Missing higher-order theoretical uncertainties in a Bayesian statistics framework

E. Bagnaschi\textsuperscript{1,2}, L. Jenniches\textsuperscript{3}

\textsuperscript{1} Sorbonne Universités, UPMC Univ Paris 06, UMR 7589, LPTHE, F-75005, Paris, France
\textsuperscript{2} CNRS, UMR 7589, LPTHE, F-75005, Paris, France
\textsuperscript{3} The Niels Bohr International Academy and Discovery Center, The Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

We consider the Bayesian framework proposed recently by Cacciari and Houdeau\textsuperscript{1} to estimate missing higher-order uncertainties (MHOU) which arise in QCD due to unknown higher-order corrections. We address some aspects which were criticised in the original model and discuss the performance of the model in the case of observables without initial-state hadrons by calculating the MHOU for a set of 19 observables at different orders and comparing them to known higher-order contributions. We then extend the method to observables with initial-state hadrons and perform the same study with a set of 20 observables. In both cases, we compare the results to those obtained with the standard procedure of scale variation. As a by-product of this analysis, we derive a heuristic confidence level (CL) for scale-variation (SV) intervals. We find that when assigning a CL of 68\% to SV intervals with the conventional choice of varying the scales within a factor of two with respect to a central scale, one risks to underestimate the uncertainties.

1 Introduction

Precision phenomenology studies aimed for by the Large Hadron Collider (LHC) physics program, require accuracy not only in experimental measurements but also in theoretical predictions. Once experimental and theoretical uncertainties are similar in size, it becomes important to be able to assess quantitatively the importance of unknown higher-order terms in perturbative calculations.

At a hadron collider like the LHC, perturbative QCD calculations are especially important and we therefore perform our study on uncertainties due to unknown higher-orders corrections in this model. In the past, they have been estimated by varying the unphysical renormalisation and factorisation scales which appear in the calculation. This approach has served the QCD community well for more than thirty years to quickly estimate the size of missing higher-order contributions, but especially for hadronic observables, it does not give reliable predictions for the size of the MHOU. In addition, this method does not provide a density profile and thus makes it impossible to assign a statistical interpretation to the intervals obtained with this procedure.

Recently, a new framework based on Bayesian statistics was proposed by Cacciari and Houdeau\textsuperscript{1} with the aim of overcoming some of these limitations. In this contribution, we present an on-going work\textsuperscript{2} which extends the CH framework and addresses some of the drawbacks of the original model. We will also propose an extension of the model to observables with initial-state hadrons. Finally, MHOU obtained with the modified Cacciari-Houdeau framework are compared to SV intervals and as a by-product of our studies, we can assign a heuristic confidence level to the standard SV intervals.
2 The Cacciari-Houdeau (CH) Bayesian model

A generic QCD observable can be written in the following form

\[ O_k(Q, \mu) \bigg|_{\mu=Q} = \sum_{n=l}^{k} \alpha_s^n(\mu) \bigg|_{\mu=Q} c_n(Q, \mu) \bigg|_{\mu=Q} \equiv \sum_{n=l}^{k} \alpha_s^n c_n, \tag{1} \]

where \( Q \) is the physical hard scale of the process and \( \mu \) is a symbol which represents the unphysical scales in the process (e.g. the renormalisation scale) that here we assume to be taken equal to \( Q \). The remainder of the series is the sum of the uncomputed higher orders

\[ \Delta_k = \sum_{n=k+1}^{\infty} \alpha_s^n c_n. \tag{2} \]

The aim of the Cacciari-Houdeau model is to provide a statistically meaningful probability distribution for \( \Delta_k \) starting from the known coefficients \( \{c_1, \ldots, c_k\} \). To achieve this, it uses a few simple assumptions and Bayesian inference to compute the posterior distribution \( f(\Delta_k|c_1, \ldots, c_n) \).

The model relies on the following hypotheses:

- All the perturbative coefficients \( \{c_1, \ldots, c_k\} \) are of the same order of magnitude \( O(c_1) \simeq \cdots \simeq O(c_k) \). This is encoded in a step function for the density distribution of the coefficients with the hidden parameter \( \bar{c} \)

\[ f(c_n|\bar{c}) = \begin{cases} 1 & \text{if } |c_n| \leq \bar{c} \\ 0 & \text{if } |c_n| > \bar{c}. \end{cases} \tag{3} \]

- The density distributions of two coefficients from the set \( \{c_i, i \in I\} \) are independent

\[ f(\{c_i, i \in I\}|\bar{c}) = \prod_{i \in I} f(c_i|\bar{c}). \tag{4} \]

- We do not have any information on the order of magnitude of the hidden parameter \( \bar{c} \). Therefore, a non-informative prior on the hidden parameter \( \bar{c} \), a log-uniform distribution, is chosen

\[ f_\epsilon(\bar{c}) = \frac{1}{2\epsilon} \frac{1}{\ln \epsilon} \frac{1}{\bar{c}} \chi_{\epsilon \leq \bar{c} \leq 1/\epsilon}. \tag{5} \]

Here \( \epsilon \) is taken to zero at the end of the computation.

If the expansion parameter is sufficiently small, we can assume that the remainder of the series is dominated by the first unknown order, i.e. \( \Delta_k \simeq \alpha_s^{k+1}c_{k+1} \). Then, the posterior of the model for \( \Delta_k \) can be computed analytically and takes in a simple form

\[ f(\Delta_k|c_1, \ldots, c_k) \simeq \left( \begin{array}{c} n_c \\ n_c + 1 \end{array} \right) \frac{1}{2\alpha_s^{k+1}\bar{c}_k} \left\{ \begin{array}{ll} 1 & \text{if } |\Delta_k| \leq \alpha_s^{k+1}\bar{c}_k \\ \left( \frac{\alpha_s^{k+1}\bar{c}_k}{|\Delta_k|} \right)^{n_c+1} & \text{if } |\Delta_k| > \alpha_s^{k+1}\bar{c}_k, \end{array} \right. \tag{6} \]

where \( n_c \) is the number of coefficients available in the computation and \( \bar{c}_k \equiv \max(|c_1|, \ldots, |c_k|) \).

3 The modified Cacciari-Houdeau (CH) model

One of the drawbacks of the original CH model is that it leads to posterior distributions of the remainder function \( \Delta_k \) which depend on the expansion parameter of the series expansion of the observable. Moreover, the coefficients satisfy the model’s hypothesis to all be of the same order of magnitude only for the “correct” choice of the expansion parameter. In perturbative QCD, the
expansion parameter is not uniquely defined. To address these issues, we introduce a parameter $\lambda$ which reflects our ignorance of the optimal expansion parameter for QCD observables. Our expression for a generic QCD observable is then given by

$$O_k = \sum_{n=1}^{k} \left( \frac{\alpha_s \lambda}{n} \right)^n (n-1)! \lambda^n c_n \equiv \sum_{n=1}^{k} \left( \frac{\alpha_s \lambda}{n} \right)^n (n-1)! d_n,$$

where we have isolated a factor $(n-1)!$ in the expansion which can be motivated from theory by looking at the behaviour of the coefficients $c_n$ for large $n$ due to renormalon chains. We assume the same priors on the modified coefficients $d_n$ as on the original $c_n$. Then, we tune $\lambda$ by measuring the performance of the model for a fixed $\lambda$ value on a given set of observables. At a given order and for a given degree of belief (DoB), we calculate for each observable the corresponding interval. Then, we compute the success rate defined as the ratio between the number of observables whose next-order is within the computed DoB interval over the total number of observables in the set. We define the optimal $\lambda$ value to be such that the success rate is equal to requested DoB.

With these modifications, the analytic expression for the posterior density distribution for $\Delta_k$ is given by

$$f(\Delta_k|d_1, \ldots, d_k) \simeq \left( \frac{n_c}{n_c+1} \right) \frac{\lambda^{k+1}}{2k!^{k+1}d_k} \left\{ \begin{array}{ll} 1 & \text{if } |\Delta_k| \leq k! \left( \frac{\alpha_s}{\lambda} \right)^{k+1} d_k \\ \frac{k!^{k+1}d_k}{(\Delta_k)^{k+1}} & \text{if } |\Delta_k| > k! \left( \frac{\alpha_s}{\lambda} \right)^{k+1} d_k \end{array} \right. \ (8)$$

where $n_c$ is the number of coefficients available in the computation and $d_k = \max(|d_1|, \ldots, |d_k|)$. This analytic expression captures the general features of the posterior distributions for $\Delta_k$ produced by this class of models: a flat top with power suppressed tails.

### 3.1 The model’s dependence on the prior on $\bar{c}$

The choice of the priors in a Bayesian model is arbitrary and subjective. In the original formulation, the prior on $\bar{c}$ was chosen as non-informative as possible, i.e. a log-uniform distribution. However, the downside of such a conservative choice is that at low perturbative orders, the tails of the posterior distribution for $\Delta_k$ are very long and give rise to very large intervals if a large DoB is given as an input to the model. As a test, we use a more informative prior. If we scale all coefficients $d_n$ by the first coefficient $d_1$, the hypothesis that all coefficients are of order unity is equivalent to the original hypothesis that all coefficients are of the same order of magnitude. Hence, we can also consider a log-normal distribution around zero

$$f(\bar{c}) = \frac{1}{\sqrt{2\pi \sigma \bar{c}}} e^{-\frac{(\log \bar{c} - \mu)^2}{2\sigma^2}}, \quad \bar{c} > 0 \ (9)$$
as a prior on the hidden parameter $\bar{c}$. Care has to be taken that the width of the log-normal distribution is chosen large enough which means in accordance with the actual variations of the coefficients.

We see from fig. 1 that indeed using a more informative prior (red curve), suppresses the tails of the posterior distributions and therefore makes the error bars smaller. Here, a log-normal distribution with $\sigma = 1$ was used. Nevertheless, we use the original log-uniform prior for our main analysis to try to be as conservative as possible and to avoid biasing our results. In addition, the log-uniform prior leads to an analytic expression which is convenient for the numerical analysis of a large set of observables and for the usage of the method in the context of the LHC physics program.

4 Results for EW precision observables

To tune the $\lambda$ parameter in the non-hadronic case, we use a set of 19 observables including hadron production in electron-positron scattering, DIS sum rules, bottom-quark and Z-boson decays, event shape variables, splitting kernels and finally Higgs decays. We show the results of the tuning procedure described in section 3 in figure 2. On the left, we have a histogram with the “optimal” values of $\lambda$ determined in a scan of DoBs between 0.05 and 0.95, separately at LO and at NLO. On the right, we plot the success rate of the model versus the requested DoB for a set of six $\lambda$ values, combined at LO+NLO. We see that there is excellent agreement between the plots. In the plot on the right, we can identify the the best value of $\lambda$ since its curve should lie close to the straight dashed line, where the success rate is equal to the input DoB. The grey band represents the 1-$\sigma$ error on the measurement of the success rate of 19 independent observables. We can use it to roughly estimate how well we can determine $\lambda$ for a finite set of observables. From both plots, we can extract an estimated interval for the optimal $\lambda$ of $[0.7, 1.1]$ at LO and in $[0.9, 1.2]$ at NLO. We observe that the model gives consistent results at LO and NLO, so we use $\lambda = 1$ as our best value of $\lambda$ for the non-hadronic case at all orders for our subsequent analysis.

In figure 3, we display the results for the Higgs decay channel $H \rightarrow \gamma\gamma$, which is particularly important for experimental searches at the LHC. We plot in blue the 68% DoB (dark-thick) and the 95% (light-thin) error bars as produced by the CH model with $\lambda = 1$. As a comparison, we plot in red the scale variation results for a rescaling factor of 2 (dark-thick) and 4 (light-thin). For this observable, with small QCD corrections, we see that the NLO-QCD contribution is contained in the 68% CH interval and in the $r = 2$ scale variation interval of the LO-QCD
result. We also observe that the 68% DoB interval is comparable to the $r = 4$ scale variation interval. On the right, we show the characteristic density profile of the CH model, a flat plateau with power suppressed tails. We observe that the $r = 2$ scale variation interval is smaller than the size of the flat top part of the distribution.

5 Extension to hadronic observables

The original CH model was formulated with observables without initial-state hadrons in mind. Extending it to observables with initial-state hadrons is not straightforward. We can write a generic hadronic observable as a convolution

$$O_k(\tau, Q) = \mathcal{L}(Q) \otimes \sum_{n=1}^{k} \alpha_s^n C_n(Q)$$

where $\mathcal{L}$ is the parton luminosity, $C_n(Q)$ is the hard-scattering coefficient function and $Q$ is the characteristic scale of the process. All the unphysical scales are taken to be equal to $Q$. In eq. (10), the equivalent of the short-scale coefficient $c_n$ is the coefficient function $C_n$, which is a distribution and not a simple number. It is therefore not possible to directly apply the CH method to it. Another issue is the presence of non-perturbative physics hidden in the parton luminosity. There are two possible ways to extended the CH to this case

1. Express the cross section as a series whose coefficients are convolutions of the coefficients of the partonic cross section and the PDFs and then proceed just as in the non-hadronic case.

2. Translate the observable to Mellin space, find the dominant moment $N^*$ of the Mellin inversion integral and apply the CH method to the coefficient function evaluated at $N^*$. Finally apply the uncertainty to the full observable taking into account a rescaling factor between the full observables and the observable in Mellin space.

In the first approach, we simply rewrite the observable as

$$O_k(\tau) = \sum_{n=1}^{k} \left( \frac{\alpha_s}{\lambda_h} \right)^n (n - 1)! h_n(\tau),$$

where $h_n$ now implicitly contains the non-perturbative physics from the parton luminosity and the perturbative terms that come from DGLAP evolution. In a first approximation, we assume that this content is equal at all orders and therefore it does not influence the size of our uncertainty intervals. The optimal expansion parameter is denoted by $\alpha_s/\lambda_h$ to distinguish it from...
the one used in the non-hadronic case. We re-tune it using the same procedure as previously and an appropriate set of observables with initial-state hadrons.

The second approach is based on the Mellin space expression of the observable. Instead of a convolution of distributions, it is now just a simple product of numbers

\[
O_k(N, Q) = \mathcal{L}(N + 1) \sum_{n=l}^{k} \frac{\alpha_s^n}{\lambda_h^n} (n-1)! \, D_n(N, Q). 
\]

It was recently demonstrated by Forte and collaborators that a dominant moment of the Mellin inversion integral exists and that a saddle point approximation around it is a good approximation of the full result (at least for Higgs-boson production in gluon fusion and for Drell-Yan processes). Moreover, it is a good approximation to use the same dominant moment at all perturbative orders. We can exploit this observation and apply the \( \text{CH} \) method to hadronic observables. More precisely, we apply the \( \text{CH} \) method to the Mellin space coefficient functions at the dominant moment \( N^* \) and calculate its uncertainty interval. The uncertainty interval for a requested DoB for the complete observable is obtained by rescaling proportionally the result to the full observable. This approach is less straightforward to implement in practice, since it requires the determination of the dominant moment for every observable. On the other hand, it has the advantage of separating clearly and explicitly the hard-scattering part from long-range physics.

6 Results for hadronic observables

By proceeding exactly as in the non-hadronic case, we tune the \( \lambda_h \) parameter using a set of hadronic observables including DIS structure functions, Higgs production, heavy-quark production, Drell-Yan with and without jets and vector-boson production. In figure 4, we show the results of this procedure. On the left, we histogram the optimal \( \lambda_h \) values at LO and at NLO. On the right, we plot the success rate of the model versus the requested DoB for six values of \( \lambda_h \). Again, we have a good agreement between the two different methods of presenting the results. We see that different from the non-hadronic case, the histograms are wider and the peaks at LO and at NLO are more separated. Nevertheless, we can extract with good approximation an interval for \( \lambda_h \) of \([0.4,0.8]\). In our further analysis, we use the central value of \( \lambda_h = 0.6 \).

In figure 5, we show the error bars for Higgs production in gluon fusion, computed with the two different extensions of the \( \text{CH} \) model to hadronic observables. The blue error bars are
7 Scale variation

As a by-product of our analysis, we have also analysed the statistical behaviour of the standard scale-variation procedure. In figure 6, we show the heuristic CL value for the scale-variation interval as a function of the rescaling factor used to generate the interval by rescaling the central value by the factor \( r \) or \( 1/r \). In the hadronic case, we have excluded the extreme combinations of factorisation and renormalisation scales because they give rise to large logarithms. From the left plot, we see that in the non-hadronic case, the LO interval reaches a heuristic CL of 68% at \( r = 4 \) and then climbs towards a flat plateau with a heuristic CL of 80% while at NLO, a plateau with a heuristic CL of 68% is reached for \( r = 2 \). The right plot shows that in the hadronic case, the LO intervals can never be interpreted as 68% intervals, while the NLO intervals reach a CL of 68% at values of \( r \) slightly bigger than 4 and then climb to a flat plateau around a heuristic CL of 80%.
8 Conclusions and Outlook

With our sets of observables with and without initial-state hadrons, we can assign a heuristic CL to scale variation. In general, values between $r = 2$ and $r = 4$ lead to a heuristic CL of about 68% except for for hadronic observables at LO, where scale-variation does not seem to reflect the size of MHOU correctly. Nevertheless, scale-variation does not provide density profiles which allow to calculate intervals at different CL’s. The CH model overcomes some of the downsides of scale variation. Not only does it give consistent results at LO and NLO, but it also naturally provides density profiles which give the intervals obtained in this formalism a statistical interpretation. The model can be extended to hadronic observables though the results for the optimal expansion parameter vary more between LO and NLO. This might be solved by introducing specific observable classes, e.g. sorting observables by their dominant production channel or whether they are more “inclusive” or “exclusive” depending on whether there are jets in the final state or not.

Another improvement of the model can be obtained by introducing a prior on the $\lambda$ parameter which reflects our ignorance of the optimal expansion parameter and to use Bayesian inference to tune $\lambda$ rather than testing a posteriori the performance of the model for different expansion parameters.

Finally, a survey of the applicability of the model to resummed and differential observables is required for a wider usage in the context of the LHC physics program.

Acknowledgements

This work was done in collaboration with Matteo Cacciari and Alberto Guffanti. E.B. was supported by the Research Executive Agency (REA) of the European Commission under the Grant Agreements PITN-GA-2010-264564 (LHCPhenoNet). L.J. was supported by the Discovery Center and the Niels Bohr Institute in Copenhagen.

References