The NPLM method:
ML for signal-agnostic searches and beyond

Marco Letizia

Machine Learning Genoa Center – Università di Genova

INFN – Sezione di Genova

Main collaborators: G. Grosso, M. Pierini, L. Rosasco, A. Wulzer, M. Zanetti

Based on: arXiv:2204.02317, arXiv:2303.05413, arXiv:2305.14137, arXiv:2408.12296

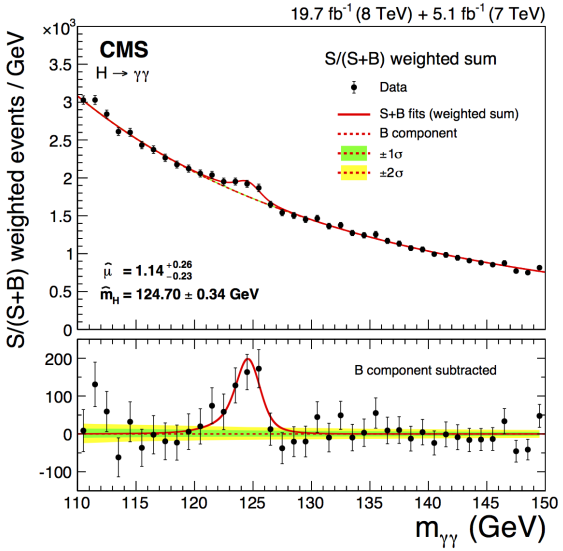
Problem definition

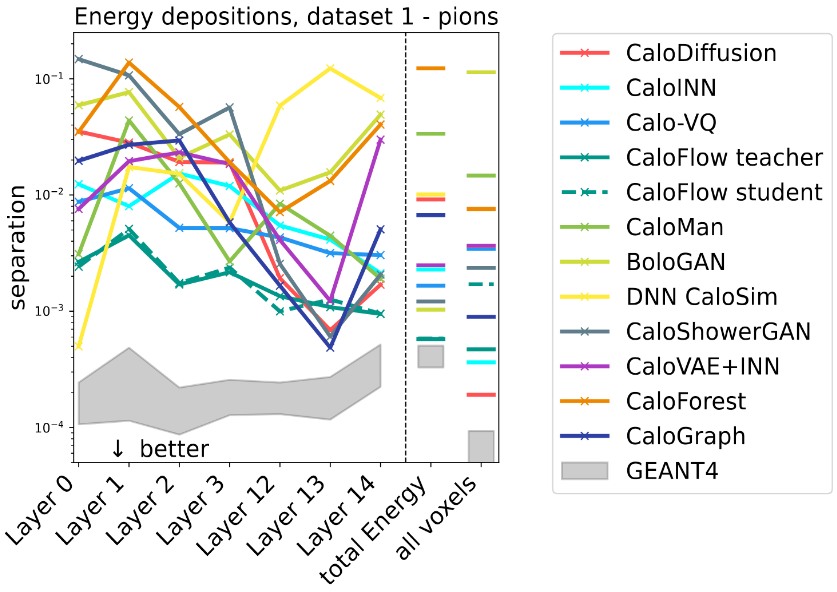
Assess the goodness of a (statistical) model against observations

Goodness-of-fit

Obiquitus problem in HEP:

- Data analysis
- Validation of simulators
- Data quality monitoring
- …





arXiv:2410.21611, Calochallenge (2022)

Problem definition

Assess the goodness of a (statistical) model against observations

Goodness-of-fit

Models are often not known in analytical form but they can be sampled

Two-sample testing

Outline

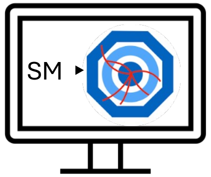
Two-sample testing

The New Physics Learning Machine

Examples

Multiple testing

Two-sample testing



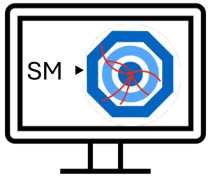
$\mathcal{D} = \{x_1, \dots, x_n\}, \qquad x \sim p_{\text{true}}$

$\mathcal{R} = \{\tilde{x}_1, \dots, \tilde{x}_m\}, \qquad \tilde{x} \sim p_{\text{R}}$

$(m \gg n)$

$H_0\colon p_{\text{true}} = p_{\text{R}}$ vs $H_1\colon p_{\text{true}} \neq p_{\text{R}}$

Two-sample testing



$\mathcal{D} = \{x_1, \dots, x_n\}, \qquad x \sim p_{\text{true}}$

$\mathcal{R} = \{\tilde{x}_1, \dots, \tilde{x}_m\}, \qquad \tilde{x} \sim p_{\text{R}}$

$(m \gg n)$

signal-aware $H_0\colon\, p_{\text{true}} = p_{\text{R}}$ vs $H_1\colon\, p_{\text{true}} = p_{\text{alt}}$, summary $z = f(x)$

$$t(\mathcal{D}) = \sum_{z \in \mathcal{D}} \log \frac{\bar{p}_{\text{alt}}(z)}{\bar{p}_{\text{R}}(z)}$$

Two-sample testing

This is a hard problem

- Large sample size $n = \mathcal{O}(10^k)\,,\ k \geq 3$
- Multivariate $x \in \mathbb{R}^d,\ d > 3$

Some traditional approaches:

- Combine univariate tests (e.g., Kolmogorov-Smirnov) $\rightarrow$ correlations
- Binning (e.g., χ^2) $\rightarrow$ curse of dimensionality

Outline

Two-sample testing

The New Physics Learning Machine

Examples

Multiple testing

The New Physics Learning Machine

Idea: data-driven alternative

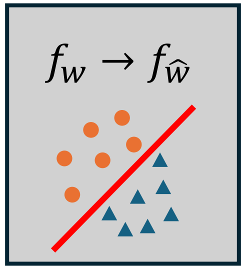
$$p_w(x) = e^{f_w(x)} p_{\mathrm{R}}(x) \quad \text{s.t.} \qquad p_{\hat{w}}(x) \approx p_{\mathrm{true}}(x)$$

$\mathcal{F} = \{f_w\}$ rich and $\hat{w}$ selected from the data (e.g., max-likelihood)

$$\rightarrow \quad t_{\hat{w}}(x) = \log \frac{p_{\hat{w}}(x)}{p_R(x)} \approx \log \frac{p_{\mathrm{true}}(x)}{p_R(x)}$$

The New Physics Learning Machine

In practice: learn the density-ratio from data using classifiers

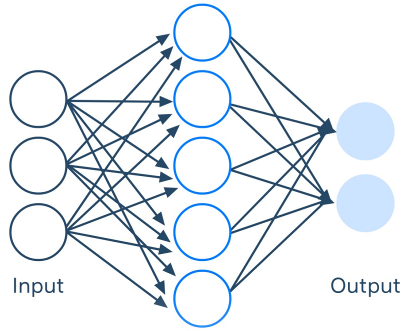


Training data $\{(x_i, y_i)\}_{i=1}^{n_{tot}}, \qquad y = \begin{cases} 0 \text{ if } x \in \mathcal{R} \\ 1 \text{ if } x \in \mathcal{D} \end{cases}$

Loss function $L(f_w) = \frac{1}{n_{tot}} \sum_i \ell(f_w(x_i), y_i) + \lambda \, \|w\|^2$

Target $f_{\hat{w}} \approx \log \frac{p_{\text{true}}(x)}{p_R(x)} \rightarrow t_{\hat{w}}(\mathcal{D}) = \sum_{x \in \mathcal{D}} f_{\hat{w}}(x)$

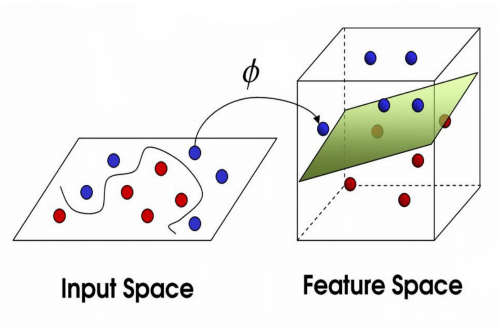
The New Physics Learning Machine



- Maximum likelihood by minimum loss with networks

D'Agnolo and Wulzer, PRD (2018)

$$\ell(f(x),y) = (1-y)\frac{N_{\mathcal{D}}}{N_{\mathcal{R}}}\left(e^{f(x)}-1\right) - y\, f(x), \qquad f_{NN}(x) = \sum_{j=1}^{u} c_j\, \sigma\left(\langle a_j, x\rangle\right)$$



- Weighted logistic regression with kernels

ML, Losapio, Rando, Grosso, Wulzer, Pierini, Zanetti, Rosasco, EPJC (2022)

$$\ell(f(x),y) = (1-y)\frac{N_{\mathcal{D}}}{N_{\mathcal{R}}}\log\left(1+e^{f(x)}\right) + y\log\left(1+e^{-f(x)}\right), \qquad f_{KM}(x) = \sum_{i=1}^{n_{tot}} w_i\, k(x, x_i)$$

The New Physics Learning Machine

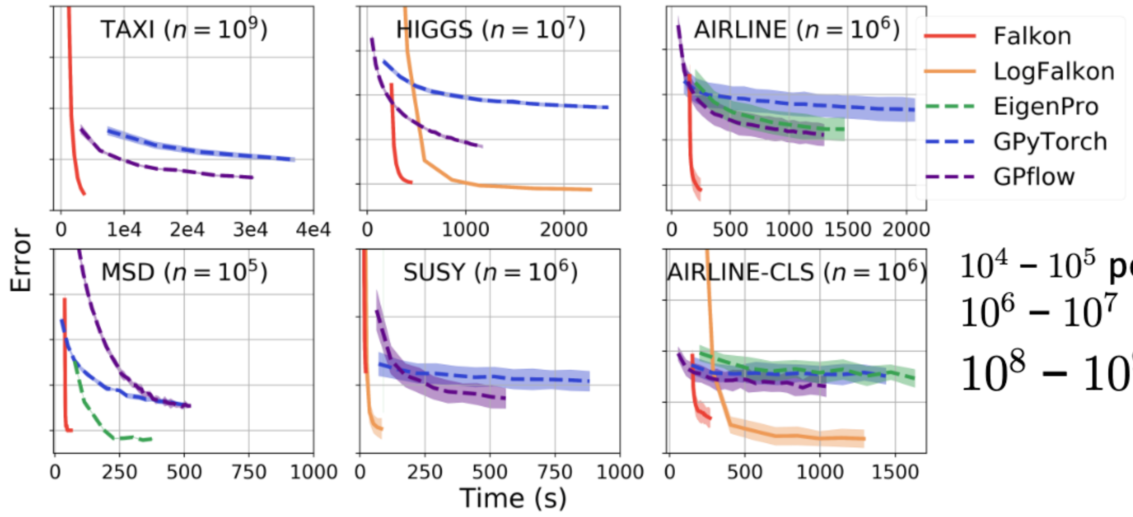
Check out
L. Rosasco's talk!

Efficient large scale kernel methods Falkon library Meanti, Carratino, Rosasco, Rudi, NeurIPS (2020)

$$f_w = \sum_{i=1}^{n} w_i \, k_\sigma(x, x_i)\,,$$

$$k_\sigma(x, x_i) = \exp - \frac{\|x - x_i\|^2}{2\sigma^2}$$

- Random projections (Nyström)
- Conjugate grad. w/ efficient preconditioning
- Efficient (multi-)GPU implementation

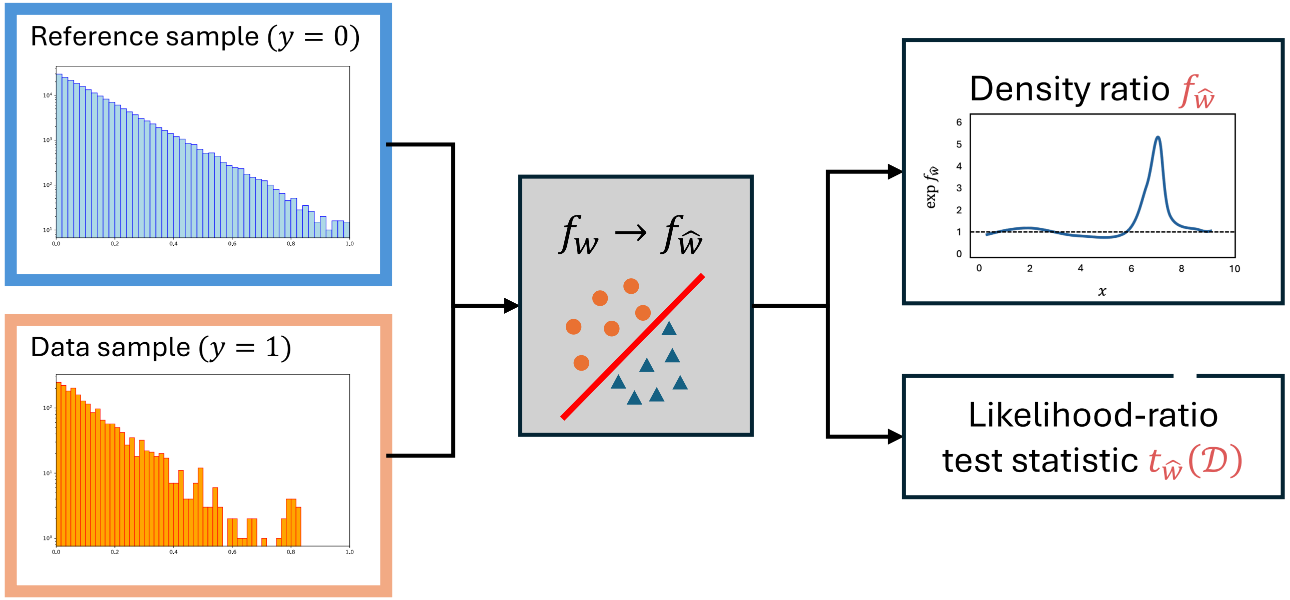


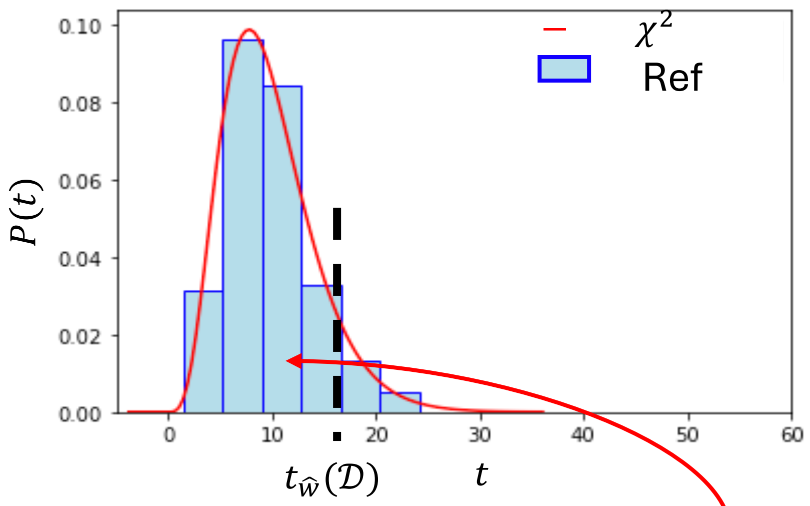
$10^4 - 10^5$ points in seconds
$10^6 - 10^7$ points in minutes
$10^8 - 10^9$ points in hours

space $\mathcal{O}(n^2) \rightarrow \mathcal{O}(n)$,

time $\mathcal{O}(n^3) \rightarrow \mathcal{O}(n\sqrt{n}\log n)$

The New Physics Learning Machine





Large $t_{\hat{w}}(\mathcal{D})$ $\to$ disagreement. How large? We need to calibrate.

- Re-train multiple times on reference-distributed pseudo-experiments (toys)

$$\begin{cases} x_1, \dots, x_n \sim p_{\mathrm{R}} \\ \tilde{x}_1, \dots, \tilde{x}_m \sim p_{\mathrm{R}} \end{cases}$$

$$\to p_{\text{value}} = \int_{t_{\hat{w}}(\mathcal{D})}^{\infty} dt\, p(t)\,, \qquad Z = \Phi^{-1}(1 - p_{\text{value}})$$

Outline

Two-sample testing

The New Physics Learning Machine

Examples

Multiple testing

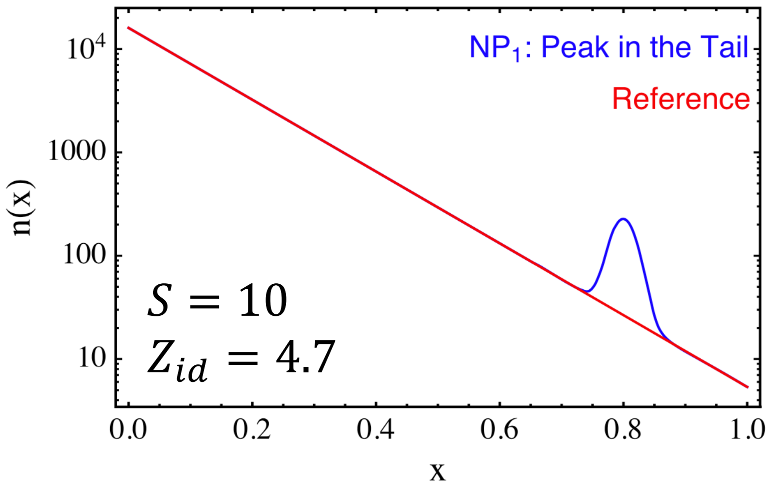
New physics searches

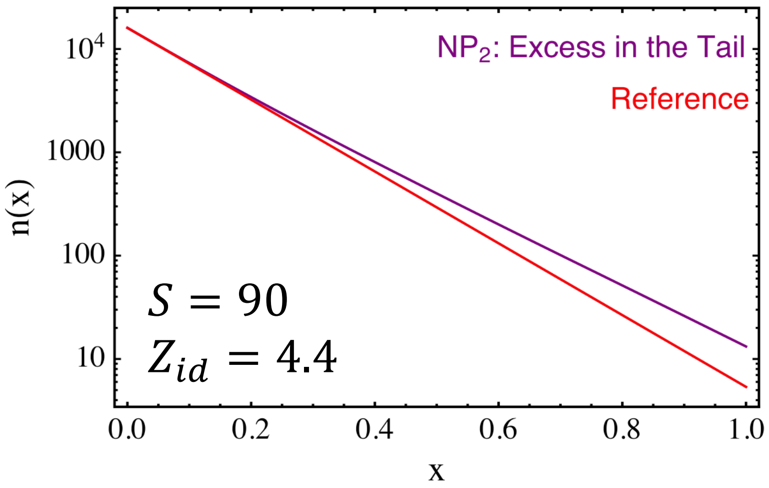
D'Agnolo, Wulzer, PRD (2018)

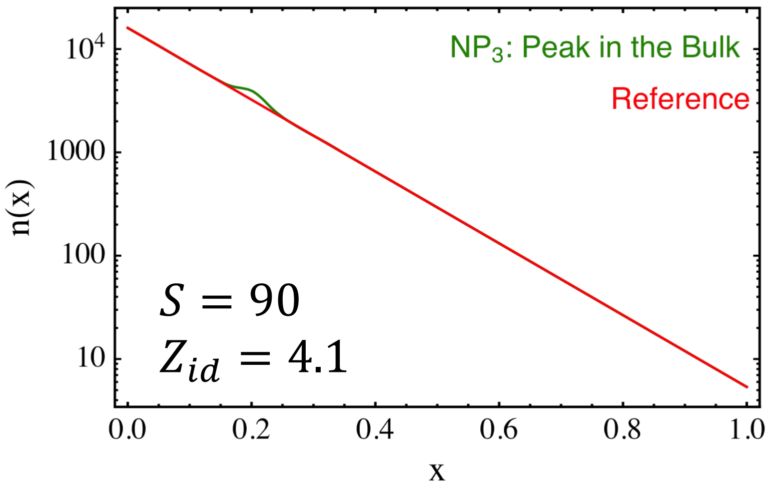
ML, Losapio, Rando, Grosso, Wulzer, Pierini, Zanetti, Rosasco, EPJC (2022)

Univariate example

$N(R) = 2000, \qquad \mathcal{N}_{\mathcal{R}} = 100 \times N(R)$

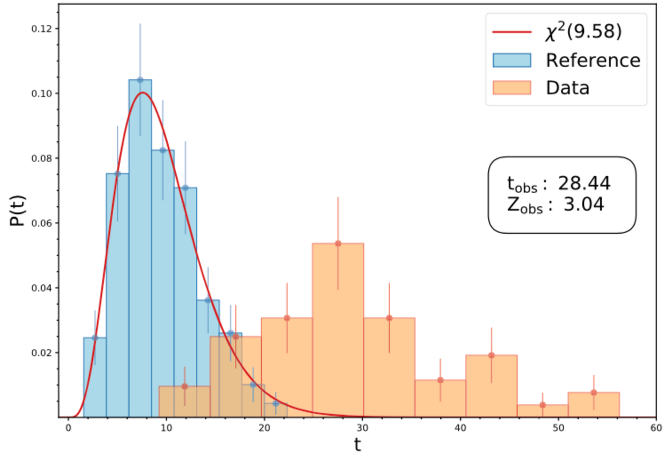


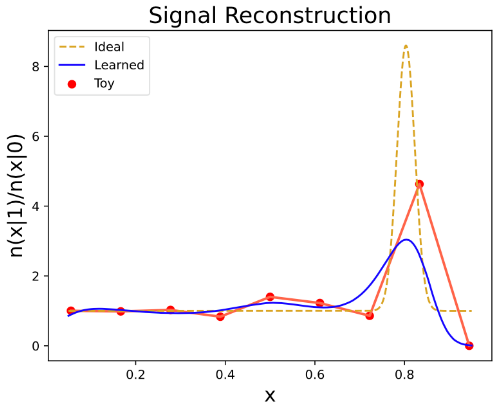


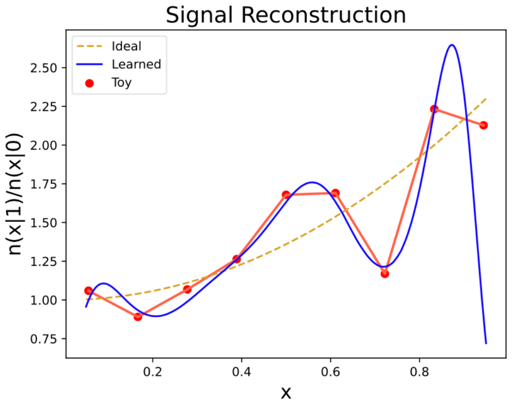


300 R-toys
100 D-toys

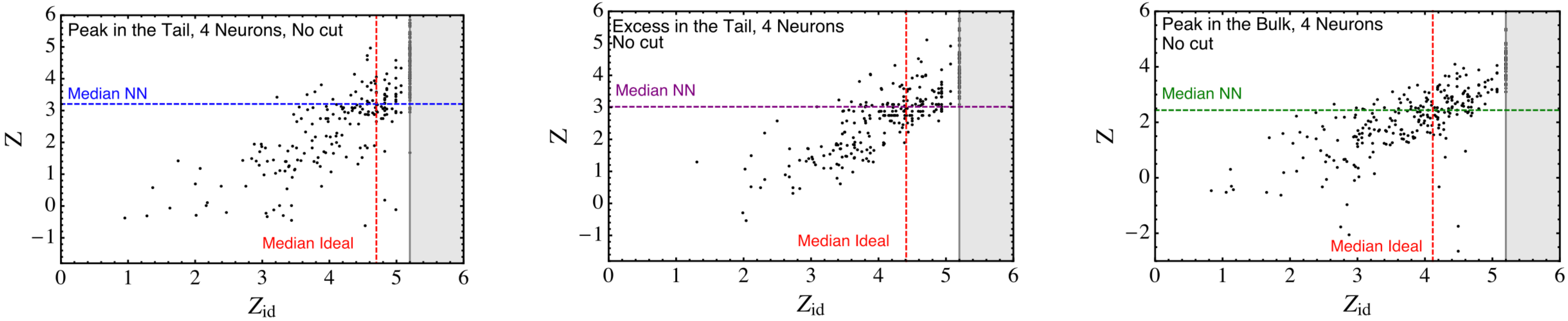
$Z_{obs} = (2.43, 3.04, 2.82)$







New physics searches



Correlation between how much tension we see, and how much there is to see.

Weakly depend on NP nature.

New physics searches

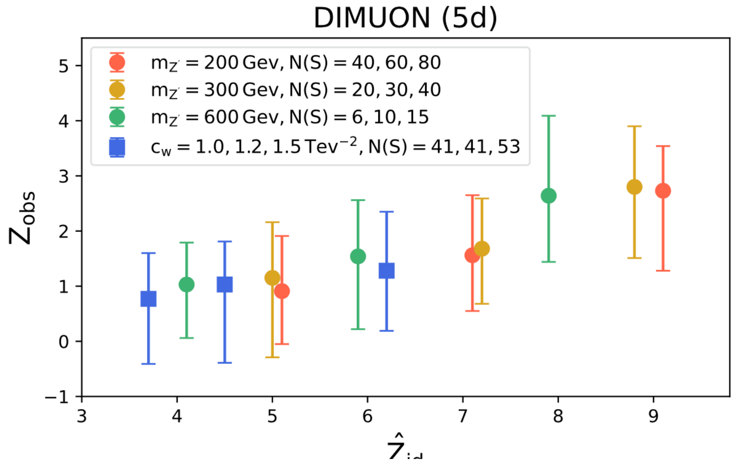
[ML, Losapio, Rando, Grosso, Wulzer, Pierini, Zanetti, Rosasco, EPJC (2022)](#)

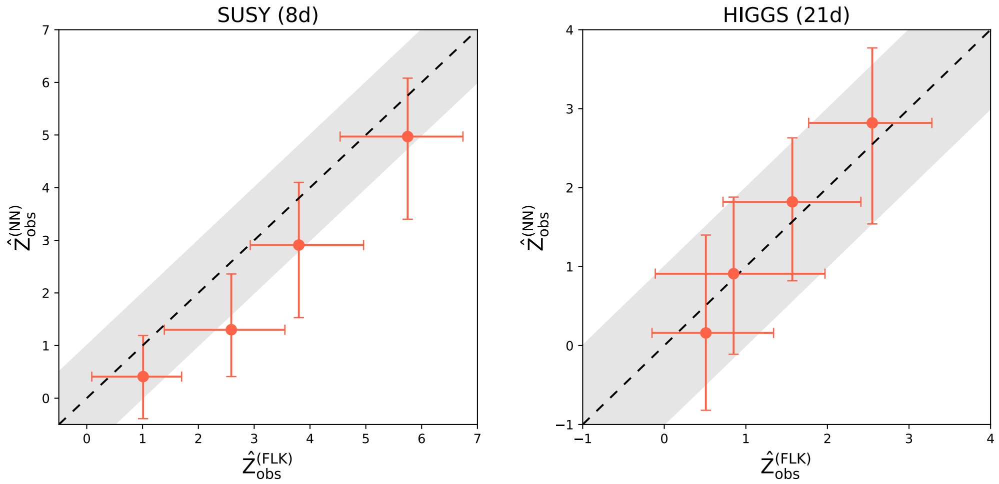
$pp \rightarrow \mu^+\mu^- \;\; [p_{T1}, p_{T2}, \eta_1, \eta_2, \Delta\phi]$,

$N(R) = 20000, \qquad \mathcal{N}_{\mathcal{R}} = 5 \times N(R)$.

SUSY (8d), HIGGS (21d)

$N(R) = 10^5, \qquad \mathcal{N}_{\mathcal{R}} = 5 \times N(R)$





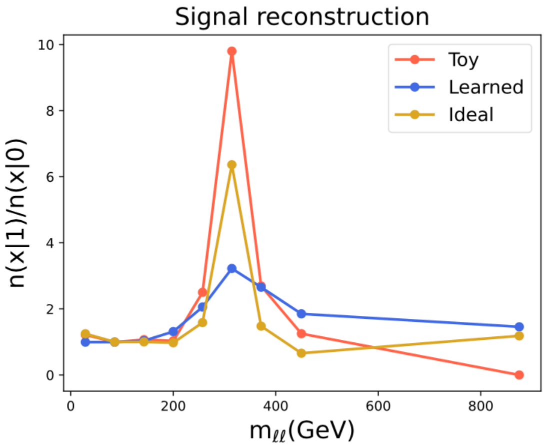


Table 1 Average training times per single run with standard deviations (low level features and reference toys). Note that time measured in hours (for NN) and seconds (for Falkon)

Model	DIMUON	SUSY	HIGGS
FLK	**(44.9 ± 3.4) s**	**(18.2 ±1.2) s**	**(22.7 ± 0.4) s**
NN	(4.23 ± 0.73) h	(73.1 ± 10) h	(112 ± 9) h

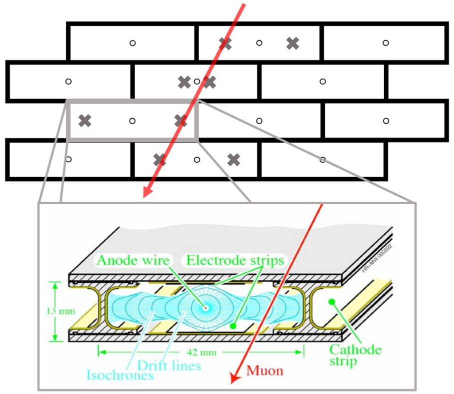
Bold values indicate the lowest for each column (lower is better)

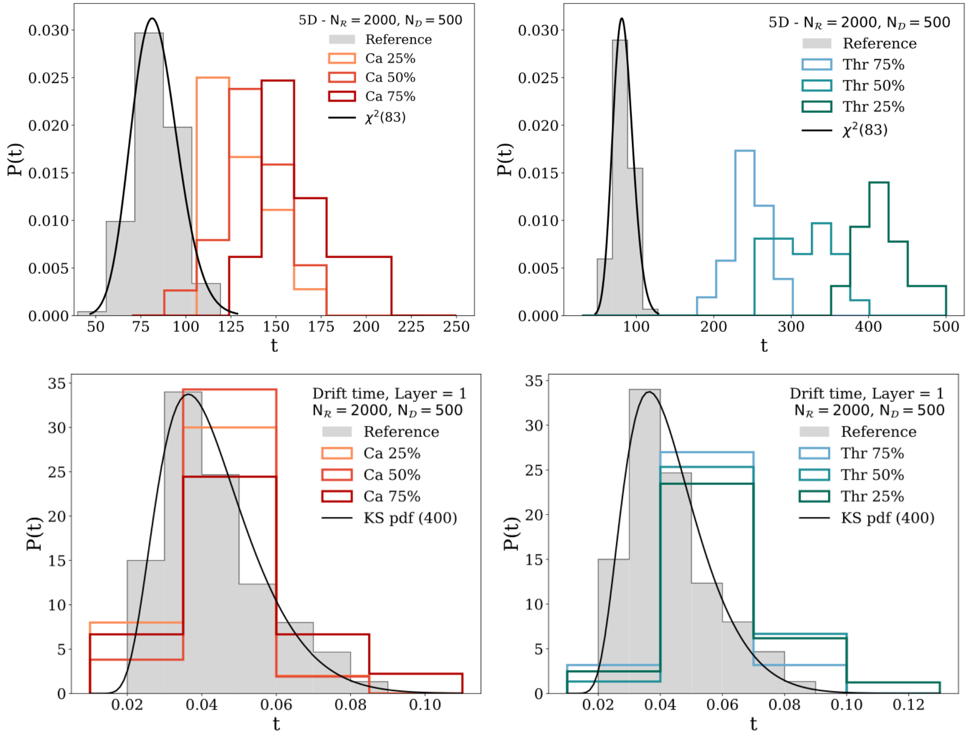
Data: https://zenodo.org/records/4442665

Data quality monitoring

[Grosso, Lai, ML, Pazzini, Rando, Rosasco, Wulzer, Zanetti, MLST (2023)](#)

Drift tube chambers from Legnaro INFN National Laboratory.





DATASET:

- Drift times (t_i): the four drift times of the muon track.
- Slope (ϕ): the angle with respect to the vertical axis.
- Reference data is collected in a controlled regime.
- Anomalies:
  - reduced voltage of cathodic strips to 75%, 50%, and 25% of their nominal value (-1.2 kV)
  - lowered front-end thresholds to 75%, 50%, and 25% of nominal value (100 mV)

$\bar{t}_{tr} \approx 0.5$ sec

Data: https://zenodo.org/records/7128223

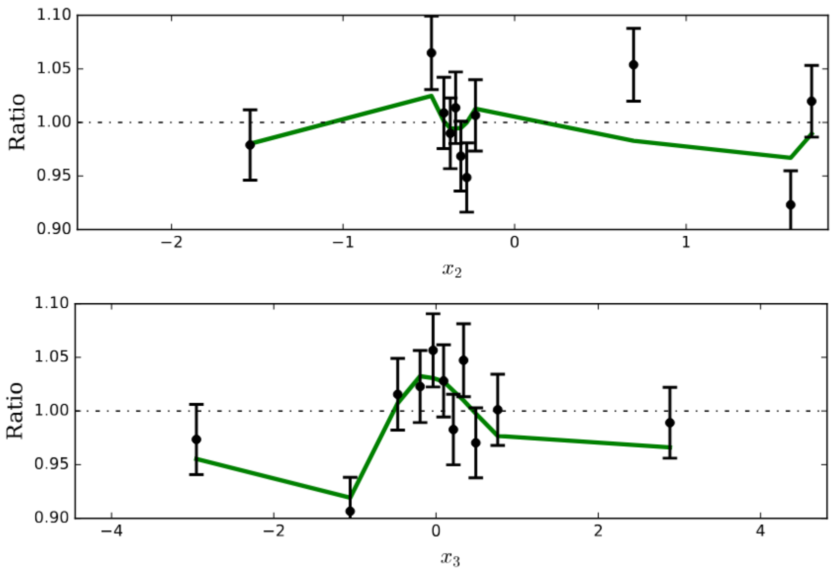
Data quality monitoring

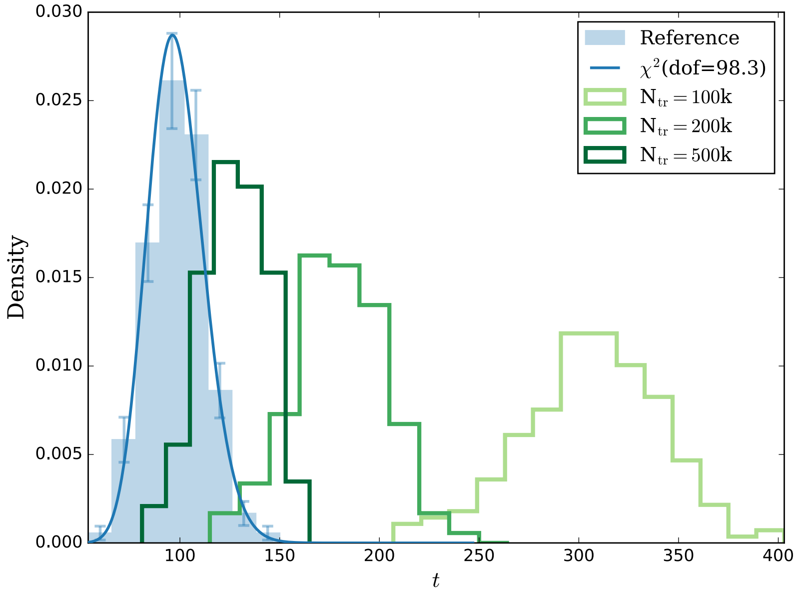
Cappelli, Grosso, Letizia, Reyes-Gonzalez, Zanetti, in preparation

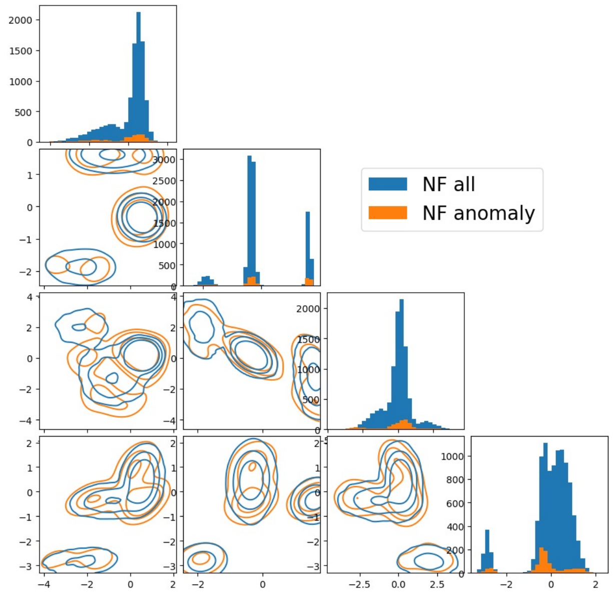
Normalising flows: RealNVP on correlated mixtures of Gaussians - $n = 10^4, m = 10^5$

PRELIMINARY

N_{tr} \ d	4	8	12	16	20	30
100k	$9.88^{+1.22}_{-1.29}$	$8.88^{+1.12}_{-1.19}$	$14.73^{+1.23}_{-0.94}$	$16.81^{+1.04}_{-1.06}$	$14.46^{+1.09}_{-0.84}$	$14.97^{+1.09}_{-0.84}$
200k	$4.79^{+1.00}_{-1.07}$	$9.90^{+0.94}_{-1.05}$	$9.56^{+1.04}_{-1.04}$	$8.34^{+0.96}_{-1.09}$	$6.45^{+0.97}_{-1.07}$	$7.32^{+0.90}_{-0.81}$
500k	$1.93^{+1.02}_{-0.99}$	$3.01^{+0.74}_{-1.13}$	$3.16^{+1.10}_{-1.02}$	$5.05^{+1.02}_{-0.99}$	$2.07^{+0.81}_{-0.97}$	$3.06^{+1.13}_{-0.86}$







Outline

Two-sample testing

The New Physics Learning Machine

Examples

Multiple testing

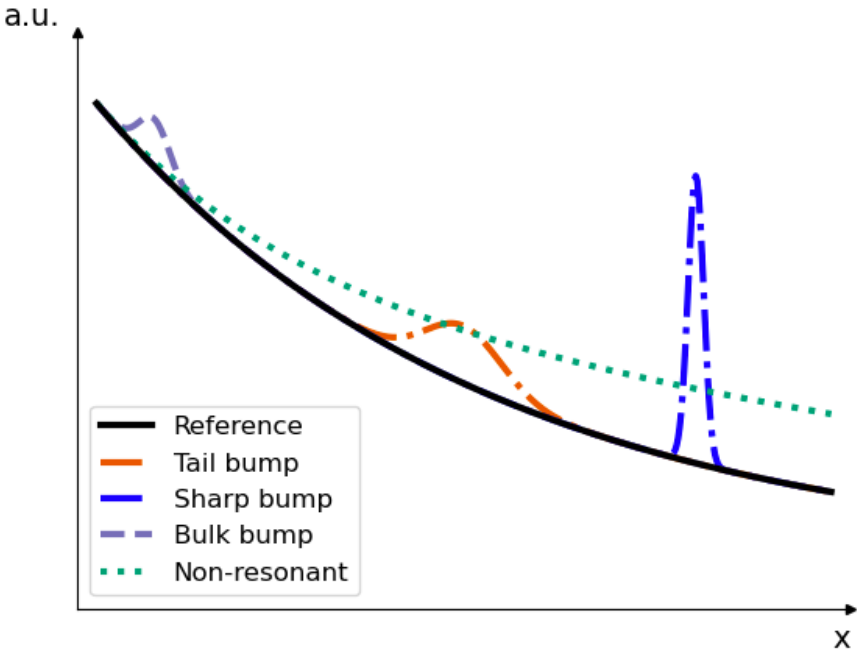
Multiple testing

Grosso, ML, EPJC (2024)

Hyperparameter tuning can favour specific signal hypotheses

$$f_{\hat{w}} = \sum_{i=1}^{n_{tot}} w_i \, k_\sigma(x, x_i) \approx \log \frac{p_{true}(x)}{p_{R(x)}},$$

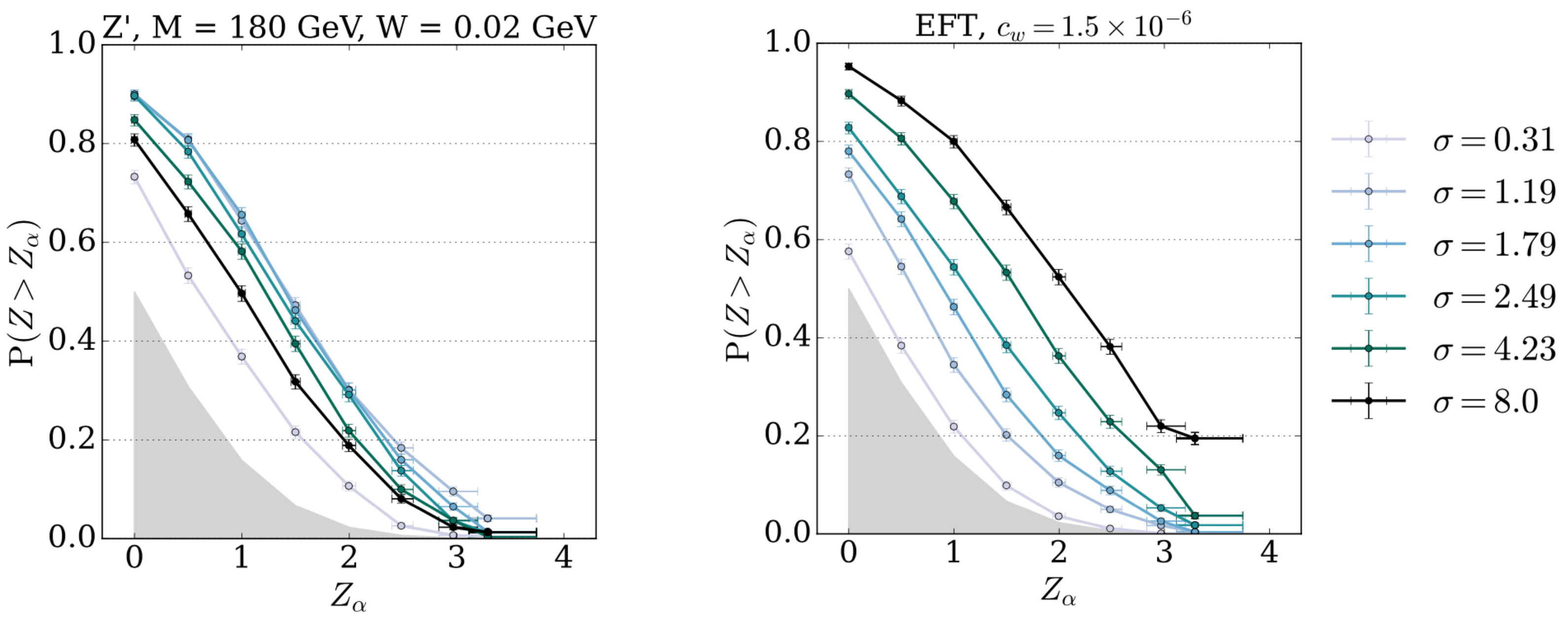
$$k_\sigma(x, x_i) = \exp \frac{\|x - x_i\|^2}{2\sigma^2}$$



Multiple testing

Grosso, ML, EPJC (2024)

Hyperparameter tuning can favour specific signal hypotheses

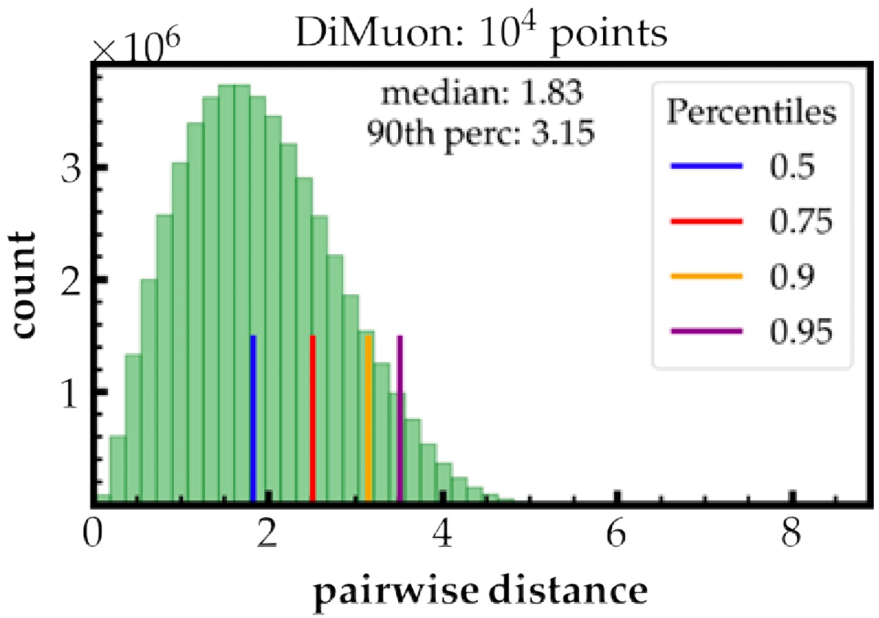


Multiple testing

Grosso, ML, EPJC (2024)

Hyperparameter tuning can favour specific signal hypotheses

→ capture average scale



Multiple testing

Grosso, ML, EPJC (2024)

Hyperparameter tuning can favour specific signal hypotheses

→ define meta-tests by aggregation

test	Z' M = 180 GeV W = 0.02 GeV	Z' M = 300 GeV W = 15 GeV	Z' M = 600 GeV W = 30 GeV	EFT $c_w = 1.5 \times 10^{-6}$
$\sigma = 0.31$	0.11 ± 0.01	0.042 ± 0.006	0.023 ± 0.005	0.036 ± 0.006
$\sigma = 1.19$	$\mathbf{0.30 \pm 0.01}$	0.35 ± 0.02	0.047 ± 0.007	0.11 ± 0.01
$\sigma = 1.79$	$\mathbf{0.30 \pm 0.01}$	0.41 ± 0.02	0.11 ± 0.01	0.16 ± 0.01
$\sigma = 2.49$	0.25 ± 0.01	$\mathbf{0.42 \pm 0.02}$	0.19 ± 0.01	0.25 ± 0.01
$\sigma = 4.23$	0.23 ± 0.01	0.41 ± 0.02	0.25 ± 0.01	0.32 ± 0.01
$\sigma = 8.0$	0.19 ± 0.01	0.31 ± 0.01	$\mathbf{0.29 \pm 0.01}$	$\mathbf{0.52 \pm 0.02}$
$\sigma = 3.0$ [9]	0.25 ± 0.02	$\mathbf{0.42 \pm 0.02}$	0.24 ± 0.02	0.30 ± 0.02
min-p	0.32 ± 0.01	0.47 ± 0.02	$\mathbf{0.28 \pm 0.01}$	$\mathbf{0.53 \pm 0.02}$
prod-p	$\mathbf{0.38 \pm 0.02}$	$\mathbf{0.53 \pm 0.02}$	0.23 ± 0.01	0.38 ± 0.02
avg-p	0.37 ± 0.02	0.46 ± 0.02	0.18 ± 0.01	0.31 ± 0.01
smax-t	0.11 ± 0.01	0.042 ± 0.006	0.023 ± 0.005	0.036 ± 0.006
HB	0.255 ± 0.003	0.40 ± 0.01	0.22 ± 0.01	0.46 ± 0.02

Table 4. MUMU 5D – $P(Z > 2)$ for different types of new physics signals. The last value of σ follows from the prescription from the original proposal [9].

Wrap up

- Efficient testing with kernel methods
- Model vs observations: analyses, DQM, generative models, MC simulators, …

- Development of library/tool https://github.com/mletizia/FalkonNPLM_1D; https://github.com/FalkonHEP
- Ongoing CMS analysis
- Evaluation of generative models Cappelli, Grosso, Letizia, Zanetti, in preparation
- Systematic uncertainties D'Agnolo, Grosso, Pierini, Wulzer, Zanetti, EPJC (2022)
- High dimensions and feature learning