September 9th 2014

Update on simulations concerning the measurement of the time-like form factors in reactions of

$$\bar{p}\,p \rightarrow \mu^{+}\mu^{-}$$

Iris Zimmermann

Johannes-Gutenberg Universität Mainz, Helmholtz Institut Mainz

Motivation

- Differential cross section[1] of the reaction $\bar{p}\,p \rightarrow \mu^+\mu^-$

  → access to the **time-like, electromagnetic form factors of the proton G_E and G_M:**

$$\frac{d\sigma}{d\cos\theta_{CM}}(s,\theta) = \frac{\alpha^2\pi}{2\cdot s}\cdot\frac{p_{l^-}}{\bar{p}}\cdot\left|G_M\right|^2\left[\frac{4M_p^2}{s}\left(1-\beta^2\cos^2\theta_{CM}\right)\cdot R^2+\left(1+\frac{4m_l^2}{s}+\beta^2\cos^2\theta_{CM}\right)\right]$$

$$R=\frac{\left|G_E\right|}{\left|G_M\right|}$$

- Individual measurement of $\left|G_E\right|$ and $\left|G_M\right|$
- **Strong hadronic background**, mainly $\bar{p}\,p \rightarrow \pi^+\pi^-$

$$\frac{\sigma\left(\mu^+\mu^-\right)}{\sigma\left(\pi^+\pi^-\right)} \propto 10^{-6}$$

→ Good muon-pion separation needed!

1) first derived by A.Zichichi et al., Nuovo Cimento XXIV,170 (1962)

Simulation & Analysis

- **Simulations** for both **signal and background** at beam momentum of 1.7 GeV/c
  - PandaRoot trunk version (linked to external packages apr13)

    Event generation*:

    Signal: $\bar{p}\,p \rightarrow \mu^+\mu^-$ **10^6 events (expected: ~0.83 * 10^6 events)**

    Background: $\bar{p}\,p \rightarrow \pi^+\pi^-$ **10^8 events (expected: ~1.63 * 10^{11} events)**

    $$|\cos(\theta_{CM})| < 0.8$$

- **Analysis**:
  - **Preselection** of the data, application of different cuts to achieve good **background suppression**
  - **Multivariate Analysis**

***Event generators developed by Manuel Zambrana, HIM.**
Simulation by A. Dbeyssi, D.Khaneft, I. Zimmermann

1) Analysis: Data Preselection

Signal: $p\,\bar{p} \rightarrow \mu^+ \mu^-$ Background: $p\,\bar{p} \rightarrow \pi^+ \pi^-$

Preselection of the data

For each event: Combination of pairs

If more than 1 pair: take “best pair” :

→ closest to 180° back-to-back production in center-of-mass system

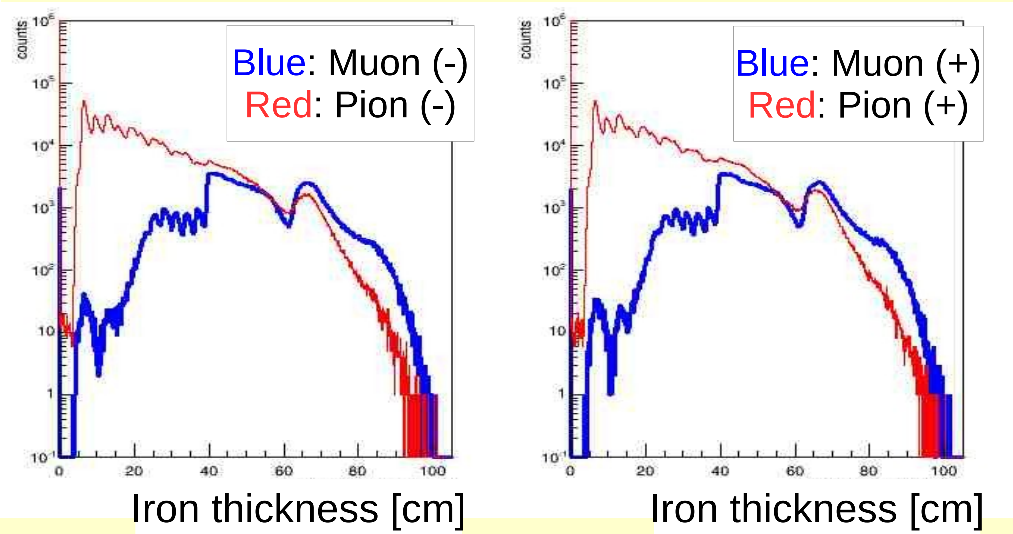
- ➢ Exclude bad fitted tracks
- ➢ Take only pairs which include 1 positive & 1 negative charge
- ➢ Both candidates must have hits in MDT

1) Analysis: Data Preselection

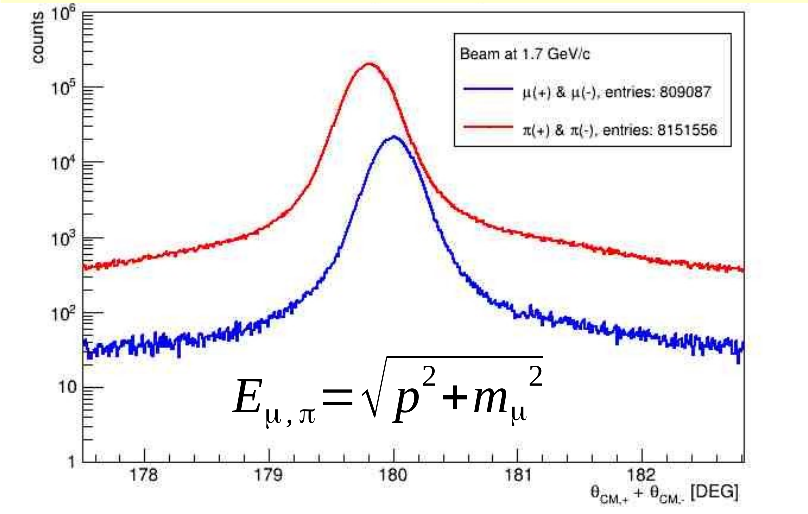
After preselection: ... some of the variables...

Muon System

Path length inside iron absorber



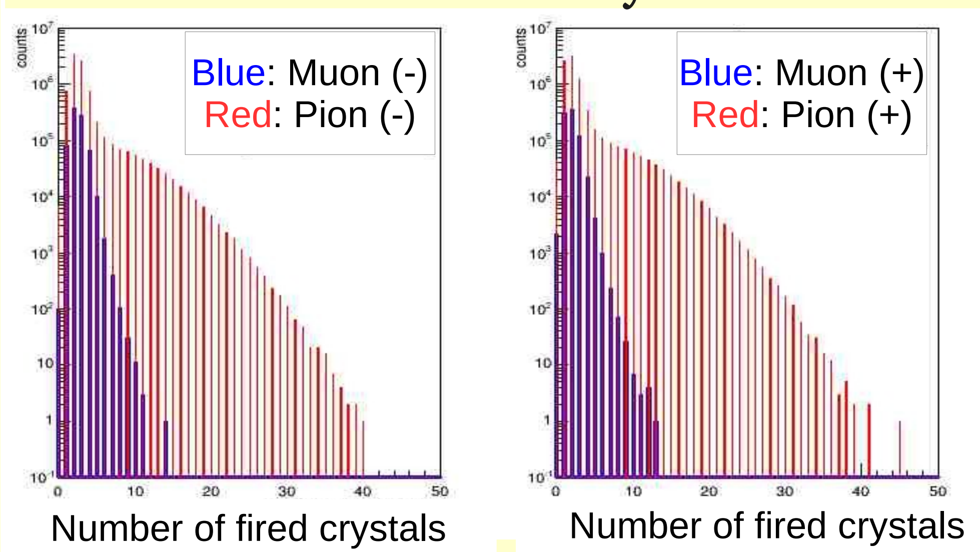
$\theta^{+}_{CM} + \theta^{-}_{CM}$



→ Next Step: Application of hard cuts

Electromagnetic Calorimeter

Number of fired crystals



2) Analysis: Hard Cuts

First scenario: Application of cuts

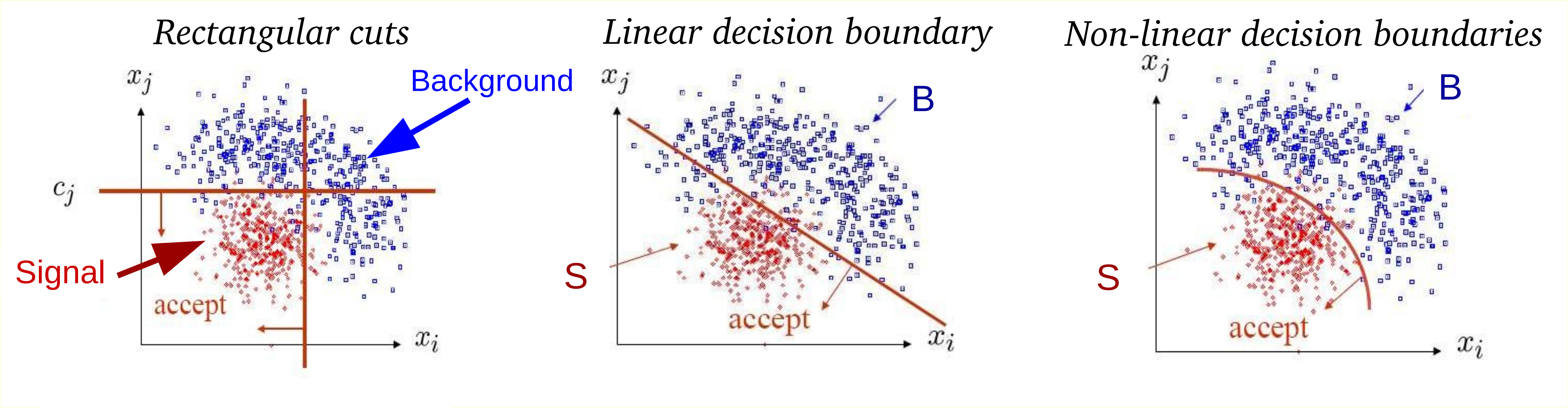
Invariant mass [GeV/c]	] 2.3 , 2.38 [
$\theta^{+}_{CM} + \theta^{-}_{CM}$ [DEG]	] 179.95 , 180.5 [
Path length inside muon system [cm]	> 42.0
$P(\mu^{+})_{MDT} ; P(\mu^{-})_{MDT}$	> 0.99
$P(\mu^{+})_{EMC} ; P(\mu^{-})_{EMC}$	> 0.05

Beam momentum: 1.7 GeV/c	#Events: Signal	#Events: Background
Monte-Carlo Simulation	10^6	$1.0425*10^8$
After Preselection	809087 (~80.9 %)	8151556 (~ 8.2%)
After Preselection & Hard Cuts	112278 (~11.2 %)	1660

Background suppression factor ~ 2 * 10^{-5} → could this be improved?

3) Analysis: Multivariate data classification (Toolkit for Multivariate Analysis)

How to find an optimal decision boundary?



Goal of Multivariate Analysis: Find *optimal decision boundary* using *many variables*,
→ optimal signal/background separation.

TMVA: Root-integrated software package for processing and evaluation of **multivariate classification methods.**

For instance:

- Artificial neural networks (**Multilayer Perceptron**)
- Support Vector Machine (**SVM**)
- Linear discriminant analysis (**Fisher**)
- Multidimensional k-nearest neighbour method (**KNN**)
- ...

3) Analysis: Multivariate data classification (Toolkit for Multivariate Analysis)

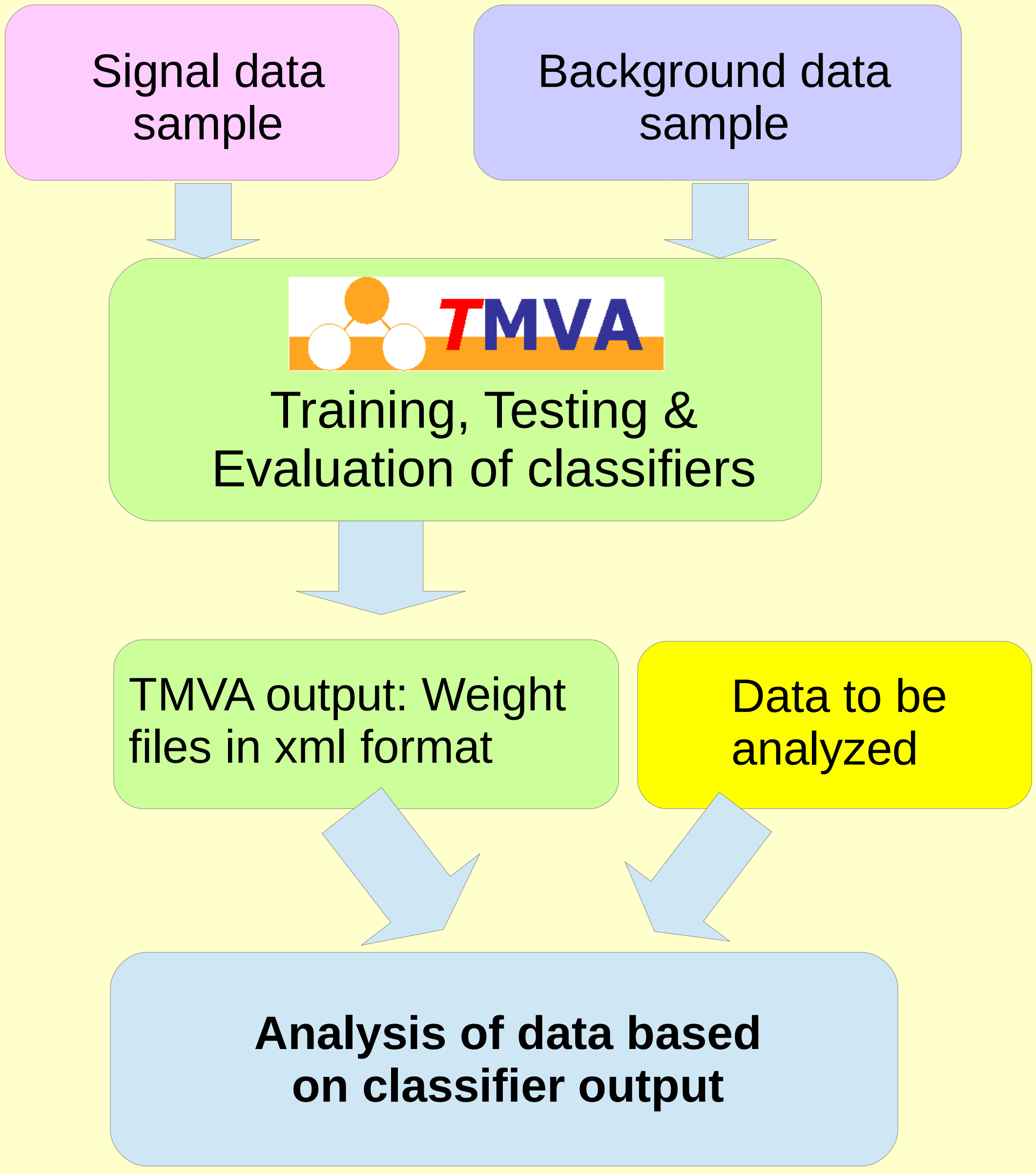
Analysis using TMVA

1) **Training phase**

- Signal and background: Set of variables (X_1,X_2, X_3, … X_n) as input for each classifier
- MVA output → weight files for each classifier
- Testing and Evaluation

2) **Application phase**

- Classification of the new data sample based on the TMVA output



3) Analysis: Multivariate data classification (Toolkit for Multivariate Analysis)

Before training: Apply cuts on training data samples

- Sum of polar production angles in CMS $\theta^{+}_{CM} + \theta^{-}_{CM} > 178.0°$ [DEG]
- Invariant Mass: $\sqrt{s} \in [2.2; 2.5]$ GeV/c

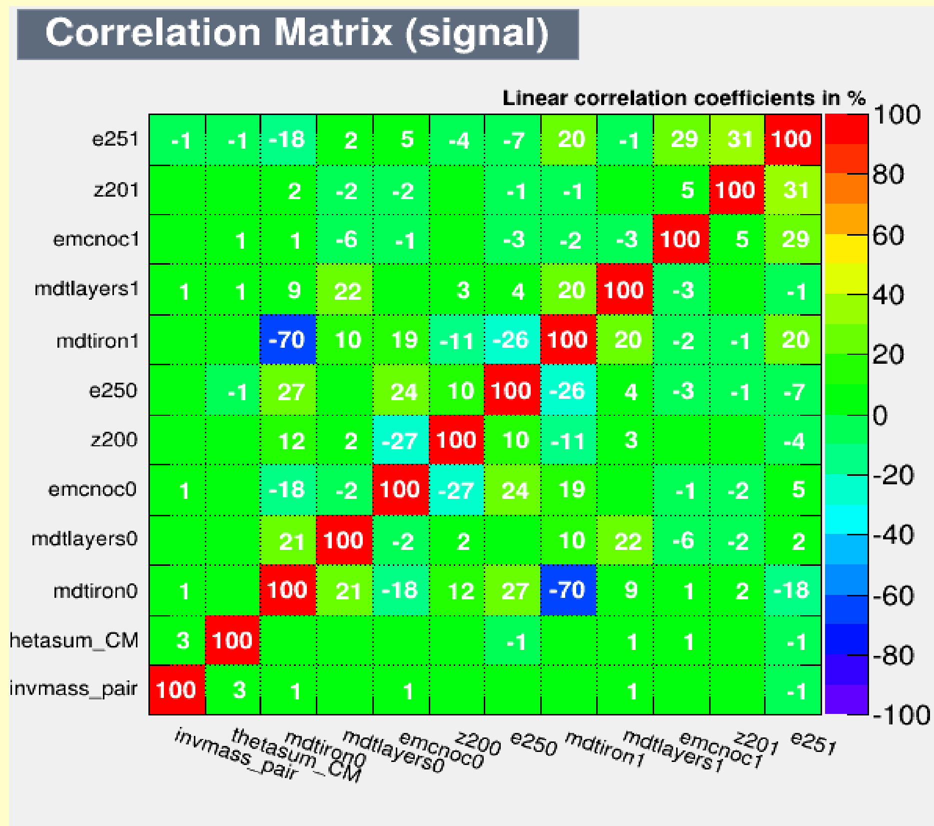
TMVA

Correlation between the variables

→ amount of variables can be reduced by excluding strongly correlated variables

A **Scenario A: In total 12 variables**

Linear correlation coefficients



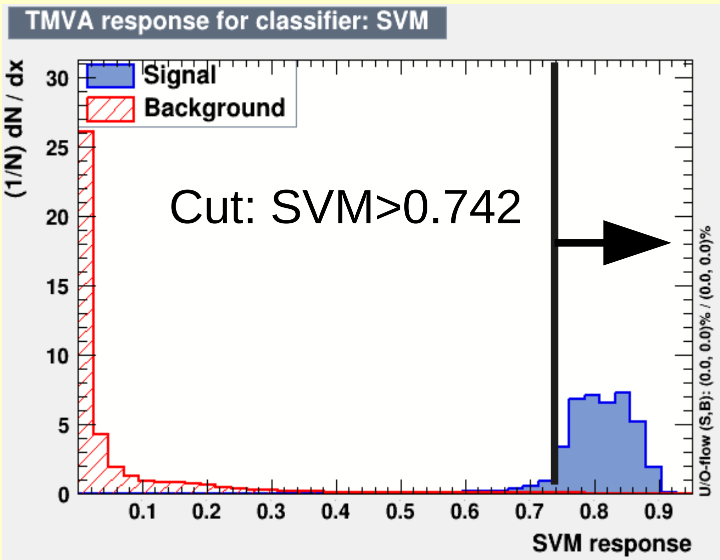
A

Scenario A:
In total 12 variables

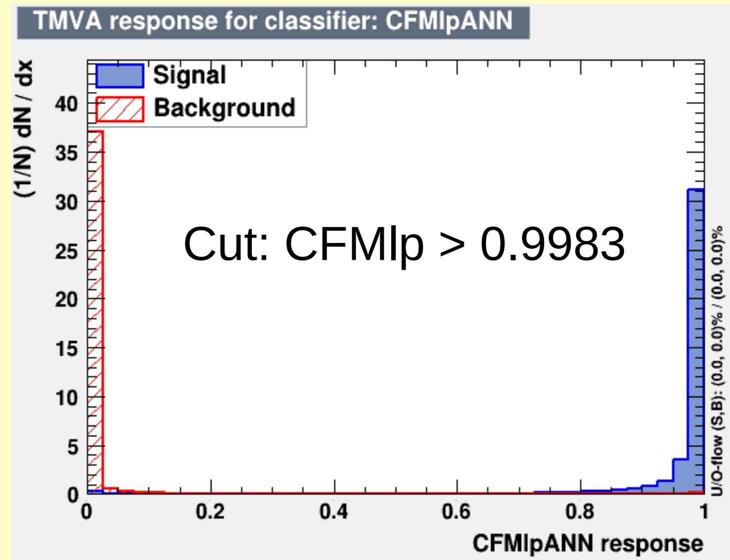
3) Analysis: Multivariate data classification

Output of classifier

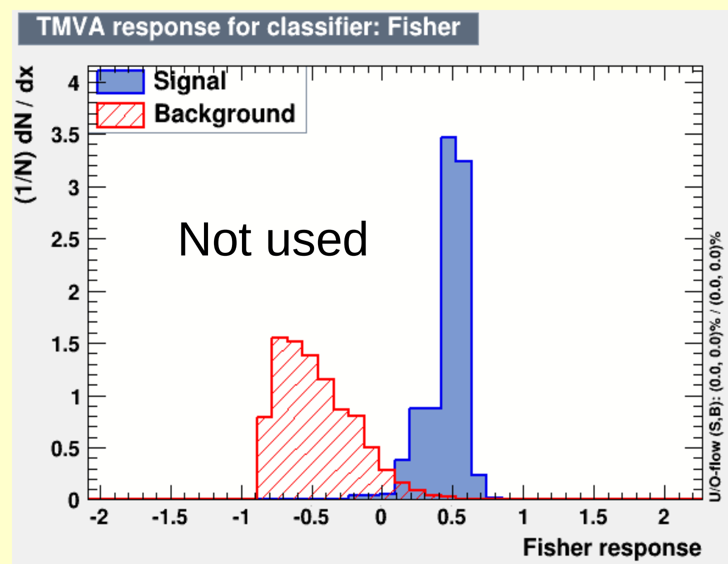
Support Vector Machine



Multilayer Perceptron



Linear discriminant analysis (Fisher)

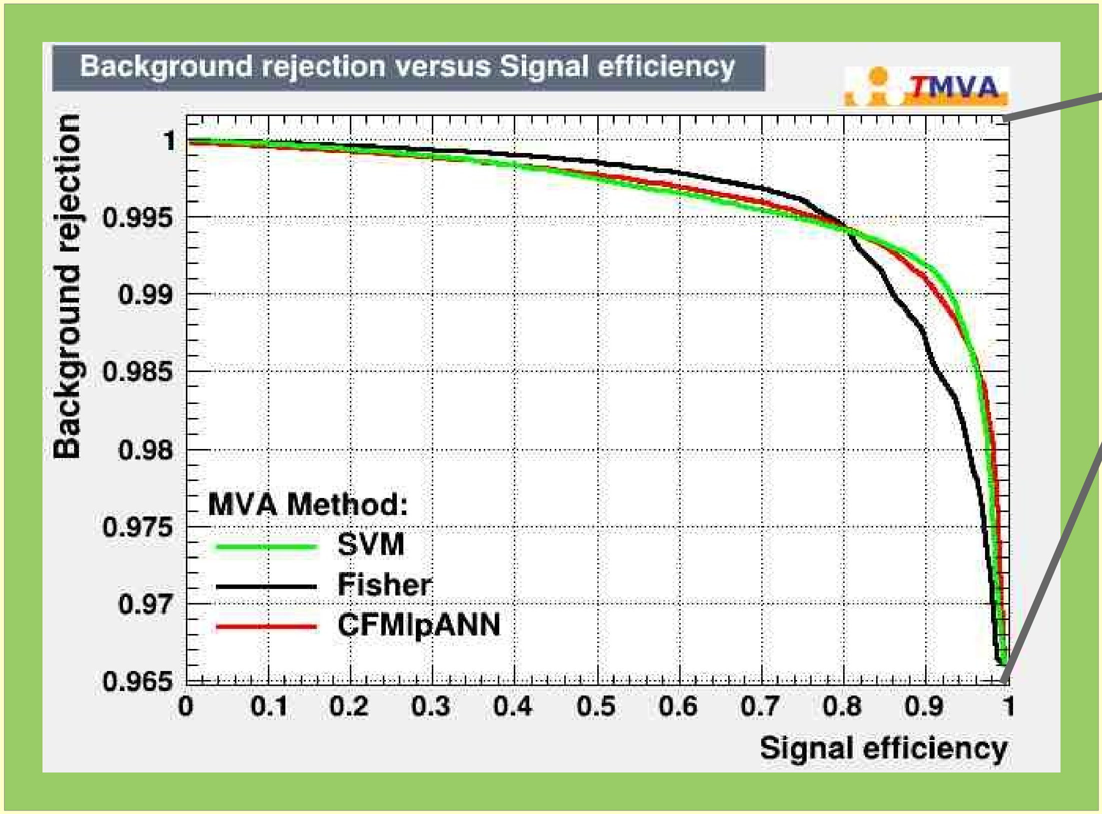


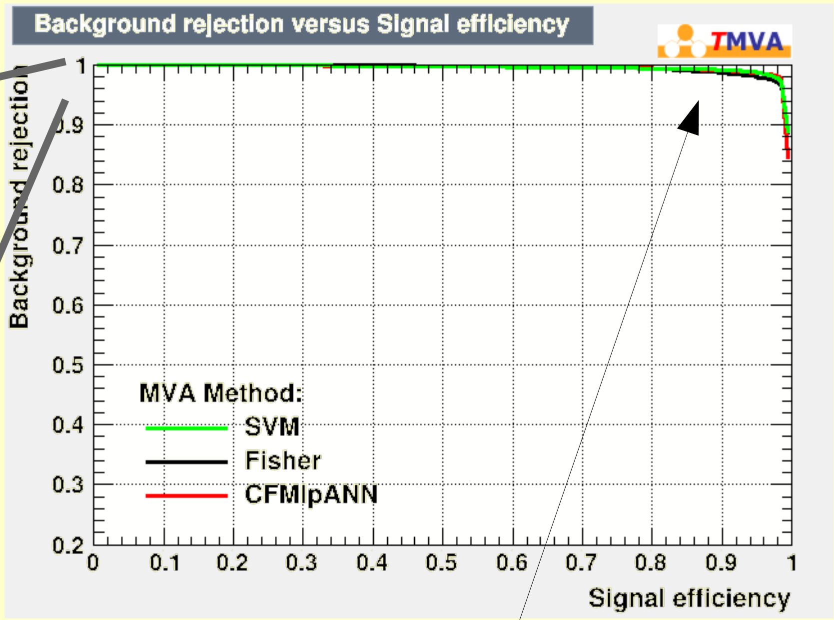
A Scenario A:
In total 12 variables

3) Analysis: Multivariate data analysis

Receiver operating characteristic (ROC curve)

The **ROC curve** characterizes the **quality of the classification procedure**.





ROC curve shows **good performance** of Fisher, CFMlpANN and SVM.

Background rejection = 1 – background efficiency

Comparison of the scenarios

Beam momentum: 1.7 GeV/c	#Events: Signal	#Events: Background
Monte-Carlo Simulation	10^6	$1.0425*10^8$
After Preselection	809 087 (~80.9 %)	8 151 556
After Preselection & Hard Cuts	112 278 (~11.2%)	1660
A **Scenario A*:** **In total 12 variables**	109 227 (~10.9%)	512
B **Scenario B*:** **In total 22 variables**	112 997 (~11.3 %)	448

* **Applied cuts:**

- Multilayer Perceptron (CFMlp > 0.9983)
- Support Vector Machine (SVM > 0.742)
- Invariant Mass $\sqrt{s} \in [2.2\,;2.5]$
- $\theta^{+}_{CM} + \theta^{-}_{CM} \in [179.99, 180.40]$

Current status:
Background suppression factor
$\sim 4.3*10^{-6}$

Expected physical signal-to-background ratio ~ 1/7.5
→ Signal extraction by background subtraction may be possible.

Summary & Outlook

- **Simulation & Analysis** for beam momentum of **1.7 GeV/c**:
  - *Signal*: $p\bar{p} \rightarrow \mu^+\mu^-$ **10^6 events**
  - *Background*: $p\bar{p} \rightarrow \pi^+\pi^-$ **10^8 events**
  - Analysis: Comparison of **hard cuts** and **multivariate classification methods** (TMVA toolkit in ROOT) for signal/background separation.
- **Hard cuts: Signal efficiency ~ 11.2% , background suppression ~2 * 10^{-5}**
- **TMVA: Signal efficiency ~ 11.3% , background suppression ~4.3 * 10^{-6}** could be obtained using **MVA methods → Signal extraction by background subtraction?**
- Test of **different scenarios**: search for optimal set of variables/cuts using TMVA classification methods is under investigation

Backup

What do we expect?

- At 1.7 GeV/c the expected number of events:

Signal: N ~**0.83 * 10^6** events of $\bar{p}\,p \rightarrow \mu^+\mu^-$

Background: N~**1.63 * 10^{11}** events of $\bar{p}\,p \rightarrow \pi^+\pi^-$

$|\cos(\theta_{CM})| < 0.8$

- For signal efficiency ~11.3 % and background suppression factor ~4.3 * 10^{-6} we could make a rough estimate:

Signal: N(after analysis)~ 93790 events

Background: N(after analysis)~ 700900 events

→ Signal-to-background ratio ~ 1 : 7.5