ACAT 2021
20th International Workshop on Advanced Computing
and Analysis Techniques in Physics Research

An *Imperfect* Machine to search for New Physics

A way to include systematic uncertainties

R.T. d'Agnolo[1], ***G. Grosso***[2,3], *M. Pierini*[3], *A. Wulzer*[2,4], *M. Zanetti*[2]

[1]Université Paris Saclay, [2]Università degli Studi di Padova, [3]CERN, [4]EFPL

November 29, 2021

Outline of the talk

New Physics Learning Machine (NPLM)

- THE ALGORITHM:
    - Absence of systematic uncertainties [1, 2]
        - Main concepts
        - Multivariate analysis setup
    - Including systematic uncertainties [3]
- EXPERIMENTS:
    - 5D analysis on a di-body final state at the LHC
- CONCLUSIONS AND OUTLOOK

[1] *Learning New Physics from a Machine*, Phys. Rev. D
[2] *Learning Multivariate New Physics*, Eur. Phys. J. C
[3] *Learning New Physics from an Imperfect Machine* 2111.13633

The algorithm

New Physics Learning Machine (NPLM)

Main Concepts (negligible uncertainties)

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

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

$\mathrm{R}_{\mathbf{0}}$: null hypothesis
$\mathrm{H}_{\mathbf{w}}$: alternative hypothesis

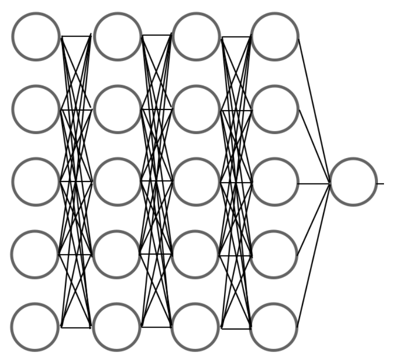
- Exploiting a Neural Network (NN) to **parametrize** the data distribution in terms of a Reference distribution ($\mathrm{R}_{\mathbf{0}}$)

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

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

New Physics Learning Machine (NPLM)

Main Concepts (negligible uncertainties)

Maximum Likelihood from minimal loss:

Test statistic

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

Loss function

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

w: trainable parameters on the NN model

D: data sample

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

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 $\sum_{x\in R} w_x = N(\mathrm{R}_0)$

New Physics Learning Machine (NPLM)

Main Concepts (negligible uncertainties)

Maximum Likelihood from minimal loss:

Test statistic

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

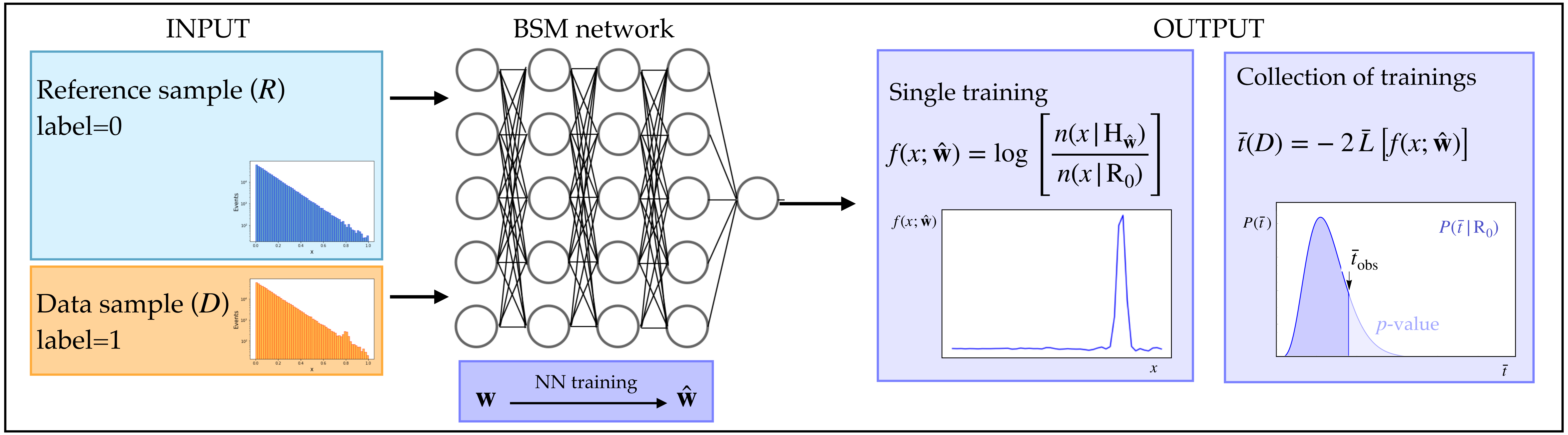
Loss function

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

$\mathbf{w}$: 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 $\sum_{x \in R} w_x = N(R_0)$



New Physics Learning Machine (NPLM)

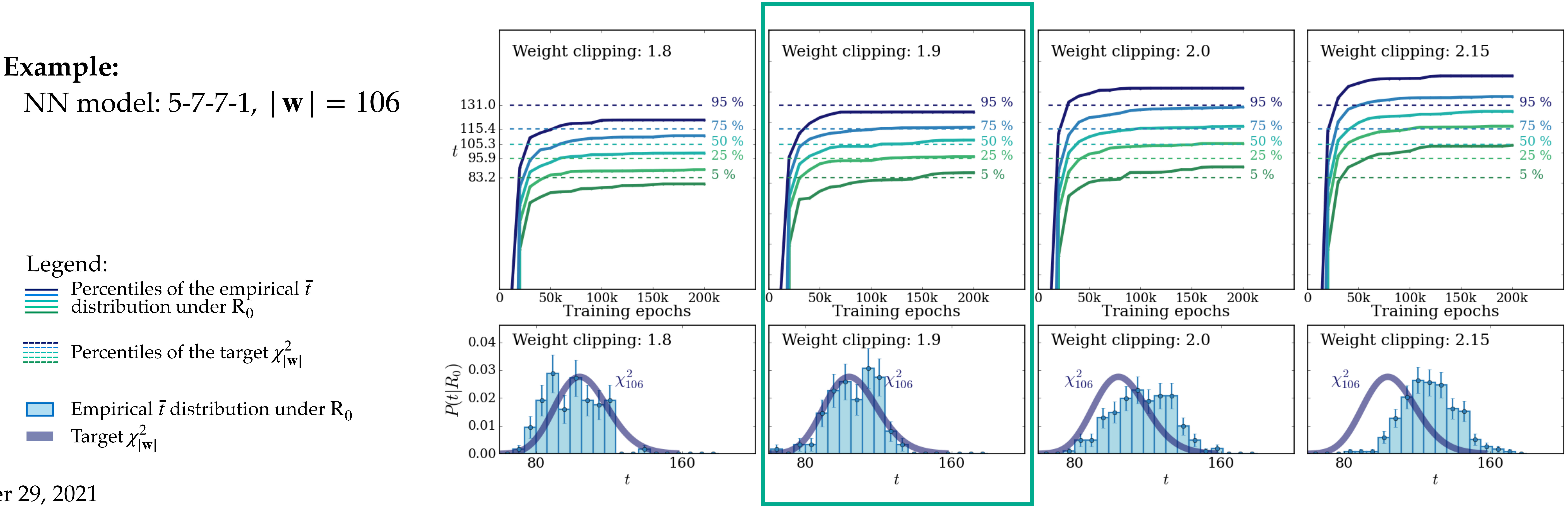
Dealing with multivariate problems (negligible uncertainties)

NN Model selection:

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, $|\mathbf{w}| = 106$



New Physics Learning Machine (NPLM)

Including systematic uncertainties

Test statistic

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

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

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

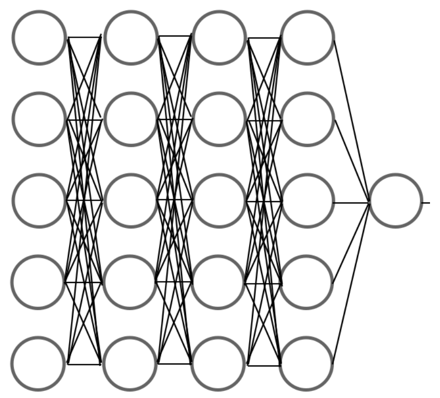
$\mathcal{D}$: data sample

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

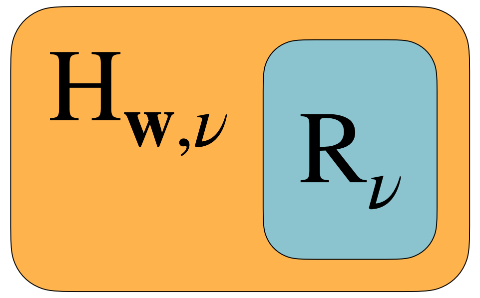
New parametrization

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

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



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



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:

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

NN 1 NN2 …

(More details later)

New Physics Learning Machine (NPLM)

Including systematic uncertainties

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 $\nu^* = \pm\sigma_\nu$:

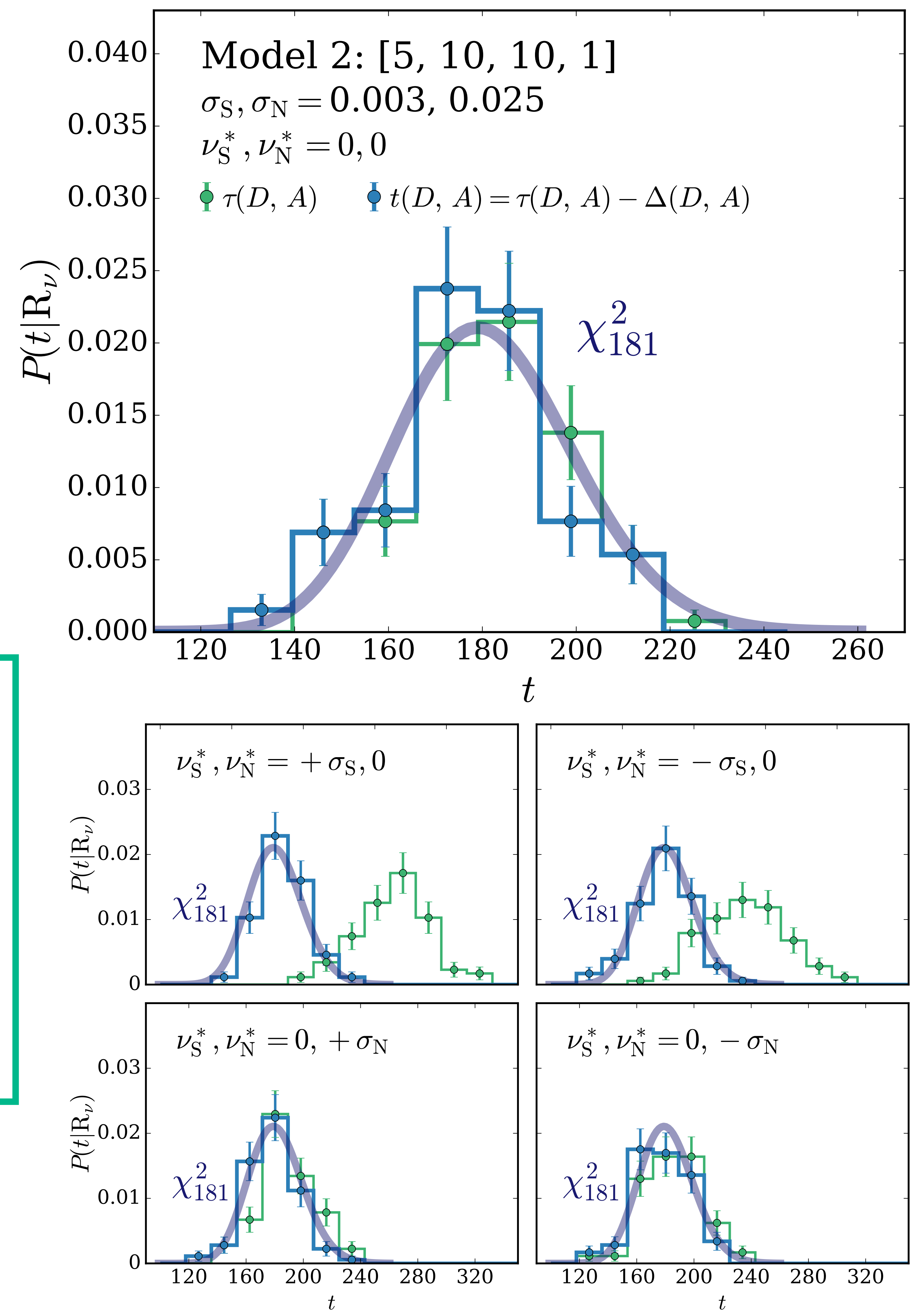
$$\mathcal{D} \sim \mathrm{R}_{\boldsymbol{\nu}^*}, \quad \boldsymbol{\nu}^* = \pm\sigma_{\boldsymbol{\nu}}$$

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 $\chi^2_{|\mathbf{w}|}$.

Example of validation plots:

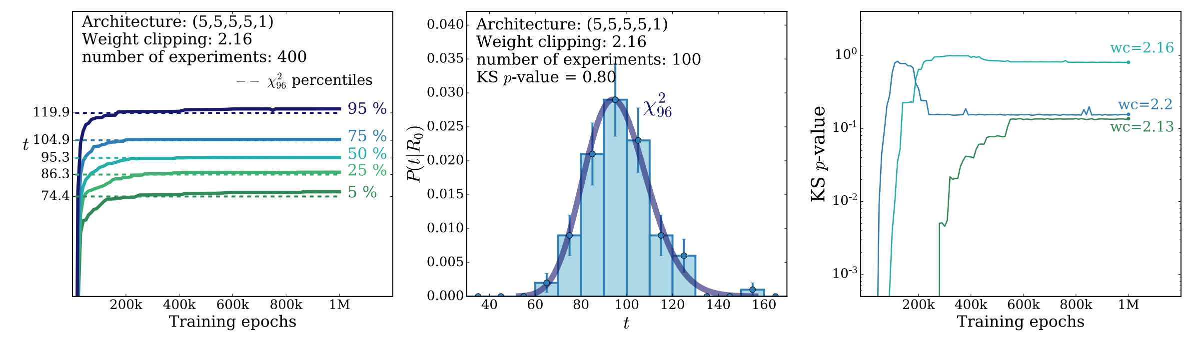


New Physics Learning Machine (NPLM)

Including systematic uncertainties

Final procedure in step:

1. **MODEL SELECTION**:
   weight clipping tuning → target $\chi^2_{|\mathbf{w}|}$;

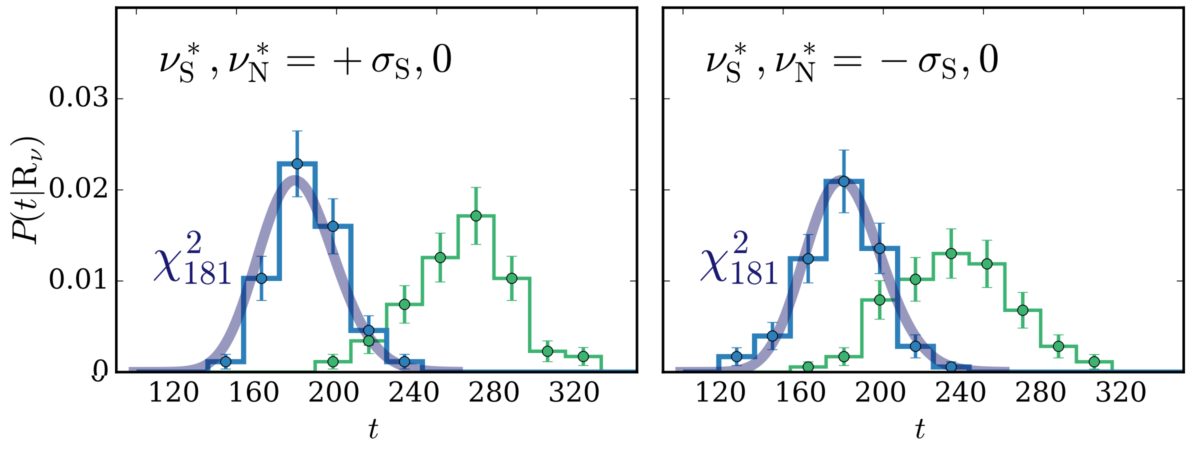


2. **NUISANCE TAYLOR'S EXPANSION LEARNING**:
   modelling $\widehat{r}(x;\nu) = \exp\left[\widehat{\delta}_1(x)\,\nu + \widehat{\delta}_2(x)\,\nu^2 + \ldots\right]$;

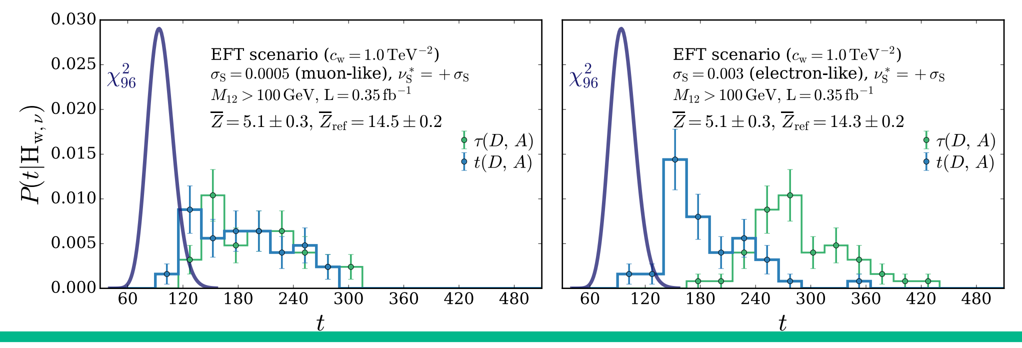
$$\widehat{r}(x;\nu) = \exp\left[\widehat{\delta}_1(x)\,\nu + \widehat{\delta}_2(x)\,\nu^2 + \ldots\right]$$

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.

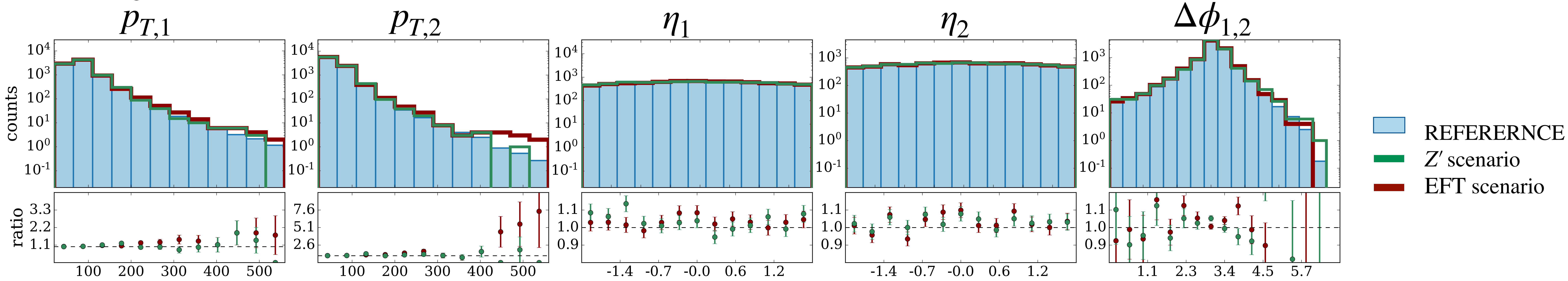


Experiments

Di-body final state at the LHC

Dataset

5D analysis — Input variables:



Reference sample (SM):

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

$$p_{T\,1,2}^{(\mathrm{b,e})} = \exp\left[\nu_{\mathrm{s}}\sigma_{\mathrm{s}}^{(\mathrm{b,e})}/\sigma_{\mathrm{s}}^{(\mathrm{b})}\right] p_{T\,1,2}^{(\mathrm{b,e})}$$

(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: $\sigma_S^{(b)} = \sigma_S^{(e)} = 3\,\%$

Di-body final state at the LHC

Dataset

New Physics benchmarks:

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: $M_{12} > 120\,\text{GeV}, L = 1.1\,\text{fb}^{-1}, \text{N(S)} = 210$

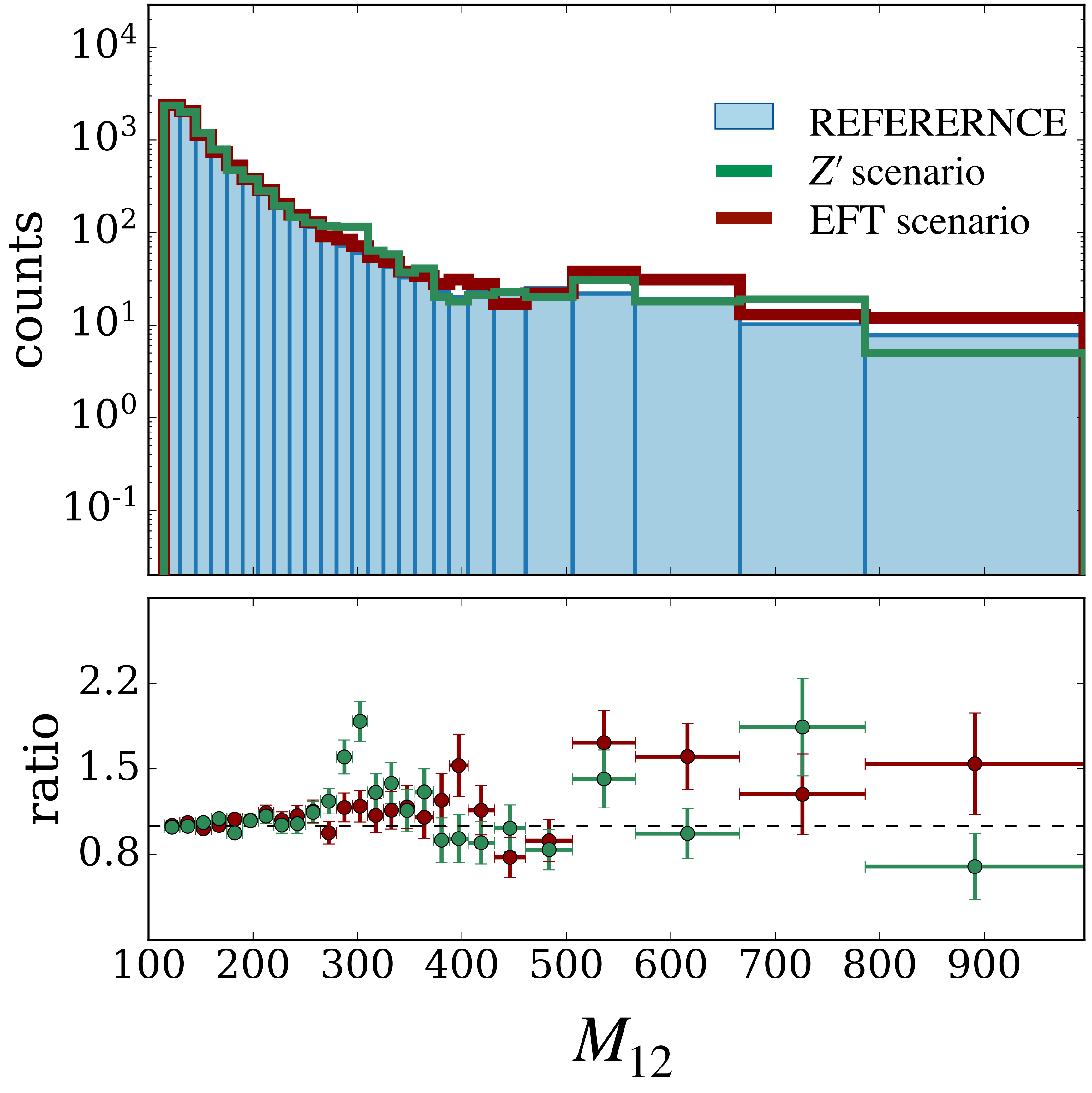
Non resonant excess in the tail of the two-body invariant mass

- **EFT scenario:** dimension-6 4-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: $M_{12} > 100\,\text{GeV}, L = 0.35\,\text{fb}^{-1}, c_W = 0.25\,\text{TeV}^{-2}$

Example:
Tau-like regime



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

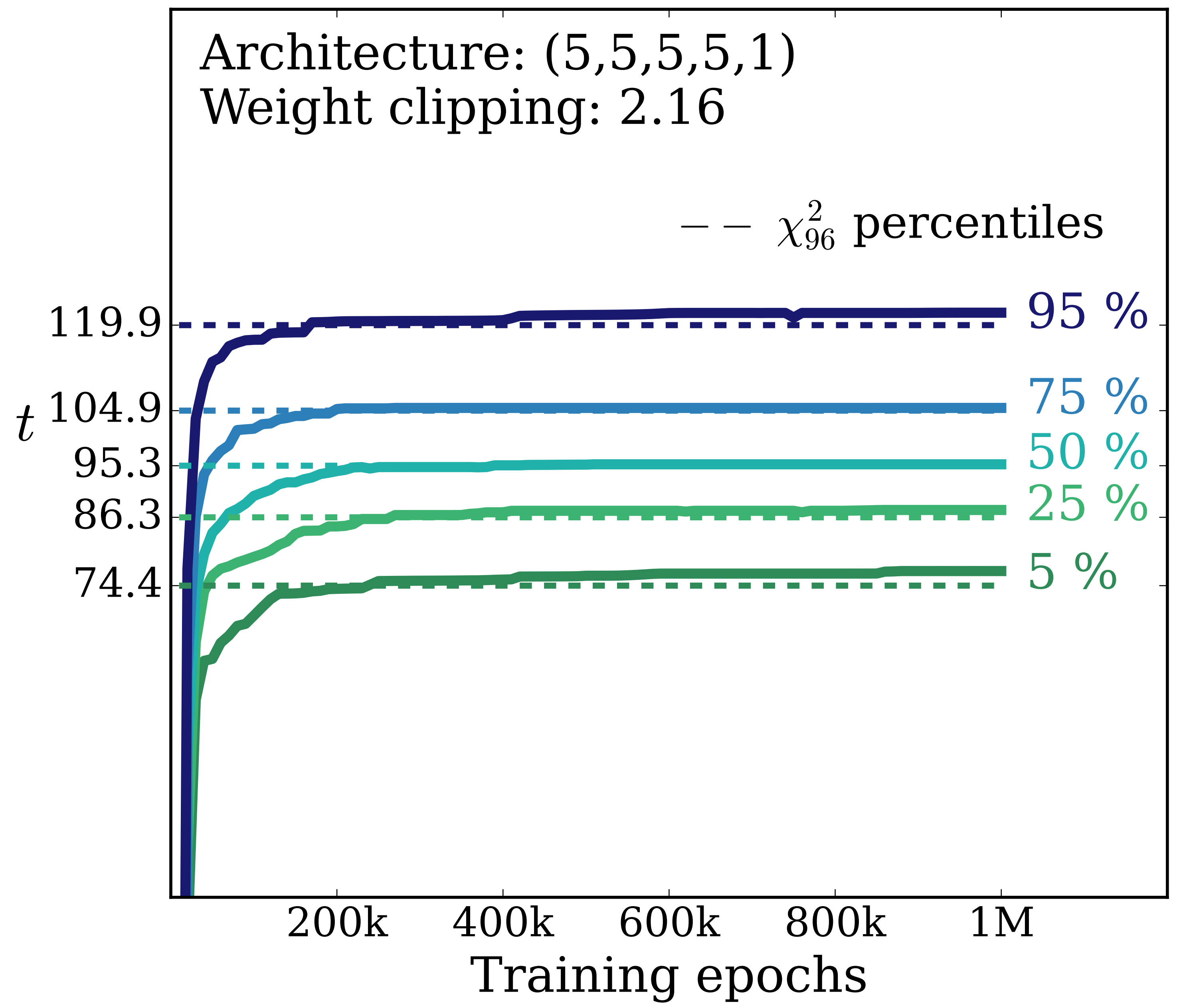
Di-body final state at the LHC

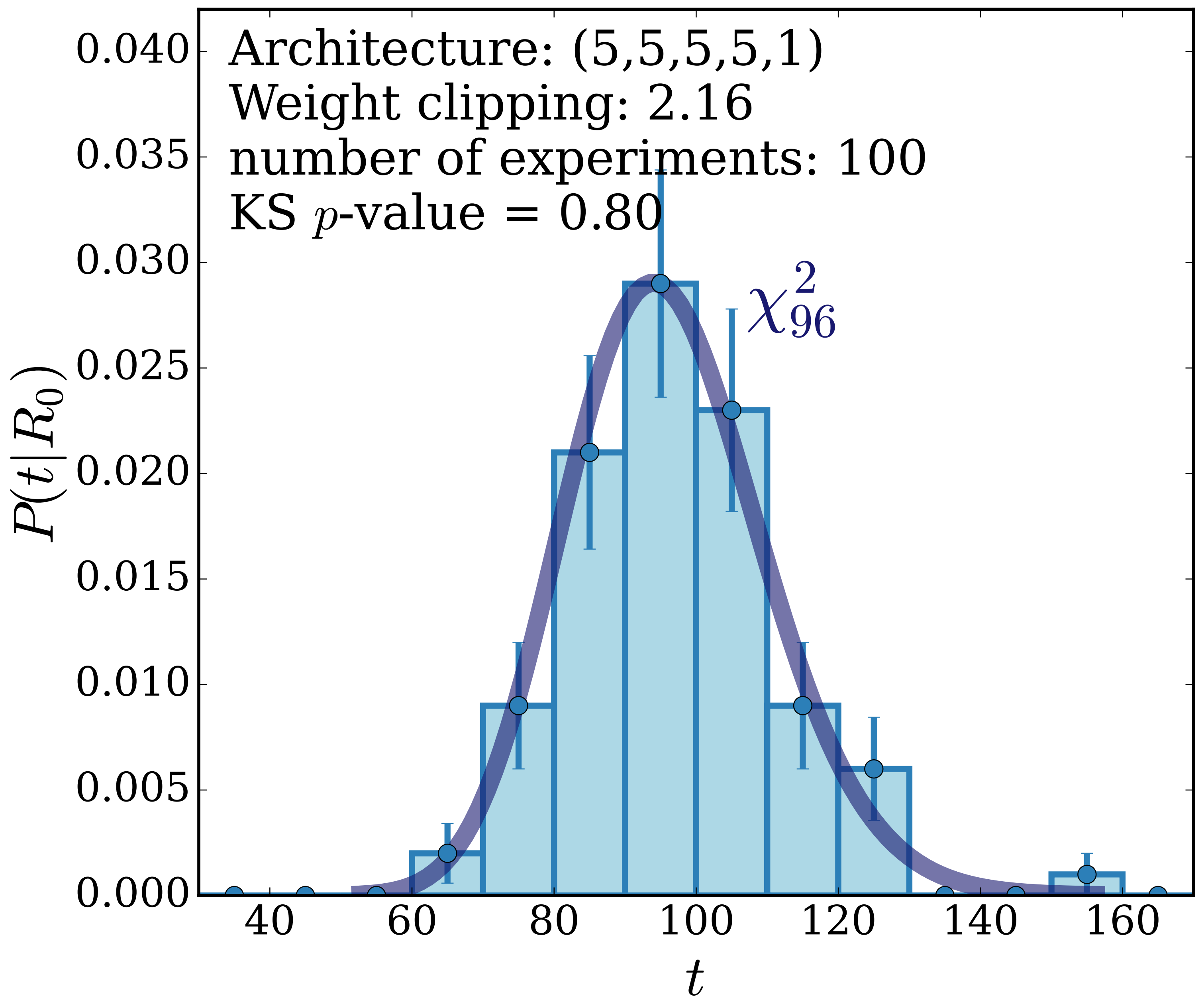
Step 1: Model selection

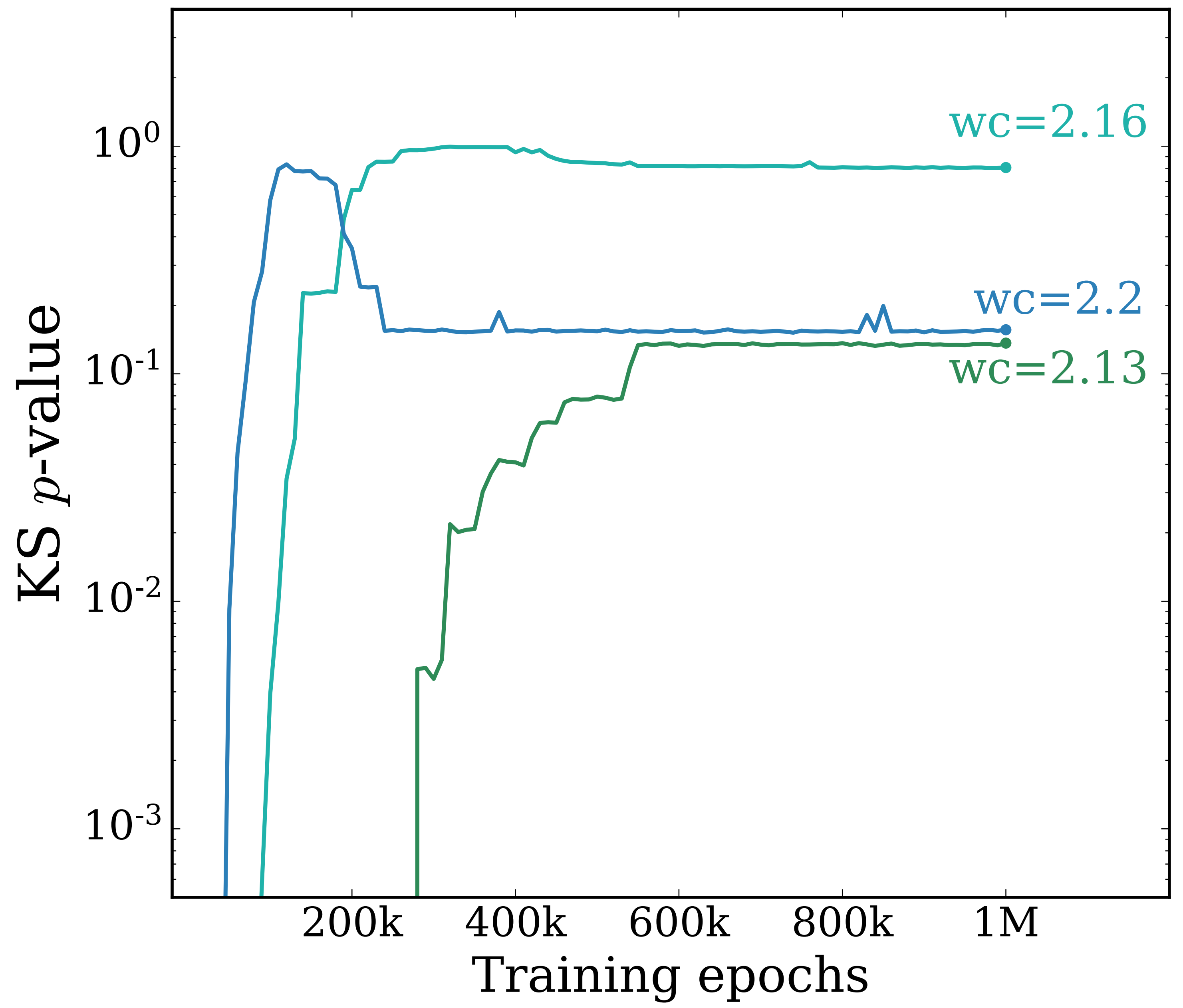
Training setup:

- Data sample N(D) $\sim$ 8500 events
- Reference sample N(R) $= 5 \times$ N(D) $\sim$ 42 500

Weight clipping tuning:







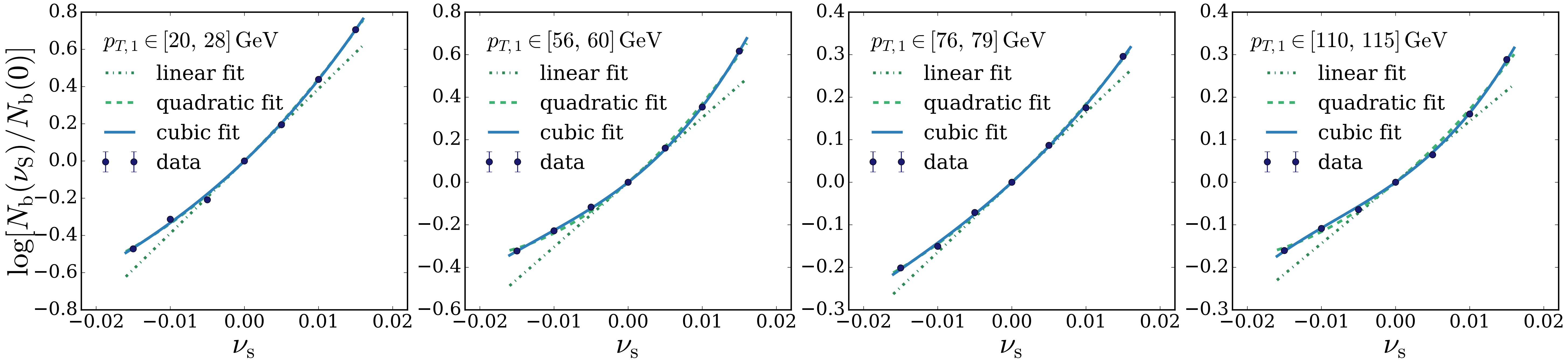
Di-body final state at the LHC

Step 2: Nuisance Taylor's expansion Learning

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

Input samples

ν_1	ν_2	...	ν_n
$S_0(\nu_1) \sim \mathrm{R}_0$	$S_0(\nu_2) \sim \mathrm{R}_0$	...	$S_0(\nu_n) \sim \mathrm{R}_0$
$S_1(\nu_1) \sim \mathrm{R}_{\nu_1}$	$S_1(\nu_2) \sim \mathrm{R}_{\nu_2}$		$S_1(\nu_n) \sim \mathrm{R}_{\nu_n}$

→ $\hat{\delta}_1$, $\hat{\delta}_2$ →

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

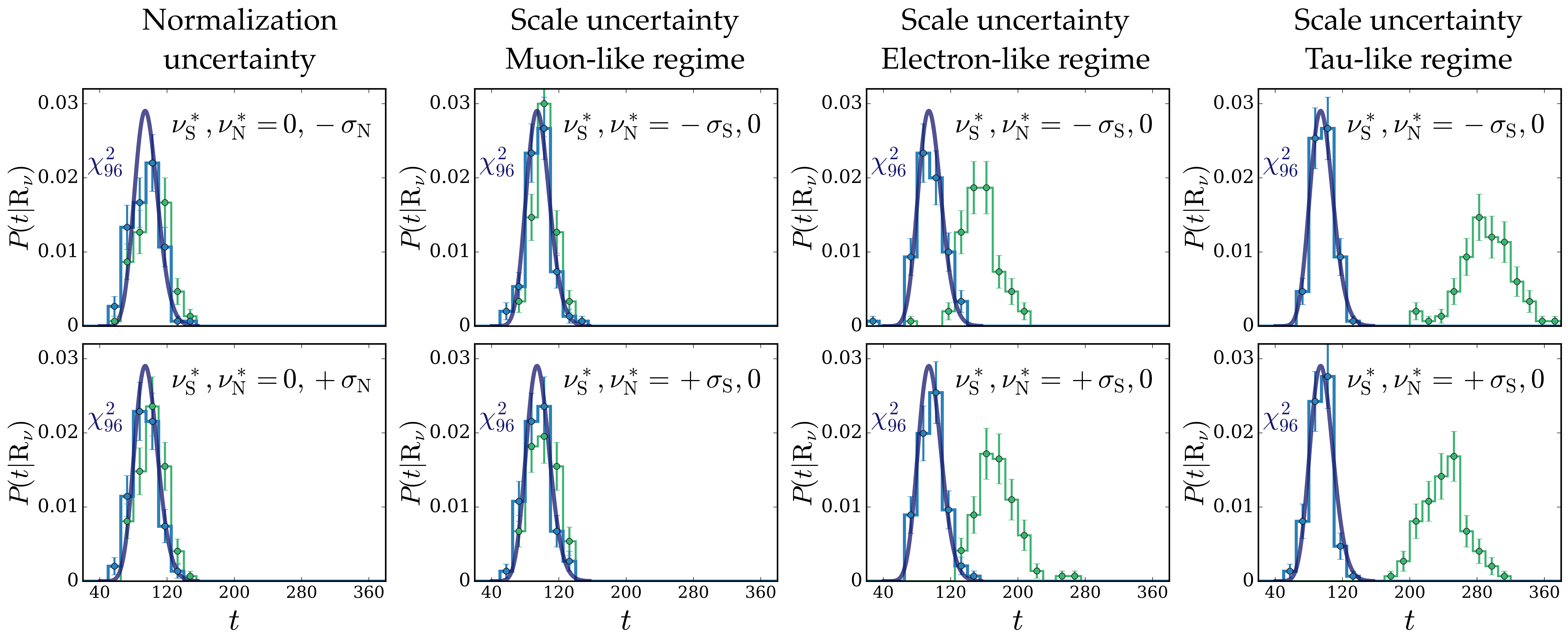
Loss function*

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

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

Di-body final state at the LHC

Step 3: Validation



Di-body final state at the LHC

Step 4: Sensitivity to New Physics scenarios

NPLM vs. **model-dependent approach**

(the signal hypothesis is known a priori and is exploited to optimise the test statistic)

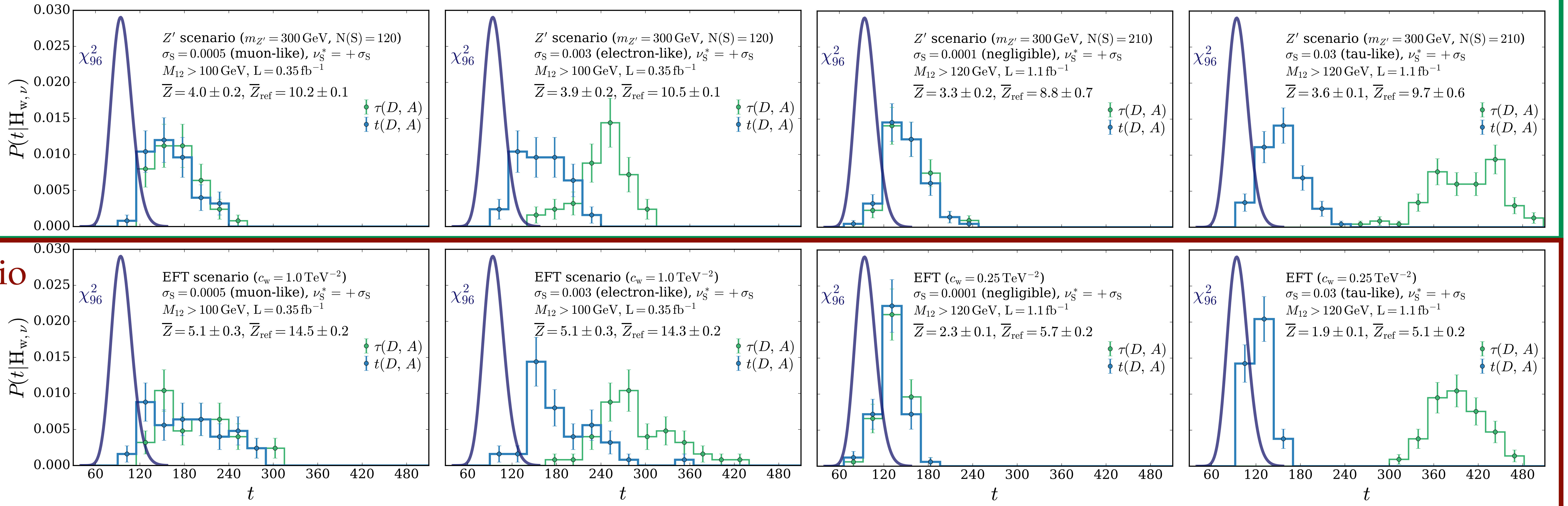
Muon-like regime | Electron-like regime | Negligible systematic uncertainties | Tau-like regime

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
- $\overline{Z}_{ref}$: Z-score from the median of the empirical reference test statistics

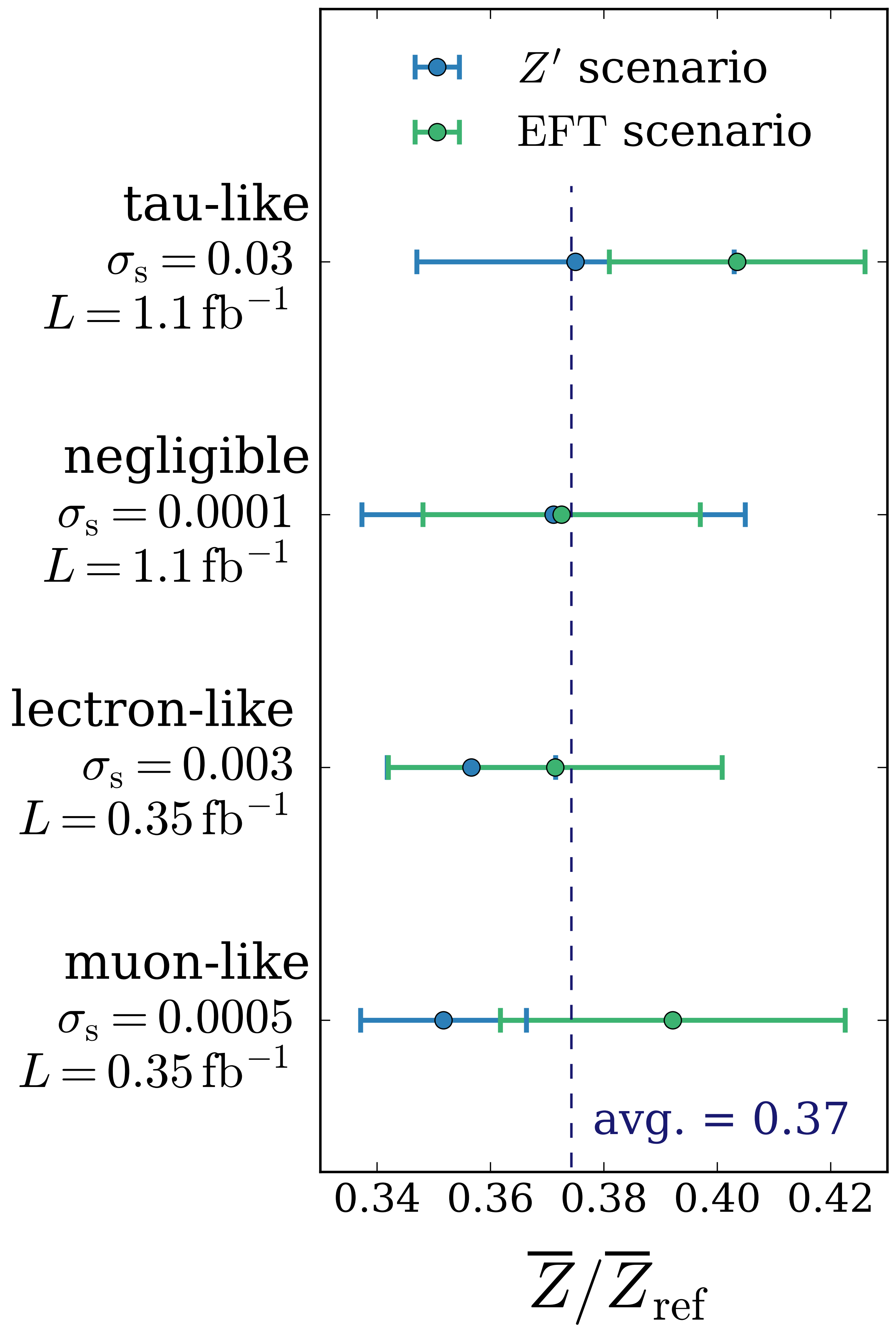
(~ 100 toys experiments)

Di-body final state at the LHC

Step 4: 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 ($\overline{Z}/\overline{Z}_{\text{ref}} = 0.37$);



Di-body final state at the LHC

Step 4: Sensitivity to New Physics scenarios

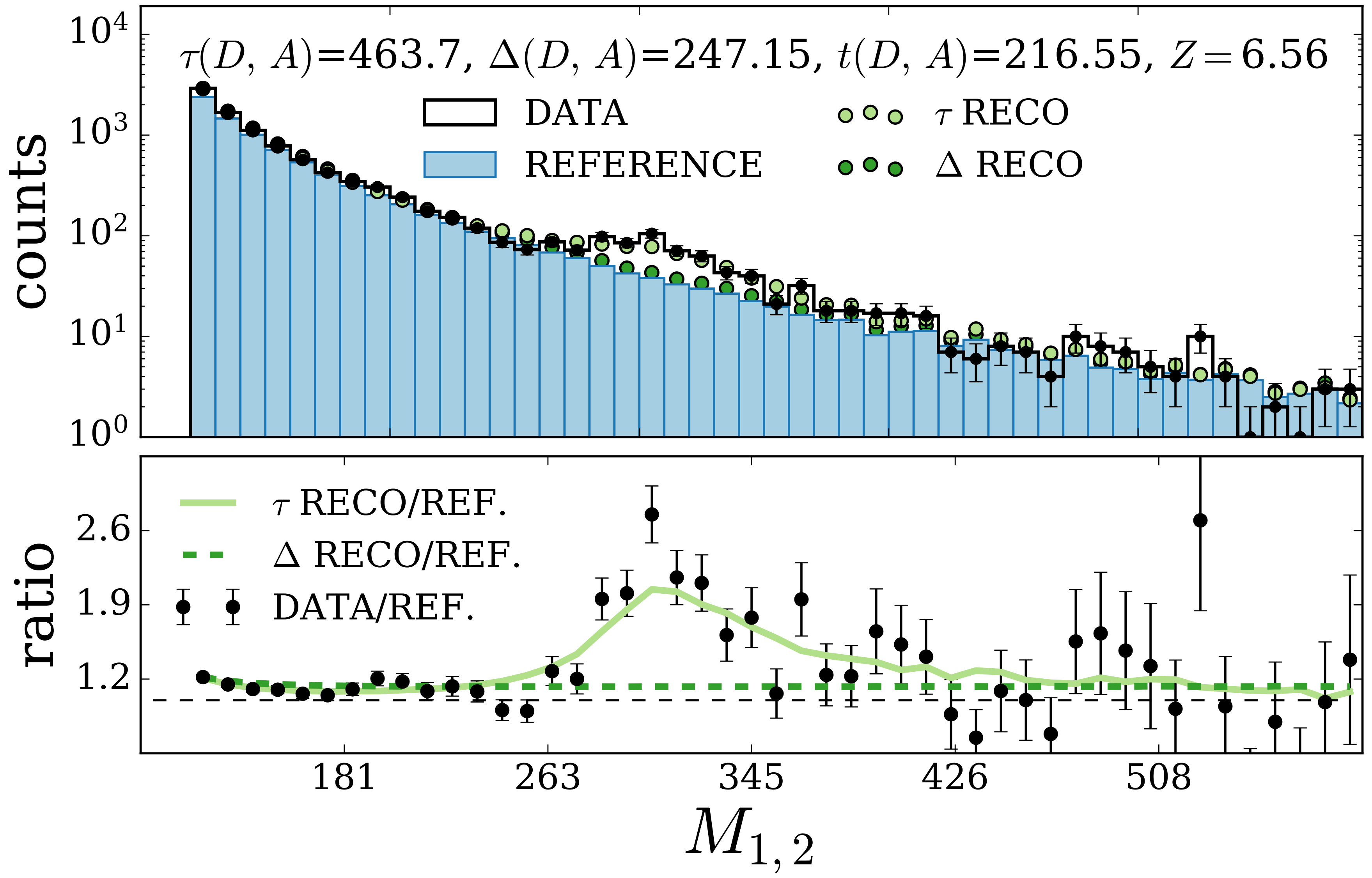
Summary of the results:

- 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 ($\overline{Z}/\overline{Z}_{\rm ref} = 0.37$);
- NPLM is able to *learn* non trivial combinations of the input variables and point to the source of the significant discrepancy.

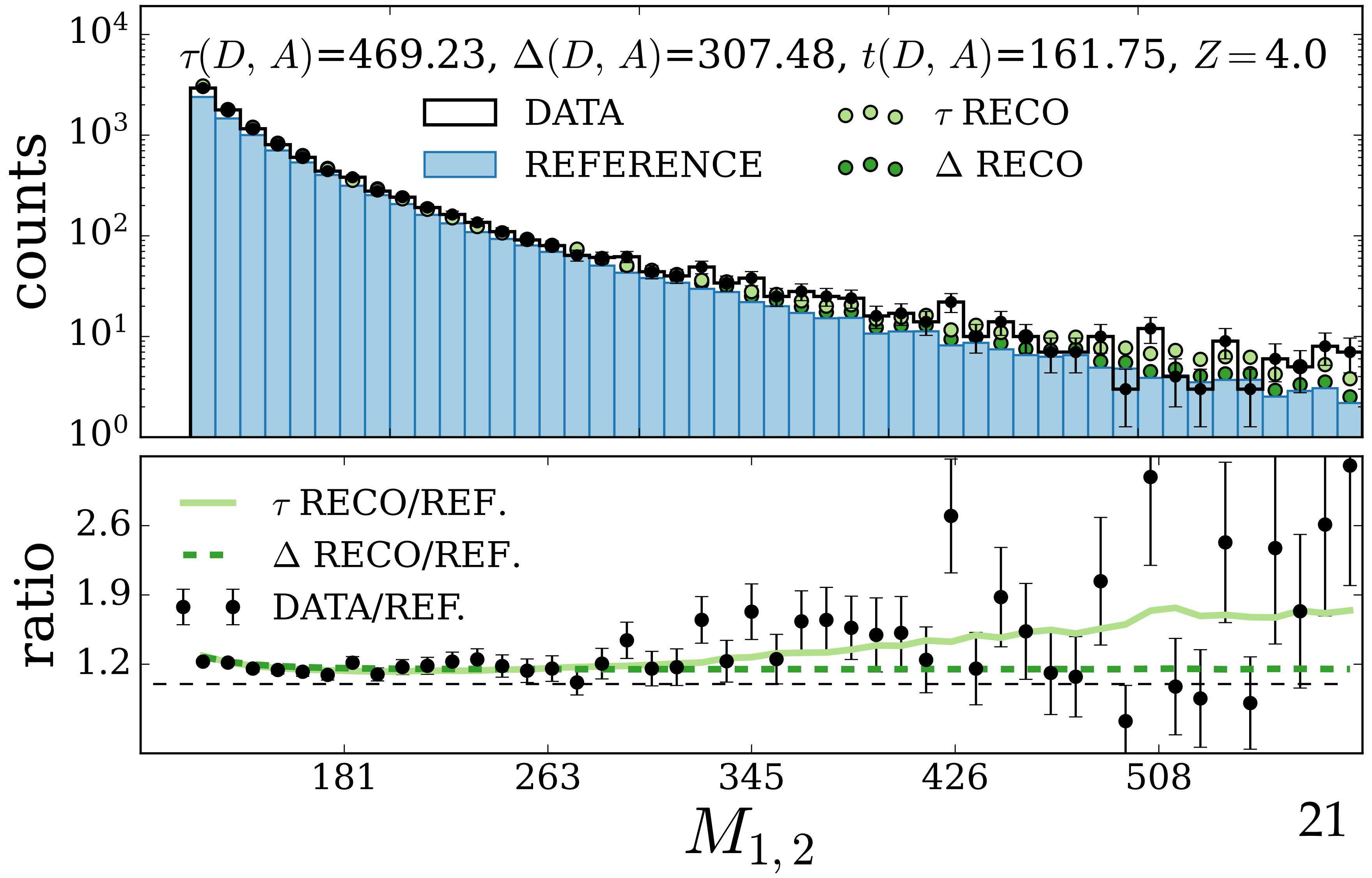
τ reconstruction: $n(x\,|\,\mathrm{H}_{\hat{\mathbf{w}},\hat{\nu}}) = n(x\,|\,\mathrm{R}_0)\,\dfrac{n(x\,|\,\mathrm{R}_{\hat{\nu}})}{n(x\,|\,\mathrm{R}_0)}\,e^{f(x;\,\hat{\mathbf{w}})}$

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

Z' scenario (tau-like regime), m = 300 GeV, N(S) = 210



EFT scenario (tau-like regime), c_W = 0.25 TeV^{-2}



Conclusions

Outlook on future perspectives

Current limitations and future developments:

- Accuracy and size of the Reference sample
- Accuracy in the (multivariate) modelling of the nuisance effects
- Training time*

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

Shared issues within the high energy Physics community that will open the way to interesting directions for future developments.

* possible solution from Kernel Methods! (See interesting poster)

NPLM is ready to be performed on a real analysis at the LHC!

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

More in:

Learning New Physics from an Imperfect Machine

(today on the arXiv [2111.13633])

Backup slides

New Physics Learning Machine (NPLM)

Dealing with multivariate problems (negligible uncertainties)

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

Wilks' theorem:

Θ_0: set of parameters describing H_0

Θ_1: set of parameters describing H_1

If $H_0 \subseteq H_1$, then under the H_0 hypothesis the test statistic

$$t(\mathcal{D}) = 2\,\log\frac{\mathcal{L}(\mathrm{H}_{\mathbf{1}}|\mathcal{D})}{\mathcal{L}(\mathrm{H}_{\mathbf{0}}|\mathcal{D})}$$

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.

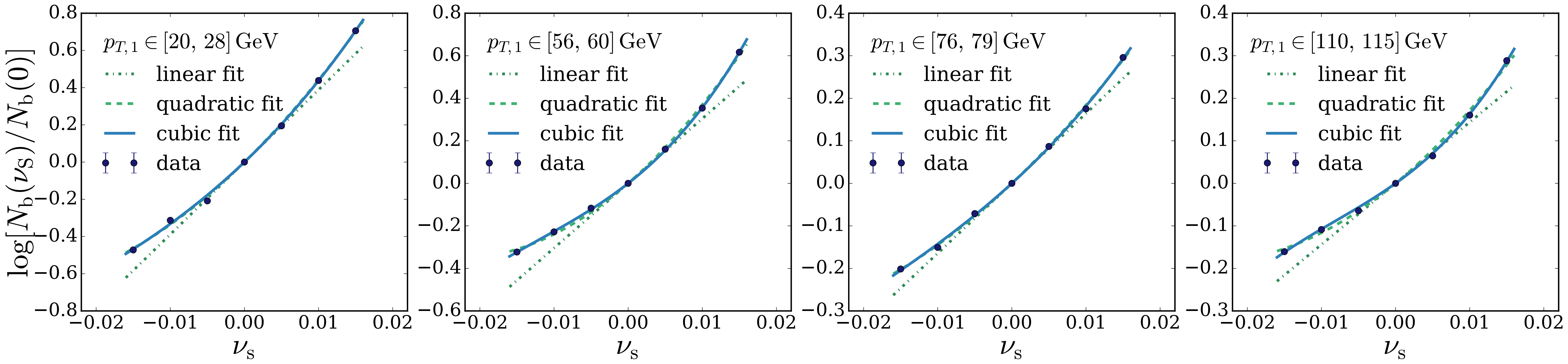
→ a **MODEL SELECTION** procedure can solve this problem!

Di-body final state at the LHC

Step 2: Nuisance Taylor’s expansion Learning

Preliminary study

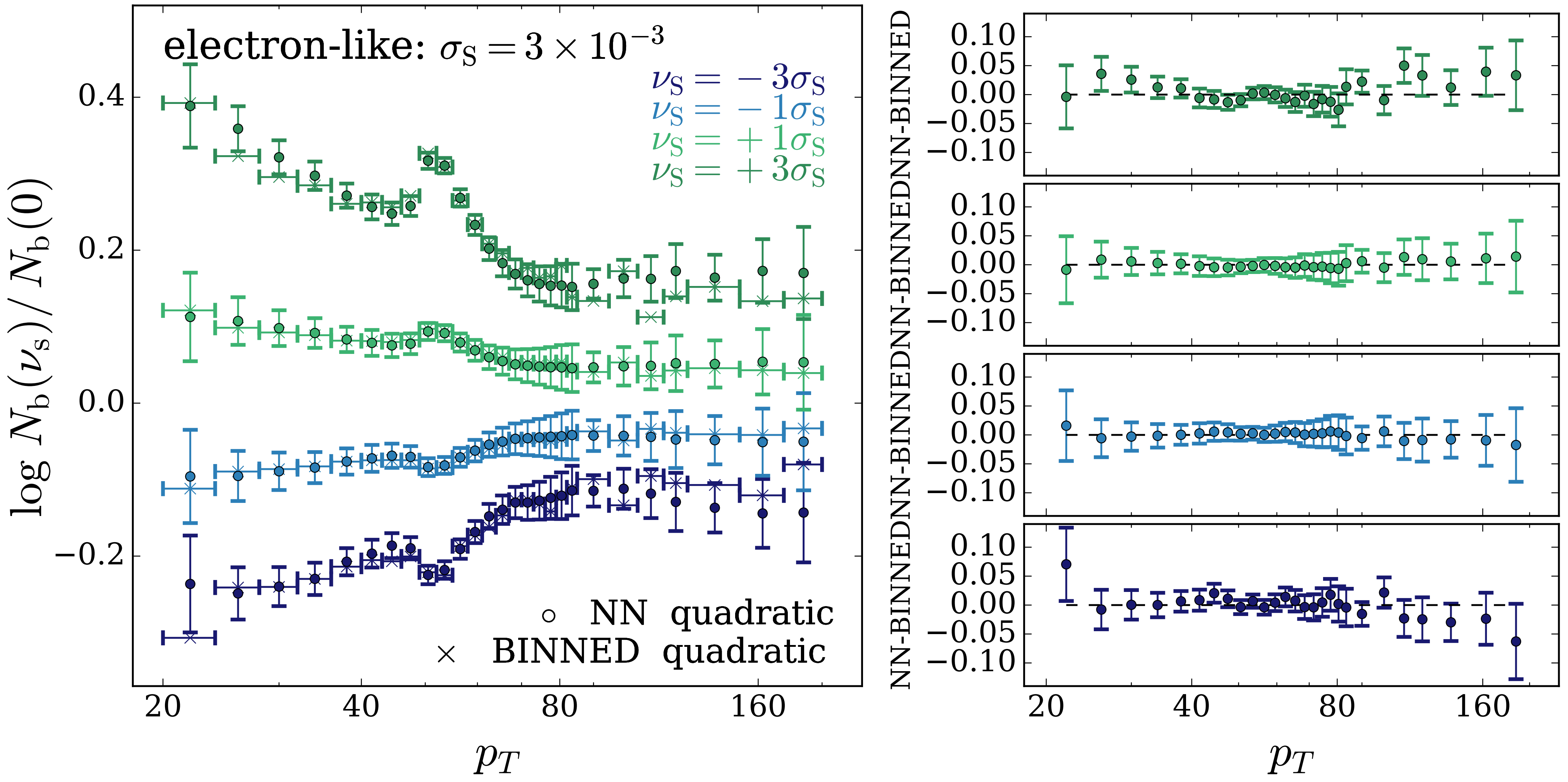
Preliminary binned analysis to determine the proper order for the Taylor’s expansion



Taylor’s expansion learning

Training a neural network model to learn each coefficient of the Taylor’s expansion of $r(x;\nu) = \dfrac{n(x\,|\,\mathrm{R}_\nu)}{n(x\,|\,\mathrm{R}_0)}$

Results:



Di-body final state at the LHC

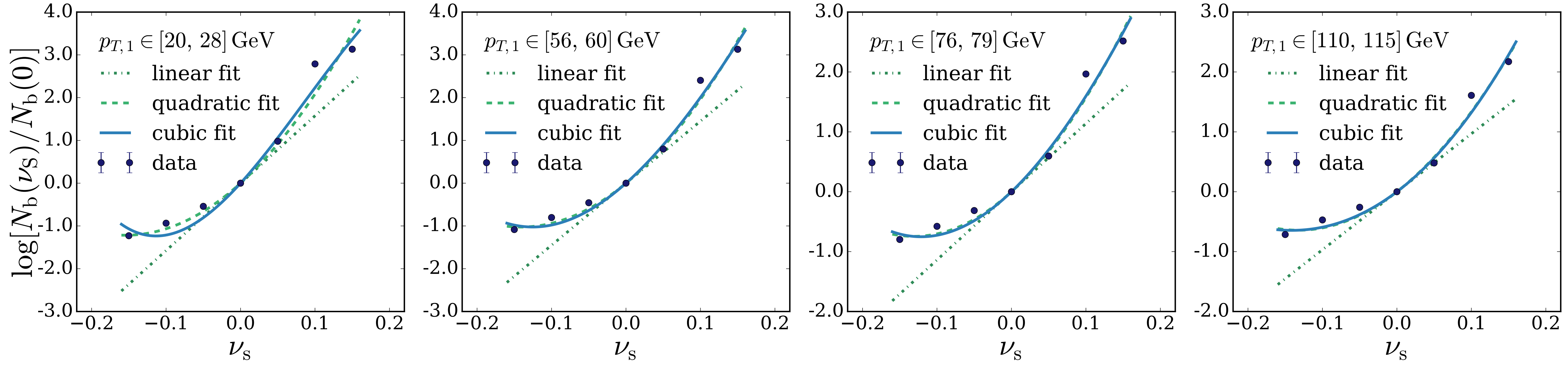
Step 2: Nuisance Taylor's expansion Learning: tau-like regime

Preliminary study

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

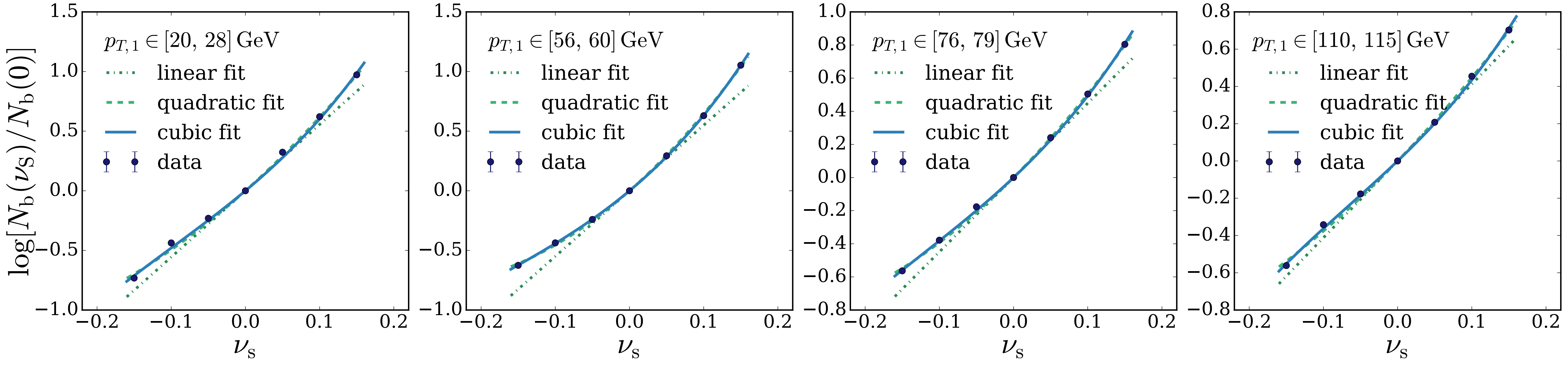
Tau-like regime

$M_{1,2} > 100\,\mathrm{GeV}$



Tau-like regime

$M_{1,2} > 120\,\mathrm{GeV}$

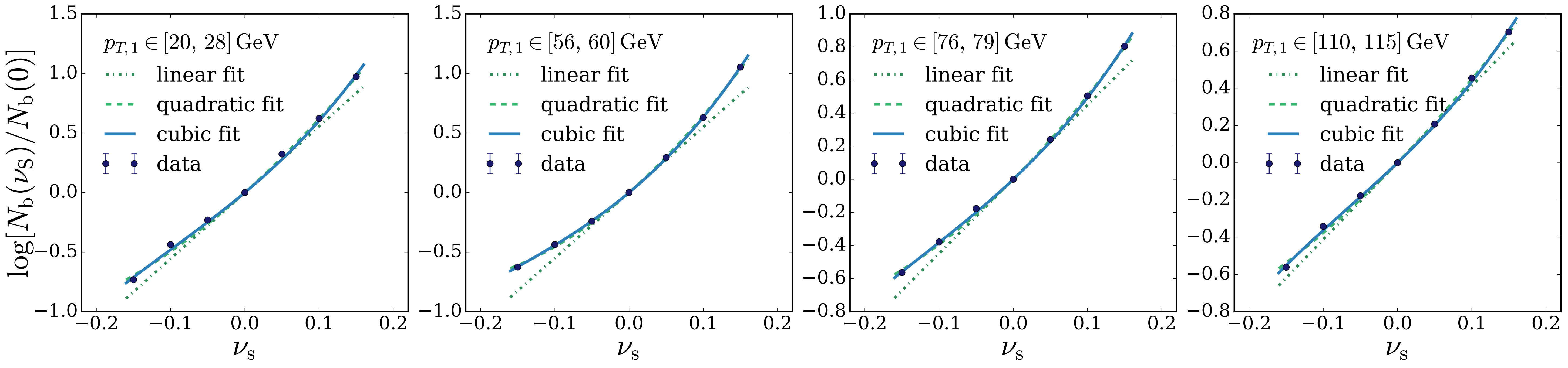


Di-body final state at the LHC

Step 2: Nuisance Taylor's expansion Learning

Preliminary study

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



Taylor's expansion learning

Training a neural network model to learn each coefficient of the Taylor's expansion of $r(x;\nu) = \dfrac{n(x\,|\,\mathrm{R}_\nu)}{n(x\,|\,\mathrm{R}_0)}$

Results: