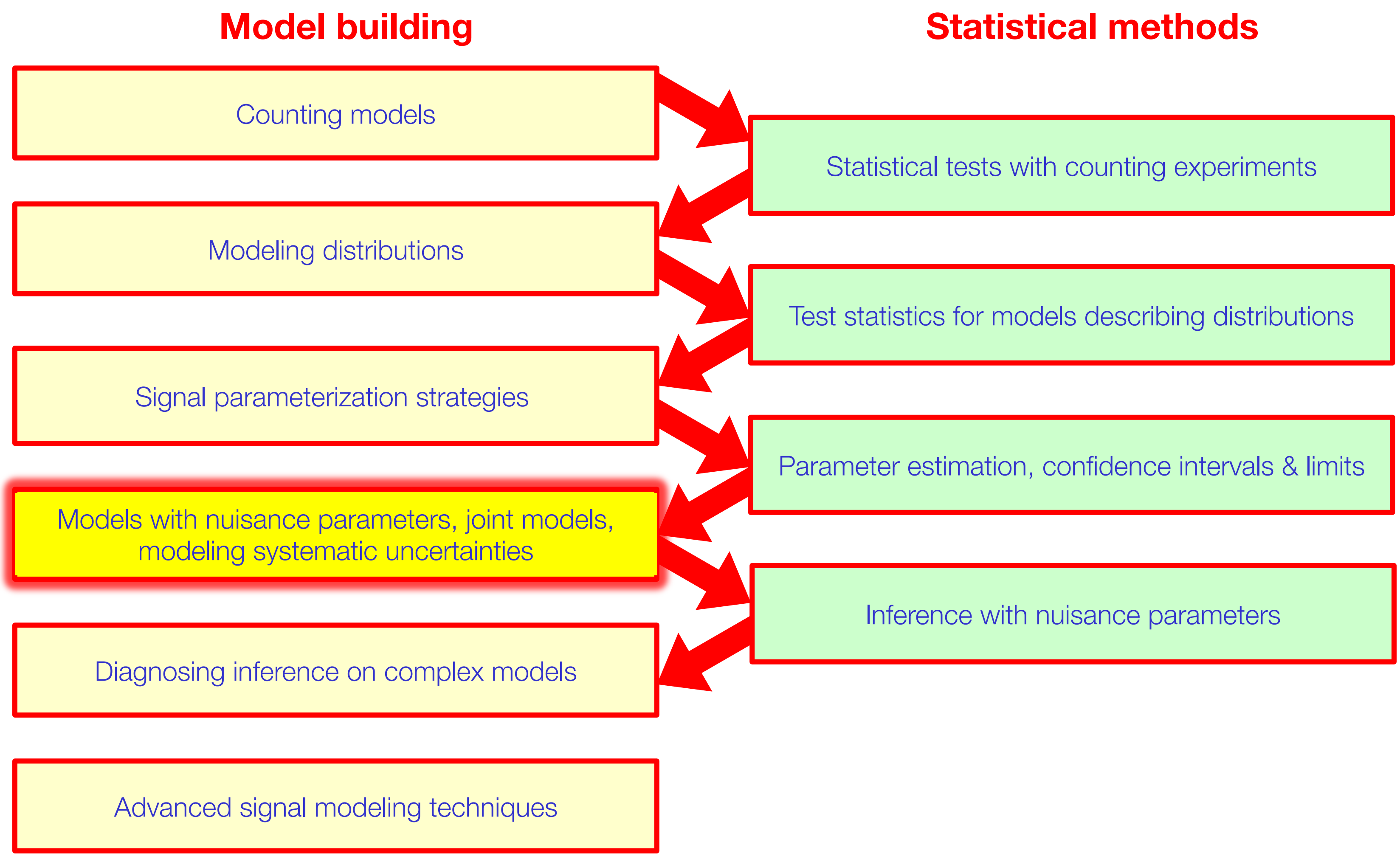
Model building 4

Models with parameters II - simultaneous fits, representing external information as subsidiary measurements ('profile likelihood fits')

Wouter Verkerke, NIKHEF

Roadmap of this course

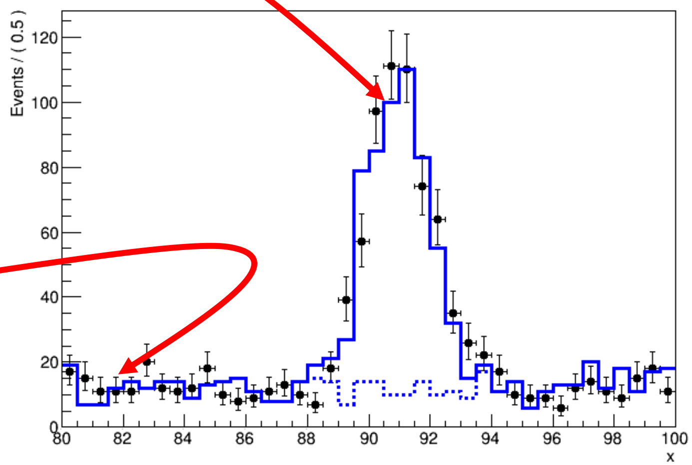
- Start with basics, gradually build up to complexity



So far we've only considered the *ideal* experiment

- The "only thing" you need to do (as an experimental physicist) is to formulate the likelihood function for your measurement
- For an ideal experiment, where signal and background are assumed to have perfectly known properties, this is trivial

$$L(\vec{N} \mid \mu) = \prod_{bins} Poisson(N_i \mid \mu \cdot \tilde{s}_i + \tilde{b}_i)$$

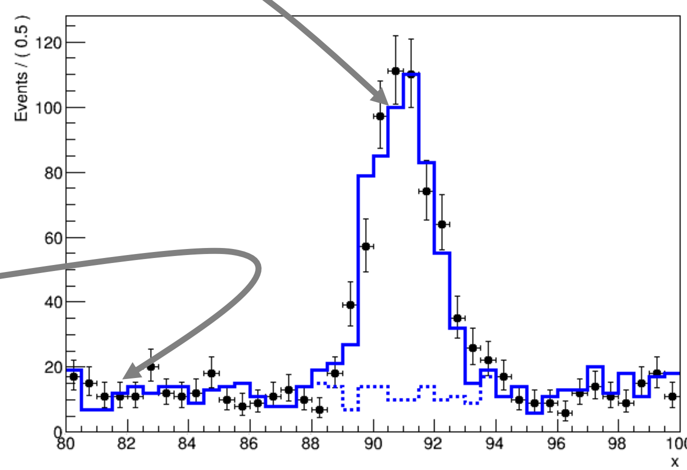


- So far only considered a single parameter in the likelihood: the physics *parameter of interest,* usually denoted as *μ*

The imperfect experiment

- In realistic measurements many effect that we don't control exactly influence measurements of parameter of interest
- How do you model these uncertainties in the likelihood?

$$L(\vec{N}\,|\,\mu) = \prod_{bins} Poisson(N_i\,|\,\mu \cdot \tilde{s}_i + \tilde{b}_i)$$

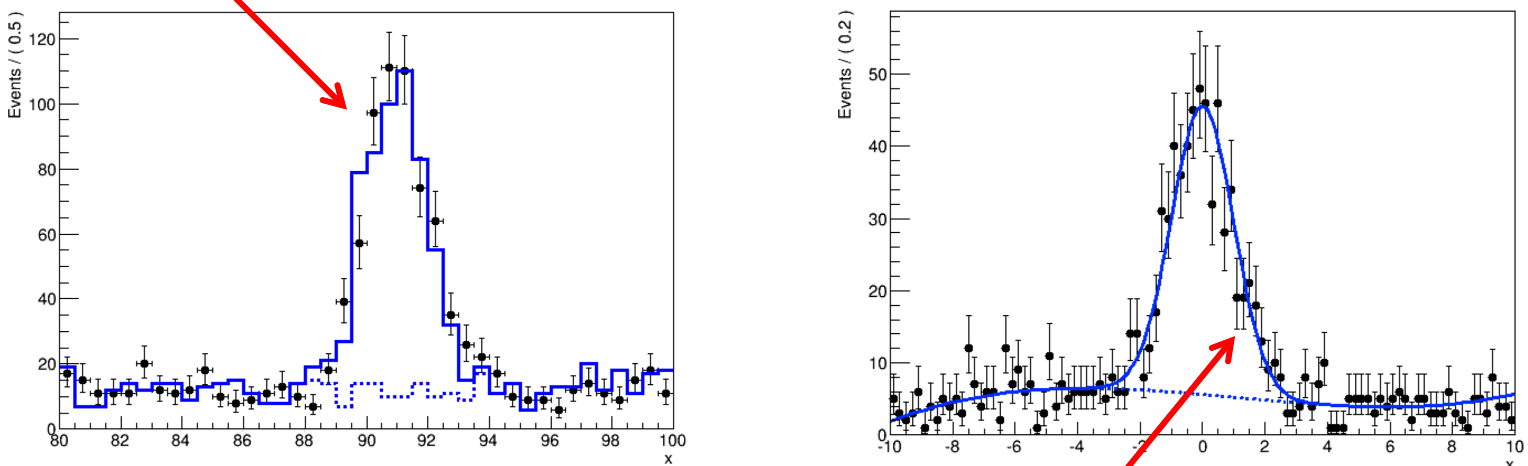


Signal and background predictions are affected by (systematic) uncertainties

Adding parameters to the model

- We can describe uncertainties in our model by adding new parameters of which the value is uncertain

$$L(\vec{N}\,|\,\mu) = \prod_{bins} Poisson(N_i\,|\,\mu \cdot \tilde{s}_i + \tilde{b}_i)$$

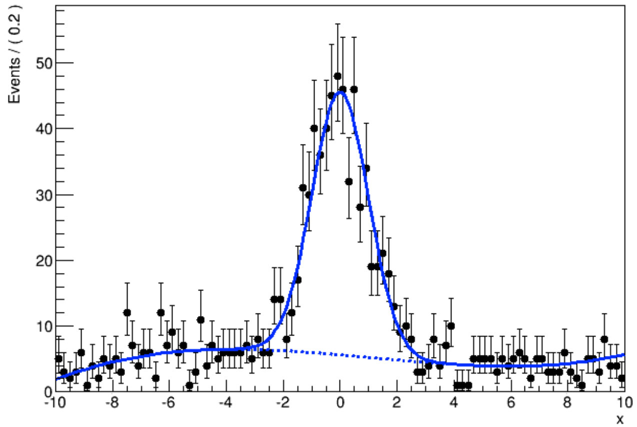


$$L(x\,|\,f,m,\sigma,a_0,a_1,a_2) = fG(x,m,\sigma) + (1-f)Poly(x,a_0,a_1,a_2)$$

- These additional model parameters are not 'of interest', but we need them to model uncertainties → 'Nuisance parameters'

What are the nuisance parameters of your *physics model*?

- *Empirical* modeling of uncertainties, e.g. polynomial for background, Gaussian for signal, is easy to do, but may lead to hard questions



$$L(x \mid f, m, \sigma, a_0, a_1, a_2) = fG(x, m, \sigma) + (1-f)Poly(x, a_0, a_1, a_2)$$

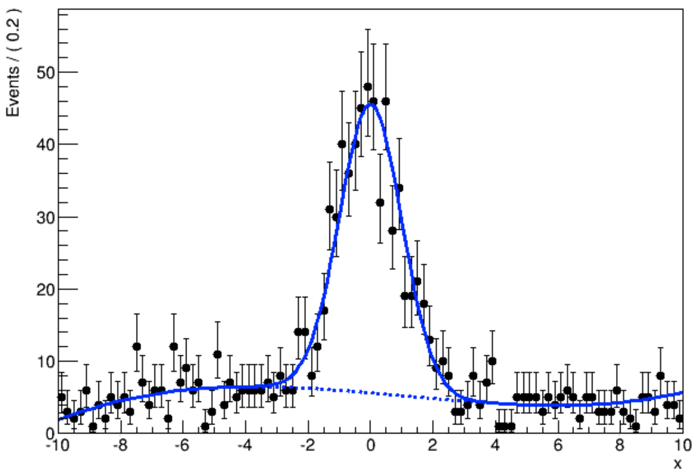
- Is your model correct? (Is true signal distr. captured by a Gaussian?)
- Is your model flexible enough? (4th order polynomial, or better 6th)?
- How do model parameters connect to known detector/theory uncertainties in your distribution?
  - what conceptual uncertainty do your parameters represent?

What information constrains nuisance parameters?

- Some datasets contain sufficient information to constrain nuisance parameters, other do not.

Example 1 – Shape fit

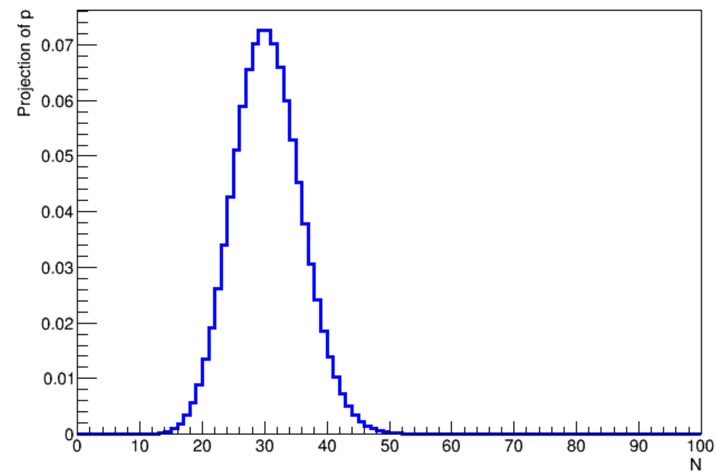
f(x|S,B)=S*Gaussian(x)+B*Uniform(x)



Sufficient information
in data to constrain both S,B

Example 2 – Counting experiment

f(N|S,B)=Poisson(N|S+B)



Insufficient information
in data to constrain both S,B
→ Need additional measurement of B

‘ex17_combined.C’

Simultaneous fits / joint likelihoods

- If >1 measurements exist that constrain (nuisance) parameters, can combine information by formulating a joint likelihood

$$L_A(x|S,B) \qquad\qquad L_B(y|B)$$

$$L(x,y|S,B)_{A+B} = L_A(x|S,B)*L_B(y|B)$$

- No constraints shapes or forms of Likelihood
  - Can combine counting measurement, shape measurement
  - Likelihoods can have same observables, different observables, all OK
  - Only condition is that parameter have same meaning in all measurements

Constraining a nuisance parameter from a control region

- Solution for Poisson counting measurement $P(N|S+B)$ with unconstrained B is to join with measurement in a control region that measures B only

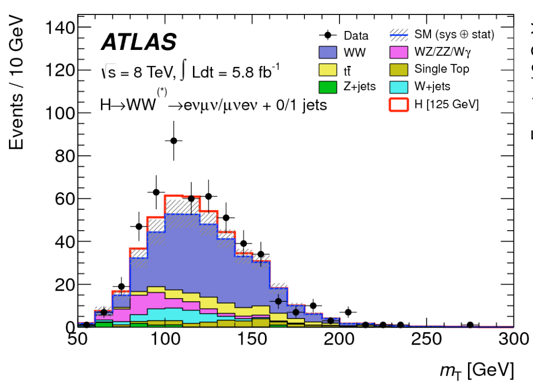
$$L_{SIG}(N_{sig}|S,B)=\mathrm{Poisson}(N_{sig}|S+B) \qquad L_{CTL}=\mathrm{Poisson}(N_{CTL}|\tau^{*}B)$$

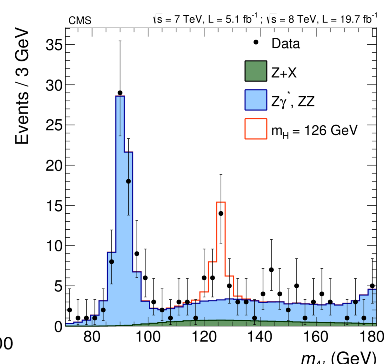
$$L_{joint}(N_{SIG},N_{CTL}|S,B)_{A+B} = \mathrm{Poisson}(N_{sig}|S+B)^{*}\mathrm{Poisson}(N_{CTL}|\tau^{*}B)$$

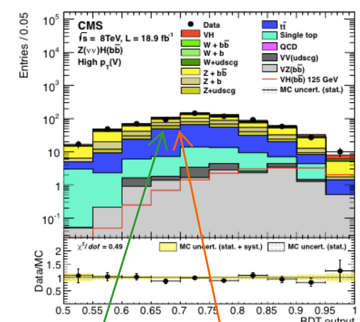
Sufficient information in joint Likelihood to solve for both S and B

Constraining parameters from >>1 region

- Inference from joint likelihood models combines information from all measurements that carry information on a given parameter
  - Can also combine many measurements that constrain the same parameter
- So can also do $L_{SIG1} + L_{SIG2} + \ldots + L_{SIGN}$ instead of $L_{SIG} + L_{CTL}$ or any combination of signal and control regions

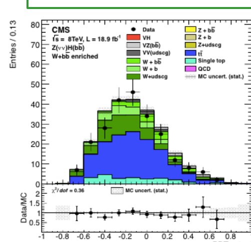


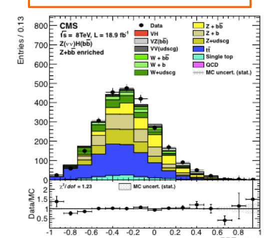




W+b(b) enriched control region

Z+b(b) enriched control region

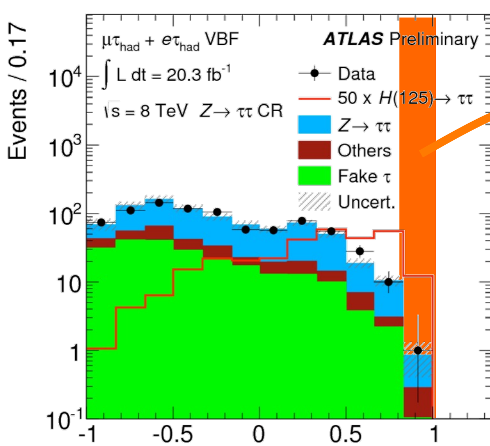


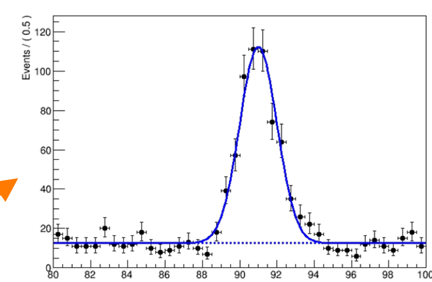


Example:
Higgs channels from ATLAS and CMS, along with the background control regions
All channels measure common Higgs signal strength modifier (=deviation of expectation from SM)

Splitting signal regions by expected purity

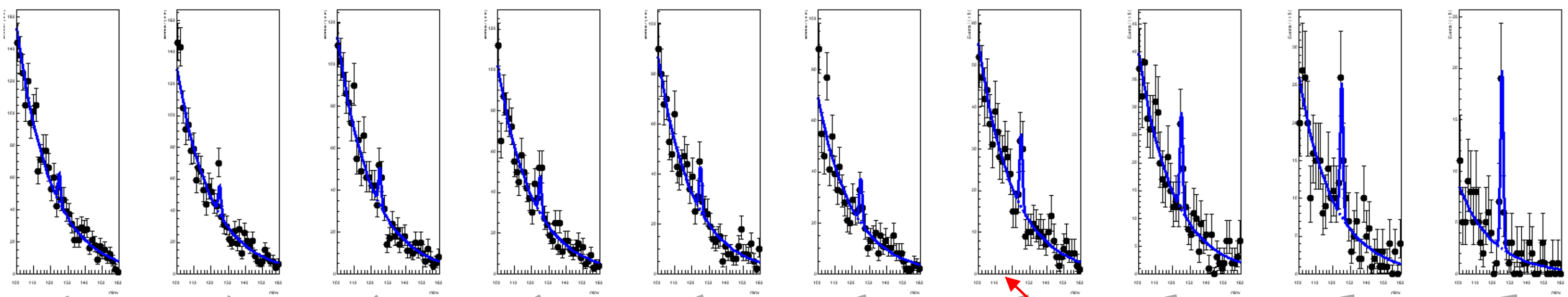
- Another common strategy that results in >>1 signal region, is to split an existing (big) signal region in multiple regions that have different expected purity
- Prototypical problem – MVA classifier sorts observed events by purity
  - If MVA shape is trusted (well understood in simulation) → fit MVA distribution
  - But MVA classification is not well trusted, then what?
- If another discriminating observable exists (e.g. invariant mass)
  - Train MVA without this observable
  - Fit 'invariant mass' in bins of MVA observable
    → Measures signal count independent of MVA prediction
  - **Exploits difference in purity across MVA prediction range without relying on its predicted distribution**





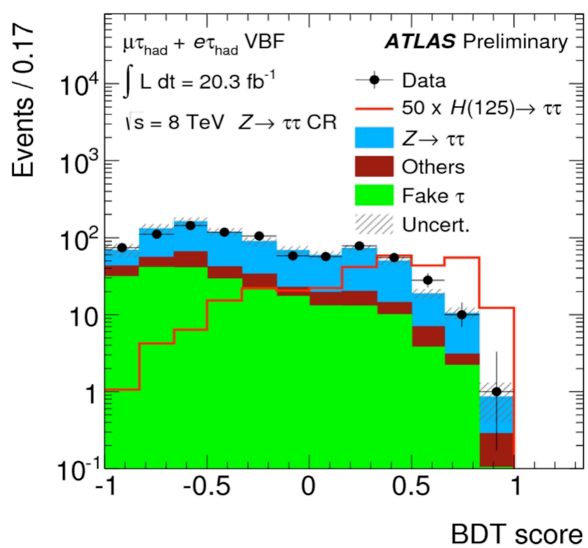
Visualization of signal region splitting

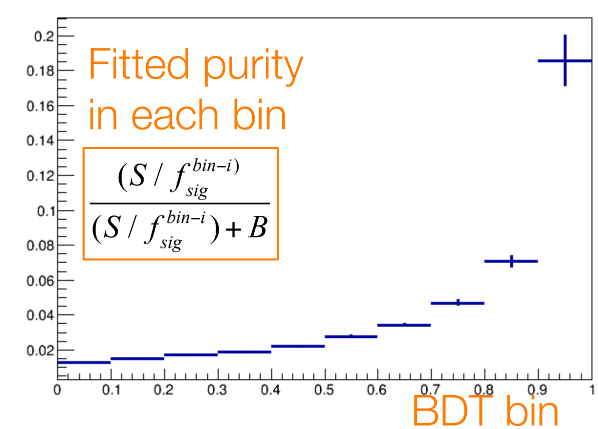
- Split data in regions by BDT score, fit each region with inv. mass



$$f_{bin-i}(m \mid S,B) = \frac{S}{f_{sig}^{bin-i}} f_S(m) + B_{bin-i} f_B(s)$$

Scale factor that ensures
that every bin interprets
S as the total signal yield





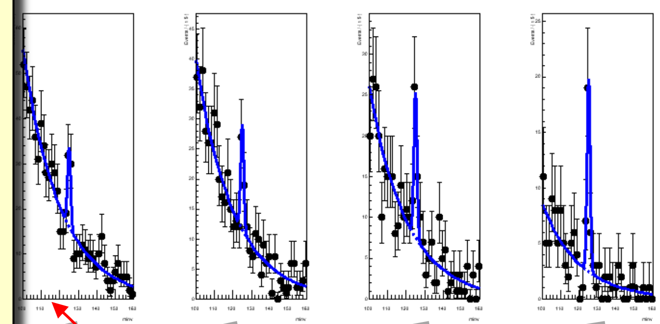
Visualization of signal region splitting

- Split data in regions by BDT score, fit each region with inv. mass

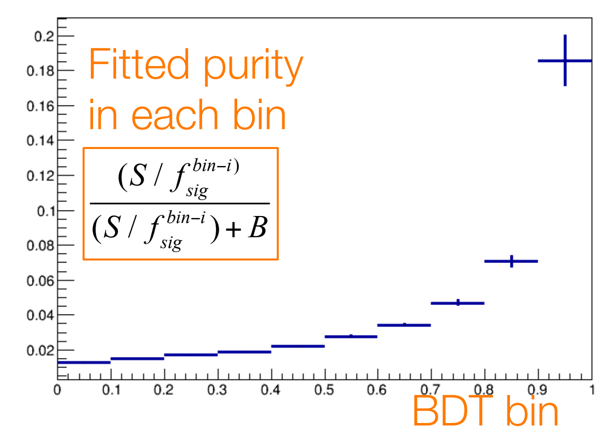
Joint PDF for this model

$$f(m,n_{BDT}\,|\,S,\vec{B}) = lookup(n_{BDT}) \begin{cases} f_{bin-0}(m\,|\,S,B_0) = \dfrac{S}{f_{sig}^{bin-0}} f_S(m) + B_{bin-0} f_B(s) \\ f_{bin-1}(m\,|\,S,B_1) = \dfrac{S}{f_{sig}^{bin-1}} f_S(m) + B_{bin-1} f_B(s) \\ f_{bin-2}(m\,|\,S,B_2) = \dfrac{S}{f_{sig}^{bin-2}} f_S(m) + B_{bin-2} f_B(s) \\ f_{bin-3}(m\,|\,S,B_3) = \dfrac{S}{f_{sig}^{bin-3}} f_S(m) + B_{bin-3} f_B(s) \\ \vdots \\ f_{bin-N}(m\,|\,S,B_N) = \dfrac{S}{f_{sig}^{bin-N}} f_S(m) + B_{bin-N} f_B(s) \end{cases}$$

```
// Construct template model
w.factory("SUM::fit_template(prod(Nsig[30,0,100],frac[1])*sig1,
                                        Nbkg[1000,0,10000]*bkg1)") ;

// Construct joint model from template clones
w.factory("SIMCLONE::fitmodel(fit_template,
                           $SplitParam({Nbkg,frac},bdtBin))") ;
```



$$n\,|\,S,B) = \frac{S}{f_{sig}^{bin-i}} f_S(m) + B_{bin-i} f_B(s)$$



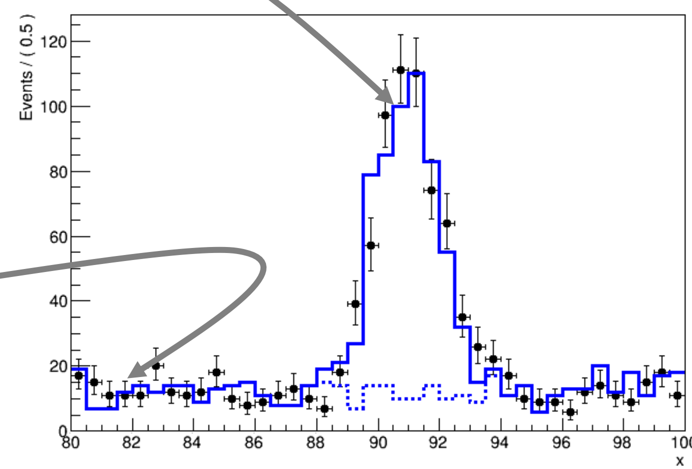
BDT score

`ex13_build_binned.C'

The imperfect experiment

- When relying on simulation templates to build models, a whole world of problems awaits when considering that simulation predictions have many systematic uncertainties associated with them?

$$L(\vec{N}\,|\,\mu) = \prod_{bins} Poisson(N_i \,|\, \mu \cdot \tilde{s}_i + \tilde{b}_i)$$



Signal and background predictions are affected by (systematic) uncertainties

The simulation workflow and origin of uncertainties

Simulation of 'soft physics' physics process

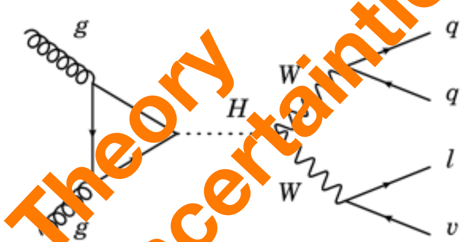
Theory uncertainties

Simulation of ATLAS detector

Detector modeling uncertainties

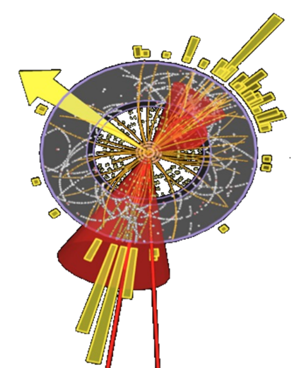
LHC data

Simulation of high-energy physics process

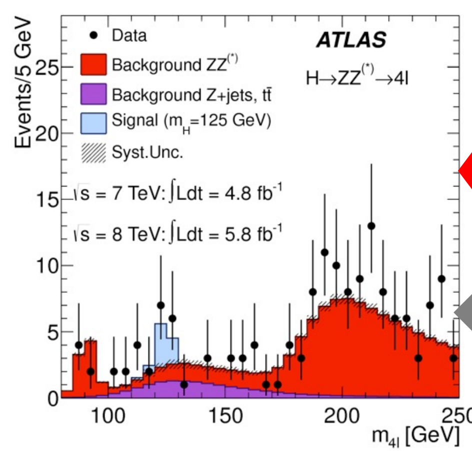


Theory uncertainties

Reconstruction of ATLAS detector



Analysis Event selection



Typical systematic uncertainties in HEP

- Detector-simulation related
  - "The Jet Energy scale uncertainty is 5%"
  - "The b-tagging efficiency uncertainty is 20% for jets with p_T<40"
- Physics/Theory related
  - The top cross-section uncertainty is 8%
  - "Vary the factorization scale by a factor 0.5 and 2.0 and consider the difference the systematic uncertainty"
  - "Evaluate the effect of using Herwig and Pythia and consider the difference the systematic uncertainty"
- MC simulation statistical uncertainty
  - Effect of (bin-by-bin) statistical uncertainties in MC samples

What can you do with *systematic* uncertainties

- As most of the typical systematic prescriptions have no immediately apparent parametric formulation in your likelihood, common approach is 'vary setting, rerun analysis, observe the difference'
- This common 'naïve' approach to assess effect of systematic uncertainties amounts to simple error propagation
- Error propagation procedure in a nutshell
  - Make nominal measurement (using your favorite statistical inference procedure)
  - Change setting in detector simulation or theory (e.g. shift Jet Calibration scale by '1 sigma' up and down) Redo measurement procedure for each shift
  - Consider propagated effect of shifted setting the systematic uncertainty

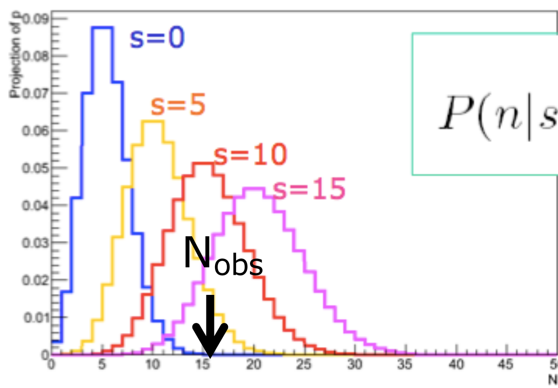
$$\mu = \underbrace{\mu_{nom} \pm \sigma_{stat}}_{\text{From statistical analysis}} \pm \underbrace{(\mu_{syst}^{up} - \mu_{syst}^{down})}_{\text{Systematic uncertainty from error propagation}} / 2 \pm \ldots$$

Pros and cons of the 'naïve' approach

- **Pros**
  - It's easy to do
  - It results in a seemingly easy-to-interpret table of systematics
- **Cons**
  - Uncorrelated source of systematic uncertainty can have correlated effect on measurement → Completely ignored
  - Magnitude of stated systematic uncertainty may be incompatible with measurement result → Completely ignored
  - You lost the connection with fundamental statistical techniques (i.e. evaluation of systematic uncertainties is completely detached from statistical procedure used to estimate physics quantity of interest) → No prescription to make confidence intervals, Bayesian posteriors etc in this way
  - No calibrated probabilistic statements possible (95% C.L.)
- 'Profiling' → Incorporate a description of systematic uncertainties in the likelihood function that is used in statistical procedures

Everything starts with the likelihood

- **All** fundamental statistical procedures are based on the likelihood function as ‘description of the measurement’



$$P(n|s+b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

NB: b is a constant in this example

Definition: the Likelihood is P(observed data|theory)

e.g. L(15|s=0)

e.g. L(15|s=10)

Frequentist statistics → **Confidence interval on s**

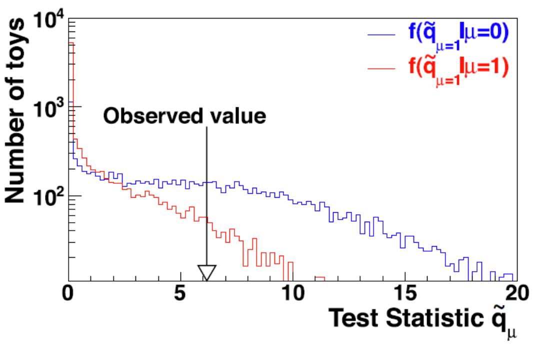
Bayesian statistics → **Posterior on s**

Maximum Likelihood → **s = x ± y**

Everything starts with the likelihood

Frequentist statistics

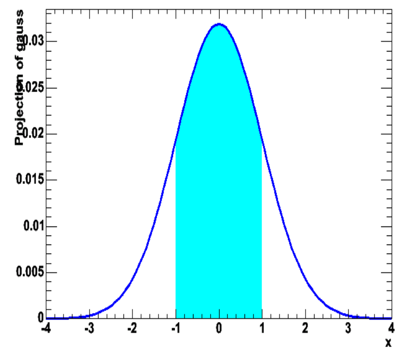
$$\lambda_\mu(\vec{N}_{obs}) = \frac{L(\vec{N}\,|\,\mu)}{L(\vec{N}\,|\,\hat{\mu})}$$



Confidence interval or p-value

Bayesian statistics

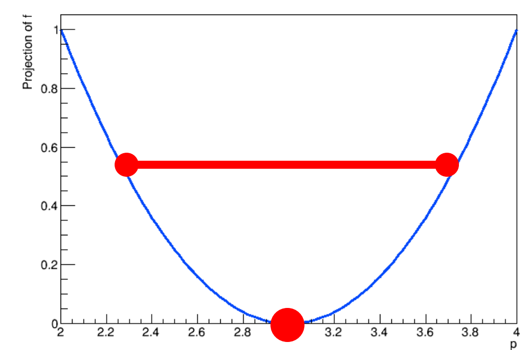
$$P(\mu) \propto L(x\,|\,\mu)\cdot\pi(\mu)$$



Posterior on s or Bayes factor

Maximum Likelihood

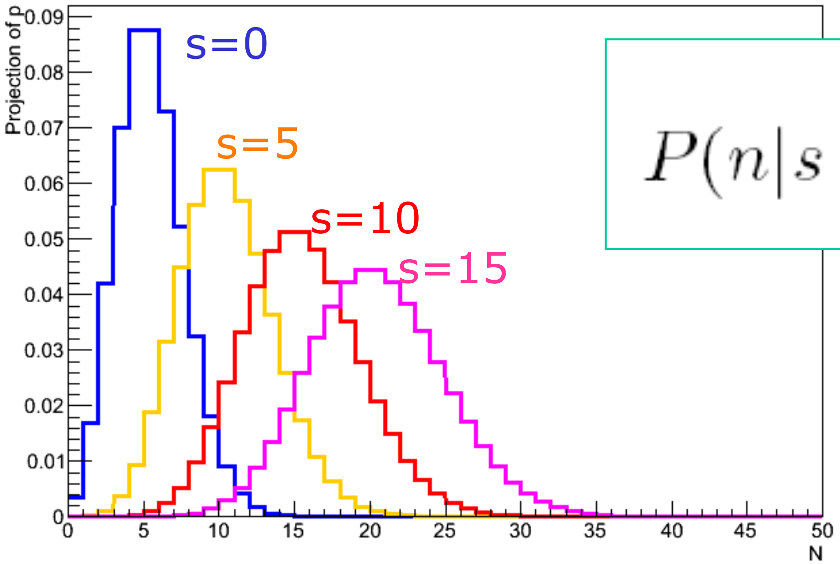
$$\left.\frac{d\ln L(\vec{p})}{d\vec{p}}\right|_{p_i=\hat{p}_i} = 0$$



s = x ± y

Introducing uncertainties – a non-systematic example

- The original model (with fixed *b*)



$$P(n|s+b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

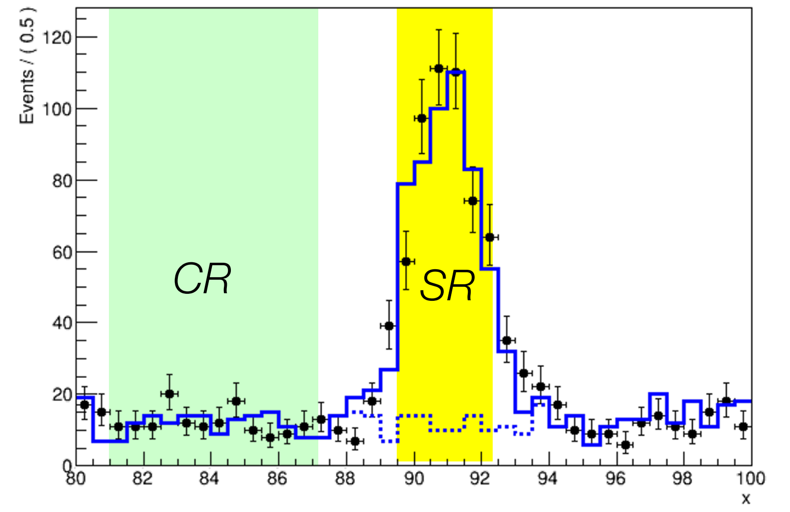
- Now consider *b* to be uncertain

L(N|s) → L(N|s,b)

- The experimental data contains insufficient to constrain both *s* and *b* → Need to add an additional measurement to constrain *b* ‘ex11_build_PoissonPoisson.C’

The sideband measurement

- Suppose your data in reality looks like this ➔



Can estimate level of background in the ‘signal region’ from event count in a ‘control region’ elsewhere in phase space

$$L_{SR}(s,b) = Poisson(N_{SR} \mid s+b)$$

$$L_{CR}(b) = Poisson(N_{CR} \mid \tilde{\tau} \cdot b)$$

NB: Define parameter ‘b’ to represents the amount of bkg is the SR.

Scale factor τ accounts for difference in size between SR and CR

“Background uncertainty constrained from the data”

- Full likelihood of the measurement (‘simultaneous fit’)

$$L_{full}(s,b) = Poisson(N_{SR} \mid s+b) \cdot Poisson(N_{CR} \mid \tilde{\tau} \cdot b)$$

'ex12_build_PoissonGaussGlobs.C'

Generalizing the concept of the sideband measurement

- Background uncertainty from sideband clearly clearly not a 'systematic uncertainty'

$$L_{full}(s,b) = Poisson(N_{SR} | s+b) \cdot Poisson(N_{CR} | \tilde{\tau} \cdot b)$$

- Now consider scenario where *b* is not measured from a sideband, but is taken from MC simulation **with an 8% cross-section 'systematic' uncertainty**

'Measured background rate by MC simulation'

$$L_{full}(s,b) = Poisson(N_{SR} | s+b) \cdot \underbrace{Gauss(\tilde{b} | b, 0.08)}_{\text{'Subsidiary measurement' of background rate}}$$

  - *We can model this in the same way, because the cross-section uncertainty is also (ultimately) the result of a measurement*

Generalize: 'sideband' → 'subsidiary measurement'

What is a systematic uncertainty?

- Concept & definitions of ‘systematic uncertainties’ originates from physics, not from fundamental statistical methodology.
  - E.g. Glen Cowans (excellent) 198pp book “statistical data analysis” does not discuss systematic uncertainties at all

- A common definition is
  - “Systematic uncertainties are all uncertainties that are not directly due to the statistics of the data”

- But the notion of ‘the data’ is a key source of ambiguity:
  - does it include control measurements?
  - does it include measurements that were used to perform basic (energy scale) calibrations?

Typical systematic uncertainties in HEP

- Detector-simulation related
  - "The Jet Energy scale uncertainty is 5%"
  - "The b-tagging efficiency uncertainty is 20% for jets with p_T<40"

Subsidiary measurement is an actual measurement
→ conceptually similar to a 'sideband' fit

- Physics/Theory related
  - The top cross-section uncertainty is 8%
  - "Vary the factorization scale by a factor 0.5 and 2.0 and consider the difference the systematic uncertainty"
  - "Evaluate the effect of using Herwig and Pythia and consider the difference the systematic uncertainty"

Subsidiary measurement unclear, but origin of prescription may well be another measurement (if yes, like sideband, if no, what is source of info?)

- MC simulation statistical uncertainty
  - Effect of (bin-by-bin) statistical uncertainties in MC samples

Subsidiary measurement is a Poisson counting experiment (but now in MC events), otherwise conceptually identical to a 'sideband fit'

Typical systematic uncertainties in HEP

- Detector-simulation related
  - “The Jet Energy scale uncertainty is 5%”
  - “The b-tagging efficiency uncertainty is 20%

Subsidiary measurement is an actual measurement → conceptually to

- P

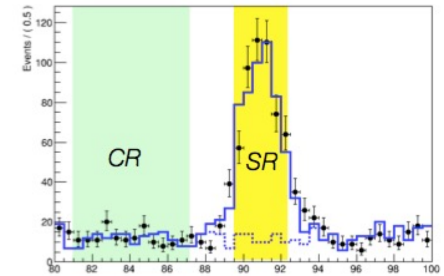
Almost all systematic uncertainties are similar in nature to ‘sidebands’ measurements of some form or shape

→ Can always model systematics like sidebands in the Likelihood

And even when the are not the (in)direct result of some measurement (certainty theory uncertainties) we can still model them in that form

- MC simulation statistical uncertainty
  - Effect of (bin-by-bin) statistical uncertainties in MC samples

Subsidiary measurement is a Poisson counting experiment (but now in MC events), otherwise conceptually identical to a ‘sideband fit’



Modeling a detector calibration uncertainty

$$L_{full}(s,b) = Poisson(N_{SR} \mid s+b) \cdot Gauss(\tilde{b} \mid b, 0.08)$$

- **Now consider a detector uncertainty**, e.g. jet energy scale calibration, which can affect the analysis acceptance in a non-trivial way (unlike the cross-section example)

Nominal calibration

Signal rate (our parameter of interest)

Assumed calibration

$$L(N,\tilde{\alpha} \mid s,\alpha) = Poisson(N \mid s + \tilde{b}(\alpha/\tilde{\alpha})\cdot 2)) \cdot Gauss(\tilde{\alpha} \mid \alpha, \sigma_\alpha)$$

Observed event count

Uncertainty on nominal calibration **(here 5%)**

Nominal background expectation from MC (a constant), obtained with a=a˜

Response function for JES uncertainty (a 1% JES change results in a 2% acceptance change)

“Subsidiary measurement” Encodes ‘external knowledge’ on JES calibration

Modeling a detector calibration uncertainty

- Simplify expression by renormalizing "subsidiary measurement"

$Gauss(\tilde{\alpha} \mid \alpha, \sigma_\alpha)$

Signal rate (our parameter of interest)

$$L(N \mid s,\alpha) = Poisson(N \mid s + \tilde{b}(1+0.1\alpha)) \cdot Gauss(0 \mid \alpha, 1)$$

Observed event count

Nominal background expectation from MC (a constant)

Response function for normalized JES parameter

[a unit change in α – a 5% JES change – still results in a 10% acceptance change]

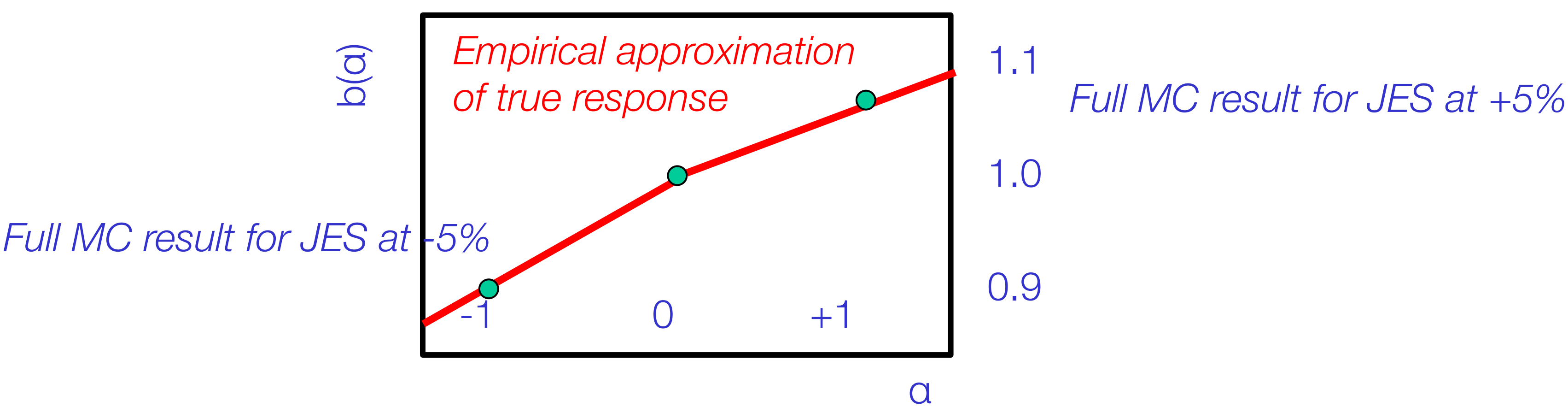
"Normalized subsidiary measurement"

The scale of parameter α is now chosen such that values ±1 corresponds to the nominal uncertainty (in this example 5%)

The response function as empirical model of full simulation

$$L(N,0\,|\,s,\alpha) = Poisson(N\,|\,s+\underbrace{b(\alpha)})\cdot Gauss(0\,|\,\alpha,1)$$

- Note that the response function is generally not linear, but can in principle *always be determined by your full simulation chain*
  - But you cannot run your full simulation chain for any arbitrary ‘systematic uncertainty variation’ → Too much time consuming
  - Typically, run full MC chain for nominal and ±1σ variation of systematic uncertainty, and approximate response for other values of NP with interpolation
  - For example run at nominal JES and with JES shifted up and down by ±5%



What is a systematic uncertainty?

- It is an uncertainty in the Likelihood of your physics measurement that is characterized deterministically, up to a set of parameters, of which the true value is unknown.
- A fully specified systematic uncertainty defines
  - 1: A set of one or more parameters
    of which the true value is unknown,
  - 2: A response model that describes the effect of those parameters on the measurement
    (sampled from full simulation, and interpolation)
  - 3: A subsidiary measurement of the parameters
    that constrains the values the parameters can take
    (implies a specific distribution: Gaussian *(default, CLT)*, Poisson *(low-stats counting)*, or otherwise)

'ex12_build_PoissonGaussGlobs.C'

Names and conventions – 'profiling' & 'constraints'

- The full likelihood function of the form

$$L(N,0 \mid s,\alpha) = Poisson(N \mid s + b(\alpha)) \cdot Gauss(0 \mid \alpha,1)$$

  is usually referred to by physicists as a 'profile likelihood', and systematics are said to be 'profiled' when incorporated this way
  - Note: statisticians use the word profiling for something else

- Physicists often refer to the subsidiary measurement as a 'constraint term'
  - This is correct in the sense that it constrains the parameter α, but this labeling commonly lead to mistaken statements (e.g. that it is a pdf for α)
  - But it is ***not*** a pdf in the NP

$$\cancel{Gauss(\alpha \mid 0,1)} \qquad Gauss(0 \mid \alpha,1)$$

Names and conventions

- The 'subsidiary measurement' as simplified form of the 'full calibration measurement' also illustrates another important point
  - The full likelihood is simply a *joint likelihood of a physics measurement and a calibration measurement* where both terms are treated on equal footing in the statistical procedure
  - In a perfect world, not bound by technical modelling constraints you would use this likelihood

$$L(N,\vec{y}\,|\,s,\alpha) = Poisson(N\,|\,s+b(1+0.1\alpha))\cdot L_{JES}(\vec{y}\,|\,\alpha,\vec{\theta})$$

    where L_{JES} is the full calibration measurement as performed by the Jet calibration group, based on a dataset y, and which may have other parameters θ specific to the calibration measurement.

- Since we are bound by technical constrains, we substitute L_{JES} with simplified (Gaussian) form, but the statistical treatment and interpretation remains the same

Gamma and logNormal distributions

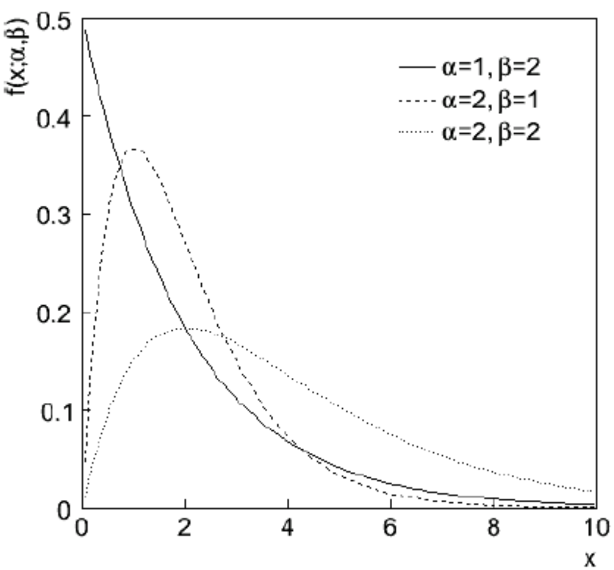
Gamma distribution

=distribution of μ resulting from a Poisson measurement L(N|μ)

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

$$E[x] = \alpha\beta$$

$$V[x] = \alpha\beta^2$$

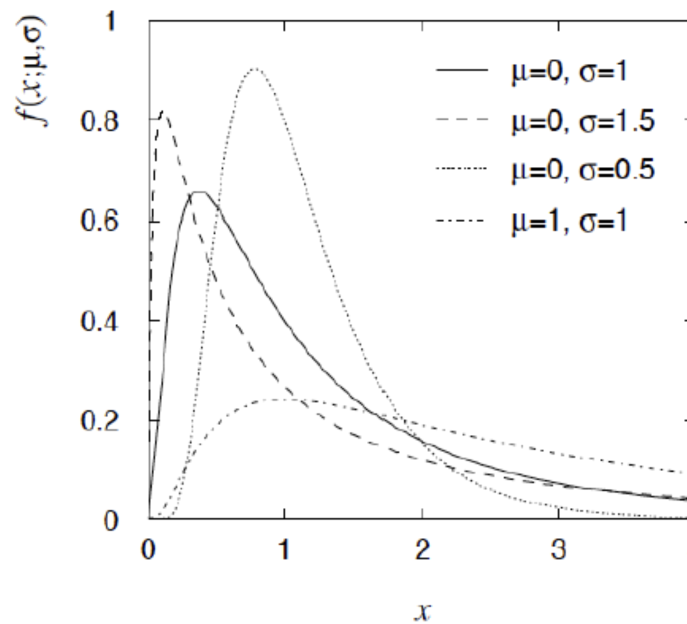


logNormal distribution

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(\frac{-(\log x-\mu)^2}{2\sigma^2}\right)$$

$$E[x] = \exp(\mu + \tfrac{1}{2}\sigma^2)$$

$$V[x] =$$
$$\exp(2\mu+\sigma^2)[\exp(\sigma^2)-1]$$



MC statistical uncertainties as systematic uncertainty

- Another example of modeling a systematic uncertainty:
  MC statistical uncertainty
- Follow same procedure again as before:
  - Define response function (this is trivial for MC statistics: it is the luminosity ratio of the MC sample and the data sample)
  - Define distribution for the 'subsidiary measurement' – This is a Poisson distribution – since MC simulation is also a Poisson process
  - Construct full likelihood ('profile likelihood')

$$L(N, N_{MC} \mid s, b) = Poisson(N \mid s+b) \cdot Poisson(N_{MC} \mid \tau \cdot b)$$

Constant factor τ = *L*(MC)/*L*(data)

- Note uncanny similarity to full likelihood of a sideband measurement!

$$L(N, N_{ctl} \mid s, b) = Poisson(N \mid s+b) \cdot Poisson(N_{ctl} \mid \tau \cdot b)$$

Modeling multiple systematic uncertainties

- Introduction of multiple systematic uncertainties presents no special issues
- Example JES uncertainty plus generator ISR uncertainty

$$L(N,0 \mid s,\alpha_{JES},\alpha_{ISR}) = P(N \mid s + \underbrace{b(1+0.1\alpha_{JES}+0.05\alpha_{ISR})}_{\text{Joint response function for both systematics}}) \cdot \underbrace{G(0 \mid \alpha_{JES},1)}_{} \cdot \underbrace{G(0 \mid \alpha_{ISR},1)}_{}$$

One subsidiary measurement for each source of uncertainty

- A brief note on correlations
  - Word "correlations" often used sloppily – proper way is to think of correlations of parameter estimators. Likelihood defines parameters α_{JES}, α_{ISR}. The (ML) estimates of these are denoted $\hat{\alpha}_{JES},\hat{\alpha}_{ISR}$
  - The ML estimators of $\hat{\alpha}_{JES},\hat{\alpha}_{ISR}$ using the Likelihood of the subsidiary measurements are uncorrelated (since the product factorize in this example)
  - The ML estimators of $\hat{\alpha}_{JES},\hat{\alpha}_{ISR}$ using the full Likelihood may be correlated. This is due to physics modeling effects encoded in the joint response function

Modeling systematic uncertainties in multiple channels

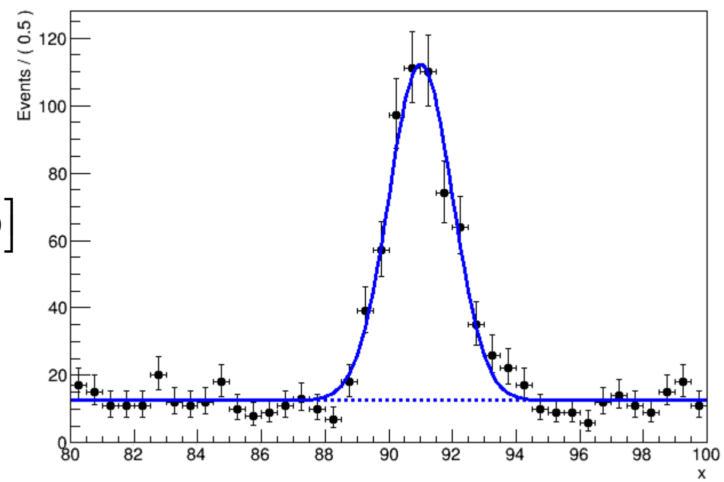
- Systematic effects that affect multiple measurements should be modeled coherently.
  - Example – Likelihood of two Poisson counting measurements

$$L(N_A, N_B \mid s, \alpha_{JES}) = P(N_A \mid s \cdot f_A + b_A \underbrace{(1+0.1\alpha_{JES})}_{\text{JES response function for channel A}}) \cdot P(N_B \mid s \cdot f_B + b_B \underbrace{(1-0.3\alpha_{JES})}_{\text{JES response function for channel B}}) \cdot \underbrace{G(0 \mid \alpha_{JES}, 1)}_{\text{JES subsidiary measurement}} \cdot$$

  - Effect of changing JES parameter α_{JES} coherently affects both measurement.
  - Magnitude and sign effect does not need to be same, this is dictated by the physics of the measurement

Introducing response functions for shape uncertainties

- Modeling of systematic uncertainties in Likelihoods describing distributions follows the same procedure as for counting models
  - Example: Likelihood modeling distribution in a di-lepton invariant mass. POI is the signal strength μ

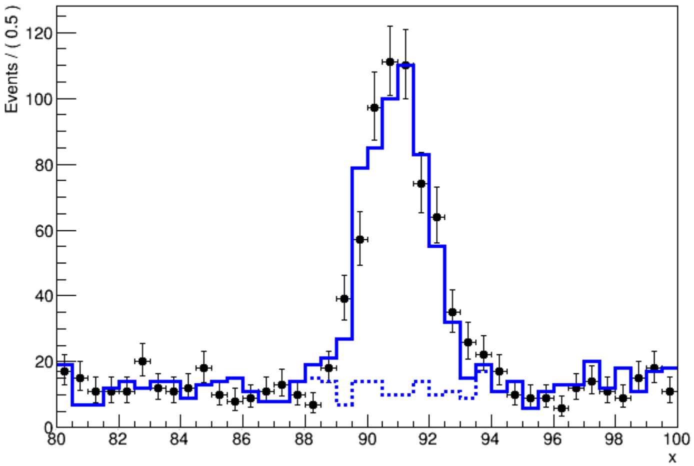


$$L(\vec{m}_{ll} \mid \mu) = \prod_i \left[\mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91, 1) + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right]$$

- Consider a lepton energy scale systematic uncertainty that affects this measurement
  - The LES has been measured with a 1% precision
  - The effect of LES on m_{ll} has been determined to a 2% shift for 1% LES change

$$L(\vec{m}_{ll} \mid \mu, \alpha_{LES}) = \prod_i \left[\mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91 \cdot \underbrace{(1 + 2\alpha_{LES}, 1)}_{\text{Response function}} + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right] \cdot \underbrace{Gauss(0 \mid \alpha_{LES}, 1)}_{\text{Subsidiary measurement}}$$

Response modeling for distributions



- For a change in the **rate**, response modeling of histogram-shaped distribution is straightforward: **simply scale entire distribution**

$$L(\vec{N}\,|\,\mu) = \prod_i Poisson(N_i\,|\,\mu\tilde{s}_i + \tilde{b}_i)$$

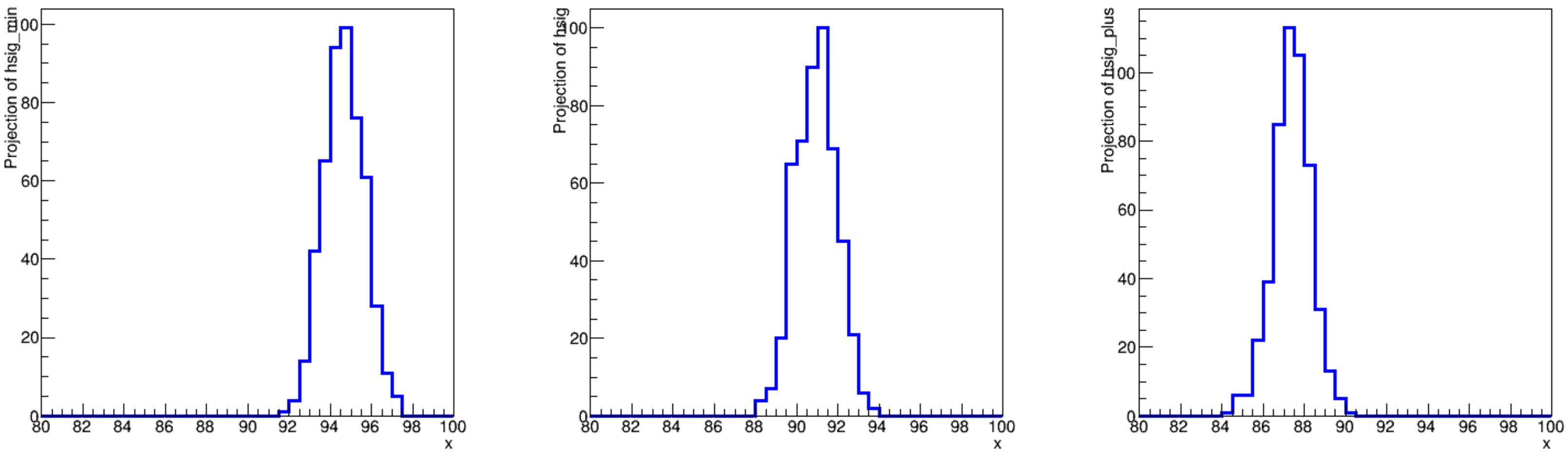
$$L(\vec{N}\,|\,\mu,\alpha) = \prod_i Poisson(N_i\,|\,\mu\tilde{s}_i \cdot \underbrace{(1+3.75\alpha)}_{\text{Response function for signal rate}} + \tilde{b}_i) \cdot \underbrace{Gauss(0\,|\,\alpha,1)}_{\text{Subsidiary measurement}}$$

- But what about a systematic uncertainty that shifts the mean, or affects the distribution in another way?

Modeling of shape systematics in the likelihood

- Effect of *any* systematic uncertainty that affects the shape of a distribution can in principle be obtained from MC simulation chain
  – Obtain histogram templates for distributions at '+1σ' and '-1σ' settings of systematic effect

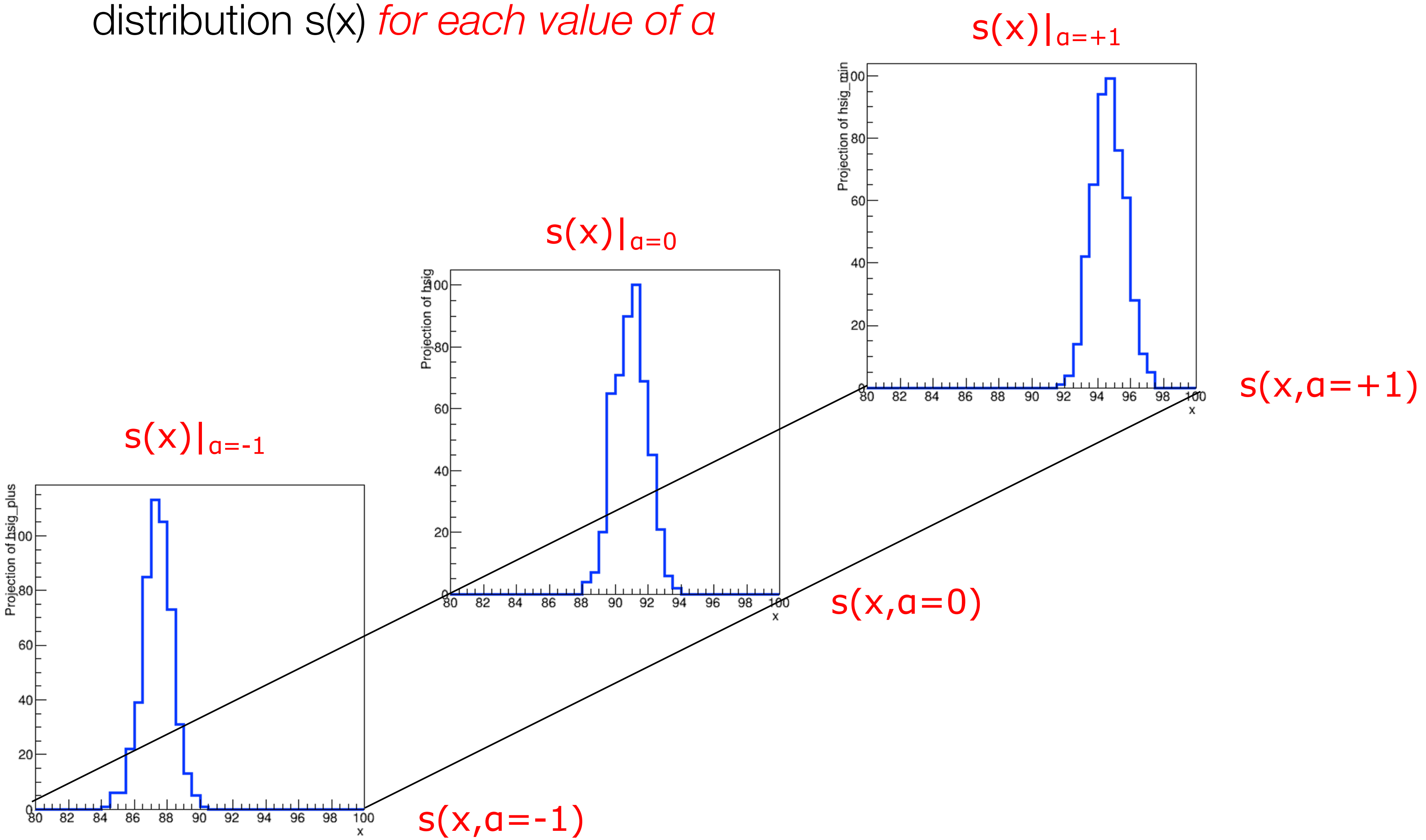
'-1σ' 'nominal' '+1σ'



- Challenge: **construct an empirical response function based on the interpolation of the shapes of these three templates**.

Need to interpolate between template models

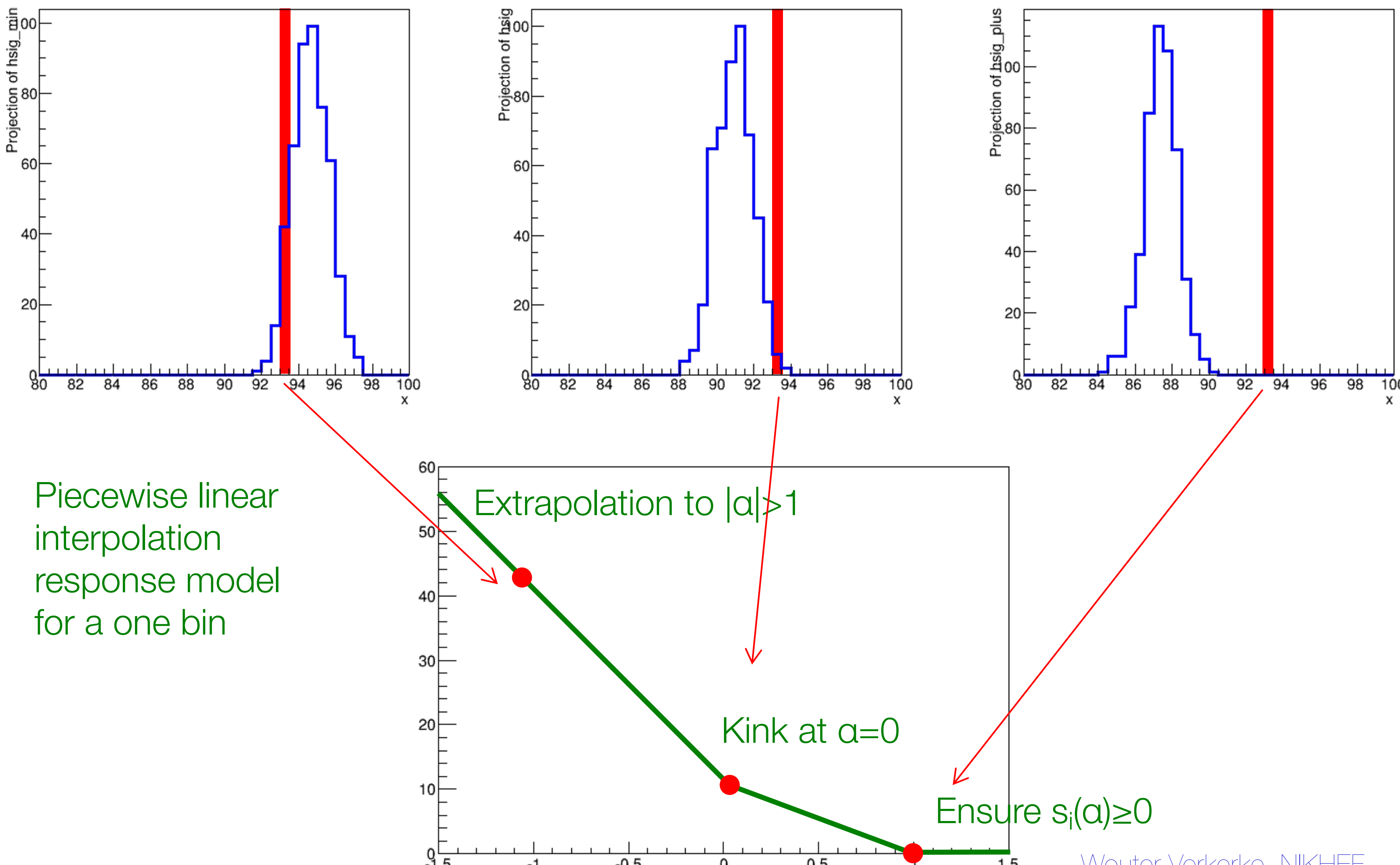
- Need to define 'morphing' algorithm to define distribution s(x) *for each value of α*



Wouter Verkerke, NIKHEF

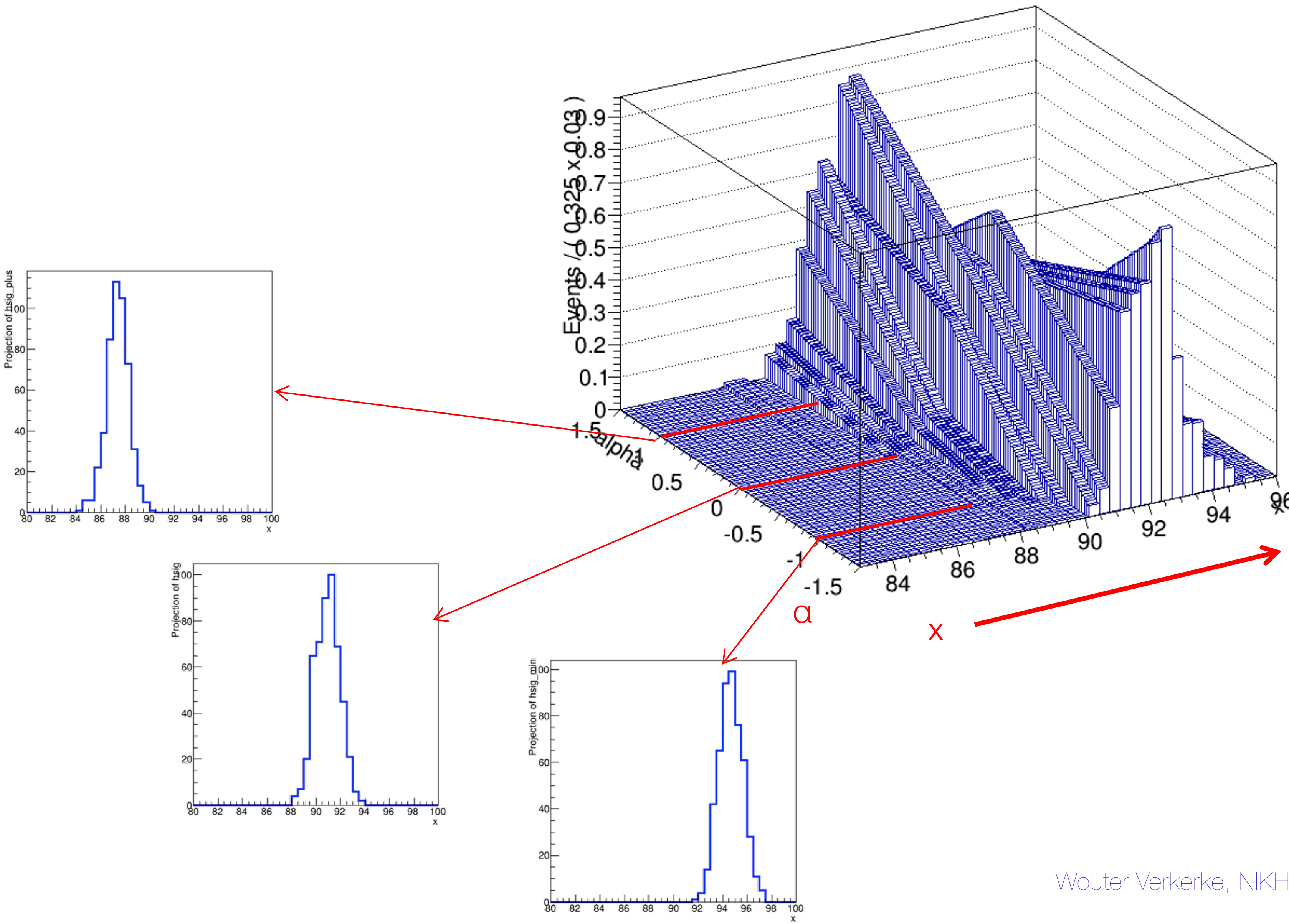
Piecewise linear interpolation

- Simplest solution is piece-wise linear interpolation for each bin

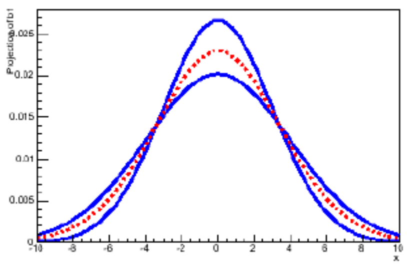
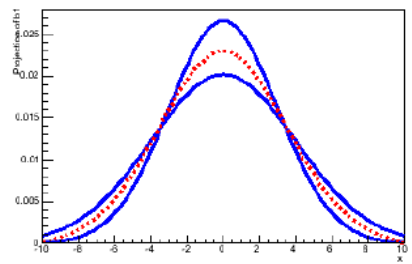
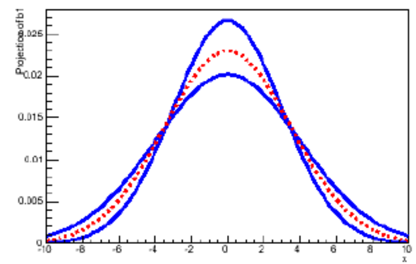
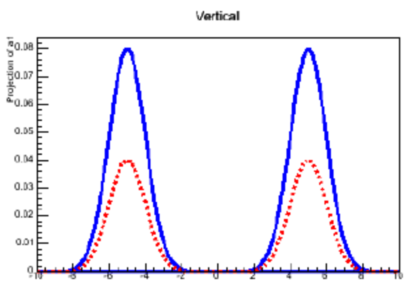
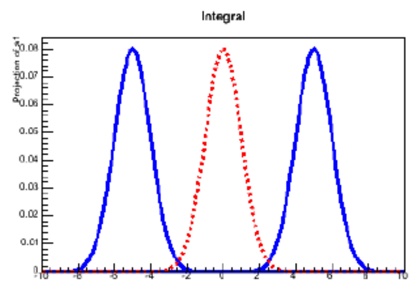
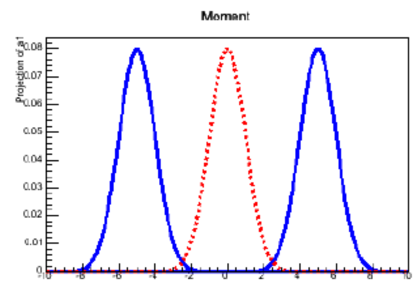
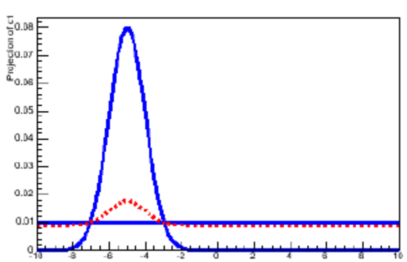
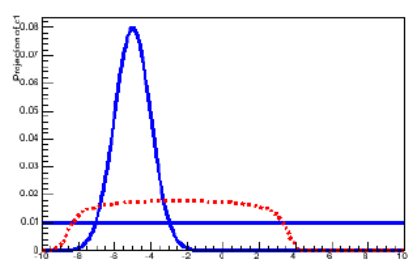
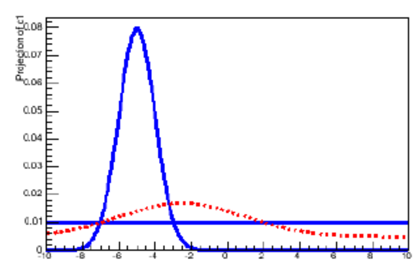


'ex15_build_binned_morphing.C'

Visualization of bin-by-bin linear interpolation of distribution

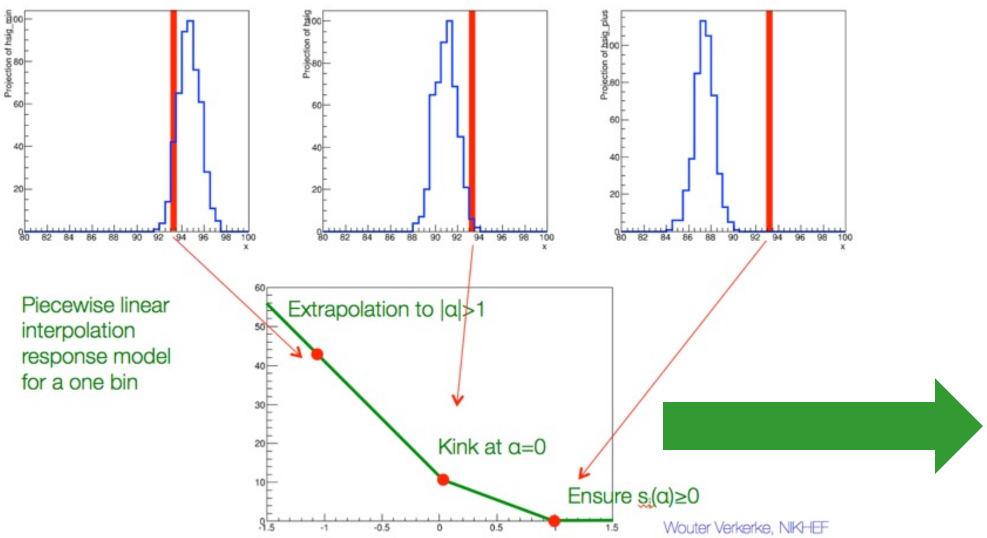


There are other morphing algorithms to choose from

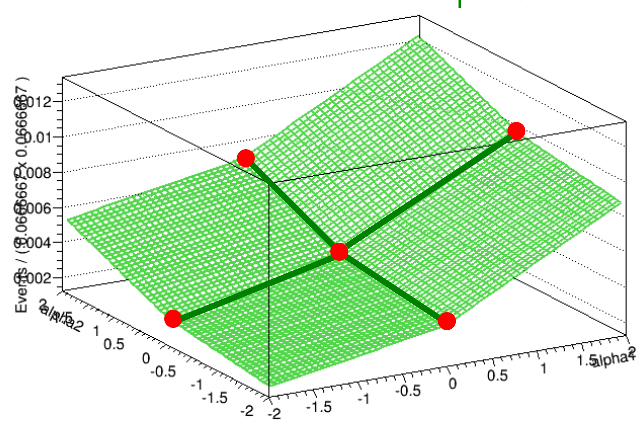
	Vertical Morphing	Horizontal Morphing	Moment Morphing
Gaussian varying width			
Gaussian varying mean			
Gaussian to Uniform (this is conceptually ambigous!)			
n-dimensional morphing?	✔	X	✔

Piece-wise interpolation for >1 nuisance parameter

- Concept of piece-wise linear interpolation can be trivially extended to apply to morphing of >1 nuisance parameter.
  - Difficult to visualize effect on full distribution, but easy to understand concept at the individual bin level

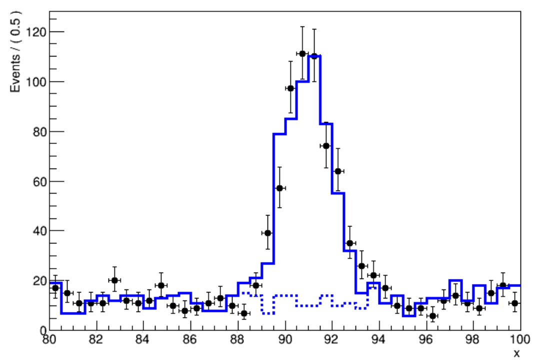


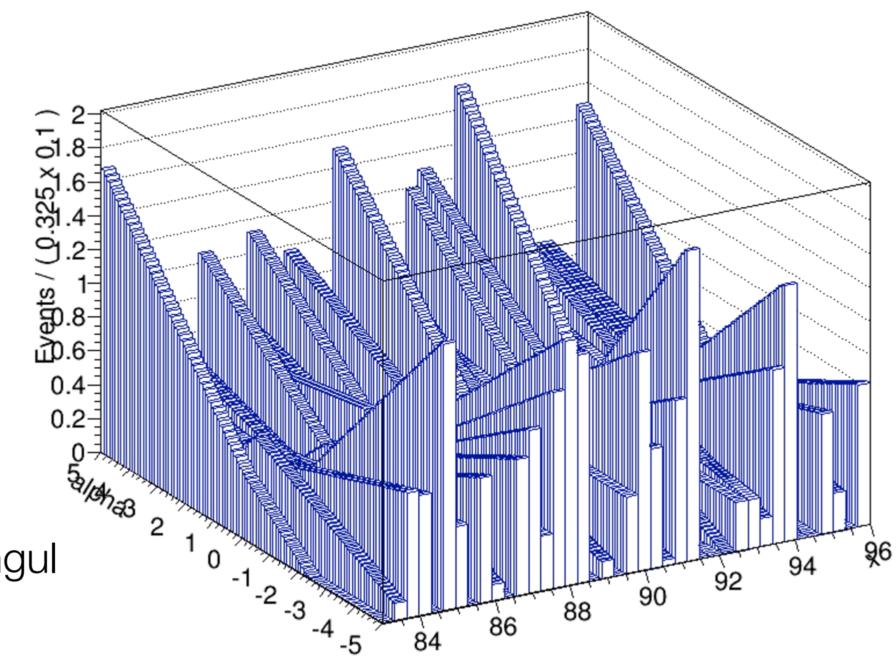
Visualization of 2D interpolation



Shape, rate or no systematic?

- Be judicious with modeling of systematic with little or no significant change in shape (w.r.t MC template statistics)
  - Example morphing of a very subtle change in the background model
  - Is this a meaningful new degree of freedom in the likelihood model?

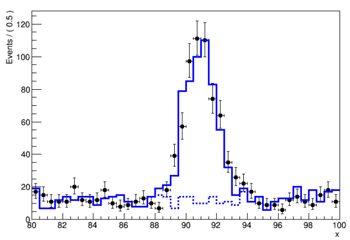




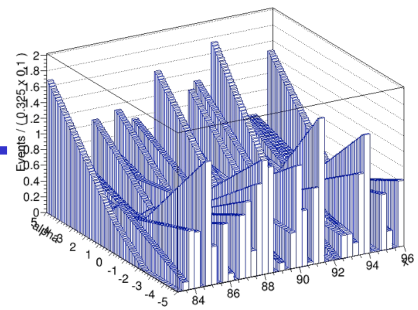
  - A χ2 or KS test between nominal and alternate template can help to decide if a shape uncertainty is meaningul
  - Most systematic uncertainties affect both rate and shape, but can make independent decision on modeling rate (which less likely to affect fit stability)

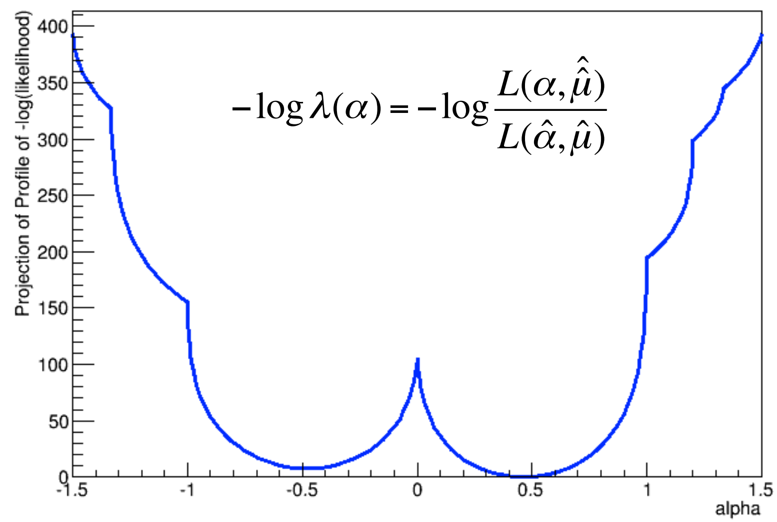
Fit stability due to insignificant shape systematics

- Shape of profile likelihood in NP α clearly raises two points



+

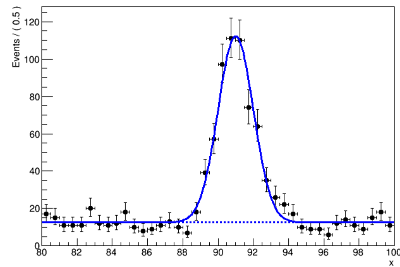




- 1) Numerical minimization process will be ‘interesting’
- 2) MC statistical effects induce strongly defined minima that are fake
  - Because for this example all three templates were sampled from the same parent distribution (a uniform distribution)

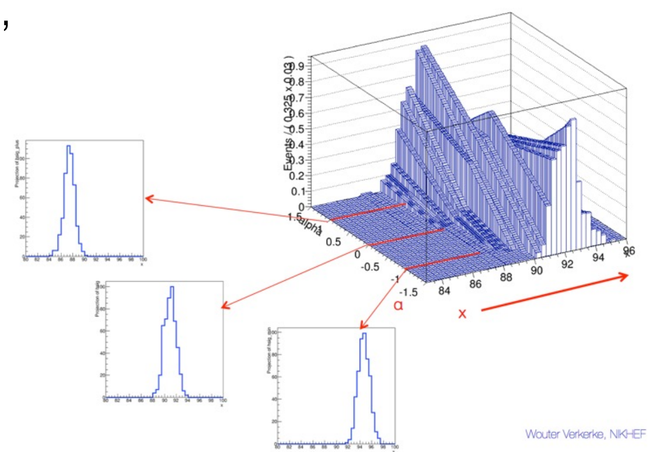
Recap on shape systematics & template morphing

- Implementation of shape systematic in likelihoods modeling distributions conceptually no different that rate systematics in counting experiments



$$L(\vec{m}_{ll} \mid \mu, \alpha_{LES}) = \prod_i \left[\mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91 \cdot (1 + 2\alpha_{LES}, 1) + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right] \cdot Gauss(0 \mid \alpha_{LES}, 1)$$

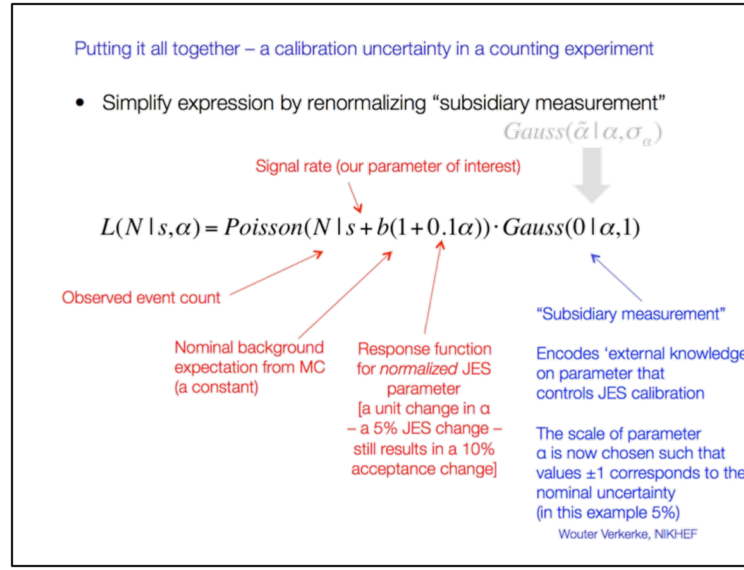
- For template modes obtained from MC simulation template provides a technical solution to implement response function
  - Simplest strategy piecewise linear interpolation, but only works well for small changes
  - Moment morphing better adapted to modeling of shifting distributions
  - Both algorithms extend to n-dimensional interpolation to model multiple systematic NPs in response function
  - Be judicious in modeling 'weak' systematics: MC systematic uncertainties will dominate likelihood



Wouter Verkerke, NIKHEF

Example 1: counting expt

- Will now demonstrate how to construct a model for a counting experiment with a systematic uncertainty



$$L(N | s,\alpha) = Poisson(N | s + b(1+0.1\alpha)) \cdot Gauss(0 | \alpha, 1)$$

```
// Subsidiary measurement of alpha
w.faxtory(“Gaussian::subs(0,alpha[-5,5],1)”) ;

// Response function mu(alpha)
w.factory(“expr::mu(‘s+b(1+0.1*alpha)’,s[20],b[20],alpha)”) ;

// Main measurement
w.factory(“Poisson::p(N[0,10000],mu)”);

// Complete model Physics*Subsidiary
w.factory(“PROD::model(p,subs)”) ;
```

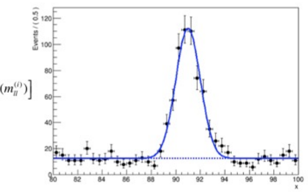
Example 2: unbinned L with syst.

- Will now demonstrate how to code complete example of the unbinned profile likelihood of Section 5:

Introducing shape systematic uncertainties

- Modeling of systematic uncertainties in Likelihood describing distributions follows the same procedure as for counting models
  - Example: Likelihood modeling distribution in a di-lepton invariant mass. POI is the signal strength μ

$$L(\vec{m}_{ll} \mid \mu) = \prod_i \left[\mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91, 1) + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right]$$



- Consider a lepton energy scale systematic uncertainty that affects this measurement
  - The LES has been measured with a 1% precision
  - The effect of LES on m_{ll} has been determined to a 2% shift for 1% LES change

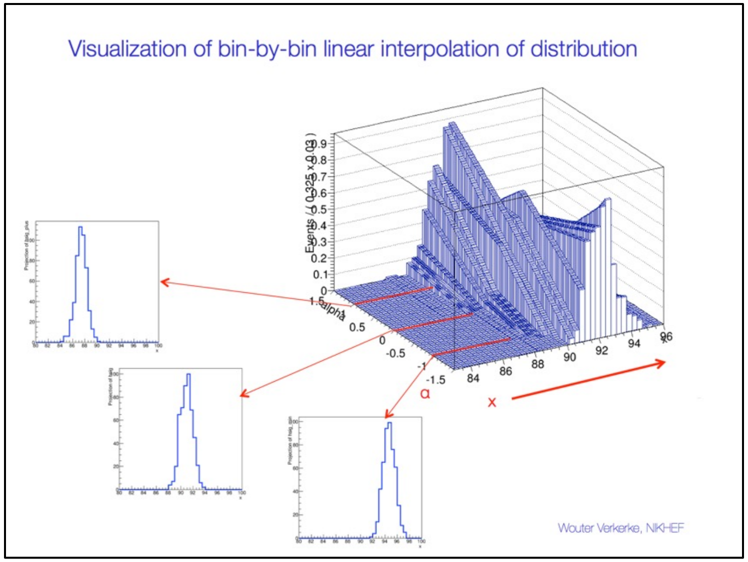
$$L(\vec{m}_{ll} \mid \mu, \alpha_{LES}) = \prod_i \left[\mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91 \cdot \underbrace{(1+2\alpha_{LES}, 1)}_{\text{Response function}} + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right] \cdot \underbrace{Gauss(0 \mid \alpha_{LES}, 1)}_{\text{Subsidiary measurement}}$$

Wouter Verkerke, NIKHEF

$$L(\vec{m}_{ll} \mid \mu, \alpha_{LES}) = \prod_i \left[\mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91 \underbrace{\cdot (1+2\alpha_{LES}), 1)}_{} + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right] \cdot \underbrace{Gauss(0 \mid \alpha_{LES}, 1)}_{}$$

```
// Subsidiary measurement of alpha
w.factory("Gaussian::subs(0,alpha[-5,5],1)");

// Response function m(alpha)
w.factory("expr::m_a("m*(1+2alpha)",m[91,80,100],alpha)") ;

// Signal model
w.factory("Gaussian::sig(x[80,100],m_a,s[1])")

// Complete model Physics(signal plus background)*Subsidiary
w.factory("PROD::model(SUM(mu[0,1]*sig,Uniform::bkg(x)),subs)") ;
```

Example 3 : binned L with syst



- Example of template morphing systematic in a binned likelihood

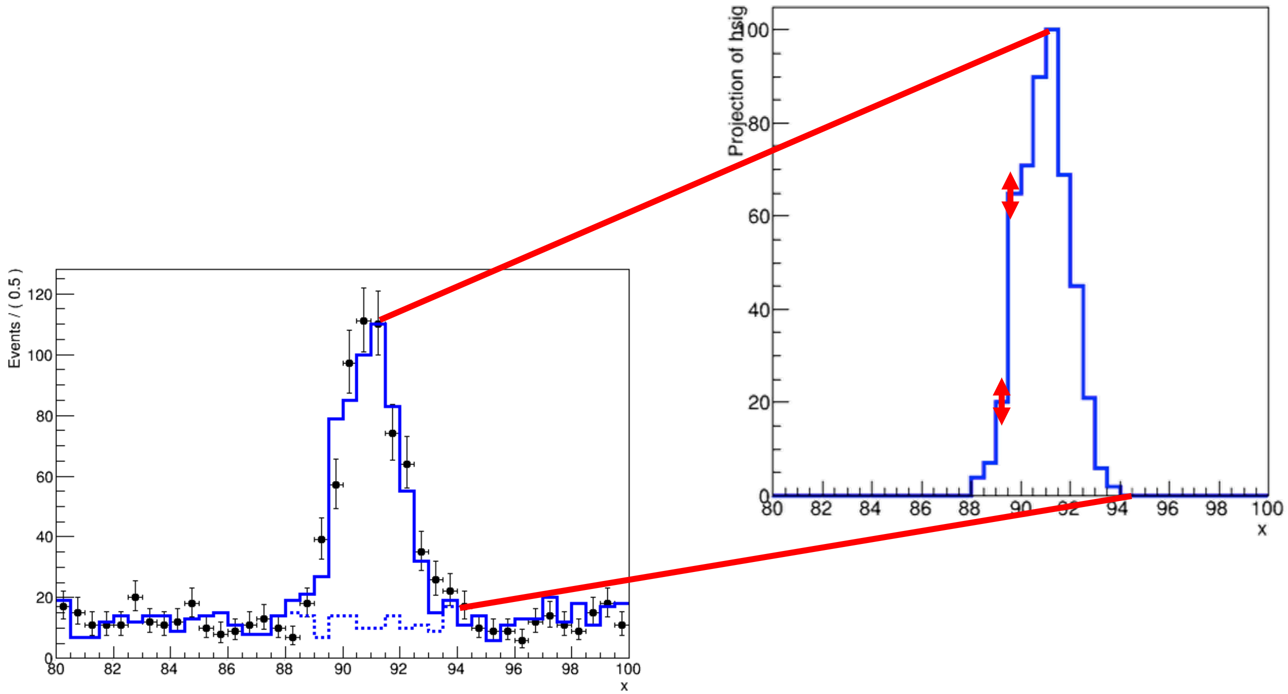
$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha\cdot(s_i^+ - s_i^0) & \forall\alpha > 0 \\ s_i^0 + \alpha\cdot(s_i^0 - s_i^-) & \forall\alpha < 0 \end{cases}$$

$$L(\vec{N}\,|\,\alpha,\vec{s}^{\,-},\vec{s}^{\,0},\vec{s}^{\,+}) = \prod_{bins} P(N_i\,|\,\underbrace{s_i(\alpha,s_i^-,s_i^0,s_i^+)})\cdot\underbrace{G(0\,|\,\alpha,1)}$$

```
// Import template histograms in workspace
 w.import(hs_0,hs_p,hs_m) ;

 // Construct template models from histograms
 w.factory(“HistFunc::s_0(x[80,100],hs_0)”) ;
 w.factory(“HistFunc::s_p(x,hs_p)”) ;
 w.factory(“HistFunc::s_m(x,hs_m)”) ;

 // Construct morphing model
 w.factory(“PiecewiseInterpolation::sig(s_0,s_,m,s_p,alpha[-5,5])”) ;

 // Construct full model
 w.factory(“PROD::model(ASUM(sig,bkg,f[0,1]),Gaussian(0,alpha,1))”) ;
```

Other uncertainties in MC shapes – finite MC statistics

- In practice, MC distributions used for template fits have finite statistics.

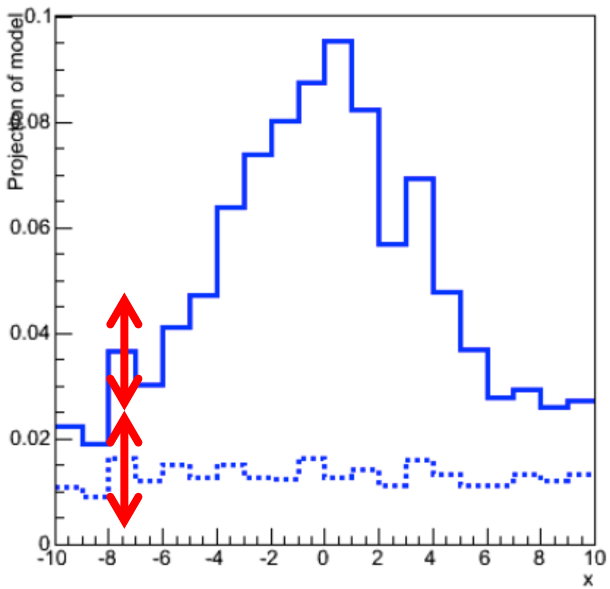


- Limited MC statistics represent an uncertainty on your model → how to model this effect in the Likelihood?

Other uncertainties in MC shapes – finite MC statistics

- Modeling MC uncertainties: *each MC bin has a Poisson uncertainty*
- Thus, apply usual 'systematics modeling' prescription.
- For a single bin – exactly like original counting measurement

Fixed signal, bkg MC prediction



$$L_{bin-i}(\mu) = Poisson(N_i \mid \mu \cdot \tilde{s}_i + \tilde{b}_i)$$

Signal, bkg
MC nuisance params

$$L_{bin-i}(\mu, s_i, b_i) = Poisson(N_i \mid \mu \cdot s_i + b_i)$$

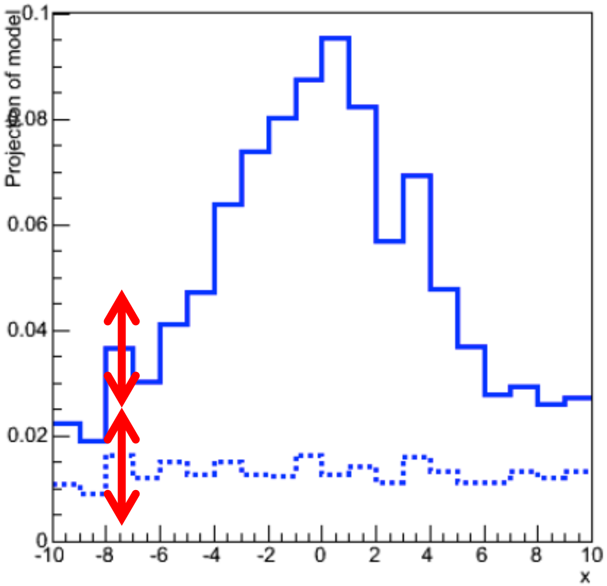
$$\cdot Poisson(N_i^{MC-s} \mid s_i)$$

$$\cdot Poisson(N_i^{MC-b} \mid b_i)$$

Subsidiary measurement for signal MC
('measures' MC prediction s_i with Poisson uncertainty)

Nuisance parameters for template statistics

- Repeat for all bins



$$L(\vec{N}\,|\,\mu) = \prod_{bins} P(N_i\,|\,\mu \cdot \tilde{s}_i + \tilde{b}_i)$$

Binned likelihood with rigid template

$$L(\vec{N}\,|\,\mu,\vec{s},\vec{b}) = \prod_{bins} P(N_i\,|\,\underbrace{\mu \cdot s_i + b_i}_{\text{Response function w.r.t. } s,\, b \text{ as parameters}})\underbrace{\prod_{bins} P(\tilde{s}_i\,|\,s_i)\prod_{bins} P(\tilde{b}_i\,|\,b_i)}_{2\text{x } N_{bins} \text{ subsidiary measurements of } s, b \text{ from } s\sim, b\sim}$$

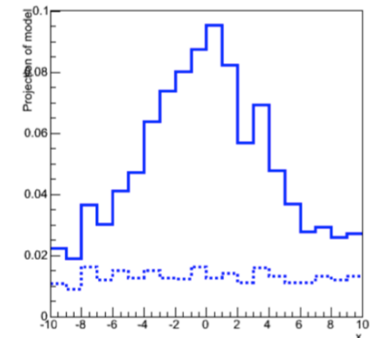
Response function w.r.t. *s*, *b* as parameters

2x N_{bins} subsidiary measurements of *s* ,*b* from s~,b~

- Result: accurate model for MC statistical uncertainty, but lots of nuisance parameters (#samples x #bins)...

The effect of template statistics

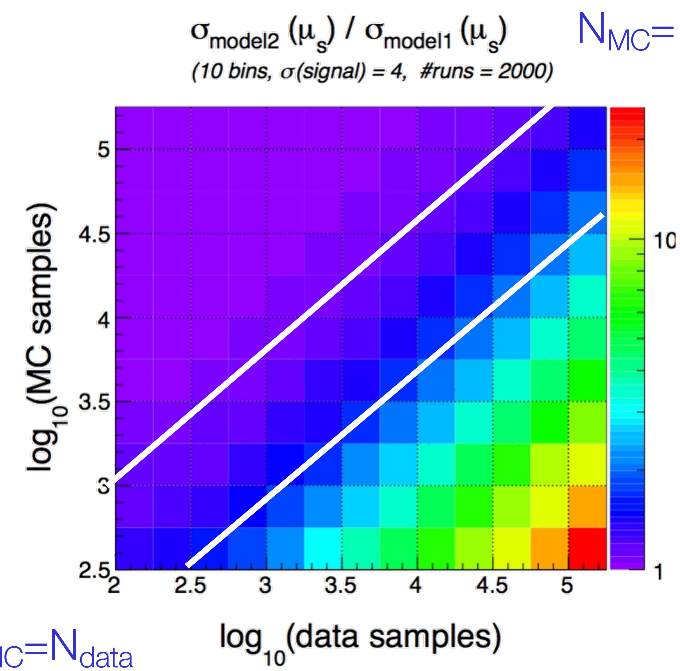
- When is it important to model the effect of template statistics in the likelihood
  - Roughly speaking the effect of template statistics becomes important when $N_{templ} < 10\text{x}\ N_{data}$ (from Beeston & Barlow)
- Measurement of effect of template statistics in previously shown toy likelihood model, where POI is the signal yield



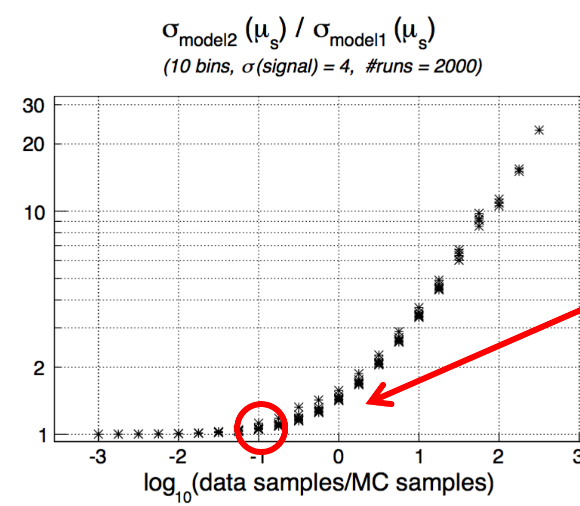
$N_{MC}=10N_{data}$

'model 1 – plain template likelihood'

'model 2 – Beeston-Barlow likelihood'



$N_{MC}=N_{data}$



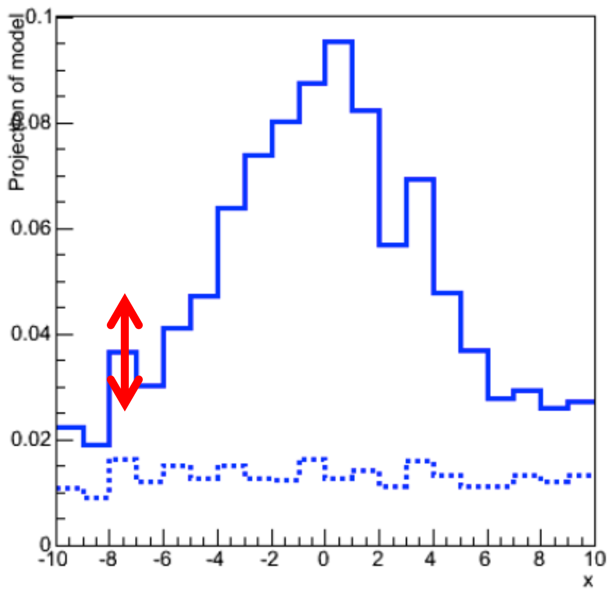
Note that even at $N_{MC}=10N_{data}$ uncertainty on POI can be underestimated by 10% without BB

Reducing the number NPs – Beeston-Barlow 'lite'

- Another approach that is being used is called 'BB' – lite
- Premise: effect of statistical fluctuations on sum of templates is dominant → Use one NP per bin instead of one NP per component per bin

'Beeston-Barlow'

$$L(\vec{N}\,|\,\vec{s},\vec{b}) = \prod_{bins} P(N_i\,|\,s_i+b_i) \prod_{bins} P(\tilde{s}_i\,|\,s_i) \prod_{bins} P(\tilde{b}_i\,|\,b_i)$$



'Beeston-Barlow lite '

$$L(\vec{N}\,|\,\vec{n}) = \prod_{bins} P(N_i\,|\,n_i) \prod_{bins} P(\tilde{s}_i+\tilde{b}_i\,|\,n_i)$$

Response function w.r.t. *n* as parameters

Subsidiary measurements of *n* from s~+b~

$$L(\vec{N}\,|\,\vec{\gamma}) = \prod_{bins} P(N_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i)) \prod_{bins} P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i))$$

Normalized NP lite model (nominal value of all γ is 1)

The interplay between shape systematics and MC systematics

- Best practice for template morphing models is to also include effect of MC systematics
- Note that that for every 'morphing systematic' there is an set of two templates that have their own (independent) MC statistical uncertainties.
  - A completely accurate should model MC stat uncertainties of all templates

$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha \cdot (s_i^+ - s_i^0) & \forall \alpha > 0 \\ s_i^0 + \alpha \cdot (s_i^0 - s_i^-) & \forall \alpha < 0 \end{cases}$$

$$L(\vec{N} \mid \alpha, \vec{s}^{\,-}, \vec{s}^{\,0}, \vec{s}^{\,+}) = \prod_{bins} P(N_i \mid \underbrace{s_i(\alpha, s_i^-, s_i^0, s_i^+)}_{\text{Morphing response function}}) \underbrace{\prod_{bins} P(\tilde{s}_i^- \mid s_i^-) \prod_{bins} P(\tilde{s}_i^0 \mid s_i^0) \prod_{bins} P(\tilde{s}_i^+ \mid s_i^+)}_{\text{Subsidiary measurements}}$$

- But has severe practical problems
  - Can only be done in 'full' Beeston-Barlow model, not in 'lite' mode, enormous number of NP models with only a handful of shape systematics…

The interplay between shape systematics and MC systematics

- Commonly chosen practical solution

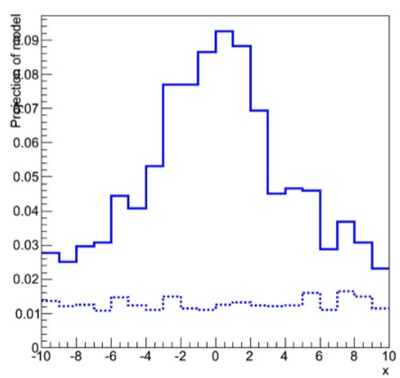
$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha\cdot(s_i^+ - s_i^0) & \forall\alpha>0 \\ s_i^0 + \alpha\cdot(s_i^0 - s_i^-) & \forall\alpha<0 \end{cases}$$

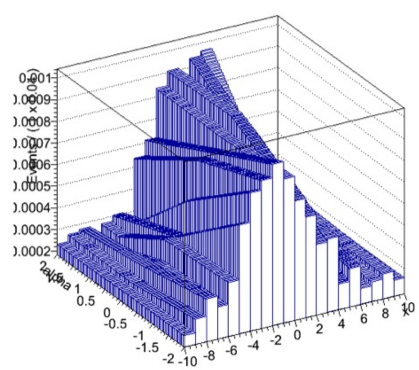
$$L(\vec{N}\,|\,\vec{s},\vec{b}) = \prod_{bins} P(N_i\,|\,\underbrace{\gamma_i\cdot[s_i(\alpha,s_i^-,s_i^0,s_i^+)+b_i]}_{\text{Morphing \& MC response function}})\prod_{bins}\underbrace{P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i\cdot[\tilde{s}_i+\tilde{b}_i])G(0\,|\,\alpha,1)}_{\text{Subsidiary measurements}}$$

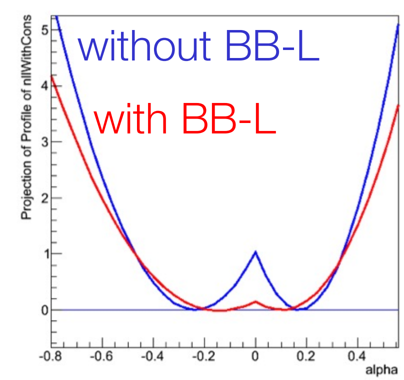
Morphing & MC response function

Subsidiary measurements

Models relative MC rate uncertainty for each bin w.r.t the nominal MC yield, even if morphed total yield is slightly different



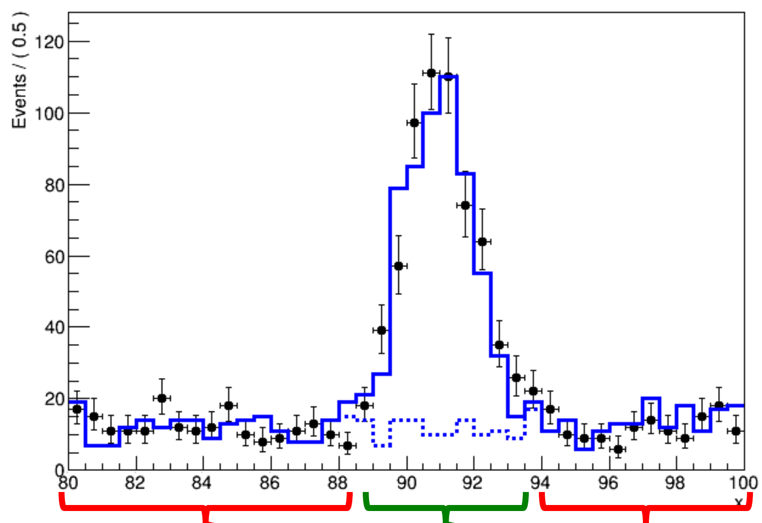




- Approximate MC template statistics already significantly improves influence of MC fluctuations on template morphing
  - Because ML fit can now 'reweight' contributions of each bin

Pruning complexity – MC statistical for selected bins

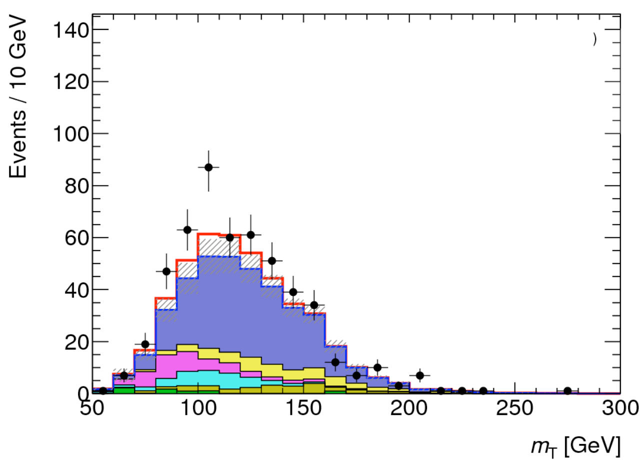
- Can also make decision to model MC statistical uncertainty on a bin-by-bin basis
  - No modeling for high statistics bins
  - Explicit modeling for low-statistics bins

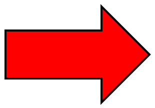


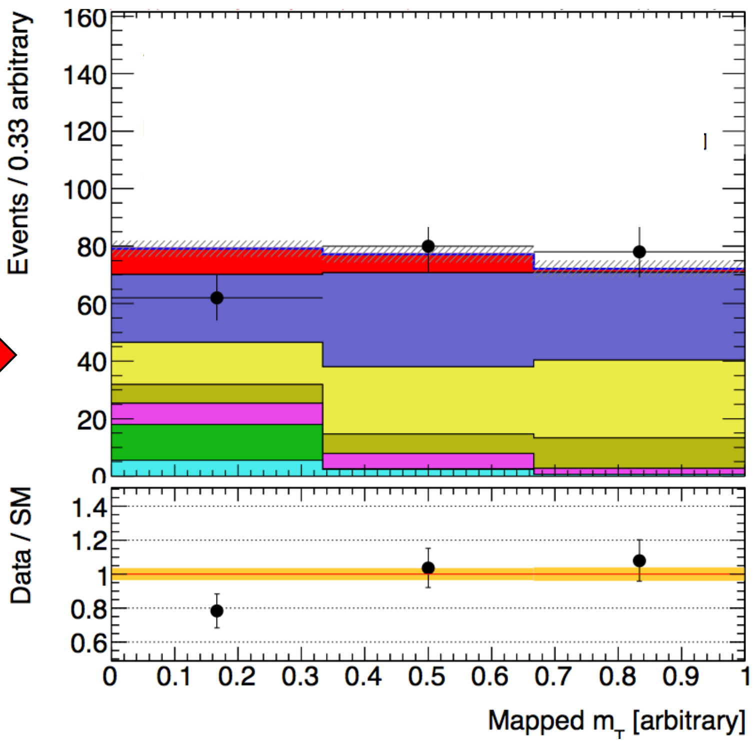
$$L(\vec{N}\,|\,\vec{\gamma}) = \prod_{bins} P(N_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i)) \prod_{low-stats\ bins} P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i)) \prod_{hi-stats\ bins} \delta(\gamma_i)$$

Adapting binning to event density

- Effect of template statistics can also be controlled by rebinning data such all bins contain expected and observed events
  - For example choose binning such that expected background has a uniform distribution (as signals are usually small and/or uncertain they matter less)







Example 4 – Beeston-Barlow light

- Beeston-Barlow-(lite) modeling of MC statistical uncertainties

$$L(\vec{N}\,|\,\vec{\gamma}) = \prod_{bins} P(N_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i)) \prod_{bins} P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i))$$

Reducing the number NPs – Beeston-Barlow 'lite'

- Another approach that is being used is called 'BB' – lite
- Premise: effect of statistical fluctuations on sum of templates is dominant → Use one NP per bin instead of one NP per component per bin

'Beeston-Barlow'

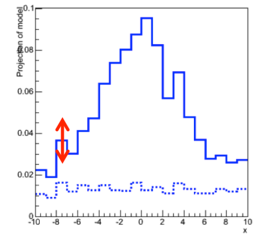
$$L(\vec{N}\,|\,\vec{s},\vec{b}) = \prod_{bins} P(N_i\,|\,s_i+b_i) \prod_{bins} P(\tilde{s}_i\,|\,s_i) \prod_{bins} P(\tilde{b}_i\,|\,b_i)$$

'Beeston-Barlow lite '

$$L(\vec{N}\,|\,\vec{n}) = \prod_{bins} P(N_i\,|\,n_i) \prod_{bins} P(\tilde{s}_i+\tilde{b}_i\,|\,n_i)$$

Response function w.r.t. n as parameters

Subsidiary measurements of n from s~+b~



$$L(\vec{N}\,|\,\vec{\gamma}) = \prod_{bins} P(N_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i)) \prod_{bins} P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i(\tilde{s}_i+\tilde{b}_i))$$

Normalized NP lite model (nominal value of all γ is 1)

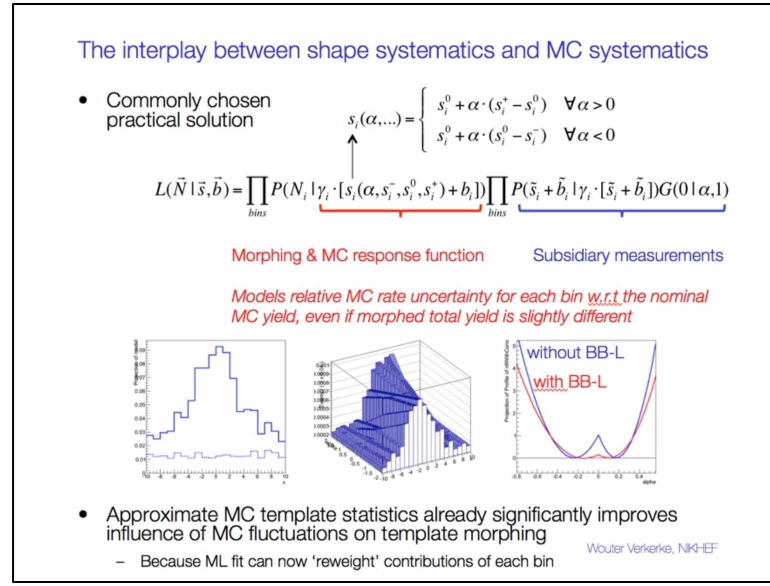
```
// Import template histogram in workspace
 w.import(hs) ;

// Construct parametric template models from histograms
// implicitly creates vector of gamma parameters
 w.factory("ParamHistFunc::s(hs)") ;

 // Product of subsidiary measurement
 w.factory("HistConstraint::subs(s)") ;

 // Construct full model
 w.factory("PROD::model(s,subs)") ;
```

Example 5 – BB-lite + morphing

- Template morphing model with Beeston-Barlow-lite MC statistical uncertainties

$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha\cdot(s_i^+ - s_i^0) & \forall\alpha > 0 \\ s_i^0 + \alpha\cdot(s_i^0 - s_i^-) & \forall\alpha < 0 \end{cases}$$

$$L(\vec{N}\,|\,\vec{s},\vec{b}) = \prod_{bins} P(N_i\,|\,\gamma_i\cdot[\underbrace{s_i(\alpha,s_i^-,s_i^0,s_i^+)}_{} + b_i])\underbrace{\prod_{bins} P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i\cdot[\tilde{s}_i+\tilde{b}_i])G(0\,|\,\alpha,1)}_{}$$




```
// Import template histograms in workspace
 w.import(hs_0,hs_p,hs_m,hb) ;

 // Construct parametric template morphing signal model
 w.factory("ParamHistFunc::s_p(hs_p)") ;
 w.factory("HistFunc::s_m(x,hs_m)") ;
 w.factory("HistFunc::s_0(x[80,100],hs_0)") ;
 w.factory("PiecewiseInterpolation::sig(s_0,s_,m,s_p,alpha[-5,5])") ;

 // Construct parametric background model (sharing gamma's with s_p)
 w.factory("ParamHistFunc::bkg(hb,s_p)") ;

 // Construct full model with BB-lite MC stats modeling
 w.factory("PROD::model(ASUM(sig,bkg,f[0,1]),
            HistConstraint({s_0,bkg}),Gaussian(0,alpha,1))") ;
```