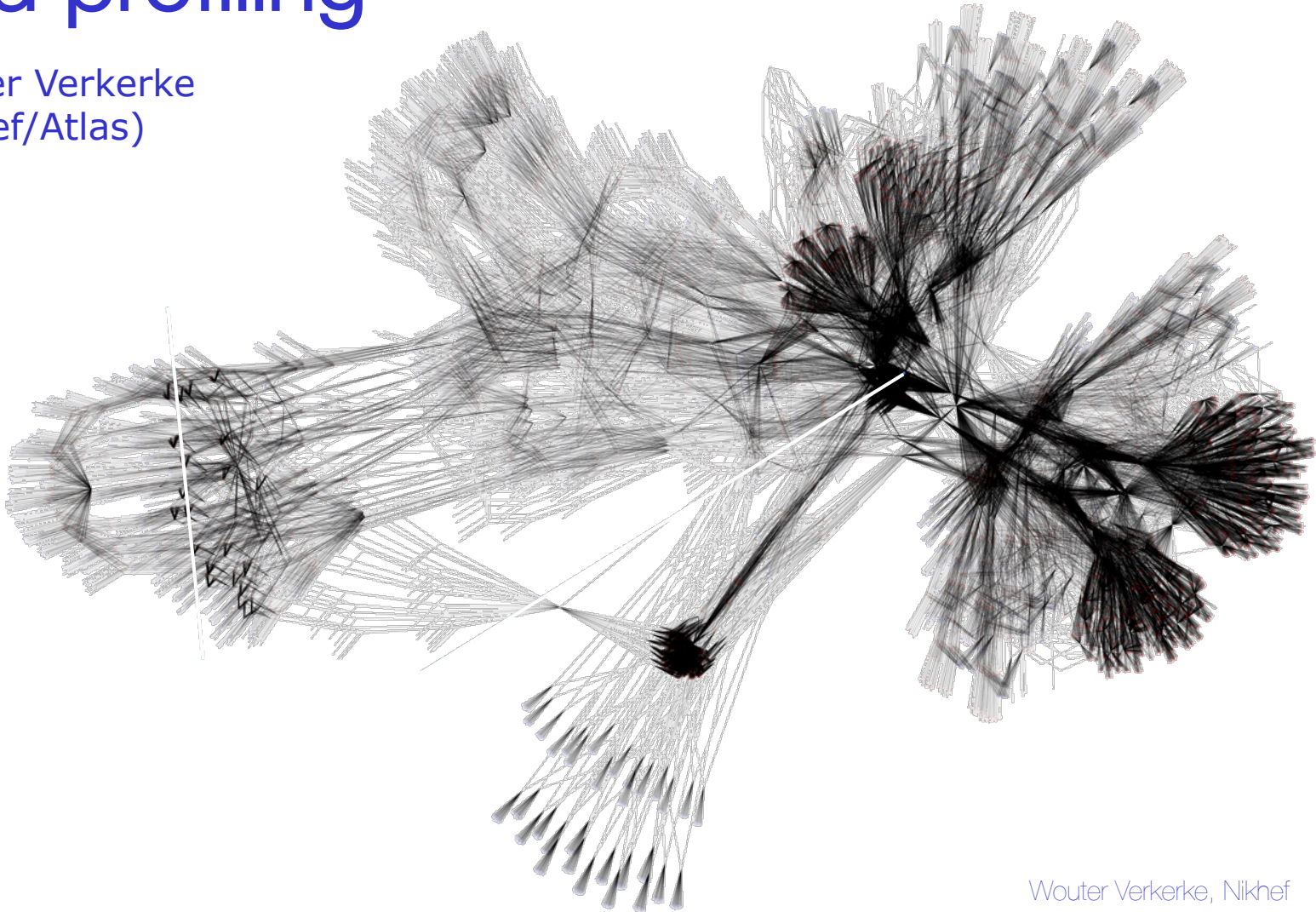


Systematic uncertainties and profiling

Wouter Verkerke
(Nikhef/Atlas)

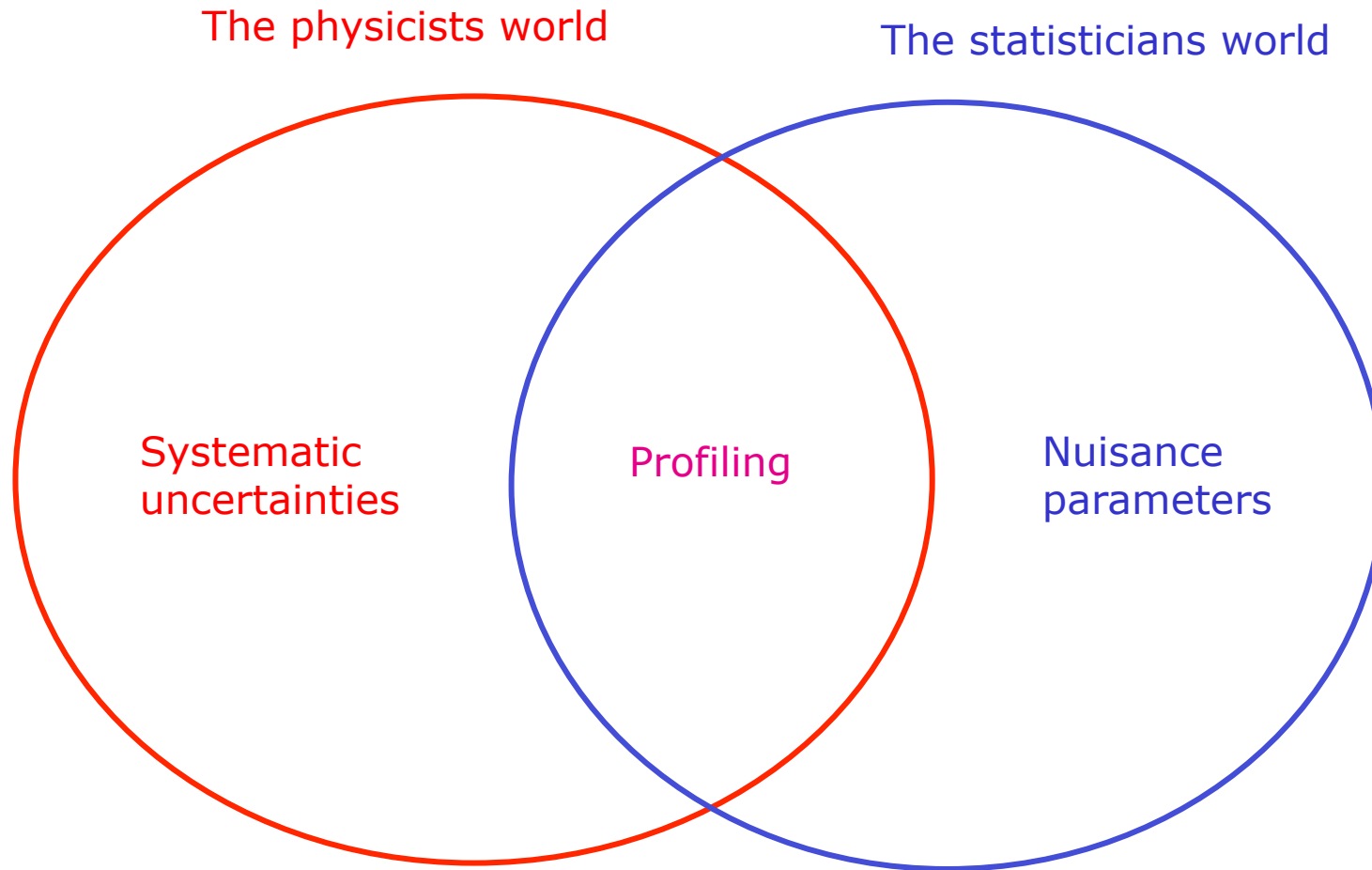


0 The scope of this course

Profiling & Systematics as part of statistical analysis

- A HEP analysis requires close integration of ‘physics concepts’ and ‘statistical concepts’
 1. Design event selection “physics”
 - Use simulated samples of signal, background to aid selection process (cuts, BDT, NN etc)
 2. Analyze (‘fit’) data in selection “statistics”
 - Measurement with statistical error, limit based on statistical uncertainty
 3. Make inventory of systematic uncertainties “physics”
 - Generally, any effect that isn’t measured constrained from your own measurement
 4. Finalize result ‘including systematics’ “statistics”
 - Variety of (empirical/fundamental) approaches to do this
 5. Interpretation “physics”
 - Better measurement, discovery etc, find mistake/sub-optimality in procedure
- Focus of this course: steps 3 and 4.
 - Practical problem: ‘physics notion’ of systematic uncertainties does not map 1-1 to a statistical procedure. Many procedures exist, some ad-hoc, some rigorous (from the statistical p.o.v.)

Profiling & Systematics as part of statistical data analysis



Outline of this course

- Outline of this course
 1. What are systematic uncertainties?
 2. Incorporating systematic uncertainties in probability models
 3. Modeling shape systematics: template morphing
 4. Tools for modelling building RooFit/RooStats and HistFactory
 5. Diagnostics: Overconstraining & choices in model parametrization

1 What are systematic uncertainties?

What are systematic uncertainties?

- 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?

Systematic uncertainty as a hidden measurement

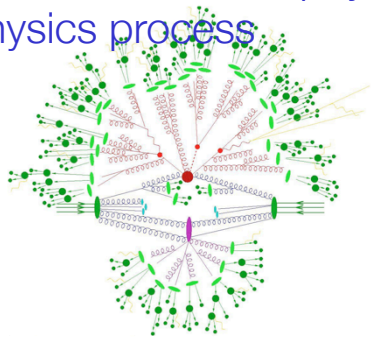
- Consider 2 examples of measurements with systematic uncertainties
- Example 1: Measuring length of an object with a ruler
 - ‘Ruler calibration uncertainty’ is systematic uncertainty on length measurement
- Example 2: Counting measurement a signal in the presence of background
 - Measurement has (Poisson) statistical uncertainty.
 - Uncertainty on rate of background process introduces a systematic uncertainty on estimate of signal rate
- Is the ‘systematic uncertainty’ just a ‘hidden measurement’?
 - Ex 1: Ruler calibration could depend on temperature and uncertainty on current temperature could be dominant component of uncertainty
 - Ex 2: Background rate could be measured by a control sample

Sources of systematic uncertainty in HEP

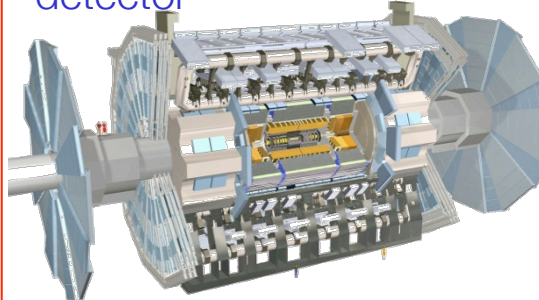
- Detector-simulation related uncertainty
 - Calibrations (electron, jet energy scale)
 - Efficiencies (particle ID, reconstruction)
 - Resolutions (jet energy, muon momentum)
- Theoretical uncertainties
 - Factorization/Normalization scale of MC generators
 - Choice of MC generator (ME and/or PS, e.g. Herwig vs Pythia)
- Monte Carlo Statistical uncertainties
 - Statistical uncertainty of simulated samples

The simulation workflow and origin of uncertainties

Simulation of 'soft physics'
physics process



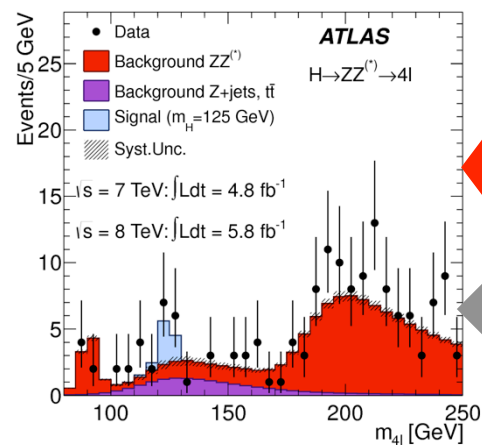
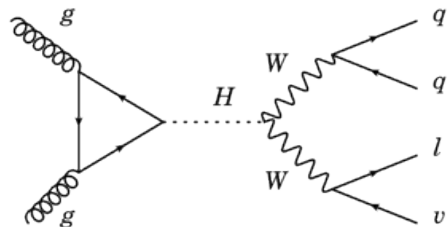
Simulation of ATLAS
detector



LHC data

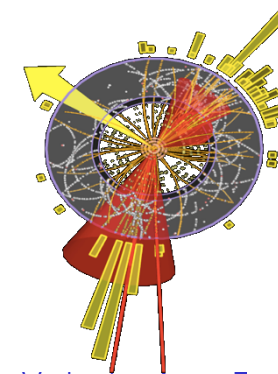


Simulation of high-energy
physics process



Analysis Event selection

Reconstruction
of ATLAS detector



Wouter Verkerke, Nikhef
Wouter Verkerke, NIKHE.

Typical specifications of systematic uncertainties

- Detector-simulation related

- “The Jet Energy scale uncertainty is 5%”
- “The b-tagging efficiency uncertainty is 20% for jets with $p_T < 40$ ”

- Theory related

- “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 related

- Usually left unspecified – but quite clearly defined as a Poisson distribution with the ‘observed number of simulated events’ as mean.
- But if MC events are weighted, it gets a bit more complicated.

- Note that specifications are often phrased as a prescription to be executed on the estimation procedure of the physics quantity of interest (‘vary and rerun...’) or can be easily cast this way.

Evaluating the effect of systematic uncertainties

- Often measurements are treated as a ‘black-box’ (e.g. as if it were a physical device that reports the measurement)
- Inspires a ‘naive’ approach to systematic uncertainty evaluation: simply propagate ‘external systematic uncertainties’ into result
 - Evaluate nominal measurement (through unspecified procedure)

$$\mu_{nom} = \hat{\mu}$$

- Evaluate measurement at ‘ ± 1 sigma’ of some systematic uncertainty

$$\mu_{up} = \hat{\mu}(syst - up)$$

$$\mu_{down} = \hat{\mu}(syst - down)$$

- Calculate systematic uncertainty on measurement through numeric error propagation

$$\sigma_{\mu}(syst) = [\mu_{up} - \mu_{down}] / 2$$

- Repeat as needed for all systematic uncertainties, add in quadrature for total systematic uncertainty.

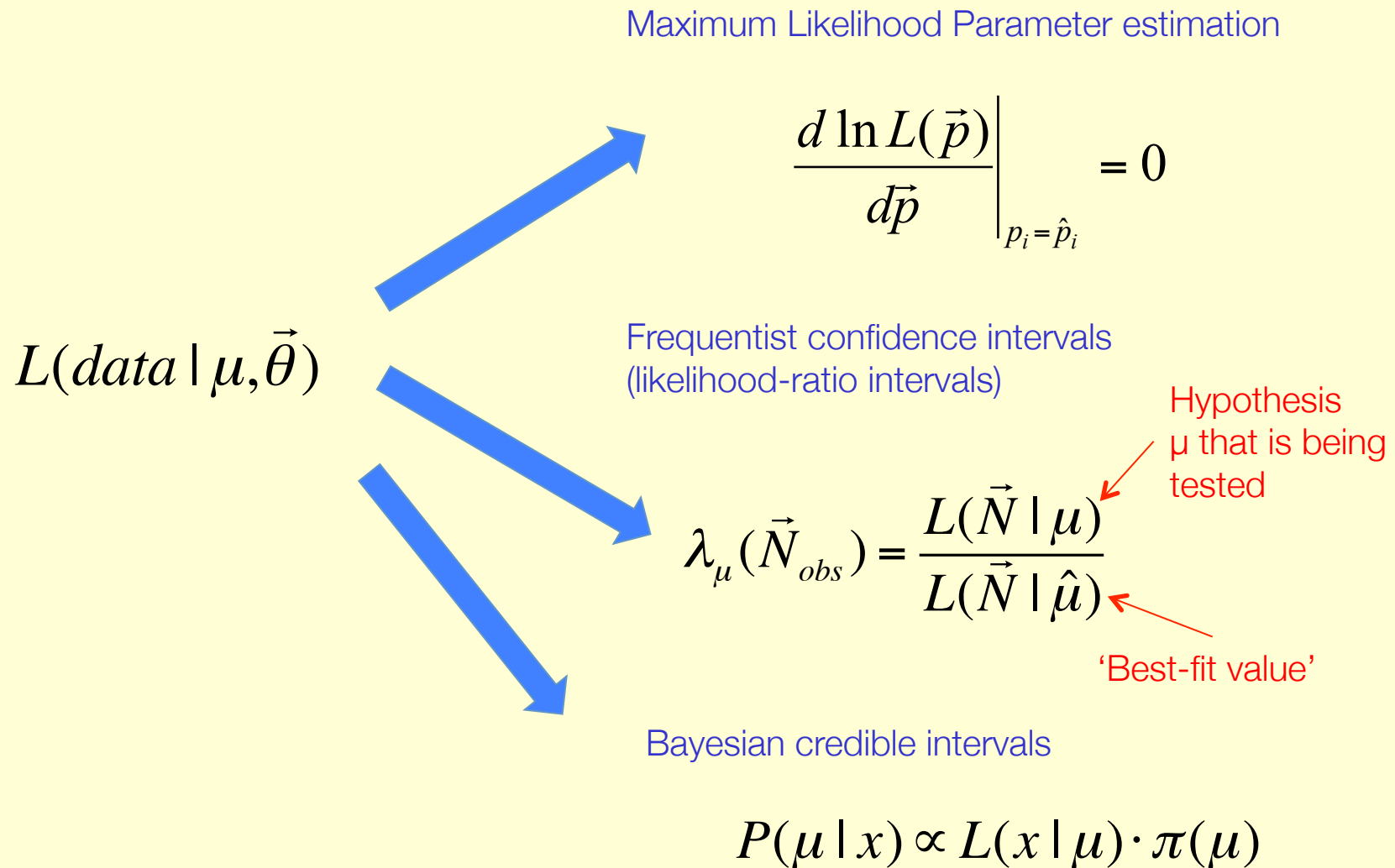
$$\mu_{meas} = \mu_{nom} \pm \sigma(JES) \pm \dots$$

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
 - A maximum likelihood measurement is really nothing like a 'device'
 - 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
 - **It's not based rigorous procedures** (i.e. evaluation of systematic uncertainties is completely detached from statistical procedure used to estimate physics quantity of interest)
 - No calibrated probabilistic statements possible (95% C.L.)
 - No known good procedure for limit setting
- 'Profiling' → Incorporate a description of systematic uncertainties in the likelihood function that is used in statistical procedures

2 Incorporating systematic uncertainties in the likelihood

The likelihood is at the basis of many statistical techniques



Introduction

- All fundamental statistical inference techniques are based on the likelihood. Thus **all aspects of a measurement – including systematic uncertainties – must be contained in the likelihood**
- Will now focus on how to express systematic uncertainties (an experimental science concept) into a likelihood (a statistical concept)
- This starts with an examination of what **we precisely mean with a systematic uncertainty.**
 - Will discuss this based on examples taken from the different classes of systematic uncertainty commonly encountered in HEP analyses
 - For illustrational clarity will for now only focus on systematic uncertainties on counting measurements (systematic uncertainties in measurements of distributions will follow later)

Modeling systematic uncertainties in the likelihood

- What is a systematic uncertainty? It consists of
 - 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.
 - 3: A distribution of possible values for the parameters
 - In practice these (response) models are often only formulated implicitly, but modeling of systematic uncertainties in the likelihood requires an explicit model
- Example of ‘typical’ systematic uncertainty prescription

“The Jet Energy Scale Uncertainty is 5%”
- Note that example does not meet definition standards above
 - Specification specifies variance of the distribution unknown parameter, *but not the distribution* itself (is it Gaussian, Poisson, something else)
 - *Response model left unspecified*

Formulating a response model

- Why does the statement

“the JES uncertainty is X%”

not a formulate a response model, while an additional statement

“If the JES is off by +X%, the energy of every jet in the event is increased by X%”

does constitute a response model?

- The first statement doesn't specify any correlation between jets with different kinematics
 - Can low pT jets be miscalibrated by -4% and high pT jets be calibrated by +5%?
 - Or must all jets be miscalibrated by exactly the same amount?
- The former interpretation would require 2 (or more) model parameters to capture the effect of the miscalibration of the simulation, the latter only one.
- Once the response model is defined, the effect of a systematic uncertainty is deterministically described, up to an (a set of) unknown strength parameter(s).

Formulating a response model

- Note that the construction of a response model for a systematic uncertainty is no different from choosing a model to describe your physics of interest
 - You define a model that deterministically describes the consequences of the underlying hypothesis, up to set of (*a priori*) unknown model parameter
- Will (for now) assume that for our example measurement the example systematic uncertainty – the Jet Energy Scale – can be correctly described with a single parameter that coherently moves the calibration of all jets in the event.
 - The correctness of such an assumption we'll revisit later (but note that this is a *physics* argument)

Modeling the strength parameter

- What do we know about distribution of the corresponding strength parameter?
 - The $\sqrt{\text{variance}}$ of the distribution was specified to be 5%
- But a variance does not completely specify a distribution
 - Does the JES measurement follow a Gaussian distribution?
 - Does the JES measurement follow a Poisson distribution?
 - Or, a ‘block-shaped’ distribution, or anything else?
- *Not* specified by “JES is 5%” prescription
 - Often not a difficult issue as detector-related uncertainties, as these since they are based on (calibration) measurements (and/or central limit theorem applies) → Gaussian or Poisson distribution
 - For theory uncertainties this can be tricky, what distribution to assume for ‘renormalization scale uncertainty’? Will come back to this later

Formalizing systematic uncertainties

- The original systematic uncertainty prescription

“the JES uncertainty is 5%”


- The formalized prescription for use in statistical analysis

“There is a calibration parameter in the likelihood of which the true value is unknown

The distribution of this parameter is a Gaussian with a 5% width

The effect of changing the calibration by 1% is that energy of all jets in the event is coherently increased by 1% ”

Putting it all together – a calibration uncertainty in a counting experiment

- Counting experiment $P(n|s + b) = \frac{(s + b)^n}{n!} e^{-(s+b)}$ 
- Background 'b' is estimated from MC simulation with some uncertainty
 - We estimate b using Monte Carlo simulation: we conclude that we expect 5.0 background events, with a negligible MC statistical uncertainty
 - But, since we use MC simulation we are sensitive to detector simulation uncertainties and theoretical cross-section uncertainties
- Ex: how to model effect of data/MC JES miscalibration uncertainty?
 - The effect of the JES calibration uncertainty is described by a single parameter that coherently moves jet calibration for all jets by same amount
 - Jet calibration group assigns a 5% Gaussian uncertainty to this parameter
 - You determine that a 1% coherent shift of jet-energy scale results in a 2% acceptance change for the background in your signal region.

'naïve approach': vary b by $\pm 2\%$ and propagate effect to s .
How do you put that in the likelihood?

Putting it all together – a calibration uncertainty in a counting experiment

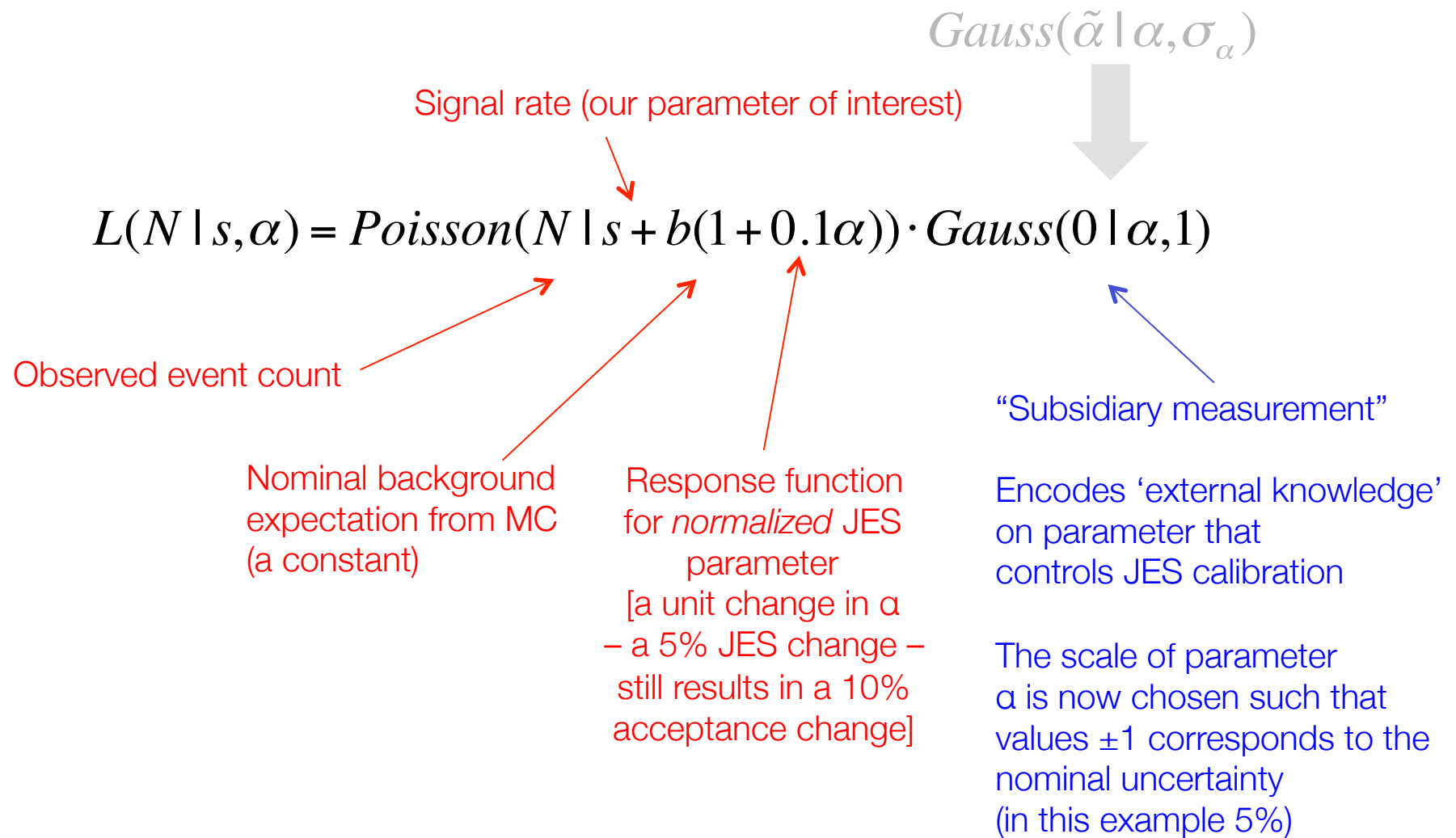
- The likelihood including systematic uncertainty $L(N | s) = \text{Poisson}(N | s + b)$

The diagram illustrates the likelihood function $L(N, \tilde{\alpha} | s, \alpha) = \text{Poisson}(N | s + b(\alpha / \tilde{\alpha}) \cdot 2) \cdot \text{Gauss}(\tilde{\alpha} | \alpha, \sigma_{\alpha})$ with the following annotations:

- Signal rate (our parameter of interest)**: Points to s in the Poisson distribution.
- Observed event count**: Points to N in the Poisson distribution.
- Nominal background expectation from MC (a constant), obtained with $a = \tilde{a}$** : Points to b in the Poisson distribution.
- Response function for JES uncertainty (a 1% JES change results in a 2% acceptance change)**: Points to the term $(\alpha / \tilde{\alpha}) \cdot 2$ in the Poisson distribution.
- Nominal calibration**: Points to α in the Gaussian distribution.
- Assumed calibration**: Points to $\tilde{\alpha}$ in the Gaussian distribution.
- Uncertainty on nominal calibration**: Points to σ_{α} in the Gaussian distribution.
- "Subsidiary measurement"**: Points to the entire Gaussian distribution term.
- Encodes 'external knowledge' on JES calibration**: Points to the entire Gaussian distribution term.

Putting it all together – a calibration uncertainty in a counting experiment

- Simplify expression by renormalizing “subsidiary measurement”



Putting it all together – a calibration uncertainty in a counting experiment

- Sources of information

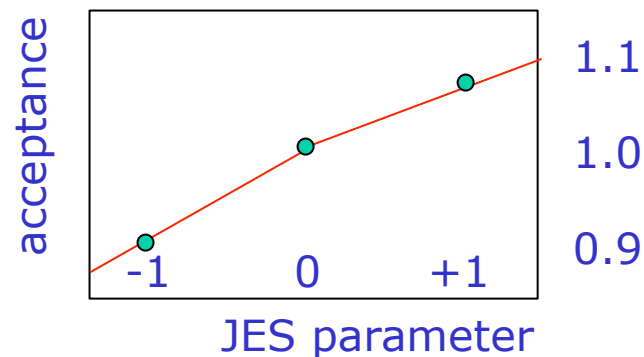
The subsidiary measurement is an implementation of information that is given to you.

It is effectively a likelihood function that ‘measures’ the JES parameter with unit Gaussian uncertainty.

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

The response function is something that *you* measure in your physics analysis.

It must be implemented as a continuous function but can be a linear interpolation, e.g. based on two or three acceptance calculations



Names and conventions

- The full likelihood function of the form

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

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 alpha, but this labeling commonly lead to mistaken statements (e.g. that it is a pdf for α)
 - It is explicitly *not* a pdf $f(\alpha|...)$. It is a (simplified) Likelihood that represents calibration measurement that measures the parameter α , based on calibration data sample that is removed in the simplification (and for which a placeholder 0 value is inserted)

~~$\text{Gauss}(\alpha | 0, 1)$~~

$\text{Gauss}(0 | \alpha, 1)$

Placeholder observable in subsidiary measurement is often called a ‘global observable’

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

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

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 constraints, we substitute L_{JES} with simplified (Gaussian) form, but the statistical treatment and interpretation remains the same

Another example – sideband measurements

- Consider again the same counting measurement

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

- Now b is estimated from a sideband measurement instead of MC simulation.

- Joint likelihood of signal count and sideband count is

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

Constant factor τ accounts for possible
size difference of signal/background region

- Nobody will consider the uncertainty on b in the signal region a systematic uncertainty (since it is constrained from side-band data), but note the similarity in the full likelihood with the ‘JES’ systematic uncertainty

$$L(N, 0 | s, \alpha_{JES}) = \text{Poisson}(N | s + b(1 + 0.1\alpha_{JES})) \cdot \text{Gauss}(0 | \alpha_{JES}, 1)$$

Sideband measurements with systematic uncertainties

- Sideband measurements can also be affected by systematic uncertainties

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

- Above model has effectively has a constant ‘response function’ implemented by the factor τ , which is ratio of bkg acceptance in SR to CR, **but this ratio estimate may be affected by detector simulation uncertainties such as JES.**
- How can we implement the effect of JES uncertainty in the ‘transport factor’ of the background estimate from CR to SR?

$$L(N, N_{ctl}, 0 | s, b, \alpha_{JES}) = \text{Poisson}(N | s + b) \cdot \text{Poisson}(N_{ctl} | \underbrace{\tau(1 + X\alpha_{JES})}_{\text{JES response model for ratio } b_{SR}/b_{CR}} \cdot b) \cdot \underbrace{\text{Gauss}(0 | \alpha_{JES}, 1)}_{\text{Subsidiary measurement of JES response parameter}}$$

JES response model for ratio b_{SR}/b_{CR}

Subsidiary measurement
of JES response parameter

MC statistical uncertainties as systematic uncertainty

- In original JES uncertainty example, the MC statistical uncertainty was ignored (since 100Mevt were available)
- What should you do if MC statistical uncertainties cannot be ignored?
- 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’)

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

Constant factor $\tau = L(\text{MC})/L(\text{data})$




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

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

MC statistical uncertainties as systematic uncertainty

- For notational convenience parameters associated with MC statistical uncertainty are expressed as renormalized γ parameters, similar to the renormalized α parameters

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


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

where b is now a constant expression
(nominal lumi-normalized event count)
and γ is a parameter with nominal value 1

- Just for fun & completeness: the full likelihood with modeling of both MC statistical uncertainty and JES uncertainty.

$$L(N | s, \alpha_{JES}, \gamma) = \text{Poisson}(N | s + (1 + X\alpha_{JES})\gamma b) \cdot \text{Poisson}(N_{MC} | \tau\gamma b) \cdot \text{Gauss}(0 | \alpha_{JES}, 1)$$

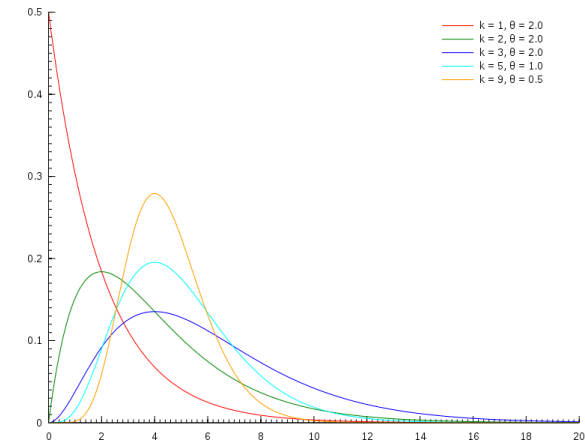
Overview of common subsidiary measurement shapes

- Gaussian $G(x|\mu, \sigma)$

- ‘Default’, motivated by Central Limit Theorem (asymptotic distribution for sum of random variables)

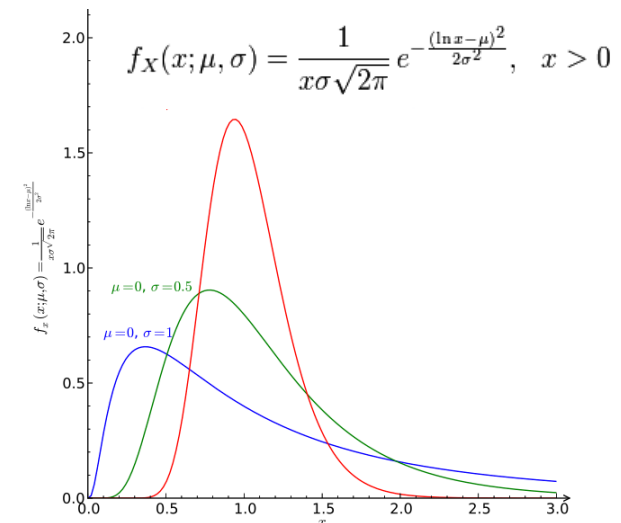
- (Rescaled) Poisson $P(N|\mu\tau)$ \longrightarrow

- Obvious choice for any subsidiary measurement that is effectively a counting experiment
- NB: For a Poisson model the distribution in μ is a Gamma distribution (posterior of Poisson)
- Scale factor τ allows to choose variance independently of mean (e.g. to account for side-band size ratio, data/mc lumi ratio)



- LogNormal $LN(x|\mu, \sigma)$ \longrightarrow

- Asymptotic distribution for product of random variables
- Appealing property for many applications is that it naturally truncates at $x=0$



Specific issues with theory uncertainties

- Modeling of **theoretical** syst. uncertainties follows familiar pattern
 - Define response
 - Define distribution for the ‘subsidiary measurement’
 - Construct full likelihood
- But **distribution of subsidiary theory measurement** can be a thorny issue
 - For detector simulation uncertainties, subsidiary measurement usually based on actual measurement → Central Limit Theorem → convergence to Gaussian distribution when measurement is based on many events
 - This argument does not always apply to theoretical uncertainties, as there may be no underlying measurement
- Example: (N)LO scale uncertainties in Matrix Element calculations
 - Typical prescription “vary to 0.5x nominal and 2x nominal and consider the difference” makes no statement on distribution
 - Yet proper statistical treatment of such an uncertainty (i.e. modeling in the likelihood) demands a specified distribution
 - Not clear what to do. You can ask theory expert, but not clear if has a well-motivated choice of distribution...
 - In any case if choice of distribution turns out not to matter too much, you just pick one.

Specific issue with theory uncertainties

- Worst type of ‘theory’ uncertainty are prescriptions that result in an observable difference that cannot be ascribed to clearly identifiable effects
- Examples of such systematic prescriptions
 - Evaluate measurement with CTEQ and MRST parton density functions and take the difference as systematic uncertainty.
 - Evaluate measurement with Herwig and Pythia showering Monte Carlos and take the difference as systematic uncertainty
- I call these ‘2-point systematics’.
 - You have the technical means to evaluate two known different configurations, but reasons for underlying difference are not clearly identified.

Specific issue with theory uncertainties

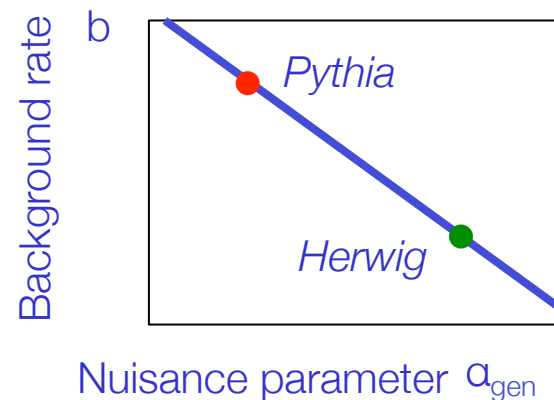
- It is difficult to define rigorous statistical procedures to deal with such 2-point uncertainties. So you need to decide
- If their estimated effect is small, you can pragmatically ignore these lack of proper knowledge and 'just do something reasonable' to model these effects in a likelihood
- If their estimated effect is large, your leading uncertainty is related to an effect that largely ununderstood effect. This is bad for physics reasons!
 - You should go back to the drawing board and design a new measurement that is less sensitive to these issues.
 - Hypothetical example:
 - * You measure an inclusive cross-section.
 - * But Pythia-Herwig effect is largest uncertainty, originates from the visible-to-inclusive acceptance factor.
 - * Does it make to publish the inclusive cross-section, or is it better to publish visible cross-section in some well-defined fiducial range?
 - * Your measurement can then contribute to further discussion and validation of various showering MC packages.

Specific issues with theory uncertainties

- Pragmatic solutions to likelihood modeling of ‘2-point systematics’
- Final solution will need to follow usual pattern

$$L(N | s, \alpha) = \text{Poisson}(N | s + b \cdot f(\alpha)) \cdot \text{SomePdf}(0 | \alpha)$$

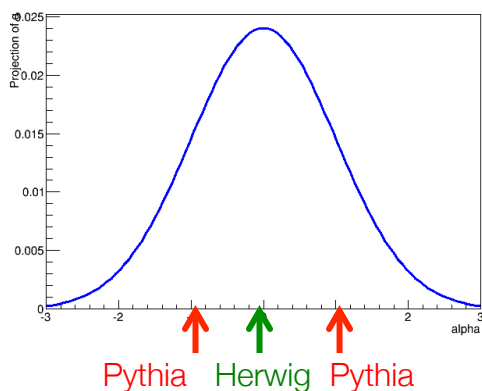
- Since underlying concept of systematic uncertainty not defined, the only option is to *define its meaning terms in terms of response in the physics measurement*
- Example
 - Estimate of bkg with Herwig = 8, with Pythia = 12
 - In the likelihood choose $b=8$ and then define $f(\alpha) = |1+4\alpha|$, so that $f(0)$ results in ‘Herwig ($b \cdot f=8$)’ and $f(\pm 1)$ results in ‘Pythia ($b \cdot f=12$)’
 - For lack of a better word you could call α now the ‘Herwigness of fragmentation w.r.t its effect on my background estimate’
- A thorny question remains: What is the subsidiary measurement for α ?
 - This should reflect you current knowledge on α .



Specific issues with theory uncertainties

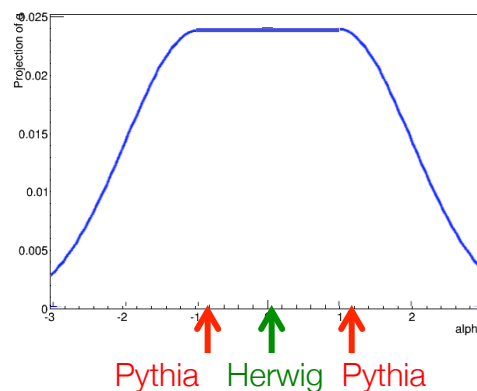
- Subsidiary measurement of a theoretical 2-point uncertainty effectively quantifies the ‘knowledge’ on these models
 - Formally staying in concepts of frequentist statistics here: likelihood of subsidiary measurement $L(x|\alpha)$ is strictly $P(\text{data}|\text{theory})$, but you ‘data’ here is not really data but something that quantifies your belief since you have no data on this problem.
 - I realize this sounds very much like “you have no idea what you’re doing”, but to some extent this is precisely the problem with 2-point systematics – you really don’t know (or decided not to care about) the underlying physics issues.
- Some options and their effects

Gaussian



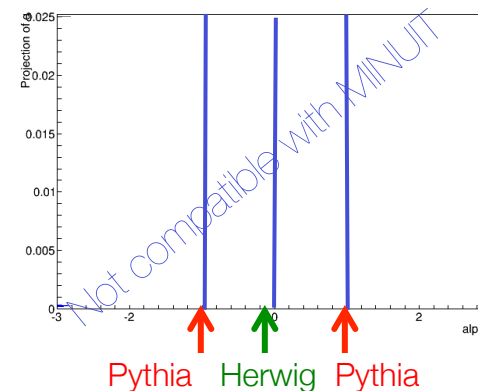
Prefers Herwig at 1σ

Box with Gaussian wings



All predictions ‘between’ Herwig and Pythia equally probable

Delta fuctions

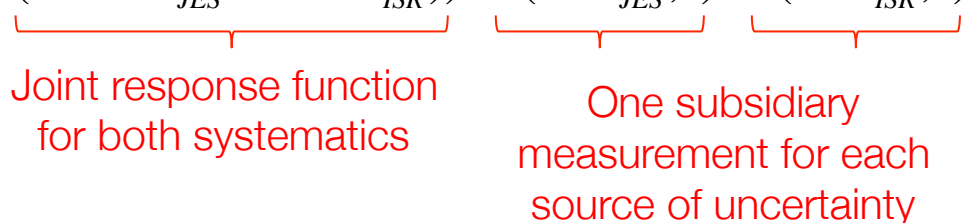


Only ‘pure’ Herwig and Pythia exist

Modeling multiple systematic uncertainties

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

$$L(N, 0 | s, \alpha_{JES}, \alpha_{ISR}) = P(N | s + b(1 + 0.1\alpha_{JES} + 0.05\alpha_{ISR})) \cdot G(0 | \alpha_{JES}, 1) \cdot G(0 | \alpha_{ISR}, 1)$$



Joint response function
for both systematics

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

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

- 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

Summary on likelihood modeling of systematic uncertainties

- To describe a systematic uncertainty in a likelihood model you need
 - A response model that deterministically describes the effect underlying the uncertainty (e.g. a change in calibration). Such a model has one or more parameters that control the strength of the effect
 - The ‘external knowledge’ on the strength of the effect is modeled as Likelihood representing the ‘subsidiary measurement’ through which this knowledge was obtained
 - Conceptually this is identical to including the likelihood of the actual calibration measurement in the likelihood of the physics analysis
 - In practice a simplified form of the measurement is included, but you must choose an explicit distribution that best represents the original measurement. For systematic uncertainties that related to external measurements (calibrations), this is often a Gaussian or Poisson distribution
- Modeling prescription can easily be repeated to extend describe effect of multiple uncertainties in multiple simultaneous measurement
 - Conceptually it is not more complicated, but technically it can get tedious. We have good tools for this → will discuss these later

Summary on likelihood modeling of systematic uncertainties

- Often the process of modeling uncertainties in the likelihood requires **information that is traditionally not provided** as part of a systematic uncertainty prescription
- **This is good thing** – your evaluation of these uncertainties otherwise relies on tacit assumptions on these. **Discuss modeling assumptions you make with the prescription ‘provider’**
- You may also learn that your measurement is strongly affect by something you don’t know (e.g. distribution of a theory uncertainty). **This is also a good thing**. This is a genuine physics problem, that you might have otherwise overlooked
- Theory uncertainty modeling can pose difficult questions
 - Usually discovered 3 days before approval deadline, tendency is to ‘be conservative’ and not think much about problem. ‘Conservative’ solution tend to be ‘naïve error propagation’ → problem gets hidden behind unspecified assumptions of that method.

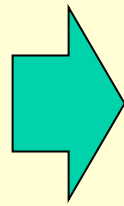
Dealing with nuisance parameters – The profile likelihood ratio

- Once we introduced systematic uncertainties as ‘nuisance parameters’, we need to account for them in the statistical inference
- For frequentist confidence intervals with LR test statistic, incorporate ‘new’ parameters θ as follows:

Likelihood for given μ

$$\lambda(\mu) = \frac{L(\mu)}{L(\hat{\mu})}$$

Maximum Likelihood



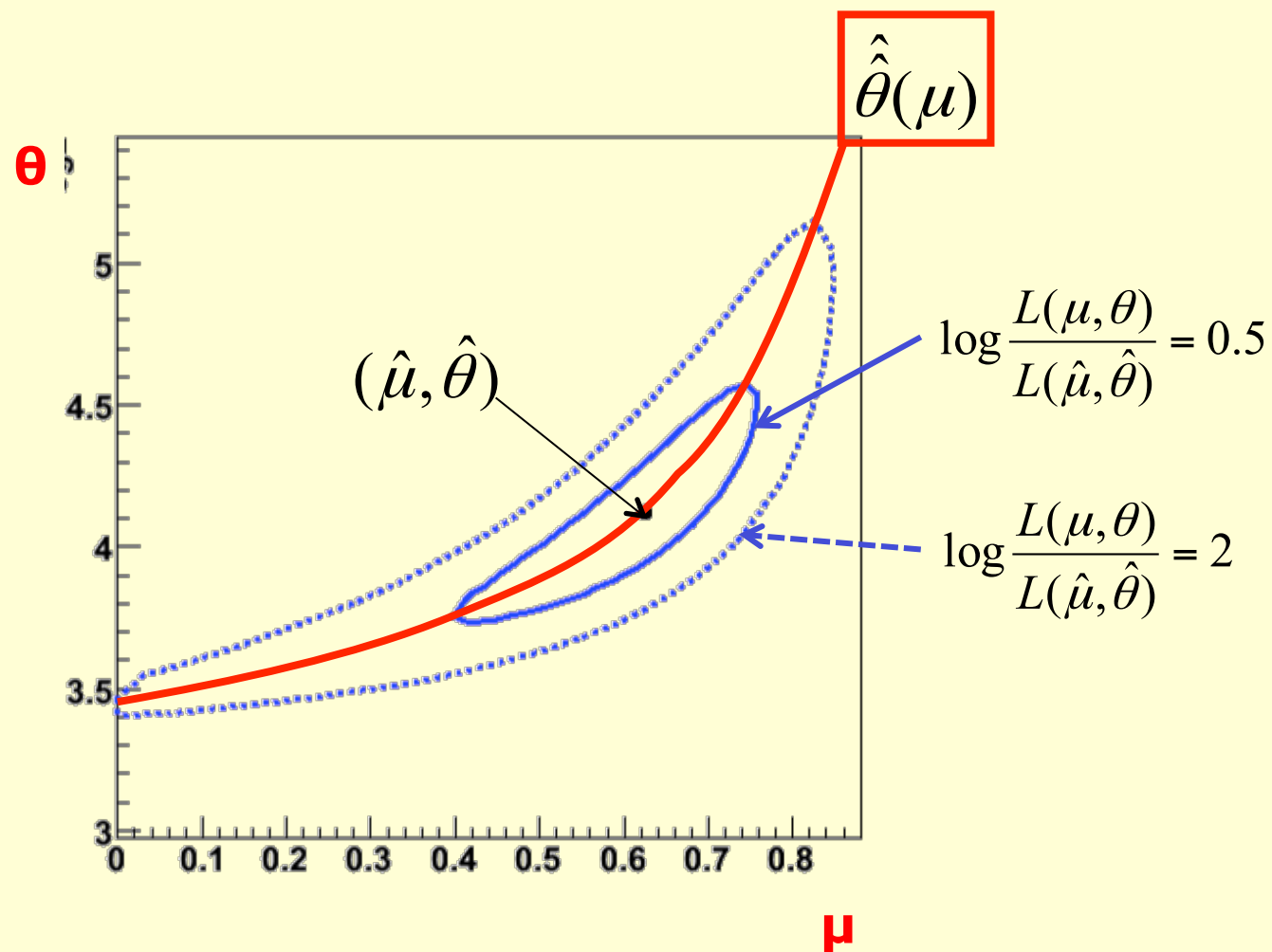
Maximum Likelihood for given μ

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$

Maximum Likelihood

- NB: value profile likelihood ratio does *not* depend on θ

Profiling illustration with one nuisance parameter

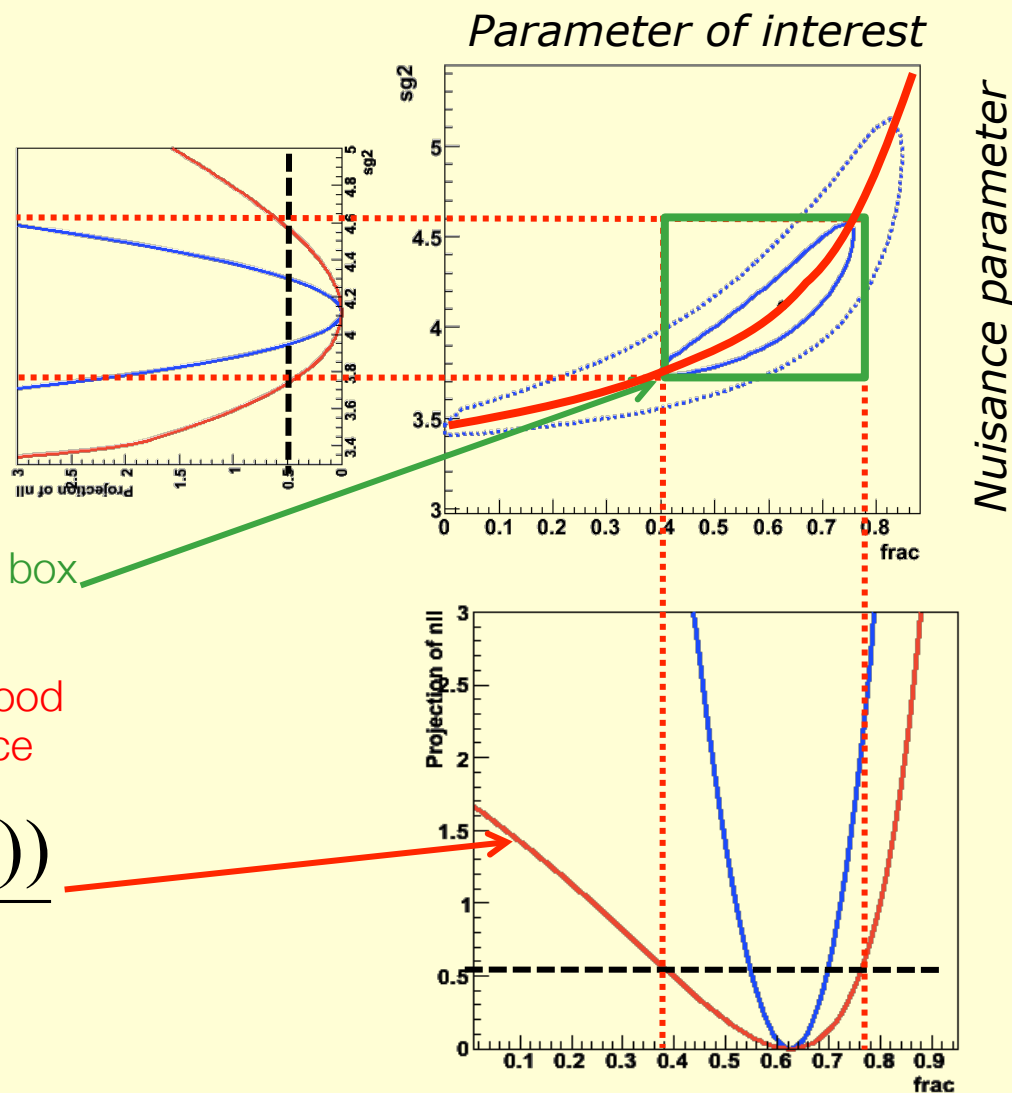


Link between MINOS errors and profile likelihood

- Note that MINOS algorithm in MINUIT gives same uncertainties as interval on Profile Likelihood Ratio

- MINOS errors is bounding box around $\lambda(s)$ contour
- Profile Likelihood = Likelihood minimized w.r.t. all nuisance parameters

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$



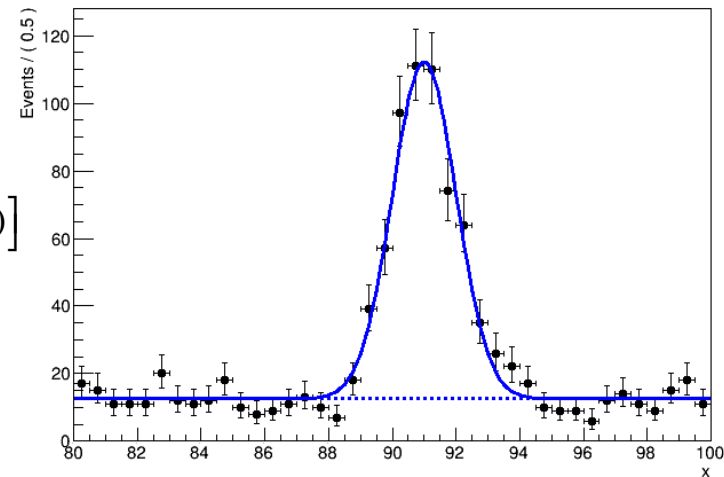
3 Modeling shape systematics: template morphing

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

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



- 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

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

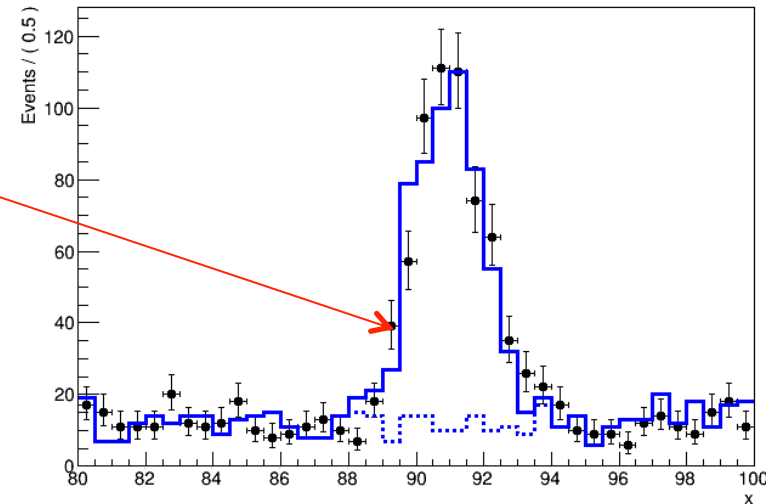
Response function

Subsidiary measurement

Analytical versus non-parametric shapes

- At hadron colliders (including), analytical distributions for signal and background shapes are usually not available
- Instead rely on MC simulation chain to obtain distribution → knowledge of distribution is a histogram of expected yields in bins of a discriminating observable

$$L(\vec{N} | \mu) = \prod_i \text{Poisson}(N_i | \mu s_i + b_i)$$



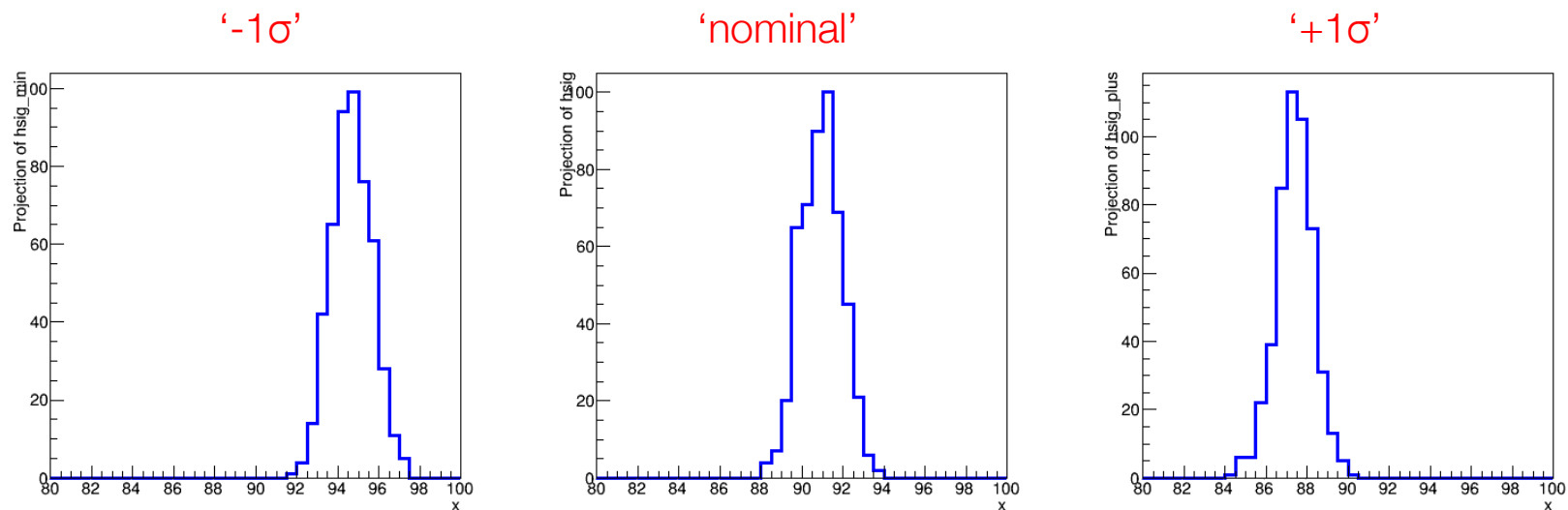
- Modeling of a *rate* systematic uncertainty is straightforward:

$$L(\vec{N} | \mu, \alpha) = \prod_i \text{Poisson}(N_i | \underbrace{\mu s_i \cdot (1 + 3.75\alpha)}_{\text{Response function}} + b_i) \cdot \underbrace{\text{Gauss}(0 | \alpha, 1)}_{\text{Subsidiary measurement}}$$

What about a systematic effect that shifts the mean?

Modeling of shape systematics in the likelihood

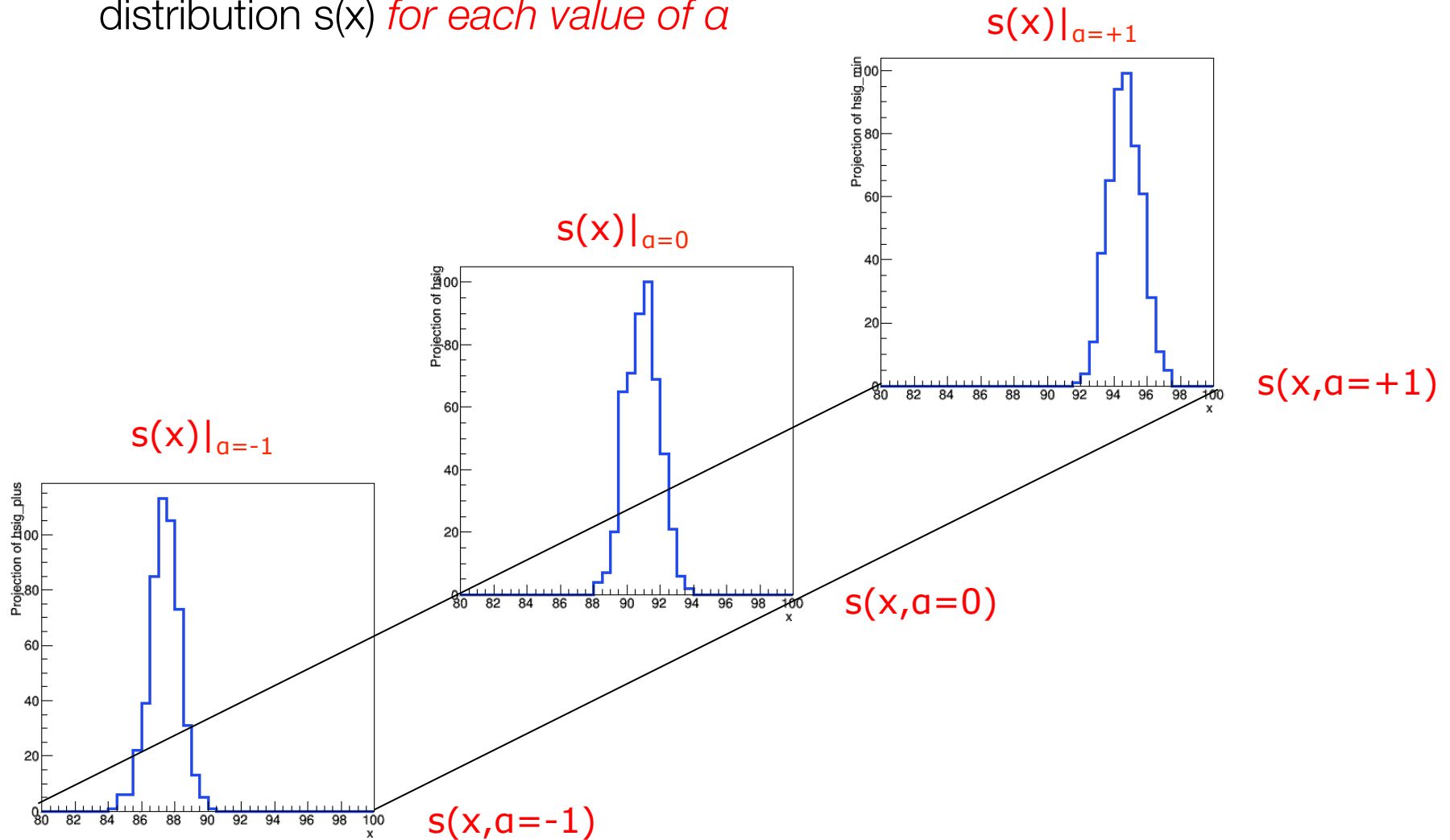
- Effect of *any* systematic uncertainty that affects the shape of a distribution can in principle be obtained from MC simulation chain
 - Obtain histogram templates for distributions at ‘ $+1\sigma$ ’ and ‘ -1σ ’ settings of systematic effect



- Now construct a response function based on the shape of these three templates.

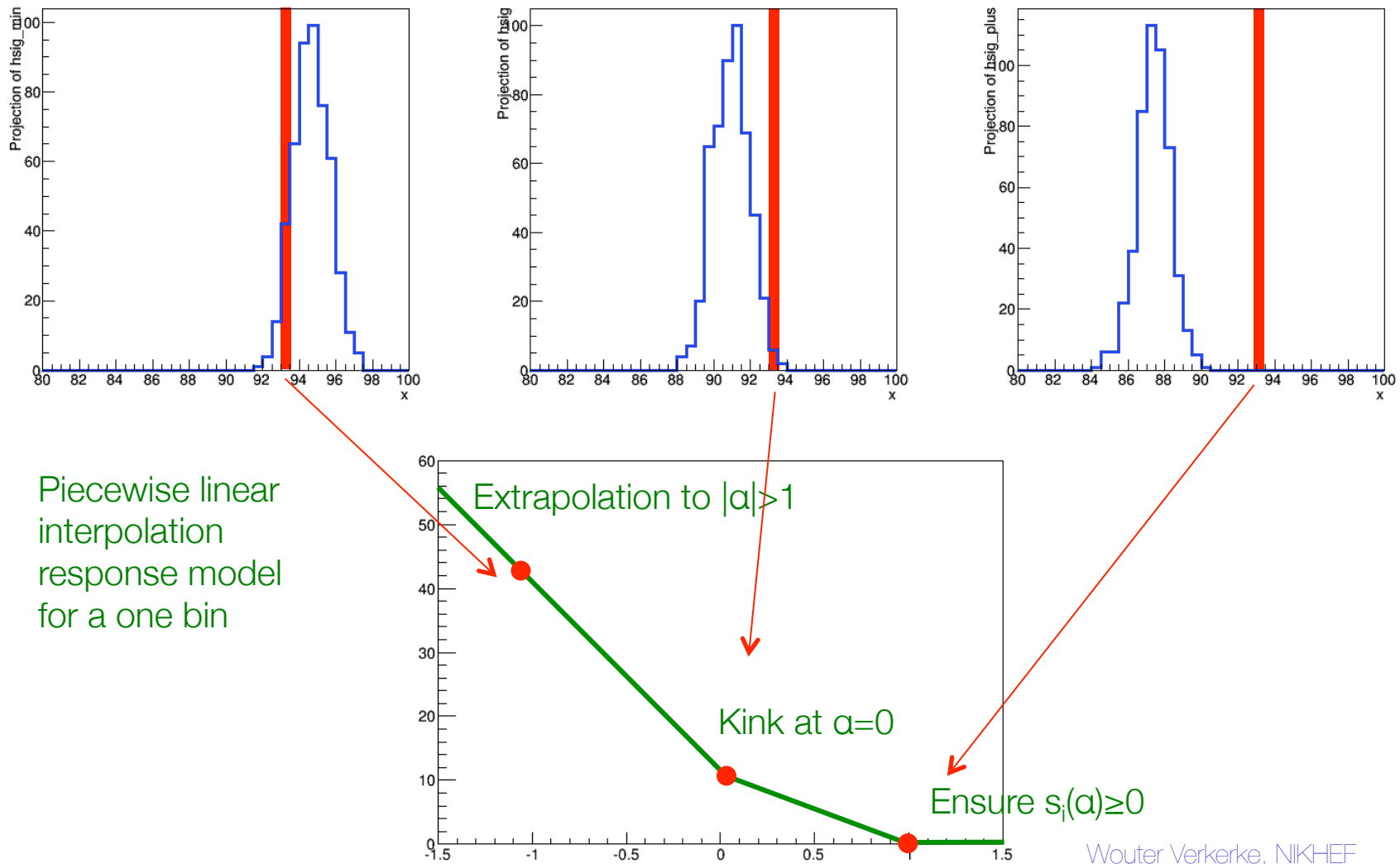
Need to interpolate between template models

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

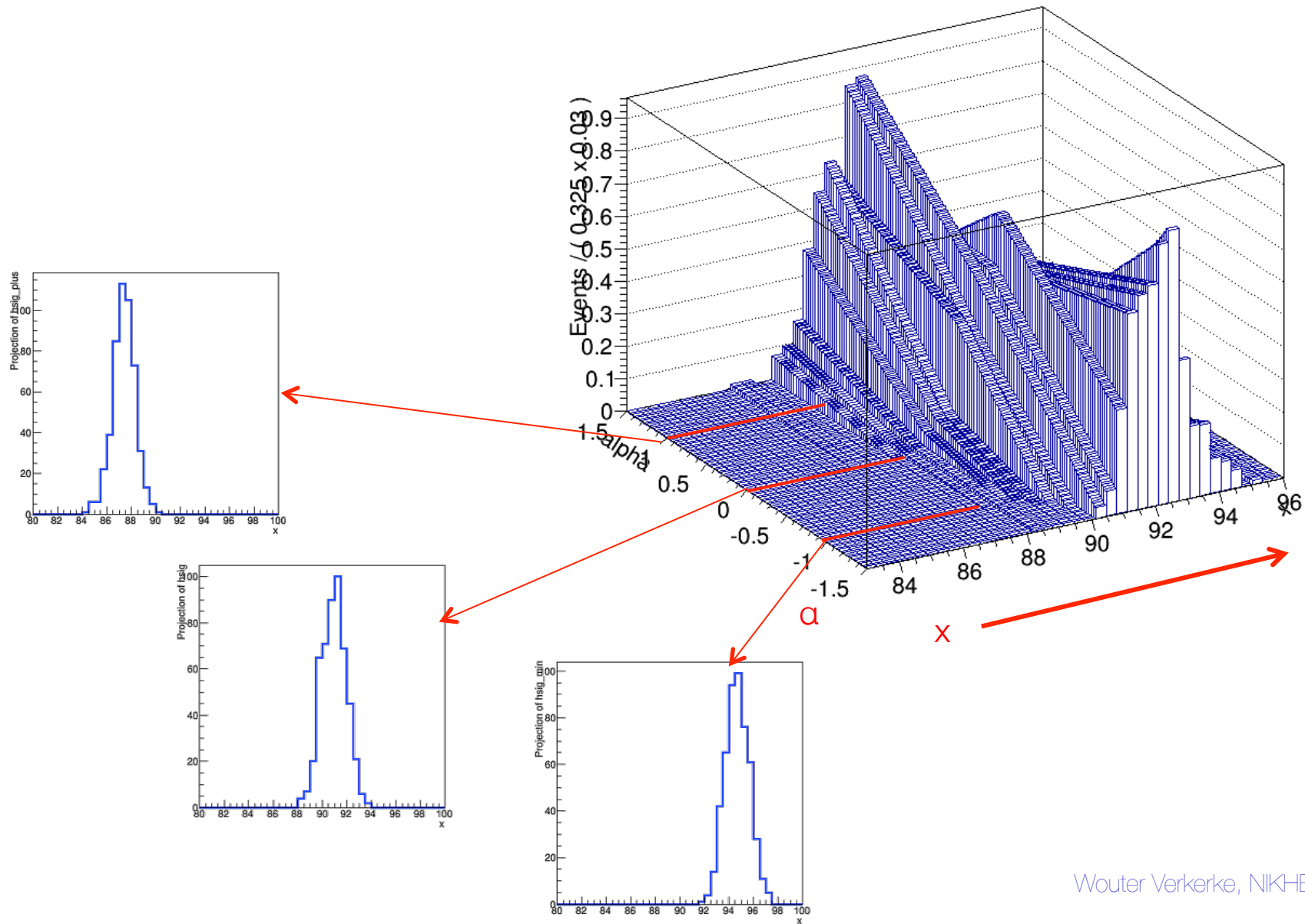


Piecewise linear interpolation

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



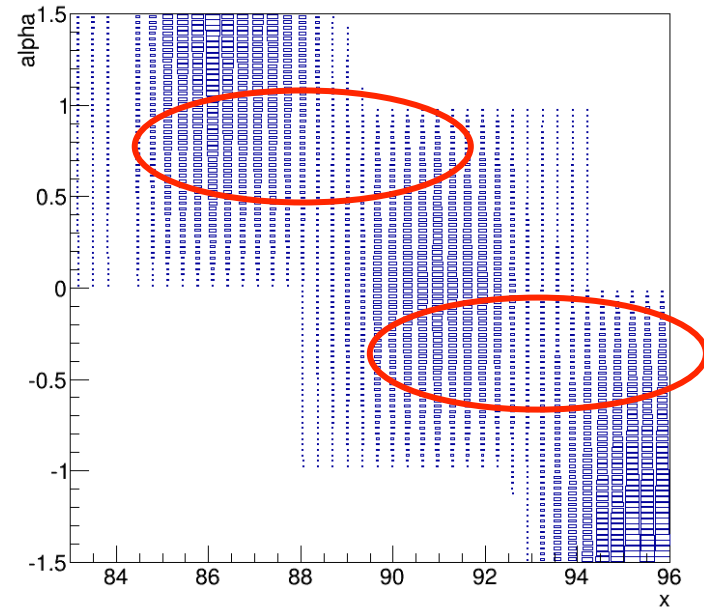
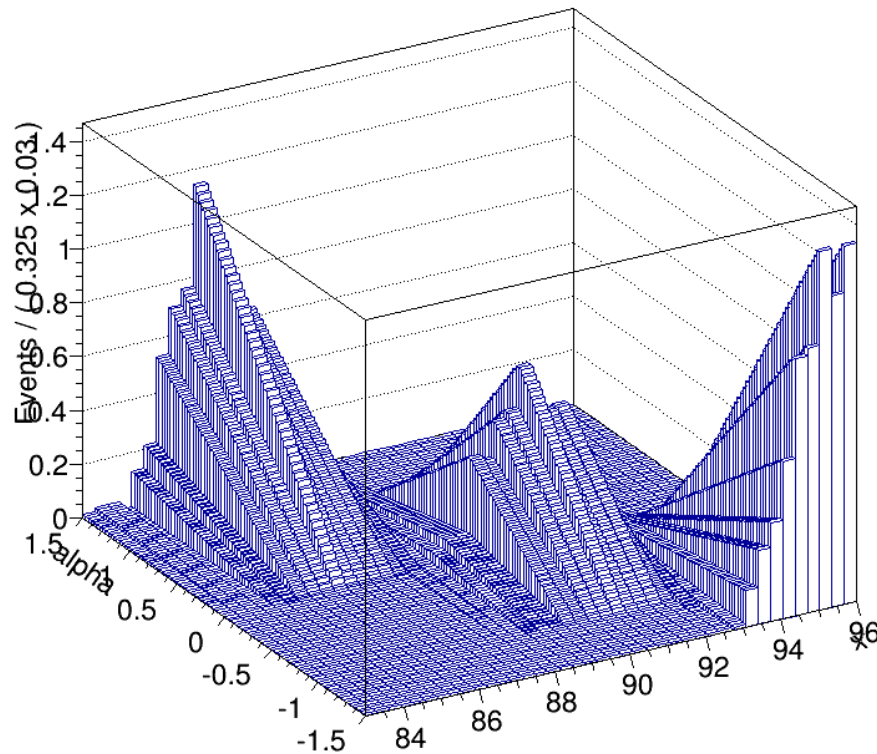
Visualization of bin-by-bin linear interpolation of distribution



Limitations of piece-wise linear interpolation

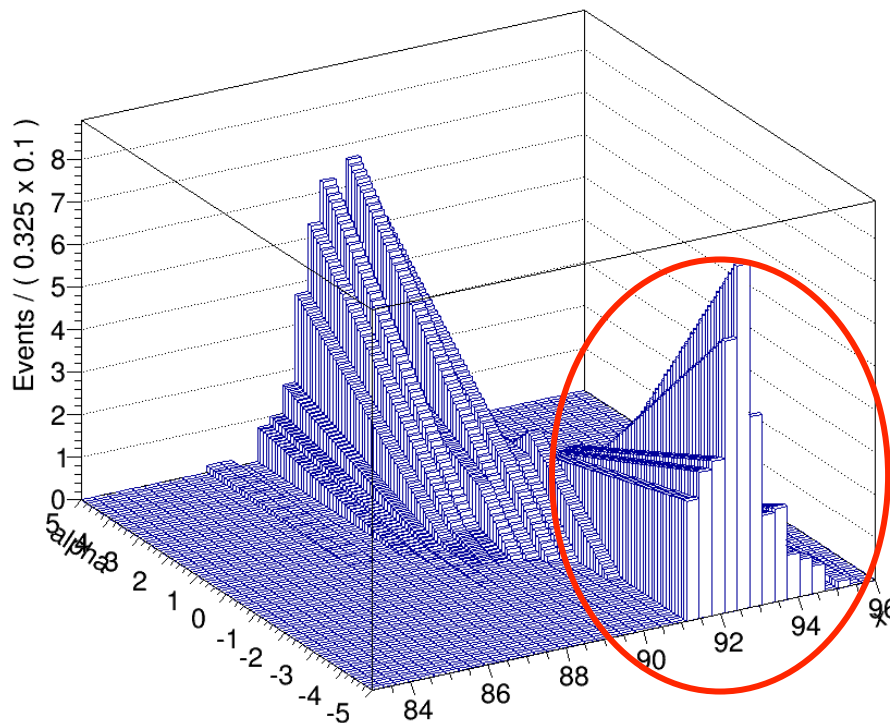
- Bin-by-bin interpolation looks spectacularly easy and simple, but be aware of its limitations
 - Same example, but with larger 'mean shift' between templates

Note double peak structure around $|\alpha|=0.5$



Limitations of piece-wise linear interpolation

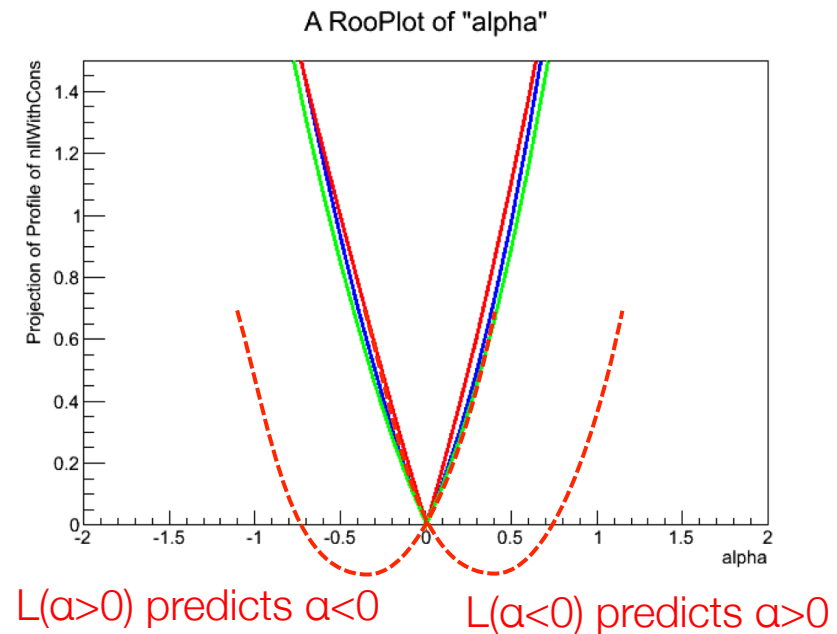
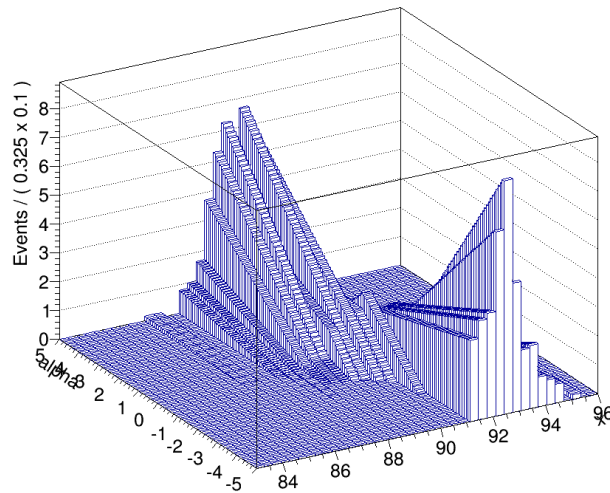
- Also be aware of extrapolation effects
 - Nuisance parameters associated to systematic uncertainties can be pulled well beyond 1σ , especially in high-significance hypothesis testing
 - Original example (with modest shift), but now visualized up to $|a|=5$



MC statistical fluctuations
amplified by extrapolation

Non-linear interpolation options

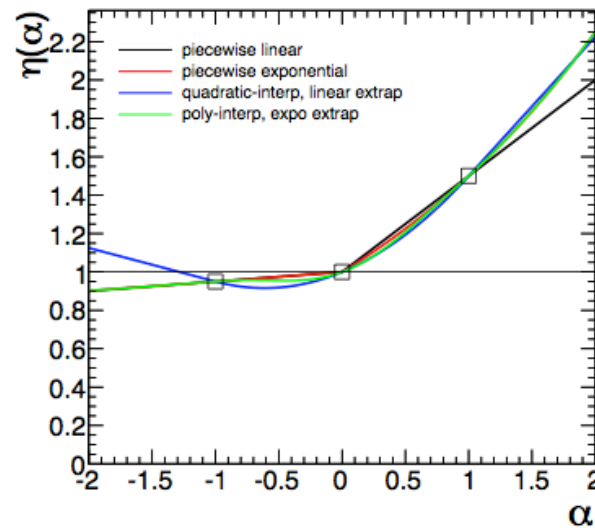
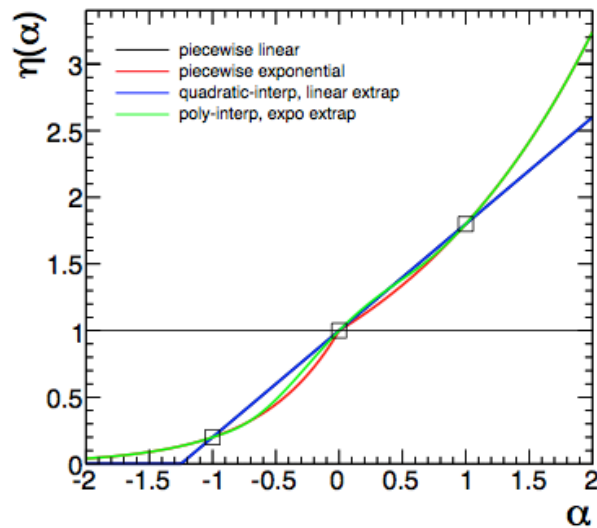
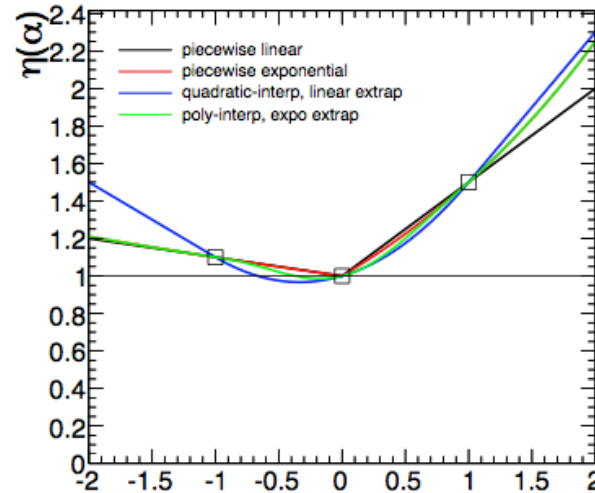
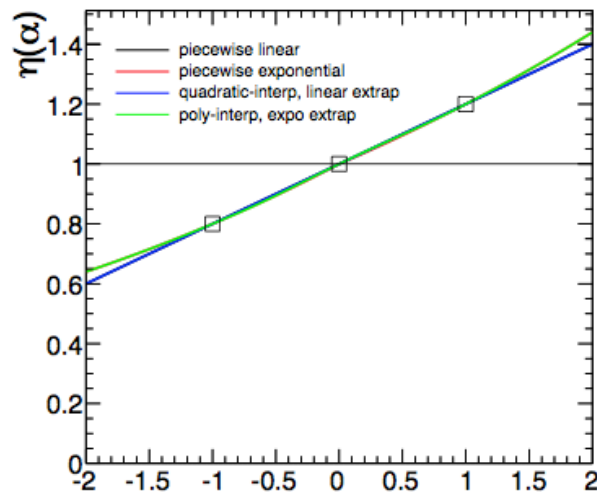
- Piece-wise linear interpolation leads to kink in response functions that may result in pathological likelihood functions



- A variety of other interpolation options exist that improve this
 - Parabolic interpolation/linear extrapolation (but causes shift of minimum)
 - Polynomial interpolation [orders 1,2,4,6]/linear extrapolation (order 1 term allows for asymmetric modeling of templates)

Non-linear interpolation options

- Comparison of common interpolation options

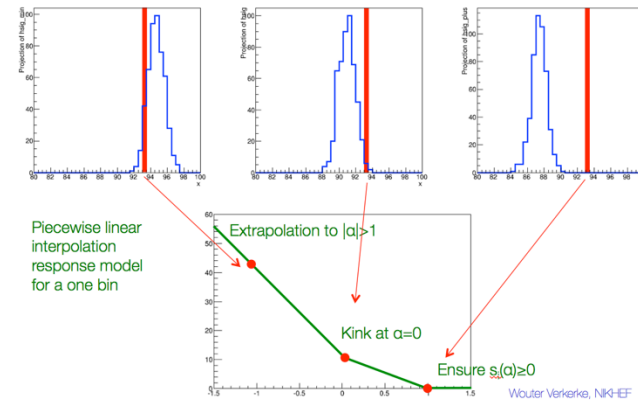


Piece-wise interpolation for >1 nuisance parameter

- 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
- One-parameter interpolation

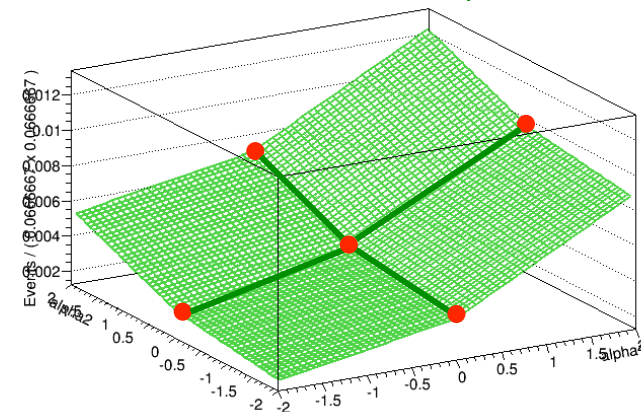
$$s_i(\alpha) = \begin{cases} s_i^0 + \alpha \cdot (s_i^+ - s_i^0) & \forall \alpha > 0 \\ s_i^0 + \alpha \cdot (s_i^0 - s_i^-) & \forall \alpha < 0 \end{cases}$$



- N-parameter interpolation

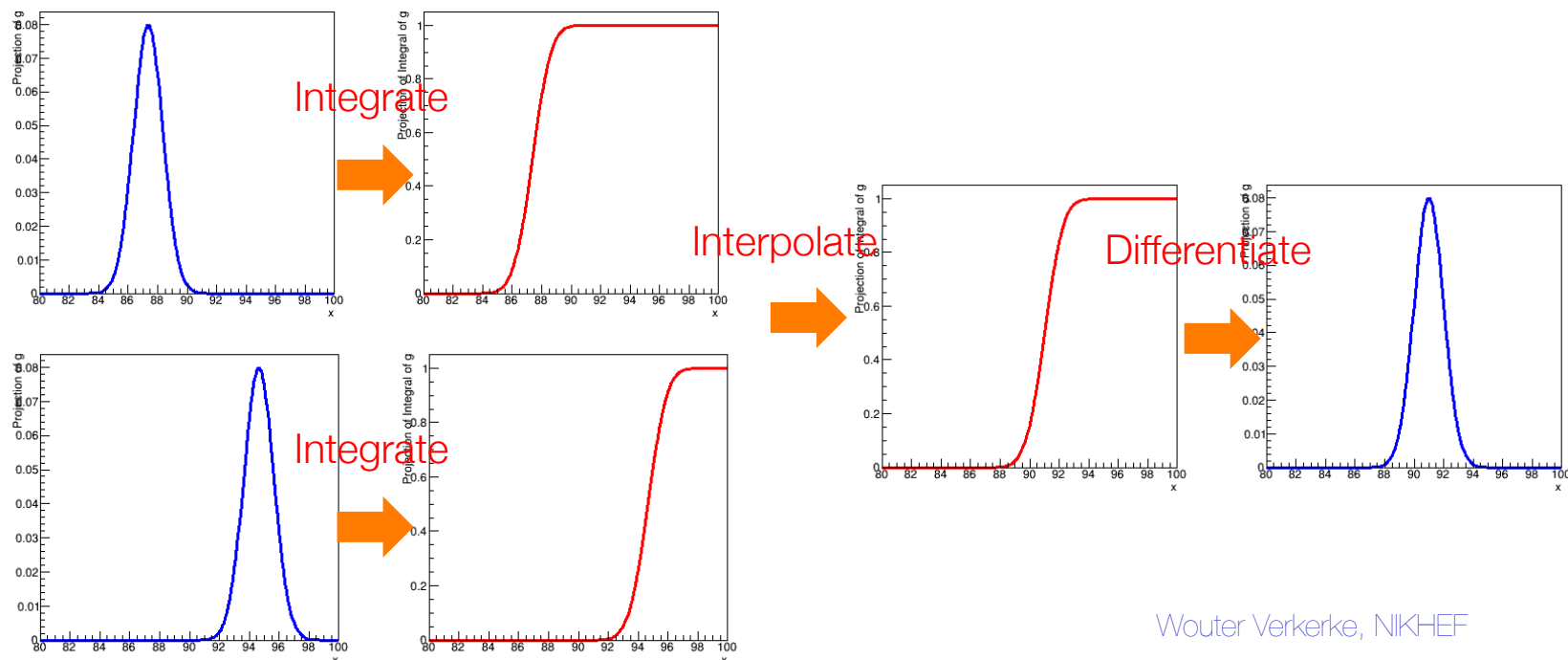
$$s_i(\vec{\alpha}) = \begin{cases} s_i^0 + \sum_j \alpha_j \cdot (s_i^{+,j} - s_i^0) & \forall \alpha > 0 \\ s_i^0 + \sum_j \alpha_j \cdot (s_i^0 - s_i^{-,j}) & \forall \alpha < 0 \end{cases}$$

Visualization of 2D interpolation



Other morphing strategies – ‘horizontal morphing’

- Other template morphing strategies exist that are less prone to unintended side effects
- A ‘horizontal morphing’ strategy was invented by Alex read.
 - Interpolates the cumulative distribution function instead of the distribution
 - Especially suitable for shifting distributions
 - Here shown on a continuous distribution, but also works on histograms
 - Drawback: computationally expensive, algorithm only worked out for 1 NP



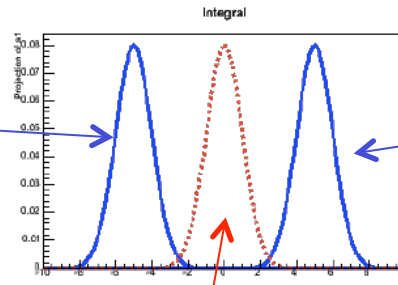
Yet another morphing strategy – ‘Moment morphing’

M. Baak & S. Gadatsch

- Given two template model $f_-(x)$ and $f_+(x)$ the strategy of moment morphing considers first two moment of template models (mean and variance)

$$\mu_- = \int x \cdot f_-(x) dx$$

$$V_- = \int (x - \mu_-)^2 \cdot f_-(x) dx$$



$$\mu_+ = \int x \cdot f_+(x) dx$$

$$V_+ = \int (x - \mu_+)^2 \cdot f_+(x) dx$$

- The goal of moment morphing is to construct an interpolated function that has linearly interpolated moments

$$\begin{aligned} \mu(\alpha) &= \alpha\mu_- + (1 - \alpha)\mu_+ \\ V(\alpha) &= \alpha V_- + (1 - \alpha)V_+ \end{aligned} \quad [1]$$

- It constructs this morphed function as combination of linearly transformed input models

$$f(x, \alpha) \rightarrow \alpha f_-(ax + b) + (1 - \alpha) f_+(cx - d)$$

- Where constants a,b,c,d are chosen such so that $f(x, \alpha)$ satisfies conditions [1]

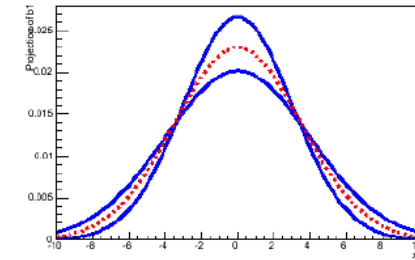
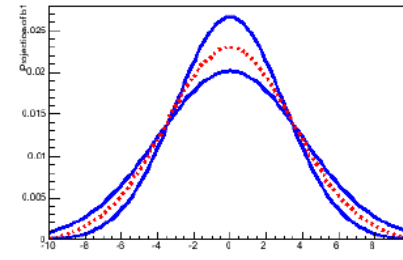
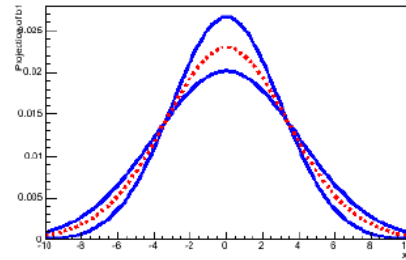
Comparison of morphing algorithms

Vertical
Morphing

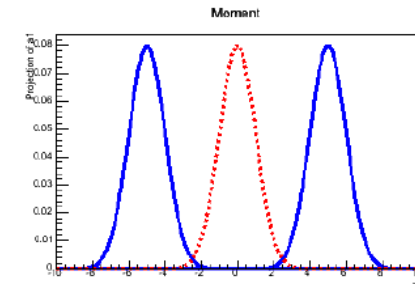
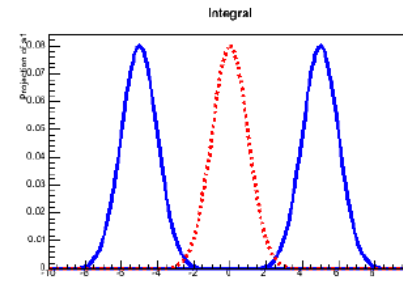
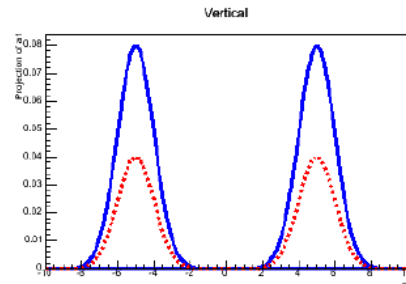
Horizontal
Morphing

Moment
Morphing

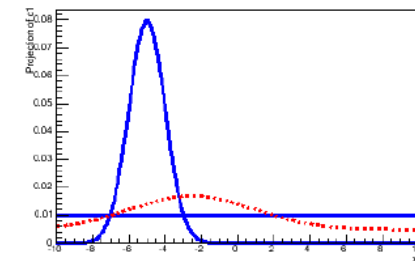
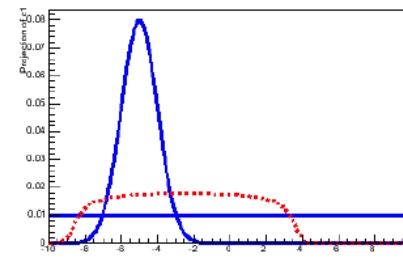
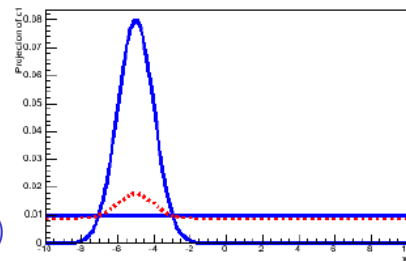
Gaussian
varying
width



Gaussian
varying
mean



Gaussian
to
Uniform
(this is
conceptually ambiguous!)

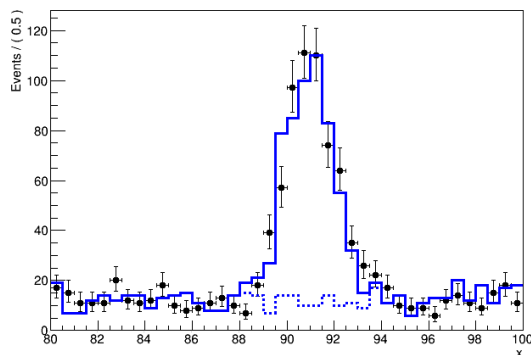


n-dimensional
morphing?

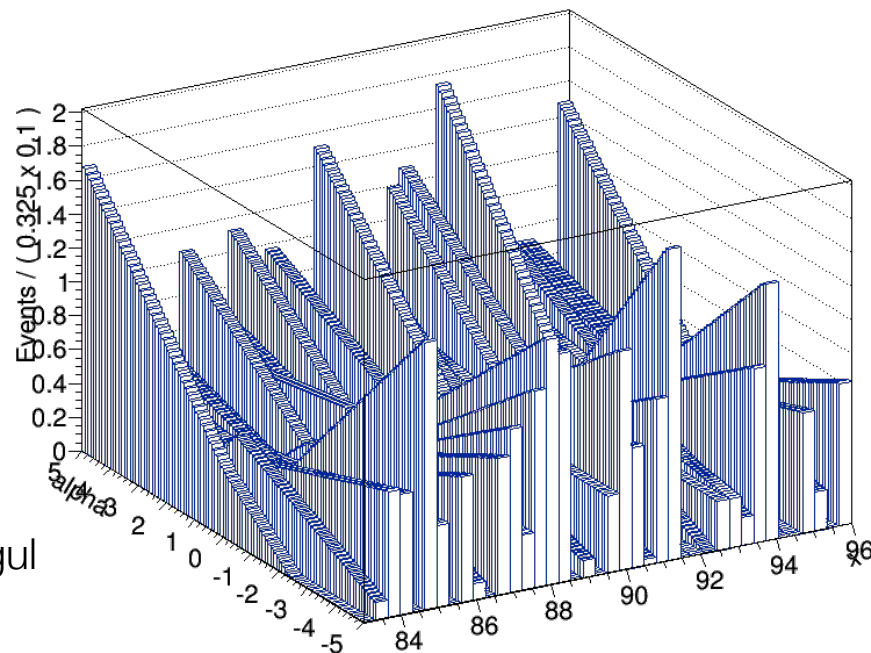


Shape, rate or no systematic?

- 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?

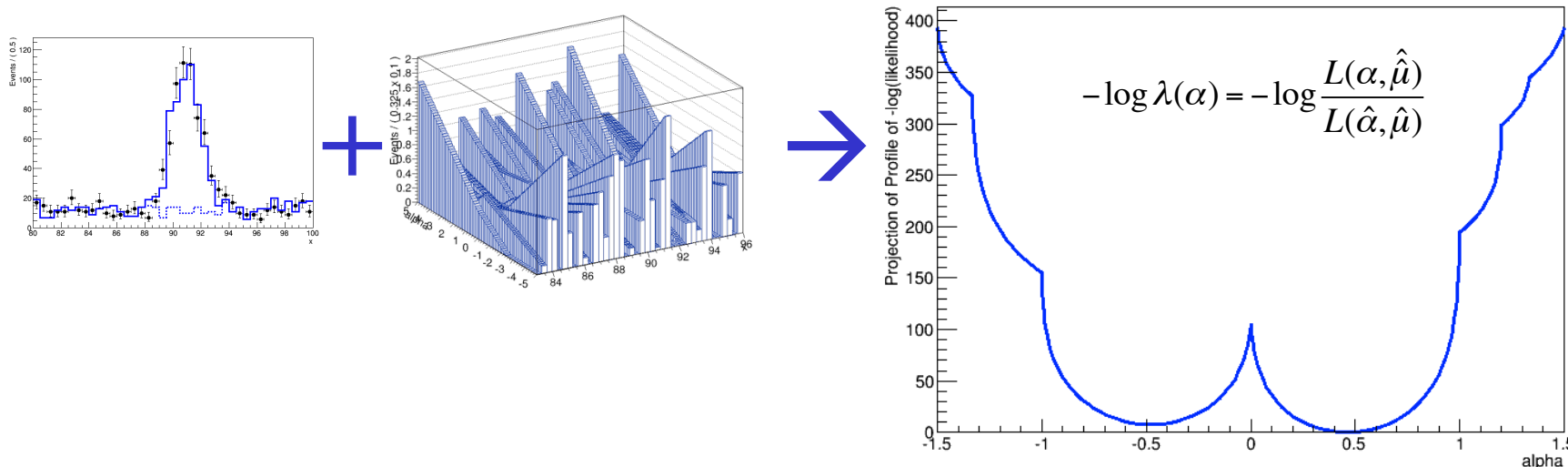


- A χ^2 or KS test between nominal and alternate template can help to decide if a shape uncertainty is meaningful
- Most systematic uncertainties affect both rate and shape, but can make independent decision on modeling rate (which less likely to affect fit stability)



Fit stability due to insignificant shape systematics

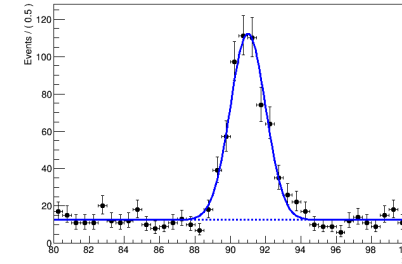
- Shape of profile likelihood in NP \propto clearly raises two points



- 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)

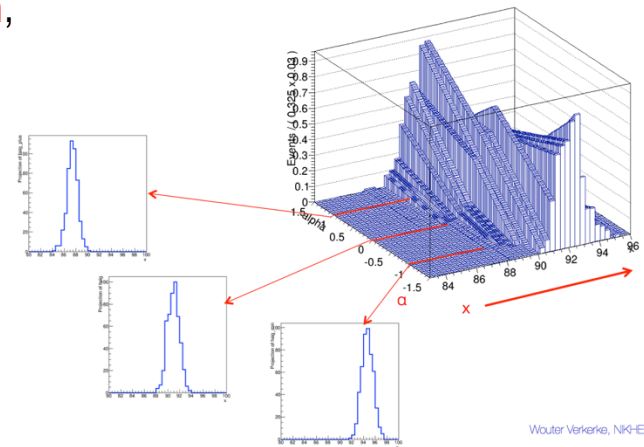
Recap on shape systematics & template morphing

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



$$L(\vec{m}_l | \mu, \alpha_{LES}) = \prod_i [\mu \cdot \text{Gauss}(m_l^{(i)}, 91 \cdot (1 + 2\alpha_{LES}, 1) + (1 - \mu) \cdot \text{Uniform}(m_l^{(i)})] \cdot \text{Gauss}(0 | \alpha_{LES}, 1)$$

- 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

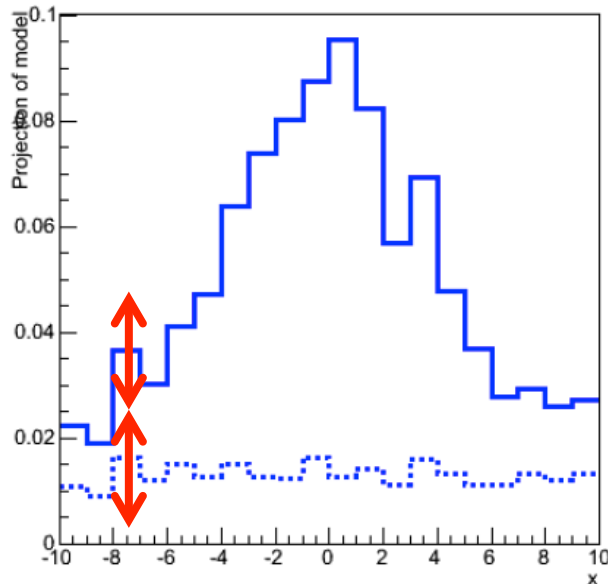


Wouter Verkerke, NIKHEF

Wouter Verkerke, NIKHEF

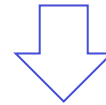
Nuisance parameters for template statistics

- Template morphing implements response function for shape systematic NPs.
- How do we model uncertainty in template due to finite MC statistics?
 - Again same concept: introduce response model in template and add subsidiary measurement to quantify MC statistical uncertainty. Conceptually straightforward for histograms: all bins are independent



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

Binned likelihood
with rigid template



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

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

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

Nuisance parameters for template statistics

- Solution of one NP per template bin conceptually straightforward, but can lead to a very large number of NPs in even the simplest models (e.g. 100 bins \rightarrow 200 NPs for a signal+background model!)
- Is this a problem? And if yes, can we do something about this?
 - It is mostly a problem because it makes numerical minimization of the likelihood tedious (many gradients to calculate, huge covariance matrix)
- Roger Barlow and Christine Beeston realized that for parameter estimation of template yields in ‘sum of template models’ (‘signal and background’) the minimization of each γ parameter can be factorized and solved with a set of n independent equations

Computer Physics Communications 77 (1993) 219–228
North-Holland

Fitting using finite Monte Carlo samples

Roger Barlow and Christine Beeston
Department of Physics, Manchester University, Manchester M13 9PL, UK

Received 27 March 1993

Computer Physics
Communications

The t_i are given by eq. (13). If d_i is zero then t_i is 1; if not then

$$\frac{d_i}{1 - t_i} = f_i = \sum_j p_j A_{ji} = \sum_j \frac{p_j a_{ji}}{1 + p_j t_i}. \quad (15)$$

If these n equations are satisfied, with eq. (14) used to define the A_{ji} , then all the $m \times n$ equations (11) are satisfied.

Merits of the Beeston-Barlow approach

- Beeston-Barlow method implemented in ROOT class TFractionFitter
 - Works great, effectively a minimization prescription, not a likelihood modeling prescription
 - Corresponding likelihood is full likelihood model shown earlier

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

- Effective computational efficiency gain also not completely clear
 - Solution of BB equation takes comparable amount of calculation compared to a numeric gradient calculation in one γ parameter, so do not expect a large gain in minimization phase of MINUT

$$\frac{d_i}{1 - t_i} = f_i = \sum_j p_j A_{ji} = \sum_j \frac{p_j a_{ji}}{1 + p_j t_i}. \quad (15)$$

- Some work on this still ongoing, but ‘plain’ BB is largely unused in LHC analyses now.

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’

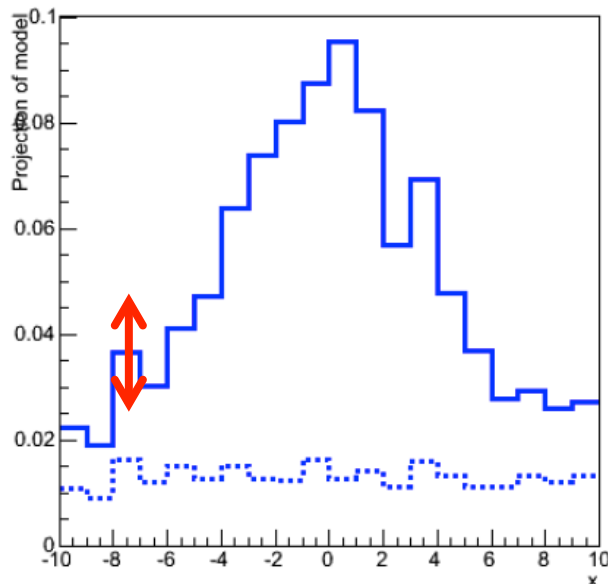
$$L(\vec{N} | \vec{s}, \vec{b}) = \prod_{bins} P(N_i | s_i + b_i) \prod_{bins} P(\tilde{s}_i | s_i) \prod_{bins} P(\tilde{b}_i | b_i)$$

‘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 $\tilde{s} + \tilde{b}$

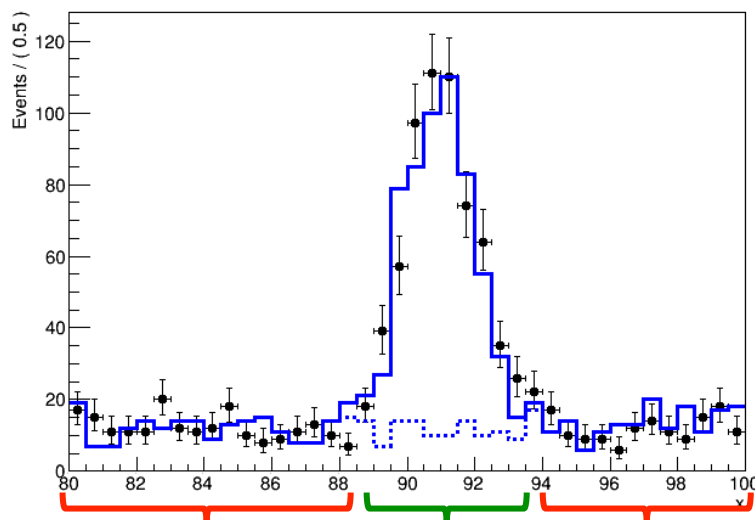


$$L(\vec{N} | \vec{\gamma}) = \prod_{bins} P(N_i | \gamma_i(\tilde{s}_i + \tilde{b}_i)) \prod_{bins} P(\tilde{s}_i + \tilde{b}_i | \gamma_i(\tilde{s}_i + \tilde{b}_i))$$

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

Pruning complexity – MC statistical for selected bins

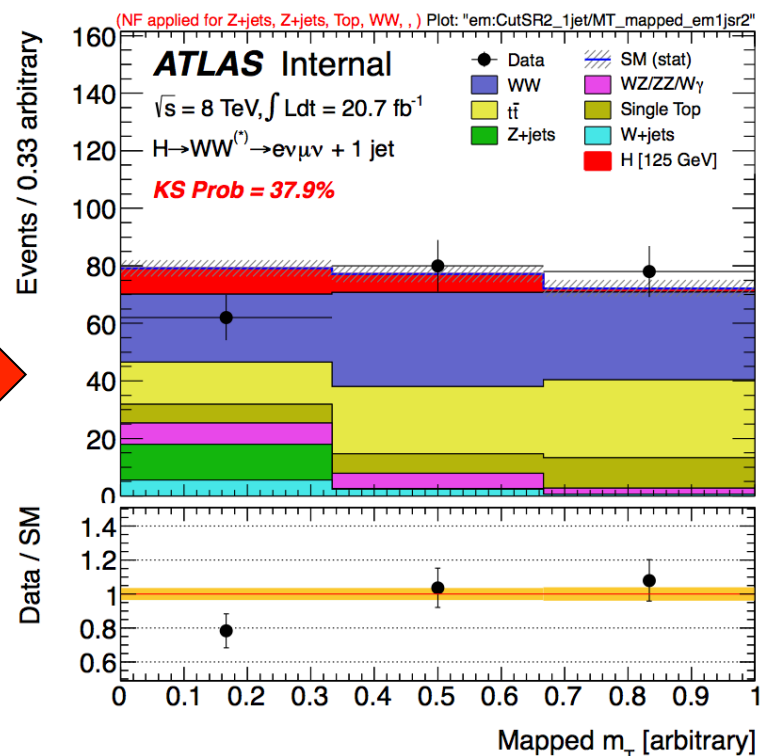
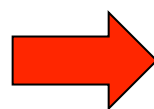
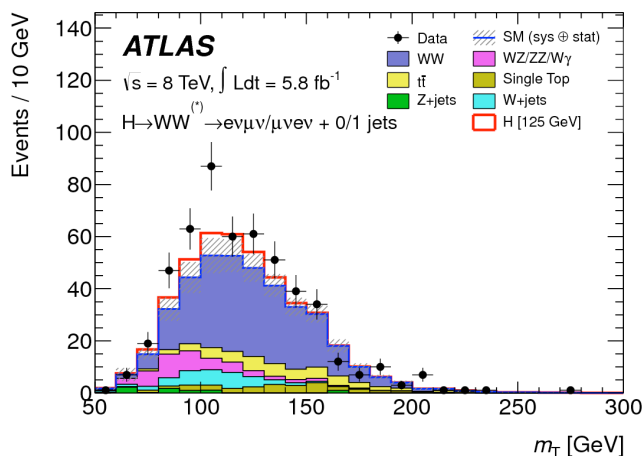
- Can also make decision to model MC statistical uncertainty on a bin-by-bin basis
 - No modeling for high statistics bins
 - Explicit modeling for low-statistics bins



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

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 of this approach in the ATLAS $H \rightarrow WW$ analysis



The interplay between shape systematics and MC systematics

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

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

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

- 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

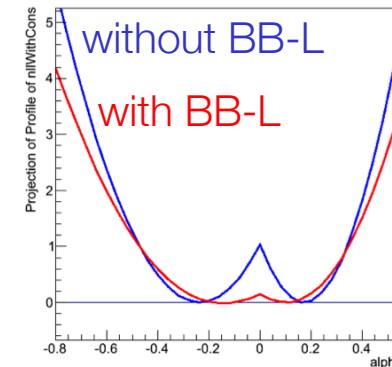
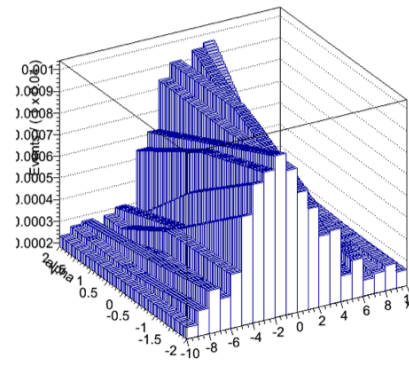
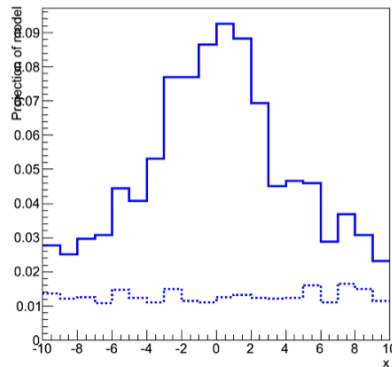
$$s_i(\alpha, \dots) = \begin{cases} s_i^0 + \alpha \cdot (s_i^+ - s_i^0) & \forall \alpha > 0 \\ s_i^0 + \alpha \cdot (s_i^0 - s_i^-) & \forall \alpha < 0 \end{cases}$$

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

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

Summary on template morphing and template statistics

- Template morphing allows to model arbitrary responses of shape systematics in template models
 - Various techniques exist, choose carefully, be wary of MC statistical effects that can dominate systematic effect
- Modeling of MC statistical uncertainties important if $N_{MC} < 10 \times N_{data}$
 - Full Beeston-Barlow likelihood most accurate, but leads to enormous number of Nuisance parameters
 - Beeston-Barlow-lite procedures gives very comparable answers if template statistics are sufficient and results in less NPs
 - Modeling of MC statistical uncertainties improves stability of template morphing algorithms

4 Modeling tools: RooFit, RooStats & HistFactory

Coding probability models and likelihood functions

- Implementation of systematic uncertainties in likelihood models typically leads to very complex probability models
- All statistical techniques discussed in Section 2,4 require numeric minimization of likelihood functions. See problem in three parts
 1. Construction of probability models and likelihood functions (always needed)
 2. Minimization of likelihood functions (for parameter estimation, variance estimate, likelihood-ratio intervals)
 3. Construction of test statistics and calculation of their distributions, construction of Neyman constructions on test statistics (p-values, confidence intervals) and calculation (MC(MC)) integrals over Likelihood (Bayesian credible intervals, Bayes factors)
- For step 2 (minimization) the HEP industry standard is MINUIT
- For steps 1, 3 good tools have been developed in the past years, will discuss these now.

RooFit, RooStats and HistFactory

Will cover RooFit and HistFactory in a bit more detail since they relate to model building – the key topic of this course

RooFit

Language for building probability models

Comprises datasets, likelihoods, minimization, toy data generation, visualization and persistence

W. Verkerke & D. Kirkby
(exists since 1999)

HistFactory

Language to simplify construction of RooFit models of a particular type: binned likelihood template (morphing) models

K. Cranmer, A. Shibata, G. Lewis, L. Moneta, W. Verkerke
(exists since 2010)

Will briefly sketch workings of RooStats

RooStats

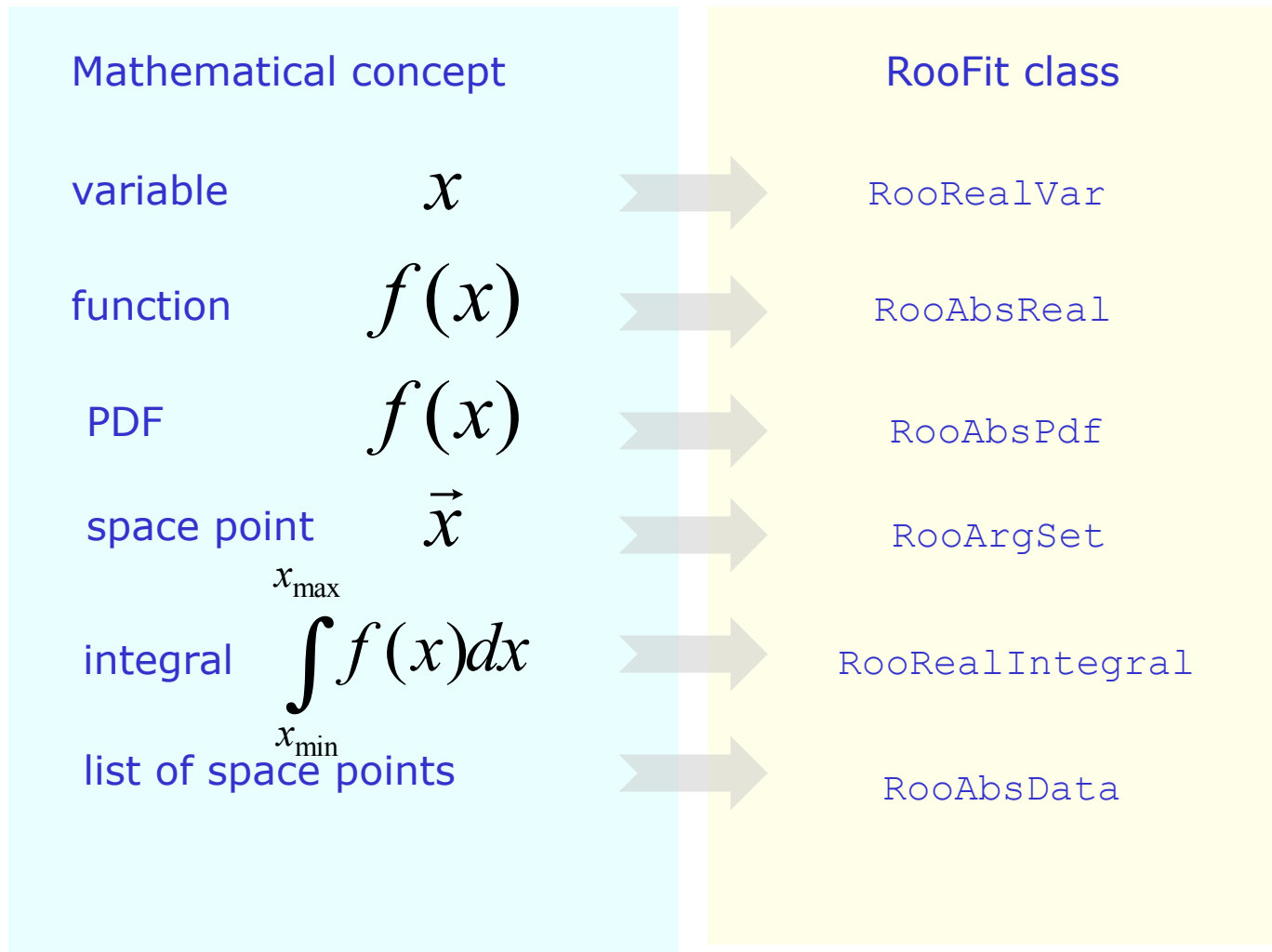
Suite of statistical tests operating on RooFit probability models

K. Cranmer, G. Schott, L. Moneta, W. Verkerke
(exists since 2008)

Wouter Verkerke, NIKHEF

RooFit core design philosophy

- Mathematical objects are represented as C++ objects



RooFit core design philosophy - Workspace

- The workspace serves a container class for all objects created

Math	$\text{Gauss}(x, \mu, \sigma)$
RooFit diagram	<pre>graph BT; x[RooRealVar x] --> g[RooGaussian g]; y[RooRealVar y] <--> g; z[RooRealVar z] --> g;</pre>
RooFit code	<pre>RooRealVar x("x","x",-10,10) ; RooRealVar m("m","y",0,-10,10) ; RooRealVar s("s","z",3,0.1,10) ; RooGaussian g("g","g",x,m,s) ;</pre>

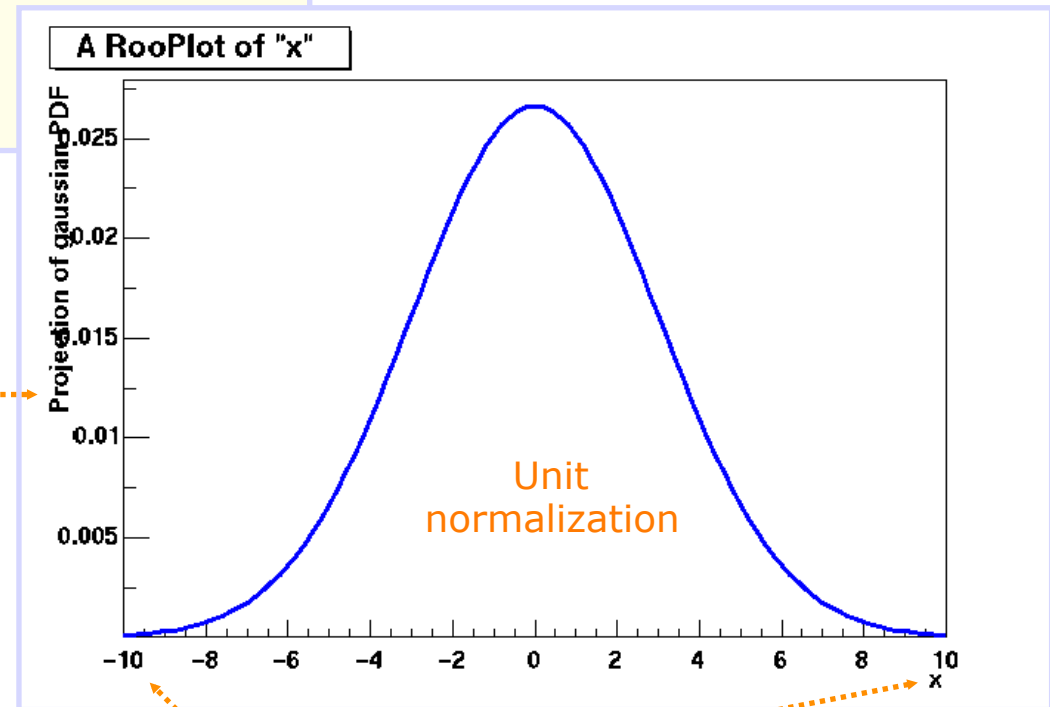
Basics – Creating and plotting a Gaussian p.d.f

Setup gaussian PDF and plot

```
// Create an empty plot frame  
RooPlot* xframe = w::x.frame() ;  
  
// Plot model on frame  
model.plotOn(xframe) ;  
  
// Draw frame on canvas  
xframe->Draw() ;
```

Axis label from gauss title

A RooPlot is an empty frame
capable of holding anything
plotted versus it variable



Plot range taken from limits of x

Basics – Generating toy MC events

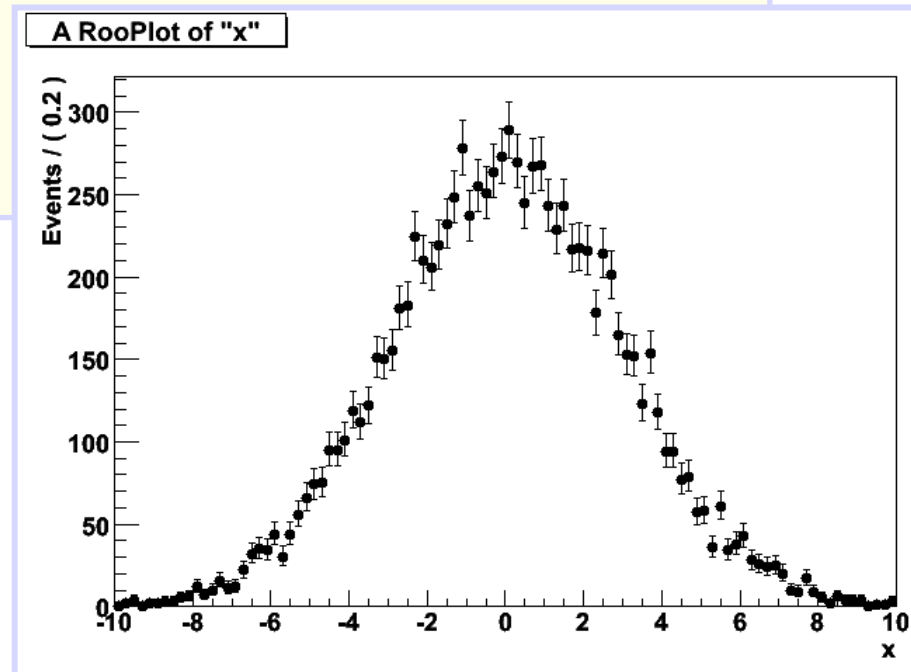
Generate 10000 events from Gaussian p.d.f and show distribution

```
// Generate an unbinned toy MC set
RooDataSet* data = w::gauss.generate(w::x,10000) ;

// Generate an binned toy MC set
RooDataHist* data = w::gauss.generateBinned(w::x,10000) ;

// Plot PDF
RooPlot* xframe = w::x.frame()
data->plotOn(xframe) ;
xframe->Draw() ;
```

Can generate both binned and unbinned datasets

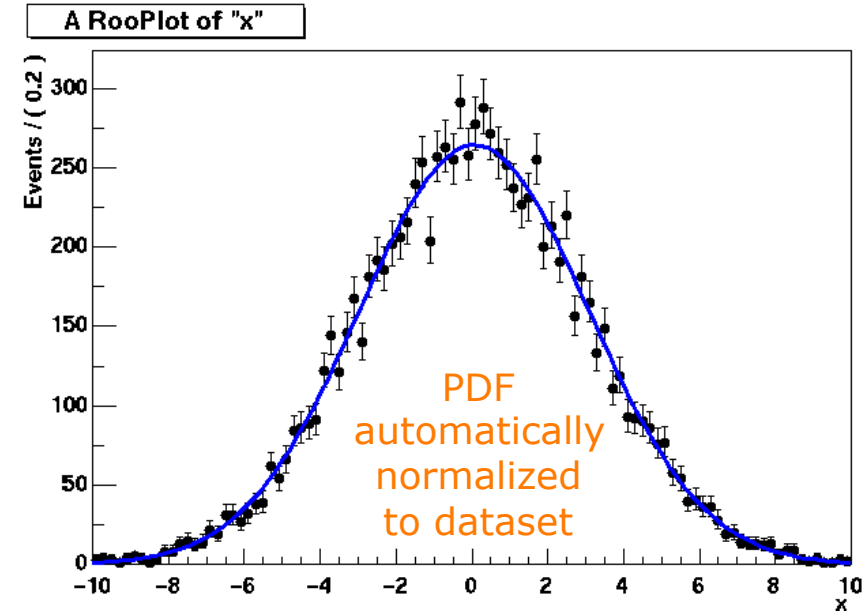


Basics – ML fit of p.d.f to *unbinned* data

```
// ML fit of gauss to data
w::gauss.fitTo(*data) ;
(MINUIT printout omitted)

// Parameters if gauss now
// reflect fitted values
w::mean.Print()
RooRealVar::mean = 0.0172335 +/- 0.0299542
w::sigma.Print()
RooRealVar::sigma = 2.98094 +/- 0.0217306

// Plot fitted PDF and toy data overlaid
RooPlot* xframe = w::x.frame() ;
data->plotOn(xframe) ;
w::gauss.plotOn(xframe) ;
```



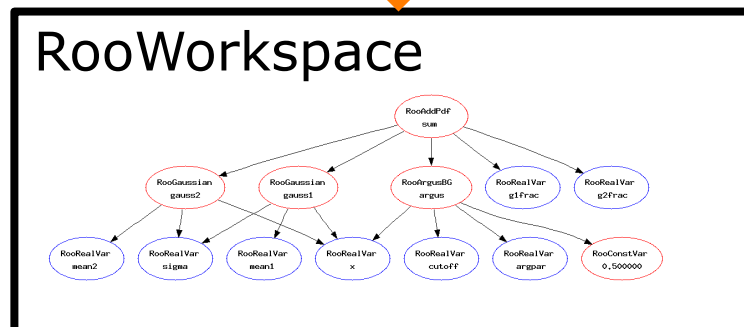
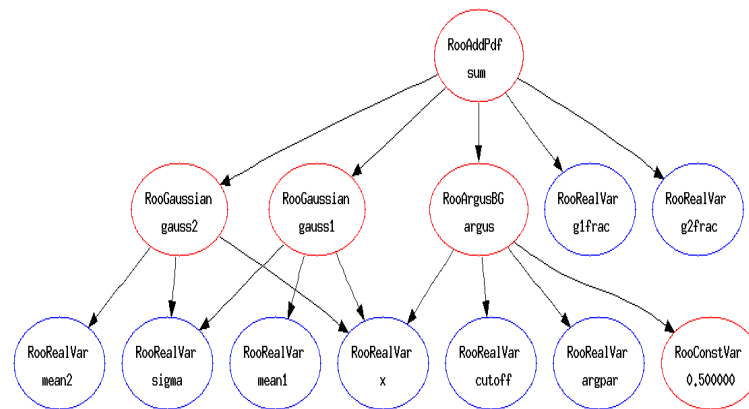
RooFit core design philosophy - Workspace

- The workspace serves a container class for all objects created

Math	Gauss(x, μ, σ)
	RooWorkspace
RooFit diagram	<pre> graph TD g[RooGaussian g] <--> x[RooRealVar x] g <--> y[RooRealVar y] g <--> z[RooRealVar z] </pre>
RooFit code	<pre> RooRealVar x("x","x",-10,10) ; RooRealVar m("m","y",0,-10,10) ; RooRealVar s("s","z",3,0.1,10) ; RooGaussian g("g","g",x,m,s) ; RooWorkspace w("w") ; w.import(g) ; </pre>

The workspace

- The workspace concept has revolutionized the way people share and combine analysis
 - **Completely** factorizes process of building and using likelihood functions
 - You can give somebody an analytical likelihood of a (potentially very complex) physics analysis in a way to the easy-to-use, provides introspection, and is easy to modify.

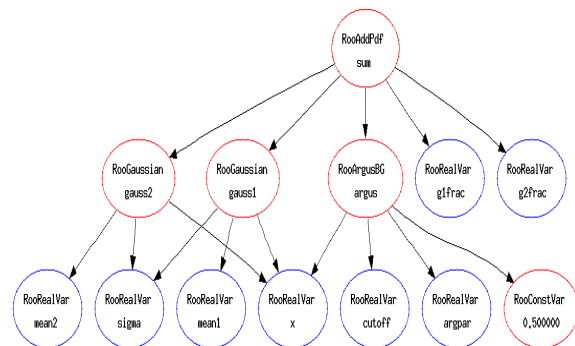
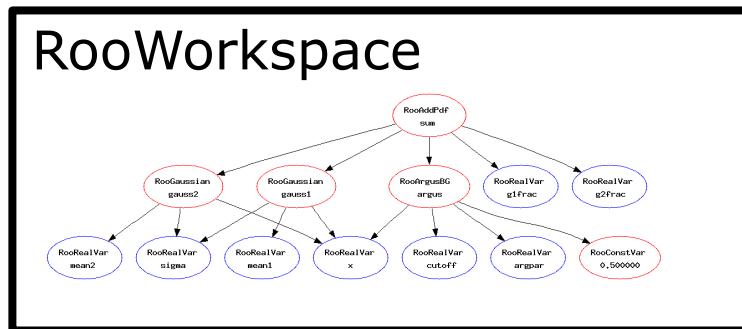


```
RooWorkspace w("w") ;  
w.import(sum) ;  
w.writeToFile("model.root") ;
```

model.root



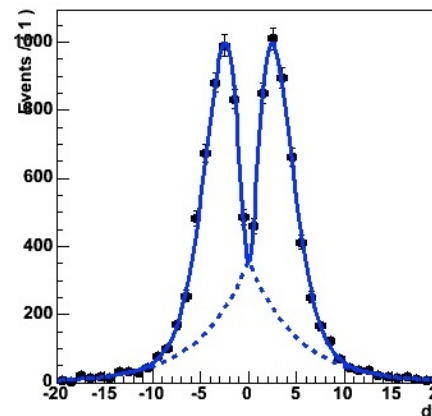
Using a workspace



```
// Resurrect model and data
TFile f("model.root") ;
RooWorkspace* w = f.Get("w") ;
RooAbsPdf* model = w->pdf("sum") ;
RooAbsData* data = w->data("xxx") ;
```

```
// Use model and data
model->fitTo(*data) ;
```

```
RooPlot* frame =
    w->var("dt")->frame() ;
data->plotOn(frame) ;
model->plotOn(frame) ;
```



outer Verkerke, NIKHEF
 inner Verkerke, NIKHEF

Accessing workspace contents

- Looking into a workspace

```
w.Print() ;

variables
-----
(mean,sigma,x)

p.d.f.s
-----
RooGaussian::f[ x=x mean=mean sigma=sigma ] = 0.249352
```

- Access components two ways

```
// 1 - Standard accessor method
RooAbsPdf* pdf = w->pdf("f") ;

// 2 - Import contents into C++ namespace in interpreter
w.exportToCint("w") ;
RooPlot* frame = w::x.frame() ;
w::f.plotOn(frame) ;
// strongly typed: w::f is 'RooGaussian&'
```


RooFit core design philosophy - Workspace

- The workspace serves a container class for all objects created

Math	Gauss(x, μ, σ)
	RooWorkspace
RooFit diagram	<pre> graph TD g[RooGaussian g] --> x[RooRealVar x] g --> y[RooRealVar y] g --> z[RooRealVar z] </pre>
RooFit code	<pre> RooRealVar x("x","x",-10,10) ; RooRealVar m("m","y",0,-10,10) ; RooRealVar s("s","z",3,0.1,10) ; RooGaussian g("g","g",x,m,s) ; RooWorkspace w("w") ; w.import(g) ; </pre>

Factory and Workspace

- *One C++ object per math symbol* provides ultimate level of control over each objects functionality, but results in lengthy user code for even simple macros
- Solution: add factory that auto-generates objects from a math-like language. Accessed through `factory()` method of workspace
- Example: reduce construction of Gaussian pdf and its parameters from 4 to 1 line of code

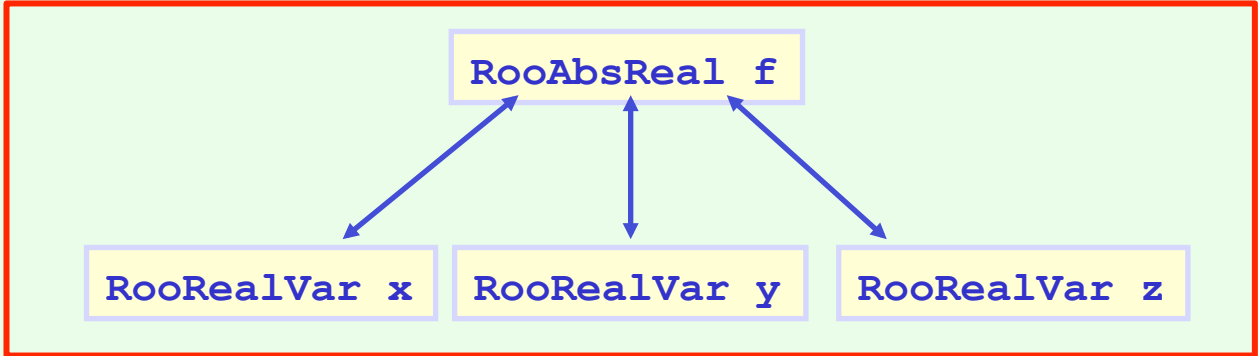
```
RooRealVar x("x","x",-10,10) ;  
RooRealVar mean("mean","mean",5) ;  
RooRealVar sigma("sigma","sigma",3) ;  
RooGaussian f("f","f",x,mean,sigma) ;  
w.import(f) ;
```



```
w.factory("Gaussian::f(x[-10,10],mean[5],sigma[3])") ;
```

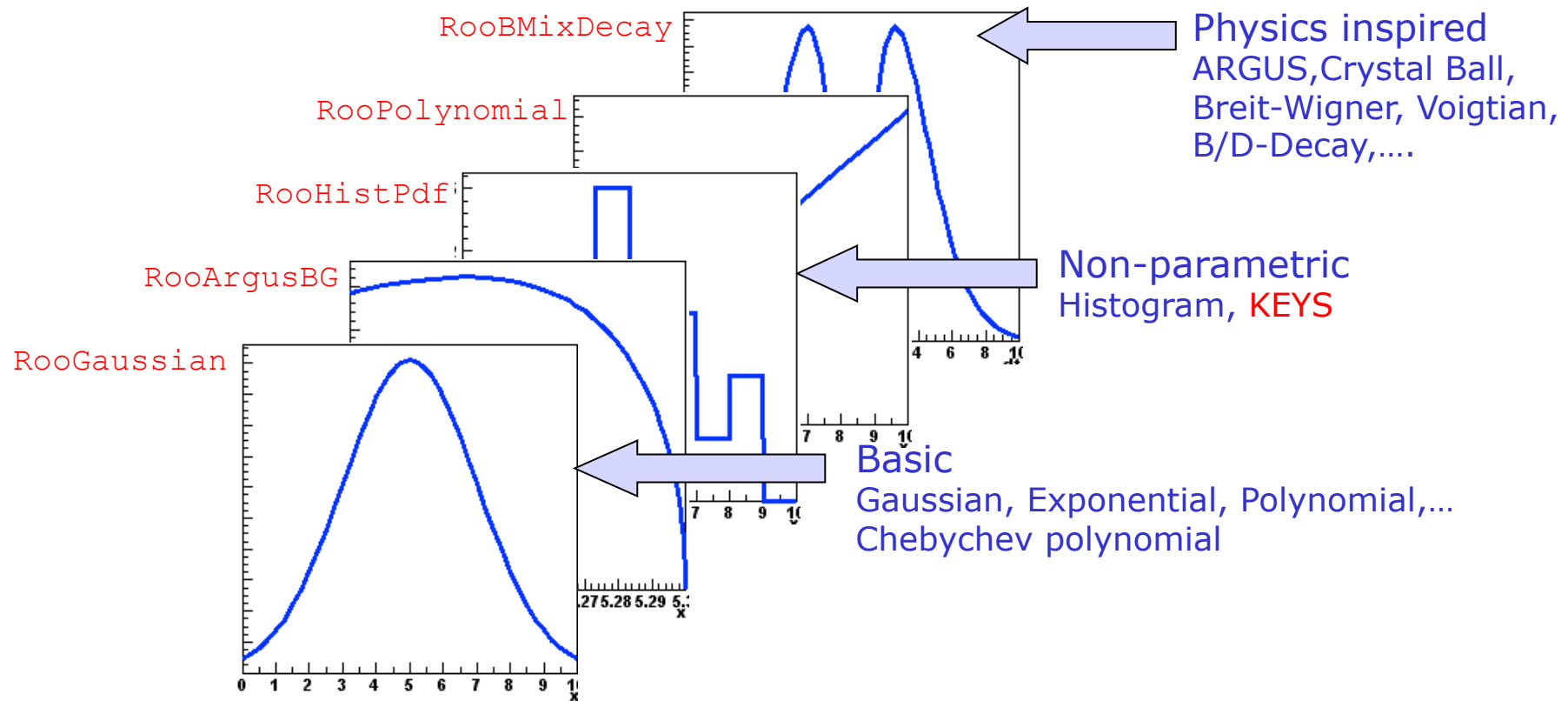

Populating a workspace the easy way – “the factory”

- The **factory** allows to fill a workspace with pdfs and variables using a simplified scripting language

Math	$\text{Gauss}(x, \mu, \sigma)$
	RooWorkspace
RooFit diagram	 <pre>graph TD; f[RooAbsReal f] <--> x[RooRealVar x]; f <--> y[RooRealVar y]; f <--> z[RooRealVar z];</pre>
RooFit code	<pre>RooWorkspace w("w") ; w.factory("RooGaussian::g(x[-10,10],m[-10,10],z[3,0.1,10])") ;</pre>

Model building – (Re)using standard components

- RooFit provides a collection of compiled standard PDF classes



Easy to extend the library: each p.d.f. is a separate C++ class

Model building – (Re)using standard components

- List of most frequently used pdfs and their factory spec

Gaussian

Gaussian::g(x,mean,sigma)

Breit-Wigner

BreitWigner::bw(x,mean,gamma)

Landau

Landau::l(x,mean,sigma)

Exponential

Exponential::e(x,alpha)

Polynomial

Polynomial::p(x,{a0,a1,a2})

Chebyshev

Chebyshev::p(x,{a0,a1,a2})

Kernel Estimation

KeysPdf::k(x,dataSet)

Poisson

Poisson::p(x,mu)

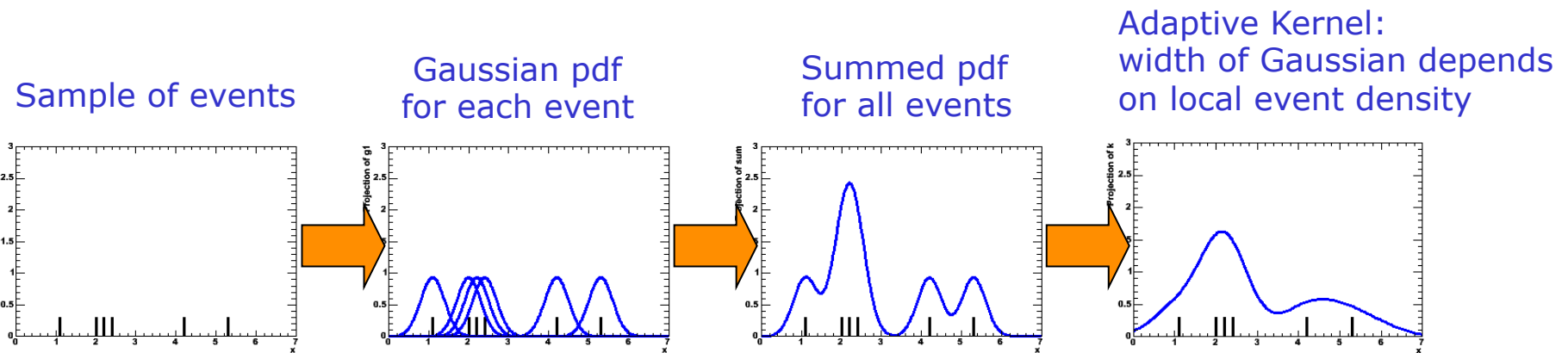
Voigtian

Voigtian::v(x,mean,gamma,sigma)

(=BW⊗G)

The power of pdf as building blocks – Advanced algorithms

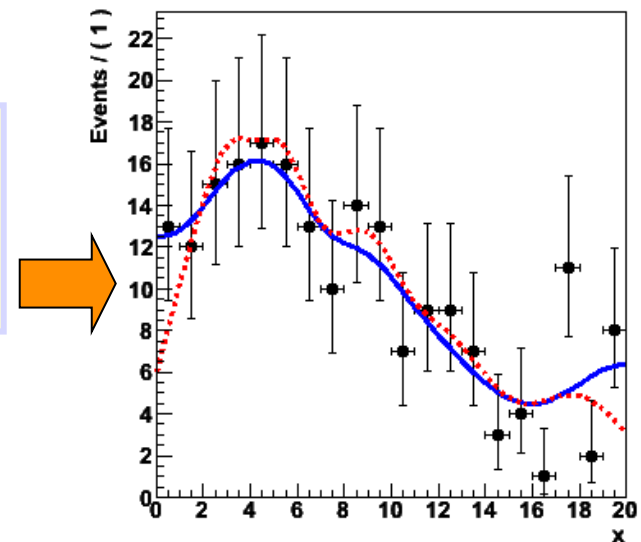
- Example: a ‘kernel estimation probability model’
 - Construct smooth pdf from unbinned data, using kernel estimation technique



- Example

```
w.import(myData,Rename("myData")) ;
w.factory("KeysPdf::k(x,myData)") ;
```

- Also available for n-D data



The power of pdf as building blocks – adaptability

- RooFit pdf classes do not require their parameter arguments to be variables, one can plug in functions as well
- Allows trivial customization, extension of probability models

class RooGaussian

also class RooGaussian!

$$Gauss(x | \mu, \sigma)$$

$$Gauss(x | \underbrace{\mu \cdot (1 + 2\alpha)}_{\text{response function}}, \sigma)$$

Introduce a response function for a systematic uncertainty

```
// Original Gaussian
w.factory("Gaussian::g1(x[80,100],m[91,80,100],s[1])")

// Gaussian with response model in mean
w.factory("expr::m_response(\"m*(1+2alpha)\",m,alpha[-5,5])") ;
w.factory("Gaussian::g1(x,m_response,s[1])")
```

NB: “expr” operates builds an interpreted function expression on the fly

The power of building blocks – operator expressions

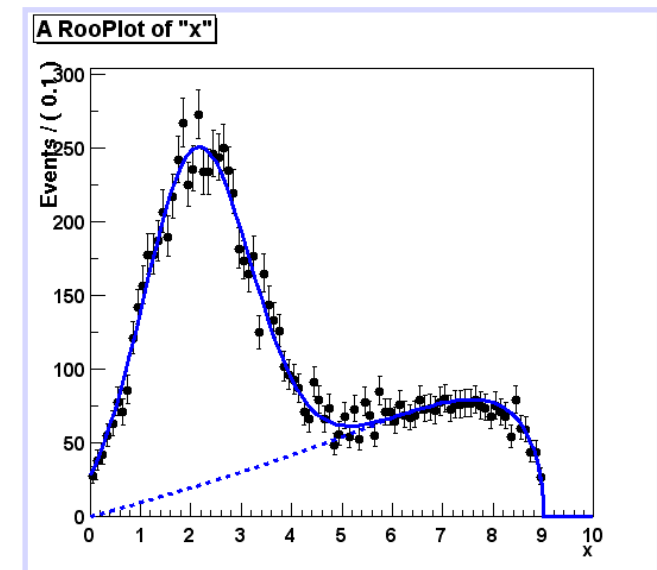
- Create a SUM expression to represent a sum of probability models

```
w.factory("Gaussian::gauss1(x[0,10],mean1[2],sigma[1])" );
w.factory("Gaussian::gauss2(x,mean2[3],sigma)" );
w.factory("ArgusBG::argus(x,k[-1],9.0)" );

w.factory("SUM::sum(g1frac[0.5]*gauss1, g2frac[0.1]*gauss2, argus)")
```

- In composite model visualization components can be accessed by name

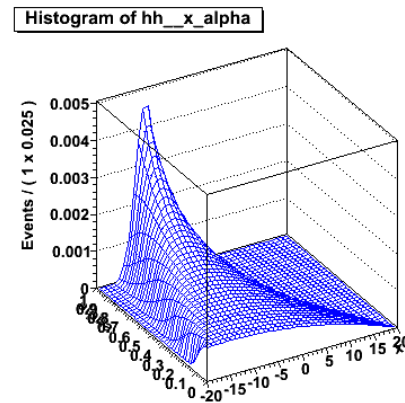
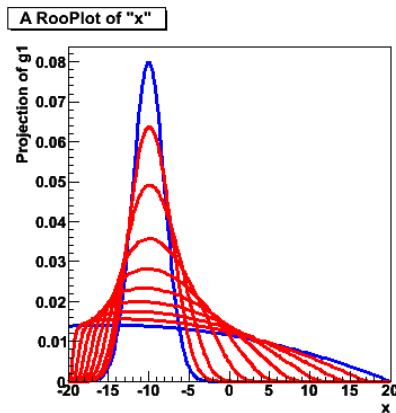
```
// Plot only argus components
w::sum.plotOn(frame,Components("argus"),
              LineStyle(kDashed)) ;
```



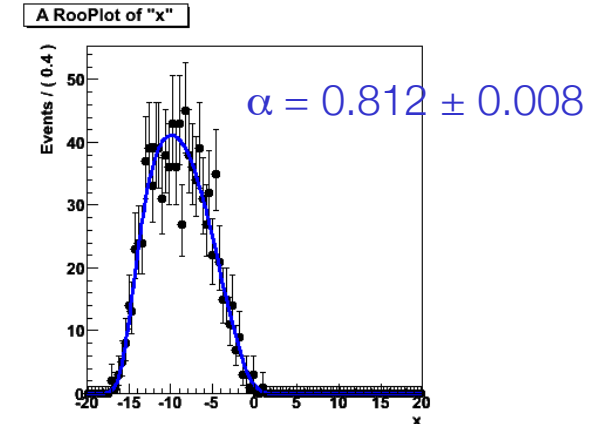
Powerful operators – Morphing interpolation

- Special operator pdfs can interpolate existing pdf shapes
 - Ex: interpolation between Gaussian and Polynomial

```
w.factory("Gaussian::g(x[-20,20],-10,2)") ;
w.factory("Polynomial::p(x,{-0.03,-0.001})") ;
w.factory("IntegralMorph::gp(g,p,x,alpha[0,1])") ;
```



Fit to data



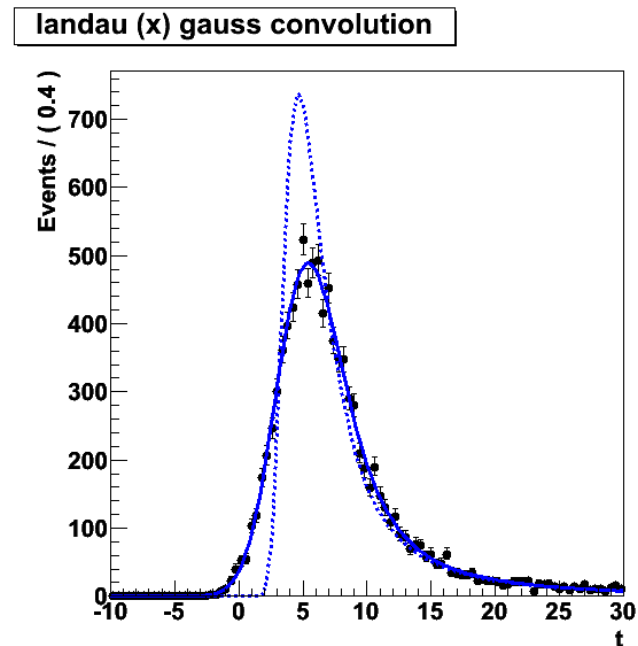
- Three morphing operator classes available
 - *IntegralMorph* (Alex Read).
 - *MomentMorph* (Max Baak).
 - *PiecewiseInterpolation* (via HistFactory)

Powerful operators – Fourier convolution

- Convolve any two arbitrary pdfs with a 1-line expression

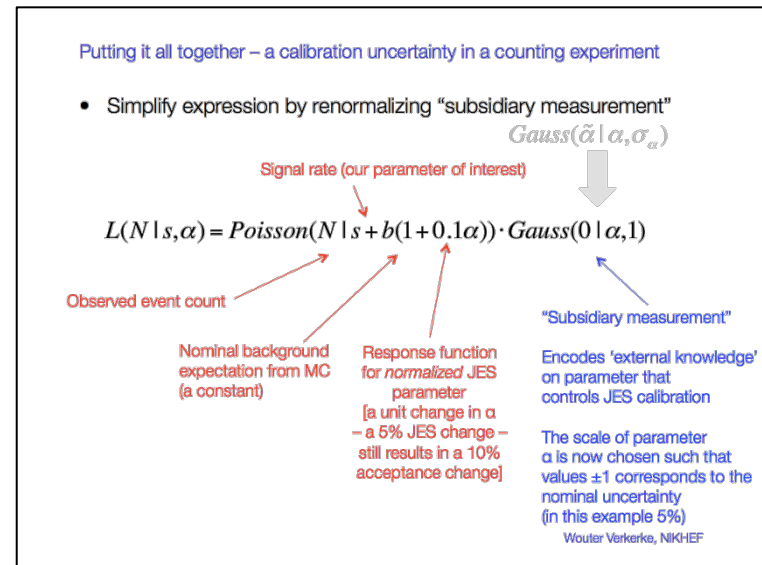
```
w.factory("Landau::L(x[-10,30],5,1)") :  
w.factory("Gaussian::G(x,0,2)") ;  
  
w::x.setBins("cache",10000) ; // FFT sampling density  
w.factory("FCONV::LGf(x,L,G)") ; // FFT convolution
```

- Exploits power of FFTW package available via ROOT
 - Hand-tuned assembler code for time-critical parts
 - Amazingly fast: unbinned ML fit to 10.000 events take ~5 seconds!



Example 1: counting expt

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



$$L(N | s, \alpha) = Poisson(N | s + \underbrace{b(1 + 0.1\alpha)}_{\text{Signal rate}}) \cdot \underbrace{Gauss(0 | \alpha, 1)}_{\text{Subsidiary measurement}}$$

```
// Subsidiary measurement of alpha
w.factory("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.

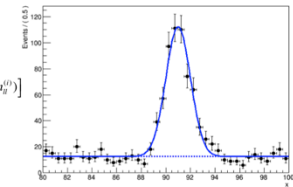
- Will now demonstrate how to code complete example of the unbinned profile likelihood of Section 5:

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

$$L(\vec{m}_l | \mu) = \prod_i [\mu \cdot \text{Gauss}(m_l^{(i)}, 91, 1) + (1 - \mu) \cdot \text{Uniform}(m_l^{(i)})]$$



- 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_l has been determined to a 2% shift for 1% LES change

$$L(\vec{m}_l | \mu, \alpha_{LES}) = \prod_i [\mu \cdot \text{Gauss}(m_l^{(i)}, 91 \cdot \underbrace{(1 + 2\alpha_{LES})}_{\text{Response function}}, 1) + (1 - \mu) \cdot \text{Uniform}(m_l^{(i)})] \cdot \underbrace{\text{Gauss}(0 | \alpha_{LES}, 1)}_{\text{Subsidiary measurement}}$$

Wouter Verkerke, Niko EF

$$L(\vec{m}_l | \mu, \alpha_{LES}) = \prod_i [\mu \cdot \text{Gauss}(m_l^{(i)}, 91 \cdot \underbrace{(1 + 2\alpha_{LES})}_{\text{Response function}}, 1) + (1 - \mu) \cdot \text{Uniform}(m_l^{(i)})] \cdot \underbrace{\text{Gauss}(0 | \alpha_{LES}, 1)}_{\text{Subsidiary measurement}}$$

```
// 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])")

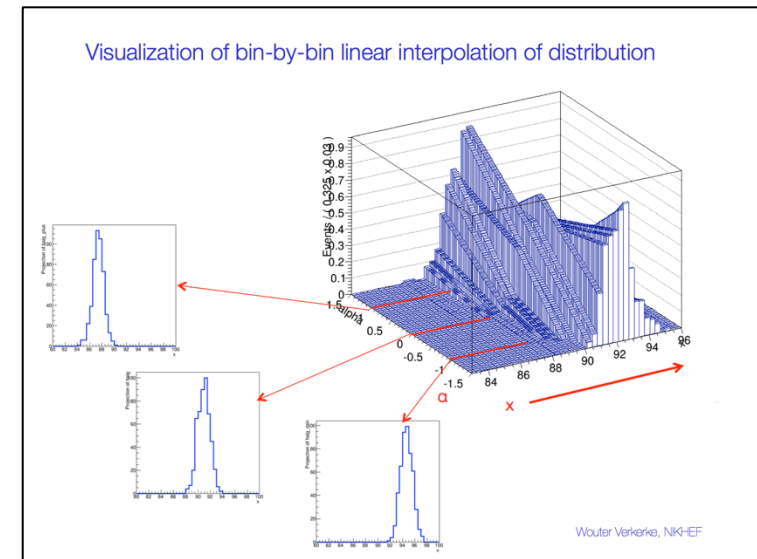
// Complete model Physics(signal plus background)*Subsidiary
w.factory("PROD::model(SUM(mu[0,1]*sig,Uniform::bkg(x)),subs)");
```


Example 3 : binned L with syst

- Example of template morphing systematic in a binned likelihood

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

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



```
// 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))" ) ;
```


Example 4 – Beeston-Barlow light

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

$$L(\vec{N} | \vec{\gamma}) = \underbrace{\prod_{bins} P(N_i | \gamma_i(\tilde{s}_i + \tilde{b}_i))}_{\text{Response function w.r.t. } n \text{ as parameters}} \underbrace{\prod_{bins} P(\tilde{s}_i + \tilde{b}_i | \gamma_i(\tilde{s}_i + \tilde{b}_i))}_{\text{Subsidiary measurements of } n \text{ from } s \sim + b \sim}$$

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

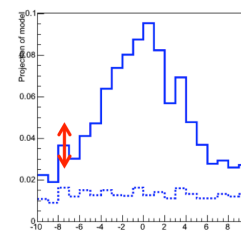
‘Beeston-Barlow’

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

‘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 $s \sim + b \sim$



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

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

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

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

The interplay between shape systematics and MC systematics

- Commonly chosen practical solution

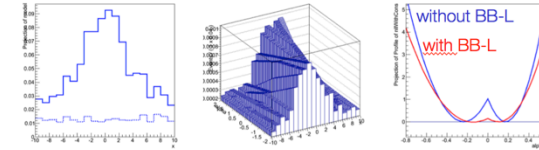
$$s_i(\alpha, \dots) = \begin{cases} s_i^0 + \alpha \cdot (s_i^+ - s_i^0) & \forall \alpha > 0 \\ s_i^0 + \alpha \cdot (s_i^0 - s_i^-) & \forall \alpha < 0 \end{cases}$$

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

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

Wouter Verkerke, Nik4EF

```
// 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)") ;

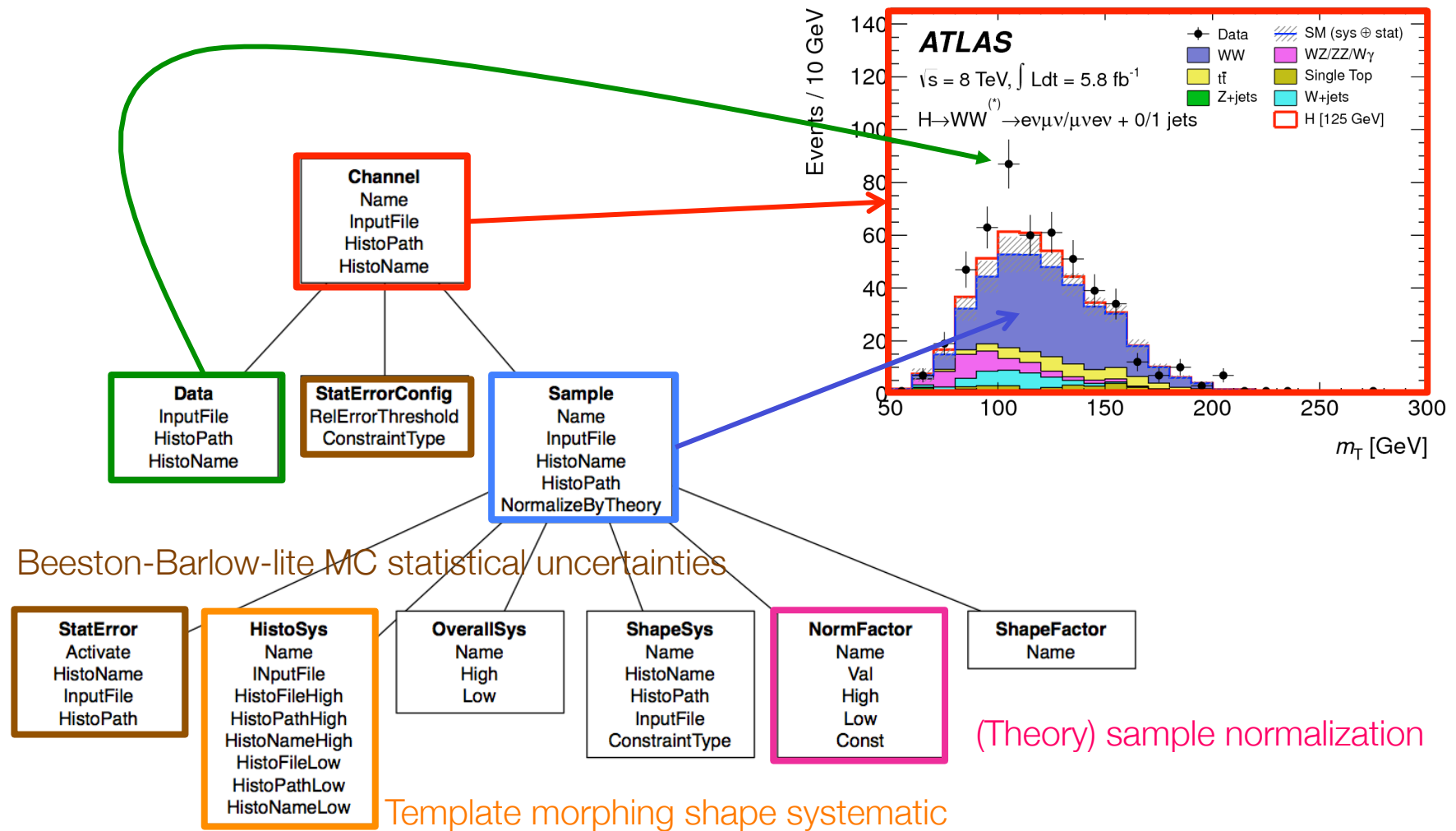
// Construct full model with BB-lite MC stats modeling
w.factory("PROD::model(ASUM(sig,bkg,f[0,1]),
                HistConstraint({s_0,bkg}),Gaussian(0,alpha,1))") ;
```


HistFactory – structured building of binned template models

- **RooFit modeling building blocks** allow to easily construct likelihood models that model shape and rate systematics with one or more nuisance parameter
 - Only few lines of code per construction
- Typical LHC analysis required modeling of 10-50 systematic uncertainties in $O(10)$ samples in anywhere between 2 and 100 channels → Need structured formalism to piece together model from specifications. **This is the purpose of HistFactory**
- **HistFactory conceptually similar to workspace factory**, but has much higher level semantics
 - Elements represent physics concepts (channels, samples, uncertainties and their relation) rather than mathematical concepts
 - Descriptive elements are represented by C++ objects (like roofit), and can be configured in C++, or alternatively from an XML file
 - Builds a RooFit (mathematical) model from a HistFactory physics model.

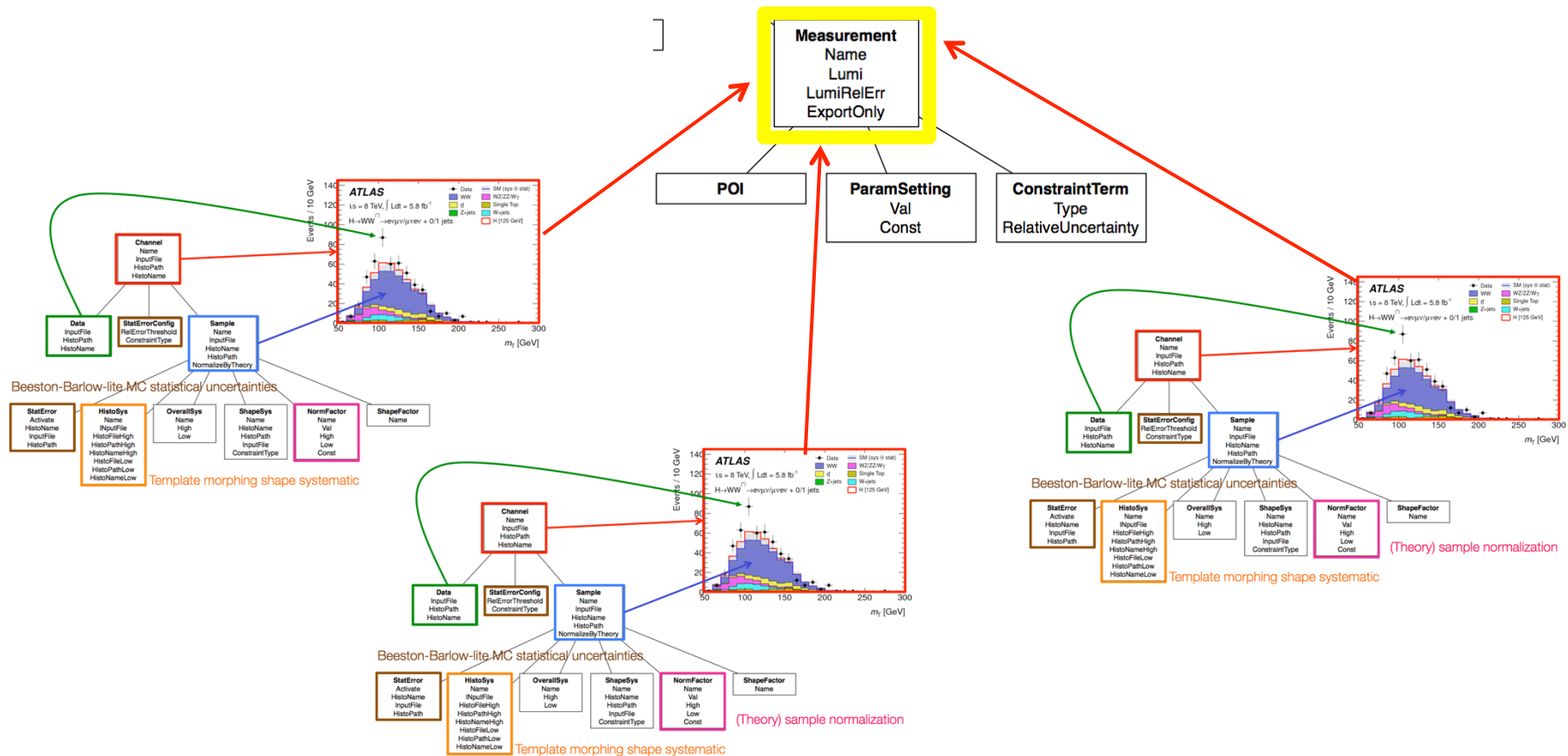
HistFactory elements of a channel

- Hierarchy of concepts for description of one measurement channel



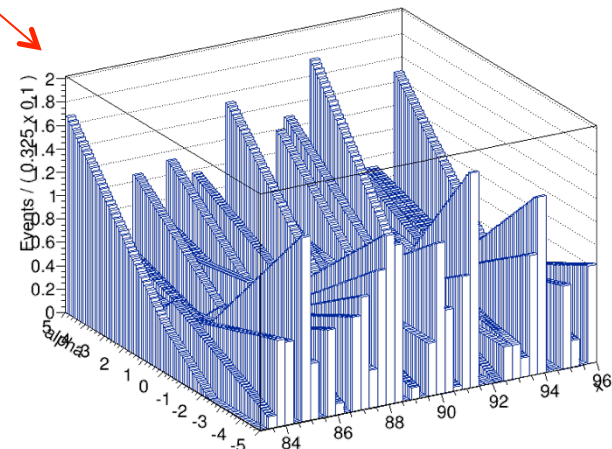
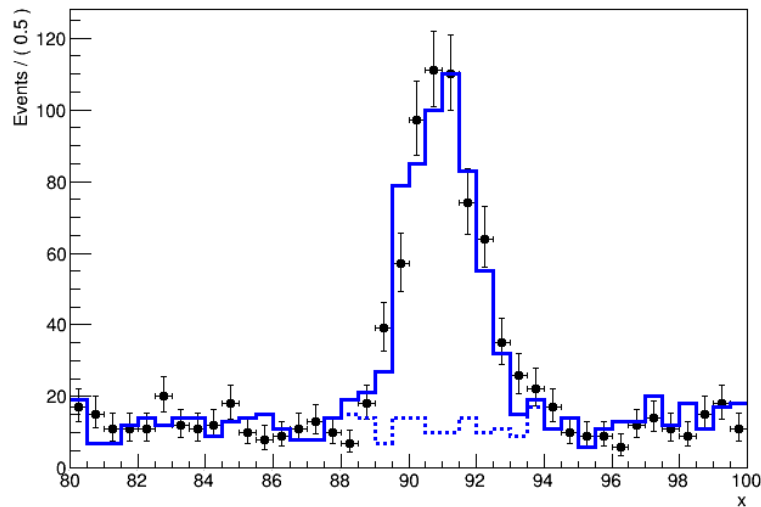
HistFactory elements of measurement

- One or more **channels** are combined to form a **measurement**
 - Along with some extra information (declaration of the POI, the luminosity of the data sample and its uncertainty)



Example of model building with HistFactory

- An example of model building with HistFactory
- Measurement consists of one channel (“VBF”)
- The VBF channel comprises
 1. A data sample
 2. A template model of two samples (“signal” and “qcd”)
 3. The background sample has a “JES” template morphing systematic uncertainty



Model building with HistFactory

```
// external input in form of TH1 shown in green
```

```
// Declare ingredients of measurement
```

```
HistFactory::Data data ;
data.SetHisto(data_hist) ;
```

```
HistFactory::Sample signal("signal") ;
signal.SetHisto(sample_hist) ;
```

```
HistFactory::Sample qcd("QCD") ;
qcd.SetHisto(sample_hist) ;
```

```
HistFactory::HistoSys hsys("QCD_JetEnergyScale") ;
hsys.SetHistoLow(sample_hist_sysdn) ;
hsys.SetHistoHigh(sample_hist_sysup) ;
qcd.AddHistoSys(hsys) ;
```

```
HistFactory::Channel channel("VBF") ;
channel.SetData(data) ;
channel.AddSample(sample) ;
```

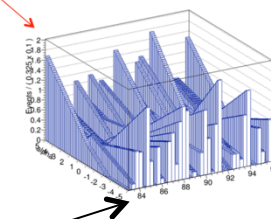
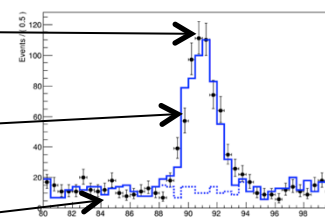
```
HistFactory::Measurement meas("MyAnalysis") ;
meas.AddChannel(channel) ;
```

```
// Now build RooFit model according to specs
```

```
HistFactory::HistoToWorkspaceFactoryFast h2w(meas) ;
RooWorkspace* w = h2w.MakeCombinedModel(meas) ;
w->Print("t") ;
w->writeToFile("test.root") ;
```

Example of model building with HistFactory

- An example of model building with HistFactory
- Measurement consists of one channel ("VBF")
- The VBF channel comprises
 1. A data sample
 2. A template model of two samples ("signal" and "qcd")
 3. The background sample has a "JES" template morphing systematic uncertainty



Wouter Verkerke, NIKHEF

Wouter Verkerke, NIKHEF

HistFactory model output

- Contents of RooFit workspace produced by HistFactory

RooFit
probability
model as
specified

```

RooWorkspace(combined) combined contents

variables
-----
(Lumi,alpha_QCD_JetEnergyScale,binwidth_obs_x_VBF_0,binwidth_obs_x_VBF_1,channelCat,
 nom_alpha_QCD_JetEnergyScale,nominalLumi,obs_x_VBF,weightVar)

p.d.f.s
-----
RooSimultaneous::simPdf[ indexCat=channelCat VBF=model_VBF ] = 0
RooProdPdf::model_VBF[ lumiConstraint * alpha_QCD_JetEnergyScaleConstraint * VBF_model(obs_x_VBF) ] = 0
RooGaussian::lumiConstraint[ x=Lumi mean=nominalLumi sigma=0.1 ] = 1
RooGaussian::alpha_QCD_JESConstraint[ x=alpha_QCD_JetEnergyScale mean=nom_alpha_QCD_JetEnergyScale sigma=1 ] = 1
RooRealSumPdf::VBF_model[ binw_obs_x_VBF_0 * L_x_sig_VBF_overallSyst_x_Exp + binw_obs_x_VBF_1 * L_x_QCD_VBF_overallSyst_x_HistSyst ] = 0
RooProduct::L_x_sig_VBF_overallSyst_x_Exp[ Lumi * sig_VBF_overallSyst_x_Exp ] = 0
RooProduct::sig_VBF_overallSyst_x_Exp[ sig_VBF_nominal * sig_VBF_epsilon ] = 0
RooHistFunc::sig_VBF_nominal[ depList=(obs_x_VBF) ] = 0
RooProduct::L_x_QCD_VBF_overallSyst_x_HistSyst[ Lumi * QCD_VBF_overallSyst_x_HistSyst ] = 0
RooProduct::QCD_VBF_overallSyst_x_HistSyst[ QCD_VBF_Hist_alpha * QCD_VBF_epsilon ] = 0
PiecewiseInterpolation::QCD_VBF_Hist_alpha[ ] = 0
RooHistFunc::QCD_VBF_Hist_alphanominal[ depList=(obs_x_VBF) ] = 0
RooHistFunc::QCD_VBF_Hist_alpha_0low[ depList=(obs_x_VBF) ] = 0
RooHistFunc::QCD_VBF_Hist_alpha_0high[ depList=(obs_x_VBF) ] = 0

datasets
-----
RooDataSet::asimovData(obs_x_VBF,weightVar,channelCat)
RooDataSet::obsData(channelCat,obs_x_VBF)

embedded datasets (in pdfs and functions)
-----
RooDataHist::sig_VBFnominalDHist(obs_x_VBF)
RooDataHist::QCD_VBF_Hist_alphanominalDHist(obs_x_VBF)
RooDataHist::QCD_VBF_Hist_alpha_0lowDHist(obs_x_VBF)
RooDataHist::QCD_VBF_Hist_alpha_0highDHist(obs_x_VBF)

parameter snapshots
-----
NominalParamValues = (nominalLumi=1[C],nom_alpha_QCD_JetEnergyScale=0[C],weightVar=0,obs_x_VBF=-4.5,Lumi=1,alpha_QCD_JetEnergyScale=0,
 binwidth_obs_x_VBF_0=1[C],binwidth_obs_x_VBF_1=1[C])

named sets
-----
ModelConfig_GlobalObservables:(nominalLumi,nom_alpha_QCD_JetEnergyScale)
ModelConfig_Observables:(obs_x_VBF,weightVar,channelCat)
ModelConfig_POI:()
globalObservables:(nominalLumi,nom_alpha_QCD_JetEnergyScale)
observables:(obs_x_VBF,weightVar,channelCat)

generic objects
-----
RooStats::ModelConfig::ModelConfig
    
```

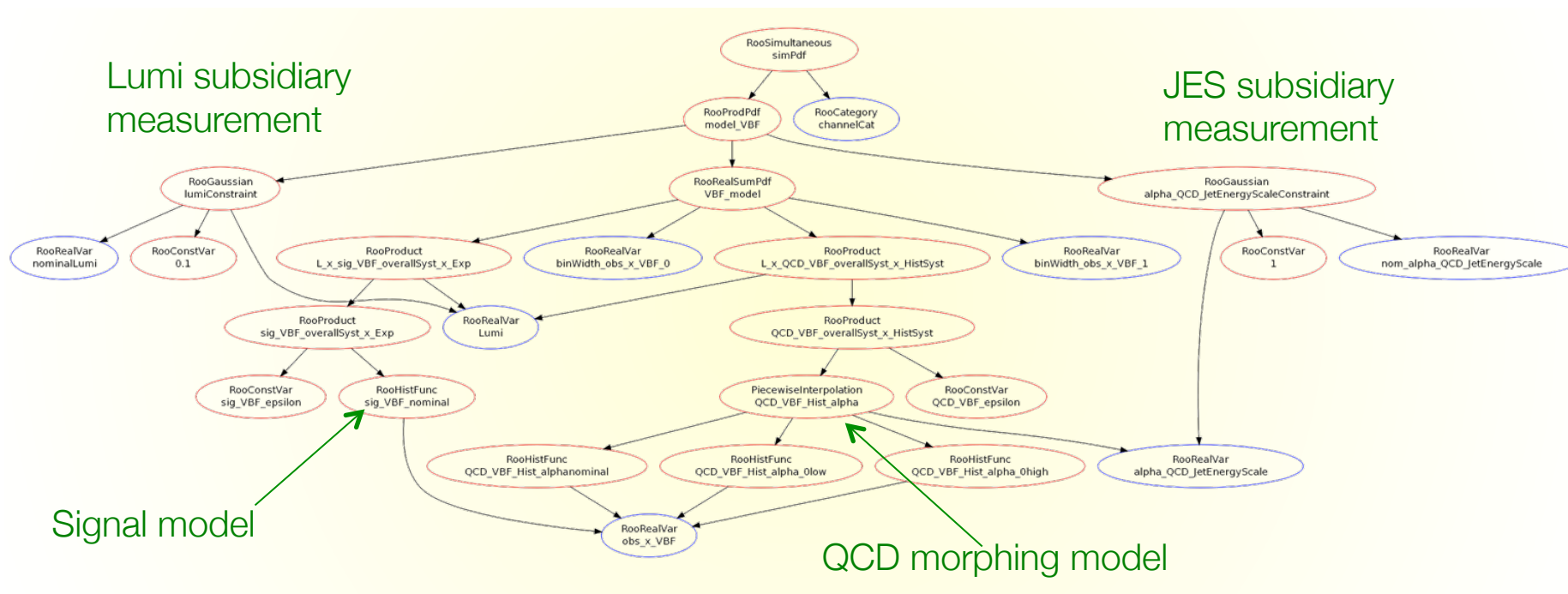
Definition of
POI, NPs,
Observables

Global observables

Universal
Model Configuration

HistFactory model structure

- RooFit object structure
 - As visualized with `simpdf::graphvizTree("model.dot")` followed by `dot -Tpng -o model.png model.dot`



- This RooFit probability model can be evaluated without knowledge of HistFactory
 - Additional (documentary) information stored in workspace specifies a uniquely specified statistical model (definition of POI, NP etc)

5 Diagnostics II: Overconstraining & choices in modeling parametrization

Role reversal of physics and subsidiary measurements

- As mention in Section 3, full (profile) likelihood treats physics and subsidiary measurement on equal footing

$$L(N, 0 | s, \alpha) = \underbrace{Poisson(N | s + b(1 + 0.1\alpha))}_{\text{Physics measurement}} \cdot \underbrace{Gauss(0 | \alpha, 1)}_{\text{Subsidiary measurement}}$$

- Our mental picture:
 - “measures s ”
 - “measures α ”
 - “dependence on α weakens inference on s ”
- Is this picture (always) correct?

Understanding your model – what constrains your NP

- The answer is no – not always! Your physics measurement may in some circumstances constrain α better than your subsidiary measurement.
- Doesn't happen in Poisson counting example
 - Physics likelihood has no information to distinguish effect of s from effect of α

$$L(N, 0 | s, \alpha) = \underbrace{Poisson(N | s + b(1 + 0.1\alpha))}_{\text{Physics measurement}} \cdot \underbrace{Gauss(0 | \alpha, 1)}_{\text{Subsidiary measurement}}$$

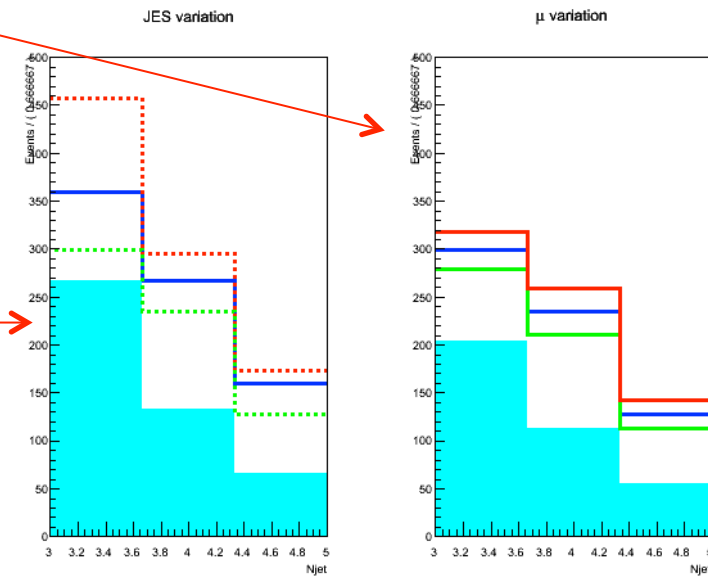
- But if physics measurement is based on a distribution or comprises multiple distributions this is well possible

Understanding your model – what constrains your NP

- A case study – measuring jet multiplicity
 - Physics observable of interest is a jet multiplicity spectrum [3j,4j,5j] after an (unspecified) p_T cut.
 - Describe data with sum of signal (mildly peaking at 4j) and a single background (exponentially falling in nj).

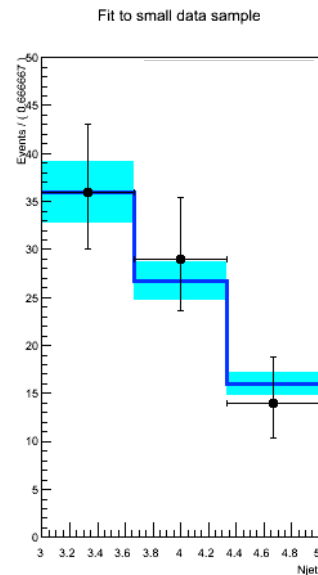
$$L(\vec{N} | \mu, \alpha_{JES}) = \prod_{i=3,4,5} \text{Poisson}(N_i | (\mu \cdot \tilde{s}_i + \tilde{b}_i) \cdot r_s(\alpha_{JES})) \cdot \text{Gauss}(0 | \alpha_{JES}, 1)$$

- POI is signal strength modifier μ .
- Jet Energy Scale is the leading systematic uncertainty
 - JES strongly affects jet multiplicity after a p_T cut,
 - Effect modeled by response function $r_s(\alpha)$
 - Magnitude of uncertainty on α constrained by subsidiary measurement

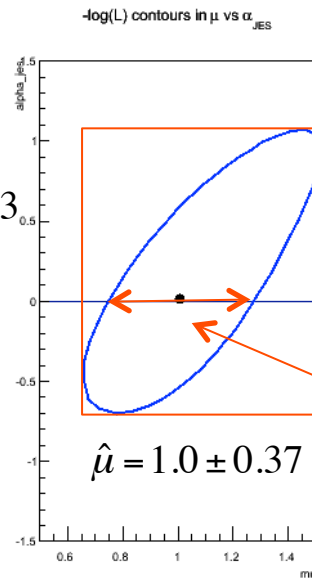


Understanding your model – what constrains your NP

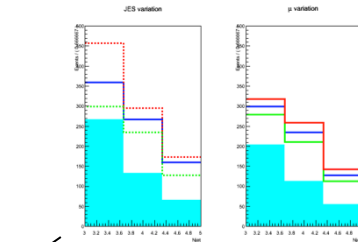
- Now measure (μ, α) from data – 80 events



$$\hat{\alpha} = 0.01 \pm 0.83$$



$$\hat{\mu} = 1.0 \pm 0.37$$



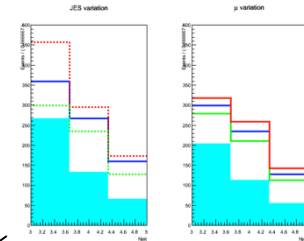
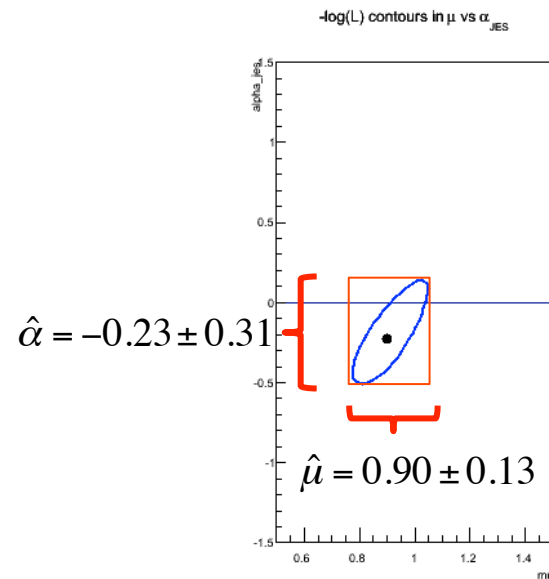
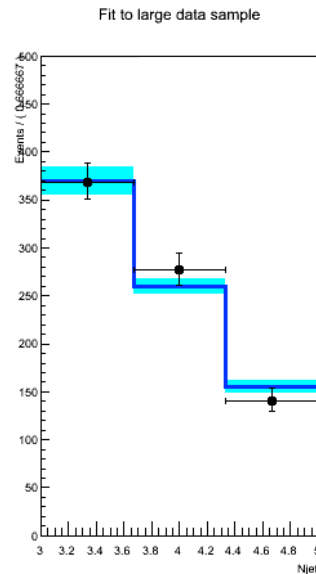
Estimators of μ , α correlated due to similar response in physics measurement

Uncertainty on μ without effect of JES

- Is this fit OK?
 - Effect of JES uncertainty propagated in to μ via response modeling in likelihood. Increases total uncertainty by about a factor of 2
 - Estimated uncertainty on α is not precisely 1, as one would expect from unit Gaussian subsidiary measurement...

Understanding your model – what constrains your NP

- The next year – 10x more data (800 events) repeat measurement with same model



Estimators of μ , α correlated due to similar response in physics measurement

- Is this fit OK?
 - Uncertainty of JES NP *much reduced* w.r.t. subsidiary meas. ($\alpha = 0 \pm 1$)
 - Because the physics likelihood can measure it better than the subsidiary measurement (the effect of μ , α are sufficiently distinct that both can be constrained at high precision)

Understanding your model – what constrains your NP

- Is it OK if the physics measurement constrains NP associated with a systematic uncertainty better than the designated subsidiary measurement?
 - From the statisticians point of view: no problem, simply a product of two likelihood that are treated on equal footing ‘simultaneous measurement’
 - From physicists point of view? Measurement is only valid if model is valid.
- Is the probability model of the physics measurement valid?

$$L(\vec{N} | \mu, \alpha_{JES}) = \prod_{i=3,4,5} Poisson(N_i | (\mu \cdot \tilde{s}_i + \tilde{b}_i) \cdot r_s(\alpha_{JES})) \cdot Gauss(0 | \alpha_{JES}, 1)$$

- Reasons for concern
 - Incomplete modeling of systematic uncertainties,
 - Or more generally, model insufficiently detailed

Understanding your model – what constrains your NP

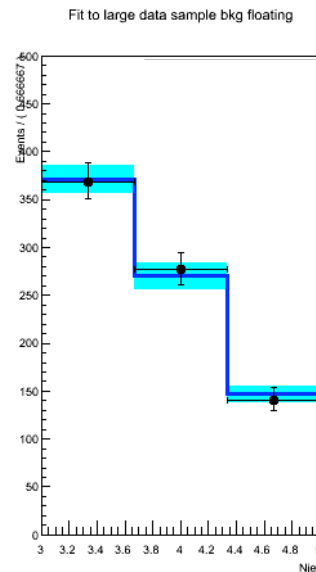
- What did we overlook in the example model?
 - The background rate has no uncertainty!
 - Insert modeling of background uncertainty

$$L(\vec{N} | \mu, \alpha_{JES}, \alpha_{bkg}) = \prod_{i=3,4,5} Poisson(N_i | (\underbrace{\mu \cdot \tilde{s}_i + \tilde{b}_i \cdot r_b(\alpha_{bkg})}_{\text{Background rate response function}}) \cdot \underbrace{r_s(\alpha_{JES})}_{\text{Background rate subsidiary measurement}})) \cdot Gauss(0 | \alpha_{JES}, 1) \cdot Gauss(0 | \alpha_{bkg}, 1)$$

Background rate
response function

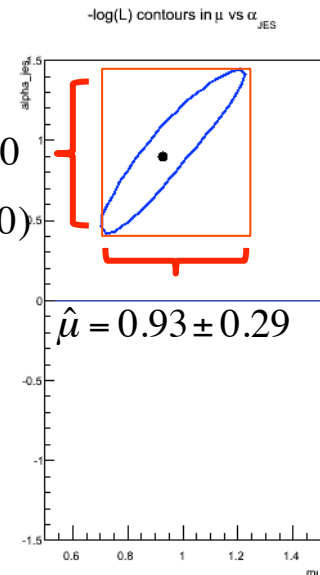
Background rate
subsidiary measurement

- With improved model accuracy estimated uncertainty on both α_{JES} , μ goes up again...
 - Inference weakened by new degree of freedom α_{bkg}
 - NB α_{JES} estimate still deviates a bit from normal distribution estimate...



$$\hat{\alpha}_{JES} = 0.90 \pm 0.70$$

$$(\hat{\alpha}_{bkg} = 1.36 \pm 0.20)$$



$$\hat{\mu} = 0.93 \pm 0.29$$

Understanding your model – what constrains your NP

- Lesson learned: if probability model of a physics measurement is insufficiently detailed (i.e. flexible) you can *underestimate* uncertainties
- Normalized subsidiary measurement provide an excellent diagnostic tool
 - Whenever estimates of a NP associated with unit Gaussian subsidiary measurement deviate from $\alpha = 0 \pm 1$ then physics measurement is constraining or biases this NP.
 - Always inspect all NPs of your fit for such signs
- Is ‘over-constraining’ of systematics NPs always bad?
 - No, sometimes there are good arguments why a physics measurement can measure a systematic uncertainty better than a dedicated calibration measurement (that is represented by the subsidiary measurement)
 - Example: in sample of reconstructed hadronic top quarks $t \rightarrow bW(qq)$, the pair of light jets should always have $m(jj)=mW$. For this special sample of jets it will possible to calibrate the JES better than with generic calibration measurement

Commonly heard arguments in discussion on over-constraining

- Overconstraining of a certain systematic is OK “because this is what the data tell us”
 - It is what the data tells you *under the hypothesis that your model is correct*. The problem is usually in the latter condition
- “The parameter α_{JES} should not be interpreted as Jet Energy Scale uncertainty provided by the jet calibration group”
 - A systematic uncertainty is always combination of response prescription and one or more nuisance parameters uncertainties.
 - If you implement the response prescription of the systematic, then the NP in your model really is the same as the prescriptions uncertainty
- “My estimate of $\alpha_{\text{JES}} = 0 \pm 0.4$ doesn’t mean that the ‘real’ Jet Energy Scale systematic is reduced from 5% to 2%”
 - It certainly means that in your analysis a 2% JES uncertainty is propagated to the POI instead of the “official” 5%.
 - One can argue that the 5% shouldn’t apply because your sample is special and can be calibrated better by a clever model, but this is a physics argument that should be documented with evidence for that (e.g. argument JES in $t \rightarrow bW(qq)$ decays)

Dealing with over-constrained systematic NPs

- Step 1 – **Diagnostics**
 - Always inspect nuisance parameters in your fit for signs of over-constraining
- Step 2 – **Analyze**
 - Are there systematic uncertainties overlooked in the construction of the likelihood that introduce unwarranted physics assumption in model that ML estimator exploits to constrain models?
 - Is your systematic uncertainty conceptually covered by a single nuisance parameter? do you perhaps need more NPs?
 - In case the physics likelihood comprises multiple samples, do you assume fully correlated responses functions, whereas sample composition should conceptually allow for some degree of decorrelation?
- Step 3 – **Solution**
 - If over-constraining is analyzed to be the result of inaccurate modeling, improve model structure, add new NPs, decompose NPs in different ways to reflect sample correlations
 - If constraint from physics is believed to be document studies as part of your physics analysis

Dealing with over-constraining – introducing more NPs

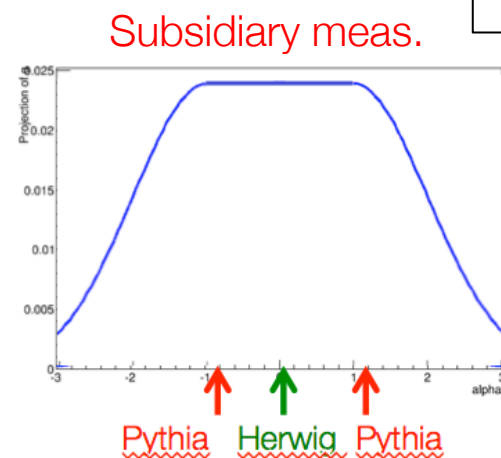
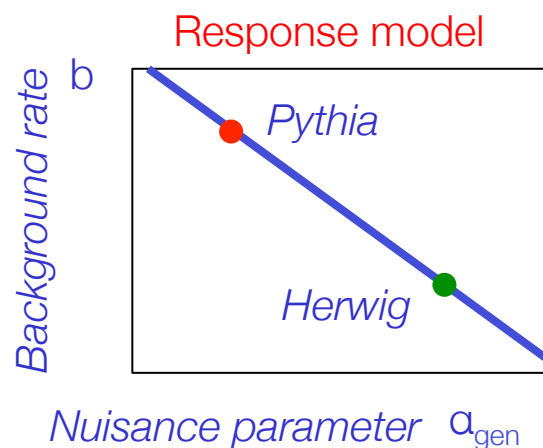
- Some systematic uncertainties are not captured well by one nuisance parameter.
- Example Jet Energy Scale
 - Statement “the JES uncertainty is 5% for all jets” does not necessarily imply that the calibration of all jets can be modeled with a single NP.
 - A single NP implies that the calibration can only be coherently off among all jets. Such an assumption allows, for example, to precisely constrain JES with a high-statistics sample of low- p_T jets, and then transport that reduced uncertainty to high- p_T jets, using the calibration scale coherency encoded in the model
 - In reality correlation between the energy scale of low- p_T and high- p_T jets is controlled by the detector design and calibration procedure and is likely a lot more complicated → Invalid modeling of systematic uncertainties often a result of ‘own interpretation’ of imprecisely formulated systematic prescription.
 - Besides this, a calibration may have multiple sources of uncertainty that were lumped together in a prescription (calibration measurements, simulation assumptions, sample-dependent effects) that would need to be individually modeled

Dealing with over-constraining – Theory uncertainties

- Over-constraining of theory uncertainties in physics measurements has different set of issues than for detector uncertainties
- **Different:** In principle it is the goal of physics measurements to constrain theory uncertainties
 - So role of physics measurement and subsidiary measurement are not symmetric: the latter quantifies some ‘degree of belief’ that is not based on an experimental measurement.
 - Likelihood of physics measurement constitutes an experimental measurement and is in principle preferred over ‘belief’
 - But question remains if physics likelihood was well designed to constrain *this particular theory uncertainty*.
- **Same:** response function and set of NPs must be able to accurately capture underlying systematic effect.
 - Sometimes easy, e.g. ‘renormalization scale’ has well-defined meaning in a given theoretical model and a clearly identifiable single associated parameter
 - Sometimes hard, e.g. ‘Pythia vs Herwig’. Not clear what it means or how many degrees of freedom underlying model has.

Dealing with ‘two-point’ uncertainties

- In discussion of rate systematics in Section 3 it was mentioned that ‘two-point systematics’ can always be effectively represented with an interpolation strategy



Specific issues with theory uncertainties

- Pragmatic solutions to likelihood modeling of ‘2-point systematics’
- Final solution will need to follow usual pattern

$$L(N | s, \alpha) = \text{Poisson}(N | s + b \cdot f(\alpha)) \cdot \text{SomePdf}(0 | \alpha)$$

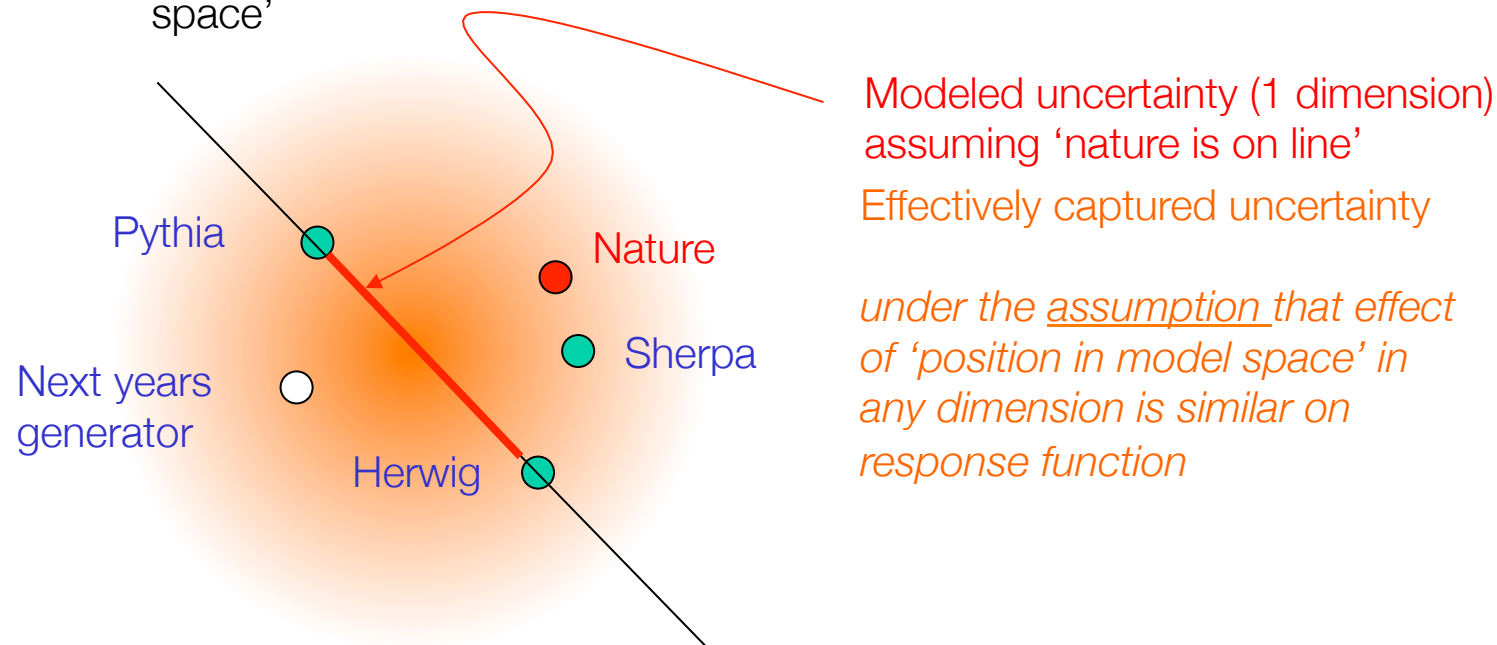
- Since underlying concept of systematic uncertainty not defined, the only option is to *define its meaning terms in terms of response in the physics measurement*
- Example
 - Estimate of bkg with Herwig = 8, with Pythia = 12
 - In the likelihood choose $b=8$ and then define $f(\alpha) = [1 + 4 \cdot \alpha]$, so that $f(0)$ results in ‘Herwig ($b, f=8$)’ and $f(1)$ results in ‘Pythia ($b, f=12$)’
 - For lack of a better word you could call α now the ‘Herwigness of fragmentation w.r.t. its effect on my background estimate’
- A thorny question remains: What is the subsidiary measurement for α ?
 - This should reflect your current knowledge on α .

Wouter Verkerke, NIKHEF

- But this argument relies crucially on the dimensional correspondence between the observable and the NP
 - The effect on a scalar observable can always be modeled with one NP
 - In other words the existence of a 3rd generator ‘Sherpa’ can always be effectively capture by the Pythia-Herwig inter/extrapolation
 - It can of course modify your subsidiary measurement (e.g. lending more credence to the Pythia outcome if its result is close, but response model is still valid)

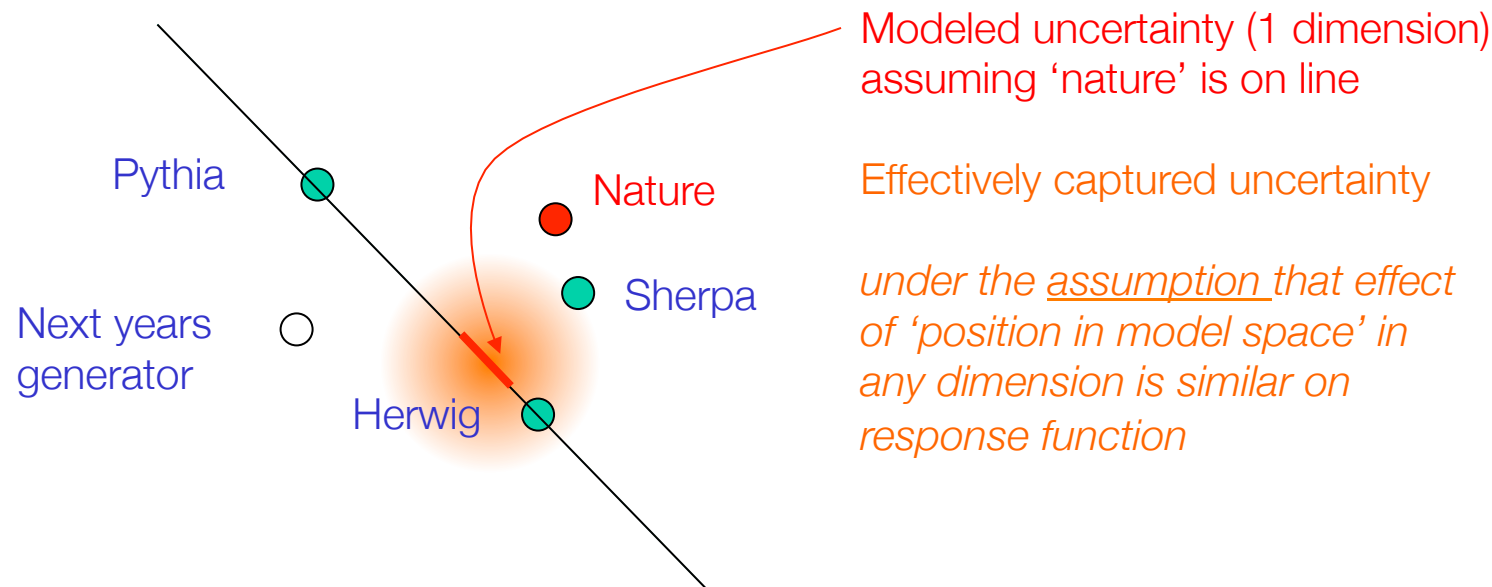
Dealing with 'two-point' uncertainties

- If '2-point' response functions models a distribution, the response corresponding to a new 'third point' is not necessarily mapped by $b(a)$ for any value of a
- This point is important in the discussion to what extent a two-point response function can be over-constrained.
 - A result $\alpha_{2p} = 0.5 \pm 1$ has 'reasonable' odds to cover the 'true generator' assuming all generators are normally scattered in an imaginary 'generator space'



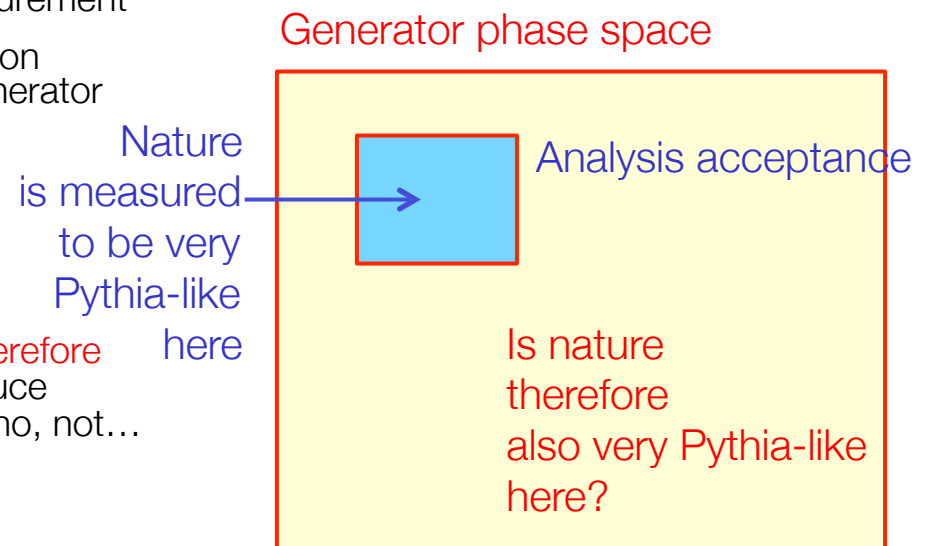
Dealing with 'two-point' uncertainties

- If '2-point' response functions models a distribution, the response corresponding to a new 'third point' is not necessarily mapped by $b(a)$ for any value of a
- This point is important in the discussion to what extent a two-point response function can be over-constrained.
 - Does a hypothetical overconstrained result $\alpha_{2p} = 0.1 \pm 0.2$ 'reasonably' cover the generator model space?



Dealing with ‘two-point’ uncertainties

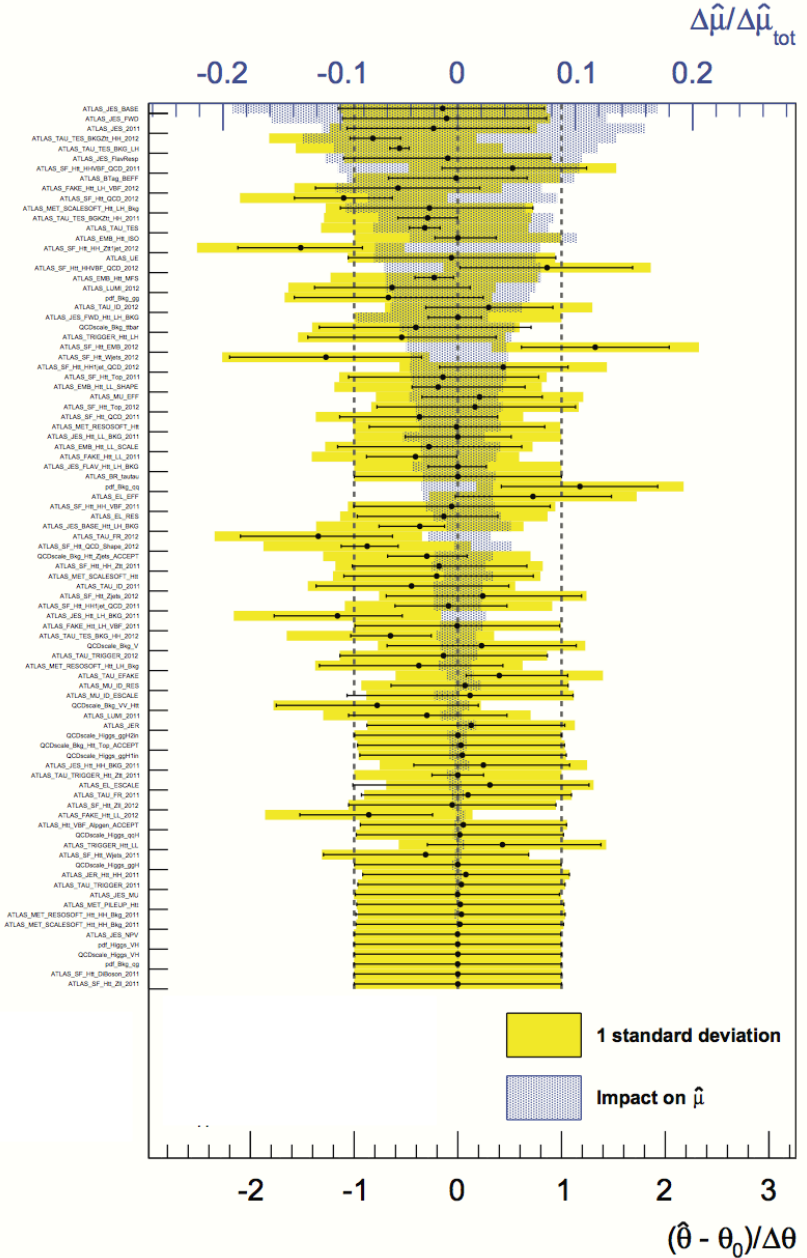
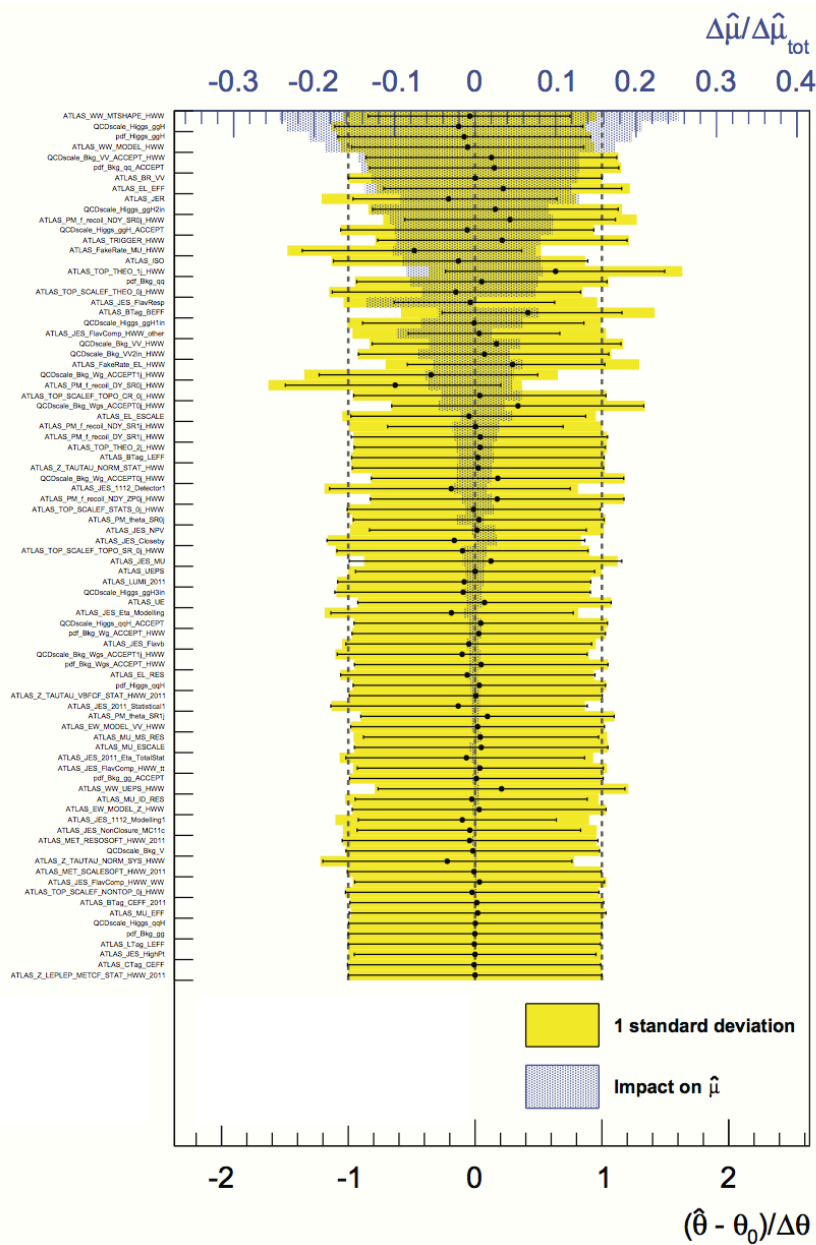
- Arguments on representativeness of sampling points of ‘2 point’ models raise questions in validity of physics models that over-constrain these
- The main problem is that you become rather sensitive to things you don’t know and quantify: the ‘dimensionality’ of the generator space.
 - To understand what you are doing you’d need to know what all degrees of freedom are (and ideally what they conceptually represent)
 - Unless you know this – trying to reduce the ‘considered space of possibilities’ is rather speculative
 - The real problem is often that you don’t really know what causes the ‘Pythia/Herwig’ effect. Unless you learn more about that there is no real progress.
- The ‘unknown dimensionality’ problem often enters a model in a seemingly standard modeling assumptions
 - Take an inclusive cross-section measurement
 - Needs to extrapolate acceptance region to full inclusive phase space using generator
→ Introduces generator systematic
 - Physics likelihood can ‘measure’ that nature inside acceptance is very Pythia-like inside using 2-point response function with 1 NP
 - Is nature in the entire phase space therefore Pythia-like? If yes, we can greatly reduce inclusive cross-section uncertainty, if no, not...



Summary

- Diagnostics over NP overconstraining provide powerful insight into your analysis model
 - An overconstrained NP indicates an externally provided systematic is inconsistent with physics measurement
 - This may point to either an incorrect response modeling of that uncertainty, to result in a genuinely better estimate of the uncertainty
 - Solution not always clear-cut, but you should be at least aware of it.
 - Note that over-constraining always points to an underlying physics issue (lack of knowledge, simplistic modeling) → Treat it as a physics analysis problem, not as a statistics problem
- Diagnostic power of profile likelihood models highlights one of the major shortcomings of the ‘naïve’ strategy of error propagation (as discussed in Section 1)
 - Physics measurement can entangle in non-trivial ways with systematic uncertainties

Example of likelihood modeling diagnostics



6

Summary & conclusions

Summary

- Modelling of systematic uncertainties in the likelihood ('profiling') is the best we know to incorporate systematic uncertainties in rigorous statistical procedures
 - Profiling requires more a 'exact' specification of what a systematic uncertainty means than traditional prescriptions → this is good thing, it makes you think about (otherwise hidden) assumption
 - It's important to involve the 'author' of uncertainty prescription in this process, as flawed assumptions can be exploited by statistical methods to arrive at unwarranted conclusions
 - Systematic uncertainties that have conceptual fuzziness ('pythia-vs-herwig') are difficult to capture in the likelihood, but this is a reflection of an underlying physics problem
 - Good software tools exist to simplify the process of likelihood modeling
 - It's important to carefully diagnose your profile likelihood models for both technical and interpretational problems ('over-constraining')