Executive Summary

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

Lucas Curtin

922 words

July 1, 2025

The internal structure of the proton, encoded in parton distribution functions (PDFs), is the cornerstone of theoretical predictions for hadronic processes at collider experiments. PDFs describe the probability densities for quarks and gluons carrying a fraction x of the proton's momentum and enter the factorisation theorem that separates calculable short distance cross sections from non-perturbative long distance physics. Uncertainties in these distributions propagate directly into predictions for Higgs boson production, electroweak boson scattering, high-p_T jet rates, and rare processes. These uncertainties limit both the precision tests of quantum chromodynamics (QCD) and the sensitivity to beyond-Standard-Model (BSM) phenomena. Conventional PDF determinations historically imposed fixed analytic forms, capable of introducing a systematic functional form bias in regions where data is sparse or absent [1]. The NNPDF framework overcame this limitation by representing each PDF flavour with a neural network trained on Monte Carlo replicas of diverse datasets thereby ensuring flexibility and statistically robust uncertainty estimates [2, 3].

In parallel, the search for new physics at energy scales beyond the direct reach of colliders has matured into the Standard Model Effective Field Theory (SMEFT) approach. The effects are encoded in higher dimensional operators weighted by Wilson coefficients. Precision extraction of these coefficients from hadronic data requires disentangling genuine BSM signals from PDF uncertainties. If this process is completed sequentially i.e. first fitting PDFs under pure QCD, then extracting Wilson coefficients residual new physics effects can be absorbed into the PDFs themselves. This in turn can bias any inference on the Wilson coefficients and undermines the validity of BSM searches [4]. To avoid this PDF-BSM degeneracy, simultaneous global fits of PDFs and Wilson coefficients have been proposed. The SIMUnet framework extends the NNPDF neural network architecture by including dedicated outputs for EFT parameters [5]. Then all parameters are trained together in a unified loss function to prevent sequential fitting artifacts. Additionally, Candido *et al.* introduced a fully Bayesian alternative that models PDFs with Gaussian process priors and performs joint inference over both PDF parameters and hyperparameters alongside SMEFT coefficients [6]. This was demonstrated on the simple non-singlet triplet $T_3(x) = u^+(x) - d^+(x)$, which doesn't include more complex features as it obeys a single DGLAP evolution equation and maps directly onto the experimental proton-deuteron structure-function difference $\Delta F_2 = F_2^p - F_2^d$. While these methodologies establish the principle of unbiased simultaneous fits, their full global implementation remains computationally demanding and methodologically complex.

This work presents `T3Net`, a minimal closure test framework that isolates the core challenge of simultaneous PDF-BSM inference in a controlled setting. Again focusing on the non-singlet combination $T_3(x)$, this toy model retains essential physics while minimising technical overhead. High statistics BCDMS measurements of F_2^p and F_2^d are combined pointwise, and their full experimental covariance is preserved [7]. NNPDF's fast-kernel tables compress DGLAP evolution and perturbative coefficient functions into a single convolution matrix, reducing each theory prediction to a simple matrix product. In essence, they construct the theoretical values of observables from combinations of reference PDFs, which can be compared to experimental results. The model outputs the momentum distribution $x\,T_3(x)$, as required by the FK tables, although the fitting is often described in terms of the underlying PDF $T_3(x)$ as the fundamental basis.

The parametrisation of $xT_3(x)$ in `T3Net` enforces known small and large x endpoint power laws via trainable exponents α and β, while the neural network captures residual shape variations. A soft valence sum rule ensures baryon number conservation, and a global L_2 penalty guards against overfitting. Monte Carlo pseudo-datasets are generated by sampling the experimental covariance around a known reference PDF, and each replica is fitted independently by minimising a combined χ^2. The resulting ensemble provides a non-parametric uncertainty band for the non-singlet distribution along with the final mean value. In initial closure tests, `T3Net` reproduces the reference $xT_3(x)$ within its one-sigma uncertainty band across the experimentally available x range. The model is then extended to fit a single Wilson coefficient C on data injected with different ansatzes at different strengths, as to trivially model some form of BSM behaviour. After the simultaneous fit, the reconstructed $\hat{C}$ remains within one-sigma of its true value but shows a minor negative bias. Moreover, comparing this simultaneous extraction to a sequential 'fixed-PDF' procedure, in which the PDF is determined first and C solved in closed form afterwards, reveals that the bias in $\hat{C}$ is appreciably larger under the simultaneous fit. Several implementation choices in the minimal `T3Net` may have contributed to the enhanced bias observed in the fitted Wilson coefficient. In particular:

- A uniform L_2 regularisation penalty applied to both the neural network weights and the scalar coefficient C, which systematically pulls C toward zero.
- A rather aggressive valence sum penalty that can dominate the loss function and indirectly suppress genuine shifts in C.
- Collinearity between the chosen SMEFT ansatz $K(x, Q^2)$ and the network basis functions, allowing the model to absorb BSM-like distortions into the PDF shape rather than into C.
- An early stopping fallback driven solely by the structure-function χ^2, potentially

terminating training before C has fully converged.

By contrast, the SIMUnet framework incorporates several features designed to avoid these pitfalls:

- Sum rules enforced via architectural constraints and Lagrange multipliers rather than a single external penalty term as per the NNPDF methodology.
- Wilson coefficients embedded directly within the neural network training graph, rather than applied as an external multiplicative factor.
- Hyperparameter optimisation carried out over both network and SMEFT parameters, ensuring balanced training and convergence that monitor both PDF accuracy and SMEFT recovery.

These distinctions suggest that the residual bias in `T3Net` arises from its simplified model design rather than from any fundamental flaw in the concept of simultaneous fits. Adopting more stable features of SIMUnet in full global analyses should restore unbiased precision in both proton structure and SMEFT parameter determinations for these smaller validation models.

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. 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

[5] 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

[6] 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

[7] 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. Tirler, 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