Machine Learning at GGI, September 7th, 2022

Learning New Physics from a Machine

model-independent analysis strategy for collider experiments

Gaia Grosso[1,2]

[1]University and INFN of Padova, [2]CERN

Based on:

Phys. Rev. D (d'Agnolo, Wulzer)

Eur. Phys. J. C (d'Agnolo, GG, Pierini, Wulzer, Zanetti)

Eur. Phys. J. C (d'Agnolo, GG, Pierini, Wulzer, Zanetti)

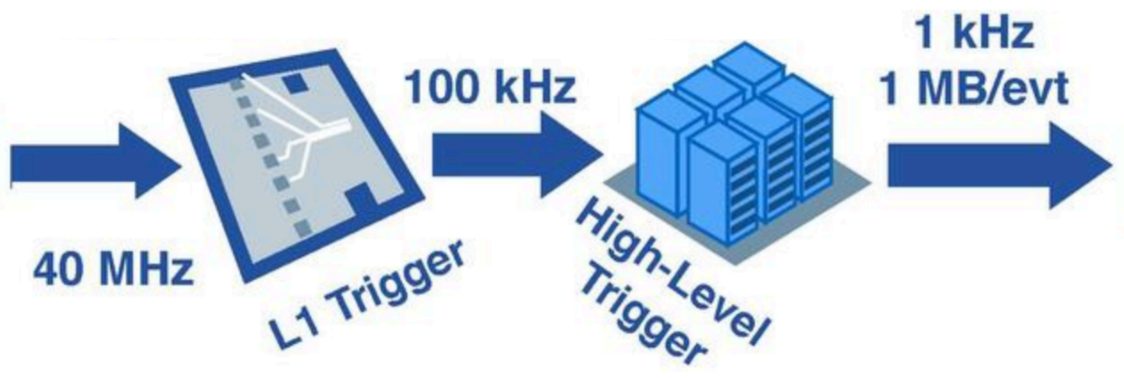
Dipartimento di Fisica e Astronomia Galileo Galilei

Search for New Physics at collider experiments

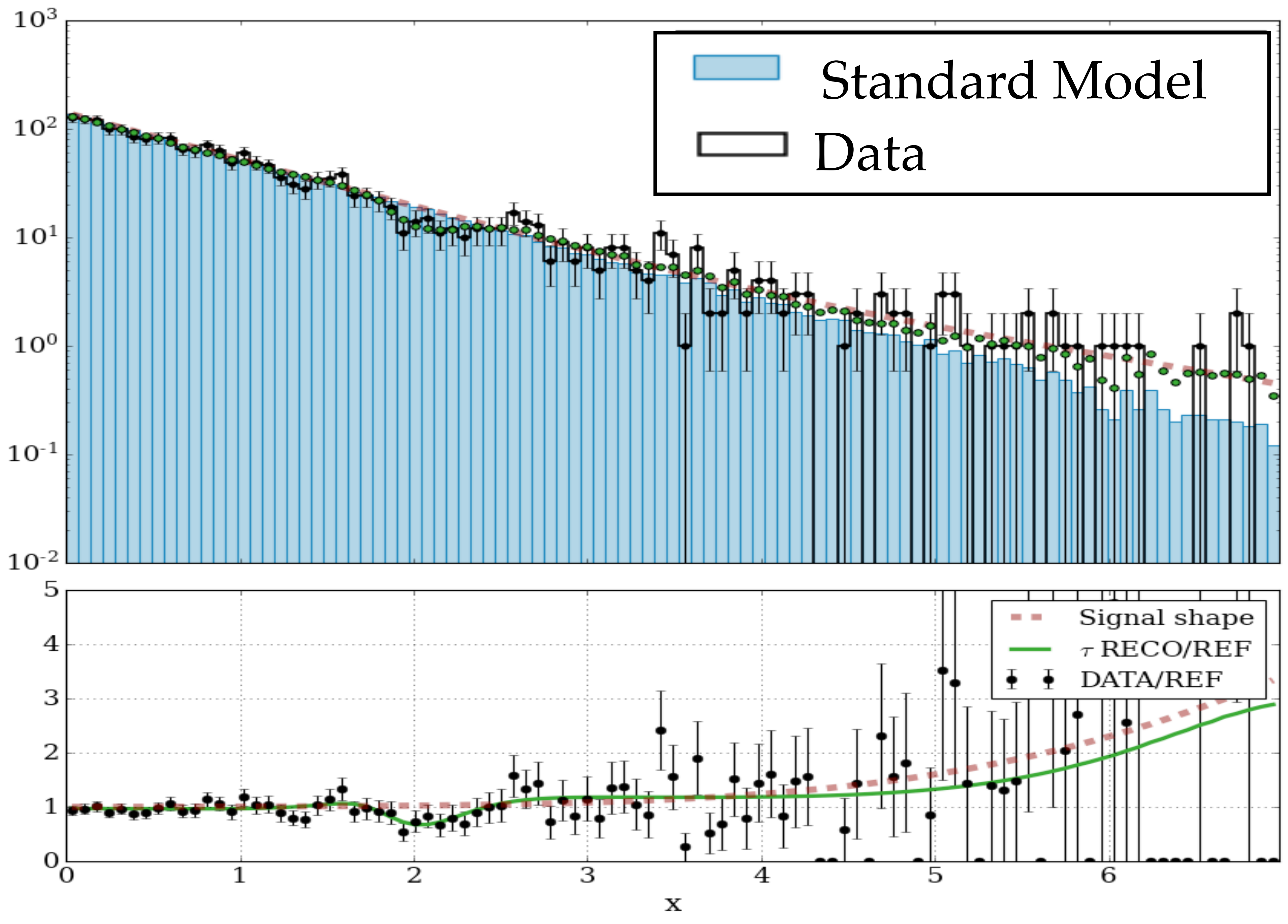
Standard Model (SM): current understanding of the world of particle physics and fundamental interactions

How well does the SM describe the data?

Experiment at LHC



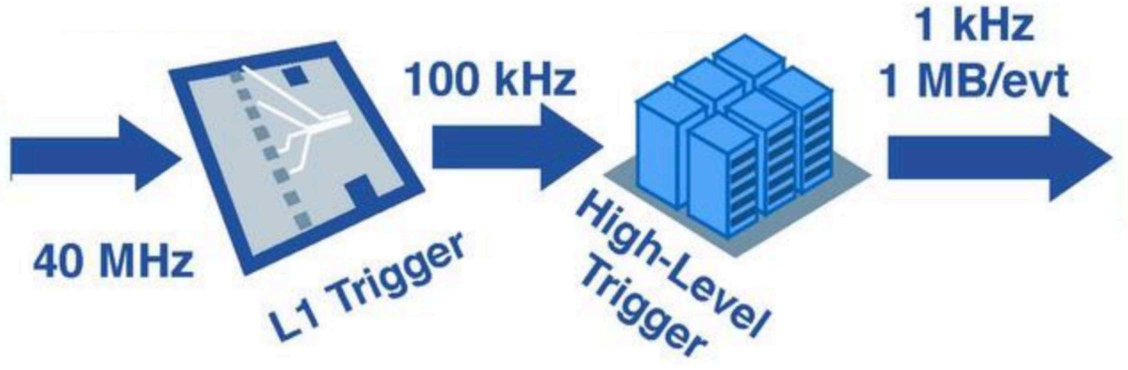
Trigger selection



Offline data analysis

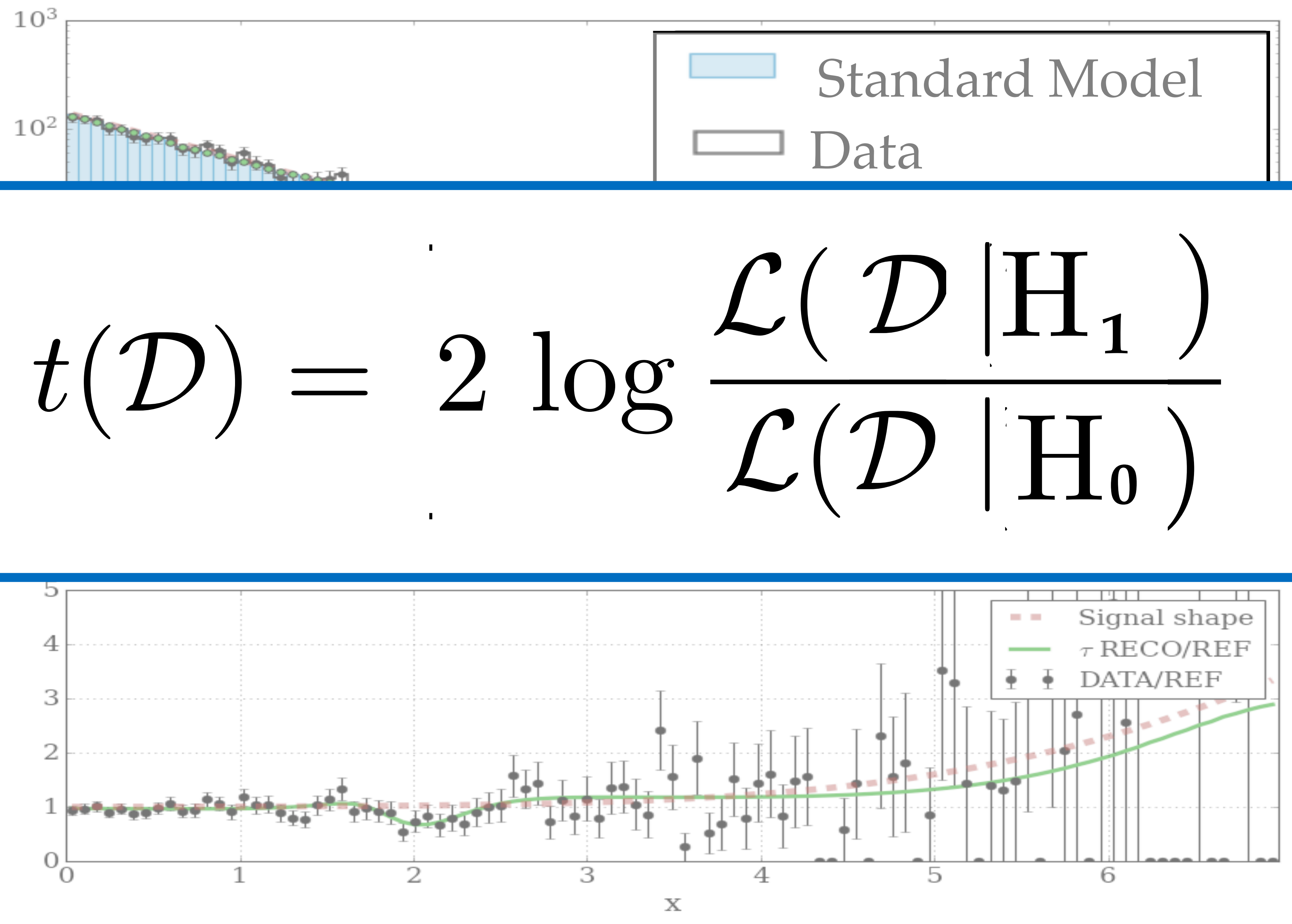
Search for New Physics at collider experiments

Experiment at LHC



Trigger selection

How well does the SM describe the data?

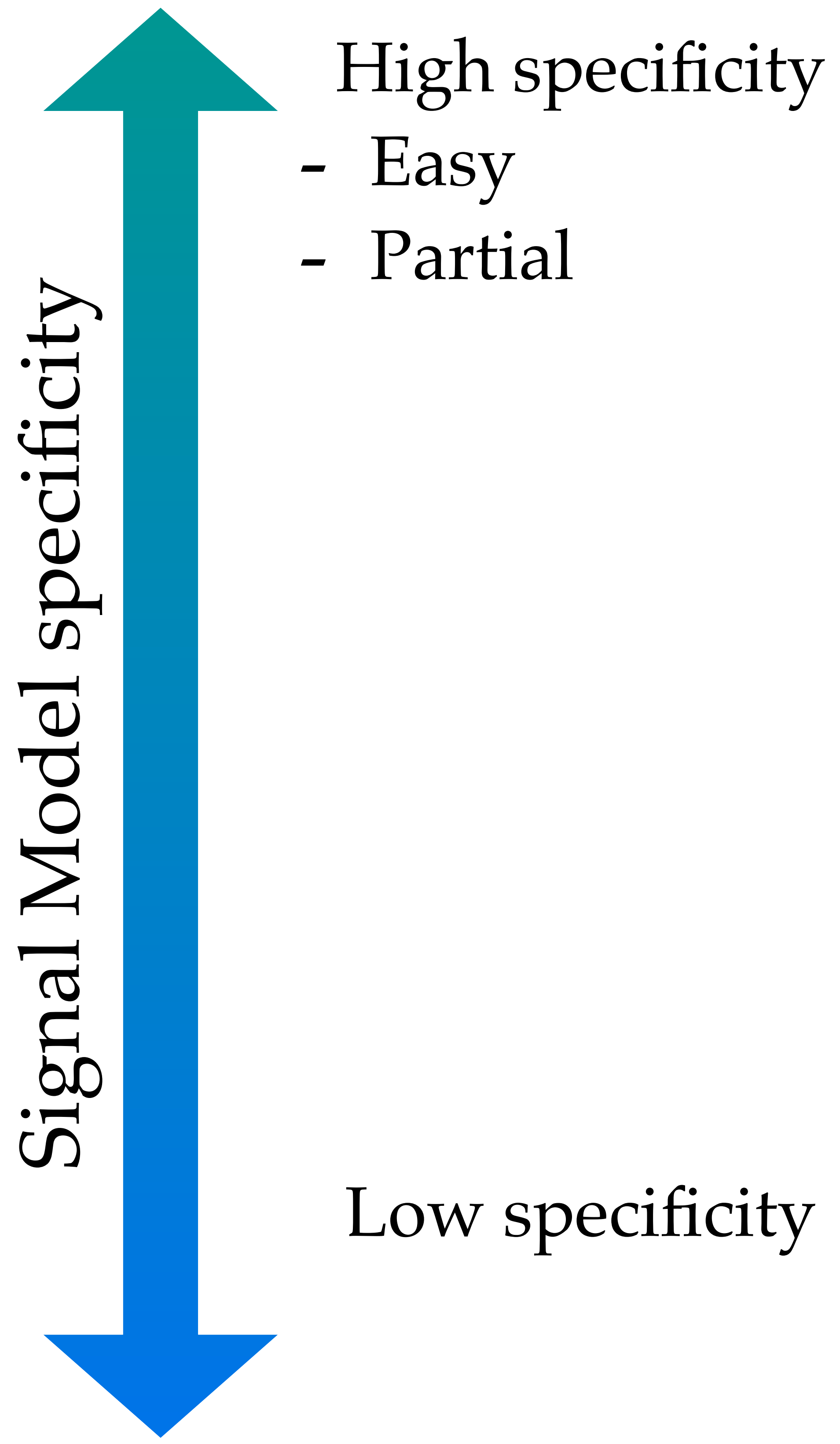


$$t(\mathcal{D}) = 2 \log \frac{\mathcal{L}(\mathcal{D} \mid H_1)}{\mathcal{L}(\mathcal{D} \mid H_0)}$$

H_0 : null hypothesis (SM)
H_1 : alternative hypothesis (NP)

Search for New Physics at collider experiments

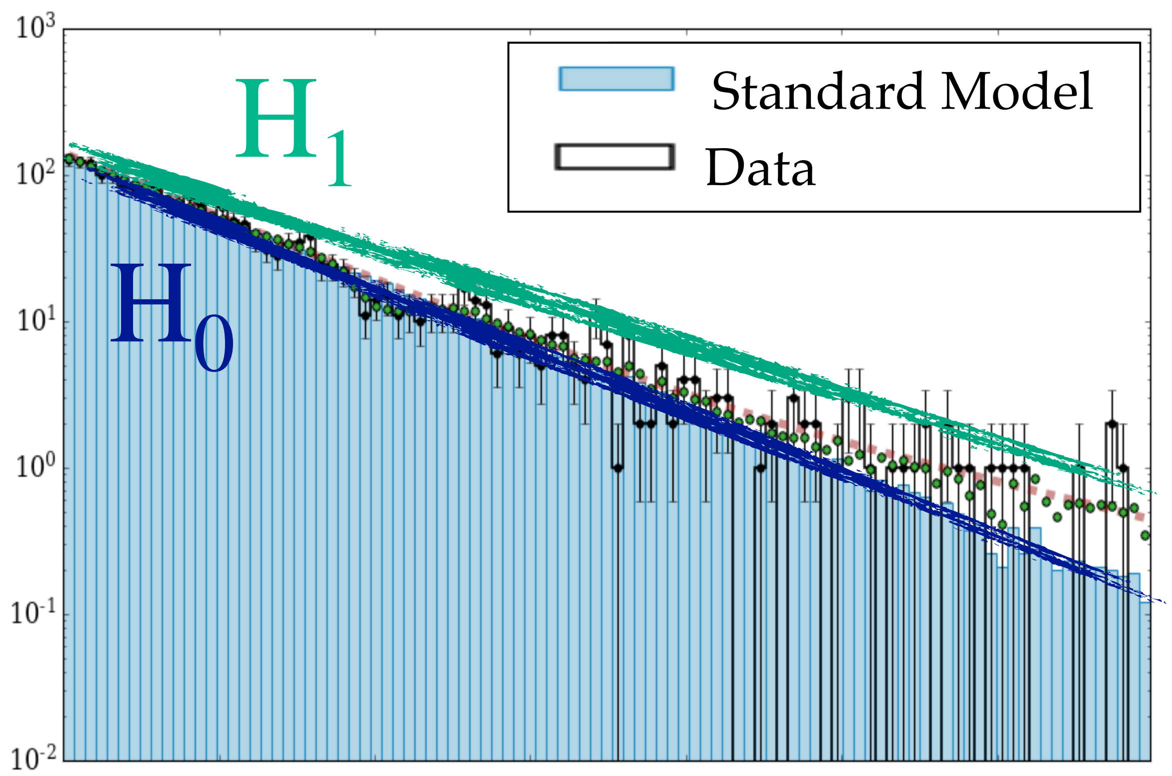
Analysis strategy



How well does the SM describe the data?

Simple hypothesis

$$t(\mathcal{D}) = 2 \log \frac{\mathcal{L}(\mathcal{D}\,|\,\mathrm{H}_1)}{\mathcal{L}(\mathcal{D}\,|\,\mathrm{H}_0)}$$



H_0 : null hypothesis (SM)
H_1 : alternative hypothesis (NP)

Search for New Physics at collider experiments

Analysis strategy

Signal Model specificity

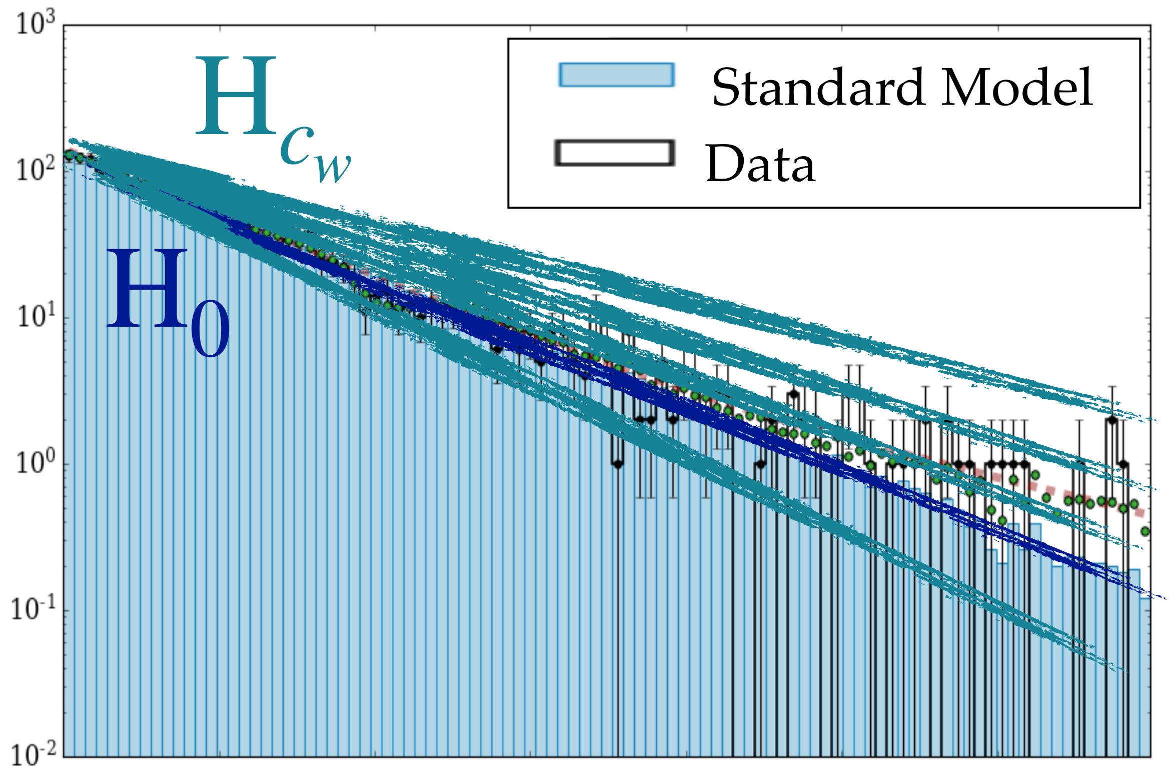
High specificity
- Easy
- Partial

Low specificity

How well does the SM describe the data?

EFT

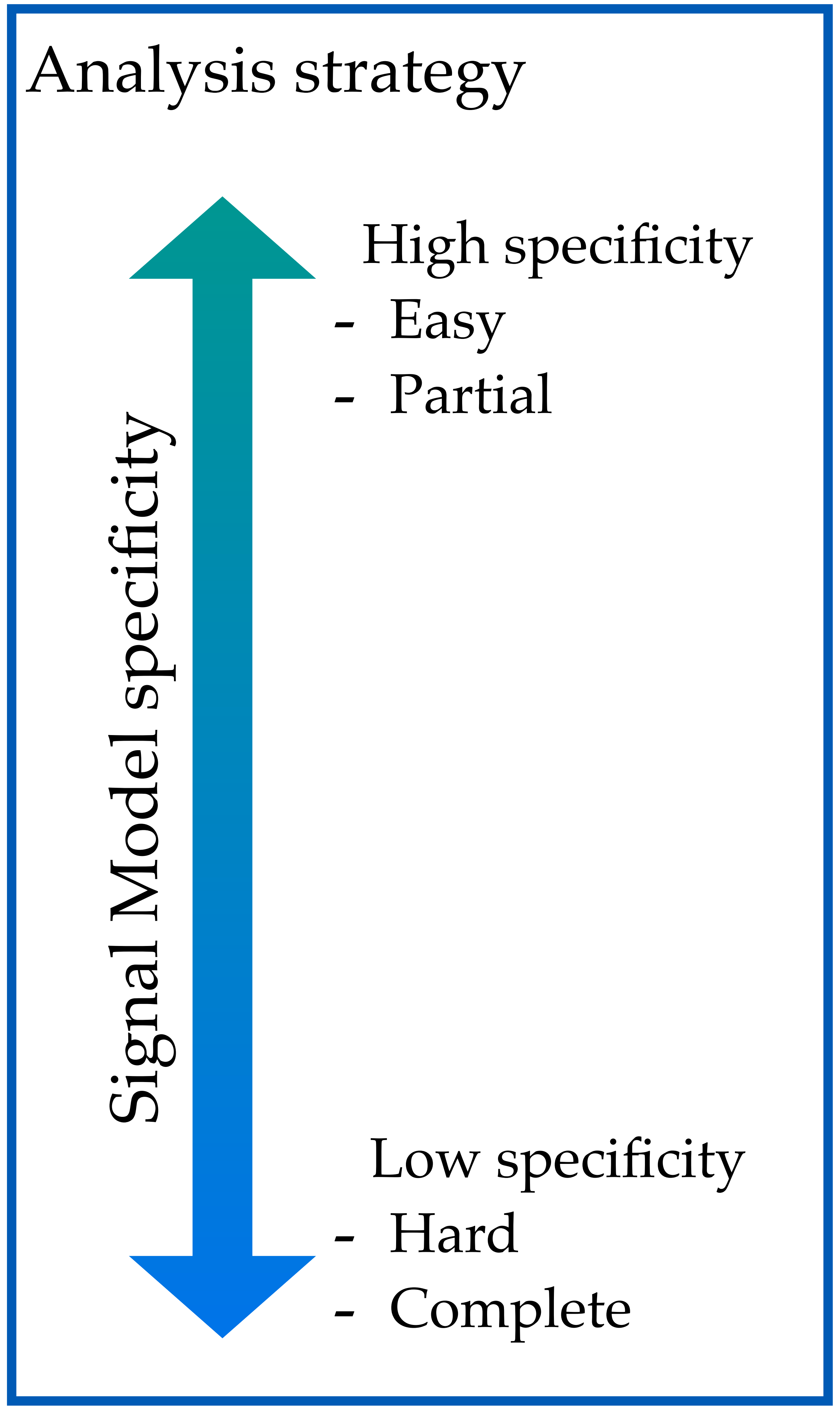
$$t(\mathcal{D}) = \max_{c_w} \left[2 \log \frac{\mathcal{L}(\mathcal{D} \,|\, \mathrm{H}_{c_w})}{\mathcal{L}(\mathcal{D} \,|\, \mathrm{H}_0)} \right]$$



See Jaco's, Robert's and Giuliano's talks tomorrow

H_0: null hypothesis (SM)
H_1: alternative hypothesis (NP)

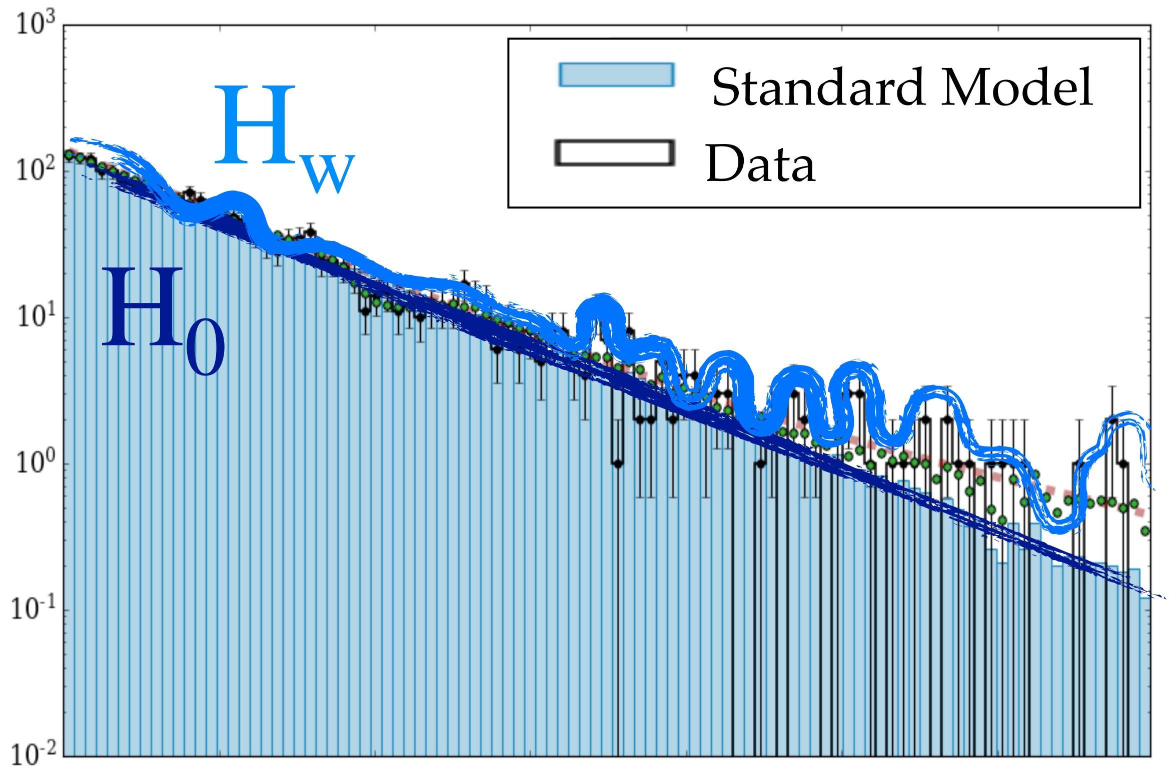
Search for New Physics at collider experiments



How well does the SM describe the data?

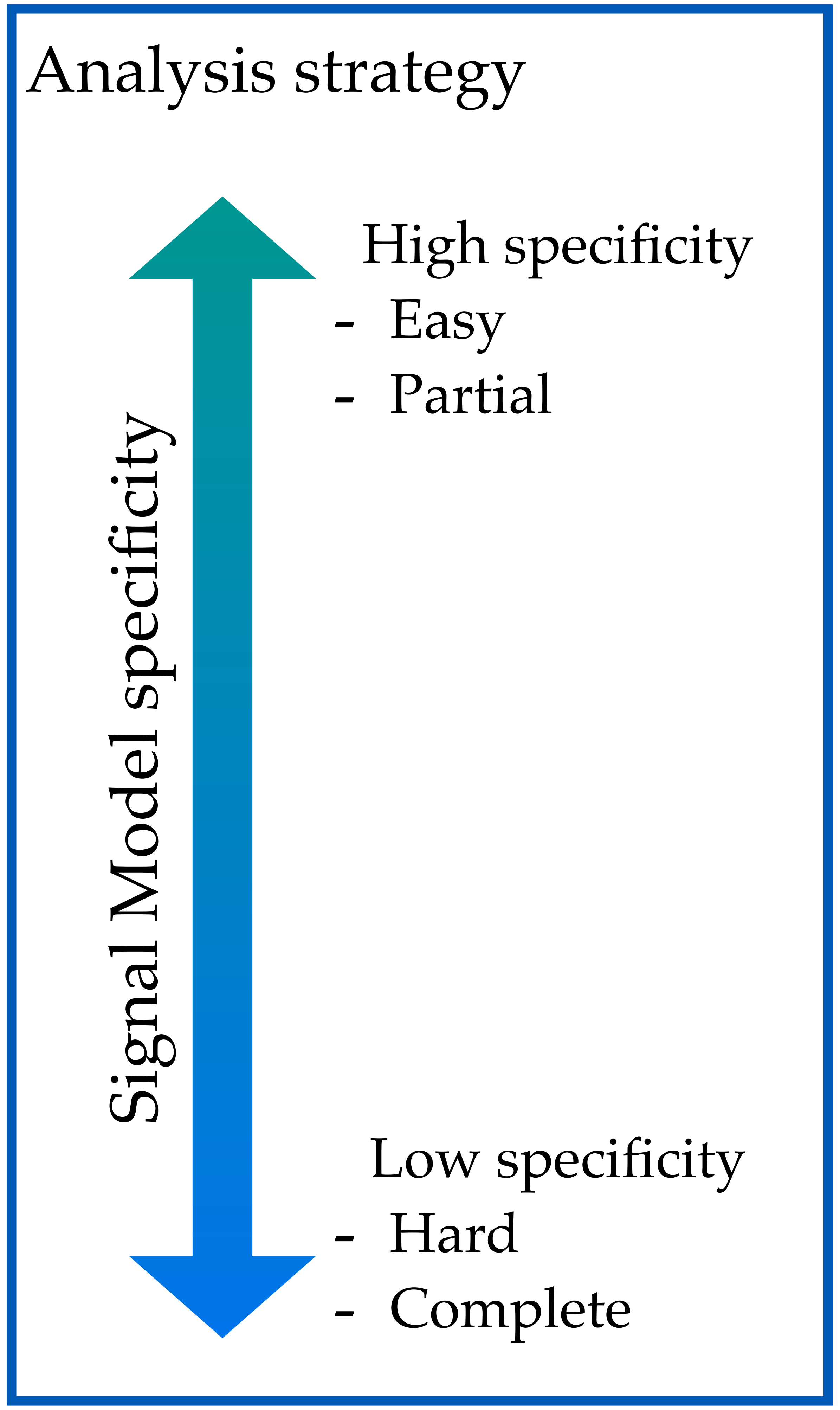
Neural networks

$$t(\mathcal{D}) = \max_{\mathbf{w}} \left[2 \log \frac{\mathcal{L}(\mathcal{D}\,|\,\mathrm{H}_{\mathbf{w}})}{\mathcal{L}(\mathcal{D}\,|\,\mathrm{H}_{\mathbf{0}})} \right]$$



$\mathrm{H}_{\mathbf{0}}$: null hypothesis (SM)
$\mathrm{H}_{\mathbf{1}}$: alternative hypothesis (NP)

Search for New Physics at collider experiments



How well does the SM describe the data?

Neural networks

$$t(\mathcal{D}) = \max_{\mathbf{w}} \left[2 \log \frac{\mathcal{L}(\mathcal{D} \,|\, \mathrm{H}_{\mathbf{w}})}{\mathcal{L}(\mathcal{D} \,|\, \mathrm{H}_{\mathbf{0}})} \right]$$

Our proposal

Learning New Physics from a Machine (NPLM)

$\mathrm{H}_{\mathbf{0}}$: null hypothesis (SM)
$\mathrm{H}_{\mathbf{1}}$: alternative hypothesis (NP)

In this talk

- **What is NPLM and how does it work?**

  Complete analysis strategy testing the data for departures from SM expectations (from data to a *p*-value for discovery, taking care of *systematic uncertainties* in the process).

  → main concept (neglecting systematic uncertainties) [1, 2]

  → including systematic uncertainties [3]

- **What is NPLM *good for*?**

  Multivariate, unbinned analysis, towards model independence (released constraints, lower level information, simultaneously sensitive to multiple signal patterns).

  → 5D analysis of a two-body final state at the LHC [3]

Link to related papers:

[1] "Learning New Physics from a Machine" Phys. Rev. D (d'Agnolo, Wulzer)
[2] "Learning Multivariate New Physics" Eur. Phys. J. C (d'Agnolo, GG, Pierini, Wulzer, Zanetti)
[3] "Learning New Physics from an Imperfect Machine" Eur. Phys. J. C (d'Agnolo, GG, Pierini, Wulzer, Zanetti)

NPLM algorithm

New Physics Learning Machine (NPLM)

Main Idea

- Goal: performing a **log-likelihood-ratio hypothesis test**
  End-to-end strategy, from the data to a p-value for the discovery (frequentistic approach)

$$t(\mathcal{D}) = \max_{\mathbf{w}} \left[2 \log \frac{\mathcal{L}(\mathcal{D} \,|\, \mathrm{H}_{\mathbf{w}})}{\mathcal{L}(\mathcal{D} \,|\, \mathrm{R}_0)} \right]$$

R_0 : reference hypothesis
$\mathrm{H}_{\mathbf{w}}$: alternative hypothesis

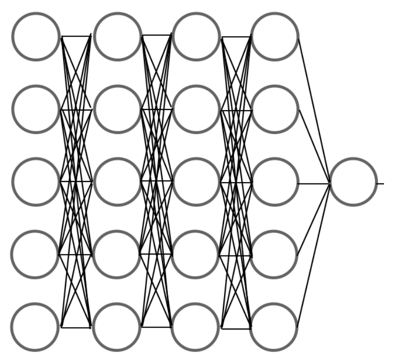
- Exploiting a Neural Network (NN) to **parametrize** the data distribution in terms of a Reference distribution (R_0)

$$n(x \,|\, \mathrm{T}) \approx n(x \,|\, \mathrm{H}_{\hat{\mathbf{w}}}) = n(x \,|\, \mathrm{R}_0)\, e^{f(x, \hat{\mathbf{w}})}$$

True (T) data distribution — Unknown

Data distribution learnt by the NN — Alternative hypothesis

Reference distribution — Null hypothesis (SM)



NN model

- **Signal-model-independent**: reduced assumptions on the signal hypothesis

"Learning New Physics from a Machine" Phys. Rev. D

New Physics Learning Machine (NPLM)

Main Idea

Maximum Likelihood from minimal loss:

Test statistic

$$\bar{t}(\mathcal{D}) = 2 \max_{\mathbf{w}} \log \left[\frac{\mathcal{L}(\mathcal{D} \mid \mathrm{H}_{\mathbf{w}})}{\mathcal{L}(\mathcal{D} \mid \mathrm{R}_{\mathbf{0}})} \right] = -2 \min_{\mathbf{w}} \{ \bar{L}[f(\cdot;\, \mathbf{w})] \}$$

Loss function

$$\bar{L}\,[f(x;\, \mathbf{w})] = -\sum_{x \in \mathcal{D}} [\, f(x;\, \mathbf{w})] + \sum_{x \in \mathcal{R}} w_x \left[e^{f(x;\, \mathbf{w})} - 1 \right]$$

$\mathbf{w}$: trainable parameters on the NN model

D: data sample

R: reference sample (built according to the R_0 hypothesis); could be weighted (w)

Assumptions:

- $N_R \gg N_D$ the statistical fluctuations of the reference sample are negligible.
- the weights of the reference sample (w) are such that the reference sample is normalised to match the data sample luminosity $\sum_{x \in R} w_x = N(R_0)$

"Learning New Physics from a Machine" Phys. Rev. D

New Physics Learning Machine (NPLM)

Main Idea

Maximum Likelihood from minimal loss:

Test statistic

$$\bar{t}(\mathcal{D}) = 2\max_{\mathbf{w}} \log\left[\frac{\mathcal{L}(\mathcal{D}\,|\,\mathrm{H}_{\mathbf{w}})}{\mathcal{L}(\mathcal{D}\,|\,\mathrm{R}_{\mathbf{0}})}\right] = -2\min_{\mathbf{w}}\{\bar{L}[f(\cdot;\,\mathbf{w})]\}$$

w: trainable parameters on the NN model

D: data sample

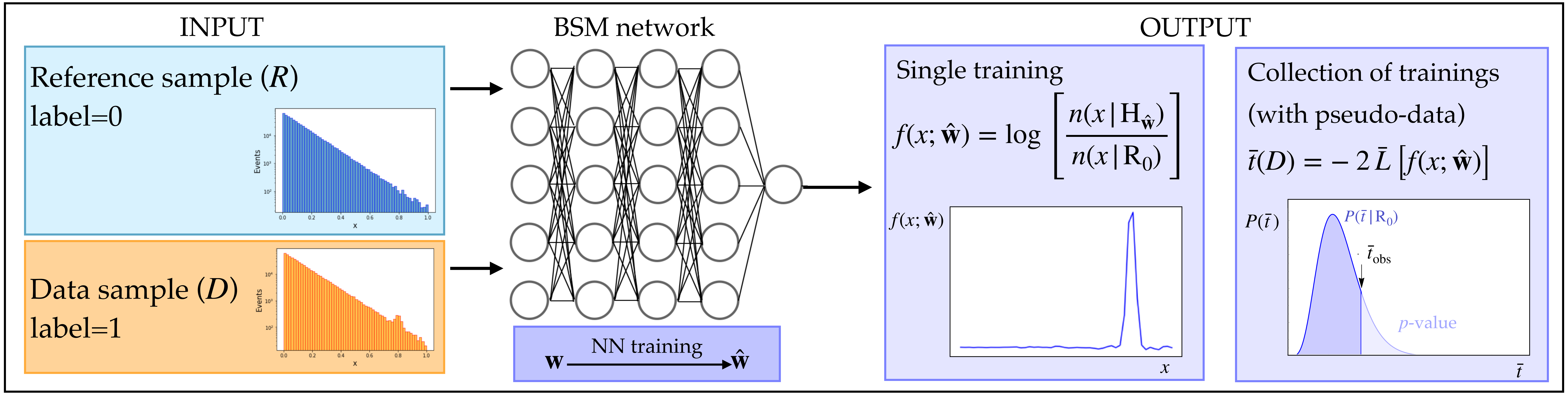
R: reference sample (built according to the R_0 hypothesis); could be weighted (*w*)

Loss function

$$\bar{L}\,[f(x;\,\mathbf{w})] = -\sum_{x\in\mathcal{D}}[\,f(x;\,\mathbf{w})] + \sum_{x\in\mathcal{R}} w_x\left[e^{f(x;\,\mathbf{w})} - 1\right]$$

Assumptions:

- $N_R \gg N_D$ the statistical fluctuations of the reference sample are negligible.
- the weights of the reference sample (*w*) are such that the reference sample is normalised to match the data sample luminosity $\sum_{x\in R} w_x = N(R_0)$



"Learning New Physics from a Machine" Phys. Rev. D

New Physics Learning Machine (NPLM)

Dealing with multivariate problems (negligible uncertainties)

Asymptotic formula for the $\bar{t}$ distribution under R_0:

Wilks-Wald theorem:

Θ_0: set of parameters describing H_0

Θ_1: set of parameters describing H_1

If $\mathrm{H}_0 \subseteq \mathrm{H}_1$, then under the H_0 hypothesis the test statistic

$$t(\mathcal{D}) = 2 \log \frac{\mathcal{L}(\mathcal{D} \,|\, \mathrm{H}_{\mathbf{w}})}{\mathcal{L}(\mathcal{D} \,|\, \mathrm{R}_{\mathbf{0}})}$$

asymptotically follows a χ^2_{df} distribution with $df = |\Theta_1| - |\Theta_0|$

→ If the Wilks' theorem hold, the target distribution for $\bar{t}$ under the R_0 hypothesis is a χ^2_{df} with $df = |\mathbf{w}|$.

Due to the **finite size** of the training samples, the **sparsity** of the data (especially in multivariate problems) and the **approximation** errors, the distribution of $\bar{t}(D)$ under R_0 does not follow the target $\chi^2_{|\mathbf{w}|}$ by default.

→ a **(NN) MODEL REGULARIZATION** procedure can solve this problem!

"Learning Multivariate New Physics" Eur. Phys. J. C

New Physics Learning Machine (NPLM)

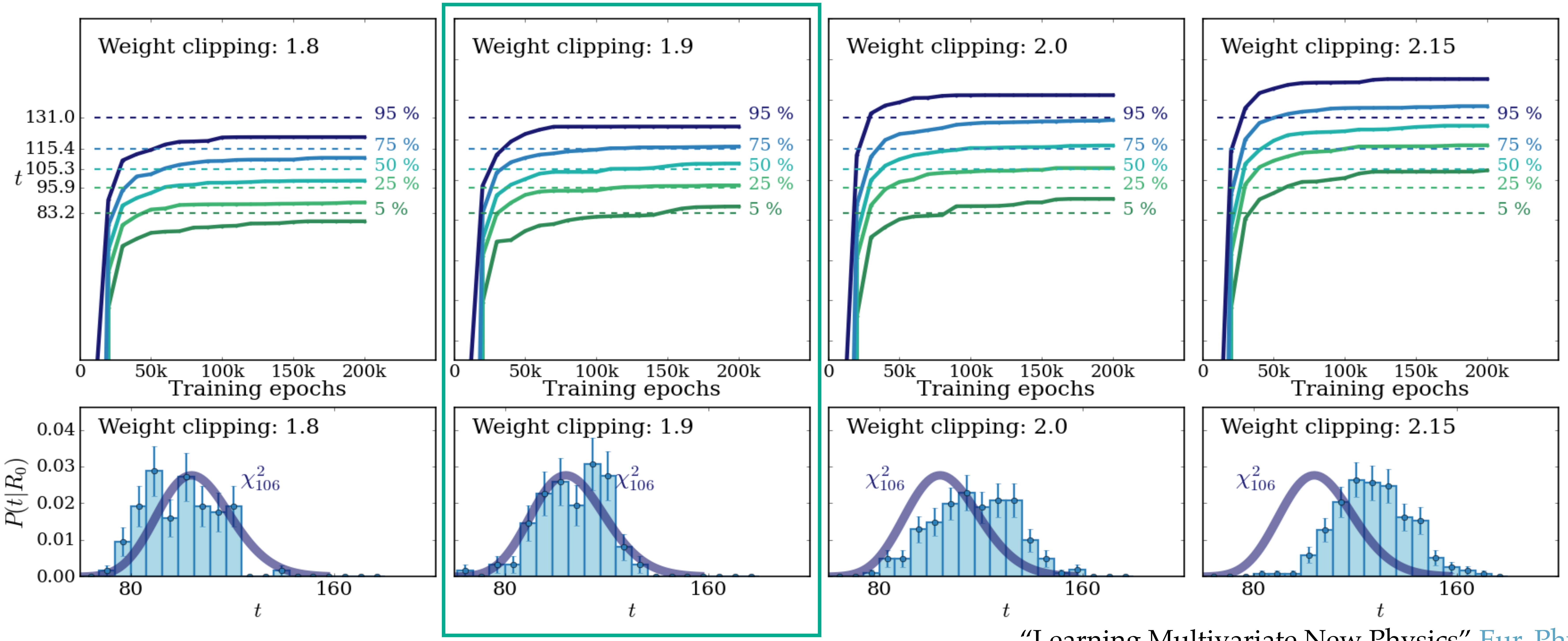
NN Model regularization:

Weight clipping parameter:
Upper boundary to the magnitude that each trainable parameter can assume during the training.

For a chosen NN architecture, tuning the weight clipping allows to recover a good agreement of the empirical distribution of $\bar{t}$ under R_0 with the **target** $\chi^2_{|\mathbf{w}|}$ distribution.

Example:
NN model: 5-7-7-1,
$|\mathbf{w}| = 106$

Legend:
- Percentiles of the empirical $\bar{t}$ distribution under R_0
- Percentiles of the target $\chi^2_{|\mathbf{w}|}$
- Empirical $\bar{t}$ distribution under R_0
- Target $\chi^2_{|\mathbf{w}|}$



"Learning Multivariate New Physics" Eur. Phys. J. C

New Physics Learning Machine (NPLM)

Including systematic uncertainties

Test statistic

$$t(\mathcal{D},\,\mathcal{A}) = 2\log\left[\frac{\max_{\mathbf{w},\,\boldsymbol{\nu}} \mathcal{L}(\mathrm{H}_{\mathbf{w},\,\boldsymbol{\nu}}|\mathcal{D},\,\mathcal{A})}{\max_{\boldsymbol{\nu}} \mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D},\,\mathcal{A})}\right] = 2\log\left[\frac{\max_{\mathbf{w},\,\boldsymbol{\nu}} \mathcal{L}(\mathrm{H}_{\mathbf{w},\,\boldsymbol{\nu}}|\mathcal{D})\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\max_{\boldsymbol{\nu}} \mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D})\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}\right]$$

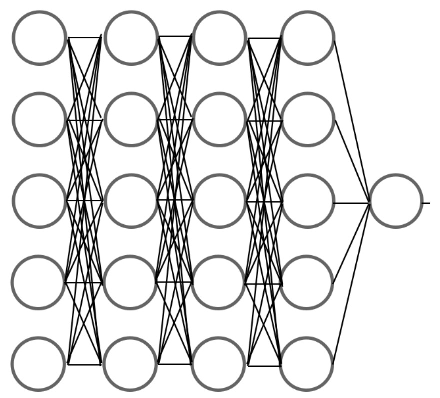
Doesn't depend on ν

- $\mathbf{w}$: trainable parameters on the NN model
- $\boldsymbol{\nu}$: set of nuisance parameters modelling the uncertainties effects
- $\mathcal{D}$: data sample
- $\mathcal{A}$: auxiliary sample (used to constrain $\boldsymbol{\nu}$)

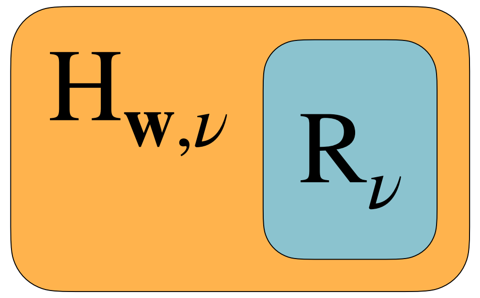
New parametrization

$$n(x\,|\,\mathrm{T}) \approx n(x\,|\,\mathrm{H}_{\hat{\mathbf{w}},\,\hat{\nu}}) = n(x\,|\,\mathrm{R}_0)\,\frac{n(x\,|\,\mathrm{R}_{\hat{\nu}})}{n(x\,|\,\mathrm{R}_0)}\,e^{f(x;\,\hat{\mathbf{w}})}$$

$n(x\,\vert\,\mathrm{T})$	$n(x\,\vert\,\mathrm{H}_{\hat{\mathbf{w}},\hat{\nu}})$	$n(x\,\vert\,\mathrm{R}_0)$	$\frac{n(x\vert\mathrm{R}_{\hat{\nu}})}{n(x\vert\mathrm{R}_0)}$	$f(x;\hat{\mathbf{w}})$
True (T) data distribution	Data distribution learnt by the NN	Reference distribution	New term containing the dependence on ν	NN model
Unknown	Alternative hypothesis	Null hypothesis	$r(x;\nu)$	



Note:
This parametrization choice guarantees $\mathrm{R}_\nu \subseteq \mathrm{H}_{\mathbf{w},\nu}$
($\mathrm{R}_\nu = \mathrm{H}_{\mathbf{w},\nu}$ for $f(\,\cdot\,;\,\mathbf{w}) \equiv 0$)



"Learning New Physics from an Imperfect Machine" Eur. Phys. J. C

New Physics Learning Machine (NPLM)

Including systematic uncertainties

Maximum Likelihood from minimal loss:

Test statistic

$$t(\mathcal{D},\,\mathcal{A}) = 2\log\left[\frac{\max_{\mathbf{w},\,\boldsymbol{\nu}} \mathcal{L}(\mathrm{H}_{\mathbf{w},\,\boldsymbol{\nu}}|\mathcal{D},\,\mathcal{A})}{\max_{\boldsymbol{\nu}} \mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D},\,\mathcal{A})}\right] = 2\log\left[\frac{\max_{\mathbf{w},\,\nu} \mathcal{L}(\mathrm{H}_{\mathbf{w},\,\boldsymbol{\nu}}|\mathcal{D})\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\max_{\boldsymbol{\nu}} \mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D})\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}\right]$$

$$= \tau(\mathcal{D},\,\mathcal{A}) - \Delta(\mathcal{D},\,\mathcal{A})$$

$\mathbf{w}$: trainable parameters on the NN model

$\boldsymbol{\nu}$: set of nuisance parameters modelling the uncertainties effects

$\mathcal{D}$: data sample

$\mathcal{A}$: auxiliary sample (used to constrain $\boldsymbol{\nu}$)

Tau term:

$$\tau(\mathcal{D},\,\mathcal{A}) = 2\max_{\mathbf{w},\,\boldsymbol{\nu}}\log\left[\frac{\mathcal{L}(\mathrm{H}_{\mathbf{w}},\,\boldsymbol{\nu}|\mathcal{D})\,\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})\,\mathcal{L}(\mathbf{0}|\mathcal{A})}\right] = -2\min_{\mathbf{w},\boldsymbol{\nu}} L\left[f(x,\mathbf{w}),\boldsymbol{\nu};\widehat{\delta}(x)\right]$$

Contains the dependence on a NN model

Delta term:

$$\Delta(\mathcal{D},\,\mathcal{A}) = 2\max_{\boldsymbol{\nu}}\log\left[\frac{\mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D})\,\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})\,\mathcal{L}(\mathbf{0}|\mathcal{A})}\right] = -2\min_{\boldsymbol{\nu}} L\left[\boldsymbol{\nu};\widehat{\delta}(x)\right]$$

Built on the knowledge of the Reference model (purely SM term)

$$r(x;\nu) = \frac{n(x\,|\,\mathrm{R}_\nu)}{n(x\,|\,\mathrm{R}_0)}$$

Taylor's expansion learning:

$$\widehat{r}(x;\nu) = \exp\left[\widehat{\delta}_1(x)\,\nu + \widehat{\delta}_2(x)\,\nu^2 + \ldots\right]$$

NN 1 NN2 …

"Learning New Physics from an Imperfect Machine" Eur. Phys. J. C

New Physics Learning Machine (NPLM)

Including systematic uncertainties

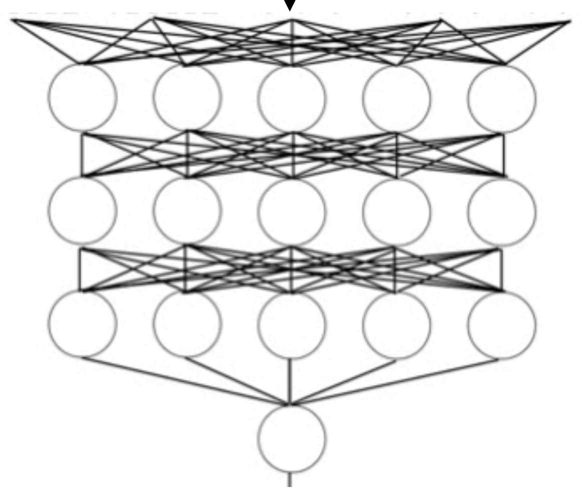
$$t(\mathcal{D}, \mathcal{A}) = \tau(\mathcal{D}, \mathcal{A}) - \Delta(\mathcal{D}, \mathcal{A})$$

τ term

$$\tau(\mathcal{D}, \mathcal{A}) = 2 \max_{\mathbf{w},\boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathrm{H}_{\mathbf{w},\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})} \right]$$

$$\tau(\mathcal{D}, \mathcal{A}) = -2 \min_{\mathbf{w},\boldsymbol{\nu}} \left\{ L_\tau \left[f(\cdot;\mathbf{w}),\ \boldsymbol{\nu};\ \widehat{\delta}(\cdot) \right] \right\}$$

$f(\cdot;\mathbf{w})$ $\quad$ $\boldsymbol{\nu};\ \widehat{\delta}(\cdot)$



$+$

$$\frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$$

Δ term

$$\Delta(\mathcal{D}, \mathcal{A}) = 2 \max_{\boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})} \right]$$

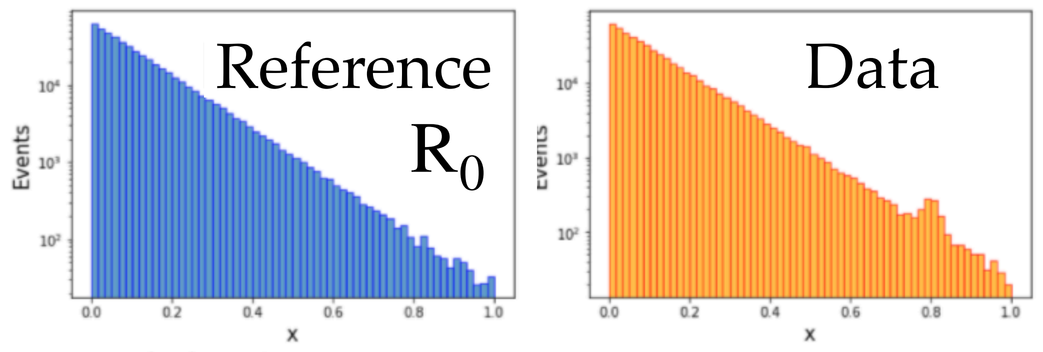
$$\Delta(\mathcal{D}, \mathcal{A}) = -2 \min_{\boldsymbol{\nu}} \left\{ L_\Delta \left[\boldsymbol{\nu};\ \widehat{\delta}(\cdot) \right] \right\}$$

$\boldsymbol{\nu};\ \widehat{\delta}(\cdot)$

$$\frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$$

Maximum likelihood from minimal loss

Training samples



New Physics Learning Machine (NPLM)

Including systematic uncertainties

$$t(\mathcal{D},\mathcal{A}) = \tau(\mathcal{D},\mathcal{A}) - \Delta(\mathcal{D},\mathcal{A})$$

τ term

$$\tau(\mathcal{D},\mathcal{A}) = 2\,\max_{\mathbf{w},\boldsymbol{\nu}} \log\left[\frac{\mathcal{L}(\mathrm{H}_{\mathbf{w},\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})}\cdot\frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})}\right]$$

Maximum likelihood from minimal loss

$$\tau(\mathcal{D},\mathcal{A}) = -2\,\min_{\mathbf{w},\boldsymbol{\nu}}\left\{L_\tau\Big[f(\cdot;\mathbf{w}),\ \boldsymbol{\nu};\ \widehat{\delta}(\cdot)\Big]\right\}$$

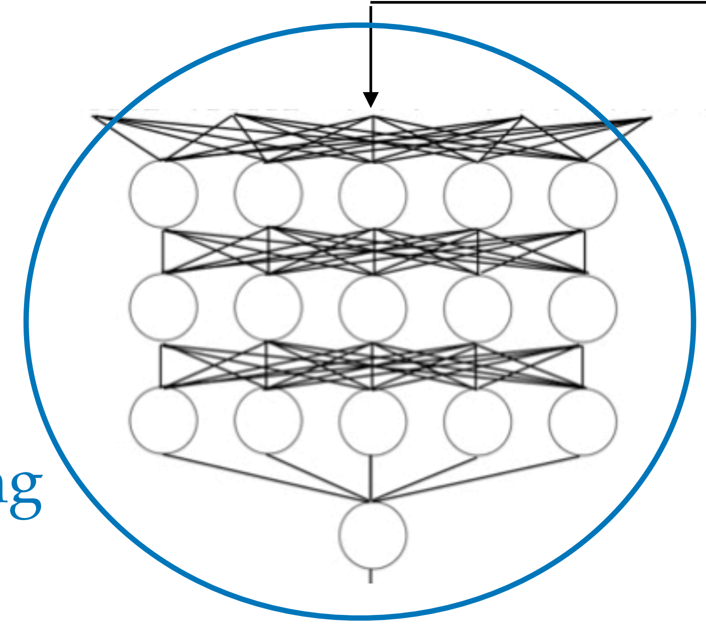
$f(\cdot;\mathbf{w})$ $\quad$ $\boldsymbol{\nu};\ \widehat{\delta}(\cdot)$

NN model

Set up

↓

Weight clipping tuning



$+$ $\quad \dfrac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$

Δ term

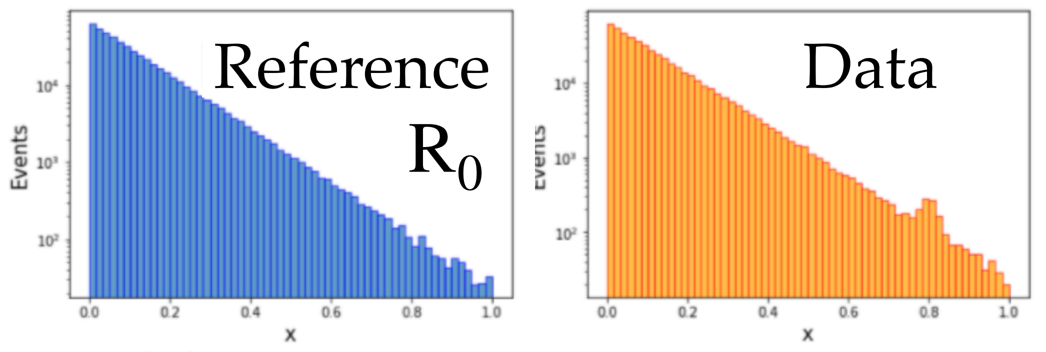
$$\Delta(\mathcal{D},\mathcal{A}) = 2\,\max_{\boldsymbol{\nu}} \log\left[\frac{\mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})}\cdot\frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})}\right]$$

$$\Delta(\mathcal{D},\mathcal{A}) = -2\,\min_{\boldsymbol{\nu}}\left\{L_\Delta\Big[\boldsymbol{\nu};\ \widehat{\delta}(\cdot)\Big]\right\}$$

$\boldsymbol{\nu};\ \widehat{\delta}(\cdot)$

$$\frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$$

Training samples



New Physics Learning Machine (NPLM)

Including systematic uncertainties

$$t(\mathcal{D},\mathcal{A}) = \tau(\mathcal{D},\mathcal{A}) - \Delta(\mathcal{D},\mathcal{A})$$

τ term

$$\tau(\mathcal{D},\mathcal{A}) = 2\,\max_{\mathbf{w},\boldsymbol{\nu}} \log\left[\frac{\mathcal{L}(\mathrm{H}_{\mathbf{w},\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})}\cdot\frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})}\right]$$

Δ term

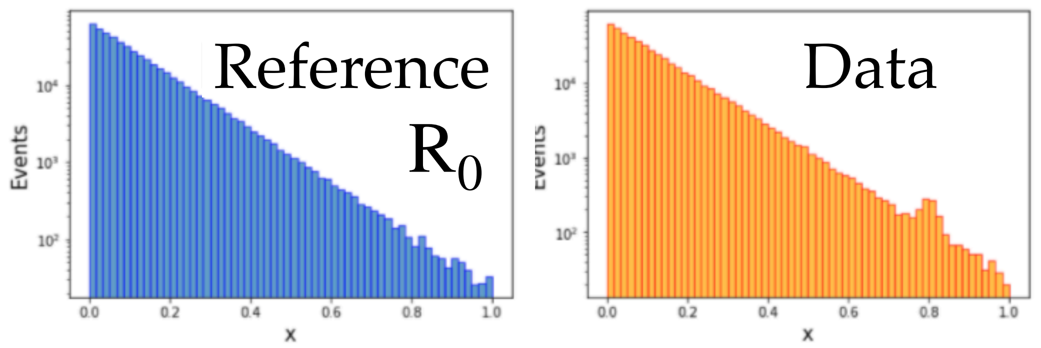
$$\Delta(\mathcal{D},\mathcal{A}) = 2\,\max_{\boldsymbol{\nu}} \log\left[\frac{\mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})}\cdot\frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})}\right]$$

Maximum likelihood from minimal loss

$$\tau(\mathcal{D},\mathcal{A}) = -2\,\min_{\mathbf{w},\boldsymbol{\nu}}\left\{L_\tau\Big[f(\cdot;\mathbf{w}),\,\boldsymbol{\nu};\,\widehat{\delta}(\cdot)\Big]\right\}$$

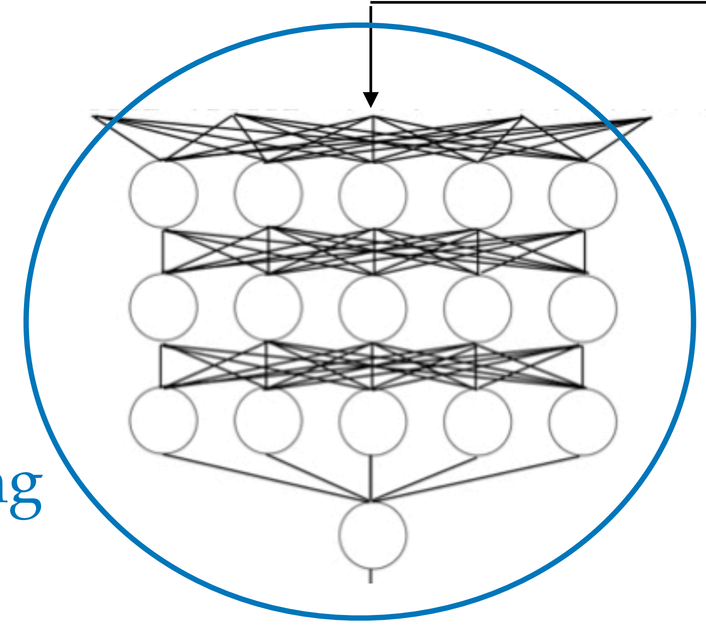
$$\Delta(\mathcal{D},\mathcal{A}) = -2\,\min_{\boldsymbol{\nu}}\left\{L_\Delta\Big[\boldsymbol{\nu};\,\widehat{\delta}(\cdot)\Big]\right\}$$

Training samples



$f(\cdot;\mathbf{w})$

NN model
Set up
↓
Weight clipping
tuning



+

$\boldsymbol{\nu};\,\widehat{\delta}(\cdot)$

$$\frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$$

$\boldsymbol{\nu};\,\widehat{\delta}(\cdot)$

$$\frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$$

New!

r modelling

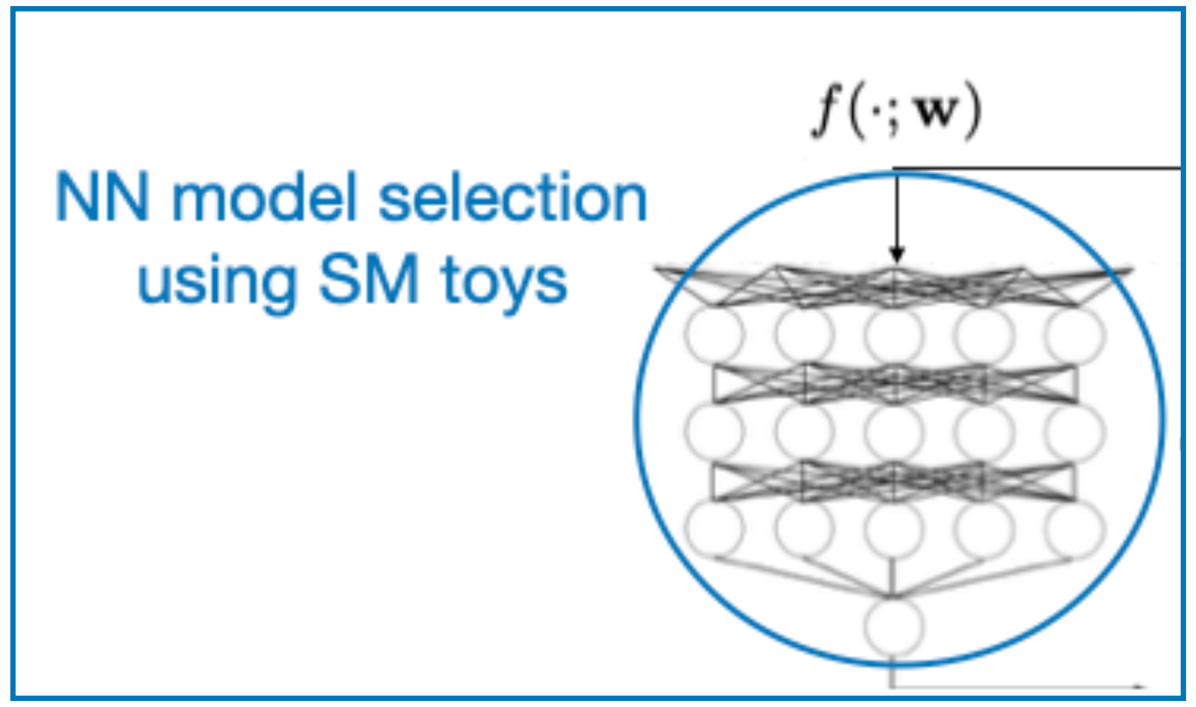
$$r(x;\boldsymbol{\nu}) \equiv \frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$$

New Physics Learning Machine (NPLM)

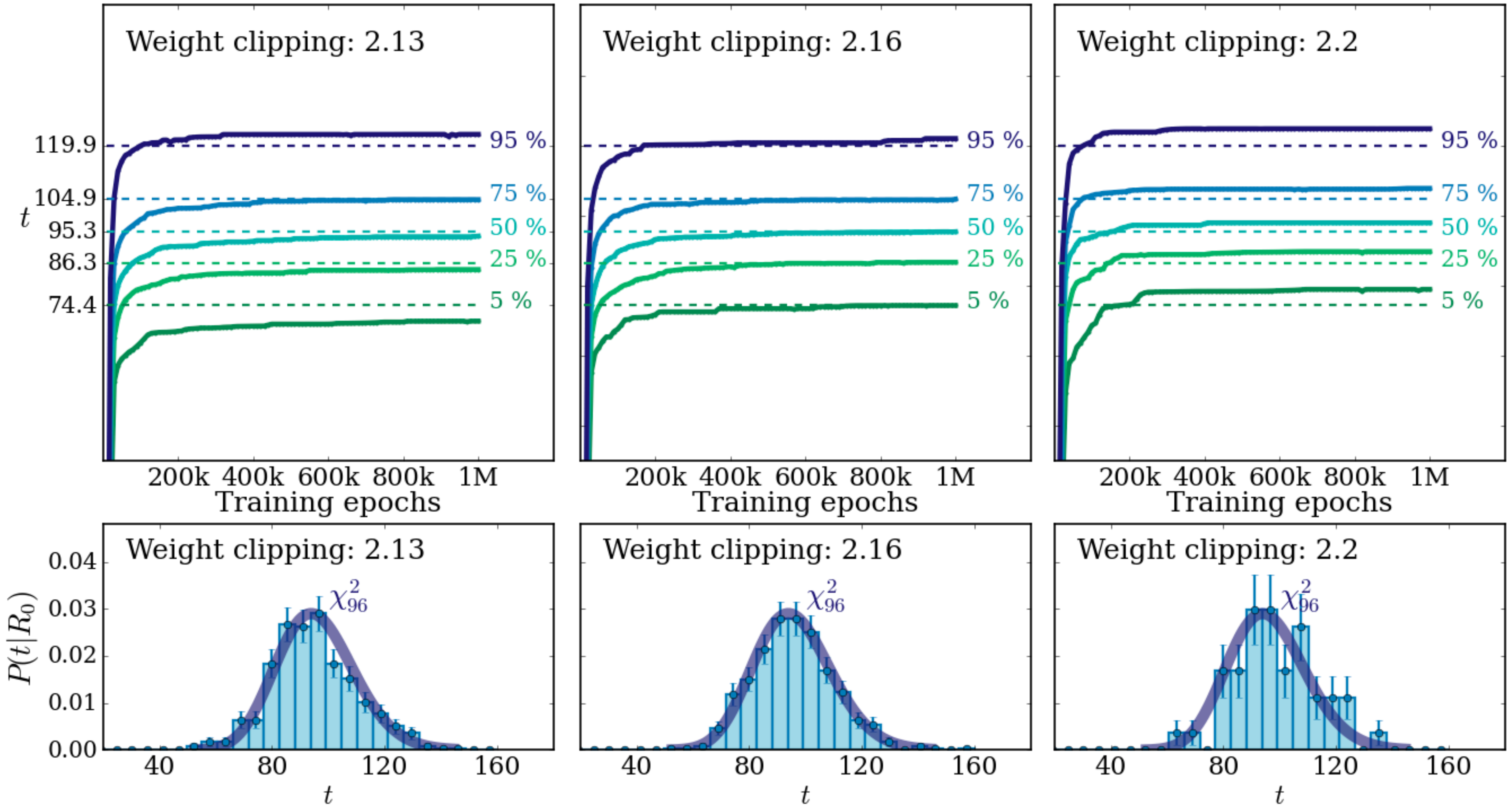
Including systematic uncertainties

NN tuning

Architecture: 5x3 (96 trainable parameters)



Weight clipping tuning, targeting a χ^2 distribution



- Loss function (no sys. unc.):

$$\bar{L} = \sum_{x \in R} w_x (e^{f(x;\,\mathrm{w})} - 1) - \sum_{x \in D} f(x;\,\mathrm{w})$$

- Both Reference and Data are sampled according to the Reference hypothesis

$$R, D \sim \mathrm{R}_0$$

Asymptotic formula: χ^2_{df}, $df = 96$

New Physics Learning Machine (NPLM)

Including systematic uncertainties

r modelling

$$r(x;\boldsymbol{\nu}) \equiv \frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})}$$

Normalization uncertainties:

Analytic description

$$r(x;\boldsymbol{\nu}) \equiv \frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})} = \exp\left[\sum_{i=1}^{N_\nu} \nu_i\right]$$

Shape uncertainties:

Taylor's expansion around the nuisance central value

$$\hat{r}(x;\nu) = \exp\left[\hat{\delta}_1(x)\,\nu + \hat{\delta}_2(x)\,\nu^2 + \dots\right]$$

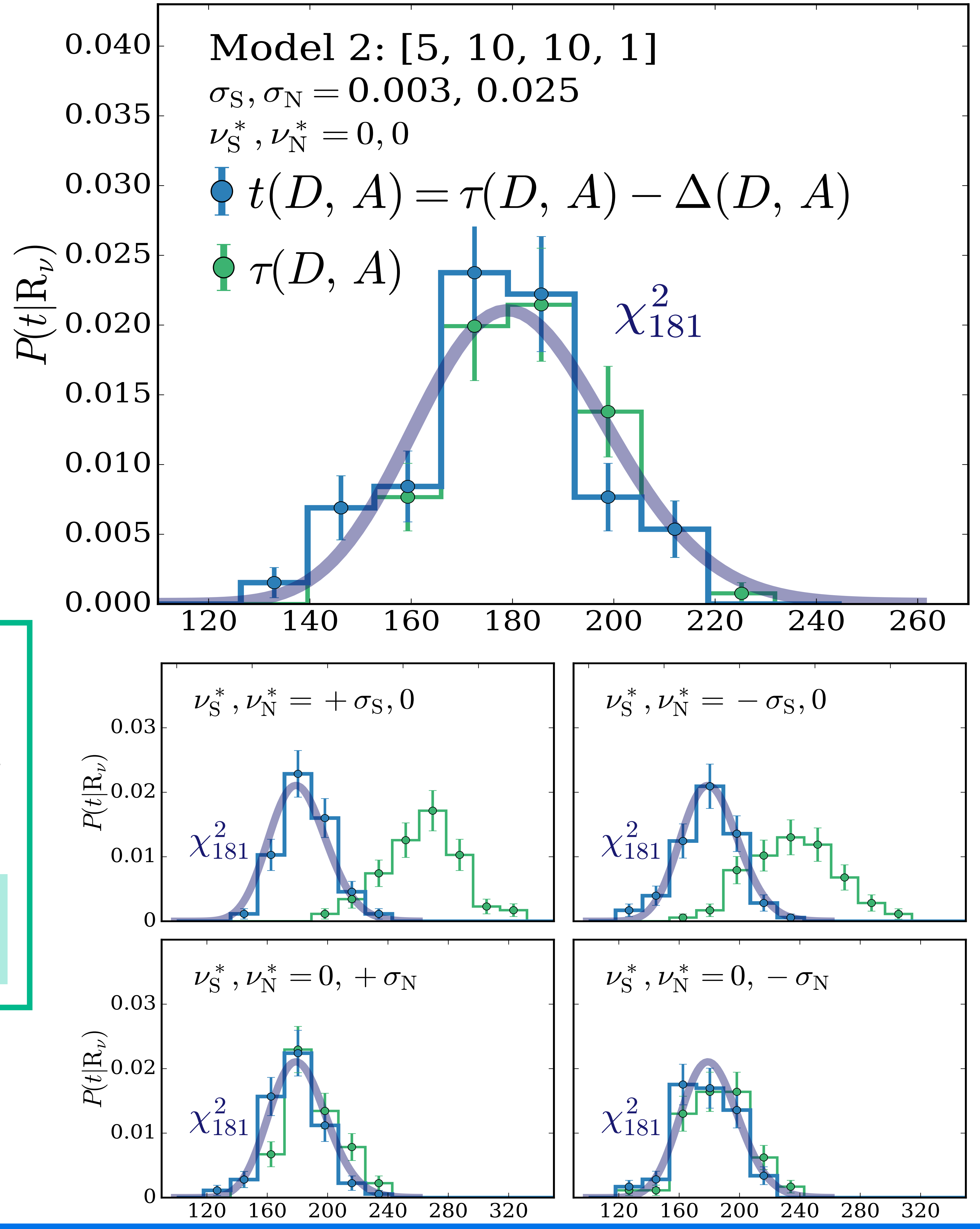
New Physics Learning Machine (NPLM)

Including systematic uncertainties

Validation of the $(\tau - \Delta)$ procedure

"Toy Data" : test the procedure on simulated toys following the Reference (SM) hypothesis with generation value for the nuisance parameters $\nu^* = \pm\sigma_\nu$:

$$\mathcal{D} \sim \mathrm{R}_{\boldsymbol{\nu}^*}, \quad \boldsymbol{\nu}^* = \pm\sigma_{\boldsymbol{\nu}}$$



The $\bar{t}$ distribution under the reference hypothesis $\mathrm{R}_{\boldsymbol{\nu}^*}$ is **compatible with the target** $\chi^2_{|\mathbf{w}|}$ for values of the true nuisance parameters within the uncertainty ($\boldsymbol{\nu}^* = \pm\sigma_{\boldsymbol{\nu}}$).

$\bar{t}$ is **independent** of the true value of the nuisance parameters!

We can build a *frequentistic* test statistic relying on the asymptotic $\chi^2_{|\mathbf{w}|}$.

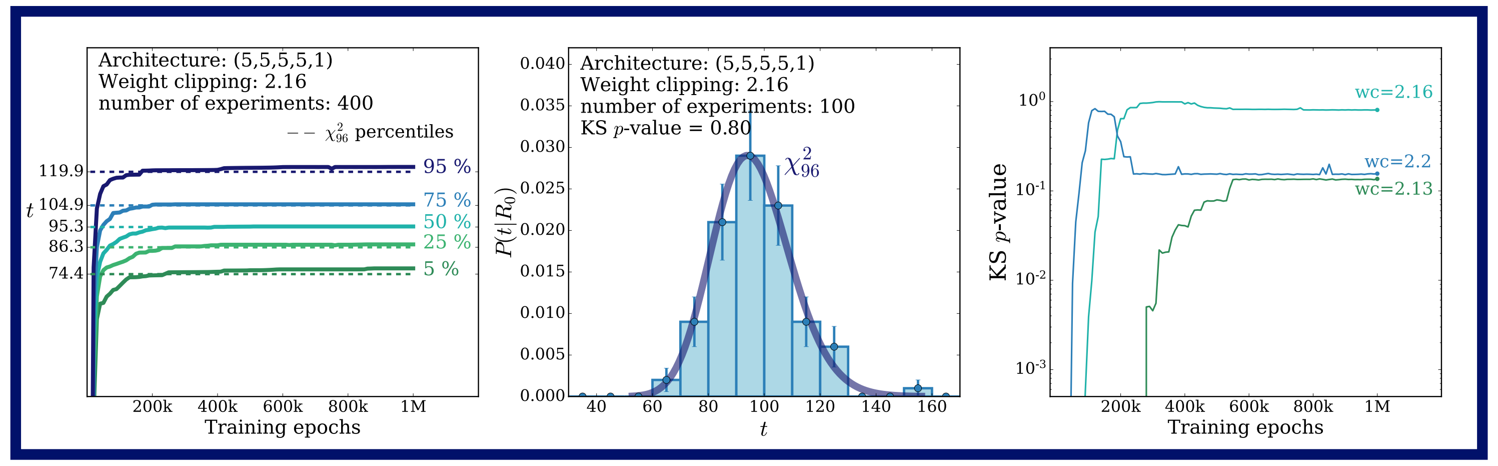
New Physics Learning Machine (NPLM)

Including systematic uncertainties

Final procedure in steps:

1. **NN MODEL SELECTION**:

   weight clipping tuning → target $\chi^2_{|\mathbf{w}|}$;



2. **NUISANCE TAYLOR'S EXPANSION LEARNING**:

   modelling $\hat{r}(x;\nu) = \exp\left[\hat{\delta}_1(x)\,\nu + \hat{\delta}_2(x)\,\nu^2 + \ldots\right]$;

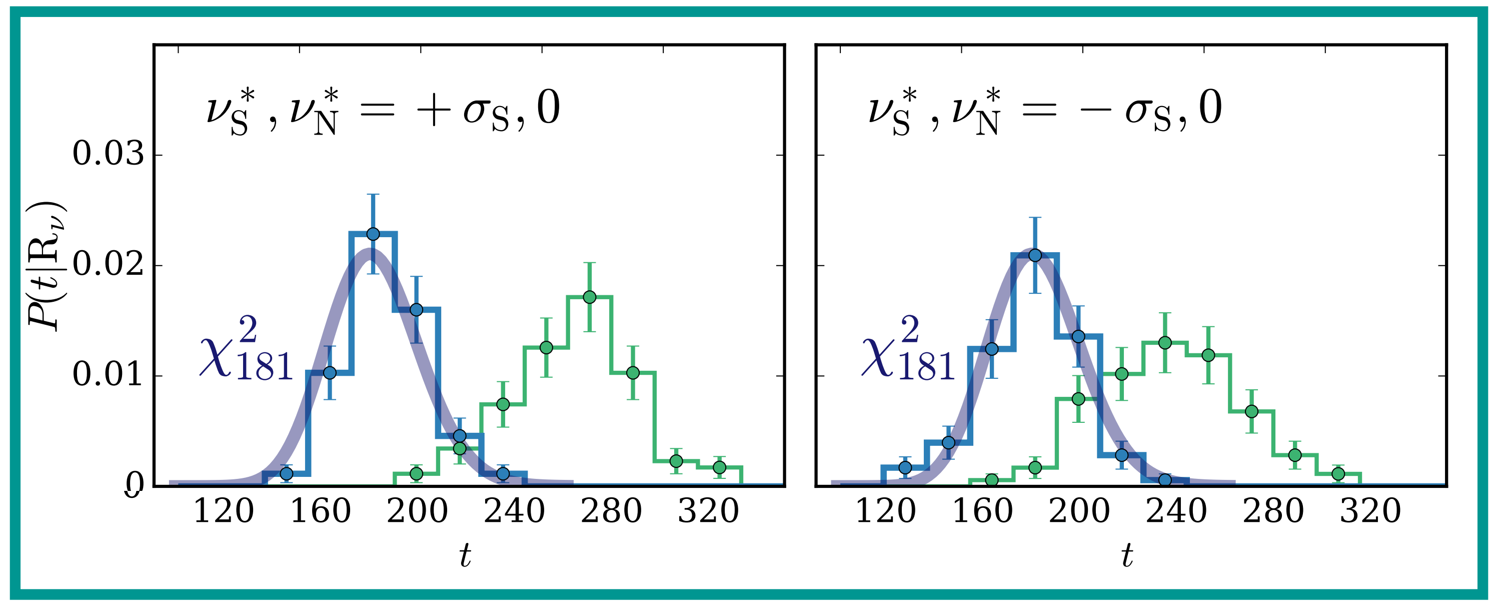
$$\hat{r}(x;\nu) = \exp\left[\hat{\delta}_1(x)\,\nu + \hat{\delta}_2(x)\,\nu^2 + \ldots\right]$$

NN 1   NN2   ...

3. **VALIDATION**:

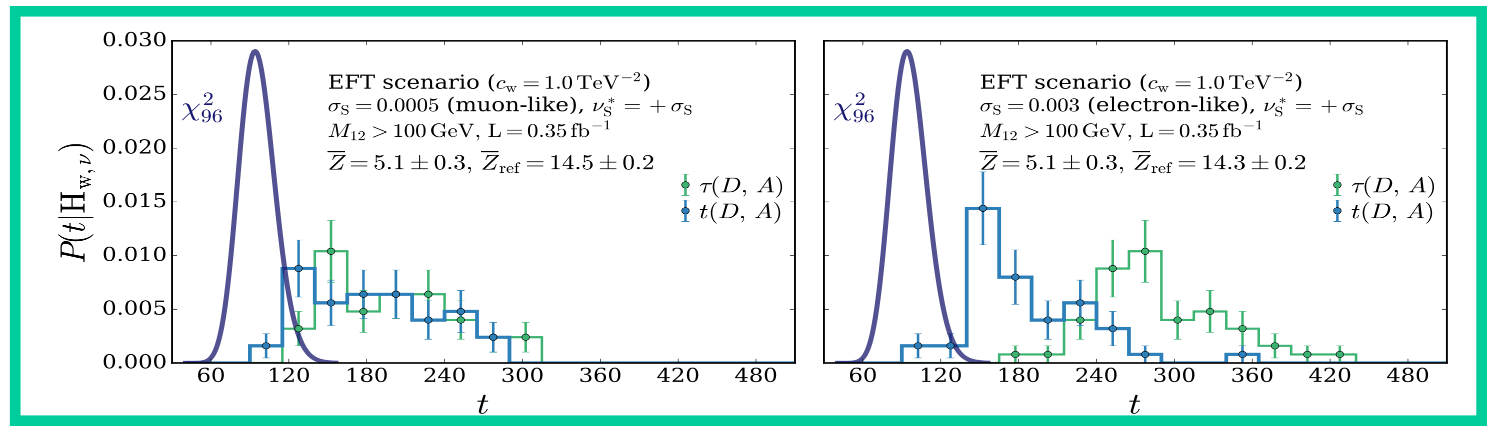
$$\mathcal{D} \sim \mathrm{R}_{\boldsymbol{\nu}^*}, \quad \boldsymbol{\nu}^* = \pm\sigma_{\boldsymbol{\nu}}$$

   Verifying that the target $\chi^2_{|\mathbf{w}|}$ is always recovered;



4. **TESTING THE DATA**:

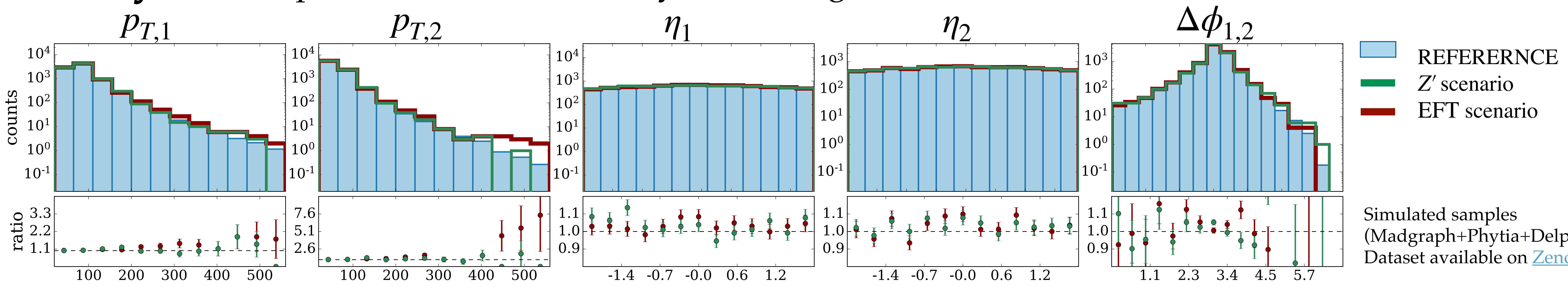
   running the procedure on real data.



Two-body final state at the LHC

Two-body final state at the LHC

5D analysis — Input: set of variables fully describing the kinematic of the final state



Simulated samples (Madgraph+Phytia+Delphes)
Dataset available on Zenodo

Uncertainties on the Reference hypothesis (SM):

- Global normalization effect: $\sigma_N = 2.5\,\%$
- Momentum scale effect:

$$p_{T\,1,2}^{(b,e)} = \exp\left[\nu_s \sigma_s^{(b,e)}/\sigma_s^{(b)}\right] p_{T\,1,2}^{(b,e)}$$

(b) barrel region $|\eta| < 1.2$, (e) endcaps region $|\eta| \geq 1.2$

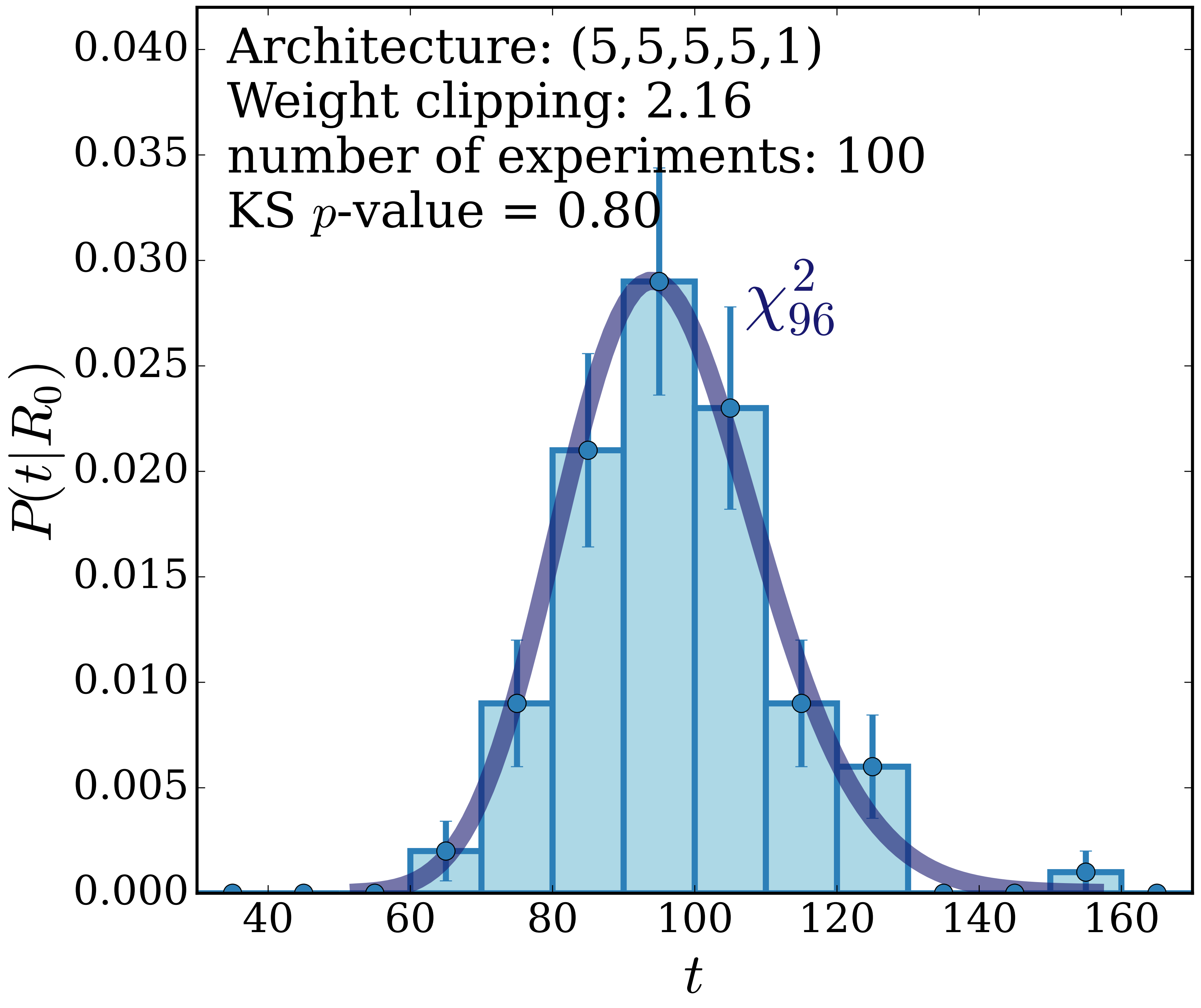
- Muon-like regime: $\sigma_S^{(b)} = 0.05\,\%$, $\sigma_S^{(e)} = 0.15\,\%$
- Electron-like regime: $\sigma_S^{(b)} = 0.3\,\%$, $\sigma_S^{(e)} = 0.9\,\%$
- Tau-like regime: $\sigma_S^{(b)} = \sigma_S^{(e)} = 3\,\%$

"Learning New Physics from an Imperfect Machine" Eur. Phys. J. C

Two-body final state at the LHC

1) NN setup for the τ term:

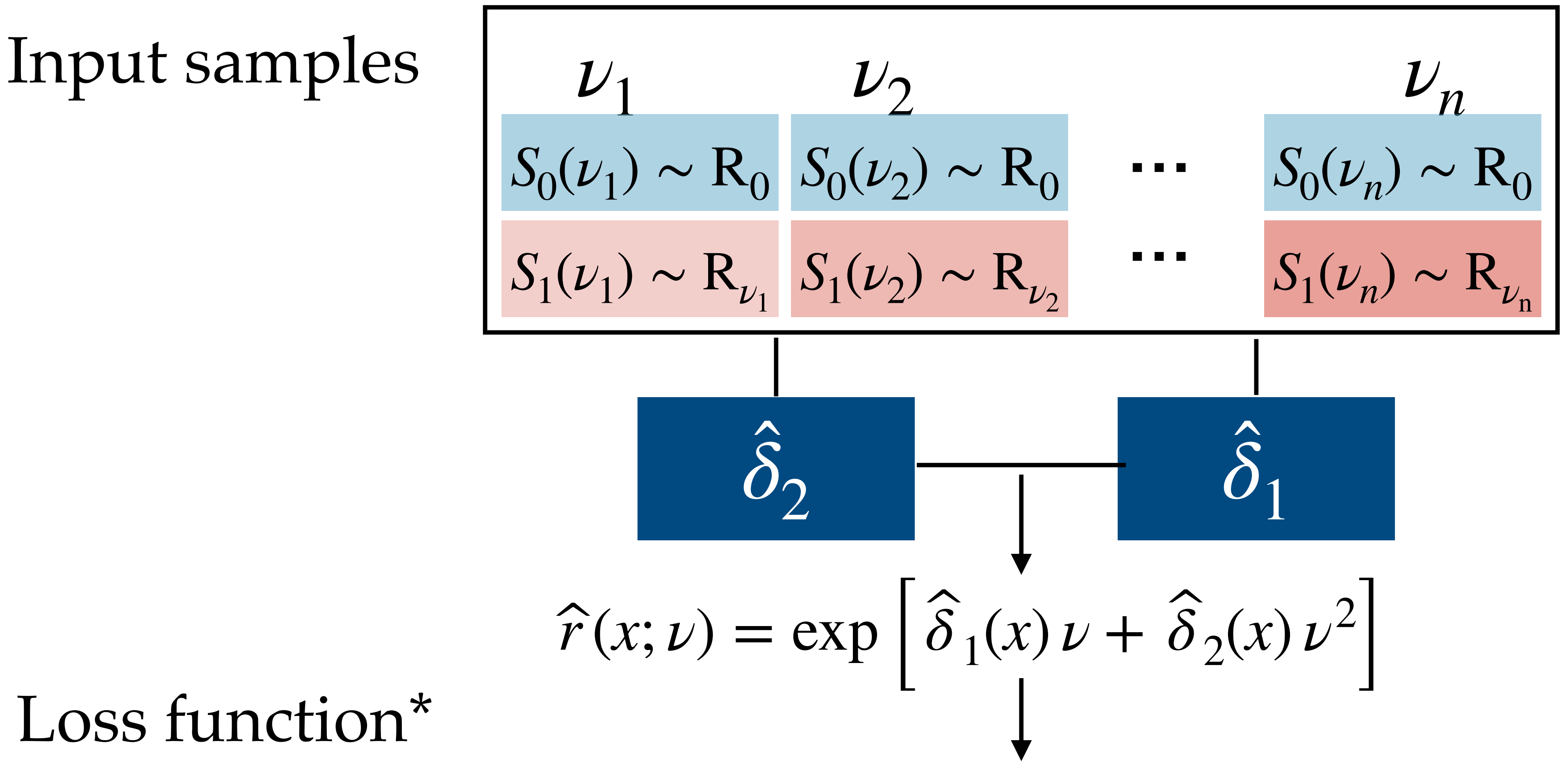
- Data sample $N(D) \sim 8500$ events
- Reference sample $N(R) = 5 \times N(D) \sim 42\,500$
- Regularised DNN (weight clipping tuning)



2) Taylor's expansion learning for $r(x, \nu)$:

Training a neural network model to learn each coefficient of the Taylor's expansion of $r(x;\nu) = \dfrac{n(x|R_\nu)}{n(x|R_0)}$

(Parametrized classifier)



$$L[\hat{\delta}(\,\cdot\,)] = \sum_{\nu_i}\left[\sum_{e\in S_0(\nu_i)} w_e c(x_e)^2 + \sum_{e\in S_1(\nu_i)} w_e\left[1 - c(x_e)\right]^2\right], \quad c(x) = \frac{1}{1+\hat{r}(x;\nu)}$$

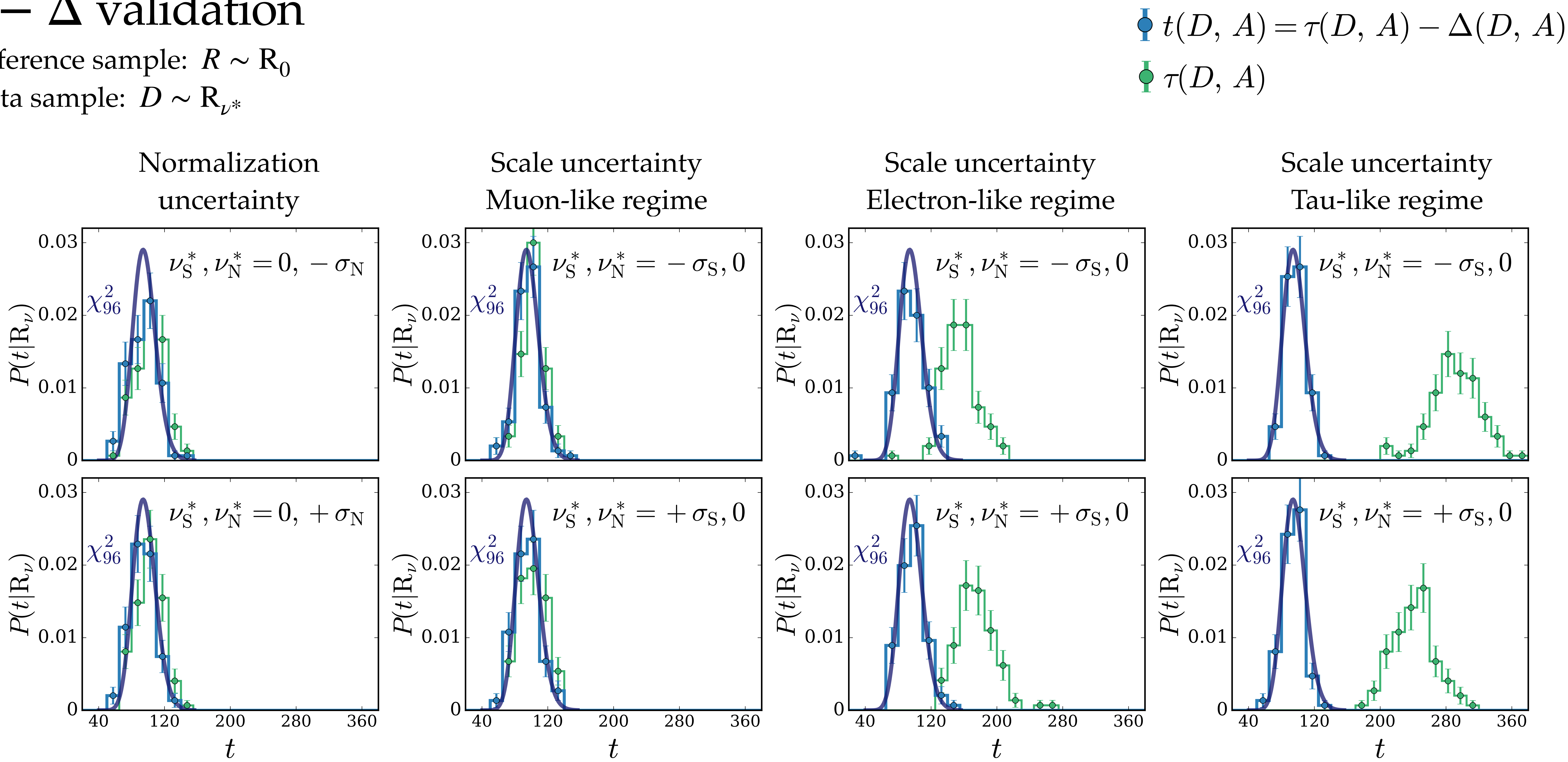
* *Parametrized classifiers for optimal EFT sensitivity* arXiv:2007.10356

Two-body final state at the LHC

$\tau - \Delta$ validation

Reference sample: $R \sim \mathrm{R}_0$

Data sample: $D \sim \mathrm{R}_{\nu^*}$



DNN [5-5-5-5-1], #trainable parameters = 96, weight clipping = 2.16

Two-body final state at the LHC

Sensitivity to New Physics scenarios

Resonance in the two-body invariant mass

- **Z' scenario:** new vector boson with the same SM coupling as the Z boson and mass of 300 GeV.
  - Muon-like, electron-like regimes:
    $M_{12} > 100\,\text{GeV}, L = 0.35\,\text{fb}^{-1}, \text{N(S)} = 120$
  - Tau-like regime:
    $M_{12} > 120\,\text{GeV}, L = 1.1\,\text{fb}^{-1}, \text{N(S)} = 210$

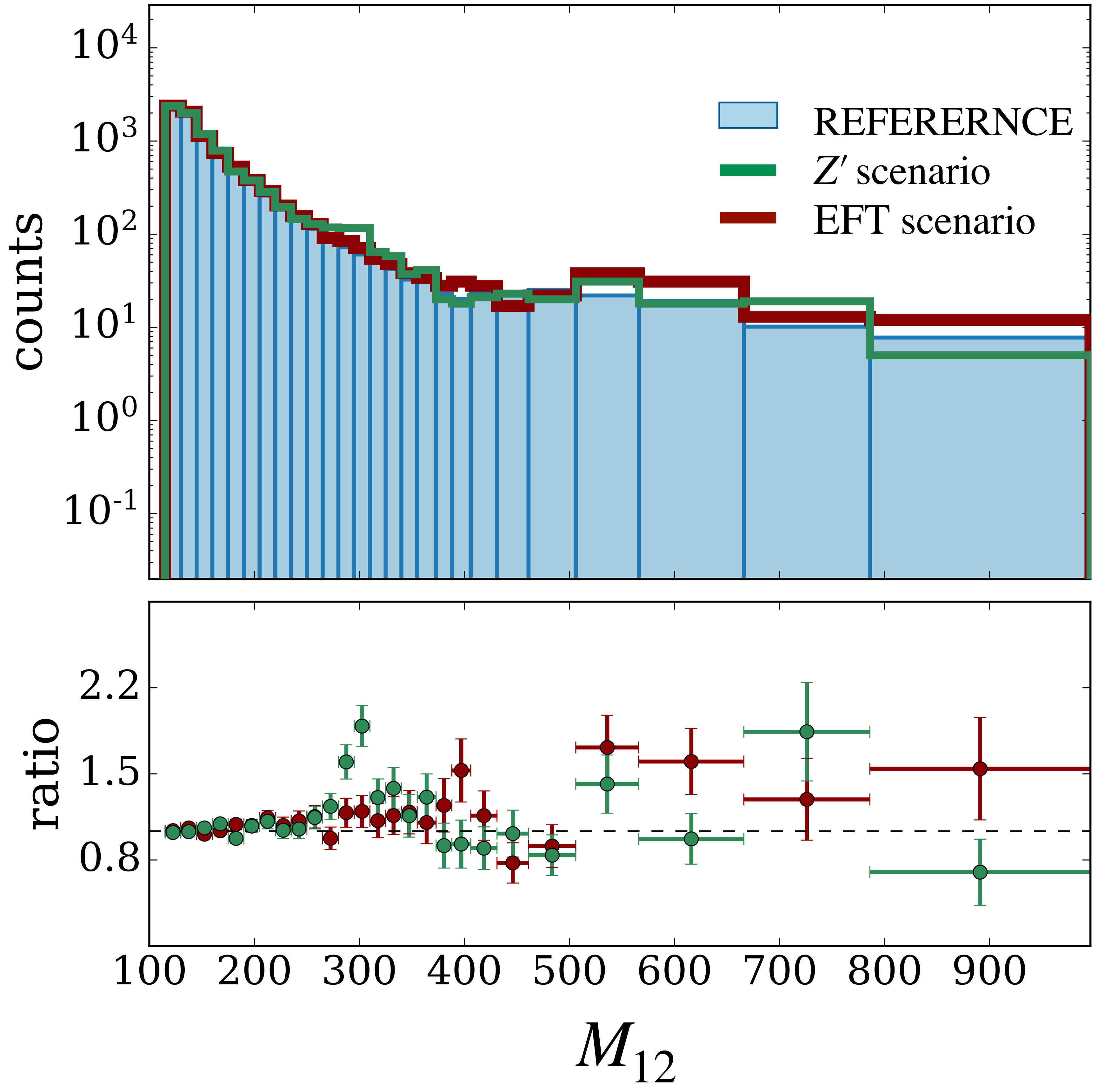
Non resonant excess in the tail of the two-body invariant mass

- **EFT scenario:** dimension-6 4-fermions-contact operator:

$$\frac{c_W}{\Lambda} J^a_{L\mu} J^\mu_{La}.$$

  - Muon-like, electron-like regimes:
    $M_{12} > 100\,\text{GeV}, L = 0.35\,\text{fb}^{-1}, c_W = 1.0\,\text{TeV}^{-2}$
  - Tau-like regime:
    $M_{12} > 120\,\text{GeV}, L = 1.1\,\text{fb}^{-1}, c_W = 0.25\,\text{TeV}^{-2}$

Example:
Tau-like regime



NOTE:
M_{12} is **not** given as an input to the algorithm!

Two-body final state at the LHC

Sensitivity to New Physics scenarios

$t(D, A) = \tau(D, A) - \Delta(D, A)$

$\tau(D, A)$

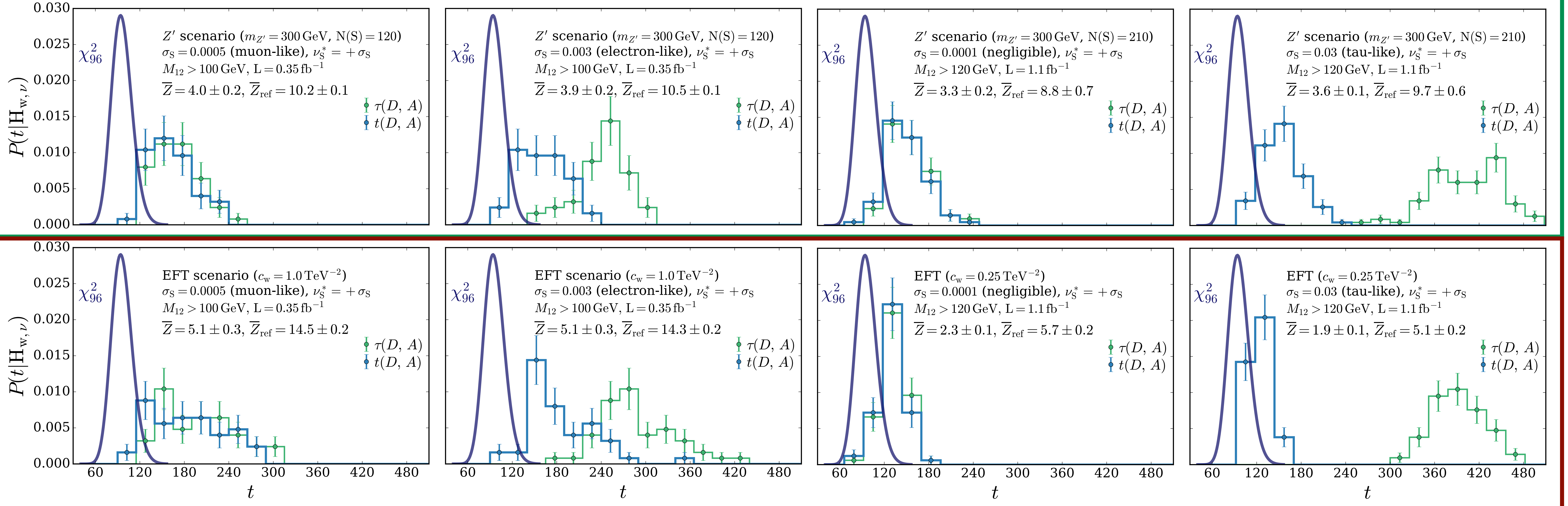
Muon-like regime | Electron-like regime | Negligible systematic uncertainties | Tau-like regime

Z' scenario

$\nu_S^* = +\sigma_S$

EFT scenario

$\nu_S^* = +\sigma_S$



Z-score: $Z = \Phi^{-1}\left[1 - p\right]$

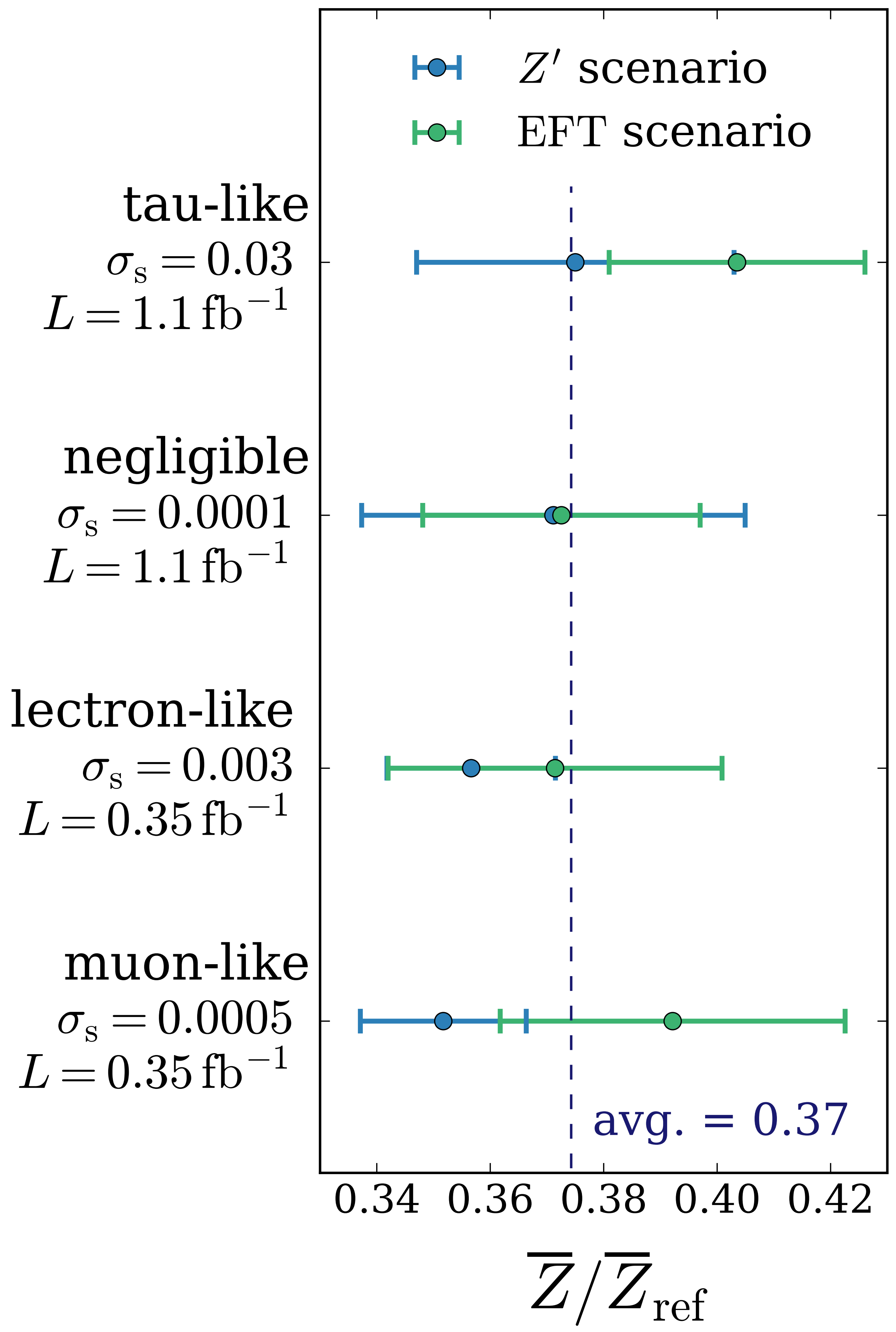
$\overline{Z}$: Z-score from the median of the empirical $t(D, A)$ distribution

Two-body final state at the LHC

Sensitivity to New Physics scenarios

Summary of the results:

- Comparable performances in the resonant and non-resonant scenarios:
  - NPLM is simultaneously sensitive to any source of New Physics;
- Comparable performances at different systematic uncertainties regimes:
  - NPLM is robust against the presence of systematic uncertainties;
  - the presence of systematic uncertainties affects NPLM in the same measure as any other hypothesis test;
- No information about the New Physics signal has been provided to the algorithm at any step of its implementation:
  - The performances of NPLM are lower than any model-dependent strategy by construction ($\overline{Z}/\overline{Z}_{\rm ref} = 0.37$);



Z-score: $Z = \Phi^{-1}\left[1-p\right]$

- $\overline{Z}$: Z-score from NPLM
- $\overline{Z}_{\rm ref}$: Z-score from a model-dependent (optimized) test statistics

Two-body final state at the LHC

Sensitivity to New Physics scenarios

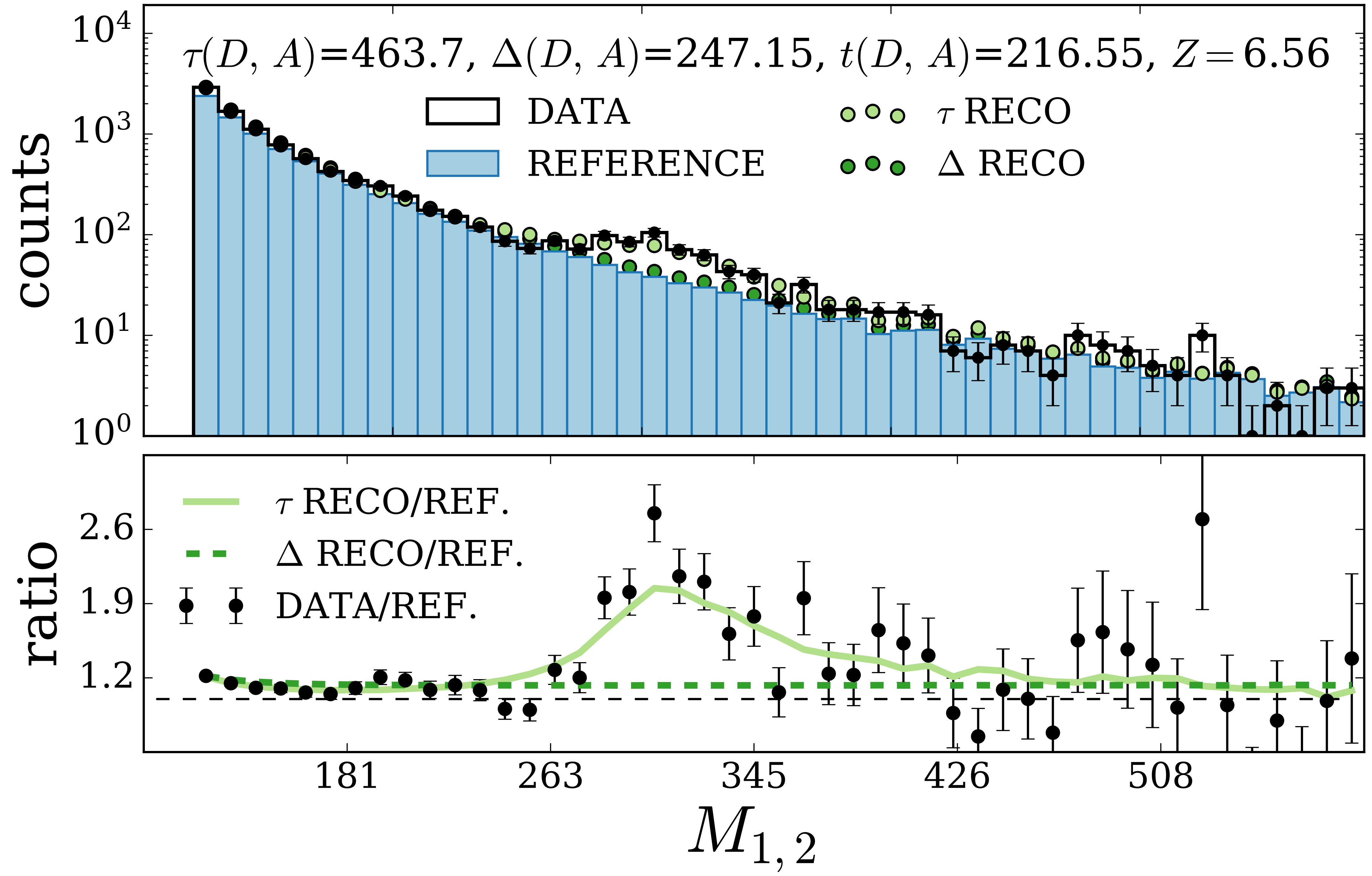
Signal reconstruction with the NN:

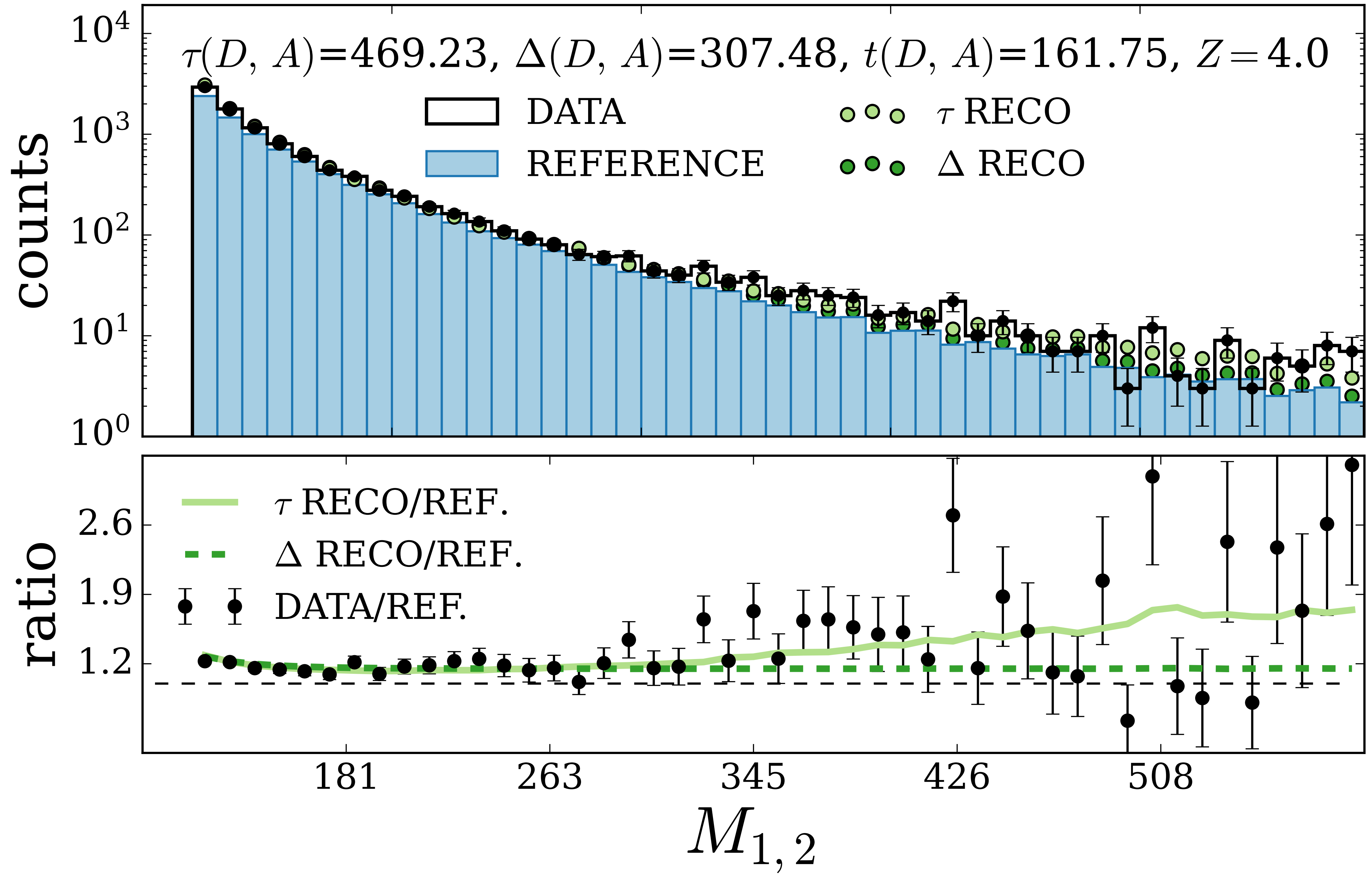
Architecture: [5-5-5-5–1] (96 dof), weigh clipping 2.15, L = 240 fb^{-1}

τ reconstruction: $n(x\,|\,\mathrm{H}_{\hat{\mathbf{w}},\hat{\nu}}) = n(x\,|\,\mathrm{R}_0)\,\dfrac{n(x\,|\,\mathrm{R}_{\hat{\nu}})}{n(x\,|\,\mathrm{R}_0)}\,e^{f(x;\,\hat{\mathbf{w}})}$

Δ reconstruction: $n(x\,|\,\mathrm{R}_{\hat{\nu}})$

NOTE:
M_{12} is **not** given as an input to the algorithm!





"Learning Multivariate New Physics" Eur. Phys. J. C

Two-body final state at the LHC

NPLM for signal model independent New Physics searches:

Ready to be performed on a real analysis at the LHC!

- ✓ Heuristic method to setup **multivariate** analysis
- ✓ Strategy to account for **systematic uncertainties**

Limitations

- Accuracy of the Reference sample
- Accuracy of the systematic uncertainties modelling
- Training time (with NN)

limit the actual **luminosity** that we are allowed to inspect, but do not obstacle the applicability of NPLM.

Next steps?

Ongoing work

- Speed up? NN $\rightarrow$ kernel models (see next talk)
- NN regularization (alternatives to weight clipping?) $\rightarrow$ performances
- NN architecture?

Ongoing work

- Speed up? NN → kernel models (see next talk)
- NN regularization (alternatives to weight clipping?) → performances
- NN architecture?

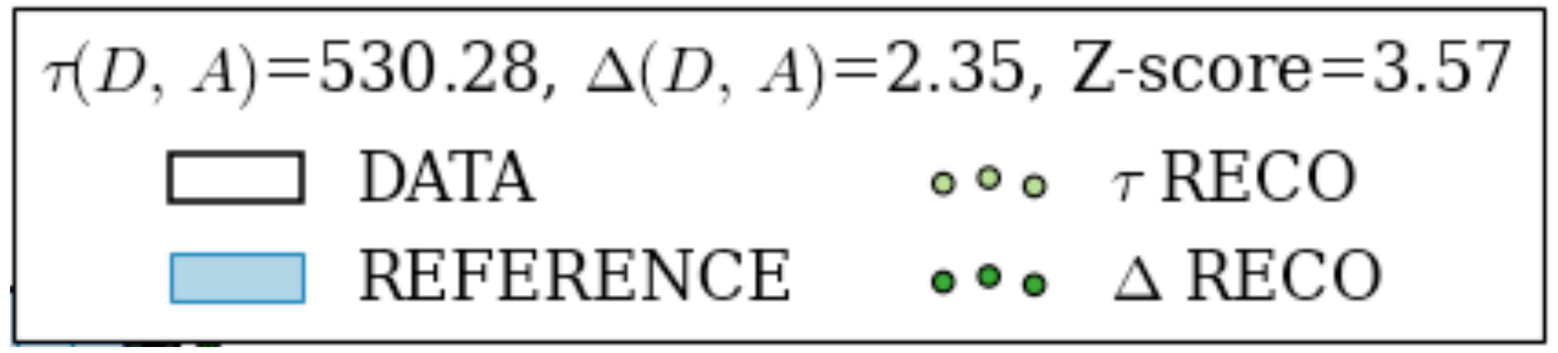
More applications

- Searching NP in harder nD problems
- Data quality monitoring (more about it in the next talk)
- Testing generators (nD)

Harder tasks: nD analysis (n large)

ZZ to 4 leptons final state (7D)

Signal reconstruction with the NN:



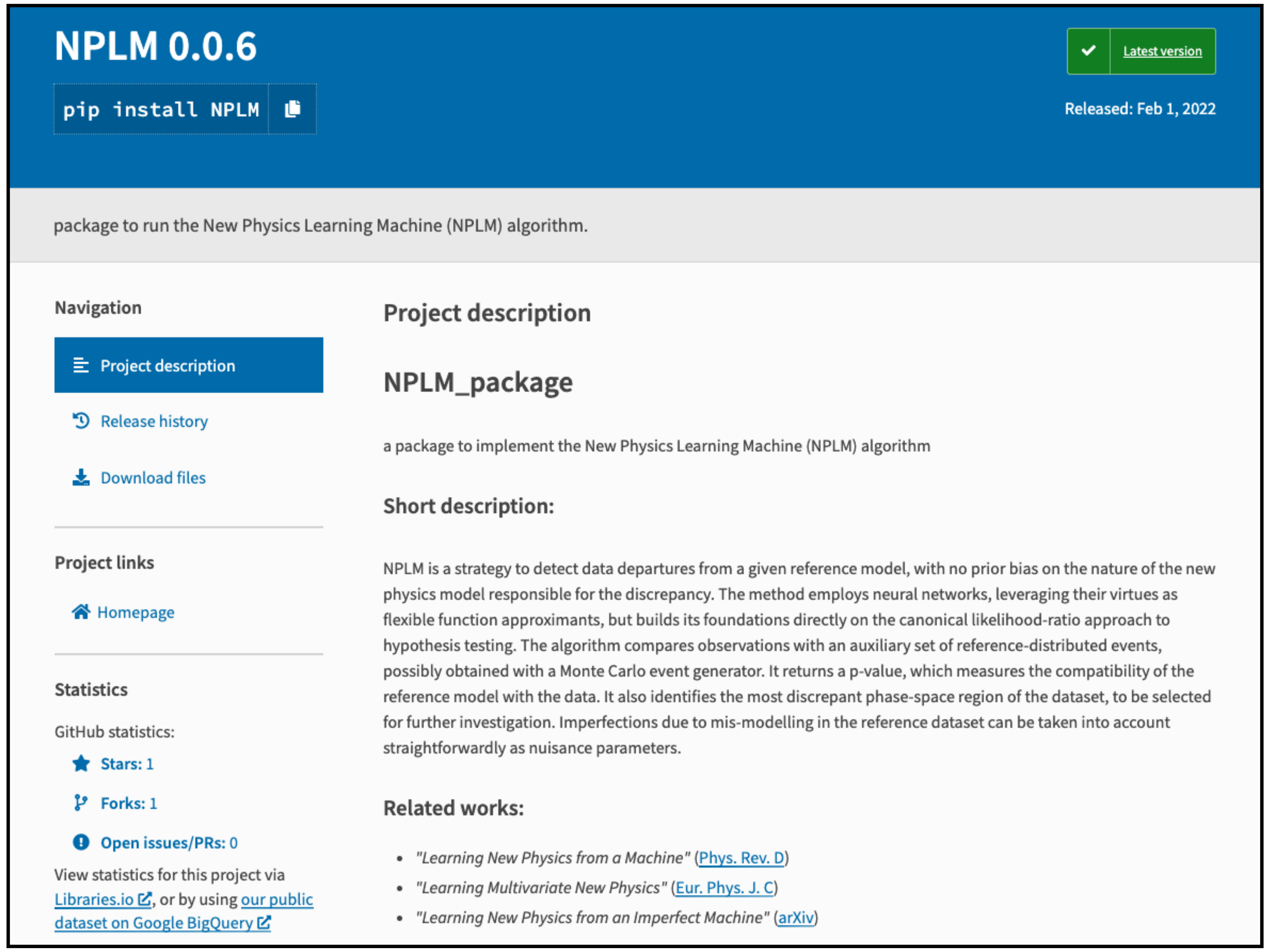
Architecture: [7-16-16-1] (417 dof), weigh clipping 1.75

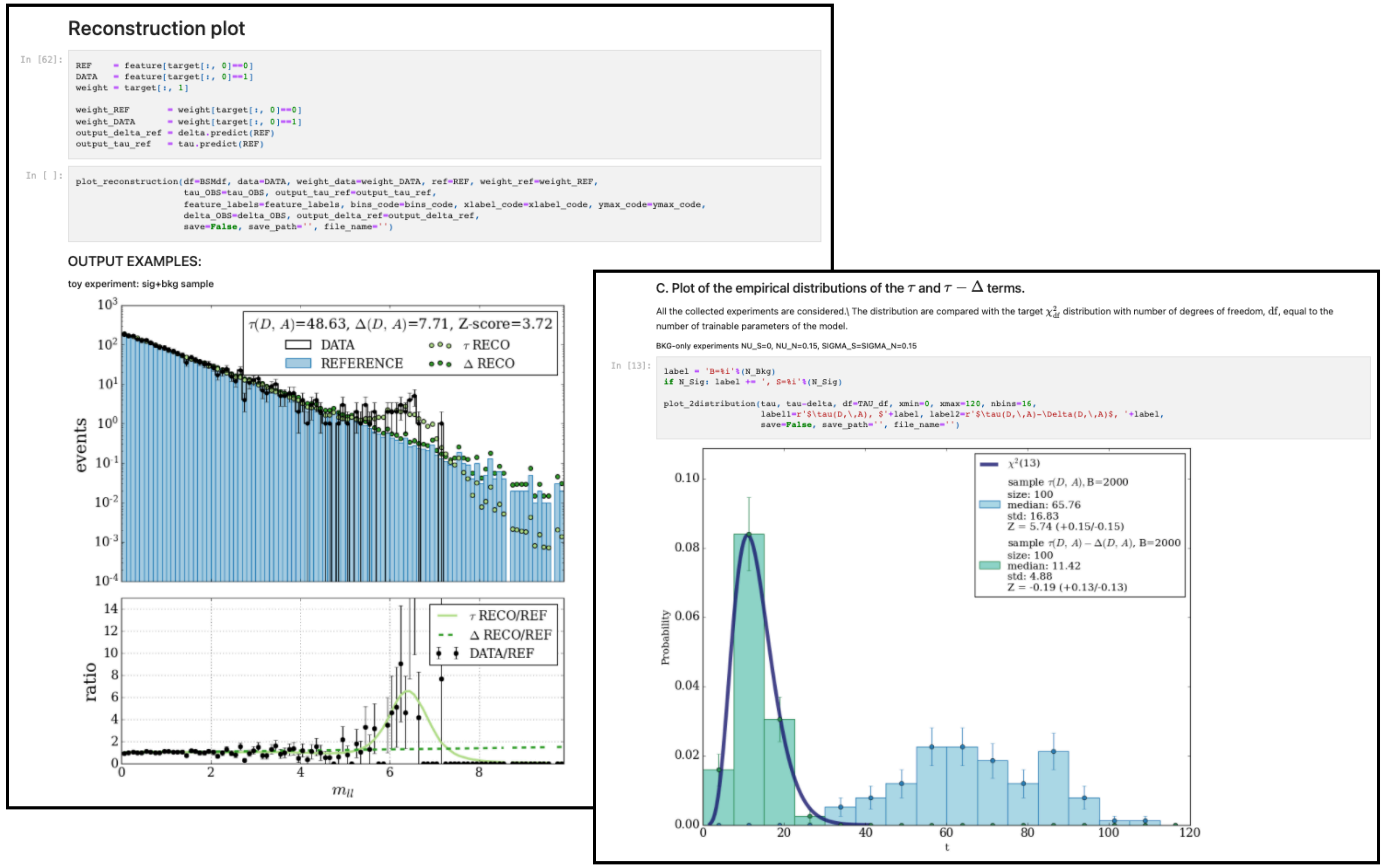
Harder tasks: *n*D analysis (*n* large)

Ongoing work for optmizing NPLM sensitivity in high dimensional problems

Getting started with NPLM

- NPLM package: python-based package to run the NPLM analysis strategy
- Tutorial on 1D toy model for getting started





Backup

New Physics Learning Machine (NPLM)

Main Concepts (negligible uncertainties)

Maximum Likelihood from minimal loss:

Test statistic

$$\bar{t}(\mathcal{D}) = 2\max_{\mathbf{w}} \log\left[\frac{\mathcal{L}(\mathrm{H}_{\mathbf{w}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})}\right]$$

w: trainable parameters on the NN model

D: data sample

R: reference sample (built according to the R_0 hypothesis); could be weighted (w_x)

$$= 2\,\log\left[\frac{e^{-N(\widehat{\mathbf{w}})}}{e^{-N(\mathrm{R})}}\prod_{x\in\mathcal{D}}\frac{n(x|\widehat{\mathbf{w}})}{n(x|\mathrm{R})}\right] = -2\,\underset{\{\mathbf{w}\}}{\mathrm{Min}}\left[N(\mathbf{w}) - N(\mathrm{R}) - \sum_{x\in\mathcal{D}} f(x;\mathbf{w})\right]$$

$$N(\mathbf{w}) = \frac{N(\mathrm{R})}{\mathcal{N}_{\mathcal{R}}}\sum_{x\in\mathcal{R}} e^{f(x;\mathbf{w})}$$

$$= -2\min_{\mathbf{w}}\{\bar{L}[f(\cdot;\,\mathbf{w})]\}$$

Loss function

$$\bar{L}\,[f(x;\,\mathbf{w})] = -\sum_{x\in\mathcal{D}}[\,f(x;\,\mathbf{w})] + \sum_{x\in\mathcal{R}} w_x\left[e^{f(x;\,\mathbf{w})} - 1\right]$$

Assumptions:

- $N_R \gg N_D$ the statistical fluctuations of the reference sample are negligible.
- the weights of the reference sample (w) are such that the reference sample is normalised to match the data sample luminosity $\sum_{x\in R} w_x = N(R_0)$

"Learning New Physics from a Machine" Phys. Rev. D

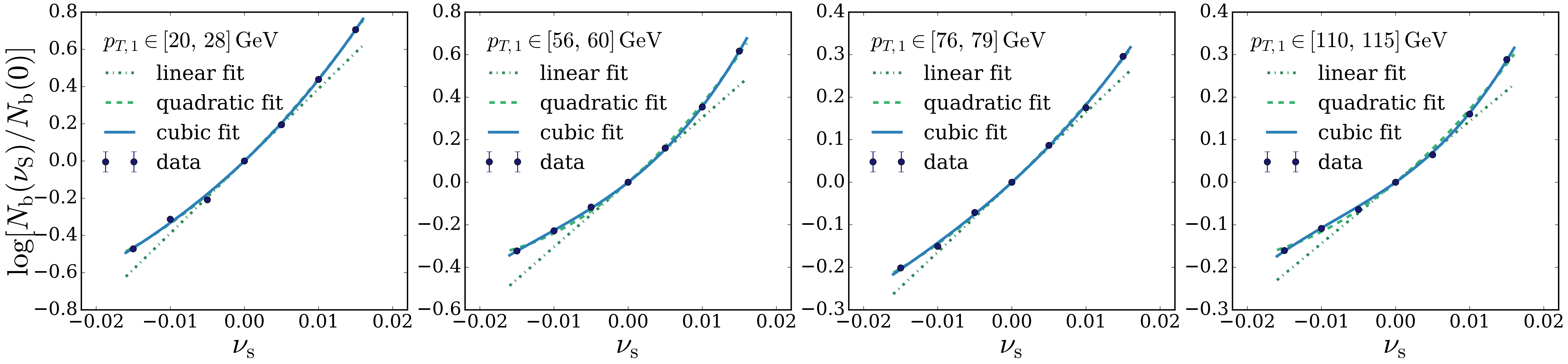
Including systematic uncertainties

Shape uncertainties: Learning the nuisance Taylor's expansion

Preliminary study

Preliminary binned analysis to determine the proper order for the Taylor's expansion

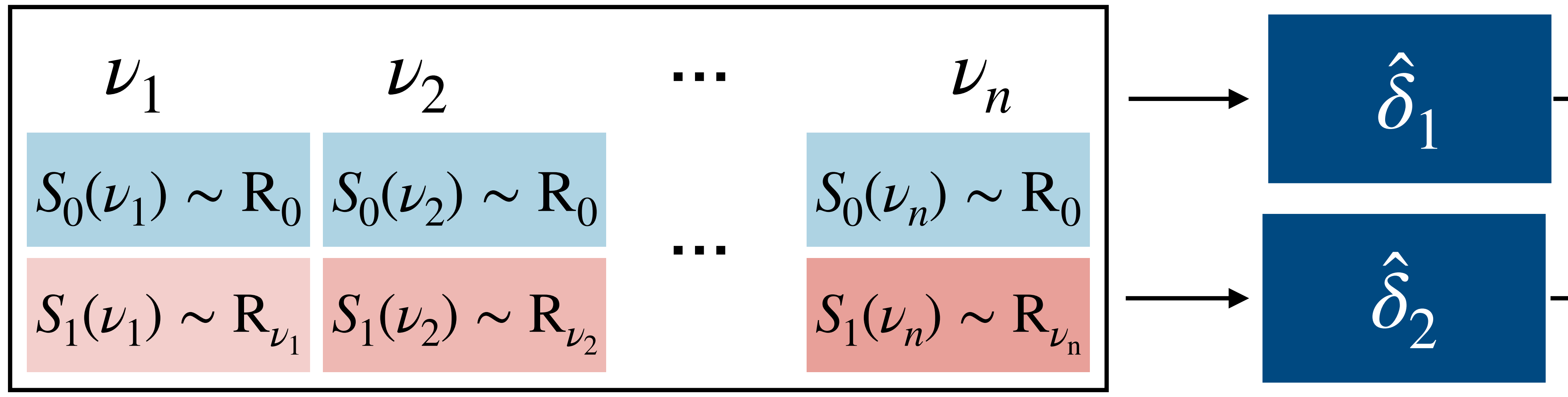
Electron-like regime



Taylor's expansion learning

Training a neural network model to learn each coefficient of the Taylor's expansion of $r(x;\nu) = \dfrac{n(x\,|\,R_\nu)}{n(x\,|\,R_0)}$

Input samples



$$\hat{r}(x;\nu) = \exp\left[\hat{\delta}_1(x)\,\nu + \hat{\delta}_2(x)\,\nu^2\right]$$

Loss function*

$$L[\hat{\delta}(\cdot)] = \sum_{\nu_i}\left[\sum_{e\in S_0(\nu_i)} w_e c(x_e)^2 + \sum_{e\in S_1(\nu_i)} w_e\left[1 - c(x_e)\right]^2\right], \qquad c(x) = \frac{1}{1+\hat{r}(x;\nu)}$$

* *Parametrized classifiers for optimal EFT sensitivity* arXiv:2007.10356

New Physics Learning Machine (NPLM)

Including systematic uncertainties