Utrecht University

Institute for Subatomic Physics

Bachelor thesis

Systematic uncertainties in the measurements of D^{*+} mesons in 7 TeV pp collisions at ALICE

Author:
Arthur Brussee
4152174

Supervisors:
Dr. Andre Mischke
Dr. Alessandro Grelli

January 2017

Abstract

Multiple studies into the quark-gluon plasma (QGP) state are underway, using the ALICE detector at the Large Hadron Collider. The QGP is important as it is understood to be the state the universe was in for a few milliseconds after the big bang. This thesis outlines why the D^{*+}meson plays a role in QGP research and describes a possible disagreement between measurement and predictions made by perturbative Quantum Chromodynamics.

The main contribution of this thesis is an analysis of the systematic uncertainties in D^{*+}yields from 7 TeV *pp* collisions measured in the ALICE detector. The systematic uncertainty is analyzed by varying a range of parameters when extracting the raw D^{*+}yield from the data, and observing the variation in this yield. The results show an assigned systematic in agreement with the statistical variation in the data. Additionally, a 'pure' systematic that is not dependant on statistical variations in the data is discussed. The presented method works entirely automatically and is applicable to other data sets at higher energies.

Contents

Chapter 1

Introduction

The Standard Model of particle physics describes the, as we currently understand, fundamental particles and forces. In the Standard Model the strong force keeps the quarks of a hadron bound together and is the base for the study of Quantum Chromodynamics (QCD). QCD is a quantum field theory describing the interactions between quarks and gluons, the mediators of the strong force. The theory introduces an $SU(3)$ gauge symmetry for each quark called the color charge. Gluons can transfer color charges between quarks. The potential energy for two quarks in this field is modeled as:

$$V(r) = -\frac{4}{3}\frac{\alpha_S(r)\hbar c}{r} + \kappa r$$

given $\alpha_S(r)$ the strong coupling constant, κ the string tension and r the distance between the two quarks. A notable property of the QCD potential is that it increases with r, unlike gravity and electrostatic potentials that decrease with r. This leads to 'quark confinement'. As quarks get further apart the potential energy rises until a quark-antiquark pair is created, as this lowers the total energy of the system. Thus quark pairs can never be pulled too far apart without creating new pairs. Quarks have never been observed on their own. Additionally, all hadrons are required to be colorless: baryons contain three differently colored quarks, mesons contain two partons with opposite color charges. A quark-gluon plasma is an exotic state of matter, predicted by QCD to exist at extremely high temperatures and densities, consisting of free quarks and gluons. Figure 1.1 shows a sketch of a transition from a hadronised state to a QGP state. In the hadron gas, the particles are confined by gluons to stay within their hadrons, whereas in a QGP the quarks are deconfined and can freely move around in the whole plasma.

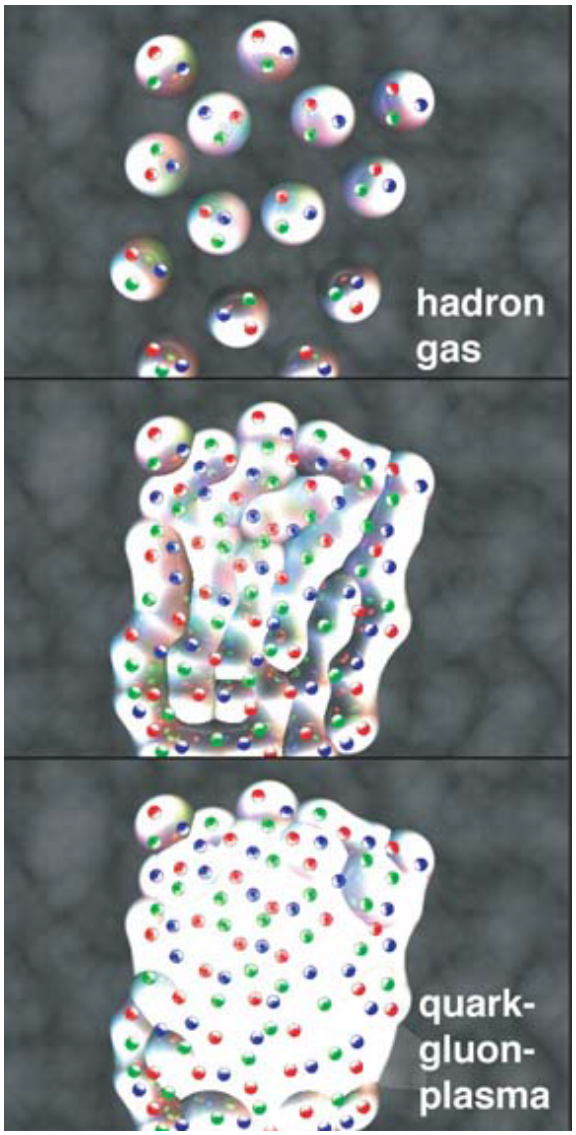

Figure 1.1: Transistion from a hadron gas to a QGP as predicted by QCD [10].

Current cosmological theories predict that the universe was in a QGP state for up to a few milliseconds after the big bang (figure 1.2).

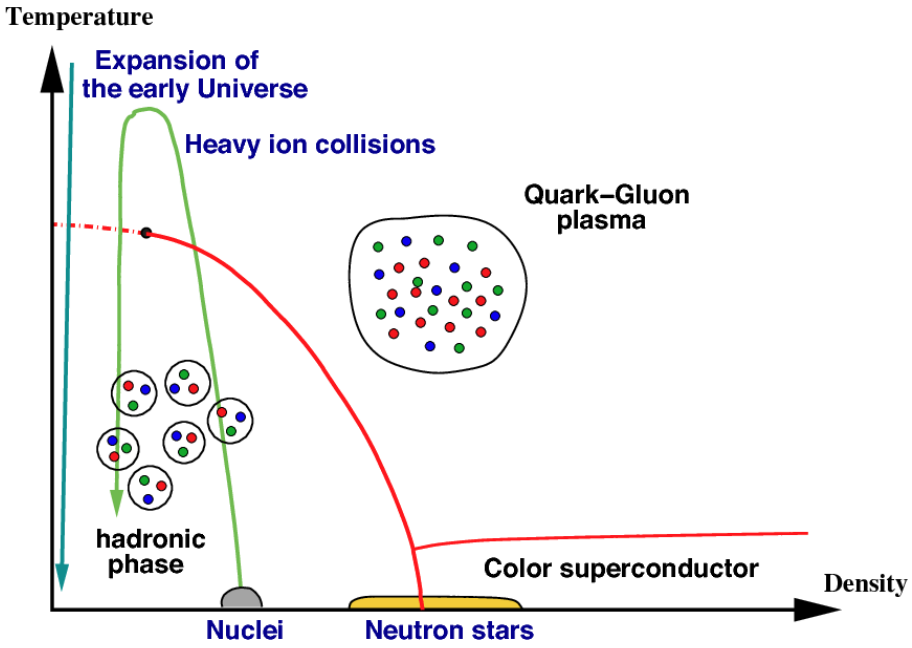

Figure 1.2: Quark phase diagram [8]. The blue arrow shows the state of the universe as it expanded over time. The green arrow shows the state of a nucleus as it collides and hadronises.

In June 2015, a research group at the LHC measured a QGP forming after colliding protons with lead nuclei [3] at high energies (5 TeV). The collisions created the temperatures needed to produce a QGP - upwards of 5.5 trillion degrees. The studies of the QGP have shown that the QGP flows like an ideal Fermi liquid, rather than flowing like a gas of quarks like initially was thought.

1.1 Probing the QGP with heavy quark-mesons

The plasmas produced by these collisions have a radius of only a couple femtometer and decay incredibly fast. Measurements therefore instead probe a QGP with particles produced in the collision. To find good candidate probes a few effects have to be considered [6]:

- Formation time: The probes need to form before the QGP forms and interact with it. Heavy quarks form earlier after the collision so they have time to interact with the QGP. Charm quarks form after $\Delta t = \pm 0.01 - 0.02$ fm/c
- Thermal production: Alongside the quarks produced by the collision, quarks can be produced thermally within the QGP, or from other decaying hadronic

matter. Both are less likely to occur for heavier quarks. Thermal production effects are not expected to significantly change the count of heavier (2nd generation) quarks.

- The 'dead cone' effect: High momenta particles radiate gluons (gluon Bremsstrahlung) and lose energy. However, radiation is suppressed for $\theta < \frac{m_q}{E_q}$ (with θ the angle between the particle and radiated gluon's momenta) - this is called the dead cone effect. A heavier particle, therefore, loses less energy by radiation.

- Mesons containing strange quarks are particularly enhanced by a QGP.

From these points, it can be seen that the family of D and B mesons are good probing candidates. The B_s^0 meson should lose less energy than the D^{*+}mesons and potentially be a better probe than the D^{*+}. However, the D^{*+}has a significantly lower rest mass than the B_s^0 (about 2.5 times less) and therefore significantly higher yields at the same energy level. The Institute for Subatomic Physics in Utrecht is however currently researching the feasibility of using B_s^0 mesons for probing.

1.2 Motivation

The nuclear modification factor (R_{AA}) describes the relation of D^{*+}yields in a collision with and a collision without a QGP forming. It is formulated as the ratio of cross-sectional D^{*+}yields (a measure of the yield independent from the detector used) in a *pp* collision with no QGP forming and the D^{*+}cross section in a *PbPb* collision with a QGP forming. The total production of D^{*+}mesons should scale with the number of nucleon-nucleon collisions ($\langle N_{coll} \rangle$) in a measurement, so R_{AA} is divided by the number of nucleon collisions to compensate. The R_{AA} is formulated as [9]:

$$R_{AA}(p_T) = \frac{d\sigma_{PbPb}/dp_T}{\langle N_{coll} \rangle d\sigma_{pp}/dp_T}$$

A R_{AA} of 1 implies the medium has had no effect on the probe production. Lower values indicate interactions of the probes with the QGP. Intuitively this ratio should be a function of $\langle N_{coll} \rangle$: For a head-on lead-lead collision, a QGP would be produced and the ratio would be lower than 1. If only 1 proton of the lead atom smashed into another proton however, the collision was essentially the same as a *pp* collision and the ratio should approach 1.

The functional dependence of the nuclear modification factor on $\langle N_{coll} \rangle$ can be calculated using a contemporary perturbative quantum chromodynamic theory called 'Fixed Order Next to Leading Log' (FONLL) [2]. However, all measurements from the LHC and other colliders have been at the upper limit of the theoretical predictions. Figure 1.3 shows the measured D^{*+}cross sections and cross sections predicted by FONLL:

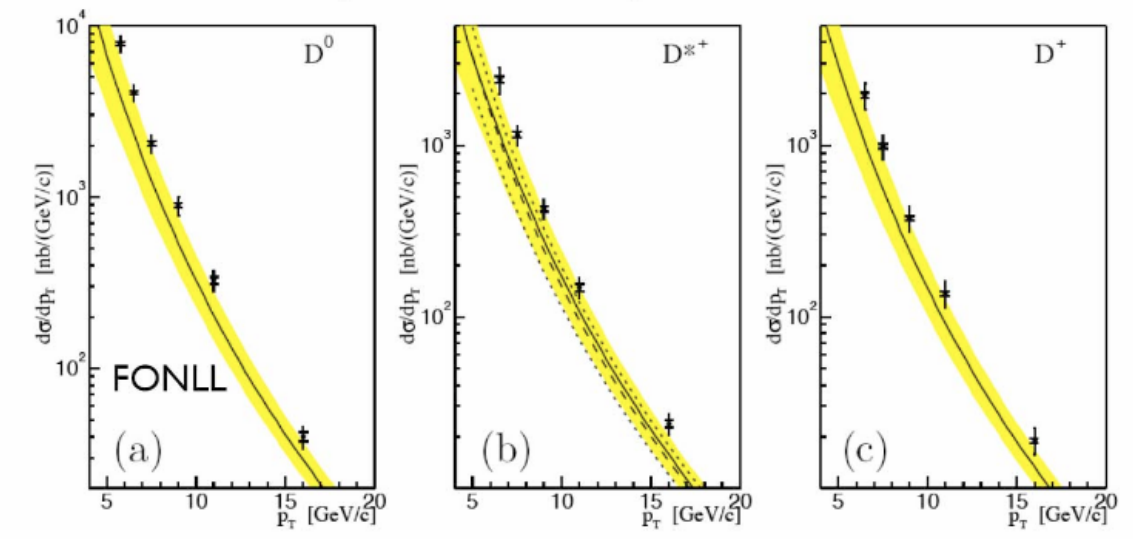

Figure 1.3: Measured cross sections and cross sections predicted by FONLL [5].

These results could still be a statistical anomaly. More precise measurements at 7 TeV and new measurements at 13 TeV will help to determine whether FONLL is indeed not matching up with measurements.

Chapter 2

Experimental setup

2.1 The ALICE detector

The LHC is a particle accelerator in Zrich run by CERN, capable of colliding particles with very high energies up to 13 TeV. These energies are needed to create a QGP and the D^{*+}probes. The D^{*+}probes decay too fast to measure directly, so instead the D^{*+}mesons are reconstructed by measuring its decay products. In this thesis the following decay path is used:

$$D^{*+} \rightarrow D^0 + \pi^+ \rightarrow K^- + \pi^+ + \pi^+$$

This decay path has a high enough yield (Branching ratio of 11.1% [11]) and can be reconstructed well. The data in this thesis is measured in the ALICE (A Large Ion Collider Experiment) detector, one of the four detectors part of the main LHC ring. ALICE was designed to study the physics of strongly interacting matter at extreme energy densities and includes many detectors to reconstruct and study the results of a collision. The main barrel of ALICE weighs 10.000 tons, is 26m long, 16m high and buried under 56m of rock.

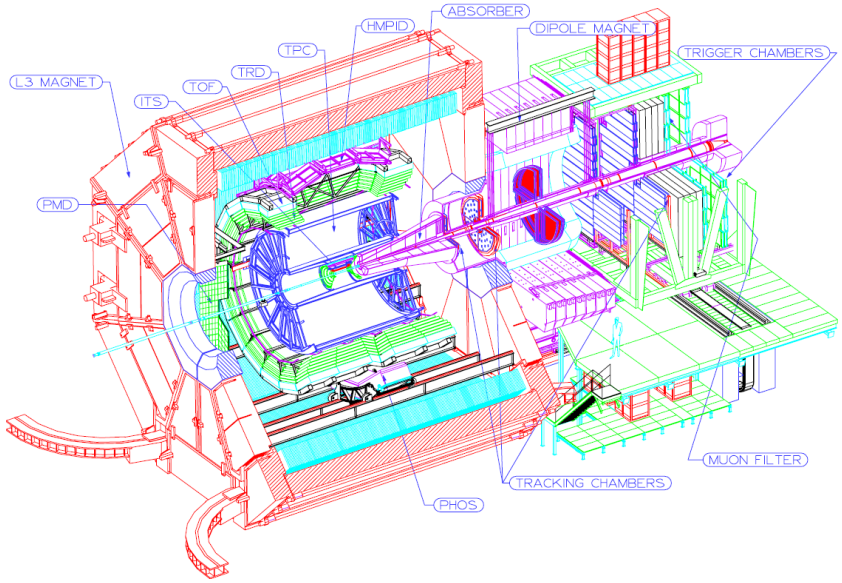

Figure 2.1: Overview of the main components in the ALICE detector [4].

Figure 2.1 shows the 6 main detectors inside ALICE, starting from the inside of the barrel these are:

- ITS: Inner Tracking System
- TPC: Time Projection Chamber
- TRD: Transition Radiation Detector
- TOF: Time of Flight detector
- PHOS: Photon Spectrometer
- HMPID: High Momentum Particle Identification Detector

For this thesis, 3 detectors are mainly important

- *ITS*: The inner tracking system is composed of six layers: Two Silicon Pixel Detectors (SPD), two middle Silicon Drift Detectors (SDD), and the two outer layers are Silicon Strip Detectors (SSD).

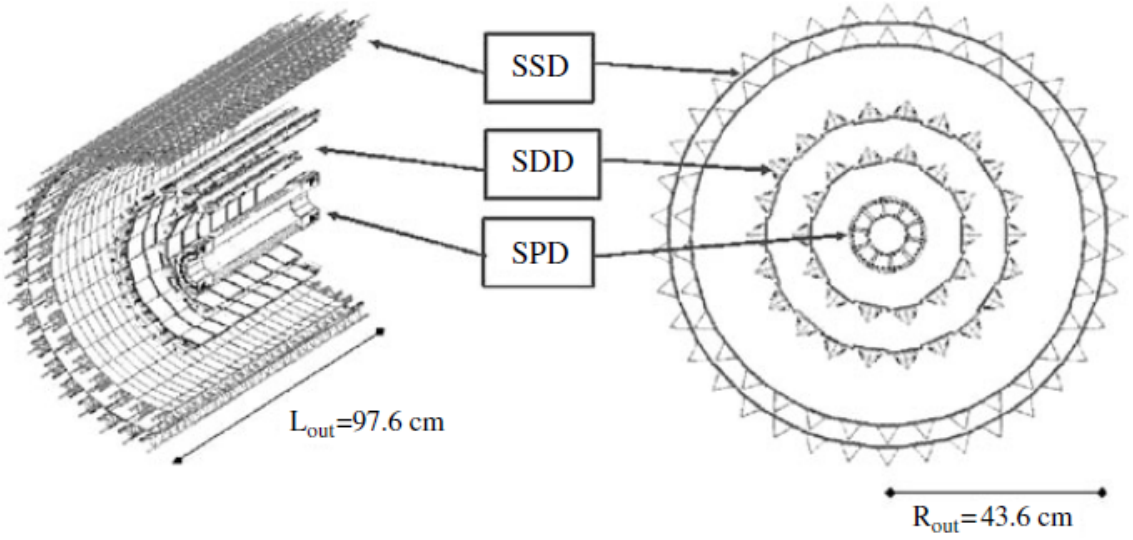

Figure 2.2: Overview of different silicon layers in the ITS [4].

These layers were set up so they provide a very high resolution near the beam pipe where the highest particles densities are expected. The ability to track particle showers moving through these layers makes the ITS important for reconstructing primary and secondary decay vertices (the positions of a particle decay, and positions of charged decay products crossing paths). It is the only detector in ALICE that can detect non-relativistic particles ($p_T < 100$ MeV). This detector can detect charged pions at $\pm 70 MeV/c$ with $> 70\%$ efficiencies. These pions are produced near the primary decay vertices in D^0 decays. Particle identification in the ITS is done by tracking energy losses ($\frac{dE}{dX}$) of a particle.

- *TPC*: The TPC is the main tracking device in the ALICE detector. It is a gas filled wire chamber. It can detect when and where a particle interacts with these wires. The TPC can detect charged hadrons and leptons with high precision. The high capacity tracking of the TPC is needed to measure secondary decay vertices. The TPC can track particles with momenta ranging from 160 MeV/c upwards to several GeV/c. Even at momenta as high as 100 GeV/c, it can still provide good resolution when combined with other tracking detectors. Like in the ITS, particle identification in the TPC is done by tracking energy losses ($\frac{dE}{dX}$) of a particle.

- *TOF*: The TOF is a detector filled with gas that interacts with electron avalanches created by charged particles passing it. By comparing delays between events produced in the gas the speed and total time of flight of the incoming charged particles can be determined. This primarily helps detect charged pions, kaons, and protons. The TOF tracks particles with momenta between 200 MeV/c to 2.5 GeV/c.

Chapter 3

Data analysis

3.1 Dataset

The data set used in this thesis was produced by $\sqrt{s} = 7$ TeV proton-proton collisions with 350 million events. It was taken from LHC pass 4 measured by the ALICE detector. The data set is split into 4 parts, LHC10b_pass4, LHC10c_pass4, LHC10d_pass4 and LHC10e_pass4.

3.2 Invariant mass fitting

As described in chapter two, the ALICE detector can detect the decay products from a D^{*+}meson decay. ALICE cannot directly measure how many D^{*+}mesons have been produced, however. Instead, all potential combinations of decay particles that form a D^{*+}meson together are considered candidate D^{*+}particles. All of the candidates have a certain invariant mass. For one particle this is the mass at rest, for multple particles the invariant mass is $m_{inv}^2 = (\Sigma_i E_i)^2 - (\Sigma_i p_i)^2$. The invariant mass is reconstructed from the momenta and invariant masses of the decay products forming the candidate. Figure 3.1 shows the number of candidates with a certain invariant mass per MeV. For a particular invariant mass the candidate mesons are correlated, showing a gaussian signal of 'real' D^{*+}particles on top of the combinatorial background.

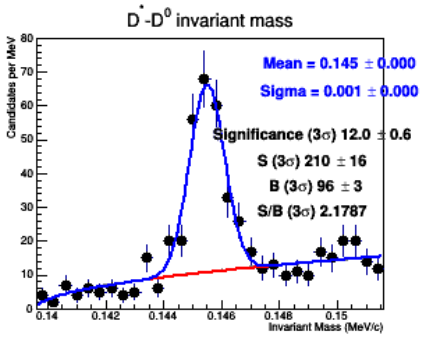

Figure 3.1: Example D^{*+}invariant mass histogram from a *pp* collision at 7 TeV.

Note that this invariant mass histogram is slightly different than a 'standard' distribution as the plotted invariant mass is the invariant mass of the D^{*+}minus the candidate D^0 mass. For D^{*+}reconstruction this produces a clearer signal.

The shape of the background can be modeled, in this thesis as a thresholded exponential function, and as mentioned the shape of the signal is expected to be a gaussian distribution. Knowing this a fit can be made for the distribution of candidates (shown as the blue line in figure 3.1, the red line shows the fitted background function). This fit extracts the parameters for the mean and σ of the gaussian signal.

Theoretically, in the reconstruction process all possible combinations of decay products that can form a D^{*+}should be considered candidates. In practice, however, the decay particles must have certain properties before it is likely they came from the same D^{*+}meson. Candidates with topologies that do not seem to match possible D^{*+}production can be discarded. So called cuts are made that exclude candidates from the invariant mass distribution based on the properties of the decay particles. Consider figure 3.2 showing a D^0 decay.

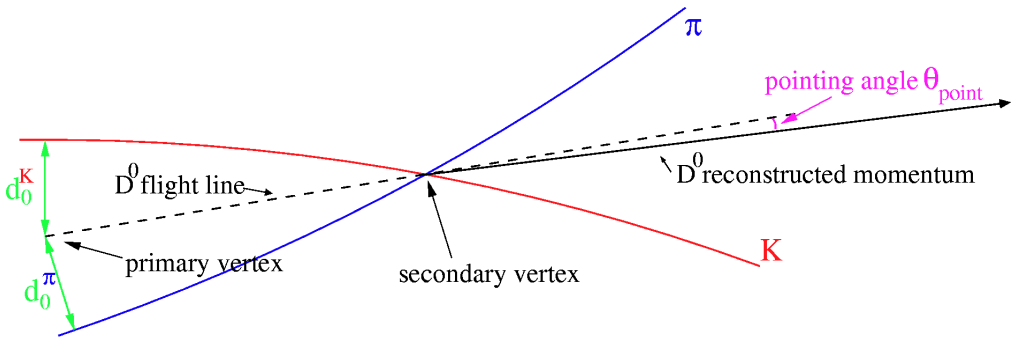

Figure 3.2: Primary and secondary vertices from a D^0 decay [6].

If the pointing angle θ_{point} is too large it can be assumed the pion and kaon were not created by the decay of the same D^0 meson. Another property that a cut can be made on is the distances shown in figure 3.2 $d_0^K \cdot d_0^\pi$. If this value is too large it again is unlikely the Kaon and Pion came from the same D^0.

The choice of these cuts to achieve the best signal to noise ratio is a complicated process in itself. Table 3.1 lists the cuts used throughout this thesis. These are the default for the data set that was used and work reasonable well for the entire data set [7] .

Table 3.1: Cut values for each p_T bin

Cut	$p_T \in [1,2]$	$p_T \in [2,3]$	$p_T \in [3,4]$	$p_T \in [4,5]$	$p_T \in [5,6]$	$p_T \in [6,7]$	$p_T \in [7,8]$	$p_T \in [8,10]$	$p_T \in [10,12]$	$p_T \in [12,16]$	$p_T \in [16,28]$
$cos(\theta_{point})$	0.80	0.82	0.90	0.90	0.80	0.70	0.70	0.75	0.80	0.75	0.70
$cos(\theta^*)$	0.90	0.90	0.80	0.80	0.90	1.00	1.00	1.00	1.00	1.00	1.00
$d_0^K \cdot d_0^\pi$ [cm*cm]	0.01	-1.5e-4	-1.9e-4	-1.6e-4	-5e-5	-1e-4	-1e-4	-1e-4	-3e-6	-3e-6	-1e-3
d_{ca} [cm]	0.050	0.035	0.030	0.030	0.080	0.100	0.100	0.100	0.0300	0.0280	0.100
d_0^K [cm]	0.10	0.10	0.10	0.10	0.07	0.10	0.10	0.10	0.10	0.10	0.15
d_0^π [cm]	0.10	0.10	0.10	0.10	0.07	0.10	0.10	0.10	0.10	0.10	0.15
p_T^K [GeV/c]	0.45	0.60	0.90	0.90	1.20	1.00	1.00	0.60	0.50	0.60	0.50
p_T^π [GeV/c]	0.45	0.60	0.90	0.90	1.20	1.00	1.00	0.60	0.50	0.60	0.50
$m_{K\pi} - m_{D0}$ [GeV/c]	0.30	0.30	0.30	0.30	0.10	0.10	0.10	0.30	0.15	0.15	0.30
min m_{inv} [GeV/c]	0.050	0.032	0.032	0.032	0.034	0.036	0.036	0.040	0.550	0.700	0.0700
min p_T of π_s [GeV/c]	0.03	0.05	0.10	0.10	0.05	0.05	0.05	0.30	0.40	0.45	0.05
max p_T of π_s [GeV/c]	0.10	0.20	0.30	0.50	100	100	100	0.80	1.10	1.50	100
$\theta_{\pi_s - D^0}$ [rad]	1.00	1.00	1.00	1.00	0.50	0.50	0.50	1.00	1.00	1.00	1.00

Besides topological cuts, all particle candidates are split into different ranges based on their transverse momentum: p_T bins. This is done because statistical variation in different bins can differ quite significantly and each bin can use different topological cuts.

For every p_T bin, a fit is constructed to extract a function for the gaussian signal. From this signal, the raw yield can be determined with methods described later. This yield is not the true amount of D^{*+}mesons produced in the collision, however. By applying cuts and through other losses in the reconstruction process, only a fraction of the D^{*+}particles is included in the measured yield. The fractions of particles not lost, so-called efficiencies, are determined by analyzing Monte Carlo simulations of the collision. With these efficiencies the normalized cross-sectional yield can be calculated, a measure of the particle production independent of the specifications of the detector used. Using this cross-sectional yield in both the *pp* collisions and *PbPb* collisions the nuclear modification factor R_{AA} can be measured and plotted against N_{coll} like described in chapter 1.

The methods to extract the raw D^{*+}yield from an invariant mass distribution leave some wiggling room on the particular details. In a perfect dataset the methods should produce a yield that is not dependent on the exact parameters used for fitting

and extracting the yield, but in practice, there are uncertainties introduced in this process. This thesis will present a measurement of these systematic uncertainties.

3.3 Parameter variations

To determine the systematic uncertainty, the parameters of the yield extraction are varied. After each variation, the yield is determined and compared to the yield measured at default parameters. This section will describe what sequences of variations are performed.

- **Range variation**: The reconstructed candidates have a range of invariant masses. However, the statistical variation of the measured background candidates gets increasingly worse approaching higher invariant masses. A cutoff is therefore chosen where there is 'enough' background for the fit to work well. This is an arbitrary choice, however. To measure the effects of this choice the cutoff is varied. The variation starts at 4 bins lower than the default cutoff up to 4 bins higher, in one bin increments. Similarly, there is a cutoff at the start of the invariant mass range. This range is similarly varied 4 bins lower to 4 higher. If these ranges were to be varied too much, the signal itself could accidentally be cut off leading to a vastly different yield. To measure just the systematic variation the range is ensured to be only varied in a small enough area so the cutoff does not get too close to the signal. The cutoff shouldn't change the overall shape of the invariant mass distribution.

- **Bin Counting**: From a fit for the gaussian signal in an invariant mass distribution an estimation of the D^{*+}yield can be made. One approach takes the yield as the integral of the gaussian signal function. Another approach is to count the number of candidates in bins that are within a few σ of the mean of the gaussian signal and subtract the number of background candidate particles. Both approaches have advantages and disadvantages. For the systematic variation, integration count and bin counting are both considered. The bin counting is done for $\sigma_{counted} \in [2, 4]$.

- **Rebinning**: To smooth out some statistical noise and to produce more reliable results, the measured invariant mass histograms are first rebinned before extracting the raw yield. Rebinning puts the candidates from a number of bins into a single bin. To measure effects of this rebinning on the yield, the

range and counting variations are performed both with 4-bin-rebinning and with 2-bin-rebinning.

In summary, 6 variation sequences are run with each sequence taking 9 samples, for a total of 54 measurements per p_T bin. The assigned systematic of a p_T bin is the RMS of these 54 variations.

3.4 Signal variation

None of the assigned systematics for a p_T bin are directly an estimate of the actual systematic uncertainty in that p_T bin. The differences in the yield caused by varying the parameters are heavily correlated with the statistical variation in the data. For example, consider some outlier point that happens to be included/excluded in the invariant mass distribution when varying the upper signal cutoff. Including/excluding this point increases/decreases the raw yield. Counting this variation as a systematic error, while counting it as a statistical error too, would effectively double the total uncertainty. Every assigned systematic error is essentially only an upper limit for the 'pure' systematic that is independent of the statistical variations.

To remedy this, all of the variations are performed on datasets with different numbers of events. As the number of events increases, and therefore have better statistics, the apparent systematic should asymptotically go down to some pure systematic.

In this thesis, as mentioned, the data used is split into four parts: B, C, D, E. The method above is applied to every permutation of these 4 passes: B, C, D, E, BC, BD, CD, ..., BCDE for a total of 15 permutations (in practice 14, as using only pass B did not contain enough entries for reliable results and was not included).

3.5 Method

To perform the analysis described, ROOT was used [1]. ROOT is a broad software analysis framework developed over years by CERN and other research groups. It includes frequently used processing functions, statistical methods, and graph drawing libraries and is programmed in C++. More specifically, AliRoot was used in this thesis, a superset of ROOT that includes functions and data specific to the ALICE detector. It includes routines to reconstruct meson decays, apply topological cuts, fit invariant mass functions and perform the final yield extraction. All of the

fits needed to measure the systematic uncertainty are performed automatically in a ROOT macro.

The highest and lowest p_T bins can sometimes contain too few events and fail to produce a good fit. These bad fits produce yields that fluctuate wildly. This could have a severe impact on the apparent systematic, while not really measuring any systematic error at all, but just measures statistical fluctuations. Therefore the fits with a significance below twp sigma are discarded and not taken into account when determining the systematic variation in any further calculations.

On rare occasions, a fit could not be found at all. This happens in cases where the statistics are just too bad for any fit to work. Failed fits are rejected as described above when calculating the assigned systematic. Losing these fits does slightly reduce the accuracy of the final systematics, but overall does not have a large impact on the results. It is a noteworthy artifact, however.

Chapter 4

Results

Before showing the obtained results, figure 4.1 illustrates what a variation sequence as described in chapter 3 looks like, in these plots the upper cutoff range is varied.

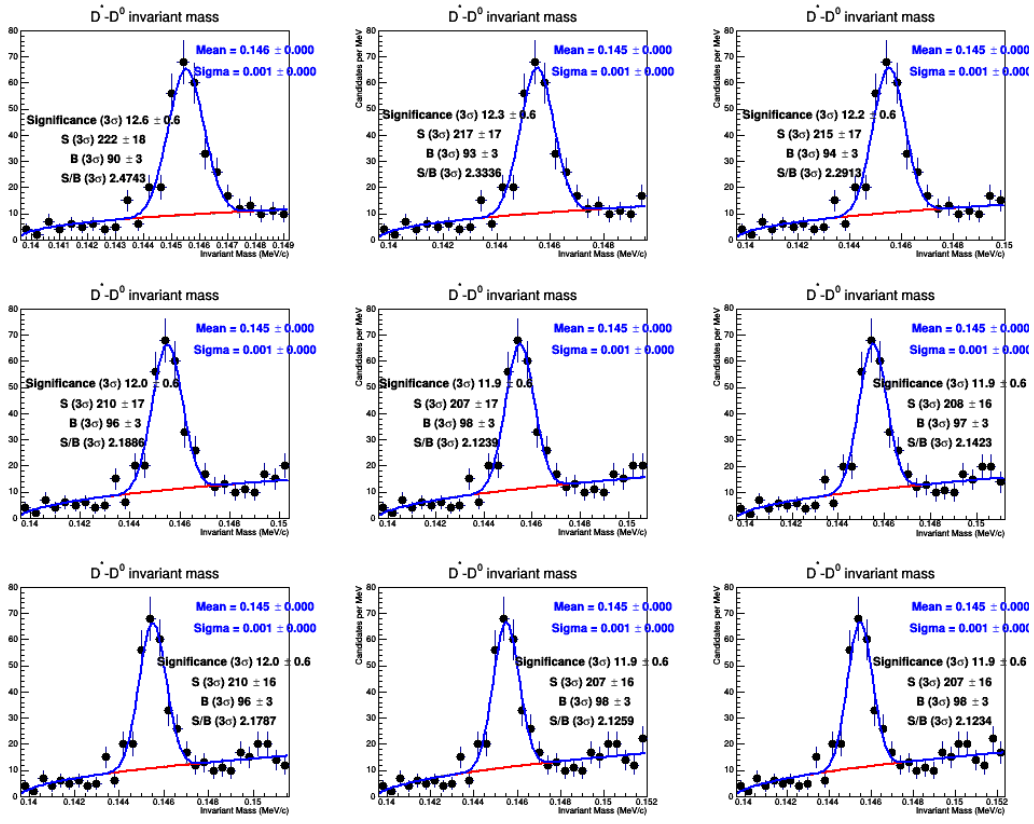

Figure 4.1: Variation sequence of the upper invariant mass cutoff. One bin is appended to the distribution in each plot. The data set used contains the full 350 million events. The distributions are filtered to show candidates with $p_T \in [6, 7]$ GeV/c. Note the yield fluctuations between plots of $\pm 5\%$ despite the similar appearance of the fit.

As described in chapter 3, in every data set 6 variation sequences taking 9 samples each are run, resulting in a total of 54 measurements per p_T bin. Figure 4.2 to 4.5 show the 54 variation measurements for every p_T bin. Each figure shows the 54 samples measured in a data set with a different number of events. Do note that like described in chapter three some points from the variation sequences can be missing if the fit failed or was too insignificant. The width of the gray band in each figure is the assigned systematic for a bin as measured by the RMS of all variations in a bin.

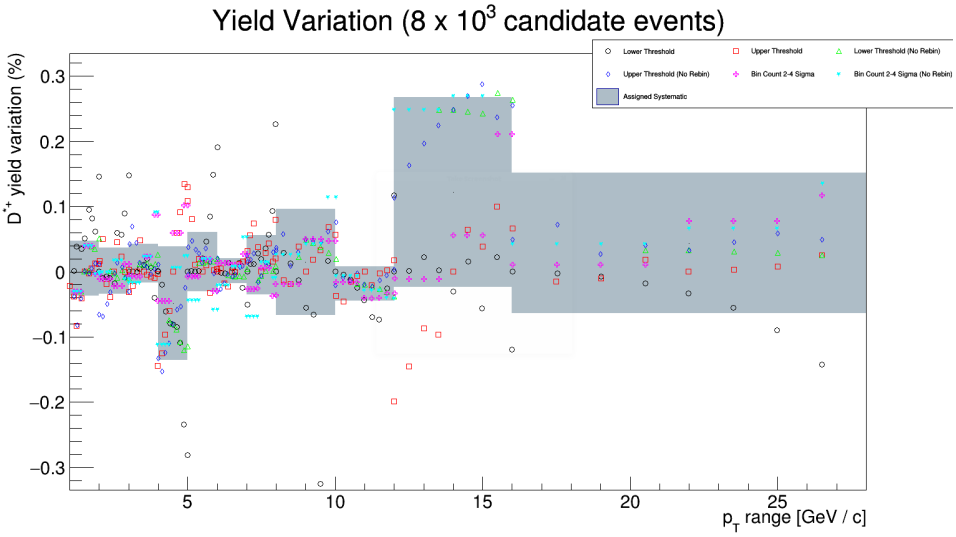

Figure 4.2: The yield variation of each for all 6 variation sequences with 8 samples each, for every p_T bin. The gray band shows the assigned systematic for a p_T bin. The data set used contains $8 \cdot 10^3$ events

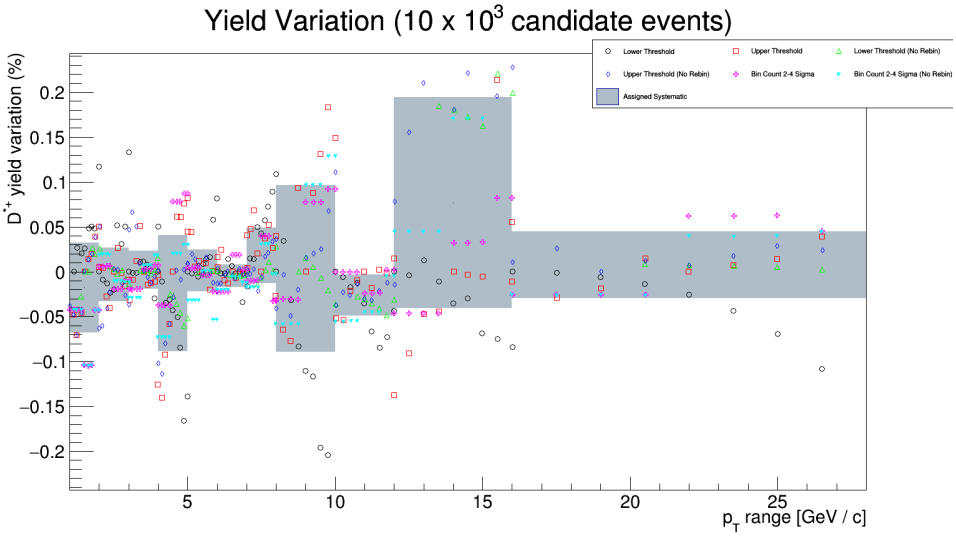

Figure 4.3: Same as figure 4.2 but with a different data set containing $10 \cdot 10^3$ candidate events

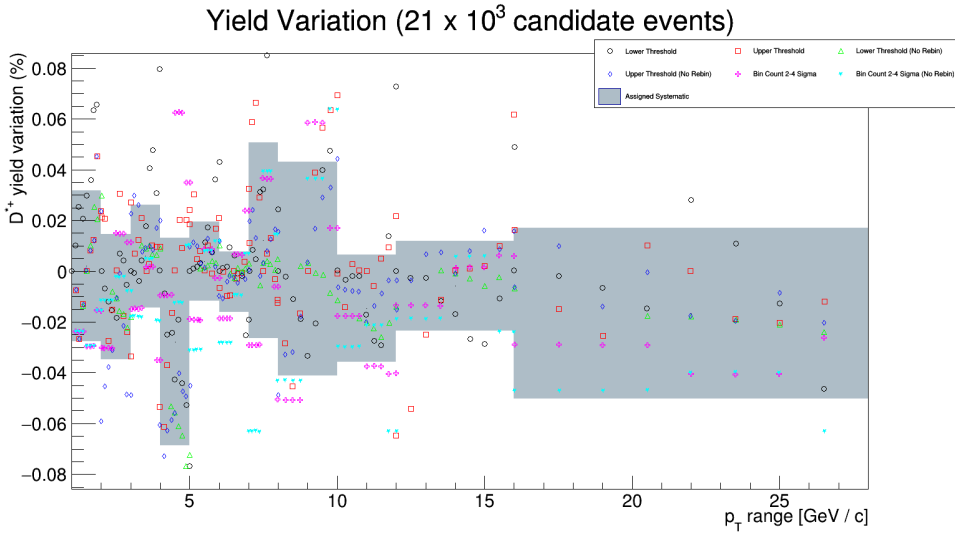

Figure 4.4: Same as figure 4.2 but with a different data set containing $21 \cdot 10^3$ candidate events

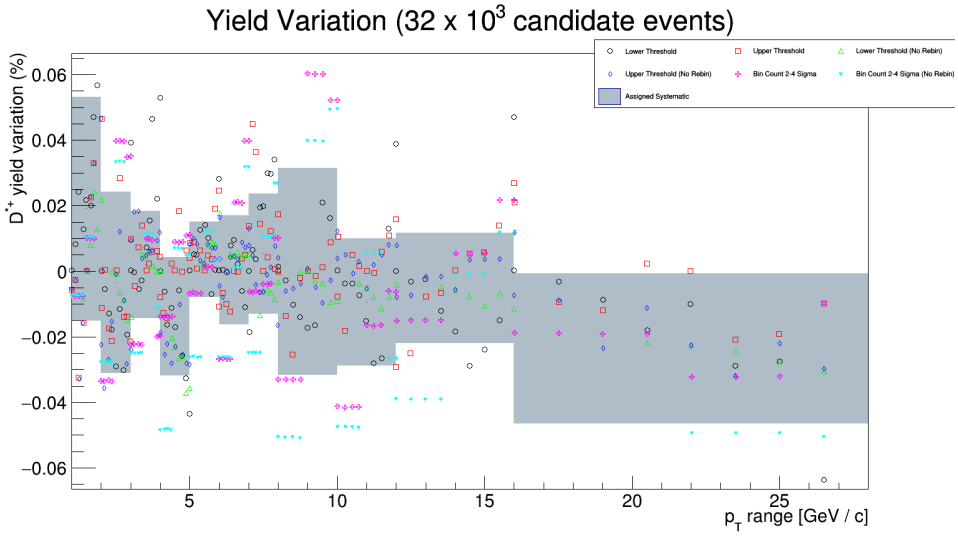

Figure 4.5: Same as figure 4.2 but with a different data set containing $32 \cdot 10^3$ candidate events

Similar plots showing measurements performed at different numbers of events are included in Appendix A.

4.1 Signal variation

As described in the chapter three, the assigned systematics in figure 4.2 to 4.5 are not direct measurements of the pure systematic. The apparent systematic is correlated

with statistical variations. To determine what the pure systematic is, a measure of how the apparent systematic changes with an increasing number of events is needed. Figure 4.6 shows the assigned systematic for each bin plotted against the number of events in that particular bin for all 14 datasets. Note that the graphs have different x-axis scales, as the number of events per bin can vary quite drastically.

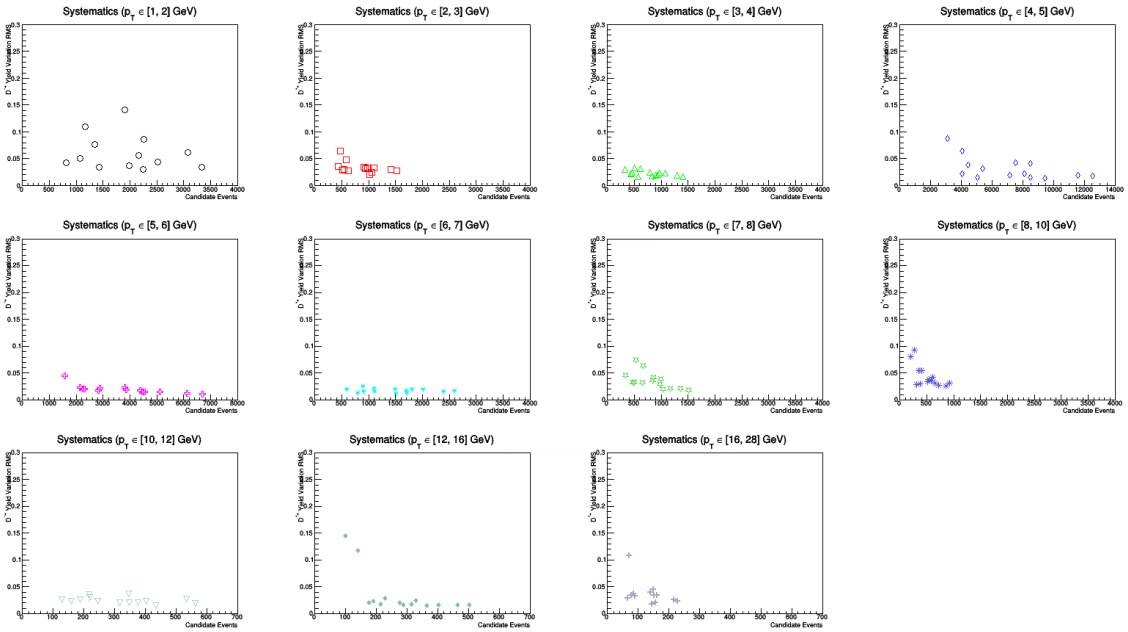

Figure 4.6: The assigned systematic of each bin set against the number of events in that bin for all 14 data sets

To get another view of how the variation changes with an increasing signal, figure 4.6 shows the variation of all 54 samples in each pt bin, plotted against the significance of the invariant mass fit of a sample.

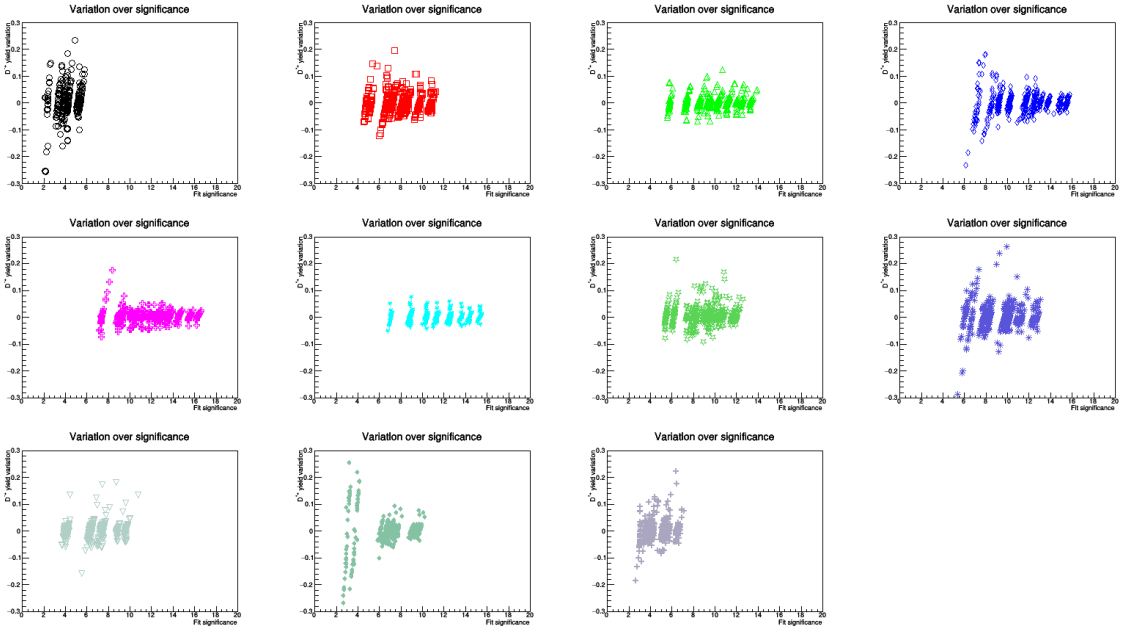

Figure 4.7: The variation of all 54 sample points plotted to the significance of the fit of each of the samples invariant mass distribution. Each plot represents a different p_T bin.

As the scale of the behavior in figure 4.6 does seem similar for all bins, figure 4.8 plots all of the p_T bins together to show a total overview. Note the abscence of events under two sigma, as these fits were discarded like described in chapter three.

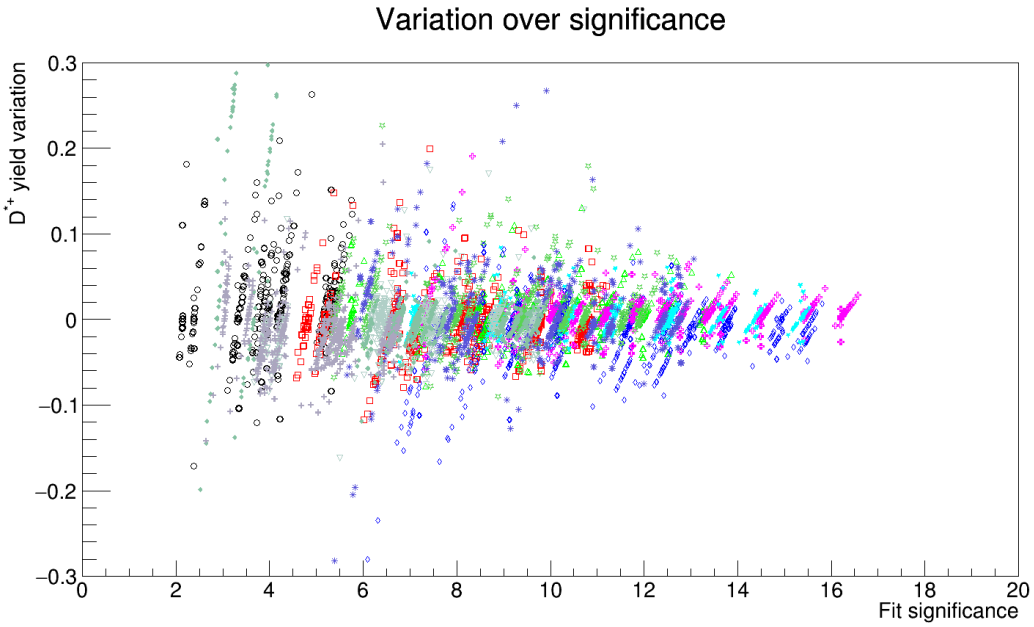

Figure 4.8: The variation of all 54 sample points plotted to the significance of the fit of each of the samples invariant mass distribution. Each color represents samples from a different p_T bin. The abscence of samples with less than two sigma significance is explained in chapter three.

Chapter 5

Study of the systematic uncertainties

Generally, the assigned systematics for all p_T bins in the full dataset are in a similar range as the statistical variation. Figure 5.1 shows the assigned systematics and the statistical variation in the same graph.

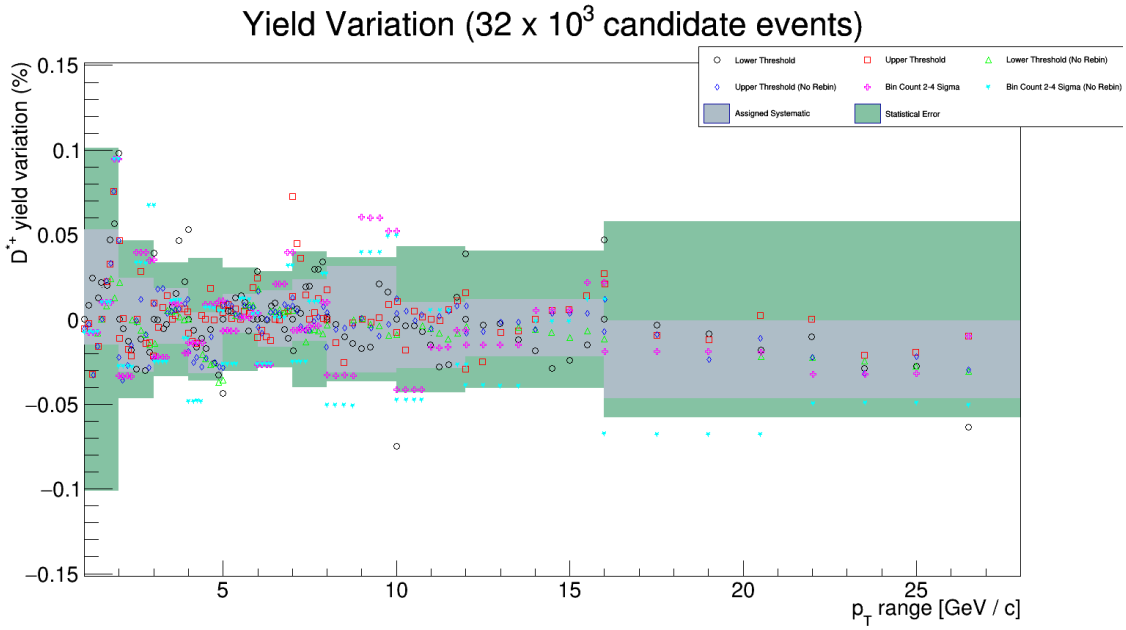

Figure 5.1: Statiscal variation of the default yield shown as the green band. Note how the gray band, the assigned systematic, falls within the statistics band for all bins.

Some specific points do fall out of this range and/or out of the range of the assigned systematic entirely, however. For lower signal counts this is easily explained

by the high statistical fluctuations in the data. Even when using the full data set, however, there still are some outliers. Nearly all of these outliers appear to come from bin-counting (with or without rebinning). This intuitively makes some sense. Outliers that show a yield that is too low come from two sigma counts. A two sigma count is expected to be about 5% too low given the shape of the Gaussian signal, and the scatter plots reflect this. A variation of 5% is already quite close to the total assigned systematic. Higher sigma bin counts do generally fluctuate around 0, like expected, but still have a few outliers too. Bin counting essentially samples whether the measured background is slightly too low or too high near the signal. A deviation in 1 of the counted bins can cause a relatively large deviation as only a few bins are taken into account for the yield, and therefore it is relatively easy for the bin counting to have outliers. Another noticeable feature in the bin counting variations are the straight lines in the sequence. This artifact is due to how this variation is performed. The sigma of the bin count is varied continuously, but the invariant mass distribution is in discrete bins. Therefore, the bin count is the same for slightly different sigmas and jumps when a new bin is added, causing the straight lines.

When varying the signal and plotting this to the assigned systematic a decrease of the assigned systematic in each of the bins is visible. This matches the behavior described in chapter 3. Also as expected the systematic does not tend to zero but rather seems to have an asymptote that approaches the 'pure' systematic. This behavior indicates that the systematic variation is indeed lower than what a direct measurement of the assigned systematic would indicate.

Some of the bins do exhibit this behavior more clearly than others, though. For example the assigned systematic for the $p_T \in [1, 2]$ GeV/c bin seems to fluctuate wildly with the signal. Some other bins don't have a decrease but seem to keep a quite constant assigned systematic despite the signal increasing. These bins seem to already be in line with the lower limit of the systematics in other bins which might explain why they don't have a downward trend. The bin with the widest range of events shows a decrease in the assigned systematic from 8.9% at 3000 events to 2.6% at 13000 events.

When plotting the yield variation to the fit significance as in figure 4.9 a similar pattern arises. Higher significance indicates less statistical variation, which results in a smaller apparent systematic and this behavior has a similar asymptote. Figure 4.9 shows some other behavior too, though. Regardless of the significance some band

within ±2.5% of variation seems consistently more densely populated indicating a 'baseline' systematic.

Another anomaly is the straight, increasing, lines visible in the graph. Figure 5.2 shows the same plot with the result of only a single variation sequence (upper cutoff threshold). It seems that plots with a higher significance correlate with a slightly higher yield. Every line also has to pass through 0 variation at some point as the sequence also samples the default parameters. Every data set has a different 'base' significance and so the plot shows 14 lines with a positive slope crossing the x-axis.

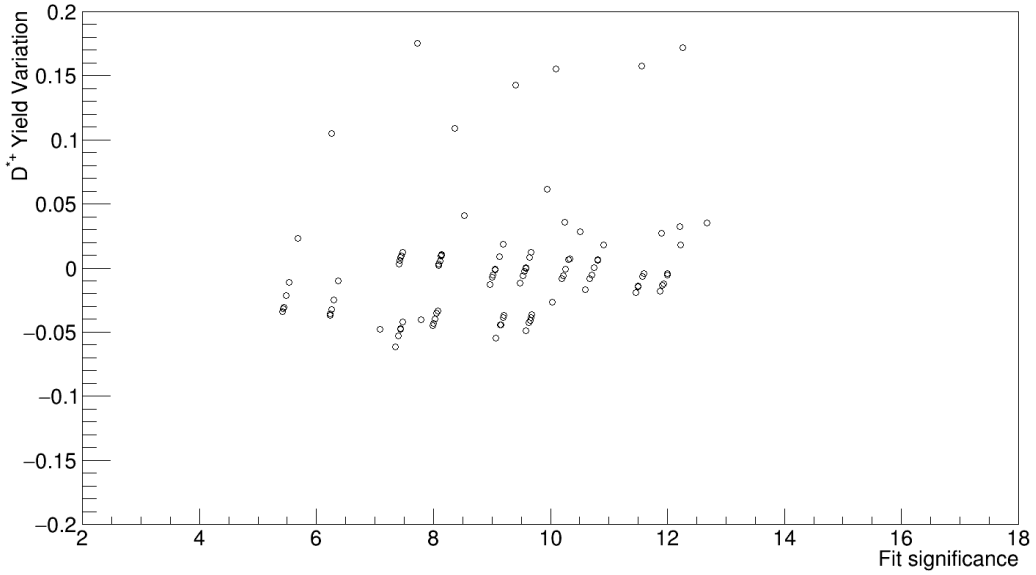

Figure 5.2: Variation plotted to the significance for only the upper cutoff variation sequence for the $p_T \in [6, 7]$ GeV/c bin. This plot shows how increased significance seems to correlate with a slightly higher yield in this sequence. Every data set corresponds to one of the dotted lines.

Other variation methods like bin counting do not show this pattern, however, so it seems that it is an artifact of varying the lower/upper cutoff threshold. Why this correlation exists is not clear at the moment.

It is tempting to look at the apparent asymptote in the plots and assign a systematic uncertainty based on that ($2.5 \pm 0.5\%$ very roughly) but more careful analysis is required before claims about the exact behavior of the apparent systematic can be made. The important takeaway is that measurement of the systematic uncertainty by varying the yield extraction parameters only ever measures an upper limit for the systematic, but not the systematic uncertainty itself.

Chapter 6

Conclusions

This thesis describes some of the theory of quark-gluon plasmas, why this state of matter is interesting and how current measurements potentially disagree with theoretical predictions. New experiments will help to determine whether the measurements indeed do not match up. This thesis gives a detailed method of measuring systematic uncertainty in the yield extraction of the D^{*+}meson. The results from this method match up with the expected ranges. Alongside the 'raw' assigned systematics a description of what would be the 'pure' systematic was given. In the future, new 13 TeV experiments and potential B^{*+}probing will further the research into the QGP.

6.1 Discussion

The variations cover quite a few of the details in the invariant mass extraction but there are more variations that could be considered (more rebinning variations, use different background fitting functions, use different topological cuts, etc). These were not deemed important for now but could be of interest for future research. The chosen p_T bins are arbitrary too but were not varied. Varying the p_T cutoffs slightly, or even sampling a few different configurations could provide an additional measure of the systematic variation.

The methods used are not limited to the particular 7 TeV pp dataset that was used. This research, however, has only analyzed the systematics for this particular data set. While there is no inherent reason why systematics would be different for different data sets at the same energy, it would be useful to verify this and see how these measurements work for different data sets. Additionally, datasets at higher

energies might have different statistical behaviors and show a different behavior for the assigned systematics too.

Another question raised is whether the yield extraction methods can be changed in a way to reduce the systematic uncertainty. The methods described here could be performed on different versions of the yield extraction routine to see if any successful changes can be made.

Bibliography

[1] CERN. ROOT a Data analysis framework. *root.cern.ch/doc/v608/*, 2016.

[2] D. Acosta et al. Measurement of Prompt Charm Meson Production Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Physical Review Letters*, Vol. 91, 2016.

[3] V. Khachatryan et al. Evidence for Collective Multiparticle Correlations in $p - Pb$ Collisions. *Physical Review Letters*, Vol. 115, June 2015.

[4] Paolo Giubellino. Perspectives of the ALICE experiment. *Brazilian Journal of Physics*, Vol. 34, 2003.

[5] Igor V. Gorelov. Heavy Flavor Production in CDF II Detector. *AIP Conf.Proc.842:271-273*, 2006.

[6] Alessandro Grelli. Probing the Quark Gluon Plasma with Heavy Mesons. *Physik Der Starken Wechselwirking*, Seminar Strong Interaction, 2015.

[7] Alessandro Grelli. Private conversation. 2016.

[8] Edmond Iancu. QCD in heavy ion collisions. *Institut de Physique Thorique de Sacla*, Slide 29, 2012.

[9] Jiangyong Jia. Influence of the nucleon-nucleon collision geometry on the determination of the nuclear modification factor for nucleon-nucleus and nucleus-nucleus collisions. *Physics Letters B*, Volume 681, Issue 4, Pages 320325, November 2009.

[10] Piotr Kukla. D^{*+} analysis in proton-proton collisions at $\sqrt{s} = 2.76$ TeV using the ALICE detector at CERN. *Institute for Subatomic Physics*, Aug 2011.

[11] R.H. Schindler. New Results on Charmed D Meson Decay. *California Institute for Technology*, 1986.

Appendix A

Appendix A

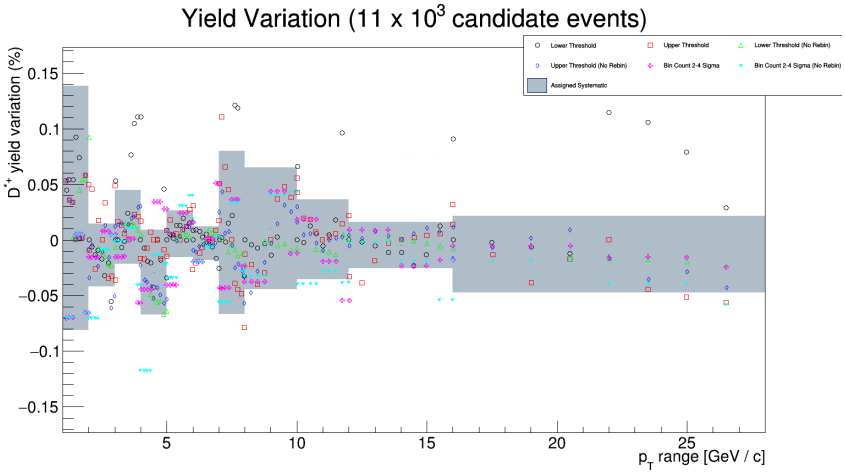

Figure A.1: The yield variation of each for all 6 variation sequences with 8 samples each, for every p_T bin. The gray band shows the assigned systematic for a p_T bin. The data set used contains $11x10^3$ events.

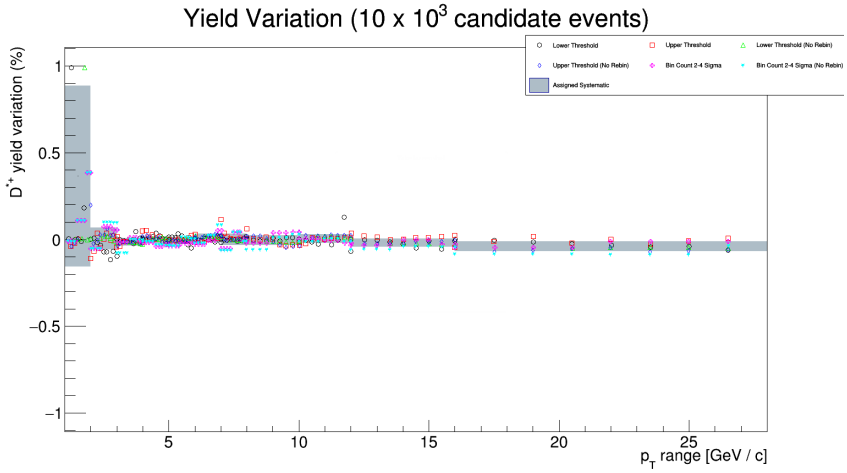

Figure A.2: Same as figure A.1 with a different data set containing $10x10^3$ candidate events.

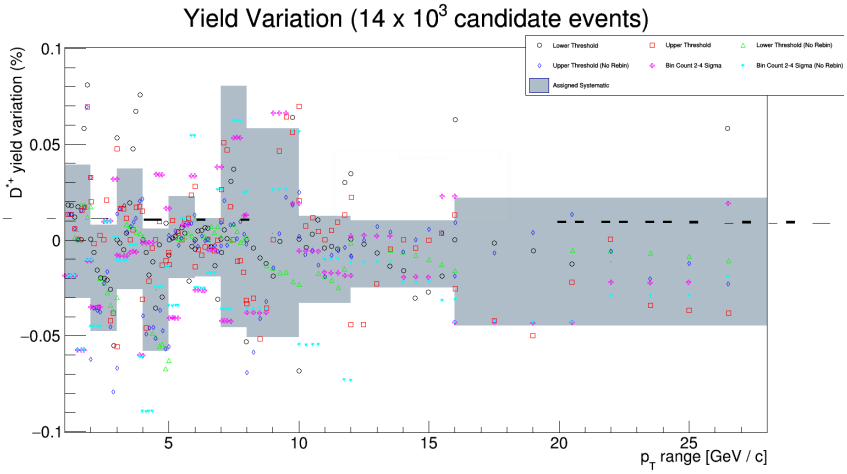

Figure A.3: Same as figure A.1 with a different data set containing $14x10^3$ candidate events.

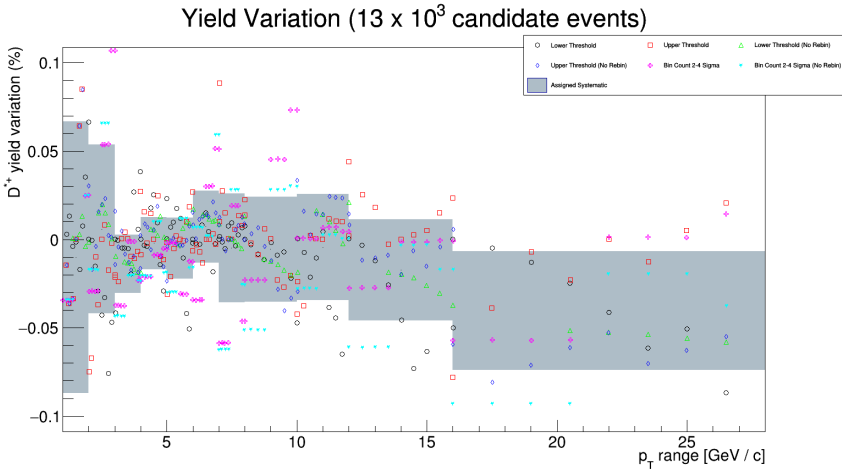

Figure A.4: Same as figure A.1 with a different data set containing $13x10^3$ candidate events.

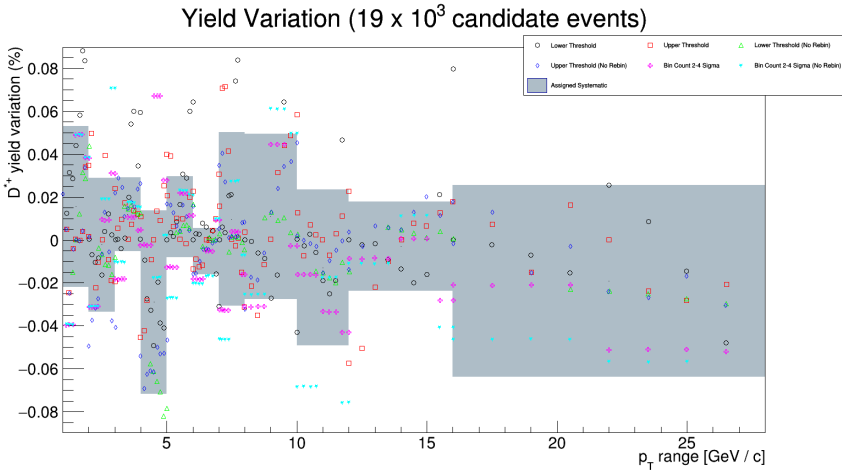

Figure A.5: Same as figure A.1 with a different data set containing $19x10^3$ candidate events.

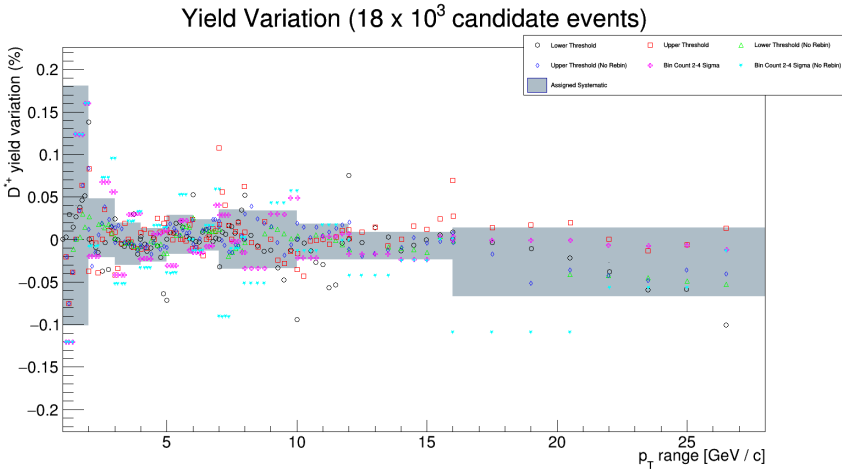

Figure A.6: Same as figure A.1 with a different data set containing $18x10^3$ candidate events.

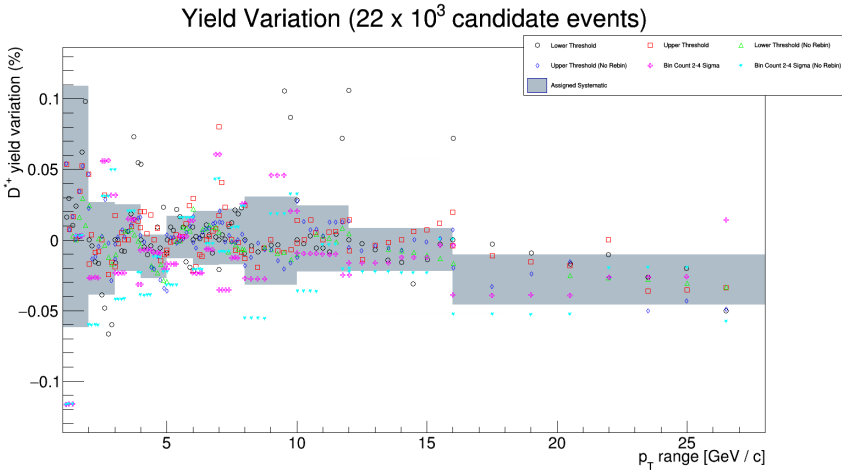

Figure A.7: Same as figure A.1 with a different data set containing $22x10^3$ candidate events.

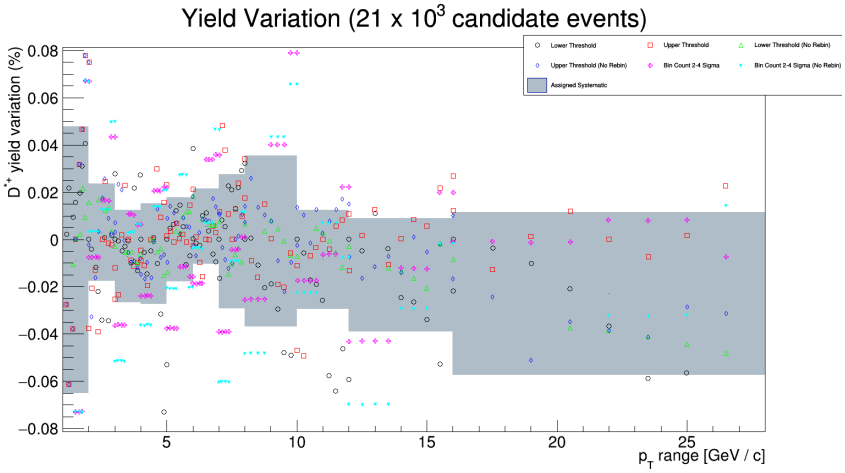

Figure A.8: Same as figure A.1 with a different data set containing $21x10^3$ candidate events.

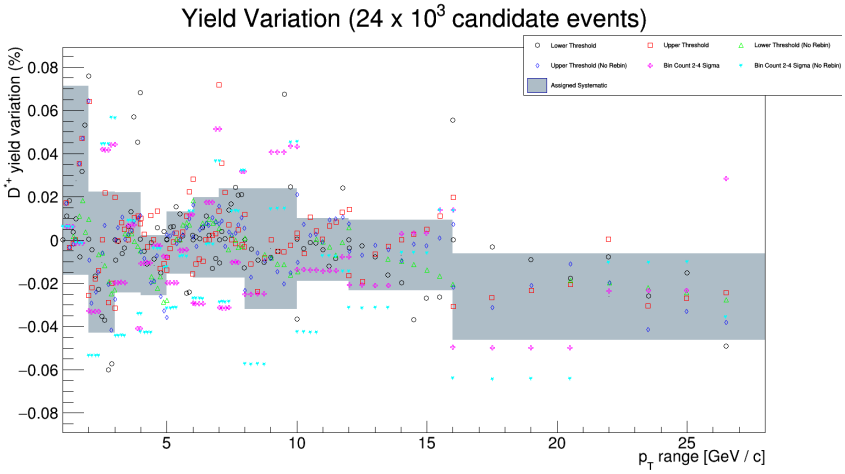

Figure A.9: Same as figure A.1 with a different data set containing $24x10^3$ candidate events.

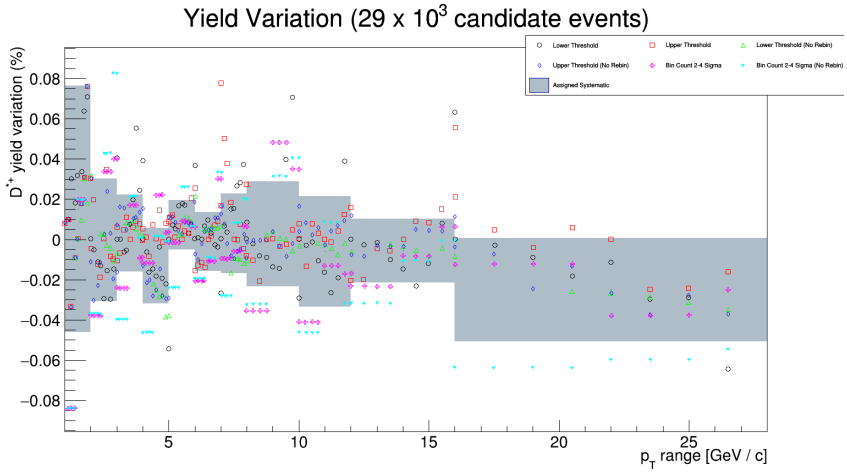

Figure A.10: Same as figure A.1 with a different data set containing $29x10^3$ candidate events.