Search for $B \to K\nu\overline{\nu}$ Decays With the Semileptonic Tag

by

Carl Vuosalo

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
(Physics)

at the

University of Wisconsin – Madison

2009

Version # 4

Search for $B \to K\nu\bar{\nu}$ Decays With the Semileptonic Tag

Carl Vuosalo

Under the supervision of Professor Sridhara Dasu

At the University of Wisconsin — Madison

Abstract

Flavor changing neutral-current transitions such as $b \to s\nu\bar{\nu}$ are absent at tree level in the Standard Model and can only occur via loop diagrams. Several new physics models may enhance the rate of these transitions. This document presents searches for the exclusive decays $B_u^+ \to K^+\nu\bar{\nu}$ and $B_d^0 \to K_S^0\nu\bar{\nu}$, which have a predicted theoretical branching fraction of $(3.8^{+1.2}_{-0.6}) \times 10^{-6}$. The presence of two neutrinos in the final state makes recognition of the signal challenging, so the full reconstruction of one B meson in the semileptonic decay channel $B \to D^{(*)}l\nu$ is used to facilitate the search for the signal in the recoiling B. This analysis uses approximately 420 fb^{-1} or 460 million $B\bar{B}$ pairs collected over runs 1-6 with the *BABAR* detector at the PEP-II B factory. This analysis finds 90%-confidence-level upper limits on the branching fractions of 1.3×10^{-5} for $B_u^+ \to K^+\nu\bar{\nu}$, 5.6×10^{-5} for $B_d^0 \to K^0\nu\bar{\nu}$, and the first upper limits on the partial branching fractions for $B_u^+ \to K^+\nu\bar{\nu}$ of 3.1×10^{-5} for K^+ CMS momentum < 1.5 GeV/c and of 0.89×10^{-5} for K^+ CMS momentum > 1.5 GeV/c. These results improve upon the previous best upper limits, which came from the Belle experiment, of 1.4×10^{-5} for $B_u^+ \to K^+\nu\bar{\nu}$ and 16×10^{-5} for $B_d^0 \to K^0\nu\bar{\nu}$. They also rule out a new-physics model of scalar dark matter for scalar particle masses below 1.8 GeV/c^2.

Sridhara Dasu (Adviser)

Contents

List of Figures

List of Tables

1   Introduction

The Standard Model (SM) [1] of particle physics provides the foundation for understanding the interactions of particles and fields. It describes three forces: electromagnetism, the strong force that gives stability to atomic nuclei, and the weak force that mediates radioactive decay. These forces are carried by particles: photons carry electromagnetism, gluons carry the strong force, and charged W bosons and neutral Z bosons carry the weak force. Two additional groups of particles, the leptons and quarks, are organized into pairs that come in three generations. The first generation of leptons includes the familiar electron and its partner, the electron neutrino. The latter two generations of charged leptons, the muon and the tau, have higher mass and are unstable. They both have partner neutrinos. Neutrinos are different from other leptons and quarks in that they only interact through the weak force, which means they are nearly undetectable, except by large, specialized neutrino detectors. The first generation of quarks includes the stable up (u) and down (d) quarks. The latter two generations are more massive and unstable: they are the charm (c) and strange (s) quarks, and the top (t) and bottom (b) quarks. All the fundamental particles of the SM also have opposites called antiparticles that have the same mass but otherwise have opposite properties. Antiparticles are denoted by a bar over the particle symbol.

Quarks are bound to each other by the strong force, which increases with distance so that lone quarks are never found experimentally. Instead, pairs of quarks form mesons, which are always unstable, and trios of quarks form baryons, some of which are stable, like the proton. A meson is always composed of a quark and an anti-quark of some type. The mesons most relevant to this dissertation are the $\Upsilon(4S)$, composed of b and $\bar{b}$, the B^+ meson, composed of $\bar{b}$ and u, the B^0 meson, composed of $\bar{b}$ and d, the K^+ meson, composed of $\bar{s}$ and u, and the K^0 meson, composed of $\bar{s}$ and d. Replacing each quark in a meson with its antiparticle

(or vice-versa) creates the anti-meson, though, in some cases like the $\Upsilon(4S)$, the meson and anti-meson are identical.

Quarks can change type via the weak interaction. The parameters that govern these transitions can be placed in a 3x3 unitary matrix called the Cabibbo-Kobayashi-Maskawa matrix [2]. These values must be determined experimentally since they are not predicted by the SM.

A powerful tool to help determine the rate of an interaction is the Feynman diagram [3]. The total interaction rate can be calculated, in principle, by summing all possible Feynman diagrams for the interaction. In loop diagrams, like those in Figure 1, all possible particles that could propagate through the loops should be considered. In the SM, there is a limited set of particles that can be found in the loops. However, if the SM is incomplete, if there are more particles beyond the SM, then these new physics particles could also appear in the loops, and interaction rates calculated under the SM might not match those observed in experiments.

In fact, there are many indications that the SM is incomplete. Neutrinos are predicted to be massless by the SM, but there is now compelling evidence neutrinos have small masses [4]. In addition, there is a lot of evidence of dark matter in the universe, but no particles in the SM can account for dark matter [5]. Furthermore, there is the general problem that the SM not only does not include gravity, and but it is also actually incompatible with general relativity. All these indications point to the need to look for new phyiscs beyond the SM.

The BABAR experiment [6] at the SLAC National Accelerator Laboratory utilized the PEP-II accelerator and BABAR detector to create a large data set of B meson decays. For most of the accelerator's run, it was tuned to produce the $\Upsilon(4S)$ resonance, which decays almost entirely into $B\bar{B}$ pairs. This large data set provides the opportunity to search for very rare decays of the B meson. One such rare decay is $B \to K\nu\bar{\nu}$, which is an attractive

target of study due to its theoretical simplicity in the Standard Model. An enhancement in the rate of this decay could be an indication of new physics. In addition, since the signature of this decay is a kaon with missing energy, the search for it could also capture decays that produce undetected dark matter particles. This thesis documents the search, using the full *BABAR* data set, for the exclusive decays $B_u^+ \to K^+ \nu \bar{\nu}$ and $B_d^0 \to K_S^0 \nu \bar{\nu}$ with the use of a B-tagging technique where the second B of the pair is reconstructed as a semileptonic decay.

As a flavor-changing neutral current (FCNC) decay, $B \to K \nu \bar{\nu}$ cannot occur at tree level in the SM but requires at least one-loop diagrams like the penguin and box diagrams shown in Figure 1. It is one of the theoretically cleanest decays in the field of FCNC processes, which provide powerful insights into the flavor dynamics of the SM, such as quark masses and mixing. The loops in the decay diagrams include quarks, leptons, and bosons, making these decays sensitive to many SM parameters.
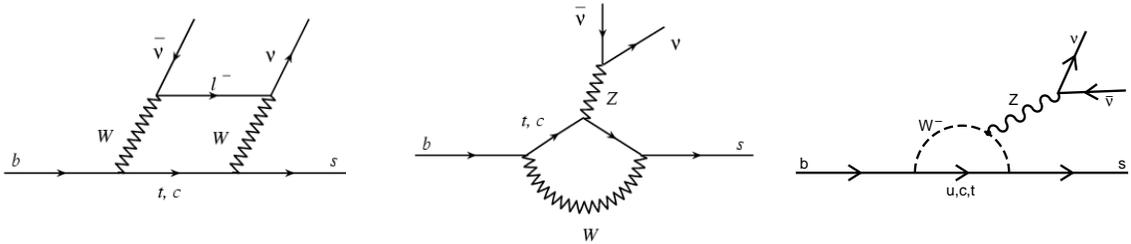
Figure 1: Lowest-order Standard Model diagrams for $b \to s \nu \bar{\nu}$.

Predictions for the branching fraction of $B \to K \nu \bar{\nu}$, according to two recent theoretical models, BHI(2001) [7] and ABSW(2009) [8], are:

$$\mathcal{B}(B \to K \nu \bar{\nu}) = (3.8^{+1.2}_{-0.6}) \times 10^{-6} \ [7] \tag{1}$$

$$\mathcal{B}(B \to K \nu \bar{\nu}) = (4.5 \pm 0.7) \times 10^{-6} \ [8] \tag{2}$$

This rate is the rate for the sum of all three neutrino flavors, and it is predicted to be

the same for both $B_u^+ \to K^+ \nu\overline{\nu}$ and $B_d^0 \to K^0 \nu\overline{\nu}$.

The current best experimental upper limits for these decays are (at the 90% confidence level):

$$\mathcal{B}(B_u^+ \to K^+ \nu\overline{\nu}) < 1.4 \times 10^{-5} \quad [9] \tag{3}$$

$$\mathcal{B}(B_d^0 \to K^0 \nu\overline{\nu}) < 16 \times 10^{-5} \quad [9] \tag{4}$$

These results derive from using a hadronic B tag on 535 million $B\bar{B}$ events in the Belle experiment.

Another FCNC decay is $K \to \pi\nu\overline{\nu}$, which is also very clean theoretically, is the target of two proposed experiments: K^0TO for $K_L \to \pi^0\nu\overline{\nu}$ [10] and NA62 for $K^+ \to \pi^+\nu\overline{\nu}$ [11]. The Feynman diagrams for $K \to \pi\nu\overline{\nu}$ are nearly identical to those for $B \to K\nu\overline{\nu}$, but, with a branching fraction on the order of 10^{-10}, this mode presents experimental challenges. Since the physics of the two modes are nearly the same, the search for $B \to K\nu\overline{\nu}$ provides a promising additional avenue for exploring FCNCs in the SM by using existing data sets while the new $K \to \pi\nu\overline{\nu}$ experiments are still in the process of being approved and constructed.

2   Theory

A B meson decay involves interactions at two energy scales: the short distance, perturbative, electro-weak interaction of the decaying quark, and the long-distance, non-pertubative, quantum chromodynamics (QCD) interactions. Different theoretical approaches are required for each energy scale. The Operator Product Expansion (OPE) conveniently conveniently separates the effects at each scale [12].

An effective Hamiltonian can be constructed using the OPE. For $B \to K\nu\bar{\nu}$, the Hamiltonian can be written as:

$$H_{eff} = \frac{G_F \alpha}{2\pi\sqrt{2}} V_{tb} V_{ts}^* C_{10}^\nu \bar{s}\gamma^\mu(1-\gamma^5)b\bar{\nu}\gamma_\mu(1-\gamma^5)\nu + h.c. \;\; [13] \tag{5}$$

G_F is the Fermi constant, α is the fine structure constant, $V_{tb}V_{ts}^*$ are the CKM matrix elements for the quark transitions in the decay, and C_{10}^ν is the short-distance Wilson coefficient related to the weak interaction. The remaining terms can be seen as a four-fermion operator, which is the local operator representing long-distance, non-perturbative, QCD effects.

C_{10}^ν is also the Wilson coefficient for $K \to \pi\nu\bar{\nu}$ decays, which shows the close relation between $B \to K\nu\bar{\nu}$ and $K \to \pi\nu\bar{\nu}$. C_{10}^ν is sometimes labelled C_L^ν or X (when multiplied by a constant).

From the Hamiltonian, the di-neutrino-invariant-mass decay spectrum can be derived. It can be written as:

$$\frac{d\Gamma}{ds} = \frac{G_F^2 \alpha^2 m_B^5}{2^8 \pi^5} \left|V_{tb}V_{ts}^*\right|^2 \left|C_{10}^\nu\right|^2 \lambda^{3/2}(r_K, s) f_+^2(s) \;\; [13] \tag{6}$$

$$\lambda \equiv 1 + r_K^2 + s^2 - 2s - 2r_K(1+s), \;\; s = m_{\nu\bar{\nu}}^2/m_B^2, \;\; r_K = m_K^2/m_B^2 \tag{7}$$

m_B is the mass of the B meson, m_K is the mass of the kaon, λ represents the kinematic terms, and f_+ is the form factor, which represents long-distance QCD effects.

As seen in Equations 1 and 2, the ABSW and BHI models have slightly different predictions for the $B \to K\nu\overline{\nu}$ branching fraction because they use different form factor models. The form factors are calculated with light-cone sum rules, but the two models use the work of different theorists; ABSW uses the work of [14], while BHI uses [15]. The models differ in the distribution of the differential branching fraction across values of $m_{\nu\overline{\nu}}^2$ – the ABSW model predicts more events with high values of $m_{\nu\overline{\nu}}^2$, as shown in Figure 2.
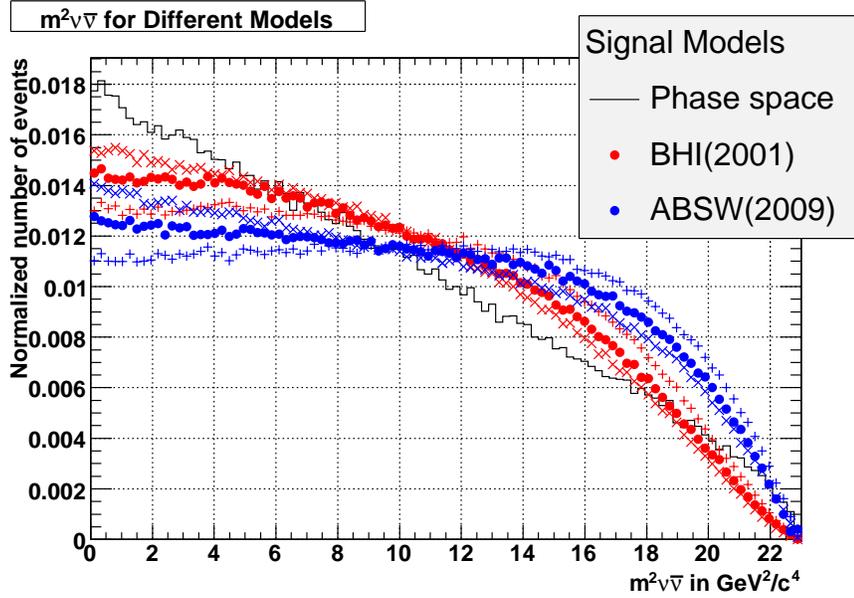
Figure 2: The di-neutrino invariant mass squared for the three signal models. For the two theoretical models, the theoretical uncertainties are shown by X's and plusses of the matching color. The theoretical uncertainties are due to uncertainties in the form factor models. Note that low values of the di-neutrino invariant mass squared correspond to high values of the signal kaon momentum, and vice-versa. See Section 4.1.2 for a description of the phase-space model.

The loops in the decay diagrams for $B \to K\nu\overline{\nu}$ (Figure 1) provide opportunities for the appearance of new physics, because new particles could occur in the loops and thereby alter

the branching fraction from the SM prediction. New physics models that could affect this decay include unparticle models [16], MSSM at large $\tan \beta$ [17], models with a single universal extra dimension [18], WIMP-less dark matter [19], and scalar WIMP dark matter [20]. This last model could increase the branching fraction by up to two orders of magnitude, so it is already constrained by current experimental limits, and stricter experimental limits will further constrain it and the other new physics models.

Other decays similar to $B \to K \nu \bar{\nu}$ also can place limits on new physics but lack some of the advantages of $B \to K \nu \bar{\nu}$. For example, $B \to K \ell^+ \ell^-$ is very similar to $B \to K \nu \bar{\nu}$, but it has three Wilson coefficients combined in its decay rate, which makes it difficult experimentally to determine their individual values. Also, there are irreducible charmonium-resonance backgrounds that overwhelm significant parts of the decay's q^2 distribution. On the other hand, the charged leptons from the decay leave clear tracks in a detector, which makes it easier to measure the branching fraction. In fact, the decay has been observed even though its branching fraction is ten times less than that of $B \to K \nu \bar{\nu}$, and the measurement is in agreement with the SM prediction:

$$\mathcal{B}(B \to K \ell^+ \ell^-) = (0.394^{+0.073}_{-0.069} \pm 0.020) \times 10^{-6} \ \ [21] \tag{8}$$

In comparison, $B \to K \nu \bar{\nu}$ is very difficult to measure experimentally but has the advantage of theoretical cleanness and simplicity, with only one Wilson coefficient contributing to the decay rate, and with no resonances interfering with the main decay. New physics signals might show up more clearly in $B \to K \nu \bar{\nu}$ than in others like $B \to K \ell^+ \ell^-$.

3 PEP-II and the *BABAR* Detector

The PEP-II collider and the *BABAR* detector are designed for the study of CP violation and rare decay processes in the B meson system and for measuring CKM parameters. CP violation is the breaking of a usual physical symmetry. Physical laws are invariant under certain transformations, like rotation. Charge conjugation (C) is the replacement of a particle with its antiparticle (or vice-versa), and the parity transformation (P) is the replacement of a right-handed particle with a left-handed one (or vice-versa). Together they are a symmetry of most particle interactions; however, certain interactions violate CP. Studying CP violation can shed light on the mystery of large CP violation and baryon number violation in the early universe that was necessary to produce the predominance of matter over antimatter in the universe today.

3.1 PEP-II

PEP-II is an asymmetric e^+e^- collider, which collides 9.0 GeV electrons and 3.1 GeV positrons from its high-energy (HER) and low-energy (LER) storage rings. These rings are fed by the SLAC Linac. The beams are precisely tuned to produce a center of mass energy equal to the mass of the $\Upsilon(4S)$ particle (10.58 GeV/c^2), which has a branching fraction to B meson pairs of nearly 100% [22]. The asymmetric beam energies cause the $\Upsilon(4S)$ system to be Lorentz-boosted by a factor $\beta\gamma = 0.56$, which is important for studying time-dependent CP violation, but much less so for analyses of rare decays.

The Lorentz boost allows measurement of the time difference in the decays to certain final states of the two B's produced by the $\Upsilon(4S)$. Since the center-of-mass system (CMS) is moving in the lab frame, this time difference creates a separation between the decay vertices along the longitudinal axis of the detector, and measurement of this separation allows calculation of an important CKM parameter and also also indicates the amount of

CP violation.

Another important feature of PEP-II is the high number of collisions (also called luminosity) it has been able to produce. This high luminosity has been enabled by the high currents produced by the two-mile-long linear accelerator and the separate beamlines for the electron and positron beams. PEP-II achieved a peak luminosity of $1.2 \times 10^{34} cm^{-2} s^{-1}$, four times its design luminosity, and a total luminosity of 557 fb^{-1} over the full lifetime of the experiment. Table 1 shows the luminosity and number of $B\bar{B}$ pairs used for this analysis. In the search for a very rare decay like $B \to K\nu\bar{\nu}$, this high luminosity has been essential.

3.2 The *BABAR* detector

BABAR [6] has many design features important to this $B \to K\nu\bar{\nu}$ analysis: hermeticity, ability to tolerate high beam luminosity, and high accuracy of particle-energy reconstruction and kaon identification. It is composed of five sub-detectors: the Silicon Vertex Tracker (SVT), the Drift Chamber (DCH), the Detector of Internally Reflected Cherenkov Light (DIRC), the Electromagnetic Calorimeter (EMC), and the Instrumented Flux Return (IFR), all shown in Figures 3 and 4.

3.2.1 Silicon Vertex Tracker

The Silicon Vertex Tracker (SVT) is a five-layer, double-sided silicon detector. It provides vertexing information and tracking and pattern recognition. The geometry of the layers is shown in Figure 5, with a design intended to maximize the angular acceptance. Figure 6 shows a side view of the SVT.

The inner three layers have a hit resolution in z of about 15 μm, while the outer layers have a resolution of 40 μm, which compares favorably to the $\sim 250 \, \mu$m mean Δz separation between the decay vertices of the two B mesons produced in an event.
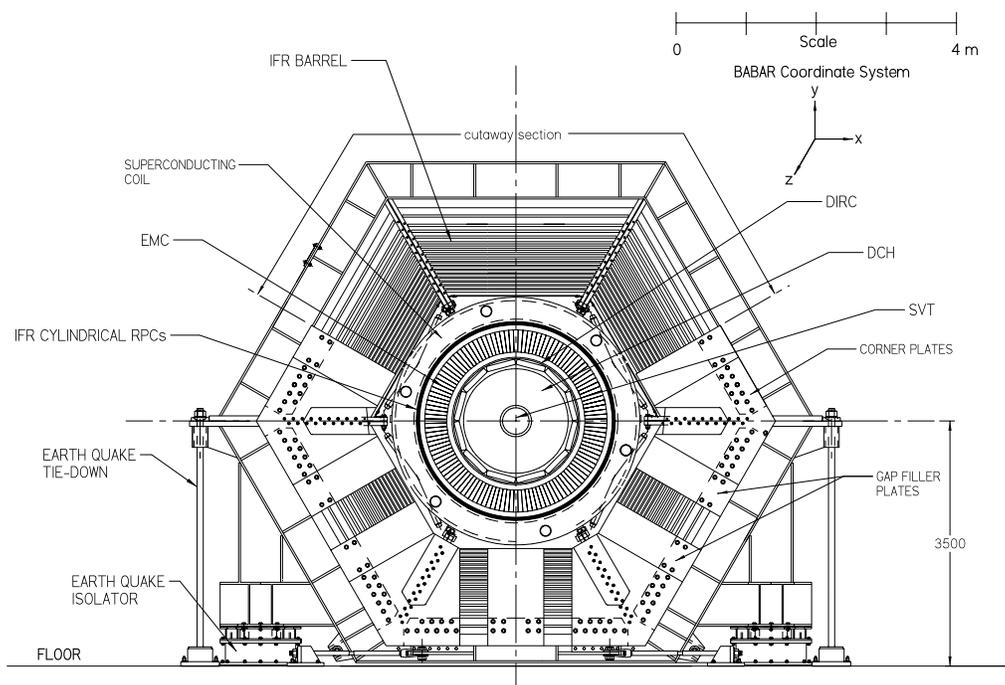
Figure 3: End view of the *BABAR* detector.

3.2.2 Drift Chamber

The drift chamber (DCH) measures particle momentum and energy loss (dE/dx). Its track-finding efficiency, averaged over momentum and angle, is about 96%, relative to the tracks found by the SVT. Figure 7 shows how dE/dx can distinguish particle types. The DCH has forty layers and is filled with a combination of helium and isobutane gas. The inert helium, being the second lightest element, minimizes energy loss in the DCH so even lower-energy particles can get through and reach the outer detectors. The basic unit of the DCH is a hexagonal cell composed of a central sense wire surrounded by six ground wires, and there are 7104 such cells in total. About half of the cells are oriented at angles to the z-axis in order to provide longitudinal tracking information. Figure 8 shows the inside of the DCH.
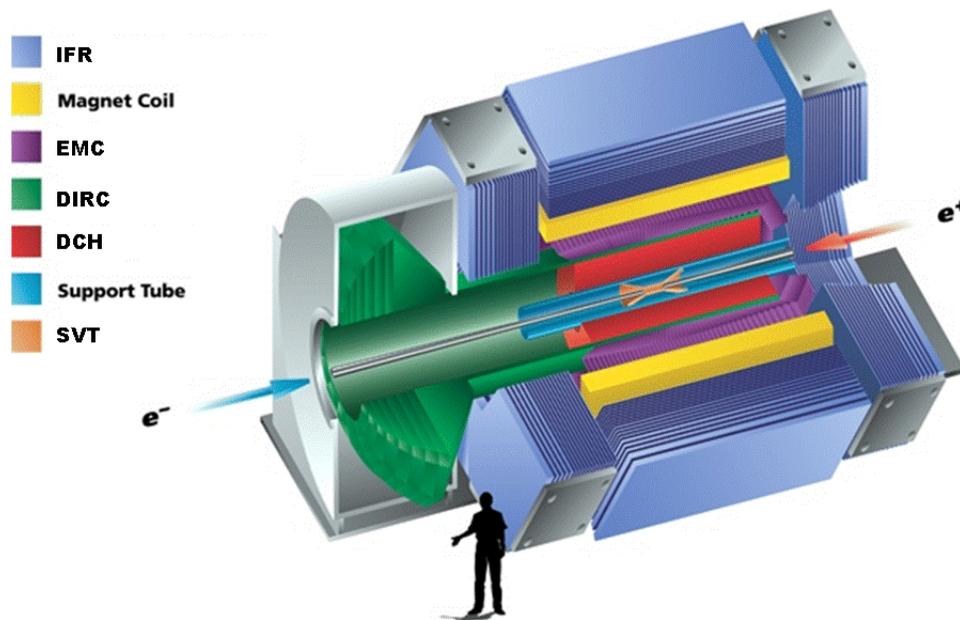
Figure 4: Layout of the *BABAR* detector.
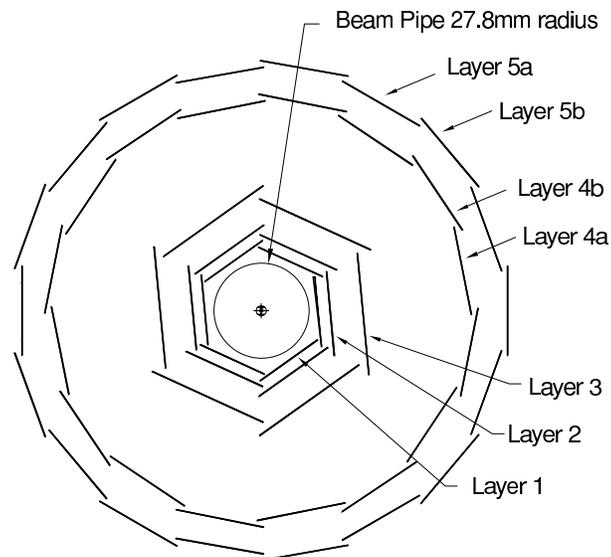
Figure 5: Layout of layers in the SVT.

3.2.3   Detector of Internally Reflected Cherenkov Light

The detector of internally reflected Cherenkov light (DIRC) measures the Cherenkov angle, track angle, and momentum of charged particles to determine their mass. Total internal
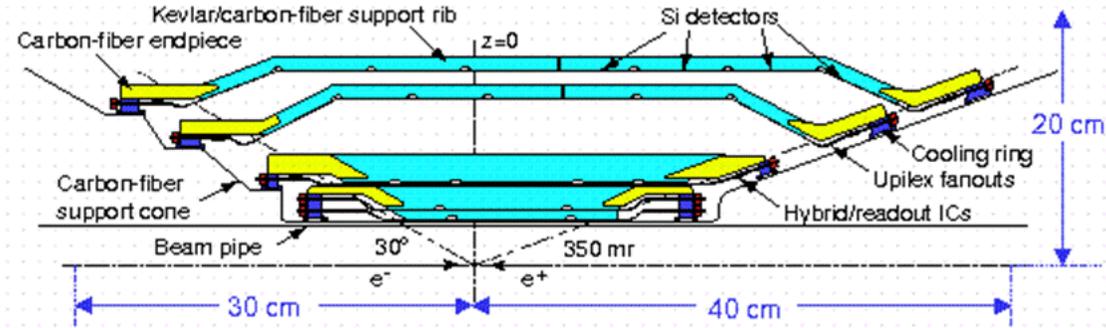
Figure 6: Side view of the SVT.
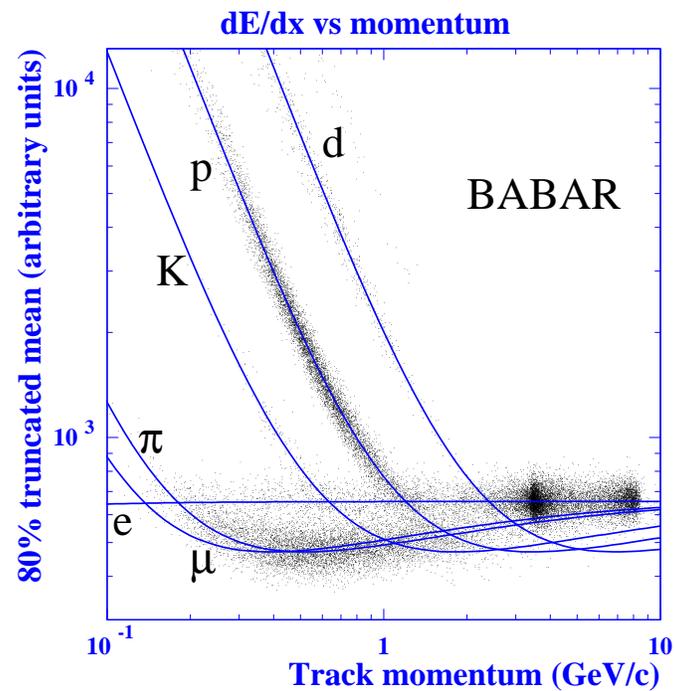
Figure 7: Energy loss in the DCH vs. track momentum.

reflection in the quartz crystals of the detector sends the Cherenkov rings to photo-multiplier tubes, as shown in Figure 9. The Cherenkov angle θ_C, defined in Equation 9, is combined with the track momentum and angle to help identify the particle. These measurements
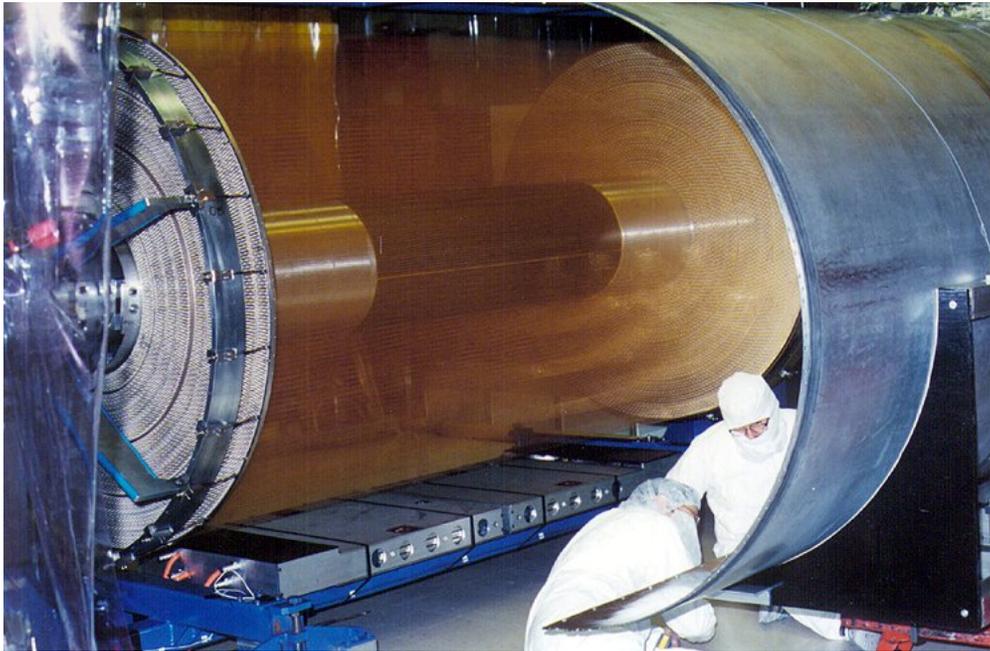
Figure 8: View of the DCH during assembly.

are particularly effective at separating charged kaons from pions at lab momenta above 0.7 GeV/c, as illustrated in Figure 10. This momentum range is where kaons from $B \rightarrow K\nu\bar{\nu}$ are expected to be found, so the DIRC makes an important contribution to this analysis.

$$cos(\theta_C) = \frac{1}{\beta n} \tag{9}$$

where $\beta = v/c$ of the particle and $n = 1.473$, the index of refraction of quartz.

3.2.4 Electromagnetic Calorimeter

The electromagnetic calorimeter (EMC) measures the energy of particles. Figure 11 shows the energy resolution of the EMC. Its channels are sensitive to energies down to 1 MeV, and 96% of photons above 20 MeV are detected. This combination of low noise and high sensitivity is crucial to the search for $B \rightarrow K\nu\bar{\nu}$ since the key characteristic of this decay is the absence of any remaining energy in the detector after all expected particles have
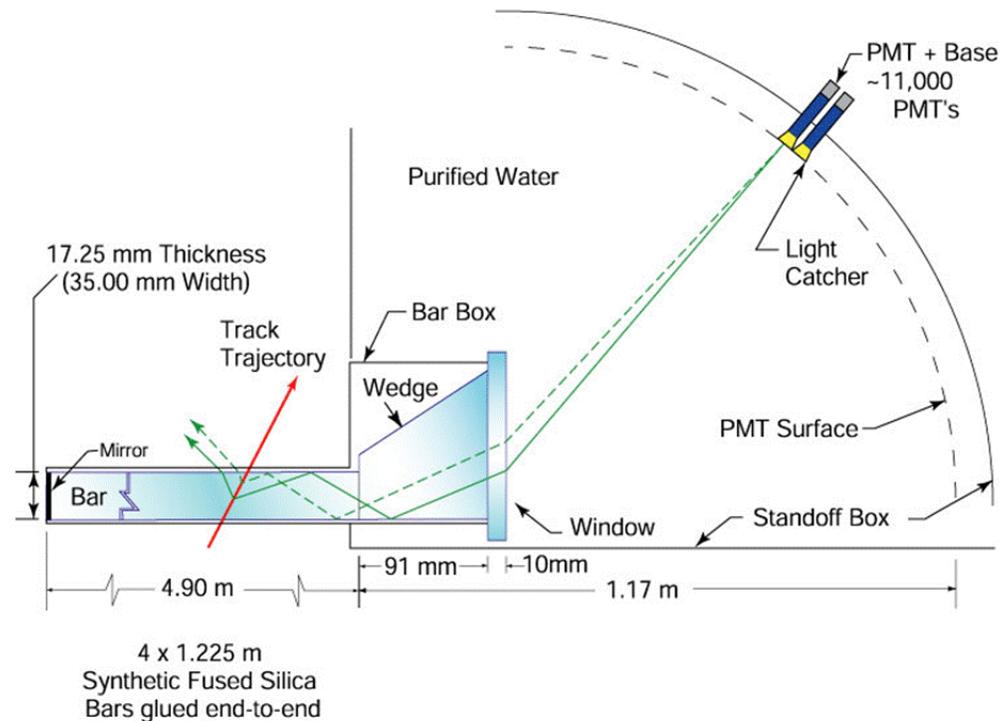
Figure 9: Geometry of the DIRC.

been reconstructed.

The EMC is composed of 6580 thallium-doped cesium iodide crystals (see Figure 12) connected to photodiodes, and the scintillation showers in the crystals help identify particles. Shower shape and energy deposition in the EMC distinguish photons, electrons, and hadrons, and allow the reconstruction of neutral pions.

3.2.5 Instrumented Flux Return

The instrumented flux return (IFR), shown in Figure 13, is designed to identify muons and neutral hadrons. It contains resistive plate chambers (RPCs) and limited streamer tubes (LSTs) interleaved with steel layers to detect the passage of muons and the showers created by hadron collisions in the steel. The outputs from the IFR, along with the other detectors,
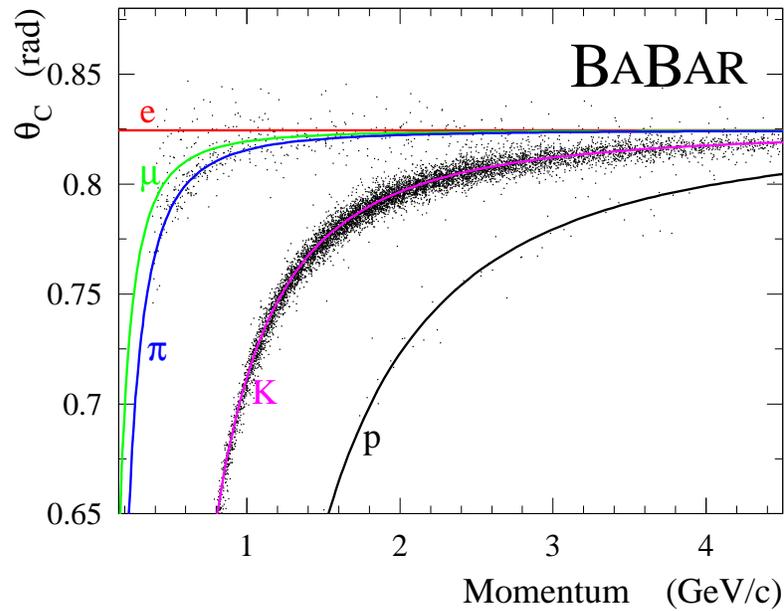
Figure 10: Cherenkov angle θ_C vs momentum for kaons (points). The solid lines correspond to the expected θ_C for other charged particle species.
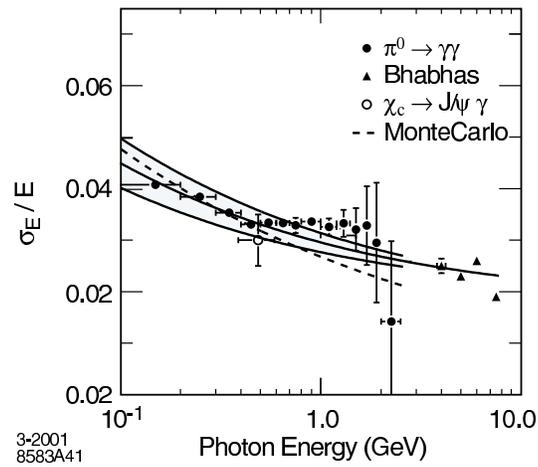
Figure 11: EMC energy resolution vs. photon energy.
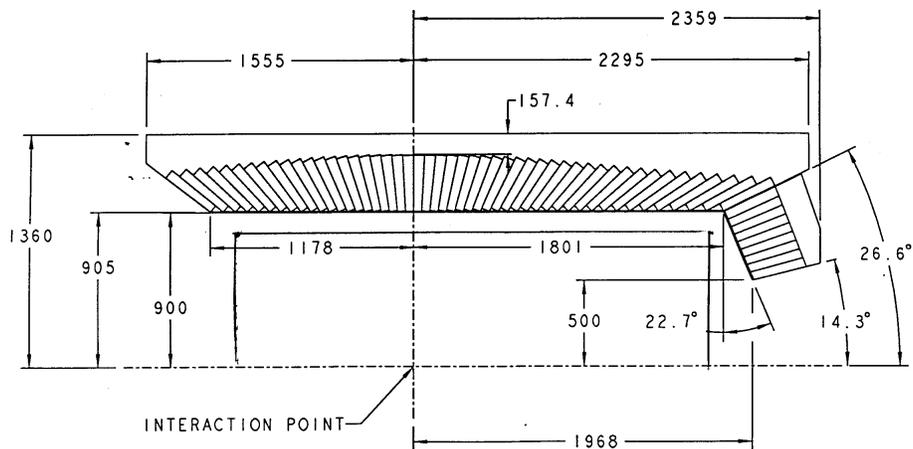
Figure 12: Side view of the crystal geometry in the electromagnetic calorimeter.
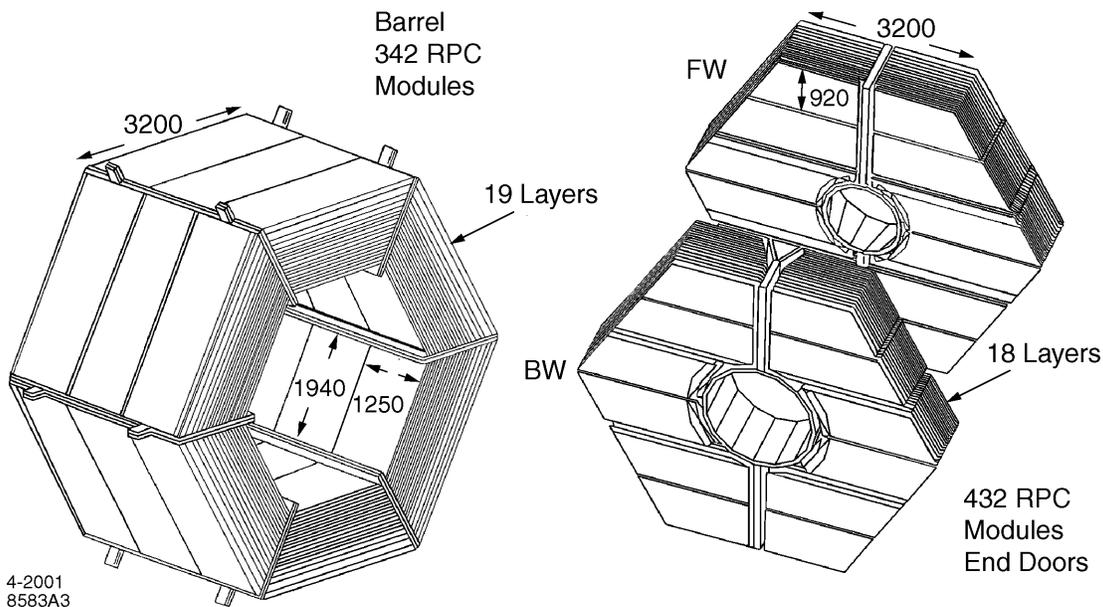
Figure 13: Geometry of the instrumented flux return.

are used by muon selectors to recognize muon tracks. The performance of two *BABAR* muon selectors in shown in Figure 14.
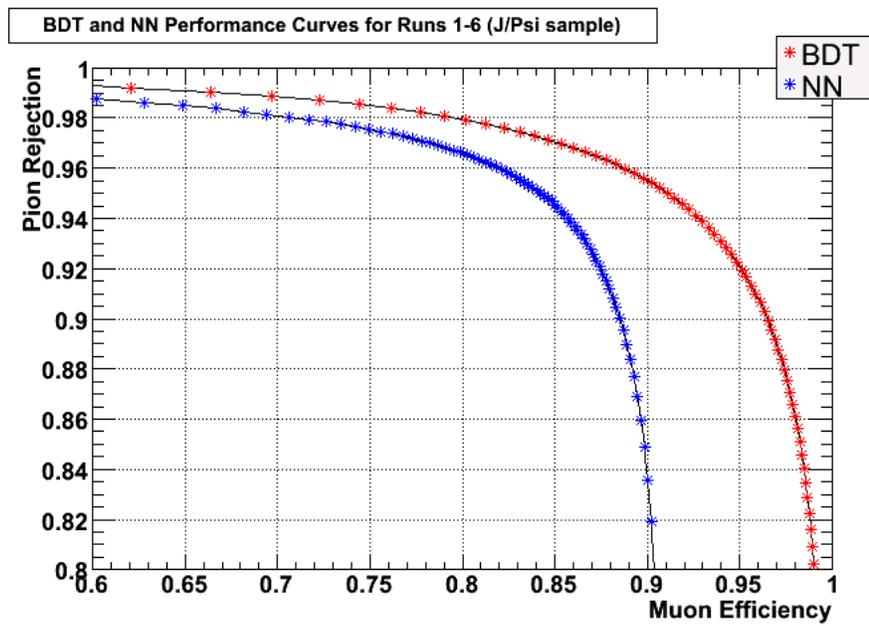
Figure 14: Performance of the muon selectors that use many IFR output variables. The right curve (red stars) is the for decision tree muon selector, and the left curve (blue stars) is for the older neural network selector. Their performance for muon identification and pion rejection is assessed using a control sample of J/Ψ events.

4   Event Selection

4.1   Event Samples

4.1.1   *BABAR* Data Samples

The full *BABAR* data set of runs 1-6 (release 22d) is used in this analysis. This data set contains 459 million $B\overline{B}$ pairs (see Table 1).

Table 1: Number of $B\overline{B}$ decays and integrated luminosity in the real data sample by run. These values are for R22dv06.

Run	No. $B\overline{B}$ (millions)	On-res $\int \mathcal{L}$ (fb^{-1})
1	22.4	20.4
2	67.4	61.1
3	35.6	32.3
4	110.4	100.3
5	147.2	133.3
6	76.0	70.3
Total	459.0±5.1	417.7

4.1.2   Simulated Data Samples

The Monte Carlo (MC) simulated event samples (Tables 2-3) used here are from the SP8 production. The MC is generated by the EvtGen software package [23]. Six distinct classes of MC simulated events are relevant to this analysis:

- $B_u^+ \to K^+ \nu\overline{\nu}$ signal events, which are generated with a phase-space model for the charged kaon momentum. This simple model imposes no structure upon the distribution of daughter particle momenta other than that required by basic conservation laws.

- $B_d^0 \to K_S^0 \nu\overline{\nu}$ signal events, which are also generated with a phase-space model for the neutral kaon momentum. This signal MC includes both $K_S^0 \to \pi^+\pi^-$ and $K_S^0 \to \pi^0\pi^0$.

Table 2: Number of MC simulated signal events and the ratio of the number of $B\bar{B}$ decays in data to the number simulated.

Mode (SP)	$\mathcal{B}(/10^{-6})$	Events	Data/MC ($/10^{-3}$)
$K^+\nu\bar{\nu}(2227)$	3.8 [7]	7845000	0.22
$K_S^0\nu\bar{\nu}(3152)$	1.9 [7]	3943000	0.22
$K^+J/\Psi(989)$	120	13431000	0.41
$K^{*+}\nu\bar{\nu}(3656)$	13 [7]	7851000	0.76
$K^{*0}\nu\bar{\nu}(2585)$	8.7 [7]	6282000	0.63
$K^{*0}\nu\bar{\nu}(7487)$	13 [7]	5270000	1.13

- $B_u^+ \to K^+ J/\Psi, J/\Psi \to l^+l^-$ signal events, used to validate the signal efficiency and derive the systematic uncertainty for the signal efficiency.

- Generic continuum events (Table 3). The samples of generic continuum events are scaled to the appropriate cross-sections.

- Generic $B\bar{B}$ events (Table 3), which are used to study combinatorial backgrounds. The generic B samples are scaled run-by-run to the B-counting statistics.

- Double-tag events, where both B's decay semileptonically.

From the above data samples two more sub-samples are derived: a wrong-tag sample for classifier validation, and an a_1^+ signal sample used for background validation. See Section 6 for details.

4.2   Semileptonic Tag Skim

The target of this analysis is a B decaying to a kaon with missing energy. This decay provides no strong signature, since kaons are produced by many types of decays. For the missing energy to be identifiable, both B's coming from the $\Upsilon(4S)$ need to be reconstructed. A typical analysis strategy, followed by this analysis, is to reconstruct one B in a common, well-known decay mode, and then search through the remainder of the event for the signal.

Table 3: Number of MC simulated generic events and the ratio of the number of $B\overline{B}$ decays (or, for continuum events, scaled cross-section) in data to the number simulated.

Mode (SP)	Nominal Cross-sec. (nb)	Events	Data/MC
Generic B^+B^- (1235)	0.55	702714000	0.327
Generic $B^0\overline{B}^0$ (1237)	0.55	692626000	0.332
Continuum $c\bar{c}$ (1005)	1.30	1088218000	0.499
Continuum uds (998)	2.09	903912000	0.966
Continuum $\tau\bar{\tau}$ (3429)	0.919	382614000	1.00
Double tag (3159)	n/a	39017000	n/a

This first B is called the tag B, and the recoiling tracks remaining in the event after its reconstruction are checked to see whether they meet the requirements for the signal mode. This analysis utilizes the semileptonic tag, defined below. In all of the following the term "signal" should be considered synonymous with "recoiling"; that is, the signal B is the B that recoils from the tag B.

The semileptonic tag B is reconstructed in three steps. First, a subset of events meeting loose criteria for semileptonic decay is created. Next, this subset, which is called the BToD1nu skim, is subjected to two steps of refinement, described in Sections 4.3 and 5, that serve to select the events with the best reconstructed tags while also applying signal criteria.

The first step, the skim, is intended to select events with the following semileptonic decays:

- $B \to Dl\nu$

- $B \to D^*l\nu$

The skim includes charged and neutral B's decaying to neutral and charged D's, respectively. The lepton can be an electron or a muon. The following shows the D-decay modes that are reconstructed and how the events are selected. The listed pion momenta are in the center-of-mass frame of the $\Upsilon(4S)$.

- $D^{*0} \to D^0 \pi^0$

  $0.135 < \text{mass}(D^{*0}) - \text{mass}(D^0) < 0.175 \, \text{GeV}/c^2$

  π^0 momentum $< 450 \, \text{MeV}/c$

- $D^{*+} \to D^0 \pi^+$

  $0.135 < \text{mass}(D^{*+}) - \text{mass}(D^0) < 0.175 \, \text{GeV}/c^2$

  π^+ momentum $< 450 \, \text{MeV}/c$

- $D^{*+} \to D^+ \pi^0$

  $0.140 < \text{mass}(D^{*+}) - \text{mass}(D^+) < 0.150 \, \text{GeV}/c^2$

  π^0 momentum $< 450 \, \text{MeV}/c$

- $D^0 \to K^+ \pi^-$

  reconstructed from K^+ and π^- with mass constraint of $\pm 60 \, \text{MeV}/c^2$ of the D^0 mass

- $D^0 \to K^+ \pi^- \pi^0$

  reconstructed from K^+, π^-, π^0 with mass constraint of $\pm 100 \, \text{MeV}/c^2$ of the D^0 mass

- $D^0 \to K^+ \pi^- \pi^+ \pi^-$

  reconstructed from K^+, π^-, π^+, π^- with mass constraint of $\pm 60 \, \text{MeV}/c^2$ of the D^0 mass

- $D^0 \to K_S \pi^+ \pi^-$, $K_S \to \pi^+ \pi^-$

  reconstructed from K_S, π^-, π^+ with mass constraint of $\pm 60 \, \text{MeV}/c^2$ of the D^0 mass

- $D^+ \to K^+ \pi^+ \pi^-$

  reconstructed from K^+, π^+, π^- with mass constraint of $\pm 60 \, \text{MeV}/c^2$ of the D^+ mass

- $D^+ \to K_S \pi^+$, $K_S \to \pi^+ \pi^-$

  reconstructed from K_S, π^-, π^+ with mass constraint of $\pm 60 \, \text{MeV}/c^2$ of the D^+ mass

To be selected by the skim, an event must have a lepton with a center-of-mass-frame momentum greater than 0.8 GeV/c.

4.3 Tag Refinement and Signal Reconstruction

The skimmed events are processed by a special-purpose software package to create ntuples where each event has a tag and a signal B. The package performs the following steps to produce the ntuples.

The first step is to explicitly reconstruct the $D^* \to D\gamma$ decays, which the skim fully includes but does not reconstruct. Photons have the following requirements:

- $LAT = [0.0, 0.8]$ (the LAT is the lateral moment of the photon shower). See Figure 219 in Appendix A.6 for a plot of the LAT distribution of a photon control sample.

- Energy in lab frame $\geq 0.1 \, \text{GeV}/c^2$

D^* candidates are then constructed from the photons and D's from the skim with the following requirements:

- $\delta M = [0.12, 0.17] \, \text{GeV}/c^2$

- D^* mass within $0.5 \, \text{GeV}/c^2$ of nominal value

The B candidates are then reconstructed from the D^*'s, the remaining D's from the skim, and the leptons by using `TreeFitter` with the requirement that the vertex probability of the χ^2 per degree of freedom be greater than or equal to 0.001.

The next step is to impose various selection requirements:

- Number of charged tracks ≤ 15

- Total charge of the event $= [-2, 2]$

- $B \to Dl\nu$ tag candidates must satisfy $cos\theta_{BY} = [-5.0, 1.5]$, as defined in Equation 10.

$$cos\theta_{BY} = \frac{2E_{beam} \cdot E_{Dl} - m_B^2 - m_{Dl}^2}{2 \, |\mathbf{p_{Dl}}| \cdot \sqrt{E_{beam}^2 - m_B^2}} \tag{10}$$

where E_{beam} is the expected beam energy, E_{Dl} is the combined Dl energy, m_B is the B mass, m_{Dl} is the combined Dl mass, and $\mathbf{p_{Dl}}$ is the combined Dl momentum (all are in the center-of-mass frame).

- $B \rightarrow D^*l\nu$ tag candidates have two requirements, calculated in the center-of-mass frame:

    For θ between the two D^* daughters (the D and a pion or photon), $\theta < 2$

    For θ between the D^* and the lepton, $cos\theta < 0$

After these steps, there may still be multiple tag candidates. For each event, the candidate chosen as the tag is the one with highest vertex probability.

4.3.1 Assignment of Signal Mode

Next the signal candidate is constructed, and it is assigned a mode based upon the first match in the sequence below. Signal candidates from signal MC are also put through this sequence, with the result that some of them are reconstructed in the wrong mode. The details of the other modes are explained below in the interests of completeness, but most do not play any further role in this analysis because they are cut away, as explained in Section 4.3.2.

The sequence is:

1. Double-tag event: the signal candidate matches all the criteria for a semileptonic tag as described previously, with no particles shared by both tags.

2. Signal tracks match a $K^{*+} \rightarrow K_S^0 \pi^+$ decay, with a reconstructed K^* mass within 0.075 GeV/c^2 of the canonical K^* mass. The K_S^0 is reconstructed from two tracks from the charged tracks list with a mass range of 0.47267 to 0.52267 GeV/c^2.

3. Signal tracks match a $K^{*0} \rightarrow K_S^0 \pi^0$ decay, with the same restrictions as item 2 above.

4. $B^0 \to K_S^0 \nu\bar{\nu}$ signal mode if the K_S^0 is reconstructed in the $K_S^0 \to \pi^+\pi^-$ mode, with the same restrictions for the K_S^0 as in item 2 above, and the additional requirements that the χ^2 per degree of freedom of the kinematic fit be at least 0.001 and the flight significance be at least 3σ.

5. Signal candidate has three tracks on the charged tracks list, and these tracks are reconstructed as an a_1^+ candidate with mass between 0.6 and 2.0 GeV/c^2.

6. Signal candidate has two tracks on the charged tracks list, and these tracks match a $K^{*0} \to K^-\pi^+$ decay.

7. If only one signal track is on the charged tracks list, then the following are checked in order until the first match is found:

   (a) Track matches a $K^{*+} \to K^+\pi^0$ decay.

   (b) Track matches a $K^{*+} \to K_S^0\pi^+, K_S^0 \to \pi^0\pi^0$ decay.

   (c) Track matches a $\rho^+ \to \pi^+\pi^0$ decay.

   (d) $B^+ \to K^+\nu\bar{\nu}$ signal mode if track is on the high-purity K^+list.

8. A signal candidate reaching this step gets labeled with a mode not used in this analysis.

The above requirements do not restrict extra charged tracks or neutrals not attached to either the signal or tag. The resulting ntuples from this process contain both K^+ and K_S^0 signal events and the other events listed above. The cuts described in Section 4.3.2 select only those events reconstructed in the desired signal mode. Table 4 shows the breakdown of the different modes that are reconstructed. Though some signal is lost by only selecting the correctly reconstructed mode, such selection is essential. Adding even one of the incorrectly reconstructed modes would increase the number of background events by at least four times while only marginally increasing the signal. For the K_S mode, only correctly reconstructed K_S events can be used because the K_S mass is an important classification variable (see

Section 5.1), and incorrectly reconstructed events have a mass for the first daughter of the recoiling B which matches its reconstructed identity, not an appropriate K_S mass.

Table 4: Percentages of skimmed signal MC that is reconstructed in the modes described in Section 4.3.1, for both K^+ and K_S signal MC. The Reconstruction Mode refers to the reconstructed identity of the first daughter of the recoiling B.

Reconstruction Mode	% of K^+ Signal MC	% of K_S Signal MC
K^+	45.0	0.7
K_S	0.6	32.7
π^+	18.2	15.5
ρ^+	11.3	8.9
D^0	9.0	7.5
K^{*+}	4.4	9.2
a_1^+	6.5	5.1
D^+	2.0	7.8
D^{*+}	0.06	4.6
K^{*0}	0.8	3.3
γ	0.1	3.7
μ^+	1.1	0.4
e^+	0.5	0.6

4.3.2 Preselection Requirement for Each Signal Mode

The next step is to apply cuts to the ntuples to prepare them for use with the event-selection decision trees described in Section 5.1. Two sets are ntuples are generated: one for the K^+ mode, and one for the K_S^0 mode. These cuts select the desired signal mode with high efficiency and loosely trim the ntuples of unnecessary background. They are applied to all ntuples, including data, signal MC, and background MC.

The cuts for the K^+ ntuples are:

- The signal B must have only one daughter, which must be a charged kaon.

- The signal B and tag lepton must have opposite charges.

- No more than two extra tracks are allowed. Extra tracks are those from the charged tracks list that are left over after the signal and tag are fully reconstructed.

The cuts for the K_S^0 ntuples are:

- The signal B must have only one daughter, which must be a K_S^0.

- No more than one extra track is allowed.

The effect of these cuts on efficiency is shown in Tables 5 and 6.

Table 5: Efficiency and signal significance for K^+ signal MC and all generic background MC at each step of the analysis process. The steps are skimming, the reconstruction of the signal mode and refinement of the tag, the preliminary cuts for ntuple production, and a cut on the BDT classifier output. The efficiencies are the total efficiencies, including all previous steps. The signal significance measure is Equation 11. See Section 5.1 for explanation of the BDT cut.

Step	K^+ Efficiency	K^+ Significance	K^+ Background Efficiency
BToDlnu Skim	0.0694±0.0001	0.01	$(9.28\pm0.00) \cdot 10^{-2}$
Signal/Tag Reconstruction	0.0148±0.0000	0.007	$(5.92\pm0.00) \cdot 10^{-3}$
Preliminary Cuts	0.00672±0.00003	0.05	$(2.11\pm0.01) \cdot 10^{-5}$
Primary BDT Cut	0.00175±0.00001	0.7	$(7\pm2) \cdot 10^{-9}$

Table 6: Efficiency and signal significance for K_S^0 signal MC and all generic background MC at each step of the analysis process. The steps are skimming, the reconstruction of the signal mode and refinement of the tag, the preliminary cuts for ntuple production, and a cut on the BDT classifier output. The efficiencies are the total efficiencies, including all previous steps. The signal significance measure is Equation 11. See Section 5.1 for explanation of the BDT cut.

Step	K_S^0 Efficiency	K_S^0 Significance	K_S^0 Background Efficiency
BToDlnu Skim	0.0578±0.0001	0.003	$(9.33\pm0.00) \cdot 10^{-2}$
Signal/Tag Reconstruction	0.0117±0.0001	0.003	$(5.96\pm0.00) \cdot 10^{-3}$
Preliminary Cuts	0.00333±0.00003	0.01	$(4.21\pm0.01) \cdot 10^{-5}$
BDT Cut	0.00060±0.00001	0.3	$(1.6\pm0.8) \cdot 10^{-9}$

5 Decision Trees for Signal/Background Discrimination

5.1 Bagger Decision Trees

A StatPatternRecognition (SPR) [24] Bagger Decision Tree (BDT) is used to select likely signal events and reject background ones. To build, optimize, and use a decision tree, the following steps are necessary:

1. Create a training set of events along with a disjoint set of events for testing.

2. Choose a list of classification variables and SPR tree parameters.

3. Build a tree using the training set and the choices from Step 2.

4. Use the testing set to evaluate tree performance by some criterion.

5. Go back to Step 2 and repeat process with new choices until performance is optimal.

The training and testing sets are composed of MC events: signal, $B^0\overline{B}^0$, B^+B^-, $c\bar{c}$, uds, $\tau\bar{\tau}$, and double-tag background, which are events in which both B's undergo semileptonic decay. The total set of events is divided in half randomly to yield the training and testing sets, with a 50% chance that any event will be put into the training set. This random division avoids any bias that a deterministic division might introduce. If the above process is performed with only one tree, any small changes in tree performance caused by different choices of tree parameters can be overwhelmed by stochastic variations related to the random division of events into the training and testing sets. In other words, a new random division into training and testing sets can change performance more than a change in tree parameters.

To reduce the effect of this variation, twenty training and testing sets are generated. The random division of the training and testing sets is performed twenty times, so each of twenty trees has its own training set and completely disjoint testing set. The performance results

from the twenty trees are averaged (as explained below) to give a more accurate measure of the performance due to the choice of classification variables and SPR parameters than would be provided by only one tree. Many cycles of tree building and evaluation are the basis for the choices of classification variables and SPR parameters described in the following paragraphs.

The choice of the number twenty for the size of the set of trees is based upon the observation that fewer than twenty trees still show significant stochastic variation, while more than twenty creates substantial overhead in disk usage and processing time with little further reduction in stochastic variation.

The use of twenty trees also largely eliminates the problem of bias in choosing tree parameters with a particular testing set. If only one tree were used, the practice of breaking the event set into three parts might be necessary: a training set, a testing set to find the optimum parameters, and a validation set to assess performance without any bias. The danger with one tree is that, because it is optimized for the testing set, the performance of the testing set will not be representative of the tree's performance with an independent set like data. However, in this analysis, with twenty trees using the same tree parameters, it is very unlikely that the twenty trees are each optimized for their testing sets, so the average performance of the testing sets is likely to be a valid measure of the tree performance. In addition, validation is performed to show that tree performance with MC matches that with data (see Section 6). Furthermore, due to the relatively small size of the event set, it is advantageous to split it only twice, into training and testing sets, rather than three times into training, testing, and validation sets.

Three kinds of trees are trained (with twenty of each kind as explained above): a primary $K^+\nu\overline{\nu}$ signal tree, a subset $K^+\nu\overline{\nu}$ signal tree that is less effective but is used for signal efficiency validation, and a $K^0_S\nu\overline{\nu}$ signal tree. Each of the three types has its own list of

classification variables, as shown in Tables 8 to 9, with Table 7 showing the definitions of the variables. The next page gives a description of the variables used. Signal and background histograms and data/MC comparisons for each variable are shown in the appendix. The primary K^+ tree uses 26 classification variables, the subset tree uses 15, and the K_S^0 tree uses 38. Tables 11 and 12 show rankings of the classification power of the variables for the primary K^+ and K_S^0 trees.

In addition to the classification variable list, there are three SPR parameters that affect a tree's performance: the choice of the figure of merit for tree building, the minimum leaf size in the tree, and the number of training cycles. SPR decision trees offer a number of different figures of merit that can optimize tree performance for particular applications. The figure of merit called the Gini index, which is $-2p(1-p)$ where p is the fraction of correctly classified events, works very well for this analysis and is used by all the trees. The best leaf size for the two K^+ trees is 50, while for the K_S^0 tree it is 35. For all three kinds of trees, the optimal number of cycles is 150.

The process of optimizing the SPR tree parameters requires a performance criterion. An SPR tree classifies events by computing a classifier output value from 0 to 1 for each input event, with values near 0 indicating background events and values near 1 indicating signal events. Figure 15 shows histograms of the classifier output for MC signal and background events in the testing set. Counting the classified events from the testing set that exceed some cut value in the classifier output gives a yield of signal and background events. Since there are twenty trees in a set, a single classifier output cut value will not work. Instead, a desired total signal efficiency is chosen. Then a cut value that yields this total signal efficiency is computed for each tree by using the testing events. Applying this cut gives a background yield for that tree, and the background yields for the twenty trees can be averaged to give a measure of the performance for that set of trees.

Description of classifier variables for both trees

- signal kaon lab momentum;
- cosine of the polar angle of (i) total missing CMS momentum, (ii) missing CMS momentum using just tag B and signal K candidates and (iii) signal K CMS momentum;
- cosine of the opening angle between the signal K candidate and (i) the event thrust vector and (ii) the tag Dl thrust vector;
- missing energy and momentum in the lab and CMS frame;
- tag B and D masses;
- tag D and lepton CMS momenta;
- tag D decay mode;
- χ^2 value of tag B vertex fit;
- $cos\theta_{BY}$ (Equation 10)
- total lab energy;
- total energy of all non-tag and non-signal charged and neutral candidates;
- number of (i) π^0, (ii) K_L and (iii) all non-tag neutral candidates;
- number of tracks not used in signal or tag reconstruction;
- normalized second Fox-Wolfram moment;
- minimum mass constructed from any three tracks.

Two additional variables are used only for the K^+ BDT:

- signal kaon CMS momentum;
- number of K_L's identified in the muon system.

Several additional variables are used only for the K_S BDT:

- signal K_S^0 mass, CMS energy;
- event missing momentum polar angle in the signal K frame;
- net charge of (i) the event, (ii) tracks matched to a calorimeter energy deposit;
- uncertainties in x of the signal K's point of closest approach to the e^+e^- interaction point, in both 3-d and a plane perpendicular to the beamline;
- $cos\theta_{BY}$ calculated by adding a single photon to the above $Dl\nu$ candidate such that $M(D^0, \gamma) - M(D^0)$ is $100 - 150 \, \mathrm{MeV}/c^2$;
- tag B CMS momentum;
- tag D number of daughters;
- tag lepton lab momentum;
- error on the x coordinate of the tag D leading pion daughter's point of closest approach to the beam spot;
- $\Upsilon(4S)$ momentum and polar angle cosine.

Table 7: Definitions of the classification variables used by the three trees.

CosThMissCM	cosine of the polar angle of the missing center-of-mass momentum
d_decaymode	decay mode of the tag D
EMiss	missing energy in the lab frame
EMissCM	missing energy in the center-of-mass frame
etotobs	total observed energy in the event (beam energy minus missing energy)
extrapi0s	the number of extra π^0's in the event
e_extra_all	total extra energy from neutral clusters and charged tracks lists
NetCharge	net charge of the event
NetChargeVisibleE	net charge of the event computed using only visible energy
PMiss	missing momentum in the lab frame
PMissCM	missing momentum in the center-of-mass frame
R2All	second Fox-Wolfram moment of the event divided by the 0th
sigbcosththrust	cosine of the polar angle between the signal K and the event thrust vector
sigbcosththrustdl	cosine of the polar angle between the signal K and the Dl thrust vector
sigbmin3invmass	minimum invariant mass that can be constructed from any three tracks in event
sigbn_extraneutrals	number of extra clusters in the EMC
sigbn_extra_emckl	number of extra K_L's detected in the EMC
sigbn_extra_ifrkl	number of extra K_L's detected in the IFR
sigbn_extra_tracks	number of charged tracks that aren't used for either signal or tag
sigbpmissthetacm	polar angle of momentum missing from signal and tag B's in center-of-mass frame
sigkcosthcm	cosine of the polar angle of the signal kaon's center-of-mass momentum vector
sigkecm	center-of-mass energy of the signal kaon
sigkmass	reconstructed mass of signal K_S
sigkp3	lab three-momentum of the signal kaon
sigkp3cm	center-of-mass three-momentum of the signal kaon
sigkpocaxy_cxx	xx covariance matrix element using the xy projection for the signal K's point of closest approach to the beam spot
sigkpoca_cxx	xx covariance matrix element for signal K's point of closest approach to beam spot
tagbchi2	χ^2 value of tag B vertex
tagbcosby	cosBY of the tag B
tagbcosbyphotonadd	like tagbcosby, but whose value has been improved (made larger) by adding the left-over photon which best improves cosBY and makes m(D^0 + photon) $-$ m(D^0) lie between 100-150 MeV/c^2
tagbmass	mass of the tag B
tagbp3cm	center-of-mass three-momentum of the tag B
tagdmass	mass of the semileptonic tag D
tagdndaus	number of daughters of the tag D
tagdp3cm	center-of-mass three-momentum of the semileptonic tag D
taglp3	lab three-momentum of the semileptonic tag lepton
taglp3cm	center-of-mass three-momentum of the semileptonic tag lepton
tagpipoca_xerr	error on the x coordinate of the tag pion's point of closest approach to the beam spot
ycosth	cosine of the polar angle of the momentum of the $\Upsilon(4S)$
yp3	momentum of the $\Upsilon(4S)$

Table 8: Classification variables for the primary $K^+\nu\bar{\nu}$ tree.

CosThMissCM	sigbn_extraneutrals
d_decaymode	sigbn_extra_emckl
EMiss	sigbn_extra_ifrkl
EMissCM	sigbn_extra_tracks
etotobs	sigkcosthcm
extrapi0s	sigkp3
e_extra_all	sigkp3cm
PMiss	tagbchi2
PMissCM	tagbcosby
R2All	tagbmass
sigbcosththrust	tagdmass
sigbcosththrustdl	tagdp3cm
sigbmin3invmass	taglp3cm

Table 9: Classification variables for the $K^0_S\nu\bar{\nu}$ tree. Variables shared with the primary $K^+\nu\bar{\nu}$ tree are in **bold**.

CosThMissCM	**sigbmin3invmass**	**tagbmass**
d_decaymode	**sigbn_extraneutrals**	tagbp3cm
EMiss	**sigbn_extra_emckl**	**tagdmass**
EMissCM	**sigbn_extra_tracks**	tagdndaus
etotobs	sigbpmissthetacm	**tagdp3cm**
extrapi0s	**sigkcosthcm**	taglp3
e_extra_all	sigkecm	**taglp3cm**
NetCharge	sigkmass	tagpipoca_xerr
NetChargeVisibleE	**sigkp3**	ycosth
PMiss	sigkpocaxy_cxx	yp3
PMissCM	sigkpoca_cxx	
R2All	**tagbchi2**	
sigbcosththrust	**tagbcosby**	
sigbcosththrustdl	tagbcosbyphotonadd	

Table 10: Classification variables for the subset tree.

d_decaymode	sigbn_extra_tracks
etotobs	tagbchi2
extrapi0s	tagbcosby
e_extra_all	tagbmass
R2All	tagdmass
sigbn_extraneutrals	tagdp3cm
sigbn_extra_emckl	taglp3cm
sigbn_extra_ifrkl	

A limitation of the primary K^+ tree is that it excludes events with low signal-kaon momentum, as shown in Figure 151. Such events are of theoretical interest because the partial branching fraction could be strongly enhanced in the low-kaon-momentum region according some new-physics models [8]. The exclusion of these low-momentum events is the result of the predominance of background in this momentum region (as seen in Figure 53) and not due to any explicit cut. To address this issue, a separate low-momentum measurement is made using the primary K^+ trees but limited to events with sigkp3cm < 1.5 GeV/c. This value appears to be the approximate limit below which nearly no events appear in the primary K^+ signal region, which is explained below. Under the ABSW signal model [8] (see Section 5.2.4) 40% of the K^+ signal MC events fall within the low-momentum limit, while 60% have higher momenta.

With values for the signal and background yields, a signal significance can be computed. Then the choice of targeted signal efficiency, and resulting background yield, can be optimized to maximize the signal significance.

Three different measures of signal significance are used in this analysis:

$$\frac{s}{\sqrt{s+b}} \tag{11}$$

$$\frac{s}{N_\sigma/2 + \sqrt{b}} \tag{12}$$

Table 11: Signal/background discrimination effectiveness (as reported by SPR) of classification variables for the primary $K^+\nu\overline{\nu}$ tree. **Delta FOM** indicates how much that variable contributes to the improvement of the figure-of-merit used by the decision tree to assess signal/background discrimination. **Splits** indicates the number of splits made by the tree on that variable. These values are averaged for the twenty trees used.

Variable	Delta FOM	Splits
e_extra_all	400.0	8689
sigkp3cm	310.9	7972
tagdmass	265.9	8016
tagbcosby	216.5	7468
sigbmin3invmass	166.0	5541
R2All	164.1	4848
PMissCM	155.5	5541
CosThMissCM	145.7	4152
tagbmass	120.7	4967
sigkp3	120.1	4223
tagbchi2	115.7	4115
taglp3cm	109.4	4269
sigkcosthcm	109.3	5383
EMissCM	106.7	4725
sigbcoststhrustdl	102.1	3537
EMiss	95.0	3430
sigbcoststhrust	86.8	2931
tagdp3cm	86.4	4329
PMiss	82.1	2901
sigbn_extraneutrals	45.5	1538
d_decaymode	38.8	1358
extrapi0s	31.0	782
etotobs	29.9	1054
sigbn_extra_emckl	10.5	377
sigbn_extra_tracks	7.1	350
sigbn_extra_ifrkl	2.2	239

Table 12: Signal/background discrimination effectiveness (as reported by SPR) of classification variables for the $K_S^0 \nu \bar{\nu}$ tree. **Delta FOM** indicates how much that variable contributes to the improvement of the figure-of-merit used by the decision tree to assess signal/background discrimination. **Splits** indicates the number of splits made by the tree on that variable. These values are averaged for the twenty trees used.

Variable	Delta FOM	Splits
e_extra_all	316.4	5374
sigkecm	204.1	4741
tagdmass	124.7	2973
PMissCM	112.7	3431
tagbp3cm	110.5	3183
R2All	99.8	2235
tagbmass	89.7	3099
sigkp3	89.0	2660
EMissCM	86.4	3114
tagbcosby	84.2	2633
tagbchi2	79.6	2143
CosThMissCM	78.2	1632
tagbcosbyphotonadd	67.3	1923
sigbcosththrustdl	66.0	1642
taglp3	57.9	1693
sigbcosththrust	53.0	1400
taglp3cm	52.9	1459
tagdp3cm	49.5	2287
tagpipoca_xerr	49.1	1720
ycosth	47.3	1409
sigkcosthcm	46.6	2671
sigbmin3invmass	46.4	1600
sigkmass	40.3	1307
yp3	38.9	966
etotobs	35.1	988
sigkpoca_cxx	34.2	1302
sigbpmissthetacm	33.3	792
PMiss	31.7	748
d_decaymode	27.9	1237
NetCharge	23.8	462
EMiss	17.8	515
sigbn_extraneutrals	17.3	523
extrapi0s	14.7	330
sigkpocaxy_cxx	12.4	619
sigbn_extra_tracks	7.5	125
tagdndaus	7.4	172
NetChargeVisibleE	6.5	192
sigbn_extra_emckl	4.9	164
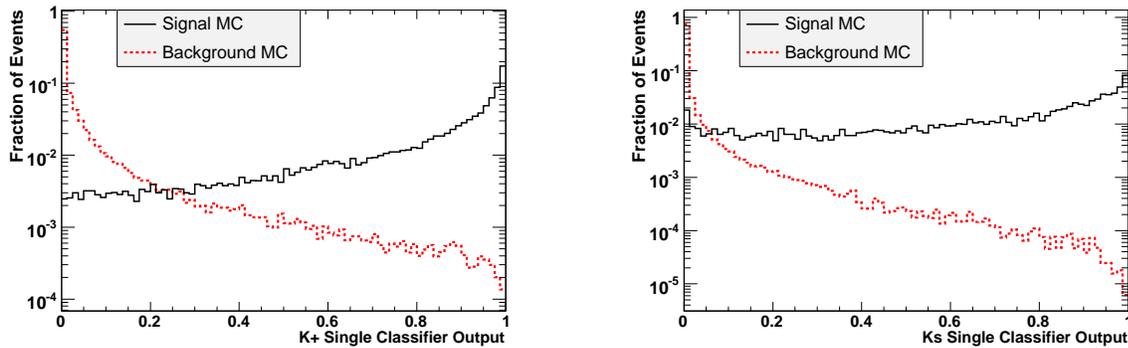
Figure 15: Histograms showing the classifier output for MC signal and background events in the testing set. The left plot is for the primary K^+ tree. The right plot is for the K_S^0 tree. Each plot is from one tree that is representative of its twenty-tree set.
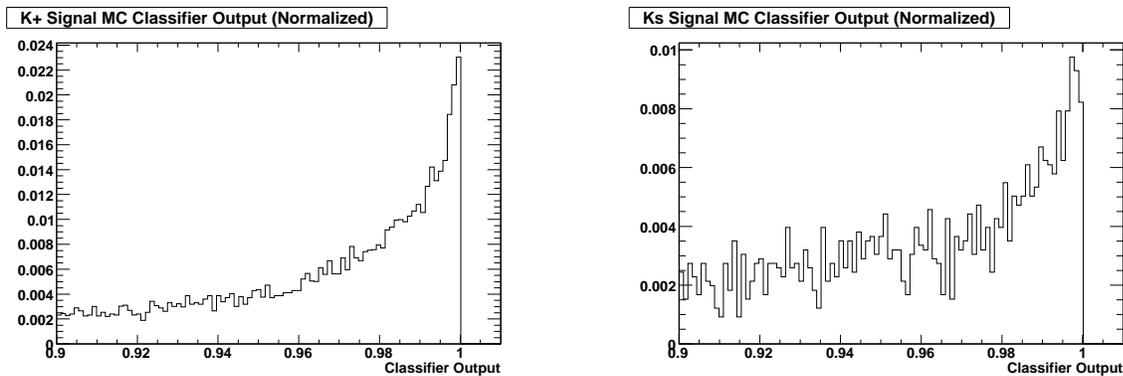
Figure 16: Histograms showing the classifier output for MC events in the testing set. Background events are too few to show up on these plots. The left plot is for the primary K^+ tree and shows only signal MC events at the high end of the classifier output. See Figure 15 with the log scale to see the background in this region. The histogram is sharply peaked at 1, indicating that the tree has strong discriminating power. The right plot is for the K_S^0 tree at the high end of the classifier output. See Figure 15 with the log scale to see the background in this region. The blunter, lower peak at 1 hints that this tree has less discriminating power than the primary K^+ shown at left. Each plot is from one tree that is representative of its twenty-tree set.

$$\frac{\epsilon_{sig}}{0.6425 + \sqrt{b}} \tag{13}$$

where s is the number of signal events, b is the number of background events, ϵ_{sig} is the total signal efficiency (the number of signal events found divided by the total number of signal events in the sample), and N_σ is the desired sigma level of discovery. Equation 11 is a rough measure of signal significance. Equation 12 is the Punzi figure of merit [25], with N_σ set to 3. Equation 13 is another version of Equation 12, with the signal efficiency used instead of the number of signal events in order to remove dependence on an assumed branching fraction, and with N_σ set to 1.285, which corresponds to a 90% confidence level.

As shown in Figures 17 to 19, all three measures show wide, flat maxima in the region of 0.16-0.20% total signal efficiency for the K^+ mode or around 0.06% efficiency for the K_S mode, so the choice of classifier output cut within this region is somewhat arbitrary.

Tables 5 and 6 show the efficiency and signal significance at each step of the analysis process.

With optimized trees and a total signal efficiency chosen to give the maximum signal significance for the trees, real data events can be classified. The averaged output of the twenty trees gives the number of real events in the signal region. Along with the predicted number background events, this value allows calculation of a limit on the branching fraction.

5.2   Signal and Background Weighting

Data and generic background MC are classified by the trees, and the agreement between the two is checked for the classifier output up to (but not including) the signal region. The K_S trees show good agreement (see Section 5.2.3), but the K^+ trees produce large discrepancies. To fix this problem, a weighting is applied to the MC, as explained in the next section.
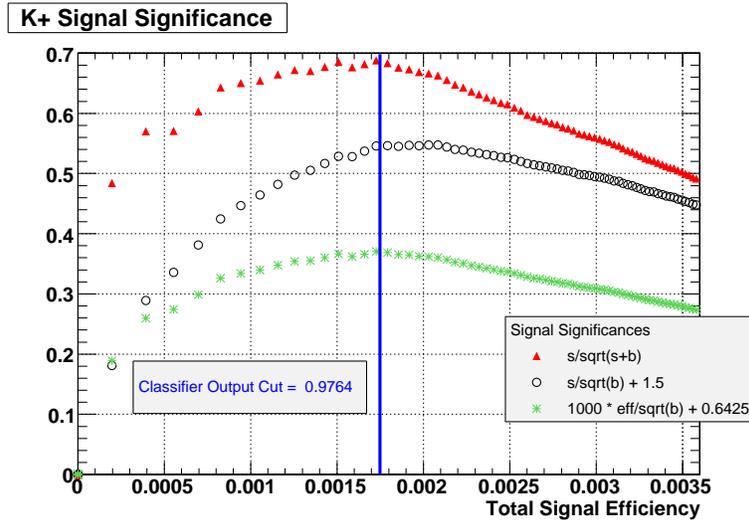
Figure 17: Measures of signal significance for the primary K^+ trees, averaged over the twenty trees. The red triangles correspond to equation 11, the black circles correspond to equation 12, and the green asterisks correspond to equation 13. The blue line shows the position of a possible cut that would approximately maximize signal significance for all three measures. The classifier output value corresponding to the blue line is shown in the box next to the line.

5.2.1 Monte Carlo Tuning for the K^+ Mode

The data and generic background MC K^+ events that result from the preliminary cuts described in Section 4.3.2 provide the starting point for calculating the MC weighting. The data events are used to populate a two-dimensional histogram of the variables `e_extra_all` and `sigkp3`. These two classifier variables are shown by Table 11 to be among the most important variables for tree performance. The binning of the histogram is made as fine as possible without creating excessively sparse bins. The lowest number of events in a bin turns out to be 77, which ensures sufficient statistics in all bins. The MC events are then put into the histogram, and the MC events in each bin are given a weight equal to the ratio in that bin of the number of data events to the luminosity-weighted number of MC events. Figure 20 shows the histogram with the calculated weights. These weights, which are referred to as
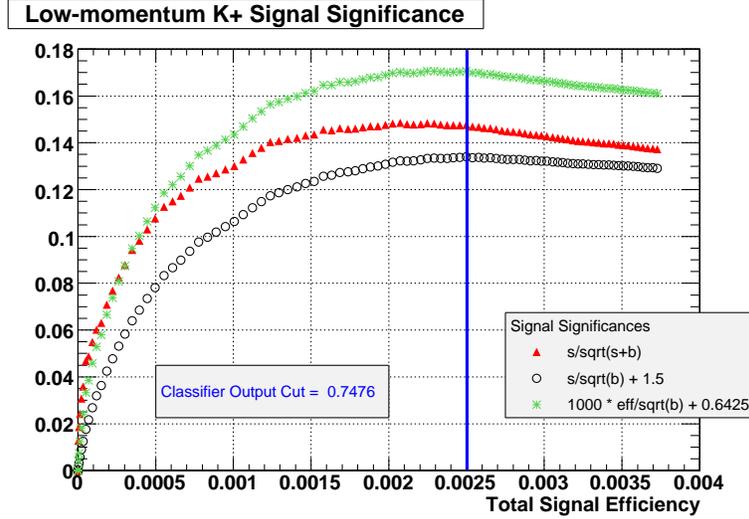
Figure 18: Measures of signal significance for the low-momentum K^+ measurement, averaged over the twenty trees. The red triangles correspond to equation 11, the black circles correspond to equation 12, and the green asterisks correspond to equation 13. The blue line shows the position of a possible cut that would approximately maximize signal significance for all three measures. The classifier output value corresponding to the blue line is shown in the box next to the line.

2-D weights in this document, act as an additional factor after the luminosity weighting of the MC events.

When these weights are applied, data/MC agreement for the primary K^+ classifier output adjacent to the signal region is very high (see Figures 21 and 22), which provides confidence that the weighted background MC accurately predicts the background in the signal region.

5.2.2 Data/MC Agreement for the Low-Momentum Measurement

The low-momentum K^+ measurement also utilizes the 2-D weighting. Figure 23 shows good data/MC agreement for this measurement, though there is a slight -5% data/MC discrepancy, which is used as a correction to the background prediction.
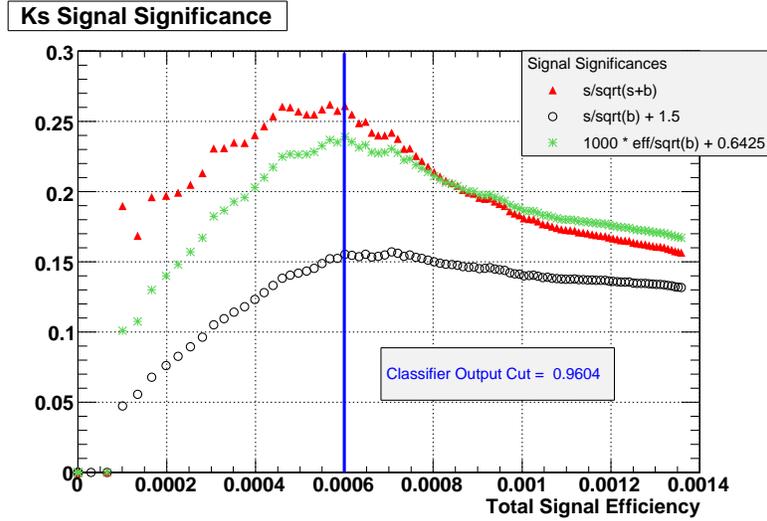
Figure 19: Measures of signal significance for the K_S^0 trees, averaged over the twenty trees. The red triangles correspond to equation 11, the black circles correspond to equation 12, and the green asterisks correspond to equation 13. The blue line shows the position of a possible cut that would approximately maximize signal significance for all three measures. The classifier output value corresponding to the blue line is shown in the box next to the line.
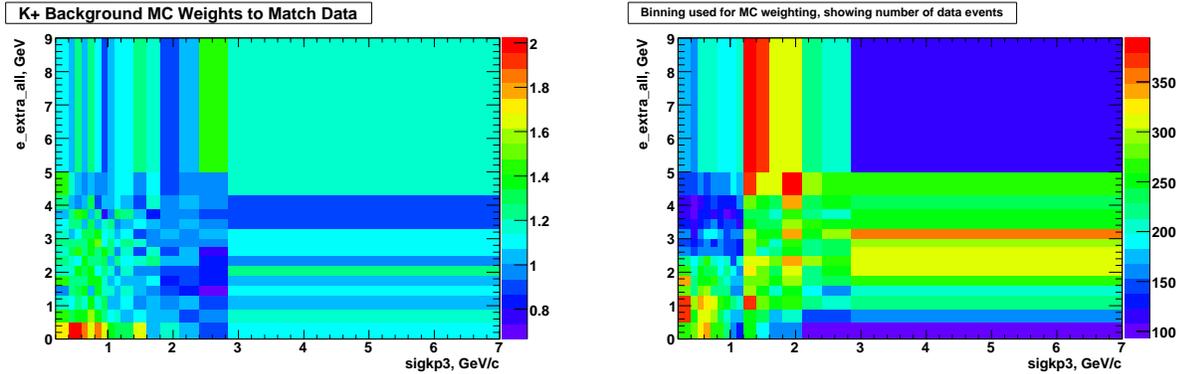
Figure 20: Two-dimensional weighting histogram for the primary K^+ mode, showing the values of the weights assigned to generic background MC events that fall in each bin on the left, and the number of data events in each bin on the right.

5.2.3   Data/MC Agreement for the K_S Mode

As shown in Figures 24 and 25, the classifier output for the K_S mode shows good data/MC agreement, even without any special weighting. The small $+10\%$ discrepancy of data over
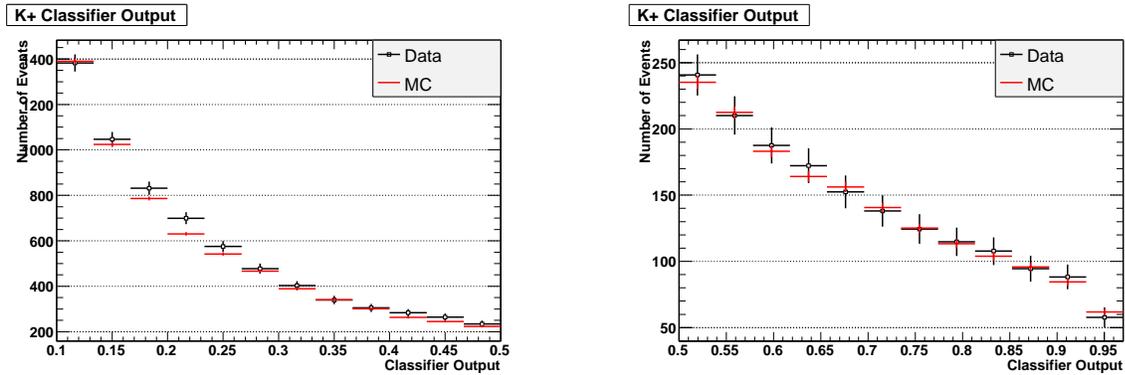
Figure 21: Classifier output for the primary K^+ trees for data and generic background MC events with the 2-D weighting. The left plot shows the range 0.1 to 0.5, and the right shows 0.5 to 0.975, the edge of the signal region. MC matches data very closely.
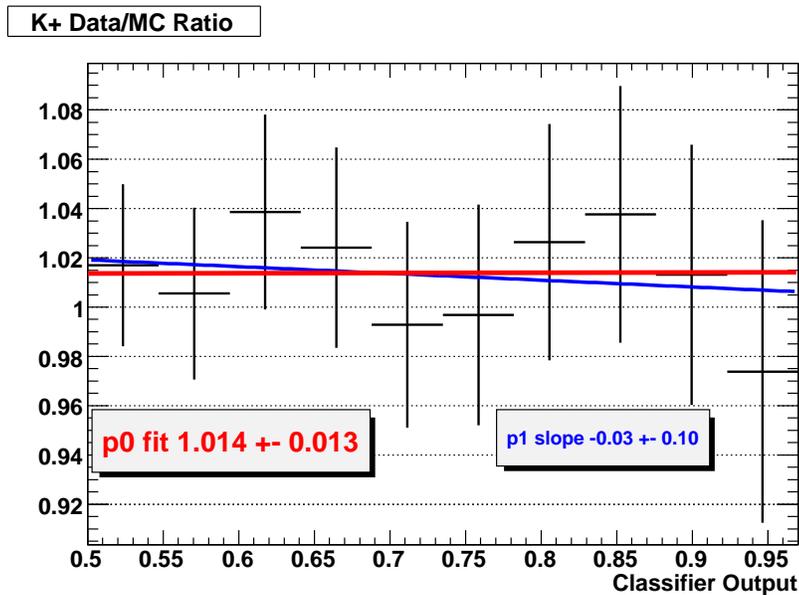
Figure 22: Data/MC ratio with the 2-D weighting applied for the primary K^+ classifier output just below the signal region. Agreement between data and MC is very good.

MC seen in Figure 25 is applied to the K_S background prediction for the signal region. As discussed in Section 7.2, this value is also used as the systematic uncertainty to the background prediction.
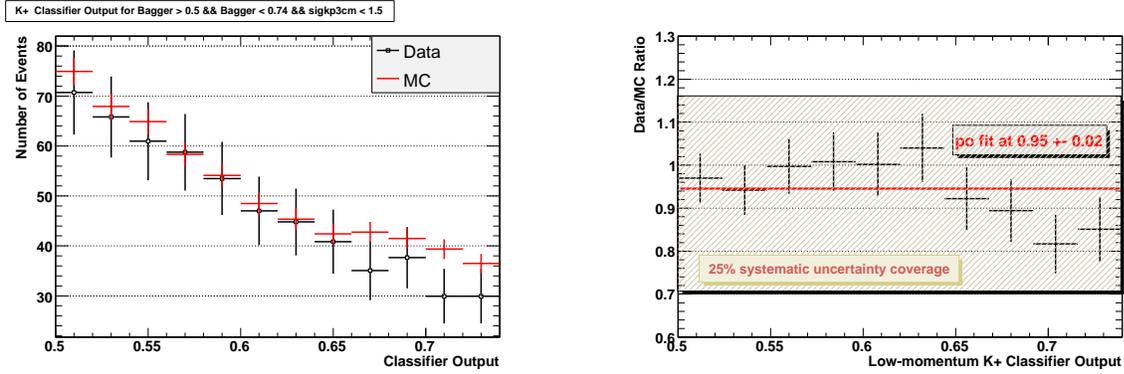
Figure 23: Classifier output for the low-momentum K^+ measurement with the 2-D weighting. The left plot shows data and background MC up to the edge of the signal region. The left plot shows the data/MC ratio. The fitted line gives the value of the ratio at 0.95 ± 0.02, so a -5% correction needs to be applied to the background prediction for this measurement. The systematic uncertainty applied to this measurement (see Section 7.2.2) is shown to demonstrate that it covers the fluctuations of the data/MC ratio.
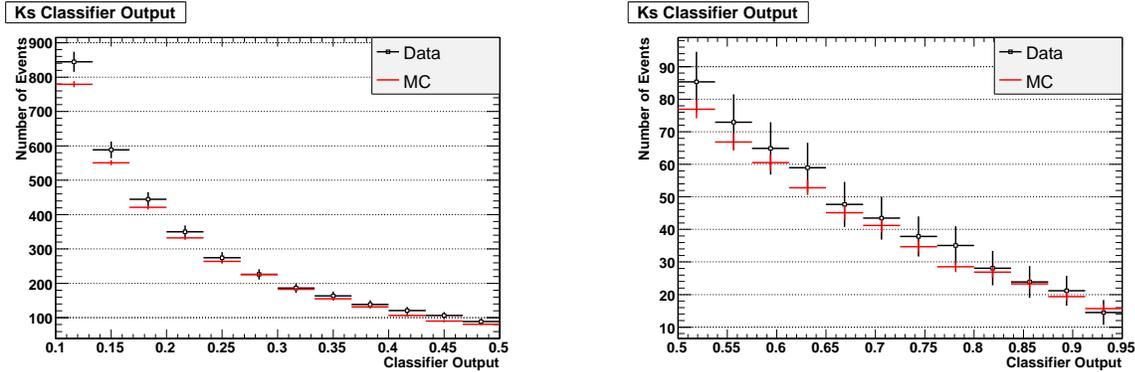
Figure 24: Classifier output for the primary K_S trees for data and generic background MC events. The left plot shows the range 0.1 to 0.5, and the right shows 0.5 to 0.95, the edge of the signal region. MC matches data very closely.

5.2.4 Signal Models and Weighting

This analysis uses three models for signal MC: phase space, BHI(2001) [7], and ABSW(2009) [8]. Phase space is the basic MC model, and BHI- and ABSW-model results are obtained by weighting the signal MC. Throughout this document the phase-space signal model should be assumed unless another model is specified. Figure 2 illustrates the three models by showing
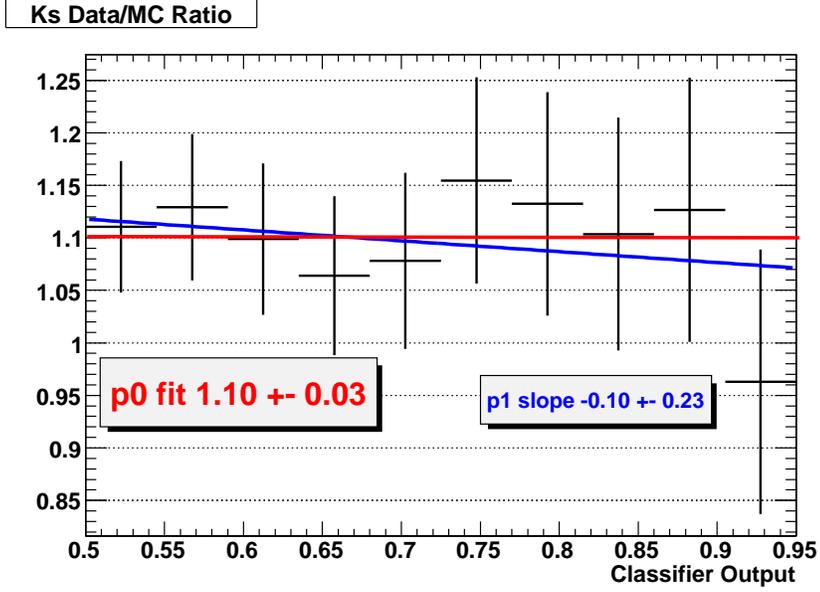
Figure 25: Data/MC ratio for the K_S classifier output just below the signal region. The small data/MC discrepancy is used as the correction to the background estimate and as the background systematic uncertainty.

the di-neutrino invariant mass squared spectrum for each.

Figure 26 shows the weights used to convert the phase-space signal MC to each of the two theoretical models. Applying the weights to the signal MC events in the signal region reduces the signal efficiency compared to the phase-space model: for both the primary K^+ and K_S modes, the BHI model reduces efficiency by 8%, while the ABSW model reduces it by 19%. Table 13 shows the signal efficiencies for each model. Section 8 gives the measured upper limits for each model.

Table 13: Total signal efficiencies under the three signal models.

	Phase-space Model	BHI Model	ABSW Model
K^+	0.175%	0.161%	0.141%
K_S	0.060%	0.055%	0.048%
High-momentum K^+	0.251%	0.241%	0.236%
Low-momentum K^+	0.260%	0.278%	0.250%
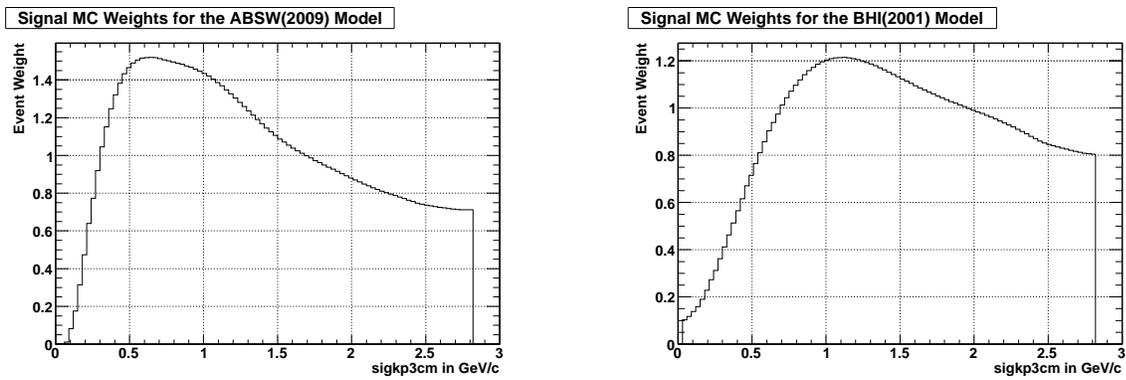
Figure 26: Weights for signal MC to convert the phase-space model to the two theoretical models.

6 Decision Tree Validation

The decision trees are trained on MC events. Checks are necessary to show that the trees classify real events similarly to MC events. Two real data data samples are used for this validation as described below.

6.1 Wrong-tag Sample

One check is to compare data and MC events where the tag B has the wrong charge to produce the signal kaon. For the K^+ trees, this tag B is the B^0, while for the K_S trees, it is B^+. Such events are pure background.

To achieve the most sensitive check of data-MC discrepancies with multiple trees, it is desirable to employ a method that ensures that the data-MC discrepancy for one tree is not canceled out by the discrepancy for another tree when they are averaged together. For this reason, the absolute data-MC differences for the trees, rather than the data/MC ratios, are used. In order to separate the effects of the classifier from data-MC discrepancies caused by skimming and reconstruction, the data-MC difference histograms are unit-normalized, so they will show only discrepancies caused by the classifier and not those caused by earlier steps of the analysis process.

For each tree, the normalized histogram of the classifier output for MC events is subtracted from that for data events. Taking the absolute value and averaging over the twenty trees produces Figure 27, which shows that the discrepancy in the classification of background B^0 events between data and MC in each bin is always less than 0.6% of the total number of events. Figure 28 shows the corresponding plot for the K_S trees, in this case using tag B^+ events. Figures 29 and 30 show similar plots where the tag particles are plotted separately from the tag anti-particles (e.g., separate plots for B^0 and $\bar{B}^0$).

Figures 31 and 32 show the data/MC difference plotted with bins covering various ranges

of classifier output values. Figures 33 and 34 are similar but show the signed differences rather than absolute values.
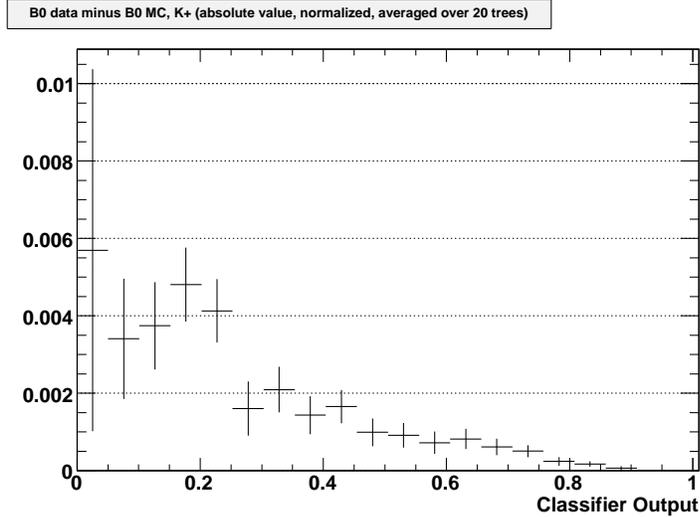
Figure 27: Normalized histogram of the difference in the classifier output for B^0 data and B^0 MC for the primary K^+ trees. The plot shows the absolute values of the normalized differences, averaged over the twenty trees. The data/MC discrepancy in each bin is always less than 0.6% of the total number of events. See text for more details.

6.2 a_1^+ Sample

Events reconstructed in the a_1^+ signal mode ($a_1^+ \to \rho^0\pi^+, \rho^0 \to \pi^+\pi^-$) offer a high-statistics control sample, which provides an opportunity to study the response of the BDT to data and MC in the signal region. These events are processed exactly like kaon events, except the cut from Section 4.3.2 which specifies that the recoiling B has a kaon daughter is changed to require the daughter be an a_1^+. $B^+ \to \tau^+\overline{\nu} \to a_1^+\nu\overline{\nu}$ is the only decay mode that, when correctly reconstructed, would produce a lone a_1^+, but this rare mode will only produce a fraction of an event in the real data set, compared with thousands of fakes. Most of the a_1^+'s in this sample are reconstruction fakes, with others coming from decays like $B \to \overline{D}^{*0}a_1^+$ (branching fraction $(1.9 \pm 0.5)\%$ [22]).
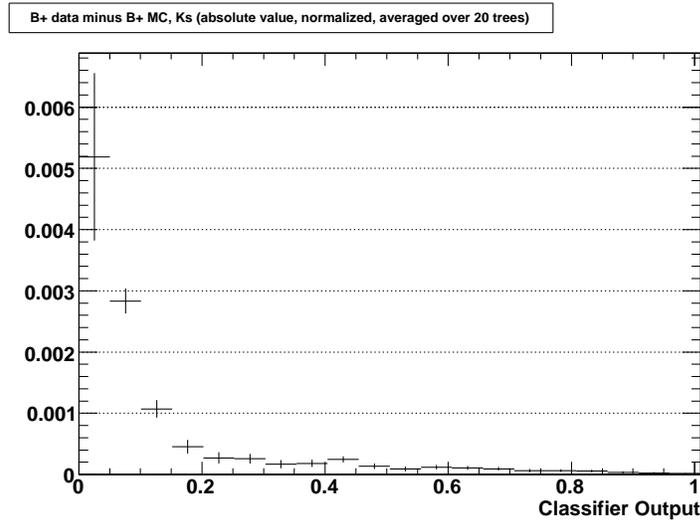
Figure 28: Normalized histogram of the difference in the classifier output for B^+ data and B^+ MC for the K_S trees. The plot shows the absolute values of the normalized differences, averaged over the twenty trees. The data/MC discrepancy in each bin is always less than 0.6% of the total number of events.
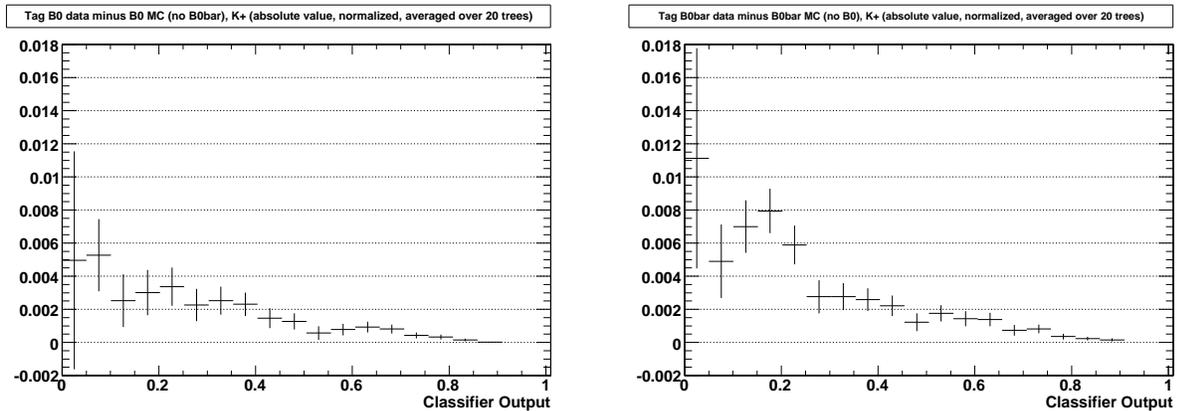
Figure 29: Normalized histograms of the difference in the classifier output for neutral B data and neutral B MC for the primary K^+ trees. The plots show the absolute values of the normalized differences, averaged over the twenty trees. The left plot is for B^0 only (no $\bar{B}^0$), while the right plot is for $\bar{B}^0$ only. The data/MC discrepancy in each bin is always less than 1.2% of the total number of events.

Figure 35 shows the data/MC classifier-output ratio for these events with the primary K^+ trees. The data/MC discrepancy is 1.05 ± 0.02, which is similar to the data/MC ratio in
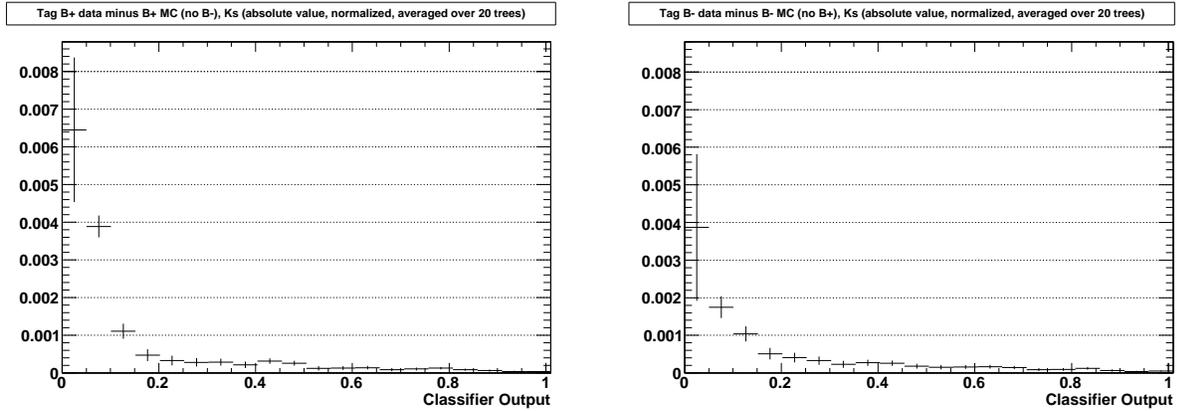
Figure 30: Normalized histograms of the difference in the classifier output for charged B data and charged B MC for the K_S trees. The plots show the absolute values of the normalized differences, averaged over the twenty trees. The left plot is for B^+ only (no B^-), while the right plot is for B^- only. The data/MC discrepancy in each bin is always less than 0.7% of the total number of events.
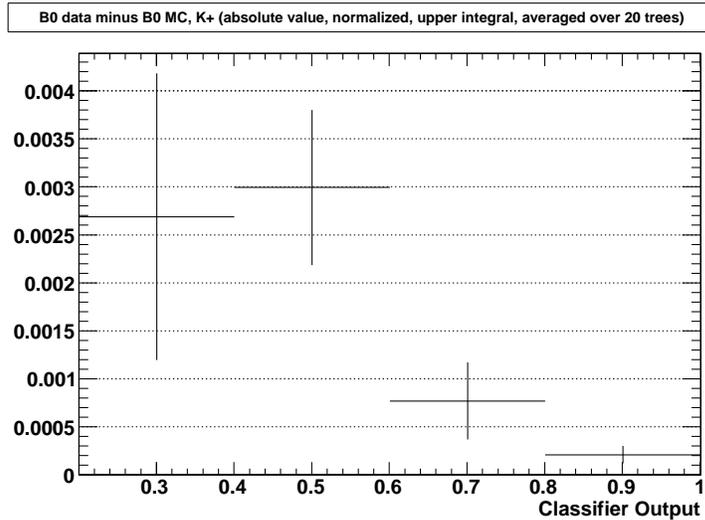
Figure 31: Plot of the difference in the integrated number of events for different classifier output values for B^0 data and B^0 MC for the K^+ trees. Each bin of the plot shows the average normalized absolute value of the difference in the number of events between B^0 data and B^0 MC for a range of classifier output values that goes from the number at the left edge of the bin to 1. The data/MC discrepancy in each bin is always less than 0.3% of the total number of events.
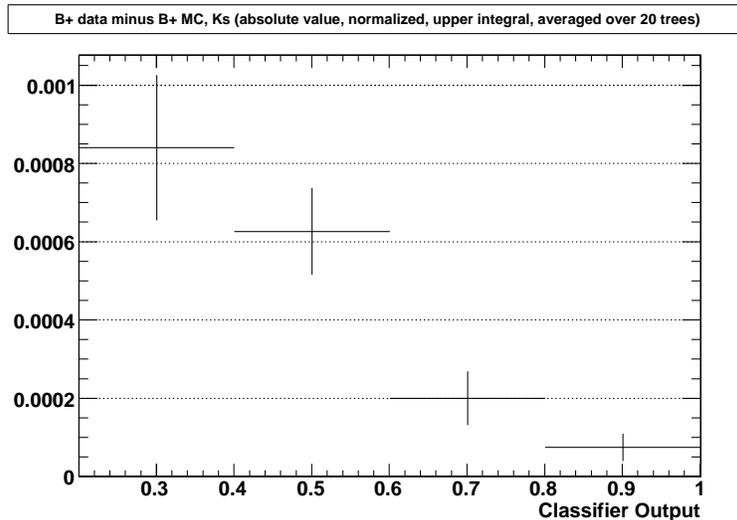
Figure 32: Plot of the difference in the integrated number of events for different classifier output values for B^+ data and B^+ MC for the K_S trees. Each bin of the plot shows the average normalized absolute value of the difference in the number of events between B^+ data and B^+ MC for a range of classifier output values that goes from the number at the left edge of the bin to 1. The data/MC discrepancy in each bin is always less than 0.1% of the total number of events.

the classifier-output sideband (Section 5.2.1). Figure 36 shows the similar ratio for the K_S trees.

6.3 J/ψ Events

The $B^+ \to K^+ J/\Psi, J/\Psi \to l^+ l^-$ mode (where the lepton pair can be either electrons or muons) provides a way to check the reliability of the signal efficiency of the trees. These events are processed almost like kaon events. To create highly pure $K J/\psi$ ntuples, the following cuts are used:

- The signal B must have two daughters, which must be a charged kaon and a J/Ψ.

- The signal B and tag lepton must have opposite charges.

- No more than two extra tracks are allowed. Extra tracks are those from the Charged-
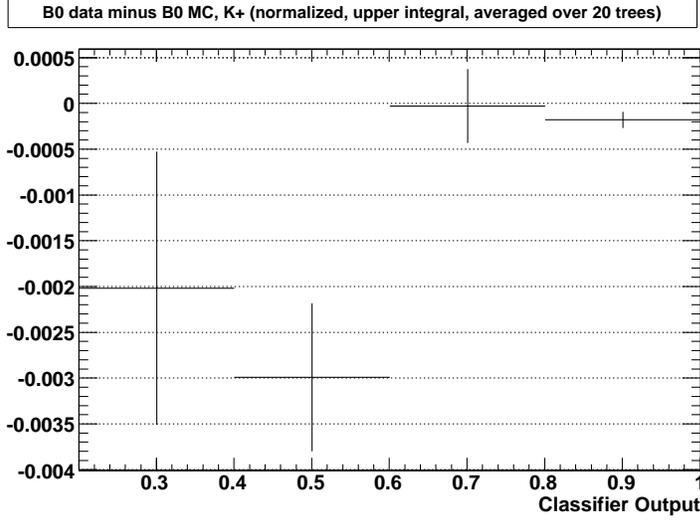
Figure 33: Plot of the difference in the integrated number of events for different classifier output values for B^0 data and B^0 MC for the K^+ trees. Each bin of the plot shows the average normalized signed value of the difference in the number of events between B^0 data and B^0 MC for a range of classifier output values that goes from the number at the left edge of the bin to 1. The data/MC discrepancy in each bin is always less than 0.35% of the total number of events.

Tracks list that are left over after the signal and tag are fully reconstructed.

- ΔE (Equation 14) must be in the following range: $-0.04 < \Delta E < 0.035\,\text{GeV}$.

- m_{ES} (Equation 15) must be in the following range: $5.2725 < m_{ES} < 5.285\,\text{GeV}$.

- m_{ll} (Equation 16) must be in the following range: $3.06 < m_{ll} < 3.12\,\text{GeV}$.

$$\Delta E = E_{SigB} - E_{beam}/2 \tag{14}$$

where E_{SigB} is the energy of the signal B and E_{beam} is the total beam energy, both in the center-of-mass frame.

$$m_{ES} = \sqrt{\frac{E_{beam}^2}{4} - p_{SigB}^2} \tag{15}$$
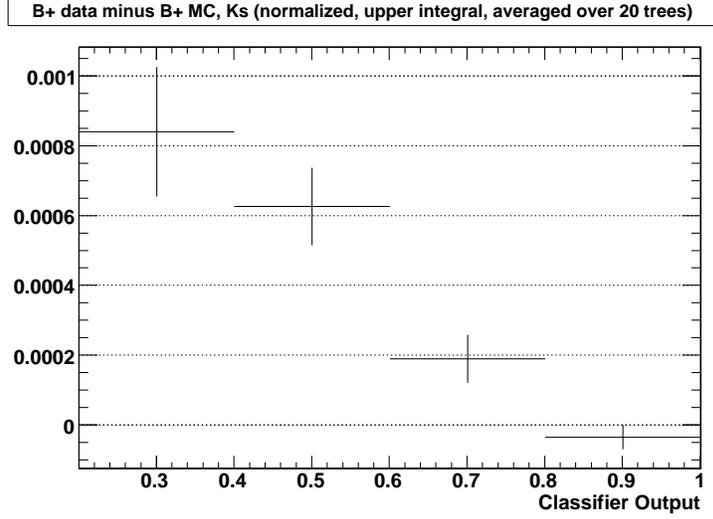
Figure 34: Plot of the difference in the integrated number of events for different classifier output values for B^+ data and B^+ MC for the K_S trees. Each bin of the plot shows the average normalized signed value of the difference in the number of events between B^+ data and B^+ MC for a range of classifier output values that goes from the number at the left edge of the bin to 1. The data/MC discrepancy in each bin is always less than 0.1% of the total number of events.
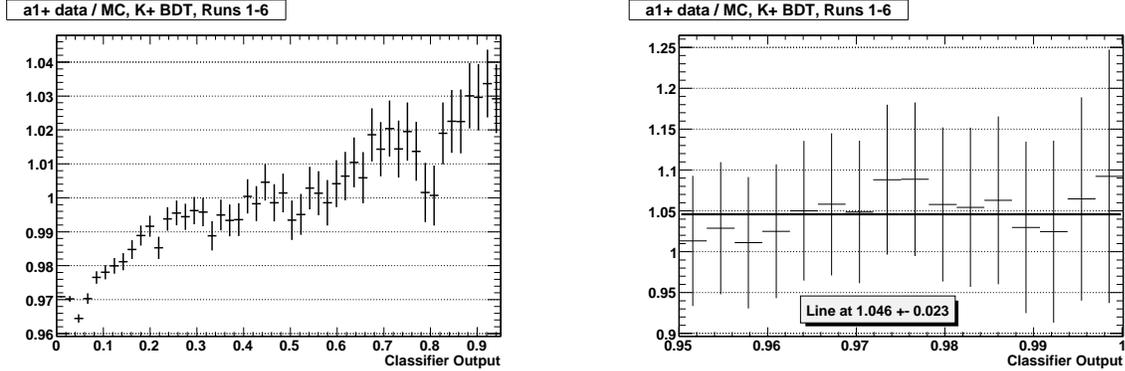
Figure 35: Data/MC ratio of the classifier output for a_1^+ events classified by the primary K^+ trees. The left plot show the full range of values up to the signal region, while the right plot shows the approximate signal region. On the right, the fitted line gives the value of the ratio as 1.05 ± 0.02.

where p_{SigB} is the three-momentum of the signal B and E_{beam} is the total beam energy, both in the center-of-mass frame.
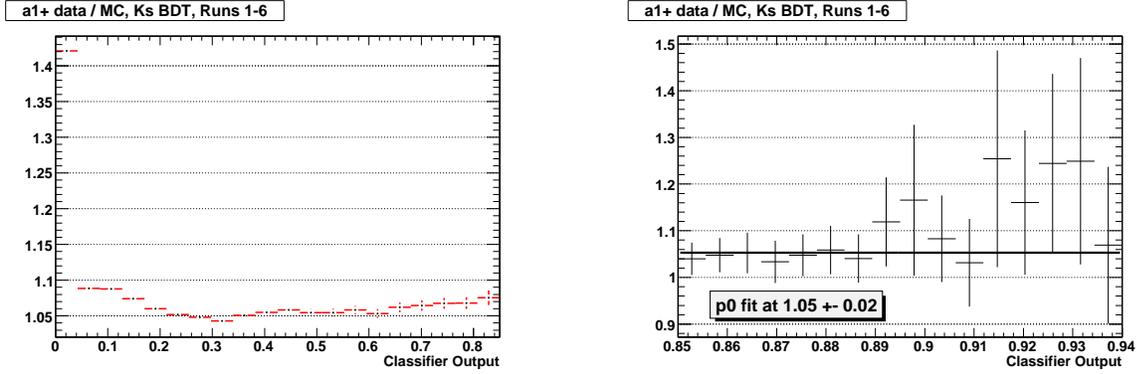
Figure 36: Data/MC ratio of the classifier output for a_1^+ events classified by the K_S trees. The left plot show the a range of values from zero to near the signal region, while the right plot shows the vicinity of the signal region. Almost no data events actually reach the signal region. On the right, the fitted line gives the value of the ratio as 1.05 ± 0.02.

$$m_{ll} = \sqrt{(p_{4l1} + p_{4l2})^2} \tag{16}$$

where p_{4l1} and p_{4l1} are the center-of-mass-frame four-momenta of the two leptons from the J/Ψ.

The result of the of the ΔE and m_{ES} cuts described above on the m_{ll} distribution in data is shown in Figure 37 .

After reconstruction and the cuts listed above, there are 211 J/Ψ data events and 52733 signal MC events, which correspond to a luminosity-weighted value of 216.9 events. The weighting is calculated using the PDG(2008) branching fraction values as follows: $B^+ \to K^+ J/\Psi$ BF = $(1.007 \pm 0.035) \cdot 10^{-3}$, $J/\Psi \to e^+ e^-$ BF = $(5.94 \pm 0.06)\%$, $J/\Psi \to \mu^+ \mu^-$ BF = $(5.93 \pm 0.06)\%$. The event numbers give a data/MC ratio of 0.97 ± 0.07, which indicates very little discrepancy between data and MC events in skimming and signal and tag reconstruction. Figure 38 shows the close agreement of data and MC.

Unfortunately, the number of available J/Ψ data events is too low to get sufficient statistics in the signal region when the events are classified with the K^+ or K_S trees. To solve this
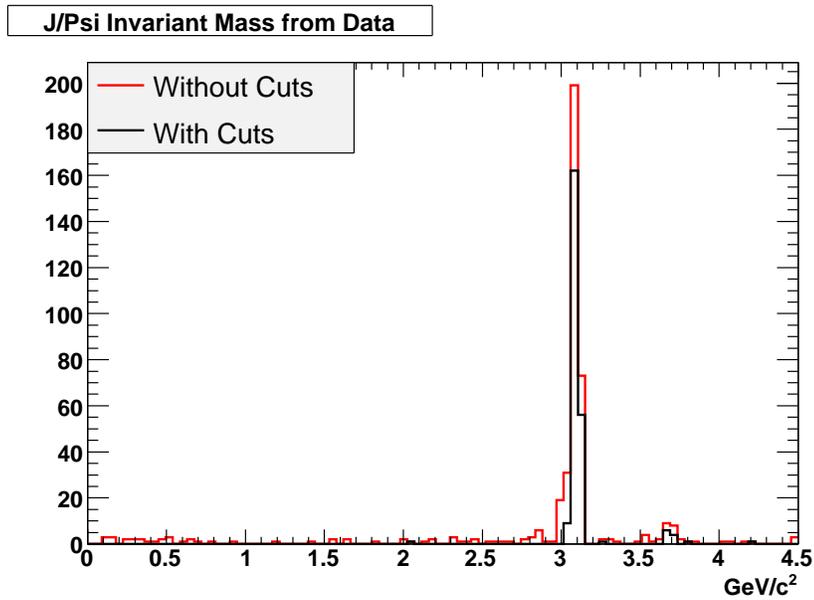
Figure 37: m_{ll} distribution for J/Ψ data events before and after the ΔE and m_{ES} cuts described in Section 6.3 are applied. All the other cuts in that section, except the m_{ll} cuts, are already applied to all these events.
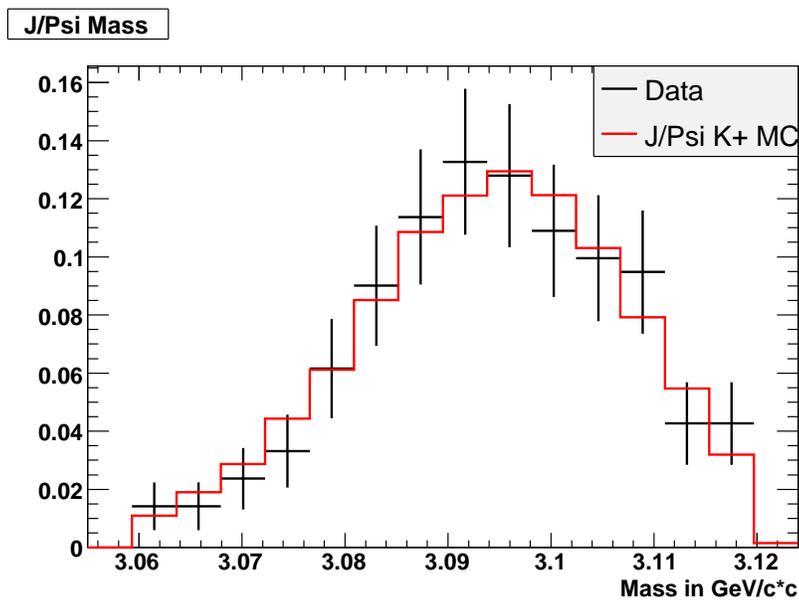
Figure 38: m_{ll} distribution for J/Ψ data and MC events after the cuts described in Section 6.3 are applied. Data and MC show close agreement.

problem, a less-discriminating subset tree is used, which utilizes a subset of the classification variables used in the other two trees. The list of classification variables used by the subset tree is shown in Table 10. The subset tree should reflect the behavior of the K^+ and K_S trees, but it allows study of J/Ψ data events in the same region where $K^+\nu\bar{\nu}$ or $K_S\nu\bar{\nu}$ events would be concentrated.

Classifying the above events with the subset trees and plotting the data/MC ratio vs. the classifier output results in Figure 39. The plot shows a data/MC discrepancy of 0.90 ± 0.03 in the signal region, which suggests that the systematic differences between data and MC in skimming, tagging, reconstruction, and BDT classification amounts to a 10% effect on the signal efficiency. Thus, 10% will be used as the total systematic uncertainty to the signal efficiency.
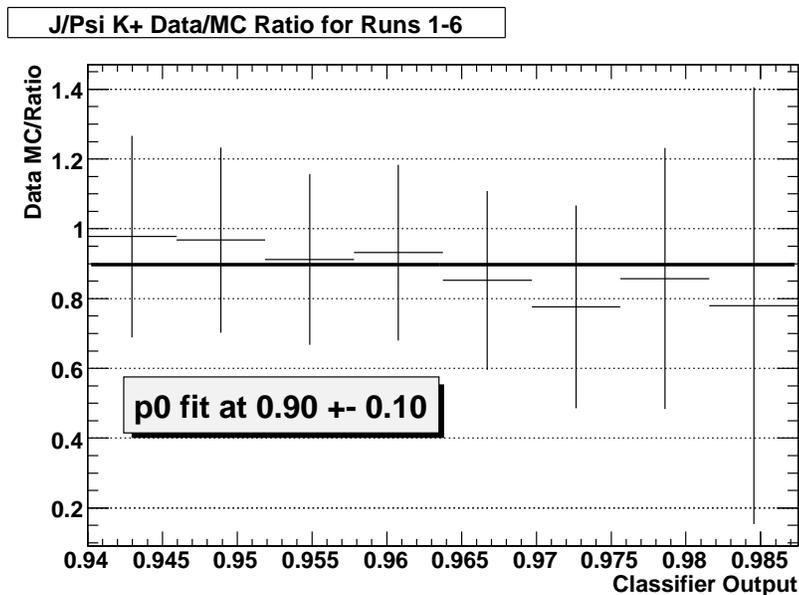
Figure 39: Data/MC ratio for J/Ψ events classified by the subset trees. The fitted line gives the value of the ratio as 0.90 ± 0.10. The plot covers only the approximate signal region in order to get a conservative estimate of the systematic uncertainty. The data/MC agreement is greater outside of the signal region. This plot is obtained by summing the classifier output histograms from the twenty trees for data and dividing by the corresponding summed histograms for luminosity-weighted MC.

7 Systematic Uncertainties

7.1 Signal Efficiency Systematic Uncertainty

As discussed in Section 6.3, processing highly pure J/Ψ events provides a way to validate the signal estimates in this analysis. Signal and tagging efficiency are inextricably linked in this analysis since the BDTs use many tag-side variables. In addition, kaon PID is part of the signal efficiency. Thus, the J/Ψ study covers the systematic uncertainty for signal efficiency, tagging efficiency, and kaon PID, and it gives the value of the uncertainty as 10%.

7.1.1 Theory Systematic Uncertainty to the Signal Efficiency

For values calculated using the ABSW or BHI models (see Section 5.2.4), a theory systematic uncertainty is necessary. Values based upon the phase-space model have no theory uncertainty included, but these values are for illustrative purposes only. As mentioned in Section 5.2.4, the BHI model reduces signal efficiency to 92% compared to phase space, while ABSW reduces it to 81%. Thus, BHI has a range of $\pm 9\%$ with regard to the other two models, while BHI is 14% above ABSW in terms of signal efficiency. Figure 2 shows that the theoretical uncertainties of each model are significantly smaller than the differences between the models. A 10% theory systematic uncertainty for each model conservatively covers the theoretical uncertainties shown in Figure 2. Added in quadrature with the signal efficiency systematic above yields a total signal efficiency systematic uncertainty of 14%.

7.2 Background Systematic Uncertainty

7.2.1 Primary K^+ Mode Background Systematic Uncertainty

The 2-D weighting described in Section 5.2.1 provides a correction to the background prediction for the K^+ mode. This correction provides a measure of how far off the MC background prediction may be from the true background in data, and thus, it will be used

as the systematic uncertainty to the background prediction. Without the 2-D weighting, the background prediction is 16.7 events; with the weighting, it is 17.6. Thus, the correction is 0.9 events, or 5%, and this value will be used as the systematic uncertainty to the background prediction for the primary K^+ mode.

7.2.2 Low-momentum Mode Background Systematic Uncertainty

The 2-D weighting described in Section 5.2.1 provides a correction to the background prediction for the low-momentum K^+ mode. This correction provides a measure of how far off the MC background prediction may be from the true background in data, and thus, it will be used as the systematic uncertainty to the background prediction. Without the 2-D weighting, the background prediction is 141 events; with the weighting, it is 187. Thus, the correction is 46 events, or 25%, and this value will be used as the systematic uncertainty to the background prediction for the low-momentum K^+ mode.

7.2.3 K_S Mode Background Systematic Uncertainty

Similar to the primary K^+ mode, the K_S mode has a correction applied to the background prediction, as discussed in Section 5.2.3. This correction will be used as the systematic uncertainty to the background prediction, for the same reasons as discussed in Section 7.2.1. For this mode, the uncertainty is 10%.

7.3 B-counting Systematic Uncertainty

As shown in Table 1, the number of $B\overline{B}$ pairs includes an uncertainty: $459.0\pm5.1 \times 10^6$. This uncertainty is 1.1%. For the calculation of the sensitivity of the branching fraction measurement, this uncertainty must be added in quadrature with the signal efficiency systematic uncertainty from Section 7.1. Since the signal efficiency systematic uncertainty is so much larger than the B-counting uncertainty, the addition has no real effect, and the total

systematic uncertainty in the sensitivity turns out equal to the signal efficiency systematic uncertainty (14%).

8 Results

8.1 Observations and Upper Limits

Table 14 shows the number of data events found in the signal region after unblinding for each of the three measurements. The table reflects the statistical uncertainties, calculated by the method in Appendix C, with the uncertainty for a single tree being the luminosity-weighted number of events over the square root of the number of MC events. The uncertainties for the numbers of excess events shown in Table 14 are calculated with the Barlow calculator [26], which is a tool that uses frequentist statistics. These uncertainties are obtained by finding the lower and upper event-number limits at the 84.1% confidence level, for a two-sided one-sigma uncertainty, and by finding the 90%-confidence-level upper limit. The probability column in the table is also calculated with the Barlow calculator by finding the confidence level for a lower limit of 0 events (or upper limit in the case of the low-momentum result since the number of excess events is negative).

Table 15 shows the resulting central values of the branching fractions, including partial branching fractions.

Tables 16 to 18 show the resulting branching fraction upper limits. The three signal models used in this analysis (phase space, ABSW, and BHI) produce different upper limits because the models differ in the distribution of signal events between the low- and high-kaon-momentum regions, as shown in Figure 2. These different distributions give rise to different signal efficiencies (see Table 13), which result in different upper limits. Note that upper limits for the results with a fractional number of observed events are interpolated from the bounding integer values. For example, the primary K^+ mode has 19.4 observed events, so its upper limits are interpolated from the limits for 19 and 20 events.

Figures 40 and 41 show histograms of the classifier output for the data events.

Table 14: Number of events in the signal region. The second column gives the number of events observed in data with the statistical uncertainty. The background column includes the statistical uncertainty, followed by the systematic, with the total uncertainty in parentheses. The excess events column includes the two-sided, one-sigma uncertainty followed by the 90%-confidence-level-upper-limit uncertainty in parentheses. The probability column shows the probability the excess events could be attributed entirely to a background fluctuation.

	Observed	Background	Excess Events	Probability
K^+	19.4 ± 4.4	$17.6 \pm 2.6 \pm 0.9(2.8)$	$1.8^{+6.2}_{-5.1}(+8.0)$	38%
Low-momentum K^+	164 ± 13	$187 \pm 10 \pm 46(47)$	$-23^{+49}_{-48}(+63)$	33%
K_S	$6.1^{+4.0}_{-2.2}$	$3.9 \pm 1.3 \pm 0.4(1.4)$	$2.2^{+4.1}_{-2.8}(+5.0)$	23%

Table 15: Central values for the branching fractions of each mode under the three signal models.

	Phase-space Model	BHI Model	ABSW Model
K^+	$(0.23^{+0.78}_{-0.64}) \times 10^{-5}$	$(0.24^{+0.84}_{-0.69}) \times 10^{-5}$	$(0.28^{+0.96}_{-0.79}) \times 10^{-5}$
K^0	$(1.6^{+2.9}_{-2.0}) \times 10^{-5}$	$(1.7^{+3.1}_{-2.1}) \times 10^{-5}$	$(2.0^{+3.7}_{-2.5}) \times 10^{-5}$
High-mom. K^+ partial BF	$(0.16^{+0.54}_{-0.44}) \times 10^{-5}$	$(0.16^{+0.56}_{-0.46}) \times 10^{-5}$	$(0.17^{+0.57}_{-0.47}) \times 10^{-5}$
Low-mom. K^+ partial BF	$(-1.9^{+4.1}_{-4.0}) \times 10^{-5}$	$(-1.8^{+3.8}_{-3.8}) \times 10^{-5}$	$(-2.0^{+4.3}_{-4.2}) \times 10^{-5}$

Table 16: Upper limits under the ABSW model [8] on the branching fractions for the primary K^+ mode, the K^0 mode, and the combination of the primary K^+ and K^0 modes, plus the low-momentum and high-momentum K^+ partial branching fractions. The limits are shown at the 90% and 95% confidence levels.

CL	K^+	K^0	K^+ & K^0	Low-mom. K^+	High-mom. K^+
90%	1.5×10^{-5}	6.5×10^{-5}	1.6×10^{-5}	3.5×10^{-5}	0.90×10^{-5}
95%	1.9×10^{-5}	7.9×10^{-5}	1.9×10^{-5}	5.1×10^{-5}	1.1×10^{-5}

Table 17: Upper limits under the BHI model [7] on the branching fractions for the primary K^+ mode, the K^0 mode, and the combination of the primary K^+ and K^0 modes, plus the low-momentum and high-momentum K^+ partial branching fractions. The limits are shown at the 90% and 95% confidence levels.

CL	K^+	K^0	K^+ & K^0	Low-mom. K^+	High-mom. K^+
90%	1.3×10^{-5}	5.6×10^{-5}	1.4×10^{-5}	3.1×10^{-5}	0.89×10^{-5}
95%	1.6×10^{-5}	6.7×10^{-5}	1.7×10^{-5}	4.6×10^{-5}	1.1×10^{-5}

Table 18: Upper limits under the phase-space model on the branching fractions for the primary K^+ mode, the K^0 mode, and the combination of the primary K^+ and K^0 modes, plus the low-momentum and high-momentum K^+ partial branching fractions. The limits are shown at the 90% and 95% confidence levels.

CL	K^+	K^0	K^+ & K^0	Low-mom. K^+	High-mom. K^+
90%	1.2×10^{-5}	5.2×10^{-5}	1.3×10^{-5}	3.4×10^{-5}	0.85×10^{-5}
95%	1.5×10^{-5}	6.2×10^{-5}	1.6×10^{-5}	4.9×10^{-5}	1.1×10^{-5}
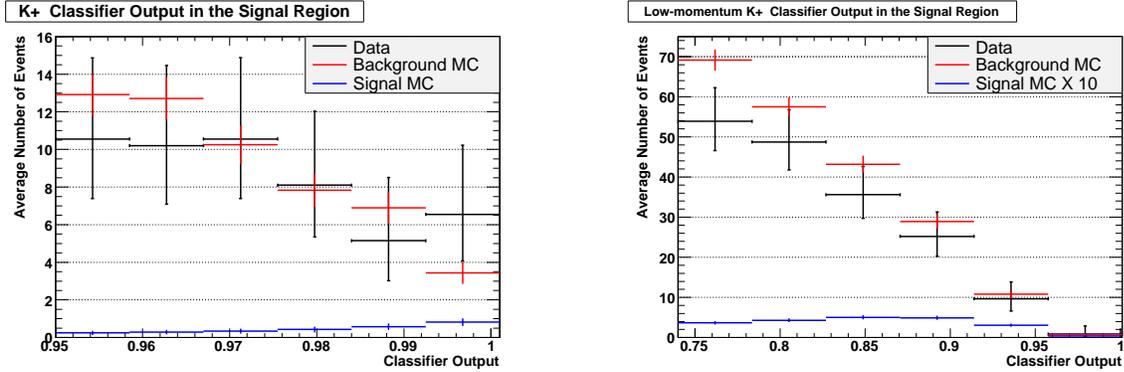
Figure 40: Data, background MC, and signal MC events in the K^+ signal region. The left plot shows the primary signal region, with signal MC scaled to the number of excess events observed. The right shows the low-momentum signal region, with the signal MC scaled arbitrarily for illustration. These plots show the average number of events for the twenty trees. Both plots include some sideband below the signal region.

8.2 Comparison with Previous Experiments

The BHI model provides the baseline for comparison with previous experiments. Figure 42 compares the results of this analysis with the previous limits, which came from the
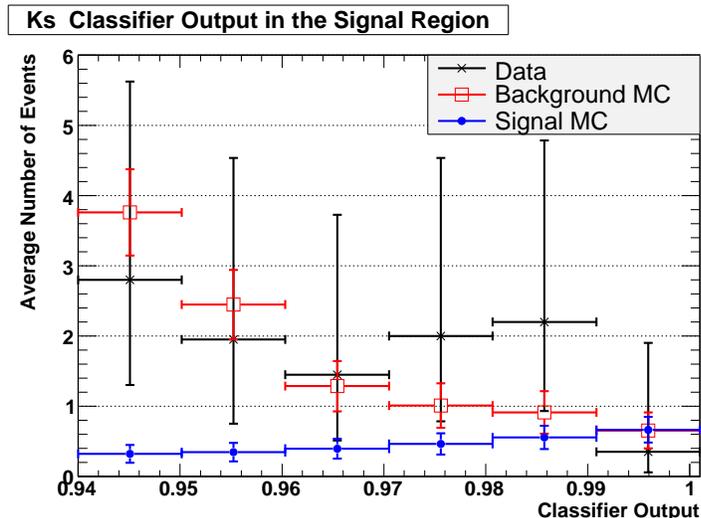
Figure 41: Data, background MC, and signal MC events in the K_S signal region, with signal MC scaled to the number of excess events observed. The plot shows the average number of events for the twenty trees, and it includes some sideband below the signal region.

previous best BABAR result [27] and the Belle experiment [9]. This analysis improves upon the previous BABAR result by almost a factor of three, provides a slight improvement over Belle to the 90%-confidence-level upper limit for the $B_u^+ \to K^+ \nu\bar{\nu}$, but reduces the upper limit for the $B_d^0 \to K^0 \nu\bar{\nu}$ mode by a factor of three, compared with Belle.

8.3   Comparison with Theoretical Models

In terms of the theoretical branching fraction under the BHI model of 3.8×10^{-6}, the $B_u^+ \to K^+ \nu\bar{\nu}$ upper limit of this analysis almost comes within three times of the theoretical value (the situation is similar with the ABSW model).

For one model of scalar WIMP dark matter [20], these results, like those of Belle, exclude dark matter particles with a mass below about 1.8 GeV/c^2, as shown in Figure 43. In the case of the other models mentioned in Section 2 (unparticle models [16], MSSM at large tan β [17], models with a single universal extra dimension [18], and WIMP-less dark matter [19]), no significant new constraints are placed.
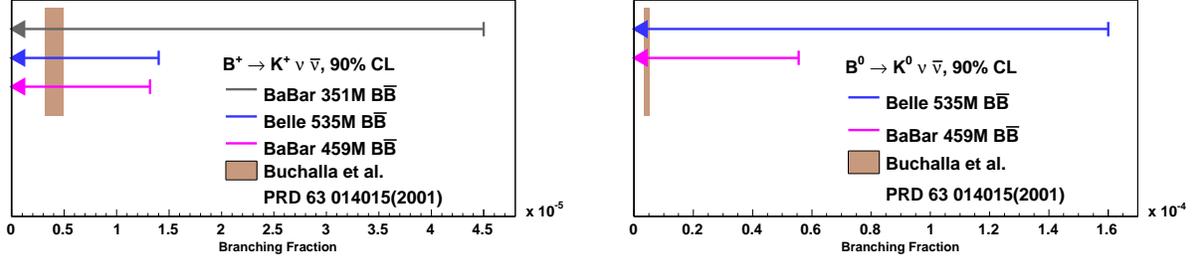
Figure 42: Comparison of BaBar and Belle branching fraction results for the decay modes in this analysis. Belle's best results are shown in blue, and the results of this analysis are shown in magenta just below. The brown box shows the theoretical prediction of the BHI model. The left plot shows the primary K^+ mode, and the right shows the K^0 mode.
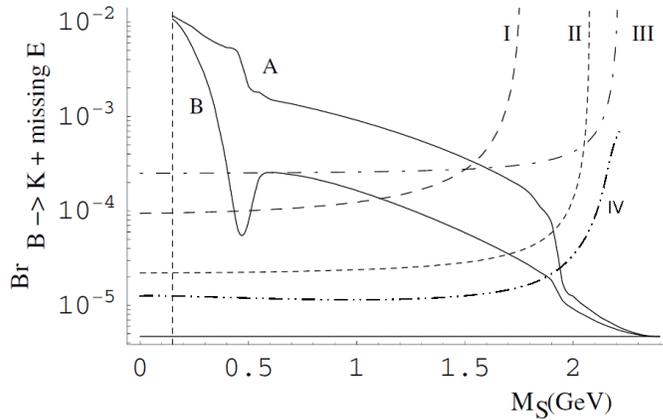
Figure 43: Branching fraction of $B \to K + missing\ energy$ vs. the mass of a scalar dark matter particle described by [20]. The model's allowed region is between curves A and B. Curve I is the limit set by an early *BABAR* $B_u^+ \to K^+\nu\overline{\nu}$ result with 82 fb^{-1} [28]. Curve II is an estimated limit projected for 250 fb^{-1} of *BABAR* data. Curve III is the limit set by a CLEO result [29]. Curve IV is the approximate limit set by the current result, which uses 418 fb^{-1}.

With regard to the $K \to \pi\nu\overline{\nu}$ mode mentioned in Section 1, the lack of any new physics signals in the results of this analysis diminish the likelihood of observing new physics in the very similar $K \to \pi\nu\overline{\nu}$ mode.

In summary, the results of this analysis show consistency with the Standard Model, confirm and extend the Belle results [9], and give no indication of new physics.

9 Conclusion and Outlook

The summary of this analysis is shown in Table 19. All the upper limits in the table are based upon the BHI model [7] for consistency. A graphical comparison of these results with the previous best Belle results [9] are shown in Figure 42.

Table 19: Summary of results. The second column gives the number of events observed in data with the statistical uncertainty. The background column includes the total uncertainty. Note the high-momentum K^+ measurement has the same number of observed and background events as the primary mode. The fourth column gives the 90%-confidence-level upper limit to the branching fraction, or partial branching fraction for the low- and high-momentum measurements. The last column shows the corresponding best upper limits reported by the Belle experiment [9]. All the upper limits in this table are based upon the BHI model [7].

Mode	Observed	Background	90% CL UL BF	Belle UL
Primary K^+	19.4±4.4	17.6±2.8	1.3×10^{-5}	1.4×10^{-5}
High-mom. K^+	Same ↑	Same ↑	0.89×10^{-5}	N/A
Low-mom. K^+	164±13	187±47	3.1×10^{-5}	N/A
$K^0(K_S)$	$6.1^{+4.0}_{-2.2}$	3.9±1.4	5.6×10^{-5}	16×10^{-5}

In the future, with the full 50 ab^{-1} data set from the proposed SuperB factory at Frascati, it should be possible to actually make a 5σ observation of the $B_u^+ \to K^+ \nu\bar{\nu}$ decay by combining results using semileptonic tags with those using hadronic tags. The $B_d^0 \to K_S^0 \nu\bar{\nu}$ mode might be more difficult to observe but might be achievable with improvements in background rejection.

Appendices

A   Additional Figures

A.1   Signal/Background Plots for Classification Variables

The following are histograms of signal and generic background MC for each classification variable. Signal and background are each unit-normalized. These MC events constitute the training and testing sets for the decision trees.

A.1.1   Background Divided by Background Type

K^+ **mode**   .
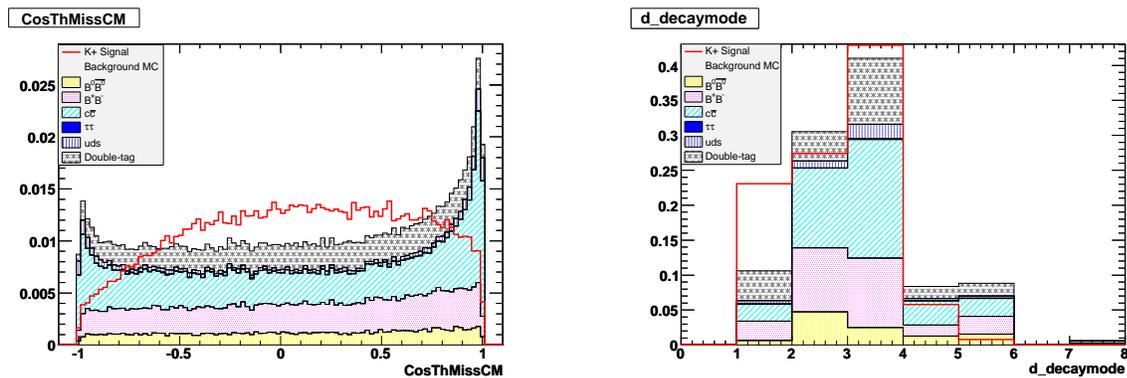
Figure 44: Signal/background histograms for CosThMissCM on left and d_decaymode on right.
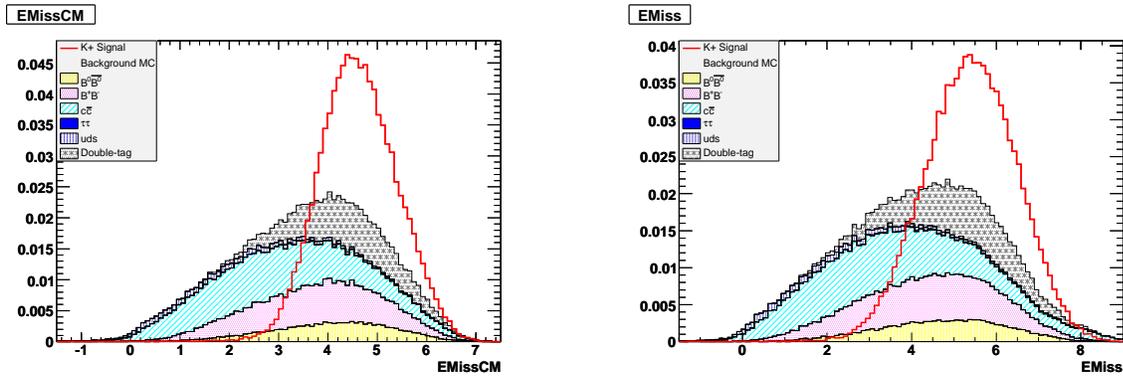
Figure 45: Signal/background histograms for EMissCM on left and EMiss on right.
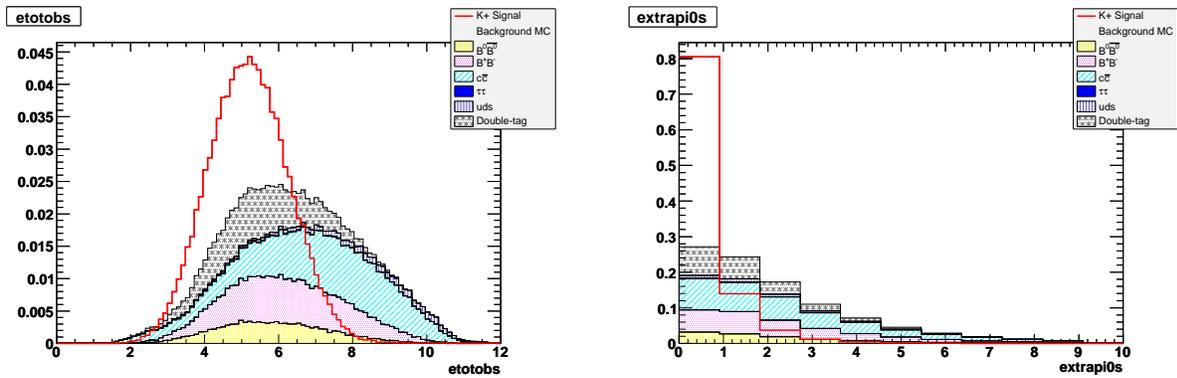
Figure 46: Signal/background histograms for etotobs on left and extrapi0s on right.
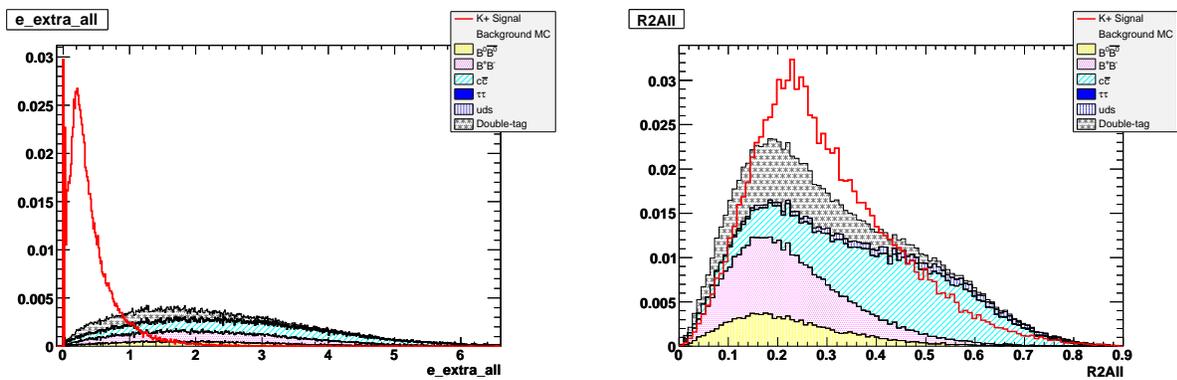
Figure 47: Signal/background histograms for e_extra_all on left and R2All on right.
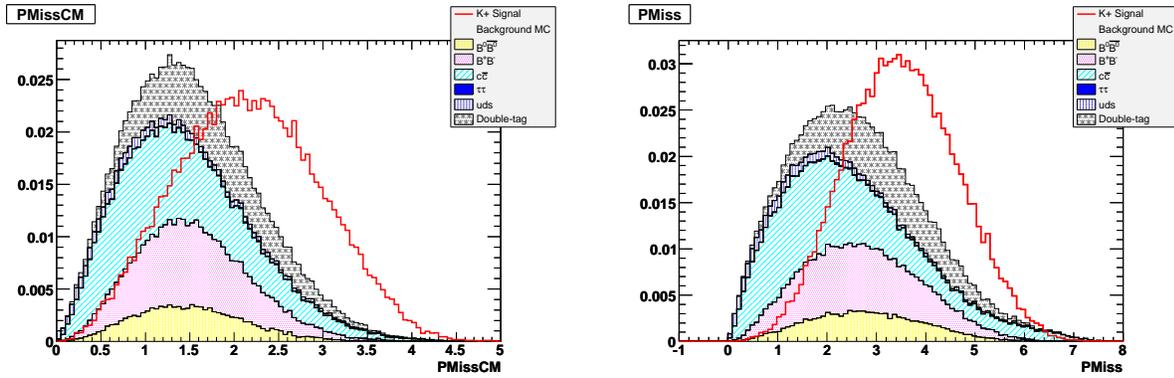
Figure 48: Signal/background histograms for PMissCM on left and PMiss on right.
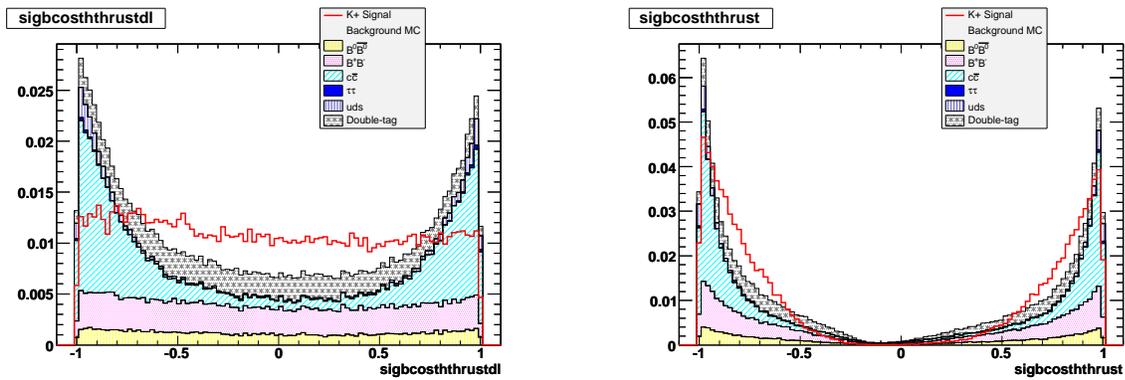
Figure 49: Signal/background histograms for sigbcosththrustdl on left and sigbcosththrust on right.
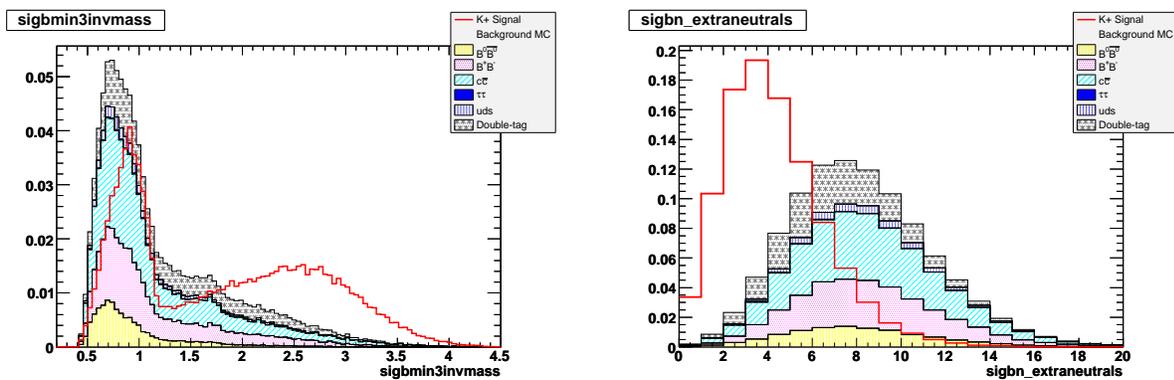
Figure 50: Signal/background histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.
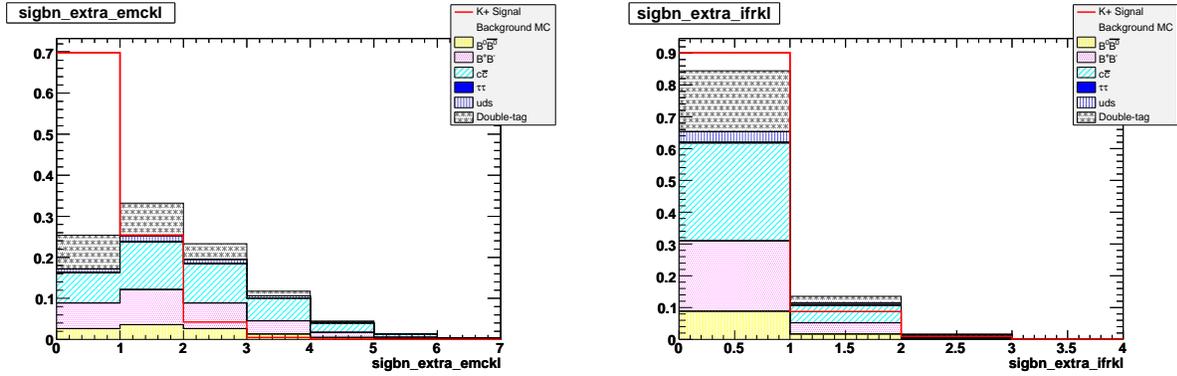
Figure 51: Signal/background histograms for sigbn_extra_emckl on left and sigbn_extra_ifrkl on right.
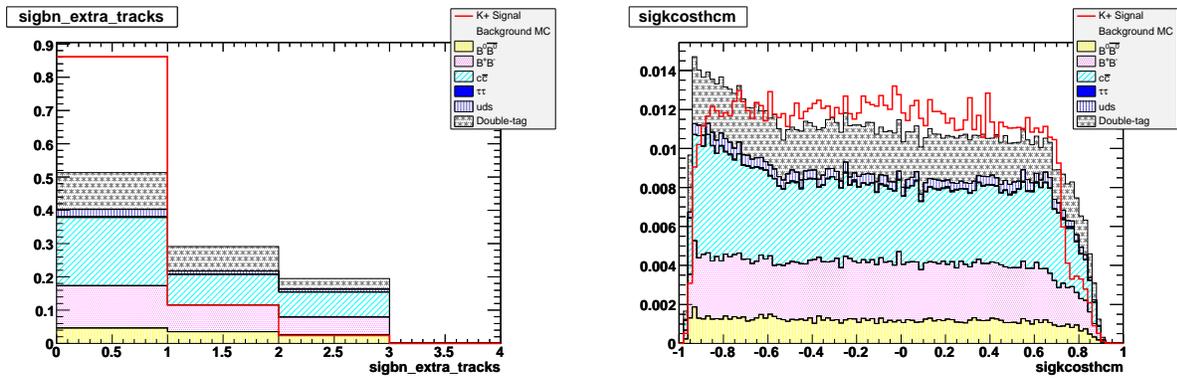
Figure 52: Signal/background histograms for sigbn_extra_tracks on left and sigkcosthcm on right.
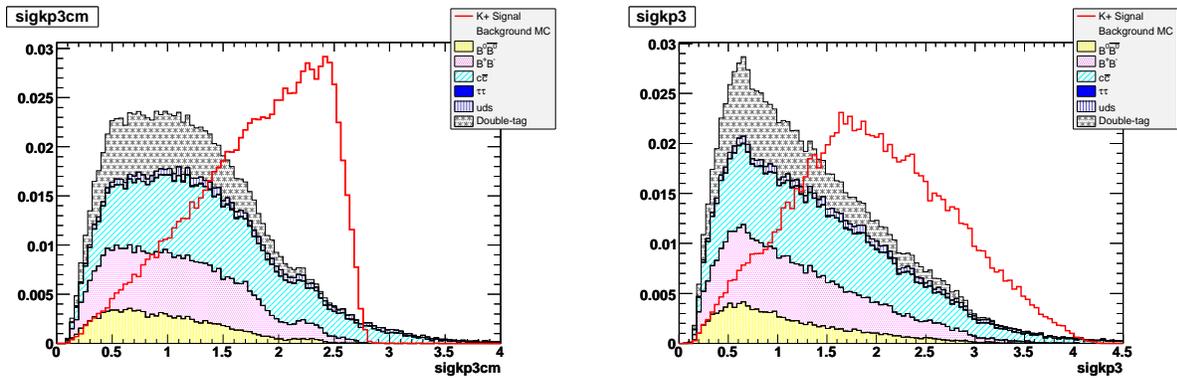
Figure 53: Signal/background histograms for sigkp3cm on left and sigkp3 on right.
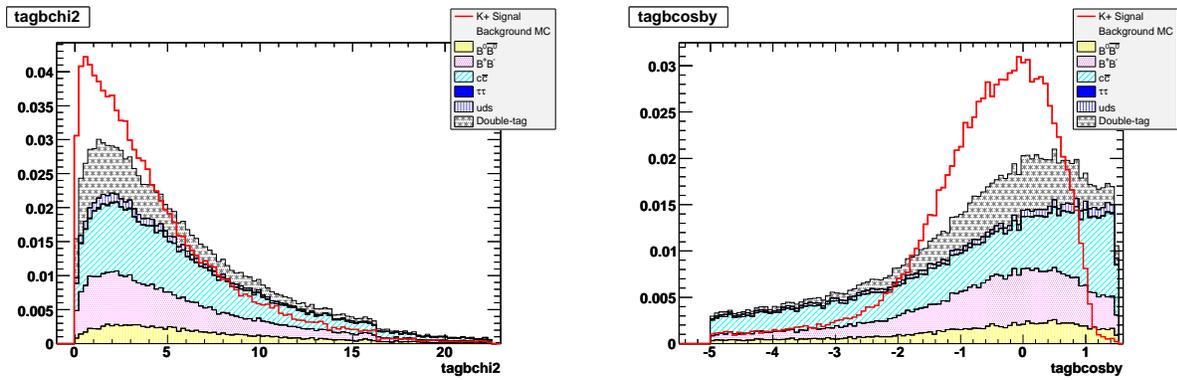
Figure 54: Signal/background histograms for tagbchi2 on left and tagbcosby on right.
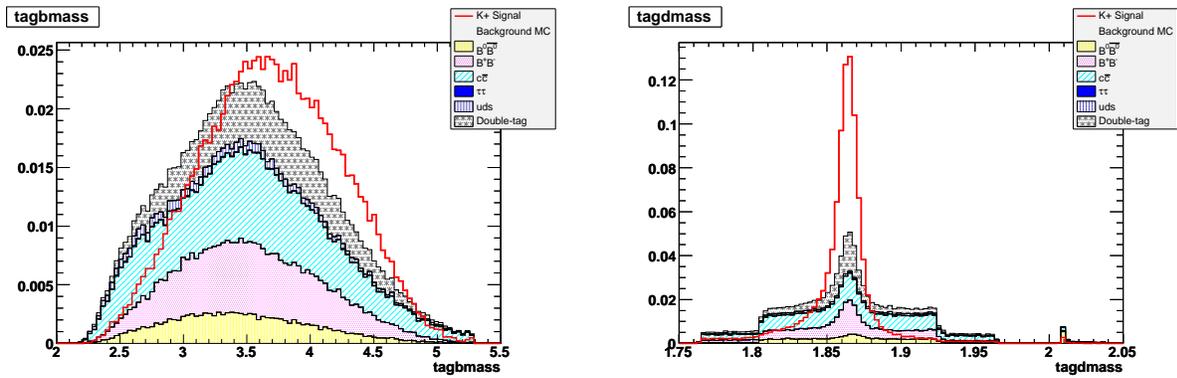
Figure 55: Signal/background histograms for tagbmass on left and tagdmass on right.
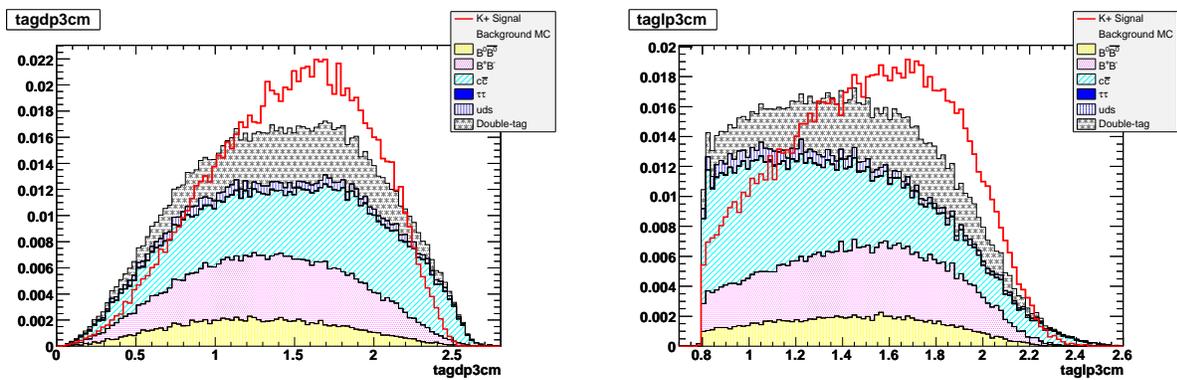
Figure 56: Signal/background histograms for tagdp3cm on left and taglp3cm on right.
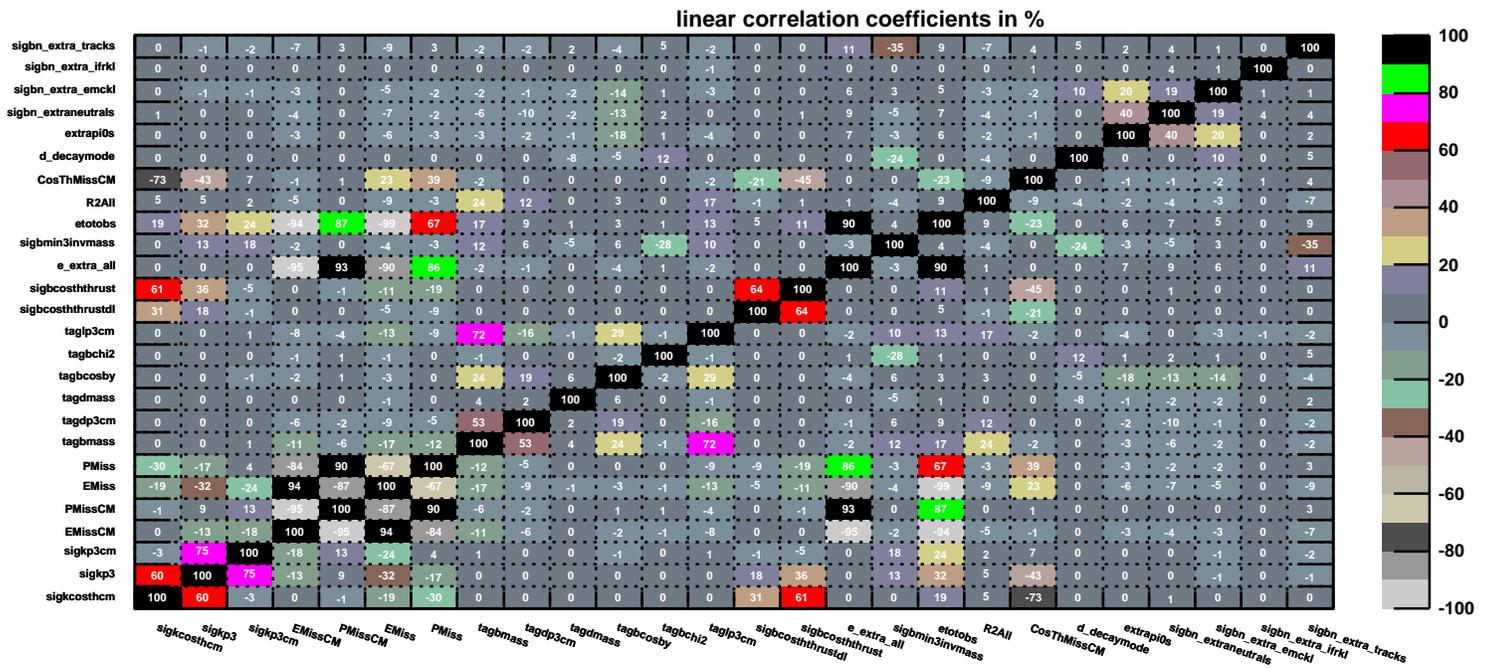
Figure 57: Correlation matrix showing the amount of correlation in signal MC between the classification variables for the K^+ trees.

K_S **Mode** .
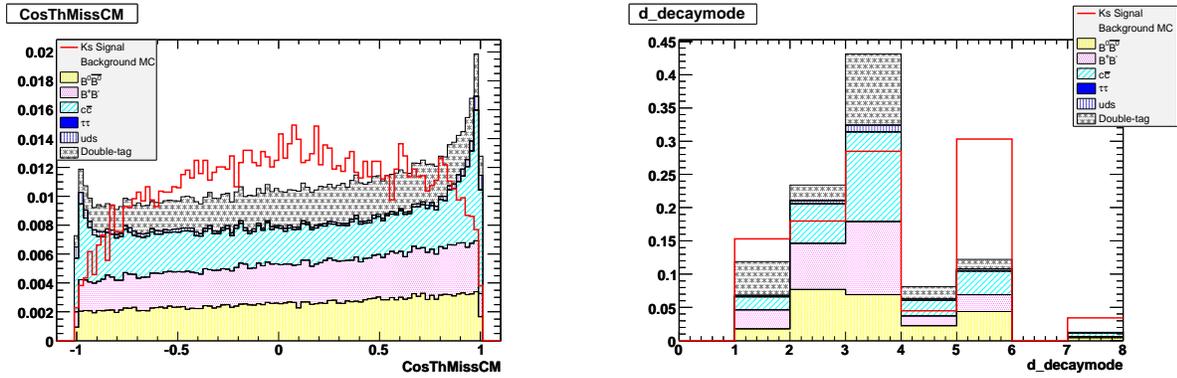
Figure 58: Signal/background histograms for CosThMissCM on left and d_decaymode on right.
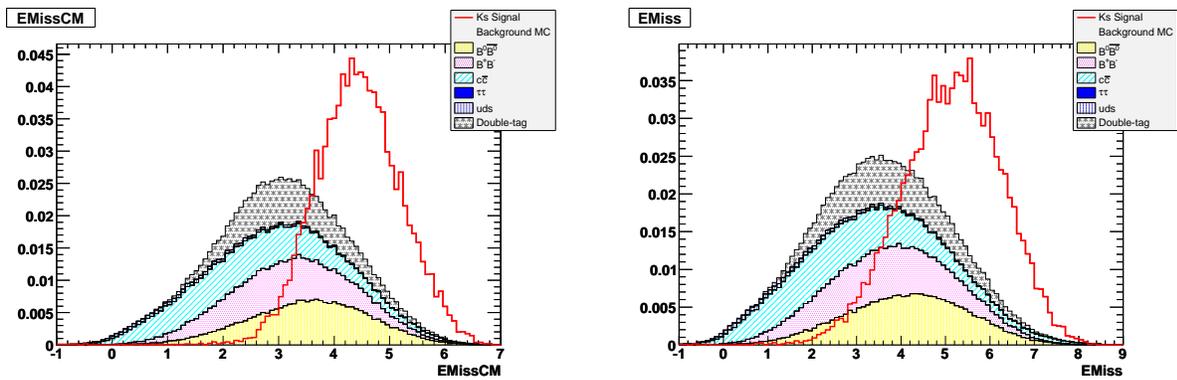
Figure 59: Signal/background histograms for EMissCM on left and EMiss on right.
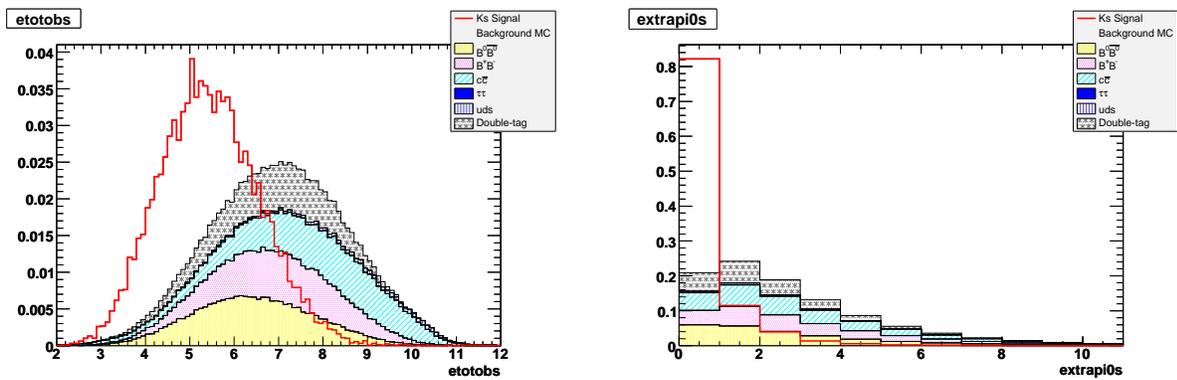
Figure 60: Signal/background histograms for etotobs on left and extrapi0s on right.
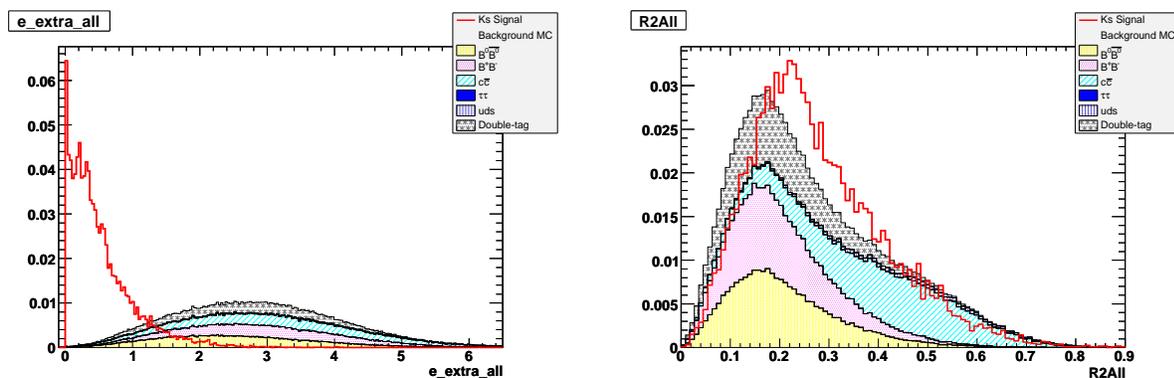
Figure 61: Signal/background histograms for e_extra_all on left and R2All on right.
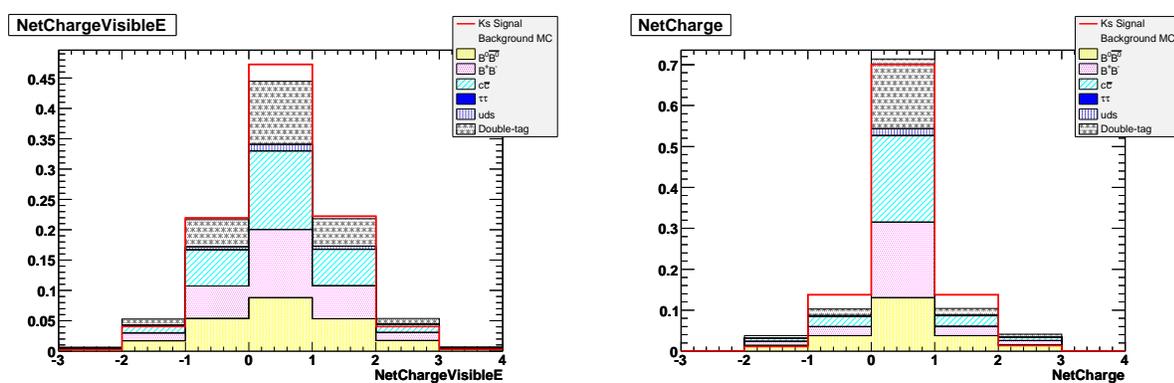
Figure 62: Signal/background histograms for NetChargeVisibleE on left and NetCharge on right.
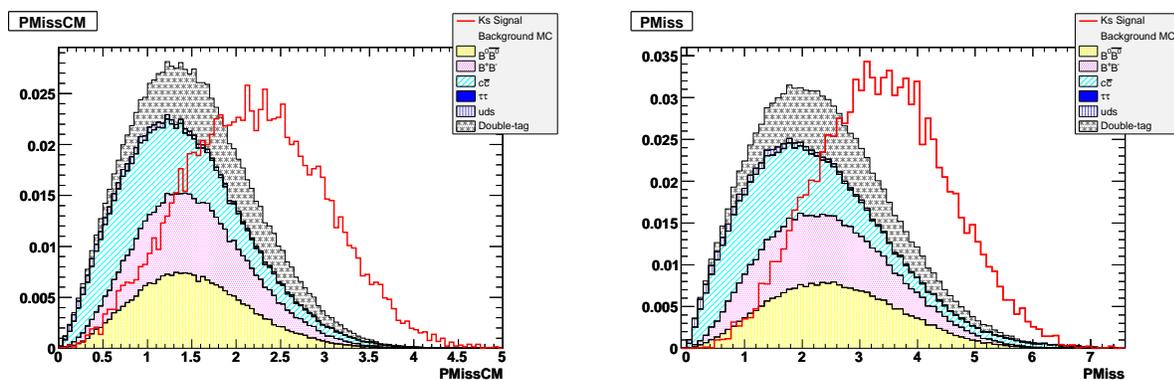
Figure 63: Signal/background histograms for PMissCM on left and PMiss on right.
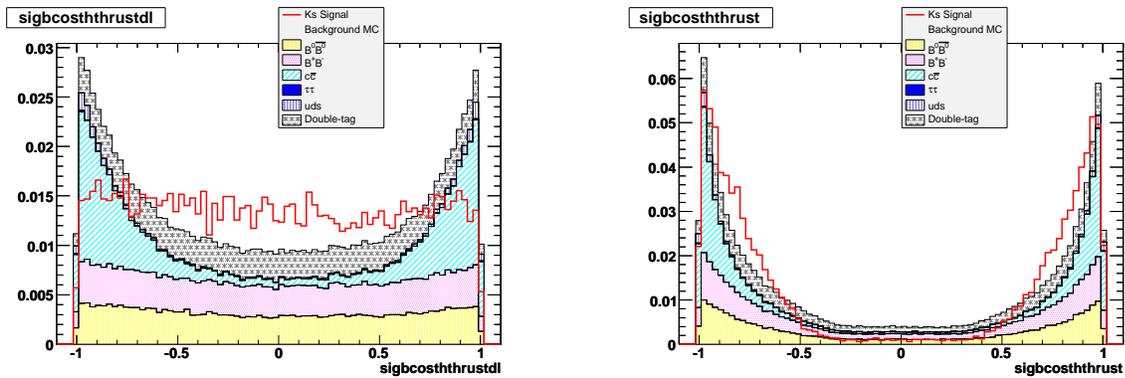
Figure 64: Signal/background histograms for sigbcoststhrustdl on left and sigbcoststhrust on right.
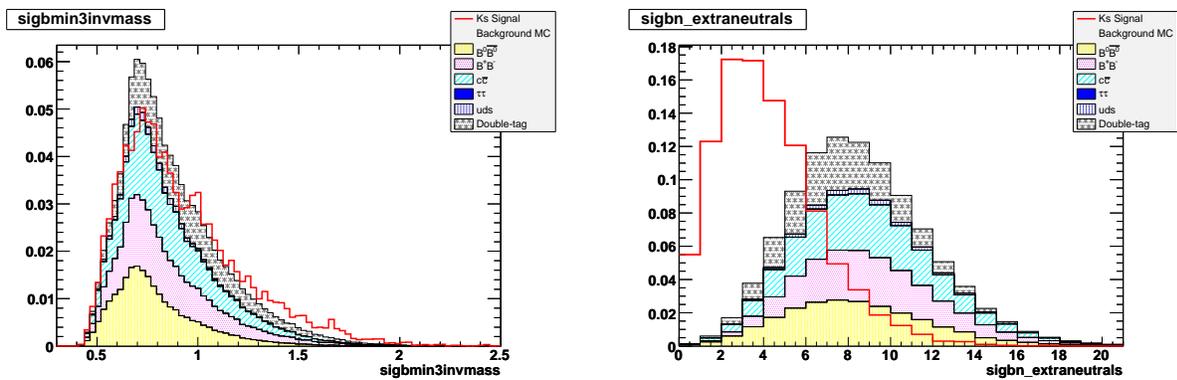
Figure 65: Signal/background histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.
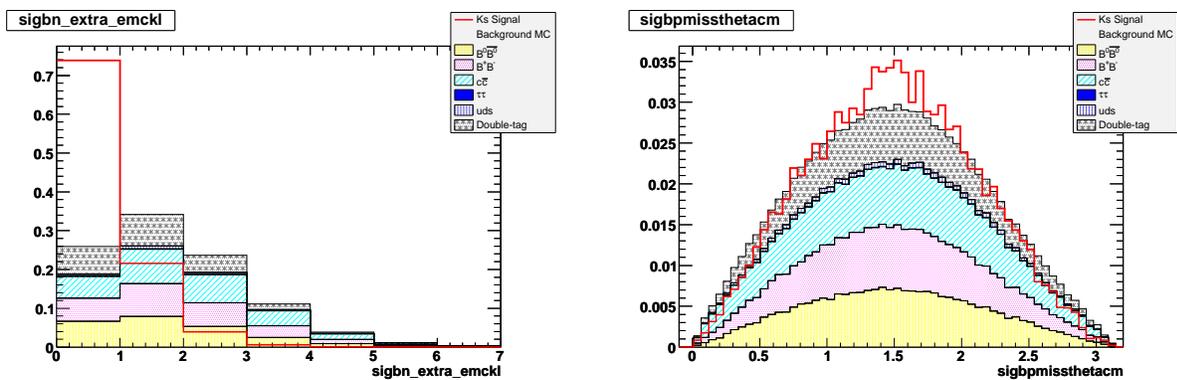
Figure 66: Signal/background histograms for sigbn_extra_emckl on left and sigbpmissthetacm on right.
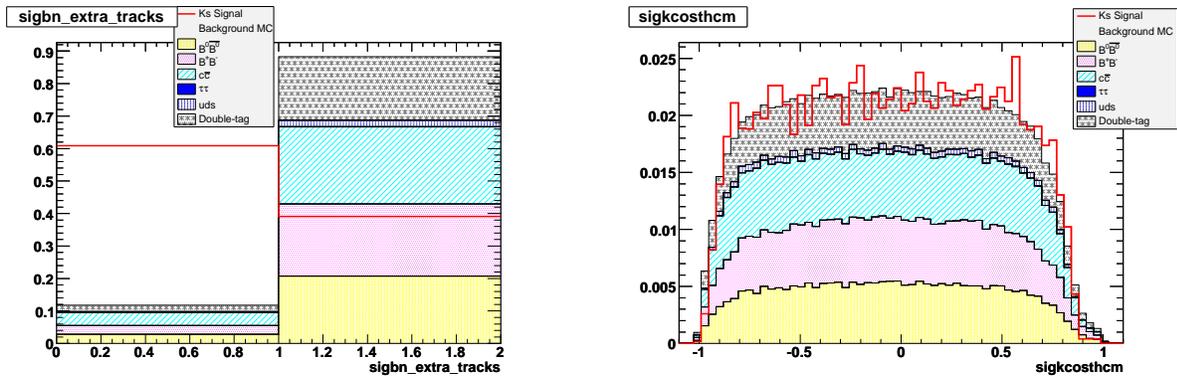
Figure 67: Signal/background histograms for sigbn_extra_tracks on left and sigkcosthcm on right.
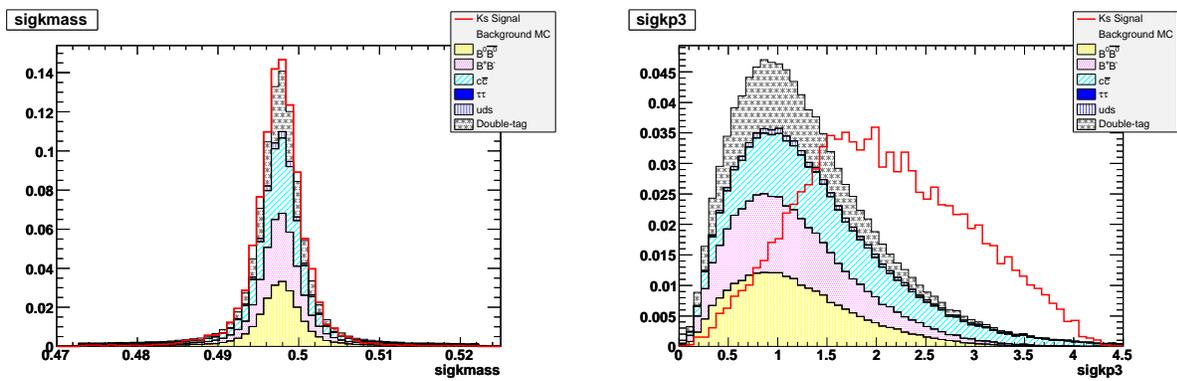
Figure 68: Signal/background histograms for sigkmass on left and sigkp3 on right.
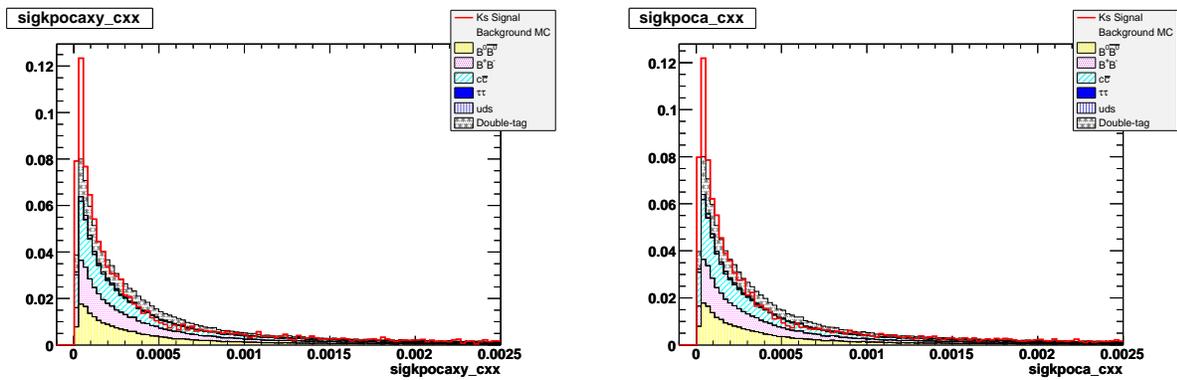
Figure 69: Signal/background histograms for sigkpocaxy_cxx on left and sigkpoca_cxx on right.
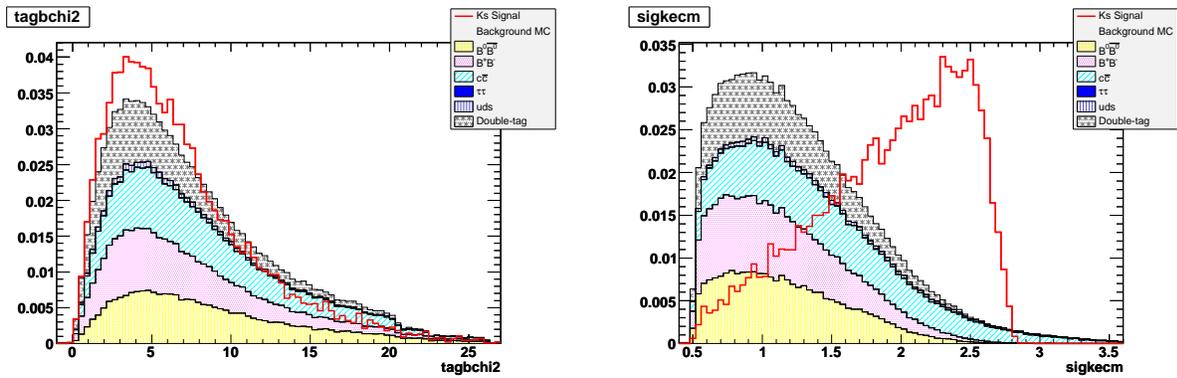
Figure 70: Signal/background histograms for tagbchi2 on left and sigkecm on right.
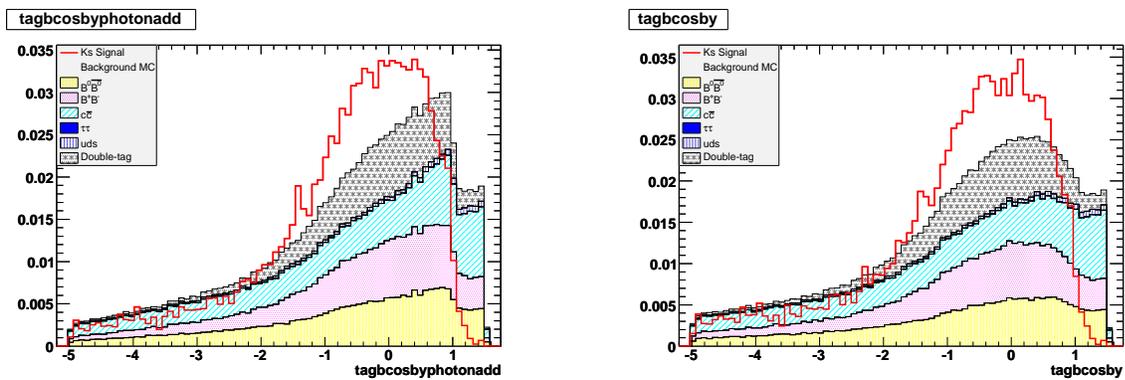
Figure 71: Signal/background histograms for tagbcosbyphotonadd on left and tagbcosby on right.
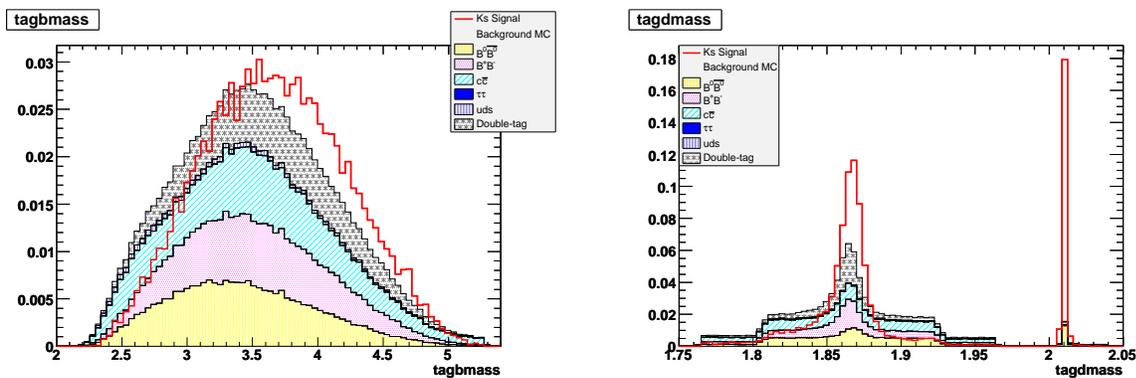
Figure 72: Signal/background histograms for tagbmass on left and tagdmass on right.
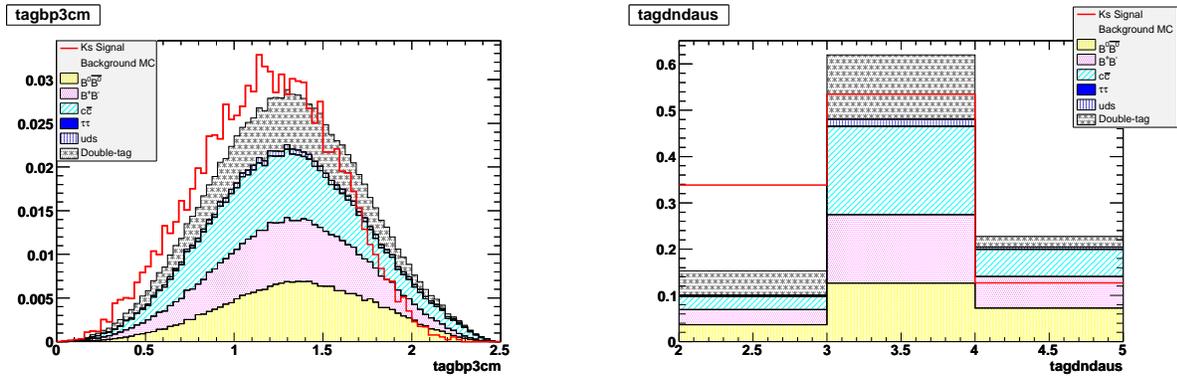
Figure 73: Signal/background histograms for tagbp3cm on left and tagdndaus on right.
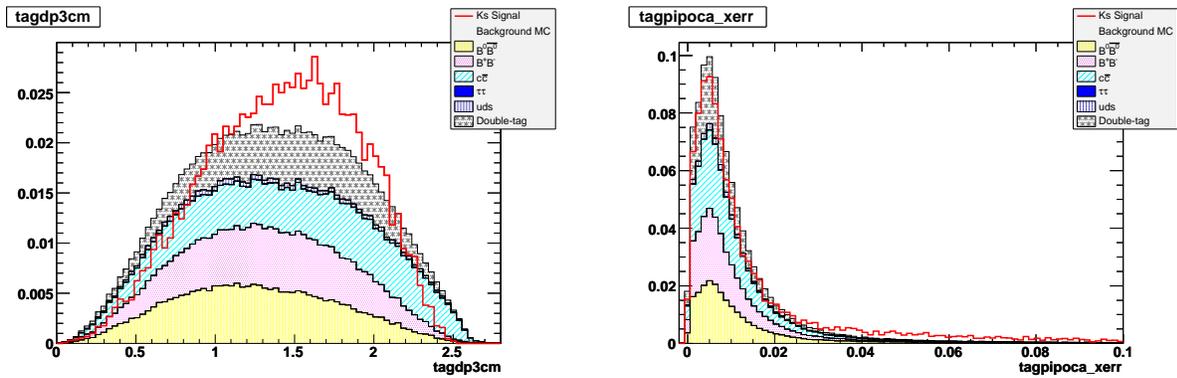
Figure 74: Signal/background histograms for tagdp3cm on left and tagpipoca_xerr on right.
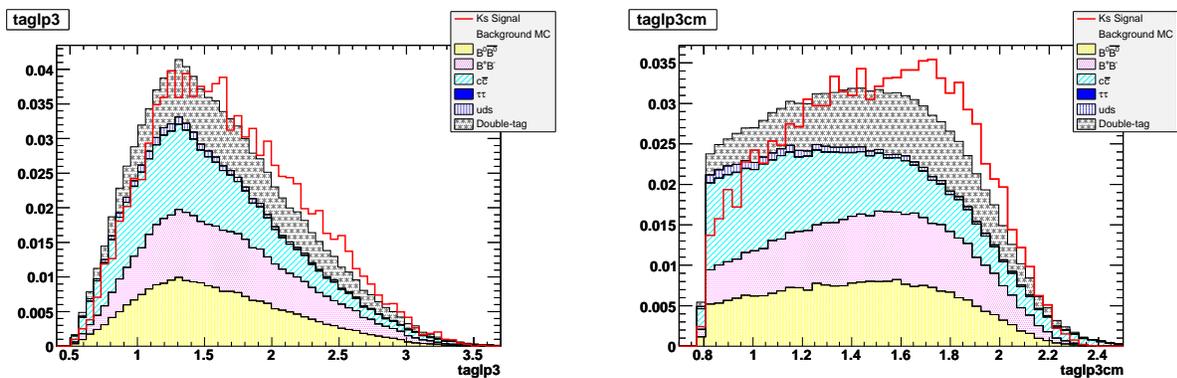
Figure 75: Signal/background histograms for taglp3 on left and taglp3cm on right.

Figure 76: Signal/background histograms for ycosth on left and yp3 on right.

Figure 77: Correlation matrix showing the amount of correlation in signal MC between the classification variables for the K_S^0 tree.

A.1.2 Background Divided by D-decay Mode

K^+ **Mode** .
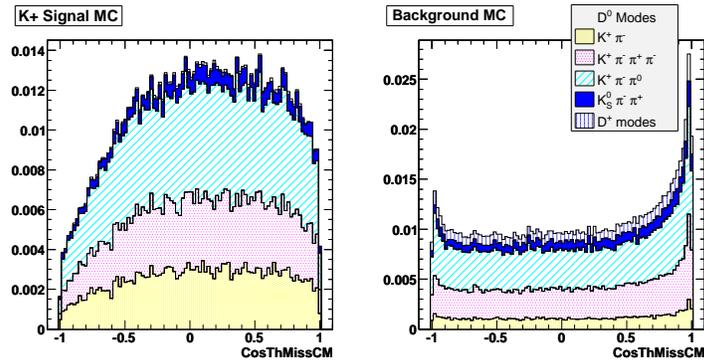
Figure 78: Signal/background histograms for CosThMissCM.
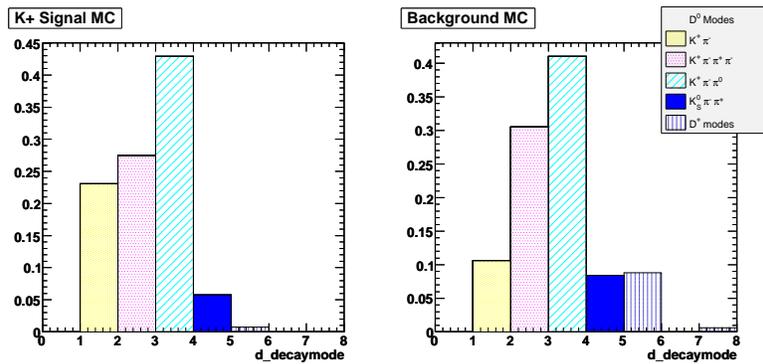
Figure 79: Signal/background histograms for d_decaymode.
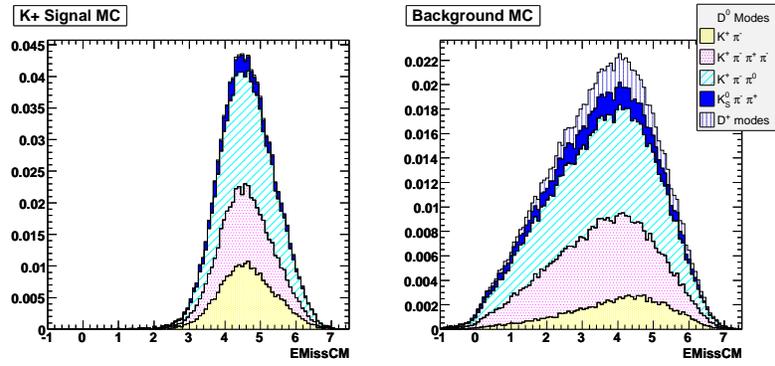
Figure 80: Signal/background histograms for EMissCM.
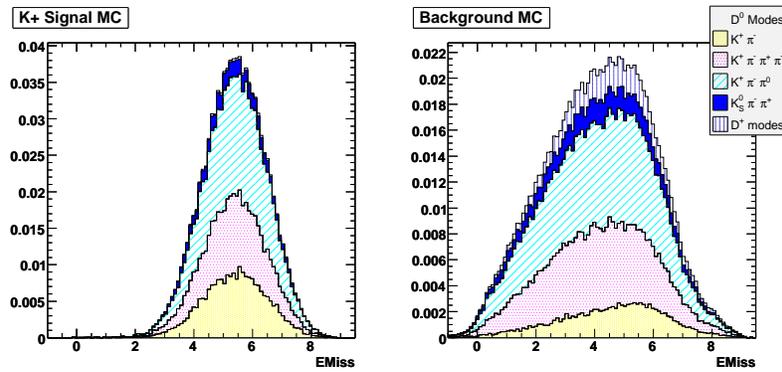
Figure 81: Signal/background histograms for EMiss.
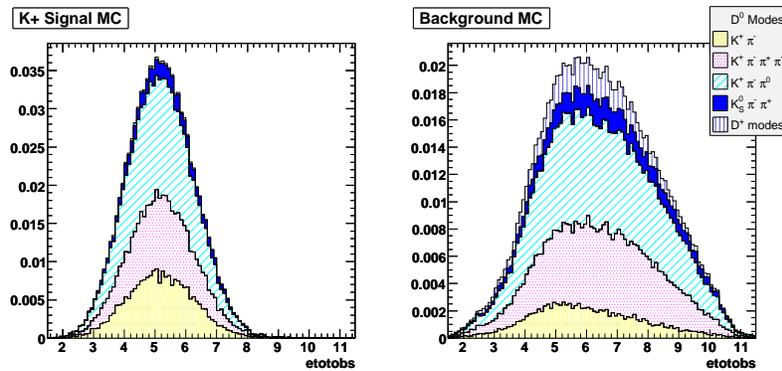
Figure 82: Signal/background histograms for etotobs.
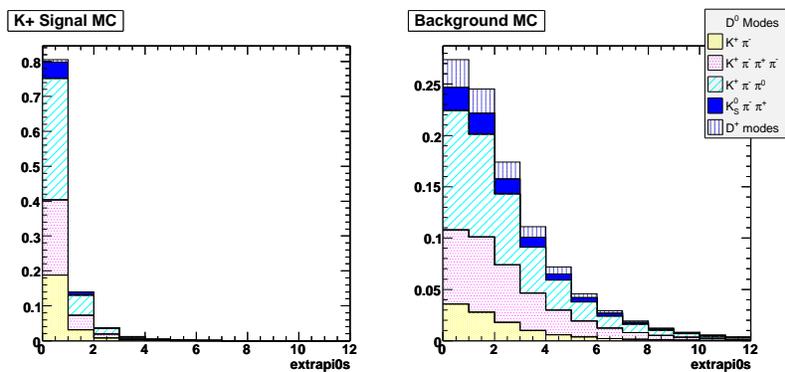
Figure 83: Signal/background histograms for extrapi0s.
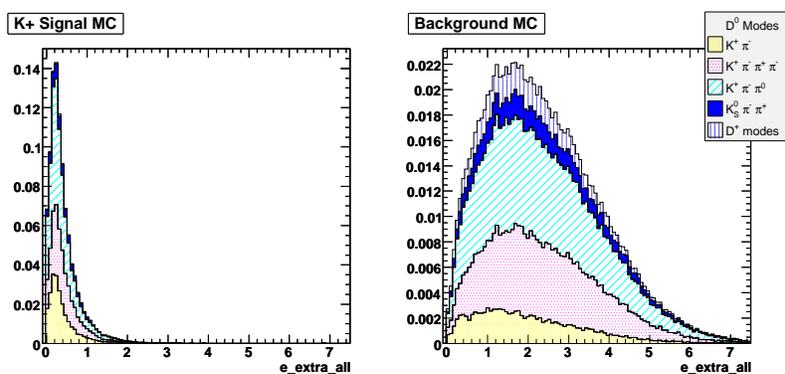
Figure 84: Signal/background histograms for e_extra_all.
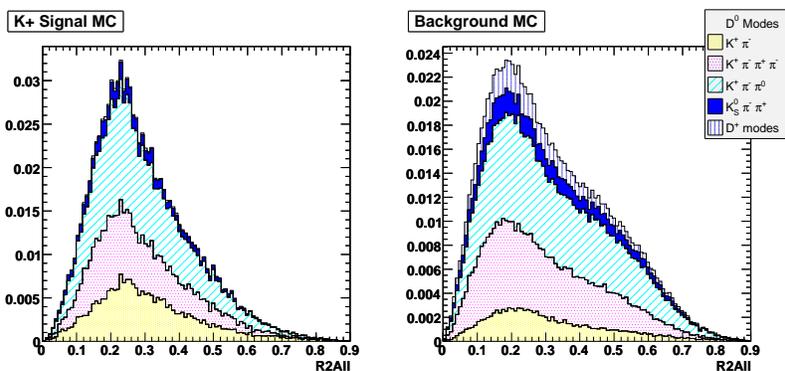
Figure 85: Signal/background histograms for R2All.
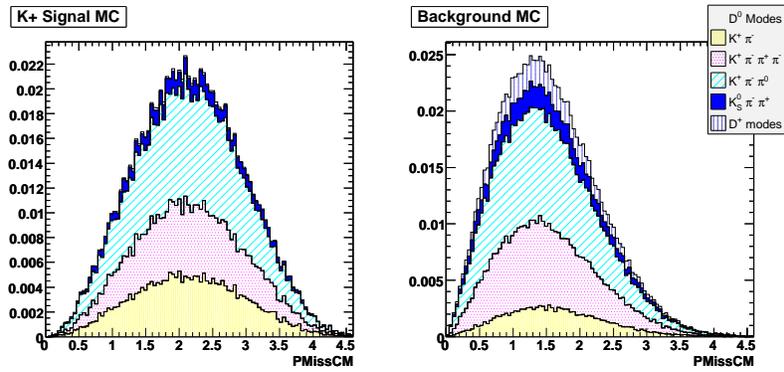
Figure 86: Signal/background histograms for PMissCM.
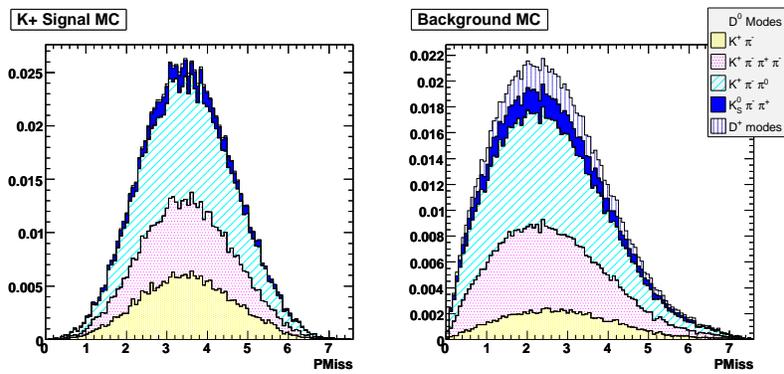
Figure 87: Signal/background histograms for PMiss.
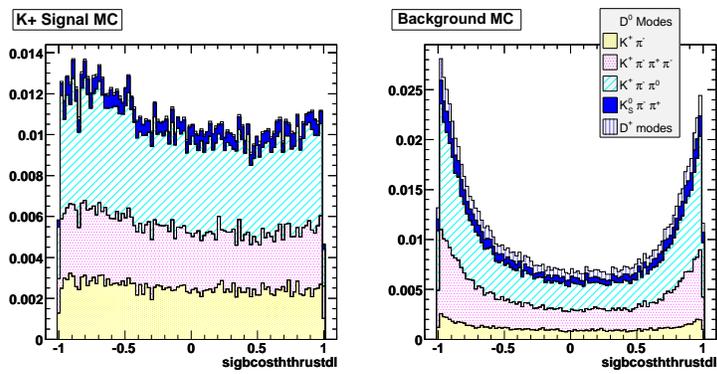
Figure 88: Signal/background histograms for sigbcoststhrustdl.
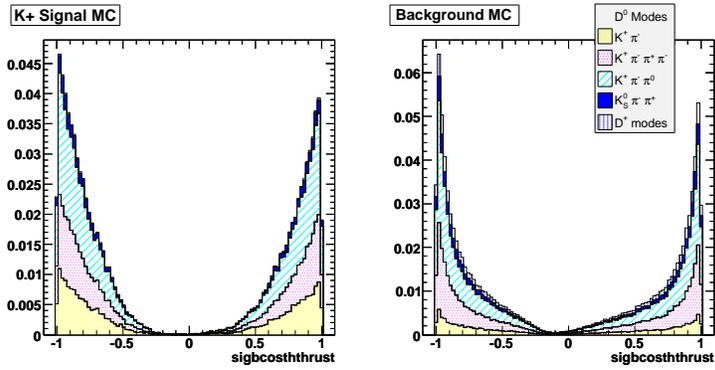
Figure 89: Signal/background histograms for sigbcosththrust.
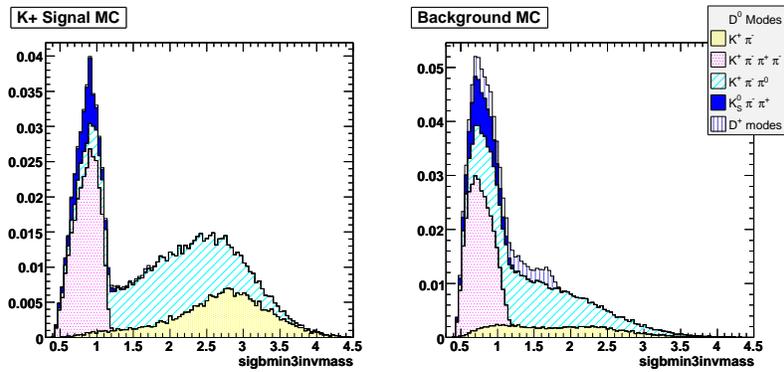
Figure 90: Signal/background histograms for sigbmin3invmass.
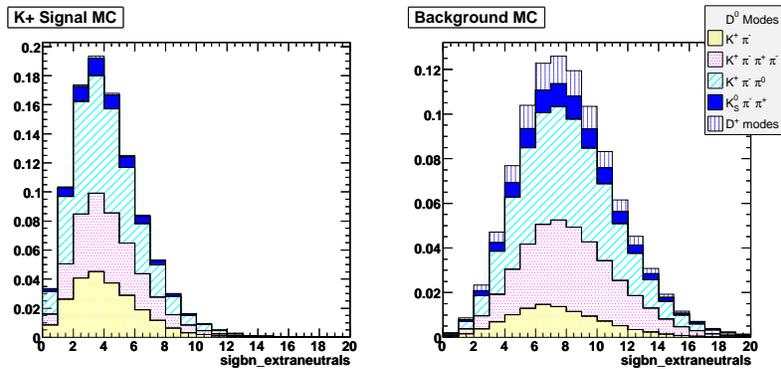
Figure 91: Signal/background histograms for sigbn_extraneutrals.
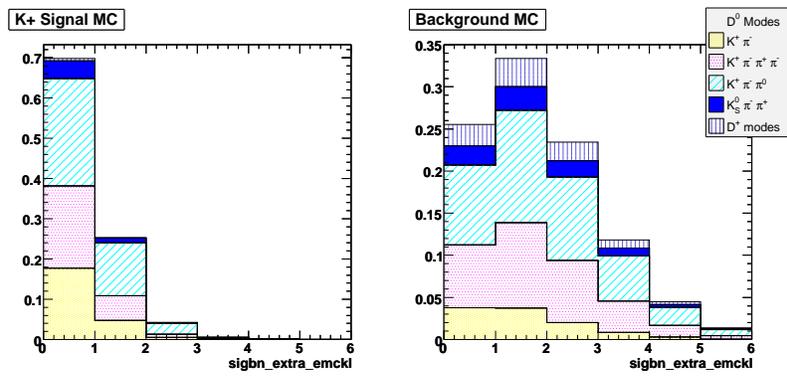
Figure 92: Signal/background histograms for sigbn_extra_emckl.
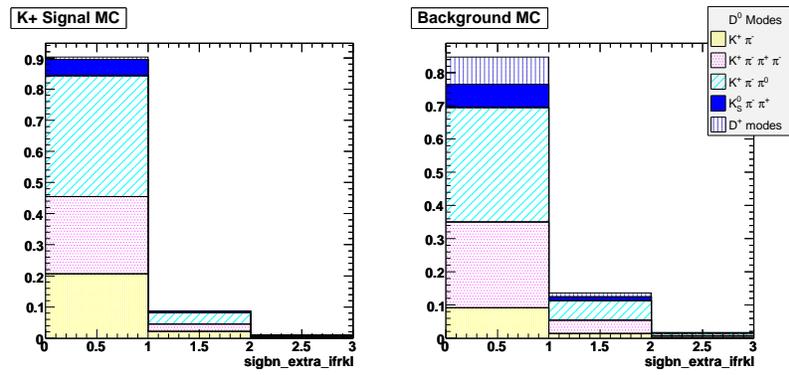
Figure 93: Signal/background histograms for sigbn_extra_ifrkl.
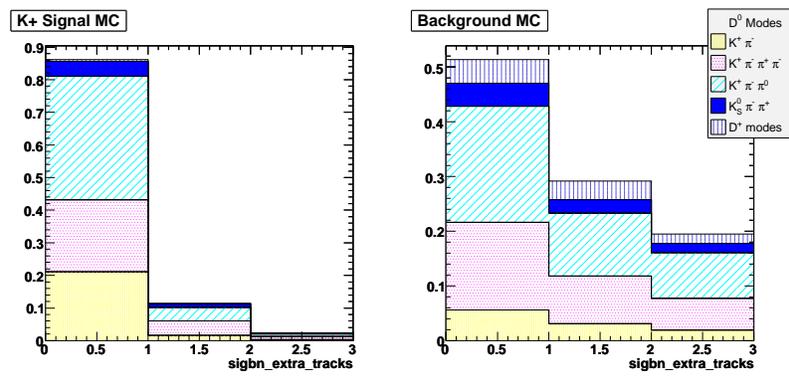
Figure 94: Signal/background histograms for sigbn_extra_tracks.
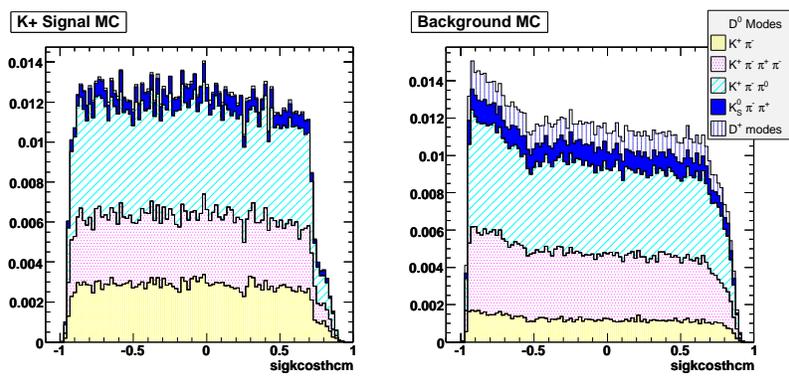
Figure 95: Signal/background histograms for sigkcosthcm.
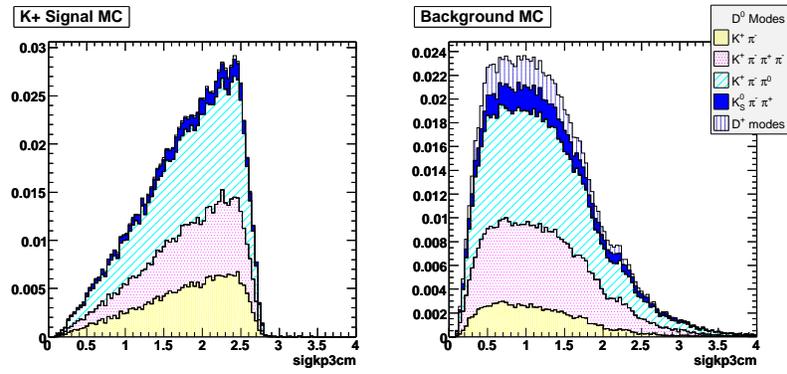
Figure 96: Signal/background histograms for sigkp3cm.
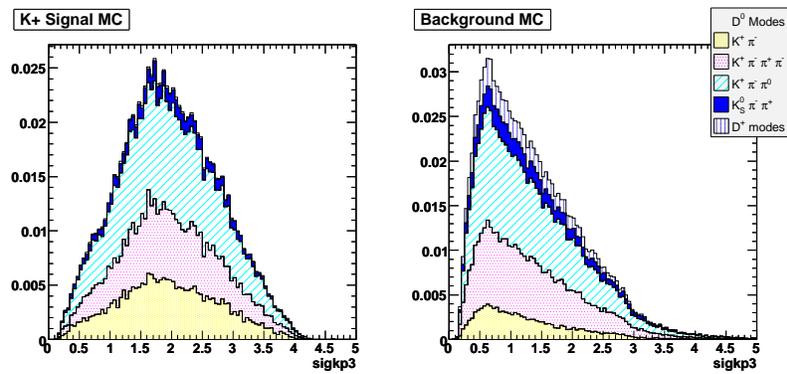
Figure 97: Signal/background histograms for sigkp3.
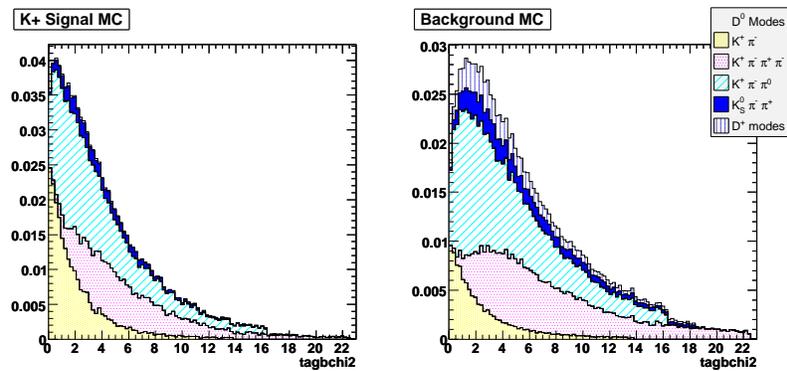
Figure 98: Signal/background histograms for tagbchi2.
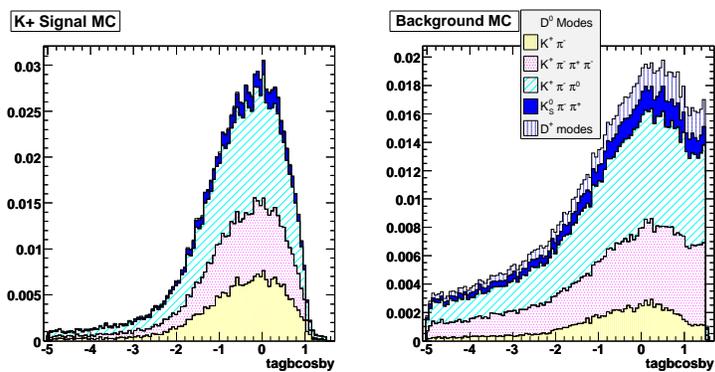
Figure 99: Signal/background histograms for tagbcosby.
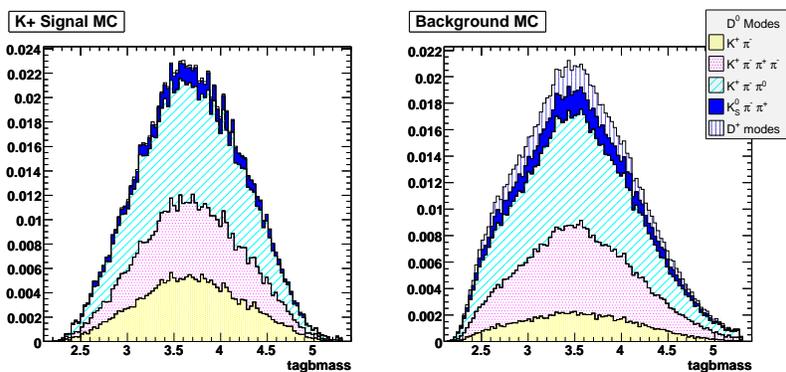
Figure 100: Signal/background histograms for tagbmass.
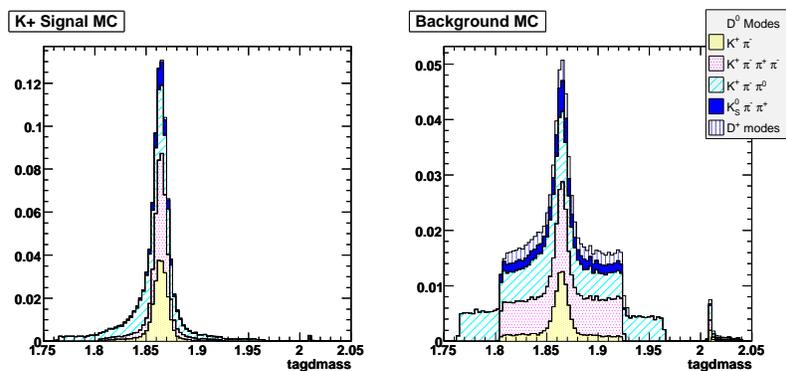
Figure 101: Signal/background histograms for tagdmass.
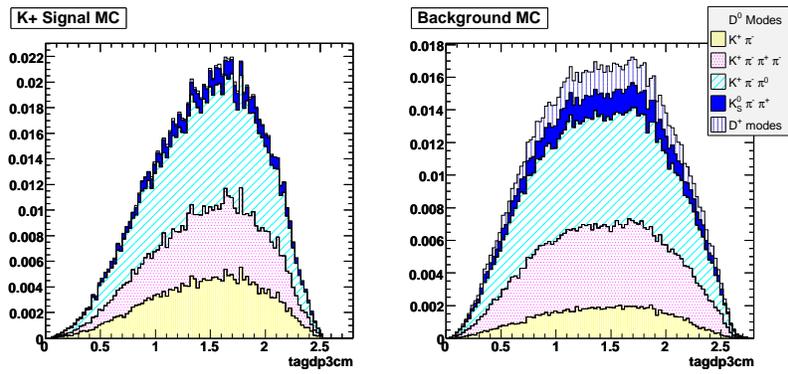
Figure 102: Signal/background histograms for `tagdp3cm`.
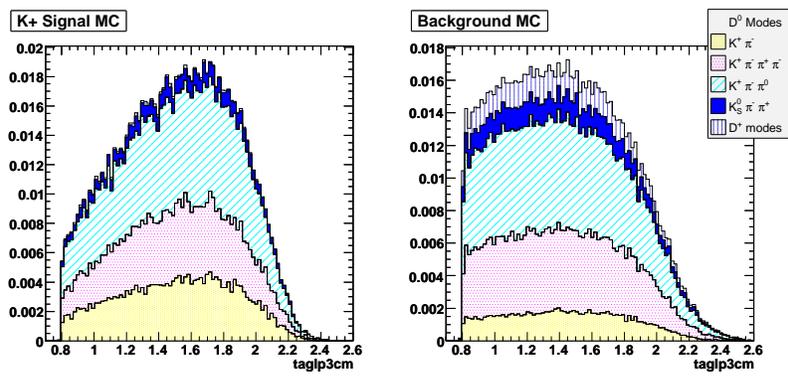
Figure 103: Signal/background histograms for `taglp3cm`.

K_S **Mode** .
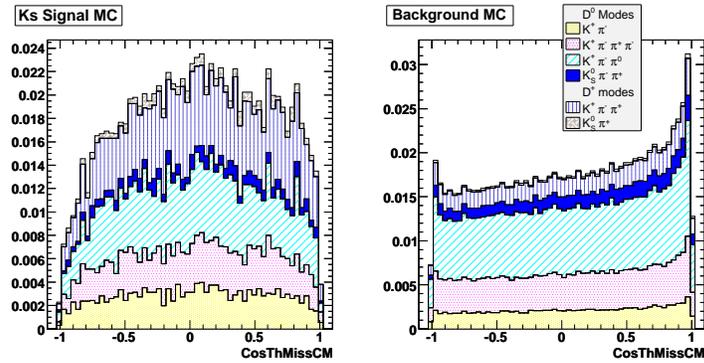
Figure 104: Signal/background histograms for CosThMissCM.
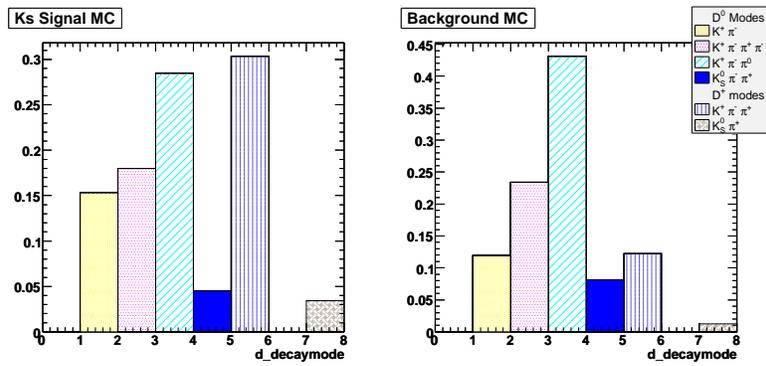
Figure 105: Signal/background histograms for d_decaymode.
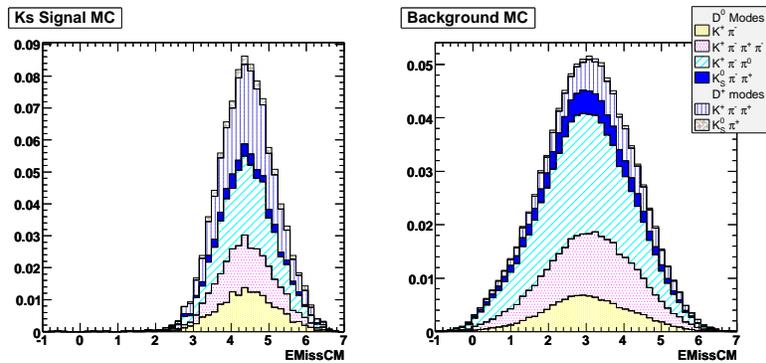
Figure 106: Signal/background histograms for EMissCM.
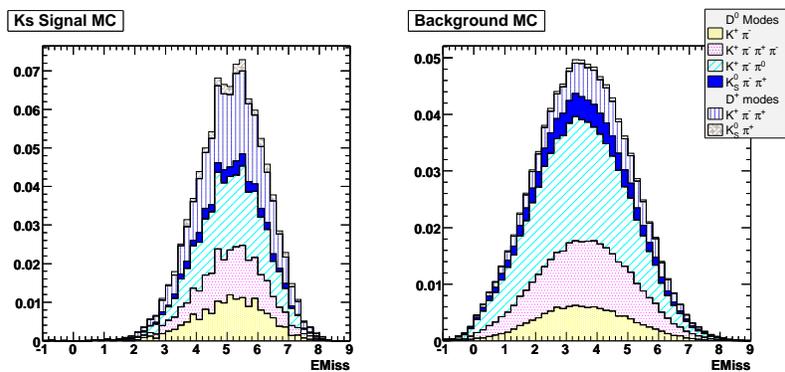
Figure 107: Signal/background histograms for EMiss.
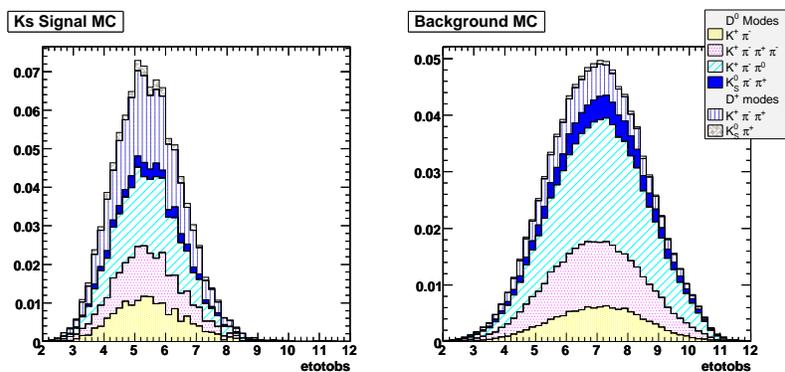
Figure 108: Signal/background histograms for etotobs.
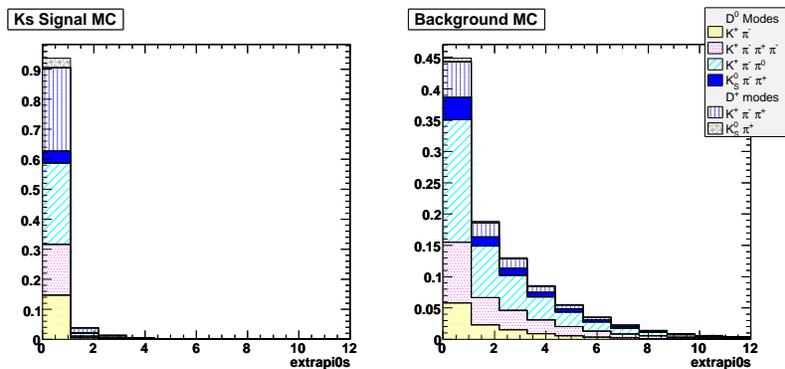
Figure 109: Signal/background histograms for extrapi0s.
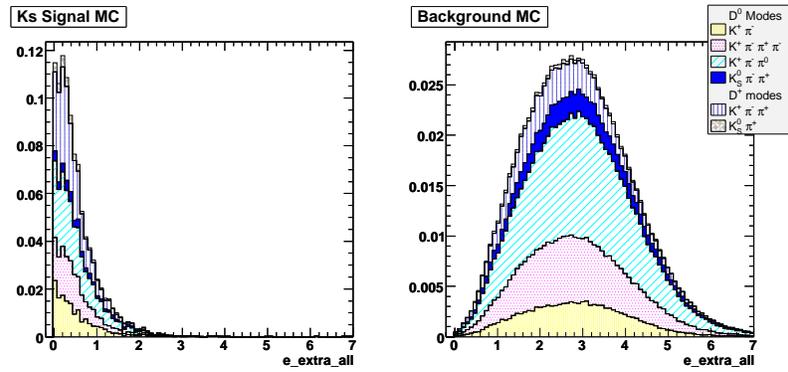
Figure 110: Signal/background histograms for e_extra_all.
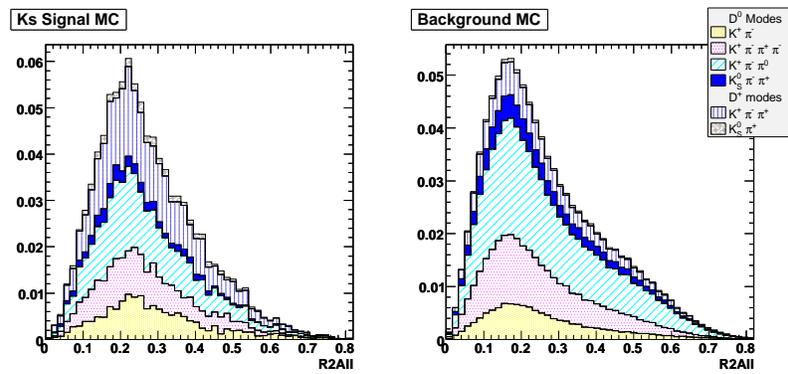
Figure 111: Signal/background histograms for R2All.
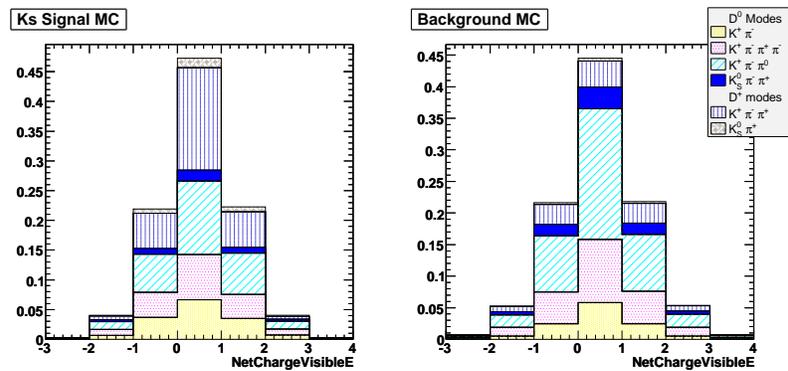
Figure 112: Signal/background histograms for NetChargeVisibleE.
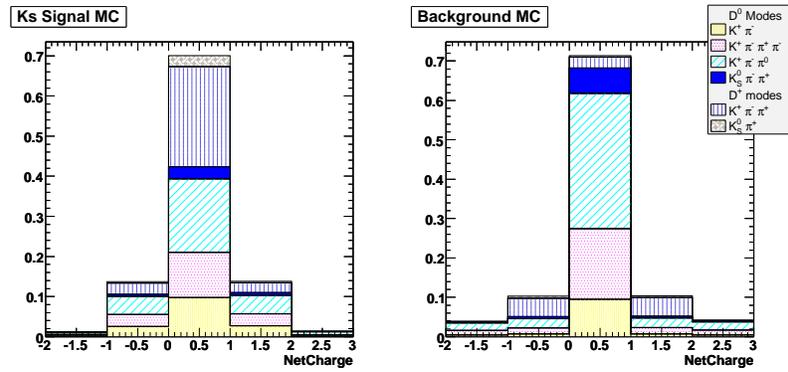
Figure 113: Signal/background histograms for NetCharge.
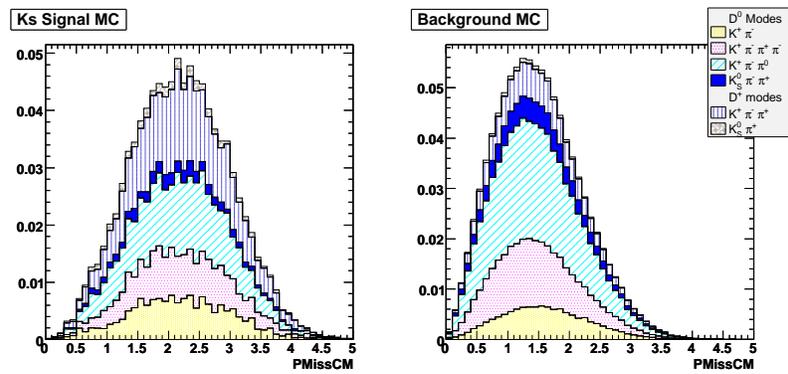
Figure 114: Signal/background histograms for PMissCM.
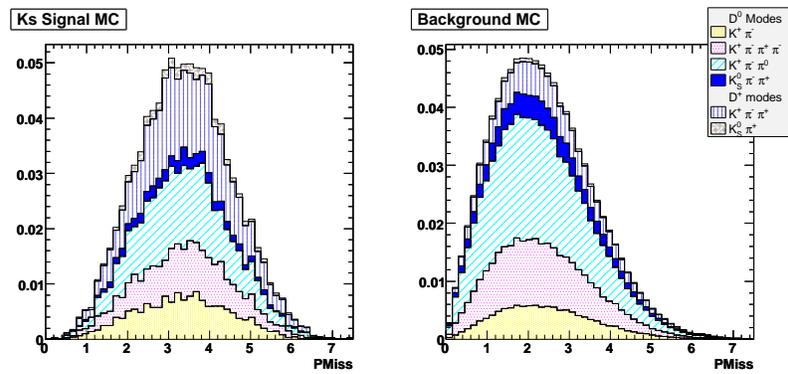
Figure 115: Signal/background histograms for PMiss.
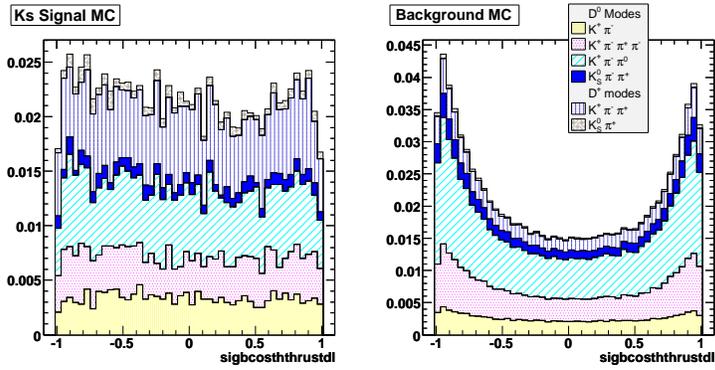
Figure 116: Signal/background histograms for sigbcoststhrustdl.
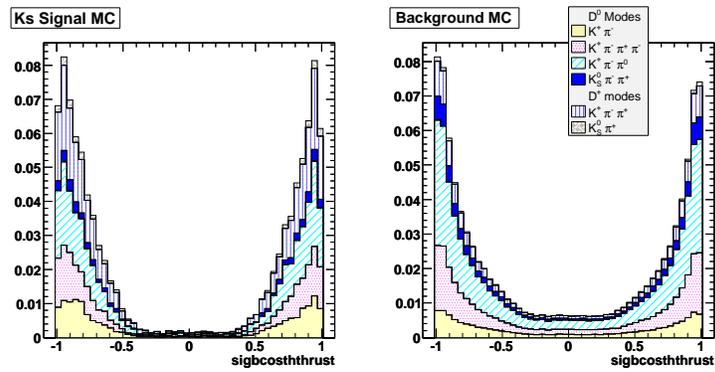
Figure 117: Signal/background histograms for sigbcoststhrust.
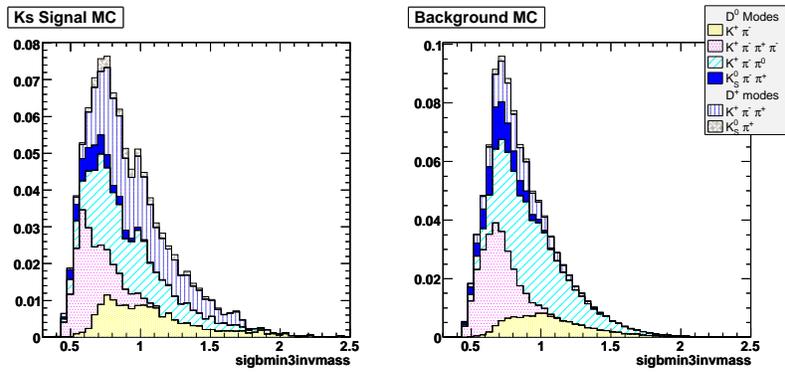
Figure 118: Signal/background histograms for sigbmin3invmass.
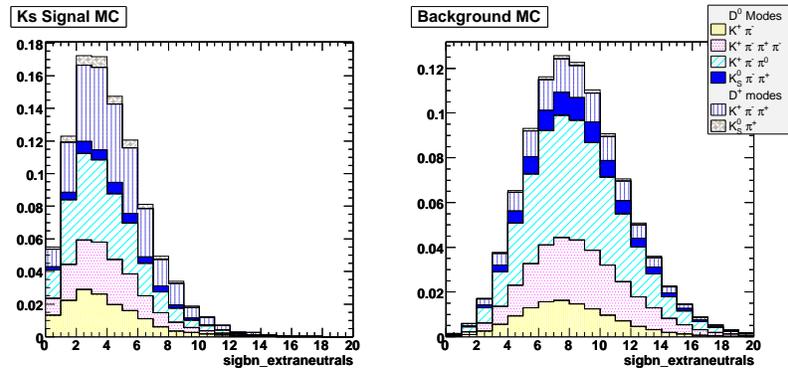
Figure 119: Signal/background histograms for sigbn_extraneutrals.
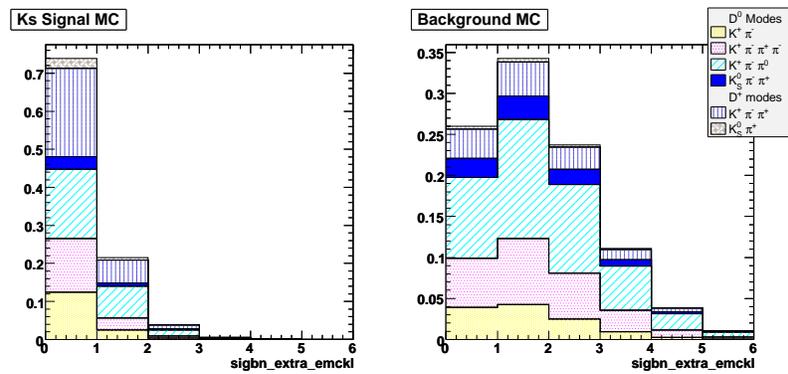
Figure 120: Signal/background histograms for sigbn_extra_emckl.
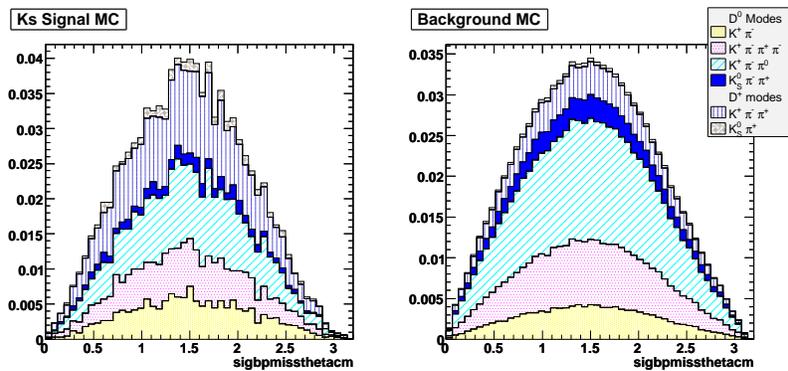
Figure 121: Signal/background histograms for sigbpmissthetacm.
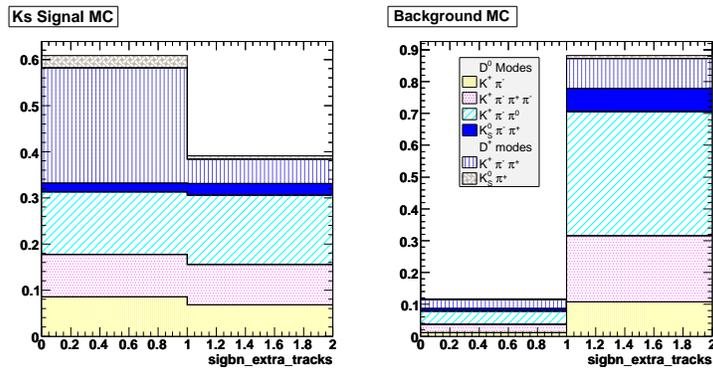
Figure 122: Signal/background histograms for sigbn_extra_tracks.
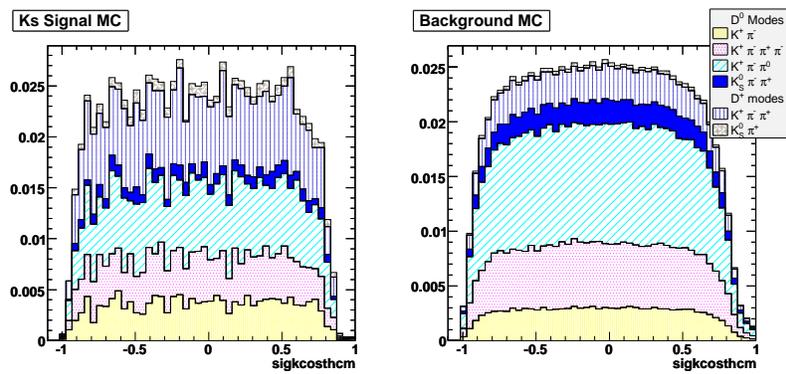
Figure 123: Signal/background histograms for sigkcosthcm.
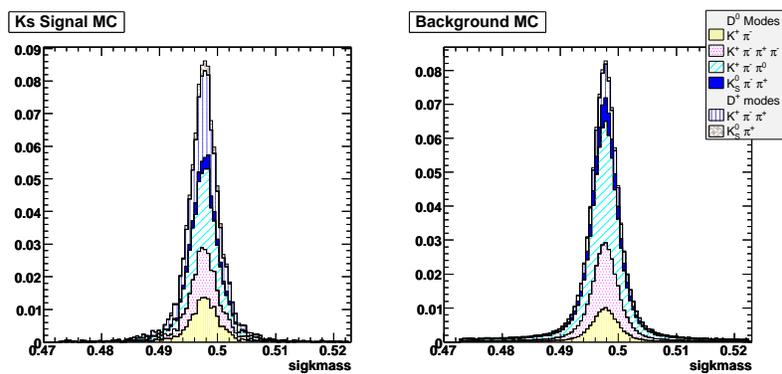
Figure 124: Signal/background histograms for sigkmass.
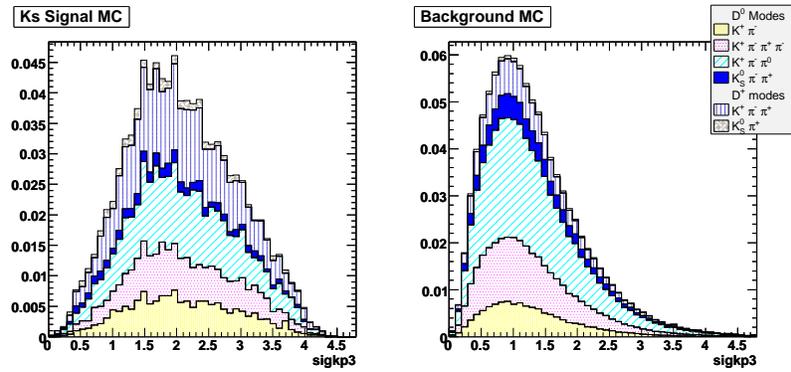
Figure 125: Signal/background histograms for sigkp3.
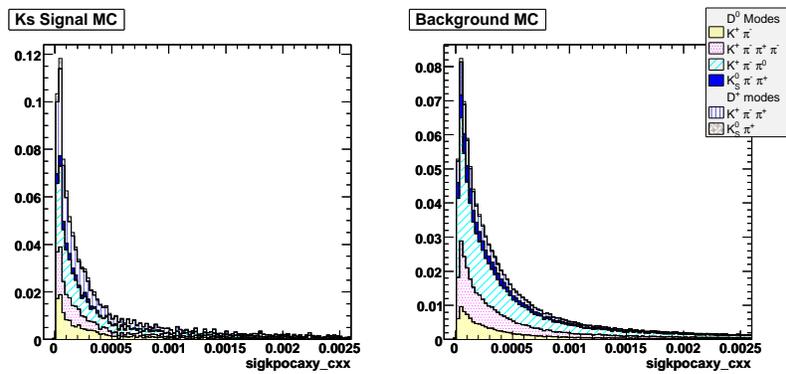
Figure 126: Signal/background histograms for sigkpocaxy_cxx.
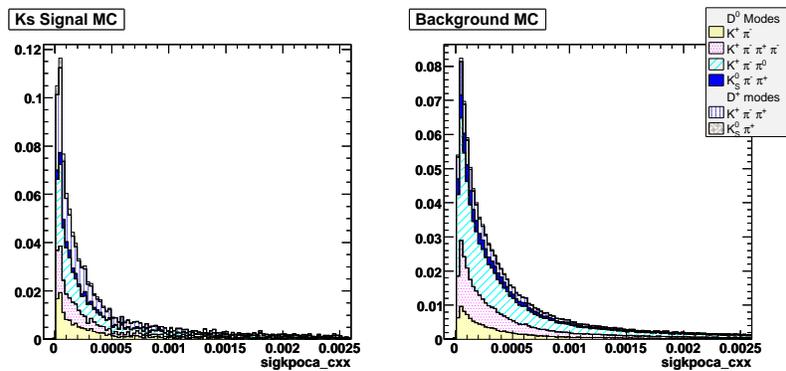
Figure 127: Signal/background histograms for sigkpoca_cxx.
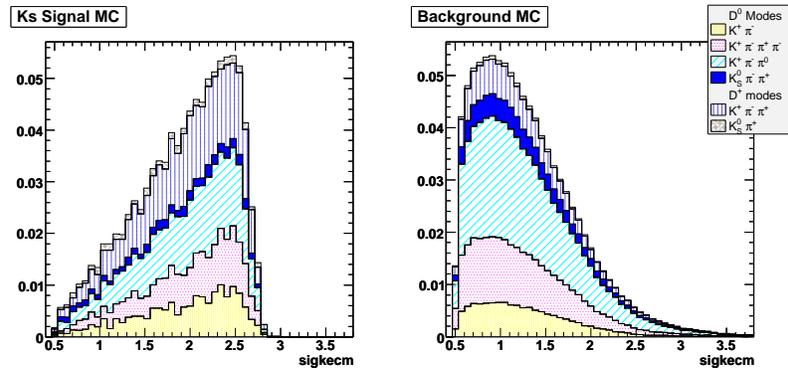
Figure 128: Signal/background histograms for sigkecm.
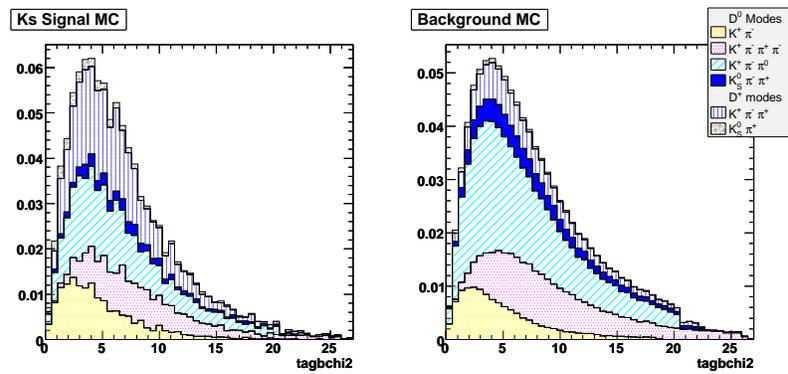
Figure 129: Signal/background histograms for tagbchi2.
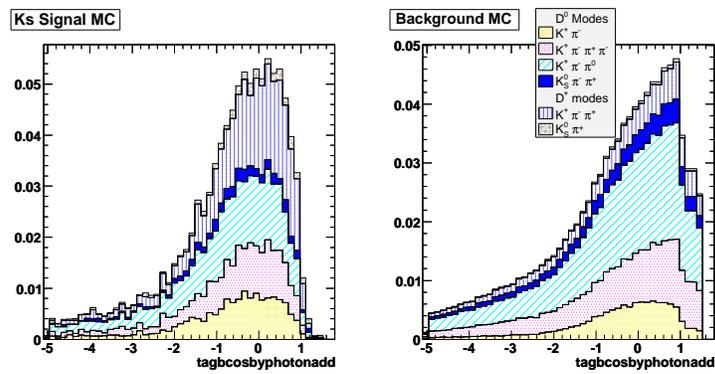
Figure 130: Signal/background histograms for tagbcosbyphotonadd.
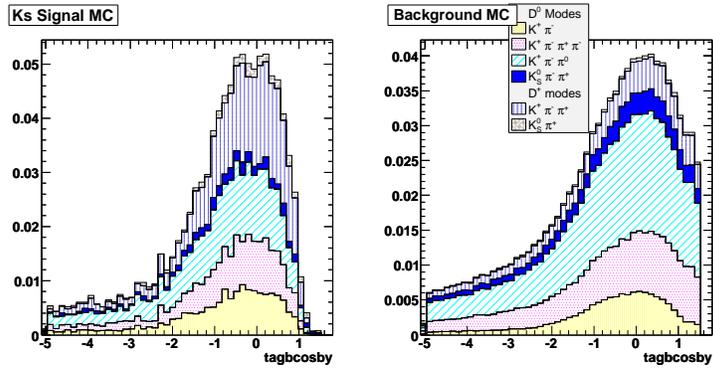
Figure 131: Signal/background histograms for tagbcosby.
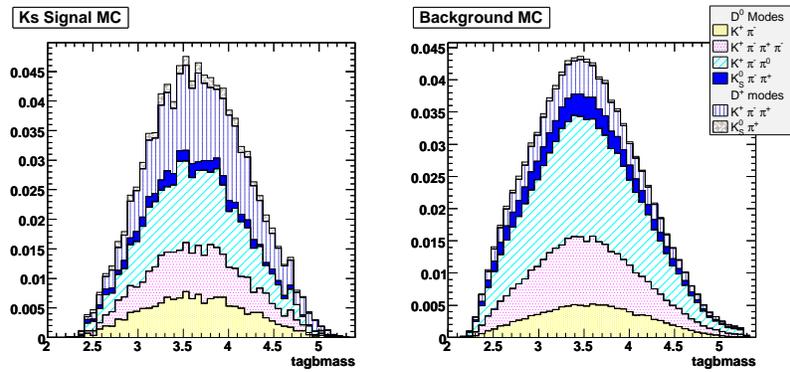
Figure 132: Signal/background histograms for tagbmass.
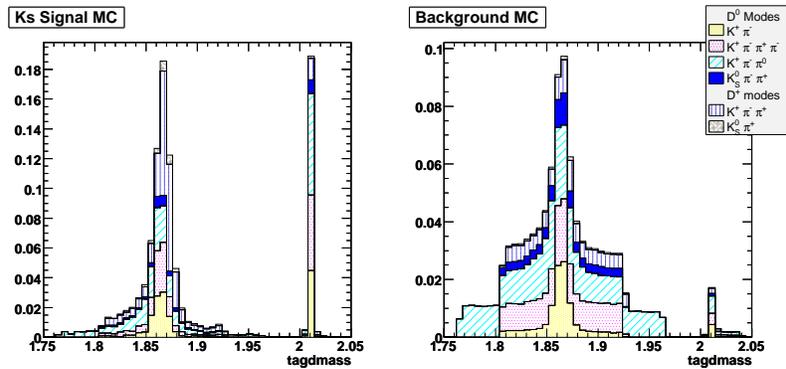
Figure 133: Signal/background histograms for tagdmass.
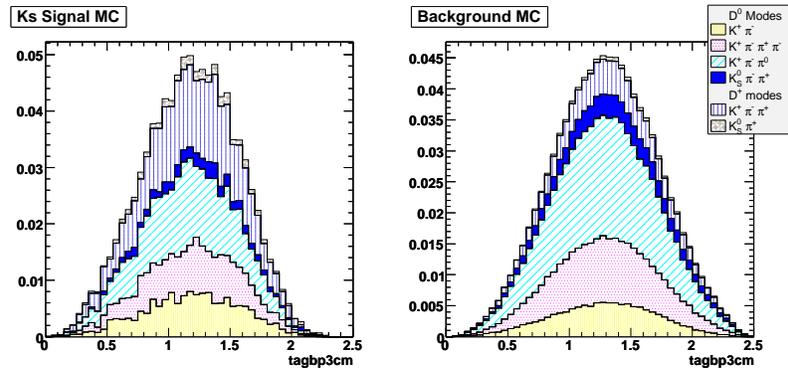
Figure 134: Signal/background histograms for tagbp3cm.
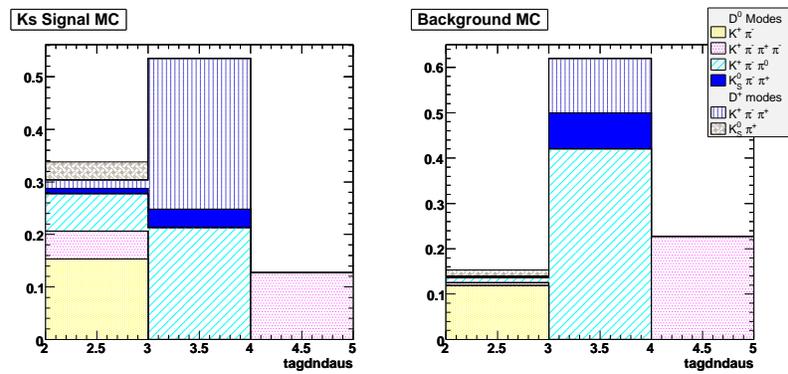
Figure 135: Signal/background histograms for tagdndaus.
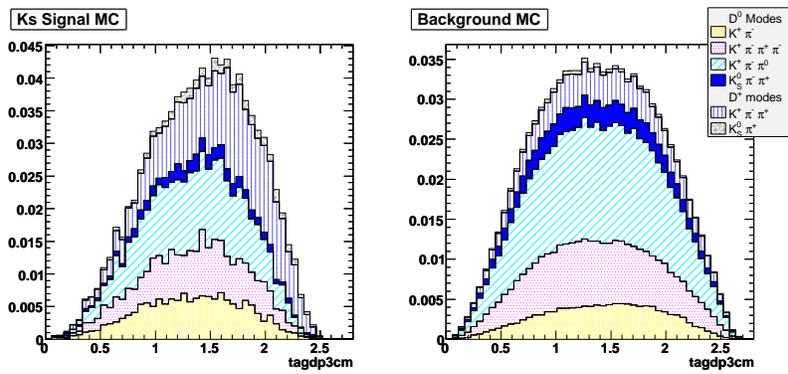
Figure 136: Signal/background histograms for tagdp3cm.
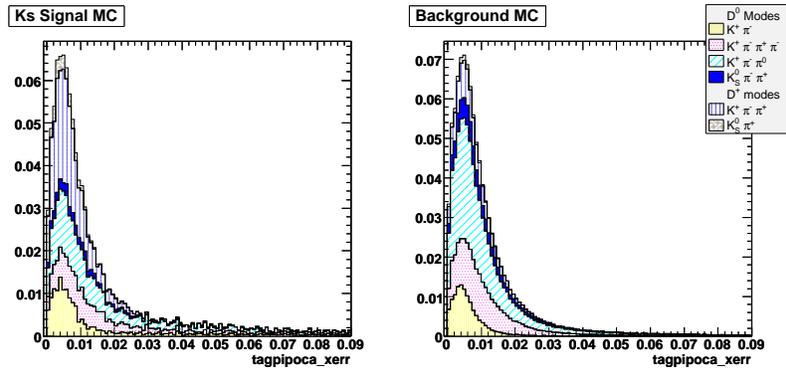
Figure 137: Signal/background histograms for tagpipoca_xerr.
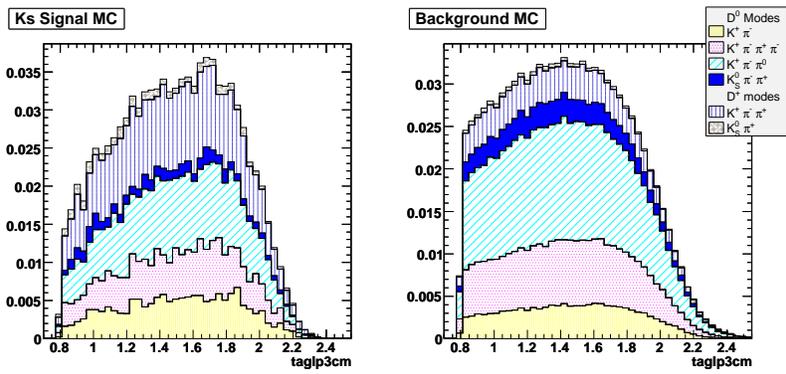
Figure 138: Signal/background histograms for taglp3cm.
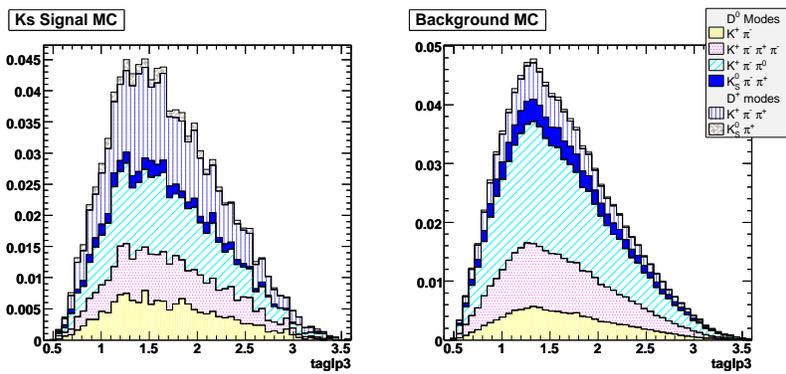
Figure 139: Signal/background histograms for taglp3.
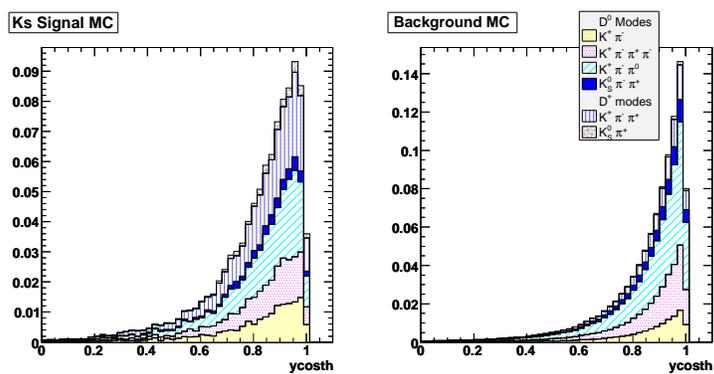
Figure 140: Signal/background histograms for ycosth.
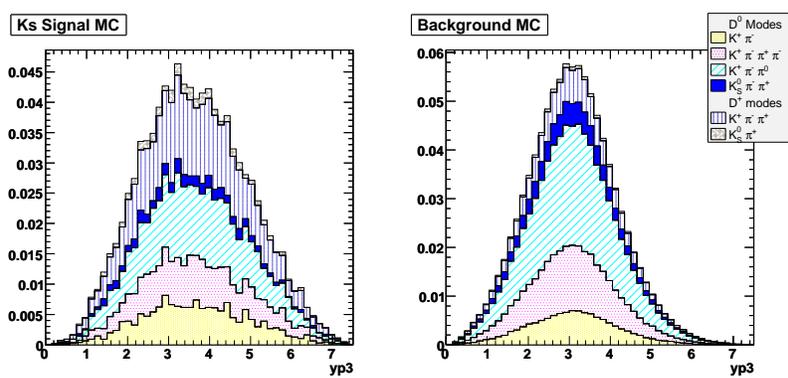
Figure 141: Signal/background histograms for yp3.

A.2 Signal MC Plots for Classification Variables after the Final BDT Cut

The following are unit-normalized histograms of signal MC after the final BDT cut for each classification variable.

A.2.1 K^+ mode
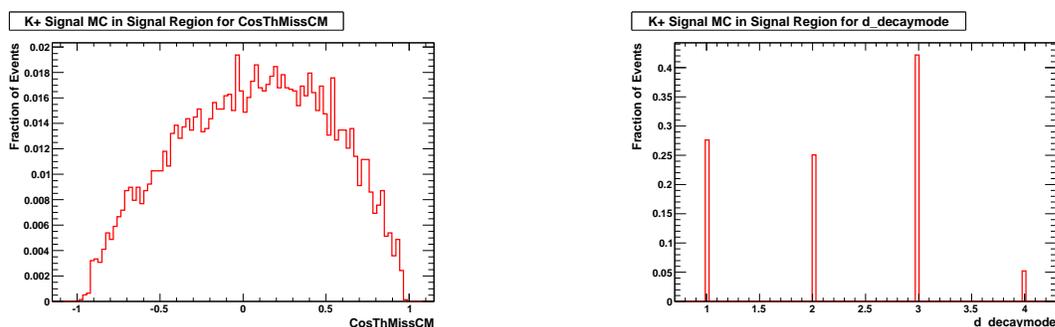
Figure 142: Signal MC histograms for CosThMissCM on left and d_decaymode on right.
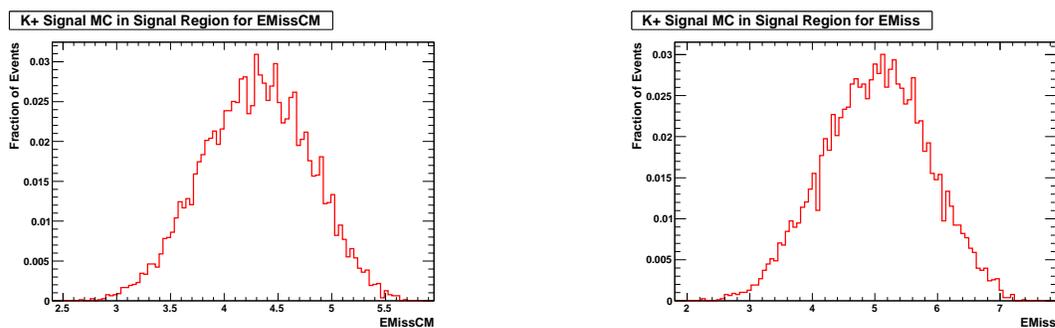
Figure 143: Signal MC histograms for EMissCM on left and EMiss on right.
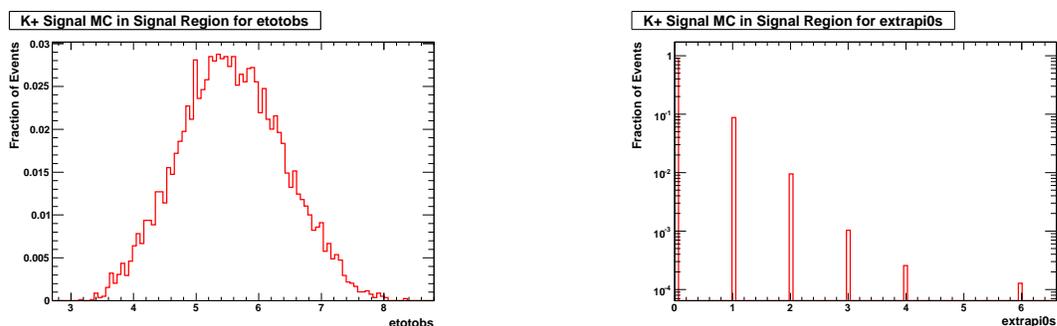
Figure 144: Signal MC histograms for etotobs on left and extrapi0s on right.
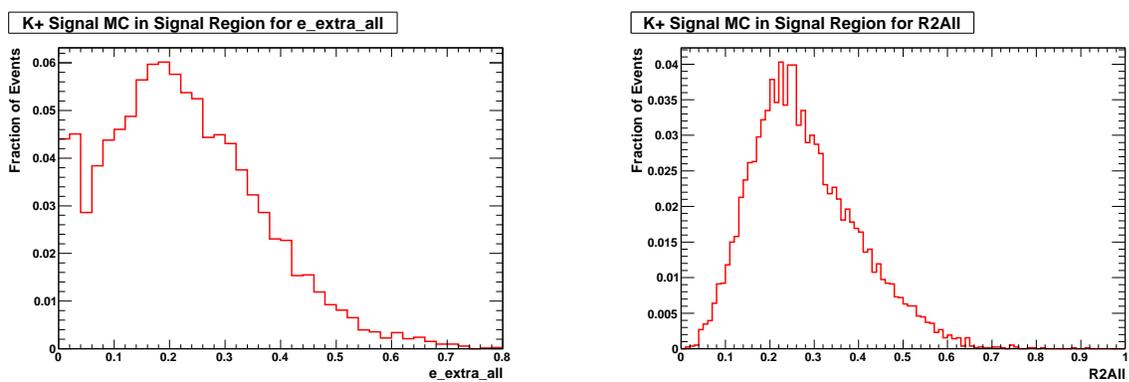
Figure 145: Signal MC histograms for e_extra_all on left and R2All on right.
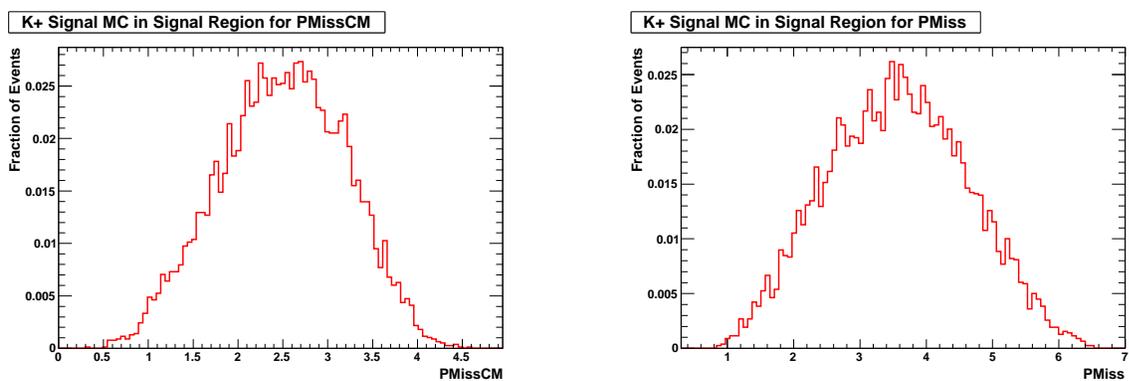
Figure 146: Signal MC histograms for PMissCM on left and PMiss on right.
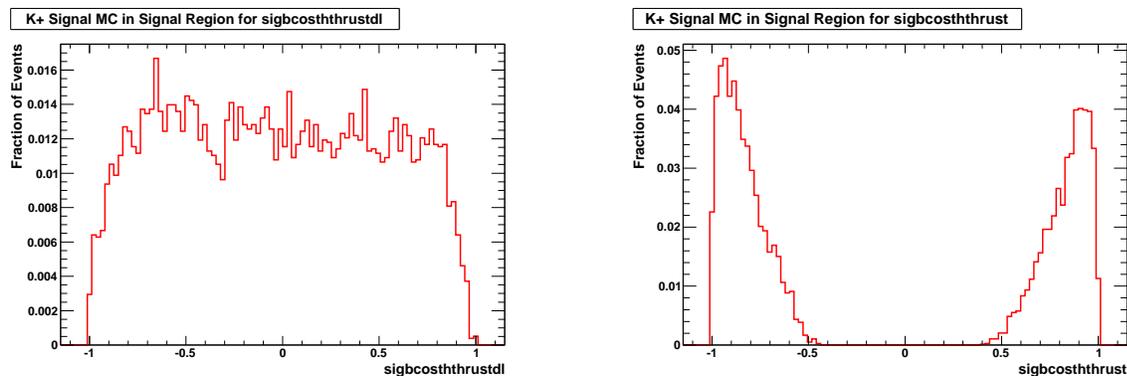
Figure 147: Signal MC histograms for **sigbcostsththrustdl** on left and **sigbcostsththrust** on right.
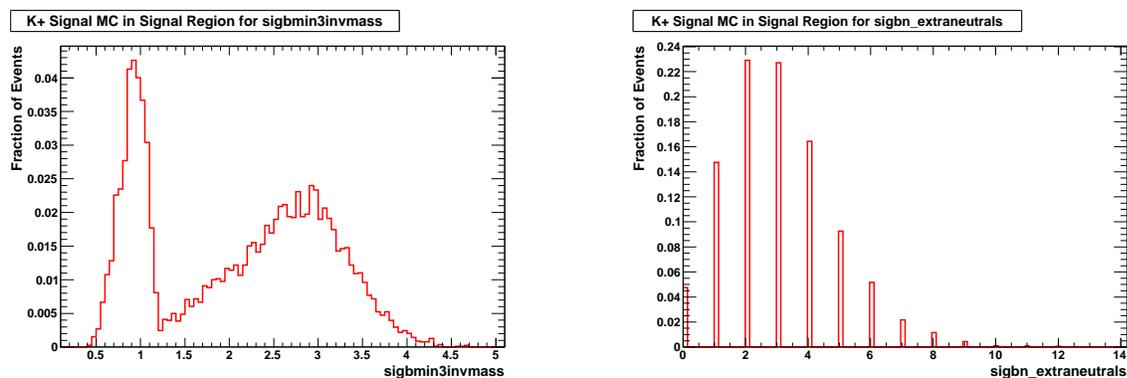
Figure 148: Signal MC histograms for **sigbmin3invmass** on left and **sigbn_extraneutrals** on right.
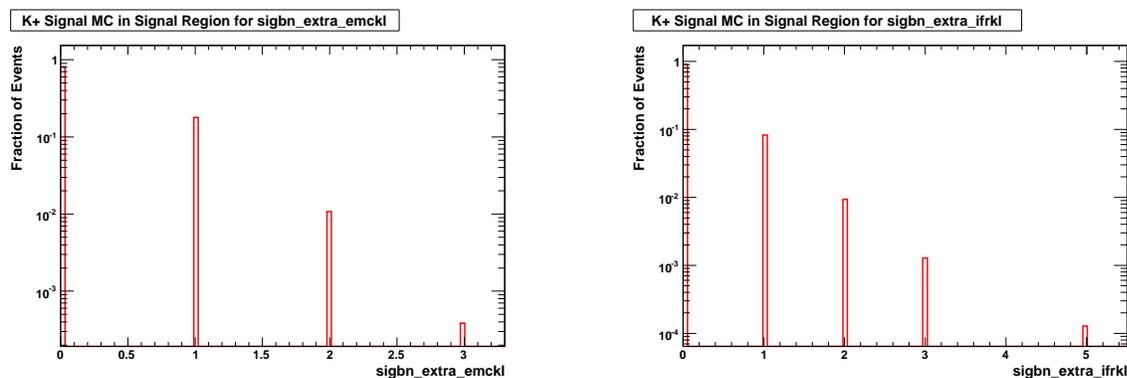
Figure 149: Signal MC histograms for **sigbn_extra_emckl** on left and **sigbn_extra_ifrkl** on right.
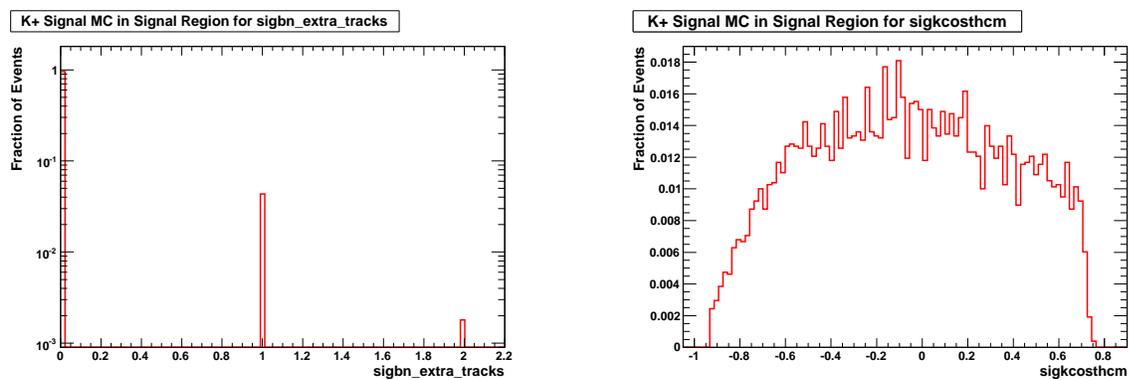
Figure 150: Signal MC histograms for **sigbn_extra_tracks** on left and **sigkcosthcm** on right.
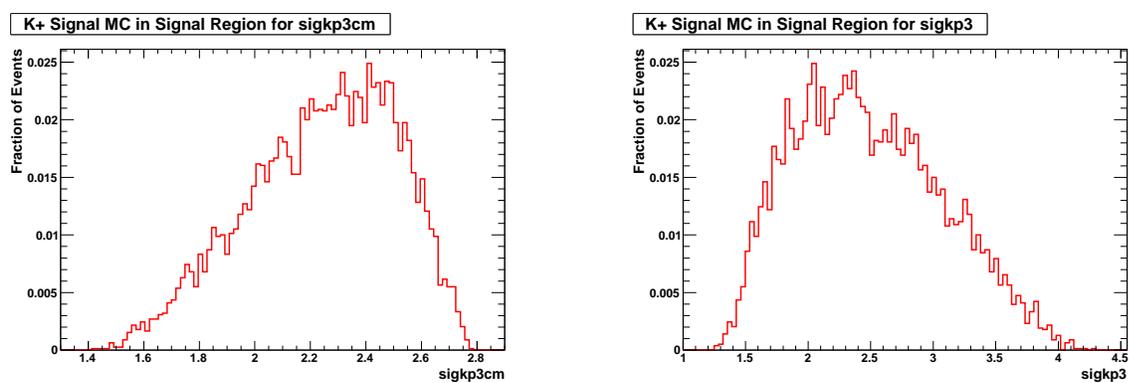
Figure 151: Signal MC histograms for **sigkp3cm** on left and **sigkp3** on right.
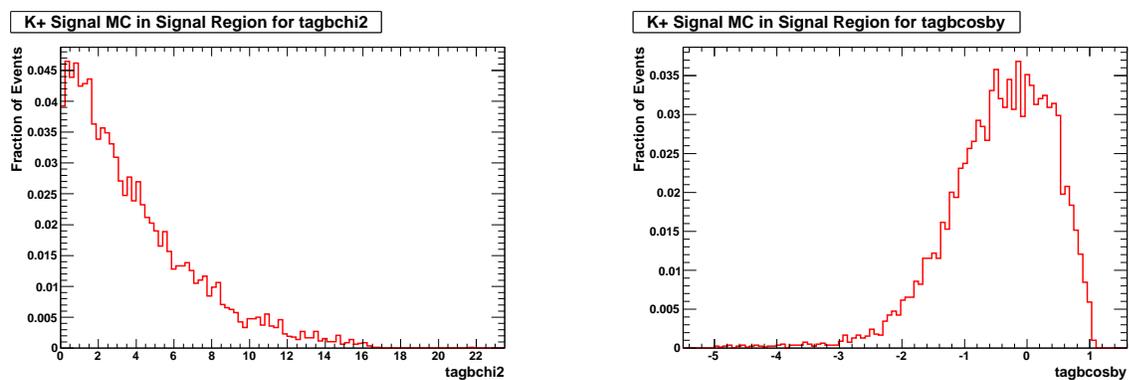
Figure 152: Signal MC histograms for **tagbchi2** on left and **tagbcosby** on right.
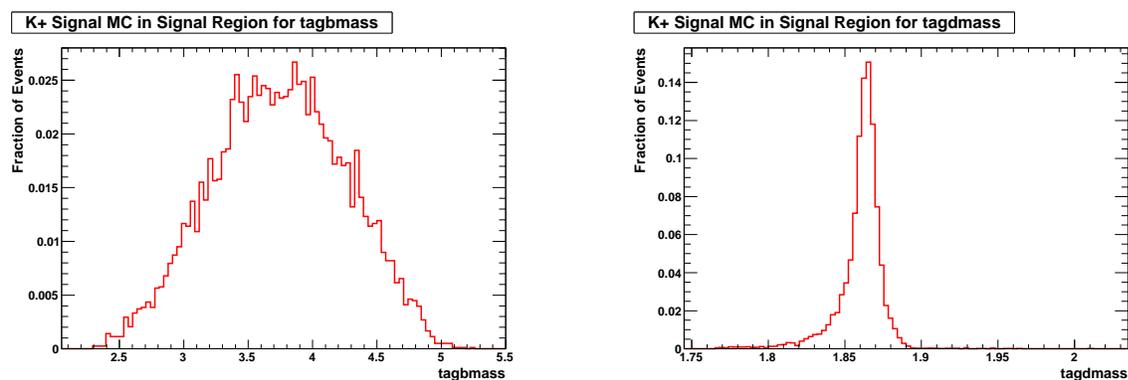
Figure 153: Signal MC histograms for tagbmass on left and tagdmass on right.
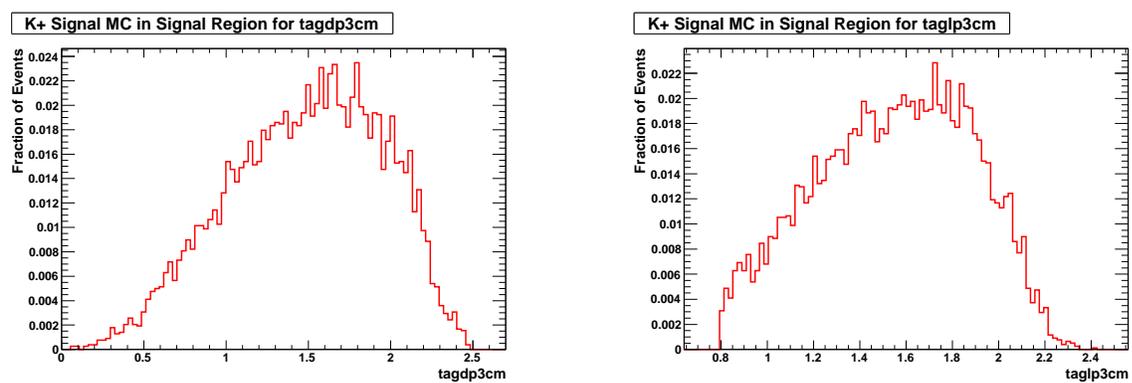
Figure 154: Signal MC histograms for tagdp3cm on left and taglp3cm on right.
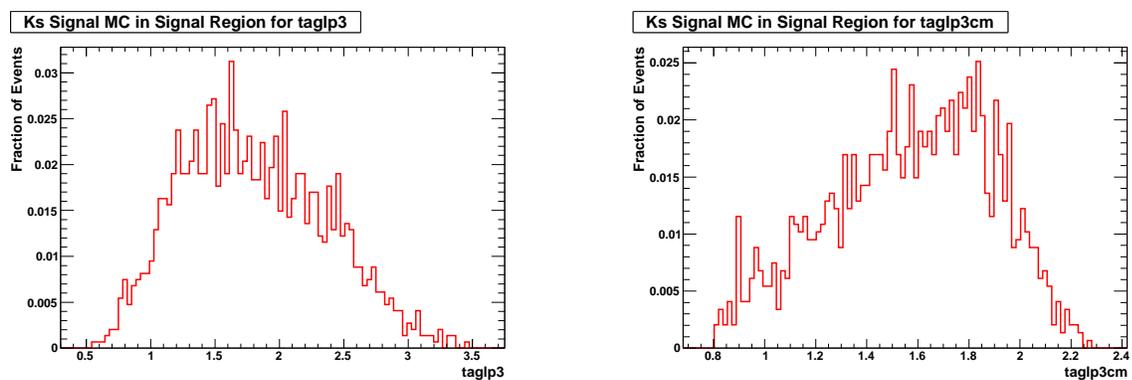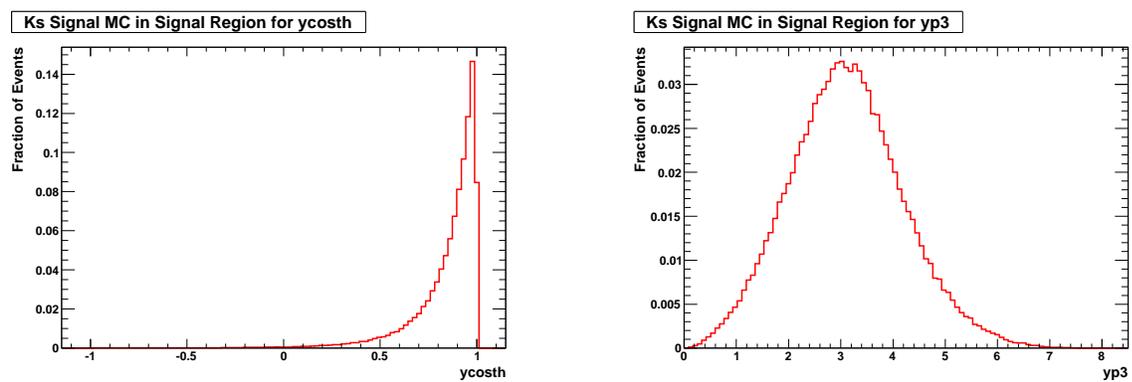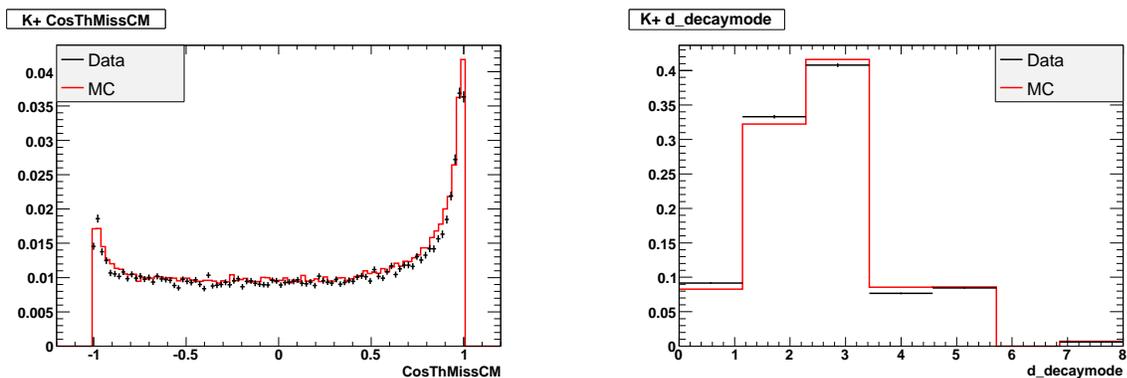
A.2.2 K_S Mode
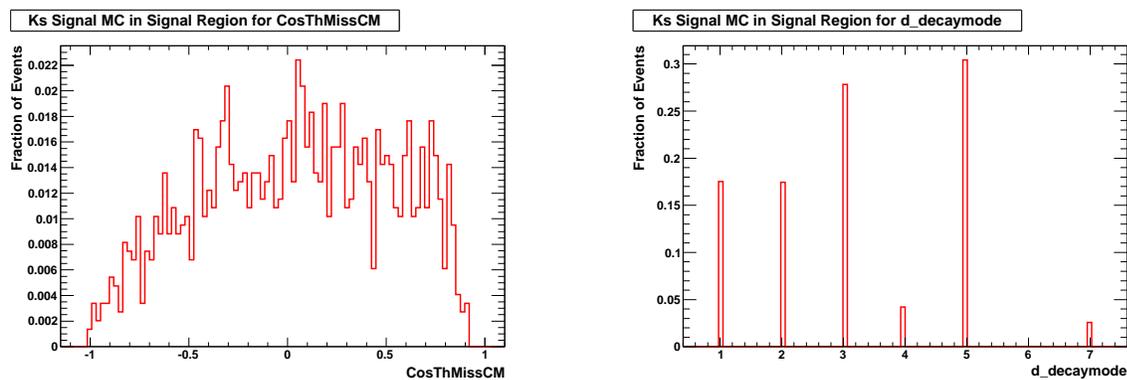
Figure 155: Signal MC histograms for CosThMissCM on left and d_decaymode on right.
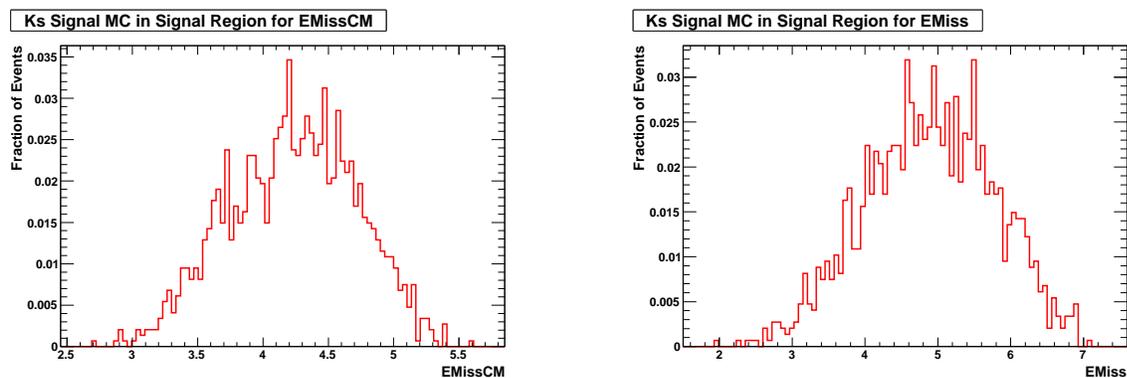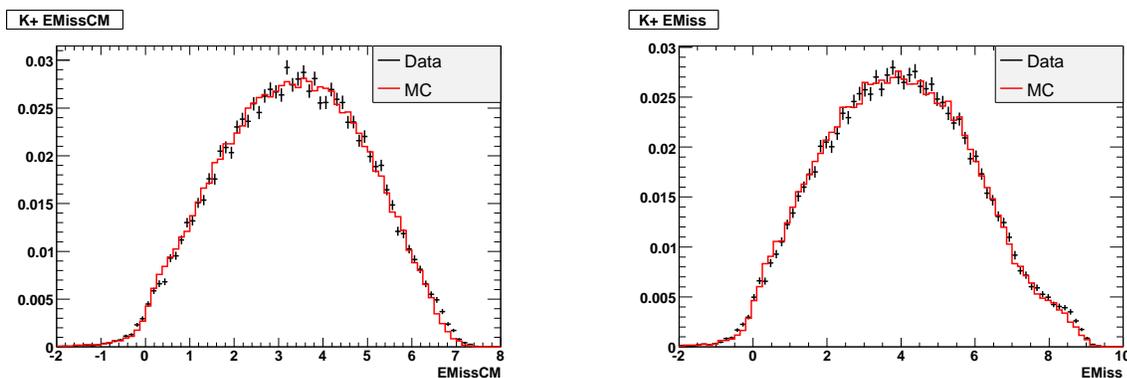

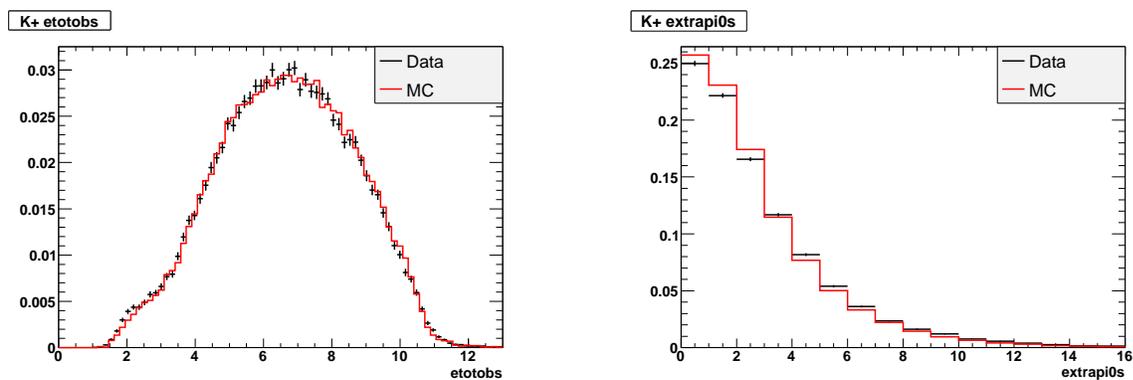Figure 156: Signal MC histograms for EMissCM on left and EMiss on right.
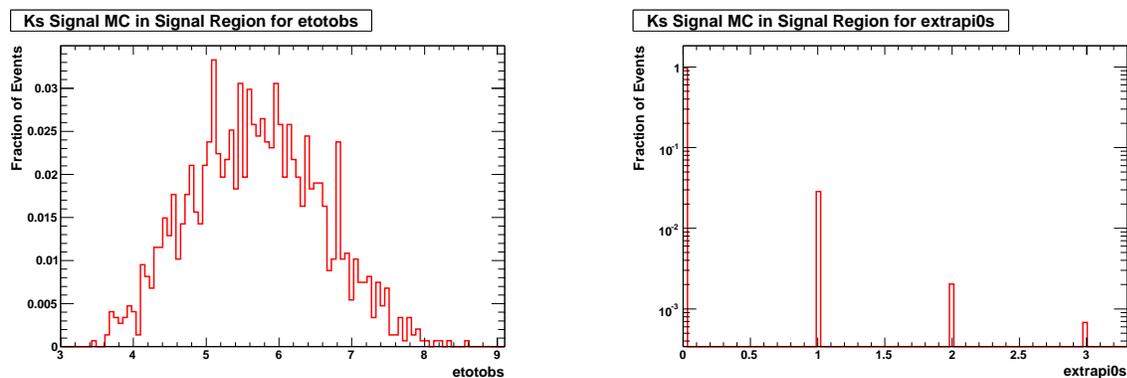
Figure 157: Signal MC histograms for etotobs on left and extrapi0s on right.
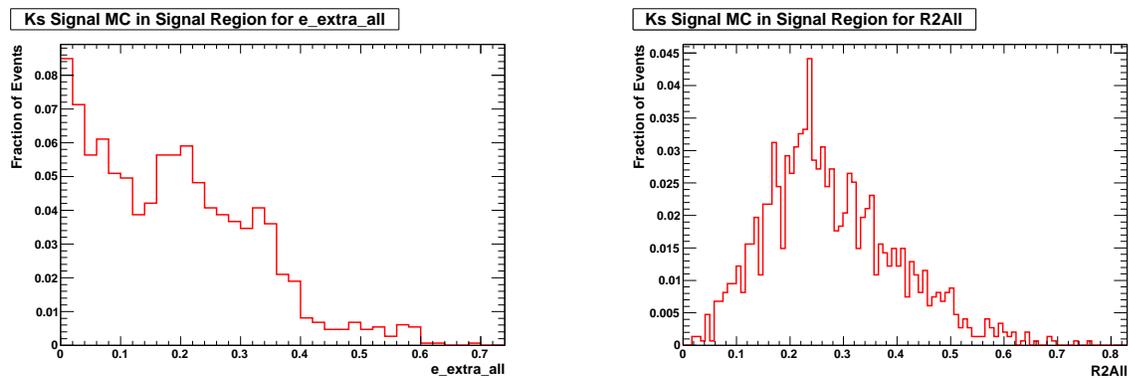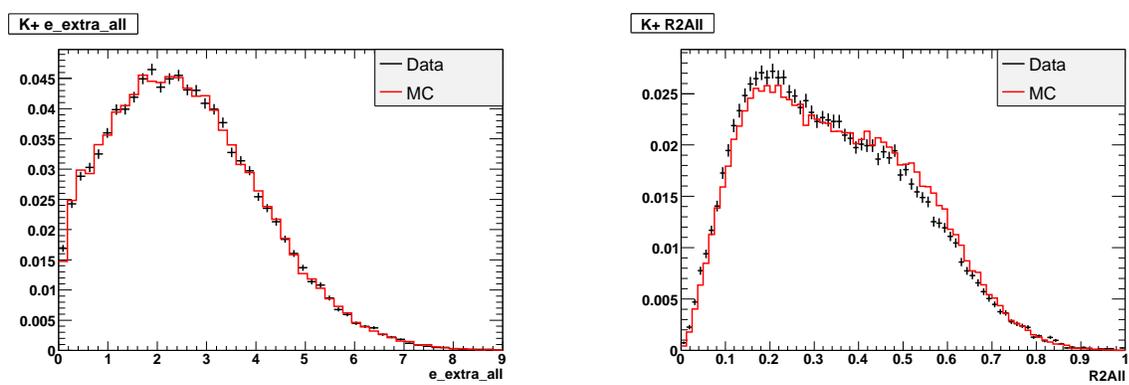
Figure 158: Signal MC histograms for e_extra_all on left and R2All on right.
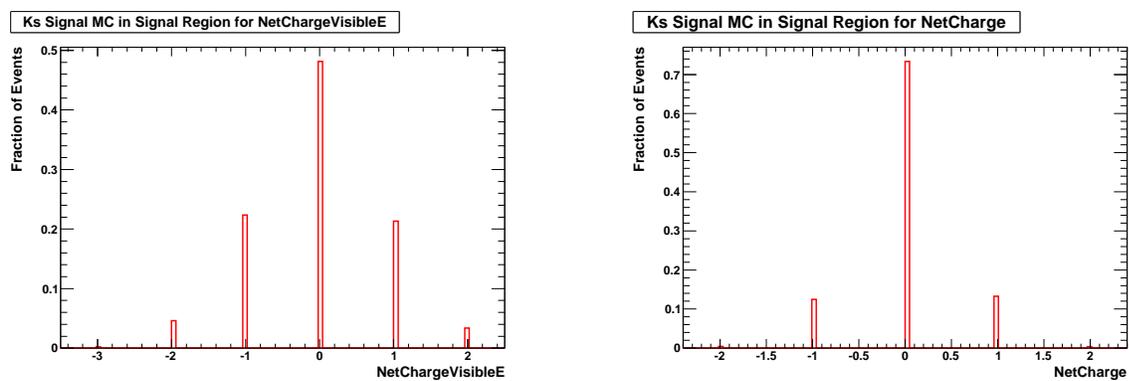
Figure 159: Signal MC histograms for NetChargeVisibleE on left and NetCharge on right.
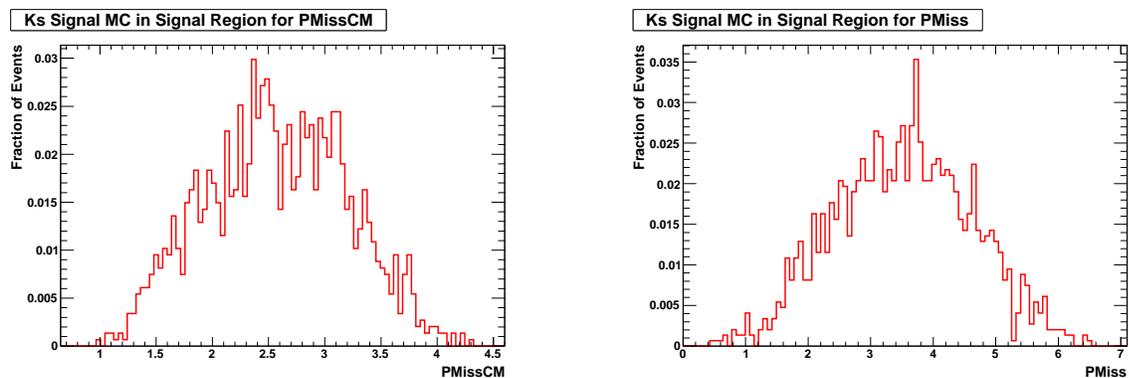
Figure 160: Signal MC histograms for PMissCM on left and PMiss on right.
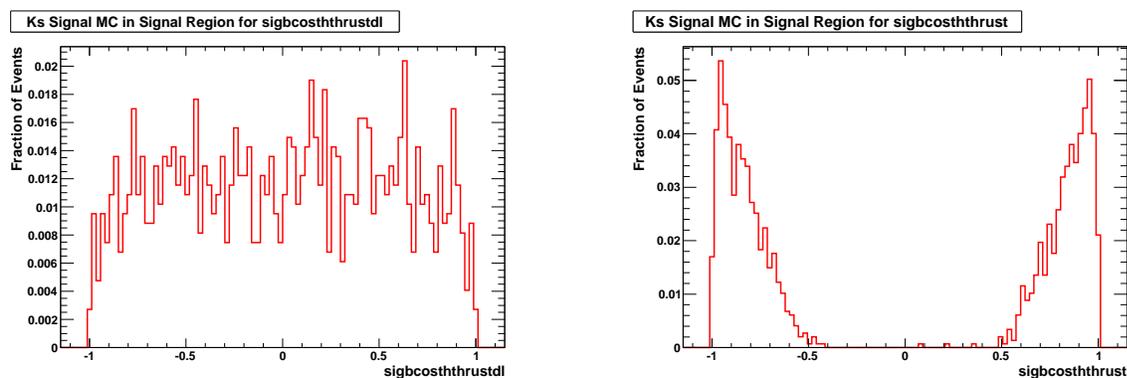
Figure 161: Signal MC histograms for **sigbcosththrustdl** on left and **sigbcosththrust** on right.
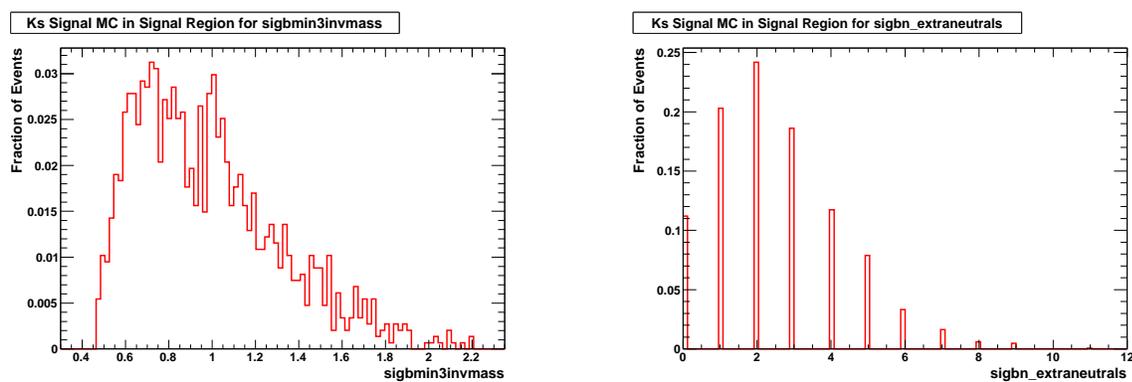
Figure 162: Signal MC histograms for **sigbmin3invmass** on left and **sigbn_extraneutrals** on right.
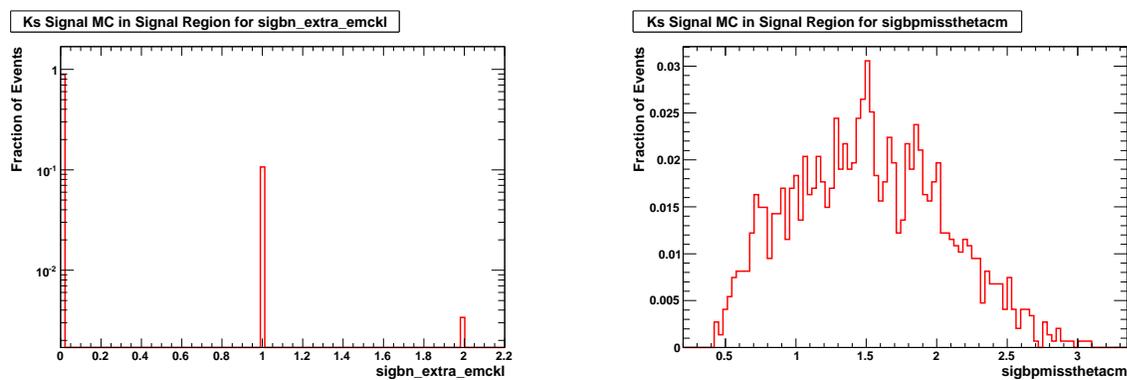
Figure 163: Signal MC histograms for **sigbn_extra_emckl** on left and **sigbpmissthetacm** on right.
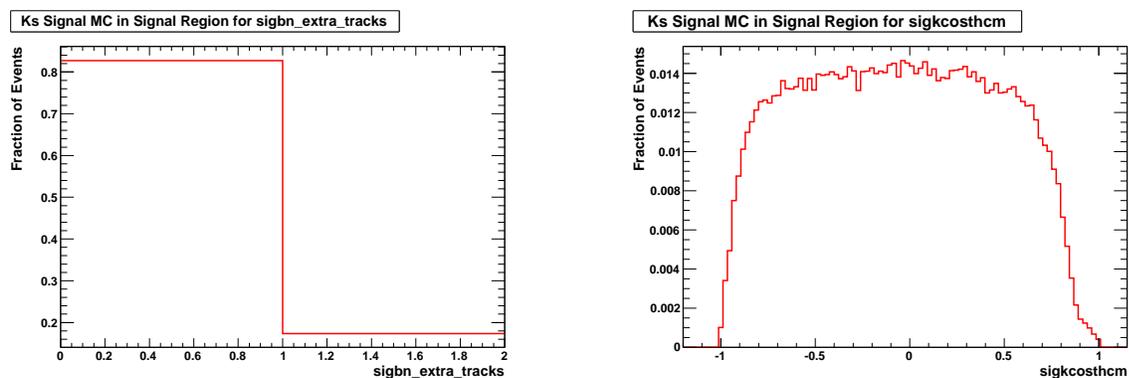
Figure 164: Signal MC histograms for sigbn_extra_tracks on left and sigkcosthcm on right.
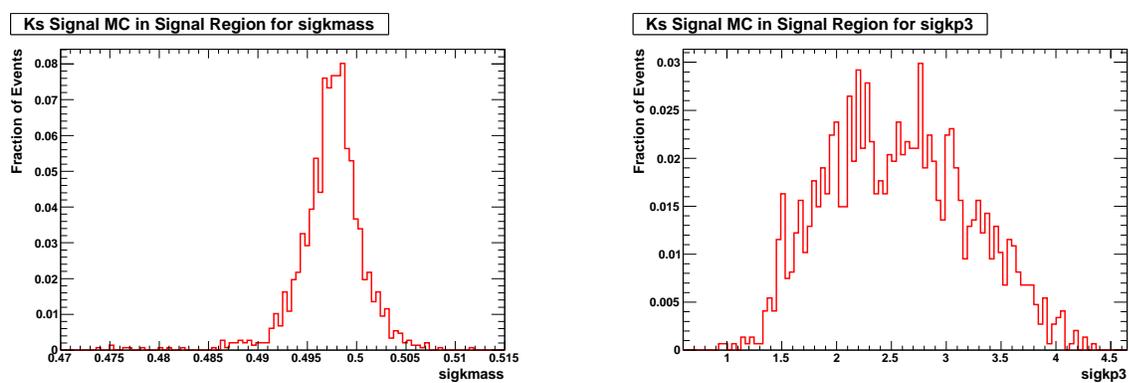
Figure 165: Signal MC histograms for sigkmass on left and sigkp3 on right.
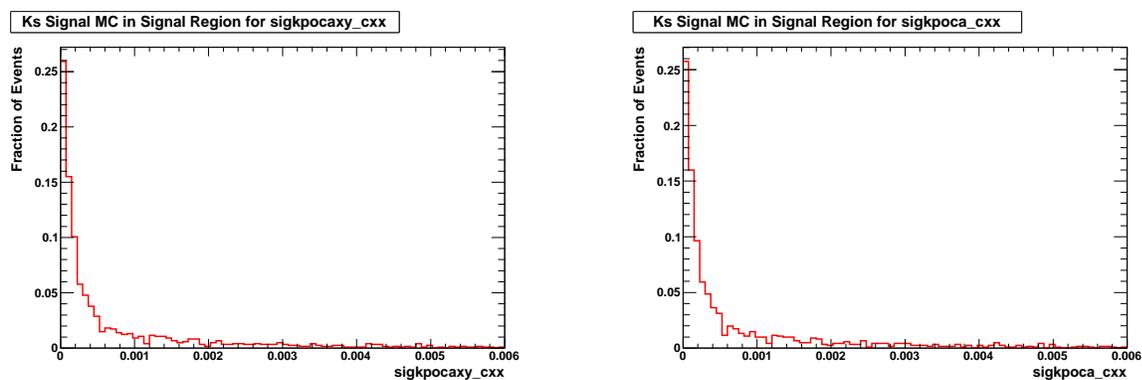
Figure 166: Signal MC histograms for sigkpocaxy_cxx on left and sigkpoca_cxx on right.
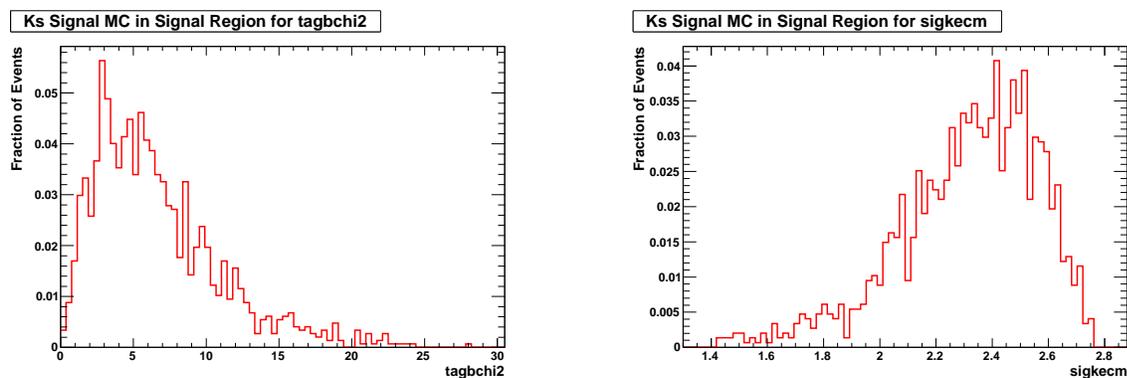
Figure 167: Signal MC histograms for **tagbchi2** on left and **sigkecm** on right.
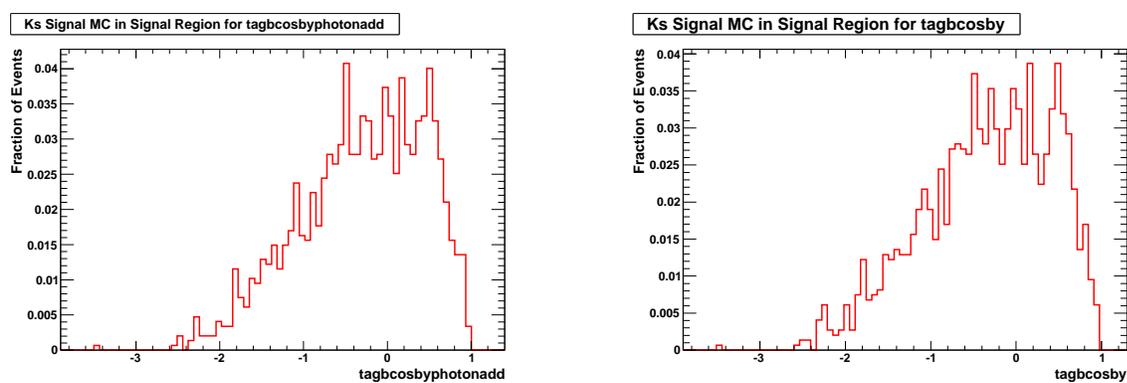
Figure 168: Signal MC histograms for **tagbcosbyphotonadd** on left and **tagbcosby** on right.
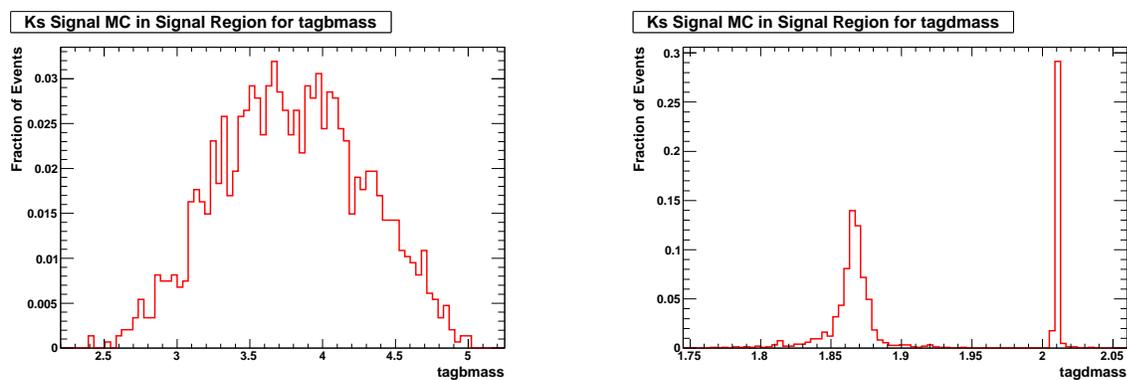
Figure 169: Signal MC histograms for **tagbmass** on left and **tagdmass** on right.
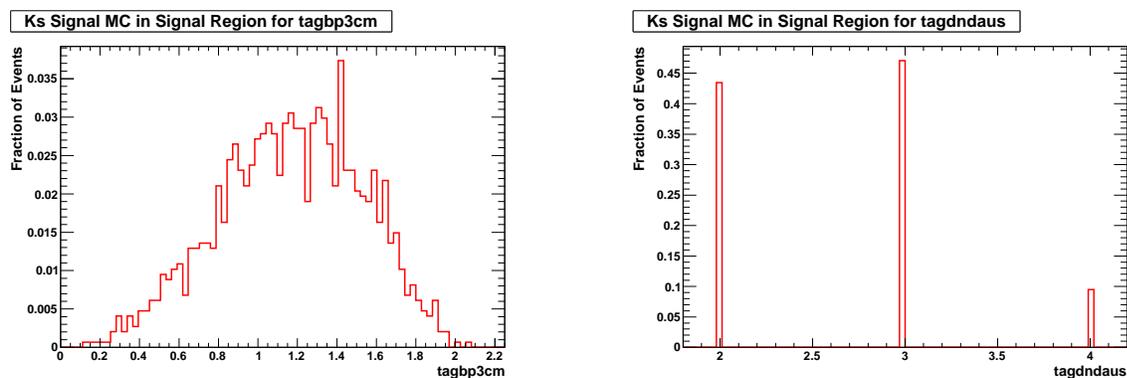
Figure 170: Signal MC histograms for **tagbp3cm** on left and **tagdndaus** on right.
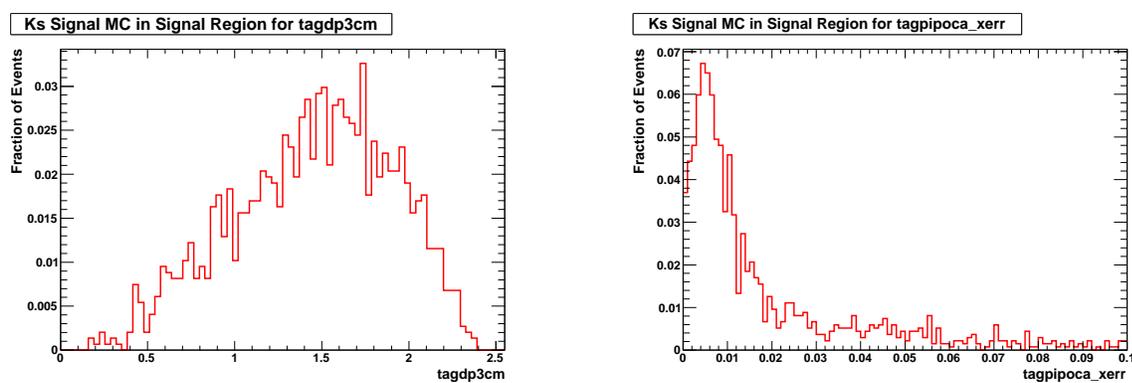
Figure 171: Signal MC histograms for **tagdp3cm** on left and **tagpipoca_xerr** on right.

Figure 172: Signal MC histograms for **taglp3** on left and **taglp3cm** on right.

Figure 173: Signal MC histograms for ycosth on left and yp3 on right.

A.3 Data/MC Correspondence for Classification Variables

The following plots show the correspondence between data and generic background MC for each classification variable. The set of events plotted is the result of the event selection described in Sections 4.2 and 4.3, and the histograms are unit-normalized.

A.3.1 K^+ mode

The background MC events in these plots are weighted as described in Section 5.2.1.

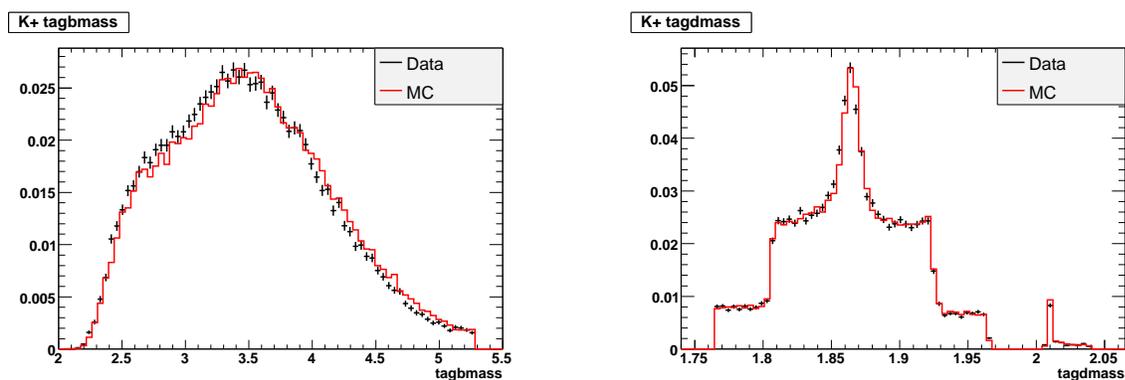Figure 174: Data/MC histograms for CosThMissCM on left and d_decaymode on right.

Figure 175: Data/MC histograms for EMissCM on left and EMiss on right.

Figure 176: Data/MC histograms for etotobs on left and extrapi0s on right.

Figure 177: Data/MC histograms for e_extra_all on left and R2All on right.
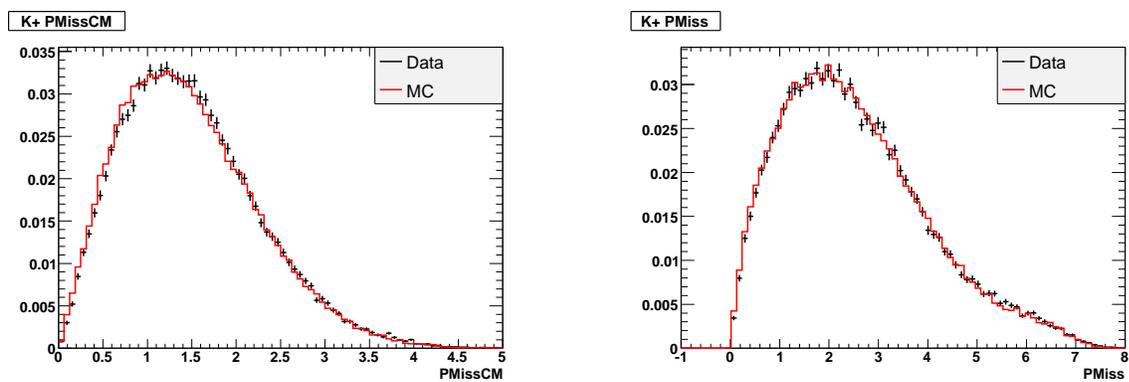
Figure 178: Data/MC histograms for PMissCM on left and PMiss on right.
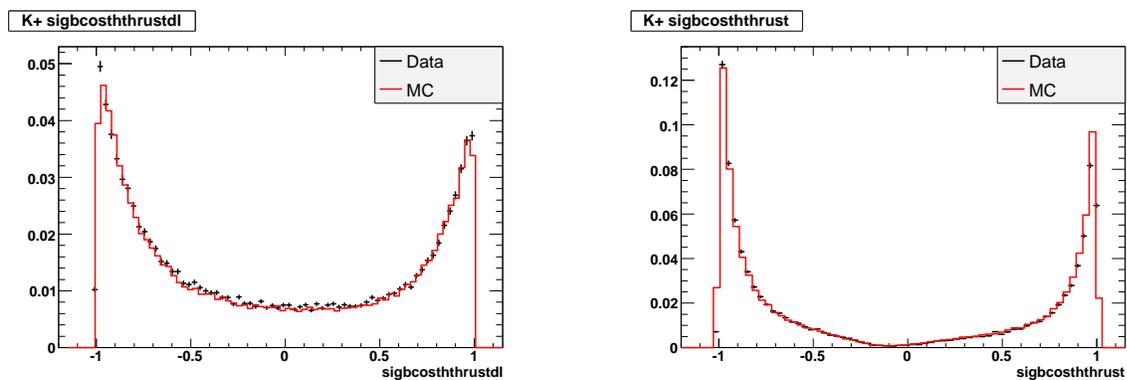
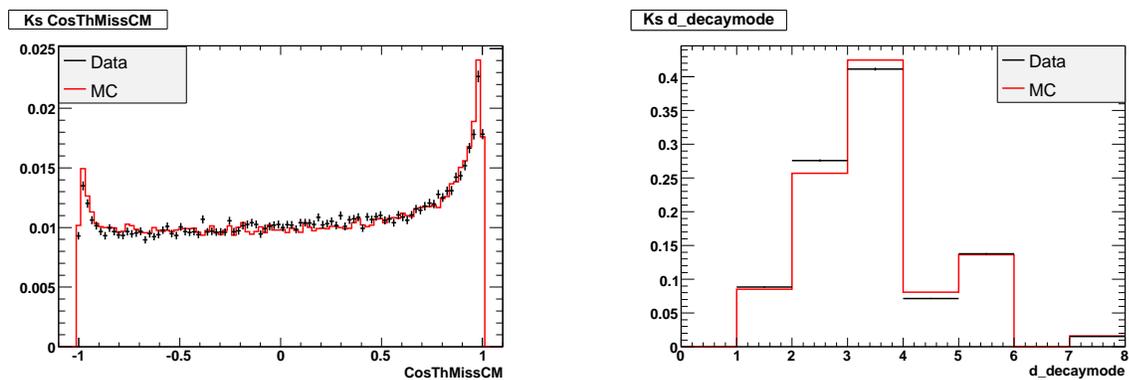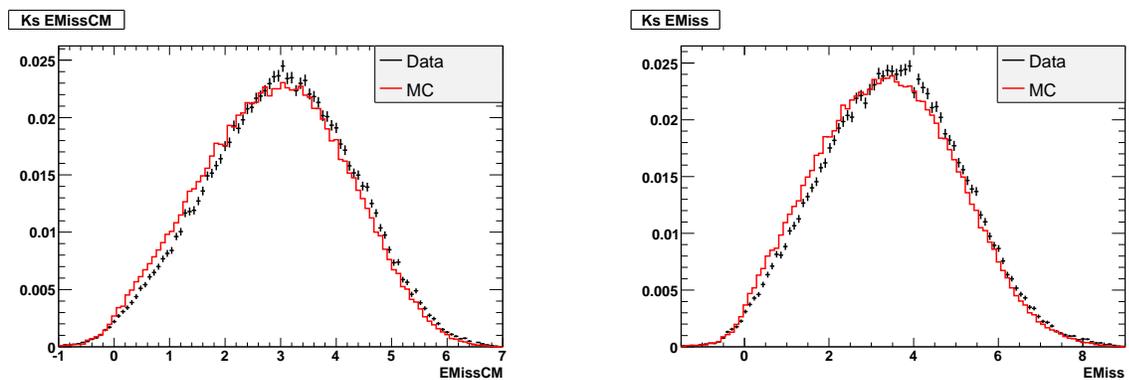Figure 179: Data/MC histograms for **sigbcosththrustdl** on left and **sigbcosththrust** on right.
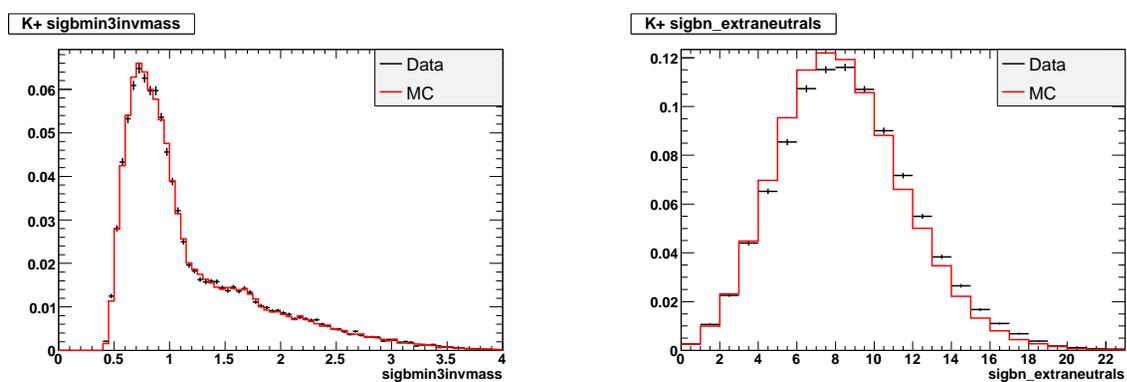
Figure 180: Data/MC histograms for **sigbmin3invmass** on left and **sigbn_extraneutrals** on right.
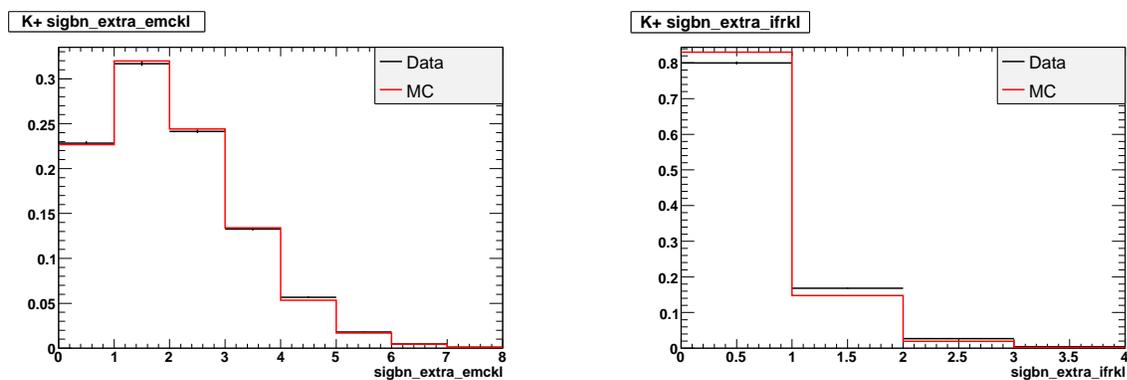
Figure 181: Data/MC histograms for **sigbn_extra_emckl** on left and **sigbn_extra_ifrkl** on right.
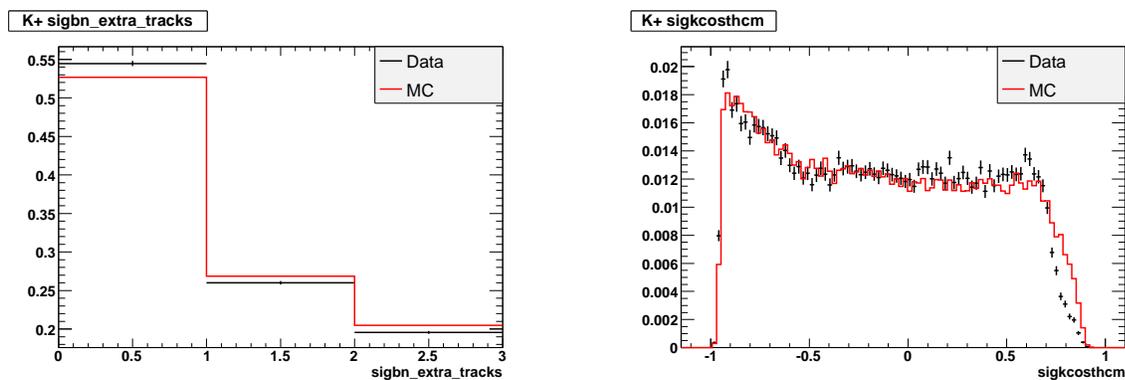
Figure 182: Data/MC histograms for sigbn_extra_tracks on left and sigkcosthcm on right.
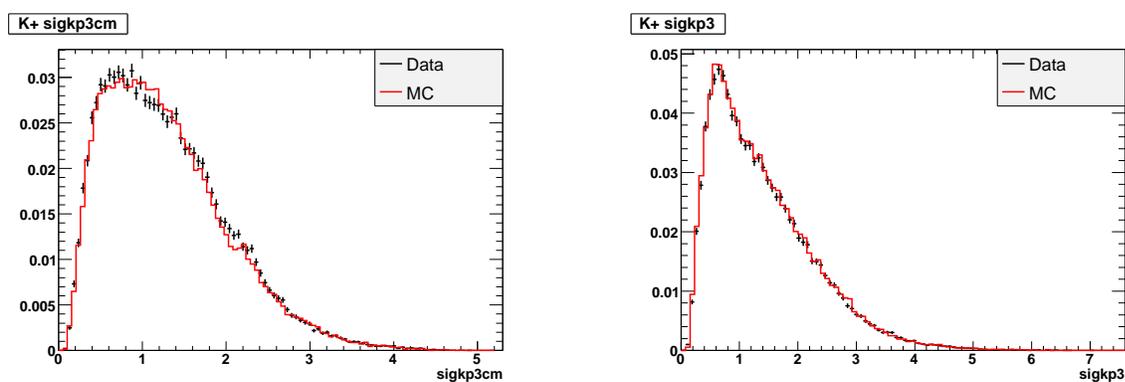
Figure 183: Data/MC histograms for sigkp3cm on left and sigkp3 on right.
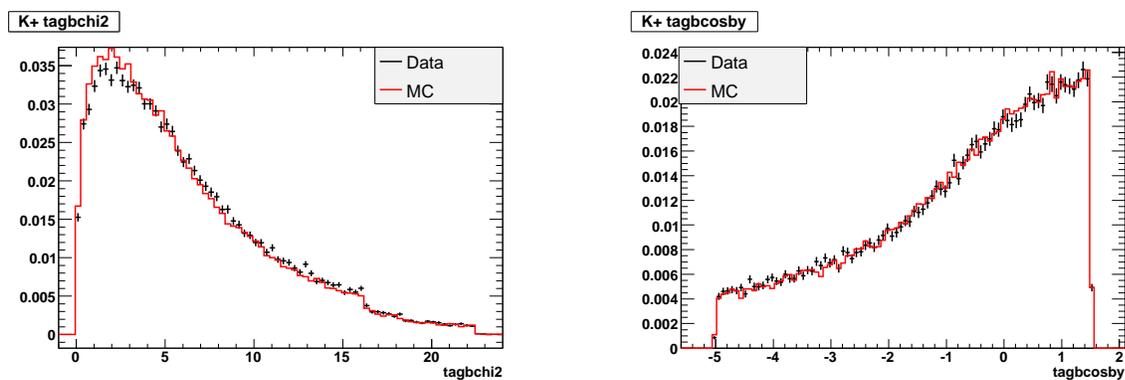
Figure 184: Data/MC histograms for tagbchi2 on left and tagbcosby on right.

Figure 185: Data/MC histograms for tagbmass on left and tagdmass on right.
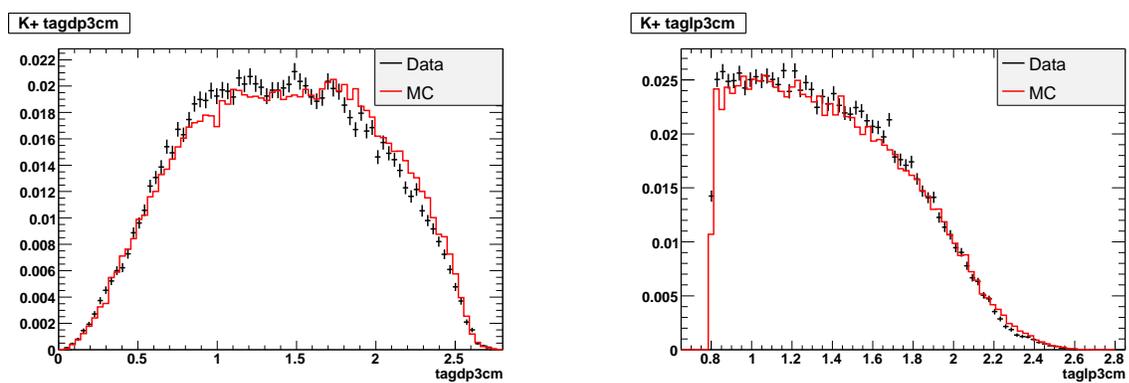
Figure 186: Data/MC histograms for tagdp3cm on left and taglp3cm on right.

A.3.2 K_S mode

Figure 187: Data/MC histograms for CosThMissCM on left and d_decaymode on right.

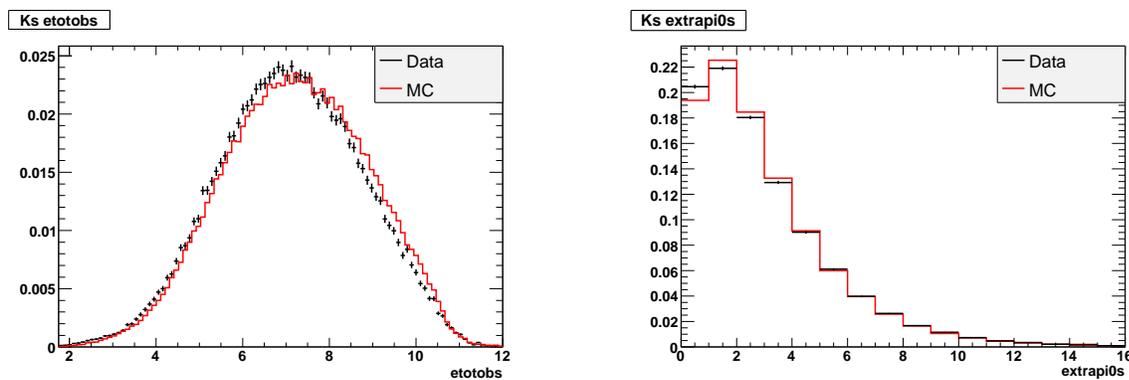Figure 188: Data/MC histograms for EMissCM on left and EMiss on right.

Figure 189: Data/MC histograms for etotobs on left and extrapi0s on right.
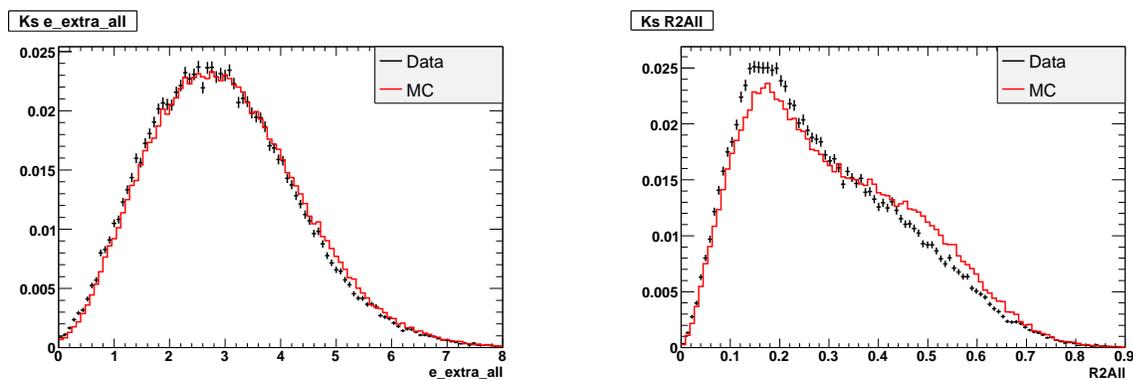
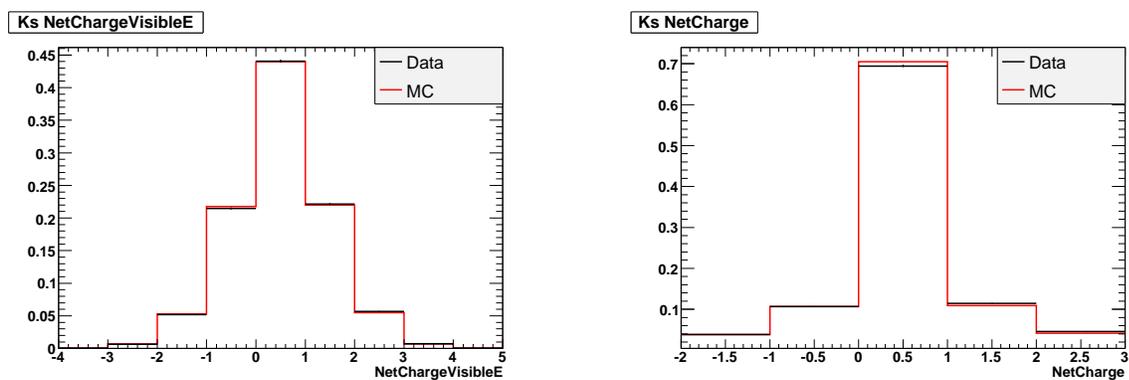Figure 190: Data/MC histograms for e_extra_all on left and R2All on right.

Figure 191: Data/MC histograms for NetChargeVisibleE on left and NetCharge on right.
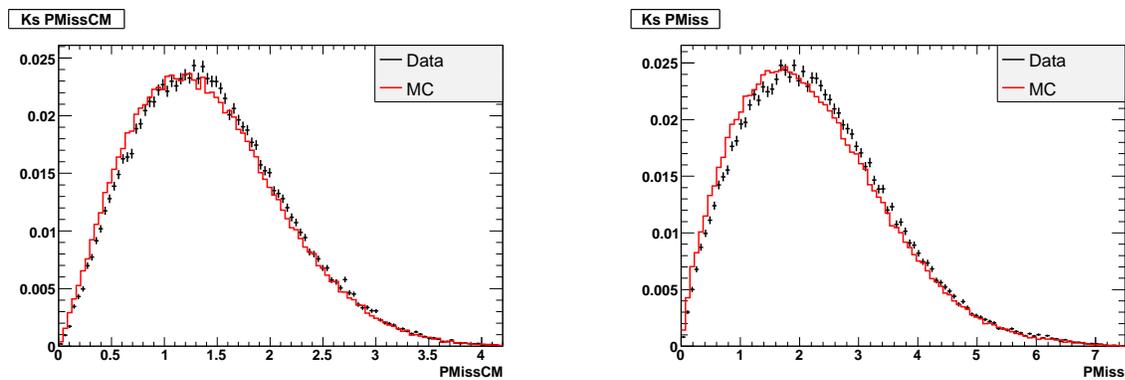

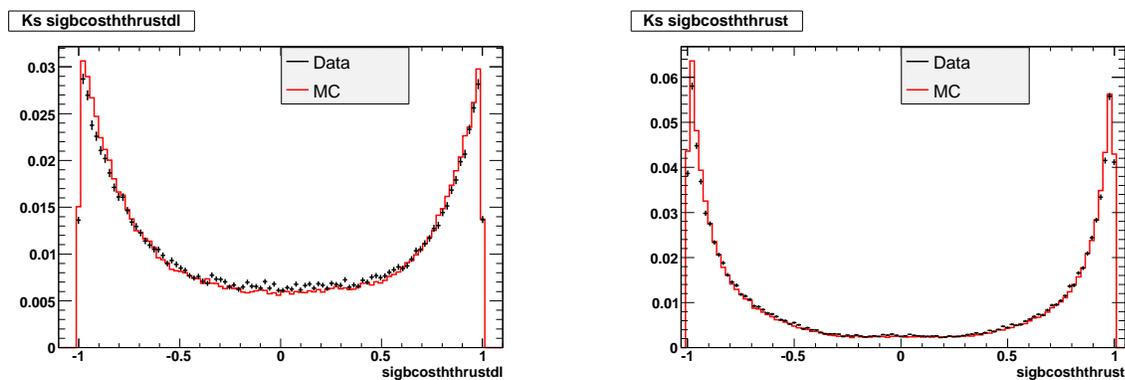Figure 192: Data/MC histograms for PMissCM on left and PMiss on right.

Figure 193: Data/MC histograms for **sigbcosththrustdl** on left and **sigbcosththrust** on right.
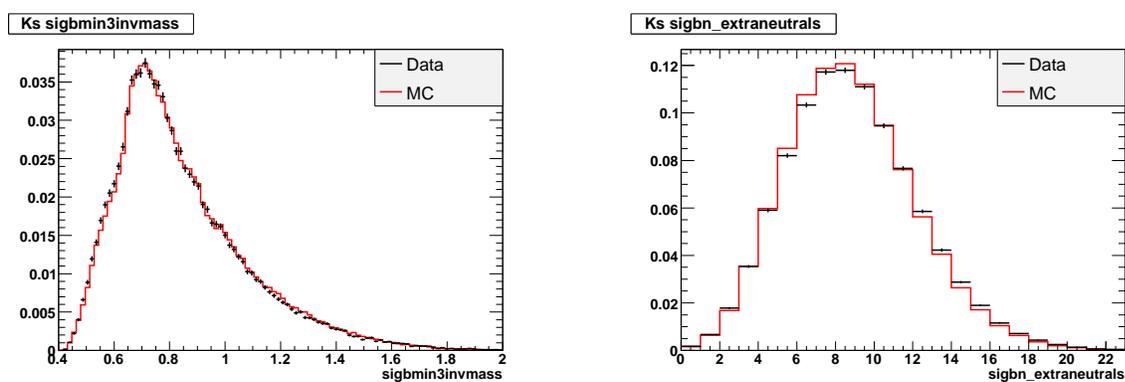

Figure 194: Data/MC histograms for **sigbmin3invmass** on left and **sigbn_extraneutrals** on right.
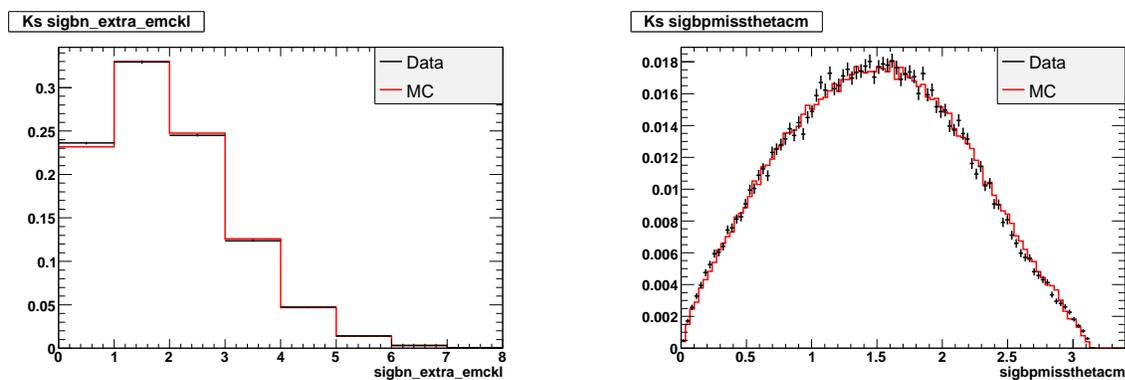

Figure 195: Data/MC histograms for **sigbn_extra_emckl** on left and **sigbpmissthetacm** on right.
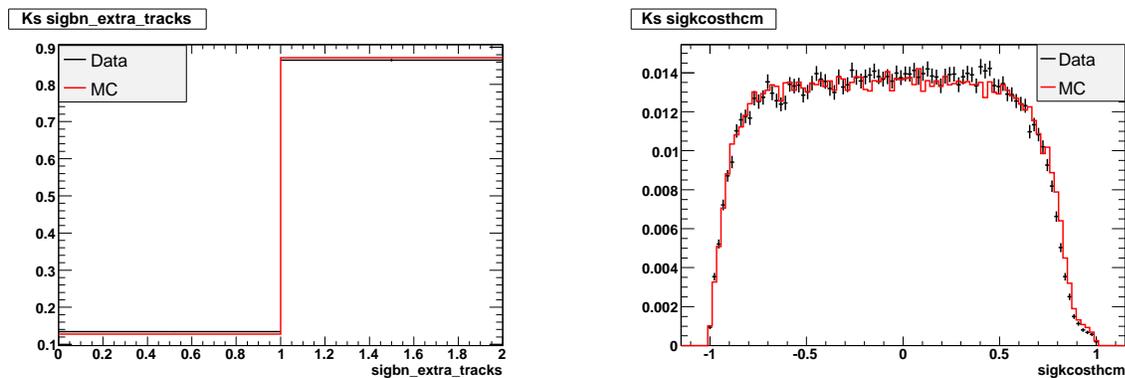
Figure 196: Data/MC histograms for sigbn_extra_tracks on left and sigkcosthcm on right.
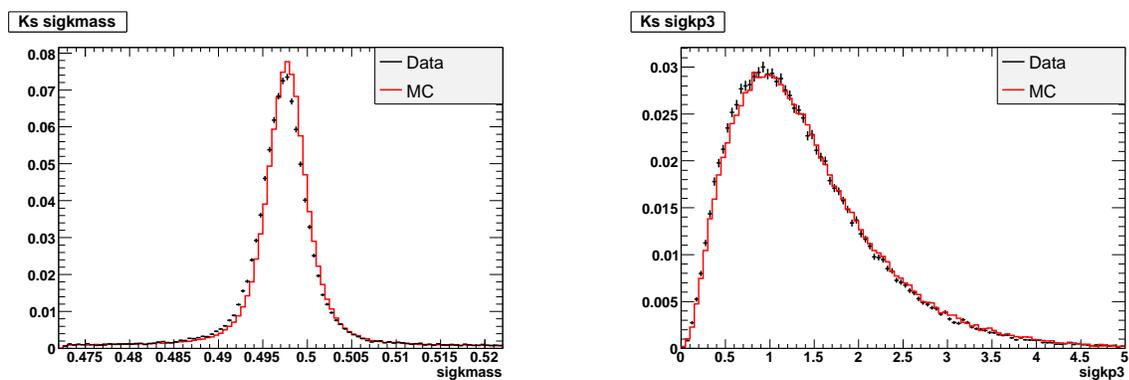

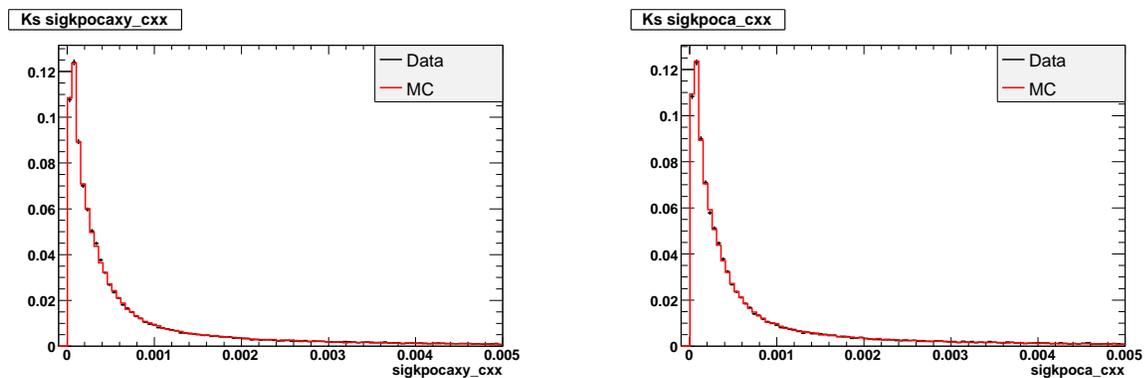Figure 197: Data/MC histograms for sigkmass on left and sigkp3 on right.

Figure 198: Data/MC histograms for sigkpocaxy_cxx on left and sigkpoca_cxx on right.

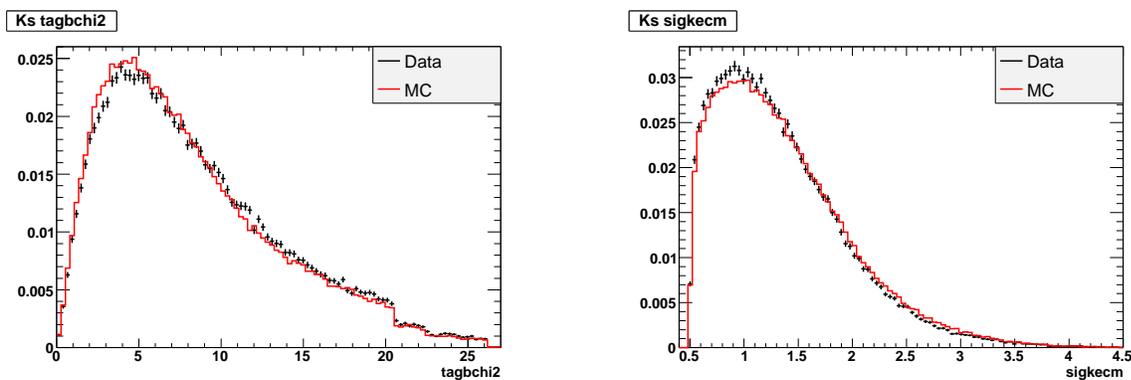Figure 199: Data/MC histograms for **tagbchi2** on left and **sigkecm** on right.
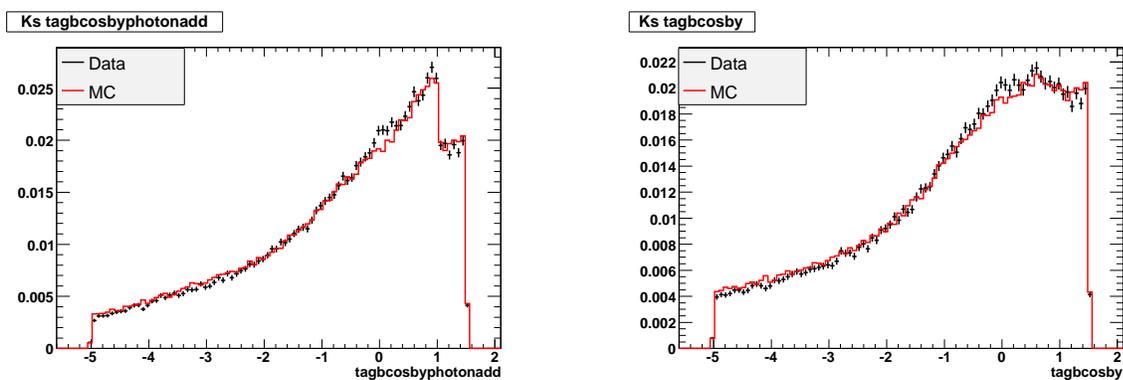
Figure 200: Data/MC histograms for **tagbcosbyphotonadd** on left and **tagbcosby** on right.

Figure 201: Data/MC histograms for **tagbmass** on left and **tagdmass** on right.

Figure 202: Data/MC histograms for **tagbp3cm** on left and **tagdndaus** on right.
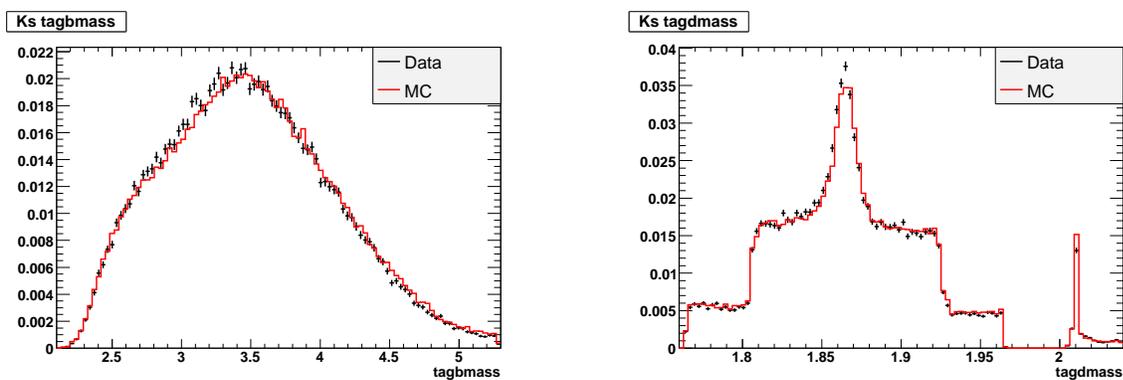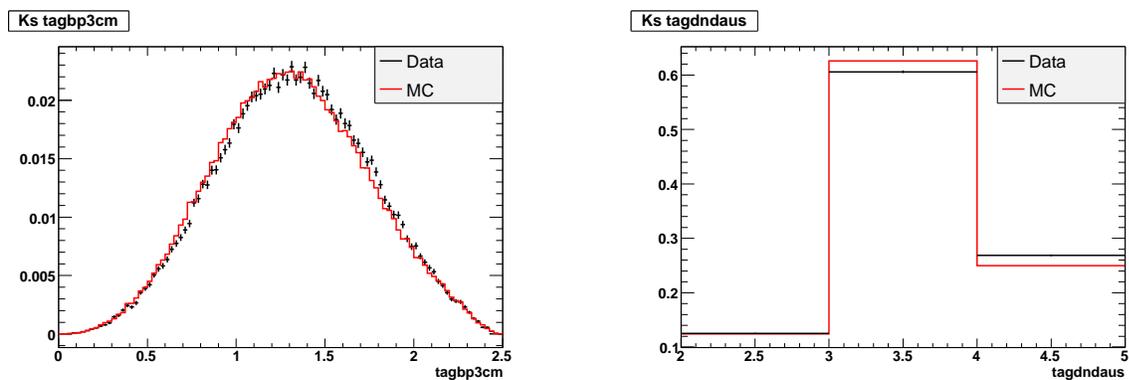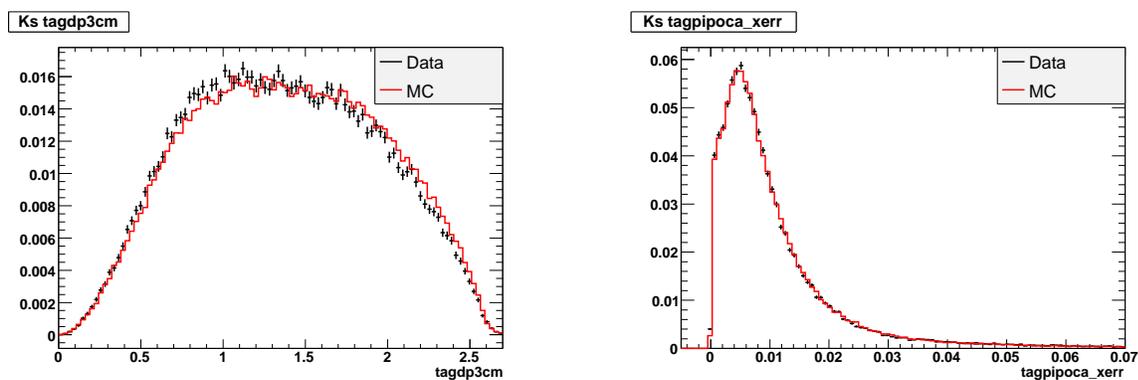

Figure 203: Data/MC histograms for **tagdp3cm** on left and **tagpipoca_xerr** on right.
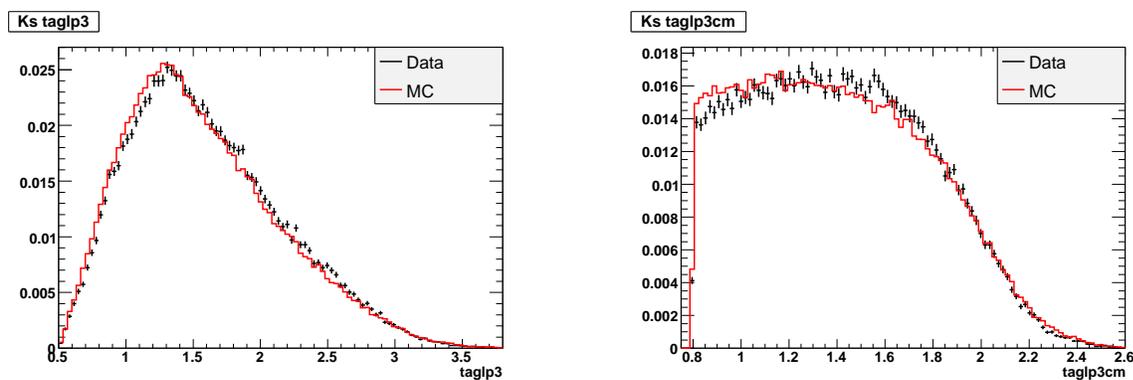
Figure 204: Data/MC histograms for **taglp3** on left and **taglp3cm** on right.
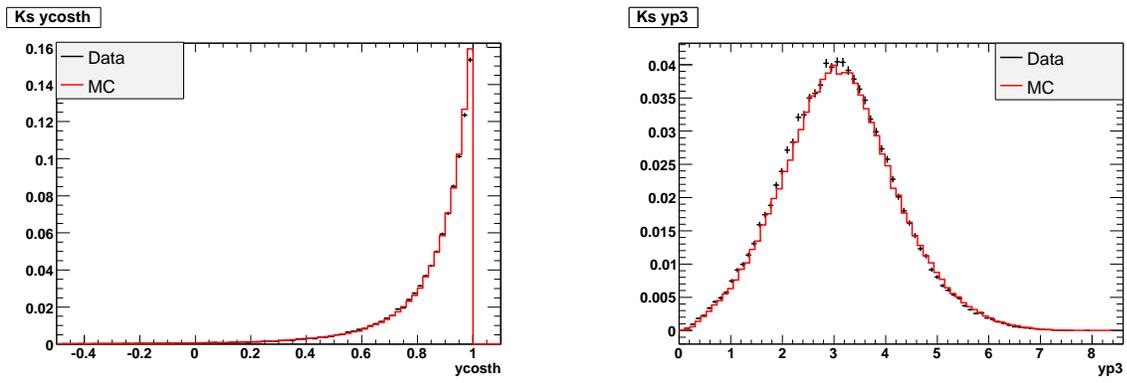
Figure 205: Data/MC histograms for ycosth on left and yp3 on right.

A.3.3 Extra Energy Detail Plots

Figure 206: e_extra_all plots for data and generic background MC events with primary K^+ classifier output values from 0.9 to 0.975 (close to the signal region). The generic background MC events have the 2-D weighting. The left plot shows data and MC histograms, while the right plot shows the data/MC ratio.
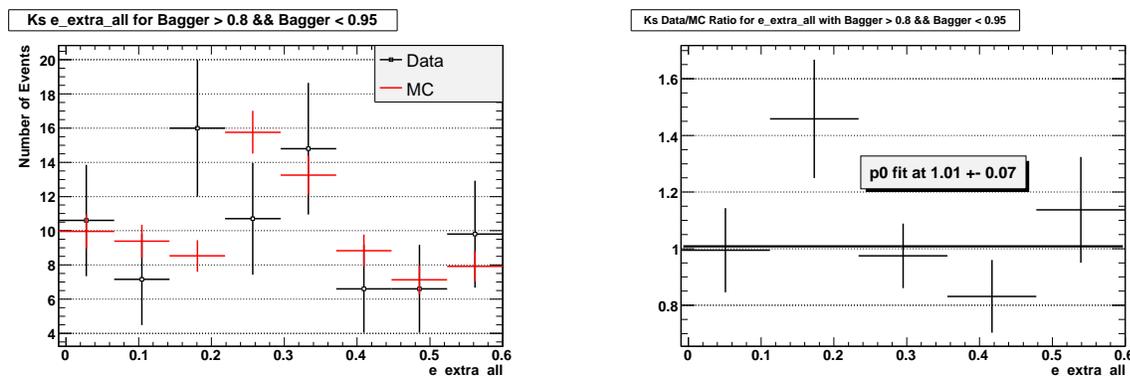
Figure 207: e_extra_all plots for data and generic background MC events with K_S classifier output values from 0.8 to 0.95 (close to the signal region). The generic background MC events are luminosity-weighted only. The left plot shows data and MC histograms, while the right plot shows the data/MC ratio.
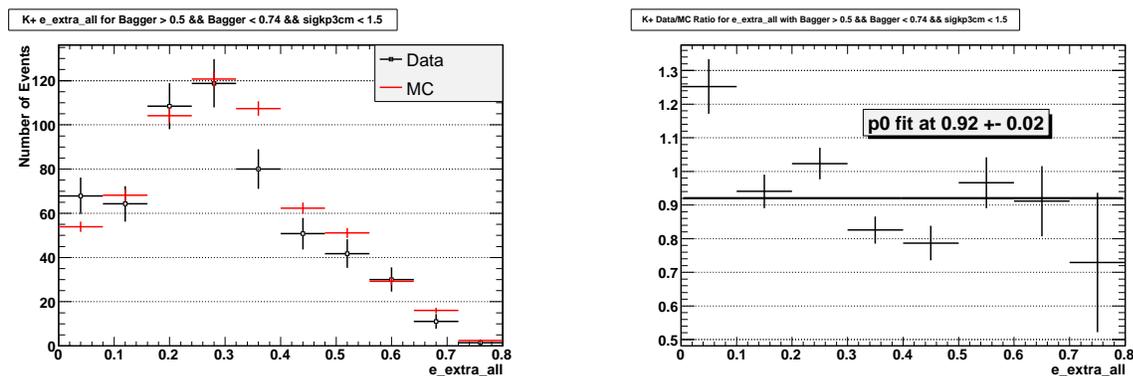
Figure 208: e_extra_all plots for low-kaon-momentum data and generic background MC events with classifier output values from 0.5 to 0.74 (close to the signal region). The generic background MC events have the 2-D weighting. The left plot shows data and MC histograms, while the right plot shows the data/MC ratio.
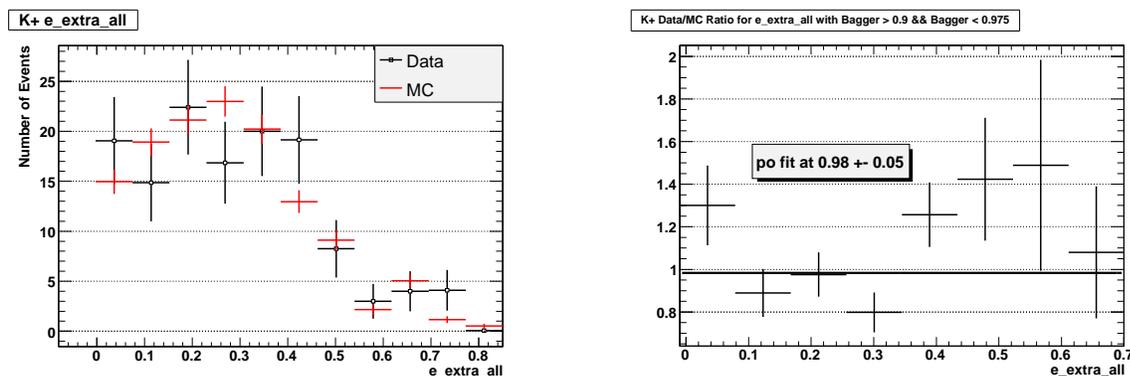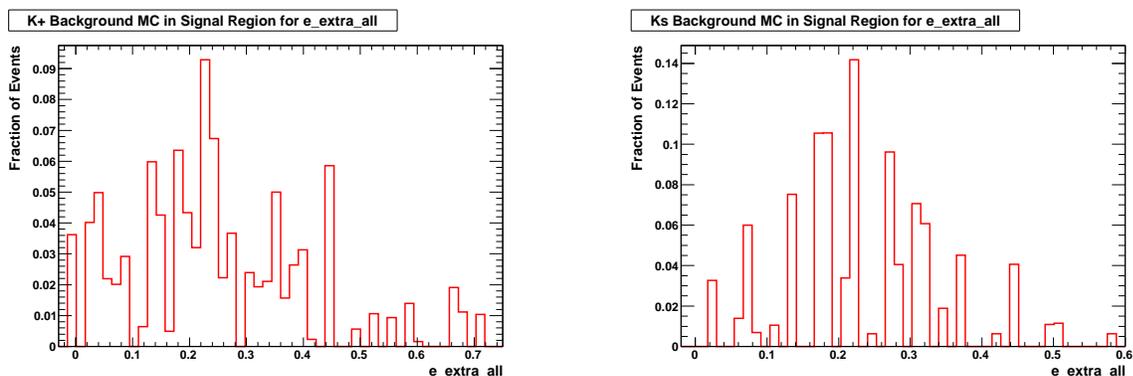

Figure 209: e_extra_all plots for background MC events in the signal region. The K^+ background MC events have the 2-D weighting. The left plot shows the K^+ background MC events, which have the 2-D weighting, while the right plot shows the K_S background.
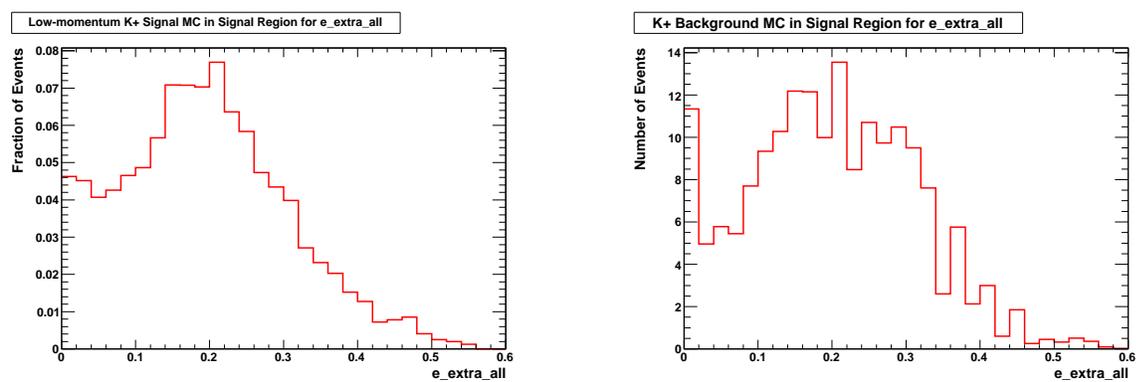
Figure 210: e_extra_all plots for MC events in the low-kaon-momentum signal region. The left plot shows signal MC, while the right one shows background MC, which has the 2-D weighting.

A.4 Data/MC Correspondence for Neutral Extra Energy
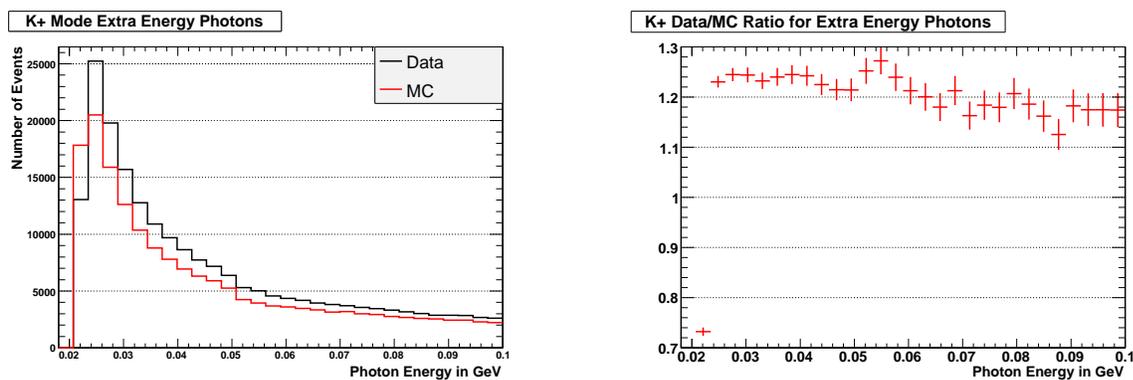
Figure 211: Correspondence between data and generic background MC for extra-energy photons for the K^+ ntuples. The MC is luminosity-weighted, but the 2-D weighting is not applied. The left plot shows data and MC histograms, and the right shows their ratio.
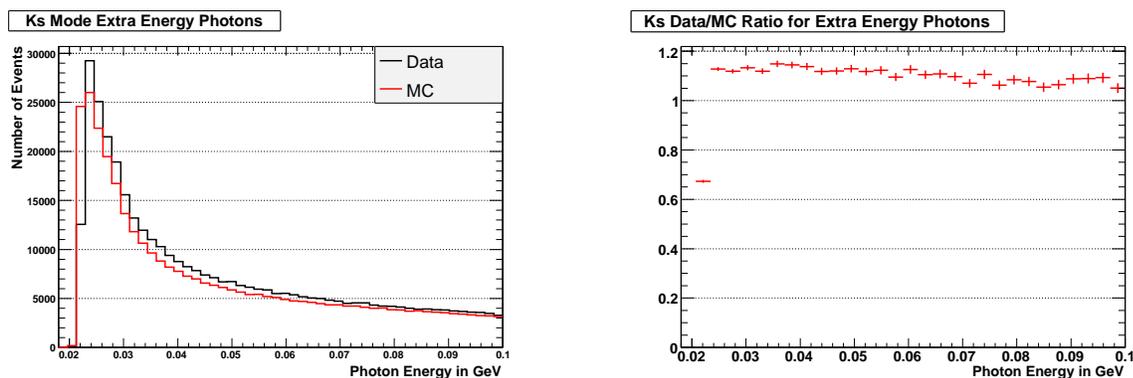
Figure 212: Correspondence between data and generic background MC for extra-energy photons for the K_S ntuples. The MC is luminosity-weighted, but the 10% correction is not applied. The left plot shows data and MC histograms, and the right shows their ratio.

A.5 Efficiency and $m^2_{\nu\bar{\nu}}$ Plots
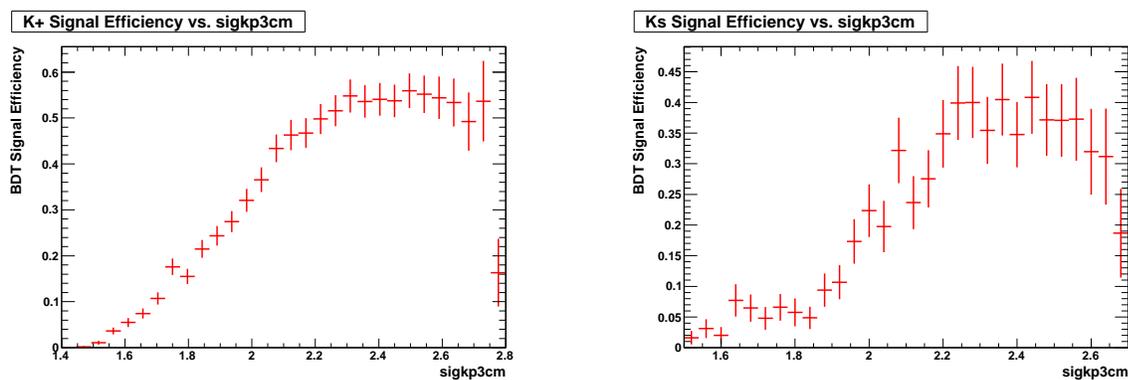
Figure 213: BDT signal efficiency vs. sigkp3cm for the primary K^+ mode on the left and the K_S mode on the right. This efficiency is for the classifier output cut alone, not any previous steps.
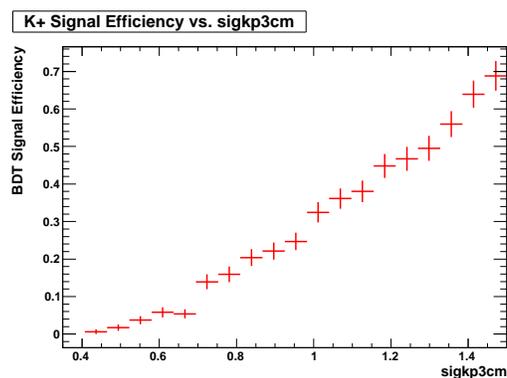
Figure 214: BDT signal efficiency vs. sigkp3cm for the low-momentum measurement. This efficiency is for the classifier output cut alone, not any previous steps.
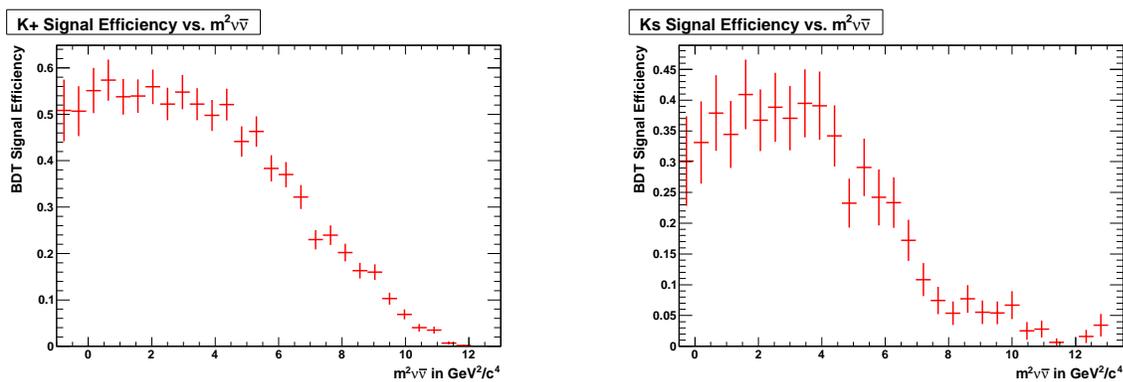
Figure 215: BDT signal efficiency vs. di-neutrino invariant mass squared for the primary K^+ mode on the left and the K_S mode on the right. This efficiency is for the classifier output cut alone, not any previous steps.
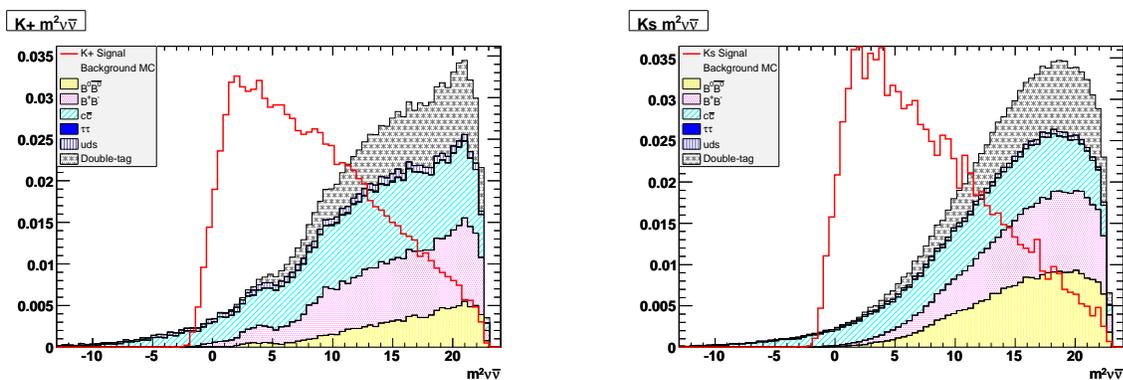
Figure 216: Di-neutrino invariant mass squared for the primary K^+ mode on the left and the K_S mode on the right. These unit-normalized plots include signal and generic background MC from the testing and training sets.
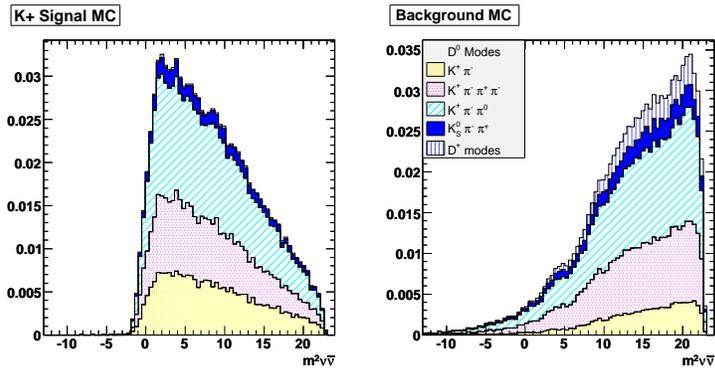
Figure 217: Di-neutrino invariant mass squared for the the primary K^+ mode. These unit-normalized plots include signal and generic background MC from the testing and training sets.
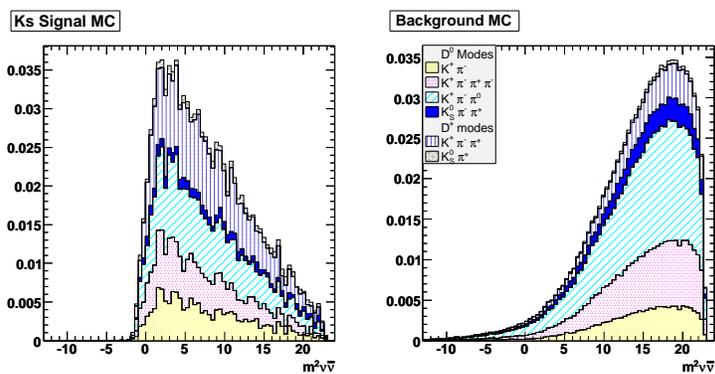
Figure 218: Di-neutrino invariant mass squared for the the K_S mode. These unit-normalized plots include signal and generic background MC from the testing and training sets.

A.6 Reconstruction Sample
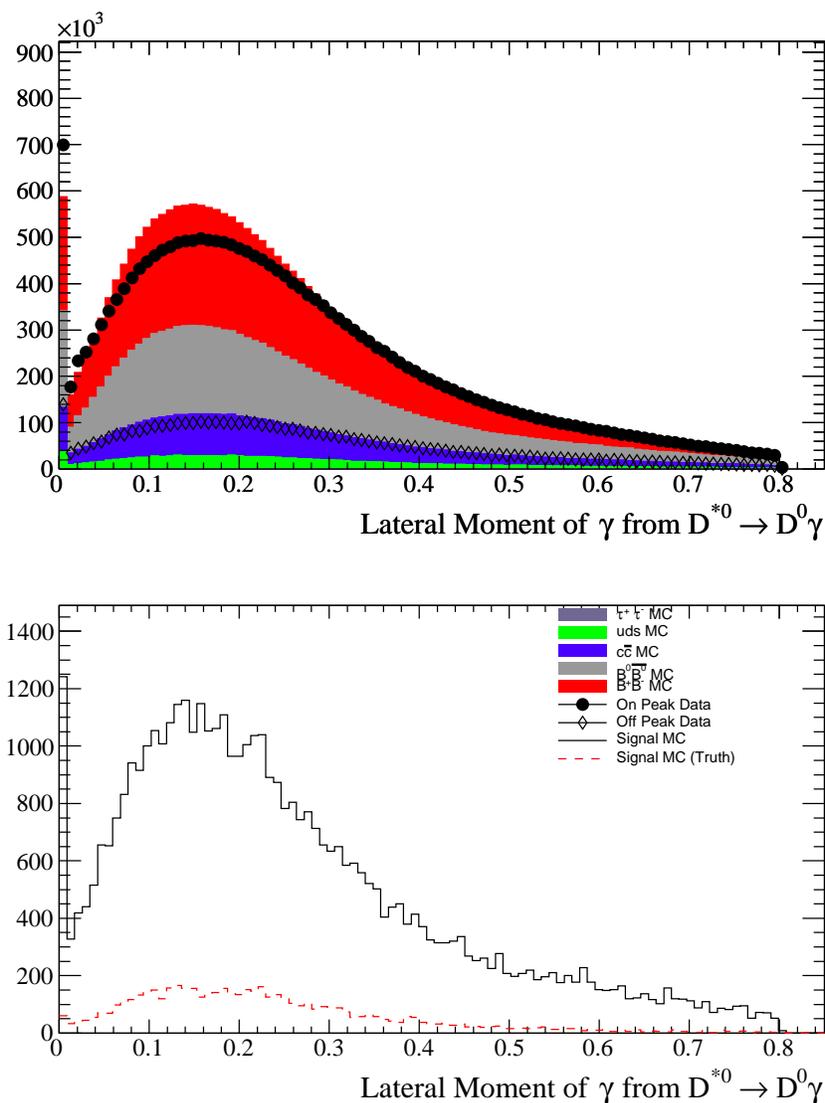
Figure 219: The lateral moment (LAT) distribution for the photon from the transition $D^* \rightarrow D\gamma$. Data and MC are shown for Runs 1-3 in release 18. The signal MC is for the process $B \rightarrow \tau\nu$. The top plot shows on- and off-resonance data overlaid on a stack of generic MC background; the bottom plot shows all reconstructed and reconstructed truth-matched signal MC.

A.7 Individual Tree Results
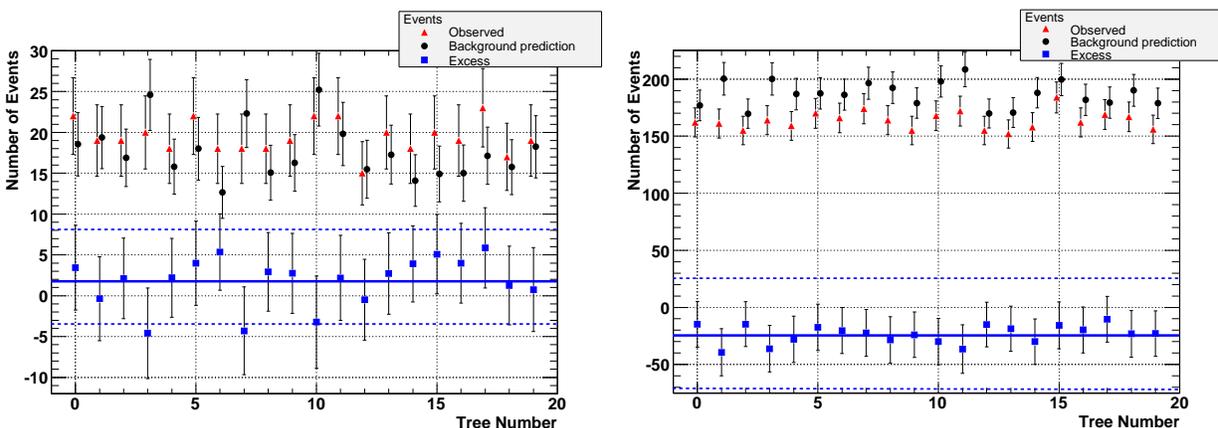
Figure 220: Individual tree results for the K^+ trees. The number of observed events (red triangles), the background prediction (black circles), and number of excess events (blue squares) are shown for each tree, numbered 0 to 19. The average number of excess events is shown by the solid blue line, with the one-sigma uncertainty marked by the dotted blue lines. The left plot the primary K^+ measurement, while the right shows the low-momentum measurement.
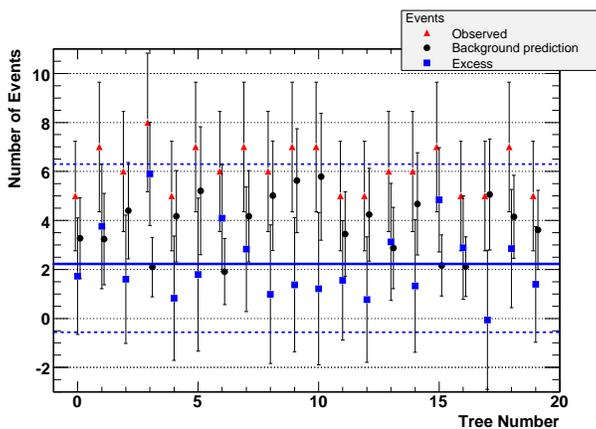
Figure 221: Individual tree results for the K_S trees. The number of observed events (red triangles), the background prediction (black circles), and number of excess events (blue squares) are shown for each tree, numbered 0 to 19. The average number of excess events is shown by the solid blue line, with the one-sigma uncertainty marked by the dotted blue lines.
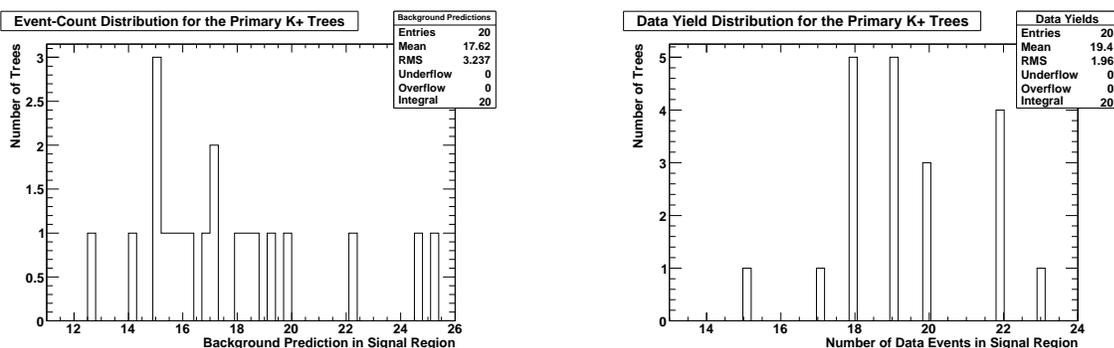
Figure 222: Individual tree results for the twenty K^+ trees. The left plot shows a histogram of the background predictions for the signal region for the twenty trees. The right shows a similar histogram of the actual data yields in the signal region.
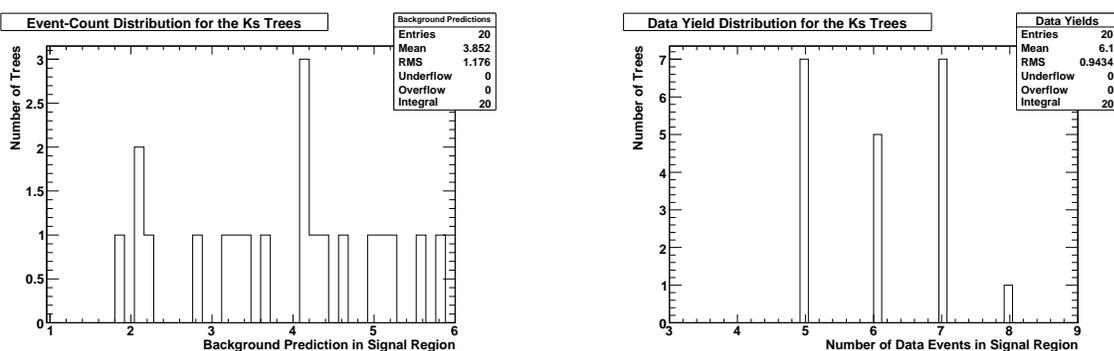
Figure 223: Individual tree results for the twenty K_S trees. The left plot shows a histogram of the background predictions for the signal region for the twenty trees. The right shows a similar histogram of the actual data yields in the signal region.
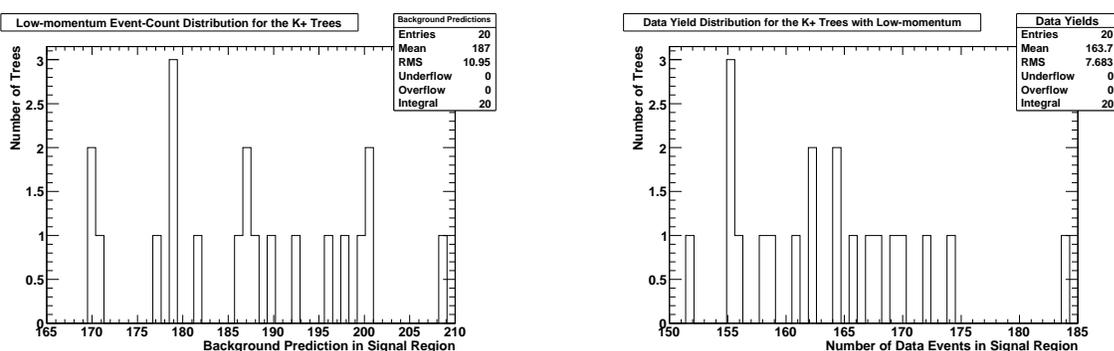
Figure 224: Individual tree results for the low-momentum measurement for the twenty K^+ trees. The left plot shows a histogram of the background predictions for the signal region for the twenty trees. The right shows a similar histogram of the actual data yields in the signal region.

B   Additional Tables

Table 20: Classifier output cut for each of the twenty primary K^+ trees to achieve the target efficiency of 0.175% for that tree.

K^+ Tree Number	Classifier Output Cut
0	0.975837
1	0.975261
2	0.976983
3	0.976564
4	0.975984
5	0.97493
6	0.974588
7	0.978404
8	0.975627
9	0.975758
10	0.9776
11	0.974686
12	0.97473
13	0.975731
14	0.975496
15	0.976173
16	0.976354
17	0.97461
18	0.975836
19	0.976497

Table 21: Classifier output cut for the low-momentum mode on each of the twenty primary K^+ trees to achieve the target efficiency of 0.25% for that tree.

K^+ **Tree Number**	**Classifier Output Cut**
0	0.742055
1	0.742881
2	0.753606
3	0.749602
4	0.746795
5	0.747271
6	0.746164
7	0.740634
8	0.753032
9	0.744778
10	0.751919
11	0.741404
12	0.748958
13	0.750554
14	0.758296
15	0.74216
16	0.744547
17	0.748664
18	0.740907
19	0.751009

Table 22: Classifier output cut for each of the twenty K_s trees to achieve the target efficiency of 0.060% for that tree.

K_s **Tree Number**	**Classifier Output Cut**
0	0.963107
1	0.959777
2	0.958332
3	0.961797
4	0.959017
5	0.961977
6	0.960557
7	0.957792
8	0.960667
9	0.959102
10	0.964032
11	0.962522
12	0.963302
13	0.961742
14	0.960317
15	0.961797
16	0.956226
17	0.959057
18	0.962832
19	0.955011

C   Statistical Uncertainty Calculation for 20 Trees

The background MC sample has limited statistics, so it is desirable to exploit its infor-
mation content to the fullest. The technique of bootstrap sampling is one effective method
of doing so. It involves resampling the same set of events to create multiple replicate sets
of the original. In this analysis, twenty bootstrap training and testing sets are generated
for each type of tree. The resulting statistical uncertainty for the output of twenty trees is
smaller than the uncertainty for one tree because the multiple values allow a more accurate
overall measurement. The formula used is given by Equation 17 [30]. This uncertainty is
used for values obtained by averaging the response of twenty trees for the twenty MC testing
sets.

$$\sigma_s = \sqrt{\frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma_i^2 + \rho\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sigma_i\sigma_j\right)} \tag{17}$$

where σ_s is the statistical uncertainty for the 20 trees together, $N = 20$, $\rho = 0.5$, and σ_i is
the uncertainty for one tree. ρ reflects the correlation between the testing sets of any two
trees, which is 50%.

This formula is derived from the standard method of combining non-independent uncer-
tainties with correlation ρ. A full derivation can be found in [30].

References

[1] F. Halzen and A. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics.* John Wiley and Sons (1984).

[2] H. Abele *et al.* Eur. Phys. J. **C33**, 1 (2004).

[3] M. Peskin and D. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press (1995).

[4] C. Giunti. Nucl. Phys. Proc. Suppl. **169**, 309 (2007).

[5] S. Dodelson. *Modern Cosmology.* Academic Press (2003).

[6] B. Aubert *et al.* (BaBar Collaboration). Nucl. Instrum. Meth. **A479**, 1 (2002).

[7] G. Buchalla, G. Hiller, and G. Isodori. Phys. Rev. D **63**, 014015 (2001).

[8] W. Altmannshofer, A. Buras, D. Straub, and M. Wick. JHEP04 **022** (2009).

[9] K. Chen *et al.* (Belle Collaboration). Phys. Rev. Lett. **99**, 221802 (2007).

[10] J. Comfort *et al.* P14: Proposal for $K_L \to \pi^0 \nu \bar{\nu}$ Experiment at J-Parc (2006). URL `http://j-parc.jp/NuclPart/Proposal_0606_e.html`.

[11] G. Anelli *et al.* (2005). CERN-SPSC-2005-013, URL `http://cdsweb.cern.ch/record/832885/files/spsc-2005-013.pdf`.

[12] E. Witten. Nuclear Physics **B122**, 109 (1977).

[13] J. H. Jeon, C. S. Kim, J. Lee, and C. Yu. Phys. Lett. **B636**, 270 (2006).

[14] P. Ball and R. Zwicky. Phys. Rev. **D71**, 014015 (2005).

[15] A. Ali, P. Ball, L. T. Handoko, and G. Hiller. Phys. Rev. **D61**, 074024 (2000).

[16] T. Aliev, A. Cornell, and N. Gaur. JHEP **0707**, 072 (2007).

[17] Y. Yamada. Phys. Rev. D **77**, 014025 (2008).

[18] P. Colangelo, F. D. Fazio, R. Ferrandes, and T. Pham. Phys. Rev. D **73**, 115006 (2006).

[19] D. McKeen. Phys. Rev. D **79**, 114001 (2009).

[20] C. Bird, P. Jackson, R. Kowalewski, and M. Pospelov. Phys. Rev. Lett. **93**, 201803 (2004).

[21] B. Aubert *et al.* (BaBar Collaboration). Phys. Rev. Lett. **102**, 091803 (2009).

[22] C. Amsler *et al.* (Particle Data Group). Physics Letters **B667**, 1 (2008).

[23] D. J. Lange. Nucl. Instrum. Meth. **A462**, 152 (2001).

[24] I. Narsky (2005). `arXiv:physics/0507157v1`.

[25] G. Punzi. In *Proceedings of PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, pages 79–83. SLAC, Stanford, CA (2003). `arXiv: physics/0308063`.

[26] R. Barlow. Computer Physics Communications **149**, 97 (2002).

[27] B. Aubert (BaBar Collaboration) (2009). `arXiv:0911.1988v2`.

[28] B. Aubert *et al.* (BaBar Collaboration). Phys. Rev. Lett. **94**, 101801 (2005).

[29] T. E. Browder *et al.* (CLEO Collaboration). Phys. Rev. Lett. **86**, 2950 (2001).

[30] R. Barlow. manchester particle physics **99**, 4 (2000). (MAN/HEP/99/4), URL `http://www.hep.man.ac.uk/preprints/1999.html`.