Parton distribution functions and machine learning

P. Zurita

in collaboration with **D. Rentería-Estrada**, **R. Hernández-Pinto** and **G. Sborlini**

Universität Regensburg

Outlook

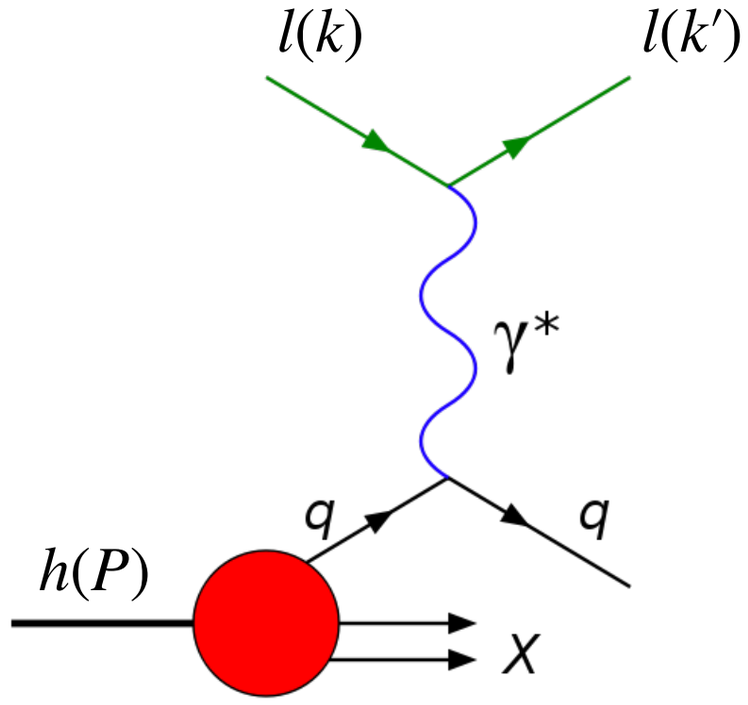
- Parton distributions functions:
  - Some history.
  - How to compute them.
  - Technical details.
- Using machine learning to learn about PDFs:
  - Motivation.
  - Accessing momentum fractions in p+p with ML (several methods and results).
- Going greener with ML (preliminary results).
- Summary

“All” about parton distribution functions

1960s: Deep(ly) Inelastic Scattering

$$Q^2 = -(k - k')^2 \qquad x = \frac{Q^2}{2P \cdot q} \qquad y = \frac{P \cdot q}{P \cdot k}$$

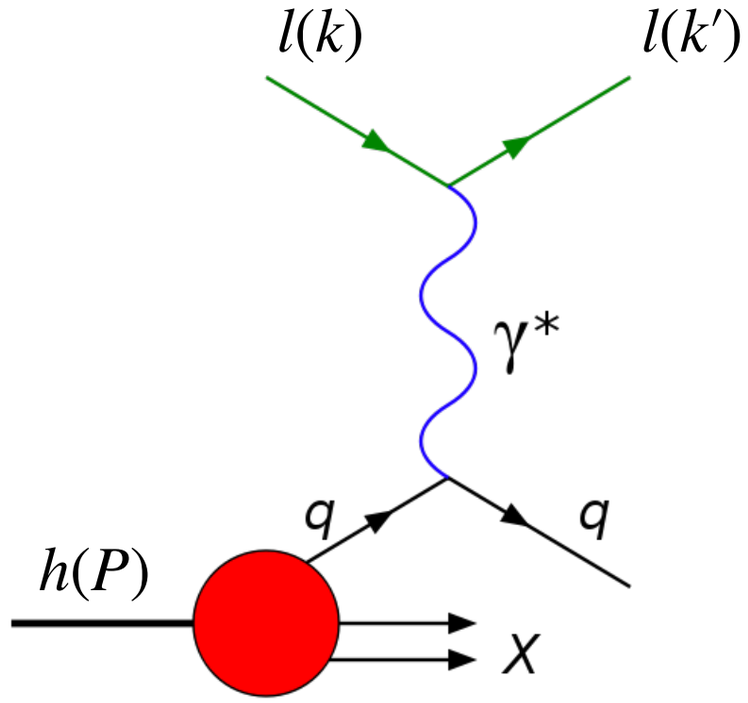
$$\frac{d^2\sigma}{dx dQ^2} \propto F_2(x, Q^2) - \frac{y^2}{1 + (1 - y)^2} F_L(x, Q^2)$$

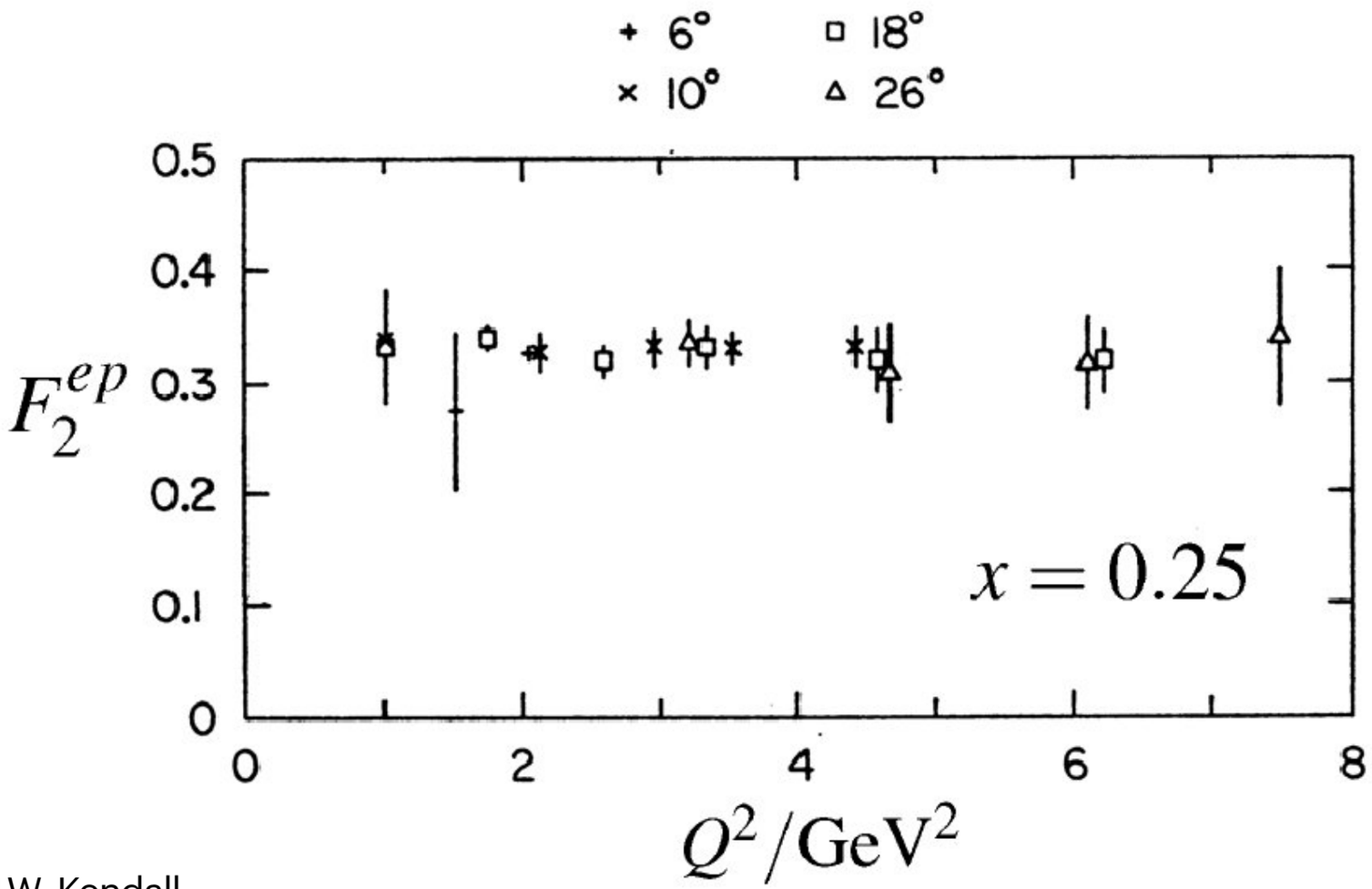


1960s: Deep(ly) Inelastic Scattering

$$Q^2 = -(k-k')^2 \qquad x = \frac{Q^2}{2P\cdot q} \qquad y = \frac{P\cdot q}{P\cdot k}$$

$$\frac{d^2\sigma}{dxdQ^2} \propto F_2(x,Q^2) - \frac{y^2}{1+(1-y)^2}F_L(x,Q^2)$$





J.T. Friedman and H.W. Kendall,
Ann.Rev.Nucl.Sci. 22 (1972) 203.

- The scaling is expected if DIS is the incoherent scattering of partons (Feynman, 1969).

$$F_2^{LO}(x) = x \sum_{i=1}^{n_f} e_i^2 \, f_{i/h}(x)$$

- $f_{i/h}(x)$ is the **probability** density of finding the parton ***i*** inside the hadron ***h*** with ***x***. These are called ***Parton Distribution Functions*** (PDFs).

- The scaling is expected if DIS is the incoherent scattering of partons (Feynman, 1969).

$$F_2^{LO}(x) = x \sum_{i=1}^{n_f} e_i^2 \, f_{i/h}(x)$$

- $f_{i/h}(x)$ is the **probability** density of finding the parton ***i*** inside the hadron ***h*** with ***x***. These are called ***Parton Distribution Functions*** (PDFs).

- At higher orders in the expansion the scaling breaks down, a dependence on the renormalisation scale μ appears.

- The scaling is expected if DIS is the incoherent scattering of partons (Feynman, 1969).

$$F_2^{LO}(x) = x \sum_{i=1}^{n_f} e_i^2 f_{i/h}(x)$$

- $f_{i/h}(x)$ is the **probability** density of finding the parton ***i*** inside the hadron ***h*** with ***x***. These are called ***Parton Distribution Functions*** (PDFs).

- At higher orders in the expansion the scaling breaks down, a dependence on the renormalisation scale μ appears.

- Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (**DGLAP**) equations give the evolution with the scale, while mixing the partons.

$$\mu^2 \frac{d}{d\mu^2} f_{a/h}(x,\mu) = \int_x^1 \frac{d\xi}{\xi} \sum_b P_{a/b}\left(\frac{x}{\xi}, \alpha_s(\mu)\right) f_{b/h}(\xi,\mu)$$

- These objects are related to the ones in the parton model and we still call them “PDFs”, but the probabilistic interpretation is not 100% accurate.

- These objects are related to the ones in the parton model and we still call them “PDFs”, but the probabilistic interpretation is not 100% accurate.

- The PDFs contain the long distance structure of hadrons.

- These objects are related to the ones in the parton model and we still call them “PDFs”, but the probabilistic interpretation is not 100% accurate.

- The PDFs contain the long distance structure of hadrons.

- They are **universal**, so any process that is an inclusive hard scattering can be written as

$$d\sigma^{DIS} = \sum_i d\sigma^{l+i\to l'} \otimes f_i \qquad d\sigma^{DY} = \sum_{i,j} d\sigma^{i+j\to l+\bar{l}} \otimes f_i \otimes f_j$$

with the same PDFs.

- These objects are related to the ones in the parton model and we still call them “PDFs”, but the probabilistic interpretation is not 100% accurate.

- The PDFs contain the long distance structure of hadrons.

- They are **universal**, so any process that is an inclusive hard scattering can be written as

$$d\sigma^{DIS} = \sum_i d\sigma^{l+i\to l'} \otimes f_i \qquad d\sigma^{DY} = \sum_{i,j} d\sigma^{i+j\to l+\bar{l}} \otimes f_i \otimes f_j$$

with the same PDFs.

Without PDFs there is no prediction!

- These objects are related to the ones in the parton model and we still call them “PDFs”, but the probabilistic interpretation is not 100% accurate.

- The PDFs contain the long distance structure of hadrons.

- They are **universal**, so any process that is an inclusive hard scattering can be written as

$$d\sigma^{DIS} = \sum_i d\sigma^{l+i\to l'} \otimes f_i \qquad d\sigma^{DY} = \sum_{i,j} d\sigma^{i+j\to l+\bar{l}} \otimes f_i \otimes f_j$$

with the same PDFs.

Without PDFs there is no prediction!

I will adhere the KISS principle and stick to collinear PDFs.

How to compute PDFs

D.E. Soper, Nucl.Phys.B (Proc.Suppl.) 53 (1997) 69

- PDFs have formal definitions in terms of operators, for example

$$f_{j/h}(x,\mu) = \frac{1}{4\pi}\int dy^- e^{-ixP^+y^-}\langle P^+, \overrightarrow{0}_T \mid \bar{\psi}_j(0,y^-,\overrightarrow{0}_T)\,\gamma^+\,\mathcal{O}\,\psi_j(0,0,\overrightarrow{0}_T) \mid P^+, \overrightarrow{0}_T\rangle_{\bar{M}S}$$

$$\mathcal{O} = \mathscr{P}\exp\left(ig\int_0^{y^-} dz^- A_a^+(0,z^-,\overrightarrow{0}_T)\,t_a \right)$$

$$P^\pm = (P^0 \pm P^3)/\sqrt{2}$$

How to compute PDFs

D.E. Soper, Nucl.Phys.B (Proc.Suppl.) 53 (1997) 69

- PDFs have formal definitions in terms of operators, for example

$$f_{j/h}(x,\mu) = \frac{1}{4\pi}\int dy^- e^{-ixP^+y^-}\langle P^+, \overrightarrow{0}_T \mid \bar{\psi}_j(0,y^-,\overrightarrow{0}_T)\,\gamma^+\,\mathcal{O}\,\psi_j(0,0,\overrightarrow{0}_T) \mid P^+, \overrightarrow{0}_T\rangle_{\bar{M}S}$$

$$\mathcal{O} = \mathscr{P}\exp\left(ig\int_0^{y^-} dz^- A_a^+(0,z^-,\overrightarrow{0}_T)\,t_a\right)$$

$$P^{\pm} = (P^0 \pm P^3)/\sqrt{2}$$

- $|P\rangle$ is the state of a **hadron**.
- ψ is the field operator of a **quark**.

How to compute PDFs

D.E. Soper, Nucl.Phys.B (Proc.Suppl.) 53 (1997) 69

- PDFs have formal definitions in terms of operators, for example

$$f_{j/h}(x,\mu) = \frac{1}{4\pi}\int dy^- e^{-ixP^+y^-}\langle P^+, \overrightarrow{0}_T \mid \bar{\psi}_j(0,y^-,\overrightarrow{0}_T)\,\gamma^+\,\mathcal{O}\,\psi_j(0,0,\overrightarrow{0}_T) \mid P^+, \overrightarrow{0}_T\rangle_{\bar{M}S}$$

$$\mathcal{O} = \mathscr{P}\exp\left(ig\int_0^{y^-} dz^- A_a^+(0,z^-,\overrightarrow{0}_T)\,t_a\right) \qquad P^\pm = (P^0 \pm P^3)/\sqrt{2}$$

- $|P\rangle$ is the state of a **hadron**.
- ψ is the field operator of a **quark**.

- So we can't compute these from first principles in pQCD, we must resort to phenomenology. We do **global** fits.

Steps for a “traditional” fit:

- Choose:
  - a factorisation scheme
  - an order in perturbation theory*
  - a starting scale Q_0 (so that above it pQCD is valid)
  - the data to be fitted
  - a heavy flavour scheme

Steps for a “traditional” fit:

- Choose:
  - a factorisation scheme
  - an order in perturbation theory*
  - a starting scale Q_0 (so that above it pQCD is valid)
  - the data to be fitted
  - a heavy flavour scheme

- Parametrise the quark and gluon distribution (or a combination of them):

$$xf_i(x, Q_0^2) = A_i x^{\alpha_i}(1-x)^{\beta_i} P(x, c_i)$$

- Solve the DGLAP equations for the measured kinematics.
- Compute the hard cross sections for the observables.
- Convolute PDFs and partonic cross-sections.

- Compute this quantity and minimise it:

$$\sum_{i,j=1}^{N_{data}} \left[\sigma_{exp} - \sigma_{th}\right]_i C_{ij}^{-1} \left[\sigma_{exp} - \sigma_{th}\right]_j = \chi^2_{test}$$

- Compute this quantity and minimise it:

$$\sum_{i,j=1}^{N_{data}} \left[\sigma_{exp} - \sigma_{th}\right]_i C_{ij}^{-1} \left[\sigma_{exp} - \sigma_{th}\right]_j = \chi^2_{test}$$

- By the CLT, each term in the sum is distributed according to the square of a standard Gaussian.
- If we use d parameters, χ^2_{test} follows a χ^2 distribution with $N_{data} - d$ degrees of freedom.
- $E[\chi^2_{test}] = N_{data} - d = N_{d.o.f.} \implies E[\chi^2_{test}]/N_{d.o.f.} = 1$

- Compute this quantity and minimise it:

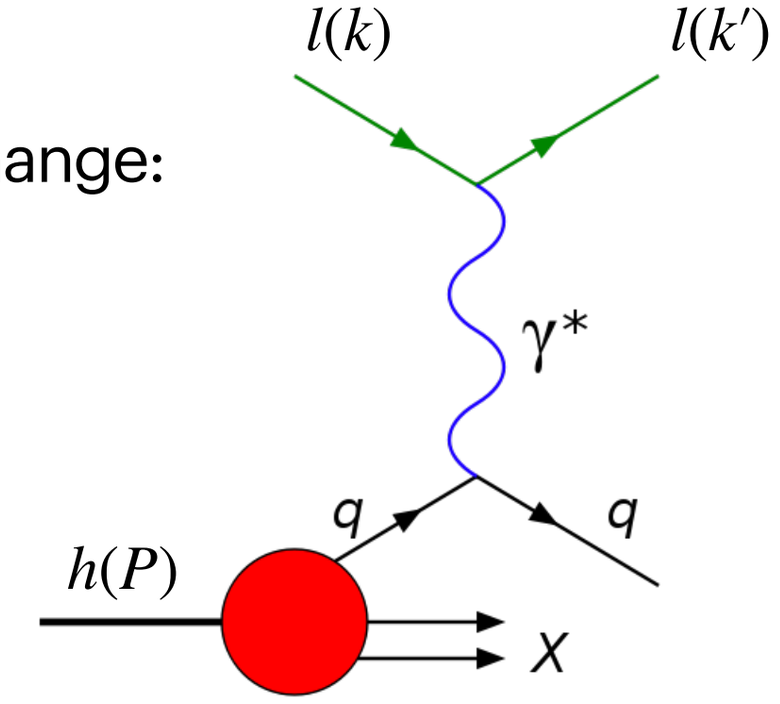
$$\sum_{i,j=1}^{N_{data}} \left[\sigma_{exp} - \sigma_{th}\right]_i C_{ij}^{-1} \left[\sigma_{exp} - \sigma_{th}\right]_j = \chi^2_{test}$$

- By the CLT, each term in the sum is distributed according to the square of a standard Gaussian.
- If we use d parameters, χ^2_{test} follows a χ^2 distribution with $N_{data} - d$ degrees of freedom.
- $E[\chi^2_{test}] = N_{data} - d = N_{d.o.f.} \implies E[\chi^2_{test}]/N_{d.o.f.} = 1$

- When we reach this point (the fit is "good enough") we can be happy and store the final parameters.
- Use some method to estimate theoretical error bands.
- Create grids in x and Q^2, and provide an interpolator for the grid. Make it publicly available (**LHAPDF**).

Some details

- The basic process is DIS with photon/Z boson exchange:



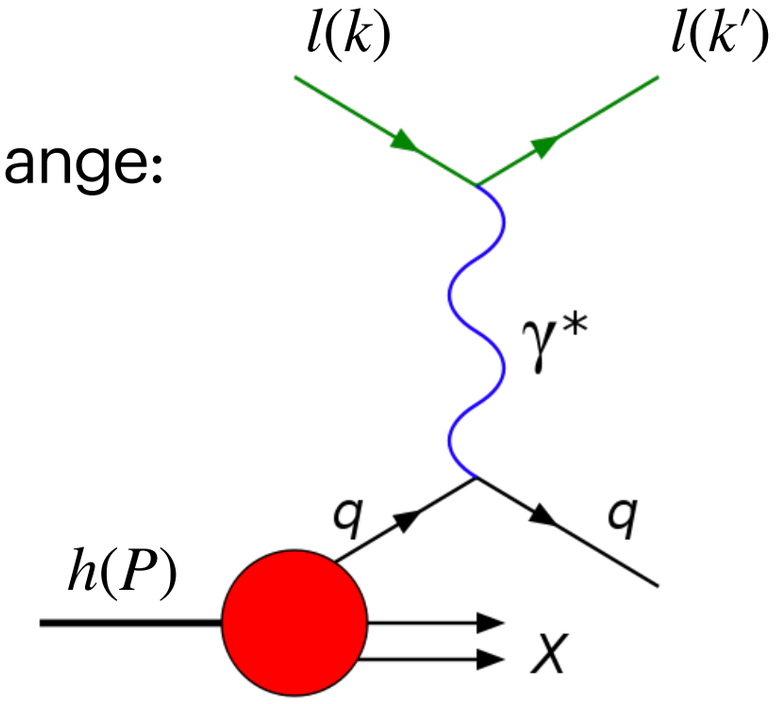
$$\frac{d^2\sigma}{dxdQ^2} \propto F_2(x,Q^2) - \frac{y^2}{1+(1-y)^2} F_L(x,Q^2)$$

$$F_2(x,Q^2) = \sum_{i=0}^{+\infty} \alpha_s^i(Q^2) \left[C_{2,q}^{i,NS} \otimes f_q^{NS} + C_{2,q}^{i,S} \otimes f_q^S + C_{2,g}^i \otimes f_g \right]$$

$$F_L(x,Q^2) = \sum_{i=1}^{+\infty} \alpha_s^i(Q^2) \left[C_{L,q}^{i,NS} \otimes f_q^{NS} + C_{L,q}^{i,S} \otimes f_q^S + C_{L,g}^i \otimes f_g \right]$$

Some details

- The basic process is DIS with photon/Z boson exchange:



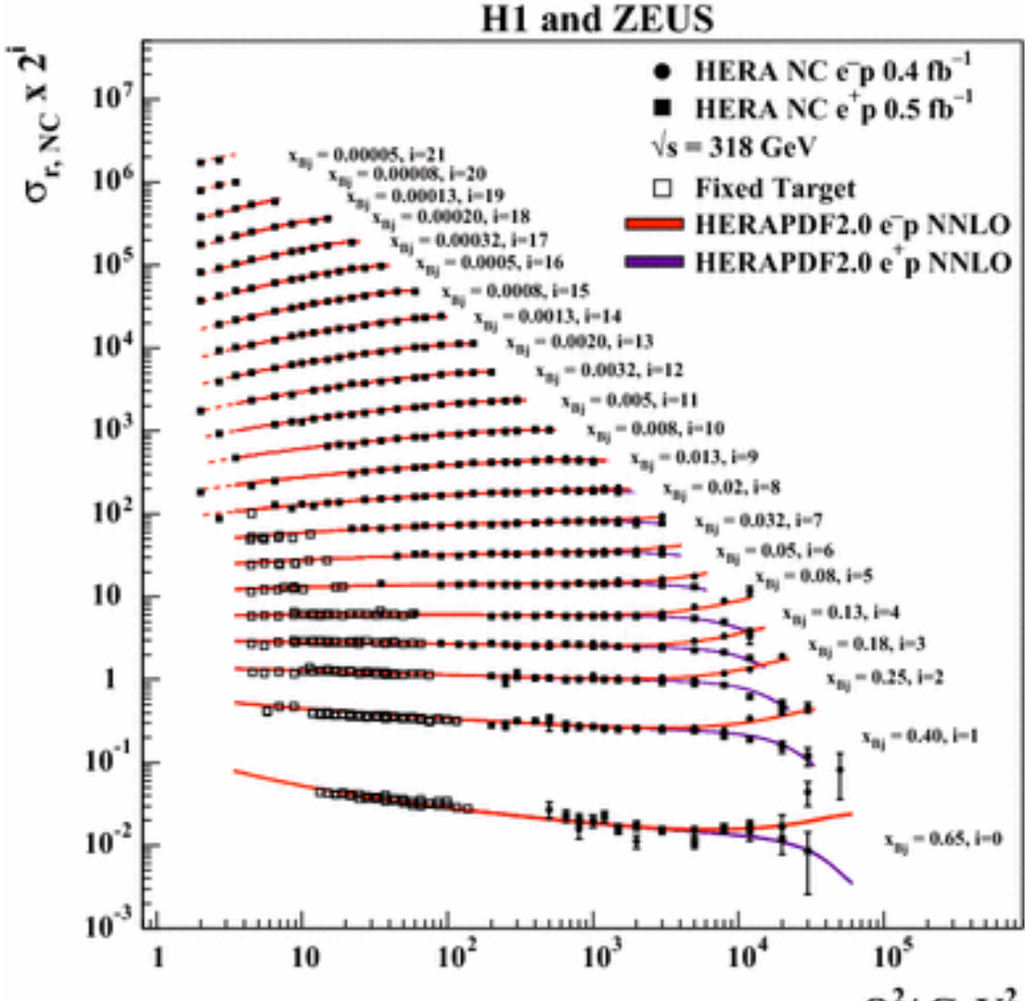
$$\frac{d^2\sigma}{dxdQ^2} \propto F_2(x,Q^2) - \frac{y^2}{1+(1-y)^2} F_L(x,Q^2)$$

$$F_2(x,Q^2) = \sum_{i=0}^{+\infty} \alpha_s^i(Q^2) \left[C_{2,q}^{i,NS} \otimes f_q^{NS} + C_{2,q}^{i,S} \otimes f_q^S + C_{2,g}^i \otimes f_g \right]$$

$$F_L(x,Q^2) = \sum_{i=1}^{+\infty} \alpha_s^i(Q^2) \left[C_{L,q}^{i,NS} \otimes f_q^{NS} + C_{L,q}^{i,S} \otimes f_q^S + C_{L,g}^i \otimes f_g \right]$$

$$f_q^{NS}(x,Q^2) = \sum_{i=1}^{n_f} e_i^2 f_i(x,Q^2)$$

$$f_q^S(x,Q^2) = \sum_{i=1}^{n_f} f_i(x,Q^2)$$



Eur.Phys.J.C 75 (2015) 12, 580

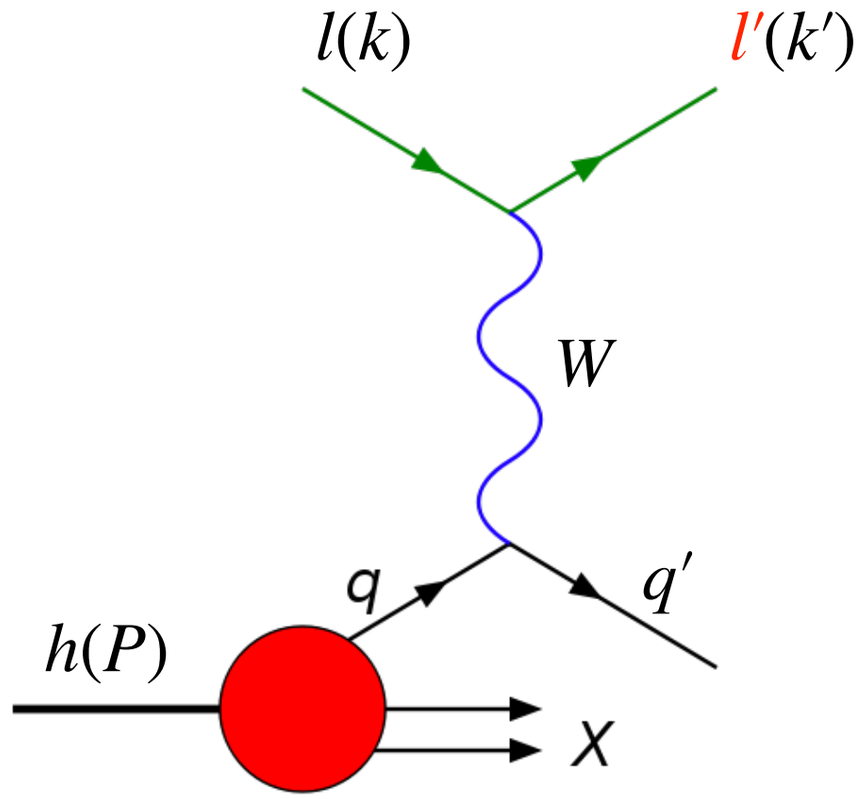
- For quarks, Neutral Current DIS is sensitive only to the non-singlet and singlet combination of the PDFs.

- For quarks, Neutral Current DIS is sensitive only to the non-singlet and singlet combination of the PDFs.

- We need to use other observables, such as Charged Current DIS to distinguish flavours.

$$\frac{d^2\sigma}{dxdQ^2} \propto F_2(x,Q^2) - \frac{y^2}{1+(1-y)^2}F_L(x,Q^2) \mp \frac{1-(1-y)^2}{1+(1-y)^2}xF_3$$

- xF_3 provides more PDF combinations, but it not enough (plus it is very complicated).



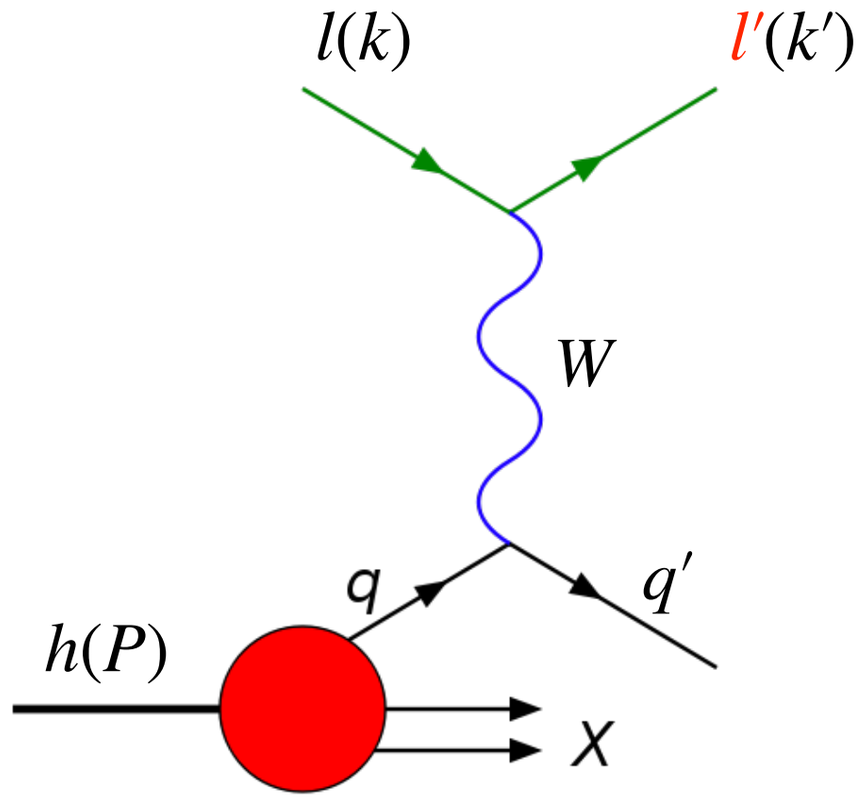
- For quarks, Neutral Current DIS is sensitive only to the non-singlet and singlet combination of the PDFs.

- We need to use other observables, such as Charged Current DIS to distinguish flavours.

$$\frac{d^2\sigma}{dxdQ^2} \propto F_2(x,Q^2) - \frac{y^2}{1+(1-y)^2}F_L(x,Q^2) \mp \frac{1-(1-y)^2}{1+(1-y)^2}xF_3$$

- xF_3 provides more PDF combinations, but it not enough (plus it is very complicated).

- We can use “neutrons”, but it is still to enough, due to the kinematic reach of the data...



- For quarks, Neutral Current DIS is sensitive only to the non-singlet and singlet combination of the PDFs.

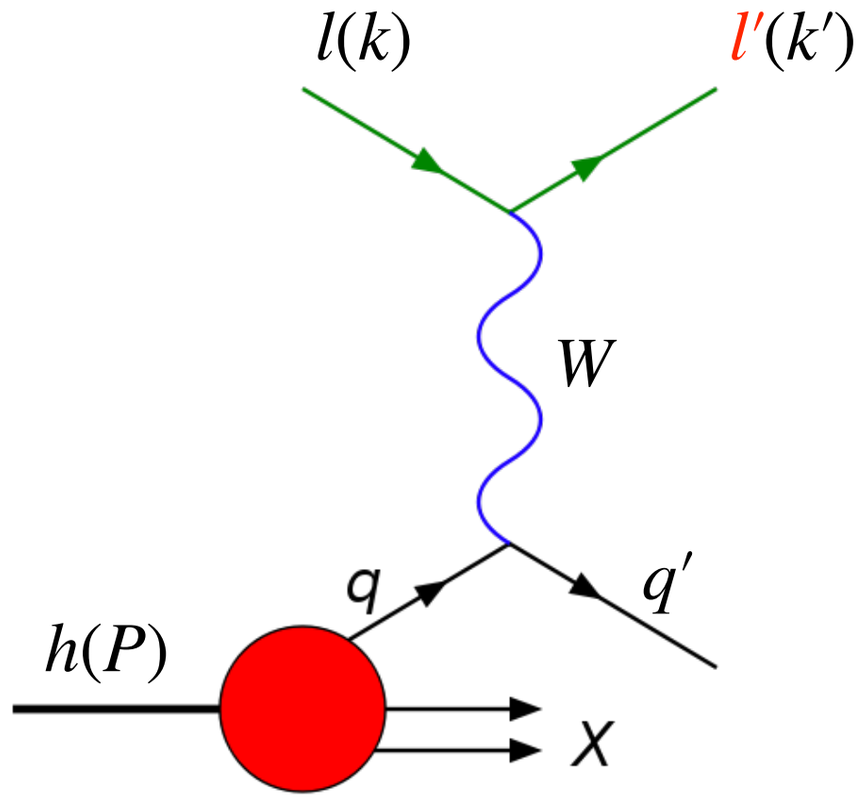
- We need to use other observables, such as Charged Current DIS to distinguish flavours.

$$\frac{d^2\sigma}{dxdQ^2} \propto F_2(x,Q^2) - \frac{y^2}{1+(1-y)^2}F_L(x,Q^2) \mp \frac{1-(1-y)^2}{1+(1-y)^2}xF_3$$

- xF_3 provides more PDF combinations, but it not enough (plus it is very complicated).

- We can use “neutrons”, but it is still to enough, due to the kinematic reach of the data…

- We use **p+p**!



- Let us see for example Drell-Yan: $$\frac{d^2\sigma}{dx_F dQ^2} = \frac{d^2\sigma^A}{dx_F dQ^2} + \frac{d^2\sigma^C}{dx_F dQ^2}$$

$$\frac{d^2\sigma^A}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \frac{d^2\hat{\sigma}^A}{dx_F dQ^2}\left[f_i(t_1,Q^2)\bar{f}_i(t_2,Q^2) + \bar{f}_i(t_1,Q^2) f_i(t_2,Q^2)\right]$$

$$\frac{d^2\sigma^C}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \left[\frac{d^2\hat{\sigma}^C}{dx_F dQ^2} f_g(t_1,Q^2)\left[f_i(t_2,Q^2) + \bar{f}_i(t_2,Q^2)\right] + (1 \longleftrightarrow 2)\right]$$

- Let us see for example Drell-Yan: $$\frac{d^2\sigma}{dx_F dQ^2} = \frac{d^2\sigma^A}{dx_F dQ^2} + \frac{d^2\sigma^C}{dx_F dQ^2}$$

$$\frac{d^2\sigma^A}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \frac{d^2\hat{\sigma}^A}{dx_F dQ^2}\left[f_i(t_1,Q^2)\bar{f}_i(t_2,Q^2) + \bar{f}_i(t_1,Q^2) f_i(t_2,Q^2)\right]$$

$$\frac{d^2\sigma^C}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \left[\frac{d^2\hat{\sigma}^C}{dx_F dQ^2} f_g(t_1,Q^2)\left[f_i(t_2,Q^2) + \bar{f}_i(t_2,Q^2)\right] + (1 \longleftrightarrow 2)\right]$$

- Now it is quadratic in the PDFs, therefore harder to extract.

- Let us see for example Drell-Yan: $\dfrac{d^2\sigma}{dx_F dQ^2} = \dfrac{d^2\sigma^A}{dx_F dQ^2} + \dfrac{d^2\sigma^C}{dx_F dQ^2}$

$$\frac{d^2\sigma^A}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \frac{d^2\hat{\sigma}^A}{dx_F dQ^2}\left[f_i(t_1,Q^2)\bar{f}_i(t_2,Q^2) + \bar{f}_i(t_1,Q^2) f_i(t_2,Q^2)\right]$$

$$\frac{d^2\sigma^C}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \left[\frac{d^2\hat{\sigma}^C}{dx_F dQ^2} f_g(t_1,Q^2)\left[f_i(t_2,Q^2) + \bar{f}_i(t_2,Q^2)\right] + (1 \longleftrightarrow 2)\right]$$

- Now it is quadratic in the PDFs, therefore harder to extract.

- More importantly, beyond LO the x_1, x_2 **are not the momentum fractions** of the partons in the hard interaction!

$$x_1, x_2 = \sqrt{\frac{M^2}{s}} e^{\pm y}$$

- Let us see for example Drell-Yan: $\dfrac{d^2\sigma}{dx_F dQ^2} = \dfrac{d^2\sigma^A}{dx_F dQ^2} + \dfrac{d^2\sigma^C}{dx_F dQ^2}$

$$\frac{d^2\sigma^A}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \frac{d^2\hat{\sigma}^A}{dx_F dQ^2}\left[f_i(t_1,Q^2)\bar{f}_i(t_2,Q^2) + \bar{f}_i(t_1,Q^2) f_i(t_2,Q^2)\right]$$

$$\frac{d^2\sigma^C}{dx_F dQ^2} = \frac{4\pi\alpha^2}{9Q^2 s}\sum_i e_i^2 \int_{x_1}^1 dt_1 \int_{x_2}^1 dt_2 \left[\frac{d^2\hat{\sigma}^C}{dx_F dQ^2} f_g(t_1,Q^2)\left[f_i(t_2,Q^2) + \bar{f}_i(t_2,Q^2)\right] + (1 \longleftrightarrow 2)\right]$$

- Now it is quadratic in the PDFs, therefore harder to extract.

- More importantly, beyond LO the x_1, x_2 **are not the momentum fractions** of the partons in the hard interaction!

$$x_1, x_2 = \sqrt{\frac{M^2}{s}} e^{\pm y}$$

definition used beyond LO (🤫 but it is not true!)

Machine learning for PDFs

Motivation

- Not being able to access the real x_1, x_2 from measuring the kinematics is not exclusive of the Drell-Yan process. It happens for all p+p collisions.

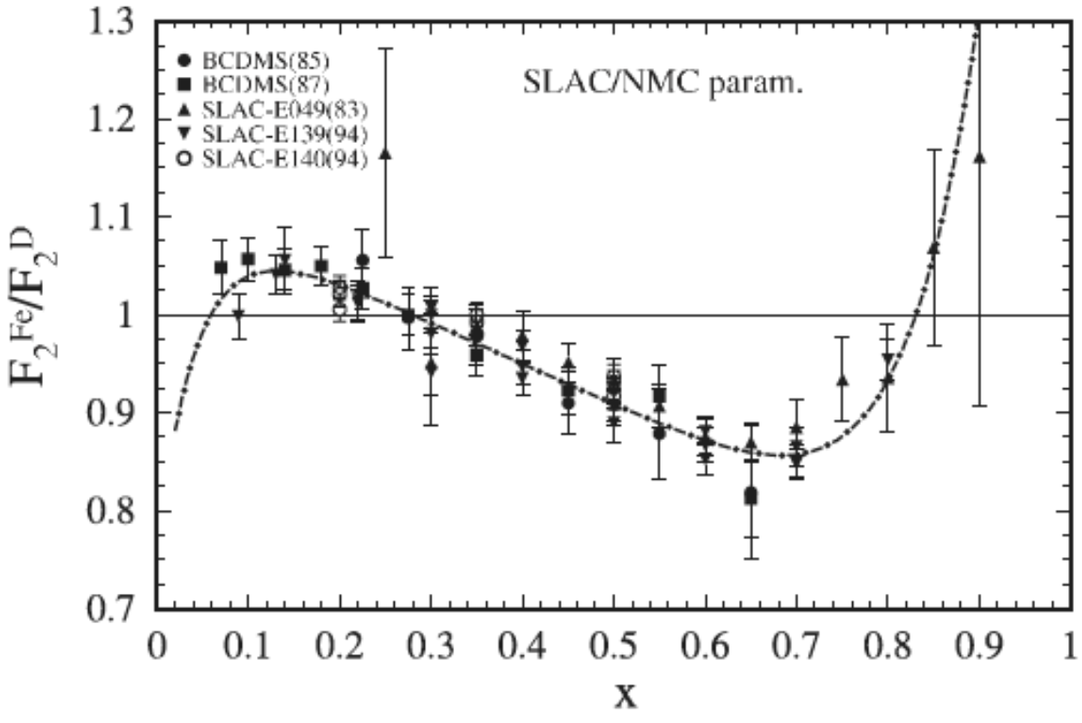
- And can have significant impact: e.g. in p+Pb collisions $x_1, x_2 = \sqrt{\frac{M^2}{s}} e^{\pm y}$

Motivation

- Not being able to access the real x_1, x_2 from measuring the kinematics is not exclusive of the Drell-Yan process. It happens for all p+p collisions.

- And can have significant impact: e.g. in p+Pb collisions

$$x_1, x_2 = \sqrt{\frac{M^2}{s}} e^{\pm y}$$



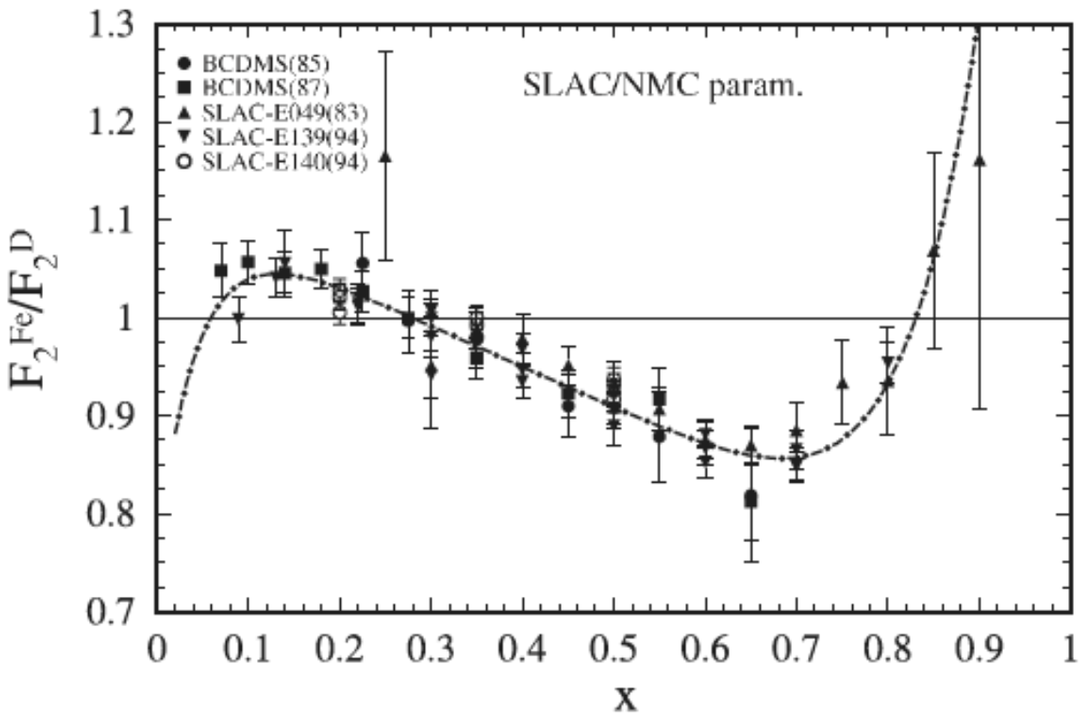
data from e+A DIS

Motivation

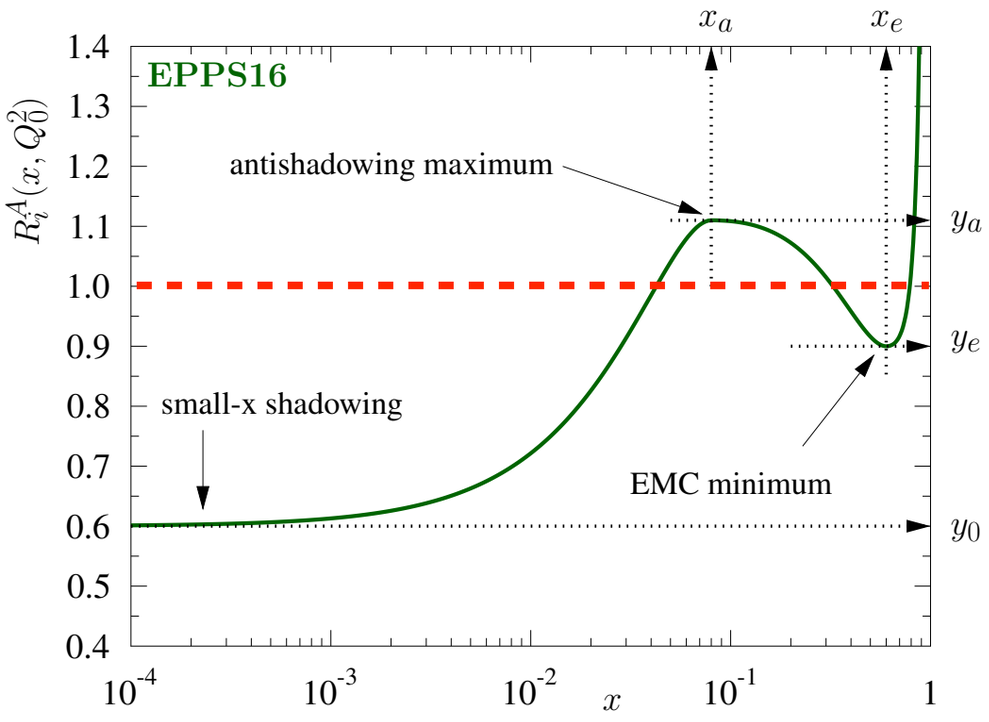
- Not being able to access the real x_1, x_2 from measuring the kinematics is not exclusive of the Drell-Yan process. It happens for all p+p collisions.

- And can have significant impact: e.g. in p+Pb collisions

$$x_1, x_2 = \sqrt{\frac{M^2}{s}} e^{\pm y}$$



data from e+A DIS



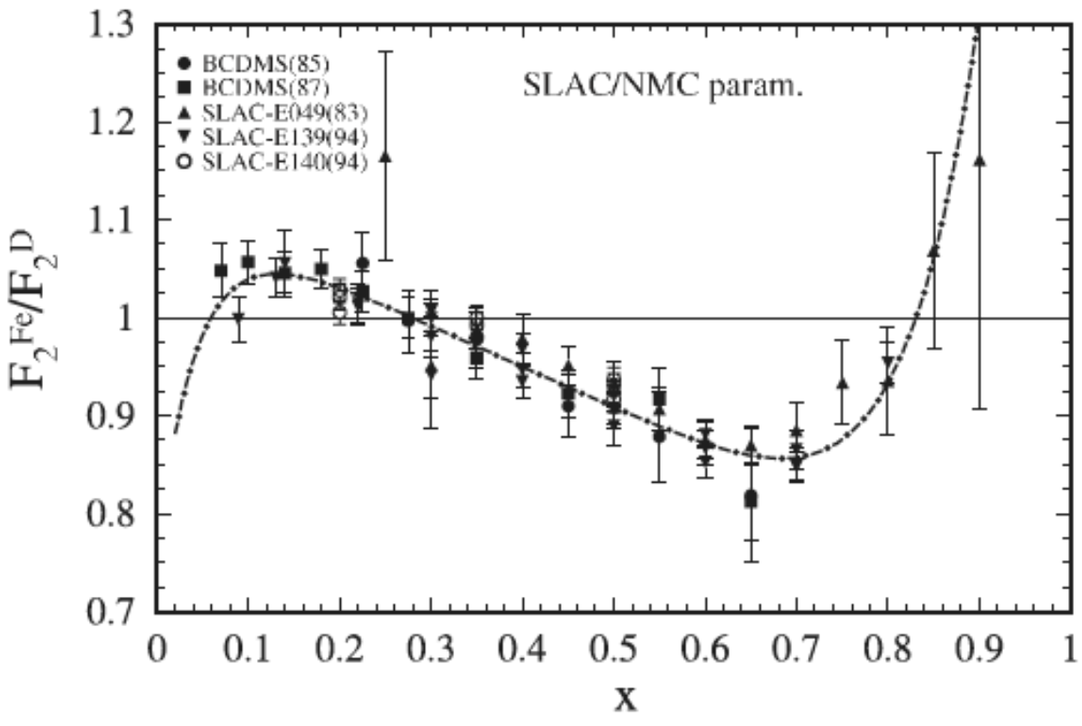
ratio of proton to nuclear PDF

Motivation

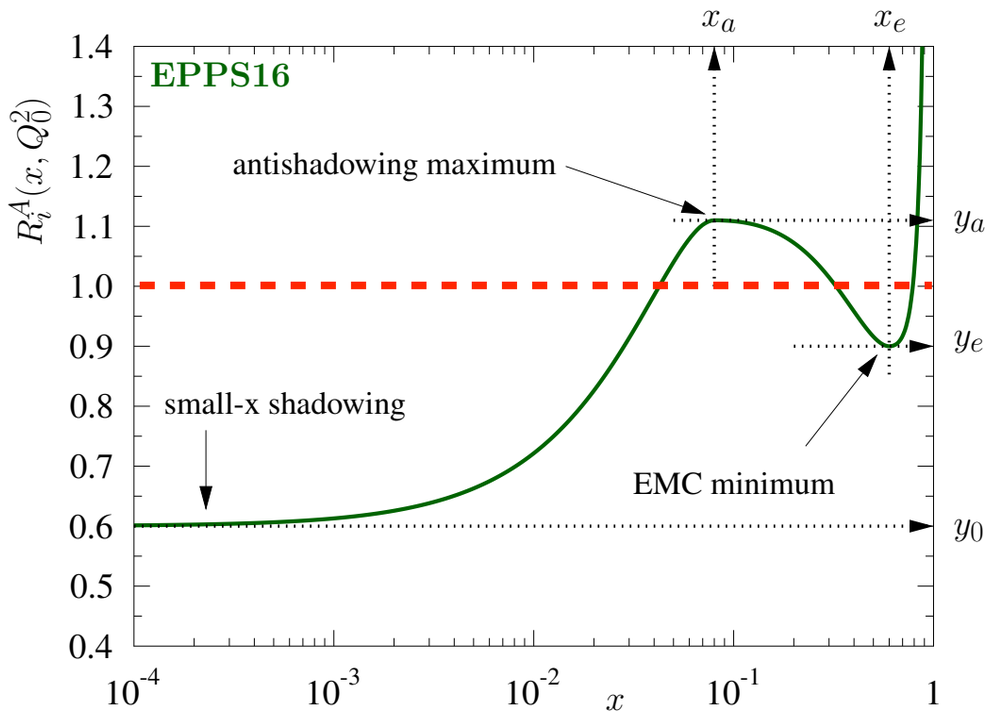
- Not being able to access the real x_1, x_2 from measuring the kinematics is not exclusive of the Drell-Yan process. It happens for all p+p collisions.

- And can have significant impact: e.g. in p+Pb collisions

$$x_1, x_2 = \sqrt{\frac{M^2}{s}} e^{\pm y}$$



data from e+A DIS



ratio of proton to nuclear PDF

- We want to use ML to find a link between the measurable quantities and the parton momentum fractions.

Accessing the kinematics using ML

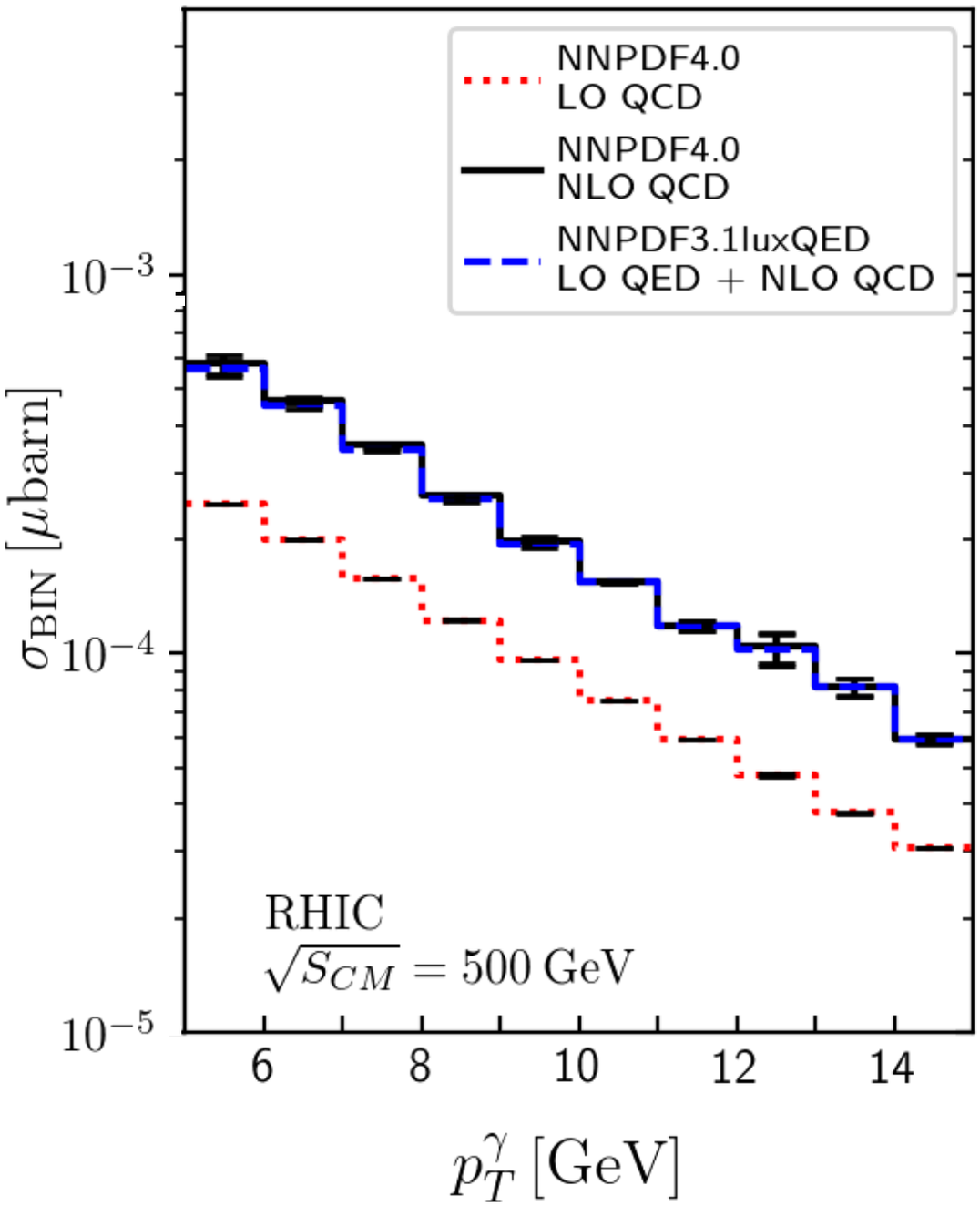
Renteria-Estrada et *al.*,
arXiv:2112.05043 [hep-ph]

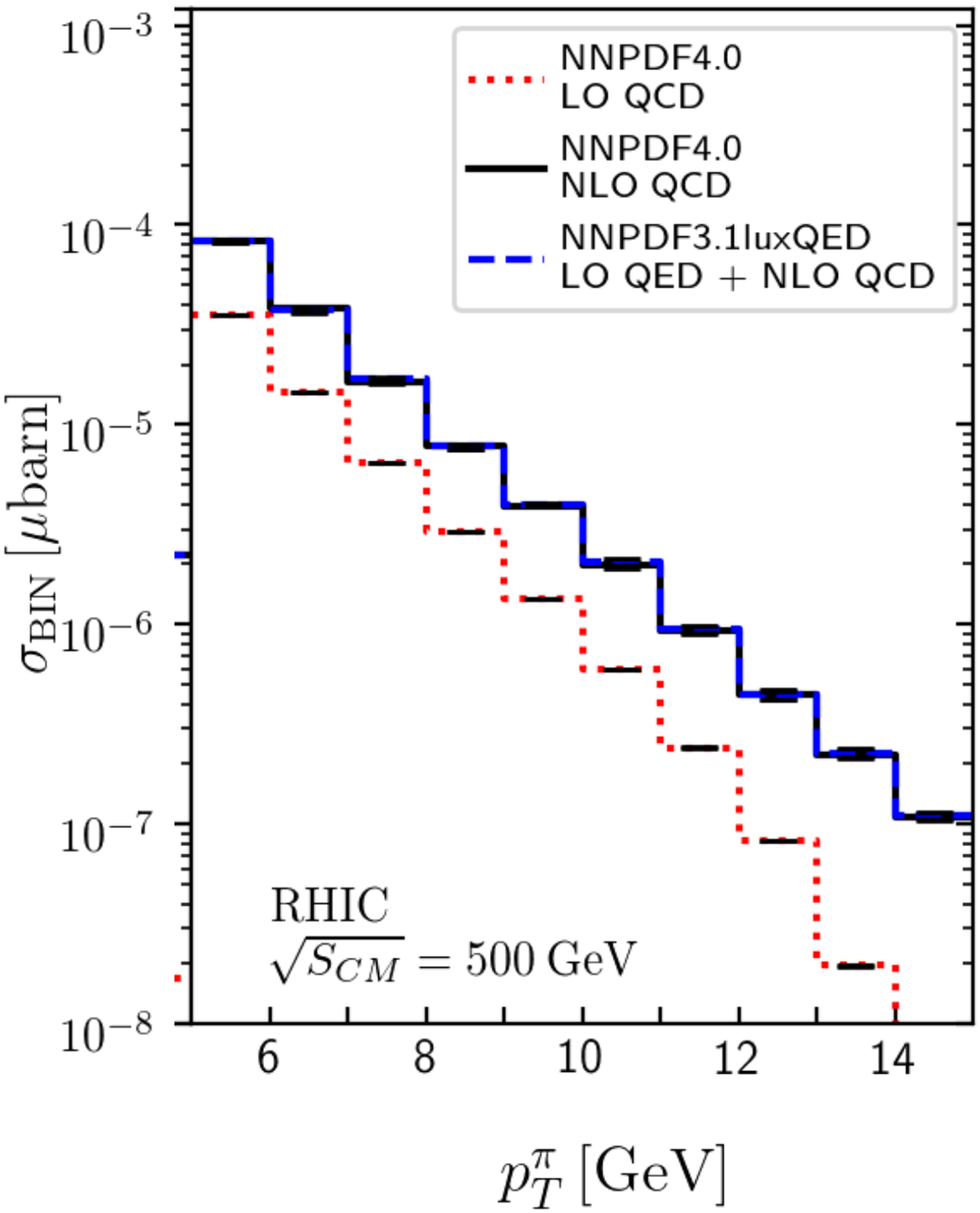
- We looked at one particular process: $p + p \rightarrow \pi^+ + \gamma$
- Reconstructed x_1, x_2 and z from momenta of π^+, γ
- For RHIC kinematics, so we could compare with previous results.

D. de Florian and G. Sborlini,
Phys.Rev.D 83 (2011) 074022

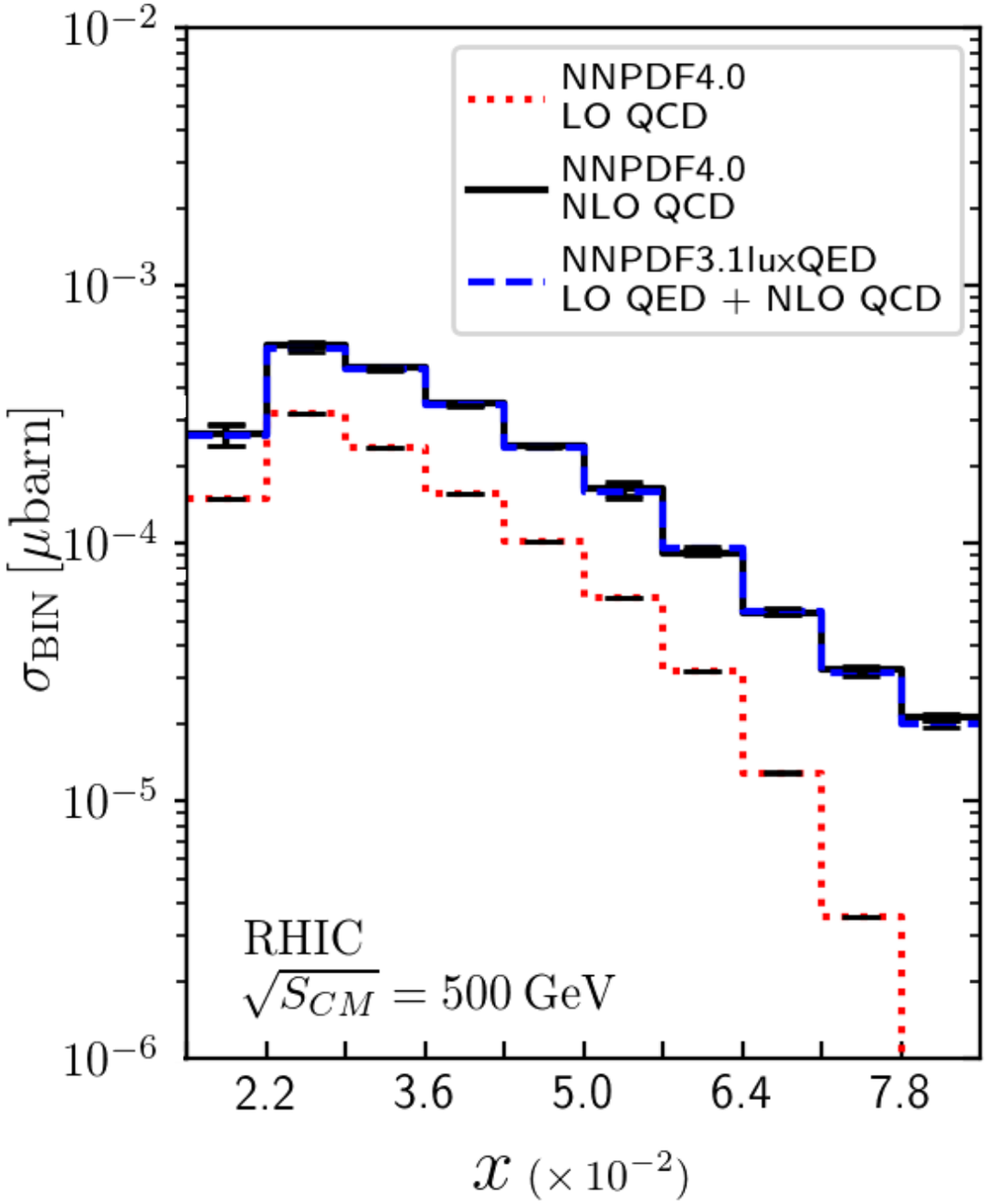
First: check the dependences on the kinematics

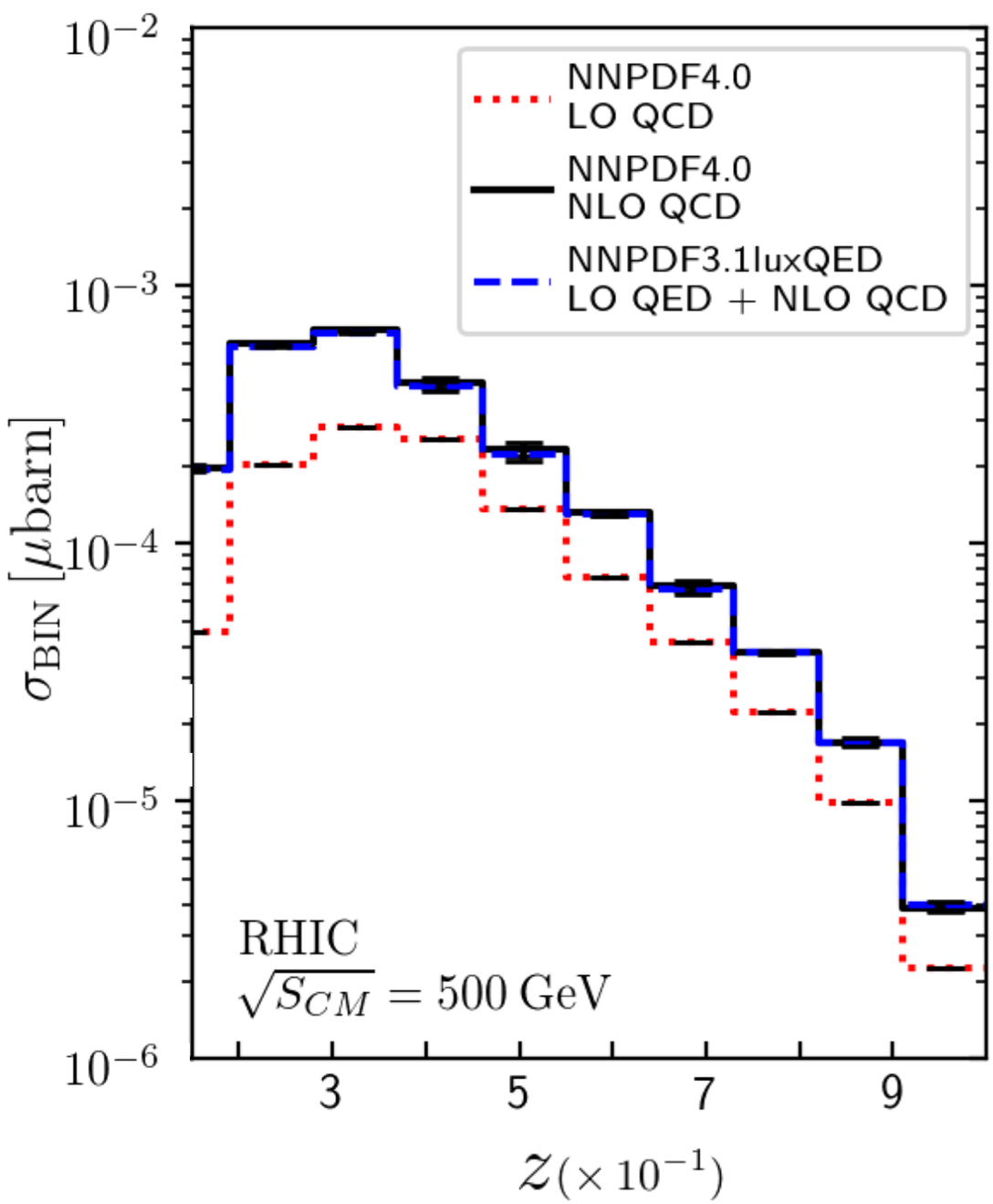
- Transverse momentum dependence:





- $x_1 = x, z$ dependences:



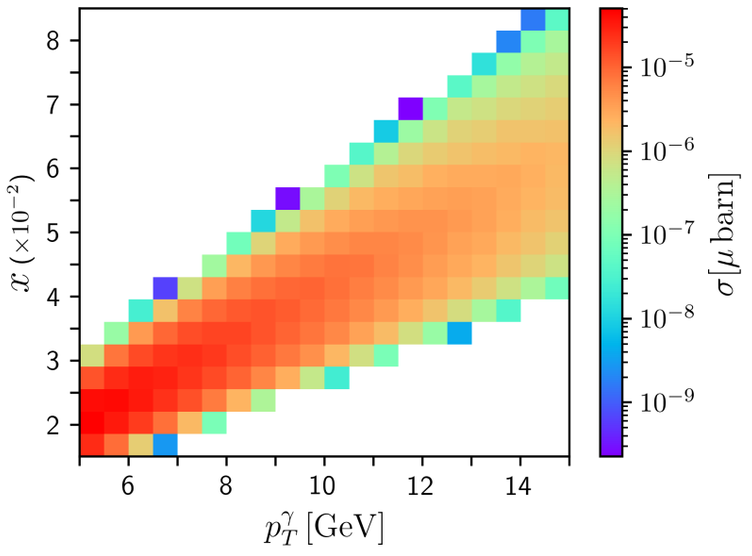


Second: check correlations

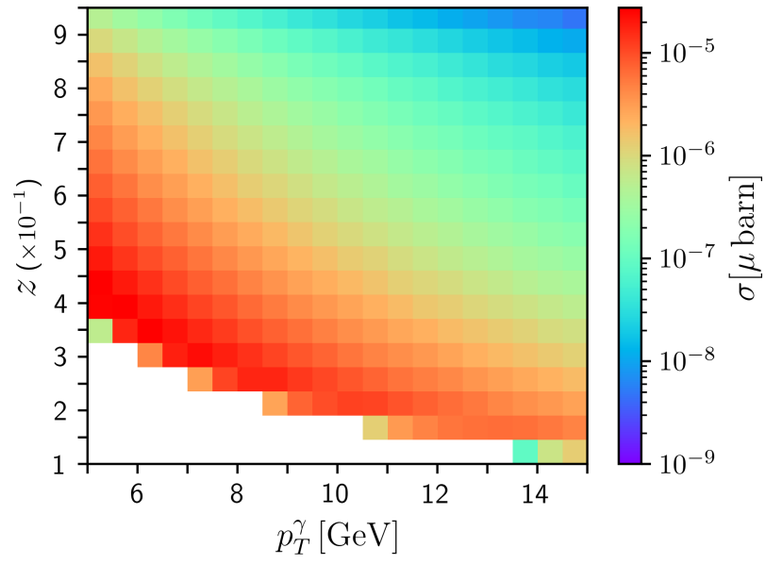
LO Kinematics

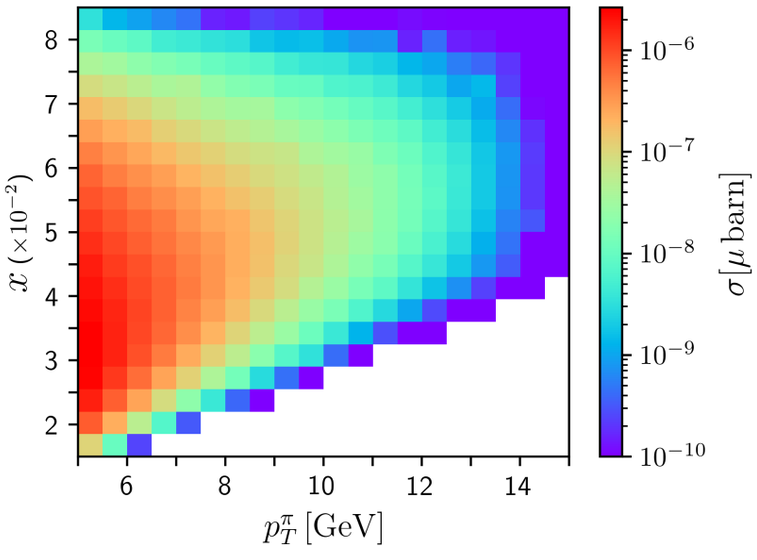
$$x_{1,2} = \frac{p_T^\gamma}{\sqrt{s}}\left(e^{\eta^{\pm\pi}} + e^{\eta^{\pm\gamma}}\right) \qquad z = \frac{p_T^\pi}{p_T^\gamma}$$

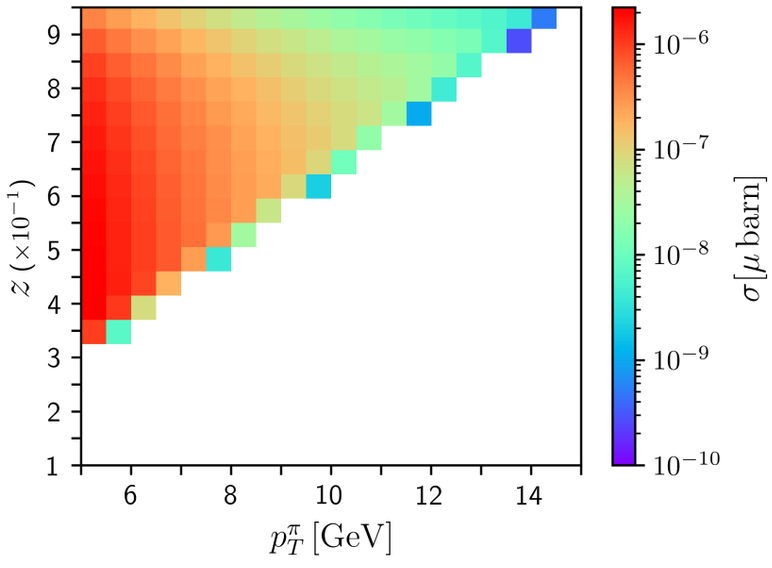
- x vs. p_T



- z vs. p_T



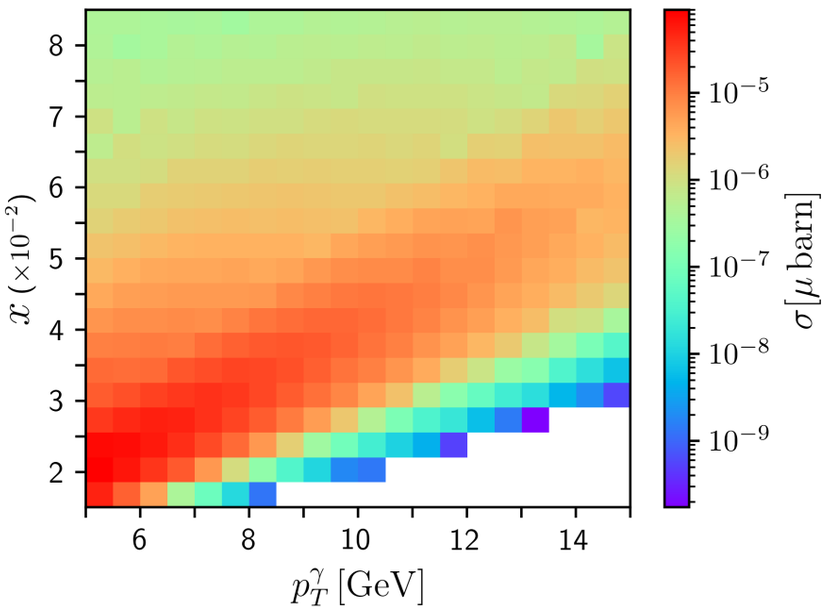




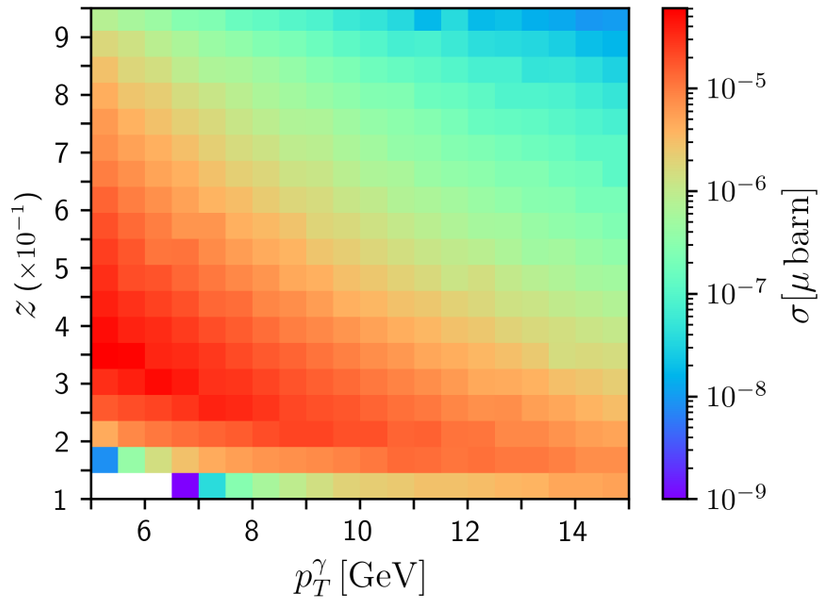
Second: check correlations

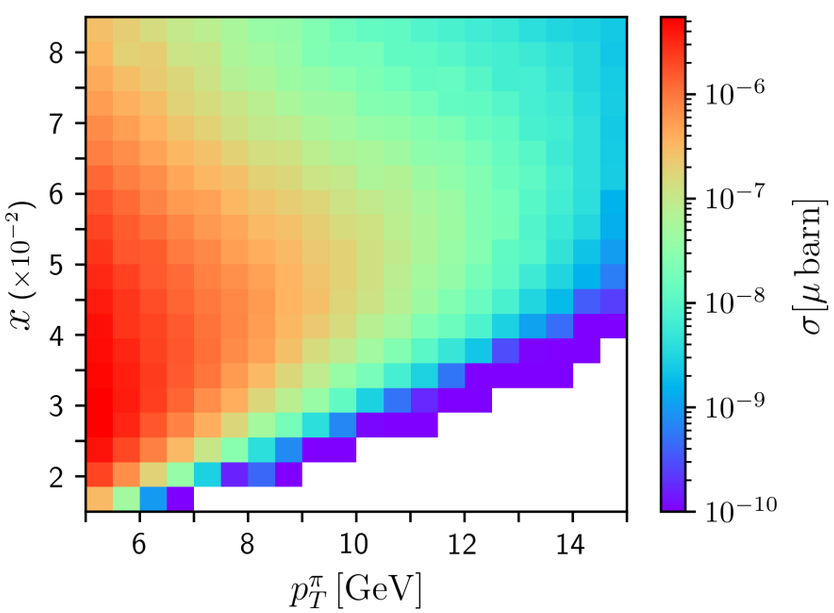
NLO Kinematics $x_{1,2} = ?$ $z = ?$

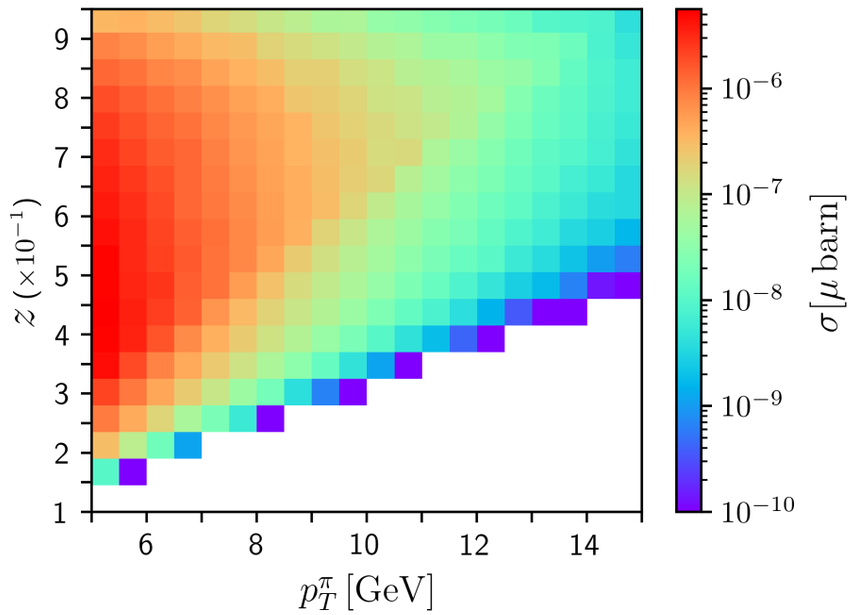
- x vs. p_T



- z vs. p_T







Kinematics: LO

$$x_{1,2}^{rec.} = \frac{p_T^{\gamma}}{\sqrt{s}} \left(e^{\pm\eta^{\pi}} + e^{\pm\eta^{\gamma}} \right)$$

Kinematics: NLO

$$x_{1,2}^{rec.} = ?$$

Kinematics: LO

$$x_{1,2}^{rec.} = \frac{p_T^{\gamma}}{\sqrt{s}}\left(e^{\pm\eta^{\pi}} + e^{\pm\eta^{\gamma}}\right)$$

Kinematics: NLO

$$x_{1,2}^{rec.} = ?$$

- Experimental collaborations used

$$x_{1,2}^{rec.} = \frac{p_T^{\gamma} e^{\pm\eta^{\pi}} - \cos(\phi^{\pi} - \phi^{\gamma}) p_T^{\gamma} e^{\pm\eta^{\gamma}}}{\sqrt{s}}$$

Kinematics: LO

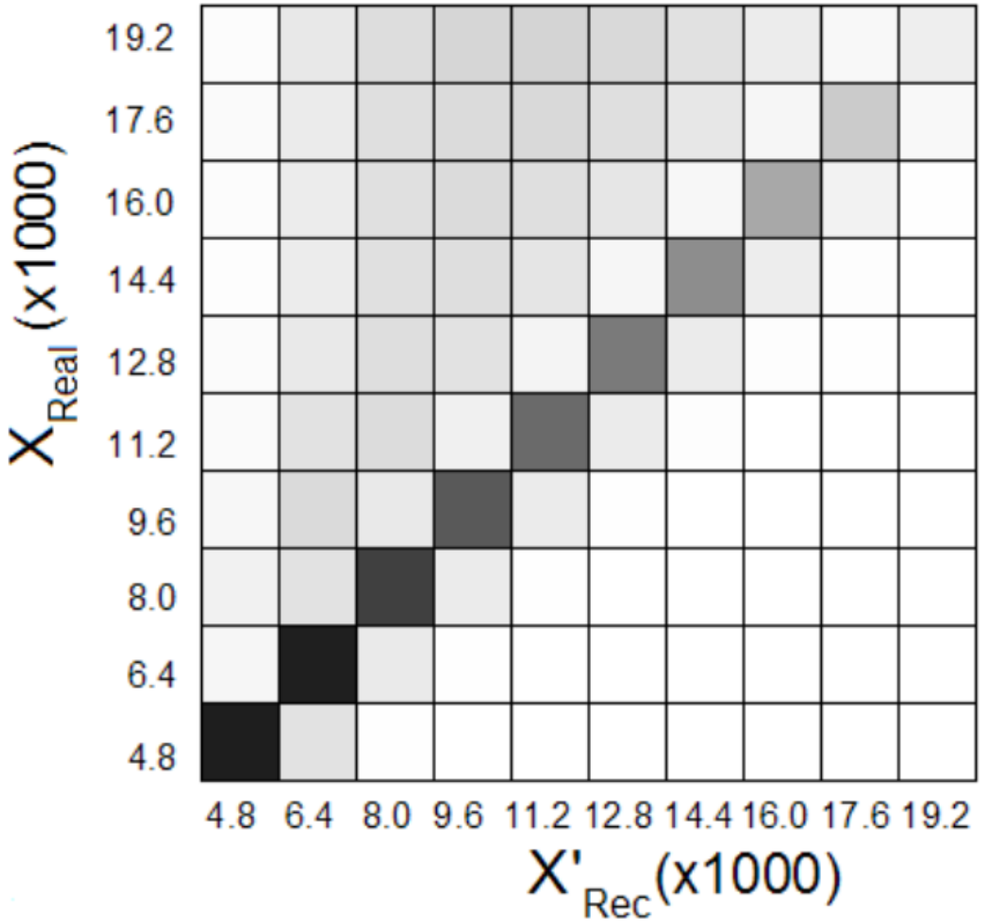
$$x_{1,2}^{rec.} = \frac{p_T^{\gamma}}{\sqrt{s}}\left(e^{\pm\eta^{\pi}} + e^{\pm\eta^{\gamma}}\right)$$

Kinematics: NLO

$$x_{1,2}^{rec.} = ?$$

- Experimental collaborations used

$$x_{1,2}^{rec.} = \frac{p_T^{\gamma} e^{\pm\eta^{\pi}} - \cos(\phi^{\pi} - \phi^{\gamma}) p_T^{\gamma} e^{\pm\eta^{\gamma}}}{\sqrt{s}}$$



D. de Florian, G. Sborlini, PRD 83, 074022

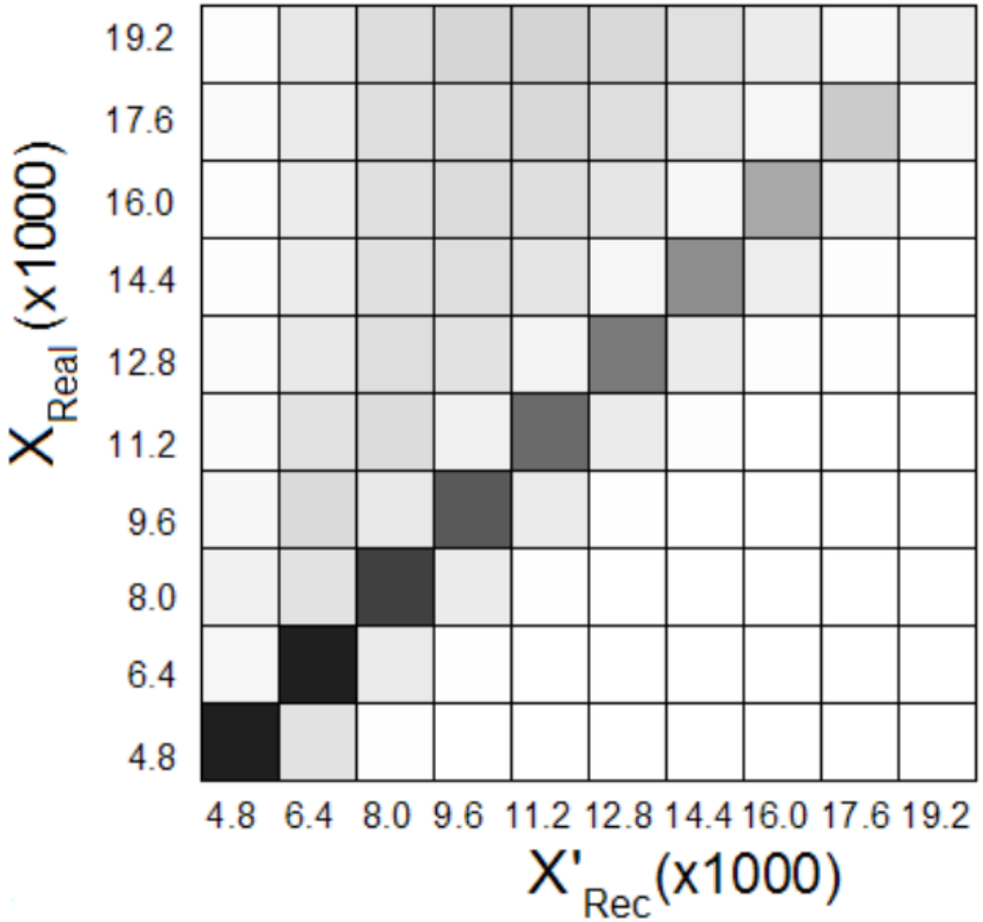
Kinematics: LO

$$x_{1,2}^{rec.} = \frac{p_T^\gamma}{\sqrt{s}}\left(e^{\pm\eta^\pi} + e^{\pm\eta^\gamma}\right) \qquad z^{rec.} = \frac{p_T^\pi}{p_T^\gamma}$$

Kinematics: NLO

$$x_{1,2}^{rec.} = ? \qquad z^{rec.} = ?$$

- Experimental collaborations used $z^{rec.} = -\cos(\phi^\pi - \phi^\gamma)\frac{p_T^\pi}{p_T^\gamma}$



D. de Florian, G. Sborlini, PRD 83, 074022

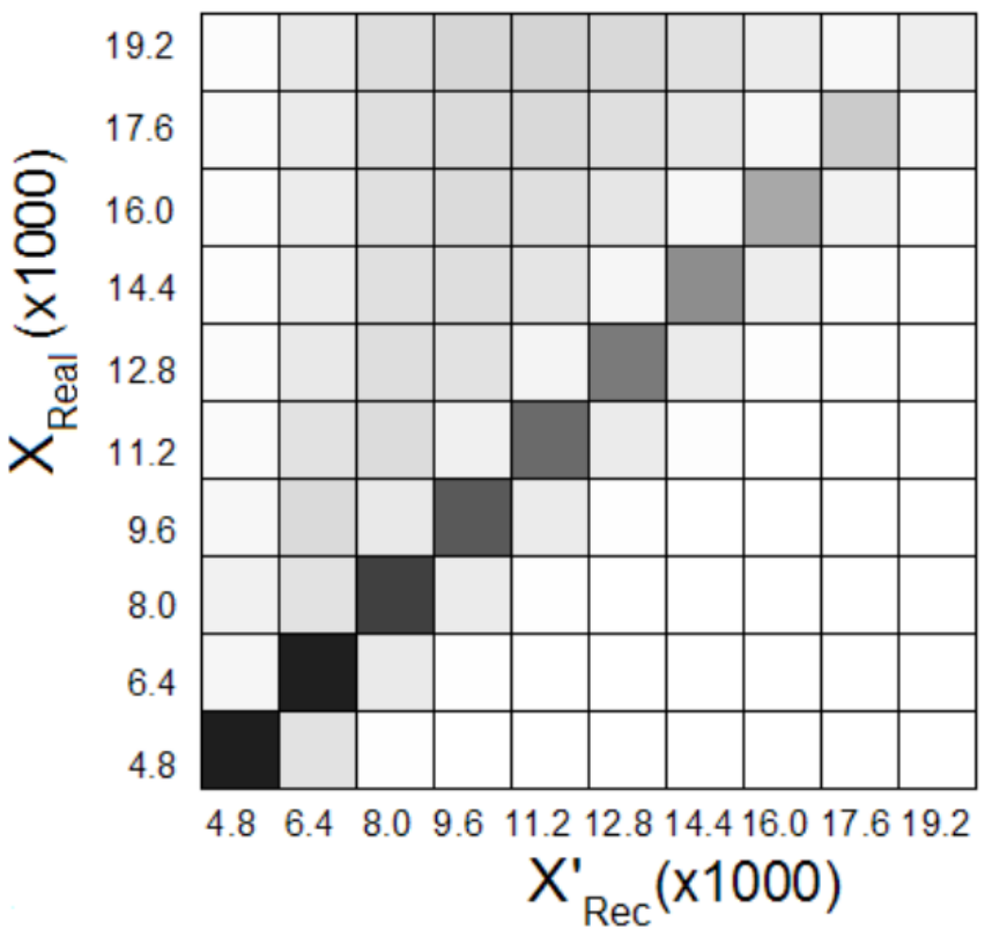
Kinematics: LO

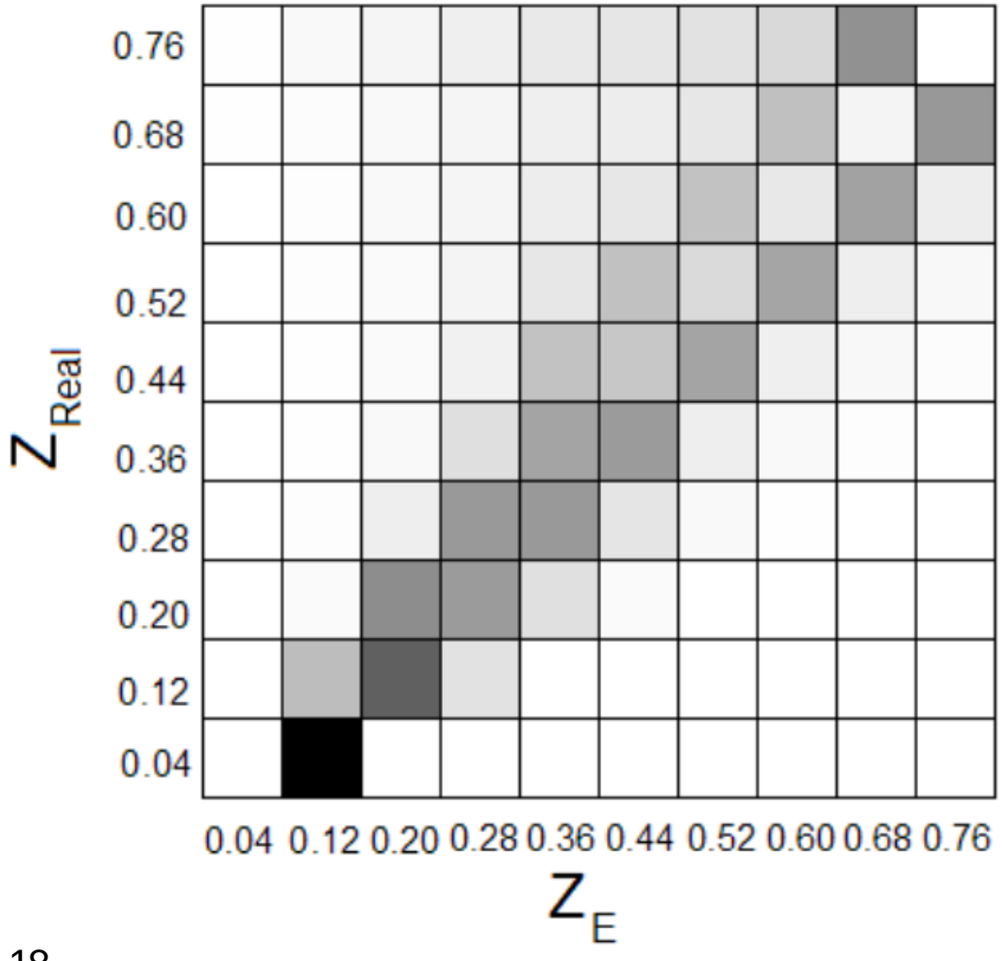
$$x_{1,2}^{rec.} = \frac{p_T^{\gamma}}{\sqrt{s}}\left(e^{\pm\eta^{\pi}} + e^{\pm\eta^{\gamma}}\right) \qquad z^{rec.} = \frac{p_T^{\pi}}{p_T^{\gamma}}$$

Kinematics: NLO

$$x_{1,2}^{rec.} = ? \qquad z^{rec.} = ?$$

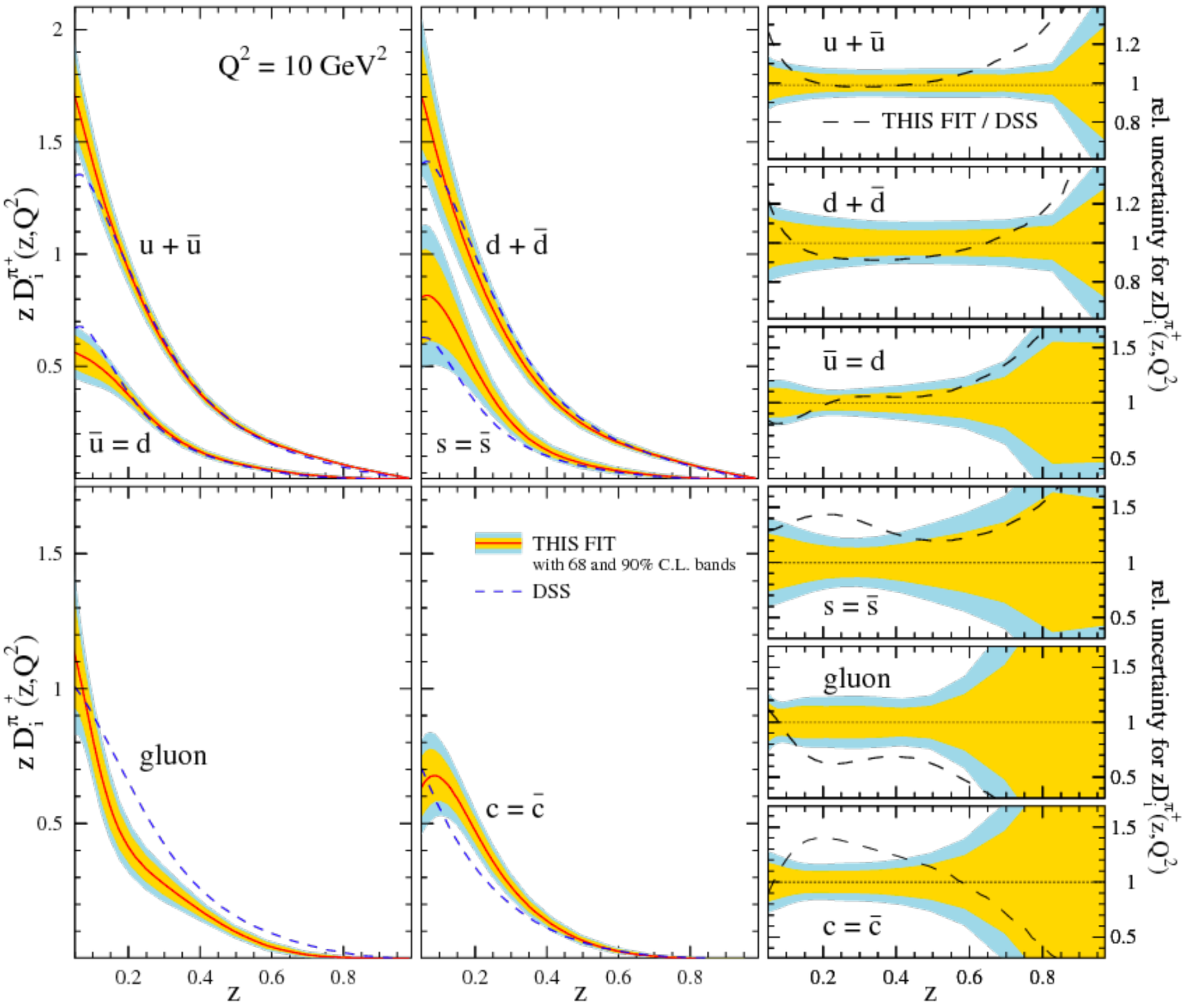
- Experimental collaborations used $z^{rec.} = -\cos(\phi^{\pi} - \phi^{\gamma})\frac{p_T^{\pi}}{p_T^{\gamma}}$





D. de Florian, G. Sborlini, PRD 83, 074022

- In the last decade PDFs and FFs have changed significantly.



D. de Florian, R. Sassot, M. Epele, R.J. Hernández-Pinto and M. Stratmann, PRD 91, 014035.

- In the last decade PDFs and FFs have changed significantly.

- The code now includes QED corrections:

  D. Rentería-Estrada, R. Hernández-Pinto, G. Sborlini, *Symmetry* 13 (2021) 6, 942

- New numerical methods/tools available with tutorials “for dummies”.

- In the last decade PDFs and FFs have changed significantly.

- The code now includes QED corrections:

  D. Rentería-Estrada, R. Hernández-Pinto, G. Sborlini, *Symmetry* 13 (2021) 6, 942

- New numerical methods/tools available with tutorials “for dummies”.

- We’re the dummies: we want to apply machine learning techniques to access the real momentum fractions and lower the assumptions.

- At NLO we have real $(2 \to 3)$ and virtual $(2 \to 2)$ contributions and counterterms $(2 \to 2)$.

- Cancellations can only happen in the MC integration when histograming.

$$\left\{\bar{p}_T^{\gamma}, \bar{p}_T^{\pi}, \bar{\eta}^{\gamma}, \bar{\eta}^{\pi}, \overline{\cos}(\phi^{\pi} - \phi^{\gamma})\right\} \in \bar{\mathscr{V}}_{EXP}$$

$$\sigma_j(\bar{p}_T^{\gamma}, \bar{p}_T^{\pi}, \bar{\eta}^{\gamma}, \bar{\eta}^{\pi}, \overline{\cos}(\phi^{\pi} - \phi^{\gamma})) = \int_{(p_T^{\gamma})_{j,MIN}}^{(p_T^{\gamma})_{j,MAX}} dp_T^{\gamma} \int_{(p_T^{\pi})_{j,MIN}}^{(p_T^{\pi})_{j,MAX}} dp_T^{\pi} \int dx_1 dx_2 dz \; d\bar{\sigma}$$

- At NLO we have real $(2 \to 3)$ and virtual $(2 \to 2)$ contributions and counterterms $(2 \to 2)$.

- Cancellations can only happen in the MC integration when histograming.

$$\left\{\bar{p}_T^{\gamma}, \bar{p}_T^{\pi}, \bar{\eta}^{\gamma}, \bar{\eta}^{\pi}, \overline{\cos}(\phi^{\pi} - \phi^{\gamma})\right\} \in \bar{\mathscr{V}}_{EXP}$$

$$\sigma_j(\bar{p}_T^{\gamma}, \bar{p}_T^{\pi}, \bar{\eta}^{\gamma}, \bar{\eta}^{\pi}, \overline{\cos}(\phi^{\pi} - \phi^{\gamma})) = \int_{(p_T^{\gamma})_{j,MIN}}^{(p_T^{\gamma})_{j,MAX}} dp_T^{\gamma} \int_{(p_T^{\pi})_{j,MIN}}^{(p_T^{\pi})_{j,MAX}} dp_T^{\pi} \int dx_1 dx_2 dz \, d\bar{\sigma}$$

- We weight the momentum fractions from the MC with the per-bin cross-section

$$(x_1)_j = \sum_i (x_1)_i \frac{d\sigma_j}{dx_1}(p_j; (x_1)_i) \qquad (z)_j = \sum_i (z)_i \frac{d\sigma_j}{dz}(p_j; (z)_i)$$

- At NLO we have real $(2 \to 3)$ and virtual $(2 \to 2)$ contributions and counterterms $(2 \to 2)$.

- Cancellations can only happen in the MC integration when histograming.

$$\left\{\bar{p}_T^{\gamma}, \bar{p}_T^{\pi}, \bar{\eta}^{\gamma}, \bar{\eta}^{\pi}, \overline{\cos}(\phi^{\pi} - \phi^{\gamma})\right\} \in \bar{\mathscr{V}}_{EXP}$$

$$\sigma_j(\bar{p}_T^{\gamma}, \bar{p}_T^{\pi}, \bar{\eta}^{\gamma}, \bar{\eta}^{\pi}, \overline{\cos}(\phi^{\pi} - \phi^{\gamma})) = \int_{(p_T^{\gamma})_{j,MIN}}^{(p_T^{\gamma})_{j,MAX}} dp_T^{\gamma} \int_{(p_T^{\pi})_{j,MIN}}^{(p_T^{\pi})_{j,MAX}} dp_T^{\pi} \int dx_1 dx_2 dz \, d\bar{\sigma}$$

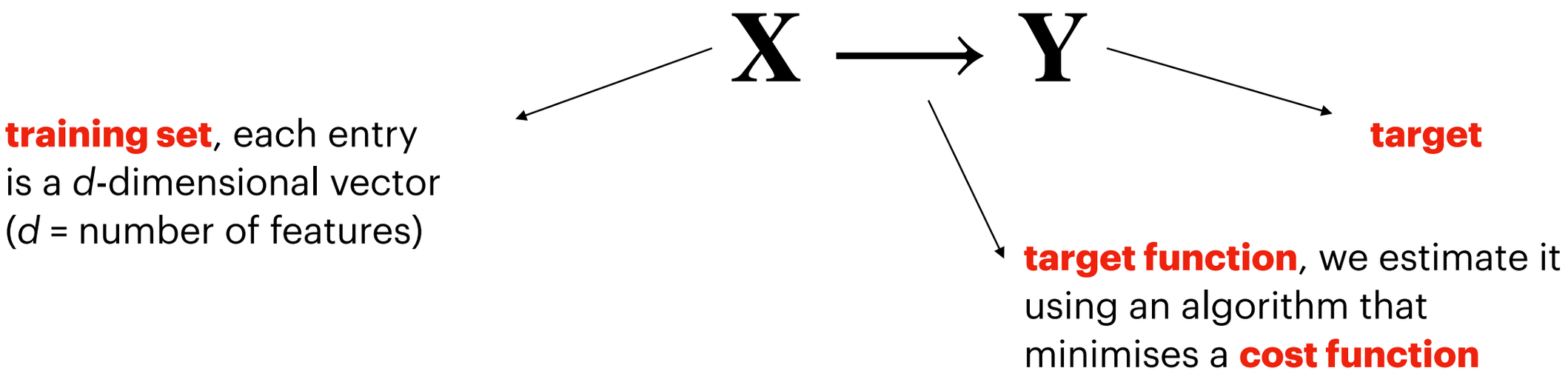
- We weight the momentum fractions from the MC with the per-bin cross-section

$$(x_1)_j = \sum_i (x_1)_i \frac{d\sigma_j}{dx_1}(p_j; (x_1)_i) \qquad (z)_j = \sum_i (z)_i \frac{d\sigma_j}{dz}(p_j; (z)_i)$$

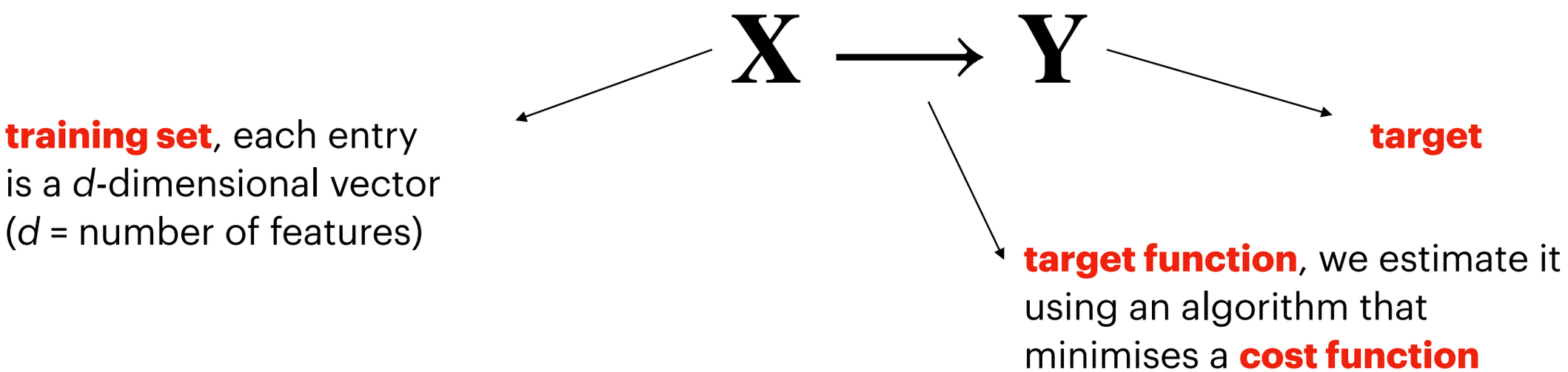
- With this we search for the mapping

$$X_{1,REC} : \ \bar{\mathscr{V}}_{EXP} \to \overline{X}_{1,REAL} = \{(x_1)_j\}$$

- In general, in ML



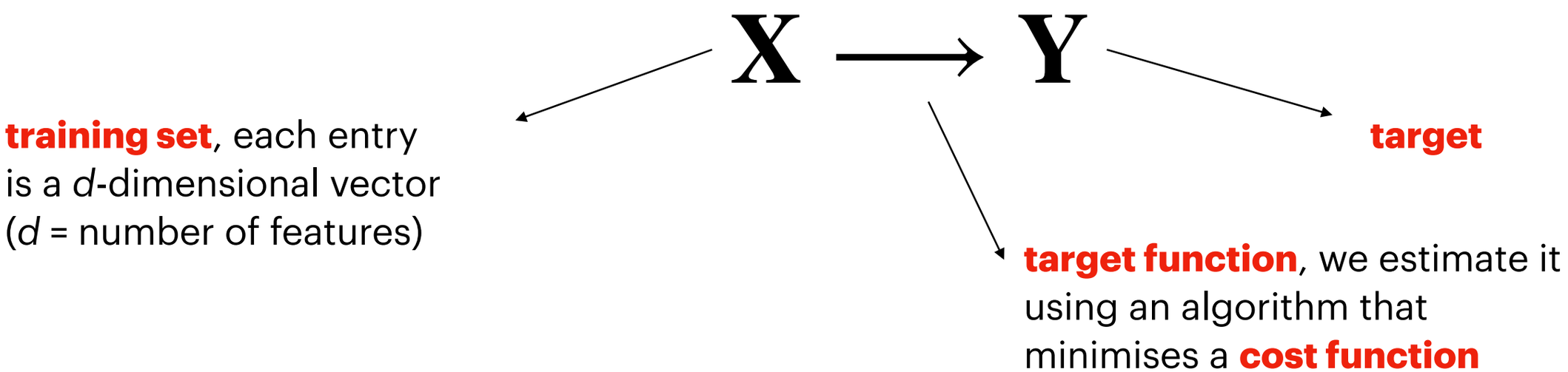
- In general, in ML



Linear regression

$$\hat{y} = \theta_0 + \theta_1 x^{(1)}$$

- In general, in ML



Linear regression

$$\hat{y} = \theta_0 + \theta_1 x^{(1)}$$

- "cheat": ***linear*** means linear ***in the parameters***.

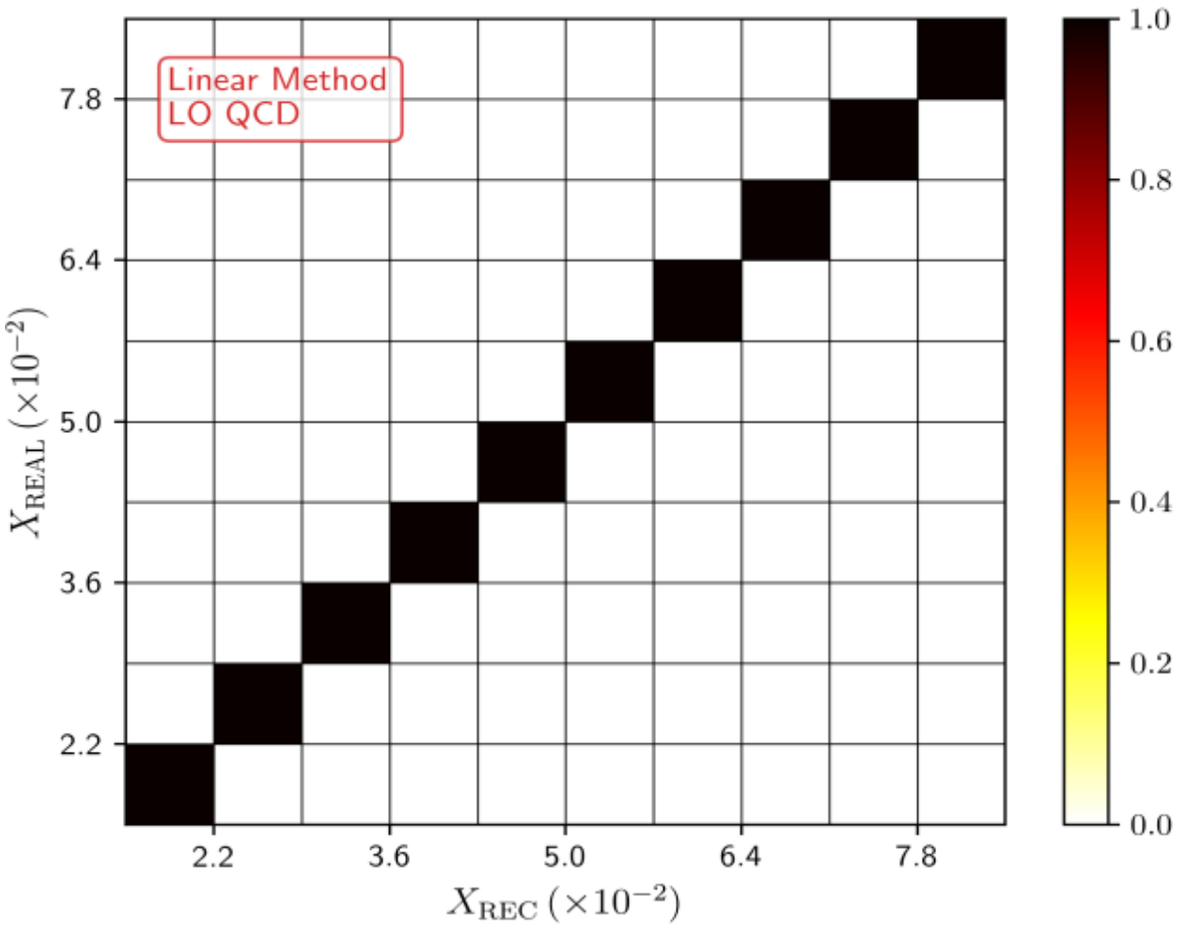
$$\left[x^{(1)}\right]^2 = x^{(2)} \longrightarrow \hat{y} = \theta_0 + \theta_1 x^{(1)} + \theta_2 x^{(2)}$$

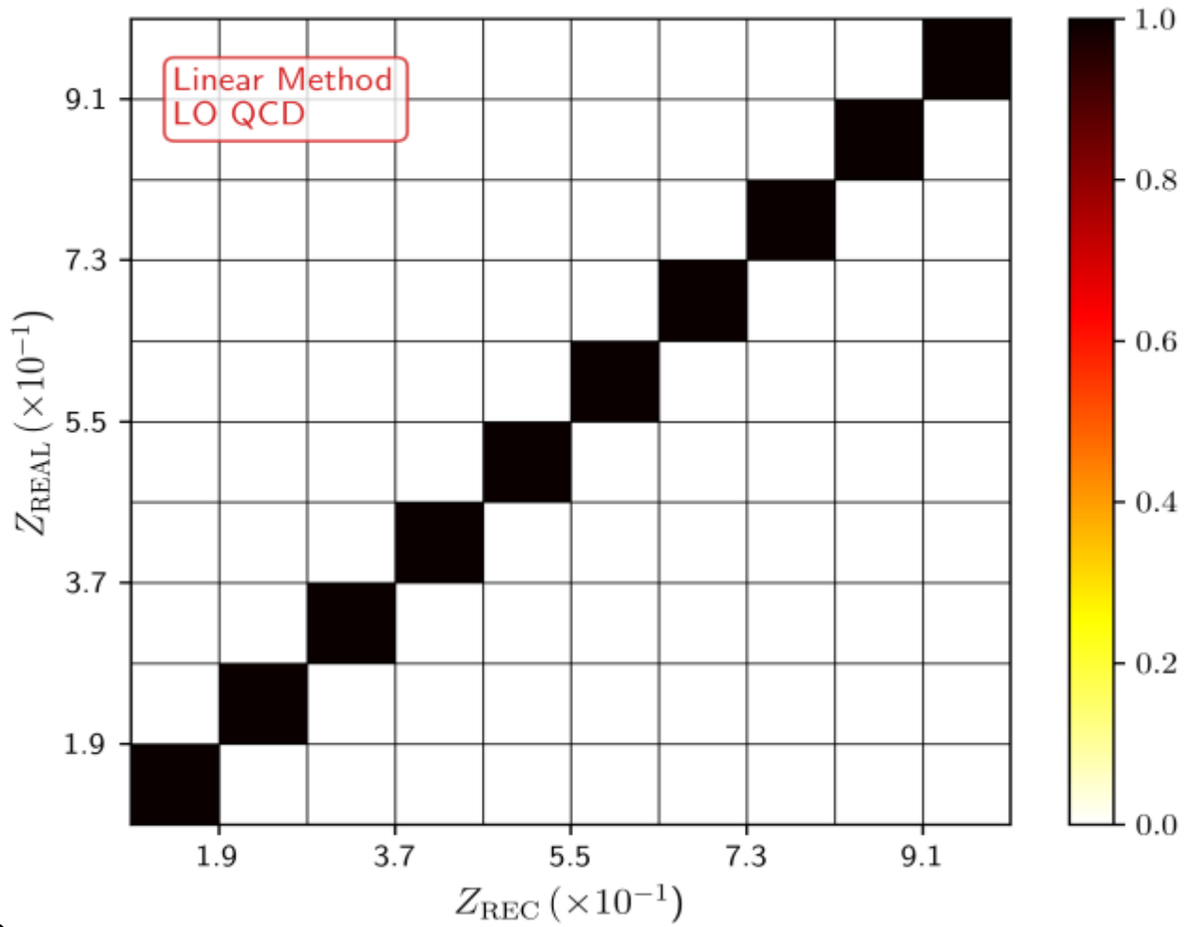
- the parameters minimise $\min_{\theta} \|\hat{y} - y\|_2^2$

- Let us start with LO and use linear regression:

- Basis: $$\mathscr{B}_{LO} = \left\{ \frac{p_T^\gamma e^{\eta^\pi}}{\sqrt{s}}, \frac{p_T^\gamma e^{\eta^\gamma}}{\sqrt{s}}, \frac{p_T^\gamma e^{-\eta^\pi}}{\sqrt{s}}, \frac{p_T^\gamma e^{-\eta^\gamma}}{\sqrt{s}}, \frac{p_T^\pi}{p_T^\gamma} \right\}$$

$$x_{1,2}^{rec.} = \frac{p_T^\gamma}{\sqrt{s}} \left(e^{\eta^{\pm\pi}} + e^{\eta^{\pm\gamma}} \right) \qquad z^{rec.} = \frac{p_T^\pi}{p_T^\gamma}$$





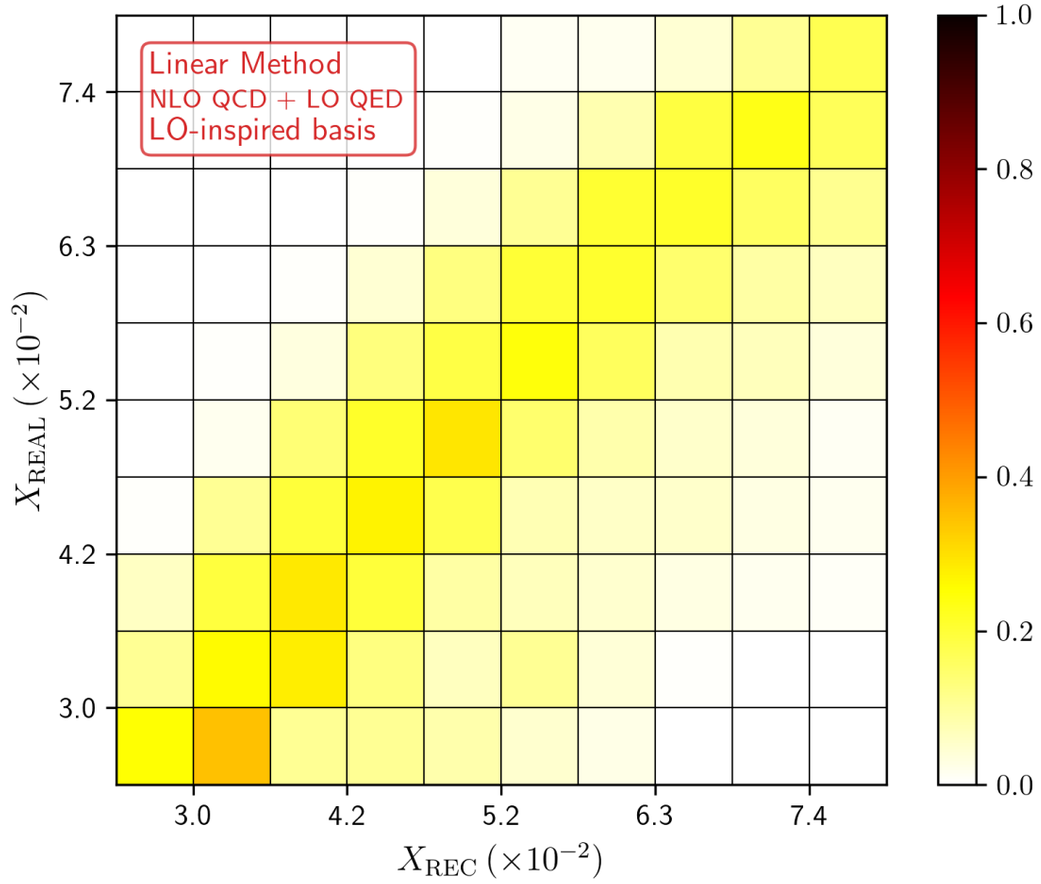
Linear regression

- We used three bases: “LO inspired”, “general”, “physically motivated”

$$x_1^{LO} = \frac{p_T^\gamma e^{\eta^\pi} + p_T^\gamma e^{\eta^\gamma}}{\sqrt{s}}$$

$$x_1^{prop.} = \frac{p_T^\gamma e^{\eta^\pi} - \cos(\phi^\pi - \phi^\gamma) p_T^\gamma e^{\eta^\gamma}}{\sqrt{s}}$$

$$\mathscr{B}_{LO} = \left\{ \frac{p_T^\gamma e^{\eta^\pi}}{\sqrt{s}}, \frac{p_T^\gamma e^{\eta^\gamma}}{\sqrt{s}}, \frac{p_T^\gamma e^{-\eta^\pi}}{\sqrt{s}}, \frac{p_T^\gamma e^{-\eta^\gamma}}{\sqrt{s}}, \frac{p_T^\pi}{p_T^\gamma} \right\}$$



$$\mathscr{B}_{\mathrm{NLO}}^{X_1} = \left\{ \frac{p_T^\gamma}{\sqrt{S_{CM}}} \exp(\eta^\gamma), \frac{p_T^\gamma}{\sqrt{S_{CM}}} \exp(\eta^\pi), \frac{p_T^\pi}{\sqrt{S_{CM}}} \exp(\eta^\gamma), \frac{p_T^\pi}{\sqrt{S_{CM}}} \exp(\eta^\pi), \frac{p_T^\gamma \mathscr{K}_5}{\sqrt{S_{CM}}} \exp(\eta^\gamma), \frac{p_T^\gamma \mathscr{K}_5}{\sqrt{S_{CM}}} \exp(\eta^\pi), \frac{p_T^\pi \mathscr{K}_5}{\sqrt{S_{CM}}} \exp(\eta^\gamma), \frac{p_T^\pi \mathscr{K}_5}{\sqrt{S_{CM}}} \exp(\eta^\pi) \right\}$$

$$\mathscr{K}_5 = \cos(\phi^\pi - \phi^\gamma)$$

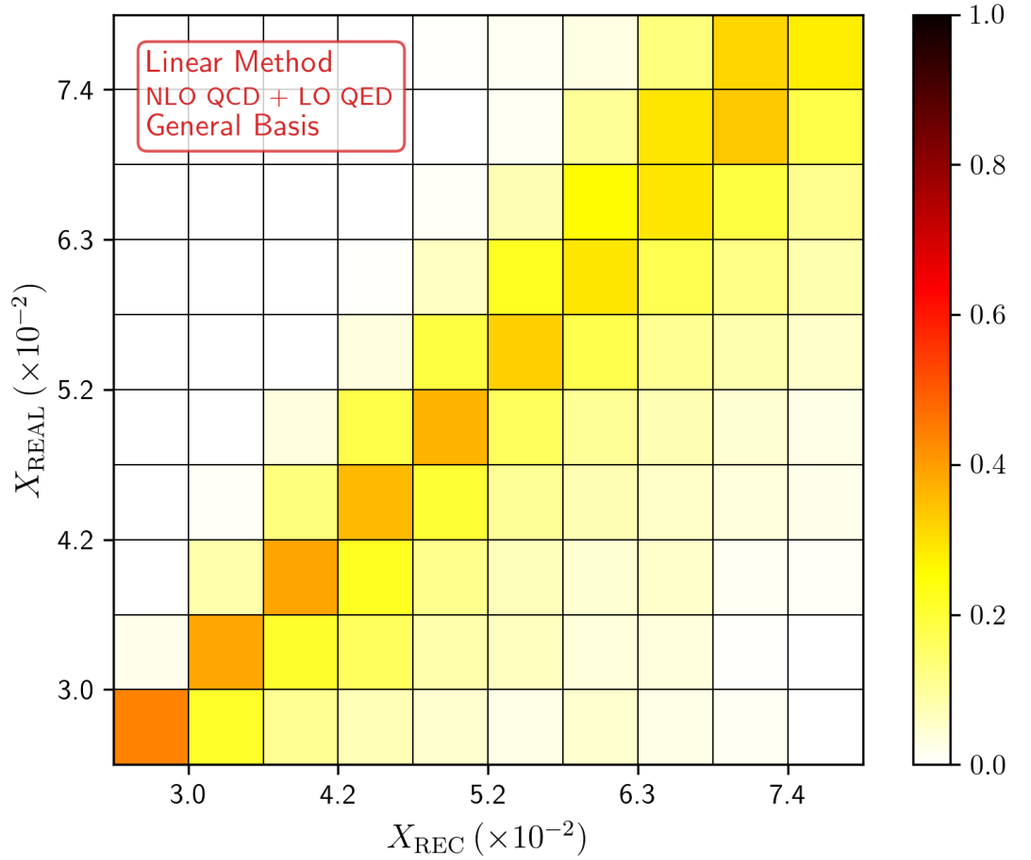
Linear regression

- We used three bases: “LO inspired”, “general”, “physically motivated”

$$\mathscr{K} = \left\{ \frac{p_T^\gamma}{\sqrt{s}}, \frac{p_T^\pi}{\sqrt{s}}, e^{\eta^\gamma}, e^{\eta^\pi}, \cos(\phi^\pi - \phi^\gamma) \right\}$$

$$X_{\mathrm{REC}} = \sum_{i=1, i\neq 5}^{9} (a_i + b_i \mathscr{K}_5)\, \mathscr{K}_i + \sum_{i\leq j, \{i,j\}\neq 5, j-i\neq 5} (c_{ij} + d_{ij} \mathscr{K}_5)\, \mathscr{K}_i \mathscr{K}_j$$

- 81 parameters in total.



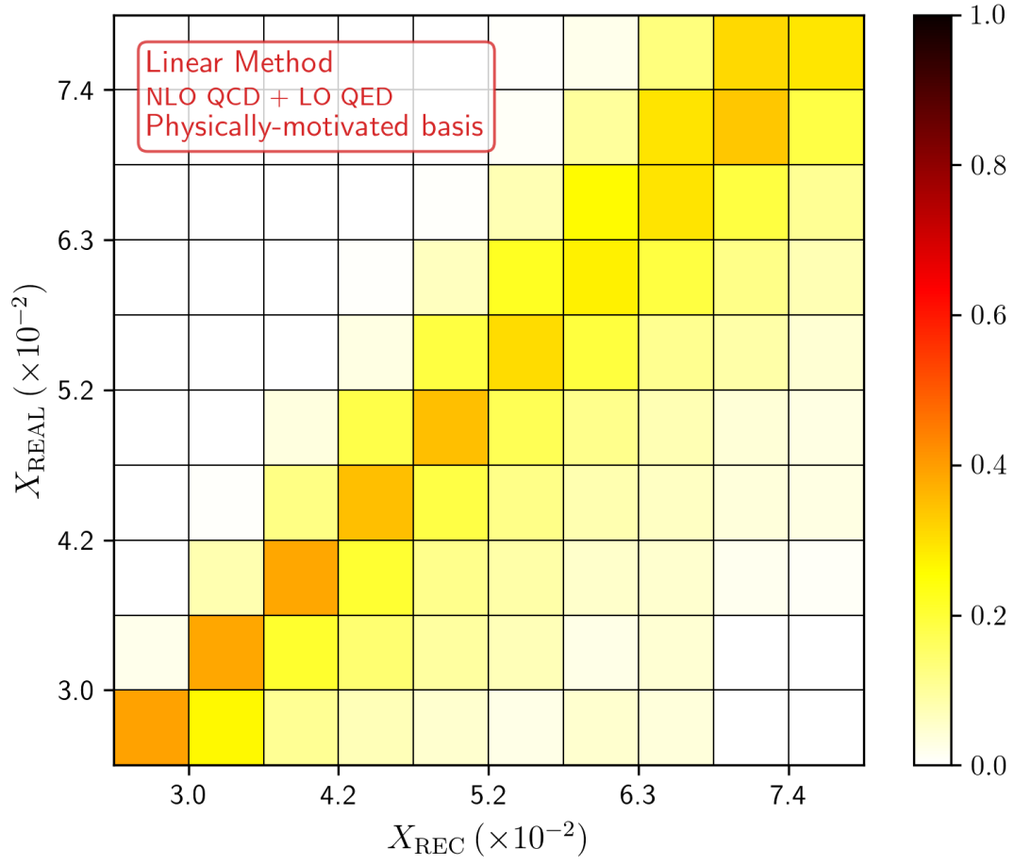
Linear regression

- We used three bases: “LO inspired”, “general”, “physically motivated”

$$\mathscr{K} = \left\{ \frac{p_T^\gamma}{\sqrt{s}}, \frac{p_T^\pi}{\sqrt{s}}, e^{\eta^\gamma}, e^{\eta^\pi}, \cos(\phi^\pi - \phi^\gamma) \right\}$$

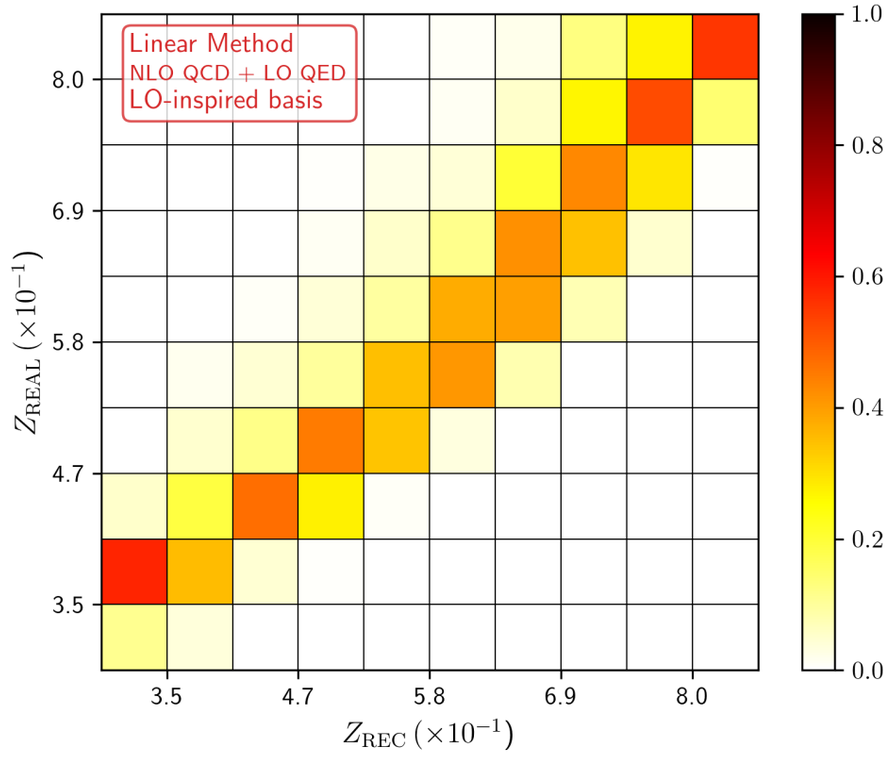
$$X_{\mathrm{REC}} = \sum_{i=1, i\neq 5}^{9} (a_i + b_i \mathscr{K}_5)\, \mathscr{K}_i + \sum_{i\leq j, \{i,j\}\neq 5, j-i\neq 5} (c_{ij} + d_{ij} \mathscr{K}_5)\, \mathscr{K}_i \mathscr{K}_j$$

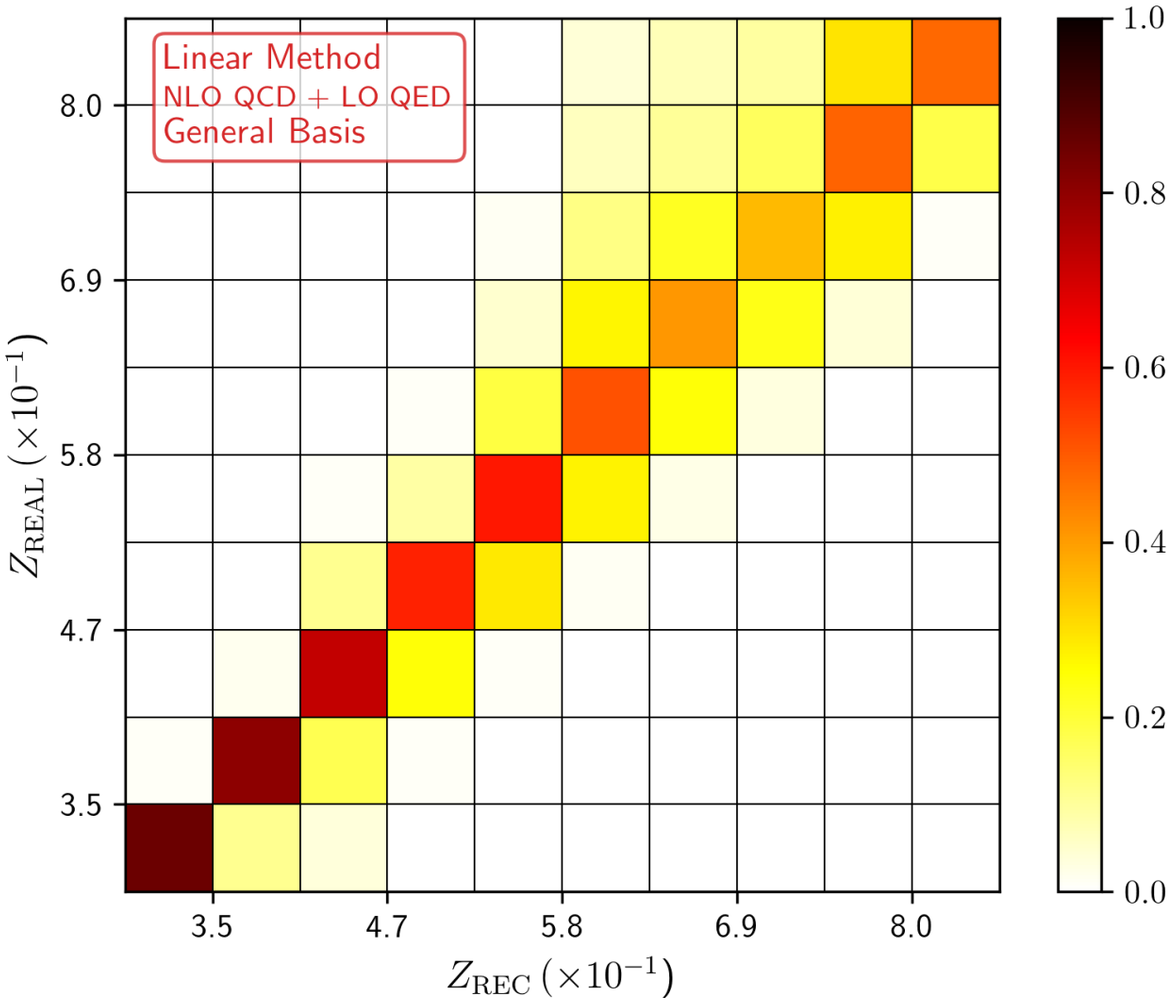
- Remove contributions w.r.t. which we see no dependence (~40 parameters).

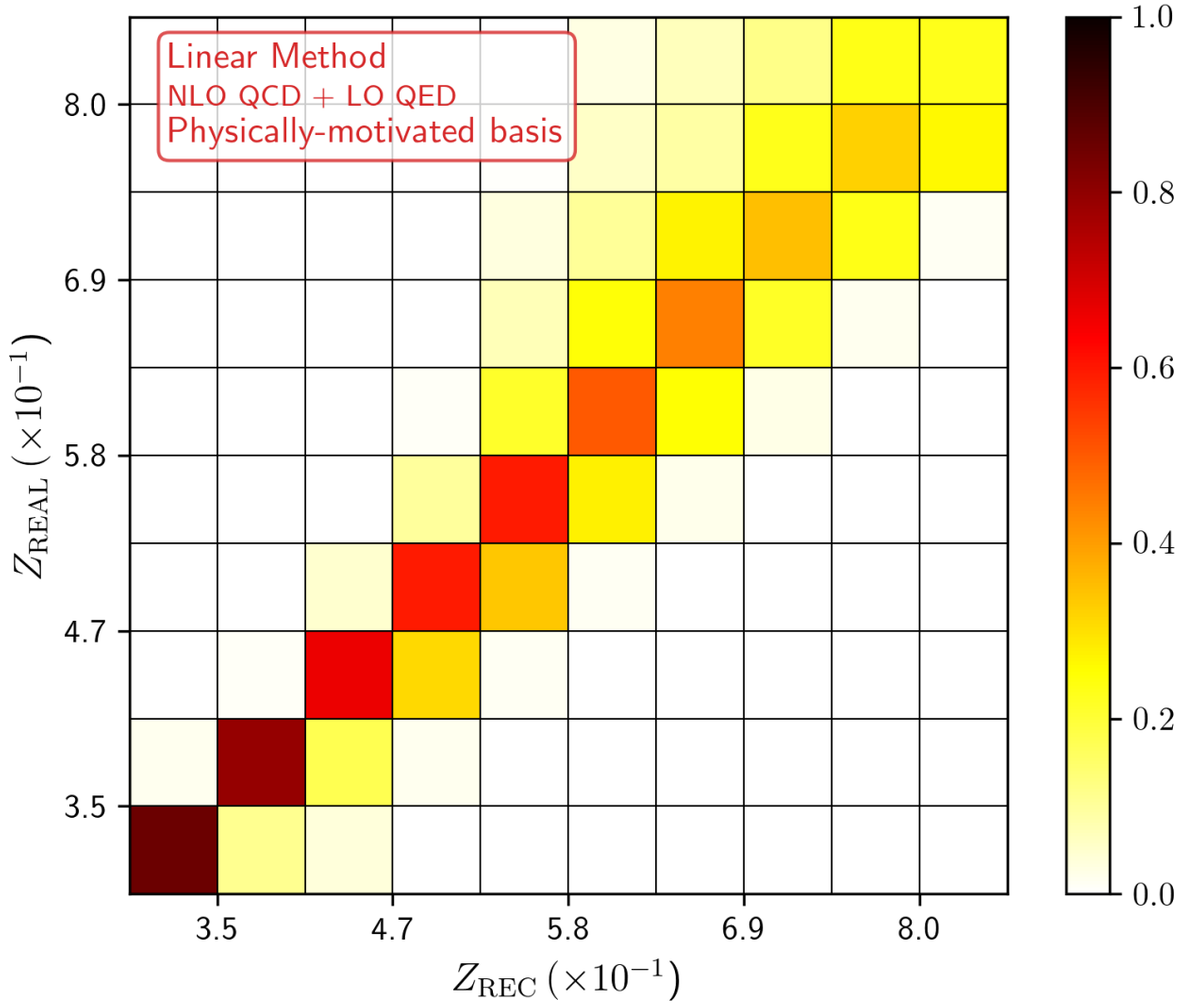


Linear regression

- For z:







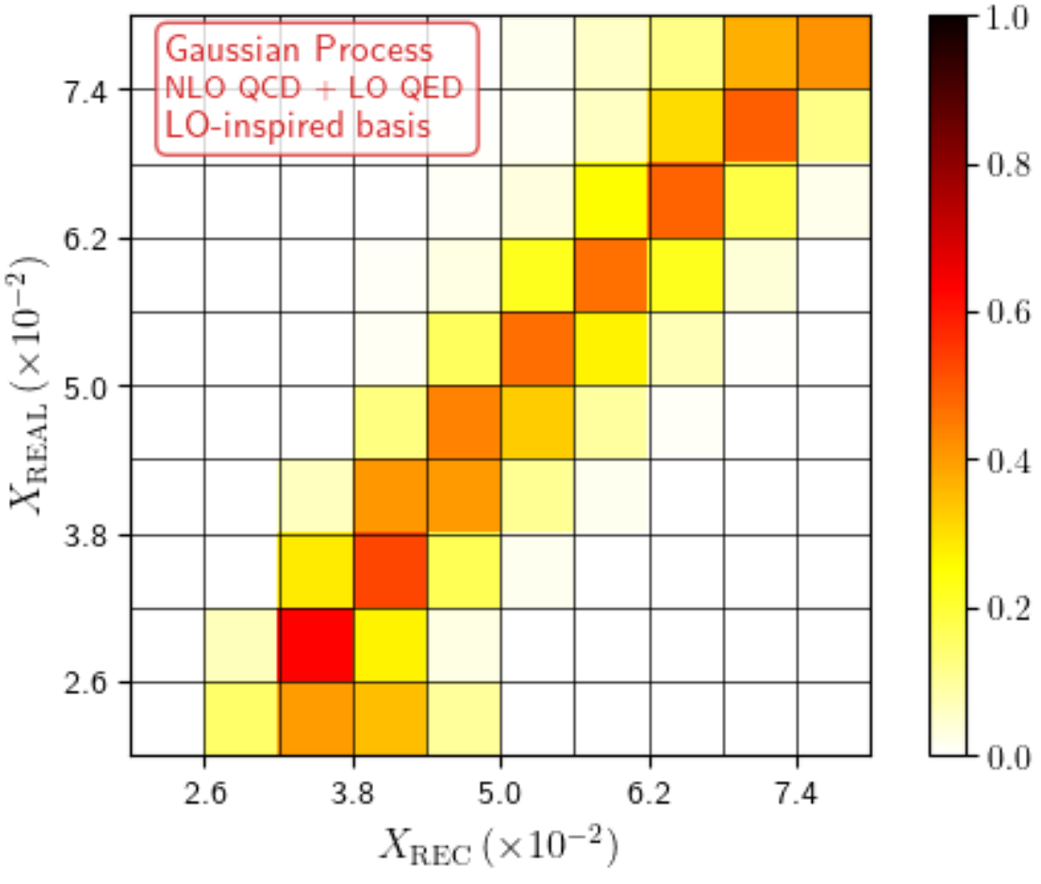
The Radial Basis Function

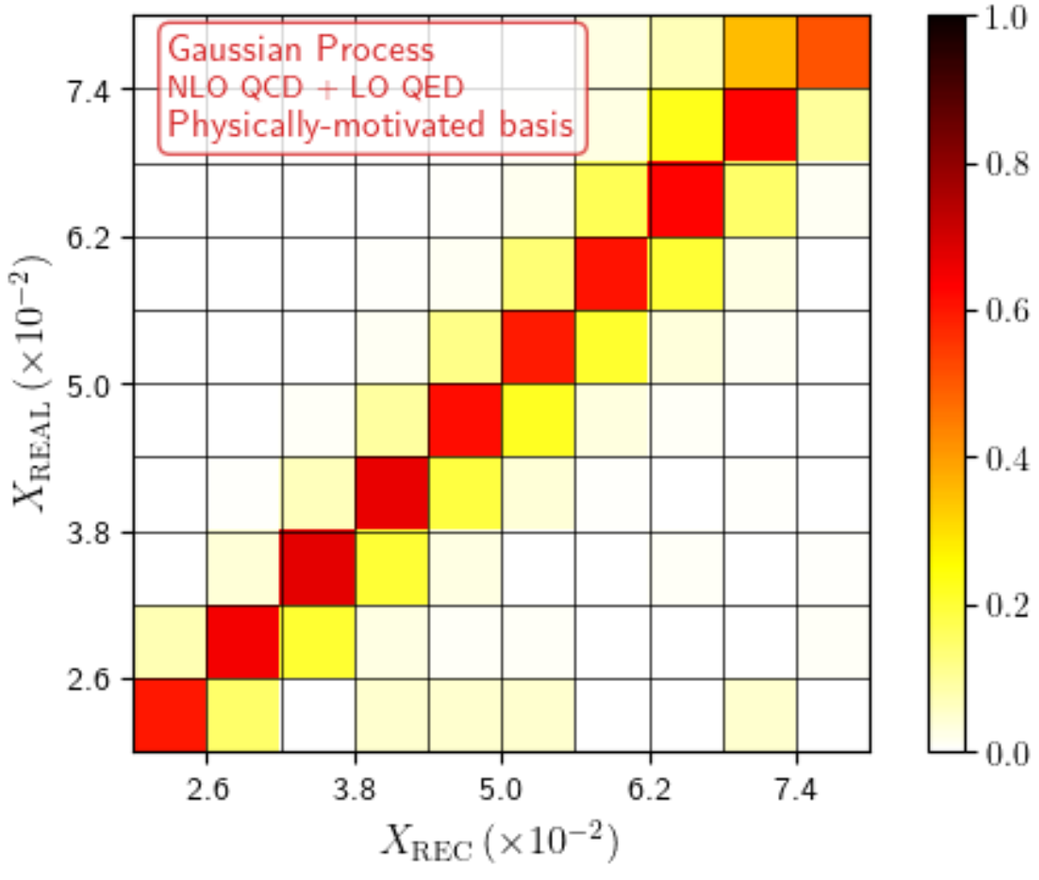
$$\mathbf{X} = \left(1, x_1, x_2\right)^T \longrightarrow \tilde{\mathbf{X}} = \left(1, x_1, x_2, x_1^2, x_1 x_2, x_2^2\right)^T$$

- How to pick $\tilde{\mathbf{X}}$? Before we did it by intuition.

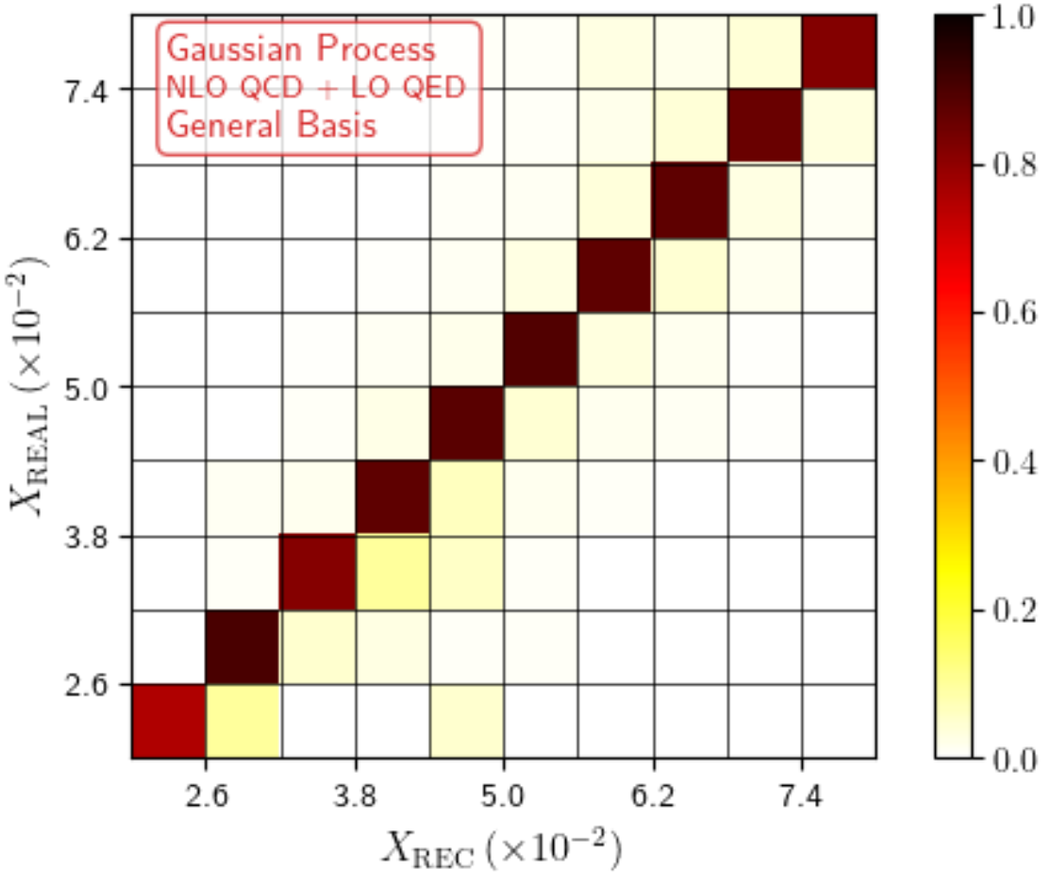
- We replace $x^{(i)}$ by $f_i(x) = e^{-\frac{\|x - x^{(i)}\|_2^2}{2l^2}}$

- This Radial Basis Function
  - requires less elements (one per feature in the basis).
  - effectively considers infinite terms.
  - is a popular form of the kernel method.



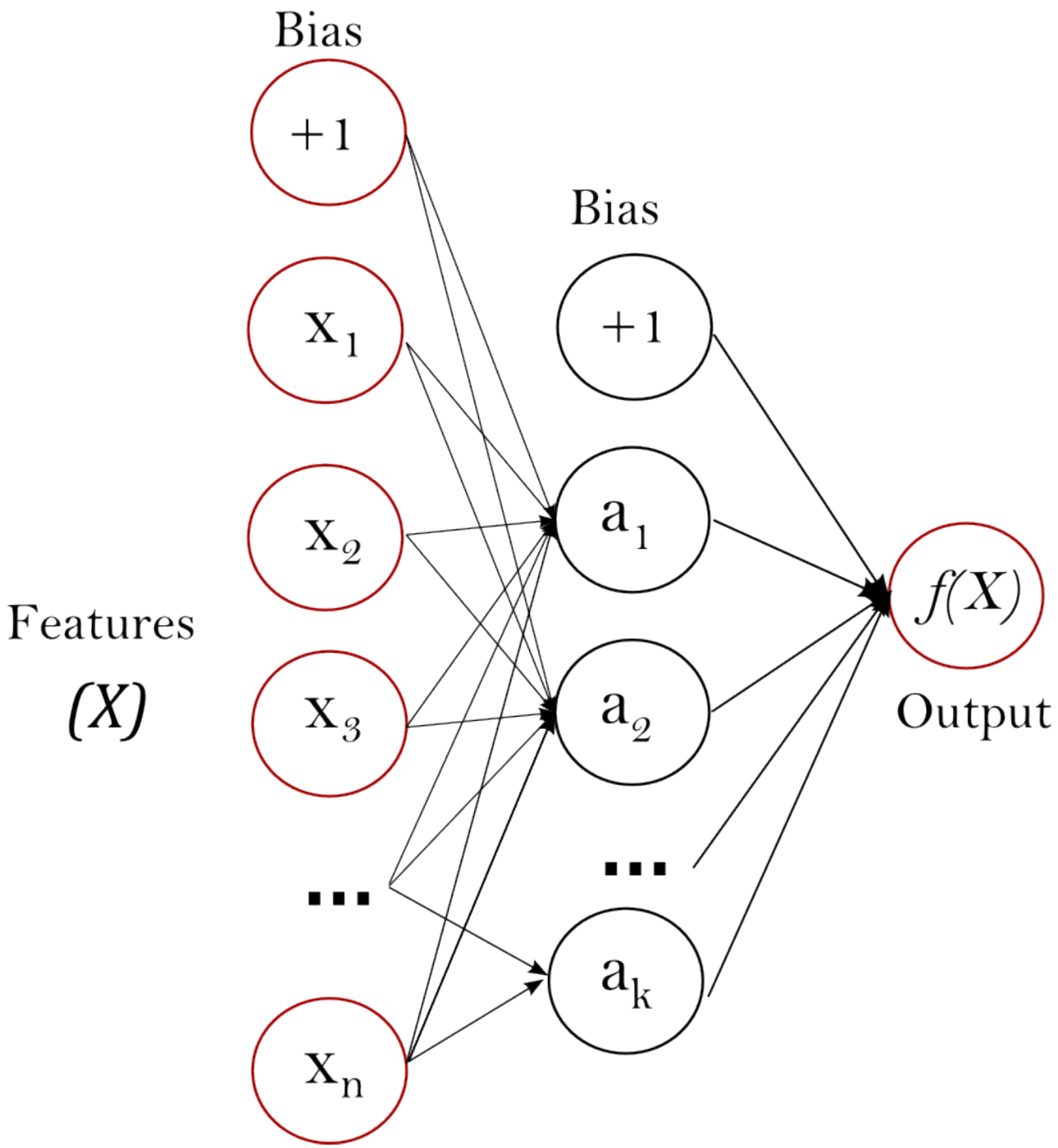


- Similar results for the reconstruction of *z*.



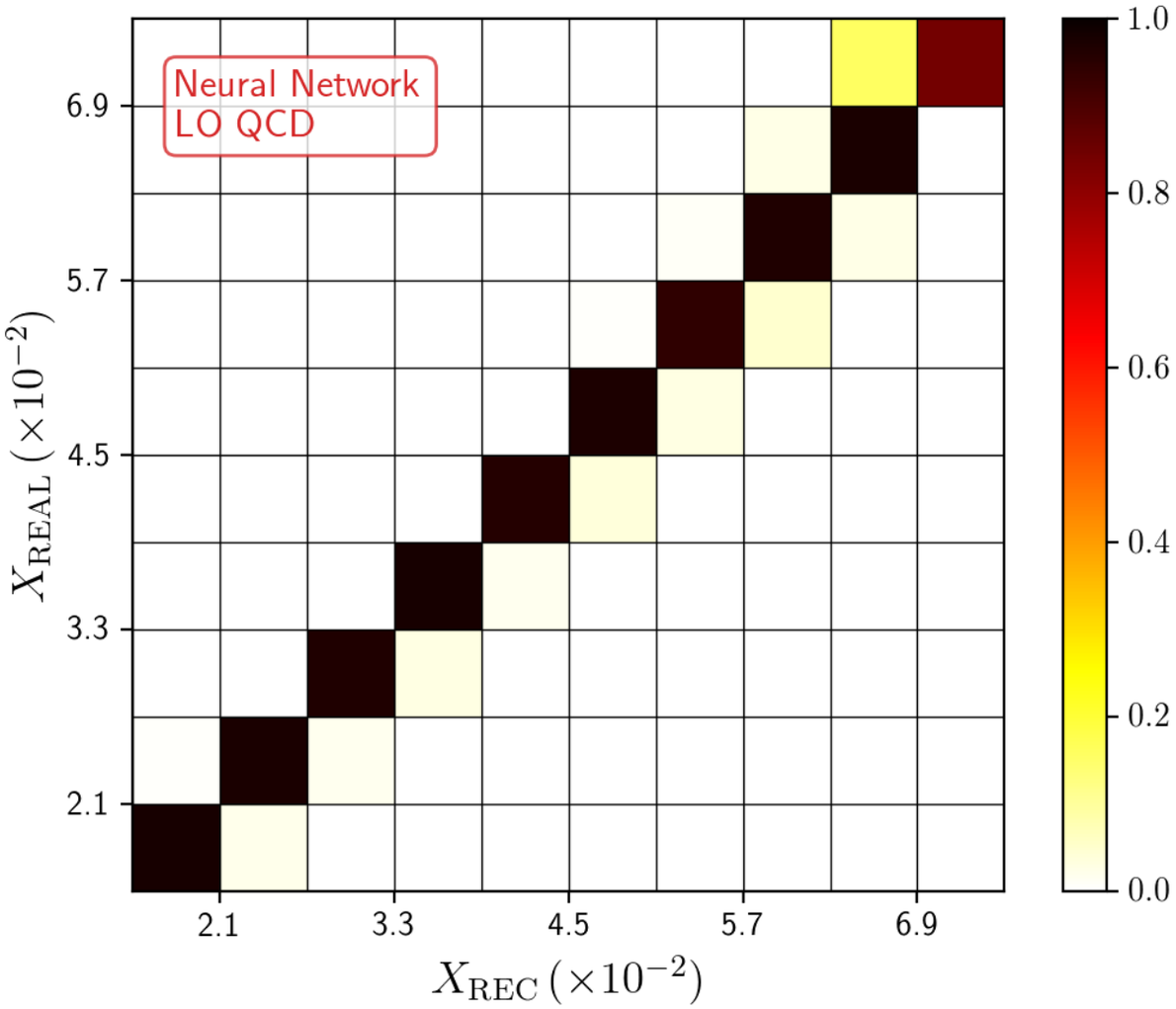
Neural networks: the basics

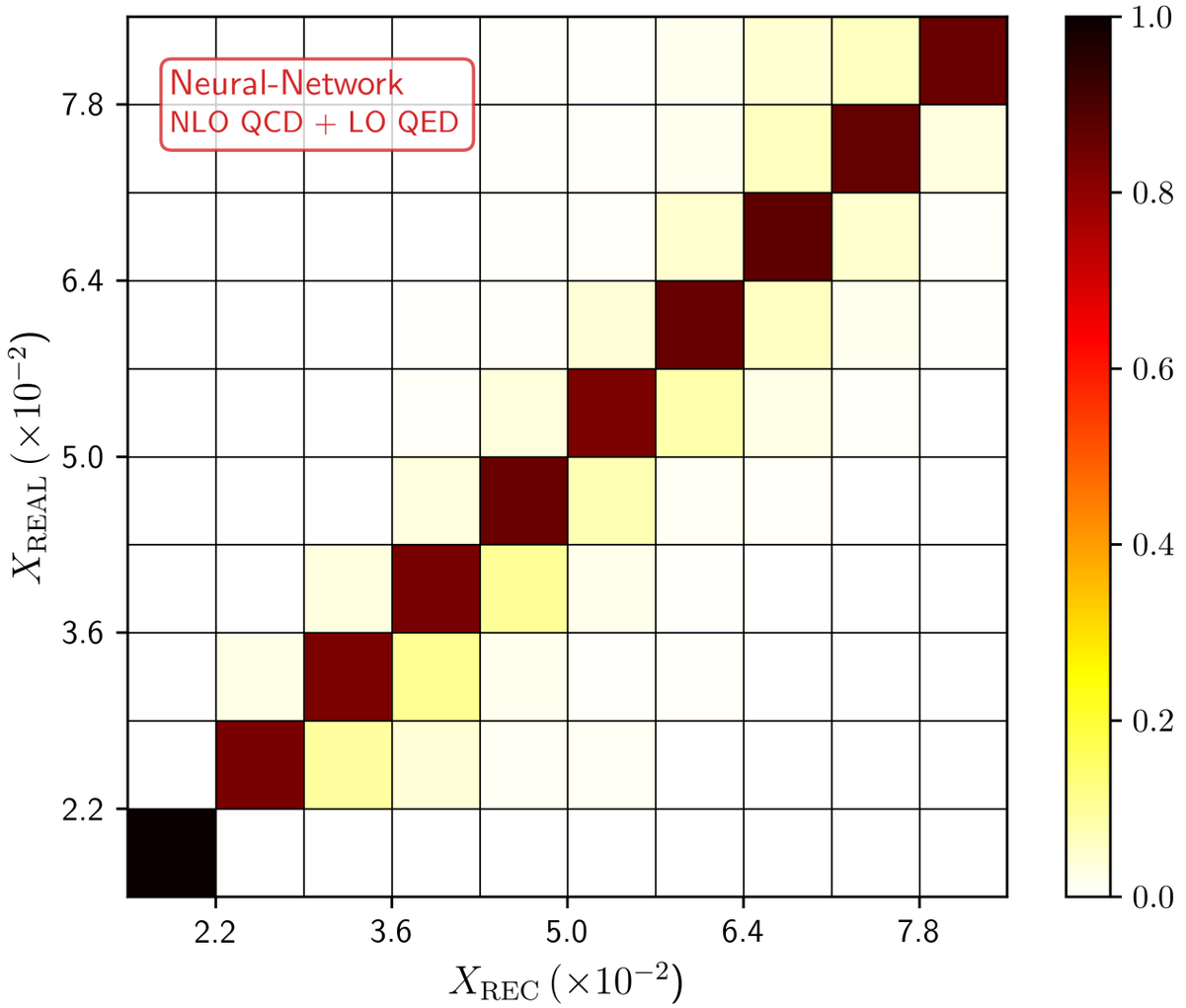
image from scikit-learn.org



- Pass from one layer to the next by applying a non-linear **activation function** to a **weighted** sum of the previous layers.
- Pros: no need to play with the basis, less human bias (but not zero!)
- Cons: the complexity of the architecture requires more time for training.
- Also, one needs to choose the architecture.

Neural networks





- For LO the complexity of the NN greatly surpasses the complexity of the problem.

Going greener with ML

- All the running of MC codes takes a long time to reach good precision.
- Apart from boring, they carry a significant environmental impact (and to our pockets given the cost of electricity).
- What can we do to make things faster using current available resources?
- Improve the codes: has to be done code by code.

Going greener with ML

- All the running of MC codes takes a long time to reach good precision.
- Apart from boring, they carry a significant environmental impact (and to our pockets given the cost of electricity).
- What can we do to make things faster using current available resources?
  - Improve the codes: has to be done code by code.
- Can we speed up the running time without touching the code?

Going greener with ML

- All the running of MC codes takes a long time to reach good precision.
- Apart from boring, they carry a significant environmental impact (and to our pockets given the cost of electricity).
- What can we do to make things faster using current available resources?
  - Improve the codes: has to be done code by code.
- Can we speed up the running time without touching the code?

Perhaps.

- Most codes require non perturbative inputs (e.g. PDFs).

- PDFs are provided as grids and functions that read the grids and interpolate over them (e.g. LHAPDF).

- And this is quite efficient, as long as we don’t need to run millions and millions of calculations.

- Most codes require non perturbative inputs (e.g. PDFs).

- PDFs are provided as grids and functions that read the grids and interpolate over them (e.g. LHAPDF).

- And this is quite efficient, as long as we don’t need to run millions and millions of calculations.

- A quick exploration shows that the time spent on the interpolation could be reduced 40-50% if we had analytical expressions for the PDFs.

- Most codes require non perturbative inputs (e.g. PDFs).

- PDFs are provided as grids and functions that read the grids and interpolate over them (e.g. LHAPDF).

- And this is quite efficient, as long as we don't need to run millions and millions of calculations.

- A quick exploration shows that the time spent on the interpolation could be reduced 40-50% if we had analytical expressions for the PDFs.

Goal: find an analytical x and Q^2 form for a set of proton PDFs.

- We are working (for now) with HERAPDF2.0

Idea and first results

- For most PDFs the x dependence at some initial scale is written as

$$f_i(x, Q_0^2) = N_i\, x^{\alpha_i}(1-x)^{\beta_i} P(x, c_{ij})$$

- We propose

$$f_i(x, Q_0^2, Q^2) = (N_i + g_{i,1}(Q^2, Q_0^2))x^{\alpha_i + g_{i,2}(Q^2, Q_0^2)}(1-x)^{\beta_i + g_{i,3}(Q^2, Q_0^2)} P(x, c_{ij} + g_{i,4}(Q^2, Q_0^2))$$

with $g_{i,j}(Q_0^2, Q_0^2) = 0$

In particular, for now, we're exploring $g_{i,4}(Q^2, Q_0^2) = 0$

- With that simplification, the ratio of the same flavour PDF at different scales is

$$R_i(x, Q_0^2, Q^2) \propto x^{g_{i,2}(Q^2,Q_0^2)}(1-x)^{g_{i,3}(Q^2,Q_0^2)}$$

- Taking logarithm

$$\ln(R_i) = \ln\left(\frac{N_i'(Q^2, Q_0^2)}{N_i}\right) + g_{i,2}(Q^2, Q_0^2)\ln(x) + g_{i,3}(Q^2, Q_0^2)\ln(1-x)$$

- With that simplification, the ratio of the same flavour PDF at different scales is

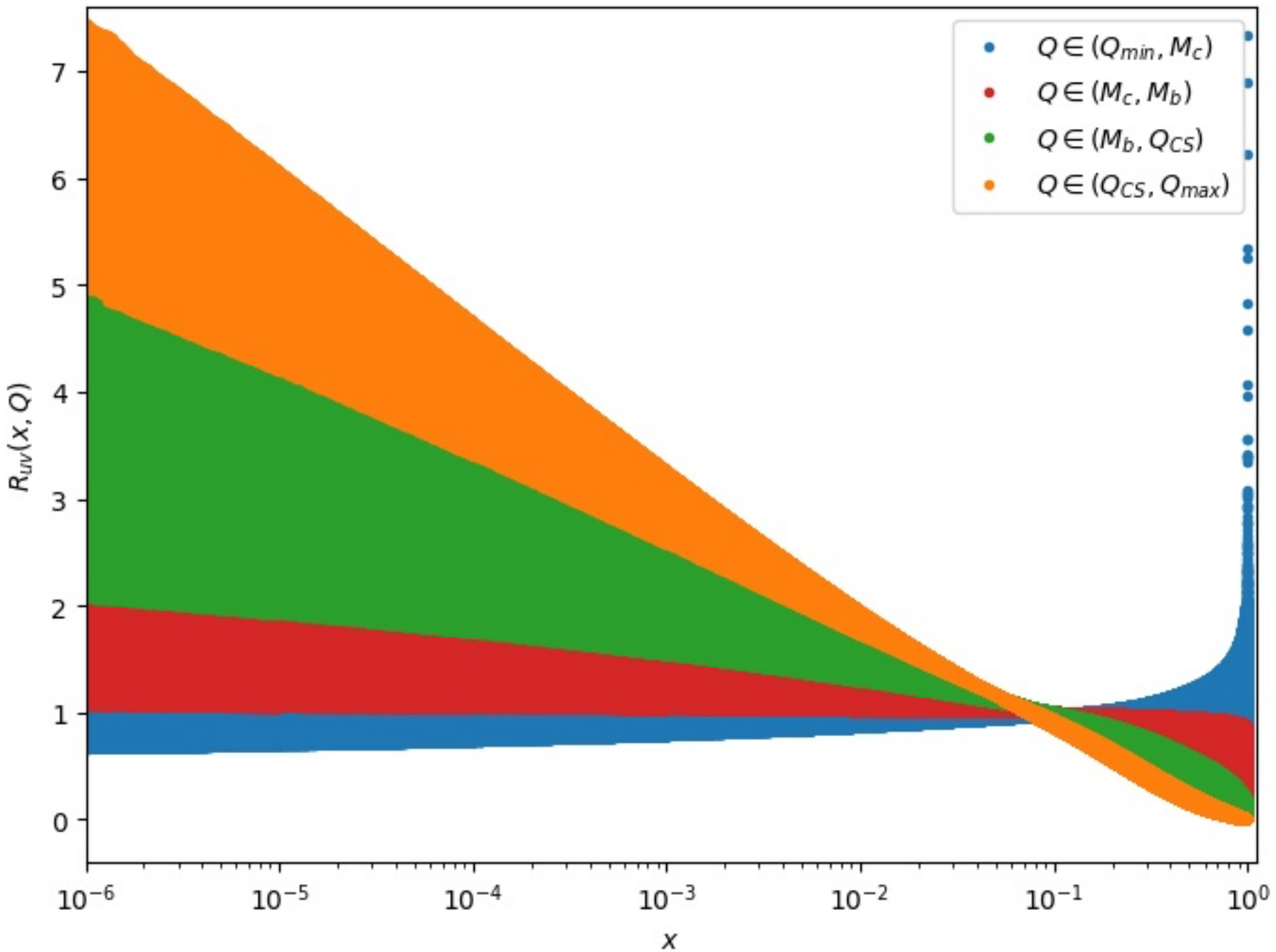
$$R_i(x, Q_0^2, Q^2) \propto x^{g_{i,2}(Q^2,Q_0^2)}(1-x)^{g_{i,3}(Q^2,Q_0^2)}$$

- Taking logarithm

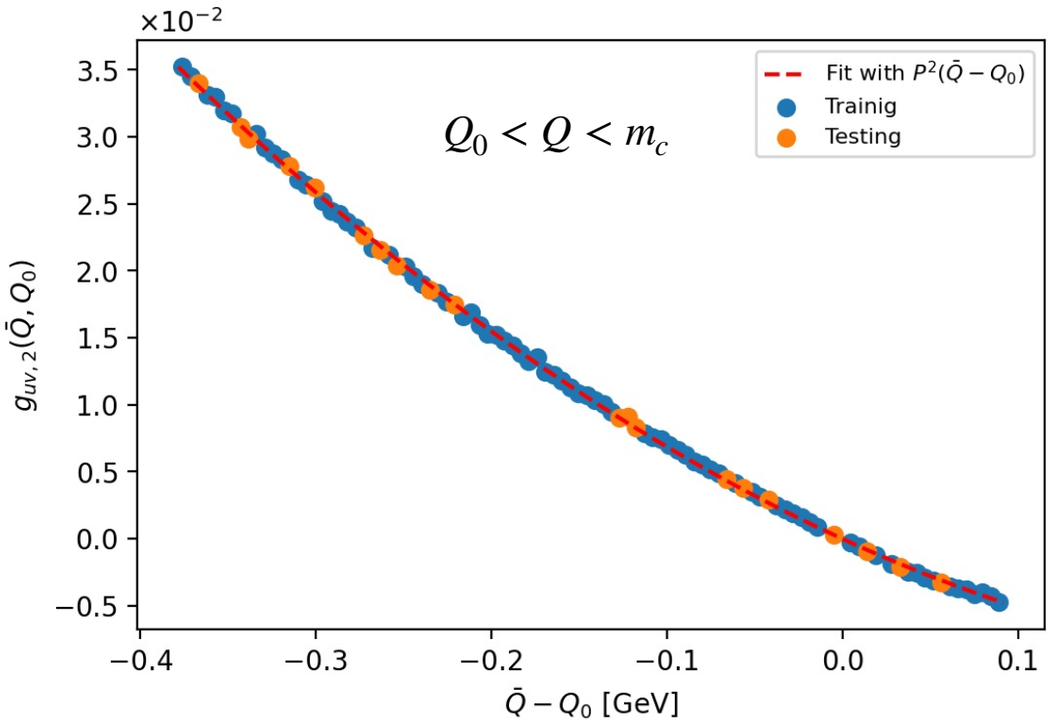
$$\ln(R_i) = \ln\left(\frac{N_i'(Q^2, Q_0^2)}{N_i}\right) + g_{i,2}(Q^2, Q_0^2)\ln(x) + g_{i,3}(Q^2, Q_0^2)\ln(1-x)$$

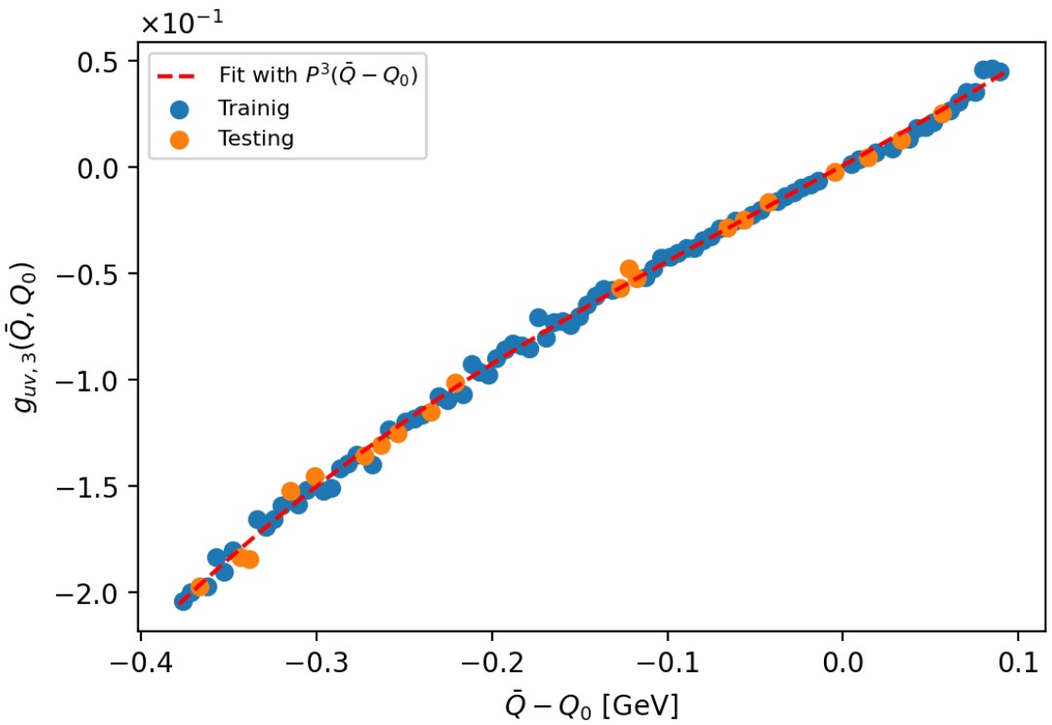
- Now we only have to find the missing functions.

First results: valence up

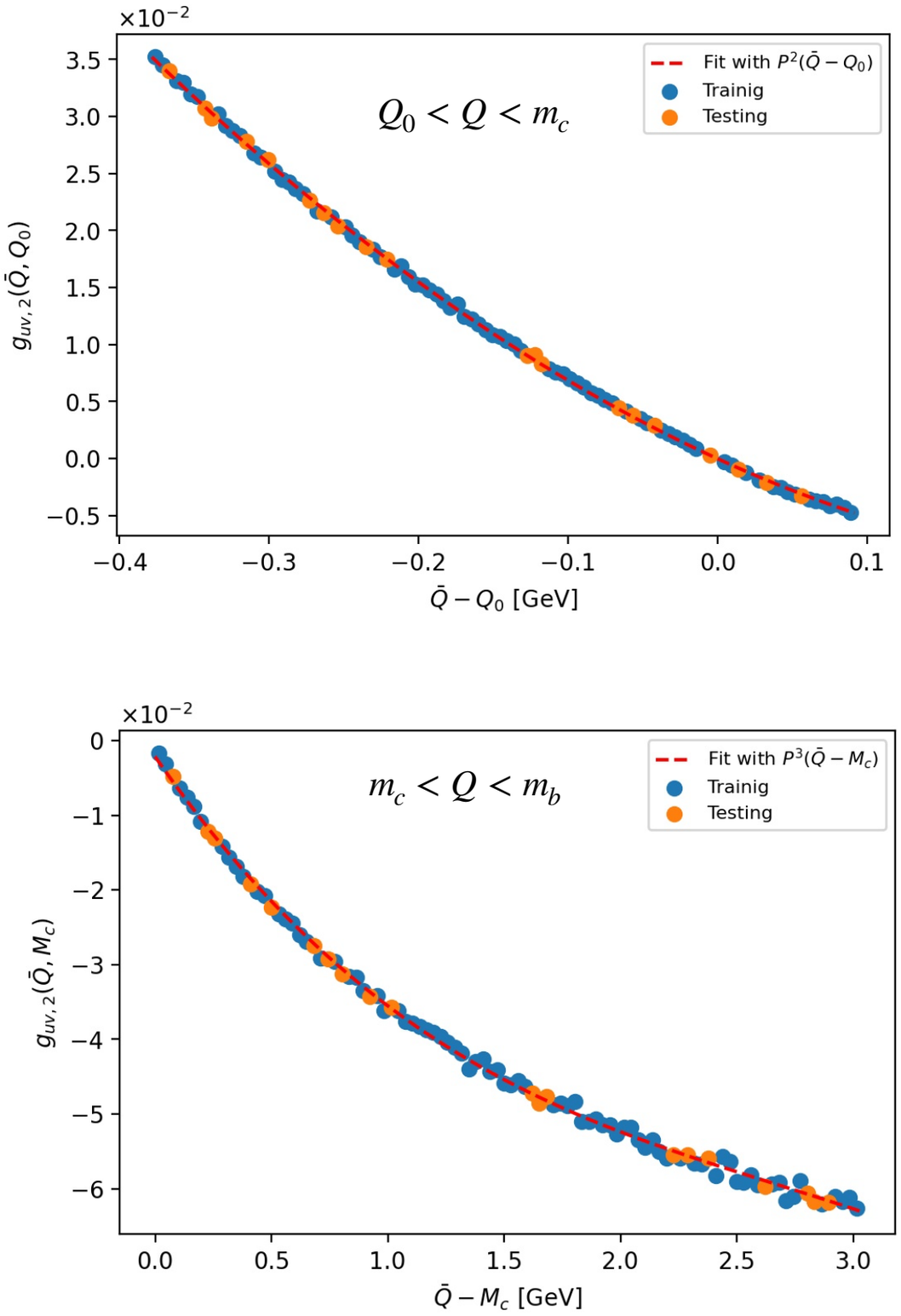


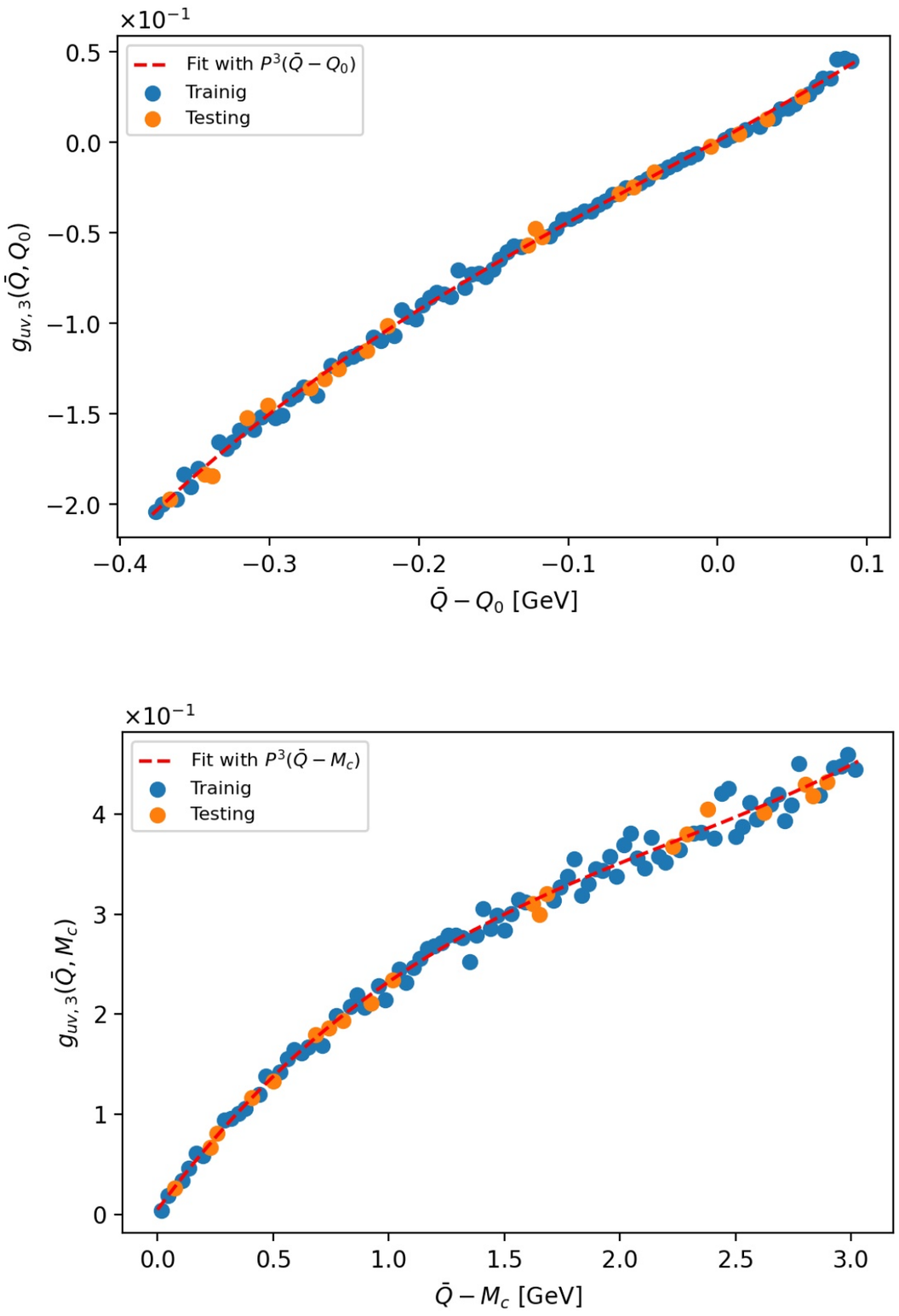
First results: valence up





First results: valence up





The GAPP Initiative

- This work is part of the GAPP Initiative.
- We aim to quantify the carbon footprint of HEP research and study ways of reducing it.
- If you are doing phenomenological studies and would like to contribute to the project, please send an email to **gapp-initiative@googlegroups.com** with:
  - **brief description of the simulation**
  - **time needed to run**
  - **hardware used**

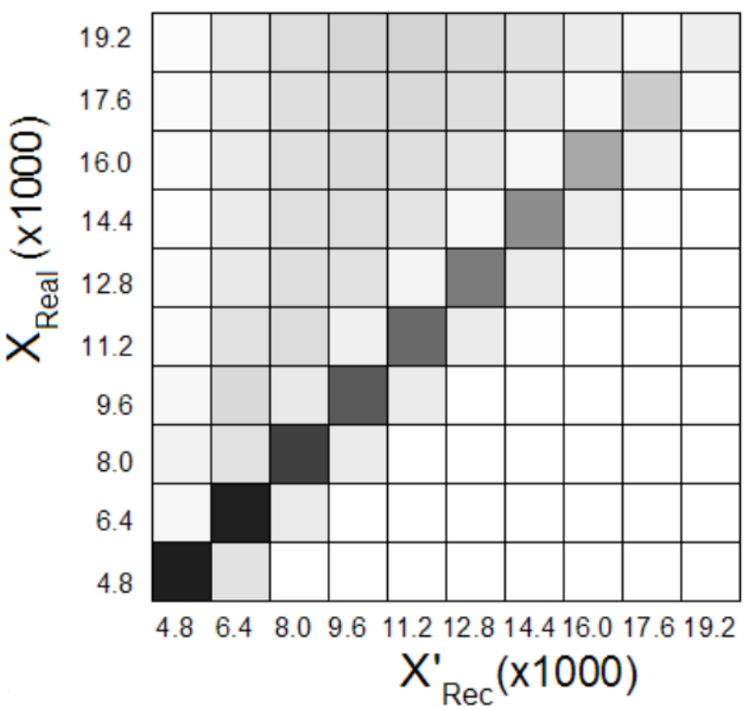
The GAPP Initiative

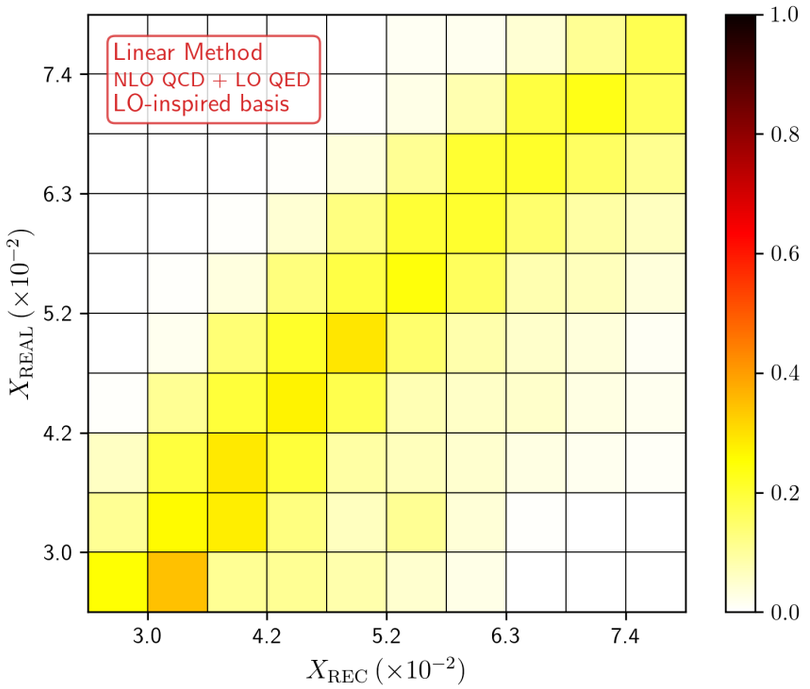
- This work is part of the GAPP Initiative.
- We aim to quantify the carbon footprint of HEP research and study ways of reducing it.
- If you are doing phenomenological studies and would like to contribute to the project, please send an email to **gapp-initiative@googlegroups.com** with:
  - **brief description of the simulation**
  - **time needed to run**
  - **hardware used**

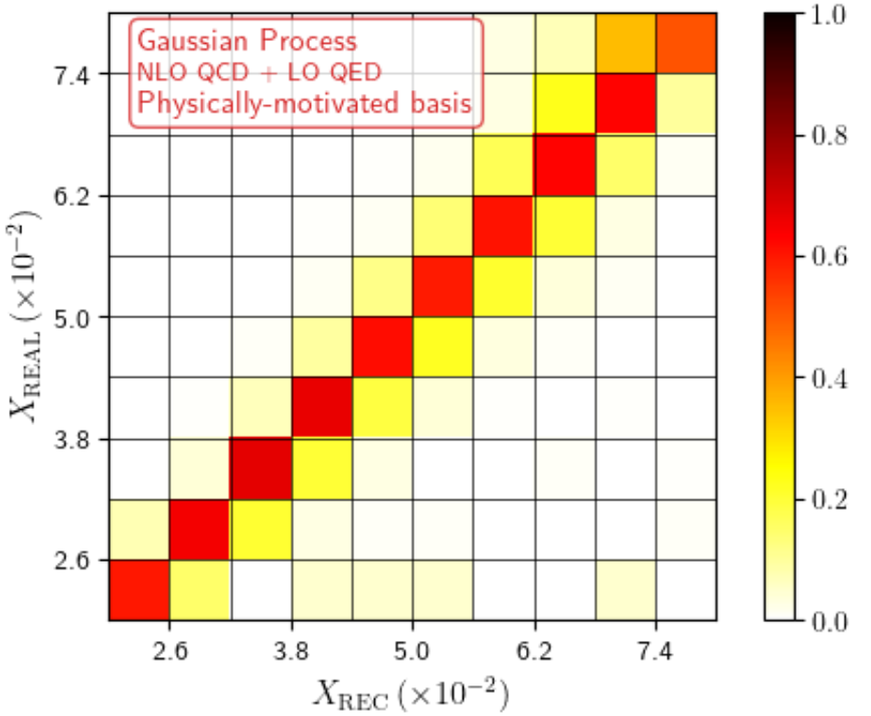
Join the GAPP!

Summary

- We have explored the applicability of ML techniques to better understand the underlying kinematics of a p+p collision.
- The methodology can be used for any process by non-experts.
- The methods applied can result in a better reconstruction than in the original work, but physical intuition can play a relevant role.

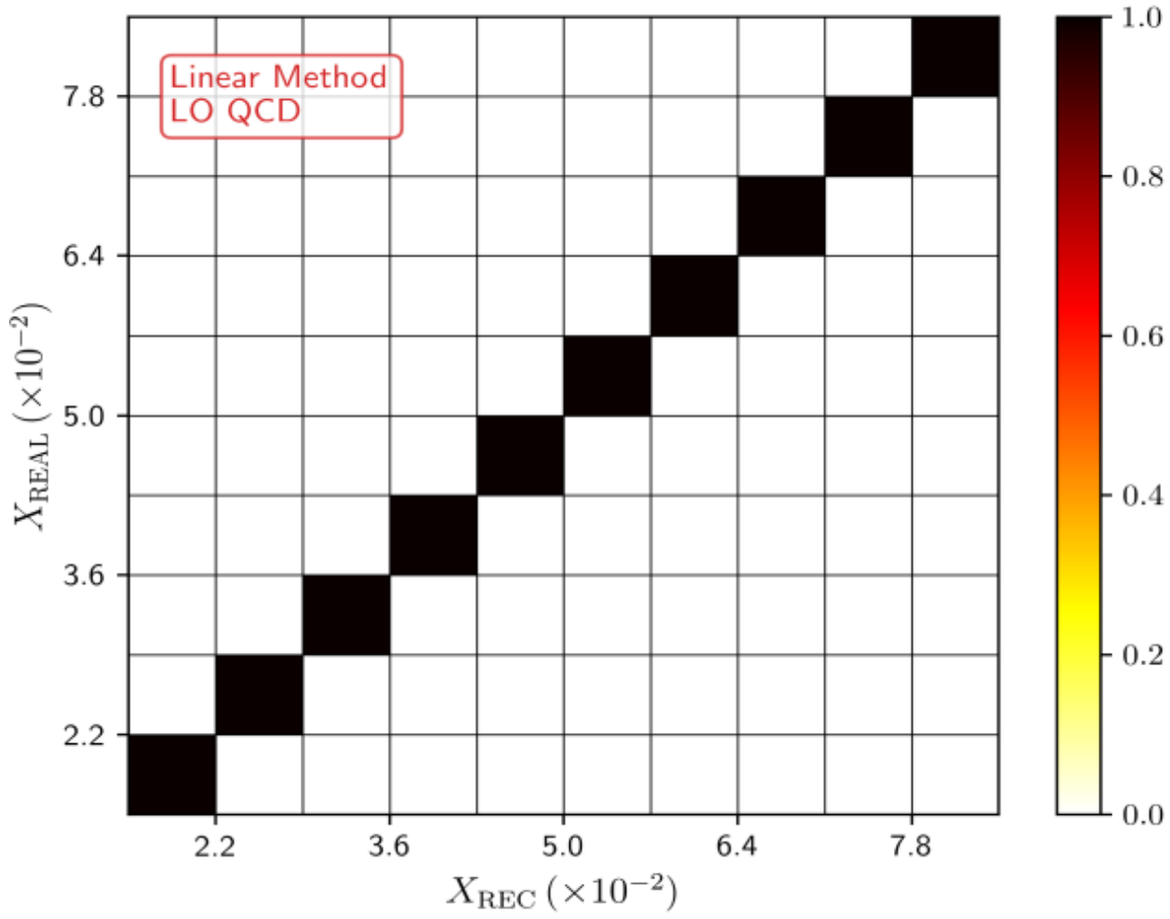


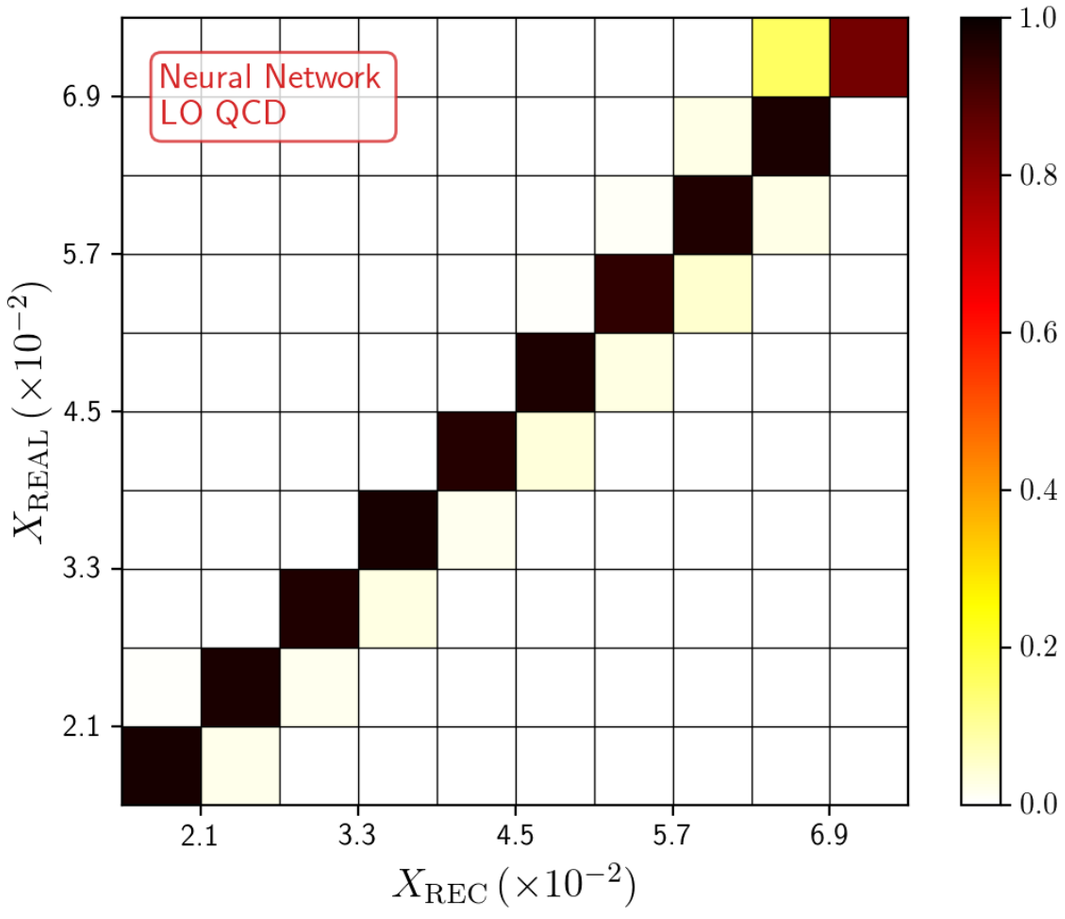




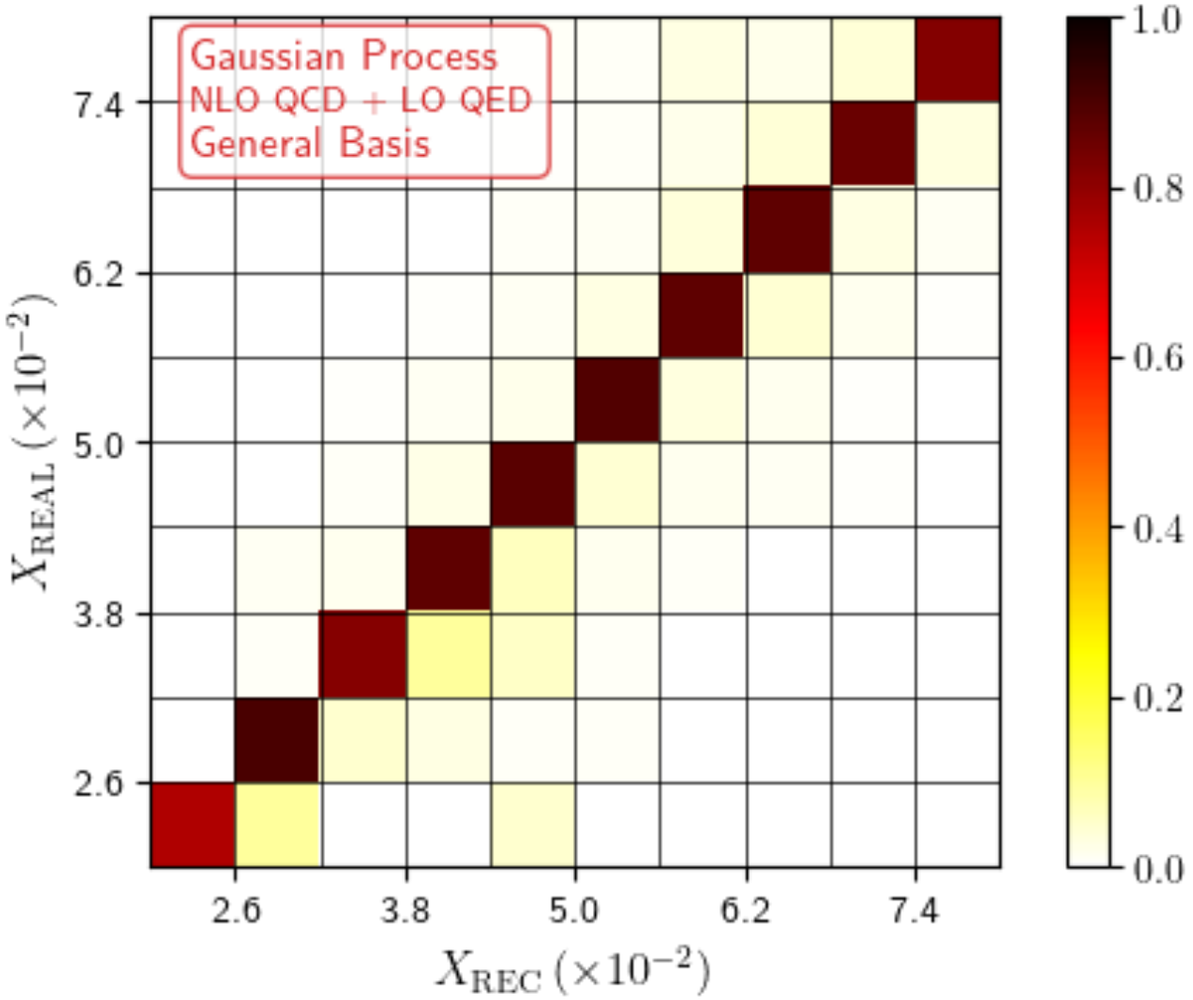
- Higher sophistication of the method does not always translate into better results. E.g.:

exact relation known

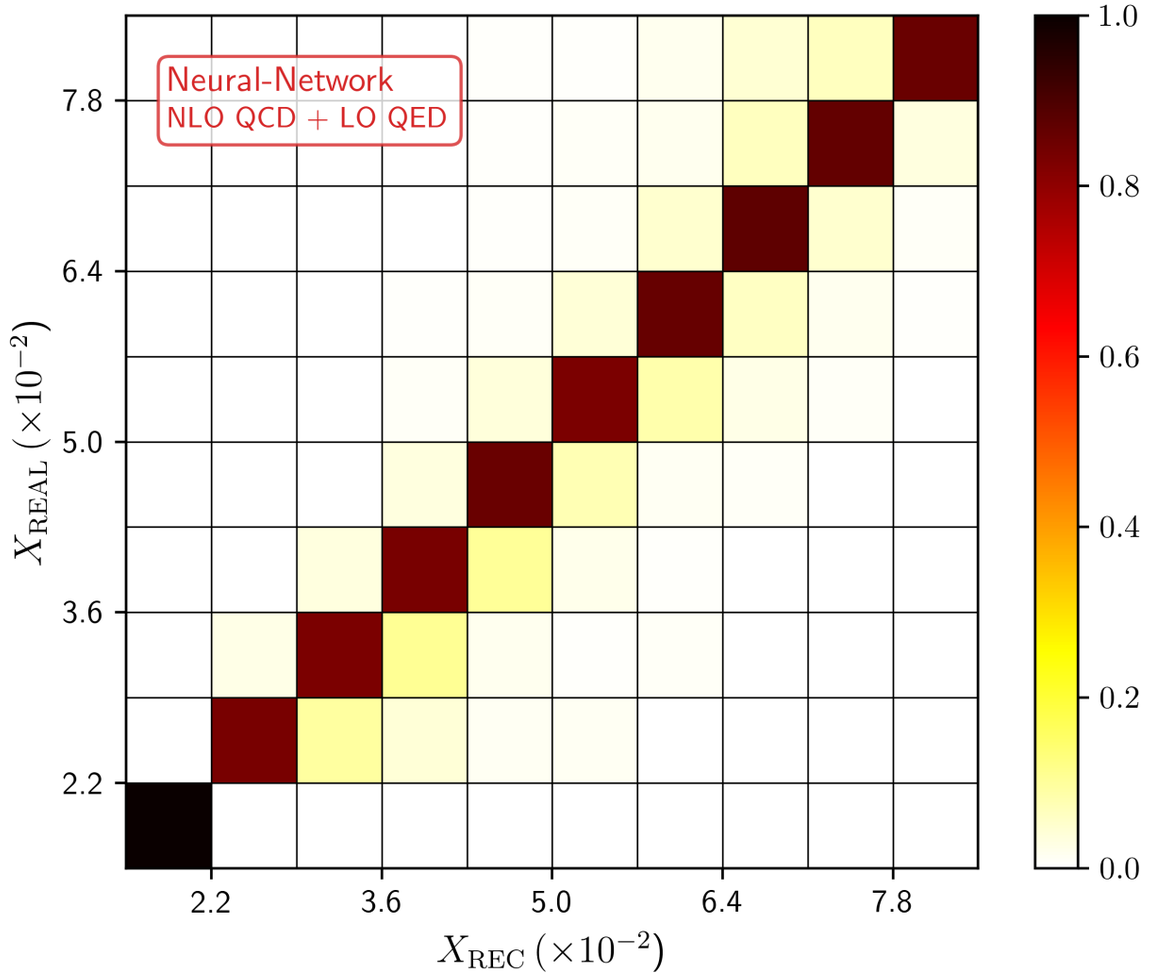




- Higher sophistication of the method does not always translate into better results. E.g.:

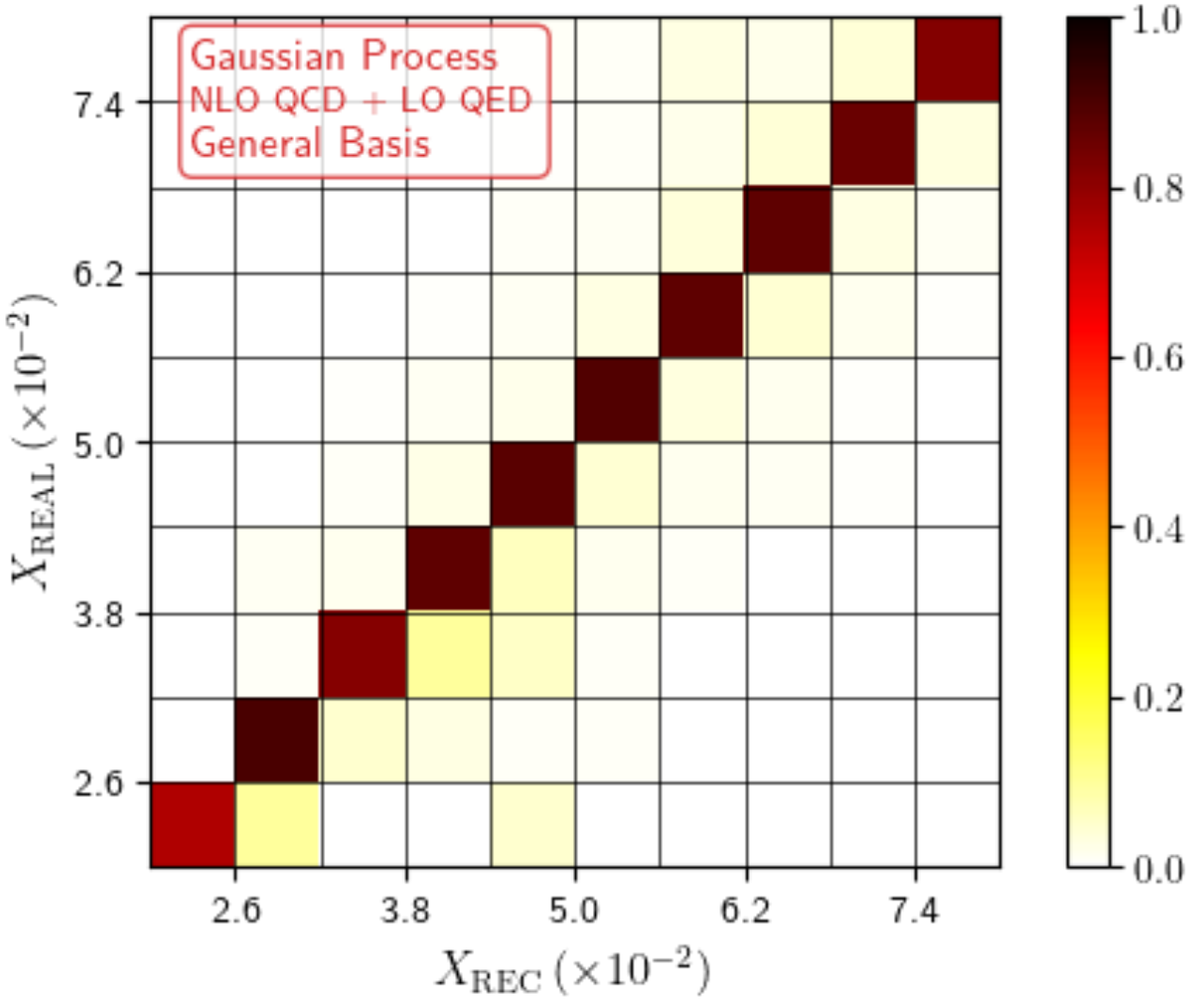


few minutes to train

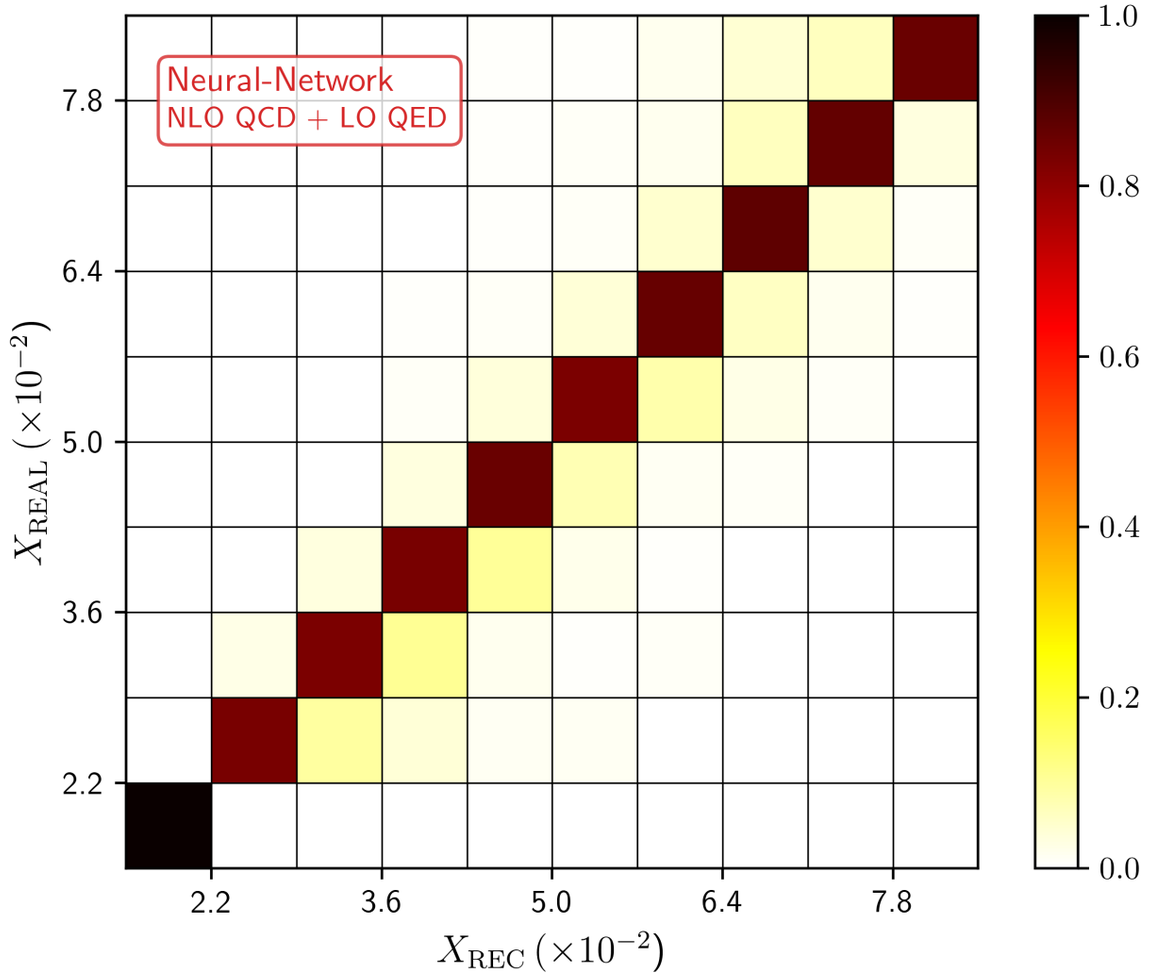


several hours to train
exploration of “good” architecture

- Higher sophistication of the method does not always translate into better results. E.g.:



few minutes to train



several hours to train
exploration of "good" architecture

- Promising steps in speeding up the calculation of codes using PDFs.

Thank you for your attention!