1^a Reunión Presencial del
Proyecto VMGRID
(ATLAS TIER2 ESPAÑA)
April 12-13, 2018

User Support and Analysis at UAM:

Measurement of inclusive isolated-photon cross sections at $\sqrt{s} = 13$ TeV using 36 fb^{-1}

J del Peso & C Glasman
Universidad Autónoma de Madrid

Contents:

Data analysis with Tier2: motivation

- Data analysis with Tier2 can be very useful for user support
  → many questions from users are related to ATLAS software for analysis or to errors on analysis programs

- Experts on data analysis procedures are candidates to join ATLAS DAST team
  → further contribution for a more efficient service to the ATLAS Collaboration

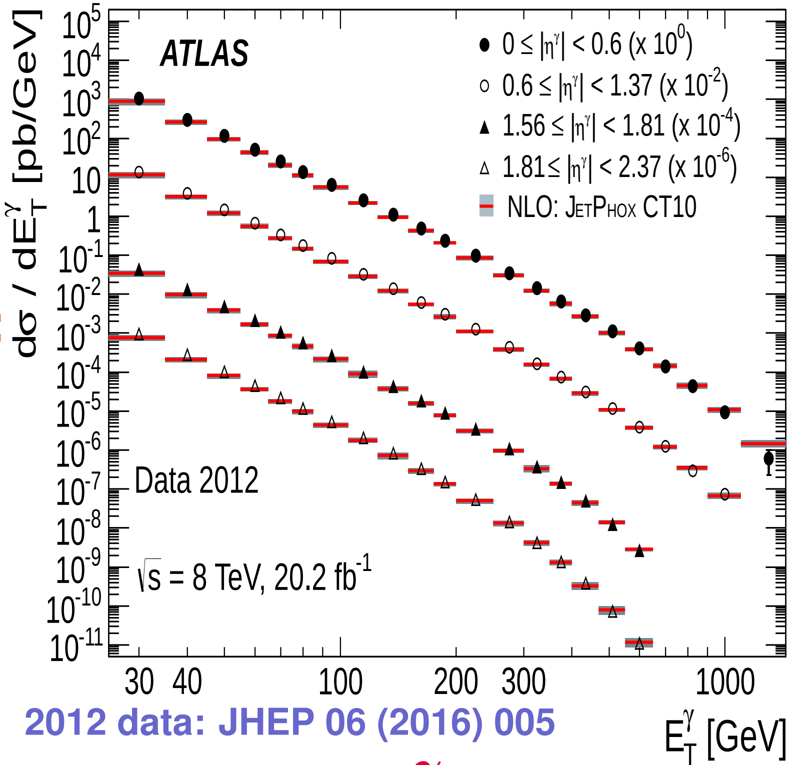
- The research team gains expertise in computing and physics
  → experience in the latest ATLAS tools for analysis can be passed to other users

- Further advantages of physics analysis:
  → this kind of activities yields to physics results for the experiment and to PhD thesis for students
  ⇒ enhancement of the research and educational capabilities of the project

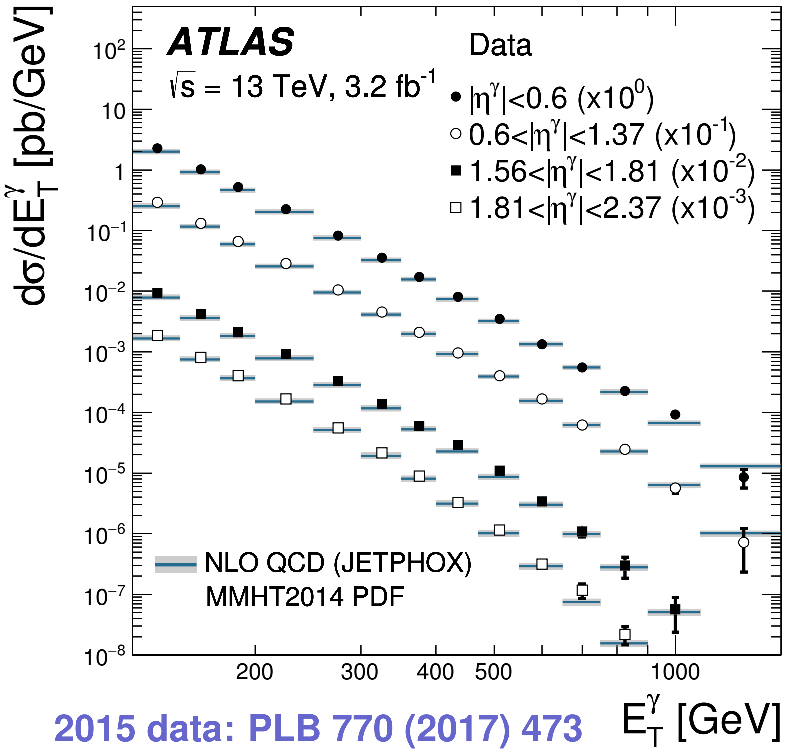
- Data analysis on two physics topics were proposed
  ⋆ prompt-photon production (UAM): status of this analysis presented here
  ⋆ top-quark production (IFIC)

Prompt-photon production: motivation

- Measurements of inclusive isolated-photon cross sections allow
  - → tests of pQCD
  - → constraints on the proton PDFs
  - → constraints on background to new particles decaying into photons
- Previous studies:



2012 data: JHEP 06 (2016) 005



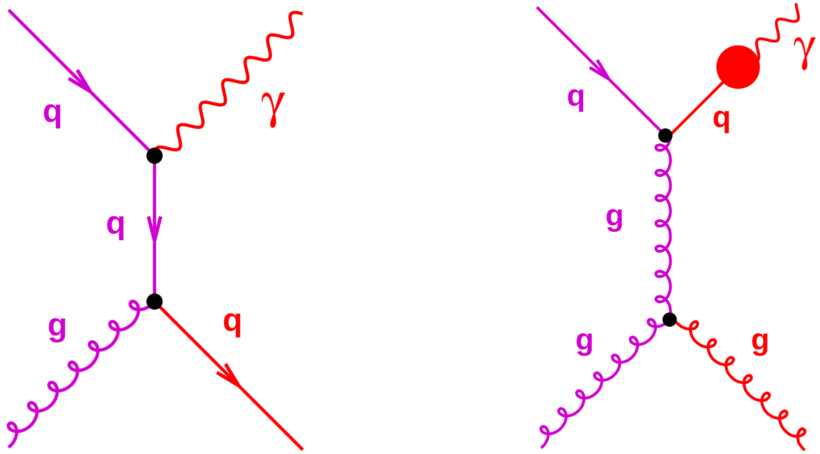
2015 data: PLB 770 (2017) 473

⇒ With 2015+2016 data, a higher E_{T}^{γ} reach is possible

→ the region where the measurements are dominated by the systematic uncertainties is extended significantly towards higher E_{T}^{γ}

Prompt photons in pp collisions at LHC

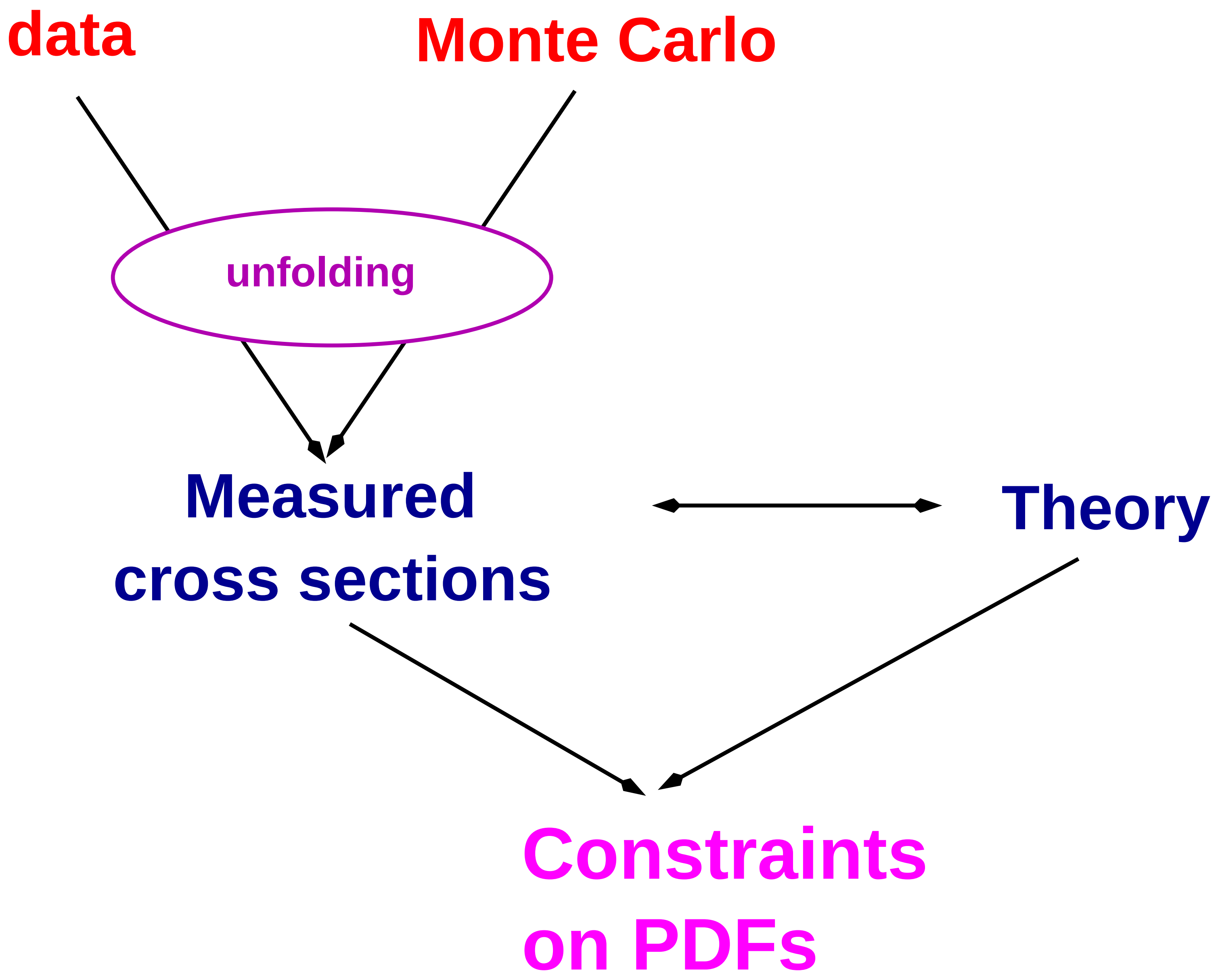
- Prompt photons in pp collisions ($pp \longrightarrow \gamma(+\text{jet}) + \text{X}$) are produced via two mechanisms:



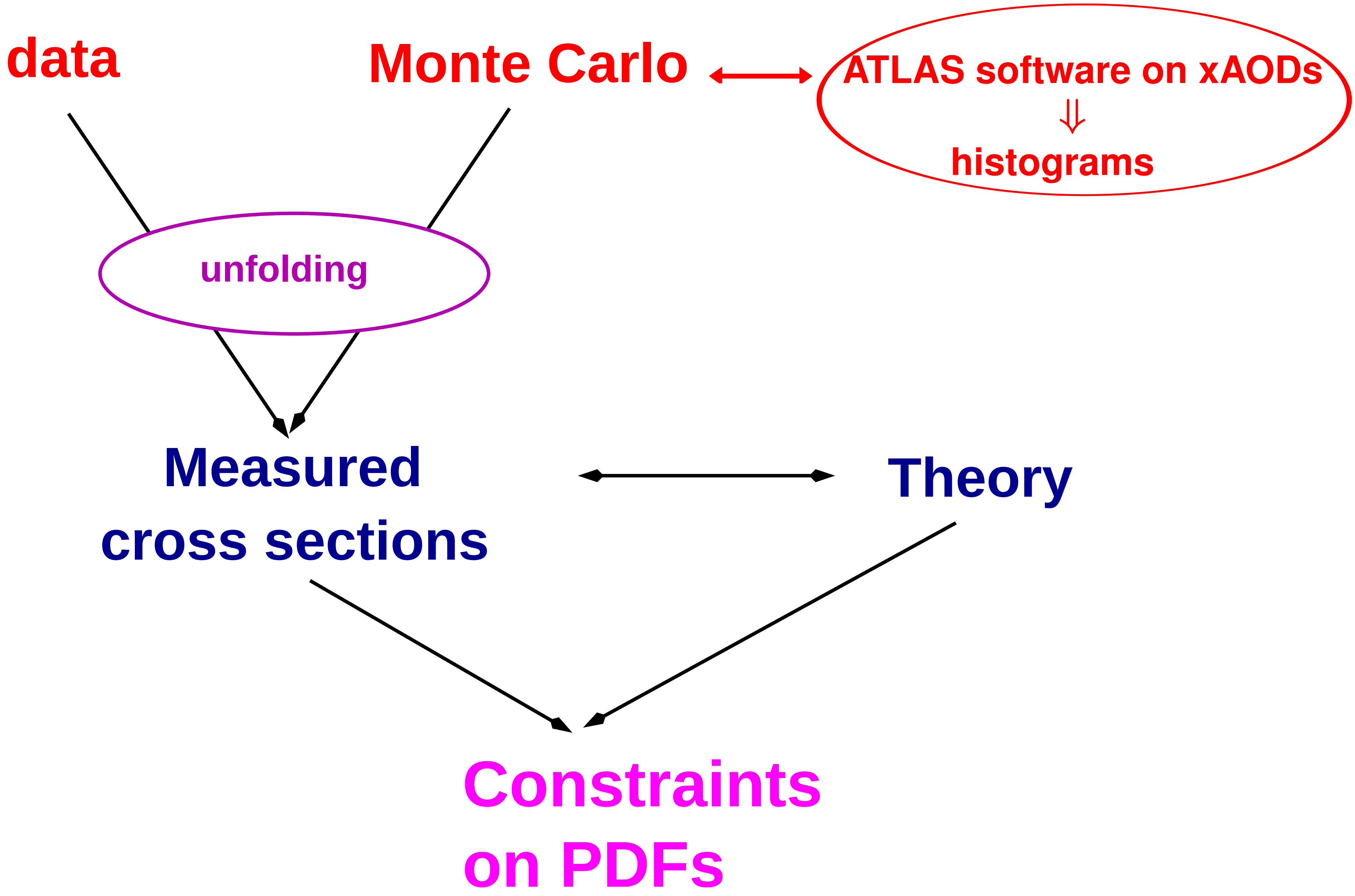
direct-photon (DP) fragmentation (F)

- In addition to prompt photons, photons are produced copiously inside jets (eg, π^0 decays)
  $\Rightarrow$ it is essential to require **isolation** to study prompt photons in hadron colliders
- This is achieved by requiring $E_{\rm T}^{\rm iso} \equiv \sum_i E_T^i < E_T^{\rm max}$, with the sum over the particles (except the photon!) inside a cone of radius R centered on the photon in the $\eta - \phi$ plane
- The isolation requirement suppresses mostly the contribution of photons inside jets and the fragmentation contribution

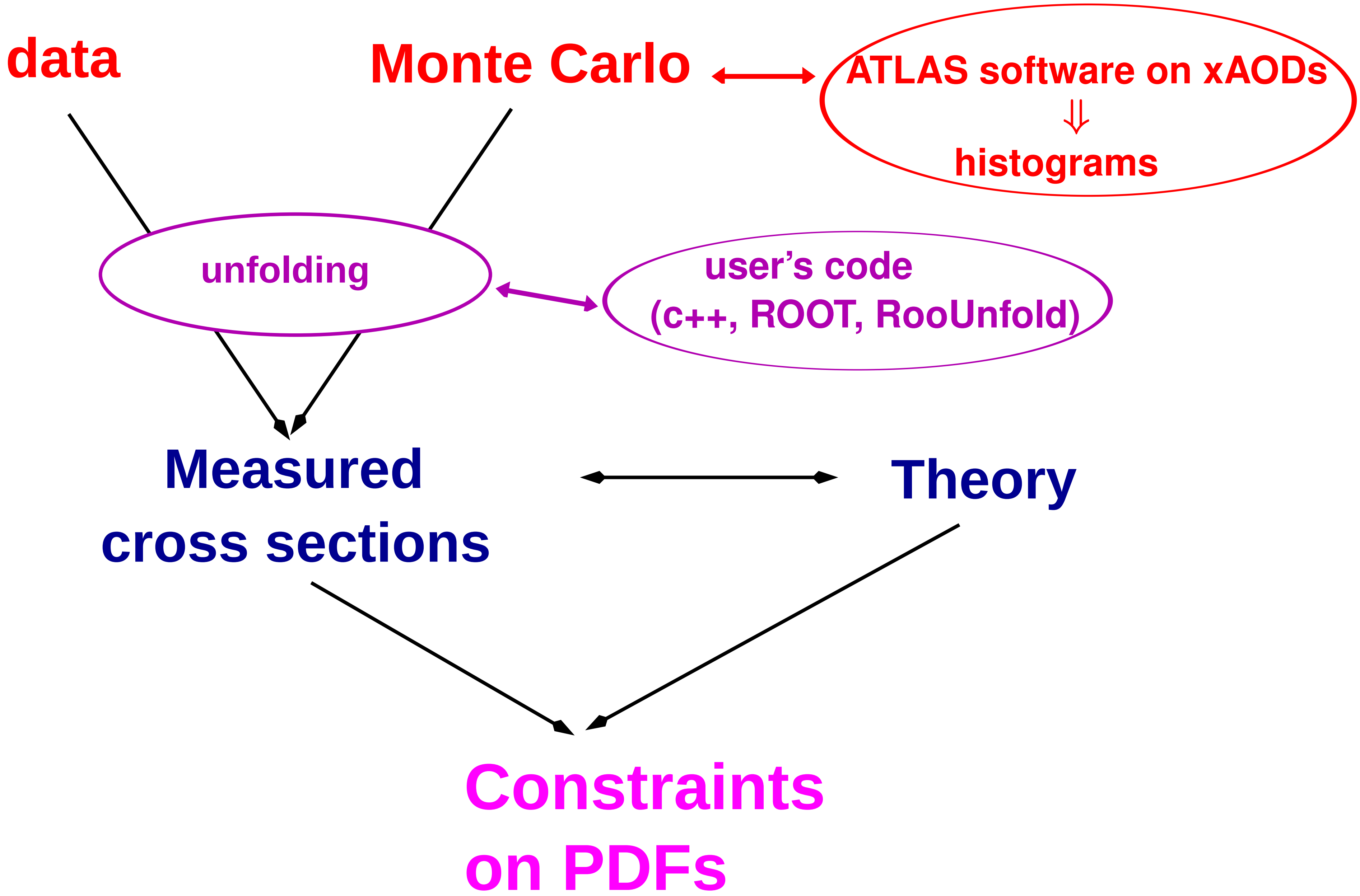
Tools for analysis



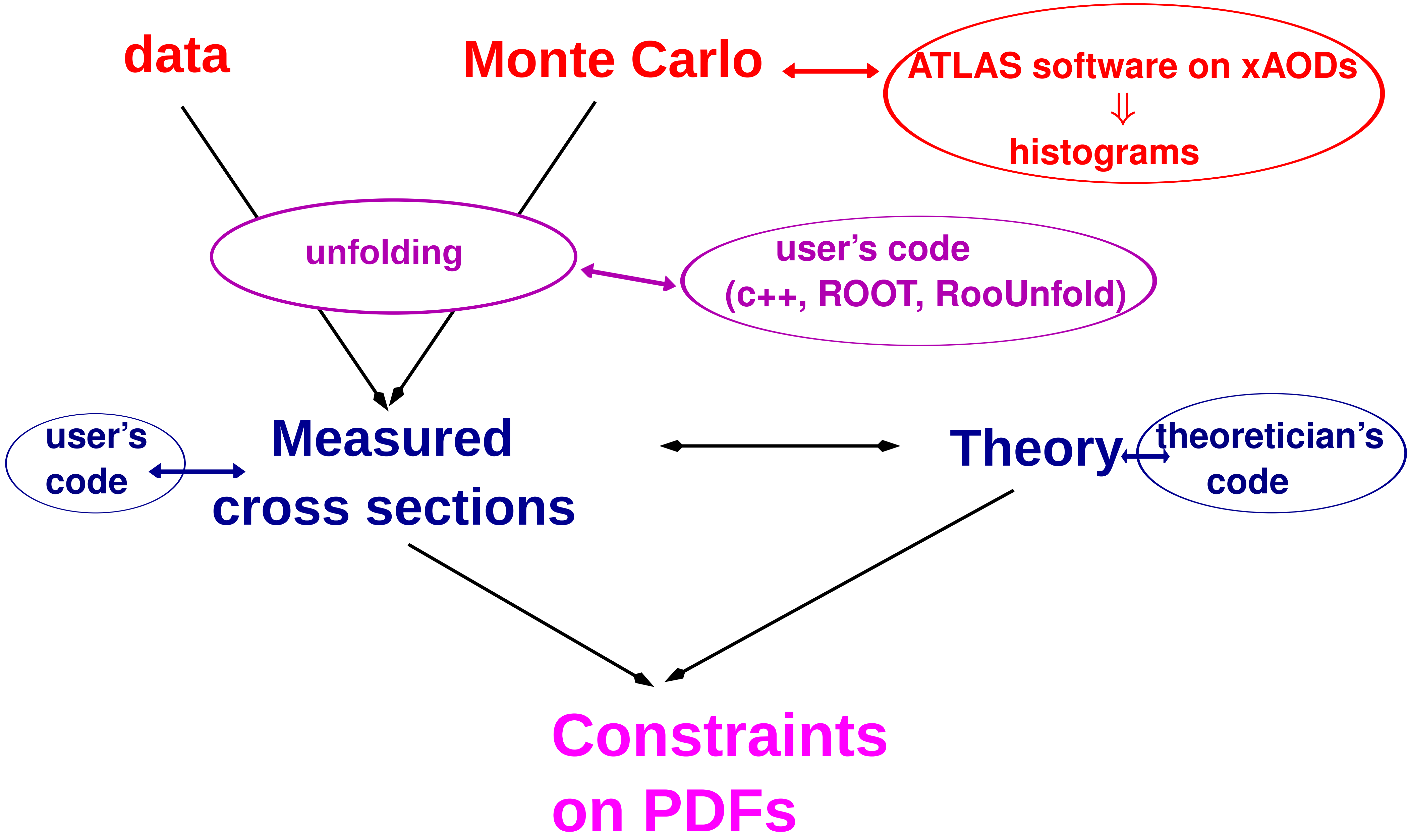
Tools for analysis



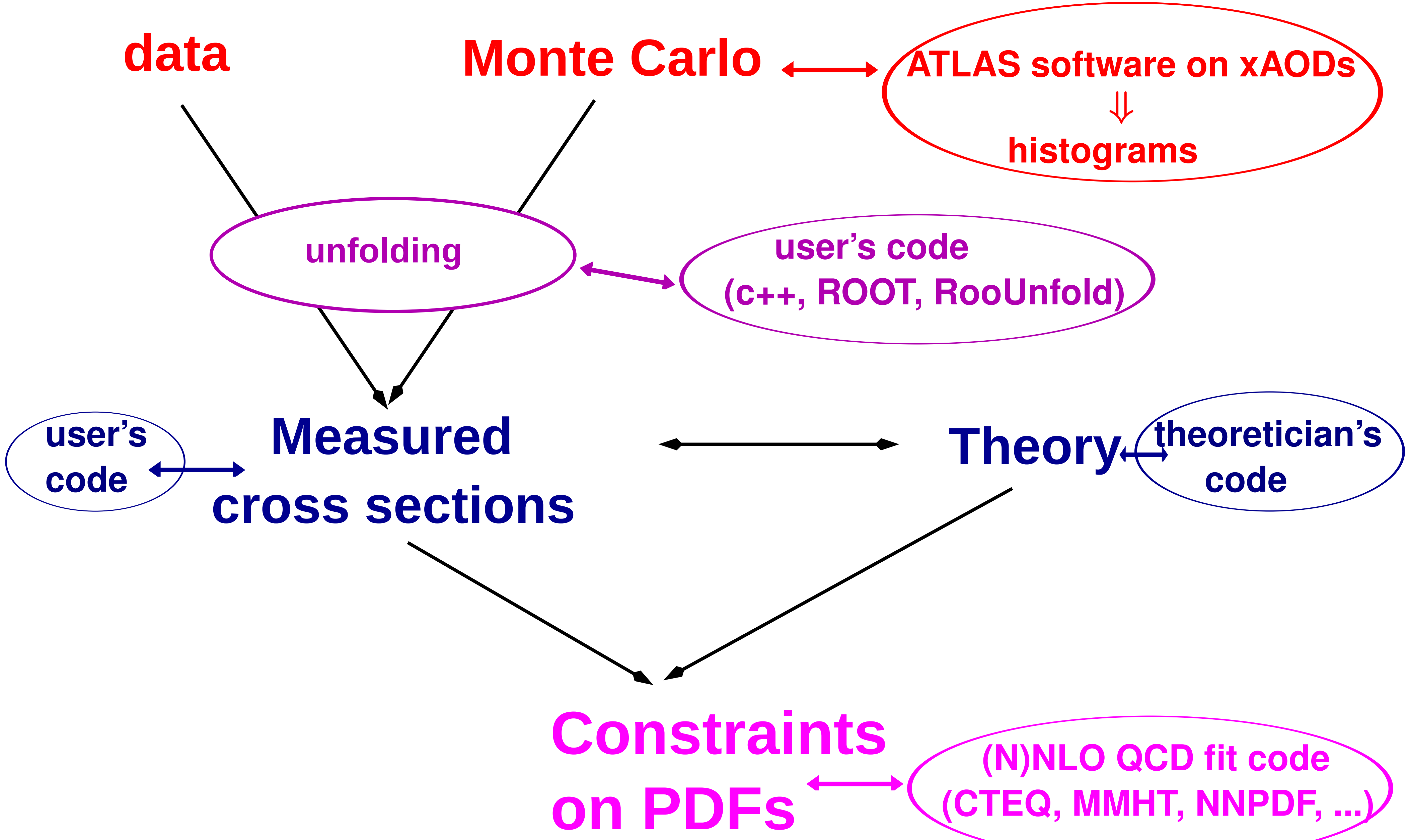
Tools for analysis



Tools for analysis



Tools for analysis



Details on data

- **Data samples:**
  → 2015; periods D-J ($\mathcal{L} = 3.21 \pm 0.07$ fb^{-1})
  (GRL: data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02_PHYS_StandardGRL_All_Good 25ns.xml)
  → 2016; periods A-K ($\mathcal{L} = 32.9 \pm 0.7$ fb^{-1})
  (GRL: data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_PHYS_StandardGRL_All_Good 25ns.xml)

- **Run on DAOD derivations (xAOD) in grid**
  → STDM2 derivations with tag p2950
  → final software release for 20.7 processing (release 2.4.42)
  → heavy use of rucio and pbook commands

- **Experience obtained:**
  – separate jobs in data periods: running jobs in parallel saves time and, when errors occur, allows a faster recovery
  – faster reprocessing time when new recommendations from CP groups are released
  – in addition, possibility to create mini-xAODs and store locally

Details on data II

- Main steps of analysis of data:
  - Data quality
  - Trigger selection
  - Photon calibration, identification and selection
  - Isolation computation and correction

- Some numbers:

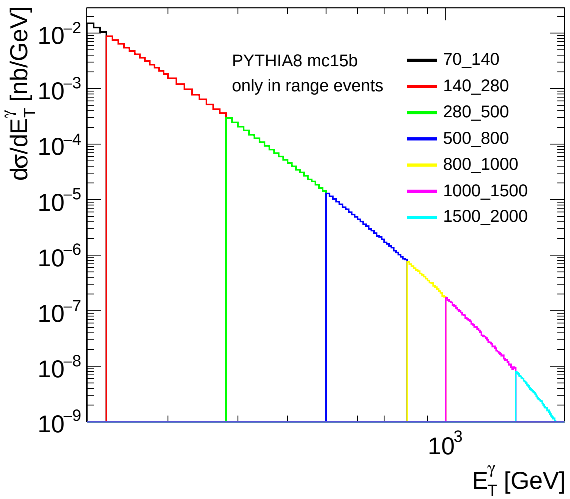
Level	Number of events	Storage	Running time
STDM2 derivation	> 300 M	22 TB (grid)	$>$ one week in grid
mini-xAODs (slimming/skimming)	> 71 M	2.1 TB (local)	1 day locally
analysis	> 26 M	histograms	minutes
used in measurement	> 7 M	histograms	seconds

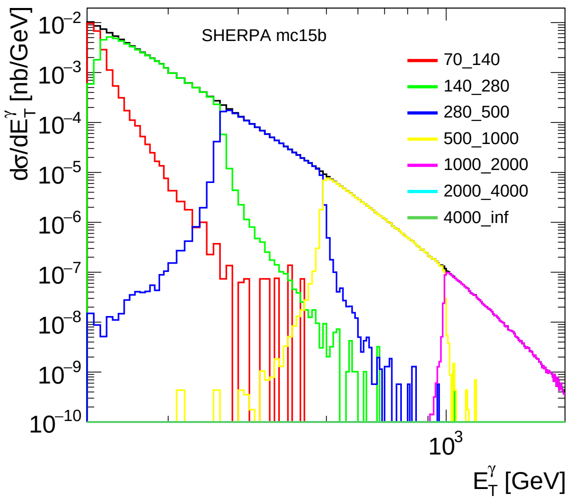
- Main problems of running data in grid:
  - $\rightarrow$ turn-around of more than a week
  - $\rightarrow$ a lot of processes fail $\rightarrow$ several retries per job are usually needed
  - $\rightarrow$ data files need frequently be rescued from deletion lists!!!

Details on Monte Carlo simulations

- Monte Carlo samples of PYTHIA and SHERPA are used to
  - → study the characteristics of signal events
  - → determine unfolding correction factors
  - → estimate hadronisation corrections to the NLO QCD calculations

- To increase efficiency of generation, samples were generated in slices of E_T^γ





- Run on DAOD derivations (xAOD) in grid
  - → STDM2 derivations with tag p2952

- Experience obtained:
  - — separate jobs in slices
  - — in addition, possibility to create ntuples and/or (partially) store locally (useful for processing of some systematic uncertainties, very time consuming)

Details on Monte Carlo simulations II

- Main steps of analysis of Monte Carlo simulations (reco and truth levels):
  - Photon calibration, identification and selection similar to data
  - Isolation computation and correction similar to data
  - Additional corrections due to mismatch between data and Monte Carlo
  - Definition of phase space for measurement done at truth level

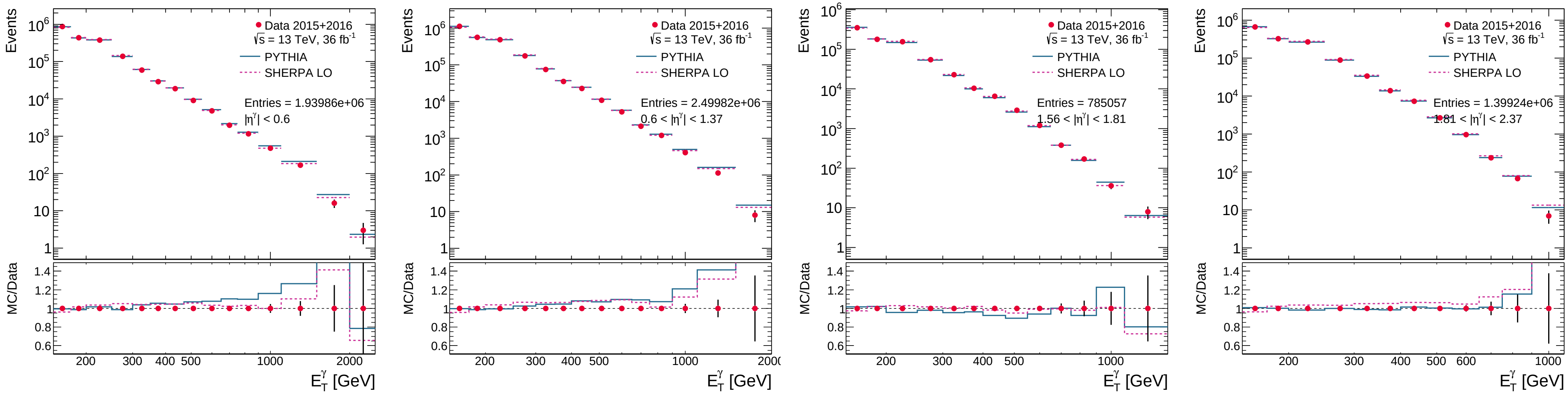
- Number of events for PYTHIA (SHERPA): $3.5\ (15)$ M

Level	Storage	Running time
STDM2 derivation	0.8 (3) TB (grid/partially local)	~ 1 day in grid/local
ntuples (only variables needed)	454 (1850) M (local)	< 0.5 hour
analysis	histograms	minutes
used in measurement	histograms	seconds

- Main problems of running Monte Carlo in grid:
  - $\rightarrow$ frequent corruption and disappearance of files
  - $\rightarrow$ jobs fail in grid but can be downloaded and run ok locally
  - $\rightarrow$ MC files need frequently be rescued from deletion lists!!!

Control plots: signal yield vs simulation

- The extracted signal yields, after background subtraction, are compared to the Monte Carlo simulations of the signal: distributions as functions of E_T^γ in different η^γ regions



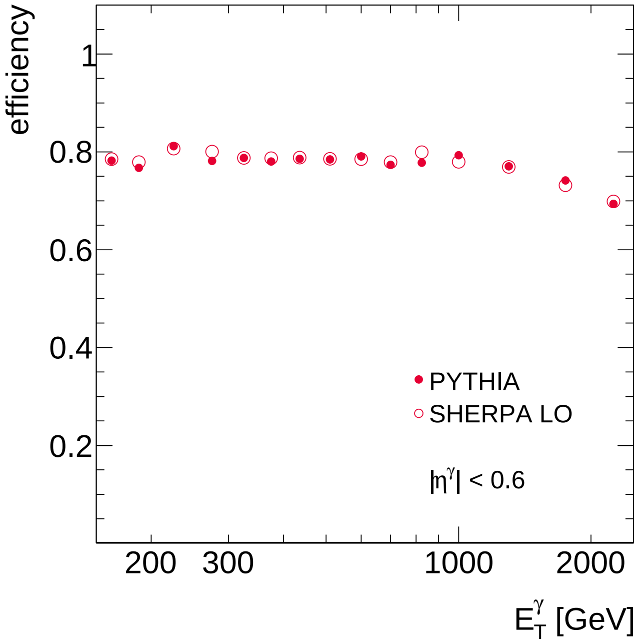
→ The data are reasonably well described by the simulation of the signal, except at high E_T^γ

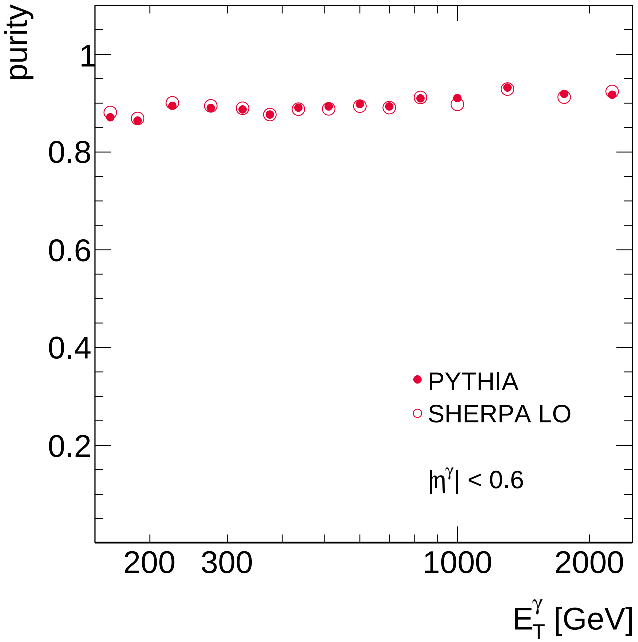
Differential cross sections

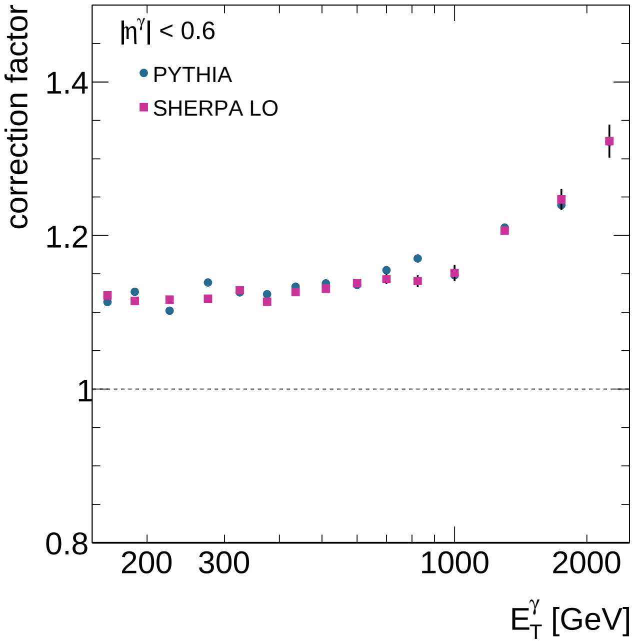
- **Cross sections measured using the bin-by-bin and Bayes' methods**
- **For bin-by-bin method, in each η^γ region:**

$$\frac{d\sigma}{dE_T^\gamma} = \frac{N^{sy} * C^{MC}}{\Delta E_T^\gamma * \mathcal{L}_{int}} \quad \text{with } C^{MC} = \frac{N^{MC}_{truth}}{N^{MC}_{reco}}$$

- **PYTHIA MC was used as nominal and SHERPA MC for systematics**
- **Reconstruction efficiency and purity and correction factors:**

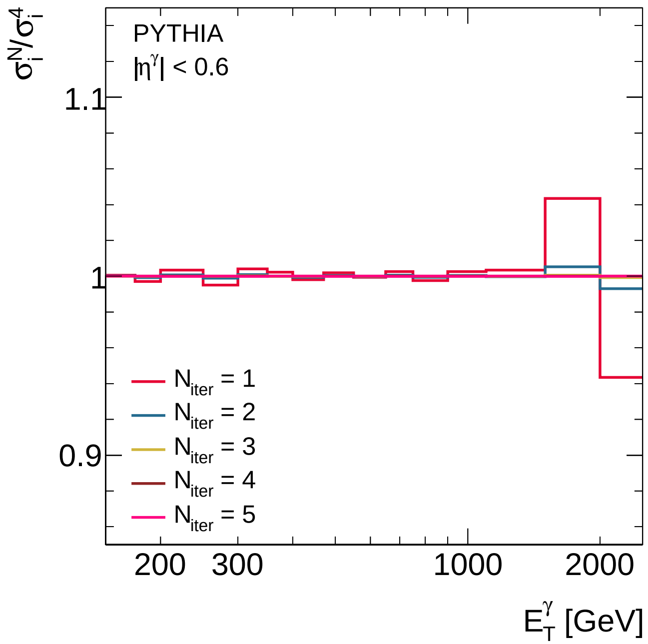


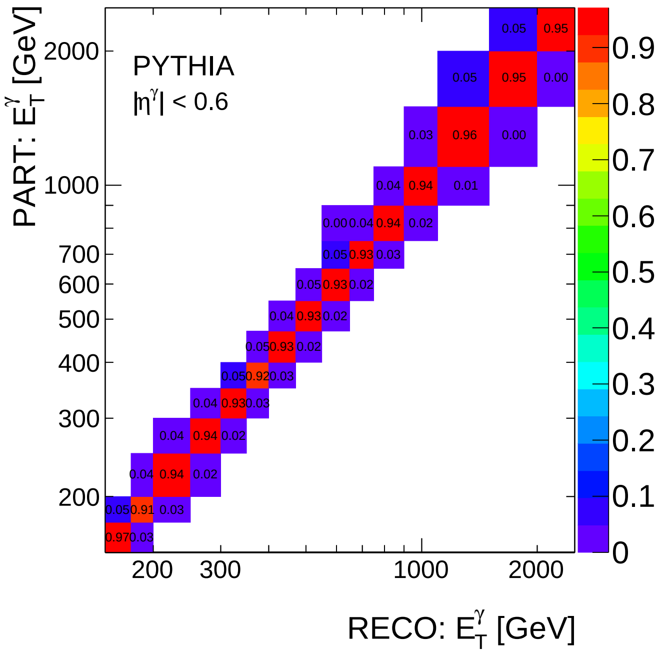


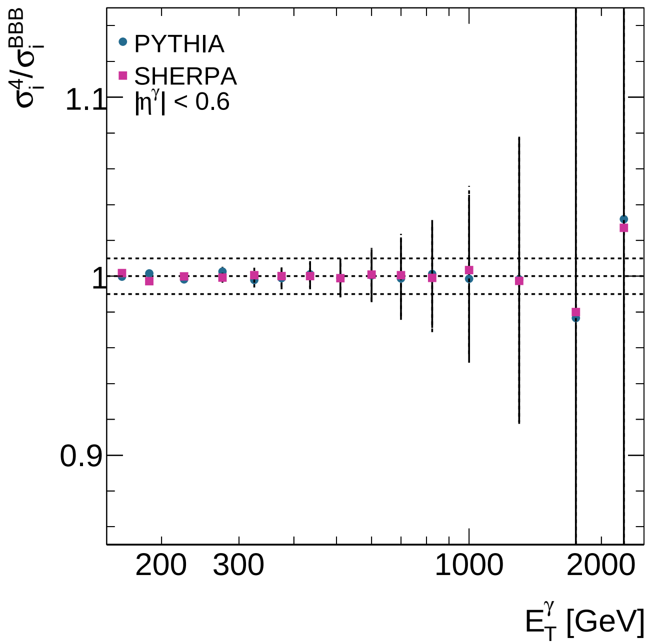


Differential cross sections

- **Cross sections measured using the bin-by-bin and Bayes' methods**
- **For Bayes' method use RooUnfold in each η^γ region with different number of iterations**
- **Number of iterations, transfer matrix and Bayes' vs BbB:**





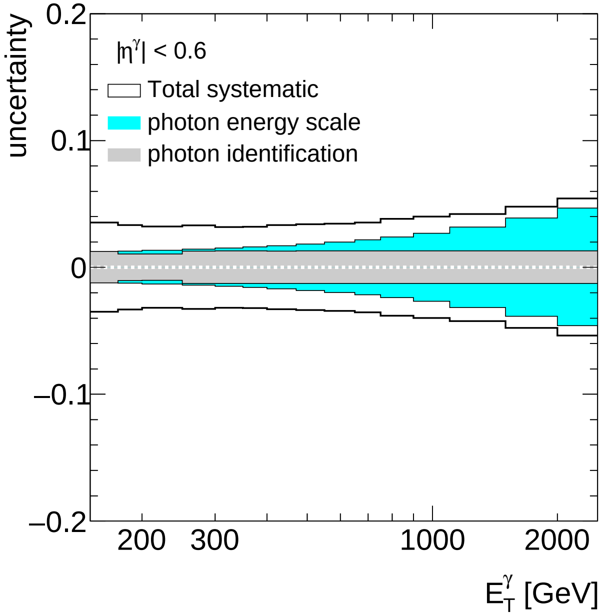


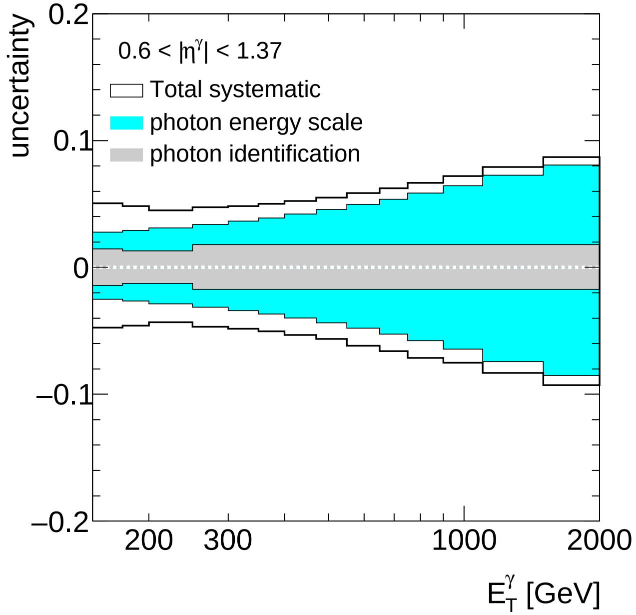
→ **Bayes' vs BbB: typically much less than 1% difference thanks to a very diagonal transfer matrix**

Systematic uncertainties

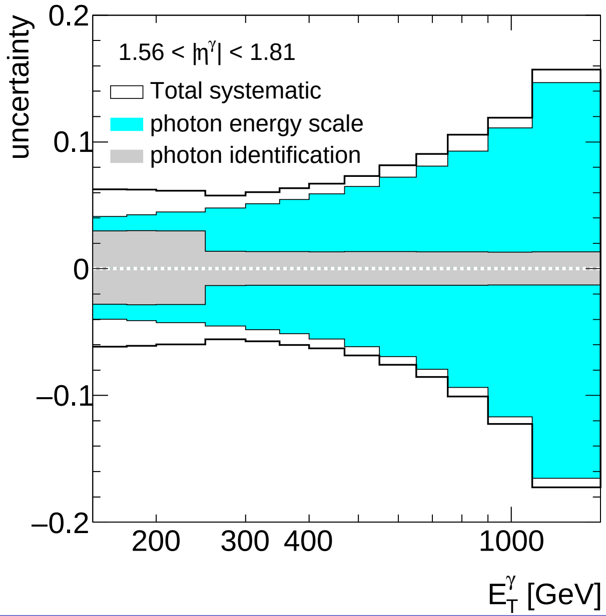
- Sources of systematic uncertainty considered:
  - → involves rerunning on data and/or Monte Carlo
  - → Signal modelling/unfolding:
    - — signal leakage fractions (Monte Carlo)
    - — correction factor and its statistical uncertainty (Monte Carlo)
  - → Background subtraction:
    - — choice of background control regions (data and Monte Carlo)
    - — identification and isolation correlation in background (data and Monte Carlo)
  - → Running conditions:
    - — pile-up (data and Monte Carlo)
    - — trigger efficiency (1%) and luminosity-measurement uncertainty (2.1%)
  - → Reconstruction:
    - — photon identification efficiency ($2^{\rm nd}$ dominant!) (Monte Carlo)
    - — isolation modelling (Monte Carlo)
  - → Photon calibration:
    - — photon energy resolution (Monte Carlo)
    - — photon energy scale (dominant!) (Monte Carlo)

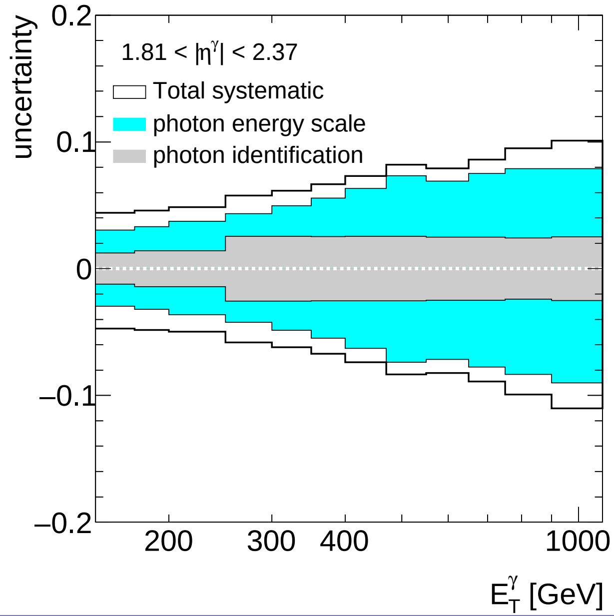
Total systematic uncertainty and main sources



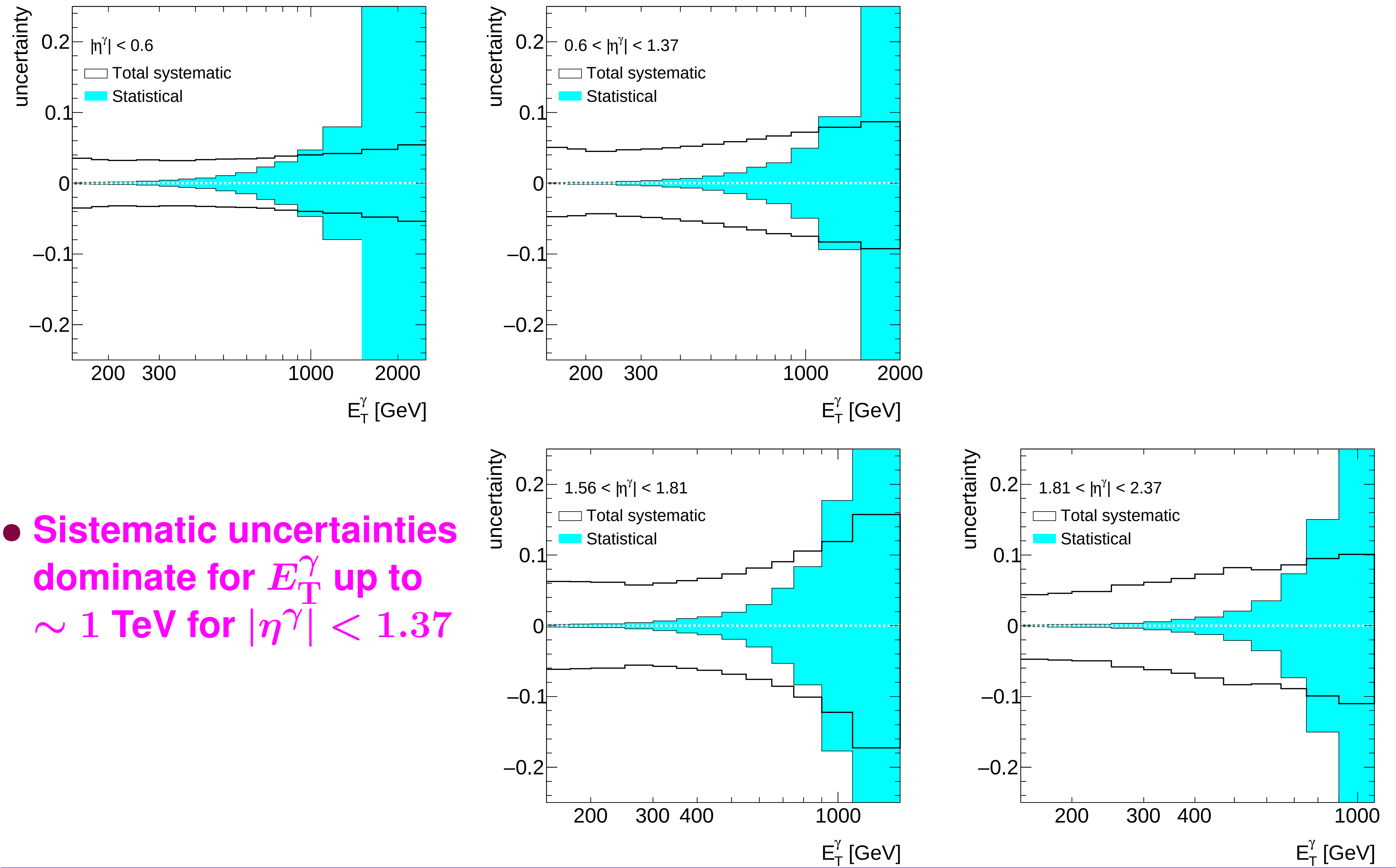


- **Dominant uncertainties:**
  - **First: photon energy scale**
  - **Second: photon identification**





Total systematic and statistical uncertainties



- **Sistematic uncertainties dominate for $E_{\rm T}^{\gamma}$ up to ~ 1 TeV for $|\eta^{\gamma}| < 1.37$**

Next-to-leading-order QCD calculations

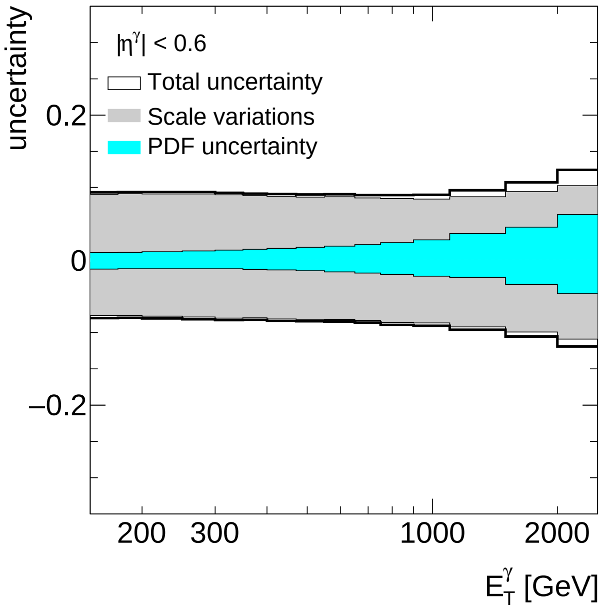
- **JETPHOX:** full fixed-order NLO pQCD predictions for direct and fragmentation contributions
  → scale: $\mu_R = \mu_F = \mu_F = E_{\rm T}^{\gamma}$
  → PDF: **MMHT2014 (also CT14, ABMP16, HERAPDF2.0, NNPDF3.0)**
  → $\alpha_{\rm s} = 0.120$
  → non-perturbative corrections for hadronisation and underlying event needed (computed using PYTHIA samples, $< 1\%$)

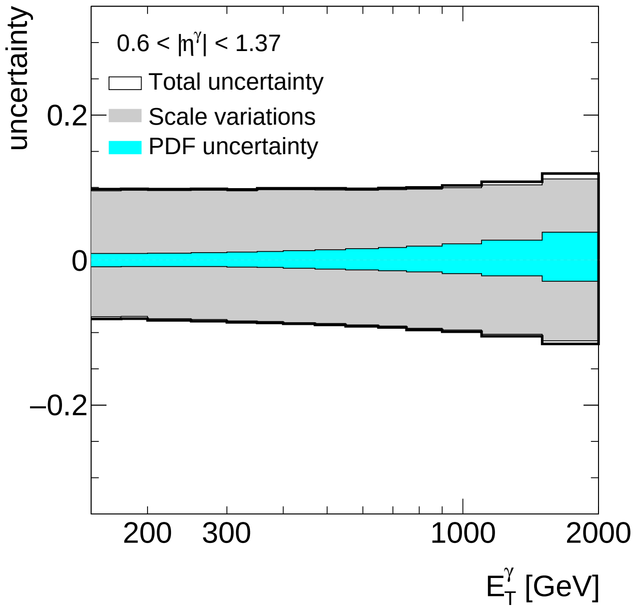
- **NLO SHERPA:** parton-level calculations for $\gamma + 1, 2$ jets at NLO and $\gamma + 3, 4$ jets at LO supplemented with parton shower; only direct contribution
  → scale: **dynamic scale setting**
  → PDF: **NNPDF3.0)**
  → $\alpha_{\rm s} = 0.118$
  → predictions at particle level

- **Theoretical uncertainties:** higher orders (dominant), PDF-induced uncertainty, uncertainty on $\alpha_{\rm s}$ and hadronisation+UE corrections (JETPHOX)

JETPHOX (S. Catani et al, JHEP 0205 (2002) 028) NLO SHERPA (T. Gleisberg et al, JHEP 0902 (2009) 007)

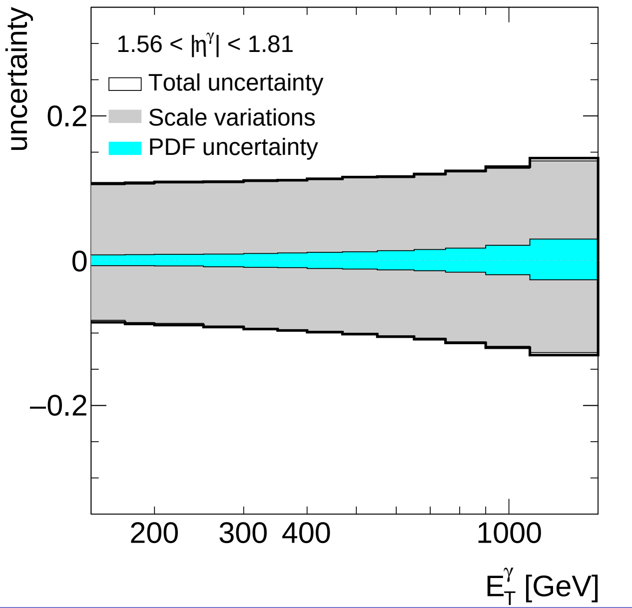
Total theoretical uncertainty and main sources

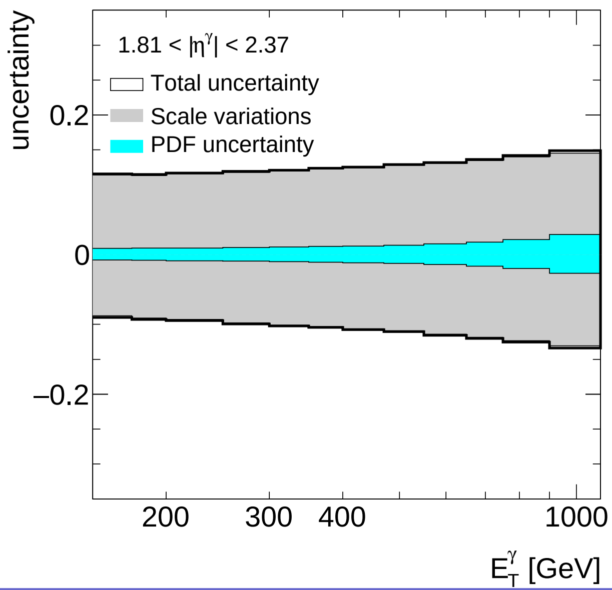




JETPHOX

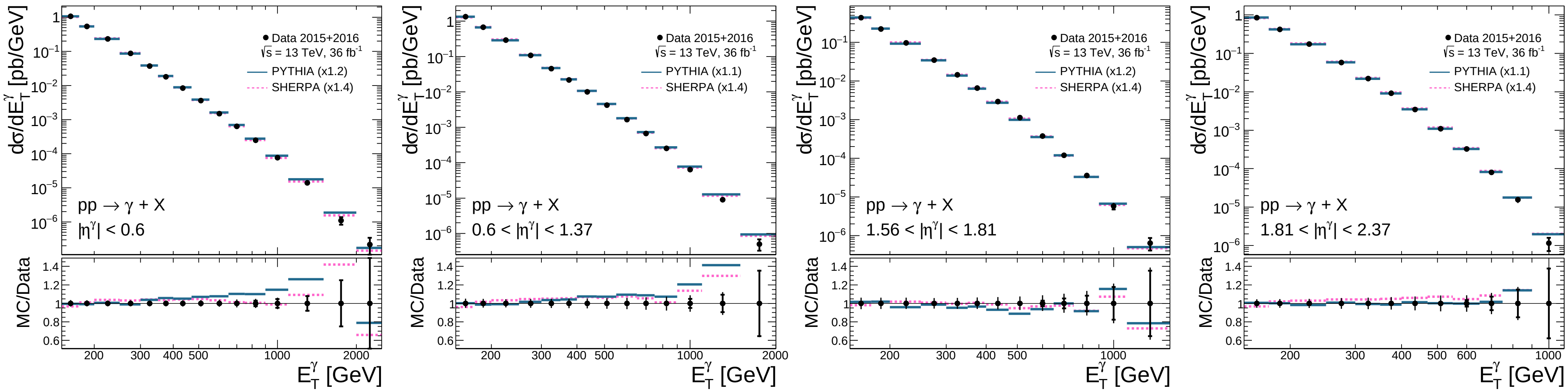
- **Dominant uncertainty:**
  - **First: higher orders**
  - **Second: PDFs**





Results: Measured cross sections vs MC predictions

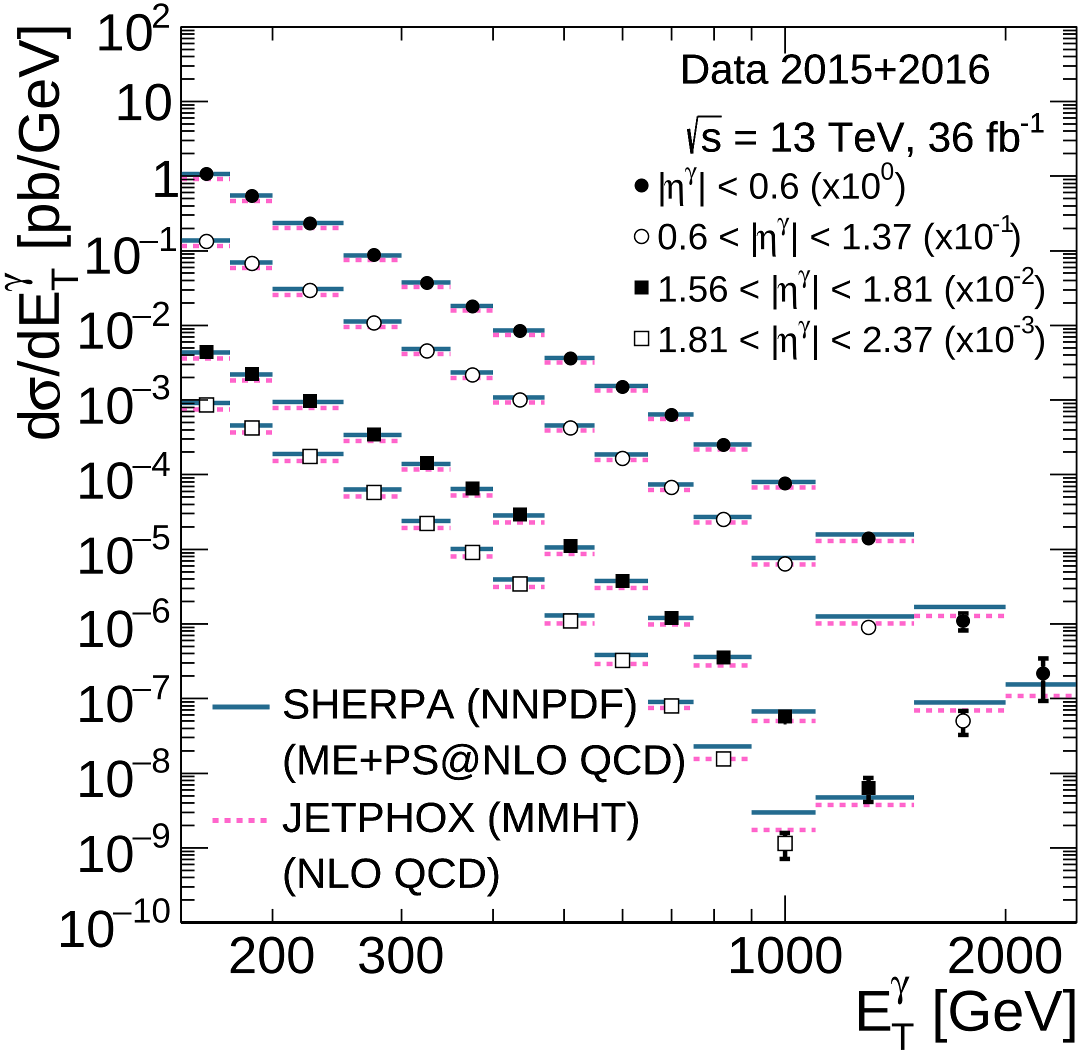
- The measured cross sections are compared to the predictions of PYTHIA and SHERPA MC calculations
  → the MC predictions are normalised to the measured integrated cross section in each η^γ region (shape comparison)
- Phase-space region: $E_T^\gamma > 150$ GeV, $|\eta^\gamma| < 2.37$ (crack region removed) and $E_T^{\rm iso} < 4.2 \cdot 10^{-3} \cdot E_T^\gamma + 4.8$ GeV



→ The shape of measured cross sections is reasonably well described by the MC predictions of PYTHIA and SHERPA, except at high E_T^γ

Results: Measured cross sections vs NLO predictions

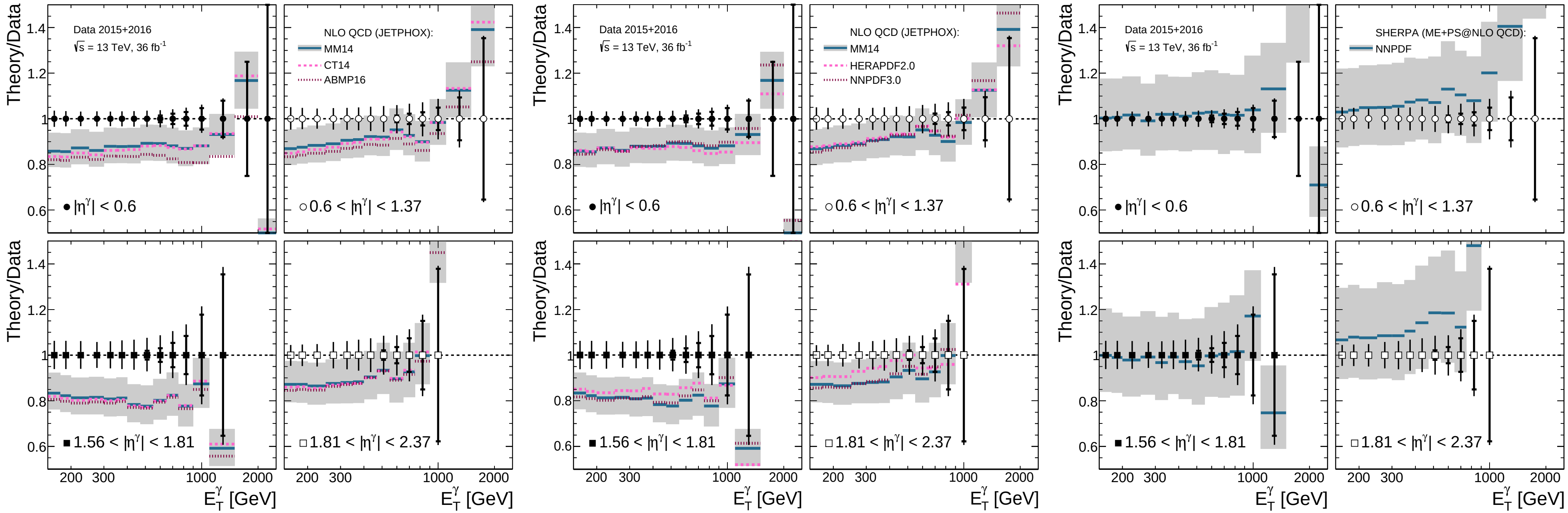
- The measured cross sections are compared to the predictions of the NLO QCD calculations



→ Measured cross sections adequately described by the NLO QCD predictions

Results: Measured cross sections vs NLO predictions

- Ratio of the NLO QCD predictions based on different PDFs to the measured cross sections



$\longrightarrow$ The NLO QCD calculations are consistent with the data within the experimental and theoretical uncertainties

$\longrightarrow$ Differences between the predictions based on different PDFs are observed in some regions of phase space $\longrightarrow$ potential to constrain the PDFs in QCD fits

Summary and conclusions

- A measurement of the cross section for inclusive isolated-photon production in pp collisions at $\sqrt{s} = 13$ TeV using an integrated luminosity of 36 fb^{-1} of 2015+2016 data was presented

- Cross sections were measured as functions of $E_{\rm T}^{\gamma}$ in different regions of η^{γ} for $E_{\rm T}^{\gamma} > 150$ GeV, $|\eta^{\gamma}| < 2.37$ and $E_{\rm T}^{\rm iso} < 4.2 \cdot 10^{-3} \cdot E_{\rm T}^{\gamma} + 4.8$ GeV

- The $E_{\rm T}^{\gamma}$ reach was extended up to 2.5 TeV (1.5 TeV in previous measurements)

- The region where the measurements are dominated by systematic uncertainties is extended up to 1 TeV (0.6 TeV in previous measurements)

- Comparison with QCD predictions:
  - the MC predictions give an adequate description of the data, except at high $E_{\rm T}^{\gamma}$
  - the NLO predictions are in agreement with the data within the experimental and theoretical uncertainties
  - differences observed between predictions based on different PDFs at high $E_{\rm T}^{\gamma}$
  - these high precision measurements have the potential to constrain further the proton PDFs in future QCD fits, especially at high $E_{\rm T}^{\gamma} \leftrightarrow$ high x

Status of analysis

- Analysis of 2015+2016 data completed by analysers (UAM group)

- Supporting documentation ready: ATL-COM-PHYS-2018-226

- EdBoard requested on plenary SM meeting on March 22nd: EdBoard has been formed last week

- The expectation is that this analysis will be finalised by the Summer

- Next step: start analysis of full Run 2 data, including 2015-2018 data
  $\rightarrow$ the expectation is that analysis of whole data set will be finalised by the end of the project