# Efficient machine learning for model-independent tests

Marco Letizia

Machine Learning Genoa Center – University of Genoa

w/ G. Grosso, M. Pierini, L. Rosasco, A. Wulzer, M. Zanetti

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

# Motivation

Traditional searches in HEP are designed around specific BSM models or signal hypotheses

→ develop signal-agnostic analysis strategies

detect *generic* discrepancies from the SM

Challenging task

- New physics effects are small or rare

→ Large scale (sensitivity)

→ Multivariate (inclusive)

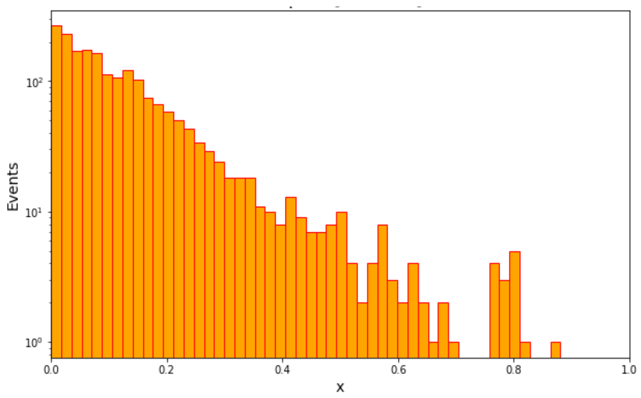
→ machine learning for a flexible and data-driven approach

# Anomaly detection as a goodness-of-fit

- Data

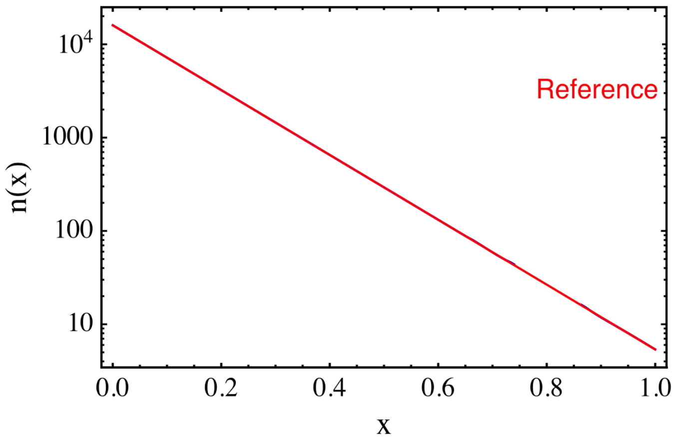
$\mathcal{D} = \{x_1, \dots, x_{N_\mathcal{D}}\}$

$x \in \mathbb{R}^d$




- Reference model

$p_R(x), \;\; N(R)$

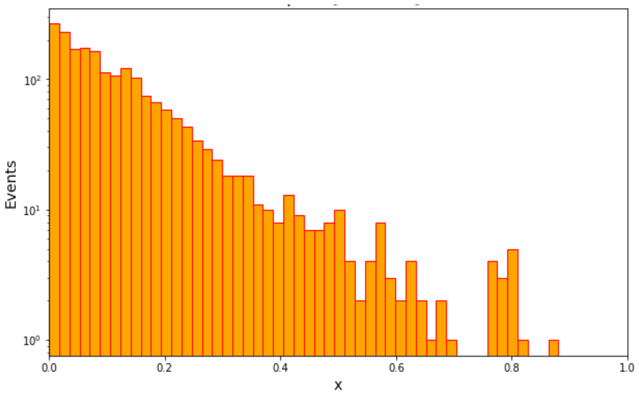


# Anomaly detection as a goodness-of-fit

- Data

$\mathcal{D} = \{x_1, \dots, x_{N_\mathcal{D}}\}$

$x \in \mathbb{R}^d$




Is the reference model a good description of the data?

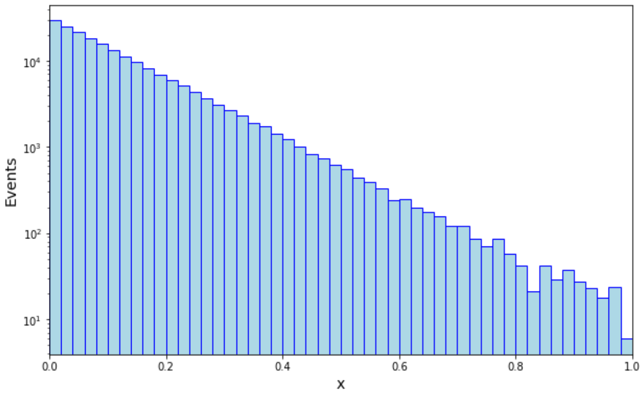
$p_{\text{true}} = p_R, \quad \text{N(true)} = N(R)$

- Reference ~~model~~ sample

$p_R(x), \quad N(R)$

$\mathcal{R} = \{\tilde{x}_1, \dots, \tilde{x}_{N_\mathcal{R}}\}$

$N_\mathcal{R} \gg N_\mathcal{D}$




Use cases:

- Signal-agnostic searches
- Data quality monitoring
- Validation of generative models

# The New Physics Learning Machine

D'agnolo and Wulzer, PRD (2019)
ML, Losapio, Rando, Grosso, Wulzer, Pierini, Zanetti, Rosasco, EPJC (2022)

## Machine learning for a signal-agnostic likelihood-ratio test

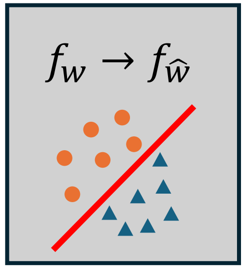
1. Likelihood-ratio test statistic

$$\mathcal{L}(\mathcal{D}|\mathrm{R}) = \frac{e^{-N(R)}}{N_{\mathcal{D}}!} \prod_{x\in\mathcal{D}} n_R(x)\,, \qquad n_R(x) = N(R)\, p_R(x)$$

$$\rightarrow\; t(\mathcal{D}) = 2\log\frac{\mathcal{L}(\mathcal{D}|\mathrm{true})}{\mathcal{L}(\mathcal{D}|\mathrm{R})} = 2\left[\sum_{x\in\mathcal{D}} f(x) - \underbrace{\frac{N(R)}{\mathcal{N}_R}\sum_{x\in\mathcal{R}}\left(e^{f(x)}-1\right)}_{N(\mathrm{true})-N(R)}\right], \qquad f = \log\frac{n_{\mathrm{true}}(x)}{n_R(x)}$$

# The New Physics Learning Machine

2. Learn the density ratio from data → classifier




Design a loss: $f_{\hat{w}} \approx \log \frac{n_{\text{true}}(x)}{n_R(x)}$

$$\ell(f_w(x), y) = (1-y)\frac{N(R)}{N_{\mathcal{R}}}\log\left(1 + e^{f_w(x)}\right) + y\log\left(1 + e^{-f_w(x)}\right)$$

$$y = \begin{cases} 0 \text{ if } x \in \mathcal{R} \\ 1 \text{ if } x \in \mathcal{D} \end{cases}$$

$$\to t_{\hat{w}}(\mathcal{D}) = 2\left[\sum_{x\in\mathcal{D}} f_{\hat{w}}(x) - \frac{N(R)}{\mathcal{N}_R}\sum_{x\in\mathcal{R}}\left(e^{f_{\hat{w}}(x)} - 1\right)\right]$$

evaluation metric
at the end of training
(in-sample)

# The New Physics Learning Machine

Check out
L. Rosasco's talk!

3. Efficient large scale kernel methods

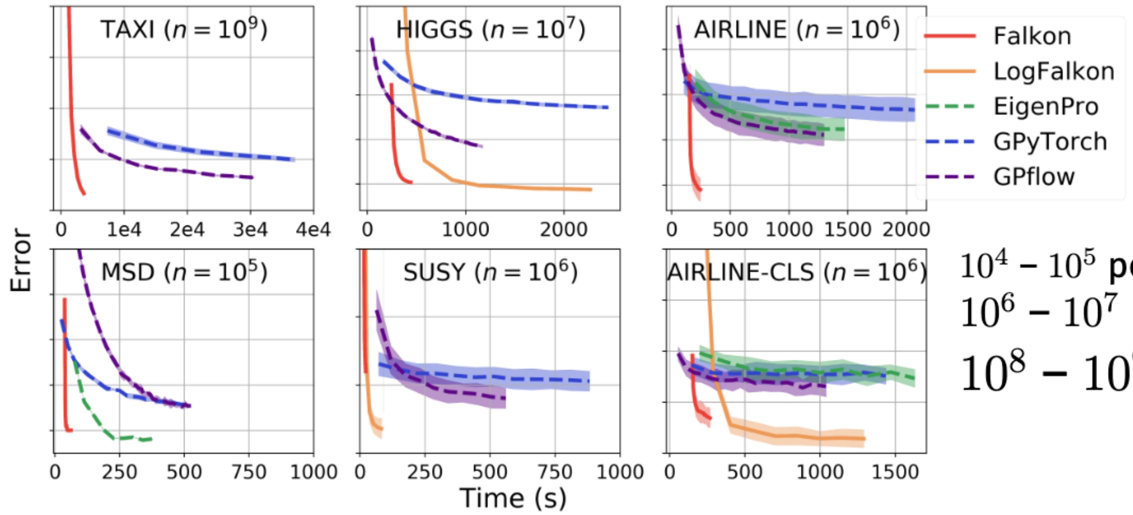
$$f_w = \sum_{i=1}^{n} w_i \, k_\sigma(x, x_i) \,,$$

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

Falkon library

Meanti, Carratino, Rosasco, Rudi, NeurIPS (2020)

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

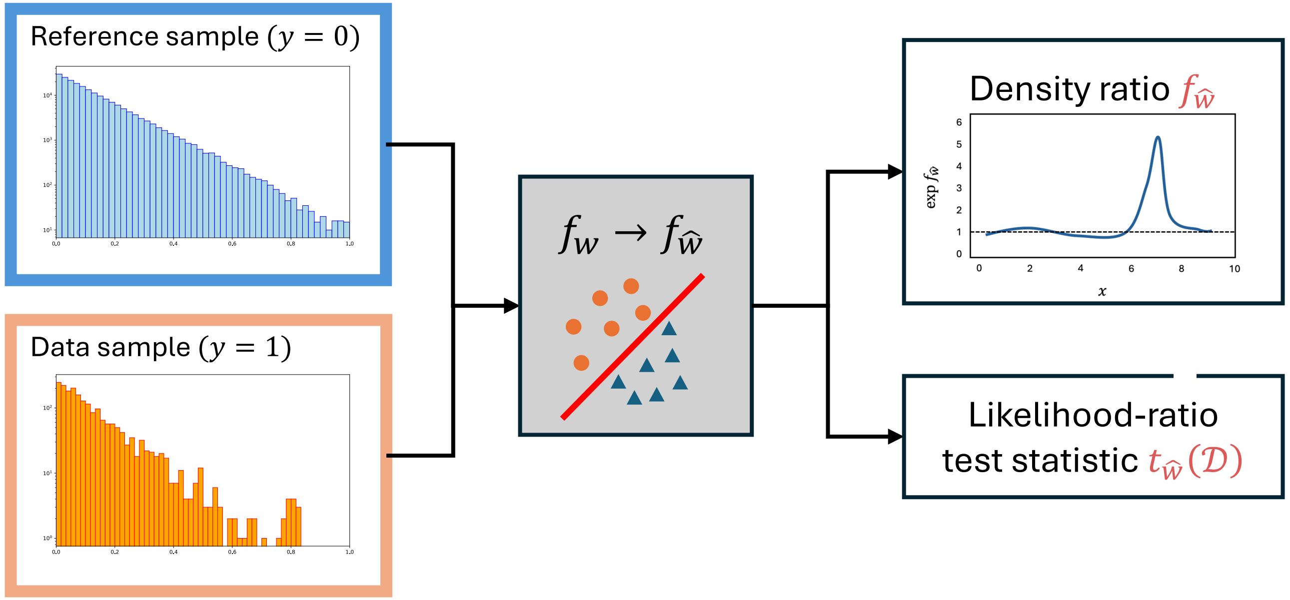


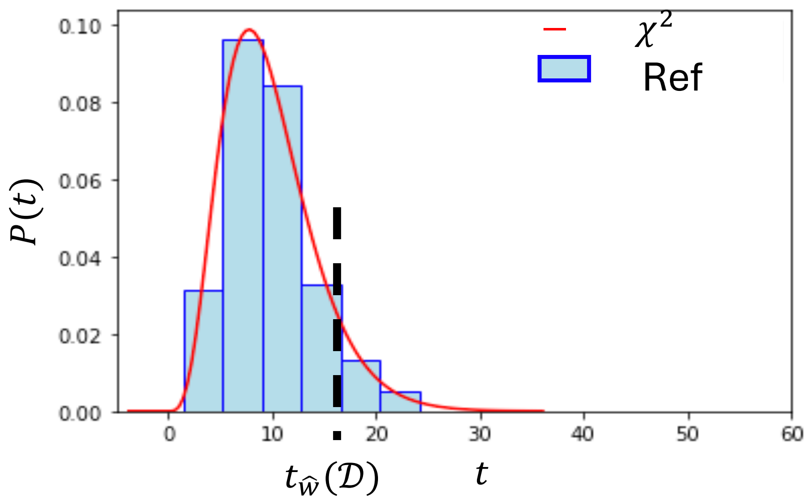

$10^4 - 10^5$ points in seconds
$10^6 - 10^7$ points in minutes
$10^8 - 10^9$ points in hours

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.

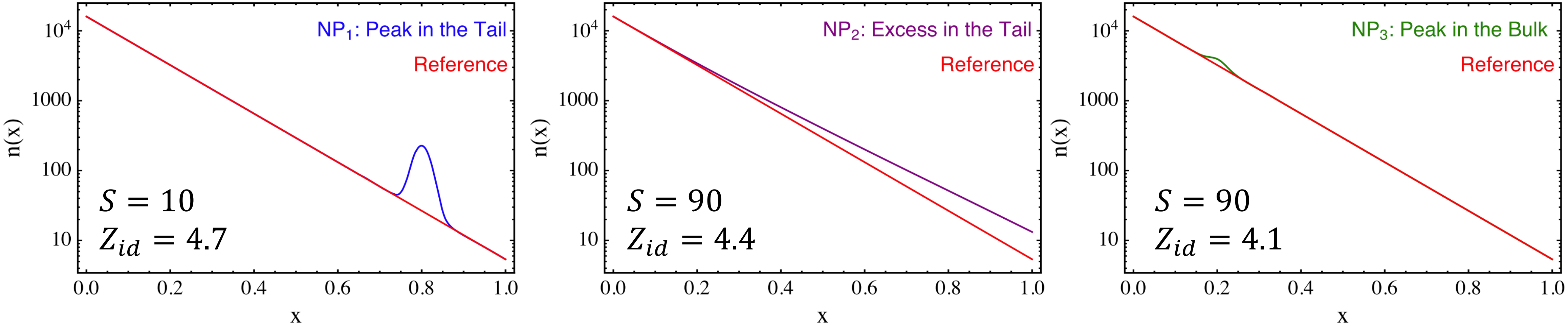
- Re-train the model multiple times on $\mathcal{D}^{(R)}$ reference-distributed data (toys).

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

# Univariate example

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

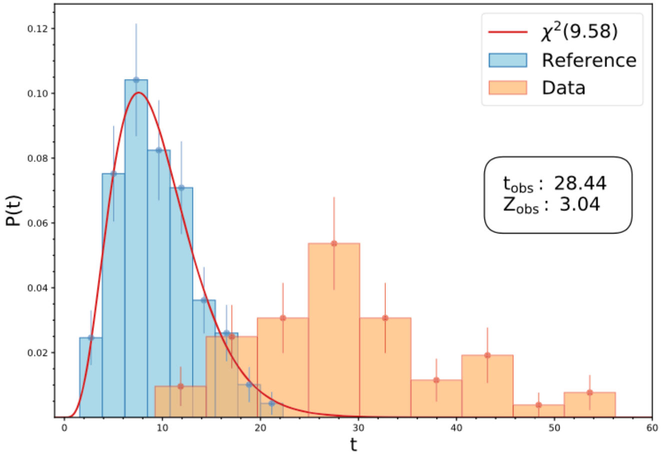
$N(R) = 2000, \quad \mathcal{N}_{\mathcal{R}} = 100 \times N(R)$

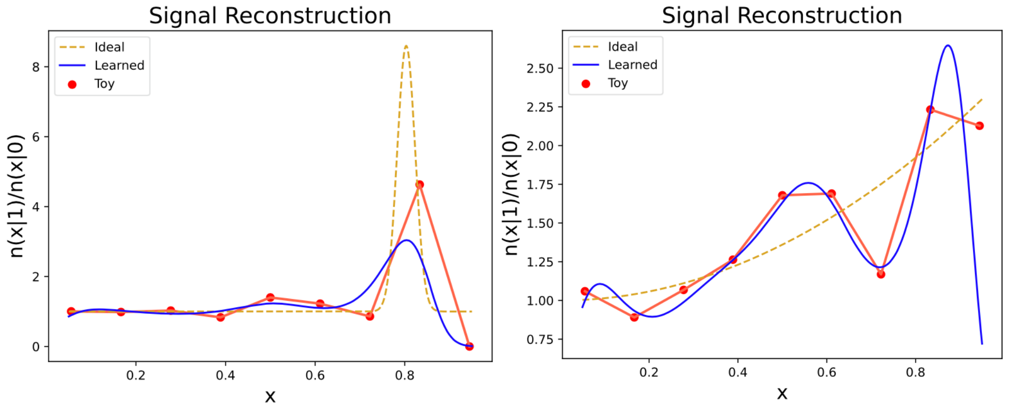



300 R-toys
100 D-toys

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

$\bar{t}_{tr} = 2.11$ sec



# Dimuon final state

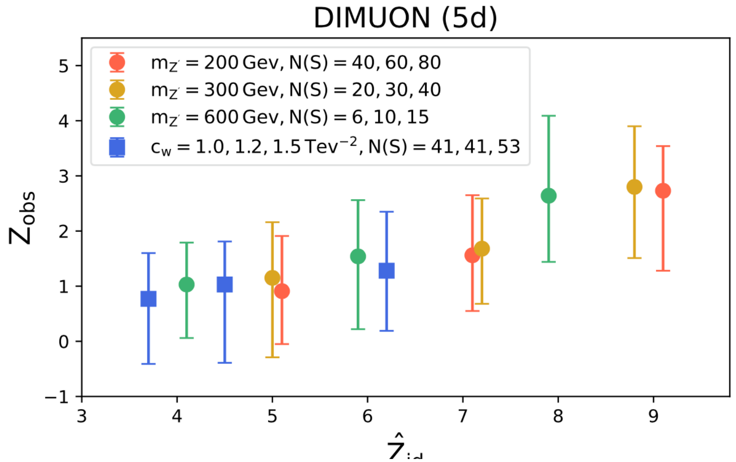
[ML, Losapio, Rando, Grosso, Wulzer, Pierini, Zanetti, Rosasco, EPJC (2022)]

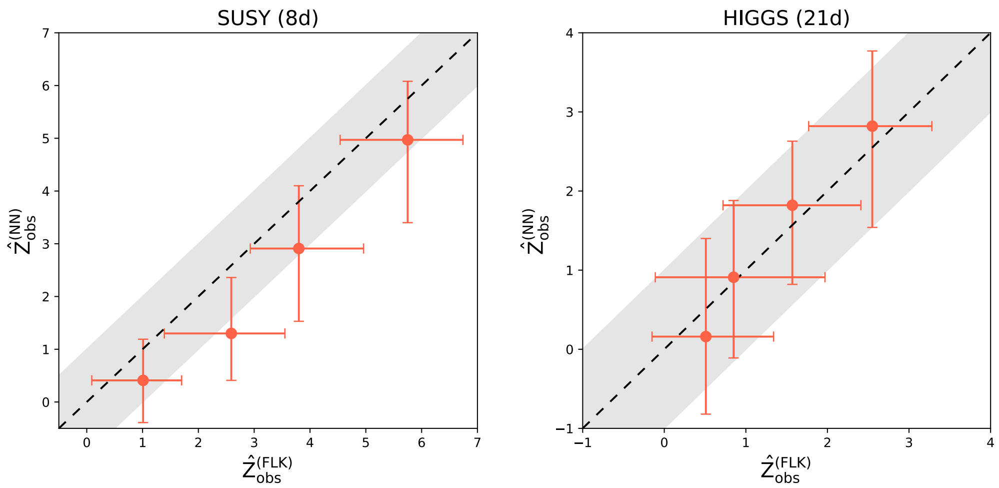
$pp \to \mu^+\mu^-\ [p_{T1}, p_{T2}, \eta_1, \eta_2, \Delta\phi],$

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

SUSY (8d), HIGGS (21d)

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



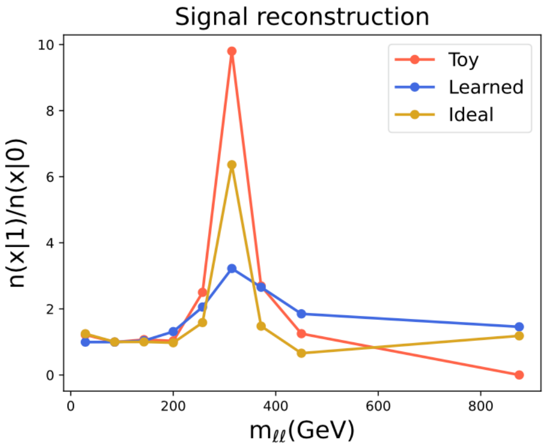




**Table 1** Average training times per single run with standard deviations (low level features and reference toys). Note that time measured in hours (for NN) and seconds (for Falkon)

| 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 |

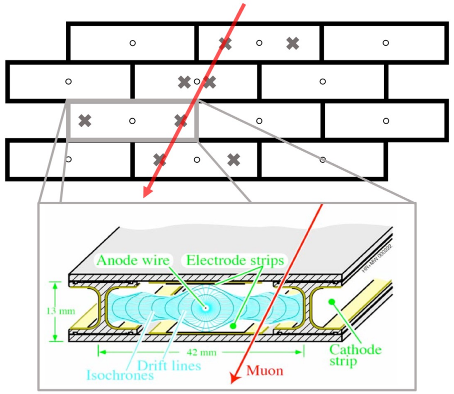
Bold values indicate the lowest for each column (lower is better)

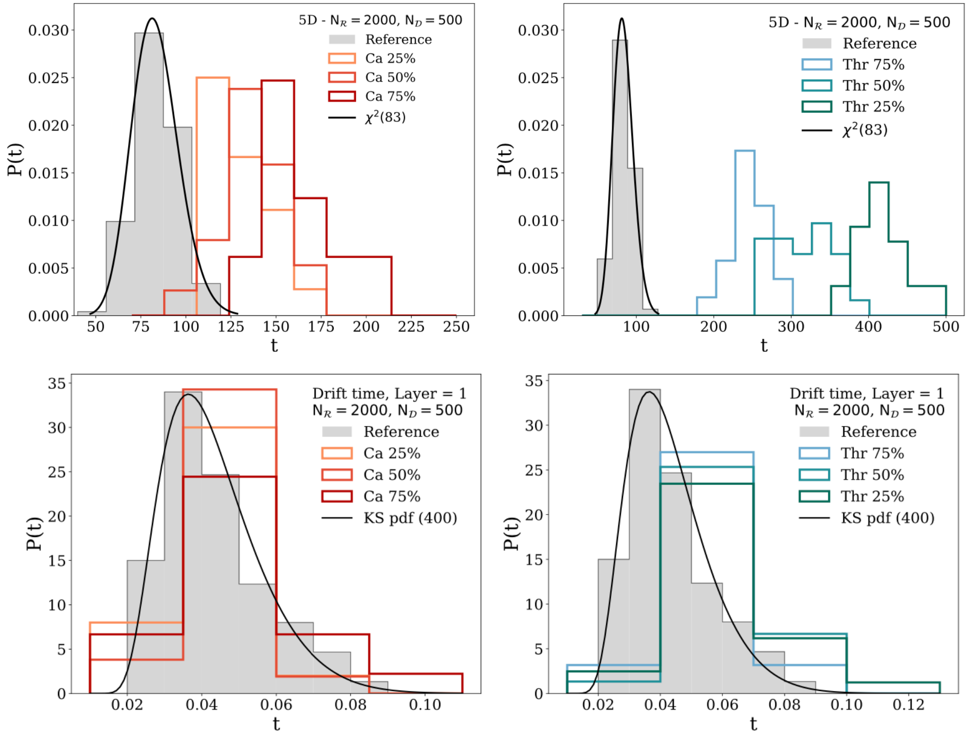
Data: https://zenodo.org/records/4442665

# 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 ($\phi$): 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)

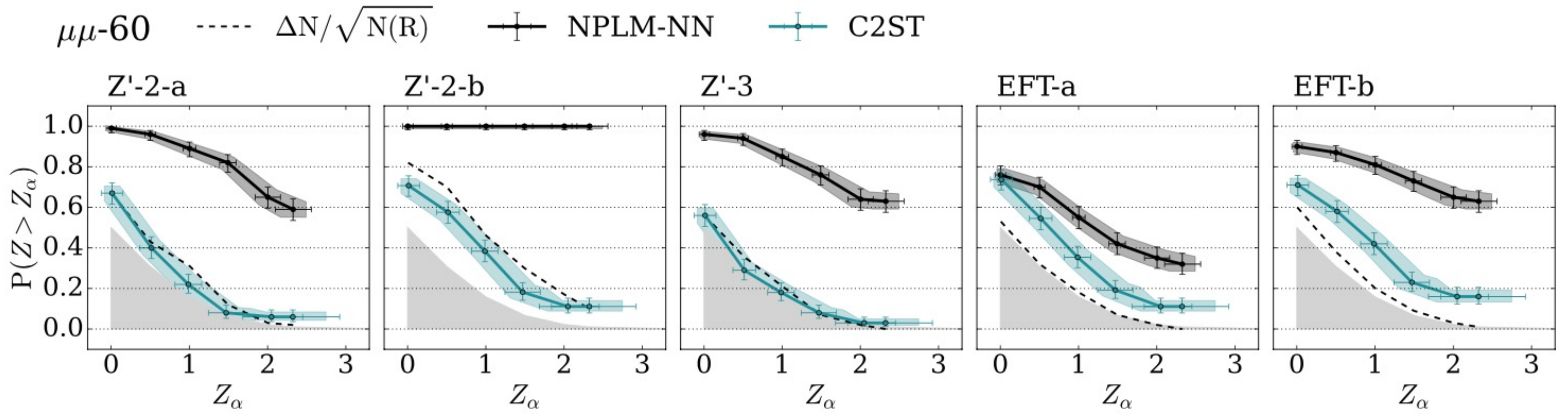
$\bar{t}_{tr} \approx 0.5$ sec

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

# Comparisons

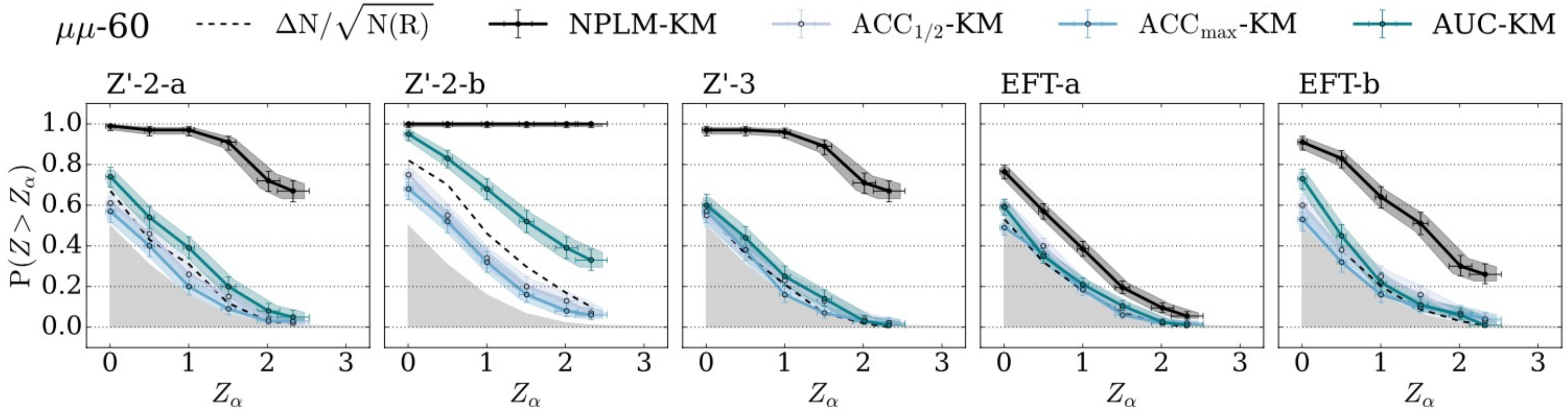
Grosso, ML, Pierini, Wulzer, SciPostPhys (2024)

CLASSIFIER 2ST

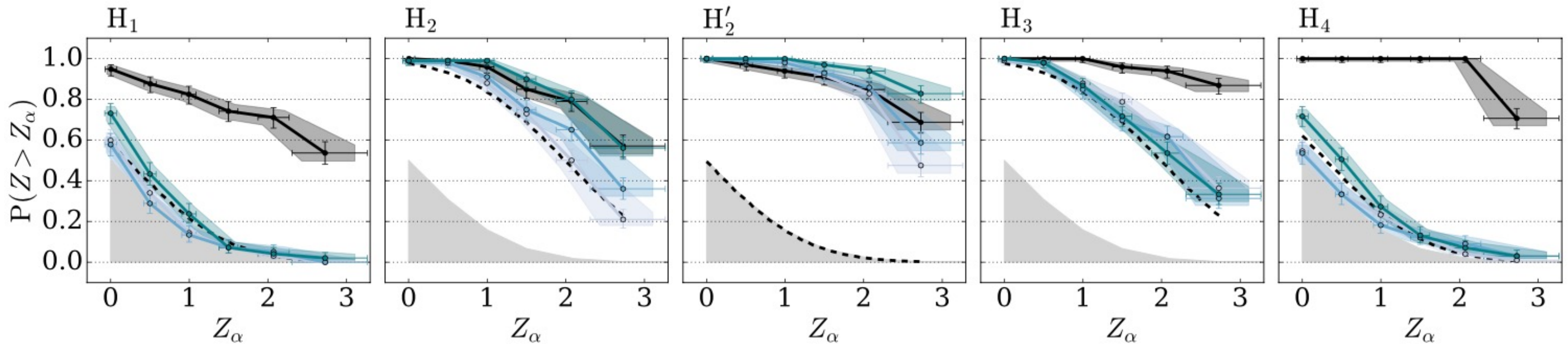



EVALUATION METRICS

Output




Input

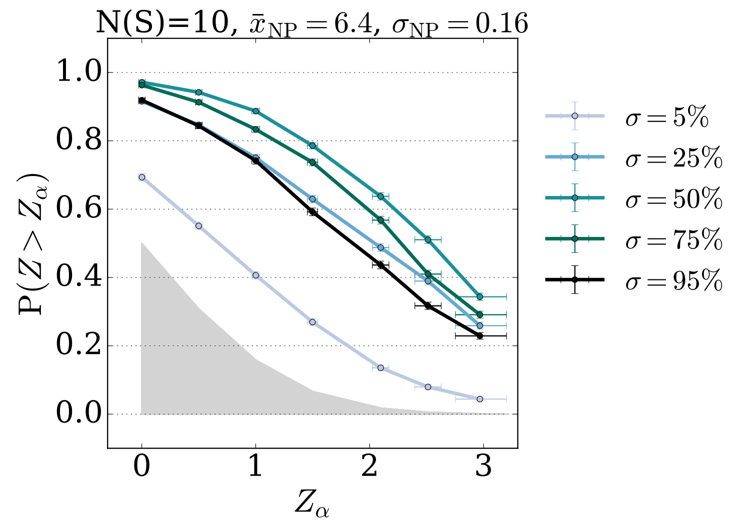



UniGe

# Multiple testing for increased sensitivity

Combine multiple tests with different hyperparameters on same data to mitigate inductive bias

G. Grosso and M. Letizia, *to appear*




| N(S) | 7 | 18 | 13 | 10 | 90 |
|---|---|---|---|---|---|
| $\bar{x}_{NP}$ | 4 | 4 | 4 | 6.4 | 1.6 |
| $\sigma_{NP}$ | 0.01 | 0.16 | 0.64 | 0.16 | 0.16 |
| $\sigma = 0.1$ | **0.015 ± 0.003** | 0.046 ± 0.005 | 0.004 ± 0.001 | 0.039 ± 0.004 | 0.37 ± 0.01 |
| $\sigma = 0.3$ | 0.013 ± 0.003 | **0.132 ± 0.008** | 0.008 ± 0.002 | 0.26 ± 0.01 | **0.67 ± 0.02** |
| $\sigma = 0.7$ | 0.005 ± 0.002 | 0.113 ± 0.007 | 0.008 ± 0.002 | **0.34 ± 0.01** | 0.66 ± 0.02 |
| $\sigma = 1.4$ | 0.004 ± 0.001 | 0.086 ± 0.007 | 0.013 ± 0.003 | 0.29 ± 0.01 | 0.54 ± 0.02 |
| $\sigma = 3.0$ | 0.007 ± 0.002 | 0.060 ± 0.005 | **0.019 ± 0.003** | 0.20 ± 0.01 | 0.38 ± 0.01 |
| min-$p$ | 0.013 ± 0.003 | 0.111 ± 0.007 | 0.013 ± 0.003 | 0.28 ± 0.01 | 0.64 ± 0.02 |
| prod-$p$ | **0.017 ± 0.003** | **0.137 ± 0.008** | **0.017 ± 0.003** | **0.32 ± 0.01** | **0.70 ± 0.02** |
| avg-$p$ | 0.005 ± 0.002 | 0.079 ± 0.006 | 0.009 ± 0.002 | 0.110 ± 0.07 | 0.55 ± 0.02 |
| smax-$t$ | 0.015 ± 0.003 | 0.046 ± 0.005 | 0.004 ± 0.001 | 0.040 ± 0.004 | 0.37 ± 0.01 |

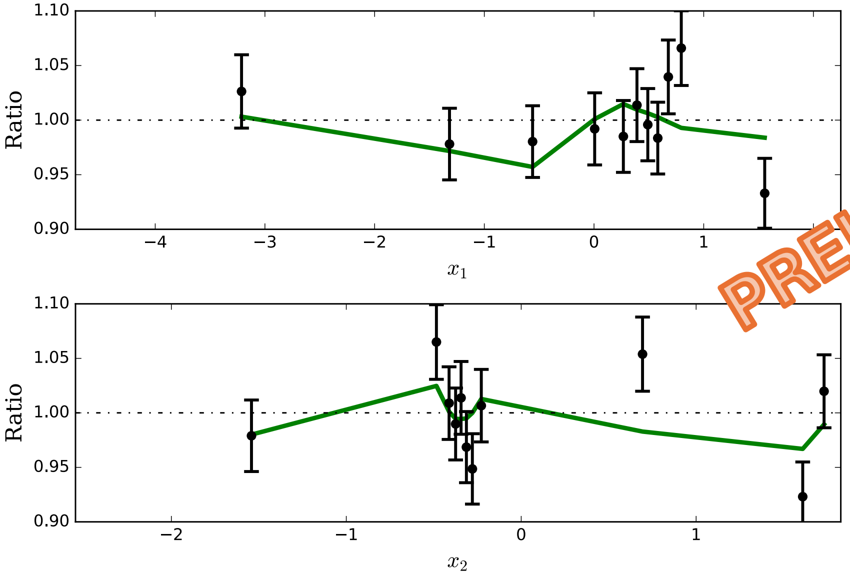
**Table 1: EXPO 1D –** probability of observing $Z \geq 3$.

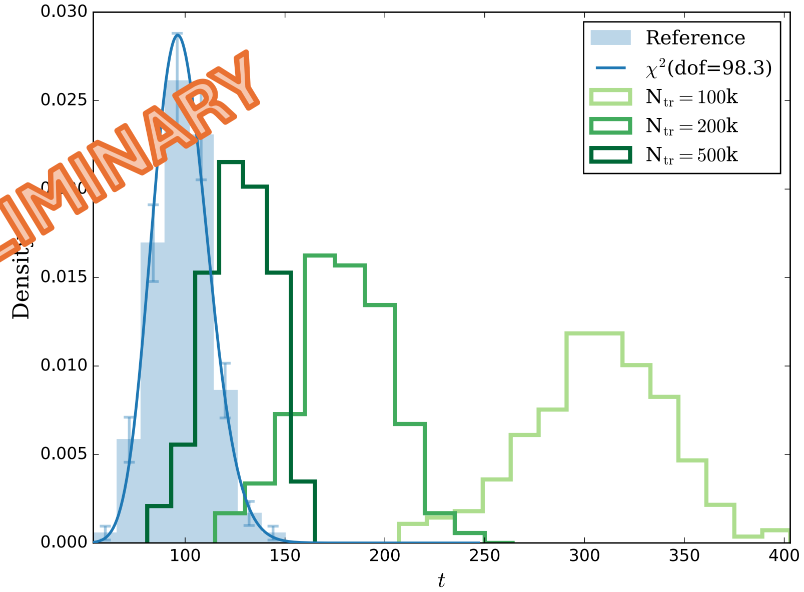
# Evaluation of generative models

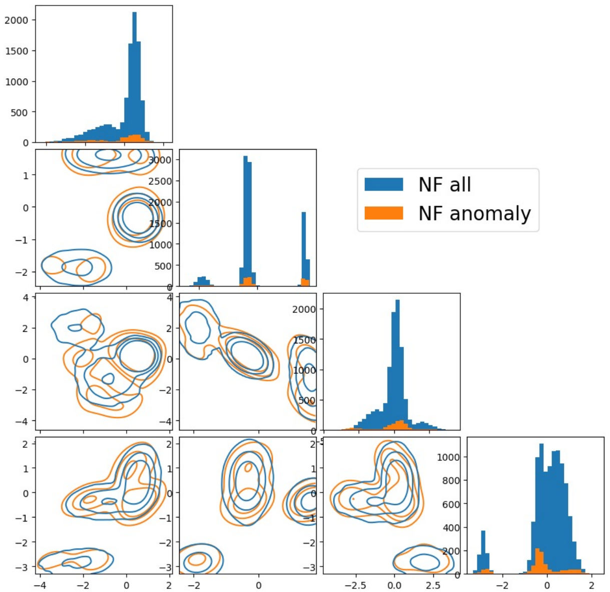
**(i**n preparation, stay tuned!)

## Normalising flows: RealNVP on correlated mixtures of Gaussians

| $N_{tr}$ \ $d$ | 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 | $1.93^{+1.02}_{-0.99}$ | $3.01^{+0.74}_{-1.13}$ | $3.16^{+1.10}_{-1.02}$ | $5.05^{+1.02}_{-0.99}$ | $2.07^{+0.81}_{-0.97}$ | $3.06^{+1.13}_{-0.86}$ |



# Wrap up

- Efficient testing with kernel methods
- Model (generator) vs observations

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