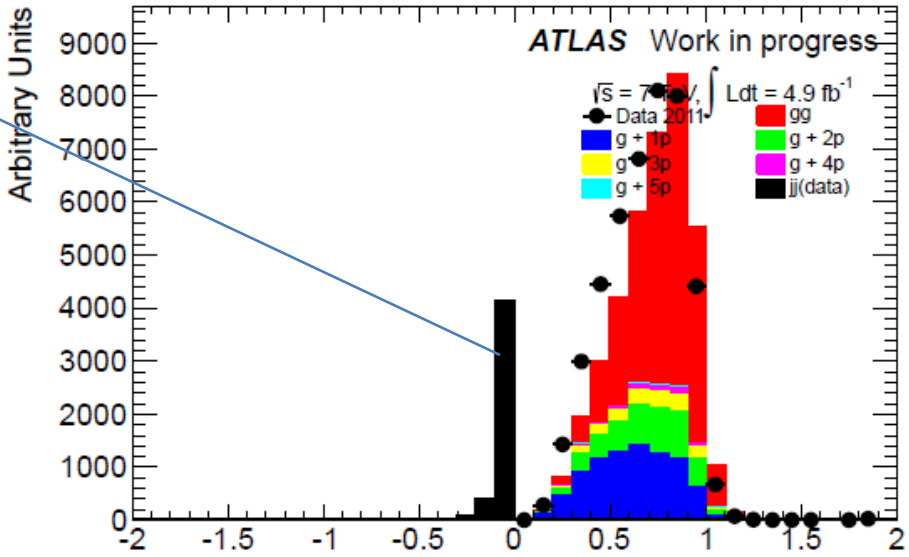
- gamma-gamma : Sherpa
- Gamma-jet : ALPGEN
- Gamma-jet: ignored.

  → Trial with data.



Get jet-jet template by
Photon NN response < 0 for both photon.
Normalize it with Marine's result:

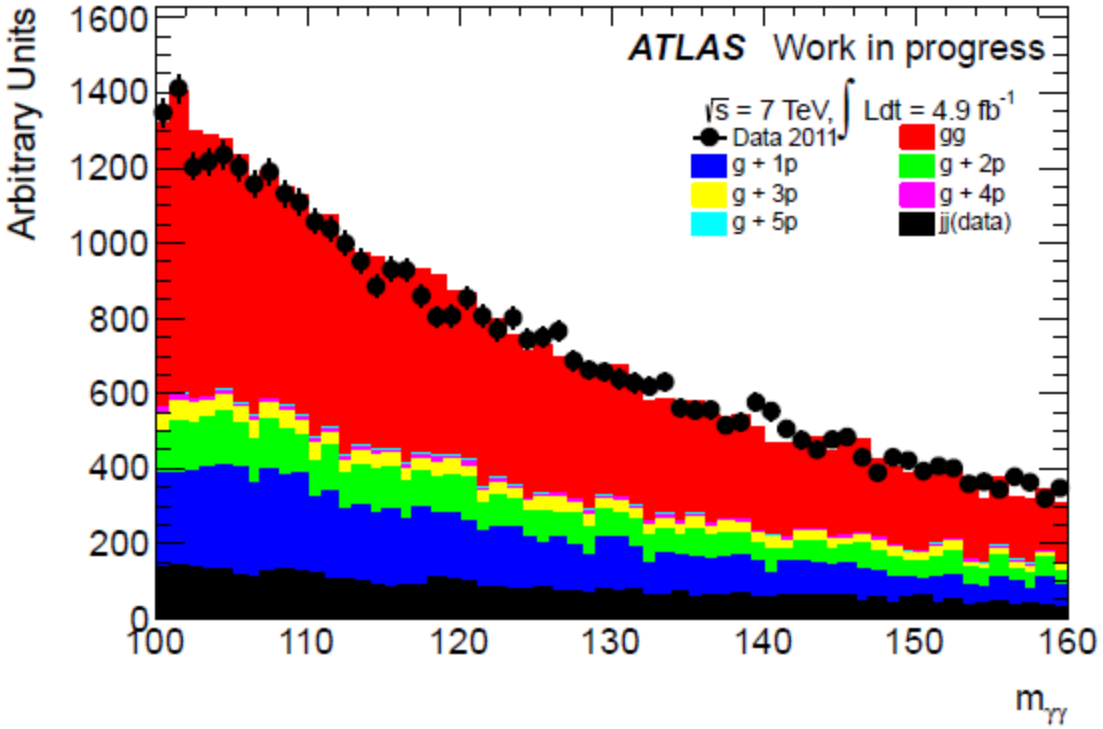
	$N_{\gamma\gamma}$	$N_{\gamma j}$	$N_{j\gamma}$	N_{jj}
Iso 4GeV, pt 30, 30	18388±164	3453±49	1392±40	469±11
Iso 4GeV, pt 40, 30	17119±156	3203±45	1145±34	320±9
Iso 7GeV, pt 30, 30	20039±178	6118±86	2290±66	1340±32
Iso 7GeV, pt 40, 30	18656±170	5737±81	1914±58	963±26
Iso 20GeV, pt 30, 30	20185±180	10390±146	3573±102	4138±98
Iso 20GeV, pt 40, 30	18801±171	9922±140	3078±93	3225±86

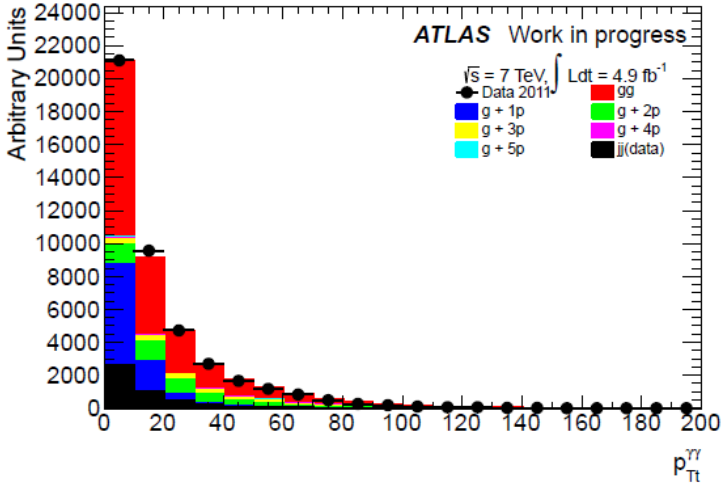


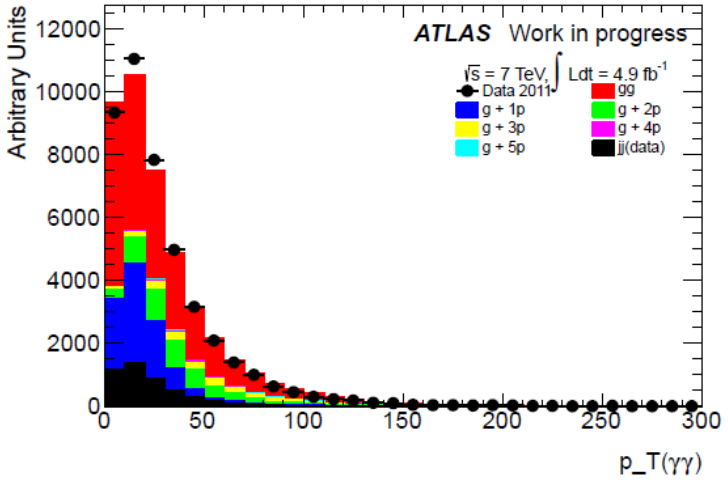
Photon ID NN response, 2nd leading photon

Some outputs

- Pt>30 GeV, TopoIso<20 GeV (both photon)
  - 100 < Mgg < 160 GeV

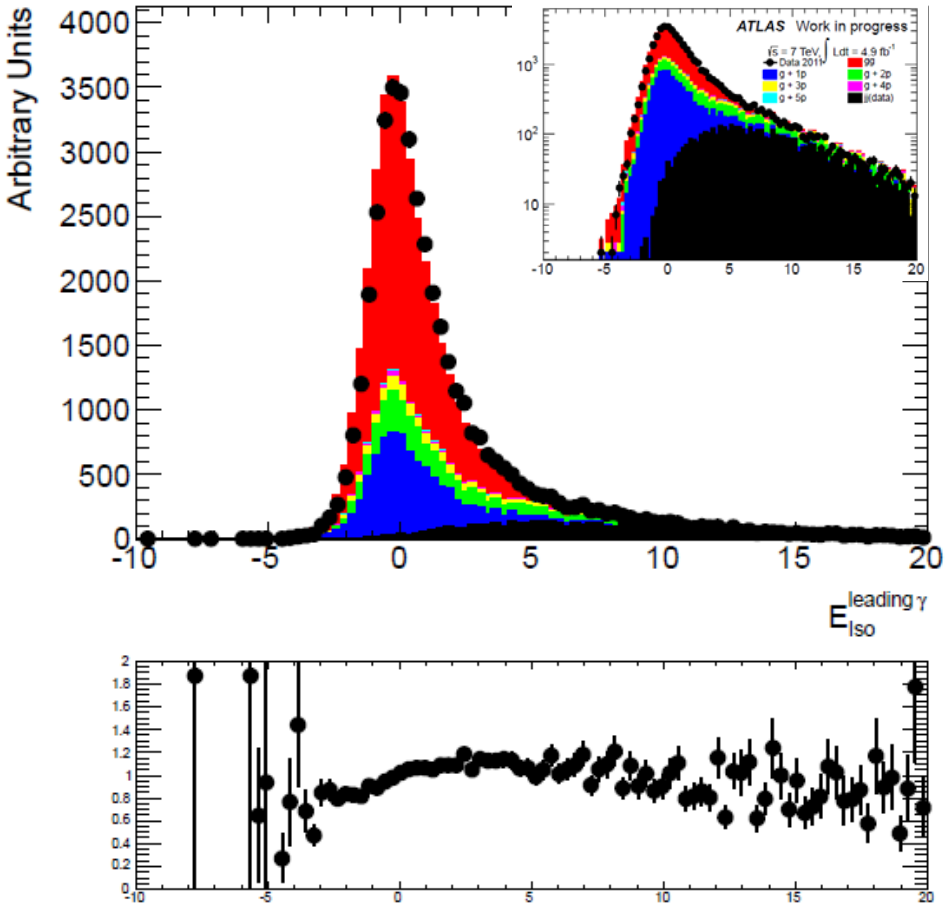


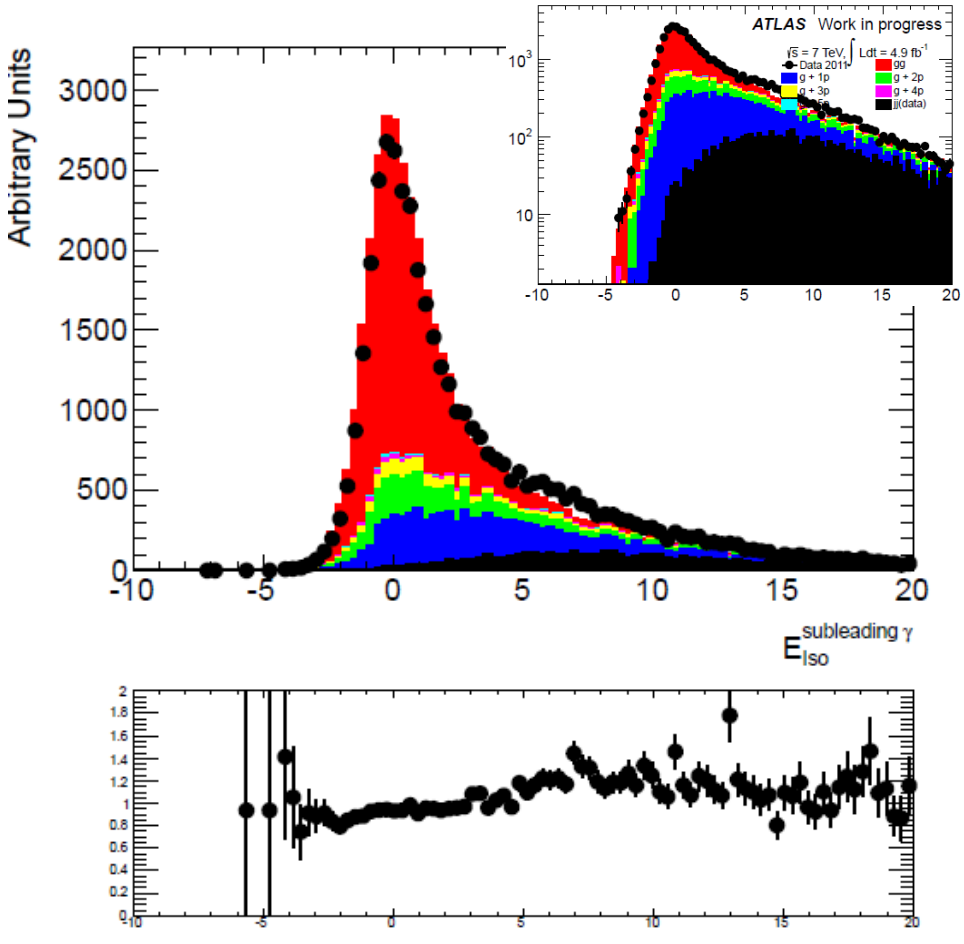




Some outputs

- Pt>30 GeV, TopoIso<20 GeV (both photon)
  - 100 < Mgg < 160 GeV

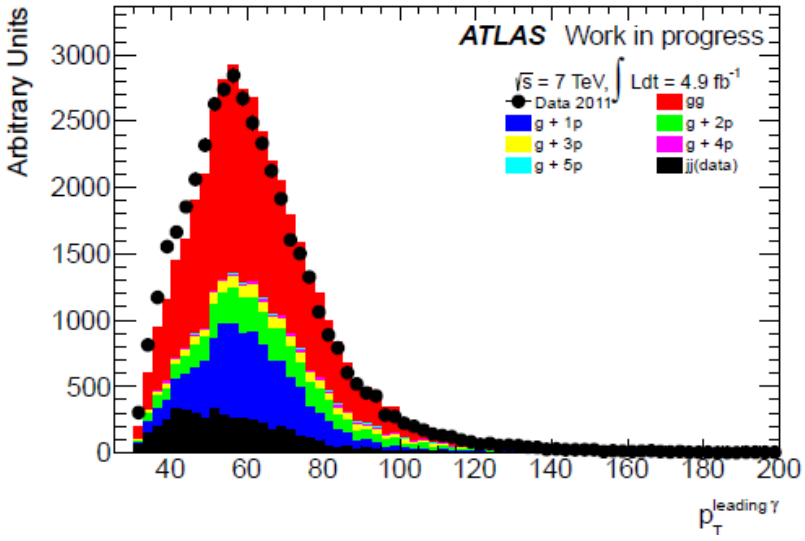


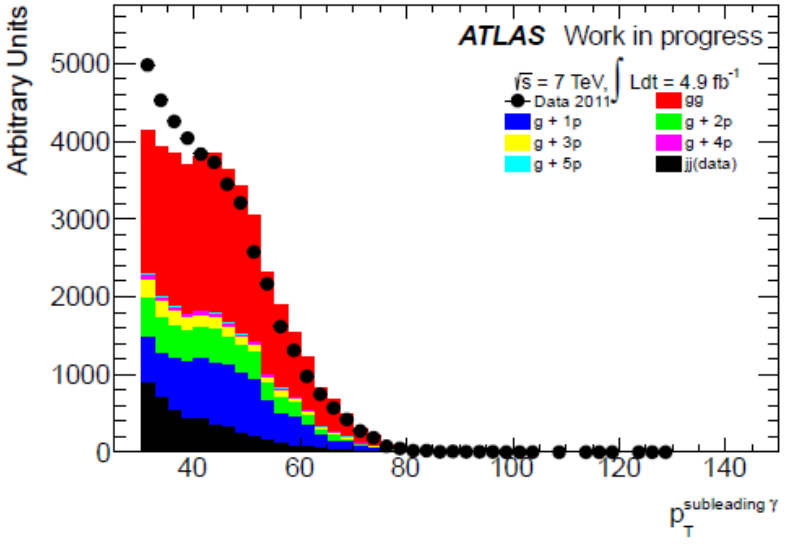


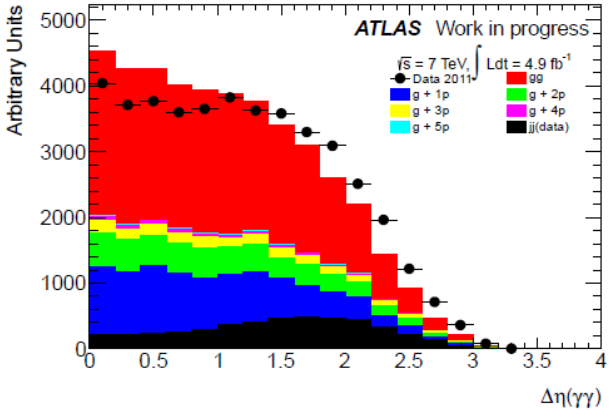
Some outputs

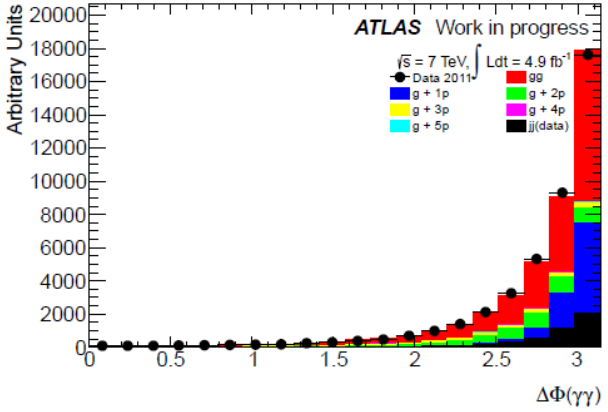
- Pt>30 GeV, TopoIso<20 GeV (both photon)
  - 100 < Mgg < 160 GeV

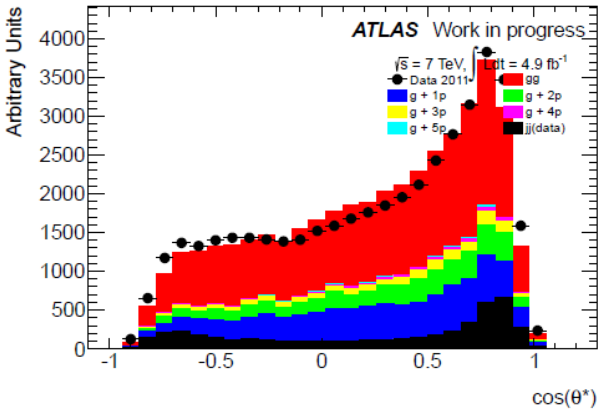
Without Deta reweighting.







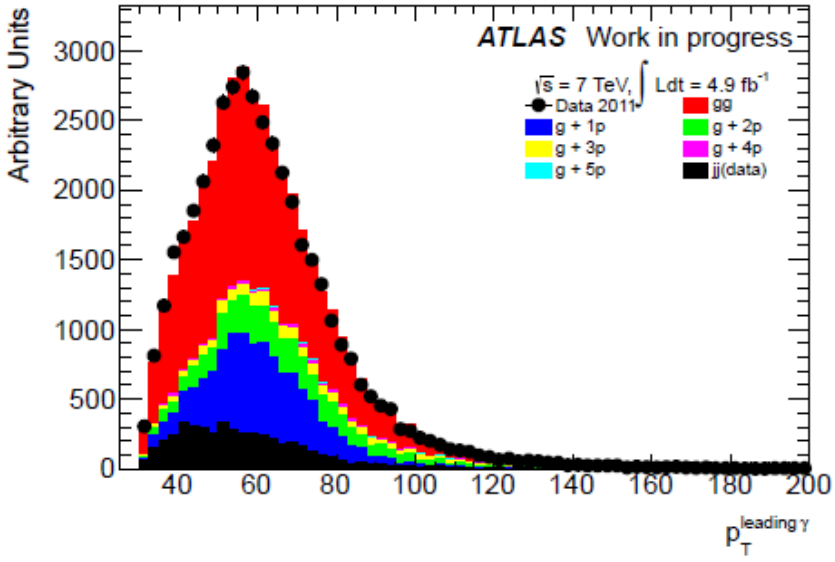


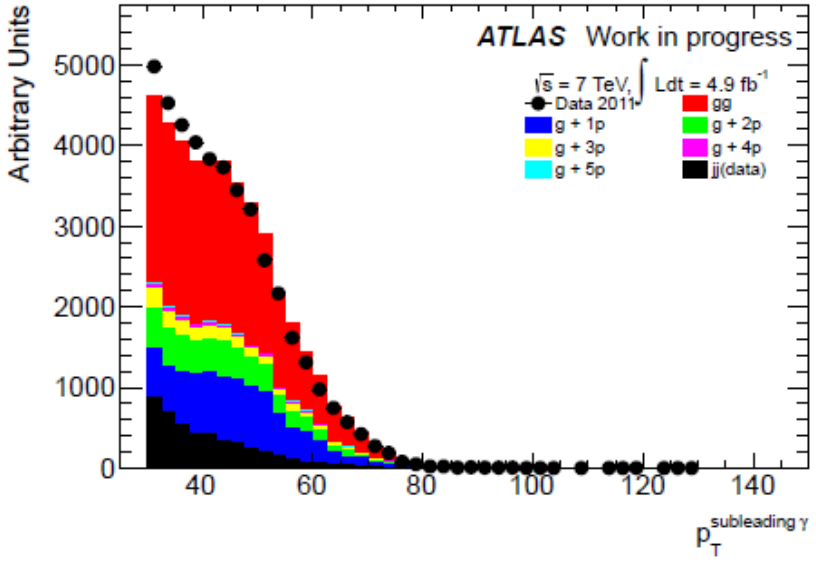


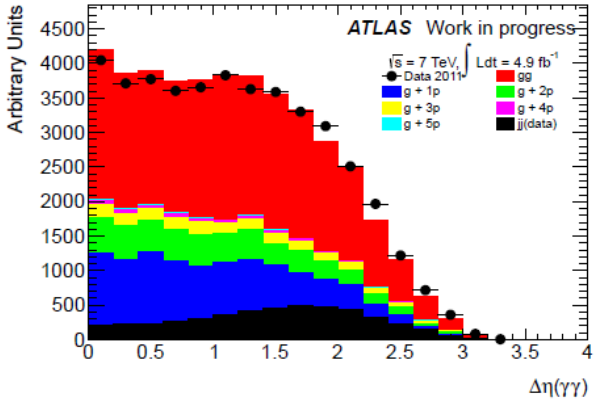
Some outputs

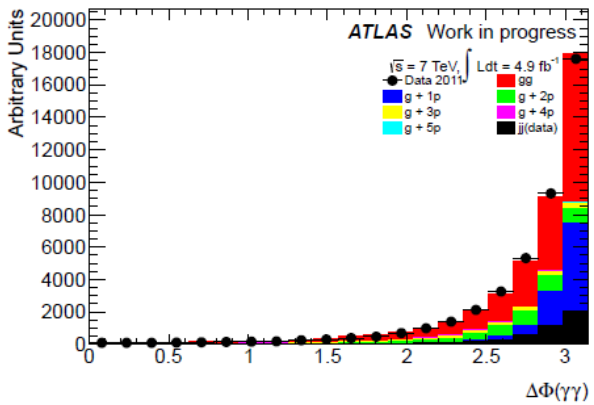
- Pt>30 GeV, TopoIso<20 GeV (both photon)
  - 100 < Mgg < 160 GeV

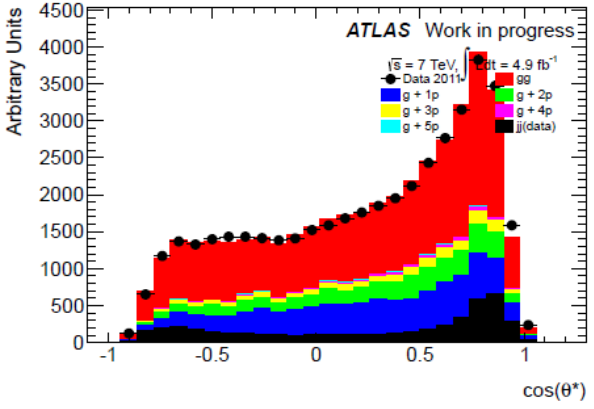
With Deta reweighting.



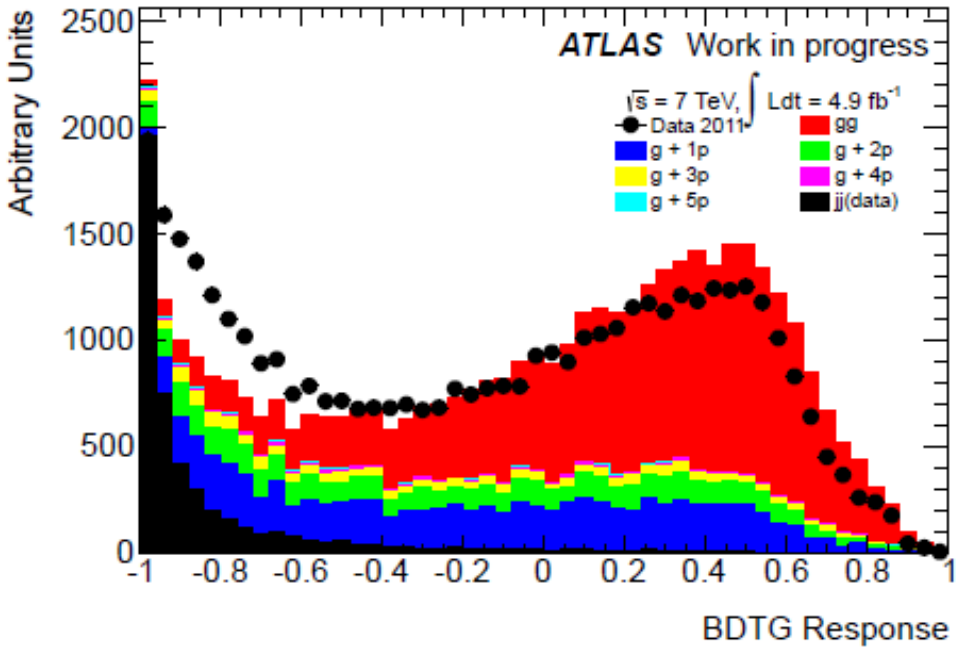








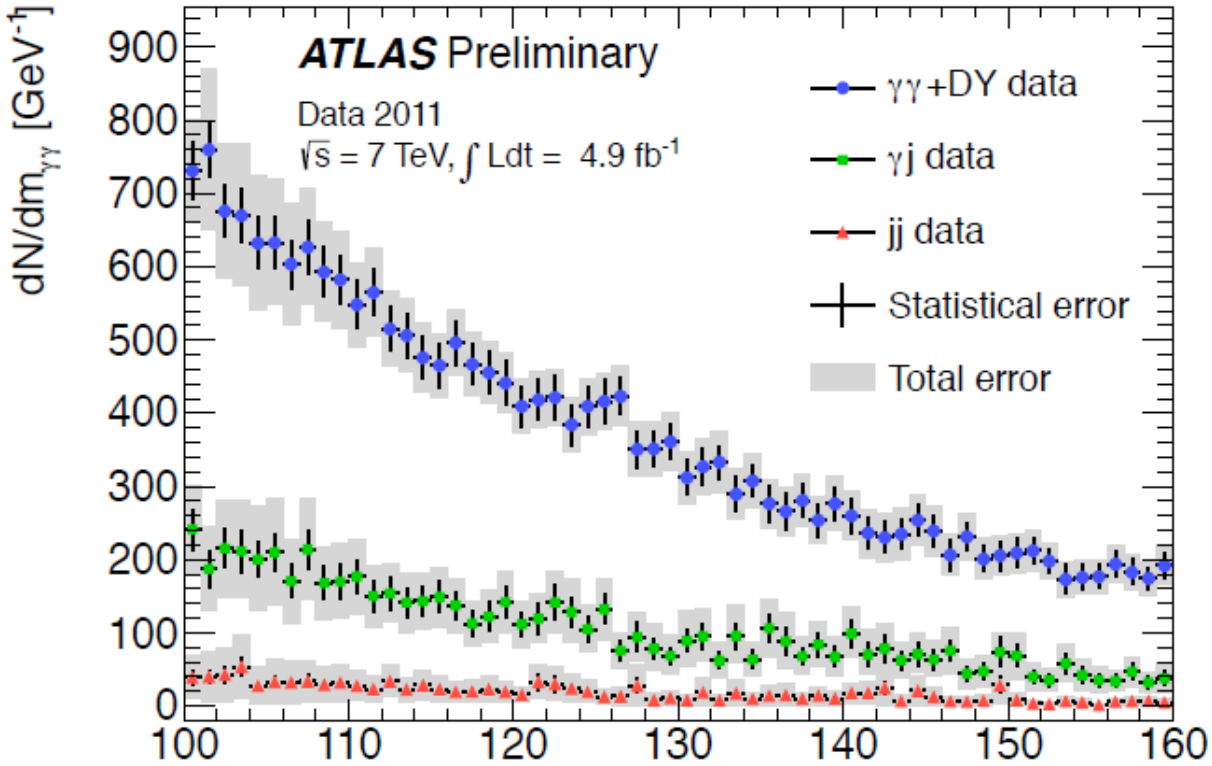
BDT output with reweighting



- Need more work to describe low BDT region.

BACKUPs!!!!

Why we want to have Monte Carlo?



- Now we estimate our background ($\gamma\gamma$, γj and jj) based on data.
- MVA needs independent sample to have a good discriminant. (i.e. can not train MVA with real data in signal region)
- Today's trial with mc11c
  - $\gamma\gamma$ → Pythia or Sherpa
  - γj: (+ jj)→ ALPGEN+Jimmy: gamma+ (1lp+2lp+3lp+4lp+5lp)
-

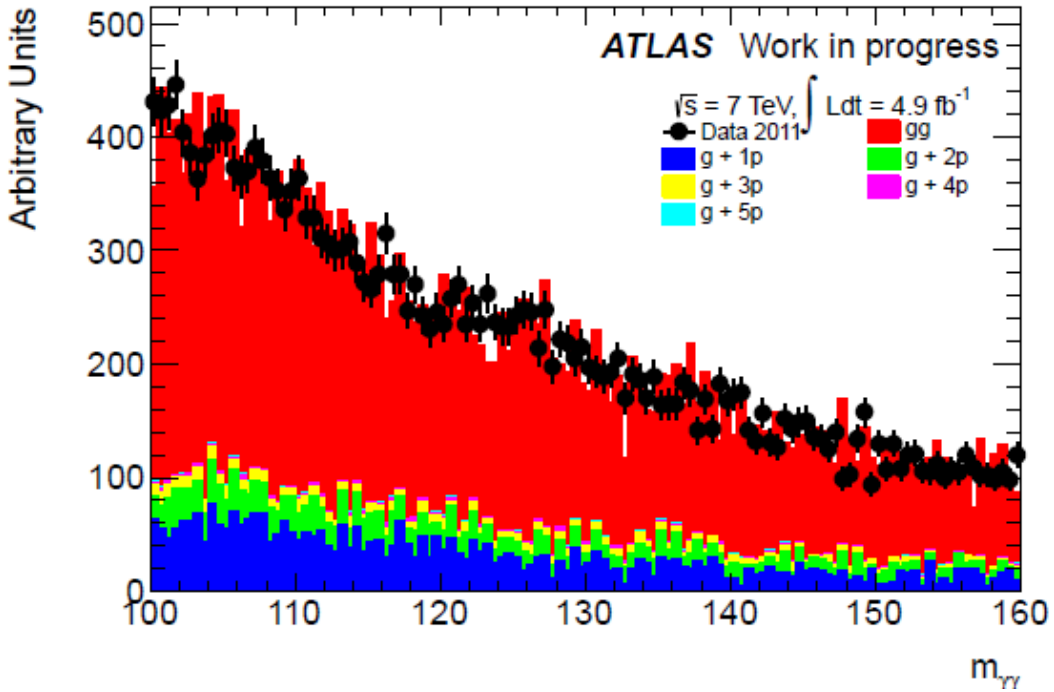
Normalization of MC samples

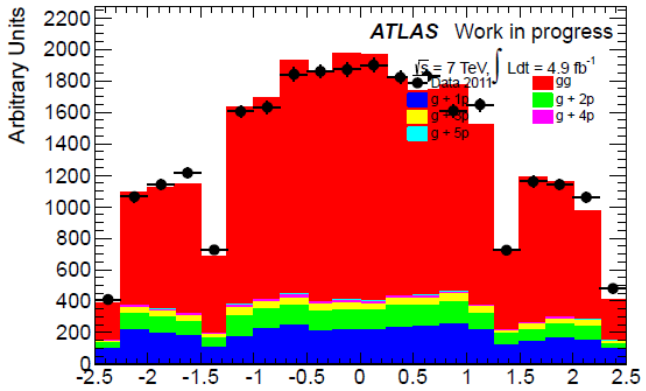
- Set number of background equal to number of data
  - NN photon, Etopoiso<4 GeV,100<mgg<160
    $pT_{\gamma 1}$>40 GeV, $pT_{\gamma 2}$>25 GeV
- Shape of $\gamma\gamma$, γj+jj from MC (mc11c)
  - $\gamma\gamma$: Pythia (0.2 M events) or Sherpa (1 M events)
  - γj : ALPGEN γ+(1lp,2lp,3lp,4lp,5lp)
    - No jet-jet sample included in today's study
    - Set sample composition based on LO cross section, # of generated MC events and filtering Efficiency
- Use Ratio of $\gamma\gamma/(\gamma\gamma+\gamma j+jj)$ **= 0.75** (from Marine) from 2D fit method

	γ+1lp	γ+2lp	γ+3lp	γ+4lp	γ+5lp
# of events	10M	10M	10M	3M	1.9M
X section (nb)	74.2	21.6	5.86	1.36	0.35
Filtering eff.	0.109	0.311	0.467	0.624	0.761

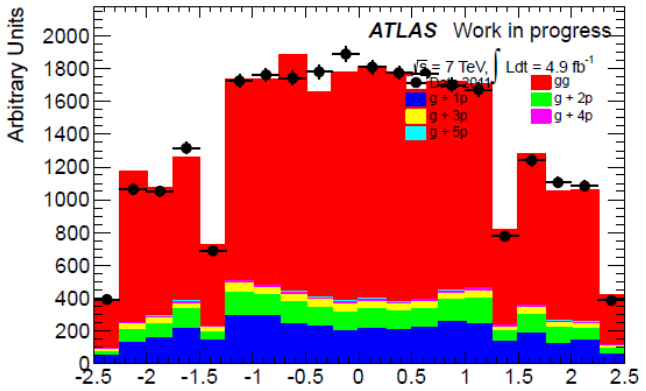
In case of γγ (Pythia) and γj+jj (Alpgen)

- Pythia can describe diphoton mass, photon eta





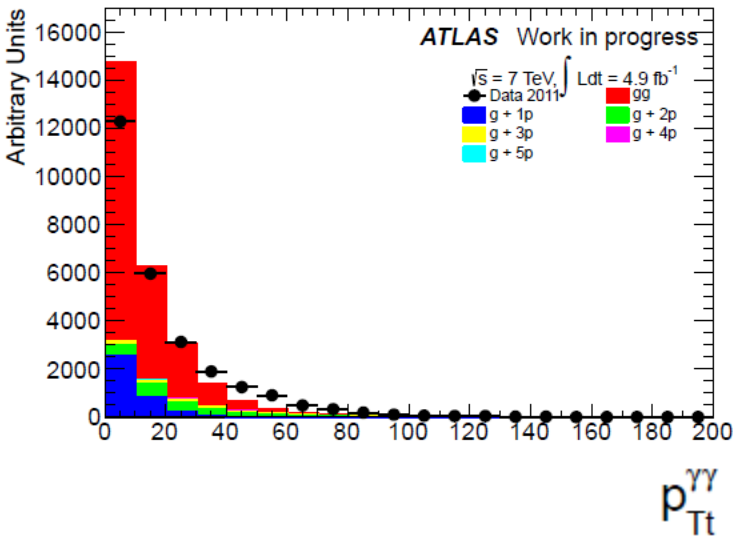
Leading photon eta

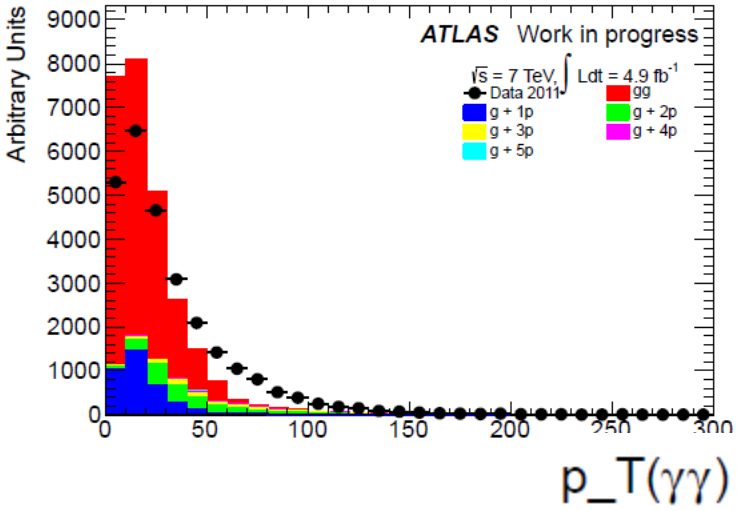


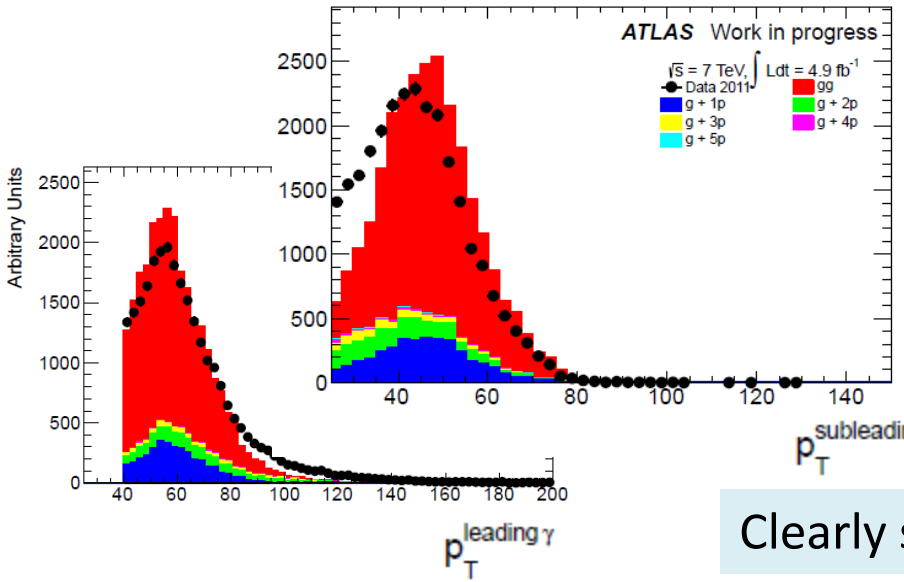
2nd Leading photon eta

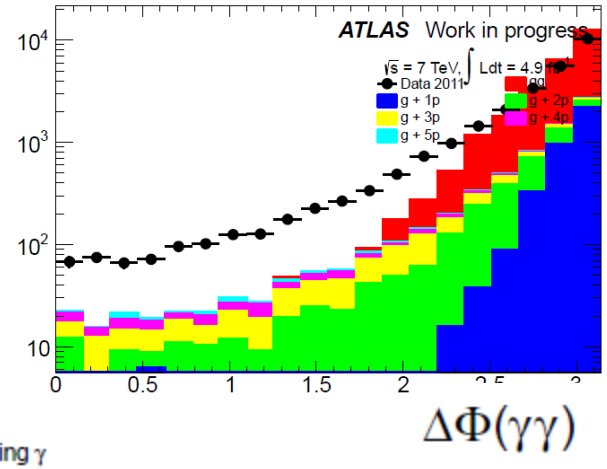
In case of γγ (Pythia) and γj+jj (Alpgen)

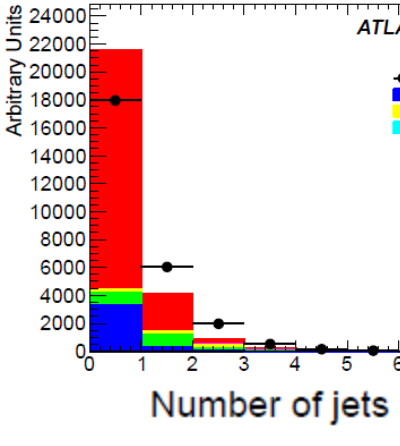
- However, **NOT** other variables







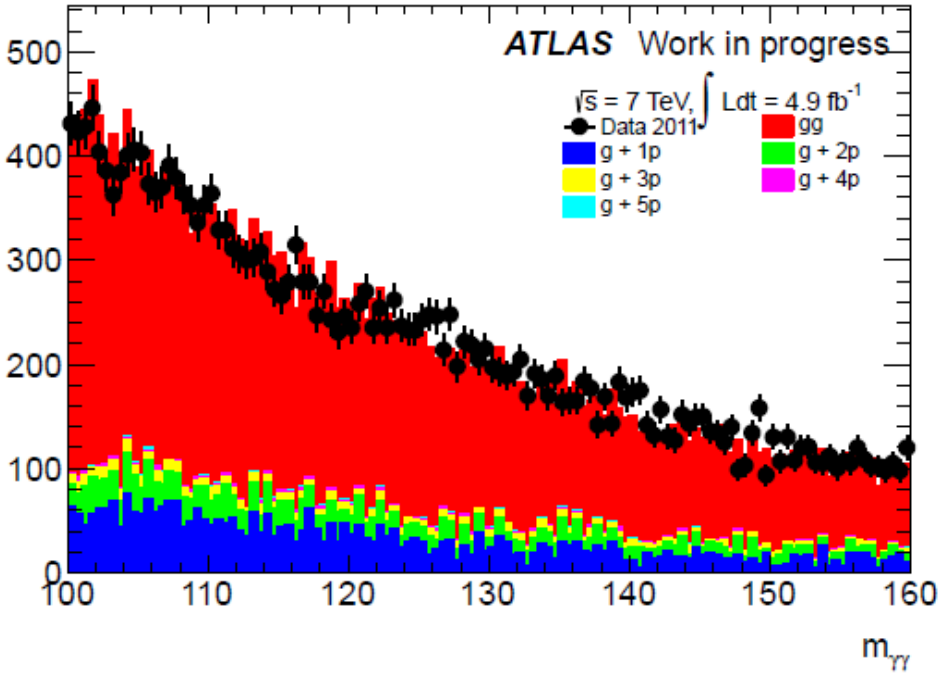


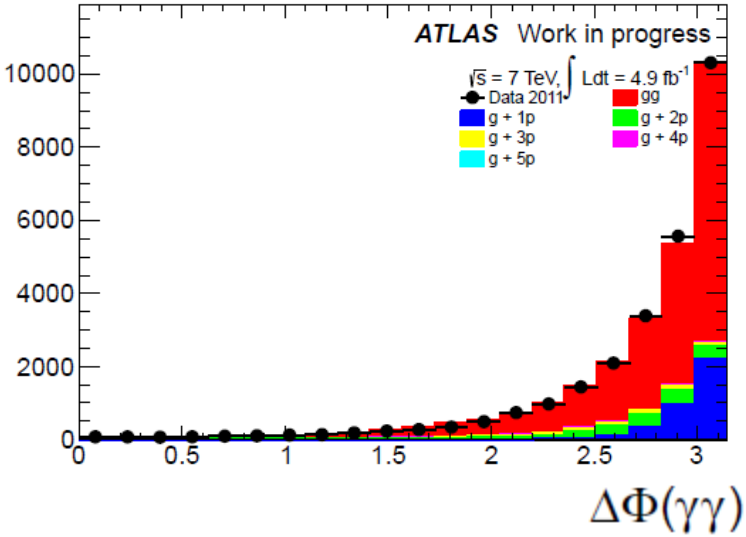


Clearly some component is missing. (i.e. recoil jets)

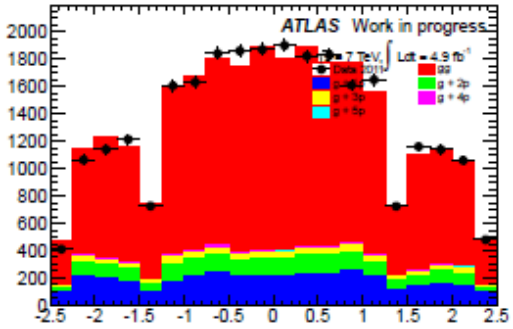
In case of γγ (Sherpa) and γj+jj (Alpgen)

- Sherpa can describe many variables very well!

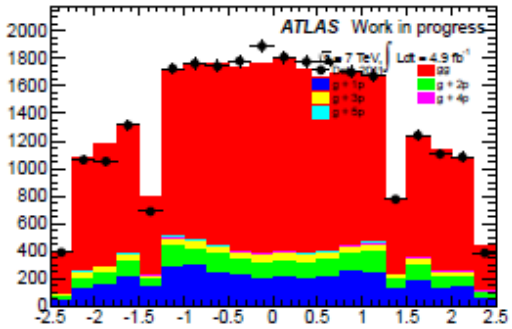




Δϕ is very good now



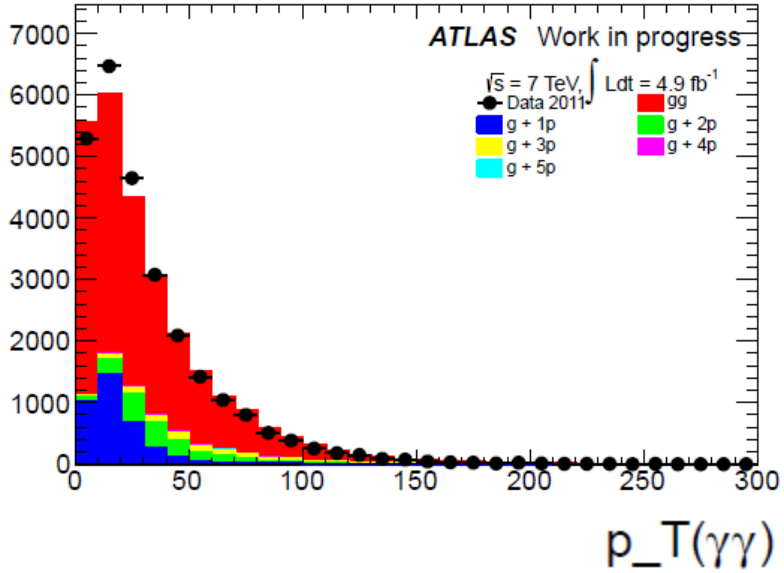
Leading photon eta

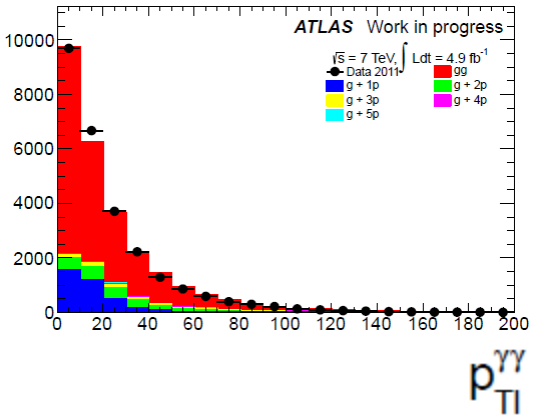


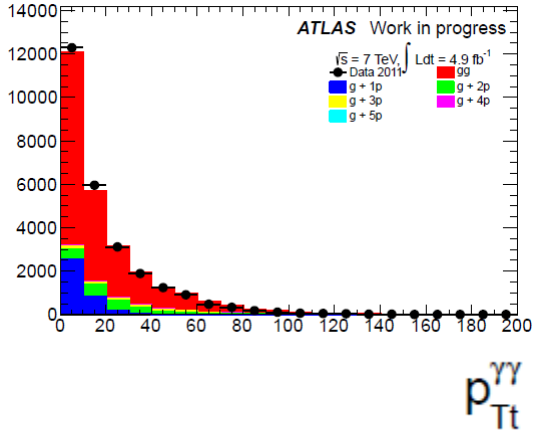
2nd Leading photon eta

In case of γγ (Sherpa) and γj+jj (Alpgen)

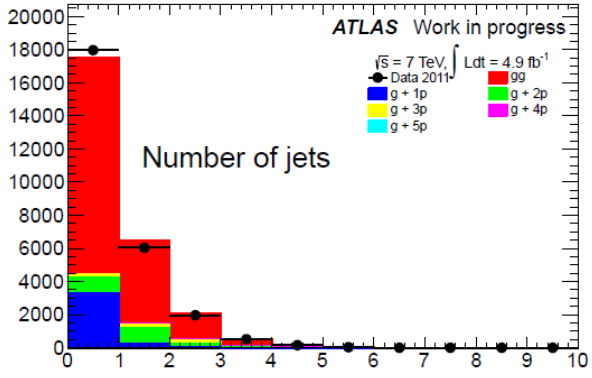
- Sherpa can describe many variables very well!

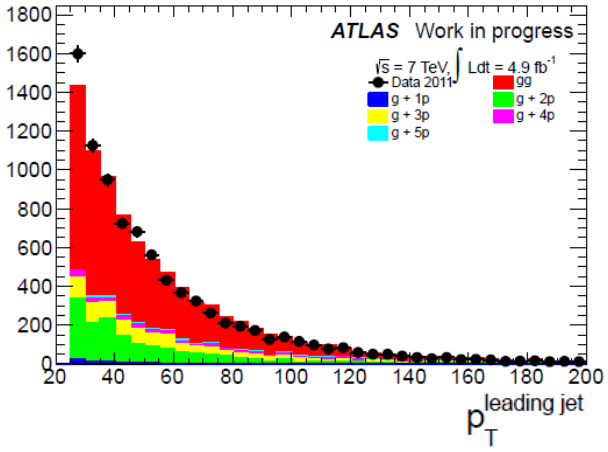


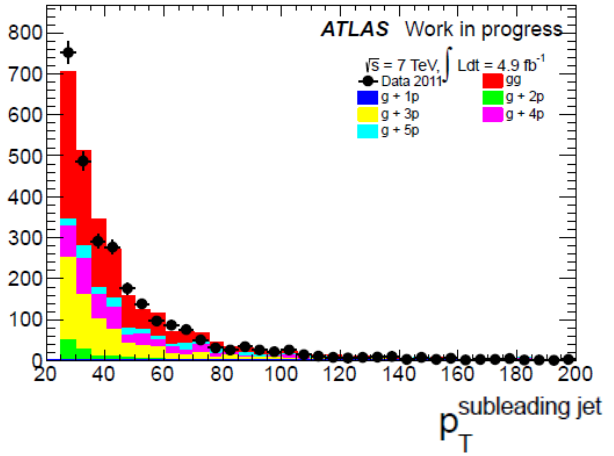




Also Jets variables

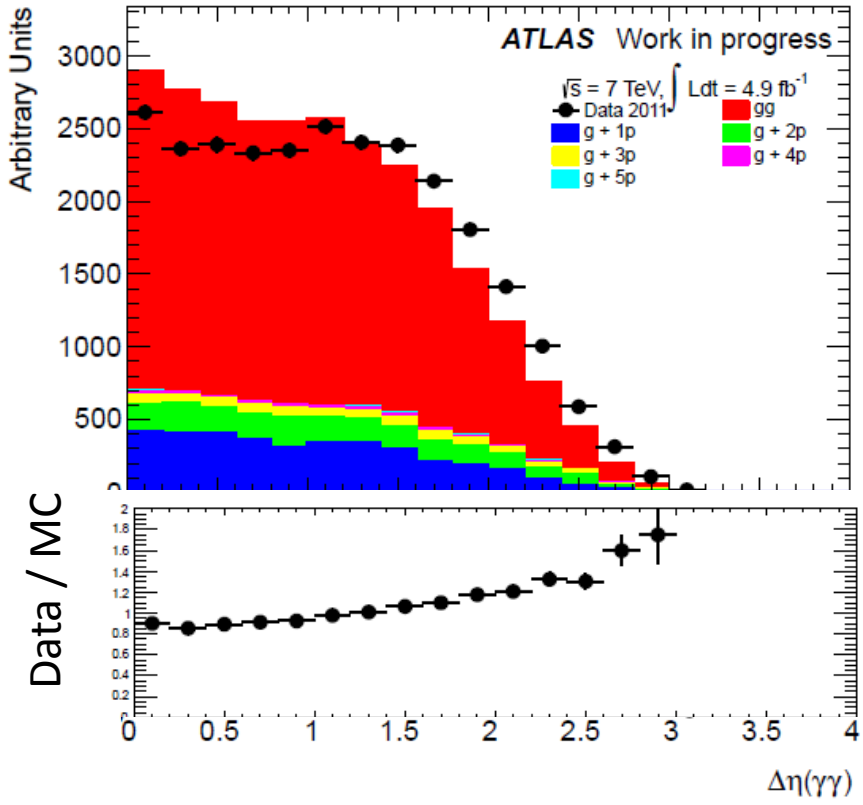


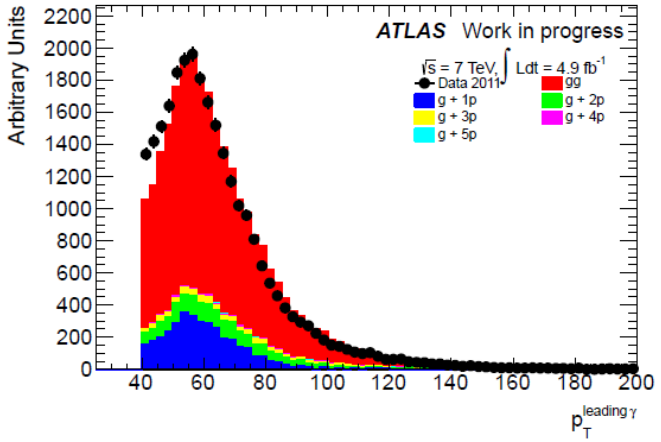


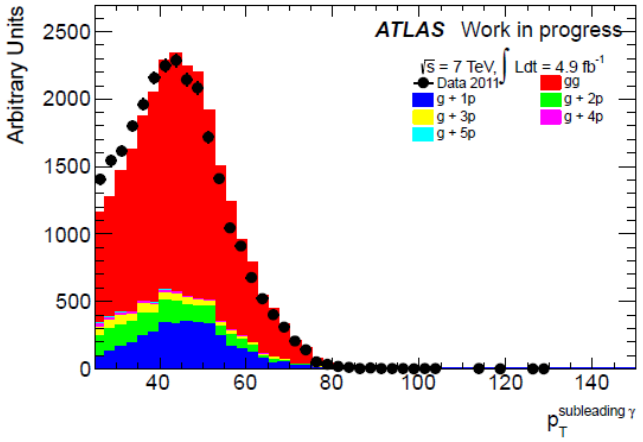


In case of γγ (Sherpa) and γj+jj (Alpgen)

- However, it's not perfect.





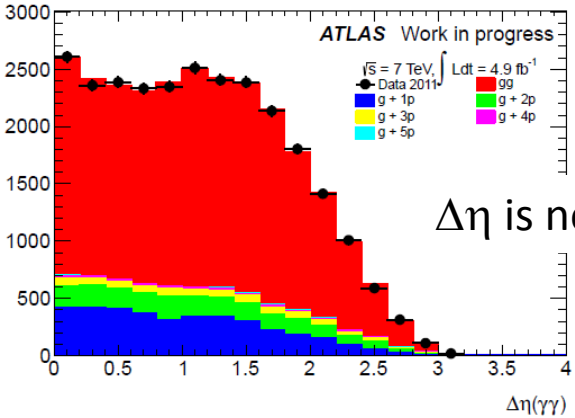


There are some possibilities:
- Still missing jet-jet component.
- Mis-modeling on Sherpa?

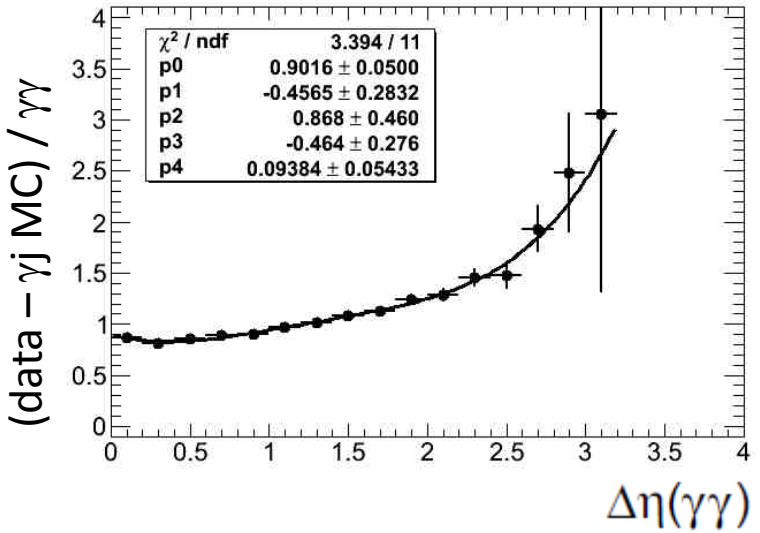
$\Delta\phi$ get improved significantly, but not for $\Delta\eta$? Let's reweight on $\Delta\eta$!

Sherpa with reweighting on $\Delta\eta(\gamma 1,\gamma 2)$

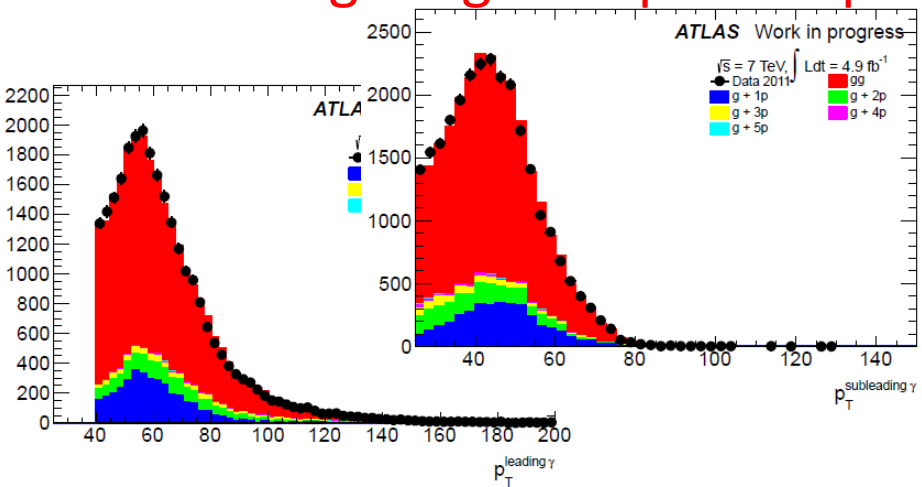
- Reweight Sherpa $\gamma\gamma$ MC on $\Delta\eta_{\gamma\gamma}$ to match data
  - Ratio: (data – γj MC) / $\gamma\gamma$
  - Fit with 4th polynomial.
- Apply only $\gamma\gamma$ MC.

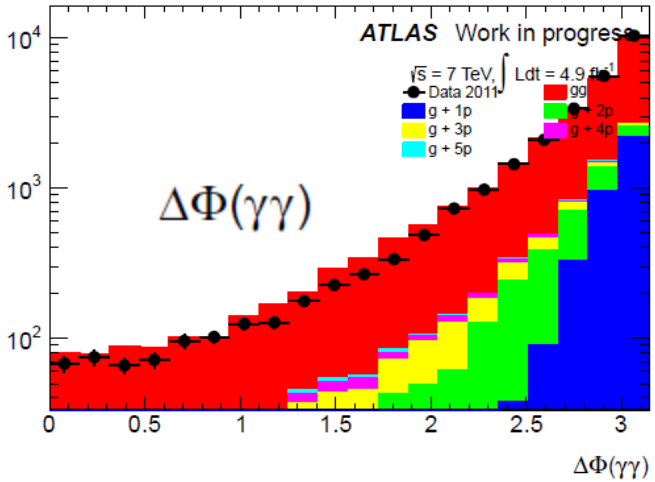


$\Delta\eta$ is now perfect.



This reweighting fixes photon pT!



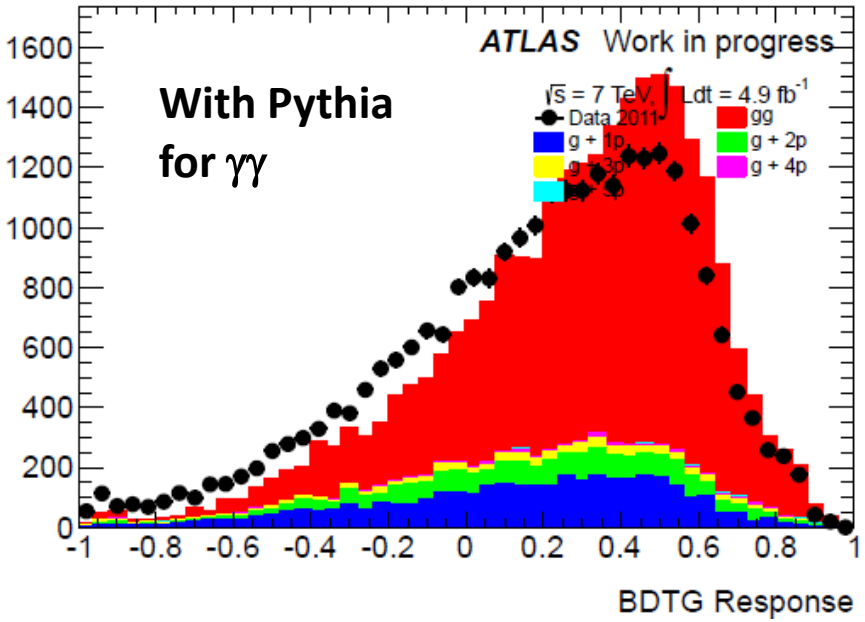


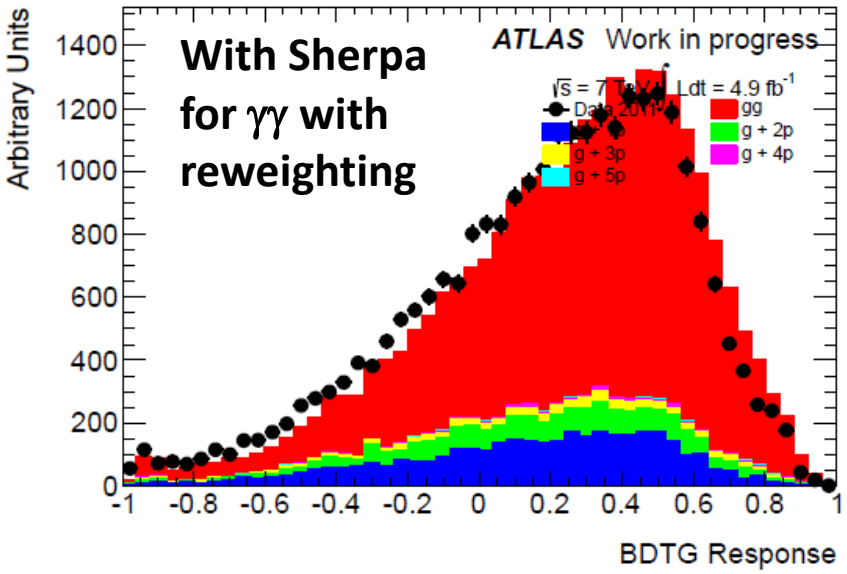
"Good" variables stay as "Good"!!

BDT distribution

- Check Olivier's BDT output (presented at HSG1 meeting at 2012/3/29
  - BDT is trained on data.

Selection in these plots:
NN photon, Etopoiso<4 GeV,
100<mgg<160, pT>40 GeV, pT>25 GeV



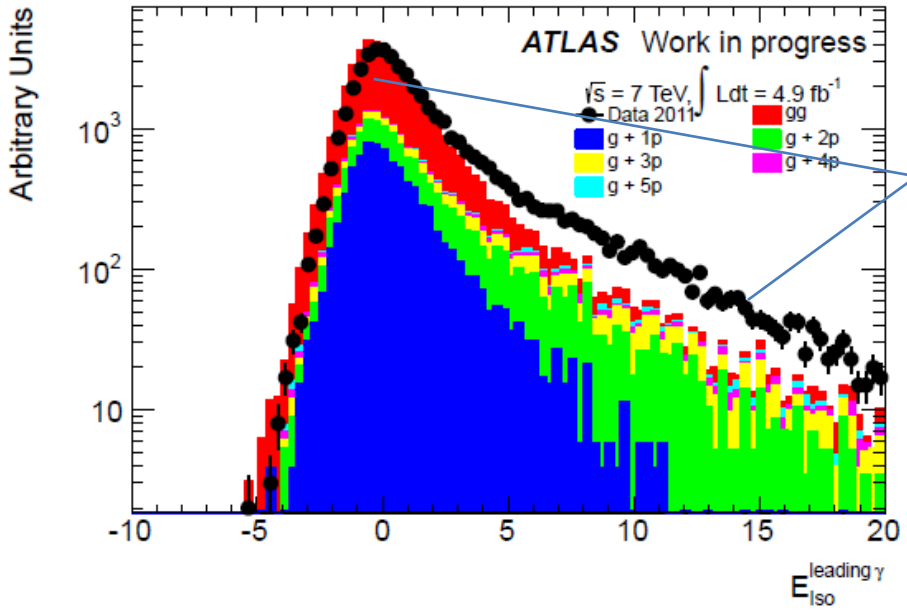


Sherpa can describe BDT output !!

(decent level, still jet-jet component is missing, and see next slide)

Isolation distribution

- Olivier and Sandro's MVA study suggests
  - Relax pT on leading photon to 25 GeV
  - Relax Eiso cut to 20 GeV
    Estimated 11% sensitivity gain with BDT respect to cut base analysis.



- jet-jet component is missing.
- Eiso shifting is not taken account. (+96 MeV, see Sandrine/Jean-Baptiste presentation in e/gamma)

Can not use Full simulated jet-jet MCs.
~ 10 MC events after photon ID cut.

- Plan
  - Will check impact of tightening Eiso cut for MVA (20 GeV→~ 10 GeV?)
  - Use data for jj component, get template from anti-tight photon ID
    - Need to consider leakage of real photon
    - 4x4 matrix method would be best choice.

Summary

- Check description of fully simulated MC sample
  - No need to use Pythia
  - Sherpa work very well
    - Additional reweighting is required
    - Propose to use $\Delta\eta$ of diphoton.
      - Reweighted MC sample describe data very well.
      - Description of BDT output improved by Sherpa in signal region.
      - Further check on BDT will be done with relaxed cut.
- Next step
  - Optimize Eiso range and photon pT range for MVA
  - Model jet-jet component from data
    - Once this done, we will study again if we need to reweight γ-jet component.