



MARIA UBIALI **UNIVERSITY OF CAMBRIDGE**

PDFS WITH BSM

ZPW2025, ZURICH

erc PBSP

European Research Council

Established by the European Commission



6TH - 8TH JANUARY 2025



OUTLINE

Introduction

- PDFs and the LHC physics programme
- ➡ The impact of LHC data
- PDFs and BSM
 - An unexpected interplay
 - ➡ PDFs can absorb New Physics
- PDFs with BSM
 - New tools for global interpretation of data
 - Simultaneous SMEFT and PDF fits
- Conclusions and outlook

INTRODUCTION

<u>#1: Theory uncertainty of SM predictions</u>



e.g. Higgs in VBF @ N3LO MSHT+NNPDF arXiv:2411.05373



<u>#1: Theory uncertainty of SM predictions</u>







1/27



High-mass final states \Leftrightarrow Large-x PDFs



Deviations from SM predictions in high energy tails (>2023)

New physics or limited understanding of proton structure?



2/27

A WEALTH OF DATA FROM THE LHC $f_i(x,\mu)$



Ball et al, 2109.02653

Perturbative QCD

Splitting functions P_{ab} known up to approximate N3LO

[Blumlein, Moch, Gehrman, von Manteufel, Sotnikov, Yang, Davies, Vogt, Bonvini, Marzani,...]

> NNPDF4.0: About 30% of input data are LHC data!





IMPACT OF LHC DATA



- HERA data crucial to constrain quark valence (up and to less extent down valence) across intermediate to small x and gluon at small x
- LHC high energy data crucial to provide additional constraints to PDFs, in particular in medium- to large-x gluon and quarks.
- Some **tension** with older fixed-target Drell-Yan and DIS data visible in the large-x region (especially gluon and anti-quarks)







CONSTRAINTS FROM THE HL-LHC



- FCC-hh will go further up by two orders of magnitude in Q² [Mangano et al 1607.01831]

Large x \leftarrow Large E and/or Large Y



• Thanks to luminosity upgrade, HL-LHC will go nearly two orders of magnitude higher in Q², populating the highenergy region, and this will allow to further constrain gluon and (anti)-quarks at large x [Khalek et al, 1810.03639]



THE PRECISION VERSUS ACCURACY CHALLENGE

Now that PDFs are getting more and more precise, it is crucial for accuracy to match precision. Highly non trivial challenges:

- Genuine inconsistencies in experimental inputs Harland-Lang et al, arXiv:2407.070944 Barontini, Costantini, De Crescenzo, Forte, MU - arXiv:2501.xxxxx
- Imperfect fitting methodology
- Inaccurate theoretical framework
 - Missing higher orders (N3LO, resummation...)
 - Other corrections (nuclear, non-perturbative effects...)
 - ➡ BSM effects focus of this talk





PDF4LHC21 study 2203.05506







PDFS AND NEW PHYSICS: AN UNEXPECTED INTERPLAY

SMEFT: A FRAMEWORK FOR INDIRECT DISCOVERIES

- EFT is a well-defined theoretical approach for indirect searches
- Assuming new physics states heavy, can write the Lagrangian with only light SM particles, BSM effects incorporated as a momentum expansion
- SMEFT: assume SM field content and gauge symmetries (apart from accidental)

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{i}^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_{j}^{N_{d8}} \frac{b_j}{\Lambda^4} \mathcal{O}_j^{(8)} + \dots$$

- Full dim-6 basis of operators under SMEFT assumptions includes 2499 operators [Grzadkowski et al, arXiv:1008.4884]
- Current SMEFT fits make flavour assumptions and restricted to a few observables/sectors & reduce the number of operators.
- Huge progress in global dim-6 SMEFT fits and matching with UV models I. Brivio's talk





EXTRACTING PARAMETERS FROM DATA

$$\chi^2 = \frac{1}{N_{\text{dat}}} \sum_{i=1}^{N_{\text{dat}}} (T_i(\{\theta\}, \{c\}) - D_i) \operatorname{cov}_{ij}^{-1} ($$

$T_i(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}, \{c\}) \otimes \hat{\sigma}_i($ **SMEFT WCs**

Parameters determining PDFs at initial scale



$(T_j(\{\theta\}, \{c\}) - D_j)$

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{i}^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_{j}^{N_{d8}} \frac{b_j}{\Lambda^4} \mathcal{O}_j^{(8)} + .$$





EXTRACTING PARAMETERS FROM DATA

 $\chi^{2} = \frac{1}{N_{\text{dat}}} \sum_{i=1}^{N_{\text{dat}}} \left(T_{i}(\{\theta\}, \{c\}) - D_{i} \right) \operatorname{cov}_{ij}^{-1} \left(T_{j}(\{\theta\}, \{c\}) - D_{j} \right)$

 $T_i(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}, \{c\}) \otimes \hat{\sigma}_i(\{c\})$ **SMEFT WCs**

Parameters determining PDFs at initial scale

✓ In a PDF fit typically

 $T_i(\{\theta\}) = \text{PDFs}(\{\theta\}, \{c=0\}) \otimes \hat{\sigma}_i(\{c=0\})$





Kassabov et al, arXiv:2303.06159





EXTRACTING PARAMETERS FROM DATA

$$\chi^2 = \frac{1}{N_{\text{dat}}} \sum_{i=1}^{N_{\text{dat}}} (T_i(\{\theta\}, \{c\}) - D_i) \operatorname{cov}_{ij}^{-1} ($$

 $T_i(\{\theta\}, \{c\}) = \operatorname{PDFs}(\{\theta\}, \{c\}) \otimes \hat{\sigma}_i($ **SMEFT WCs**

Parameters determining PDFs at initial scale

✓ In a PDF fit typically

 $T_i(\{\theta\}) = \text{PDFs}(\{\theta\}, \{c=0\}) \otimes \hat{\sigma}_i(\{c=0\})$

In a fit of SMEFT Wilson Coefficients

 $T_i(\{c\}) = \text{PDFs}(\{\theta = \bar{\theta}\}, \{c = 0\}) \otimes \hat{\sigma}_i(\{c\})$



$(T_j(\{\theta\}, \{c\}) - D_j)$

$$\left\{\left\{C\right\}\right\} \qquad \qquad \mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{i}^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_{j}^{N_{d8}} \frac{b_j}{\Lambda^4} \mathcal{O}_j^{(8)} + \dots \right\}$$







PDF AND SMEFT INTERPLAY

- (SM)EFT fits are performed by assuming a priori that PDFs are SM-like.
- environment might intertwine them



PDFs are low-scale quantities extracted from experimental data at all scales, without considering any potential high-energy bias in theory predictions due to new physics.

In principle low-scale physics is separable from high-scale physics, but complexity of LHC



PDF AND SMEFT INTERPLAY

- From the point of view of PDF fits:
 - How to make sure that new physics effects are not inadvertently fitted away in a PDF fit?
- From the point of view of BSM indirect searches (or specifically in SMEFT fits):
 - What PDFs should be used? Are conservative PDFs good enough? What about flavour?

 [PBSP+..., in progress]
 - How would the bounds change if PDF fits included the same operators that are fitted in SMEFT fit?

[Top sector: PBSP + Rojo, 2303.06159]

[Drell-Yan sector: PBSP + Greljo : 2104.02723, Iranipour + MU: 2201.07240]

[Jet and Dijets: Greljo, Hammou, Merlotti, Smolkovic, MU, in progress]

PBSP

MU + Zahari Kassabov, Maeve Madigan, Luca Mantani, James Moore, Manuel Morales Alvarado, Ella Cole, Elie Hammou, Mark Costantini, Francesco Merlotti







DATA OVERLAP



- Top pair production and single top data included in SMEFT analysis [Hartland et al 1901.05965] [Ellis et al 2012.02779]
- Drell-Yan high mass data [Farina et al 1609.08157, Torre et al 2008.12978]
- ➡ Inclusive Jets data [Alte et al 1711.07484]
- Dijets data [Krauss et al 1611.00767] [Alioli et al 1706.03068] [Bordone et al 2103.10332]
- →Overlap enhanced in HL-LHC projections [Abdul Khalek et al1810.03639]





- Naive solution: remove all data above $Q > Q_{high}$ for example $Q_{high} \ge m_{top}$
- PDFs at large x would rely on rather old DIS data and on ability to define uncertainties in the extrapolation region

• However away all high energy data would result in dramatic uncertainty increase and loss of precision. Knowledge of

CAN PDFS ABSORB NEW PHYSICS?



✓ PDF fitting methodology can be tested via closure test (in the data region) [Del Debbio, Giani, Wilson, 2111.05787, Harland-Lang et al, 2407.07944] and future test (in the extrapolation region) [Cruz-Martinez, Forte, Nocera, 2103.08606]. Closure tests assess methodology robustness and efficiency & faithfulness of uncertainty estimate. Input the "true" PDFs, generate MC data according to the "truth" with exp. uncertainty and check if what you get out of the fit corresponds to the truth

 \checkmark Can we test if we can "safely" add high energy data using this statistical test?



CAN PDFS ABSORB NEW PHYSICS?



- Imagine that on top of the "true" PDFs one inject the "true" BSM model in the MC data
- Generate artificial MC data assuming "true" law of nature =

"true" PDFs + "true" BSM model

 $D_i(\{\bar{\theta}\},\{\bar{c}\})$

- Fit PDFs assuming SM
- Can PDFs absorb signs of new physics?

E. Hammou, Z. Kassabov, M. Madigan, M. Mangano, L. Mantani, J. Moore, M. Morales, MU





15/27

2307.10370

CAN PDFS ABSORB NEW PHYSICS? THE W' CASE AT HL-LHC

- The fit-quality of the global fit is <u>unchanged</u> even with signal from Mw' = 13.8 TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Beyond this point , the fit quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan MC data.
- Already for Mw' = 13.8 TeV the qq~ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large-x compensate or "fit away" the effect of New Physics and we would not know in a real fit.





HL-LHC HM DY 14 TeV - charged current - e ch.





CAN PDFS ABSORB NEW PHYSICS? THE W' CASE AT HL-LHC

- The fit-quality of the global fit is <u>unchanged</u> even with signal from Mw' = 13.8 TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Beyond this point , the fit quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan MC data.
- Already for Mw' = 13.8 TeV the qq~ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large-x compensate or "fit away" the effect of New Physics and we would not know in a real fit.



E. Hammou, HEFT 2024, Bologna





CAN PDFS ABSORB NEW PHYSICS? THE W' CASE AT HL-LHC

- The fit-quality of the global fit is <u>unchanged</u> even with signal from Mw' = 13.8 TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Beyond this point , the fit quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan MC data.
- Already for Mw' = 13.8 TeV the qq~ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large-x compensate or "fit away" the effect of New Physics and we would not 10^{-1} know in a real fit.

Consequence #2: Would see New Physics effects where there are none (for example in WW)

E. Hammou et al, arXiv: 2307.10370









PDFS WITH BSM: SIMULTANEOUS FITS



SIMULTANEOUS FITS OF PDFS AND SM/BSM PARAMETERS

- Simultaneous fits of PDFs and SM/BSM parameters pursued by several groups and experimental collaborations.
- Capture the correlation between PDFs of the proton and parameters that are otherwise neglected.
- Effort towards a new generation of global simultaneous fits

$$T_i(\{\theta\}, \{c\}) = PDFs(\{\theta\}, \{c\})$$



 $\otimes \hat{\sigma}_i(\{c\})$



Liu, Sun, Gao, 2201.06586

Forte, Kassabov 2001.04986

CMS collaboration, 2111.10431







SIMUNET: A TOOL FOR SIMULTANEOUS FITS

- How to perform simultaneous fits of PDFs and SM/BSM parameters?
- SimuNET yields simultaneous fit of PDFs and SMEFT coefficients, it does not have limit in number of parameters that can be fitted alongside PDFs at the initial scale! Extendable to SM parameters!



S. Iranipour, MU (arXiv: 2201.07240)

M. Costantini, E Hammou, M. Madigan, L. Mantani, J. Moore, M. Morales, MU (arXiv: 2402.03308)

Linear dim-6 operator

$$T(\hat{\theta}) = \Sigma(\{c_n\}) \cdot L^0(\theta) = T^{\text{SM}}(\theta) \cdot \left(1 + \sum_{n=1}^N c_n R_{\text{SMF}}^{(n)}\right)$$

$$T^{\rm SM}(\theta) = \Sigma^{\rm SM} \cdot L^0$$

$$\begin{aligned} \text{Ouadratic dim-6 operato} \\ T(\hat{\theta}) &= T^{\text{SM}}(\theta) \cdot \left(1 + \sum_{n=1}^{N} c_n R_{\text{SMEFT}}^{(n)} + \sum_{1 \le n \le m \le N} c_{nm} R_{\text{SM}}^{(n)} \right) \end{aligned}$$





 $c_n c_m$

SIMUNET: A TOOL FOR SIMULTANEOUS FITS

SimuNET released open-source with detailed documentation & allow users https://github.com/HEP-PBSP/SIMUnet

- ✓ PDF only fits NNPDF4.0 methodology
- ✓ SMEFT only fits linear SMEFT
- ✓ Simultaneous SMEFT & PDFs linear SMEFT









SIMUNET: A TOOL FOR SIMULTANEOUS FITS

SimuNET released open-source with detailed documentation & allow users https://github.com/HEP-PBSP/SIMUnet

- ✓ PDF only fits NNPDF4.0 methodology
- ✓ SMEFT only fits linear SMEFT
- ✓ Simultaneous SMEFT & PDFs linear SMEFT

```
dataset_inputs:
- {dataset: NMC, frac: 0.75}
- {dataset: ATLASTTBARTOT7TEV, cfac: [QCD], simu_fac: "EFT_NLO"}
- {dataset: CMS_SINGLETOPW_8TEV_TOTAL, simu_fac: "EFT_NLO", use_fixed_predictions: True}
```

Inject any new physics model in the data and check robustness against PDF absorbing it

```
dataset_inputs:
- {dataset: LHCB_Z_13TEV_DIELECTRON, frac: 0.75, cfac: ['QCD']}
- {dataset: CMSDY1D12, frac: 0.75, cfac: ['QCD', 'EWK'], contamination: 'EFT_LO'}
```



M. Costantini, E Hammou, M. Madigan, L. Mantani, J. Moore, M. Morales, MU arXiv: 2402.03308





SIMUNET: INPUT DATA



Total of ~ 5000 input datapoints, some constraining only SMEFT, some only PDFs, some both SMEFT & PDFs

O(4100) data from DIS, low-energy DY, jets, di-jets, W and Z production, Z pT +

O(400) top data (pair production, single top, charge asymmetry, associated V+t production) and high energy DY data

O(300) Higgs strength and EWPO from LEP

GLOBAL FI

Simultaneous analysis of PDFs and 40 operators in the top+Higgs+EWPO+di-boson sector using simuNET methodology shows that WCs are stable, while PDF uncertainties broaden and PDF fitted simultaneously alongside WCs sits nicely between PDF-only fit without top data and PDF fit including all data - sign of partial NP absorption Exploring New Physics model affecting the gluons through the jets [Hammou,Greljo, Merlotti, Smolkovic, MU - in progress] [Costantini, Lonigro, Mantani, Nocera, MU - in progress]

IMUNET: APPLICATION TO DRELL-YAN SECTOR

✓ Simultaneous analysis of PDFs and Drell-Yan sector Wilson coefficient in the context of universal parameters of DIS + DY (including HL-LHC projections) using simuNET method shows if HL-LHC projections were generated by using a SM model, then WCs bounds broaden and PDF uncertainties change significantly once SMEFT effects allowed in theory predictions entering PDF fit [Greljo et al 2104.02723] [Iranipour, MU 2201.07240]

SIMUNET: APPLICATION TO DRELL-YAN SECTOR

E. Hammou's slides @Moriond 2024

✓ Simultaneous analysis of PDFs and Drell-Yan sector Wilson coefficient in the context of universal parameters of DIS + DY (including HL-LHC projections) using simuNET method shows that if HL-LHC projections were generated by using a W' BSM model, the simultaneous fit would be able to find the "true" SM PDFs and the "true" SMEFT value! [Costantini et al - in preparation]

Case II: HL-LHC MC data generated assuming law of Nature = W' model

BEYOND MONTE CARLO AND NEURAL NETWORKS

- In the quadratic SMEFT fit observed disagreement
- Towards a general Bayesian methodology for PDF [Costantini, Mantani, Moore, MU - POD parametrisation] [Costantini, Del Debbio, Giani, Mantani, MU - GP parametrisation]

M. Costantini, M. Madigan, L. Mantani, J. Moore - 2404.10056

Artwork by: <u>@qftoons</u>

CONCLUSIONS AND OUTLOOK

as we move to the High-Luminosity LHC phase

PDFs and BSM

- Important to disentangle large-x from high-energy / low-energy (for example FPF, EIC) physics coupling to 3rd generation
- PDFs with BSM: simultaneous fits
 - operators has been developed
 - Open source code https://github.com/HEP-PBSP/SIMUnet
 - Effect of simultaneous fits on PDFs and SMEFT depends on the sector
 - Carlo sampling (Bayesian PDFs) [Costantini, Madigan, Mantani, Moore arXiv:2404.10056]
 - Bayesian methodologies can clarify aspects of prior dependence also in PDF fits!

• Interplay between indirect new physics searches via (SM)EFT fits and PDFs is going to be more and more relevant

Identified a "naive" UV scenario such that the high-mass HL-LHC invariant mass can absorb NP and quantified effects More realistic scenarios to explore, especially large-x gluons (jets, top) and flavour observables sensitive to new

General simuNET methodology for simultaneous fits linear SMEFT + PDFs for an arbitrary number of SMEFT

Quadratic SMEFT corrections: need a new methodology to determine joint probability distribution beyond Monte

THANK YOU FOR YOUR ATTENTION

EXTRA MATERIAL

DRELL-YAN DATA PROJECTIONS @HL-LHC

• Add HL-LHC projections for both NC and CC in PDF fit

$$\sigma_i^{\text{hllhc}} \equiv \sigma_i^{\text{th}} \left(1 + \lambda \delta_{\mathcal{L}}^{\text{exp}} + r_i \delta_{\text{tot},i}^{\text{exp}} \right) , \qquad i = 1, \dots, n_{\text{bin}}$$
$$\delta_{\text{tot},i}^{\text{exp}} = \left(\left(\delta_i^{\text{stat}} \right)^2 + \sum_{j=1}^{n_{\text{sys}}} \left(f_{\text{red},j} \delta_{i,j}^{\text{sys}} \right)^2 \right)^{1/2}$$

HL-LHC HM DY 14 TeV - charged current - electron channel

+ muon channel

SMEFT AND PDF CORRELATIONS

HOW TO AVOID NEW PHYSICS CONTAMINATION?

- LHCb on-shell at high rapidity data do not help as quark probed at large x, antiquark at small x
- Need more accurate low-energy/large-x constraining measurements to really disentangle such effects
- energies (hence no "BSM-bias" from heavy new physics) and large-x

Low energy fixed target Drell-Yan data constrain antiquarks at large x, if they were more precise, tension with new-physics contaminated HL-LHC data would be evident!

• Strong motivation for synergy with ep experiments (FCC-eh, EIC and Forward Physics Facility) probing lower

Inclusion of FPF pseudo-data (FASERv2, advSNDv) and EIC structure functions measurements along with HL-LHC (DY NC and DY CC) pseudodata would prevent **BSM-induced bias!**

Hammou, MU, arXiv:2410.00963

19/27

DISCOVERY THROUGH PRECISION: A DARK PHOTON EXAMPLE

- Pre-LHC studies: what is there was a light SUSY coloured partner?
- A light SUSY Parton would modify DGLAP equation and running of α_s
- Comparison to data excludes any light coloured parton on increasing mass range as more (and more precise) data are included in the global PDF fit

DISCOVERY THROUGH PRECISION: A DARK PHOTON EXAMPLE

M. McCullough, J. Moore, MU, arXiv:2203.12628

- Idea: now PDFs are known very precisely, and their uncertainties will continue to reduce in the near future with the HL-LHC, could we do the same for a colourless particle too?
- If there was a lepto-phobic dark photon weakly coupled to quarks via effective Lagrangian

$$\mathcal{L}_{
m int} = rac{1}{3} g_B \bar{q} \not B q$$
 n

it would appear among the partons of the proton.

• To include the dark photon as a constituent of the proton: compute the dark photon splitting functions, and add them to DGLAP evolution. Starting from an appropriate initial-scale ansatz (dark photon generated dynamically off quarks and antiquarks at threshold) and a reference PDF set, evolve using the modified DGLAP equations

$m_R \in [2, 80]$ GeV

$$B \qquad P_{ij} = \left(\frac{\alpha_s}{2\pi}\right) P_{ij}^{(1,0,0)} + \left(\frac{\alpha_s}{2\pi}\right)^2 P_{ij}^{(2,0,0)} + \left(\frac{\alpha_s}{2\pi}\right)^3 P_{ij}^{(3,0,0)} \\ + \left(\frac{\alpha}{2\pi}\right) P_{ij}^{(0,1,0)} + \left(\frac{\alpha_s}{2\pi}\right) \left(\frac{\alpha}{2\pi}\right) P_{ij}^{(1,1,0)} + \left(\frac{\alpha}{2\pi}\right)^2 P_{ij}^{(0,2,0)} \\ + \left(\frac{\alpha_B}{2\pi}\right) P_{ij}^{(0,0,1)} + \cdots, \qquad \alpha_B \sim 0.001$$

DISCOVERY THROUGH PRECISION: A DARK PHOTON EXAMPLE

Precise HL-LHC data can indirectly constrain parameter space of the dark photon in a competitive way compared to direct searches

- The presence of the dark Parton would modify the evolution of standard quarks and gluon.
- Interesting that dark photon effect in PDFs is dominant but was not taken into account in recent publication about evidence for dark photons in DIS data [N. T. Hunt-Smith et al arXiv:2302.11126]