$Z \rightarrow \tau\tau$
New method for background estimation from data

Gordon Fischer

Deutsches Elektronen Synchrotron - DESY

M -> tautau Workshop

Outline

- Problems with current methods
- General description
- Solutions for the equation system
- First values
- Outlook

Problems with current methods

- OS/SS or ABCD promise a good background estimation
- But: many uncertainties (modified cuts for control regions, correct conclusions from control to signal region, assumptions for shapes, correct determination of electro-weak background is MC dependend, no background estimation on reconstruction level)
- The introduced method can avoid such problems (but of course causes new problems)

General description: fake rates

- The idea is to use **fake rates** for DiJets, W+jets, and Z+jets events
- Background for $Z \to \tau\tau$ is QCD, W+Jets, and Z+Jets (additional bkg negligible)
- OS: Signal+QCD+W+Z
- SS: QCD+W+Z
  (assume for tighter tau ID that the number of signal in SS region is negligible)
- **DiJets** fake rates: use studies on dijet fake rates
- **Z+Jets** fake rates: use studies from Zee+Jet
- **W+Jets** fake rates: use studies from Zee+Jet
  since Jets from W or Z events are identical (quark jets) $\rightarrow$ can use the same results

General description: $Z \to \ell\ell$ background

- For $Z \to ee$+Jets (also for $Z \to \mu\mu$+Jets) two possible scenarios:
  - the jet fakes the τ_h
  - the lepton fakes the τ_h
  $\Longrightarrow Z \to \ell\ell$ background is composed of Z = $Z_{\ell->\tau} + Z_{jet->\tau}$
- Use lepton fake rate studies via $Z \to \ell\ell$
- With $\frac{\#(\ell->\tau)}{\#(jet->\tau)}$ the relative contribution can be determined
- Of course kinematic effects have to be investigated

General description: basic idea

- For a certain Tau ID we can write:
  I) OS = S1 + QCD_{OS} + $W_{_{OS}}$ + Z_{OS}
  II) SS = QCD_{SS} + $W_{_{SS}}$ + Z_{SS}
- For the second Tau ID we can write:
  I) OS (ID 1)= S1 + QCD_{OS} + $W_{_{OS}}$ + Z_{OS}
  II) SS (ID 1)= QCD_{SS} + $W_{_{SS}}$ + Z_{SS}
  III) OS (ID 2)= S2 + A $\times$ QCD_{OS} + B $\times$ $W_{_{OS}}$ + C $\times$ Z_{OS}
  IV) SS (ID 2)= A $\times$ QCD_{SS} + B $\times$ $W_{_{SS}}$ + C $\times$ Z_{SS}
- S1 and S2 expresses the signal and the relative fake rates A = $\frac{FR(TauID2)}{FR(TauID1)}$ for the DiJet fake rate (analogically with B and C for W and Z events) describes the background
- For the full description we need 12 equations
  $\rightarrow$ equations = unknown variables

General description: mathematics

- Can write: **A** $\times \vec{X}$ = $\vec{b}$ with **A** = coefficient matrix, $\vec{b}$ = measured values and $\vec{X}$ the vector which includes the solutions.
- To solve this linear equation system the Gauss-Seidel iteration is used These iteration is optimal to solve under-determined matrices with uncertain coefficients
- Gauss-Seidel always converges if matrix is (main) diagonal dominant $\rightarrow$ modify the equation system with respect to these issue
- Main feature of Gauss-Seidel iteration:
  $x_i^k = \frac{b_i - \sum_{j<i} a_{ij} x_j^{(k)} - \sum_{j>i} a_{ij} x_j^{(k-1)}}{a_{ii}}$
  Firstly, the computations appear to be serial. Since each component of the new iterate depends upon all previously computed components, the updates cannot be done simultaneously
  Secondly, the new iterate $x^{(k)}$ depends upon the order in which the equations are examined. If this ordering is changed, the components of the new iterates (and not just their order) will also change.
  (source: *http://mathworld.wolfram.com/Gauss-SeidelMethod.html*)

Solution for linear equation system

- Working example: all coefficients = 1 (the fake rates), all values for signal + background = 100
- Need six different Tau IDs (cut based, multi variable techniques for different medium and tight selection)
- Make matrix diagonal dominant → solution converges

$$
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{pmatrix}
\times
\begin{pmatrix}
S1 = 100 \\ S2 = 100 \\ S3 = 100 \\ S4 = 100 \\ S5 = 100 \\ S6 = 100 \\ QCD\_OS = 100 \\ W\_OS = 100 \\ Z\_OS = 100 \\ QCD\_SS = 100 \\ W\_SS = 100 \\ Z\_SS = 100
\end{pmatrix}
=
\begin{pmatrix}
ID1\_OS = 400 \\ ID2\_OS = 400 \\ ID3\_OS = 400 \\ ID4\_OS = 400 \\ ID5\_OS = 400 \\ ID6\_OS = 400 \\ ID1\_SS = 300 \\ ID2\_SS = 300 \\ ID3\_SS = 300 \\ ID4\_SS = 300 \\ ID5\_SS = 300 \\ ID6\_SS = 300
\end{pmatrix}
\quad (1)
$$

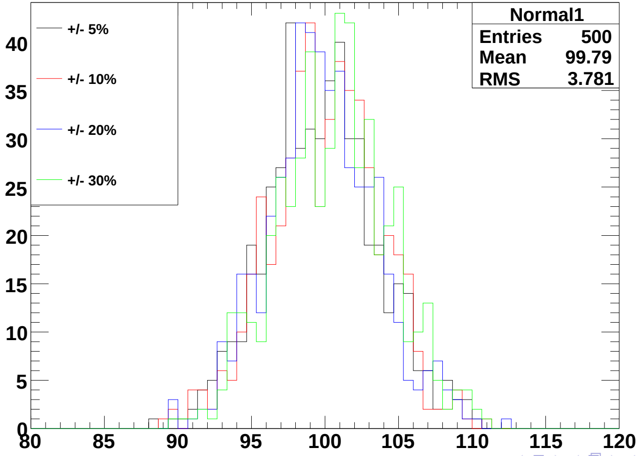
Solution for linear equation system

- Assume that the correct starting vector $\vec{x}$ was found
  Note: the correct solution is the solution were the starting vector yields the minimized error and minimized number of iteration steps (if the matrix is diagonal dominant, all starting vectors will result into the correct solution plus an uncertainty)
- Apply statistical error on $\vec{b}$ and run over different uncertainties on coefficients to study the effect of errors for the solution of the equation system
  Important: equation systems strongly sensitive to uncertain coefficients!
- Apply gauss normal distribution and solve equation system n times

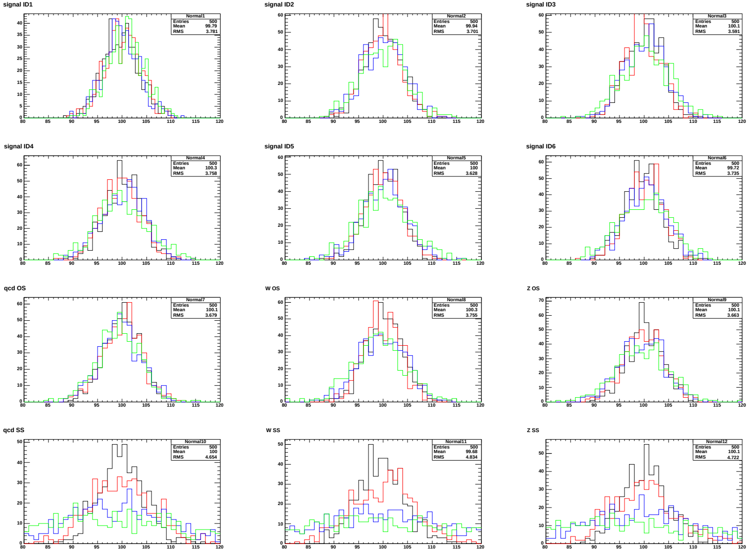
Uncertainty for signal and background

- Solution for different levels of uncertainties coefficient matrix(5% to 30%)
- The number of signal for one Tau ID (first entree in $\vec{x}$ is relatively stable)

signal ID1



Uncertainty for signal and background (all 12 values in $\vec{x}$)



Further studies

- Method can be modified for different studies
  a) Apply OS/SS rescaling factors to reduce the number of equations and determine S/B on reconstruction level
  b) Determine relative Tau efficiencies and S/B for $W \to \tau\nu$
  c) Determine visible mass shapes with respect to the S/B for $Z \to \tau\tau$

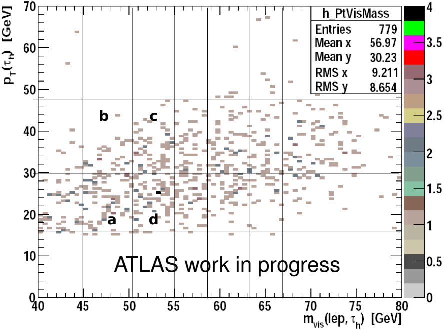
a) Apply OS/SS rescaling factors to reduce the number of equations and determine S/B on reconstruction level

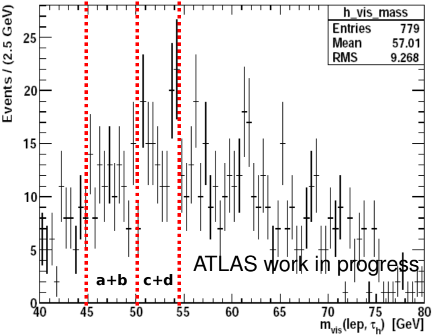
- The two equations for a certain Tau ID can be expressed as
  I) OS = S1 + QCD_{OS} + W_{OS} + Z_{OS}
  II) SS = a $\times$ QCD_{OS} +b $\times$ W_{OS} + c $\times$ Z_{OS}
- with a,b, and c as the SS rescaling factor (obtained from control regions)
- good cross check of OS/SS method and this method
- If number of equations is reduced $\rightarrow$ new variables can be determined
- Use the reconstruction fake rates to determine S/B before applying Tau ID
  I) OS = S1 + QCD_{OS} + W_{OS} + Z_{OS}
  II) SS = X + a $\times$ QCD_{OS} +b $\times$ W_{OS} + c $\times$ Z_{OS}
- X expresses the effect from charge mis-identified Taus (signal in the SS region)
- Possible to determine S/B on reconstruction level optimizes the ID efficiency determination

b) Determine relative Tau efficiencies and S/B for $W \rightarrow \tau\nu$

- Since all values can be determined it is also possible to estimate e.g. the relative efficiencies
  $\frac{\varepsilon^{ID}_{tight}}{\varepsilon^{ID}_{medium}}$ = S1/S2
- Can be used for S/B background estimation in $W \rightarrow \tau\nu$ channel
- Define the following equation system
  I) (ID 1)= S1 + QCD + W
  II) (ID 2)= a $\times$ S1 + A $\times$ QCD + B $\times$ W
  III) (ID 3)= b $\times$ S1 + C $\times$ QCD + D $\times$ W
  IV) (ID 4)= c $\times$ S1 + E $\times$ QCD + F $\times$ W
- With the fake rates A..F and the relative efficiencies a,b, and c
- Note: for the $W \rightarrow \ell\nu$ we need also $\ell- > tau + jet- > tau$ fake rates as described previously

c) Determine visible mass shapes for $Z \rightarrow \tau\tau$





- Fake rates are Jet p_T dependend, so S/B for OS could be:
  region a=Signal+QCD+W+Z
  region b=4*Signal+2*QCD+1.5*W+3*Z
- Interpolate the S/B for different regions to the visible mass window

Summary and Outlook

Summary

- The described method uses the fake rates for gluon and quark jets
- Background estimation from data without MC input possible
- Uncertainties under control
- Use results from different fake rate studies

Outlook

- Summarize all available fake rates
- Then performance on data
- Compare with results from OS/SS and ABCD methods

Additional information: expected fake rates

- **Prilimanary** study from Almut Pingel on data
- Just to show that different Tau IDs have different fake rates $\pm$ uncertainty

ATLAS work in progress				
tight	fake-rates	fake-rates	fake-rates	fake-rates
probe Jet p_T [GeV]	30.0 - 40.0	40.0 - 50.0	50.0 - 60.0	60.0 - 70.0
cutbased	0.044567	0.060821	0.068628	0.079702
BDT	0.015621	0.023945	0.025210	0.033244
LLH	0.024910	0.018198	0.011207	0.011567
medium	fake-rates	fake-rates	fake-rates	fake-rates
probe Jet p_T [GeV]	30.0 - 40.0	40.0 - 50.0	50.0 - 60.0	60.0 - 70.0
cutbased	0.177852	0.219443	0.236965	0.2470
BDT	0.046168	0.062284	0.062925	0.069973
LLH	0.059841	0.045256	0.030376	0.028141
ATLAS work in progress				

Additional information: effect from correlations

- Chose CutBased tight and CutBased medium (e.g. QCD fake rate)
  Let assume we have for OS
  CutBased tight: 100 QCD events
  and for
  CutBased medium: 250 QCD events
- Of course all 100 events in tight are also in medium.
- But also fake rates are correlated. That means, the fake rate studies should give a result like FR(medium)/FR(tight) = 2.5 $\pm$ uncertainty
- Also all correlated uncertainties have to be considered