

A new look at the estimation of the PDF uncertainties in the determination of electroweak parameters at hadron colliders

Emanuele Bagnaschi^{1,*} and Alessandro Vicini^{2,†}

¹*Paul Scherrer Institut, CH-5232 Villigen, Switzerland*

²*TH Department, CERN 1 Esplanade des Particules, Geneva 23, CH-1211, Switzerland and*

Dipartimento di Fisica “Aldo Pontremoli”, University of Milano

and INFN Sezione di Milano, Via Celoria 16, 20133 Milano, Italy

(Dated: October 11, 2019)

We discuss the determination of electroweak parameters from hadron collider observables, focusing on the W boson mass measurement. We revise the procedures adopted in the literature to include in the experimental analysis the uncertainty due to our imperfect knowledge of the proton structure. We show how the treatment of the proton parton density functions (PDFs) uncertainty as a source of systematic error, leads to the automatic inclusion in the fit of the bin-bin correlation of the kinematic distributions with respect to PDF variations. In the case of the determination of M_W from the charged lepton transverse momentum distribution, we observe that the inclusion of this correlation factor yields a strong reduction of the PDF uncertainty, given a sufficiently good control over all the other error sources. This improvement depends on a systematic accounting of the features of the QCD-based PDF model, and it is achieved relying only on the information available in current PDF sets. While a realistic quantitative estimate requires to take into account the details of the experimental systematics, we argue that, in perspective, the proton PDF uncertainty will not be a bottleneck for precision measurements.

INTRODUCTION

The values of the W boson mass m_W and of the sinus of the leptonic effective weak mixing angle $\sin^2 \theta_{\text{eff}}^{\ell}$ are very precise predictions in the electroweak (EW) sector of the Standard Model (SM) and allow stringent tests at the level of the quantum corrections. The measurements of these two parameters at the Tevatron and at LHC indicate [1–6] the imperfect knowledge of the proton structure as one of the main sources of systematic uncertainty of theoretical origin. The latter affects the computation of the templates used in the fit of the kinematic distributions and eventually the determination of the EW parameters.

The proton collinear parton distribution functions (PDFs) suffer from different uncertainties of experimental as well as theoretical origin. The impact of the error of the data from which the PDFs are extracted is represented by sets of functions, Hessian eigenvectors or Monte Carlo replica, that span in a statistically significant way the functional space of all possible parameterisations. The propagation of the experimental error in the prediction of any observable is achieved by simply repeating the evaluation of the latter several times, with all the available members of the PDF set; mean and standard deviation are eventually computed, the latter being the propagation of the experimental error to the observable under study. All the members of a PDF set share some common theoretical features, like the fact that they all obey the perturbative QCD (pQCD) evolution equations and sum rules, and are thus correlated with each other. While this correlation is automatically included in the propagation of the experimental PDF error to the

prediction of any observable, the determination of a parameter extracted from the simultaneous fit of several observables requires a careful discussion.

In the case of the W boson mass determination, the role of the PDFs has been discussed in the articles presenting the experimental results, for those PDF sets used in the analyses, while a more general comparison of different parameterisations has been presented in Refs. [7, 8]; in all cases the common outcome is that an uncertainty $\Delta^{\text{PDF}} m_W$ at the 10 (20) MeV level is expected in the lepton-pair transverse mass (lepton transverse momentum) case, with the precise value depending on several details of the analyses and on which parameterisations are included in the study. All these studies considered the fit of a kinematic distribution, by combining the information of different bins weighed by their statistical and systematic errors, but neglecting any bin-bin correlation with respect to PDF variations. In Ref. [9] the dependence of the uncertainty on the rapidity range included in the acceptance region was exploited to quantify the benefit given by an m_W measurement at LHCb to the final combination of all the available results, in terms of a reduced PDF uncertainty. The possibility of a systematic extraction of very precise information about the Drell-Yan parton-parton luminosities has been studied in Refs. [10–14], aiming at a better modelling of the initial state and to a consequent reduction of the PDF error on m_W . The impact of measurements at different colliders and energies has been scrutinised in Ref. [15].

We plan to revise the propagation of the PDF uncertainties of experimental origin, in the determination of a parameter obtained via the fit of a kinematic observable.

BIN-BIN CORRELATION AND TEMPLATE FIT DEFINITIONS

At hadron colliders m_W is determined in the charged-current (CC) Drell-Yan (DY) process, from the measurement of observables such as the charged-lepton transverse momentum $d\sigma/dp_\perp^\ell$ and the lepton-pair transverse mass distributions, for which the Jacobian peak enhances the sensitivity to the position of the pole of the W propagator. The finite rapidity detector acceptance and other kinematical constraints induce a sensitivity of the shape of these observables, defined in the transverse plane, to the initial state proton collinear PDFs.

The PDF uncertainty, represented by a set of replica with N_{rep} members, affects the normalisation but also the shape of the observables. Different bins of the same distribution are correlated with respect to a PDF replica variation, as it can be seen in Fig. 1, because of kinematic constraints and due to the theoretical framework shared by all the replica. In Fig. 1 it is quite evident

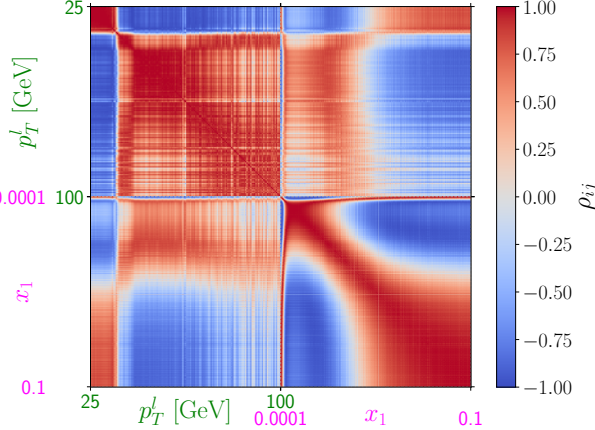


FIG. 1. Correlation with respect to PDF replica variations of the differential cross sections with respect to the charged lepton transverse momentum and to the partonic x_1 variable. Results obtained with normalised distributions. We show the self correlation of $d\sigma/dx_1$ (lower right), of $d\sigma/dp_\perp^\ell$ (upper left) and the cross correlation of the two distributions. Fluctuations due to finite MC statistics are visible as a stripe-like pattern in the plots.

the sudden and strong change of sign of the $d\sigma/dp_\perp^\ell$ self-correlation across the Jacobian peak at $p_\perp^\ell \sim 40$ GeV; the self-correlation of the $d\sigma/dx_1$ distribution also signals the existence of two partonic x ranges, below and above $x \sim 4 \cdot 10^{-3}$; the cross correlations thus establish a link between the parton-parton luminosities, i.e. the source of the PDF uncertainty, and the $d\sigma/dp_\perp^\ell$ distribution, from which m_W is determined, with a non-trivial underlying correlation pattern.

The determination of a Lagrangian parameter from a kinematic distribution via a template fit requires the choice of a Lagrangian density (in our case the SM one)

and of a tool that simulates the observables computed in that model in a well defined setup. The simulation tool is fully specified by the choice of a proton PDF parameterisation, while the parameter, e.g. m_W , is left free to vary when comparing to the experimental data. In this construction the PDF replica represent a one parameter family of models to analyse the data.

The equivalence of the replica in the proton description represents a source of theoretical systematic error, when we try to determine m_W from the fit of a kinematic distribution. We account for this systematics in the following χ^2 definition:

$$\chi_k^2 = \sum_{i \in bins} \frac{((\mathcal{T}_{0,k})_i - (\mathcal{D}^{exp})_i - \sum_{r \in \mathcal{R}} \alpha_r (\mathcal{S}_{r,k})_i)^2}{\sigma_i^2} + \sum_{r \in \mathcal{R}} \alpha_r^2 \quad (1)$$

where, in the bin i of the distribution, we have the following quantities: $\mathcal{T}_{0,k}$ is our fitting model based on the average replica 0 of the PDF set $\mathcal{R}$ and it has been computed with the k -th W -boson mass hypothesis $m_{W,k}$; $\mathcal{D}^{exp}$ is the experimental value and σ_i^2 is its statistical error; the differences $\mathcal{S}_{r,k} = \mathcal{T}_{r,k} - \mathcal{T}_{0,k}$ are computed for each member r of the PDF set and are treated as nuisances with fit parameters α_r . Since the templates are in general affected by statistical Monte Carlo and experimental errors, we should take that into account by considering $\sigma_i^2 = (\sigma_i^{exp,stat})^2 + (\sigma_i^{MC})^2$.

By repeating the minimisation of χ_k^2 , with respect to the α_r , for different values of $m_{W,k}$, the minimum of the sequence labelled by k selects the preferred m_W value and the $\Delta\chi^2 = 1, 4, 9$ rules identify the 1,2,3 standard deviations intervals due to the PDF uncertainty. For a given $m_{W,k}$, the minimum of the χ^2 expression in Eq. 1 can be written [16] with the bin-bin covariance matrix computed with respect to PDF variations and including the statistical and systematic error contributions [17].

$$\chi_{k,min}^2 = \sum_{(r,s) \in bins} (\mathcal{T}_{0,k} - \mathcal{D}^{exp})_r (C^{-1})_{rs} (\mathcal{T}_{0,k} - \mathcal{D}^{exp})_s \quad (2)$$

$$C = \Sigma_{PDF} + \Sigma_{stat} + \Sigma_{MC} + \Sigma_{exp,syst} \quad (3)$$

$$(\Sigma_{PDF})_{rs} = \langle (\mathcal{T} - \langle \mathcal{T} \rangle_{PDF})_r (\mathcal{T} - \langle \mathcal{T} \rangle_{PDF})_s \rangle_{PDF} \quad (4)$$

$$\langle \mathcal{O} \rangle_{PDF} \equiv \frac{1}{N_{cov}} \sum_{l=1}^{N_{cov}} \mathcal{O}^{(l)} \quad (5)$$

where Σ_{stat} is a diagonal matrix with the statistical variances on each bin of the distribution, estimated for a given integrated luminosity $\mathcal{L}$, Σ_{MC} is the diagonal matrix of the squared Monte Carlo error of the templates and N_{cov} is the number of PDF replica used to compute the PDF covariance matrix [18]. We introduce in the full covariance matrix an additional term $\Sigma_{exp,syst}$ to account for experimental systematics, although their faithful description depends on the details of each experiment. In

Eq. 3 we approximate $\Sigma_{exp,syst}$ by using the CMS detector model presented in Ref. [19]. We stress that in this note all the replica are treated as equivalent, i.e. we do not anticipate the impact that future measurements may have in reducing the PDF uncertainty. The approach that we are proposing to include the PDF uncertainty on an EW parameter has to be compared with what has been used in the past, e.g. in Refs. [7, 8], where the analysis relied on the minimisation of a χ^2 defined as

$$\chi^2_{k,r,no-cov} = \sum_{i \in bins} (\mathcal{T}_{0,k} - \mathcal{D}_r)_i^2 / \sigma_i^2 \quad (6)$$

treating the contributions of different bins as independent and weighing them with their statistical error; the templates were generated with the central PDF replica 0, for different mass hypotheses k ; the distributions, computed with N_{rep} different replica, were treated as independent pseudodata and the minimisation was repeated separately for each of them; the resulting N_{rep} preferred $m_{W,r}$ values were eventually analysed computing mean value and standard deviation and ignoring the associated values of $\chi^2_{k,r,no-cov}$; the standard deviation was taken as the estimate of the PDF uncertainty. A similar χ^2 definition, including only diagonal contributions, has been used up to now by the experimental collaborations at Tevatron and LHC.

NUMERICAL RESULTS

We perform all the simulations using the CC-DY event generator provided in the POWHEG-BOX [20, 21], showered with Pythia8.2 [22], setting $\sqrt{S} = 13$ TeV. We restrict ourselves to W^+ production without hindering the generality of our arguments. We apply the acceptance cut $|\eta_l| < 2.5$. We use for our analysis the PDF set NNPDF30_nlo_as_0118_1000 [23], featuring $N_{rep} = 1000$ replica.

In Eq. 2, the templates are computed using the replica 0 of the PDF set, scanning m_W with a 1 MeV spacing in the interval $m_W \in [80.035, 80.735]$ GeV. We let the distribution computed with the central replica 0 of the PDF set, and with a fixed $m_{W,0} = 80.385$ GeV value, play the role of the experimental data $\mathcal{D}_{exp}$; this choice does not spoil the validity of the method and of the conclusion and offers a sanity check on the fit results. The covariance matrix is evaluated with the N_{rep} replica. We checked that the dependence of the covariance matrix on m_W , in the interesting range of ± 20 MeV around the central value of 80.385 GeV, is small and therefore we neglected it in the numerical analysis. The statistical error on the pseudodata is estimated assuming 2 different luminosities, 1, and 300 fb^{-1} .

Since the value of the PDF uncertainty affecting the m_W determination is sensitive to the fit window

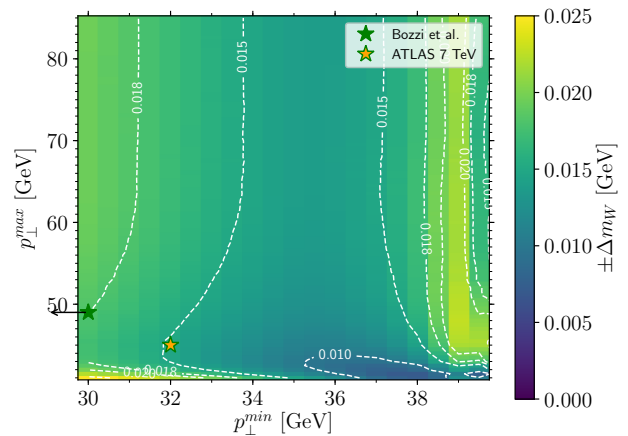


FIG. 2. PDF error estimated by computing the standard deviation of the m_W values corresponding to the minima of the fit of 200 replica onto the template pseudodata represented by the central replica.

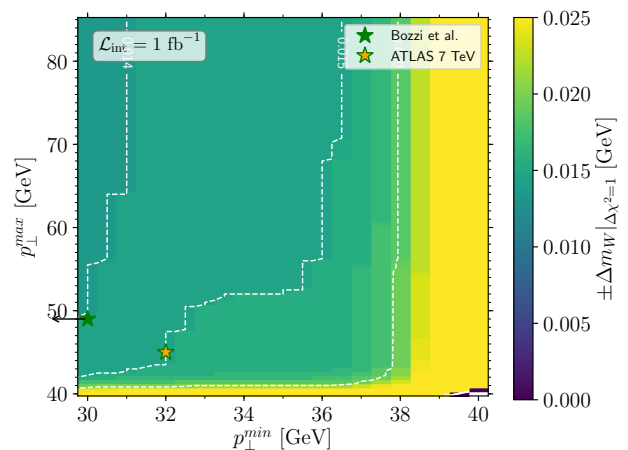


FIG. 3. PDF error as a function of the fit window expressed by its minimum and maximum $p_\perp^\ell$ values. Error estimated from a fit of shape distributions and using a covariance matrix obtained by summing the PDF one with a statistical (diagonal) error on the pseudodata corresponding to 1 fb^{-1} .

$[p_\perp^{min}, p_\perp^{max}]$, we perform a scan in the two values $p_\perp^{min,max}$ and plot, for each point in this plane, the uncertainty value corresponding to the half-width of the $\Delta\chi^2 = 1$ interval.

To present a comparison with the previous approaches, we perform an analysis using the prescription of Eq. 6, using 200 replica, this time with a fixed $m_{W,0} = 80.385$ GeV value, as distinct pseudodata distributions; we generate the templates with the replica 0. In Fig. 2 we show the analysis of distributions normalised to the cross section integrated in the fitting interval. The results, consistent with those presented in Ref. [8], show a weak sensitivity to the upper limit of the fit window, but a clear dependence on its lower limit.

In Figs. 3, 4, we present the results based on Eq. 2, in

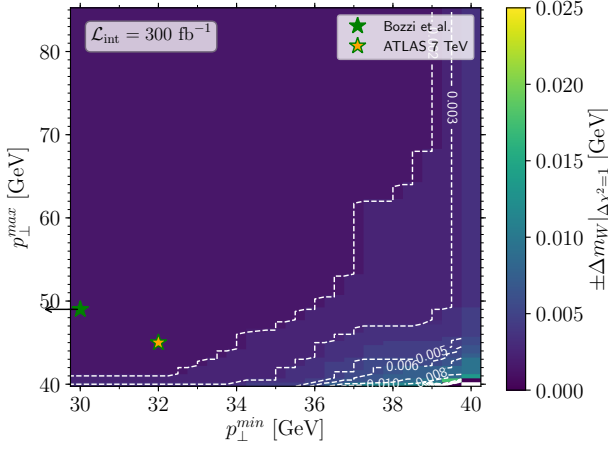


FIG. 4. Same as in Fig. 3, but assuming $\mathcal{L}_{\text{int}} = 300 \text{ fb}^{-1}$.

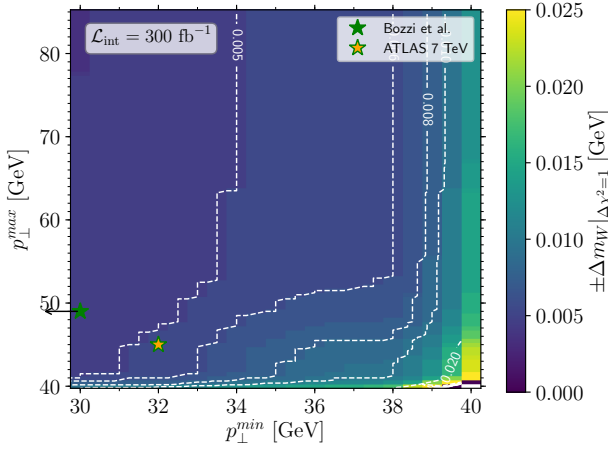


FIG. 5. Same as in Fig. 4, but including also a Monte Carlo error on the templates corresponding to 10^{10} events.

the case of normalised distributions, assuming an experimental integrated luminosity $\mathcal{L}_{\text{int}}$ respectively equal to 1, and 300 fb^{-1} , and no template Monte Carlo error. Fig. 5 also corresponds to 300 fb^{-1} , but we now include a Monte Carlo error extrapolated to a statistics of 10^{10} events [24]. The statistical error is dominant in Fig. 3, while it becomes negligible at high luminosity, putting in evidence a strong reduction of the PDF uncertainty, down to the $\mathcal{O}(1 \text{ MeV})$ level. The Monte Carlo error of the templates has a visible impact, as shown in Fig. 5, and would become negligible in a sample with 200B events. We remark the weak sensitivity of the results to the fit window.

We eventually show in Fig. 6 the results corresponding to 300 fb^{-1} and a systematic error on the muon momentum reconstruction simulated via the model of Ref. [19]. The covariance matrix used in Fig. 5 is added to the one coming from the detector effects, estimated using 100 toys. The negative impact in the description of the peak region balances the improved control on the tails of the distribution, increasing the size of the total error. We

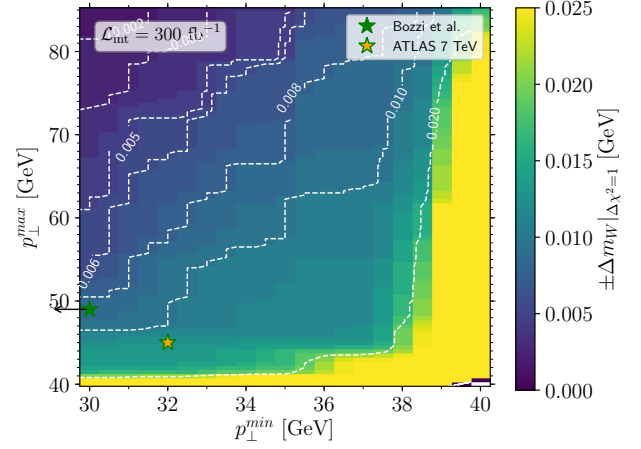


FIG. 6. Same as in Fig. 4, but including also detector effects modelled according to Ref. [19].

have checked that a reduction by a factor of 10 of the Gaussian smearing of the lepton momentum, would lead to uncertainties close to the ones shown in Fig. 4. Other sources of theoretical systematics, such as perturbative QCD or parton shower uncertainties, could become one of the limiting factor for the m_W determination, and will be considered in a future publication.

We observe that this approach strongly reduces the impact of PDF uncertainties because of the specific structure of the bin-bin PDF covariance matrix Σ_{PDF} of the $p_{\perp}^{\ell}$ distribution, with the presence of quite distinct blocks formed by the bins below and above the Jacobian peak [25]. The eigenvalues spectrum of Σ_{PDF} covers more than 7 orders of magnitude between the largest and the smallest elements in absolute value. The broad range of the eigenvalues induces a very narrow shape of the χ^2 distribution as a function of m_W , implying a strong penalty factor for all the templates that do not perfectly overlay their peak position with the one of the data. The penalty applies to the differences in the tails of the $p_{\perp}^{\ell}$ distribution, while, at the same time, an excellent sensitivity to m_W , at the 1 MeV level given by the templates granularity, is preserved, as we explicitly verified as a sanity check of the approach. The important role played by Σ_{PDF} is partially smeared by the interplay between PDF and statistical and systematic errors. Since $C = \Sigma_{PDF} + \Sigma_{\text{stat}} + \Sigma_{MC} + \Sigma_{\text{exp,syst}}$, at low luminosities or low template accuracy the statistical error has a non-trivial interplay with the PDF error, yielding larger uncertainties than the values obtained for each class of errors alone; at high-luminosities, with highly-accurate templates, instead we approach the limit $C \simeq \Sigma_{PDF}$ and the corresponding strong uncertainty reduction.

Similar comments apply to the inclusion of the experimental systematic errors.

The PDF uncertainty band of the $p_{\perp}^{\ell}$ distribution is given by a combination of perturbative and non-

perturbative effects, which can not be analytically separated; although the pQCD elements (DGLAP equations, QCD sum rules) in the proton description can not be qualified as uncertainty sources, they nevertheless enter in the generation of the uncertainty band, because of their entanglement with the data. The covariance matrix allows the effective encoding of a substantial piece of information of pQCD origin, which should not be qualified as uncertainty, and includes it in the fit. The description of the proton in terms of a QCD-inspired model and the representation of the uncertainty via Monte Carlo replicae are thus the two elements allowing the PDF uncertainty reduction. The discussion of the PDF sets representing the associated uncertainty via Hessian eigenvectors will be presented in a future publication.

In conclusion, we have studied the theoretical systematic error due to the PDF uncertainty, focusing on the determination of the W boson mass from the DY $d\sigma/dp_\perp^\ell$ distribution. We included this systematics in the χ^2 that we use in the data fitting, achieving the automatic inclusion of the bin-bin correlation with respect to PDF variations. We observe a drastic reduction of the PDF uncertainty on m_W , which we explain as a consequence of the strong kinematic correlation, of pQCD origin, of the bins above and below the Jacobian peak of the distribution. The interplay of the PDF with the statistical and experimental systematic errors yields non trivial results, when the statistics is limited and systematics not fully understood. We consider this approach promising in view of the reduction of one of the bottlenecks limiting so far the high-precision determination of m_W at hadron colliders. The formulation of Eq. 2 is well suited for a direct and efficient inclusion of the PDF uncertainty in the analysis of the experimental data. The use of this information should not be limited to the fit of m_W , but it should also be part of the determination of any Lagrangian parameter derived in the analysis of LHC observables.

ACKNOWLEDGEMENTS

We thank L. Bianchini, S. Camarda, E. Manca, M. Mangano, G. Rolandi and L. Silvestrini for useful discussions. We also thank the DESY IT department for making us available the computational resources of the BIRD/NAF2 cluster, which have been extensively used for this work. A.V. is supported by the European Research Council under the European Unions Horizon 2020 research and innovation Programme (grant agreement number 740006).

* emanuele.bagnaschi@psi.ch

- [†] alessandro.vicini@mi.infn.it
- [1] T. A. Aaltonen *et al.* (CDF, D0), Phys. Rev. **D88**, 052018 (2013), arXiv:1307.7627 [hep-ex].
 - [2] M. Aaboud *et al.* (ATLAS), Eur. Phys. J. **C78**, 110 (2018), [Erratum: Eur. Phys. J. C78,no.11,898(2018)], arXiv:1701.07240 [hep-ex].
 - [3] T. A. Aaltonen *et al.* (CDF, D0), Phys. Rev. **D97**, 112007 (2018), arXiv:1801.06283 [hep-ex].
 - [4] T. A. collaboration (ATLAS), ATLAS-CONF-2018-037 (2018).
 - [5] A. M. Sirunyan *et al.* (CMS), Eur. Phys. J. **C78**, 701 (2018), arXiv:1806.00863 [hep-ex].
 - [6] R. Aaij *et al.* (LHCb), JHEP **11**, 190 (2015), arXiv:1509.07645 [hep-ex].
 - [7] G. Bozzi, J. Rojo, and A. Vicini, Phys. Rev. **D83**, 113008 (2011), arXiv:1104.2056 [hep-ph].
 - [8] G. Bozzi, L. Citelli, and A. Vicini, Phys. Rev. **D91**, 113005 (2015), arXiv:1501.05587 [hep-ph].
 - [9] G. Bozzi, L. Citelli, M. Vesterinen, and A. Vicini, Eur. Phys. J. **C75**, 601 (2015), arXiv:1508.06954 [hep-ex].
 - [10] E. Manca, O. Cerri, N. Foppiani, and G. Rolandi, JHEP **12**, 130 (2017), arXiv:1707.09344 [hep-ex].
 - [11] E. Bagnaschi, F. Maltoni, A. Vicini, and M. Zaro, JHEP **07**, 101 (2018), arXiv:1803.04336 [hep-ph].
 - [12] A. Bacchetta, G. Bozzi, M. Radici, M. Ritzmann, and A. Signori, Phys. Lett. **B788**, 542 (2019), arXiv:1807.02101 [hep-ph].
 - [13] S. Farry, O. Lupton, M. Pili, and M. Vesterinen, (2019), arXiv:1902.04323 [hep-ex].
 - [14] L. Bianchini and G. Rolandi, JHEP **05**, 044 (2019), arXiv:1902.03028 [hep-ph].
 - [15] *Prospects for the measurement of the W-boson mass at the HL- and HE-LHC*, Tech. Rep. ATL-PHYS-PUB-2018-026 (CERN, Geneva, 2018).
 - [16] L. Demortier (CDF), CDF/MEMO/STATISTICS/PUBLIC/8661 (1999).
 - [17] A similar approach has been adopted in Ref. [26] for the α_s measurement.
 - [18] In the external expectation value we divide by $N_{cov} - 1$.
 - [19] E. Manca, *Validation of the muon momentum resolution in view of the W mass measurement with the CMS experiment*, Ph.D. thesis, INFN, Pisa (2016).
 - [20] L. Barze, G. Montagna, P. Nason, O. Nicrosini, and F. Piccinini, JHEP **04**, 037 (2012), arXiv:1202.0465 [hep-ph].
 - [21] L. Barze, G. Montagna, P. Nason, O. Nicrosini, F. Piccinini, and A. Vicini, Eur. Phys. J. **C73**, 2474 (2013), arXiv:1302.4606 [hep-ph].
 - [22] T. Sjstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].
 - [23] R. D. Ball *et al.* (NNPDF), JHEP **04**, 040 (2015), arXiv:1410.8849 [hep-ph].
 - [24] Our simulations were performed with only 15 millions events, due to our constrained computational resources. For the illustrative purpose of this note we rescaled the Monte Carlo error to the target statistics.
 - [25] A large N_{rep} value avoids the caveats of Ref. [27].
 - [26] V. Khachatryan *et al.* (CMS), Eur. Phys. J. **C75**, 186 (2015), arXiv:1412.1633 [hep-ex].
 - [27] C. Michael, Phys. Rev. **D49**, 2616 (1994), arXiv:hep-lat/9310026 [hep-lat].