Probabilistic definition of the perturbative theoretical uncertainty from missing higher orders

Marco Bonvini

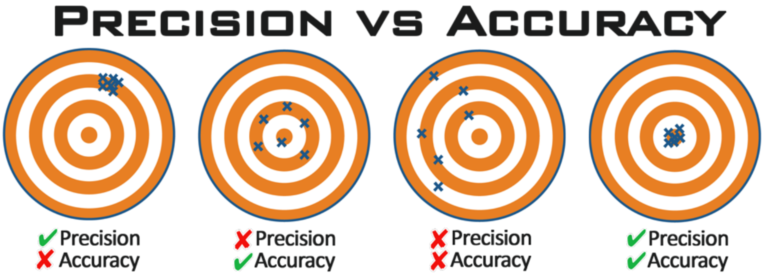
INFN, Rome 1 unit

UZH and ETH, Zürich (zoom), 14 December 2020

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

Istituto Nazionale di Fisica Nucleare
Sezione di ROMA

Precision is nothing without accuracy



Precision = small uncertainty
Accuracy = reliable uncertainty

A very precise (small uncertainty) determination of a cross section which is far from the "true" value is not good for anyone...

A **realistic** determination of the theory uncertainties is preferred/mandatory!

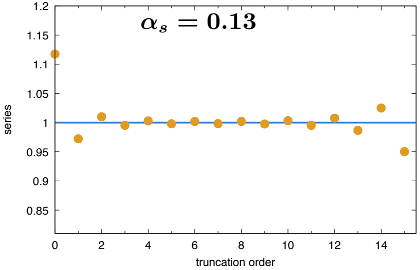
Theoretical prediction within perturbation theory

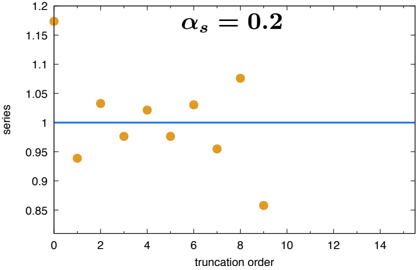
An observable Σ is computed using perturbation theory as

$$\Sigma \simeq \sum_{k=0}^{n} c_k \alpha_s^k + \mathcal{O}(\alpha_s^{n+1})$$

Perturbative expansions are divergent, and assumed to be **asymptotic** to Σ
This implies that up to some order k_{asympt} adding orders improves the approximation; beyond that order, the divergence of the series is manifest

Consider for example $\frac{1}{\alpha_s} \exp\left(\frac{1}{\alpha_s}\right) \Gamma\left(0, \frac{1}{\alpha_s}\right)$ " = " $\sum_{k=0}^{\infty} (-1)^k k! \alpha_s^k$





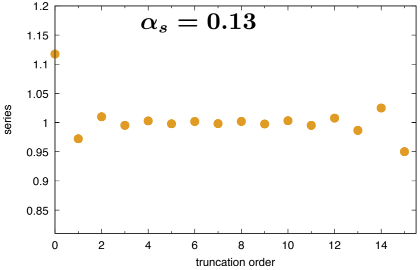
Theoretical prediction within perturbation theory

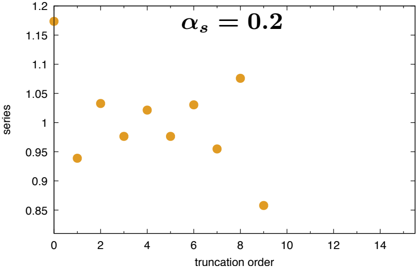
An observable $\boldsymbol{\Sigma}$ is computed using perturbation theory as

$$\Sigma \simeq \sum_{k=0}^{n} c_k \alpha_s^k + \mathcal{O}(\alpha_s^{n+1})$$

Perturbative expansions are divergent, and assumed to be **asymptotic** to $\boldsymbol{\Sigma}$
This implies that up to some order k_{asympt} adding orders improves the approximation; beyond that order, the divergence of the series is manifest

Consider for example $\dfrac{1}{\alpha_s} \exp\left(\dfrac{1}{\alpha_s}\right) \Gamma\left(0, \dfrac{1}{\alpha_s}\right)$ " = " $\displaystyle\sum_{k=0}^{\infty} (-1)^k k! \alpha_s^k$



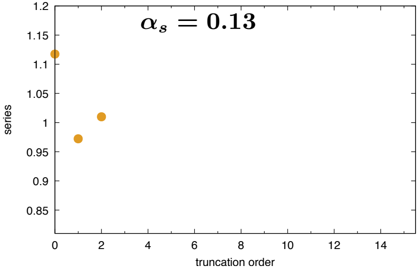


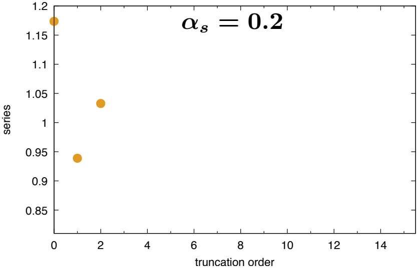
An observable $\boldsymbol{\Sigma}$ is computed using perturbation theory as

$$\boldsymbol{\Sigma \simeq \sum_{k=0}^{n} c_k \alpha_s^k + \mathcal{O}(\alpha_s^{n+1})}$$

Perturbative expansions are divergent, and assumed to be **asymptotic** to $\boldsymbol{\Sigma}$
This implies that up to some order k_{asympt} adding orders improves the approximation; beyond that order, the divergence of the series is manifest

Consider for example $\boldsymbol{\frac{1}{\alpha_s}\exp\left(\frac{1}{\alpha_s}\right)\Gamma\left(0,\frac{1}{\alpha_s}\right)}$ "$=$" $\boldsymbol{\sum_{k=0}^{\infty}(-1)^k k!\,\alpha_s^k}$





Perturbation theory and missing higher orders (MHO)

An observable $\boldsymbol{\Sigma}$ can be written as

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathrm{N}^{n}\mathrm{LO}} + \boldsymbol{\Delta}_{\mathrm{MHO}} + \boldsymbol{\Delta}_{\text{non-pert}}$$

where

- $\boldsymbol{\Sigma}_{\mathrm{N}^{n}\mathrm{LO}} = \sum_{k=0}^{n} c_k \alpha_s^k$ next-to-next-....-to-leading order
- $\boldsymbol{\Delta}_{\mathrm{MHO}} = \sum_{k=n+1}^{k_{\text{asympt}}} c_k \alpha_s^k$ are the missing higher orders
- $\boldsymbol{\Delta}_{\text{non-pert}}$ contains non-perturbative contributions

We believe (and we assume) that

$$|\boldsymbol{\Delta}_{\mathrm{MHO}}| \gg |\boldsymbol{\Delta}_{\text{non-pert}}|$$

and thus focus on $\boldsymbol{\Delta}_{\mathrm{MHO}}$ from now on

How can we estimate $\boldsymbol{\Delta}_{\mathrm{MHO}}$?

the canonical method

Unphysical scales to probe MHO

Renormalization in QFT introduces a dependence on a unphysical scale μ

Physical observables are independent of μ

$$\mu \frac{d}{d\mu} \Sigma = 0$$

However, perturbative computations have a residual dependence on μ, which is formally of higher order

$$\Sigma_{\mathrm{N}^{n}\mathrm{LO}}(\mu) = \sum_{k=0}^{n} c_k(\mu) \alpha_s^k(\mu)$$

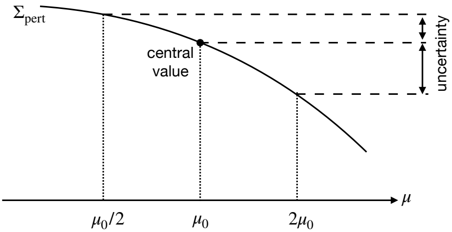
$$\mu \frac{d}{d\mu} \Sigma_{\mathrm{N}^{n}\mathrm{LO}}(\mu) = \mathcal{O}(\alpha_s^{n+1}) = \mathcal{O}(\Delta_{\mathrm{MHO}})$$

Idea: use the scale dependence to probe higher orders

Canonical method: Scale Variation

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

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

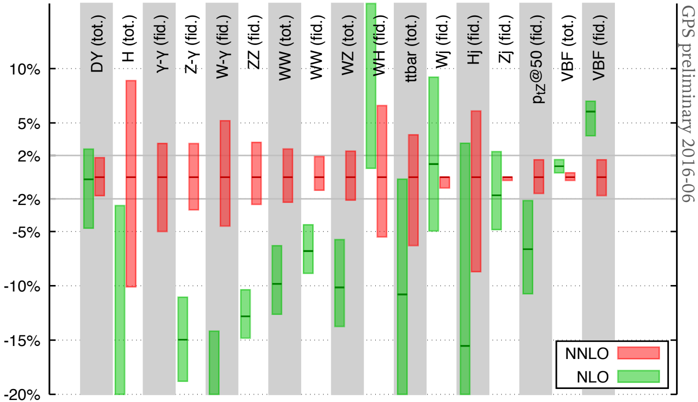


Caveats:

Which central scale μ_0?
How much should I vary the scale?
How do I deal with stationary points?
How do I interpret the uncertainty probabilistically?

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

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**

This would also allow for a statistically meaningful comparison of theory predictions with data (e.g. precision tests of the Standard Model, or PDF fits)

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

Frequentist approach to probability → requires repeatable events → no way...

Bayesian approach → probability defined as the **degree of belief** of an "event"

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

Acquiring information → 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 $\boldsymbol{\Sigma}$:

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

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

the breakthrough

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

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

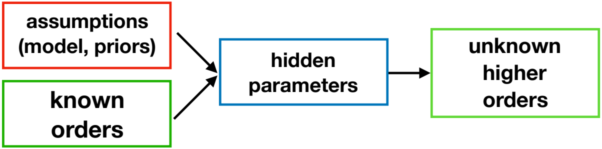
"We make the assumption that all the coefficients c_k in a perturbative series share some sort of upper bound $\bar{c} > 0$ to their absolute values, specific to the physical process studied. The calculated coefficients will give an estimate of this $\bar{c}$, restricting the possible values for the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

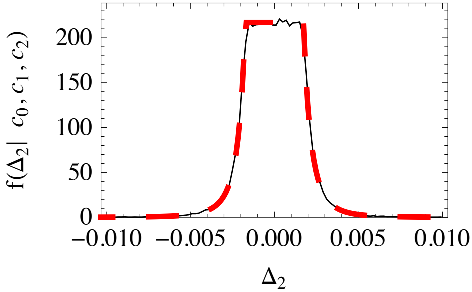
$$P(\text{pars}|\text{known } c_k) \propto P(\text{known } c_k|\text{pars})P_0(\text{pars})$$

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

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

The typical output of the Cacciari-Houdeau method is a distribution for $\boldsymbol{\Delta}_{\mathrm{MHO}}$

$$P(\boldsymbol{\Delta}_{\mathrm{MHO}}|\boldsymbol{c_0},\boldsymbol{c_1},\boldsymbol{c_2}) \simeq P(\boldsymbol{\alpha_s^3 c_3}|\boldsymbol{c_0},\boldsymbol{c_1},\boldsymbol{c_2})$$



One can then compute statistical quantities like degree-of-belief (DoB) intervals, standard deviation, ...

But the probability distribution is the actual result, and the experimentalist can use it directly

Limitations of the CH model

The Cacciari-Houdeau model has some caveats too:

- it assumes a convergent perturbative behaviour (bounded by a geometric series)

$$|c_k| \leq \bar{c} \qquad \Rightarrow \qquad \left|\sum_k c_k \alpha_s^k\right| \leq \sum_k |c_k| \alpha_s^k \leq \sum_k \bar{c} \alpha_s^k$$

  - acceptable from an asymptotic expansion point of view
  - attempt to treat the series as factorially divergent [Bagnaschi,Cacciari,Guffanti,Jenniches 1409.5036]

- if the coefficients grow as a power, $c_k \sim \eta^k$, which is very likely, the method cannot perform well
  - Cacciari-Houdeau proposed a modified version with η accounted for
  - in [Bagnaschi,Cacciari,Guffanti,Jenniches 1409.5036] η is determined from a survey on various observables
  - in an alternative approach [Forte,Isgrò,Vita 1312.6688] the value of η is fitted

- it still depends on the choice of the "central" scale μ_0

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

Notational change

More general expansion

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

Rather important:

- emphasises that the LO does not play a role, but only sets the size
- $\delta_k(\mu)$ are dimensionless
- $\delta_0(\mu) = 1$ independent of μ (very important for dealing with scale dependence, see later)
- if a series starts at order $\alpha_s^{k_0}$, this sum keeps starting from $k = 0$
- the coupling does no longer appear explicitly
- it can describe a more general perturbative expansion, e.g. a resummed expansion

model 1

sect. 4

Model 1: Geometric behaviour model (improved Cacciari-Houdeau)

Generalized condition that accounts for a possible power growth

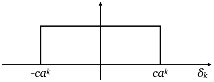
$$|\delta_k(\mu)| \le c a^k \qquad \forall k < k_{\text{asympt}} \qquad\qquad \text{CH: } \left|c_k \alpha_s^k\right| \le \bar{c}\alpha_s^k$$

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:

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

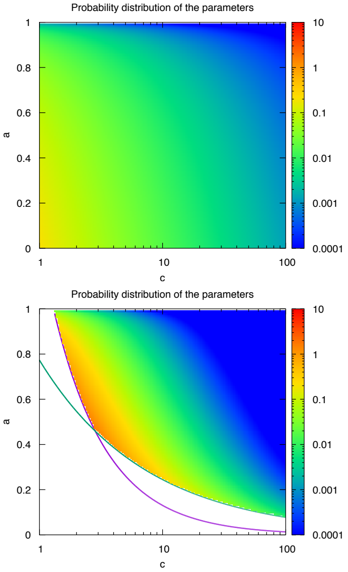
Inference scheme:

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

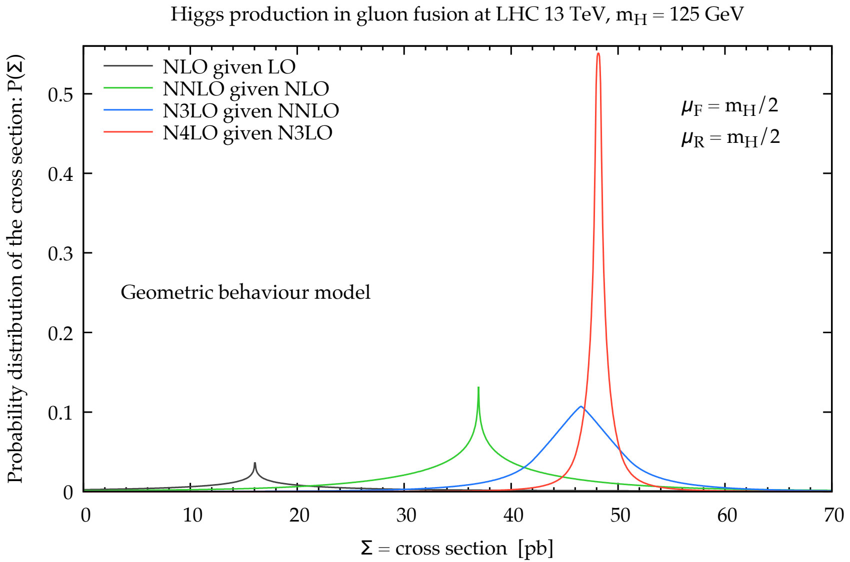
Final output:

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

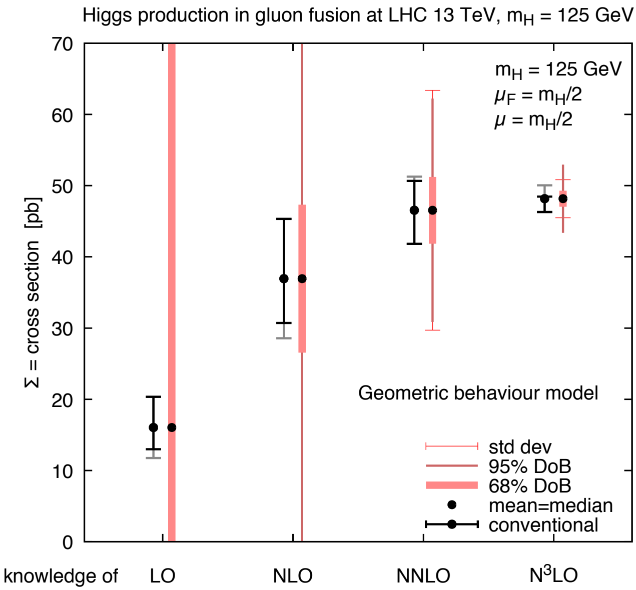
Posterior of c, a for Higgs production in gluon fusion



Higgs in gluon fusion at LHC: probability distributions



From distributions to statistical estimators



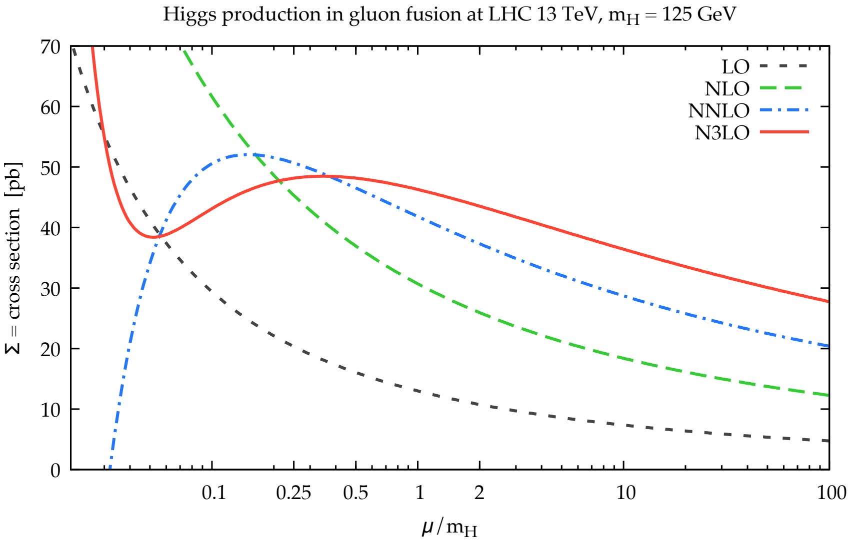
conventional result: canonical scale variation

new result: geometric behaviour model

model 2

sect. 5

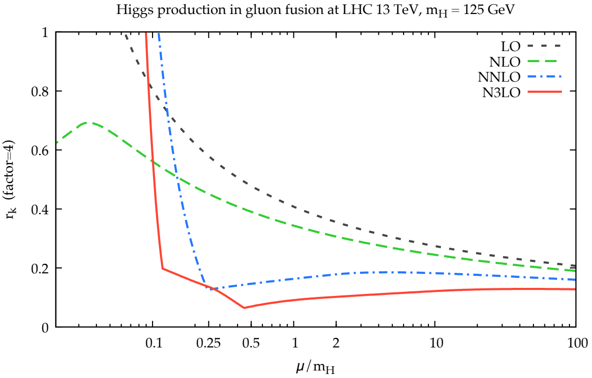
Higgs production in gluon fusion at LHC



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:

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



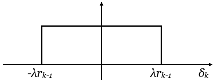
I propose the condition

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

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:

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

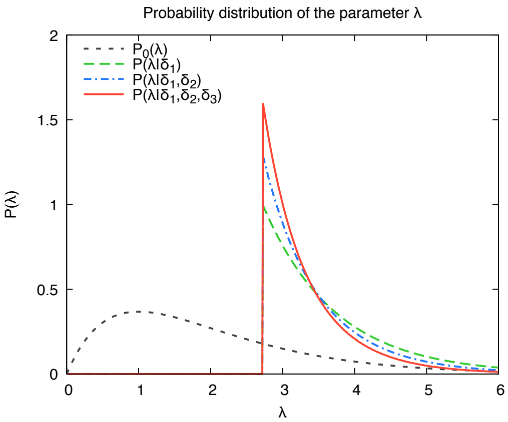
Inference scheme:

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

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

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

Posterior of λ for Higgs production in gluon fusion

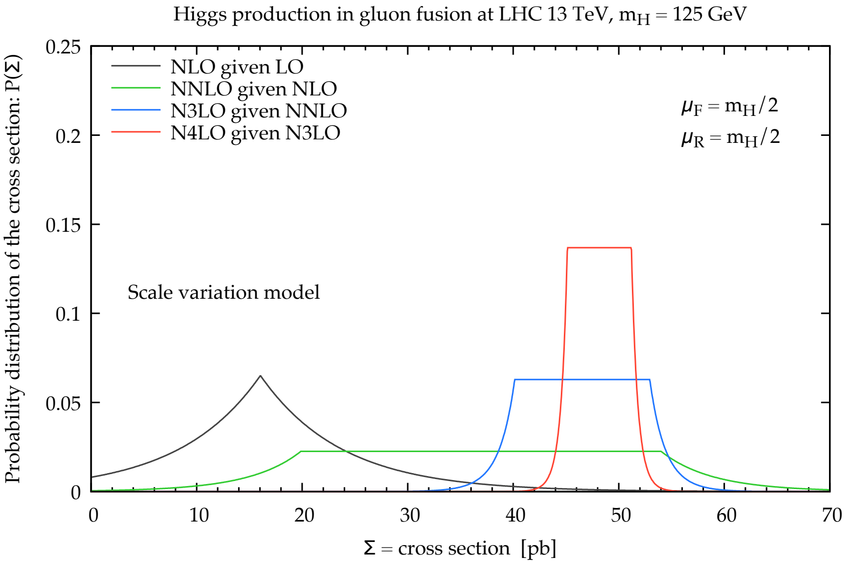


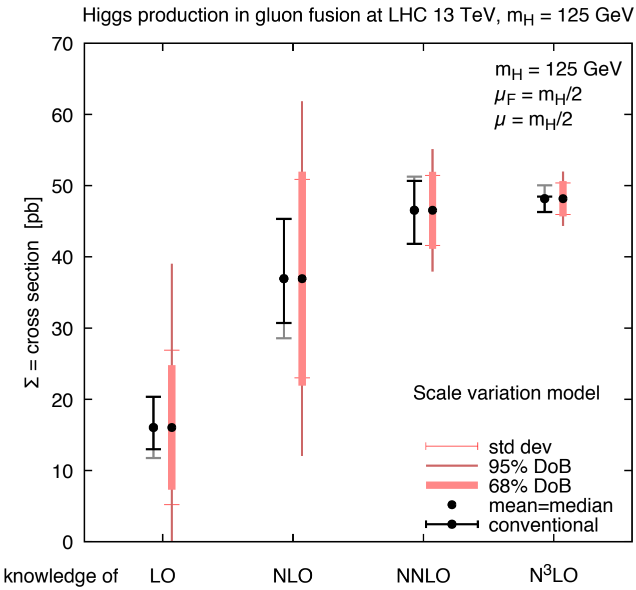
The first non-trivial order ($\boldsymbol{\delta_1}$) sets the lower limit of $\boldsymbol{\lambda}$

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

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

Higgs in gluon fusion at LHC: probability distributions





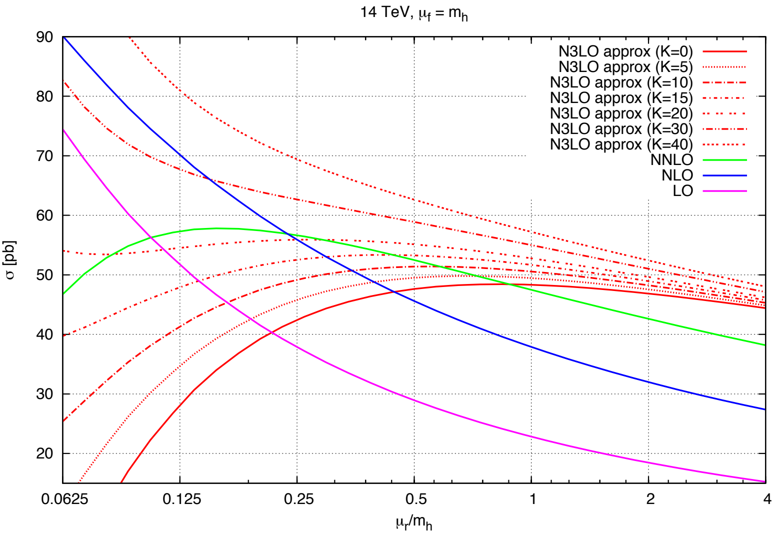
conventional result: canonical scale variation

new result: scale variation inspired model

model 3

appendix B.4

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:

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

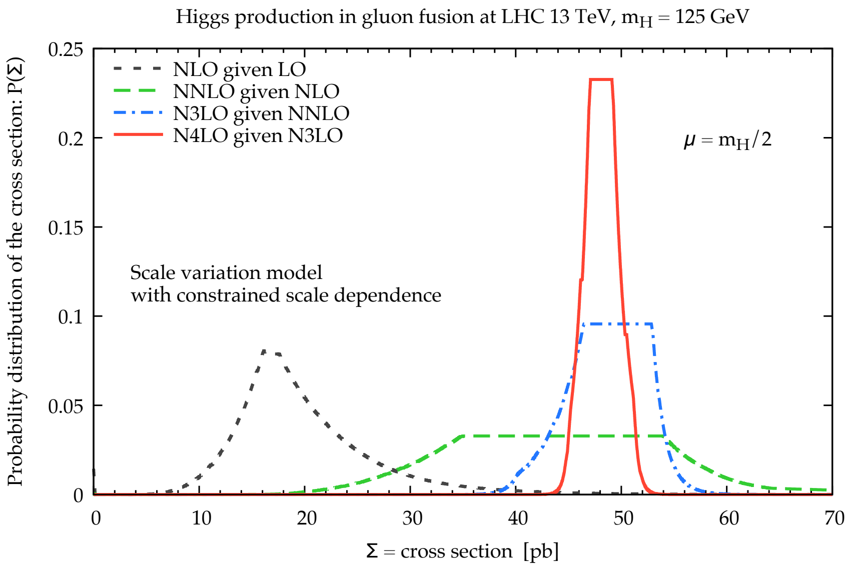
that depends on two hidden parameters λ, η

The implementation of these condition is more difficult (see appendix B.4)

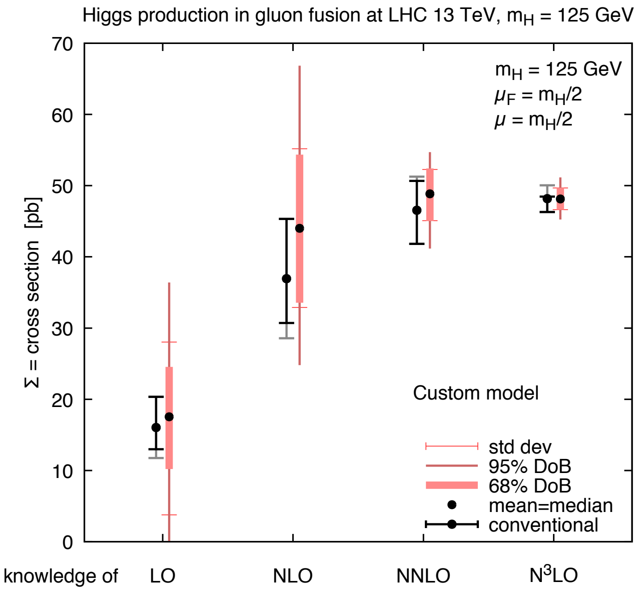
New prior:

$$P(\eta|\mu) = e^{-\eta}\theta(\eta)$$

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


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

From distributions to statistical estimators



conventional result: canonical scale variation

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

more models

appendix B.2, B.3, B.5, B.6

Combining models and inventing new ones

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

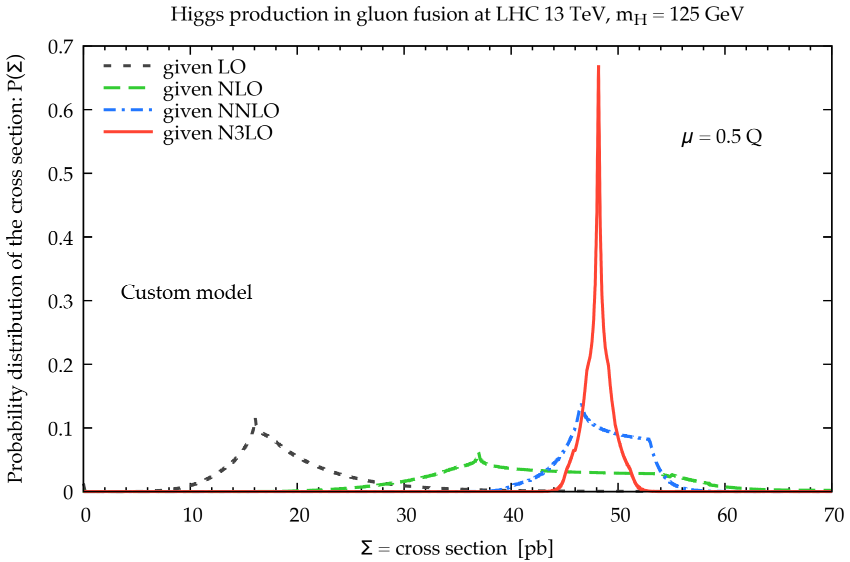
So far we have seen three conditions

$$|\delta_k(\mu)| \le c a^k$$
$$|\delta_k(\mu)| \le \lambda r_{k-1}(\mu)$$
$$|r_k(\mu)| \le \eta r_{k-1}(\mu)$$

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

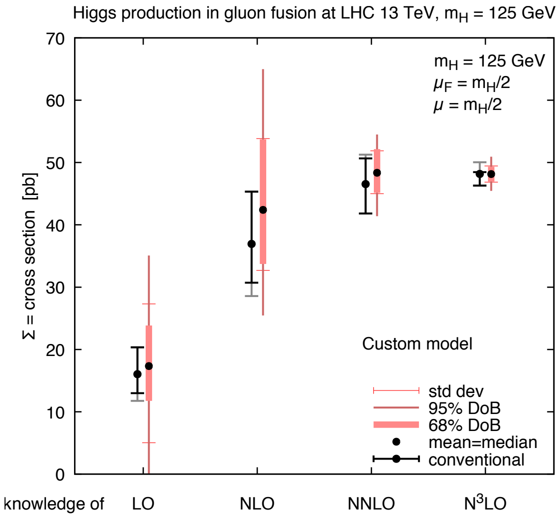
The models are implemented in a code named `THunc`, that provides a *custom model* feature to implement any customized model

Putting all conditions together....



Higgs production in gluon fusion at LHC 13 TeV, $m_H = 125$ GeV
given LO
given NLO
given NNLO
given N3LO
$\mu = 0.5$ Q
Custom model
Probability distribution of the cross section: $P(\Sigma)$
Σ = cross section [pb]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
10
20
30
40
50
60
70

From distributions to statistical estimators



dealing with scale dependence

sect. 6

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!

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

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

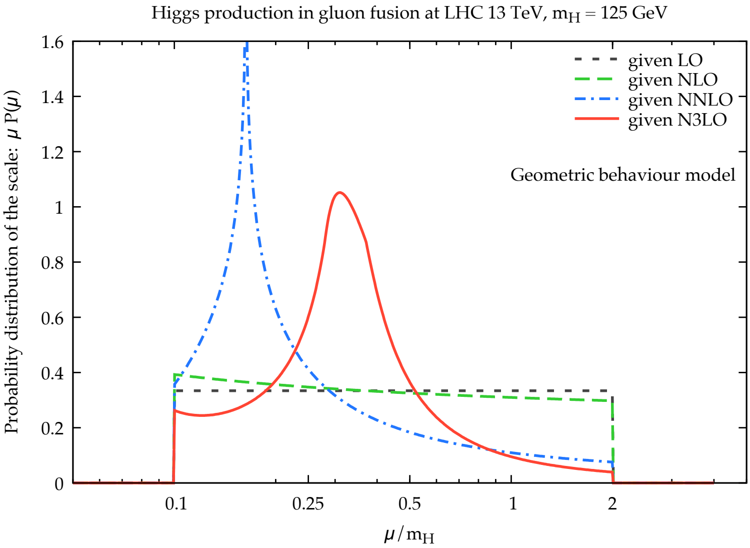
where $P(\mu|\delta_0,\ldots,\delta_n)$ is the posterior distribution for μ given the known orders (which depends on the model)

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

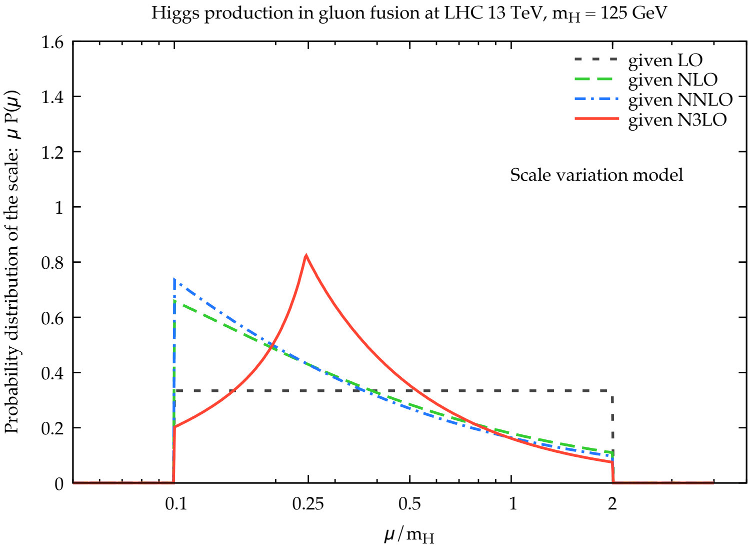
The prior $P_0(\mu)$ contains our prejudices on what are the most appropriate scales, but the results are largely independent of the precise form and size of the prior
$\Rightarrow$ **a lot of arbitrariness is removed!**

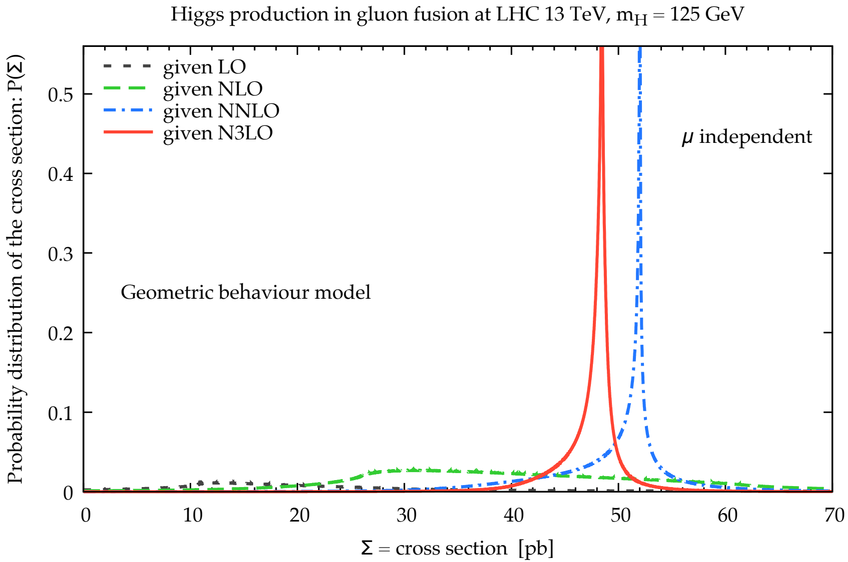
Note: it is crucial to use the dimensionless δ_k coefficients, such that $\delta_0 = 1$, otherwise the LO will also contribute to the inference on the scale, giving non-sense results (see sect. 6.2)

Posterior distribution for the scale μ

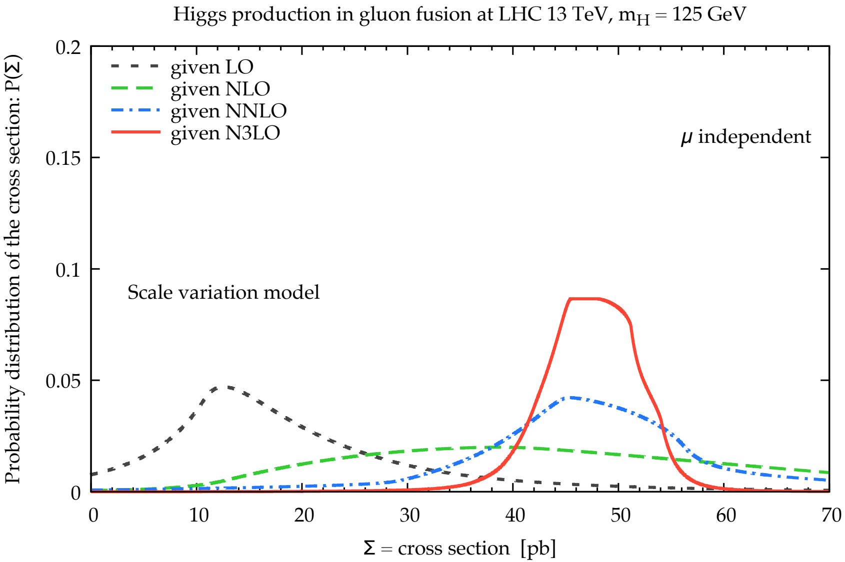


Posterior distribution for the scale μ

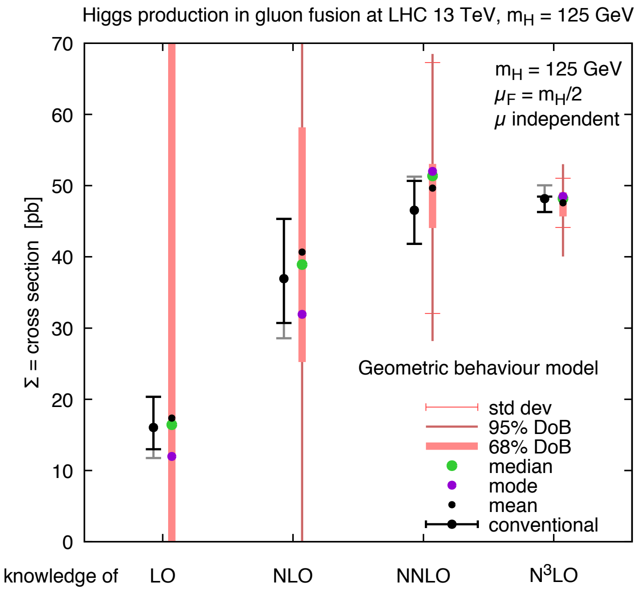



Higgs production in gluon fusion at LHC 13 TeV, $m_H = 125$ GeV
given LO
given NLO
given NNLO
given N3LO
μ independent
Geometric behaviour model
Probability distribution of the cross section: P(Σ)
0
0.1
0.2
0.3
0.4
0.5
0
10
20
30
40
50
60
70
Σ = cross section [pb]

Higgs in gluon fusion at LHC: scale independent distributions



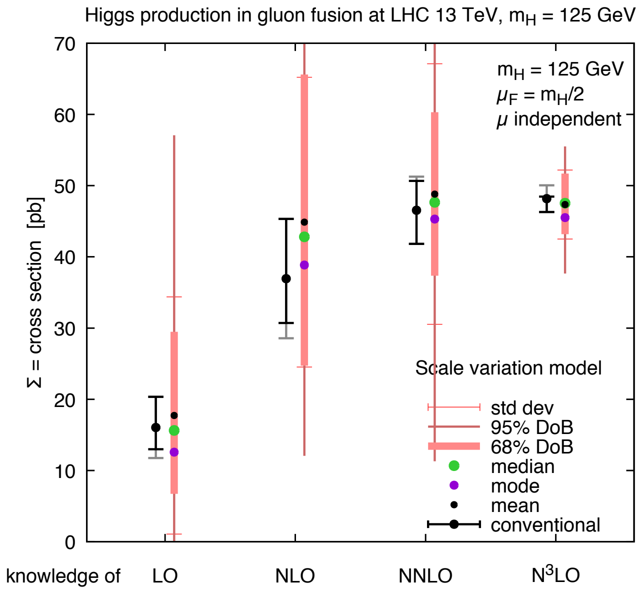
Higgs in gluon fusion at LHC: final results



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

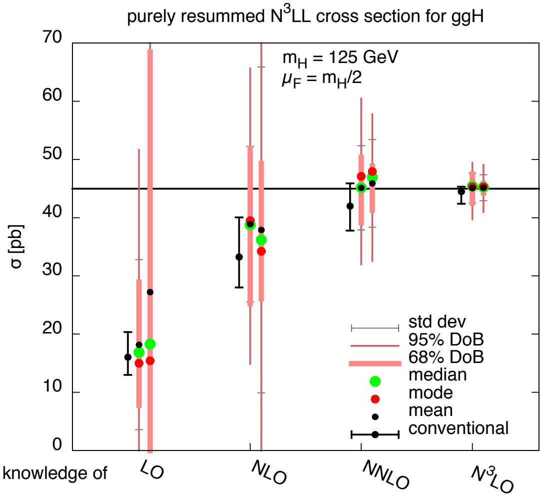
validation

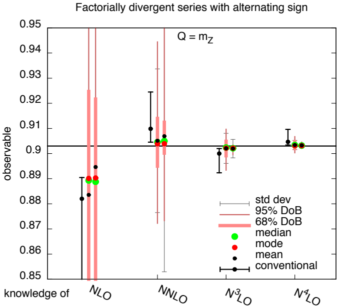
sect. 7

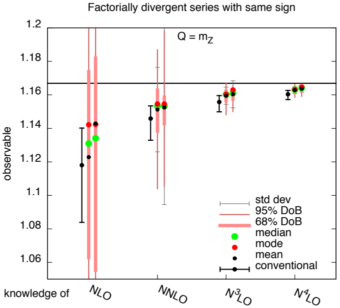
Validation

Tests to series with known sums to verify the goodness of the model:

- TOY: convergent series $\sum_k (A\alpha_s)^k \cos(Bk)$
- TOY: factorially divergent series with alternating signs $\sum_k (-1)^k k! \alpha_s^k$
- TOY: factorially divergent series with equal signs $\sum_k k! \alpha_s^k$
- anharmonic oscillator in Quantum Mechanics
- purely resummed ggH at N^{3}LL, expanded in powers of α_s







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

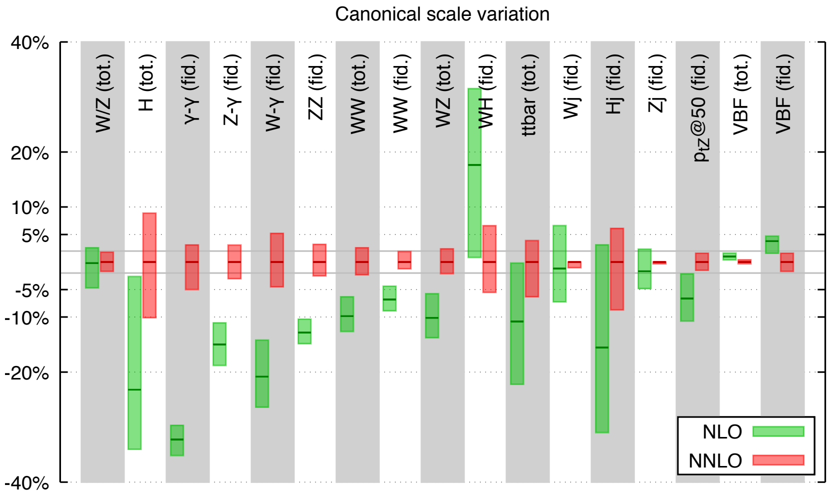
conclusions

Summary

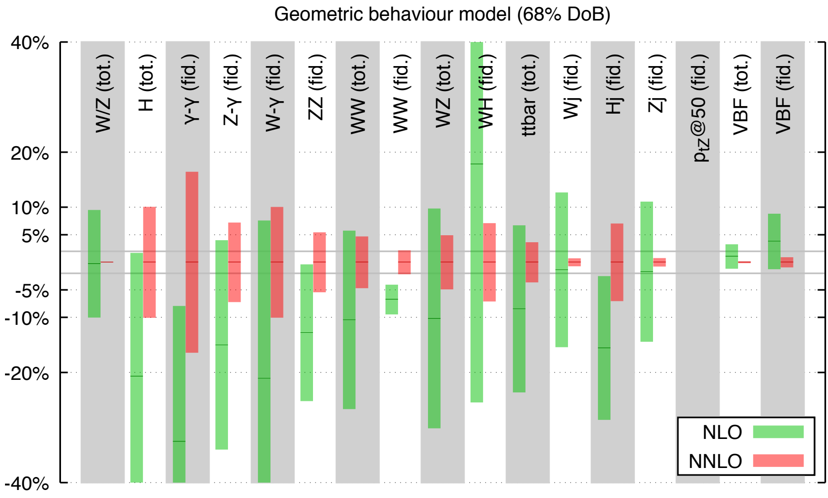
Key message: it is possible to define theory uncertainties from MHO in a probabilistic way, which is reliable and less arbitrary than the canonical scale-variation approach

- **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`
- **Correlations?**
  - correlations between kinematic points of the same observable/process, or between processes are fundamental
  - no unique way to implement them, need to decide how correlations arise
  - interesting proposal by F.Tackmann [SCET2019]
- **Dynamical scales** $\rightarrow$ straightforward extension

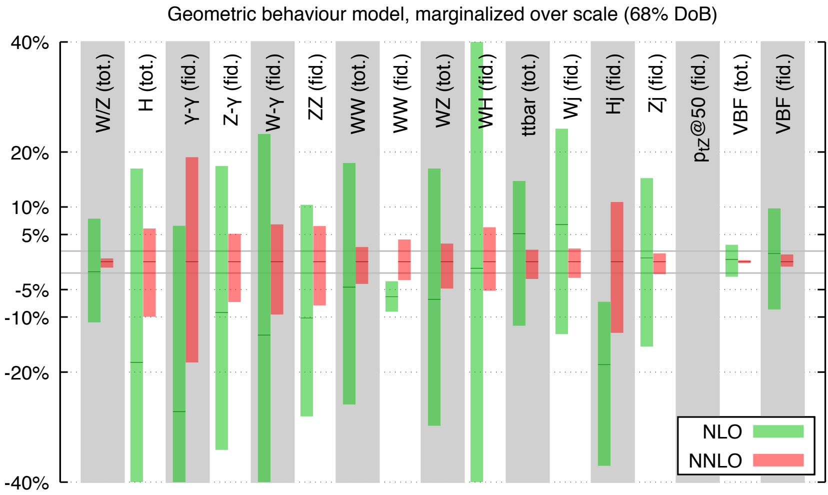
Gavin Salam's plot



Gavin Salam's plot



Gavin Salam's plot



Recommendation

So many models.... which should you choose?

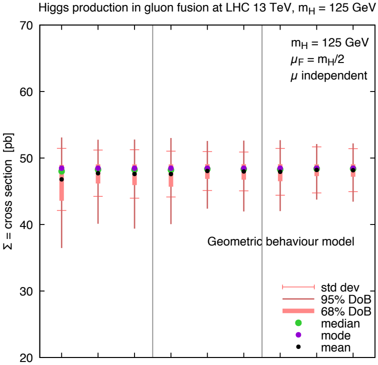
- **The honest answer:**
  The user should decide, based on his own *beliefs*, which model and prior better suit the given perturbative expansion.
  It is fundamental to state this choice very clearly.

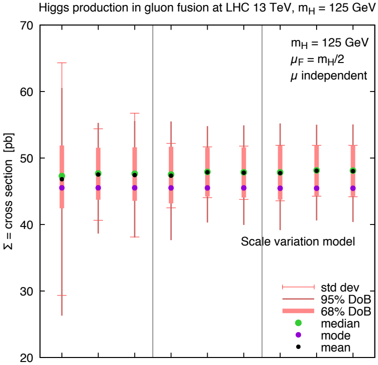
- **A recommendation (i.e., my own preference):**
  - Use the geometric behaviour model with marginalization over the scale μ as default. It works always well and leads to decently precise results. It is also fast.
  - Consider also the scale variation model with marginalization over μ for cross check.
  - For more aggressive application, mixing all models would lead to the best performance (more precise results), but it is slow.

But please, do not buy this as a recipe!

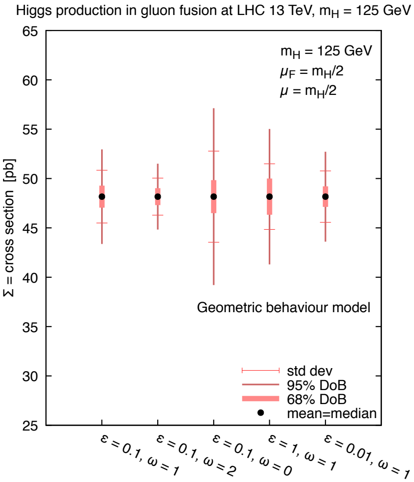
Backup slides

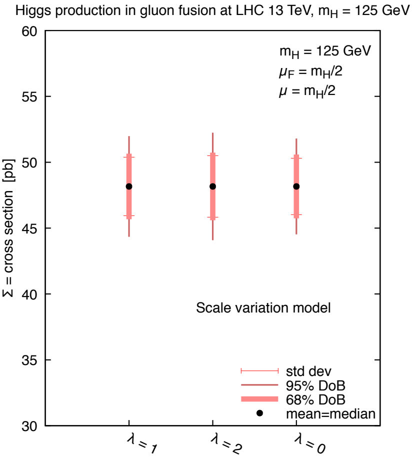
Scan of priors for the scale μ





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

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