N3PDF

Machine Learning • PDFs • QCD

European Research Council

Established by the European Commission

Resummation and Machine Learning Techniques Towards Precision Phenomenology at the LHC

Tanjona R. Rabemananjara

Supervisor: **Stefano Forte**

PhD Defense, Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy. December 20th, 2021.

Acknowledgment: This project has received funding from the European Unions Horizon 2020 research and innovation programme under the grant agreement number **740006**.

UNIVERSITÀ DEGLI STUDI DI MILANO

DIPARTIMENTO DI FISICA

INFN

Istituto Nazionale di Fisica Nucleare

OUTLINE

Combining Resummations

Transverse Momentum Distribution

$$h_1 + h_2 \to F(M) + X \qquad (M : \text{Invariant Mass})$$

QCD factorization theorem as a main guiding principle:

$$\frac{d\sigma_F}{dp_{T,F}^2}\left(p_{T,F},\alpha_s\right) = \frac{1}{M^2}\sum_{a,b}\int_\tau^1 \frac{d}{dx}\mathcal{L}_{ab}\left(\frac{\tau}{x}\right)\frac{d\hat{\sigma}_{ab,H}}{dp_{T,F}^2}(x,p_{T,F},\alpha_s)$$

in which the partonic part is expanded as a series in α_s

$$\frac{d\hat{\sigma}_{ab,F}}{dp_{T,F}^2}\left(x,p_{T,F},\alpha_s\right) = \sigma_F^{(0)}\left\{\underbrace{1}_{\text{LO}} + \underbrace{\alpha_s\mathcal{C}_{ab}^{(1)}}_{\text{NLO}}\left(x,p_{T,F}\right) + \underbrace{\alpha_s^2\mathcal{C}_{ab}^{(2)}}_{\text{NNLO}}\left(x,p_{T,F}\right) + \cdots\right\}$$

Perturbative computations assume that $\mathcal{C}_{ab}^{(n)}$ are **WELL-BEHAVED**. **What happens when the smallness of α_s is compensated by large logarithms $(\alpha_s^n\mathrm{L}^m \sim 1)$?**

$$\mathcal{C}^{(1)} = c_{21}\mathrm{L}^2 + c_{11}\mathrm{L}$$
$$\mathcal{C}^{(2)} = c_{42}\mathrm{L}^4 + c_{32}\mathrm{L}^3 + c_{22}\mathrm{L}^2 + c_{21}\mathrm{L}$$

$$\mathrm{L} = \ln\left(\frac{M^2}{p_{T,F}^2}\right)$$

Resumming logarithmic enhancement to all order

Conjugate spaces:

- Bypass convolution $\Longrightarrow$ Mellin Space:

$$\sum_{a,b}\int_{\tau}^{1}\frac{d}{dx}\mathcal{L}_{ab}\left(\frac{\tau}{x}\right)\frac{d\hat{\sigma}_{ab}}{dp_T^2}(x)\longrightarrow\sum_{ab}\mathcal{L}(N)\frac{d\hat{\sigma}_{ab}}{dp_T^2}(N)$$

- Factorize δ-constraint $\Longrightarrow$ Fourier Space:

$$\int d^2\vec{p}_T\exp\left(-\mathrm{i}\vec{b}\cdot\vec{p}_T\right)\delta\left(\vec{p}_T-\sum_{k=1}^{n}\vec{p}_{T,k}\right)\longrightarrow\prod_{k=1}^{n}\exp\left(-\mathrm{i}\vec{b}\cdot\vec{p}_{T,k}\right)$$

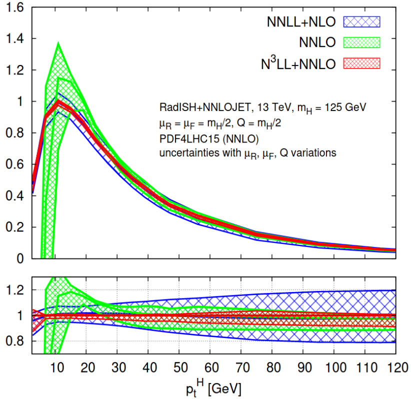
Exponentiation:

$$\frac{d\hat{\sigma}_{ab,F}}{dp_T^2}(N,b)=\sigma_F^{(0)}\mathcal{H}(N)\exp\left(\underbrace{\mathrm{L}f_1(\alpha_s\mathrm{L})}_{\text{LL}}+\underbrace{f_2(\alpha_s\mathrm{L})}_{\text{NLL}}+\underbrace{\alpha_s f_3(\alpha_s\mathrm{L})}_{\text{NNLL}}+\cdots\right)$$

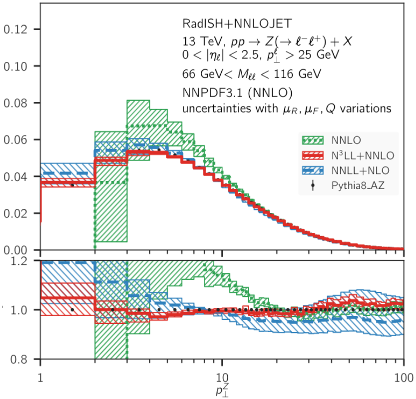
where the logs become $\mathrm{L}=\ln\left(b^2M^2\right)$.

State of the art for small-p_T resummation

[Bizon et al. ('18)]

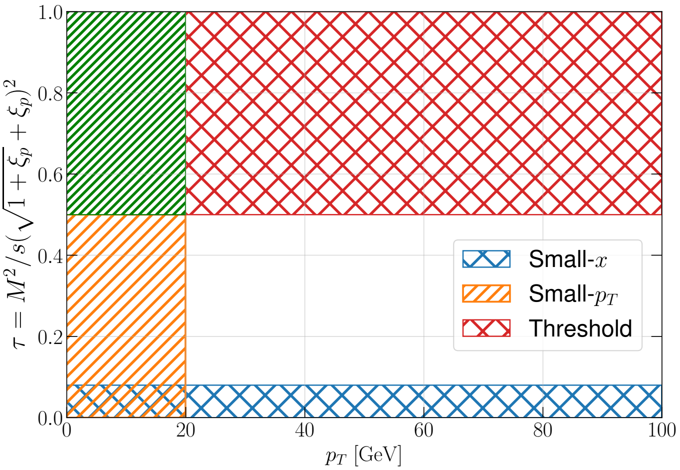


[Bizon et al. ('19)]



- **WELL-BEHAVED** distribution at small-p_T. Matched results yield good predictions in medium and large-p_T
- N3LL+NNLO corrections amount to about **5-10%** in the Jacobian peak
- Very **good convergence** of the predictions at different perturbative orders: N3LL+NNLO bands are entirely contained in NNLL+NLO

What about the other logarithmic enhancement?



$x \to 1$: **Threshold Limit** $\quad \xi_p \equiv p_T^2/M^2 \to 0$: **Small-pT Limit** $\quad x \to 0$: **High-energy Limit**

(Consistently) Supplement small-p_T resummation with threshold logarithms

(1) **Reproduces standard small-p_T** (CFG) resummation in the small-p_T limit

(2) **Reproduces threshold resummation** to some given logarithmic accuracy in the soft limit

(3) **Leads to the total cross section upon integration over p_T**:

$$\exp\left\{S(\alpha_s^n \mathrm{L}^m)\right\}_{\mathrm{L}\to 0} = 1 \longrightarrow \int_0^\infty dp^2_{p_T,F}\left(\frac{d\hat{\sigma}_F}{dp^2_{T,F}}\right) = \hat{\sigma}_F^{\mathrm{TOT}}$$

(In CFG's formulation, this is enforced by the **Unitarity Constraint**)

Some previous attempts

Joint resummation for Higgs (NLL) and VB (NNLL) **[Kulesza et al. ('03)][Marzani & Theeuwes ('17)]**

Soft and collinear logs **jointly** resummed in $\ln \chi(N, b)$

$$\chi(N, b) = \frac{bM}{2\mathrm{e}^{-\gamma_E}} + \frac{N\mathrm{e}^{\gamma_E}}{1 + \eta \frac{bM}{2N}}$$

(✓) **Reproduces standard small-p_T resummation**

(✓) **Leads to the total cross section upon integration over p_T**

Some previous attempts

Joint resummation for Higgs (NLL) and VB (NNLL) **[Kulesza et al. ('03)][Marzani & Theeuwes ('17)]**

Soft and collinear logs **jointly** resummed in $\ln \chi(N, b)$

$$\chi(N, b) = \frac{bM}{2\mathrm{e}^{-\gamma_E}} + \frac{N\mathrm{e}^{\gamma_E}}{1 + \eta \frac{bM}{2N}}$$

(✓) **Reproduces standard small-p_T resummation**

(✗) **Reproduces threshold resummation**

(✓) **Leads to the total cross section upon integration over p_T**

Some previous attempts

Joint resummation for Higgs (NLL) and VB (NNLL) **[Kulesza et al. ('03)][Marzani & Theeuwes ('17)]**

Soft and collinear logs **jointly** resummed in $\ln \chi(N, b)$

$$\chi(N, b) = \frac{bM}{2\mathrm{e}^{-\gamma_E}} + \frac{N\mathrm{e}^{\gamma_E}}{1 + \eta \frac{bM}{2N}}$$

(✓) **Reproduces standard small-p_T resummation**

(✗) **Reproduces threshold resummation**

(✓) **Leads to the total cross section upon integration over p_T**

SCET **[Li et al. ('16)] [Lustermans et al. ('16)]**

(✓) **Reproduces standard small-p_T resummation**

(✓) **Reproduces threshold resummation**

Some previous attempts

Joint resummation for Higgs (NLL) and VB (NNLL) **[Kulesza et al. ('03)][Marzani & Theeuwes ('17)]**

Soft and collinear logs **jointly** resummed in $\ln \chi(N,b)$

$$\chi(N,b) = \frac{bM}{2\mathrm{e}^{-\gamma_E}} + \frac{N\mathrm{e}^{\gamma_E}}{1+\eta\frac{bM}{2N}}$$

(✓) **Reproduces standard small-p_T resummation**

(✗) **Reproduces threshold resummation**

(✓) **Leads to the total cross section upon integration over p_T**

SCET **[Li et al. ('16)] [Lustermans et al. ('16)]**

(✓) **Reproduces standard small-p_T resummation**

(✓) **Reproduces threshold resummation**

(✗) **Leads to the total cross section upon integration over p_T**

Soft-improved small-p_T (SIpT) resummation

MAIN INGREDIENTS:

Factorization of the Phase Space **[Forte & Muselli & Ridolfi ('17)]**

The factorization of the Phase Space is performed by taking the small-p_T (or $b \to \infty$) while keeping N/b in order to preserve the threshold behavior at the inclusive level.

New treatment of the Evolutions

In order to compute N^{k}LL soft-improved small-p_T resummation, the large-N behavior of the evolution has to be included up to N^{k}LO (as opposed to N^{k-1}LO in standard small-p_T).

New prescription to compute the Fourier-Mellin inverse (?)

MAIN RESULTS: **[Forte & Muselli & Ridolfi ('17)]**

New argument for the Logarithms

The modified argument takes into account both the small-$p_{T,H}$ and threshold limit

$$\chi = \frac{b^2 M^2}{b_0^2} + \bar{N}^2 \quad \text{with} \quad b_0 = 2\mathrm{e}^{-\gamma_E} \quad \text{and} \quad \bar{N}^2 = N\mathrm{e}^{\gamma_E}$$

which **interpolates** between $b^2 Q^2$ when $b \to \infty$ and $\bar{N}^2$ when $b = 0$.

Structure of the SIpT resummed expression

Resummed expression:

$$\frac{\mathrm{d}\hat{\sigma}^{\text{cons}}_{ab}}{\mathrm{d}\xi_p}(N,\chi,\alpha_s) = \sigma^0_c\,\overline{\mathrm{H}}_c\left(\frac{\bar{N}^2}{\chi},\alpha_s\left(Q^2\right)\right) C_{ci}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right)\right) C_{cj}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right)\right)$$
$$U_{ia}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right),\alpha_s\left(\mu_F^2\right)\right) U_{jb}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right),\alpha_s\left(\mu_F^2\right)\right)\exp\left(S_c\left(N,\chi,\mu_R^2\right)\right)$$

Hard function $\bar{\mathrm{H}}$:

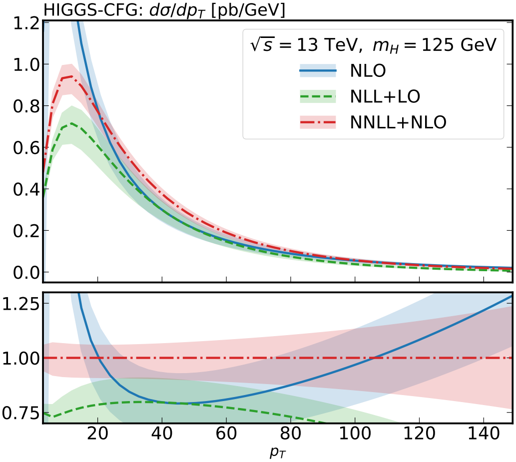
$$\overline{\mathrm{H}}_c\left(\frac{\bar{N}^2}{\chi},\alpha_s\right) = \mathrm{H}_c\left(\alpha_s\right) + A^{(1)}_c\,\mathrm{Li}_2\left(\frac{\bar{N}^2}{\chi}\right)\alpha_s + \mathcal{O}\left(\alpha_s^3\right)$$

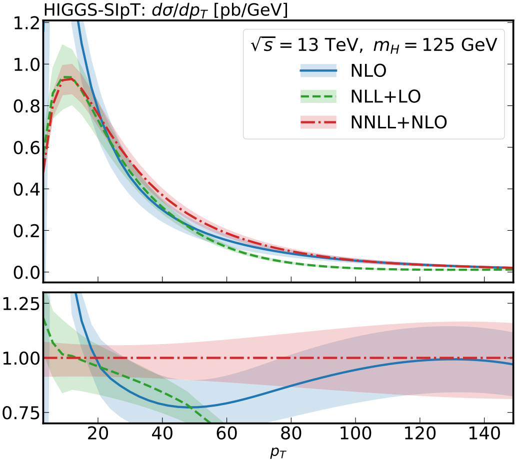
Sudakov exponent (Process-Independent):

$$S_c(\chi,N) = -\int_{\frac{Q^2}{\chi}}^{Q^2}\frac{dq^2}{q^2}\left[A_c\left(\alpha_s\left(q^2\right)\right)\ln\frac{Q^2}{q^2} + B_c\left(\alpha_s\left(q^2\right)\right)\right] + \alpha_s^2\int_{\frac{Q^2}{N^2}}^{Q^2}\beta_0 A^{(1)}_c\mathrm{Li}_2\left(\frac{\bar{N}^2}{\chi}\right)$$

Landau Pole+Additional Singularities $\Longrightarrow\ \nexists$ Fourier/Mellin Inverse $\Longrightarrow$ Borel Summation

CFG vs. SIpT: Results for Higgs at LHC

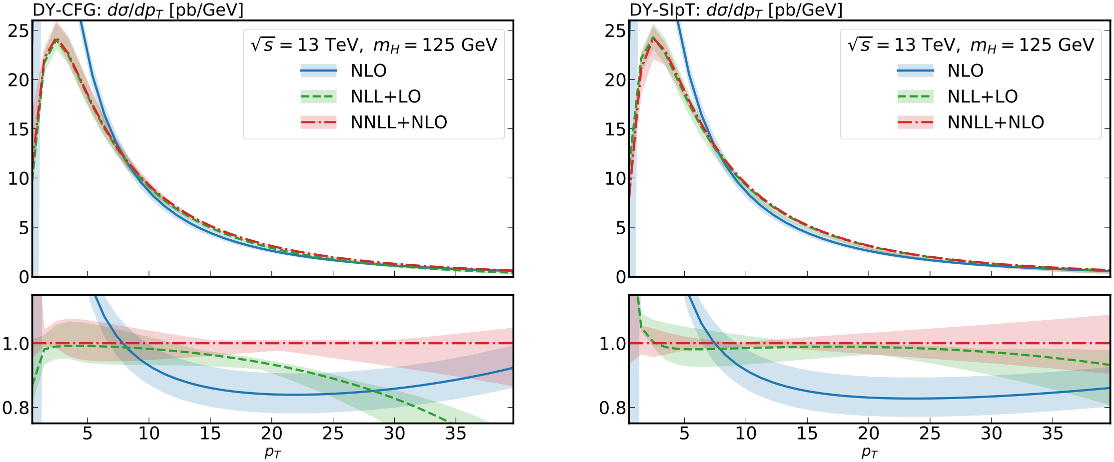




◇ **Faster perturbative convergence in the small-p_T region** ◇

◇ **Improvement of the large-p_T behaviour** ◇

CFG vs. SIpT: Results for Z Boson at Tevatron II



◇ **Improvement of the large-p_T behaviour** ◇

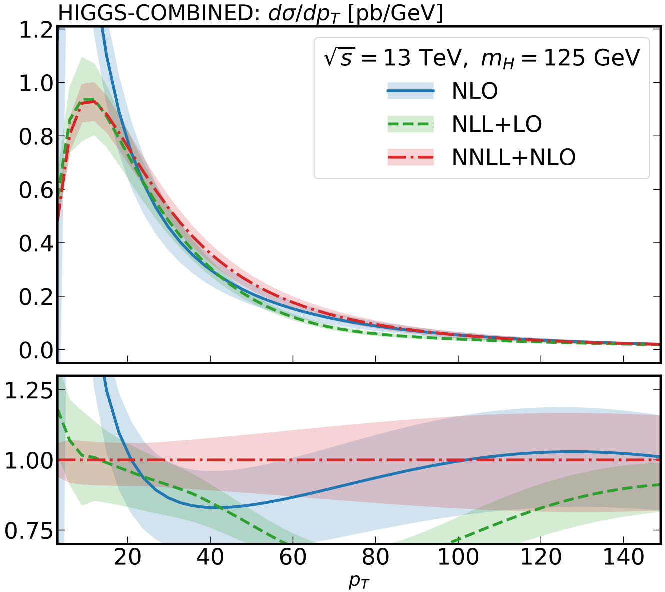
Fully Combined Resummation

- **Not all soft logarithms are taken into account by SIpT**: these are soft logarithms emitted at large angle & p_T-suppressed initial state radiations.
- Therefore, one needs to combine it with the pure threshold resummed expression through a **profile matching function** such that the resulting expression reproduces small-p_T and threshold resummation in the resp. limit.

$$\frac{\mathrm{d}\hat{\sigma}_{ab}}{\mathrm{d}\xi_p}(N,\xi_p,\alpha_s) = \left(1 - \mathrm{T}\left(N,\xi_p\right)\right)\frac{\mathrm{d}\hat{\sigma}^{\mathrm{tr}\star}_{ab}}{\mathrm{d}\xi_p}(N,\xi_p,\alpha_s) + \mathrm{T}\left(N,\xi_p\right)\frac{\mathrm{d}\hat{\sigma}^{\mathrm{th}}_{ab}}{\mathrm{d}\xi_p}(N,\xi_p,\alpha_s)$$

- For small-p_T, T gets rid of the $\xi_p \to 0$ **singularity** and only keeps $\mathrm{d}\hat{\sigma}^{\mathrm{tr}\star}_{ab}$ which contributes to the total cross-section.
- For finite p_T and large-N, T gets rid of the $N \to \infty$ **singularity** and only keeps $\mathrm{d}\hat{\sigma}^{\mathrm{th}}_{ab}$ which reproduces the soft behavior.

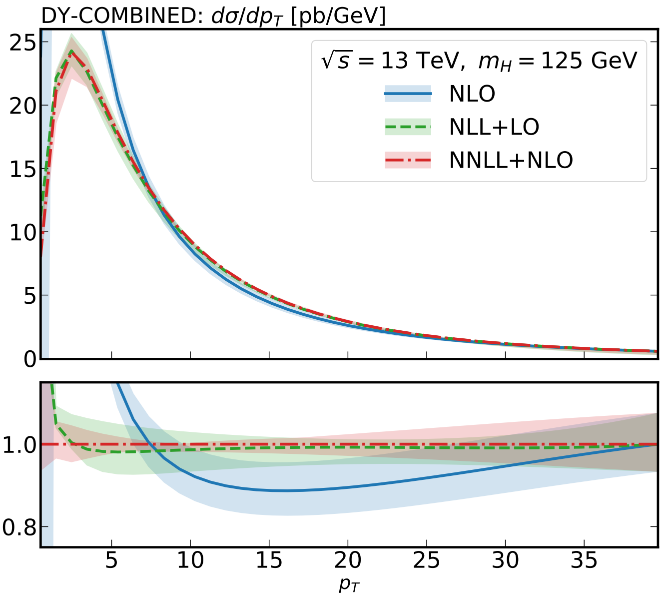
CFG vs. Combined: Results for Higgs at LHC



Faster perturbative convergence around Jacobian peak

Better agreement with FO predictions in medium and large-p_T

CFG vs. Combined: Results for Z Boson at Tevatron II



Better agreement with FO predictions in medium and large-p_T

Missing Higher Orders (MHO)

Predictions in Perturbation Theory

An observable Σ is computed in perturbation theory as:

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

The perturbative expansions are asymptotic to Σ, i.e. (up to some order) increasing in powers of α_s improves the approximation.

$$\Sigma \simeq \Sigma_{\mathrm{N}^n\mathrm{LO}} + \Delta\Sigma_{\mathrm{MHO}}$$

How to estimate $\Delta\Sigma_{\mathrm{MHO}}$?

Predictions in Perturbation Theory

An observable Σ is computed in perturbation theory as:

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

The perturbative expansions are asymptotic to Σ, i.e. (up to some order) increasing in powers of α_s improves the approximation.

$$\Sigma \simeq \Sigma_{\mathrm{N}^n\mathrm{LO}} + \Delta\Sigma_{\mathrm{MHO}}$$

How to estimate $\Delta\Sigma_{\mathrm{MHO}}$?

Renormalization in QFT introduces an unphysical dependence μ, and despite the fact that RGE states that physical observables are independent of μ ($\mu\partial\Sigma/\partial\mu = 0$), residual μ-dependence appear in perturbative computations.

$$\mu\frac{\partial}{\partial\mu}\Sigma_{\mathrm{N}^n\mathrm{LO}} = \mathcal{O}(\alpha_s^{n+1}) = \mathcal{O}\left(\Delta\Sigma_{\mathrm{MHO}}\right)$$

An observable Σ is computed in perturbation theory as:

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

The perturbative expansions are asymptotic to Σ, i.e. (up to some order) increasing in powers of α_s improves the approximation.

$$\Sigma \simeq \Sigma_{\mathrm{N}^n\mathrm{LO}} + \Delta\Sigma_{\mathrm{MHO}}$$

How to estimate $\Delta\Sigma_{\mathrm{MHO}}$?

Renormalization in QFT introduces an unphysical dependence μ, and despite the fact that RGE states that physical observables are independent of μ ($\mu\partial\Sigma/\partial\mu = 0$), residual μ-dependence appear in perturbative computations.

$$\mu \frac{\partial}{\partial \mu} \Sigma_{\mathrm{N}^n\mathrm{LO}} = \mathcal{O}(\alpha_s^{n+1}) = \mathcal{O}\left(\Delta\Sigma_{\mathrm{MHO}}\right)$$

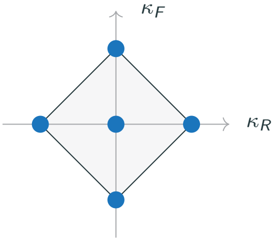
$\Longrightarrow$ **Use the unphysical scale μ to probe MHO**

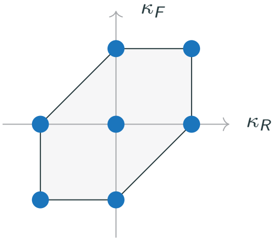
Scale variation prescription

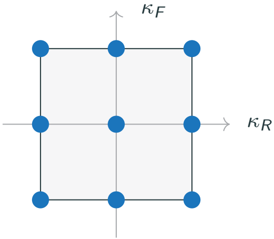
CANONICAL METHOD: Variation by a factor of 2 around a **central scale** μ_0.

$$\Sigma \simeq \Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu_0) \pm \max_{\substack{\mu_{\min} \leq 2\mu_0 \\ 2\mu_{\max} \geq \mu_0}} |\Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu) - \Sigma_{\mathrm{N}^n\mathrm{LO}}(\mu_0)|$$

For a multi-scale process involving the **renormalization scale** ($\mu_R = \kappa_R \mu_0$) and the **factorization scale** ($\mu_F = \kappa_F \mu_0$), there exists various (most common) prescriptions:

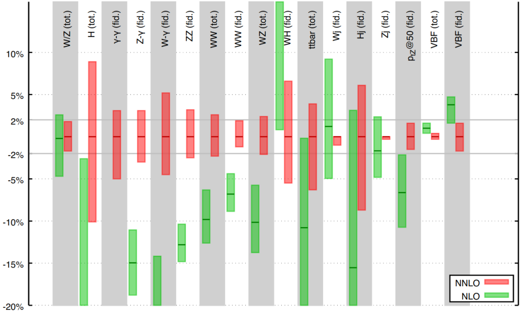


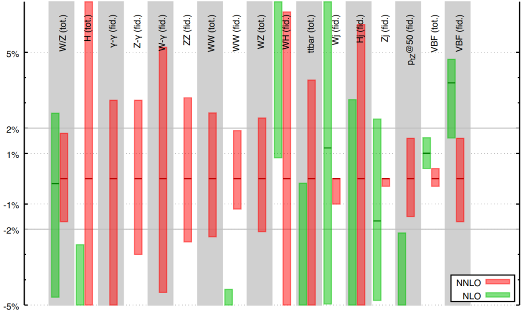




How good is the Scale Variation method?

[Gavin Salam's Slides ('16 PSR,'19)]





- NNLO predictions just barely reach **1%** and for many processes the scale band is $\sim \pm$**2%**
- Only **3/17** cases in which the NNLO central values are contained in NLO uncertainty band

Pros & Cons of Scale Variation

ADVANTAGES:

- **Renormalization Group Invariance** ensures that as the order **increases**, the scale dependence **decreases**
- Lead to smooth functions hereby incorporating correlations between nearby regions in the phase space
- **Universal** and therefore can be applied to various processes

CAVEATS:

- Lack of **Probabilistic Interpretation**
- Ambiguity in defining the central scale and the ranges at which the scale should vary
- Do not account for new singularities appearing at higher-orders

Approximate Higher-Order with Resummations

Consider the gg-channel in $gg \longrightarrow \mathrm{H}$ (HEFT):

$$\left[\frac{d\hat{\sigma}_{gg}}{dp_T}\right]^{\mathrm{N}^{n+1}\mathrm{LO}}(N,\xi_p) = \left[\frac{d\hat{\sigma}_{gg}}{dp_T}\right]^{\mathrm{N}^{n}\mathrm{LO}}(N,\xi_p) + \frac{d\hat{\sigma}_{gg}^{(n+1)}}{dp_T}(N,\xi_p)$$

where the $(n+1)$-th order:

$$\boxed{\frac{d\hat{\sigma}_{gg}^{(n)}}{dp_T}(N,\xi_p) = \frac{d\hat{\sigma}_{gg}^{\mathrm{he},(n)}}{dp_T}(N,\xi_p) + \frac{d\hat{\sigma}_{gg}^{\mathrm{th},(n)}}{dp_T}(N,\xi_p).}$$

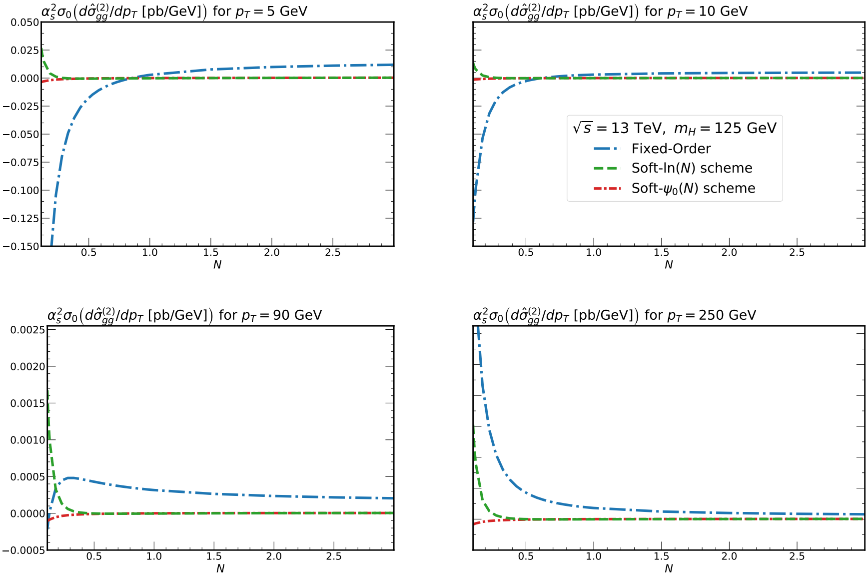
Only holds if small-N (BFKL) behaviours are not spoiled by large-N (soft), and vice-versa.

Treatment of Soft part: (N-soft approximation is expected not to be good at finite N)

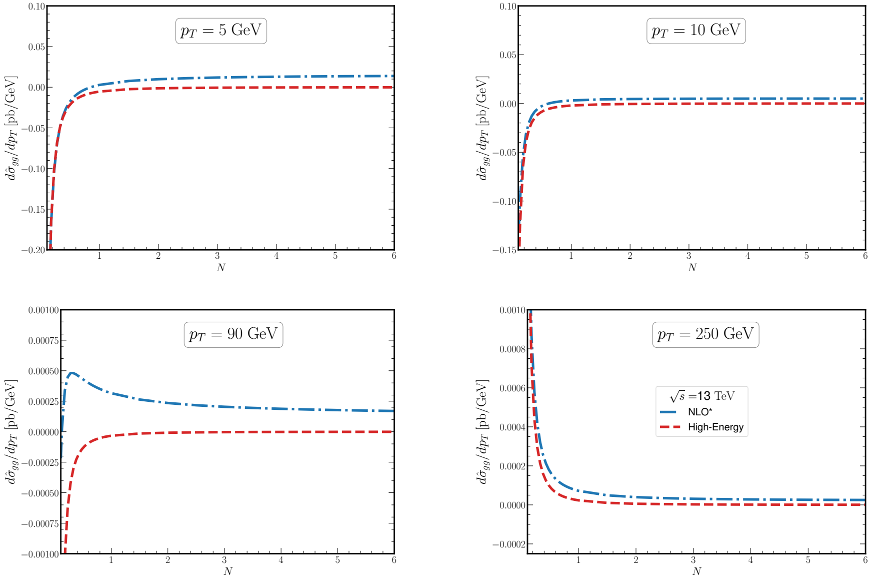
Resummation: $$\frac{d\hat{\sigma}_{gg}^{\mathrm{th},(n)}}{dp_T}(N) = \sum_{n=0}^{\infty} \alpha_s^n \sum_{k=0}^{2n} \mathcal{C}_{n,k} \ln^k N, \qquad \mathcal{M}^{-1}(\ln N) \approx \left(\frac{\ln(\ln 1/x)}{\ln 1/x}\right)_+$$

FO: $$\int_0^1 dx\, x^{N-1} \left(\frac{\ln(1-x)}{1-x}\right)_+ = \frac{1}{2}\left(\gamma_E^2 + \frac{\pi^2}{6}\right) + \gamma_E \psi_0(N) + \frac{1}{2}\left(\psi_0^2(N) - \psi_1(N)\right),$$

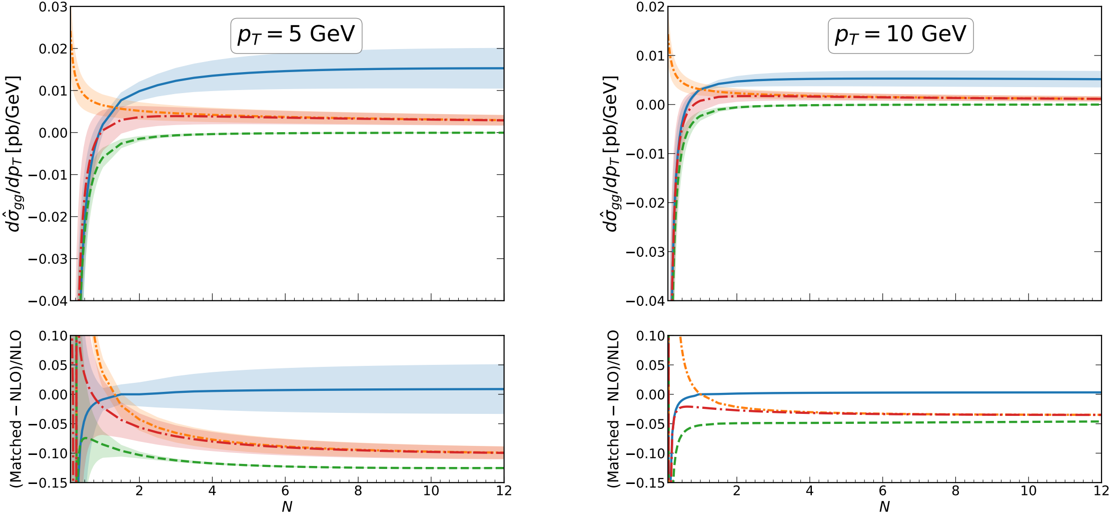
Soft Part: Validation at NLO



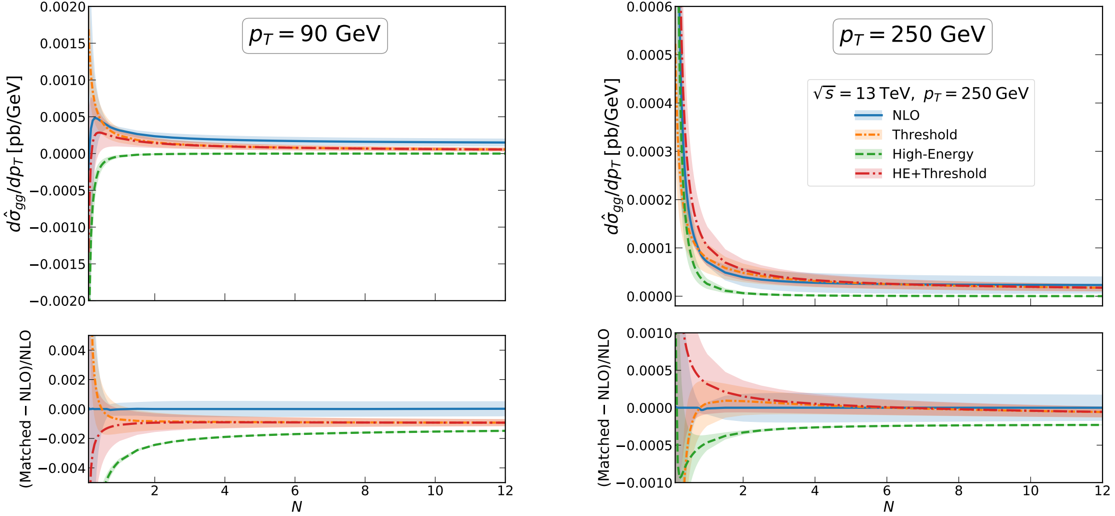
High-Energy Part: Validation at NLO



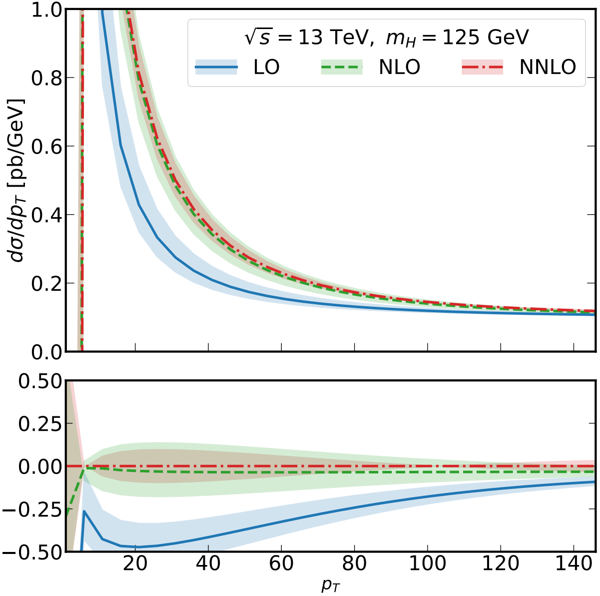
Threshold+High-Energy: Validation at NLO



Threshold+High-Energy: Validation at NLO



NNLO Approximation: Hadonic result

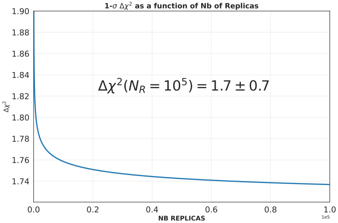


- **Good perturbative convergence: only increase by ∼2-5% wrt NLO**
- **NNLO error fully contained in the NLO uncertainty band**

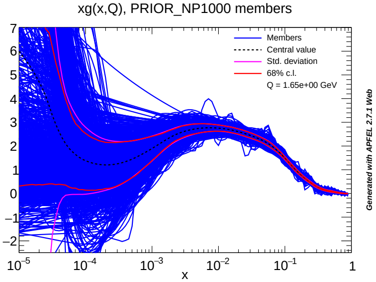
Generative Modellings for PDFs

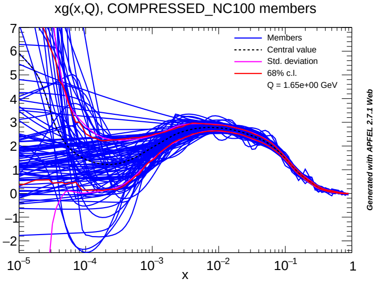
Motivations

- $\Delta\chi^2$ strongly depends on the **number of Monte Carlo (MC) replicas**.
- **Drawback:** Dealing with very large samples of replicas can become unpractical.



Solution: Reduce size of MC PDF sets with **no/minimal loss of information**.





Compression of Probability Distributions

Statement of the Problem:

How to find a specific subset of the original replicas such that the statistical distance between the original and the compressed probability distribution is minimal?

Figure of merit to quantify distinguishability between two Monte Carlo PDF sets:

$$\mathrm{ERF} = \frac{1}{\mathrm{N}_{\mathrm{est}}} \sum_k \frac{1}{\mathrm{N}_k} \sum_i \left(\frac{\mathrm{C}^k(x_i) - \mathrm{P}^k(x_i)}{\mathrm{P}^k(x_i)} \right)^2$$

where:

- P^k is the value of the estimator k for the **Prior**.
- C^k is the value of the same estimator for the **Compressed**.
- N_k is the **normalization** factor for the estimator k.

List of possible estimators:

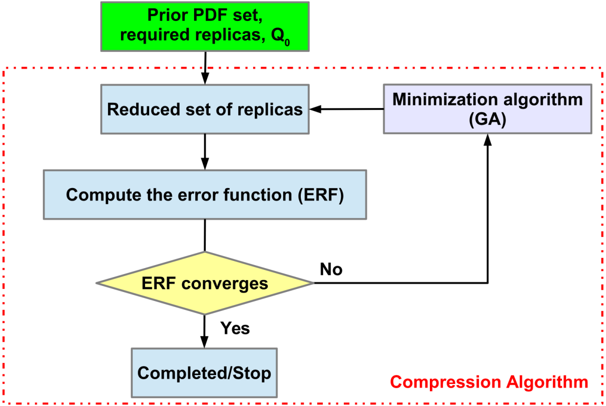
✓ Central Value

✓ Higher Moments (standard deviation, Kurtosis, Skewness)

✓ Kolmogorov-Smirnov distance

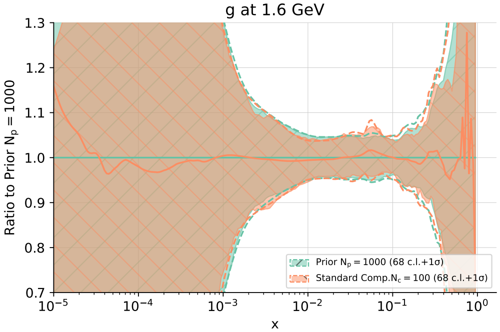
✓ Correlation between pairs of flavours

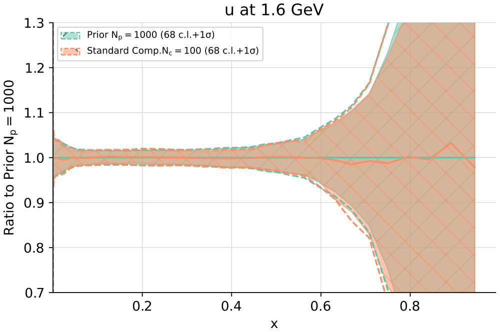
Workflow design of a standard compression algorithm [S. Carrazza et al., 2015]



Prior vs. Compressed set

The main concept works! ($N_p = 1000 \longrightarrow N_c = 100$)

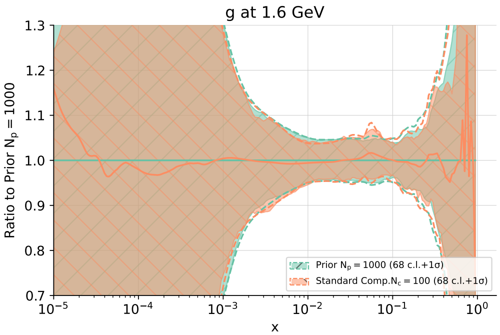


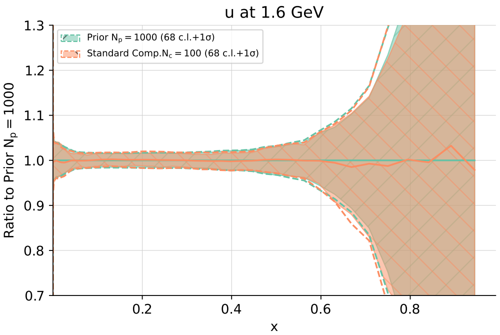


$N_c = 100$ **captures the statistical information of** $N_p = 1000$

Prior vs. Compressed set

The main concept works! $(N_p = 1000 \longrightarrow N_c = 100)$





$N_c = 100$ **captures the statistical information of** $N_p = 1000$

BUT CAN WE DO BETTER?

Why GANs?

Observation: Standard compression algorithm just extracts samples that present small fluctuations and which reproduce best the statistical properties of the Prior.

Consideration: Efficiency of the compression can be improved by generating additional **(synthetic)** replicas that contain less fluctuations.

Why GANs?

Observation: Standard compression algorithm just extracts samples that present small fluctuations and which reproduce best the statistical properties of the Prior.

Consideration: Efficiency of the compression can be improved by generating additional **(synthetic)** replicas that contain less fluctuations.

How?

Assuming **PRIOR**$\sim \mathcal{P}_R$: Generate **SYNTHETIC**$\sim \mathcal{P}_\theta$ with GANs such that $\mathcal{P}_\theta \sim \mathcal{P}_R$.

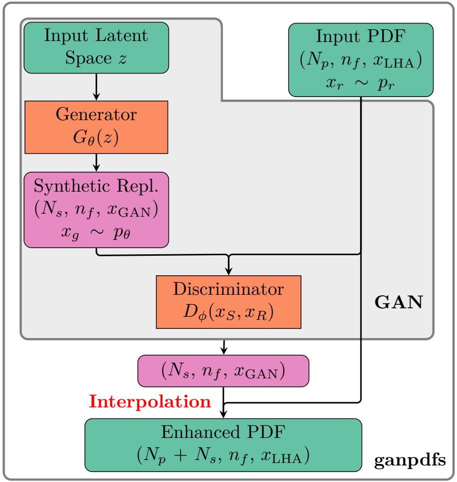
Why GANs?

Observation: Standard compression algorithm just extracts samples that present small fluctuations and which reproduce best the statistical properties of the Prior.

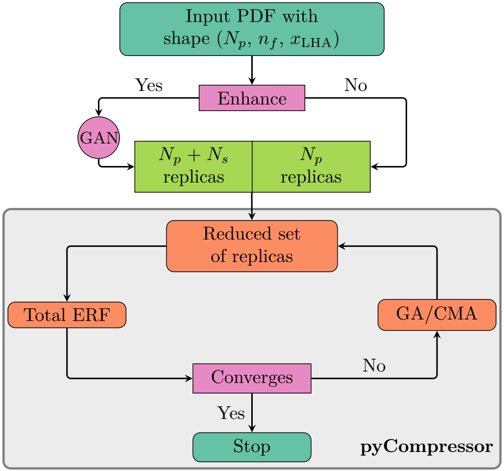
Consideration: Efficiency of the compression can be improved by generating additional **(synthetic)** replicas that contain less fluctuations.

How?

Assuming **PRIOR**$\sim \mathcal{P}_R$: Generate **SYNTHETIC**$\sim \mathcal{P}_\theta$ with GANs such that $\mathcal{P}_\theta \sim \mathcal{P}_R$.

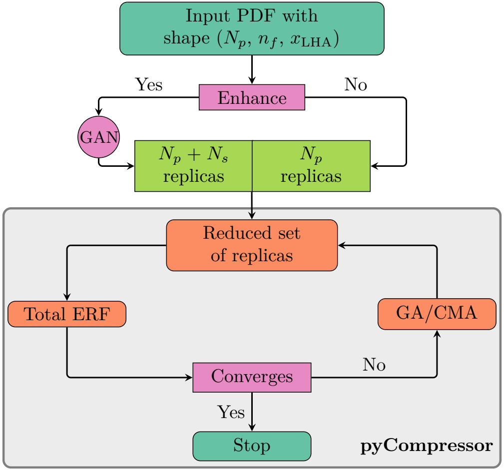


Incorporating ganpdfs into a Compressor



Samples for the compressed set $\mathbb{C}$ are now drawn from $(N_p + N_s)$ **while the minimization is still performed w.r.t the prior $\mathcal{P}$**.

Incorporating ganpdfs into a Compressor



Samples for the compressed set C are now drawn from $(N_p + N_s)$ **while the minimization is still performed w.r.t the prior** $\mathcal{P}$.

$\Longrightarrow$ The compression algorithm has to be:

- efficient to avoid the possibility of being stuck in a local minima
  $\Rightarrow$ **Add various minimizers**
- fast enough in order to perform more iterations
  $\Rightarrow$ **Modify code design**

The Art Hyperparameter Optimization

How can we make sure that the set of hyperparameters are well-chosen in order to get the best results out of a Model?

The Art Hyperparameter Optimization

How can we make sure that the set of hyperparameters are well-chosen in order to get the best results out of a Model?

"... evaluating the quality of generated ~~images~~ samples with human vision is expensive and cumbersome, biased [...] difficult to reproduce, and does not fully reflect the capacity of the models."

[Pros & Cons of GANs Evaluation Measures, 2018]

How can we make sure that the set of hyperparameters are well-chosen in order to get the best results out of a Model?

"... evaluating the quality of generated ~~images~~ samples with human vision is expensive and cumbersome, biased [...] difficult to reproduce, and does not fully reflect the capacity of the models."
[Pros & Cons of GANs Evaluation Measures, 2018]

Solution:

Perform a **Hyperparameter Optimization Scan**:

- Define a reward function to assess the model.
  `ganpdfs` uses a Frechet Inception Distance (FID): **[M. Heusel et al., 2018]**

$$\mathrm{FID} = \frac{1}{2n_f + 1} \sum_{i=-n_f}^{n_f} ||\mu_r^{(i)} - \mu_s^{(i)}||^2 + \mathrm{Tr}\left(\Sigma_r^{(i)} + \Sigma_s^{(i)} - 2\sqrt{\Sigma_r^{(i)}\Sigma_s^{(i)}}\right).$$

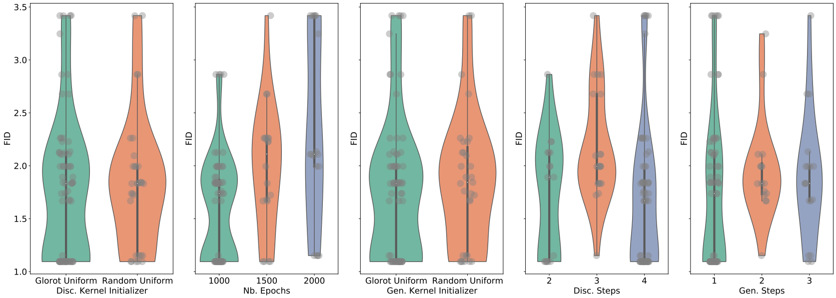
- Scan over thousands of hyperparameter combinations:
  `ganpdfs` relies on the **Tree-Structured Parzen Estimator (TPE)** algorithm to scan over the parameter space. **[J. Bergstra et al., 2013]**

Example of a Hyperparameter Scan

There are over 15 hyperparameters to be taken into account:

Generator	Discriminator	Others
nb. layers & nodes	nb. layers & nodes	size of x_{GAN} grid
kernel initializer	weights clipping	Gaussian noise
activation function	Gradient Penalty loss & weight	nb. epochs
...	...	...

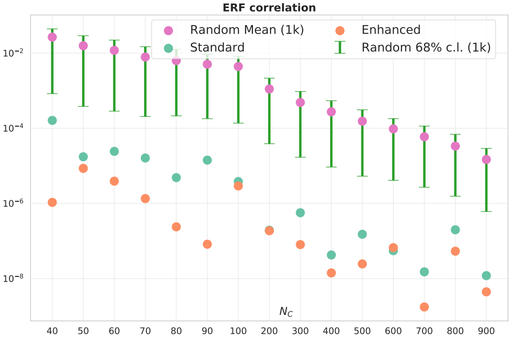
Visual representation of a hyperparameter scan:

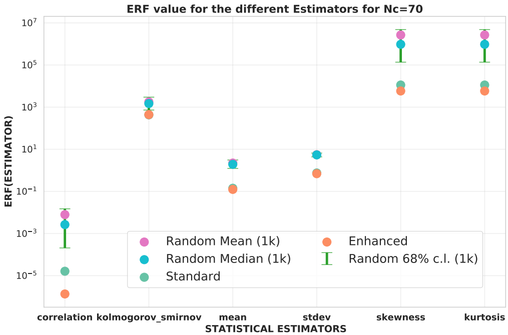


Efficiency of GAN-enhanced Compression: Individual Estimators

SETUP: $(N_p = 1000 + N_s = 2000) \longrightarrow \{N_c\}$

Non-normalized ERFs:

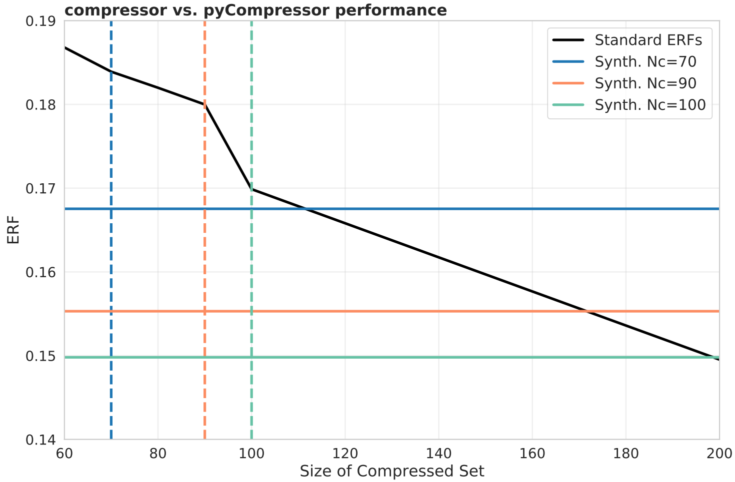




✓ Both **standard** and **GAN-enhanced** compression methodologies outperform any random selection.

✓ **GAN-enhanced** approach yields better compression for all the individual estimators.

Efficiency of GAN-enhanced Compression: Total ERF



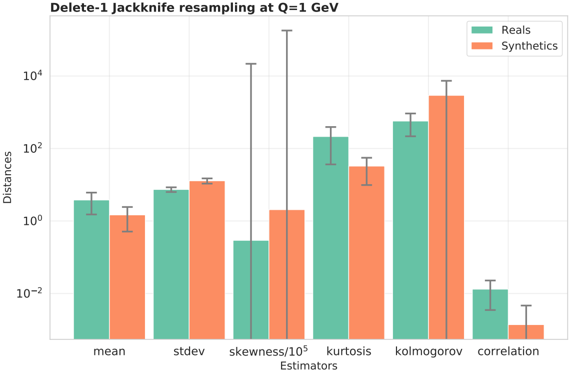
$$N_c(\text{GAN-Enhanced}) = 70 \sim N_c(\text{Standard}) = 110$$

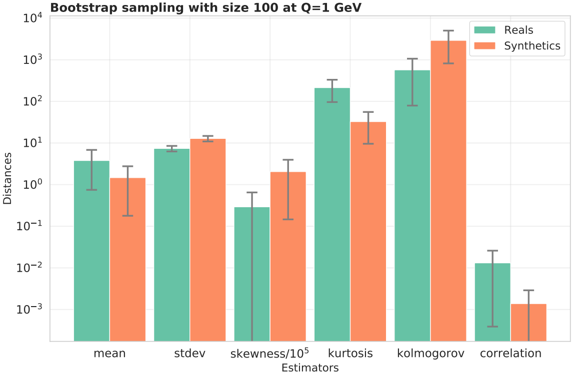
GANs for finite size effect?

Consider 3 different fits:

- 2 disjoint fits S_1 and S_2 with $N = 500$ replicas
- GAN fit S_3 with with $N = 500$ replicas determined from $N_0 = 100$

Compare $\mathrm{D}(S_1, S_2)$ and $\mathrm{D}(S_1, S_3)$ using different resampling methodologies.





Conclusions

Resummation:

- **SIpT** helps elucidate the relation between collinear and soft logarithms
- **SIpT** accelerates perturbative convergences of the Higgs $p_{T,H}$ spectra in the small-p_T regions
- **Combined resummation** yields more reliable results beyond small-p_T

Approximation:

- Resummed expressions could be potential tools to estimate/approximate MHO
- Extend results beyond HEQT and/or to DY processes (currently ongoing work)

GANs for PDFs:

- GANs can be used to generate synthetic MC PDF replicas
- `ganpdfs-pyCompressor` outperforms standard compression by providing a compressed set with smaller number of replicas and a more adequate representation of the original probability distribution
- GANs can potentially be used to address **finite size effects** in PDFs

"The second main type of machine learning is the descriptive or unsupervised learning approach. Here we are only given inputs, and the goal is to find interesting patterns in the data. [...] This is a much less well-defined problem, since we are not told what kinds of patterns to look for, and there is no obvious error metric to use."

K. P. Murphy, *Machine Learning: A Probabilistic Perspective*

THANK YOU

BACKUP

Phase Space Factorization

Move to conjugate space

- Mellin space: trade the convolution for a normal product

$$\frac{\mathrm{d}\hat{\sigma}_{ab}}{\mathrm{d}\xi_p}(N, \xi_p, \alpha_s) = \int_0^1 \mathrm{d}x\, x^{N-1} \frac{1}{x} \frac{\mathrm{d}\hat{\sigma}_{ab}}{\mathrm{d}\xi_p}(x, \xi_p, \alpha_s)$$

- Fourier space: factorize constraints in δ-terms

$$\frac{\mathrm{d}\hat{\sigma}_{ab}}{\mathrm{d}\xi_p}(N, b, \alpha_s) = \pi \int_0^\infty d\xi_p\, J_0(bp_T) \frac{\mathrm{d}\hat{\sigma}_{ab}}{\mathrm{d}\xi_p}(N, \xi_p, \alpha_s)$$

Factorization of the Phase Space

The phase space for n emissions factorizes in Mellin-Fourier space:

$$d\Phi_{n+1}\left(p_1, p_2; p, k_1, \ldots, k_n\right) = \frac{8\pi^{3-\epsilon} Q^{2n}}{\left[4(2\pi)^{2-\epsilon}\right]^{n+1}} \frac{\hat{\tau}}{\Gamma(1-\epsilon)} d\xi_p \int db^2 \left(bp_{\mathrm{T}}\right)^{-\epsilon} b^{2n\epsilon}$$
$$J_{-\epsilon}\left(bp_{\mathrm{T}}\right) \left(\ldots J_{-\epsilon}\left(bk_{\mathrm{T_n}}\right) \frac{\left(bk_{\mathrm{T_n}}\right)^{-\epsilon} d\xi_n dz_n}{\sqrt{(1-z_n)^2 - \frac{4}{z_1^2 \ldots z_{n-1}^2} \xi_n \hat{\tau}}} \right) \delta\left(\hat{\tau} - z_1 \ldots z_n\right) + \mathcal{O}\left(\frac{1}{b}\right)$$

Phase Space Factorization

In the context of SSpT, the phase space factor is expanded in the ξ_i limit as

$$\frac{1}{\sqrt{(1-z)^2-4\xi}} \to \left(\frac{1}{1-z}\right)_+ - \frac{1}{2}\delta(1-z)\ln\xi$$

However, power counting in Fourier-Mellin space shows that

$$\mathcal{FM}\left[\frac{1}{\sqrt{(1-z)^2-4\xi}}\right] = \frac{2}{b^2Q^2}\left(1-\frac{4N^2}{b^2Q^2}+\frac{16N^4}{b^4Q^4}+\dots\right) \tag{1}$$

$\Longrightarrow$ **To preserve the threshold limit at inclusive level, small-p_T resummation must be performed by taking the limit $b \to \infty$ at fixed N/b.**

The factorized phase space then writes:

$$d\Phi_{n+1}\left(p_1,p_2;p,k_1,\dots,k_n\right) = \frac{8\pi^{3-\epsilon}Q^{2n}}{\left[4(2\pi)^{2-\epsilon}\right]^{n+1}}\frac{\hat{\tau}}{\Gamma(1-\epsilon)}d\xi_p\int db^2 J_0\left(bp_{\mathrm{T}}\right)$$
$$\left(\prod_{i=1}^{n} J_{-\epsilon}\left(bk_{\mathrm{i}}\right)\frac{(bk_i)^{-\epsilon}d\xi_i dz_i}{\sqrt{\left(1-z_i\right)^2-4z_i\xi_i}}\right)\delta\left(\hat{\tau}-z_1\dots z_n\right)+\mathcal{O}\left(\frac{1}{b}\right)+\mathcal{O}\left(\frac{1}{N}\right)$$

Problem of Fourier-Mellin inverse transform

In addition to the **Landau Pole** that prevents the existence of a Mellin inverse. TIpT exhibits more complicated singularities due to the interplay between N and b in the logs.

1st Trick: Expand the resummed expression as a series in $\bar{\alpha}_s \mathrm{L}$ ($\bar{\alpha}_s = \alpha_s \beta_0$):

$$\frac{\mathrm{d}\hat{\sigma}'}{\mathrm{d}\xi_p}\left(N, \bar{\alpha}_s \ln \chi\right) = \sum_{k=0}^{\infty} h_k\left(N\right) \bar{\alpha}_{\mathrm{s}}^k \ln^k \chi = \sum_{k=0}^{\infty} h_k\left(N, \alpha_{\mathrm{s}}\right) \bar{\alpha}_{\mathrm{s}}^k \ln^k\left(\bar{N}^2 + \frac{\hat{b}^2}{b_0^2}\right)$$

and perform the b integration term-by-term:

$$\frac{\mathrm{d}\hat{\sigma}'}{\mathrm{d}\xi_p}\left(N, \xi_p\right) = \sum_{k=0}^{\infty} h_k\left(N, \alpha_s\right) \frac{k!}{2\pi\mathrm{i}} \oint_H \frac{\mathrm{d}\xi}{\xi} M\left(N, \xi_p, \xi\right) \left(\frac{\bar{\alpha}_s}{\xi}\right)^k$$

where $$M\left(N, \xi_p, \xi\right) = \int_0^{\infty} \mathrm{d}\hat{b} \frac{\hat{b}}{2} J_0\left(\hat{b}\sqrt{\xi_p}\right) \ln^k\left(\bar{N}^2 + \frac{\hat{b}^2}{b_0^2}\right)$$

Problem: The series does not converge!

Problem of Fourier-Mellin inverse transform

2nd Trick: Sum the series using Borel method (i.e. $\alpha_s \to \frac{w^k}{k!}$):

$$\mathcal{B}\left[\frac{\mathrm{d}\hat{\sigma}'}{\mathrm{d}\xi_p}(N,\xi_p)\right](w) = \frac{1}{2\pi\mathrm{i}}\sum_{k=0}^{\infty} h_k(N)\oint_H \frac{\mathrm{d}\xi}{\xi} M(N,\xi_p,\xi)\left(\frac{w}{\xi}\right)^k$$

$$= \frac{1}{2\pi\mathrm{i}}\oint_H \frac{\mathrm{d}\xi}{\xi} M(N,\xi_p,\xi)\frac{\mathrm{d}\hat{\sigma}'}{\mathrm{d}\xi_p}\left(N,\frac{w}{\xi}\right)$$

and then perform a Borel inverse transform:

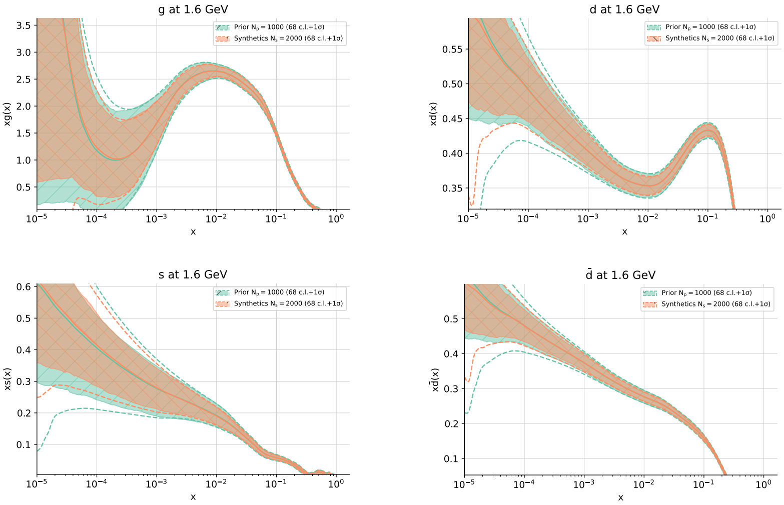
$$\frac{\mathrm{d}\hat{\sigma}'}{\mathrm{d}\xi_p}(N,\xi_p) = \int_0^{\infty} \mathrm{d}w\, \mathrm{e}^{-w}\mathcal{B}\left[\frac{\mathrm{d}\hat{\sigma}'}{\mathrm{d}\xi_p}(N,\xi_p)\right](w)$$

Since the above integral diverges at ∞, we have to cut the integration at a given **cutoff** C.

Finally, the Mellin inverse transform is performed using the standard **Minimal Prescription**.

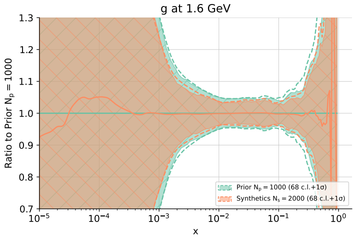
Prior vs. Synthetic: PDFs

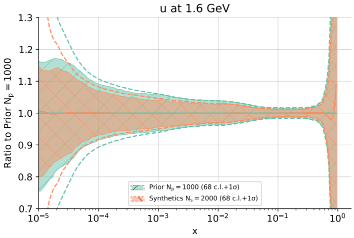
Samples of $N_s = 2000$ synthetic replicas generated from $N_p = 1000$ prior. The plots are the results of a hyperparameter scan with 1000 trials.



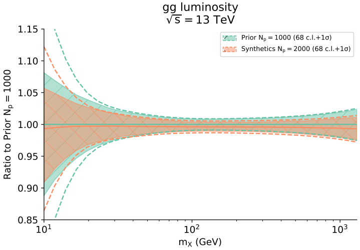
Prior vs. Synthetic: Normalized PDFs & Luminosities

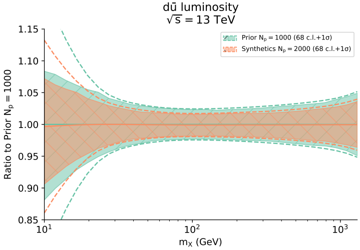
Normalized PDFs:





Luminosities:





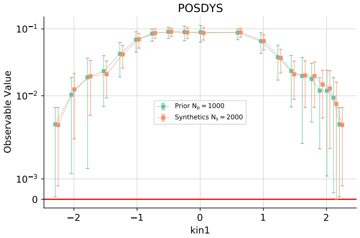
Prior vs. Synthetic: Physical Constraints

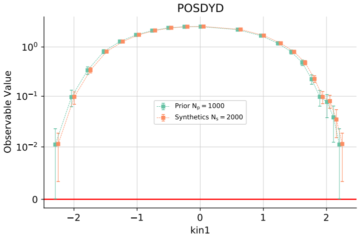
The generated synthetic samples satisfy physical constraints:

Sum Rules:

Prior | Synthetic	mean	std
momentum	0.9968 | 0.9954	7.315×10^{-4} | 1.907×10^{-3}
u_v	1.985 | 1.992	3.122×10^{-2} | 3.788×10^{-2}
d_v	0.9888 | 0.9956	3.764×10^{-2} | 3.796×10^{-2}
s_v	3.249×10^{-3} | 2.073×10^{-4}	3.547×10^{-2} | 4.833×10^{-2}

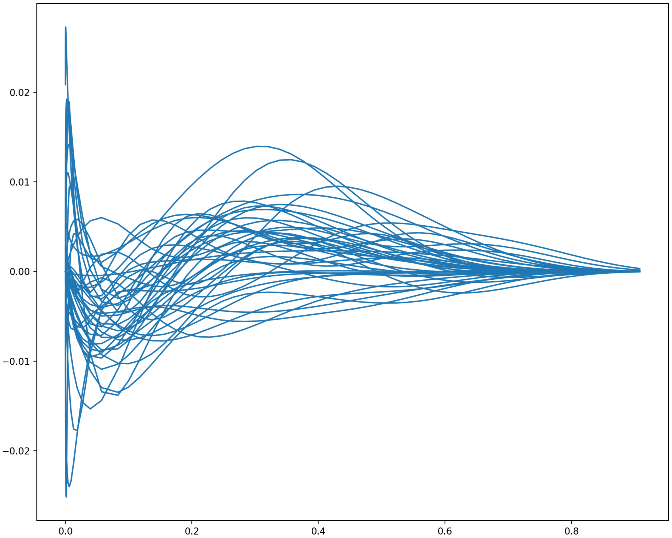
Positivity:





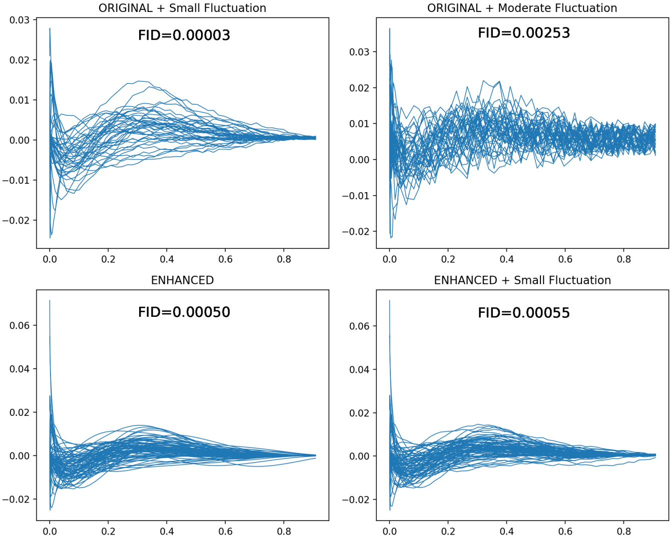
Intuition for FID

Given some probability distributions:



Intuition for FID

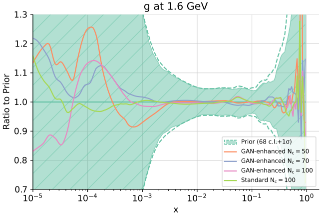
The FID can provide an estimation on the similarity:

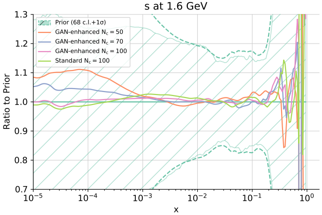


Prior vs. GAN-enhanced

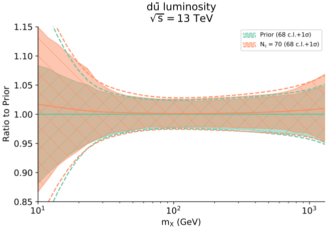
$$N_p = 1000 \rightarrow \text{GAN} \rightarrow (N_p + N_s) = 3000 \rightarrow N_c = 70, 100$$

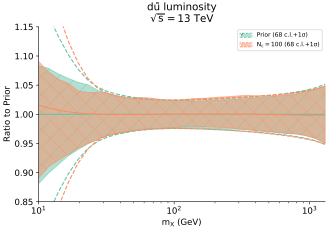
Normalized PDFs:





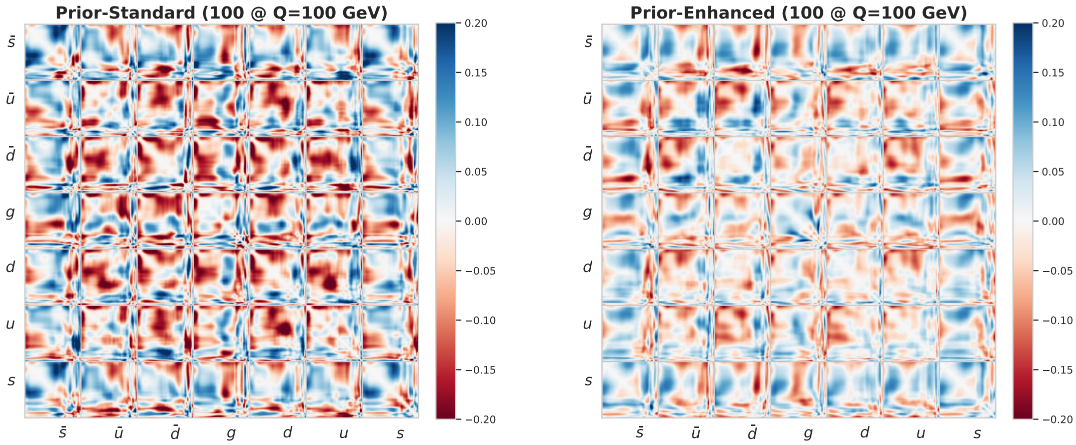
Luminosities:





Prior vs. GAN-enhanced: Correlation

Compression to $N_c = 100$:



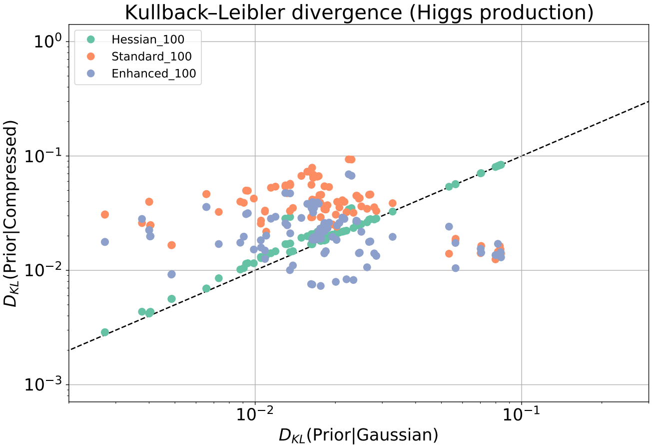
Phenomenological Implications: Accuracy & (Non) Gaussianity

Phenomenological Implications:

Higgs boson production fully differential in either $p_T^H = [0, 200]$ GeV or $y_H = [-2.5, 2.5]$.

Process:	$gg \to H$	VBF $H2j$	HW	HZ	$Ht\bar{t}$

[Les Houches, 2015]



Deliverables

The implementation of the tools presented here are OPEN-SOURCE:

Resummation & Approximation:

- **HpT-MON:** Higgs transverse momentum distribution in momentum and Mellin space.
  github.com/N3PDF/HpT-MON
- **HpT-N3LO:** Approximation of the Higgs transverse momentum distribution at NNLO.
  github.com/N3PDF/HpT-N3LO

GANs for PDFs:

- **ganpdfs:** Generation of synthetic MC PDF replicas with GANs.
  github.com/N3PDF/ganpdfs n3pdf.github.io/ganpdfs/
- **pyCompressor:** Fast compression of MC PDF replicas.
  github.com/N3PDF/pycompressor n3pdf.github.io/pycompressor/