

MENU 2026

The 17th International Conference on Meson-Nucleon Physics and the Structure of the Nucleon

Bayesian Gaussian Reconstruction of TMDs:

An exact approach to solve Inverse Problem

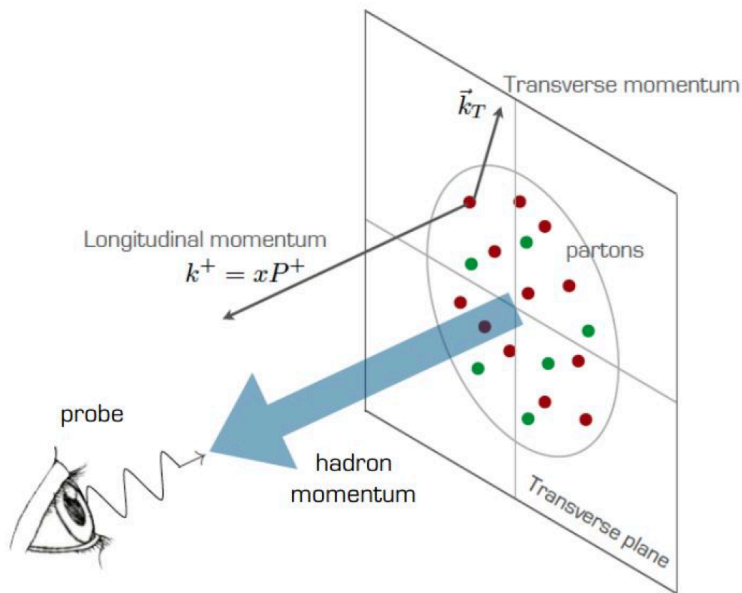
Poonam Choudhary

In collaboration with: Patrick C. Barry, Jianwei Qiu, Nobuo Sato, Marco Zaccheddu

Jefferson Lab, USA



Transverse Momentum Distributions(TMDs)



Credit picture: A. Bacchetta

$f_1(x)$
1D structure
in momentum space

$f_1(x, k_T^2)$
3D structure
in momentum space

SIDIS

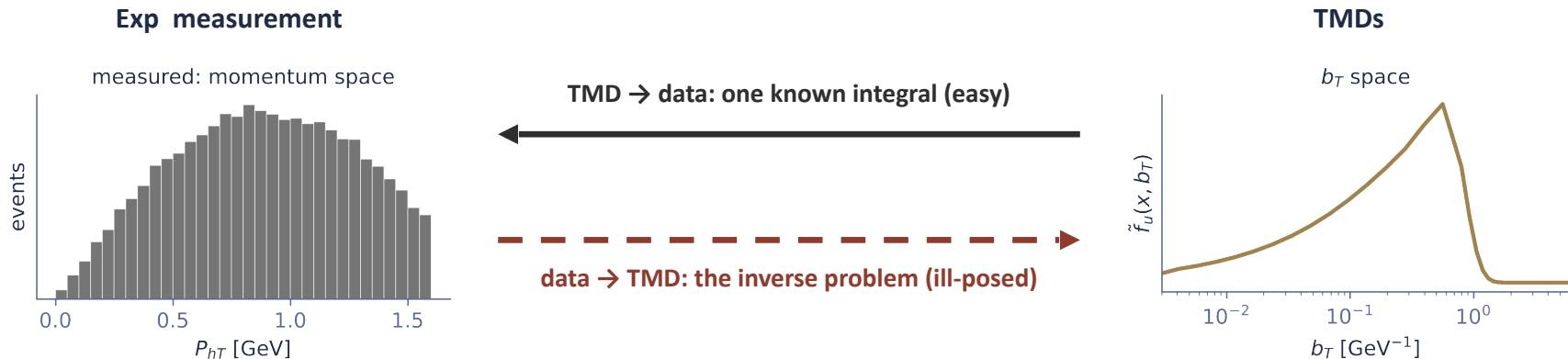
$$e + N \rightarrow e + h + X$$

- Beyond collinear
- three dimensional picture
- Transverse structure
- Spin momentum correlation
- Momentum momentum correlation

Data

- Jefferson Lab (Hall B/C) CLAS 12
- COMPASS (CERN)
- HERMES (DESY)
- Future EIC (BNL)

Inverse Problem



$$\sigma(x, Q^2, z, P_{hT}) \propto \int_0^\infty db_T b_T J_0\left(b_T \frac{P_{hT}}{z}\right) \tilde{f}(x, b_T; Q) \tilde{D}(z, b_T; Q)$$

Why ill-posed

the integral averages over b_T : different TMDs give the same data, so the data alone do not fix the TMD.

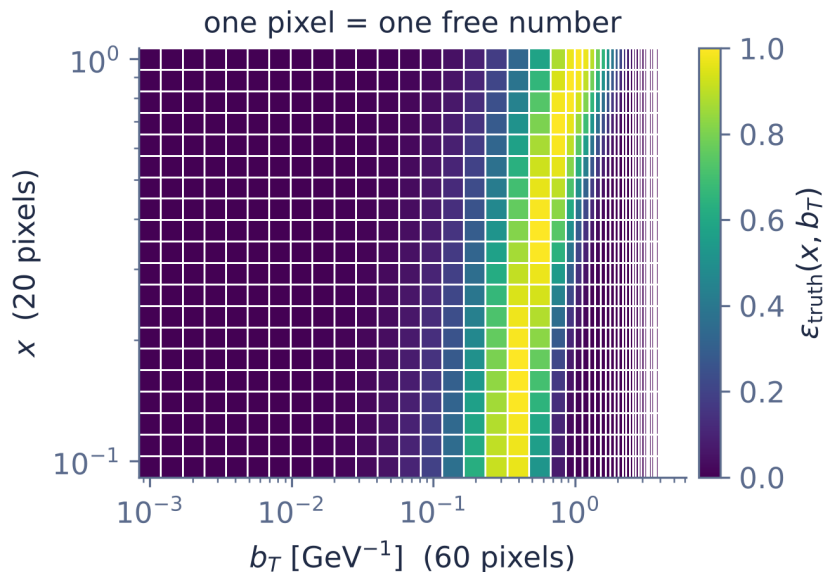
Standard approach

assume a functional form and fit a few parameters (JAM, NNPDF, MAP ART, DSSV): the form fills in what the data cannot fix.

This talk

no functional form: pixels in b_T with a stated prior. The posterior is exact and shows what the data determine, and what they cannot.

Pixel Based Approach



$$\epsilon = (\epsilon(x_i, b_j)), \quad i = 1..20, j = 1..60 \Rightarrow 1200 \text{ unknowns}$$

What are pixels ?

- Inspired by digital imaging
- Pixels $\rightarrow$ An image of discrete units
- A grid in (x, b_T)
- A literal 2d image of nucleon internal structure
- No functional form

Why pixels

the data, not a formula, decide where they can; nothing is imposed between pixels.

Talk by Wally Melnitchouk

Pixel TMDs with generative AI: M. Zaccheddu et al., arXiv:2605.06606

TMD Factorization in SIDIS

$$\frac{d\sigma}{dx dQ^2 dz dP_{hT}^2} = \sigma_0(x, Q^2) \sum_q e_q^2 \int_0^\infty \frac{db_T b_T}{2\pi} J_0\left(\frac{b_T P_{hT}}{z}\right) \tilde{f}_q(x, b_T; Q) \tilde{D}_q(z, b_T; Q)$$

$$\tilde{f}_q(x, b_T; Q) = f_q(x, \mu_{b_*}) S_{\text{pert}}(b_*; Q) e^{g_K(b_T) \ln(Q/Q_0)} [R_f + (1 - R_f) \epsilon(x, b_T)]$$

Truth used to generate the data

$$\epsilon_{\text{truth}}(x, b_T) = e \left(\frac{b_T}{\lambda_f(x)} \right)^2 \exp \left[- \left(\frac{b_T}{\lambda_f(x)} \right)^2 \right]$$

- collinear PDFs & FFs (JAM20 NLO)
- S_{pert} perturbative Sudakov
- **Fixed** FF non-perturbative model



How ϵ enters linearly in cross section ?

$$\sigma_v(\epsilon) = \sum_j K_{vj} [R_f + (1 - R_f) \epsilon]_j = \mathbf{a}_v \cdot \epsilon + b_v$$

$$\mathbf{t}(\epsilon) = \mathbf{A} \epsilon + \mathbf{b}$$

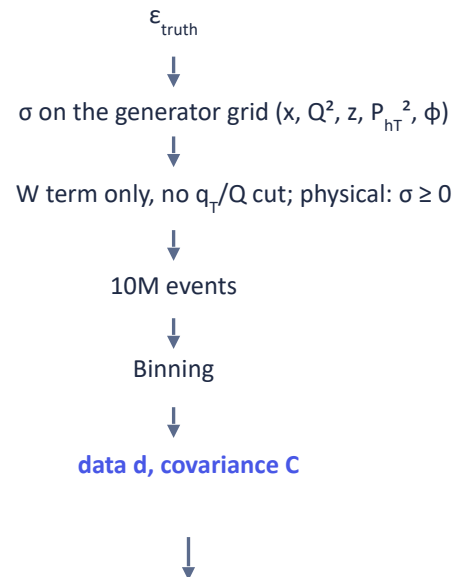
$$t_i(\epsilon) = \sum_v w_{iv} \sigma_v(\epsilon) = \underbrace{\left(\sum_v w_{iv} \mathbf{a}_v \right)}_{\text{row } i \text{ of } \mathbf{A}} \cdot \epsilon + \underbrace{\sum_v w_{iv} b_v}_{b_i}$$

- $\mathbf{A}$ and $\mathbf{b}$ are independent of choice of ϵ
- Find $\mathbf{A}$ and $\mathbf{b}$ for all bins



The whole theory = one matrix $\mathbf{A}$ and one vector $\mathbf{b}$, built once

Referenced Events and binning



Matrices $\mathbf{A}$, $\mathbf{b}$, $\mathbf{d}$, $\mathbf{C}$

Bayesian Gaussian Inference

Prior on pixels

$$\pi(\epsilon) \propto \exp\left[-\frac{1}{2} \epsilon^T \Sigma_0^{-1} \epsilon\right], \quad \Sigma_0^{-1} = \tau^{-2} I + \lambda_b D_b^T D_b + \lambda_x D_x^T D_x$$

Likelihood: Gaussian errors, covariance C of the bins

$$p(\mathbf{d} \mid \epsilon) \propto \exp\left[-\frac{1}{2} (\mathbf{d} - A\epsilon - \mathbf{b})^T C^{-1} (\mathbf{d} - A\epsilon - \mathbf{b})\right]$$

Bayes theorem:

$$p(\epsilon \mid \mathbf{d}) = \frac{p(\mathbf{d} \mid \epsilon) \pi(\epsilon)}{p(\mathbf{d})}$$

$\mathbf{d} - \mathbf{t}(\epsilon) = \mathbf{d} - \mathbf{b} - A\epsilon$ should be linear in ϵ

Step by step

a) Multiply likelihood $\times$ prior, with $\mathbf{r} \equiv \mathbf{d} - \mathbf{b}$

$$-2 \ln P = (\mathbf{r} - A\epsilon)^T C^{-1} (\mathbf{r} - A\epsilon) + \epsilon^T \Sigma_0^{-1} \epsilon + \text{const}$$

b) Expand: only ϵ^2 and ϵ terms (linearity)

$$= \epsilon^T H \epsilon - 2 \epsilon^T \mathbf{h} + \text{const}, \quad H \equiv A^T C^{-1} A + \Sigma_0^{-1}, \quad \mathbf{h} \equiv A^T C^{-1} \mathbf{r}$$

c) Complete the square

$$= (\epsilon - \mu)^T H (\epsilon - \mu) + \text{const}, \quad \mu = H^{-1} \mathbf{h}$$

$$P(\epsilon \mid \mathbf{d}) = \mathcal{N}(\mu_{\text{post}}, \Sigma_{\text{post}}), \quad \Sigma_{\text{post}} = (A^T C^{-1} A + \Sigma_0^{-1})^{-1}, \quad \mu_{\text{post}} = \Sigma_{\text{post}} A^T C^{-1} (\mathbf{d} - \mathbf{b})$$

Exact Gaussian posterior

Posterior covariance

Posterior mean

Closure tests: A way to validate methodology

Data side

fixed theory with ϵ_{truth} (one specific function)

→ pseudo-data $\mathbf{d}$

Fit side

same fixed theory with free pixels ϵ

→ posterior for ϵ

closure: posterior $\approx$
 ϵ_{truth} ?

The loop closes if what went into the data comes back out of the fit: only ϵ differs between the two sides, the theory is the same.

Level 0: noiseless

Data

$$\mathbf{d} = A\epsilon_{\text{truth}} + \mathbf{b}$$

Tests

- Is the truth recovered?
- Where do the data decide, where the prior?

Level 1: noisy

Data

$$\mathbf{d} = A\epsilon_{\text{truth}} + \mathbf{b} + \eta, \text{ 100 experiments}$$

Tests

- Are the bands honest (coverage)?
- Bias vs. scatter from noise

Level2: Pseudo data truth model

Data

10M generated events, binned

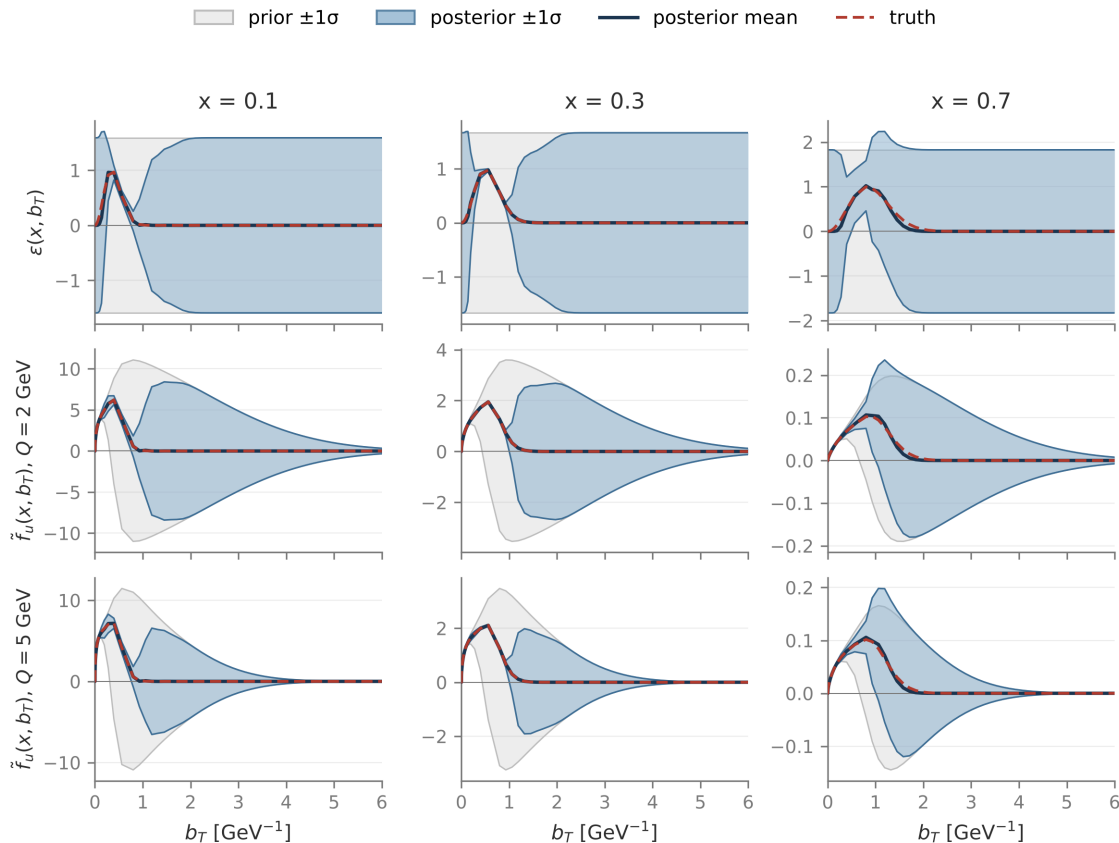
Tests

- physical data: $\sigma \geq 0$ everywhere (positive probability)
- σ is non linear

*Why: real data have no truth. Only a known truth shows whether the method, its bands and its prior can be trusted. **Level 3: Real data***

Closure-test levels: NNPDF, R. D.
Ball et al., JHEP 04 (2015) 040

Results: Level 0



Prior: independent pixels, $\tau = 2$

mean 0, no link between neighbouring pixels

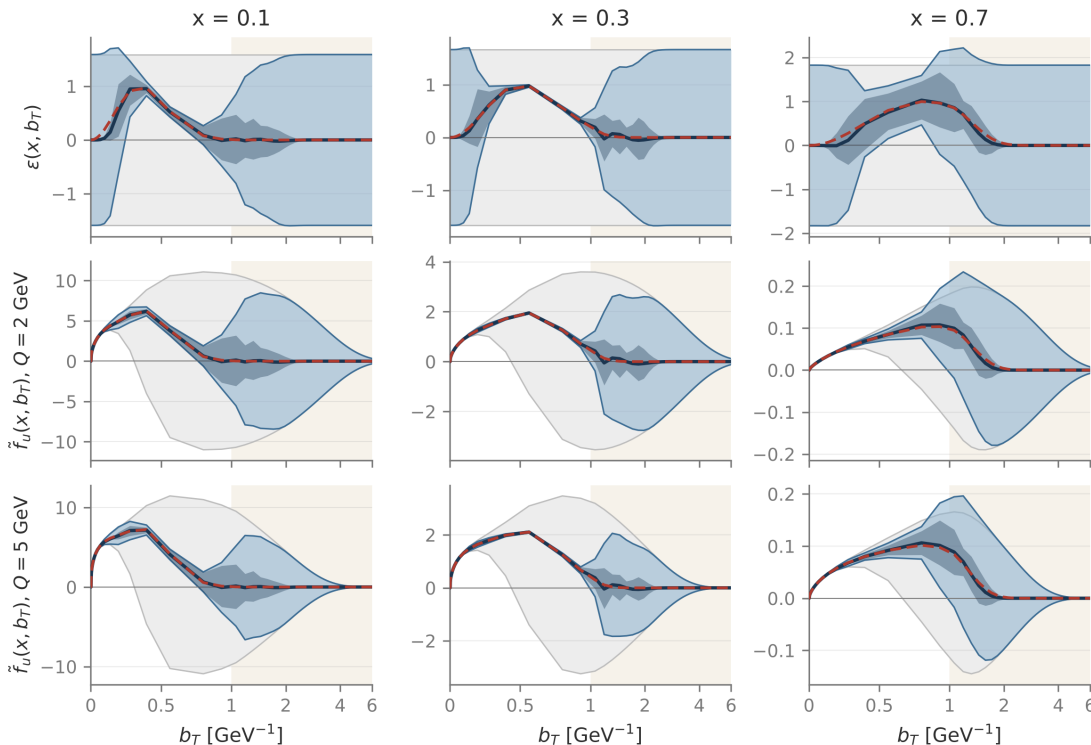
Posterior

- Small b_T : ϵ switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE → **expected**
- Window $b_T \approx 0.1\text{--}1.5$: truth recovered, band $\approx \frac{1}{2}$ of the prior → **honest but broad**
- Large $b_T \gtrsim 2.5$: data silent: posterior = prior → **not constrained**

Honest (truth inside the band), not narrow: nothing ties neighbouring pixels.

Results: Level 1

prior $\pm 1\sigma$
 mean of 100 posteriors
 spread of the 100 means
 large b_T (log scale)
 posterior $\pm 1\sigma$
 truth



Level 1 noise

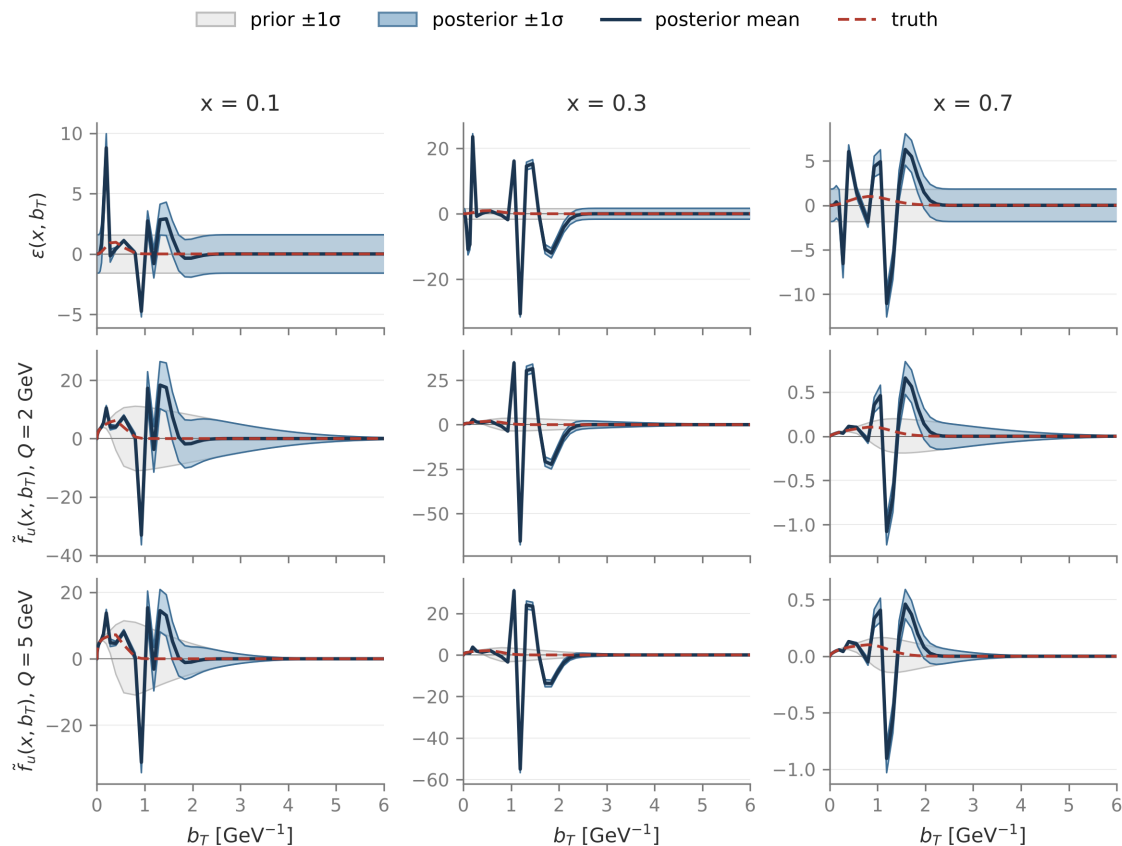
- $\eta \sim \mathcal{N}(0, \mathbf{C})$: $\mathbf{C}$ = covariance of the 10M-event bins ($\approx$ Poisson)
- **100 draws**: 100 independent $\eta = 100$ pseudo-experiments, each fitted separately
- **Prior**: smooth, $\tau = 2$, $\lambda = 30$

Posterior

- **Small b_T** : ϵ switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE $\rightarrow$ **expected**
- **Window**: mean moves $\approx 0.08 < \text{band } 0.18$; truth inside in 94% $\rightarrow$ **good**
- **Large b_T** : posterior = prior $\rightarrow$ **not constrained**

Unbiased (average of 100 = truth) and honest ($\chi^2/n = 0.84$): noise moves each answer by less than its band.

Results: Level 2



Event data, all 787 bins, linear theory

- **Events are physical:** $\sigma \geq 0$ everywhere
- **Linear theory is not:** $\sigma < 0$ at high P_{HT} , even at the truth

Posterior

- **Small b_T :** ϵ switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE → **expected**
- **Window:** mean oscillates far from the truth; narrow band around a wrong answer, $\chi^2/n \approx 370$ → **bad**
- **Large b_T :** distorted up to $b_T \approx 2.5$, then prior → **bad**

Not a failure of the method: data and theory disagree where the linear theory is unphysical.

Gaussian Reconstruction on SIDIS cross section

Reference events and bins side

- $\sigma = W$ term only (TMD region), no Y term
- no q_T/Q cut: W is used up to high P_{hT} , beyond its validity
- there $\sigma_v = a_v \cdot \varepsilon + b_v < 0$, even for $\varepsilon_{\text{truth}}$
- physical data and theory need $\sigma \geq 0$: **this breaks linearity**

The data side is there are no unphysical bins or x section

Theory side:

- In a linear theory if we put any positivity constraint, it's Linearity breaks.
- there are bins which have negative cross section from theory.
- There are unphysical bins.

Case 1: remove unphysical bins

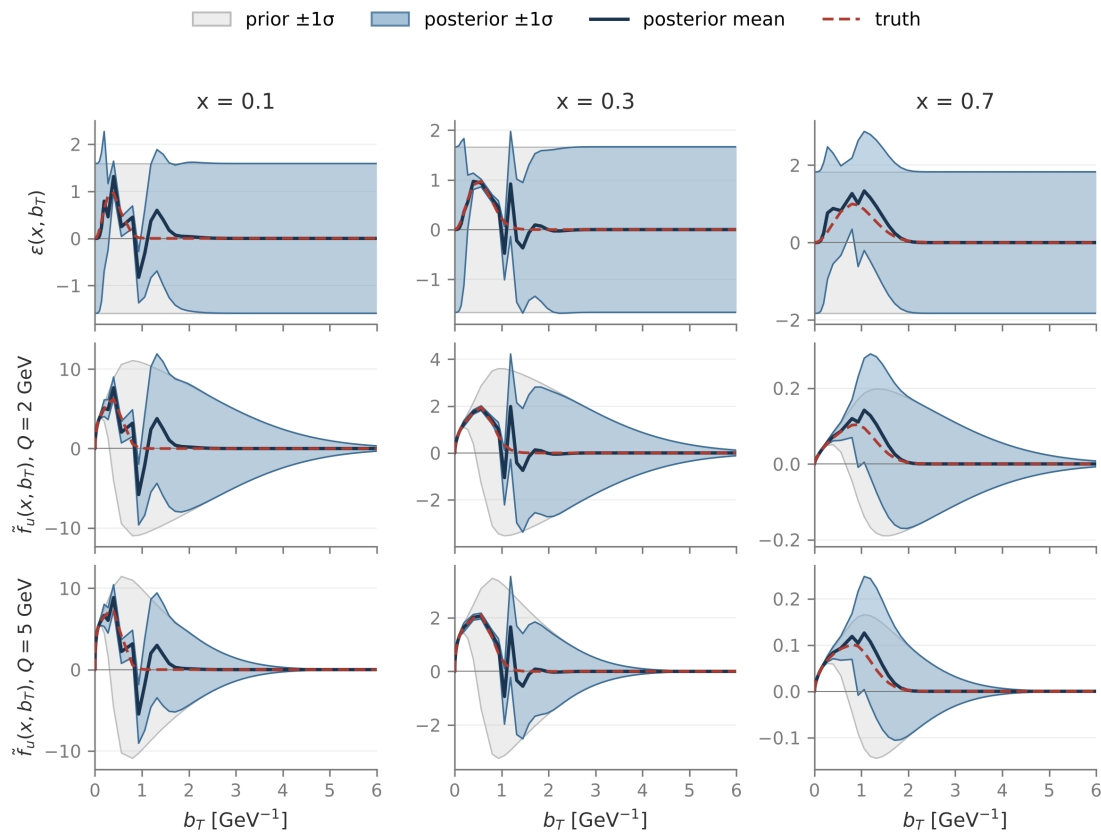
- drop every bin with a $\sigma < 0$ grid point
- no kinematic cut, theory unchanged
- remaining bins exactly linear: same exact method

Case 2: keep all bins, physical theory

- theory = only physical probability and cross section (non-linear in ε)
- Gauss–Newton: fix which points are positive $\rightarrow$ linear (A_s, b_s)
- exact Gaussian solve, update the pattern, repeat
- converged: Gaussian band around the best fit

Both reuse the same Gaussian reconstruction.

Case 1: Unphysical bins removed



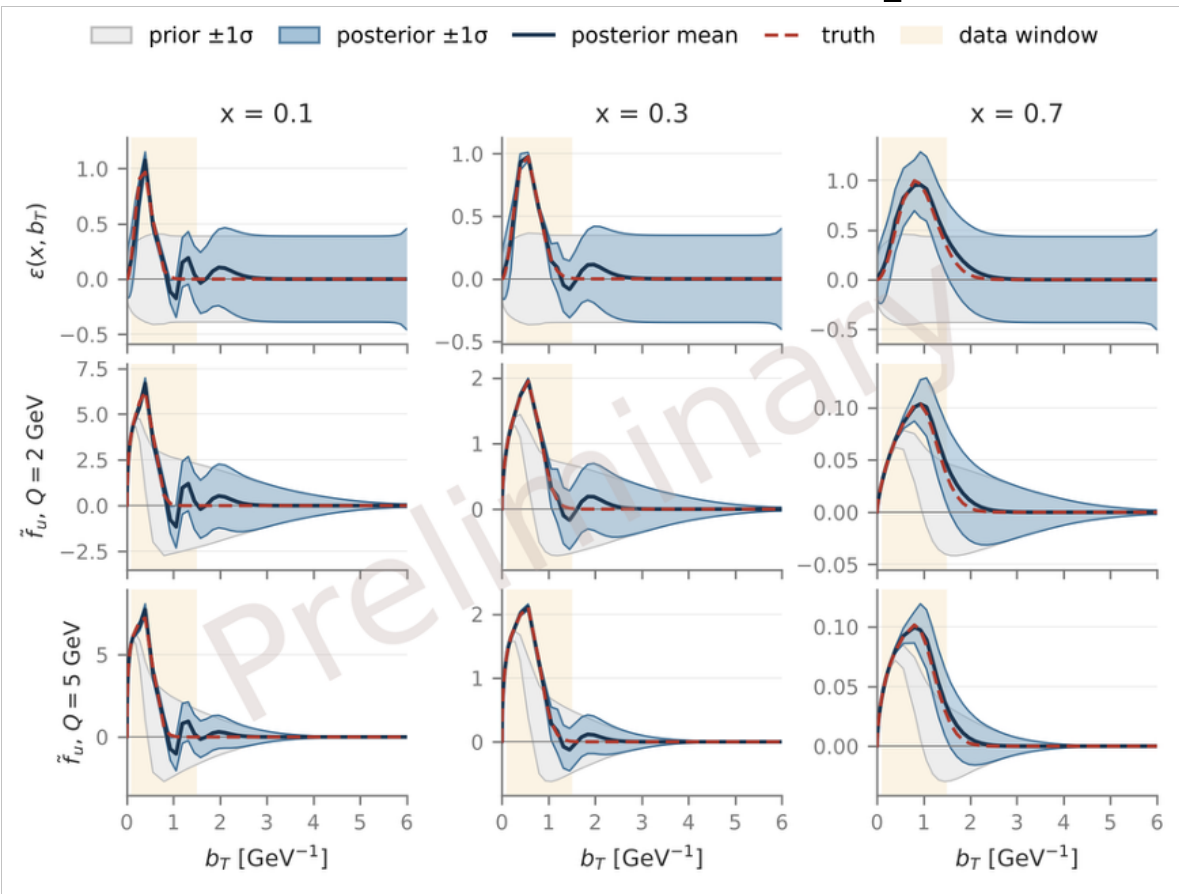
Prior: independent pixels, $\tau = 2$

Posterior

- **Small b_T :** ϵ switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE → **expected**
- **Window:** mean follows the data but wiggles from pixel to pixel → **bad**
- **Large b_T :** posterior = prior, very broad → **not constrained**
- **Why:** nothing ties neighbouring pixels, so noise in weakly seen patterns goes straight into ϵ

Fix: penalise roughness in the prior → smooth (Tikhonov) prior.

Results: Case 1 +Smooth prior



Smooth (Tikhonov) prior: $\tau = 2$, $\lambda = 30$

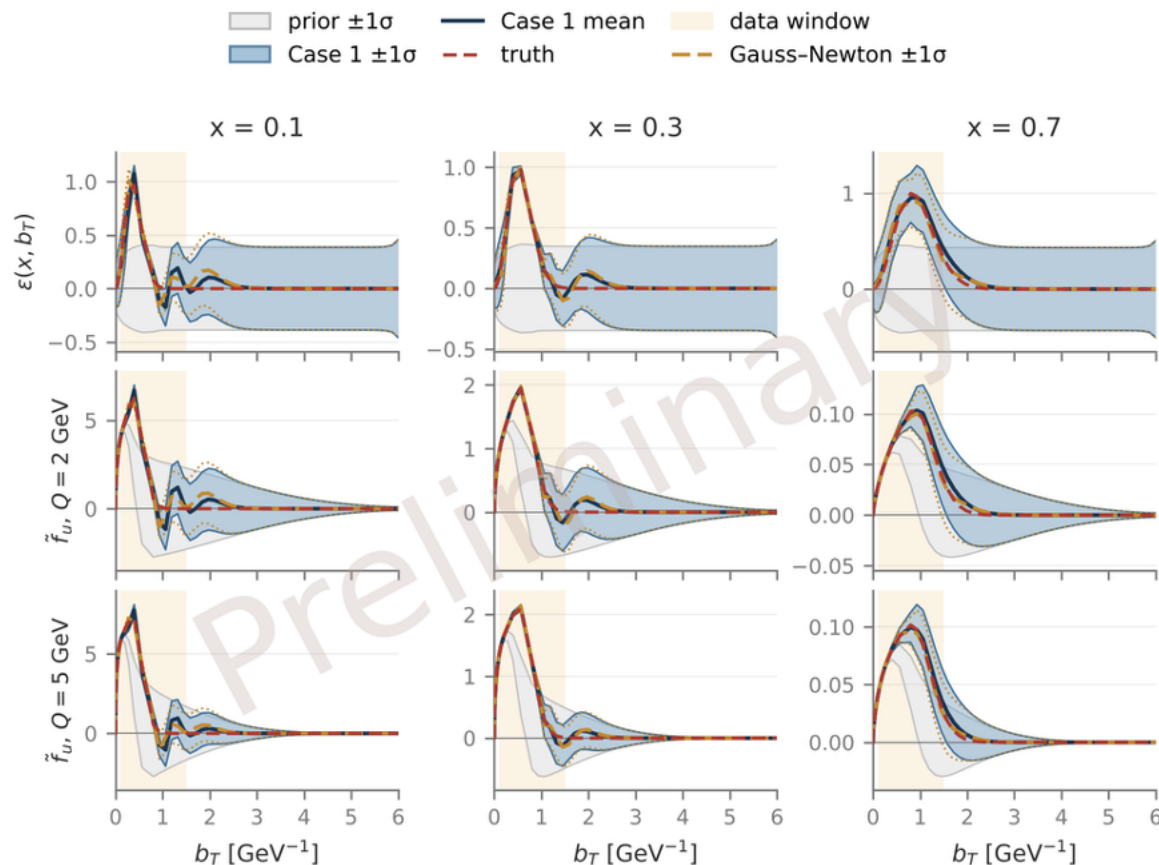
neighbouring pixels tied in b_T and x : a gradient penalty suppresses oscillations

Posterior

- **Small b_T :** ε switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE → **expected**
- **Window:** truth recovered, band $\approx 5\times$ narrower; 93% inside → **good**
- **Edge $b_T \approx 1\text{--}2.5$:** small wiggles, inside the band → **acceptable**
- **Large b_T :** posterior = prior → **not constrained**

Honest and narrow ($\chi^2/n = 0.91$, 598 bins):
the linear method is exact on the physical bins.

Case2 : Gauss–Newton Method



All 787 bins, physical theory $\sigma \geq 0$

- **Gauss–Newton:** fix which grid points are positive $\rightarrow$ linear (A_S, b_S) $\rightarrow$ exact Gaussian solve $\rightarrow$ repeat
- same answer from any start; agrees with Case 1 within the bands

Posterior

- **Small b_T :** ε switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE $\rightarrow$ **expected**
- **Window:** truth inside the band (93%), $\chi^2/n = 0.92 \rightarrow$ **good**
- **Edge / large b_T :** same wiggles, then prior $\rightarrow$ **not constrained**

Why wiggles at the edge and no resolution at large b_T ? $\rightarrow$ SVD (next).

Singular Value Decomposition (SVD)

1 Decompose the response matrix

$$A = U S V^T, \quad A \mathbf{v}_k = s_k \mathbf{u}_k$$

- **A**: theory matrix, bins scaled by their errors
- $\mathbf{v}_k$: a pattern of pixels in (x, b_T)
- $\mathbf{u}_k$: the pattern it makes in the bins (P_{hT})
- s_k : how many σ the bins move per unit of $\mathbf{v}_k$

2 Split the pixels by s_k

- $s_k > 1$: data beat the noise $\rightarrow$ observable $\rightarrow$ **data decide**
- $s_k < 1$: below the noise $\rightarrow$ weak / null space $\rightarrow$ **prior decides**

$$\epsilon = P_{\text{obs}} \epsilon + P_{\text{null}} \epsilon, \quad P_{\text{obs}} = \sum_{s_k > 1} \mathbf{v}_k \mathbf{v}_k^T$$

3 What it means

$$A \epsilon = A \epsilon_{\text{obs}} + \underbrace{A \epsilon_{\text{null}}}_{< 1\sigma \text{ in every bin}} \approx A \epsilon_{\text{obs}}$$

- the null part moves no bin by more than 1σ : no method can recover it from these data
- tells us where to trust the result, and where the band is the prior

SVD shows what the data can see, and what they cannot.

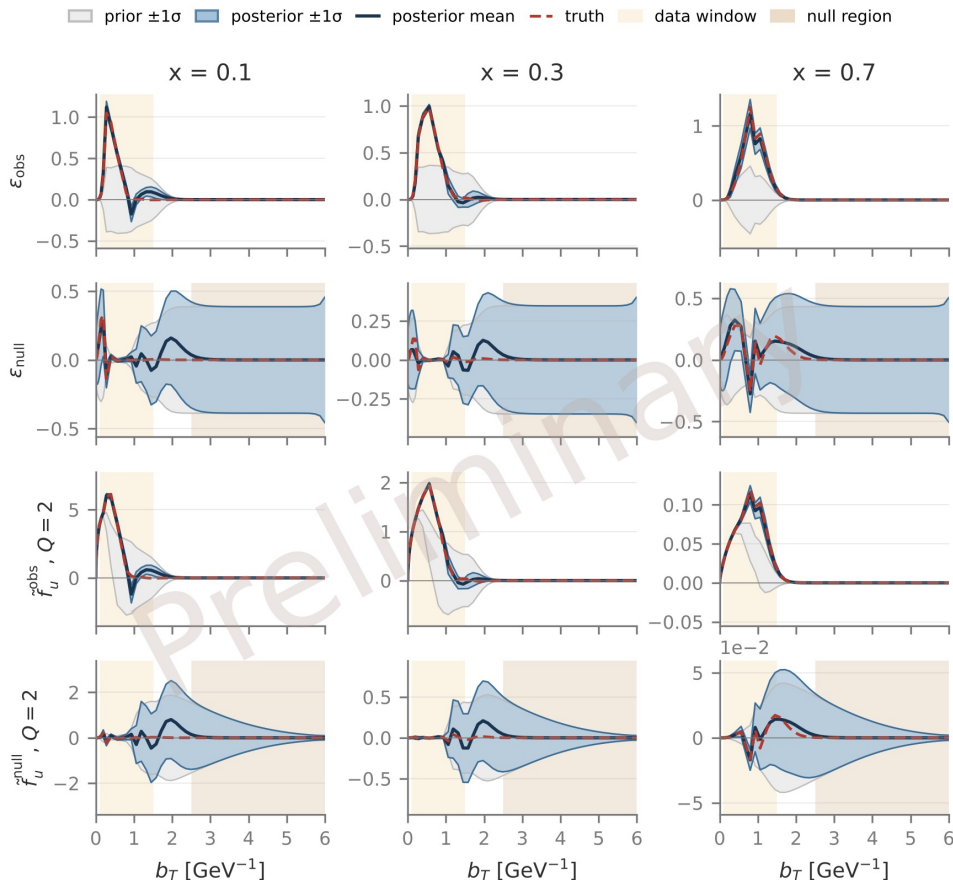
If only the first p singular values are non-zero:

$$V_{\text{obs}} = [\mathbf{v}_1, \dots, \mathbf{v}_p]$$
$$V_{\text{null}} = [\mathbf{v}_{p+1}, \dots, \mathbf{v}_N]$$

$$P_{\text{obs}} = V_{\text{obs}} V_{\text{obs}}^T = \sum_{i=1}^p \mathbf{v}_i \mathbf{v}_i^T$$

$$P_{\text{null}} = V_{\text{null}} V_{\text{null}}^T = \sum_{i=p+1}^N \mathbf{v}_i \mathbf{v}_i^T$$

TMDs in Null Space



Final result (Gauss–Newton), split by SVD

- **Observable part** : posterior = truth, narrow $\rightarrow$ good
- **Null part**: posterior $\approx$ prior; the truth's null part is not recovered, by any method

Where

- **Small b_T** : not missing: ϵ switched off by design, TMD = OPE
- **Window edge $b_T \approx 1\text{--}2.5$** : weakly seen patterns $\rightarrow$ wiggles, inside the band
- **Large $b_T \gtrsim 2.5$** : the FF non-perturbative factor suppresses the signal $\rightarrow$ null space

To shrink the null space: data more sensitive to large b_T (small P_{hT} , wider Q^2 range, FF).

Summary and Outlook

- SIDIS data in momentum space. To extract TMDs from data, it is **inverse problem**.
- Standard approaches are based on parameterization and costly.
- Introduce a pixel based, model and parameterization independent non perturbative TMD in x , b_T .
- Pixel + A theory linear in ϵ + Bayesian inference $\rightarrow$ give an exactly Gaussian posterior
- Pixels learned in small to moderate b_T but at high b_T model could not learn from data, which is null space
- For a given data and setup, no method can learn in that regime.
- The null space can get modified with Q^2 and different choice of TMD.
- The method can be checked for real experimental data where there is no truth model.

Exact, fast and transparent: a benchmark for TMD extraction and for machine-learning methods on the same inputs.

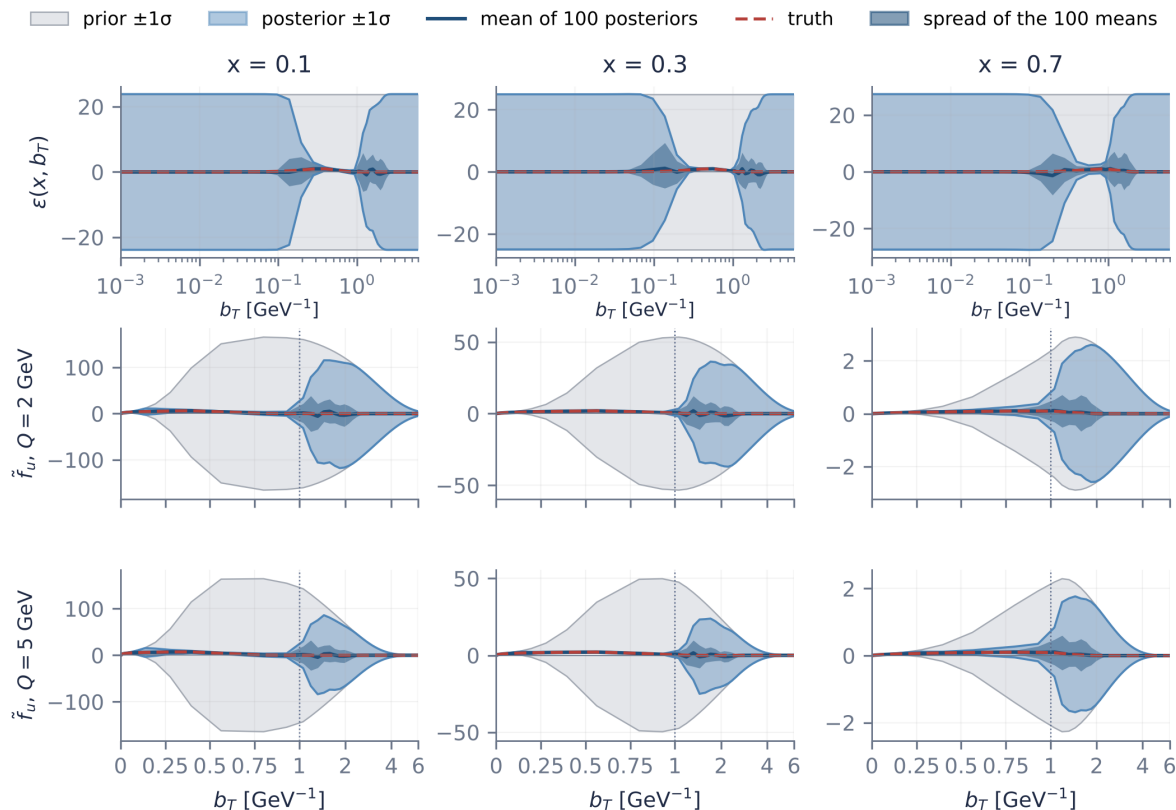
Back up

Backup: why other methods cannot use the exact solution

$t(\epsilon) = A\epsilon + b$ is linear in ϵ . Write $\epsilon = g(\theta)$ with a formula or a network, and $t(\theta) = A g(\theta) + b$ is nonlinear in θ : no square to complete.

	Unknowns	Theory in the unknowns	Posterior / computation	What it pays
Parameterized fit (MAP, SV19, JAM)	few parameters θ in a chosen $f_{\text{NP}}(x, bT; \theta)$	nonlinear in θ	not Gaussian: Hessian approximation, replicas or MCMC	the formula fills the null space; bands reflect the form, often too narrow where data are silent
Neural network / GAN	network weights w (thousands)	strongly nonlinear in w	no closed form: training, ensembles of seeds or replicas	architecture, initialization and training fill the null space implicitly; costly; coverage hard to verify
This talk: pixels	pixel values $\epsilon(x, bT)$	linear: $A\epsilon + b$	exactly Gaussian: one linear solve	prior must be stated (and is visible); theory must be linear in ϵ ($\sigma \geq 0$ constraint $\rightarrow$ Gauss–Newton)

Backup: wider independent prior ($\tau = 30$), 100 noisy experiments



Prior: independent pixels, width $\tau = 30$. Rows: ε , TMD at $Q = 2$ and 5 GeV .

Prior: independent, $\tau = 30$

very weak assumption, no link between neighbours.

Window

band ≈ 8 in ε ; the 100 means scatter widely: noise is amplified.

TMD

huge bands above $b_T \approx 1$.

Honest but useless: the width reflects the weak prior, not the data. Widening an independent prior does not help.

Singular Value Decomposition (SVD)

Decompose the raw forward matrix as

$$A = U S V^T, \quad A \mathbf{v}_k = s_k \mathbf{u}_k$$

v_k is pixels coefficient

s_k magnitude of response in data space



U and V are orthogonal Matrices

- U data space in kT
- V model space or pixel space in (x, bT)
- S singular diagonal matrix with values s_k



SVD is a diagnostic tool to Analyze mapping b/w data Space and pixel space through response matrix.

If only the first p singular values are non-zero:

$$V_{\text{obs}} = [\mathbf{v}_1, \dots, \mathbf{v}_p]$$

$$V_{\text{null}} = [\mathbf{v}_{p+1}, \dots, \mathbf{v}_N]$$

$$P_{\text{obs}} = V_{\text{obs}} V_{\text{obs}}^T = \sum_{i=1}^p \mathbf{v}_i \mathbf{v}_i^T$$

$$P_{\text{null}} = V_{\text{null}} V_{\text{null}}^T = \sum_{i=p+1}^N \mathbf{v}_i \mathbf{v}_i^T$$

We can split the model into observable and null space.

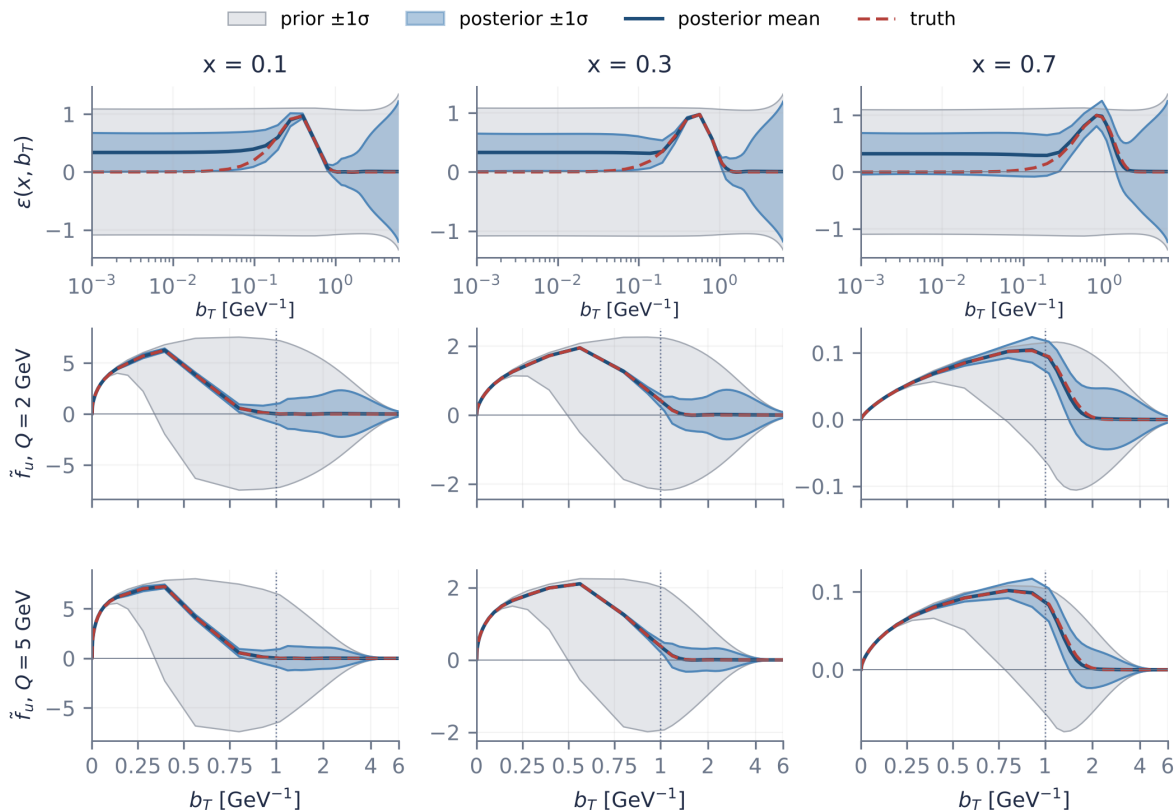
$$\epsilon = P_{\text{obs}} \epsilon + P_{\text{null}} \epsilon, \quad P_{\text{obs}} = \sum_{s_k > 1} \mathbf{v}_k \mathbf{v}_k^T$$

Only observable space can be constructed from data, rest is null

$$A \epsilon = A \epsilon_{\text{obs}} + \underbrace{A \epsilon_{\text{null}}}_{\approx 0} \approx A \epsilon_{\text{obs}}$$

$A v_i = 0$ for all modes in null space

Backup: wide smooth prior ($\tau = 30$): overconfident where ε is off



Prior: smooth, width $\tau = 30$, smoothing $\lambda = 30$ in b_T and x . Rows: ε , TMD at $Q = 2$ and 5 GeV .

Smooth prior

neighbouring pixels similar in b_T and x ($\lambda = 30$), width $\tau = 30$.

Window $b_T \approx 0.1\text{--}1.5$

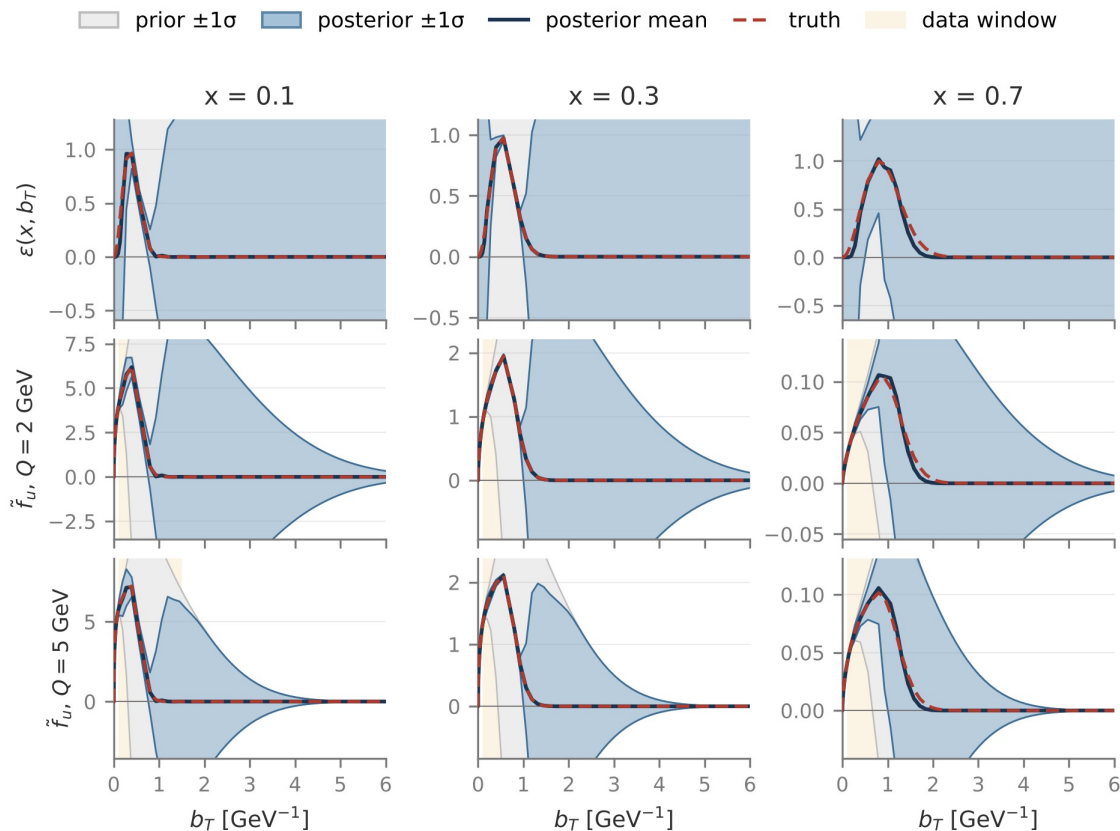
ε and TMD recovered; band much narrower than the prior.

Small b_T

ε switched off, yet the band shrinks below the prior: smoothing carries the window value inward.

Narrow in the window, but smoothing everywhere also invents information where the theory does not use ε .

Results: Level 0



Prior: independent pixels, $\tau = 2$

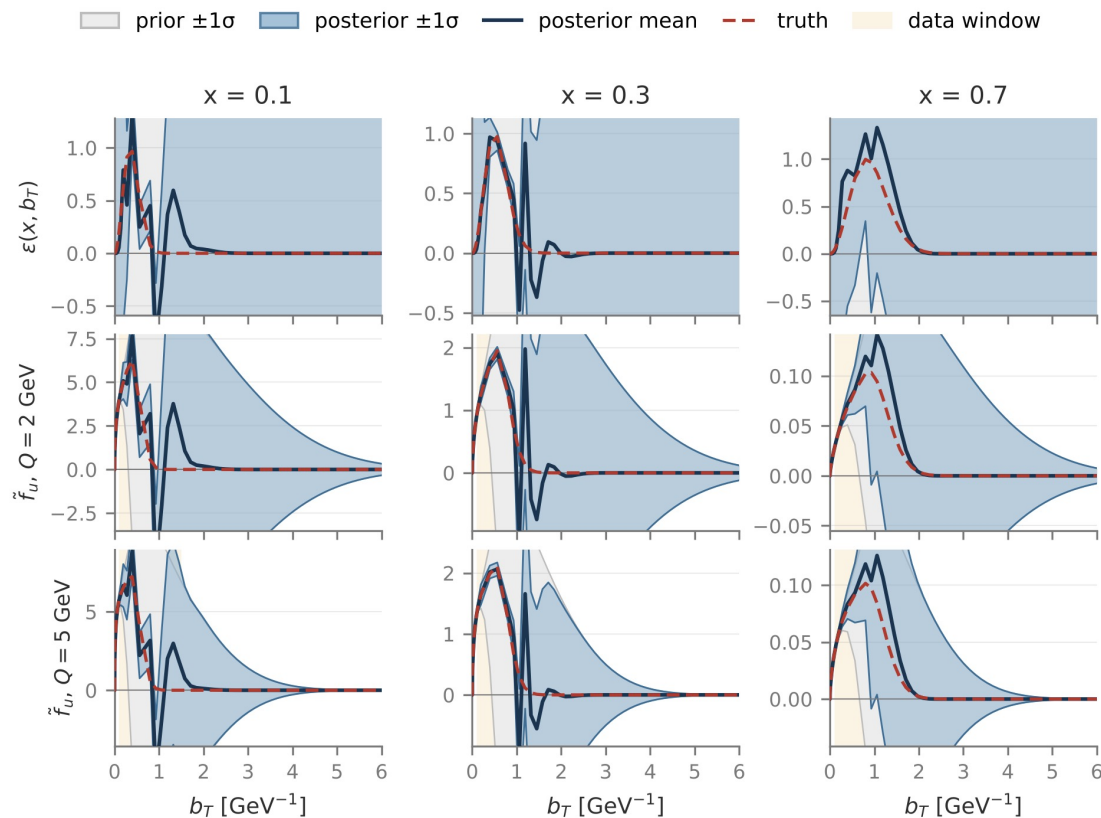
mean 0, no link between neighbouring pixels

Posterior

- **Small b_T :** ε switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE → **expected**
- **Window $b_T \approx 0.1\text{--}1.5$:** truth recovered, band $\approx \frac{1}{2}$ of the prior → **honest but broad**
- **Large $b_T \gtrsim 2.5$:** data silent: posterior = prior → **not constrained**

Honest (truth inside the band), not narrow:
nothing ties neighbouring pixels.

Case 1: Unphysical bins removed



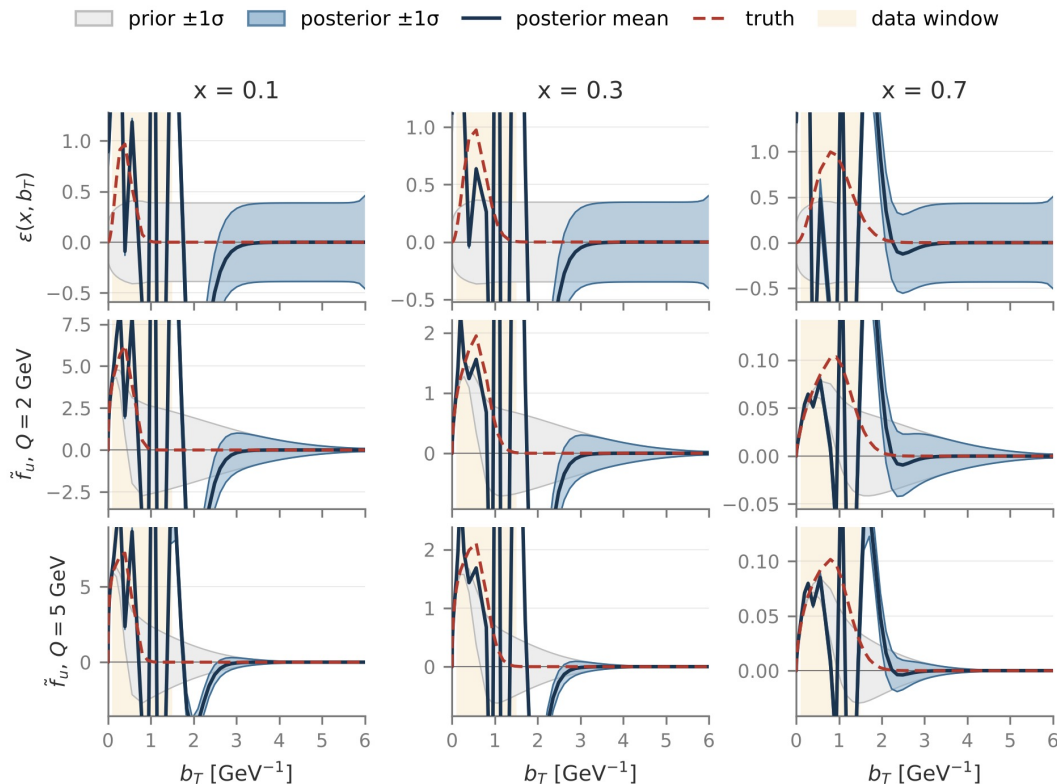
Prior: independent pixels, $\tau = 2$

Posterior

- **Small b_T :** ϵ switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE → **expected**
- **Window:** mean follows the data but wiggles from pixel to pixel → **bad**
- **Large b_T :** posterior = prior, very broad → **not constrained**
- **Why:** nothing ties neighbouring pixels, so noise in weakly seen patterns goes straight into ϵ

Fix: penalise roughness in the prior → smooth (Tikhonov) prior.

Results: Level 2



Event data, all 787 bins, linear theory

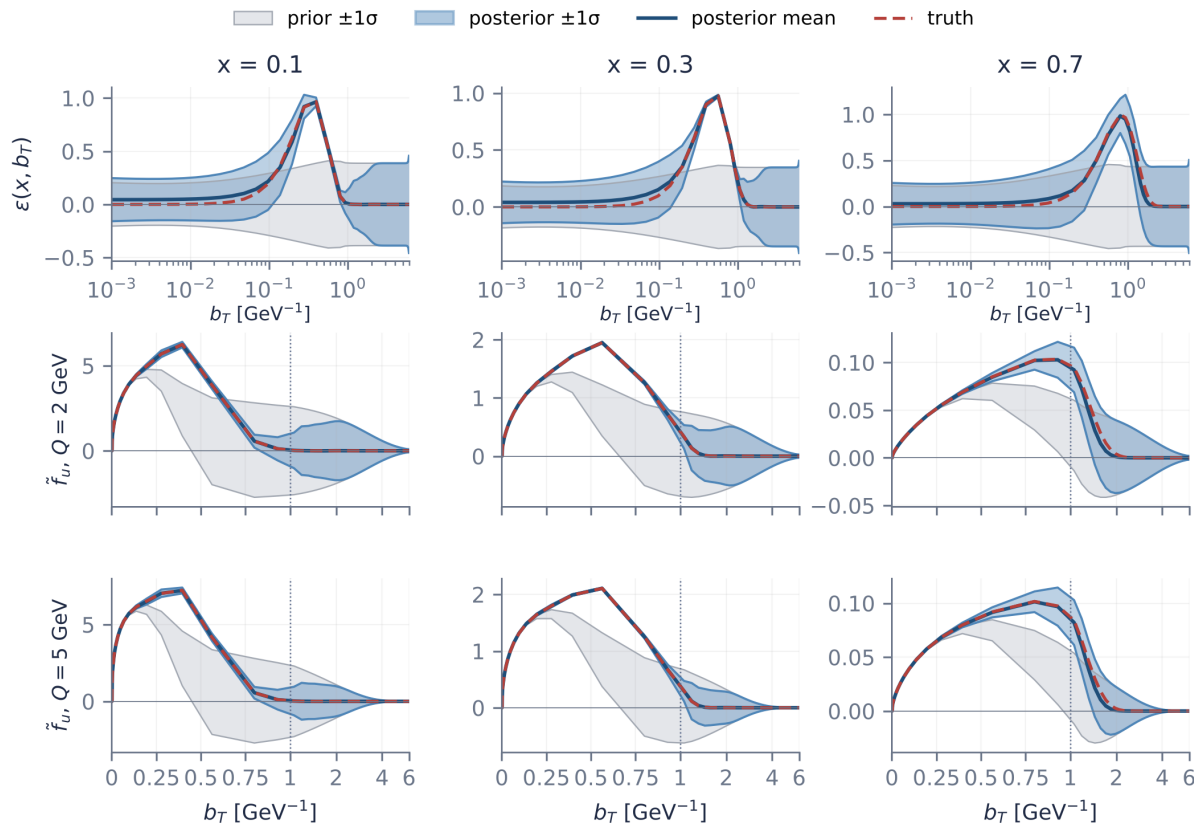
- **Events are physical:** $\sigma \geq 0$ everywhere
- **Linear theory is not:** $\sigma < 0$ at high P_{hT} , even at the truth

Posterior

- **Small b_T :** ϵ switched off by design ($1 - R_f \approx 0$): posterior = prior, TMD = OPE → **expected**
- **Window:** mean oscillates far from the truth; narrow band around a wrong answer, $\chi^2/n \approx 370$ → **bad**
- **Large b_T :** distorted up to $b_T \approx 2.5$, then prior → **bad**

Not a failure of the method: data and theory disagree where the linear theory is unphysical.

Closure test: Level 0.2



Smooth prior

same width as before ($\tau = 2$); neighbouring pixels tied in b_T and x ($\lambda = 30$).

Window $b_T \approx 0.1\text{--}1.5$

ε and TMD recovered; band $\approx 5\times$ narrower than with independent pixels.

Outside the window

data silent: posterior = prior (small b_T : ε switched off; large b_T : FF suppression).

Honest and narrow. λ is set by the data (window smoothness); τ only sets the prior where the data are silent. Used from here on.