Beyond Scale Variations
Theory Uncertainties from Nuisance Parameters.

Frank Tackmann

Deutsches Elektronen-Synchrotron

Les Houches, June 14, 2019

Theory Uncertainties and Correlations.

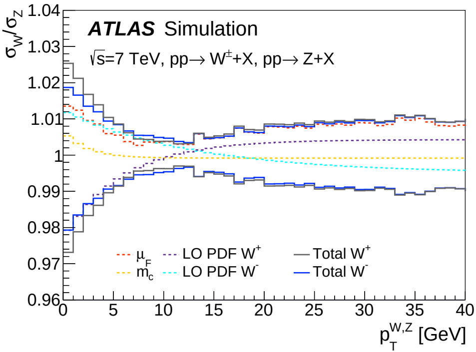
Reliable theory uncertainties are essential for any precision studies and interpretation of experimental measurements

- Especially when theory uncertainties $\gtrsim$ experimental uncertainties
- Correlations can have significant impact
  - In fact, whenever one combines more than a single measurement, one should ask how the theory uncertainties in the predictions for each measurement are correlated with each other
  - Correlations between different points in a spectrum
  - Correlations between processes, observables, ...
- So far we have (mostly) been skirting the issue
  - However, experimentalists have to treat theory uncertainties like any other systematic uncertainty, and in absence of anything better they have to make something up based on naive scale variations
  - In likelihood fits, some (possibly enveloped) scale variation impact will get treated as a free nuisance parameter and floated in the fit

Example: Measurement of the W Mass.

Small $p_T^W < 40\,\mathrm{GeV}$ is the relevant region for m_W

- Needs very precise predictions for p_T^W spectrum
- $\simeq 2\%$ uncertainties in p_T^W translate into $\simeq 10\,\mathrm{MeV}$ uncertainty in m_W
- Direct theory predictions for p_T^W are insufficient



⇒ Strategy: Exploit precisely measured Z p_T spectrum to get best possible description for W

- Regardless how precisely $\mathrm{d}\sigma(W)/\mathrm{d}p_T$ can be calculated directly, one always wants to exploit Z data to maximize precision

Example: Extrapolating from Z to W.

$$\underbrace{\frac{\mathrm{d}\sigma(W)}{\mathrm{d}p_T}}_{\text{needed}} = \underbrace{\left[\frac{\mathrm{d}\sigma(Z)}{\mathrm{d}p_T}\right]_{\text{measured}}}_{\text{measure precisely}} \times \underbrace{\left[\frac{\mathrm{d}\sigma(W)/\mathrm{d}p_T}{\mathrm{d}\sigma(Z)/\mathrm{d}p_T}\right]_{\text{theory}}}_{\substack{\text{calculate precisely} \\ \text{theory uncertainties cancel}}}$$

- Ratio is just a proxy
  - More generally: Combined fit to both processes
  - Tuning Pythia on Z and using it to predict W is one example of this

Example: Extrapolating from Z to W.

$$\underbrace{\frac{\mathrm{d}\sigma(W)}{\mathrm{d}p_T}}_{\text{needed}} = \underbrace{\left[\frac{\mathrm{d}\sigma(Z)}{\mathrm{d}p_T}\right]_{\text{measured}}}_{\text{measure precisely}} \times \underbrace{\left[\frac{\mathrm{d}\sigma(W)/\mathrm{d}p_T}{\mathrm{d}\sigma(Z)/\mathrm{d}p_T}\right]_{\text{theory}}}_{\substack{\text{calculate precisely}\\ \text{theory uncertainties cancel}}}$$

- Ratio is just a proxy
  - More generally: Combined fit to both processes
  - Tuning Pythia on Z and using it to predict W is one example of this
- Crucial Caveat: Cancellation fundamentally relies on theory correlations
  - Take 10% theory uncertainty on $\mathrm{d}\sigma(W)$ and $\mathrm{d}\sigma(Z)$
    - $\rightarrow$ 99.5% correlation yields 1% uncertainty on their ratio
    - $\rightarrow$ 98.0% correlation yields 2% uncertainty on their ratio – 2$\times$ larger!
- One of many examples, this happens whenever experiments extrapolate from some control region or process to the signal region

What is a Scale Variation?

What is a Scale Variation?

It is not automagically a theory uncertainty!

(in case you didn't pay attention during Stefano's talk)

So What is a Scale Variation?

It is a (continuous) change of perturbative scheme, i.e., a different way to expand the same quantity

$$\epsilon = \alpha_s(\mu) \qquad \to \qquad \sigma = c_0 + \epsilon\, c_1 + \epsilon^2\, c_2 + \cdots$$

$$\tilde{\epsilon} = \alpha_s(\tilde{\mu}) \qquad \to \qquad \sigma = c_0 + \tilde{\epsilon}\, \tilde{c}_1 + \tilde{\epsilon}^2\, \tilde{c}_2 + \cdots$$

- The all-order result is the same and scheme independent
  - Truncated expansions are scheme dependent, so their difference *might* give us a feeling about the possible size of the missing $\epsilon^2 c_2 + \cdots$ terms
- It also *might not*
  - Many examples where this is not quantitatively reliable
  - Often, the main reason is that there are new structures in c_2 that are not present in c_1 (new partonic channels, new kinematic dependences, ...)
  - Side note: Using the shift from the previous order has the same caveat

What About Correlations?

Correlations only come from common sources of uncertainties

✓ “Straightforward” for unc. due to input parameters ($\alpha_s(m_Z)$,)

Scale variations are inherently ill-suited for correlations

✗ Scales are not physical parameters with an uncertainty that can be propagated

  ✗ They are not the underlying source of uncertainty

  ✗ Scale variation reduces at higher order not because the scales become better known but because the cross section becomes less dependent on them

✗ A priori, scale variations do not imply true correlations between different kinematic regions or different processes

✗ Taking an envelope is not a linear operation and so does not propagate

⇒ In my mind, trying to decide how to (un)correlate scale variations in the end only treats a symptom, but not the actual problem

A Possible Solution.

$$\sigma = c_0 + \alpha_s(\mu)\big[c_1 + \alpha_s(\mu)\, c_2 + \cdots\big]$$

Identify the actual source of uncertainty

- The unknown higher-order corrections: $\alpha_s(\mu)\, c_2 + \cdots$

Parametrize and vary the unknown

- We often know quite a lot about the general structure of c_2
  - μ dependence, color structure, partonic channels, kinematic structure, ...
- Suitably parametrize the missing pieces
  - Simplest case: c_2 is just a number
  - More generally, have to parametrize an unknown function
- Common/independent pieces between different predictions determine the correlations between them

Provides Numerous Advantages.

Immediately get all benefits of parametric uncertainties

- ✓ Encode correct correlations
- ✓ Can be propagated straightforwardly
  - Including Monte Carlo, BDTs, neural networks, ...
- ✓ Can be consistently included in fits (and in principle be constrained by data)
  - Allows using control measurements to reduce theory uncertainties
- ✓ Can correctly correlate theory uncertainties between measurement and interpretation

Provides Numerous Advantages.

Immediately get all benefits of parametric uncertainties

- ✓ Encode correct correlations
- ✓ Can be propagated straightforwardly
  - Including Monte Carlo, BDTs, neural networks, ...
- ✓ Can be consistently included in fits (and in principle be constrained by data)
  - Allows using control measurements to reduce theory uncertainties
- ✓ Can correctly correlate theory uncertainties between measurement and interpretation

Additional theory benefits compared to scale variations

- Much easier to scrutinize since all assumptions are fully exposed
- Can fully exploit all partially known higher-order information
- Can explicitly account for new structures appearing at higher order
- Typically there will be multiple parameters
  - Much safer against accidental underestimates due to multiple parameters
  - Due to central-limit theorem, total theory uncertainty becomes Gaussian

Application to p_T Resummation.

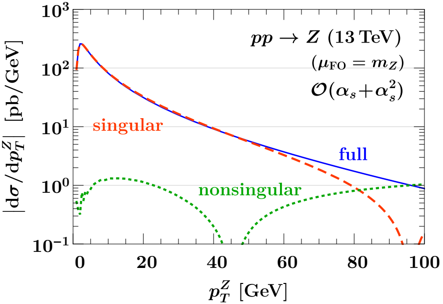
[arxiv:19xx.sooon]

Small-p_T Power Expansion.

Define scaling variable $\tau \equiv p_T^2/m_V^2$ and expand in powers of τ

$$\begin{aligned}\frac{d\sigma}{d\tau} &= \delta(\tau) + \alpha_s\Big[\frac{\ln\tau}{\tau} + \frac{1}{\tau} + \delta(\tau) + f_1^{\rm nons}(\tau)\Big] \\ &\quad + \alpha_s^2\Big[\frac{\ln^3\tau}{\tau} + \frac{\ln^2\tau}{\tau} + \frac{\ln\tau}{\tau} + \frac{1}{\tau} + \delta(\tau) + f_2^{\rm nons}(\tau)\Big] \\ &\quad + \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \ddots + \dots\Big] \\ &= d\sigma^{(0)}/d\tau + \mathcal{O}(\tau)/\tau\end{aligned}$$

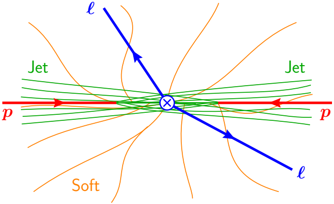
- For small $\tau \ll 1$
  - Logarithmic terms completely dominate perturbative series
  - Their all-order structure is actually simpler and more universal, which allows their resummation
  - Also holds the key for a rigorous treatment of theory correlations



Factorization and Resummation.

Leading-power spectrum factorizes into
hard, collinear, and soft contributions, e.g. for p_T

$$\frac{\mathrm{d}\sigma^{(0)}}{\mathrm{d}\vec{p}_T} = \sigma_0\, H(Q,\mu) \int \mathrm{d}^2\vec{k}_a\, \mathrm{d}^2\vec{k}_b\, \mathrm{d}^2\vec{k}_s$$
$$\times\, B_a(\vec{k}_a, Qe^{Y}, \mu, \nu)\, B_b(\vec{k}_b, Qe^{-Y}, \mu, \nu)$$
$$\times\, S(\vec{k}_s, \mu, \nu)\, \delta(\vec{p}_T - \vec{k}_a - \vec{k}_b - \vec{k}_s)$$



- Each function is a renormalized object with an associated RGE
  - Structure depends on type of variable but is universal for all hard processes

⇒ Dependence on p_T and Q is fully determined to all orders by a coupled system of differential equations
  - Their solution leads to resummed predictions
  - Each resummation order (only) requires as ingredients anomalous dimensions and boundary conditions entering the RG solution

Simplest Example: Multiplicative RGE.

All-order RGE and its solution

$$\mu \frac{\mathrm{d}H(Q,\mu)}{\mathrm{d}\mu} = \gamma_H(Q,\mu)\, H(Q,\mu)$$

$$\Rightarrow \qquad H(Q,\mu) = H(Q) \times \exp\Big[\int_Q^\mu \frac{\mathrm{d}\mu'}{\mu'} \gamma_H(Q,\mu')\Big]$$

Necessary ingredients

- Boundary condition

$$H(Q) = 1 + \alpha_s(Q)\, h_1 + \alpha_s^2(Q)\, h_2 + \cdots$$

- *Anomalous dimension*

$$\gamma_H(Q,\mu) = \alpha_s(\mu)\big[\Gamma_0 + \alpha_s(\mu)\,\Gamma_1 + \cdots\big] \ln\frac{Q}{\mu} + \alpha_s(\mu)\big[\gamma_0 + \alpha_s(\mu)\,\gamma_1 + \cdots\big]$$

⇒ Resummation is determined by coefficients of three fixed-order series

- True regardless of how RGE is solved in more complicated cases

Theory Nuisance Parameters.

Perturbative series at leading power is determined to all orders by a coupled system of differential equations (RGEs)

→ Each resummation order only depends on a few semi-universal parameters

→ Unknown parameters at higher orders are the actual sources of perturbative theory uncertainty

	boundary conditions			anomalous dimensions			
order	h_n	s_n	b_n	γ_n^h	γ_n^s	Γ_n	β_n
LL	h_0	s_0	b_0	–	–	Γ_0	β_0
NLL′	h_1	s_1	b_1	γ_0^h	γ_0^s	Γ_1	β_1
NNLL′	h_2	s_2	b_2	γ_1^h	γ_1^s	Γ_2	β_2
N^{3}LL′	h_3	s_3	b_3	γ_2^h	γ_2^s	Γ_3	β_3
N^{4}LL′	h_4	s_4	b_4	γ_3^h	γ_3^s	Γ_4	β_4

- Basic Idea: Use them as theory nuisance parameters
  - ✓ Vary them independently to estimate the theory uncertainties
  - ✓ Impact of each independent nuisance parameter is fully correlated across all kinematic regions and processes
  - ✓ Impact of different nuisance parameters is fully uncorrelated
- Price to Pay: Calculation becomes quite a bit more complex

How to Vary What.

- Level 1: At given order vary parameters around their known values

$$c_0 + \alpha_s(\mu)\big[c_1 + \alpha_s(\mu)\, c_2 + \cdots\big] \;\to\; c_0 + \alpha_s(\mu)(c_1 + \tilde{\theta}_1)$$

  - Simpler but perhaps less robust

- Level 2: Implement the full next order in terms of unknown parameters

$$c_0 + \alpha_s(\mu)\big[c_1 + \alpha_s(\mu)\, c_2 + \cdots\big] \;\to\; c_0 + \alpha_s(\mu)\big[c_1 + \alpha_s(\mu)\, \theta_2\big]$$

  - More involved, but also more robust, allowing for maximal precision

- In general, can also have combination of both

How to Vary What.

- Level 1: At given order vary parameters around their known values

$$c_0 + \alpha_s(\mu)\big[c_1 + \alpha_s(\mu)\, c_2 + \cdots\big] \;\to\; c_0 + \alpha_s(\mu)(c_1 + \tilde{\theta}_1)$$

  - Simpler but perhaps less robust

- Level 2: Implement the full next order in terms of unknown parameters

$$c_0 + \alpha_s(\mu)\big[c_1 + \alpha_s(\mu)\, c_2 + \cdots\big] \;\to\; c_0 + \alpha_s(\mu)\big[c_1 + \alpha_s(\mu)\, \theta_2\big]$$

  - More involved, but also more robust, allowing for maximal precision

- In general, can also have combination of both

Note: Some parameters are actually functions of additional variables

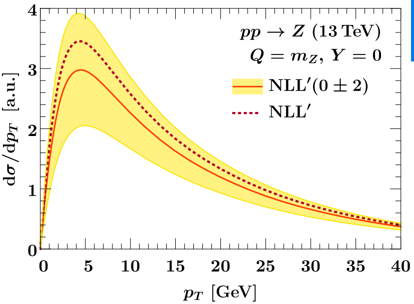
- E.g. beam function constants, auxiliary dependences (jet radius, ...)
- In general, might have to parametrize an unknown function
  - Can e.g. expand/parametrize in terms of appropriate functional basis or known dependence

$Z\ p_T$ Spectrum.

For illustration use

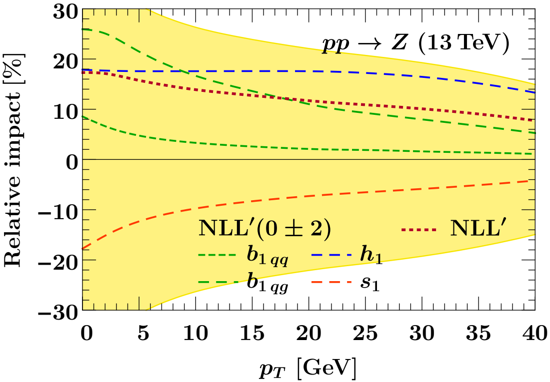
- Level 1: $\tilde{\theta}_i = (0 \pm 0.25) \times c_i$
- Level 2: $\theta_i = (0 \pm 2) \times c_i$

  (with the true values for c_i)



Relative impact of different nuisance parameters

- h_1
- b_1: $q \to q$, $g \to q$
- s_1

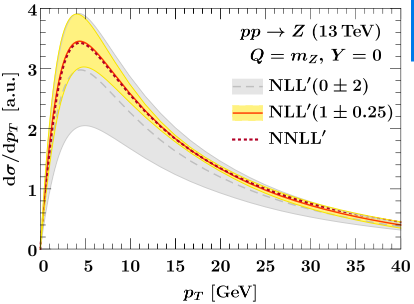


$Z\ p_T$ Spectrum.

For illustration use

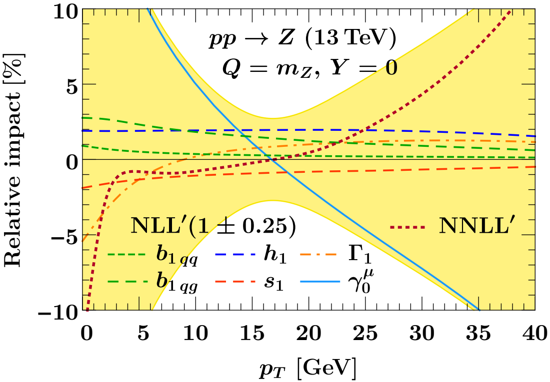
- Level 1: $\tilde{\theta}_i = (0 \pm 0.25) \times c_i$
- Level 2: $\theta_i = (0 \pm 2) \times c_i$

  (with the true values for c_i)



Relative impact of different nuisance parameters

- h_1
- γ_0^μ
- b_1: $q \to q$, $g \to q$
- Γ_1
- s_1

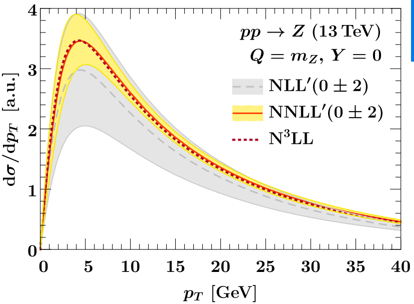


$Z\ p_T$ Spectrum.

For illustration use

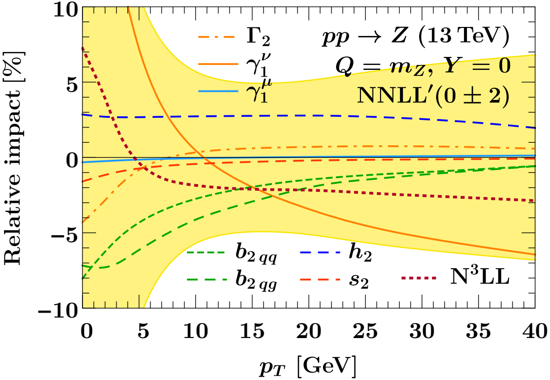
- Level 1: $\tilde{\theta}_i = (0 \pm 0.25) \times c_i$
- Level 2: $\theta_i = (0 \pm 2) \times c_i$

  (with the true values for c_i)

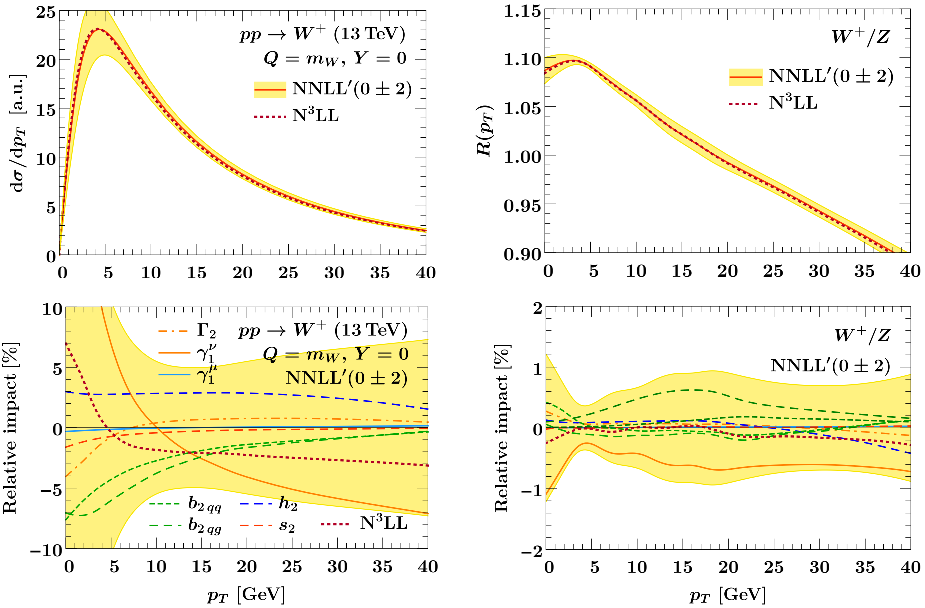


Relative impact of different nuisance parameters

- h_2
- γ_1^μ
- b_2: $q \to q$, $g \to q$
- Γ_2
- γ_1^ν
- s_2

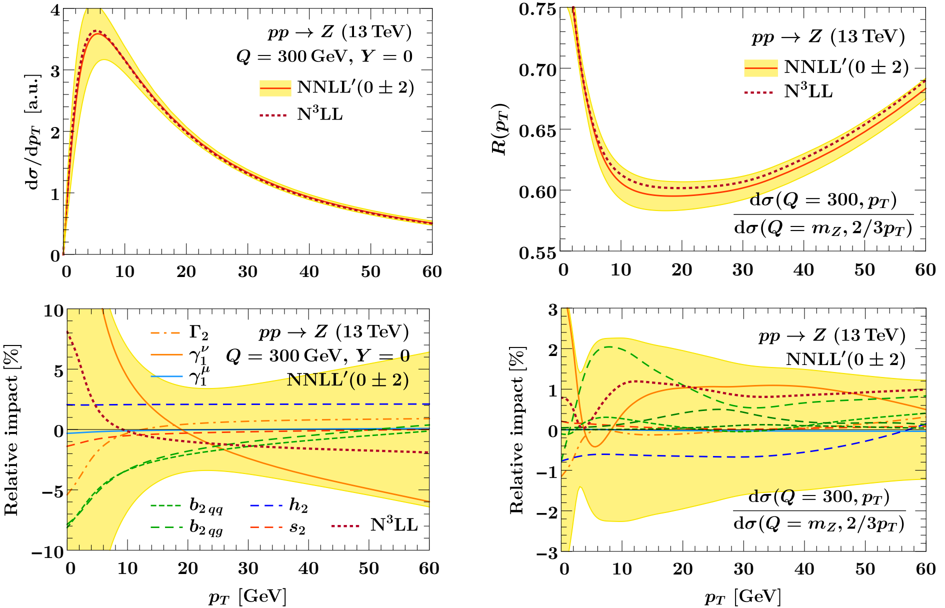


W vs. Z.



$pp \to W^+$ (13 TeV)
$Q = m_W$, $Y = 0$
NNLL′(0 ± 2)
N³LL
dσ/dp_T [a.u.]
W^+/Z
$R(p_T)$
Γ_2
γ_1^ν
γ_1^μ
$b_{2\,qq}$
$b_{2\,qg}$
h_2
s_2
Relative impact [%]
p_T [GeV]

Drell-Yan at High Q vs. Z Pole.



Summary.

A theory prediction without an uncertainty is about as useful as a measurement without an uncertainty

- Uncertainties need to be reliable (small is not good enough ...)

Theory nuisance parameters overcome many problems of scale variations

- Allow to reliably quantify perturbative theory uncertainties
- In particular encode correct correlations
  - Between different p_T values, Q values
  - Between different partonic channels, hard processes
  - Between different variables ($\vec{p}_T$, p_T^{jet}, ϕ^*, ...),
- Can be propagated straightforwardly
  - Including Monte Carlo, BDTs, neural networks, ...

$\Rightarrow$ A plethora of possibilities to explore ...