Machine Learning approaches for measurements and calibrations in the context of the W mass analysis at the LHC

Olmo Cerri (Caltech), **Nicolò Foppiani (Harvard University)**, Luigi Rolandi (SNS)

Supervised by: Josh Bendavid, Patrick Janot, Maurizio Pierini, ... (CERN)

Once upon a time...

Outline

- Motivation for a precise W mass measurement
- The W mass measurement at the LHC
- Machine Learning approaches for the measurement of the hadronic recoil

Machine learning approaches:

- New experimental definition of the recoil based on a semi-parametric regression
- Machine learning approach for the calibration of the recoil

It is a thesis work developed by Nicolò and Olmo, and can be found at:
https://cds.cern.ch/record/2281312
https://cds.cern.ch/record/2285935

Motivation for a precise measurement of the mass of the W boson

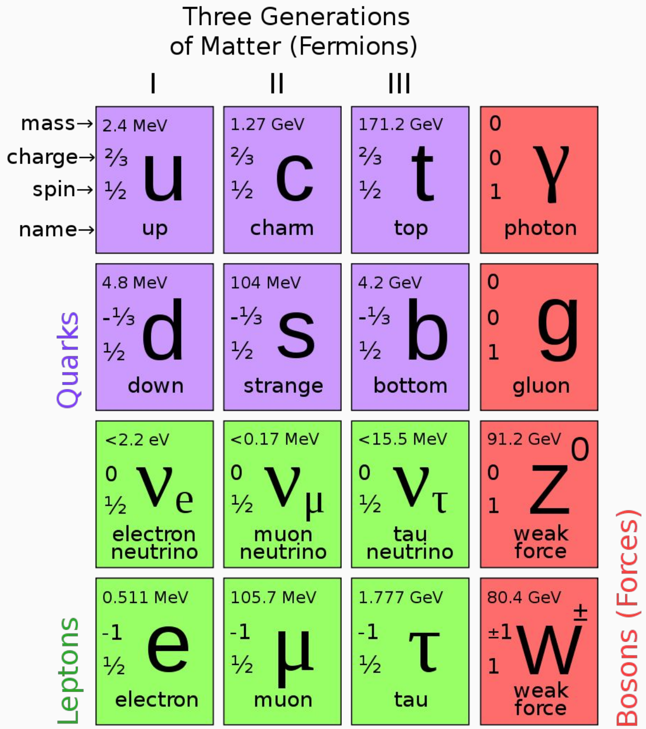
The weak interaction in the standard model

The weak interaction:

- Responsible for the β-decay
- Fermi theory at low energy
- Electroweak theory at high energy
  - Carried by the γ, Z, and W bosons

The W boson:

- Electromagnetic charged
- mass ~ 80 GeV
- Interacts with
  - Quarks q-q’
  - lepton-neutrino



M_W values: theory vs experiments

The standard model is predictive given:

- 3 quantities at tree level
- 6 quantities at one loop

Collective fit of all measurements:

➢ Test of the theory

➢ Inconsistency may unveil new physics through quantum effects

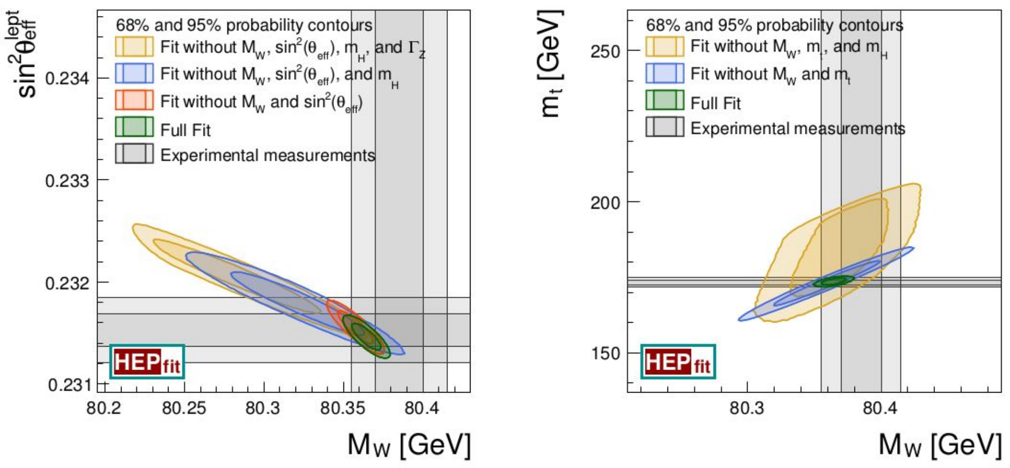
M_W is an interesting case

Average of exp measurements (without ATLAS):

➔ M_W^{exp} = 80.385 ± **0.015** GeV

Theory prediction:

➔ M_W^{theo} = 80.362 ± **0.008** GeV



Strong motivation to improve the experimental precision!!!

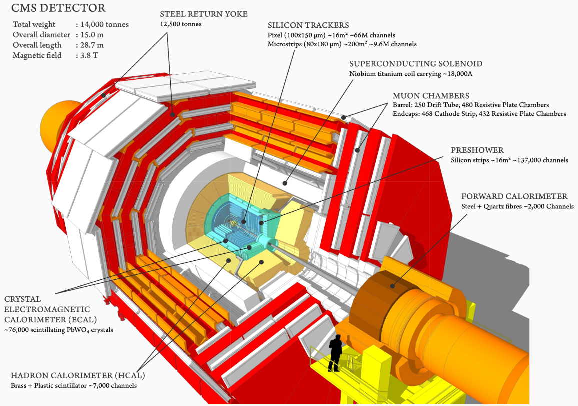
LHC collected a lot of W bosons already!

Measuring M_W with precision ~ 10^{-4} needs

- Huge sample of W decays
- Control systematic uncertainties

The Large Hadron Collider (LHC):

- ➢ Proton proton collisions @13 TeV
- ➢ W bosons copiously produced
- ➢ The CMS experiments collected 400M W boson decays during 2016 only



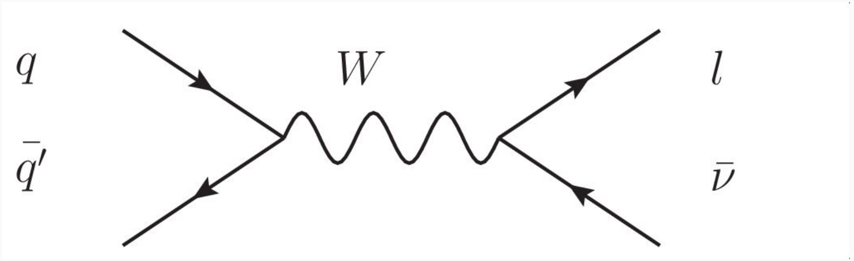
The challenge is now given by the systematic uncertainties!!!

The W mass measurement at the LHC

W boson production at the LHC

Main process:

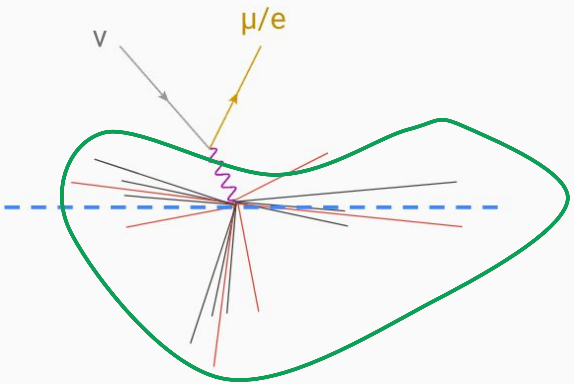
➢ qq’ ⟶ W



Main decay channel

➢ W ⟶ $\ell\nu$

➢ Small background



Other particles are produced in the collision:

➔ Collectively named **recoil**

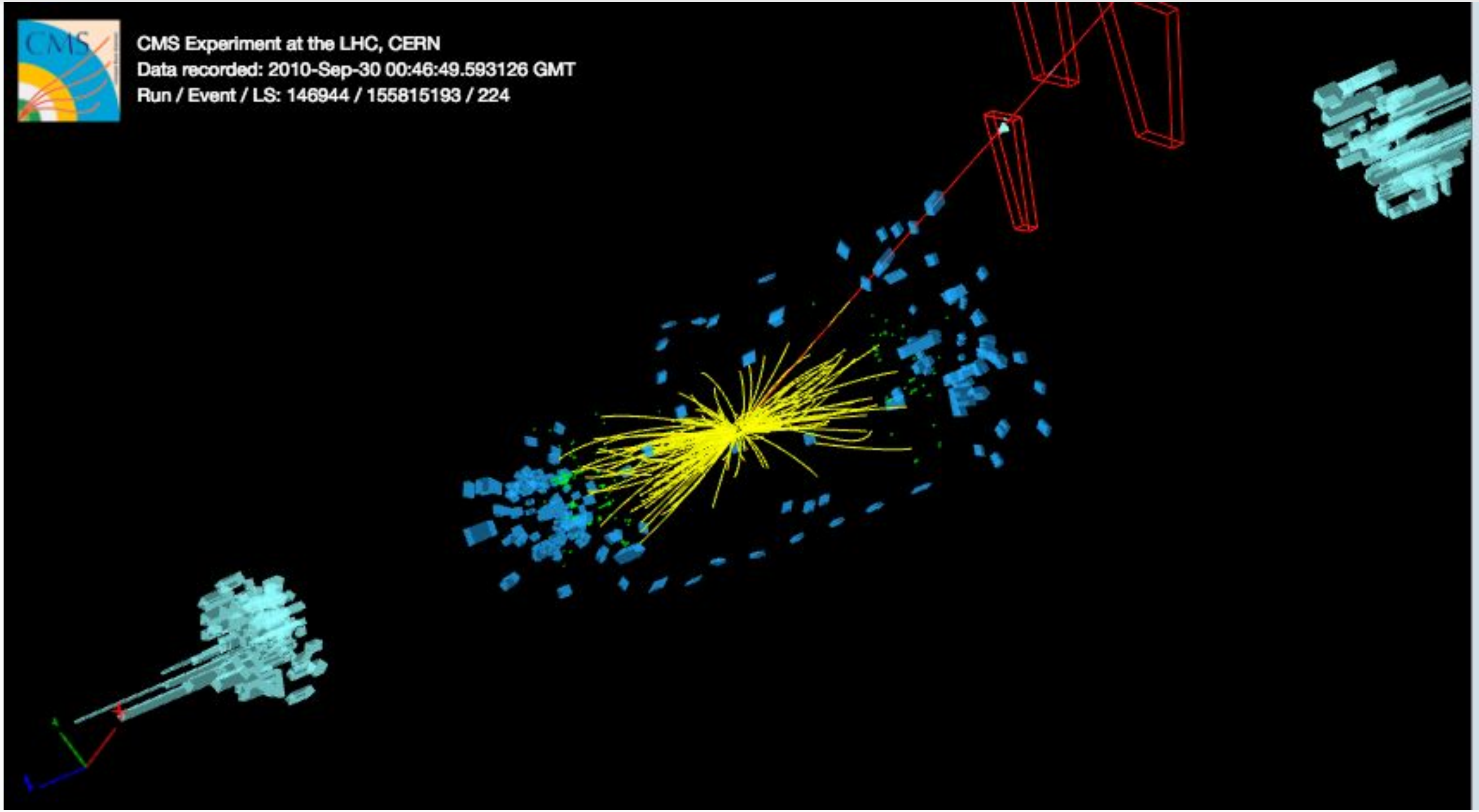
Processes responsible for that:

➔ Hadronization of ISR

➔ Underlying event

➔ NLO terms

How an event looks



Extract the mass: template fit

The invariant mass cannot be reconstructed

➔ Hard to measure the neutrino momentum on the beam axis

Need for a template fit

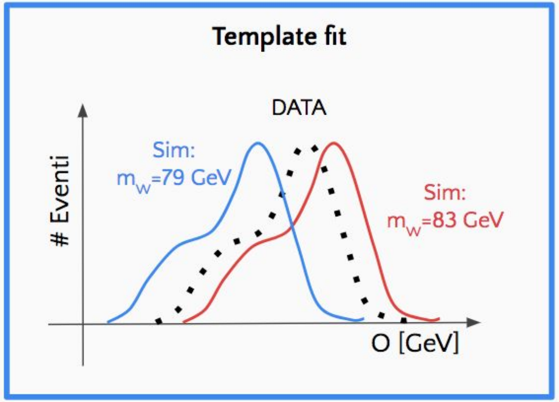
➔ Consider an experimental variable correlated with the mass of the W boson

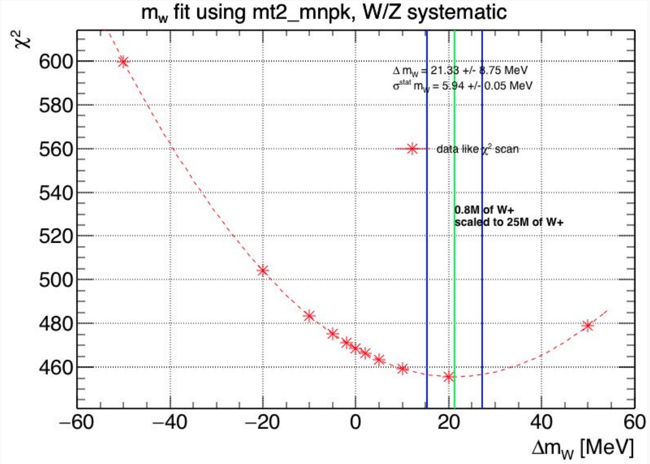
➔ Model the expected distribution for different values of m_W

➔ Maximum likelihood fit point by point

Uncertainties in the W production model

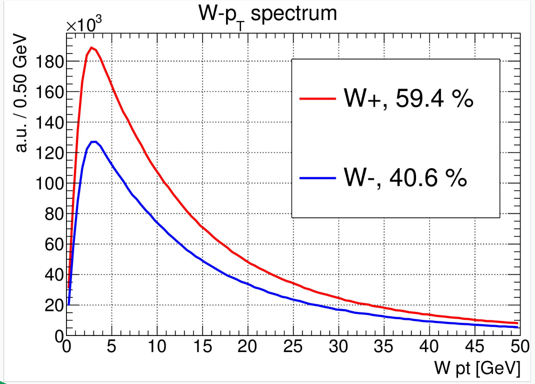
➔ Systematic uncertainties

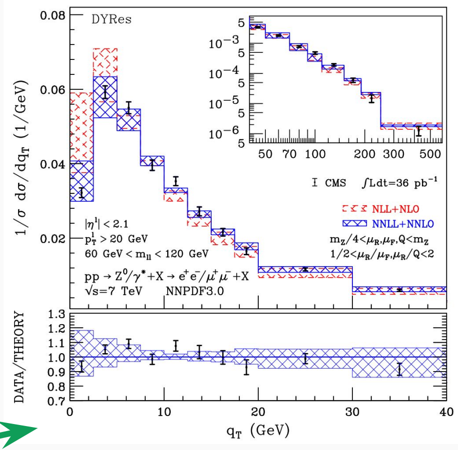




p_T-W is one of the biggest uncertainties

- ➔ p_T-W is hard to predict:
  - ◆ Typical momentum of 5 GeV, non perturbative QCD
  - ◆ Theory uncertainties ~ 10%
- ➔ One of the dominant uncertainties in the analysis





	CDF/p_T	CDF/M_T	ATLAS/p_T	ATLAS/M_T
Statistical	16	15	7.2	9.6
Lepton Scale and Resolution	7	7	6.5	6.5
Recoil Scale and Resolution	5.5	6	2.5	13
Backgrounds	4	3.5	4.6	8.3
PDFs	9	10	9	10.2
W transverse momentum model / QCD	9	3	8.3	9.6
Photon Radiation/EWK	4	4	5.7	3.4

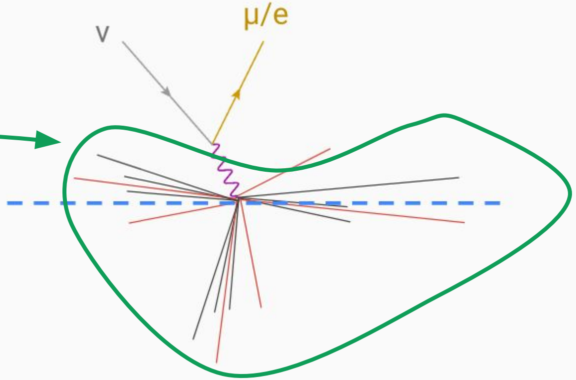
How to defeat the uncertainty on p_T-W?

- ➔ Ask to theorists: improve the calculations and simulations
  - ◆ People are working, but it's hard
- ➔ Use experimental variables as less dependent as possible on p_T-W
  - ◆ Let's try to build it!

From the kinematics of the final state:

Recoil-momentum = W momentum

- Impossible to measure the longitudinal part
- Momentum conservation on the transverse plane still helpful
- Perfect measurement p_T recoil = p_T-W



The transverse mass

- The invariant mass is independent of p_T-W
  - But I cannot measure the longitudinal momentum of the neutrino
- The transverse mass

$$M_T^2 = 2(p_\mu|\vec{p_\mu} + \vec{h}| + p_\mu^2 + \vec{p_\mu} \cdot \vec{h})$$

  - Invariant mass on the transverse plane
  - Less sensitive to p_T-W, but it depends on the resolution of the recoil measurement

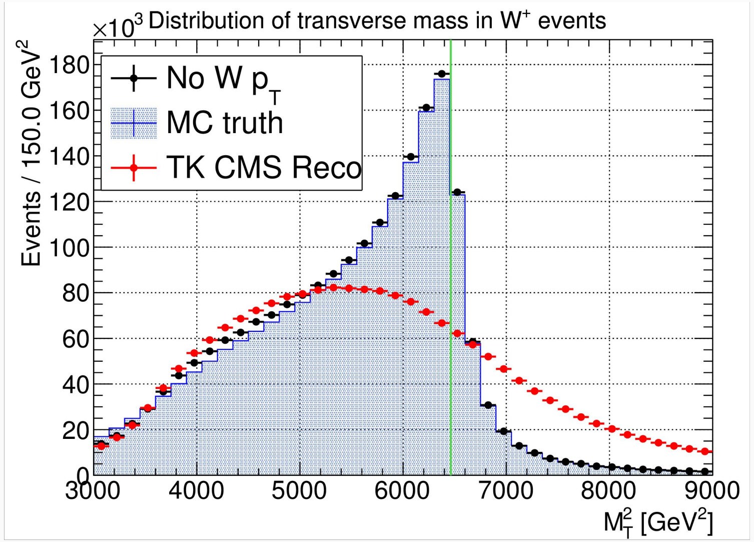
Better measurement of the hadronic recoil

=

The transverse mass is less dependent on p_T-W

=

Smaller systematic uncertainty



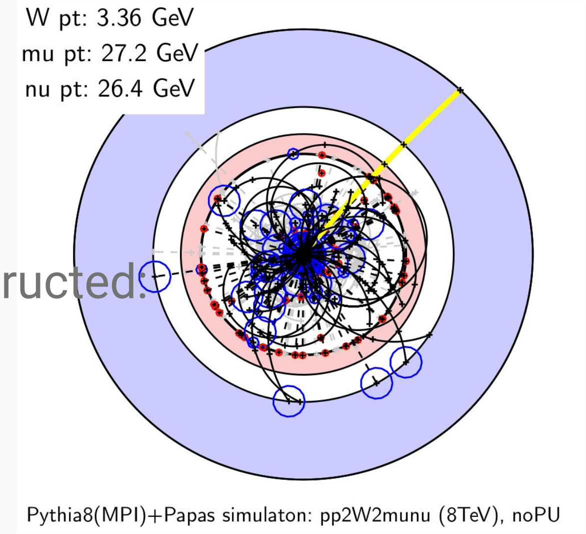
A new recoil definition

A semi parametric regression implemented with deep neural networks

Experimental definition of the recoil

- ➔ Experimentally it is a set of particles
- ➔ For what concerns the transverse mass we need p_T and Φ
- ➔ The recoil definition is the abstract function that maps the experimentally measured particles to p_T and Φ

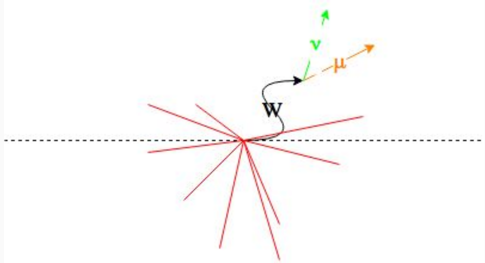
- Yellow track: muon
- All the other particles experimentally reconstructed: recoil



Definitions under study

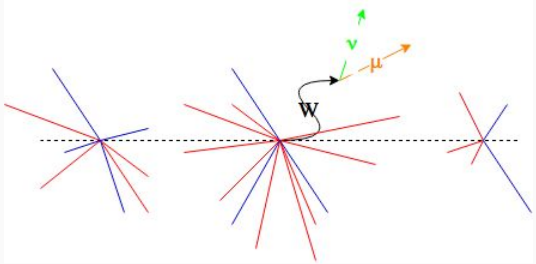
Track recoil (TK):

- Association of the charged tracks to the primary vertex
- Vectorial sum of the momenta of these tracks
- Needs to be calibrated to the proper scale



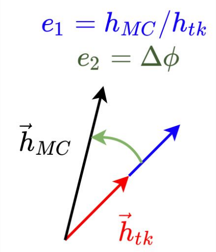
Particle Flow recoil (PF):

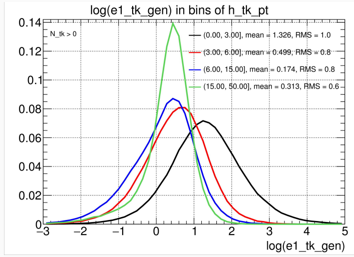
- Vectorial sum of the momenta of all reconstructed particles (but the lepton from the W decay)
- For pileup vertices the sum is zero on average, for the primary vertex is = p_T-W
- Needs to be calibrated to the proper scale



The new idea (MNPK):

- Define some experimental features describing the recoil: eg. number of charged particles, leading particle momentum, etc.
- Combine these features to obtain a better measurement





Let's be formal

x: Experimental observables of the recoil (n features)

y: True value of the recoil (magnitude and angle, 2D)

Experimental definition: $\hat{y} = f(x)$

Eg: track recoil

- $\mathbf{x} = \{ h_{TK}^{x}, h_{TK}^{y} \}$
- $\mathbf{y} = \{ h_{true}^{x}, h_{true}^{y} \}$
- $\hat{\mathbf{y}} = I(\mathbf{x}) = \mathbf{x}$

Vectorial sum of all particles from the primary vertex

$$\vec{h}_{TK} = \sum_{\text{PV tracks}} \vec{p}_i$$

What if I have a longer series of x?

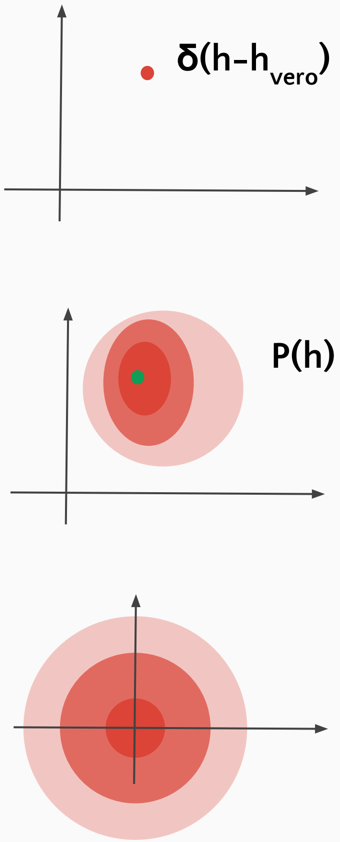
p(y|x) is the largest information I can get

The correlation between **x** and **y** defines a PDF on an event by event basis

Complete correlation
E.g.: $\mathbf{h} \in \mathbf{x}$

Finite correlation

Zero correlation



Perfect measurement of the recoil

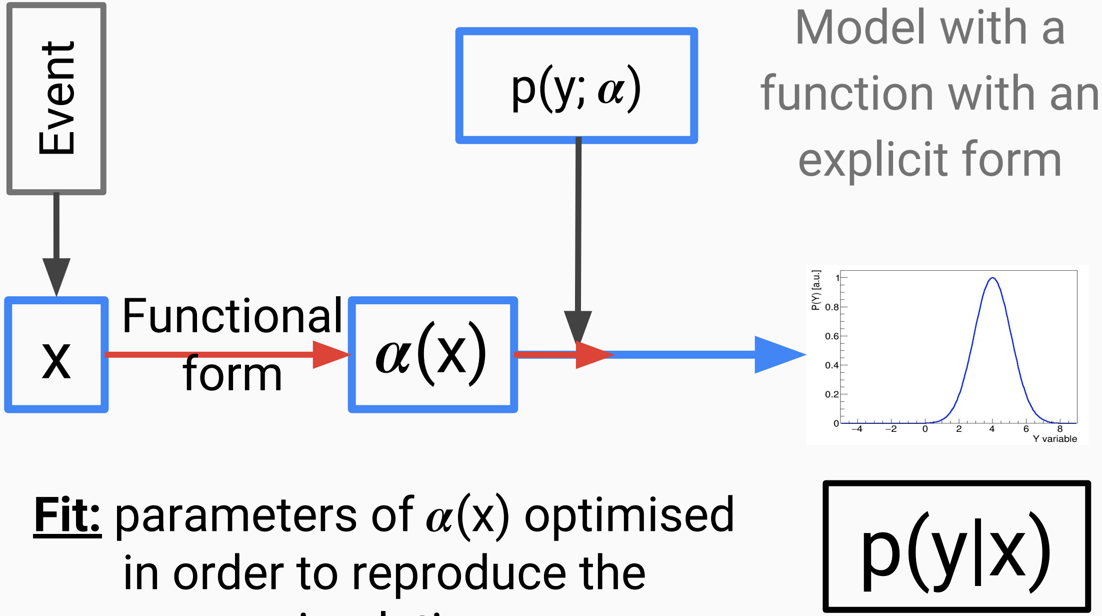
I get a PDF point estimate

No information added

How to estimate p(y|x)? (for each value of x)

From the simulation, using a **semi-parametric regression** (it's machine learning!)

- Model the distribution p(y|x) = p(y|x ; α)
- Assume a functional form for α(x)
- Fit α(x) with the simulation: for each event **x** and **y** are know
- In the data **x** is measured **predict the pdf of y**

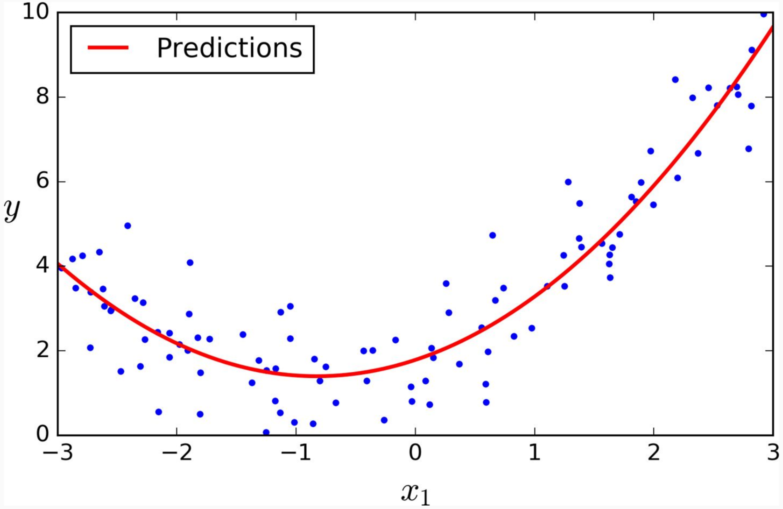


Model with a function with an explicit form

Fit: parameters of α(x) optimised in order to reproduce the simulation

Eg: the least squares fit

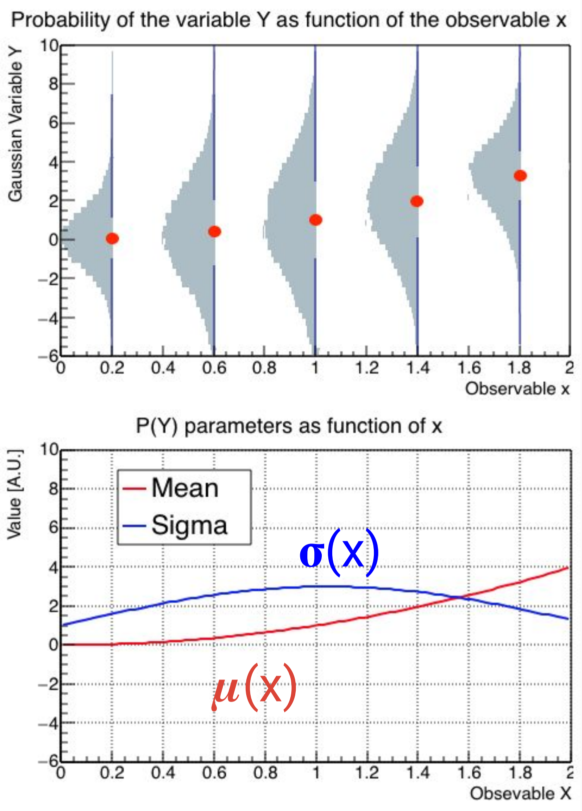
- Training set: $\{x_i, y_i\}_{i=1,..,N}$
- Test set: $\{x_i\}_{i=1,..,M}$
- Assume $p(y|x) \sim N(\mu(x), \sigma^2=1)$
- Model $\mu(x) = \mu(x; \mathbf{a})$
  - Eg: polynomial with coefficient **a**
- Compute maximum likelihood
  - Or the negative-log-likelihood (sometimes called χ^2)
  - Find the values of **a** that minimises it
- Use $\mu(x; a)$ to predict $p(y|x_i)$ in the test set



Picture credit: Hands-On Machine Learning with Scikit-Learn and TensorFlow by Aurélien Géron

Eg: p(y|x) gaussian with non constant σ

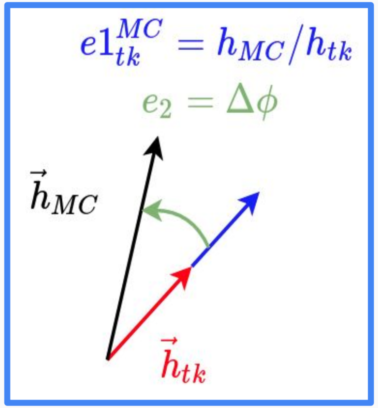
- Training set: $\{x_i, y_i\}_{i=1,..,N}$
- Test set: $\{x_i\}_{i=1,..,M}$
- Assume $p(y|x) \sim N(\mu(x), \sigma^2(x))$
- Model $\mu(x) = \mu(x; \mathbf{a})$, $\sigma^2(x) = \sigma^2(x; \mathbf{b})$
  - Eg: polynomial with coefficient **a, b**
- Compute maximum likelihood
  - Find the values of **a** and **b** that minimises it
- Use $\mu(x; \mathbf{a})$ and $\sigma^2(x; \mathbf{b})$ to predict $p(y|x_i)$ in the test set



Our case

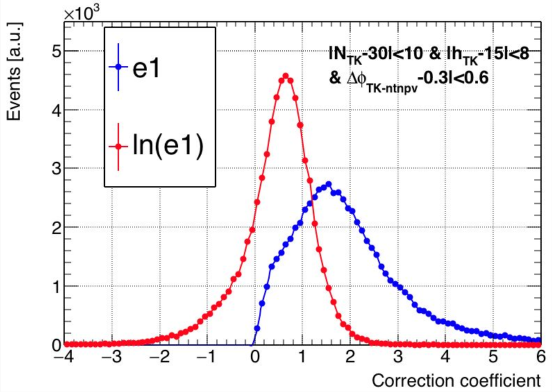
Need for an analytic function that models p(**y**|**x**)

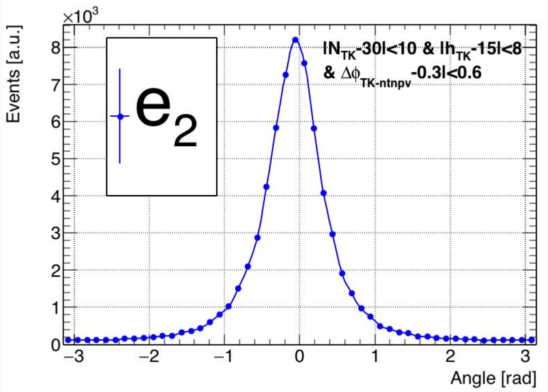
- Easier using correction coefficients
- Easier with 2 one-dim functions
  - One for each component
  - Fewer parameters, easier to converge
  - neglect correlation
- Effectively correcting TK recoil



$$P\left(\vec{h} \mid x\right) \sim f_1(\ln e_1 \mid x) \times f_2(e_2 \mid x)$$

Distributions in a bin of finite width

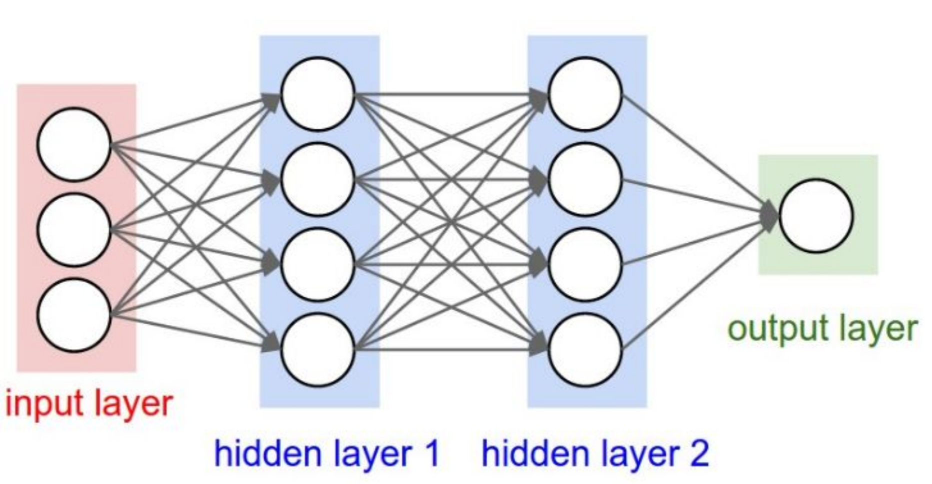


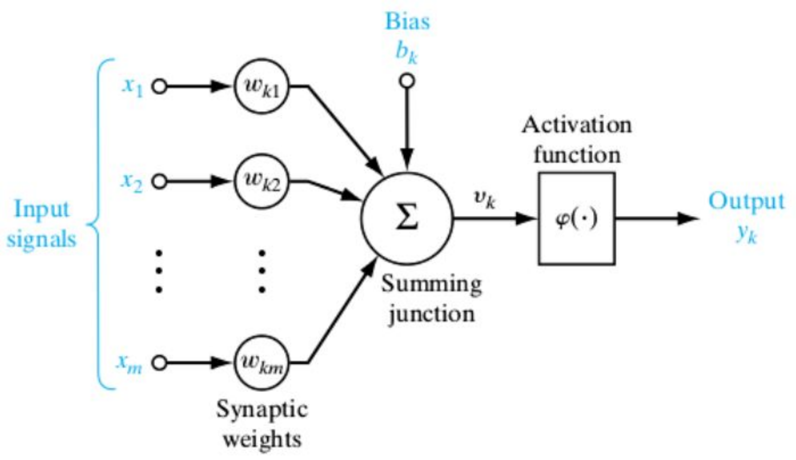


Need for a very versatile function: deep neural networks (DNN)

Arbitrarily non-linear functions:

- Built from small units called neurons, organised in several layers
- It can be proved that, as long as enough neurons and layers are provided, a DNN can approximate whatever non-linear function
- Easy to compute derivatives (chain rule)
- Easy to optimise O(1000-10000) parameters at the same time



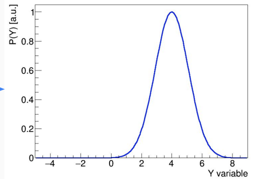


Does it work (converge)?

I should check the convergence in each corner of the feature space
Large feature space ⇒ Check convergence globally

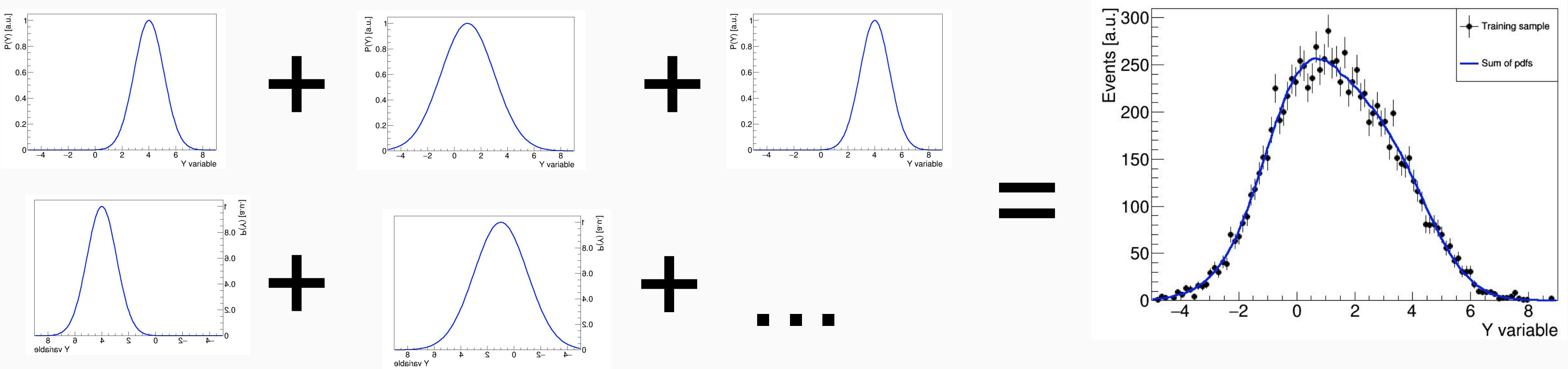
Example with the Gaussian slide 22

x (features) → **Regression** →

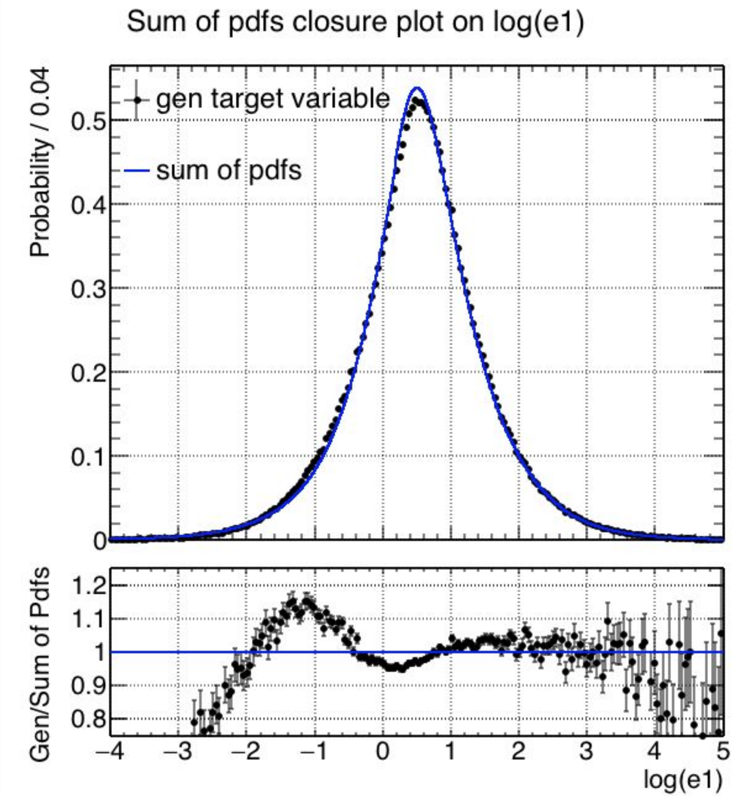


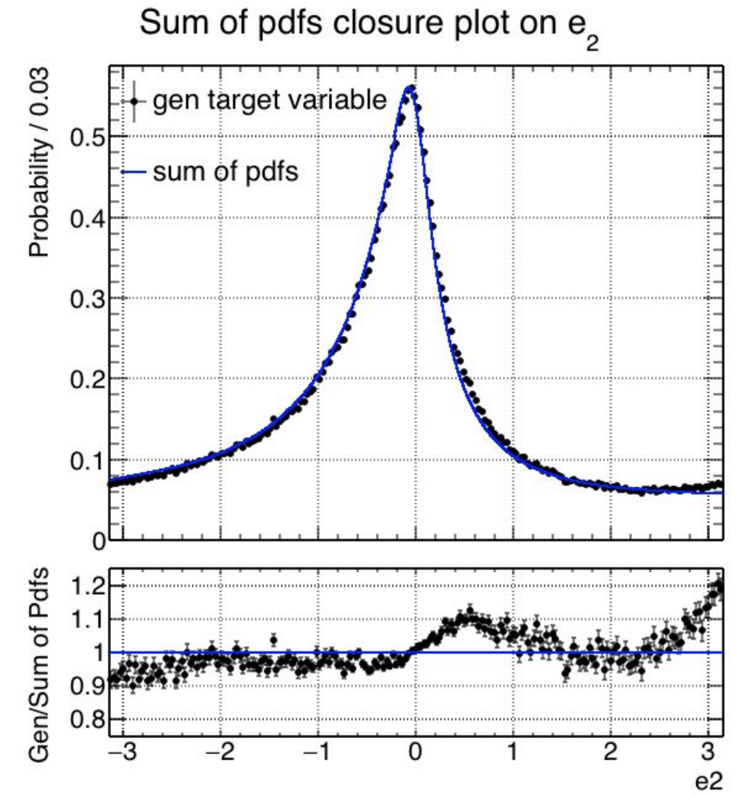
Predict the PDF
in one event

Obtain the PDF of y summing up the predicted PDFs



Sum of pdfs: our case

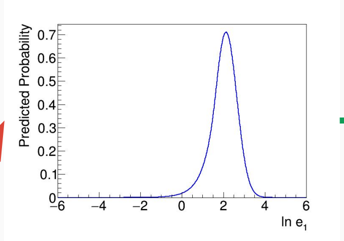




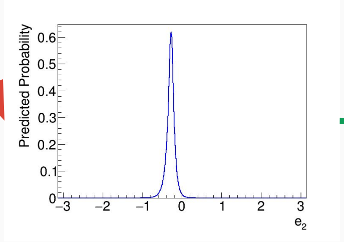
Estimators

Correction coefficients

Event features **x**



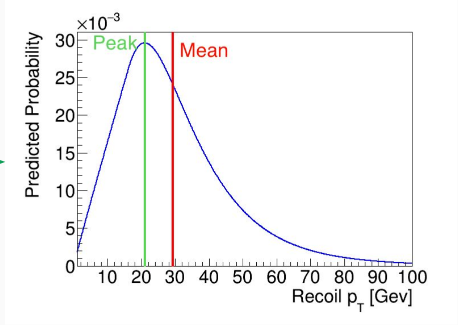
Regression

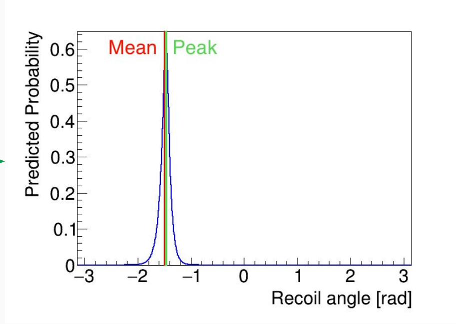


$$h_{exp} = h_{TK} \cdot e^{\ln e_1}$$

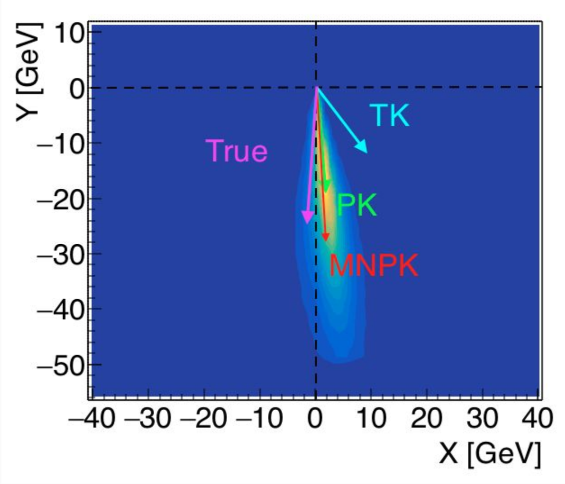
$$\phi_{exp} = \phi_{TK} + e_2$$

Recoil pdf





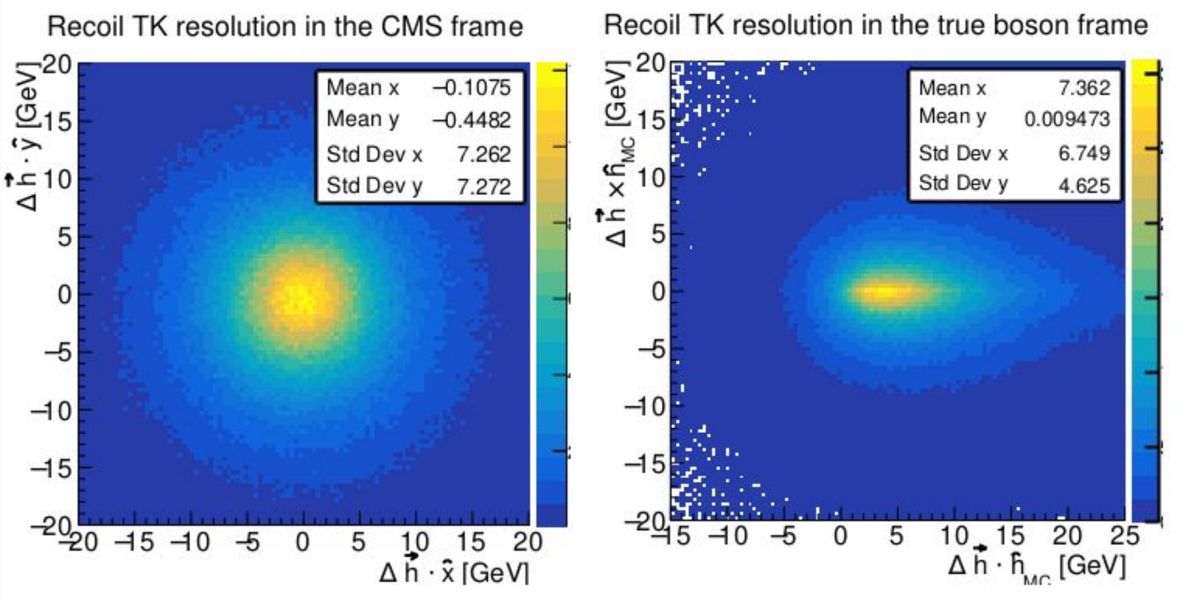
Point estimate



Which definition is better?

Not easy a priori: the recoil is a 2D object, and the resolution is hard to define.

Resolution plot in two different frames

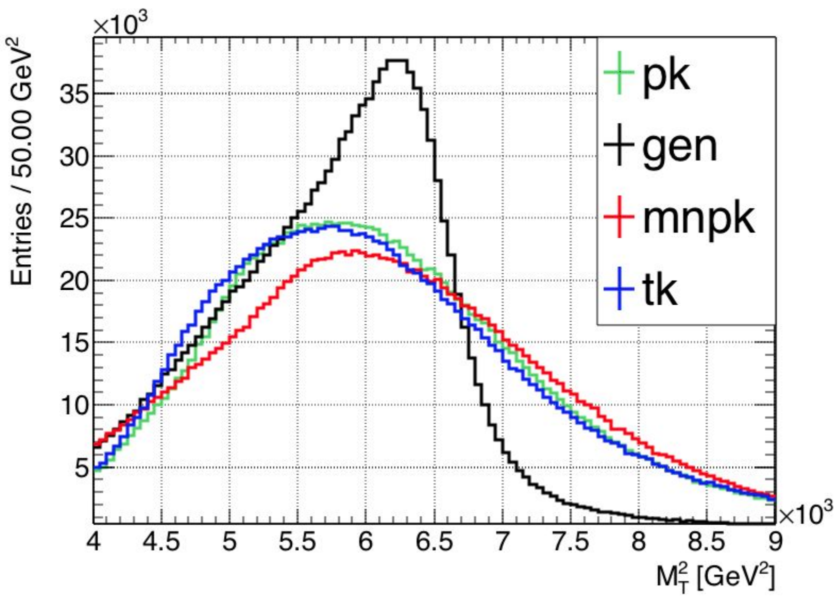


Hard to use as a figure of merit

Need to evaluate the uncertainties on the final fit!

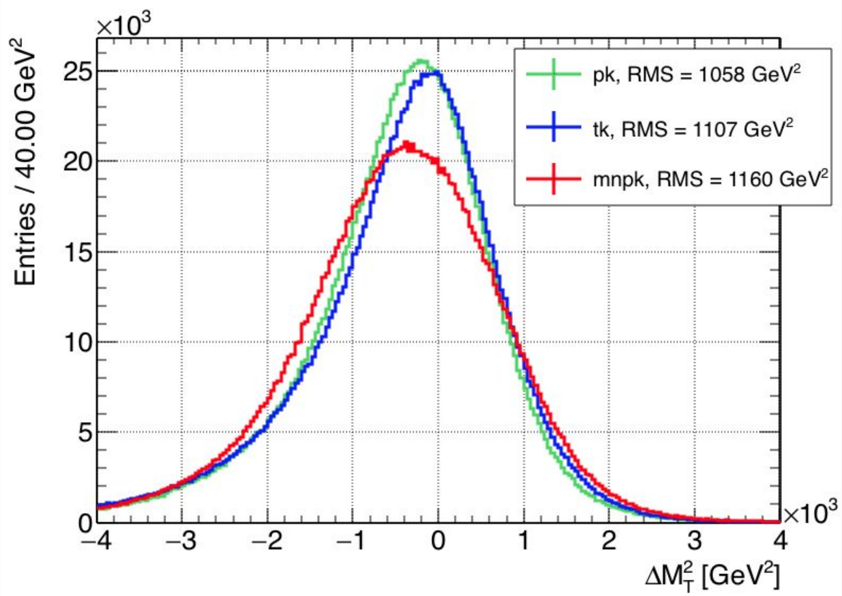
Effect on the transverse mass

Distribution used for the fit



Hard to tell which one is better...

Distribution of the resolution



It does not tell too much unfortunately… probably related to the statistical uncertainty

Evaluate the systematic uncertainties

Need to evaluate systematic uncertainties

Better measurement of the recoil:

➔ Expect a **smaller uncertainty related to p_T-W**

...But...

More information used in the transverse mass:

➔ **Larger uncertainty due to the modelling** of these variables

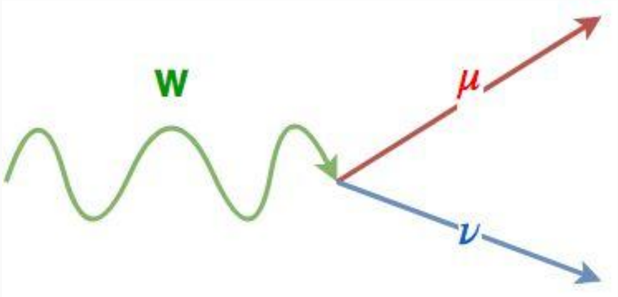
Factorisation of the systematic uncertainties

The transverse mass

$$M_T^2 = M_T^2(p_T^l, p_T^h, \Delta\phi_{l,h})$$

depends only on

- Lepton p_T
- Recoil p_T
- Relative angle



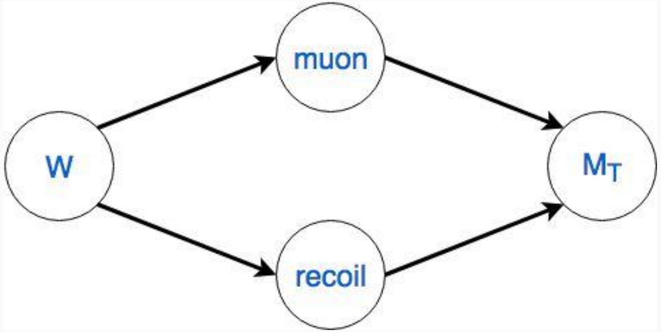
Lepton and recoil are correlated

➔ **The lepton comes from the W decay**

➔ **Recoil momentum = W momentum**

Experimental uncertainties are **uncorrelated**

How to decorrelate the theory ones?



Muon and recoil are conditionally independent given the W kinematics

Factorisation of the systematic uncertainties

Once fixed the W kinematics (z), muon and recoil are independent

$$f(M_T) = J \times \int f_l(p_T^l, \Delta\phi_{l,W}|z) f_h(p_T^h, \Delta\phi_{h,W}|z) f_W(z) dz$$

Jacobian of M_T^2

lepton pdf given W kinematics

recoil pdf given W kinematics

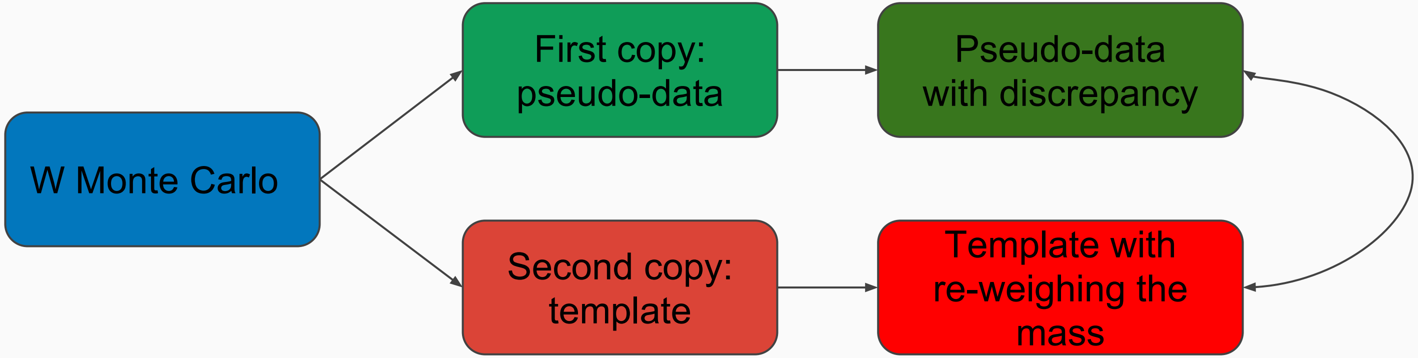
pdf of W kinematics

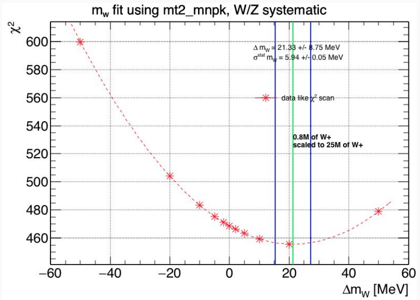
W kinematics

For our purposes W kinematics is just p_T and p_L

Evaluate the size of the systematic uncertainties

Convert a discrepancy data/MC to a bias on M_W





Template fit

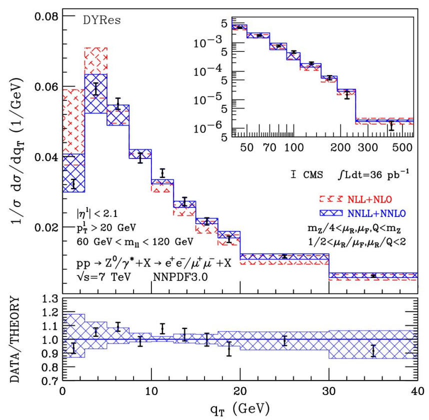
Advantages:

- ➔ Isolate the statistical and the other systematic uncertainties
- ➔ The position of the minimum is the expected bias
- ➔ The width of the parabola is the expected statistical uncertainties
- ➔ The uncertainty on the estimate of the bias can be evaluated with bootstrap

Which definition is the best?

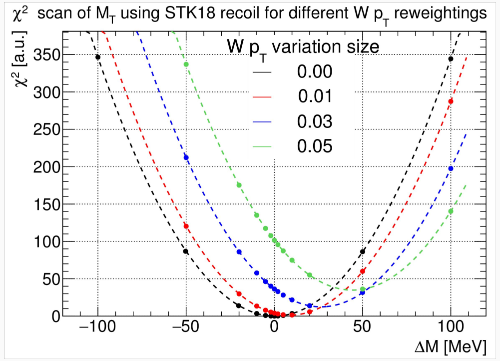
Effect of discrepancies in $f_W(z)$

$$f(M_T) = J \times \int f_l(p_T^l, \Delta\phi_{l,W}|z) f_h(p_T^h, \Delta\phi_{h,W}|z) f_W(z) dz$$



Add **discrepancies between data and simulation in the p_T-W spectrum**

➔ ~ uncertainty on the theory prediction

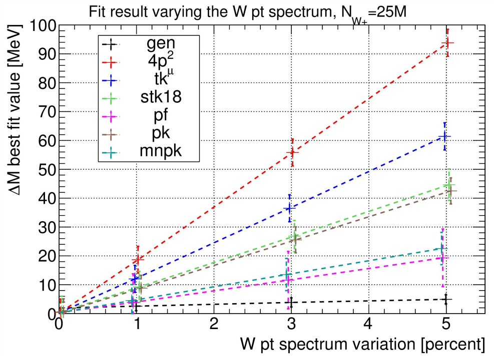


Likelihood scan for different value of the discrepancy:

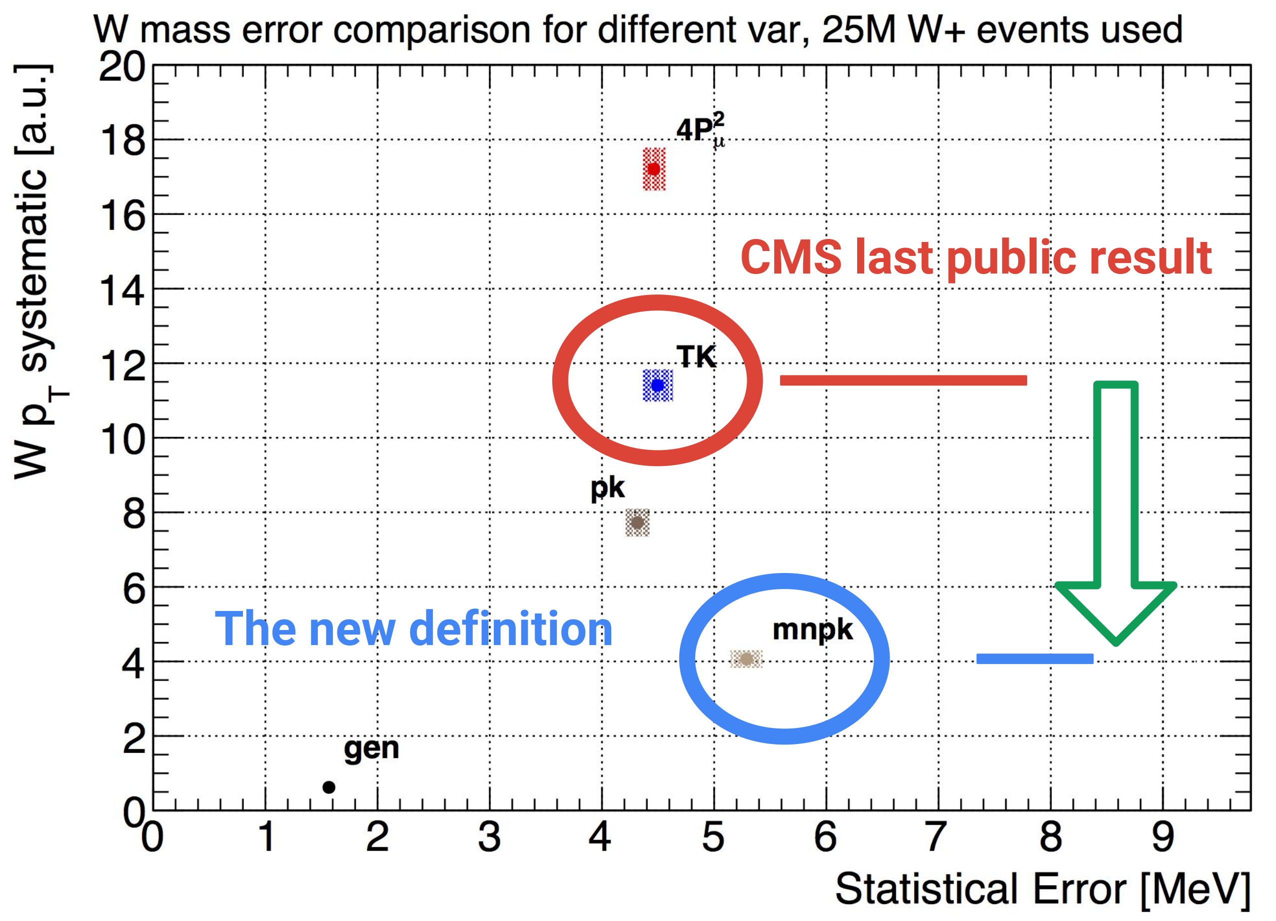
systematic uncertainty

=

Bias for unit of discrepancy



Which definition is the best?



Almost a factor 3 improvement in the p_T-W systematic uncertainty!

Recoil calibration

N-dimensional k-conditional quantile morphing,
based on Boosted Decision Trees

Recoil modeling

$$f(M_T) = J \times \int f_l(p_T^l, \Delta\phi_{l,W}|z) f_h(p_T^h, \Delta\phi_{h,W}|z) f_W(z) dz$$

Distribution of the measured recoil, once fixed the true value of recoil momentum

Recall: $y = \{p_T^h, \Delta\Phi_{h,W}\}$, $z = \{p_T^W, p_L^W\}$

$f(y|z)_{DATA} \neq f(y|z)_{MC}$ implies systematic uncertainty due to the modeling

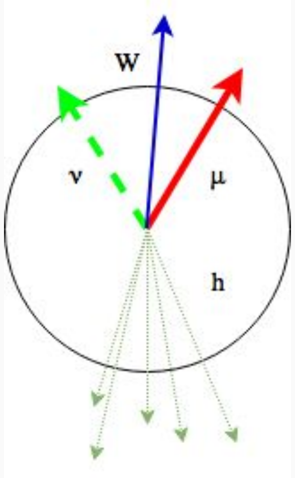
Questions:

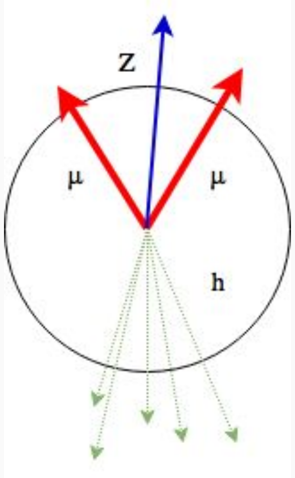
1) How large is this systematic?
2) If large, how can I correct (calibrate) for it?

A tool to calibrate the recoil: the Z leptonic events

Z leptonic decay: $Z \longrightarrow ll$

- Z and W bosons are similar for production mode
- Z⇒ll: the kinematics of the Z can be well reconstructed in the data
- I can access f(y|z) in both data and simulation



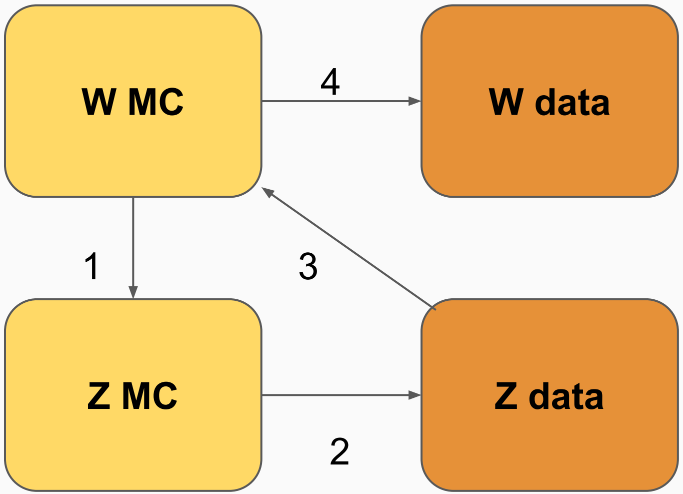


Correct the recoil "mismodeling" with Z events

Correct the distribution f(y|z) for each value of z

=

For each event: $y \longrightarrow y' = T(y, z)$ such that $f(y'|z)_{MC} = f(y|z)_{DATA}$



Strategy:

➔ compare Z_{MC} and Z_{data}

➔ derive corrections

➔ apply on W_{MC}

Let's build T(y, z)

Several problems:

1) **y** is bidimensional (correlation is important!)
2) Conditional means making bins, hard with many dimensions

The program:

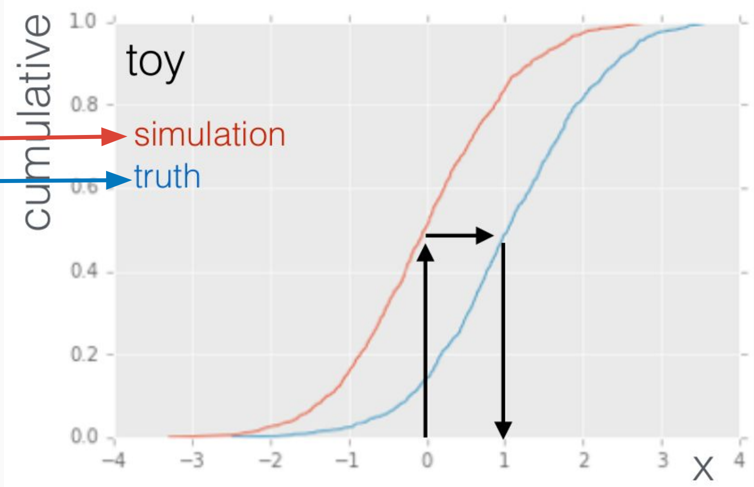
1) 1D quantile morphing (N=1, k=0)
2) Make it N-dimensional
3) Make it k-conditional and unbinned using quantile regression BDT based

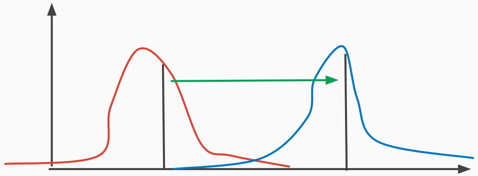
1) 1D quantile morphing (N=1, k=0)

Purpose: find a function that transforms $f_{MC}(y) \to f_{DATA}(y)$

It is a change of variables, but I know the PDFs only through a sample

- Build cumulative distributions $F_{MC}(X)$ $F_{DATA}(X)$
- Associate points with the same quantile values: $X \to F_{DATA}^{-1}(F_{MC}(X))$





2) Extension to 2 (or N) variables

One by one:

- Take the first variable and do 1D quantile morphing
- Take the second variable and do 1D quantile morphing, conditional to the first one
- Take the third variable and do 1D quantile morphing, conditional to the first and the second ones
- Iterate until the last variable

Bottom line:

A N-dimensional morphing can be seen as a sequence of N-1, N-2, .., 0 conditional morphing it is enough to have a 1D k-conditional quantile morphing

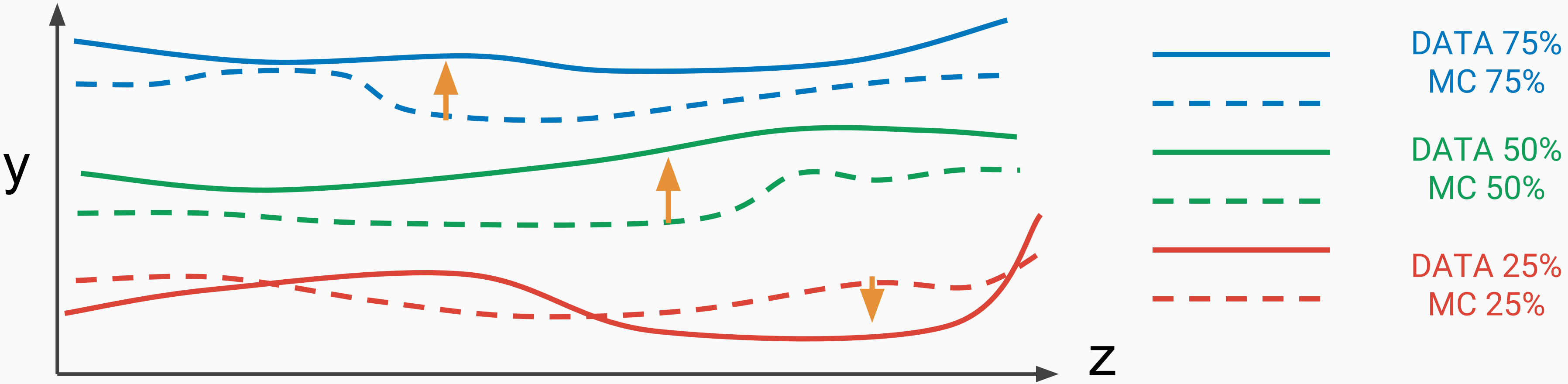
3) The hard step: do it conditional

Typically "conditional = in bins of"

- ➔ Impossible with many dimensions, as the number of bins scales as $\#bins^{k\text{-}dim}$

Need for an unbinned way **to tell the quantile**:

- ➔ a function $f(z; \tau)$ for some values of the quantile τ
- ➔ If I have $f_{DATA}(z; \tau)$ and $f_{MC}(z; \tau)$ for many values of τ (eg: 10%, 20%, ..., 90%) the matching is done



The solution: quantile regression with Boosted Decision Trees

REMINDER: regression = technique to fit parameters $\alpha(x)$ of a PDF f(y|x)

- Least squares: regression of the mean
- Recoil definition: 4 parameters that fully parametrise the distribution
- **Quantile regression: regression of the quantile $\tau(x)$**

How does it work?

Proper Loss function: $Q_{loss}(\tau) = \sum_{y_i < f(x_i)} \tau(y_i - f(x_i)) + \sum_{y_i > f(x_i)} (1-\tau)(f(x_i) - y_i)$

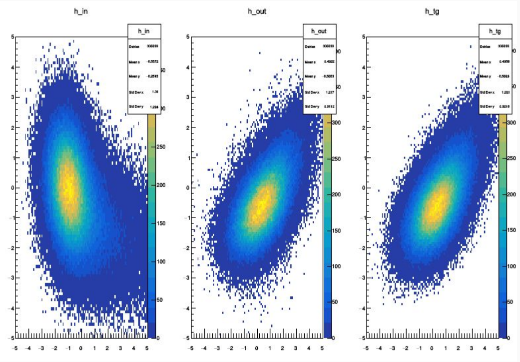
Need for a versatile arbitrarily non-linear function

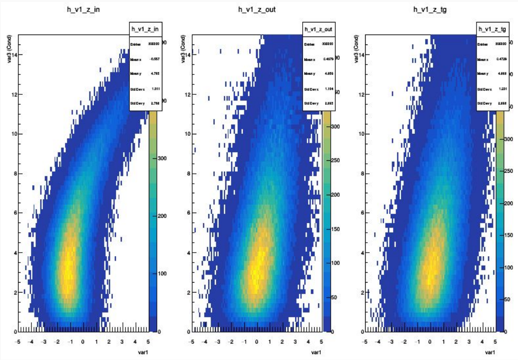
- Neural networks would do the job
- Boosted decision trees found to work better in this case

Summary: the procedure

- A set of “quantile regressions”, for 10 different values (5%, 10%, 20%, etc.)
- Input of regressions are the k variables
- For each value of the quantile τ, I can predict y_τ on data and MC
- The association is y_τ^{MC}->y_τ^{DATA}

Proof of concept on Gaussian variables: N=1, k=1





Summary: the application

Tool to solve a new and general problem:

Morph N-dimensional distribution conditional to k variables

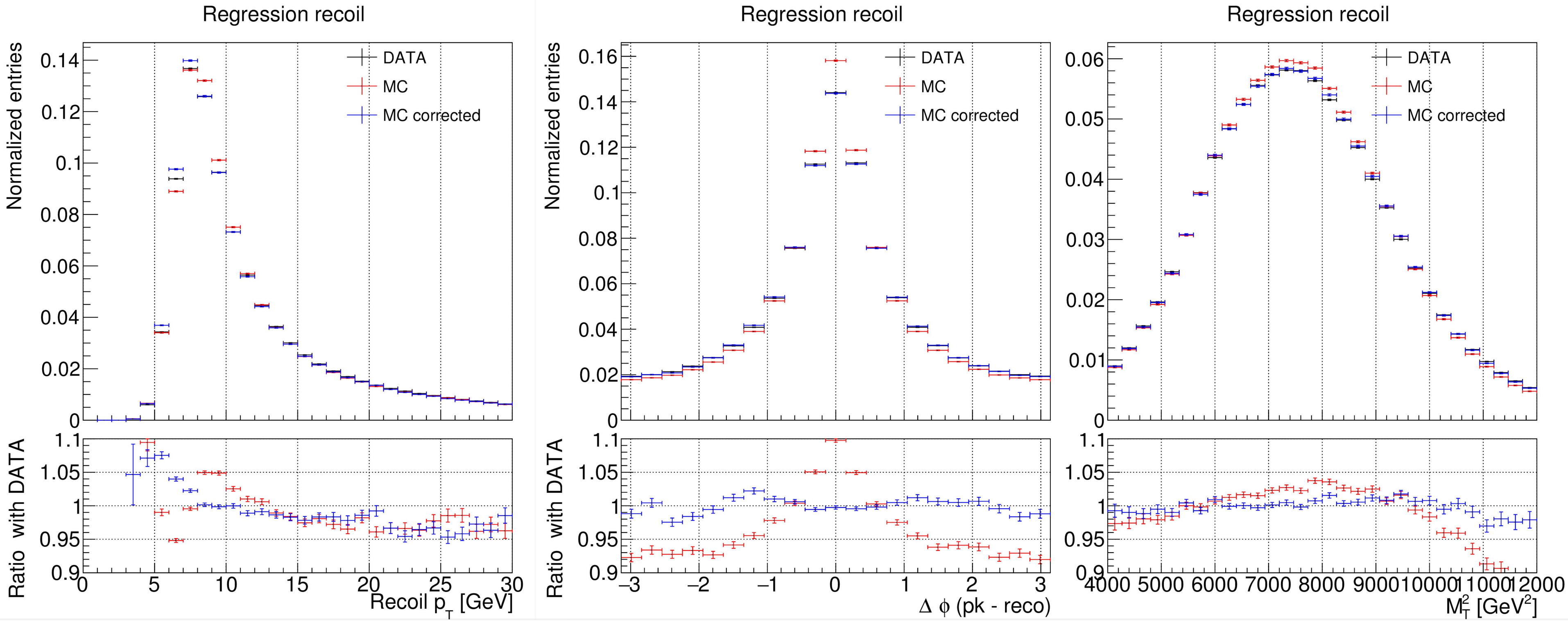
With this tool we build a function y’ = T(y, z) such that

$$f_{MC}(y|z) \rightarrow f_{MC}(y'|z) = f_{DATA}(y|z)$$

Now I can use the tool in order to:

- ➢ Derive and apply a correction to the W Monte Carlo
- ➢ Estimate the bias if no correction is applied
- ➢ Estimate the systematic after the correction is applied.

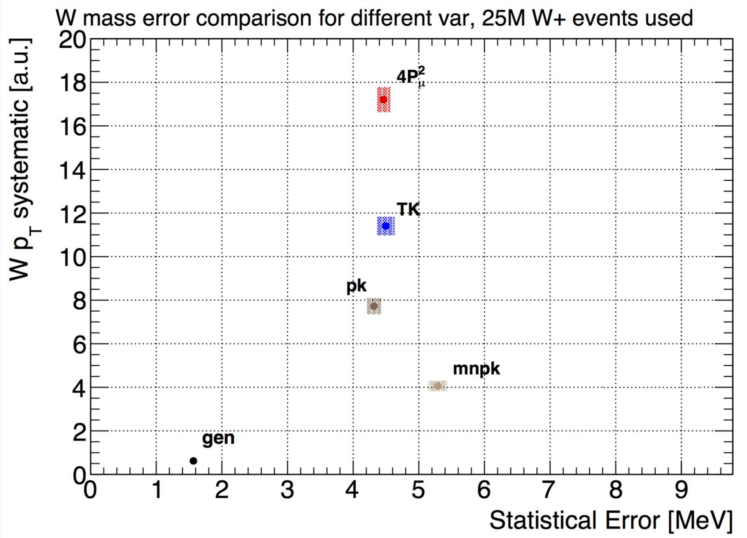
Before and after morphing



Results on systematic uncertainties

Trade off: more information used in the recoil implies:

Smaller systematic uncertainty due to W-p_T



MNPK = 1/3 di TK

Larger systematic due to the modeling

TK: 29 ± 11 MeV

MNPK: 140 ± 14 MeV

MNPK = 5 x TK

I can correct for it

TK: -11 ± 10 MeV

MNPK: -14 ± 7 MeV

MNPK ≃ TK

Conclusions

The W mass measurement in this historical moment

- A new precision measurement of the W mass is needed
- The systematic uncertainties are **the problem**
- The systematic uncertainty on the W-p_T spectrum is one of the biggest

Machine learning approach

- New experimental definition of the recoil
  - Semi-parametric regression with custom loss function
  - Based on Deep Neural Networks
- Methodology to decompose the systematic uncertainties
- Calibration of the simulation to reduce modeling uncertainties
  - Multi-dimensional extension of the quantile morphing
  - Based on quantile regressions, implemented on Boosted Decision Trees

Thank you!

Special thanks to:

SCUOLA NORMALE SUPERIORE

CERN

INFN
Istituto Nazionale Fisica Nucleare Sezione di Pisa

CMS

BACKUP

Electroweak Precision Observables (EWPO)

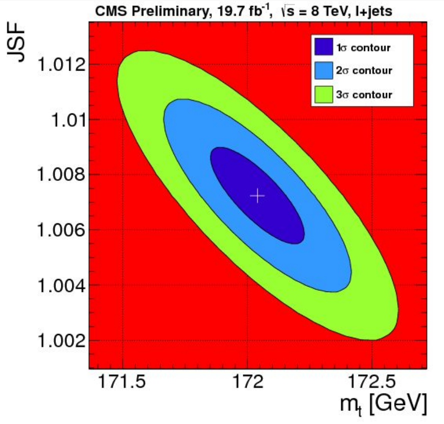
The Standard Model predicts a precise set of relations between observables

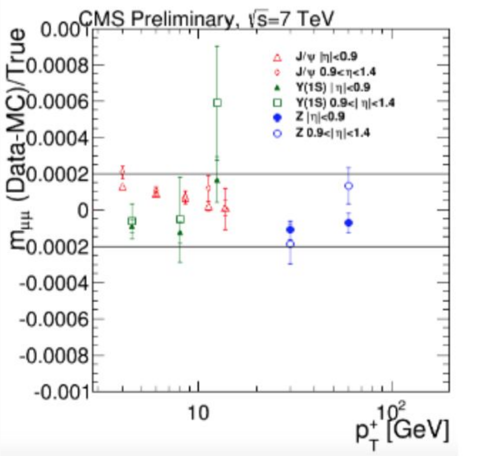
The comparison (EW fit) between the prediction and the measurement of the EWPOs is a severe test for the Standard Model

$\alpha \quad G_F \quad M_Z$ —Tree→ $$M_W^2 = \frac{\pi\alpha}{\sqrt{2}G_F \sin^2\theta_W}$$

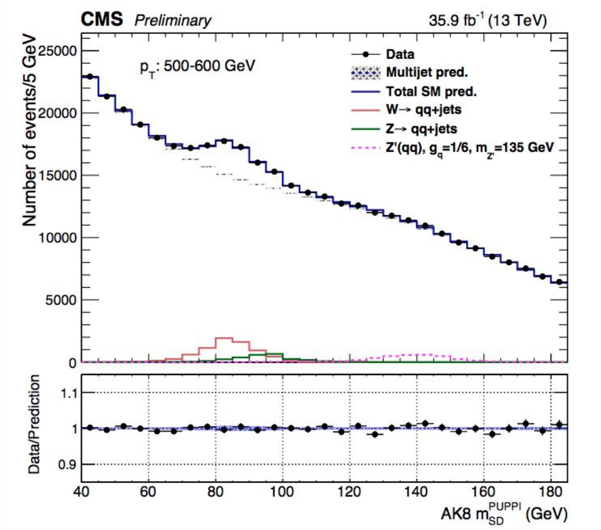
$\alpha_s \quad m_t \quad M_H$ —Corr.→

I decadimenti leptonici: una scala più precisa





Anche in regioni particolari dello spazio delle fasi il S/B per W→ qq è molto basso.
A parità di precisione necessita di molta più statistica



La precisione sulla scala della misura dell'oggetto fisico dominante da la precisione della misura

- Muon momentum scale: $2 \cdot 10^{-4}$
- Jet energy scale: $3 \cdot 10^{-3}$

Systematic uncertainties in the template fit

Many ingredients in the simulation:

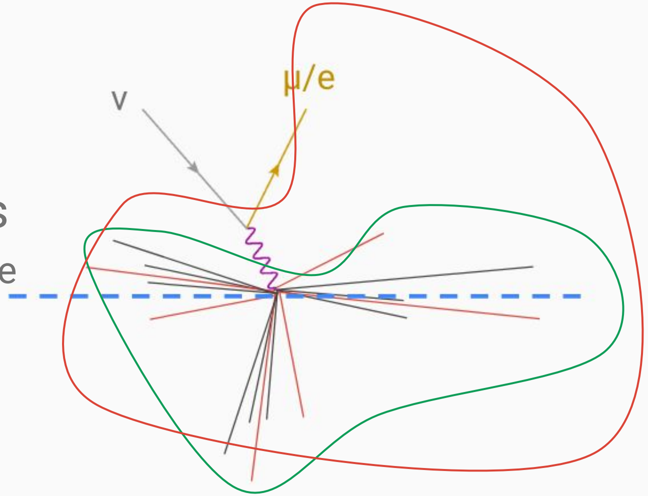
- Physics of the process, from the initial to the final state
  - Incoming protons and PDFs
  - Kinematics of the W boson
  - Polarisation of the W boson and decay
  - Hadronization of quarks and gluons
  - Pileup
- Simulation of the detector
  - Acceptance
  - Efficiency
  - Trigger
  - Energy/Moentum scale and Resolution

The uncertainty in the prediction of the physics of the process or of the model of the detector **translates** to a systematic uncertainty

MET vs recoil

Example:

- Lepton variables: θ and p_T
  - Depend only on the kinematics and polarisation of the W
- Variable of the recoil: number of charged particels
  - Depends only on the kinematics of the W against which the recoil is recoiling
- For a fixed value of the kinematics of the W
  - Muon and recoil have fixed distribution
  - They are uncorrelated



The same is not true when I consider the neutrino

- Experimentally the neutrino is called MET: sum of lepton and recoil
- Is correlated with the lepton by definition
- They are not conditionally independent given the W kinematics

The transverse mass

$$M^2 = 2p_\mu p_\nu[\cosh(\eta_\mu - \eta_\nu) - \cos\Delta\phi]$$

Cannot measure $p_L(\nu)$

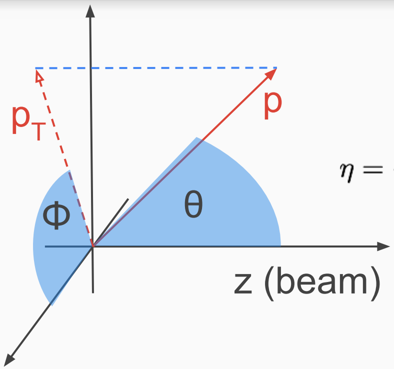
⇒ Cannot measure $\Delta\eta = \eta_\mu - \eta_\nu$

$\Delta\eta = 0$

$$M_T^2 = 2p_\mu p_\nu(1 - \cos\Delta\phi)$$

$$\vec{p}_\nu = -\vec{h} - \vec{p}_\mu$$

$$M_T^2 = 2p_\mu \left|\vec{p}_\mu + \vec{h}\right| + 2p_\mu^2 + 2\vec{p}_\mu \cdot \vec{h}$$



$$\eta = -\ln\left(\tan\frac{\theta}{2}\right)$$

If neutrino and muon have the same angular distance from the beam axis ⇒ $M = M_T$

Tutti gli impulsi sono intesi come le proiezioni sul piano trasverso

Scelta delle osservabili (2/2): L'impulso del muone

$m_\mu = 0$ (approx ~ $3 \cdot 10^{-6}$)

$p^*_\mu = m_W/2 \approx 40$ GeV
$h = p_W \approx 5$ GeV

⇒

$p^*_\mu \approx p_\mu$
$h << p_\mu$

$$M_T^2 = 2p_\mu \left|\vec{p}_\mu + \vec{h}\right| + 2p_\mu^2 + 2\vec{p}_\mu \cdot \vec{h}$$

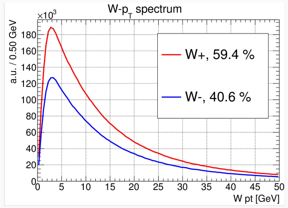
⇒

$$M_T^2 \approx 4p_\mu^2$$

Lo stesso impulso trasverso del muone p_μ è una buona osservabile per estrarre il valore di m_W

Lepton p_T

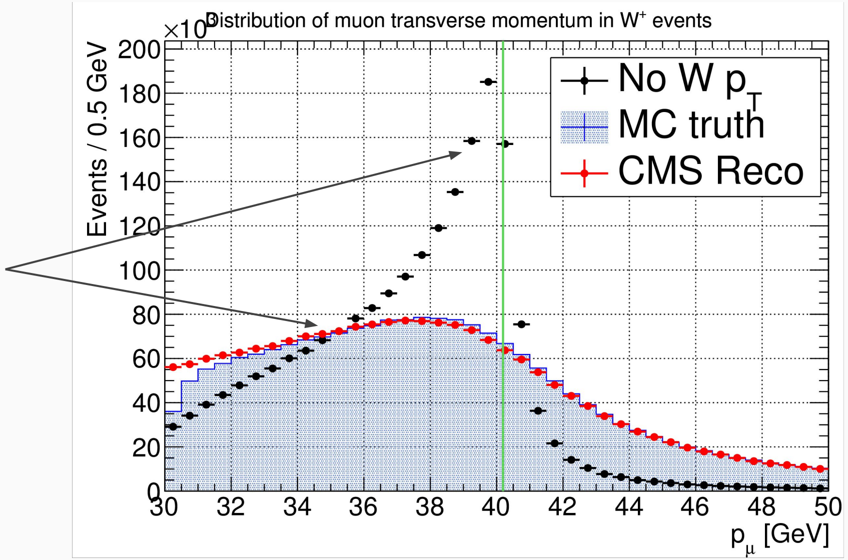
The transverse momentum of the lepton is sensitive to the W mass



➔ The maximum is at $m_W/2$ in case of not p_T-W
➔ With p_T-W the distribution gets broader

Systematic uncertainties:

- Production of W and its decay
  - W polarisation
  - p_T-W
  - Final state radiation
- Detector simulation
  - Scale of transverse momentum

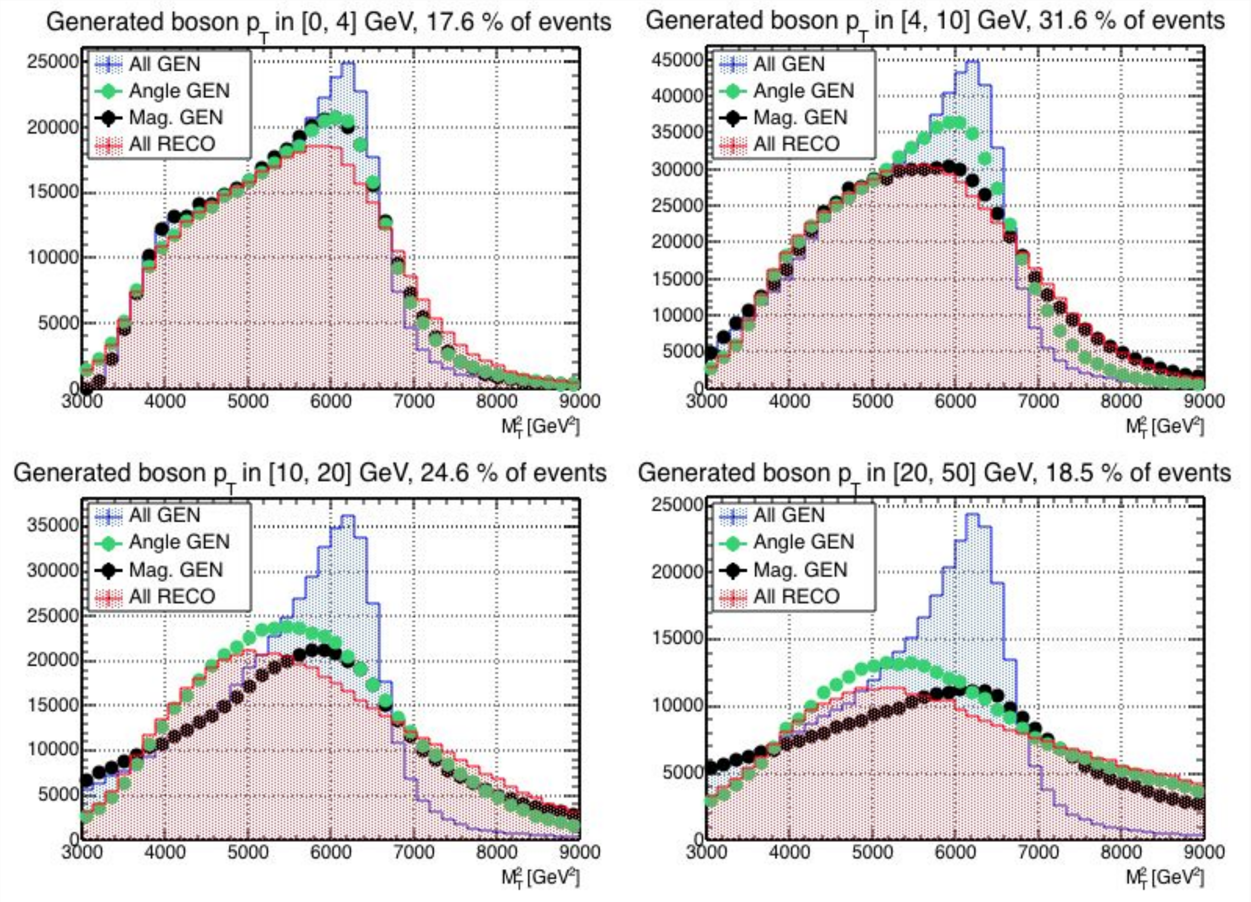


Amount of events in the samples

Final state only muon-neutrino

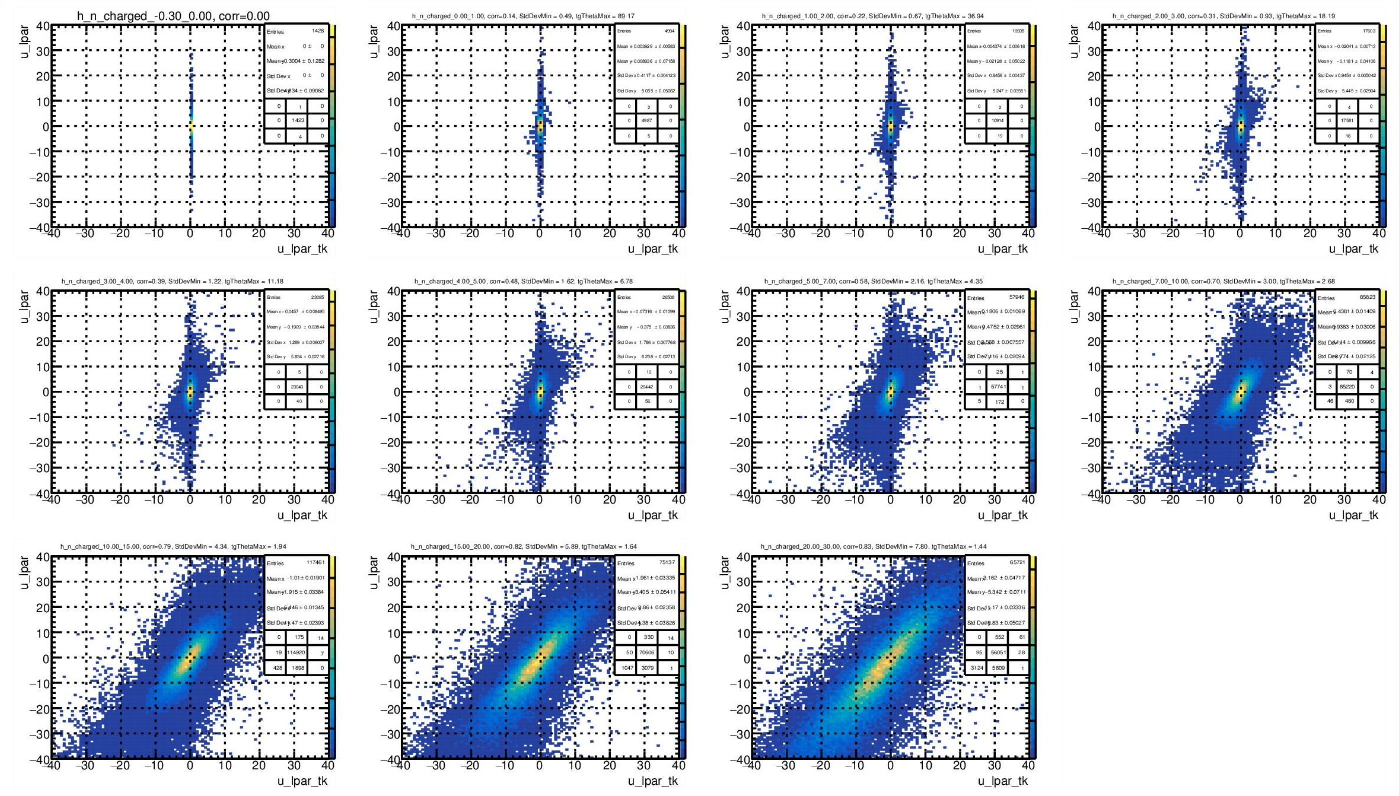
Esprimento	√s [TeV]	Eventi selezionati	note
CMS (W+)	8 (19 fb^-1)	25 M	120M prodotti * 0.2
CMS (W+)	13 (70 fb^-1)	~ 150	sigma_W e’ doppia
ATLAS (W)	7 (4.6 fb^-1)	8 M	
CDF (W)	2 (2.2 fb^-1)	0.6 M	

Importance of the angle in the recoil



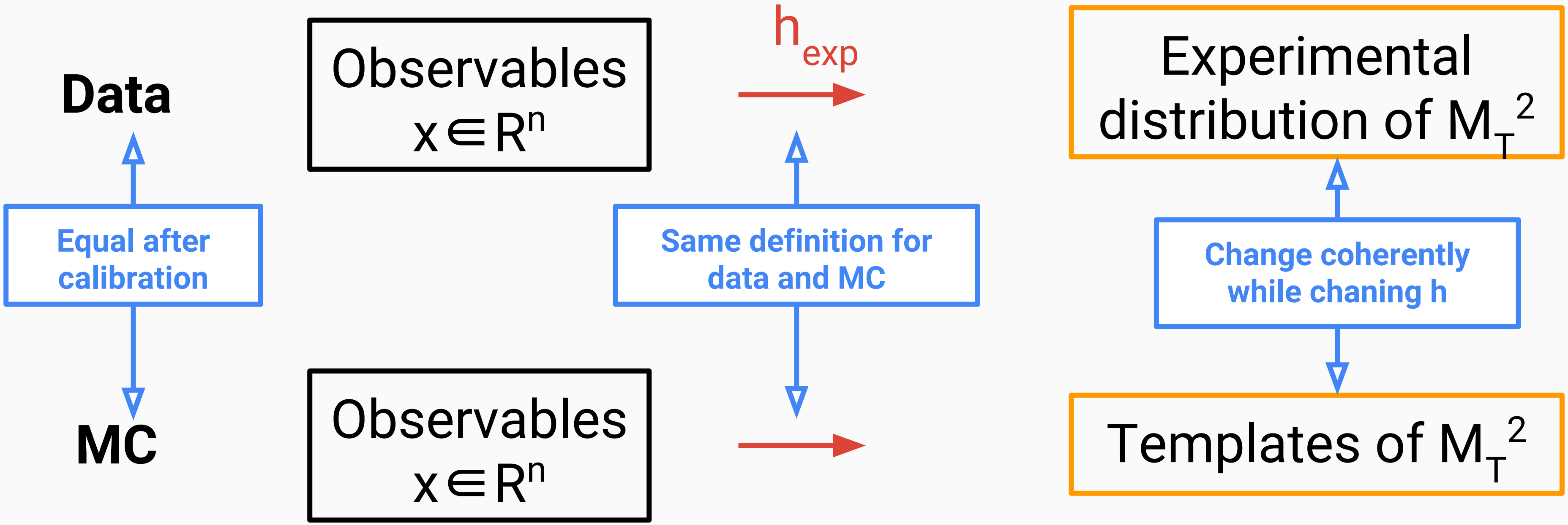
Correlation between recoil and observables

Correlation between measured and true recoil for different values of the number of charged tracks.



No bias due to the recoil definition

Every function is fine, as long as it is applied on both data and Monte Carlo



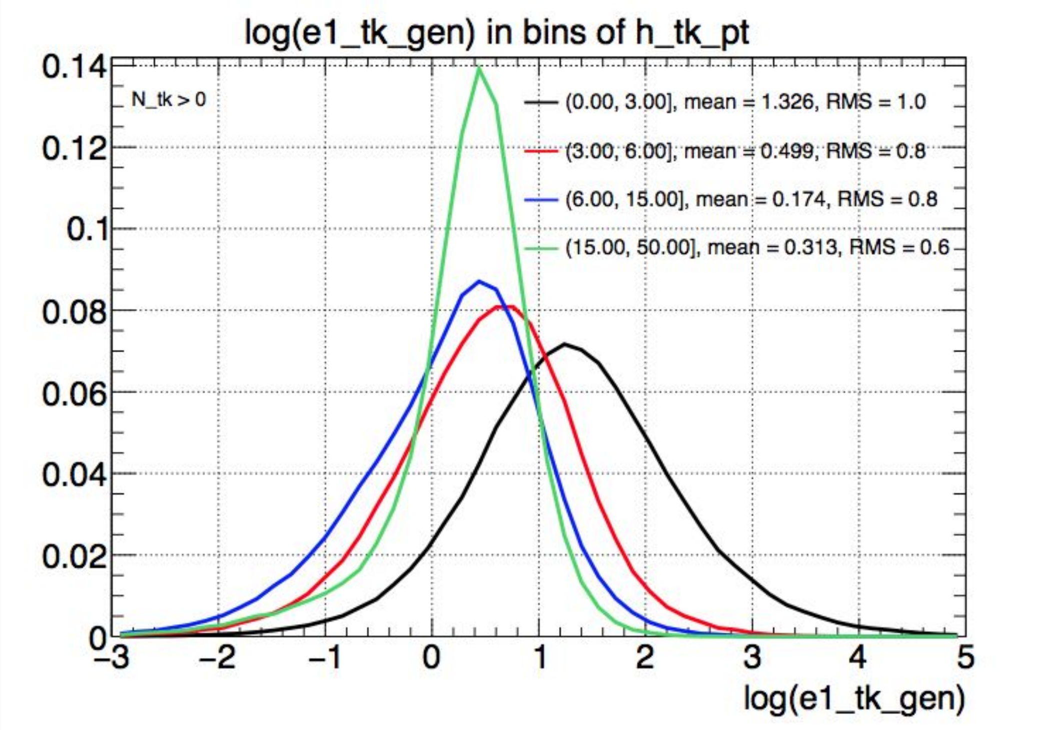
E.g.: If the training sample does not represent the data ⇨ Estimator of the recoil less powerful, but still no bias

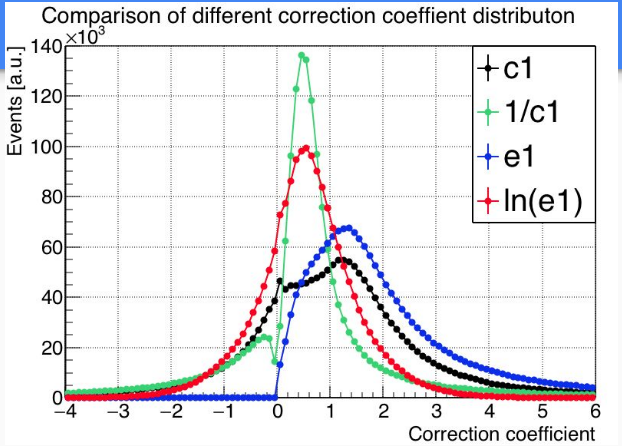
Input variables

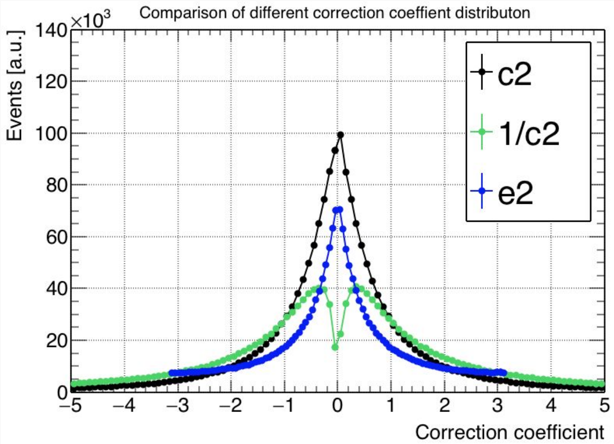
Symbol	Formula	Notes	$\ln(e1)$	$e2$
h_{tk}	$\lvert\vec{h}_{TK}\rvert$		✓	✓
ϕ_{tk}		Azymuthal angle of TK recoil	✓	✓
R_{ntnpv}	$\ln\left(h_{ntnpv}/h_{TK}\right)$		✓	✓
R_m	$\ln(m_{TK}/h_{TK})$	m_{TK} is the invariant mass of the TK recoil	✓	✓
R_{lTK}	$\ln(p_T^{\text{leading TK}}/h_{TK})$	leading TK is track with highest p_T	✓	✓
R_{lnt}	$\ln(p_T^{\text{leading nt}}/h_{TK})$	leading nt is neutral cluster with highest p_T	✓	✓
S_{TK}	$\dfrac{\left\lvert\sum_{i\in TK}\vec{p}_T^{(i)}\right\rvert}{\sum_{i\in TK}p_T^{(i)}}$	Vector over scalar sum of particles $\vec{p}_T$ belonging to TK recoil	✓	✓
S_{nt}	"	Same as above but with neutral clusters	✓	✓
S_{PF}	"	Same as above but with all PF candidates	✓	✓
	$\cos(\phi_{tk}-\phi_{\text{ntnpv}})$		✓	
	$\lvert\Delta\phi_{TK-ntnpv}\rvert$	Modulus of angle between TK and ntnpv recoil		✓
	$\varsigma\cdot\Delta\phi_{TK-PF}$	Angle between TK and PF recoil		✓
	$\varsigma\cdot\Delta\phi_{TK-\text{leading TK}}$	Angle between TK and leading track		✓
N_{TK}		# of tracks from PV	✓	✓
N_{vtx}		# of reconstructed vertexes	✓	✓

$$\varsigma = sgn(\Delta\phi_{TK-ntnpv})$$

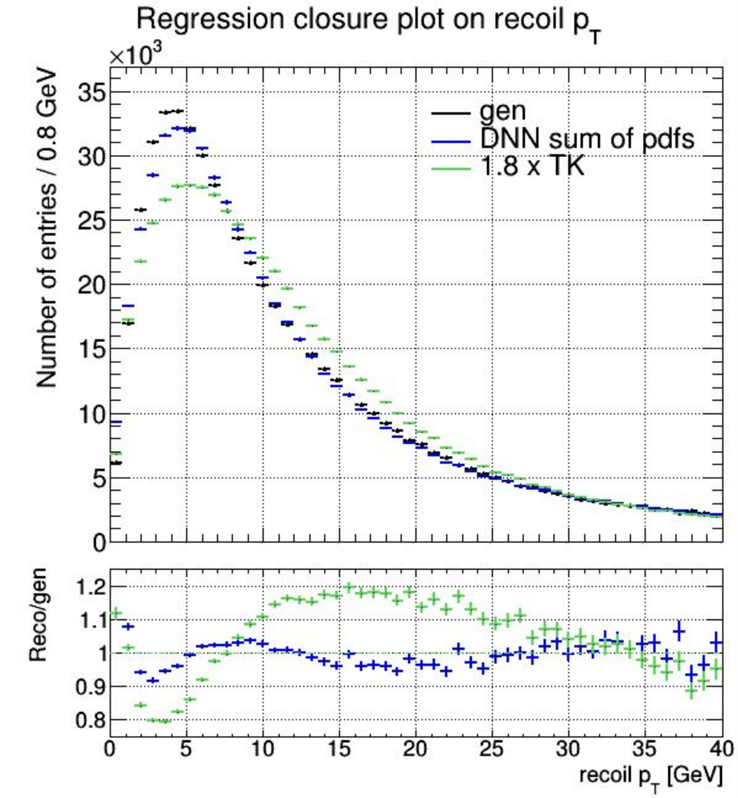
Integrate distribution of the corrections

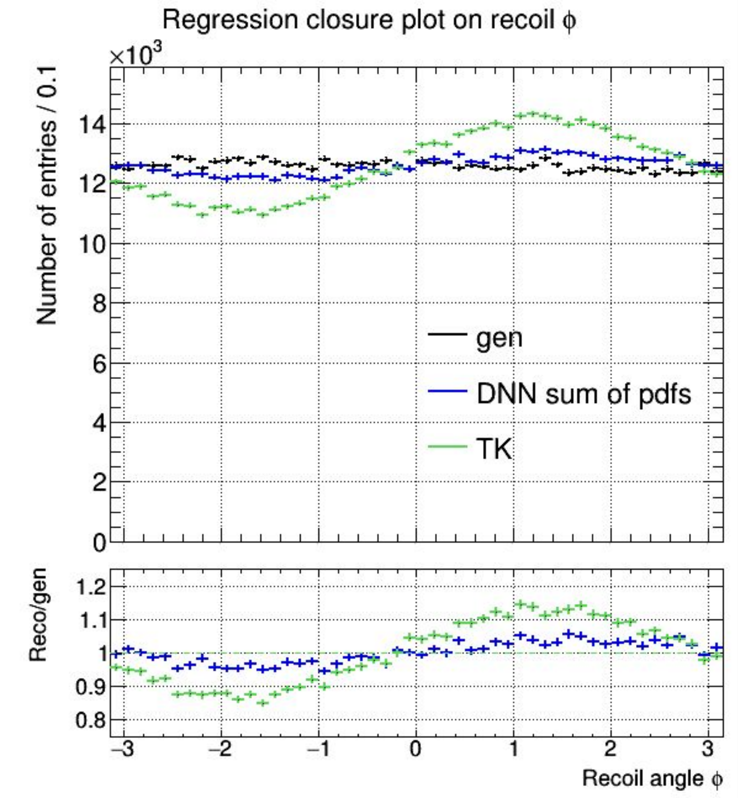




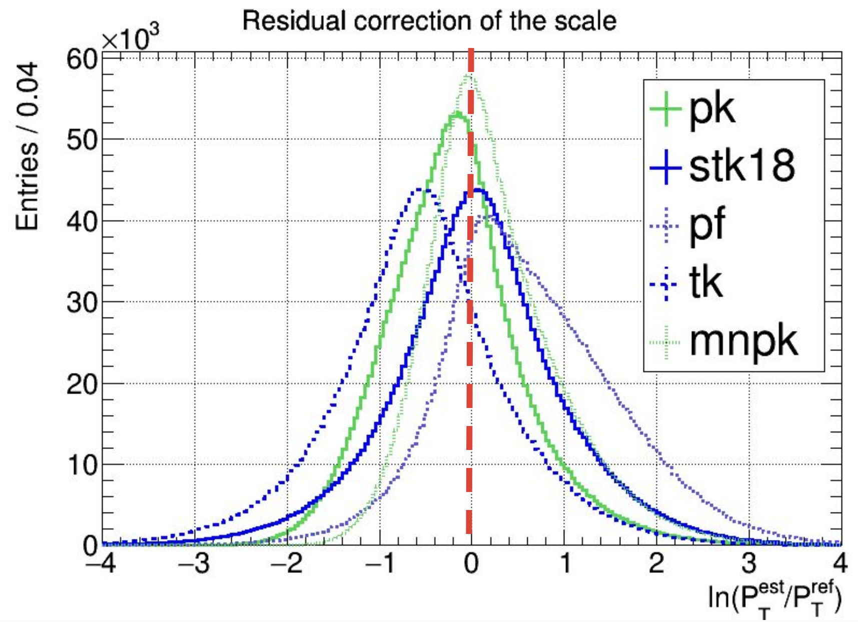


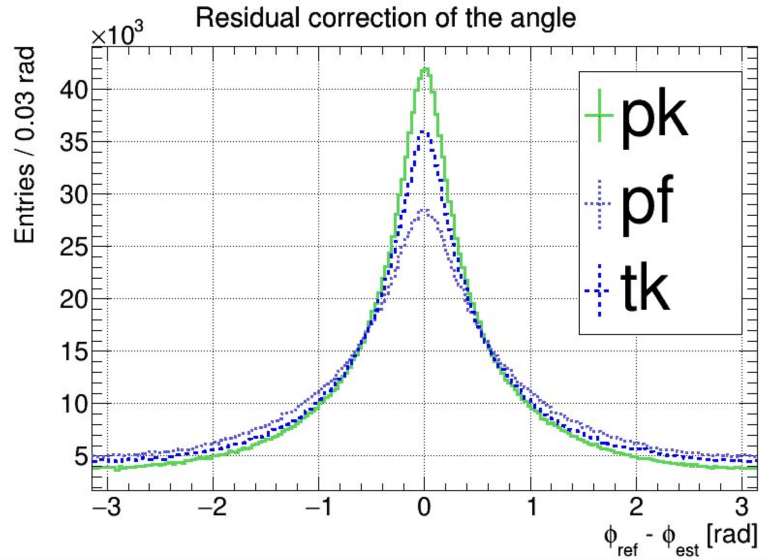
Sum of pdfs: magnitude and angle of the recoil





Residual correction





RMS: 0.96 (STK18) ⇒ 0.74 (MNPK)

RMS: 1.32 (TK) ⇒ 1.24 (PK)

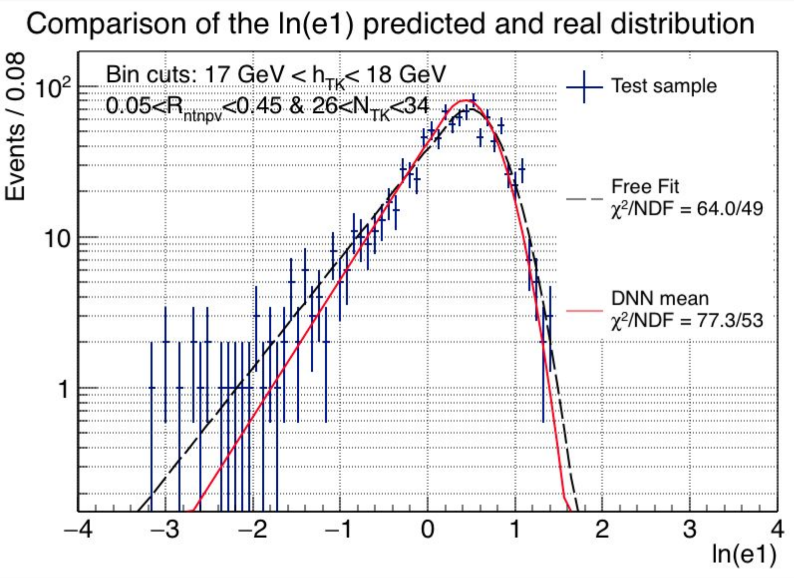
Regression: functions and loss function

$$P(y|x) = P(\ln(e1)|x_1) \times P(e2|x_2)$$

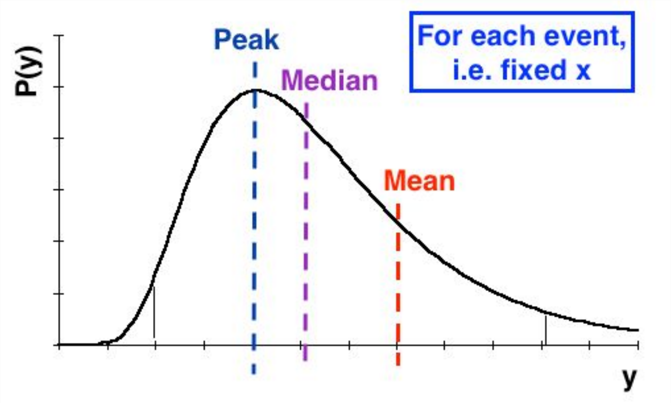
$$f_{lne1}(y|\vec{\alpha}_{lne1}(x_1)) = \begin{cases} \frac{1}{N} e^{\alpha_1^2/2} e^{\alpha_1 t}, & \text{if } t < -\alpha 1 \\ \frac{1}{N} e^{t^2/2}, & \text{if } -\alpha_1 \leq t \leq \alpha_2 \\ \frac{1}{N} e^{\alpha_2^2/2} e^{-\alpha_2 t}, & \text{if } t > \alpha_2 \end{cases}$$

$$\text{loss} = \ln \mathcal{L} = \sum_i \ln \left(f_{lne1}(\ln(e1^i)|\vec{\alpha}_{lne1}(x_1^i)) \right)$$

$$N = N(\sigma, \alpha_1, \alpha_2) = \sqrt{\frac{\pi}{2}} \sigma \left[\text{erf}\left(\frac{\alpha_2}{\sqrt{2}}\right) - \text{erf}\left(\frac{\alpha_1}{\sqrt{2}}\right) \right] + \frac{e^{-\frac{\alpha_2^2}{2}}}{\alpha_2} \sigma + \frac{e^{-\frac{\alpha_1^2}{2}}}{\alpha_1} \sigma$$



The estimators



$$\phi_{PK} = \phi_{TK} + \mu_{e_2}$$

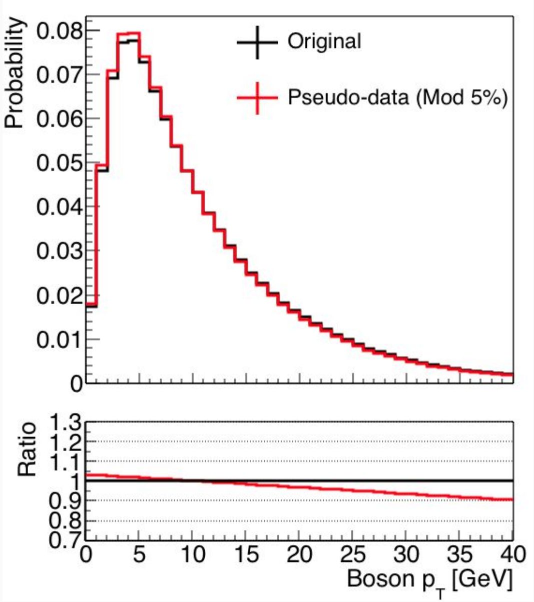
$$h_{PK} = h_{TK} \cdot \theta\left(\alpha_1 - \sigma\right) \cdot \exp\left(\mu - \sigma^2\right)$$

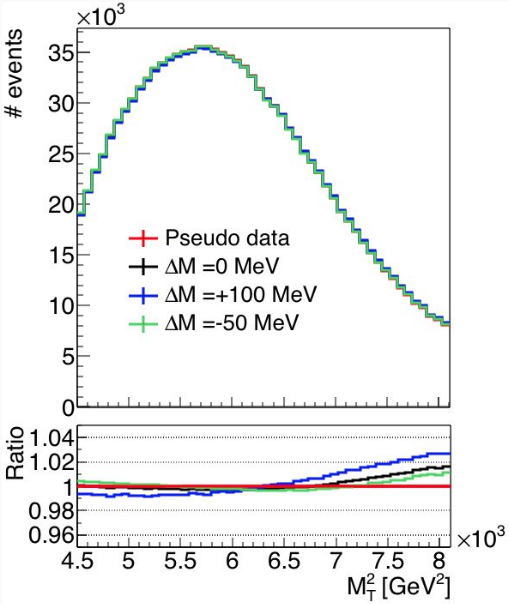
$$h_{MN} = h_{TK} \cdot \left[\frac{\exp(\mu - \alpha_1\sigma - \alpha_1^2/2)}{1 + \alpha_1/\sigma} + e^{\sigma^2/2+\mu}\sqrt{\frac{\pi}{2}}\sigma\left(\mathrm{erf}\left(\frac{\alpha_2 - \sigma}{\sqrt{2}}\right) - \mathrm{erf}\left(\frac{-\alpha_1 - \sigma}{\sqrt{2}}\right)\right) + \frac{\exp(\mu + \alpha_2\sigma - \alpha_2^2/2)}{\alpha_2/\sigma - 1}\right]$$

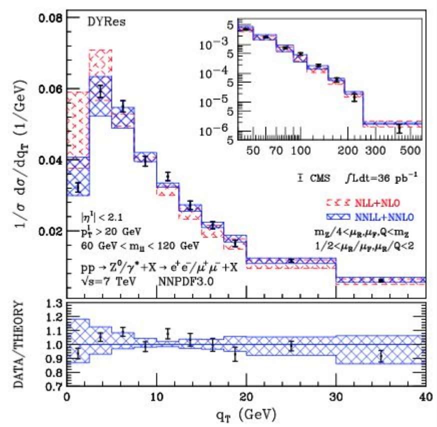
Modification of the pseudo-dati

- Spectrum modified by small amount
- Effects on M_T^2 are hard to tell by eye

Uncertainty from the literature







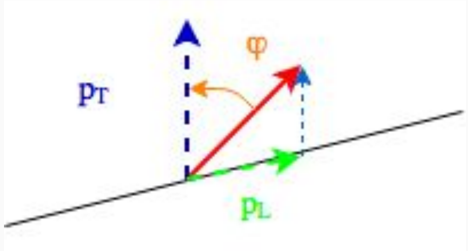
Parametrising the 4-momentum of the W: the “z” variables

The correlation between W and recoil starts from the conservation of 3-momentum

3D momentum of the recoil = 3D momentum of the W

Describe the 4-momentum of the W:

- Invariant mass M
    - The recoil does not depend on the W mass
- Transverse and longitudinal momentum p_T and p_L
- Azimuthal angle Φ
    - Negligible in the limit of detector symmetry around Φ
    - Parametrise with the angle with respect to the W
        - No dependence on the angle Φ of the W



Two variables describing the W kinematics in the context of the recoil: p_T and p_L

Understanding the discrepancies

Experimentally the recoil = set of reconstructed particles

Described by some variables **x**

Eg: x = $\{h_{TK}, N_{TK}, h_{NT}\}$

The definition of the recoil y = R(x) → $$f(y|z) = \frac{dx}{dR} f(x|z)$$

Discrepancies in f(y|z) depend on:

- discrepancies in f(x|z)
- Recoil definition R(x)

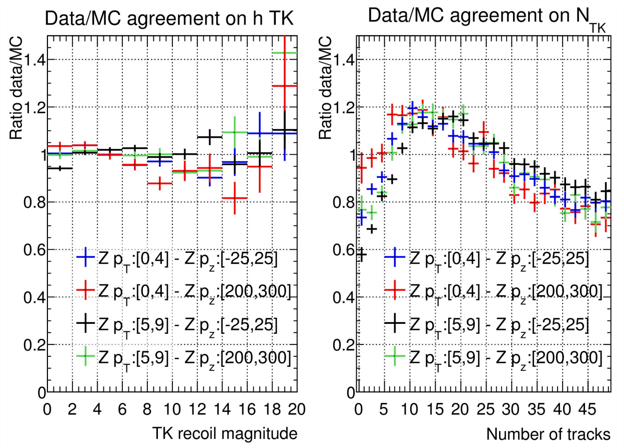
More variables

=

Larger discrepancy

=

Larger systematic uncertainty



Discrepancies data/MC are studied on the Z events

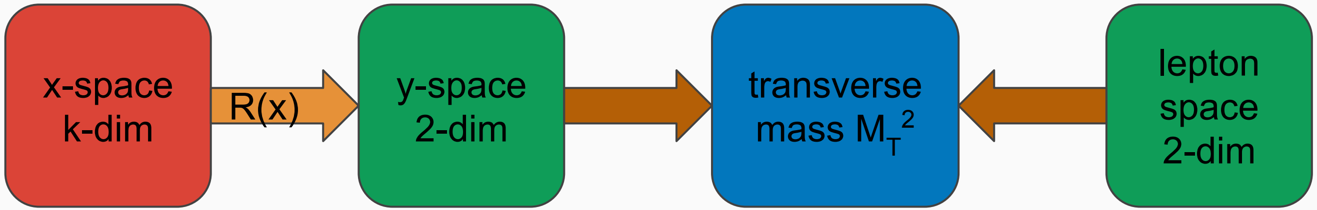
Make the MC similar to the data

The x space can be arbitrarily large

- ➔ In the case of our regression: 13 dimensions
- ➔ I need to correct the variables in a correlate
- ➔ Correct a space with many dimensions (k) is hard

The **y** space is always bidimensional (p_T and Φ)

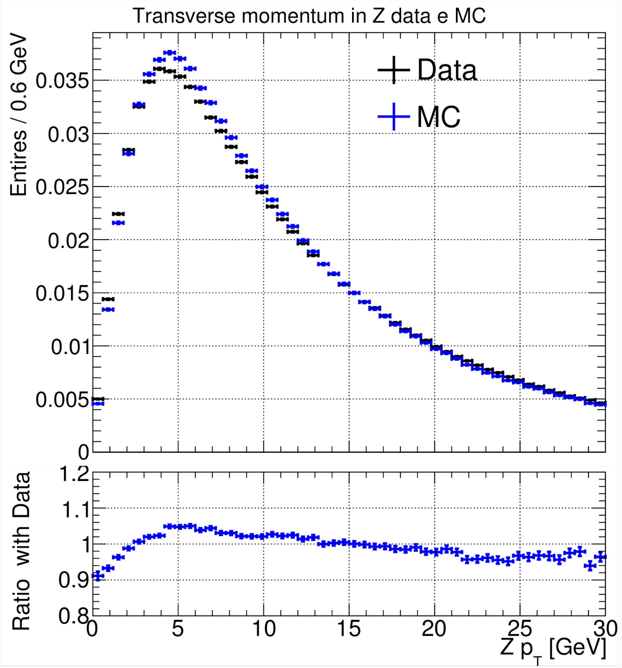
- ➔ I can do an effective correction on this space
- ➔ The correlation is extremely important

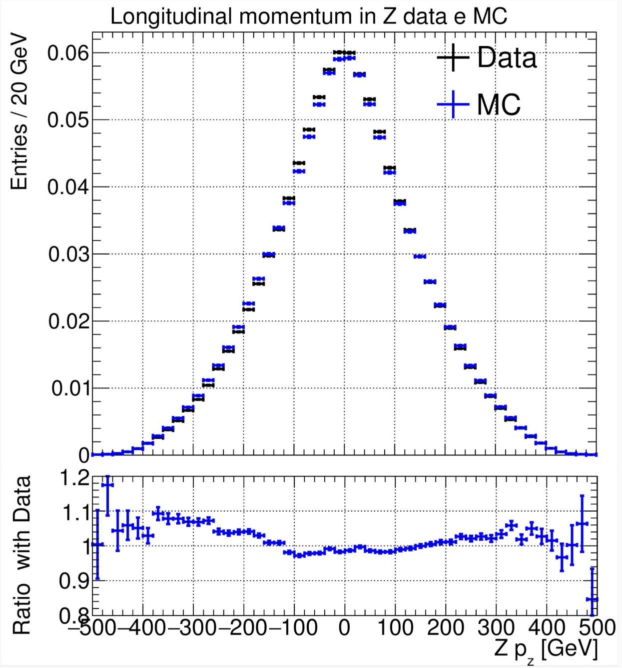


Correct f(y|z) instead of f(x|z)!

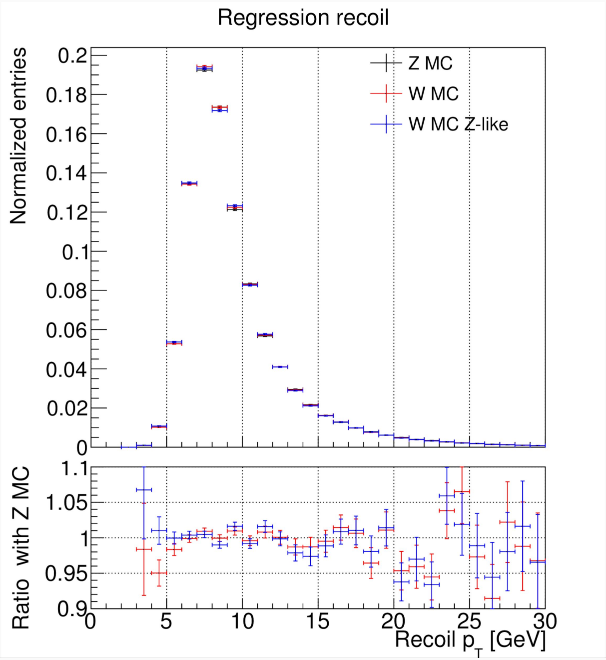
Z p_T and Z p_L

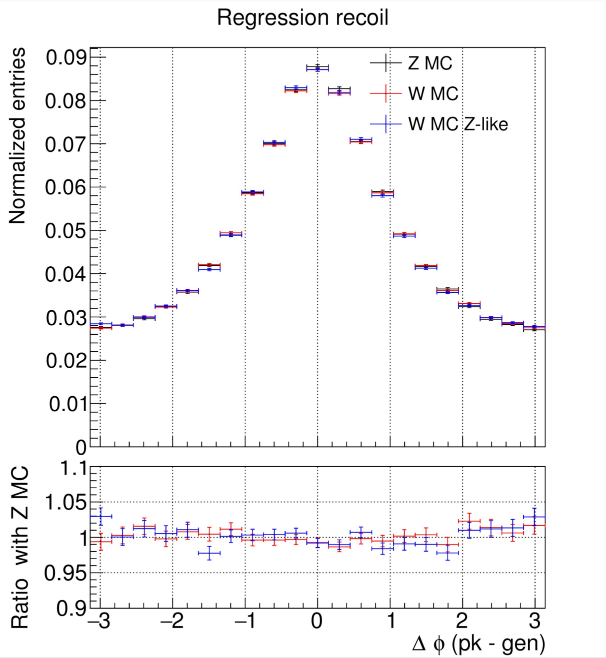
Subtlety: there are differences between data and MC on the Z spectrum





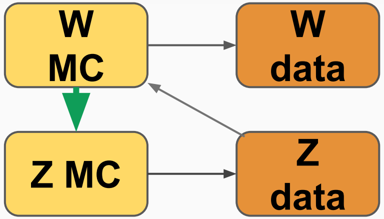
Z-like variables

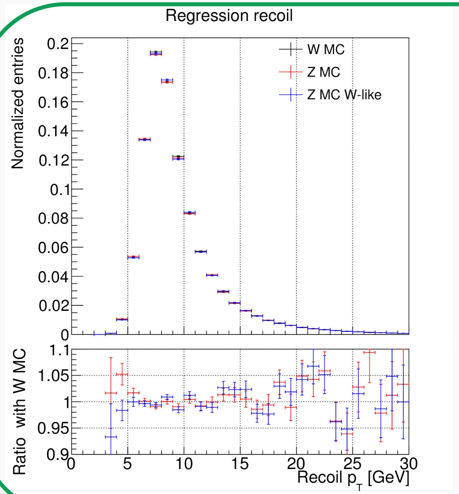


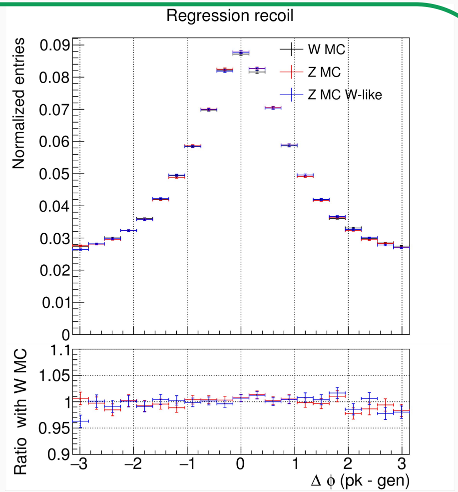


First step: apply the correction to the W - first problem

- ➢ Z and W should have no difference:
  - ○ The conservation of momentum is universal
- ➢ Build a morphing that transforms $f_Z(y|z)$ to $f_W(y|z)$
  - ○ If there is no difference the morphing is the identity







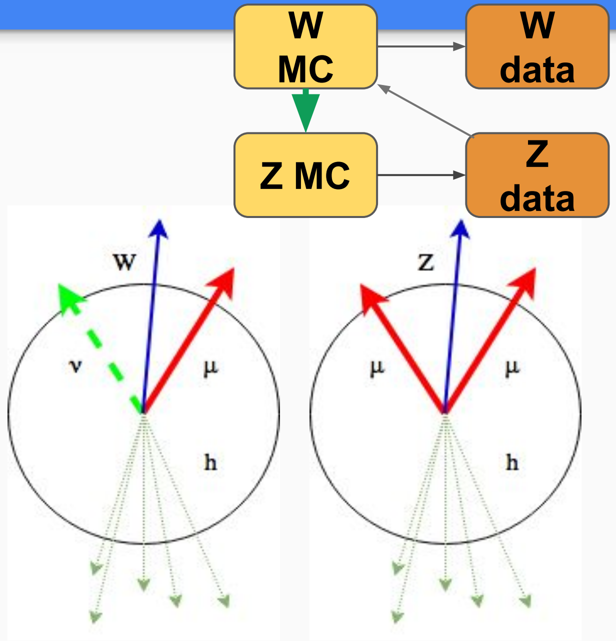
Bias

MNPK: 2 ± 12 MeV

No discrepancy

Bias agrees with 0

First step: apply the correction to the W - second problem

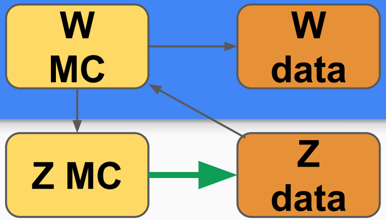


- In the morphing $Z_{MC} \longrightarrow Z_{DATA}$ $z = \{p_T^W, p_L^W\}$ reconstructed with the two leptons
  - In the W_{MC} I only have z "gen level"
- Test on Z_{MC}
  - Apply wrt true variables
  - Execute a fit to evaluate the bias

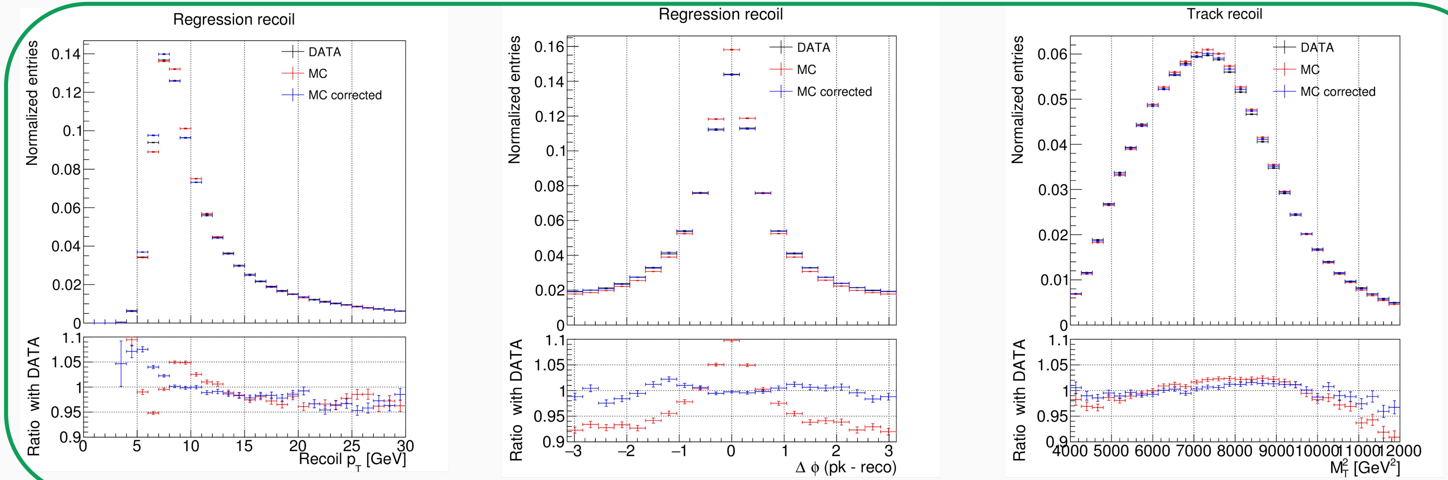
Bias
TK: -5 ± 7 MeV
MNPK: -12 ± 11 MeV

In agreement with 0!

Second step: uncertainty before correction



1) morph Z_{MC} to Z_{DATA}



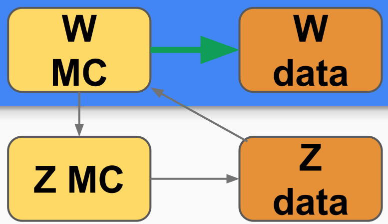
2) fit: extract the bias

TK: 29 ± 11 MeV

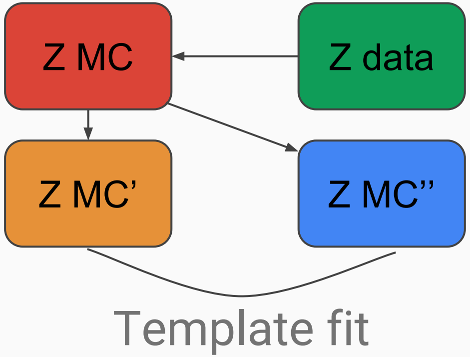
MNPK: 140 ± 14 MeV

Need to apply a correction

Third step: residual systematic uncertainties



- ➔ Do I have residual bias I can remove
- ➔ How much is the final systematic uncertainty?



Study on the Z:

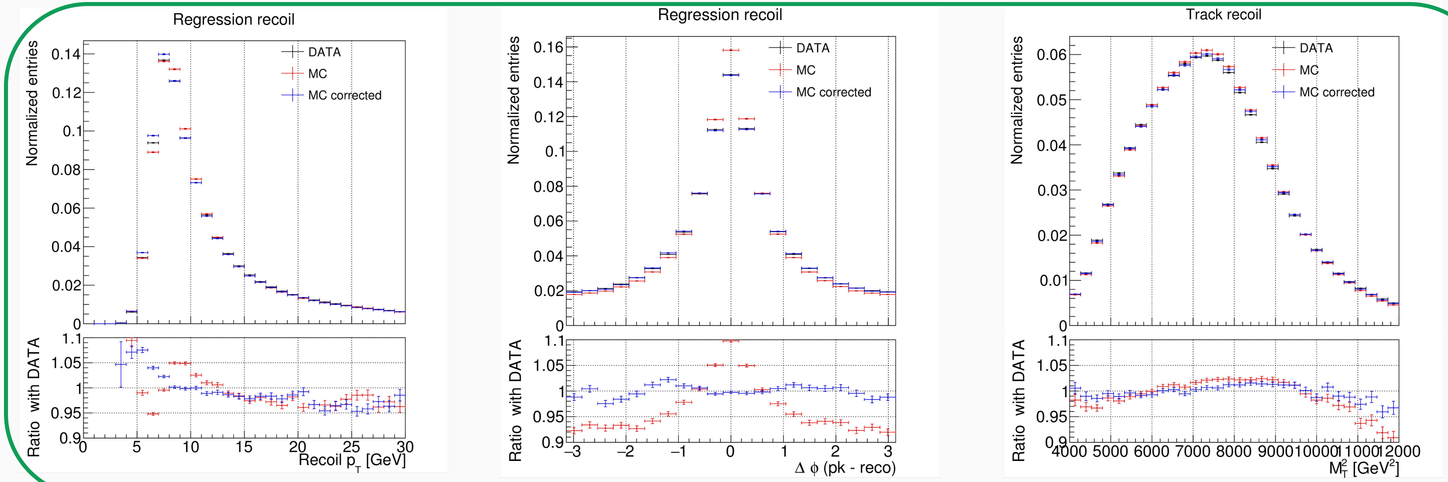
- A new morphing which transforms Z_{MC} to Z_{MC}'
  - Apply it to Z_{MC} and get Z_{MC}''
- Result fit:
  - TK -11 ± 10 MeV
  - MNPK -14 ± 7 MeV

Not in perfect agreement with 0, need for more investigation!

First step: variabili data-like

Make a morphing and apply on Z_{MC}, fit to estimate before applying the correction: need for a correction!!!

1. Build a morphing which transforms Z_{MC} to Z_{DATA}
   a. The variables z = {p_T^W, p_L^W} are obtained from the momenta of the two leptons
2. Apply to Z_{MC} and get the y "data-like", in the sample Z_{MC}'
3. Fit between Z_{MC} (template) and Z_{MC}' (pseudo-data)



Bias [MeV]

TK: 29 ± 11

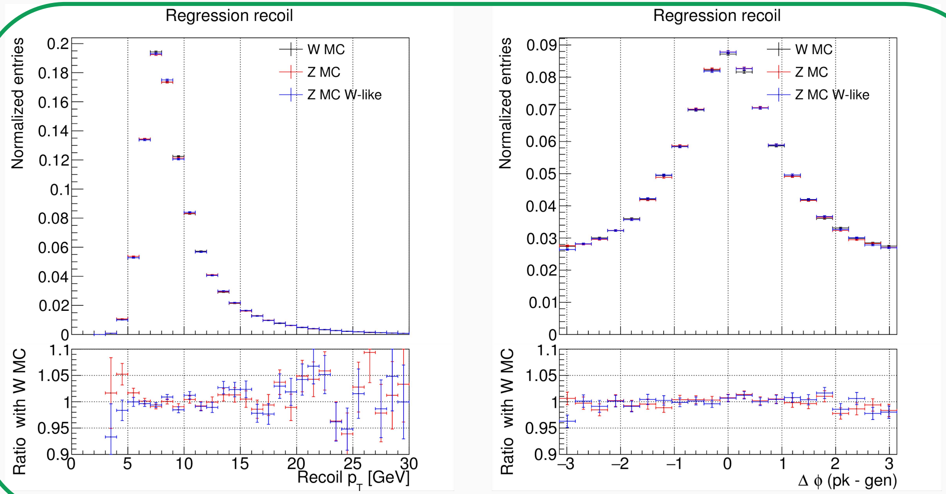
PF: -80 ± 24

MNPK: 140 ± 14

Need for a correction!

Second step: apply the correction to the W - problem 1

- The production mechanisms and masses of Z and W are different
  - It is not obvious that $f_W(y|z) = f_Z(y|z)$
  - It make non-sense to apply a correction
- Build a morphing that transforms $f_Z(y|z)$ to $f_W(y|z)$
  - With no difference the morphing is the identity
  - Fit to estimate the effect



Bias

MNPK: 2 ± 12 MeV

No discrepancy

Bias agrees with 0

Second step: apply the correction to the W - problem 2

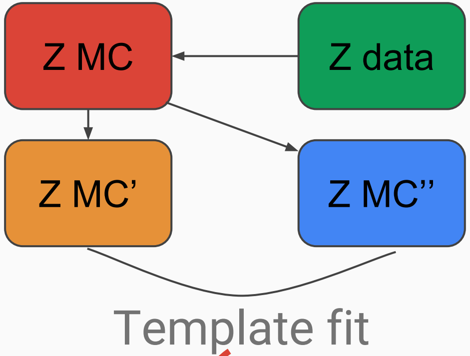
- In the morphing $Z_{MC} \longrightarrow Z_{DATA}$ I use $z = \{p_T^W, p_L^W\}$ reconstructed with thw two leptons
  - Present in both samples
  - No similar variables in the W_{MC}
    - I can use gen level variables such as $z = \{p_T^W, p_L^W\}$
- Test on Z_{MC}
  - On this sample I have $\{p_T^W, p_L^W\}$ both true and reco with the leptons
  - Apply morphing wrt to gen level rather than reco variables
  - Fit to estimate the bias on M_W
- No agreement with 0
  - I can smear the W kinematics at gen level to reproduce the same effect

Bias
TK: -5 ± 7 MeV
MNPK: -12 ± 11 MeV

→

For both definitions I have agreement with 0, I can proceed

Third step: residual systematics



I can apply the morphing on the W_{MC}

- ➔ Residual systematics I can correct for?
- ➔ If not, size of systematic uncertainty

Study on the Z:

- Build a new morphing that transforms Z_{MC} to Z_{MC}'
  - Apply to Z_{MC} and get Z_{MC}''
- If the morphing works well $Z_{MC}'' = Z_{MC}'$, then $Z_{MC}' = Z_{DATA}$
- Evaluate with a fit
  - TK -11 ± 10 MeV
  - MNPK -14 ± 7 MeV

Not in perfect agreement, need for more investigation

Further check: compare Z_{MC}' with Z_{DATA}

- ➢ A new morphing between the two samples -> Z_{MC}'''
- ➢ Study the discrepancy between Z_{MC}' and Z_{MC}'''