GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

BMBF-Forschungsschwerpunkt "Elementarteilchenphysik mit dem CMS-Experiment"

Physik an der TeV-Skala mit dem Large Hadron Collider

Compact Muon Solenoid

FSP102

RooStatsCms

Karlsruhe Institute of Technology

PHYSICS AT THE TERASCALE
Strategic Helmholtz Alliance

HELMHOLTZ GEMEINSCHAFT

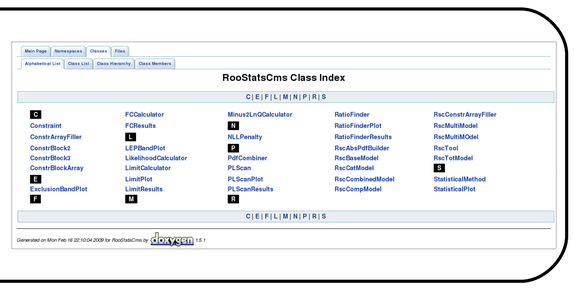
A tool for analysis modelling, combination and statistical studies

Danilo Piparo, Grégory Schott, Günter Quast – Universität Karlsruhe

danilo.piparo@cern.ch, gregory.schott@cern.ch

Introduction

- **RooStatsCMS (RSC)** is a software framework based on RooFit technology.
- Scope: modelling and combination of analysis channels and statistical studies.
- Classes design: splitting in CPU intensive jobs on batch systems or on the GRID.
- Analyses, systematics and correlations characterised in configuration files.
- Analyses combination: sharper limits and earlier discoveries.



RooFit

- Toolkit for data modelling, fitting and toy MC studies
- Collection of basic Pdfs, allows composition of those for complex models
- Part of ROOT standard release
- See RooFit poster in this session!

Modelling and combination

Models creation as RooFit objects is carried out with an advanced model factory

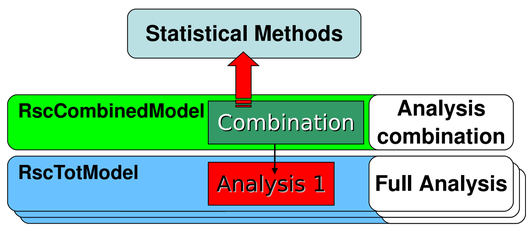
Describe the model in an ASCII configuration file (*datacard*). With 2 lines of C++ code build the objects. This innovative approach has three advantages: easy results sharing, factorise analyses modelling from C++ code and usage of the same model with the different statistical methods at disposal. Inspection of the model is possible with diagrams (*graphviz*) and website creation.

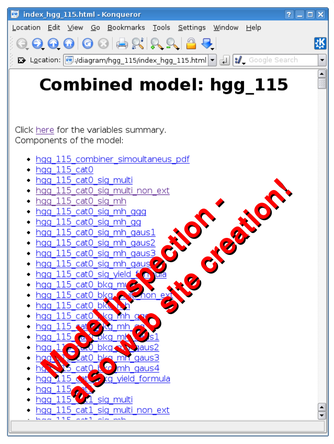
```
# The combined model
[hzz4l]
   model = combined
   components = hzz_4mu, hzz_4e

# H -> ZZ -> 4mu
[hzz_4mu]
   variables = x // Dummy for a counting analysis
   x = 0 L(0 – 1)

[hzz_4mu_sig]
   hzz_4mu_sig_yield = 62.78 L(0 - 200)

[hzz_4mu_sig_x]
   model = yieldonly

[hzz_4mu_bkg]
   yield_factors_number = 2

   yield_factor_1 = scale
   scale = 1 C L (0 - 3)
   scale_constraint = Gaussian,1,0.041

   yield_factor_2 = bkg_4mu
   bkg_4mu = 19.93 C

[hzz_4mu_bkg_x]
   model = yieldonly
……
```

DataCard

```
RscCombinedModel mymodel ("hzz4l");
RooAbsPdf* sb_pdf=mymodel.getPdf();
```

C++



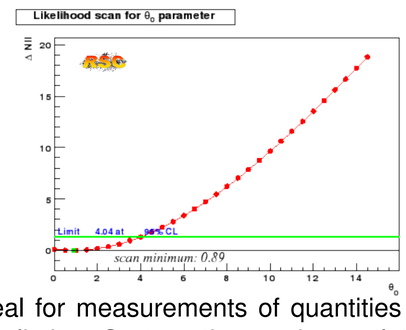


Model inspection - also web site creation!

Statistical Methods

RooStatsCms provides implementation of different statistical methods to allow determination of confidence or credibility regions, exclusion limits and (expected) significance. Here a selection of them is illustrated:

Profile likelihood method



Ideal for measurements of quantities and for limits. Systematics and correlations treated via profiling.

It can also be used for estimation of significances

Profiling

- Penalty term in the likelihood ($\log L_T = \log L_{base} + \log L_P$) e.g.

  1. One uncorrelated Gaussian constraint
  $\log L_P \sim 0.5\cdot(m-m_0)^2/\sigma_m^2$

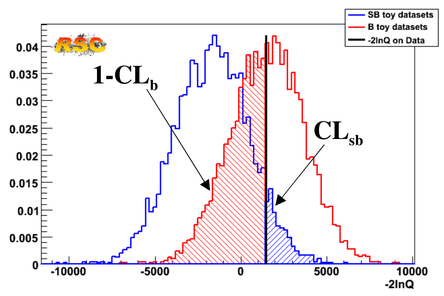
  2. Correlated Gaussian constraints
  $\log L_P \sim 0.5\cdot(m-m_0)^T \cdot V^{-1} \cdot (m-m_0)$, V is correlation matrix

- No toys: a few fits done to get the profile

Marginalisation

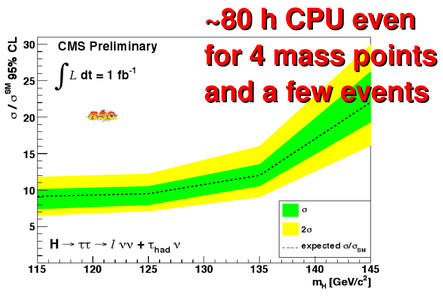
- MC phase-space integration
- Lots of toy experiments

Separation of Hypotheses



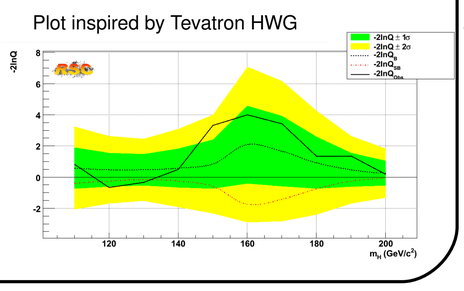
Quantify the separation of a *null hypothesis* (background only) and an *alternative hypothesis* (signal + bkg). Many toy Monte Carlo experiments to be performed: very CPU intensive. Systematics treated via marginalisation.

Formulation of *upper limit estimation* in terms of hypotheses separation



Recreate plot of expected test statistic distributions at different working points

Plot inspired by Tevatron HWG



The structure of the package

Statistical Methods – Mother: StatisticalMethod class

LimitCalculator | PLScan | FCCalculator

Statistical Results – Mother: StatisticalResult class

LimitResults | PLScanResults | FCResults

"Sum" the results: batch/Grid jobs retrieval easier

Statistical Plot – Mother: StatisticalPlot class

LimitPlot | PLScanPlot (add also FC curves)

Lots of jobs, almost no I/O, CPU intensive: Grid is the ideal solution!

Grid

Local batch systems

Classes holding constraints and correlations among them:

- Constraint
- ConstrBlock
- ConstrBlockArray

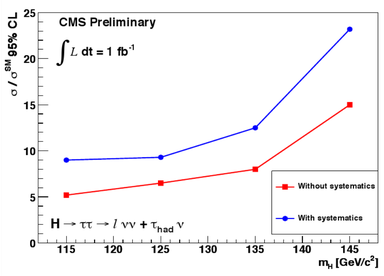
} Mother: NLLPenalty

More advanced graphics routines and helper classes:

- LEPBandPlot
- ExclusionBandPlot
- RatioFinder

CMS Examples

VBF H → ττ



HIG-08-008

M	N Sig (12% sys)	N Bkg (30% sys)
115	1.6	45.2
125	1.4	45.2
135	1.1	45.2
145	0.6	45.2

- Vector Boson Fusion channel
- Integrated luminosity: 1 fb^{-1} @ 14 TeV
- Small signal on a significant background

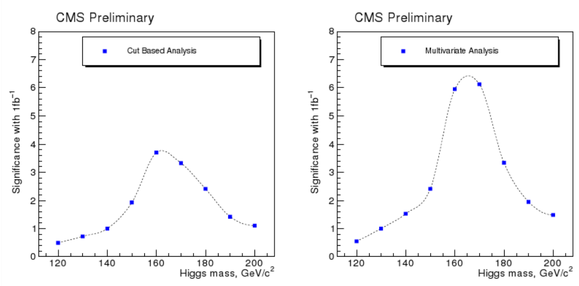
No discovery expected with this integrated luminosity

Which model can be excluded with the data accumulated?

Formulate problem in terms of hypotheses separation

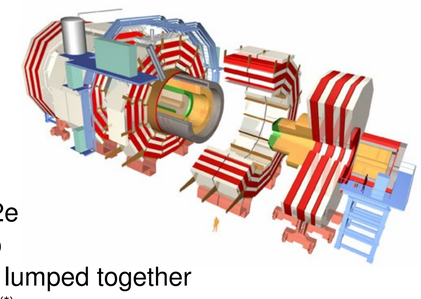
→ Use modified frequentist method increasing expected signal events number until „enough" separation is reached (CL_s=5%)

H → WW(*) → lνlν

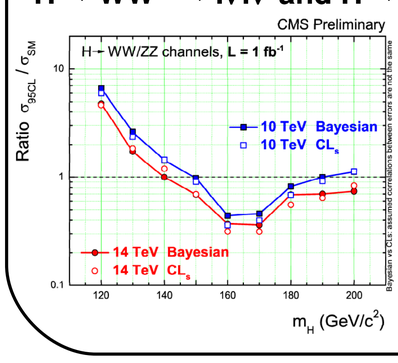


HIG-08-006

- Combination: 3 channels - ee, eμ, μμ
- Cut based and neural network analyses
- Integrated luminosity: 1 fb^{-1} @ 14 TeV
- Discovery: expected significance calculated with profile likelihood



H → WW(*) → lνlν and H → ZZ(*) → 4l



- H → ZZ(*) → 4l: 4μ, 4e, 2μ2e
  - 3 modes: similar S/B ratio
  - Signal/Background yields lumped together
- Combination with H → WW(*)
- Cut and count analyses
- CM energies studied: 14 and 10 TeV (1fb^{-1})
- CL_s technique and Bayesian approach compared
- Good agreement despite slightly different model of correlations and statistical methods

Further developments

- Further developments in RooStats project in collaboration with ATLAS
- RooStats is part of ROOT - see Lorenzo Moneta's talk!
  - Classes factorisation suggested by RooStatsCms structure
  - *LimitCalculator* imported under the name of *HybridCalculator* for ROOT 5.22 release
  - High level model factory import ongoing
- Interfacing to BAT (Bayesian Analysis Toolkit) in *BCRooInterface* class

BAT

References:

- RooStatsCms Website: www-ekp.physik.uni-karlsruhe.de/~RooStatsCms
- CMS Physics Analysis Note: cms-physics.web.cern.ch/cms-physics/public/HIG-08-008-pas.pdf
- CMS Physics Analysis Note: cms-physics.web.cern.ch/cms-physics/public/HIG-08-006-pas.pdf
- ROOT: http://root.cern.ch/
- RooFit: http://roofit.sourceforge.net/
- RooStats: https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome
- BAT: http://www.mppmu.mpg.de/bat