Theory Uncertainties in Extracting α_s from the $Z\ p_T$ Spectrum.

Frank Tackmann

Deutsches Elektronen-Synchrotron

EWWG General Meeting
July 11, 2024

European Research Council
Established by the European Commission

WIP with Thomas Cridge and Giulia Marinelli
[arXiv:240x.yyyyy]

Outline.

Scale Variations and Theory Correlations.

Scale Variations in a Nutshell.

Theory uncertainty due to inexactness of our prediction

- We have a series expansion in a small quantity α

$$f(\alpha) = f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3\,\alpha^3 + \cdots$$

- We make a prediction based on first few known terms

$$f^{\text{predicted}} = f_0 + f_1\,\alpha \pm \Delta f \qquad \text{with} \qquad \Delta f = f_1\,b_0\,\alpha^2 + \mathcal{O}(\alpha^3)$$

Scale Variations in a Nutshell.

Theory uncertainty due to inexactness of our prediction

- We have a series expansion in a small quantity α

$$f(\alpha) = f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3\,\alpha^3 + \cdots$$

- We make a prediction based on first few known terms

$$f^{\text{predicted}} = f_0 + f_1\,\alpha \pm \Delta f \qquad \text{with} \qquad \Delta f = f_1\,b_0\,\alpha^2 + \mathcal{O}(\alpha^3)$$

We effectively account for inexactness by approximating $f_2 \approx f_1\,b_0$

✓ Resulting Δf is indeed $\mathcal{O}(\alpha^2)$

✗ Nothing guarantees that this is a good approximation (often it is not)

- f_2 usually has more complex structure than just $f_1 \times \mathbf{const}$

✗ $b_0 \sim \beta_0 \ln(2\mu/\mu)$ is *not* a parameter with a true value that f depends on

- No value for it might ever capture the true result (happens regularly)

Theory Correlations.

Correlations can be crucial once several predictions are used in combination

- Prototype of many data-driven methods or any type of combined fit

$$\underbrace{f(y_i)}_{\text{wanted}} = \underbrace{\Big[g(y_j)\Big]_{\text{measured}}}_{\text{measure precisely}} \times \underbrace{\left[\frac{f(y_i) \pm \Delta f}{g(y_j) \pm \Delta g}\right]_{\text{predicted}}}_{\text{theory uncertainties cancel}}$$

  - Cancellation of theory uncertainties is often assumed or taken for granted
  - But obviously relies crucially on precise correlation between Δf and Δg

Theory Correlations.

Correlations can be crucial once several predictions are used in combination

- Prototype of many data-driven methods or any type of combined fit

$$\underbrace{f(y_i)}_{\text{wanted}} = \underbrace{\Big[g(y_j)\Big]_{\text{measured}}}_{\text{measure precisely}} \times \underbrace{\left[\frac{f(y_i) \pm \Delta f}{g(y_j) \pm \Delta g}\right]_{\text{predicted}}}_{\text{theory uncertainties cancel}}$$

- ▶ Cancellation of theory uncertainties is often assumed or taken for granted
- ▶ But obviously relies crucially on precise correlation between Δf and Δg
- ▶ For example: Take a 10% uncertainty for both Δf and Δg, then

a correlation of	99.5%	98%	95.5%	87.5%
yields a reduction by a factor of	10	5	3.33	2
and an uncertainty on the ratio of	1%	2%	3%	5%

$\Rightarrow$ The Challenge: How to account for correlation between Δf and Δg?

- ▶ Depends on the extent to which inexactness in $f(y_i)$ and $g(y_i)$ are related

What About Correlations?

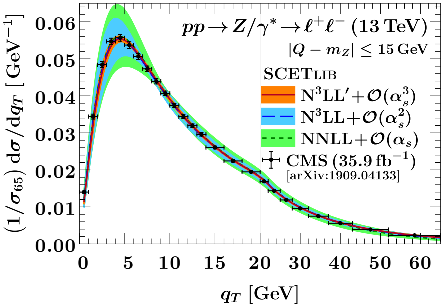
$$f(\alpha) = f_0 + f_1\,\alpha \pm \Delta f \qquad \text{with} \qquad \Delta f = f_1\, b_0\, \alpha^2 + \cdots$$
$$g(\alpha) = g_0 + g_1\,\alpha \pm \Delta g \qquad \text{with} \qquad \Delta g = g_1\, b_0\, \alpha^2 + \cdots$$

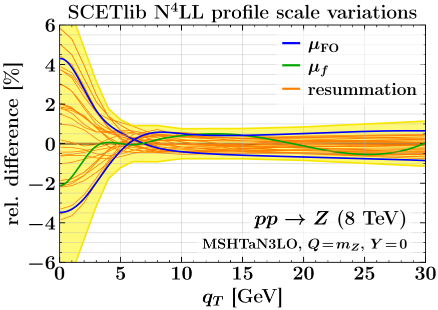
How are Δf and Δg correlated?

- We don't know – the scale variation method simply does not tell us
  - Correlations require a common uncertain parameter
    (or more generally a common source of uncertainty)
  - b_0 (or μ) is not a common or uncertain parameter, we just made it up
  - ✗ A priori, scale variations *do not* imply correct correlations
- Best we can do is *assume* some theoretically motivated but still *ad hoc* correlation model that we impose on Δf and Δg

⇒ Probably the most severe shortcoming of scale variations

Scale Variations for Differential Spectrum.





Now $f(\alpha; x)$ is some differential spectrum in x, e.g. $p_T^Z \equiv q_T$

- Its $\Delta f(x)$ comes from envelope of various scale variations
  - Take $f(\alpha) \equiv f(\alpha; x_1)$ and $g(\alpha) = f(\alpha; x_2)$ to be spectrum at different points in x
  - We don't know their correct correlation
  - ✗ A priori, scale variations *do not* imply correct shape uncertainties

⇒ How to interpret and propagate this envelope?

Theory Nuisance Parameters.

What We Should be Doing.

Step 1: Identify the actual source of uncertainty

$$f(\alpha) = f_0 + f_1\,\alpha \underbrace{+ f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)}_{\text{source of the theory uncertainty}}$$

What We Should be Doing.

Step 1: Identify the actual source of uncertainty

$$f(\alpha) = f_0 + f_1\,\alpha \underbrace{+ f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)}_{\text{source of the theory uncertainty}}$$

Step 2: Explicitly parametrize and include the (leading) source of uncertainty

$\mathrm{N}^{1+1}\mathrm{LO}$: $$f^{\mathrm{predicted}}(\alpha) = f_0 + f_1\,\alpha + f_2(\theta_2)\,\alpha^2$$

- In terms of unknown but well-defined parameters θ_n, which are the *theory nuisance parameters (TNPs)*
  - Simplest: Use f_2 itself: $f_2(\theta_2) \equiv \theta_2$
  - Better: Account for known internal structure of f_2 (color, partonic channels, ...)
- Sufficient to include the next term
  - We always assume that expansion converges, so f_3 is not yet relevant

What We Should be Doing.

Step 1: Identify the actual source of uncertainty

$$f(\alpha) = f_0 + f_1\,\alpha \underbrace{+ f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)}_{\text{source of the theory uncertainty}}$$

Step 2: Explicitly parametrize and include the (leading) source of uncertainty

$\mathrm{N}^{2+1}\mathrm{LO}$: $$f^{\text{predicted}}(\alpha) = f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3(\theta_3)\,\alpha^3$$

- In terms of unknown but well-defined parameters θ_n, which are the *theory nuisance parameters (TNPs)*
  - Simplest: Use f_2 itself: $f_2(\theta_2) \equiv \theta_2$
  - Better: Account for known internal structure of f_2 (color, partonic channels, ...)
- Sufficient to include the next term
  - We always assume that expansion converges, so f_3 is not yet relevant
  - When f_2 becomes known (or strongly constrained), need to include $f_3(\theta_3)$

What We Should be Doing.

Step 1: Identify the actual source of uncertainty

$$f(\alpha) = f_0 + f_1\,\alpha \underbrace{+ f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)}_{\text{source of the theory uncertainty}}$$

Step 2: Explicitly parametrize and include the (leading) source of uncertainty

$\mathrm{N}^{2+1}\mathrm{LO}$: $$f^{\text{predicted}}(\alpha) = f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3(\theta_3)\,\alpha^3$$

- In terms of unknown but well-defined parameters θ_n, which are the *theory nuisance parameters (TNPs)*
  - Simplest: Use f_2 itself: $f_2(\theta_2) \equiv \theta_2$
  - Better: Account for known internal structure of f_2 (color, partonic channels, ...)
- Sufficient to include the next term
  - We always assume that expansion converges, so f_3 is not yet relevant
  - When f_2 becomes known (or strongly constrained), need to include $f_3(\theta_3)$

Step 3: Vary all θ_i to account for correctly correlated theory uncertainty

Application to Drell-Yan p_T Spectrum.

Structure of p_T dependence is known to all orders (up to small power corrections)

$$p_T \frac{d\sigma}{dp_T} = \Big[H \times B_a \otimes B_b \otimes S\Big](\alpha_s; L \equiv \ln p_T/m_Z) + \mathcal{O}(p_T^2/m_Z^2)$$

- Each factor depends on a *boundary condition* and *anomalous dimensions* (solution to a coupled RGE system)

$$F(\alpha_s, L) = F(\alpha_s)\, \exp \int_0^L dL' \Big\{\Gamma[\alpha_s(L')]\, L' + \gamma_F[\alpha_s(L')]\Big\}$$

- We're left with several independent (scalar) perturbative series (plus QCD beta function and splitting functions)
  - N^{2+1}LL: $F(\alpha_s) = F_0 + \alpha_s F_1 + \alpha_s^2 F_2(\theta_2^F) + \cdots$

    $$\gamma_F(\alpha_s) = \alpha_s \gamma_{F0} + \alpha_s^2 \gamma_{F1} + \alpha_s^3 \gamma_{F2}(\theta_2^\gamma) + \cdots$$

    $$\Gamma(\alpha_s) = \alpha_s \Gamma_0 + \alpha_s^2 \Gamma_1 + \alpha_s^3 \Gamma_2 + \alpha_s^4 \Gamma_3(\theta_3^\Gamma) + \cdots$$

  - analogously for N^{3+1}LL, etc.

$\Rightarrow$ Remaining task: How exactly to define and vary the θ_i?

Theory Uncertainties via TNPs.

ML fits: $$L(y, \theta_i) = P(d|y, \theta_i) \times \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(u_i - \theta_i)^2}{2\sigma_i^2}\right]$$

χ^2 fits: $$\chi^2(y, \theta_i) = \sum_d \frac{[d - f_d^{\mathrm{predicted}}(y, \theta_i)]^2}{\sigma_d^2} + \sum_i \frac{(u_i - \theta_i)^2}{\sigma_i^2}$$

Standard method of including systematic unc. via nuisance parameters

- Auxiliary (real or imagined) measurements provide constraint on θ_i
  - $u_i =$ best estimate of θ_i (from an actual measurement or our best guess)
  - $\sigma_i =$ uncertainty on u_i (the estimated "systematic uncertainty")
- We *do not* need a precise estimate of the true value for each θ_i
  - Typically our best-guess central value will be $u_i = 0$
  - Generically we can still have $f_2(\theta_2 = 0) \neq 0$
- We *do* need an estimate of σ_i for each θ_i (the systematic "theory uncertainty")
  - i.e., how is θ_i allowed to vary around u_i (if otherwise unconstrained)

$\Rightarrow$ Sufficient to understand the *typical, generic size* of θ_i (or equivalently f_2)

TNP Parameterization.

cross sections, boundary conditions: $$F(\alpha_s) = 1 + \sum_{n=1} \Big(\frac{\alpha_s}{4\pi}\Big)^n F_n$$

anomalous dimensions: $$\gamma(\alpha_s) = \sum_{n=0} \Big(\frac{\alpha_s}{4\pi}\Big)^{n+1} \gamma_n$$

- Parametrize n-loop coefficients as

$$F_n(\theta_n) = 4C_r(4C_A)^{n-1}(n-1)!\,\theta_n^F(n_f)$$

$$\gamma_n(\theta_n) = 2C_r(4C_A)^n\,\theta_n^\gamma(n_f)$$

  - $C_r C_A^{n-1}$ = leading n-loop color factor

- Expect θ_n to be $\mathcal{O}(1)$ numbers $\quad\rightarrow\quad \theta_i = 0 \pm \mathcal{O}(1)$
  - We can of course check by looking at known n-loop coefficients

TNP Parameterization.

cross sections, boundary conditions: $$F(\alpha_s) = 1 + \sum_{n=1} \Big(\frac{\alpha_s}{4\pi}\Big)^n F_n$$

anomalous dimensions: $$\gamma(\alpha_s) = \sum_{n=0} \Big(\frac{\alpha_s}{4\pi}\Big)^{n+1} \gamma_n$$

- Parametrize n-loop coefficients as

$$F_n(\theta_n) = 4C_r(4C_A)^{n-1}(n-1)!\,\theta_n^F(n_f)$$

$$\gamma_n(\theta_n) = 2C_r(4C_A)^n\,\theta_n^\gamma(n_f)$$

  - $C_r C_A^{n-1}$ = leading n-loop color factor

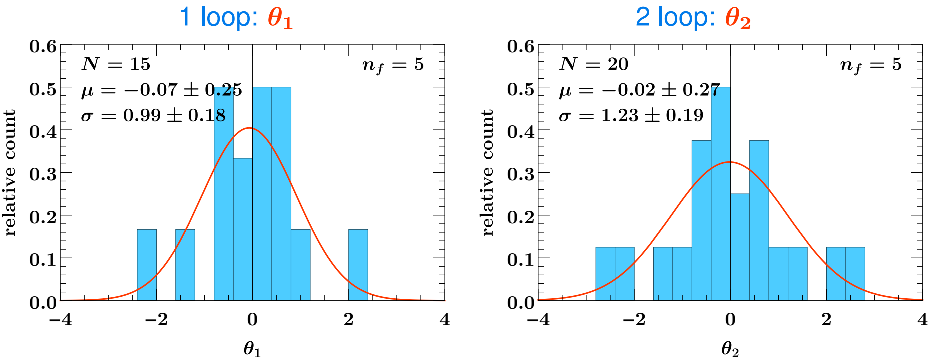
- Expect θ_n to be $\mathcal{O}(1)$ numbers $\rightarrow$ $\theta_i = 0 \pm \mathcal{O}(1)$
  - We can of course check by looking at known n-loop coefficients

In statistics terms: QCD has an (unknown) population of θ_n^F and θ_n^γ (for each n)

- How are they distributed?

$\Rightarrow$ We can find out from population sample

Distribution of TNPs: Boundary Conditions.



$$F_n(\theta_n) = 4C_r(4C_A)^{n-1}(n-1)!\,\theta_n^F(n_f)$$

Estimate θ_n^F population distribution from sample of known, independent series

✓ Good fit to a Gaussian with mean 0 and variance 1

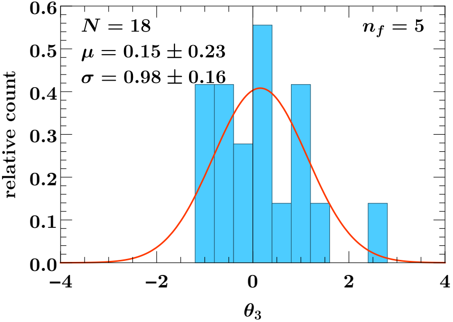
✓ Provides well-defined and reliable estimate: $u_i = 0$ with $\sigma_i = 1$

Fineprint: Strong n_f dependence, for $n_f \to 0$ variance increases to $\sigma_i \sim n$

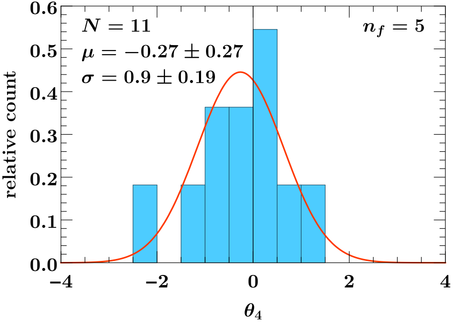
$\to$ Nontrivial correlation between different n_f, can be estimated from sample

Distribution of TNPs: Boundary Conditions.

3 loop: θ_3



4 loop: θ_4



$$F_n(\theta_n) = 4C_r(4C_A)^{n-1}(n-1)!\,\theta_n^F(n_f)$$

Estimate θ_n^F population distribution from sample of known, independent series

✓ Good fit to a Gaussian with mean 0 and variance 1

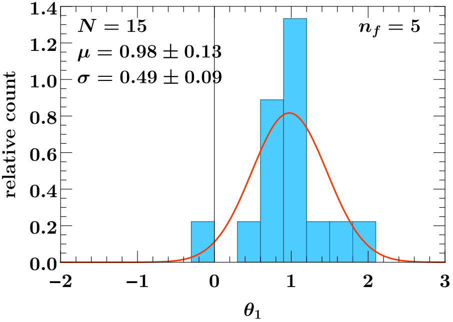
✓ Provides well-defined and reliable estimate: $u_i = 0$ with $\sigma_i = 1$

Fineprint: Strong n_f dependence, for $n_f \to 0$ variance increases to $\sigma_i \sim n$

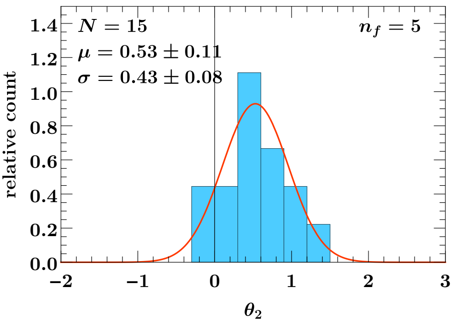
→ Nontrivial correlation between different n_f, can be estimated from sample

Distribution of TNPs: Anomalous Dimensions.

2 loop: θ_1



3 loop: θ_2



$$\gamma_n(\theta_n) = 2C_r(4C_A)^n\,\theta_n^\gamma(n_f)$$

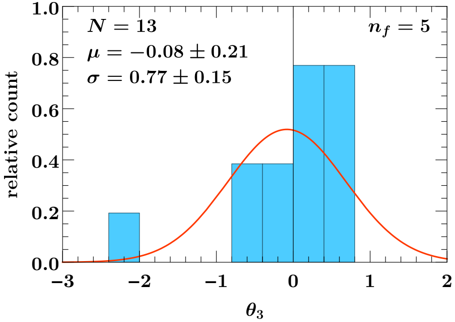
Estimate θ_n^γ population distribution from sample of known, independent series

✓ Good fit to a Gaussian, now with variance ~ 0.5 and mean $\neq 0$

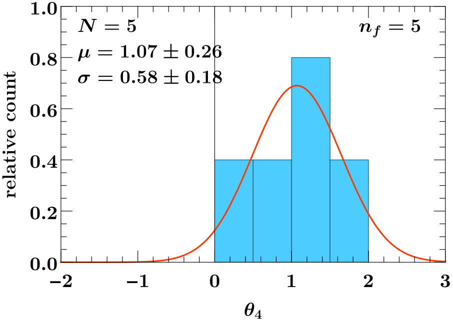
- In the following will use $u_i = 0$ with $\sigma_i = 1$
  (just for simplicity, as we will mostly care about θ_3)

Distribution of TNPs: Anomalous Dimensions.

4 loop: θ_3



5 loop: θ_4



$$\gamma_n(\theta_n) = 2C_r(4C_A)^n\, \theta_n^\gamma(n_f)$$

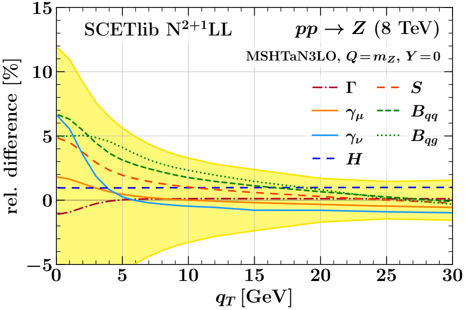
Estimate θ_n^γ population distribution from sample of known, independent series

✓ Good fit to a Gaussian, now with variance ~ 0.5 and mean $\neq 0$

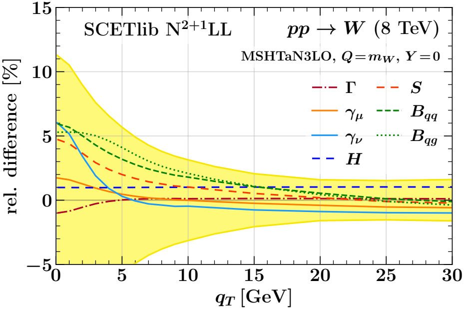
- In the following will use $u_i = 0$ with $\sigma_i = 1$
  (just for simplicity, as we will mostly care about θ_3)

Application to Drell-Yan p_T Spectrum.

relative impact for Z



relative impact for W



Vary each θ_i independently

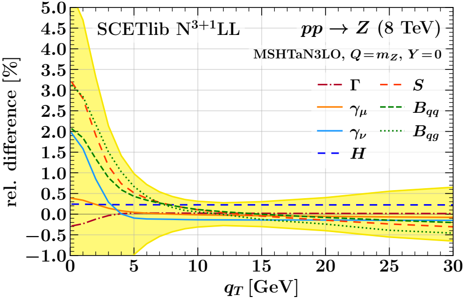
- ✓ Add in quadrature to get total uncertainty
- ✓ Correlations in p_T and Q and between W and Z are correctly captured
  - Each θ_i fully correlated, different θ_i uncorrelated

Fineprint:

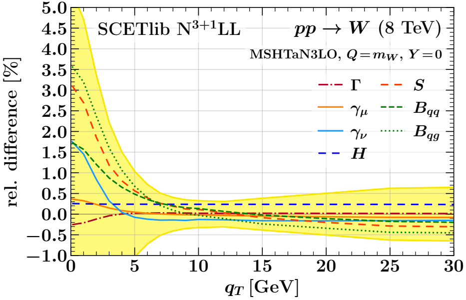
- Beam boundary conditions B_{qj}: Using $f_n = (0 \pm 1.5) \times f_n^{\mathrm{true}}$ here
- DGLAP splitting functions not varied here (count as noncusp anom. dims.)
- Hard boundary conditions H: No singlet corrections (enter for Z but not W)

Application to Drell-Yan p_T Spectrum.

relative impact for Z



relative impact for W



Vary each θ_i independently

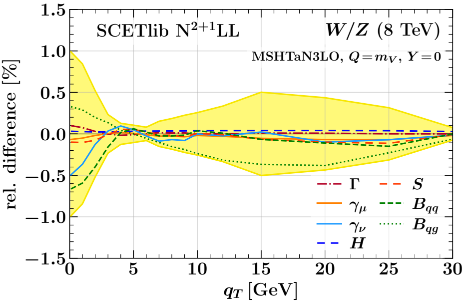
- ✓ Add in quadrature to get total uncertainty
- ✓ Correlations in p_T and Q and between W and Z are correctly captured
  - Each θ_i fully correlated, different θ_i uncorrelated

Fineprint:

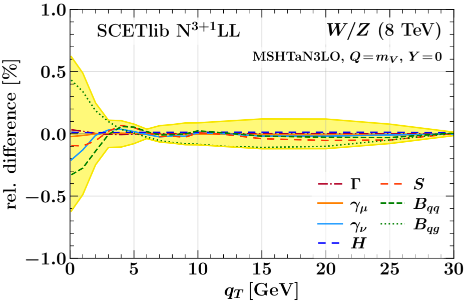
- Beam boundary conditions B_{qj}: Using $f_n = (0 \pm 1.5) \times f_n^{\text{true}}$ here
- DGLAP splitting functions not varied here (count as noncusp anom. dims.)
- Hard boundary conditions H: No singlet corrections (enter for Z but not W)

Application to Drell-Yan p_T Spectrum.

relative impact for W/Z



relative impact for W/Z



Vary each θ_i independently

- ✓ Add in quadrature to get total uncertainty
- ✓ Correlations in p_T and Q and between W and Z are correctly captured
  - Each θ_i fully correlated, different θ_i uncorrelated

Fineprint:

- Beam boundary conditions B_{qj}: Using $f_n = (0 \pm 1.5) \times f_n^{\mathrm{true}}$ here
- DGLAP splitting functions not varied here (count as noncusp anom. dims.)
- Hard boundary conditions H: No singlet corrections (enter for Z but not W)

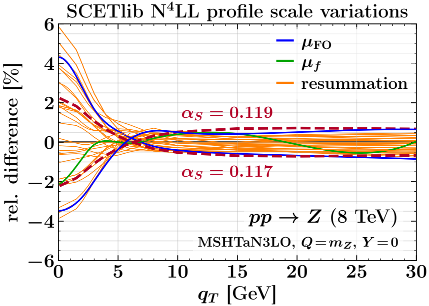
Extracting α_s from Z p_T spectrum.

Toy Study Setup.

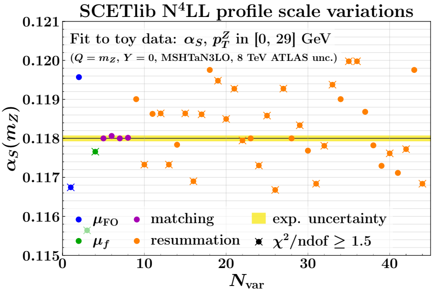
We perform a toy study of fitting α_s from $Z\ p_T$ using Asimov fits

- Goals
  - Obtain *expected* theory uncertainty on α_s in a controlled setting
  - Try out TNPs in real life ...
- Theory input
  - SCETlib resummed-only at N^{2+1}LL and N^{3+1}LL
  - Nonperturbative model at small p_T – not discussed here
  - Neglecting $\mathcal{O}(p_T^2/m_Z^2)$ (FO matching), quark-mass effects, QED effects
    $\rightarrow$ Okay for toy studies, important to include in fit to real data
- Toy data
  - Central values from central theory prediction with $\alpha_s(m_Z) = 0.118$
  - Uncertainties from recent ATLAS 8 TeV inclusive measurement including full correlations (integrated over $|Y| \leq 1.6$)
  - 9 p_T points in [0, 29] GeV corresponding to ATLAS bins
    (Fixed $Q = m_Z$, $Y = 0$ for simplicity, integrating in q_T, Q, Y makes practically no difference)

Scanning over Scale Variations.



$\Rightarrow$



Each variation provides a trial 100% (anti)correlated correlation model

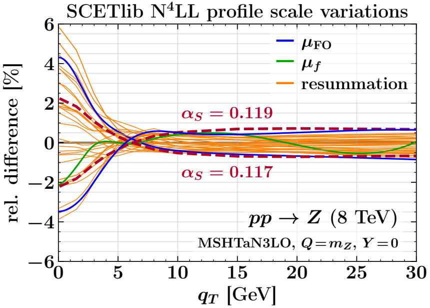
- Correlation model strongly impacts the result (as expected ...)
  - How to interpret this?

    Sum of envelopes: $\boldsymbol{\Delta_{\rm total} = \sqrt{\Delta^2_{\rm FO} + \Delta^2_{\rm resum} + \Delta_{\rm match}} \sim 2.6 \times 10^{-3}}$

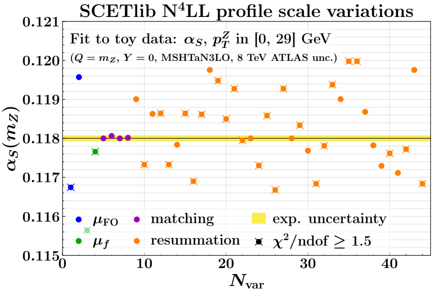
    Max envelope: $\boldsymbol{\Delta_{\rm total} \sim 2.1 \times 10^{-3}}$

  - Fit should be able to decide whether to allow or constrain some theory excursion vs. changing fitted POIs (here α_s) to compensate

Scanning over Scale Variations.



⇒



Each variation provides a trial 100% (anti)correlated correlation model

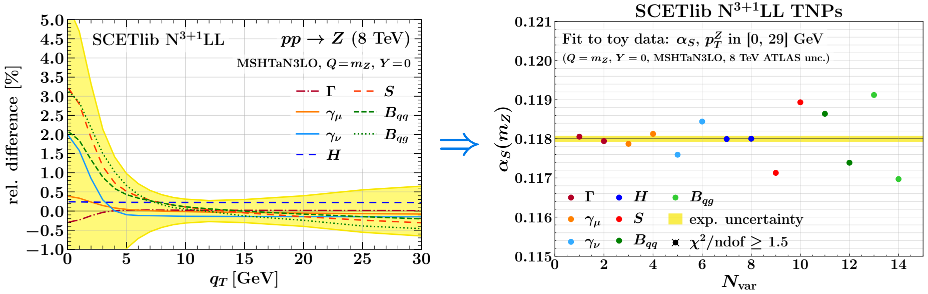
- Correlation model strongly impacts the result (as expected ...)
  - How to interpret this?

    Sum of envelopes: $\Delta_{\text{total}} = \sqrt{\Delta^2_{\text{FO}} + \Delta^2_{\text{resum}} + \Delta_{\text{match}}} \sim 2.6 \times 10^{-3}$

    Max envelope: $\Delta_{\text{total}} \sim 2.1 \times 10^{-3}$
  - Fit should be able to decide whether to allow or constrain some theory excursion vs. changing fitted POIs (here α_s) to compensate

⇒ Upshot: Scale variations are just not sufficient for this purpose

Scanning with TNPs.



- Repeat fit for each TNP variation
  - TNPs correctly capture independent uncertainty sources and correlations
  - Well-defined interpretation

Sum in quadrature: $\Delta_{\rm total} = 1.59 \times 10^{-3}$

- Still does not let the fit decide between moving theory vs. α_s
  - Amounts to neglecting possible correlations between TNPs and α_s

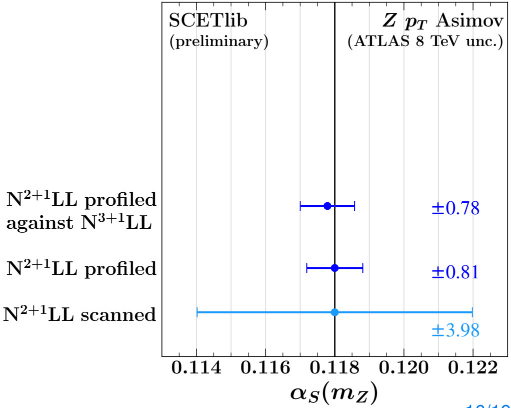
Profiling TNPs.

TNPs are real parameters, so it is perfectly okay to profile them in the fit

- Include all TNPs in the fit
  - With Gaussian prior constraint of $\theta_i = 0 \pm 1$
  - Accounts for correlations between theory uncertainties and fitted POIs
  - Allows data to constrain TNPs and reduce theory uncertainty

- At N^{2+1}LL:
  - Huge reduction due to profiling
  - Overconstrained?

(Uncertainties in units of 10^{-3})



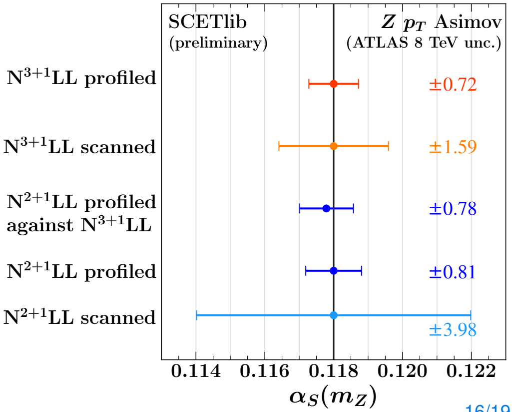
Profiling TNPs.

TNPs are real parameters, so it is perfectly okay to profile them in the fit

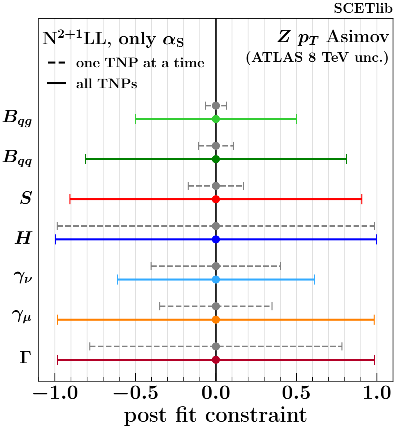
- Include all TNPs in the fit
  - With Gaussian prior constraint of $\theta_i = 0 \pm 1$
  - Accounts for correlations between theory uncertainties and fitted POIs
  - Allows data to constrain TNPs and reduce theory uncertainty

(Uncertainties in units of 10^{-3})

- At N^{2+1}LL:
  - Huge reduction due to profiling
  - Overconstrained?
- At N^{3+1}LL:
  - Still sizeable reduction
  - More reasonable ...

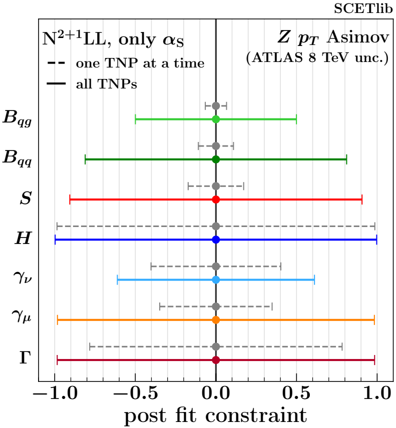


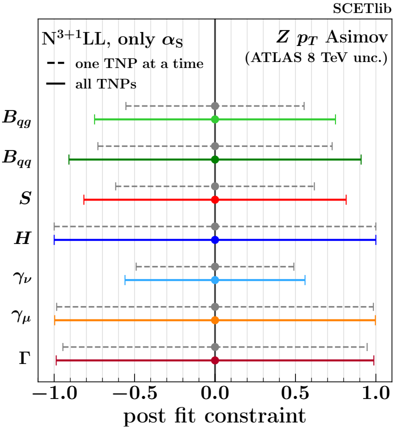
Constraints on TNPs.



- At N^{2+1}LL: TNPs are strongly constrained by data
  - Theory accuracy is insufficient → next order becomes relevant

Constraints on TNPs.





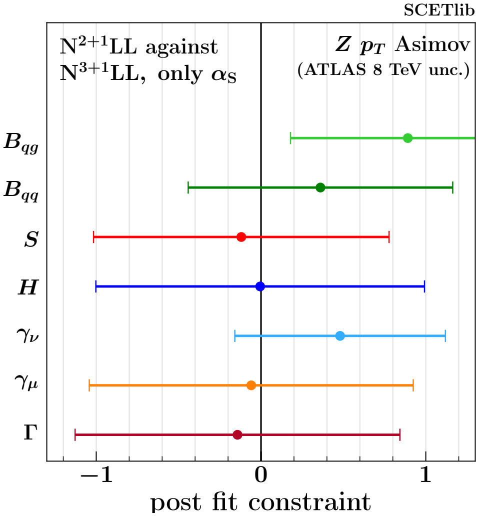
- At N^{2+1}LL: TNPs are strongly constrained by data
  - Theory accuracy is insufficient $\to$ next order becomes relevant
- At N^{3+1}LL: TNPs are somewhat constrained by data
  - Theory accuracy likely sufficient $\to$ next order not relevant (yet)

Constraints on TNPs.

Another test we can do

- Profile N^{2+1}LL theory model against N^{3+1}LL toy data
- Some TNPs get strongly pulled

⇒ Indicates again that N^{2+1}LL is insufficient for data precision



Summary.

Interpretation of LHC precision measurements requires theory predictions with reliable uncertainties and in particular correct correlations

Scale variations

- Neither particularly reliable nor can they do correlations
  - → One cannot rely on them for shape uncertainties which unfortunately is exactly what is often done
  - → Insufficient for extracting α_s from small-p_T Z, since correlations are critical

Theory nuisance parameters overcome many limitations of scale variations

- Provide truly parametric theory uncertainties that
  - ✓ Encode correct correlations
  - ✓ Can be consistently propagated everywhere (fits, MCs, neural networks, ...)
  - ✓ Can be consistently profiled and constrained by data
- Price to pay: Obviously not as “easy and cheap” as scale variations
- First toy studies show that TNPs work as advertised
  - ► Should allow for competitive α_s extraction by profiling theory uncertainties

Acknowledgments.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE)

erc

European Research Council

Established by the European Commission