UNIVERSITÀ DI TORINO

INFN
Istituto Nazionale di Fisica Nucleare

Andrea Signori

University of Turin and INFN

Closure tests of TMDs

MAP meeting, Pavia

April 28th, 2026

- Kamil Laurent
  (now PhD student at Nikhef)

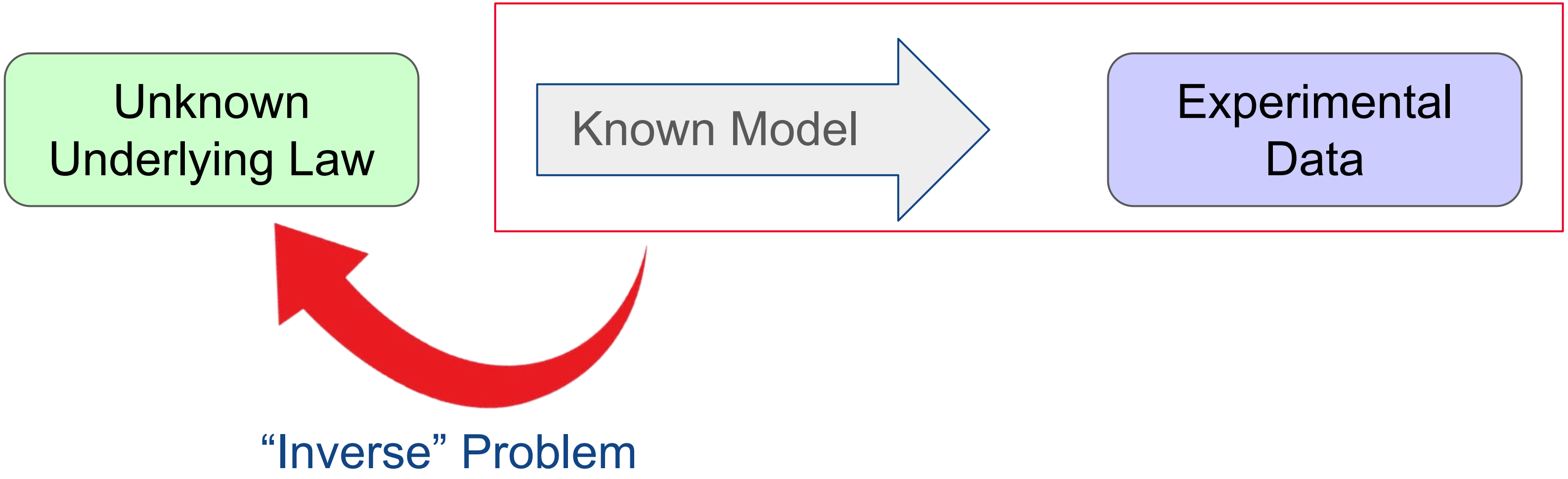
- Mariaelena Boglione

- Emanuele R. Nocera

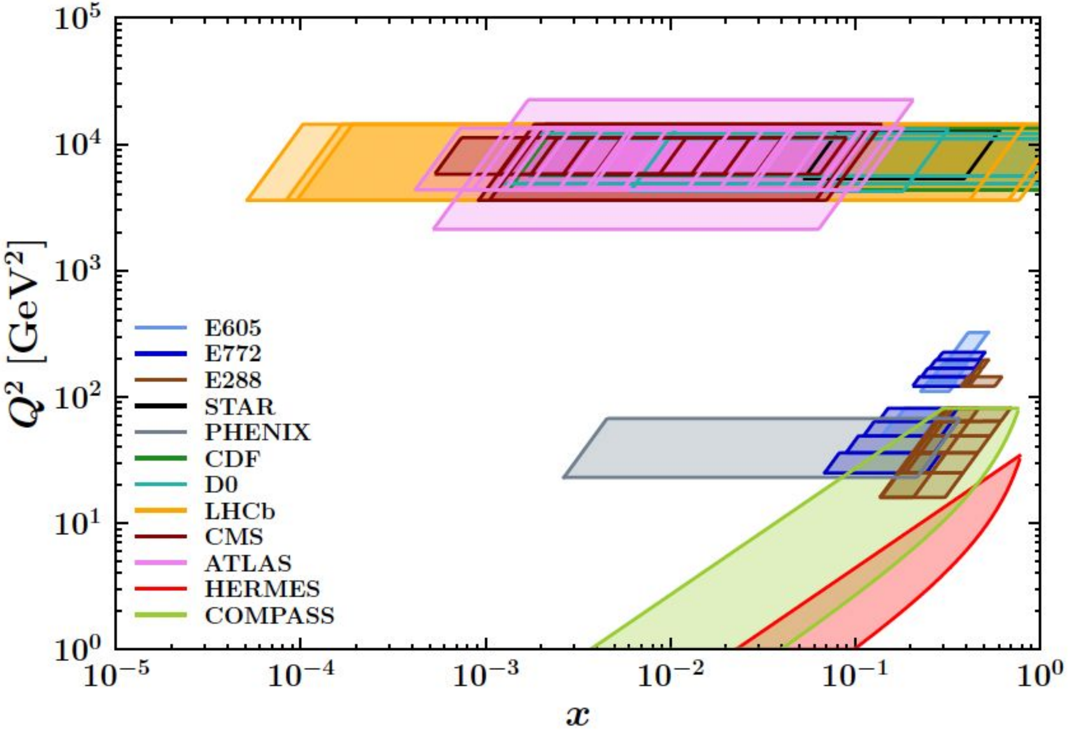
- Andrea Signori

The Inverse Problem

The statistical analysis we performed applies to any framework dealing with inverse problems.



TMDs from data: an inverse problem

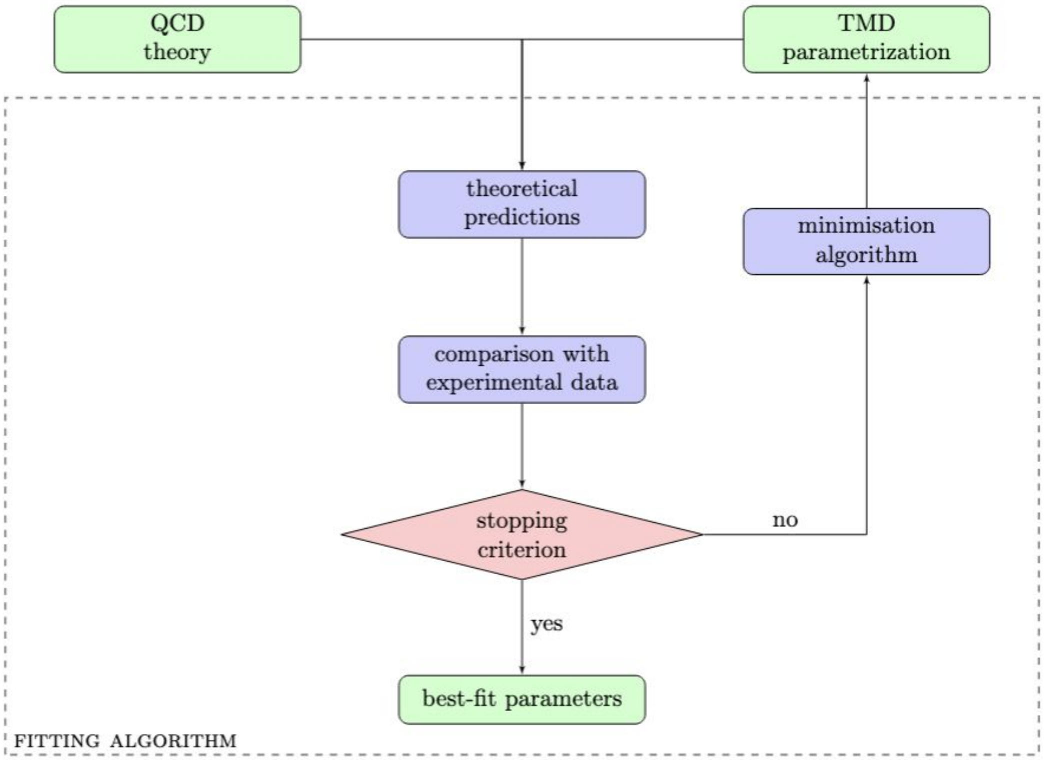


[Figure from Bacchetta et al. JHEP 10 (2022) 127]

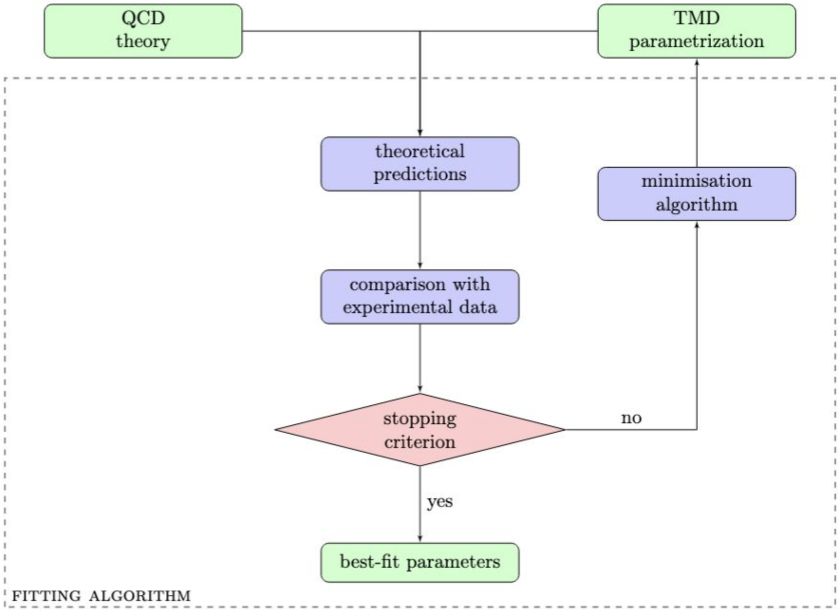
Fitting Frameworks

We analysed these frameworks:

- PV19
- MAP22
- MAPNN



Unanswered Questions



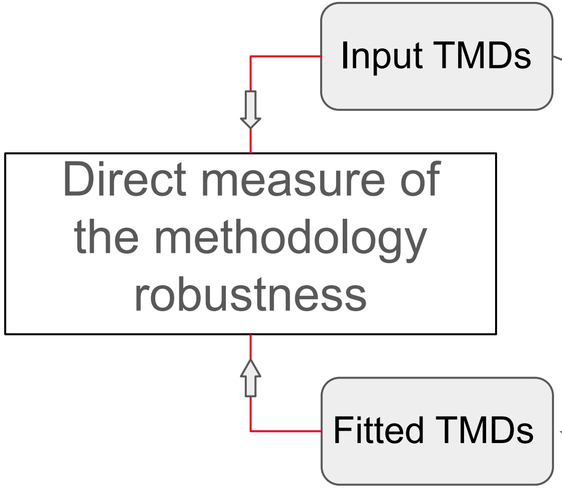
How close are the fitted TMDs to the real law?

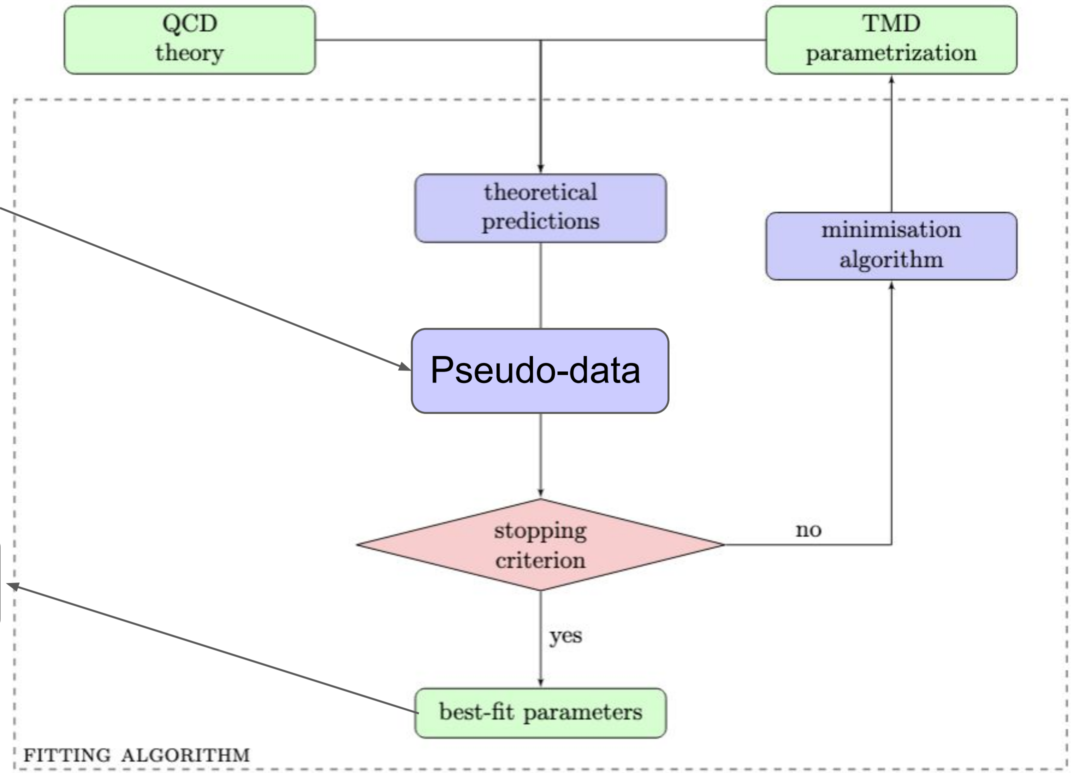
How much of the declared TMD uncertainties derive from the framework?

Are the declared TMD uncertainties statistically faithful?

Closure Testing

Closure Testing

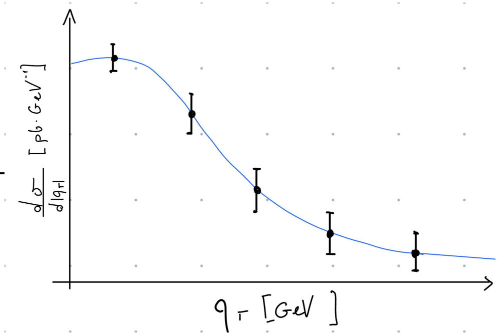




[Del Debbio et al. Eur.Phys.J.C 82 (2022) 4, 330]

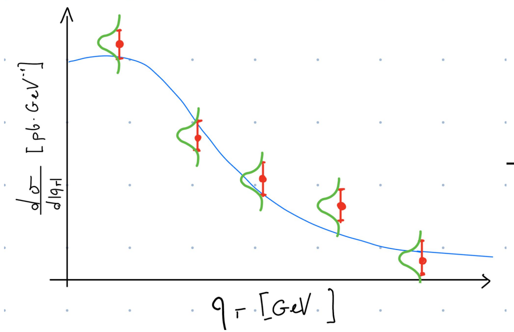
Levels of Fluctuation

Level 0 (no fluctuations)

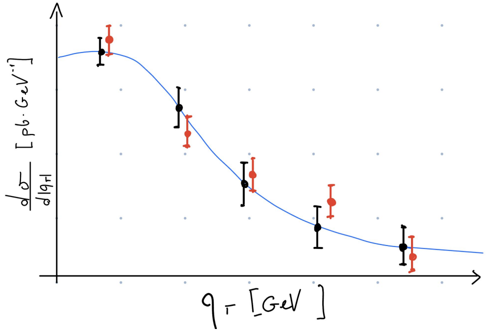


Pseudo Data

Gaussian fluctuation



N_{rep} Monte Carlo replicas

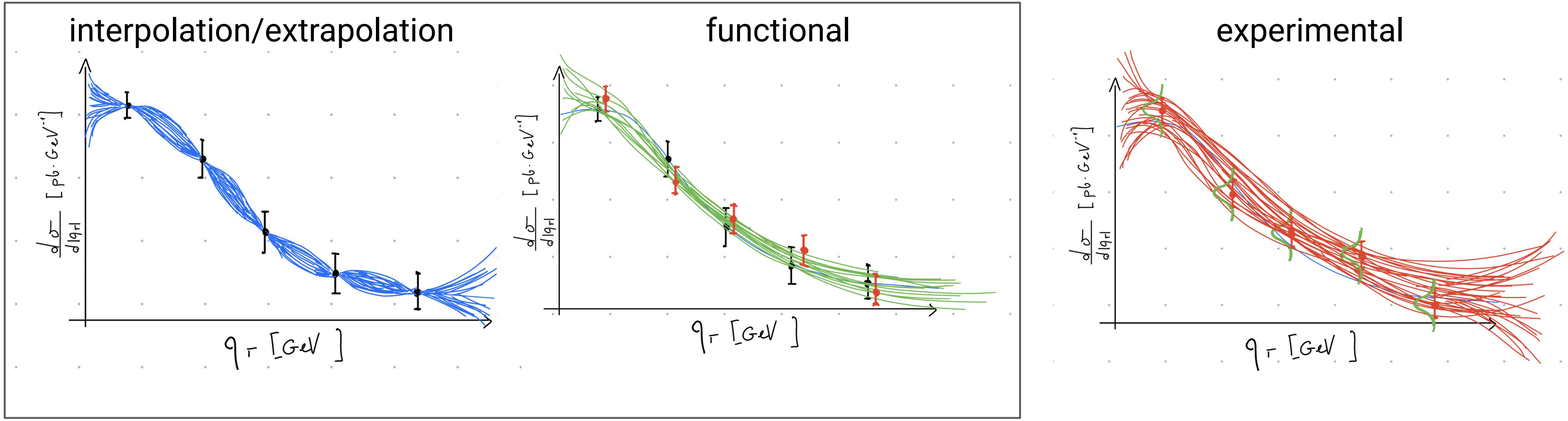


Level 2 (L2)
(multi closure test)

Level 1 (L1)

Levels of Uncertainty

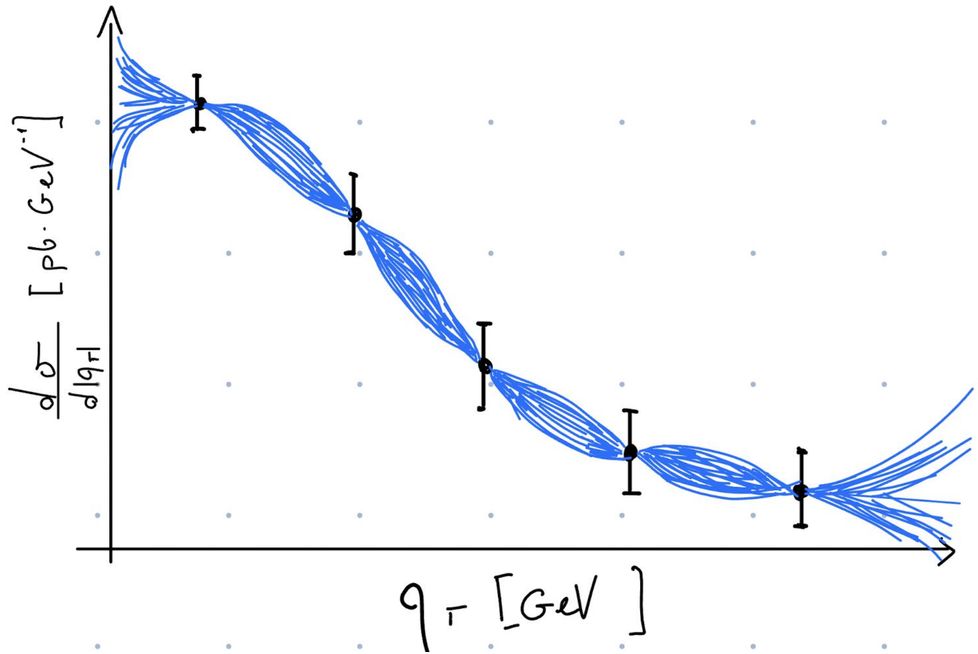
Each level of fluctuation introduces a different level of uncertainty. We can fit L0, L1 and L2 data to characterize the uncertainty components.



framework-dependent uncertainties

[Ball et al. JHEP 04 (2015) 040]

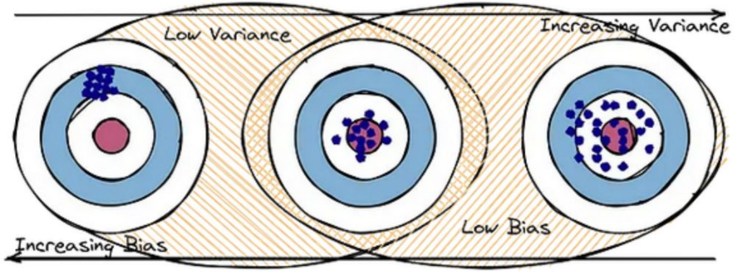
Level 0 Test



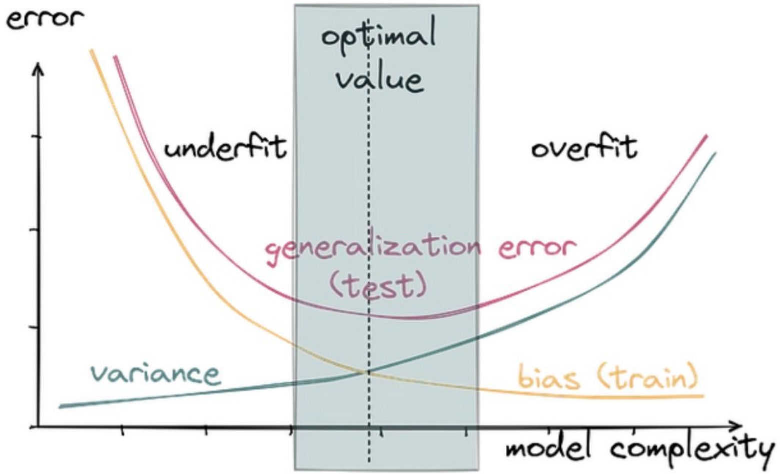
From a test on L0 data we can assess two aspects:

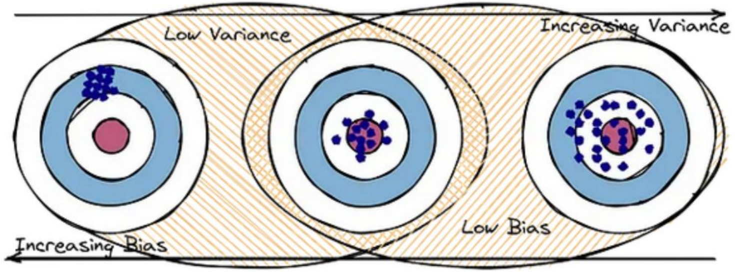
- **Frameworks flexibility**: is the framework able to reproduce the solution with $\chi^2 = 0$?
- **Interpolation/extrapolation uncertainty**, given by the finiteness of the dataset

Multi Closure Tests



Multi Closure Tests





A large number of fits (e.g. 50 fit with 100 replicas) to determine:

- Bias-Variance Ratio:

$$\mathcal{R}_{bv} = \sqrt{\frac{\mathbb{E}[bias]}{\mathbb{E}[variance]}} \approx 1$$

- TMD uncertainty faithfulness:

$$\xi_{1\sigma} \approx 0.683$$

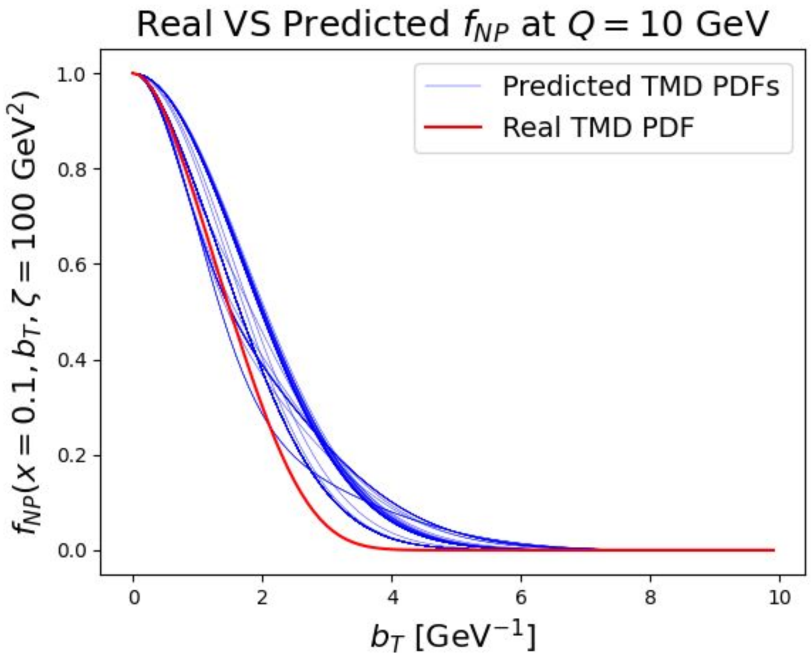
Results

Are the Methodologies Flexible?

Test	Mean χ^2	Best MSD	Closure
$\mathrm{MAP22oMAP22}_{L0}$	$\mathcal{O}(10^{-14})$	/	yes
$\mathrm{PV19oPV19}_{L0}$	$\mathcal{O}(10^{-13})$	/	yes
$\mathrm{MAP22oPV19}_{L0}$	0.152	< 0.0015	no
$\mathrm{PV19oMAP22}_{L0}$	0.017	< 0.00015	no
$\mathrm{MAP22oMIX24}_{L0}$	0.155	< 0.004	no
$\mathrm{PV19oMIX24}_{L0}$	$\mathcal{O}(10^{-7})$	/	no

- same input and fitting parametrization
- different input and fitting parameterizations

Are the Methodologies Flexible?



input: PV19, *fit*: MAPTMD22

This picture outlines two difficulties (MAP22oPV19):

- **Generalization**:
  the frameworks cannot generalize well
- **Minimizer**:
  the minimizer cannot find the best solution in 100% of the cases

Bias-Variance Tradeoff

PV19oPV19 (L2)

- Quantile and bias-variance ratio:

$$\mathcal{R}_{bv} = 1.577 \pm 0.068$$
$$\xi_{1\sigma} = 0.486 \pm 0.012$$

- We observe signs of **underfitting**
- The TMD uncertainties are **underestimated**

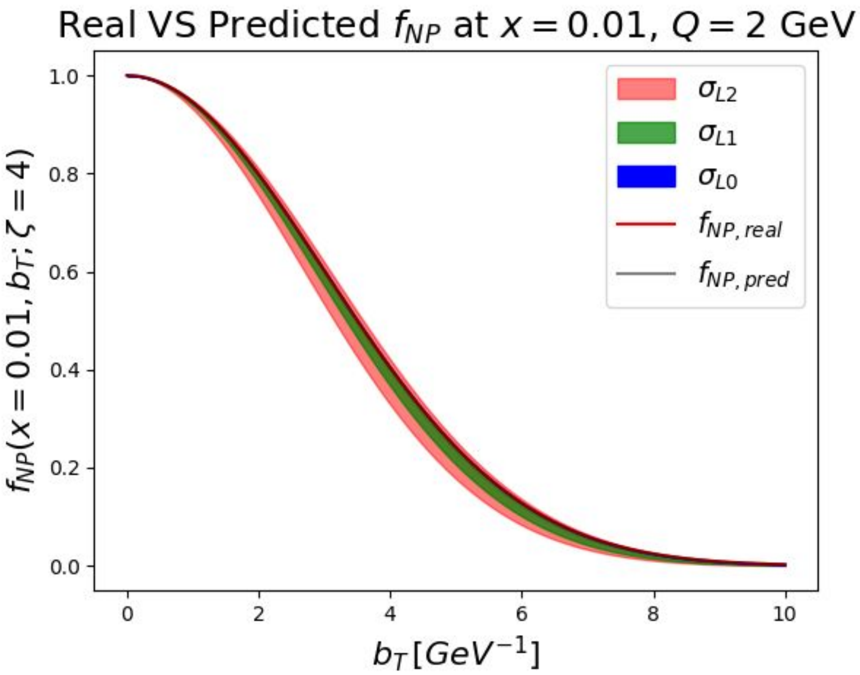
MAPTMD22oMAPTMD22 (L2)

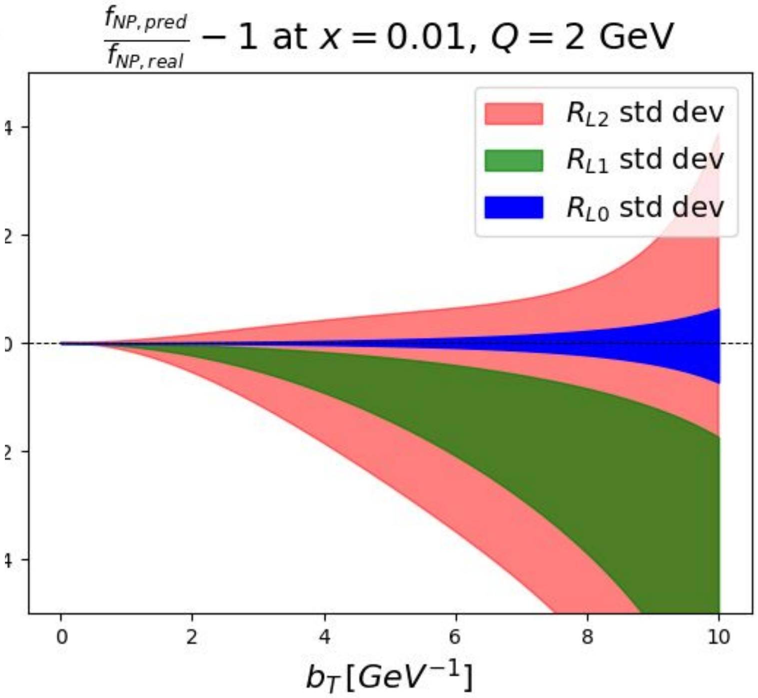
- Quantile and bias-variance ratio:

$$\mathcal{R}_{bv} = 0.978 \pm 0.045$$
$$\xi_{1\sigma} = 0.688 \pm 0.010$$

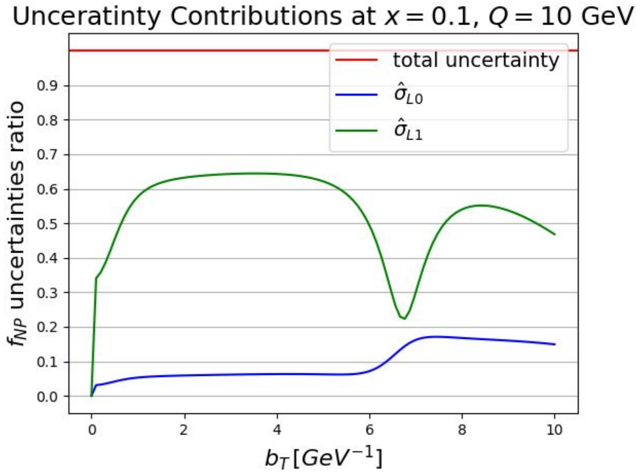
- The model is **optimized**
- The uncertainty estimates are **statistically faithful**

MAPTMD22 Uncertainties

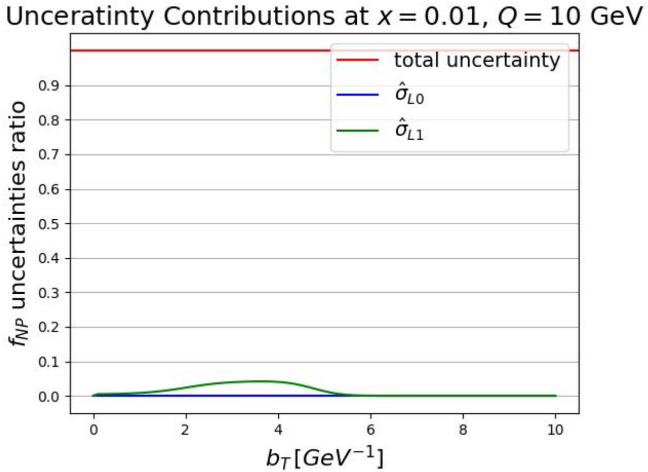




Uncertainty Characterization (L0-L1-L2)



MAPTMD22 uncertainty contributions



PV19 uncertainty contributions

More data

More precise data

Better predictivity

The NN case

L2

Test Name	Input	Fit Parameterization	$\mathcal{R}_{bv} \pm \sigma_{\mathcal{R}_{bv}}$	$\xi_{1\sigma} \pm \sigma_{\xi_{1\sigma}}$
MAPNNoMAPNN	$u^{\text{real}}_{\text{MAPNN}}$	MAPNN	0.970 ± 0.097	0.714 ± 0.043
MAPNNoMAP22	$u^{\text{real}}_{\text{MAP22}}$	MAPNN	1.032 ± 0.077	0.699 ± 0.050
MAPNNoPV19	$u^{\text{real}}_{\text{PV19}}$	MAPNN	0.510 ± 0.195	0.675 ± 0.051

X

Table 4.3. Results of three multi-closure tests in which we use the MAPNN framework to fit ensembles of L2 pseudo-data generated with different parameterizations.

The NN case

L2

Test Name	Input	Fit Parameterization	$\mathcal{R}_{bv} \pm \sigma_{\mathcal{R}_{bv}}$	$\xi_{1\sigma} \pm \sigma_{\xi_{1\sigma}}$
MAPNNoMAPNN	$u^{\text{real}}_{\text{MAPNN}}$	MAPNN	0.970 ± 0.097	0.714 ± 0.043
MAPNNoMAP22	$u^{\text{real}}_{\text{MAP22}}$	MAPNN	1.032 ± 0.077	0.699 ± 0.050
MAPNNoPV19	$u^{\text{real}}_{\text{PV19}}$	MAPNN	0.510 ± 0.195	0.675 ± 0.051

X

Table 4.3. Results of three multi-closure tests in which we use the MAPNN framework to fit ensembles of L2 pseudo-data generated with different parameterizations.

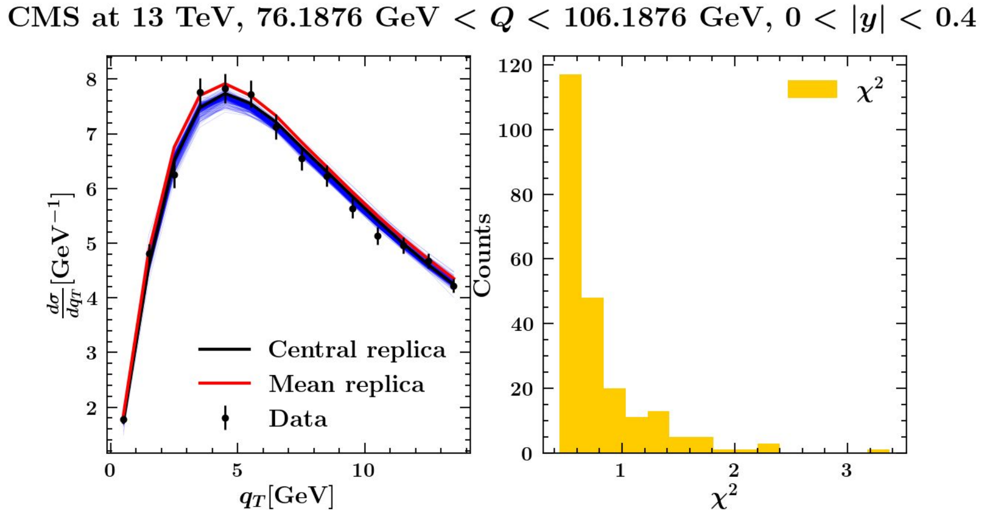
contradiction?

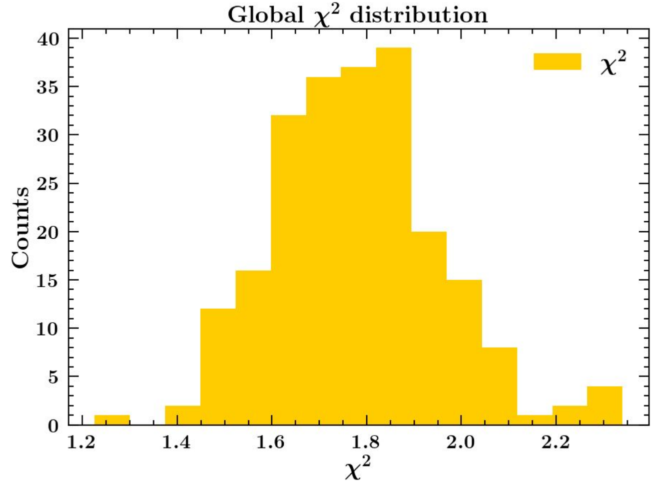
L0

Test Name	Input	Fit Parameterization	$\langle \chi^2 \rangle \pm \sigma_{\chi^2}$	Closure
MAPNNoMAPNN_{L0}	$u^{\text{real}}_{\text{MAPNN}}$	MAPNN	0.0648 ± 0.043	no
MAPNNoMAP22_{L0}	$u^{\text{real}}_{\text{MAP22}}$	MAPNN	0.0199 ± 0.0452	no
MAPNNoPV19_{L0}	$u^{\text{real}}_{\text{PV19}}$	MAPNN	0.0275 ± 0.0251	no

Table 4.4. Results of three Level 0 tests in which we use the MAPNN framework to fit pseudo-data generated with different parameterizations.

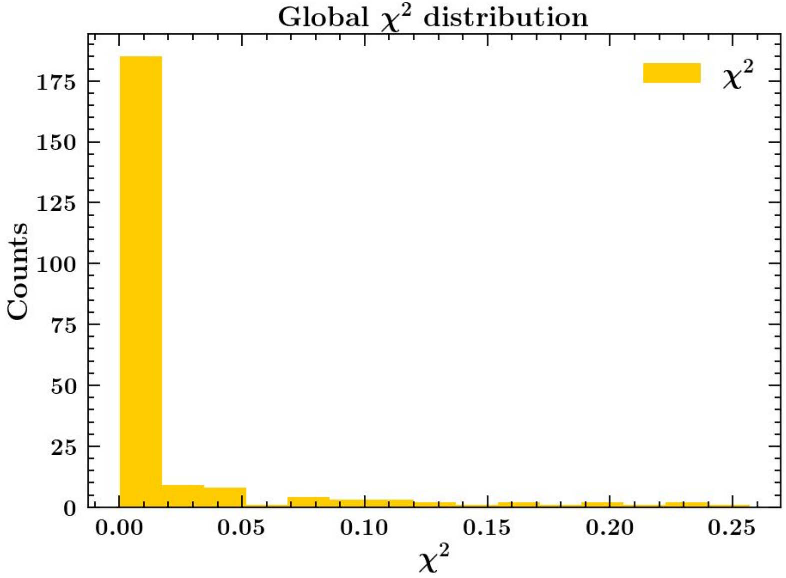
MAPNNoMAP22: Level 2



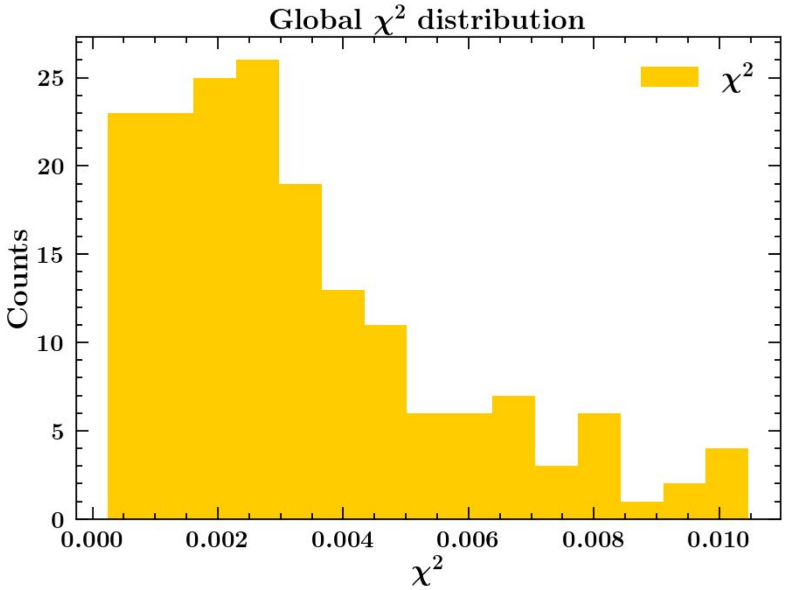


Everything looks fine at the observable level, what happens **at TMD level**?

MAPNNoMAP22: Level 0

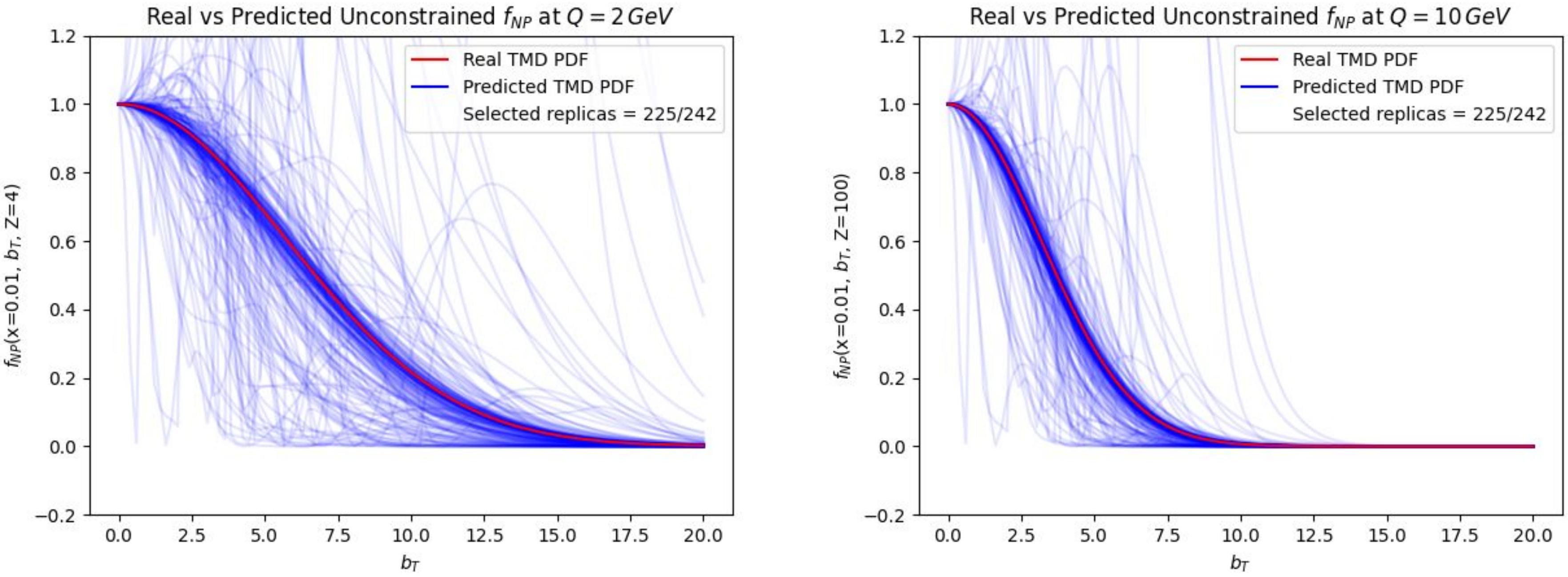


225 L0 replicas



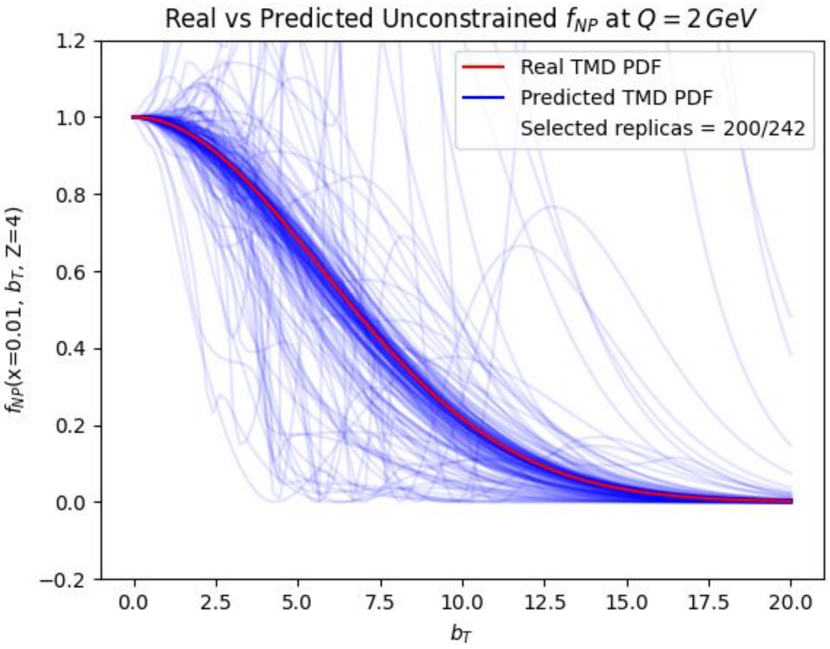
175 L0 replicas

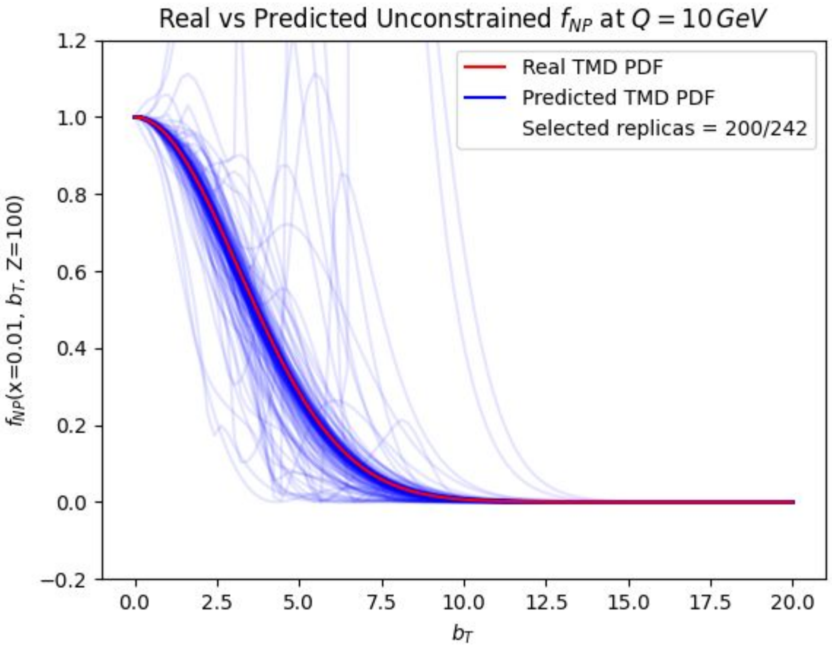
MAPNNoMAP22: Level 0



input: MAP22, *fit*: MAPNN

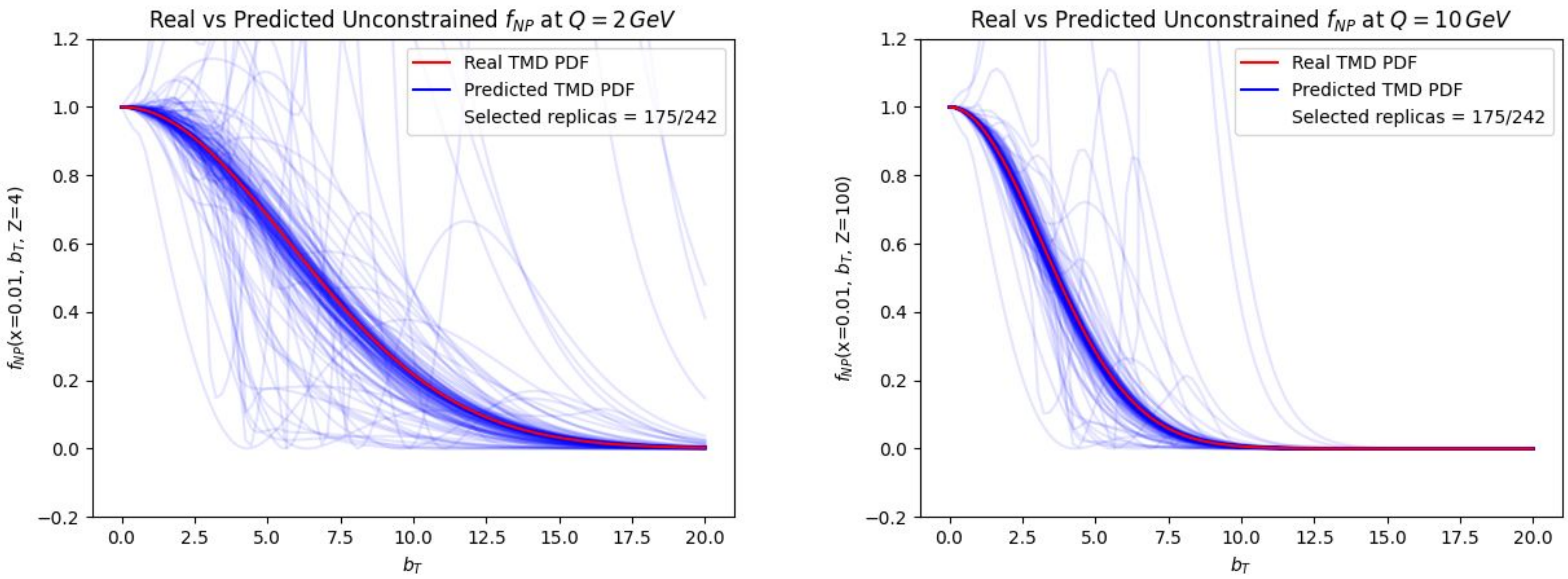
MAPNNoMAP22: Level 0





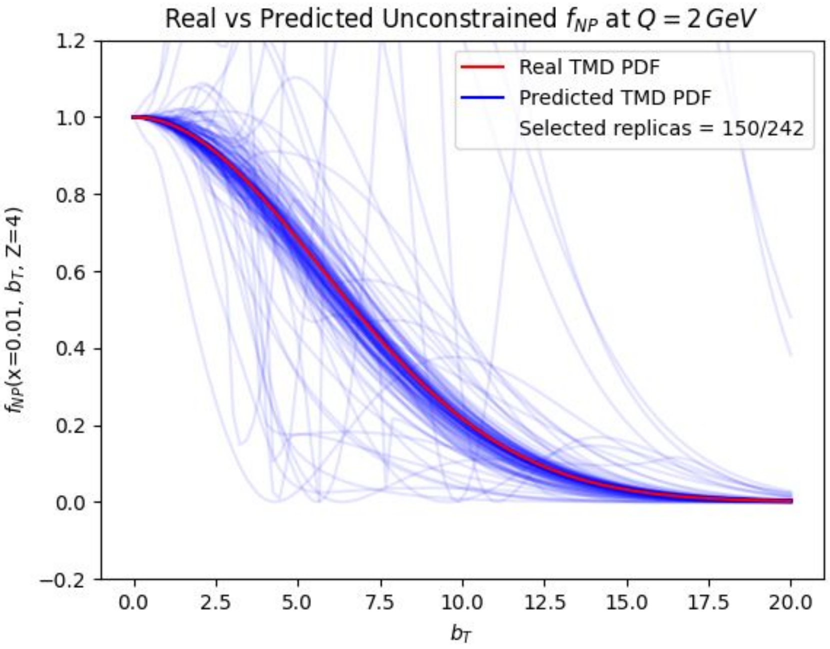
input: MAP22, *fit*: MAPNN

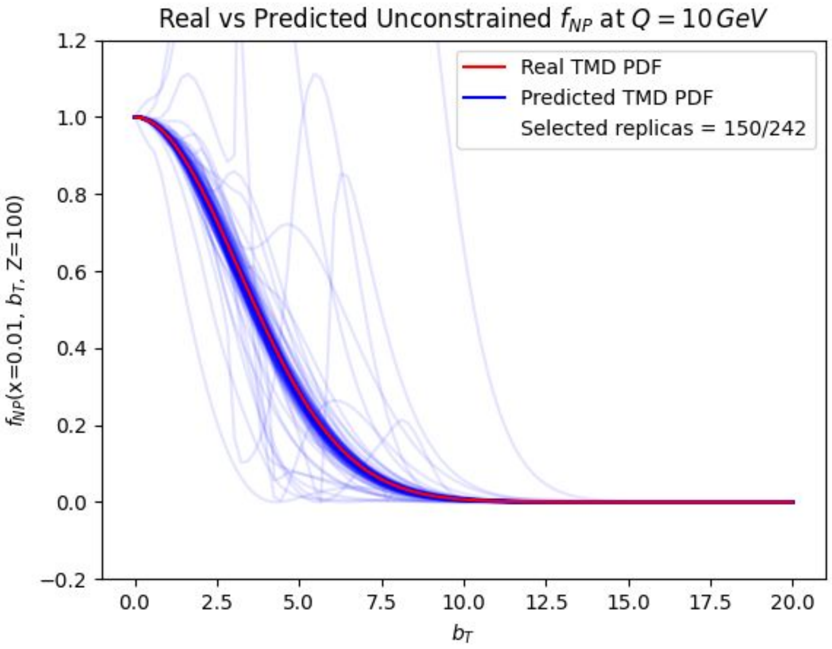
MAPNNoMAP22: Level 0



input: MAP22, *fit*: MAPNN

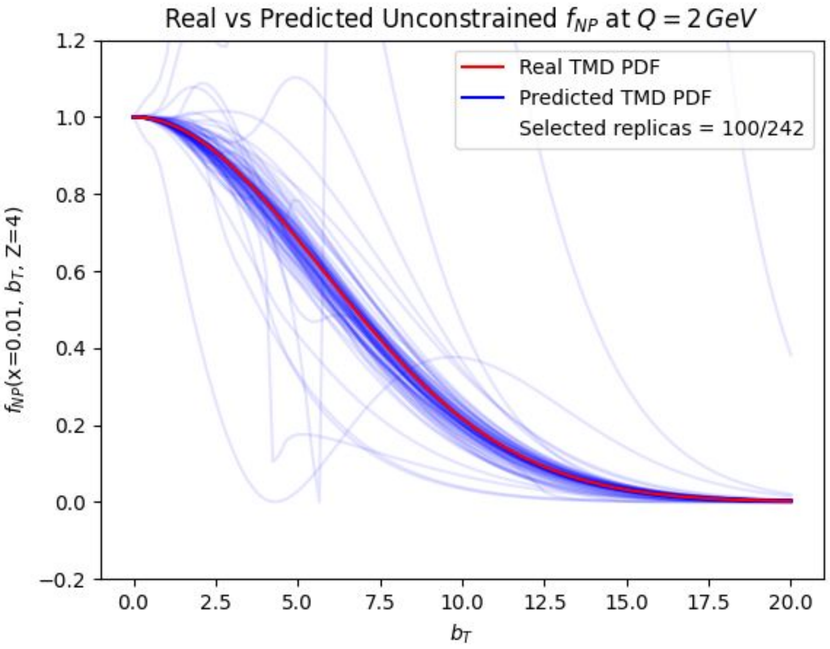
MAPNNoMAP22: Level 0

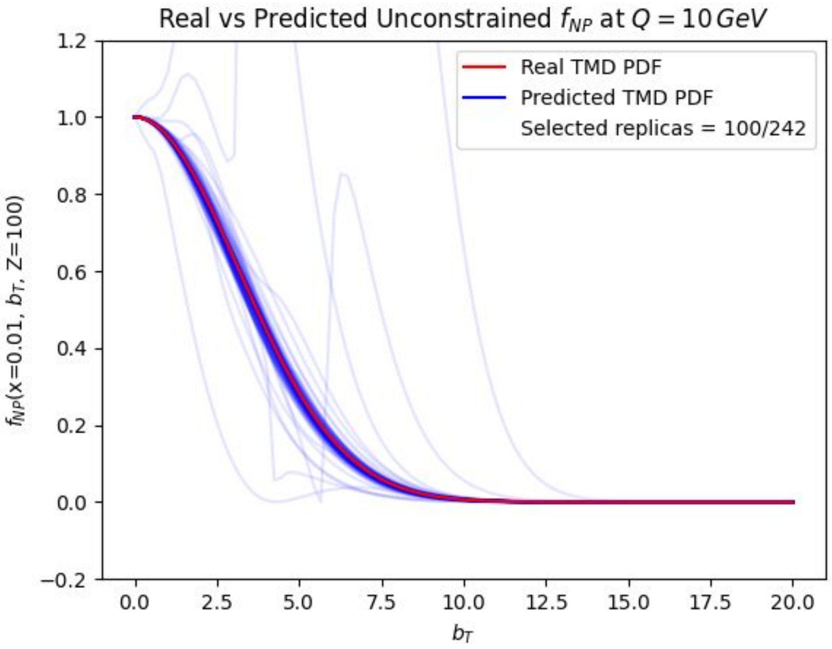




input: MAP22, *fit*: MAPNN

MAPNNoMAP22: Level 0





input: MAP22, *fit*: MAPNN

MAPNNoMAP22: Level 0

Why can a L0 test fail?

- too rigid **functional form**

AND/OR

- the **minimizer** can't find the solution in 100% of the cases

What do we observe?

- The real TMD lies in the chosen functional space

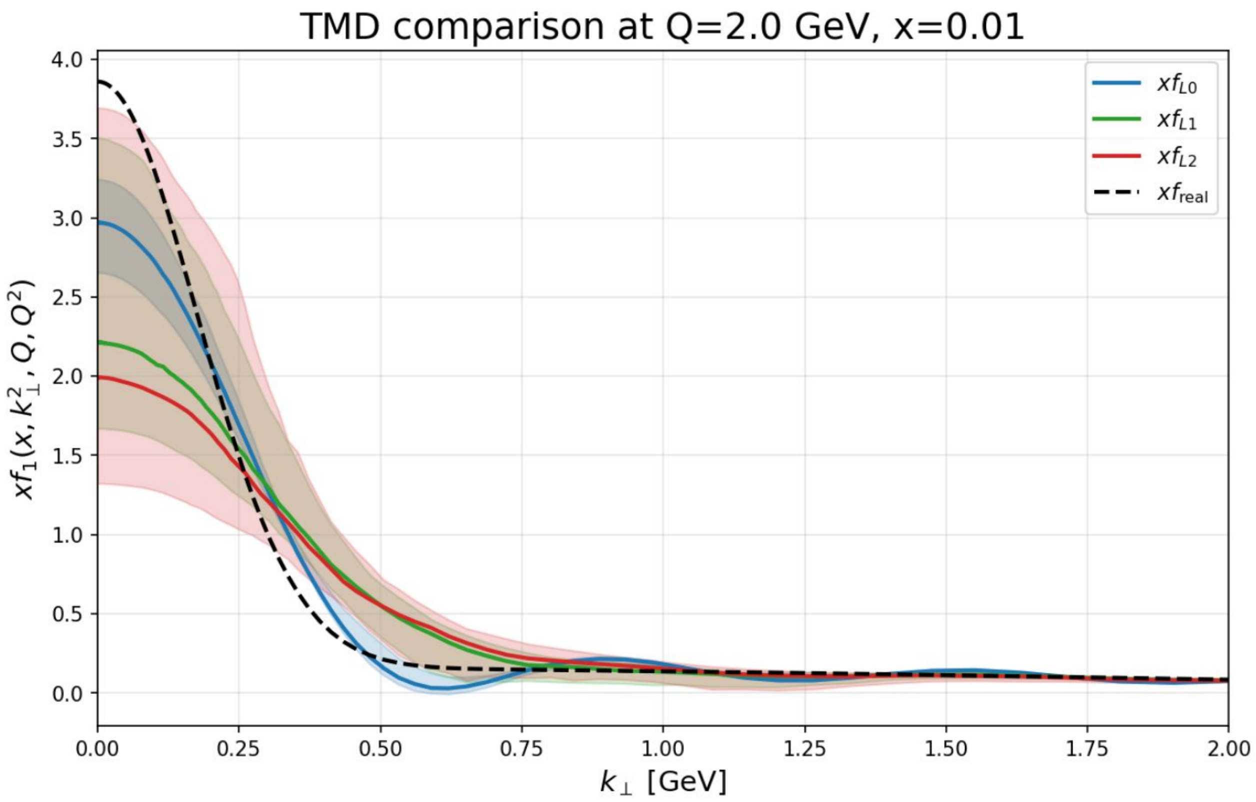
SO

- **it's the minimizer**'s fault

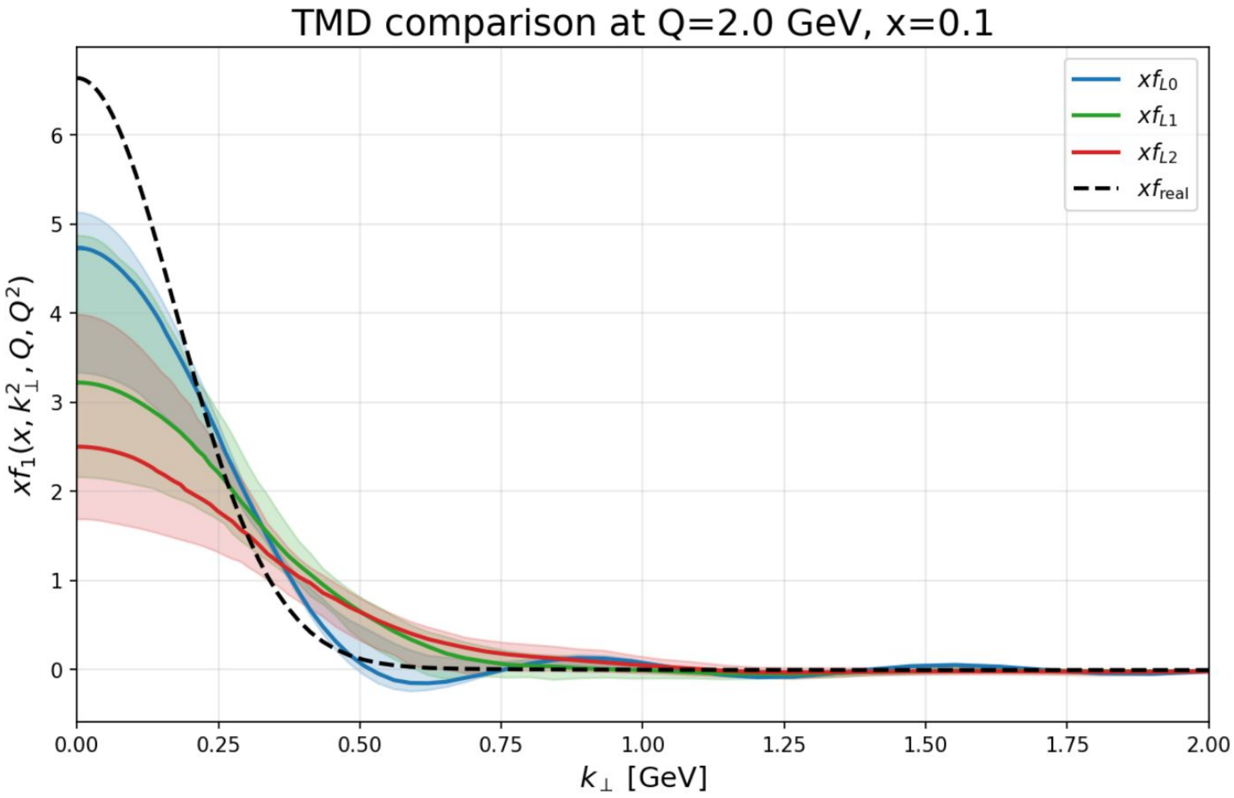
AND

- possibility: implementation of arc length estimator?

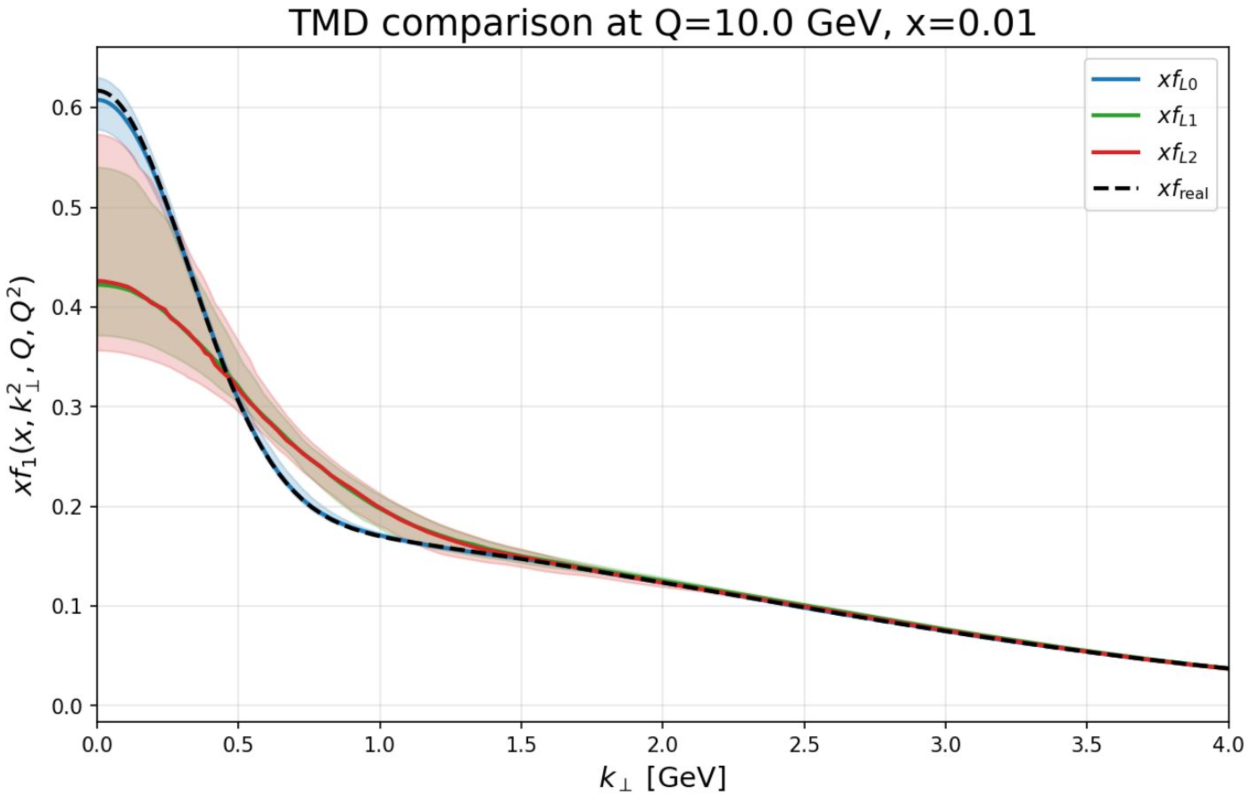
MAPNNoMAP22 - work in progress



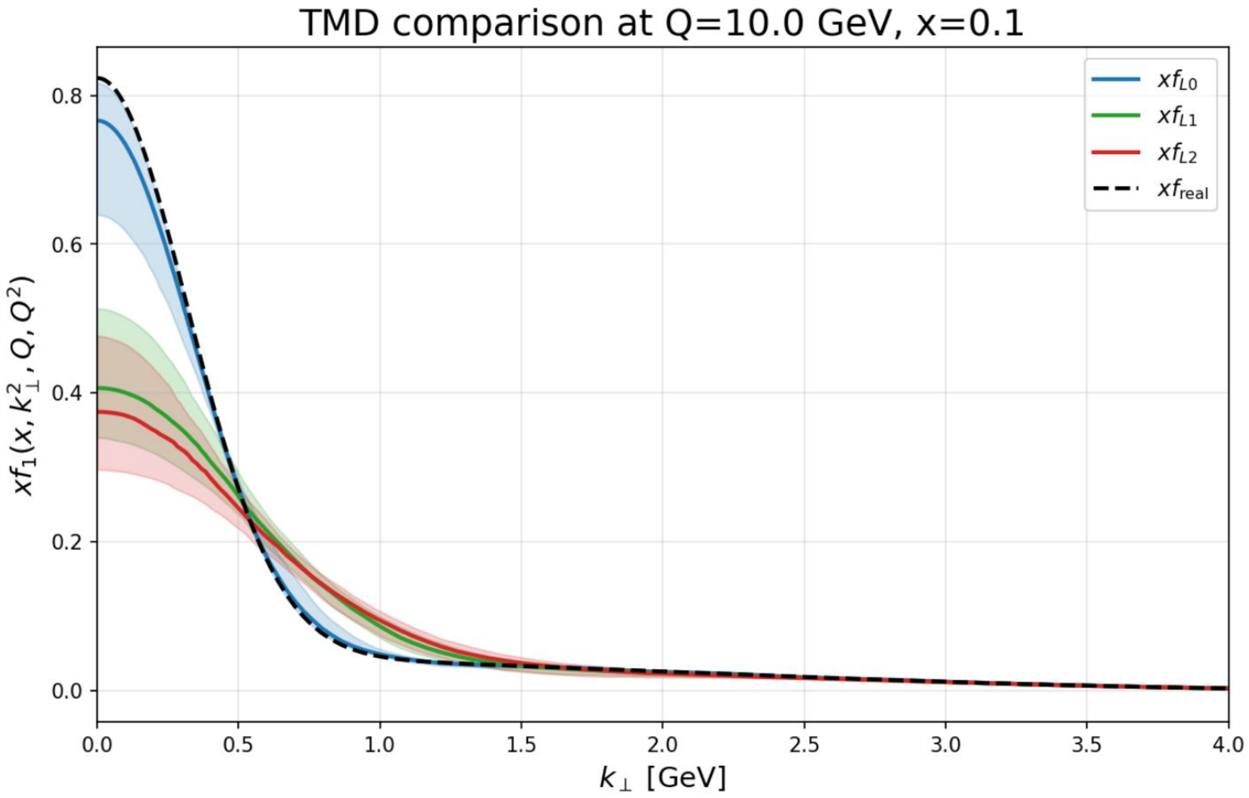
MAPNNoMAP22 - work in progress



MAPNNoMAP22 - work in progress



MAPNNoMAP22 - work in progress



Partial conclusions and outlook

1. **PV19** underfits and the uncertainties are underestimated
2. **MAPTMD22** is optimized (but they don't generalize well)
3. The interpolation/extrapolation uncertainties are subdominant in both frameworks, meaning **more precise data** should increase the predictivity of the models
4. The **NN** is statistically sound (see L2 tests) but seems unable to capture the input law (see L0 tests)
5. Before moving further with NNs **we should improve the methodology**