SLAC-R-945

Search for $B \to K\nu\overline{\nu}$ Decays With a 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

2010

Work supported in part by US Department of Energy contract DE-AC02-76SF00515.
SLAC National Accelerator Laboratory, Menlo Park, CA 94025

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

Carl Vuosalo
Under the supervision of Professor Sridhara Dasu
At the University of Wisconsin-Madison

Flavor-changing neutral-current transitions such as $b \to s\nu\overline{\nu}$ 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 $B_u^+ \to K^+\nu\overline{\nu}$ and $B_d^0 \to K_S^0\nu\overline{\nu}$, 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 $B\overline{B}$ 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 $B_u^+ \to K^+\nu\overline{\nu}$, 5.6×10^{-5} for $B_d^0 \to K^0\nu\overline{\nu}$, and the first upper limits on the partial branching fractions for $B_u^+ \to K^+\nu\overline{\nu}$ 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 $B_u^+ \to K^+\nu\overline{\nu}$ and 16×10^{-5} for $B_d^0 \to K^0\nu\overline{\nu}$. They also rule out a new physics model of scalar dark matter for scalar particle masses below 1.7 GeV/c^2.

Acknowledgements

There are many that I must thank for their generous contributions to creation of this dissertation. My deepest gratitude goes to my adviser, Sridhara Dasu, and to Kevin Flood, my closest colleague. I also wish to thank the other helpful members of the Wisconsin BaBar group, Richard Prepost, Henry Band, and Jonathan Hollar. My thanks also go to Steve Sekula for his guidance and the software that is the foundation of this analysis. Further thanks go to my Review Committee, Steve Robertson, Steve Playfer, and Bertrand Echenard, and to the members of my thesis committee, Matthew Herndon, Frank Petriello, and Arun Yethiraj. Thanks also to the current and former conveners of the Leptonic Bottom & Charm Analysis Working Group, Silke Nelson, Guglielmo De Nardo, and Paul Jackson, and to the members of that group, especially Malachi Schram. Finally, I must thank the entire BaBar Collaboration, our PEP-II colleagues, and the SLAC National Accelerator Laboratory, who made possible this research.

Contents

List of Figures

List of Tables

1 Introduction

The Standard Model (SM) [1, 2, 3, 4, 5] 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 tied to particles termed bosons: 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. Quarks and leptons occur in two forms: particle and antiparticle, the latter having the same mass as the particle but being opposite in some properties like charge. 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. 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 flavor via a charged current governed by the weak interaction's charged W boson. The parameters that govern these transitions can be placed in a 3x3 unitary matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [6, 7]. These values must be determined experimentally since they are not predicted by the SM. In contrast to the charged current, quark interactions with the neutral Z boson do not change flavor. Despite the fact that the flavor mixing caused by the weak interaction means that weak eigenstates differ from mass eigenstates, for quark interactions with the Z, the mixing cancels out via the Glashow-Iliopoulis-Maiani (GIM) mechanism [8], so that weak eigenstates and mass eigenstates become identical, and no flavor changes can occur. The GIM mechanism is a reflection of the unitarity of the CKM matrix.

A powerful tool to help determine the rate of an interaction is the Feynman diagram [9]. 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 assumed to be massless by the SM, but there is now compelling evidence neutrinos have small masses [10]. In addition, there is strong evidence for dark matter (DM) in the universe, but no particles in the SM can account for it [11]. Furthermore, the mass of the Higgs boson, the only particle predicted by the SM but not yet observed, is divergent when calculated

under the SM, so corrections to the mass calculation due to new-physics effects are expected to make the Higgs mass finite and compatible with other SM constraints [12]. Additionally, there is the general problem that the SM not only does not include gravity, but it is also actually incompatible with general relativity. All these indications point to the need to look for new physics beyond the SM.

A prominent candidate for resolving some of the problems of the SM is a theory called Supersymmetry, which posits that each fundamental SM particle has a supersymmetric partner. There are many variations of supersymmetry, with one called the Minimal Supersymmetric extension of the Standard Model (MSSM) serving as foundation for most of the others [12]. Arising naturally out of many supersymmetric models is a stable, massive particle that interacts with SM particles only via the weak force. Such a Weakly Interacting Massive Particle (WIMP) is a good candidate for a DM particle.

The *BABAR* experiment [13] 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 $B \to K\nu\overline{\nu}$, 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 $B_u^+ \to K^+\nu\overline{\nu}$ and $B_d^0 \to K_S^0\nu\overline{\nu}$ 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, $B \to K\nu\overline{\nu}$ 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 of FCNC decays, 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 vector bosons, making these decays sensitive to many SM parameters.

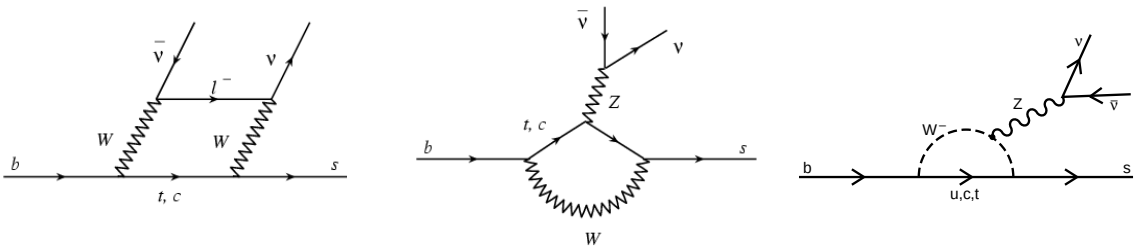

Figure 1: Lowest-order Standard Model diagrams for $b \to s\nu\overline{\nu}$.

Predictions for the branching fraction of $B \to K\nu\overline{\nu}$, according to two recent theoretical models, BHI(2001) [14] and ABSW(2009) [15], are:

$$\mathcal{B}(B \to K\nu\overline{\nu}) = (3.8^{+1.2}_{-0.6}) \times 10^{-6} \ \ [14] \tag{1}$$

$$\mathcal{B}(B \to K\nu\overline{\nu}) = (4.5 \pm 0.7) \times 10^{-6} \ \ [15] \tag{2}$$

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 $B^0_d \to K^0\nu\overline{\nu}$. The rate for $B^0_d \to K^0_S\nu\overline{\nu}$ is predicted to be exactly half of the $B^0_d \to K^0\nu\overline{\nu}$ rate, with the other half being $B^0_d \to K^0_L\nu\overline{\nu}$, which is extremely difficult to reconstruct. The differences between the BHI and ABSW models are discussed in Section 2.

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

$$\mathcal{B}(B^+_u \to K^+\nu\overline{\nu}) < 1.4 \times 10^{-5} \ \ [16] \tag{3}$$

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

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

Another theoretically clean FCNC decay, $K \to \pi\nu\overline{\nu}$, is targeted by two proposed experiments: K^0TO for $K_L \to \pi^0\nu\overline{\nu}$ [17] and NA62 for $K^+ \to \pi^+\nu\overline{\nu}$ [18]. 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, electroweak 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 separates the effects at each scale [19, 20].

An effective Hamiltonian can be constructed using the OPE. For $B \to K\nu\overline{\nu}$, the basic form of Hamiltonian under the SM is:

$$H_{eff} = C_L^{SM}\mathcal{O}_L \text{ with } \mathcal{O}_L = (\bar{s}_L\gamma^\mu b_L)(\bar{\nu}\gamma_\mu(1-\gamma^5)\nu) \quad [21] \tag{5}$$

A more detailed version of the effective Hamiltonian is:

$$H_{eff} = \frac{G_F\alpha}{2\pi\sqrt{2}}V_{tb}V_{ts}^* C_{10}^\nu \bar{s}\gamma^\mu(1-\gamma^5)b\overline{\nu}\gamma_\mu(1-\gamma^5)\nu + h.c. \ [22] \tag{6}$$

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 [19]. 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 $K \to \pi\nu\overline{\nu}$ decays, which shows the close relation between $B \to K\nu\overline{\nu}$ and $K \to \pi\nu\overline{\nu}$. C_{10}^ν is sometimes labelled C_L^ν or X (when multiplied by a constant). It is an electroweak V-A (Vector minus Axial vector) Wilson coefficient, with a value dependent upon the top quark mass [23]. Some theoretical calculations for the value of C_{10}^ν are -6.38 ± 0.06 [15], -6.6 [14], and -7 [24]. The theoretical uncertainty in the value of C_{10}^ν is small because of the absence of long-distance effects and the fact that QCD corrections are quite small [25]. The main uncertainty comes from uncertainty in the value

of the top quark mass.

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

$$\frac{d\Gamma}{ds} = \frac{G_F^2 \alpha^2 m_B^5}{2^8 \pi^5} \left|V_{tb} V_{ts}^*\right|^2 \left|C_{10}^{\nu}\right|^2 \Lambda^{3/2}(r_K, s) f_+^2(s) \quad [22] \tag{7}$$

$$\Lambda \equiv 1 + r_K^2 + s^2 - 2s - 2r_K(1+s),\ s = m_{\nu\bar{\nu}}^2/m_B^2,\ \ r_K = m_K^2/m_B^2 \tag{8}$$

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.

A rough calculation of the branching fraction using Equation 7 gives a result within an order of magnitude of the more precise values given by Equations 1 and 2:

$$G_F = 1.2 \times 10^{-5}, \alpha = 7.3 \times 10^{-3}, m_B = 5.3 \text{ GeV}, V_{tb} = 1, V_{ts} = 0.04, C_{10}^{\nu} \cong -6$$

$$\Gamma_B = 1/1.6 \times 10^{-12}\text{s},\ \bar{h} = 6.6 \times 10^{-25}\text{GeV s.}\quad \text{Assume } \Lambda \sim \mathcal{O}(1), f_+ \sim \mathcal{O}(1).$$

$$\text{BF} = \frac{\Gamma_{B\to K\nu\bar{\nu}}}{\Gamma_B} \cong \frac{(1.2 \times 10^{-5} \cdot 7.3 \times 10^{-3})^2 \cdot 5.3^5 \cdot 0.04^2 \cdot 36}{2^8 \pi^5} \cdot \frac{1.6 \times 10^{-12}\text{GeV s}}{6.6 \times 10^{-25}\text{GeV s}} \cong 6 \times 10^{-5}$$

As seen in Equations 1 and 2, the ABSW and BHI models have slightly different predictions for the $B \to K\nu\bar{\nu}$ branching fraction, due largely to the fact that they use different form factor models. The form factors are calculated with light-cone sum rules, but the two models use the results of different papers; ABSW uses [26], while BHI uses [27]. The BHI model is older, is referenced by the $B \to K\nu\bar{\nu}$ literature, and provides a basis of comparison with previous $B \to K\nu\bar{\nu}$ results, while the ABSW model was not published until 2009 and uses an improved form factor model that includes one-loop radiative corrections to the expansion of the OPE. The ABSW and BHI models differ in the distribution of the differential branching fraction across values of $m_{\nu\bar{\nu}}^2$ – the ABSW model predicts more events with

high values of $m^2_{\nu\bar{\nu}}$, as shown in Figure 2. (The phase-space model included on the plot is described in Section 5.1.2. It is not fully realistic but its simplicity provides a baseline for comparison with the other models.)

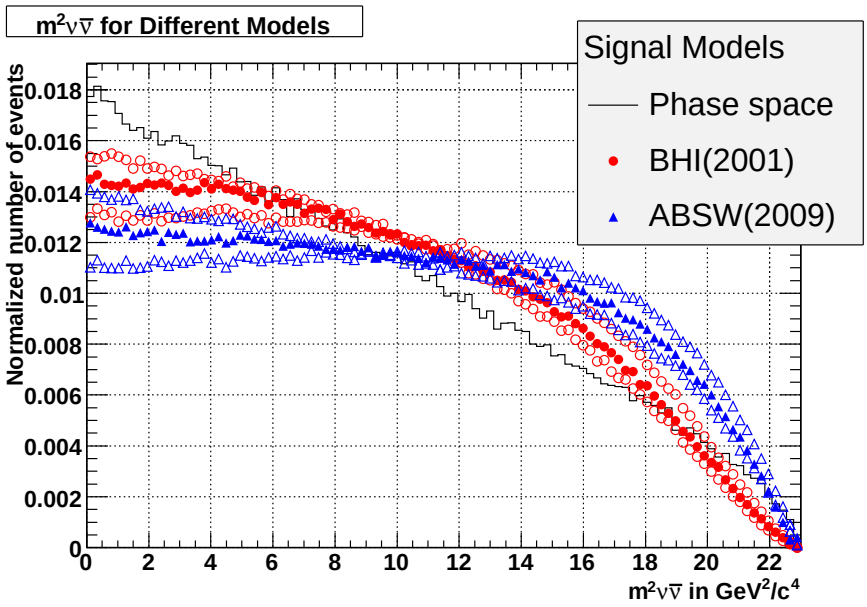

Figure 2: The di-neutrino invariant mass squared for the three signal models. For the two theoretical models, the theoretical uncertainties are shown by open markers of the matching shape. The theoretical uncertainties are due to uncertainties in the form factor models. Note that low values of the di-neutrino invariant mass squared correspond to high values of the signal kaon momentum, and vice-versa. See Section 5.1.2 for a description of the phase-space model.

$B \to K\nu\overline{\nu}$ has several theoretical features that make it an important target of study. With only one Wilson coefficient and operator, theoretical uncertainties are limited to the value of a single Wilson coefficient and form factor, rather than multiple interfering coefficients and operators seen in other decays. In addition, the transition to neutrinos is a very clean process lacking long-distance effects. Third, if the branching fraction could actually be measured, the low theoretical uncertainties could allow a calculation of the important V_{ts} CKM parameter. Alternatively, the branching fraction could be used to provide an experimental value of C^{ν}_{10},

which is shared by the $K \to \pi\nu\overline{\nu}$ decay, and then significant CKM parameters could be calculated if the $K \to \pi\nu\overline{\nu}$ branching fraction could be measured. Finally, there are only two specific ways for new physics to change the $B \to K\nu\overline{\nu}$ decay: either a change in the value of C_{10}^{ν} or the addition of a new Wilson coefficient and operator [21], as explained below.

The loops in the decay diagrams for $B \to K\nu\overline{\nu}$ (Figure 1) provide opportunities for the appearance of new physics, if new particles were to appear in the loops. One new physics model that could increase the $B \to K\nu\overline{\nu}$ branching fraction is the MSSM [15, 24]. In this model several new supersymmetric particles can contribute to the decay: the charged Higgs($H^{\pm}$), the chargino ($\tilde{\chi}^{\pm}$), the neutralino ($\tilde{\chi}^{0}$), the gluino ($\tilde{g}$), and squarks ($\tilde{u}, \tilde{c}, \tilde{t}, \tilde{d}, \tilde{s}, \tilde{b}$). Figures 3 and 4 show Feynman diagrams with the new particles. With reference to Equation 5, the effective Hamiltonian becomes:

$$H_{eff} = C_L\mathcal{O}_L + C_R\mathcal{O}_R \text{ with } \mathcal{O}_R = (\bar{s}_R\gamma^{\mu}b_R)(\bar{\nu}\gamma_{\mu}(1-\gamma^5)\nu) \tag{9}$$

The new particles contribute to the new Wilson coefficients C_L and C_R:

$$C_L = C_L^{SM} + C_L^{\tilde{g}} + C_L^{\tilde{\chi}^{\pm}} + C_L^{\tilde{\chi}^{0}} + C_L^{H^{\pm}} \tag{10}$$

$$C_R = C_R^{\tilde{g}} + C_R^{\tilde{\chi}^{\pm}} + C_R^{\tilde{\chi}^{0}} + C_R^{H^{\pm}} \tag{11}$$

Among these contributions, only those from the chargino are likely to be comparable to the SM. The rest are constrained by experimental limits from $b \to s\gamma$ and $B_s \to \mu^+\mu^-$ decays. The modified $B \to K\nu\overline{\nu}$ branching fraction is:

$$3.5 \times 10^{-6} < \mathcal{B}(B_u^+ \to K^+\nu\overline{\nu}) < 5.8 \times 10^{-6} \text{ [15]} \tag{12}$$

This branching fraction range is based upon the MSSM parameter values shown in Table 1, with parameter set I giving the higher value and parameter set II the lower.

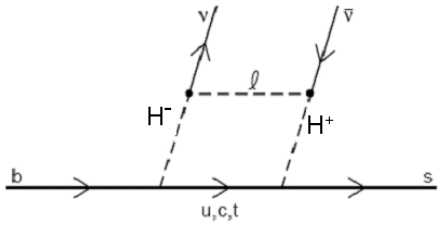

Figure 3: New-physics Feynman diagram for $b \to s$ decay that has charged Higgs bosons in the box.

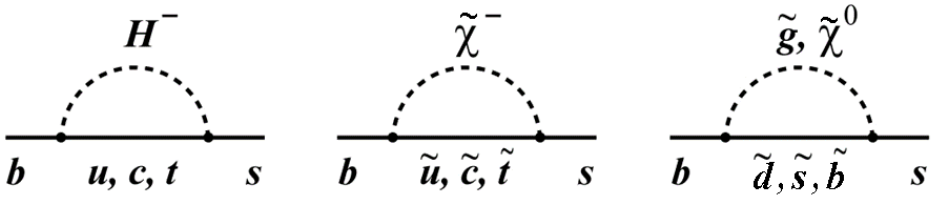

Figure 4: Feynman diagrams for $b \to s$ decay that have new physics particles in the loops. The emission of the Z boson from the loop is omitted for simplicity. The left diagram shows a charged Higgs, the middle shows a chargino and squarks, and the right shows a gluino or neutralino and squarks.

Another new physics possibility is scalar WIMP dark matter [28, 29]. Like neutrinos, stable DM particles would leave no tracks in a detector, so the decay of a B to a kaon and DM could give a signature similar to $B \to K\nu\overline{\nu}$. Thus, a branching fraction measurement for $B \to K\nu\overline{\nu}$ could include the DM decays. However, there could be significant differences between

Table 1: MSSM parameters used to calculate the branching fraction range in Equation 12. tanβ is the ratio of the vacuum expectation values of the two neutral Higgs fields. μ is the supersymmetry-conserving higgsino mass parameter. M_2 is the Wino mass. $m_{\tilde{Q}}$ and $m_{\tilde{U}}$ are the masses of the left- and right-handed up-type squarks, respectively. A_t is the trilinear coupling of the stop squark. $(\delta_u^{RL})_{32}$ is the helicity- and flavor-changing mass insertion. Dimension-full quantities are in GeV.

Parameter Set	tanβ	μ	M_2	$m_{\tilde{Q}}$	$m_{\tilde{U}}$	A_t	$(\delta_u^{RL})_{32}$
I	5	500	800	500	400	-800	0.75
II	5	120	700	400	800	-700	-0.5

the two decay processes, such as differences in the kaon angular or momentum distribution, that could greatly diminish any contribution from DM decays to a search sensitive only to specific features of $B \to K\nu\overline{\nu}$ decays.

Figure 5 shows Feynman diagrams for the DM decay. The matrix element for this decay is shown in Equation 13. With a sufficiently high value of the DM Wilson coefficient C_{DM}, based upon the parameter λ/m_h^2, the DM decay could could increase the branching fraction by up to two orders of magnitude. However, Figure 6 shows how branching fraction upper limits already constrain this decay and the allowed mass of the DM particles. These constraints are calculated by multiplying the upper limit by a ratio involving the phase-space integral shown in Equation 14. This ratio is $F(m_S)$ calculated with no experimental minimum kaon momentum over $F(m_S)$ calculated with a minimum kaon momentum determined by the experiment.

$$\mathcal{M}_{B\to KSS} = C_{DM} m_b \frac{m_B^2 - m_K^2}{m_b - m_s} f_0(q^2) \text{ with } C_{DM} = \frac{\lambda}{m_h^2} \frac{3G_F^2 V_{ts}^* V_{tb}}{64\pi^2} \frac{m_t^2}{m_W^2} \tag{13}$$

m_b is the b quark mass, and m_s is the s quark mass, f_0 is the form factor, λ is the scalar DM coupling constant, m_h is the Higgs boson mass, m_t is the top quark mass, and m_W is the W boson mass.

$$F(m_S) = \int_{\hat{s}_{min}}^{\hat{s}_{max}} f_0(\hat{s})^2 I(\hat{s}, m_S) d\hat{s} \left[\int_{\hat{s}_{min}}^{\hat{s}_{max}} f_0(\hat{s})^2 I(\hat{s}, 0) d\hat{s} \right]^{-1} \tag{14}$$

$$\text{with } \hat{s} = (p_B - p_K)^2 \text{ and}$$

$$I(\hat{s}, m_S) = \left[\hat{s}^2 - 2\hat{s}(m_B^2 + m_K^2) + (m_B^2 - m_K^2)^2\right]^{\frac{1}{2}} \left[1 - 4m_S^2/\hat{s}\right]^{\frac{1}{2}}$$

m_S is the mass of the scalar DM particle, p_B is the lab-frame momentum of the B, and p_K is the lab-frame momentum of the kaon. $\hat{s}_{min}$ is determined by m_S, while $\hat{s}_{max}$ can be set based

upon the lowest kinematically allowed kaon momentum or the minimum kaon momentum to which a particular experiment is sensitive. Note that $F(0) = 1$ and $F(m_S) = 1$ for $m_S > \frac{1}{2}(m_B - m_K)$.

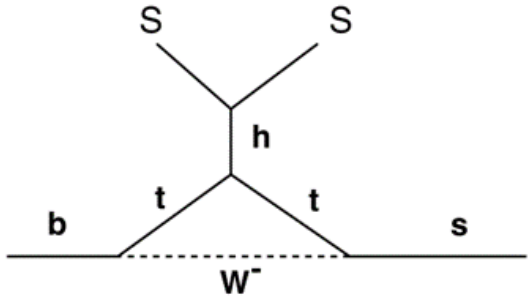

Figure 5: Feynman diagram for $b \to s$ decay that produces two scalar DM particles.

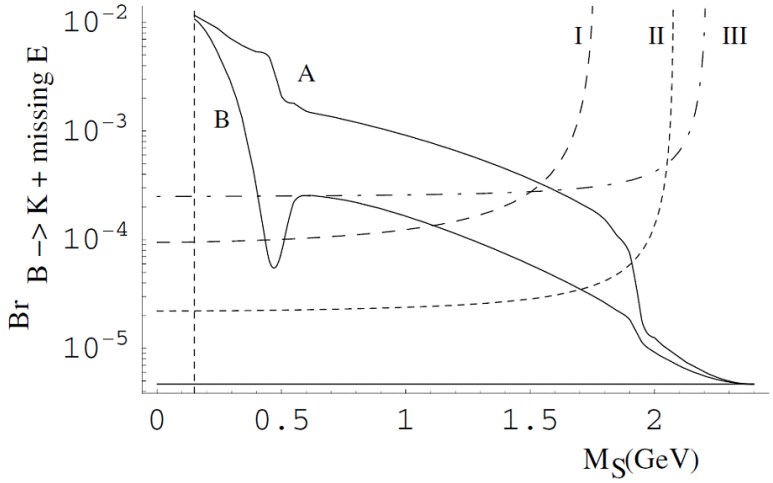

Figure 6: Branching fraction of $B \to K$ + *missing energy* vs. the mass of a scalar DM particle described by [28]. The model's allowed region is between curves A and B, which are determined by calculations of cosmic DM abundance. Curve I is the limit set by an early *BABAR* $B_u^+ \to K^+ \nu\overline{\nu}$ result [30] with 82 fb^{-1}, an upper limit of 7.0×10^{-5}, and a cutoff of $p_K > 1.5$ GeV/c. Curve II is an estimated limit projected for 250 fb^{-1} of *BABAR* data, with an upper limit of 2.1×10^{-5} and a cutoff of $p_K > 1.0$ GeV/c. Curve III is the limit set by a CLEO result [31], with an upper limit of 24×10^{-5} and a cutoff of $p_K > 0.7$ GeV/c.

Figure 7 shows the effect on the $B \to K\nu\overline{\nu}$ differential branching fraction of a generic scalar DM model with $m_S = 1.1$ GeV/c^2 [15]. The plot shows that the scalar contribution to the differential branching fraction only turns on when there is sufficient energy to produce the scalar particles; that is, at low kaon momentum.

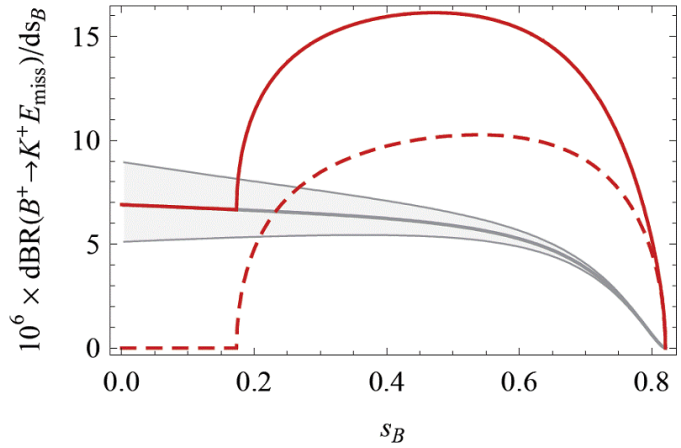

Figure 7: Differential branching fraction for $B \to K$ + *missing energy* [15]. s_B is the invariant mass squared of the undetected particles, either neutrinos or scalar DM, divided by the square of the B mass. The thick gray curve is the theoretical prediction for $B_u^+ \to K^+\nu\overline{\nu}$, and the grey shading around that curve indicates the theoretical uncertainty. The dotted red curve is $B^+ \to K^+SS$, and the solid red curve is the sum of the two.

Some other new physics models that could also affect this decay include the following. Unparticle models [32] and WIMP-less dark matter [33] would produce invisible unparticles or DM that could increase the apparent branching fraction of $B \to K\nu\overline{\nu}$. The current experimental upper limit of 1.4×10^{-5} from Belle already constrains these models, and stronger limits will further constrain their parameters. Models with a single universal extra dimension [34] change the Wilson coefficient and are currently constrained by the 1.4×10^{-5} Belle upper limit to exclude $1/R < 100$ GeV, where R is the radius of the compactified extra dimension.

Other decays similar to $B \to K\nu\overline{\nu}$ also can place limits on new physics but lack some

of the advantages of $B \to K\nu\overline{\nu}$. For example, $B \to K\ell^+\ell^-$ is very similar to $B \to K\nu\overline{\nu}$, 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 make it possible to fully reconstruct the event. In fact, the decay has been observed even though its branching fraction is ten times less than that of $B \to K\nu\overline{\nu}$, and the measurement is in agreement with the SM prediction:

$$\mathcal{B}(B \to K\ell^+\ell^-) = (0.394^{+0.073}_{-0.069} \pm 0.020) \times 10^{-6} \quad [35] \tag{15}$$

In comparison, $B \to K\nu\overline{\nu}$ 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 no resonances obscuring the main decay. New physics signals might show up more clearly in $B \to K\nu\overline{\nu}$ 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 and low-energy 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% [12]. 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 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 linear accelerator and the separate storage rings 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 2 shows the luminosity and number of $B\overline{B}$ pairs used for this analysis. In the search for a very rare decay like $B \to K\nu\overline{\nu}$, this high luminosity has been essential.

3.2 The *BABAR* detector

BABAR [13] has many design features important to this $B \to K\nu\overline{\nu}$ 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 8 and 9.

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 10, with a design intended to maximize the angular acceptance. Figure 11 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.

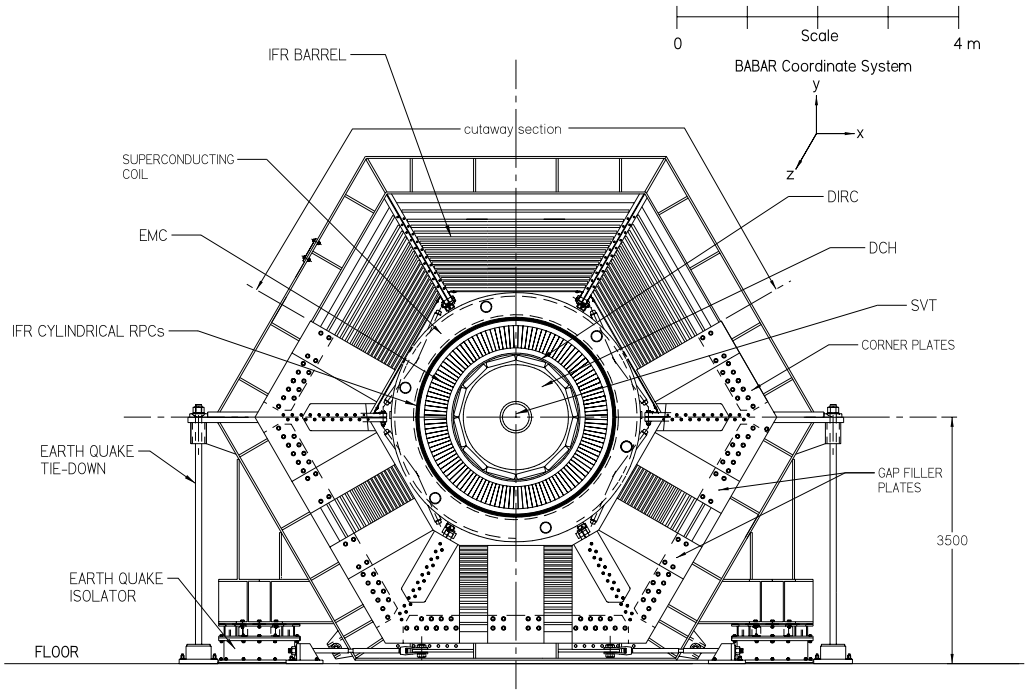

Figure 8: End view of the *BABAR* detector.

3.2.2 Drift Chamber

The drift chamber (DCH) measures particle momentum and energy loss (dE/dx). Figure 12 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 13 shows the inside of the DCH.

The track-finding efficiency of the DCH, averaged over momentum and angle, is about 96%, relative to the tracks found by the SVT. The highly effective tracking of particles

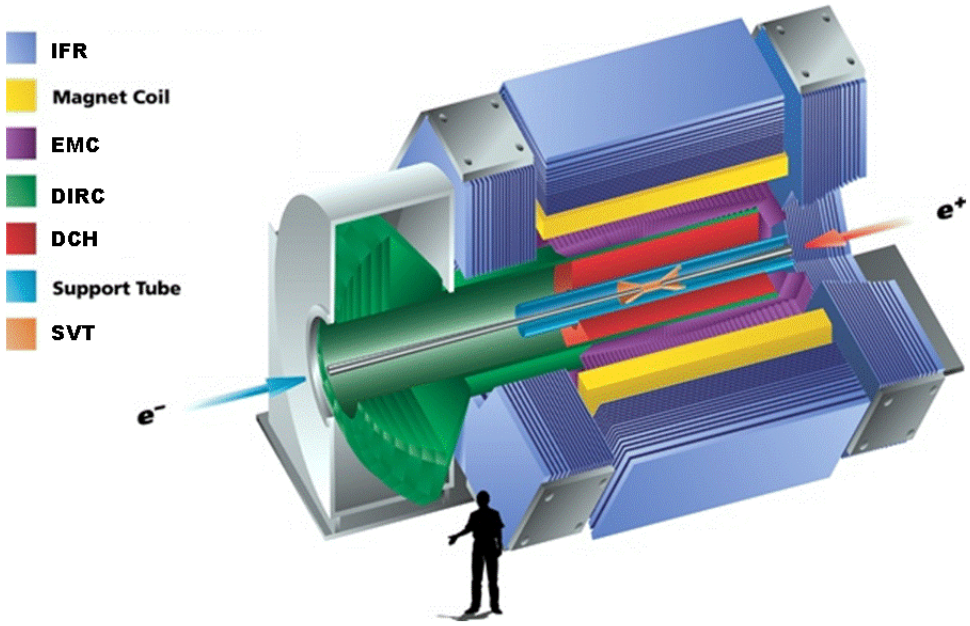

Figure 9: Layout of the *BABAR* detector.

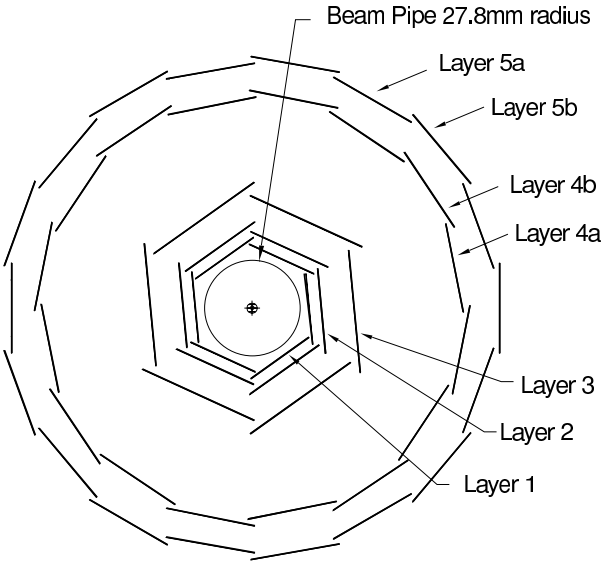

Figure 10: Layout of layers in the SVT.

provided by the SVT and DCH is essential to this analysis, which requires the reconstruction of B mesons from tracks in the events.

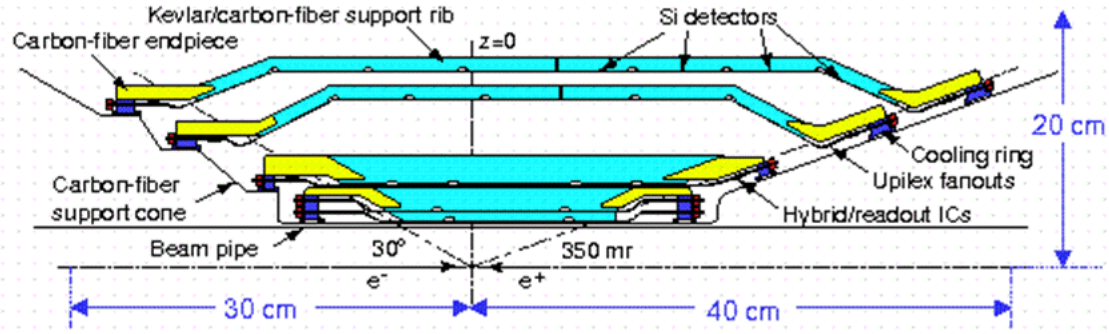

Figure 11: Side view of the SVT.

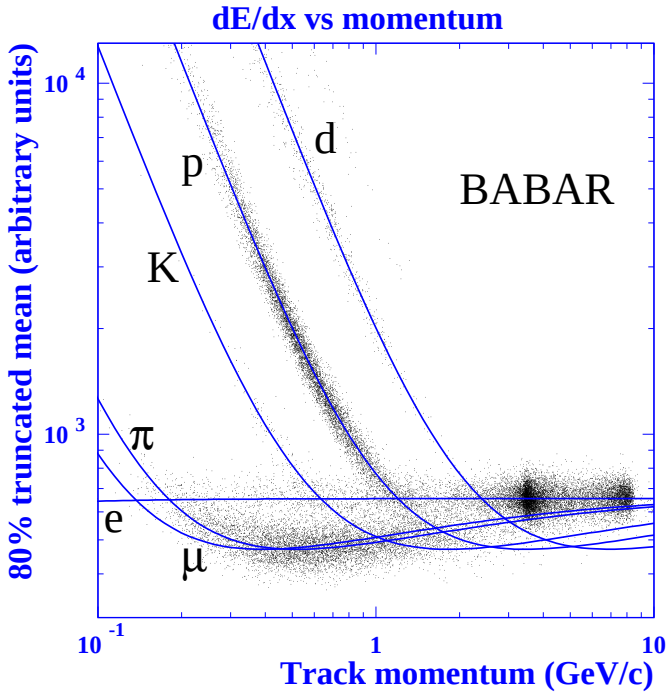

Figure 12: Energy loss in the DCH vs. track momentum.

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

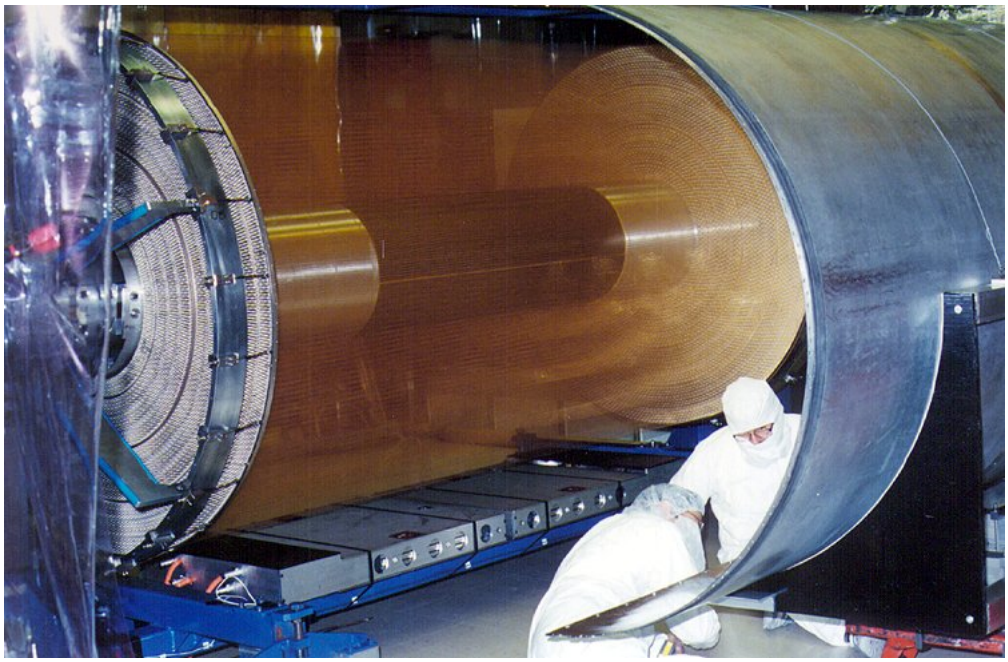

Figure 13: View of the DCH during assembly.

reflection in the quartz crystals of the detector sends the Cherenkov rings to photo-multiplier tubes, as shown in Figure 14. The Cherenkov angle θ_C, defined in Equation 16, is combined with the track momentum and angle to help identify the particle. These measurements are particularly effective at separating charged kaons from pions at lab momenta above 0.7 GeV/c, as illustrated in Figure 15. This momentum range is where kaons from $B \to K\nu\overline{\nu}$ are expected to be found, so the DIRC makes an important contribution to this analysis.

$$cos(\theta_C) = \frac{1}{\beta n} \tag{16}$$

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 16 shows the energy resolution of the EMC. Its channels are sensitive to energies down to

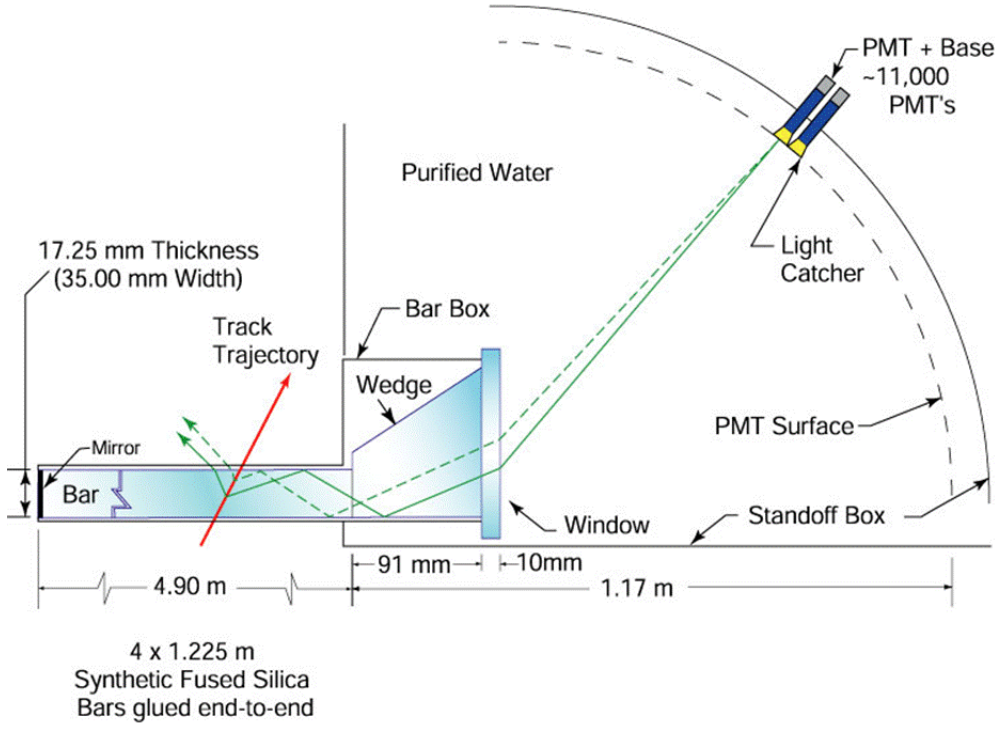

Figure 14: Geometry of the DIRC.

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 $B \to K\nu\overline{\nu}$ since the key characteristic of this decay is the absence of any remaining energy in the detector after all expected particles have been reconstructed.

The EMC is composed of 6580 thallium-doped cesium iodide crystals (see Figure 17) 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.

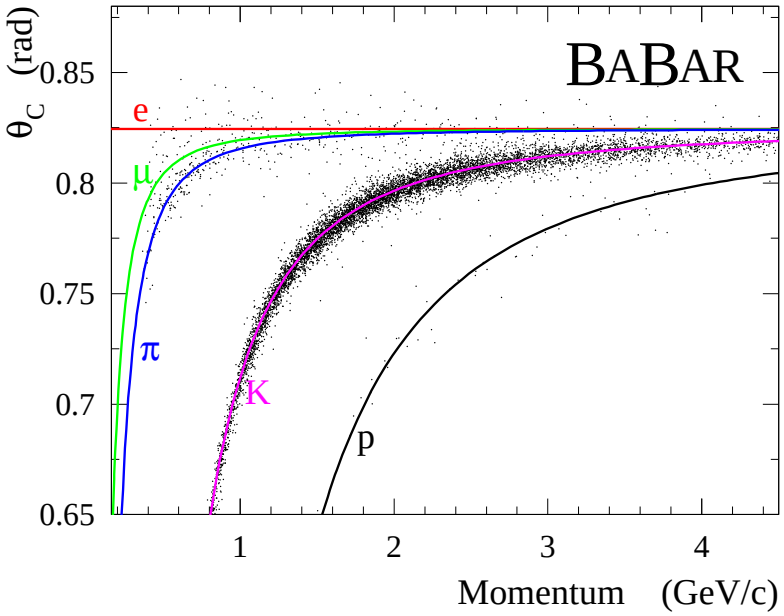

Figure 15: Cherenkov angle θ_C vs momentum for kaons (points). The solid lines correspond to the expected θ_C for other charged particle species.

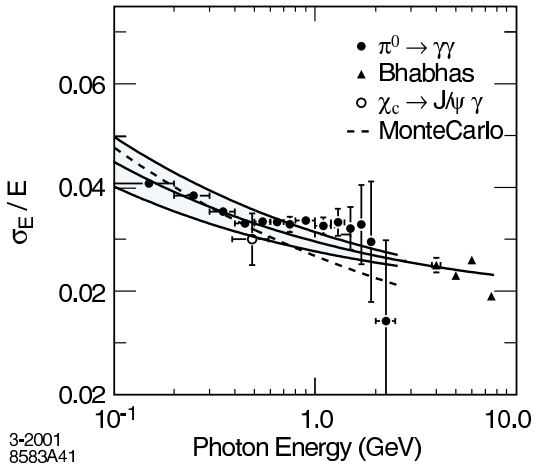

Figure 16: EMC energy resolution vs. photon energy.

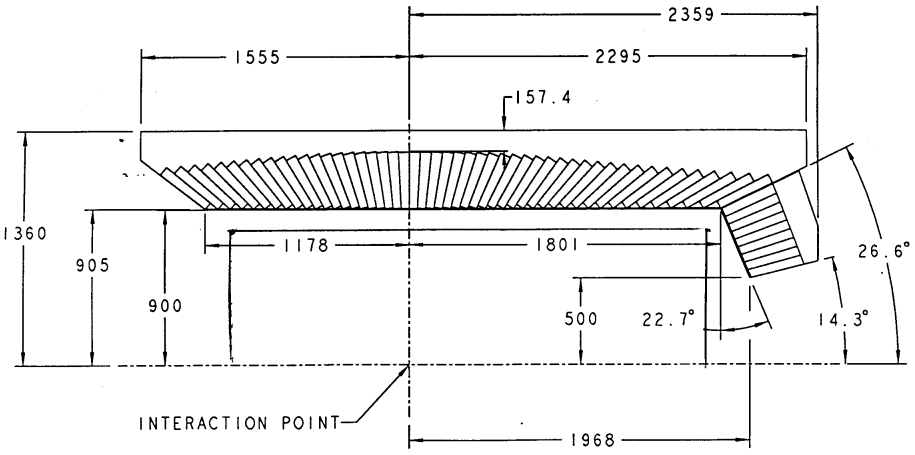

Figure 17: Side view of the crystal geometry in the electromagnetic calorimeter.

3.2.5 Instrumented Flux Return

The instrumented flux return (IFR), shown in Figure 18, 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, are used by muon selectors to recognize muon tracks. The performance of two *BABAR* muon selectors in shown in Figure 19.

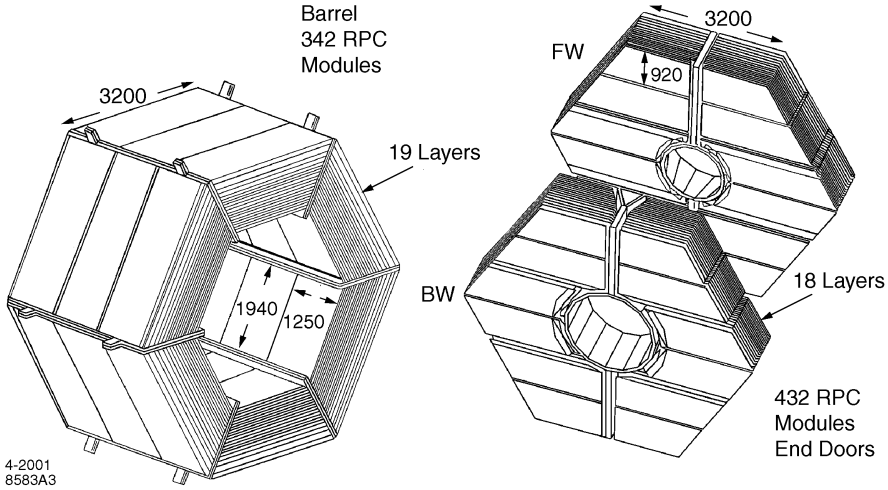

Figure 18: Geometry of the instrumented flux return.

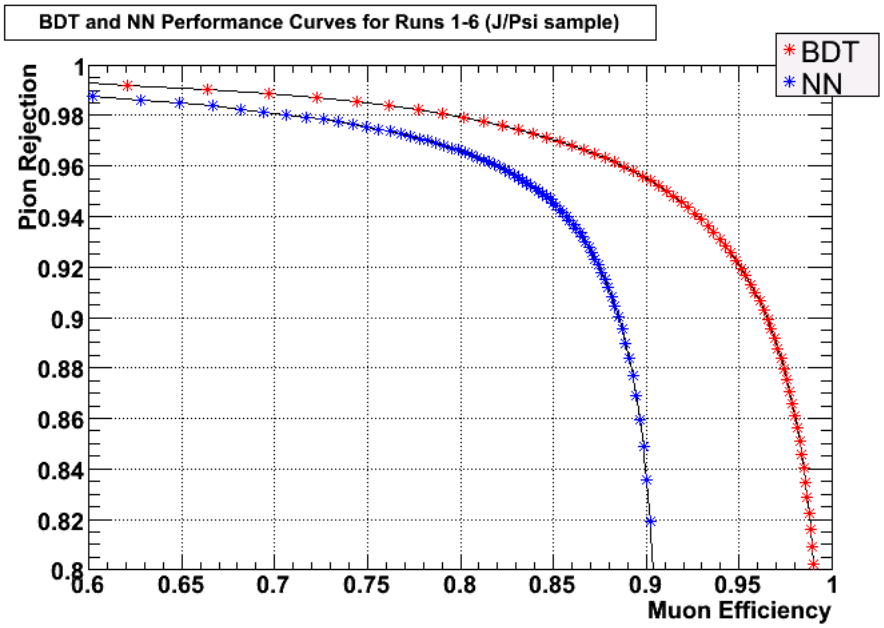

Figure 19: Performance of the muon selectors that use many IFR output variables. The right curve (red stars) is for the decision tree muon selector, and the left curve (blue stars) is for the older neural network selector. Their performance for muon identification and pion rejection is assessed using a control sample of J/Ψ events.

4 Analysis Overview

This analysis uses a blind analysis protocol to ensure that the final results are unbiased. This protocol requires that real data events that could include the desired signal be excluded from the analysis until after all analysis procedures have been developed and validated. The major steps of this analysis are as follows. This sequence is followed for both the $B_u^+ \to K^+\nu\overline{\nu}$ and $B_d^0 \to K_S^0\nu\overline{\nu}$ modes.

1. Two types of simulated events are produced: signal events and background events.

2. From these sets, those events are selected whose reconstruction makes them appear signal-like.

3. Decision trees are trained on these selected sets to select signal events and reject background.

4. A signal region is defined based on the decision tree output, and for this signal region, a background prediction with statistical and systematic uncertainties is made.

5. The decision trees are validated by verifying that their performance on real data and simulated events is the same, with any small discrepancies accounted for as systematic uncertainties.

6. Unblinding of the signal region can now occur since the analysis method has been finalized and validated. Real data events possibly containing the signal are classified by the decision trees.

7. The number of data events in the signal region is obtained and compared with the predicted background, and upper limits are calculated.

The following sections detail these steps.

5 Event Selection

5.1 Event Samples

5.1.1 *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 2).

Table 2: 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

5.1.2 Simulated Data Samples

The Monte Carlo (MC) simulated event samples (Tables 3-4) used here are from the SP8 production. The MC is generated by the EvtGen software package [36], and then it is processed through the GEANT4 detector simulation [37]. Six distinct classes of MC simulated events are relevant to this analysis:

- $B_u^+ \rightarrow K^+\nu\overline{\nu}$ 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.

Table 3: Number of MC simulated signal events and the ratio of the number of $B\overline{B}$ decays in data to the number simulated.

Mode (SP)	$\mathcal{B}(/10^{-6})$	Events	Data/MC $(/10^{-3})$
$K^+\nu\overline{\nu}(2227)$	3.8 [14]	7845000	0.22
$K^0_S\nu\overline{\nu}(3152)$	1.9 [14]	3943000	0.22
$K^+J/\Psi(989)$	120	13431000	0.41

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

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

- Generic continuum events (Table 4). The samples of generic continuum events are scaled to the appropriate cross sections.

- Generic $B\bar{B}$ events (Table 4), 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 7 for details.

5.2 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.

Table 4: 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

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. 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 [38, 39], 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 `BToDlnu` skim, is subjected to two steps of refinement, described in Sections 5.3 and 6, 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 $\pm 60\,\mathrm{MeV}/c^2$ of the D^+ mass

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

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

To be selected by the skim, an event must have a lepton with a CMS momentum greater than 0.8 GeV/c.

5.3 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. The photons from the decay have the following requirements:

- $LAT = [0.0, 0.8]$ (the LAT is the lateral moment of the photon shower). See Figure 160 in Appendix A.6 for a plot of the LAT distribution of a photon control sample.

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

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

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

- D^* mass within $0.5\,\mathrm{GeV}/c^2$ 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 17.

$$cos\theta_{BY} = \frac{2E_{beam} \cdot E_{Dl} - m_B^2 - m_{Dl}^2}{2\,|\mathbf{p_{Dl}}\,| \cdot \sqrt{E_{beam}^2 - m_B^2}} \tag{17}$$

  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 CMS).

- $B \to D^*l\nu$ tag candidates have two requirements, calculated in the CMS:

  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.

5.3.1 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 5.3.2.

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 $K^{*+} \to K^0_S\pi^+$decay, with a reconstructed K^* mass within 0.075 GeV/c^2 of the canonical K^* mass. The K^0_S 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 $K^{*0} \to K^0_S\pi^0$ decay, with the same restrictions as item 2 above.

4. $B^0 \to K^0_S\nu\overline{\nu}$ signal mode if the K^0_S is reconstructed in the $K^0_S \to \pi^+\pi^-$ mode, with the same restrictions for the K^0_S 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 $K^{*+} \to K^0_S\pi^+, K^0_S \to \pi^0\pi^0$ decay.

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

(d) $B^+ \to K^+\nu\overline{\nu}$ 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^0_S signal events and the other events listed above. The cuts described in Section 5.3.2 select only those events reconstructed in the desired signal mode. Table 5 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 6.1), 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 5: Percentages of skimmed signal MC that is reconstructed in the modes described in Section 5.3.1, 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

5.3.2 Preselection Requirements 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 6.1. Two sets are ntuples are generated: one for the K^+ mode, and one for the K^0_S 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^0_S ntuples are:

- The signal B must have only one daughter, which must be a K^0_S.
- No more than one extra track is allowed.

The effect of these cuts on efficiency is shown in Tables 6 and 7.

Table 6: 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 18. See Section 6.1 for explanation of the BDT cut.

Step	K^+ **Efficiency**	K^+ **Significance**	K^+ **Background Efficiency**
BToDlnu Skim	0.0694±0.0001	0.01	$(9.28\pm0.00)\cdot10^{-2}$
Signal/Tag Reconstruction	0.0148±0.0000	0.007	$(5.92\pm0.00)\cdot10^{-3}$
Preliminary Cuts	0.00672±0.00003	0.05	$(2.11\pm0.01)\cdot10^{-5}$
Primary BDT Cut	0.00175±0.00001	0.7	$(7\pm2)\cdot10^{-9}$

Table 7: Efficiency and signal significance for K^0_S 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 18. See Section 6.1 for explanation of the BDT cut.

Step	K^0_S **Efficiency**	K^0_S **Significance**	K^0_S **Background Efficiency**
BToDlnu Skim	0.0578±0.0001	0.003	$(9.33\pm0.00)\cdot10^{-2}$
Signal/Tag Reconstruction	0.0117±0.0001	0.003	$(5.96\pm0.00)\cdot10^{-3}$
Preliminary Cuts	0.00333±0.00003	0.01	$(4.21\pm0.01)\cdot10^{-5}$
BDT Cut	0.00060±0.00001	0.3	$(1.6\pm0.8)\cdot10^{-9}$

6 Decision Trees for Signal/Background Separation

6.1 Bagger Decision Trees

A StatPatternRecognition (SPR) [40] Bagger Decision Tree (BDT) [41] 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 the 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 7). 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 ways 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 9 to 11, with Table 8 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^0_S tree uses 38. Tables 12 and 13 show rankings of the classification power of the variables for the primary K^+ and K^0_S 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^0_S 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 20 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 and (ii) 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 17)
- 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^0_S mass, CMS energy;
- polar angle of missing CMS momentum using just tag B and signal K candidates;
- 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 $Dl\nu$ 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 8: 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 0^{th}
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 $\mathrm{m}(D^0 + \mathrm{photon}) - \mathrm{m}(D^0)$ 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 tag D
tagdndaus	number of daughters of the tag D
tagdp3cm	center-of-mass three-momentum of the tag D
taglp3	lab three-momentum of the tag lepton
taglp3cm	center-of-mass three-momentum of the 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 lab momentum of the Υ(4S)
yp3	lab momentum of the Υ(4S)

Table 9: Classification variables for the primary $K^+\nu\overline{\nu}$ 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 10: Classification variables for the $K^0_S\nu\overline{\nu}$ tree. Variables shared with the primary $K^+\nu\overline{\nu}$ 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 11: 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 92. 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 [15]. The exclusion of these low-momentum events is the result of the predominance of background in this momentum region (as seen in Figure 58) 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 [15] (see Section 6.2.4) 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.

Table 12: 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
sigbcosththrustdl	102.1	3537
EMiss	95.0	3430
sigbcosththrust	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 13: Signal/background discrimination effectiveness (as reported by SPR) of classification variables for the $K^0_S\nu\overline{\nu}$ 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	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
sigbcostthrustdl	66.0	1642
taglp3	57.9	1693
sigbcostthrust	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

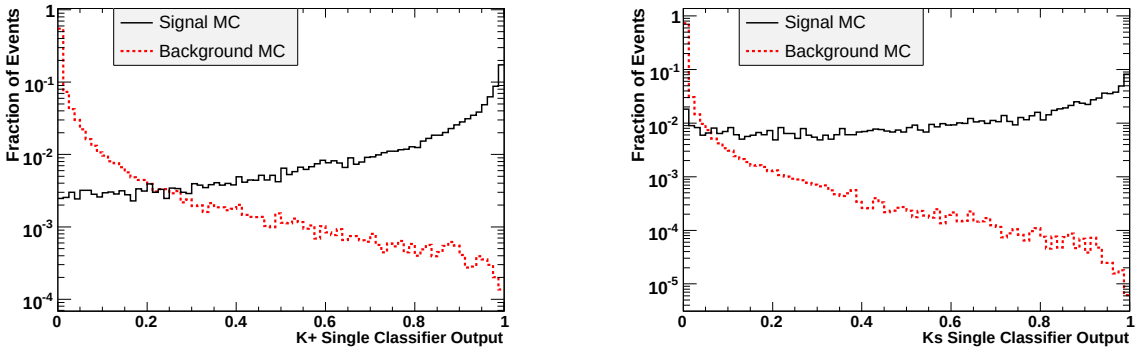

Figure 20: Histograms showing the classifier output for MC signal and background events in the testing set. The left plot is for the primary K^+ tree. The right plot is for the K^0_S tree. Each plot is from one tree that is representative of its twenty-tree set.

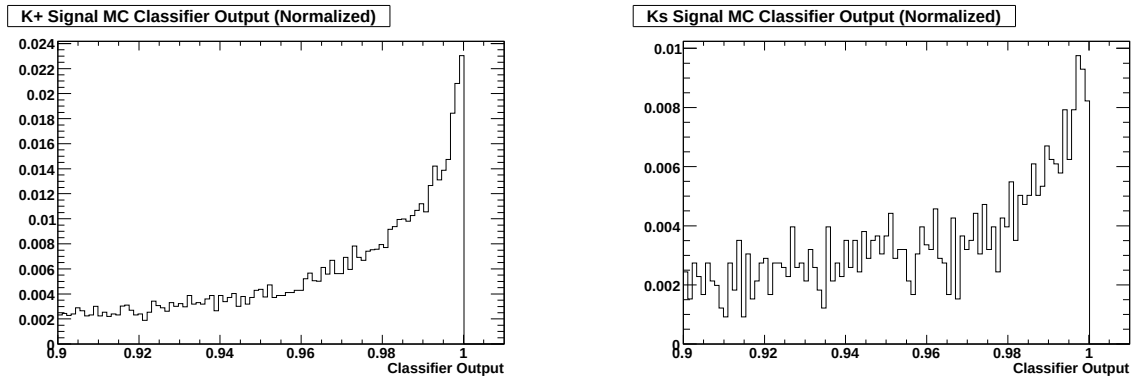

Figure 21: Histograms showing the classifier output for MC events in the testing set. Background events are too few to show up on these plots. The left plot is for the primary K^+ tree and shows only signal MC events at the high end of the classifier output. See Figure 20 with the log scale to see the 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^0_S tree at the high end of the classifier output. See Figure 20 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.

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

$$\frac{s}{\sqrt{s+b}} \tag{18}$$

$$\frac{s}{N_\sigma/2+\sqrt{b}} \tag{19}$$

$$\frac{\epsilon_{sig}}{0.6425+\sqrt{b}} \tag{20}$$

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 18 is a rough measure of signal significance. Equation 19 is the Punzi figure of merit [42], with N_σ set to 3. Equation 20 is another version of Equation 19, 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 22 to 24, 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 6 and 7 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 data events in the signal region. Along with the predicted number background events, this value allows calculation of a limit on the branching fraction.

6.2 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

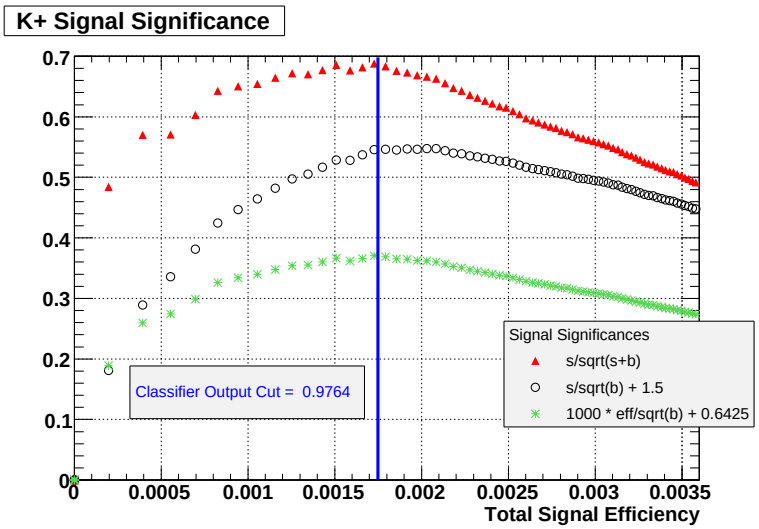

Figure 22: Measures of signal significance for the primary K^+ trees, averaged over the twenty trees. The red triangles correspond to equation 18, the black circles correspond to equation 19, and the green asterisks correspond to equation 20. The blue line shows the position of a cut that would approximately 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. This plot is based upon the phase-space model.

trees show good agreement (see Section 6.2.3), but the K^+ trees produce large discrepancies. To fix this problem, a weighting is applied to the MC, as explained in the next section.

6.2.1 Monte Carlo Tuning for the K^+ Mode

The data and generic background MC K^+ events that result from the preliminary cuts described in Section 5.3.2 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 12 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

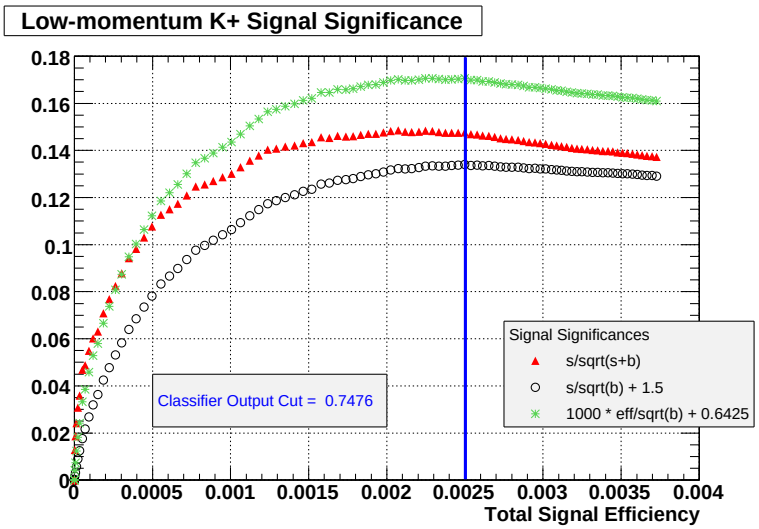

Figure 23: Measures of signal significance for the low-momentum K^+ measurement, averaged over the twenty trees. The red triangles correspond to equation 18, the black circles correspond to equation 19, and the green asterisks correspond to equation 20. The blue line shows the position of a cut that 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. This plot is based upon the ABSW model [15], since there are more events in the low-momentum region for that model than in the other two models.

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 25 shows the histogram with the calculated weights. These weights, which are referred to as 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 26 and 27), which provides confidence that the weighted background MC accurately predicts the background in the signal region.

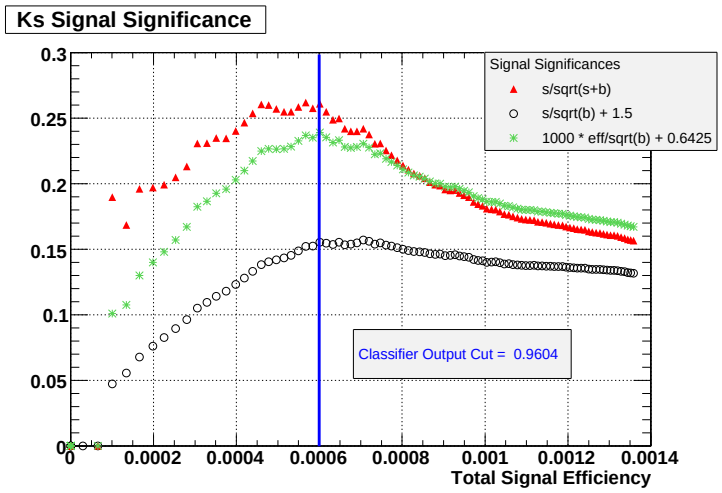

Figure 24: Measures of signal significance for the K^0_S trees, averaged over the twenty trees. The red triangles correspond to equation 18, the black circles correspond to equation 19, and the green asterisks correspond to equation 20. The blue line shows the position of a cut that would approximately 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. This plot is based upon the phase-space model.

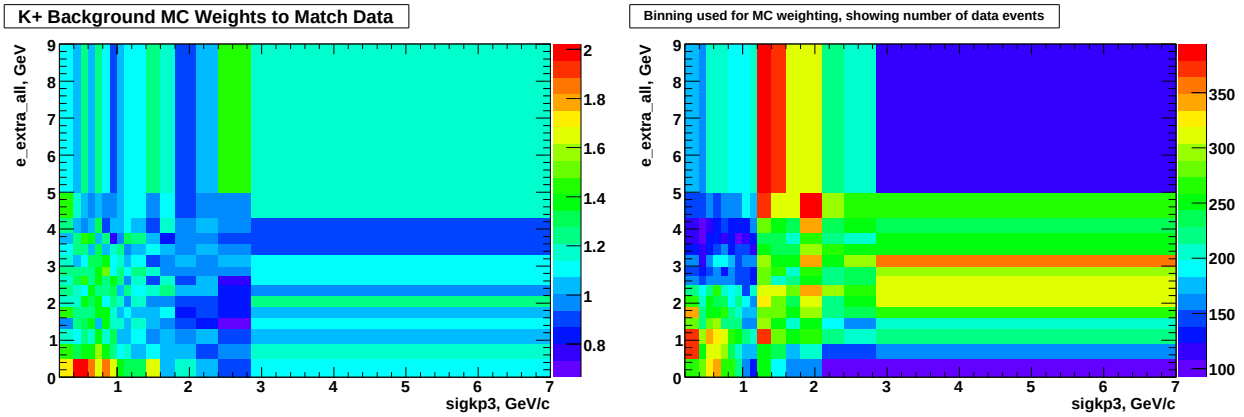

Figure 25: Two-dimensional weighting histogram for the primary K^+ mode, showing the values of the weights assigned to generic background MC events that fall in each bin on the left, and the number of data events in each bin on the right.

6.2.2 Data/MC Agreement for the Low-momentum Measurement

The low-momentum K^+ measurement also utilizes the 2-D weighting. Figure 28 shows good data/MC agreement for this measurement, though there is a slight -5% data/MC

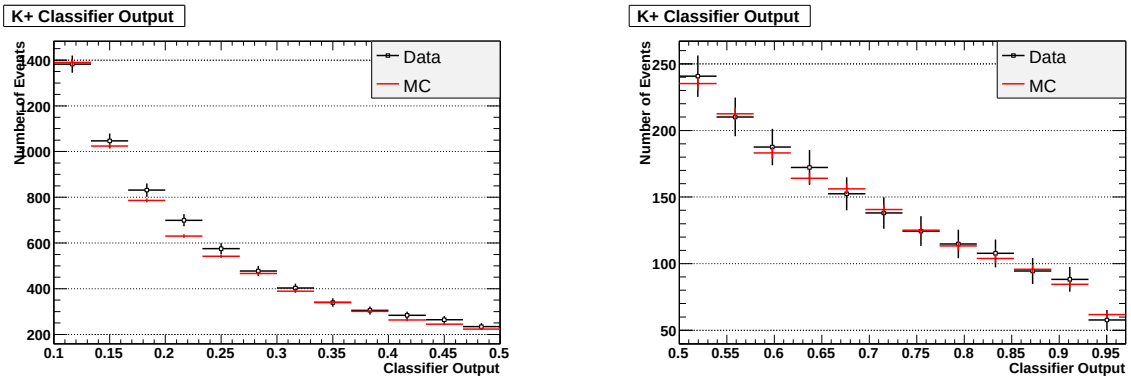

Figure 26: Classifier output for the primary K^+ trees for data and generic background MC events with the 2-D weighting. The left plot shows the range 0.1 to 0.5, and the right shows 0.5 to 0.975, the edge of the signal region. MC matches data very closely.

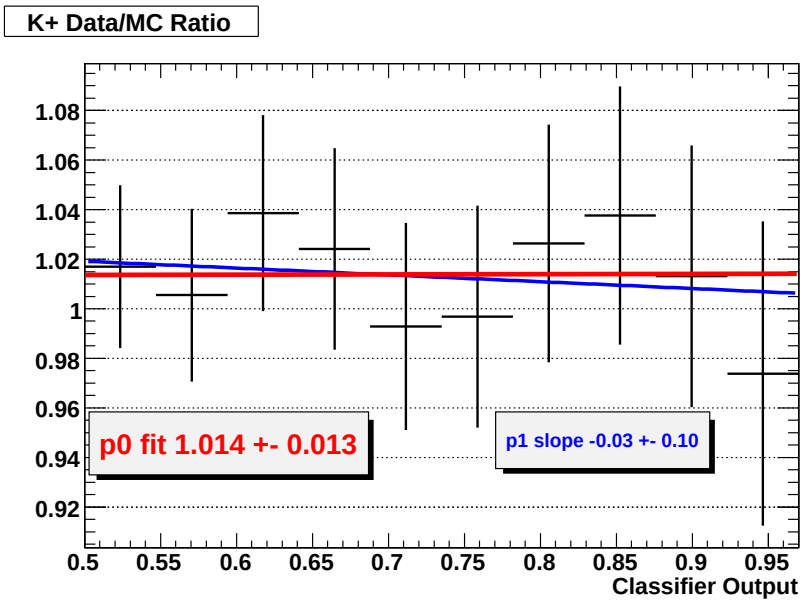

Figure 27: Data/MC ratio with the 2-D weighting applied for the primary K^+ classifier output just below the signal region. Agreement between data and MC is very good.

discrepancy, which is used as a correction to the background prediction.

6.2.3 Data/MC Agreement for the K_S Mode

As shown in Figures 29 and 30, 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

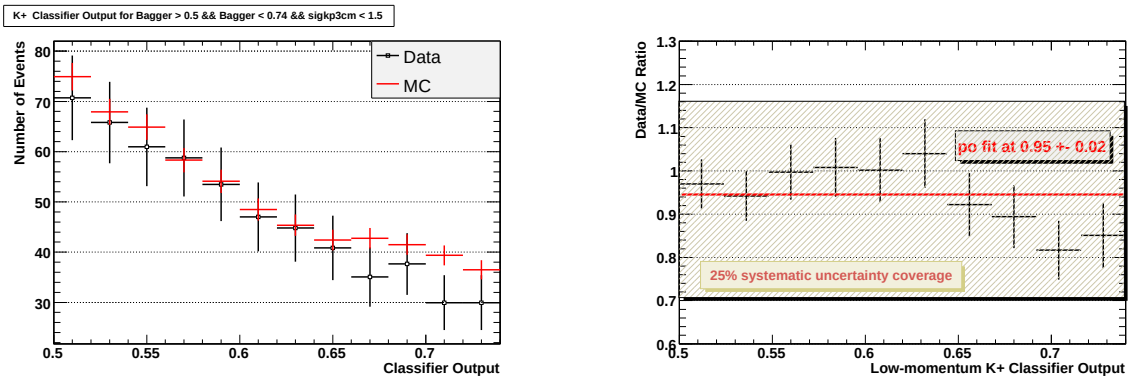

Figure 28: Classifier output for the low-momentum K^+ measurement with the 2-D weighting. The left plot shows data and background MC up to the edge of the signal region. The right plot shows the data/MC ratio. The fitted line gives the value of the ratio at 0.95 ± 0.02, so a -5% correction needs to be applied to the background prediction for this measurement. The systematic uncertainty applied to this measurement (see Section 8.2.2) is shown to demonstrate that it covers the fluctuations of the data/MC ratio.

MC seen in Figure 30 is applied to the K_S background prediction for the signal region. As discussed in Section 8.2, this value is also used as the systematic uncertainty to the background prediction.

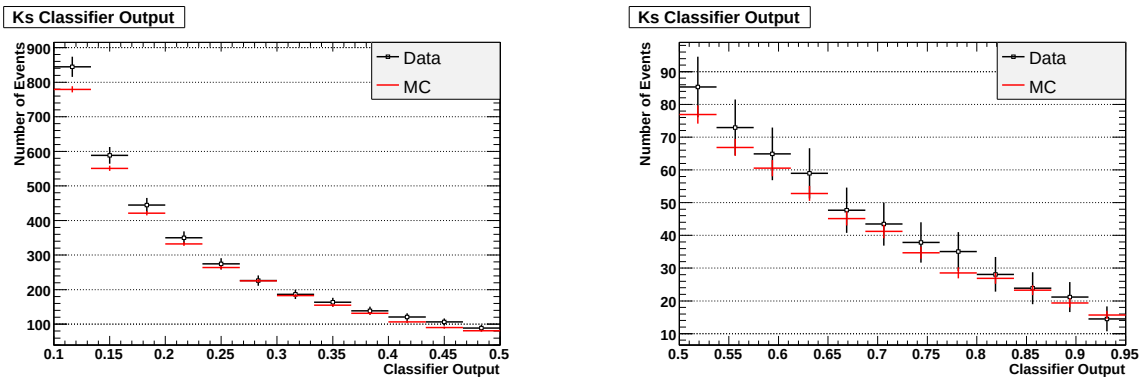

Figure 29: Classifier output for the primary K_S trees for data and generic background MC events. The left plot shows the range 0.1 to 0.5, and the right shows 0.5 to 0.95, the edge of the signal region. MC matches data very closely.

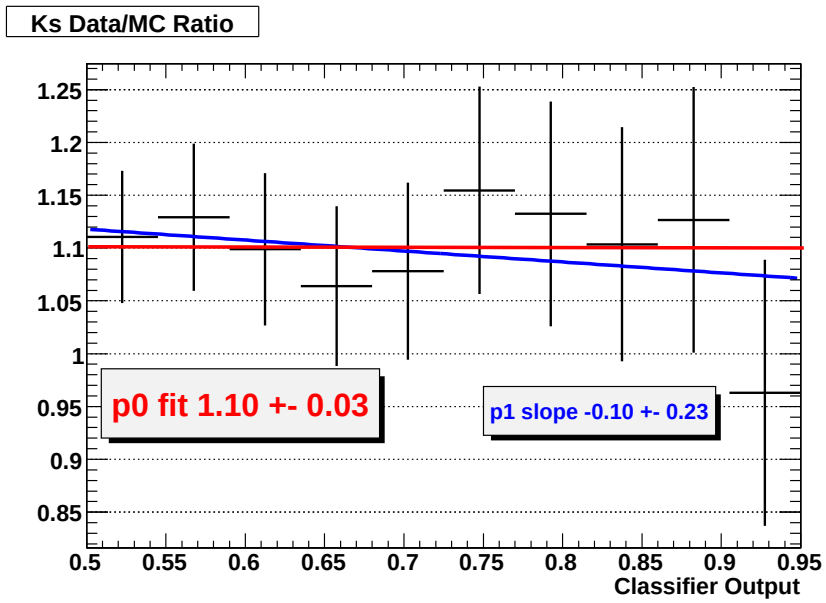

Figure 30: Data/MC ratio for the K_S classifier output just below the signal region. The small +10% data/MC discrepancy is used as the correction to the background estimate and as the background systematic uncertainty.

6.2.4 Signal Models and Weighting

This analysis uses three models for signal MC: phase space, BHI(2001) [14], and ABSW(2009) [15]. 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 the di-neutrino invariant mass squared spectrum for each.

Figure 31 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 14 shows the signal efficiencies for each model. Section 9 gives the measured upper limits for each model.

The signal efficiencies for the high-momentum K^+ measurement are calculated by only considering events where sigkp3cm > 1.5 GeV/c. The high-momentum classifier output cuts are the same ones used for the primary K^+ measurement, which means the background predictions are identical for the two measurements. The only difference between the two is the signal efficiency.

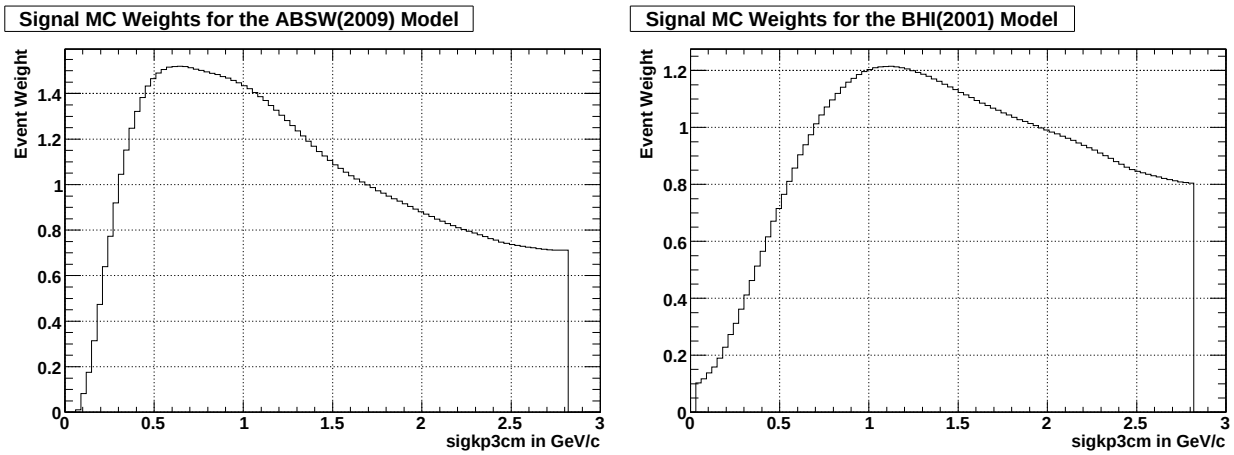

Figure 31: Weights for signal MC to convert the phase-space model to the two theoretical models.

Table 14: Total signal efficiencies in percent under the three signal models. The listed uncertainties are statistical only.

	Phase-space Model	**BHI Model**	**ABSW Model**
K^+	0.175 ± 0.001	0.161 ± 0.001	0.141 ± 0.001
K_S	0.060 ± 0.001	0.055 ± 0.001	0.048 ± 0.001
High-momentum K^+	0.251 ± 0.001	0.241 ± 0.001	0.236 ± 0.001
Low-momentum K^+	0.260 ± 0.001	0.278 ± 0.001	0.250 ± 0.001

7 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. Three real data samples are used for this validation as described below.

7.1 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 32, 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 33 shows the corresponding plot for the K_S trees, in this case using tag B^+ events. Figures 34 and 35 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 36 and 37 show the data/MC difference plotted with bins covering various ranges

of classifier output values. Figures 38 and 39 are similar but show the signed differences rather than absolute values.

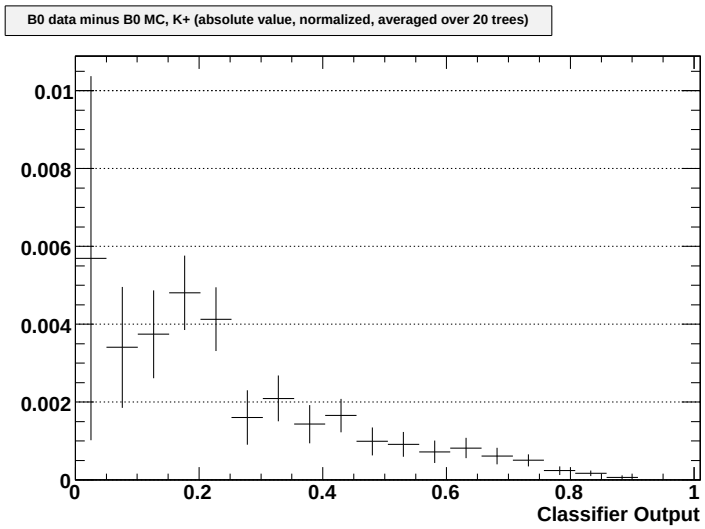

Figure 32: Normalized histogram of the difference in the classifier output for B^0 data and B^0 MC for the primary K^+ trees. The plot shows the absolute values of the normalized differences, averaged over the twenty trees. The data/MC discrepancy in each bin is always less than 0.6% of the total number of events. See text for more details.

7.2 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 5.3.2 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 $B \to \bar{D}^{*0}a_1^+$ (branching fraction $(1.9 \pm 0.5)\%$ [12]).

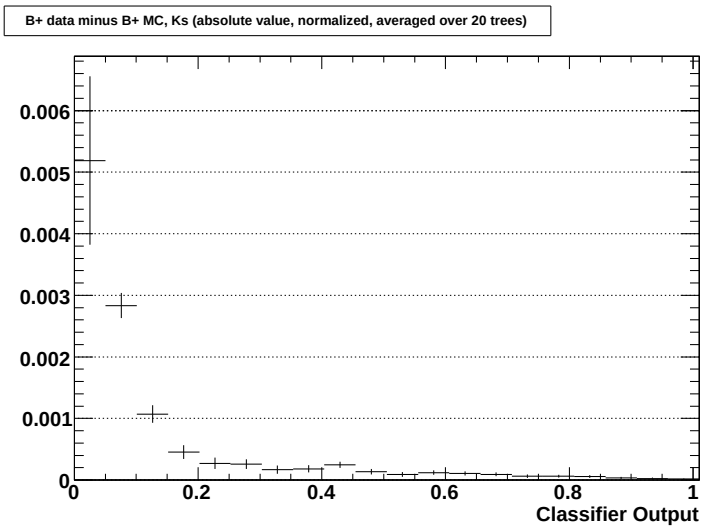

Figure 33: Normalized histogram of the difference in the classifier output for B^+ data and B^+ MC for the K_S trees. The plot shows the absolute values of the normalized differences, averaged over the twenty trees. The data/MC discrepancy in each bin is always less than 0.6% of the total number of events.

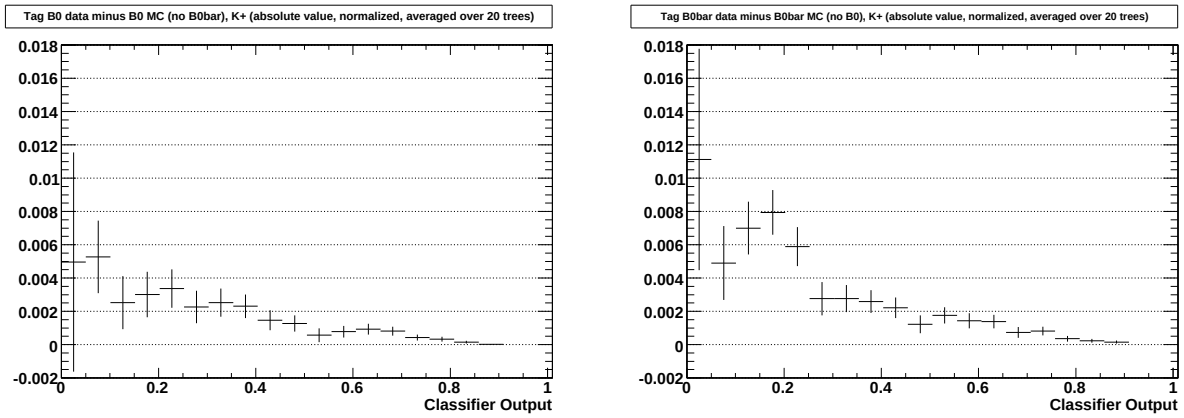

Figure 34: Normalized histograms of the difference in the classifier output for neutral B data and neutral B MC for the primary K^+ trees. The plots show the absolute values of the normalized differences, averaged over the twenty trees. The left plot is for B^0 only (no $\bar{B}^0$), while the right 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 40 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

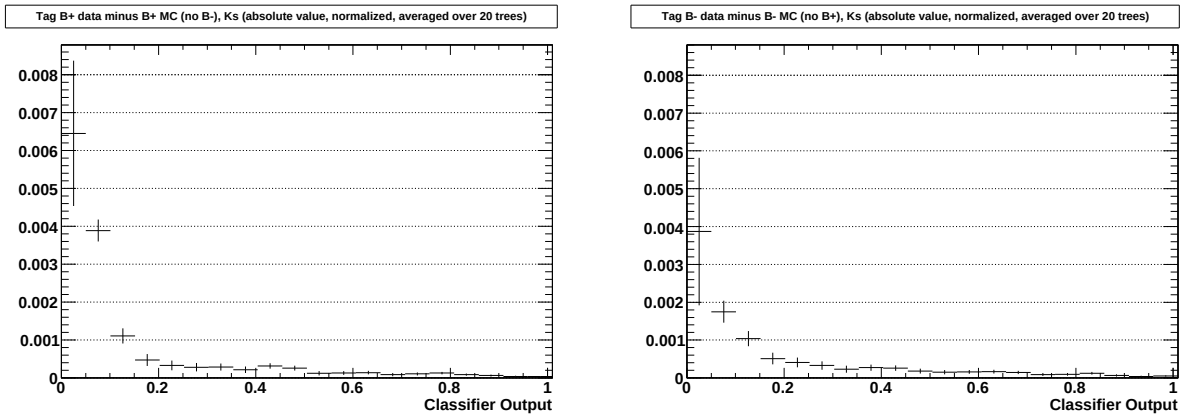

Figure 35: Normalized histograms of the difference in the classifier output for charged B data and charged B MC for the K_S trees. The plots show the absolute values of the normalized differences, averaged over the twenty trees. The left plot is for B^+ only (no B^-), while the right plot is for B^- only. The data/MC discrepancy in each bin is always less than 0.7% of the total number of events.

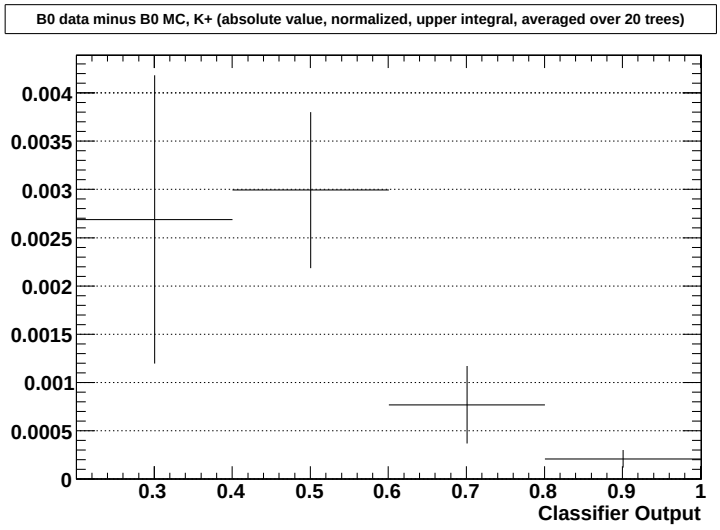

Figure 36: Plot of the difference in the integrated number of events for different classifier output values for B^0 data and B^0 MC for the K^+ trees. Each bin of the plot shows the average normalized absolute value of the difference in the number of events between B^0 data and B^0 MC for a range of 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.

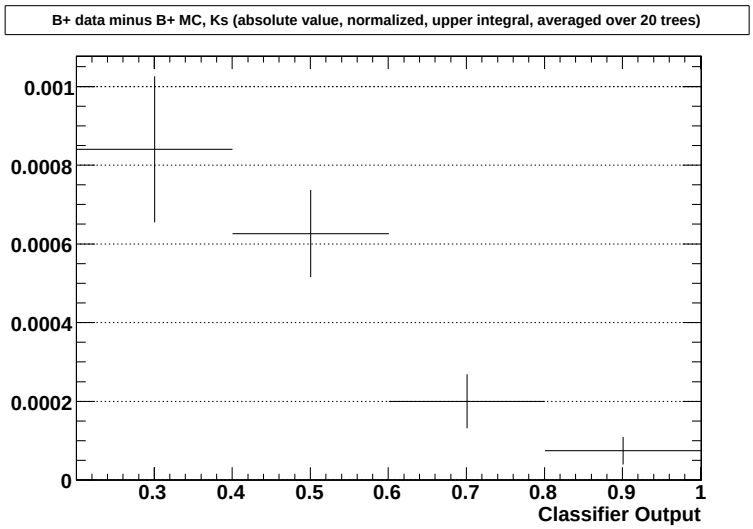

Figure 37: Plot of the difference in the integrated number of events for different classifier output values for B^+ data and B^+ MC for the K_S trees. Each bin of the plot shows the average normalized absolute value of the difference in the number of events between B^+ data and B^+ MC for a range of 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 6.2.1). Figure 41 shows the similar ratio for the K_S trees.

7.3 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 KJ/ψ 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

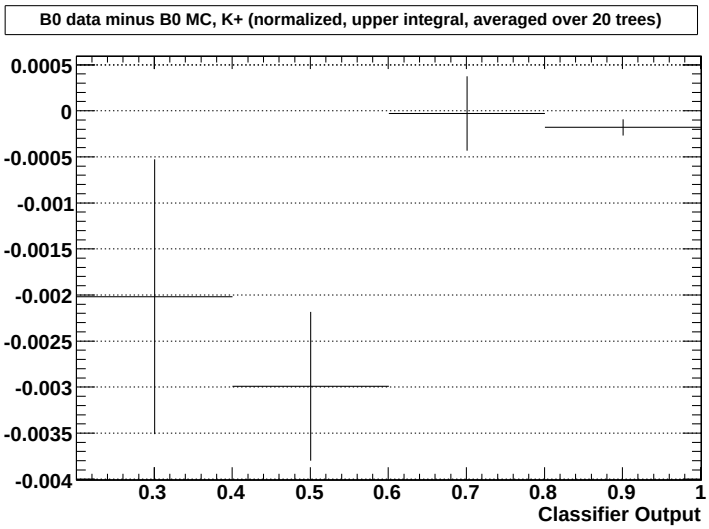

Figure 38: Plot of the difference in the integrated number of events for different classifier output values for B^0 data and B^0 MC for the K^+ trees. Each bin of the plot shows the average normalized signed value of the difference in the number of events between B^0 data and B^0 MC for a range of classifier 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 21) must be in the following range: $-0.04 < \Delta E < 0.035\,\mathrm{GeV}$.
- m_{ES} (Equation 22) must be in the following range: $5.2725 < m_{ES} < 5.285\,\mathrm{GeV}$.
- m_{ll} (Equation 23) must be in the following range: $3.06 < m_{ll} < 3.12\,\mathrm{GeV}$.

$$\Delta E = E_{SigB} - E_{beam}/2 \tag{21}$$

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.

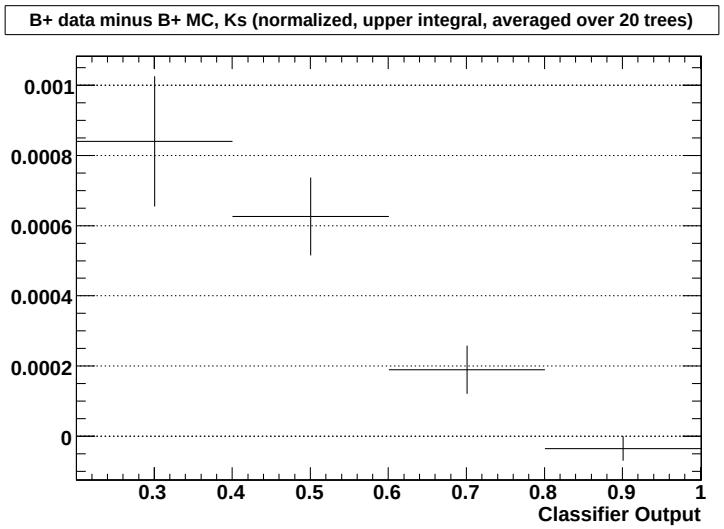

Figure 39: Plot of the difference in the integrated number of events for different classifier output values for B^+ data and B^+ MC for the K_S trees. Each bin of the plot shows the average normalized signed value of the difference in the number of events between B^+ data and B^+ MC for a range of classifier 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.

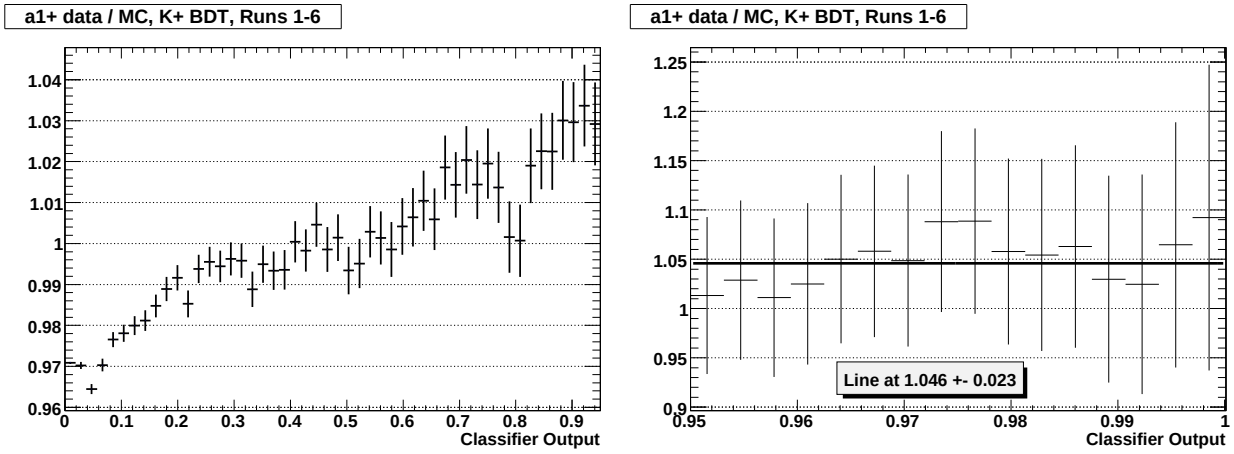

Figure 40: Data/MC ratio of the classifier output for a_1^+ events classified by the primary K^+ trees. The left plot show the full range of values up to the signal region, while the right plot shows the approximate signal region. On the right, the fitted line gives the value of the ratio as 1.05 ± 0.02.

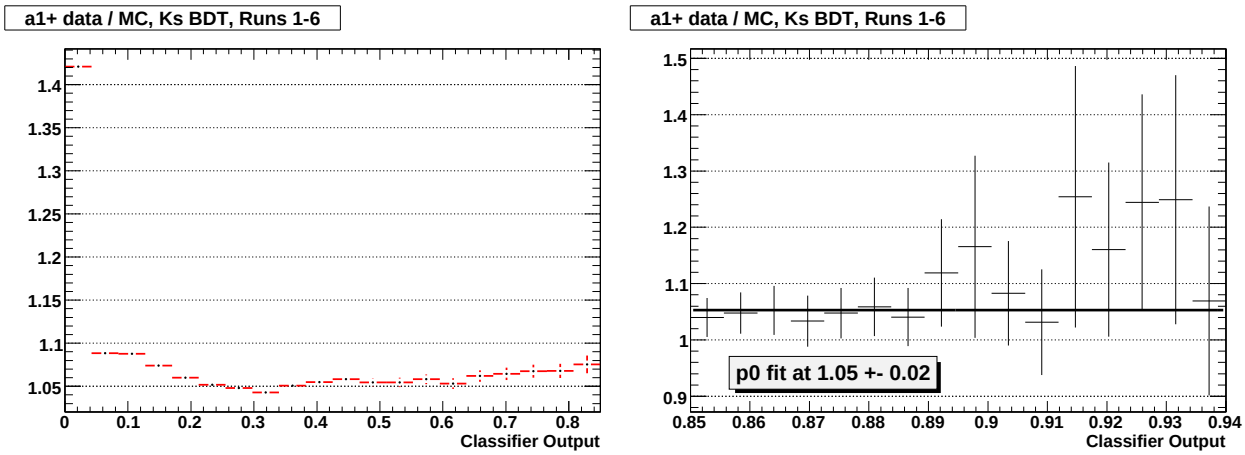

Figure 41: Data/MC ratio of the classifier output for a_1^+ events classified by the K_S trees. The left plot show the a range of values from zero to near the signal region, while the right plot shows the vicinity of the signal region. Almost no data events actually reach the signal region. On the right, the fitted line gives the value of the ratio as 1.05 ± 0.02.

$$m_{ES} = \sqrt{\frac{E_{beam}^2}{4} - p_{SigB}^2} \tag{22}$$

where p_{SigB} is the three-momentum of the signal B and E_{beam} is the total beam energy, both in the CMS.

$$m_{ll} = \sqrt{(p_{4l1} + p_{4l2})^2} \tag{23}$$

where p_{4l1} and p_{4l2} are the CMS 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 42 .

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 indi-

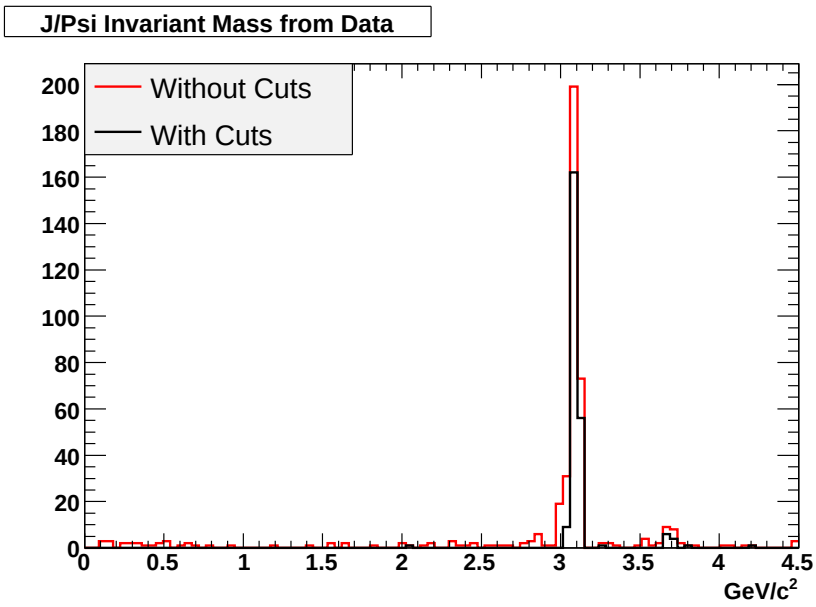

Figure 42: m_{ll} distribution for J/Ψ data events before and after the ΔE and m_{ES} cuts described in Section 7.3 are applied. All the other cuts in that section, except the m_{ll} cuts, are already applied to these events.

cates very little discrepancy between data and MC events in skimming and signal and tag reconstruction. Figure 43 shows the close agreement of data and MC.

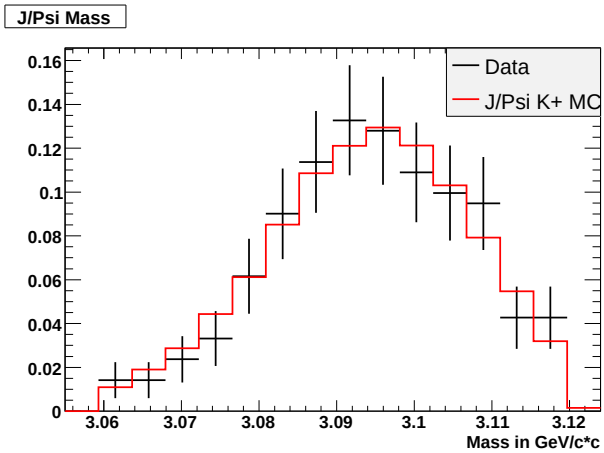

Figure 43: m_{ll} distribution for J/Ψ data and MC events after the cuts described in Section 7.3 are applied. Data and MC show close agreement.

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 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 11. The subset tree should reflect the behavior of the K^+ and K_S trees, and it allows study of J/Ψ data events in the same region where $K^+\nu\overline{\nu}$ or $K_S\nu\overline{\nu}$ 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 44. 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 systematic uncertainty to the signal efficiency.

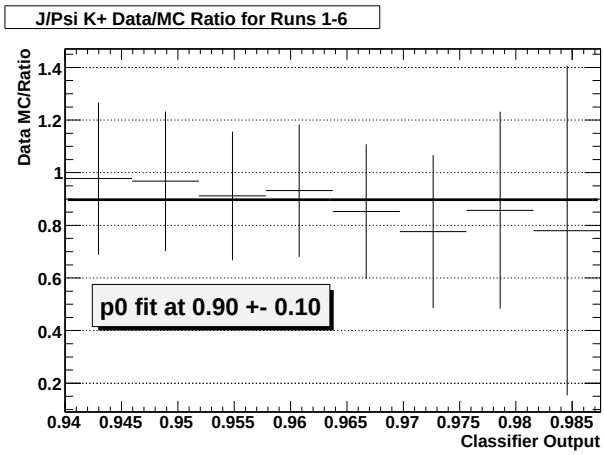

Figure 44: Data/MC ratio for J/Ψ events classified by the subset trees. The fitted line gives the value of the ratio as 0.90 ± 0.10. The plot covers only the approximate signal region in order to get a conservative estimate of the systematic uncertainty. The data/MC agreement is greater outside of the 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.

8 Systematic Uncertainties

8.1 Signal Efficiency Systematic Uncertainty

As discussed in Section 7.3, 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%.

8.1.1 Theory Systematic Uncertainty of the Signal Efficiency

For values calculated using the ABSW or BHI models (see Section 6.2.4), 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 6.2.4, 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%.

8.2 Background Systematic Uncertainty

8.2.1 Primary K^+ Mode Background Systematic Uncertainty

The 2-D weighting described in Section 6.2.1 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.

8.2.2 Low-momentum Mode Background Systematic Uncertainty

The 2-D weighting described in Section 6.2.1 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.

8.2.3 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 6.2.3. This correction will be used as the systematic uncertainty to the background prediction, for the same reasons as discussed in Section 8.2.1. For this mode, the uncertainty is 10%.

8.3 B-counting Systematic Uncertainty

As shown in Table 2, the number of $B\overline{B}$ pairs includes an uncertainty: $459.0 \pm 5.1 \times 10^6$. 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 8.1. 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%).

9 Results

9.1 Observations and Upper Limits

Table 15 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 15 are calculated with the Barlow calculator [43], 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 16 shows the resulting central values of the branching fractions, including partial branching fractions.

Tables 17 to 19 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 14), 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.

The combined K^+ and K^0 limits shown in Tables 17 to 19 are calculated using the

frequentist method of the Barlow calculator. The combined limit is higher than the K^+ limit by itself due to the high K^0 limit, which comes from the observation of 2.2 excess K_S events when the SM branching fraction predicts only 0.5 events. In comparison, a Bayesian method [44] of combining limits will never be worse than the lowest limit, and it can give a better combined limit if the limits being combined are close to each other. In this case, because the K^+ limit is so much less than the K^0 limit, the Bayesian method gives a combined limit equal to the K^+ limit.

Figures 45 and 46 show histograms of the classifier output for the data events.

Table 15: 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±2.6±0.9(2.8)	$1.8^{+6.2}_{-5.1}(+8.0)$	38%
Low-momentum K^+	164±13	187±10±46(47)	$-23^{+49}_{-48}(+63)$	33%
K_S	$6.1^{+4.0}_{-2.2}$	3.9±1.3±0.4(1.4)	$2.2^{+4.1}_{-2.8}(+5.0)$	23%

Table 16: 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 17: Upper limits under the ABSW model [15] 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 18: Upper limits under the BHI model [14] 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 19: 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}

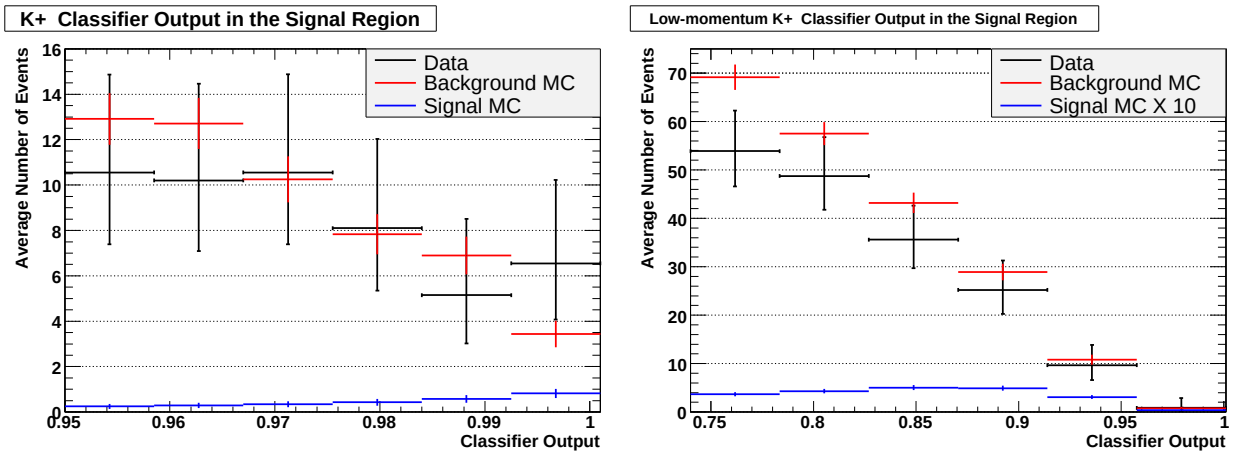

Figure 45: Data, background MC, and signal MC events in the K^+ signal region. The left plot shows the primary signal region, with signal MC scaled to the number of excess events observed. The right shows the low-momentum signal region, with the signal MC scaled arbitrarily for illustration. These plots show the average number of events for the twenty trees. Both plots include some sideband below the signal region.

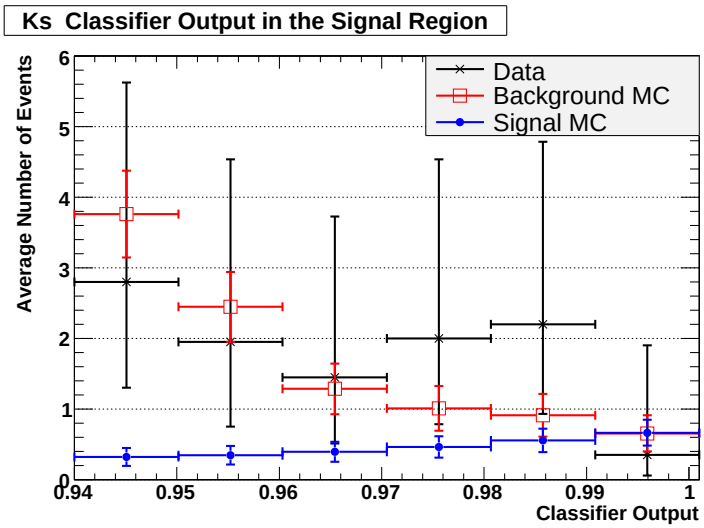

Figure 46: Data, background MC, and signal MC events in the K_S signal region, with signal MC scaled to the number of excess events observed. The plot shows the average number of events for the twenty trees, and it includes some sideband below the signal region.

9.2 Comparison with Previous Experiments

The BHI model provides the baseline for comparison with previous experiments. Figure 47 compares the results of this analysis with the previous limits, which came from the

previous best *BABAR* result [45] and the Belle experiment [16]. 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 $B_u^+ \to K^+\nu\overline{\nu}$, but reduces the upper limit for the $B_d^0 \to K^0\nu\overline{\nu}$ mode by a factor of three, compared with Belle.

9.3 Comparison with Theoretical Models

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

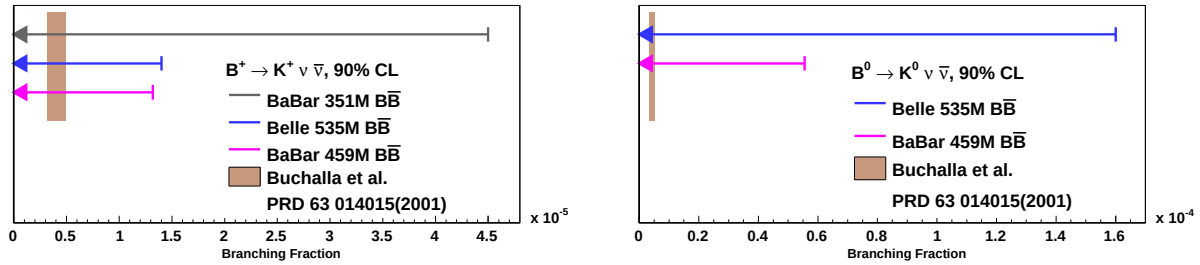

Figure 47: Comparison of BaBar and Belle branching fraction results for the decay modes in this analysis. Belle's best results are shown in blue, and the results of this analysis are shown in magenta just below. The brown box shows the theoretical prediction of the BHI model. The left plot shows the primary K^+ mode, and the right shows the K^0 mode.

For the model of scalar WIMP dark matter [28] discussed in Section 2 and shown in Figure 6, the kaon momentum range of the results must be considered. The minimum kaon momentum included in a result helps determine the position of its curve on the exclusion plot. The primary K^+ result only covers kaon momentum down to 1.5 GeV/c (in both the lab and CMS frames). To cover a larger range, the low-momentum and high-momentum K^+ results can be combined, which gives a 90% confidence level upper limit on the branching fraction of 3.6×10^{-5} (for all signal models). This limit is higher than the other results and thus is not listed in the tables earlier in this section, but its advantage is that it covers a

kaon momentum range down to 1.0 GeV/c, as seen in Figure 155. Plotting both results on the exclusion plot gives Figure 48, which shows that these results, like those of Belle, exclude dark matter particles with a mass below about 1.7 GeV/c^2.

With regard to the MSSM model [15], the result of this analysis is a 90% confidence level branching fraction upper limit of 1.5×10^{-5} under the ABSW signal model, which is still quite far above the MSSM prediction of 5.8×10^{-6}, which also uses the ABSW signal model. Much more data, on the order of 50 ab^{-1}, would be required to obtain an upper limit closer to the MSSM prediction.

In the case of the other models mentioned in Section 2 (unparticle models [32], models with a single universal extra dimension [34], and WIMP-less dark matter [33]), the new upper limit from this analysis of 1.3×10^{-5} (under the BHI signal model) only slightly increases the constraints on these new physics models compared to Belle's upper limit of 1.4×10^{-5}.

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 improve upon the Belle results [16], and provide constraints to several new physics models.

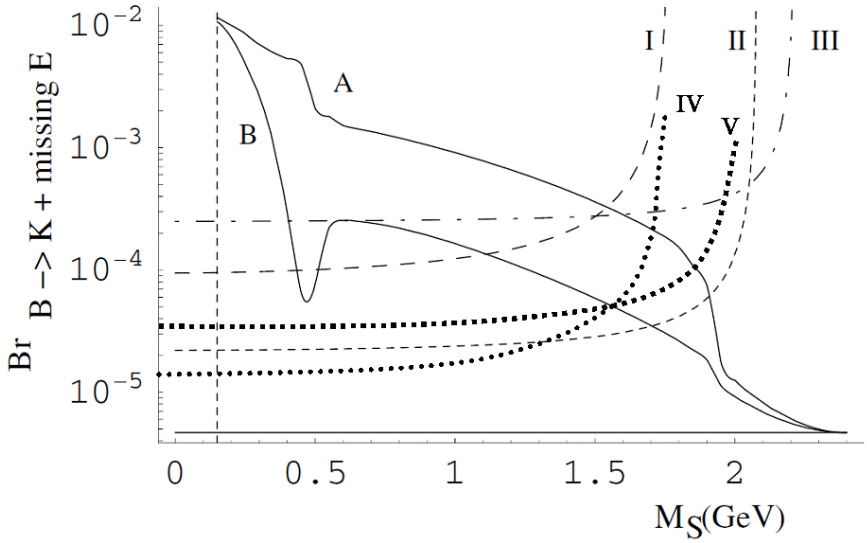

Figure 48: Branching fraction of $B \to K$ + *missing energy* vs. the mass of a scalar DM particle described by [28]. The model's allowed region is between curves A and B. Curve I is the limit set by an early *BABAR* $B_u^+ \to K^+\nu\overline{\nu}$ result [30] with 82 fb^{-1}, an upper limit of 7.0×10^{-5}, and a cutoff of $p_K > 1.5$ GeV/c. Curve II is an estimated limit projected for 250 fb^{-1} of *BABAR* data, with an upper limit of 2.1×10^{-5} and a cutoff of $p_K > 1.0$ GeV/c. Curve III is the limit set by a CLEO result [31], with an upper limit of 24×10^{-5} and a cutoff of $p_K > 0.7$ GeV/c. Curve IV is the result of this analysis with 418 fb^{-1} of data, an upper limit of 1.3×10^{-5}, and a cutoff of $p_K > 1.5$ GeV/c. Curve V combines the low- and high-kaon-momentum results of this analysis to give upper limit of 3.6×10^{-5} and a cutoff of $p_K > 1.0$ GeV/c.

10 Conclusion and Outlook

The summary of this analysis is shown in Table 20. All the upper limits in the table are based upon the BHI model [14] for consistency. A graphical comparison of these results with the previous best Belle results [16] are shown in Figure 47.

Table 20: 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 [16]. All the upper limits in this table are based upon the BHI model [14].

	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}

These results rule out a model of scalar dark matter [28] with scalar particle masses of less than 1.7 GeV/c^2, as shown in Figure 48. Other new-physics models have not yet been ruled out, but this analysis finds no discrepancy with the SM.

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 $B^+_u \to K^+\nu\overline{\nu}$ decay by combining results using semileptonic tags with those using hadronic tags. The $B^0_d \to K^0_S\nu\overline{\nu}$ 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 K^+ mode

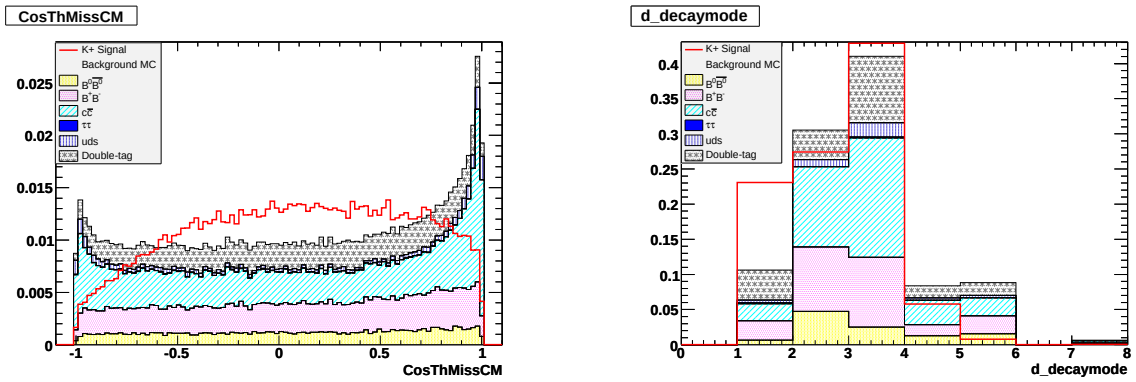

Figure 49: Signal/background histograms for CosThMissCM on left and d_decaymode on right.

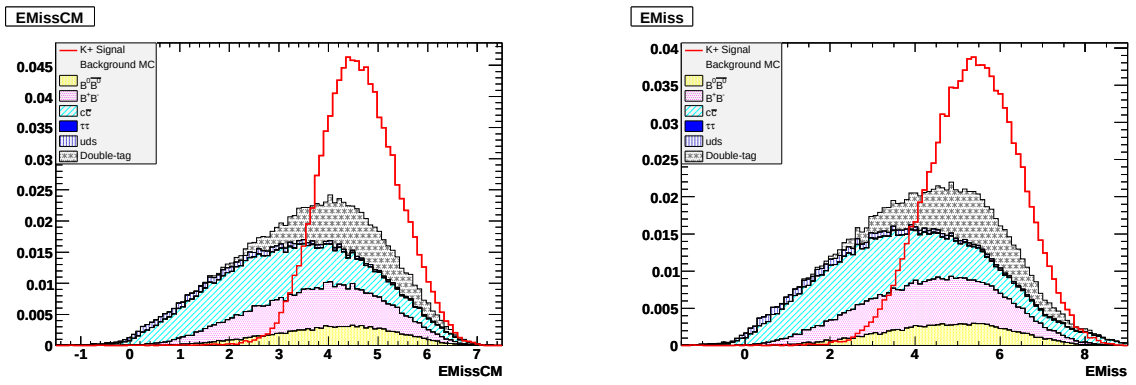

Figure 50: Signal/background histograms for EMissCM on left and EMiss on right.

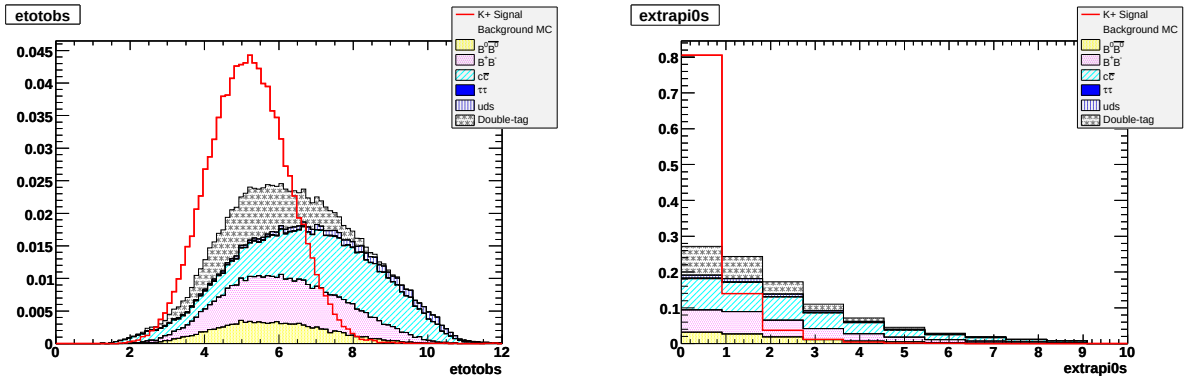

Figure 51: Signal/background histograms for etotobs on left and extrapi0s on right.

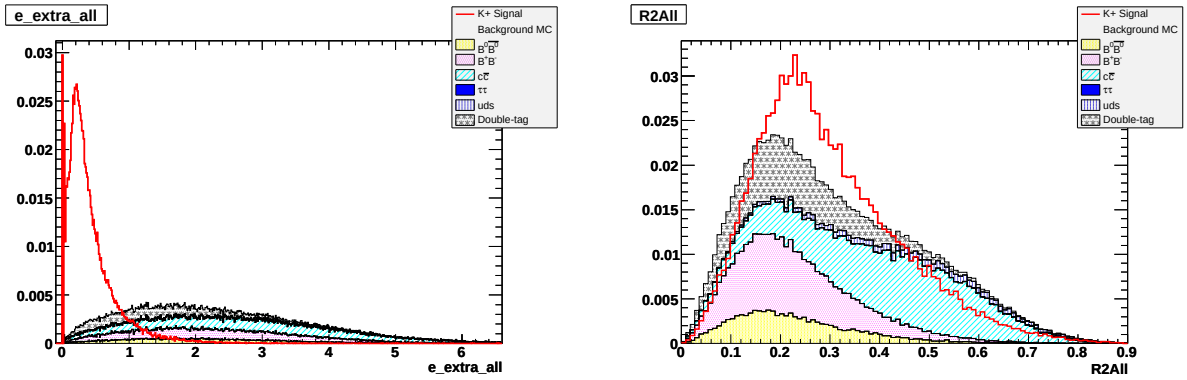

Figure 52: Signal/background histograms for e_extra_all on left and R2All on right.

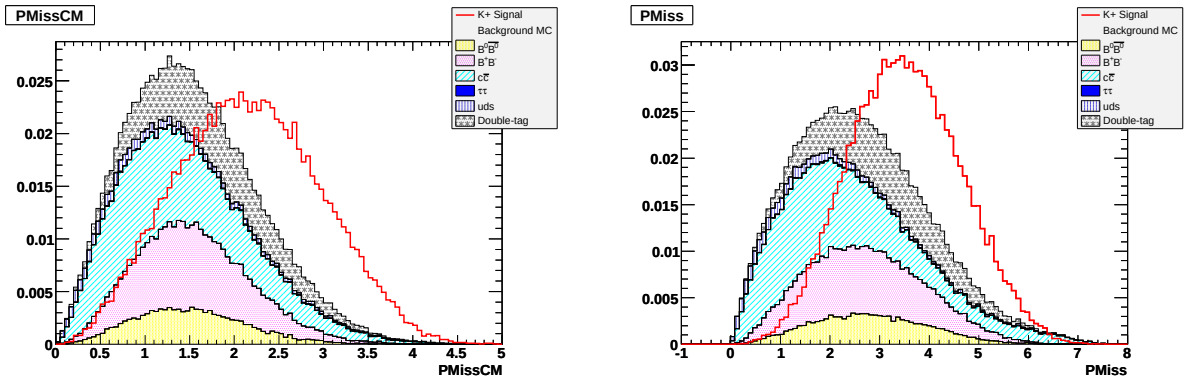

Figure 53: Signal/background histograms for PMissCM on left and PMiss on right.

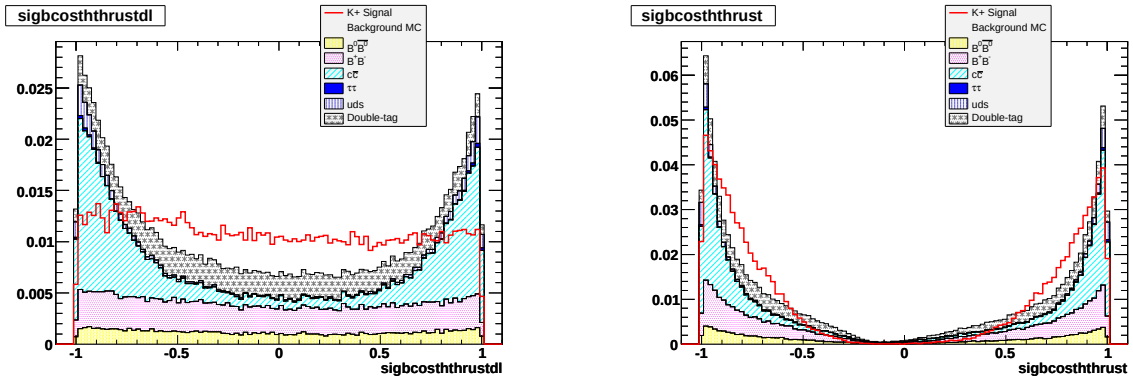

Figure 54: Signal/background histograms for sigbcostthrustdl on left and sigbcostthrust on right.

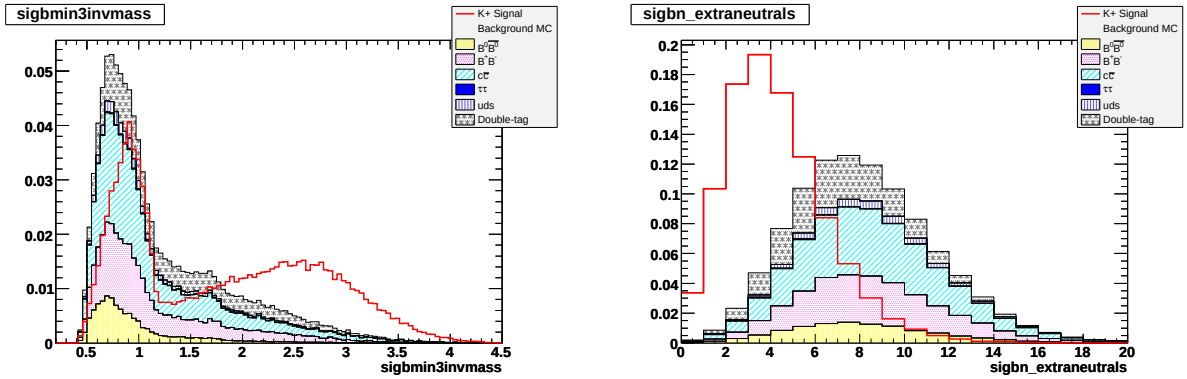

Figure 55: Signal/background histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.

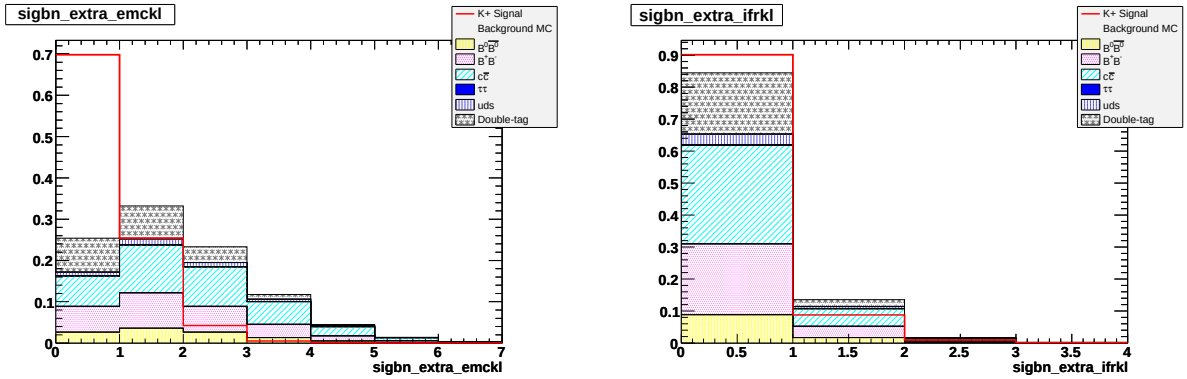

Figure 56: Signal/background histograms for sigbn_extra_emckl on left and sigbn_extra_ifrkl on right.

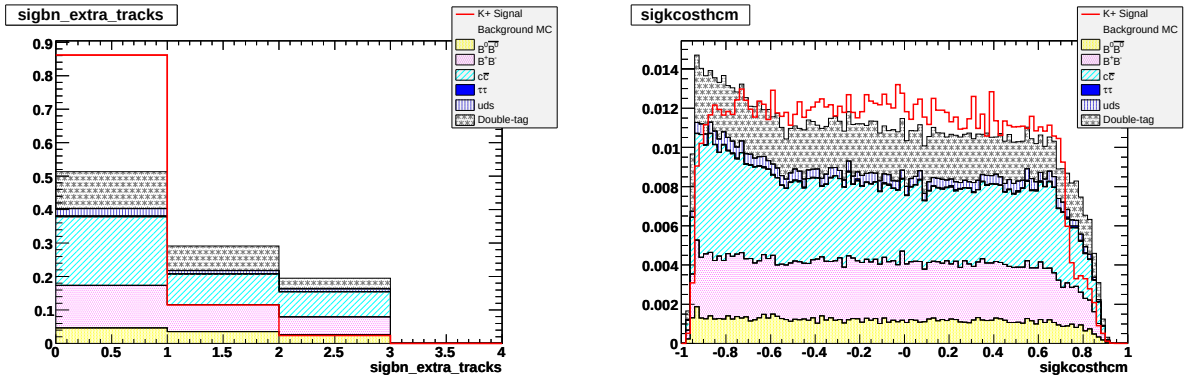

Figure 57: Signal/background histograms for sigbn_extra_tracks on left and sigkcosthcm on right.

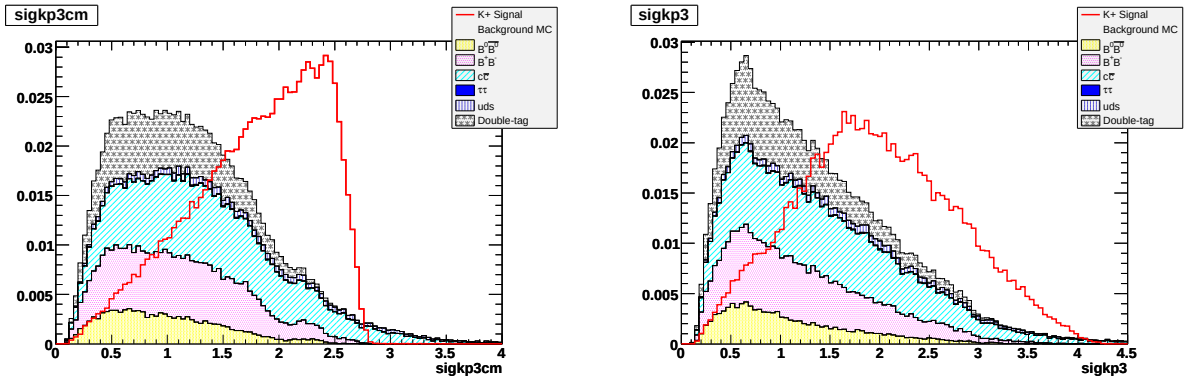

Figure 58: Signal/background histograms for sigkp3cm on left and sigkp3 on right.

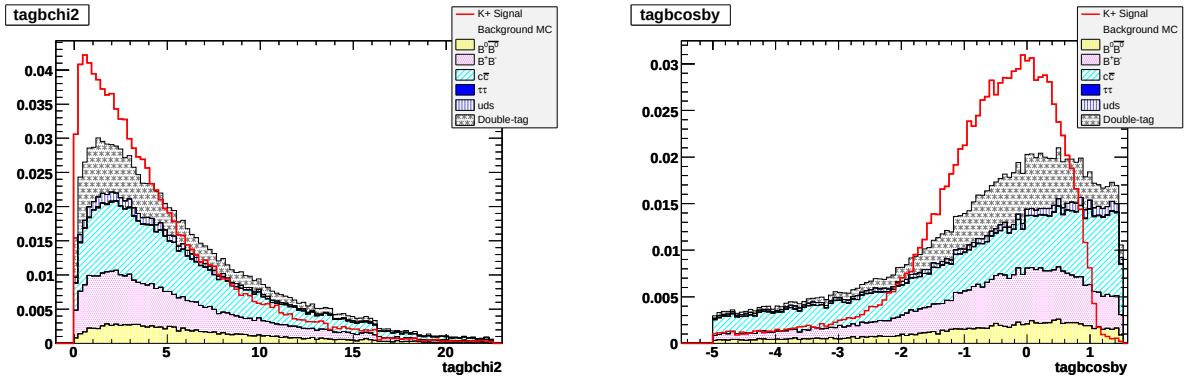

Figure 59: Signal/background histograms for tagbchi2 on left and tagbcosby on right.

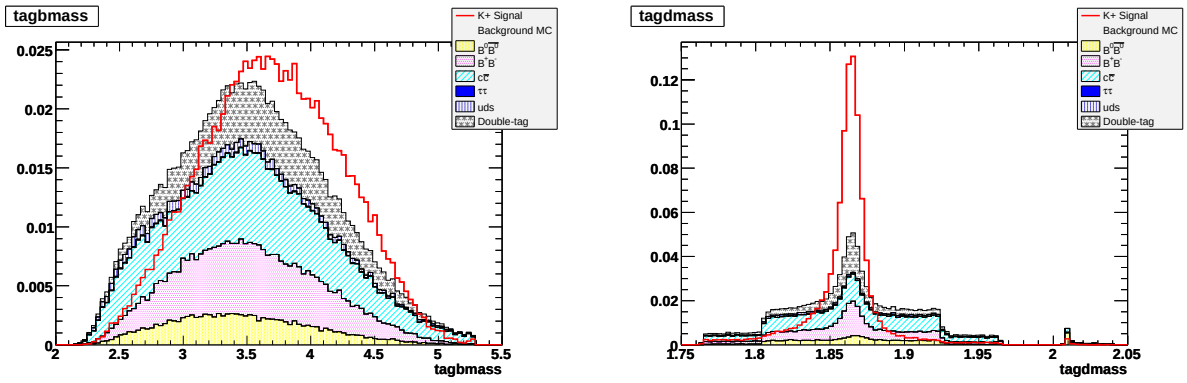

Figure 60: Signal/background histograms for tagbmass on left and tagdmass on right.

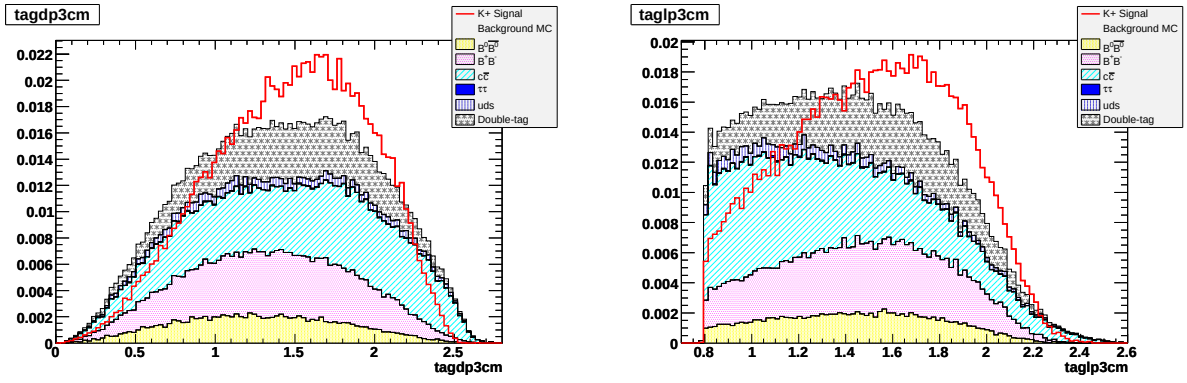

Figure 61: Signal/background histograms for tagdp3cm on left and taglp3cm on right.

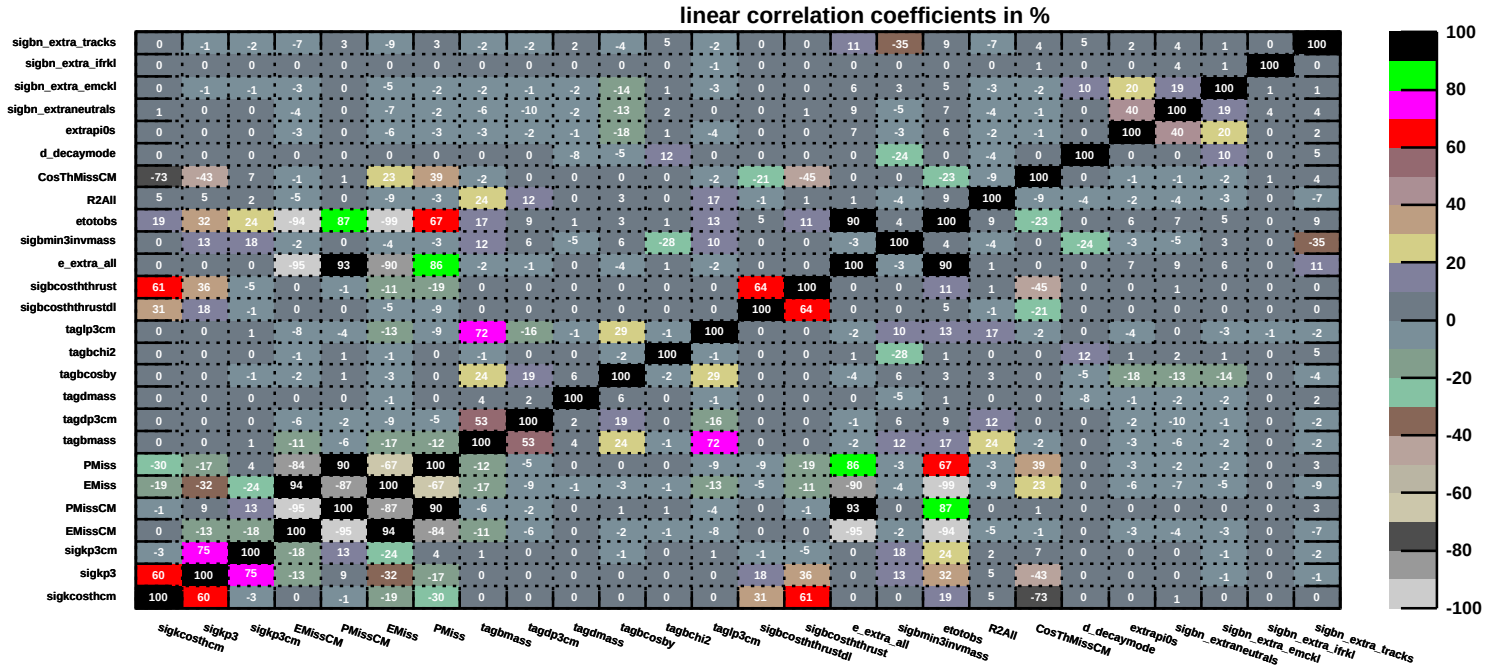

Figure 62: Correlation matrix showing the amount of correlation in signal MC between the classification variables for the K^+ trees.

A.1.2 K_S Mode

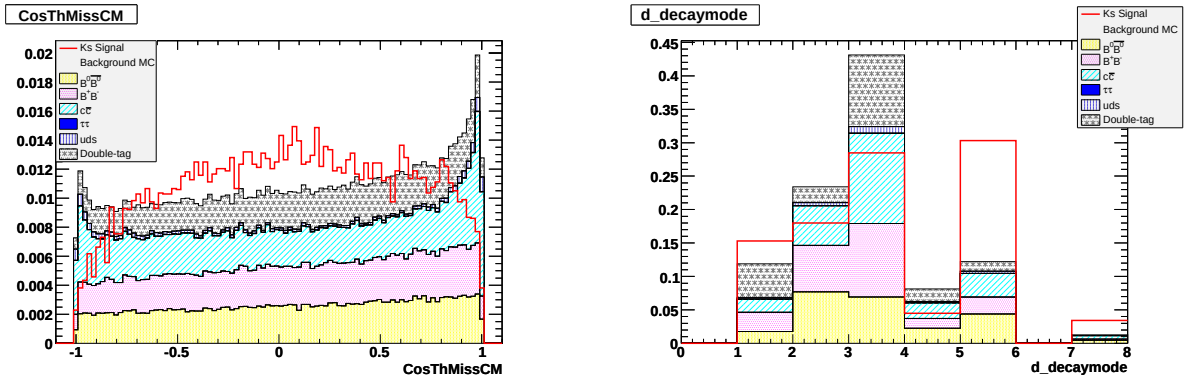

Figure 63: Signal/background histograms for CosThMissCM on left and d_decaymode on right.

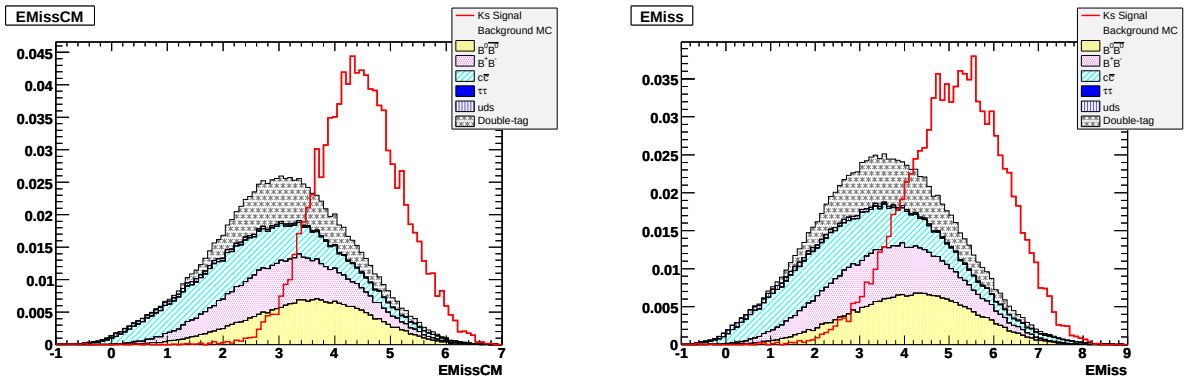

Figure 64: Signal/background histograms for EMissCM on left and EMiss on right.

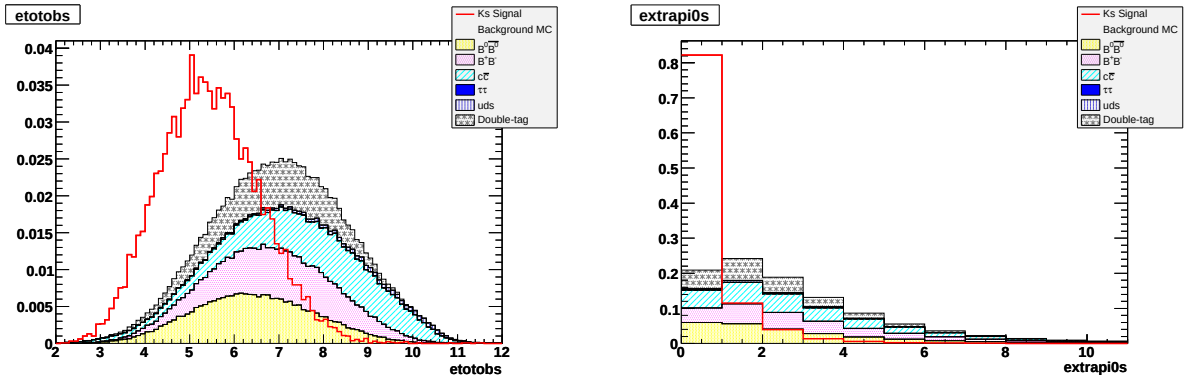

Figure 65: Signal/background histograms for etotobs on left and extrapi0s on right.

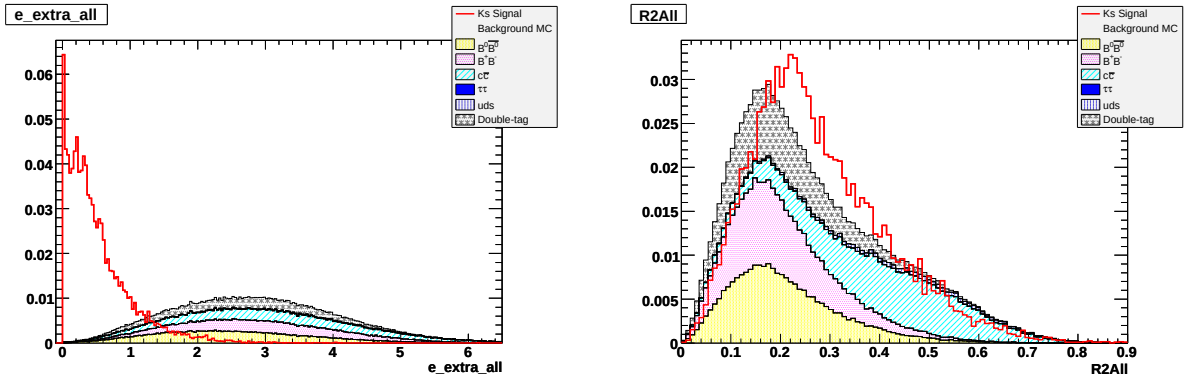

Figure 66: Signal/background histograms for e_extra_all on left and R2All on right.

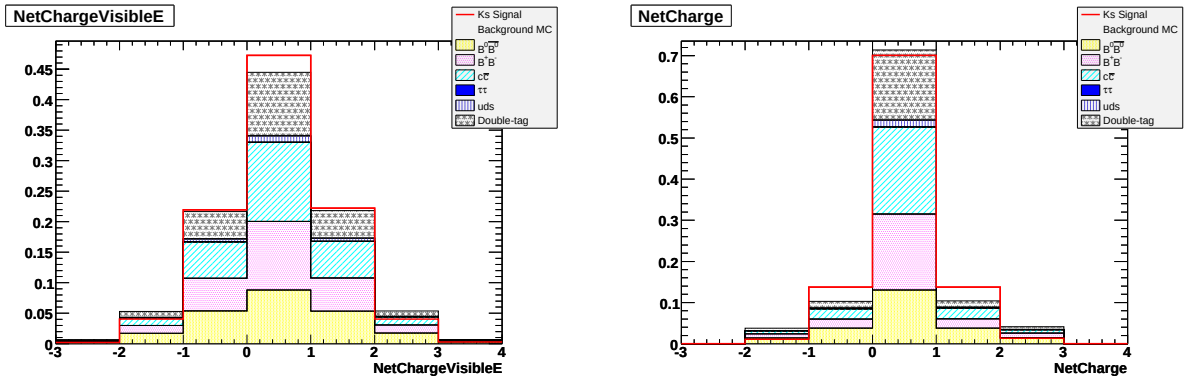

Figure 67: Signal/background histograms for NetChargeVisibleE on left and NetCharge on right.

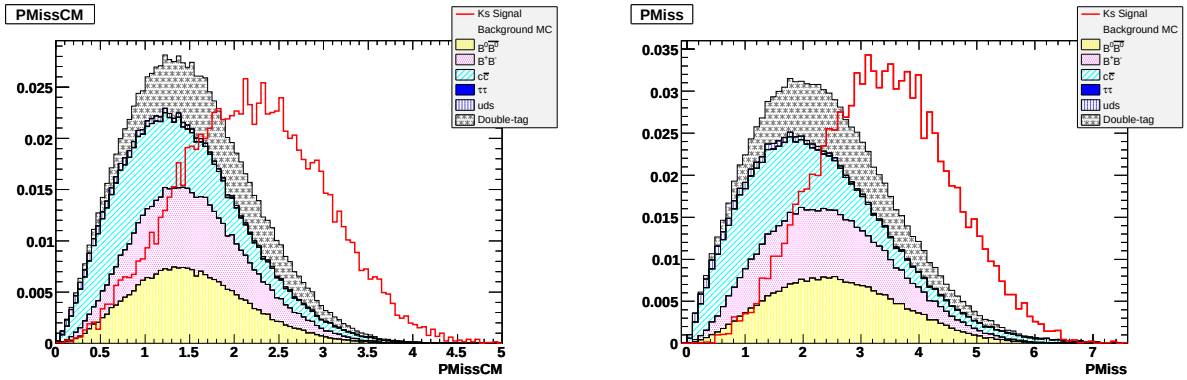

Figure 68: Signal/background histograms for PMissCM on left and PMiss on right.

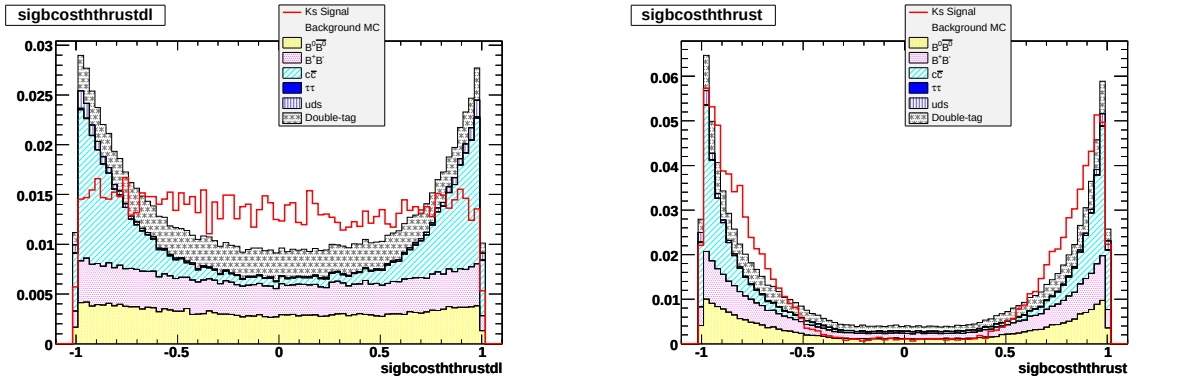

Figure 69: Signal/background histograms for sigbcosththrustdl on left and sigbcosththrust on right.

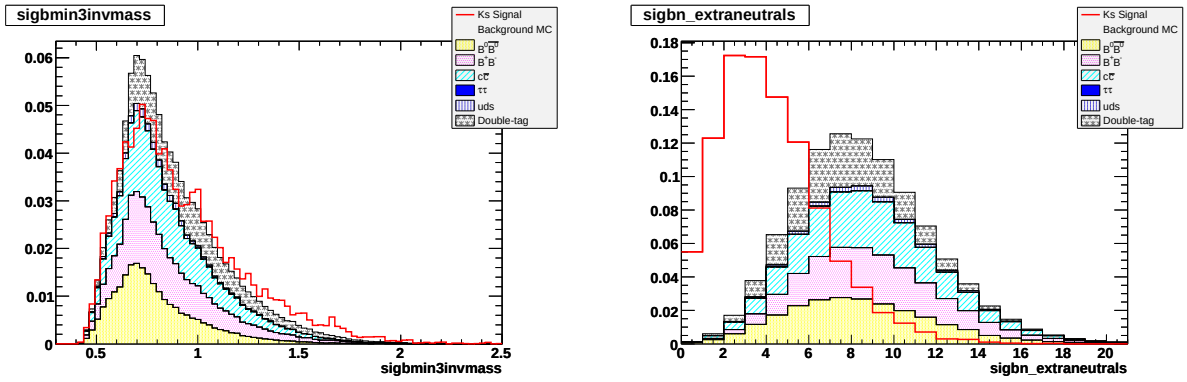

Figure 70: Signal/background histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.

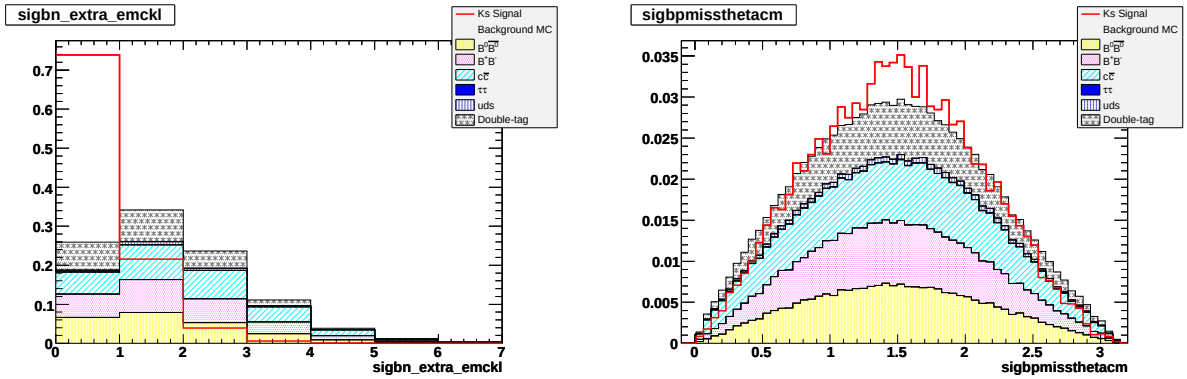

Figure 71: Signal/background histograms for sigbn_extra_emckl on left and sigbpmissthetacm on right.

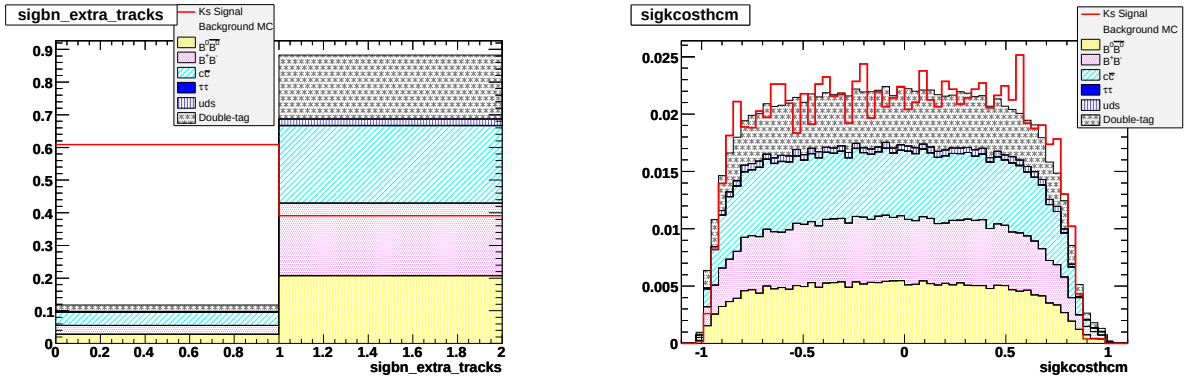

Figure 72: Signal/background histograms for sigbn_extra_tracks on left and sigkcosthcm on right.

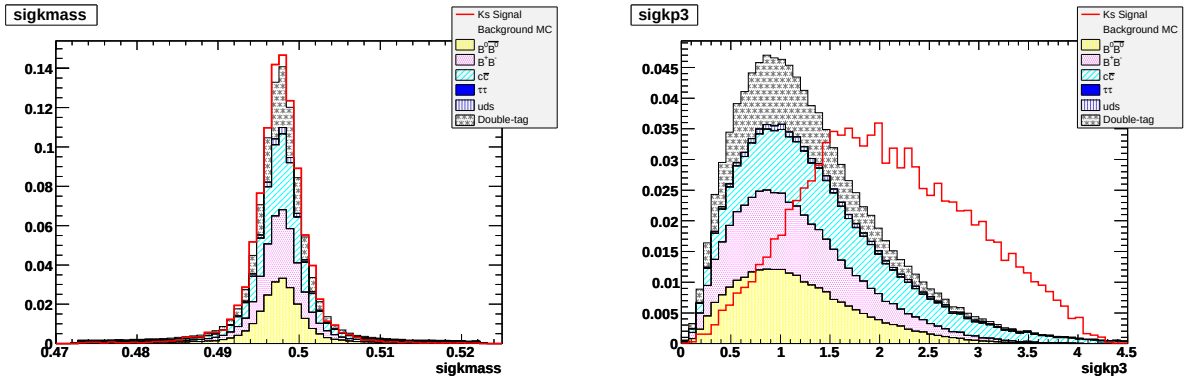

Figure 73: Signal/background histograms for sigkmass on left and sigkp3 on right.

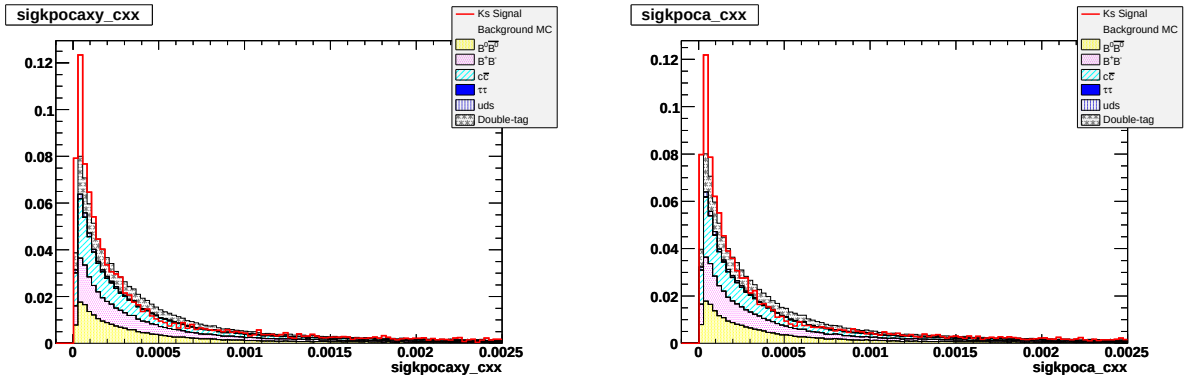

Figure 74: Signal/background histograms for sigkpocaxy_cxx on left and sigkpoca_cxx on right.

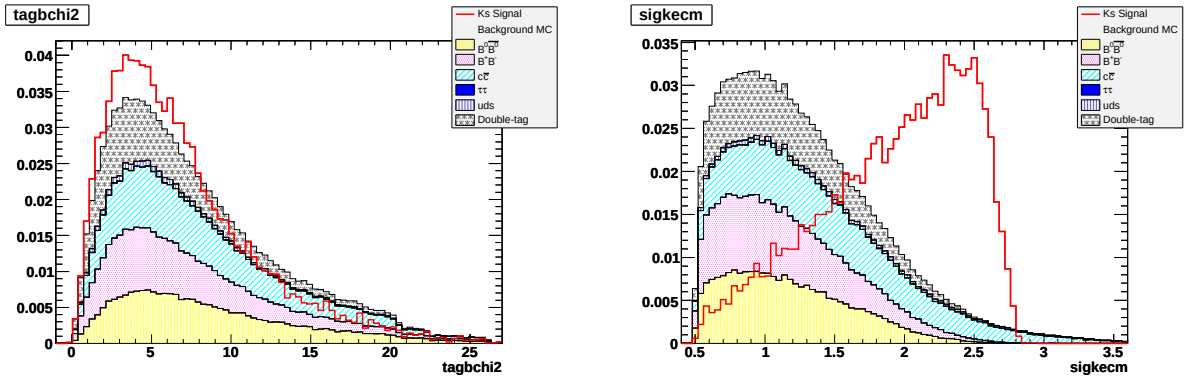

Figure 75: Signal/background histograms for tagbchi2 on left and sigkecm on right.

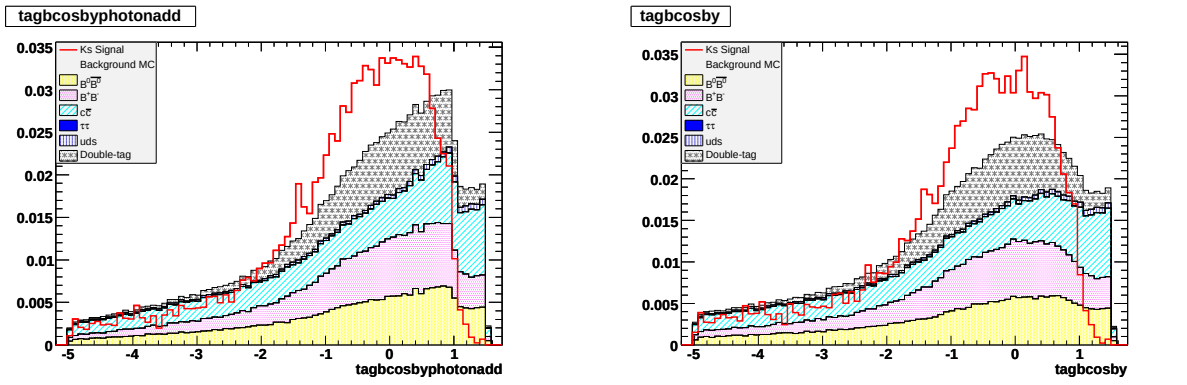

Figure 76: Signal/background histograms for tagbcosbyphotonadd on left and tagbcosby on right.

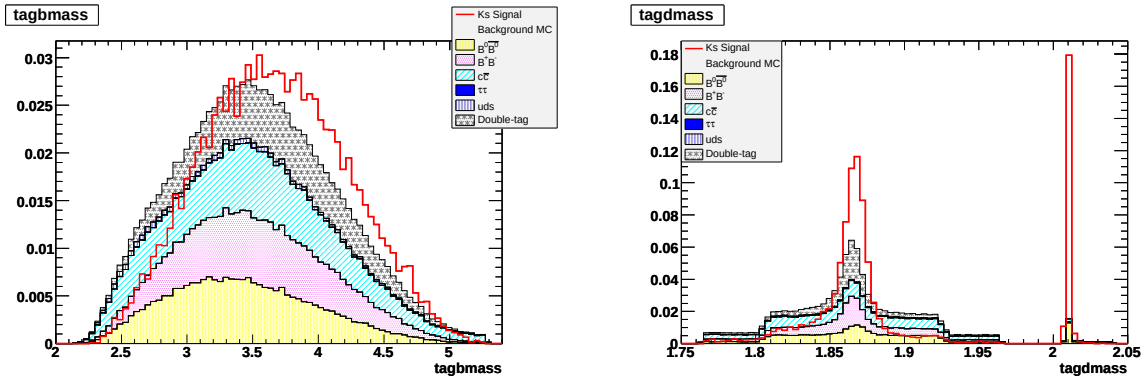

Figure 77: Signal/background histograms for tagbmass on left and tagdmass on right.

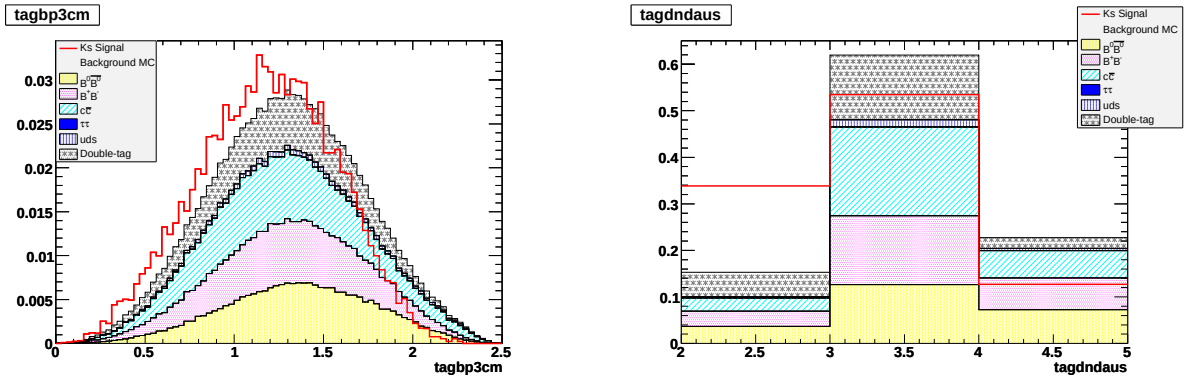

Figure 78: Signal/background histograms for tagbp3cm on left and tagdndaus on right.

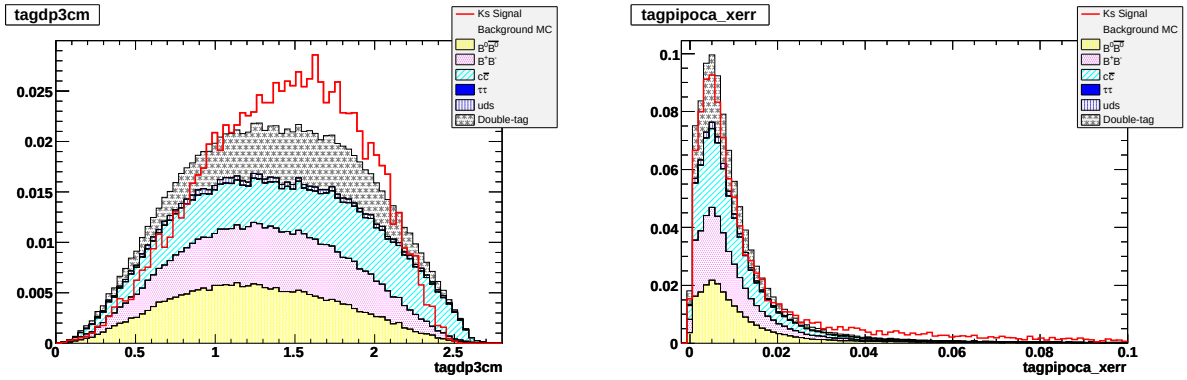

Figure 79: Signal/background histograms for tagdp3cm on left and tagpipoca_xerr on right.

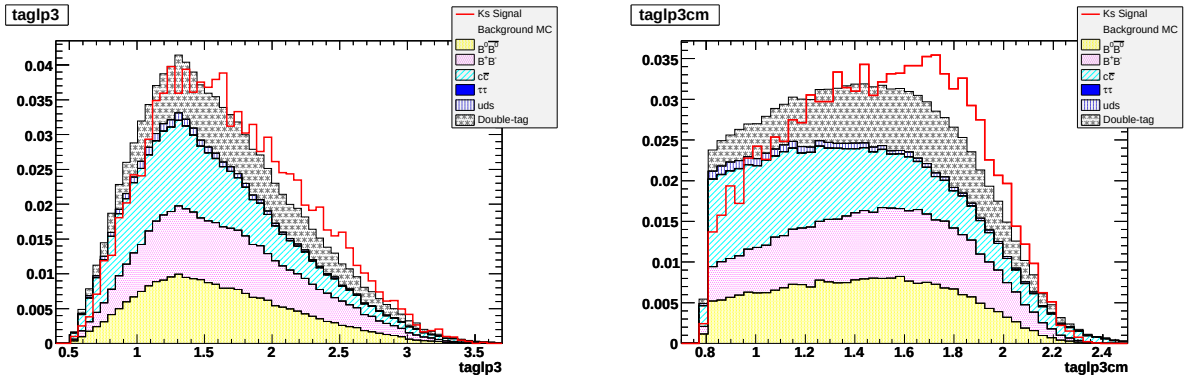

Figure 80: Signal/background histograms for taglp3 on left and taglp3cm on right.

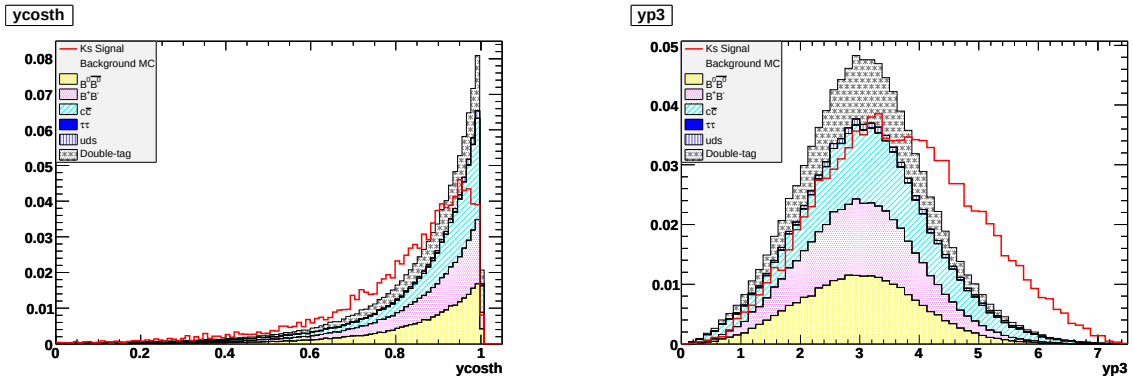

Figure 81: Signal/background histograms for ycosth on left and yp3 on right.

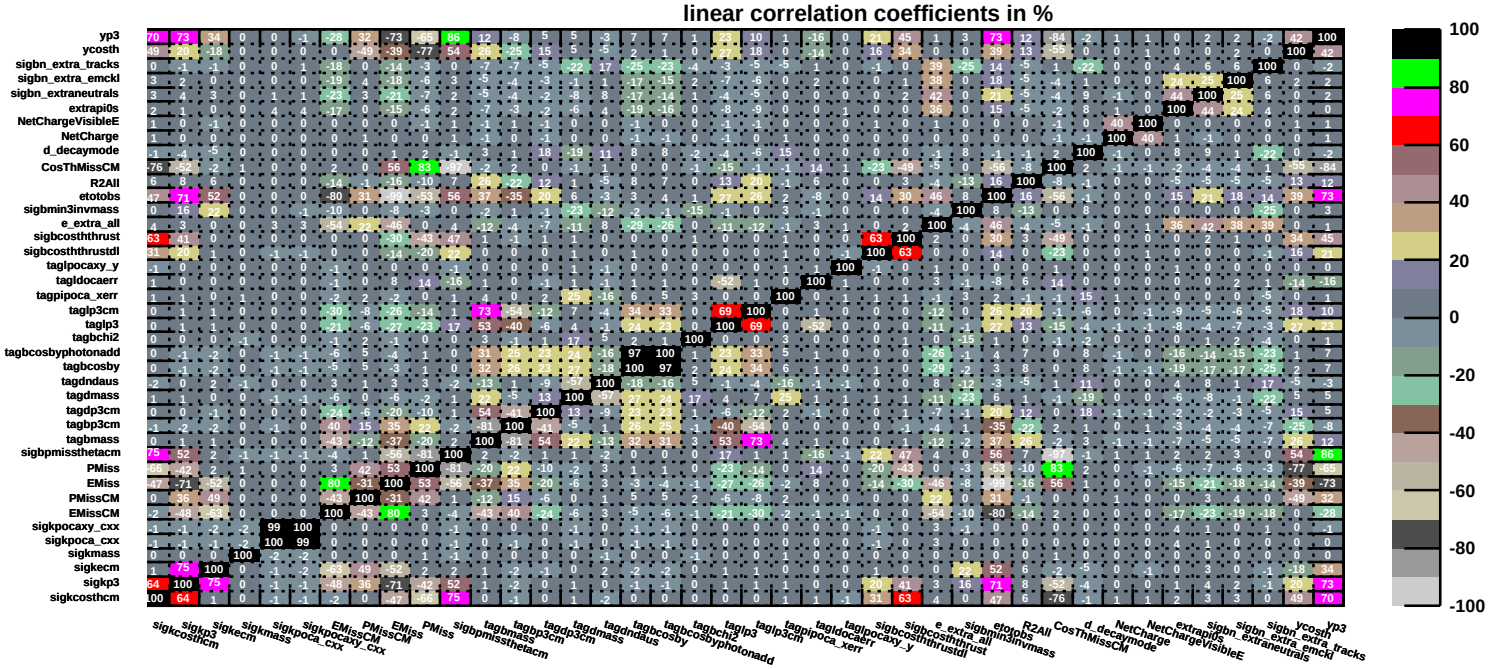

Figure 82: Correlation matrix showing the amount of correlation in signal MC between the classification variables for the K^0_S tree.

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

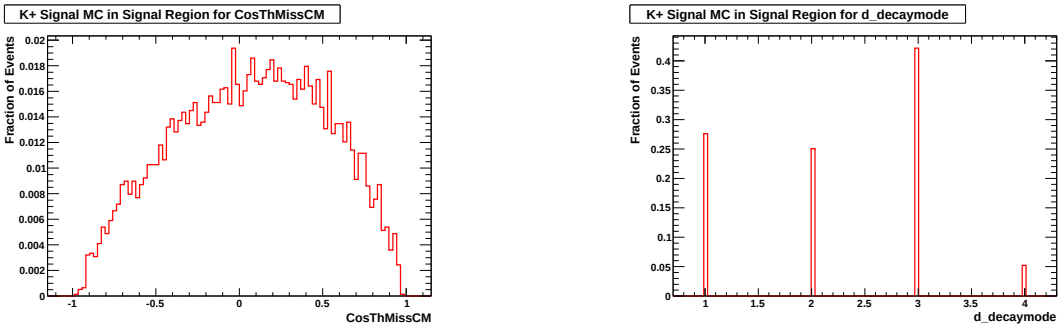

Figure 83: Signal MC histograms for CosThMissCM on left and d_decaymode on right.

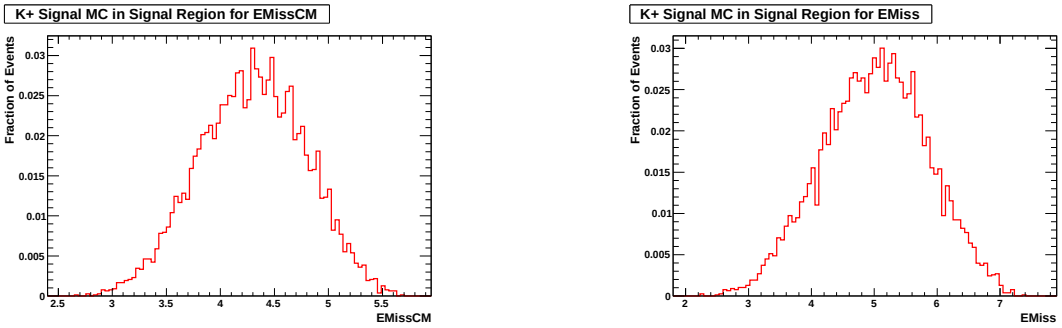

Figure 84: Signal MC histograms for EMissCM on left and EMiss on right.

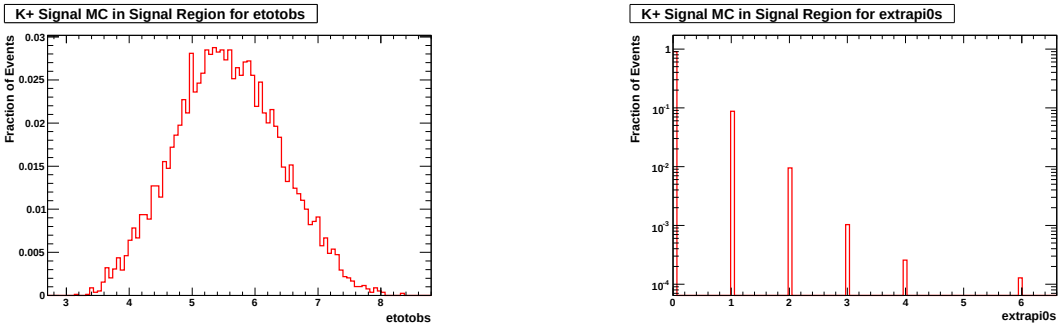

Figure 85: Signal MC histograms for etotobs on left and extrapi0s on right.

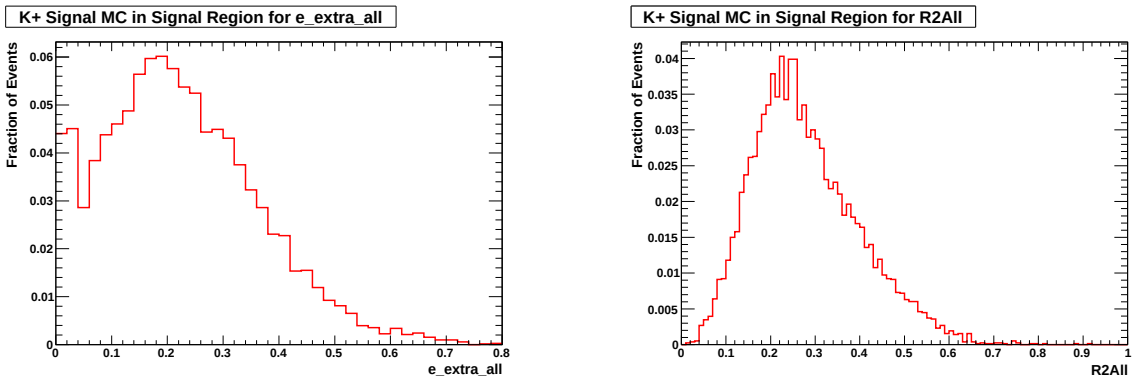

Figure 86: Signal MC histograms for e_extra_all on left and R2All on right.

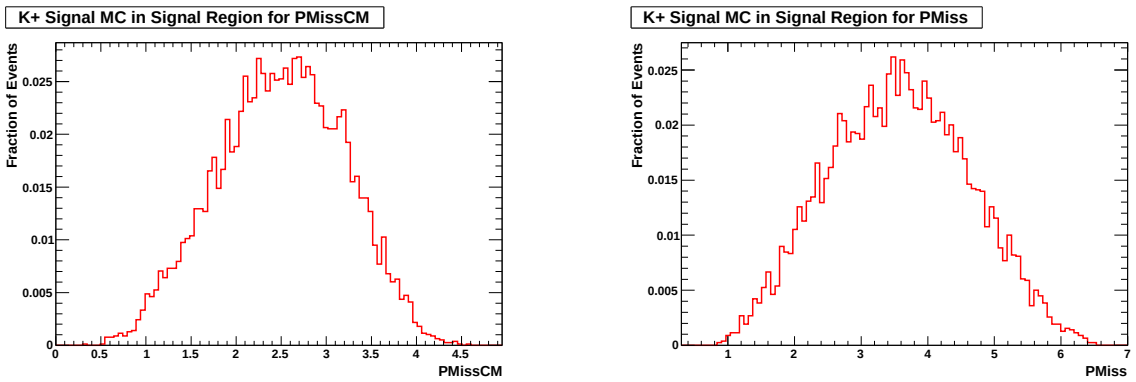

Figure 87: Signal MC histograms for PMissCM on left and PMiss on right.

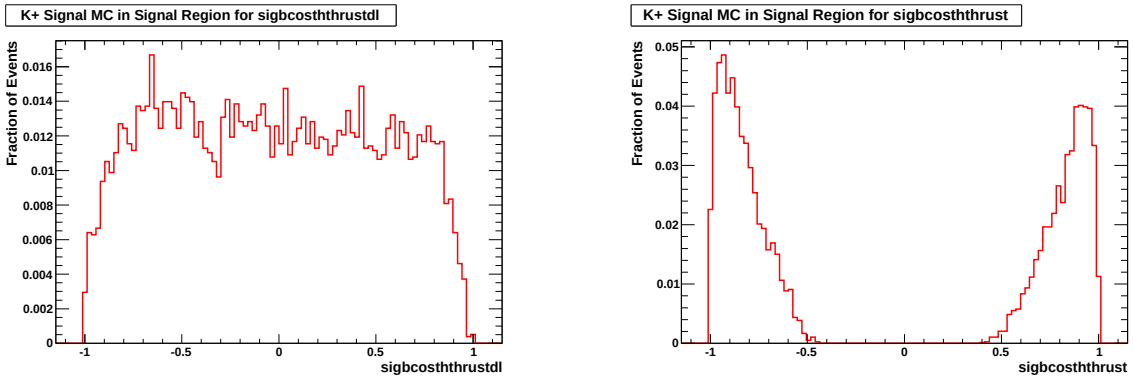

Figure 88: Signal MC histograms for sigbcostthrustdl on left and sigbcostthrust on right.

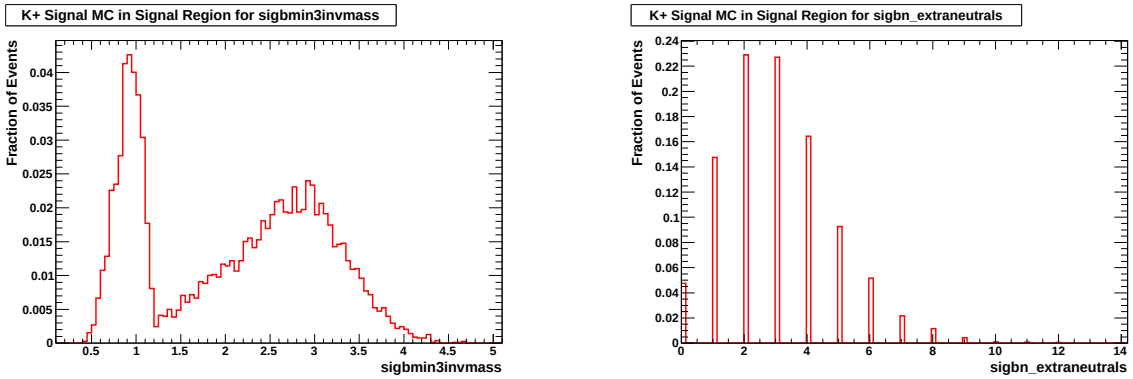

Figure 89: Signal MC histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.

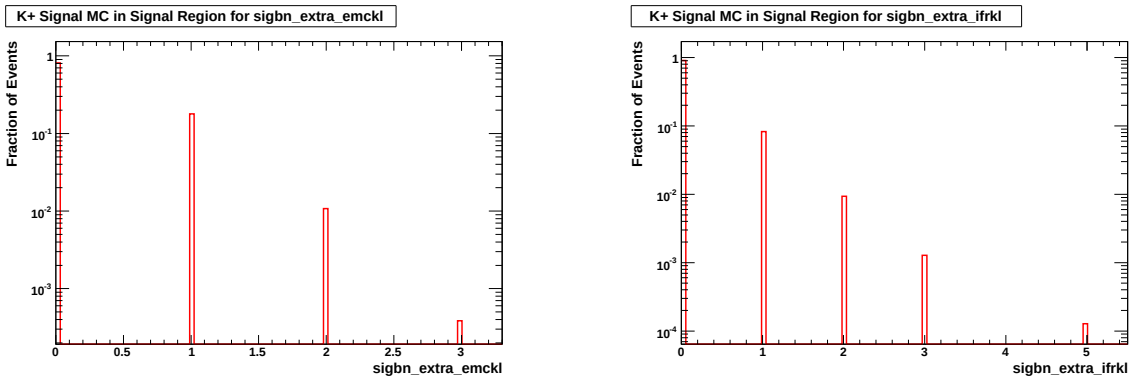

Figure 90: Signal MC histograms for sigbn_extra_emckl on left and sigbn_extra_ifrkl on right.

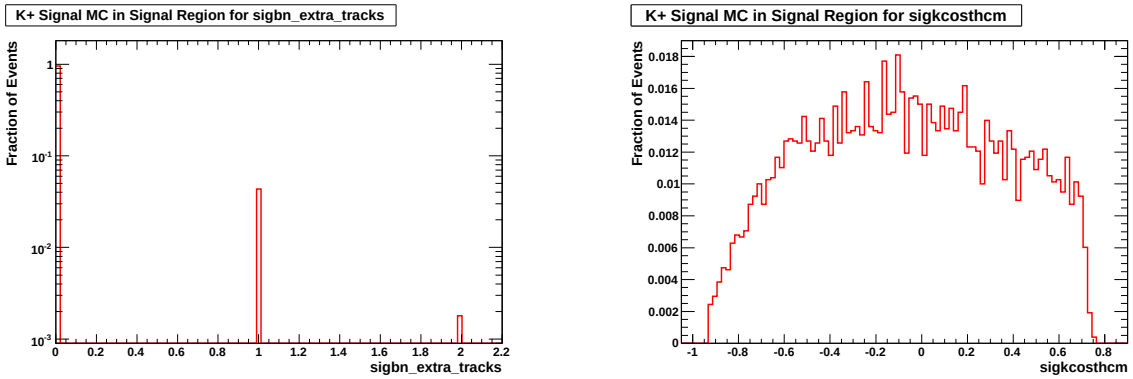

Figure 91: Signal MC histograms for sigbn_extra_tracks on left and sigkcosthcm on right.

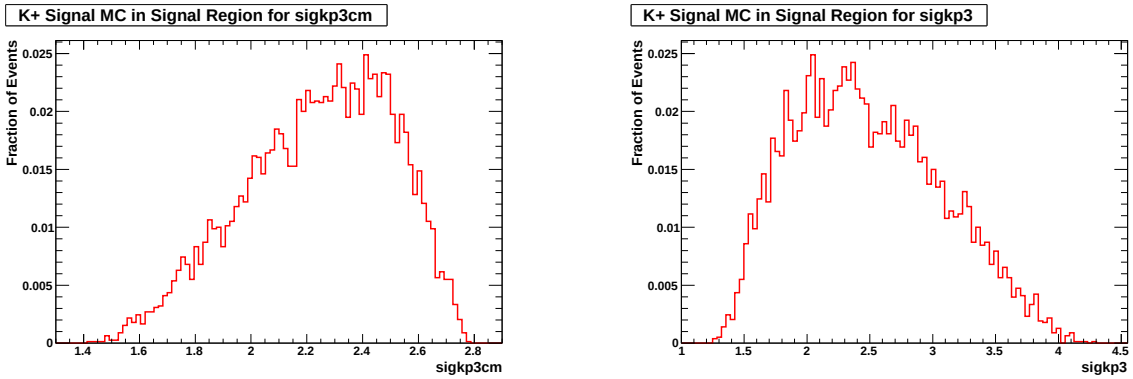

Figure 92: Signal MC histograms for sigkp3cm on left and sigkp3 on right.

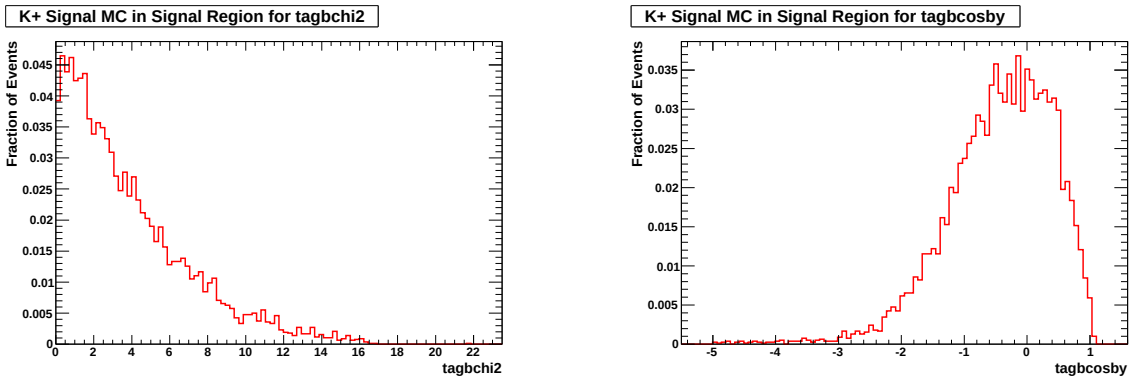

Figure 93: Signal MC histograms for tagbchi2 on left and tagbcosby on right.

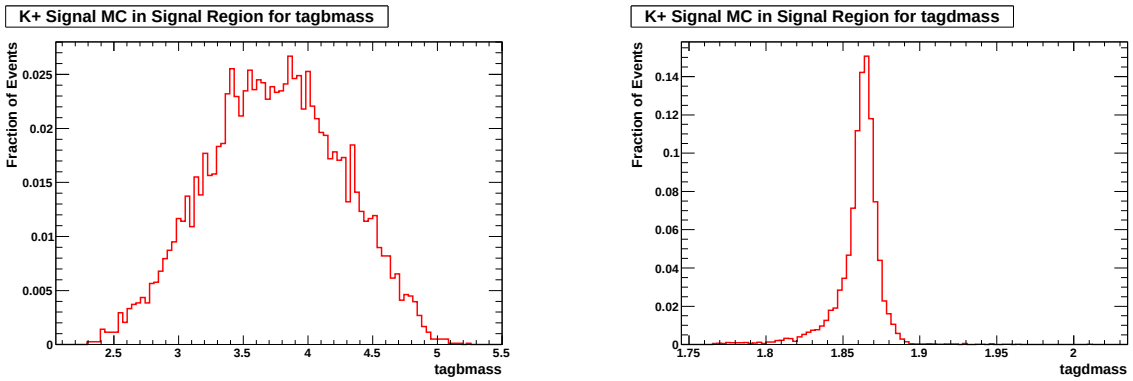

Figure 94: Signal MC histograms for tagbmass on left and tagdmass on right.

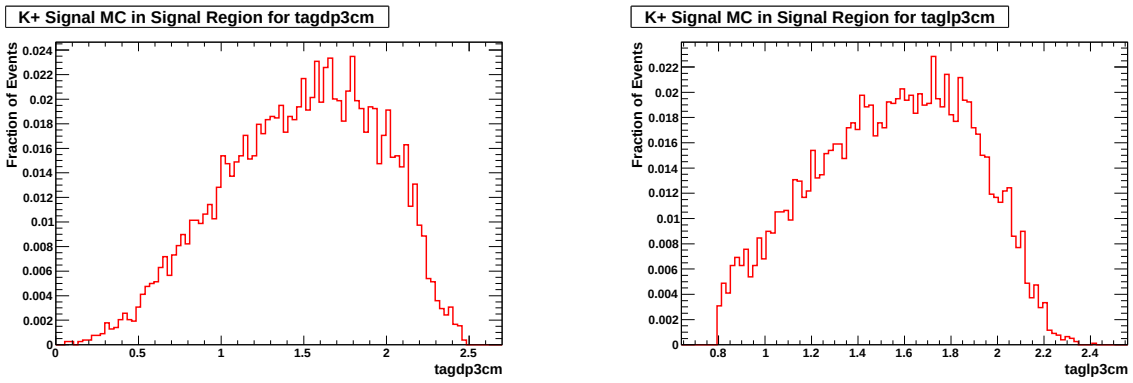

Figure 95: Signal MC histograms for tagdp3cm on left and taglp3cm on right.

A.2.2 K_S Mode

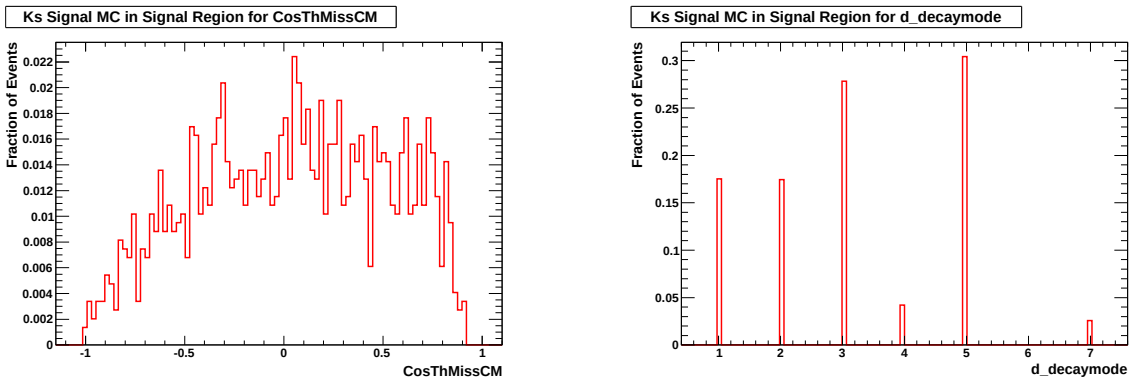

Figure 96: Signal MC histograms for CosThMissCM on left and d_decaymode on right.

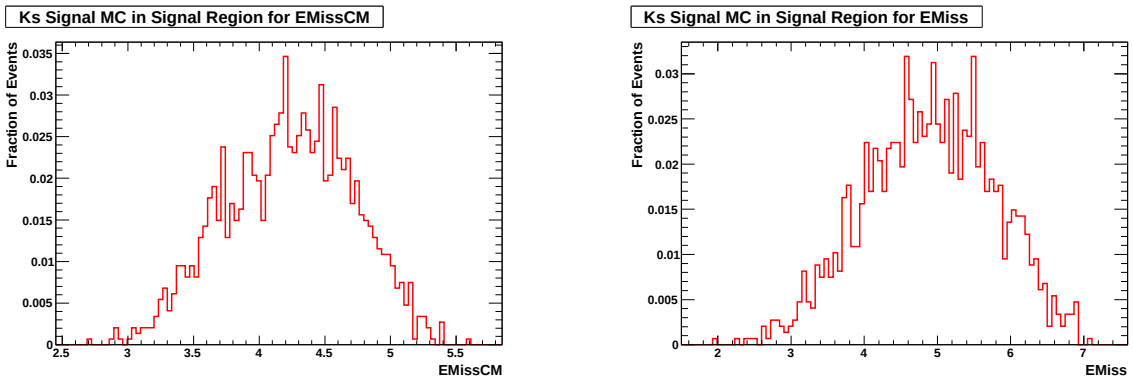

Figure 97: Signal MC histograms for EMissCM on left and EMiss on right.

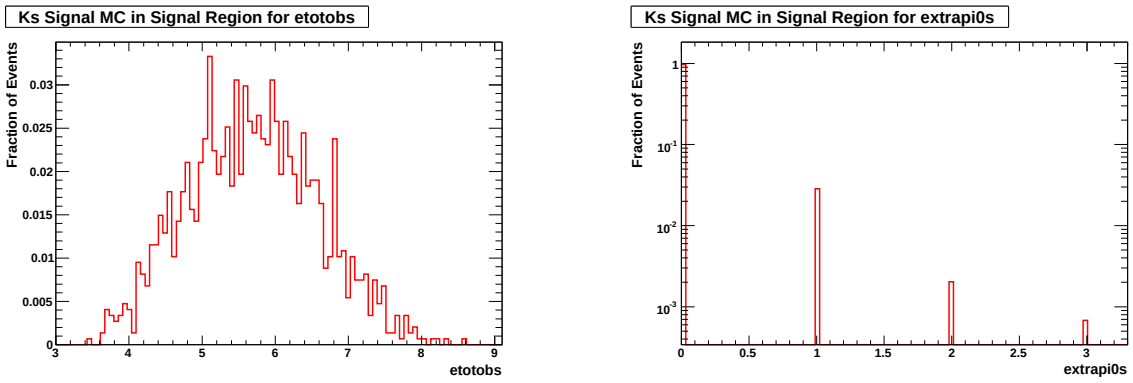

Figure 98: Signal MC histograms for etotobs on left and extrapi0s on right.

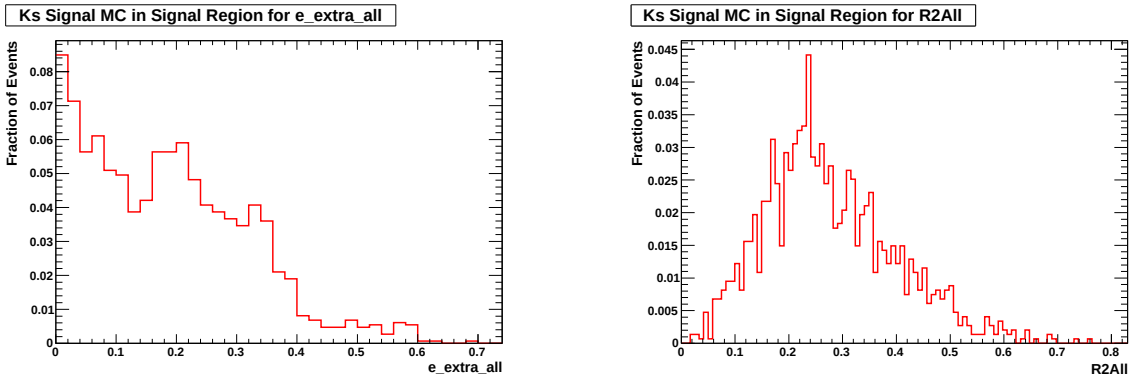

Figure 99: Signal MC histograms for e_extra_all on left and R2All on right.

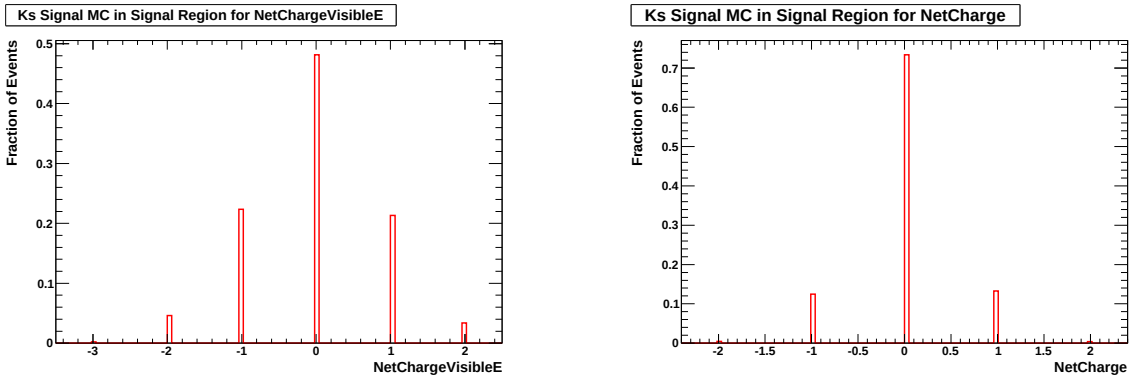

Figure 100: Signal MC histograms for NetChargeVisibleE on left and NetCharge on right.

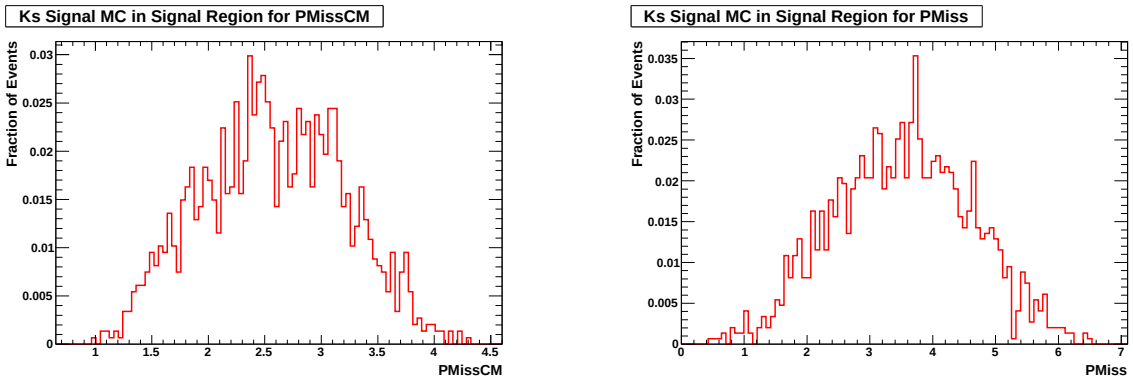

Figure 101: Signal MC histograms for PMissCM on left and PMiss on right.

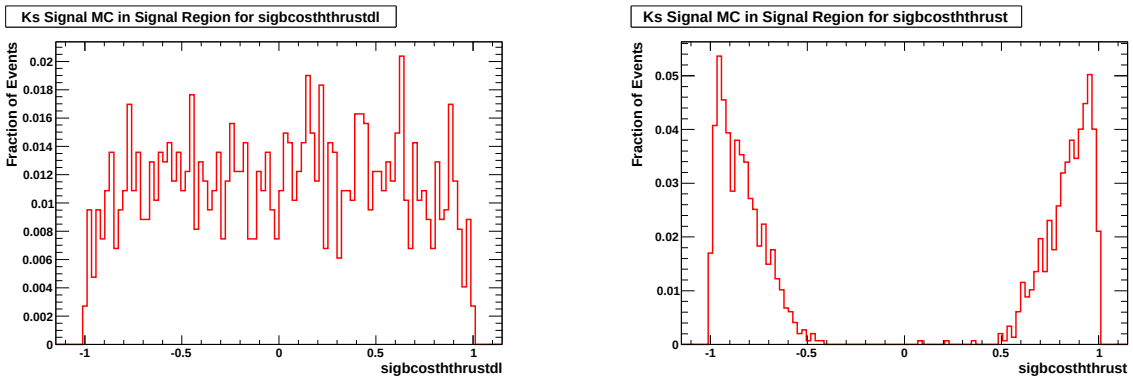

Figure 102: Signal MC histograms for sigbcostthrustdl on left and sigbcostthrust on right.

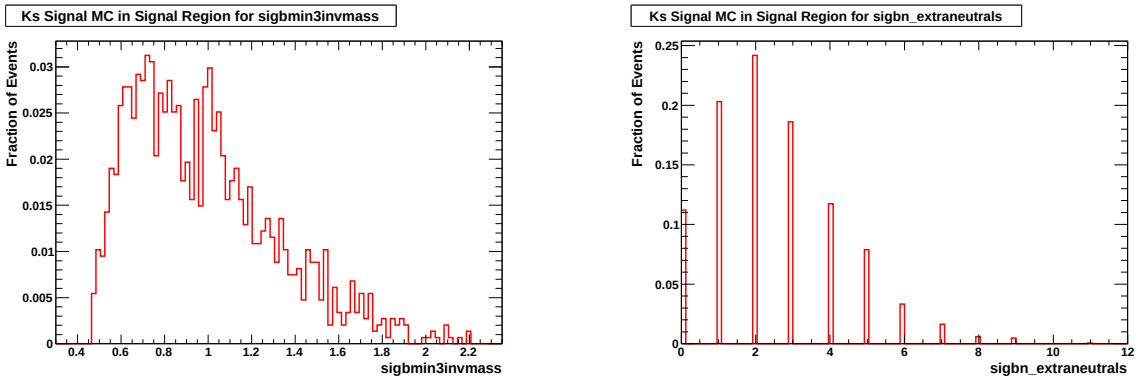

Figure 103: Signal MC histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.

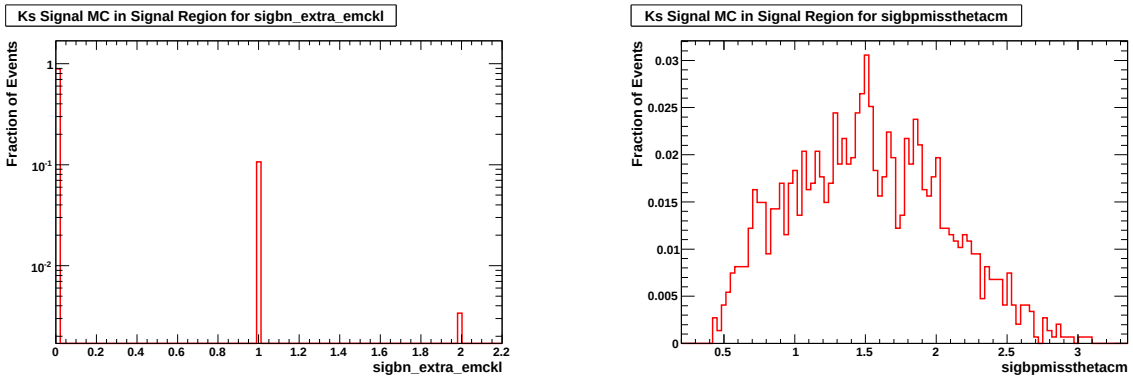

Figure 104: Signal MC histograms for sigbn_extra_emckl on left and sigbpmissthetacm on right.

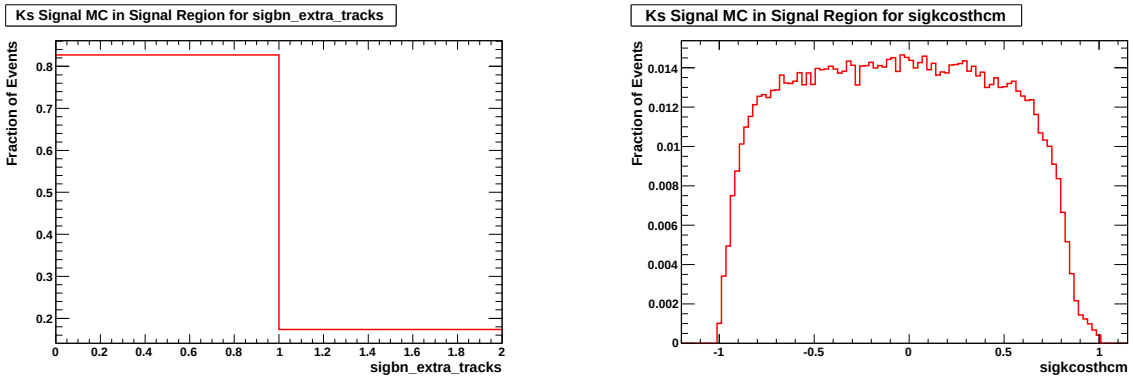

Figure 105: Signal MC histograms for sigbn_extra_tracks on left and sigkcosthcm on right.

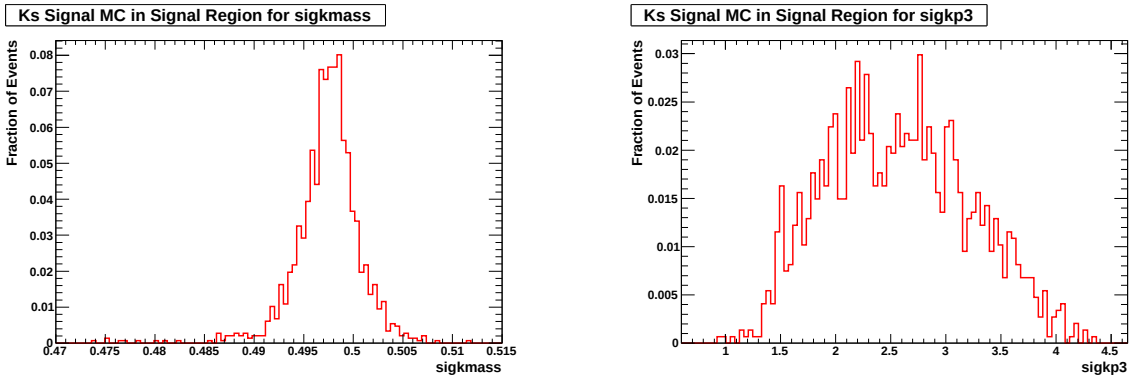

Figure 106: Signal MC histograms for sigkmass on left and sigkp3 on right.

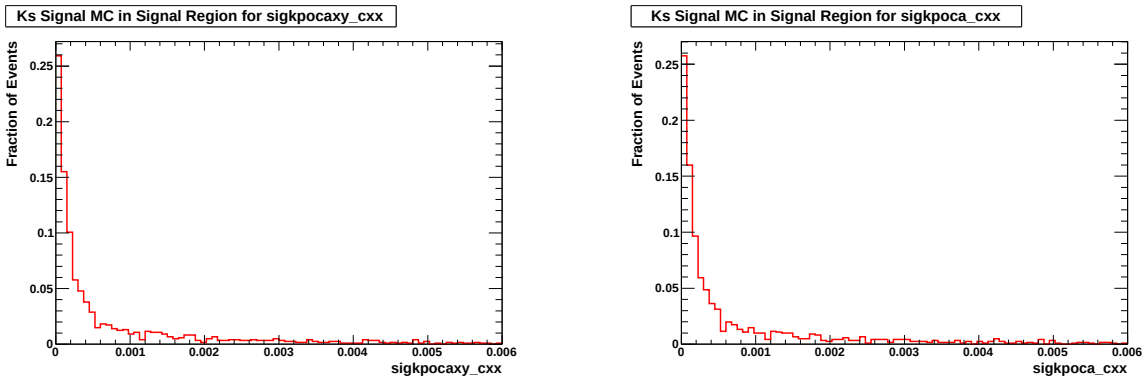

Figure 107: Signal MC histograms for sigkpocaxy_cxx on left and sigkpoca_cxx on right.

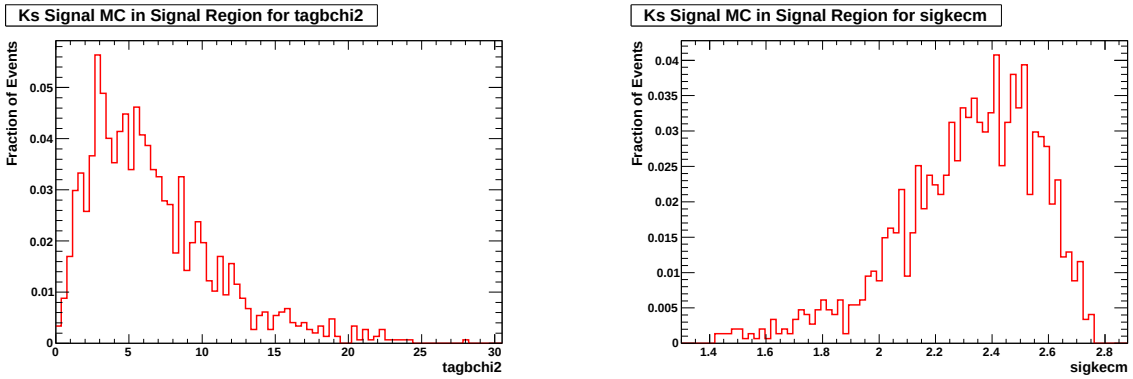

Figure 108: Signal MC histograms for tagbchi2 on left and sigkecm on right.

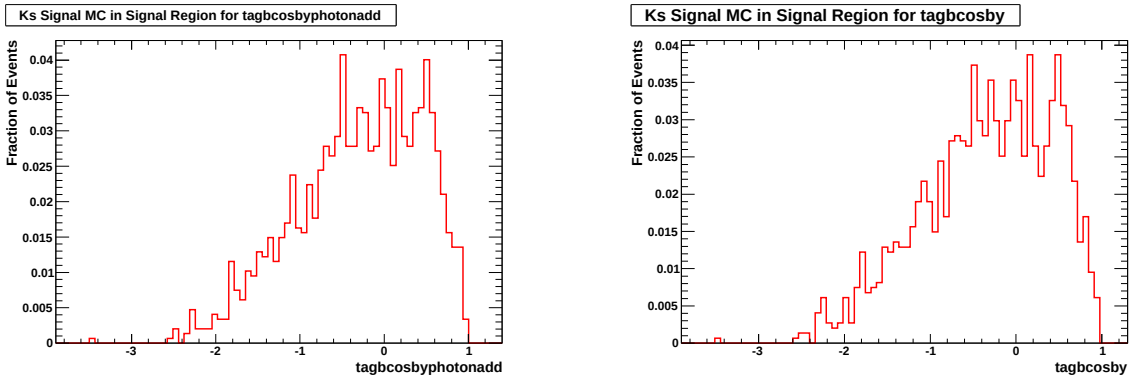

Figure 109: Signal MC histograms for tagbcosbyphotonadd on left and tagbcosby on right.

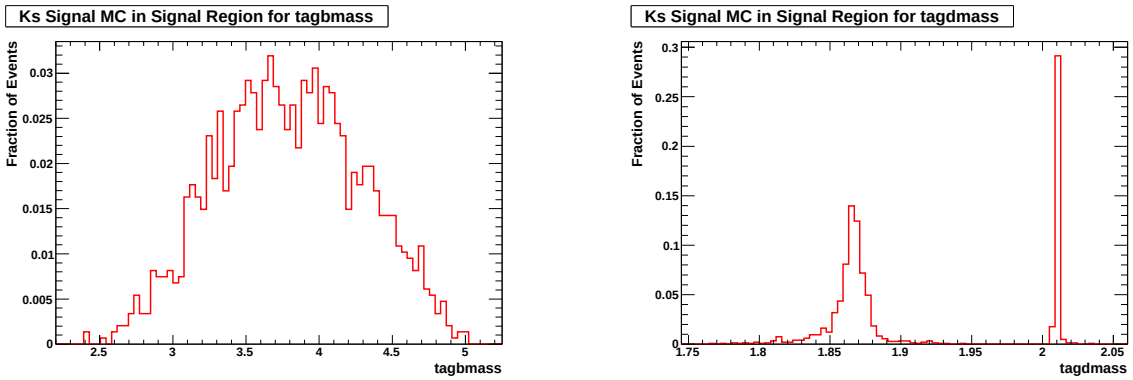

Figure 110: Signal MC histograms for tagbmass on left and tagdmass on right.

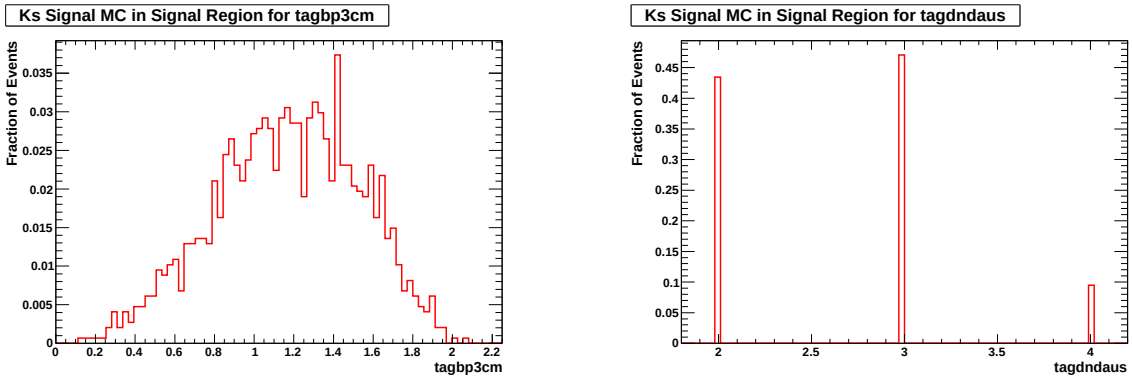

Figure 111: Signal MC histograms for tagbp3cm on left and tagdndaus on right.

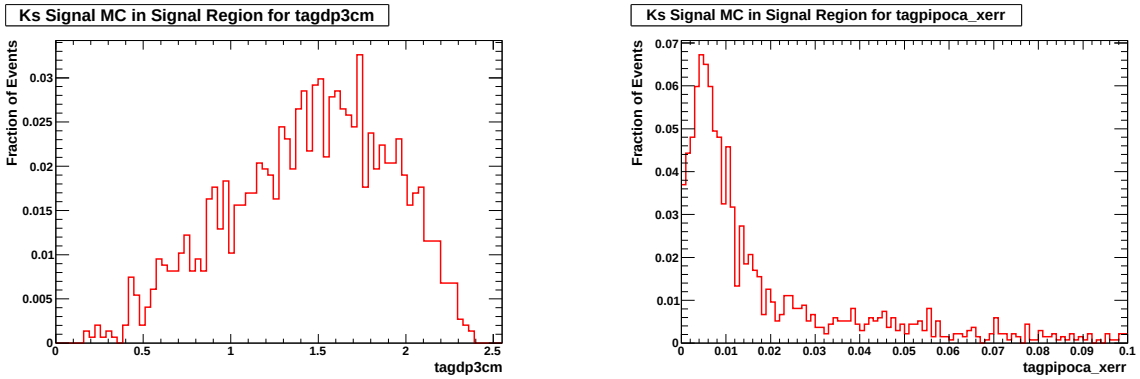

Figure 112: Signal MC histograms for tagdp3cm on left and tagpipoca_xerr on right.

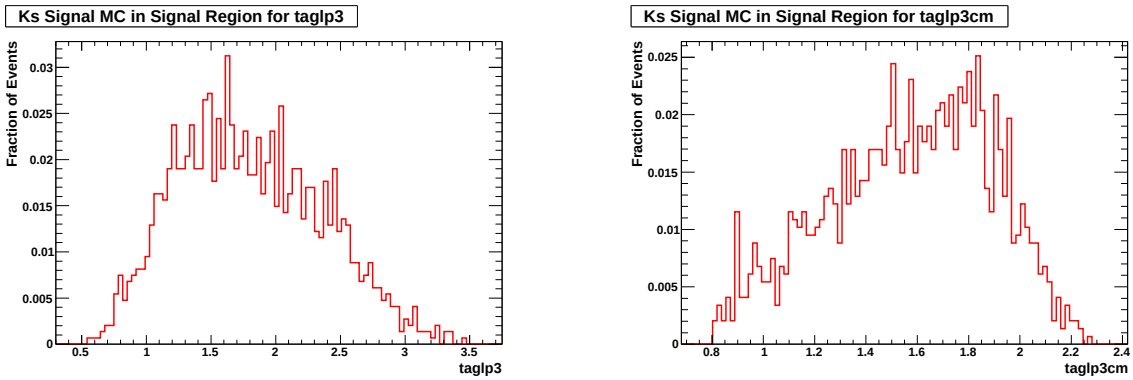

Figure 113: Signal MC histograms for taglp3 on left and taglp3cm on right.

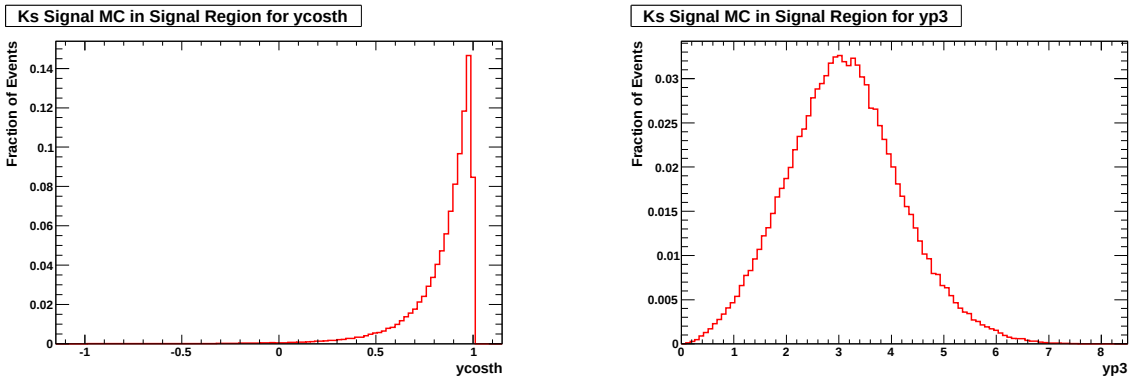

Figure 114: Signal MC histograms for ycosth on left and yp3 on right.

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 5.2 and 5.3, and the histograms are unit-normalized.

A.3.1 K^+ mode

The background MC events in these plots are weighted as described in Section 6.2.1.

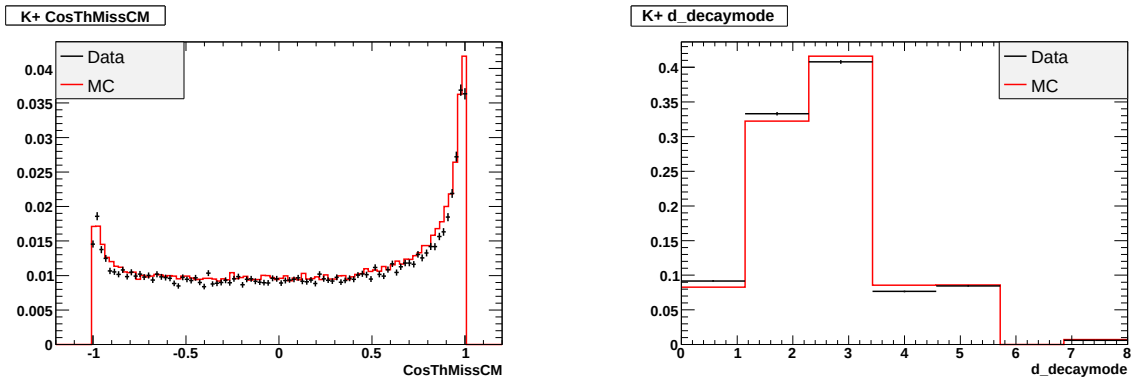

Figure 115: Data/MC histograms for CosThMissCM on left and d_decaymode on right.

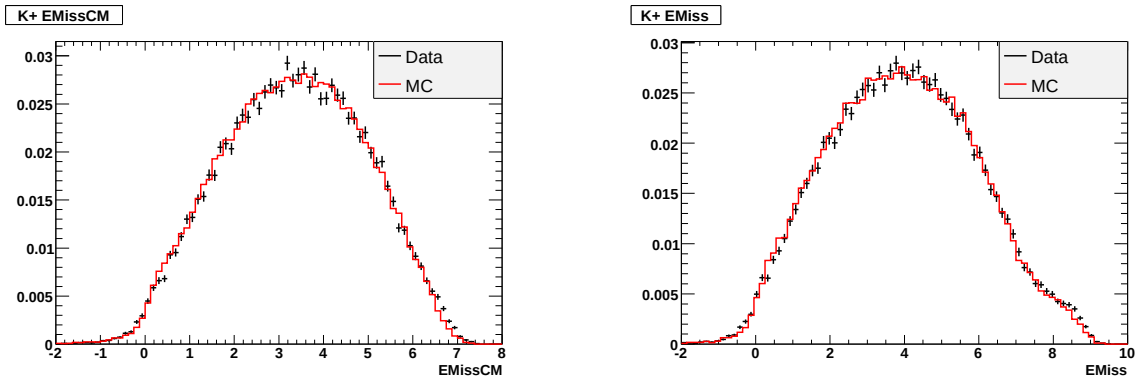

Figure 116: Data/MC histograms for EMissCM on left and EMiss on right.

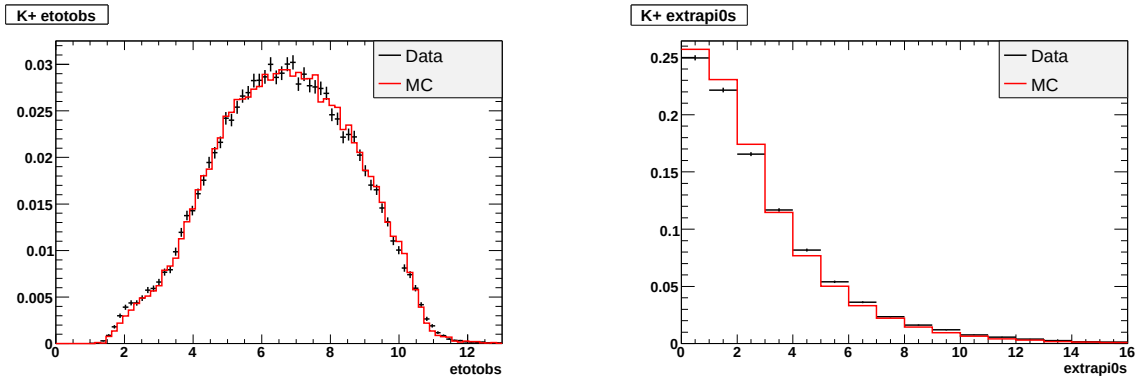

Figure 117: Data/MC histograms for etotobs on left and extrapi0s on right.

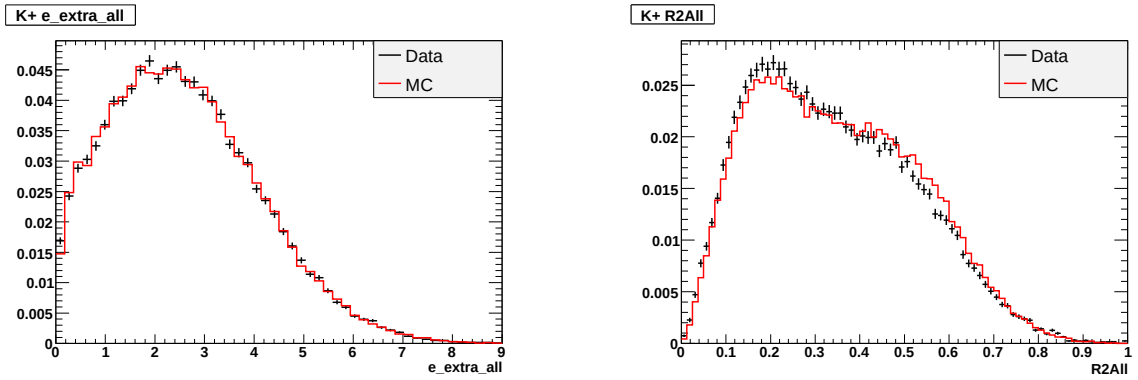

Figure 118: Data/MC histograms for e_extra_all on left and R2All on right.

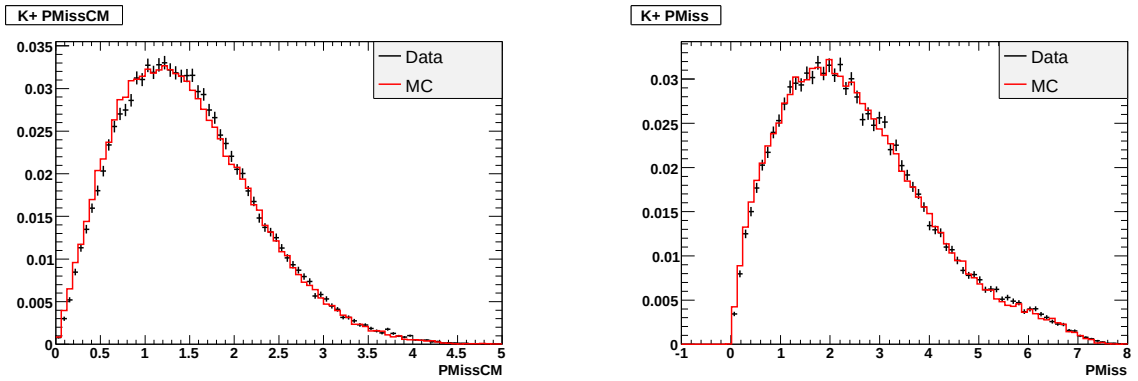

Figure 119: Data/MC histograms for PMissCM on left and PMiss on right.

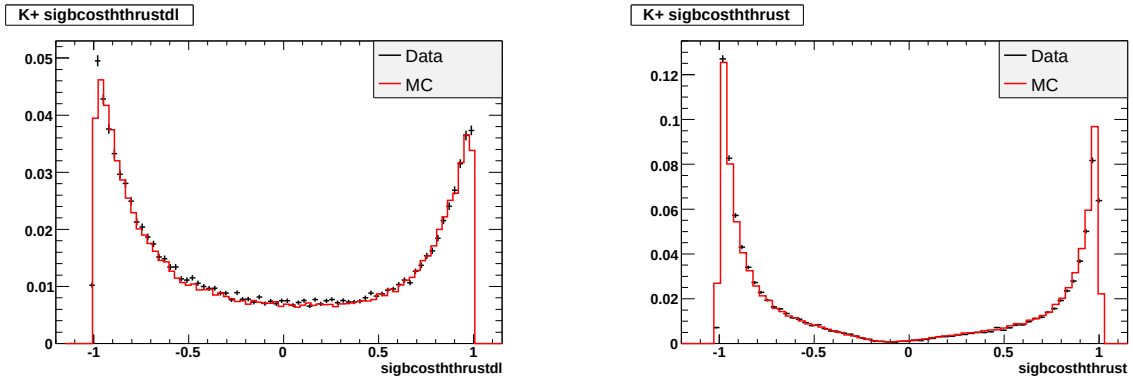

Figure 120: Data/MC histograms for sigbcostthrustdl on left and sigbcostthrust on right.

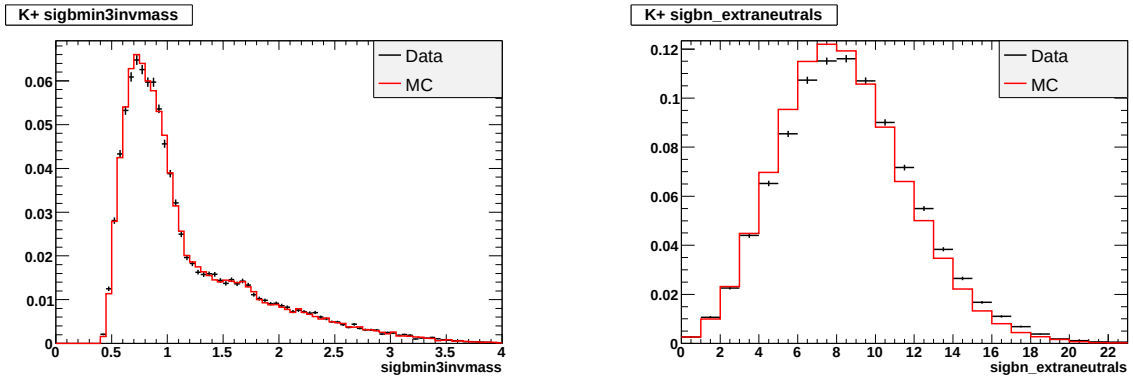

Figure 121: Data/MC histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.

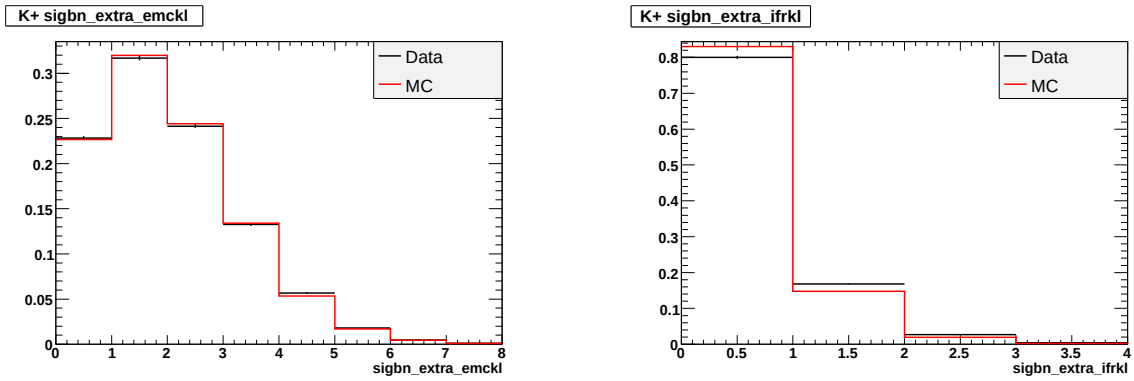

Figure 122: Data/MC histograms for sigbn_extra_emckl on left and sigbn_extra_ifrkl on right.

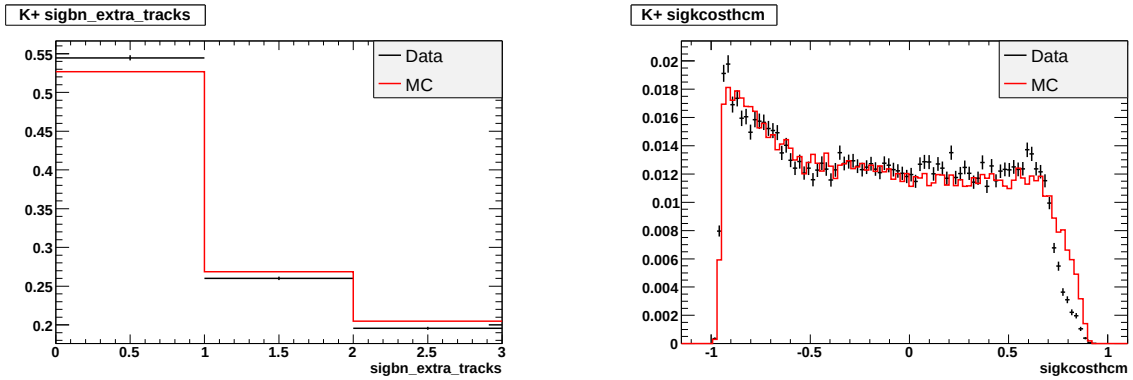

Figure 123: Data/MC histograms for sigbn_extra_tracks on left and sigkcosthcm on right.

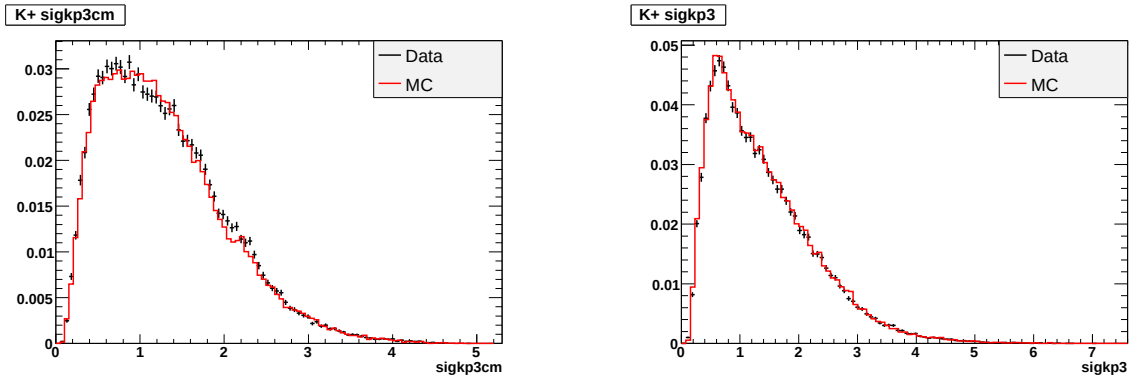

Figure 124: Data/MC histograms for sigkp3cm on left and sigkp3 on right.

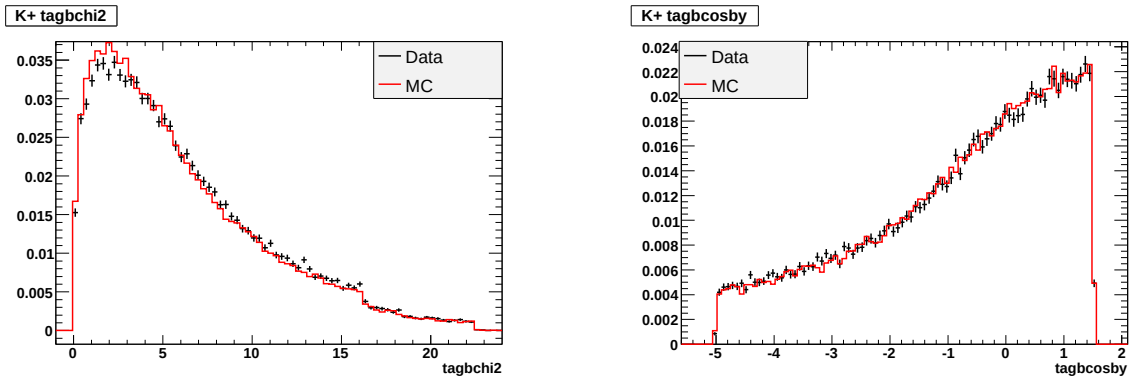

Figure 125: Data/MC histograms for tagbchi2 on left and tagbcosby on right.

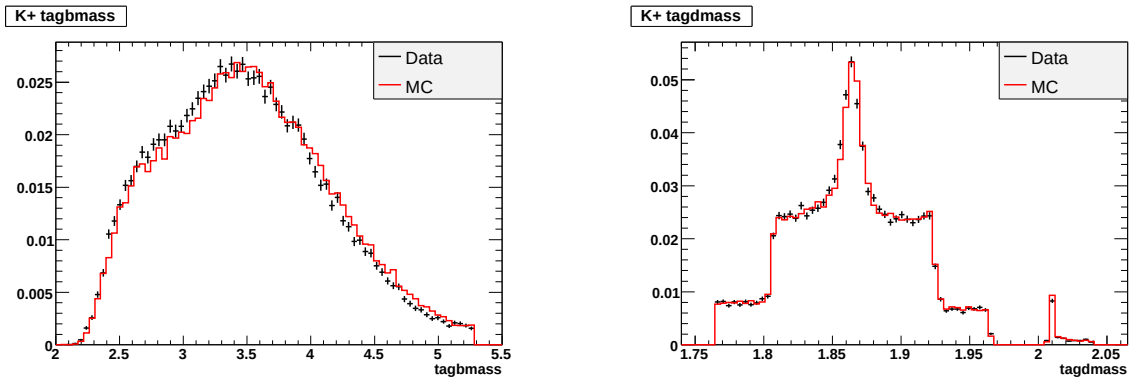

Figure 126: Data/MC histograms for tagbmass on left and tagdmass on right.

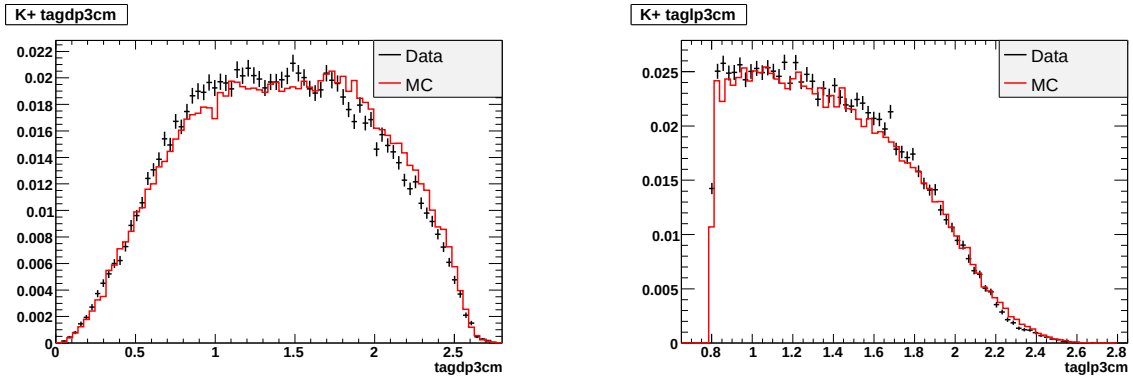

Figure 127: Data/MC histograms for tagdp3cm on left and taglp3cm on right.

A.3.2 K_S mode

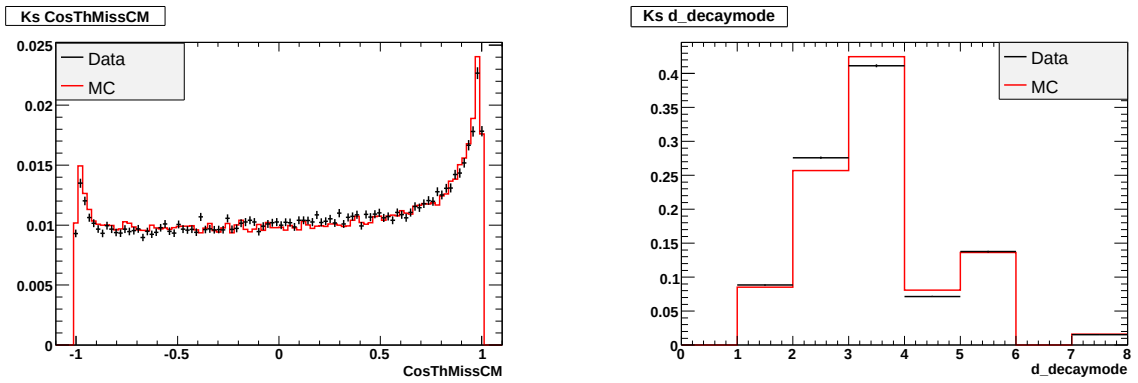

Figure 128: Data/MC histograms for CosThMissCM on left and d_decaymode on right.

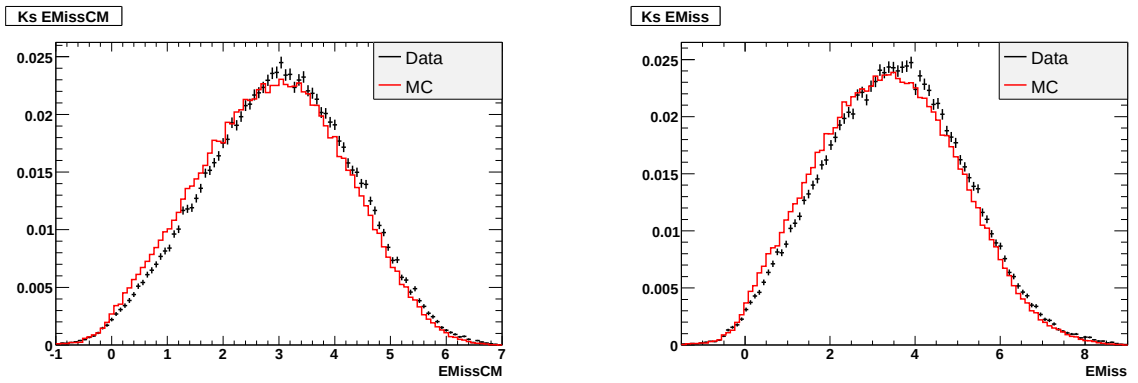

Figure 129: Data/MC histograms for EMissCM on left and EMiss on right.

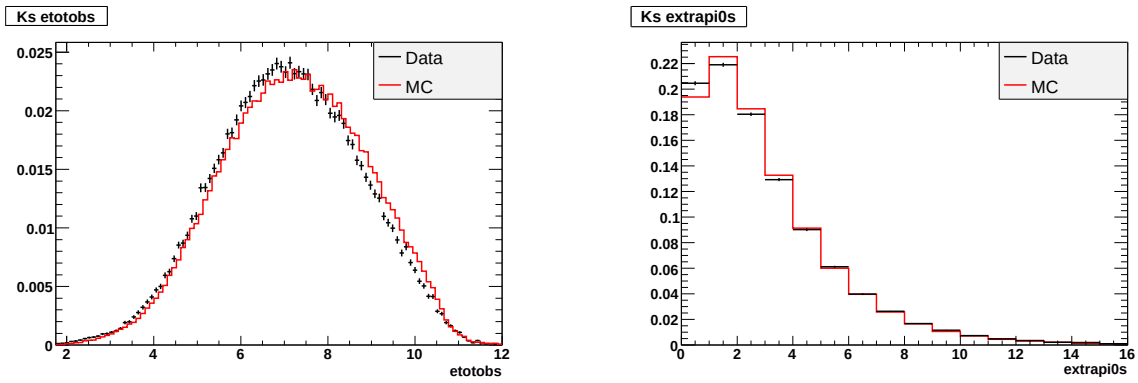

Figure 130: Data/MC histograms for etotobs on left and extrapi0s on right.

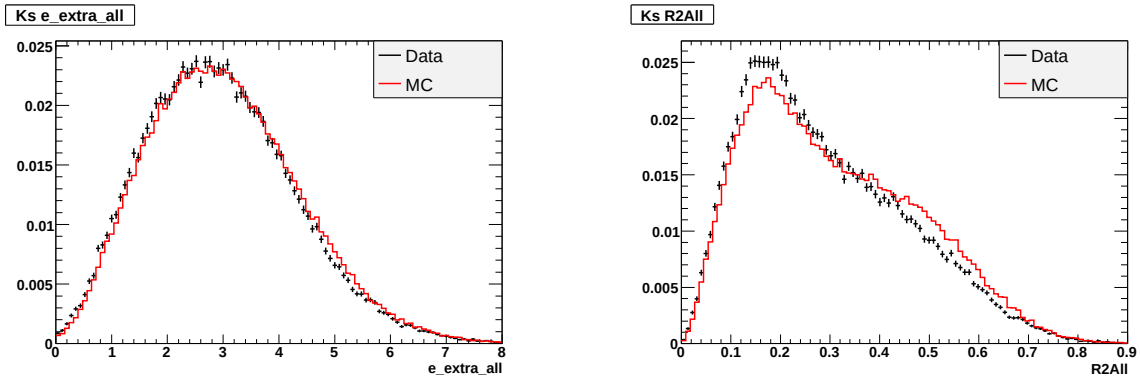

Figure 131: Data/MC histograms for e_extra_all on left and R2All on right.

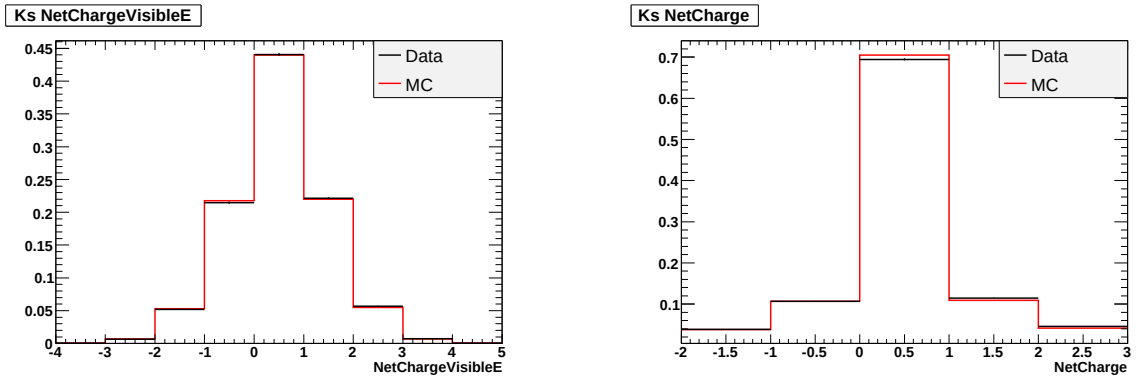

Figure 132: Data/MC histograms for NetChargeVisibleE on left and NetCharge on right.

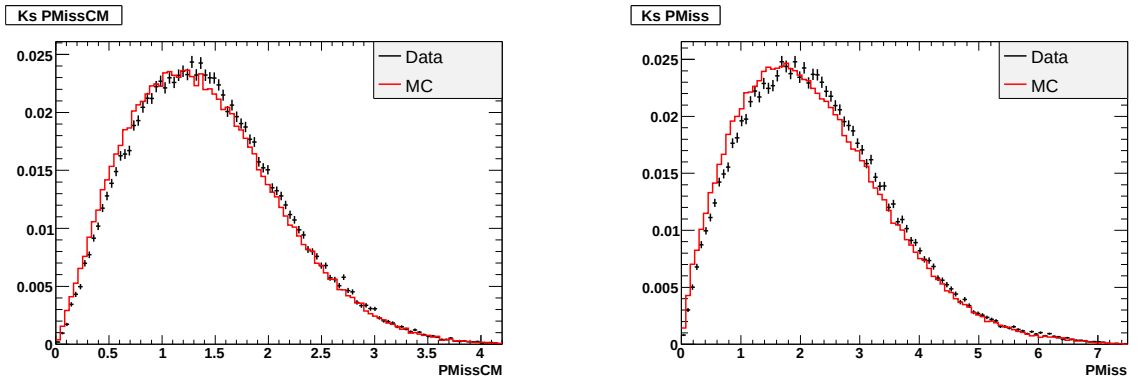

Figure 133: Data/MC histograms for PMissCM on left and PMiss on right.

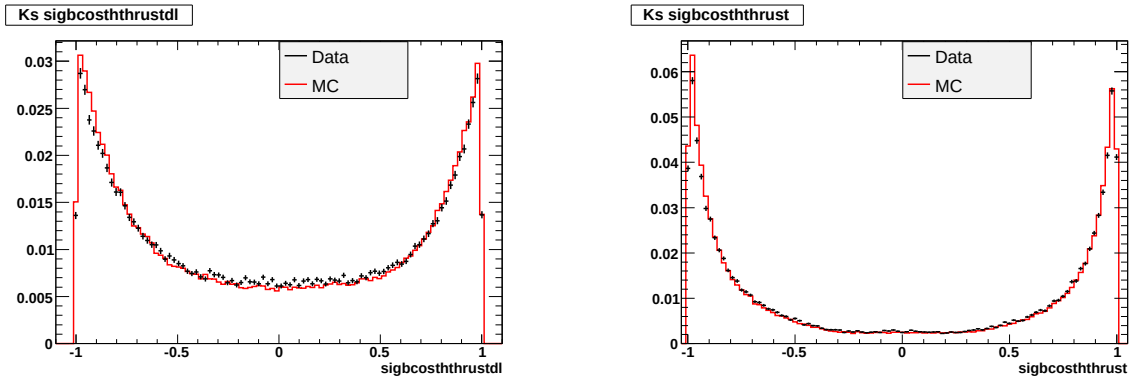

Figure 134: Data/MC histograms for sigbcostthrustdl on left and sigbcostthrust on right.

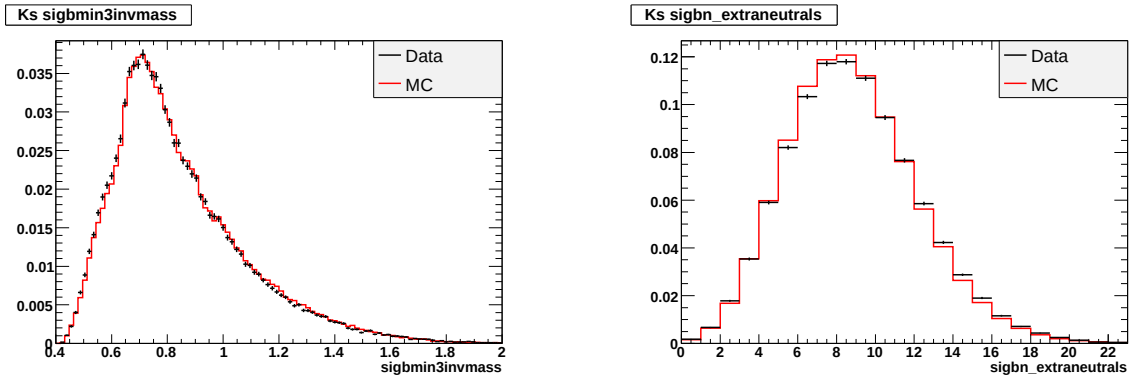

Figure 135: Data/MC histograms for sigbmin3invmass on left and sigbn_extraneutrals on right.

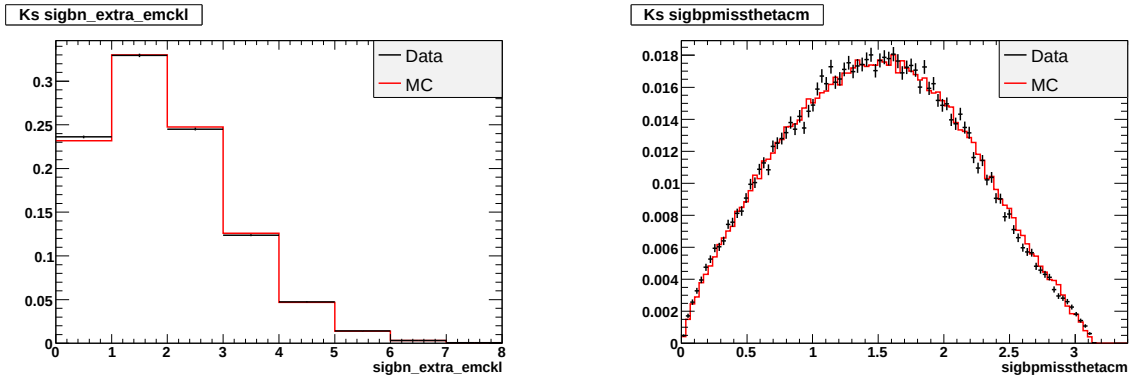

Figure 136: Data/MC histograms for sigbn_extra_emckl on left and sigbpmissthetacm on right.

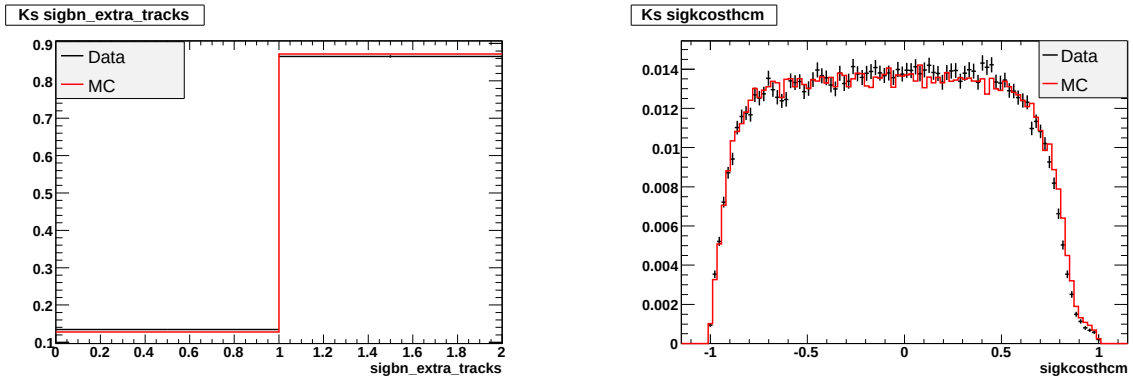

Figure 137: Data/MC histograms for sigbn_extra_tracks on left and sigkcosthcm on right.

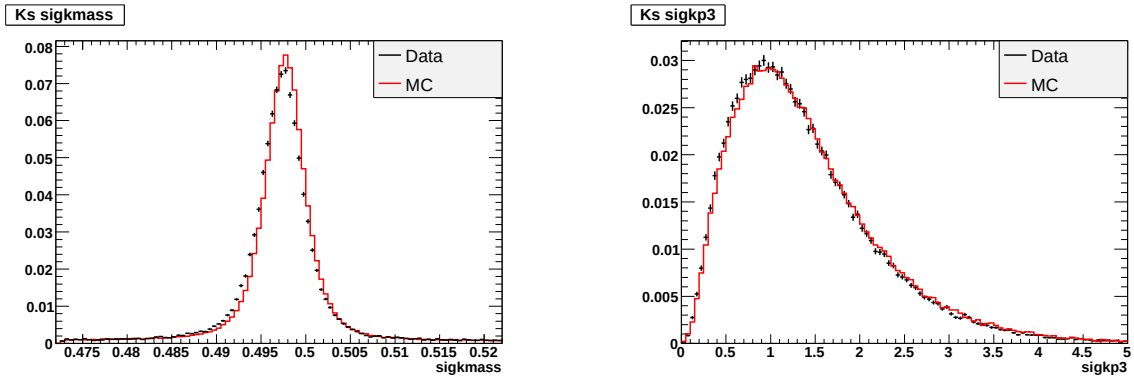

Figure 138: Data/MC histograms for sigkmass on left and sigkp3 on right.

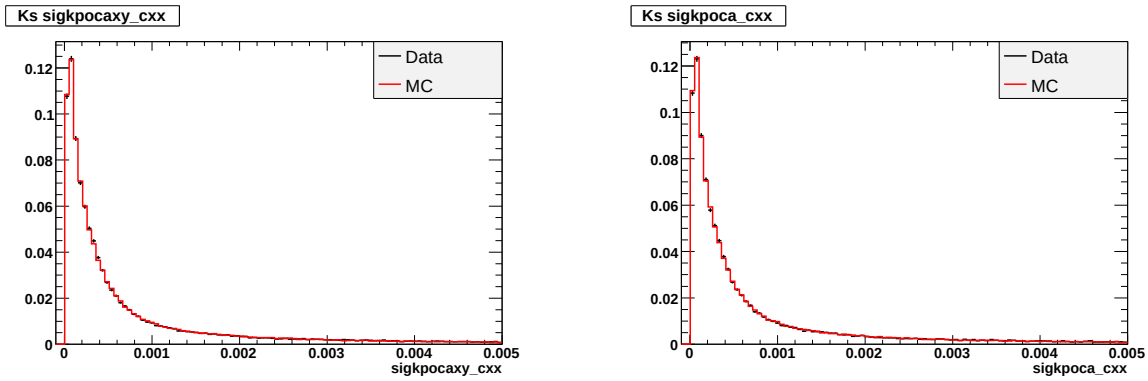

Figure 139: Data/MC histograms for sigkpocaxy_cxx on left and sigkpoca_cxx on right.

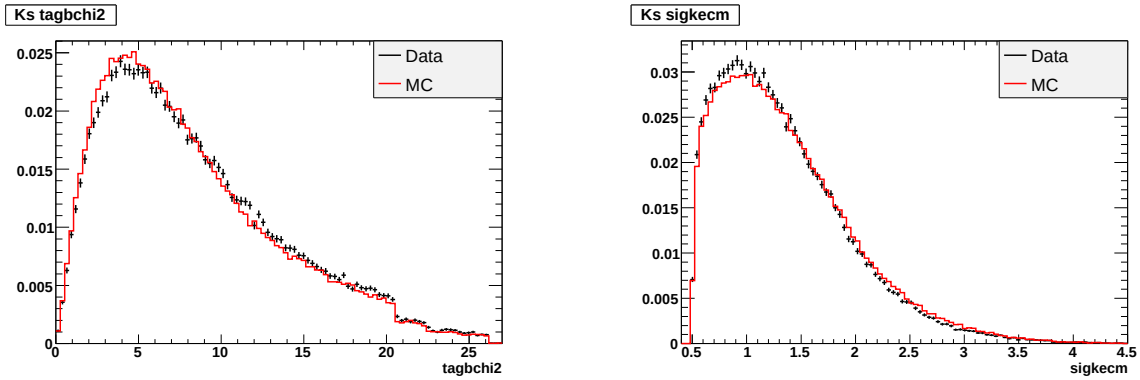

Figure 140: Data/MC histograms for tagbchi2 on left and sigkecm on right.

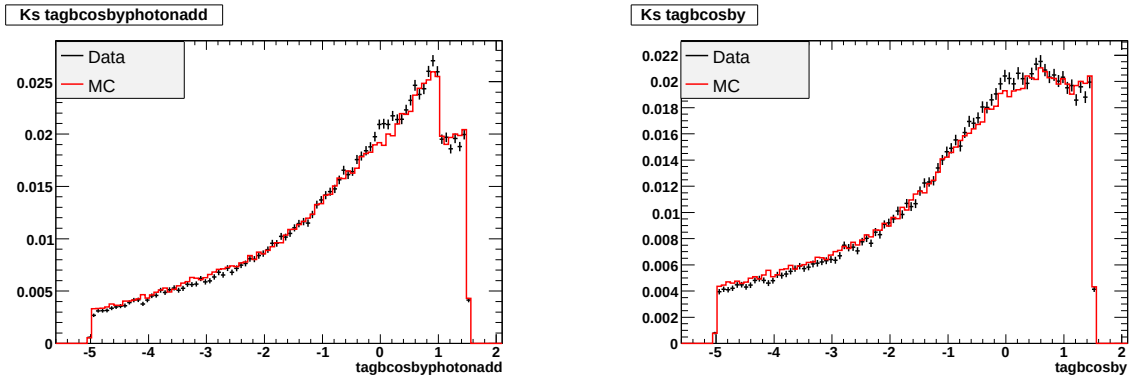

Figure 141: Data/MC histograms for tagbcosbyphotonadd on left and tagbcosby on right.

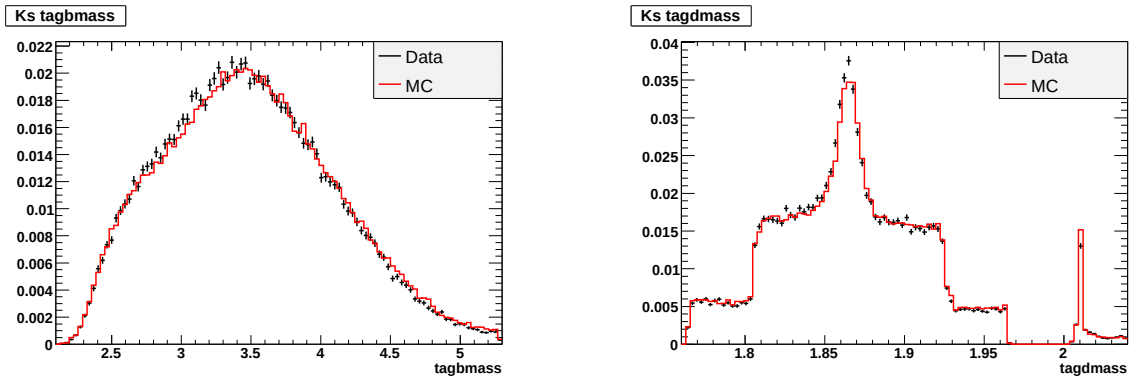

Figure 142: Data/MC histograms for tagbmass on left and tagdmass on right.

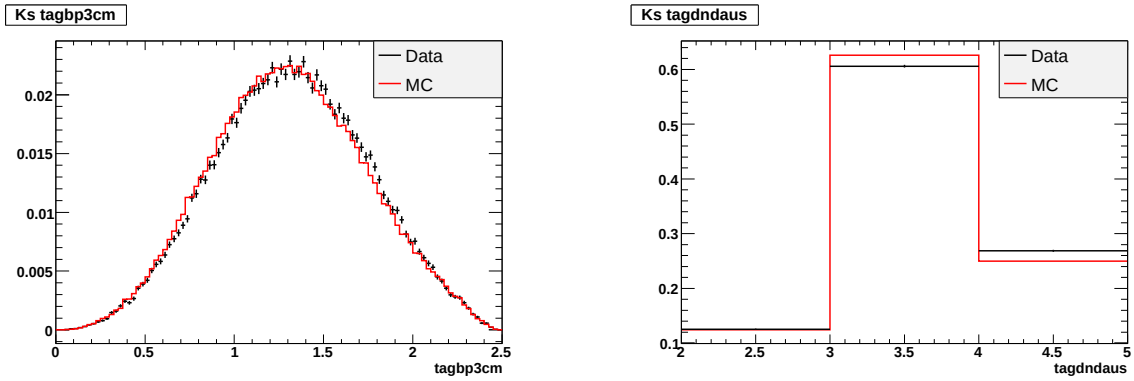

Figure 143: Data/MC histograms for tagbp3cm on left and tagdndaus on right.

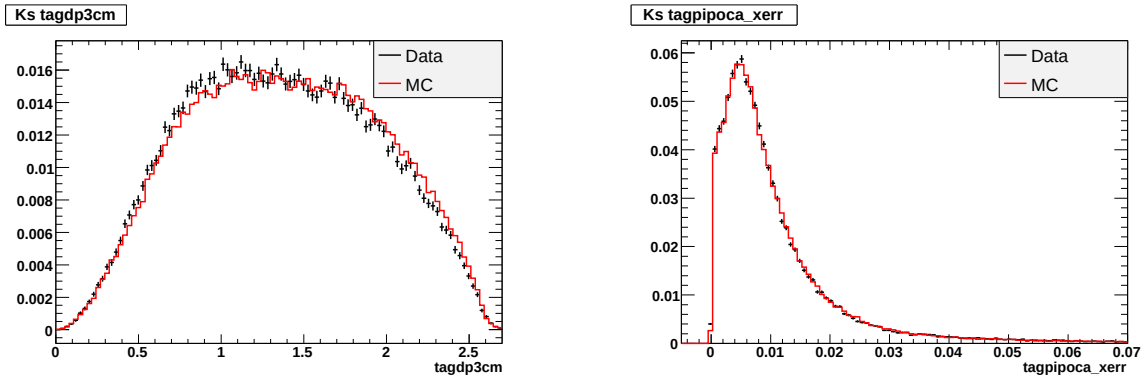

Figure 144: Data/MC histograms for tagdp3cm on left and tagpipoca_xerr on right.

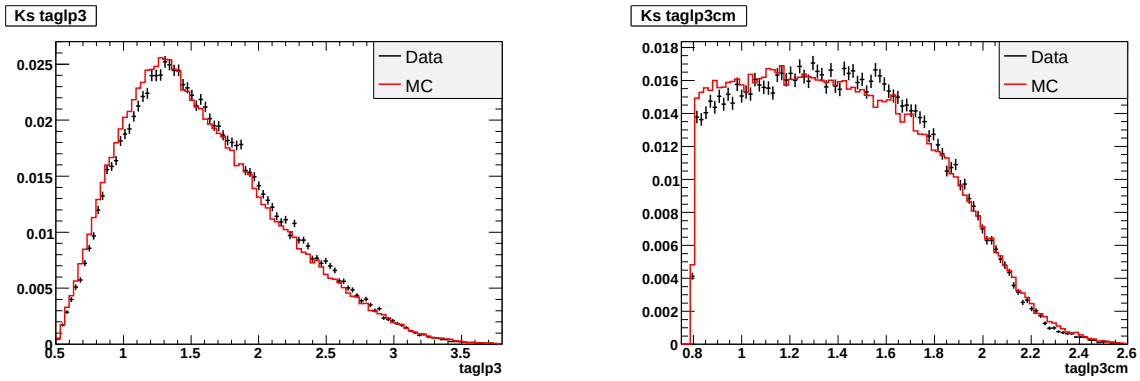

Figure 145: Data/MC histograms for taglp3 on left and taglp3cm on right.

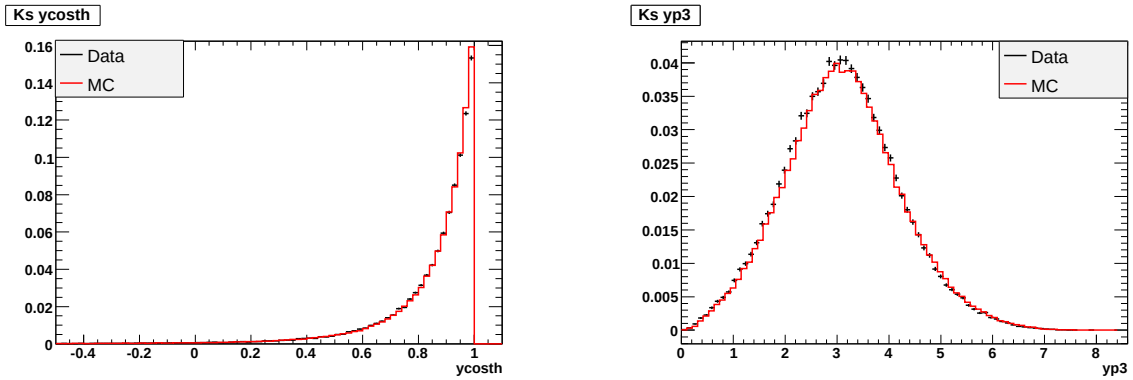

Figure 146: Data/MC histograms for ycosth on left and yp3 on right.

A.3.3 Extra Energy Detail Plots

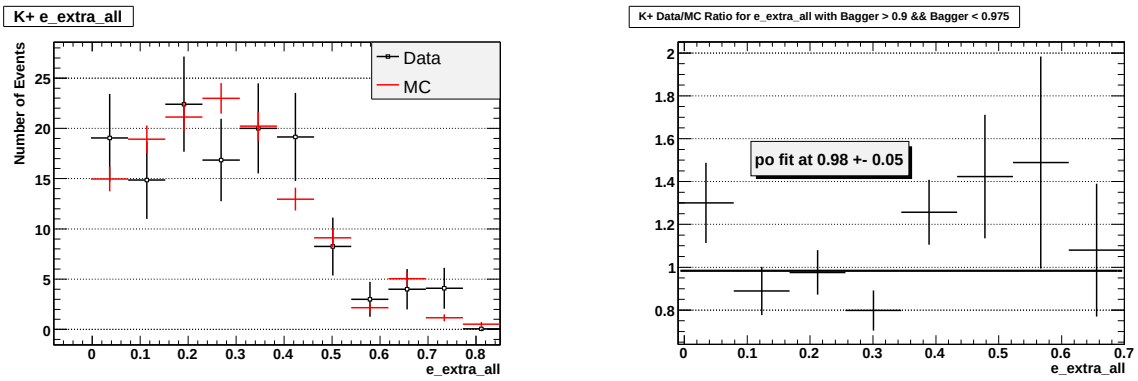

Figure 147: e_extra_all plots for data and generic background MC events with primary K^+ classifier output values from 0.9 to 0.975 (close to the signal region). The generic background MC events have the 2-D weighting. The left plot shows data and MC histograms, while the right plot shows the data/MC ratio.

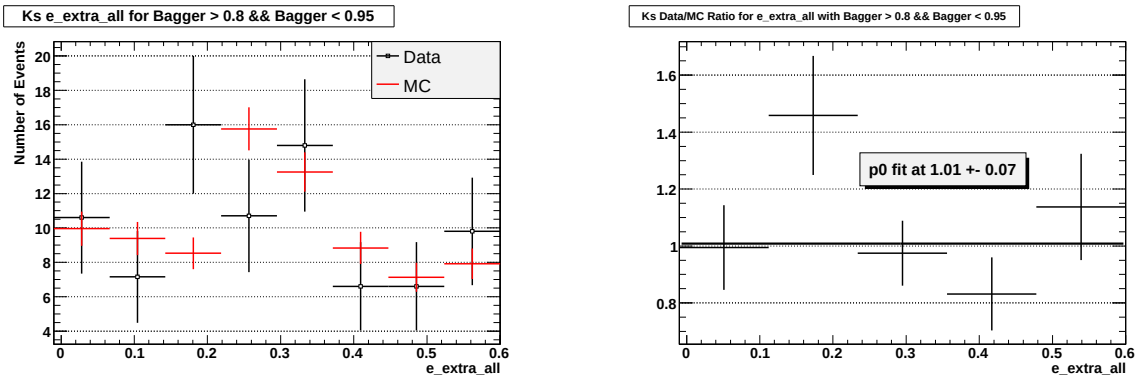

Figure 148: e_extra_all plots for data and generic background MC events with K_S classifier output values from 0.8 to 0.95 (close to the signal region). The generic background MC events are luminosity-weighted only. The left plot shows data and MC histograms, while the right plot shows the data/MC ratio.

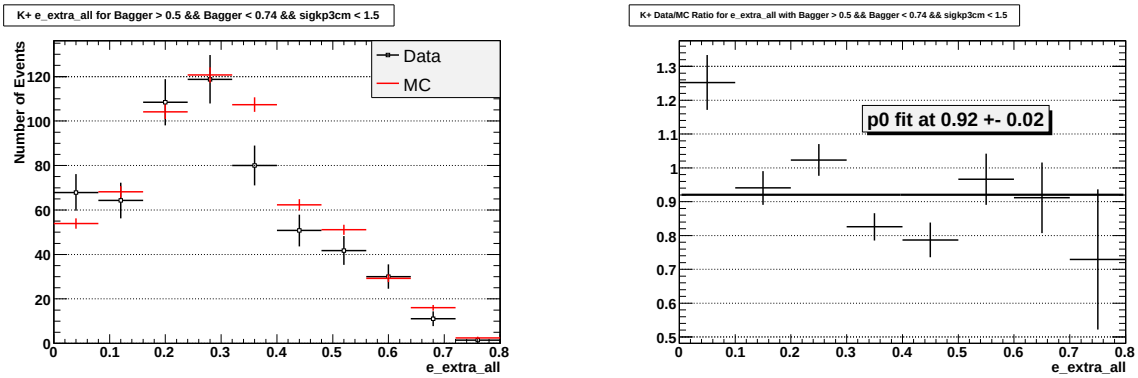

Figure 149: e_extra_all plots for low-kaon-momentum data and generic background MC events with classifier output values from 0.5 to 0.74 (close to the signal region). The generic background MC events have the 2-D weighting. The left plot shows data and MC histograms, while the right plot shows the data/MC ratio.

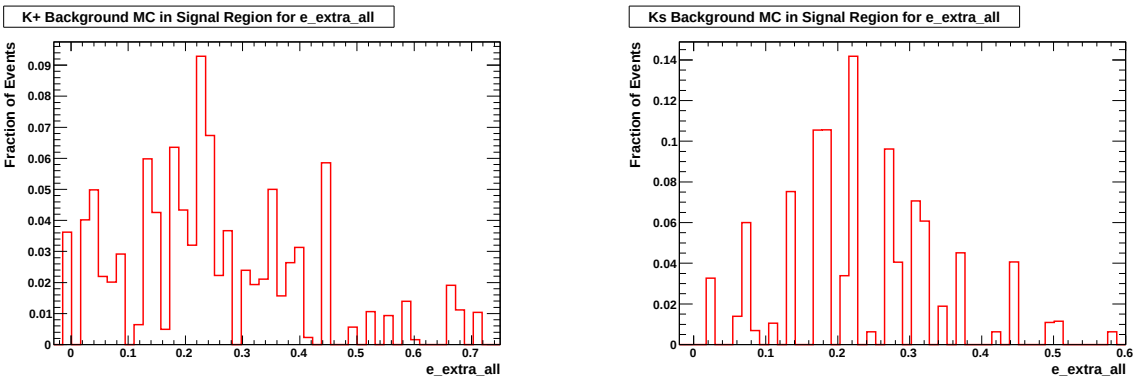

Figure 150: e_extra_all plots for background MC events in the signal region. The K^+ background MC events have the 2-D weighting. The left plot shows the K^+ background MC events, which have the 2-D weighting, while the right plot shows the K_S background.

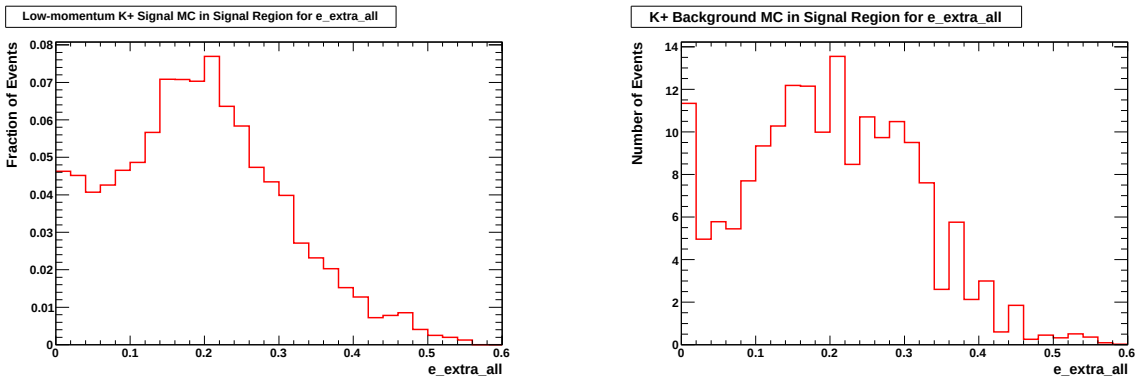

Figure 151: e_extra_all plots for MC events in the low-kaon-momentum signal region. The left plot shows signal MC, while the right one shows background MC, which has the 2-D weighting.

A.4 Data/MC Correspondence for Neutral Extra Energy

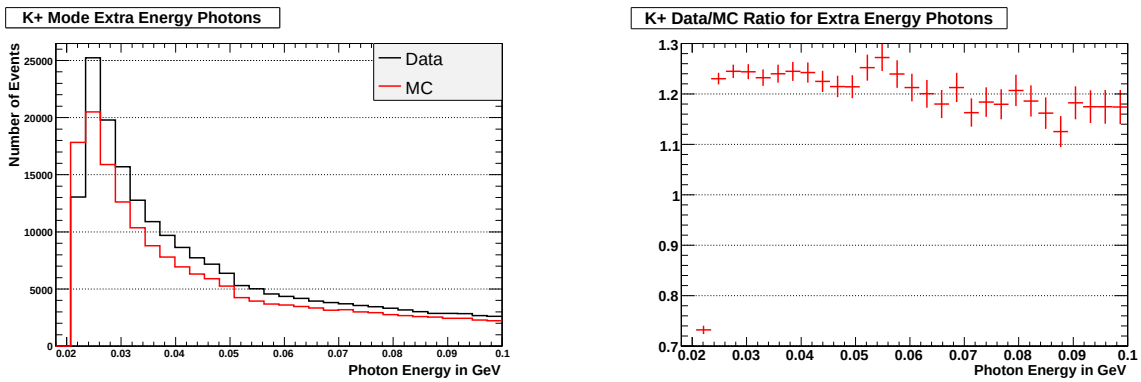

Figure 152: Correspondence between data and generic background MC for extra-energy photons for the K^+ ntuples. The MC is luminosity-weighted, but the 2-D weighting is not applied. The left plot shows data and MC histograms, and the right shows their ratio.

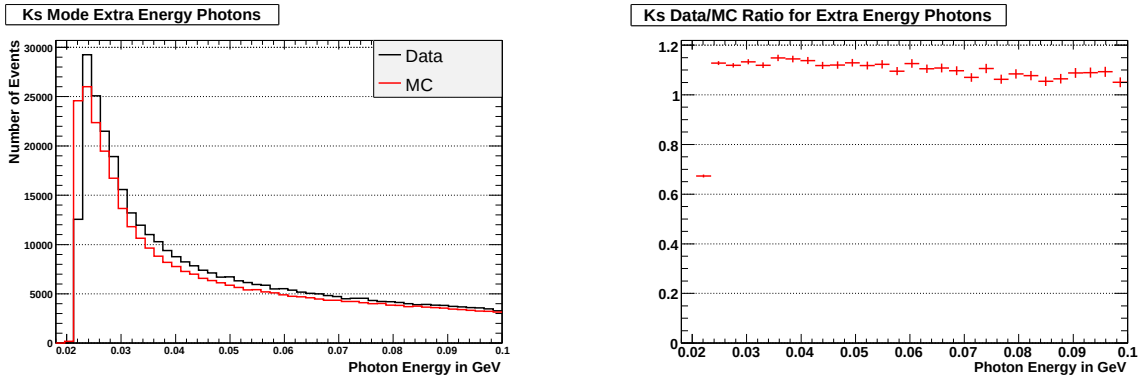

Figure 153: Correspondence between data and generic background MC for extra-energy photons for the K_S ntuples. The MC is luminosity-weighted, but the 10% correction is not applied. The left plot shows data and MC histograms, and the right shows their ratio.

A.5 Efficiency and $m^2_{\nu\bar{\nu}}$ Plots

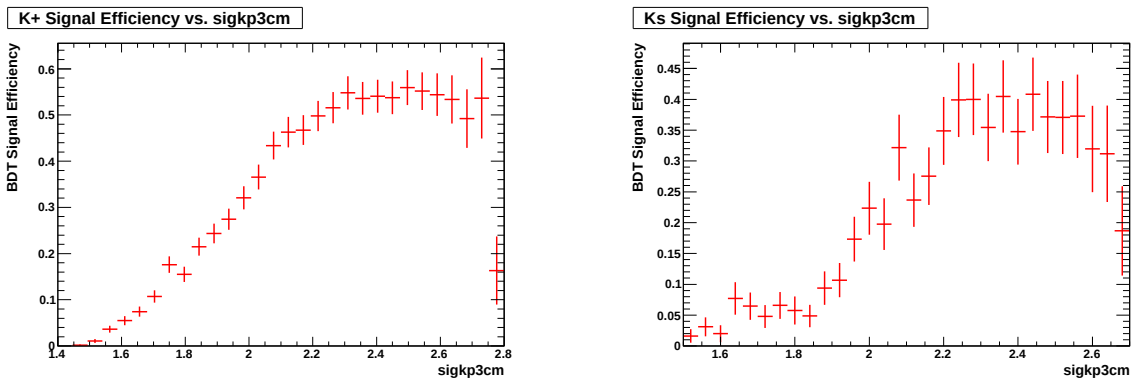

Figure 154: BDT signal efficiency vs. sigkp3cm for the primary K^+ mode on the left and the K_S mode on the right. This efficiency is for the classifier output cut alone, not any previous steps.

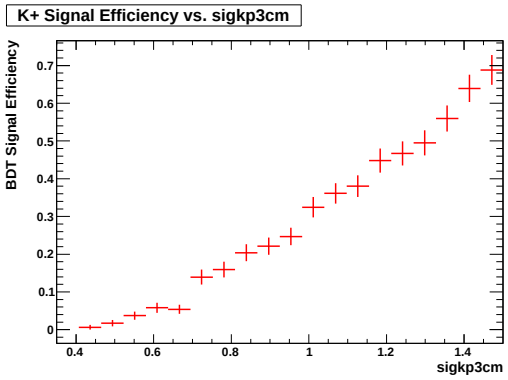

Figure 155: BDT signal efficiency vs. sigkp3cm for the low-momentum measurement. This efficiency is for the classifier output cut alone, not any previous steps.

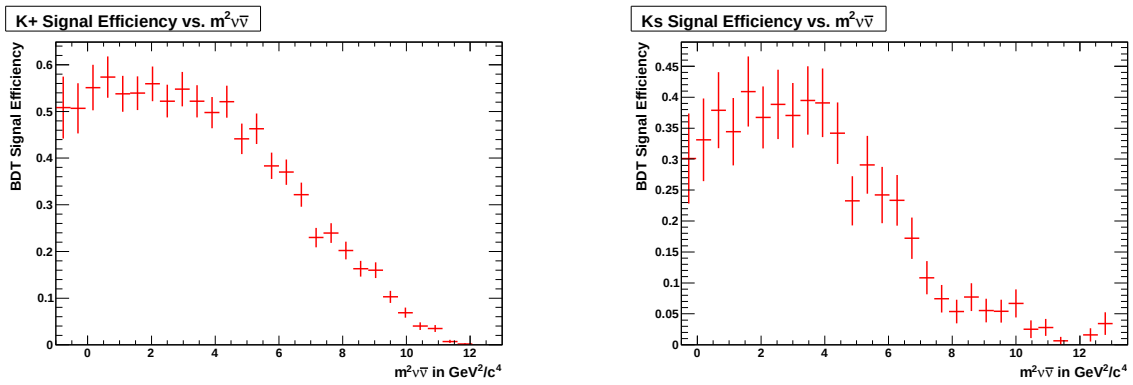

Figure 156: BDT signal efficiency vs. di-neutrino invariant mass squared for the primary K^+ mode on the left and the K_S mode on the right. This efficiency is for the classifier output cut alone, not any previous steps.

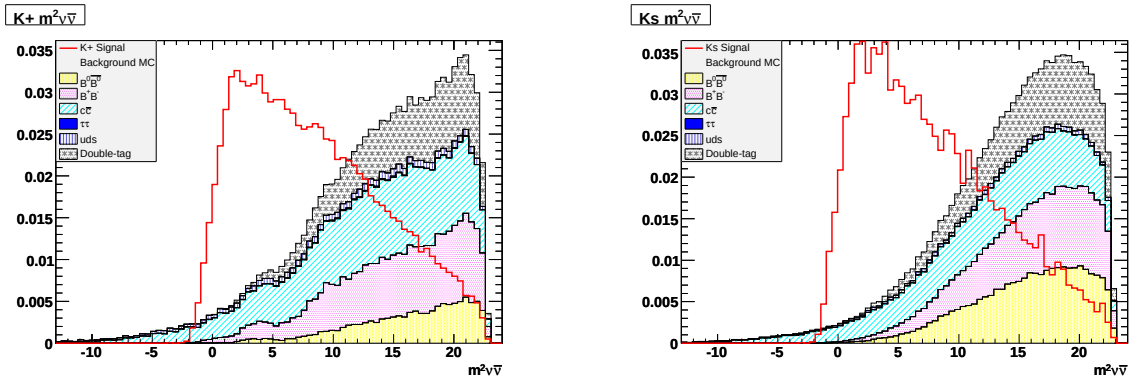

Figure 157: Di-neutrino invariant mass squared for the primary K^+ mode on the left and the K_S mode on the right. These unit-normalized plots include signal and generic background MC from the testing and training sets.

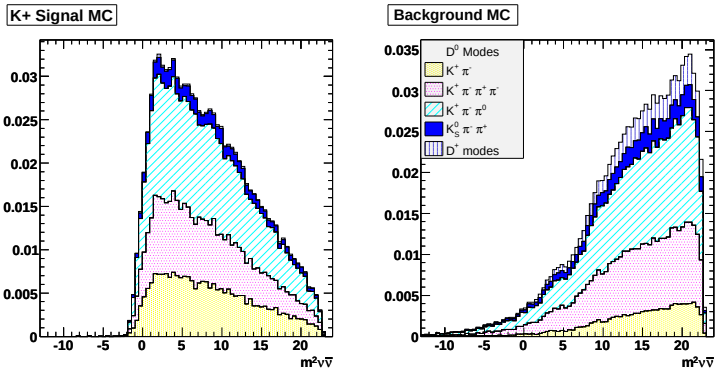

Figure 158: Di-neutrino invariant mass squared for the the primary K^+ mode. These unit-normalized plots include signal and generic background MC from the testing and training sets.

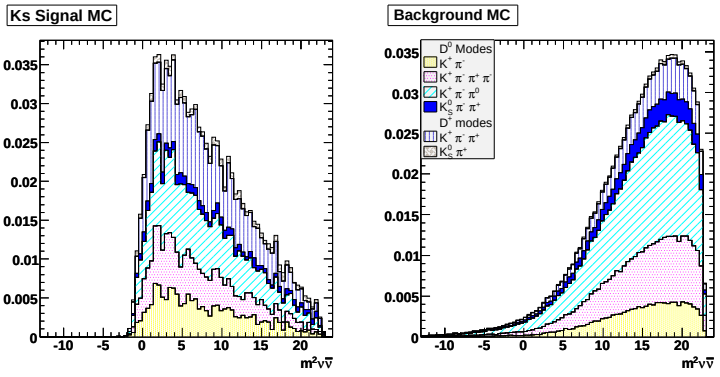

Figure 159: Di-neutrino invariant mass squared for the the K_S mode. These unit-normalized plots include signal and generic background MC from the testing and training sets.

A.6 Reconstruction Sample

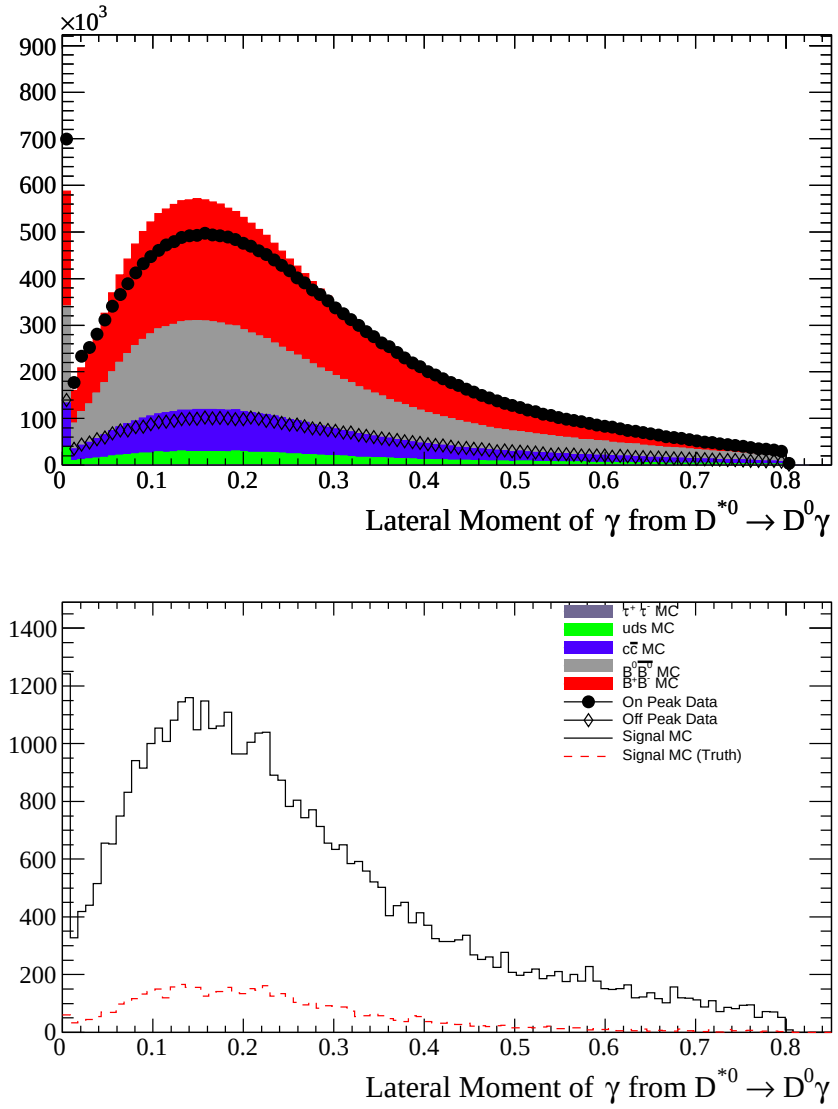

Figure 160: The lateral moment (LAT) distribution for the photon from the transition $D^* \to D\gamma$. Data and MC are shown for Runs 1-3 in release 18. The signal MC is for the process $B \to \tau\nu$. The top plot shows on- and off-resonance data overlaid on a stack of generic MC background; the bottom plot shows all reconstructed and reconstructed truth-matched signal MC.

A.7 Individual Tree Results

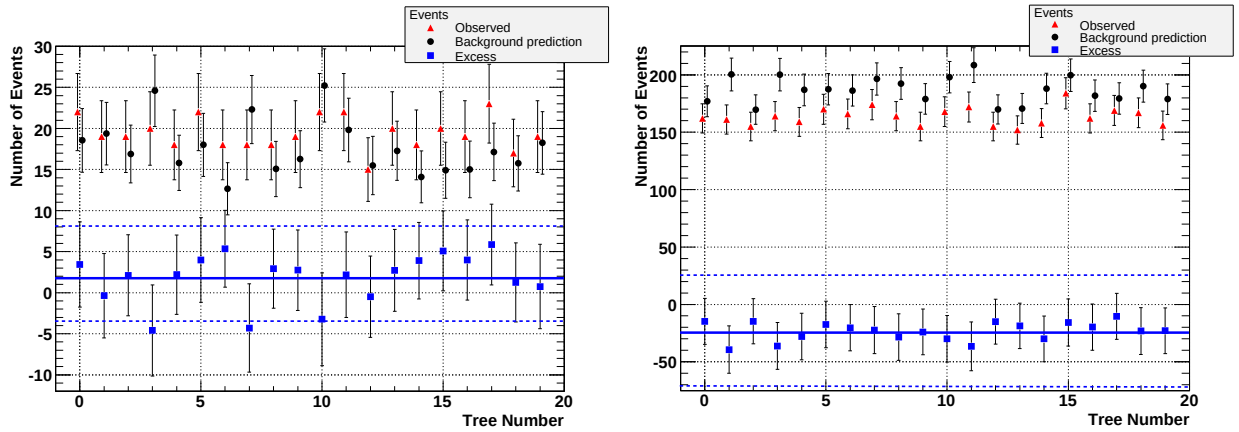

Figure 161: Individual tree results for the K^+ trees. The number of observed events (red triangles), the background prediction (black circles), and number of excess events (blue squares) are shown for each tree, numbered 0 to 19. The average number of excess events is shown by the solid blue line, with 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.

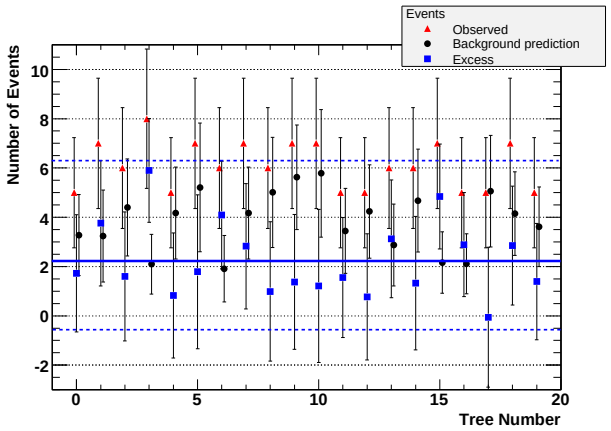

Figure 162: Individual tree results for the K_S trees. The number of observed events (red triangles), the background prediction (black circles), and number of excess events (blue squares) are shown for each tree, numbered 0 to 19. The average number of excess events is shown by the solid blue line, with the one-sigma uncertainty marked by the dotted blue lines.

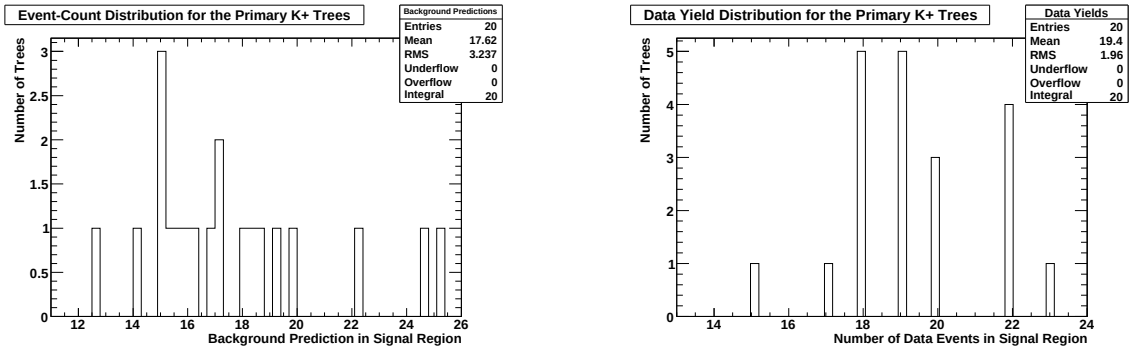

Figure 163: Individual tree results for the twenty K^+ trees. The left plot shows a histogram of the background predictions for the signal region for the twenty trees. The right shows a similar histogram of the actual data yields in the signal region.

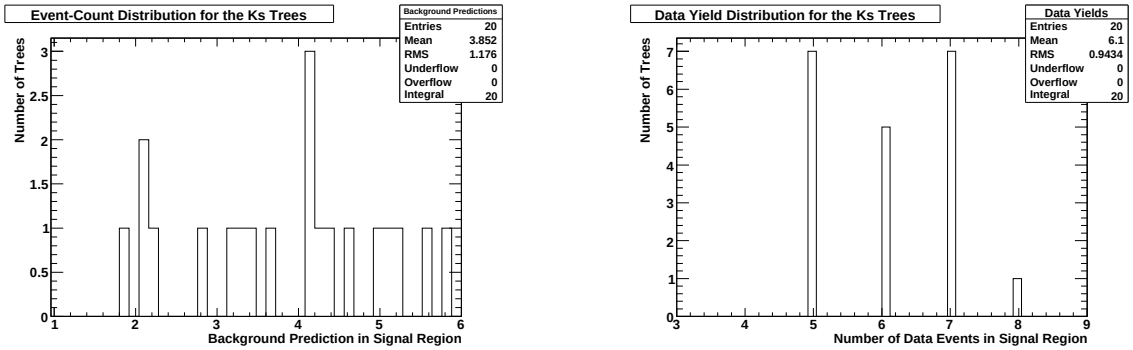

Figure 164: Individual tree results for the twenty K_S trees. The left plot shows a histogram of the background predictions for the signal region for the twenty trees. The right shows a similar histogram of the actual data yields in the signal region.

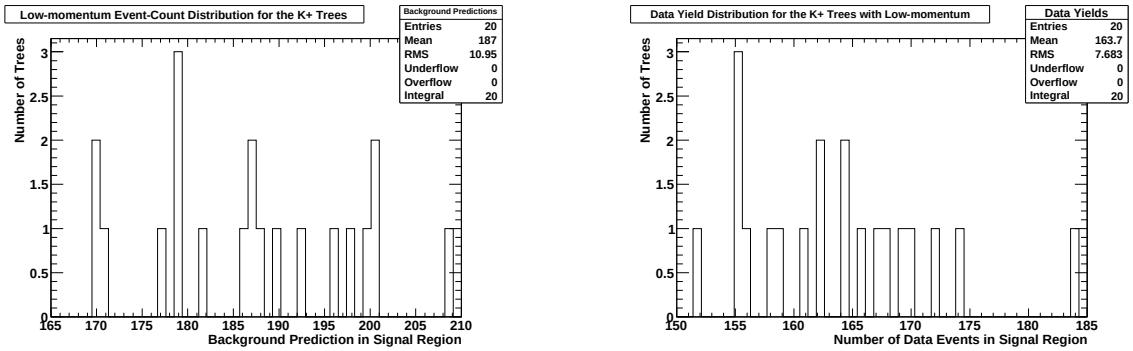

Figure 165: Individual tree results for the low-momentum measurement for the twenty K^+ trees. The left plot shows a histogram of the background predictions for the signal region for the twenty trees. The right shows a similar histogram of the actual data yields in the signal region.

B Additional Tables

Table 21: 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 22: Classifier output cut for the low-momentum mode on each of the twenty primary K^+ trees to achieve the target efficiency of 0.25% (under the ABSW model) 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 23: 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 information content to the fullest. The technique of bootstrap sampling [46] 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 24 [47]. This uncertainty is used for values obtained by averaging the response of twenty trees for the twenty MC testing sets.

$$\sigma_s = \sqrt{\frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma_i^2 + \rho\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sigma_i\sigma_j\right)} \tag{24}$$

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 uncertainties with correlation ρ. A full derivation can be found in [47].

References

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

[2] S. L. Glashow. Nucl. Phys. **22**, 579 (1961).

[3] S. Weinberg. Phys. Rev. Lett. **19**, 1264 (1967).

[4] A. Salam Originally printed in *Svartholm: Elementary Particle Theory, Proceedings Of The Nobel Symposium Held 1968 At Lerum, Sweden*, Stockholm 1968, 367-377.

[5] P. W. Higgs. Phys. Rev. Lett. **13**, 508 (1964).

[6] H. Abele *et al.* Eur. Phys. J. C **33**, 1 (2004).

[7] M. Kobayashi and T. Maskawa. Prog. Theor. Phys. **49**, 652 (1973).

[8] S. L. Glashow, J. Iliopoulos, and L. Maiani. Phys. Rev. D **2**, 7, 1285 (1970).

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

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

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

[12] C. Amsler *et al.* (Particle Data Group). Physics Letters B **667**, 1 (2008).

[13] B. Aubert *et al.* (BaBar Collaboration). Nucl. Instrum. Meth. A **479**, 1 (2002).

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

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

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

[17] 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`.

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

[19] E. Witten. Nuclear Physics B **122**, 109 (1977).

[20] K. G. Wilson and W. Zimmermann. Commun. Math. Phys. **24**, 87 (1972).

[21] P. Colangelo, F. D. Fazio, P. Santorelli, and E. Scrimieri. Physics Letters B **395**, 339 (1997).

[22] J. H. Jeon, C. S. Kim, J. Lee, and C. Yu. Phys. Lett. B **636**, 270 (2006).

[23] G. Buchalla, A. J. Buras, and M. E. Lautenbacher. Rev. Mod. Phys. **68**, 1125 (1996).

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

[25] T. M. Aliev and C. S. Kim. Phys. Rev. D **58**, 013003 (1998).

[26] P. Ball and R. Zwicky. Phys. Rev. D **71**, 014015 (2005).

[27] A. Ali, P. Ball, L. T. Handoko, and G. Hiller. Phys. Rev. D **61**, 074024 (2000).

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

[29] C. Bird, R. V. Kowalewski, and M. Pospelov. Mod. Phys. Lett. A **21**, 457 (2006).

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

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

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

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

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

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

[36] D. J. Lange. Nucl. Instrum. Meth. A **462**, 152 (2001).

[37] S. Agostinelli *et al.* (GEANT4 Collaboration). Nucl. Instrum. Meth. A **506**, 250 (2003).

[38] B. Aubert *et al.* (BaBar Collaboration). Phys. Rev. D **76**, 052002 (2007).

[39] B. Aubert *et al.* (BaBar Collaboration). Phys. Rev. D **79**, 091101 (2009).

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

[41] L. Breiman. Machine Learning **24**, 2, 123 (1996).

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

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

[44] O. Helene. Nucl. Instrum. Meth. A **390**, 383 (1997).

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

[46] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall (1993).

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