Step: a tool to perform tests of smoothness on differential distributions

Terascale alliance meeting

Patrick L.S. Connor[1] Radek Žlebčík[2]

[1]Universität Hamburg, [2]Charles University

24 November 2021

Bundesministerium für Bildung und Forschung

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

CDCS
CENTER FOR DATA AND COMPUTING IN NATURAL SCIENCES

Introduction

Motivation

Introduction

Step library

***S**moothness **T**est with **E**xpansion of **P**olynomials*

⟶ automated 1D fits with correlations (optional) and early stopping (optional)

Motivation

- Test the **quality** of a differential distribution in terms of the "alignment" of the points around a smooth function.
- Here we will discuss it in the **context** of the QCD interpretation of inclusive jet production[a].
- But **applications** beyond QCD interpretation or even HEP are very likely.

[a] e.g. Toni's presentation in which context it was developed.

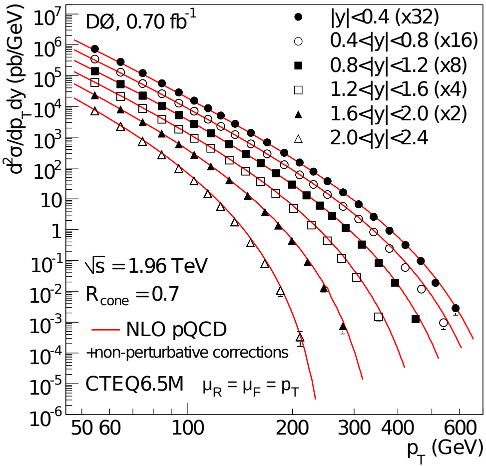
Applications in this presentation

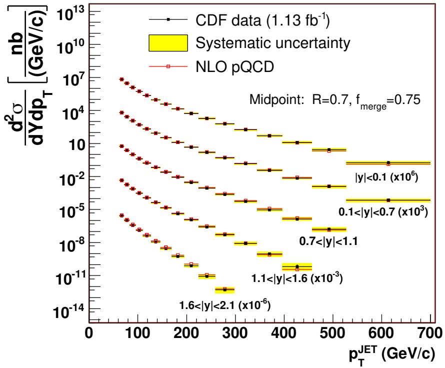
Tevatron DØ & CDF ($p\bar{p}$ at $\sqrt{s} = 1.96\ \mathrm{TeV}$)

LHC CMS & ATLAS (pp at $\sqrt{s} = 8\ \mathrm{TeV}$)

⟶ These data have been used in QCD fits by PDF collaborations [1, 2, 3].

Motivation





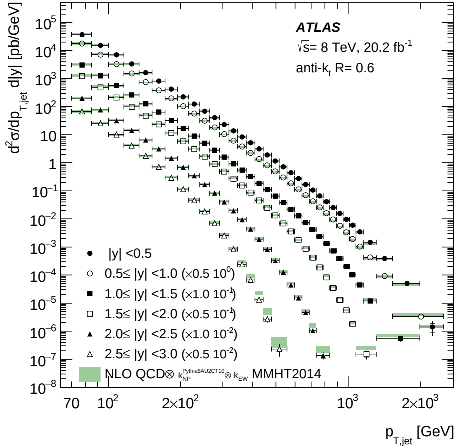
Idea

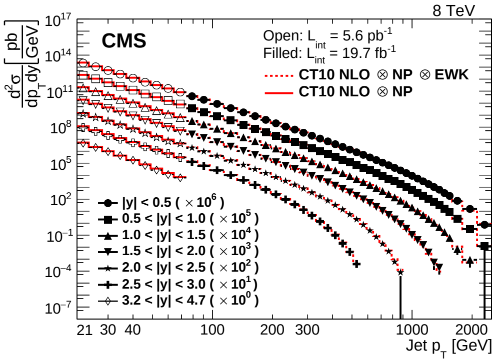
« Physics is smooth. »

« Logarithmic scales can hide monsters. »

⟶ Relevant for **any differential measurement** with “large” number of bins.

Motivation





Test of smoothness

Find an adequate analytical function to fit the spectrum in order to find deviations **beyond** bin-to-bin fluctuations.

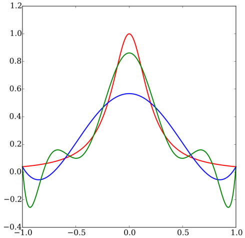
$\longrightarrow$ Applicable in different circumstances, as for instance:

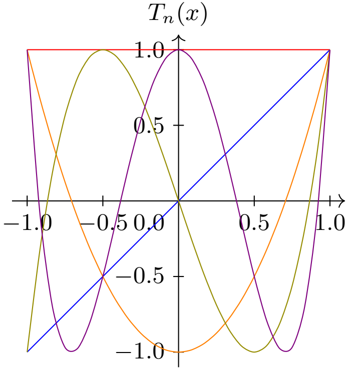
- Impact from bin-to-bin correlations.
- Smooth systematic uncertainties.
- Check combination of different triggers.
- Smooth binned corrections.

Method

Chebyshev polynomials

General form





$T_0(x) = 1$, $T_1(x) = x$, $T_2(x) = 2x^2 - 1$,
$T_3(x) = 4x^3 - 3x$,
$T_4(x) = 8x^4 - 8x^2 + 1$, etc.

Chebyshev polynomials

Difficulties with standard polynomials

1. Runge phenomenon
2. Many parameters

$\longrightarrow$ fits based on standard polynomials $\sum a_i x^i$ difficult...

Definition (first kind)

$$P_n(x) = \sum_{i=0}^{n} a_i T_i(x) \tag{1}$$

where $T_0(x) = 1, \quad T_1(x) = x$ (2)

and $T_{i+1}(x) = 2xT_i(x) - T_{i-1}(x)$ (3)

Interesting properties

1. Fits with Chebyshev polynomials more robust against Runge phenomenon.
2. P_n is a good approximation of P_{n+1}.

General form

Chebyshev fit of inclusive jet p_T spectrum

$$f_n(p_\mathrm{T}) = \exp\left(\sum_{i=0}^{n} b_i T_i\left(2\frac{\log p_\mathrm{T} / \log p_\mathrm{T}^\mathrm{min}}{\log p_\mathrm{T}^\mathrm{max} / \log p_\mathrm{T}^\mathrm{min}} - 1\right)\right) \quad (4)$$

- The exponential and the logarithm are used because of the **steeply falling** nature of spectrum[a]
- Then **rescaling** is necessary as Chebyshev polynomials only cover $[-1, 1]$.
- n is the **number of parameters** (one unity larger than the degree).

[a]They may of course be removed in case of other differential distributions.

Remark

The **only assumption** is a smooth behaviour...
$\longrightarrow$ No assumption on the physical nature of the fitted data!

General form

Objective function

$$\chi_n^2 = \min_{b_{i\leq n}} \left[\left(\mathbf{x} - \mathbf{y}_{b_{i\leq n}}\right)^\intercal \mathbf{V}^{-1} \left(\mathbf{x} - \mathbf{y}_{b_{i\leq n}}\right) \right] \tag{5}$$

- $\mathbf{x}$ corresponds to the binned differential distribution;
- $\mathbf{y}_{b_{i\leq n}}$ corresponds to the values of the smooth fit evaluated at, for instance, the centre of the bins for a given set of parameters $b_{i\leq n}$;
- $\mathbf{V}$ is the covariance matrix of the binned differential distribution.

Algorithm

0. The two first parameters are initialised from the first and last points of the spectrum whereas all other parameters are fixed to 0.
1. **Fit** with all released parameters.
2. **Stop** if one of the following statements is satisfied[a]:
   - i. the χ^2 is compatible with the number of degrees of freedom (ndf);
   - ii. the χ^2/ndf is no longer decreasing;
   - iii. or the number of released parameters has reached the maximum allowed.
3. **Release** the next parameter and go back to item 1.

[a] The two first criteria can be deactivated.

Application

- CDF
- DØ
- CMS
- ATLAS
- Comparison

Application

Datasets

Tevatron Run 2 DØ & CDF ($\mathrm{p\bar{p}}$ at $\sqrt{s} = 1.96\ \mathrm{TeV}$)

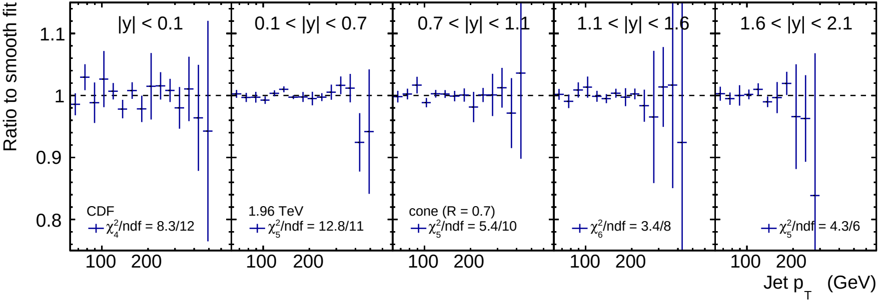
LHC Run 1 CMS & ATLAS (pp at $\sqrt{s} = 8\ \mathrm{TeV}$)

$\longrightarrow$ These data have been used in QCD fits by PDF collaborations [1, 2, 3].

Bin-to-bin uncertainties

detector	stat. corr.	syst. unc.
CDF [4]		
DØ [5]		×
CMS [6]	×	×
ATLAS [7]	×	

CDF

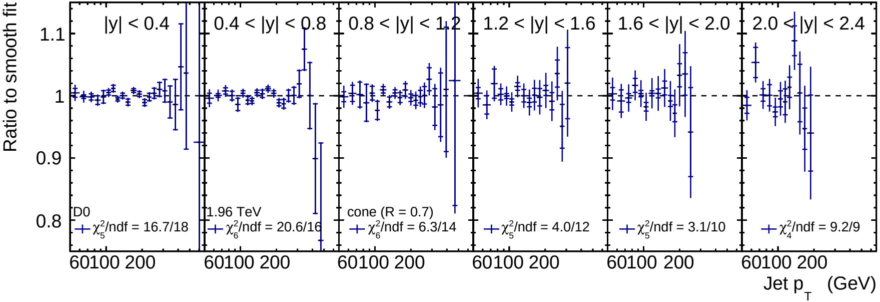


Figure

Steps due to the trigger strategy visible at 146 GeV (96 GeV) in 2nd (3rd) y bin.
$\longrightarrow$ Seems to be negligible in front of the statistical uncertainties.

	$\|y\| < 0.1$	$0.1 < \|y\| < 0.7$	$0.7 < \|y\| < 1.1$	$1.1 < \|y\| < 1.6$	$1.6 < \|y\| < 2.1$
2	163.16 ± 0.38	1896.56 ± 0.38	1165.27 ± 0.39	1441.15 ± 0.41	1224.13 ± 0.47
3	7.03 ± 0.39	97.04 ± 0.39	67.30 ± 0.41	62.69 ± 0.43	41.67 ± 0.50
4	0.70 ± 0.41	8.82 ± 0.41	9.89 ± 0.43	1.69 ± 0.45	1.97 ± 0.53
5	0.56 ± 0.43	1.17 ± 0.43	0.54 ± 0.45	0.43 ± 0.47	0.72 ± 0.58
6	0.55 ± 0.45	0.65 ± 0.45	0.60 ± 0.47	0.42 ± 0.50	0.42 ± 0.63

DØ



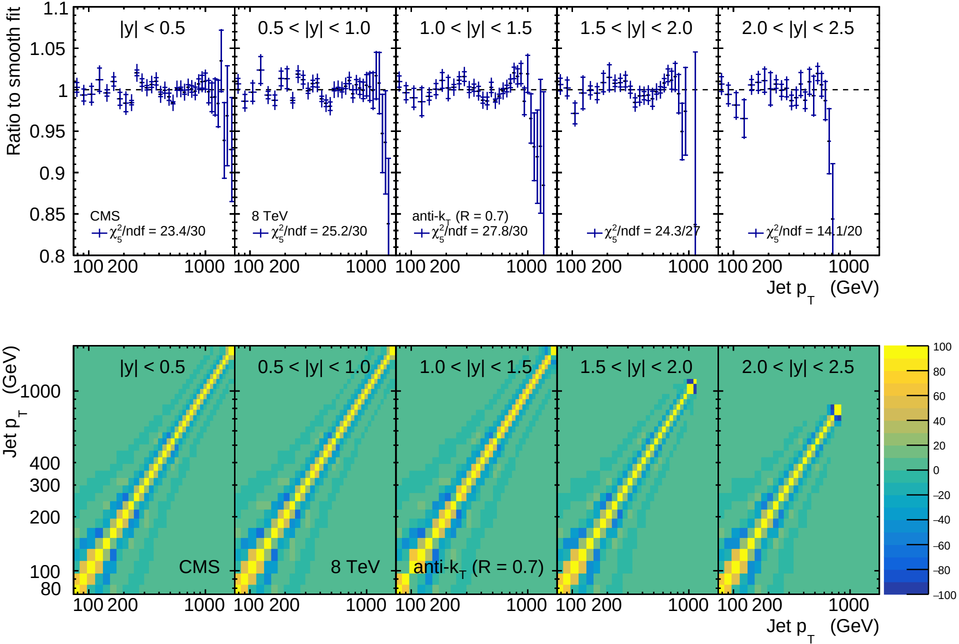
Figure

Additional bin-to-bin uncorrelated systematic uncertainty seems slightly overestimated at $|y| > 0.8$.

	$\|y\| < 0.4$	$0.4 < \|y\| < 0.8$	$0.8 < \|y\| < 1.2$	$1.2 < \|y\| < 1.6$	$1.6 < \|y\| < 2.0$	$2.0 < \|y\| < 2.4$
2	811.57 ± 0.31	1016.79 ± 0.32	795.84 ± 0.33	432.24 ± 0.37	386.47 ± 0.39	329.43 ± 0.43
3	59.47 ± 0.32	76.89 ± 0.32	54.23 ± 0.34	35.55 ± 0.38	27.01 ± 0.41	14.65 ± 0.45
4	6.45 ± 0.32	7.31 ± 0.33	3.35 ± 0.35	1.47 ± 0.39	2.04 ± 0.43	1.02 ± 0.47
5	0.93 ± 0.33	1.44 ± 0.34	0.50 ± 0.37	0.33 ± 0.41	0.31 ± 0.45	0.42 ± 0.50
6	0.88 ± 0.34	1.29 ± 0.35	0.45 ± 0.38	0.35 ± 0.43	0.32 ± 0.47	0.42 ± 0.53

Step
P.L.S. Connor

CMS



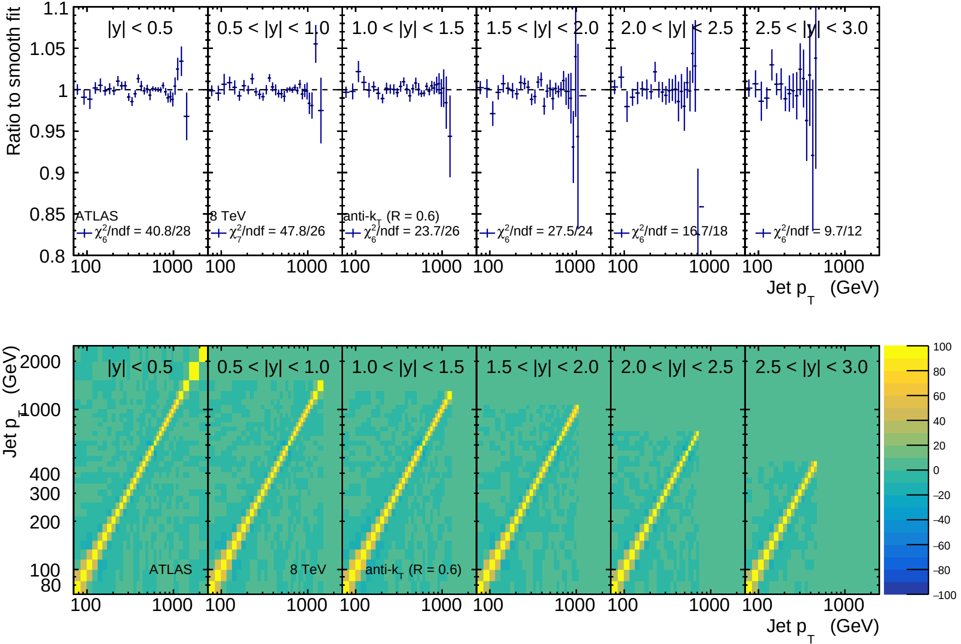
CMS

Figure

- Only the **phase space** relevant for QCD interpretation has been considered.
- **Statistical correlations** are provided directly by CMS and accounted for in each fit.
- Steps related to the **unfolding** may still be seen but are more difficult to find due to correlations and to the unfolding.
- The additional 1% **bin-to-bin uncorrelated systematic uncertainty** is necessary but likely too conservative (wavy shape).

	$\lvert y\rvert < 0.5$	$0.5 < \lvert y\rvert < 1.0$	$1.0 < \lvert y\rvert < 1.5$	$1.5 < \lvert y\rvert < 2.0$	$2.0 < \lvert y\rvert < 2.5$
2	841.51 ± 0.25	1019.44 ± 0.25	1442.31 ± 0.25	1965.81 ± 0.26	1943.62 ± 0.29
3	47.77 ± 0.25	66.72 ± 0.25	115.51 ± 0.25	185.72 ± 0.26	187.95 ± 0.30
4	7.52 ± 0.25	7.55 ± 0.25	14.07 ± 0.25	21.57 ± 0.27	17.79 ± 0.31
5	0.78 ± 0.26	0.84 ± 0.26	0.93 ± 0.26	0.90 ± 0.27	0.70 ± 0.32
6	0.76 ± 0.26	0.65 ± 0.26	0.50 ± 0.26	0.47 ± 0.28	0.41 ± 0.32

ATLAS


Ratio to smooth fit
|y| < 0.5
0.5 < |y| < 1.0
1.0 < |y| < 1.5
1.5 < |y| < 2.0
2.0 < |y| < 2.5
2.5 < |y| < 3.0
ATLAS
8 TeV
anti-k$_T$ (R = 0.6)
χ^2_6/ndf = 40.8/28
χ^2_7/ndf = 47.8/26
χ^2_6/ndf = 23.7/26
χ^2_6/ndf = 27.5/24
χ^2_6/ndf = 16.7/18
χ^2_6/ndf = 9.7/12
Jet p$_T$ (GeV)
Jet p$_T$ (GeV)
|y| < 0.5
0.5 < |y| < 1.0
1.0 < |y| < 1.5
1.5 < |y| < 2.0
2.0 < |y| < 2.5
2.5 < |y| < 3.0
ATLAS
8 TeV
anti-k$_T$ (R = 0.6)
Jet p$_T$ (GeV)

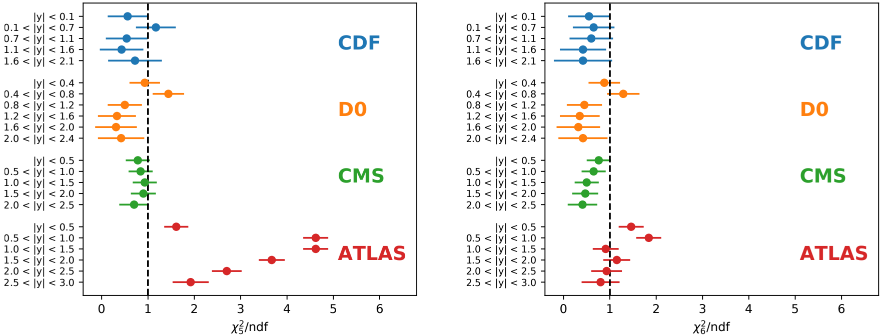
ATLAS

Figure

- **Statistical correlations** are accounted for using replica method.
- The statistical uncertainties seem insufficient to cover the **deviations** from a smooth behaviour in the two first rapidity bins, and no additional bin-to-bin uncorrelated systematic uncertainties are provided.
- This might be related to the issues observed by several PDF collaborations [1, 2, 3].

	$\lvert y\rvert < 0.5$	$0.5 < \lvert y\rvert < 1.0$	$1.0 < \lvert y\rvert < 1.5$	$1.5 < \lvert y\rvert < 2.0$	$2.0 < \lvert y\rvert < 2.5$	$2.5 < \lvert y\rvert < 3.0$
2	2188.98 ± 0.25	3033.97 ± 0.25	4115.24 ± 0.26	3860.63 ± 0.27	2722.43 ± 0.30	1142.21 ± 0.35
3	99.77 ± 0.25	178.11 ± 0.26	325.30 ± 0.26	322.57 ± 0.27	261.13 ± 0.31	92.98 ± 0.37
4	17.57 ± 0.26	27.64 ± 0.26	49.63 ± 0.27	37.35 ± 0.28	17.49 ± 0.32	9.03 ± 0.38
5	1.61 ± 0.26	4.62 ± 0.27	4.62 ± 0.27	3.67 ± 0.28	2.70 ± 0.32	1.92 ± 0.39
6	1.46 ± 0.27	1.84 ± 0.27	0.91 ± 0.28	1.15 ± 0.29	0.93 ± 0.33	0.80 ± 0.41
7	1.46 ± 0.27	1.84 ± 0.28	0.92 ± 0.28	1.18 ± 0.29	0.73 ± 0.34	0.86 ± 0.43

Comparison



Interpretation

- Steps due to the **triggers** are visible (e.g. CDF & CMS).
- Certain **bin-to-bin decorrelated uncertainties** are likely overestimated (DØ in forward region; CMS in all regions) or underestimated (ATLAS).
- Need for higher number of parameters in central region of ATLAS measurement might be related to difficulties encountered in QCD interpretation.

Reference

Article

Source code

Article

arXiv:2111.09968v1 [hep-ph] 18 Nov 2021

Step: a tool to perform tests of smoothness on differential distributions based on Chebyshev polynomials of the first kind

Patrick L.S. Connor[a,1] Radek Žlebčík[b]
[a] *Institut für Experimentalphysik*
Center for Data Computing in natural Sciences, Universität Hamburg
Hamburg, Germany
[b] *Charles University, Prague, Czech Republic*
E-mail: patrick.connor@desy.de

Abstract: We motivate and describe a method based on fits with Chebyshev polynomials to test the smoothness of differential distributions. We also provide a header-only tool in C++ called Step to perform such tests. As a demonstration we use the tool in the context of the measurement of inclusive double-differential cross section in the jet transverse momentum and rapidity at the Tevatron and LHC.

Keywords: smoothness, step, Chebyshev polynomials, inclusive jet, correlations, D0, CDF, CMS, ATLAS, library, C++

[1] Corresponding author

Outline

- Motivation and description of the method
- Application on real data

⟶ Essentially the content of this presentation.

Status

- Posted on arXiv [8].
- Plan to submit it to a journal in the coming hours/days.

Reference

Source code

How-to

- Just download the header `Step.h` from the GitLab repository [9], then the following call is enough to get a ROOT TF1 object:

```
TF1 * f = Step::GetSmoothFit<log,exp>(hist,n);
```

- Here is an example of a more advanced call (but more complex isn't possible):

```
TF1 * f = Step::GetSmoothFit (h,cov,im,iM,n,1,true,cout);
```

- Most general prototype:

```
TF1 * GetSmoothFit (TH1 * h, TH2 * cov, int im, int iM,
        int maxdegree, double nSigmaStop = 1, bool autoStop = true,
        std::ostream& stream = std::cout);
```

⟶ Additional information on the fit performance is available using the static STL container `Step::chi2s`.

Examples

All results shown in the presentation and in the article may be reproduced.

Summary & Conclusions

Summary & Conclusions

- We have presented an original method to test the quality of a differential distribution.
- We have applied the method to several measurements used by PDF collaborations.
- We have provided a package and briefly instructed how to use it.

Thank you for your attention!

Back-up

Acronyms I

ATLAS A Toroidal LHC ApparatuS. 3, 11, 18

CDF Collider Detector at Fermilab. 3, 11, 18

CMS Compact Muon Solenoid. 3, 11, 15, 18

DØ after the location of the detector. 3, 11, 18

HEP High-Energy Physics. 3

PDF Parton Distribution Function. 3, 11, 17, 23

QCD Quantum Chromodynamics. 3, 11, 15, 18

STL Standard Template Library. 21

Step
P.L.S. Connor

Acronyms
References
Visiting card

References I

Tie-Jiun Hou et al. "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC". In: **Phys. Rev. D** 103 (1 Jan. 2021), p. 014013. DOI: 10.1103/PhysRevD.103.014013. URL: https://link.aps.org/doi/10.1103/PhysRevD.103.014013.

L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. "The impact of LHC jet data on the MMHT PDF fit at NNLO". In: **The European Physical Journal C** 78.3 (Mar. 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-5710-7. URL: http://dx.doi.org/10.1140/epjc/s10052-018-5710-7.

Rabah Abdul Khalek et al. "Phenomenology of NNLO jet production at the LHC and its impact on parton distributions". In: **The European Physical Journal C** 80.8 (Aug. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-8328-5. URL: http://dx.doi.org/10.1140/epjc/s10052-020-8328-5.

T. Aaltonen et al. "Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p anti-p Collider Using a Cone-Based Jet Algorithm". In: **Phys. Rev. D** 78 (2008). [Erratum: Phys.Rev.D 79, 119902 (2009)], p. 052006. DOI: 10.1103/PhysRevD.78.052006. arXiv: 0807.2204 [hep-ex].

UHH

Step
P.L.S. Connor

Acronyms
References
Visiting card

References II

Victor Mukhamedovich Abazov et al. "Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: **Phys. Rev. D** 85 (2012), p. 052006. DOI: 10.1103/PhysRevD.85.052006. arXiv: 1110.3771 [hep-ex].

Vardan Khachatryan et al. "Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV". In: **JHEP** 03 (2017), p. 156. DOI: 10.1007/JHEP03(2017)156. arXiv: 1609.05331 [hep-ex].

ATLAS Collaboration. "Measurement of the inclusive jet cross-sections in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: **JHEP** 09 (2017), p. 020. DOI: 10.1007/JHEP09(2017)020. arXiv: 1706.03192 [hep-ex].

Patrick L. S. Connor and Radek Žlebčík. **Step: a tool to perform tests of smoothness on differential distributions based on Chebyshev polynomials of the first kind**. 2021. arXiv: 2111.09968 [hep-ph].

Patrick L.S. Connor and Radek Žlebčík. **Step repository**. https://gitlab.cern.ch/step. 2021.

Patrick L.S. Connor

patrick.connor@desy.de
Universität Hamburg
`https://www.desy.de/~connorpa`

MIN-Fakultät
Institut für Experimentalphysik
Tel.: +49 40 8998-2617
Geb.: DESY Campus 68/121

Center for Data and Computing
in natural Sciences
Tel.: +49 42838-6109
Geb.: Notkestraße 9