Probabilistic theory uncertainty from missing higher orders

Marco Bonvini

INFN, Rome 1 unit

IRN Terascale, LNF, 16 April 2024

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

Istituto Nazionale di Fisica Nucleare
Sezione di ROMA

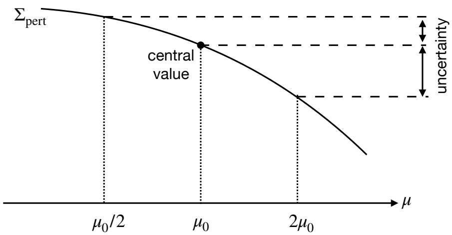
Unphysical (renormalization) scale dependence in perturbative computations is of higher order

$$\Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu) = \sum_{k=0}^{n} c_k(\mu)\alpha_s^k(\mu) \qquad \mu\frac{d}{d\mu}\Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu) = \mathcal{O}(\alpha_s^{n+1})$$

Canonical Scale Variation:

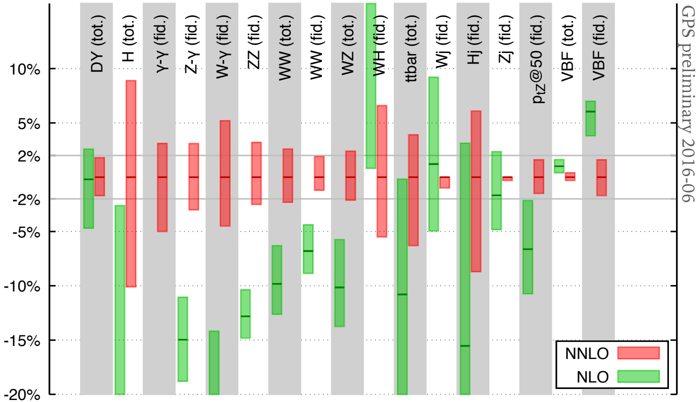
Variation by a factor of 2 about a "central" scale μ_0

$$\Sigma \approx \Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu_0) \pm \max_{\mu_0/2\leq\mu\leq2\mu_0} \left|\Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu) - \Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu_0)\right|$$



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**

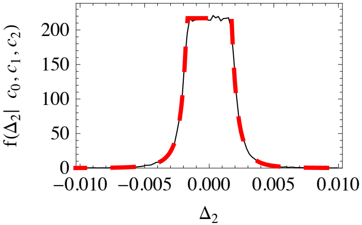
Cacciari and Houdeau [1105.5152] proposed a probabilistic model for the interpretation of theory uncertainties, based on the behaviour of the perturbative expansion

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

"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."

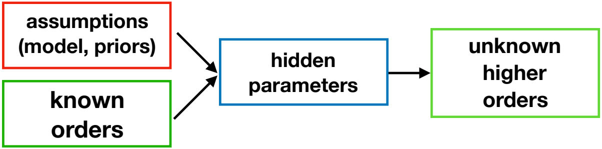
In other words, the model assumes that

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



Bayesian inference

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

- 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

New model (1): Geometric behaviour model (aka improved CH)

More general expansion

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

CH model assumes that δ_k behave as α_s^k

$$\Sigma_{\text{LO}}(\mu)\,|\delta_k(\mu)| \leq \bar{c}\,\alpha_s^k$$

Power growth of the coefficients $c_k \sim \eta^k$ is very likely:

- Cacciari-Houdeau proposed a modified version with η accounted for
- in [Bagnaschi,Cacciari,Guffanti,Jenniches 1409.5036] η is determined from a survey
- in an alternative approach [Forte,Isgrò,Vita 1312.6688] the value of η is fitted

---

My proposal: geometric behaviour model

$$|\delta_k(\mu)| \leq c\,a^k$$

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]

Constructing a "scale-independent" result

The method just described still needs to chose a renormalization scale μ: if I change the scale, the result changes. How can we get rid of the scale?

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

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

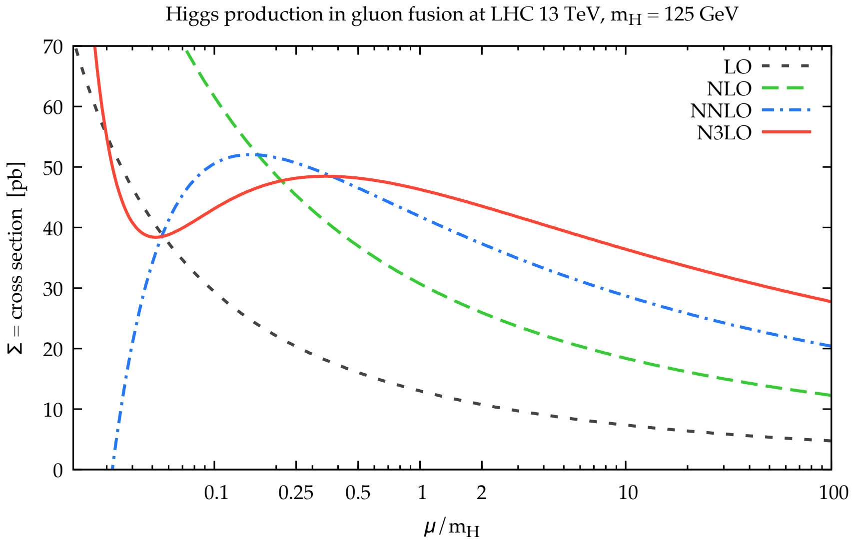
where $P(\mu|\delta_0, ..., \delta_n)$ is the posterior distribution for μ given the known orders (which depends on the model)

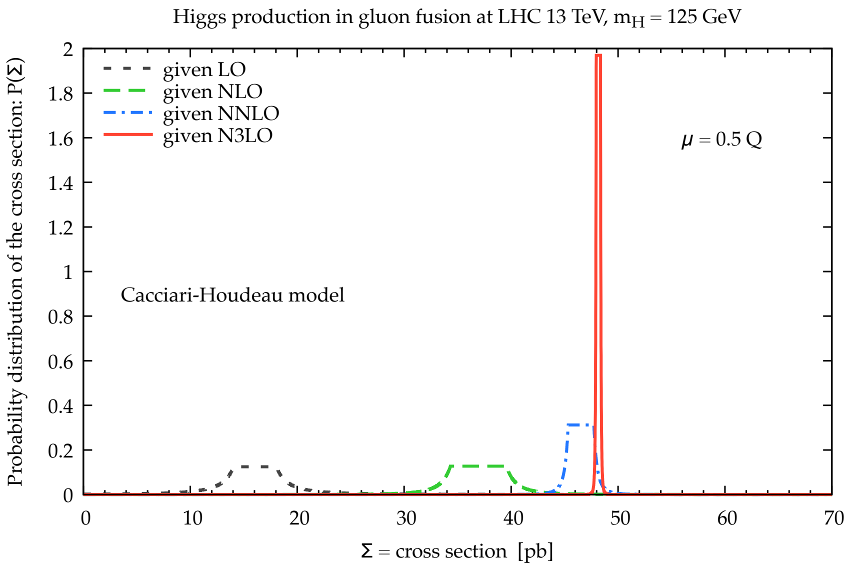
The prior $P_0(\mu)$ contains our prejudices on what are the most appropriate scales, but the results are largely independent of the precise form and size of the prior
$\Rightarrow$ **a lot of arbitrariness is removed!**

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.

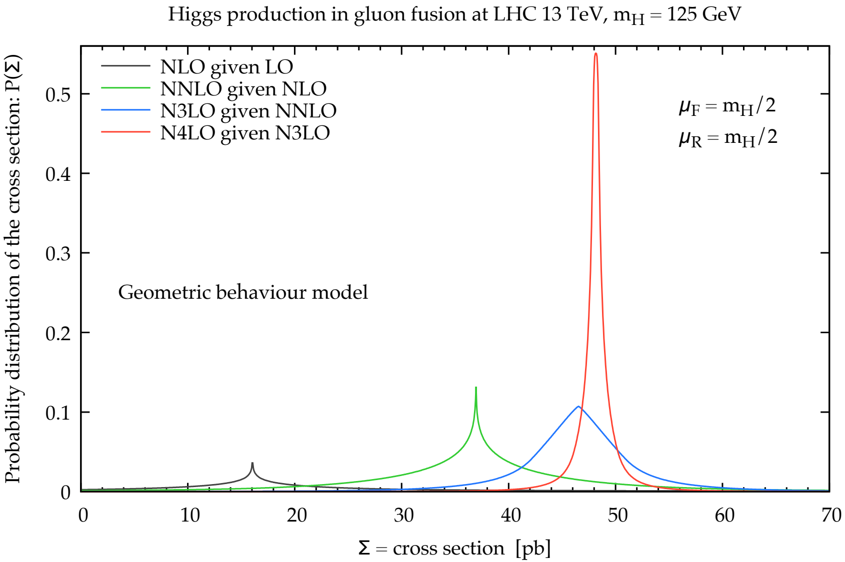
A case study: Higgs production in gluon fusion (ggH) at LHC



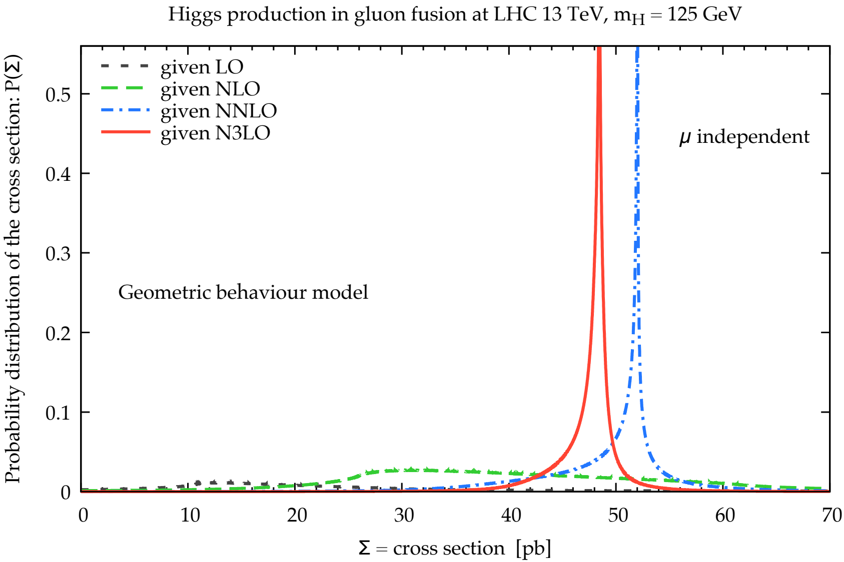


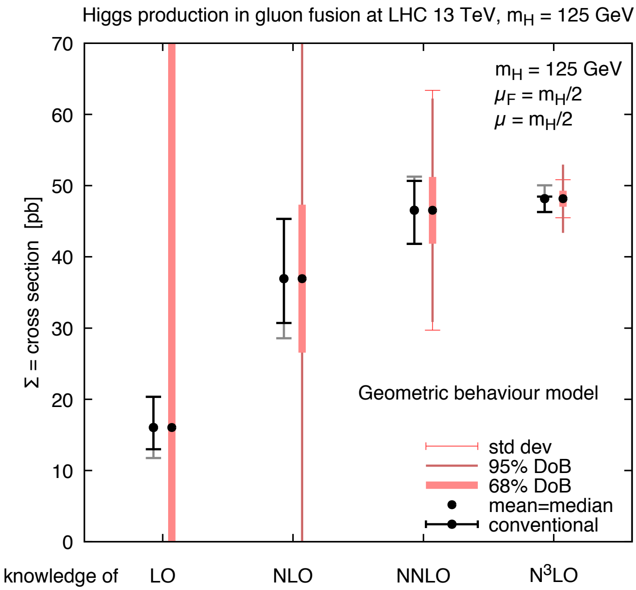
Higgs production in gluon fusion at LHC 13 TeV, m_H = 125 GeV
given LO
given NLO
given NNLO
given N3LO
$\mu = 0.5$ Q
Cacciari-Houdeau model
Probability distribution of the cross section: P(Σ)
Σ = cross section [pb]
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
10
20
30
40
50
60
70

ggH at LHC: probability distributions (geo)



ggH at LHC: probability distributions (geo, marginalized over μ)

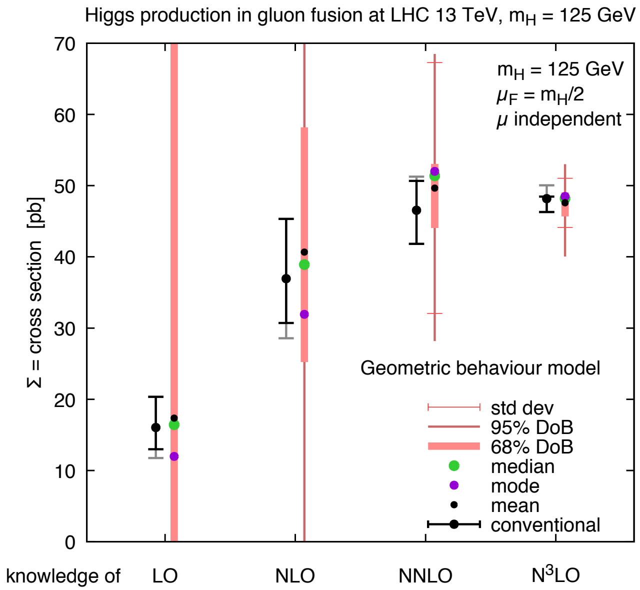




conventional result: canonical scale variation

new result: geometric behaviour model

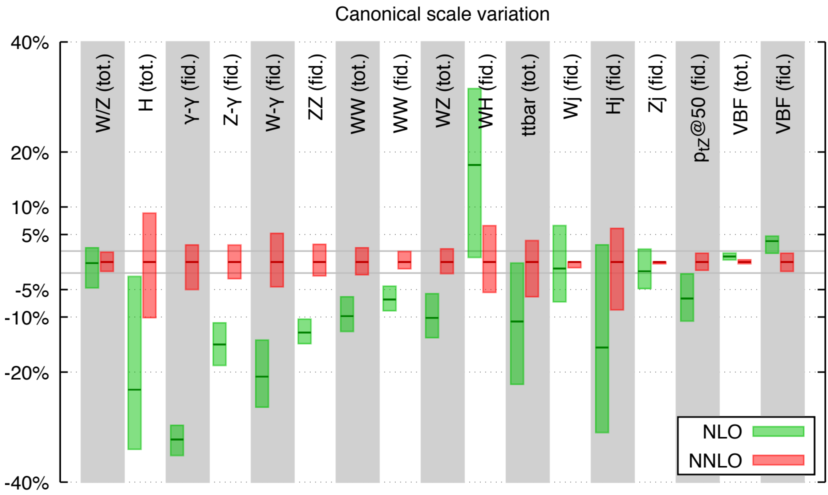
ggH at LHC: statistical estimators (geo, marginalized over μ)



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

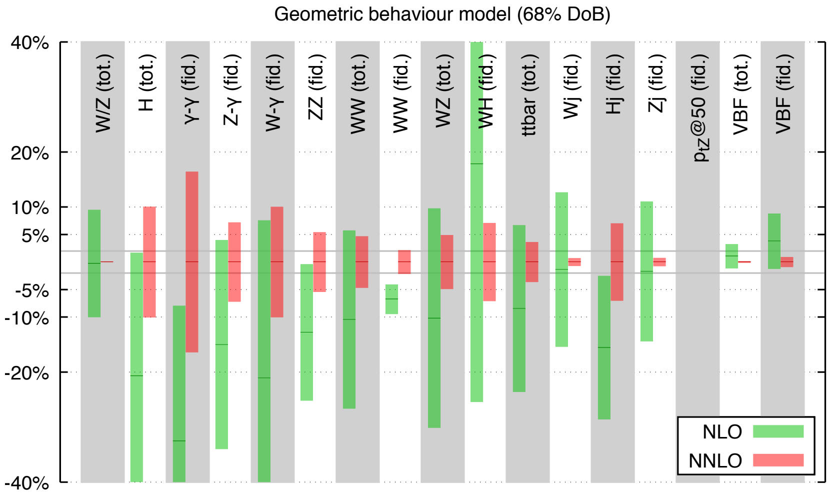
new result: geometric behaviour model

Canonical scale variation vs Geometric behaviour model

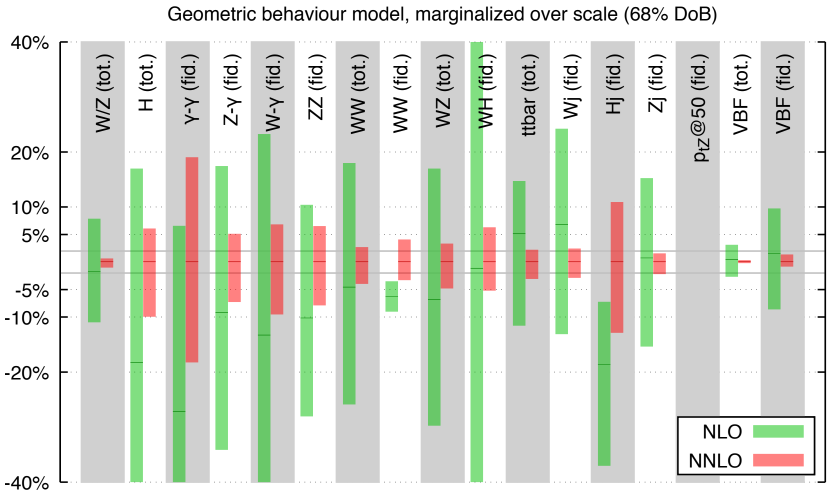


(courtesy of Gavin Salam)

Canonical scale variation vs Geometric behaviour model



Computed with THunc



Computed with THunc

Correlations

Still need 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

Crucial observation: correlations from MHO are due to similarities in the form of the perturbative expansions

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])

Thinking in progress...

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, better with constrained scale dependence
    (see Extra material)
  - other possible variants and combinations
    (see Backup)

- 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

Extra content

(if I have time)

New model (2): Scale variation inspired model

Scale dependence probes higher orders... why not using it?

Idea (inspired by canonical scale variation): assume that the size of the higher order is comparable with the size of the scale dependence

Definition: "scale dependence numbers" r_k

$$r_k(\mu) \simeq \left| \mu \frac{d}{d\mu} \log \Sigma_{\mathrm{N}^k\mathrm{LO}}(\mu) \right|$$

measure the scale dependence of Σ

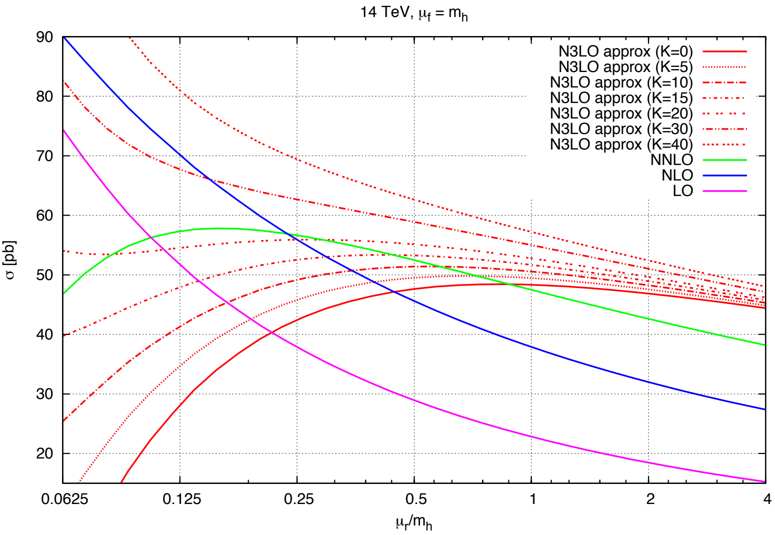
My proposal: scale variation model

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

depends on one hidden parameter λ

Canonical scale variation is approximately recovered for $\lambda = \log 2$

Not all higher orders are good...



[Buehler,Lazopoulos 1306.2223]

New model (3): Constrained scale dependence

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!

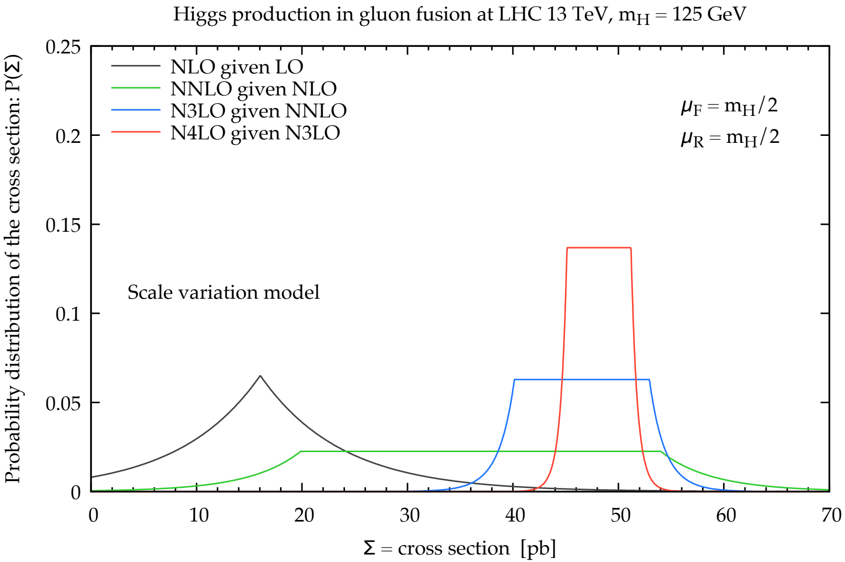
Two conditions:

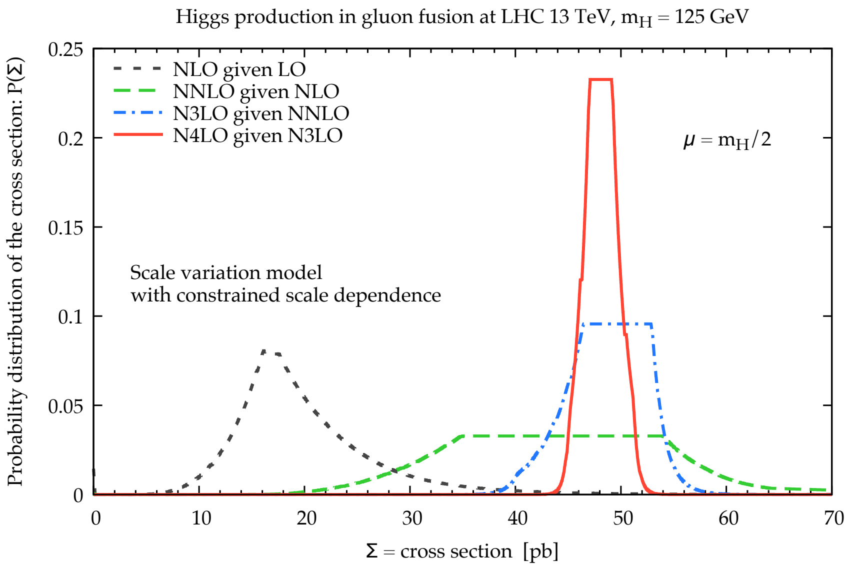
$$
\begin{aligned}
|\delta_{k+1}(\mu)| &\leq \lambda r_k(\mu) \\
|r_{k+1}(\mu)| &\leq \eta r_k(\mu)
\end{aligned}
$$

that depends on two hidden parameters λ, η

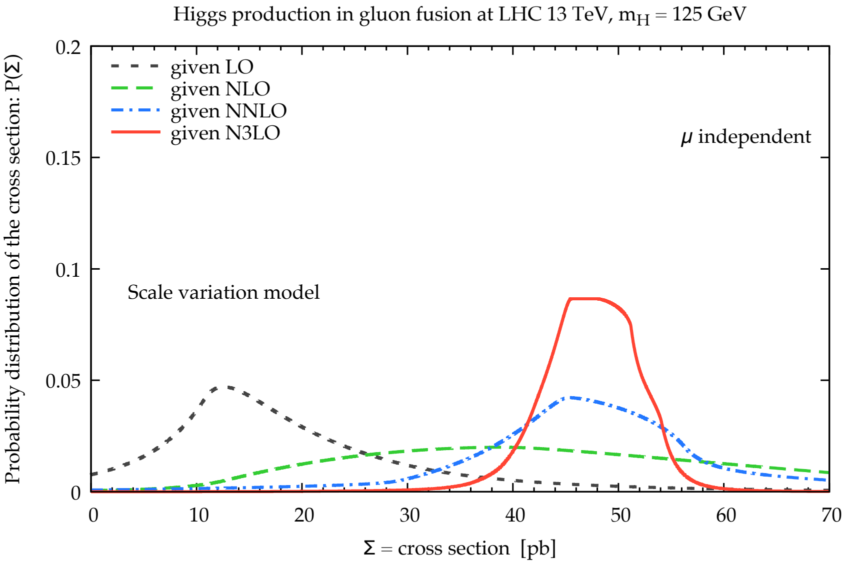
Leads to more stable and narrower results
(but the implementation is numerical, hence slow)

ggH at LHC: probability distributions (scale)

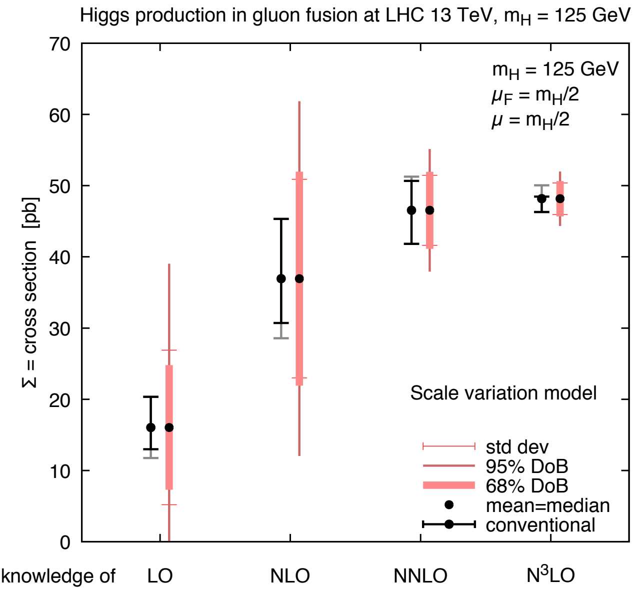



Higgs production in gluon fusion at LHC 13 TeV, $m_H = 125$ GeV
Probability distribution of the cross section: P(Σ)
NLO given LO
NNLO given NLO
N3LO given NNLO
N4LO given N3LO
$\mu = m_H/2$
Scale variation model
with constrained scale dependence
0.25
0.2
0.15
0.1
0.05
0
0
10
20
30
40
50
60
70
Σ = cross section [pb]

ggH at LHC: probability distributions (scale, marginalized over μ)

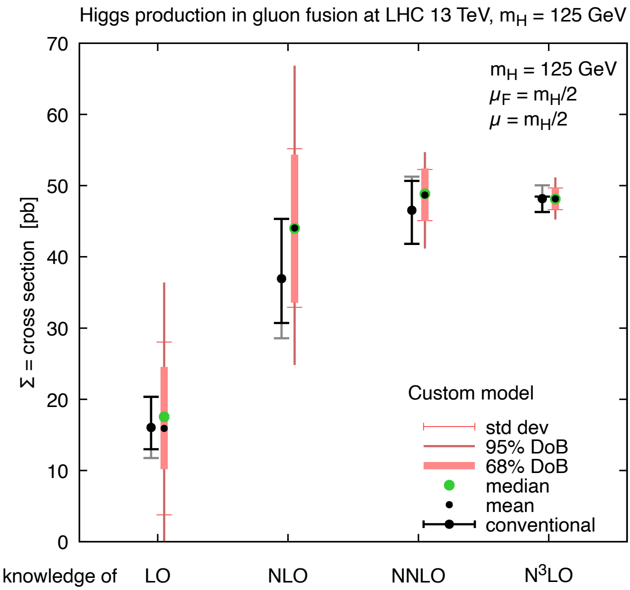


ggH at LHC: statistical estimators (scale)



conventional result:
canonical scale
variation

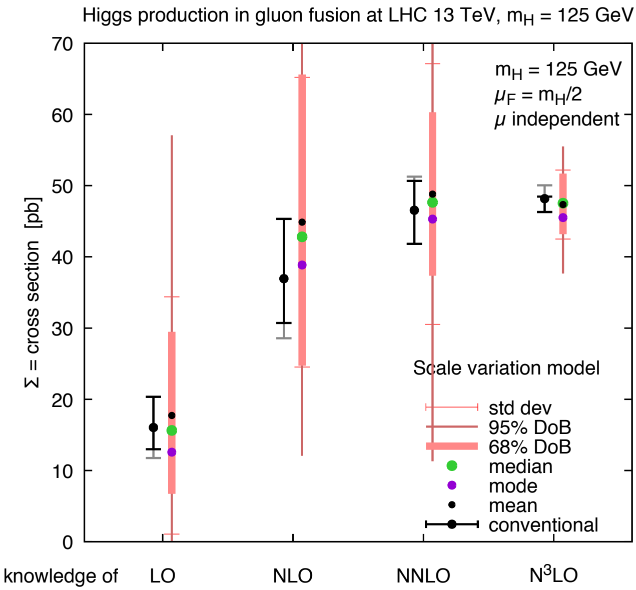
new result:
scale variation
inspired model



conventional result: canonical scale variation

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

ggH at LHC: statistical estimators (scale, marginalized over μ)



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

Backup slides

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

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

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

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

Acquiring information $\to$ 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* $\boldsymbol{\Sigma}$*"*, and its probability distribution will be a function of $\boldsymbol{\Sigma}$:

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

Information $=$ perturbative expansion of the observable.
Bayes theorem $\to$ improve the knowledge on the observable, namely update the distribution of $\boldsymbol{\Sigma}$.

Generalized condition that accounts for a possible power growth

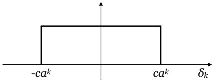
$$|\delta_k(\mu)| \leq c a^k \qquad \forall k < k_{\text{asympt}} \qquad\qquad \text{CH: } \left|c_k \alpha_s^k\right| \leq \bar{c}\alpha_s^k$$

depends on two hidden parameters c, a

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

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:

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

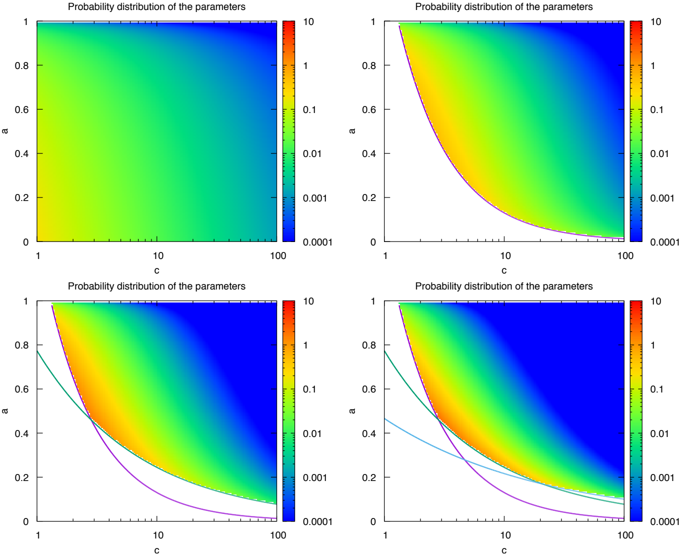
Inference scheme:

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

Final output:

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

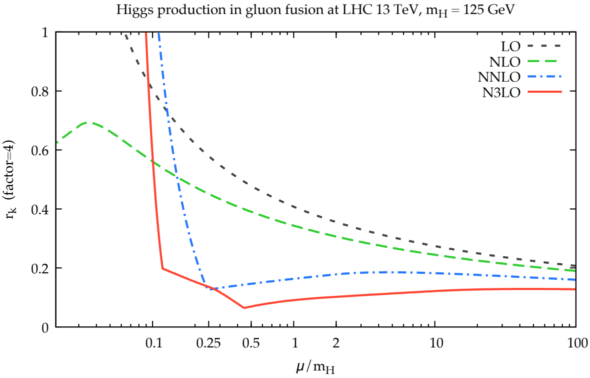
Posterior of c, a for Higgs production in gluon fusion



$$\boldsymbol{\delta_0 + \delta_1 + \delta_2 + \delta_3 + \ldots = 1 + 1.36 + 0.85 + 0.35 + \ldots}$$

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:

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



Model 2: Scale variation inspired model

I propose the condition

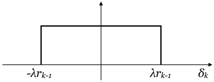
$$|\delta_{k+1}(\mu)| \le \lambda r_k(\mu) \qquad \forall k < k_{\text{asympt}}$$

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:

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

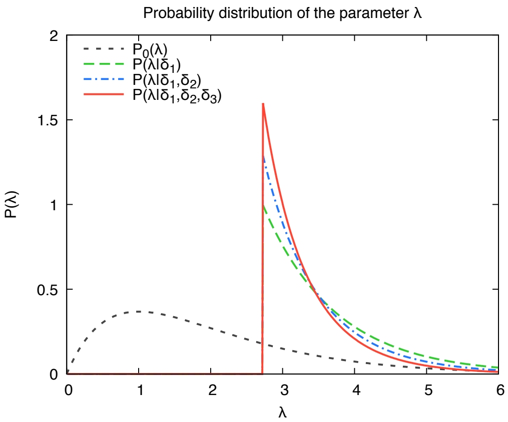
Inference scheme:

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

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

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

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

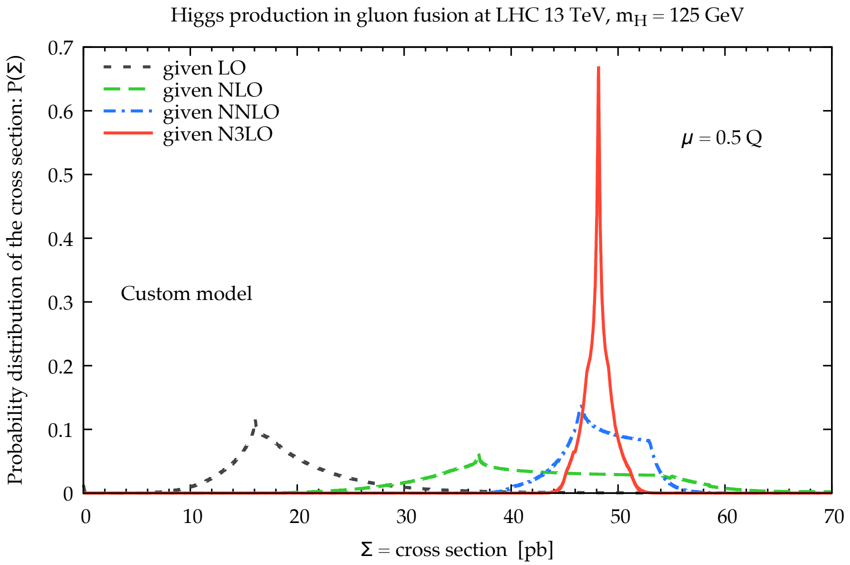
$$
\begin{aligned}
|\delta_k(\mu)| &\leq c a^k \\
|\delta_k(\mu)| &\leq \lambda r_{k-1}(\mu) \\
|r_k(\mu)| &\leq \eta r_{k-1}(\mu)
\end{aligned}
$$

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

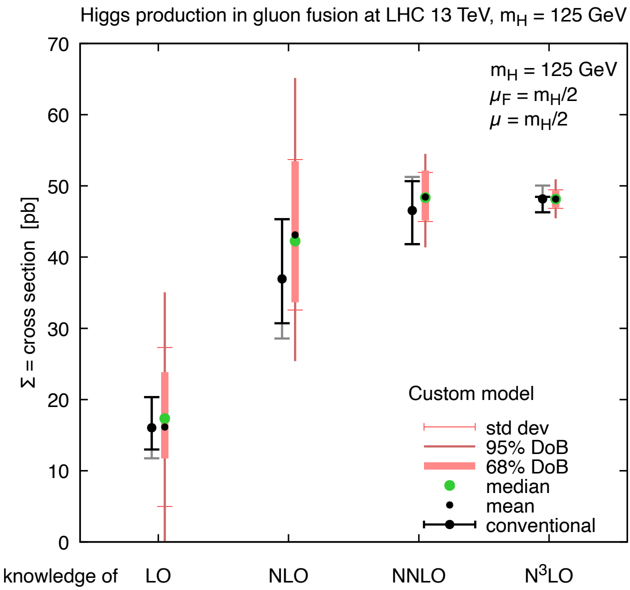
The models are implemented in a code named `THunc`, that provides a *custom model* feature to implement any customized model

Putting all conditions together....

Higgs in gluon fusion at LHC: probability distributions



From distributions to statistical estimators



It's a generalisation of the geometric behaviour model,

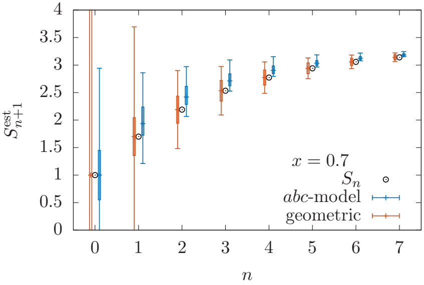
$$\text{geo:} \quad |\delta_k(\mu)| \le c a^k \qquad\qquad abc: \quad -c+b \le \frac{\delta_k(\mu)}{a^k} \le c+b$$

depends on three hidden parameters a, b, c

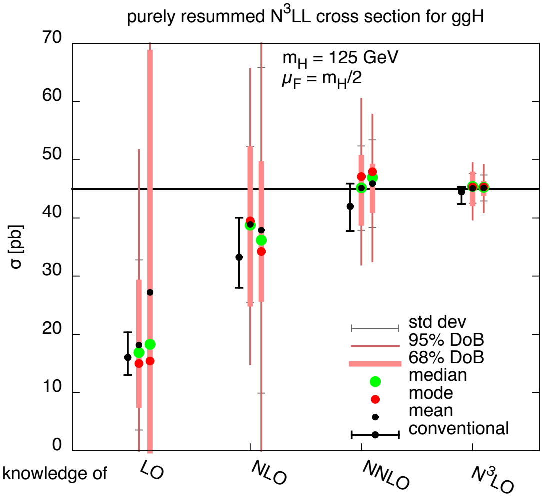
They keep requiring $|a| \le 1$, but the sign can be negative (to describe alternating sign series)

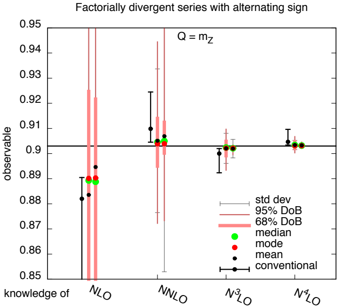
Moreover the b parameter accounts for asymmetric behaviour

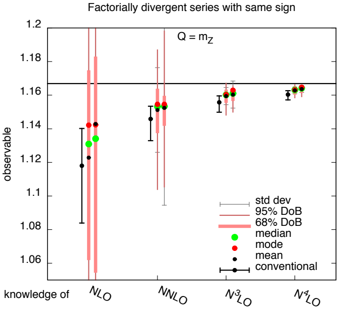
Comparison for $\sum_{k\ge 0} x^k$, $x = 0.7$



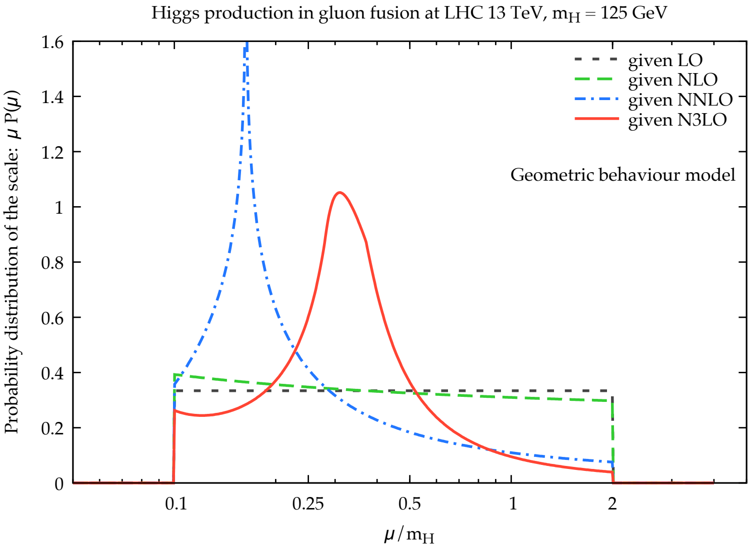
Note: I have proposed a different way to account for a sign pattern, which can be applied to any symmetric model (app. B.5)



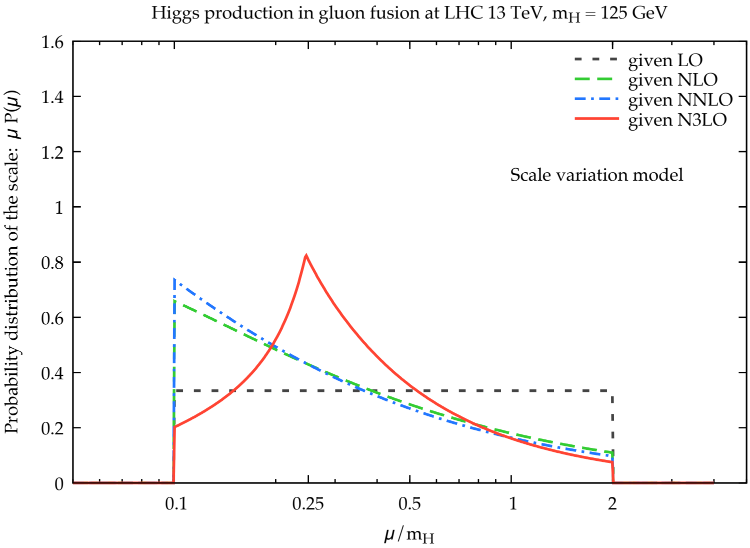




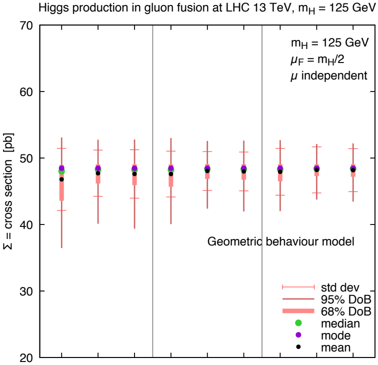
- purely resummed ggH at N^{3}LL, expanded in α_s
- factorially divergent series $\sum_k (-1)^k k! \alpha_s^k(m_Z)$
- factorially divergent series $\sum_k k! \alpha_s^k(m_Z)$

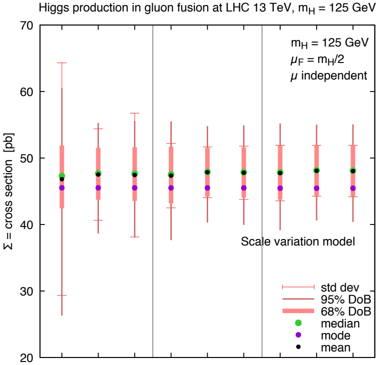

Higgs production in gluon fusion at LHC 13 TeV, $m_H = 125$ GeV
given LO
given NLO
given NNLO
given N3LO
Geometric behaviour model
Probability distribution of the scale: $\mu\, P(\mu)$
μ / m_H
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0.1
0.25
0.5
1
2

Posterior distribution for the scale μ

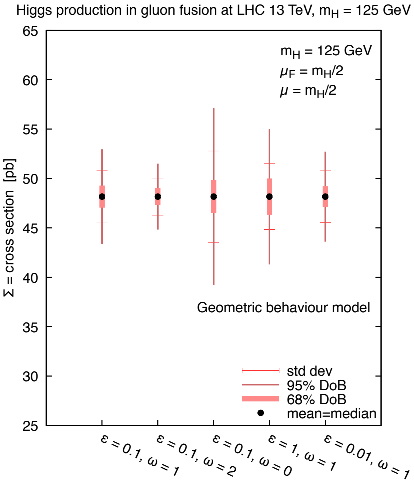


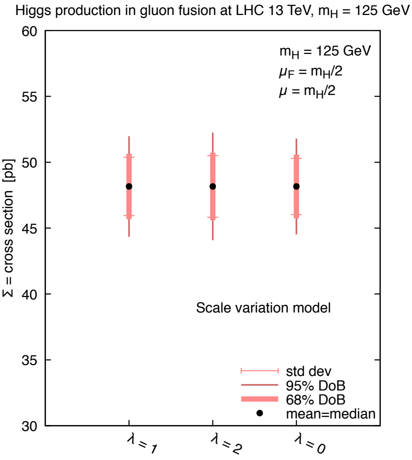
Scan of priors for the scale μ





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 $c_0, ..., c_n$ orders

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

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

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