

# CMS Data Analysis

## Constraining Top Mass Using Differential Cross Section Measurements from $m_{t\bar{t}}$ and $pT_{top}$

Gevy Cao

Queen's University, Kingston, ON, CANADA

Supervisor: James Keaveney

September 6, 2017



### Abstract

The field of particle physics is entering the era of precision measurements, with the LHC running more powerful than before at the centre of mass energy 13 TeV. In this report, I present the techniques used to constrain the mass of the heaviest elementary particle – the top quark. With the differential cross section measurements of  $m_{t\bar{t}}$  and  $pT_{top}$  using the full 2016 data set taken at the LHC, and the theory predictions at six top masses with next-to-leading-order (NLO) accuracy, the parameter  $\chi^2$  is calculated at each mass point and fitted with a parabolic function. The mass of the top quark is extracted at the minimum  $\chi^2$ , and the mass values at minimum  $\chi^2 + 1$  on either side of the curve constitute the uncertainty band on the top mass. Using the next-to-next-to-leading-order (NNLO) and NLO predictions at 173.3 GeV, a bin-by-bin kfactor is taken to be the ratio of the cross sections of NNLO to NLO at each bin. The kfactor is then applied to all the theory mass points to obtain the approximated NNLO predictions. The  $\chi^2$  curve is calculated and fitted using the same procedure as in NLO predictions. With the data uncertainties, theory uncertainties and statistical correlations between bins in data folded into a covariance matrix, the top mass extraction yields  $172.5 \pm 0.9$  GeV from  $pT_{top}$  measurements and  $171.5 \pm 0.7$  GeV from  $m_{t\bar{t}}$  measurements. The only set of measurements needed to complete the analysis is the systematic correlations between bins in the data set, which have not arrived in time for the timeline of this project.

## **Acknowledgement**

I would like to express my utmost gratitude towards my supervisor James Keaveney for his helpful discussions and guidance, Dr. Oleksandr Zenaiev and Dr. Katerina Lipka for their technical support and expertise in the topic, the summer school organizers for every effort they have made to provide such an incredible experience and the international office staff for their help with my visa. Lastly, I would like to extend my thanks to my supervisor in Canada, Tony Noble, for inspiring me into the field of particle physics and his continuous support towards my career choices.

# I. Introduction

The top quark is known to be the heaviest elementary particle in the standard model (SM) of particle physics, raising many curiosities on its properties, especially its large mass. With the most recent data taken at the LHC at  $\sqrt{s}=13\text{ TeV}$  throughout the year 2016, the mass of the top quark can be determined more precisely than before. Precision measurements on the mass of the top quark can provide insight to the stability of SM. If  $m_t \leq 171.22$  or  $m_t \geq 177$  GeV, then the universe is unstable by the current theory of SM, probing the margin for the search for new physics[1].

At the LHC, top quarks are predominantly produced in  $t\bar{t}$  pairs through strong interactions. Measurements are made in different decay channels of the W boson, including the dilepton channel, lepton+jet channel and the all-jets channel. The theory calculations are performed to LO, NLO, NNLO and NNLL accuracy.

Measurements for the analysis of the top quark are performed on jets<sup>1</sup> and unfolded to the observables of the top quarks. In Section II of this report, I summarize the cross section measurements on the observables,  $m_{t\bar{t}}$  and  $p_{T_{\text{top}}}$ , and theory predictions used in the analysis. Section III demonstrates the analysis techniques to extract the top pole mass from the data described in Section II. The results of the analysis are summarized in Section IV, followed by discussion and conclusion in Section V. Finally, plots that can be used as a comparison to the results in Section IV are shown in Section VI.

# II. Data Set and Theory Predictions

The 1D normalised and absolute differential cross sections at parton and particle level using full 2016 dataset collected by a colleague, Mykola Savitskyi, contains differential cross section measurements and associated uncertainties. In this data set, the absolute parton-level measurements of  $m_{t\bar{t}}$  and  $p_{T_{\text{top}}}$  are used to extract the top mass. Corresponding theory predictions with the exact same binning at six top masses – 168, 170, 172, 173.3, 174 and 176 GeV, at NLO accuracy generated using MCFM by Oleksandr Zenaiev are used to extract top mass.

The data measurements are represented in Tables 1 and 2.

*Table 1: cross section data measurements for  $m_{t\bar{t}}$*

300	380	1.596350	1.931580	9.534570
380	470	3.371460	1.218940	9.094710
470	620	1.683000	1.150370	6.910700
620	820	0.530573	1.728320	6.256120
820	1100	0.123834	2.885360	7.419680
1100	1550	0.020963	4.079740	7.163170
1550	2500	0.001673	11.541500	20.179700

---

1 A narrow cone of hadrons, can be easily measured and tracked by the hadron calorimeter of a detector.

Table 2: cross section data measurements for  $pT_{top}$

mtmin[GeV]	mtmax[GeV]	ds/dmt [pb/GeV]	stat[%]	syst[%]
0	65	3.42498	1.13311	7.92430
65	125	5.12114	0.979177	7.86025
125	200	2.80269	0.914268	6.92297
200	290	0.849737	1.23811	6.05947
290	400	0.189046	1.55142	6.32172
400	550	0.0334429	3.89697	7.99319

NNLO predictions are not available in all six masses. Therefore, predictions at NNLO accuracy at the mass 173.3 GeV by Alexander Mitov[2] are used to extract kfactors. The bin-by-bin ratio of NNLO/NLO cross section at top mass 173.3 GeV is used as kfactor and applied to the NLO predictions of all other masses to obtain the estimated NNLO predictions, using Equation 2. The NNLO predictions have the same binning with the NLO and data measurements for  $pT_{top}$ , and similar binning for  $mt\bar{t}$  (with the absence of the last bin). In the case of  $mt\bar{t}$ , a weighted average of the cross sections from two different binnings are applied to the first bin using Equation 1, and the discrepancy between the NLO centre value ( $mt = 1350$  GeV) and NNLO centre value ( $mt = 1325$  GeV) of the second last bin was ignored. The bins in between agree perfectly.

Equation 1: weighted average cross section  $mt\bar{t}$  bin1

$$xsec(mt_{340}) = 0.75/0.8 * xsec(mt_{300-375}) + 0.05/0.8 * xsec(mt_{340-380})$$

Equation 2: Approximate NNLO cross section for each bin

$$xsec_{NNLO,i} = xsec_{NLO,i} \times kfactor_i, \text{ for bin } i$$

The data measurements and NLO predictions are plotted in the upper graph of Figures 1 & 2, and their ratio is plotted in the lower graph.

# Data & Predictions NLO pT

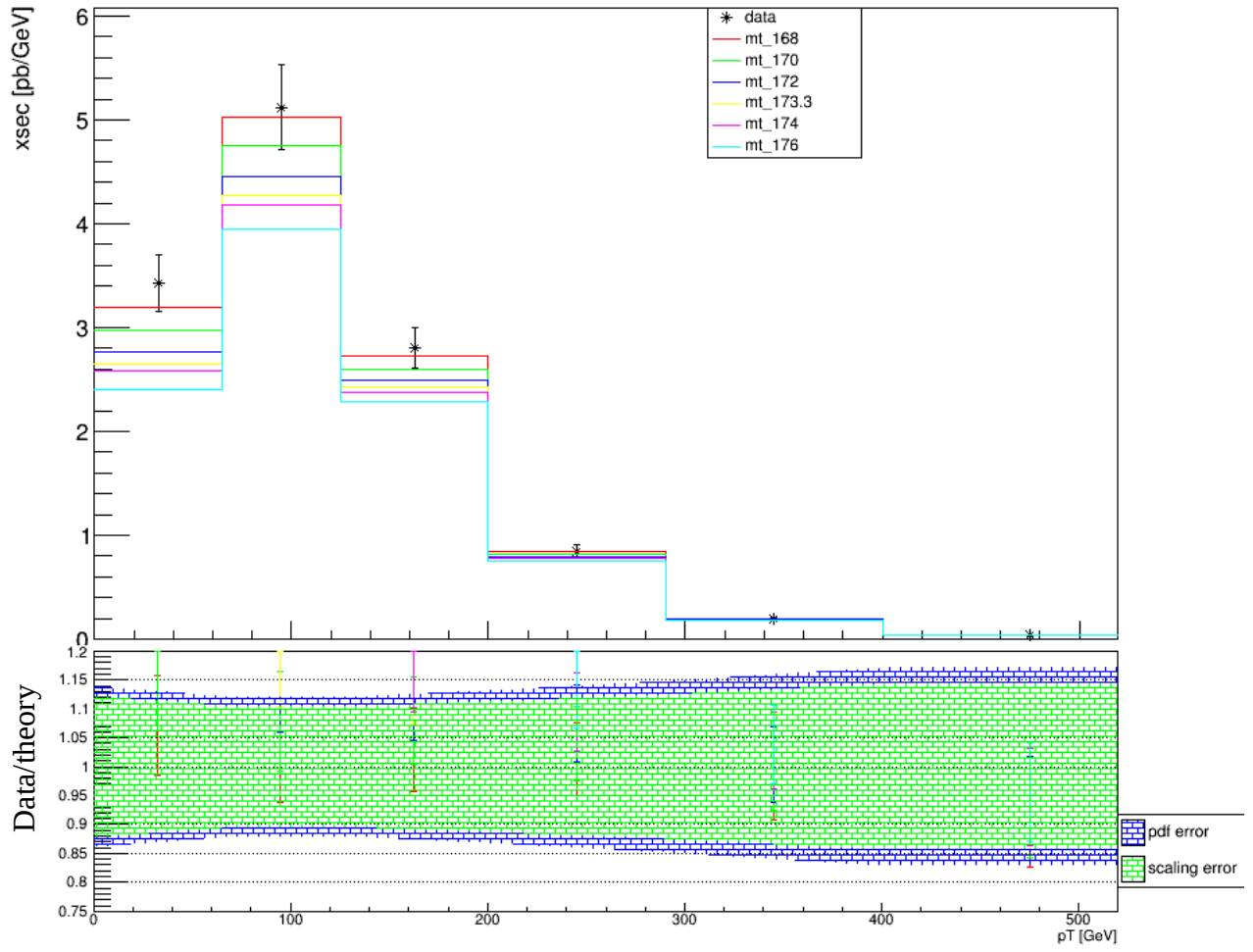


Figure 1: data & NLO predictions for  $pT_{top}$

## Data & Predictions NLO mtt

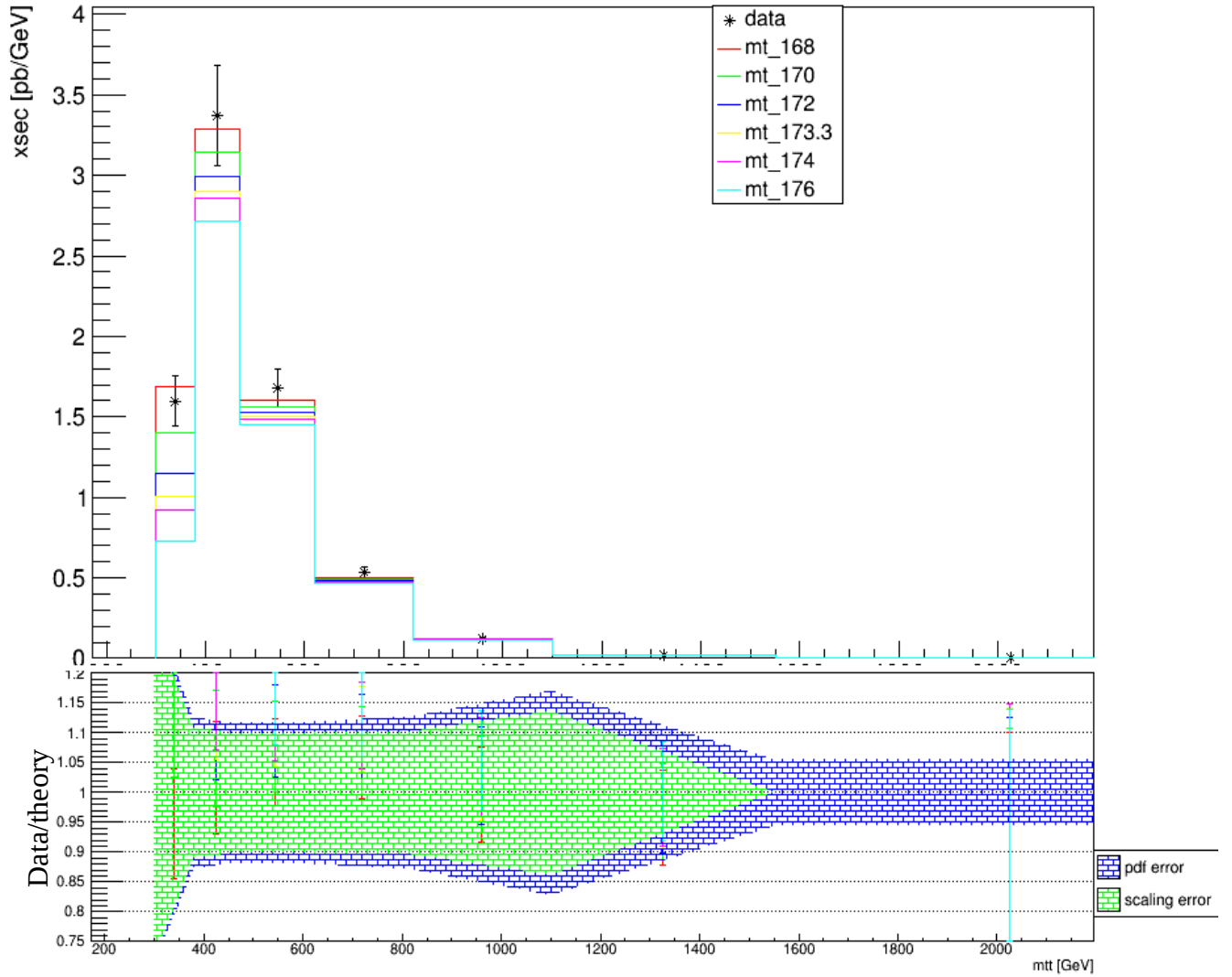


Figure 2: data & NLO predictions for  $mt\bar{t}$

It is important to note that for  $pT_{top}$ , the  $k$ factors are exact since there're no discrepancy on binnings. For  $mt\bar{t}$ , the  $k$ factors between bins 2 and 5 are exact;  $k$ factors for bins 1 and 6 are approximated and the NNLO prediction for bin 7 is missing. The scale uncertainties for both NLO and NNLO predictions in  $mt\bar{t}$  are missing as well due to the absence of available predictions on that bin. However, the effects of the last bin on the analysis are minor.

The data measurements and NLO theory predictions with  $k$ factor applied to them are plotted in Figures 3 & 4.

# Data & Predictions after $k$ pT

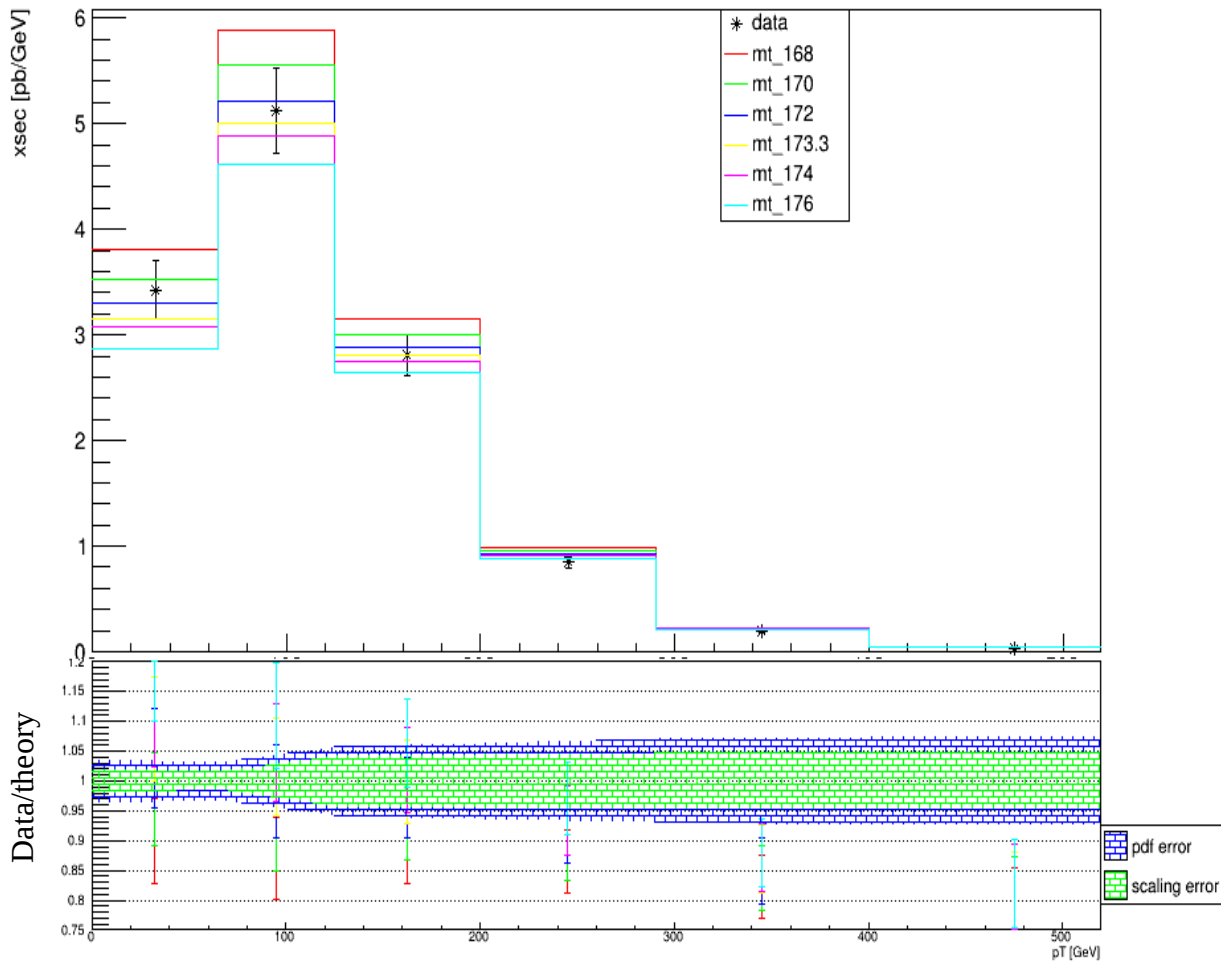


Figure 3: data measurements & NLO predictions with  $k$ factors applied for  $pT_{top}$

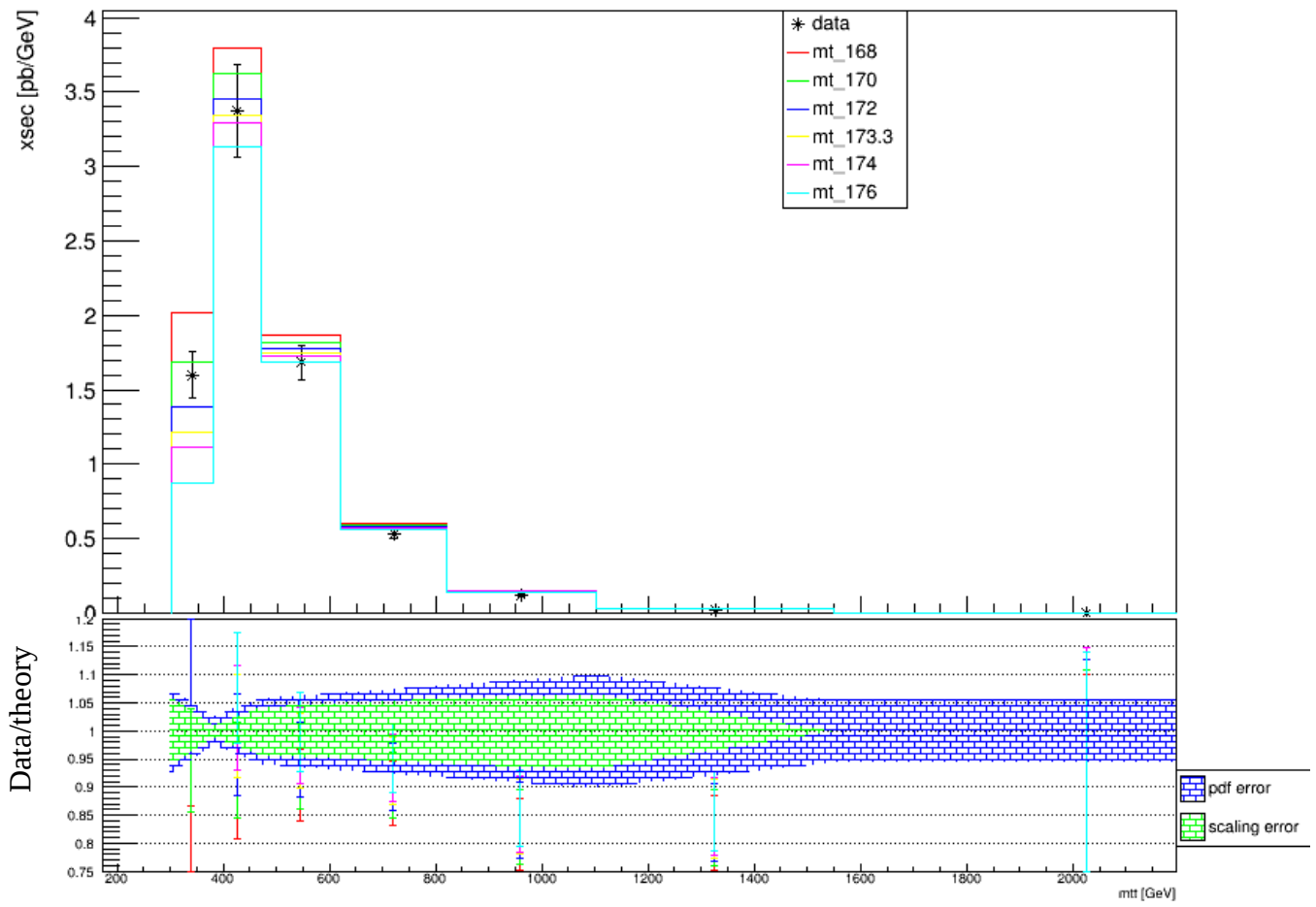


Figure 4: data measurements & NLO predictions with  $k$ factor applied for  $mt\bar{t}$

### III. Analysis

The analysis exploits two methods to cross check the results. One is to manually calculate the parameter  $\chi^2$ , defined in Equation 3. As a first step, the uncertainties on the data and theory are added in quadrature to constitute the uncertainty on total cross sections. However, in reality, the bins in the data measurements are correlated. The correlation can be incorporated into the covariance matrix defined in Equation 4.

Equation 3: chi2 definition in manual extraction for each theory mass  $m_k$

$$\chi_{m_k}^2 = \sum_{i,j=\text{bins}} (data_i - theory_{i,m_k}) * \sigma_{tot,ij}^{-1} * (data_j - theory_j)$$

$data_i / theory_{i,m_k}$ : cross sections corresponding to data in bin  $i$  and theory for mass  $m_k$  in bin  $i$ .

Equation 4: Covariance matrix elements

$$\begin{aligned} \sigma_{tot,ij} &= \rho_{ij} \sigma_{dat,i} \sigma_{dat,j}, \text{ for } i \neq j \\ \sigma_{tot,ij} &= \rho_{ij} \sigma_{dat,i} \sigma_{dat,j} + \sigma_{theo,i}^2, \text{ for } i = j \\ \sigma_{theo} &= \sqrt{(\sigma_{pdf}^2 + \sigma_{scale}^2)} \end{aligned}$$

$\sigma_{\text{pdf}}$  = uncertainties from the parton density functions (pdfs) evaluated by xfitter

$\sigma_{\text{scale}}$  = uncertainties by varying the scale of the experiment from Mitov

$\sigma_{\text{dat}}$  = uncertainties on the data

$\rho_{ij}$  = statistical correlation coefficient between bin  $i$  and  $j$ . By definition,  $\rho_{ij} = 1$  for  $i = j$ .

$\chi^2$  is calculated at each theory mass point using the data set in Tables 1 & 2, and fitted with a parabolic curve. The mass of the top quark is extracted using Equation 5.

*Equation 5: top quark mass extraction from chi2*

$$mt = mt(\chi_{\min}^2)$$

$$\sigma_{mt} = mt(\chi_{\min}^2 + 1) - mt$$

The other method is to use xfitter to evaluate  $\chi^2$ . The data in Tables 1 & 2, and theory predictions by Oleksandr Zenaiev are provided as inputs to xfitter, while specifying the order of the predictions (NLO or NNLO). The advantage of xfitter is that it alters the pdf parameters after the pdfs have been fitted to data. As a result, the theory predictions are modified. In the manual extraction, no action is taken on the set of pdfs. However, xfitter does not yet take the covariance matrix into account. Therefore, the two methods yield slightly different results.

## IV. Results

The  $\chi^2$  curves using NLO predictions are shown in Figures 5, 6 & 7. The curves fit poorly, making extraction of top mass difficult due to the lack of mass points to the left of the minimum  $\chi^2$  in the manual extraction method. The extremely poor fit from mtt by xfitter could be caused by the inaccurate kfactor in the first bin and the incorrect set of pdfs.

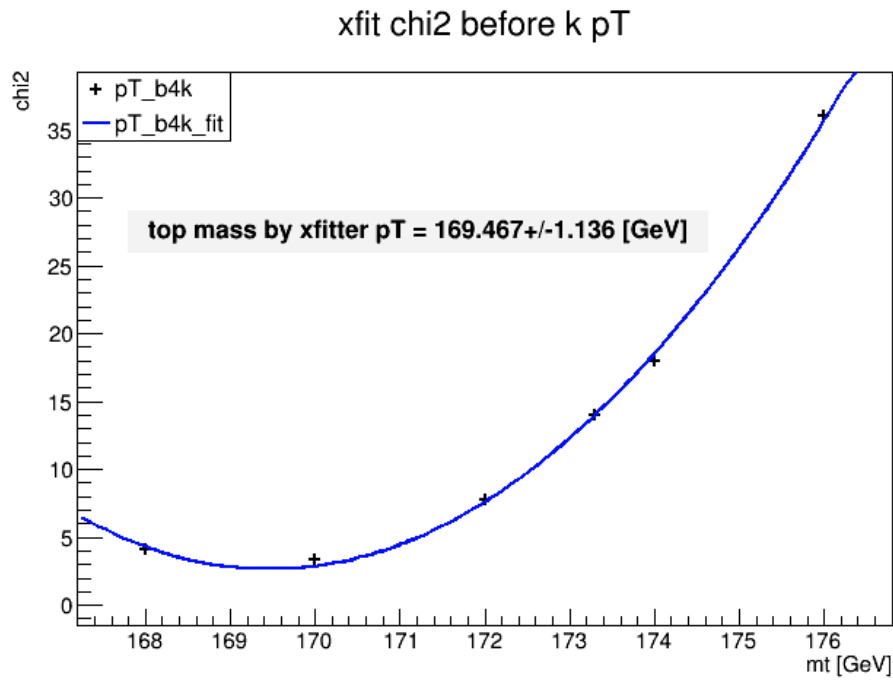


Figure 5: chi2 curve before  $k$ factor for  $pT_{top}$  measurements by xfitter

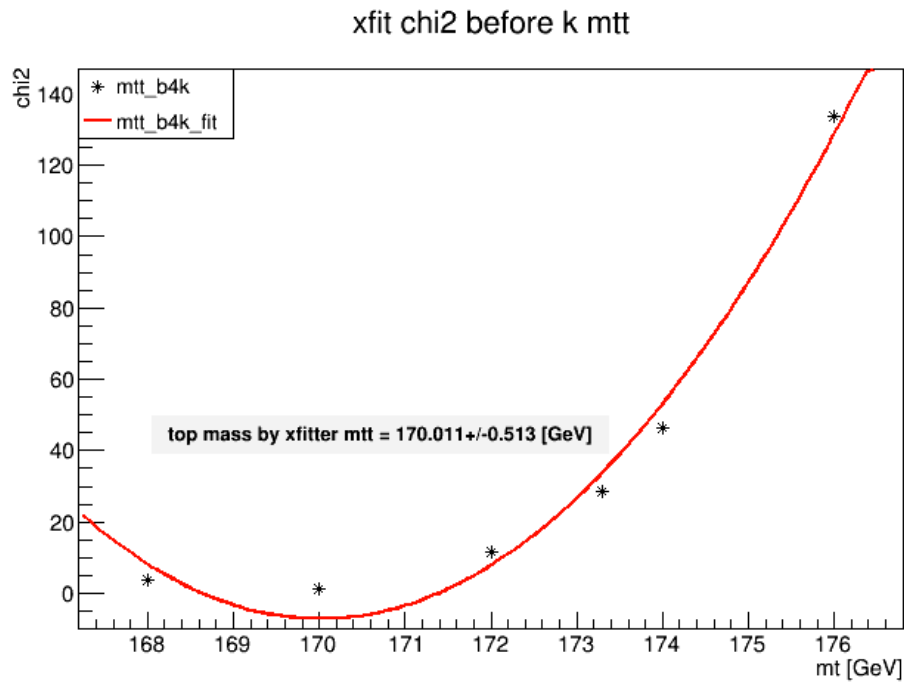
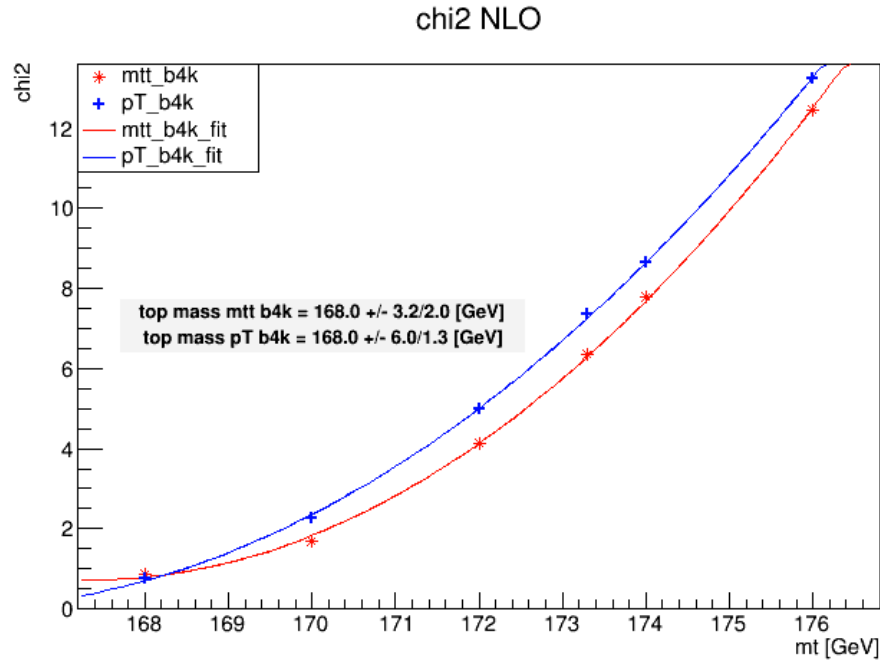


Figure 6: chi2 curve before  $k$ factor for  $mtt$  measurements by xfitter



*Figure 7: chi2 curve before kfactor in manual method*

After kfactor is applied, the calculated  $\chi^2$  and the parabolic fits from the manual extraction, along with the extracted top mass, are plotted in Figure 8.

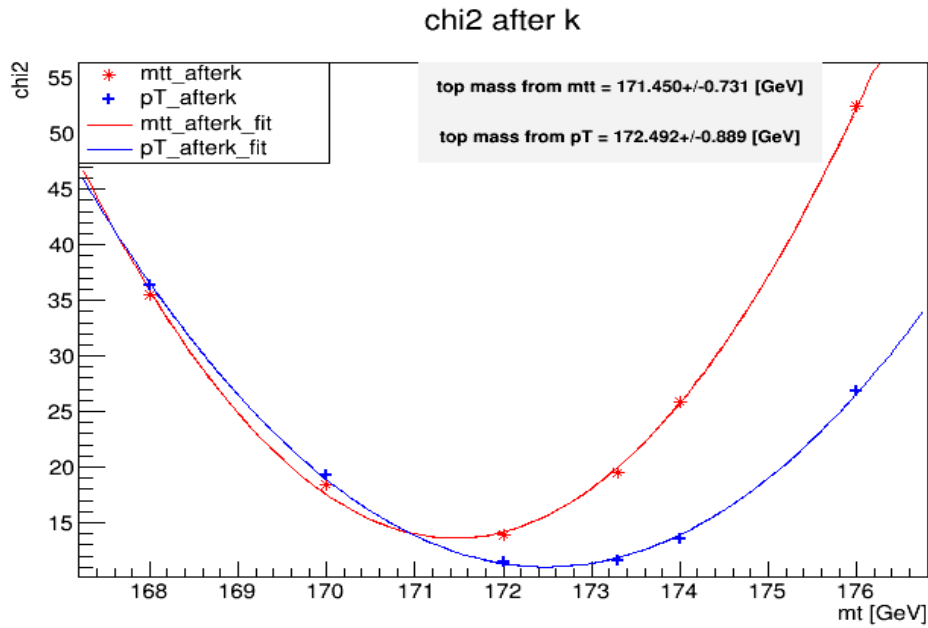


Figure 8: Manual calculation of  $\chi^2$  after  $k$ factor has been applied to NLO predictions

To complete the analysis, one needs to add the systematic correlation coefficients to the covariance matrix.

Xfitter only takes the data and pdf uncertainties into account. The scale uncertainties were externalized by manually adding the uncertainty on predictions and repeating the fits, and the correlations between bins have not been incorporated. The results from xfitter are plotted in Figures 9 & 10.

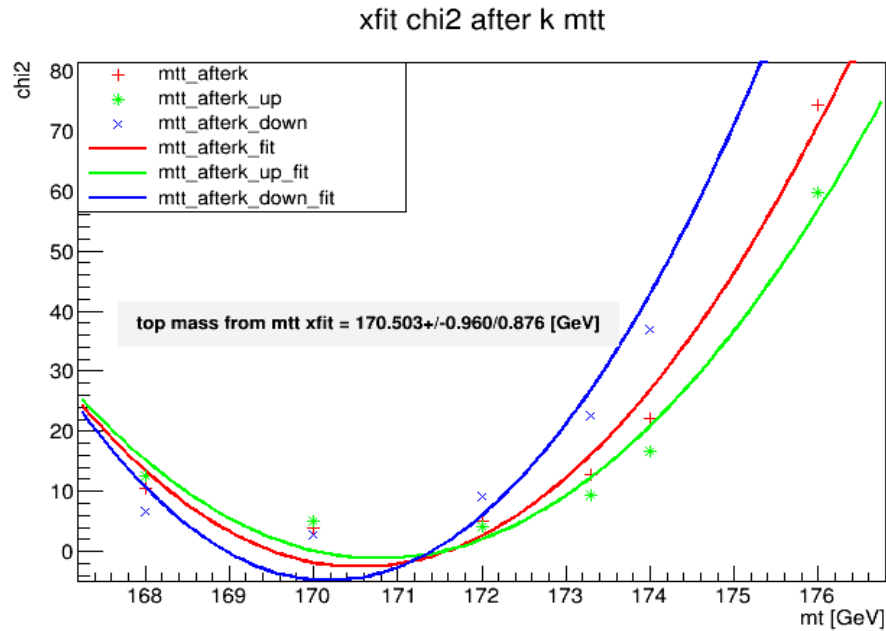


Figure 9:  $\chi^2$  calculation of  $\chi^2$  for  $m_{tt}$  after  $k$ factor has been applied to NLO predictions

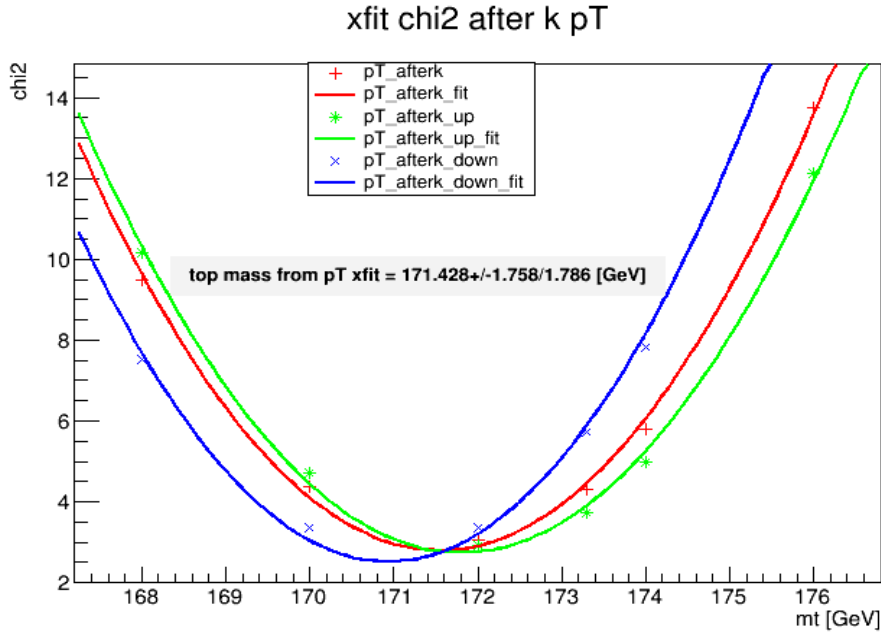


Figure 10: *xfitter* calculation of  $\chi^2$  for  $pT_{top}$  after *kfactor* has been applied to NLO predictions

As a direct comparison to *xfitter* results, the manual extraction without using the covariance matrix (the uncertainties are therefore added in quadrature) is plotted in Figure 11.

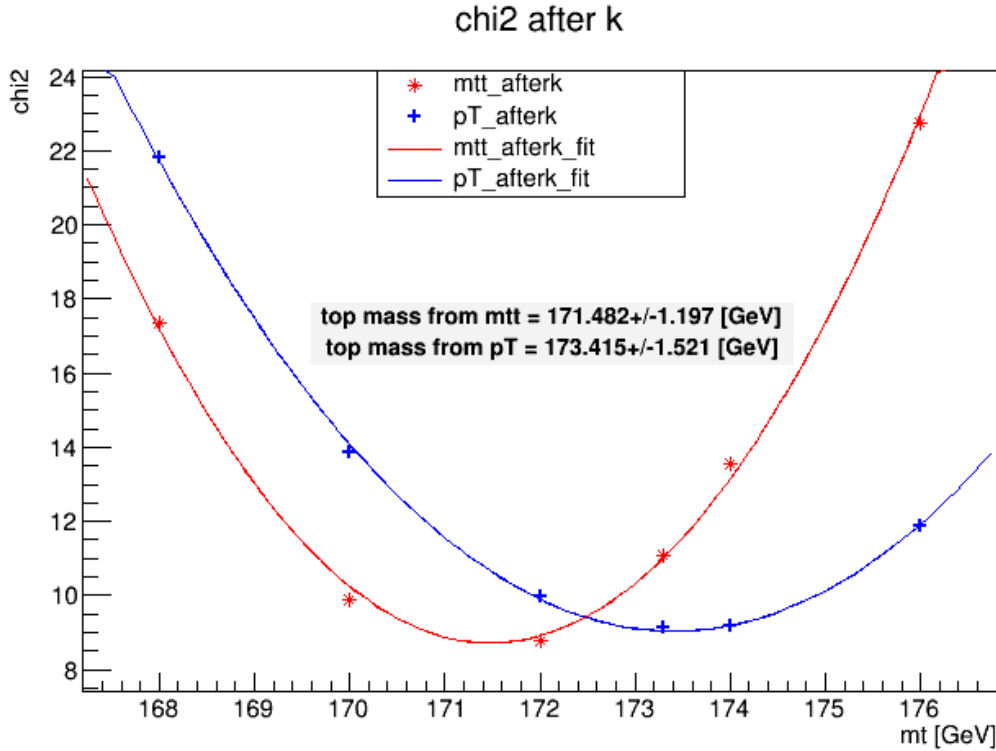


Figure 11: Manual calculation of  $\chi^2$  by adding uncertainties in quadrature

A summary of all the results shown is presented in Table 3. The last column shows a list of items needed to complete the analysis at NNLO accuracy. The results from NLO predictions are imprecise as expected.

Table 3: Top mass extraction results summary

Order	method/mt by:	$pT_{top}$ [GeV]	$m_{t\bar{t}}$ [GeV]	Needed for Completion
NNLO	<b>xfitter</b>	171.4+/-1.8	170.5+/-1.0	-NNLO pdfs for afterk calculations -Covariance Matrix
NNLO	<b>manual (quadrature)</b>	173.4+/-1.5	171.5+/-1.2	-Covariance Matrix
NNLO	<b>manual (covariance)</b>	172.5+/-0.9	171.5+/-0.7	- Systematic correlations to be added to covariance matrix
NLO	<b>xfitter</b>	169.5+/-1.1	170.0+/-0.5	
NLO	<b>manual (quadrature)</b>	168.0+/-6.0/1.3	168.0+/-3.2/2.0	

## V. Discussion and Conclusion

In summary, the top mass extracted from the  $pT_{\text{top}}$  measurements is more significant than that from the  $mtt$  measurements, because  $pT_{\text{top}}$  has perfect binning agreement between NLO and NNLO predictions. One can improve the analysis by applying the  $k$ factors separately to each mass point, though it would require NNLO predictions with the same binning at each of the six masses, which will take significant amount of time. The worthiness of producing these extra NNLO predictions needs to be discussed.

The xfitter results differ from the manual results in the fact that xfitter alters the pdf parameters, which in turn modifies the theory predictions.

In the current analysis, the pdfs used are at NLO accuracy. For the  $\chi^2$  after applying  $k$ factors, NNLO pdfs should be used, as the cross sections with  $k$ factor are an approximation to NNLO predictions. Once the pdf set is corrected for NLO predictions after  $k$ factor, the next step would be to input the full covariance matrix to xfitter. For the manual extraction, no action is needed because the method does not modify the pdf parameters. The last step necessary to complete analysis for the manual method is to incorporate systematic correlations to the covariance matrix. Then the results can be compared in a meaningful manner.

## VI. Appendix

The  $\chi^2$  is also calculated for the inclusive cross section measurements, shown in Figure 12. The inclusive cross section at NNLO can be computed easily in theory predictions. Once the differential cross section analysis is complete, the results can be compared to the inclusive cross section result.

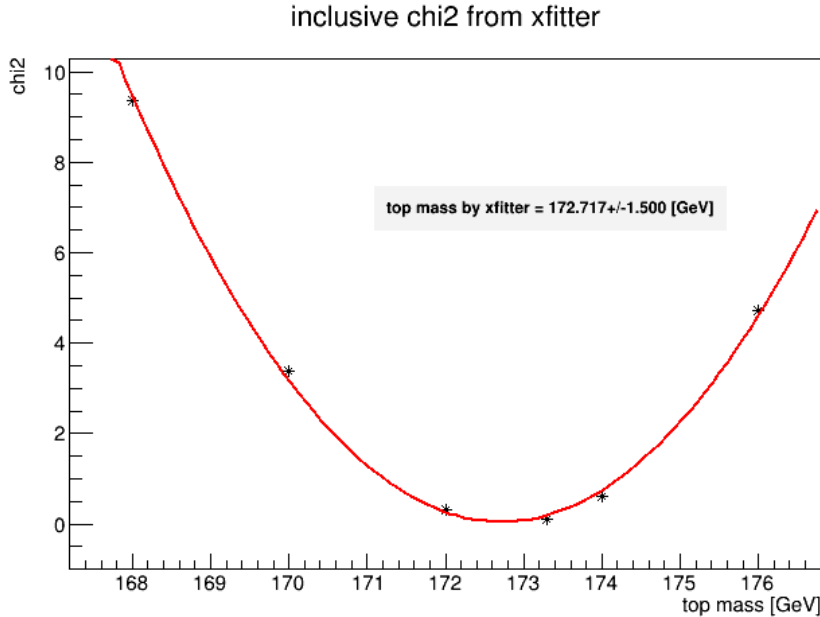


Figure 12:  $\chi^2$  curve for inclusive cross section by xfitter

## **Bibliography**

- 1: Anders Andreass, Matthew D.Schwartz, Reducing the Top Quark Mass Uncertainty with Jet Grooming, 2017
- 2: Czakon Michal, Heymes David, Mitov Alexander, Dynamical scales for multi-TeV top-pair production at the LHC, 2016