

Precision-Machine Learning for the Matrix Element Method

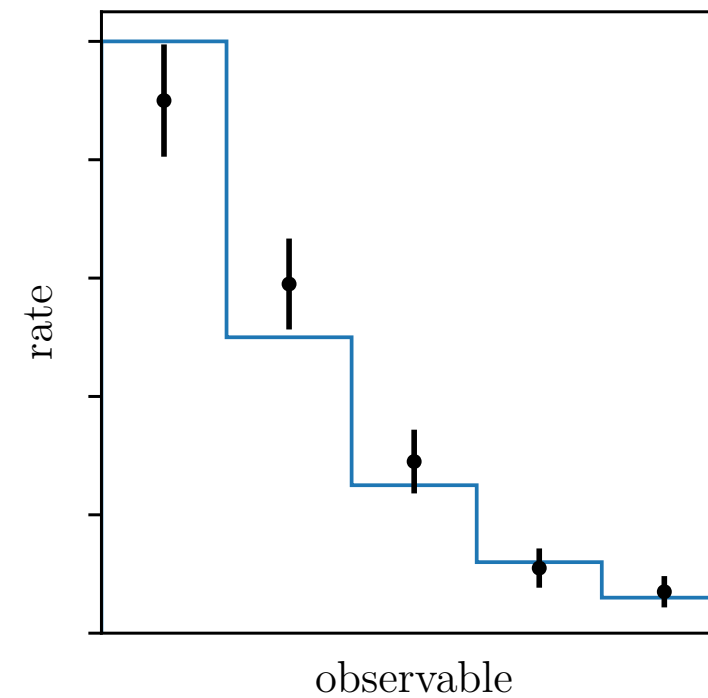
Theo Heimel
April 2024

Institut für theoretische Physik
Universität Heidelberg

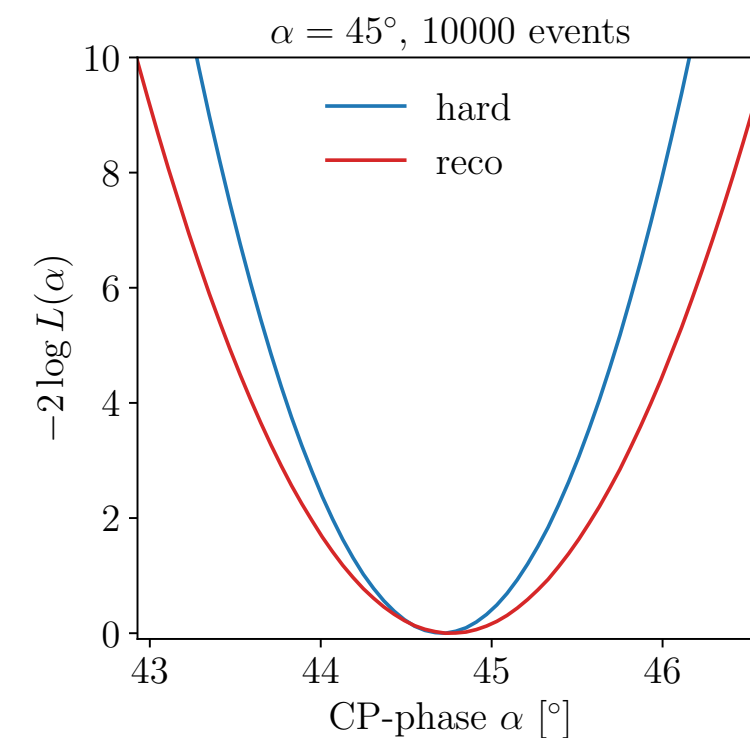


UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

How can we extract all the
available information from LHC data?



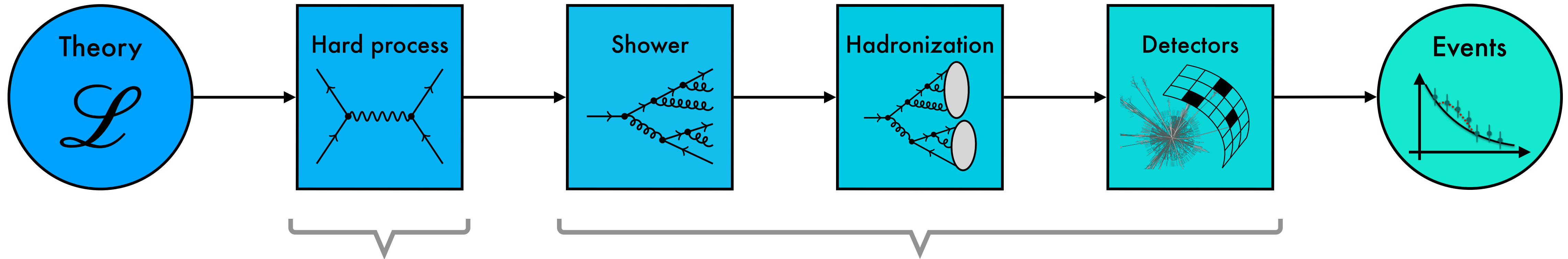
Binned, low-dimensional data
→ **loss of information**



Use theory knowledge to
extract likelihood

→ **matrix element method**

Precision-Machine Learning for the Matrix Element Method



known from theory

likelihood intractable
 → use machine learning

Precision-Machine Learning for the Matrix Element Method
 Theo Heinzel, Nathan Huetsch, Ramon Winterhalder, Tilman Plehn, Anja Butter

Classical analysis

- hand-crafted observables
- limited data
- loss of information

How can we extract all the available information from LHC data?

Matrix Element Method (MEM)

- Based on first principles
- estimates uncertainties reliably
- optimal use of information
- perfect for processes with few events

$$p(x_{\text{reco}} | \alpha) = \int d\mathcal{M}_{\text{hard}} p(\mathcal{M}_{\text{hard}} | \alpha) p(x_{\text{reco}} | \mathcal{M}_{\text{hard}}) \epsilon(\mathcal{M}_{\text{hard}})$$

Efficient MC integration
 importance sampling with normalizing flow
 $p_{\text{hard}} = p(\mathcal{M}_{\text{hard}} | \alpha)$

Theory knowledge
 diff. cross-section
 $\frac{1}{\sigma} \frac{d\sigma}{d\mathcal{M}_{\text{hard}}}$

Transfer function
 density estimation; normalizing flow and transformer

Acceptance function
 learn with simple classifier network

Learning the transfer function

LHC example
 Single top and Higgs production with anomalous CP-phase α
 Hadronic decay of top: $t\bar{t} \rightarrow (b\bar{b})(\gamma\gamma) + \text{OCD jets}$

- low total cross section (few events)
- low relation of rate
- kinematic observables still sensitive
- ideal use case for MEM

Results

- smooth and well-calibrated likelihoods, both for low and high event counts
- close to optimal information
- uncertainty bands; MC integration error & systematic error from limited training statistics (BNN)

Transformers
 correlations between momenta, combinatorics

normalizing flow
 likelihood for individual momenta

Bayesian networks
 estimate training uncertainties

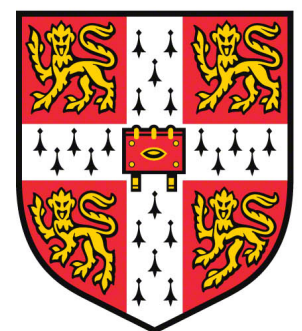
Come to my poster to see how this can be done with transformers, normalizing flows, classifiers and neural importance sampling!

PDF uncertainties in the presence of inconsistent training data

Mark N. Costantini

EuCAIF Conference 2024, Amsterdam

POSTER: #100



**UNIVERSITY OF
CAMBRIDGE**



**Funded by
the European Union**



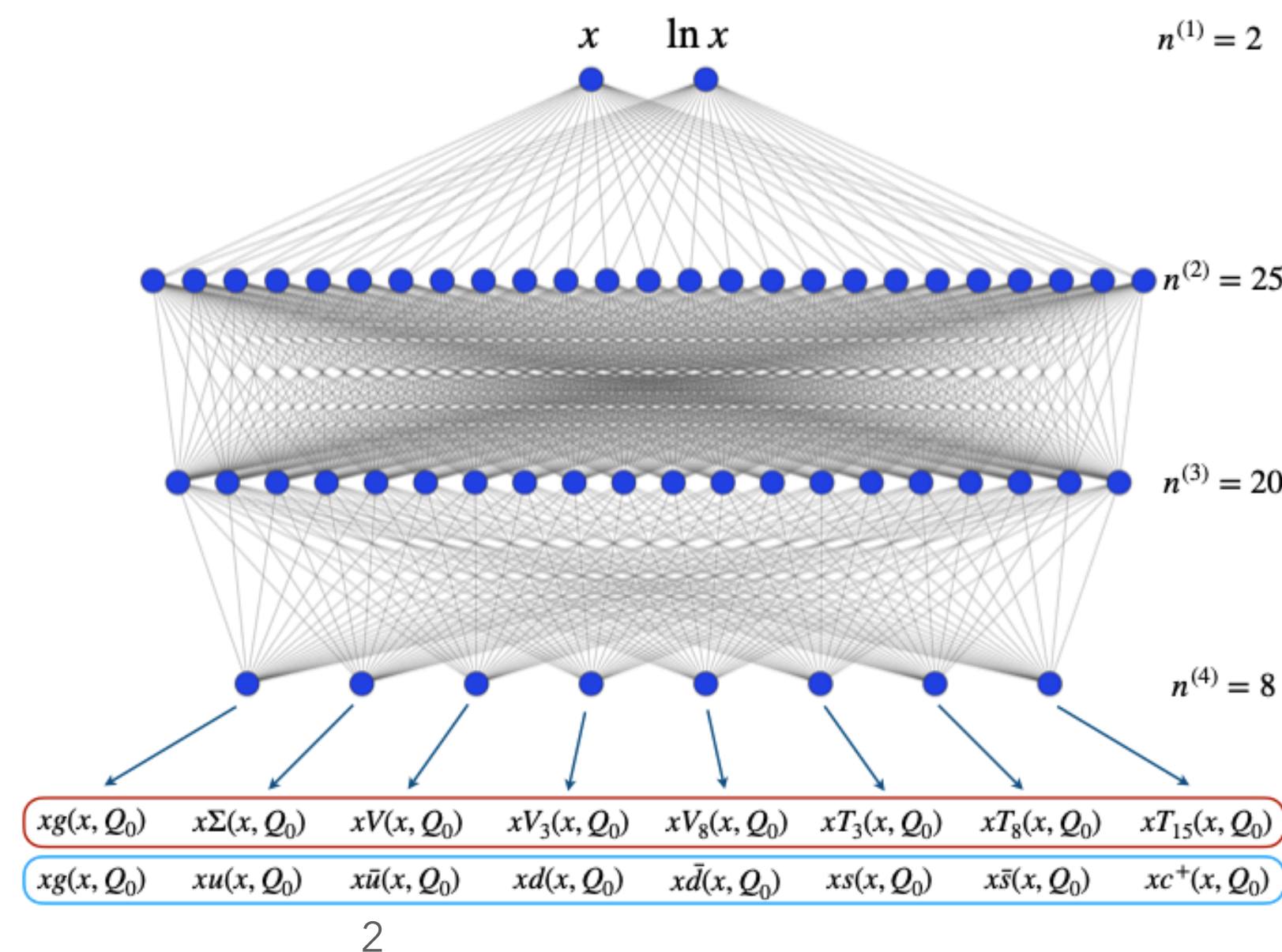
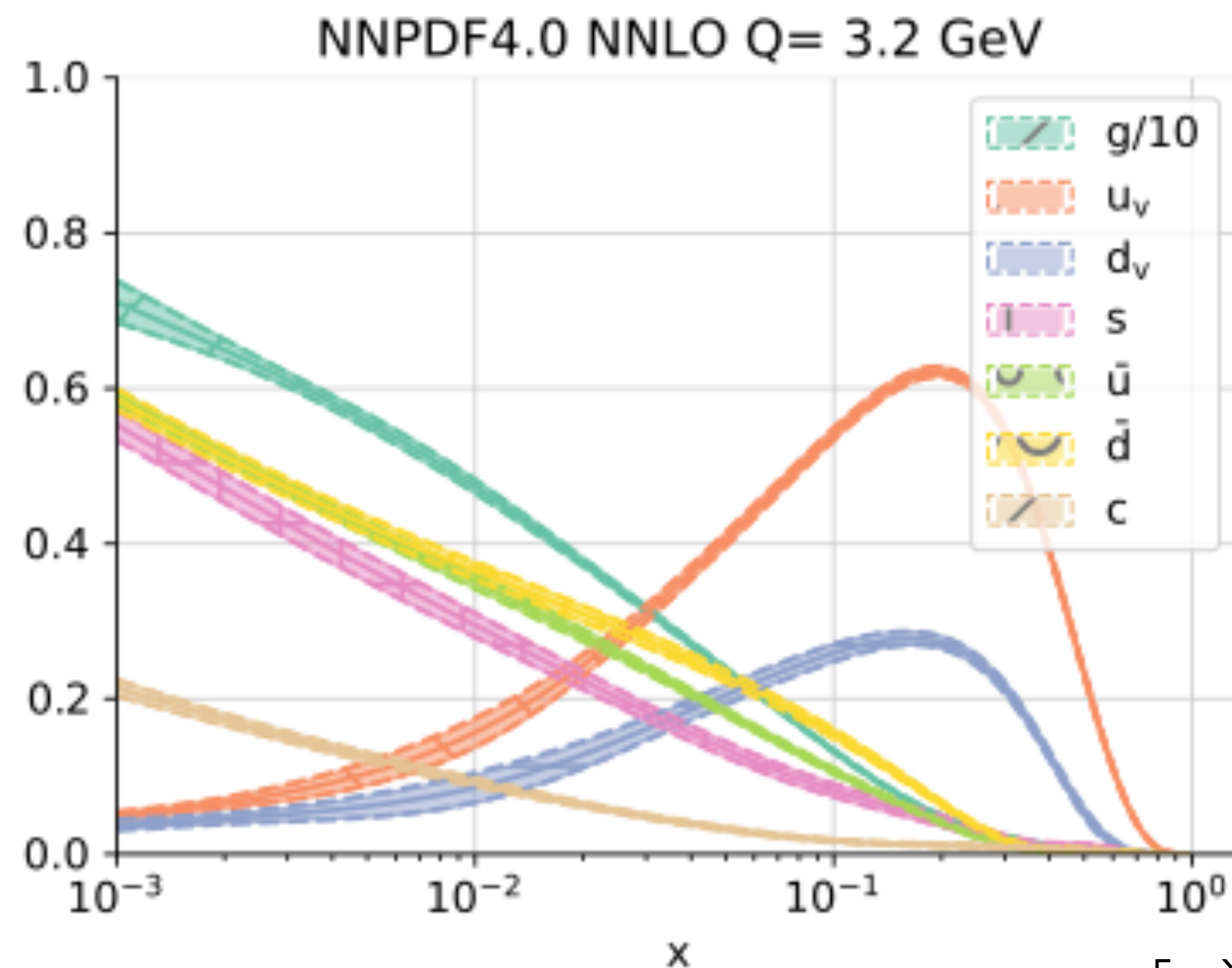
European Research Council
Established by the European Commission

NNPDF

PBSP 

Parton Distribution Functions and their uncertainties

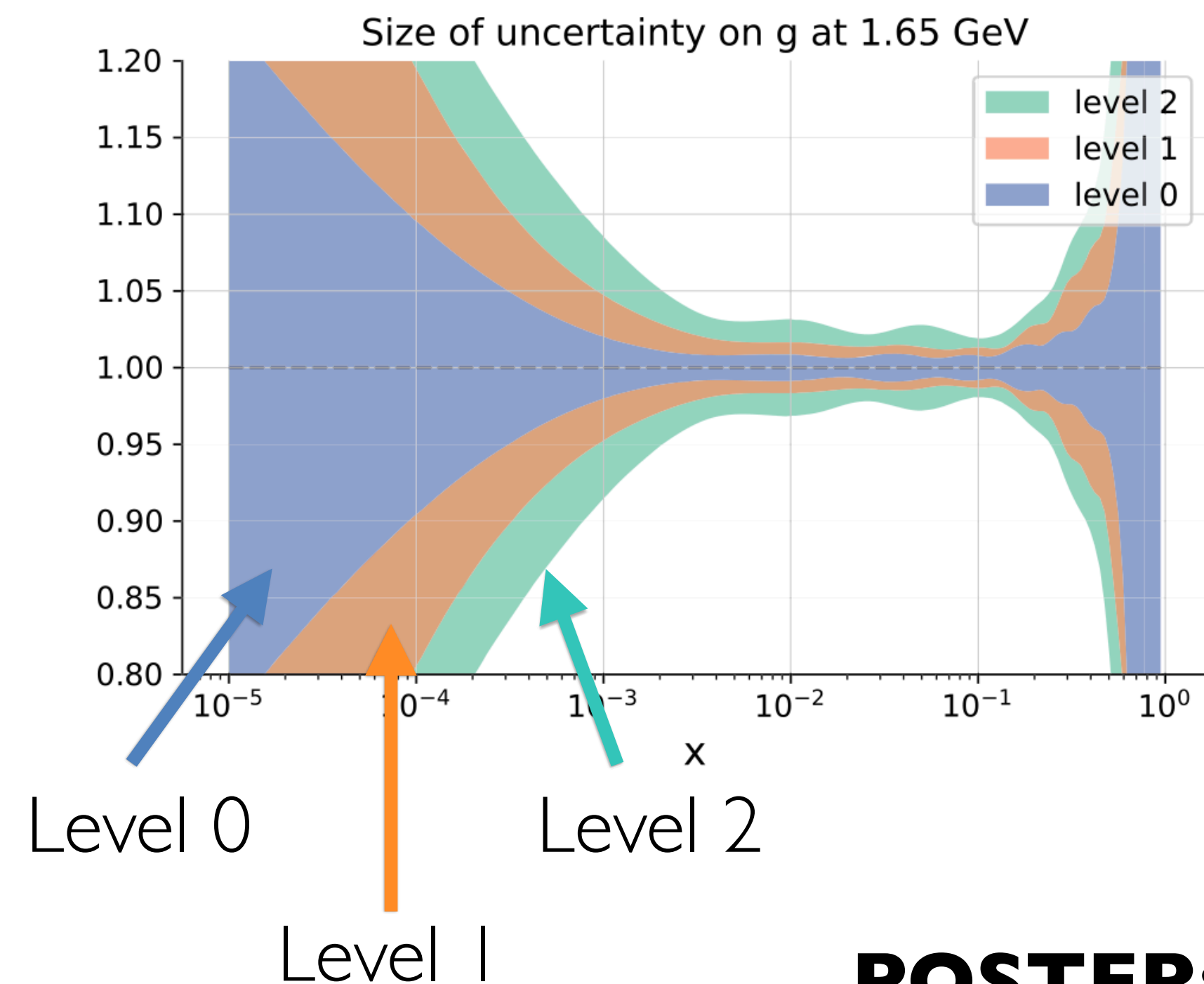
- PDFs parameterise the structure of the proton in terms of its subnuclear constituents
- NNPDF: PDF model is parameterised by a dense Neural Network and uncertainties are estimated through the Monte Carlo replica method [arXiv: 2404.10056]
- Fitted from data from several experimental sources → several sources of inconsistencies can affect fit



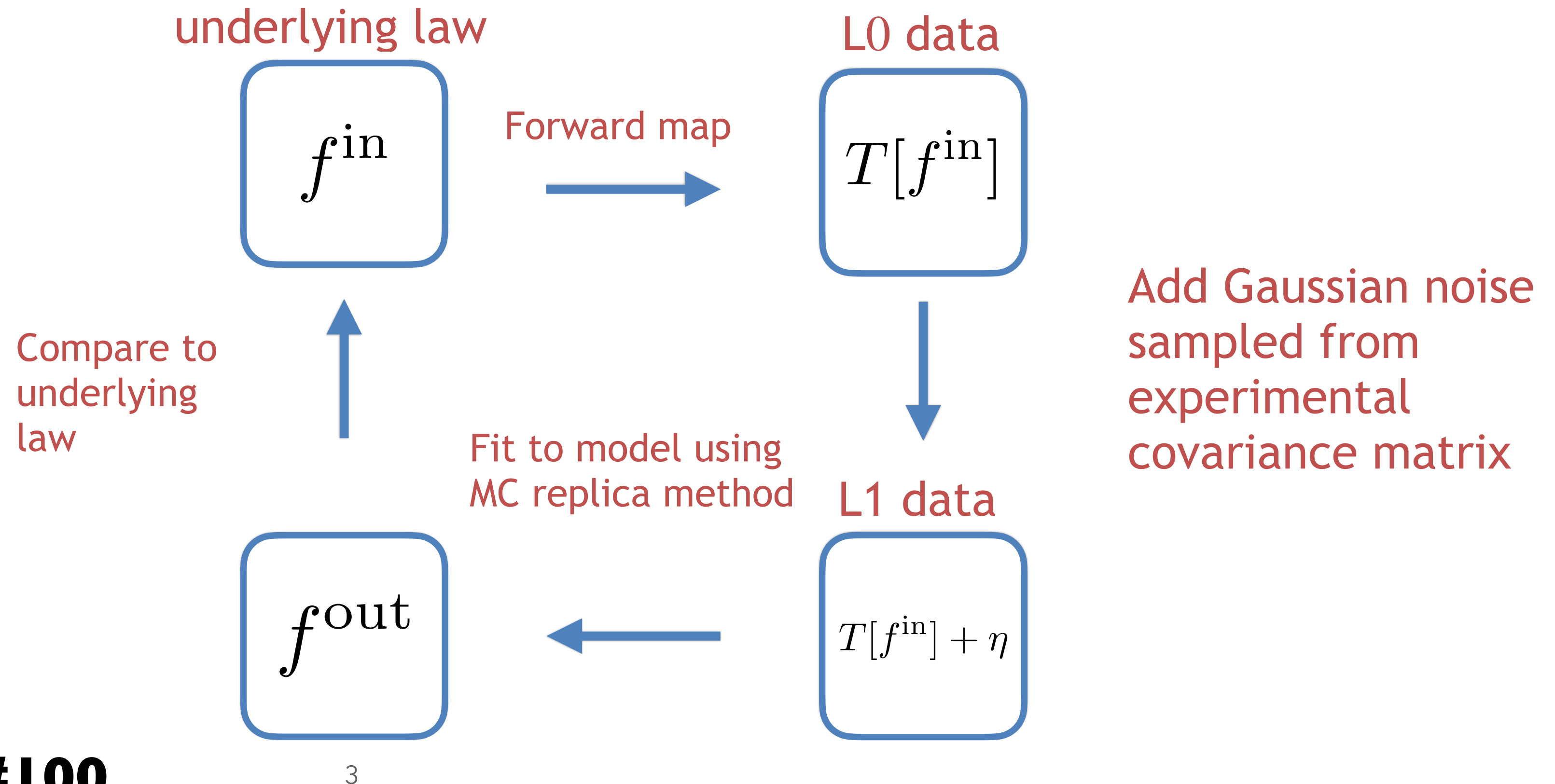
POSTER: #100

Inconsistent training data

- Inconsistency is defined as a missing or underestimated experimental uncertainty
- Closure test: powerful framework to test fitting methodology [arXiv: 2111.05787]



POSTER: #100





**Università
di Genova**

Evaluating Generative Models with non-parametric two-sample tests

Samuele Grossi

Amsterdam, EuCAIFCon, 30th April 2024

Based on work in collaboration with Riccardo Torre

Two-sample test in high energy physics

Two-Sample Test: understand if two independent data samples are drawn from the same probability density function (PDF)

- **PARAMETRIC:** Some assumptions on the underlying distributions of the samples are needed to perform the test
- **NON-PARAMETRIC:** Only the data are used to perform the test, without any assumption on the underlying distributions

In high energy physics: two-sample test to compare data provided by two different generators.
Example: Powheg/MadGraph vs Neural Networks

Objectives and procedure

Purpose of the work: provide a systematic analysis of non-parametric two-sample test using different evaluation metrics

- Use reference distributions to :
 - Extract toy samples to build null hypothesis distributions
 - Perform the likelihood-ratio test
- Carry out non-parametric tests using different metrics. Compare performances to the likelihood-ratio

Particle physics  high dimensional datasets

We focused on *univariate integral probability measure* metrics:
easy to implement, fast results, scale well with the dimensions

Estimation of ML model uncertainty in particle physics event classifiers

Miguel Cárdenas Montes¹ miguel.cardenas@ciemat.es

José M. Hernández¹ jose.hernandez@ciemat.es Julia Vázquez Escobar¹ julia.vazquez@ciemat.es

¹Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT)

EuCAIFCon 2024

May 30, 2024



GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES

Ciemat

Centro de Investigaciones
Energéticas, Medioambientales
y Tecnológicas



[Julia Vázquez-Escobar, J.M. Hernández, Miguel Cárdenas-Montes, Computer Physics Communications, 268, \(2021\)](#)

Methods

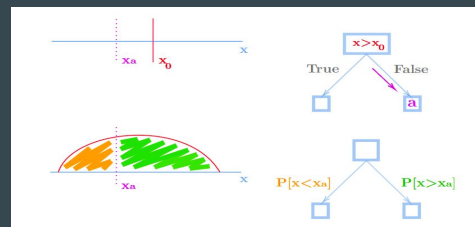
Bayesian approx.

$$\tilde{\mathbb{E}}[\mathbf{y}^*] := \frac{1}{T} \sum_{t=1}^T \mathbf{f}^{\hat{\omega}_t}(\mathbf{x}^*) \xrightarrow{T \rightarrow \infty} \mathbb{E}_{q_{\theta^*}(\mathbf{y}^*|\mathbf{x}^*)}[\mathbf{y}^*]$$

$$\widetilde{\text{Var}}[\mathbf{y}^*] := \tau^{-1} \mathbf{I} + \frac{1}{T} \sum_{t=1}^T \mathbf{f}^{\hat{\omega}_t}(\mathbf{x}^*)^T \mathbf{f}^{\hat{\omega}_t}(\mathbf{x}^*) -$$

$$- \tilde{\mathbb{E}}[\mathbf{y}^*]^T \tilde{\mathbb{E}}[\mathbf{y}^*] \xrightarrow{T \rightarrow \infty} \text{Var}_{q_{\theta^*}(\mathbf{y}^*|\mathbf{x}^*)}[\mathbf{y}^*]$$

Probabilistic RF

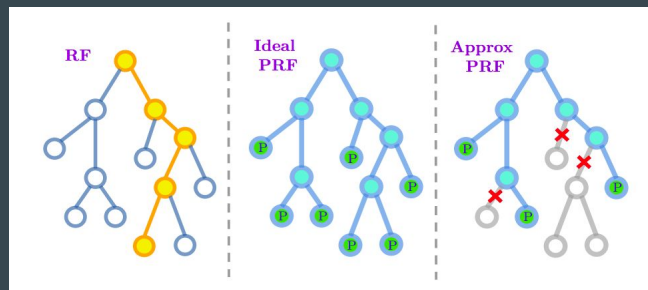


Local ensembles

Proposition 1.1 Let Δ_{θ} be the projection of a random perturbation with mean zero and covariance proportional to the identity $\epsilon \cdot \mathbf{I}$ into the ensemble subspace spanned by $\{\xi_{(j)} : j > m\}$. Let P_{Δ} be the linearized change in prediction induced by the perturbation

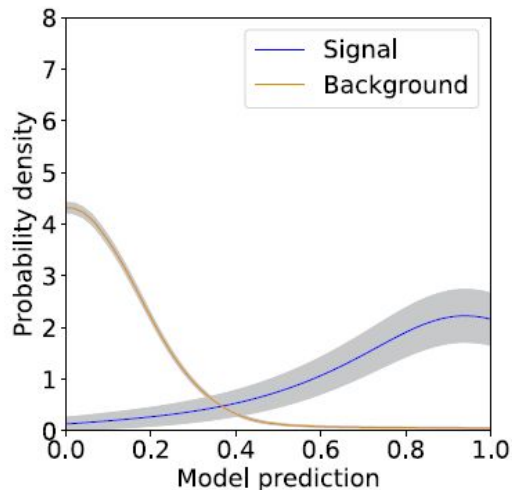
$$P_{\Delta}(x') := g_{\theta^*}(x')^{\top} \Delta_{\theta} \approx \hat{y}(x', \theta^* + \Delta_{\theta}) - \hat{y}(x', \theta^*)$$

then $\mathcal{E}_m(x') = \epsilon^{-1/2} \cdot SD(P_{\Delta}(x'))$.

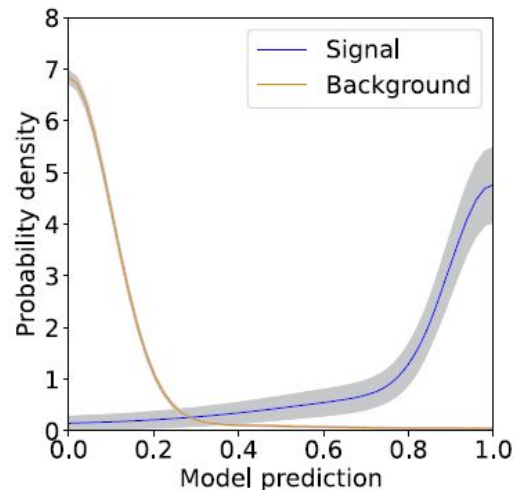


Results

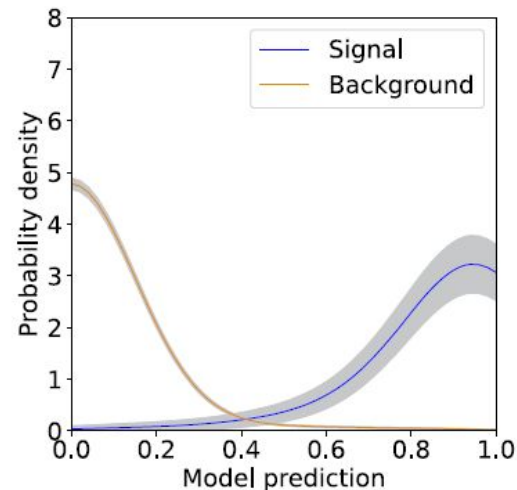
The probability density functions of the classification parameter for true signal and background events are shown.



(a) Local Ensembles



(b) Probabilistic Random Forest



(c) Bayesian approximation

Results

Model	AUC
Probabilistic Random Forest	0.969 ± 0.005
Local Ensembles	0.951 ± 0.006
Bayesian Approximation	0.990 ± 0.001

