



A Monte Carlo study of the rare B -meson decay $B^\pm \rightarrow K^\pm \tau \tau$ at the Belle experiment.

Jo-Frederik Krohn, University of Hamburg, Germany

September 8, 2015

Abstract

An analysis in the intend to set an expected upper limit to the branching ratio of the rare B -meson decay into a K -meson and two τ leptons was performed to study flavor changing neutral currents $b \rightarrow sll$ on official Belle monte carlo data. An expected upper limit of $\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau) \leq 1.19 \cdot 10^{-3}$ was found. This limit was found using multivariate analysis methods to perform a counting experiment on simulated data.

Contents

1	Introduction	1
2	The Belle experiment	1
3	Decay Channels	1
4	Statistical methods	1
4.1	Upper Limit calculation	1
4.2	Branching Ratio	3
4.3	Boosted Decision Trees	3
4.4	Full reconstruction	4
5	Analysis	4
5.1	Event selection	5
5.2	Pre-cuts	5
5.3	Analysis of classifiers	5
5.4	Choice of variables to feed the classifier	6
5.5	Background contribution	7
5.6	Data MC comparison	7
5.7	Upper limit calculation	8
5.8	Systematics	8
6	Conclusion	8

1 Introduction

This is a monte carlo study of the expected branching ratio of the process $B^\pm \rightarrow K^\pm \tau \tau$ in the context of the DESY Summerschool. This analysis was performed on official Belle monte carlo data.

The decay is aimed at flavor changing neutral currents which are forbidden in the standard model on tree level. Yet they exist in higher order loop diagrams and are therefore strongly suppressed.

This decay in particular is interesting as deviations of 3.7σ from the Standard Model (SM) have been observed in an angular analysis of the decay of B mesons to a Kaon and two leptons $B \rightarrow K^* l l$ [1], a small anomaly has recently been found in the angular analysis of $B \rightarrow K^* \mu^+ \mu^-$ [2] and the analysis of the ratio $R_K = \mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-) / \mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$ as well showed a discrepancy of 2.6σ to SM calculations. According to [3] this might hint for physics beyond the SM. The decay $B^\pm \rightarrow K^\pm \tau \tau$ has not been deeply analyzed in this context and might therefore give new insight to the meaning of these deviations. A short investigation at the BaBar experiment resulted in a limit of $\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau) \leq 3.3 \cdot 10^{-3}$ [4]. Hence it is interesting to study the sensitivity of experiments like Belle and Belle 2 to such processes.

2 The Belle experiment

The Belle experiment is located in Tsukuba, Japan at KEK. It collected a total integrated Luminosity of 710 fb^{-1} while running on the $\Upsilon(4S)$ resonance. This corresponds to $771 \cdot 10^6$ produced $B\bar{B}$ pairs. The $\Upsilon(4S)$ resonance decays with $> 96\%$ chance into $B\bar{B}$ which is why experiments running on this resonance are called B -factories. These B -mesons can then be used for high precision measurements.

3 Decay Channels

There are different possibilities for $b \rightarrow s l l$, penguin diagrams and box diagrams. They are depicted in the Feynman Graphs in figure 1. The decay rate into the final state is sensitive to the existence of new gauge bosons mediating the decay, as for example the existence of charged Higgs bosons depicted in figure 1 (d).

4 Statistical methods

4.1 Upper Limit calculation

Upper limits are calculated in dependency of a certain degree of confidence. In the case of absent backgrounds they can be found from the equation:

$$1 - \alpha = \sum_{k=0}^n P(k; \mu) \quad (1)$$

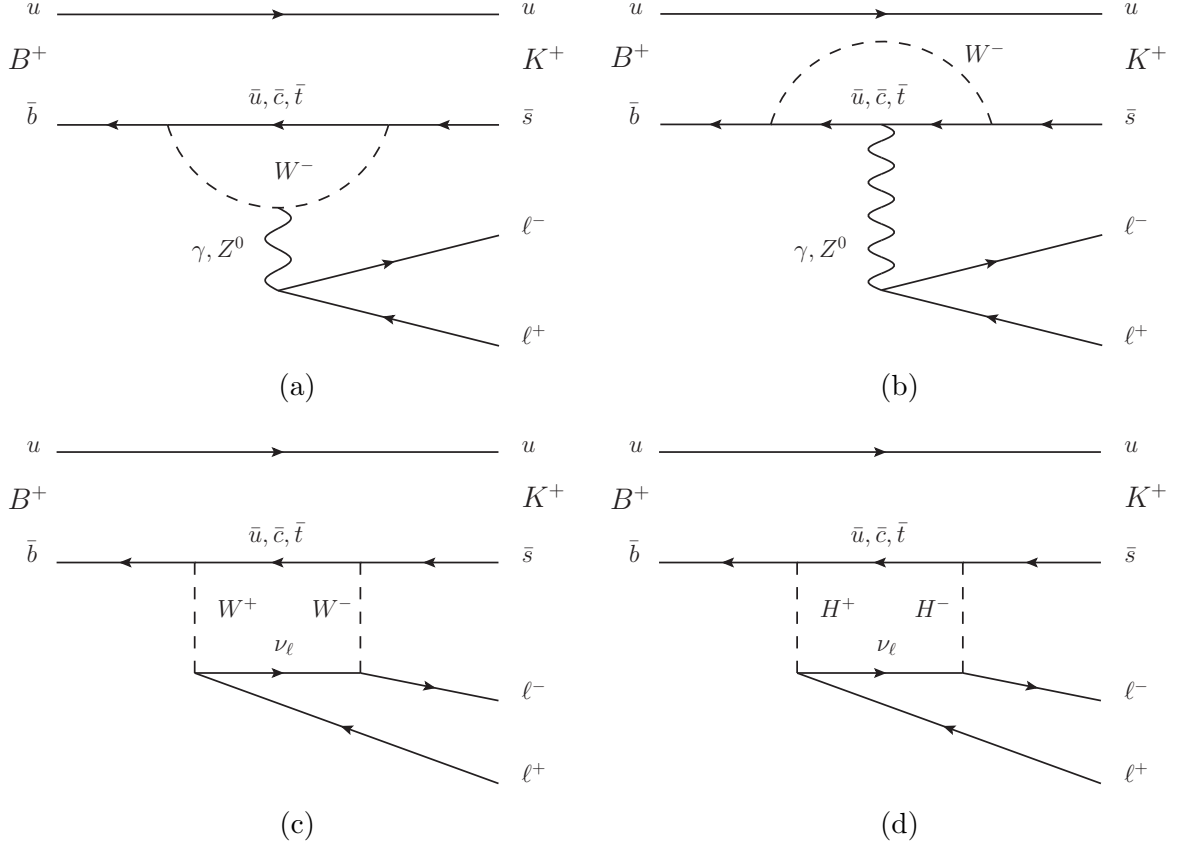


Figure 1: Penguin diagrams (a) and (b), and box diagrams (c) and (d) of the B-meson decay to two leptons and a K-meson in the final state. Diagram (a), (b) and (c) are possible from SM calculations, while (d) is only possible in beyond standard model models containing at least 2 charged Higgs-bosons.

Where α is the chosen confidence limit, $1 - \alpha = 90\%$ in our case, n is the number of observed events, μ is the unknown true expectation value and P is the underlying probability density function.

Thus for an expected $n = 0$ and an underlying Poisson distribution, a confidence limit of $CL = 90\%$ would yield an upper limit for $\mu \leq 2.3$. This means the true value of μ could still be ≤ 2.3 at $CL = 90\%$ confidence even though it was measured to be zero. In the case of present background the picture becomes more complicated. Downward fluctuations of the background could hide signals and upward fluctuation could mimic them. This is even more problematic as the expected number of signal events in our case is zero, while the expected background is large and gaussian. To handle this, the method suggested by W.Rolke [5] is used, as it is implemented in the ROOT class TRolke.

The likelihood function used to model a gaussian background and a poissonian signal is:

$$L(\mu, b, \sigma | x, y) = \frac{(\mu + b)^x}{x!} \cdot \exp(-(\mu + b)) \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(y - b)^2}{2\sigma^2}\right) \quad (2)$$

Here μ denotes the signal ratio, b the background ratio, σ the uncertainty on b and (x, y) are observed values for background and signal. The maximum likelihood estimators $(\hat{\mu}, \hat{b})$ can be calculated by maximizing $L(\mu, b, \sigma|x, y)$ with respect to μ and b . It is assumed that the log-likelihood distribution $-2 \log \lambda$ with :

$$\lambda(\mu|x, y) = \frac{L(\mu, \hat{b}(\mu)|x, y)}{L(\hat{\mu}, \hat{b}|x, y)} \quad (3)$$

follows a χ^2 -distribution. The limits can then be extracted by minimizing this function. Moving left and right with respect to the minimum of the resulting distribution, one can find the points where the function increases by the given α percentile. A confidence limit of $1 - \alpha = 90\%$ corresponds to an increase of this function by 2.706.

4.2 Branching Ratio

The branching ratio depends on the true number of decays $N_{true}^{B^\pm \rightarrow K^\pm \tau \tau}$ in the observed channel and the total number of decaying B-meson pairs N_{true}^{BB} by the relation:

$$\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau) = \frac{N_{true}^{B^\pm \rightarrow K^\pm \tau \tau}}{N_{true}^{BB}}. \quad (4)$$

Since $N_{true}^{B^\pm \rightarrow K^\pm \tau \tau}$ is intrinsically unknown it has to be replaced by the upper limit number on the observable events $N_{obs, UL}^{B^\pm \rightarrow K^\pm \tau \tau}$ and the efficiency ϵ of the classifier to gain an upper limit of B as:

$$\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau)^{UL} \leq \frac{N_{obs, UL}^{B^\pm \rightarrow K^\pm \tau \tau}}{\epsilon \cdot 2 \cdot f_+ \cdot N_{true}^{BB}}. \quad (5)$$

The efficiency depends on the classifier fitting and has to be calculated on MC by:

$$\epsilon = \frac{N_{obs}^{B^\pm \rightarrow K^\pm \tau \tau}}{N_{true}^{B^\pm \rightarrow K^\pm \tau \tau}}. \quad (6)$$

The factor $f_+ = 0.51$ is used to account for the number of charged B -Mesons from the $\Upsilon(4s)$ decay.

4.3 Boosted Decision Trees

Decision Trees are used to model a chain of multiple if-else decisions. A node in a tree resembles a certain decision if an input is member of a class A or B. Using several nodes a whole tree of decisions can be modeled. The classes of the lowest level of the graph are called leafs. These give the final decision if a processed data point is member of a certain class, embodied by the leaf it was associated with.

The decision tree is feed by a set of variables and evaluates their importance based on their entropy and possible information gain. The most important ones are then taken as nodes of the tree, where if-else conditions are used to navigate through the tree.

Boosted Decision Trees (BDT) are strong classifiers. They rely on an ensemble of weak classifiers. These give a vote if a data point is member of a certain class. By evaluation of the votes of a large number, typically 100, of weak classifiers, a strong classifier is constructed. The BDT uses a technique named boosting to generate these classifiers. A BDT is trained on a data set. After an initial fitting an evaluation of the discrimination power of the created classifier is done. Those data points which were wrongly classified gain an increased weight in the next training such that the next classifier created by an additional fitting has a higher focus on these points. This process is repeated until a maximum number of n iterations is reached and n classifiers are created. The final classification of a data point is then done as a weighted vote by these n weak classifiers and is normalized to an output $x \in [0, 1]$, such that x can be interpreted as a degree of belief that a certain object is member of a certain class.

Standard decision trees are very susceptible for overfitting as their number of degrees of freedom (= the depth of the tree) is generally arbitrary. BDT are less susceptible for overfitting because they rely on weak classifiers who themselves have a small number of degrees of freedom and a fixed depth of usually approximately 4 layers.

4.4 Full reconstruction

In order to probe invisible final states like for example those containing neutrinos, a full reconstruction of one of the B -mesons is performed. This was done using a neuron network - NeuroBayes - to classify candidates events.

Since the center of mass energy of the event corresponds to the mass of the $\Upsilon(4s)$ resonance, the initial state is known very well and it is easy to reconstruct the momentum of the B -meson, once one of the B -mesons is identified. The efficiency of B -tagging is around 0.2%. NeuroBayes gives an output $x \in [0, 1]$ such that it can be interpreted as a degree of belief in the probability of correct tagging. By cutting on this output one can increase the purity of the candidate sample. In the case of two identified B -mesons the one with the higher NeuroBayes output is chosen as B_{tag} .

5 Analysis

This analysis is a counting experiment on monte carlo (MC) data with the aim to set an expected upper limit on the branching ratio of the decay $B^\pm \rightarrow K^\pm \tau \tau$. Multivariate classifiers are used to distinguish between signal and background. A gradient boosted decision tree (BDT) was found to perform best. It is trained on MC signal and background samples and then used to classify an independent sample from which the upper limit is calculated. The signal is reconstructed in the "ExtraEnergyECL" variable, the residual energy in the calorimeter after removing all clusters from tag and signal particles. The systematics have not been treated.

5.1 Event selection

The signal event selection is done using the following steps. The first step in is to identify the B_{tag} -meson. Once it is identified, all tracks related to it are removed from the event. The remaining tracks then must be related to the signal B -meson. There are two cases that have to be distinguished. The B_{tag} can be a B^\pm or a B^0 . We only consider the case of B^\pm . In this case we require a K^+ -track and two charged tracks. The K^+ -track is identified by cutting on the likelihood-hypothesis to this being a π^+ . The τ -tracks then are found from demanding two oppositely charged tracks in the detector and no more tracks or entries in the calorimeter. The same holds for the charge conjugated case.

5.2 Pre-cuts

Soft precuts have been applied to "BTagNBout" and "BTagMbc" before fitting the BDT. The precuts are used to reduce the amount of background and to increase the computation speed of the fitting. "BTagNBout" was cut such that $BTagNBout > 10^{-4}$ and $BTagMbc > 5.27 [GeV]$. The applied cuts are depicted in figure 2 and 3.

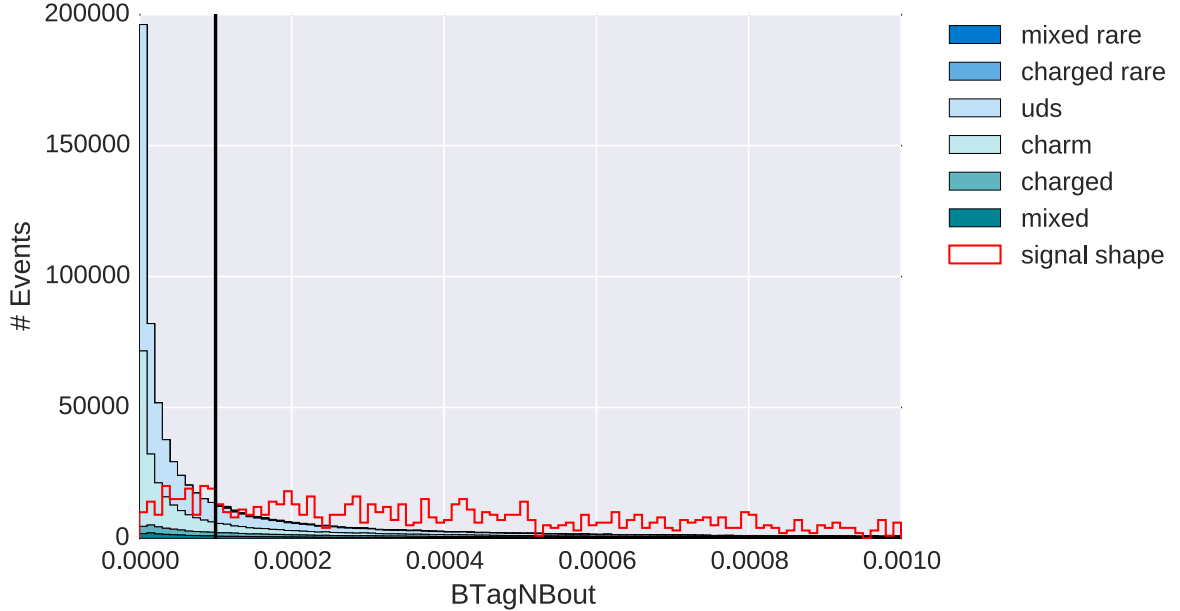


Figure 2: Histogram of the "BTagNBout" variable. This variable describes the NeuroBayes output for the B_{tag} -candidate. The contributions to the background are displayed in different color. The black line indicates the pre-cut of $BTagNBout > 10^{-4}$ that has been applied on this variable.

5.3 Analysis of classifiers

To find the best classifier for the signal-background discrimination a set of classifiers has been trained on background and a signal events to learn how to discriminate those. The

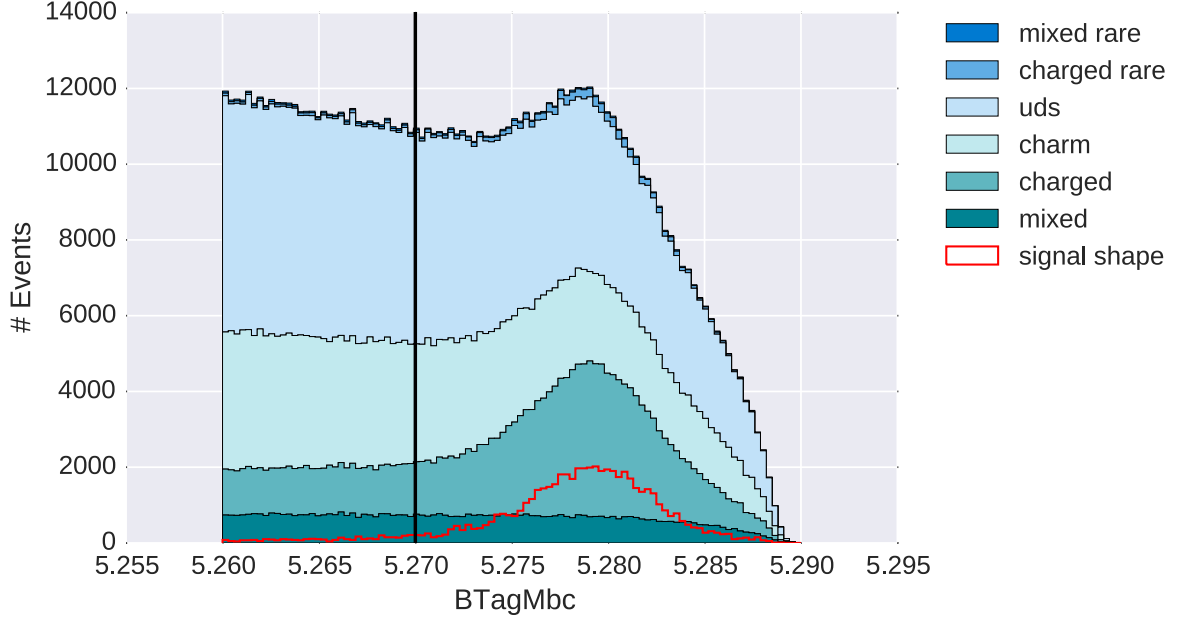


Figure 3: Histogram of the "BTagMbc" variable. This variable describes the mass of the B_{tag} -candidate. The contributions to the background are displayed in different colors. The black line indicates the pre-cut of $BTagMBC > 5.27 [GeV]$ that has been applied on this variable.

one with the best receiver operating characteristic (=purity versus efficiency behavior) then is chosen as the tool for the analysis. The classifiers investigated were a gradient boosted decision tree, a decision tree stump - a gradient boosted decision tree with a maximum depth of one layer, an adaptive boosted decision tree, a fisher discriminant and NeuroBayes. To investigate the behavior, purity and efficiency of a classification is plotted against each other while altering a cut on the output of the classifier. The results are depicted in figure 4. A gradient boosted decision tree was found to perform best.

5.4 Choice of variables to feed the classifier

The set of variables that has been chosen for the classification has been defined during the event selection step. The variables are displayed and their meaning are tabulated in table 1.

Their importance for the classification is depicted in figure 5. The measure for the importance is the relative frequency of occurring in the BDT's subtrees. This can be thought of as a measure of information gain from that variable.

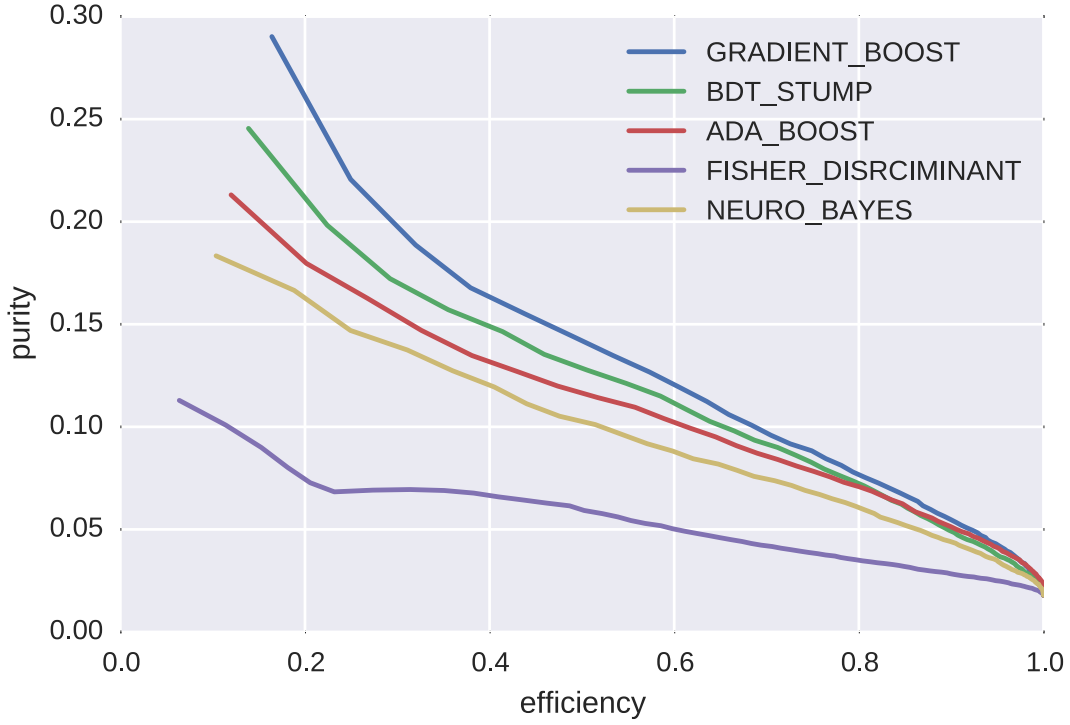


Figure 4: Purity versus efficiency of different classifiers. A gradient boosted decision tree has the best purity to efficiency characteristics.

5.5 Background contribution

The main contributions to the background originate from decays $e^+e^- \rightarrow \bar{u}u, \bar{d}d, \bar{s}s$ called 'uds', $e^+e^- \rightarrow c\bar{c}$ called 'charm' and $e^+e^- \rightarrow \bar{B}^+B^-$ called 'charged', as well as $e^+e^- \rightarrow \bar{B}^0B^0$ called mixed. There are also contributions from rare decays, however these are very small. The contributions in all variables is depicted in figure 6, figure 7 and figure 8.

5.6 Data MC comparison

After a set of suitable variables was chosen they have been compared to real data in order to check the accuracy of MC-modeling for each variable. A Kolmogorov-Test has been performed to check the overlap of all data points in a variable. It has been found that in a subset of variables the data is not well described by MC. These variables have been removed from the set of variables feed to the classifier. The comparison is depicted in figure 9 and figure 10.

The features which were assumed to be well described by both MC and data are depicted with their importance ranking in figure 11.

5.7 Upper limit calculation

The BDT then has been fitted with a MC-stream containing background and signal shape. The previously determined set of variables that agree with real data has been used in this fit. After that the upper limit for N_{obs} has been calculated using the method of W.Rolke [5] and the TRolke class as it is implemented in ROOT. The branching ratio has then been calculated using formula 5. The upper limit on \mathcal{B} was computed by varying the cut on the BDT output and the signal window size. In each step the classifier efficiency was calculated according to formula 6. The result from this computation is depicted in figure 12. The minimum in this plane is the expected upper limit. The upper limit as well as the other results found are depicted in table 2. The expected upper limit is $\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau) \leq 1.19 \cdot 10^{-3}$ on 90% confidence level. The result of varying the signal window is depicted in figure 13. The main contribution to the background is due to decays containing K_L which escape identification.

5.8 Systematics

Systematic uncertainties arise from the MC models to simulate the events, the detector simulation and uncertainties to the efficiency of B -tagging, classifying and the number of total $B\bar{B}$ -events that can be expected. Those from MC, detector simulation and B -tagging can be joined in an efficiency and treated as an uncertainty on that. But this quantity is still unknown. The uncertainty on the classifier efficiency is believed to be $\approx 2\%$. The uncertainty on the number of B -mesons is known with $\pm 1,36\%$. However, the systematics are not treated in this analysis.

6 Conclusion

An analysis aiming to set an expected upper limit to the branching ratio of the rare B -meson decay into a K -meson and two τ leptons was performed. This analysis is based on a counting experiment using multivariate classifying methods on preprocessed monte carlo data.

A boosted decision tree was chosen as the classifying tool for this analysis. A set of suitable variables was feed to the classifier. This set was determined by the quality of simulation reproducing data. An expected upper limit of $\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau) \leq 1.19 \cdot 10^{-3}$ on 90% confidence level was found.

The systematic uncertainties are not included in this calculation but are expected to alter the result by $\approx 10\%$. This has to be validated in the future.

The result is an improvement by approximately a factor three when comparing to the result of $\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau) \leq 3.3 \cdot 10^{-3}$ (90% CL) produced on BaBar data.

Table 1: Variables chosen for the classification and their meaning.

Variable	Meaning
NRemainGamma	number of remaining photons in the event
NRemainKs	number of remaining K_S
NRemainKL	number of remaining K_L
NRemainTracks	number of remaining tracks
Btag_costThetaB	$\cos(\theta)$ of the tag- B -meson
dist_to_IP	distance the secondary vertex to the interaction point
Ch0_Pstar_inBSigRest	momentum of the zeroth child in the B_{sig} restframe
chi2	χ^2 -value of the vertex fit
Qvalue	reconstructed mass of the B_{tag} with children mass subtracted
ChargeHighMomentumTrack	sign of the charge of the highest momentum track
AngleMissingMomentumToBeam	the angle of missing momentum track relative to the beam
Ch1_AngleToBeam	child one angle with respect to beam
Ch2_AngleToBeam	child two angle with respect to beam
DistToOtherB_dz_significance	significance of the distance to the other B -meson in z -direction
DistToOtherB_dz	distance to the other B -meson in z -direction
BTagMbc	reconstructed mass of the tag- B -meson
pt	transversal momentum of the signal- B -meson
MissingMass	missing mass of the event calculated from the B masses
Ch01_Angle	angle between child 0 and child 1
q2_tag.k	barbar paepr suchen
MomentumAsymmetrie	momentum asymmetry of the event
Ch0_energy	energy of child 0
Ch0_pt	transverse momentum of child 0
CosThetaB	$\cos(\theta)$ of the reconstructed B -meson
R2	ratio of the second to the zeroth Fox-Wolfram-Moment
Ch1_NBout	NeuroBayes output for child 1
Ch01_InvMassScaled	invariant mass of child 0 and 1
Ch02_InvMassScaled	invariant mass of child 0 and 2
Ch12_InvMassScaled	invariant mass of child 1 and 2
Ch1_PseudoHelAng	cosine of angle of the pseudo helicity of child 1
Ch2_PseudoHelAng	cosine of the angle of the pseudo helicity of child 2
Children_nboutprod	product of the NeuroBayes ouput of child 1 and 2
Emiss	missing energy of the event
Mmiss	missing mass of the event calculated from all reconstructed tracks
BTagDE	ΔE of the tag- B -meson
BTagNBout	NeuroBayes output for the tag- B -meson

Table 2: Results from the upper limit calculation.

results	
upper limit on branching ratio:	$\mathcal{B}(B^\pm \rightarrow K^\pm \tau \tau) \leq 1.19 \cdot 10^{-3}$
classifier cut:	$BDT_{out} \geq 0.129$
signal window cut:	$ExtraEnergyECL \leq 0.324 [GeV]$
classifier efficiency:	$\epsilon_{class} = 0.42$

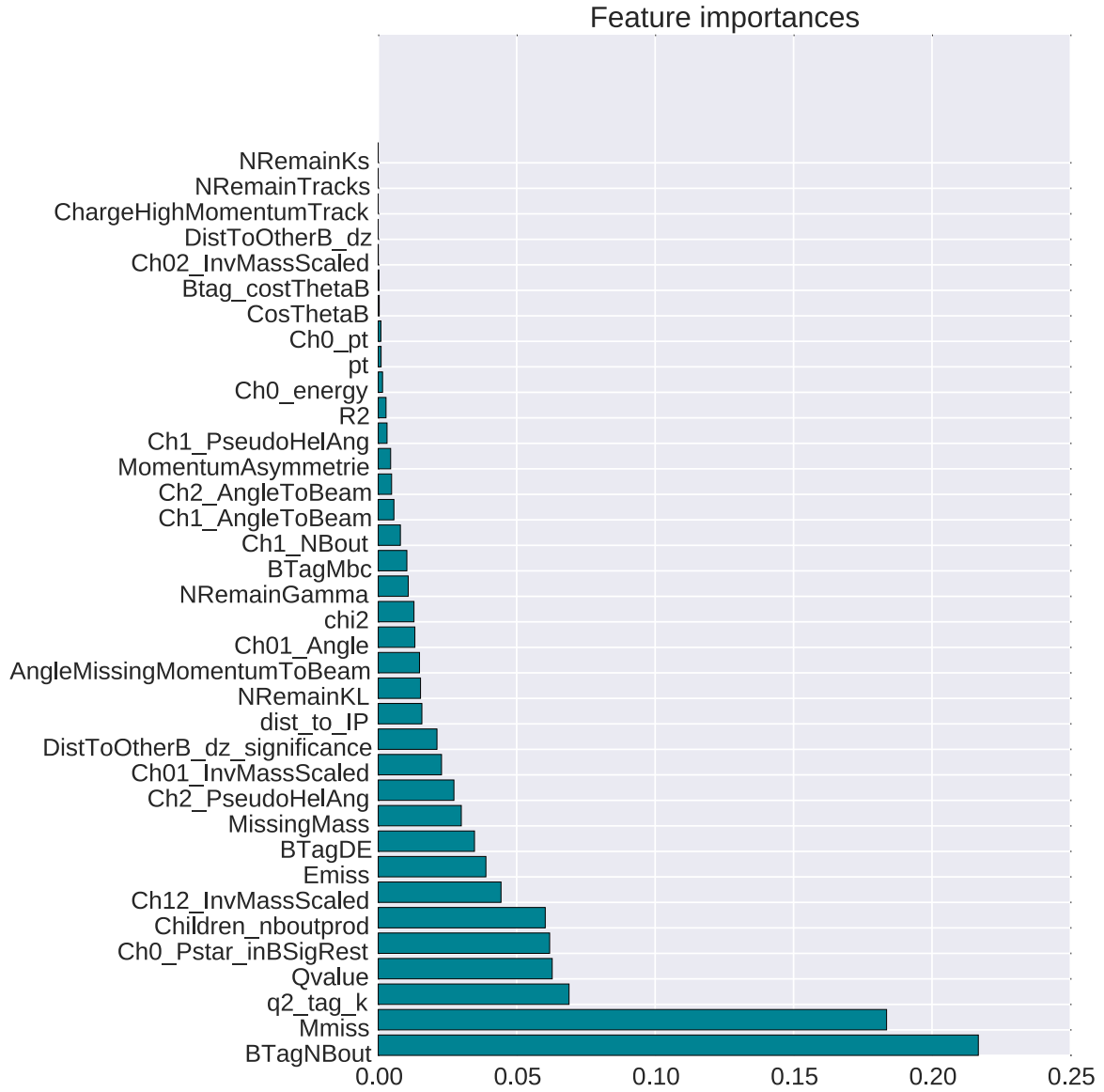


Figure 5: Feature importance according to the BDT. The X-Axis displays the relative frequency of occurring in the subtrees.

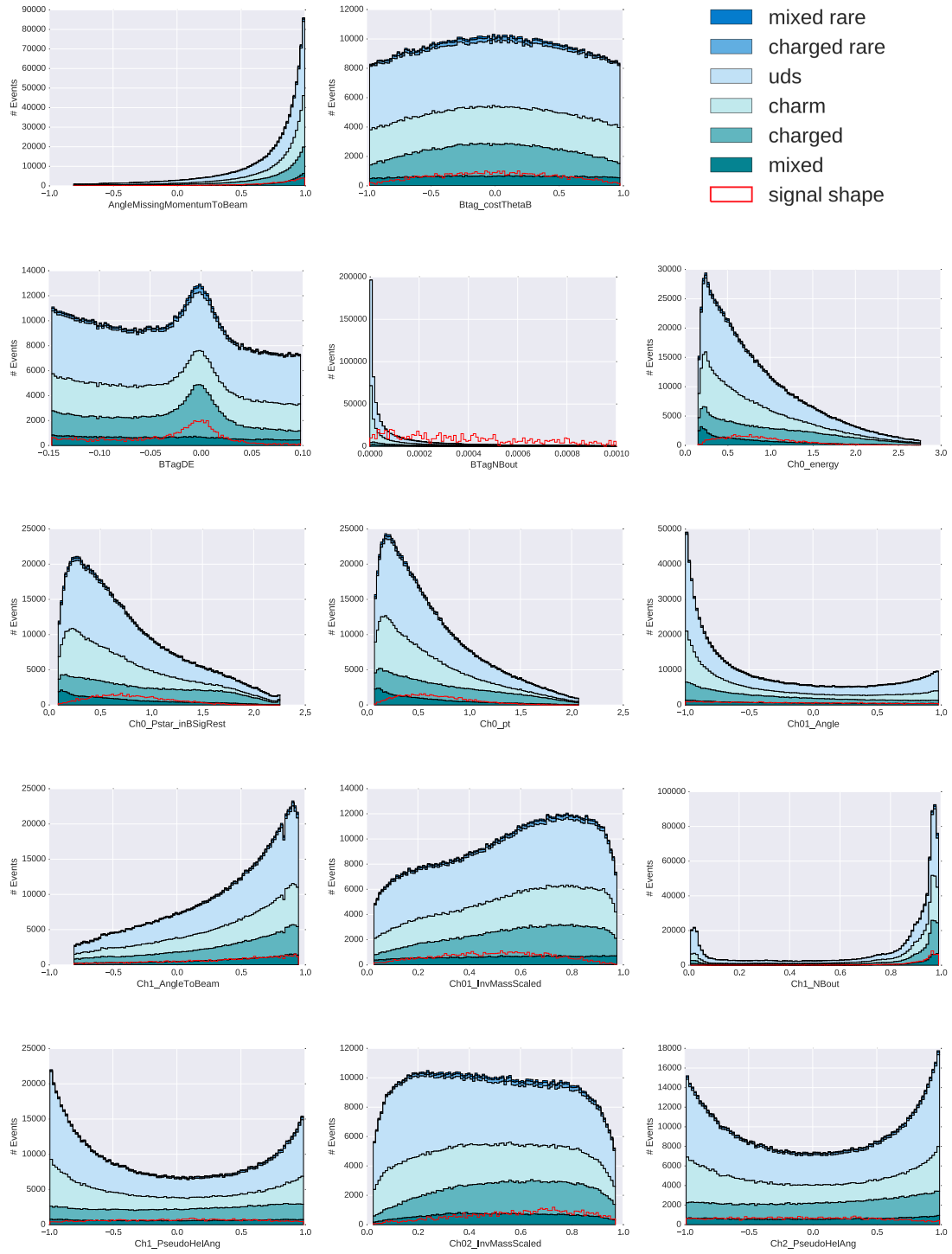


Figure 6: Background contributions in the different variables.

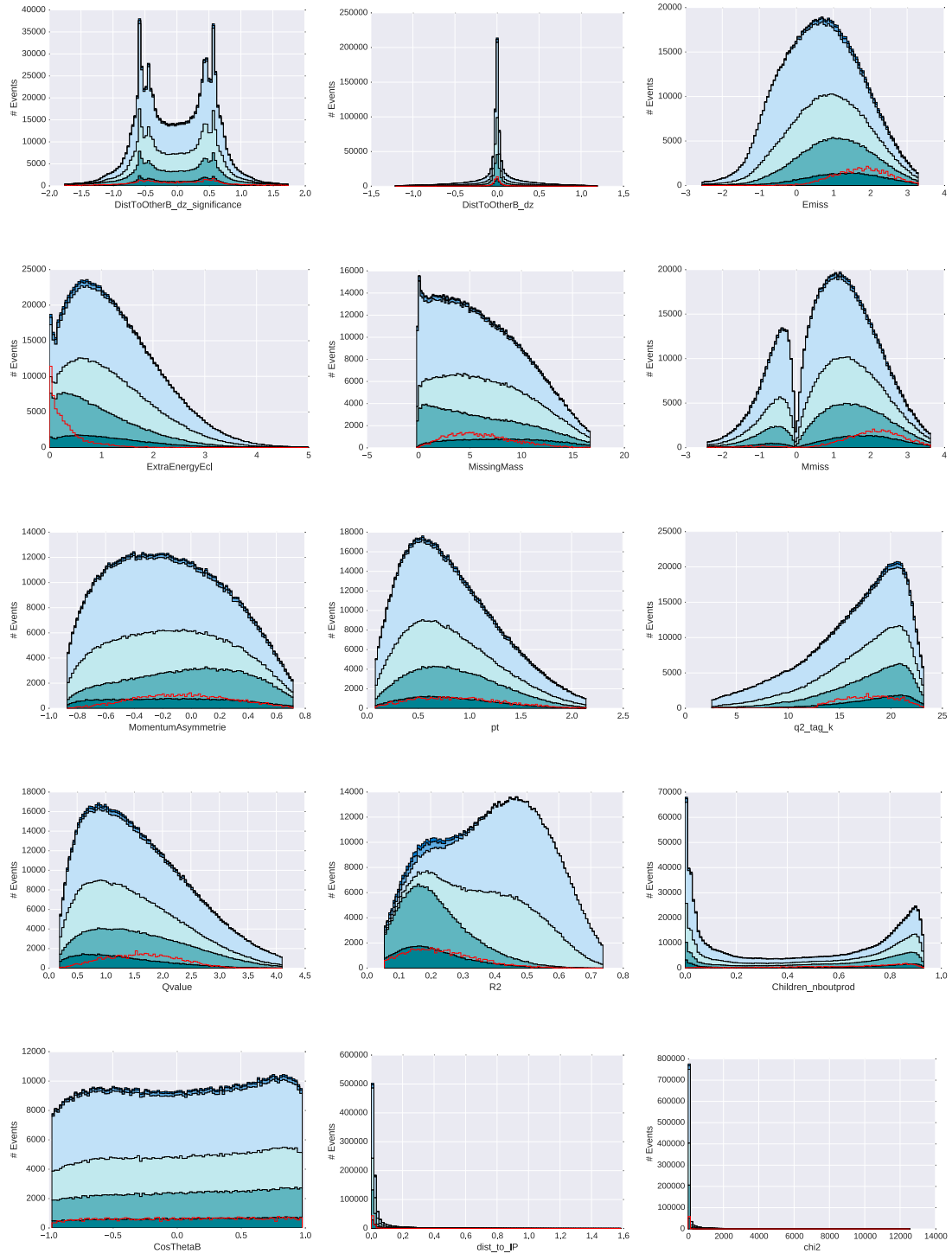


Figure 7: Background contributions in the different variables. The legend can be found in figure 6.

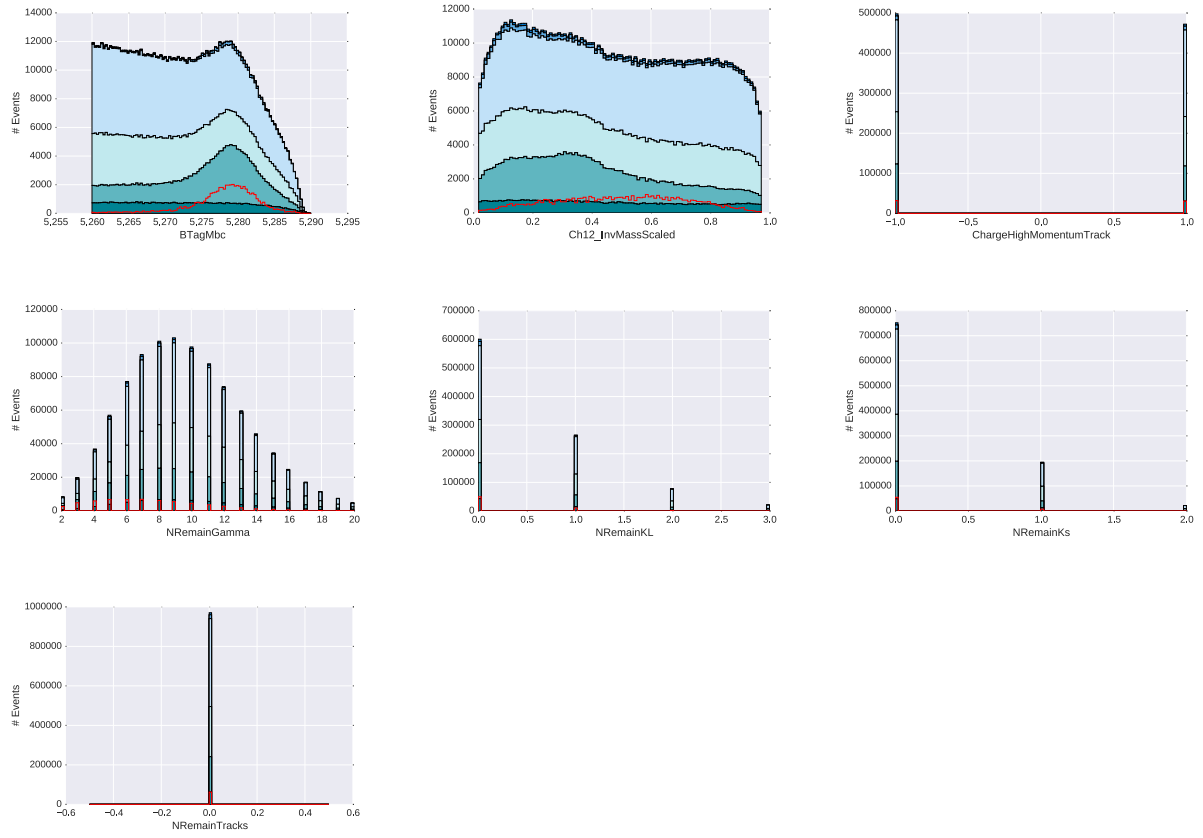


Figure 8: Background contributions in the different variables. The legend can be found in figure 6.

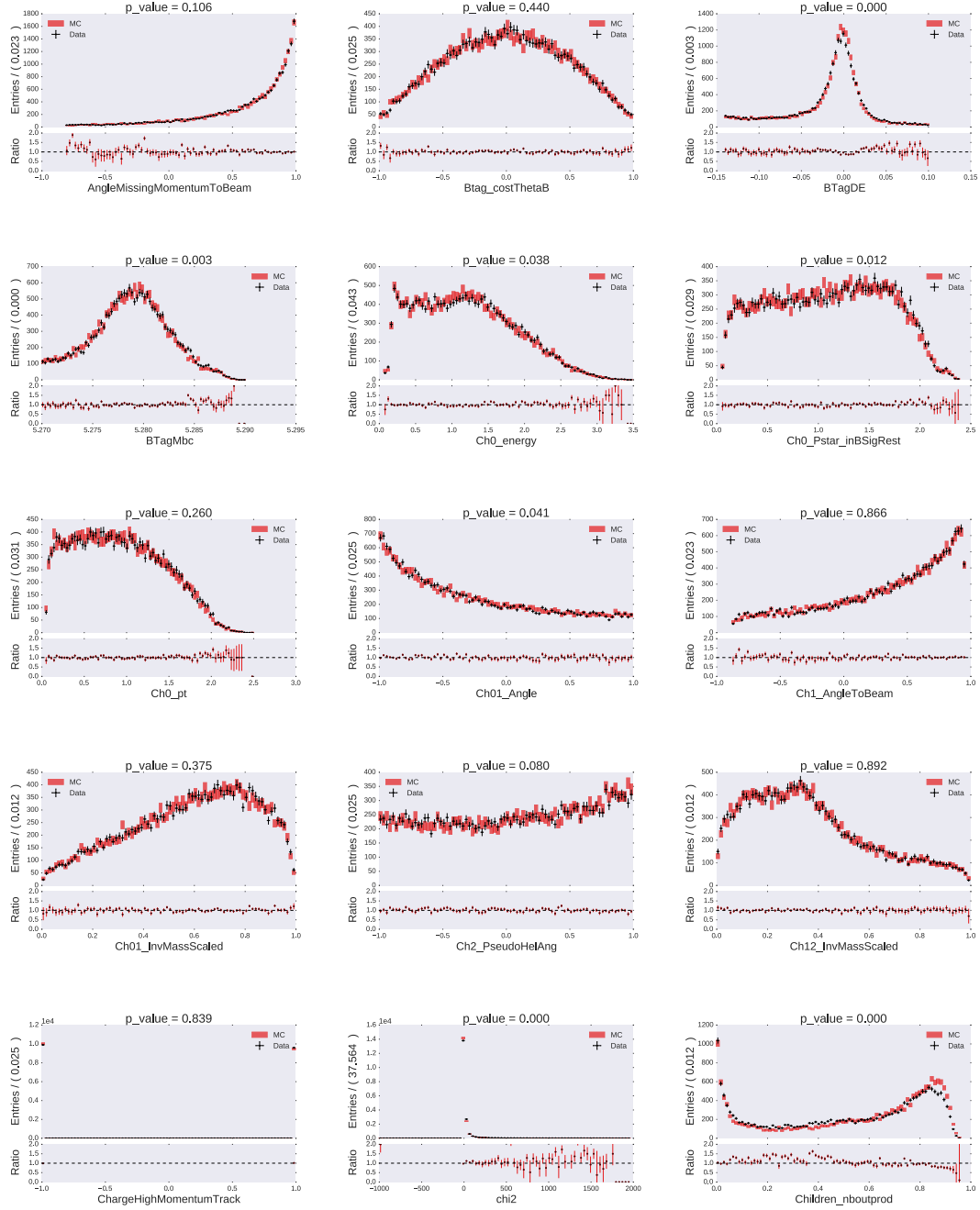


Figure 9: Comparison of real data with MC using the same event selection. A Kolmogorov-Test has been performed to measure the quality of MC for the different variables.

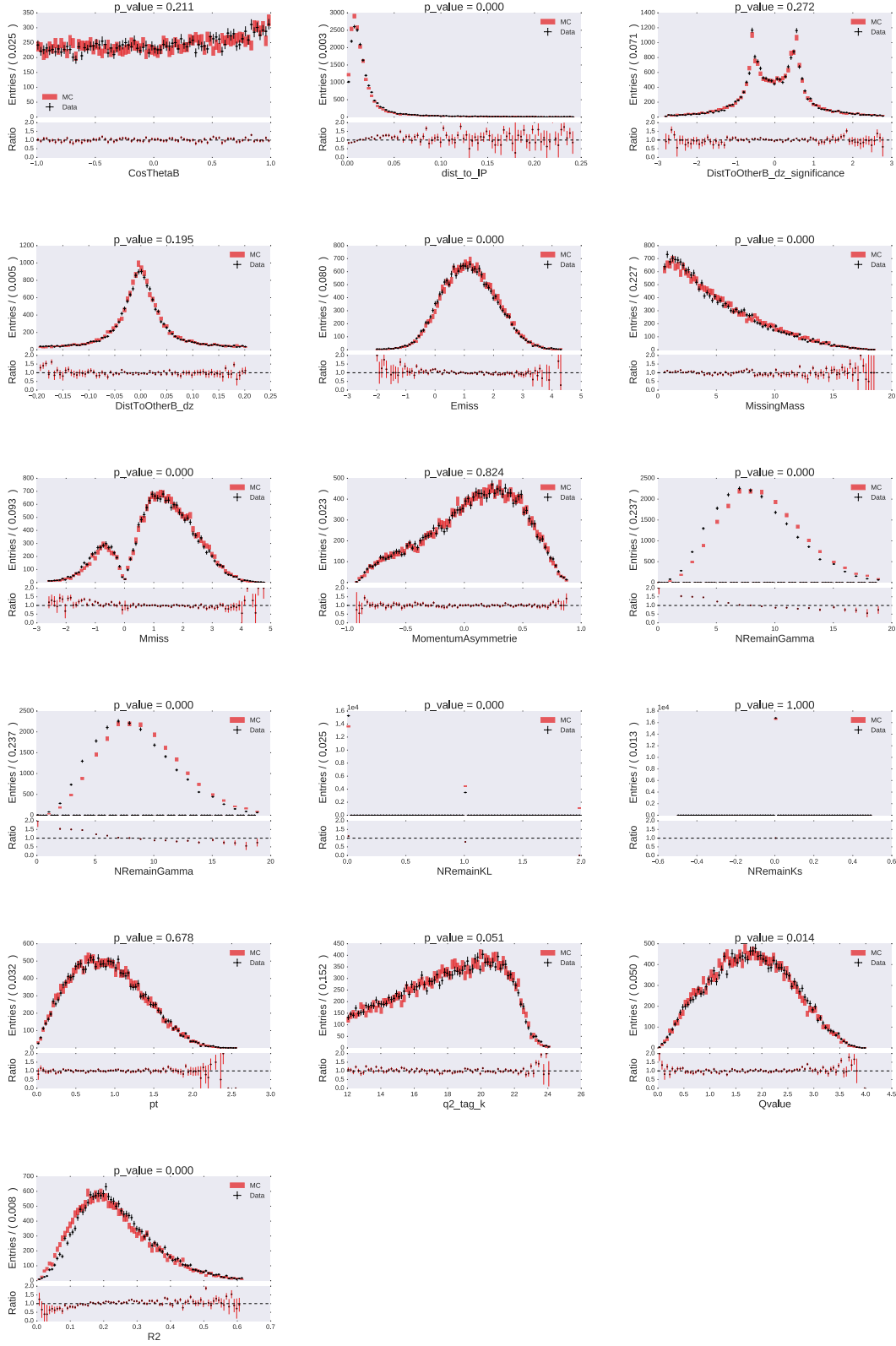


Figure 10: Comparison of real data with MC using the same event selection. A Kolmogorov-Test has been performed to measure the quality of MC for the different variables.

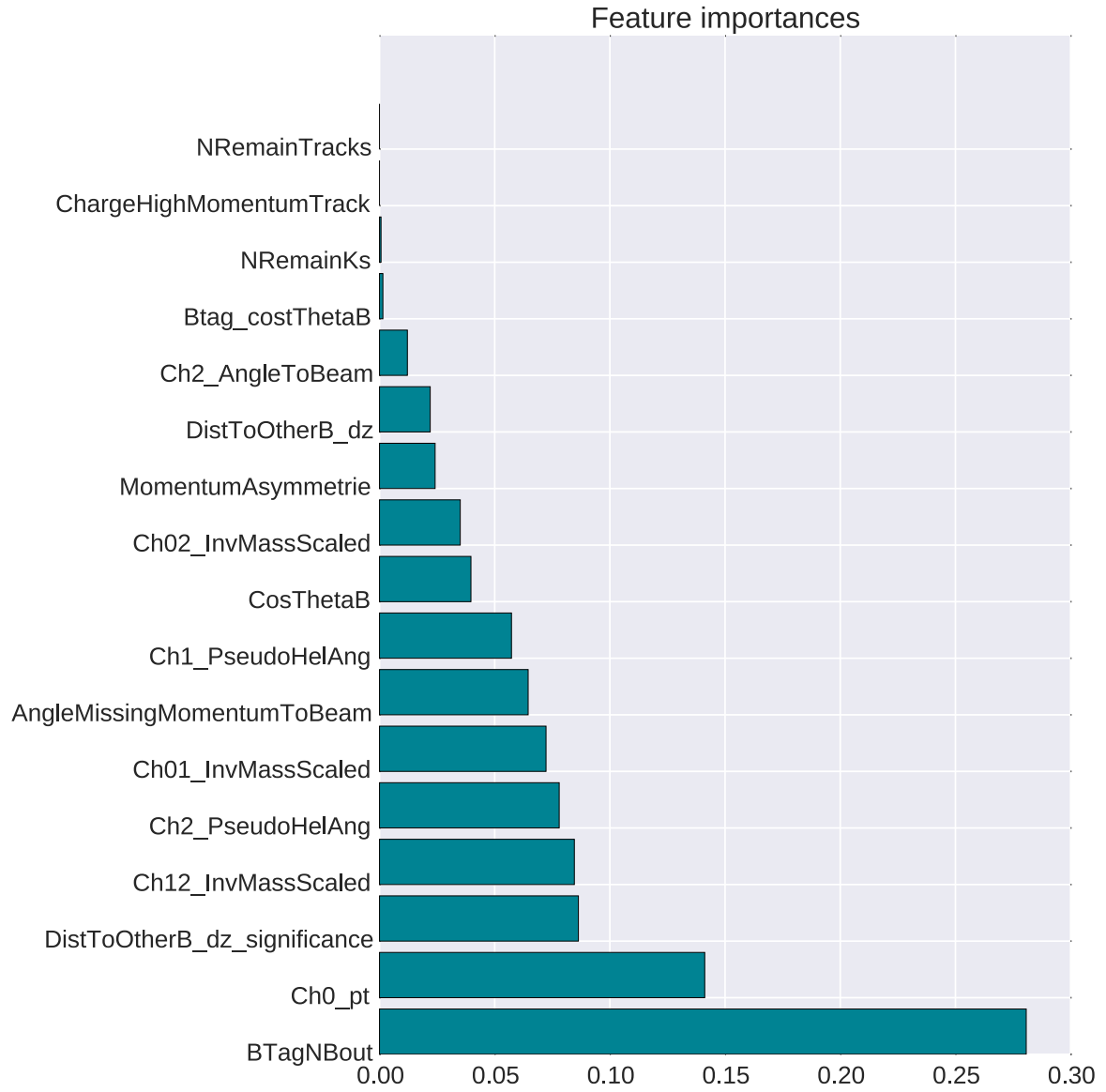


Figure 11: Feature importance according to the BDT. The X-Axis displays the relative frequency of occurring in the subtrees. This set of variables was chosen after the data-MC comparison as all features are well described in MC and data.

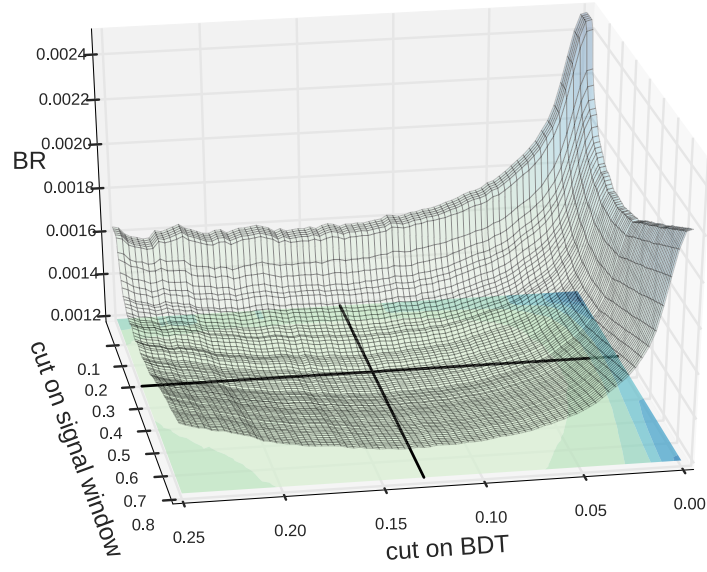


Figure 12: Computed branching ratio depending on the cut applied to the BDT output and the signal window size. The signal window is described by the variable "ExtraEnergyECL". The minimum point in this topography is the expected upper limit on the branching ratio. The cuts producing this minimum are indicated with a black line.

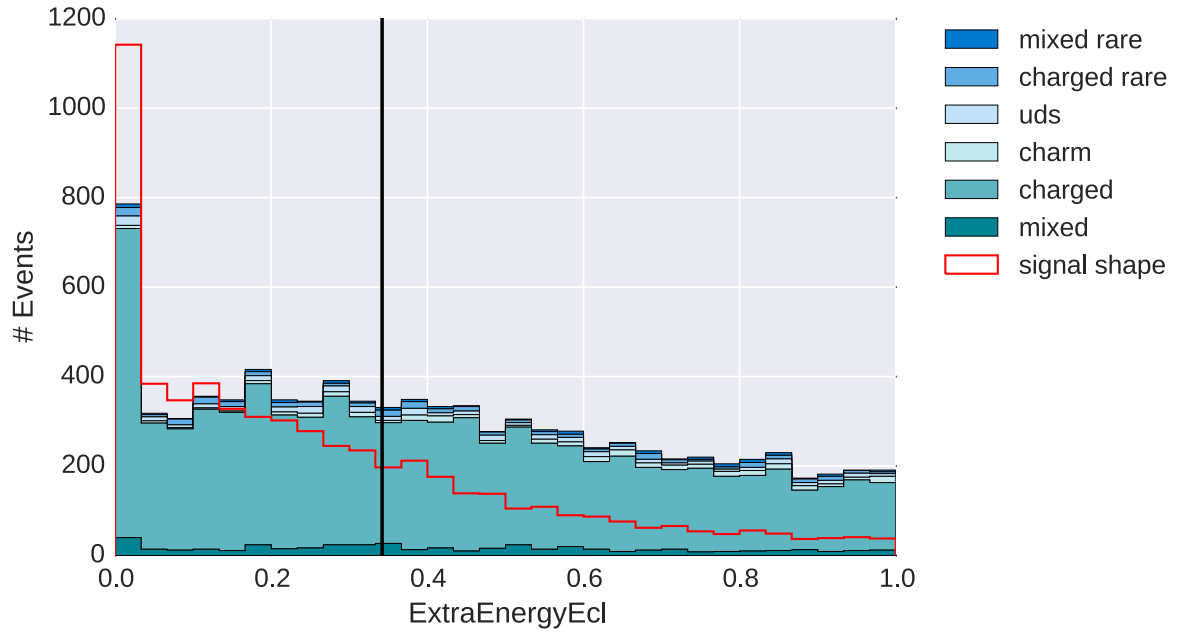


Figure 13: The resulting cut from varying the signal window. The numerical value found is: $ExtraEnergyECL \leq 0.324 [GeV]$. The main contribution to the background is due to decays containing K_L which escape identification.

References

- [1] LHCb-CONF-2015-002 (2015) *LHCb collaboration*
- [2] Phys. Rev. Lett. 111 (2013) 191801 *R. Aiji et al. [LHCb Collaboration]*
- [3] arXiv:1410.4545v2 (2014) *T.Hurth*
- [4] POS (ICHEP 2010) 234 *K. Flood*
- [5] Nucl. Instrum. Meth. A551: 493-503 (2005) *W.Rolke, A. Lopez, J., Conrad and Fred James*