Theory uncertainties from missing higher orders in PDF fits

Marco Bonvini

INFN, Rome 1 unit

DIS 2022, Santiago de Compostela, 3 May 2022

mostly based upon: Eur.Phys.J.C 80 (2020) 10, 989 — arXiv:2006.16293

Istituto Nazionale di Fisica Nucleare
Sezione di ROMA

Good reasons for including theory uncertainties in PDF fits

- more reliable determination of PDF uncertainty
- perhaps reduced PDF uncertainties in some regions
  (if fixed-order predictions are inaccurate and, e.g., NNLO theory for DIS favours a smaller gluon in some region of x while NNLO Drell-Yan production favours a larger gluon in the same region, the resulting NNLO fit will likely have a larger uncertainty for the gluon in that region)
- possibility to perform higher order (N^3LO) PDF fits without having all ingredients at that order
  (for instance, if a process is known at NNLO only, it can be included anyway because, within uncertainty, it will be compatible with the unknown N^3LO)

For all this to work theory uncertainties must be reliable

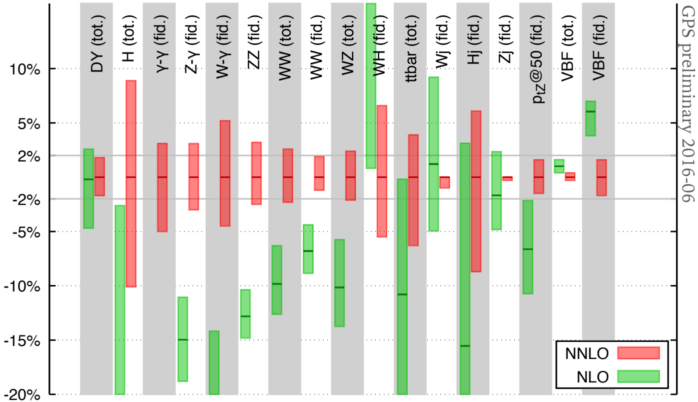
State of the art

- NNPDF 3.1 studies (at NLO) [1905.04311] [1906.10698]
- Consistent use of theory uncertainties in PDF fits and cross section predictions [Harland-Lang,Thorne 1811.08434] [Ball,Pearson 2105.05114]
- ongoing studies (MCscales, MSHT20) *(previous talks)*

All this is based on scale variation as the tool to quantify theory uncertainties from missing higher orders (MHO)

WHAT PRECISION AT NNLO?

Slide from Gavin Salam, PSR 2016



For many processes NNLO scale band is ~±2%

Though only in 3/17 cases is NNLO (central) within NLO scale band…

11

Beyond canonical scale variation

Also caveats of canonical scale variation:

- the result depends on the central scale chosen
- the variation by a factor of 2 is arbitrary
- no probabilistic interpretation

New definition of theory uncertainties from missing higher orders:

- reliable
- less dependent on arbitrary assumptions
- probabilistically well defined

Ideally, theory uncertainty from MHO should be a **probability distribution**

A probabilistic definition in this context can only be based on a **Bayesian approach**

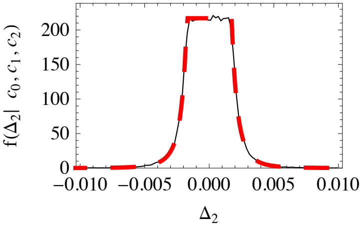
Cacciari and Houdeau [1105.5152] proposed a probabilistic model for the interpretation of theory uncertainties, based on the behaviour of the perturbative expansion

$$\Sigma = \sum_k c_k \alpha_s^k$$

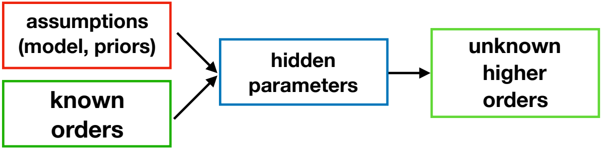
"We make the assumption that all the coefficients c_k in a perturbative series share some sort of upper bound $\bar{c} > 0$ to their absolute values, specific to the physical process studied. The calculated coefficients will give an estimate of this $\bar{c}$, restricting the possible values for the unknown c_k."

In other words, the model assumes that

$$|c_k| \leq \bar{c} \qquad \forall k$$



Inference scheme



Inference on the unknown coefficients c_k

$$P(\text{unknown } c_k|\text{known } c_k) = \int d\text{pars}\, P(\text{unknown } c_k|\text{pars}) P(\text{pars}|\text{known } c_k)$$

in terms of the posterior distribution of the hidden parameters

$$P(\text{pars}|\text{known } c_k) \propto P(\text{known } c_k|\text{pars}) P_0(\text{pars})$$

which depends on the prior distribution $P_0(\text{pars})$ and on the model through the likelihood $P(c_k|\text{pars})$

Cacciari-Houdeau: $P(c_k|\bar{c}) \propto \theta(\bar{c} - |c_k|), \quad P_0(\bar{c}) \propto 1/\bar{c}$

Recent progress: my proposal(s)

- CH probabilistic framework is good (probably the only way to define probabilistically a theory uncertainty from missing higher orders)
- better model assumptions on the behaviour of the expansion
- do not forget scale dependence:
  - **as a tool**, to gain further information on missing higher orders (as in canonical scale variation)
  - **as an issue**, due to the need of choosing a scale

Model 1:
geometric behaviour model

Model 2:
scale variation model

Other models:
variants, combinations, ...

a unified probabilistic way to deal with scale dependence

Main models at fixed scale

More general expansion

$$\Sigma = \Sigma_{\mathrm{LO}}(\mu) \sum_{k\geq 0} \delta_k(\mu) \qquad\qquad \Sigma_{\mathrm{LO}}(\mu)\delta_k(\mu) = c_k(\mu)\alpha_s^k(\mu)$$

Proposed models:

- geometric behaviour model (improved CH)

$$|\delta_k(\mu)| \leq c a^k \qquad\qquad \text{CH: } \left|c_k \alpha_s^k\right| \leq \bar{c}\alpha_s^k$$

  depends on two hidden parameters c, a, it accounts for a possible power growth of the coefficients within the model

  Asymmetric variant, called abc model, proposed in [Duhr,Huss,Mazeliauskas,Szafron 2106.04585]

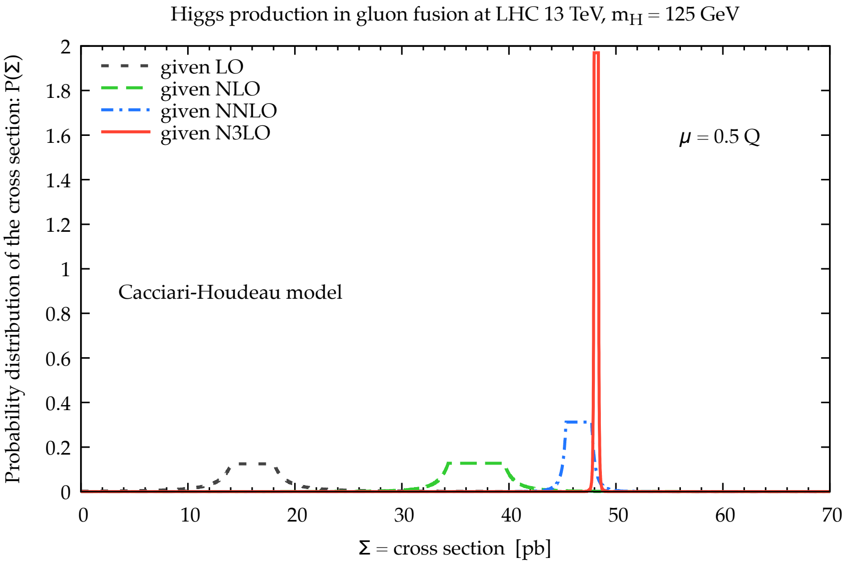
- scale variation model (inspired by canonical scale variation)

$$|\delta_{k+1}(\mu)| \leq \lambda r_k(\mu) \qquad\qquad r_k(\mu) \simeq \left|\mu \frac{d}{d\mu} \log \Sigma_{\mathrm{N}^k\mathrm{LO}}(\mu)\right|$$

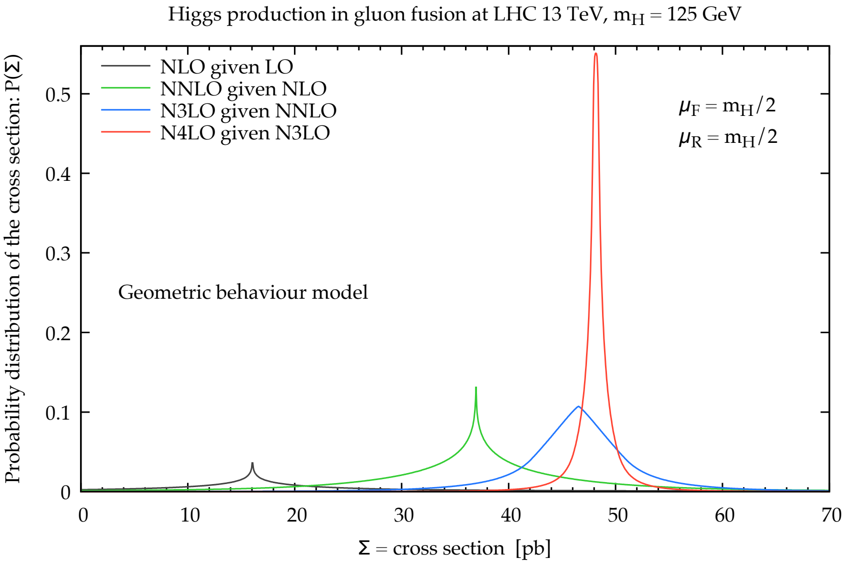
  depends on one hidden parameter λ, canonical scale variation is approximately recovered for $\lambda = \log 2$

- others...

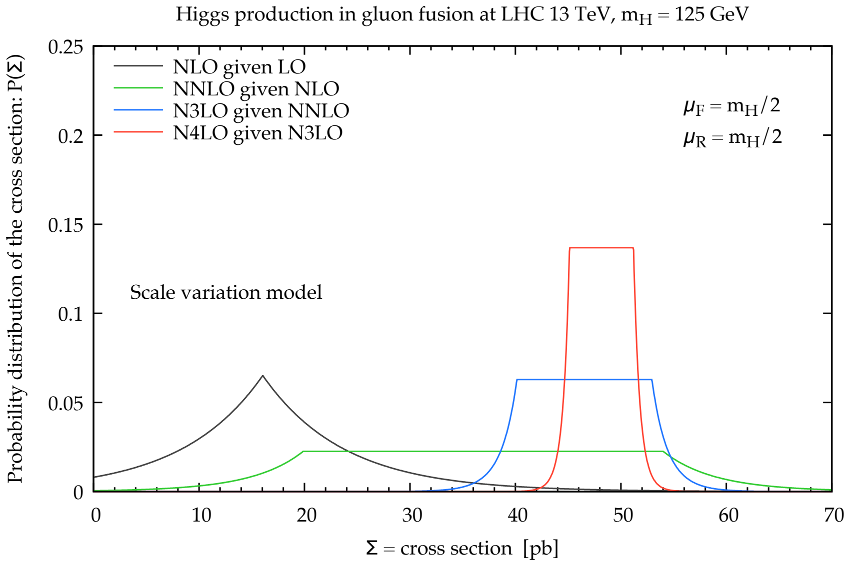
Higgs in gluon fusion at LHC: probability distributions



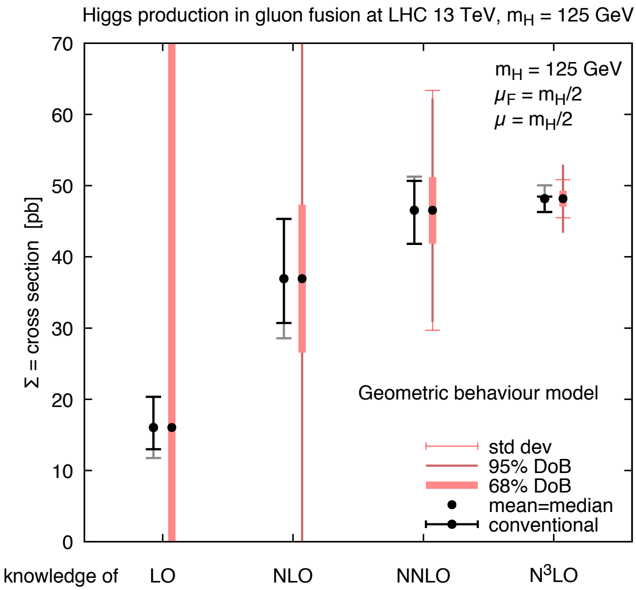
Higgs in gluon fusion at LHC: probability distributions



Higgs in gluon fusion at LHC: probability distributions

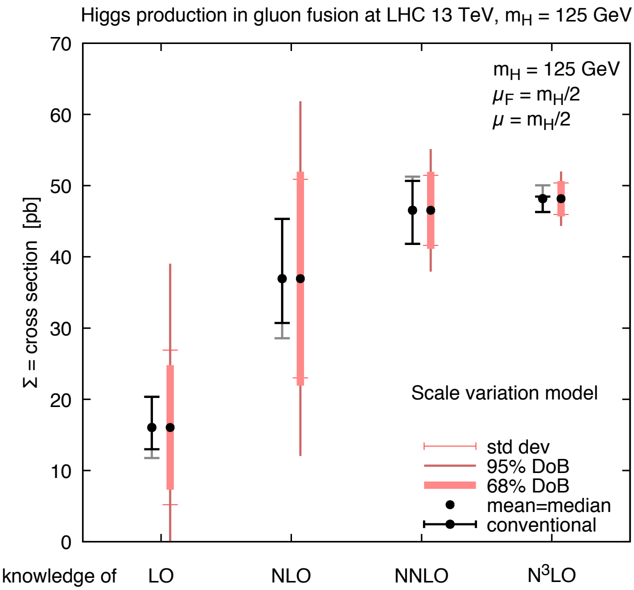


From distributions to statistical estimators



conventional result: canonical scale variation

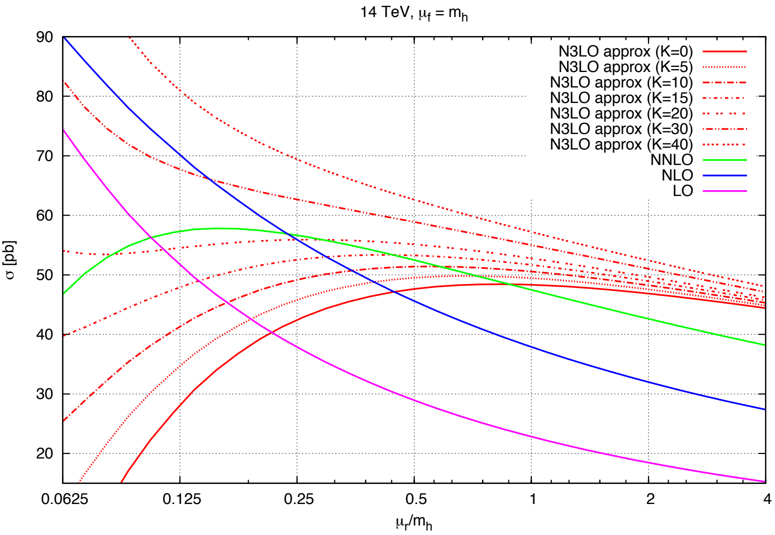
new result: geometric behaviour model



conventional result:
canonical scale
variation

new result:
scale variation
inspired model

Not all higher orders are good...



[Buehler,Lazopoulos 1306.2223]

Another way of using scale dependence as a tool

Because $r_k(\mu) = \mathcal{O}(\alpha_s^{k+1})$, they should also behave perturbatively

Idea: require perturbativity of the $r_k(\mu)$ as a model condition!

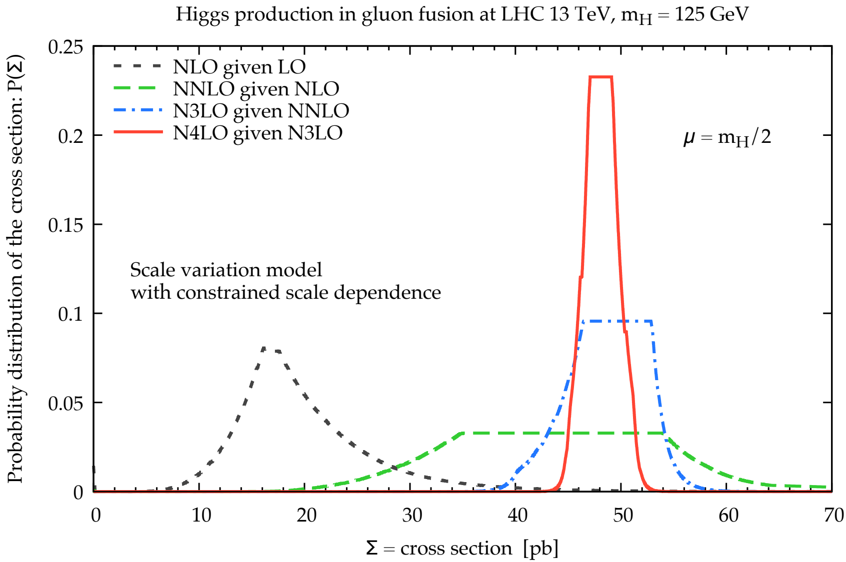
Two conditions:

$$|\delta_{k+1}(\mu)| \leq \lambda r_k(\mu)$$
$$|r_{k+1}(\mu)| \leq \eta r_k(\mu)$$

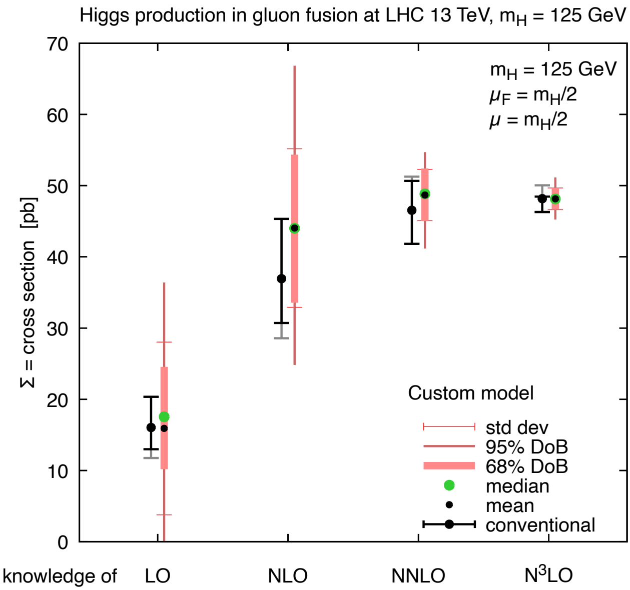
that depends on two hidden parameters λ, η

Leads to more stable and narrower results
(but the implementation is numerical, hence slow)

Higgs in gluon fusion at LHC: probability distributions



From distributions to statistical estimators



conventional result: canonical scale variation

new result:
scale variation inspired model with contraints on higher order scale dependence

Scale dependence as an issue

The results presented so far depend on the scale μ: if I change the scale, the result changes

But any scale is in principle acceptable, so what can I do?

Two options:

- either I have a way to select an "optimal" scale ✗
- or I need to combine in some way the results at different scales ✓

First option is simpler, provided such a criterion exists

There are various proposal in the literature: BLM, PMS, PMC, POEM, ...
PMC (principle of maximal conformality) is the most widespread, and authors claim it leads to basically zero scale ambiguity in the final prediction
However, this conclusion has been criticized, and the ambiguity of the PMC method is likely comparable to the canonical scale ambiguity [Kataev,Mikhailov 1408.0122]

[Kataev,Mikhailov 1607.08698] [Chawdhry,Mitov 1907.06610]

We go for the second option!

Constructing a "scale-independent" result

Basic idea: treat the unphysical scale μ as a parameter of the model, and simply marginalize over it

$$P(\Sigma|\delta_0, ..., \delta_n) = \int d\mu \; P(\Sigma|\delta_0, ..., \delta_n, \mu) \, P(\mu|\delta_0, ..., \delta_n)$$

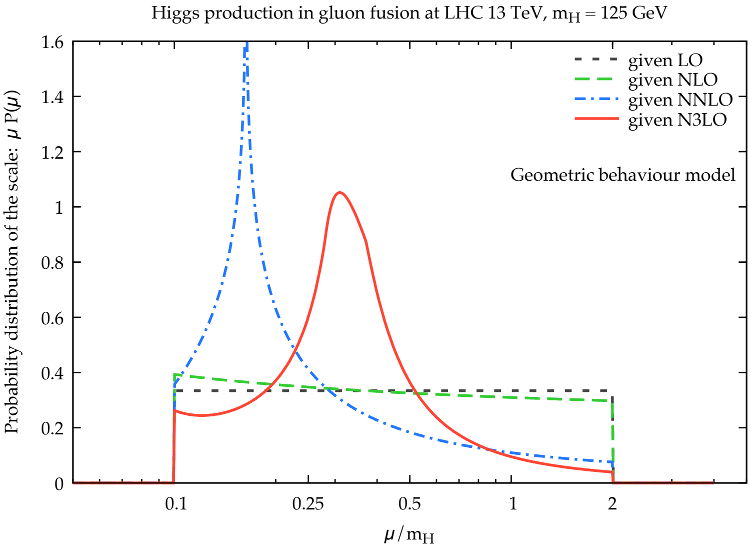
where $P(\mu|\delta_0, ..., \delta_n)$ is the posterior distribution for μ given the known orders (which depends on the model)

The prior $P_0(\mu)$ contains our prejudices on what are the most appropriate scales, but the results are largely independent of the precise form and size of the prior
$\Rightarrow$ **a lot of arbitrariness is removed!**

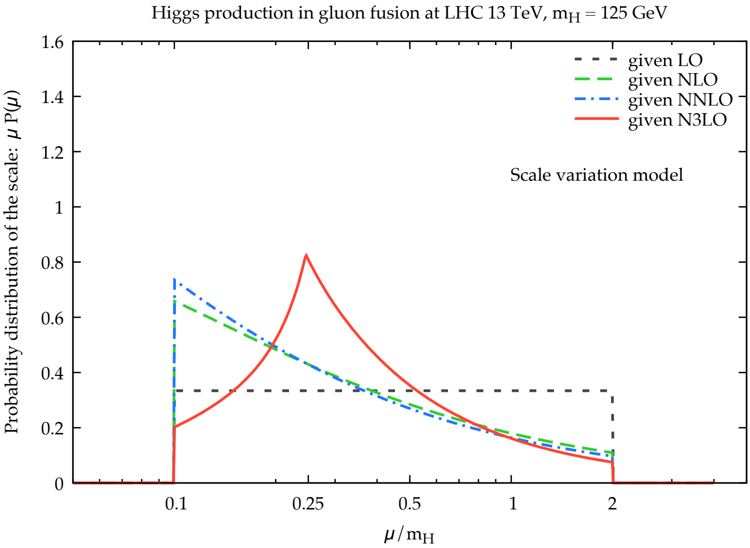
In this approach, inference on μ selects the values that give the best convergence properties *according to the model*

In [Duhr,Huss,Mazeliauskas,Szafron 2106.04585] they propose an alternative way denoted "scale averaging". Rather than treating the scale as a model parameter, they integrate over it using a weight function $w(\mu)$, so there is no inference on μ. I personally find it less powerful.

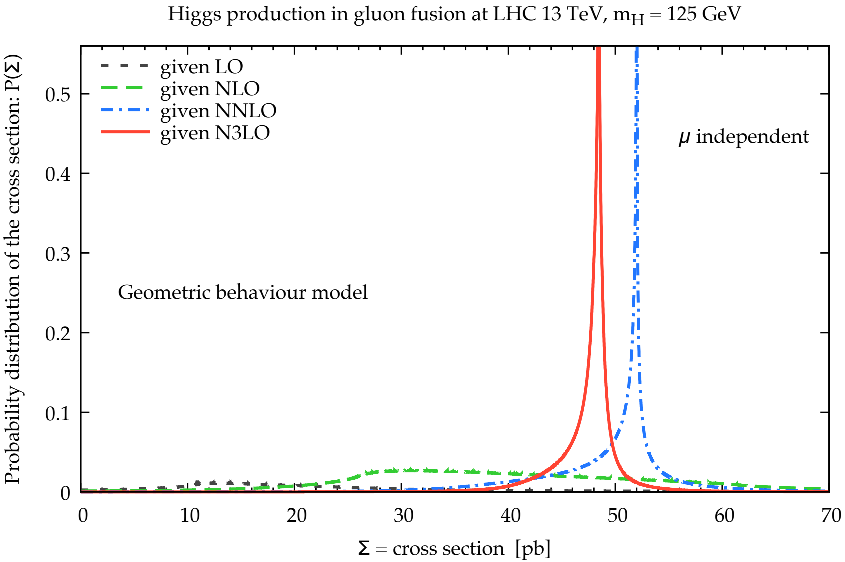
Posterior distribution for the scale μ



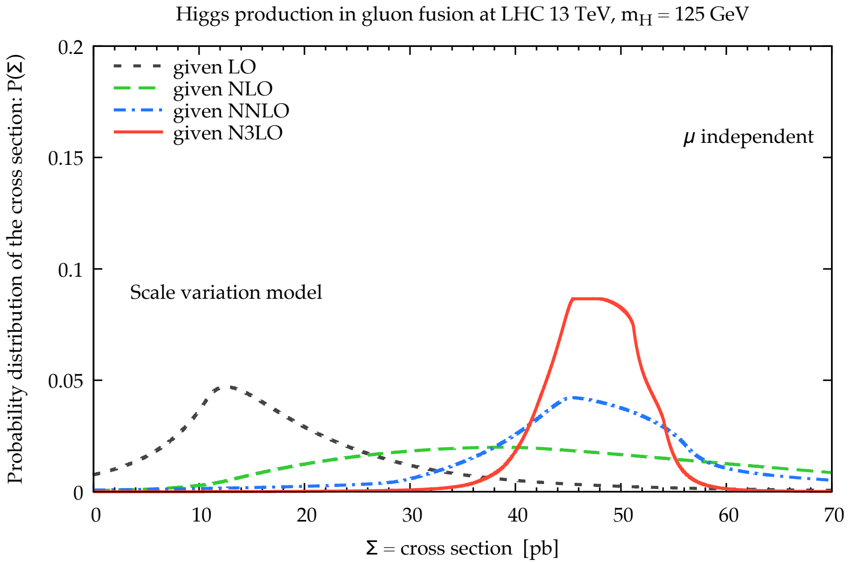
Posterior distribution for the scale μ



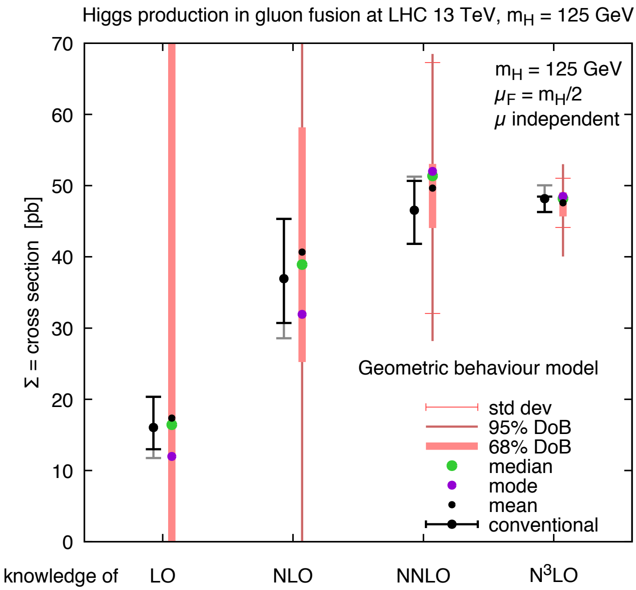
Higgs in gluon fusion at LHC: scale independent distributions



Higgs in gluon fusion at LHC: scale independent distributions



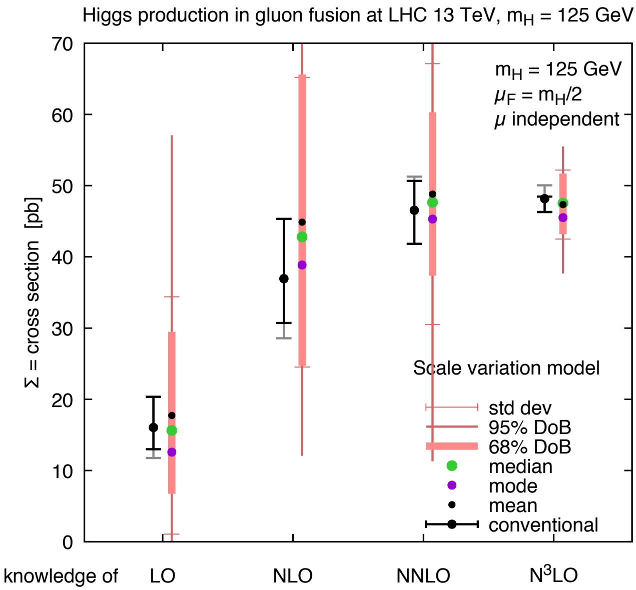
Higgs in gluon fusion at LHC: final results



conventional result: canonical scale variation by a factor of 2 about $\mu_R = m_H/2$ (best convergence properties)

new result: geometric behaviour model

Higgs in gluon fusion at LHC: final results



conventional result: canonical scale variation by a factor of 2 about $\mu_R = m_H/2$ (best convergence properties)

new result: scale variation inspired model

Correlations

The last step for a PDF fit is to account for correlations:

- between different bins of the same observable
- between different observables of the same process
- between different processes

No unique way to do so

A simple way is to use the hidden parameters, including the scale μ, to correlate the predictions

However, better (but more complicated) ways can be considered (see e.g. an interesting proposal by F.Tackmann [SCET2019])

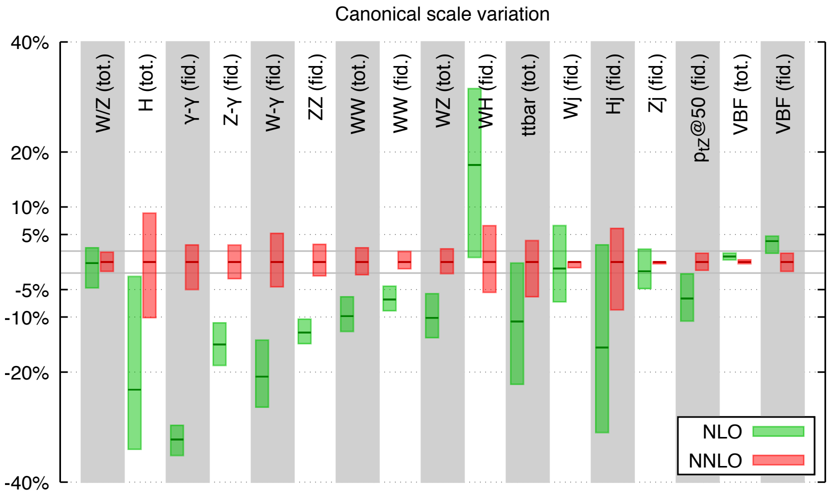
I'm happy to start collaborations with the PDF fitting community on this!

Summary

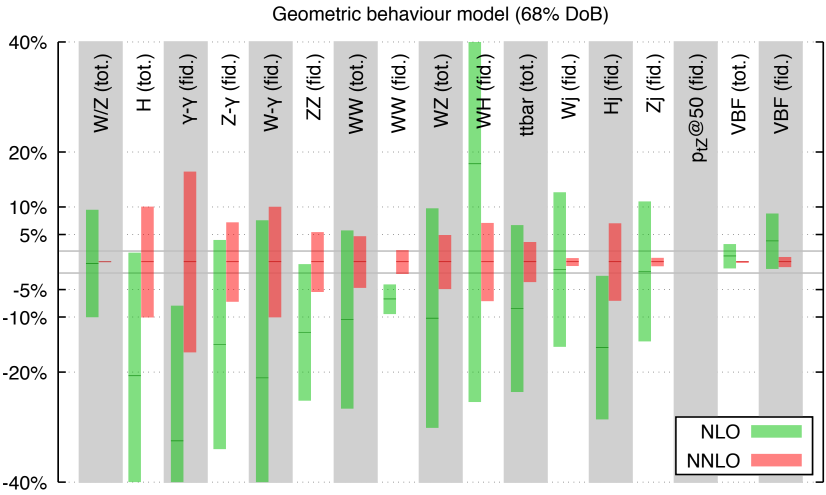
Key message: it is possible to define theory uncertainties from MHO in a probabilistic way, which is reliable and less arbitrary than the canonical scale-variation approach

- **New statistical models for theory uncertainties**:
  - an improved version of Cacciari-Houdeau (geometric behaviour model)
  - a model inspired by scale variation
  - other variants and combinations
- A novel way to obtain **scale-independent results**
- **Public code**: `THunc` `www.roma1.infn.it/~bonvini/THunc`
  see also my "competitors" code `MiHO`: `github.com/aykhuss/miho`
- **Correlations**
  - various ideas, to be discussed, implemented, and tested

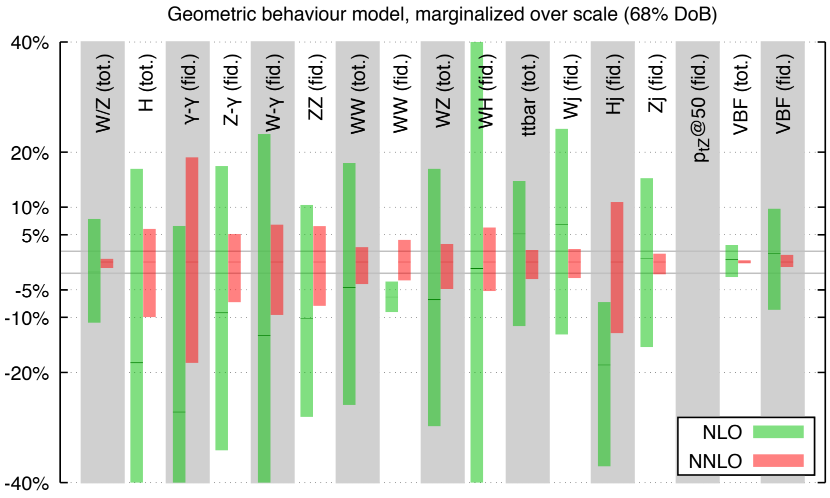
Gavin Salam's plot



Gavin Salam's plot



Gavin Salam's plot



Backup slides

How can a theory uncertainty from missing higher orders be probabilistic?

Frequentist approach to probability $\rightarrow$ requires repeatable events $\rightarrow$ no way...

Bayesian approach $\rightarrow$ probability defined as the **degree of belief** of an "event"

Initially no information $\rightarrow$ the probability of an event is given by a *prior* distribution, which encodes our subjective and arbitrary prejudices.

Acquiring information $\rightarrow$ changes the degree of belief through inference (Bayes theorem), making it less and less dependent on the prior.

see e.g. G.D'Agostini, Bayesian reasoning in data analysis

"Event" means something that can happen in different ways with different likelihoods.
In our case, the "event" is *"the observable takes the value* Σ*"*, and its probability distribution will be a function of Σ:

$$P(\Sigma|\text{information}, \text{hypotheses})$$

Information = perturbative expansion of the observable.
Bayes theorem $\rightarrow$ improve the knowledge on the observable, namely update the distribution of Σ.

Generalized condition that accounts for a possible power growth

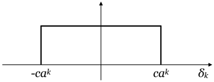
$$|\delta_k(\mu)| \le c a^k \qquad \forall k < k_{\text{asympt}} \qquad\qquad \text{CH: } \left|c_k \alpha_s^k\right| \le \bar{c}\alpha_s^k$$

depends on two hidden parameters c, a

It accounts for a possible power growth of the coefficients within the model!

Likelihood:

$$P(\delta_k | c, a, \mu) \propto \theta(c a^k - |\delta_k(\mu)|) =$$



namely all values of δ_k within the allowed range are equally likely

Prior:

$$P(c, a|\mu) \propto \frac{\theta(c-1)}{c^{1+\epsilon}} \times (1-a)^\omega \theta(a)\theta(1-a), \qquad \epsilon = 0.1, \quad \omega = 1$$

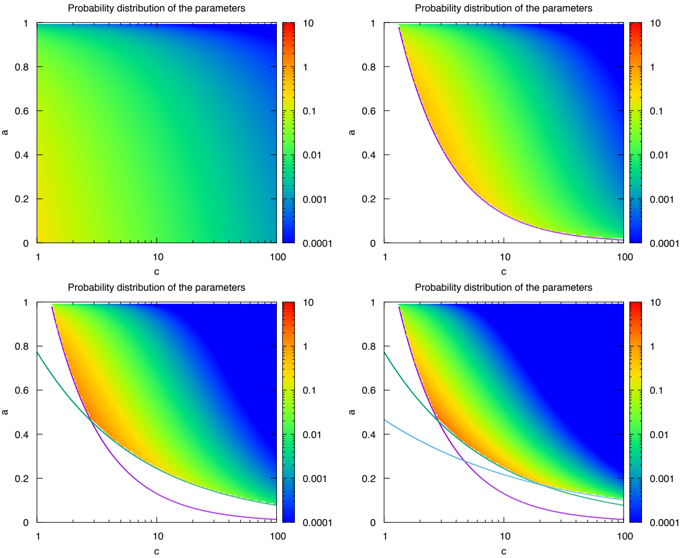
Inference scheme:

$$\underbrace{\delta_0, \ldots, \delta_n}_{\text{known}} \xrightarrow{\text{inference}} c, a \xrightarrow{\text{inference}} \underbrace{\delta_{n+1}, \delta_{n+2}, \ldots}_{\text{unknown}} \xrightarrow{\text{sum}} \Sigma$$

Final output:

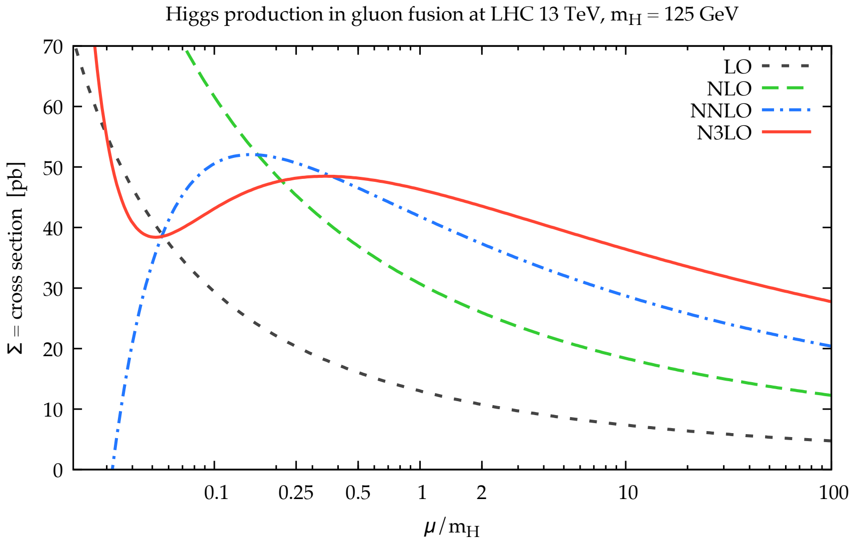
$$P(\Sigma | \delta_0, \ldots, \delta_n, \mu, \text{model}_1)$$

Posterior of c, a for Higgs production in gluon fusion



$$\delta_0 + \delta_1 + \delta_2 + \delta_3 + \dots = 1 + 1.36 + 0.85 + 0.35 + \dots$$

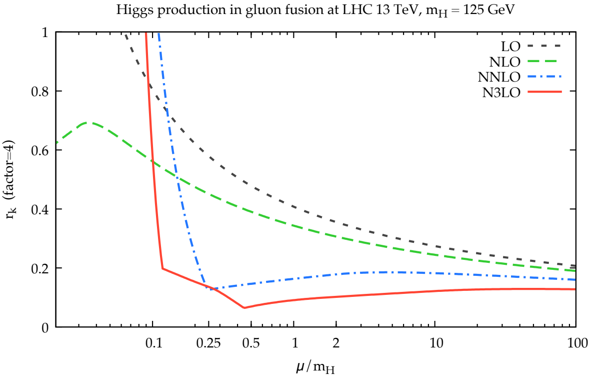
Higgs production in gluon fusion at LHC



Defining a good scale-dependence estimator

I want to define a model that uses scale variation.
I need a dimensionless number (to be compared to δ_k) that probes higher orders:

$$r_k(\mu) \simeq \left| \mu \frac{d}{d\mu} \log \Sigma_{\mathrm{N}^k\mathrm{LO}}(\mu) \right| = \mathcal{O}(\alpha_s^{k+1}) = \mathcal{O}(\delta_{k+1}(\mu))$$



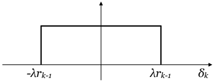
I propose the condition

$$|\delta_{k+1}(\mu)| \leq \lambda r_k(\mu) \qquad \forall k < k_{\text{asympt}}$$

that depends on one hidden parameter λ
Canonical scale variation is approximately recovered for $\lambda = \log 2$

Likelihood:

$$P(\delta_k | r_{k-1}, \lambda, \mu) \propto \theta(\lambda r_{k-1} - |\delta_k(\mu)|) =$$



namely all values of δ_k within the allowed range are equally likely

Prior:

$$P(\lambda|\mu) \propto \lambda^\gamma e^{-\lambda} \theta(\lambda), \qquad \gamma = 1$$

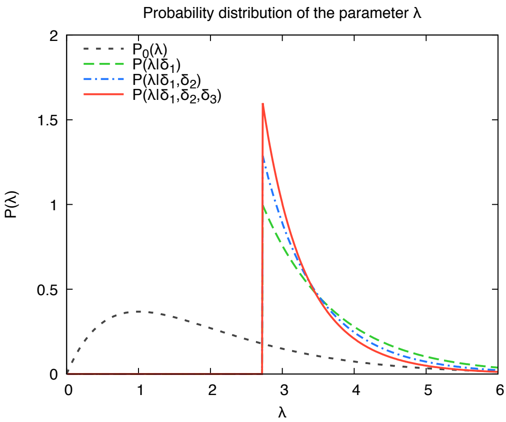
Inference scheme:

$$\underbrace{\delta_0, \ldots, \delta_n, r_0, \ldots, r_{n-1}}_{\text{known}} \xrightarrow{\text{inference}} \lambda \xrightarrow{\text{inference}+r_n} \underbrace{\delta_{n+1}}_{\text{unknown}} \xrightarrow{\text{sum}} \Sigma_{\text{N}^{n+1}\text{LO}}$$

in this case only the first missing higher order can be predicted:

$$P(\Sigma_{\text{N}^{n+1}\text{LO}} | \delta_0, \ldots, \delta_n, r_0, \ldots, r_n, \mu, \text{model}_2)$$

Posterior of λ for Higgs production in gluon fusion



The first non-trivial order ($\boldsymbol{\delta_1}$) sets the lower limit of $\boldsymbol{\lambda}$

$\rightarrow$ stable but possibly non optimal (overestimating uncertainty)

Improvable allowing violation of the bound (see appendix B.3)

Combining models and inventing new ones

Models can be combined together, requiring two or more conditions at the same time

So far we have seen three conditions

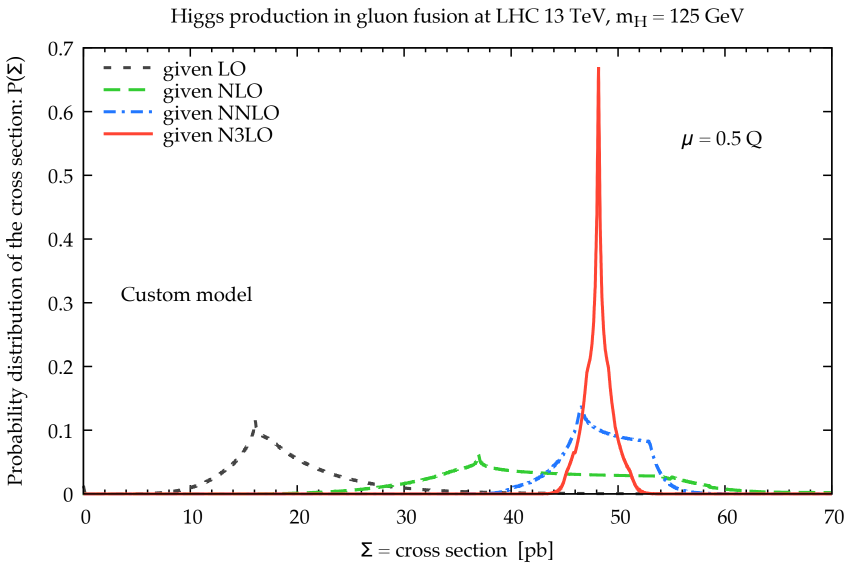
$$
\begin{aligned}
|\delta_k(\mu)| &\leq c a^k \\
|\delta_k(\mu)| &\leq \lambda r_{k-1}(\mu) \\
|r_k(\mu)| &\leq \eta r_{k-1}(\mu)
\end{aligned}
$$

that are not contradictory and can thus hold at the same time

The models are implemented in a code named `THunc`, that provides a *custom model* feature to implement any customized model

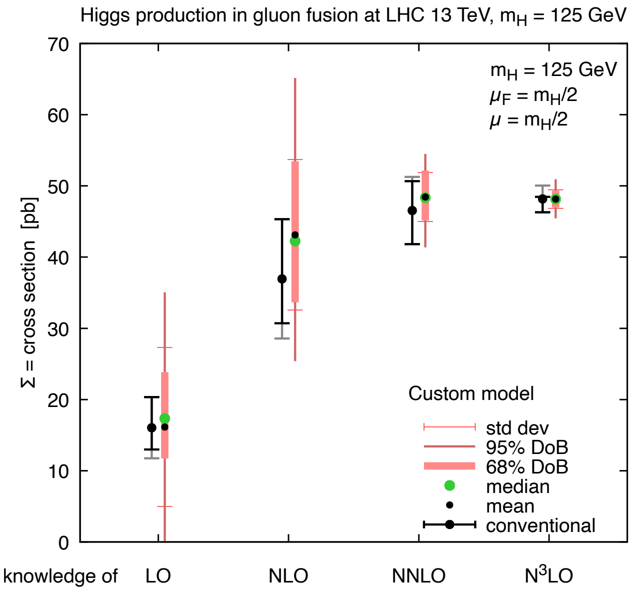
Putting all conditions together....

Higgs in gluon fusion at LHC: probability distributions



go to slide **??**

From distributions to statistical estimators



It's a generalisation of the geometric behaviour model,

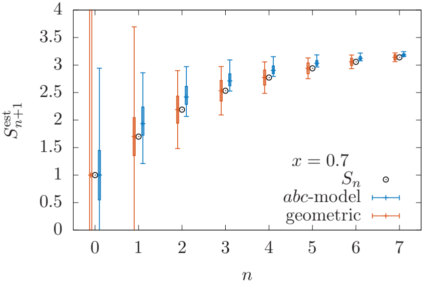
$$\text{geo:} \quad |\delta_k(\mu)| \le c a^k \qquad\qquad abc: \quad -c + b \le \frac{\delta_k(\mu)}{a^k} \le c + b$$

depends on three hidden parameters a, b, c

They keep requiring $|a| \le 1$, but the sign can be negative (to describe alternating sign series)

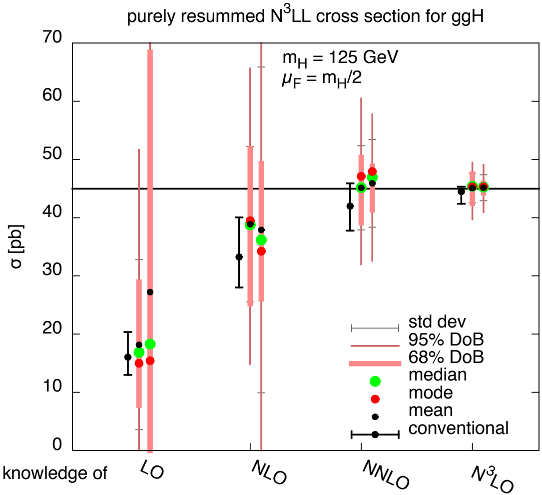
Moreover the b parameter accounts for asymmetric behaviour

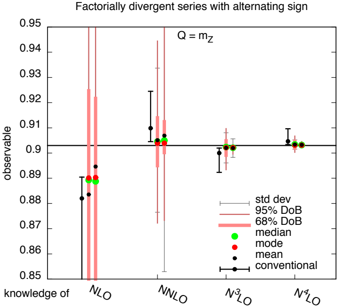
Comparison for $\sum_{k\ge 0} x^k$, $x = 0.7$

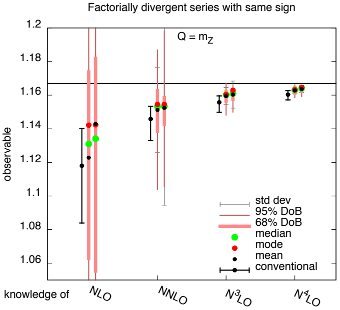


Note: I have proposed a different way to account for a sign pattern, which can be applied to any symmetric model (app. B.5)

Validation using known sums

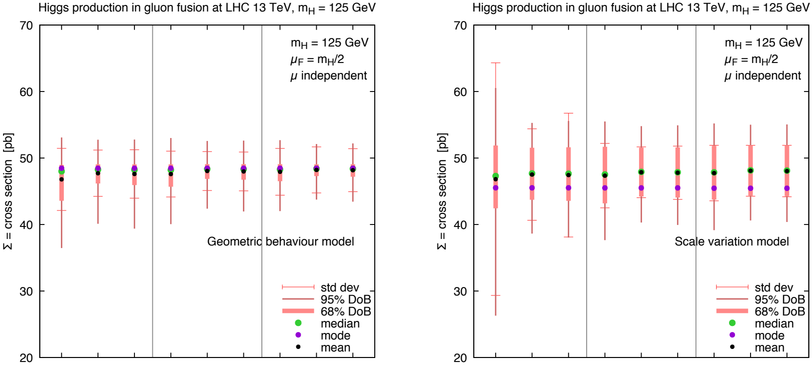




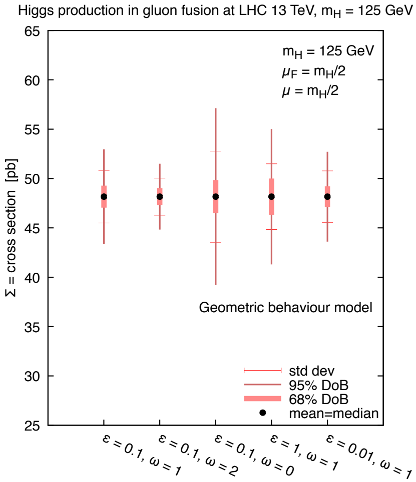


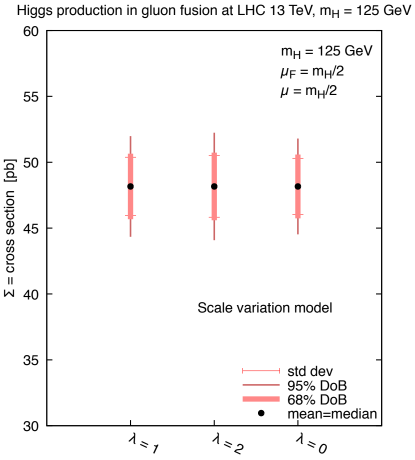
- purely resummed ggH at N^{3}LL, expanded in α_s
- factorially divergent series $\sum_k (-1)^k k! \alpha_s^k(m_Z)$
- factorially divergent series $\sum_k k! \alpha_s^k(m_Z)$

Scan of priors for the scale μ



Scan of priors for the model parameters





Explicit inference procedure in Cacciari-Houdeau

Probability of a missing higher order coefficient c_k given the knowledge of the first $c_0, \dots, c_n$ orders

$$\begin{aligned} P(c_k|c_0, \dots, c_n) &= \frac{P(c_k, c_0, \dots, c_n)}{P(c_0, \dots, c_n)} \qquad\qquad (k > n) \\ &= \frac{\int d\bar{c}\, P(c_k, c_0, \dots, c_n, \bar{c})}{\int d\bar{c}\, P(c_0, \dots, c_n, \bar{c})} \\ &= \frac{\int d\bar{c}\, P(c_k, c_0, \dots, c_n|\bar{c}) P_0(\bar{c})}{\int d\bar{c}\, P(c_0, \dots, c_n|\bar{c}) P_0(\bar{c})} \\ &= \frac{\int d\bar{c}\, P(c_k|\bar{c}) P(c_0|\bar{c}) \cdots P(c_n|\bar{c}) P_0(\bar{c})}{\int d\bar{c}\, P(c_0|\bar{c}) \cdots P(c_n|\bar{c}) P_0(\bar{c})} \end{aligned}$$

having used

$$P(A, B) = P(A|B)P(B), \qquad P(A) = \int dB\, P(A, B)$$

The probability for the full observable is given by

$$P(\Sigma|c_0, \dots, c_n) = \int dc_{n+1} dc_{n+2} \cdots P(c_{n+1}, c_{n+2}, \dots|c_0, \dots, c_n)\, \delta\left(\Sigma - \sum_{k=0}^{\infty} c_k \alpha_s^k\right)$$