Data Reuse for MC tuning and validation

Forum on the interpretation of the LHC results for BSM studies @ CERN

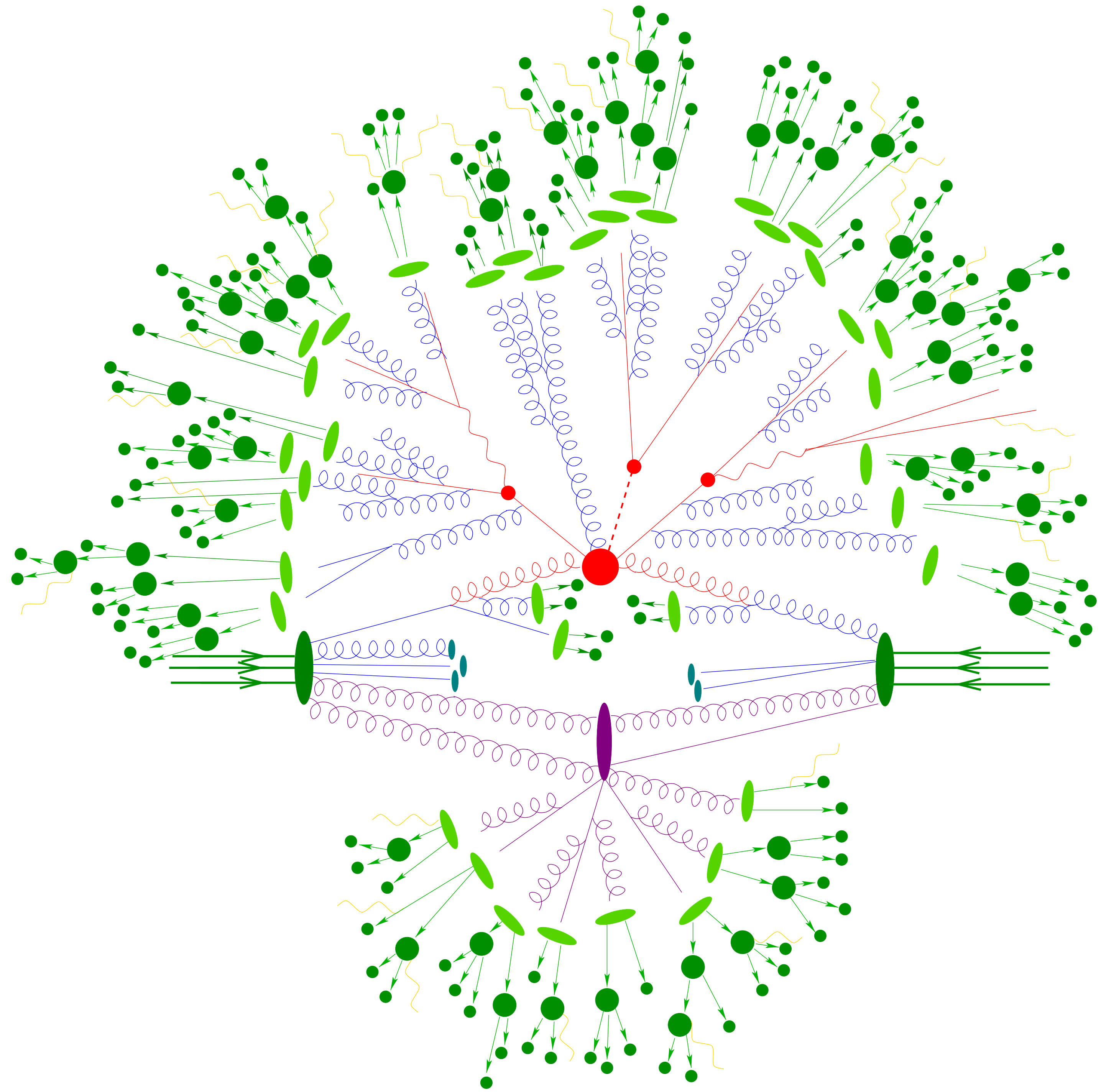
Peter Meinzinger, Uni Zürich, 26 Feb 2025
with input from Frank Krauss, Ilkka Helenius and Stefan Kiebacher

General-purpose event generators

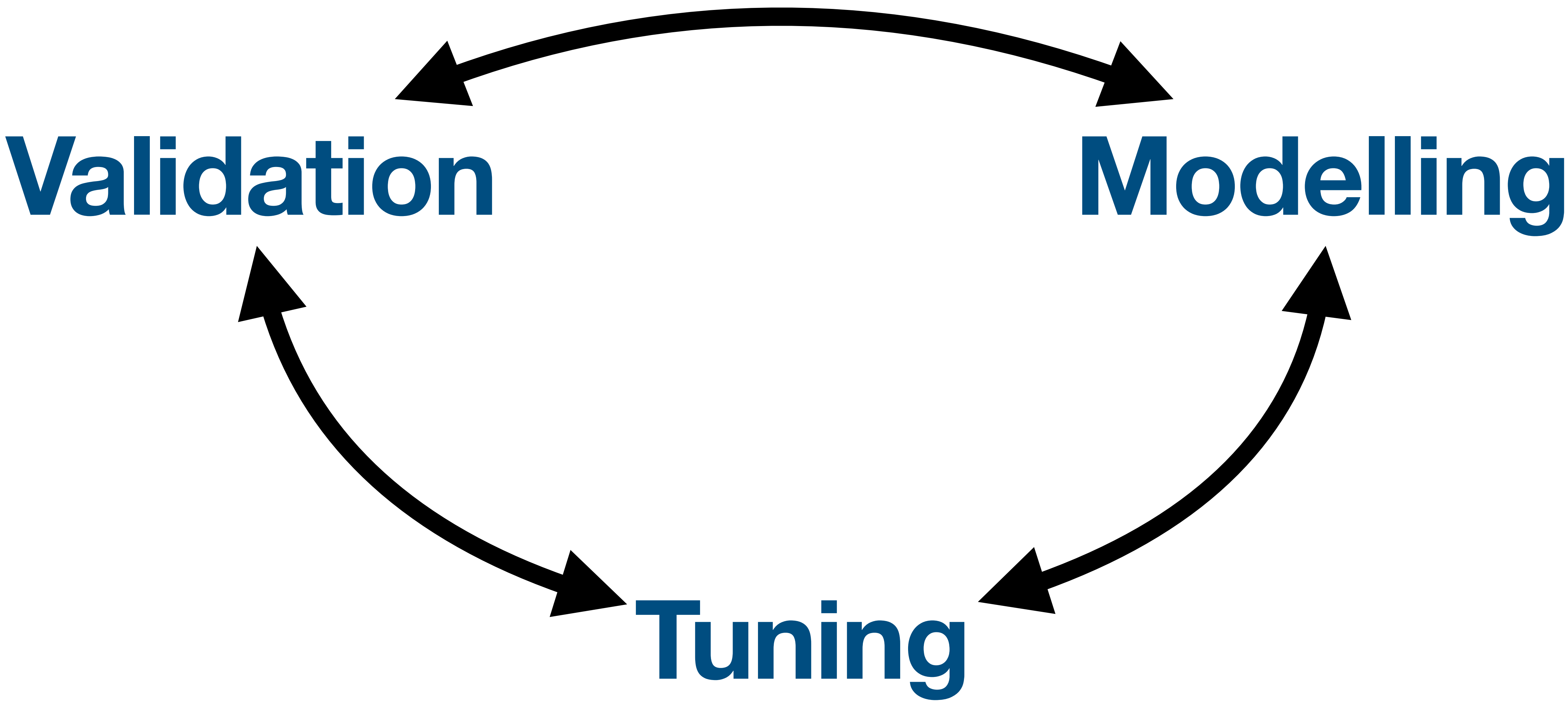
- they incorporate both **perturbative** and **non-perturbative/ beyond-factorisation** physics

- **“divide et impera”** strategy

- non-perturbative/ beyond-factorisation components need tuning to data



Three use-cases



Some guidance on photoproduction

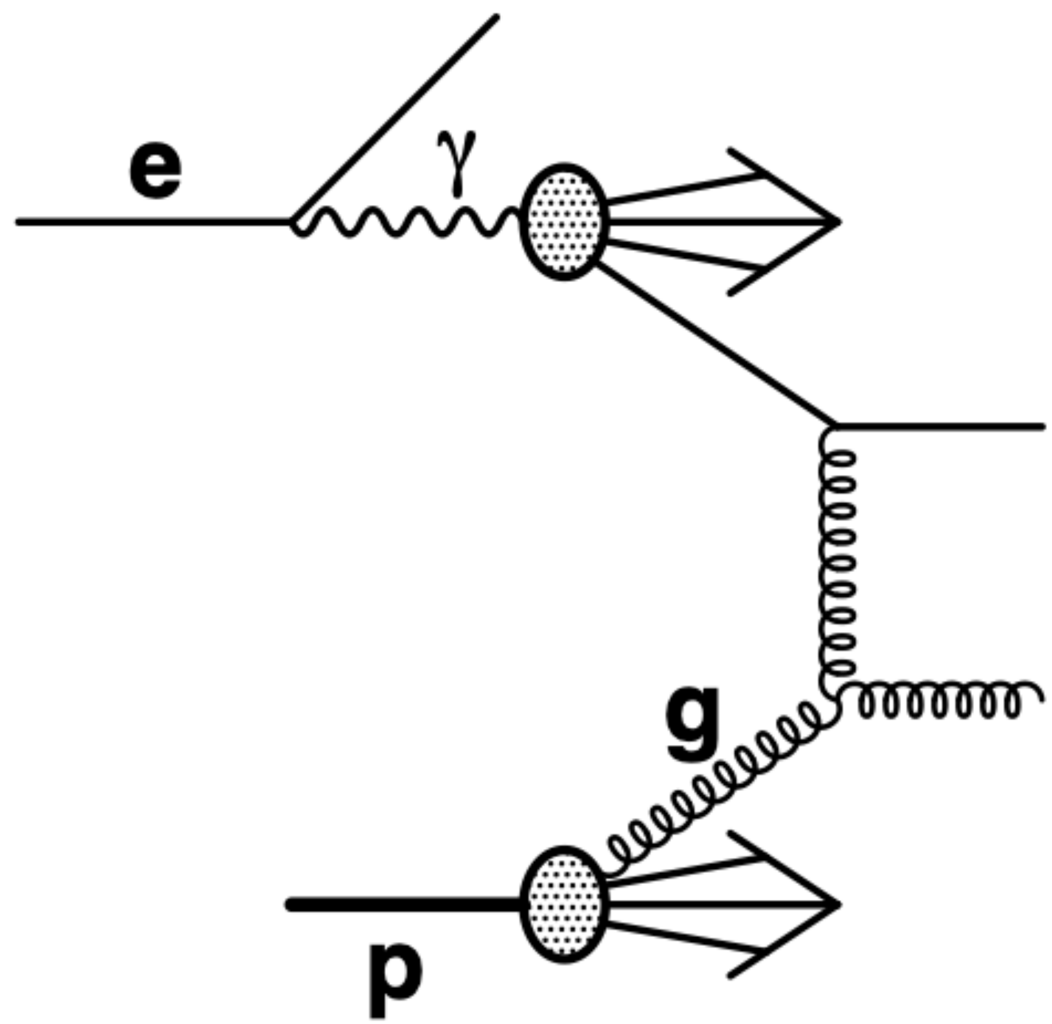
- Consider ep colliders like HERA

  ➡ high photon virtuality:
  Deep Inelastic Scattering

  ➡ low photon virtuality:
  photoproduction

- In photoproduction, the photon might behave like a hadron

- If the proton stays intact, the event is also classified as diffractive



taken from [1308.3368]

Validation

Validation

Example: diffractive photoproduction at HERA

Krauss, PM, [2407.02133]

- Diffraction = proton stays intact (or dissociates into low-mass system)

- For high Q^2 (diffractive DIS), factorisation into ME and Diffractive PDF has been proven in [hep-ph/9709499]

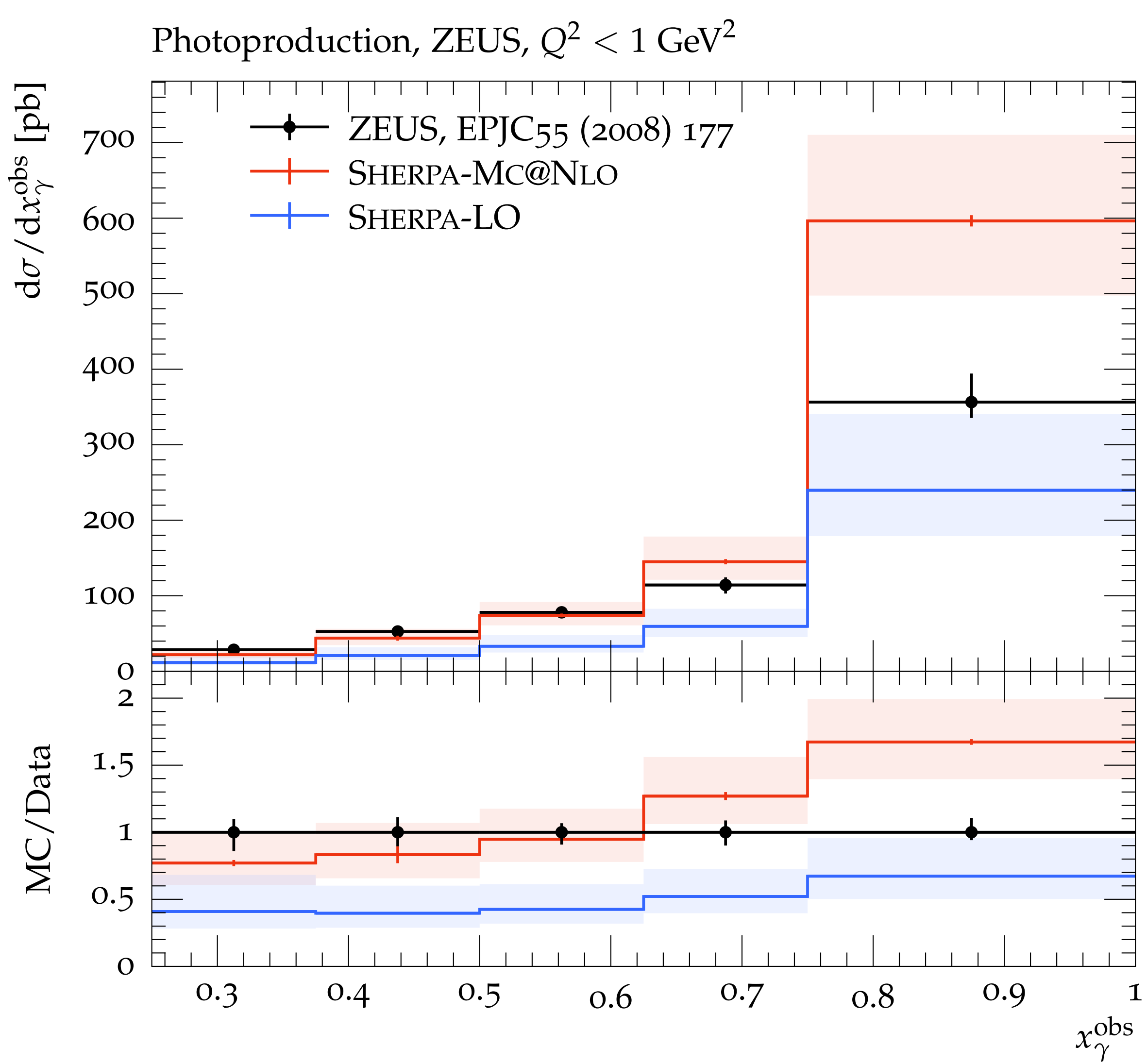
- For low Q^2 (diffractive photoproduction), no theorem for factorisation

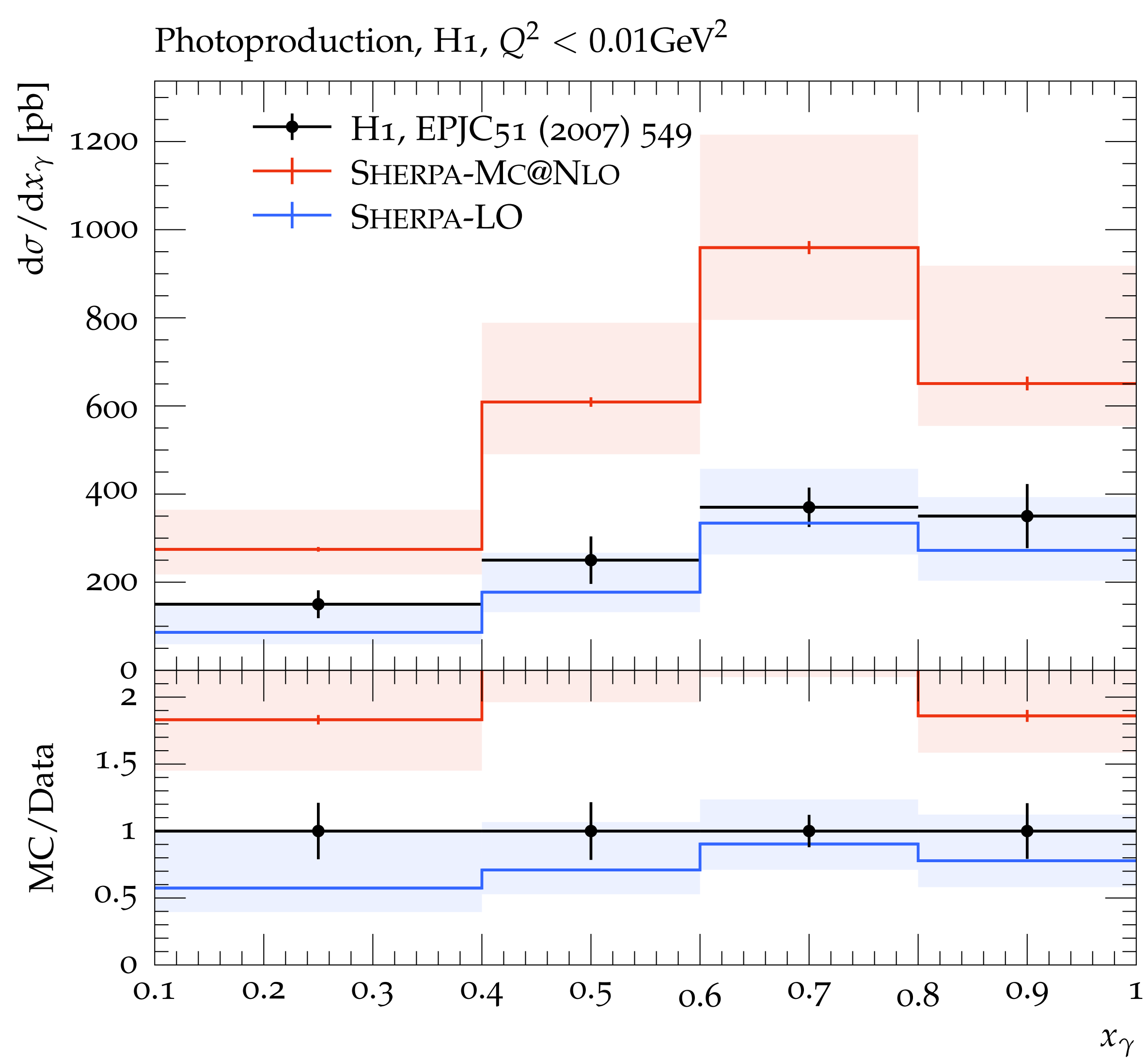
Validation

Example: diffractive photoproduction at HERA

Krauss, PM, [2407.02133]

- compare two measurements, by H1 (2007) and ZEUS (2008), with MC@NLO predictions using Diffractive PDFs



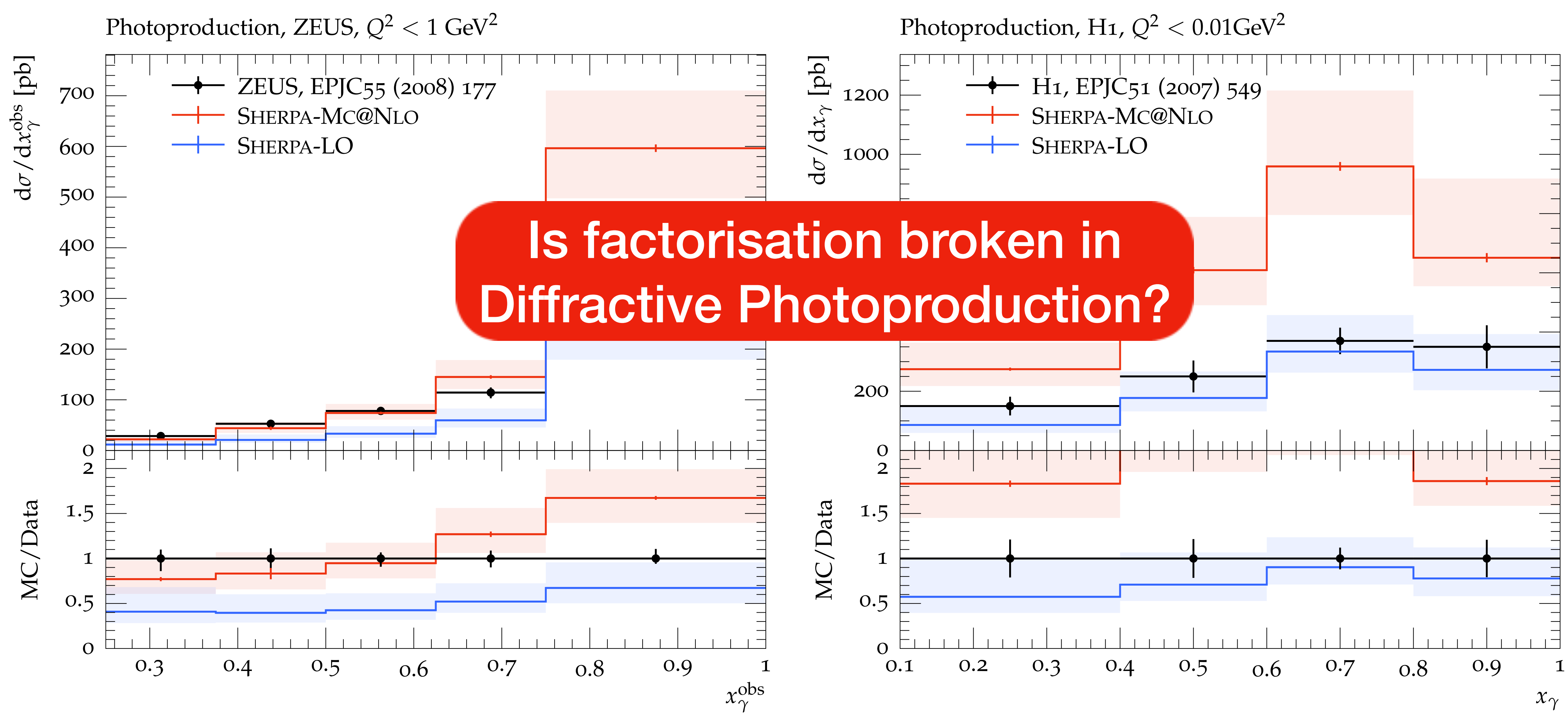


Validation

Example: diffractive photoproduction at HERA

Krauss, PM, [2407.02133]

- compare two measurements, by H1 (2007) and ZEUS (2008), with MC@NLO predictions using Diffractive PDFs



Validation

Example: diffractive photoproduction at HERA

Is factorisation broken in Diffractive Photoproduction?

Factorisation breaking in diffractive dijet photoproduction at HERA

A.B. Kaidalov[1], V.A. Khoze[2,3,a], A.D. Martin[2], M.G. Ryskin[2,3]

Abstract We discuss the factorisation breaking observed in diffractive dijet photoproduction by the H1 and ZEUS collaborations at HERA. By considering the effects of rapidity gap survival, hadronisation, migration and NLO contributions, we find that the observed data are compatible with theoretical expectations.

[0911.3716]

Factorisation breaking in diffractive dijet photoproduction at HERA?

Radek Žlebčík, Karel Černý, Alice Valkárová[a]

Abstract Recent experimental data on dijet cross sections in diffractive photoproduction at HERA collider are analysed with an emphasis on QCD factorisation breaking effects. The possible sources of the different conclusions of H1 and ZEUS collaborations are studied.

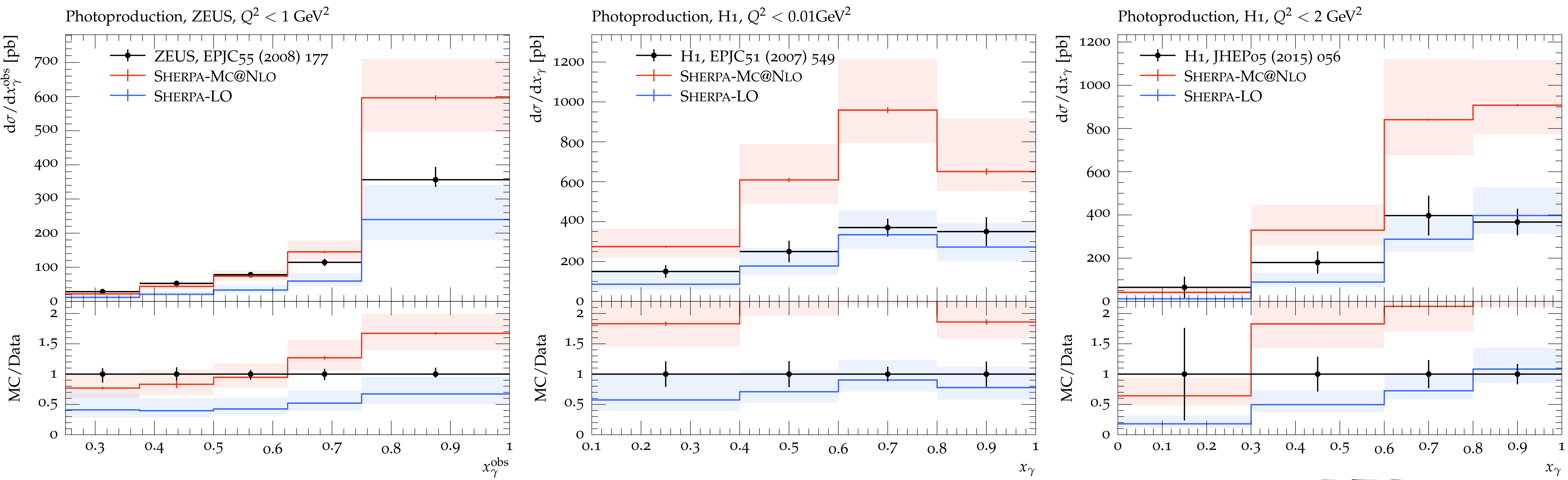
[1102.3806]

Validation

Example: diffractive photoproduction at HERA

Krauss, PM, [2407.02133]

Luckily, there were more measurements, e.g. by H1 in 2015:



Is factorisation broken in Diffractive Photoproduction?

Probably yes, as it can not be explained by NLO, parton shower, or hadronisation corrections

Model development

Model development

Example: multiple interactions of photons

Butterworth et al., [2408.15842]

- the Electron-Ion Collider will need calculations for low-Q^2 data

➡ photoproduction with a hadron-like behaviour includes Multiple-Parton Interactions (MPI)

➡ different MPI models possible

What is the $\sqrt{s}$-scaling behaviour of MPIs?

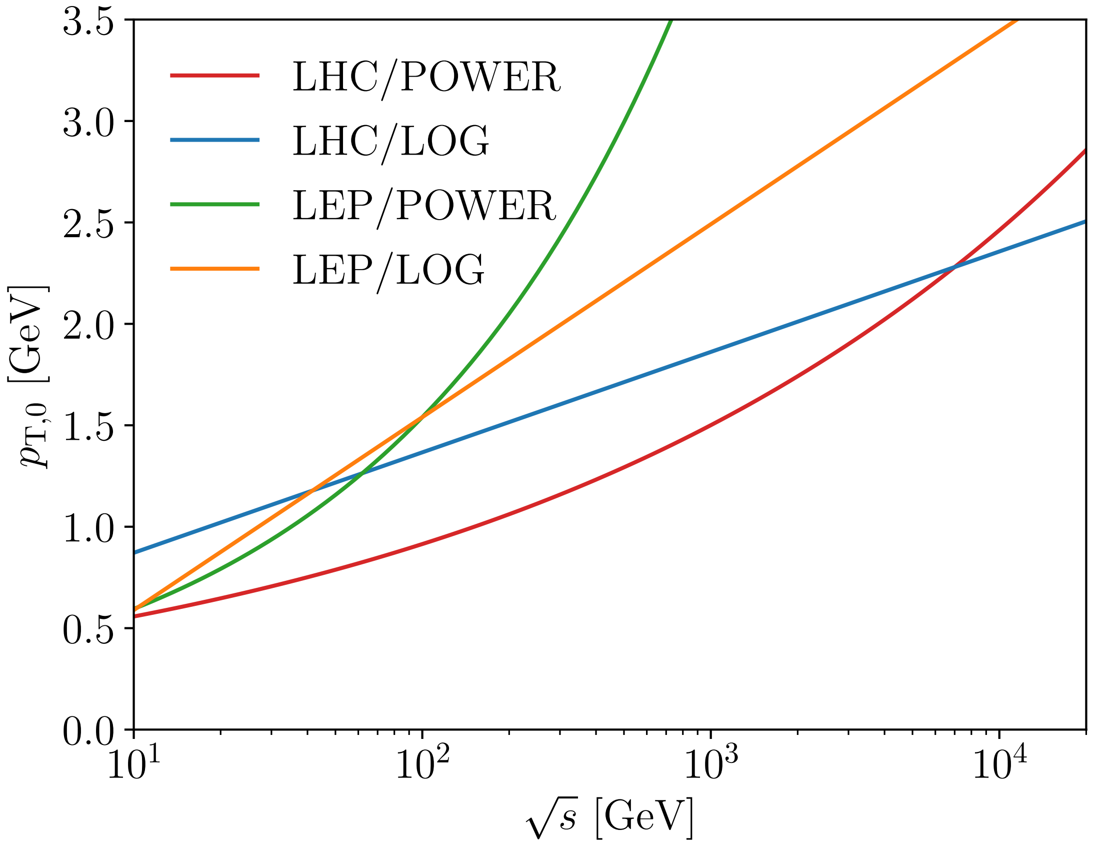
Can MPIs at LHC (pp), HERA (γp) and LEP ($\gamma\gamma$) be described on the same footing?

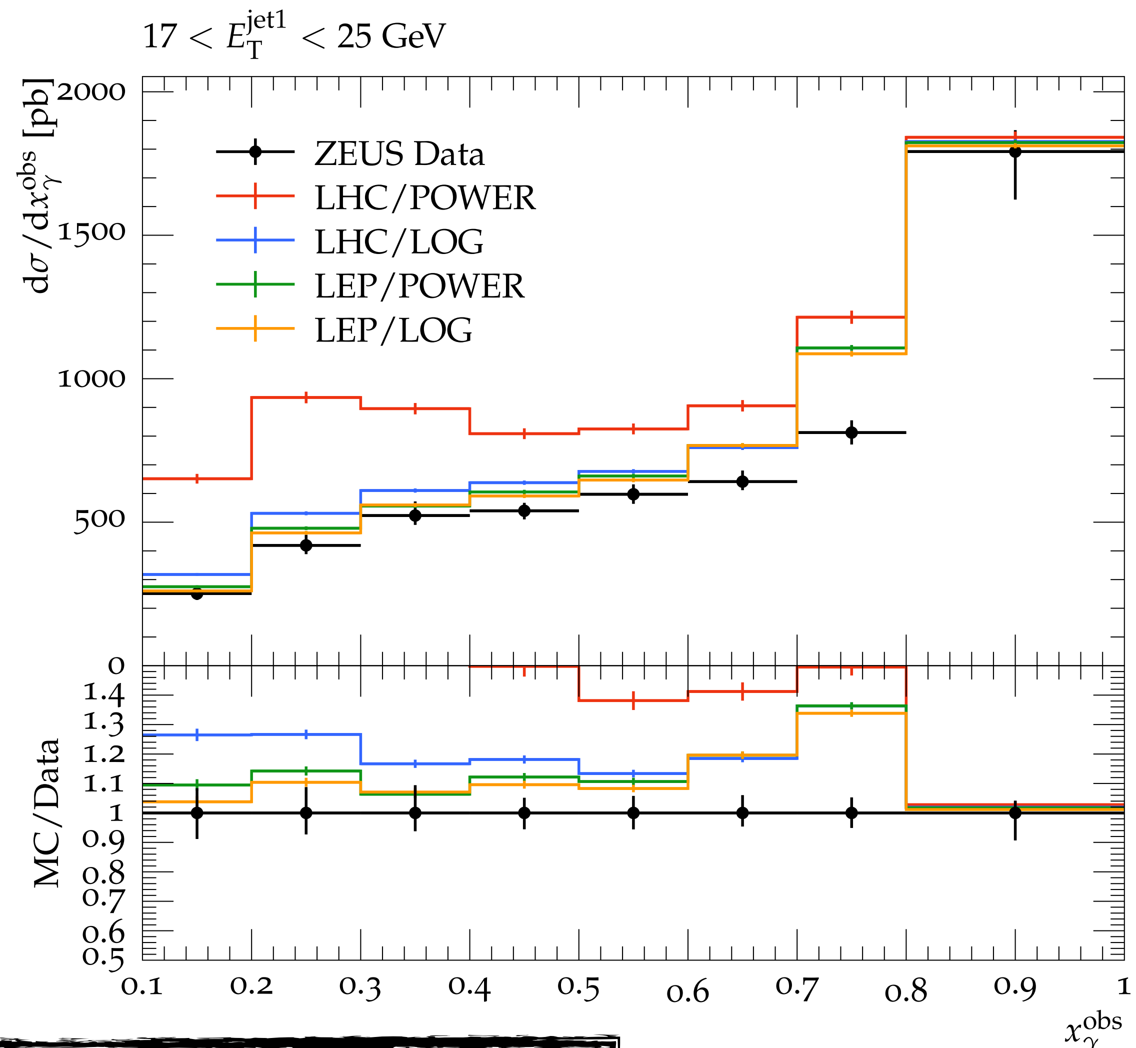
Model development

Example: multiple interactions of photons

Butterworth et al., [2408.15842]

comparing different scaling laws of MPI parameter $p_{T,0}$ in Pythia against data:





With the current model, it is hard to get all the data, pp, γp and $\gamma\gamma$, right

Tuning

Tuning

- Important for **hadronisation**, **colour reconnections**, **underlying event** (aka. multiple-parton interactions) and **beam remnant** modelling
- Want the best description of data without overfitting

Tuning

- Important for **hadronisation**, **colour reconnections**, **underlying event** (aka. multiple-parton interactions) and **beam remnant** modelling

- Want the best description of data without overfitting

How many free parameters do we use?

How do we assess uncertainties?

What data do we fit to?

How global/local do we fit to data?

How to deal with discrepancies?

Tuning

Current procedure (roughly)

see for example [1404.5630], [2203.11385], [2312.05175]

1. Tune **hadronisation to LEP** data
   - Event shape, identified particle momentum distribution/ multiplicities
2. Tune **colour reconnections to LHC and Tevatron** (and maybe LEP)
   - Multiplicities, $\langle p_T \rangle (N_{\mathrm{ch}})$
3. Tune **MPI to LHC and Tevatron**
   - p_T distributions, multiplicities in pseudorapidity

Selection of data, observables and weights subject to personal judgement

Tuning

Uncertainties increasingly important

Les Houches 2023, SM WG, [2406.00708]

- Interplay of parton-shower cutoff and hadronisation pose systematic uncertainty
- Currently, two-point variation, i.e. comparing Pythia and Herwig, is used as estimate; this is not very robust
- Important for measurements of jet substructure
- More Rivet analyses needed, esp. from ALICE;
  Correlation measurements might be key to unveil short-comings of current models
- Idea: a standardized basis of datasets+weights for tuning?

Comparison to data is crucial

for MC development

The HEP-Cedar (and friends) pipeline has become the standard in the event generator community:

[HEPCedar GitLab]

- Rivet: analysis framework for comparison to other MC and data
- YODA: compact and flexible histogramming format
- HepData: open and easy access to data
- HepMC3: event file format
- Professor/Apprentice: tuning

Summary

Take-away points

- MC calculations move on, but new experimental data is not easily measured
- HERA, LEP and other legacy colliders offer a lot of information and (overlooked) measurements, however, the analyses need to be reproducible
- It is vital to preserve measurements from previous experiments for validation, model development and tuning