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

Mark Mattson, Paul Karchin, and Nagesh Kulkarni
Wayne State University, Detroit, MI
CDF/DOC/BOTTOM/CDFR/8879
Version 1.4

October 6, 2007

Abstract

We present a measurement of $R_{\mathcal{B}}(t)$, the time-dependent ratio of the branching fraction for the rare decay $D^0 \to K^+\pi^-$ to that for the Cabibbo-favored decay $D^0 \to K^-\pi^+$. The rare decay will be composed of the doubly Cabibbo suppressed D^0 decay, and $D^0 - \bar{D^0}$ mixing. If $R_{\mathcal{B}}$ changes as decay time increases, this is evidence for charm mixing. We present a method to get the decay time-binned ratio, which is fit to obtain the mixing parameters (R_D, y', x'^2). The results are compared to what is expected if there were no charm mixing, to set a limit on mixing or to make an observation.

This note describes the time-dependent measurement using Two Track Trigger data with 1.5 fb^{-1} marked good for B physics. (The details for the previous analysis for the time-independent measurement with an integrated luminosity of 0.35 fb^{-1} are in CDF note 7116.) We have approximately 3.0 million CF D^0 decays and 12.7 thousand DCS/mixing decays from D^* decays coming from the primary vertex.

The best fit for the mixing parameters is $R_d = 3.04 \pm 0.55(\times 10^{-3}), y' = 8.54 \pm 7.55(\times 10^{-3})$, and $x'^2 = -0.12 \pm 0.35(\times 10^{-3})$. The no-mixing point ($y' = x'^2 = 0$) is excluded with a significance equivalent to 3.8 Gaussian standard deviations. This measurement is competitive with the current world-best results by Belle and BaBar.

Versions:

- 1.0 Document posted for pre-blessing.
- 1.1 Pre-blessing document, with unblinded data plots and the modification of the non-prompt WS correction (section 4.2.1).
- 1.2 Document posted for blessing. The modification of the non-prompt WS correction (section 4.2.1) was rolled back. Unblinded data plots updated. Updated section on statistics to include some checks of our Bayesian method. Section 3.2.2 will now be called "RS Background at 1.83 GeV", and has more text.
- 1.3 Document posted before 1st paper draft. Data included through period 10 ($1.5fb^{-1}$). Ch 2 includes the study to validate the data sets (periods 8-10). Ch 3 has the recent "extra lump" study. Ch 4 updated B-decay correction. Ch 5 contour info expanded. "Future work" appendix removed. Appendix data results updated.
- 1.4 Fixed a few typos and outdated numbers. Added section 5.6 for p-value and coverage using simple simulation.

Contents

List of Figures

List of Tables

1 Introduction

Since the discovery of the charm quark in 1974, physicists have been searching for the mixing (or oscillation) of neutral charm mesons between particle and anti-particle states. The analogue process was discovered for kaons in 1962 and for B_d mesons in 1987. The year 2007 has seen landmark new results on mixing: observation of B_s mixing by CDF and evidence for D^0 mixing from Belle and BaBar.

In the standard model, the decay $D^0 \to K^+\pi^-$ proceeds through a doubly Cabibbo-suppressed (DCS) tree diagram (see Fig. 1) and possibly through a "mixing" process in which the D^0 changes into a $\overline{D}^0$. (In this section, discussion of a decay reaction implicitly includes the charge conjugate process, unless otherwise noted.) The DCS decay rate depends on CKM factors as well as the magnitude of SU(3) flavor symmetry violation. Mixing may occur through two distinct types of second-order weak processes. In the first, shown in Fig. 2, the D^0 decays into a virtual ("long-range") intermediate state such as $\pi^+\pi^-$, which subsequently decays into a $\overline{D}^0$. The second type, shown in Fig. 3, is a short range process, with either a "box" or "penguin" topology. It is not established whether long range mixing occurs. Its strength depends on SU(3) flavor symmetry violation. Short range mixing is negligible in the SM. However, exotic weakly interacting particles could enhance the short range mixing and provide a signature of new physics.

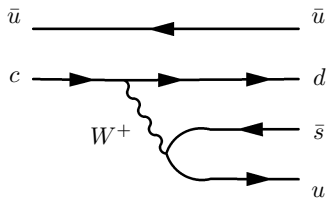


Figure 1: Feynman diagram for doubly- Cabibbo- suppressed decay.

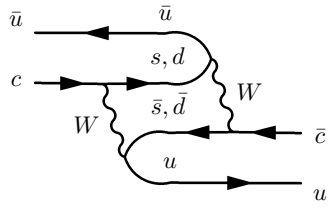


Figure 2: D^0 mixing through a virtual "long-range" intermediate state.

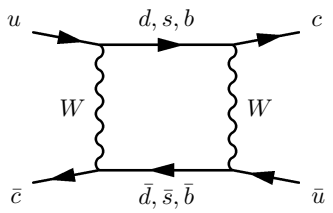


Figure 3: Highly suppressed standard model short-range D^0 mixing via a box diagram.

The experimental method we are exploring at CDF exploits the decay chain $D^{*+} \to D^0\pi^+$. The subsequent decay $D^0 \to K^-\pi^+$ is Cabibbo-favored and experimentally well established. Doubly Cabibbo-suppressed decay and mixing can result in the decay $D^0 \to K^+\pi^-$. These so-called *wrong*

sign decays are identified by the decay chain $D^{*+} \to D^0\pi^+$, $D^0 \to K^+\pi^-$ in which the charge of the pion from the D^{*+} decay is opposite to the charge of the pion from the D^0 decay. The pion from the D^{*+} decay is called the *tagging* pion because its charge identifies the state of the neutral D meson as particle (D^0) or anti-particle ($\overline{D}^0$). The tagging pion is also called the *slow pion* because its laboratory momentum is much less than that of the D^0 or $\overline{D}^0$. This kinematic property is a result of the small Q-value (145.4 MeV) of the D^{*+} decay. In the subsequent discussion, the term *right sign* will denote the Cabibbo favored decay, $D^0 \to K^-\pi^+$.

The ratio $R(t)$ of wrong-sign to right-sign decay rates can be expressed [1] as a simple quadratic function of proper time t under the assumption of CP conservation and small values for the parameters x and y. The parameter x is defined in terms of the mass difference Δm between the heavy and light mass eigenstates and the parameter y involves the mass width difference $\Delta\Gamma$ between these eigenstates according to,

$$x = \Delta m/\Gamma \quad \text{and} \quad y = \Delta\Gamma/2\Gamma$$

where Γ is the average mass width of the mass eigenstates. Under the assumptions stated above,

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

The parameters x' and y' are linear combinations of x and y according to the relations,

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

where δ is a strong interaction phase.

1.1 Experimental Status

Last year, CDF published a measurement [2] of R_B, the ratio of branching fractions of wrong-sign and right-sign decays,

$$R_{\mathcal{B}} = \mathcal{B}(D^0 \to K^+\pi^-)/\mathcal{B}(D^0 \to K^-\pi^+) \tag{2}$$

This ratio is given by the ratio of the time-integrals of the corresponding decay rates, yielding,

$$R_{\mathcal{B}} = R_D + \sqrt{R_D}y' + \frac{x'^2 + y'^2}{2}. \tag{3}$$

Thus $R_{\mathcal{B}}$ is sensitive to the three physics parameters R_D, x' and y', but does not provide a separate measure of them. This is possible from the time-dependent measurement described above. Previous time-dependent measurements have been reported for R_D and have set limits on x'^2 and y'. The existing measurements, significance and/or limits are summarized in Table 1 and are discussed below.

Experiment	$R_D(10^{-3})$	$y'\,(10^{-3})$	$x'^2\,(10^{-3})$	No Mixing Signficance
CLEO[3]	$4.8 \pm 1.2 \pm 0.4$	$-25\ ^{+14}_{-16} \pm 3$ $-58 < y' < 10$	$0 \pm 15 \pm 2$ $x'^2 < 0.82$	–
BaBar[4]	$3.03 \pm 0.16 \pm 0.10$	$9.7 \pm 4.4 \pm 3.1$ –	$-0.22 \pm 0.30 \pm 0.21$ –	10^{-4} (3.9 standard deviations)
Belle[5]	3.64 ± 0.17	$0.6\ ^{+4.0}_{-3.9}$ $-9.9 < y' < 6.8$	$0.18\ ^{+0.21}_{-0.23}$ $x'^2 < 0.72$	3.9%

Table 1: Published results for charm mixing parameters. The uncertainties are statistical then systematic, except for Belle's numbers which have both. Below the best fit results are the 95% C.L. on the mixing parameters. The CLEO result from 2000 is listed to how recent measurements by BaBar and Belle have improved the measurements on the charm mixing parameters (R_D, y', x'^2). These results use $D^0 \to K\pi$ decays only.

At the Moriond Conference on March 13, 2007, the Belle [6] and BaBar [7] collaborations reported evidence for $D^0 - \overline{D}^0$ mixing, using different techniques. The results are now published [8, 4].

BaBar sees a direct mixing signal using the same decay chain as studied here. So far, for this decay chain, Belle has reported only limits for x'^2 and y'. Thus, the BaBar evidence remains unconfirmed and has provided the motivation for the accelerated analysis of CDF data presented here.

It is also worth noting that the evidence for D^0 - $\bar{D}^0$ mixing presented by Belle is based on two methods, both different than the one discussed here for CDF and BaBar. One of the Belle methods is a comparison of the decay time distributions for D^0 to CP-eigenstates K^+K^- and $\pi^+\pi^-$ with the decay time distribution for D^0 to the CP-mixed state $K^-\pi^+$. They measure a non-zero value of y and hence indirect evidence of mixing.

The other method used by Belle is a time-dependent measurement of the Dalitz decay $D^0 \to K^0_s\pi^+\pi^-$ [9]. They report a preliminary non-zero value for x and a value for y consistent with zero. The non-zero value for x constitutes direct evidence for mixing.

There are no recent measurements to compare with either of Belle's new results - they remain unconfirmed.

2 Data Sample

This note describes the time-dependent measurement using Two Track Trigger data with 1.5 fb^{-1} marked good for B physics. (The details for the previous analysis for the time-independent measurement with an integrated luminosity of 0.35 fb^{-1} are in CDF note 7116.)

2.1 BStNtuples

To speed up processing, we are using the BStNtuple files [10]. These are the B-Physics ntuples built from the production StNtuples, with an emphasis on quantities used in heavy flavor physics analyses. We then make our own ntuple files with quantities specific to this analysis.

The oldest bsntuple set (xbhd0d) was made before a necessary reconstruction ($B \to \pi\pi$) was included in the processing. Fortunately, Ivan Furic pointed out the files made by Christoph Paus for the MIT group. We are using those files for the 0d sample, and the "standard" BsNtuples for 0h and 0i. We have confirmed that the Paus files for 0h and 0i produce the same results as the standard files, so we are confident that the 0d set is acceptable. When the official production 0d sample is reprocessed, we can start using that data set. The BStNtuple catalogs are located on the offline CDF disk area, cdfopr/cafdfc/ directory: paus/h77jh0 (0d data), bottom/bmix-60/h77jg0 (0h), bottom/bmix-70/h77jm0 (0i), and bottom/hadr-80/xbhdii (0i for data past $1fb^{-1}$).

2.2 Analysis Ntuples

The BStNtuple files handle track reconstruction and vertexing, but we still need to make our own ntuple files with quantities specific to this analysis. The BStNtuple has reconstructions for D^0 and D^*, which we cannot use. These reconstruction blocks are made for all events in the production StNtuple, but are not used to select which events are written to BStNtuple file. As such, starting from these reconstruction blocks resulted in less than 50% of the expected signal.

Instead, we are starting from the $B \to \pi\pi$ block. This makes a vertex from the oppositely charged TTT trigger tracks, and any event with a candidate is written out. The events in this block are required to have the decay vertex with $L_{xy}/\sigma_{xy} > 4$, but there is no mass selection. All available D^0s will be in the BStNtuple files, after changing the two tracks to be $K\pi$.

The pion track block is used to get extra pions, which are combined with the D^0 candidate to form a D^* candidate. The mass difference ($\Delta m = m_{K\pi\pi^*} - m_{K\pi} - m_{\pi^*}$) can be anywhere from 0 - 2 GeV. We only retain candidates with $\Delta m < 30$ MeV, which will retain all possible D^* decays, and provide enough sideband events to observe the background distribution.

2.3 Data Sets

Table 2 lists the data sets we are using. The luminosity is based on the Data Quality Monitoring page, using the online luminosity times the percent "good" for B-physics. We are using the Two Track trigger. Track reconstruction uses ISL and L00 silicon. The D^0 yield depended on the running conditions and trigger prescales. The number is used as a sanity check. We detected a bug in our code early when the yield for one of the data periods dropped to 500.

Data Set	Data Periods	Run Range	Online Lum (pb^{-1})	"Good" Lum	Dates	D^0 per pb^{-1}
xbhd0d	0	138425-186598	470	362	02/2002 - 08/2004	1953
xbhd0h	1, 2, 3, 4	190697-203799	429	346	12/2004 - 09/2005	2295
xbhd0i	5, 6, 7	203819-212133	265	244	09/2005 - 02/2006	1578
xbhdii	8, 9	217990-2228596	385	343	06/2006 - 09/2006	1583
xbhdii	10	228664-233111	241	215	09/2006 - 01/2007	1344

Table 2: CDF Data Sets. Online luminosity and luminosity marked good for B physics are taken from the Data Quality Monitoring Home Page.

To further validate the data, we looked at the $K\pi$ mass distribution and the dE/dX particle identification. Figure 4 shows the D^0 signal shape distribution, for the five data periods listed in

Table 2. When we analyze the entire data sample, the signal shape will work well with events from all five data periods.

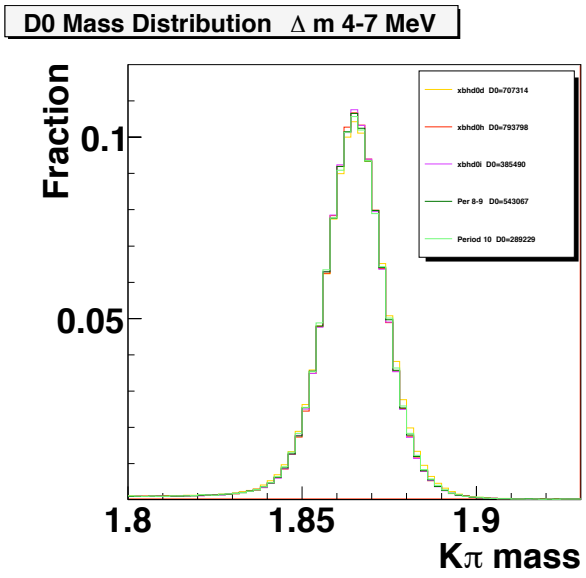


Figure 4: D^0 signal shape distributions for the data sets.

Figure 5 shows the particle identification distributions. The period 10 results are different from the other results, with the peak narrower and shifted approximately half a unit negative. Period 10 was the start of data taking with dE/dX turned off for the innermost COT superlayers (due to increased instantaneous luminosity at the start of accelerator stores). The adjustment may be correctable, but we have not checked that. Fortunately, the part id cut that we use (described in the next chapter) still works with period 10. Figure 6 shows the part id cut variable distribution. For right-sign D^*s, the variable is required to be negative. Period 10 has 84.8% of the events passing this requirement, compared to the 0d set (85.0%) and the remainder of the data (86%). Including period 10 will not bias the results, and can be treated the same as the other data periods (within errors).

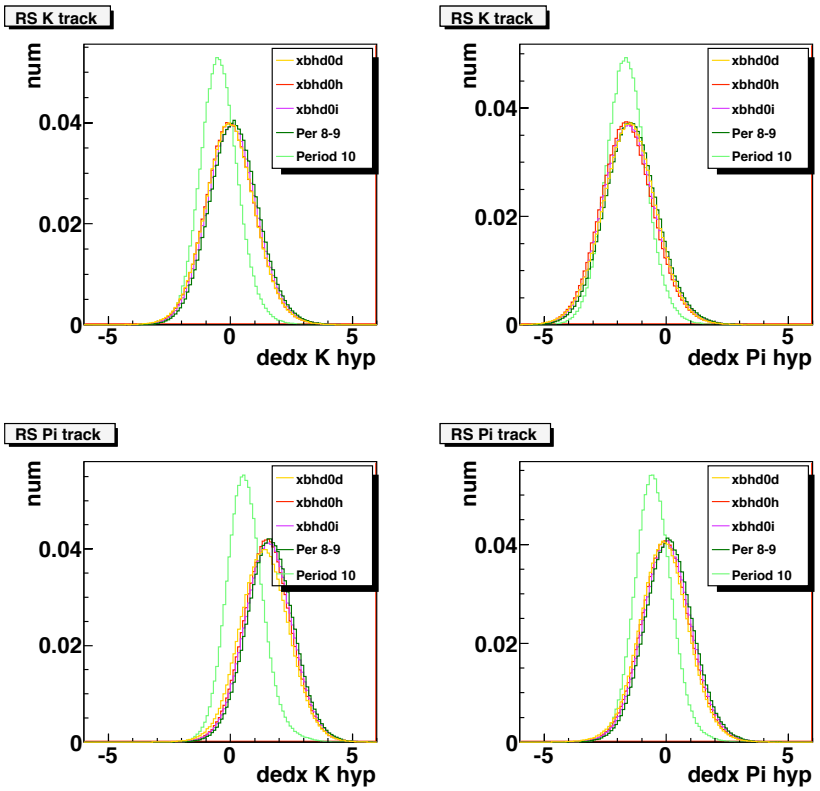


Figure 5: Particle identification distributions for the data sets. Right-sign D^0 events are used. The top row is the K track, the bottom row is the π track. The left side is dE/dX with a kaon track hypothesis, the right side with a pion track hypothesis.

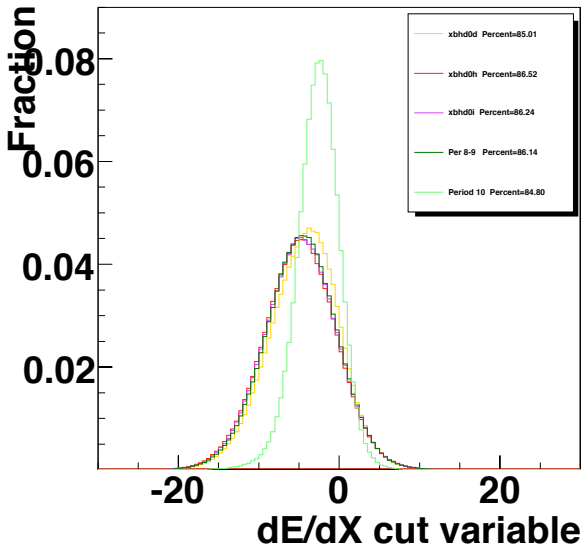


Figure 6: Part id cut variable distributions for the data sets. The part id variable is described in section 3.2.3

3 Method to Get D^* Yields

This chapter will give the analysis strategy, with the reasons for the choices. The next chapter will describe the results from data. The aim is to get the ratio of WS/RS D^*s as a function of proper decay time. With the relatively large number of WS signal events in the CDF sample, we decided to measure the binned ratio of RS to WS events as a function of lifetime. An advantage of this is the detector efficiency versus lifetime will cancel in the ratio.

The measurement for D^0-$\bar{D}^0$ mixing for this analysis will be reconstructing D^* events coming from the primary vertex. The reasons for using those events are:

- The charge of the soft pion coming from the decay $D^* \rightarrow D^0\pi^*$ will tag whether the charm meson started out as a D^0 (for π^+) or a $\bar{D}^0$ (for π^-).

- The charge of the pion from the $D^0 \rightarrow K\pi$ decay, along with the charge of the π^*, determines what type of candidate the event is. "Right-Sign" (RS) events, where both pions have the same charge, are candidates for Cabibbo-Favored charm decays. "Wrong-Sign" (WS) events, where the pions have opposite charge, are candidates for DCS decays and mixing.

- D^*s decay strongly. For D^*s produced at the primary vertex, we can use the distance from the primary vertex to the $K\pi$ secondary vertex to obtain the D^0 decay time.

- Although the Two-Track Trigger was optimized for beauty decays, CDF has a copious amount of $D^0 \rightarrow K\pi$, with a large signal-to-background. Requiring a D^* improves the signal purity.

3.1 General Strategy

The analysis starts by reconstructing the all-hadronic mode $D^{*+} \rightarrow D^0\pi^+$, $D^0 \rightarrow K^-\pi^+$. (Unless otherwise mentioned, all modes include the charge conjugate decay.) Right-Sign (RS) and Wrong-Sign (WS) candidates are handled in separate binned fits. The following fits are done in succession, with the signal (and uncertainty) from the previous fit being used as the distribution for the next fit.

- Number of candidates vs. $K\pi$ mass: The result is the number of D^0s.

- Number of D^0s vs. $(m_{D^*} - m_{D^0} - m_\pi)$ mass difference: The result is the number of D^*s.

- Number of D^*s vs. decay time, for events outside the impact parameter cut: This information will be used to correct for D^*s produced in secondary (beauty) decays, which will not have accurate decay times.

- Number of prompt D^*s vs. proper decay time: The fit for this distribution will provide the estimate of the charm mixing parameters.

The idea of the sequence of fits is to produce a distribution with a distinct narrow signal distribution that can be separated from a non-peaked background p.d.f. While it is possible to do a global unbinned maximum likelihood fit, some of the background distributions have complicated distributions when all variables are used ($K\pi$ mass, mass difference, i.p., decay time). This method requires less knowledge of the background distibutions that are fit early and discarded. For example, non-D^0 background has a complicated mass difference p.d.f., but this does not need to be known since that class of events are not present in our mass difference fit.

All fits are done with the ROOT fit integral option. The $K\pi$ fits use log-likelihood. The other fits use least chi-square fitting. The ROOT fitter will return an error code on about 1% of the $K\pi$ and mass difference fits. In those cases, we use sideband subtraction to get the number of signal events and the uncertainty.

Figure 7 shows the expected mass distribution shapes for the various types of events, which will be discussed in the next sections.

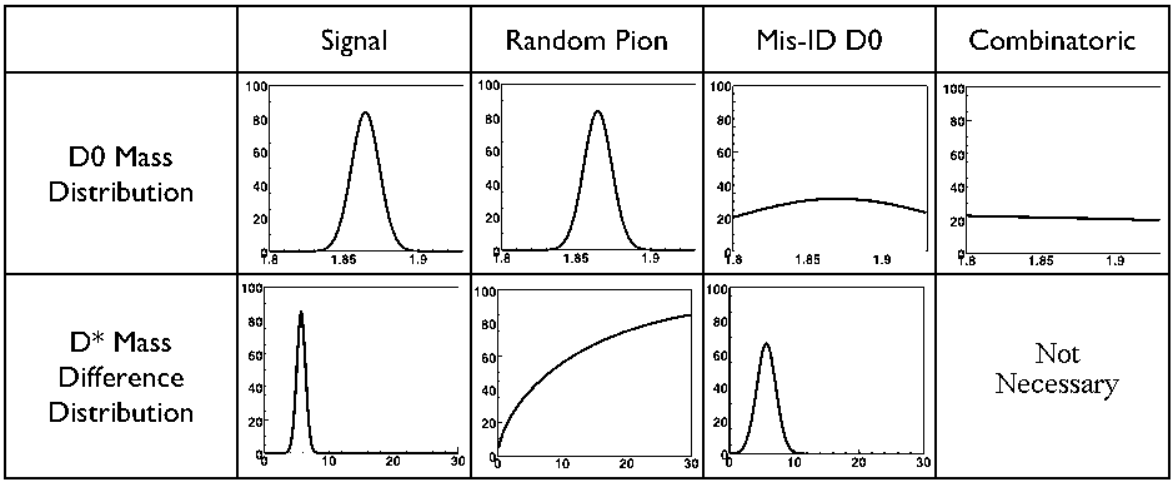


Figure 7: Event distributions for signal and different types of backgrounds as a function of $K\pi$ mass or $(D^* - D^0 - \pi)$ mass difference. The plots are cartoons, not actual data. The D^0 $(K\pi)$ mass distribution has the range from 1.80 to 1.92 GeV. The mass difference distribution has the range from 0 to 30 MeV. Combinatoric background events are not present in our mass difference yield plots, so we have not studied that distribution.

3.2 $K\pi$ Mass Plot

The $D^0 \rightarrow K\pi$ signal distribution will show up as an nearly gaussian peak, with a width on the order of 9 MeV. The actual shape is narrower at the center and wider in the tails, compared to a single gaussian. The shape will be fixed from a fit to a time-summed RS distribution and used for all individual RS and WS $K\pi$ fits. The time-summed $K\pi$ plot has a D^* mass difference cut of $4 < \Delta m < 9$ MeV applied to improve signal purity. The signal function is two gaussians with independent means and widths. The two additional parameters are an overall signal amplitude, and the relative size of the two gaussians with respect to each other.

The true signal shape is more complicated, but we have not found an alternate signal shape (with a reasonably small number of parameters) that works as well as this one. The fit chisquare will get slightly worse as signal statistics increase, but this has already been observed in other CDF analyses.

3.2.1 Combinatoric Background

We need to consider background events in the mass plot, most of which are where one or both tracks do not belong to a $D^0 \rightarrow K\pi$ decay. An earlier study of D^0 mesons by CDF[11] used a linear (decreasing) function to model this background. We used the same function for the previous time-independent measurement, but allowed the parameters to float for each fit, rather than assuming a particular slope. Pure combinatoric background has candidates formed from a random selection of tracks that happen to pass our selection criteria. This is a well understood phenomena in any physics analysis. The data from this $1fb^{-1}$ sample shows that we need to use a quadratic function for the background. A fit for the Wrong-Sign $K\pi$ plot has a chisquare per degree of freedom of 7.79 with a linear background fit, which changes to 1.28 for a quadratic fit (fig. 8).

3.2.2 $D^0 \rightarrow KK, \pi\pi$ Background

The fit range is limited from 1.80 - 1.92 GeV, to exclude background from $D^0 \rightarrow KK, \pi\pi$ decays that are reconstructed as $K\pi$ candidates. Assigning the wrong particle mass to the decay tracks will move the reconstructed mass for those decays out of this search window, as seen in figure 9.

3.2.3 Mis-assigned $D^0 \rightarrow K\pi$ Background

Every $K^-\pi^+$ reconstruction can also have the mass assignments reversed, to generate a $K^+\pi^-$ candidate. A correct mass assignment will result in a narrow peak, the width being determined by detector

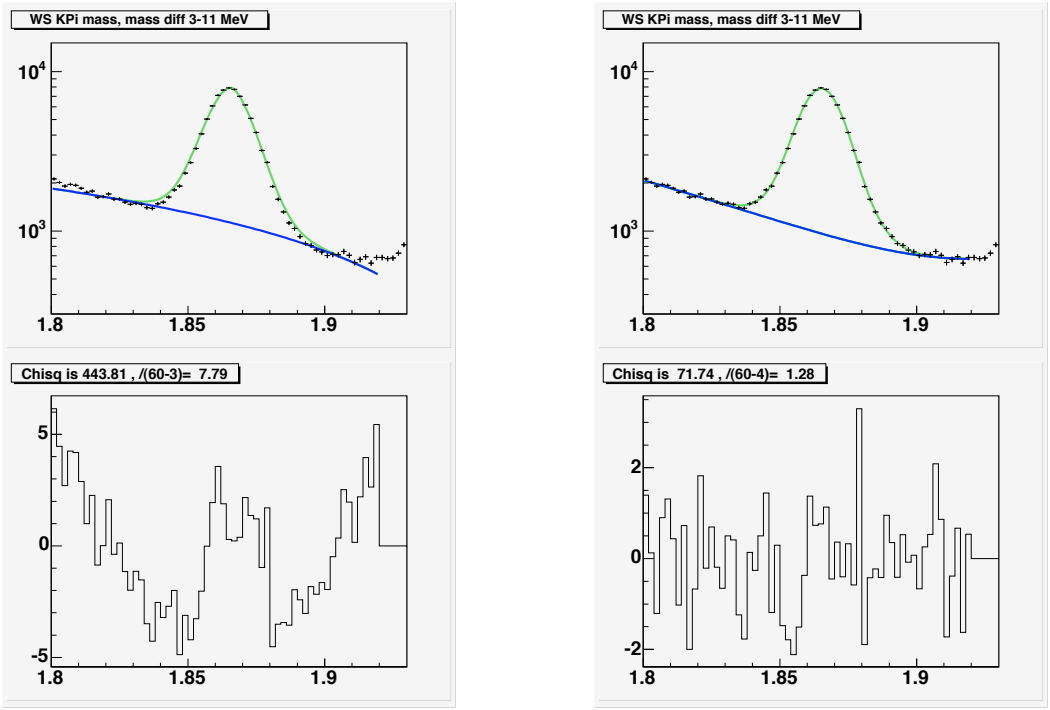


Figure 8: WS candidates from data are plotted as a function of $K\pi$ mass (top plots). The blue line is a fit to the background, the green line is signal plus background. The left plots use a linear background function, the right plots have a quadratic background. The pulls (difference between the data points and the fit, divided by the square root of the number of events in the bin) are in the bottom plots. A D^* mass difference cut of $3 < \Delta m < 11$ MeV is applied to improve signal purity. No selection on decay time is made. The signal shape is fixed from a fit to the RS distribution.

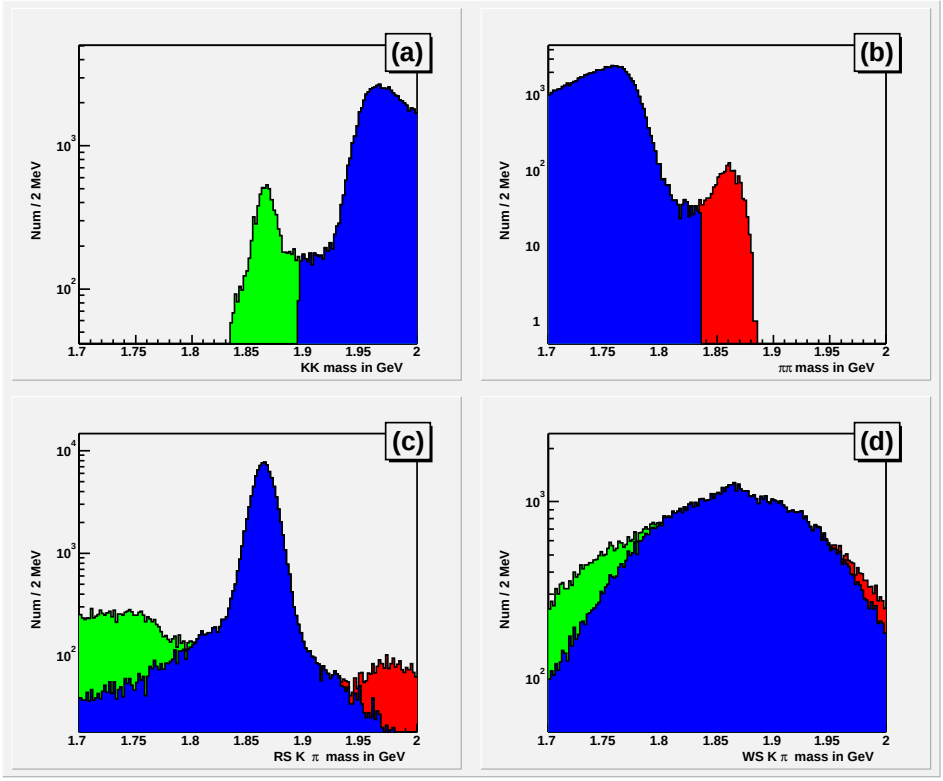


Figure 9: Events from the 0.35 fb-1 data sample have their invariant mass recalculated, assigning the tracks as KK (a) or $\pi\pi$ (b). Events within 30 MeV of the $D^0 \to KK$ mass in (a) are colored green. Events within 30 MeV of the $D^0 \to \pi\pi$ mass in (b) are colored red. All others are colored blue. Those events are then shown as cumulative distributions for RS $K\pi$ candidates (c) and WS $K\pi$ candidates (d).

resolution. (The physical D^0 width is negligible compared to detector resolution.) A mis-assigned reconstruction will result in a broad (gaussian) distribution several times wider than a correct assignment. Since the number of RS D^*s is so much larger than both the general background and the WS D^* signal, we have to be concerned about mis-assigned Cabibbo-Favored D^0s overwhelming the WS signal.

Figure 10 shows candidates plotted with both RS and WS interpretations. The CF D^0 decays dominate the WS plot. Any WS signal would be completely obscured by these candidates. We apply a kinematic cut and a particle identification cut to reduce this background.

The kinematic cut uses the opposite assignment mass. The opposite assignment refers to switching the K and π assignments of the D^0 tracks. When plotting the WS mass, events are excluded if they have a RS mass consistent with being a D^0. In the case of plots showing the RS mass, the cut is applied to the candidate's WS mass. Since the wrongly reconstructed D^0s have a distribution almost 10 times wider than correct D^0s, this will remove most of background while still retaining most of the correctly reconstructed events.

The particle identification cut uses dE/dX information. For CDF II, it refers to the ionization energy loss as particles pass through the COT. (The details of how dE/dX is done are too complicated for this note, but are described in other papers[12, 13].) This analysis uses the dE/dX variable which is expressed as $dE/dX = \log(dedx_{measured}/dedx_{predicted})$. For a correct hypothesis, the $Z = (dE/dX)/\sigma_{dE/dX}$ variable is a unit gaussian centered at zero. For a pion hypothesis, a real kaon will have a Z gaussian distribution centered at -1.6 with a width of 1.25. For a kaon hypothesis, a real pion will be centered at 1.3 with a width of 0.9. With $Z_{K(\pi)}(1)$ being the dE/dX value for the first track with a kaon (pion) hypothesis, and $Z_{\pi(K)}(2)$ being the dE/dX value for the second track with a pion(kaon) hypothesis, we compare $(Z_K^2(1) + Z_\pi^2(2))$ to $(Z_\pi^2(1) + Z_K^2(2))$ and only accept the more probable hypothesis.

The kinematic and part. id. cuts affect the same background, so they are optimized together. With our choice for the part. id. cut, the kinematic cut rejects opposite assignment mass within 20 MeV of the D^0 mass. If we did not use the part id cut, the optimal kinematic cut would be closer to 30 MeV.

These cuts were retained from the time-independent analysis. When more data past 1 fb^{-1} becomes available, a re-optimization may produce (slightly) better results.

The cuts reduce the number of mis-assigned RS D^*s (MRS) in the WS plot to less than 1%, but this number is still comparable to the expected WS signal. We calculate the expected number of mis-assigned events N_{MRS}:

$$N_{MRS} = N_{RS} \frac{f_k(RS \to WS)}{f_k(RS \to RS)} \frac{f_p(RS \to WS)}{f_p(RS \to RS)}, \tag{4}$$

where

- N_{RS} is the number of RS D^*s
- $f_k(RS \to WS) = 3.62 \pm 0.01\%$ is the fraction of RS D^*s that survive the WS kinematic cut
- $f_k(RS \to RS) = 77.41 \pm 0.07\%$ is the fraction of RS D^*s that survive the RS kinematic cut
- $f_p(RS \to RS) = 86.58 \pm 0.08\%$ is the fraction of RS D^*s that survive the RS particle id cut
- $f_p(RS \to WS) = (1 - f_p(RS \to RS))$ is the fraction of RS D^*s that survive the WS particle id cut

The efficiency numbers were determined from fits to RS $K\pi$ distributions with both kinematic and particle id cuts, with only the kinematic cut, and without either cut. The ratio of RS D^0s with and without the cut determine the efficiency for that cut.

The number of MRS events are then included in the WS $K\pi$ fit. The calculated number will be different for each time bin, depending on the RS signal for the same decay time range. The MRS distribution is a gaussian, determined from toy MC (mean 1.866 GeV, width 82 MeV), and the amplitude given by N_{MRS}. The MRS distribution width can be fit from data before kinematic and part id cuts are applied. After the cuts, the background amplitude is small enough to make it difficult for the fitter to distinguish between this broad gaussian background and the general combinatoric background. Toy MC shows that applying the kinematic cut increases the MRS width by several MeV. (The distribution is still Gaussian.)

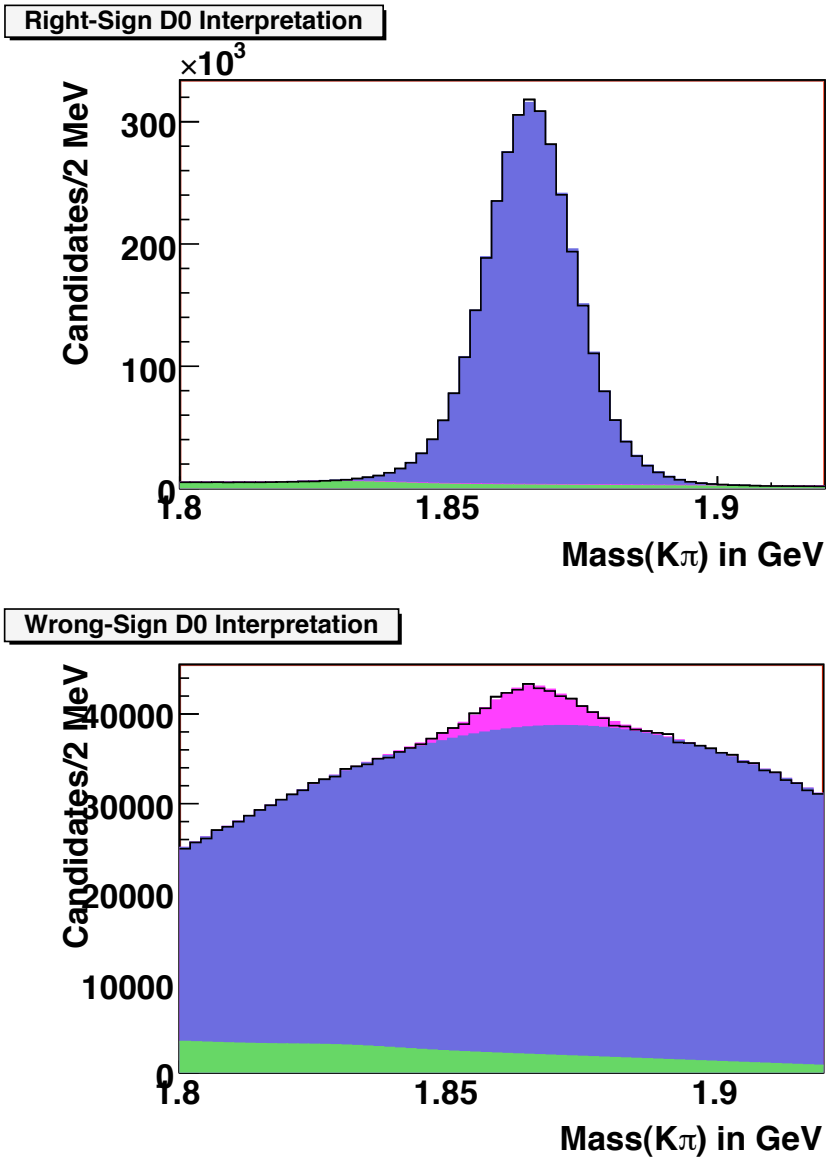


Figure 10: RS (top) and WS (bottom) mass plots for D^0 candidates from data, before the OAM cut is applied. A good RS D^* mass difference was required. The events that are in the top plot are re-interpreted as WS to fill the bottom plot. The blue area (CF D^0s) is the same in both plots, when the under- and over-flow bins are included in the bottom plot. The red areas, from D^0s correctly reconstructed with the WS interpretation, are the same in both plots. The green area is the linear background, which is fit separately for both plots. The narrow signal and the broad gaussian shapes have the same shape parameters for both fits. The D^0s include both real D^*s and D^0s with a fake tag.

3.2.4 RS Background at 1.83 GeV

The RS $K\pi$ fit suggests an additional gaussian background lump is needed at 1.83 GeV. A fit for the Right-Sign $K\pi$ plot has a chisquare per degree of freedom of 24.2 with a quadratic background fit without an additional gaussian, which changes to 1.89 when the lump is added to the fit (fig. 11). The lump in the RS plot is about two orders of magnitude smaller than the signal amplitude; it looks larger when using a log scale. For the analysis, we allow RS fits to have a floating parameter for the lump amplitude, and do not use the lump in WS fits (zero amplitude).

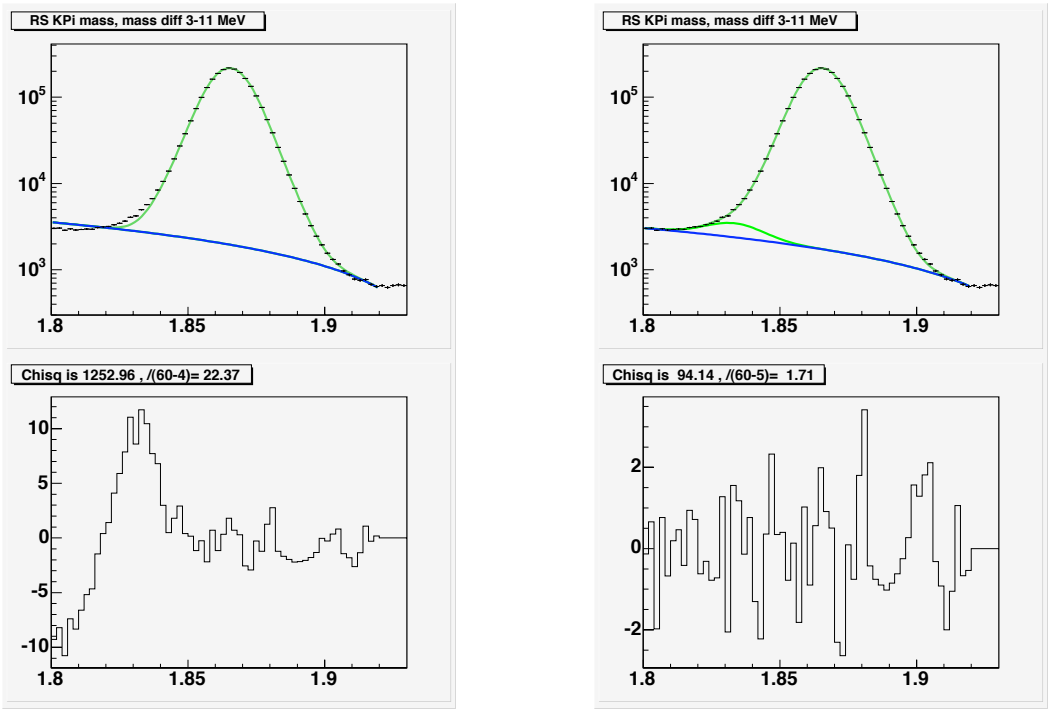


Figure 11: RS candidate $K\pi$ mass, similar to the WS plots in Fig. 8, with a quadratic function for the background. The right plot has an additional Gaussian background lump added to the fit. The lump is at 1.832 GeV (determined from the fit). The titles of the pull plots have the wrong number for the degrees of freedom, since it does not account for the D^0 signal shape parameters, which are determined from these fits. The left plot should be $\chi^2/\text{dof} = 24.6$, the right plot is $\chi^2/\text{dof} = 1.96$.

We tried fitting both RS and WS distributions with the extra lump, as summarized in table 3. The "RS signal" region is almost all CF D^*. Both the RS and WS upper sidebands will be CF D^0 decays with a random pion. The "WS signal" has comparable amounts of signal and random pion background. The WS fits are compatible with zero amplitude for the extra lump, and incompatable (at the 3 sigma level) with the RS lump amplitudes. The WS fit pull distribution in figure 8 does not indicate any deviations at 1.83 GeV. This suggests that the extra lump is not a part of the D^0 signal distribution, as it would be present for both RS and WS plots.

We applied dE/dX track cuts to investigate reflections, as summarized in table 4. An incorrect dE/dX hypothesis moves the Z/σ_Z distribution away from zero. If the extra lump was due to misidentification, we would expect an enhancement in the lump on one side or the other from zero. Within statistics, the lump:D^0 ratio is constant, which suggests that the extra lump is from $K\pi\pi$.

We divided the data by decay time to investigate the lifetime, as summarized in table 5. The lump:D^0 ratio fluctuates, but there is no obvious trend. With the limited statistics, the lump has charm-like lifetime.

A suggestion was made that this is the energy-loss tail of the signal distribution. If that was true, this lump would scale with both the RS and WS (main) signal distribution, which is not what we see from the fits. A possibility is that this comes from incomplete reconstruction of D_s/D^+ mesons, as

Mass Difference Region	D^0 Amplitude	Lump Amplitude	Ratio of Lump to Signal (10^{-3})	Floating Fit χ^2 / dof	Zero Fit χ^2 / dof	0.6% Fit χ^2 / dof
RS - D^* Signal	2804 K	16.4 K	5.86 ± 0.15	103.8 / 55	1357 / 56	103.8 / 56
RS - D^* Sideband	284.6 K	1833	6.44 ± 1.26	71.5 / 55	94.5 / 56	71.6 / 56
WS - D^* Signal	60.5 K	0.	0.	69.7 / 55	69.7 / 56	74.5 / 56
WS - D^* Sideband	264.3 K	629	2.39 ± 1.36	80.5 / 55	81.4 / 56	91.0 / 56

Table 3: $K\pi$ fits for time-summed data done with an extra lump at 1.83 GeV. The D^* signal region are events with a mass difference between 4 and 9 MeV. The signal region will include background from real D^0s plus a random pion. The sideband has a mass difference range from 15-30 MeV. The three fits for the χ^2 performed are with the lump amplitude as a floating parameter, set to zero, and set to 0.6% of the D^0 amplitude.

RS Track	dE/dX Range	D^0 Sig	Lump	Lump/Sig Ratio (10^{-3})
K	$Z/\sigma_Z < 0$	1.57 M	9236	5.88
K	$Z/\sigma_Z > 0$	1.23 M	7136	5.78
π	$Z/\sigma_Z < 0$	1.24 M	7237	5.83
π	$Z/\sigma_Z > 0$	1.56 M	9240	5.91
π^*	$Z/\sigma_Z < 0$	1.34 M	7867	5.86
π^*	$Z/\sigma_Z > 0$	1.46 M	8626	5.90

Table 4: Extra lump study results with dE/dX track cuts. Right-sign events are used. A correct dE/dX hypothesis is used for the tracks given. The π^* is the pion from the D^* decay. Period 10 data is not used.

Decay Range	D^0 Sig	Lump	Lump/Sig Ratio (10^{-3})
2-3	456 K	2572	5.64 ± 0.43
3-4	755 K	5068	6.71 ± 0.34
4-5	698 K	4218	6.05 ± 0.36
5-7	793 K	4622	5.83 ± 0.29
7-10	428 K	2770	6.47 ± 0.42

Table 5: Extra lump study results with events divided by decay time (in D^0 lifetimes).

seen in figure 12. The Belle thesis that described this background had it as WS-only, not RS. Rob Harr has suggested that we check other possibilities, like $\Lambda_c \to p\pi\pi$ with particle misidentification. The dE/dX track study was done in response to this suggestion, and appears to rule out reflections. Stephen Wolbers suggested we look at semi-leptonic decays, like $D^0 \to K\mu\nu$, but any the decay mode would have to explain the differences between RS and WS results.

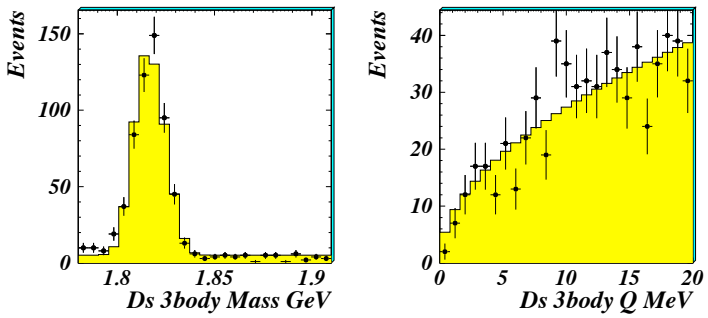


Figure 12: Belle WS background from D_s^+/D^+, taken from figure 4.7 page 53 of the thesis by Li Jin [14]. The events are MC. The description of the $K\pi$ shape mentions that the peak is correlated with the mass difference (event by event).

The extra Gaussian was added purely to improve the χ^2 of the RS $K\pi$ mass fits, and makes an almost imperceptible difference in the number of D^0s. Since this will not affect the results, we will include a systematic uncertainty from the lump (described in section 4.4) . This lump will be investigated more in the long term, but is a low priority for a fast-tracked result.

3.3 D^* Mass Difference Plot

To get the number of D^*s, we use the mass difference variable $\Delta m = M(D^0\pi^*) - M(D^0) - M(\pi^*)$. (The D^0 mass is not constrained.) The yield of D^0s is plotted versus mass difference, with a $K\pi$ fit for each bin of mass difference. The quadratic background parameters for each $K\pi$ fit are floating.

The mass difference yield plot will contain two types of events: D^* signal, and D^0s that pick up a random pion to form a fake D^* candidate. The function used for the signal was obtained by trial and error, by trying to minimize the chi-square of the fit and to avoid features in the residual plots. The signal function is a gaussian plus an additional pseudo-Gaussian asymmetric function. With seven parameters $(A, B, C, D, \mu_1, \mu_2, \sigma)$ and the mass difference Δm, for $\Delta m < \mu_1$ the signal function is:

$$A\left[\exp\left(B(\Delta m - \mu_1)\right) + C\exp\left(-(\Delta m - \mu_2)^2/\sigma^2\right)\right] \tag{5}$$

For $\Delta x > \mu_1$

$$A\left[\exp\left(D(\frac{\mu_1}{\Delta m} - 1)\right) + C\exp\left(-(\Delta m - \mu_2)^2/\sigma^2\right)\right] \tag{6}$$

The background is from D^0s that combine with an unassociated pion to form a (fake) D^* candidate. These random pions could be from the primary vertex, or some other decay that is independent of the D^0. This background has a distribution of the mass difference Δm raised to a power. The background shapes are independent of time, since random pions will not have knowledge of the D^0 decay time.

The signal shape parameters and the RS background power parameter are determined from a fit to the time-summed RS mass difference yield plot, as shown in the next chapter (Fig 18, left plot). The signal shape is fixed for all subsequent RS and WS mass difference fits. The RS power is fixed for all subsequent RS mass difference fits. Then the time-summed WS mass difference yield plot is fit to get the WS background power (Fig 18, right plot). The WS power is fixed for all subsequent WS mass difference fits.

3.4 General Cuts

We apply a set of selection cuts to the ntuple, to improve WS D^*signal to background. The WS uncertainty dominates the RS contribution to the ratio uncertainty. The D^* yield technique from the

previous sections is used on the time-summed data. The number of WS D^*s is kept blinded, so we use the RS D^* signal scaled by $R_D = 0.004$. The variable of merit is the scaled RS signal divided by the WS signal uncertainty (not blinded), which will be referred to as the WS significance.

First, the kinematic (opposite assignment) and particle id (dE/dX) cuts are varied to improve the significance. Those cut values are fixed, and used from that point forward. We examine many other ntuple variables to see if they affect the WS significance. The promising candidates are chosen, and cut values assigned. Two more rounds of optimization are done, where the additional cuts are applied, and the WS significance re-examined to see if the optimal cut values have changed. Figure 13 shows an example optimization plot for the dE/dX information on the π^* (the pion coming from the D^* decay), which was done on the last round of optimization.

Since we were asked to fast-track this analysis, the cuts being used were not optimized with the 1.5fb^{-1} sample. The kinematic and part id cut values were retained from the settings used for the time-independent result (with 0.35 fb^{-1}). The other cuts were optimized early in the analysis, with about half the data. (Some variables that were used for the time-independent result are not available in the BStNtuple, so we had to find viable alternative cuts.) The D^* yield technique has evolved since then, in particular the signal and background functions have changed slightly. In the future, especially as more data becomes available, we will redo the optimization of the cuts. This is not critical, as our current cuts still improve the WS significance.

These are the ntuple cuts in use:

- The opposite assignment mass cut is 20 MeV.
- All three tracks are required to have dE/dX information.
- When looking at WS (RS) events, the particle identification WS (RS) hypothesis must be more consistent than the RS (WS) hypothesis.
- The reconstructed D^0 candidate must have an (unsigned) impact parameter d_0 less than 100 microns.
- The pion coming from the D^* vertex must have d_0 less than 600 microns.
- The pion coming from the D^* vertex must have a point of closest approach to the primary vertex less than 1.5 cm along the Z axis.
- The pion coming from the D^* vertex must have a part. id. Z value less than 2.2.

The transverse impact parameter d_0 is defined as the distance of closest approach, in the transverse plane, of a track (or reconstructed particle) to the primary vertex. When we are looking at a distribution, we used the signed d_0. In all other cases (like the cuts), we use $|d_0|$.

3.5 Decay Time Distribution

We divide the data into time bins, so we can observe any changes in the WS/RS ratio as a function of D^0 decay time. Guided by toy MC with a realistic detector efficiency versus decay time, we have chosen 20 bins:

- 13 bins of a quarter D^0 lifetime width, for decay times from 0.75 - 4.0 D^0 lifetimes
- 4 bins of a half D^0 lifetime width, for decay times from 4.0 - 6.0 D^0 lifetimes
- 2 bins of a D^0 lifetime width, for decay times from 6.0 - 8.0 D^0 lifetimes
- 1 bin of two D^0 lifetimes width, for decay times from 8.0 - 10.0 D^0 lifetimes

The bin sizes are not optimized, but a by-hand attempt to balance sufficient WS statistics with a simple scheme. We have > 20 WS D^*s in the longest decay time bin, which has the smallest number of signal events. Fig. 14 has toy MC simulations of the WS/RS ratio decay time distribution. The ratio in each time bin is the result of fits to the WS and RS D^* yield plots.

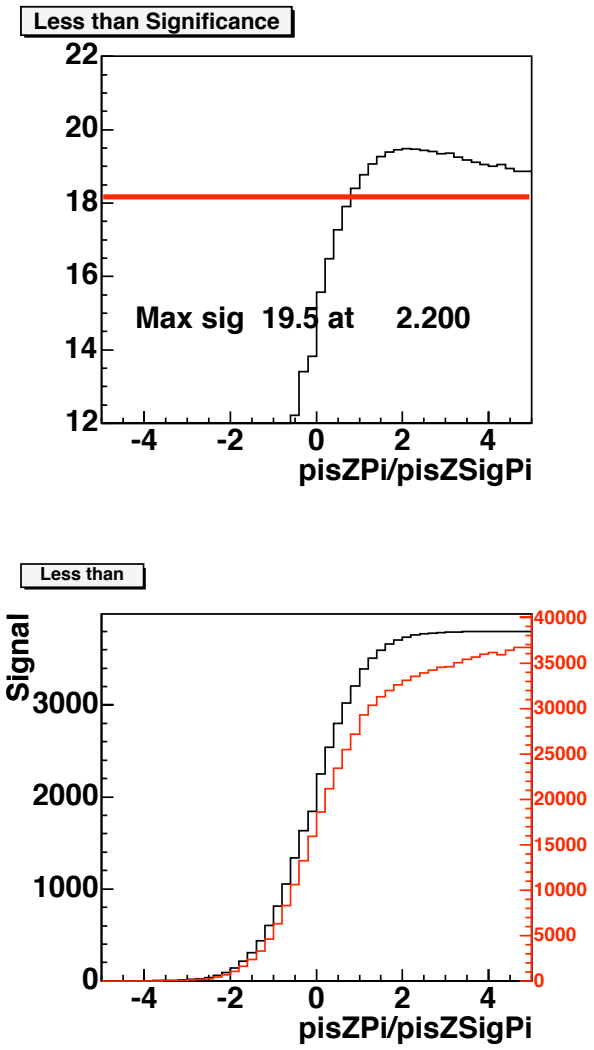


Figure 13: Cut optimization plots checking dE/dX cuts on the π^*. The Z value uses a pion hypothesis. The x-axis is the cut value, requiring events to have that value of Z or less. The top plot y-axis is the (blinded) WS significance. The red line is the significance if we did not use this cut. The bottom plot shows the expected signal (black) and background (red) levels for the cut values.

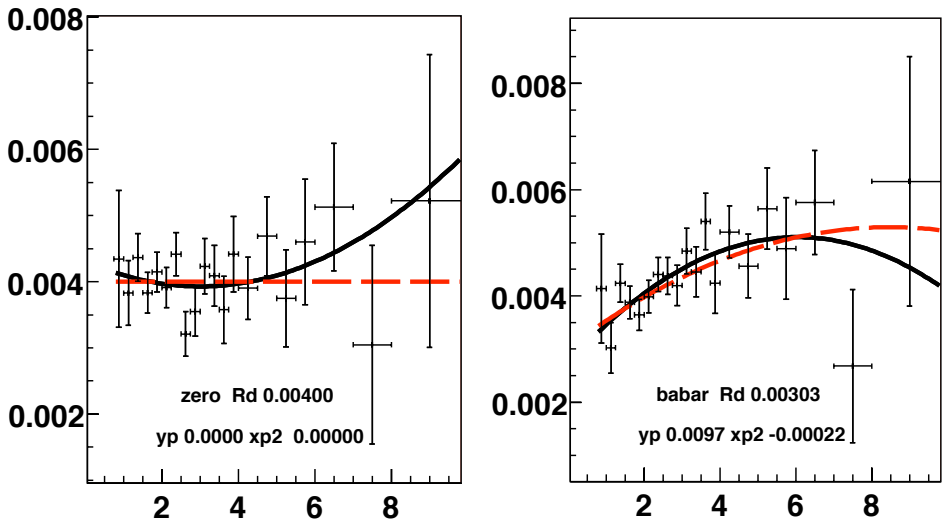


Figure 14: Simulation of the WS/RS ratio versus decay time (in D^0 lifetimes). The black line is the best fit to the simulated data points, while the red line is drawn using the mixing parameters that were put into the toy MC. The left plot was generated with no mixing, the right plot starts with BaBar's best fit (including the non-physical $x'^2 < 0$).

3.6 D^*s From Secondary Decays

The D^0 decay time is measured from the primary vertex. If the D^* is produced at a secondary vertex, like from a $B \to D^* X$, the decay time will be inaccurate. The number of non-prompt (not produced at primary vertex) D^*s turns out to be small compared to the prompt charm. Rather than try to correct the decay time for these events (which would involve considerable effort for little signal gain), we remove the contribution from non-prompt D^*s from our signal. (We are not aware of any sources of non-prompt D^* other than beauty decays)

To get a realistic decay time distribution for the toy MC, we examined the detector efficiency versus decay time for data RS D^*s. At the longest decay times (greater than 8 D^0 lifetimes) the efficiency curve appeared to increase instead of continuing the expected slow decrease. This was the first evidence that we had non-prompt D^*s. This was further validated when Ivan Furic discussed charm coming from B decays at the PSP subgroup meeting [15]. Figure 15 has two plots taken from the presentation. D^0s decaying to two tracks have a narrow d_0 distribution if they are produced at the primary vertex. Non-prompt charm produced from B decays will have a wider d_0 distribution.

3.6.1 Non-prompt D^* Study

The D^* yield technique described previously will include both prompt and non-prompt D^*s. This section describes the method to estimate the non-prompt background that should be subtracted to get the (prompt) RS and WS signal. We studied the distribution of the impact parameter (IP or d_0) vs decay time for D^0 mesons. Figure 16 gives an example of the distribution.

D^* produced at a secondary vertex will have a wider d_0 distribution compared to D^*s produced at primary vertex. The prompt distribution shape is constant for all the time bins. The non-prompt distribution width increases with time. The non-prompt width should converge to the prompt width at short decay times, since a very short lived B decay would be close to the primary vertex. The RS and WS prompt distributions are the same, as the two track kinematics are identical. The non-prompt distributions must be the same for RS and WS. The non-prompt width is wider than the prompt width due to the B decay, which doesn't depend on the charm decay mode.

Our initial study (which is in version 1.2 of this CDF note), used a prompt distribution shape to be the same as the $D^0 \to \mu\mu$ MC sample. The data was divided into 9 RS (decay) time bins and 5 WS time bins. The normal D^* yield technique with 20 time bins would result in some of the IP distributions having too few events to provide reliable conclusions. To help keep the WS/RS ratio blinded, there were no time bins that are the same for both RS and WS. The time-binned IP

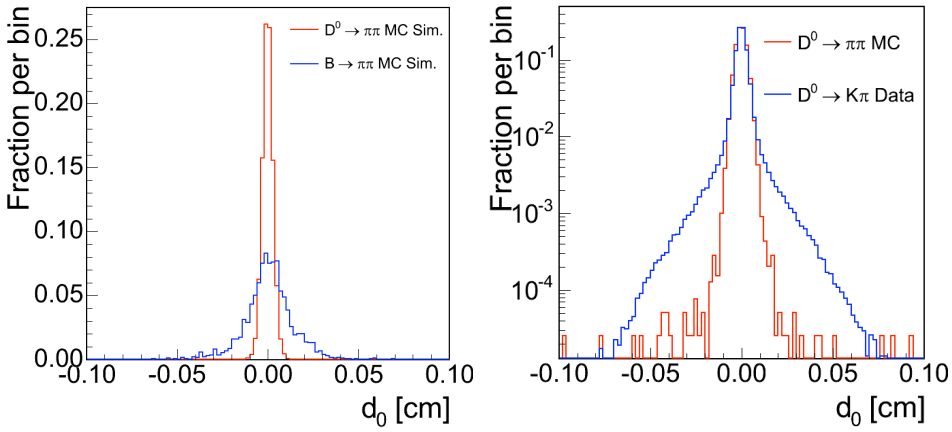


Figure 15: Plots of impact parameter distributions from the $D^0 \to \mu\mu$ analysis, taken from Ivan Furic's PSP meeting slides. The left plot is MC showing a wider d_0 distribution for B decays than D^0 decays. The right plot shows the difference between prompt charm MC and (their) data. They are using $D^0 \to \mu\mu$ MC instead of $K\pi$, but trigger tracks have enough momentum that the change in the d_0 distribution should be small.

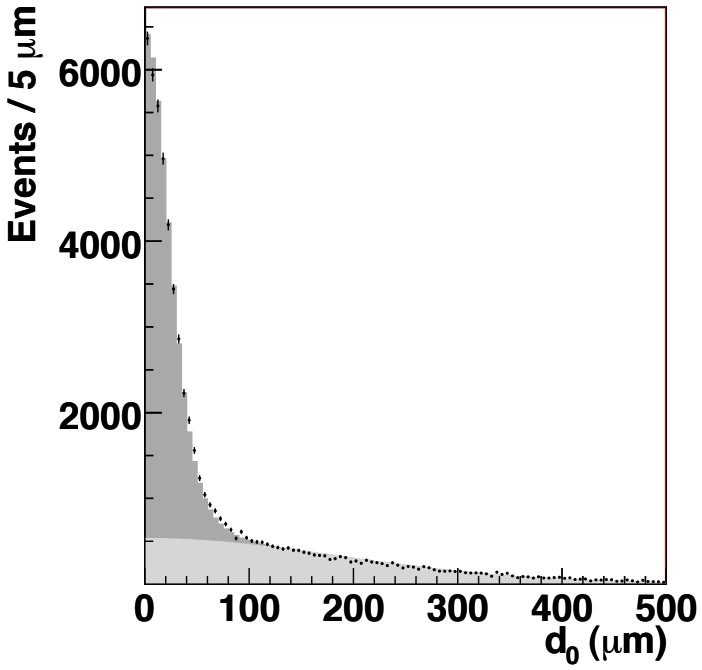


Figure 16: Impact parameter distribution for decay times between 5-6 D^0 lifetimes. The points are data, the light grey is the fit for events from B-decays, and the dark grey is the fit for prompt events.

distributions were fit to measure the non-prompt Gaussian width as a function of decay time. Those values were parameterized to get the non-prompt width at any decay time.

With the information gained from the initial study, we changed the study to use RS data in the same 20 time bins that we use for data. Since the RS signal:background is so large, we are using ($K\pi$) sideband subtraction to get the d_0 signal distribution. The d_0 plot now has 5 micron bins, instead of 50.

Appendix C shows fits for the time-binned impact parameter distributions. 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. 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 ($p_0 + p_1 t + p_2 t^2 + p_3 t^3$) fits the non-prompt Gaussian widths. The first fit parameter (p_0) is fixed to 27.9 μm, 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. (We still need to use the yield technique to get the WS amplitudes.)

3.6.2 Correction For Non-prompt D^*

(This section describes the original correction method. The current method is discussed in the next chapter.)

The data has been processed with an impact parameter cut of 100 μm, to get the WS/RS D^* ratio versus proper decay time. This sample still contains non-prompt D^*s, which will not have the correct measured decay time. We repeat the D^* yield technique on data with events outside the IP cut. With the number of D^*s outside the cut, and the expected non-prompt IP distribution obtained from the preceding section, we can extrapolate the amount of background inside of the IP cut for each time bin. The number of D^*s (within the IP cut) in each time bin is then corrected by:

$$N^{inside}_{non-prompt} = N^{outside}_{non-prompt} \times \frac{\int dx G(x;t)}{1 - \int dx G(x;t)} \tag{7}$$

where x is the impact parameter, and we integrate the gaussian IP distribution for the non-prompt D^*s inside the IP cut, using the width at decay time t from the parameterization described in the previous section. The appropriate amounts of non-prompt D^*s is subtracted for the RS and WS mass difference yield plot time bins. The fit uncertainties (for signal yields inside and outside the IP cut) are added in quadrature.

No further corrections are needed. We could attempt to correct for the detector acceptance versus decay time for the WS and RS distributions, to get individual lifetime distributions. However, that is not our goal, and we will not show the separate RS and WS distributions in public. We will only show the ratio, where the detector acceptance will cancel since it is the same factor for the numerator and the denominator.

4 Analysis of Data

The last chapter gave the analysis strategy. This chapter shows the results of analyzing the 1.5fb^{-1} data. After we have the WS/RS ratio versus decay time R_i in each time bin i, it is fit with a quadratic formula to get the three mixing parameters in Eq. 1 (reprinted here for convenience):

$$R_{\mathcal{B}}(t) = R_D + \left[y'\sqrt{R_D}\right] t + \left[\frac{1}{4}(x'^2 + y'^2)\right] t^2$$

where t is the proper decay time in units of the D^0 lifetime.

4.1 Fixed Shapes

Figures 17-18 shows the fits for the time-summed RS $K\pi$ plot, the RS mass difference, and the WS mass difference. These plots use the data with $|d_0| < 60\mu$m. The signal peak in that RS $K\pi$ plot has 3.0 million D^0s. The signal shape is two gaussians, with means of 1.8652$\pm$0.00001 and 1.8648$\pm$0.00003 GeV (respectively) and widths of 7.16 and 11.63 MeV. The second (wider) gaussian has 36.7% of the signal area. This D^0 signal shape is then fixed for all subsequent $K\pi$ fits.

The D^* signal shape parameters are fixed from the fit to the time-summed RS mass difference yield plot. The plot has 3.033 million $\pm$21 thousand D^*s. The same fit also gives the RS background power term (0.458 ± 0.005), which is fixed for the RS time-binned fits. The D^* shape is then fixed for all subsequent mass difference fits. The time-summed WS mass difference yield plot is fit to get the WS power term (0.469 ± 0.003), which is fixed for the WS time-binned fits. The WS plot has 12768 ± 281 D^*s.

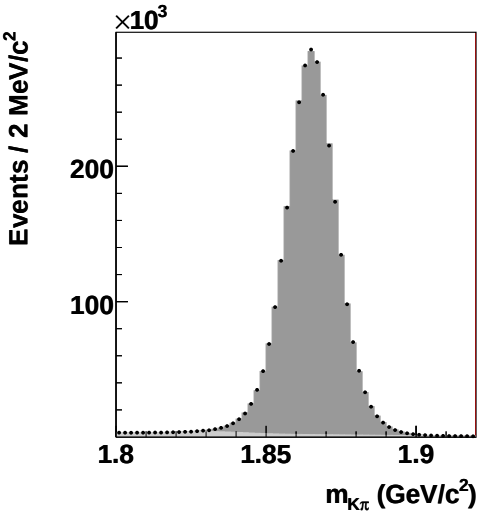


Figure 17: RS $K\pi$ Mass Distribution, with a D^* mass difference cut of $4 < \Delta m < 9$ MeV and $d_0 \leq 60\mu$m applied to improve signal purity.

4.2 RS and WS D^* Yields

After fixing the shapes for the D^0 signal, D^* signal, and the (RS, WS) mass difference backgrounds from fits to the time-summed data, those shapes are used for the fits where the data is divided into time bins. For the RS candidates, we fit 20 mass difference yield plots (one for each time bin), with 60 $K\pi$ fits for each yield plot. Figure **??** has the χ^2/dof distribution for the 1200 RS $K\pi$ fits, with the average value being 1.0. We then fit the 1200 WS $K\pi$ fits. Figure 19 has the χ^2/dof distribution for the 1200 WS $K\pi$ fits, with the average value being 1.0.

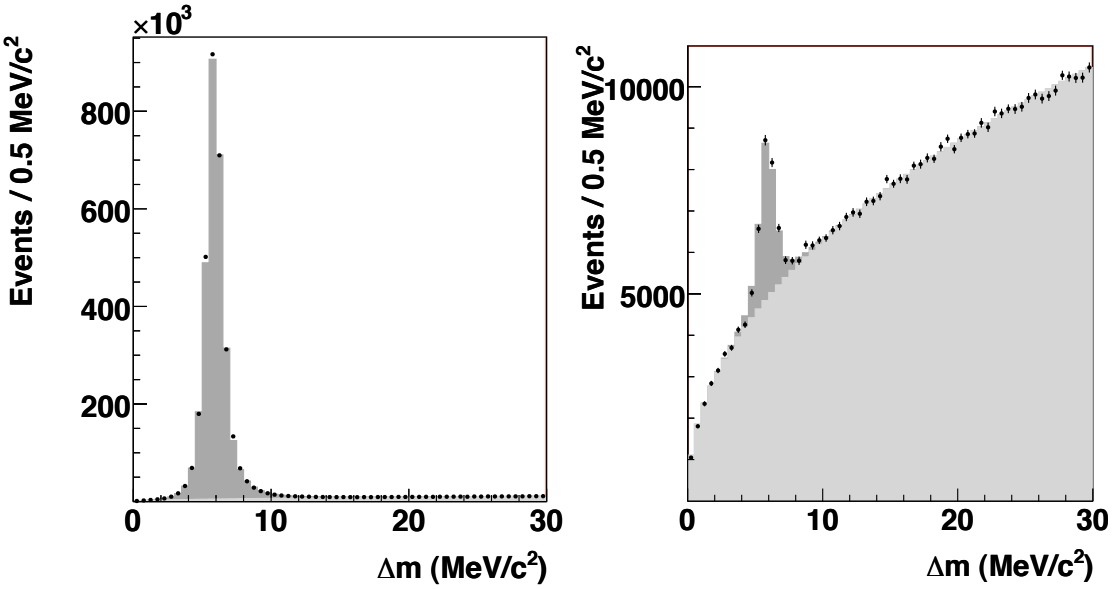


Figure 18: RS and WS mass difference yield plots.

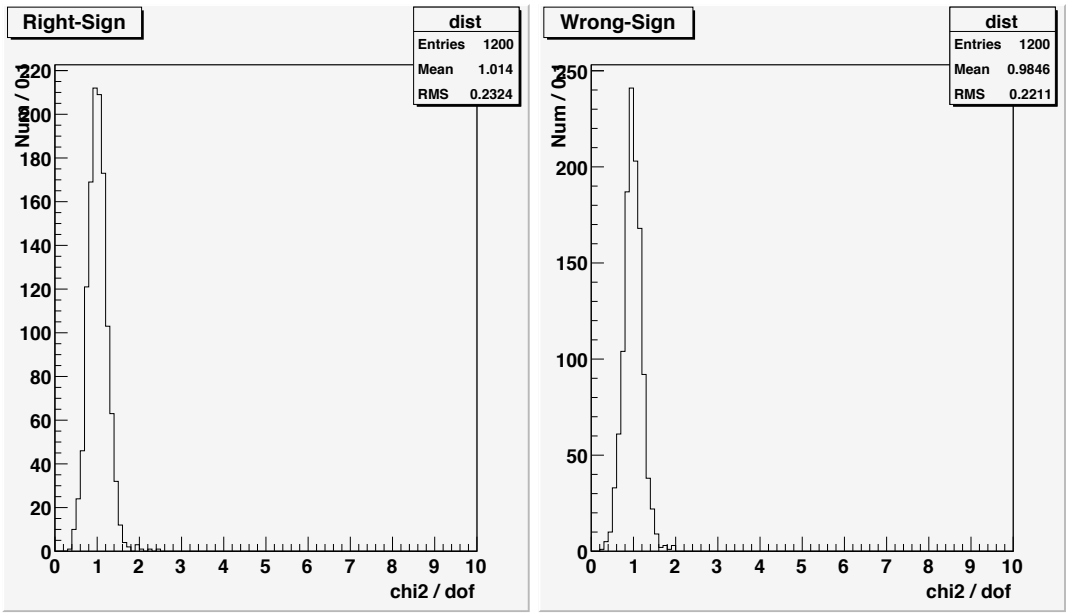


Figure 19: The χ^2/dof for the 1200 RS and 1200 WS $K\pi$ fits.

The twenty RS and twenty WS mass difference plots are fit to get the number of D^*s in each time bin. The distributions, fits, and residuals are in appendix A.

The yield method is repeated with events outside our impact parameter cut of 60 μm. The D^0 signal, D^* signal, and the mass difference background shapes are still fixed to the same values that were obtained from the time-summed fits done with events inside the IP cut.

4.3 Modified B-Decay Correction

During the preparation for the blessing of this analysis, two issues were discovered that led us to modify the method used to account for non-prompt D^*s. The first was due to the WS fits for $|d_0| > 100\mu$m, which had large uncertainties of the same order of magnitude as the uncertainties on the WS prompt fits. This is due to low signal statistics and a higher relative background level outside of the IP cut. The second was raised by Thomas Kuhr, who pointed out that the signal distribution might have tails past the IP cut. Although this fraction is small, it might be comparable to the amount of non-prompt events.

Since we have to account for signal and background in both IP regions, we have more flexibility on where to put the IP cut. The value of 100 μm was chosen by eye to include the prompt distribution. By moving the cut in, we improve the statistics for events for fitting D^* in the outer d_0 region. We used a simple (toy MC) model to predict the statistics for different values of the IP cut. The impact parameter study results from section 3.6.1 are used for the prompt and non-prompt shapes. The model suggested that a value of 60 μm would reduce our mixing parameter uncertainties by 10-15%. We redid the analysis (with the 1.0 fb^{-1}) with the results blinded, but looked at the uncertainties. The uncertainties were reasonably close to the model, and more importantly, better than what we had with the previous IP cut value.

For each time bin i, we use the non-prompt IP Gaussian shape to get the fraction f_i with $|d_0| \leq 60\mu$m, and the fraction g_i with $60 < |d_0| \leq 500\mu$m. We also have the fractions f_p and g_p for the prompt distribution. The yield technique is used to get the RS D^*s for the inner (n_i) and outer (o_i) IP regions. With those numbers, we can get the number of prompt and non-prompt D^*s:

- RS prompt $= (n_i\, g_i \,-\, o_i\, f_i)/(f_p\, g_i \,-\, f_i\, g_p)$
- RS non-prompt $= (o_i\, f_p \,-\, n_i\, g_p)/(f_p\, g_i \,-\, f_i\, g_p)$

The uncertainties on the number of D^*s and the IP fractions are propagated to get the uncertainty on the number of prompt events. A similar calculation is done for the WS D^*s. The numbers used for this procedure are given in more detail in Appendix C. Our ratio uncertainties are still limited by the WS D^* statistics.

4.4 Systematic Uncertainties

This section lists the systematic uncertainties for this result, most of which are already included in the value returned from the ROOT fitter. The CDF Statistics Committee [16] points out that a fit to data includes 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).

4.4.1 Signal Shapes

The true distribution of the D^0 and D^* signal shapes is complicated, but well matched by the simple signal functions we have chosen. The true signal distributions are the same for RS and WS, since they have identical kinematics. 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.

4.4.2 $K\pi$ 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 χ^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.

4.4.3 $K\pi$ Lump at 1.83 GeV

We observe a change in the number of RS events of $\Delta R/R = 0.34\%$ when the extra lump background is fixed to zero (instead of floating for all $K\pi$ fits). A systematic uncertainty of the same amount will be added to the WS D^* yields for each time bin.

4.4.4 Mass Difference

We do assign an additional uncertainty due to fixing the mass difference background power term from the time-summed fit. As a check, we changed the power term by $\pm 1\sigma$, based on 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 caused a negligible change in the ratio, in particular because the signal-background is so high. The WS change in the ratio was on average $\Delta R/R = 0.80\%$. For the fit to the time-summed data, the ratio change was roughly a quarter of the uncertainty returned by the fitter.

4.4.5 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. Our kinematic and particle identification cuts greatly reduce this background, and the $K\pi$ mass distribution for this background only has slight curvature under the signal. We turned off the correction entirely, and observed a change in the WS signal of 105 events for the time-summed fit. This by itself is small compared to the uncertainty on the WS signal. The uncertainty on the correction is about 1% of the correction itself, which is negligible.

4.4.6 Non-prompt D^* Correction

We do not assign any additional uncertainty due to correction for non-prompt D^*s, Our results already include uncertainty from the fits for events outside of the IP cut. The uncertainty on the correction is of the same order as the uncertainty on the number of events inside the IP cut. 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.

4.4.7 Time Resolution

There is no additional systematic uncertainty due to uncertainty in the decay time. CDF has very good time resolution, smaller than the width of the time bins. We used the toy MC (that made the time distributions in Fig. 14), and examined the effect of turning off time-smearing. The change in the ratio with and without time-smearing turns out to be negligible compared to the fitter uncertainty. It is possible that the WS and RS lifetime have a noticeable systematic uncertainty, but any such effect would cancel in the WS/RS ratio.

4.5 WS/RS Prompt D^* Ratio

The corrected WS/RS ratio versus decay time was blinded to help minimize bias as the analysis method was being developed. After the method was preblessed, the unblinded ratio time distribution was made. Table 6 has the corrected WS/RS ratios for the time bins. All systematic uncertainties have been included. Fig. 20 shows fits to data: normal (all parameter space), limited to the physically allowed region ($x'^2 \geq 0$), no-mixing ($y' = x'^2 = 0$), and a normal fit with the first and last two data points dropped (check of fit stability). The normal fit has a correlation of -0.984 for y', x'^2.

Time Bin	WS/RS Ratio
0.75- 1.00	3.733784 ± 1.223251
1.00- 1.25	2.696761 ± 0.494228
1.25- 1.50	4.417567 ± 0.351018
1.50- 1.75	3.584056 ± 0.299417
1.75- 2.00	3.650940 ± 0.279834
2.00- 2.25	4.334846 ± 0.289116
2.25- 2.50	4.117611 ± 0.290144
2.50- 2.75	4.076093 ± 0.340159
2.75- 3.00	4.464356 ± 0.372993
3.00- 3.25	3.963153 ± 0.397021
3.25- 3.50	4.063845 ± 0.454334
3.50- 3.75	5.141623 ± 0.509611
3.75- 4.00	5.826085 ± 0.622706
4.00- 4.50	4.456788 ± 0.490784
4.50- 5.00	5.169453 ± 0.708326
5.00- 5.50	4.663368 ± 0.823720
5.50- 6.00	6.035865 ± 1.121994
6.00- 7.00	5.896054 ± 1.317134
7.00- 8.00	4.343844 ± 2.579420
8.00-10.00	9.356841 ± 5.957180

Table 6: WS/RS ratios for data.

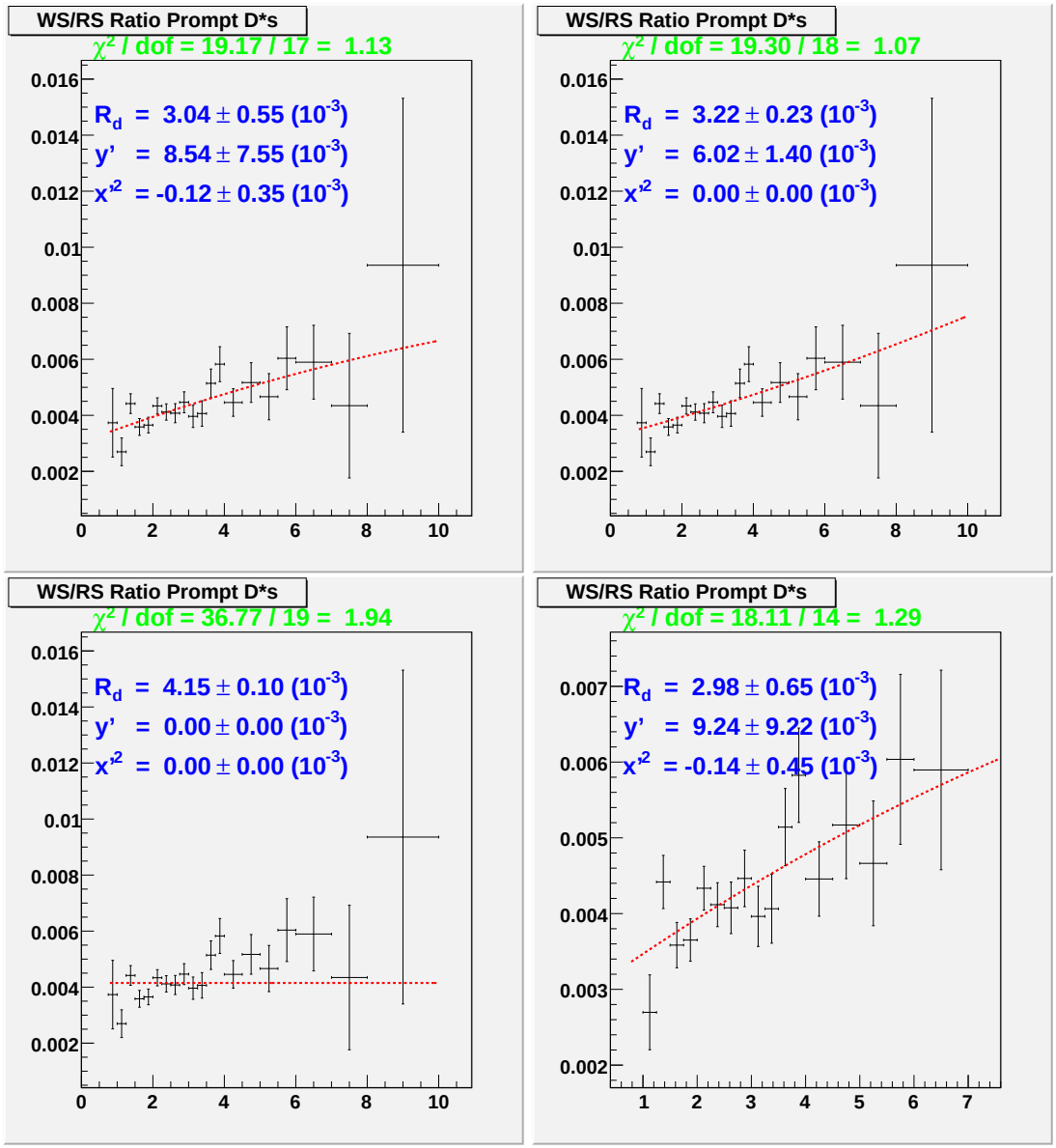


Figure 20: WS/RS ratio versus proper decay time for data. The distribution of RS and WS D^*s were corrected to remove non-prompt D^*s. The red line the lowest chisquare fit for the data. The upper left plot uses a quadratic fit, the upper right plot is restricted to physically allowed region, the lower left plot is a no-mixing fit, and the lower right is the quadratic fit with the first and last two data points removed.

5 Interpretations of the WS/RS Ratio Distribution

We use the decay time distribution of the (prompt) WS/RS ratio to present three results. A fit to the distribution gives the three mixing parameters (R_D, y', x'^2). We construct confidence probability contours in a two-dimensional plot for (y', x'^2). Finally, a limit on how the result excludes the no-mixing point $(y' = x'^2 = 0)$, as a probability or an equivalent number of (gaussian) sigmas.

5.1 Time-Dependent Ratio

The ratio of WS to RS decays as a function of time is given in Eq. 1. Fig. 14 from a previous chapter showed toy MC simulations of the time-binned ratios and the fits. The right-hand fit of that figure, with the MC starting with BaBar's reported mixing parameters, has a fit χ^2 of 14.5 with 17 d.o.f. (20 time bins, 3 mixing parameters).

Table 7 lists the uncertainties in the mixing parameters, based on toy MC simulation. Since the CDF data requires a detached secondary vertex, our measurement of R_D (ratio at zero decay time) is expected to have larger uncertainties than Belle and BaBar.

(The uncertainty study was done without including the correction from non-prompt D^*s. Since this is only to get an estimate on the results with 1 fb^{-1}, it makes more sense to simply wait for the unblinded fit results rather than delay the blessing to redo this study.)

Experiment	$R_D(10^{-3})$	$y'(10^{-3})$	$x'^2(10^{-3})$
Belle	3.64 ± 0.17	0.6 ± 4.0	0.18 ± 0.22
BaBar	3.03 ± 0.16	9.7 ± 4.4	-0.22 ± 0.30
CDF Simulations			
No Mixing	± 0.48	± 4.8	± 0.17
Belle	± 0.48	± 5.1	± 0.18
BaBar	± 0.49	± 6.2	± 0.29

Table 7: Mixing parameters uncertainties with CDF toy MC. The Belle and BaBar values are from their best fits to data, with statistical uncertainties. The CDF values are from toy MC simulation, with MC parameters set to no-mixing ($R_D = 4 \times 10^{-3}, y' = x'^2 = 0$), or the best fit values from Belle and BaBar. The uncertainties are from the fitter, which include systematic uncertainties. This study was done without including the contribution from non-prompt D^*s.

5.2 Probability Contours

Since measurements have only recently excluded the no mixing result, this lifetime study will be aimed at finding the region of the mixing parameters x' and y' with posterior probabilities at a certain value. (This is similar to the frequentist 95% confidence limit.) This section discusses the Bayesian method. A good overview is in the Statistics section of the PDG Particle Physics Booklet [17].

So far, experimental results have constrained the limits for the mixing parameters, with only the most recent results being consistent with no mixing. The 95% confidence limit will be calculated, in addition to the best parameter fit to the data. The standard way to present this information is to make a two dimensional plot of y' versus x'^2 with the 95% c.l. contours.

Each article that discusses statistics uses a slightly different notation. We will set up some terms, to make the discussion of statistics in later sections easier.

The term ϕ will refer to a set of mixing parameters, that we wish to test. The measurable quantities will be the ratio $R_i(\phi)$ of WS to RS signal in each time bin i. The term ϕ_0 refers the "true" value from nature (if this model describes nature). The best fit from data will be $\hat{\phi}$. The uncertainties on the ratio in each time bin are given by σ_i.

If we start with a prediction for the mixing parameters ϕ, then Eq. 1 gives the predicted ratio for any particular time. For each time bin, we compare the prediction $E_i(\phi)$ with R_i to calculate the chi-square for that fit:

$$\chi^2(R_i|\phi) = \sum_i \left[R_i - E_i(\phi)\right]^2 / \sigma_i^2 \tag{8}$$

We will assume that the likelihood is related to the χ^2 by

$$L(R_i|\phi) = exp\left[-\chi^2(R_i|\phi)/2\right]. \quad (9)$$

The two-dimensional plot for (y', x'^2) does not include the third mixing parameter R_D. We treat that as a nuisance parameter, and integrate over a range of R_D values (0.002-0.007). The range of values for R_D are large enough that the $L(R_i|\phi)$ will approach zero well before the R_D limits are reached.

The last terms we will use are to describe areas in parameter space. We will use D to describe the set of possible mixing parameter points. This will include unphysical points, like $x'^2 < 0$. The term D_P is the subset of D that are physically allowed ($D_P \subset D$). We will want to construct $D_{c.l.} \subset D_P$ so that it satisfies the requirements to be a confidence limit on the values of the mixing parameters. This is in contrast to BaBar, where they chose their confidence limits to include the non-physical region ($x'^2 < 0$).

5.2.1 Comments on Bayesian and Frequentist Limits

The CDF Statistics Committee recommendations for limits start with the statement:

> There are many recipes for calculating limits. In many cases it is not a question of one method being correct, and the others wrong. It is crucial, however, to be aware of the properties of one's method; and also to be very explicit in the paper about the technique used.

For our analysis, we will be using a Bayesian approach, instead a frequentist method. The Statistics Committee mention: "Of the methods that we have investigated so far, we find that the Bayesian technique is the most practical method currently available, especially for situations involving nuisance parameters..." [18] The method described in following section is straightforward to implement using the ratio-decay time data points.

The frequentist approach would involve massive computer resources to generate many simulations of data. Unlike a counting experiment with Poisson statistics, we have a continuum of values for the three mixing parameters, each point requiring a sufficient number of pseudo-experiments. A realistic MC would be impractical, due to analysis time and the number of events (a few million RS D^*) needed for each pseudo-experiment.

Even with toy MC for the ratio distributions, there is still the question of validating that the toy MC is sufficiently accurate compared to data. The analysis only uses toy MC explicitly in one place (mass distribution for mis-assigned $K\pi$ events), and the results have a negligible systematic uncertainty for this dependence. The analysis method was deliberately chosen to use the data directly whenever possible, to speed up the analysis and to have the numbers depend on tuning Monte Carlo. A toy MC with the accuracy necessary for the frequentist approach is additional work beyond what is needed to get the ratios. (A toy MC to test the statistics method, with some choice of signals and background, is easier to make than a toy MC meant to perfectly reflect data.)

5.2.2 Bayesian Intervals

To get the credible interval for a certain probability level α, we want to get the posterior p.d.f. $p(\phi|R_i)$ which is the degree of belief for the mixing parameters (ϕ) to take on values in a certain region, given the data R_i. This can be calculated from the likelihood distribution $L(R_i|\phi)$, using Bayes' theorem. The posterior p.d.f is

$$p(\phi|R_i) = L(R_i|\phi)\,\pi(\phi)/m(\phi). \quad (10)$$

$\pi(\phi)$ is the prior p.d.f for the mixing parameters. For our situation, we will assume that the p.d.f is a constant value for the region of parameter space that we are looking at (which is the same as figure 21), and zero outside. The prior is flat for (R_d, y', x'^2). We could consider using the best mixing results from Belle or BaBar, but besides potentially biasing our credible interval, it is not clear that their parameters are consistent with each other. For non-physical values (like x'^2 less than zero) or values far from expected (like $x', y' >$ a few per cent), $\pi(\phi)$ will be zero. The marginal distribution $m(\phi)$ ensures that the probability is normalized, with $m(\phi) = \int L(R_i|\phi')\pi(\phi')d\phi'$. The likelihood drops off rapidly as prediction (ϕ) moves away from the best fit mixing ($\hat{\phi}$), so we do not need the computer to calculate likelihoods that are far away from the expected mixing values.

The mixing parameter R_d and systematic uncertainties are treated as nuisance parameters. We remove the posterior dependence on those parameters by integrating over the nuisance parameters.

Next, we want to find a region of D of possible mixing parameter values which will contain α of the probability:

$$\alpha = \int_D p(\phi|R_i)d\phi \tag{11}$$

Since our likelihood distribution is a grid of possible ϕ values instead of a continuum, we will be accumulating a set of points into D until the sum of the probabilities equals α. The order of which points belong to D is somewhat arbitrary. A convenient choice is the highest posterior density (HPD) interval, where the probability $p(\phi|R_i)$ for every point within the interval D is higher than every point outside the interval.

Figure 21 shows the α credible intervals for two simulations.

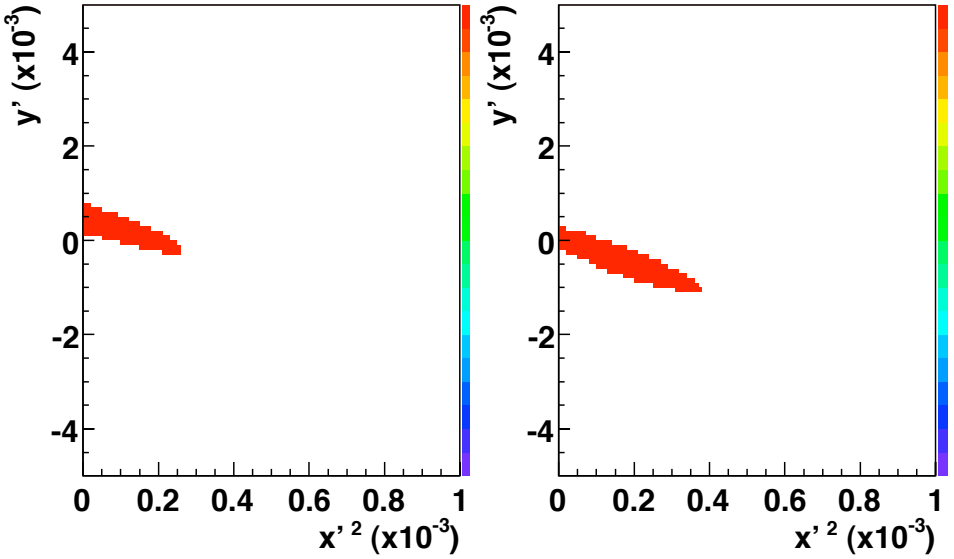


Figure 21: The Bayesian 95% credible intervals for two toy MC simulations of the CDF WS/RS ratio versus decay time. The left plot uses the BaBar best fit result for the mixing parameters, including the nonphysical value for x'^2. The right plot uses no mixing ($y' = x'^2 = 0$).

5.2.3 Check for Prior Cut-Off

For the $1fb^{-1}$ data sample, our prior is set to zero for $|y'| > 5\%$ or $x'^2 > 0.1\%$. This represents the size of the grid of values we are using in ROOT when computing the probabilities. The Statistics Committee pointed out the potential danger if the limits were sensitive to the prior cut-off. We doubled the limits (10% and 0.2%), doubled the number of bins in each axis (to keep the bin sizes constant), and repeated the procedure to get the probabilities. There was no difference in the probabilities. As a check, we reduced the limits to 4% and 0.08%, and started to see small changes in the probabilities, which is what we would expect if our cut-off limits were too small. Our current limits appear to be safe.

5.2.4 Check for Prior Distribution

We were asked to try a different (reasonable) prior distribution, to see how strongly it would affect the distribution of the probabilities. Figure 22 shows the probabilities for a prior flat in x'^2 and a prior flat in x'. This was the simplest alternate prior we could try. Although there are small changes in the probability values (since it is a different prior), qualitatively the contours have the same shape (once the change in the x-axis is taken into account) . The no-mixing probability changes from 0.11% (3.26 sigma) to 0.05% (3.48 sigma), so as long as we are not too close to 0.27% (3.0 sigma), the choice of prior will not affect our result.

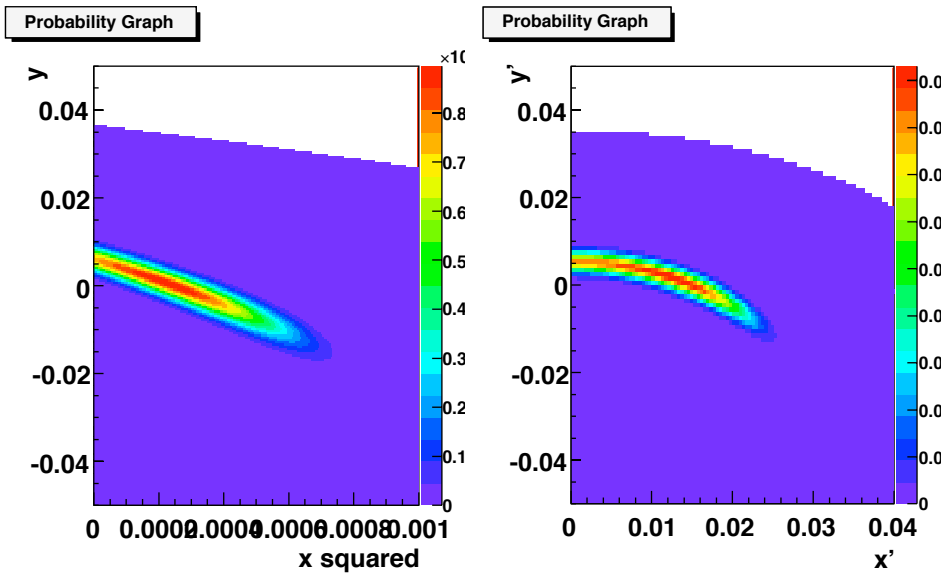


Figure 22: The Bayesian probabilities using a prior flat in x'^2 (left) and flat in x' (right). Both plots use data for $1fb^{-1}$. The x-axis of the plots are different.

We also see from toy MC with different starting mixing parameters, that the contours remain centered on the best fit for the mixing parameters. The choice of prior is not noticeably influencing the location of the contours.

On the Statistics Committee FAQ web page, they mention:

> "If you do not like highest posterior density intervals because of their lack of invariance under reparametrization, use likelihood intervals instead. These are intervals with a given credibility content and such that the likelihood of any point inside the interval is larger than the likelihood of any point outside it. Likelihood intervals have been shown to be robust against variations in the prior, see L.A. Wasserman, "A robust Bayesian interpretation of likelihood regions," Ann. Statist. 17, 1387 (1989)."

Contours using a difference in log-likelihoods have the same qualitative shape and size as what we see using our flat prior.

5.2.5 Check for Coverage

We were asked to test the coverage of our contours (a frequentist check). We used a toy MC for the $1fb^{-1}$ data set, starting with $(y' = x'^2 = 0)$. Figure 23 shows the comparison between the probabilities computed from a single simulation (using the Bayesian method described above), and the distribution of the (best fit) measured mixing parameters from 250K simulations that started with no-mixing. The chosen single simulation had the best fit parameters closest to the MC starting parameters. The MC set distribution and the (single sim) Bayesian probabilities are qualitatively similar.

Table 8 shows the coverage for this set of simulations. The Bayesian probabilities are close the coverage. While it is (MC) statistics limited, the trend is for the contour using the entire parameter space to overcover, while the contour for the physically allowed region uncovers.

Between the first and second paper drafts, we will make a higher statistics toy MC sample, tuned for the $1.5fb^{-1}$ data set. The recent improvement in the ratio uncertainties, as described in section 4.3, requires more adjustment of the toy MC. Once the new (no-mixing) MC sample is ready, we will recheck the coverage, and calculate a frequentist p-value. This timetable is similar to what BaBar did, as they published contours (using the difference of log-likelihoods), and calculated a p-value afterwards.

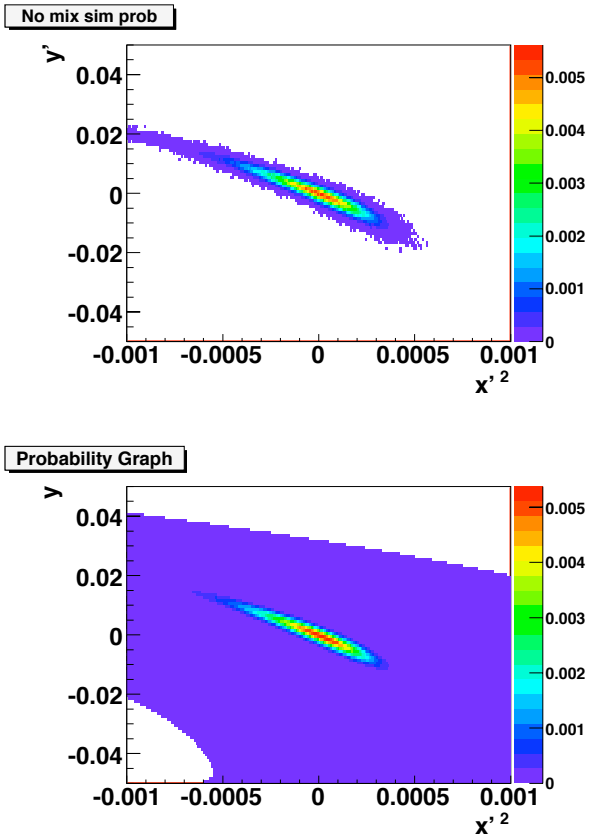


Figure 23: The probabilites from the toy MC starting from no-mixing. The top plot is the distribution of the best fit mixing points from 250K simulations. The bottom plot is from the one simulation closest to the original no-mixing point, with the probabilities computed from the same method we use for data.

Standard Deviations	Bayesian Prob.	Physical Region	All x'^2
0.5	38.293 %	38.0 %	38.8 %
1.0	68.269 %	67.6 %	68.9 %
1.5	86.639 %	86.1 %	87.4 %
2.0	95.450 %	95.2 %	95.9 %

Table 8: Coverage of Bayesian contours for $1fb^{-1}$ toy MC. 250K simulations were made, generated with the no-mixing point. The signals and uncertainties were (only) approximately tuned for the $1fb^{-1}$ data set. The first two columns are the expected fraction of events α. The last two columns are the fraction of the simulations that were found in each of the contours. The physical region excludes $x'^2 < 0$, while the "all x'^2" results includes the unphysical region.

5.3 No-mixing Posterior Probability

We will compare our results to the no-mixing hypothesis. The Statistics Committee in their FAQ: "It is nevertheless possible to use a Bayesian posterior distribution to calculate quantities that can be interpreted as evidence against a given hypothesis... It is also possible to define a Bayesian significance level in this case, as one minus the largest alpha for which theta0 is not inside the interval."

For this analysis, the point θ_0 is the no-mixing point ($y' = x'^2 = 0$). We build up the largest Bayesian probability α contour that excludes the mixing point, and have the no-mixing probability as (1- α). This probability is converted to an equivalent number of standard deviations (for a Gaussian pdf). Figure 24 shows a simulation with 99.9715% credible interval, which is the highest probability that excludes the no-mixing point. This probability is equivalent to the area outside 3.63 standard deviations for a unit Gaussian.

As a cross-check, we can use the difference in log-likelihoods between the best fit point and the no-mixing point. The probability is computed assuming that the difference has a χ^2 distribution for (3-1 =) 2 parameters.

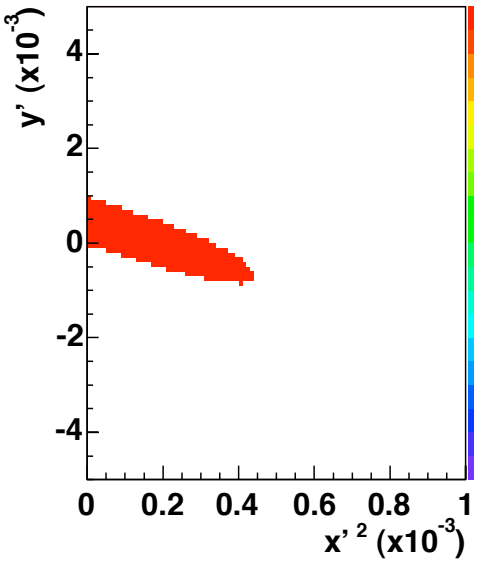


Figure 24: Highest probability bayesian credible interval that excludes the no-mixing point, for a simulation using the BaBar best fit parameters and the number of D^* events for 1 fb^{-1} . This corresponds to a probability of 99.9715%.

5.4 Systematic Uncertainties

Eq. 8 uses the fit uncertainties on the ratio, which already includes some systematic uncertainties along with the statistical uncertainties. After we determine what additional systematic uncertainties there are beyond what is included in the fitter results, we will want to include them in the fit and the credible interval. We can make a change in the analysis method, and redo the fit to determine the systematic uncertainty. The difference in the fit results with the systematic change, compared to the fit with the standard method, would be used to determine the systematic uncertainty.

The credible interval would also include systematic uncertainties, which would be treated as nuisance parameters. When the systematic effect on the measured ratio for each time bin is known, the likelihood would then become:

$$L(R_i|\phi) \sim \prod_i^n \int_j ds_j P(s_j) \left[\exp(-\frac{1}{2} \left[R_i(\theta) + \Delta R_i(s_j) - E_i(\phi) \right]^2 / \sigma_i^2) \right] \tag{12}$$

For this equation, s_j is the variable describing the systematic uncertainty j. ΔR_i difference in the measured ratio from the standard result in the i time bin due to the systematic uncertainty j. $P(s_j)$

is the probability for the systematic uncertainty to have a particular value, which we will assume has a gaussian distribution.

There were two largest systematic uncertainties that were not part of the fitter are the RS "extra lump" (if it is present in the WS events) and the WS mass difference. Although the systematic uncertainty for the WS mass difference is correlated for the different time bins, the effect is small enough that it is simpler (computationally) to treat them as uncorrelated. For each time bin, we add $\Delta R_i/R_i = \pm 0.80\%$ (mass difference) and $\Delta R_i/R_i = \pm 0.34\%$ ("extra lump") in quadrature to the uncertainty returned from the fitter. The change in the uncertainty is almost unnoticeable.

5.5 Unblinded Results From Data

Figure 25 has the RS and WS D^* distributions as a function of decay time. The mixing parameters from the fits are summarized in table 9. The largest contour for the physically allowed region that excludes the no-mixing point has $(1-\alpha) = 1.47 \times 10^{-4}$, which is equivalent to 3.8 sigma. The contour for the entire parameter space has $(1-\alpha) = 1.15 \times 10^{-4}$, which is equivalent to 3.8 sigma. Figure 26 has the Bayesian contours for data. As a cross-check, the difference in log-likelihood between the no-mixing point and the best fit in the physical region is 17.47, which for a χ^2 distribution with 2 d.o.f. has a probability of 1.61×10^{-4}, or 3.8 sigma. (The equivalent number of standard deviations are different if we go to 2 decimal places.)

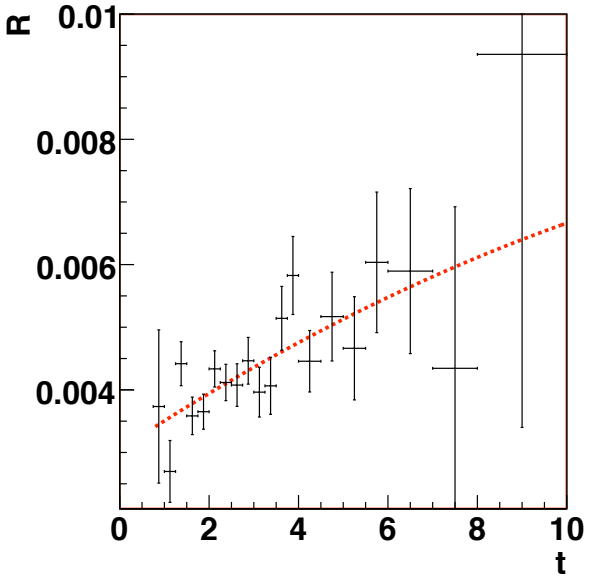


Figure 25: WS/RS Ratio Versus Proper Decay Time Distributions For Data. The distribution of RS and WS D^*s were corrected to remove non-prompt D^*s. The red line is a quadratic fit to the data points, with the mixing parameters $R_d = 3.04 \times 10^{-3}, y' = 8.54 \times 10^{-3}, and x'^2 = -0.12 \times 10^{-3}$.

Experiment	$R_D(10^{-3})$	$y'(10^{-3})$	$x'^2(10^{-3})$	Fit χ^2
Best fit	3.04 ± 0.55	8.54 ± 7.55	-0.11 ± 0.35	19.17
Physical	3.22 ± 0.23	6.02± 1.4	0	19.30
No-Mix	4.15 ± 0.10	0	0	36.77

Table 9: Best mixing parameter fits for $1.5 fb^{-1}$.

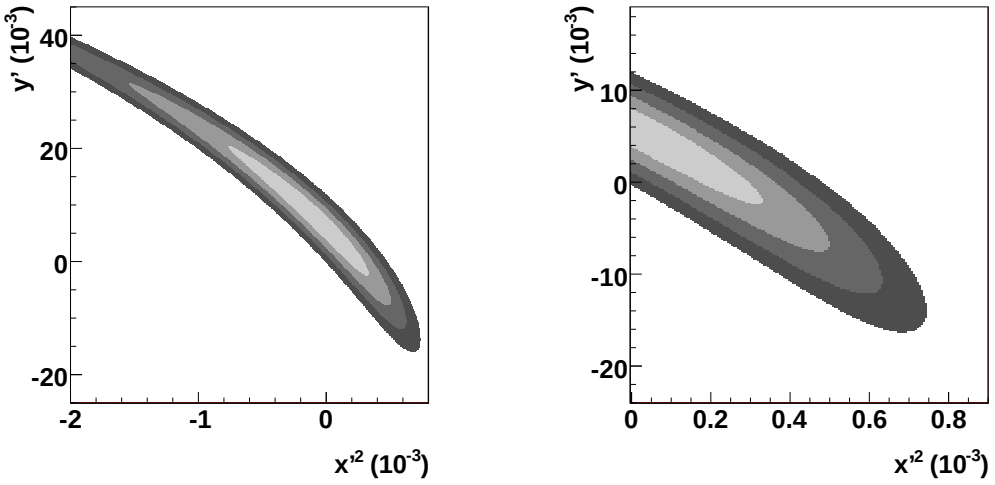


Figure 26: Bayesian contours from data, the posterior probabilities equivalent to 1, 2, 3, and 4 Gaussian sigmas. The left plot has the $y' - x'^2$ parameter space, the right plot is restricted to the physically allowed region ($x'^2 \geq 0$).

5.6 Coverage and P-Value

As a sanity check, we obtained the (frequentist) coverage and p-value for our Bayesian contours. Our toy MC would require significant computing time to get these results (let alone a full MC), so we are following a suggestion from Ivan Furic. The measurements of each time bin ratio are uncorrelated, and the uncertainties include all known statistical and systematic sources. We use a simple model, where the ratios are set to a specific value for each time-bin, then each time bin value is moved randomly using a Gaussian distribution with a width equal to the error bar for that time bin. Each pseudo-experiment is an independent set of these randomized ratios.

To get the coverage, the mean values of the time-bin ratios are set using the mixing equation with our best fit parameters. We get a best fit for each of the pseudo-experiments, and add a point to a (x'^2, y') histogram. We take the Bayesian contours made from data, and compare how many of the pseudo-experiments ended up in each of the contours. The results are in table 10. The Bayesian contours have reasonable frequentist coverage.

Standard Deviations	Bayesian Prob.	Physical Region	All x'^2
1.0	68.269 %	67.529 %	69.770 %
2.0	95.450 %	95.229 %	95.878 %
3.0	99.730 %	99.711 %	99.741 %
4.0	99.994 %	99.992 %	99.994 %

Table 10: Coverage of Bayesian contours for $1.5 fb^{-1}$ with the simple model. 5M simulations were made, starting with the best fit parameters. The first two columns are the expected fraction of events α. The last two columns are the fraction of the simulations that were found in each of the contours. The physical region excludes $x'^2 < 0$, while the "all x'^2" results includes the unphysical region.

To get the p-value using the simple model, the mean values of the time-bin ratios are set to the best no-mixing fit value of 4.15×10^{-3}. For each pseudo-experiment, we make fits for no-mixing and the best (physically allowed) fit. The χ^2 difference for the two fits is compared to the difference obtained from data (17.47 units). The fraction of pseudo-experiments with the same χ^2 difference as

data or greater determines the p-value. The results are shown in figure 27. The p-value is equivalent in significance to 3.8 standard deviations, which is the same as Bayesian contours.

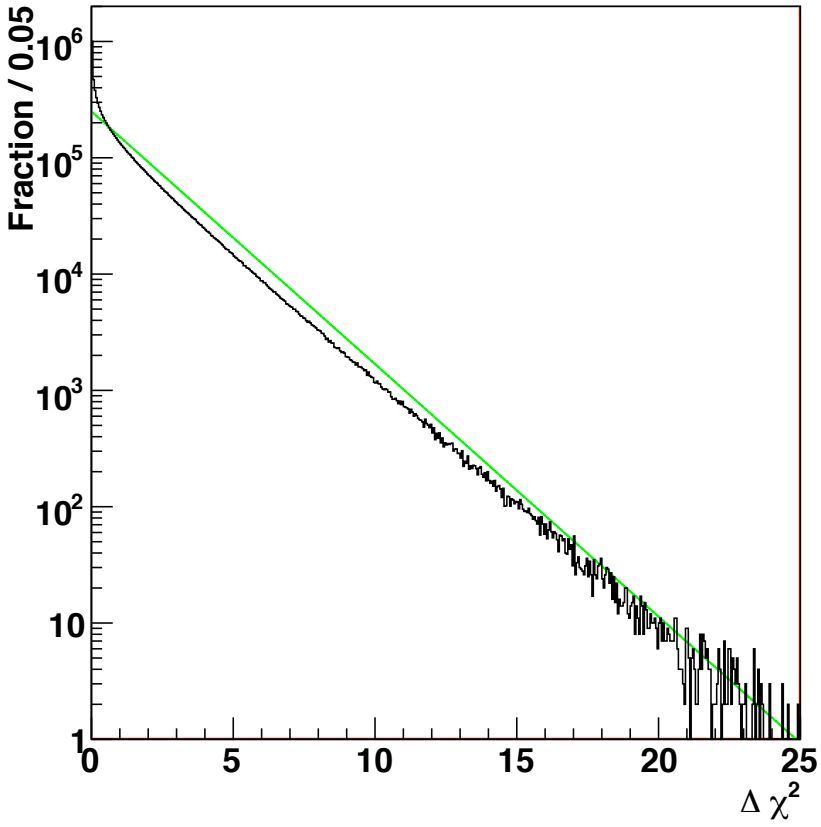


Figure 27: P-value histogram using the simple model. 1312 out of 10 million pseudo-experiments had $\Delta\chi^2 >= 17.47$ units, for a p-value of 1.31×10^{-4}. The green line is a χ^2 distribution for 2 degrees of freedom. The pile-up of pseudo-experiments near zero is due to restricting the"best-fit" to be in the physically allowed region ($x'^2 \geq 0$).

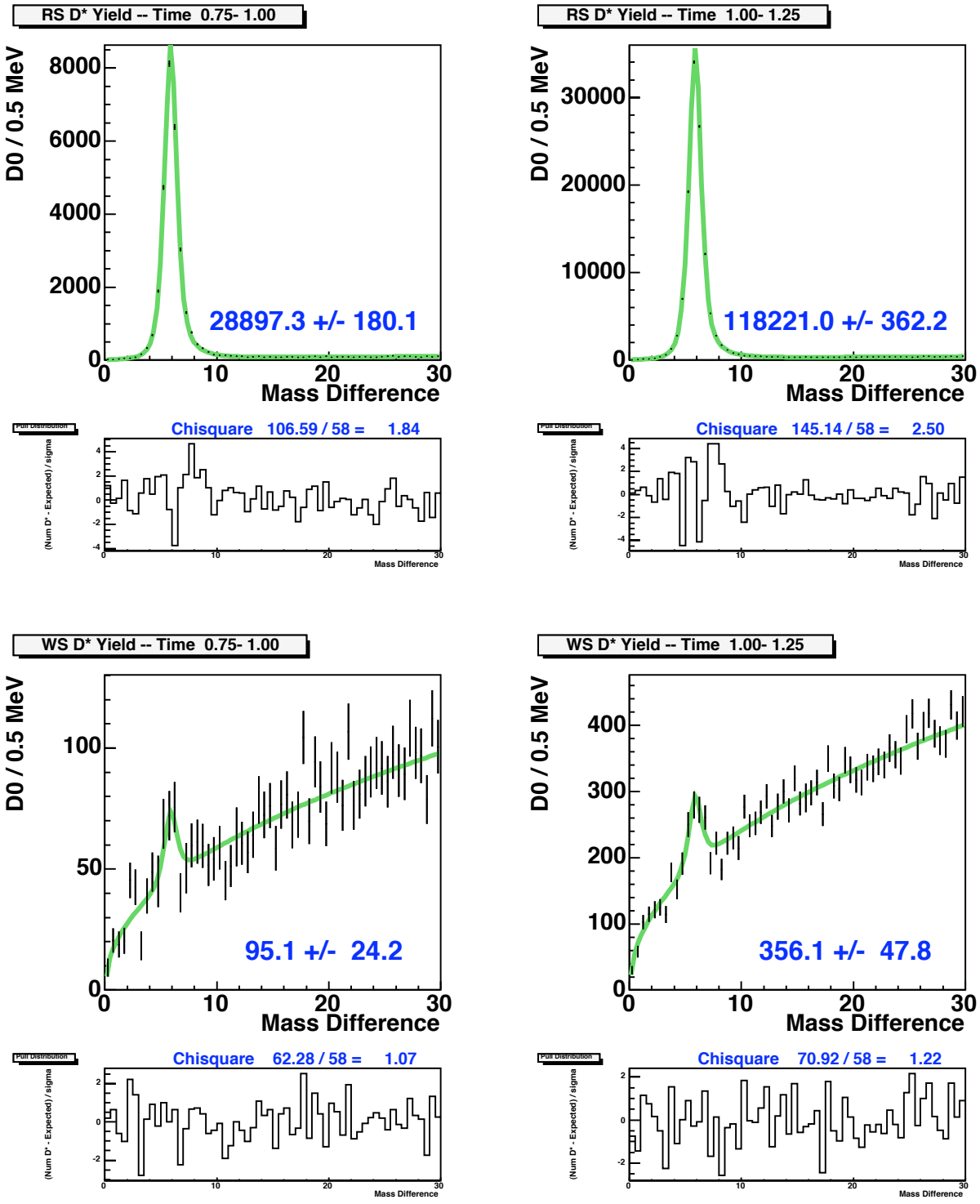
Appendices

A Mass Difference Fits

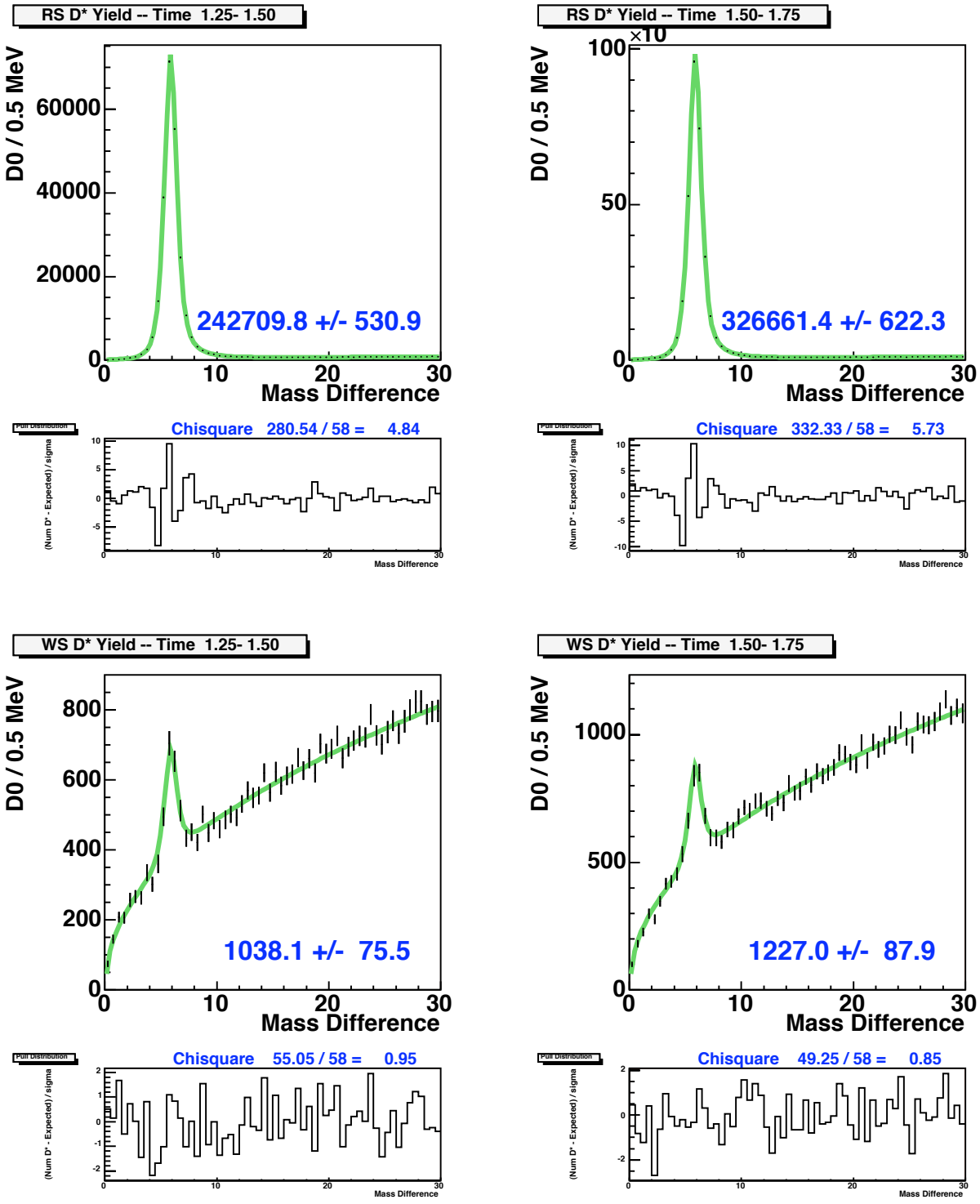
The following pages show the fits for the mass difference plots for events $|d_0| \leq 60$ microns. For each plot, the x-axis is the mass difference in MeV. The y-axis is the number of D^0s per 0.5 MeV wide bin. The title shows the decay time range, in terms of the D^0 lifetime. The black points (with error bars) are the D^0 yields and uncertainties from the fits to the $K\pi$ mass plots. The green line is the fit. The blue number is the signal with uncertainty. Below the distribution is the pulls for the data compared to the fit, along with the fit χ^2. The top row on each page is for Right-Sign events, the bottom row is for Wrong-Sign events. There are 20 time bins.

The fit uses the ROOT "chi-square" and integral options. Each fit only has two free parameters: the (signal) amplitude for D^*s, and the (background) amplitude for D^0 + fake tagging pion. The signal and background shapes are fixed beforehand from a fit to the time-summed data.

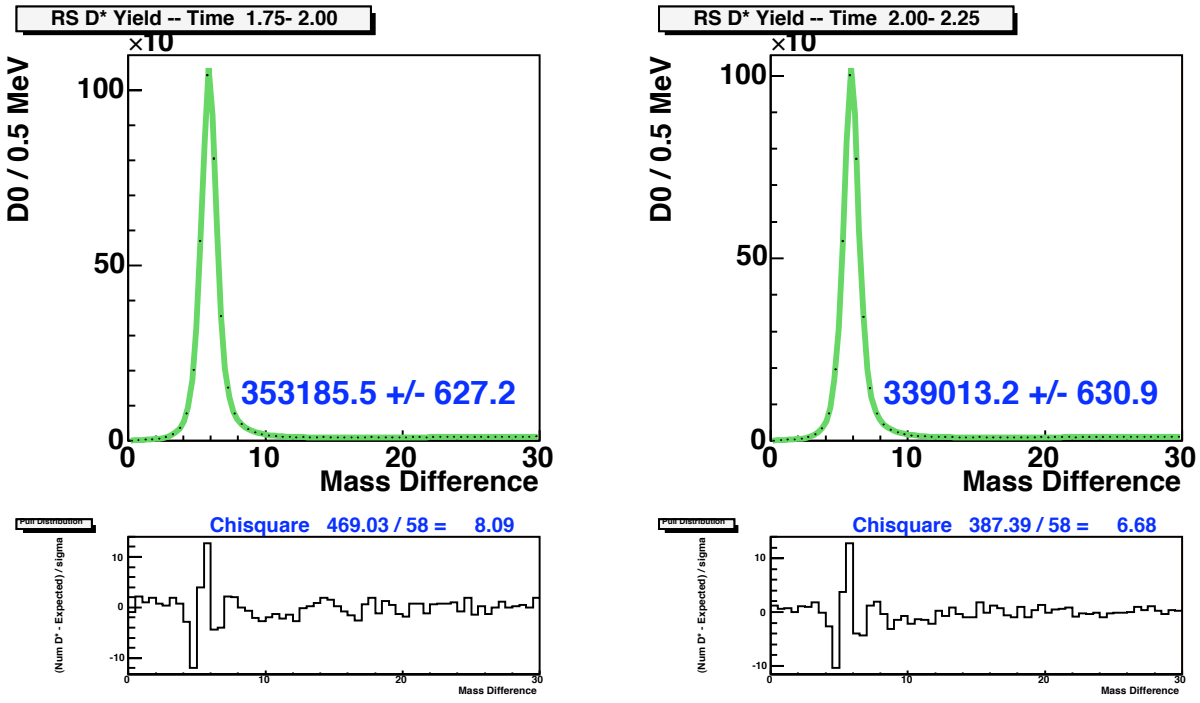
Appendix A: Δm Yield Fits Inside IP Cut

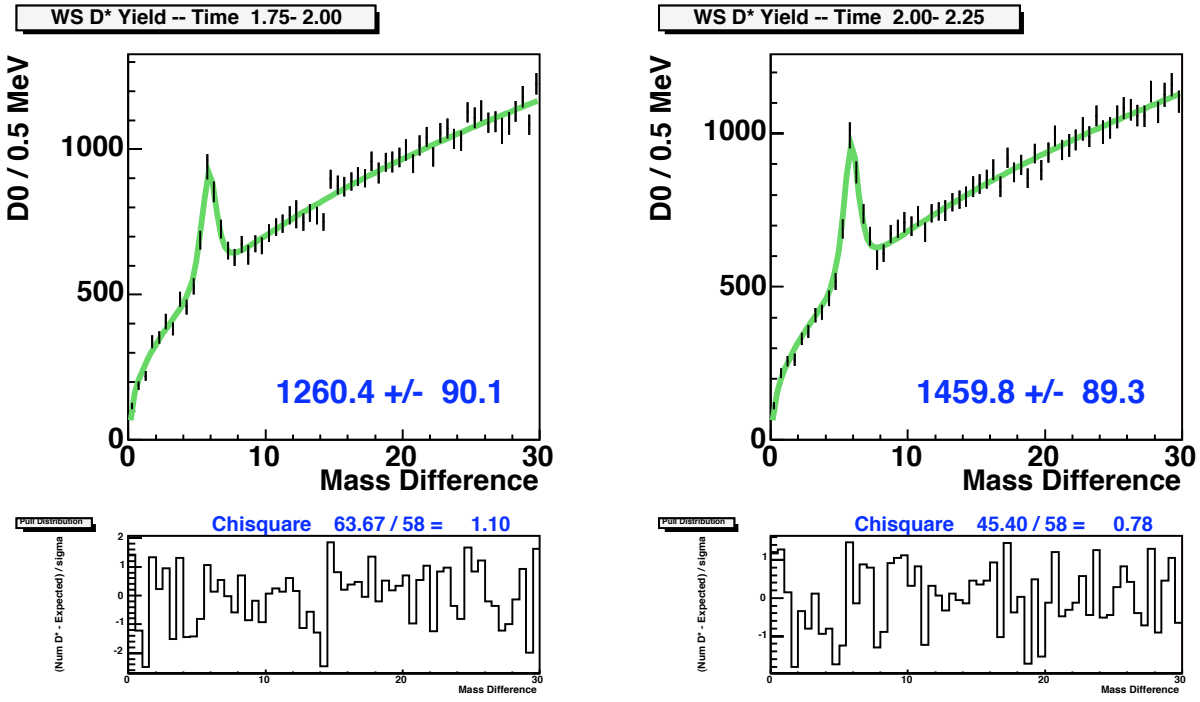


Appendix A: Δm Yield Fits Inside IP Cut

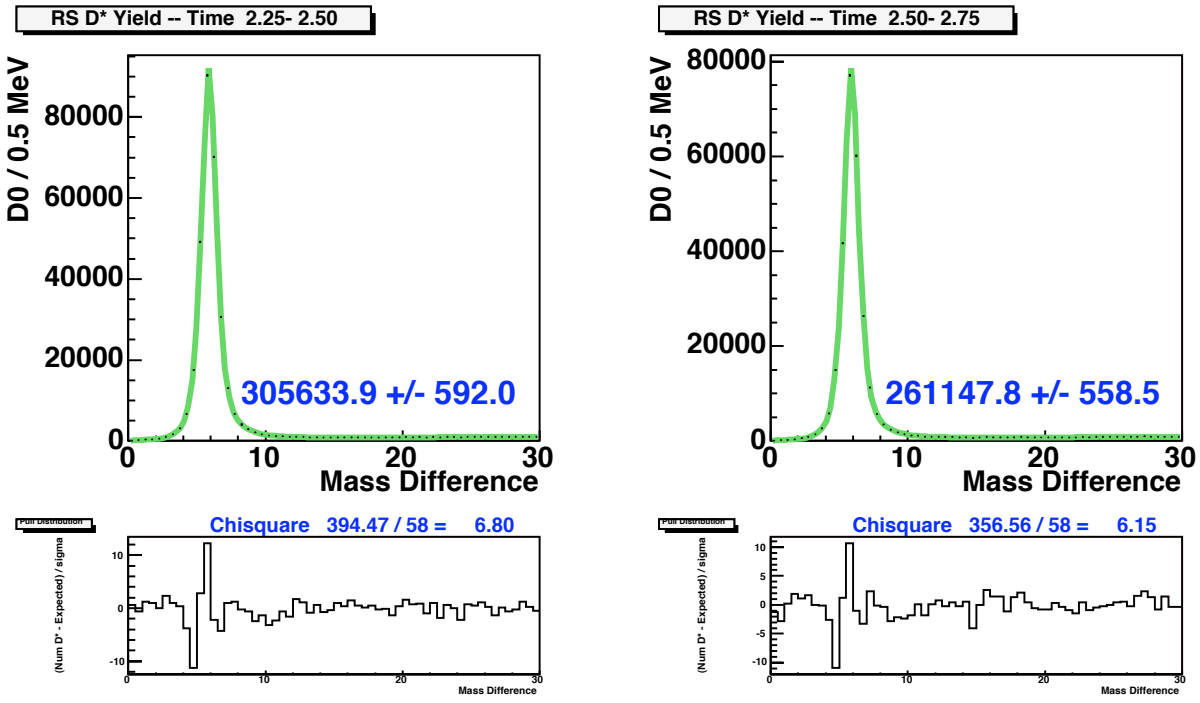


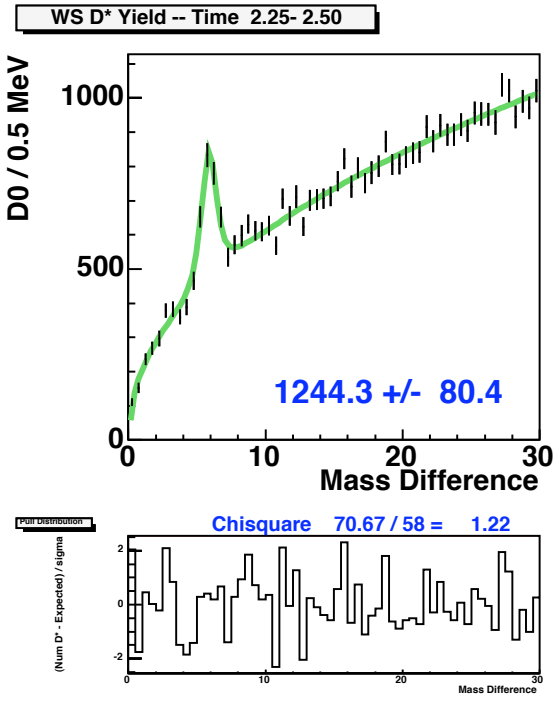
Appendix A: Δm Yield Fits Inside IP Cut

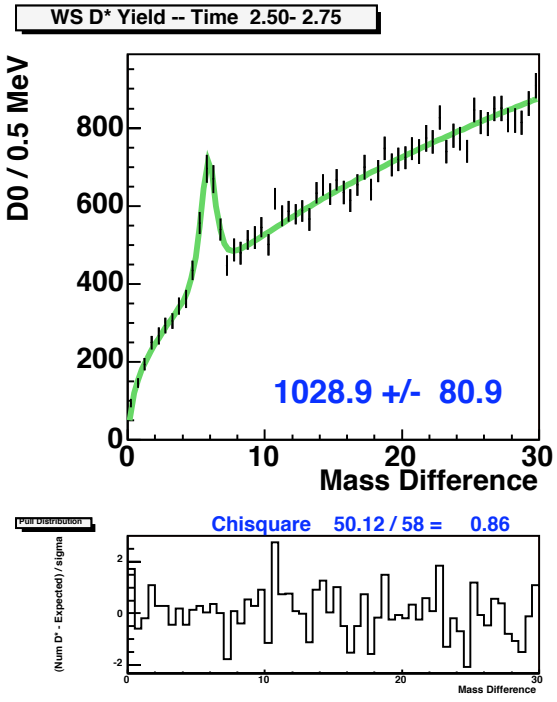




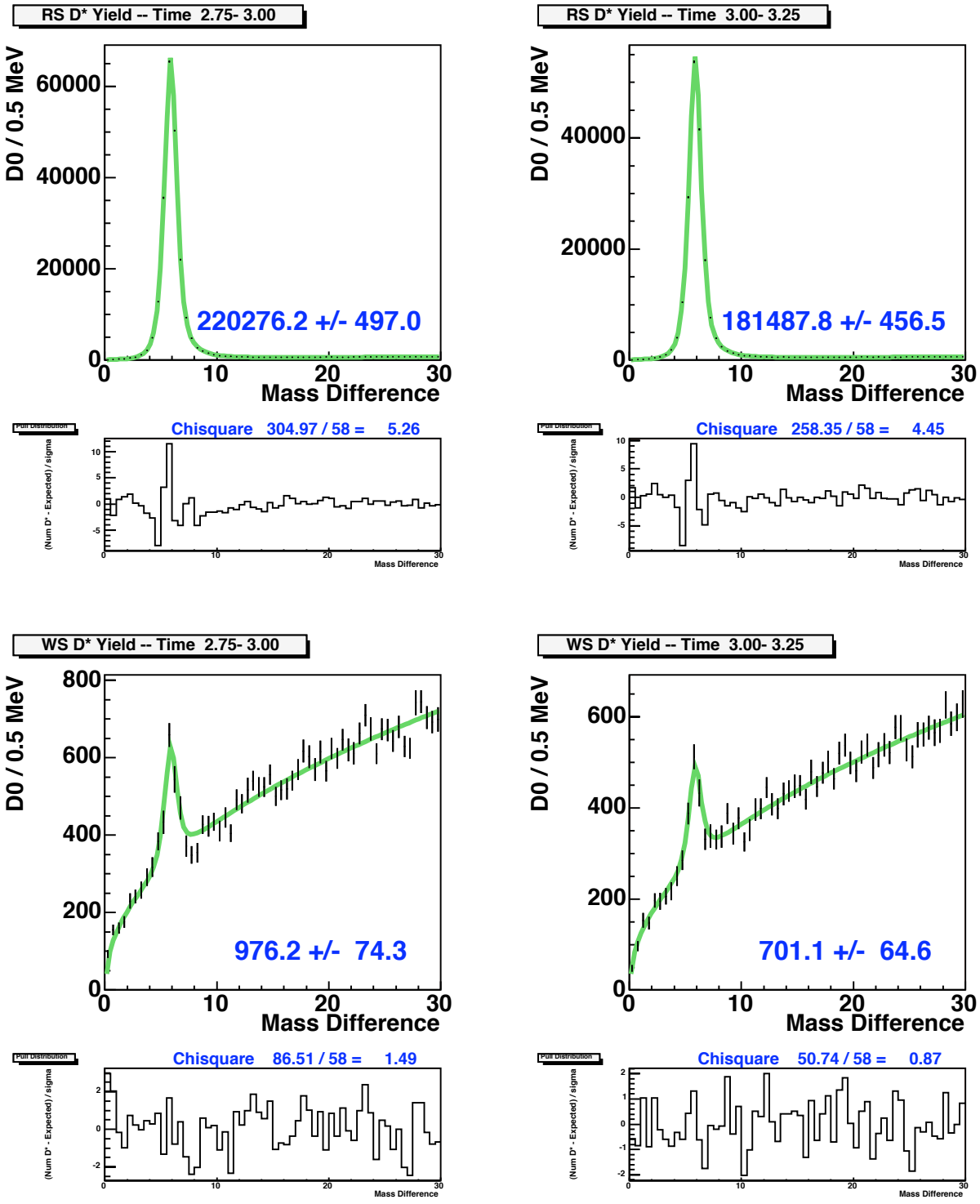
Appendix A: Δm Yield Fits Inside IP Cut



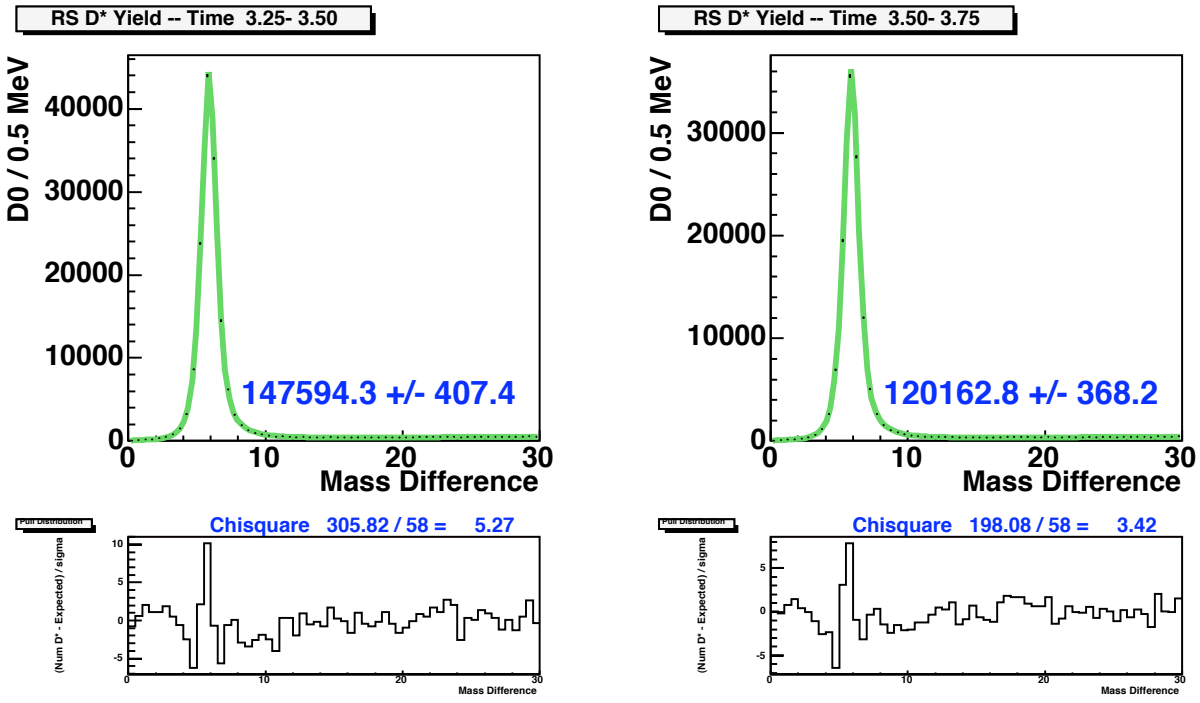


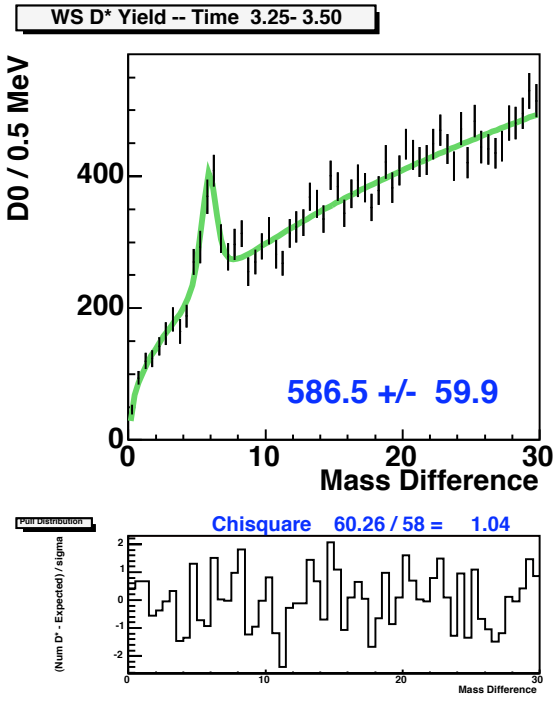


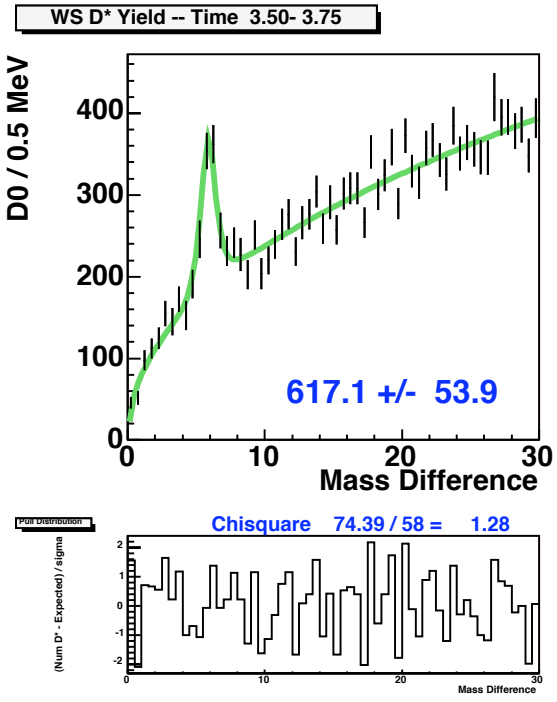
Appendix A: Δm Yield Fits Inside IP Cut



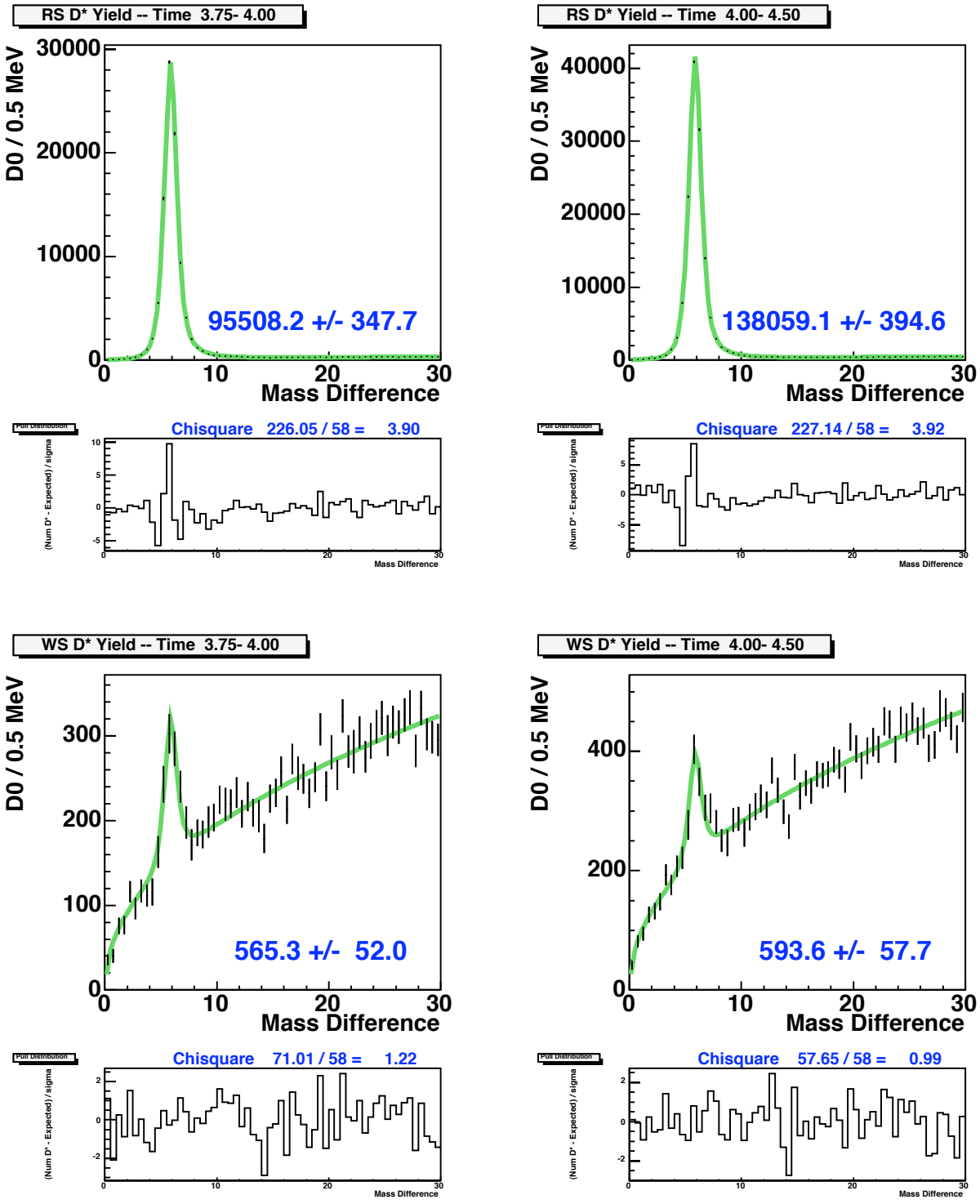
Appendix A: Δm Yield Fits Inside IP Cut



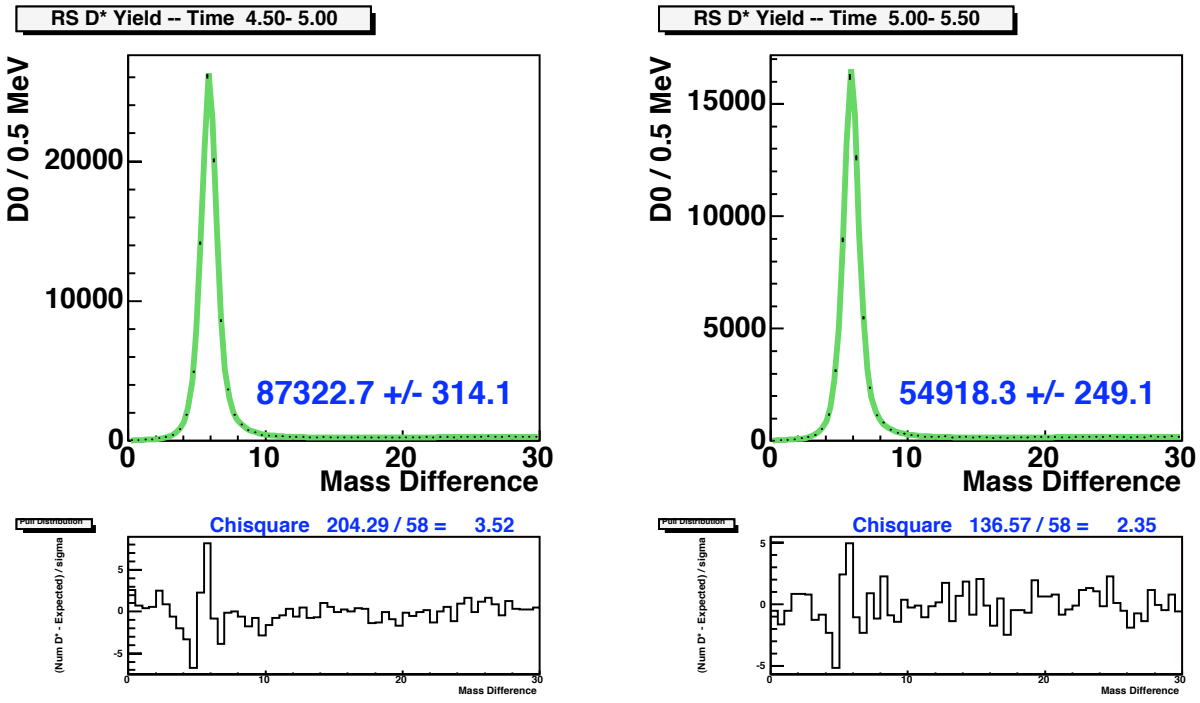


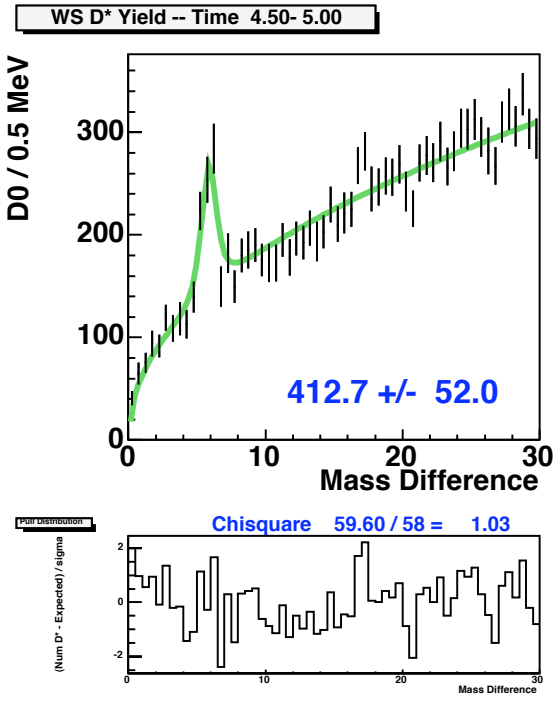


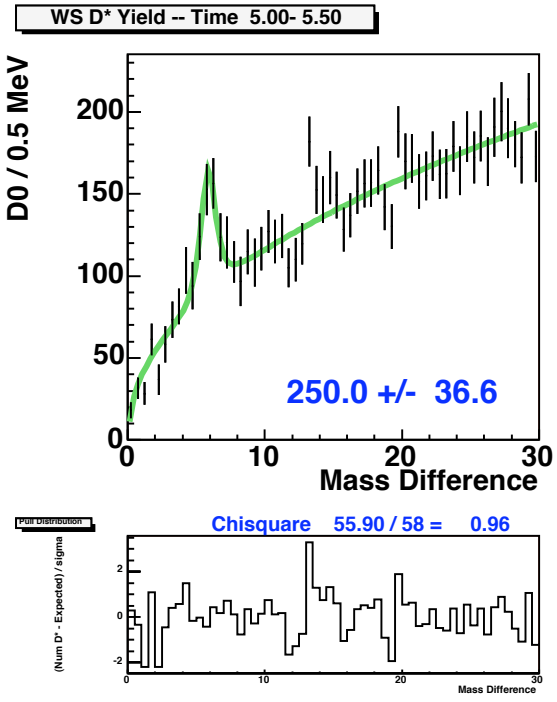
Appendix A: Δm Yield Fits Inside IP Cut



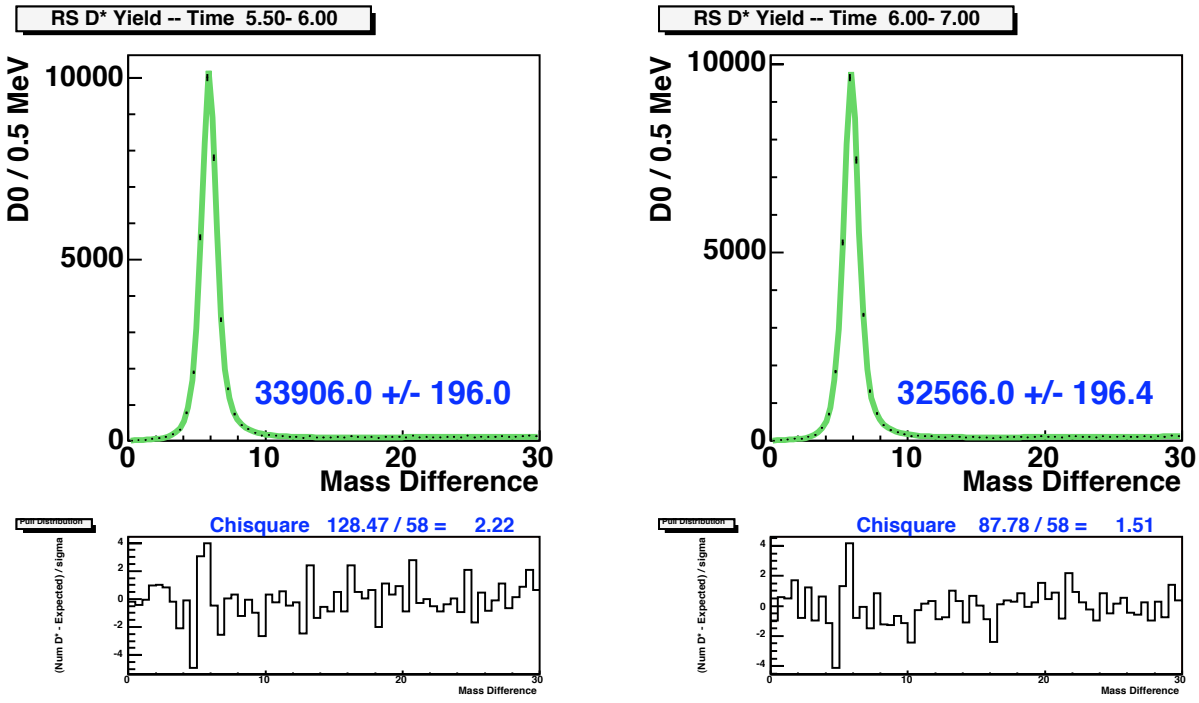
Appendix A: Δm Yield Fits Inside IP Cut

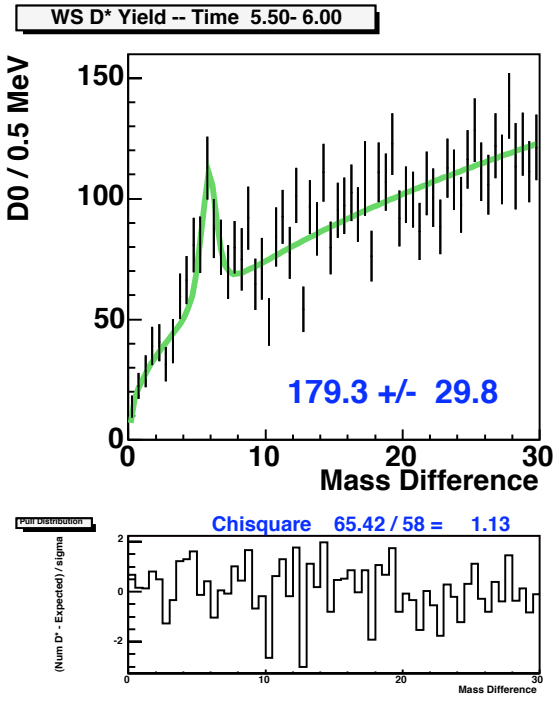


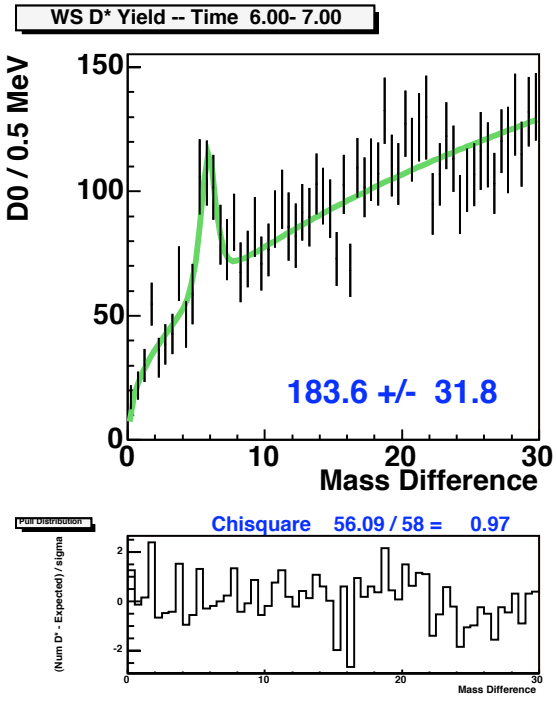




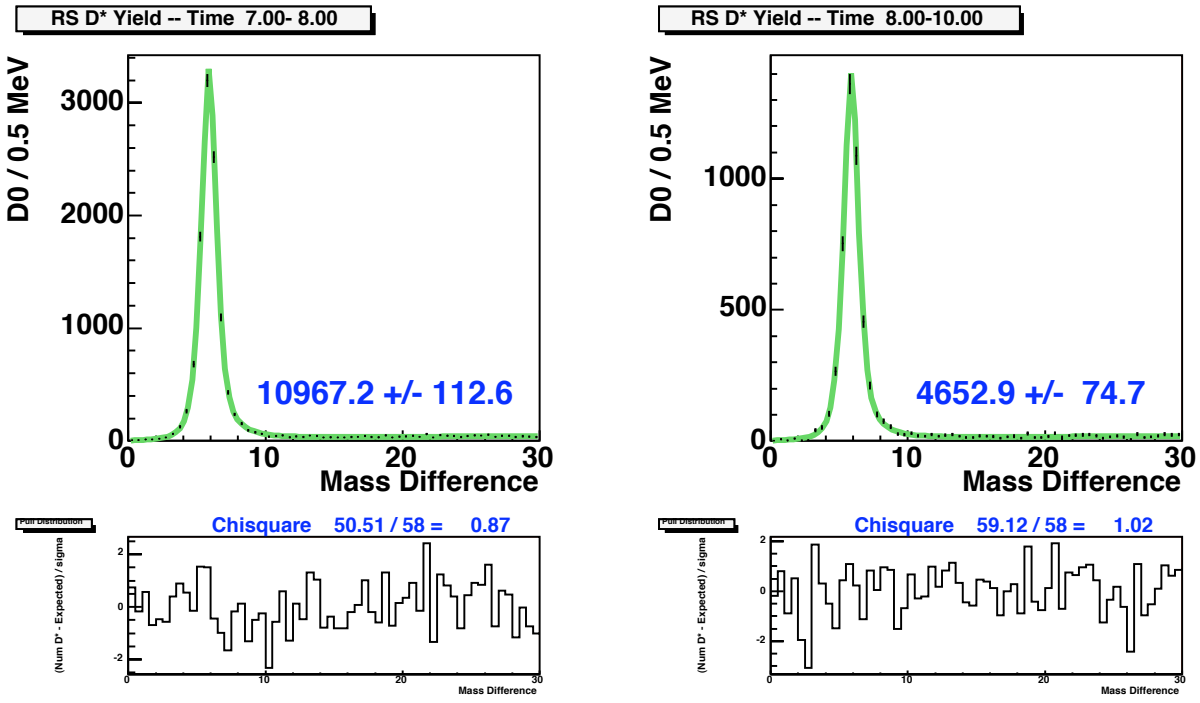
Appendix A: Δm Yield Fits Inside IP Cut

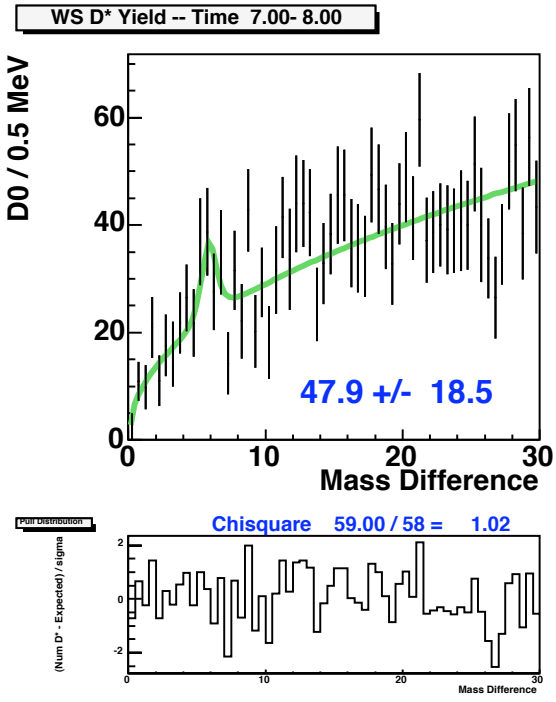


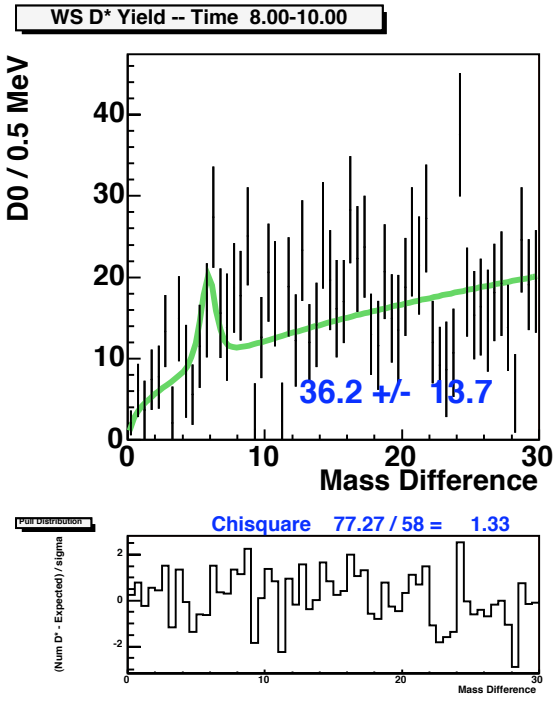




Appendix A: Δm Yield Fits Inside IP Cut



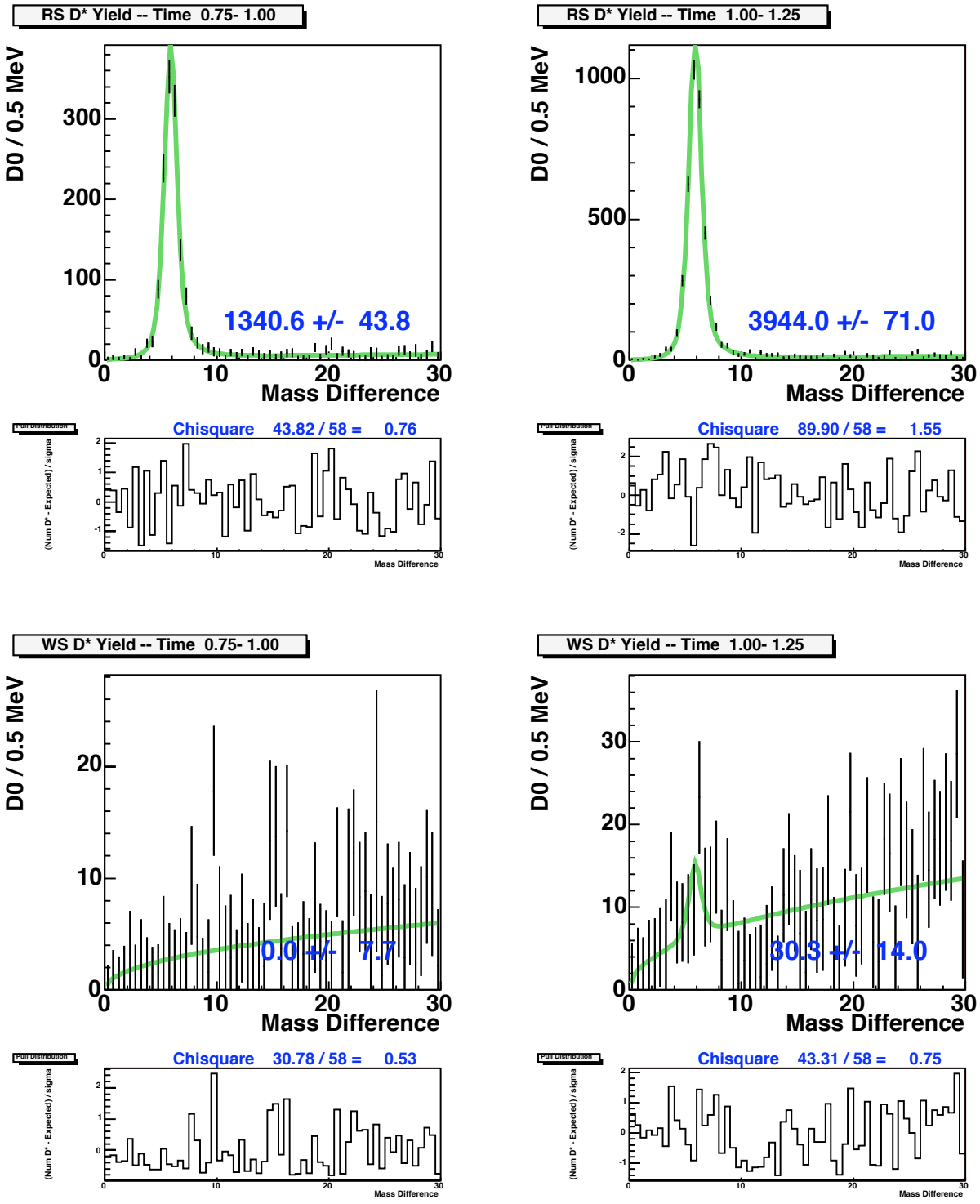




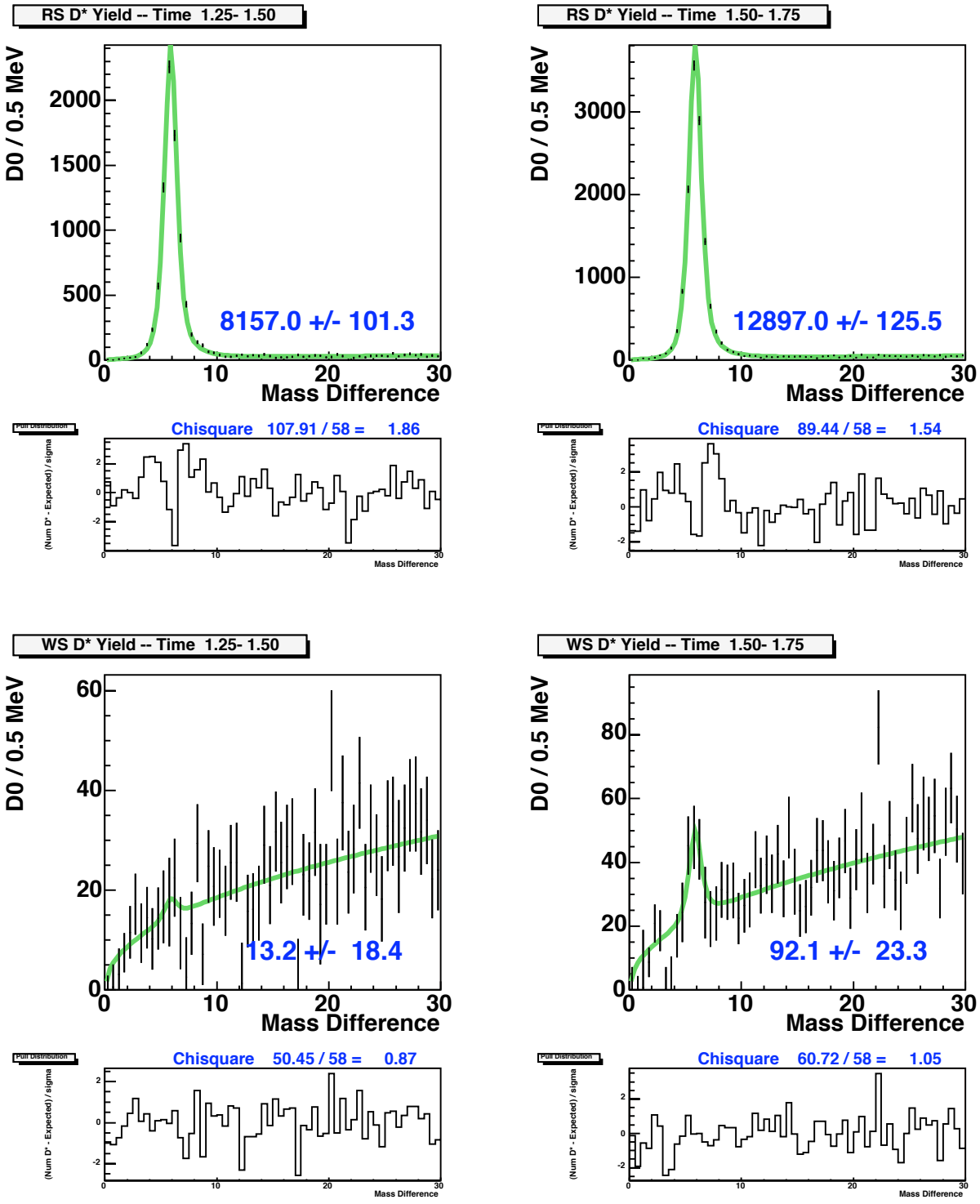
B Mass Difference Fits For Events Outside the IP Cut

This appendix is similar to Appendix A, except that the D^* yields are for events with $|d_0| > 60$ microns. This region will have a higher concentration of D^* not produced at the primary vertex, and is used to calculate the amount of prompt D^*s (at all IP values).

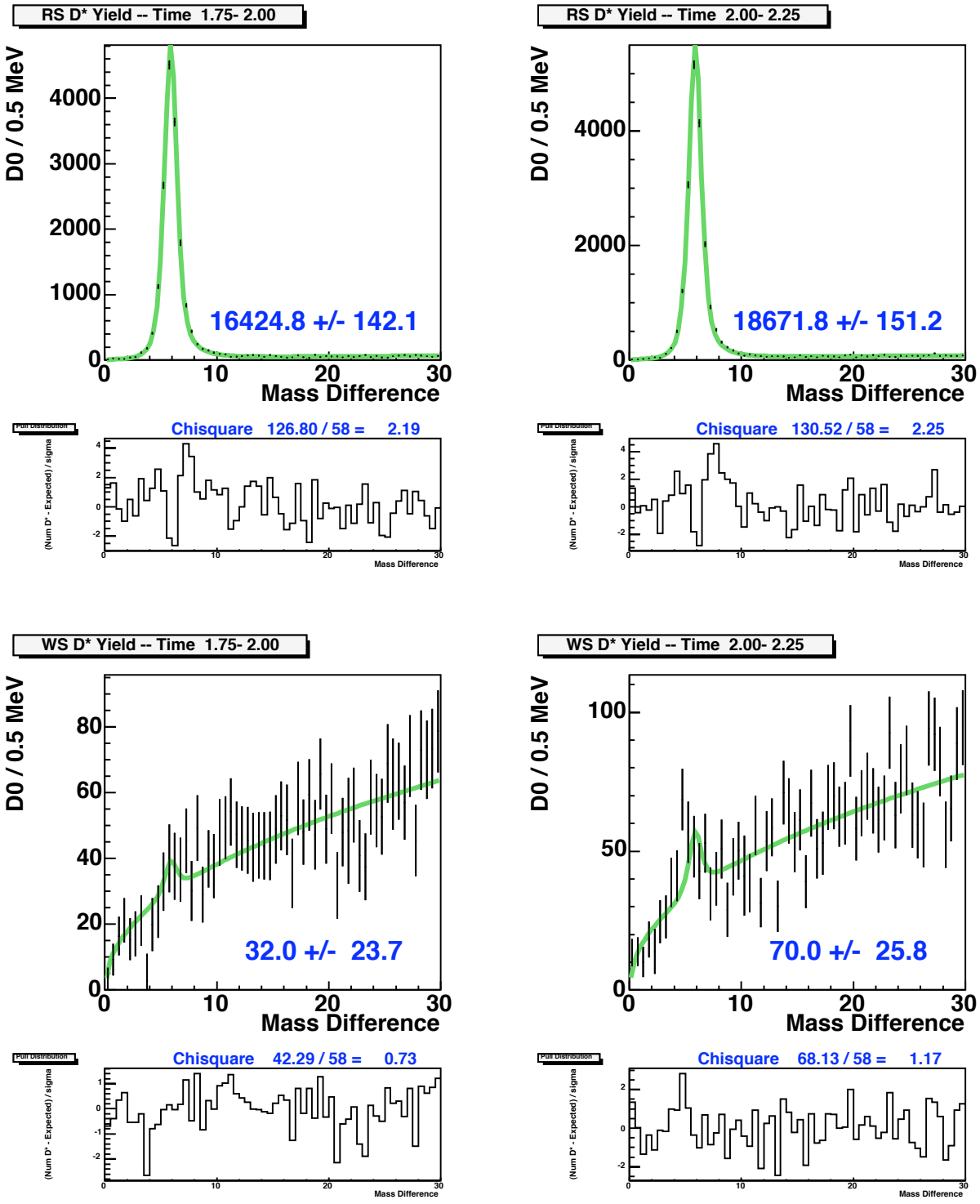
Appendix B: Δm Yield Fits Outside IP Cut



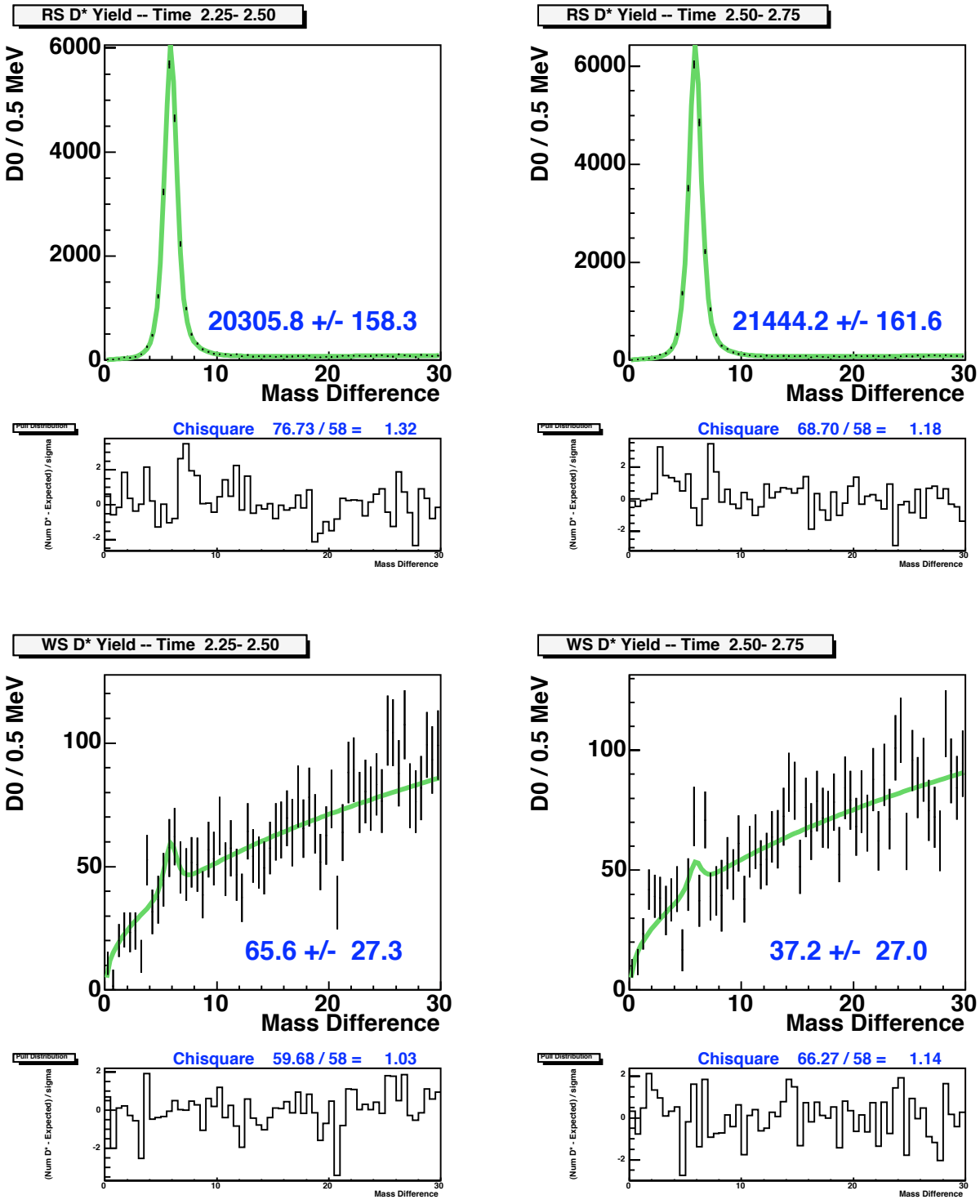
Appendix B: Δm Yield Fits Outside IP Cut



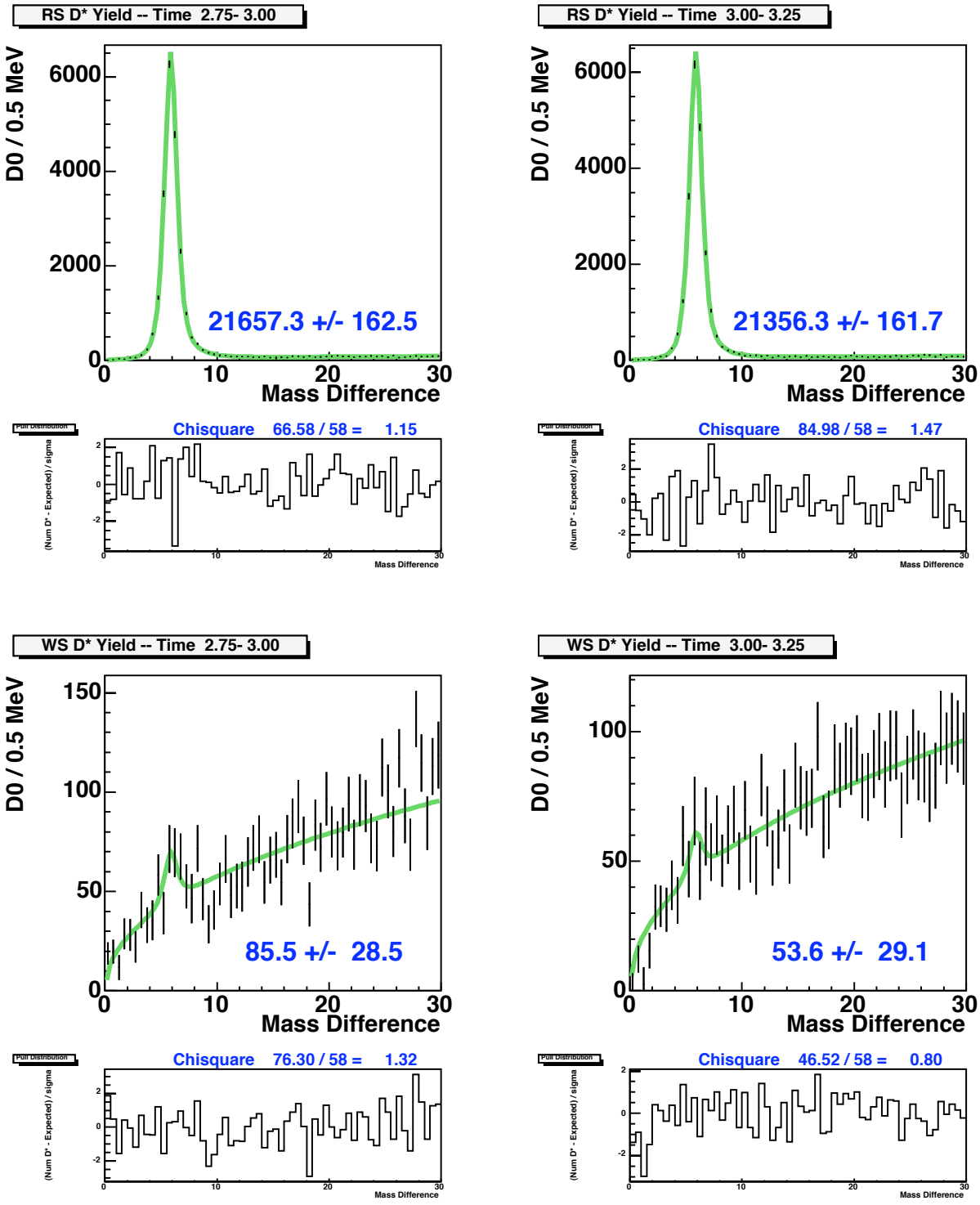
Appendix B: Δm Yield Fits Outside IP Cut



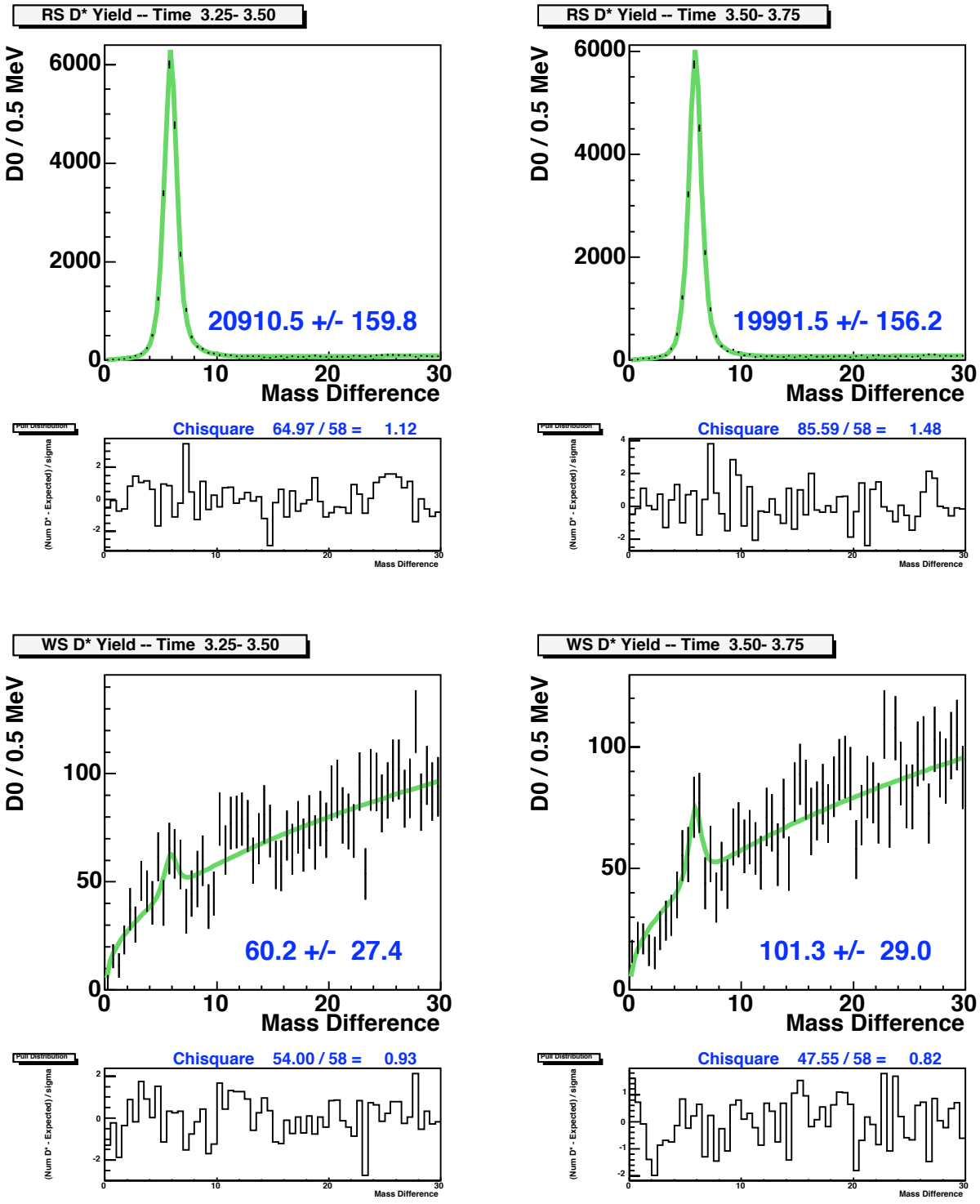
Appendix B: Δm Yield Fits Outside IP Cut



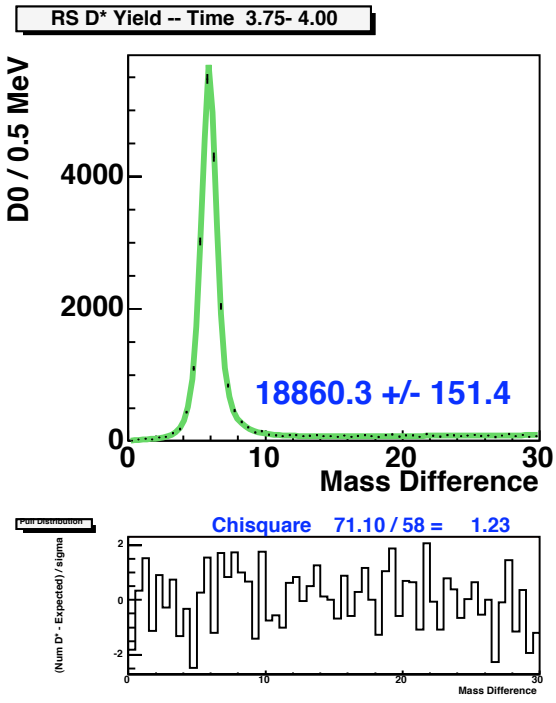
Appendix B: Δm Yield Fits Outside IP Cut

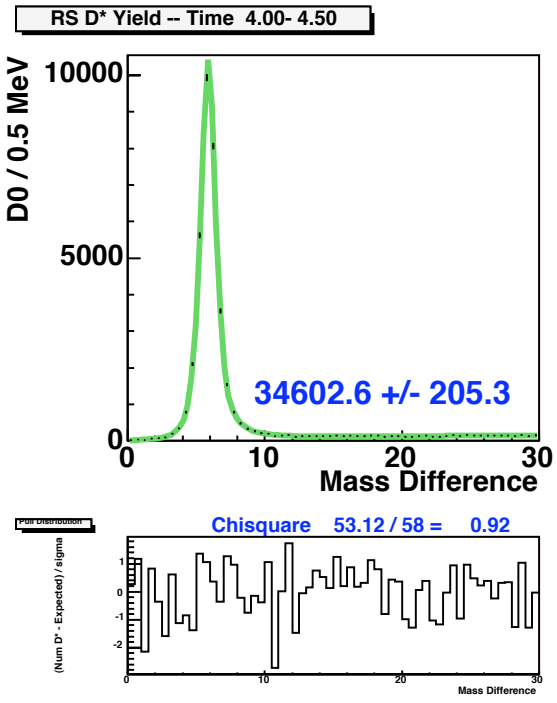


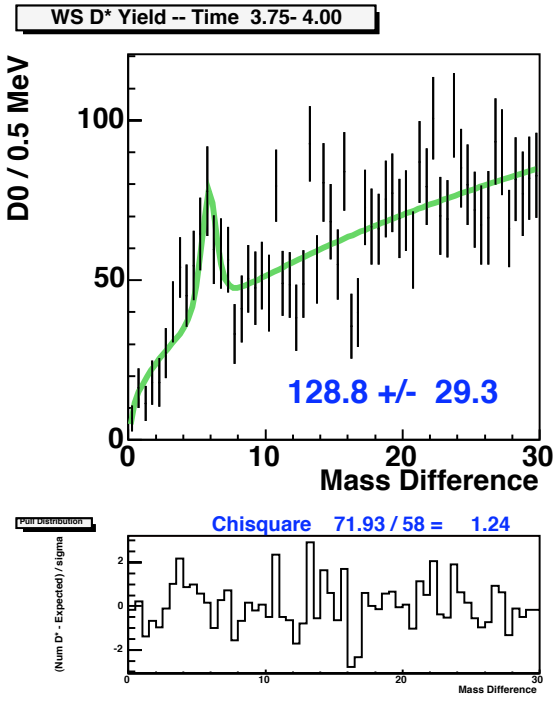
Appendix B: Δm Yield Fits Outside IP Cut

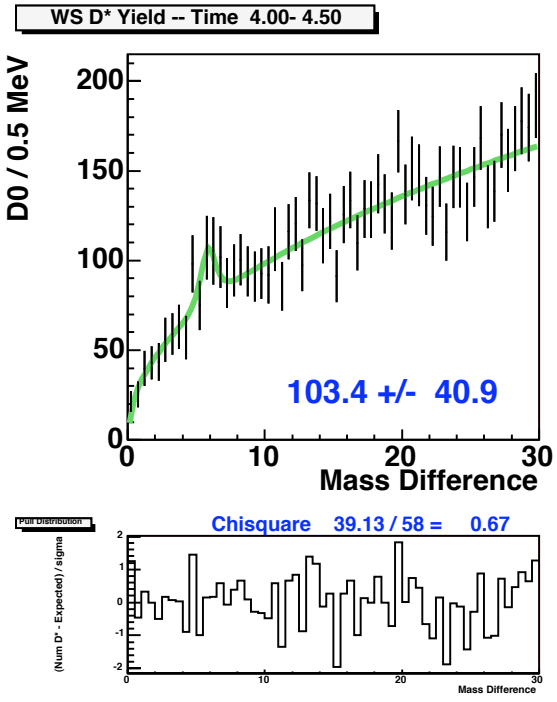


Appendix B: Δm Yield Fits Outside IP Cut

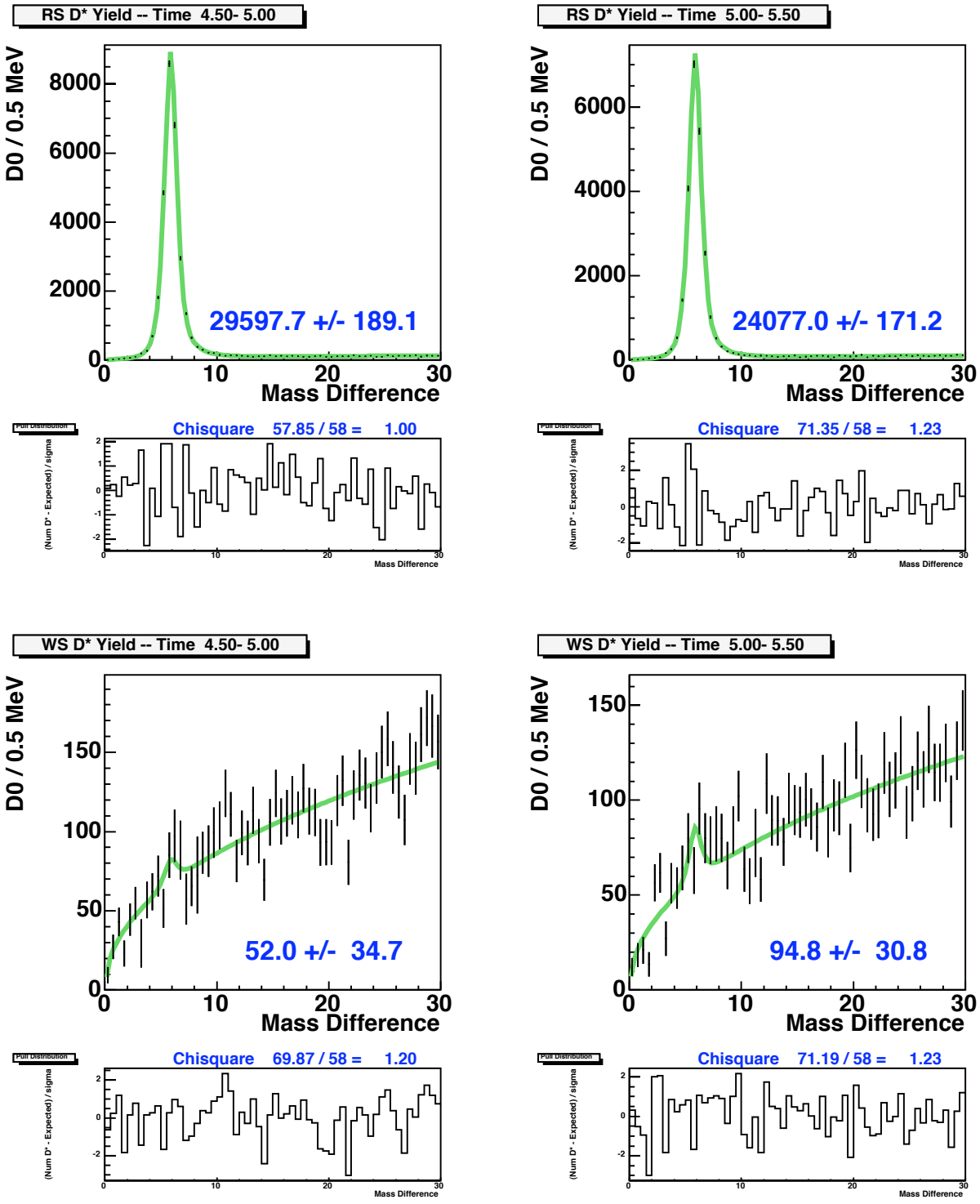




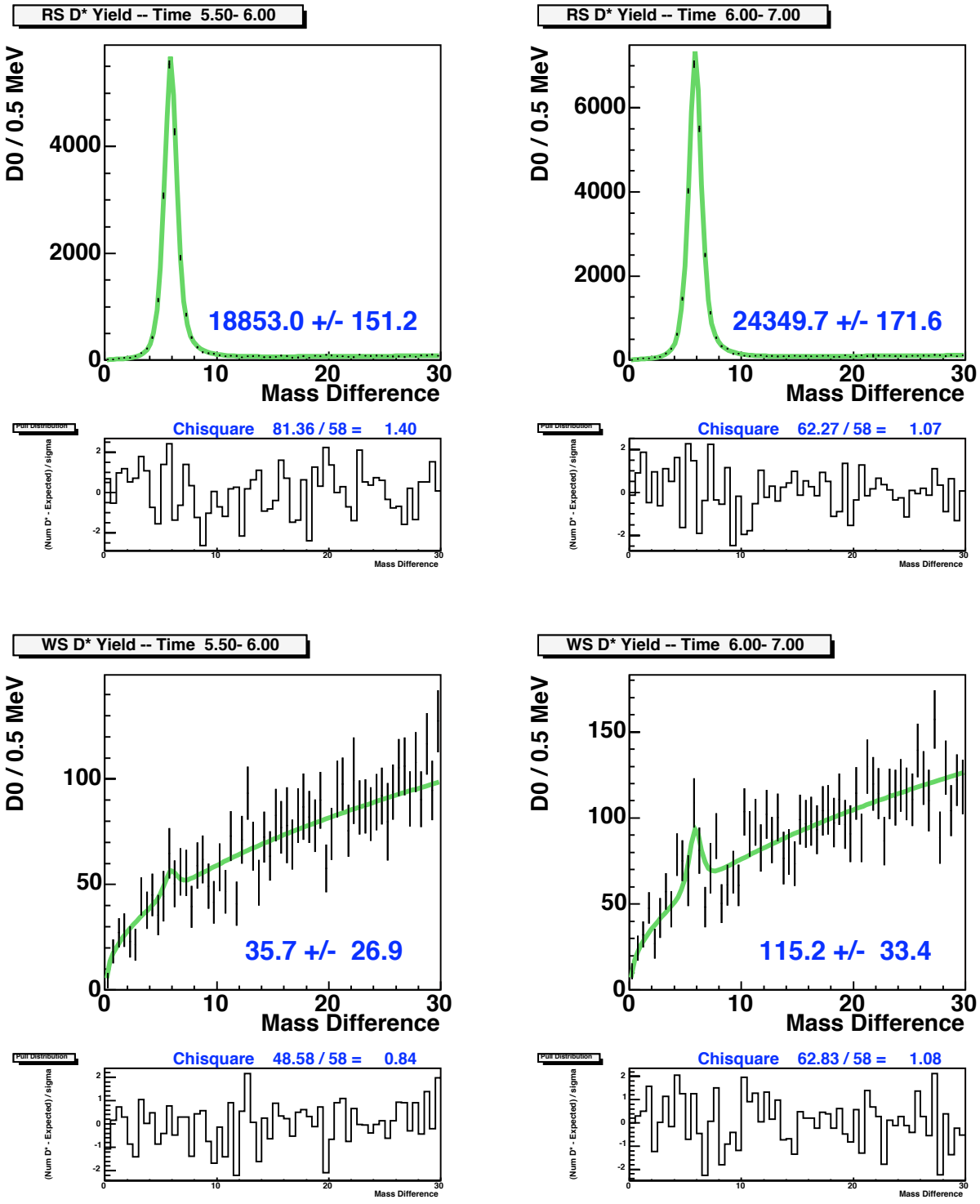




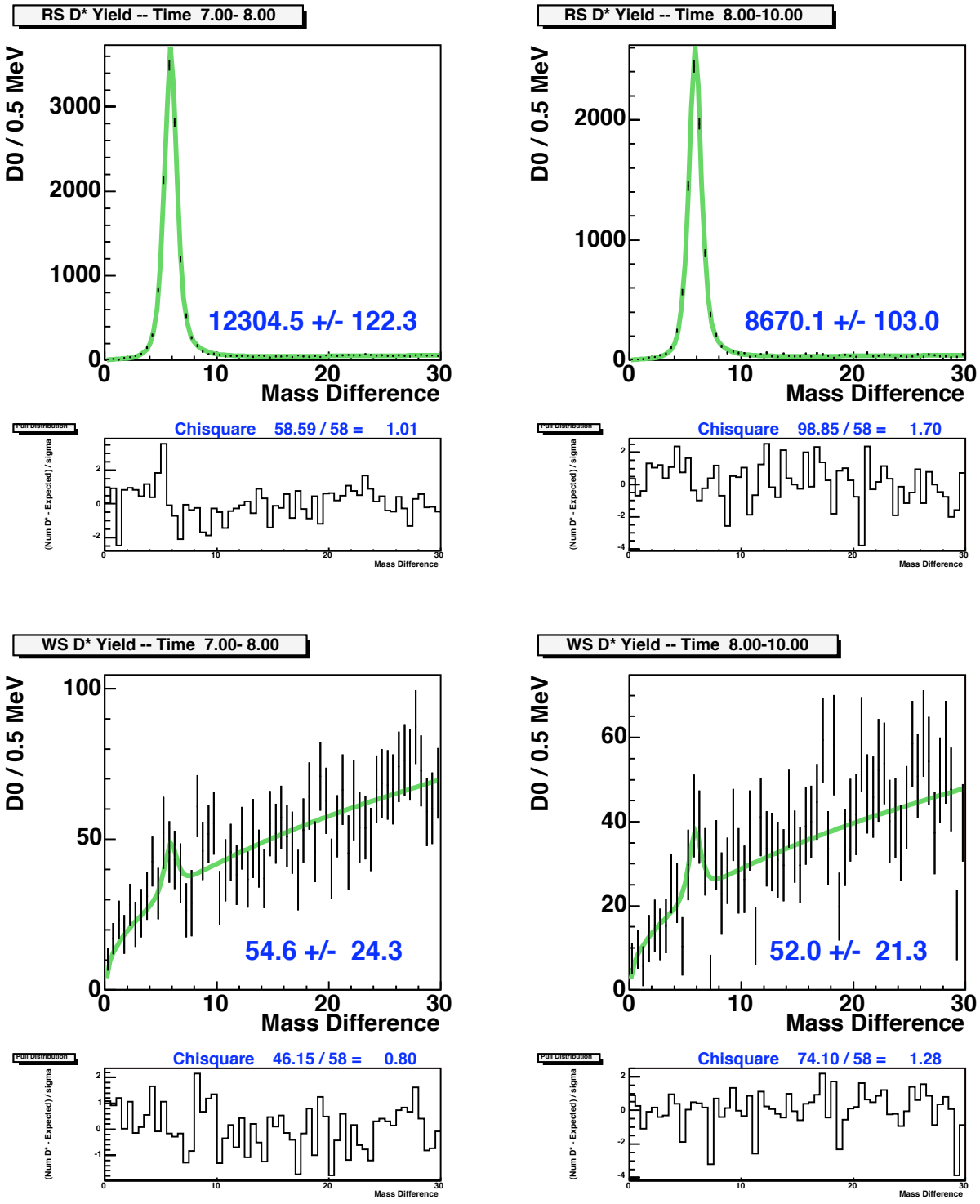
Appendix B: Δm Yield Fits Outside IP Cut



Appendix B: Δm Yield Fits Outside IP Cut



Appendix B: Δm Yield Fits Outside IP Cut



C Non-Prompt D^* Correction Details

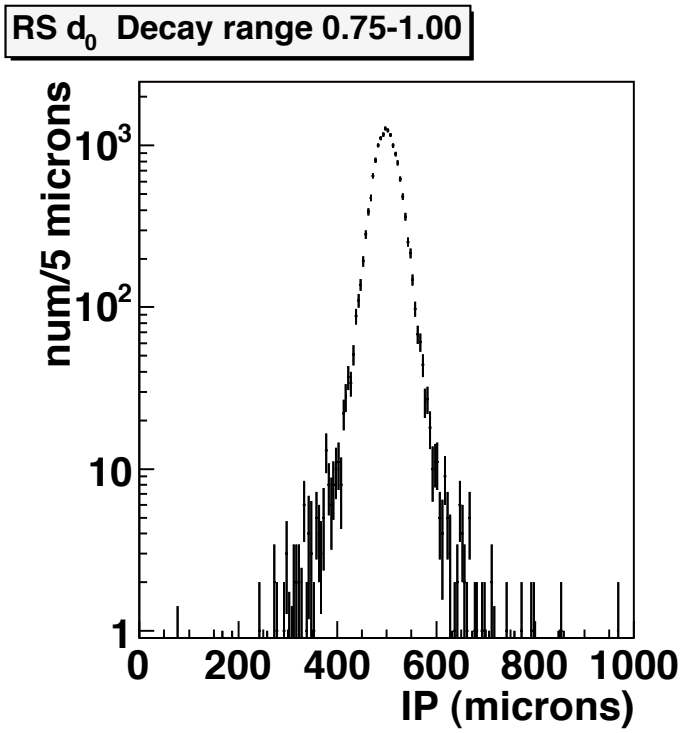
This appendix shows the data results for the B-decay correction, starting with the fits for the time-binned impact parameter distributions. We use RS data, with $(K\pi)$ sideband-subtraction. The distribution for each time bin is fit with a double Gaussian (sum of two Gaussians) for the prompt (signal)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

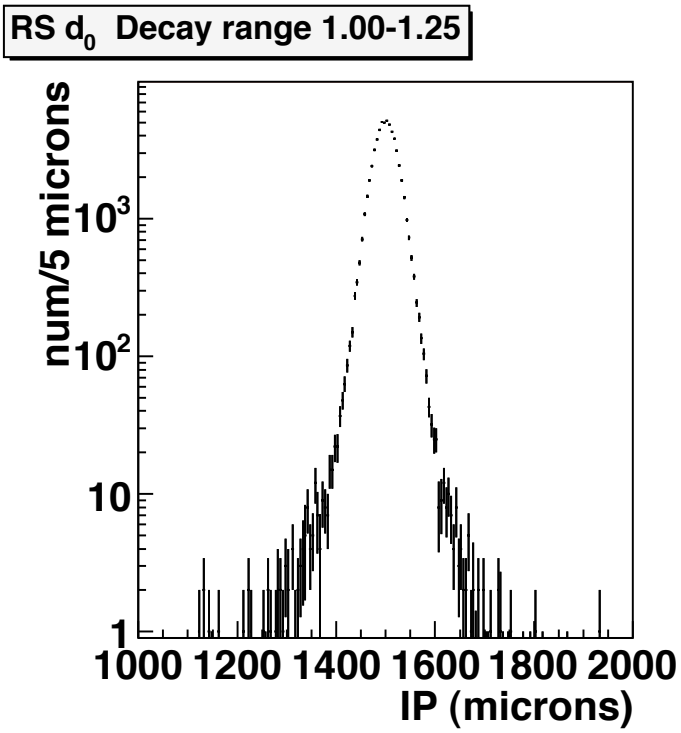
Each plot covers the (signed) d_0 values from -500 to +500 microns (ignore the x-axis numbers). The first 4 time bins are excluded, since the fitter was having problems distinguishing between prompt and non-prompt distributions. The other 16 bins were fit simultaneously, using the same prompt shape for all time bins. The blue curve is the non-prompt fit, the green curve is prompt plus non-prompt.

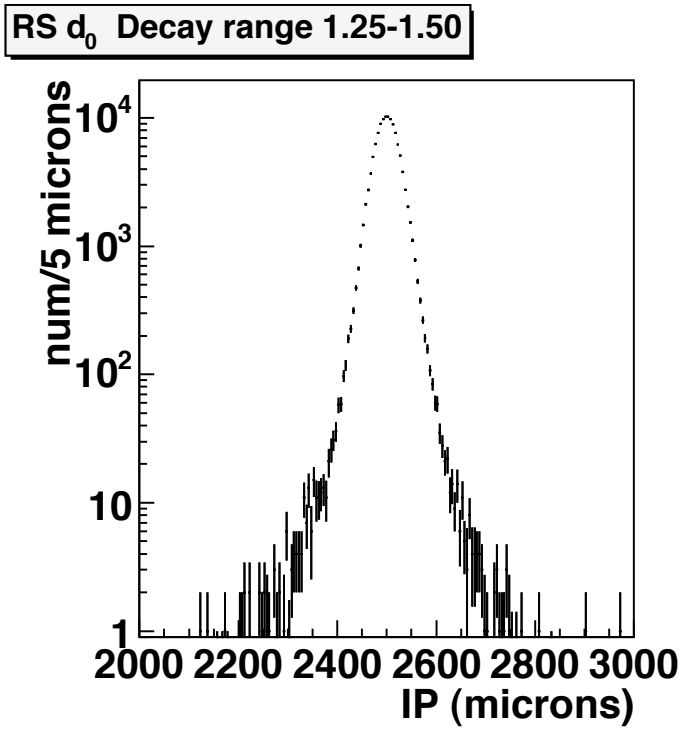
The signal fit parameters are: $f = 27.816 \pm 0.743\,\%$ (fraction of the second Gaussian to the total amplitude), $\sigma_1 = 21.374 \pm 0.087\,\mu$m (width of the first Gaussian), and $\sigma_2 = 39.412 \pm 0.387\,\mu$m (width of the second Gaussian). The parameterization for the non-prompt Gaussian width is a 3rd order polynomial: $p_0 = 27.9\,\mu$m, $p_1 = 31.769 \pm 0.854\,\mu$m, $p_2 = 0.089 \pm 0.317\,\mu$m, and $p_3 = -0.1025 \pm 0.0285\,\mu$m.

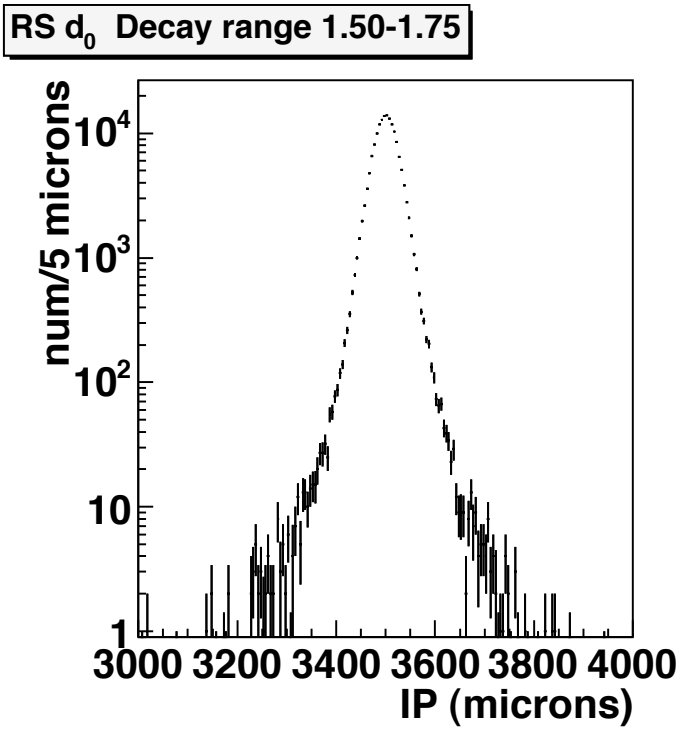
Following the procedure in section 4.3 to get the number of D^*s produced at the primary vertex. We get the number of D^*s (for each time bin) for two impact parameter regions: $|d_0| \leq 60\mu$m (inside), and $60 < |d_0| \leq 500\mu$m (outside). The amount of the prompt and non-prompt distributions in each IP region is summarized in table 11. The D^* numbers are summarized in tables 12-13.

Appendix C: RS d_0 Distributions By Decay Time

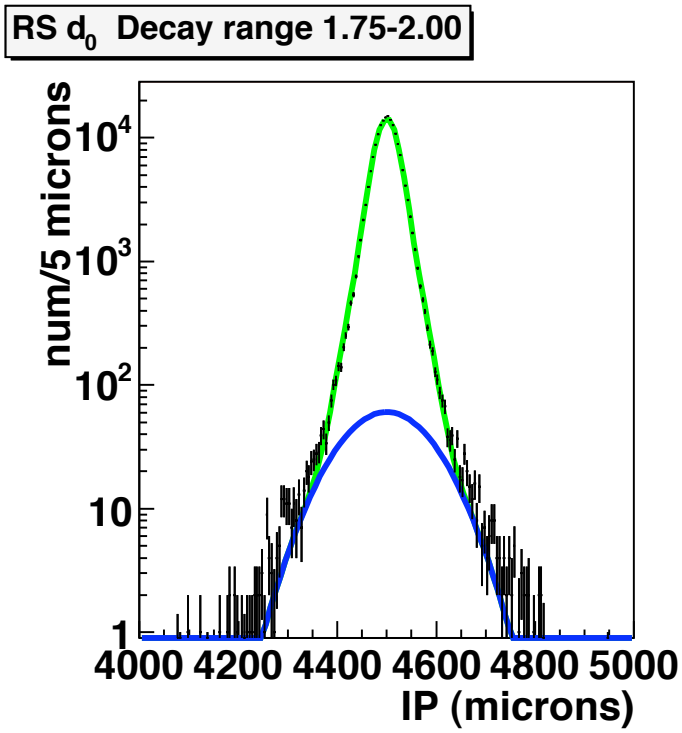


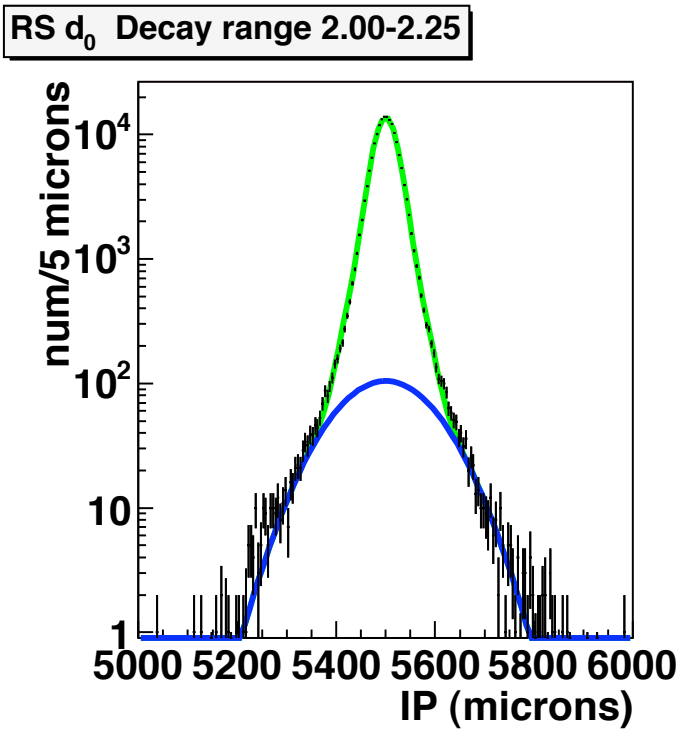


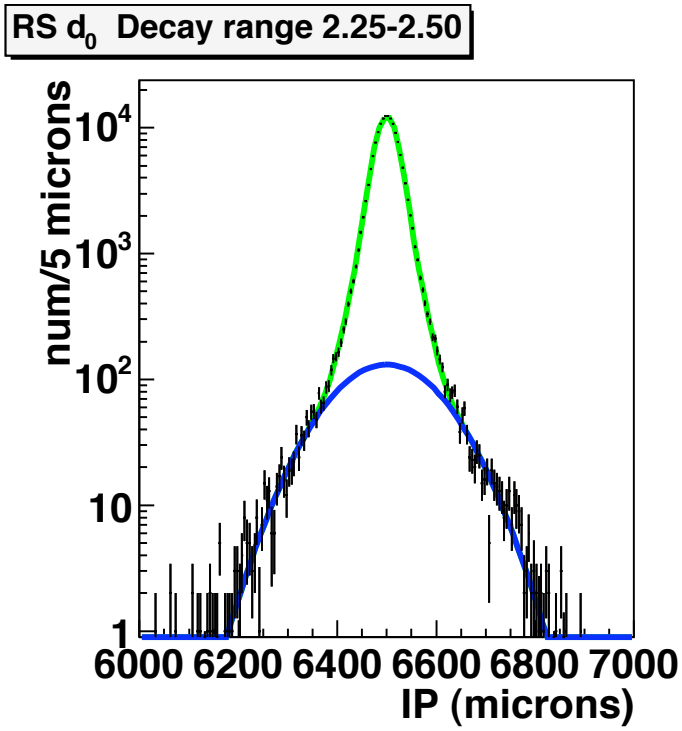


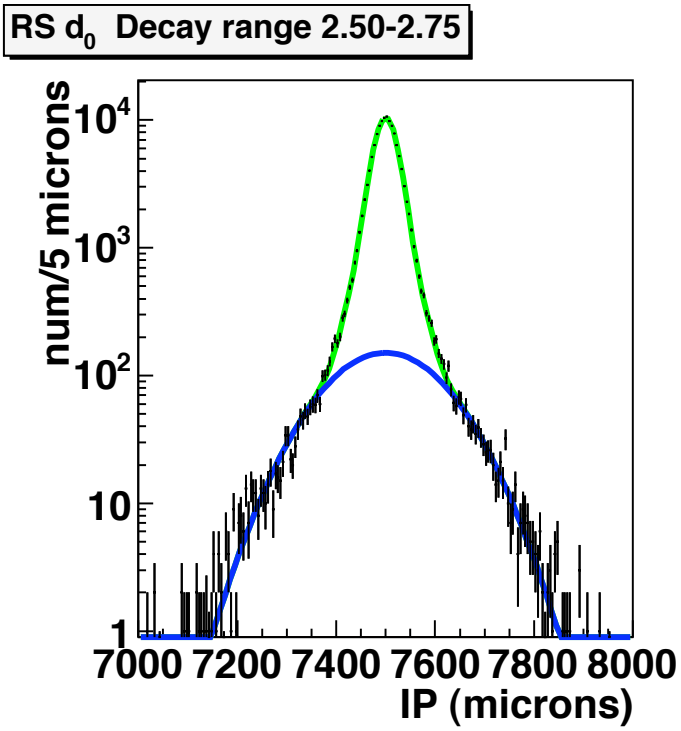


Appendix C: RS d_0 Distributions By Decay Time

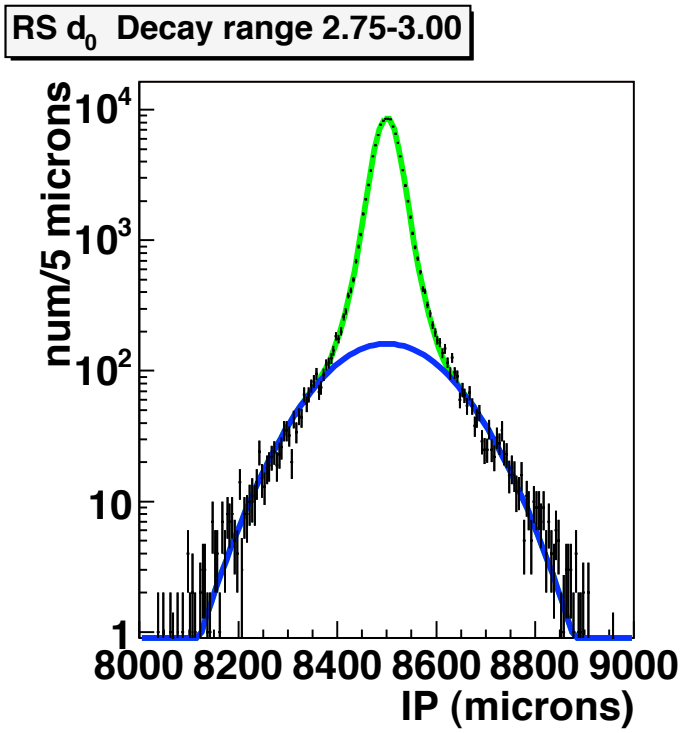


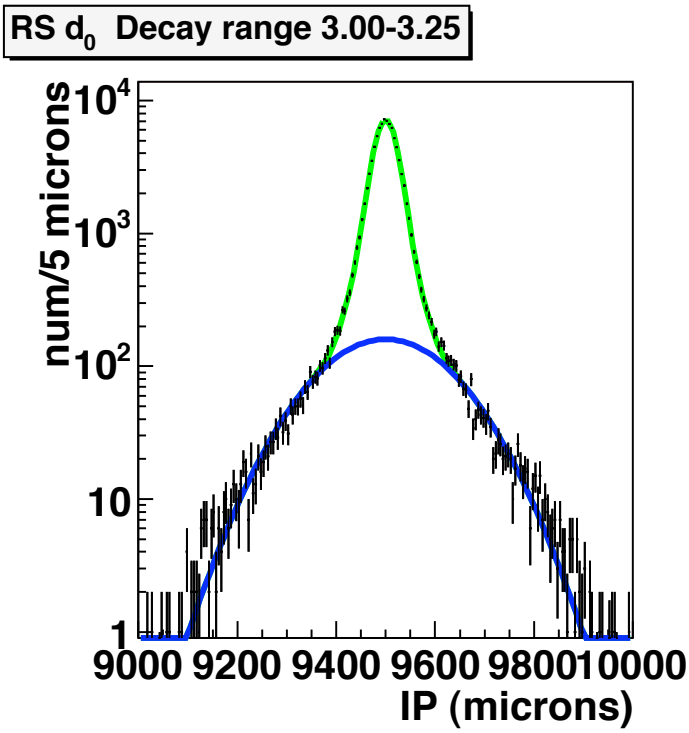


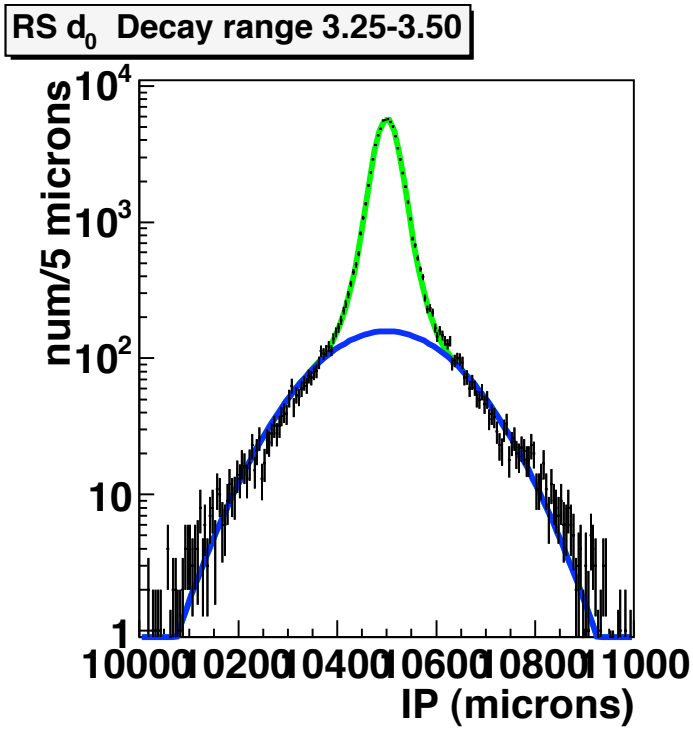


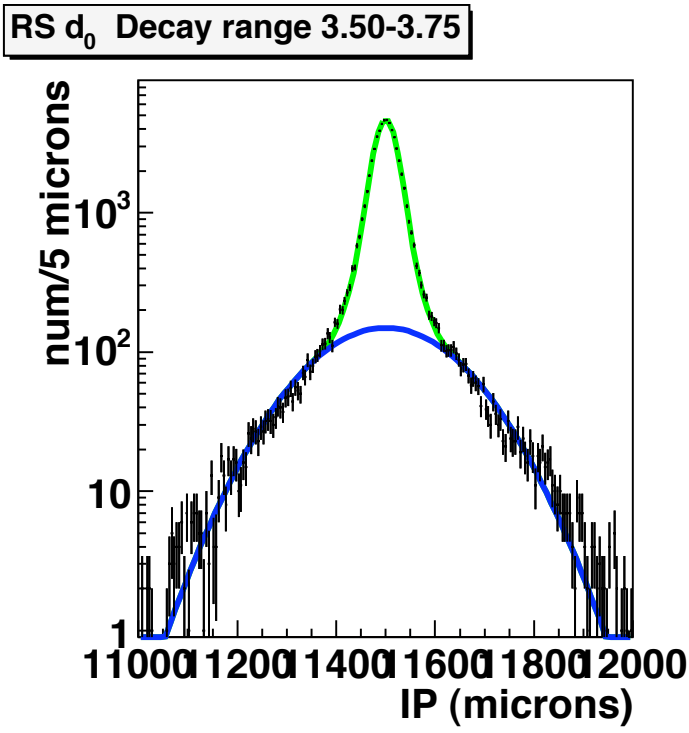


Appendix C: RS d_0 Distributions By Decay Time

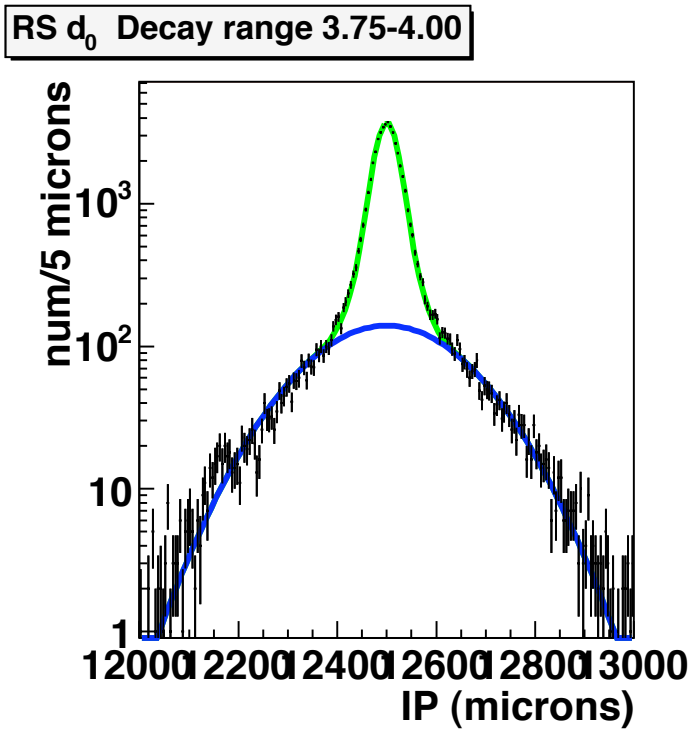


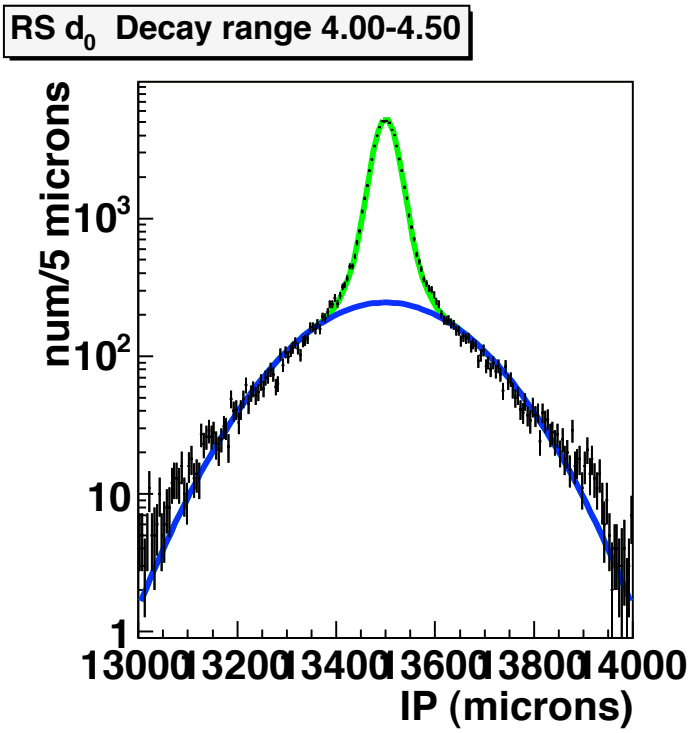


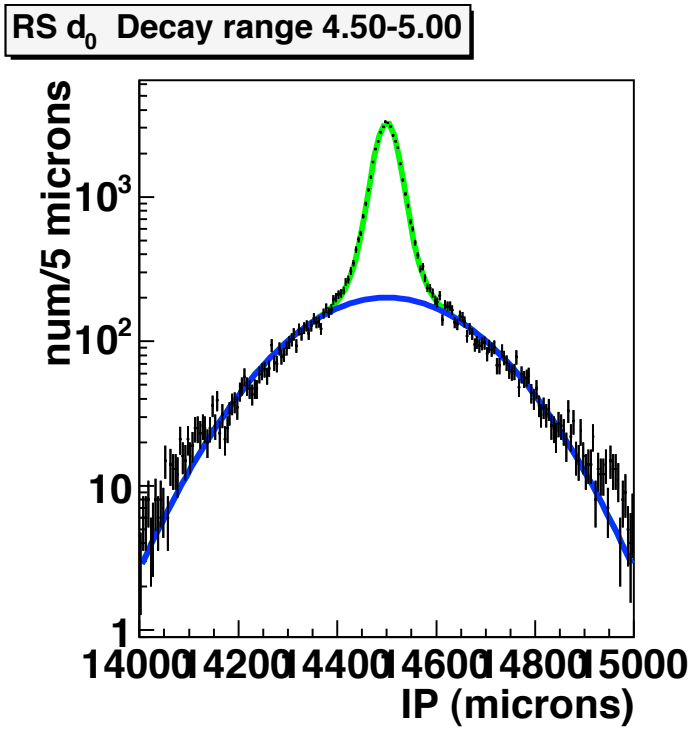


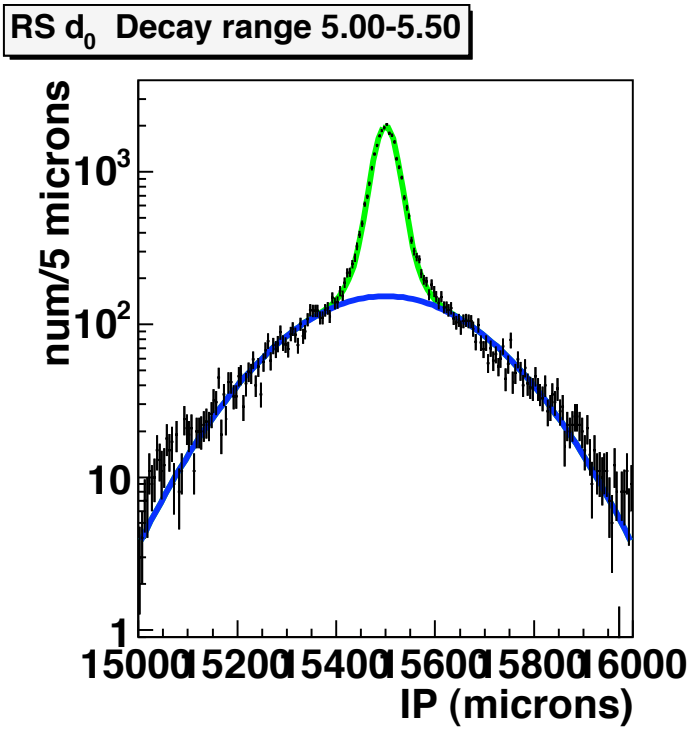


Appendix C: RS d_0 Distributions By Decay Time

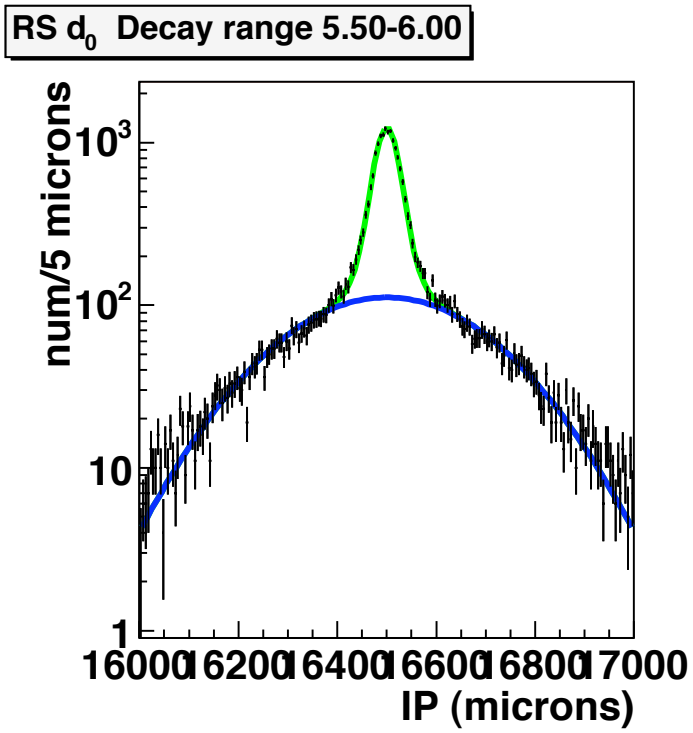


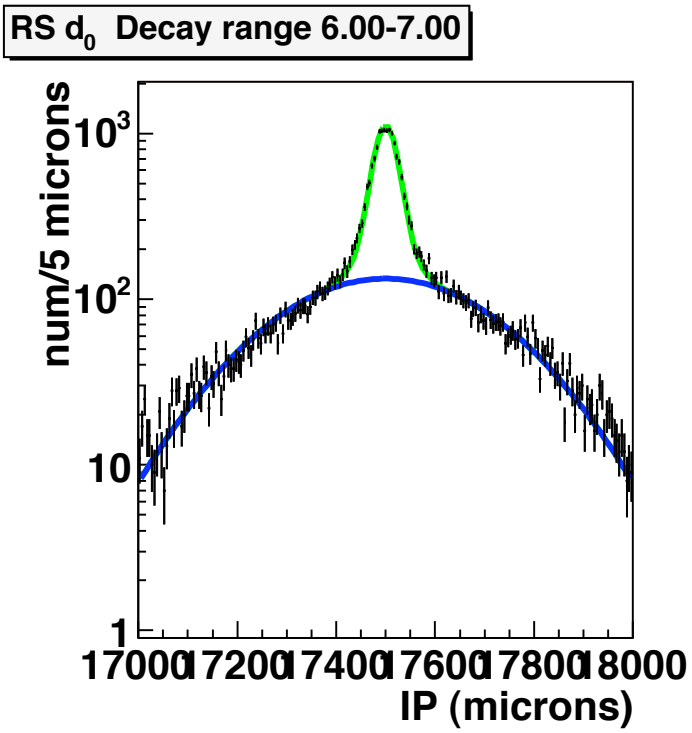


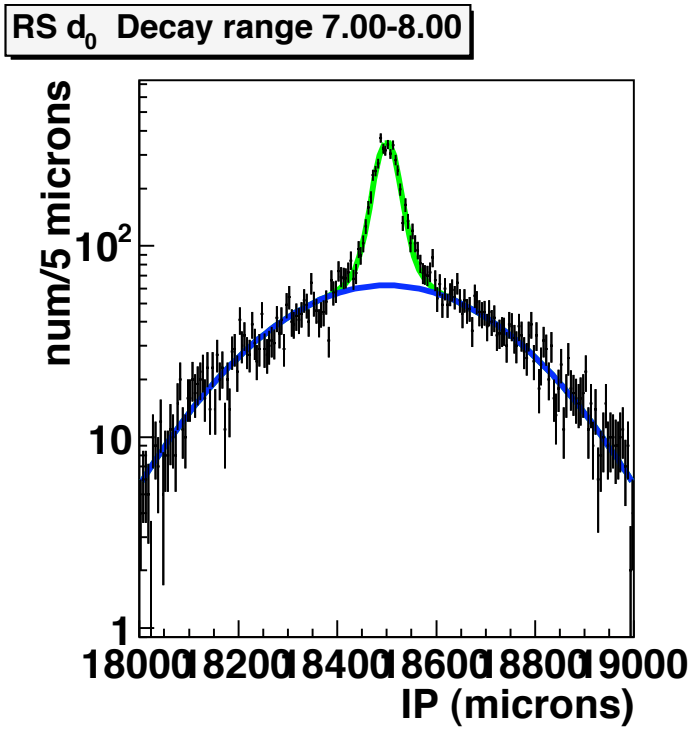


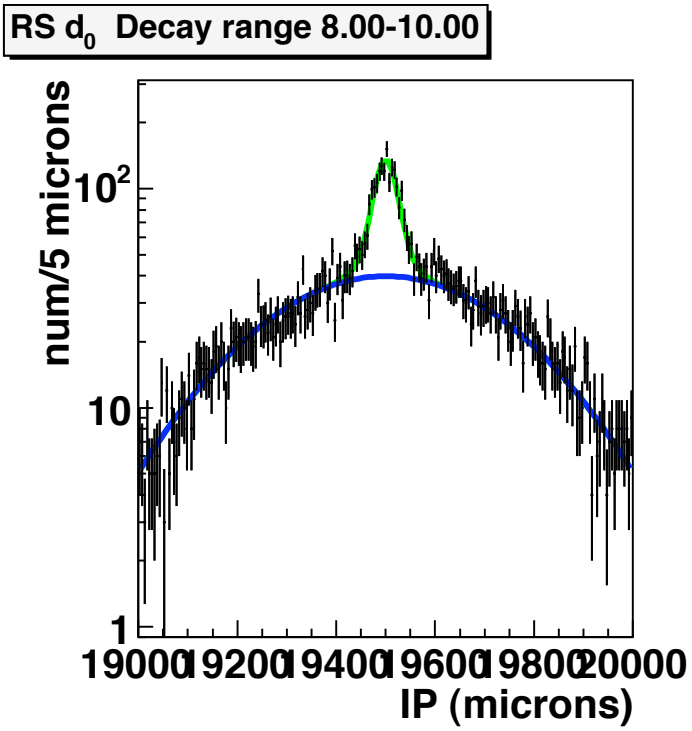


Appendix C: RS d_0 Distributions By Decay Time









Time Bin	Fraction In (f_i)	σ_{f_i}	Fraction Out (g_i)
prompt distribution	0.96081	0.00039	0.03019
0.75-1.00	0.718720	0.004590	0.281280
1.00-1.25	0.654550	0.004650	0.345450
1.25-1.50	0.598790	0.004410	0.401210
1.50-1.75	0.550670	0.004020	0.449330
1.75-2.00	0.509120	0.003570	0.490880
2.00-2.25	0.473130	0.003110	0.526870
2.25-2.50	0.441810	0.002680	0.558190
2.50-2.75	0.414410	0.002290	0.585580
2.75-3.00	0.390310	0.001950	0.609670
3.00-3.25	0.369010	0.001660	0.630930
3.25-3.50	0.350100	0.001430	0.649740
3.50-3.75	0.333230	0.001250	0.666440
3.75-4.00	0.318130	0.001130	0.681230
4.0-4.5	0.298290	0.001020	0.700300
4.5-5.0	0.276140	0.000960	0.720620
5.0-5.5	0.257930	0.000960	0.735970
5.5-6.0	0.242850	0.000970	0.747180
6.0-7.0	0.224890	0.001020	0.757830
7.0-8.0	0.207620	0.001330	0.764110
8.0-10.0	0.192440	0.002760	0.765180

Table 11: Prompt and non-prompt distribution fractions. σ_{f_i} is the uncertainty on f_i based on the uncertainties of the parameterization for the non-prompt Gaussian width. Except for the longer decay times, $f_i + g_i = 1$.

Time Bin	RS In	RS Out	RS Prompt	RS Non-P
0.75- 1.00	28210.6 ± 175.8	1308.8 ± 42.7	28891.9 ± 245.2	627.5 ± 178.5
1.00- 1.25	115411.5 ± 353.6	3850.3 ± 69.4	121950.8 ± 454.8	-2689.1 ± 274.2
1.25- 1.50	236941.8 ± 518.3	7963.1 ± 98.9	249420.4 ± 657.2	-4515.5 ± 383.7
1.50- 1.75	318898.3 ± 607.5	12590.5 ± 122.5	332465.4 ± 755.1	-976.6 ± 431.1
1.75- 2.00	344792.1 ± 612.3	16034.5 ± 138.7	356634.1 ± 750.5	4192.5 ± 431.0
2.00- 2.25	330956.6 ± 615.9	18228.1 ± 147.6	339868.0 ± 735.3	9316.8 ± 405.4
2.25- 2.50	298370.6 ± 577.9	19823.3 ± 154.5	304025.7 ± 679.1	14168.2 ± 375.8
2.50- 2.75	254941.6 ± 545.2	20934.6 ± 157.8	257344.8 ± 628.9	18526.7 ± 343.9
2.75- 3.00	215041.3 ± 485.2	21142.7 ± 158.7	215341.0 ± 554.4	20835.5 ± 315.0
3.00- 3.25	177174.8 ± 445.7	20848.8 ± 157.9	175889.6 ± 502.8	22116.0 ± 289.6
3.25- 3.50	144086.8 ± 397.7	20413.6 ± 156.0	141592.8 ± 445.1	22869.9 ± 268.3
3.50- 3.75	117307.1 ± 359.4	19516.4 ± 152.5	114208.1 ± 399.1	22553.8 ± 249.1
3.75- 4.00	93238.5 ± 339.4	18412.2 ± 147.8	89717.1 ± 373.1	21840.7 ± 231.8
4.00- 4.50	134778.1 ± 385.3	33780.4 ± 200.5	127254.7 ± 428.5	41017.0 ± 307.6
4.50- 5.00	85247.5 ± 306.7	28894.4 ± 184.7	78069.9 ± 338.0	35668.5 ± 268.2
5.00- 5.50	53613.2 ± 243.1	23504.9 ± 167.1	47511.8 ± 266.2	29141.0 ± 233.7
5.50- 6.00	33100.3 ± 191.3	18405.0 ± 147.6	28221.9 ± 208.1	22823.8 ± 200.7
6.00- 7.00	31792.1 ± 191.8	23771.1 ± 167.5	25474.4 ± 208.5	29342.3 ± 223.1
7.00- 8.00	10706.6 ± 109.9	12012.1 ± 119.4	7550.6 ± 119.2	14766.2 ± 154.7
8.00-10.00	4542.3 ± 73.0	8464.1 ± 100.5	2403.6 ± 85.4	10352.3 ± 131.1

Table 12: RS D^* numbers.

Time Bin	WS In	WS Out	WS Prompt	WS Non-P
0.75- 1.00	92.8 ± 23.6	0.0 ± 7.5	107.9 ± 35.3	-15.0 ± 29.9
1.00- 1.25	347.6 ± 46.7	29.6 ± 13.6	328.9 ± 60.3	48.4 ± 43.2
1.25- 1.50	1013.4 ± 74.3	12.9 ± 17.9	1101.8 ± 87.5	-75.6 ± 48.3
1.50- 1.75	1197.9 ± 86.4	89.9 ± 22.7	1191.6 ± 99.5	96.2 ± 53.9
1.75- 2.00	1230.5 ± 88.6	31.2 ± 23.1	1302.0 ± 99.8	-40.4 ± 49.8
2.00- 2.25	1425.1 ± 88.0	68.3 ± 25.1	1473.3 ± 98.2	20.1 ± 50.1
2.25- 2.50	1214.7 ± 79.2	64.1 ± 26.7	1251.9 ± 88.2	26.9 ± 49.7
2.50- 2.75	1004.5 ± 79.5	36.3 ± 26.3	1049.0 ± 87.5	-8.2 ± 46.7
2.75- 3.00	953.0 ± 73.0	83.5 ± 27.9	961.4 ± 80.3	75.1 ± 47.2
3.00- 3.25	684.4 ± 63.3	52.3 ± 28.4	697.1 ± 69.8	39.6 ± 46.3
3.25- 3.50	572.5 ± 58.7	58.8 ± 26.7	575.4 ± 64.3	55.7 ± 42.2
3.50- 3.75	602.4 ± 52.9	98.9 ± 28.4	587.2 ± 58.2	113.8 ± 43.5
3.75- 4.00	551.8 ± 50.9	125.7 ± 28.6	522.7 ± 55.8	154.3 ± 42.9
4.00- 4.50	579.5 ± 56.5	100.9 ± 40.0	567.1 ± 62.4	112.1 ± 58.1
4.50- 5.00	402.9 ± 50.9	50.7 ± 33.9	403.6 ± 55.3	48.1 ± 47.7
5.00- 5.50	244.0 ± 35.8	92.6 ± 30.1	221.6 ± 39.1	112.9 ± 41.2
5.50- 6.00	175.0 ± 29.2	34.8 ± 26.3	170.3 ± 31.6	37.0 ± 35.2
6.00- 7.00	179.2 ± 31.1	112.5 ± 32.6	150.2 ± 33.5	137.3 ± 42.6
7.00- 8.00	46.8 ± 18.1	53.3 ± 23.8	32.8 ± 19.5	65.5 ± 30.3
8.00-10.00	35.4 ± 13.4	50.8 ± 20.8	22.5 ± 14.3	61.7 ± 26.0

Table 13: WS D^* numbers.

References

[1] BaBar Collaboration, B. Aubert *et al.*, *Search for $D^0 - \overline{D}^0$ Mixing and a Measurement of the Doubly Cabibbo-suppressed Decay Rate in $D^0 \to K\pi$ Decays*, Phys. Rev. Lett. **91**, 171801 (2003), hep-ex/0304007.

[2] CDF Collaboration, A. Abulencia *et al.*, *Measurement of the Ratio of Branching Fractions $\mathcal{B}(D^0 \to K^+\pi^-)/\mathcal{B}(D^0 \to K^-\pi^+)$ using the CDF II Detector*, Phys. Rev. D **74**, 031109 (R) (2006).

[3] CLEO, R. Godang et al., *Search for $D^0 - \bar{D}^0$ Mixing*, Phys. Rev. Lett. **84**, 5038 (2000), hep-ex/0001060.

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

[5] Belle Collaboration, L. M. Zhang et al., *Improved Constraints on D^0-$\bar{D}^0$ Mixing in $D^0 \to K^+\pi^-$ Decays from the Belle detector*, Phys. Rev. Lett. **96**, 151801 (2007), hep-ex/0601029.

[6] Marko Starič, *Search for $D^0 - \overline{D}^0$ mixing at Belle and BaBar*, talk at XLII Rencontres de Moriond, La Thuile, Italy (13 March, 2007).

[7] BaBar Collaboration, Kevin Flood, *Evidence for $D^0 - \overline{D}^0$ Mixing*, talk at XLII Rencontres de Moriond, La Thuile, Italy (13 March, 2007).

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

[9] Belle Collaboration, K. Abe *et al.*, *Measurement of $D^0 - \overline{D}^0$ mixing in $D^0 \to K^0_s\pi^+\pi^-$ decays*, arXiv:0704.1000v1 (7 April, 2007).

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

[11] S. D'Auria et al., *Relative Branching Fractions and CP-violating Decay Rate Asymmetries in Cabibbo Suppressed Decays of the D^0 Meson*, CDF-Note 6391, 2004.

[12] Shin-Shan. Yu et al., *COT dE/dx Measurement and Corrections*, CDF-Note 6361, 2003.

[13] Saverio D'Auria et al., *Track-based calibration of the COT specific ionization*, CDF-Note 6932, 2004.

[14] Li Jin, *Search for $D^0 - \overline{D}^0$ mixing in $D^0 \to K^+\pi^-$ decay*, PhD thesis, University of Science and Technology of China, 5 June, 2004.

[15] I. K. Furic et al., *$D^0 \to \mu\mu$ Update*, PSP Meeting, March 27, 2007.

[16] http://www cdf.fnal.gov/physics/statistics/statistics_recommendations.html, in particular, 'What is Systematics' by Giovanni Punzi, and 'Systematic Uncertainties' by Craig Blocker.

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

[18] The CDF Statistics Committee, *Recommendations Concerning Limits*, CDF-Note 7739, 2005.