# Analysis of $D^0 - \bar{D}^0$ Mixing

Nagesh Kulkarni, Mark Mattson, and Paul Karchin
Wayne State University, Detroit, MI

CDF/DOC/BOTTOM/CDFR/9806
Version 1.1

*June 15, 2009*

**Abstract**

The purpose of this CDF note is to present the new results in $D^0 - \bar{D}^0$ mixing analysis using $D^0 \to K\pi$ decays. We measure the time dependent ratio of doubly Cabibbo suppressed decay rate to Cabibbo favored decay rate.

Our previous result updated in July 2008 excluded no-mixing hypothesis with 4.27 Gaussian standard deviations using approximately 2.3 fb$^{-1}$ of integrated B-physics luminosity. In this note we present our attempts to improve the analysis method by introducing new cuts and applying Artificial Neural network method. The new results from the addition of approximately 1.5 fb$^{-1}$ of new data are presented. Systematic uncertainties are updated. The new results show reduced statistical and systematic uncertainties. We compute the expected mixing significance from toy Monte Carlo simulations and expect to exclude no-mixing (x' =y' = 0) with $\approx$ 5 Gaussian standard deviation.

# Contents

**1 Motivation** **5**

**2 Introduction** **6**

**3 Current Experimental Status** **7**

**4 The CDF experiment** **10**

**5 Data Samples** **11**

**6 Qualitative comparison of the old and new data sets** **11**
6.1 $dE/dx$ variable distribution comparison . . . . . 13
6.2 $D^0$ mass distribution comparison . . . . . 13

**7 Analysis and Results** **16**

**8 Toy Monte Carlo simulations of WS/RS ratios and estimate of mixing significance** **18**
8.1 $\delta\chi^2$ method to estimate probabilities . . . . . 18
8.2 Bayesian contour method . . . . . 18

**9 Attempts to improve the standard analysis method** **18**
9.1 $D^0$ isolation variable cut . . . . . 18
9.2 Multiple candidate background . . . . . 19

**10 Artificial Neural Network method** **21**
10.1 Neural network method for $D^0 - \bar{D}^0$ mixing analysis . . . . . 21
10.2 Training sample preparation . . . . . 22
10.3 Input parameters and settings . . . . . 22

**11 Lump study** **24**

**12 Systematic Uncertainties** **26**
12.1 Signal Shapes . . . . . 26
12.2 $K\pi$ Background . . . . . 26
12.3 Mis-assigned Background Correction . . . . . 26
12.4 Non-prompt $D^*$ Correction . . . . . 27
12.5 Mass Difference . . . . . 27
12.6 $K\pi$ Lump at 1.83 GeV . . . . . 27

**13 Conclusion** **31**

**Appendices**

**A $K\pi$ mass fit comparison for the old and the new data** **32**

**B Time slices1:period 0 to 21 data** **32**

**C Time slices2: period 17 to 21 data** **43**

**D Time slices3:period 0 to 16 data** **54**

**E Some variables used as Neural network inputs** **65**

**F Non-Prompt $D^*$ Correction Details** **67**

# List of Figures


1 Babar, Belle, and CDF results using wrong sign hadronic decay channel ($D^0 \to K\pi$ ). The first plot is Babar's measurement of $x'^2$ and $y'$ . The second plot is the Belle's result in which $x'^2$ is constrained to positive values. The last two plots are CDF's July 2008 results with and without restrictions on $x'^2$ . . . . . 7

2 Results from Belle: The top three plots are the comparison of lifetime differences of $D^0$ CP eigenstates and CP mixed states. The first two plots are the decay time distributions from $D^0 \to KK$ and $D^0 \to \pi\pi$ channels. The third plot shows decay time distribution using $D^0 \to K\pi$ decay. The last plot is Belle's measurement of $x$ and $y$ from time dependent Dalitz analysis. . . . . 8

3 $dE/dx$ variable for period 0 to 9 data (blue) and for period 10 to 21 data (red). The change in the distribution is due to the detector changes made to handle higher initial instantaneous luminosity. . . . . 13

4 dE/dx variable distributions for $K$ and $\pi$ tracks each with $K$ hypothesis and $\pi$ hypothesis. Red: Period 17 to 21 data, blue:period 0 to 16 data. The red and blue curves in all four plots almost overlap and show the consistency of the new data and the old data after the detector change. . . . . 14

5 The first plot shows m(K$\pi$) distribution comparison. Red curve shows period 17 to 21 data and the blue curve shows period 0 to 16 data. The second plot is also $D^0$ mass distribution comparison in a different way as described in the text. . . . . 15

6 Ratio plots: The first plot is the WS/RS ratio from period 0 to 16 data. The second plot is from period 17 to 21 data. The last plot is from period 0 to 21, in which only uncertainties on the mixing parameters are presented to follow the blinding procedure. 17

7 Simulated experiment and $\delta\chi^2$ distribution: The left plot show single simulation of WS/RS ratio. The second plot shows the $\delta\chi^2$ distribution from 1000 simulated experiments. . . . . 18

8 Bayesian contours from simulated experiment. The no-mixing point is on $4.92\sigma$ contour. 19

9 WS $D^*$ yield vs. Isolation of $D^0$ cut. The yield is almost flat. It appears that the $D^0$ isolation variable is not useful as a standard cut. . . . . 19

10 Scatter plot showing WS and RS mass distribution. The plot shows WS signal dominated by RS background. . . . . 21

11 A typical artificial neural network has input nodes, synapses, middle layer with hidden nodes, and output node. The network finds a discrimination function by adjusting weights on each of the synapses. . . . . 22

12 Signal and background samples used for training. The red corners show the region from where background sample is extracted. The central horizontal bar shows the signal region. 23

13 Neural network training output: Signal is show in red color and background in black color. Signal gradually increases and background decreases with NN output. Maximum signal significance is found at NN cut $> 0.10$. . . . . 24

14 $K\pi$ mass fit with and without extra Gaussian in addition to quadratic function for background. The green curve show fit without Gaussian. . . . . 25

15 $K\pi$ mass for "good" RS (left) and WS (right) candidates, fit without an extra Gaussian (lump). . . . . 28

16 $K\pi$ mass for "good" RS (left) and WS (right) candidates, fit with an extra Gaussian (lump) that is treated as part of the signal shape. . . . . 29

17 $K\pi$ mass for "good" RS (left) and WS (right) candidates, fit with an extra Gaussian (lump) that is not considered signal. The RS and WS fits have independent amplitudes for the extra Gaussian, although they have the same mean and width. The WS fit has zero for the "lump" amplitude. . . . . 30

18 Time integrated $K\pi$ mass fits using the standard double Gaussian fit function. Left plot is for period 0 to 16 data and the right plot is for period 17 to 21 data. The first three parameters are listed in table 7 . . . . . 32

19 Some of the neural network inputs: 1. Transverse momentum of kaon, 2. Transverse momentum of pion, 3. Transverse momentum of soft pion, 4. The Z0 variable-the point of closest approach along the beam line -for soft pion, 5. Impact parameter for soft pion. In all plots red data shows distribution for signal and black data shows background distribution. . . . . 66

# List of Tables

1 Summary of three results from wrong sign hadronic decay channel. . . . . 7
2 Bstntuple data sets and corresponding date ranges . . . . . 12
3 RS $D^*$ yields per unit luminosity in the old and the new data . . . . . 12
4 The ratio of the RS $D^*$ yields from the old and the new data sets in different time bins 12
5 WS $D^*$ significance using the standard method and various neural network cuts. The maximum signal significance is at NN cut $> 0.10$ . . . . . 24
6 $\chi^2$ /ndf for with and without extra Gaussian for the old and the new data. . . . . 25
7 The comparison of the first fit parameters of $K\pi$ mass fits for the old and the new data. The mean and the sigma for one Gaussian are shown. There is no shift in $D^0$ mass distribution in the new data. . . . . 32
8 Prompt and Non-Prompt Distribution Fractions . . . . . 73

# 1 Motivation

The $D^0 - \bar{D^0}$ mixing is now fairly established with the contributions from the BaBar, Belle, CDF, and Cleo experiments. The first evidence of $D^0 - \bar{D^0}$ mixing came in 2006 from BaBar [1], which was immediately confirmed by the CDF experiment in March 2007 [2]. Both the experiments analyzed $D^0 \to K\pi$ decay channel. Belles first result [3] that measured life time differences of CP eigenstates and CP mixed states was published at the same time as BaBars. They also published the mixing using the time dependent Dalitz analysis method [4]. Later, the BaBar experiment performed similar measurements as Belles [5]. Previously the Celo experiment contributed to the searh by measuring individual mixing parameter $y$ and the strong phase difference $\delta$ [6]. The Focus experiment has also analyzed wrong sign hadronic decay channel [7].

The combined result from all the experiments excludes no-mixing with more than 6.7 Gaussian standard deviations ($\sigma$) and thus provides a clear observation of $D^0 - \bar{D^0}$ mixing [8]. However, there is no single experiment that provides above 5 $\sigma$ observation. BaBar and Belle experiments exclude no-mixing with slightly more than 3 Gaussian standard deviation. In July 2008, CDF presented the result that excludes no-mixing with 4.27 Gaussian standard deviation. There are also no signs of CP violation from any of the above experiments.

At CDF we initiated our efforts to improve the result immediately after our publication in 2007. We started with exploring new ways to reduce backgrounds by introducing new cuts and improving fits. We tried the application of Artificial Neural Network (ANN) in the hope that it will bring significant improvement over the standard method.

At this time we have almost 4 fb$^{-1}$ of integrated B-physics luminosity data at our disposal. This brings us to a unique position where we can provide the world's most significant measurement of the mixing parameters and may provide the first observation of the mixing as a single experiment, if it exists in nature. For these reasons, we would like to present the new $D^0 - \bar{D^0}$ mixing results in this note.

The next section briefly introduces the $D^0 - \bar{D^0}$ mixing theory. Section 3 summarizes the current experimental status and provides the world average of the mixing parameter values. Section 4 introduces the main detector components relevant to this analysis. The description of the new results starts from section 5 in which we start comparing the old and the new datasets in terms of luminosities. Section 6 presents a qualitative comparison of the old and the new data sets. In that section we show that the Right sign $D^*$ yield is decreasing with time but other than that, the new data is consistent with the old data. Section 7 outlines the analysis method briefly for completeness and presents the new measurements of the mixing parameters. Section 8 describes two methods based on toy Monte Carlo (MC) simulations to estimate the expected mixing significance. Section 9 describes our various attempts to improve the standard analysis method. Section 10 is devoted to the description of analysis using Artificial Neural Network method. The section describes the basic idea, design, training and testing of the neural network. Section 11 is about a quick analysis performed to study the extra lump in RS $K\pi$ mass distribution. section 12 describes updated systematic uncertainties. Finally, we summarize the work in the last section 13. Appendix A shows $K\pi$ mass fit comparison of the old and the new data sets, appendices B to D provide time sliced $K\pi$ and $D^*$ yield plots. In appendix E we show the signal and backgound distributions for some of the input variables used in neural network analysis. In appendix F, the calculations for non prompt $D^*$ are described.

This note is based on the two previous CDF notes 7116 [9] and 8879 [10] describing time independent analysis and time dependent analysis respectively. Since these notes describe the theory and methods at length, in this note they are outlined only briefly for the purpose of completeness. The emphasis is more on the new studies.

# 2 Introduction

In this section, we briefly discuss the $D^0 - \bar{D^0}$ mixing formalism, introduce the mixing parameters and explain the goal of the analysis.

Analogous to $K^0 - \bar{K^0}$ and $B^0 - \bar{B^0}$ systems, the $D^0 - \bar{D^0}$ systems can be described by the Schrodinger equation

$$i\frac{\partial}{\partial t}\begin{pmatrix} D^0(t) \\ \bar{D^0}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} \end{pmatrix}\begin{pmatrix} D^0(t) \\ \bar{D^0}(t) \end{pmatrix}$$

Where, $\mathbf{M}$ and $\mathbf{\Gamma}$ are the Hermitian matrices. The $\Gamma$ term ensures the particle decay. The CPT invariance require $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$. In case of negligible CP violation, the mass eigenstates are also the CP eigenstates and can be written as linear combination of flavor eigenstates.

Solving the equation gives time evolution of the $D^0 - \bar{D^0}$ systems which will have oscillatory exponential terms showing oscillations of $D^0$ into $\bar{D^0}$ and vice-versa. The slight difference in masses and decay widths of the two mass eigenstates governs the mixing. The mixing parameters are then defined as

$$x = \frac{m_1 - m_2}{\Gamma} = \frac{\delta m}{\Gamma} \quad \text{and} \quad y = \frac{\Gamma_1 - \Gamma_2}{2\Gamma} = \frac{\delta\Gamma}{2\Gamma}$$

The goal of the analysis is to measure these mixing parameters.

In weak interactions, the $D^0 \to K\pi$ decay proceeds through a doubly Cabibbo suppressed (DCS) decay, involving two 'cdw' vertices or through $D^0 - \bar{D^0}$ mixing followed by Cabibbo flavored (CF) decay. Unlike $K^0 - \bar{K^0}$ and $B^0 - \bar{B^0}$ systems, the $D^0 - \bar{D^0}$ system involves $d$ quark and so the mixing rate is small compared to $K^0 - \bar{K^0}$ and $B^0 - \bar{B^0}$ systems due to CKM factors. The mixing can occur through either long-range processes or through short-range box diagrams. In long-range processes, the $D^0$ decays into some intermediate state such as $\pi^+\pi^-$ and then evolves into $\bar{D^0}$ . In short range processes, the $D^0$ decay proceeds through penguin topology involving $d$, $s$ and $b$ quarks. The current theoretical limits on the mixing parameters from short range calculations are $x, y < 10^{-3}$ [11], .

The mixing parameters can be measured experimentally in several ways. Some of the important methods employed by various experiments are outlined in the next section. In our analysis, we use $D^0 \to K\pi$ channel and exploit the decay chain $D^* \to D^0\pi$ ; $D^0 \to K\pi$ . We define the decay chains as wrong sign or right sign depending on the relative charges of the tagging pion and that from $D^0$ . In particular, the decay chain $D^{*+} \to D^0\pi^+$ , $D^0 \to K^+\pi^-$ is identified as wrong sign and the decay chain $D^{*+} \to D^0\pi^+$ , $D^0 \to K^-\pi^+$ is identified as right-sign. The same holds for charge conjugates.

Following the formalism given in PDG review on $D^0 - \bar{D^0}$ mixing [12], the time dependent ratio of wrong sign to right sign decay rates is given by

$$r(t) \quad = \quad R_D + \sqrt{R_D}y'(t) + \frac{(x'^2 + y'^2)}{4}(t)^2 \tag{1}$$

where, $R_D$ is the the squared modulus of the ratio of DCS to CF amplitudes and $x'$ and $y'$ are the rotated $x$ and $y$ by strong phase factor $\delta$ as:

$$x' = x\cos\delta + y\sin\delta \quad \text{and} \quad y' = -x\sin\delta + y\cos\delta$$

The decay time $t$ is measured in terms of $D^0$ proper decay time.

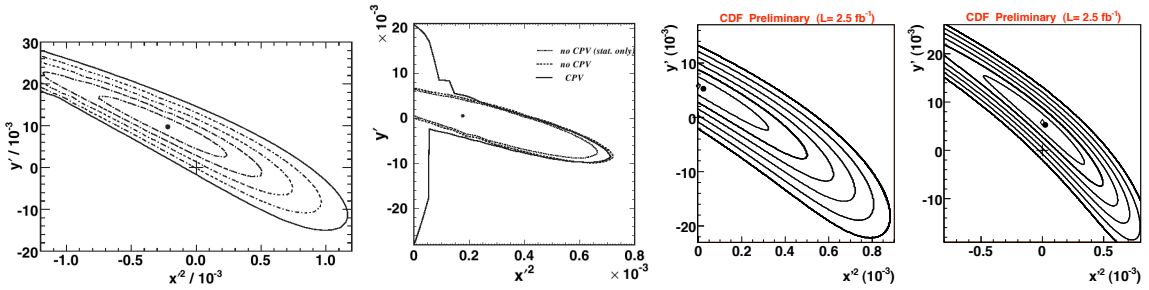



**Figure 1:** Babar, Belle, and CDF results using wrong sign hadronic decay channel ($D^0 \to K\pi$ ). The first plot is Babar's measurement of $x'^2$ and $y'$ . The second plot is the Belle's result in which $x'^2$ is constrained to positive values. The last two plots are CDF's July 2008 results with and without restrictions on $x'^2$ .

| Experiment | $R_D$ $(10^{-3})$ | $y(10^{-}3)$ | $x'$ $(10^{-3})$ | Mixing significance |
|---|---|---|---|---|
| BaBar | $3.03 \pm 0.19$ | $9.7 \pm 5.4$ | $-0.22 \pm 0.37$ | 3.9 |
| Belle | $3.64 \pm 0.17$ | $0.6 \pm 4.0 \pm 3.9$ | $0.18 \pm 0.21 \pm 0.23$ | 2.0 |
| CDF (2007) | $3.04 \pm 0.55$ | $8.5 \pm 7.6$ | $-0.12 \pm 0.35$ | 3.8 |
| CDF (2008-not published) | $3.36 \pm 0.48$ | $5.27 \pm 6.01$ | $0.02 \pm 0.26$ | 4.27 |

**Table 1:** Summary of three results from wrong sign hadronic decay channel.

# 3 Current Experimental Status

In this section we summarize the important experimental results in $D^0 - \bar{D}^0$ mixing.

The BaBar, Belle, and CDF experiments provide the evidence of mixing by analyzing wrong sign hadronic decay channel $D^0 \to K\pi$ . The BaBar experiment excludes no-mixing hypothesis with 3.8 Gaussian standard deviation using likelihood contours. In their measurement, $x'^2$ and $y'$ are allowed to take negative values. The Belle experiment performed similar studies but by restricting $x'^2$ value to physically allowed region. Their result excluded no-mixing with 2.0 Gaussian standard deviation using Feldman-Cousins method. The CDF experiment used Bayesian probability contours and presented the results with and without restricting $x'^2$ to physically allowed region. These results are shown in figure 1.

The Belle and BaBaR experiments also analyze decay time distributions of $D^0$ meson from CP eigenstates $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ and compare them with the decay time distribution from CP mixed eigenstate $D^0 \to K^-\pi^+$ . From this analysis they measure the individual value of mixing parameter $y_{CP}$ given by:

$$y_{CP} = y\cos\phi - \frac{1}{2}\mathrm{A_M}\sin\phi$$

where, $A_M$ describes CP violation in mixing and $\phi$ describes interference between mixing and decays. With no CP violation, $A_M$ and $\phi$ will be zero and $y_{CP}$ will be the same as y-the difference of decay widths of two flavor eigenstates. The measurement of $y_{CP}$ found by both the BaBaR and the Belle experiment provides evidence of mixing with 3 Gaussian standard deviation. Figure 2 shows the plots from the Belle experiment.

The third significant result was from time-dependent Dalitz analysis again from the Belle experiment using $D^0 \to K_S\pi^+\pi^-$ channel. The BaBar experiment also performed similar studies using $D^0 \to K^+\pi^-\pi^0$ decays. In Figure 2 the Belle's measurement of x and y from the time evolution of Dalitz distributions is shown in the last plot.

The Cleo experiment provided measurement of mixing parameter $y$ and strong phase difference $\delta$ using $\psi(3770) \to D^0\bar{D}^0$ channel. Other than these there were other approaches in which semileptonic $D^0$ decay channels were analyzed.

Table 1 summarizes the numerical values of $x'^2$ and $y'$ provided by wrong sign hadronic decay channel analyses from the recent experiments.

The heavy flavor averaging group (HFAG) provides the current average values of the mixing parameters by combining all of the published results as

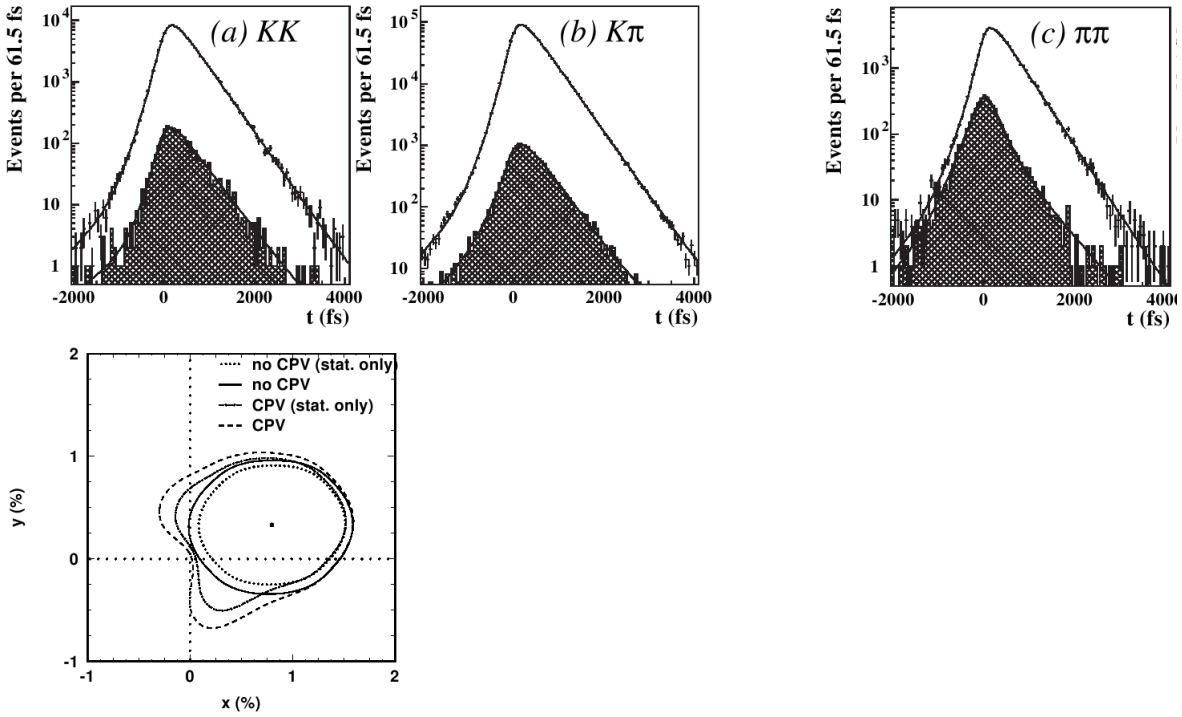



**Figure 2:** Results from Belle: The top three plots are the comparison of lifetime differences of $D^0$ CP eigenstates and CP mixed states. The first two plots are the decay time distributions from $D^0 \to KK$ and $D^0 \to \pi\pi$ channels. The third plot shows decay time distribution using $D^0 \to K\pi$ decay. The last plot is Belle's measurement of $x$ and $y$ from time dependent Dalitz analysis.

$$x = 0.97 \pm 0.27 \pm 0.29 \quad \text{and} \quad y = 0.78 \pm 0.18 \pm 0.19$$

These values correspond to exclusion of no-mixing hypothesis with 6.7 Gaussian standard deviation. With this background, we now present the CDF results.

# 4 The CDF experiment

The CDF II detector at the Fermilab Tevatron collider is a general purpose detector which detects proton anti-proton collisions at $\sqrt{s} = 1.96$ TeV. The detector components relevant to this analysis are the silicon vertex detector (SVX),Intermediate Silicon Layer (ISL), silicon Layer zero (L00), and the drift chamber (COT) surrounded by a superconducting magnet producing a magnetic field of 1.4 T. These components together provide the physics quantities required to reconstruct wrong sign and right sign events. The other detector parts relevant to B-physics data sets are: silicon vertex tracker (SVT), Extremely fast tracker (XFT), Cherenkov luminosity counter (CLC), Central muon detector (CMU), central muon upgrade detector (CMP), and central muon extension detector (CMX).

The CDF data acquisition system has a three level trigger system described by L1, L2, L3 triggers. The data passing the selected trigger requirements is processed offline.

Here we define some of the important physics quantities which will be frequently referred throughout the text.

Decay length significance for a particle is defined as

$$L_{xy} = \frac{\mathbf{r} \cdot \mathbf{P_T}}{P_T}$$

where $\mathbf{r}$ is the distance between the primary and $D^0$ vertices and $\mathbf{P_T}$ is the transverse momentum vector. In our analysis we sometimes use decay length significance $L_{xy}/\sigma_{xy}$, where $\sigma_{xy}$ is the uncertainty on $L_{xy}$ Impact parameter for a particle is the distance of the closest approach from the primary vertex and is defined as

$$d_0 = \frac{\hat{\mathbf{z}} \cdot (\mathbf{r} \times \mathbf{P_T})}{P_T}$$

where $\hat{\mathbf{z}}$ is the unit vector along z-axis.

# 5 Data Samples

This section describes data samples, trigger requirements, and luminosity calculations. The detector components and physics quantities mentioned here are introduced in the previous section.

The data samples used in $D^0 - \bar{D^0}$ mixing analysis are derived from B-physics hadronic data sets passing two track trigger (TTT) requirements. These requirements are described in the CDF note 6526. At level 1, the TTT trigger requires two XFT tracks with transverse momentum $P_T > 2.05$ GeV/$c$ and scalar sum of two $P_T$ s $> 5.5$ GeV/$c$ . At level 2, two SVT tracks are required to have impact parameters between 0.12 mm and 1.0 mm. At level 3, the intersection of two tracks in $r - \phi$ plane projected on the net momentum vector of the two tracks is required to be 200 $\mu$m or greater from the beam line.

The events accepted by the above trigger requirements are written to $B \to \pi\pi$ data block of Bstntuple data sets [13]. The offline processing is done periodically by processing the data collected over a certain time period. Table 2 shows the Bstntuple data sets used in this analysis with corresponding run numbers, date range over which the data was collected and integrated luminosities. The luminosity values are obtained from the CDF II online run summary database (http::/www-cdfonline.fnal.gov/daq/runSummary/). For B-physics the following run bits are set true: Required silicon, Good run (core components), CLC, L1, L2, L3, COT, CMU, CMP, CMX, SVX, and ISL. These settings yield the B-physics luminosity approximately 70 % total online luminosity.

For daily use we make our own $D^0 - \bar{D^0}$ mixing specific data sets from selected Bstntupe data sets. The mini-data sets are small, manageable and allow faster processing. They are given shorter names as shown in the sixth column of the table 2. Each mini-dataset comprises several root files. To be able to run a medium type job on computer farm, the size of each file is limited to 10 million events.

Table 3 compares the data periods used for July 2008 result (we will refer to this data set as old data) and the new data periods available after that result. The July 2008 result used period 0 to 16 data collected from 2002 to 2008 corresponding to integrated B-physics luminosity $\approx 2.23$ fb$^{-1}$ . The new data added for the current analysis includes period 17 to 21 collected from February 2008 to December 2008 and adds $\approx 1.5$fb$^{-1}$ of data. The last row shows the total data used for this analysis.

# 6 Qualitative comparison of the old and new data sets

To qualitatively compare the old and the new data sets, we obtain the right sign $D^*$ yields per unit luminosity. The $D^*$ yields are obtained using the yield script described in the next section. The table 3 compares the RS $D^*$ yields for the data sets. We observe that the RS $D^*$ yield per unit luminosity has decreased from 1.9 million to 1.5 million. The $D^*$ yield is decreasing with time. This is consistent with the findings from the other CDF groups.

We can use these numbers to estimate what to expect at the end. With 4.2 million RS $D^*$ events in July 2008, no-mixing was excluded with $\sigma_{old} = 4.27$. Now we have 5.6 million RS $D^*$ s. If we assume that the Gaussian standard deviation is proportional to $\sqrt{N}$, where N is the number of events, then we may expect $\sigma_{New} = \sqrt{\frac{N_{New}}{N_{old}}} \times \sigma_{old}$. Then with the above numbers we get $\sigma_{New} = 4.94$. This tells us that we can expect to exclude no mixing with $\approx 5\sigma$ if we do not change the analysis method. We may expect higher significance if the application of ANN brings some improvements in yield.

To see how the yield is affected as a function of $D^0$ decay time, we look at the ratio of RS $D^*$ yields of the old and the new data sets. This is shown in table 4. The table shows the ratio of the RS $D^*$ yields for the new and old data sets in each time bin. As we can see, there are no large variations in the ratio of the yields as function of decay time and they are consistent with the time integrated ratio of the yields (the last row).

Apart from $D^*$ yield decrease with luminosity, there are no other major differences observed between the old and new data sets. We studied particle identification ($dE/dx$ variable) distribution (please refer to CDF note 8879 section 3.2.3 for the definition of $dE/dx$ variable) and $D^0$ mass distributions to qualitative compare different data sets.

| Data set | Periods | Run range | Dates | Lumi ($fb^{-1}$) | Mini data set | Number of files |
|---|---|---|---|---|---|---|
| hadr-80/xbhdid | 0 | 138425-186598 | 2002.02.09-2004.08.22 | 0.37 | x0d | 27 |
| cdfpbnt/xbhdih | 1-4 | 190697-203799 | 2004.12.16-2005.09.04 | 0.35 | x0h | 21 |
| cdfpbnt/xbhdii | 5-10 | 203819-233111 | 2005.09.05-2007.01.30 | 0.84 | x0i | 30 |
| cdfpbnt/xbhdij | 11-13 | 233133-246231 | 2007.01.30-2007.05.13 | 0.39 | x0j | 15 |
| cdfpbnt/xbhdik | 14-17 | 252836-261005 | 2007.11.04-2008.04.16 | 0.46 | x0k | 15 |
| cdfpbnt/xbhdfm | 18-21 | 261119-271047 | 2008.04.18-2009.01.01 | 1.31 | x0m | 26 |

**Table 2:** Bstntuple data sets and corresponding date ranges

| Period | Run range | Dates | Lumi ($fb^{-1}$) | RS $D^*$ (Millions) | RS $D^*$ yield/lumi |
|---|---|---|---|---|---|
| 0-16 | 138425-258879 | Feb 09, 2002 - Feb 27, 2008 | 2,229,205.26 | 4.21 | 1.9 M |
| 17-21 | 258880-271047 | Feb 28, 2008 - Jan 01, 2009 | 1,492,480.97 | 1.30 | 0.87 M |
| 0-21 | 138425-271047 | Feb 09, 2002 - Jan 01, 2009 | 3,721,686.23 | 5.6 | 1.5 M |

**Table 3:** RS $D^*$ yields per unit luminosity in the old and the new data

| Time bin | Yield(old) P0 to P16 ($10^6$) | Yield (new) P17 to P21 ($10^6$) | Ratio (Yield (new)/Yield (old) |
|---|---|---|---|
| 1 | $0.04 \pm 0.0020$ | $0.01 \pm 0.0012$ | $0.36 \pm 0.04$ |
| 2 | $0.16 \pm 0.0041$ | $0.05 \pm 0.0025$ | $0.33 \pm 0.02$ |
| 3 | $0.33 \pm 0.0058$ | $0.11 \pm 0.0033$ | $0.32 \pm 0.01$ |
| 4 | $0.44 \pm 0.0068$ | $0.14 \pm 0.0040$ | $0.31 \pm 0.01$ |
| 5 | $0.48 \pm 0.0071$ | $0.15 \pm 0.0042$ | $0.31 \pm 0.01$ |
| 6 | $0.46 \pm 0.0069$ | $0.14 \pm 0.0042$ | $0.31 \pm 0.01$ |
| 7 | $0.41 \pm 0.0070$ | $0.13 \pm 0.0039$ | $0.30 \pm 0.01$ |
| 8 | $0.35 \pm 0.0061$ | $0.11 \pm 0.0034$ | $0.30 \pm 0.01$ |
| 9 | $0.30 \pm 0.0056$ | $0.09 \pm 0.0031$ | $0.30 \pm 0.01$ |
| 10 | $0.25 \pm 0.0051$ | $0.07 \pm 0.0028$ | $0.30 \pm 0.01$ |
| 11 | $0.20 \pm 0.0046$ | $0.06 \pm 0.0025$ | $0.30 \pm 0.01$ |
| 12 | $0.16 \pm 0.0041$ | $0.05 \pm 0.0023$ | $0.30 \pm 0.02$ |
| 13 | $0.13 \pm 0.0037$ | $0.04 \pm 0.0020$ | $0.30 \pm 0.02$ |
| 14 | $0.19 \pm 0.0044$ | $0.06 \pm 0.0024$ | $0.30 \pm 0.01$ |
| 15 | $0.12 \pm 0.0035$ | $0.04 \pm 0.0019$ | $0.30 \pm 0.02$ |
| 16 | $0.07 \pm 0.0028$ | $0.02 \pm 0.0015$ | $0.29 \pm 0.02$ |
| 17 | $0.05 \pm 0.0022$ | $0.01 \pm 0.0012$ | $0.31 \pm 0.03$ |
| 18 | $0.04 \pm 0.0022$ | $0.01 \pm 0.0012$ | $0.30 \pm 0.03$ |
| 19 | $0.015 \pm 0.00126$ | $0.004 \pm 0.00068$ | $0.29 \pm 0.05$ |
| 20 | $0.006 \pm 0.00084$ | $0.002 \pm 0.00047$ | $0.31 \pm 0.08$ |
| Total | $4.21 \pm 0.0322$ | $1.30 \pm 0.0135$ | $0.31 \pm 0.00$ |

**Table 4:** The ratio of the RS $D^*$ yields from the old and the new data sets in different time bins

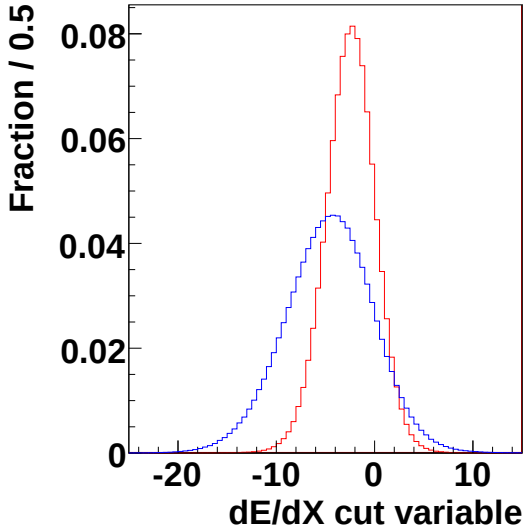



**Figure 3:** $dE/dx$ variable for period 0 to 9 data (blue) and for period 10 to 21 data (red). The change in the distribution is due to the detector changes made to handle higher initial instantaneous luminosity.

## 6.1 $dE/dx$ variable distribution comparison

Data taking with $dE/dx$ was turned off for the innermost COT superlayers due to increased initial instantaneous luminosity. Due to this change the $\frac{dE}{dx}$ distribution for period 10 onwards data is different than the $\frac{dE}{dx}$ distribution for period 0 to 9 data.This comparison is shown in figure **??**. The blue curve shows the $\frac{dE}{dx}$ variable for period 0 to 9 data and the red curve shows period 10 to 21. We use $dE/dx < 0$ cut to get RS signal. Despite the different distributions, both curves have 85 % of the RS events below our cut of 0. The slight difference in efficiency between those two curves will not affect the ratio. There is a new $dE/dx$ calibration but it is not used in Bstntuple data sets and we continue to use the same old calibration. We do not attempt to correct the period 10 to 21 data for the new calibration since it will not make a noticeable difference.

To qualitative compare the old and the new data, we plot normalized $dE/dx$ distributions for period 10 to 16 and for period 17 to 21 data sets on top of each other. Figure **??** shows the plots for $K$ and $\pi$ tracks each with $K$ hypothesis and $\pi$ hypothesis. The red curve shows period 17 to 21 data and blue curve shows period 10 to 16 data and the two curves in all four plots are almost overlapping. This shows that the new data is consistent with the part of old data after the detector changes.

## 6.2 $D^0$ mass distribution comparison

We also compare $D^0$ mass distribution to qualitatively compare the old and the new data. In figure 5, the first plot shows the binned fit for $K\pi$ mass distribution for period 0 to 16 data and for period 17 to 21 data shown in blue and red colors respectively. The second plot in the figure shows a plot in which we plot $\Delta M_{D^*}$ on x-axis and on y-axis we have integral of $K\pi$ distribution in range between $1.865 - \Delta m$ and $1.865 + \Delta m$ GeV/$c$ . The both distributions show no differences.

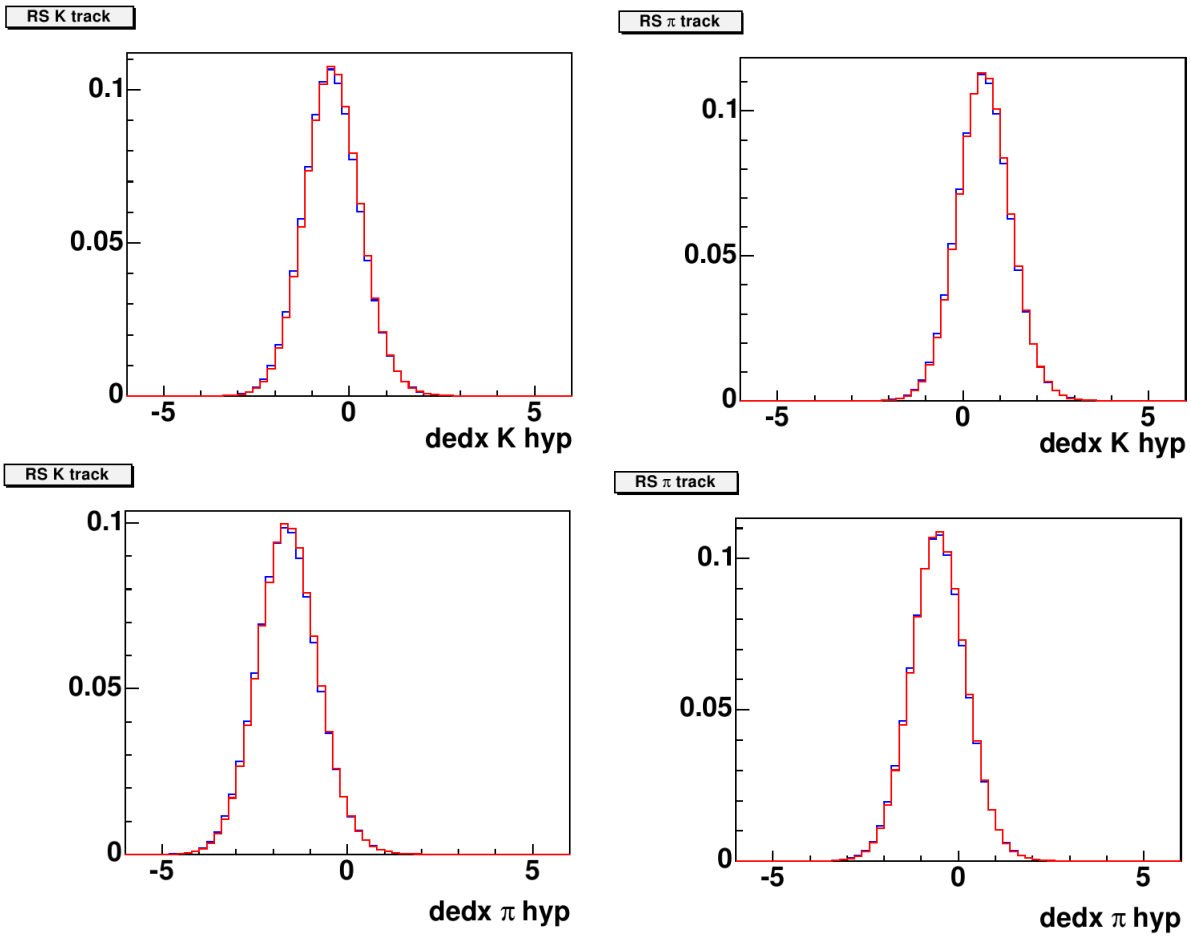



**Figure 4:** dE/dx variable distributions for $K$ and $\pi$ tracks each with $K$ hypothesis and $\pi$ hypothesis. Red: Period 17 to 21 data, blue:period 0 to 16 data. The red and blue curves in all four plots almost overlap and show the consistency of the new data and the old data after the detector change.

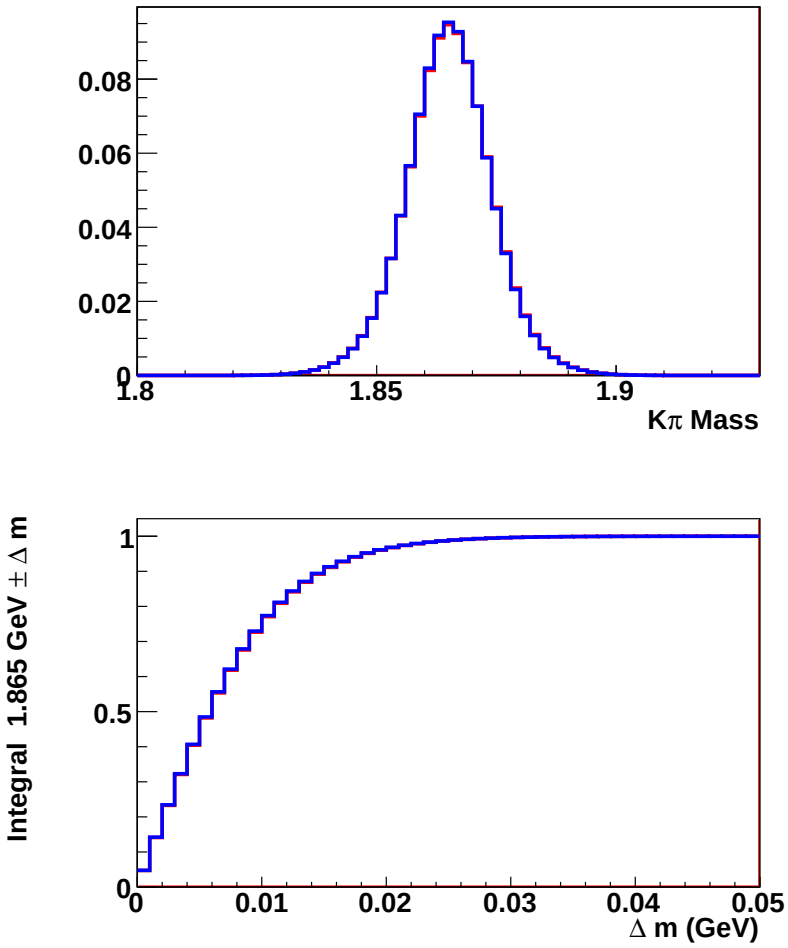



**Figure 5:** The first plot shows m(K$\pi$) distribution comparison. Red curve shows period 17 to 21 data and the blue curve shows period 0 to 16 data. The second plot is also $D^0$ mass distribution comparison in a different way as described in the text.

# 7 Analysis and Results

This section describes the analysis using the previous analysis method without modification. The method is described in detail in earlier CDF notes, so here we will outline only the basic ideas.

As described before, the goal of the analysis is to measure the ratio of wrong sign to right sign $D^*$ s as a function of $D^0$ decay time.

The wrong sign and right sign events are reconstructed from two oppositely charge particle tracks that satisfy two track hadronic trigger requirements. The two tracks are given $K$ and $\pi$ assignments and $D^0$ vertex is formed. A third track requiring $P_T > 0.3$ GeV/$c$ is considered as pion and combined with $D^0$ track to reconstruct $D^*$ vertex. For each event we compute RS and WS invariant mass from $K\pi$ and $\pi K$ interpretation and the data is stored in mini-data sets specific to $D^0 - \bar{D}^0$ mixing.

An opposite mass assignment cut is applied to remove the background from mis-identification of kaon and pion tracks. This appears as RS background in WS signal. The cut is determined from data and removes m($K\pi$ )$\pm$20 MeV/$c$ window from the WS mass distribution.

The particle ionization energy loss cut which provides particle identification (PID) is used to distinguish pion from kaons for all three tracks. The PID cut helps reducing opposite mass assignment background. Other than these two cuts, a transverse decay length significance cut and impact parameter cuts on the tagging pion reduces combinatorial backgrounds. These cuts are summarized below.

- Transverse momentum of soft pion cut for $D^*$ vertex reconstruction : $P_T(\pi_s) \geq 0.3$ GeV/$c$ .
- Opposite mass assignment cut to reduce background from mis-identification of kaon and pion: $|\mathrm{m(K}\pi) - 1.8645| < 0.02$ GeV/$c^2$ . This cut removes RS background from WS signal.
- Particle identification cut on $dE/dx$ variable is used to identify kaon and pion tracks.
- To reduce the background from non $D^*$ decay tracks we apply a series of cuts as: $D^0$ Decay length significance $L_{xy}/\sigma_{xy} > 4$, Tagging pion impact parameter $d_0 < 500\mu$m, and the point of closest approach along the beam line for the tagging pion must be less than 1.5 cm.

For the cut optimization procedure and other details, please refer to CDF notes 7116 and 8879.

The analysis method proceeds as follows: The RS and WS data is divided in 20 time bins in terms of $D^0$ lifetimes. The data in each time bin is further divided in 60 mass difference ($\Delta M_{D^*}$ ) bins. For each $\Delta M_{D^*}$ bin, we fit $K\pi$ mass distribution. This gives the number of $D^0$ vs $\Delta M_{D^*}$ distribution. We fit this distribution to obtain number of $D^*$ s as a function of decay time. The time integrated m($K\pi$ )and $\Delta M_{D^*}$ distributions are used to fix the shapes for signal and backgrounds. To reduce the background from non-prompt $D^*$ s not originating at the primary vertex, we apply Impact parameter cut. The details of the non-prompt $D^*$ correction procedure are given in appendix H.

Thus, we finally obtain the ratio of prompt WS to RS $D^*$ s as a function of decay time. Fitting this distribution with equation 1 gives the values of $x'$ , $y'$ , and $R_D$ and statistical uncertainties on these values. Most systematic uncertainties are canceled in the WS/RS ratio.

Following this method, we obtain the ratio plots for the period 0 to 16 data, Period 17 to 21 data and period 0 to 21 total data. These plots are shown in Figure 6.

Uncertainties on the measured mixing parameters have reduced in the new analysis when we compared period 0 to 21 and period 0 to 16 data. There is around 17 % reductions in uncertainties on $y'$ and approximately 16 % reduction in that for $x'$ . So the average improvement is 16.5%. If we propagate this improvement in mixing significance, we should expect 16.5% improvement over 4.27 $\sigma$-the July 2008 result. This gives expected mixing significance of 4.98 sigma, which is consistent with our previous estimate from the number of RS $D^*$ events. This again tells us that we are right on the edge of 5 sigma significance.

In the next section we try to estimate mixing significance more systematically using toy Monte Carlo studies.

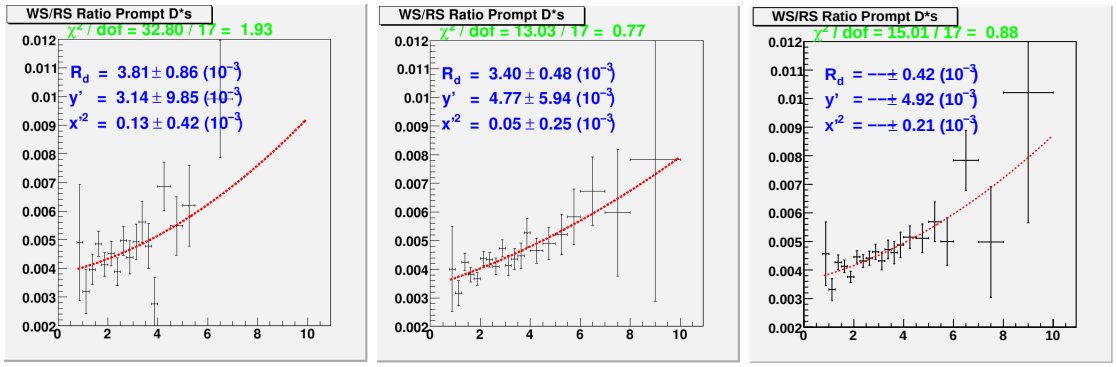



**Figure 6:** Ratio plots: The first plot is the WS/RS ratio from period 0 to 16 data. The second plot is from period 17 to 21 data. The last plot is from period 0 to 21, in which only uncertainties on the mixing parameters are presented to follow the blinding procedure.

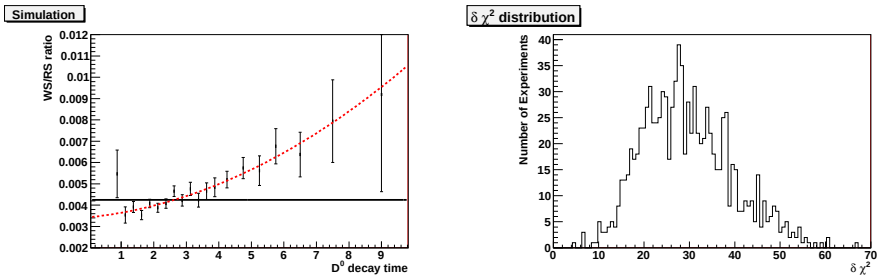



**Figure 7:** Simulated experiment and $\delta\chi^2$ distribution: The left plot show single simulation of WS/RS ratio. The second plot shows the $\delta\chi^2$ distribution from 1000 simulated experiments.

# 8 Toy Monte Carlo simulations of WS/RS ratios and estimate of mixing significance

Here we describe two Monte Carlo simulation methods to estimate the expected significance of $D^0 - \bar{D}^0$ mixing from the statistical uncertainties in the new measurements.

In this simulation method we take the values of $R_D$ , $x'$ , and $y'$ from July 2008 result and compute the value of R(t) from equation 1 for each of the 20 times bins (center of the bin). For each data point we add random noise with Gaussian distribution having standard deviation equal to the uncertainty on the data point in corresponding time bin in Period 0 to 21 data ratio plot. We fit this simulated data again using equation 1 but now $R_D$ , $x'$ and $y'$ are allowed to float. This gives new values of the mixing parameters. We also fit the data with $x'$ and $y'$ fixing to zero. This fit corresponds to no mixing. Then we find the difference between the $\chi^2$ of two fits.

## 8.1 $\delta\chi^2$ method to estimate probabilities

We make a thousand simulations and get a $\delta\chi^2$ distribution. Using the numerical mean of the $\delta\chi^2$ distribution we compute the probability for two degrees of freedom. The calculations are based on incomplete Gamma function P$(a, x)$ where $a$ =ndf/2 and $x = \chi^2/2$. P$(a, x)$ represents the probability that the observed chi-squared for a correct model should be less than the value of $\chi^2$ .

The first plot in figure 7 shows one simulation in which black colored fit is the best fit and red fit corresponds to no mixing with $x'$ and $y'$ fixed to zero. The second plot in figure 7 shows a $\delta\chi^2$ distribution from one thousand simulated experiments. The numerical mean we get from this distribution is 30.053 which correspond to significance of 5.125 Gaussian standard deviations for 2 degrees of freedom. The expected significance is again $\approx 5\sigma$, which is consistent with our previous rough calculations.

## 8.2 Bayesian contour method

In this method we generate the Bayesian probability contours for a single simulated experiment. The Bayesian contours are computed in the same way as it was done for the real data. Figure 8 shows the result of one such experiment. In the resulting plot, the no-mixing point lies on the 5 sigma contour.

# 9 Attempts to improve the standard analysis method

We tried to improve the standard analysis method by adding new cuts and by finding ways to reduce background. Two such studies are mentioned here: The $D^0$ isolation cut and the multiple candidate background study.

## 9.1 $D^0$ isolation variable cut

We tried to add a new cut using $D^0$ isolation variable where the isolation variable is defined as

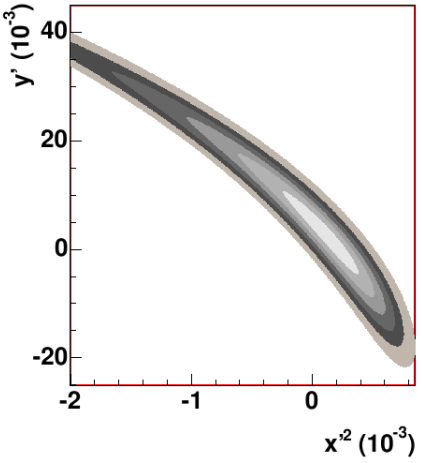



**Figure 8:** Bayesian contours from simulated experiment. The no-mixing point is on $4.92\sigma$ contour.

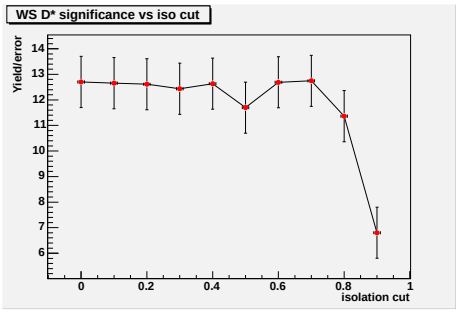



**Figure 9:** WS $D^*$ yield vs. Isolation of $D^0$ cut. The yield is almost flat. It appears that the $D^0$ isolation variable is not useful as a standard cut.

$$I_t = \frac{P_T^{\text{cand}}}{P_T^{\text{cand}} + \sum_{\Delta R<0.4} P_T}$$

where,

- $I_t$: Isolation of tracks.
- $P_t^{\text{cand}}$ : Transverse momentum of candidate particle.
- $P_T$ : Transverse momentum of all other tracks except descendants of candidate particle.
- $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where $\eta$ is the pseudo-rapidity and $\phi$ is the azimuthal angle in $\eta - \phi$ space of CDF II detector. $\Delta R < 0.4, 0.7$ or 1.0.

For a fixed radius $\Delta R = 0.4$, we computed WS $D^*$ syield for 10 isolation cuts using the standard yield technique. Figure 9 shows the WS $D^*$ yield/uncertainty, where uncertainty comes from the statistical error on the first fit parameter of $K\pi$ mass fit and is equivalent to $\sqrt{S+B}$. As we see from the figure 9, the yield is almost flat. This tells us that isolation variable may not be very useful as a standard cut. We will leave this variable for the neural network to learn from it.

## 9.2 Multiple candidate background

We found that in our data we have a few events having the same $D^0$ candidate but different soft pion. To study these multiple candidate events, we processed a total 146 thousand events which were filtered in after applying $\frac{dE}{dx}$ cut, $K\pi$ mass cut, and mass difference cut on a small fraction of

x0h data set. In 146000 events, we found 4745 candidates having one event with the same $D^0$ but different soft pion, 140 events had two such events and no events were found with more than 2 multiple candidates. In summary, there are only about 3% events in our data which have multiple candidates. We decided not to change our method for such a small background. Initially we also began to study the $K\pi$ mass fit comparison using the multiple candidate events and single candidate events; but since the backgrounds is negligible we did not push those efforts too far.

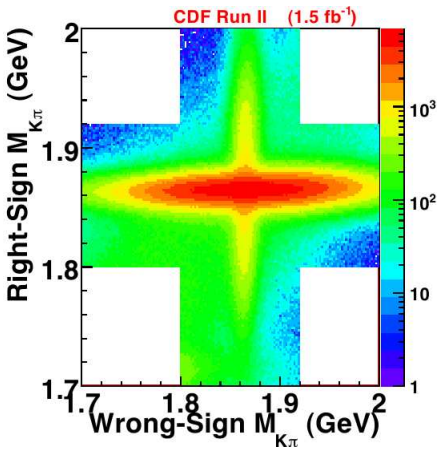



**Figure 10:** Scatter plot showing WS and RS mass distribution. The plot shows WS signal dominated by RS background.

# 10 Artificial Neural Network method

In our efforts to improve the $D^0 - \bar{D^0}$ results we attempted application of Artificial Neural Network. In this section we describe this method.

## 10.1 Neural network method for $D^0 - \bar{D^0}$ mixing analysis

There are basically four types of backgrounds in our data. The best way to understand these background is the wrong-sing vs right sign scatter plot as shown in figure 10. The plot shows wrong sign $K\pi$ mass plotted on x-axis and right sign $K\pi$ mass plotted on y-axis. The four white corners correspond to $D^0 \to KK$ and $D^0 \to \pi\pi$ background which is removed with a $K\pi$ mass cut. The second type of background is from mis-identification of kaon and pion tracks which appears as right sign background in wrong sign signal. The vertical oval in the center is our wrong sign signal which is dominated by right sign background -the horizontal central oval. We apply the opposite mass assignment cut to remove most of this background. It is the m($K\pi$ )mass cut that removes a 40 MeV/$c$ window centered at the $D^0$ mass 1.865 GeV/$c^2$ . We also apply the particle identification cut (dE/dx variable) to identify kaons and pions which also helps removing this background. The third type of background is from random pion in $D^* \to D^0\pi$ reconstruction. Mass difference cut $\Delta M_{D^*} < 20$ MeV/$c^2$ helps to remove most of this background. The fourth background is combinatorial background which arises from any other particle coming from the primary vertex. Particle identification cut helps to remove this background to some extent. Even after applying all possible cuts our data can still have some background from random pion and combinatorial background. We attempt to use artificial neural network cut method to reduce this background.

Our hope that a neural network can reduce combinatorial and random pion background is based on the neural networks ability to learn a complicated non linear multidimensional discrimination function in hyper surface of its input parameter space. Neural network makes such a precise signal and background discrimination possible by reducing multidimensional cuts to one dimensional neural network output cut. This may bring improvement in our signal significance over the standard cut optimization method.

The artificial neural network algorithm finds the non-linear discrimination function by adjusting hundreds of internal parameters. The algorithm can be graphically visualized as a set of input nodes, middle layers (called hidden layers), connecting links called synapses and the output nodes as shown in figure 11. The algorithm starts with a random weight assigned to each synapse. At each subsequent layer, the linear combination of the sigmoid functions is computed. The sigmoid function is defined as $\frac{1}{1+e^w}$ where $w$ is the weight on an input synapse. The output of sigmoid function lies between -1 and +1. The linear combination of these sigmoid functions will be a non-linear function. If the problem is classification problems as in most particle physics problems, then only one output node is required. The output at this node will be between -1 or +1. If the neural network is used to find probability

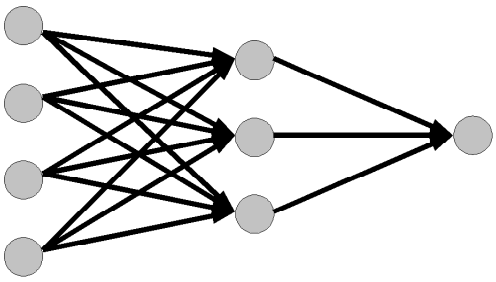


**Figure 11:** A typical artificial neural network has input nodes, synapses, middle layer with hidden nodes, and output node. The network finds a discrimination function by adjusting weights on each of the synapses.

distribution function, then more than one output nodes may be required. There are no specific rules to decide the number of hidden layers and the number of nodes in each hidden layer. In most cases, only one hidden layer is sufficient and a rule of thumb says that the number of nodes in the hidden layer should be equal to twice the number of input variables.

In the training stage, a set of input values from a known data sample is given at the input nodes of the neural network and the target output is set to +1 for signal and -1 for background. If the neural network output at the end node is different from the target output then the difference is back-propagated and the weights at all synapses are adjusted. Several error minimization algorithms are available and choice of the algorithm depends on the type of the problem. The learning process is repeated for all sets of inputs until the weights on the synapses become stable. At this stage the network is said to be trained and can be used for signal and background separation in new data.

For our purpose we used a commercial neural network software product called NeruoBayes [14] which is made available for CDF users under special license and is available only of fcdflnx machines. We performed this analysis at fcdflnx4.fnal.gov machine at /cdf/spool/nageshk/NBD0/ folder.

## 10.2 Training sample preparation

For neural network training we need signal and background samples. Many neural network analyses use Monte Carlo generated training samples. But since in our case it impossible to simulate all kinds of backgrounds,the background samples taken from real data would be much more accurate. For this reason we decide to extract our training samples from real data.

To get the signal data we applied standard $\frac{dE}{dx}$ and impact parameter cuts and require $K\pi$ mass to be $|m(D^0\text{RS})| < 20$ MeV/$c$ and $D^*$ mass difference $\Delta M_{D^*} < 20$ MeV/$c$ . To extract background sample we applied standard $\frac{dE}{dx}$ cut but not the impact parameter cut and require having wrong sign $K\pi$ mass either between 1.8 and 1.835 GeV/$c^2$ or between 1.895 and 1.928 GeV/$c^2$ windows. Wrong sign mass difference $\Delta M_{D^*}$ (WS) $> 30$ MeV/$c^2$ cut was required to remove most signal events keeping random pion background. In addition, we apply right sign mass cut of 70 MeV form RS $D^0$ mass to eliminate opposite mass assignment background. The final regions in the mass distributions from where the signal and background are extracted can be best illustrated by figure 12. The central horizontal distribution shows the signal region and four corner boxes show the background region. Even from a very narrow signal region we get a large number of signal events compared to background events. Typically we get signal 8 to 10 times larger than background.

## 10.3 Input parameters and settings

We followed a rule that the variables used to obtain the training data such as $D^0$ mass, mass difference, dEdx and impact parameter of $D^0$ , will not be used as inputs for the neural network for the reason that there will not much for the neural network to learn from these variables. For each variable available in our mini-data set, we studied the distribution for signal and background and decided to keep the variable as NN input if the difference between the two distributions is large. The final 11 variables are summarized below.

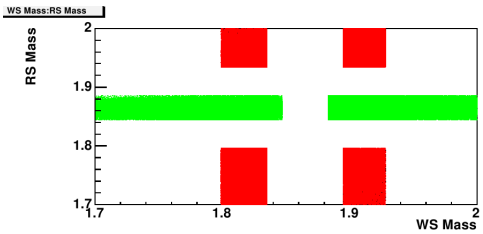



**Figure 12:** Signal and background samples used for training. The red corners show the region from where background sample is extracted. The central horizontal bar shows the signal region.

For kaon, pion, and the tagging pion we used impact parameter, transverse momentum, and Z0 variable (the distance of closes approach along beam line). For $D^0$ meson we used transverse and z components of decay length significance defined earlier.

We also used isolation variable for $D^0$ meson where the definition of isolation variable is given earlier. In addition we used eta variable for the tagging poin only. The signal and background distributions for kaon and pion eta variables were not significantly different and so they are not used as NN inputs. Apart from these 11 variables, NeuroBayes require one extra unnamed input variable for internal use.

The plots for the signal and background distributions for these variables are shown in Appendix E. As we see from these plots, some variables such as transverse momentum of kaon and pions have large differences in their signal and background distributions. We expect Neural network to learn more from these inputs than others. The output of neural network training provides information about the final weights assigned to each input synapse. This information is consistent with our expectation in that the synapses corresponding to kaon and pion transverse momentum inputs were always given relatively higher weights. Interestingly, we found that the $D^0$ isolation variable which we found not very useful in the standard analysis is also given significant weight and removing it from the inputs reduced the neural network's discrimination power.

The script Teacher.cc and Expert.cc are compiled to form Teacher.exe and Expert.exe executable. Since the purpose of the neural network in our case is to separate signal from background, this becomes a classification problem and so the neural network is designed to have single output node with output varying between -1 and +1. The default setting of NeuroBayes program uses only 80 % of the training sample events for training and leaves 20 % for testing. This setting is kept unchanged. Number of epochs is set to 200, where epoch is the number of iterations or events after which the weights are updated. This number is obtained from trial and error to give maximum yield at highest learning speed. The number of training iterations is set to 15. The error minimization method used is BFGS which stands for Broyden, Fletcher, Goldfarb, Shanno learning method. The other default settings are unchanged. For 11 input variables we set 22 nodes in hidden layer and one node in output layer. The topology of this network is shown in appendix B.

The teacher executable reads events from signal and background data files prepared for training. Only 1.5 million total (signal + background) events are allowed for training. Since we have very few background events compared to signal events, we wanted to use all available background events. For the period 0 to 21 data, we get total 5.3 million signal events and 0.65 million background events. We use all 0.65 million background events and only 0.85 million signal events to maintain the total number 1.5 million. The signal to background ratio is then 1.3.

To train the neural network we provide the neural network with the signal events with target output value set to +1.0 and background events with the target output value set to 0.0. The neural network starts training process and after the training is over, the expertise(which include final weights on synapses) is stored in file D0expertise1.nb. The executable also produces histograms of output node value (NN output) for each event in signal and background training sample. Figure 13 shows the training output in which signal is shown in red and background is shown in black. We observe that signal gradually increases and background decreases with NN output from -1.00 to +1.00. From figure 13 we expect to get maximum integrated signal and minimum integrated background if we require the NN output to be slightly greater than 0.0.

The Expert executable takes D0expertise1.nb file and new data which we want to classify as inputs.

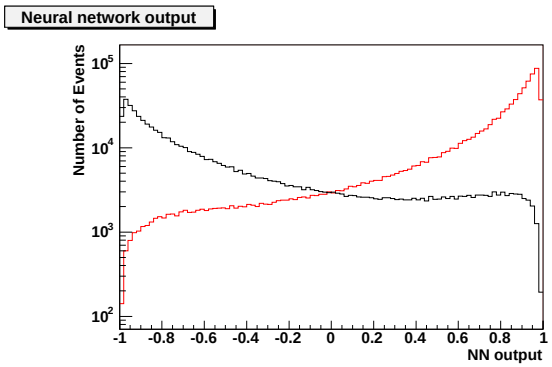



**Figure 13:** Neural network training output: Signal is show in red color and background in black color. Signal gradually increases and background decreases with NN output. Maximum signal significance is found at NN cut $> 0.10$.

| Analysis | WS $D^*$ syield | Uncertainty $\delta$ | Ratio:Yield/$\delta$ |
|---|---|---|---|
| Standard method | 24056.82 | 374.40 | 64.25 |
| NN cut $> 0.0$ | 24363.45 | 376.61 | 64.69 |
| NN cut $> 0.05$ | 238888.9 | 368.61 | 64.80 |
| NN cut$> 0.10$ | 23481.15 | 360.88 | 65.06 |
| NN cut $> 0.15$ | 22838.22 | 352.92 | 64.71 |
| NN cut $> 0.20$ | 22133.21 | 344.70 | 64.20 |

**Table 5:** WS $D^*$ significance using the standard method and various neural network cuts. The maximum signal significance is at NN cut $> 0.10$

The executable takes the input variables from the new data in the same order as they were given for training and computes the NN output value for each event using the saved expertise.

We apply a neural network cut for which we expect to get maximum signal to background ratio and the data passing that cut is used to find WS $D^*$ yield using the yield technique described previously. The yield technique gives the number of WS $D^*$ sas a value of the first fit parameter. The uncertainty delta on the first parameter can be considered as equivalent to $\sqrt{(S+B)}$, where S is signal and B is background. We use the ratio of yield/$\delta$ to compare the Neural network result with that from the standard analysis method. Table 5 shows the WS $D^*$ significance we get from the standard method and various neural network cuts. We see from the table that we get maximum signal significance for neural network cut ¿ 0.10. This is close to our expected NN cut value.

As we see, the neural network gives only 1.5 % improvement over the standard method. When we started with lesser data, we got 5 to 6 % improvements over the standard method but after adding the total data from period 0 to 21 the improvement drops to 1.5 %. From this we conclude that this is a negligible improvement and may turn out to be just a statistical fluctuation. Nevertheless, we tried running the complete analysis to generate ratio plots using a neural network classified data to see if the result looks sensible. The ratio plot was sensible and uncertainties on the values of the mixing parameters were slightly reduced as expected. This means that, although at this time neural network does not give major improvements it has potential to yield better results. In future we may try to use realistic Monte Carlo data or throw in more input variables to bring more improvements.

# 11 Lump study

The CDF note 8879 describes the extra lump that appears in RS $K\pi$ mass fit at 1.83 GeV/$c$ (section 3.2.4). An extra Gaussian in the quadratic background fit function was required which improved $\chi^2$ per number of degrees of freedom. We want to study if the significance of lump increased or

| $K\pi$ mass Fit | $\chi^2$ /ndf for Period 0 to 16 data | $\chi^2$ /ndf for Period 0 to 21 data |
|---|---|---|
| Without Gaussian | 852.65/50 = 17.05 | 983.5/50 = 19.67 |
| With Gaussian | 129.60/48 = 2.70 | 137.59/48 = 2.86 |

**Table 6:** $\chi^2$ /ndf for with and without extra Gaussian for the old and the new data.

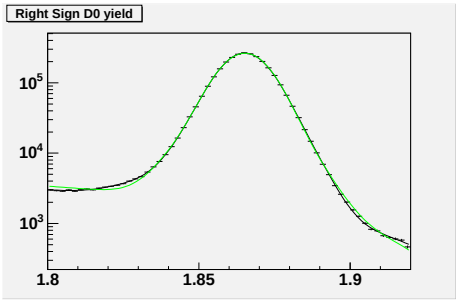



**Figure 14:** $K\pi$ mass fit with and without extra Gaussian in addition to quadratic function for background. The green curve show fit without Gaussian.

decreased with more data, to see if it is a real effect. We fit the RS $K\pi$ mass distribution with and without extra Gaussian for period 0 to 16 data and for period 00 to 21 data separately. Table 6 lists the $\chi^2$ /ndf for both the data sets.

The difference in $\chi^2$ /ndf of two fits for period 0 to 16 data is 17.05-2.70 = 14.35 and that for period 0 to 21 data is 19.67-2.86= 16.81. The increase in $\delta\chi^2$ in the new data set suggests increase in the extra lump with time. Figure 14 shows the two fits.

We tried to search the cause of the extra lump through some realistic Monte Carlo simulation studies. One idea was to test semileptonic $D^0 \to Kl\nu$ decay. For a quick look we generated 3000 $D^* \to D^0\pi^+$ , $D^0 \to K^+\mu^+\nu_\mu$ simulated events using Pythia model. We find that the $D^0$ mass reconstructed from this decay falls below 1.8 GeV/$c^2$ and after applying mass difference $m_{D^*} - m_{D^0} - m_\pi$ cut, it cannot cause a lump at 1.83 GeV/$c^2$ where it was found in the real RS $D^0$ data. Similar results found in other decay chains we tried. None of them had mean invariant mass close to 1.83 GeV/$c^2$ . The cause of the extra lump is still a mystery.

# 12 Systematic Uncertainties

The systematic uncertainties for the 1.5 $fb^{-1}$ analysis are described on pp. 26 - 27 of CDF note 8879. This section will repeat some of the descriptions, and show the differences with the current analysis.

Most of the systematic uncertainties for this result are already included in the values returned from the ROOT fitter. The CDF Statistics Committee points out that a fit to data will include uncertainty about the fit shapes, as long as those parameters are floating. Adding more uncertainty to the function shape would be a form of double counting. This is not true if the parameters are fixed beforehand (from Monte Carlo or a fit to the time-summed data). We examined the possible sources of systematic uncertainty, described in the following sections. Two additional terms are added (in quadrature) to the ratio uncertainties, due to the mass difference background function and the extra RS "lump" in the $K\pi$ mass fits.

Most of the studies use "good" RS $D^*$. This means RS candidates with a mass difference near the D* peak ($4 < \Delta m < 9$ MeV) and an (unsigned) impact parameter less than 60 $\mu$m, for all decay times. The number of "good" RS $D^*$ is determined from a fit.

## 12.1 Signal Shapes

The $D^0$ signal shape is determined from a fit to "good" RS $D^*$. This shape is fixed, and used as the signal shape for all subsequent RS and WS $K\pi$ mass fits. The $D^*$ signal shape is determined from a fit to the time-integrated mass difference plot.

The mass distributions of the $D^0$ and $D^*$ candidates well matched by the signal functions we have chosen. The true signal distributions are the same for RS and WS, since they have identical kinematics ($D^0 \to K\pi$). We use the same signal shapes for RS and WS. While there might be systematic uncertainties in the RS or WS signal yield separately, we expect it to be a common multiplicative factor for both WS and RS, which will cancel in the WS/RS ratio.

## 12.2 $K\pi$ Background

We expect an effect due to uncertainty in the background shapes for the $K\pi$ plots. Since the signal-to-background ratio is different for WS and RS, the ratio could be affected as well. However, since we allow the quadratic polynomial parameters to float for every $K\pi$ fit, the systematic uncertainty is already included in the signal uncertainty returned by the fitter. (The signal uncertainty is larger than a similar unbinned likelihood fit, or binned fits where the background shape is fixed a priori.)

Other analyses usually assume that the distribution in one variable (like $K\pi$ mass) is independent from another variable (such as mass difference). We do not make that assumption, so both the shape and amplitude of the $K\pi$ background can change depending on the mass difference and decay time.

Another point to consider is that we can't have an alternate background function as a cross-check. We are using a quadratic polynomial to fit the background. The $K\pi$ fits agree with the data points, with an average $\chi^2$/dof of 1.0. Any differences between our distribution and the "true" background from data must be small. Any difference in shape could be treated as a Taylor expansion, which would also make it a polynomial (and included in the fit). We have tried including a third order term to the polynomial. There was no noticeable changes in the signal yields, and all fits that we looked at had the third order term set to zero.

## 12.3 Mis-assigned Background Correction

There is no additional systematic uncertainty due to correcting for mis-assigned RS $D^*$s that show up as background in the WS plots. The correction is given by Eq. 4 on page 14 of CDF note 8879. Our kinematic and particle identification cuts greatly reduce the MRS background in the WS $K\pi$ mass plots, and the $K\pi$ mass distribution for this background only has slight curvature under the signal. For the previous result, the change in the WS/RS ratio due to not using the MRS correction was smaller than the uncertainty returned by the fitter, and the uncertainty on the correction itself (which would be the systematic uncertainty) was O(1 %).

The efficiency of the particle ID cut, determined by the change in the number of "good" RS $D^*$ with and without the PID cut, is now $85.99 \pm 0.07\%$, compared to $86.58 \pm 0.08\%$ for our previous result. The MRS correction in the analysis method has been modified to take this into account. Since

the PID cut will have affect RS and WS signal equally, the cut does not affect the ratio. If the cut efficiency changed greatly, for instance O(10 %) instead of the measured 0.6 %, we would consider re-optimizing the cut value to get better signal-to-background discrimination.

## 12.4 Non-prompt $D^*$ Correction

We do not assign any additional uncertainty due to correction for non-prompt $D^*$s, Our results already include uncertainties from the fits to the RS IP distributions in the different time bins. (For this study, "good" $D^*$ do not have the IP requirement, but use sideband subtraction with the $K\pi$ mass.) For long decay times, the number of $D^*$s from $B$ decays is a significant fraction of the total number of $D^*$s.

We do vary the gaussian width of the non-prompt distribution, which has been parameterized. The width is changed by $\pm 1\sigma$, and the change in the corrected ratio observed. The change was negligible compared to the fitter uncertainty.

## 12.5 Mass Difference

The background distributions for the mass difference plots are described by a power term ($\Delta m^x$, where $x$ is approximately 0.5), which is determined from fits to the RS and WS mass difference plots.

We do assign an additional uncertainty due to fixing the mass difference background power term from the time-integrated fit. As a check, we changed the power term by $\pm 1\sigma$, where $\sigma$ is the uncertainty from the fit to the time-summed data, and then observed the change in the WS/RS ratio in each time bin compared to the best fit power term. The RS background is very small compared to the signal, so caused a negligible change in the ratio. The WS signal was refit for the twenty decay time fits, and the average change in the ratio was $|\Delta R/R| = 0.46\%$. (The systematic uncertainty used for the 1.5 $fb^{-1}$ analysis was $|\Delta R/R| = 0.80\%$.)

## 12.6 $K\pi$ Lump at 1.83 GeV

An extra Gaussian term is added to the RS $K\pi$ fit, as described on pp. 16-18 of CDF note 8879. This is background in the RS only. Since we have not been able to determine a decay that would produce it, we have added a systematic uncertainty to be on the safe side. Using the time-integrated "good" $D^*$, there is a change in the WS/RS ratio of $\Delta R/R = 0.07\%$ when this "lump" is treated as part of the $D^0$ signal shape. (The value of $\Delta R/R = 0.34\%$ from the 1.5 $fb^{-1}$ analysis was overestimated.) A systematic uncertainty of the same amount will be added to the WS $D^*$ yields for each time bin.

Since we had to recalculate the systematic uncertainty for this source, it might be useful to go over the reasons why it is included. Figure 15 shows the RS and WS time-integrated fit for "good" $D^0$s. RS fit has 9 free parameters: 6 for the double Gaussian signal, and 3 for the (quadratic) background. The WS fit has 4 free parameters: 1 for the signal amplitude, and 3 for the background. The WS signal shape is determined from the RS fit. The RS fit, besides having an enormous $\chi^2$, shows structure in the pull distribution.

To improve the RS $K\pi$ fit, we added an additional Gaussian. The parameters for that "lump" Gaussian are a mean of 1.83 GeV, width of 10 MeV, and an amplitude of 0.6 % compared to the signal Gaussian. Figure 16 shows the RS and WS fits when this "lump" is determined from the RS fit, and treated as part of the signal Gaussian. (The WS fit has no additional parameters, because the signal shape is determined from the RS fit.) The improvements in the RS fit pull distribution and $\chi^2$ are compelling reasons for including the extra Gaussian in the $K\pi$ fit.

We also tried a fit where the extra Gaussian is treated as a background term. The mean and width are the same for both RS and WS, but the amplitudes are allowed to float. Figure 17 shows those RS and WS fits. The WS fit $\chi^2$ has improved by 25.4 units with the addition of a parameter for the "lump" amplitude (independent of the RS). This excludes the hypothesis that the "lump" is signal, at the $5.0\sigma$ level. In fact, the WS fit put the amplitude to zero. (The ROOT fit was not allowed to have a negative amplitude.) When looking at the WS $K\pi$ fits divided by mass difference or decay time, the "lump" amplitude was always consistent (within errors) with zero. This led to the decision to not have the "lump" in the WS $K\pi$ fits, and to have the lump amplitude be an independent background parameter for RS $K\pi$ fits.

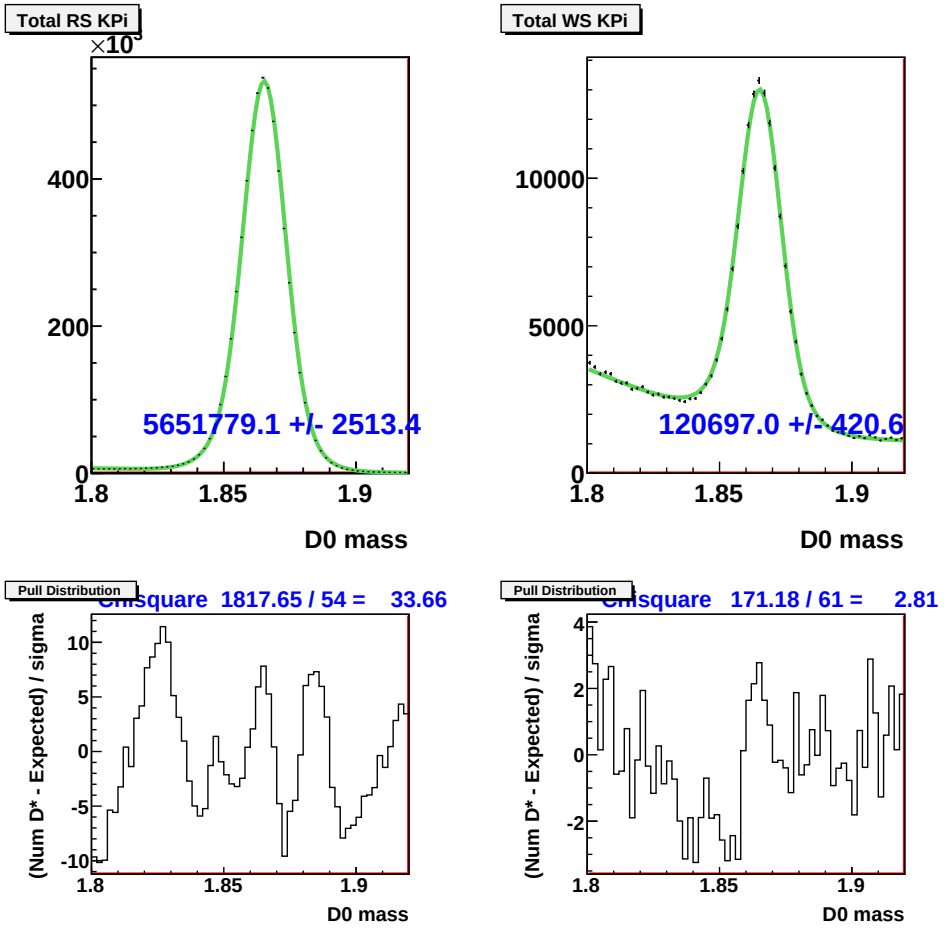



**Figure 15:** $K\pi$ mass for "good" RS (left) and WS (right) candidates, fit without an extra Gaussian (lump).

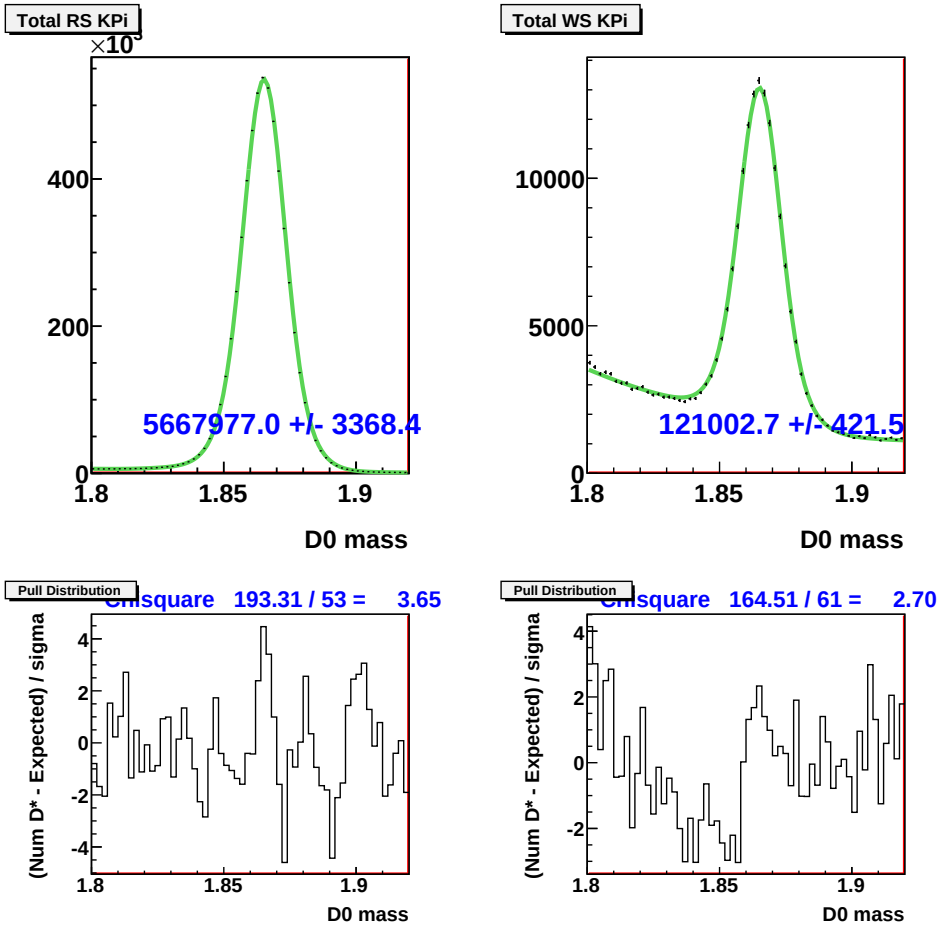



**Figure 16:** $K\pi$ mass for "good" RS (left) and WS (right) candidates, fit with an extra Gaussian (lump) that is treated as part of the signal shape.

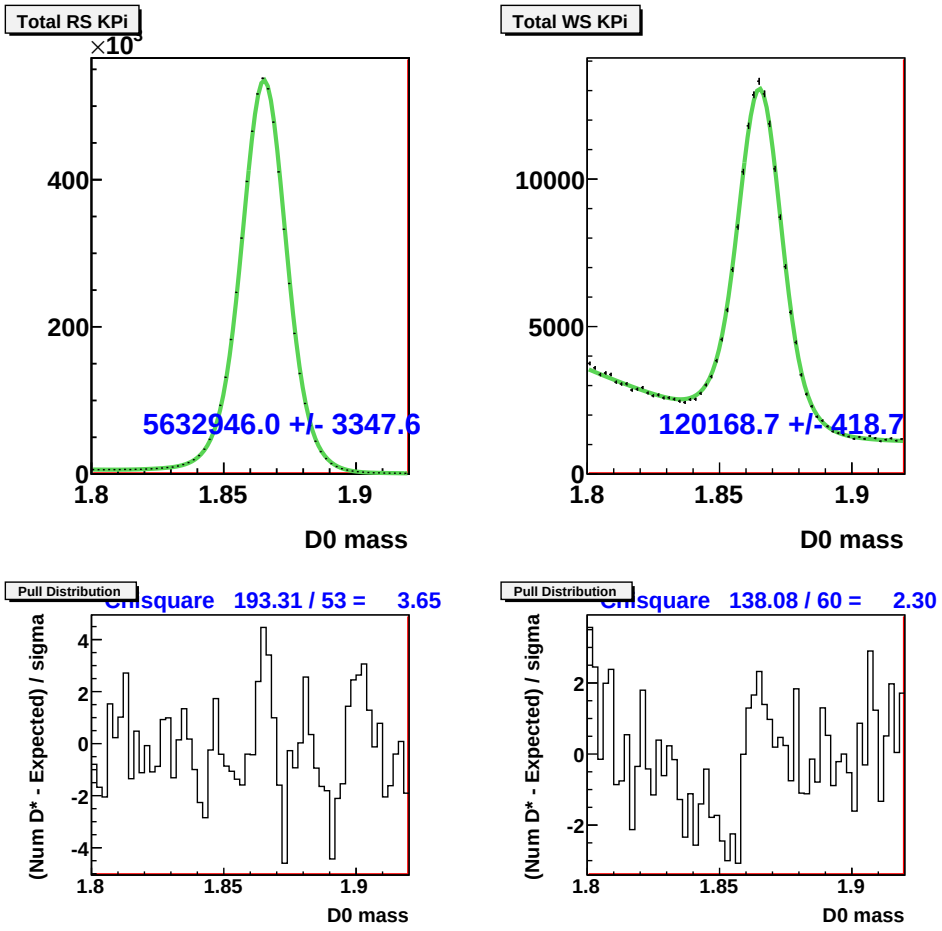



**Figure 17:** $K\pi$ mass for "good" RS (left) and WS (right) candidates, fit with an extra Gaussian (lump) that is not considered signal. The RS and WS fits have independent amplitudes for the extra Gaussian, although they have the same mean and width. The WS fit has zero for the "lump" amplitude.

# 13 Conclusion

We made four major attempts to improve our previous (standard) analysis method. We tried to introduce a new cut using $D^0$ isolation variable which we did not find very useful in the standard analysis method but found useful in Neural Network analysis. We found that the background from multiple candidate events is negligible and it may not be worth pursuing it further.

Results from the application of artificial neural network were comparable to the standard analysis method results. This confirms that our cuts are optimal.

In summary, we do not propose any changes in the standard analysis method at this time, and present new result mainly from the addition of new data. We now have total 3.7 $\text{fb}^{-1}$ integrated B-Physics luminosity data available. The new data is qualitatively consistent with our old data that was used in July 2008 result but we found that the RS $D^*$ yield is decreasing with time. We presented the new WS/RS ratio plots using the total available data. We also recalculated the systematic uncertainties. The following are the uncertainties on the mixing parameter from using all available data.

$$R_D = - - - \pm 0.42(10^{-3}), y' = - - - \pm 4.92(10^{-3}), x'^2 = - - - \pm 0.21(10^{-3})$$

The addition of new data reduces uncertainties on the mixing parameters from previous measurements by 16 to 17 %.

We tried to estimate the expected mixing significance from rough calculations using RS $D^*$ yield and from more systematic Monte Carlo simulations using $\chi^2$ difference method and Bayesian contour method. From each of these methods we see that we may expect to exclude no-mixing with approximately 5 Gaussian standard deviations. We are standing right on the edge from where we may be able to observe $D^0 - \bar{D^0}$ mixing as single experiment, if it exists in nature. We also have a good chance of providing the world's most significant measurement of $D^0 - \bar{D^0}$ mixing parameters.

# Appendices

## A $K\pi$ mass fit comparison for the old and the new data

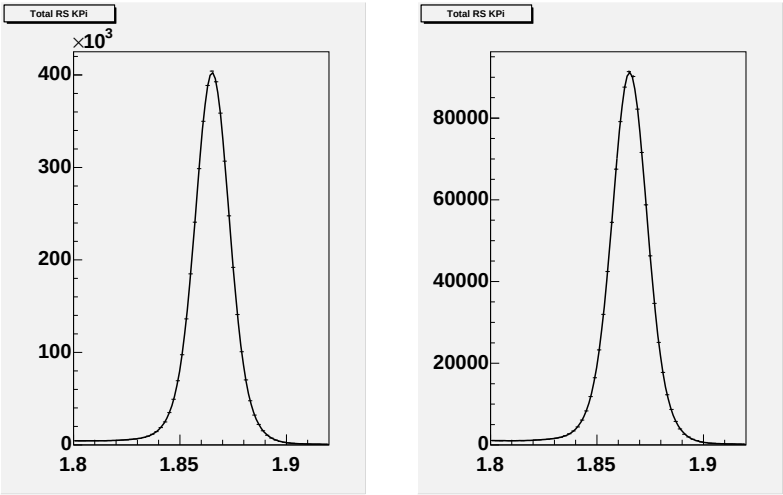


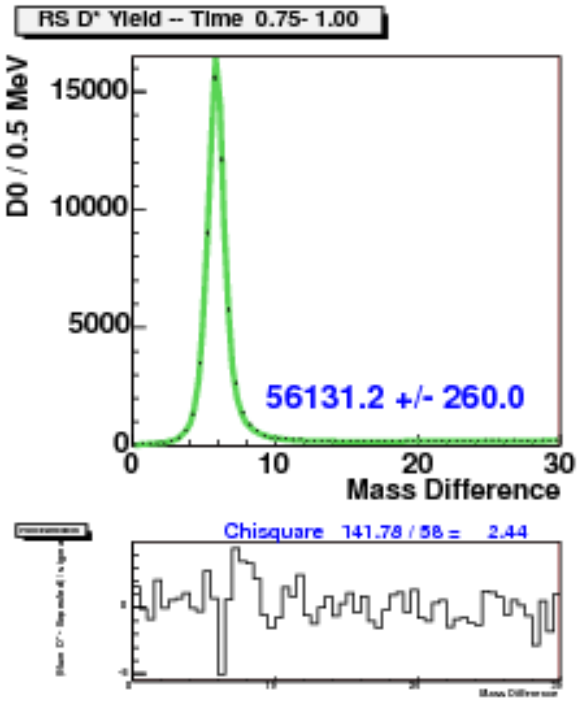

**Figure 18:** Time integrated $K\pi$ mass fits using the standard double Gaussian fit function. Left plot is for period 0 to 16 data and the right plot is for period 17 to 21 data. The first three parameters are listed in table 7

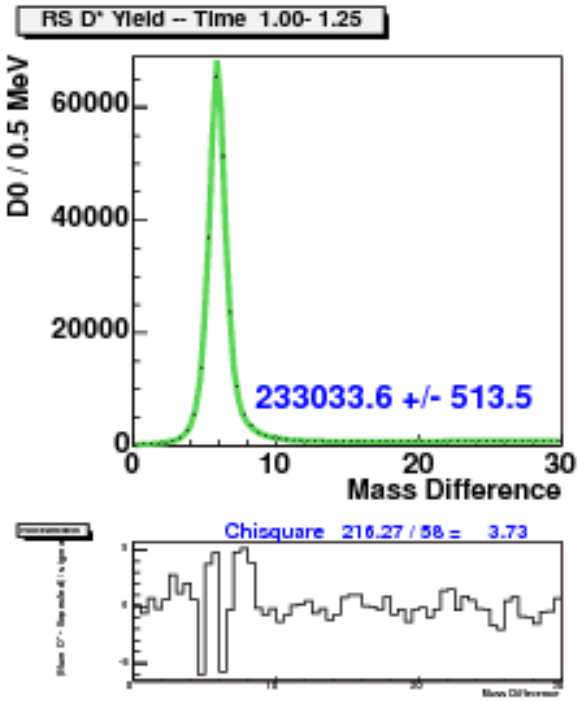
| Parameter | Period 0 to 16 | Period 17 to 21 |
|---|---|---|
| Number 1 | $4.20695 \pm 0.001352(10^6)$ | $1.295182 \pm 0.001352(10^6)$ |
| Mean 1 | $1.865214 \pm 0.000021$ GeV/$c^2$ | $1.865385 \pm 0.000021$ GeV/$c^2$ |
| Sigma 1 | $0.007192 \pm 0.000044$ GeV/$c^2$ | $0.007369 \pm 0.000044$ GeV/$c^2$ |
| Number 2 | $0.366765 \pm 0.013326$ | $0.383522 \pm 0.013326$ |
| Mean 2 | $1.864871 \pm 0.000064$ GeV/$c^2$ | $1.864961 \pm 0.000064$ GeV/$c^2$ |
| Sigma 2 | $1.619948 \pm 0.008501$ GeV/$c^2$ | $1.607657 \pm 0.008501$ GeV/$c^2$ |

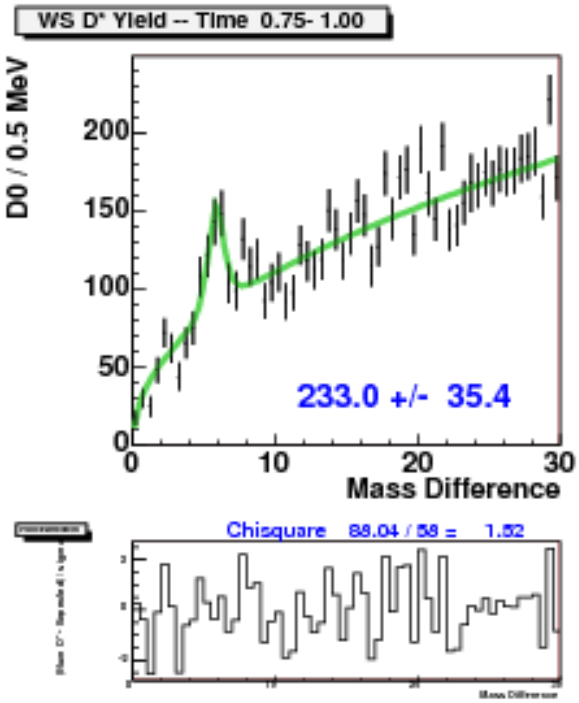
**Table 7:** The comparison of the first fit parameters of $K\pi$ mass fits for the old and the new data. The mean and the sigma for one Gaussian are shown. There is no shift in $D^0$ mass distribution in the new data.

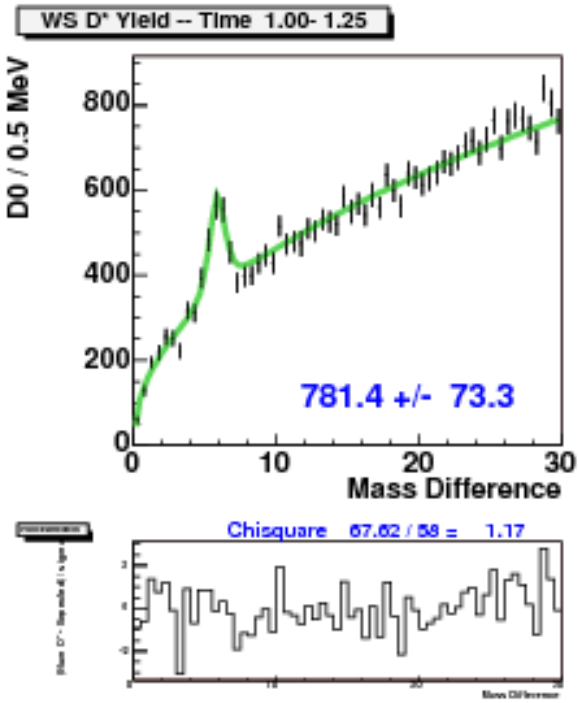
## B Time slices1:period 0 to 21 data

Appendix B:Title

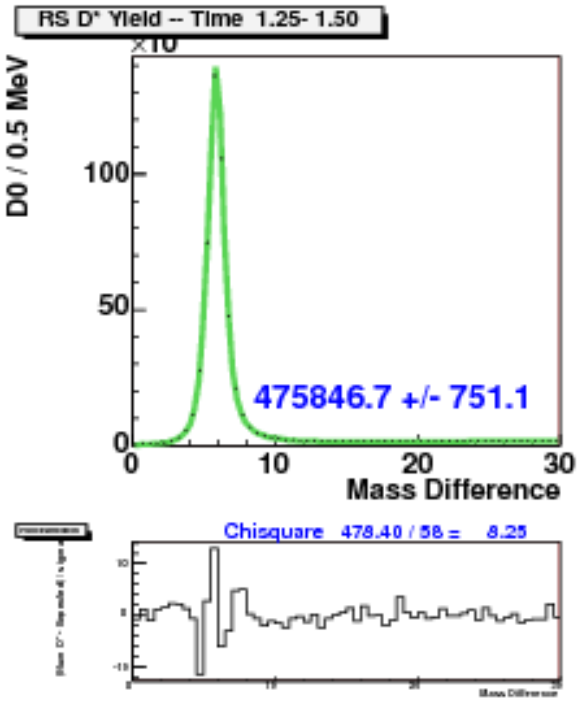


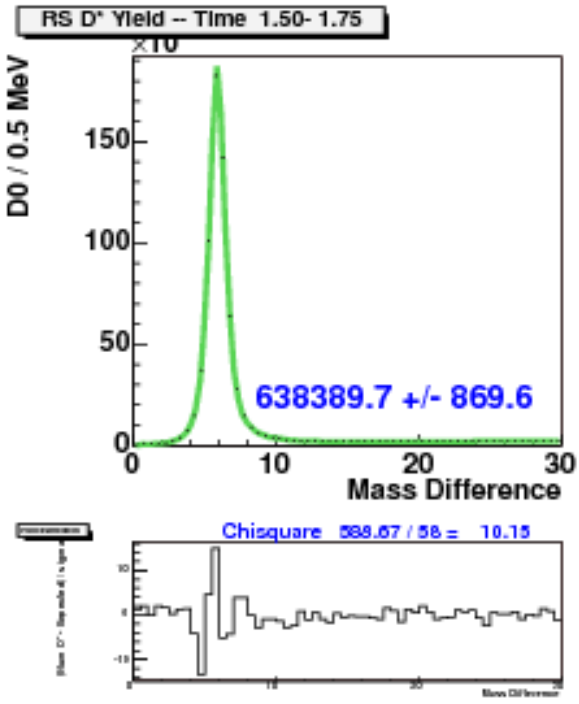



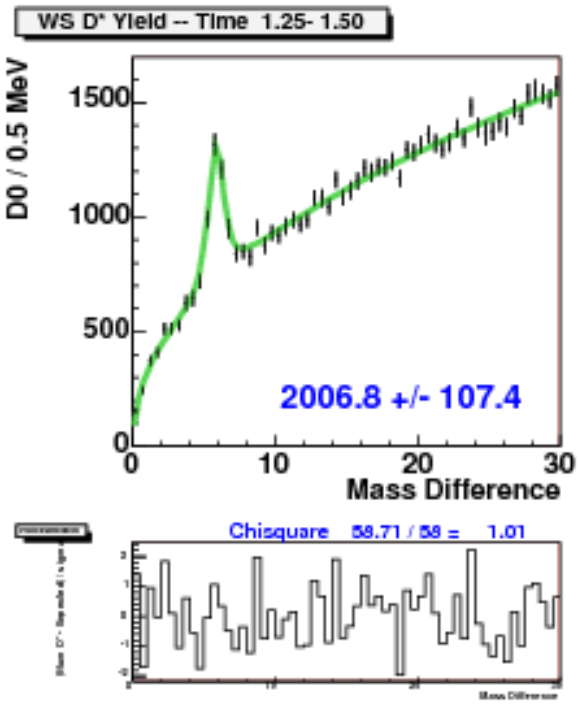



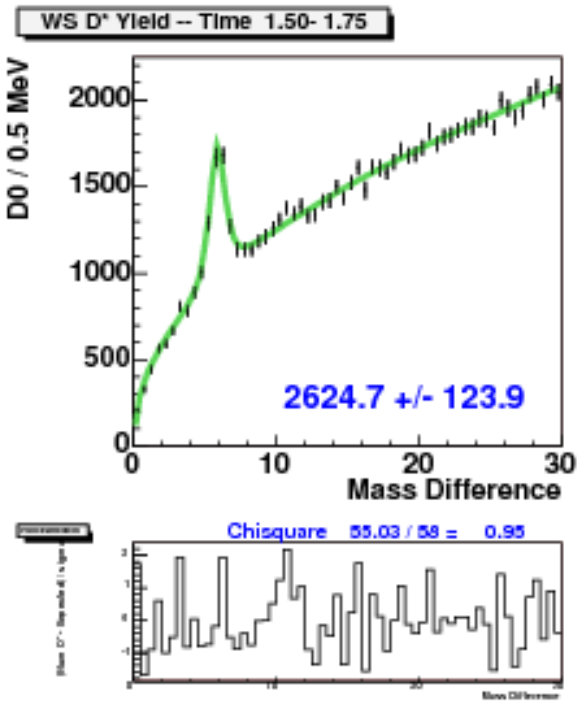



Appendix B:Title

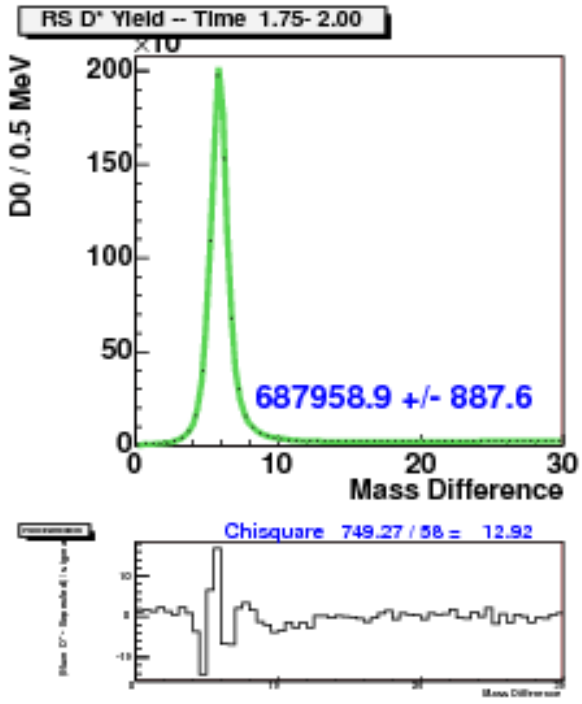


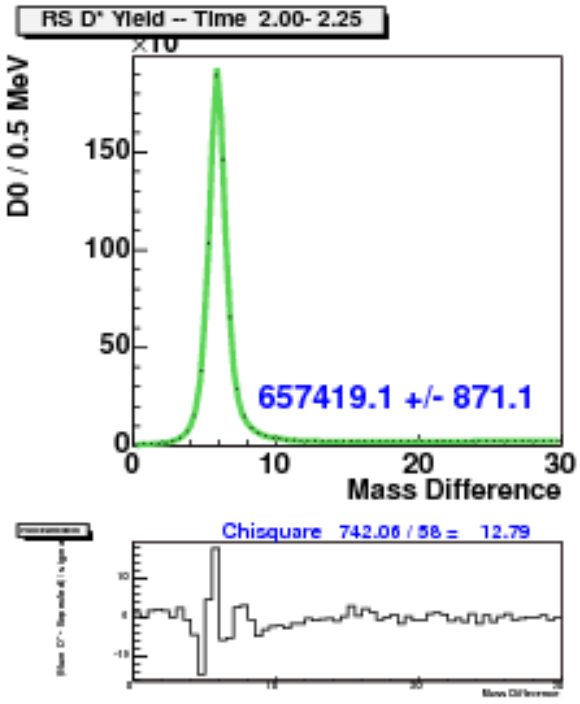



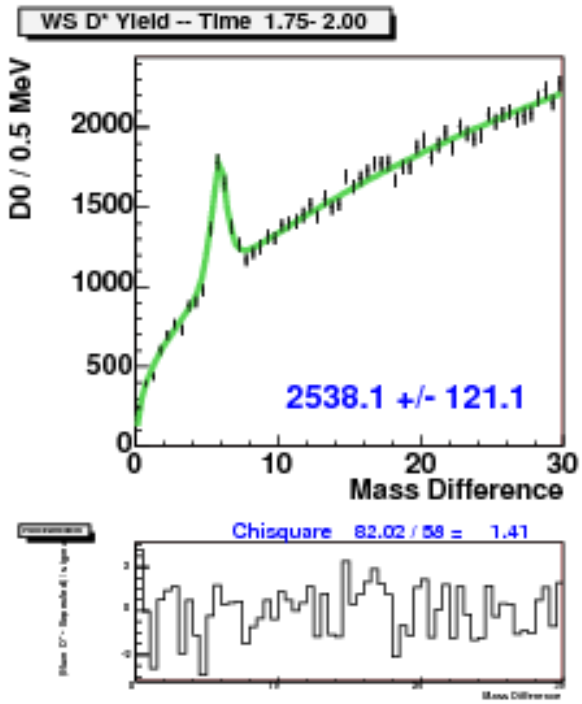



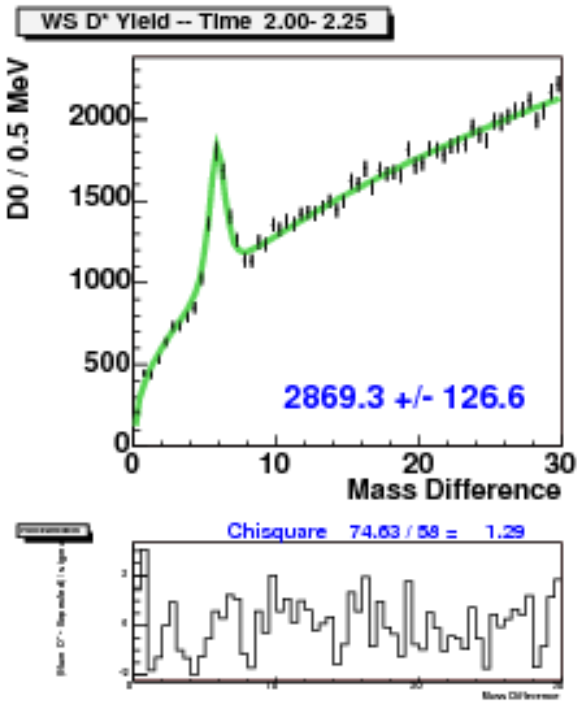



Appendix B:Title

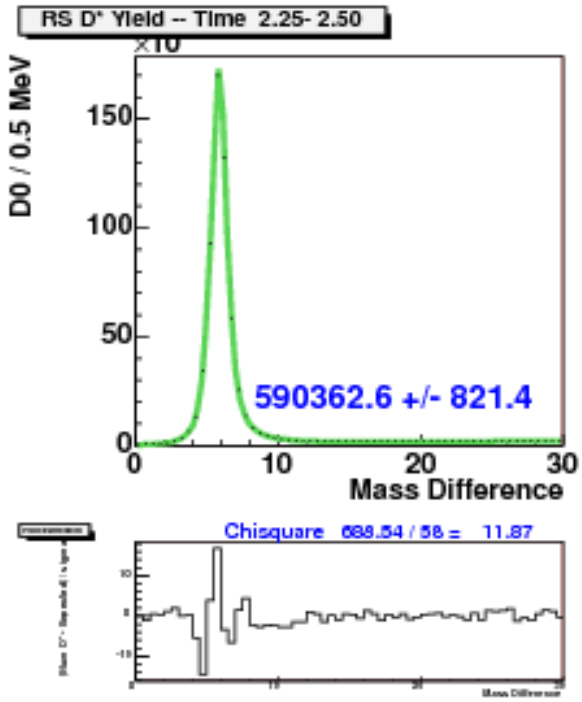


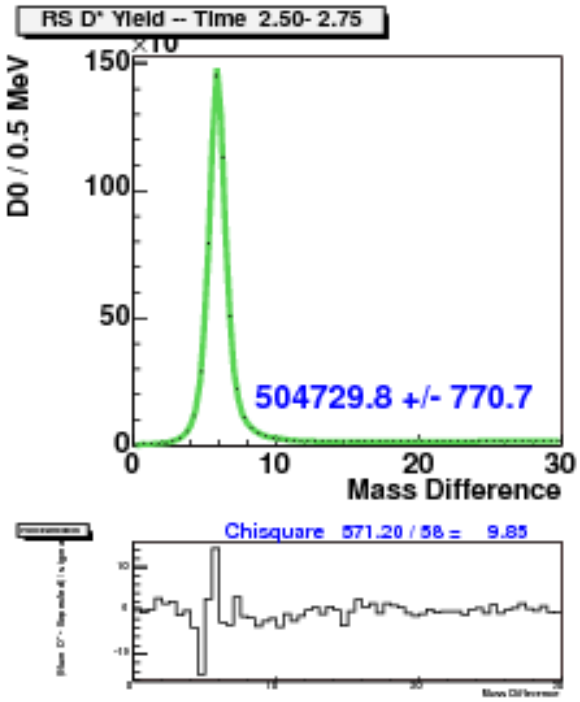



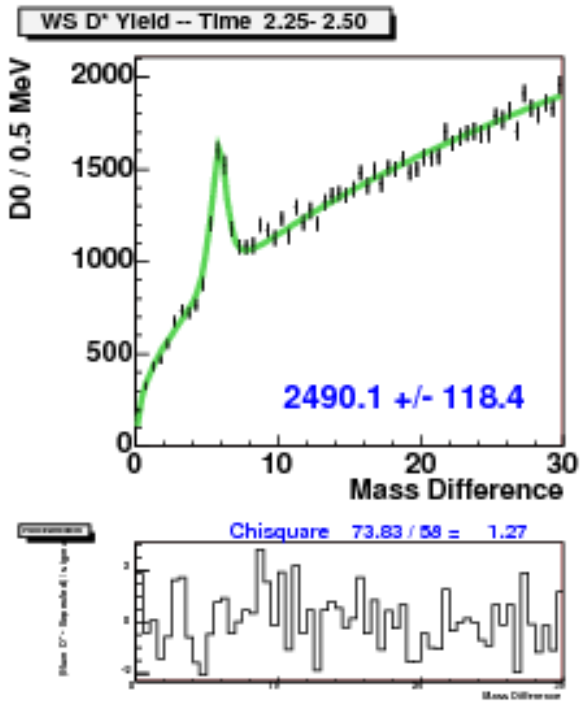



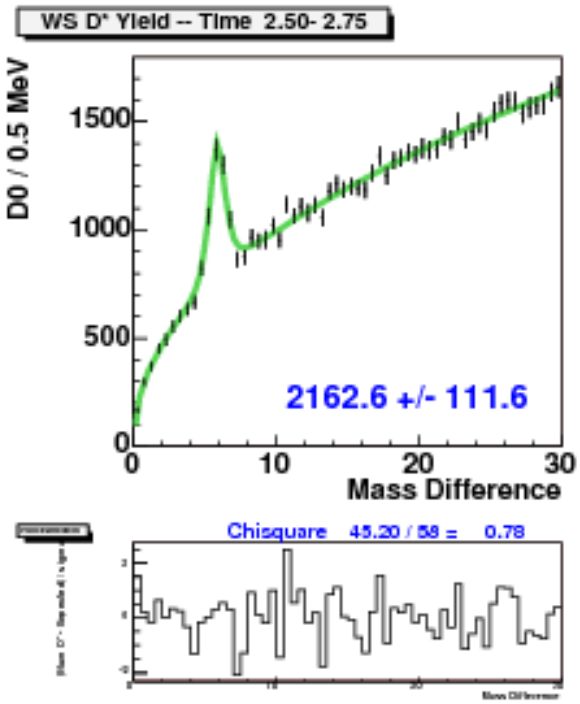



Appendix B:Title

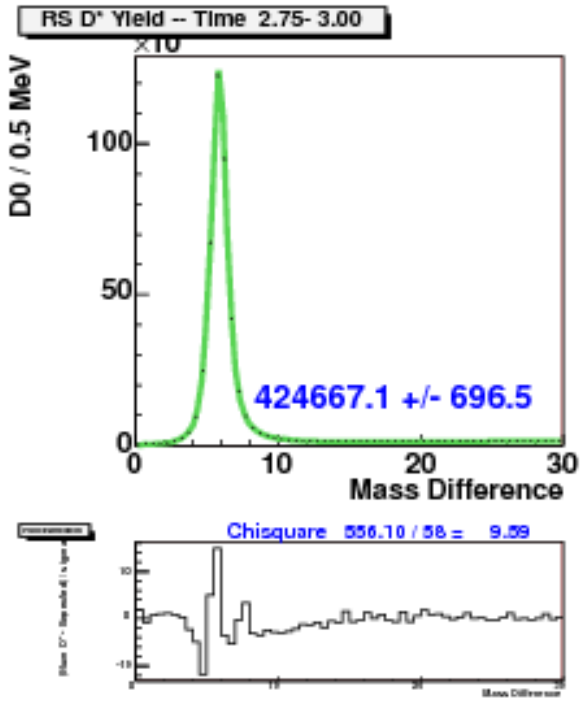


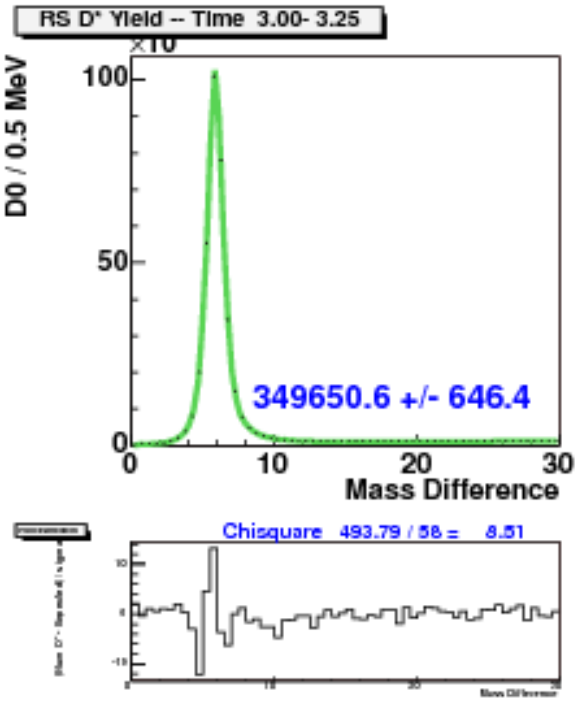



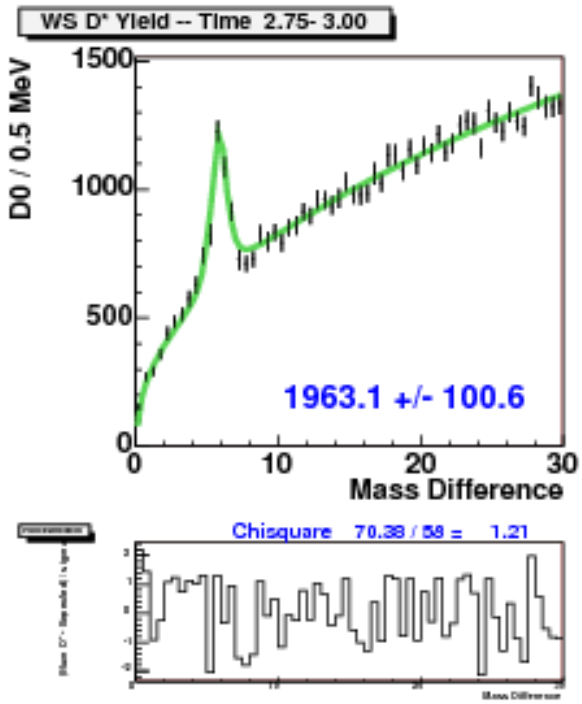



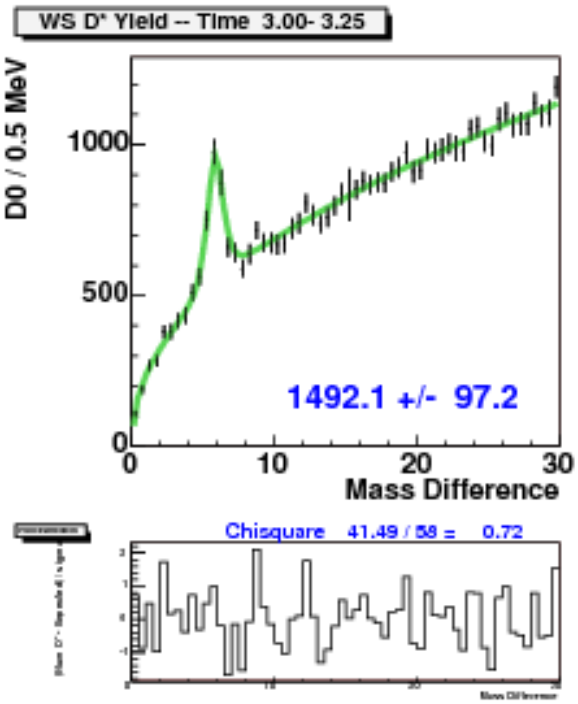



Appendix B:Title

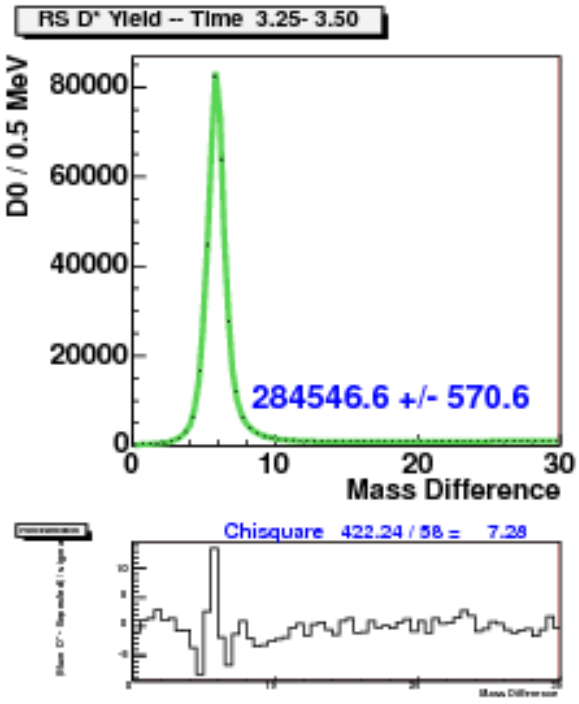


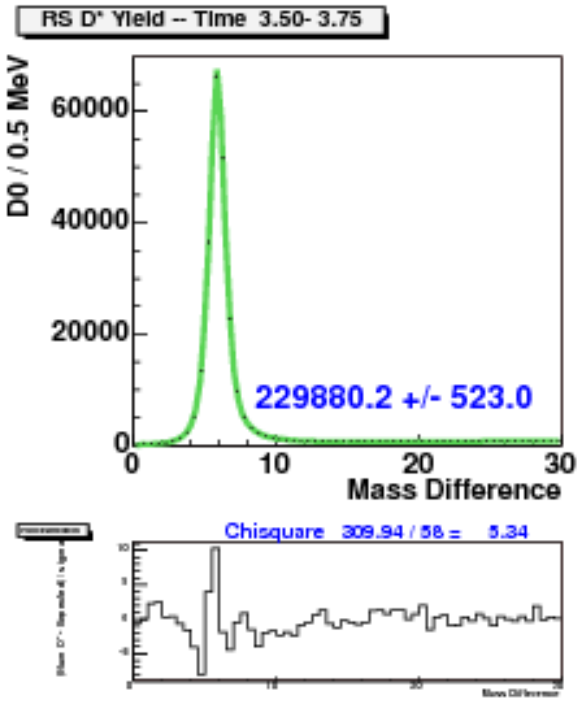



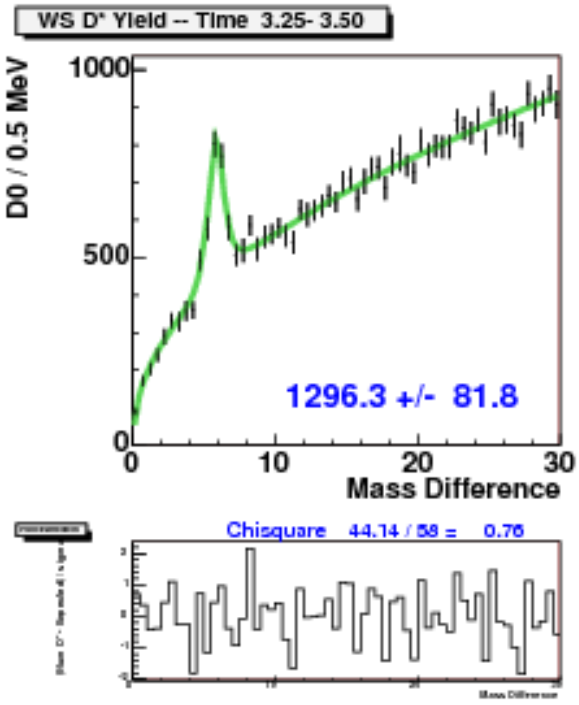



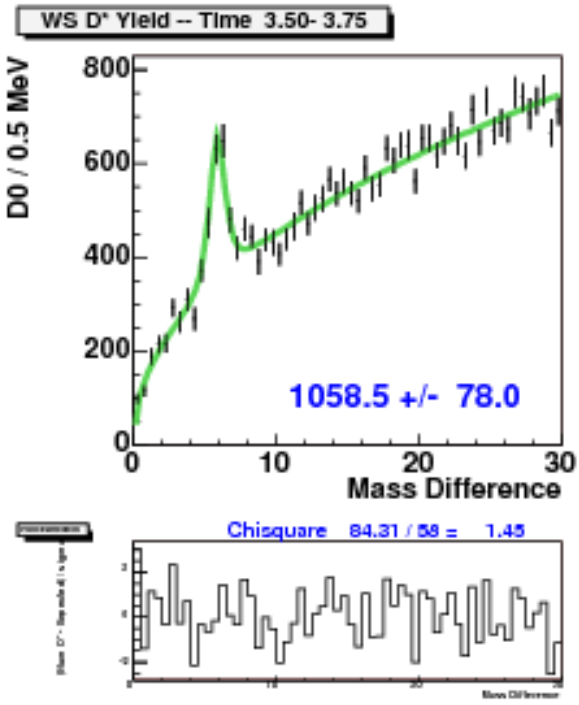



Appendix B:Title

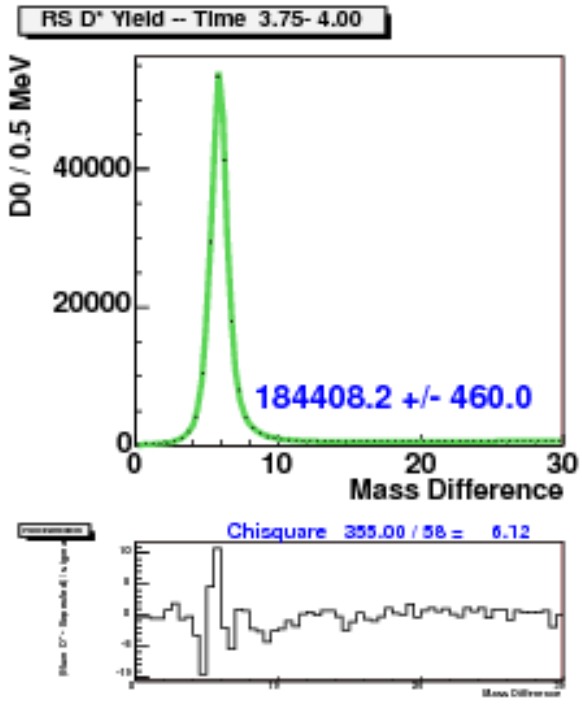


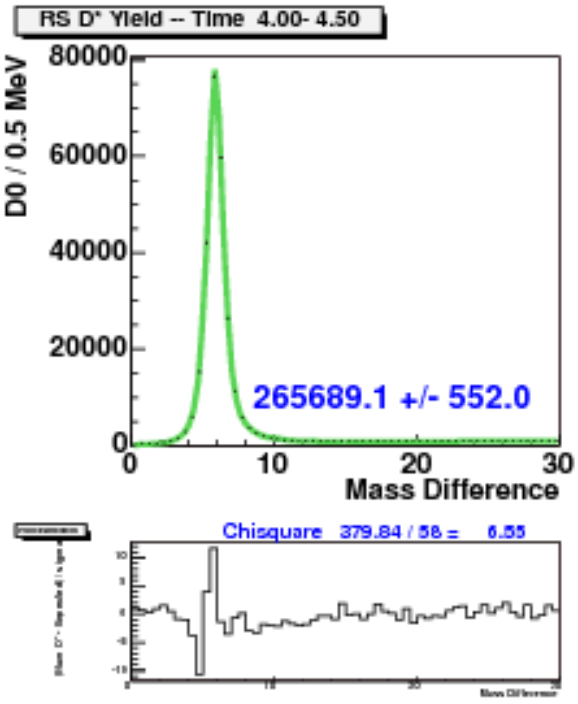



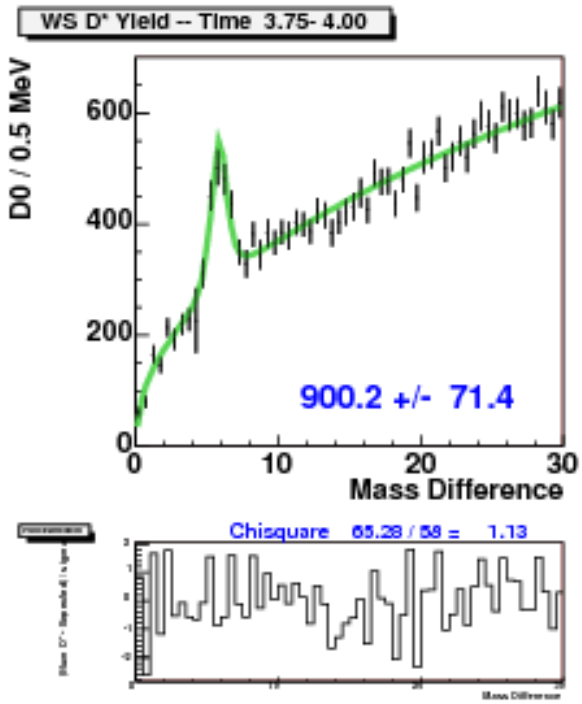



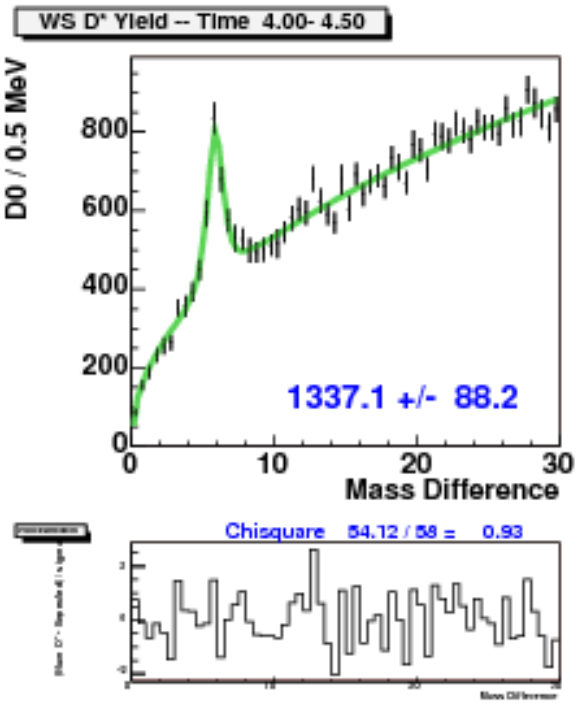



Appendix B:Title

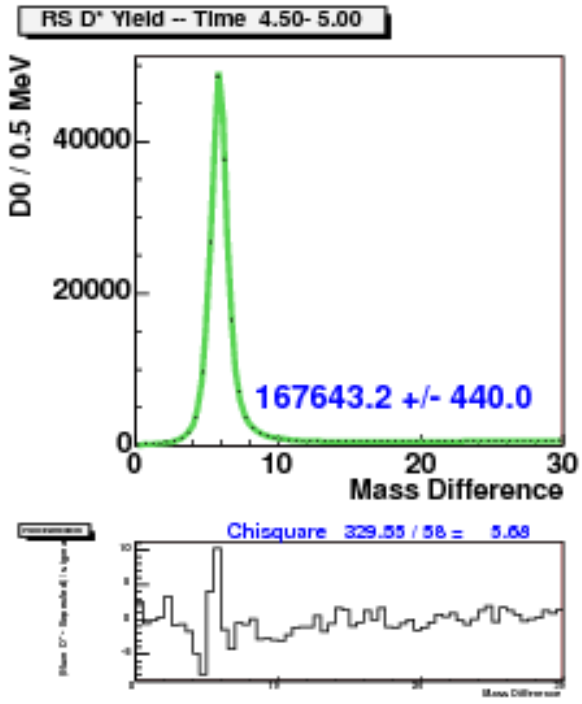


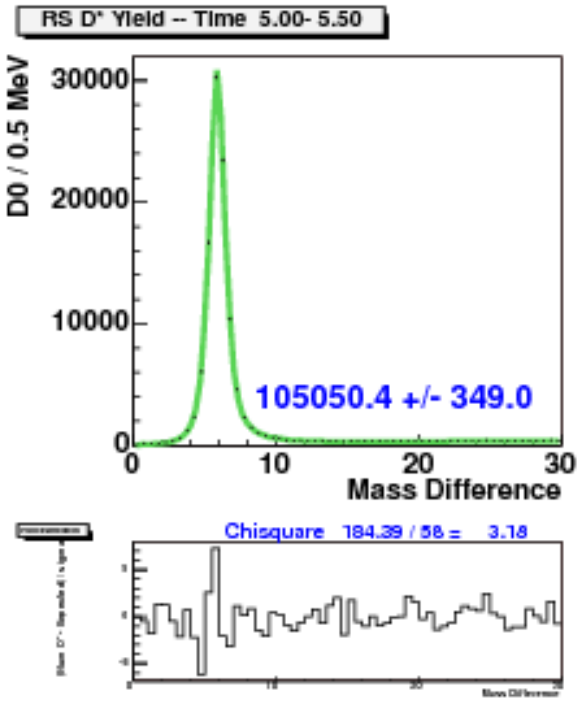



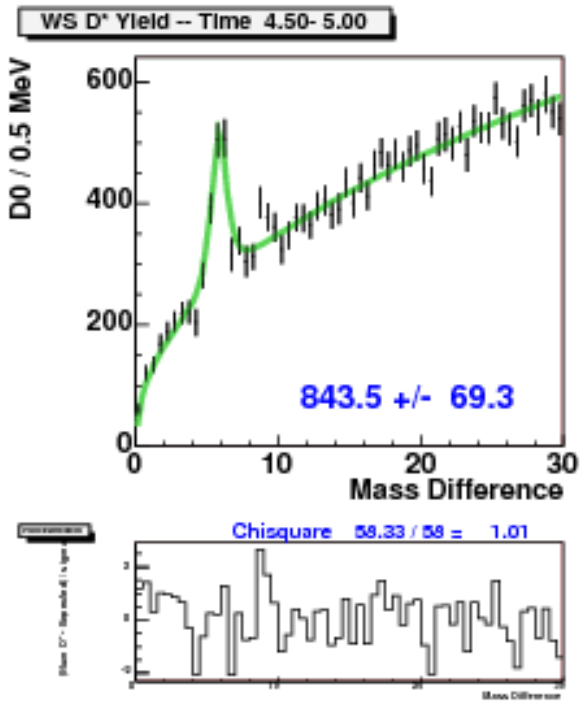



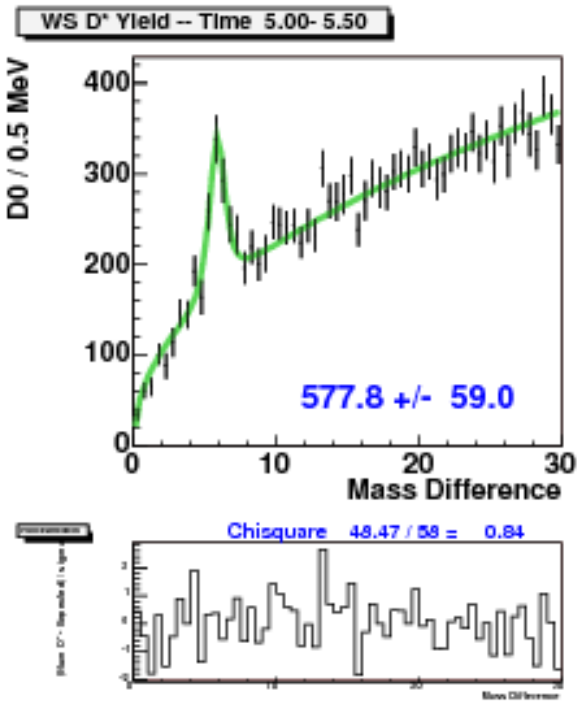



Appendix B:Title

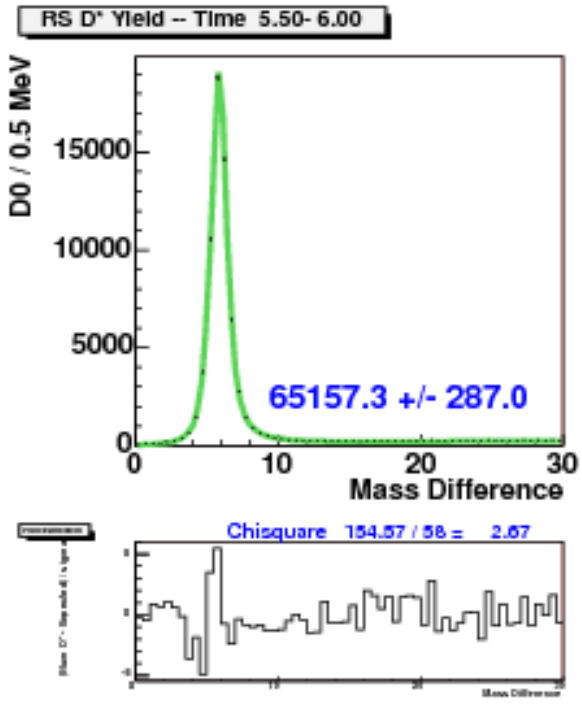


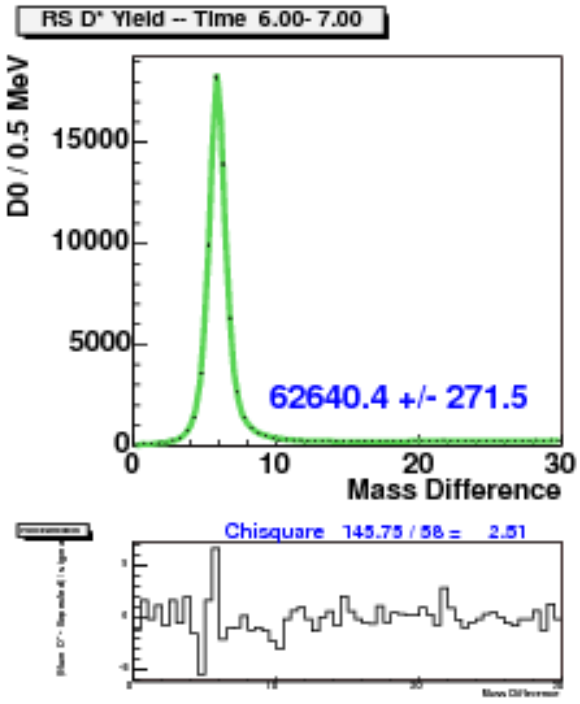



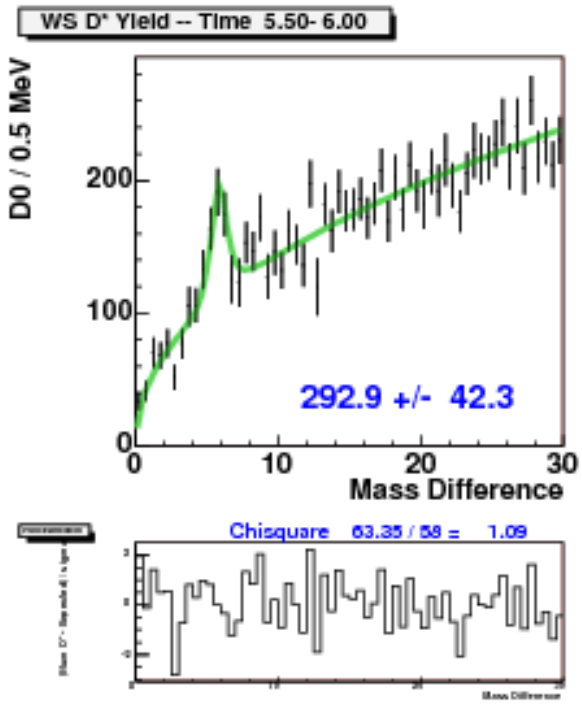



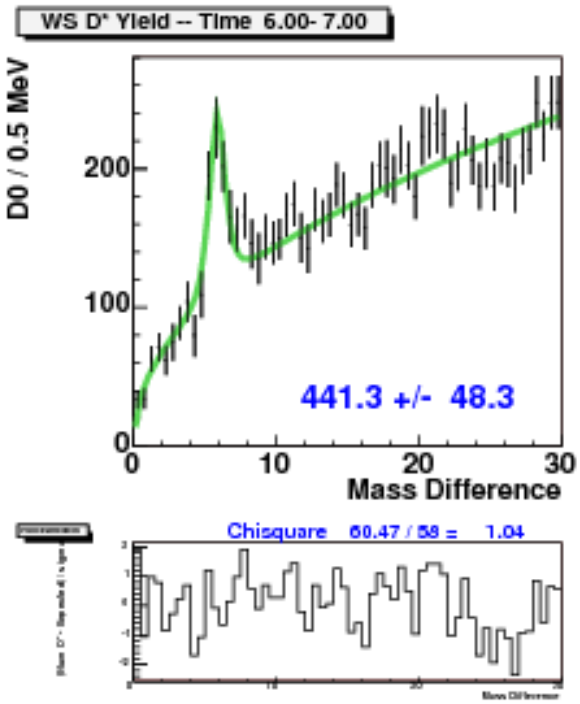



Appendix B:Title

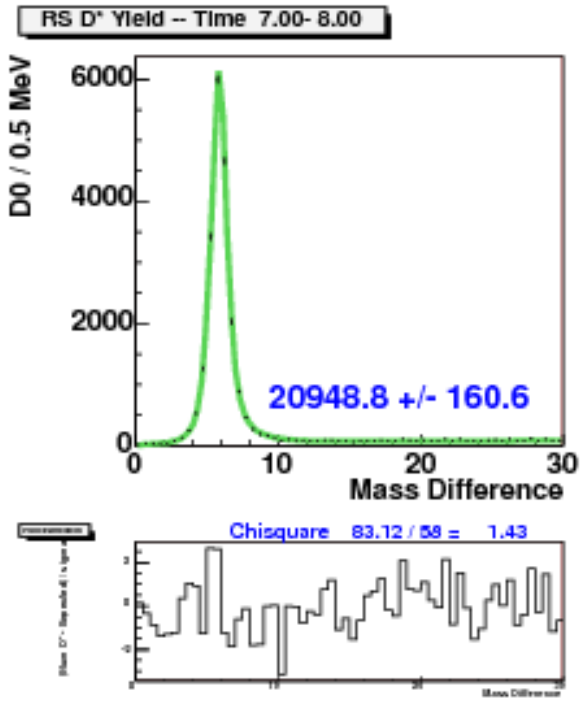


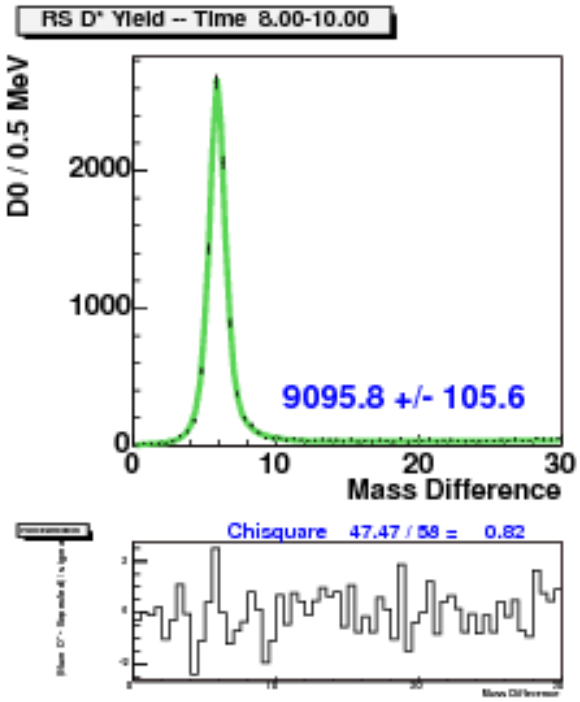



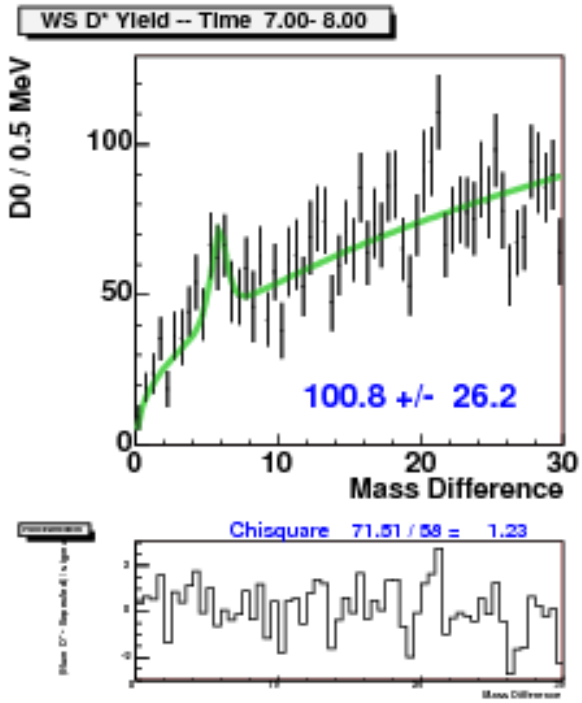



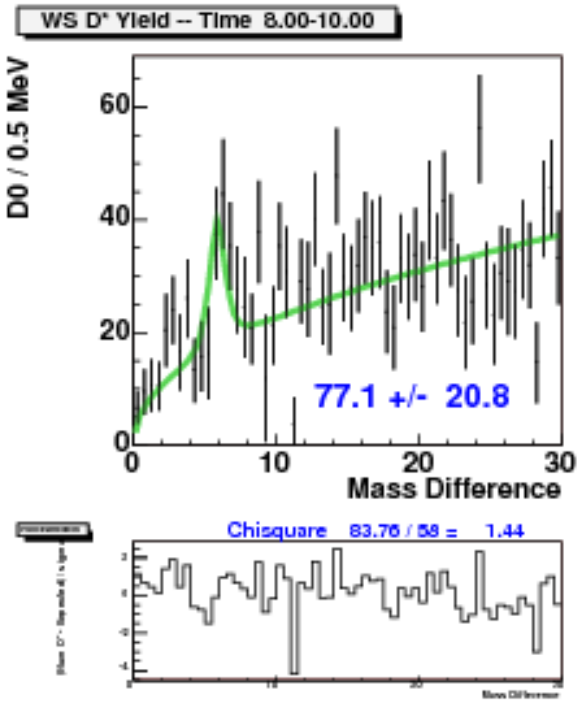



Appendix B:Title



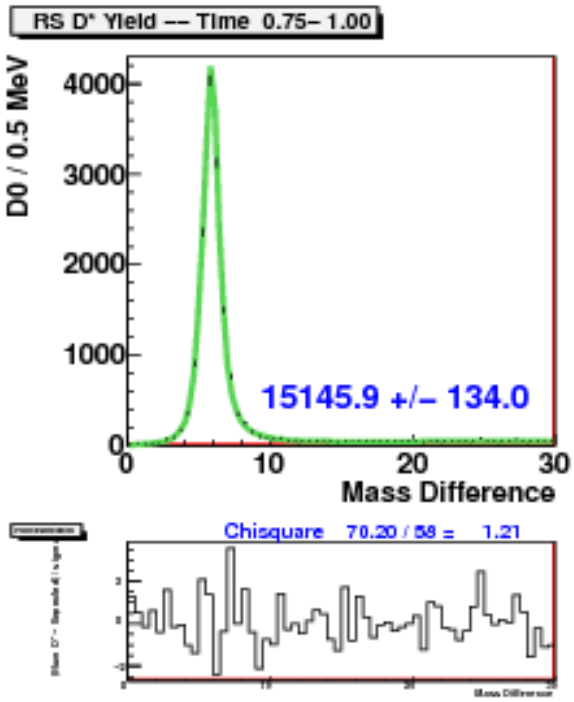



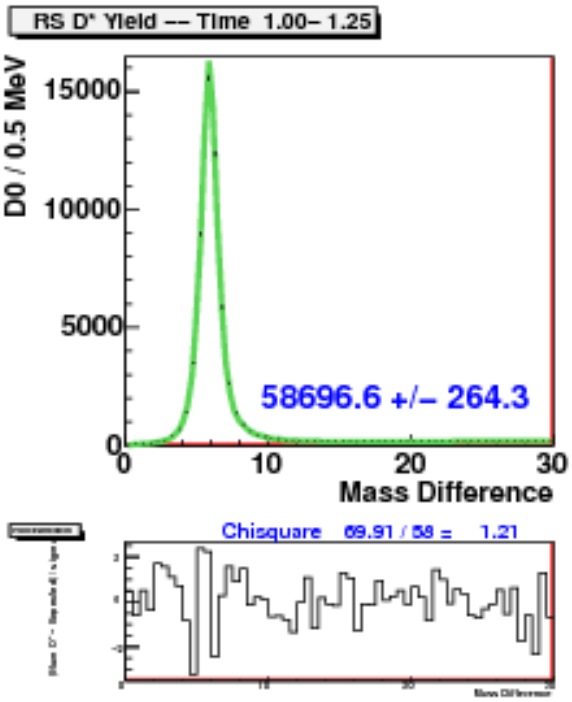



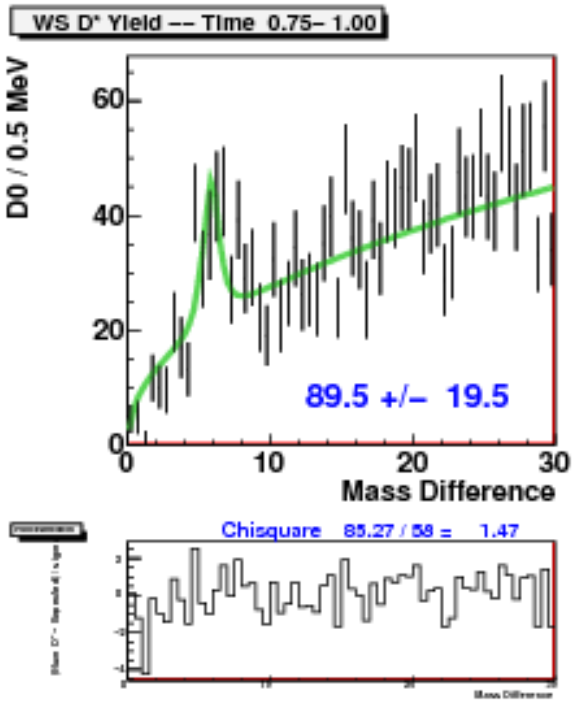



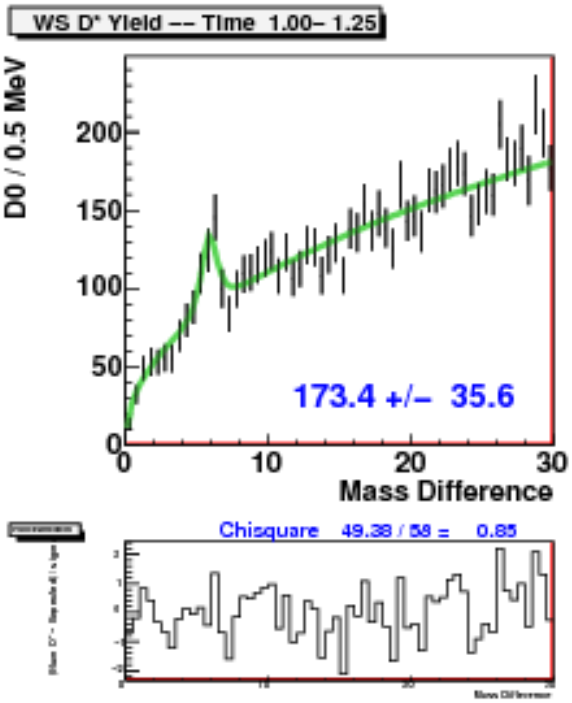
# C Time slices2: period 17 to 21 data

Appendix C:Title

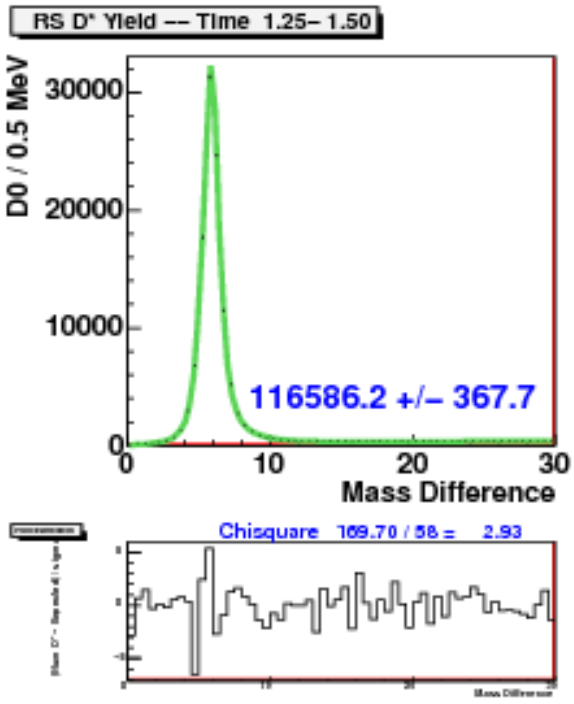


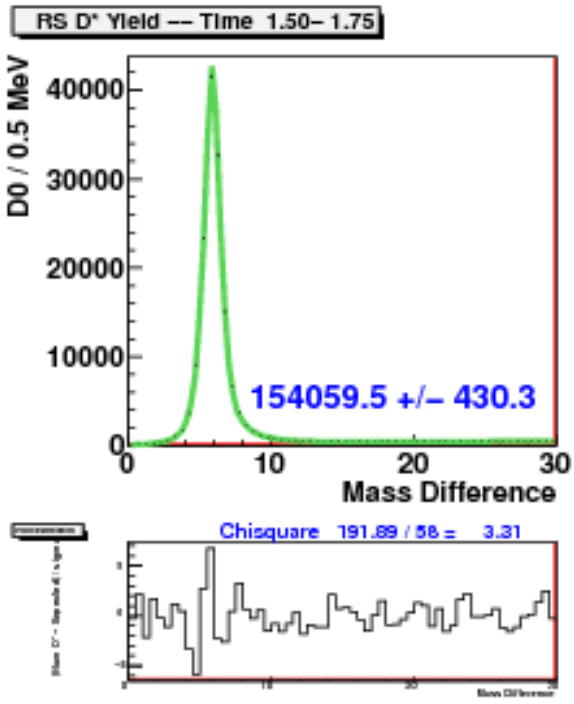



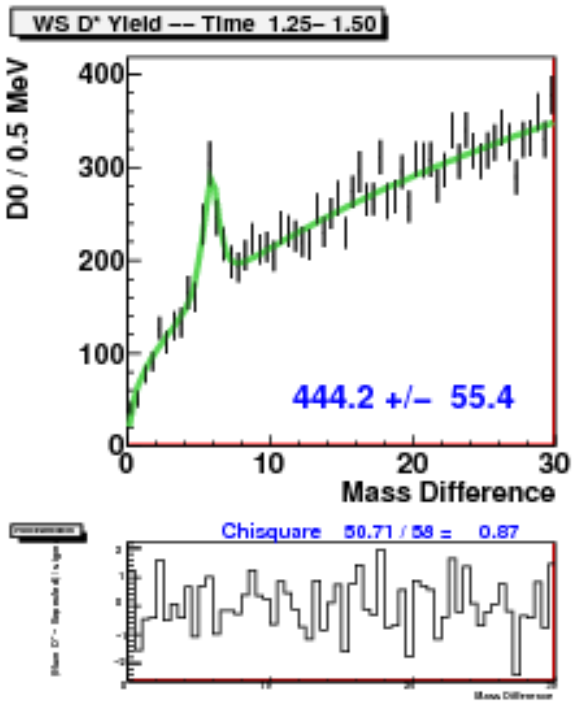



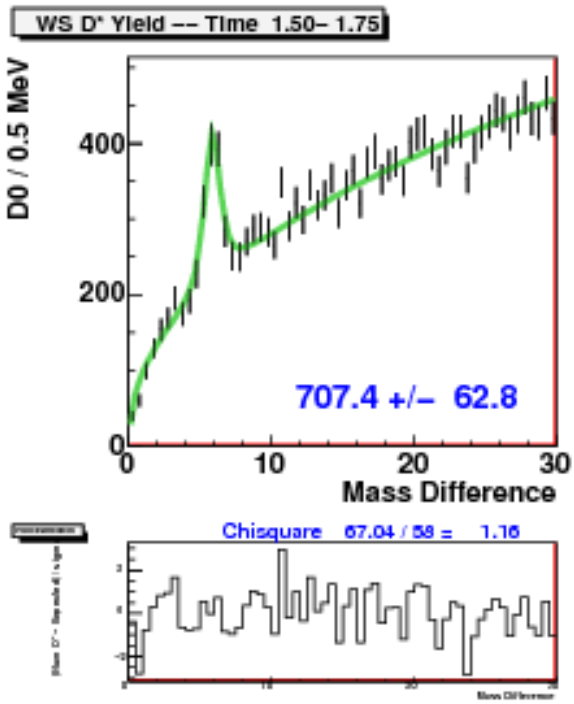



Appendix C:Title

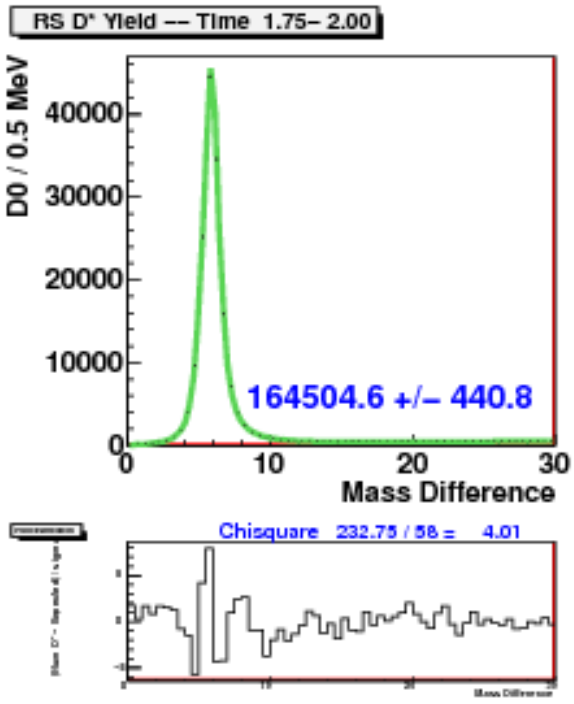


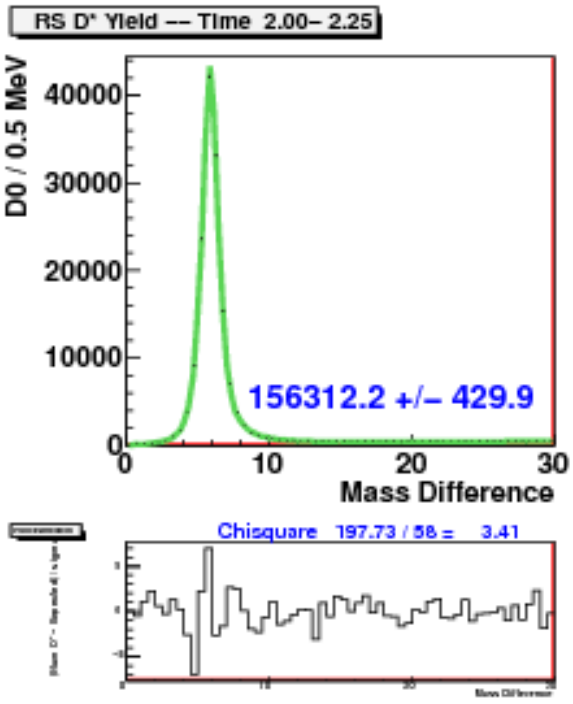



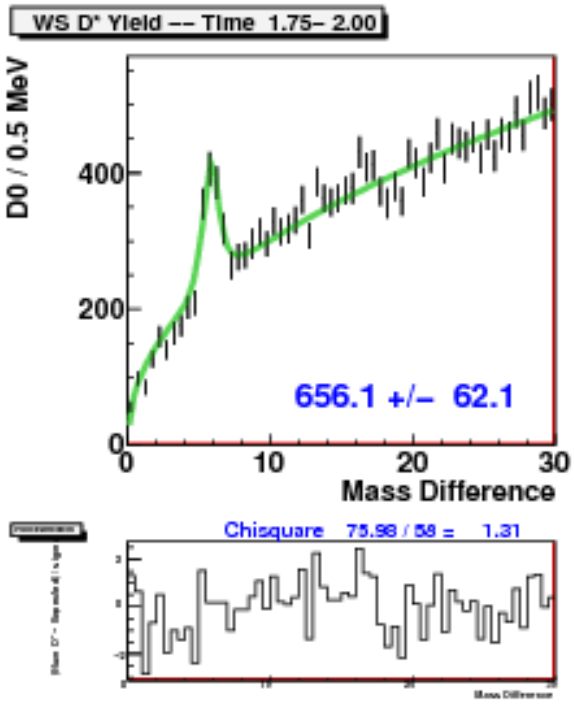



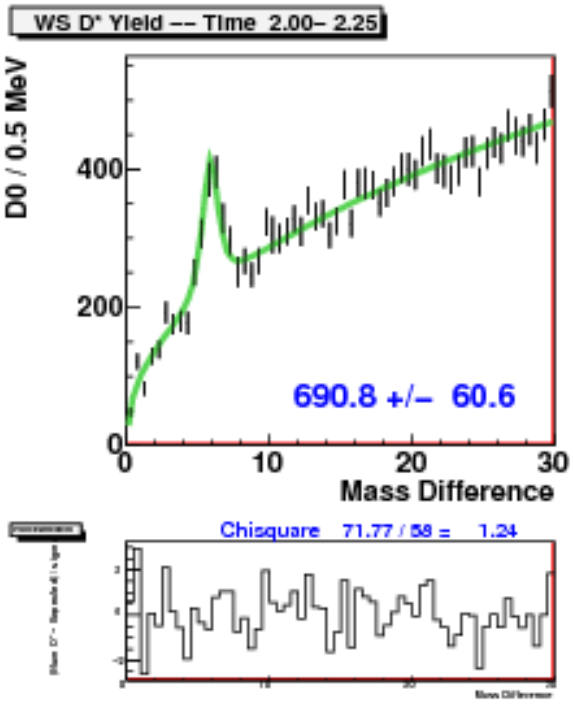



Appendix C:Title

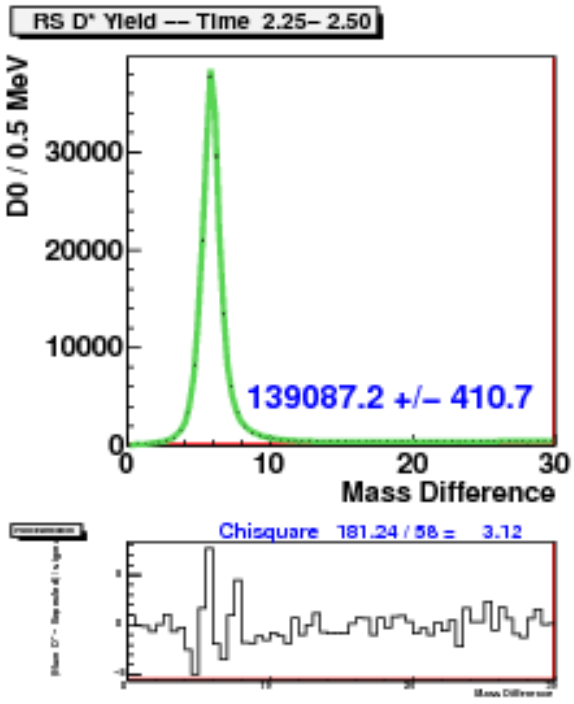


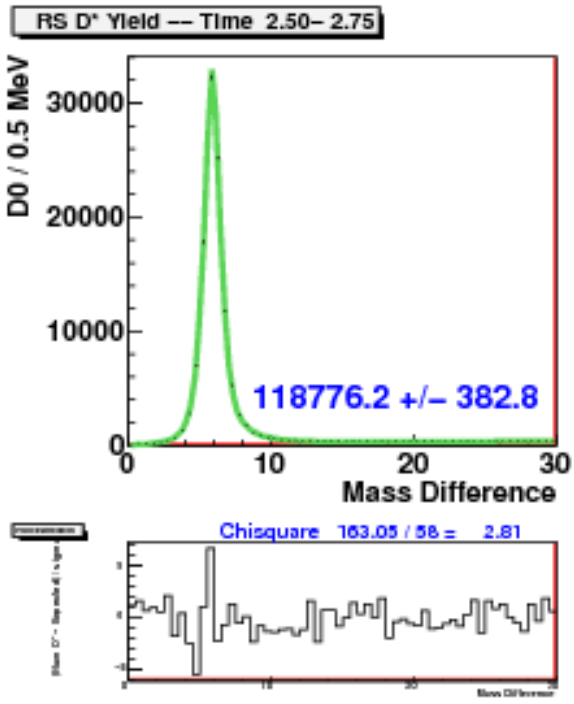



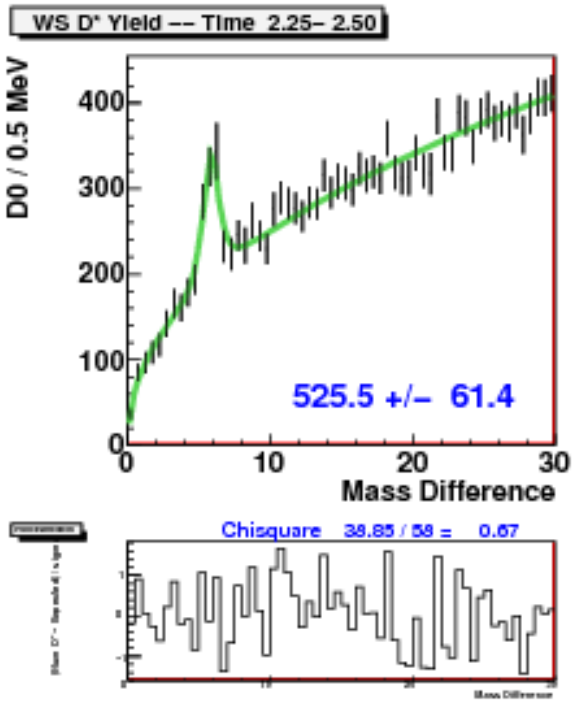



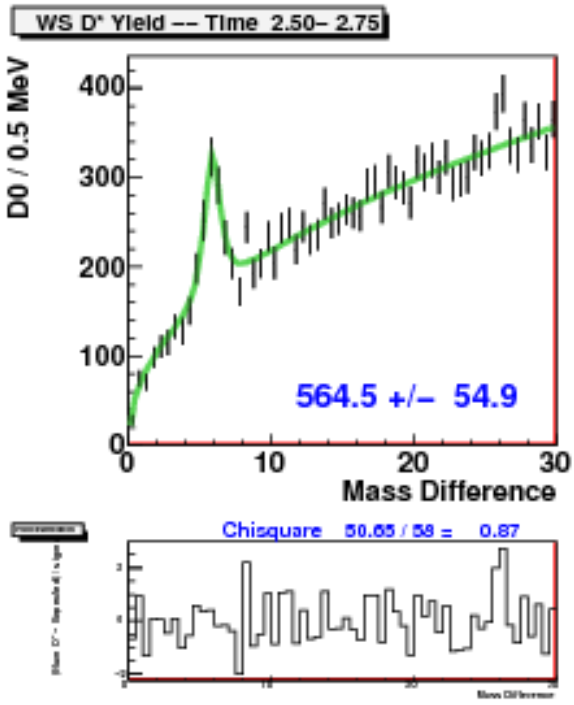



Appendix C:Title

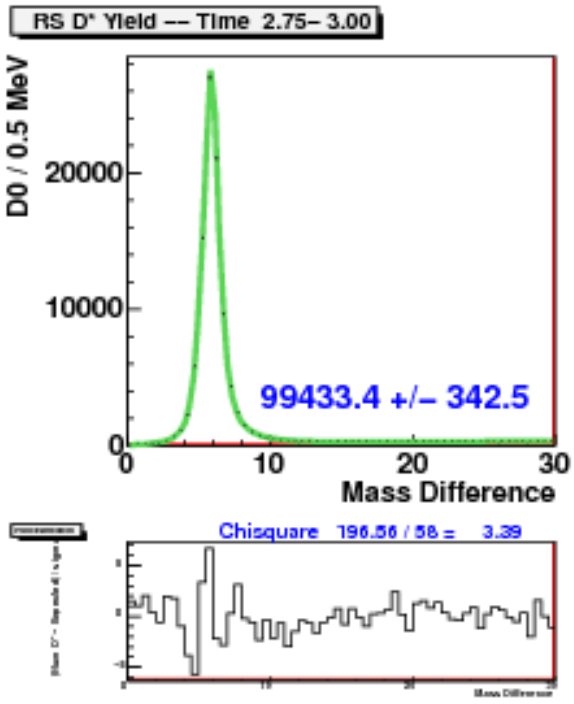


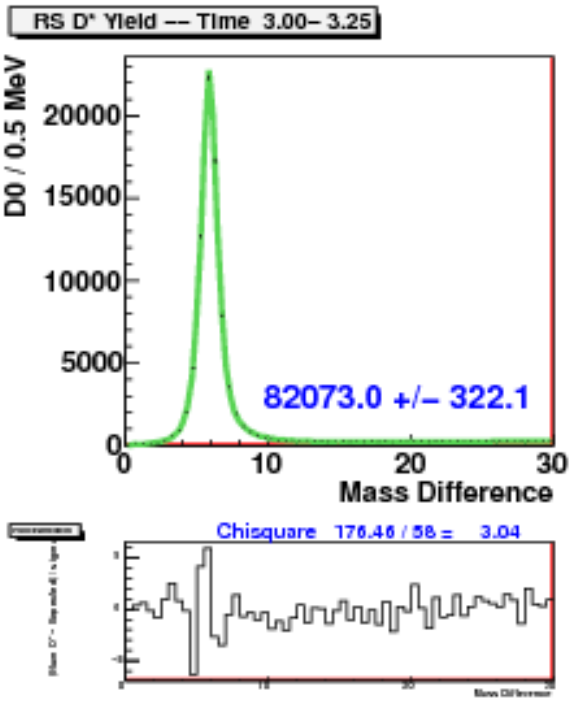



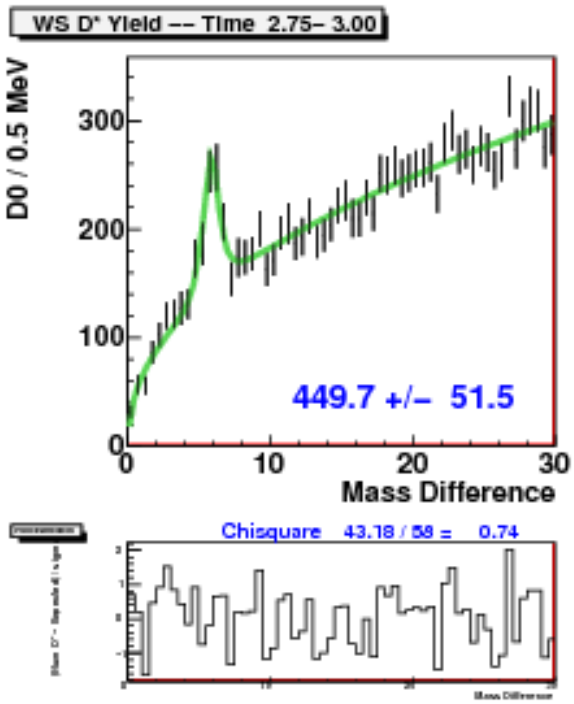



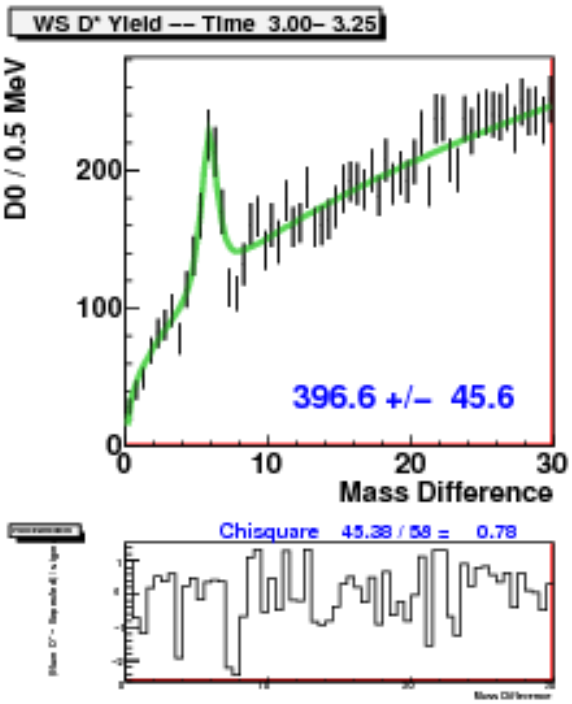



Appendix C:Title

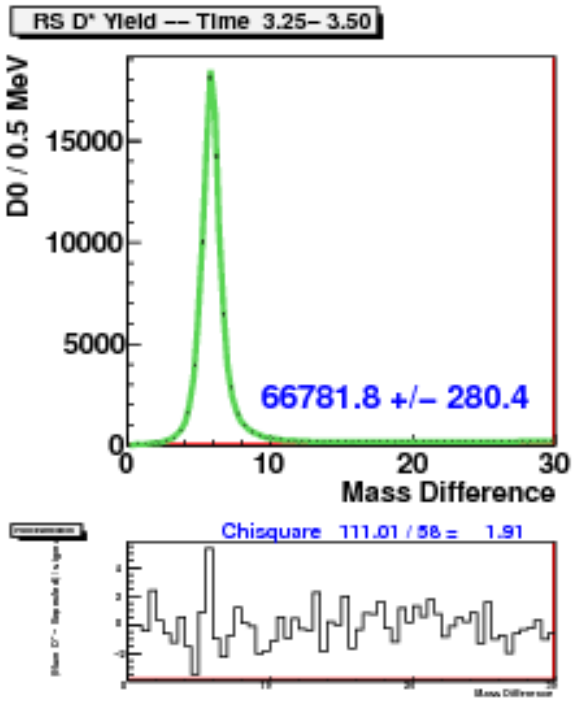


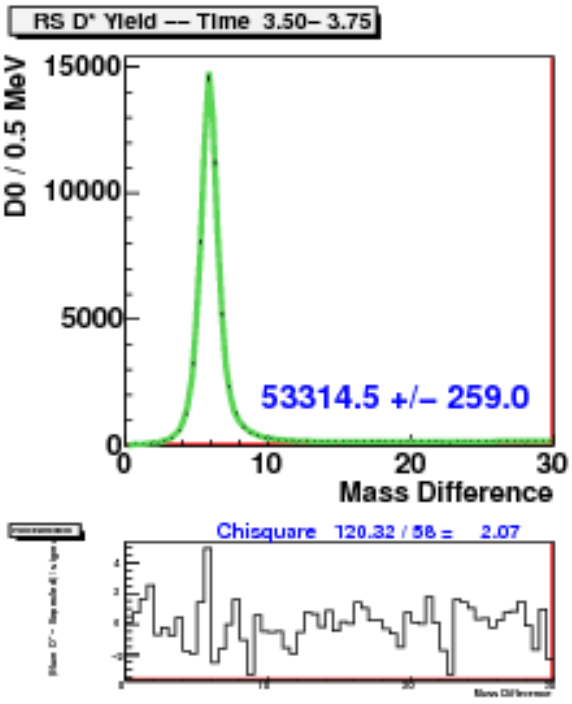



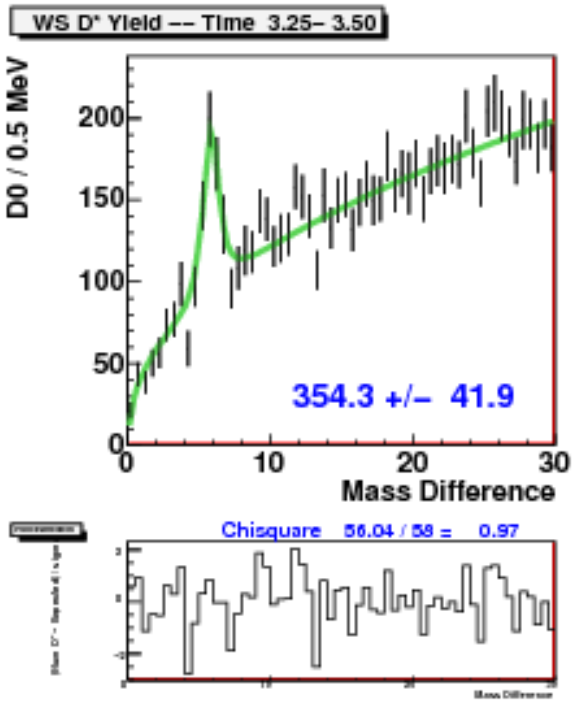



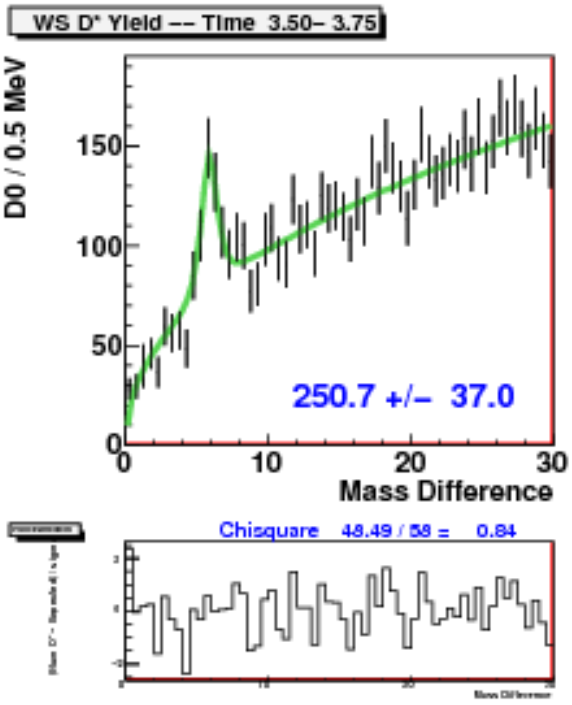



Appendix C:Title

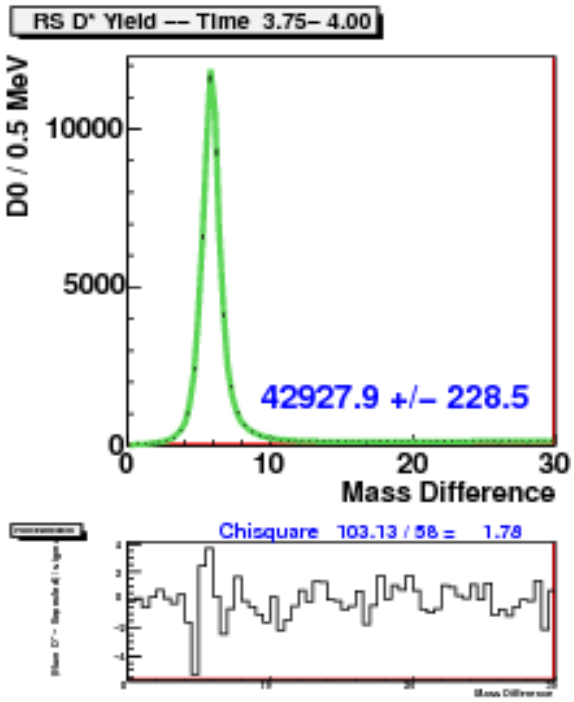


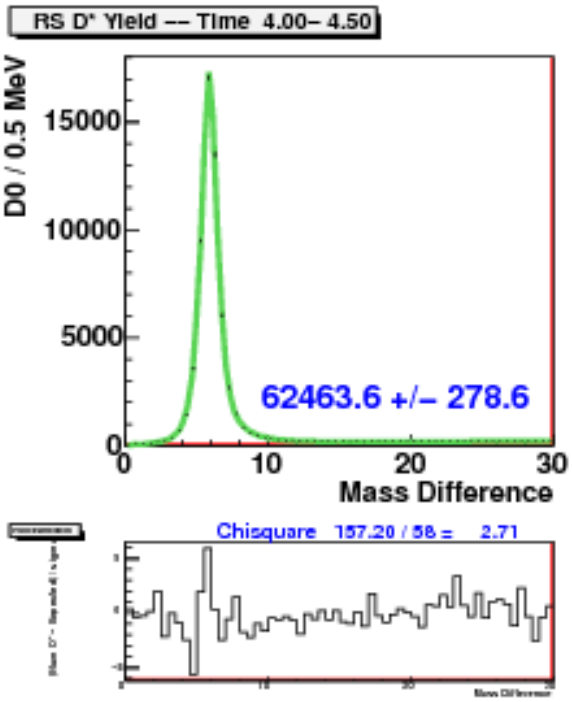



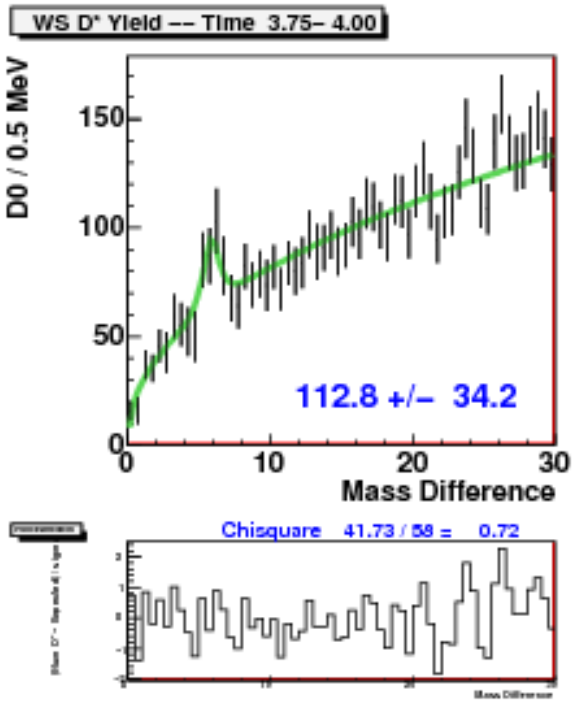



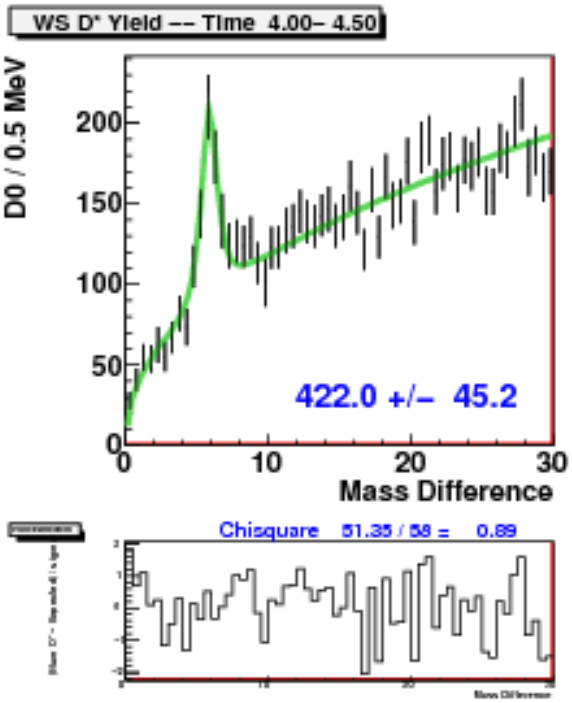



Appendix C:Title

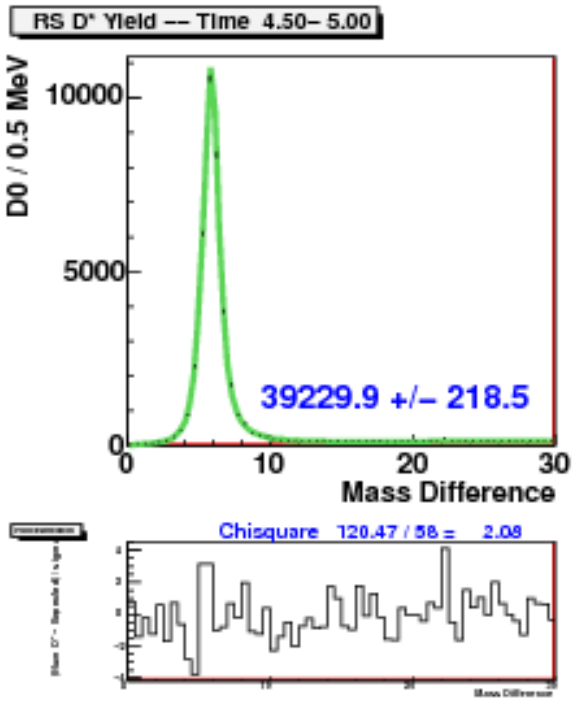


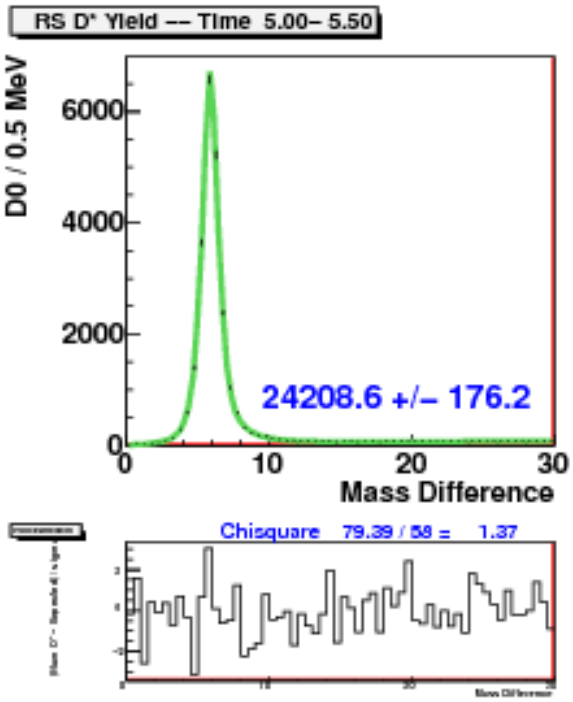



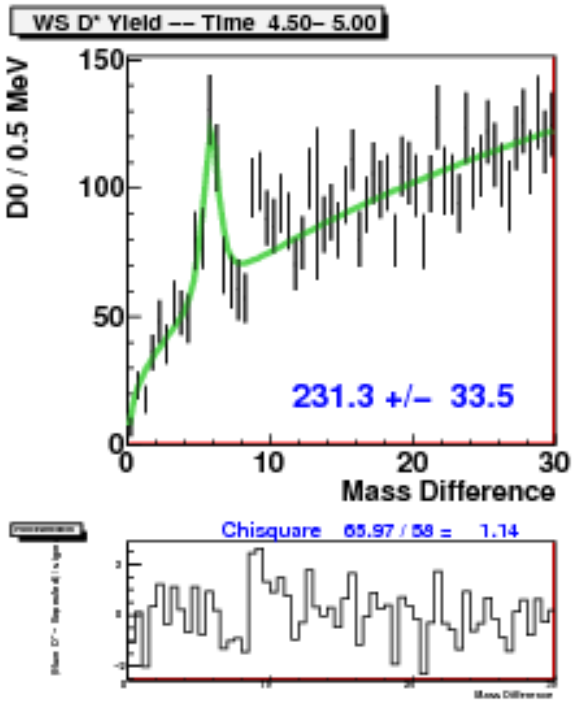



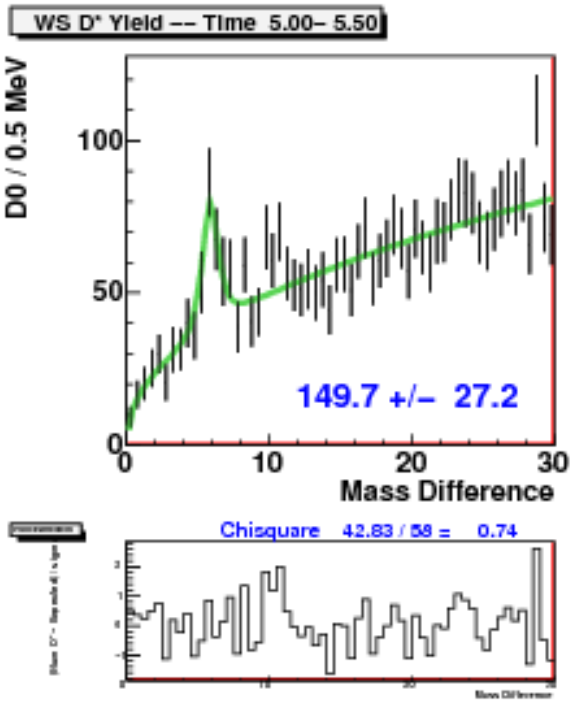



Appendix C:Title

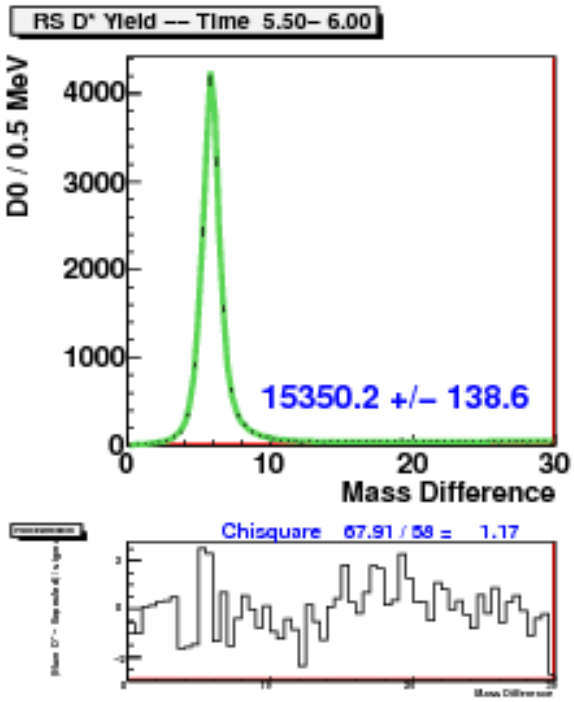


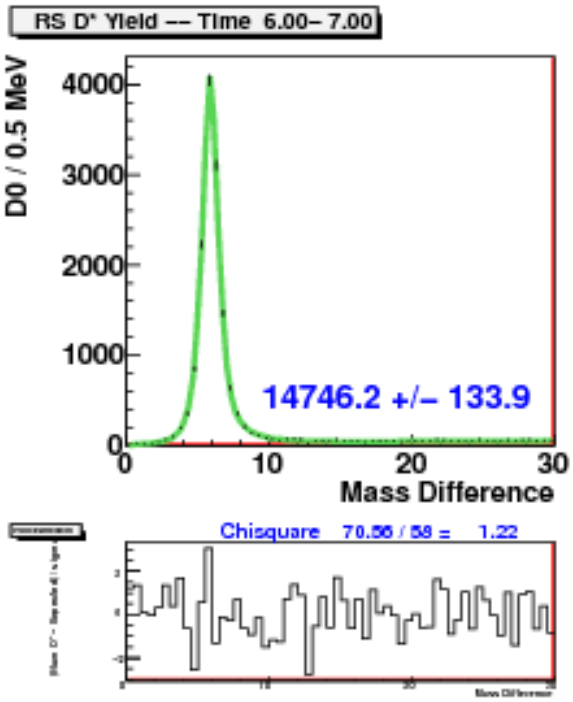



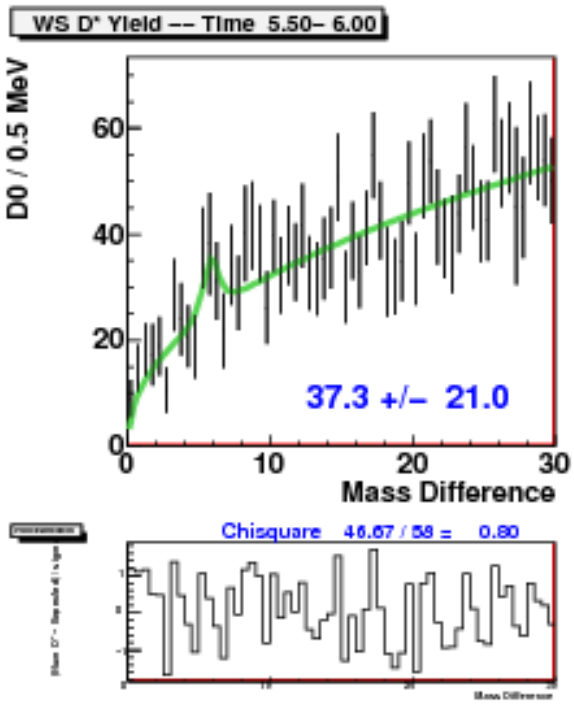



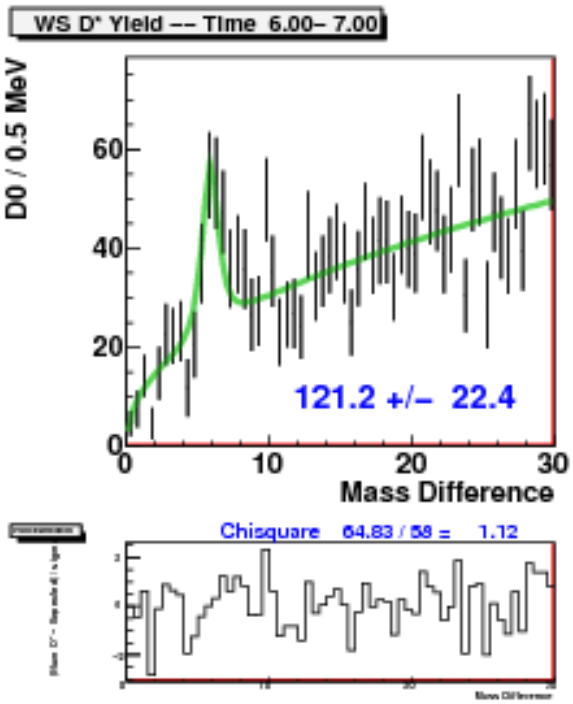



Appendix C:Title

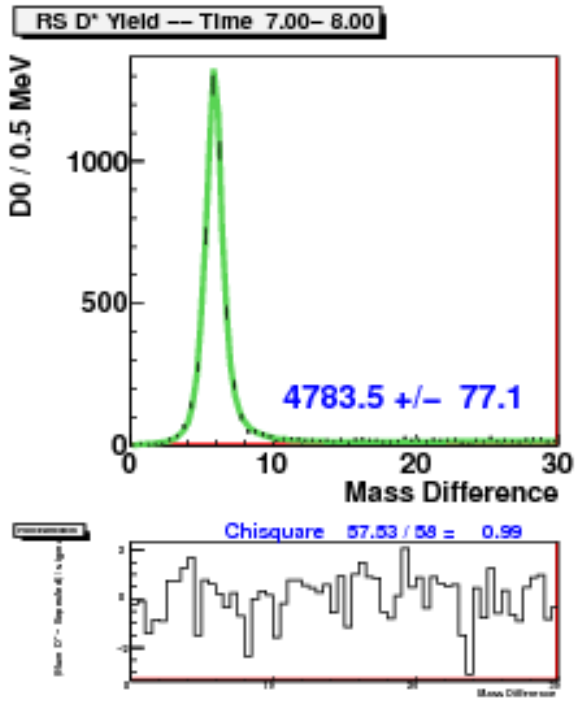


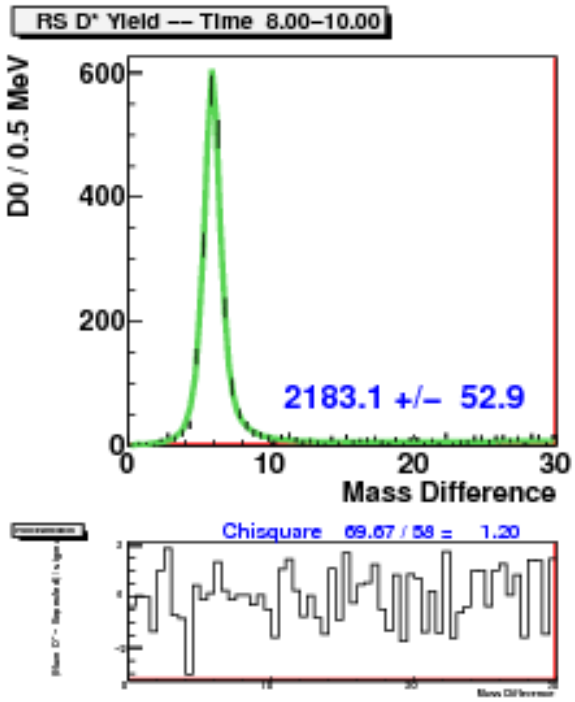



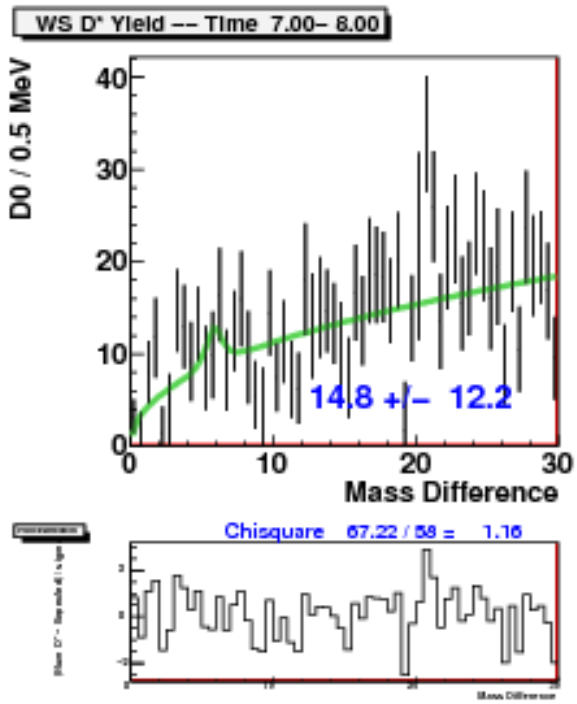



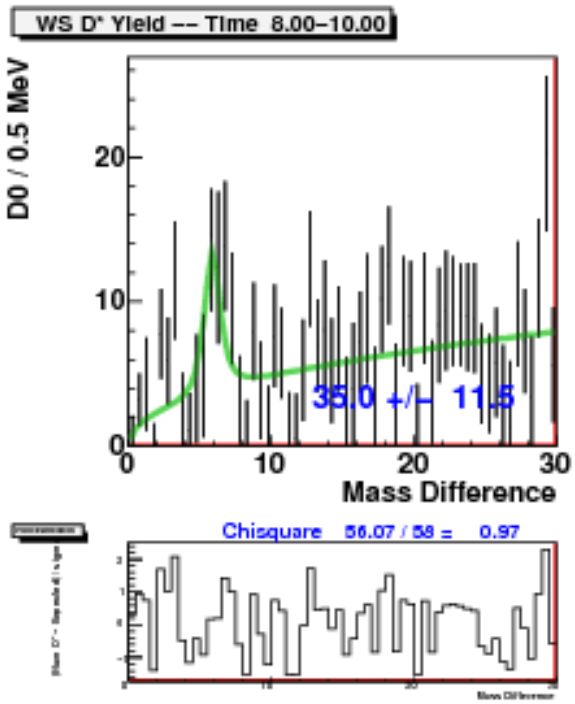



Appendix C:Title



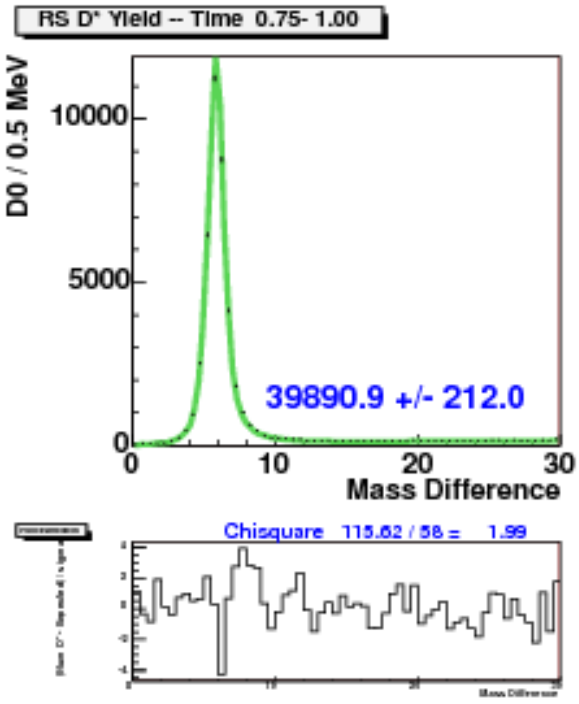



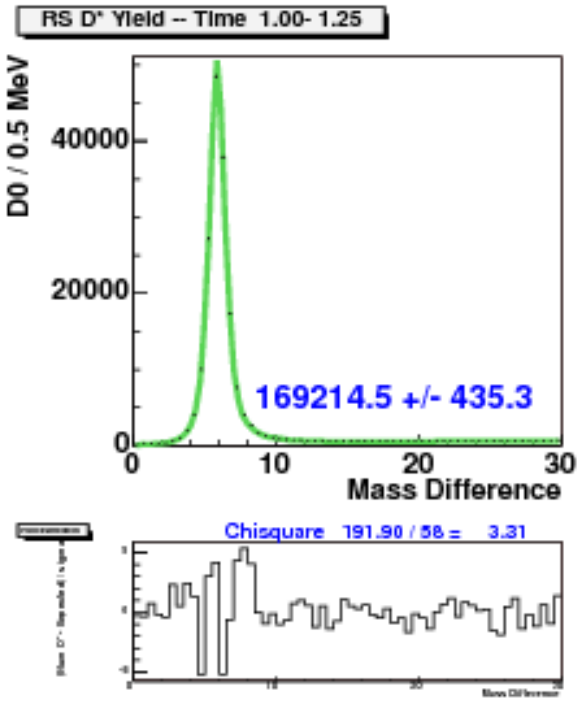



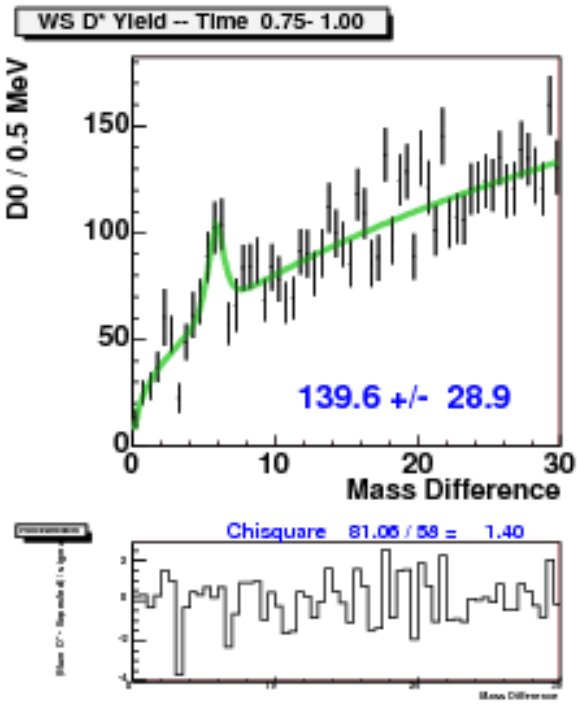



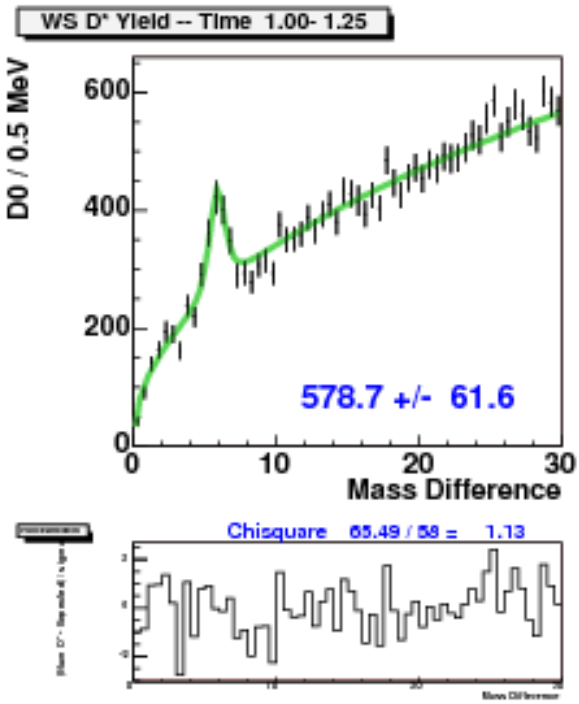
# D Time slices3:period 0 to 16 data

Appendix D:Title

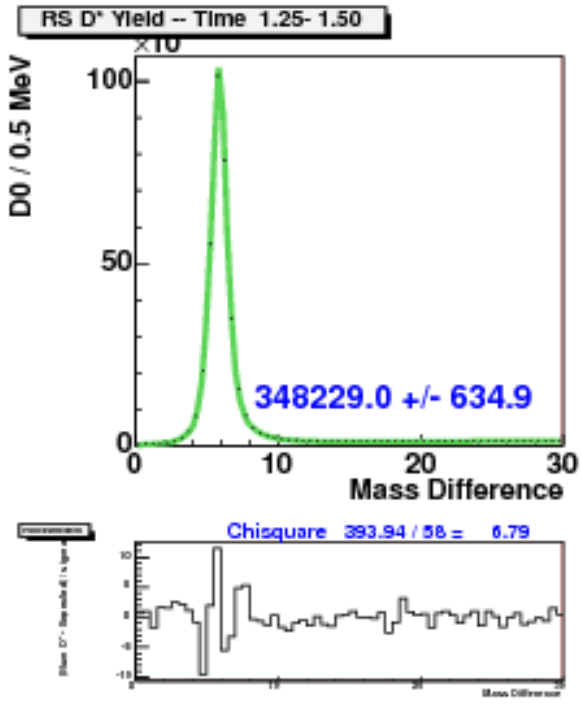


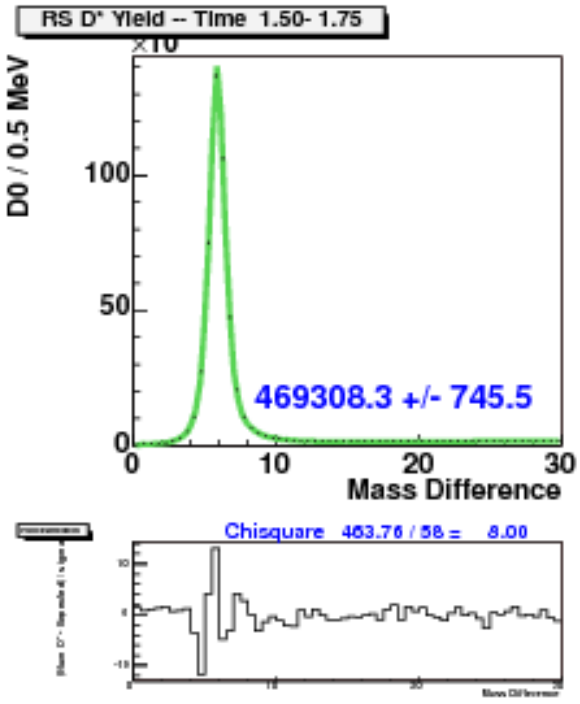



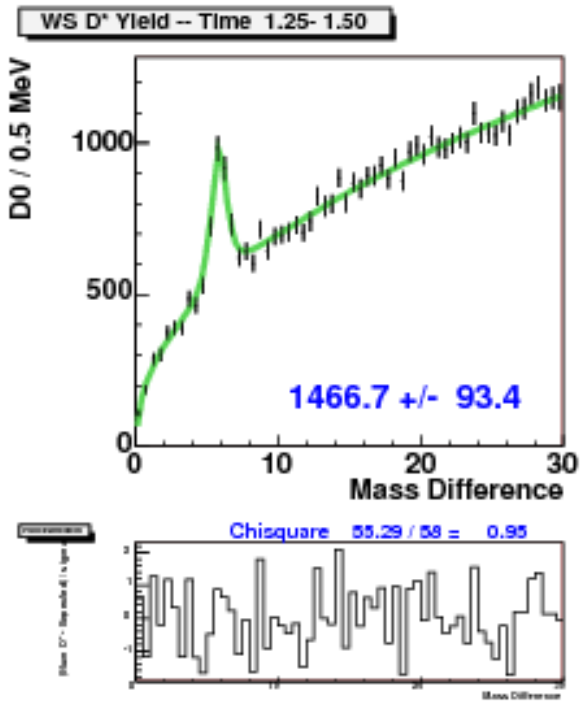



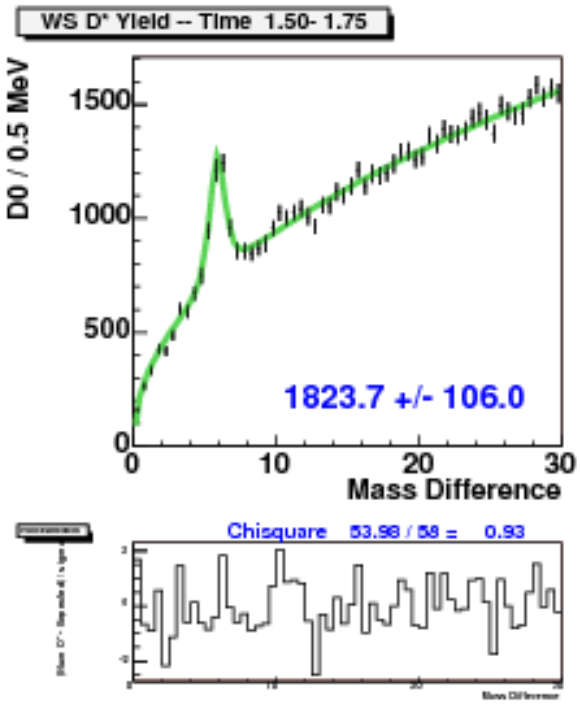



Appendix D:Title

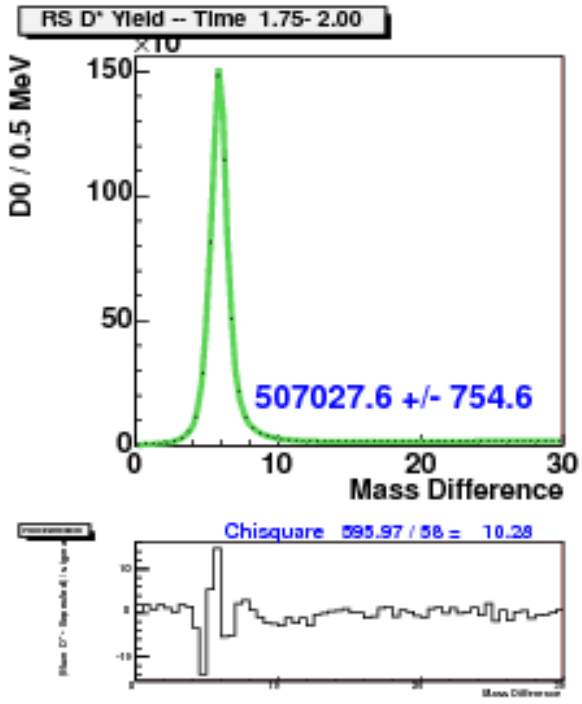


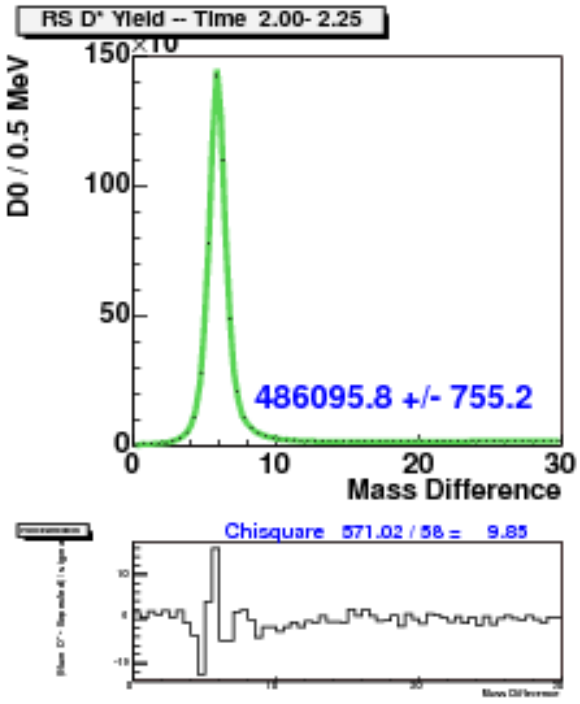



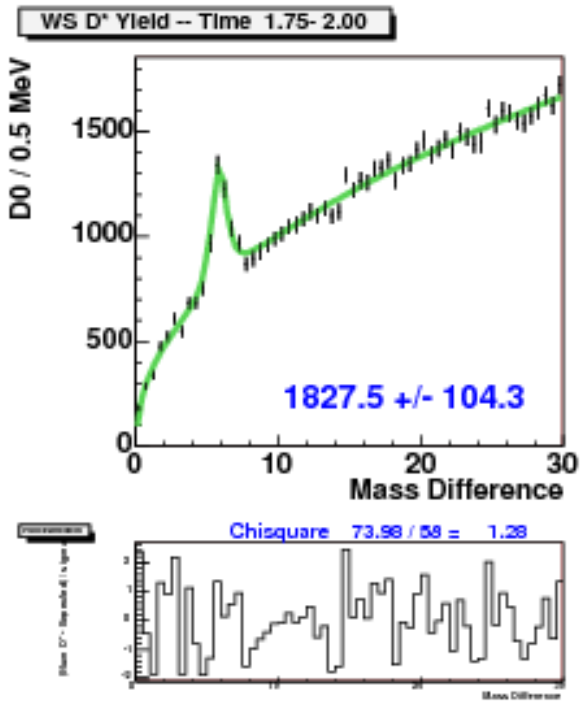



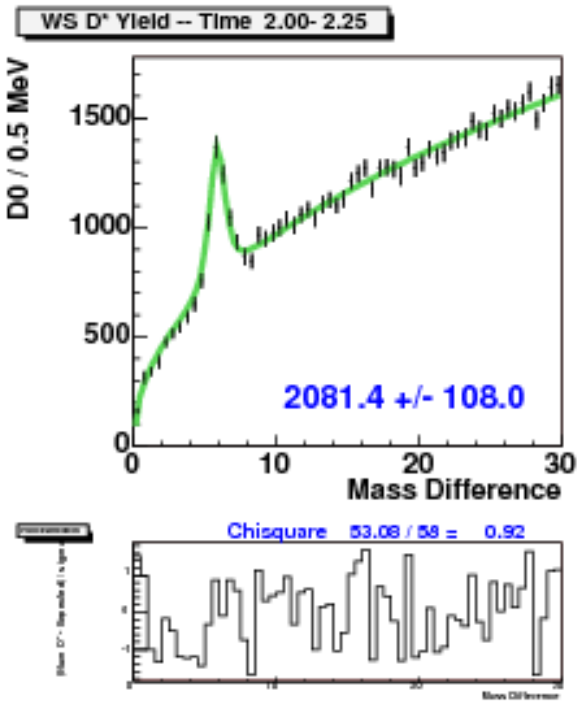



Appendix D:Title

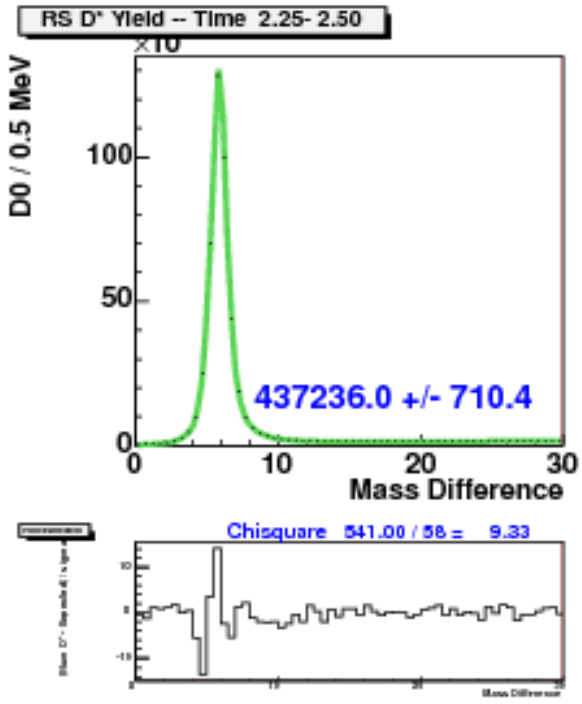


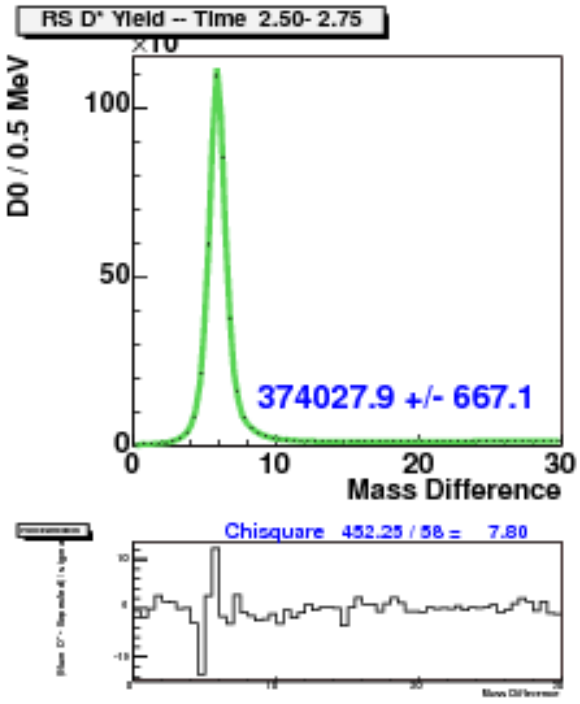



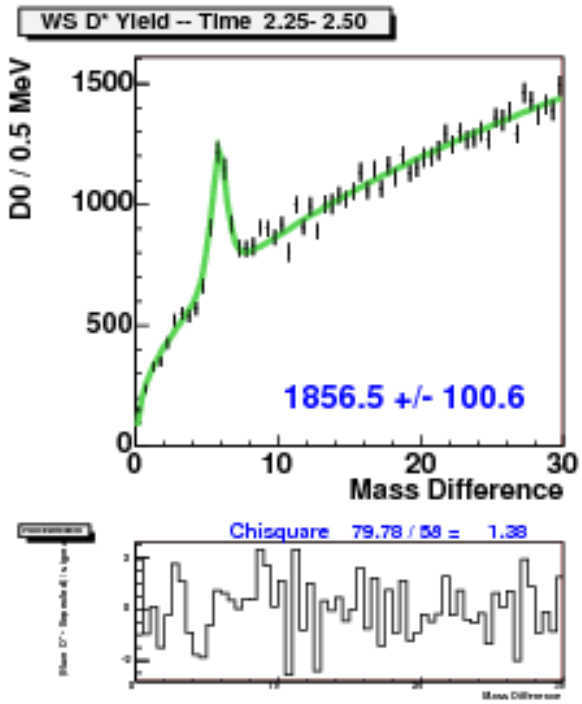



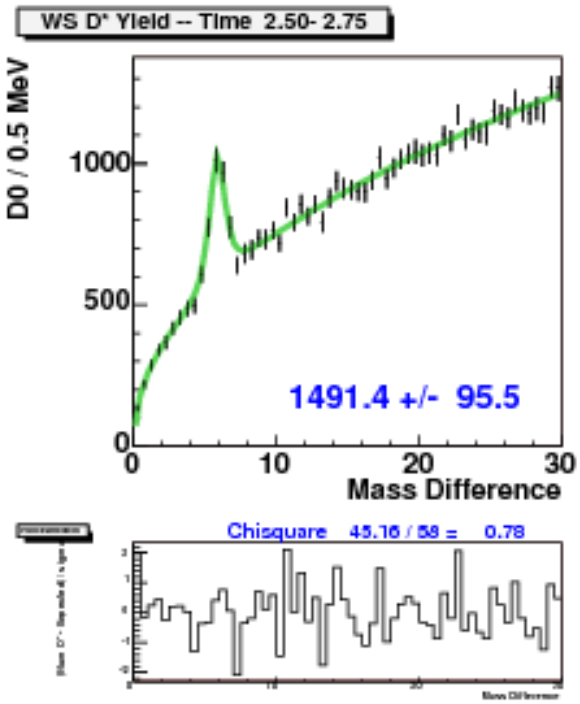



Appendix D:Title

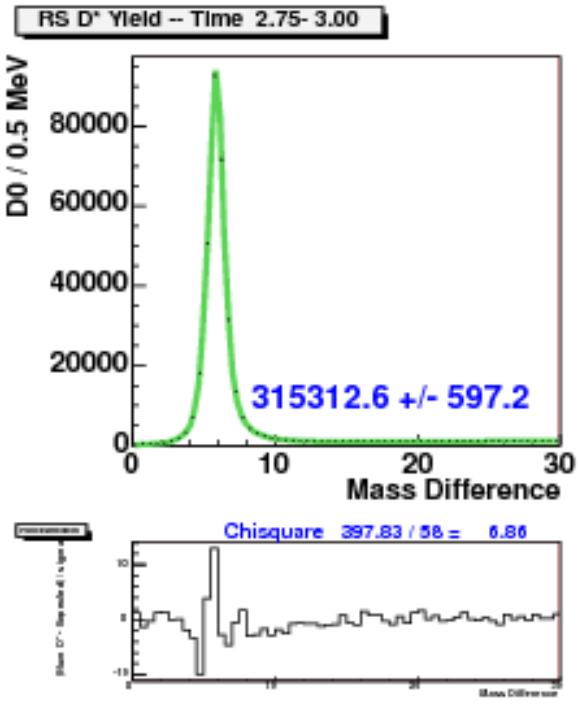


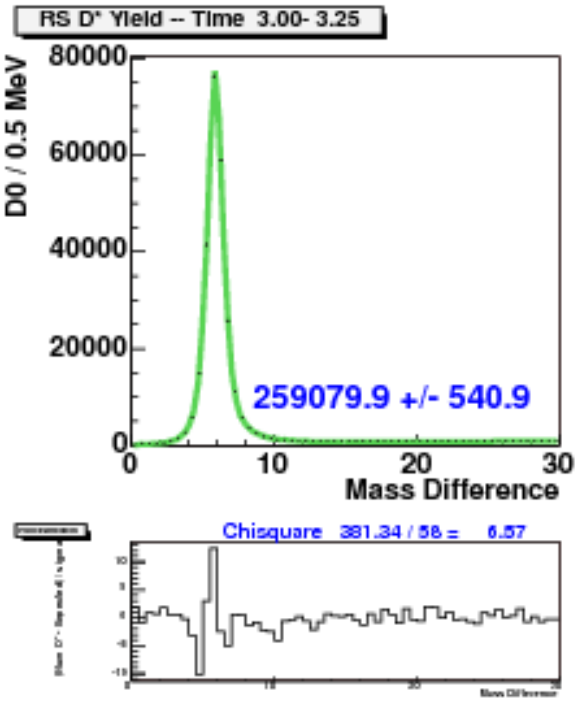



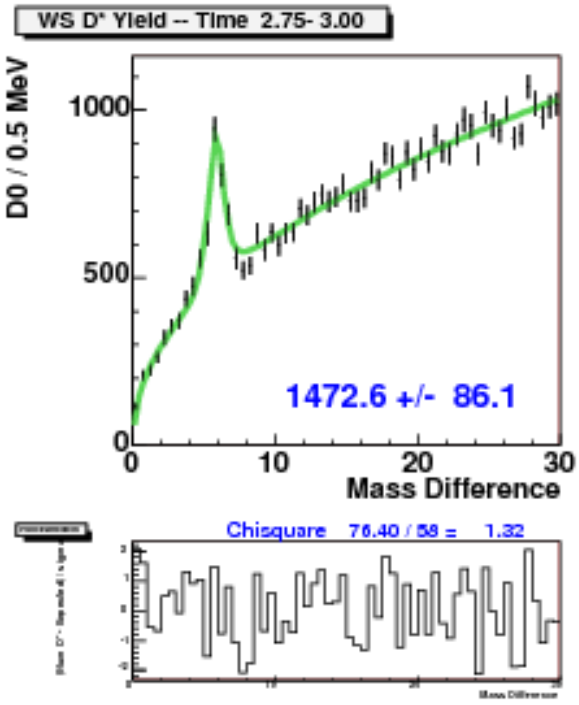



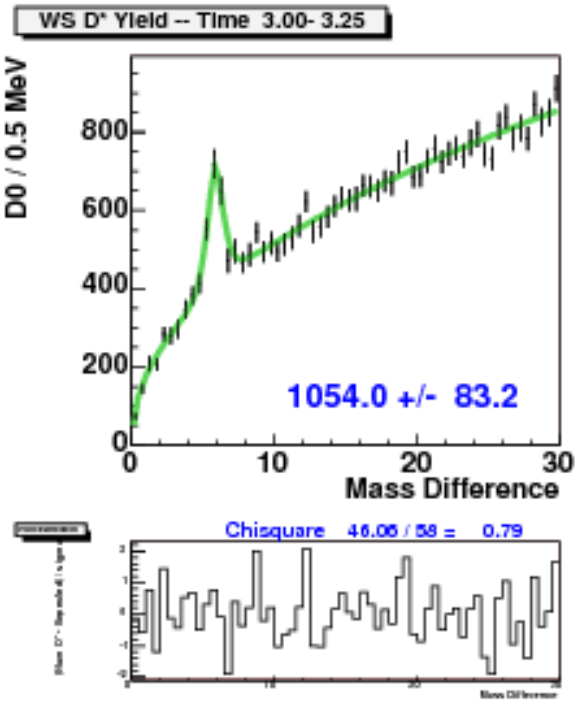



Appendix D:Title

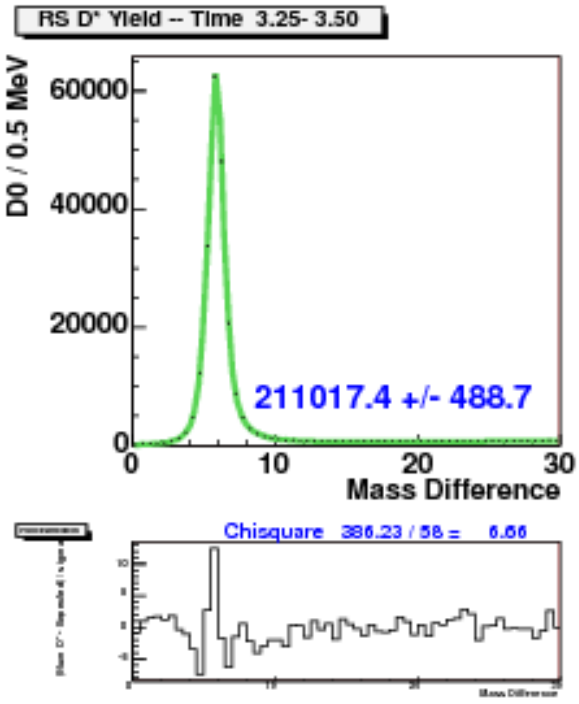


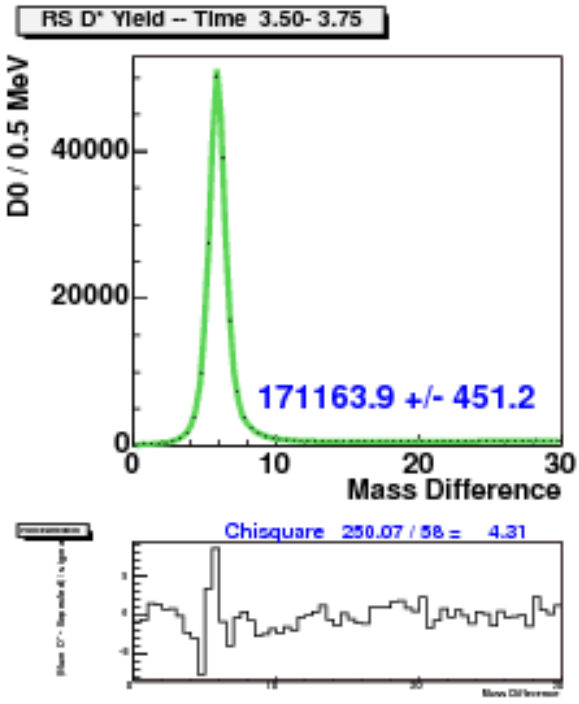



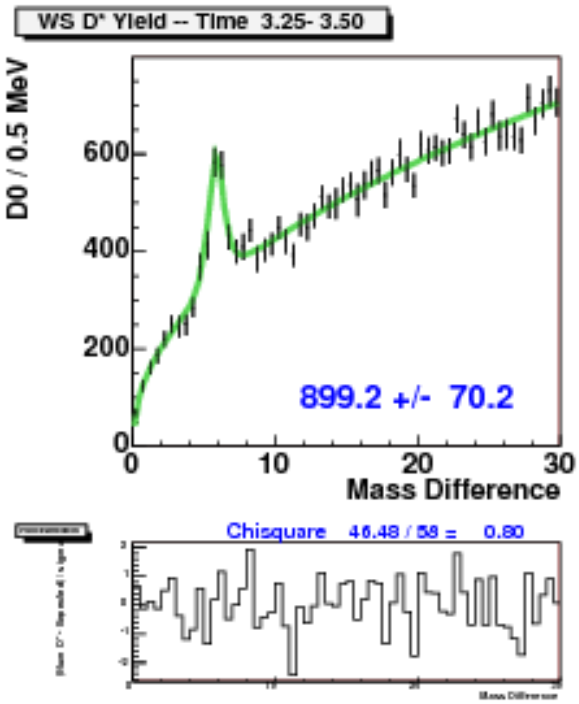



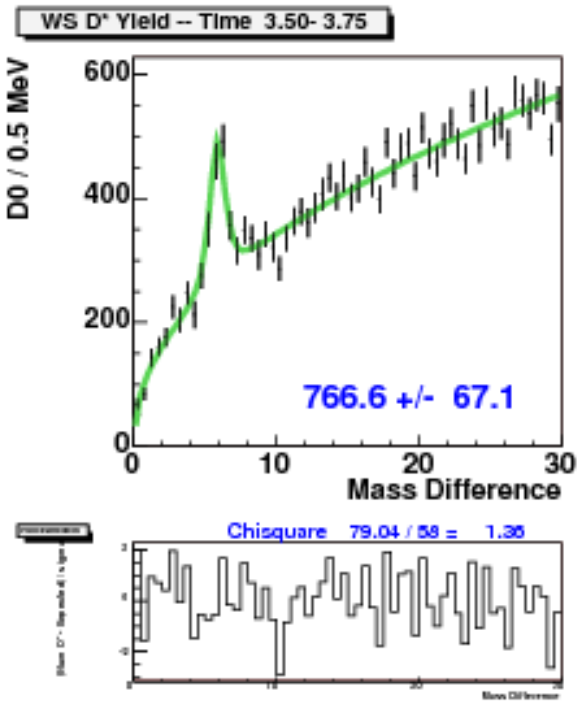



Appendix D:Title

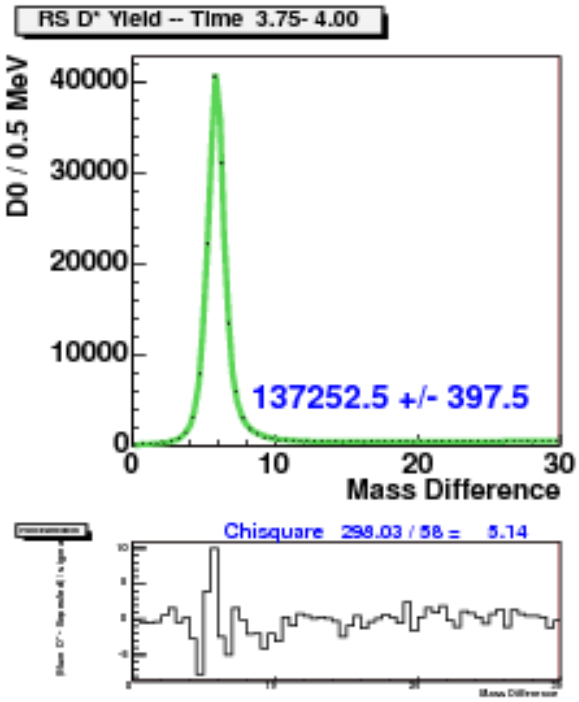


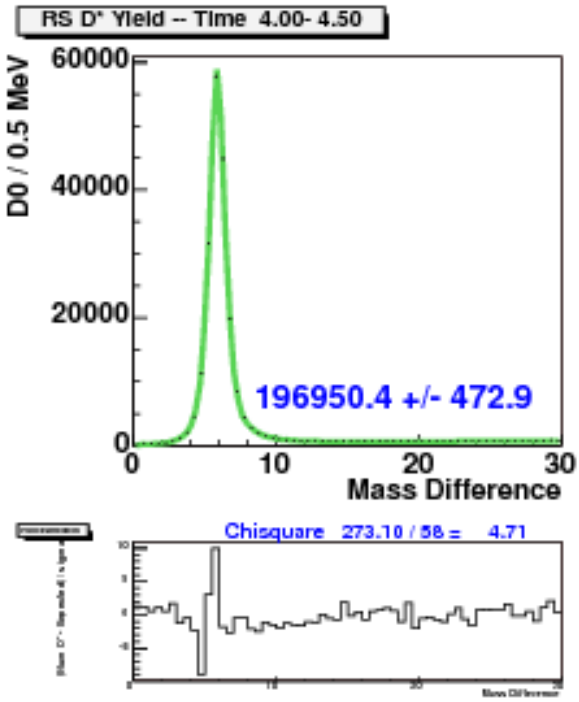



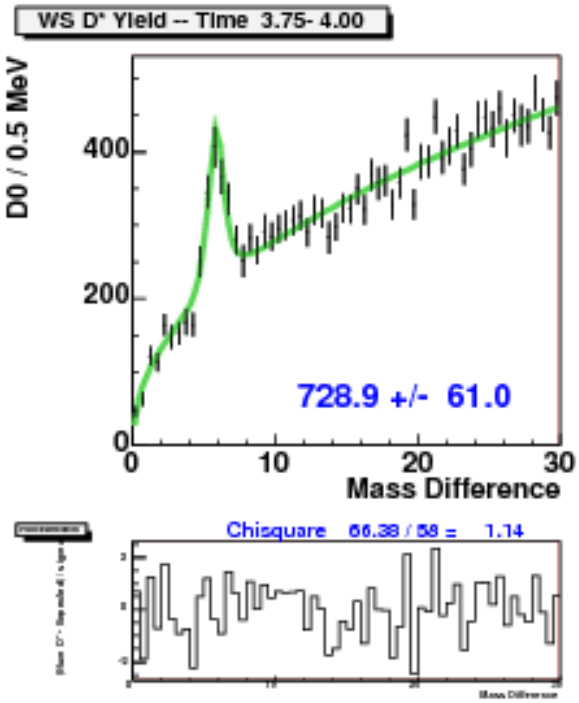



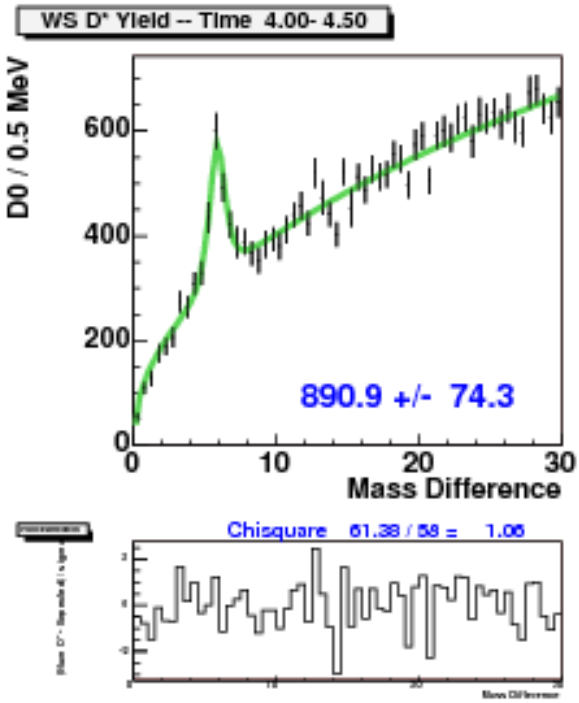



Appendix D:Title

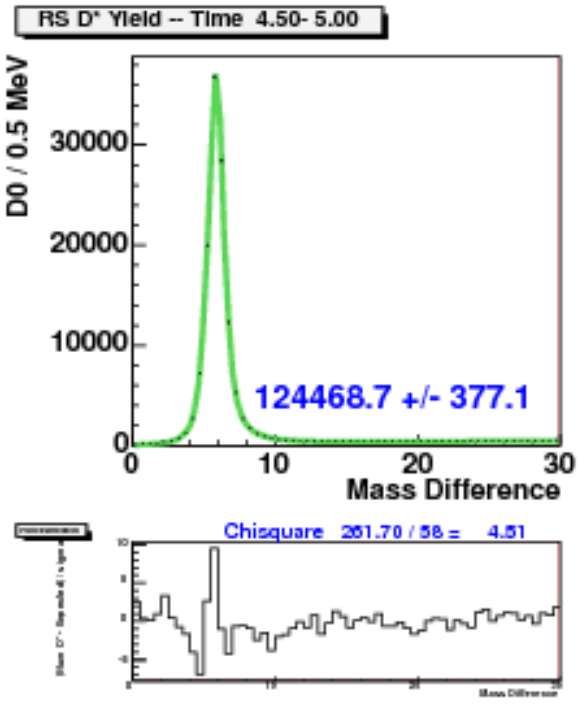


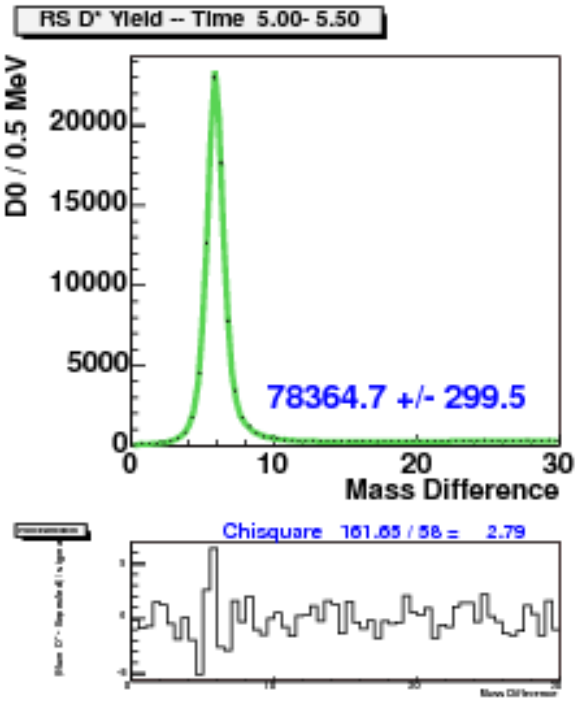



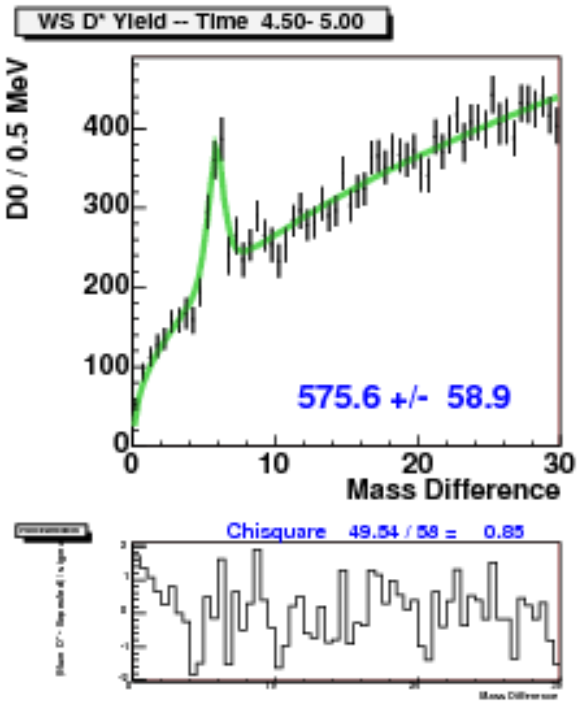



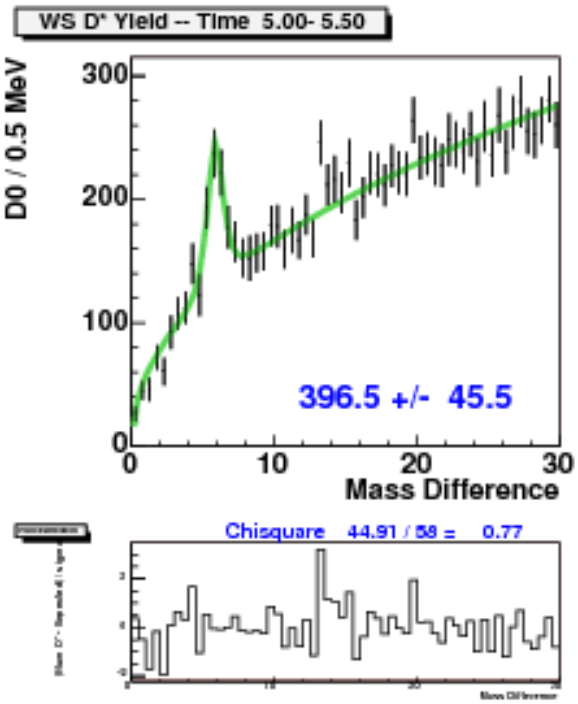



Appendix D:Title

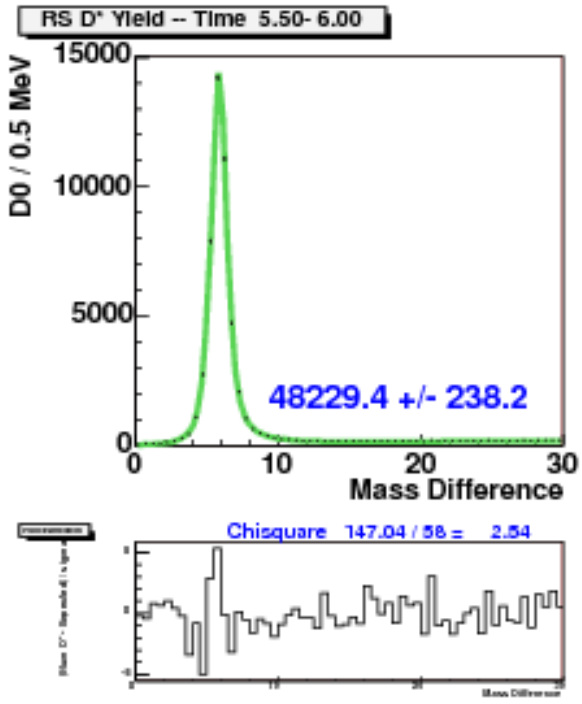


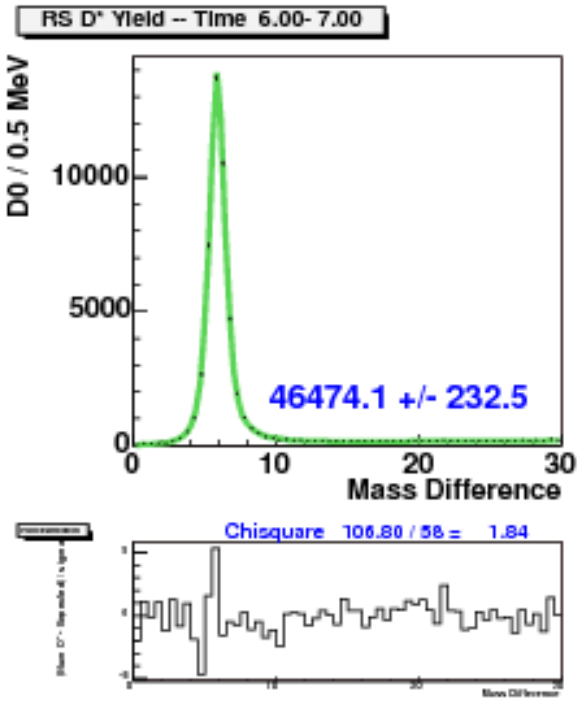



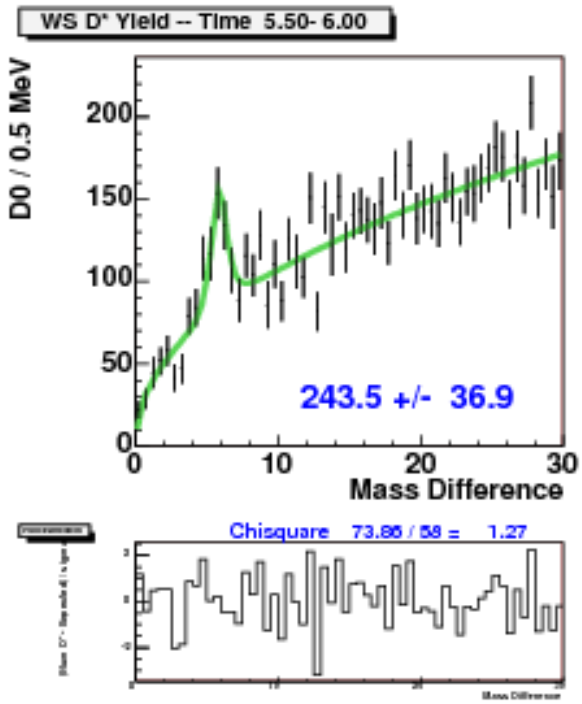



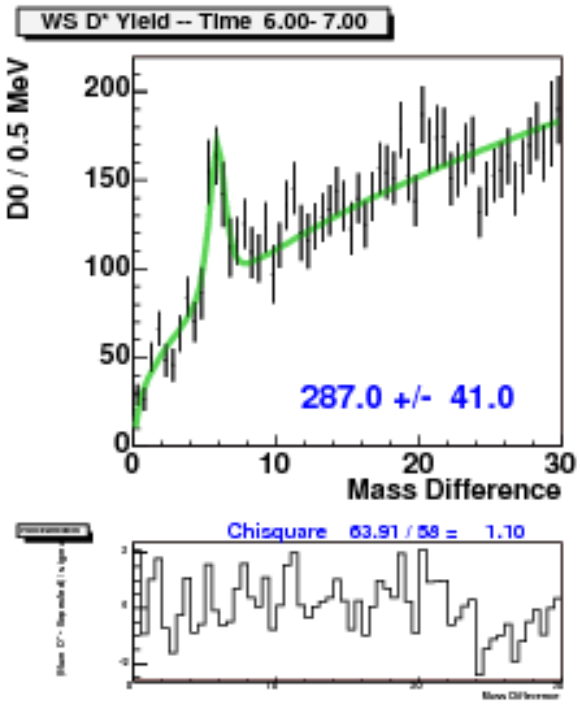



Appendix D:Title

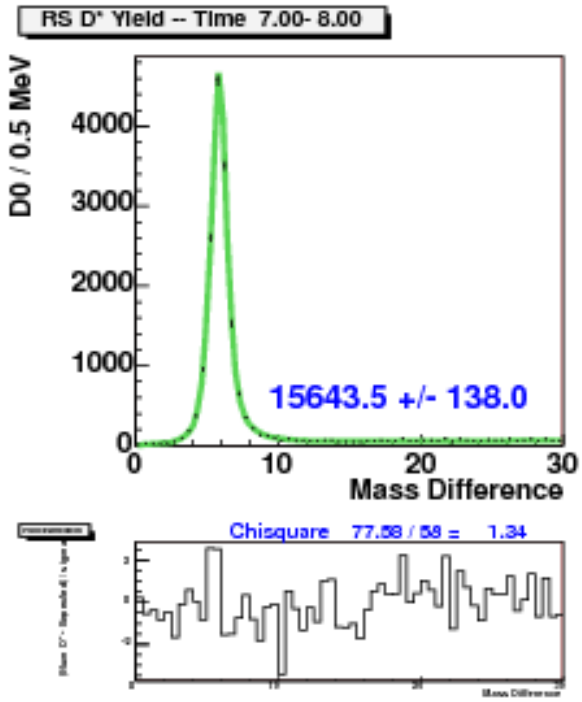


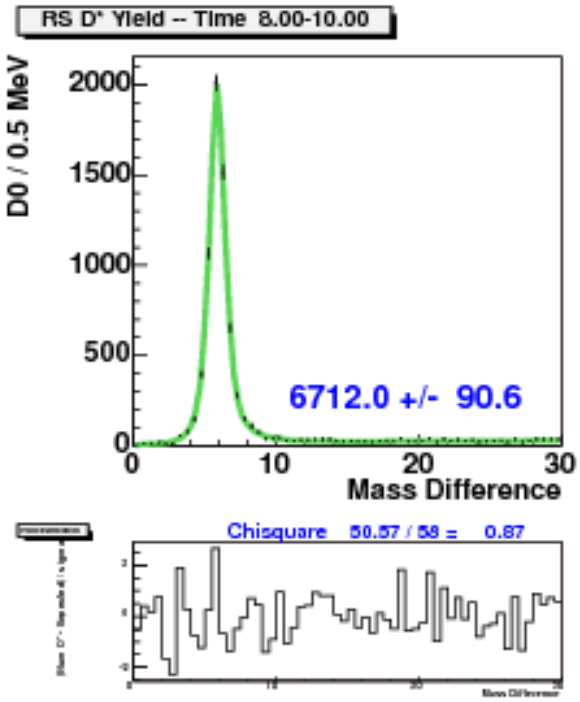



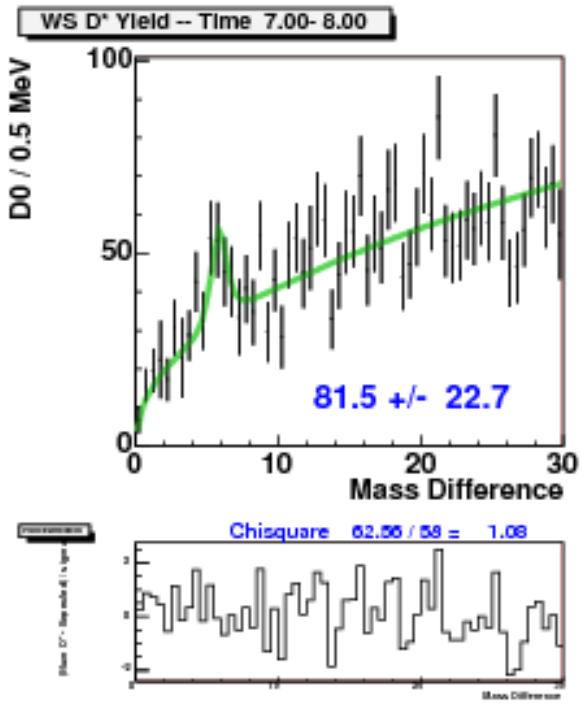



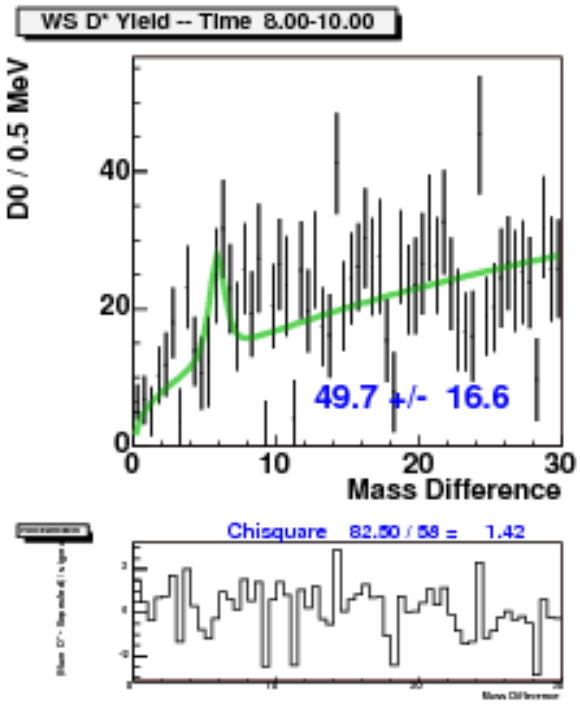



Appendix D:Title



# E Some variables used as Neural network inputs

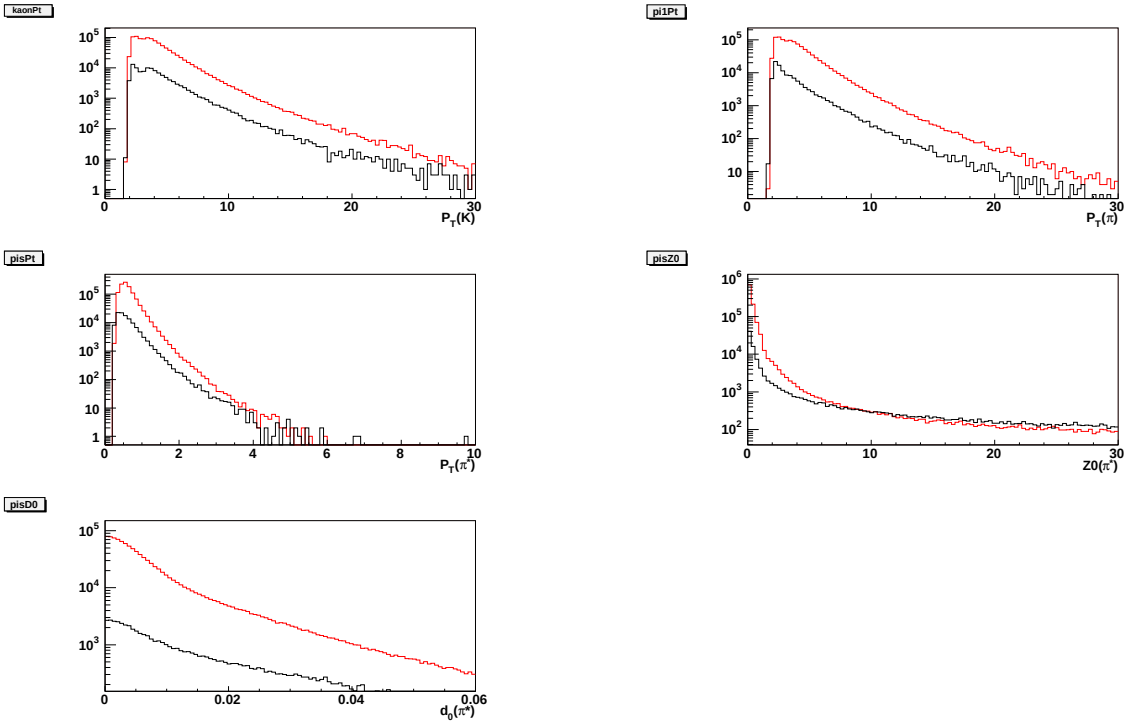



**Figure 19:** Some of the neural network inputs: 1. Transverse momentum of kaon, 2. Transverse momentum of pion, 3. Transverse momentum of soft pion, 4. The Z0 variable-the point of closest approach along the beam line -for soft pion, 5. Impact parameter for soft pion. In all plots red data shows distribution for signal and black data shows background distribution.

# F Non-Prompt $D^*$ Correction Details

This appendix is similar to Appendix D of CDF note 8879, and shows the data results for the $B$-decay correction, starting with the fits for the time-binned impact parameter distributions. We use RS data, with $(K\pi)$ sideband-subtraction. The distribution for each time bin is fit with a double Gaussian (sum of two Gaussians) for the prompt (signal) peak, and a single Gaussian for the non-prompt distribution.

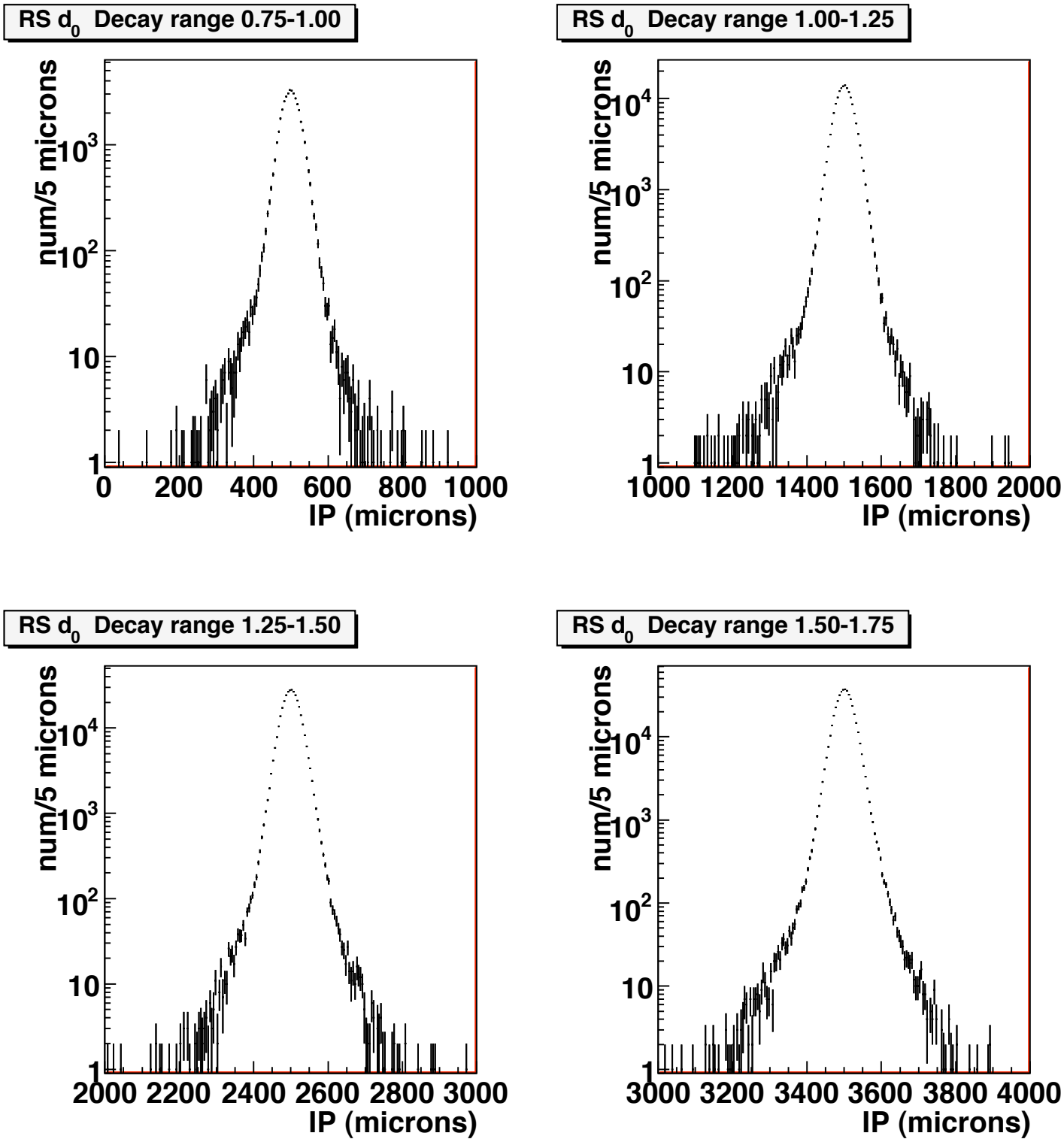
From discussions with some of the authors of the $D^0 \to \mu\mu$ analysis (CDF note 9226), the impact parameter distribution for promptly produced $D^0$ is the same for all decay times, and $D^0$ produced as a secondary decay (like from $B$ decays) have IP distributions that get wider with increasing decay time. They did MC studies to confirm this.

Each plot covers the (signed) $d_0$ values from -500 to +500 microns (ignore the x-axis numbers). The first four time bins are excluded from the fit, as the ROOT fitter could not distinguish the prompt and non-prompt shapes. The remaining 16 distributions were fit simultaneously, using the same signal shape. A polynomial of order 3 ($p_0 + p_1 t + p_2 t^2 + p_3 t^3$) is used to parametrize the non-prompt Gaussian widths as a function of decay time. The first fit parameter ($p_0$) is fixed to 27.9 $\mu m$, which is the Gaussian width when the prompt shape width is fit with a single instead of double Gaussian. The other fit parameters ($p_1, p_2$, and $p_3$) are allowed to float. The signal and background amplitudes are used to allow the fit to converge, and are not used elsewhere. The blue curve is the non-prompt fit, the green curve is prompt plus non-prompt.

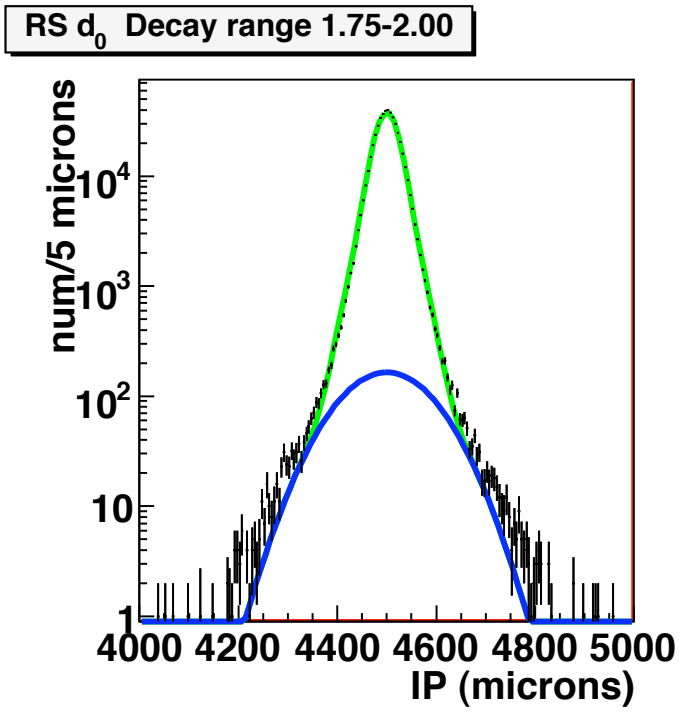
The signal fit parameters are: $f = 27.985 \pm 0.453\,\%$ (fraction of the second Gaussian to the total amplitude), $\sigma_1 = 22.024 \pm 0.054\,\mu$m (width of the first Gaussian), and $\sigma_2 = 40.347 \pm 0.235\,\mu$m (width of the second Gaussian). The parameterization for the non-prompt Gaussian width is a 3rd order polynomial: $p_0 = 27.9\,\mu$m, $p_1 = 33.676 \pm 0.549\,\mu$m, $p_2 = -0.498 \pm 0.208\,\mu$m, and $p_3 = -0.0516 \pm 0.0190\,\mu$m.

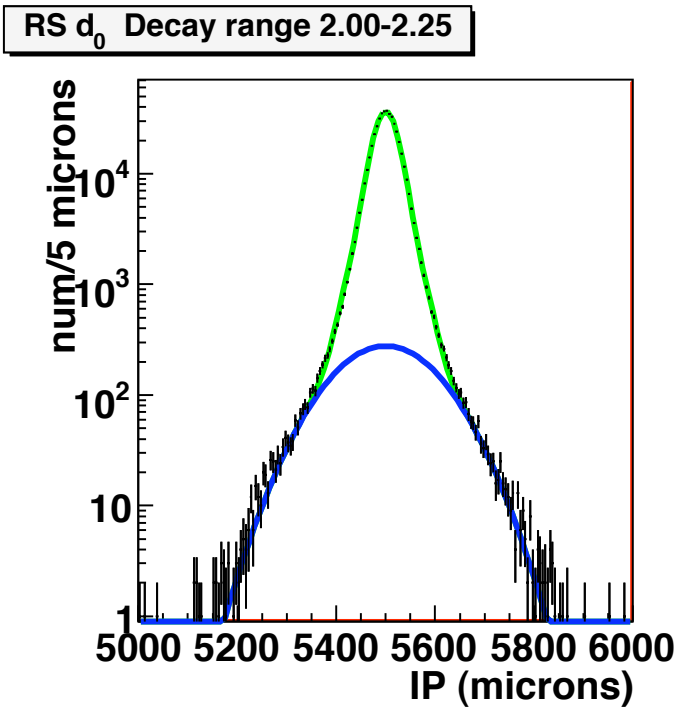
We get the number of $D^*$s (for each time bin) for two impact parameter regions: $|d_0| \leq 60\mu$m (inside), and $60 < |d_0| \leq 500\mu$m (outside). These are used to correct for background from $D^*$ not produced at the primary vertex. The amount of the prompt and non-prompt distributions in each IP region is summarized in table 8.

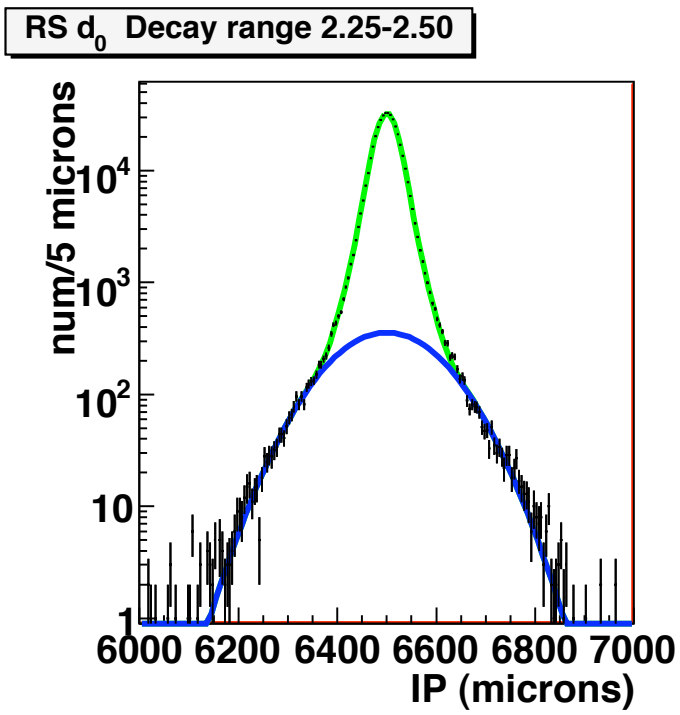
Appendix F: RS $d_0$ Distributions By Decay Time

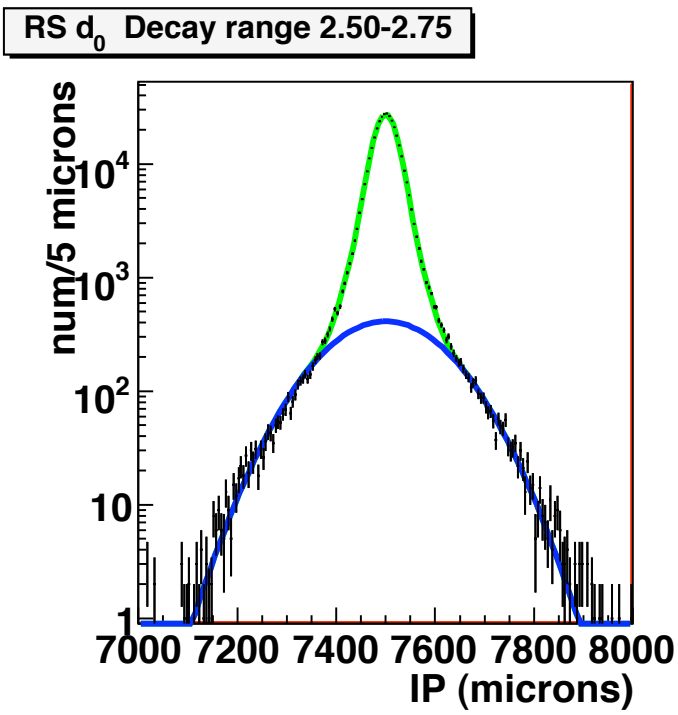


Appendix F: RS $d_0$ Distributions By Decay Time

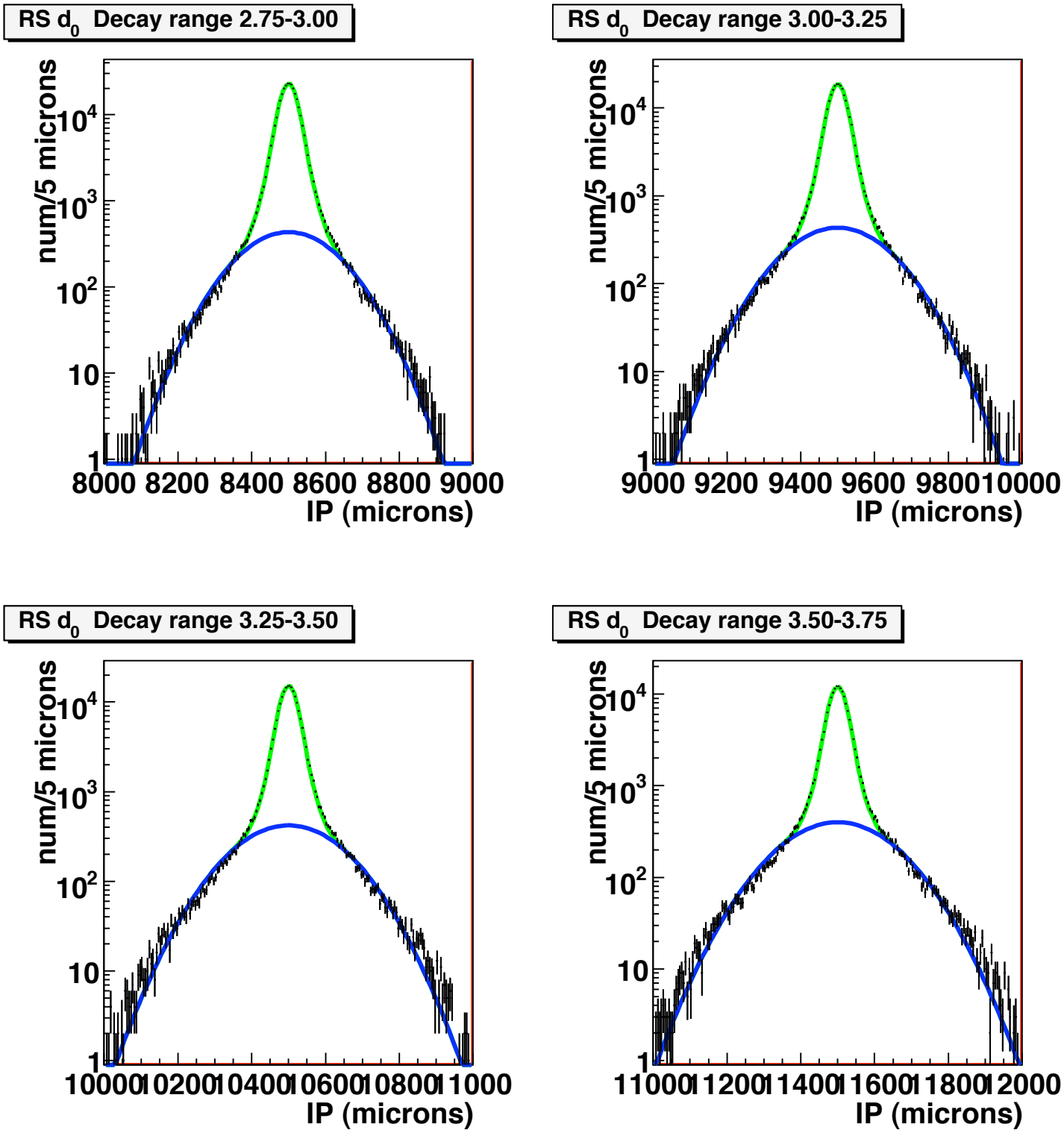


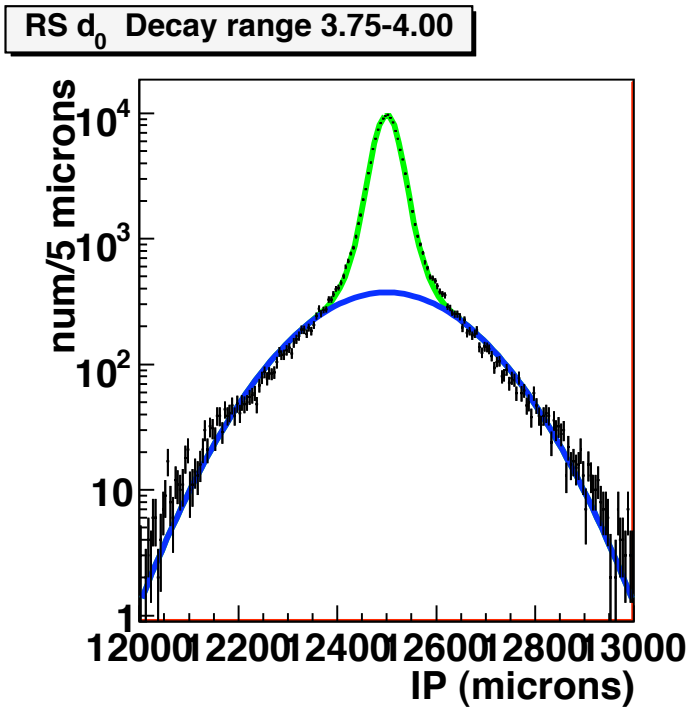



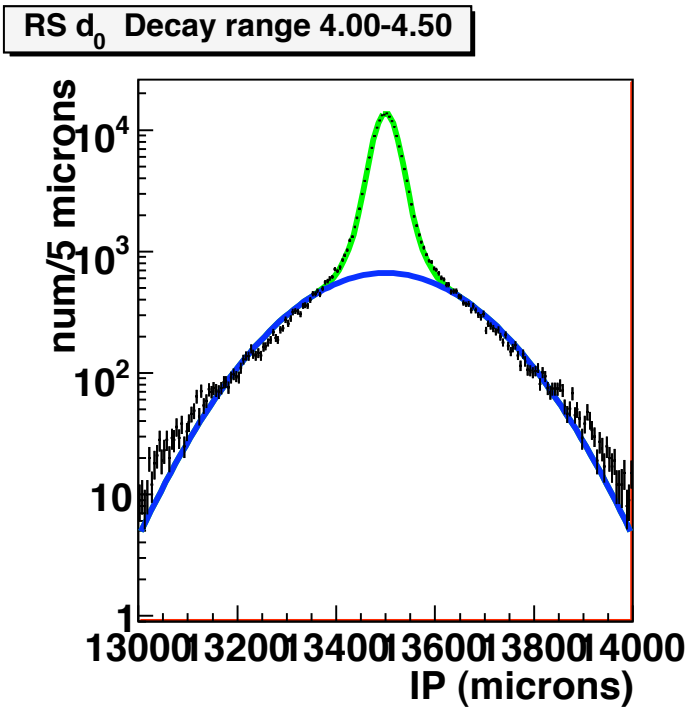



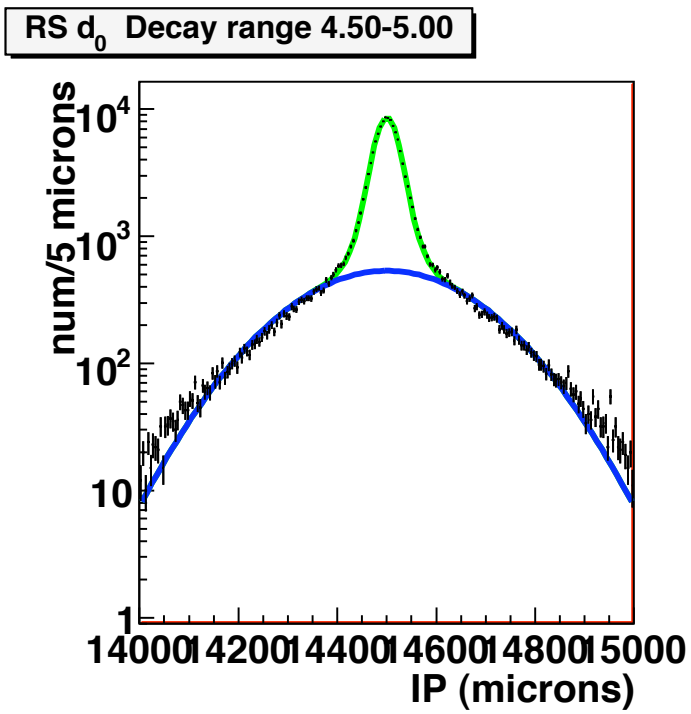
## Appendix F: RS $d_0$ Distributions By Decay Time

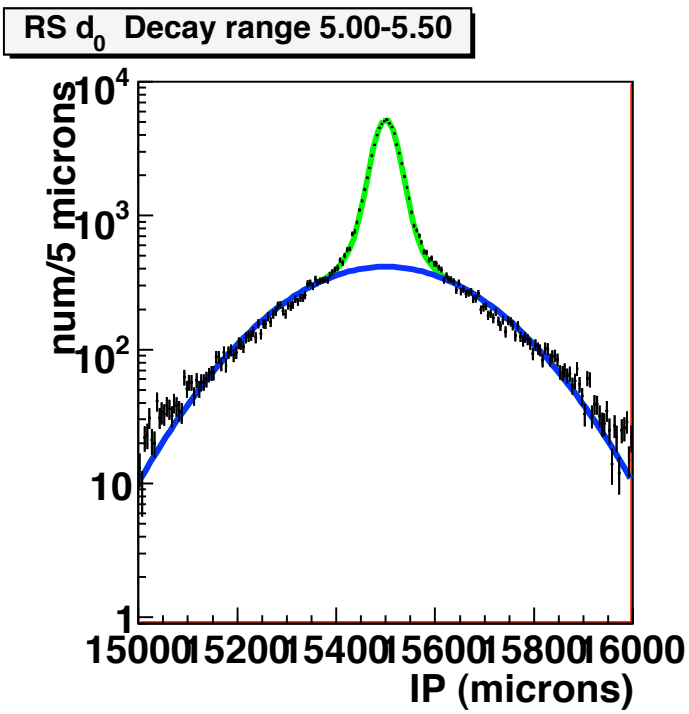


Appendix F: RS $d_0$ Distributions By Decay Time

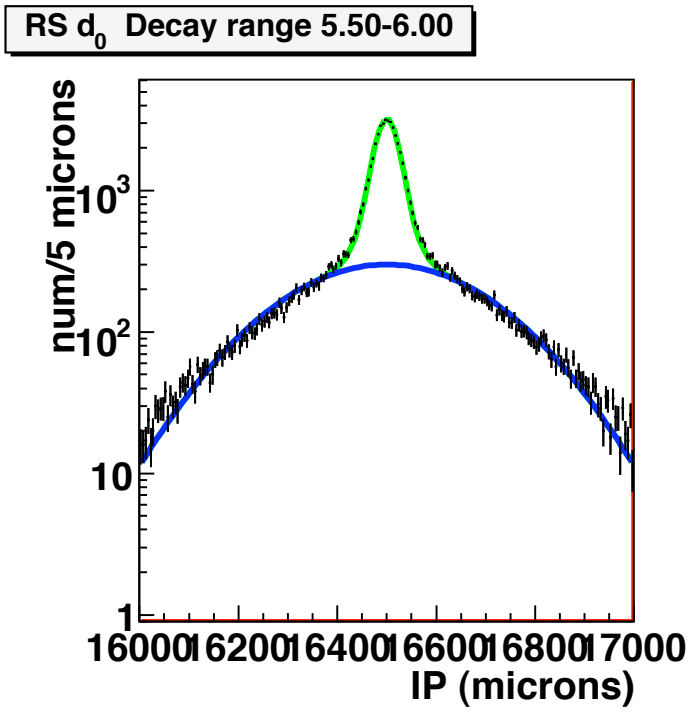


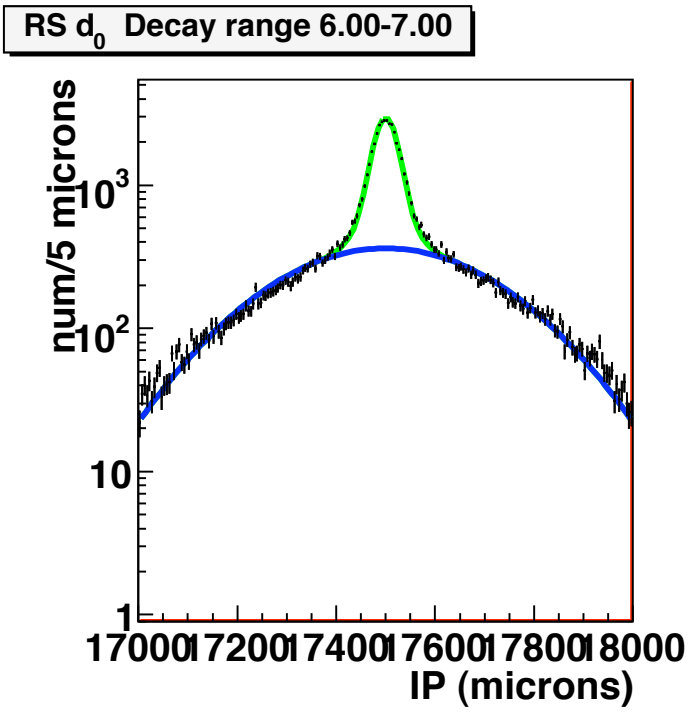



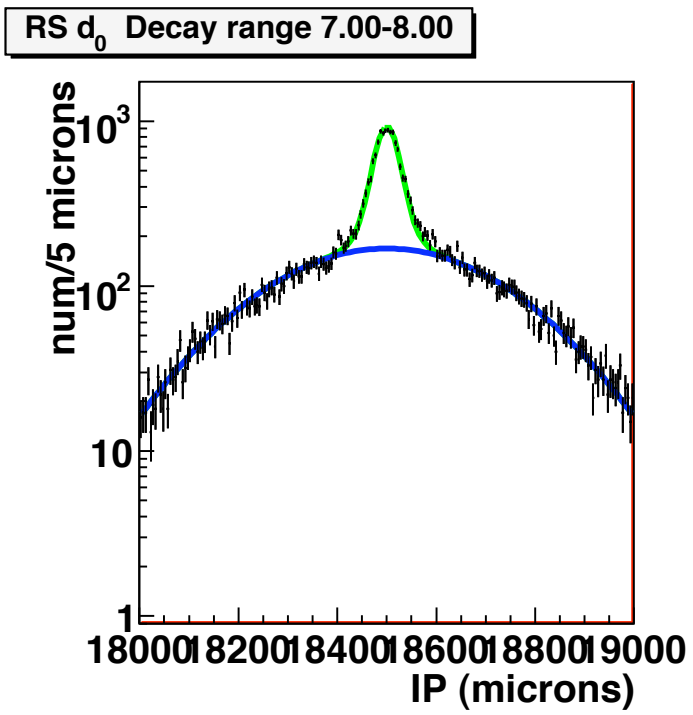



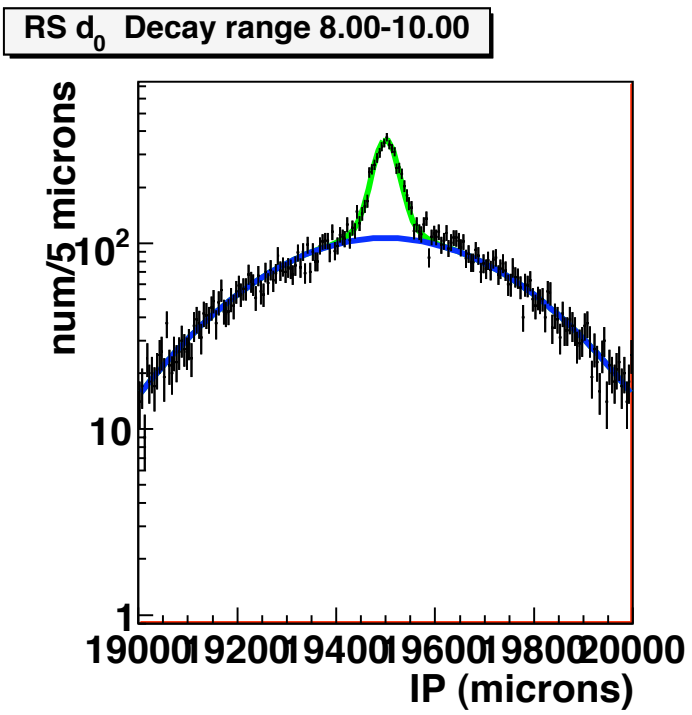



Appendix F: RS $d_0$ Distributions By Decay Time



| Time Bin | Fraction In ($f_i$) | $\sigma_{f_i}$ | Fraction Out ($g_i$) |
|---|---|---|---|
| prompt distribution | 0.95702 | 0.00024 | 0.04298 |
| 0.75-1.00 | 0.70800 | 0.00271 | 0.29200 |
| 1.00-1.25 | 0.64350 | 0.00304 | 0.35650 |
| 1.25-1.50 | 0.58812 | 0.00321 | 0.41188 |
| 1.50-1.75 | 0.54072 | 0.00329 | 0.45928 |
| 1.75-2.00 | 0.50007 | 0.00330 | 0.49993 |
| 2.00-2.25 | 0.46502 | 0.00328 | 0.53498 |
| 2.25-2.50 | 0.43462 | 0.00323 | 0.56538 |
| 2.50-2.75 | 0.40809 | 0.00318 | 0.59191 |
| 2.75-3.00 | 0.38479 | 0.00311 | 0.61521 |
| 3.00-3.25 | 0.36421 | 0.00304 | 0.63579 |
| 3.25-3.50 | 0.34593 | 0.00298 | 0.65407 |
| 3.50-3.75 | 0.32963 | 0.00292 | 0.67037 |
| 3.75-4.00 | 0.31501 | 0.00285 | 0.68499 |
| 4.0-4.5 | 0.29576 | 0.00277 | 0.70424 |
| 4.5-5.0 | 0.27418 | 0.00267 | 0.72582 |
| 5.0-5.5 | 0.25631 | 0.00259 | 0.74369 |
| 5.5-6.0 | 0.24137 | 0.00252 | 0.75863 |
| 6.0-7.0 | 0.22327 | 0.00244 | 0.77673 |
| 7.0-8.0 | 0.20521 | 0.00238 | 0.79479 |
| 8.0-10.0 | 0.18744 | 0.00239 | 0.81256 |

**Table 8:** Prompt and non-prompt distribution fractions. $\sigma_{f_i}$ is the uncertainty on $f_i$ based on the uncertainties of the parameterization for the non-prompt Gaussian width.

# References


[1] BaBar Collaboration, B. Aubert *et al.*, *Evidence for $D^0 - \overline{D}^0$ Mixing*, Phys. Rev. Lett. **98**, 211802 (2007).

[2] CDF, T. Aaltonen et al., *Evidence for $D^0 - \bar{D}^0$ mixing using the CDF II Detector*, Phys. Rev. Lett. **100**, 121802 (2008), 0712.1567.

[3] Belle Collaboration, M. Starič *et al.*, *Evidence for $D^0 - \overline{D}^0$ Mixing*, Phys. Rev. Lett. **98**, 211803 (2007).

[4] BELLE, K. Abe et al., *Measurement of D0-D0bar mixing in D0-¿Ks pi+ pi- decays*, Phys. Rev. Lett. **99**, 131803 (2007), 0704.1000.

[5] BABAR, B. Aubert et al., *Measurement of $D^0 - \bar{D}^0$ mixing from a time- dependent amplitude analysis of $D^0 \to K^+\pi^-\pi^0$ decays*, (2008), 0807.4544.

[6] CLEO, R. Godang et al., *Search for D0 anti-D0 mixing*, Phys. Rev. Lett. **84**, 5038 (2000), hep-ex/0001060.

[7] FOCUS, J. M. Link et al., *Measurement of the branching ratio of the decay D0 –¿ pi- mu+ nu relative to D0 –¿ K- mu+ nu*, Phys. Lett. **B607**, 51 (2005), hep-ex/0410068.

[8] Heavy Flavor Averaging Group, E. Barberio et al., *Averages of b−hadron and c−hadron Properties at the End of 2007*, (2008), 0808.1297.

[9] Mark Mattson and Paul Karchin, *Analysis of $D^0 - \bar{D}^0$ Mixing in the $K\pi$ Channel*, CDF-Note 7116, 2006.

[10] Mark Mattson, Paul Karchin, and Nagesh Kulkarni, *Analysis of $D^0 - \bar{D}^0$ Mixing in the $K\pi$ Channel*, CDF-Note 8879, 2007.

[11] Adam F. Falk, Yuval Grossman, Zoltan Ligeti, Yosef Nir, and Alexey A. Petrov, *The D0 - anti-D0 mass difference from a dispersion relation*, Phys. Rev. **D69**, 114021 (2004), hep-ph/0402204.

[12] Particle Data Group, Yao, W.-M. and others, *Review of Particle Physics*, Journal of Physics G **33**, 1+ (2006).

[13] http://www cdf.fnal.gov/htbin/twiki/bin/view/BStntuples/WebHome.

[14] http://www.phi tps.de.