Multichannel number counting experiments

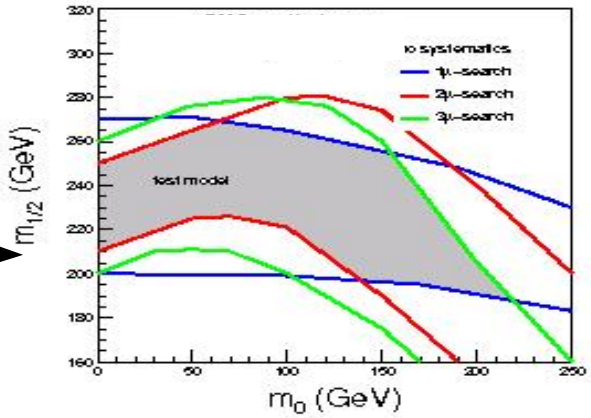
V.Zhukov
M.Bonsch

KIT, Karlsruhe

Multichannel searches at LHC

Beyond SM(eg. SUSY) can manifest in different signal topologies.
LHC detectors can identify different objects (lepton, jets, MET, etc)
→ consider different **exclusive** topologies 1) independently,
and then 2) combine them.

Eg. mSUGRA 95%CL for different topologies



Combination has clear benefits:

- Extend statistics by combining many (hundredths) searches with relatively low signal efficiency each, increases sensitivity.
- Consistent treatment of systematics
- Consistency check of different measurements can be used to constrain H_1 model parameters

… and challenges:

- How to combine/split topologies in presence of correlated systematics?
- How to optimize selection(S/B) for each channel for optimum combination?
- Which method to use for confidence intervals, hypothesis testing?
- Effect of the initial and boundary conditions (zero observation, unphysical systematics, truncation, range for common signal, flip-flop for combinations)
- When it make sense to use auxiliary measurements and how to combine them?....

Framework

- **Use RooFit/Stats framework in root_v27.6 *)**
  Likelihood formalism, common workspace...

- **Trust the RooStat implementations.**
  Experimentalist approach, no attempt to improve the existing codes.
  Bugs(features) are not excluded...

- **Simple number counting test model to check basic behavior of different methods.**
  Run thousands jobs on cluster.
  Show some part of results...

*) Thanks to RooStat developers and CMS Stat committee for consultations

Statistical model (as in RooFit)

Single channel (simple model for demonstration):

N_{obs} observation

seff efficiency of signal s=[0, 3*(Nobs+Nbkg)]

N_{bkg} background expectation

sigmb relative background uncertainties.

Here consider truncated Gaussian b' =[0., 5*sigmb].

Other shapes(Lognormal, Gamma) have similar qualitative behavior.

*Poisson::signal(Nobs, s*seff+b')*
Gaussian::sysb(b', Nbkg, sigmb)
("nuis", " b' ")
L(n|x)= PROD::model1(signal,sysb)

Combined model:

Systematics: use either individual nuisance or common systematics

L(n|x)=
PROD::combmodel(model1,model2,....)
("nuis","b1', b2',...")
("nuis", " b' ") //common systematics

Auxiliary measurement:

'Data-driven' background estimation.
Constrain background in the signal region by auxiliary measurements.
Introduce extra Poisson term and systematics on *tau=b/c* relating signal(b) and sideband control regions (c)

*Poisson::signal(Nobs, s*effs+c'*tau')*
Poisson::aux(Naux, c')
Gaussian::systau(tau ', tau, sigmtau)
("nuis", " c' , tau' ")
L(n|x)= PROD::model(signal, aux, systau)

Statistical methods

Profile Liklehood (PLC):

Based on Lhood ratio and Wilk's theorem

Minuit for nuisance

```
ProfileLikelihoodCalculator plc(*data, *smodel);
 plc.SetConfidenceLevel(cls);
 LikelihoodInterval* plInt = plc.GetInterval();
 pl_L= plInt->LowerLimit( *w->var("s") );
 pl_U= plInt->UpperLimit( *w->var("s") );
HypoTestResult* plh=plc.GetHypoTest();
pl_sig=plh->Significance();
```

Bayes credible intervals (Bayes):

use flat prior on signal here

Numerical integration

```
BayesianCalculator bc(*data, *smodel);
 bc.SetTestSize(1.-cls);
 bc.SetLeftSideTailFraction(0.5); //0 for central
 SimpleInterval* bInt = bc.GetInterval();
 bayes_L= bInt->LowerLimit( );
 bayes_U= bInt->UpperLimit( );
```

Binominal significance Z_Bi

Correspondance of *on/off* and *sigmb* problem

Analytical for single channel(arx.0702156)

```
double tau=_Nbkg/(sigmb*Nbkg*sigmb*Nbkg);
double noff=tau*_Nbkg;
double pBi=TMath::
BetaIncomplete(1/(1.+tau),_Nobs,noff+1.);
double Z_Bi=sqrt(2.)*TMath::ErfInverse(1-2.*pBi);
```

Unified, Feldman-Cousins(FC):

Neyman construction with ordering

Minuit for nuisances

```
FeldmanCousins fc(*data, *smodel);
 fc.SetConfidenceLevel(cls);
  fc.FluctuateNumDataEntries(false);
  fc.UseAdaptiveSampling(true);
  fc.SetNBins(100);
  PointSetInterval*fcInt=(PointSetInterval*)
  fc.GetInterval();
   fc_L= fcInt->LowerLimit( *w->var("s")  );
   fc_U= fcInt->UpperLimit( *w->var("s")   );
```

Hybrid (Hyb)

modified Cousins-Highland.

MC toys for nuisance integration, use CLs

```
HybridCalculatorOriginal hyb(*data,*smodel,*bmodel);
 hyb.PatchSetExtended(false);
 hyb.SetTestStatistic(1);
 hyb.UseNuisance(true);
 hyb.SetNuisancePdf(*w->pdf("prior_nuis"));
 hyb.SetNuisanceParameters(*w->set("nuis"));
 HypoTestInverter myInv(hyb,s);
myInv.UseCLs(true); myInv.SetTestSize(1.0-cls);
hyb.SetNumberOfToys(5000);
myInv.RunAutoScan(lr1,lr2,myInv.Size()/2.,0.01,1);
HypoTestInverterResult* results = myInv.GetInterval();
    hyb_U = results->UpperLimit();
    hyb_L = results->LowerLimit();
 HybridResult* hcResult = hyb.GetHypoTest();
 hyb_significance = hcResult->Significance();
```

RooStat presentations:

Confidence intervals

- Calculate central 95%CL upper and lower limits and one sided upper limit versus Nobs and Nbkg with PLC*), FC and Bayes.
- Test frequentist coverage (Neyman construction) for PLC, FC (should cover) and Bayes credible (not really motivated)

Use different models:

1) Single channel without and with Gaussian rel. systematics $\sigma_b = 0 - 0.5$

   Compare different methods (with different treatment of systematics)

2) Combined *Nch=5* identical channels with *Nobs/Nch, Nbkg/Nch, Seff/Nch* without systematics and the same individual systematics, or correlated. This 'split' combined model should be equivalent to the single channel with the *Nobs* and *Nbkg*.

3) Combined *Nch* channels, but with observations only in one channel *Nobs(1)=Nobs,* others are *Nobs(2,3,4,5..)=0*.

   Check treatment of 'zero' observation.

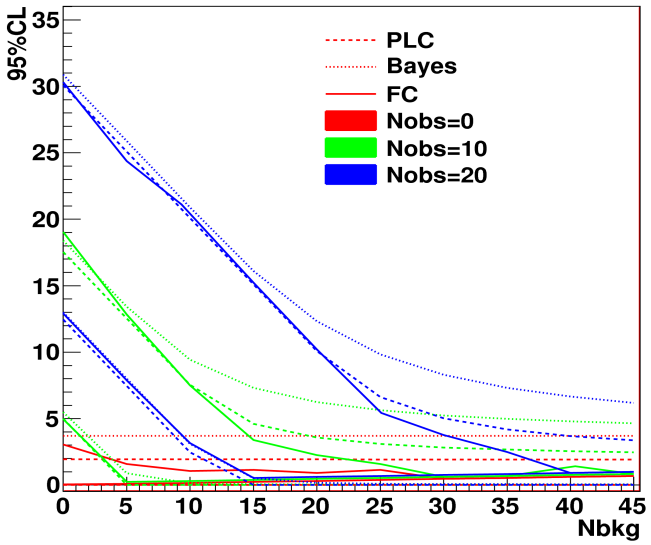
*) upper limits for PLC are calculated with the offset of CL to 90%.

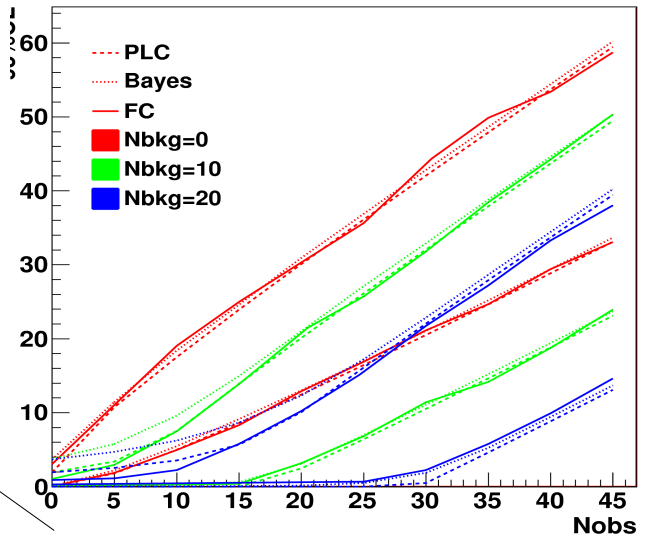
Confidence intervals for single channel

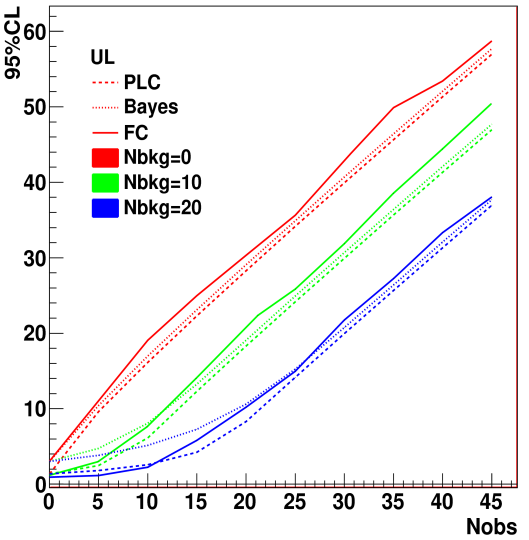
1. CL vs Nbkg, Nobs, σ_b=0 (no systematics)

Central intervals

One sided Upper limit



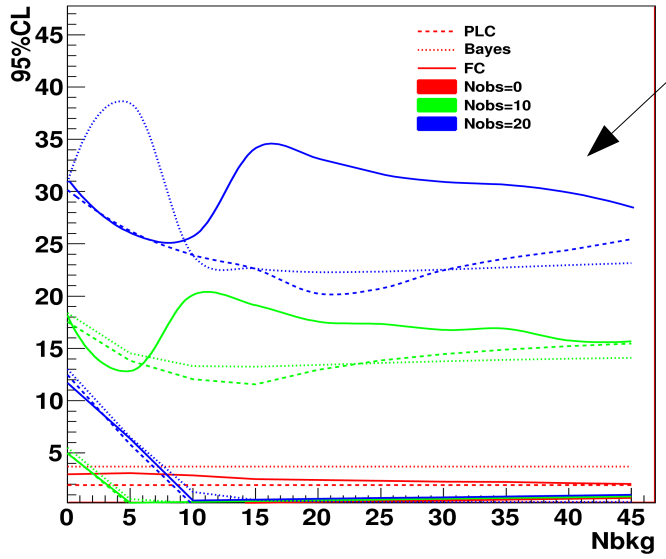




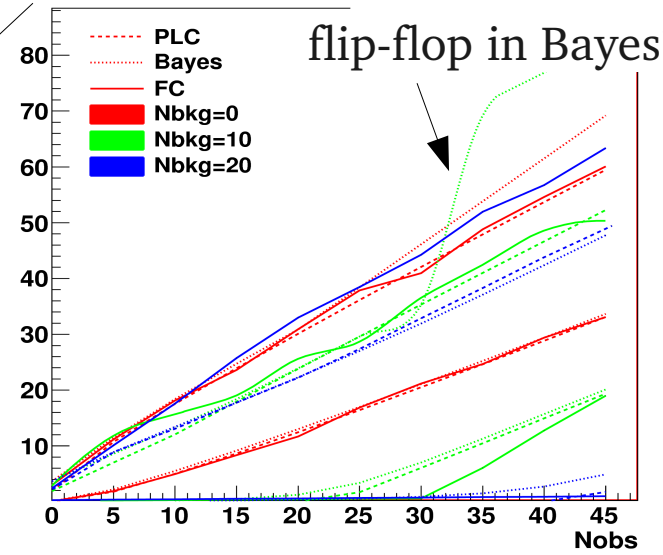
Some differences at large Nbkg>20 small Nobs<20

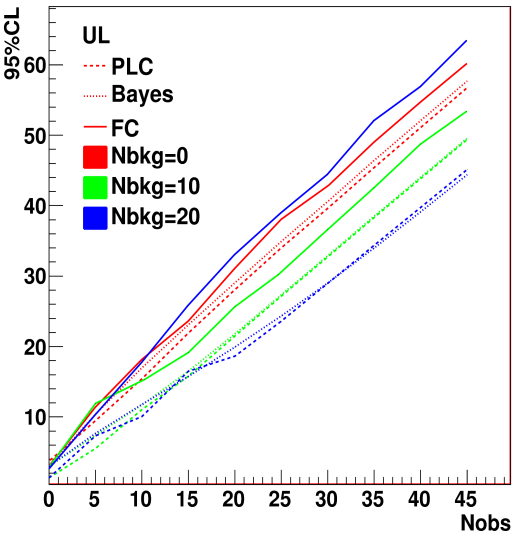
Systematics boundaries?

2. Nbkg, Nobs, σ_b=0.5



flip-flop in Bayes?



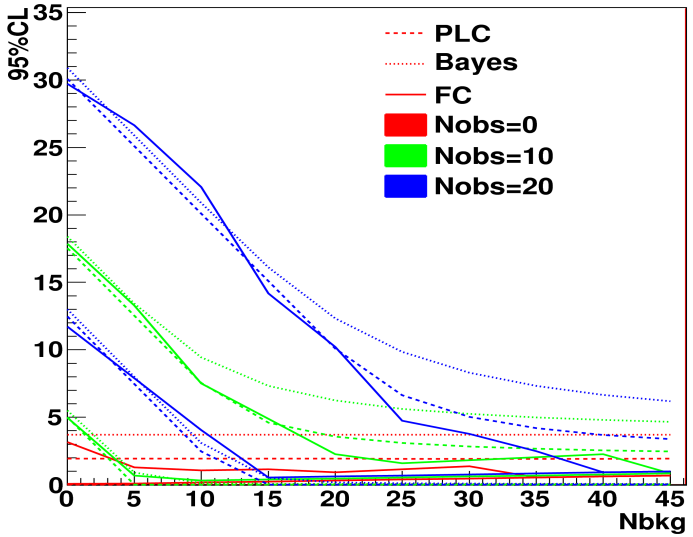


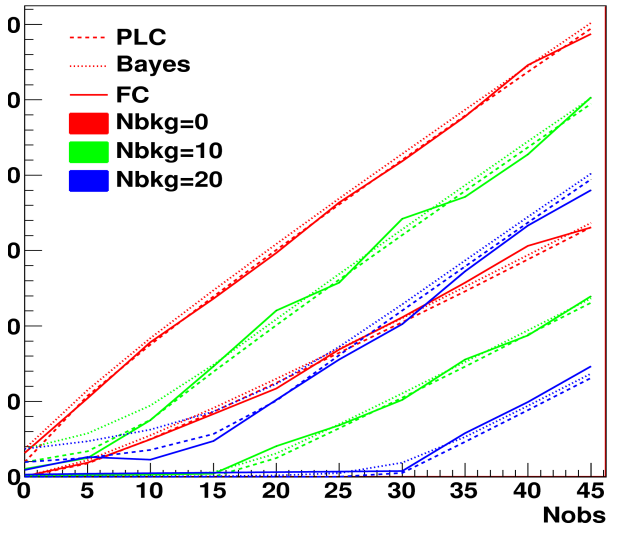
Intervals for the combined model

1. Nbkg, Nobs, σ_b=0 (no systematics)

Central intervals

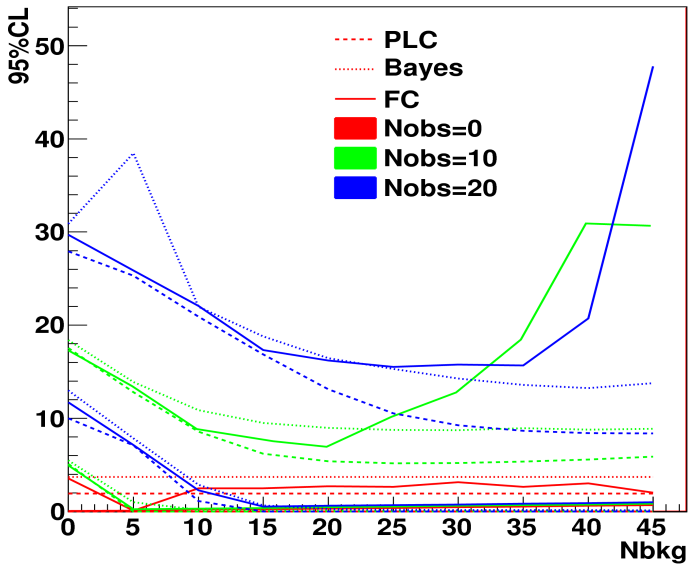
5 identical channels

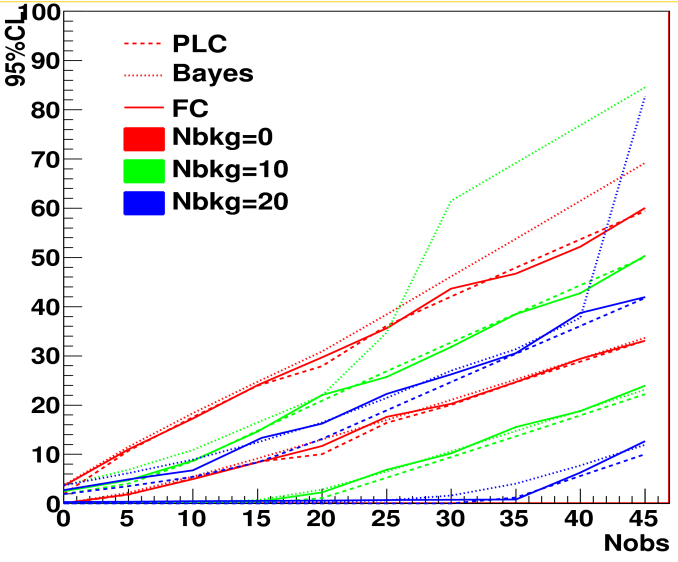




Without systematics:
similar to single channel (as expected)
<u>For large Nbkg:</u>
Methods differ
<u>For Nobs=0:</u>
PLC doesn't work(no Wilks!)
Bayes is not sensitive to bkg
FC improves with large Nbkg

2. Nbkg, Nobs, σ_b=0.5 (non correlated)



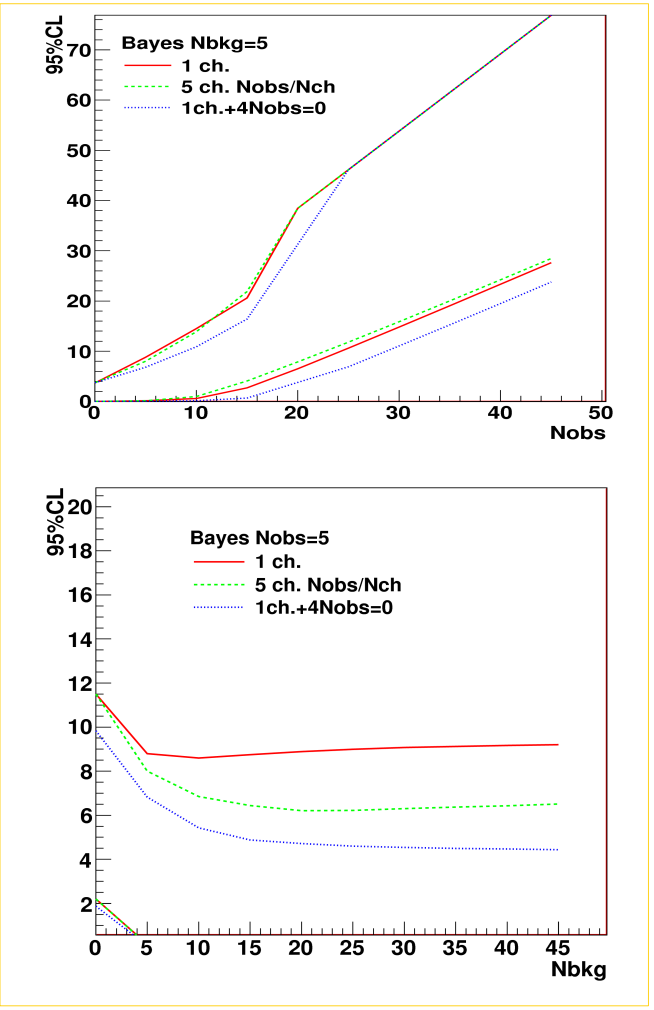


With systematics:
Effect of systematics is changed or moved to higher Nbkg.
Not equivalent to single channel.

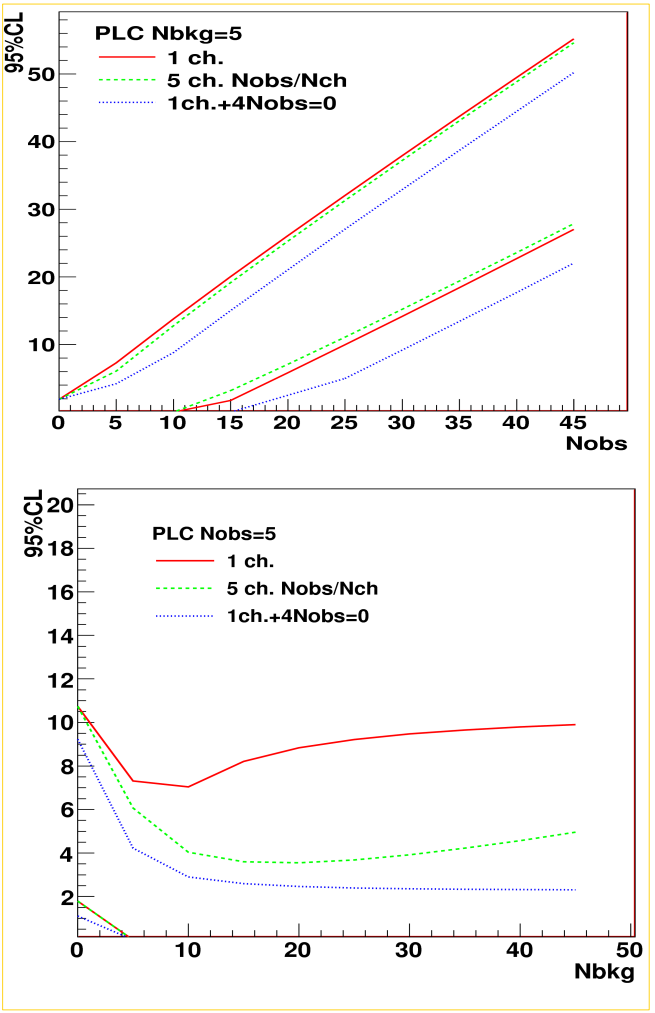
Intervals for the single channel and combined

More comparison (with systematics, without systematics there is no difference)

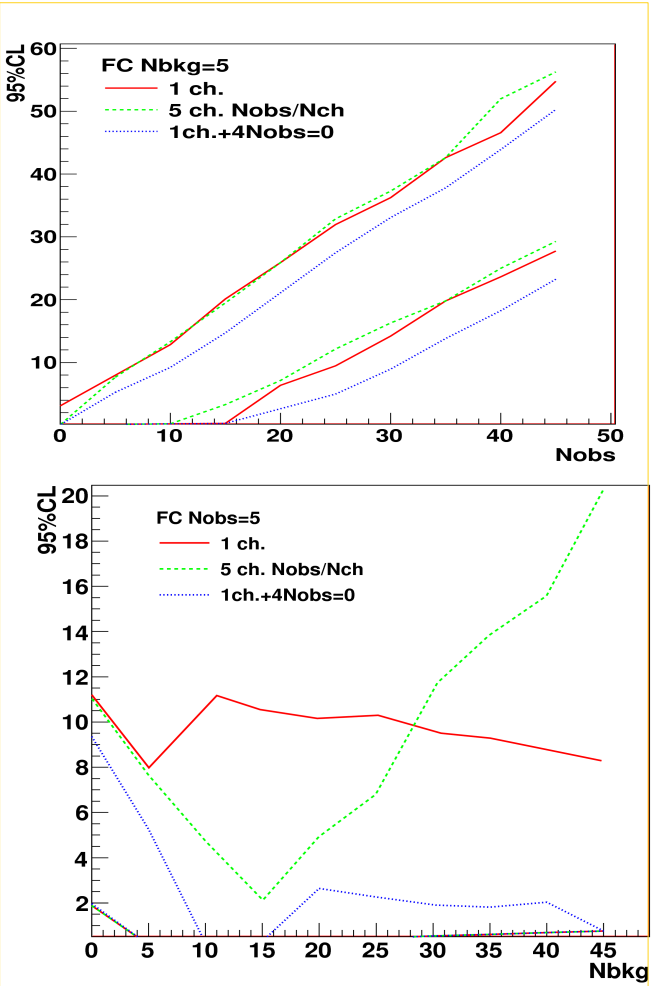
Bayes credible



Profile Lhood



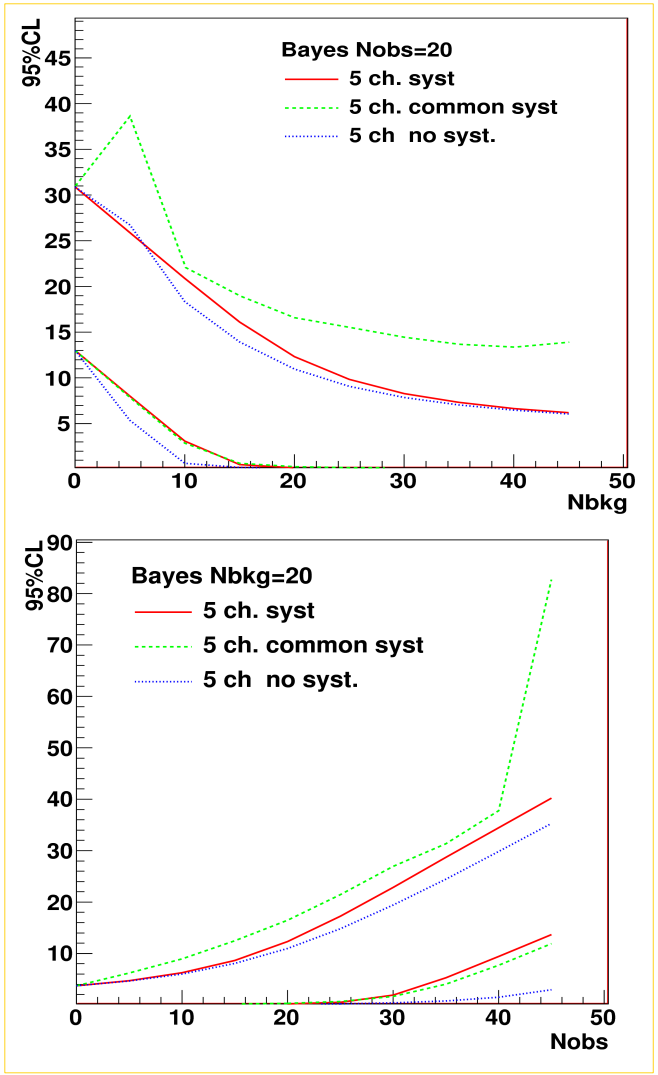
Feldman-Cousins

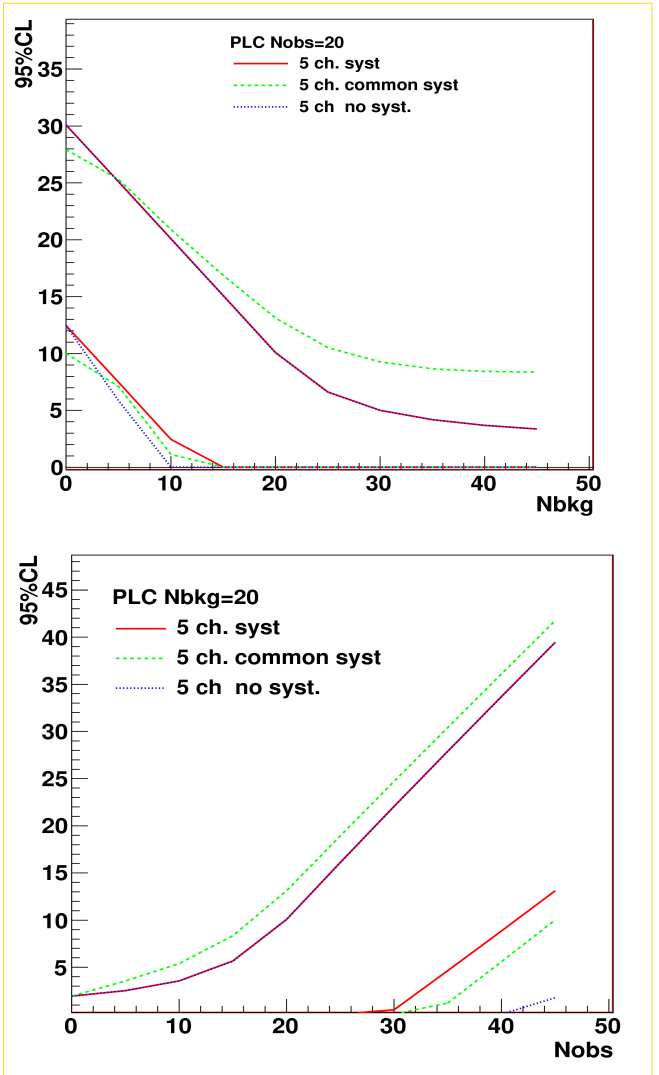


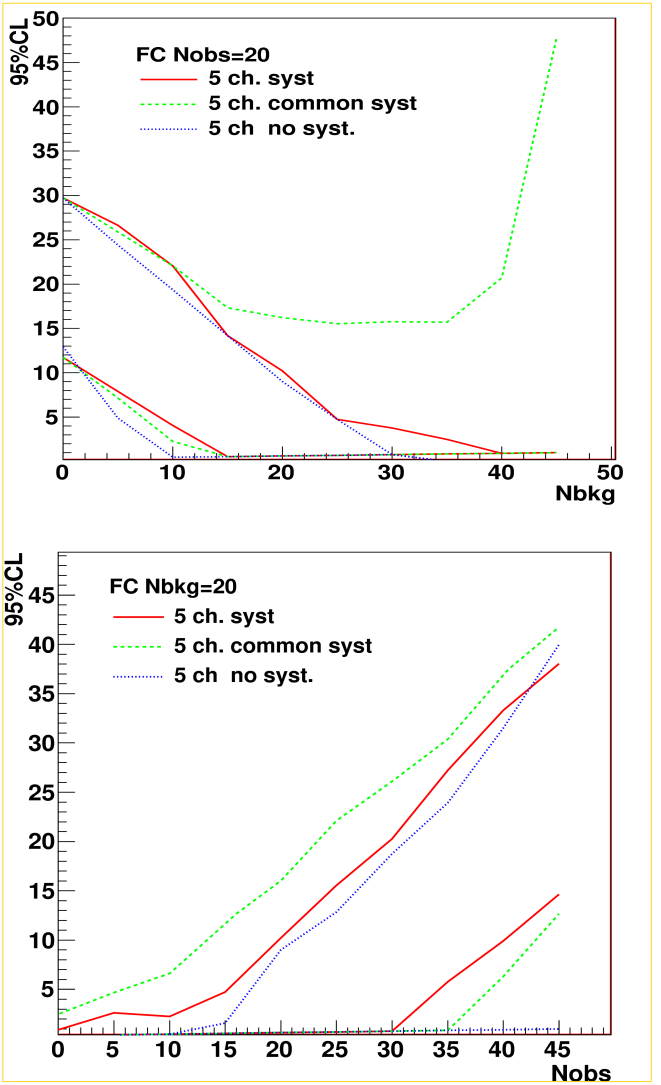
- Good agreement for single channel and split model(5 identical channels) at small Nbkg but significant difference with large Nbkg for all methods, especially FC.
- Tighter limits for the only one channel with observations.

Confidence intervals with correlated systematics

Combined model: 5 channels





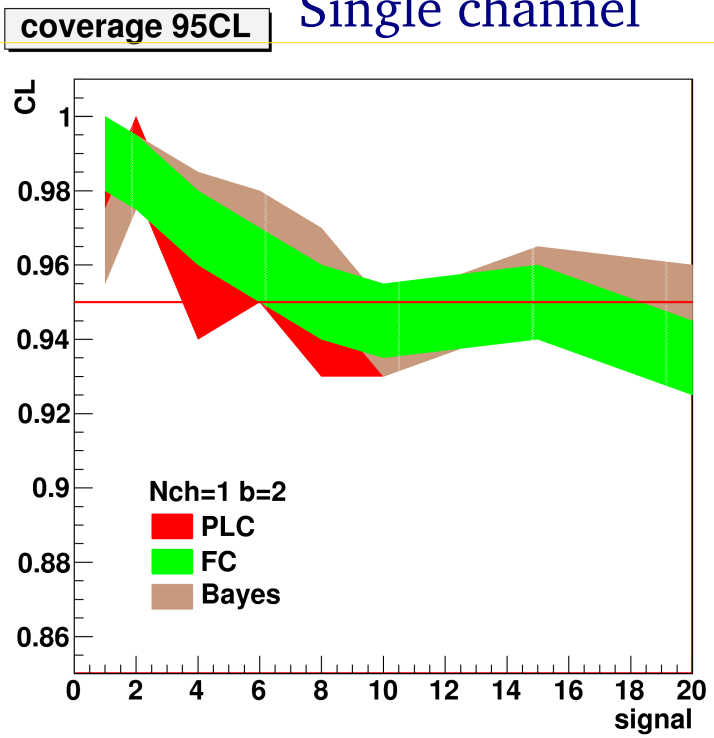


- Correlated systematics can have significant effect in Bayesian and FC intervals at large Nbkg, Nobs and lesser for PLC .

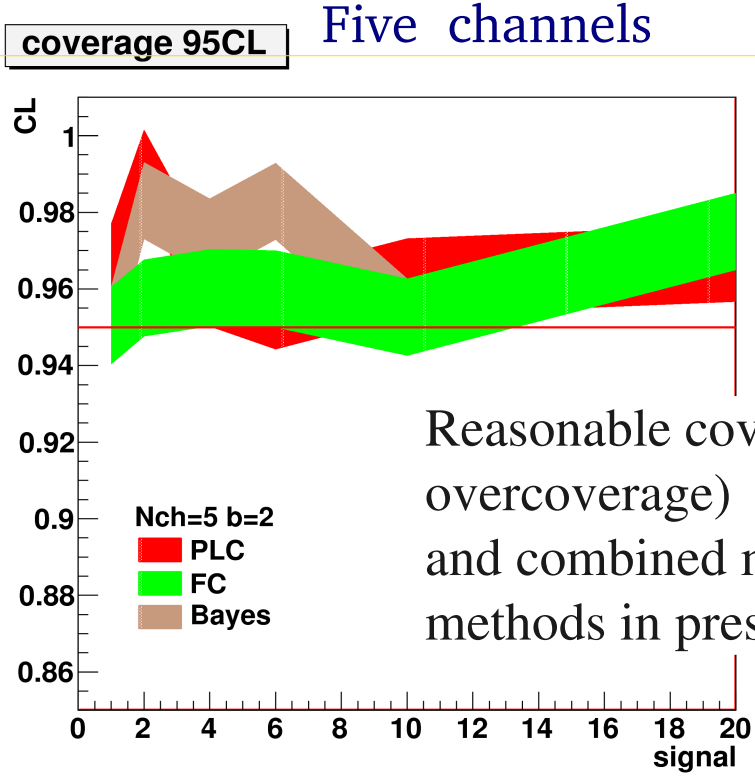
Coverage of central intervals

Single channel

200toys per point

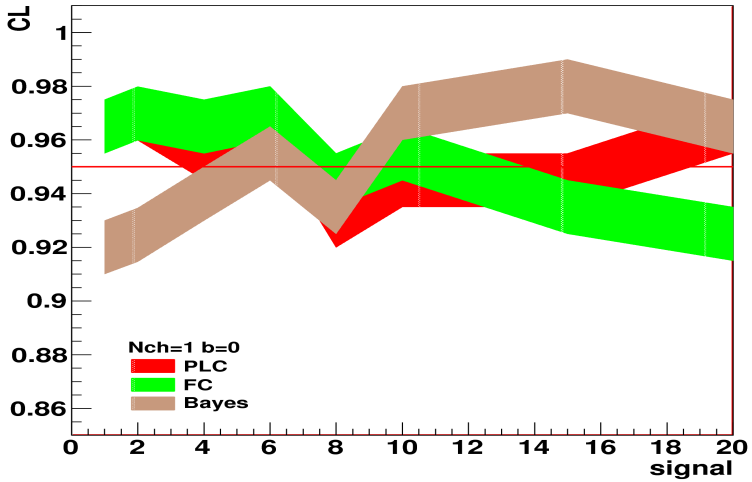


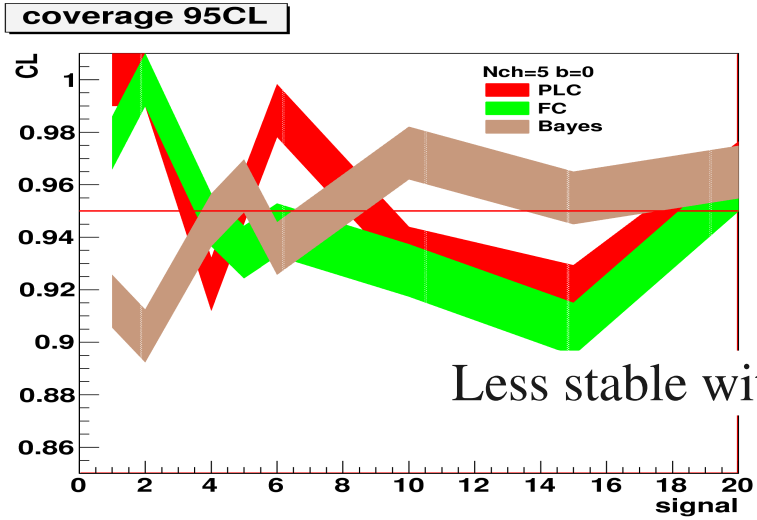
Five channels



Nbkg=2
sigmb=0.5
Signal s=0-20
(s=2 → Z~5)

Reasonable coverage(some overcoverage) in single channel and combined model for all methods in presence of background





Nbkg=0
sigmb=0.5

Less stable without background

Hypothesis testing

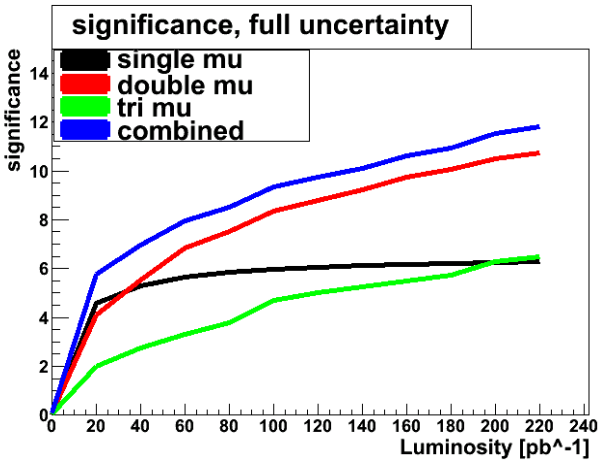
Consider channels with different S/B and different systematics(Gauss, Gamma), calculate significance with PLC, Hybrid, Z_Bi

1) **Saturation behavior**.
Increase of statistics(lumi)

$$Z_{\text{inf}} \rightarrow \frac{S}{B * \sigma_{sys}}$$

Important for combination of channels with different S/B and systematics.
Defines selection optimization strategy for each combination.
Related: 'coverage' of methods used for significance

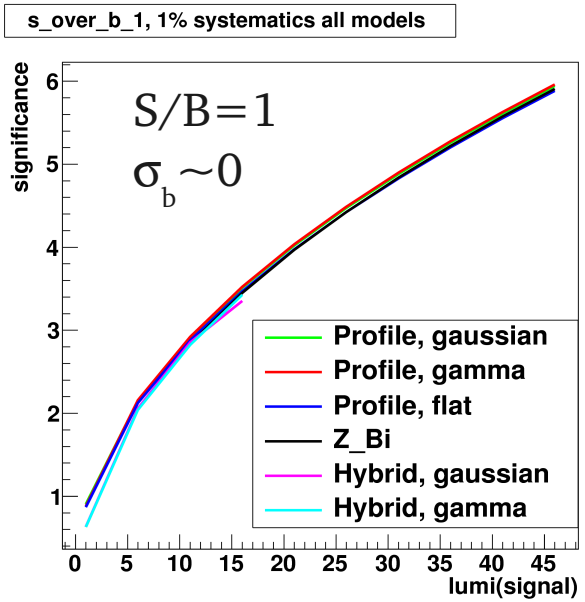
Eg. mSUGRA search

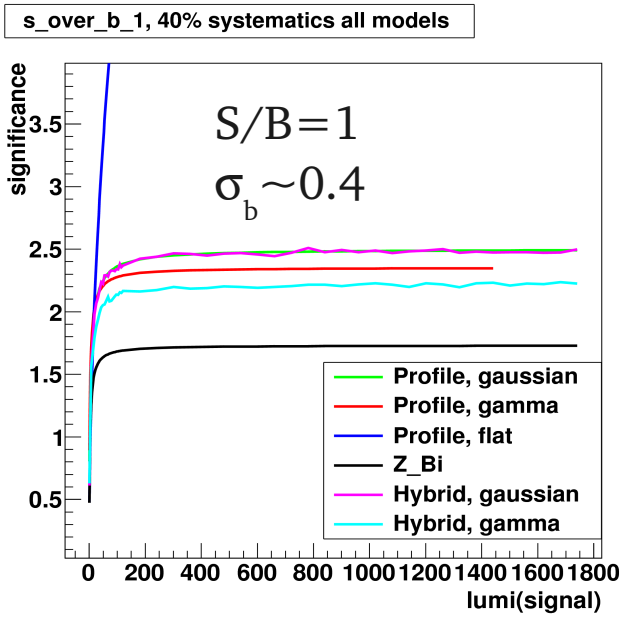


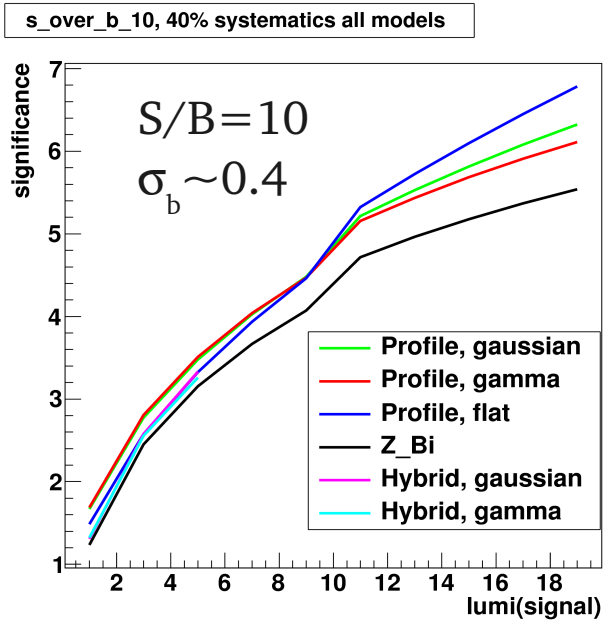
2) **Systematics correlations**.
Correlations can decrease or increase significance of combined model (similar to auxiliary measurements). What are the conditions?

3) **Auxiliary measurements** (data driven background estimation)
Can constrain some model parameters(bkg) with extra measurements
Provided there is some benefit comparing with MC truth uncertainties

Significance vs luminosity

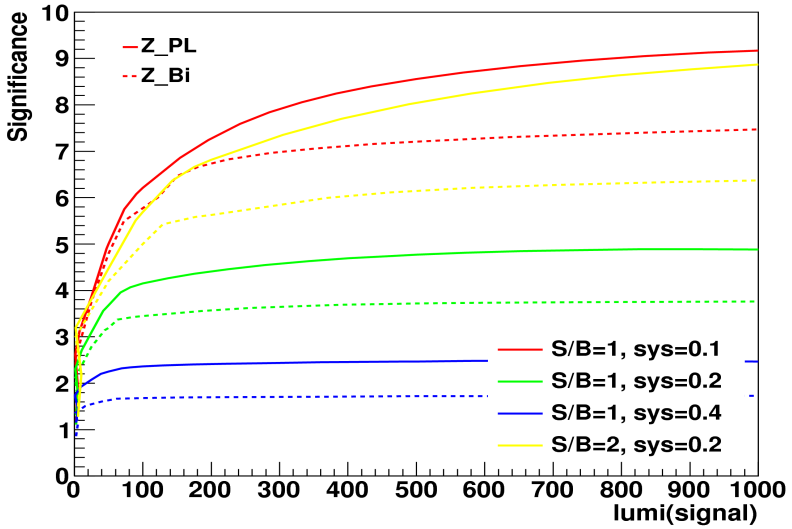






Without systematics:
excellent agreement among methods

With systematics :
Some differences and different asymptotic behavior



Saturation for models with different S/B and systematics.

- Good scaling for PLC, S/B is the right parameters for optimization of selection.
- Z_Bi is always below
  (needs modification in on/off-sigmb formulas)

Correlated systematics

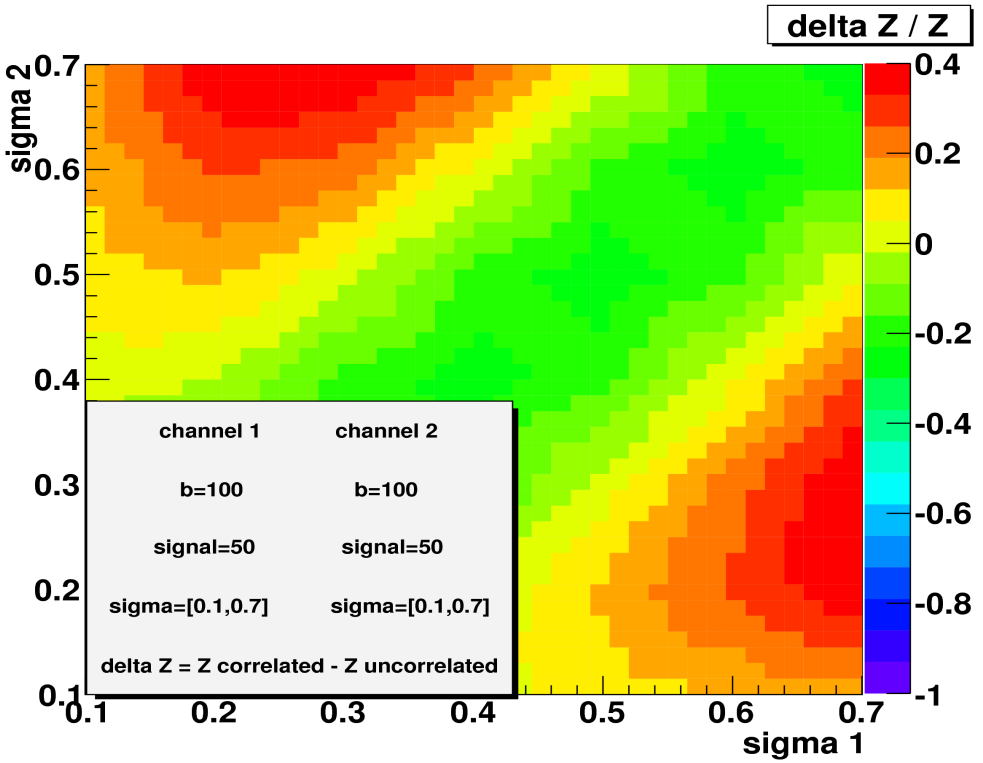
Without correlations the combined significance $Z \sim \Sigma\, z/\sqrt{n}$, with systematics depends...

Eg.

Consider 2 channels with the same signal $s_1 = s_2 = 50$ and bkg $b_1 = b2 = 100$

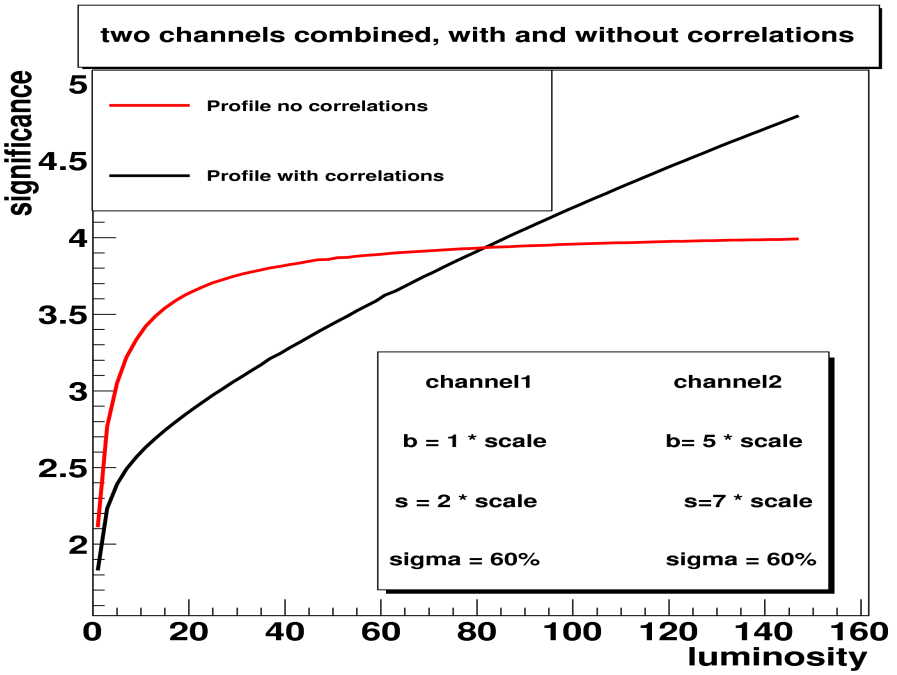
and systematics sigmb=0.1-0.7 correlated (Z_{cor}) and uncorrelated (Z_{uncor}):

$\Delta Z/Z = (Z_{cor} - Z_{uncor})/Z_{uncor}$ (use PLC)



When channels are in disbalance the correlation can be beneficial (red regions) otherwise ~20% decrease(green)

The effect of correlations on significance can change with luminosity



Combined significance of 2 ch. S/B=2 and S/B=1.4 sigmb=0.5 with b1=1 and b2=5.

Auxiliary measurements

Main idea: reduces systematics in absolute normalization, but systematics in shape remains via tau factor τ=bkg-in-signal-region /bkg-in-sideband

Data driven(dd) bkg constrain can be used if the main uncertainties in MC are expected from this absolute normalization.

It can be beneficial when $\sigma_b(MC)^2 > n_{ax} + \sigma_\tau(MC)^2$ (n_{ax} measurement in side band)

And:

Have to compare aux measurements itself with MC predictions anyway.

1. If there is a large difference can't use dd bkg because the σ_τ (MC) is undefined
2. If there are no difference why we need it?

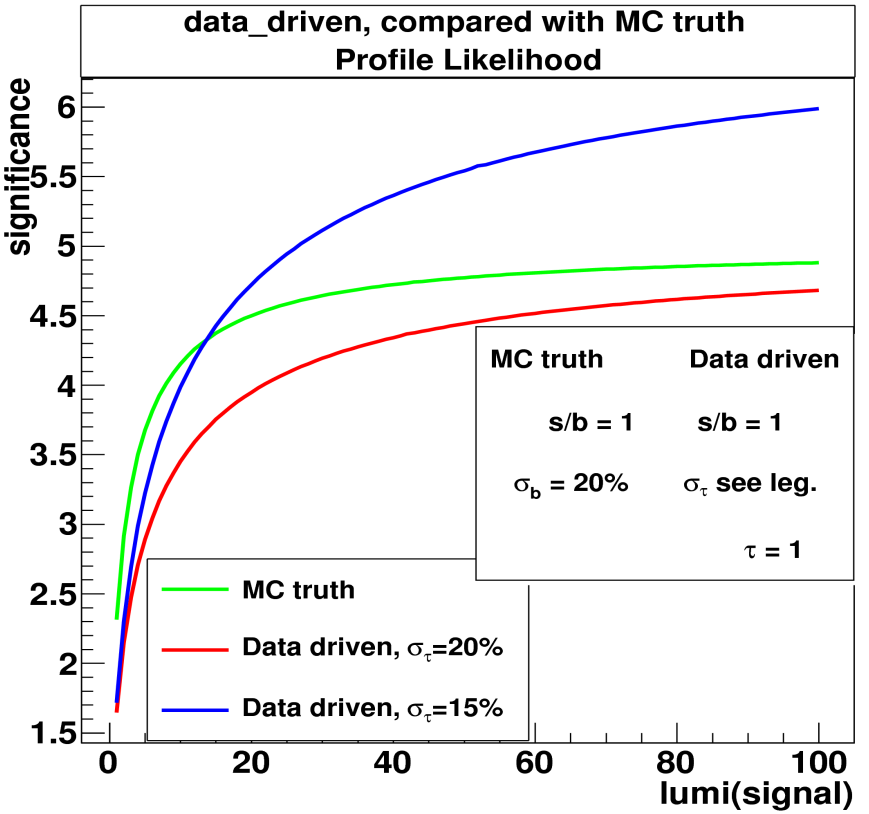
Better (and more correct) to use all measurements in G.O.F test to tune the MC.

Eg.

Consider 3 models (S/B=1)

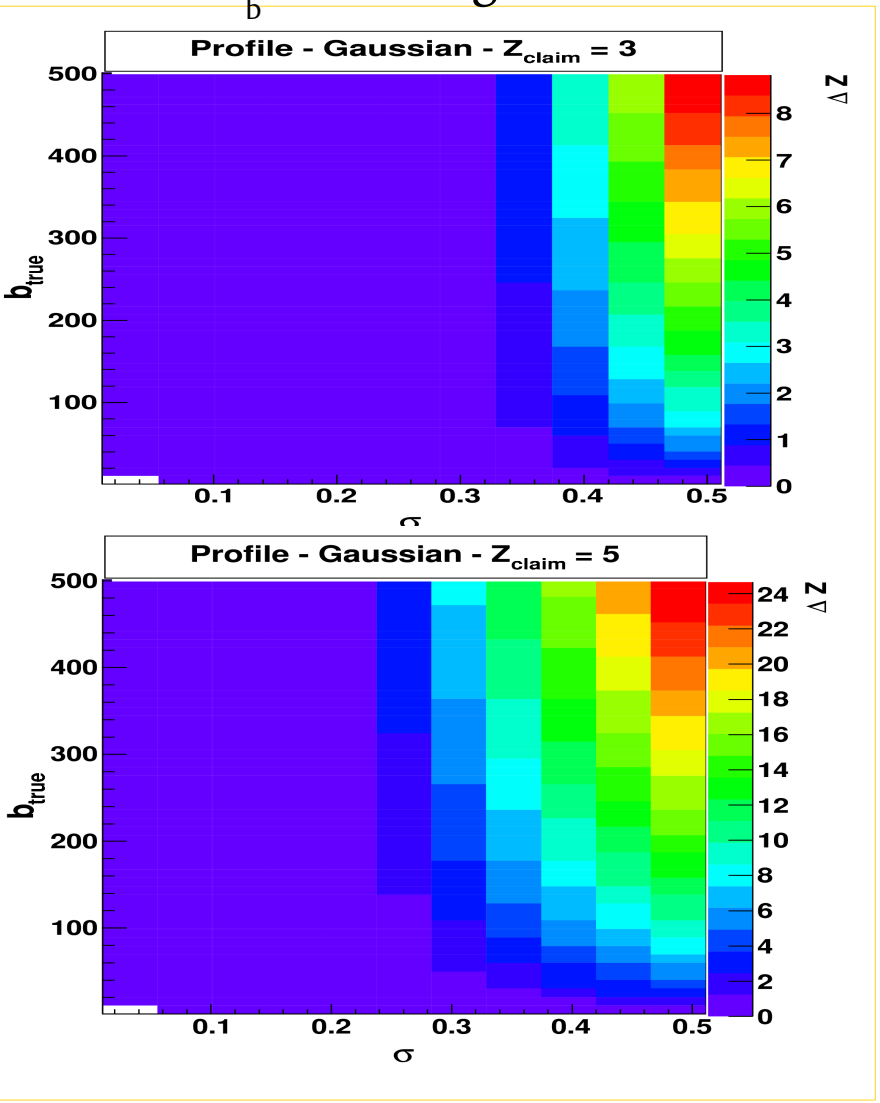
1. $\sigma_b(MC)=0.2$
2. dd bkg. with $\sigma\tau=0.2$
3. dd bkg with $\sigma\tau=0.15$

Evolution of significance (PLC) with luminosity.

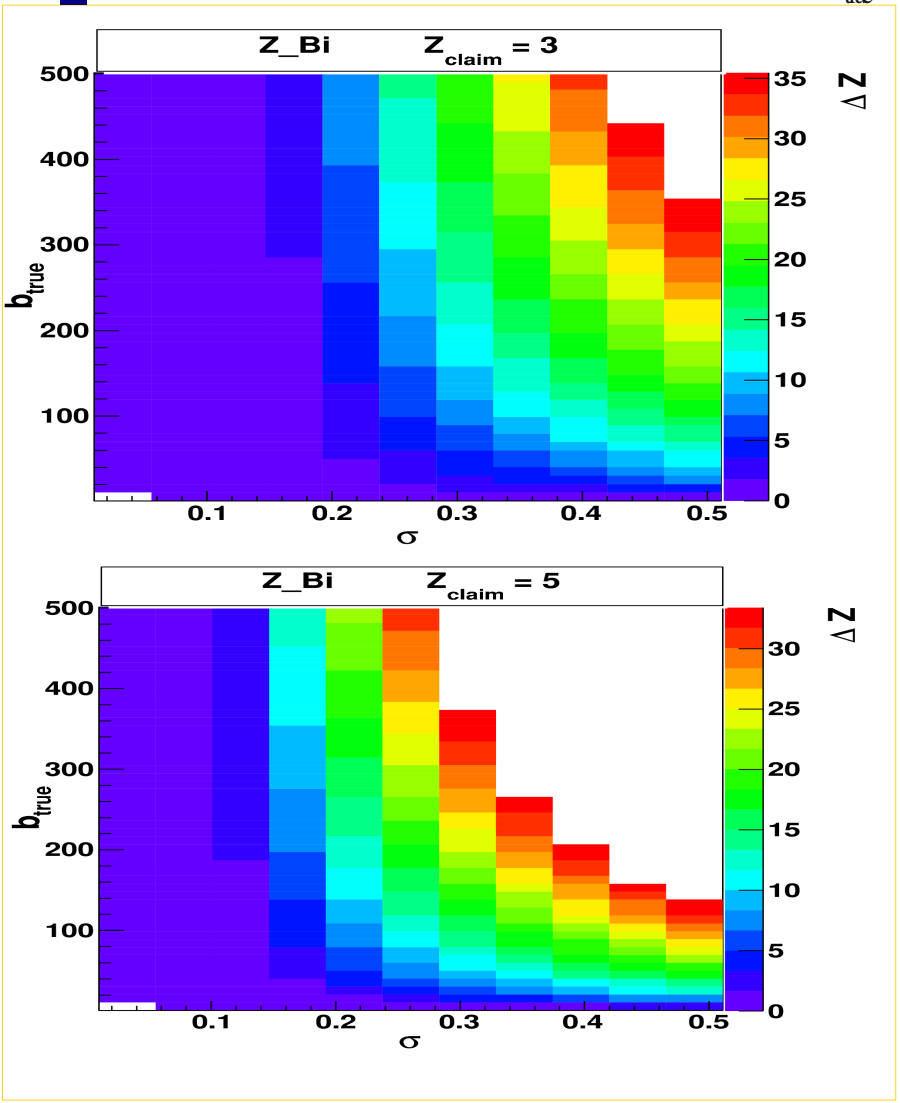


Coverage test of significance

PLC : σ_b =0 -0.5 gaussian



Z_Bi: $\Delta Z = Z_{true} - Z_{claim}$



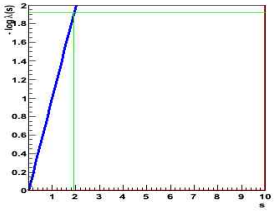
- Huge overcoverage for high systematics (>0.3), especially for Z_Bi getting even worse for high significance (Z>3) and sensitive to the systematics shape. Is that what you call conservative?

Summary

1. RooFit/Stats is an excellent tool for statistical modeling.
   used in real physics analysis (yet for rather simple models)

2. Different methods:
ProfileLikelihood, Feldman-Cousins, Bayes calculator gives rather similar results for confidence intervals without systematics (Poisson).
Similarly the Hybrid, LikelihoodRatio and Hybrid methods are in good agreement in hypothesis testing without systematics.

With the combined multichannel model and some realistic systematics results start to diverge substantially.



Confidence intervals:

- ProfileLhood doesn't work with zero observation and gives least conservative limits often close to undercoverage
- Feldman-Cousins deliver conservative limits al low bakcground with good coverage though sensitive to the systematics boundaries.

In case of zero observation it gives more optimistic upper limits for high background expectation.

- Bayesian intervals(flat prior) gives conservative limits but the coverage is not guarantied. Doesn't depend on bkg for zero observation.

The result are sensitive to the prior and the range of expected signal.

Summary(cont.)

Hypothesis testing:

- Likelihood ratio is working well for single channel and combined models even at low observation and bkg expectations and high systematics(but sym.)
- Hybrid method is consistent with Likelihood but somewhat more conservative
- The Z_Bi gives most conservative limits, huge overcoverage for high systematics

3. When combining different (exclusive) channels one has to consider:

- split search topologies to have least correlations in bkg. Systematics
- split topologies to have complimentary sensitivity in model parameters space
- optimize selection (S/B) to have similar evolution with statistics, i.e balance of stat and systematic uncertainties
- use data driven background estimation when most of uncertainties are in the absolute normalization(rarely the case) otherwise use as a control measurement in g.o.f. If still using, evaluate systematics in tau factor.
- check consistency among all channels before combining them (eg. CL_b for excl.)
- present results with different methods and, ideally, perform coverage test.