Symbolic Regression and Precision LHC Physics

M. Morales Alvarado*, J. Bendavid, D. Conde, V. Sanz, M. Ubiali

*INFN Trieste

mmorales@sissa.it

1 Introduction

- In high energy physics, fundamental interactions can be described by the *Standard Model* (SM) of particle physics to a good extent.
- The SM is tested at the world's most powerful microscope: the *Large Hadron Collider* (LHC). Theoretical predictions and observables are compared [1].

- The question: *Can simple, accurate, and useful formulas be derived from noisy, complex, collider data?*

2 Methodology

- We use *symbolic regression* (SR) in the PySR [2] library.
- Formulas are represented as expression trees and are evolved via crossover and mutations to generate new trees, Fig. 1.

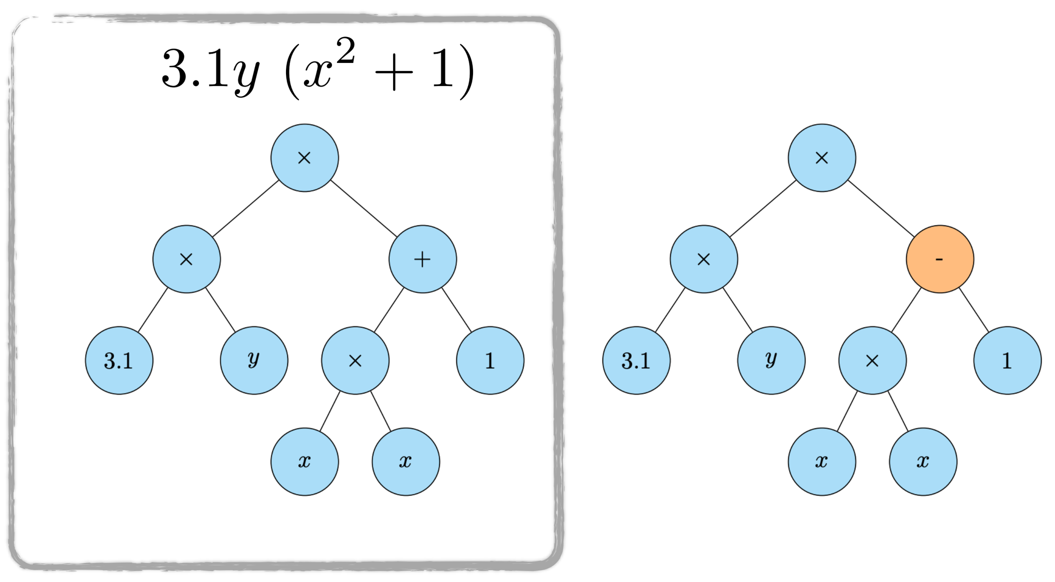

Fig. 1: Expression tree, mutation, and crossover.

- Trees are chosen as to maximise accuracy and minimise complexity.

3 Results

- *Equation recovery in quantum electrodynamics, Fig. 2.*

$$e^- e^+ \to \gamma^* \to \mu^- \mu^+ \qquad \frac{d\sigma}{d\cos\theta} = \frac{\pi\alpha^2}{2s}(1+\cos^2\theta)$$

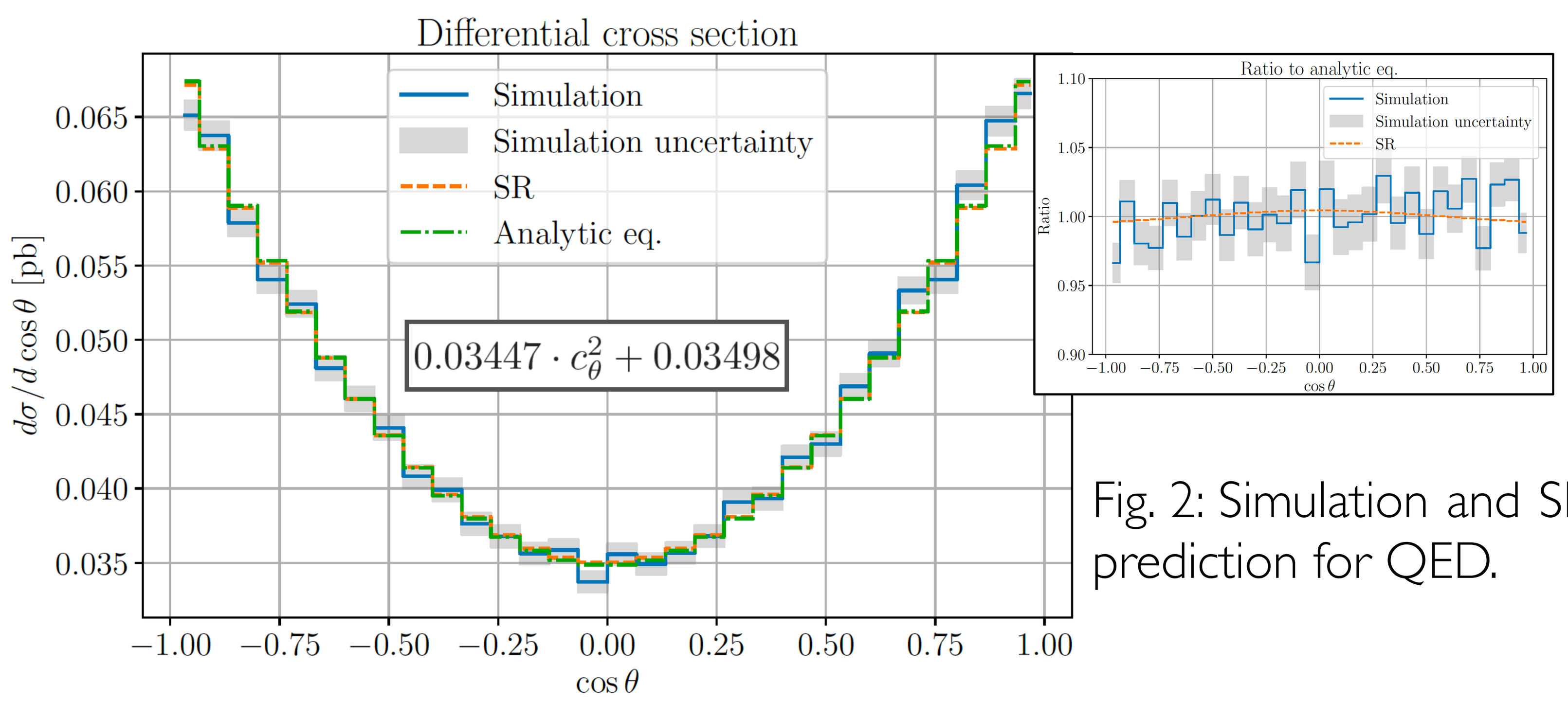

Fig. 2: Simulation and SR prediction for QED.

- *Equation discovery in quantum chromodynamics [3], Fig. 3.*

$$p\,p \to \gamma^* \to \mu^- \mu^+ \qquad \frac{d^2\sigma}{dM dy} = \frac{8\pi\alpha^2}{9sM} F(M,y)$$

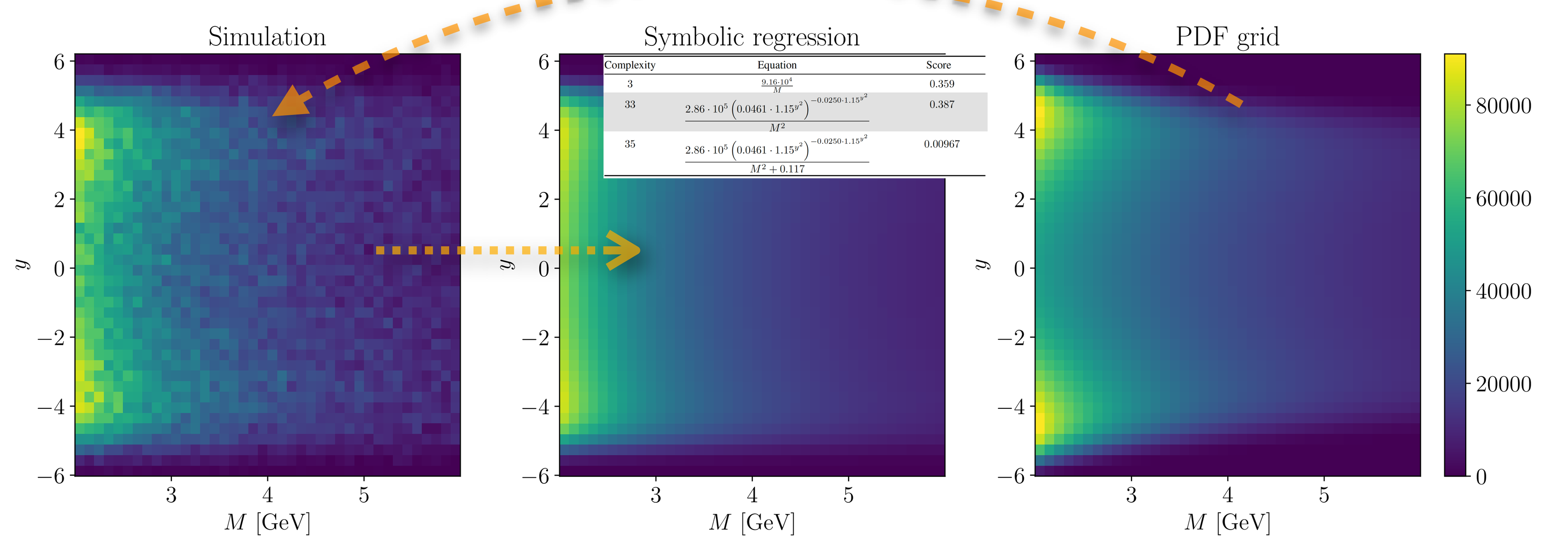

Fig. 3: Simulation, SR fit, and underlying F (PDF grid).

- *Equation discovery: angular coefficients in Drell-Yan, Figs. 4, 5, and 6.*

$$\frac{d^5\sigma}{dp_T\,dy\,dm\,d\cos\theta\,d\phi} = \frac{3}{16\pi}\frac{d^3\sigma^{U+L}}{dp_T\,dy\,dm}\left[(1+\cos^2\theta) + \sum_{i=0}^{7} P_i(\theta,\phi)\,A_i\right] \qquad A_i(p_T, y, m)$$

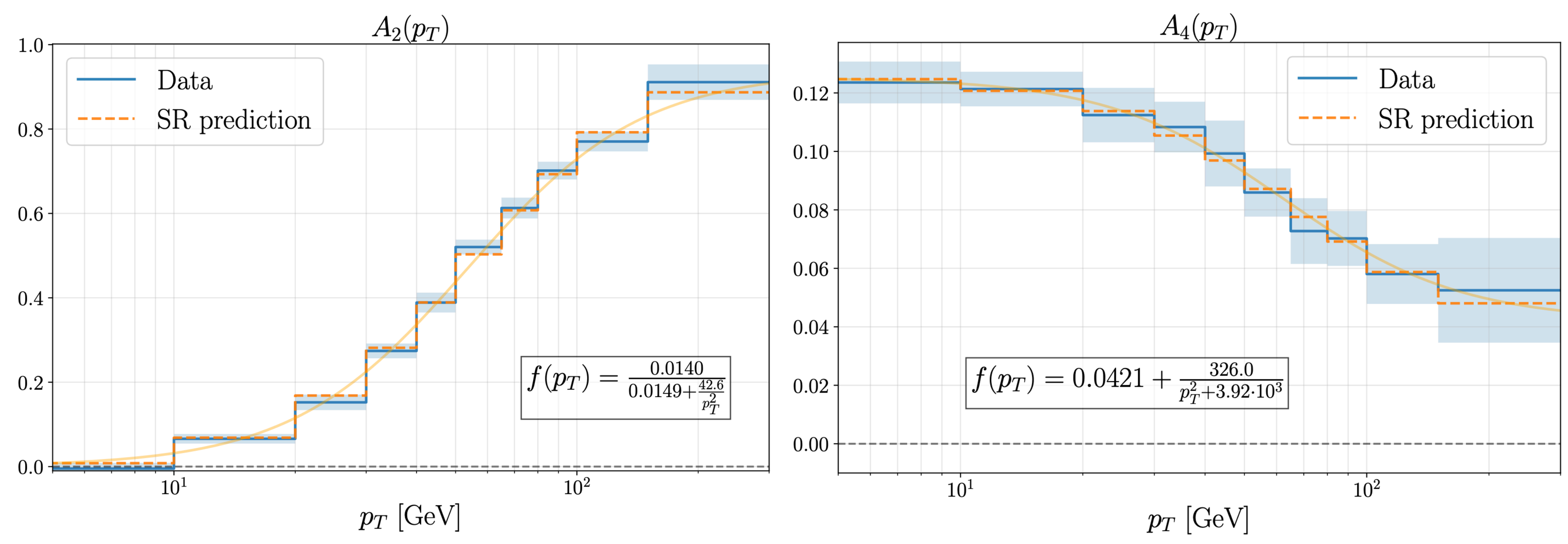

Fig. 4: Data and SR prediction in 1D distributions in p_T .

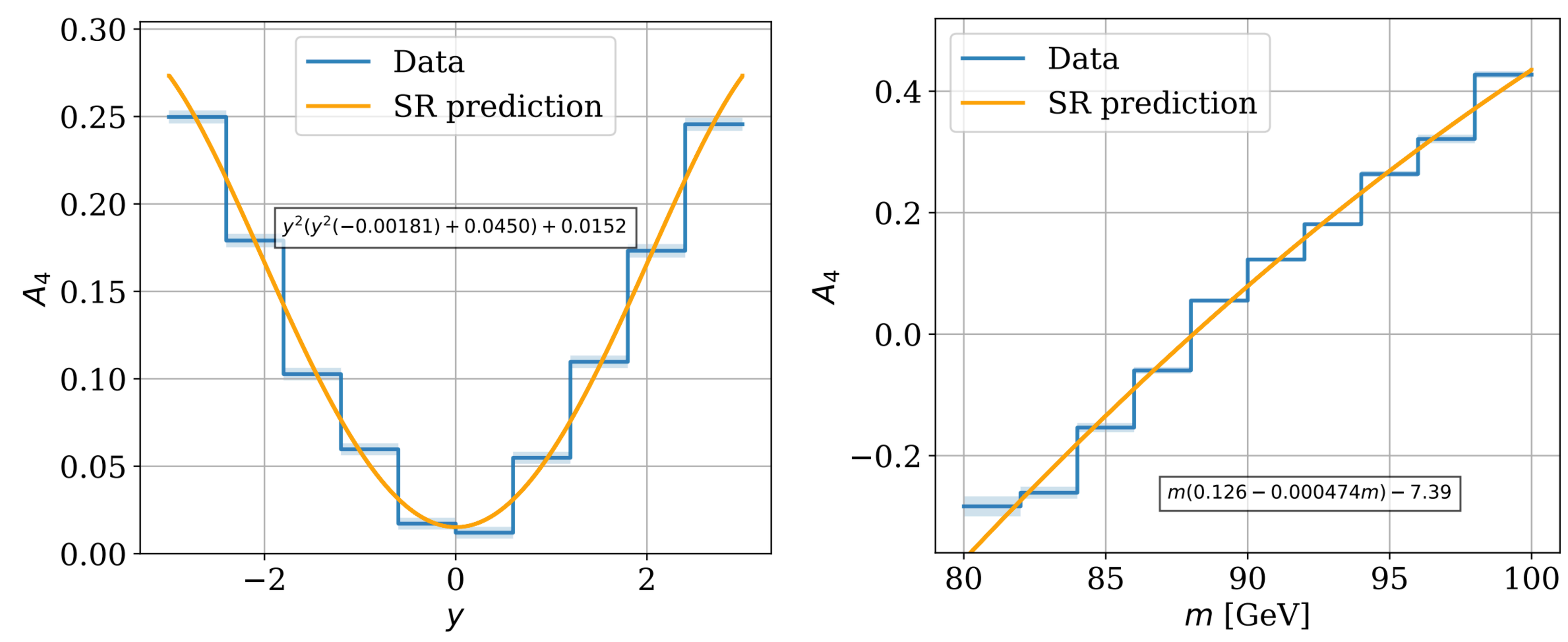

Fig. 5: Data and SR prediction in 1D distributions in y and m.

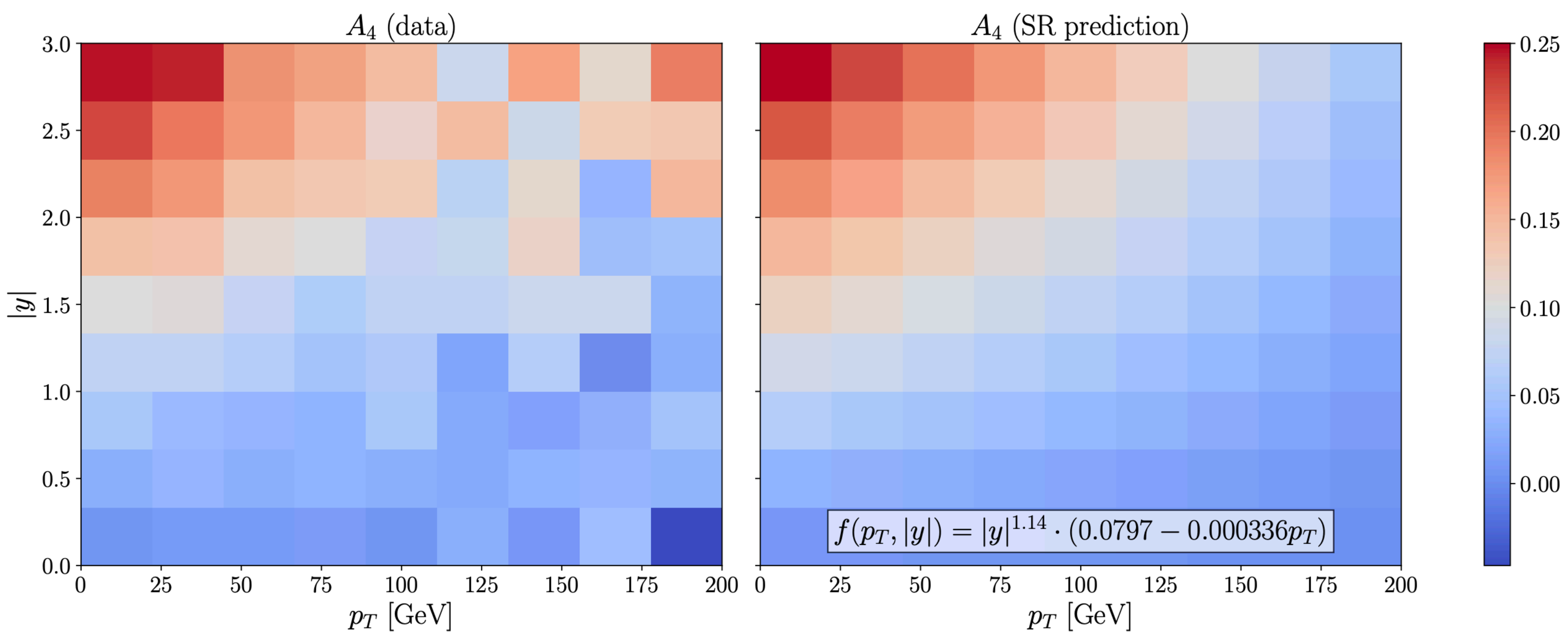

Fig. 5: Data and SR prediction in 2D distributions in $(p_T, |y|)$.

The answer:

SR can combine the power of machine learning with the intuition of simple equations to describe collider data easily and efficiently.

arXiv 2508.0098

[1] Weinberg, S. *The Quantum Theory of Fields* (1996).
[2] Cranmer, M. Interpretable Machine Learning with PySR and SymbolicRegression.jl, arXiv 2305.01582.
[3] QCD and Collider Physics, R. K. Ellis, W. J. Stirling, B. R. Webber (1996).