

T3Net: Simultaneous Extraction of the Non-Singlet PDF $T_3(x)$ and a Wilson Coefficient C

Lucas Curtin

6592 words

July 1, 2025

Abstract

Precision predictions for hadron colliders require unbiased parton distribution functions (PDFs) and a clear separation between proton structure and genuine beyond-Standard-Model (BSM) effects. The NNPDF framework eliminates functional form bias [1] by fitting neural networks to Monte Carlo replicas of wide-ranging datasets [2, 3]. SIMUnet incorporates theory parameters directly into the fit to prevent BSM signals from being absorbed into the PDFs [4]. A complementary Bayesian strategy was recently proposed by Candido *et al.* [5], employing Gaussian processes as flexible priors and performing full inference over PDF parameters and hyperparameters. Building on these approaches, T3Net replaces the Gaussian process prior with a neural network and similarly focuses on the non-singlet combination

$$T_3(x) = u^+(x) - d^+(x),$$

extracted from the proton-deuteron structure-function difference. Pseudo-data generated with realistic experimental uncertainties undergo closure tests to confirm that the fit faithfully reproduces the reference distribution within its uncertainty band. However, introducing a single additional theory parameter to model generic BSM distortions, in turn, reveals a bias-variance trade-off. PDF uncertainties contract along with the amount of correct coverage, and the BSM parameter itself is systematically underestimated. These artifacts can be traced to the arguably naive construction of T3Net, which focused too heavily on simplicity. This minimal setting exposes potential pitfalls of simultaneous fits and suggests targeted refinements to improve the performance of these simpler validation models.

Table of Contents

1	Introduction	3
2	Theoretical Background	4
2.1	Parton distribution functions (PDFs)	4
2.2	Collinear factorisation	5
2.3	DGLAP evolution	7
2.4	Integral constraints (sum rules)	8
3	Standard PDF Determination Frameworks	9
3.1	NNPDF	10
3.2	Beyond-Standard-Model effects in hadronic processes	11
3.3	SIMUnet	13
4	Simplified Toy Model	13
4.1	Proton and deuteron structure functions	14
5	Data Preparation and Fast-Kernel Convolution	16
5.1	Assembling the non-singlet dataset	16
5.2	Fast-kernel convolution matrix	16
6	Neural Network Model for T_3	17
6.1	Replica-based training framework	17
6.2	Real vs pseudo-data comparison	19
7	Preliminary Results	20
8	Extension to Beyond-Standard-Model Corrections	21
8.1	Choice of ansatz functions	21
8.2	Goals and theoretical expectations	22
9	Closure Tests	22
9.1	Recovery of the Wilson coefficient	23
9.2	Endpoint exponent stability	24
9.3	Fixed vs. simultaneous fit: Impact on PDF Reconstruction	24
9.4	Fixed vs. simultaneous fit: Wilson Coefficient Extraction	26
10	Discussion and Conclusion	27
10.1	Potential model issues	27
10.2	Comparison with SIMUnet and methodological lessons	28
10.3	Future research directions	29

1 Introduction

Parton distribution functions (PDFs) encode the probability of finding a quark or gluon carrying a fraction x of the proton momentum and form a critical input for precision collider phenomenology. Conventional PDF determinations impose fixed analytic ansatzes which can introduce functional form bias, a systematic error when you assume a particular parameter shape for a model [1]. This can lead to underestimated uncertainties in regions not directly constrained by the data. The NNPDF methodology avoids this limitation by representing each PDF flavour with a neural network trained on Monte Carlo replicas of diverse measurements such as: deep-inelastic scattering, Drell-Yan, jets, electroweak bosons. This methodology is capable of yielding a flexible shape basis and robust uncertainty bands [2, 3]. Concurrently, searches for physics beyond the Standard Model (BSM) increasingly use the Standard Model Effective Field Theory (SMEFT). This involves encoding potential deviations from the Standard Model with Wilson coefficients that in turn multiply higher-dimensional operators. A traditional two step procedure would consist of first fitting PDFs under pure QCD (quantum chromodynamics), then extracting SMEFT coefficients from any residual deviations. This, however, risks absorbing genuine BSM effects into the fitted PDFs, manifesting as a PDF-BSM degeneracy [6]. To address this, SIMUnet extends the PDF fit to include theory parameters directly, enabling simultaneous inference of PDFs and BSM couplings as to try and avoid this signal absorption [4]. Full implementation of simultaneous PDF-SMEFT fits in a global analysis is difficult as it involves: multiple flavors, numerous operators, and diverse datasets. Inspired by the Bayesian inference framework of Candido *et al.* [5], which uses Gaussian processes (GPs) as flexible priors and performs joint parameter-hyperparameter sampling, T3Net adopts a streamlined test case to isolate key behaviours. Rather than a GP prior, T3Net employs a neural network to model the same non-singlet distribution

$$T_3(x) = u^+(x) - d^+(x)$$

This serves as an ideal probe for simultaneous PDF-BSM inference, since it obeys a single DGLAP evolution equation (Sec. 2.3) and maps directly onto the proton-deuteron structure function difference (Sec. 4.1). The central objective of this study is to determine how flexible parametrisations and theoretical constraints affect the ability to detect small BSM induced distortions. To this end, realistic pseudo-data was generated from a known reference distribution. A neural network captured deviations from small- x and large- x power-law behaviours and the valence sum rule was enforced via penalty terms. Closure tests demonstrated that, in the absence of BSM effects, the network reproduced the reference $T_3(x)$ within its uncertainty band. With confirmation that the model works in idealised situations, a single BSM theory parameter was then introduced to model generic distortions. This controlled extension revealed a subtle bias-variance trade off. Uncertainties on T_3 shrank and correct coverage degraded, while the additional parameter was systematically underestimated. This in turn allowed for a closer appraisal of the SIMUnet methodology and the benefits and pitfalls of simultaneous fits.

2 Theoretical Background

To better understand the work of this study, a theoretical explanation of the underlying physics is included in the follow section, largely guided by the helpful work of Devenish and Cooper-Sarkar [7]. Additionally, a glossary is included in Appendix 10.3.

2.1 Parton distribution functions (PDFs)

Parton distribution functions (PDFs) describe how the proton's momentum is shared among its constituents, quarks and gluons. For each parton flavour i , the PDF

$$f_i(x, Q^2)$$

gives the probability density that a parton of type i carries a fraction x of the proton's momentum, as probed at a resolution (energy) scale Q^2 .

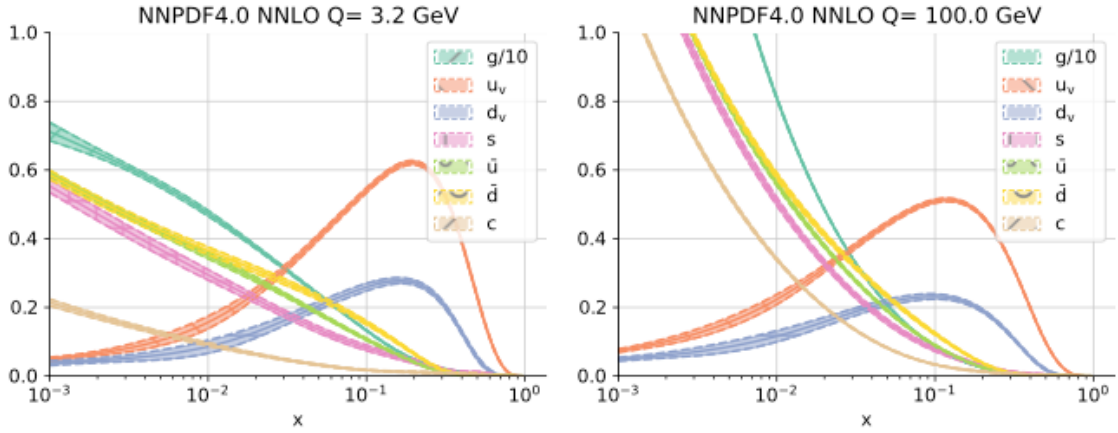


Figure 1: Parton distribution functions from the NNPDF4.0 global fit at next-to-next-to-leading order (NNLO), shown at two scales $Q = 3.2$ GeV (left) and $Q = 100$ GeV (right) [3]. The curves indicate the likelihood of finding each parton type with momentum fraction x .

Key variables

- **Momentum fraction** x also called the Bjorken variable, $x \in [0, 1]$ represents the share of the proton's momentum carried by a single parton in the direction of motion.
- **Resolution scale** Q^2 determined by the squared momentum transfer in a high energy collision, Q^2 acts like the magnification of a microscope. Larger Q^2 resolves finer structure inside the proton.

Figure 1 highlights several general features of proton structure:

-
- **Valence distributions:** The proton's core consists of two up-valence quarks and one down-valence quark. Their distributions,

$$u_v(x) = u(x) - \bar{u}(x) \quad d_v(x) = d(x) - \bar{d}(x)$$

peak around $x \approx 0.2$, indicating that each valence quark typically carries about one fifth of the proton's momentum.

- **Sea quarks and gluons:** Quantum fluctuations produce a 'sea' of quark-antiquark pairs (for example $\bar{u}, \bar{d}, s, \bar{s}$) and many gluons. These appear mainly at low x , since gluon splitting and soft gluon emission are more frequent at small momentum fractions.
- **Endpoint suppression:** All PDFs vanish as $x \rightarrow 1$. It is therefore unlikely for a single parton to carry nearly all of the proton's momentum, as the remainder must be shared among other constituents.
- **Scale dependence:** Raising the resolution scale Q^2 tends to shift distribution strength from larger to smaller x , revealing additional quark-gluon activity. The precise form of this evolution is governed by perturbative QCD and is discussed in Sec. 2.3.

2.2 Collinear factorisation

In high energy collisions of hadrons A and B , the total probability for producing a final state X can be expressed as a product of two contributions: how quarks and gluons are distributed inside each hadron and how those quarks and gluons interact to form X . This concept, known as collinear factorisation, is written as

$$\sigma_{AB \rightarrow X} = \sum_{i,j} \int_0^1 dx_1 dx_2 f_i^A(x_1, Q^2) f_j^B(x_2, Q^2) \hat{\sigma}_{ij \rightarrow X}(x_1, x_2, Q^2) \quad (1)$$

The terms in this equation are:

- $f_i^A(x, Q^2)$ is the parton distribution function giving the probability density for finding a parton of type i inside hadron A , carrying fraction x of its momentum at scale Q . These functions capture all long distance physics that cannot be calculated using perturbation theory and must be determined from experimental measurements.
- The integrals over x_1 and x_2 sum over every possible way of sharing the two hadrons' momenta between their partons.
- $\hat{\sigma}_{ij \rightarrow X}(x_1, x_2, Q^2)$ is the partonic cross section for scattering partons i and j into final state X at scale Q . It is computed in quantum chromodynamics as an expansion in the strong coupling constant

$$\alpha_s(Q^2) = \frac{g_s^2(Q)}{4\pi},$$

which sets the strength of quark-gluon interactions. Quantum effects cause α_s to decrease at larger Q , so that when Q is high the series

$$\hat{\sigma} = \hat{\sigma}^{(0)} + \alpha_s \hat{\sigma}^{(1)} + \alpha_s^2 \hat{\sigma}^{(2)} + \dots$$

converges. The term $\hat{\sigma}^{(0)}$ is called leading order and includes only the simplest diagrams without extra gluons or loops. The term multiplied by α_s is the next order contribution, incorporating one additional gluon emission or single loop virtual effects. Higher powers of α_s are successive higher order corrections, each improving accuracy at the cost of more complex calculations.

- The sum over i and j runs over all quark, antiquark and gluon species in the initial state.

The true cross section $\sigma_{AB \rightarrow X}$ does not depend on the arbitrary scale Q , but truncating the series at finite order leaves a small residual dependence. Choosing Q to match the characteristic momentum transfer of the hard scattering keeps $\alpha_s(Q)$ small and minimises the effect of omitted higher order corrections. Collinear factorisation thus separates long distance physics contained in universal PDFs from short distance scattering, computed in perturbation theory to the desired order in α_s . This separation allows the same PDFs to predict a wide range of processes simply by supplying the appropriate partonic cross sections.

Example: Drell-Yan lepton-pair production

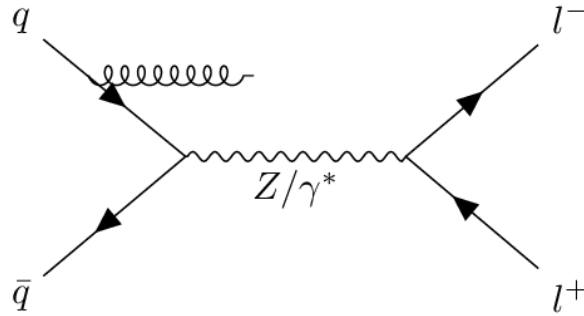


Figure 2: Leading order Drell-Yan process $q\bar{q} \rightarrow \gamma^*/Z \rightarrow \ell^+\ell^-$ [8].

As an illustration of collinear factorisation, consider the production of a charged lepton pair

$$p + p \rightarrow \ell^+\ell^- + X$$

which at lowest order proceeds via quark-antiquark annihilation into a virtual photon or Z boson:

$$q(x_1P) + \bar{q}(x_2P) \rightarrow \gamma^*/Z \rightarrow \ell^+\ell^-$$

Here x_1 and x_2 are the momentum fractions carried by the annihilating partons and P is the proton momentum. Figure 2 shows the corresponding Feynman diagram. Inserting this subprocess into the factorisation formula (Eq. 1)

$$\sigma_{pp \rightarrow \ell^+ \ell^-} = \sum_{i,j} \int_0^1 dx_1 dx_2 f_i(x_1, Q^2) f_j(x_2, Q^2) \hat{\sigma}_{ij \rightarrow \ell^+ \ell^-}(x_1, x_2, Q^2)$$

with $Q^2 = M^2$ (the lepton-pair invariant mass squared) demonstrates three key points:

1. $f_q(x_1, Q^2)$ and $f_{\bar{q}}(x_2, Q^2)$ encode the non-perturbative likelihood of finding the annihilating partons in each proton at those momentum fractions.
2. $\hat{\sigma}_{q\bar{q} \rightarrow \ell^+ \ell^-}(\hat{s}, Q^2)$ captures all the short distance dynamics of the electroweak subprocess. Once computed, for example at leading or next-to-leading order in the electroweak and strong couplings, it can then be reused unchanged.
3. The convolution integrals and sum over quark flavours assemble contributions from every possible initial state configuration, yielding a prediction for both the total rate and differential distributions in M or rapidity. Since the same PDFs enter every process, their experimental determination (and associated uncertainties) propagates directly into the Drell-Yan prediction without altering the underlying factorisation framework.

2.3 DGLAP evolution

The PDFs $f_i(x, Q^2)$ evolve with the resolution scale Q^2 because a higher energy probe can resolve additional emissions and splittings of quarks and gluons inside the proton. This evolution is described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations:

$$\frac{\partial f_i(x, Q^2)}{\partial \ln Q^2} = \sum_j \int_x^1 \frac{dz}{z} P_{ij}(z, \alpha_s(Q^2)) f_j\left(\frac{x}{z}, Q^2\right) \quad (2)$$

Here:

- $\partial f_i / \partial \ln Q^2$ is the rate at which the probability of finding parton i with momentum fraction x changes as the resolution Q^2 varies.
- The variable z (with $x \leq z \leq 1$) represents the fraction of the parent parton's momentum retained after it splits. A parent of flavour j carrying fraction x/z can emit radiation (for example $q \rightarrow qg$ or $g \rightarrow gg$) and produce a daughter of flavour i carrying fraction x .
- The splitting function $P_{ij}(z, \alpha_s)$ gives the probability density for the transition $j \rightarrow i$ with momentum share z . Its form is calculated in perturbative QCD and depends on the strong coupling $\alpha_s(Q^2)$.
- The convolution integral over z sums all possible parent momenta that can lead to a daughter with momentum x .

Intuitive picture of parton evolution

- At low Q^2 , the probe mainly ‘sees’ the three valence quarks in the proton and a few gluons.
- Increasing Q^2 is like increasing the magnification of a microscope. Each quark can radiate gluons and gluons can split into further quark-antiquark pairs or into two gluons.
- This branching process creates a cascade or ‘shower’ of partons, predominantly at smaller momentum fractions x .
- As a result, the probability to find many low x partons grows, while the high x valence peaks broaden and decline in amplitude.

Solving the DGLAP equations step by step in $\ln Q^2$ yields the full set of PDFs at any desired resolution, given an initial distribution at some starting scale Q_0^2 .

2.4 Integral constraints (sum rules)

The shape of each PDF $f_i(x, Q^2)$ is constrained not only by data but also by fundamental conservation laws. These laws take the form of exact integral relations, or sum rules, which must hold at every scale Q^2 . Imposing them ensures that the fitted PDFs respect the total momentum and quantum numbers of the proton. This will be particularly important later on to ensure that the outputs of the model are physically consistent.

Momentum sum rule

Conservation of four-momentum requires that all partons together carry the proton’s entire momentum. Mathematically,

$$\sum_i \int_0^1 x f_i(x, Q^2) dx = 1$$

Here $x f_i(x)$ is the momentum density of parton i . The integral over x sums its share of the proton momentum and the sum over i runs through all quark, antiquark and gluon flavours. If this rule were violated, the fit would either lose momentum (sum < 1) or create momentum (sum > 1) out of nothing.

Valence-quark number sum rules

The proton’s baryon number and flavour content fix the net number of valence quarks (as opposed to sea quarks). Denoting quark and antiquark distributions by $q_i(x)$ and $\bar{q}_i(x)$. The net valence number of flavour i is

$$\int_0^1 [q_i(x, Q^2) - \bar{q}_i(x, Q^2)] dx = N_i$$

where N_i is known ($N_u = 2$, $N_d = 1$ for the proton). This condition guarantees that the fitting procedure cannot invent or destroy valence quarks i.e. the area under the difference curve should always equal the correct amount.

Additional flavour constraints

In many analyses, further assumptions impose equalities or vanishing integrals for heavier flavours. For example:

- $\int_0^1 [s(x) - \bar{s}(x)] dx = 0$, enforcing a symmetric strange sea.
- Vanishing charm content below its production threshold.

Such constraints reflect theoretical expectations or experimental findings about the proton's sea-quark composition. Since $T_3(x) = u^+(x) - d^+(x)$ contains no strange, charm or heavier flavour terms, sea-symmetry and threshold constraints on s, \bar{s}, c, \dots drop out and can be omitted in the non-singlet fit.

Role in PDF fitting

Applying sum rules has two main benefits:

1. *Physical consistency.* Ensures that the fitted PDFs obey exact conservation laws, preventing unphysical distortions.
2. *Improved stability.* Reduces fit uncertainties by removing degrees of freedom that would otherwise be unconstrained at the kinematic extremes ($x \rightarrow 0$ or 1).

3 Standard PDF Determination Frameworks

Global QCD analyses extract PDFs by fitting a broad array of experimental measurements, exploiting the factorisation theorem (Eq. 1), DGLAP evolution (Eq. 2), and sum rules (Sec. 2.4) to ensure theoretical consistency. Typical inputs include:

- **Deep-inelastic scattering (DIS):** fixed target experiments (BCDMS, SLAC, NMC) and HERA which constrain quark and gluon distributions over a wide range of x and Q^2 .
- **Drell-Yan lepton pairs:** pp and $p\bar{p}$ collider data from the LHC, sensitive to flavour separation via W^\pm and Z rapidity distributions.
- **Inclusive jets and high p_T processes:** LHC jet and photon production, which provide complementary constraints on the gluon PDF at medium and large x .

- **Electroweak boson and top-quark production:** precision W, Z and $t\bar{t}$ cross sections, reducing flavour and scale uncertainties in the global fit.

3.1 NNPDF

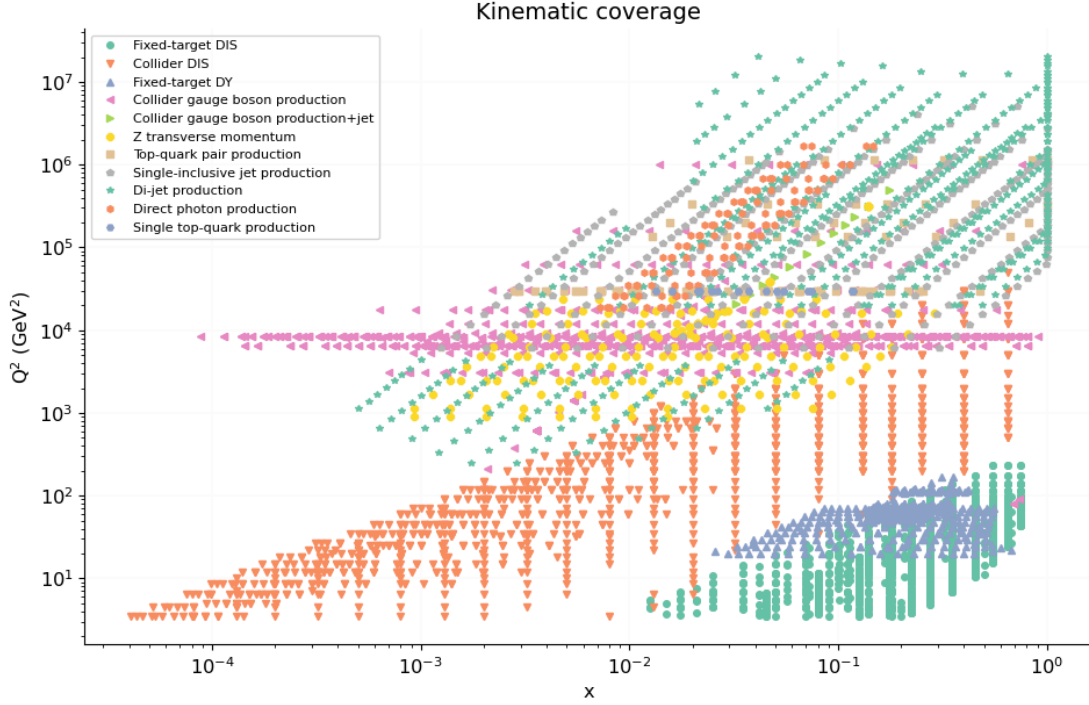


Figure 3: Kinematic coverage of the datasets entering the NNPDF4.0 global fit, plotted in the (x, Q^2) plane. Fixed target DIS (green circles) and collider DIS (orange triangles) span low to moderate Q^2 and x , while Drell-Yan, gauge boson, jet, photon, and top quark measurements extend the reach to higher scales. New data included in NNPDF4.0 are outlined in black. This comprehensive coverage ensures that PDFs are constrained across the majority of the kinematic region relevant for LHC phenomenology [3].

In the NNPDF approach [2, 3], bias from fixed analytic forms is avoided by representing each PDF $f_i(x, Q_0^2)$ at the input scale Q_0^2 with a neural network. Figure 3 illustrates how different processes probe distinct regions of Bjorken x and scale Q^2 , which is essential for disentangling valence, sea and gluon components. This network is sufficiently flexible to capture the true underlying shape, subject only to the theoretical constraints of factorisation, evolution, and sum rules. The key steps are:

1. **Monte Carlo Replica Generation:** Starting from the experimental central values y_k and their covariance matrix C_{kk} , an ensemble of pseudo-datasets is created by sampling each data point according to its uncertainties and correlations. This produces N_{rep} statistically independent replicas that reflect the full distribution of possible measurements.

-
2. **Neural-Network Training:** For each replica r , the neural network parameters are adjusted to minimise a loss function. This is typically the χ^2 which essentially measures the difference between the theory predictions $T_k^{(r)}$ and the pseudo-data $y_k^{(r)}$.
 3. **Ensemble Construction:** Once all replicas have been fitted, the resulting collection of neural network outputs at any (x, Q^2) point constitutes a Monte Carlo ensemble. The PDF central value is the ensemble mean and the uncertainty is given by the standard deviation across replicas. This procedure automatically propagates experimental uncertainties, correlations and methodological variability into the final PDF errors.
 4. **Validation and Closure Tests:** To verify that the neural parametrisation and training procedure do not introduce artifacts, extensive validation is performed. Fits to generated data with known answers ('closure tests') and cross-validation on withheld subsets ensure that the networks generalise correctly and that uncertainties are neither underestimated nor inflated.

3.2 Beyond-Standard-Model effects in hadronic processes

Heavy particles with mass around a scale Λ cannot be produced when the collision energy E satisfies $E \ll \Lambda$. Their presence is still felt through small corrections to known interactions, which can be summarised by an *effective Lagrangian*:

$$\mathcal{L}_{\text{eff}} = \underbrace{\mathcal{L}_{\text{SM}}}_{\text{all Standard Model forces}} + \sum_a \frac{C_a}{\Lambda^2} \underbrace{\mathcal{O}_a^{(6)}}_{\text{new contact interaction}}$$

where:

\mathcal{L}_{SM} contains the familiar electromagnetic, weak and strong interactions among quarks, gluons, leptons, and the Higgs field.

$\mathcal{O}_a^{(6)}$ is a *dimension-six operator* built from six powers of fields and derivatives. Such operators encode the simplest ways in which heavy particles can produce point-like interactions among Standard Model fields.

Λ is the *heavy physics scale*, the typical mass of the new particles. Below this scale, those particles do not appear explicitly, but their effects survive through the operators $\mathcal{O}_a^{(6)}$.

C_a are *Wilson coefficients*: plain numbers that set the strength and sign of each new interaction. A value $C_a = 0$ means no contribution from that operator whereas a non-zero C_a signals new physics.

$\frac{1}{\Lambda^2}$ suppresses each correction by the square of the heavy physics scale.

In practical terms, measured scattering rates at various energies are compared to the Standard Model prediction. Any consistent deviation can be described by a non-zero ratio C_a/Λ^2 , revealing the presence and approximate strength of heavy new particles without producing them directly.

Four-Quark Contact Interaction: Interpretation and Cross-Section Example

The operator

$$\mathcal{O}_q = (\bar{q} \gamma^\mu q) (\bar{q} \gamma_\mu q)$$

describes a direct, point-like interaction of four quarks. It can be unpacked as follows:

1. **Quark fields:** $q(x)$ denotes a quark field at space-time point x , $\bar{q}(x)$ is its Dirac adjoint (often called the ‘antiquark’ field).
2. **Quark current:** γ^μ ($\mu = 0, 1, 2, 3$) are matrices encoding the spin structure of a spin- $1/2$ particle. The combination $\bar{q} \gamma^\mu q$ is a *current*, representing the flow of quark quantum numbers (e.g. electric charge) in direction μ .
3. **Contact interaction:** Multiplying two currents and summing over μ ,

$$(\bar{q} \gamma^\mu q) (\bar{q} \gamma_\mu q)$$

represents two quark currents interacting at a single point, i.e. no intermediate particle propagates between them.

When this operator appears in the effective Lagrangian with Wilson coefficient C_q and heavy physics scale Λ , it adds a correction to the partonic cross section of the form

$$\Delta\sigma_q(\hat{s}) = \kappa \frac{C_q}{\Lambda^2} \hat{s}$$

where:

- $\Delta\sigma_q(\hat{s})$ is the deviation from the Standard Model rate at partonic energy squared \hat{s} .
- κ is a factor from phase space integration and spin colour sums.
- The factor \hat{s} shows that the effect grows linearly with energy.

Numerical Illustration For $\Lambda = 10 \text{ TeV}$, $C_q = 1$ and $\kappa = 1$,

$$\frac{C_q}{\Lambda^2} = 10^{-2} \text{ TeV}^{-2}, \quad \sqrt{\hat{s}} = 1 \text{ TeV} \quad \implies \quad \frac{\Delta\sigma_q}{\sigma_{\text{SM}}} \approx 1\%$$

A one percent excess in four quark scattering at $\sqrt{\hat{s}} = 1 \text{ TeV}$ would therefore signal this interaction.

3.3 SIMUnet

Standard two step approaches (first determining PDFs with NNPDF, then fitting EFT coefficients on fixed PDFs) can suffer from a *PDF-BSM degeneracy*. This is when small BSM induced shifts in high energy cross sections can be absorbed into distortions of the fitted PDFs. This biases the extracted Wilson coefficients. SIMUnet avoids this bias by treating PDF shapes and EFT parameters on an equal footing:

- **Joint amplitude model:** Instead of a separate PDF fit, a single neural network is constructed whose outputs are

$$\{f_i(x, Q_0^2)\} \quad \text{and} \quad \{C_a/\Lambda^2\}$$

This ensures that any tension in the data can be shared optimally between changes in proton structure and genuine new physics effects.

- **Unified global fit:** All measurements (DIS, Drell-Yan, jets, electroweak bosons, top-quark pair production) enter one combined loss function, again typically a global χ^2 . The network minimises this over both the PDF parameters and the Wilson coefficients simultaneously, avoiding the artifact of sequential fitting.

By exploring the full multidimensional likelihood at once, SIMUnet automatically assigns genuine high energy discrepancies either to alterations in the parton distributions or to non-zero C_a/Λ^2 . This simultaneous inference delivers:

- *Unbiased Wilson-coefficient bounds* free from PDF-induced artifacts.
- *Accurate PDF uncertainties* that include correlations with new physics parameters.
- *A single consistent framework* for including theoretical systematics (scale variations, missing higher orders) alongside experimental errors.

4 Simplified Toy Model

Applying SIMUnet to a full global fit would require simultaneous determination of all quark, antiquark and gluon distributions together with multiple Wilson coefficients. This would be difficult to implement in an independent standalone model. Singlet and gluon distributions would mix under DGLAP evolution and a global dataset would have to span lots of kinematic regions (combinations of x and Q^2). Training hundreds of Monte Carlo replicas under these conditions would demand lots of computing time and could obscure the origins of any fitting issues. To retain the essential features of a simultaneous fit while minimising complexity, this toy model fits an individual non-singlet PDF

$$T_3(x, Q^2) = [u(x, Q^2) + \bar{u}(x, Q^2)] - [d(x, Q^2) + \bar{d}(x, Q^2)]$$

which by construction removes any flavour ‘blind’ component (the sea and the gluon) and therefore obeys a single DGLAP equation without mixing into other distributions. The following discussion goes through the theory behind how this PDF can be measured from DIS data, again supported by the work from [7].

4.1 Proton and deuteron structure functions F_2^p and F_2^d

To determine the non-singlet distribution $T_3(x)$ from experiment, measurements of the proton and deuteron structure functions are combined. Deep-inelastic scattering (DIS) data can yield the structure functions $F_2^p(x, Q^2)$ and $F_2^d(x, Q^2)$. By taking their difference isolates the up-down quark excess and thus directly probes $x T_3(x)$.

Theory of DIS structure functions

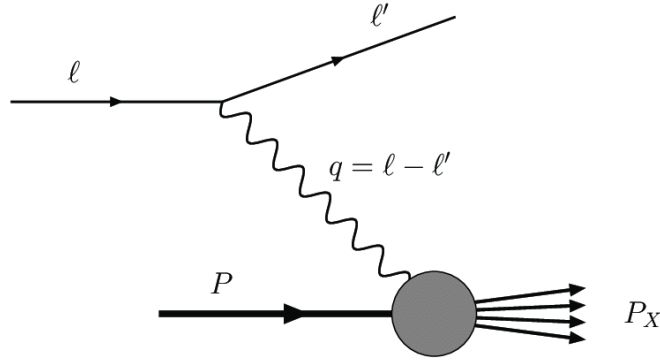


Figure 4: Schematic of deep-inelastic scattering (DIS): a lepton ℓ exchanges a virtual photon of momentum q with a hadronic target of momentum P , producing a scattered lepton ℓ' and hadronic debris P_X [9].

Deep-inelastic scattering probes the partonic structure of a nucleon by firing a high energy lepton of four-momentum k at a target of four-momentum P . The lepton exchanges a photon with momentum q (so that $Q^2 \equiv -q^2 > 0$) and the target fragments into hadronic debris P_X . Key kinematic variables are constructed from the variables defined in Sec. 2.1:

$$Q^2 = -q^2, \quad x = \frac{Q^2}{2 P \cdot q}, \quad y = \frac{P \cdot q}{P \cdot k}$$

Where y is the ‘inelasticity’ of the interaction and captures the fractional energy loss of the lepton. The experimentally measured cross section is

$$\frac{d^2\sigma}{dx dQ^2} = \frac{2\pi\alpha^2}{x Q^4} \left[(1 + (1 - y)^2) F_2(x, Q^2) - y^2 F_L(x, Q^2) \right] + \mathcal{O}(\text{EW interference})$$

- α is the electromagnetic coupling,

-
- $F_2(x, Q^2)$ is the dominant structure function, sensitive to the sum of quark plus antiquark PDFs,
 - $F_L(x, Q^2)$ is the longitudinal structure function (suppressed except at high y),
 - the factor $1+(1-y)^2$ arises from transverse photon exchange and y^2 from longitudinal exchange.

At leading order in QCD:

$$F_2(x, Q^2) = x \sum_i e_i^2 [q_i(x, Q^2) + \bar{q}_i(x, Q^2)]$$

with e_i the parton's electric charge ($+2/3$ for u , $-1/3$ for d , etc.). Thus F_2 measures a charge-weighted sum of quark and antiquark momentum distributions. The deuteron per nucleon structure function satisfies

$$F_2^d(x, Q^2) = \frac{1}{2} [F_2^p(x, Q^2) + F_2^n(x, Q^2)], \quad F_2^n \text{ obtained from } F_2^p \text{ by } u \leftrightarrow d.$$

Define

$$\Delta F_2(x, Q^2) \equiv F_2^p(x, Q^2) - F_2^d(x, Q^2).$$

Then

$$\begin{aligned} \Delta F_2 &= F_2^p - \frac{1}{2}(F_2^p + F_2^n) = \frac{1}{2}(F_2^p - F_2^n) \\ &= \frac{1}{2} x \left[e_u^2 (u^+ - d^+) + e_d^2 (d^+ - u^+) \right] = \frac{1}{2} (e_u^2 - e_d^2) x T_3(x, Q^2) \\ &= \frac{1}{2} \cdot \frac{1}{3} x T_3(x, Q^2) = \frac{1}{6} x T_3(x, Q^2), \end{aligned}$$

where

$$u^+(x) \equiv u(x) + \bar{u}(x) \quad d^+(x) \equiv d(x) + \bar{d}(x) \quad T_3(x) \equiv u^+(x) - d^+(x)$$

Hence

$$x T_3(x, Q^2) = 6 [F_2^p(x, Q^2) - F_2^d(x, Q^2)].$$

5 Data Preparation and Fast-Kernel Convolution

The extraction of the non-singlet combination

$$\Delta F_2(x, Q^2) = F_2^p(x, Q^2) - F_2^d(x, Q^2)$$

proceeds in three stages: assembling a fully correlated proton-deuteron dataset, constructing a fast-kernel convolution matrix, and generating pseudo-data via a correlated closure test.

5.1 Assembling the non-singlet dataset

Proton and deuteron structure function measurements were obtained from the ValidPhys API [10], which provides the legacy BCDMS datasets for F_2^p and F_2^d with internal kinematic cuts applied. The proton data came from the BCDMS measurement of F_2^p in [11] and the deuteron counterpart uses the same collaboration's deuteron analysis. Matching points (x_i, Q_i^2) were identified and merged so that

$$y_i^{\text{data}} = F_2^p(x_i, Q_i^2) - F_2^d(x_i, Q_i^2), \quad i = 1, \dots, N_{\text{data}}$$

Statistical and systematic uncertainties, along with their correlations, were retrieved in their covariance matrices

$$C_{pp}, C_{dd}, C_{pd}$$

and combined into

$$C_{yy} = C_{pp} + C_{dd} - 2C_{pd}$$

A diagonal regulator $\epsilon \times \text{diag}(C_{yy})$ with $\epsilon \sim 10^{-6}$ - 10^{-4} guarantees positive definiteness and stable inversion for uncertainty propagation later on.

5.2 Fast-kernel convolution matrix

A traditional deep-inelastic scattering fit requires recalculating the convolution of parton distributions with perturbative coefficient functions and DGLAP evolution at every iteration. Fast-kernel (FK) tables instead encode all of these operations into fixed weights

$$K_{i\alpha f}^{(h)}, \quad h \in \{p, d\}, \quad i = 1, \dots, N_{\text{data}}, \quad \alpha = 1, \dots, N_{\text{grid}}, \quad f \in \{\text{flavours}\}$$

Each table entry multiplies the input scale momentum distribution $x T_3(x_\alpha, Q_0^2)$ rather than the bare PDF, so that the predicted structure-function difference becomes

$$\Delta F_2(x_i, Q_i^2) = \sum_{f, \alpha} [K_{i\alpha f}^p - K_{i\alpha f}^d] \kappa_f [x_\alpha T_3(x_\alpha, Q_0^2)]$$

Here

$$\kappa_f = \begin{cases} +1, & f = u, \bar{u}, \\ -1, & f = d, \bar{d}, \\ 0, & \text{otherwise,} \end{cases} \quad i = 1, \dots, N_{\text{data}}, \quad \alpha = 1, \dots, N_{\text{grid}}$$

Defining the combined FK matrix

$$W_{i\alpha} = \sum_f [K_{i\alpha f}^p - K_{i\alpha f}^d] \kappa_f$$

and the discrete input vector $\{x_\alpha T_3(x_\alpha, Q_0^2)\}_{\alpha=1}^{N_{\text{grid}}}$, the theory prediction reduces to the single matrix multiplication

$$\mathbf{y}^{\text{theory}} = W \mathbf{xT}_3, \quad (\mathbf{xT}_3)_\alpha = x_\alpha T_3(x_\alpha, Q_0^2)$$

Thus, FK tables combine and weight the input momentum distribution across all relevant flavors and kinematic points to yield the theoretical expectation for the structure-function difference in a single and highly efficient calculation .

6 Neural Network Model for T_3

A neural network is then used to infer the non-singlet distribution $T_3(x)$ directly from the pseudo-data for $F_2^p - F_2^d$. The parametrisation combines explicit theoretical expectations at the kinematic endpoints with a flexible interpolating function in the interior.

$$T_3(x) = \frac{f_{\text{pred}}(x)}{x}, \quad f_{\text{pred}}(x) = x^\alpha (1-x)^\beta \Phi(x; \theta)$$

where the factors x^α and $(1-x)^\beta$ enforce the known small- x rise and large- x suppression of $xT_3(x)$. The remainder $\Phi(x; \theta)$ is required to be strictly positive and captures any deviations from these simple power laws. Trainable exponents α and β allow the data to determine the precise asymptotic slopes rather than fixing them by hand, while the neural network parameters θ encode local variations of the distribution. Predicting xT_3 rather than just T_3 is just a convention due to the inputs from ValidPhys, and are referred to each-other interchangeably despite the former technically being a momentum distribution. The fitting of the momentum distribution is hereafter often described in terms of the underlying PDF $T_3(x)$ as the fundamental basis i.e. the quality of a PDF fit to distinguish from say the quality of the extraction of the Wilson coefficient. This is to ensure the clear distinction of what is being fed into the model, whilst also being able to use SIMUnet familiar terminology.

6.1 Replica-based training framework

To map experimental uncertainties into PDF uncertainties and to validate that the procedure is unbiased, training proceeds in four conceptual stages:

1. Generation of pseudo-data replicas

An ensemble of N_{rep} alternative datasets is created by perturbing the central theory prediction with correlated fluctuations.

$$y^{(r)} = y_{\text{theory}} + \eta^{(r)}, \quad \eta^{(r)} \sim \mathcal{N}(0, C_{yy})$$

Here $y_{\text{theory}} = W x T_3^{\text{ref}}$ is computed from a reference PDF and C_{yy} is the full data covariance. Each replica thus provides a plausible realisation of the experiment, including both statistical and systematic errors.

2. Choice of loss function

For each replica, the network prediction $y_{\text{model}} = W x T_3$ is compared to $y^{(r)}$ via

$$\chi^2 = (y_{\text{model}} - y^{(r)})^T C_{yy}^{-1} (y_{\text{model}} - y^{(r)})$$

To enforce the valence sum rule on the restricted $x \in [x_{\min}, x_{\max}]$ grid, a penalty term

$$\lambda_{\text{sum}} (I_{\text{fit}} - I_{\text{ref}})^2$$

is added where

$$I_{\text{fit}} = \int_{x_{\min}}^{x_{\max}} T_3(x) dx, \quad I_{\text{ref}} = \int_{x_{\min}}^{x_{\max}} T_3^{\text{ref}}(x) dx$$

The total loss is

$$L = \chi^2 + \lambda_{\text{sum}} (I_{\text{fit}} - I_{\text{ref}})^2$$

with λ_{sum} being a hyperparameter that balances how strictly the baryon number is conserved compared to the overall fit quality of $x T_3$.

3. Independent replica fits and monitoring of χ_{red}^2

Each pseudo-dataset is fitted independently by minimising L . The reduced chi squared,

$$\chi_{\text{red}}^2 = \chi^2 / N_{\text{data}}$$

is monitored for each fit and is expected to lie near unity if the model describes the imposed fluctuations within their uncertainties. Fits with χ_{red}^2 significantly different from 1 indicate either numerical instability or over/under-fitting and are excluded from the ensemble.

4. Ensemble averages and uncertainty estimation

The collection of fitted distributions $\{T_3^{(r)}(x)\}$ forms a Monte Carlo ensemble. At each x , the ensemble mean

$$\langle T_3(x) \rangle = \frac{1}{N_{\text{rep}}} \sum_{r=1}^{N_{\text{rep}}} T_3^{(r)}(x)$$

provides the central value, while the standard deviation across replicas quantifies the propagated experimental uncertainty.

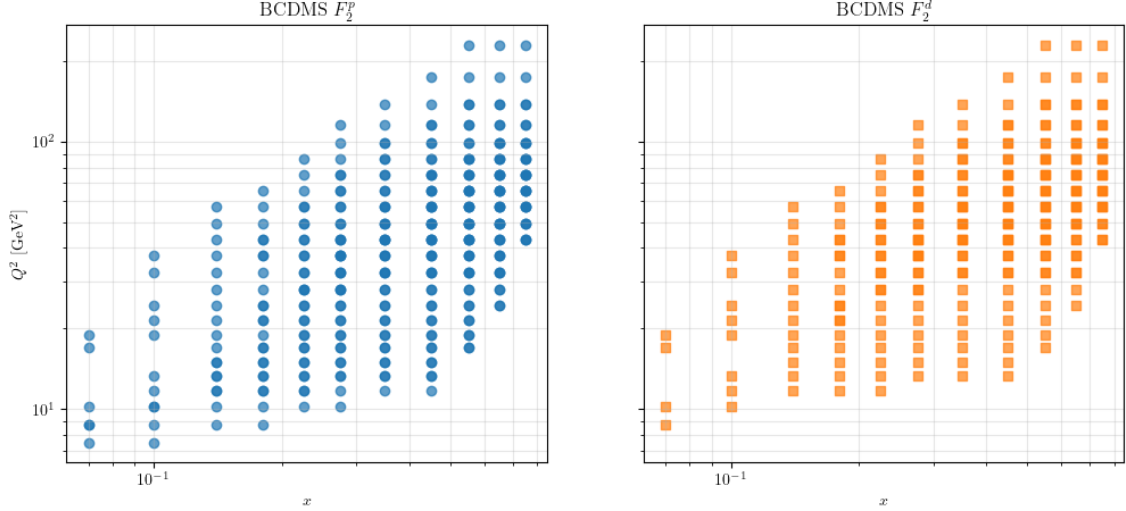


Figure 5: Kinematic coverage of the BCDMS proton (\circ) and deuteron (\square) measurements in the (x, Q^2) plane.

6.2 Real vs pseudo-data comparison

Figure 5 shows the kinematic reach of the BCDMS proton in the (x, Q^2) plane. The combined dataset covers lots of distinct regions for both x and Q^2 but does not extend to the full range $x \in [0, 1]$. This limited coverage will complicate any attempt to pin down endpoint behaviour of the PDF. Figure 6 compares the theory prediction y_{theory} with both a single pseudo-data replica and the actual BCDMS measurements of $F_2^p - F_2^d$. In contrast, the real data exhibit systematic deviations. These arise because the pseudo-data include only the experimental uncertainties encoded in C_{yy} , whereas the real measurements are also affected by nuclear effects described in [12]:

- **Fermi motion and binding:** The momentum distribution of bound nucleons in the deuteron, described by the wave function $\Psi_d(\mathbf{p})$, smears the structure functions via a convolution over nucleon momentum.
- **Off-shell correction:** Bound nucleons are slightly off their mass shell ($p^2 \neq M^2$), inducing a relative modification $\delta f(x)v$ of the nucleon structure functions, where $v = (p^2 - M^2)/M^2$.
- **Meson-exchange currents and shadowing:** At $x \lesssim 0.1$, two-nucleon exchange currents and coherent (shadowing) effects become non-negligible.

In a more realistic fit these nuclear corrections could be implemented, thereby reducing the residual discrepancy between real and pseudo-data. However, for the purposes of

testing the core methodologies, the majority of the analysis is focused on the generated pseudo-data.

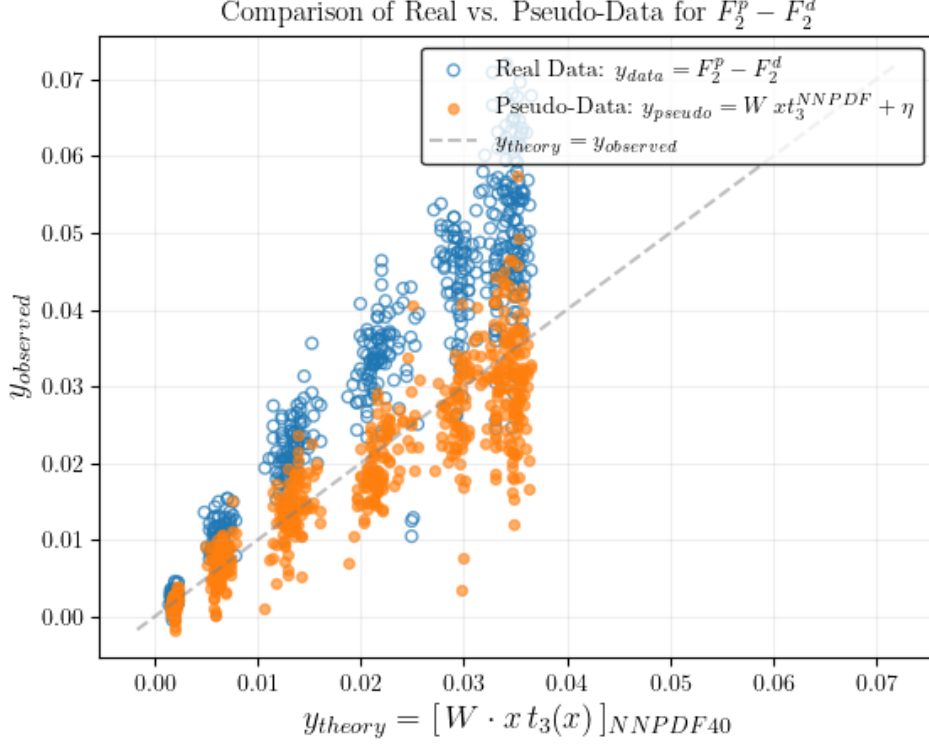


Figure 6: Filled dots: one pseudo-data replica of $\Delta F_2 = F_2^p - F_2^d$ generated around the theory line y_{theory} . Open circles: actual BCDMS measurements. The dashed diagonal is y_{theory} . This was generated using the style of the public code from [5]

7 Preliminary Results

The left panel of Figure 7 displays the fit to the actual BCDMS structure function difference $F_2^p - F_2^d$. The ensemble mean struggles to capture the overall shape of the reference distribution across the well-measured region. This is largely to be expected due to the previously seen deviation from the experimental data and the pseudo-data. For training, the λ_{sum} parameter was set to 0 as it wouldn't constrain the fit to the correct reference integral. Overall, further corrections would indeed need to be made on the experimental data before there would be cohesive agreement between the two distributions. The right panel shows the closure test performed on the central value generated from the NNPDF4.0 reference. As this was the reference that the pseudo-data was generated from, the ensemble mean understandably lies almost on top of the reference curve and only 6% of points aren't contained in the 1σ band. This shows that in this idealised scenario T3Net can capture the momentum distribution xT_3 successfully.

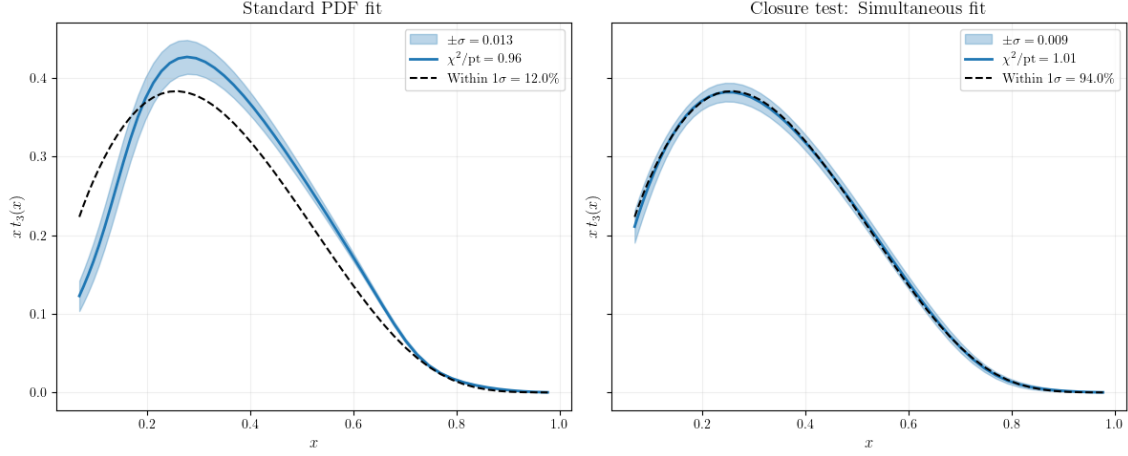


Figure 7: Comparison of fitted non-singlet distributions $xT_3(x)$ obtained from the real BCDMS data (left) and from one pseudo-data replica (right). Also included are the reduced chi squared scores and the percentage of the ‘true’ NNPDF points that lie within one σ .

8 Extension to Beyond-Standard-Model Corrections

Within the EFT framework, the total cross section or structure function may be written at leading order in $1/\Lambda^2$ as

$$y_i = y_i^{\text{SM}} + \frac{C}{\Lambda^2} \Delta y_i^{(1)} + \mathcal{O}(\Lambda^{-4}) \approx y_i^{\text{SM}} [1 + C K_i]$$

where C is the dimensionless Wilson coefficient of a single dimension-six operator and K_i encodes its interference pattern at the kinematic point (x_i, Q_i^2) . Factoring out the SM prediction $y_i^{\text{SM}} = W f_{\text{pred}}$ emphasises that the BSM effect enters as a small, multiplicative distortion.

8.1 Choice of ansatz functions

Since the precise EFT correction $\Delta y_i^{(1)}$ is not known without a full operator level calculation, simple normalised ansatzes K_i are adopted to capture generic kinematic behaviour. Two forms are considered:

$$K_i^{(1)} = \frac{(Q_i^2 - Q_{\min}^2)^2}{\max_j (Q_j^2 - Q_{\min}^2)^2}, \quad K_i^{(2)} = \frac{x_i(1-x_i)(Q_i^2 - Q_{\min}^2)}{\max_j [x_j(1-x_j)(Q_j^2 - Q_{\min}^2)]}$$

where Q_{\min}^2 is the lowest scale in the dataset. Both ansatzes satisfy $\max_i |K_i| = 1$ so that the fitted coefficient C directly measures the relative strength of the BSM contribution. Ansatz 1 emphasises high- Q^2 sensitivity, while ansatz 2 also weights mid- x regions where data density is greatest.

8.2 Goals and theoretical expectations

Introducing a single Wilson coefficient C alongside the non-singlet PDF parametrisation allows a unified test for BSM effects that cannot be absorbed into reshaping $T_3(x)$. The extended prediction ansatz is

$$y_i(C) = [W x T_3(x)]_i [1 + C K_i]$$

where K_i is a normalised interference shape. In summary, the three training procedures can be as follows:

$$y_i^{\text{train}} = \begin{cases} y_i^{\text{real}}, & \text{(Standard PDF fit),} \\ y_i^{\text{pseudo}} \sim \mathcal{N}(W x t_3^{\text{true}}, C_{yy}), & \text{(Closure test),} \\ y_i^{\text{BSM}} \sim \mathcal{N}(W x t_3^{\text{true}} [1 + C_{\text{true}} K], C_{yy}), & \text{(Simultaneous Fit).} \end{cases}$$

When $C_{\text{true}} = 0$, $\hat{C} \approx 0$ should be easily retrieved with uncertainty set by the data covariance, or else there is a serious oversight in the model's design. For $C_{\text{true}} \neq 0$, unbiased recovery of C_{true} would suggest control of PDF-BSM degeneracies. All parameters $(\theta, \alpha, \beta, C)$ are then finally determined by minimising

$$\chi^2 + \lambda_{\text{sum}} (I_{\text{fit}} - I_{\text{ref}})^2 = \sum_{i,j} [y_i(C) - y_i^{\text{train}}] (C_{yy}^{-1})_{ij} [y_j(C) - y_j^{\text{train}}] + \lambda_{\text{sum}} (I_{\text{fit}} - I_{\text{ref}})^2$$

in one simultaneous optimisation.

9 Closure Tests

Figure 8 displays two rows of sensitivity scans corresponding to the two ansatz functions. In each panel the solid curve denotes the ensemble mean of the fitted non-singlet distribution, while the shaded region represents the $\pm 1\sigma$ uncertainty band. Overall to, one σ , the retrieved distribution is stable whilst fitting alongside this new parameter C . Furthermore, for $C_{\text{true}} = 0$, both distributions for each ansatz have the same metrics for fit quality. This is a good sign as the ansatzes may be capable of ‘soaking up’ parts of the data that the network struggles with, even when there is no actual injection, but initially this does not appear to be the case. Interestingly, the uncertainty associated with each PDF fit does not seem to increase along with the magnitude of the injected signal. This can be further explored by analysing the distributions of the recovered wilson coefficients, to see if there is a scale dependence there.

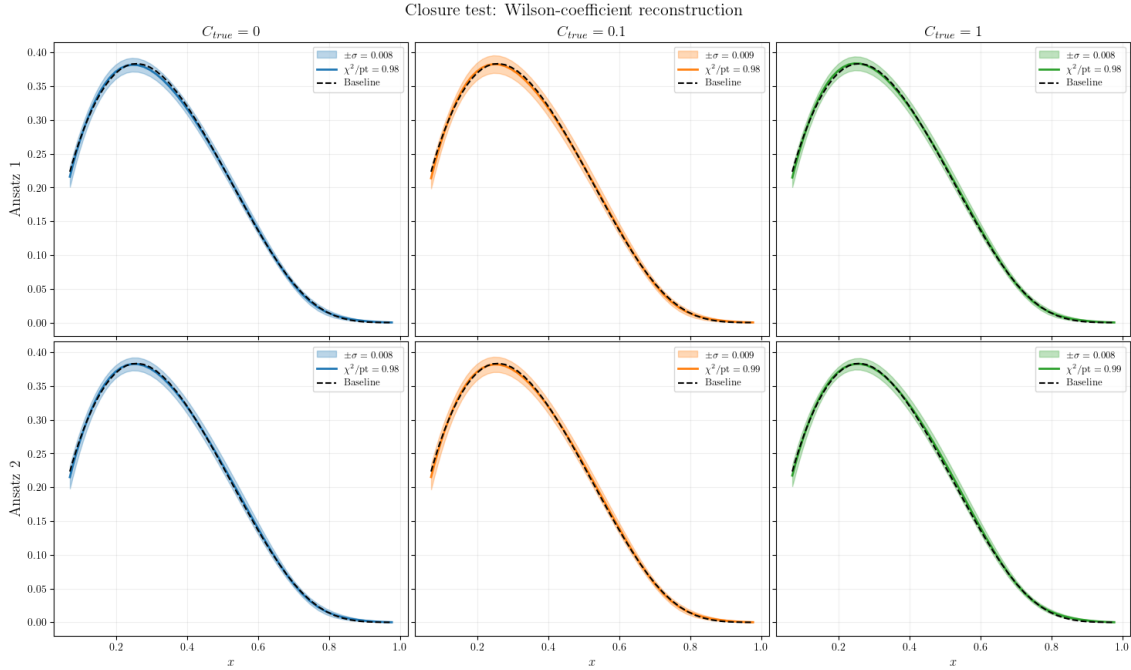


Figure 8: Sensitivity scans for ansatz 1 (top row) and ansatz 2 (bottom row) at $C_{\text{true}} = 0$, 0.1 and 1.0. Each panel shows the ensemble mean of $xT_3(x)$ (solid line), its $\pm 1\sigma$ uncertainty band (shaded region) and the reference distribution (dashed line), with annotated metrics: average uncertainty $\langle\sigma\rangle$, mean χ^2/pt and percentage of points within one standard deviation.

9.1 Recovery of the Wilson coefficient

Figure 9 shows the distribution of the fitted Wilson coefficient C_{fit} over the replica ensemble for each injected signal strength C_{true} . Each histogram is peaked near the corresponding C_{true} , demonstrating that the extraction of C is on average successful to one standard deviation. Nonetheless, a small systematic shift is apparent: the ensemble mean of C_{fit} lies below the injected value in all panels. This residual bias can possibly be attributed to the flexibility of the neural network, which allows some of the BSM-induced deformation to be absorbed into changes of the non-singlet shape itself. This would in turn reduce the portion of the deviation that must be accounted for by the explicit parameter C . Successful simultaneous fits should avoid this PDF-BSM degeneracy, so there is possibly an oversight in the design of T3Net model. This could be from the endpoint exponent parameters α and β , as they are also allowed to float with each fit. Furthermore, the absolute uncertainty interestingly seems to be quite stable again as the magnitude of the signal is increased (see Appendix 10.3 for a formal proof on why this may occur).

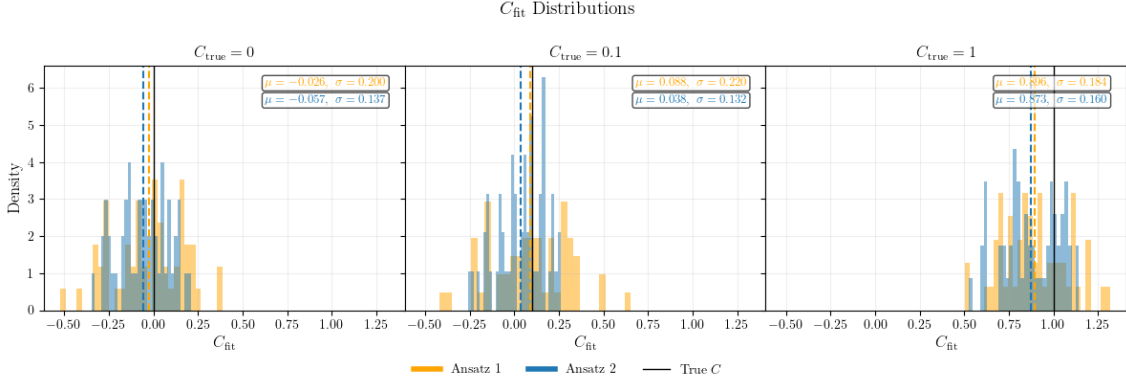


Figure 9: Distributions of fitted Wilson coefficients C_{fit} for ansatz 1 (orange) and ansatz 2 (blue) at $C_{\text{true}} = 0, 0.1$ and 1.0 . Each histogram is normalised to unit area, with dashed vertical lines marking the sample mean and accompanying text indicating the mean and standard deviation. The black vertical line denotes the true injected value.

9.2 Endpoint exponent stability

Figure 10 illustrates the joint distributions of the endpoint exponents α and β . For ansatz 1, all mean points lie well within each other’s uncertainty ellipses, indicating that floating the Wilson coefficient C has a negligible effect on the small and large x behaviour. In the case of ansatz 2, however, the 1σ ellipse at the largest injection $C_{\text{true}} = 1.0$ fails to fully contain the no BSM pseudo-data mean, even though the reverse inclusion holds. This asymmetry could reflect the slightly greater mid- x flexibility of the second ansatz, which in turn could cause a shift in the fitted α and β values. Overall, the two ellipses still overlap substantially, and the absolute shifts amount to only a few percent of the exponent magnitudes. Since endpoint exponents are often held fixed in other frameworks (e.g. NNPDF or SIMUnet), their observed stability here suggests that the bias in C_{fit} does not arise from endpoint behaviour.

9.3 Fixed vs. simultaneous fit: Impact on PDF Reconstruction

Figure 11 contrasts two approaches applied to the same pseudo-data ensemble with no true BSM signal ($C_{\text{true}} = 0$):

- *Fixed-PDF analysis* the network is trained with $C \equiv 0$ throughout.
- *Simultaneous fit* the network is again trained with C as an additional free parameter.

A key observation is that the average PDF uncertainty

$$\langle \sigma_{\text{PDF}}(x) \rangle$$

is smaller when performing a simultaneous fit of both the PDF parameters and the Wilson coefficient C . Naively, the introduction of this extra free parameter would be expected to

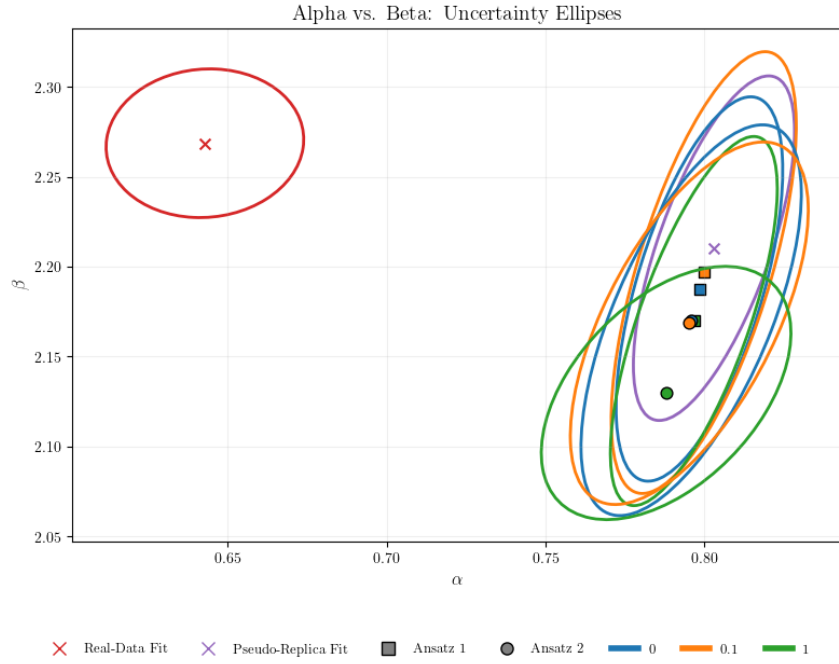


Figure 10: Uncertainty ellipses and mean markers for endpoint exponents (α, β) . Crosses in the central panel correspond to fits without BSM ($C_{\text{true}} = 0$) on real and pseudo-replica data. Squares and circles in the surrounding panels denote the mean exponents for ansatz 1 and ansatz 2 respectively, at each C_{true} . Ellipses represent the 1σ contour of the joint (α, β) distribution.

increase the overall variance of the fit, since each added degree of freedom opens a new direction in parameter space for statistical fluctuations. In practice, variability that was previously absorbed by the PDF parameters alone may now be reallocated into the single coefficient C . The PDF itself then sees less residual variance and its uncertainty band narrows

$$\langle \sigma_{\text{PDF}}(x) \rangle_{\text{simultaneous}} < \langle \sigma_{\text{PDF}}(x) \rangle_{\text{PDF-only}}$$

The initial expectation should therefore be reframed: adding C does not uniformly inflate uncertainties but may instead redistribute variance from the PDF into the coefficient, tightening the PDF band while inducing a bias in the extracted Wilson coefficient.

Comparison of $t_3(x)$: Fixed-PDF analysis vs Simultaneous fit ($C_{\text{true}} = 0$)

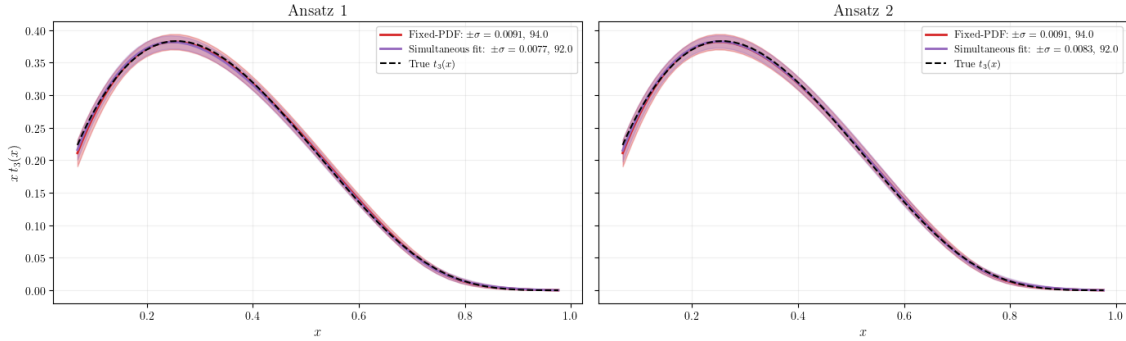


Figure 11: Reconstruction of $x T_3(x)$ for fixed-PDF versus simultaneous- C fit at $C_{\text{true}} = 0$, shown for ansatz 1 (left) and ansatz 2 (right).

9.4 Fixed vs. simultaneous fit: Wilson Coefficient Extraction

The decrease in variance for the simultaneous fit on the PDF itself would suggest that there would be a contrary increase in variance on the corresponding Wilson Coefficient distribution. Figure 12 overlays two estimators of the Wilson coefficient in the closure test at $C_{\text{true}} = 0$:

- *Analytic fixed-PDF extraction:* the non-singlet PDF is held fixed at its mean shape f_{fixed} , obtained from the previous fixed fit. Then the BSM correction can be solved for as a single linear parameter C in

$$y = A + C B + \eta, \quad A = W f_{\text{fixed}}, \quad B = A \odot K,$$

and minimises

$$\chi^2(C) = (y - A - C B)^T C_{yy}^{-1} (y - A - C B).$$

The closed-form solution is

$$C_{\text{fixed}} = \frac{B^T C_{yy}^{-1} (y - A)}{B^T C_{yy}^{-1} B},$$

- *Simultaneous Fit*: the same simultaneous fit from section 9.3

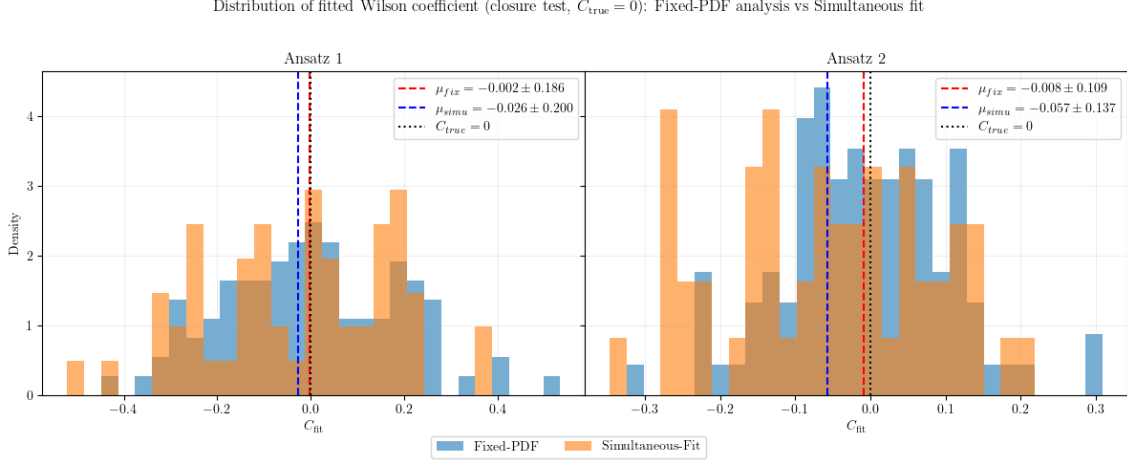


Figure 12: Histograms of C_{fixed} (red) and C_{joint} (blue) over the replica ensemble, with dashed lines marking the respective means μ_{fixed} and μ_{joint} and the dotted line indicating $C_{\text{true}} = 0$.

Both estimators are centred near zero, confirming a lack of significant bias on average. However, the simultaneous distribution is slightly broader and its mean μ_{simu} is offset further from zero than μ_{fix} for both ansatzes. The simultaneous distribution is therefore likely underestimating the uncertainty in the PDF, as it propagates it into the Wilson coefficient. Furthermore, the initial expectations of this study are not matched as the fixed ‘sequential’ approach also seems to have less bias in the final result.

10 Discussion and Conclusion

Overall, T3Net recovers the non-singlet PDF and the Wilson coefficient to within one standard deviation, but two important caveats emerge. First, when C is fitted simultaneously with the PDF, the distribution of C_{fit} exhibits a consistent downward bias relative to C_{true} . Second, compared to the analytic fixed-PDF fit with the ‘sequential’ solving for C , the simultaneous fit produces both a larger bias and a broader spread in C_{fit} . In an ideal scenario, it would be expected that freeing C inflates the PDF uncertainty without introducing a PDF-BSM degeneracy. Instead, the observed behaviour indicates that variance is being reallocated from the PDF into the Wilson coefficient, amplifying both its bias and its uncertainty.

10.1 Potential model issues

The design of T3Net was centred around simplicity, so hundreds of runs of multiple training variations could be performed to draw meaningful conclusions. However, several features

of the model and training procedure may have introduced a systematic bias in the fitted Wilson coefficient.

1. Uniform regularisation on all parameters

Applying an L_2 penalty to every trainable parameter, including the scalar C , implicitly biases C towards zero at each optimisation step. Even when the data contained a genuine non-zero Wilson coefficient, the regulariser could continually ‘shrink’ it, possibly leading to the observed downward bias.

2. Over-strong sum-rule penalty relative to χ^2

The valence sum constraint enters the loss as

$$\lambda_{\text{sum}} (I_{\text{fit}} - I_{\text{ref}})^2, \quad \lambda_{\text{sum}} = 10,000.$$

With typical χ^2 values around a hundred, this term can dominate unless the integral is almost exact. For example, if the deviation between the two integrals is 0.1, the loss contribution from this step can make up over 50% of the loss. Although a lower value ($\lambda_{\text{sum}} = 1,000$) led to unstable performance at the peak, the chosen $\lambda_{\text{sum}} = 10,000$ still may over constrain the total area of $T_3(x)$, indirectly suppressing any shift that might otherwise be absorbed by C .

3. Collinearity between the ansatz and the network basis

The preprocessing functions $x^\alpha(1-x)^\beta$ and the MLP activations can already reproduce many smooth distortions in the data. If the chosen ansatz $K(x, q^2)$ lies largely within that function space, the optimiser can explain BSM deviations by adjusting the MLP weights rather than by moving C . This multicollinearity can then drive the fitted coefficient towards zero.

4. Early stopping driven solely by χ^2 on the observable

The training halts when the validation χ^2 on y ceases to improve, without explicitly monitoring the recovery of C . Since the network can absorb most discrepancies via its PDF branch the optimal χ^2 can occur at a suboptimal C and training terminates before the coefficient has fully converged.

10.2 Comparison with SIMUnet and methodological lessons

Despite investigating the methodology of SIMUnet in this study, there are still fundamental differences between the two approaches. Firstly, SIMUnet constructs its predictions by convolving full NNLO QCD and NLO FK tables with the neural network PDF parametrisation and then inserting SMEFT corrections via a dedicated combination layer. This layer carries trainable weights for both the linear interference terms $R^{(n)}$ and the quadratic self-interference terms $R^{(n,m)}$. By contrast, T3Net again models new physics effects through a single shape function $K(x, Q^2)$

$$y_{\text{pred}}(x, Q^2) = W t_3(x) [1 + C K(x, Q^2)]$$

This ansatz doesn't consider $\mathcal{O}(C^2)$ contributions, treats C outside the core network and applies it as a multiplicative factor in the training loop. Additionally, SIMUnet enforces sum rules and PDF positivity via architectural constraints and Lagrange multipliers built into the loss as per the NNPDF methodology. **T3Net** relies on a large external penalty λ_{sum} that as discussed may over-constrain the PDF shape. Finally, SIMUnet employs hyperparameter optimisation across its parameters, while **T3Net** uses a simple fixed architecture. In summary, **T3Net** not being able to avoid PDF-BSM degeneracy does not condemn the viability of simultaneous fits, but rather highlights how oversights in the model's construction can lead to their failure.

10.3 Future research directions

There are several ways to improve the sensitivity and robustness of simultaneous **T3Net** fits in the SMEFT setting. Firstly, separating the regularisation of the Wilson coefficient from that of the neural network weights prevents all parameters being uniformly pulled towards zero. Assign distinct L_2 penalties to the SMEFT coupling and to the network parameters, or adopt hierarchical Bayesian priors to preserve numerical stability. At the same time, replacing the rigid valence sum constraint with a softer constraint that allows for small deviations when they improve the overall fit. Broadening the experimental input is equally important to distinguish real new physics signals from fitting artefacts. Incorporating measurements such as Drell-Yan, jets, electroweak and heavy flavour production would probe different parton channels in complementary ways. This richer dataset could help cross-check potential SMEFT signals across diverse processes. On the modelling side, updating the neural network architecture can increase flexibility and capture subtle correlations. Mixture-of-experts layers or attention mechanisms allow different subnetworks to specialise in specific kinematic regions, while physics informed subnetworks embed known theoretical constraints directly into the model. Such designs better represent the connection between PDF shapes and the kinematic patterns generated by SMEFT operators, overall improving the extraction of Wilson coefficients.

AI Declaration

Generative AI tools (ChatGPT) were employed during the preparation of this thesis. In particular, AI support contributed to:

- Preliminary drafting of prose in multiple sections (introductions, methodological summaries, and conclusions), which were then fully rewritten and expanded by the author. It did not perform any scientific analysis in of itself, but rather summarise what code does and captions for figures for example.
- LaTeX formatting and styling: document structure, section headings, figure placement, table layouts and bibliography setup.
- Generation and polishing of all plotting routines and LaTeX style figure captions, ensuring publication ready quality.
- Suggestions for inline code comments, skeletons for data pipeline scripts and initial outlines of training loops, subsequently reviewed and reimplemented by the author.
- Outline of key derivations and references to publicly accessible materials. All mathematical steps were independently derived, verified and integrated by the author.
- Grammar and UK-English consistency checks across text, captions and code documentation.

Every AI-generated suggestion has been carefully reviewed, substantially edited or replaced, and fully validated by the author. All scientific content (such as the design, implementation, execution of neural network models, data generation procedures, numerical analyses and interpretation of results) was independently developed and verified by the author.

Glossary of Terms

Loop A closed path in a Feynman diagram that corresponds to virtual particles circulating in an intermediate state. This then contributes to higher-order corrections in perturbation theory.

Valence The quark content that determines the quantum numbers of a hadron (for example, two up quarks and one down quark in the proton), as opposed to sea quarks or gluons.

Rapidity A kinematic variable defined by

$$y = \frac{1}{2} \ln[(E + p_z)/(E - p_z)],$$

where E is the energy and p_z the momentum component along the beam axis.

Invariant mass The Lorentz-invariant quantity

$$m_{\text{inv}}^2 = E^2 - \mathbf{p}^2,$$

computed from the total energy E and three-momentum \mathbf{p} of a particle.

Transverse momentum p_T The component of momentum perpendicular to the beam axis,

$$p_T = \sqrt{p_x^2 + p_y^2},$$

used to characterise particle production in collider experiments.

Four-momentum The relativistic momentum four-vector

$$p^\mu = (E, \mathbf{p}),$$

combining the energy E and three-momentum \mathbf{p} into a single Lorentz-covariant object.

Datasets and Preprocessing Details

- **BCDMS proton DIS data dataset** = `BCDMS_NC_NOTFIXED_P_EM-F2`, `variant = legacy`, loaded via `validphys.API.loaded_commondata_with_cuts` with `use_cuts = internal` and `theoryid = 208`.
- **BCDMS deuteron DIS data dataset** = `BCDMS_NC_NOTFIXED_D_EM-F2`, same variant, cuts and theory ID as for the proton.
- **Internal cuts** All BCDMS data are subject to the experiment's default quality and kinematic cuts via `use_cuts = internal` in the ValidPhys API.
- **Covariance computation** Combined systematic covariance obtained from `API.dataset_inputs_covmat_from_systematics` for the proton and deuteron inputs; sub-blocks C_{pp}, C_{dd}, C_{pd} extracted and merged according to matching (x, q^2) indices.
- **Fast-kernel tables** Loaded with `validphys.fkparser.load_fktable` from set names "BCDMSP" and "BCDMSD" at `theoryID=208`. No additional c -factors (`cfac=()`).
- **PDF baseline for closure** LHAPDF set "NNPDF40_nnlo_as_01180", replica 0, used to compute the reference $xt_3^{\text{true}}(x)$ at the common input scale Q_0^2 .
- **Grid information** The (x_α) grid for convolution extracted from `fk_p.xgrid`, matching both proton and deuteron FK tables.
- **Data matching** Proton and deuteron tables merged on (x, q^2) via `pandas merge`, producing N_{data} matched points with entries `entry_p`, `entry_d` used to index into the FK arrays.

-
- **Kinematic range** BCDMS covers roughly $0.07 \lesssim x \lesssim 0.8$ and $20 \text{ GeV}^2 \lesssim Q^2 \lesssim 200 \text{ GeV}^2$. Only points with matching (x, q^2) appear in the non-singlet dataset.
 - **Random seeds** Pseudo-data and replica sampling use NumPy's `default_rng` for reproducibility (seed = 451) and torch manual seeds proportional to the replica index for training splits. This is mainly a measure in case this script is later expanded.

Constancy of the statistical uncertainty on C

Consider a prediction of the form

$$y_{\text{pred}}(C) = y_{\text{SM}} [1 + C K]$$

with data covariance C_{yy} . The chi-squared is

$$\chi^2(C) = [y - y_{\text{pred}}(C)]^T C_{yy}^{-1} [y - y_{\text{pred}}(C)]$$

Its second derivative with respect to C is

$$\frac{d^2 \chi^2}{dC^2} = 2 K^T y_{\text{SM}} C_{yy}^{-1} y_{\text{SM}} K$$

which is independent of C . Hence, near the minimum,

$$\sigma_C^2 \approx \left[\frac{1}{2} \frac{d^2 \chi^2}{dC^2} \right]^{-1} = (K^T y_{\text{SM}} C_{yy}^{-1} y_{\text{SM}} K)^{-1}$$

showing that the statistical uncertainty σ_C does not depend on the true signal strength C_{true} .

References

- [1] R. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. Latorre, A. Piccione, J. Rojo, and M. Ubiali, “A determination of parton distributions with faithful uncertainty estimation (vol 809, pg 1, 2009),” *Nuclear physics b*, vol. 816, no. 1-2, pp. 293–293, Jul. 2009.
- [2] R. D. Ball, S. Carrazza, J. Cruz-Martinez, L. Del Debbio, S. Forte, T. Giani, S. Iranipour, Z. Kassabov, J. I. Latorre, E. R. Nocera, R. L. Pearson, J. Rojo, R. Stegeman, C. Schwan, M. Ubiali, C. Voisey, M. Wilson, and N. Collaboration, “An open-source machine learning framework for global analyses of parton distributions,” *The European Physical Journal C*, vol. 81, no. 10, p. 958, 2021. [Online]. Available: <https://doi.org/10.1140/epjc/s10052-021-09747-9>
- [3] R. D. Ball, S. Carrazza, J. Cruz-Martinez, L. Del Debbio, S. Forte, T. Giani, S. Iranipour, Z. Kassabov, J. I. Latorre, E. R. Nocera, R. L. Pearson, J. Rojo, R. Stegeman, C. Schwan, M. Ubiali, C. Voisey, and M. Wilson, “The path to proton structure at 1% accuracy: Nnpdf collaboration,” *The European Physical Journal C*, vol. 82, no. 5, May 2022. [Online]. Available: <http://dx.doi.org/10.1140/epjc/s10052-022-10328-7>
- [4] S. Iranipour and M. Ubiali, “A new generation of simultaneous fits to lhc data using deep learning,” *Journal of High Energy Physics*, vol. 2022, no. 5, p. 32, 2022. [Online]. Available: [https://doi.org/10.1007/JHEP05\(2022\)032](https://doi.org/10.1007/JHEP05(2022)032)
- [5] A. Candido, L. D. Debbio, T. Giani, and G. Petrillo, “Bayesian inference with gaussian processes for the determination of parton distribution functions,” *The European Physical Journal C*, vol. 84, no. 7, p. 716, 2024. [Online]. Available: <https://doi.org/10.1140/epjc/s10052-024-13100-1>
- [6] S. Forte and Z. Kassabov, “Why α_s cannot be determined from hadronic processes without simultaneously determining the parton distributions,” *The European Physical Journal C*, vol. 80, no. 3, Feb. 2020. [Online]. Available: <http://dx.doi.org/10.1140/epjc/s10052-020-7748-6>
- [7] R. Devenish and A. Cooper-Sarkar, *Deep Inelastic Scattering*. Oxford University Press, 11 2003. [Online]. Available: <https://doi.org/10.1093/acprof:oso/9780198506713.001.0001>
- [8] S. Mortier, “Exploring top-quark and z boson coupling in top pair production in association with a z boson,” Ph.D. dissertation, 07 2020.
- [9] W. Youcai, “Measurement of single-spin asymmetry in charged kaon electroproduction on a transversely polarized 3he target,” 02 2010.
- [10] S. Zahari Kassabov, “Validphys 2 guide — data.nnpdf.science,” <https://data.nnpdf.science/validphys-docs/guide.html#overview>, [Accessed 29-06-2025].

-
- [11] A. C. Benvenuti, D. Bollini, G. Bruni, T. Camporesi, L. Monari, F. L. Navarria, A. Argento, J. Cvach, W. Lohmann, L. Piemontese, P. Závada, A. A. Akhundov, V. I. Genchev, I. A. Golutvin, Y. T. Kiryushin, V. G. Krivokhizhin, V. V. Kukhtin, R. Lednicky, S. Némécek, P. Reimer, I. A. Savin, G. I. Smirnov, J. Strachota, G. G. Sultanov, P. T. Todorov, A. G. Volodko, D. Jamnik, R. Kopp, U. Meyer-Berkhout, A. Staude, K. M. Teichert, R. Tirlir, R. Voss, C. Zupancic, M. Cribier, J. Feltesse, A. I. Milshtein, A. Ouraou, P. Rich-Hennion, Y. Sacquin, G. Smadja, and M. Virchaux, “A high statistics measurement of the proton structure functions $F_2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2 ,” *Phys. Lett. B*, vol. 223, pp. 485–489, 1989. [Online]. Available: <https://cds.cern.ch/record/194232>
- [12] S. Alekhin, S. Kulagin, and R. Petti, “Nuclear effects in the deuteron and global qcd analyses,” *Physical Review D*, vol. 105, no. 11, Jun. 2022. [Online]. Available: <http://dx.doi.org/10.1103/PhysRevD.105.114037>