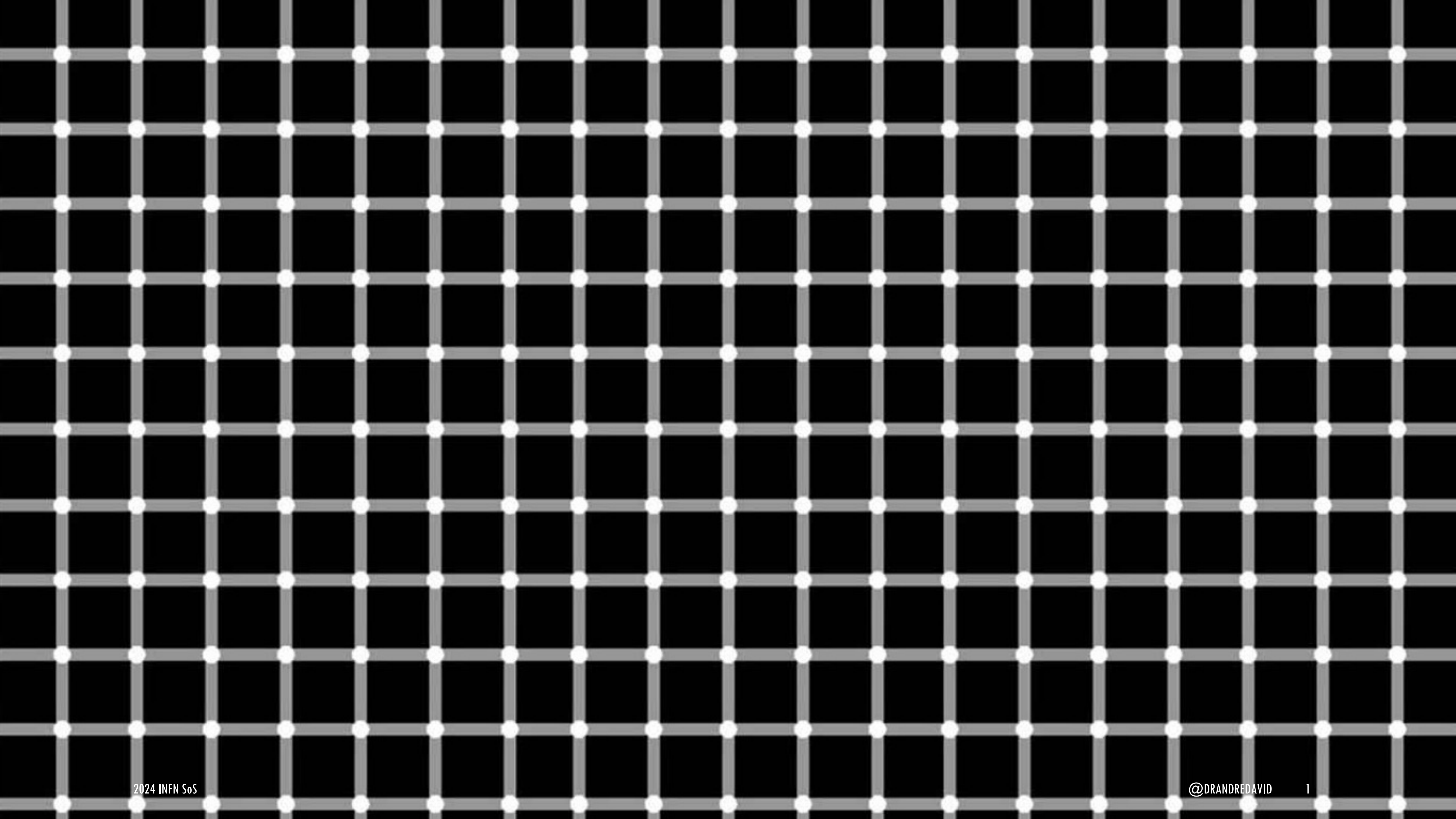


SYSTEMATIC UNCERTAINTIES AND NUISANCE PARAMETERS

A. David (CERN and IST-Lisboa)

ABOUT THE SPEAKER

Detector builder, Nature scrutiniser, Engineer botherer, and Theory disprover.

PhD in NA60, until 2013 in CLOUD, and CMS since 2006:

- CMS ECAL, 1st LHC single isolated photon cross-section measurement, then two photons, then one Higgs, then COMBINE Higgs analyses, then Higgs co-convener.
- Presently building CMS HGCAL for Phase 2 #LifeWithHexagons.

Profiled as many nuisance parameters as there are millionaires in the world.

LIMITED ACCEPTANCE WARNING

👷 **I am not a statistician,** just a physicist trying to be more accurate.

- Since 2000 that physicists and statisticians meet in PHYSTAT to figure out many of these issues.

🪟 This view is **incomplete** and **has biases.**

- Both are my fault, not that of the sources I used.

🦄 *My goal: share some of what I have found useful when building an analysis.*

BEFORE ~~DINNER~~ ANDARE ALLA SPIAGGIA

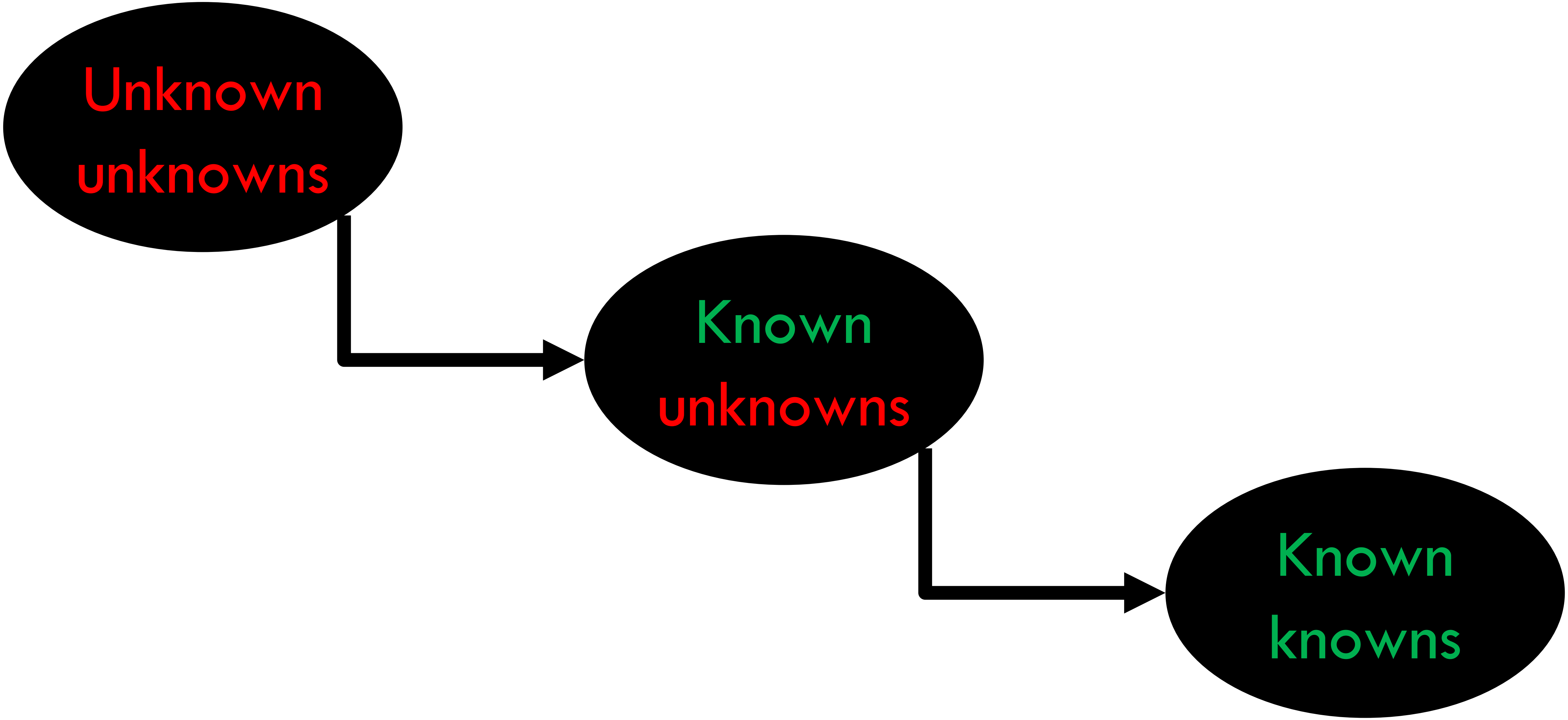
1. 📏 What are systematic uncertainties.
2. ⚗️ How we account for them using nuisance parameters.
3. ⚙️ How frequentist inference is done in the present of such beasts.
4. 🧐 How to live with them and answer your reviewers' questions.
5. 🗺️ Walkthroughs of practices and caveats from experience in the wild.

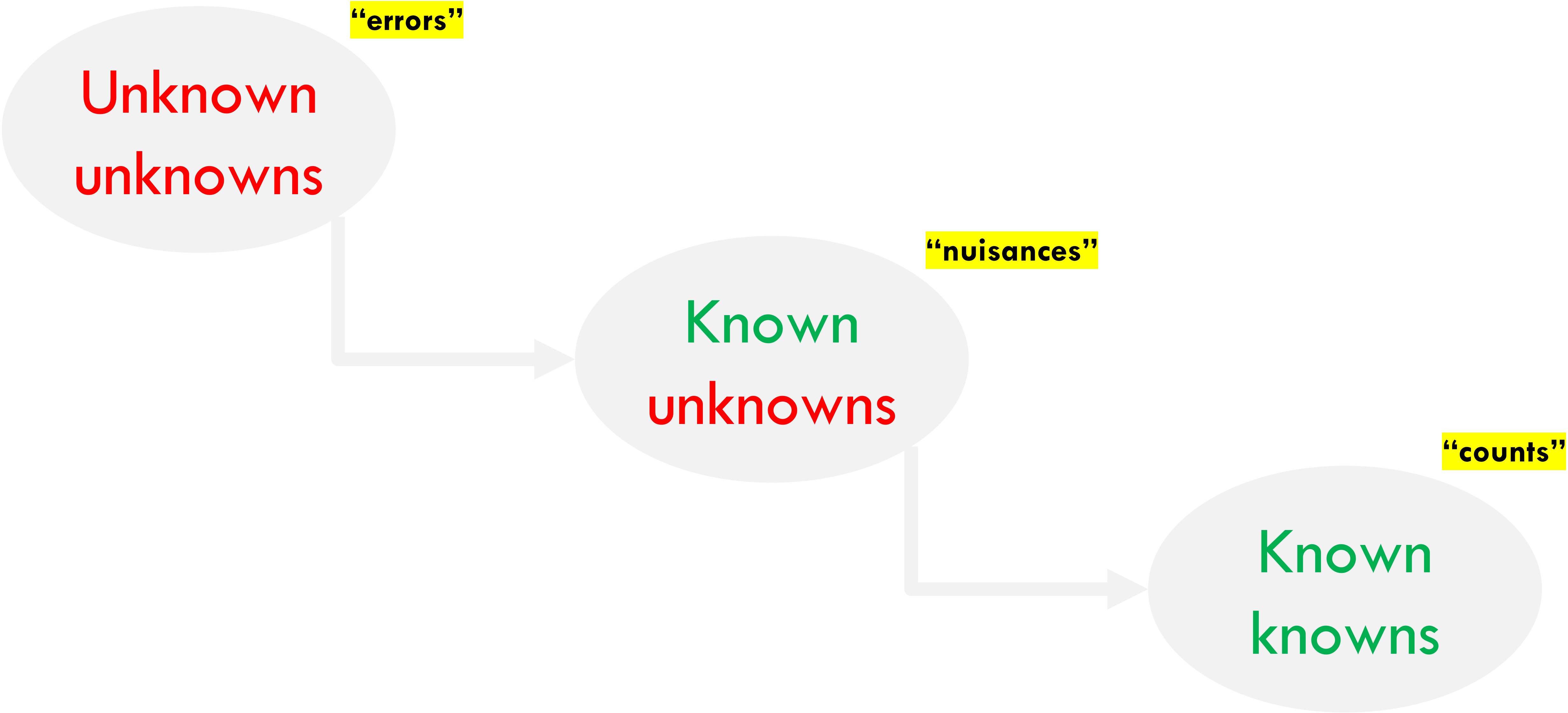
☝🏻 ***Interrupt whenever you have a question.***
Worst that can happen is to discuss it alla spiaggia.

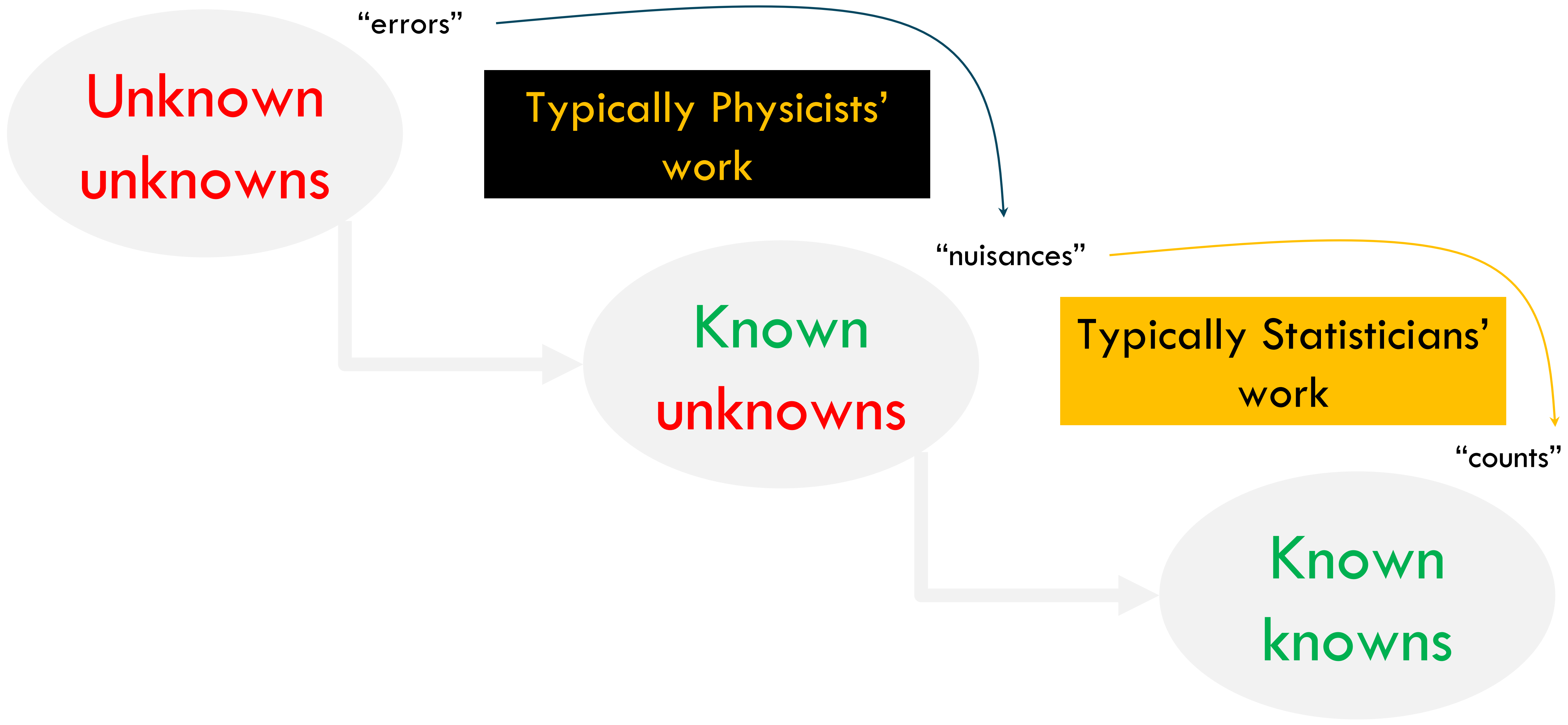
ℹ️ Personal disclaimer: *publish frequentist results and take Bayesian decisions.*

SYSTEMATIC UNCERTAINTIES AND NUISANCE PARAMETERS

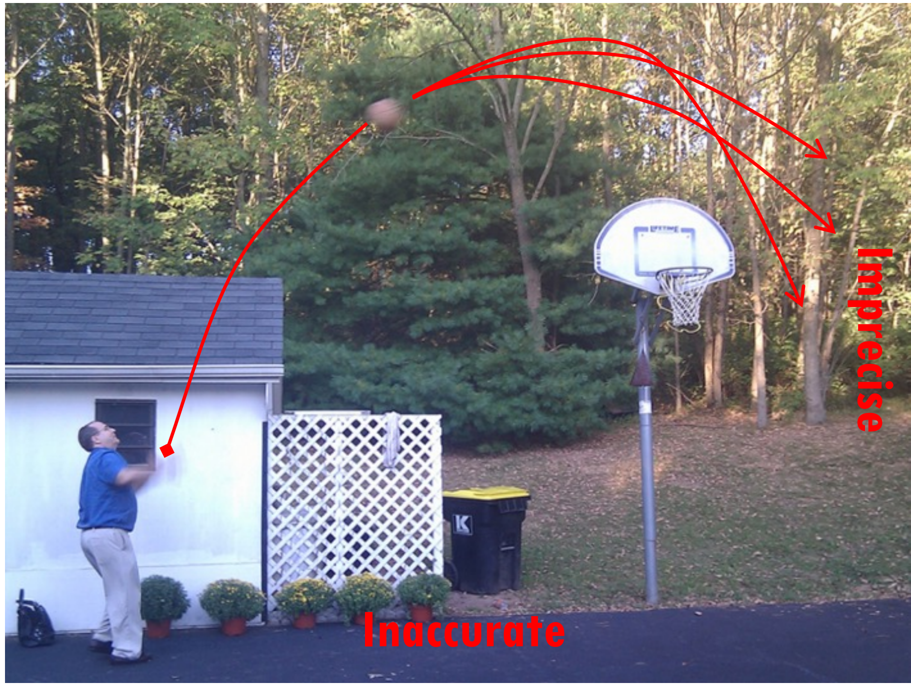
A. David (CERN and IST-Lisboa)


Unknown unknowns
Known unknowns
Known knowns

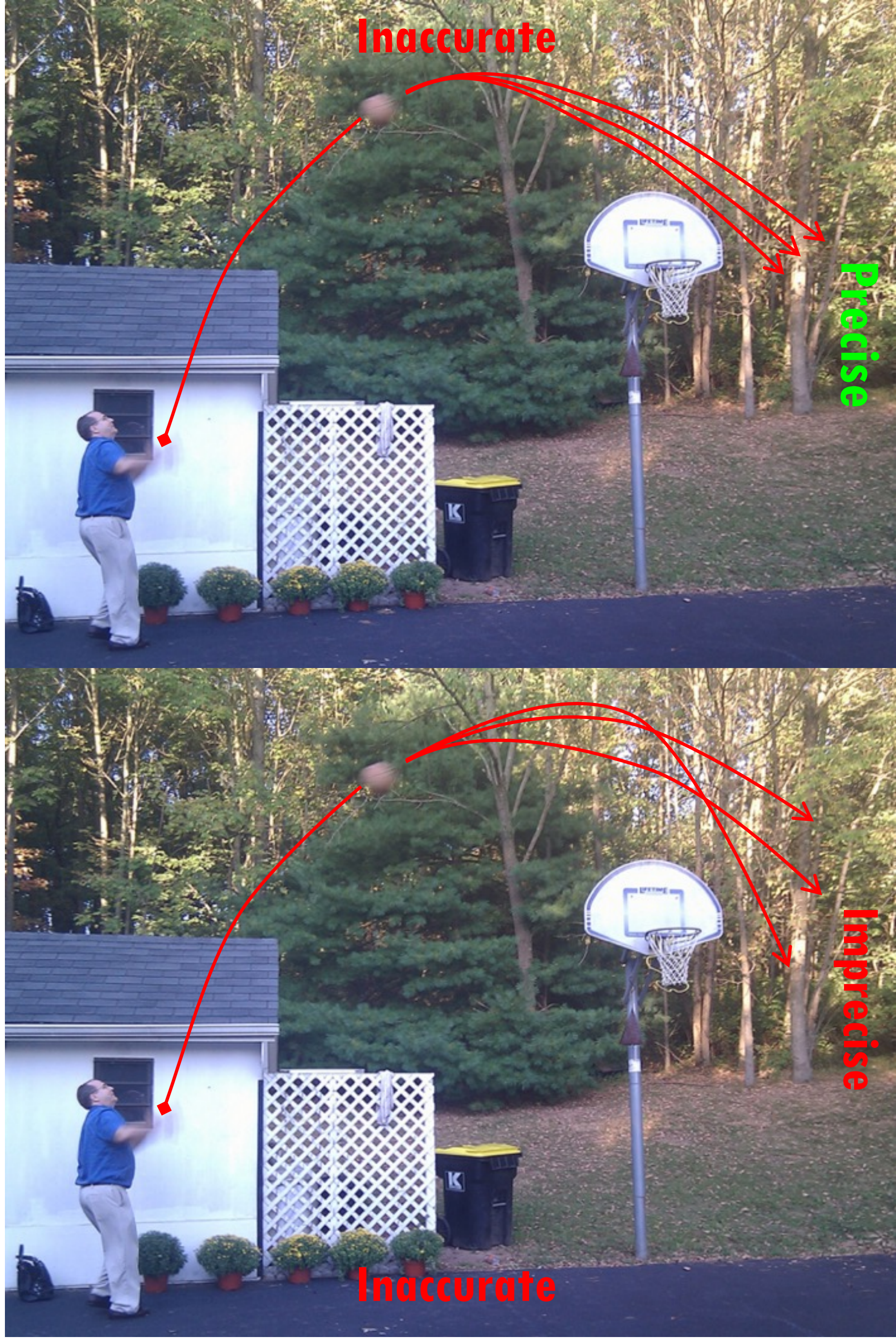

"errors"
Unknown unknowns
"nuisances"
Known unknowns
"counts"
Known knowns


"errors"
Unknown unknowns
Typically Physicists' work
"nuisances"
Known unknowns
Typically Statisticians' work
"counts"
Known knowns

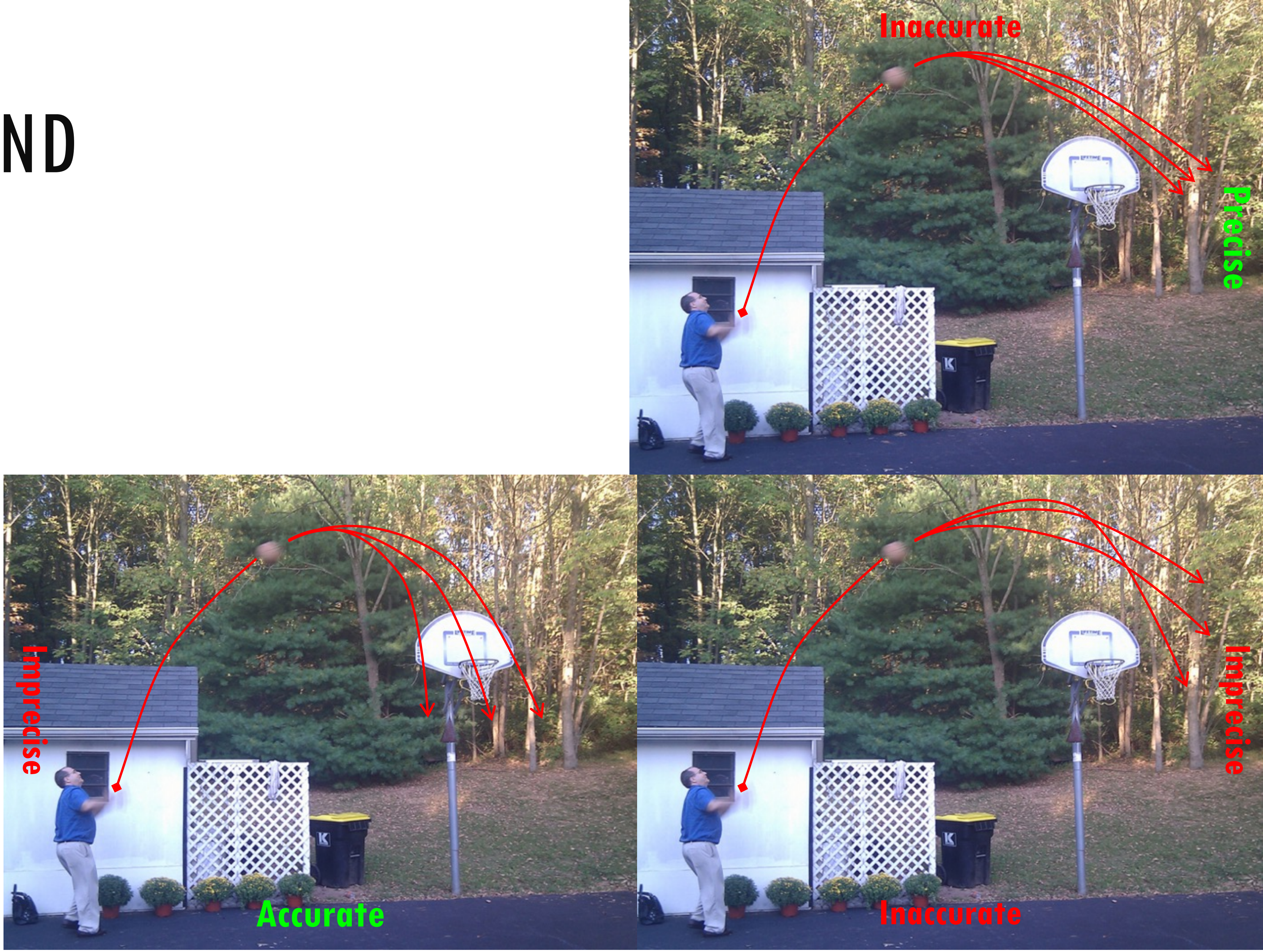
ACCURACY AND PRECISION



ACCURACY AND PRECISION



ACCURACY AND PRECISION

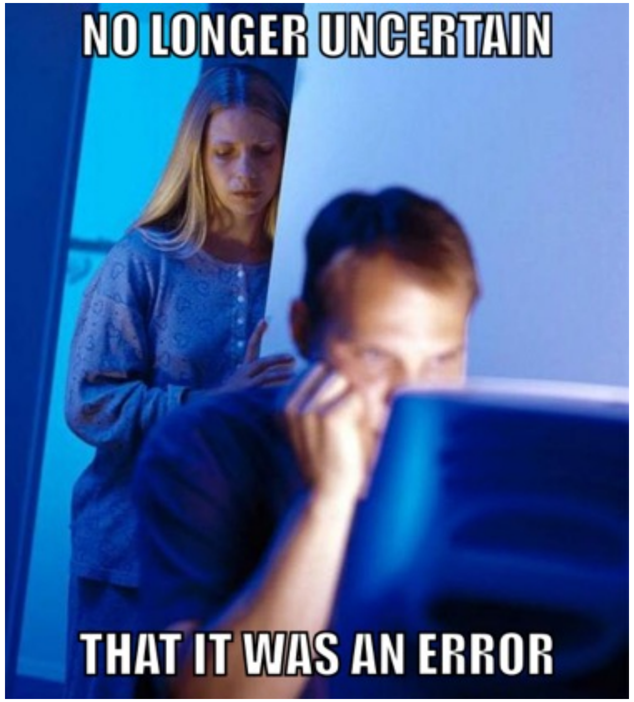


ACCURACY AND PRECISION

TWO WORDS ON ERROR AND UNCERTAINTY

What I call **error** is the result of a **bias** or **mistake.**

What I call **uncertainty** is the degree to which something is (un)known.



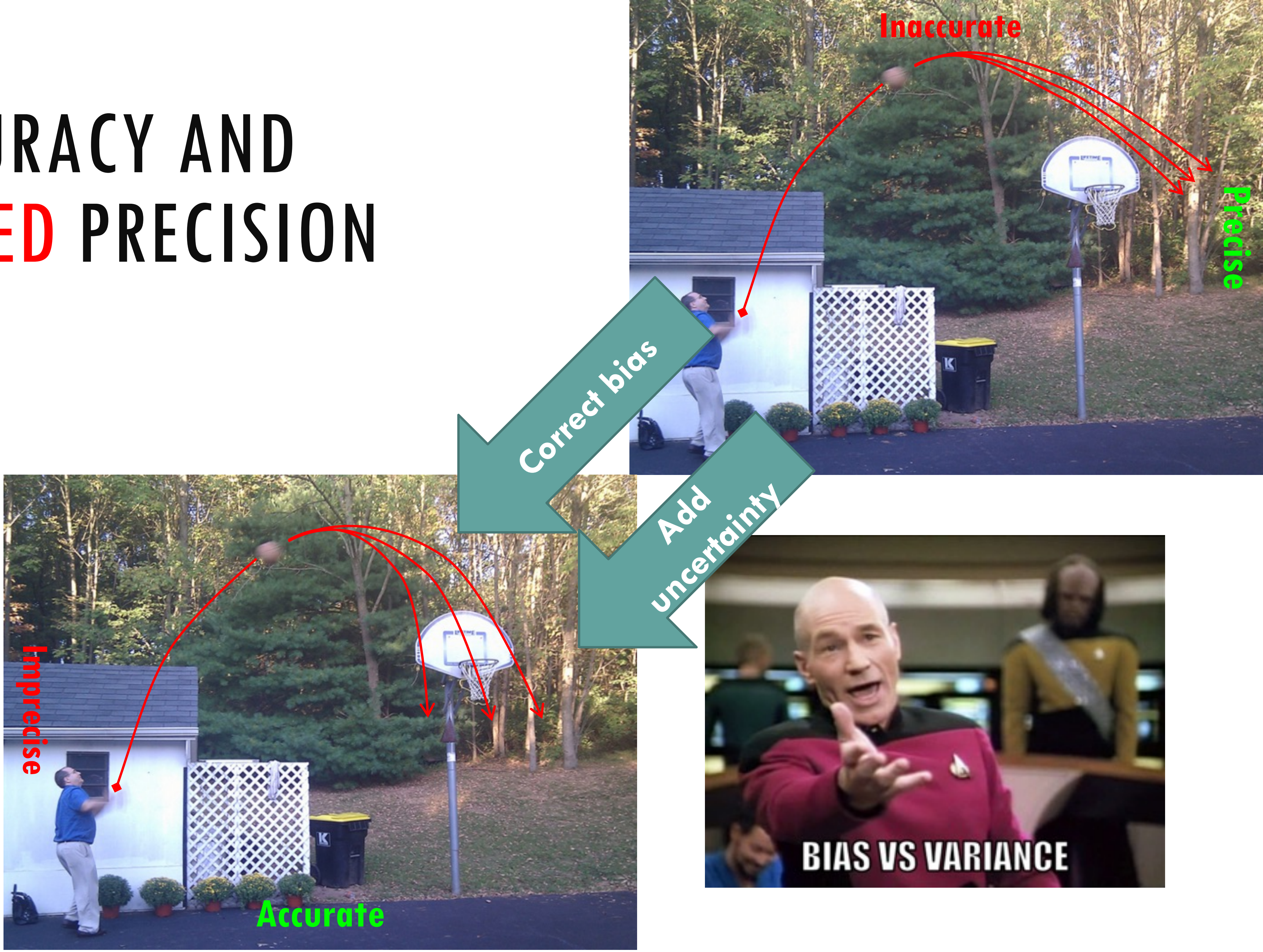
I think it's a mistake to call errors uncertainties.

E.g., experimentalists correct for systematic effects (biases).

- Corrections come with added uncertainty.

BETTER ACCURACY AND DETERIORATED PRECISION

BETTER ACCURACY AND DETERIORATED PRECISION



EXAMPLE: METAL RULER ON A HOT DAY

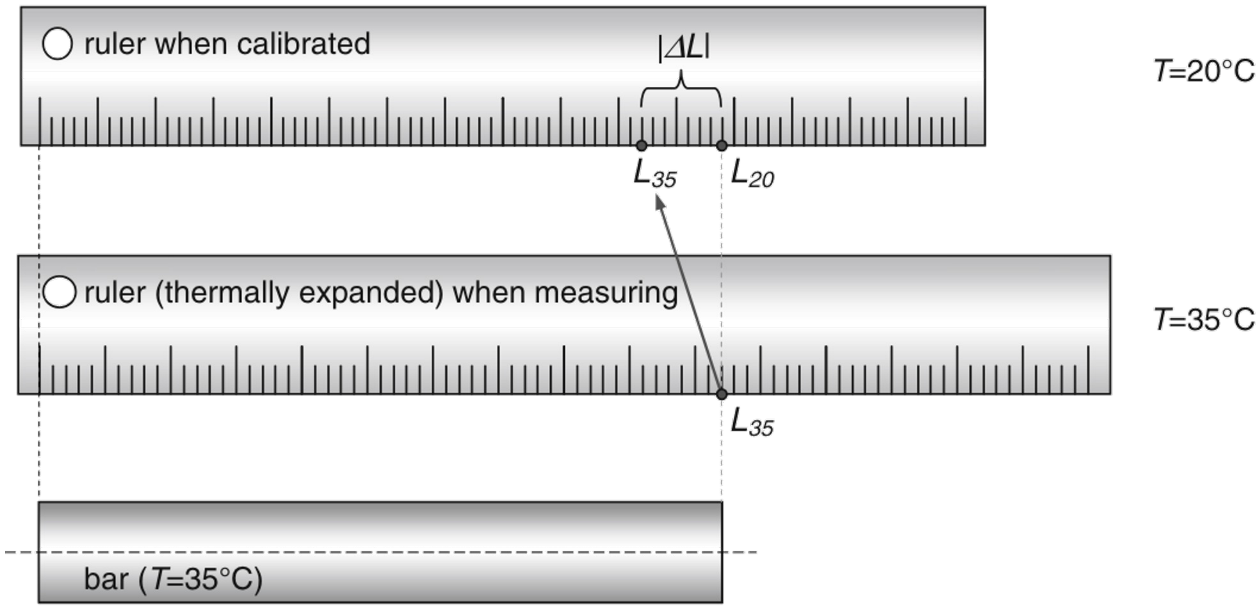
Givens:

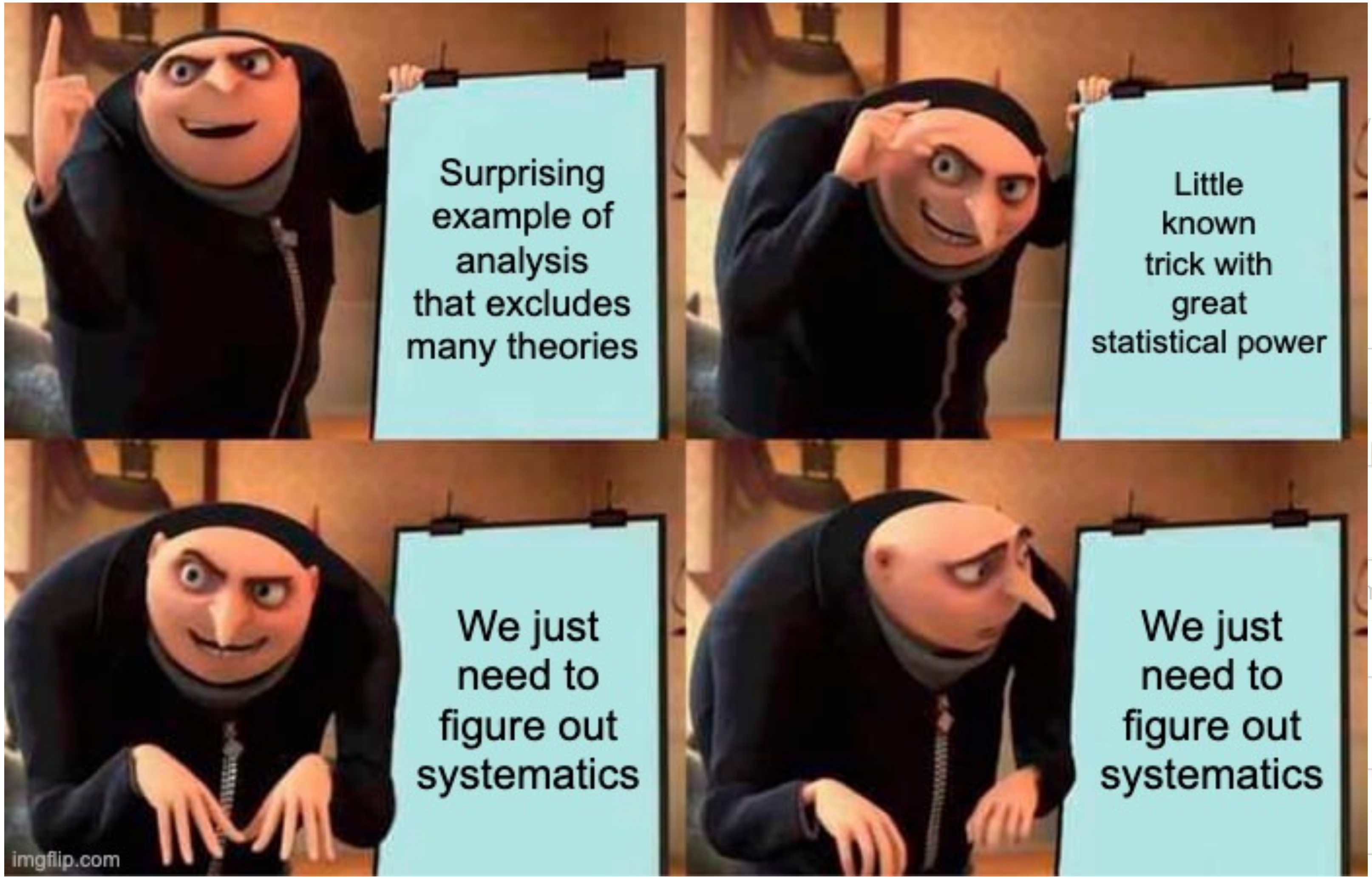
- Bar measured with ruler at T_1 = 35 °C.
- Metal ruler calibrated at T_0 = 20 °C.
- Ruler thermal expansion coefficient is α.
  - Measured in some way that will have uncertainty δα.

Estimates:

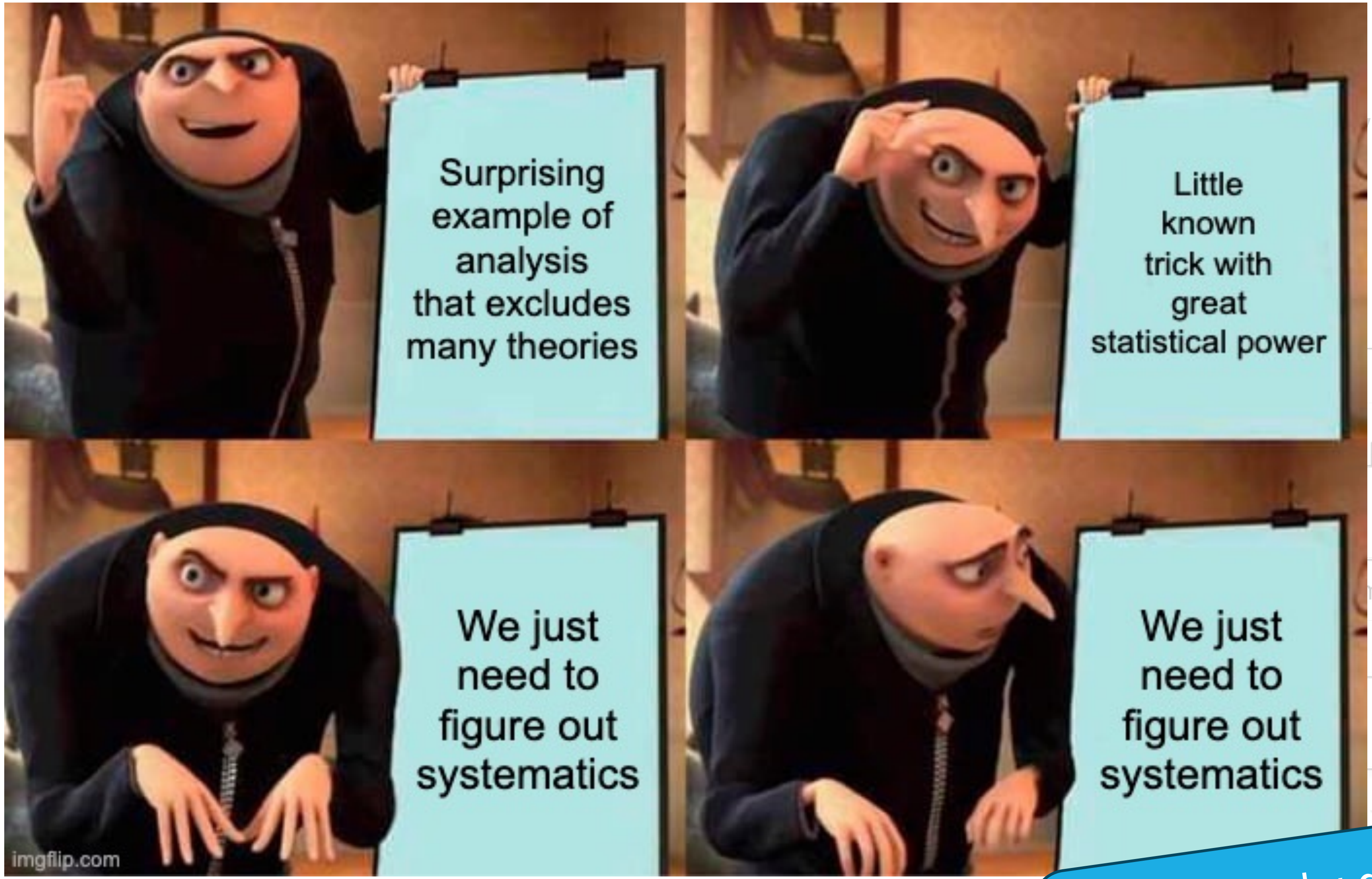
- Central value: $L_{bar} = \mathbf{L_{20} = L_{35} [1 + \alpha (T_0 - T_1)]}$
  - Reduced bias !
- Unc. on L_{bar} includes **uncertainties on L_{35}, α, T_1.**
  - Additional uncertainty !

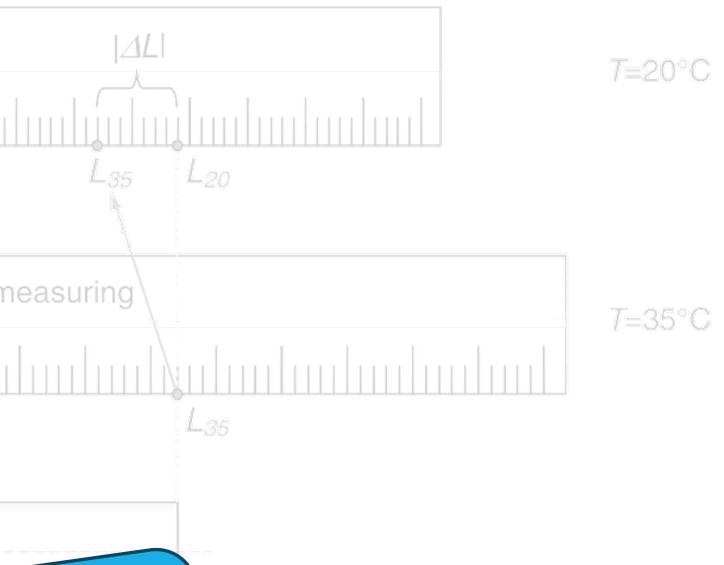
In practice: make δα and δT_1 contributions smaller than that from δL_{35}.





In practice: make $\delta\alpha$ and δT_1 contributions smaller than that from δL_{35}.

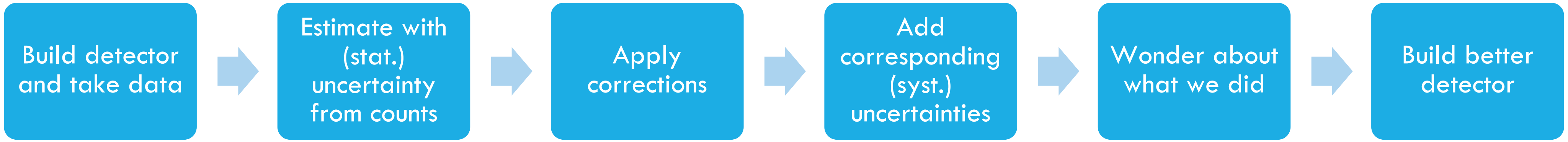




In practice: make $\delta\alpha$ and δT_1 contributions smaller than that from δL_{35}.

The reason why analyses are 10% inspiration and 90% perspiration

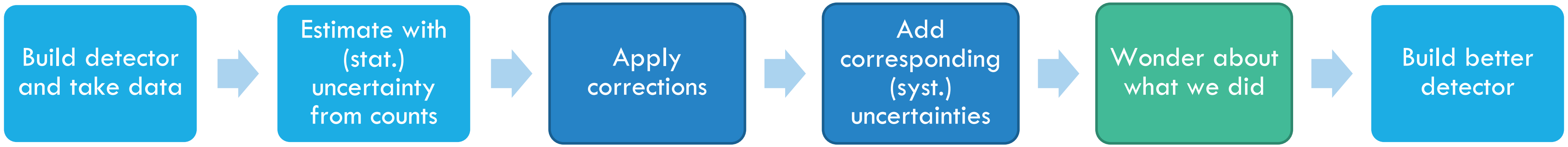
OUR SYSTEMATIC PLIGHT WITH BIAS



OUR SYSTEMATIC PLIGHT WITH BIAS

Build detector and take data → Estimate with (stat.) uncertainty from counts → Apply corrections → Add corresponding (syst.) uncertainties → Wonder about what we did → Build better detector

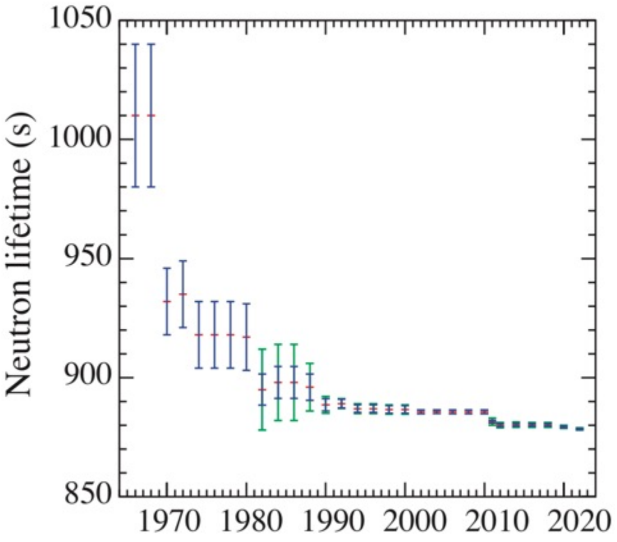
OUR SYSTEMATIC PLIGHT WITH BIAS

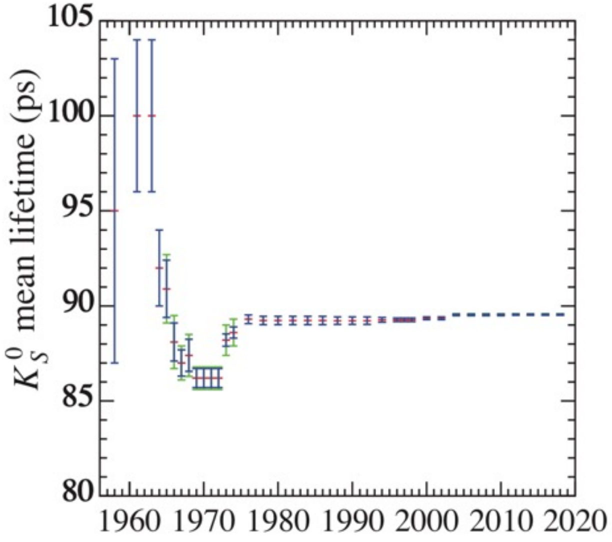


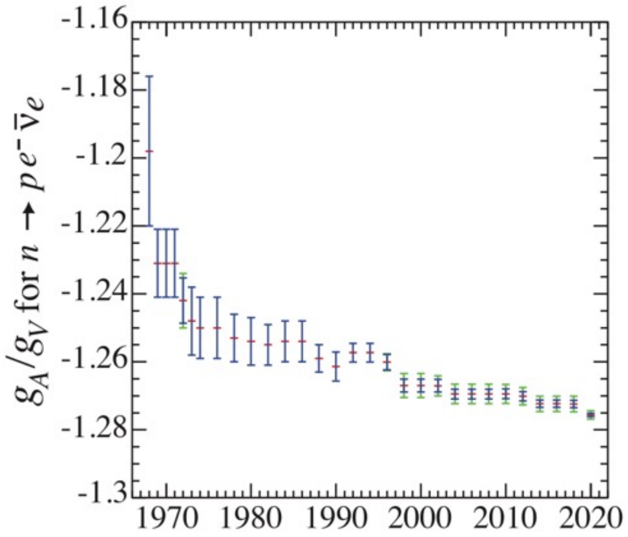
OUR SYSTEMATIC PLIGHT WITH BIAS

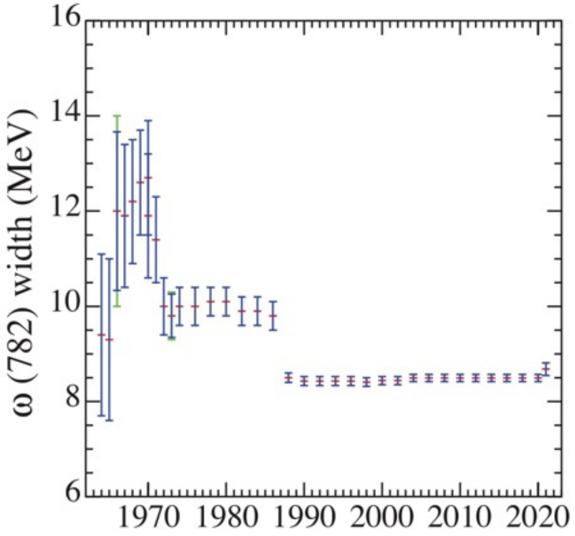
Build detector and take data → Estimate with (stat.) uncertainty from counts → Apply corrections → Add corresponding (syst.) uncertainties → Wonder about what we did → **Build better detector**

LOOKING BACK TO LOOK FORWARD



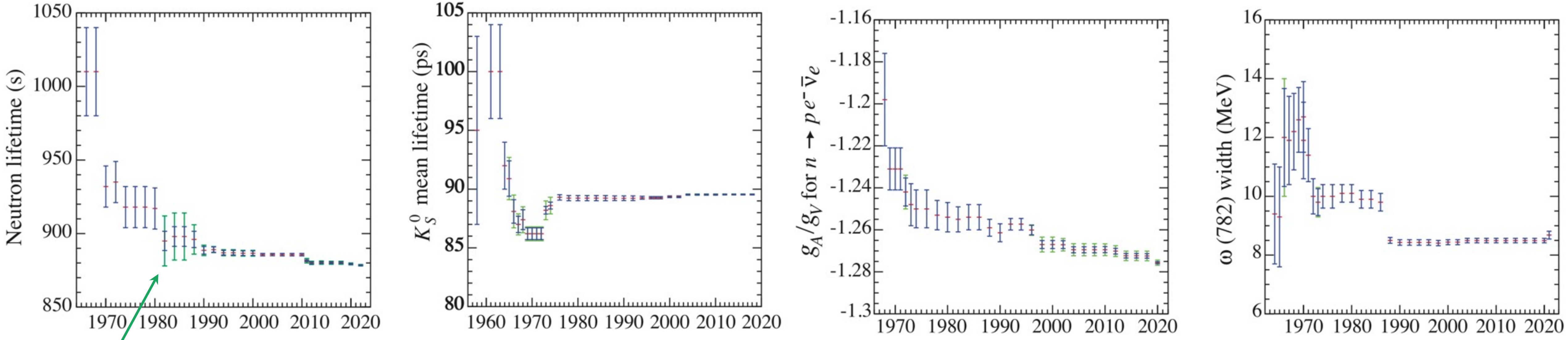




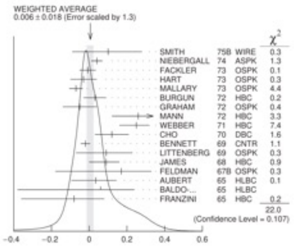


Yes, The PDG refers to uncertainties as "errors". 🤷‍♂️

LOOKING BACK TO LOOK FORWARD

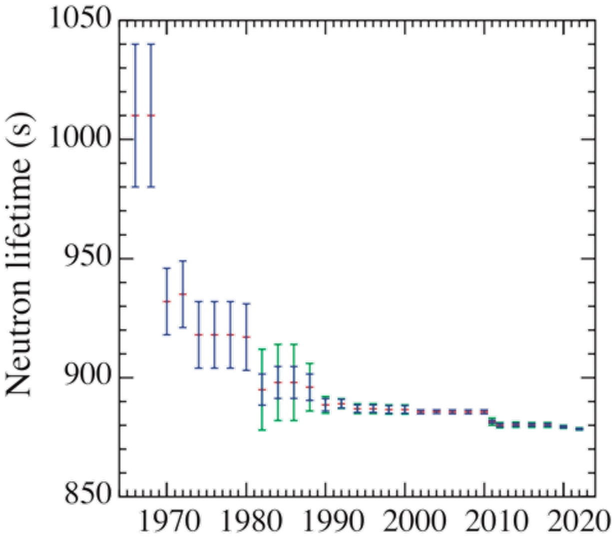


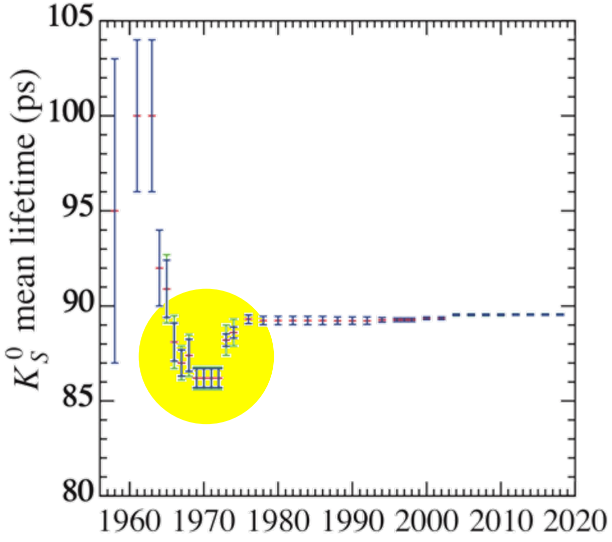
PDG scales uncertainties to deal with discrepant measurements.

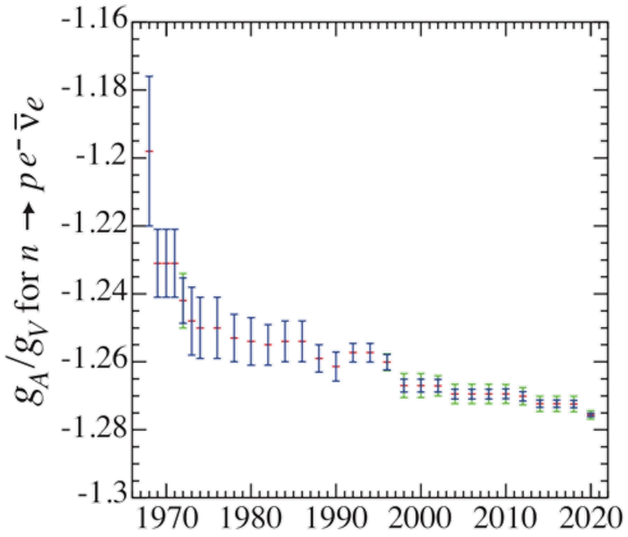


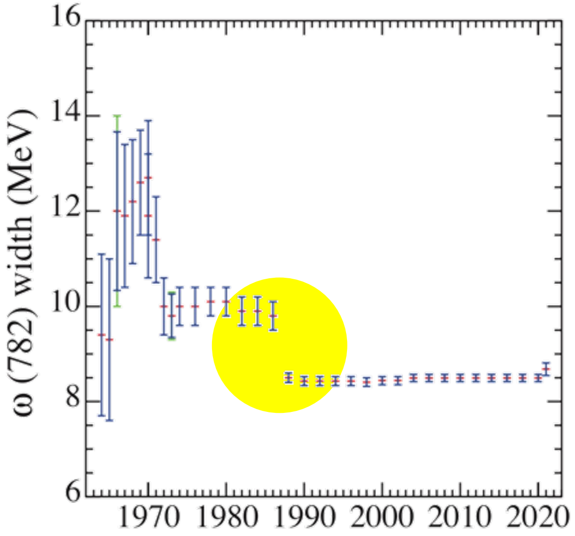
Yes, The PDG refers to uncertainties as "errors". 🤷

LOOKING BACK TO LOOK FORWARD









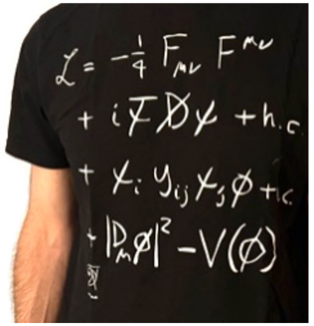
A brief history of early Particle Data Group averages is given in Ref. [5]. Our History Plots show the time evolution of some of our values of a few particle properties. Sometimes large changes occur. These usually reflect the introduction of significant new data or the discarding of older data. Older data are discarded in favor of newer data when it is felt that the newer data have smaller systematic errors, or have more checks on systematic errors, or have made corrections unknown at the time of the older experiments, or simply have much smaller errors. Sometimes, the scale

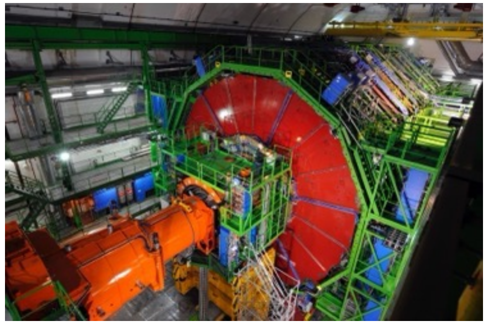
Yes, The PDG refers to uncertainties as "errors". 🤷🏻‍♂️

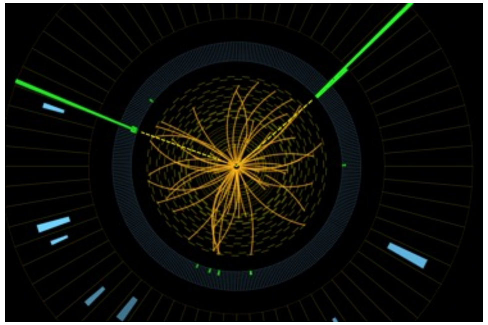
SYSTEMATIC UNCERTAINTIES AND NUISANCE PARAMETERS

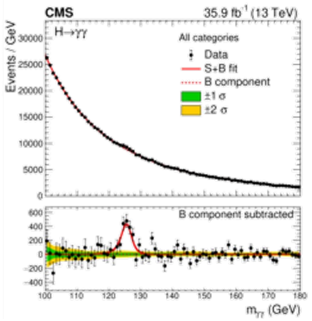
A. David (CERN and IST-Lisboa)

ANATOMY OF A LHC MEASUREMENT









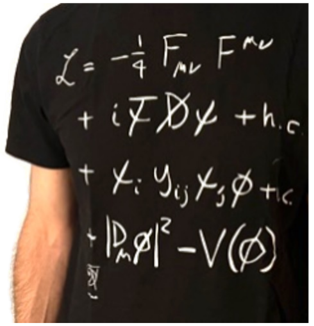
A theory
(e.g. the SM)

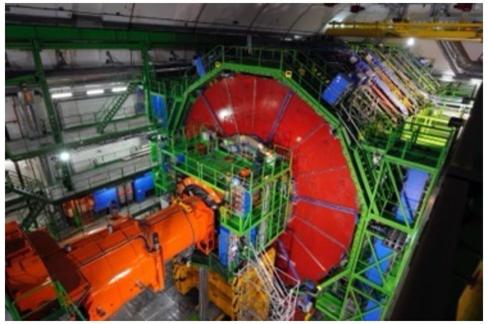
A detector
(with response that can be simulated)

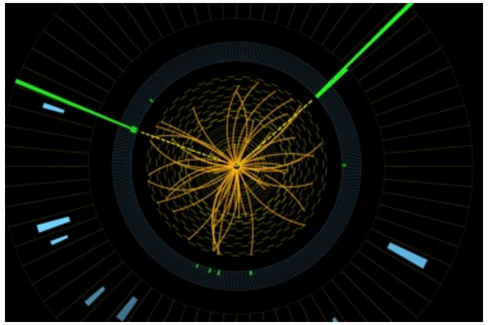
Event data

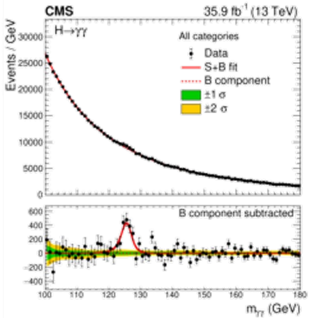
Summary data
("observables")

ANATOMY OF A LHC MEASUREMENT









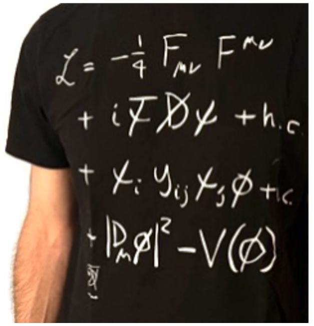
Simulation → **Reconstruction**

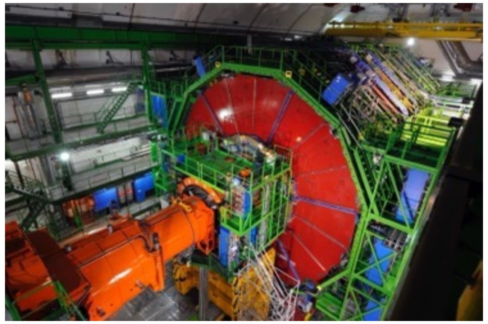
A theory (e.g. the SM)	A detector (with response that can be simulated)	Event data	Summary data (“observables”)

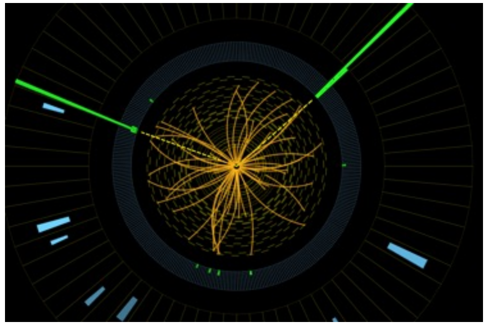
Simulation encompasses both theory and experiment aspects.

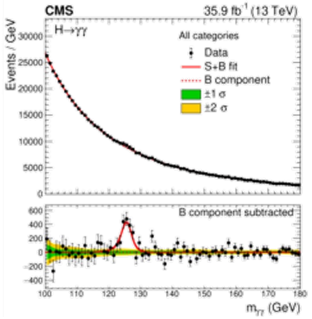
Reconstruction also includes any aggregation (binning) or transformation (machine learning, calibration).

ANATOMY OF A LHC MEASUREMENT









Simulation

Reconstruction

A theory
(e.g. the SM)

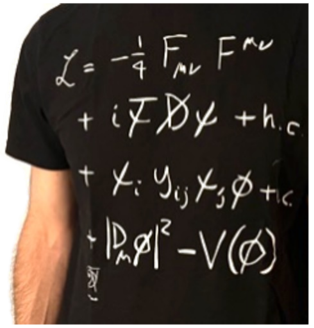
A detector
(with response that can be simulated)

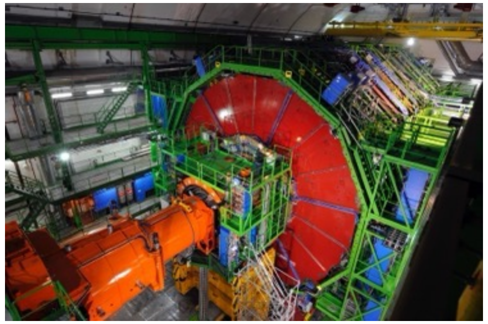
Event data

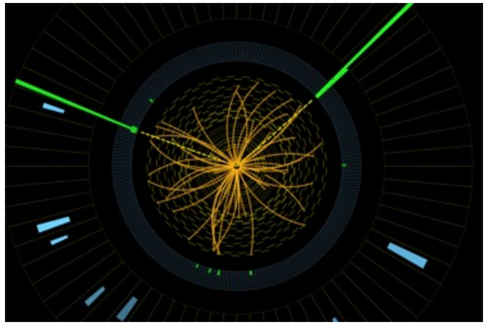
Summary data
("observables")

Inference (an inverse problem)

IN PRACTICE...



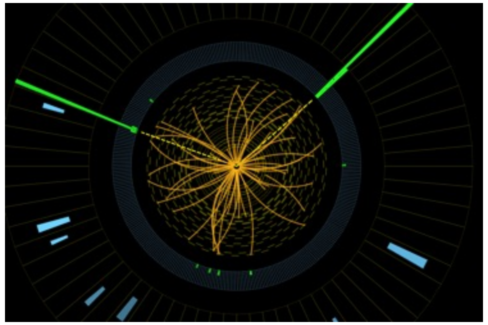


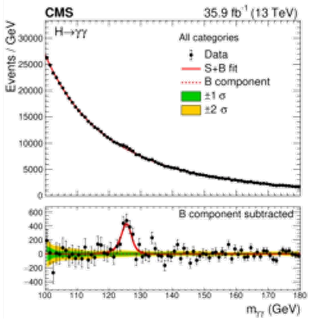


Simulation with imperfections (ν) using Monte Carlo sampling →

A theory (with parameters μ)	A detector $f(\, e \,;\, \mu, \nu \,)$	Event data $\{\, e \,\}$

IN PRACTICE...



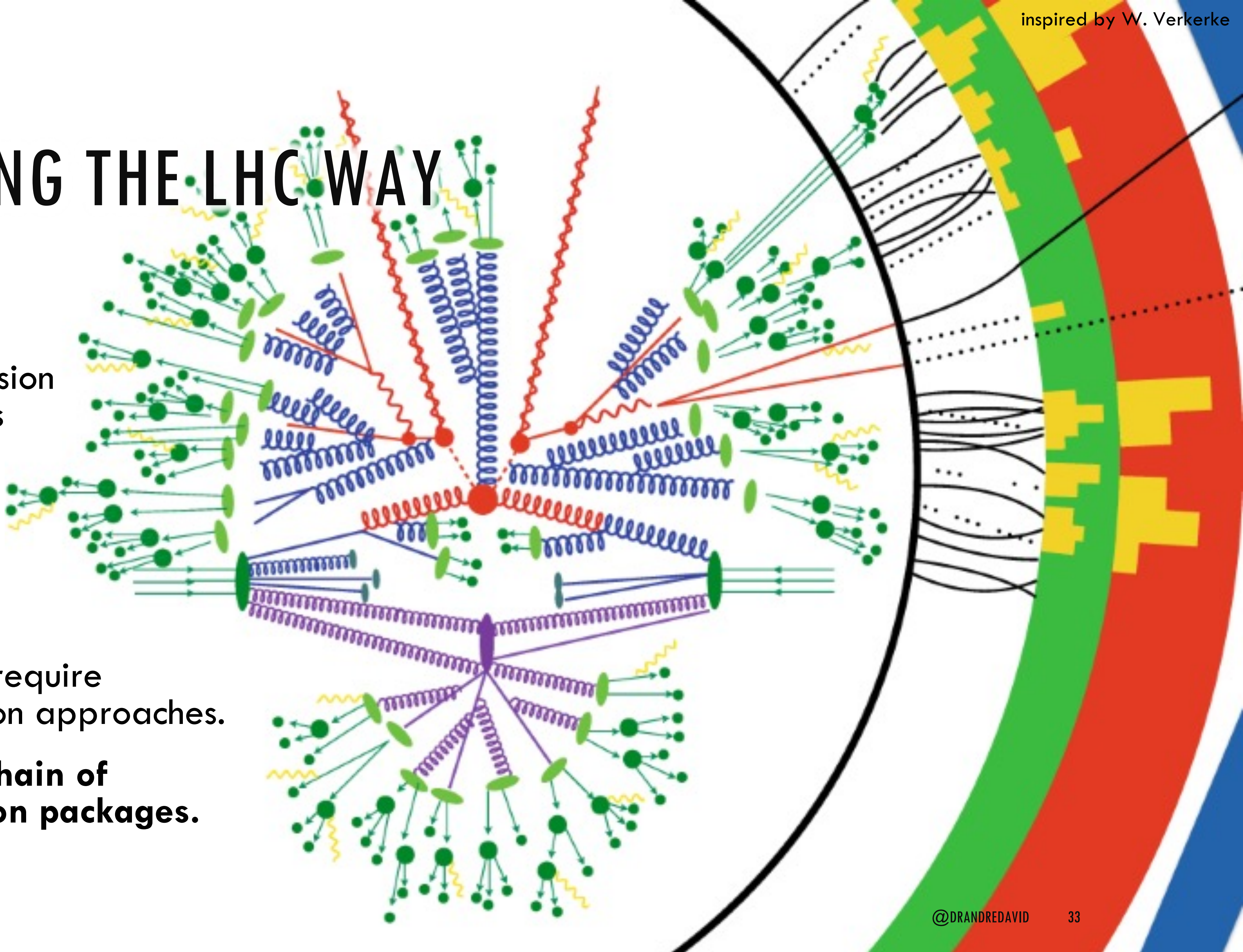


Imperfect reconstruction

Event data
{ e }

Summary data
{ x }

STEPS ALONG THE LHC WAY



Simulation of proton-proton collision at the LHC involves processes at many energy scales.

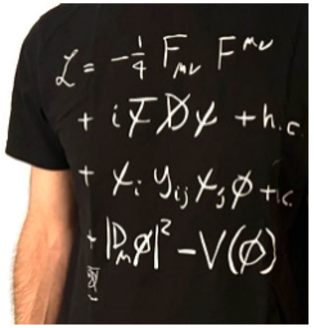
Different regimes require separate calculation approaches.

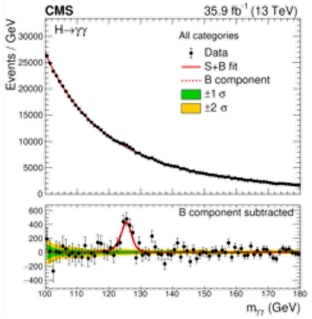
Implemented as chain of separate simulation packages.

IMPERFECTIONS ALONG THE LHC WAY

Ingredient	Estimate from…	Uncertainty estimates include…
Proton structure	Empirical; no first principles calculation.	Fit method and statistical uncertainties.
Matrix element calculation	Perturbative calculation from theory, resummed or fixed-order.	Missing higher orders.
Parton shower	Perturbative shower development.	Matching energy scale to Matrix element calculation.
{Hadroniz,Fragment}ation	MC simulation based on empirical models.	Tuneable parameters and different implementations.
Detector simulation	GEANT4 or Fast simulation.	GEANT tunes, parameterizations of digitisation.
Object reconstruction and identification	Custom-built algorithms (e, μ, τ, γ, b/c/q/g-jets, …).	Data-driven calibration uncertainties.

ANATOMY OF A MEASUREMENT





All the imperfections (ν)

A theory
(with parameters μ)

Summary data
{ x }

Inference on μ based on $p(x, y; \mu, \nu)$

🤔 What's y and where did that come from ?

THE STATISTICAL MODEL

The statistical model for inference is a function of the data given all parameters (Φ),

$$p(\text{data}; \vec{\Phi})$$

can be factorised into **primary data, x,** and **auxiliary observables, *y*.**
For *k* systematic uncertainties, each y_k is paired with a nuisance parameter ν_k:

$$p(\vec{x}, \vec{y}; \vec{\Phi}) = p(\vec{x}; \vec{\mu}, \vec{\nu}) \prod_k p_k(y_k; \nu_k)$$

where p_k are the probability distribution functions of the auxiliary observables.

⚠ *$\vec{x}$ is **one** data point (that can be multidimensional).*

THE STATISTICAL MODEL

The "counts" and their estimates' dependency on all parameters.

$$p(\vec{x},\vec{y};\vec{\Phi}) = p(\vec{x};\vec{\mu},\vec{\nu}) \prod_k p_k(y_k;\nu_k)$$

Information on how well any "nuisance" is known.

SIMPLE COUNTING EXPERIMENT

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin              ch1
observation      0
# ----------
bin              ch1
process          ppX
process          0
rate             1.47
```

- Single bin, zero observed events.
- **ppX** signal, 1.47 events expected.
  - $\mu = r$ is ppX rate modifier. $r = 1$ recovers theory prediction.

$$p(n, \vec{y}; r, \vec{v}) = \frac{\lambda(r,\vec{v})^n}{n!} \mathrm{e}^{-\lambda(r,\vec{v})}$$

$\rightarrow$ $n = 0$

$\rightarrow$ $\lambda(r, \vec{v}) = r\, 1.47$

SIMPLE COUNTING EXPERIMENT

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin             ch1
observation     0
# ----------
bin             ch1   ch1   ch1
process         ppX   WW    tt
process         0     1     2
rate            1.47  0.64  0.22
```

- Single bin, zero observed events.
- **ppX** signal.
  - $\mu = r$ is ppX rate modifier. $r = 1$ recovers theory prediction.
- WW and tt contributions with 0.64 and 0.22 events expected.

$$p(n,\vec{y};r,\vec{v}) = \frac{\lambda(r,\vec{v})^n}{n!} e^{-\lambda(r,\vec{v})}$$

→ $n = 0$

→ $\lambda(r,\vec{v}) = r\,1.47 + 0.22 + 0.64$

SIMPLE COUNTING EXPERIMENT

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin            ch1
observation    0
# ----------
bin            ch1   ch1   ch1
process        ppX   WW    tt
process        0     1     2
rate           1.47  0.64  0.22
# ----------
lumi     lnN   1.11  1.11   1.11
```

- Single bin, zero observed events.
- **ppX** signal.
  - ***μ* = *r*** is ppX rate modifier. $r = 1$ recovers theory prediction.
- **WW** and **tt** backgrounds.
- **lumi:** all processes from simulation and since $N = \sigma \mathcal{L}$ all yields affected by luminosity measurement uncertainty.

$$p(n, \vec{y}; r, \vec{\nu}) = \frac{\lambda(r,\vec{\nu})^n}{n!} e^{-\lambda(r,\vec{\nu})} \frac{1}{2\pi} e^{-(\nu_{\mathrm{lumi}} - y_{\mathrm{lumi}})^2}$$

→ $n = 0, y_{\mathrm{lumi}} = 0$

→ $\lambda(r, \vec{\nu}) = r\, 1.47\, (1.11)^{\nu_{\mathrm{lumi}}} + 0.22\, (1.11)^{\nu_{\mathrm{lumi}}} + 0.64\, (1.11)^{\nu_{\mathrm{lumi}}}$

SIMPLE COUNTING EXPERIMENT

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin             ch1
observation     0
# ----------
bin             ch1   ch1   ch1
process         ppX   WW    tt
process         0     1     2
rate            1.47  0.64  0.22
# ----------
lumi     lnN    1.11  1.11   1.11
xs       lnN    1.20   -      -
```

- Single bin, zero observed events.
- **ppX** signal.
  - ***μ* = *r*** is ppX rate modifier. $r = 1$ recovers theory prediction.
- **WW** and **tt** backgrounds.
- **lumi**: all processes affected by luminosity measurement uncertainty.
- **xs: ppX has theoretical uncertainty on cross-section (σ).**

$$p(n,\vec{y};r,\vec{\nu}) = \frac{\lambda(r,\vec{\nu})^n}{n!} e^{-\lambda(r,\vec{\nu})} \frac{1}{2\pi} e^{-(\nu_{\mathtt{lumi}}-y_{\mathtt{lumi}})^2} e^{-(\nu_{\mathtt{xs}}-y_{\mathtt{xs}})^2}$$

$\rightarrow$ $n = 0, y_{\mathtt{lumi}} = y_{\mathtt{xs}} = 0$

$\rightarrow$ $\lambda(r,\vec{\nu}) = r\, 1.47\, (1.11)^{\nu_{\mathtt{lumi}}} (1.2)^{\nu_{\mathtt{xs}}} + 0.22\, (1.11)^{\nu_{\mathtt{lumi}}} + 0.64\, (1.11)^{\nu_{\mathtt{lumi}}}$

SIMPLE COUNTING EXPERIMENT

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin             ch1
observation     0
# ----------
bin             ch1   ch1   ch1
process         ppX   WW    tt
process         0     1     2
rate            1.47  0.64  0.22
# ----------
lumi     lnN    1.11  1.11   1.11
xs       lnN    1.20   -      -
nWW      gmN 4   -    0.16    -
```

- Single bin, zero observed events.
- **ppX** signal.
  - $\mu = r$ is ppX rate modifier. $r = 1$ recovers theory prediction.
- **WW** and **tt** backgrounds.
- **lumi**: all processes affected by luminosity measurement uncertainty.
- **xs**: ppX has theoretical uncertainty on cross-section (σ).
- **nWW: WW yield estimated from 4 simulated events.**

$$p(n,\vec{y};r,\vec{\nu}) = \frac{\lambda(r,\vec{\nu})^n}{n!}e^{-\lambda(r,\vec{\nu})}\frac{1}{2\pi}e^{-(\nu_{\mathrm{lumi}}-y_{\mathrm{lumi}})^2}e^{-(\nu_{\mathrm{xs}}-y_{\mathrm{xs}})^2}\frac{(\nu_{\mathrm{nWW}})^{y_{\mathrm{nWW}}}}{y_{\mathrm{nWW}}!}e^{-\nu_{\mathrm{nWW}}}$$

→ $n = 0, y_{\mathrm{lumi}} = y_{\mathrm{xs}} = 0, \text{and } y_{\mathrm{nWW}} = 4$

→ $\lambda(r,\vec{\nu}) = r\,1.47\,(1.11)^{\nu_{\mathrm{lumi}}}(1.2)^{\nu_{\mathrm{xs}}} + 0.22\,(1.11)^{\nu_{\mathrm{lumi}}} + 0.64\,(1.11)^{\nu_{\mathrm{lumi}}}\frac{\nu_{\mathrm{nWW}}}{0.64}$

ℹ️ 0.64 yield ≡ 0.16 factor × 4 counts

SIMPLE COUNTING EXPERIMENT

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin            ch1
observation    0
# ----------
bin            ch1   ch1   ch1
process        ppX   WW    tt
process        0     1     2
rate           1.47  0.64  0.22
# ----------
lumi    lnN    1.11  1.11   1.11
xs      lnN    1.20   -      -
nWW     gmN 4   -    0.16    -
```

- Single bin, zero observed events.
- **ppX** signal.
  - $\mu = r$ is ppX rate modifier. $r = 1$ recovers theory prediction.
- **WW** and **tt** backgrounds.
- **lumi**: all processes affected by luminosity measurement uncertainty.
- **xs**: ppX has theoretical uncertainty on cross-section (σ).
- **nWW**: WW yield estimated from 4 simulated events.

$$p(n,\vec{y};r,\vec{\nu}) = \frac{\lambda(r,\vec{\nu})^n}{n!} e^{-\lambda(r,\vec{\nu})} \frac{1}{2\pi} e^{-(\nu_{\mathrm{lumi}}-y_{\mathrm{lumi}})^2} e^{-(\nu_{\mathrm{xs}}-y_{\mathrm{xs}})^2} \frac{(\nu_{\mathrm{nWW}})^{y_{\mathrm{nWW}}}}{y_{\mathrm{nWW}}!} e^{-\nu_{\mathrm{nWW}}}$$

ℹ️ Bin rate estimate λ does not depend on y_k, only ν_k !

→ $n = 0, y_{\mathrm{lumi}} = y_{\mathrm{xs}} = 0$, and $y_{\mathrm{nWW}} = 4$

→ $\lambda(r,\vec{\nu}) = r\,1.47\,(1.11)^{\nu_{\mathrm{lumi}}}(1.2)^{\nu_{\mathrm{xs}}} + 0.22\,(1.11)^{\nu_{\mathrm{lumi}}} + 0.64\,(1.11)^{\nu_{\mathrm{lumi}}}\frac{\nu_{\mathrm{nWW}}}{0.64}$

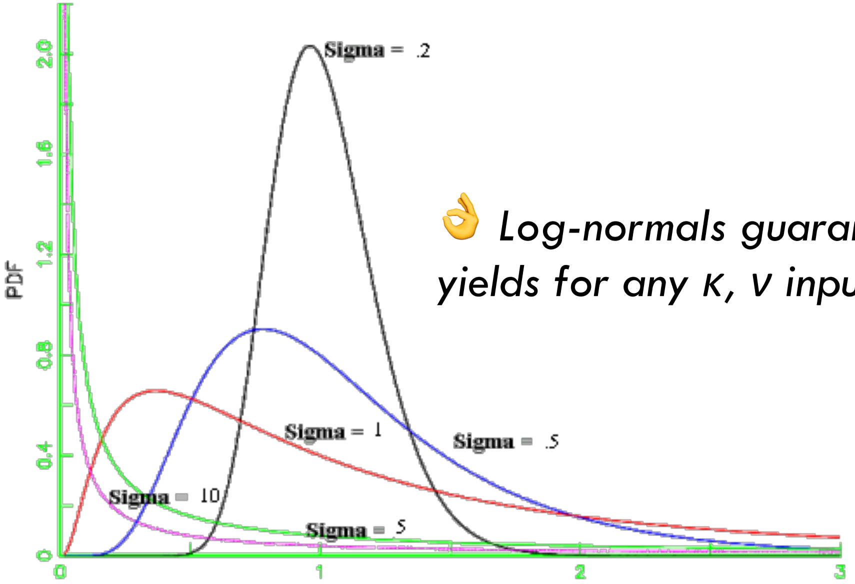
UNCERTAINTIES, EFFECTS, AND NUISANCES

Uncertainty type	Directive	Inputs	Multiplicative factor, $f(\nu)$	$p(y;\nu)$	Default values
Log-normal	`lnN`	`kappa`	κ^{ν}	$\mathcal{N}(y;\nu,1)$	$\nu = y = 0$

For a log-normal systematic **uncertainty that scales yields…**

… the **effect** on the process normalisation.

… the **probability distribution function** appended to the statistical model.



👌 *Log-normals guarantee positive yields for any κ, ν input values.*

You may also hear this p.d.f. referred to as "constraint term".
*This is because **we can include external constraints** to the statistical model by appending more terms here.*

UNCERTAINTIES, EFFECTS, AND NUISANCES

Uncertainty type	Directive	Inputs	Multiplicative factor, $f(\nu)$	$p(y;\nu)$	Default values
Log-normal	`lnN`	`kappa`	κ^{ν}	$\mathcal{N}(y;\nu,1)$	$\nu = y = 0$
Asymmetric log-normal	`lnN`	`kappaDown,` `kappaUp`	$\left(\kappa^{\text{Down}}\right)^{-\nu}$ if $\nu < -0.5$, $\left(\kappa^{\text{Up}}\right)^{\nu}$ if $\nu > 0.5$, $e^{\nu K\left(\kappa^{\text{Down}},\kappa^{\text{Up}},\nu\right)}$ otherwise.*	$\mathcal{N}(y;\nu,1)$	$\nu = y = 0$
Log-uniform	`lnU`	`kappa`	κ^{ν}	$\mathcal{U}(y,1/\kappa,\kappa)$	$\nu = y = \frac{1}{2}(\kappa + 1/\kappa)$
Gamma	`gmN`	`N, alpha`†	ν/N	$\mathcal{P}(y;\nu)$	$\nu = N+1, y = N$‡

*$K\left(\kappa^{\text{Down}},\kappa^{\text{Up}},\nu\right) = \frac{1}{8}\left[4\ln\left(\kappa^{\text{Up}}/\kappa^{\text{Down}}\right) + \ln\left(\kappa^{\text{Up}}\kappa^{\text{Down}}\right)\left(48\nu^5 - 40\nu^3 + 15\nu\right)\right]$ ensures that the multiplicative factor and its first and second derivatives are continuous for all values of ν, and reduces to a log-normal for $\kappa^{\text{Down}} = 1/\kappa^{\text{Up}}$.

†The rate value for the affected process must be equal to $N\alpha$.

‡The default value for the nuisance parameter is set to the mean of a gamma distribution with parameters $\kappa = N+1$, $\lambda = 1$, as defined in Ref. [20].

THE LIKELIHOOD FUNCTION

For *d* entries in the data set we tack on more “counts” terms to define the likelihood function:

$$\mathcal{L}(\vec{\Phi}) = \prod_{d} p(\vec{x}_d; \vec{\mu}, \vec{\nu}) \prod_{k} p_k(y_k; \nu_k)$$

THE LIKELIHOOD FUNCTION

For *d* entries in the data set we tack on more “counts” terms to define the likelihood function:

$$\mathcal{L}(\vec{\Phi}) = \prod_{d} p(\vec{x}_d; \vec{\mu}, \vec{\nu}) \prod_{k} p_k(y_k; \nu_k)$$

But what is this good for ?

THE LIKELIHOOD FUNCTION

For *d* entries in the data set we tack on more "counts" terms to define the likelihood function:

$$\mathcal{L}(\vec{\Phi}) = \prod_d p(\vec{x}_d; \vec{\mu}, \vec{\nu}) \prod_k p_k(y_k; \nu_k)$$

1. **Frequentist** inference: **ν profiled** in a likelihood ratio:

$$q_{\text{LEP}}(\mu) = -2\ln\left(\frac{\mathcal{L}(\mu=0,\vec{\nu}_0)}{\mathcal{L}(\mu,\vec{\nu}_0)}\right) \qquad q_{\text{TEV}}(\mu) = -2\ln\left(\frac{\mathcal{L}(\mu=0,\hat{\hat{\vec{\nu}}}(0))}{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}\right)$$

$$\tilde{q}_{\text{LHC}}(\mu) = \begin{cases} -2\ln\left(\frac{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\vec{\nu}})}\right) & \text{if } 0 \leq \hat{\mu} \leq \mu, \\ -2\ln\left(\frac{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(0,\hat{\hat{\vec{\nu}}}(0))}\right) & \text{if } \hat{\mu} < 0, \\ 0 & \text{if } \hat{\mu} > \mu, \end{cases}$$

2. **Bayesian** inference: ***ν_k* marginalised/averaged** over their priors π_k:

$$\mathcal{L}_{\text{int}}(\vec{\mu}) = \int \mathcal{L}(\vec{\Phi}) \prod_k \pi_k(\nu_k) d\vec{\nu}, \quad p_k(\nu_k|y_k) \propto p_k(y_k;\nu_k)\pi_k(\nu_k)$$

FREQUENTIST INFERENCE WITHOUT NUISANCES

Statistical methodology in particle physics is (very) predominantly **frequentist**.

Notion of **coverage** is central to define uncertainties (68%, 95%).

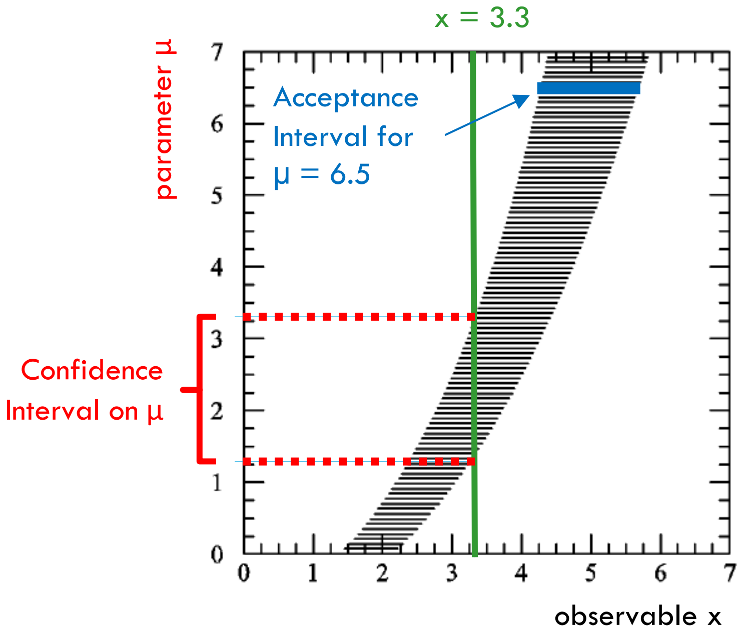
Computational procedures for frequentist methodology quite different from Bayesian: *influences practical aspects of how systematics uncertainties are modelled.*

30-second nutshell reminder of Frequentist approach:

- Observations **{ x }** summarized by test statistic **q(μ)**, typically a **likelihood ratio** testing for compatibility of the data with a certain hypothesis $\mu = \mu_0$.
- Knowing the distribution of **q(μ)** under given hypotheses $\mu = \mu_i$ define a **acceptance interval** that contains 68% of the observed outcomes.
- A confidence belt maps the acceptance interval for each value of μ, and allows to construct a **confidence interval** in **μ** for a given observed value of **q(μ)**.

FREQUENTIST UNCERTAINTIES IN HEPP

Single measurement interval inversion



Neyman construction of the confidence belt

Acceptance intervals defined by

$$P(x_{low} < x < x_{high}; \mu) = \int_{x_{low}}^{x_{high}} p(x;\mu)\, dx \geq 1 - \alpha$$

where $\mathbf{1 - \alpha}$ is the **confidence level.**

ℹ️ Procedure in a nutshell:

1. For a given μ generate distribution of x, p(x; μ).
2. Use p(x; μ) to determine x_{low} and x_{high} and make horizontal line.
   - NB: acceptance interval depends on 1-α choice and can be one-sided (for limits).
3. Repeat for many values of μ to construct the belt.
4. For a given x = 3.3 look up the confidence interval for μ from the belt.

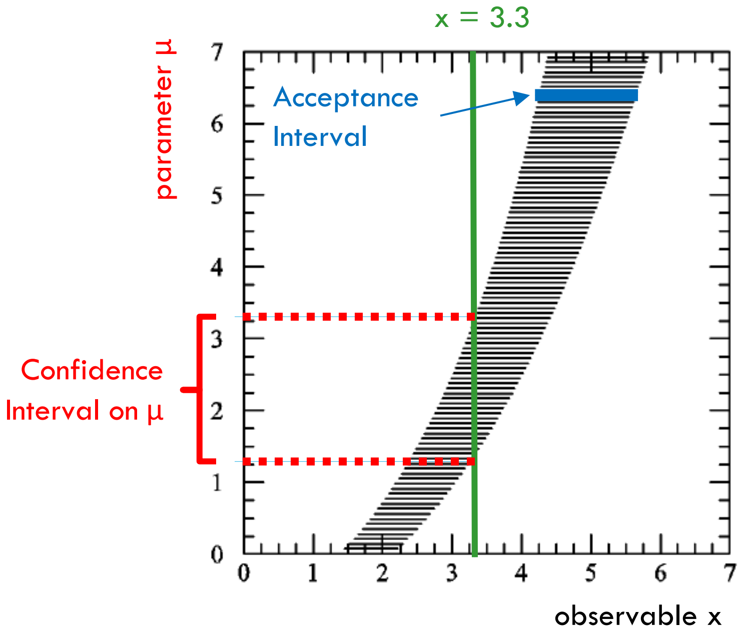
(Detailed step-by-step in backup.)

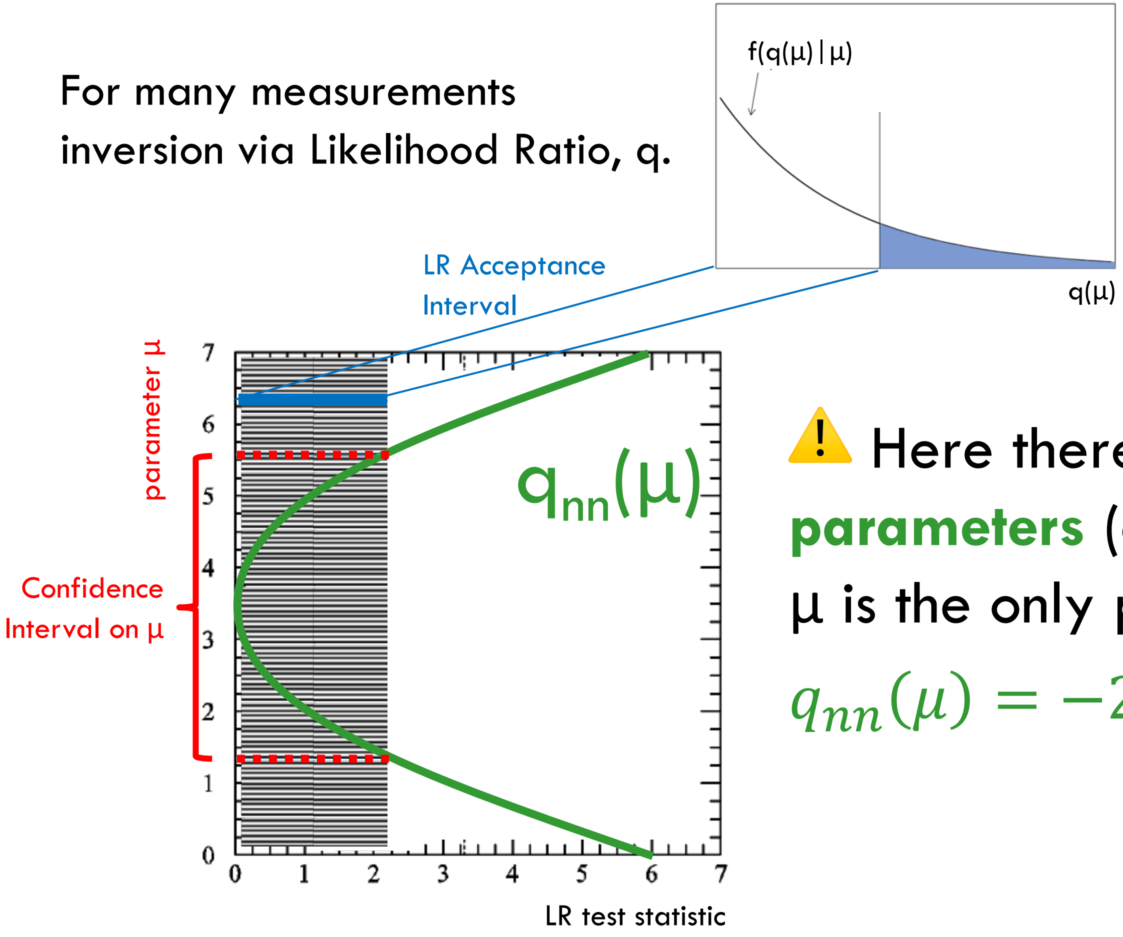
FREQUENTIST UNCERTAINTIES IN HEPP

Single measurement interval inversion

For many measurements inversion via Likelihood Ratio, q.

More on the CLs criterion in backup.



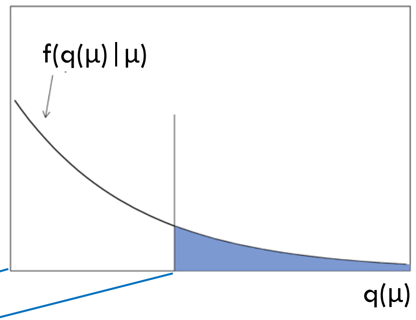


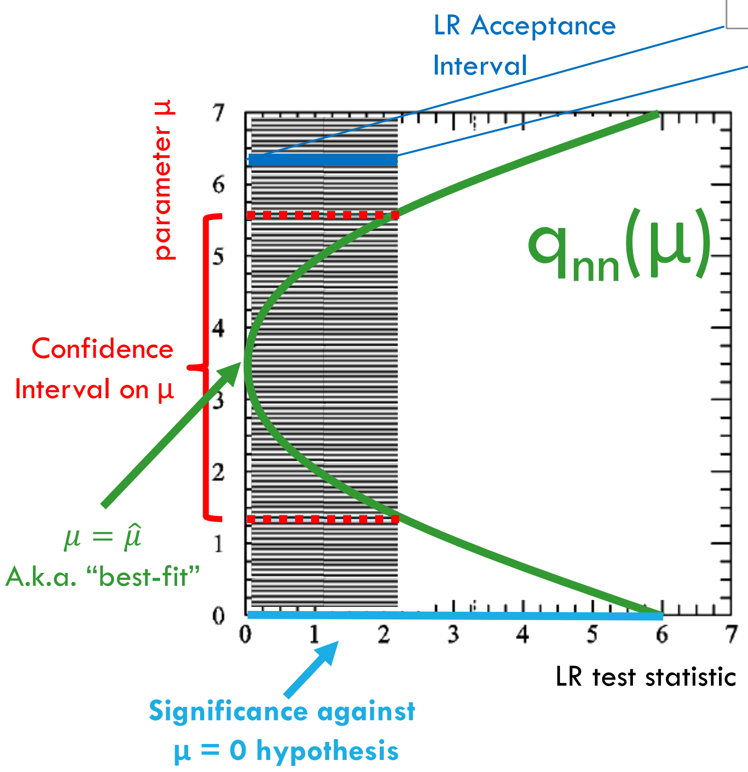
⚠ Here there are **no nuisance parameters** (cf. D. van Dyk's 2.3). μ is the only parameter and

$$q_{nn}(\mu) = -2\ln\frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})}$$

FREQUENTIST UNCERTAINTIES IN HEPP

For many measurements inversion via Likelihood Ratio, q.

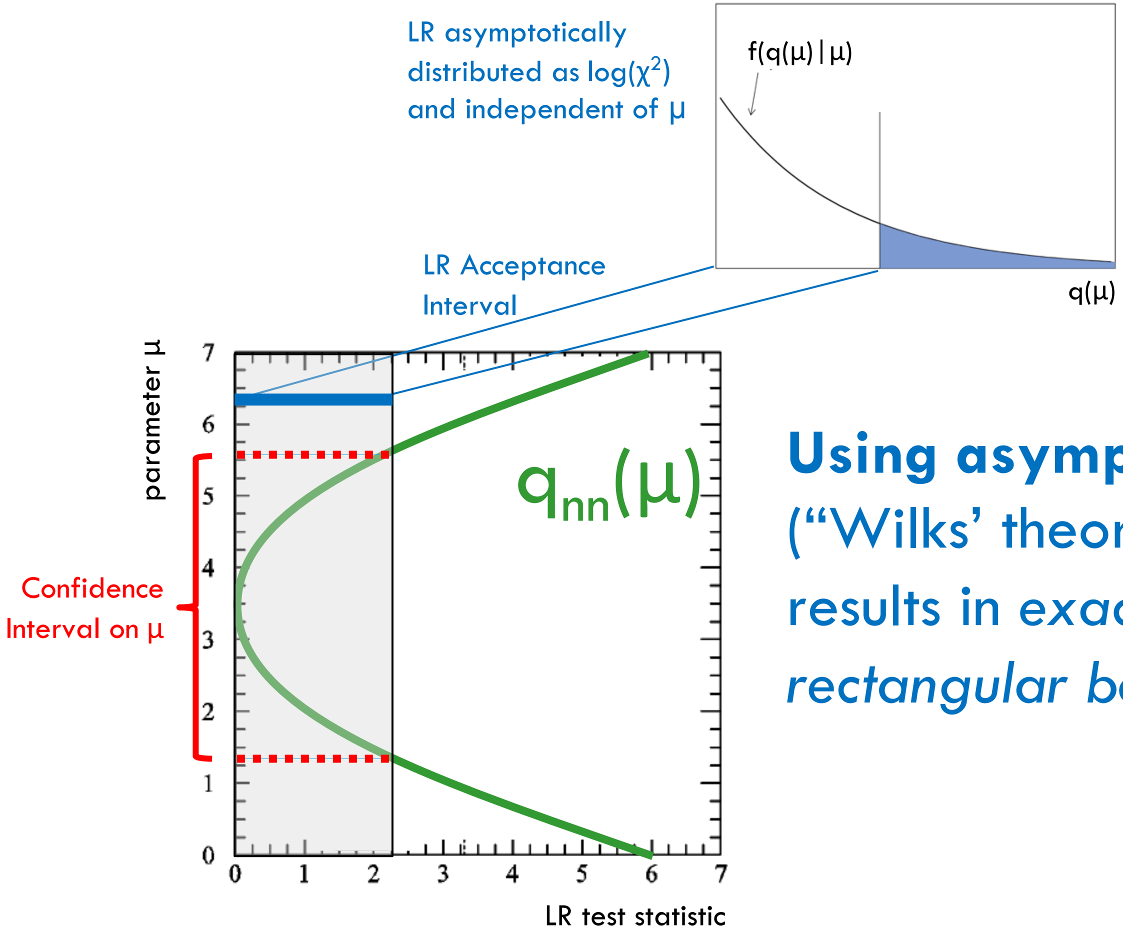




Here there are **no nuisance parameters** (cf. D. van Dyk's 2.3). μ is the only parameter and

$$q_{nn}(\mu) = -2\ln\frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})}$$

ASYMPTOTIC APPROXIMATION



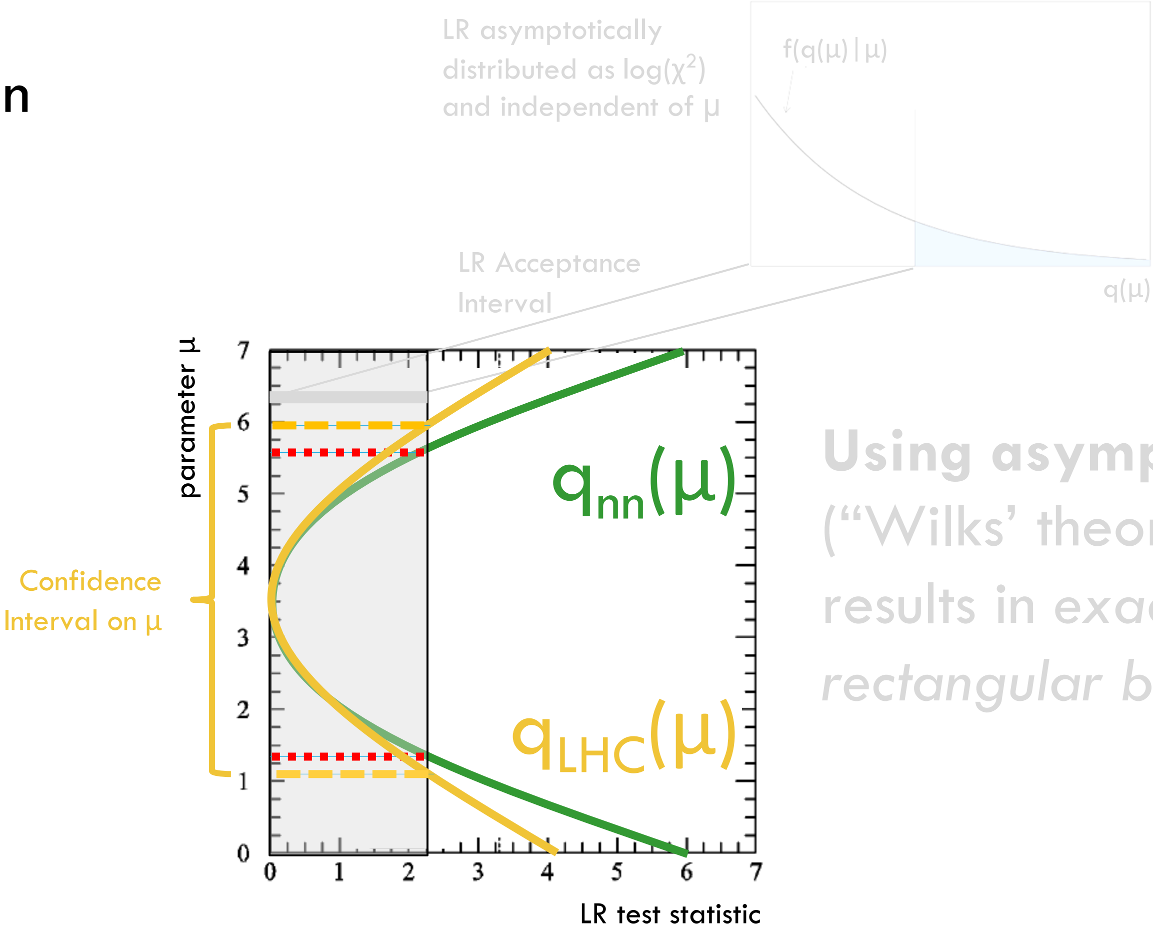
Using asymptotics
("Wilks' theorem")
results in *exactly*
rectangular belt

ADDING NUISANCES

Including nuisance parameters in **profile likelihood ratio** **broadens the interval**

$$q_{nn}(\mu) = -2\ln\frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})}$$

$$\tilde{q}_{LHC}(\mu) = -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\vec{\nu}})}$$



Using asymptotics ("Wilks' theorem") results in *exactly rectangular belt*

ADDING NUISANCES

Including nuisance parameters in **profile likelihood ratio** **broadens the interval**

The workhorse of inference with nuisances at the LHC.

- Confidence regions, confidence intervals, and upper limits.
- Beware caveats concerning the asymptotic approximation.

Computing $\tilde{q}_{LHC}(\mu)$ is **relatively cheap** even when dimension of $\vec{\nu}$ is large.

- No practical penalty to introduce many nuisance parameters.
- Many LHC analyses have 10^2 to 10^3.
- Combined CMS+ATLAS analyses can reach 10^4.

$$\tilde{q}_{LHC}(\mu) = -2 \ln \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\vec{\nu}})}$$

$\hat{\vec{\nu}}$ is the overall best-fit value of $\vec{\nu}$, i.e. when $\mu = \hat{\mu}$.

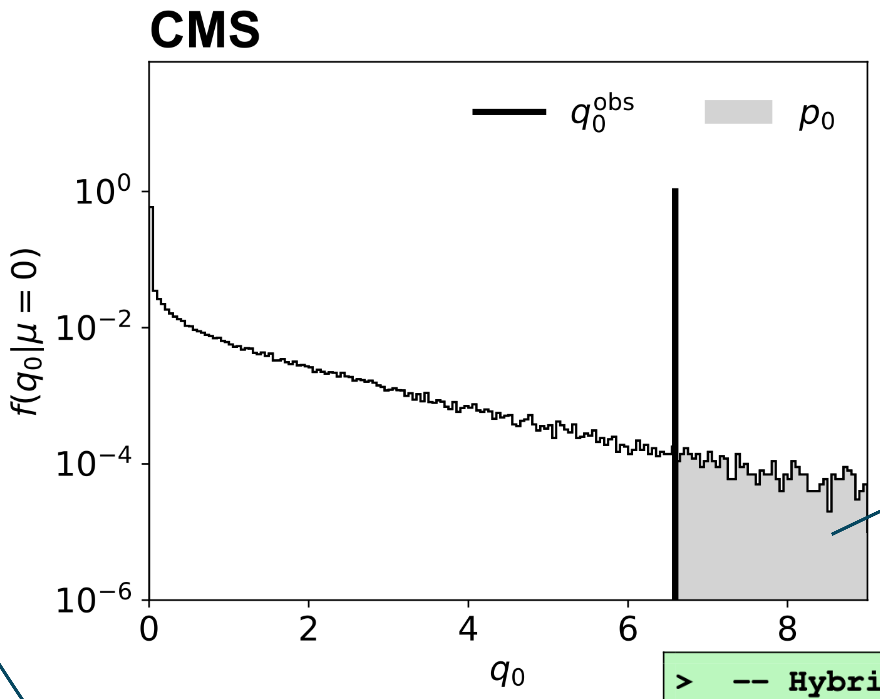
$\hat{\hat{\vec{\nu}}}(\mu)$ is the best-fit value of $\vec{\nu}$ for a specific value of μ.

THE $\mu = 0$ CASE – "SIGNIFICANCE"

How unlikely are the data under the $\mu = 0$ hypothesis?

- Asymptotics also available.

$$q_0 = \tilde{q}_{LHC}(\mathbf{0}) = -2\ln\frac{\mathcal{L}(\mathbf{0},\hat{\hat{\vec{\nu}}}(\mathbf{0}))}{\mathcal{L}(\hat{\mu},\hat{\vec{\nu}})}$$



$$p_0 = \int_{q_0^{obs}}^{\infty} f(q_0|0)\,\mathrm{d}q_0$$

```
>  -- Hybrid New --
> Significance: 2.54397  -0.0146063/+0.015170
> Null p-value: 0.00548 +/- 0.000233452
> Done in 5.95 min (cpu), 7.57 min (real)
```

```
>  -- Significance --
> Significance: 2.56729
> Done in 0.00 min (cpu), 0.00 min (real)
```

SIMPLE COUNTING EXPERIMENT – SIGNIFICANCE

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin             ch1
observation     0
# ----------
bin             ch1   ch1   ch1
process         ppX   WW    tt
process         0     1     2
rate            1.47  0.64  0.22
# ----------
lumi    lnN     1.11  1.11   1.11
xs      lnN     1.20   -      -
nWW     gmN 4    -    0.16    -
```

```
silverlining-851.local
y.cern.ch/cms-cloud/combine-standalone:latest          2.6m < Tue May 1
t$ combine data/tutorials/CAT23001/datacard-2-template-analysis.txt -M S

California & Stanford University
            All rights reserved, please read http://roofit.sourceforge.net/license.txt

 <<< Combine >>>
 <<< v9.2.1 >>>
>>> Random number generator seed is 123456
>>> Method used is Significance

RooFit v3.60 -- Developed by Wouter Verkerke and David Kirkby
            Copyright (C) 2000-2013 NIKHEF, University of California & Stanford University
            All rights reserved, please read http://roofit.sourceforge.net/license.txt


 -- Significance --
Significance: 0
Done in 0.01 min (cpu), 0.01 min (real)
(base) mambauser@130e6b27ce0e:/code/HiggsAnalysis/CombinedLimit$
```

Makes sense !

SIMPLE COUNTING EXPERIMENT – BEST-FIT

```
~                                                                    silverlining-851.local
! ~  docker run --rm -it --platform linux/amd64 gitlab-registry.cern.ch/cms-cloud/combine-standalone:latest
(base) mambauser@f02cb7968bcb:/code/HiggsAnalysis/CombinedLimit$ combine data/tutorials/CAT23001/datacard-2-template-analysis.txt -M M
ultiDimFit

RooFit v3.60 -- Developed by Wouter Verkerke and David Kirkby
                Copyright (C) 2000-2013 NIKHEF, University of California & Stanford University
                All rights reserved, please read http://roofit.sourceforge.net/license.txt

 <<< Combine >>>
 <<< v9.2.1 >>>
>>> Random number generator seed is 123456
>>> Method used is MultiDimFit

RooFit v3.60 -- Developed by Wouter Verkerke and David Kirkby
                Copyright (C) 2000-2013 NIKHEF, University of California & Stanford University
                All rights reserved, please read http://roofit.sourceforge.net/license.txt

Doing initial fit:

 --- MultiDimFit ---
best fit parameter values:
   r :    +0.000
Done in 0.01 min (cpu), 0.01 min (real)
(base) mambauser@f02cb7968bcb:/code/HiggsAnalysis/CombinedLimit$
```

Makes sense !

SIMPLE COUNTING EXPERIMENT – LIMITS

```
~ ~                                                                silverlining-851.local
~ docker run --rm -it --platform linux/amd64 gitlab-registry.cern.ch/cms-cloud/combine-standalone:latest     3.46h  Mon May 1
3 15:44:24 2024
(base) mambauser@350bb2cf0367:/code/HiggsAnalysis/CombinedLimit$ combine data/tutorials/CAT23001/datacard-1-counting-experiment.txt

RooFit v3.60 -- Developed by Wouter Verkerke and David Kirkby
                Copyright (C) 2000-2013 NIKHEF, University of California & Stanford University
                All rights reserved, please read http://roofit.sourceforge.net/license.txt

 <<< Combine >>>
 <<< v9.2.1 >>>
>>> Random number generator seed is 123456
>>> Method used is AsymptoticLimits

RooFit v3.60 -- Developed by Wouter Verkerke and David Kirkby
                Copyright (C) 2000-2013 NIKHEF, University of California & Stanford University
                All rights reserved, please read http://roofit.sourceforge.net/license.txt


 -- AsymptoticLimits ( CLs ) --
Observed Limit: r < 1.6516
Expected  2.5%: r < 0.9613
Expected 16.0%: r < 1.4257
Expected 50.0%: r < 2.3438
Expected 84.0%: r < 4.0440
Expected 97.5%: r < 6.8621

Done in 0.01 min (cpu), 0.01 min (real)
(base) mambauser@350bb2cf0367:/code/HiggsAnalysis/CombinedLimit$
```

Runs the full asymptotic frequentist machinery for 95% CL upper limit on rate modifier on both expected and observed data.

HALF-TIME

📏 **All imperfections should give rise to an uncertainty.**

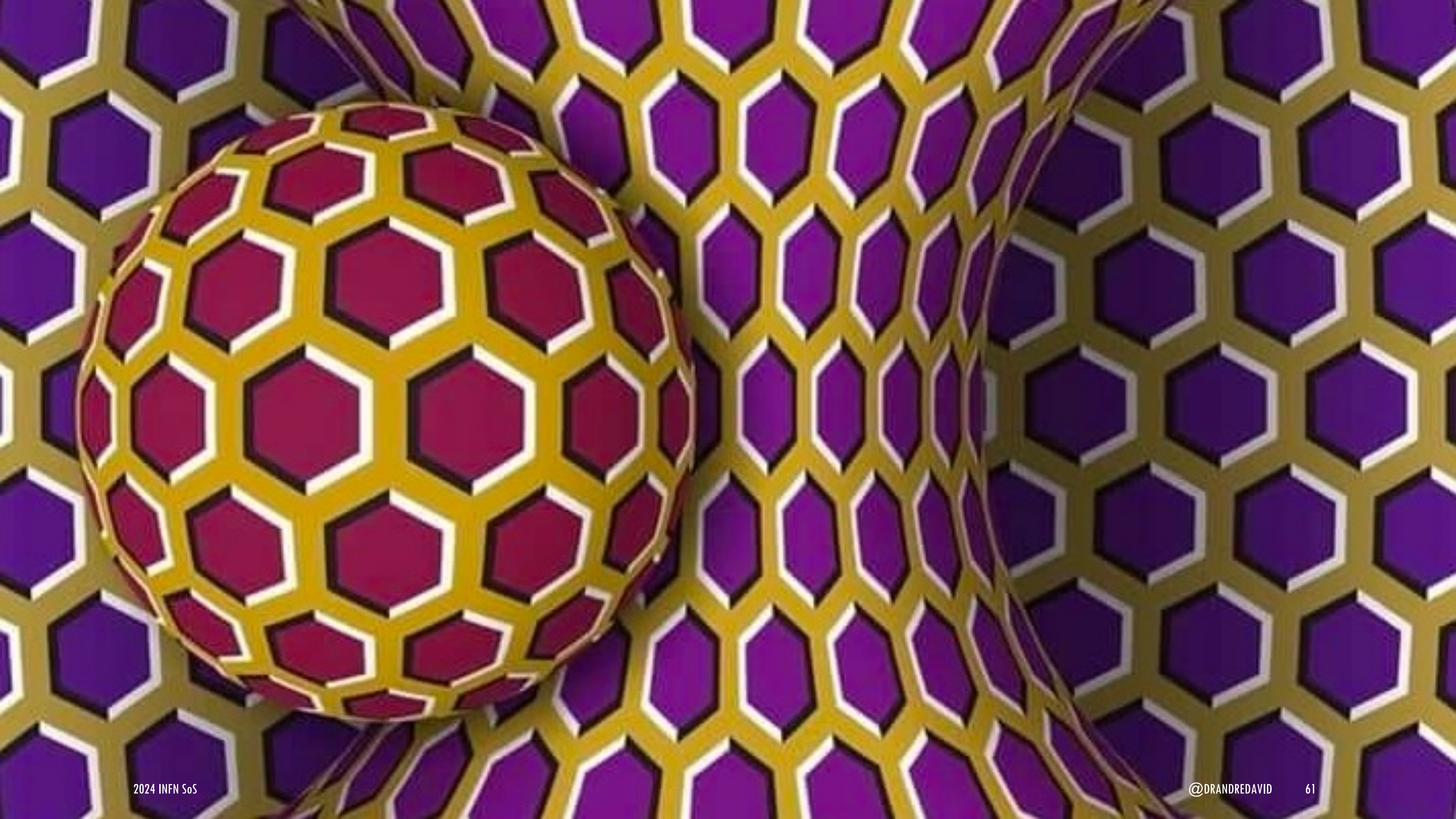
- Some are more important than others.
- Some are important to one inference but not another.

📚 **There are decades of practice on how to include them in the statistical model.**

- Make use of the tools that already exist.

⛓️‍💥 **Both Frequentist and Bayesian methods fail in some cases.**

- Make sure to understand their limitations and strengths.



ONE LISTING OF NUISANCE PARAMETERS

Luminosity

Detector and per-particle type

- Acceptance
- Efficiency and misidentification
- Energy scales
- Energy resolutions

Templates of processes:

- Theory total cross section uncertainty
- Theory modelling uncertainties
- Limited MC statistics

Empirical process shape modelling

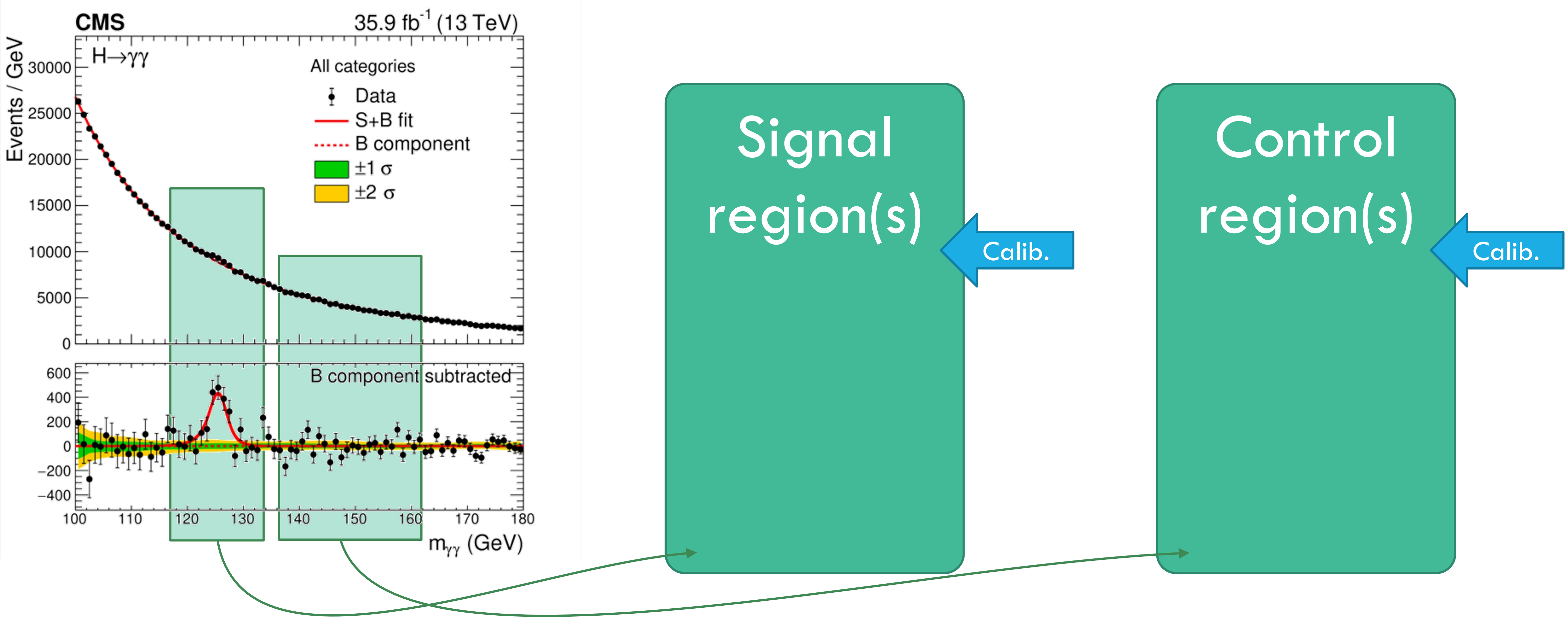
- Parameterisations
- Non-parametric smoothing
- Morphing of templates

Nuisance parameters can be constrained by:

- Detector calibration data
- Control samples with different event selection
- The data distributions
- Measurements from other experiments
- Theory calculations

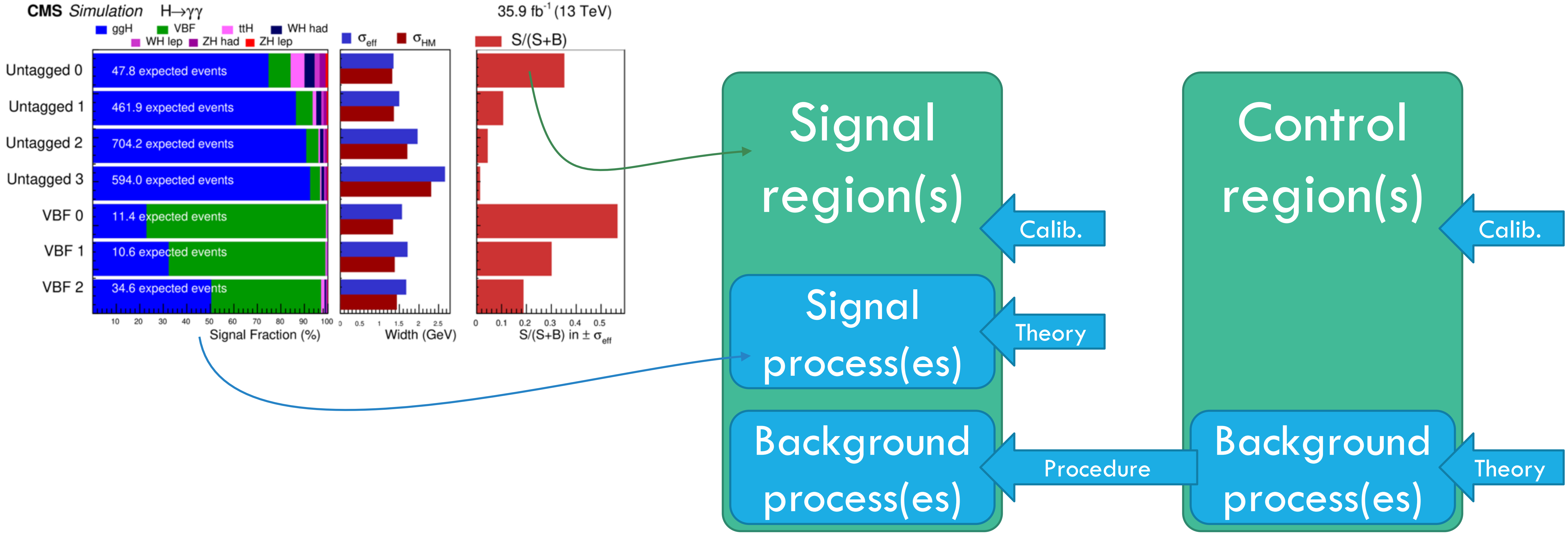

Physicists
"Counts"
Systematics and Nuisances

ERRORS – NUISANCES – COUNTS



arXiv:2002.06398

ERRORS – NUISANCES – COUNTS



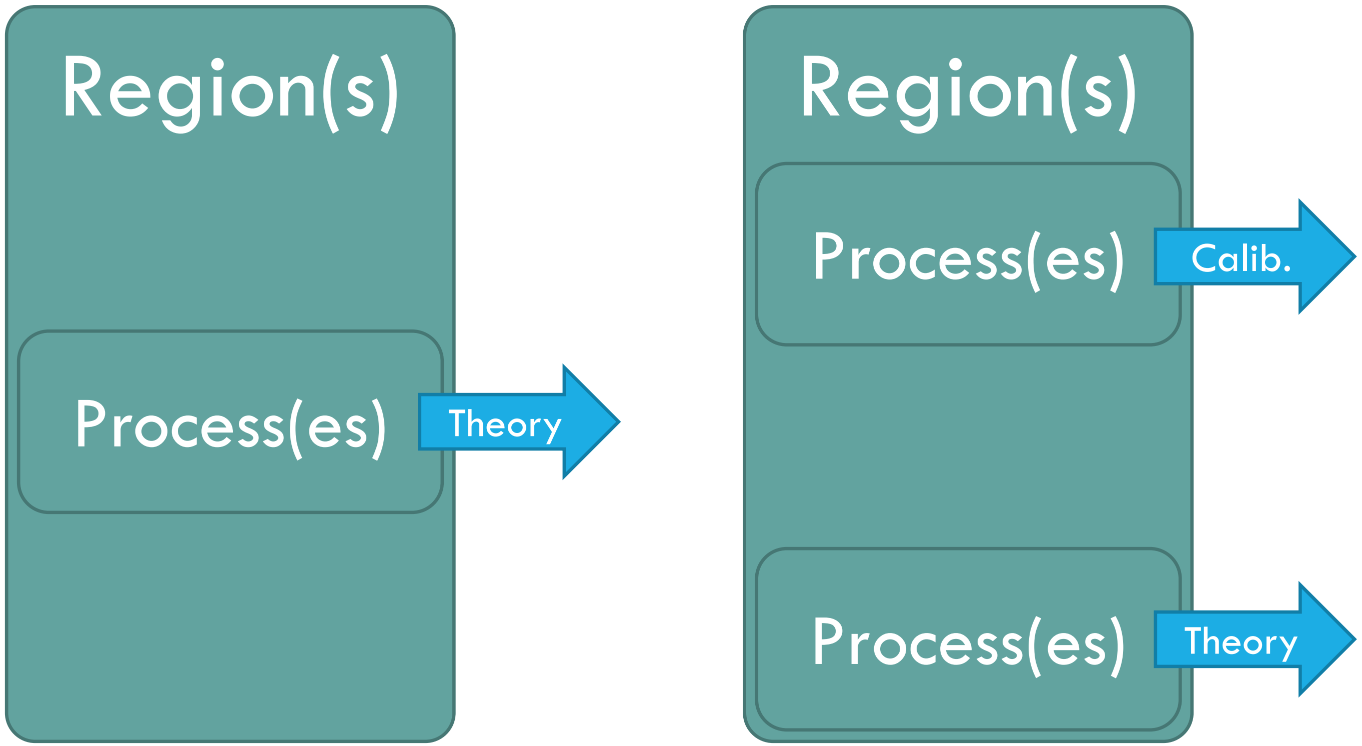
A REVERSIBLE PROCESS

Data also used to:

- Calibrate.
- Constrain theory parameters.
- Constrain non-perturbative inputs.
  - Perennial concern that parton distribution function fits may subsume BSM physics effects.

Same events ≠
(Double-counting = Double-dipping)

- Avoiding circularity always in the back of our minds.



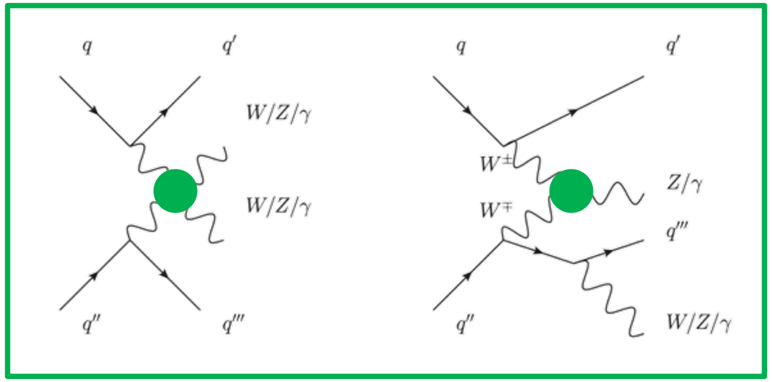
BEYOND S AND B – PROCESSES

Quantum-mechanically indistinguishable example.

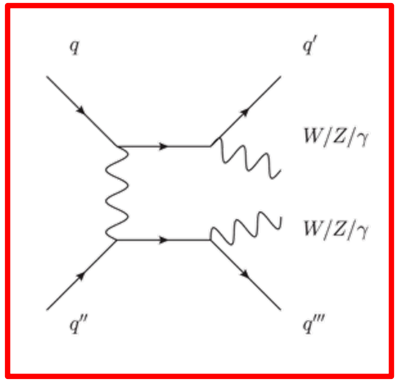
- Use interference as systematic?
- Avoid interfering phase space?
- Estimate effects on the total?
- …

Generally-speaking there are:

- Processes sensitive to the inference you want to make.
- Processes that are not.
  - Some you can estimate from MC.
  - Others may be better estimated from data.
  - Many have an impact on the power of your inference.
- Detector limitations (like noise).



More interesting



Less interesting

or $\mu^+\mu^-$, a photon, and two jets are selected. The electroweak component is measured with observed and expected significances of 4.1 standard deviations. The fiducial cross-section for electroweak production is measured to be $\sigma_{Z\gamma jj-\text{EW}} = 7.8 \pm 2.0$ fb, in good agreement with the Standard Model prediction.

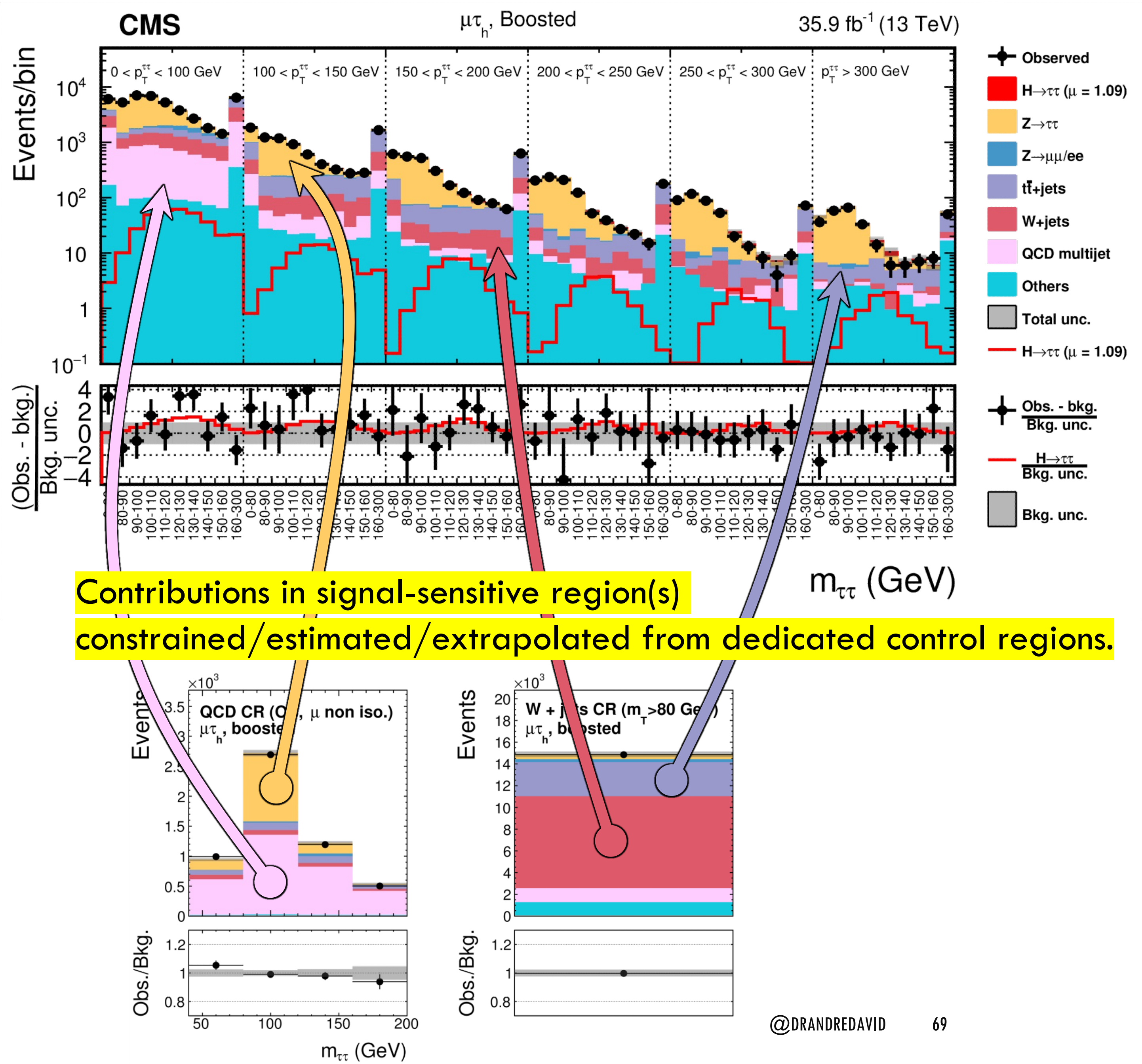
MadGraph5_aMC@NLO 2.3.3 MC cross-section prediction in the fiducial region ($\sigma^{\text{fid., MC}}_{Z\gamma jj-\text{EW}}$). Because the effect of interference between the $Z\gamma jj$–QCD and the $Z\gamma jj$–EW processes is not accounted for in the $Z\gamma jj$–QCD contribution, the observed cross-section $\sigma^{\text{fid.}}_{Z\gamma jj-\text{EW}}$ formally corresponds to electroweak production plus the interference effects.

CONTRIBUTIONS, NOT CONTAMINATION

Simultaneous inference accounts for all processes in all regions.

Pure regions not as important as independent ones.

- **Covering similar kinematics minimises extrapolation systematic uncertainties !**



Contributions in signal-sensitive region(s) constrained/estimated/extrapolated from dedicated control regions.

THE ASIMOV DATASET

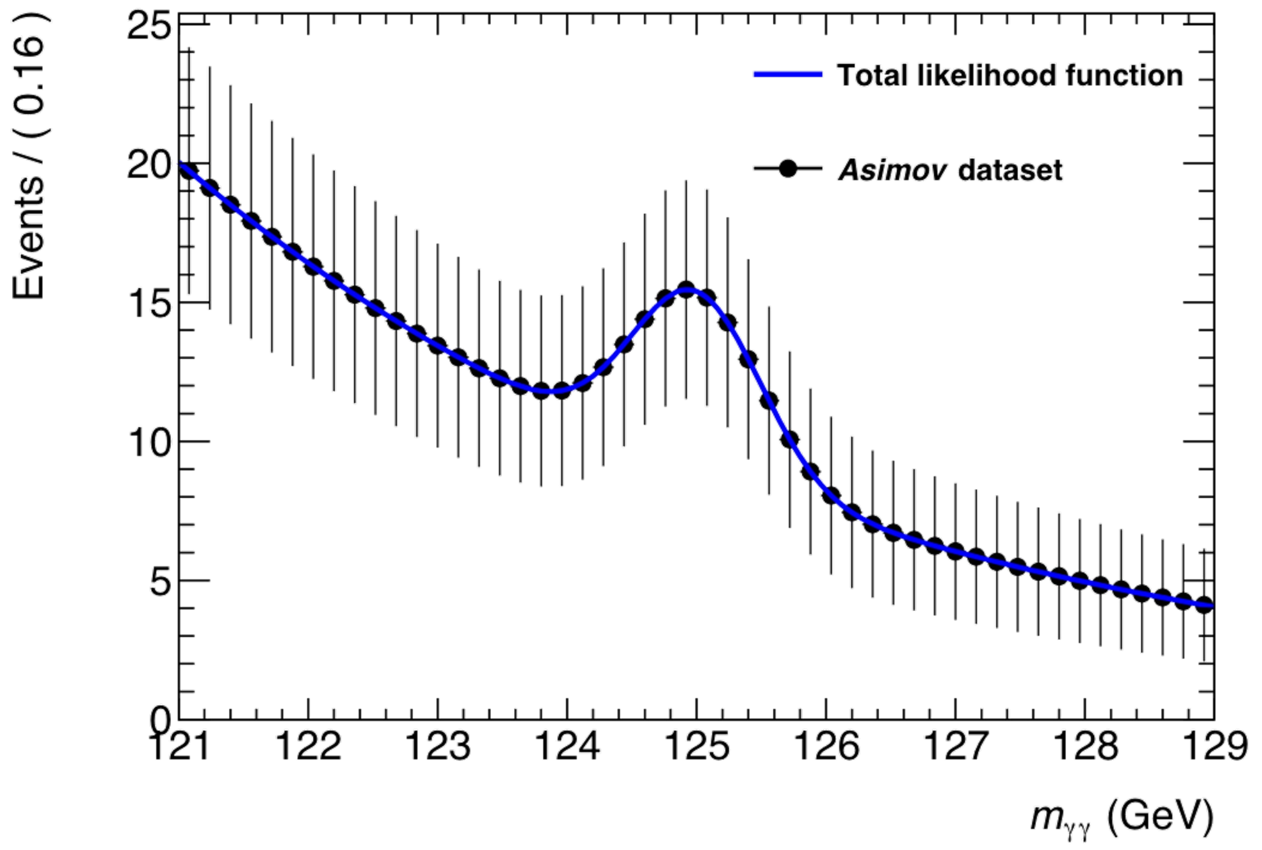
Procedure

- Fix all parameters of the **model**.
- Estimate corresponding **expected counts**.

Zero statistical fluctuations.

- Crucial property to explore the model's power without (Nature's) randomness added.

First used by CCGV for median significance and inspired by I. Asimov's "Franchise" short story.



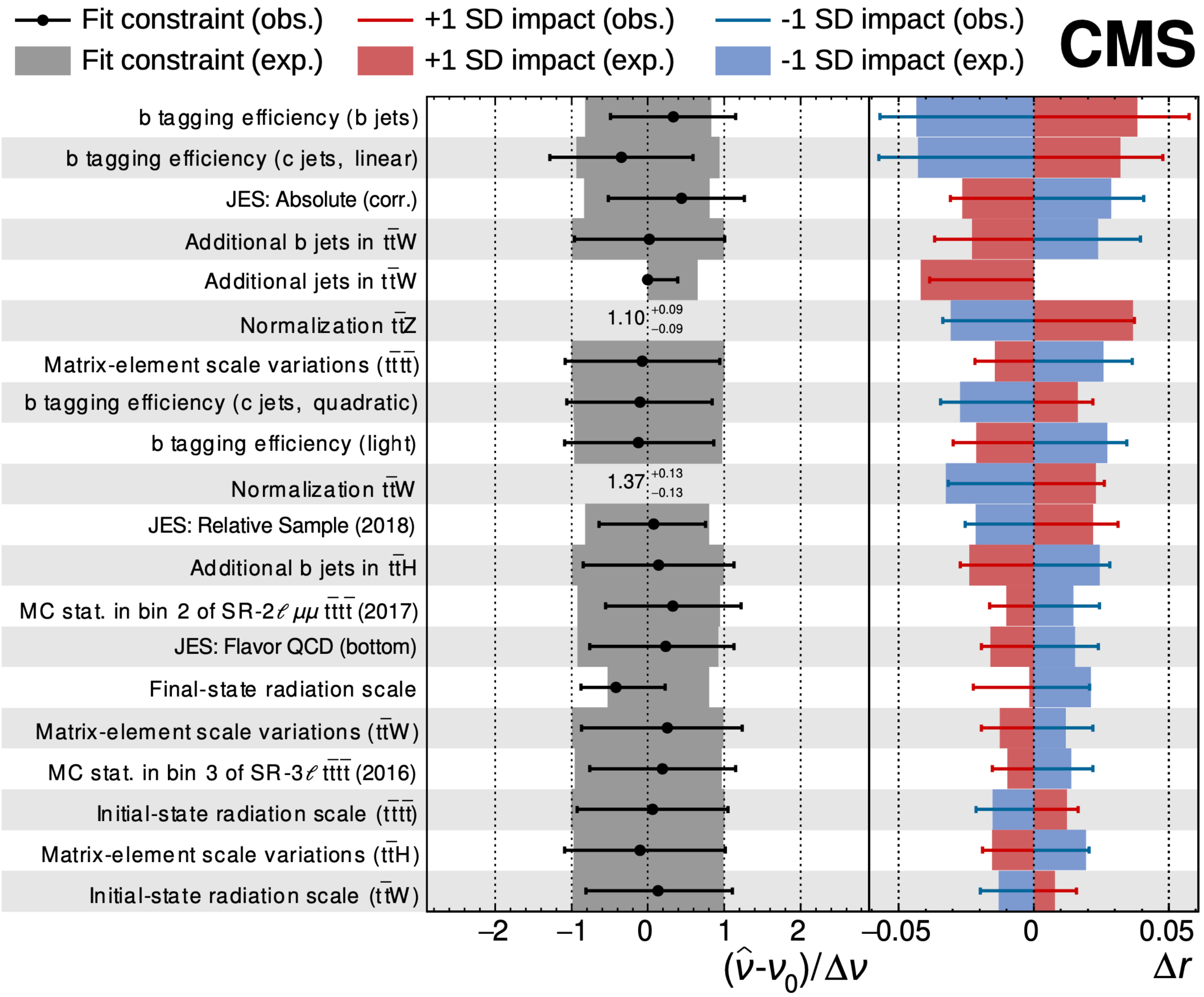
arXiv:2204.06614

PULLS, CONSTRAINTS, IMPACTS

Essential diagnostics that the tools can produce for you.

They cannot tell you whether you are missing something in your statistical model.

They provide insight on how the statistical model and the data interact and affect your inference.



arXiv:2204.06614

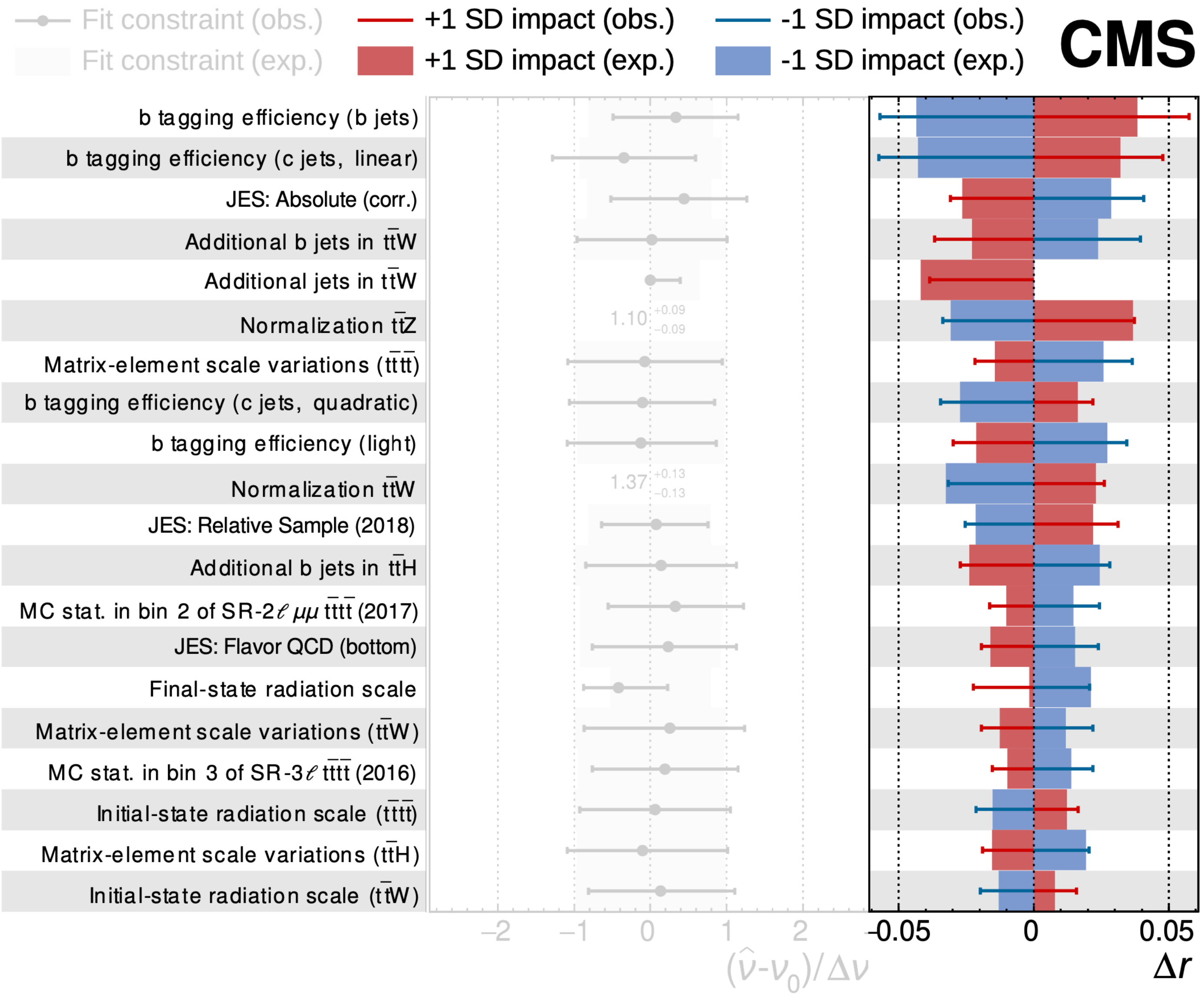
PULLS, CONSTRAINTS, IMPACTS

How much does μ change when changing ν ?

- Fix ν_i to $\widehat{\nu_i} \pm \Delta\nu_i$ and refit all other parameters $(\mu, \nu_{k\neq i})$.
  - $\Delta\nu$ typically chosen as 1 standard deviation value.
- Impact of $\hat{\nu} + \Delta\nu$ in red.
- Impact of $\hat{\nu} - \Delta\nu$ in blue.

Two different cases shown:

- **Full boxes** are the impact *after* the fit to the Asimov dataset ("expected").
  - ⚠️ "expected" does not imply "correct"; it's just a reflection of what to expect with this choice of statistical model and observables.
- **Lines** show the impact after the fit to collision data ("observed").



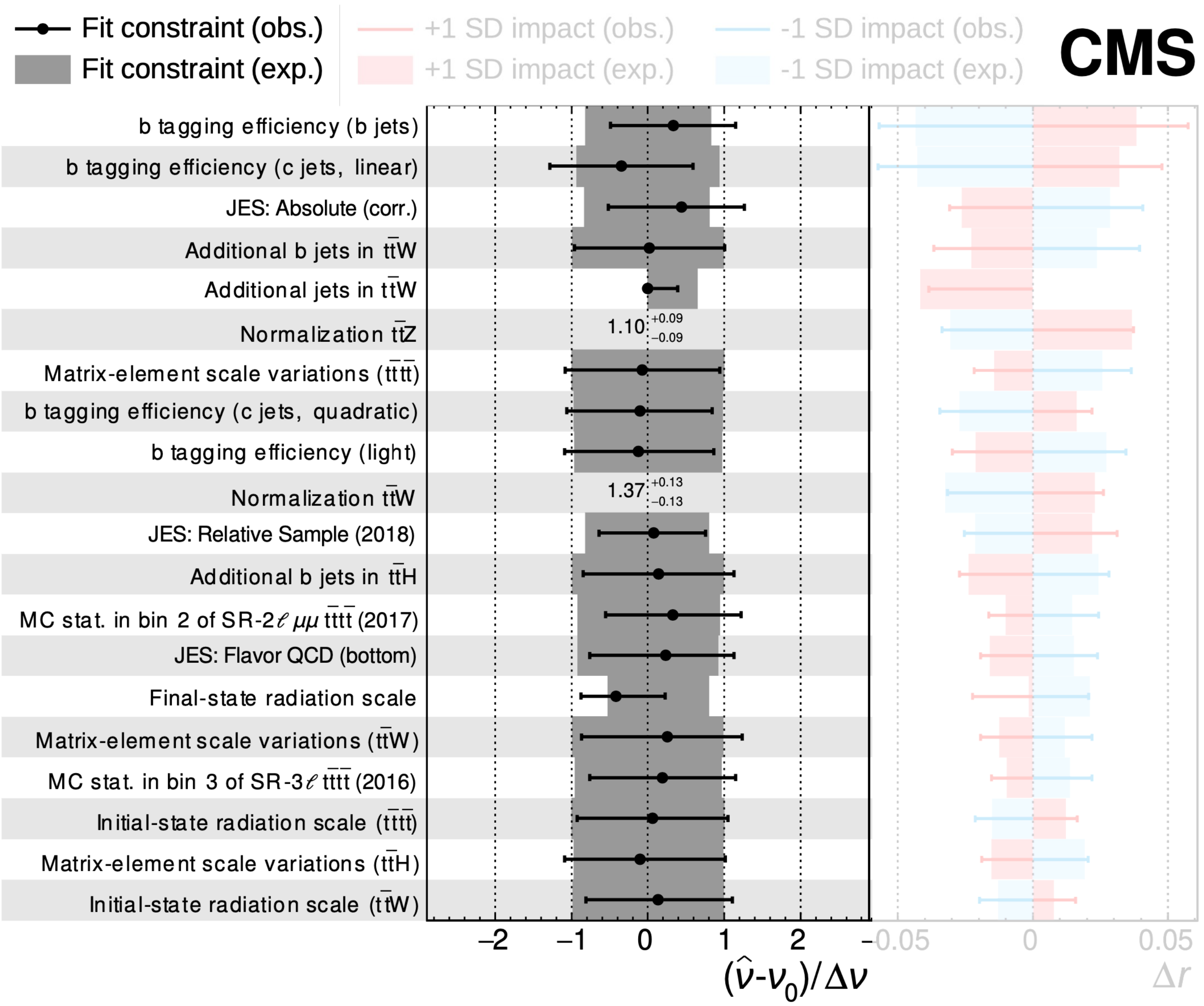
arXiv:2204.06614

PULLS, CONSTRAINTS, IMPACTS

How much information is gained on each ν ?

Three things shown:

- ± 1 represents what you put in the model.
- **Grey bars** are the uncertainty *after* the fit to the Asimov dataset ("expected").
  - ⚠️ "expected" does not imply "correct"; it's just a reflection of what to expect with this choice of statistical model and observables.
- **Black uncertainties** are the uncertainty after the fit to collision data ("observed").

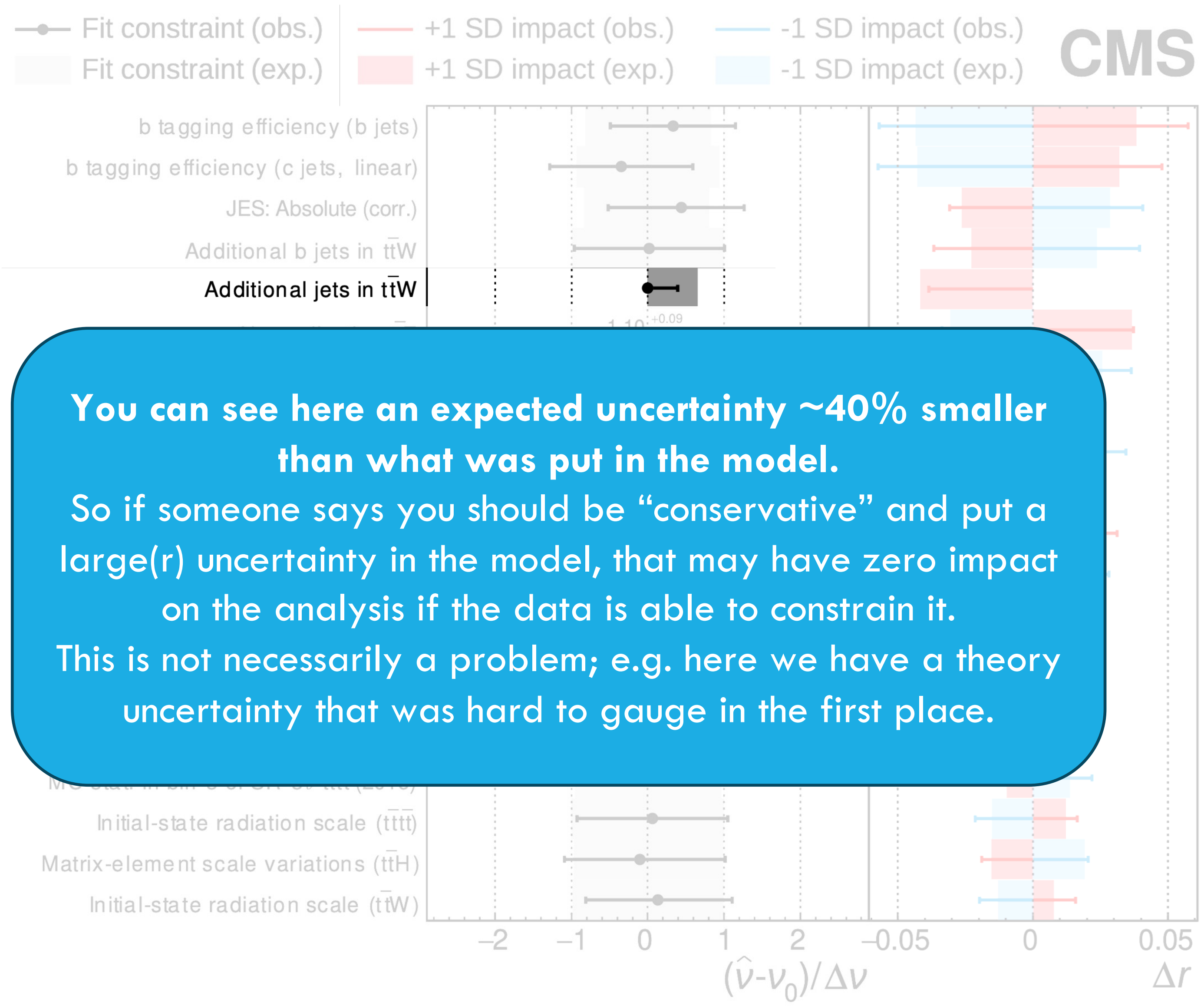


PULLS, CONSTRAINTS, IMPACTS

How much information is gained on each ν ?

Three things shown:

- ± 1 represents what you put in the model.
- **Grey bars** are the uncertainty *after* the fit to the Asimov dataset ("expected").
  - ⚠️ "expected" does not imply "correct"; it's just a reflection of what to expect with this choice of statistical model and observables.
- **Black uncertainties** are the uncertainty after the fit to collision data ("observed").



You can see here an expected uncertainty ~40% smaller than what was put in the model.

So if someone says you should be "conservative" and put a large(r) uncertainty in the model, that may have zero impact on the analysis if the data is able to constrain it.

This is not necessarily a problem; e.g. here we have a theory uncertainty that was hard to gauge in the first place.

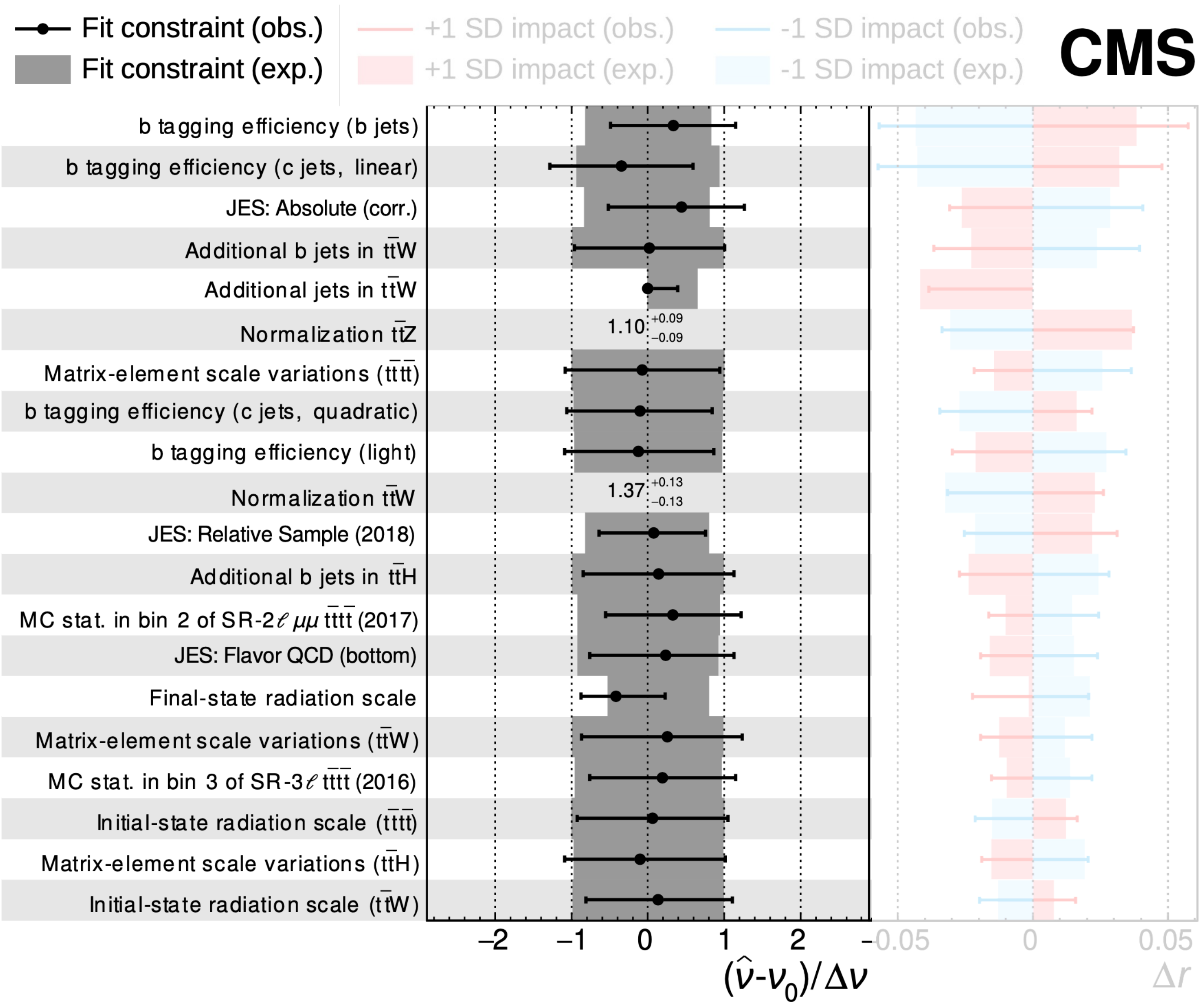
arXiv:2204.06614

PULLS, CONSTRAINTS, IMPACTS

How much information is gained on each ν ?

Different datasets tell us different things:

- Pre-fit expected (Asimov)
  - "What power is this model expected to have ?"
- Pre-fit toy data sets
  - "What do statistical fluctuations imply for that power ?"
- Post-fit expected (Asimov)
  - From partial data, e.g. only (some) CRs.
  - From whole data: CRs + SRs.
- Post-fit observed
  - Partial data or whole data.
  - Include statistical fluctuations.
- Post-fit toy data sets
  - "How unlucky was the observed ?"

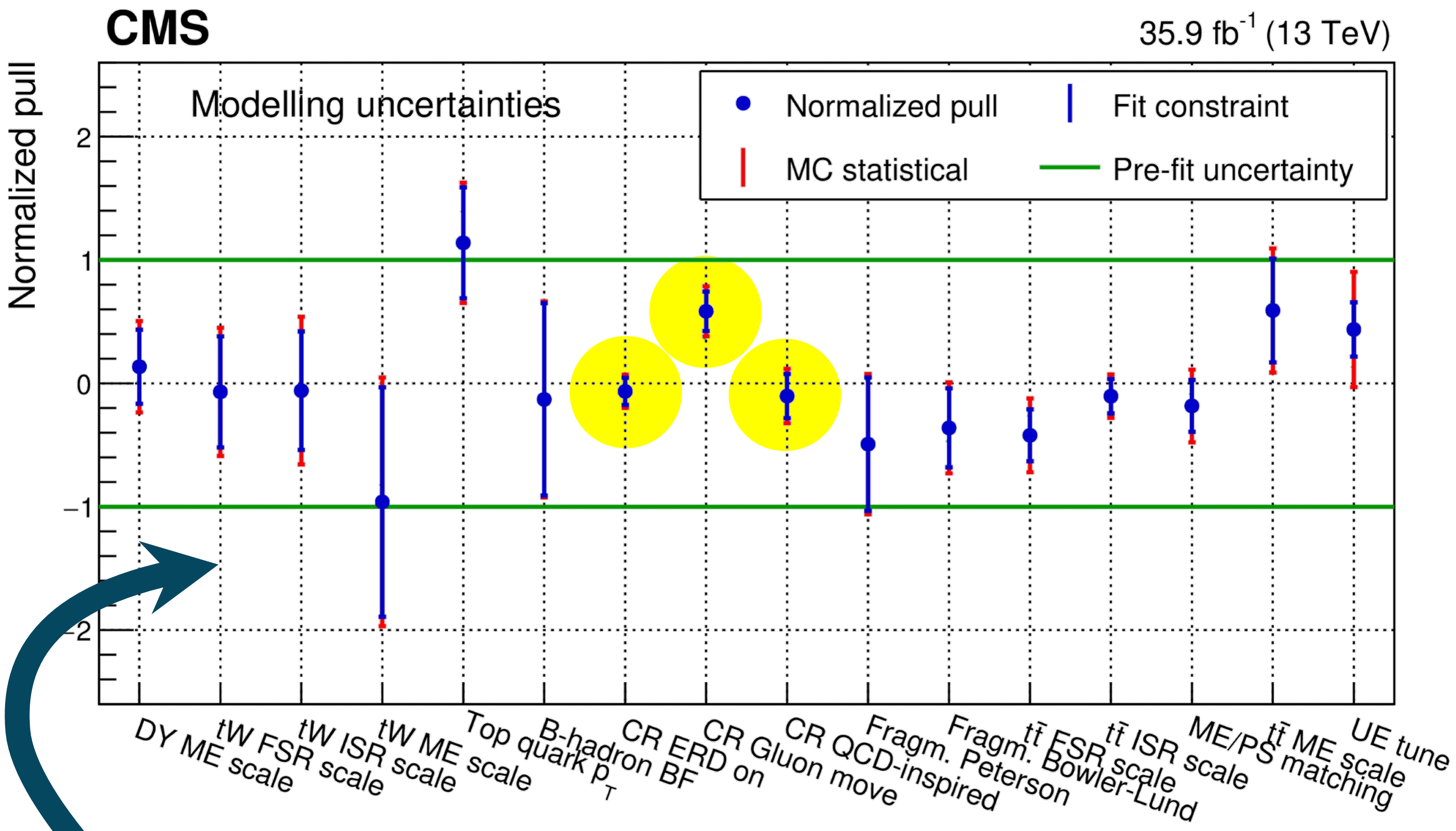


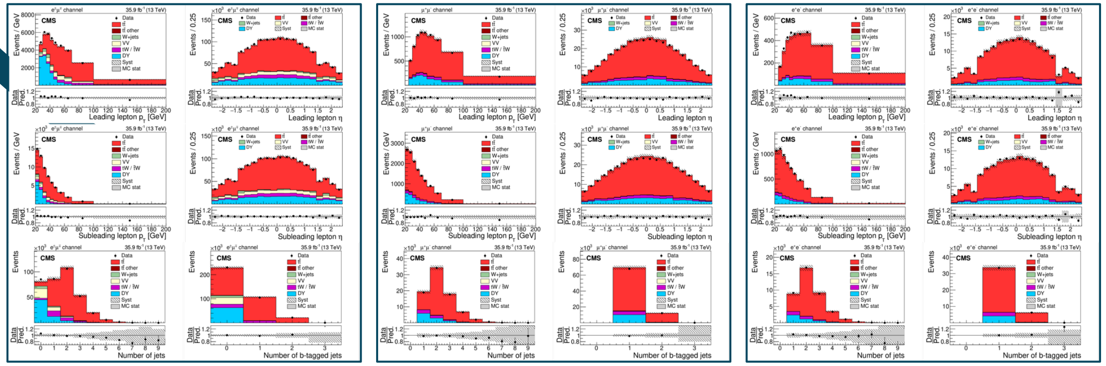
CONSTRAINING NUISANCES ?

From $t\bar{t}$ cross-section measurement: $e^{\pm}\mu^{\mp}$, $\mu^{+}\mu^{-}$, and $e^{+}e^{-}$.

“effects of **colour reconnection (CR)** processes on the top quark final state”

Is constraining these ok ?



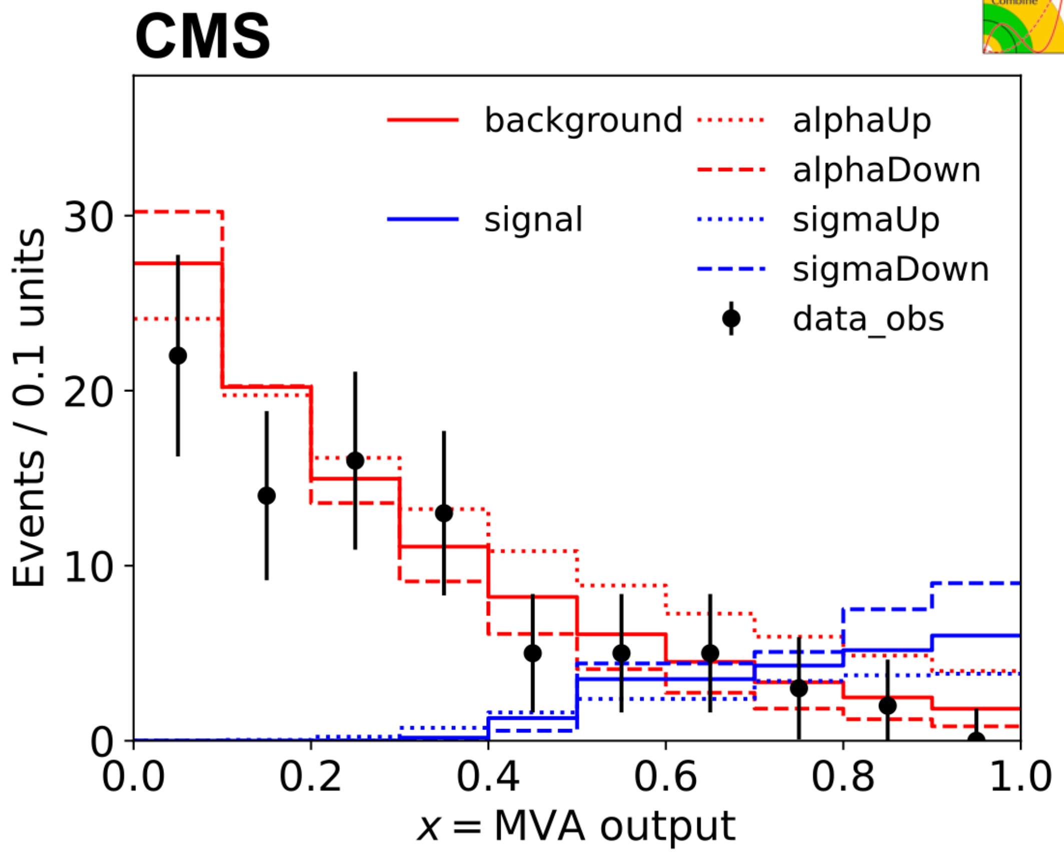


arXiv:2204.06614

USE THE TOOLS TO SEE THE **INGREDIENTS**

Example analysis with two templates, each with a systematic shape variation:

- **Observable x is a MVA output:** more background at lower values and more signal at larger values.
  - One can see from the data that there is probably no signal in Nature.
- Signal syst. uncertainty controlled by nuisance parameter **σ**.
- Background syst. uncertainty controlled by nuisance parameter **α**.



arXiv:2204.06614

USE THE TOOLS TO SEE THE **EFFECTS**

Top panel – bin contents:

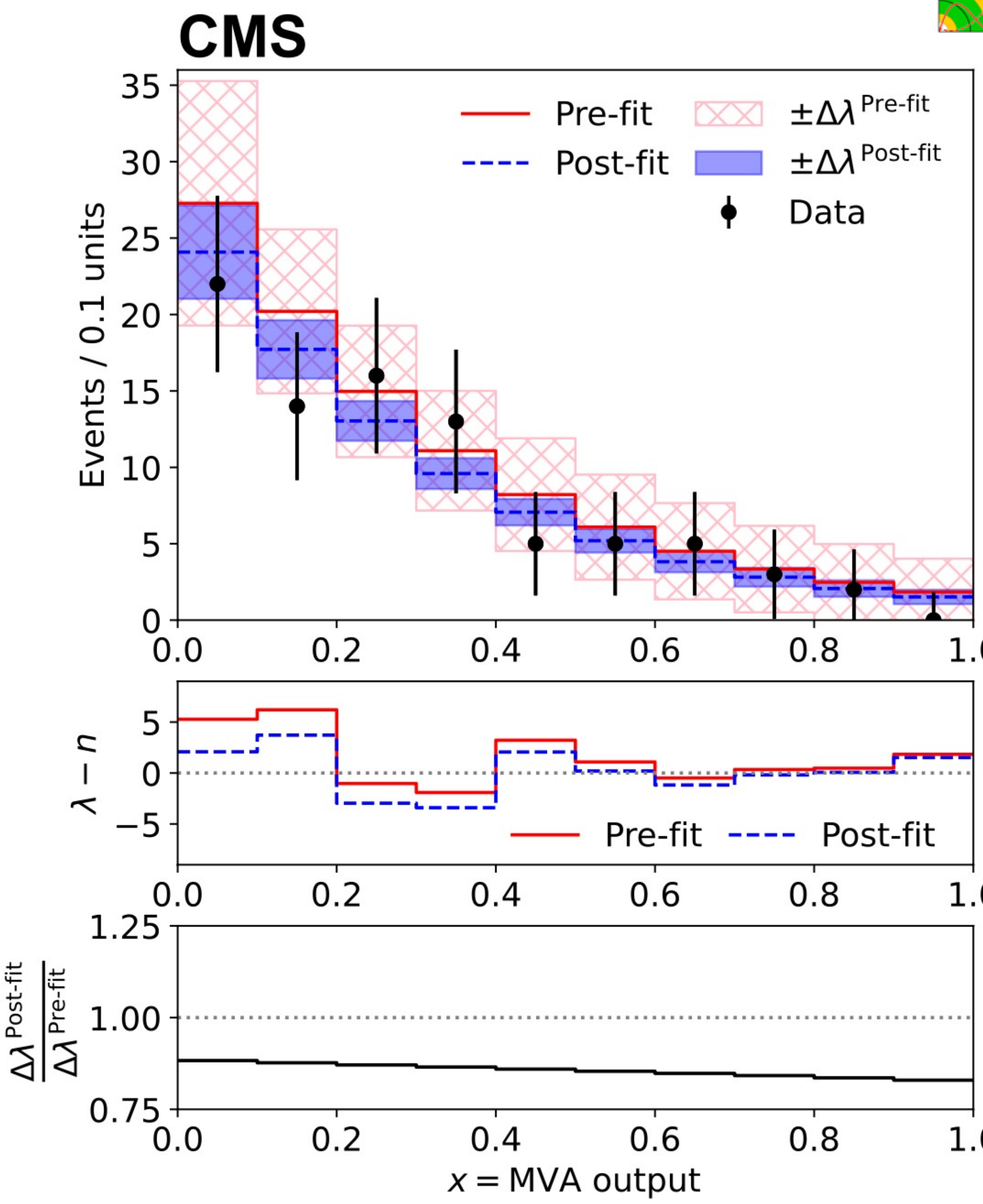
- **Black** points: observed value.
- **Red**: pre-fit estimate and uncertainty.
- **Blue**: ditto post-fit.

Middle panel – bin contents differences:

- Difference between estimates (λ) and observations (n) both pre-fit and post-fit.
- *Post-fit reduces discrepancies between model and data.*

Lower panel – estimate uncertainties:

- Ratio between the estimated uncertainty post-fit and estimated uncertainty pre-fit.
- *Post-fit reduces estimated uncertainties.*



ASTRONOMERS AND CALIBRATORS

All-in-one in HEPP but not universal.

- Also makes HEPP papers have very long, uninformative, author lists.

Cases in LHC where "interpreters" are "calibrators".

Cases where "interpretation" is blunted to not step beyond "calibration" stated ability.

My rule of thumb: if an analysis constrains a calibration-provided nuisance parameter, stop and think.

- And then possibly take action.

ALTERNATIVES AND MORPHING

Some alternatives are **physical deformations** with meaning.

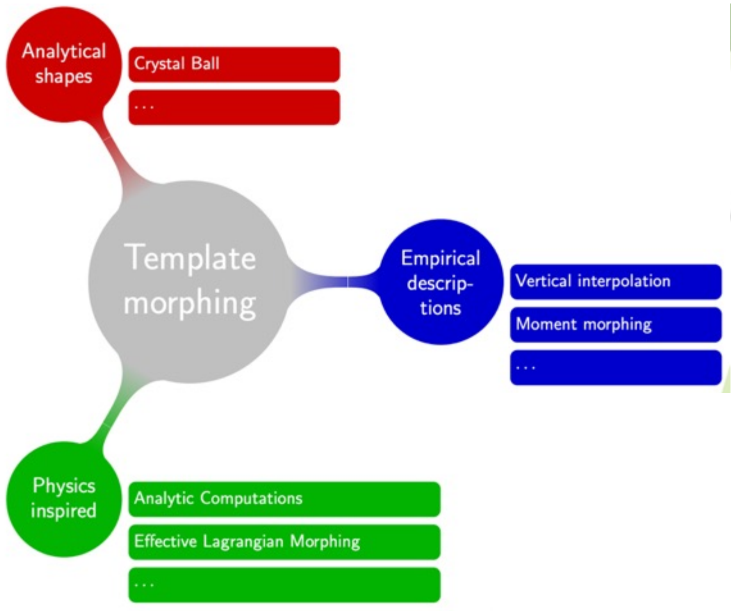
- "Average"/morphing makes sense.

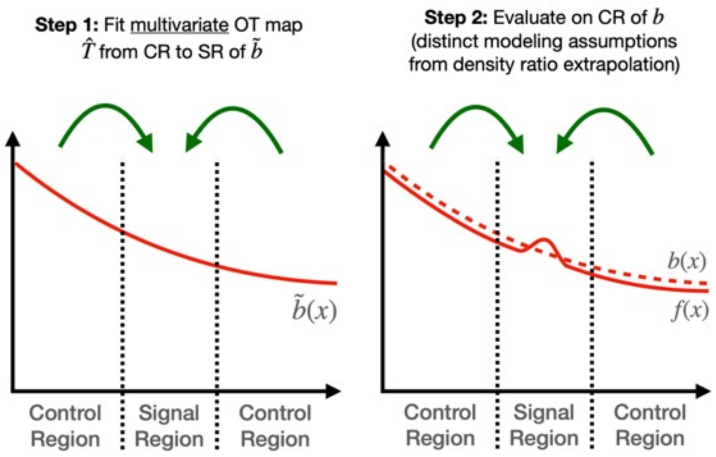
Some alternatives are really **just alternatives**.

- And if they end up mattering we'll likely throw one out as unphysical. (Cousins)

Perturbative theory uncertainties are a whole different beast altogether.

- Limited but non-zero knowledge on the next term.

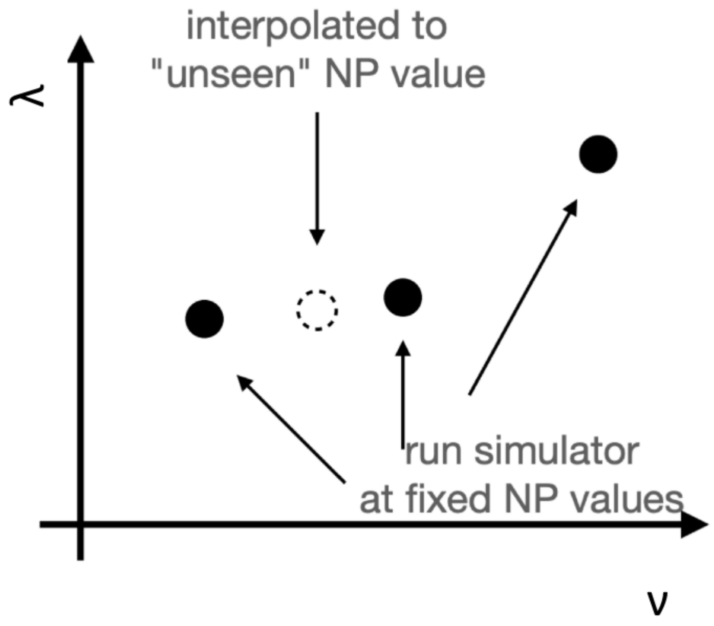


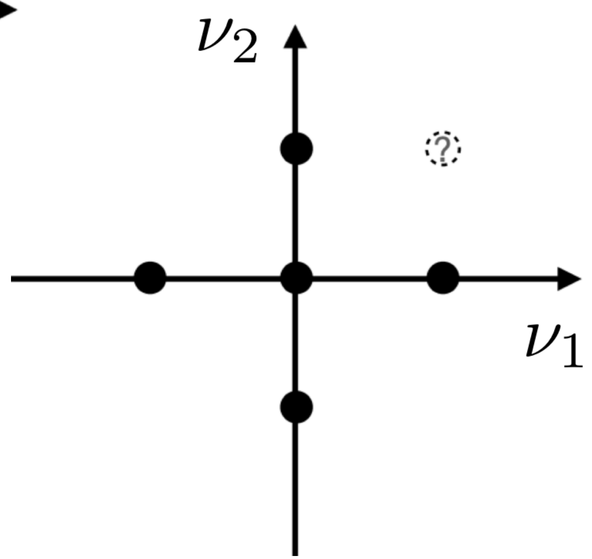


MORPHING

Nuisance parameters are continuous, but samples are generated only a specific values.

Not just **interpolating**, but also **extrapolating**.



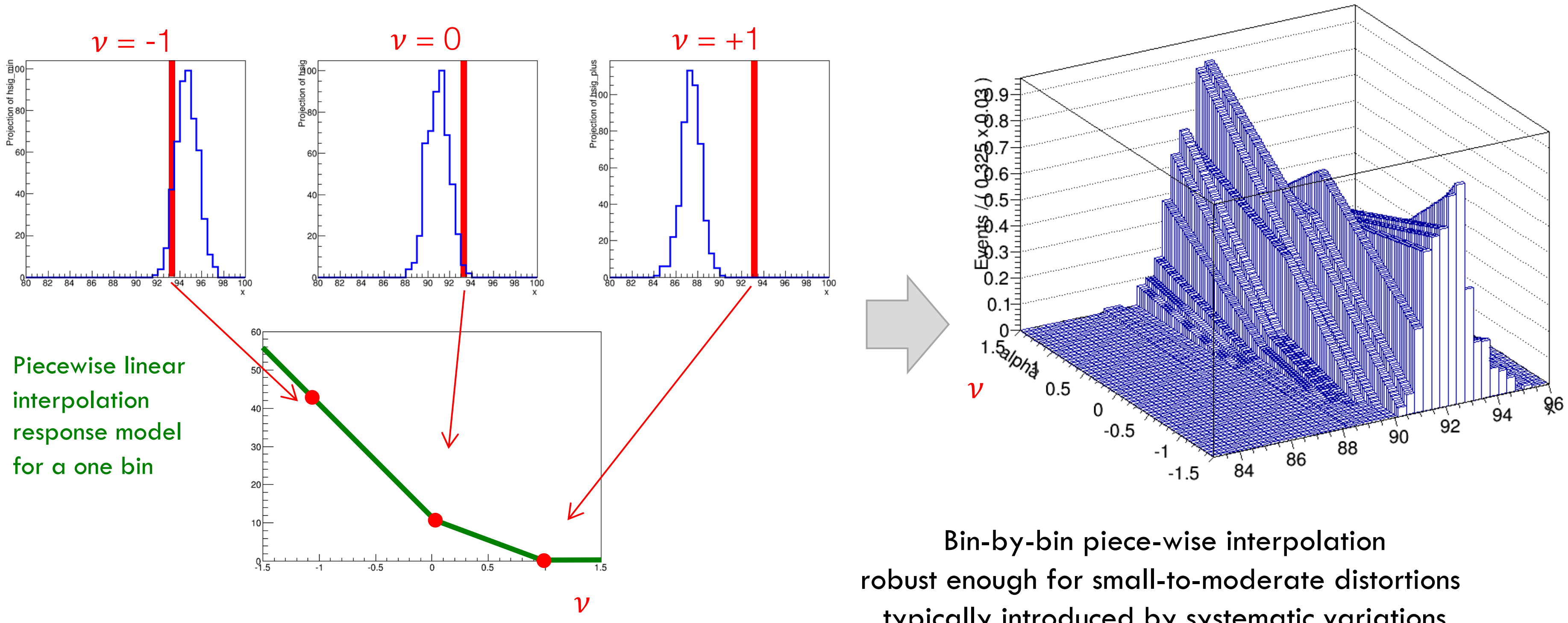


MORPHING

Main workhorse at LHC.

- Several limitations known.

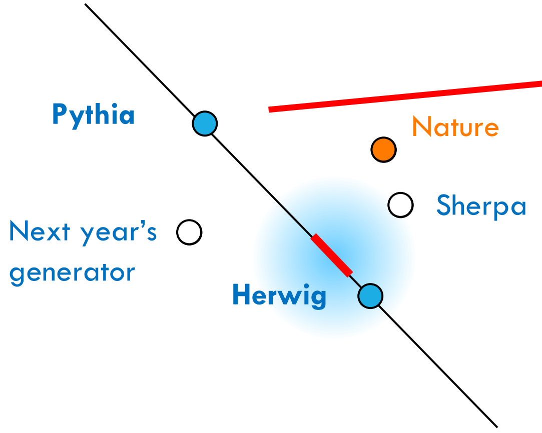
Now modern technology for multi-dimensional morphing now available.

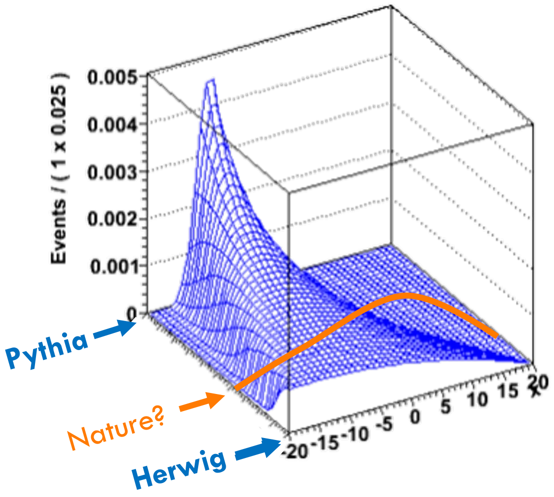


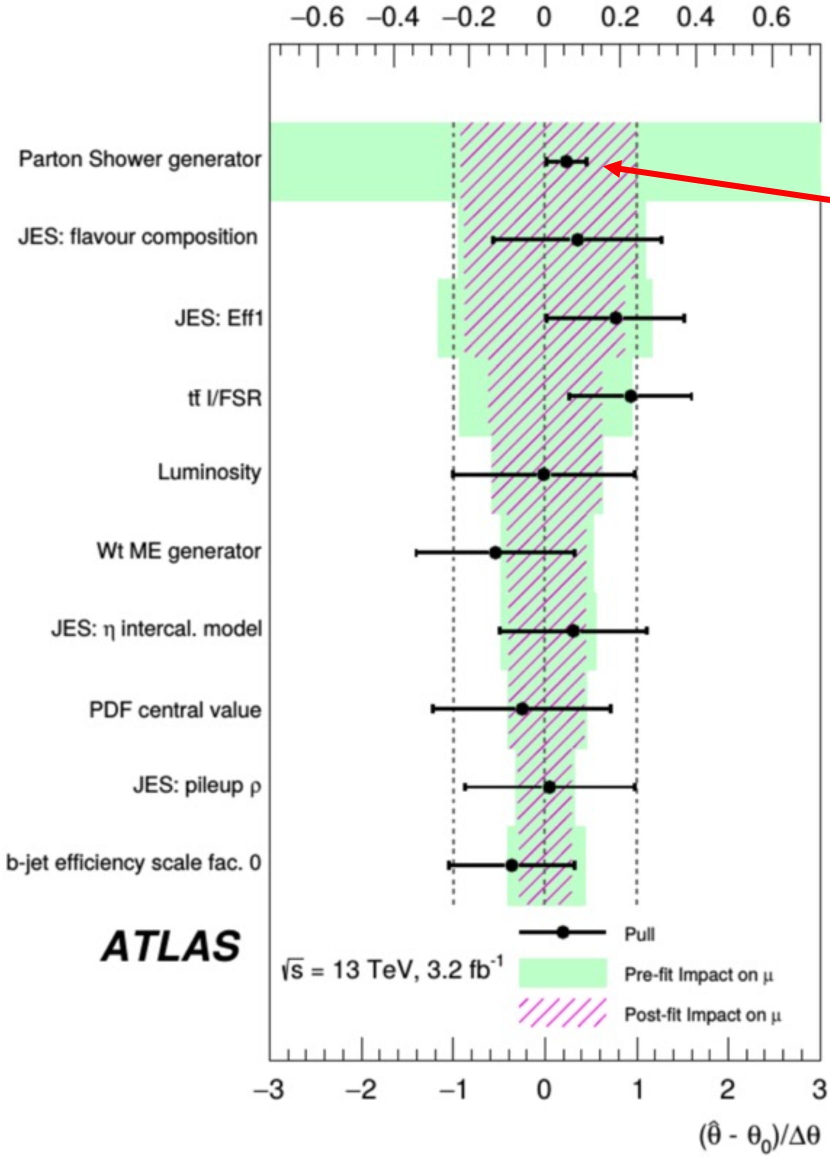
Bin-by-bin piece-wise interpolation robust enough for small-to-moderate distortions typically introduced by systematic variations

ALTERNATIVES

Are we really gaining any knowledge about Nature ?





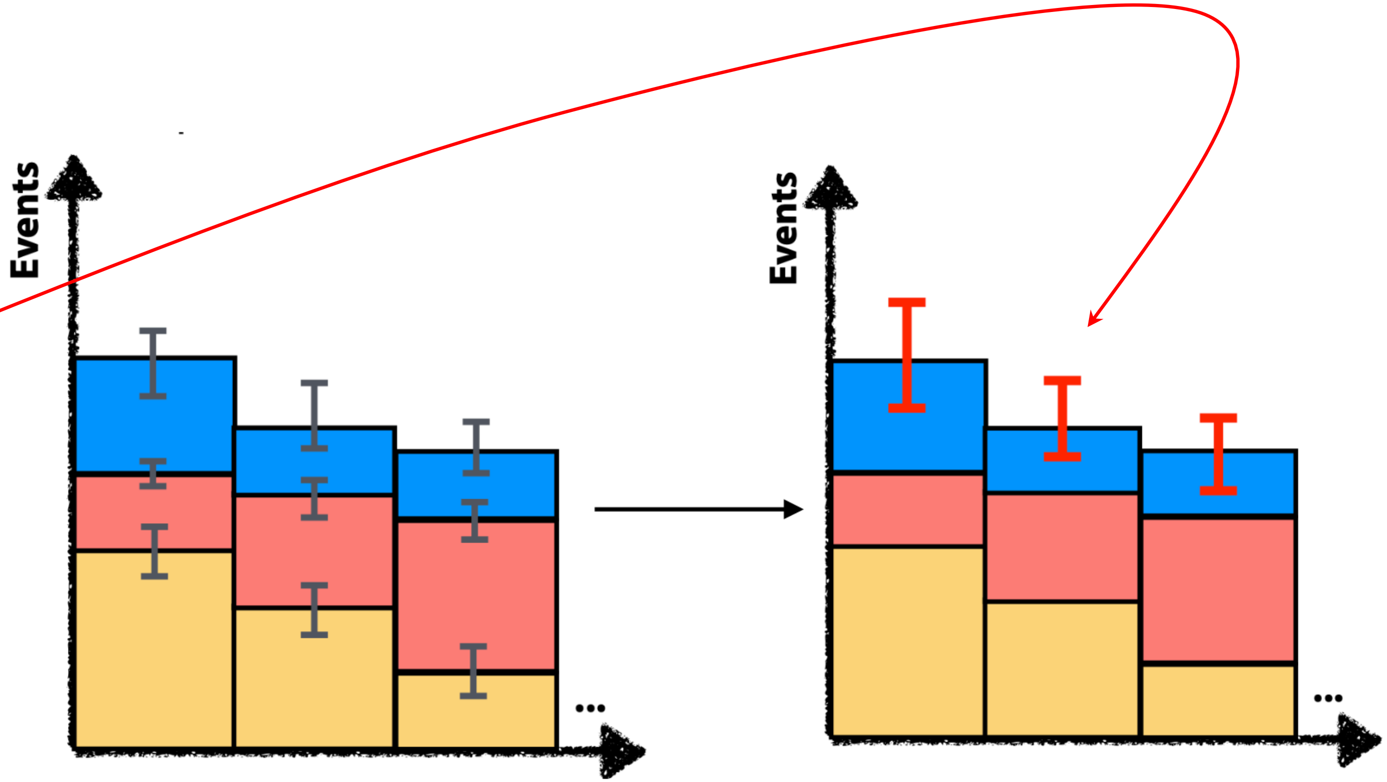


v with 10× reduced uncertainty

LIMITED SIMULATION STATISTICS

Account for uncertainties due to the limited sample size used when creating templates.

- Barlow-Beeston procedure widely available.
- "Lite" version reduces number of nuisance parameters.
  - No one prescription on how to merge nuisances.



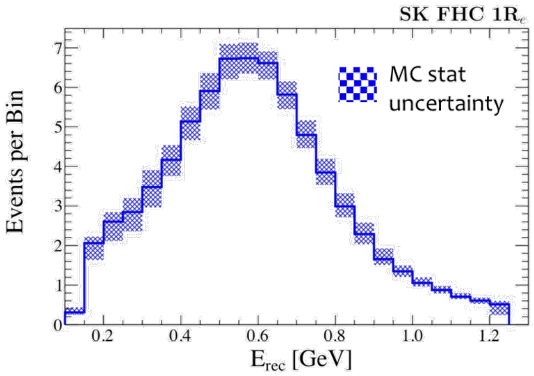
Barlow-Beeston method
Comput. Phys. Comun 77 (1993) 219

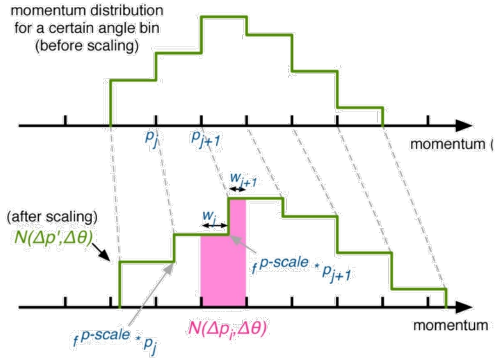
Barlow-Beeston-**lite** method
arXiv:1103.0354

LIMITED SIMULATION STATISTICS

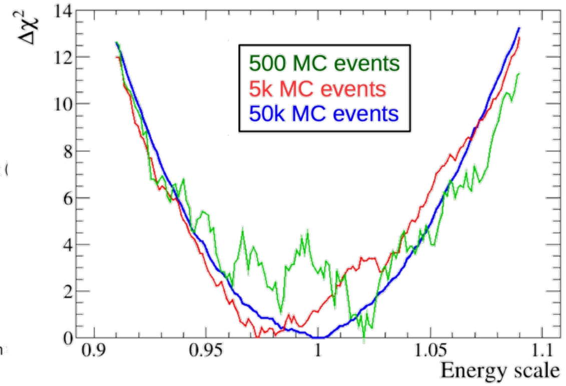
Effects from “horizontal” migration of events.

- Induce “vertical” effects.
- **Not** related to bin-by-bin uncertainties (Barlow-Beeston).

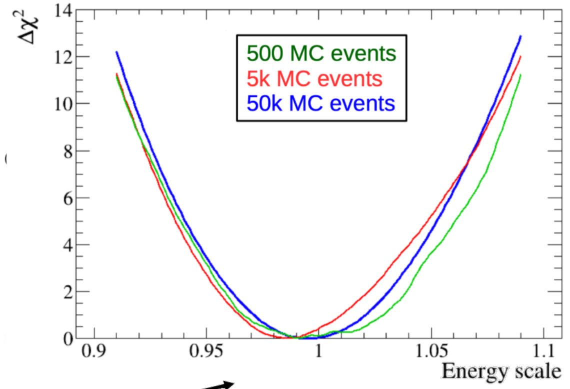




Shift energy scale in simulation and calculate migration between neighboring bins



Jumps in χ^2 due to events jumping between bins



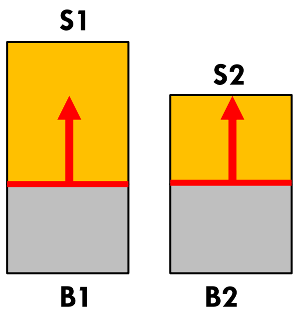
Regularisation can help smooth out these effects: e.g distribute events across-bins (assign ”width” to each event)

"BEING CONSERVATIVE"

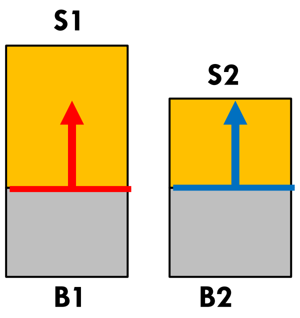
⚠️ Also: mass measurements and scale uncertainties.

Should uncertainties be correlated or uncorrelated ? **It depends.**

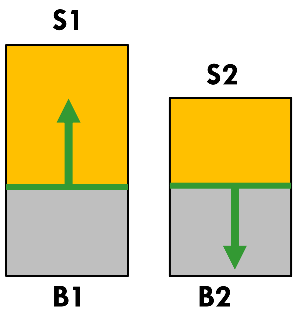
Consider two bins, **Bi** with yields **Si**.



NP: 10% bkg uncertainty
correlated modeling



NP1: 10% bkg unc. – bin 1
NP2: 10% bkg unc. – bin 2



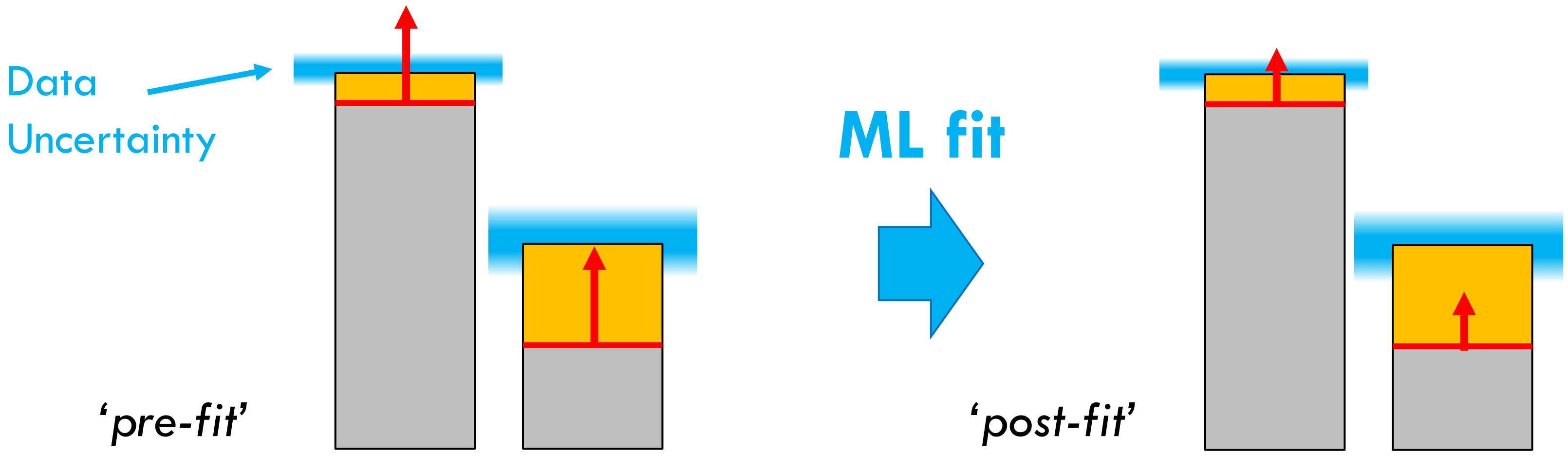
NP: 10% bkg uncertainty
anti-correlated model

POI ∝ S1 + S2	Conservative	Appropriate ?	**Very Optimistic**
POI ∝ S1 / S2	**Very Optimistic**	Appropriate ?	Conservative

"BEING CONSERVATIVE"

Should uncertainties be correlated or uncorrelated ? **It depends.**

Consider bins with very different numbers of "counts".



One NP representing
10% bkg uncertainty
correlated effect in both bins

One NP representing
5% bkg uncertainty
correlated effect in both bins

Uncertainty reduction in both bins from contraining power of left bin alone !

ALTERNATIVES AND MORPHING

Some alternatives are physical deformations with meaning.

- "Average"/morphing makes sense.

Some alternatives are really just alternatives.

- And if they end up mattering we'll likely throw one out as unphysical. (Cousins)

Perturbative theory uncertainties are a whole different beast altogether.

- Limited but non-zero knowledge on the next term.

Parametrize and estimate the actual source of the uncertainty: $f''(0)$

$$f(x) = f(0) + f'(0)\,x \underbrace{+ f''(0)\,\frac{x^2}{2} + \mathcal{O}(x^3)}_{\text{source of the theory uncertainty}}$$

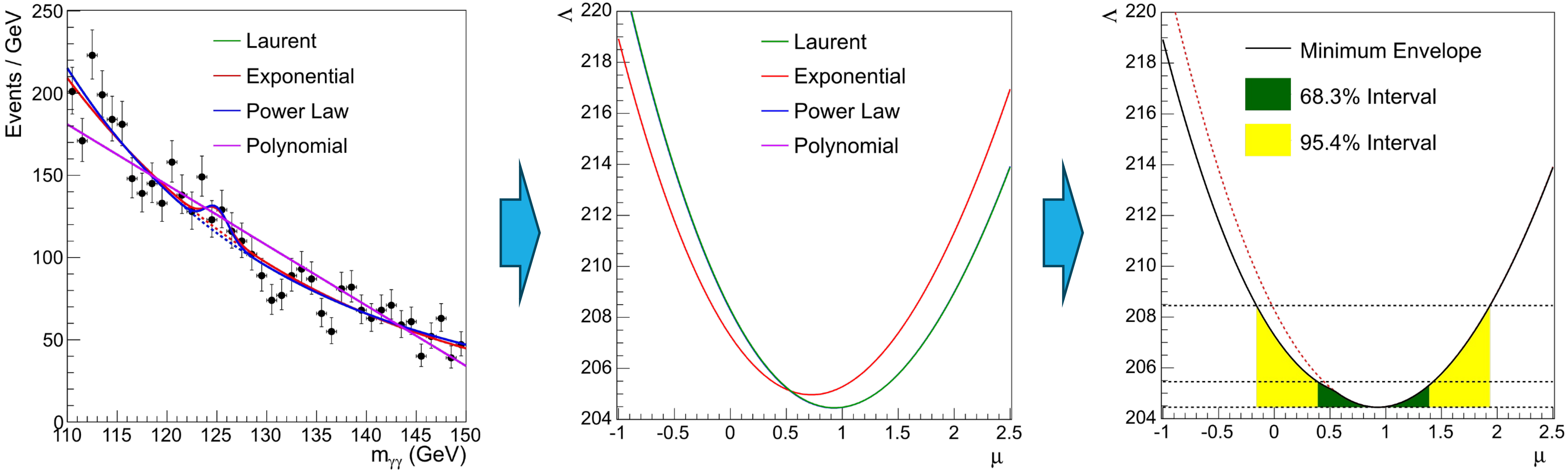
- We typically know a lot about the general structure of $f''(0)$ even without explicitly calculating it

DISCRETE PROFILING

Invented in CMS $H \to \gamma\gamma$ to deal with **different parametric background choices.**

- ATLAS uses "spurious signal" (different) method.

Extensively validated to not bias inference.



THE 🪄 OF DISCRETE PROFILING

Perhaps there is **hope to understand discrete profiling** in the model selection context.

- How do you feel about model averaging being the (weighted) average of estimates across different models?

Unavoidable comparison with spurious signal; **both are prescriptions using statistical uncertainty under the signal as the gauge**

- Discrete profiling functions are chosen to have bias smaller than O(10%) stat. unc.
- Spurious signal is chosen on similar basis and added to signal model.

Model	MLE of θ	Log likelihood at MLE	AIC	Akaike Weight
Linear	0.13	-174.79	355.58	≈ 0
Quadratic	0.84	-155.85	319.70	0.29
Cubic	1.14	-154.81	319.62	0.30
Quartic	1.21	-154.02	320.05	0.24
Quintic	1.18	-153.84	321.67	0.11
Sextic	1.24	-153.83	323.65	0.040
Septic	1.25	-153.34	324.69	0.024

A WELCOME SYSTEMATISATION

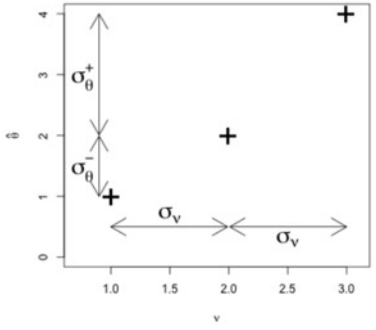
OPAT vs APAST

Combination of measurements vs combined measurement.

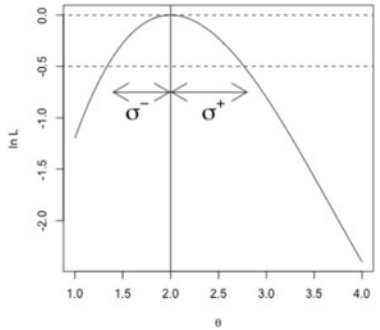
Discussion on **simplified likelihoods**:

- Taylor expansion seems to be founded.
- For PDFs a whole different story: cumulants, saddle point approximation, etc.

"Systematic" OPAT systematics evaluation

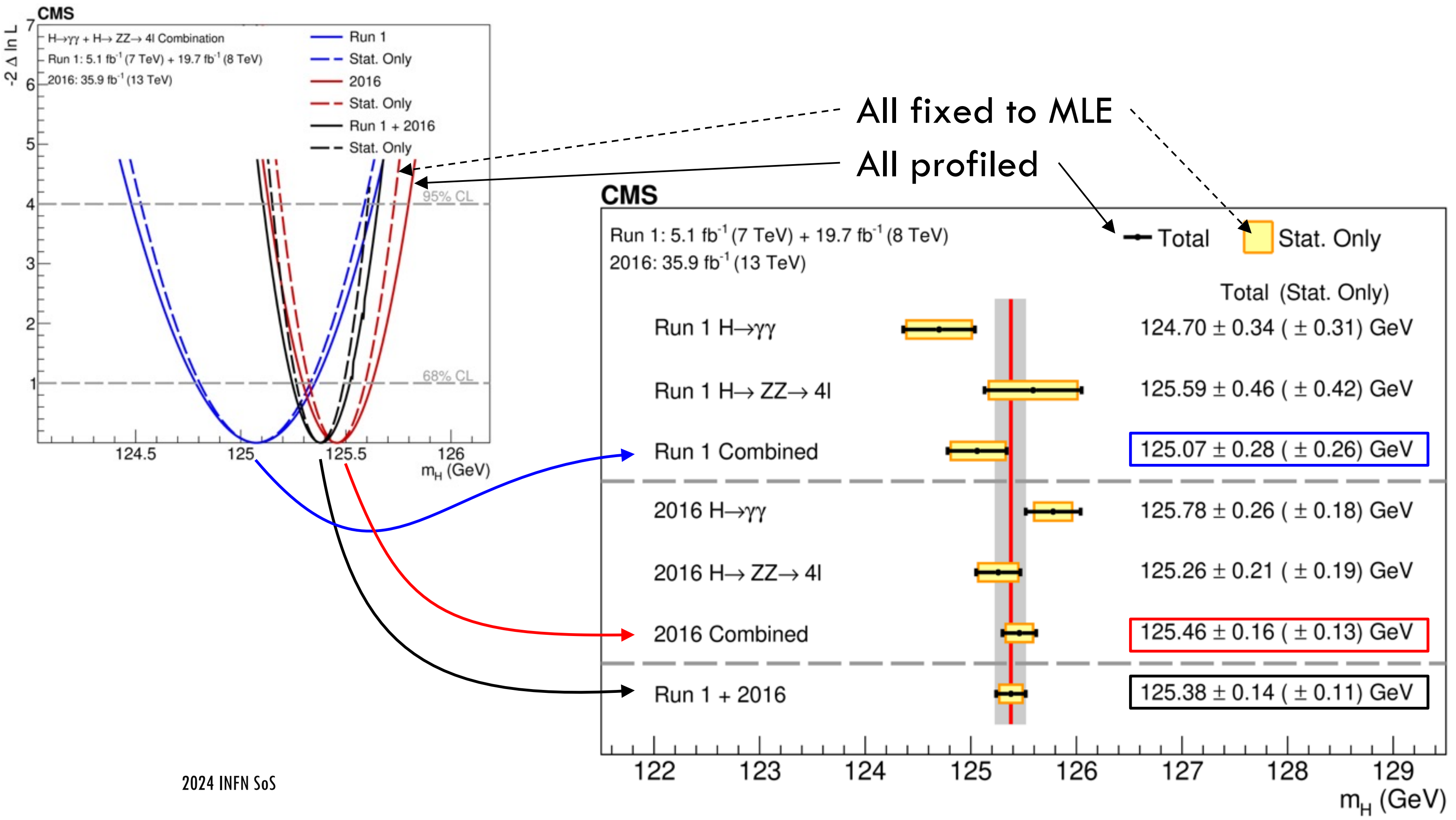


"Statistical" From ML estimation



From PDFs	From Likelihoods
Error is variance of result	Error is 68% central CL
You are probably Combining Errors, in quadrature + skew	You are probably Combining Results
Goodness of fit is irrelevant	Compatibility vital & straightforward
You are probably not combining results (but you can if you work at it)	You are probably not combining errors (you can if you work at it, but not in quadrature)
"Systematic" Asymmetric Error formulæ	"Statistical" Asymmetric Error formulæ

APAT – ALL PARAMETERS AT A TIME



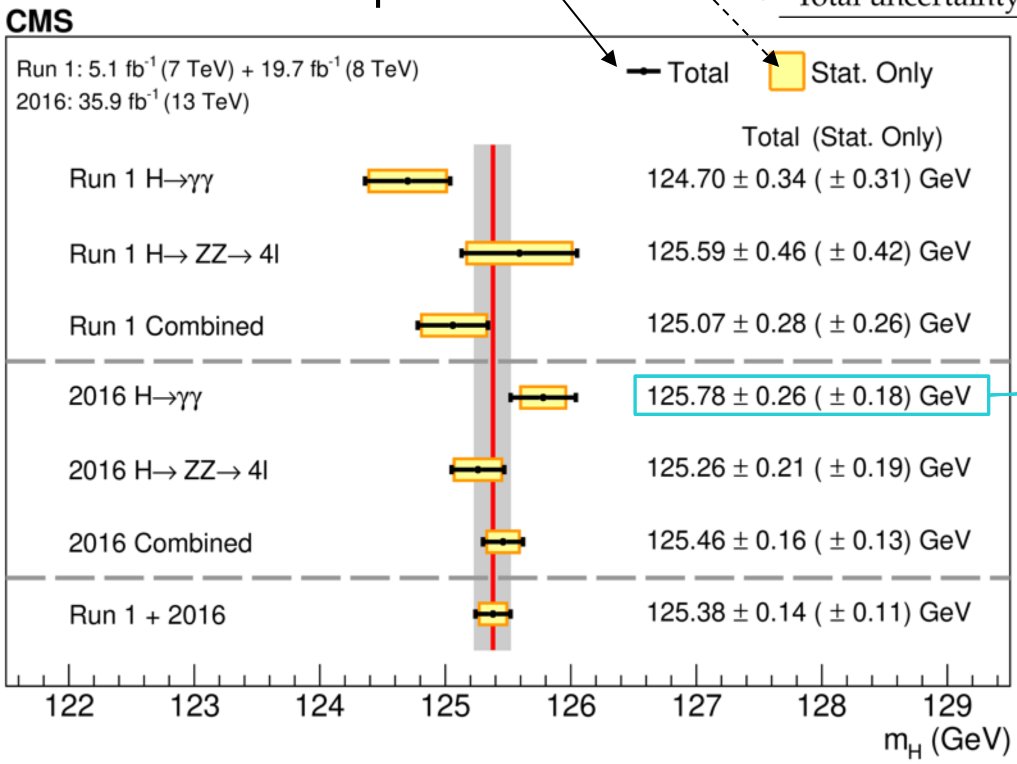
APAT – ALL PARAMETERS AT A TIME

Source	Contribution (GeV)
Electron energy scale and resolution corrections	0.10
Residual p_T dependence of the photon energy scale	0.11
Modelling of the material budget	0.03
Nonuniformity of the light collection	0.11
Total systematic uncertainty = Total unc. ⊖ Stat. unc.	0.18
Statistical uncertainty	0.18
Total uncertainty	0.26

Subsets profiled

All fixed to MLE

All profiled



BETTER ASYMPTOTICS

Sine qua non for ”errors on errors” that can benefit all.

Correction can also be used as coverage diagnostic tool.

I wonder what happens in asymmetric cases…

$$r_\mu^* = r_\mu + \frac{1}{r_\mu}\log\frac{q_\mu}{r_\mu} = \frac{r_\mu - \mathrm{E}[r_\mu]}{\mathrm{V}[r_\mu]^{1/2}} + \mathcal{O}(n^{-3/2})$$

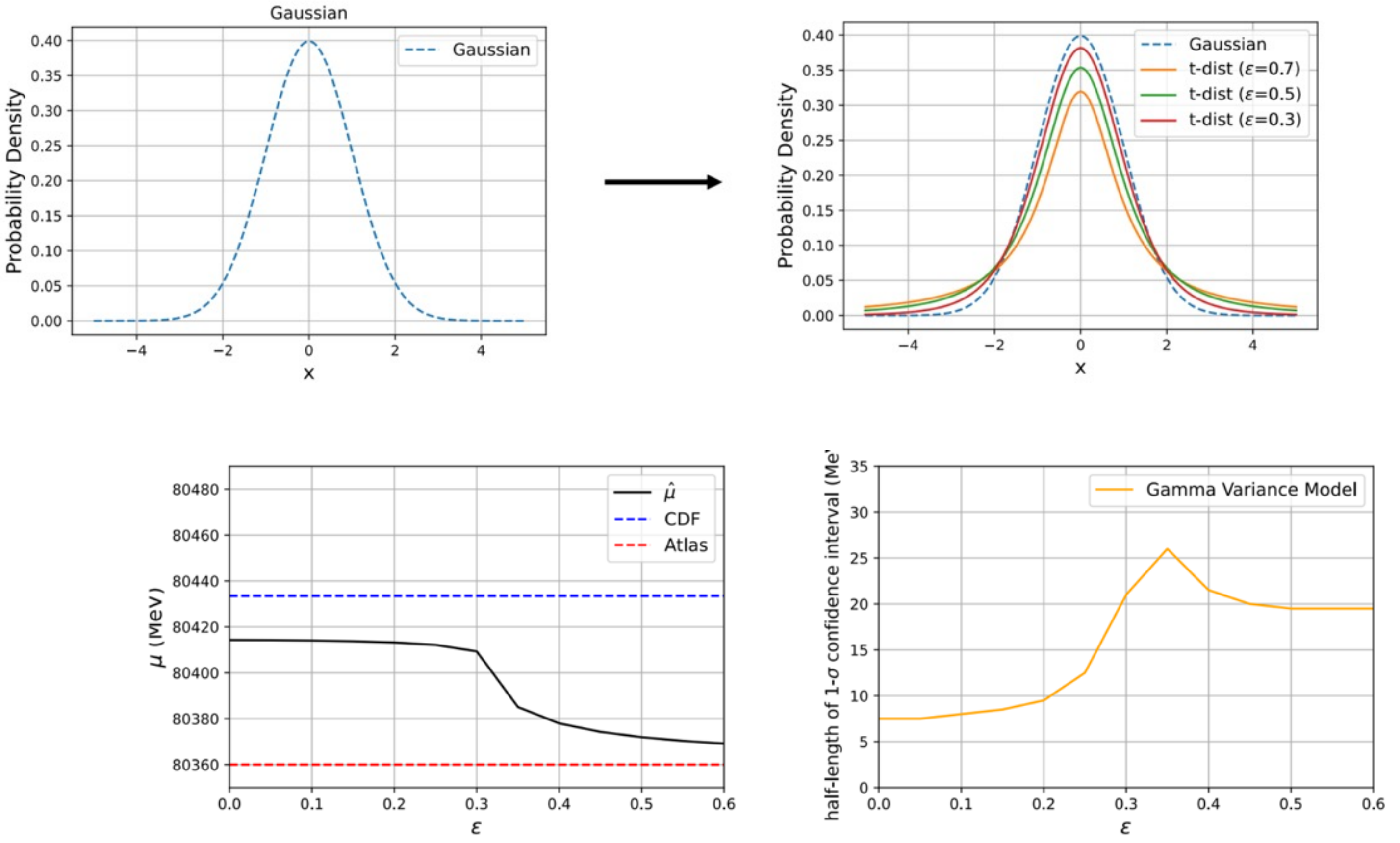
$$r_\mu^* \sim \mathcal{N}(0,1) + \mathcal{O}(n^{-3/2})$$

HOW CERTAIN IS THAT UNCERTAINTY?

“Unleash the tails !”

Discussion focused on applying these foremost to theory inputs.

- For exp. uncs. I wonder what the evaluation experiment_k by physicist_i would be.
  - Especially when k = i.
- Lots of interesting ideas to pursue to understand how it deals with outliers.
- I know at least one theorist seriously studying the method.



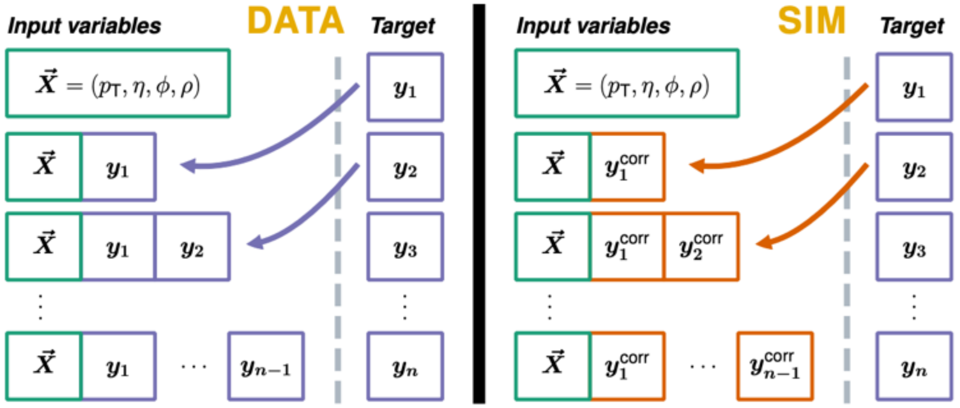
"LAST MILE" CORRECTIONS

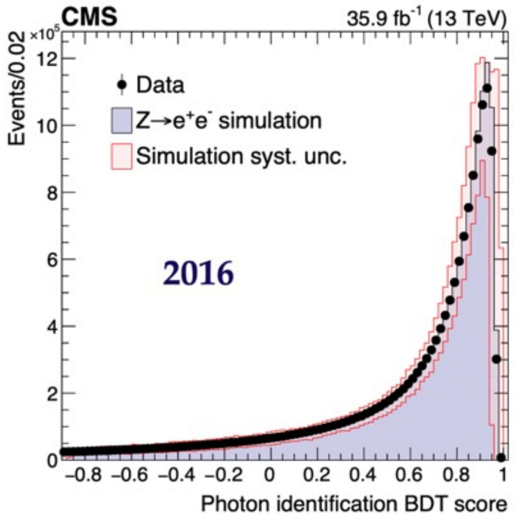
Simulation imperfections can have substantial impact on inference.

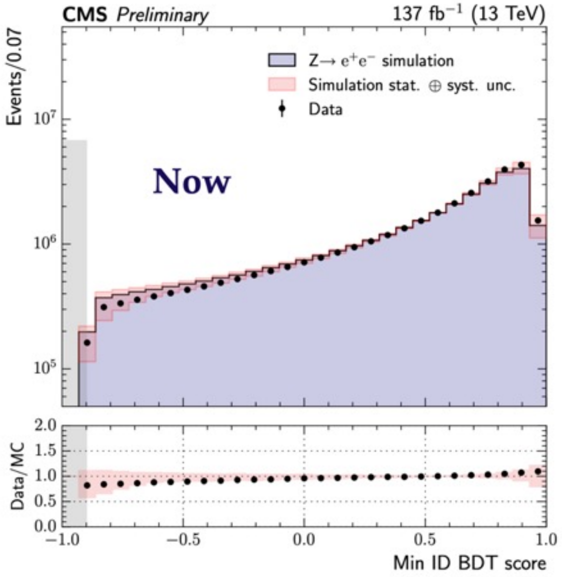
Example evolution with time:
- "Multiply and smear".
- 1-D quantile regression.
- Chained quantile regression.

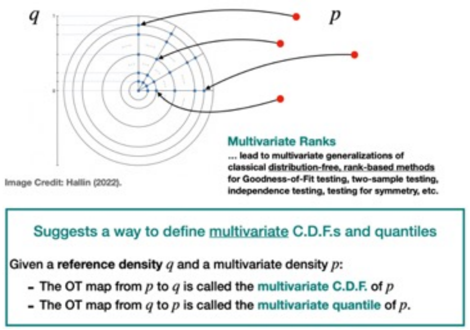
Is this the best that can be done?

Heard at PHYSTAT 2023:
- Multi-dim. quantile regression.
- Multi-dim. CDF.
- Optimal Transport maps.









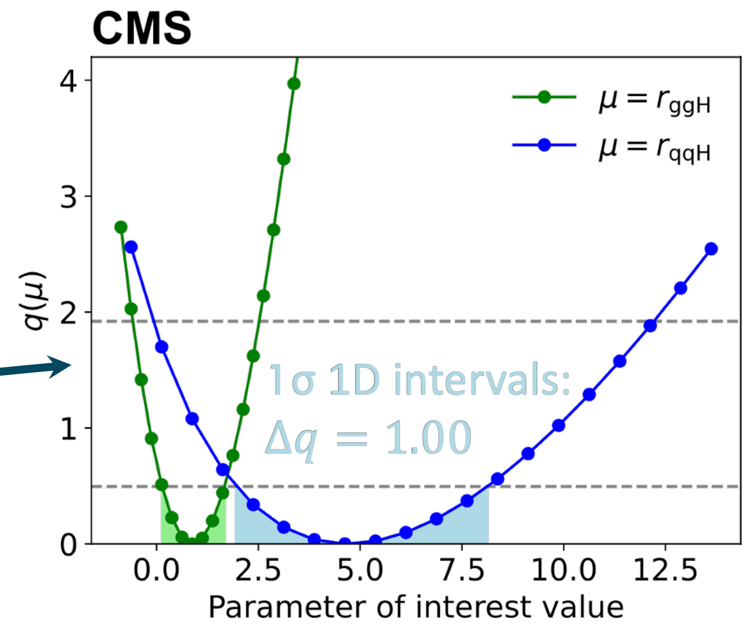
≥1 PARAMETERS OF INTEREST

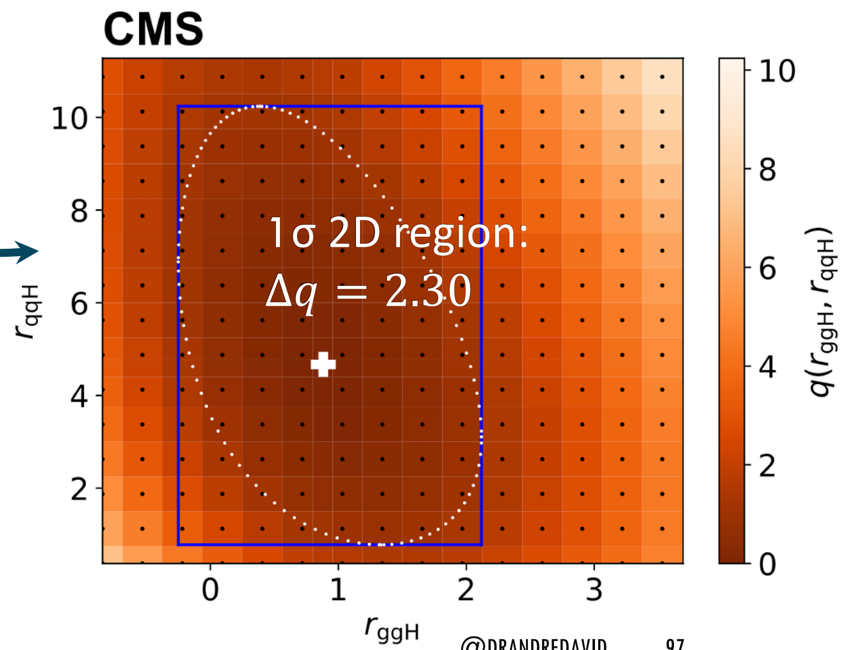
Thanks to Wald and Engle !

Joint inference can and should be done.

Same asymptotics but thresholds for regions depend on number of parameters.

$(1-\alpha)$ (%)	$M=1$	$M=2$
68.27	1.00	2.30
90.	2.71	4.61
95.	3.84	5.99
95.45	4.00	6.18
99.	6.63	9.21
99.73	9.00	11.83





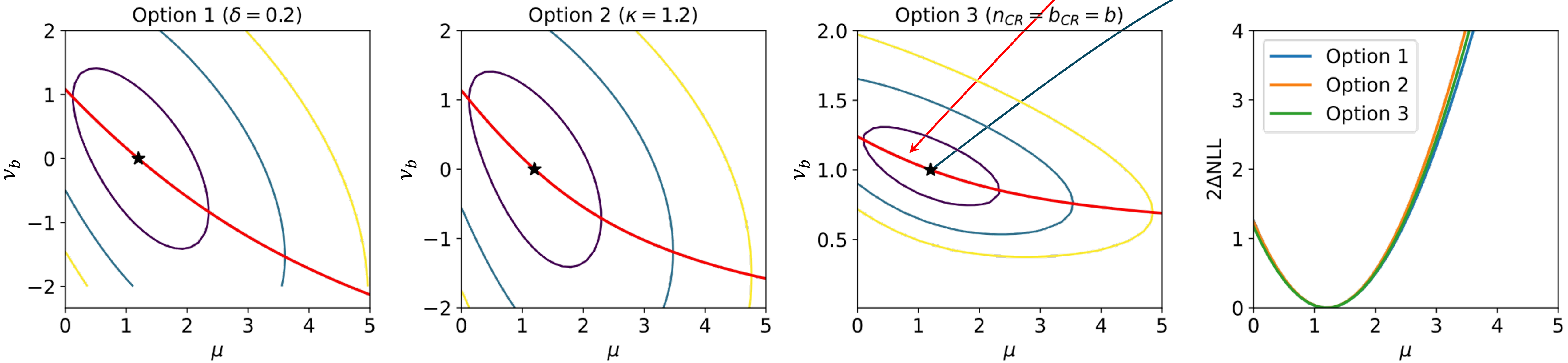
KNOW YOUR APPROXIMATIONS

Background with 20% uncertainty modelled in three different ways.

- Option 1 (Gaussian) $p(x;\mu,\nu_b) = P\big(x;\mu\cdot s + b(1+\delta\cdot\nu_b)\big)\cdot N(y;\nu_b,1).$
- Option 2 (Log-normal) $p(x;\mu,\nu_b) = P(x;\mu\cdot s + b\cdot\kappa^{\nu_b})\cdot N(y;\nu_b,1).$
- Option 3 (Gamma) $p(x;\mu,\nu_b) = P(x;\mu\cdot s + b\cdot\nu_b)\cdot P(n_{CR};b_{CR}\cdot\nu_b).$

$$q(\mu) = -2\ln\frac{\mathcal{L}(\mu,\widehat{\widehat{\nu_b}}(\mu))}{\mathcal{L}(\hat{\mu},\widehat{\nu_b})}$$

$s = 25, b = 25, x = 37$

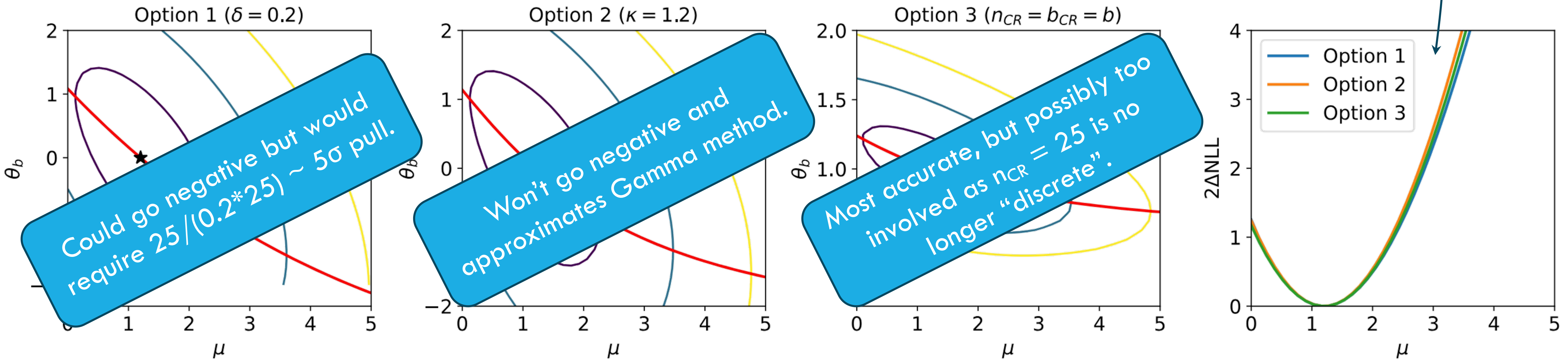


SAVED BY THE C.L.T.

Background with 20% uncertainty modelled in three different ways:

- Option 1 (Gaussian) $p(x;\mu,\nu_b) = P\big(x;\mu\cdot s + b(1+\delta\cdot\nu_b)\big)\cdot N(y;\nu_b,1).$
- Option 2 (Log-normal) $p(x;\mu,\nu_b) = P(x;\mu\cdot s + b\cdot\kappa^{\nu_b})\cdot N(y;\nu_b,1).$
- Option 3 (Gamma) $p(x;\mu,\nu_b) = P(x;\mu\cdot s + b\cdot\nu_b)\cdot P(n_{CR};b_{CR}\cdot\nu_b).$

$s = 10, b = 25, x = 37 \Rightarrow$ Small difference in the inference **for this case.**



PARTING THOUGHTS

😅 **You'll spend a lot of time correcting measurements.**

- So make friends with the uncertainties that come with those corrections.

🫙 **Whenever you have low counts, be very careful.**

- If you have zero counts, welcome to the club.

⚡ **Profiling nuisances has great power.**

- Whether the power to constrain them is licit or not is another matter.

Vamos a la playa

Brano di Righeira

[Righeira Vamos a la playa 1983 - YouTube](https://www.youtube.com)

https://www.youtube.com › watch

Testo

Vamos a la playa, oh oh oh oh
Vamos a la playa, oh oh oh oh
Vamos a la playa, oh oh oh oh
Vamos a la playa, oh oh… Testo completo

Fonte: LyricFind

FOR DISCUSSION

ACKNOWLEDGEMENTS

All who have contributed to more than two decades of PHYSTAT wisdom-building.

- https://phystat.github.io/Website/

CMS's collective wisdom distilled into the COMBINE tool.

- https://arxiv.org/abs/2404.06614
- https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit

Speakers at the 2021 and 2023 PHYSTAT workshops on systematic uncertainties.

- https://indico.cern.ch/event/1051224
- https://www.birs.ca/events/2023/5-day-workshops/23w5096

Wouter Verkerke's decades of material on the topic of modelling.

- *Dankjewel !*

NOT ENOUGH TIME TO COVER DETAILS OF...

Multidimensional modern morphing.

"Marginalizing versus Profiling of Nuisance Parameters" arXiv:2404.17180

Diagonalization and externalisation of uncertainties.

Sampling nuisance parameters and constructing toy datasets.

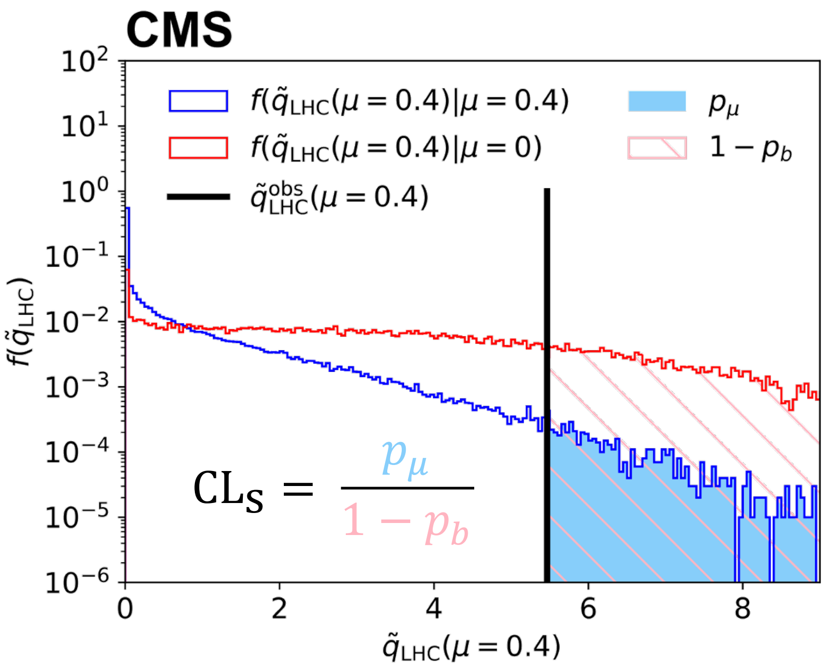
The galaxy of asymptotics:

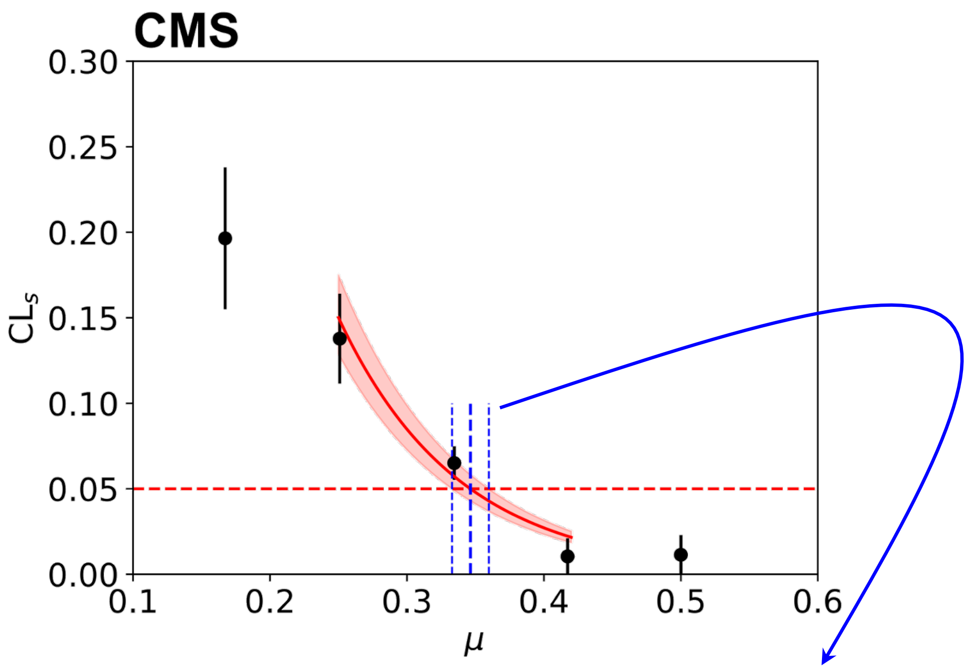
- Wilks (single parameter),
- Wald and Engle (multiple parameters), and
- Chernoff and Self-Liang (parameters at boundaries).

The Trials Factor or Look Elsewhere Effect and Machine Learning.

THE CL_S CRITERION – 95% CL LIMIT EXAMPLE

Motivation and description in PDG RPP 40.4.2.4.






```
>   -- Hybrid New --
> Limit: r < 0.346362 +/- 0.0134581 @ 95% CL
> Done in 0.31 min (cpu), 0.32 min (real)
```

$$p_\mu = \begin{cases} \int_{q_x^{obs}(\mu)}^{\infty} f(q_x(\mu)|\mu)\,dq & \text{if x=LHC,} \\ \int_{-\infty}^{q_x^{obs}(\mu)} f(q_x(\mu)|\mu)\,dq & \text{if x=TEV or LEP,} \end{cases}$$

$$p_b = \begin{cases} \int_{0}^{q_x^{obs}(\mu)} f(q_x(\mu)|0)\,dq_x & \text{if x=LHC,} \\ \int_{q_x^{obs}(\mu)}^{\infty} f(q_x(\mu)|0)\,dq_x & \text{if x=TEV or LEP,} \end{cases}$$

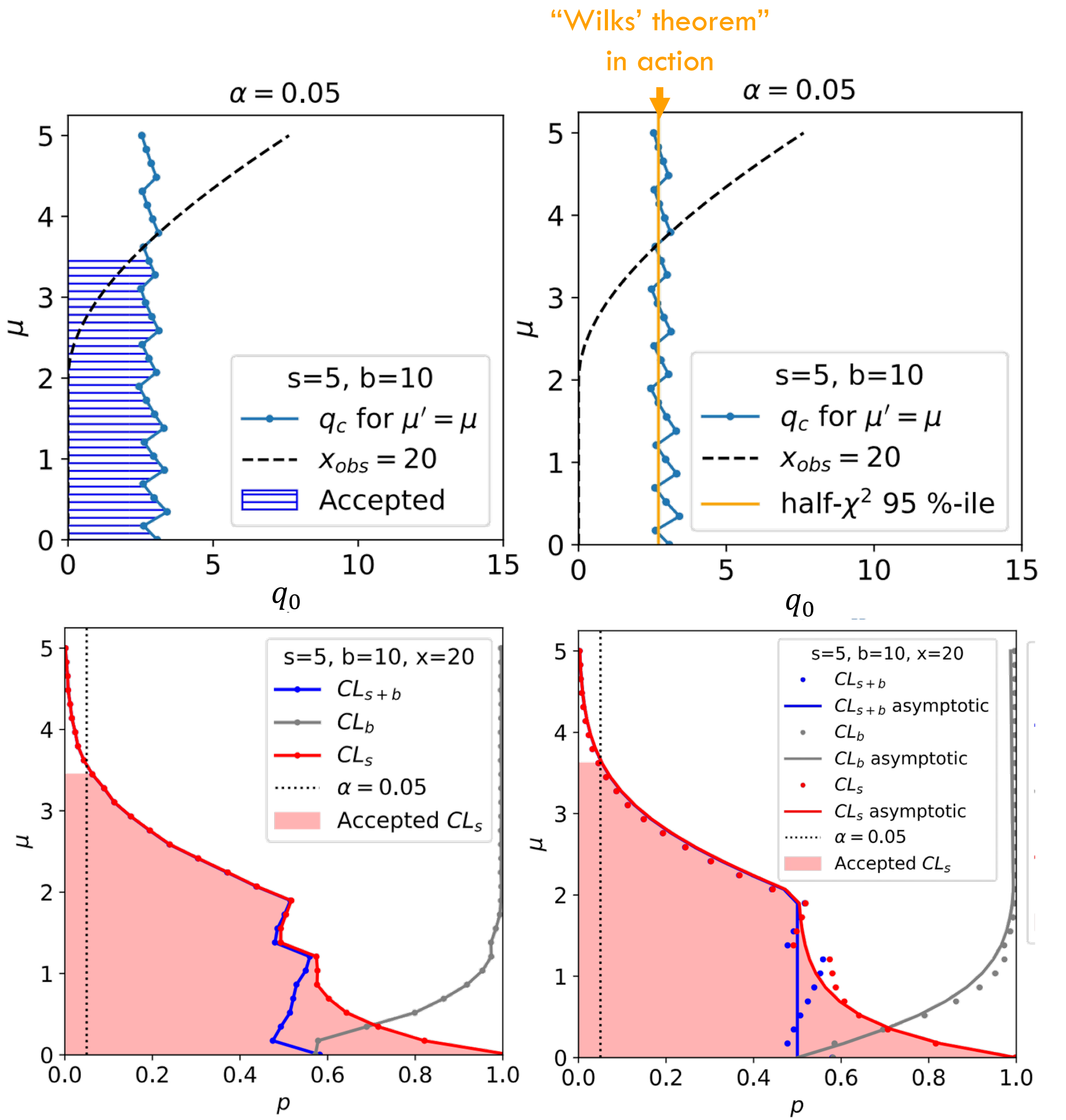
POISSON EXAMPLE WITHOUT NUISANCES

Model: $p(x; \mu s + b) = \frac{(\mu s+b)^x e^{-(\mu s+b)}}{x!}$

- Known (fixed!) background.

Test statistic: $q_0(\mu) = -2 \ln \frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})}$

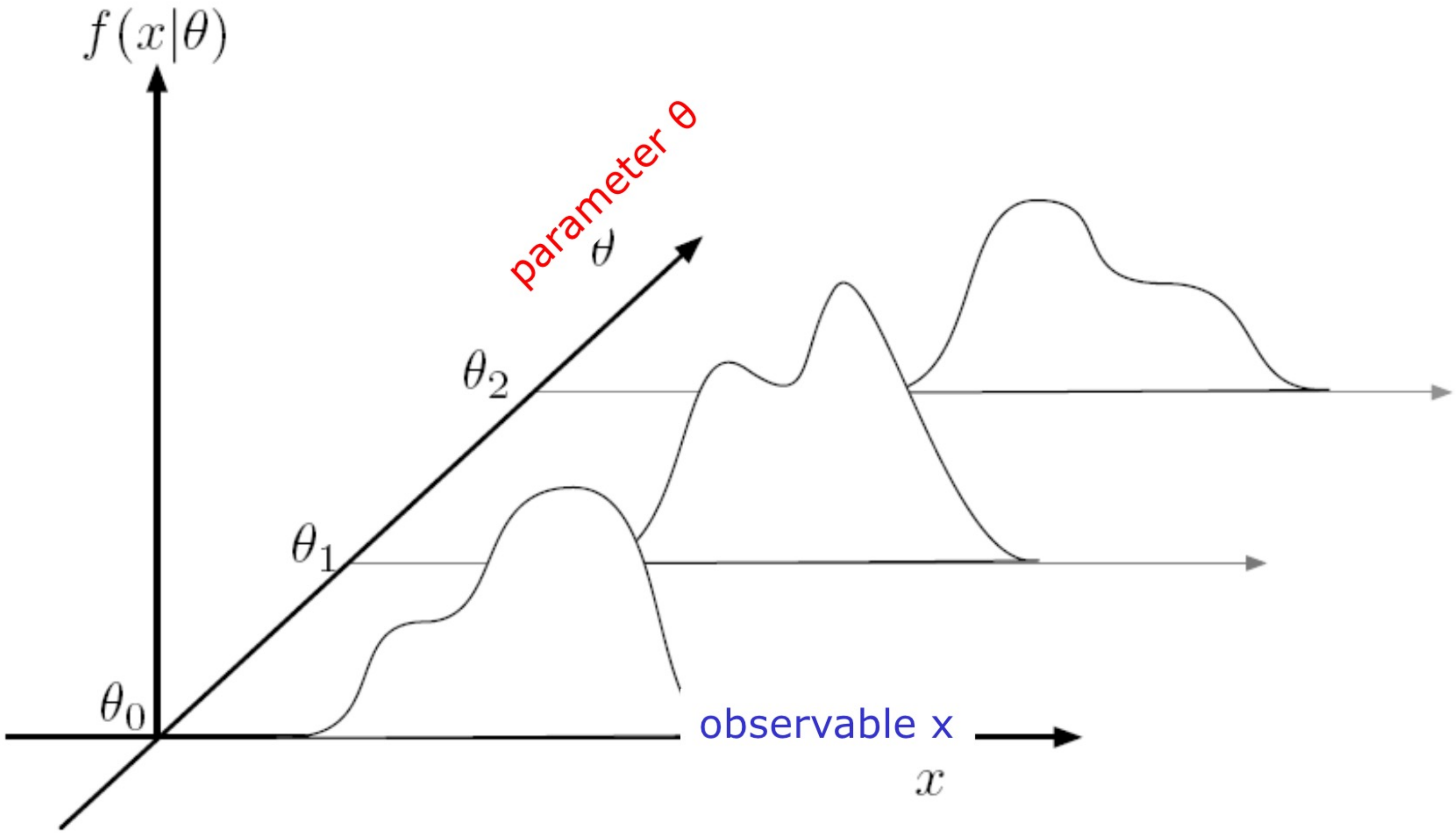
- Jagged behaviour of q_0 due to Poisson discrete nature, not by limited toy statistics (10^4 in this case).



NEYMAN CONSTRUCTION

NEYMAN CONSTRUCTION DECONSTRUCTED

- Simplest experiment: one measurement (x), one theory parameter (θ)
- For each value of parameter θ, determine distribution in in observable x



from W. Verkerke

NEYMAN CONSTRUCTION DECONSTRUCTED

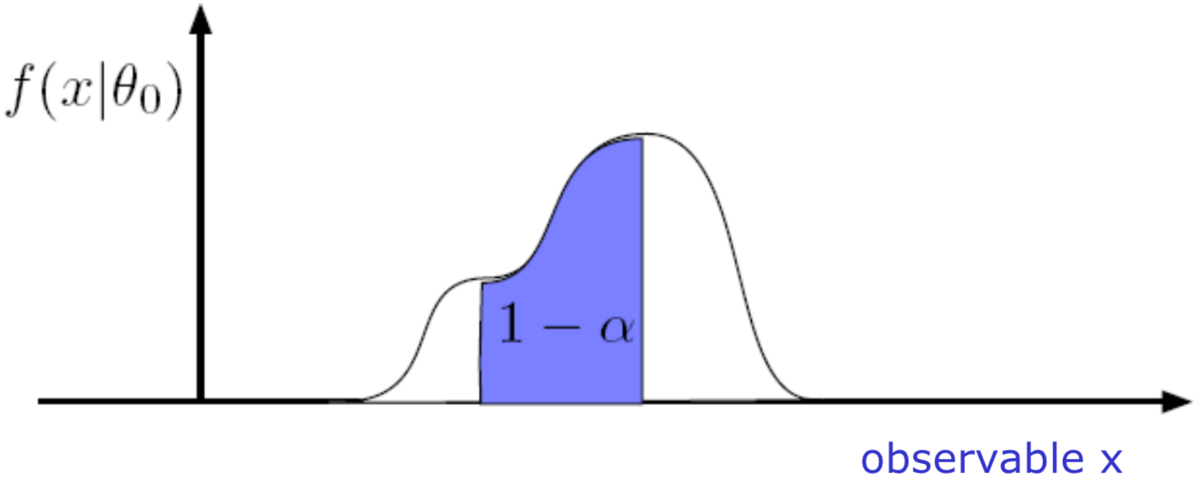
⚠️ This $1-\alpha$ region is:

- **two-sided** for **intervals,** and
- **one-sided** for **limits.**

ℹ️ $1-\alpha = 68\%$ constructs $\pm 1\sigma$ intervals.

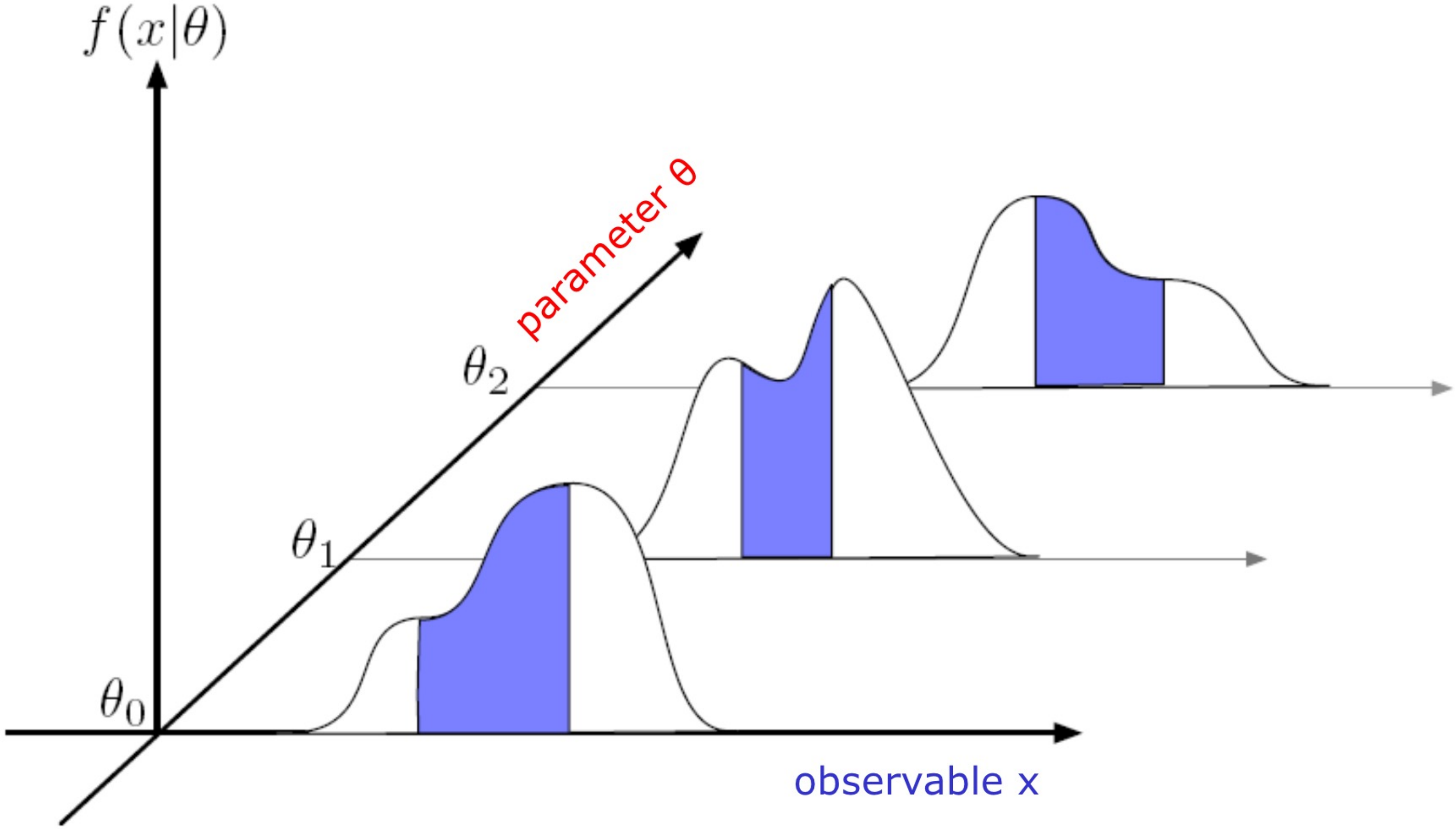
- Focus on a slice in θ
  - For a 1-α confidence Interval, define *acceptance interval* that contains 100%-α of the distribution

pdf for observable x
given a parameter value θ_0



NEYMAN CONSTRUCTION DECONSTRUCTED

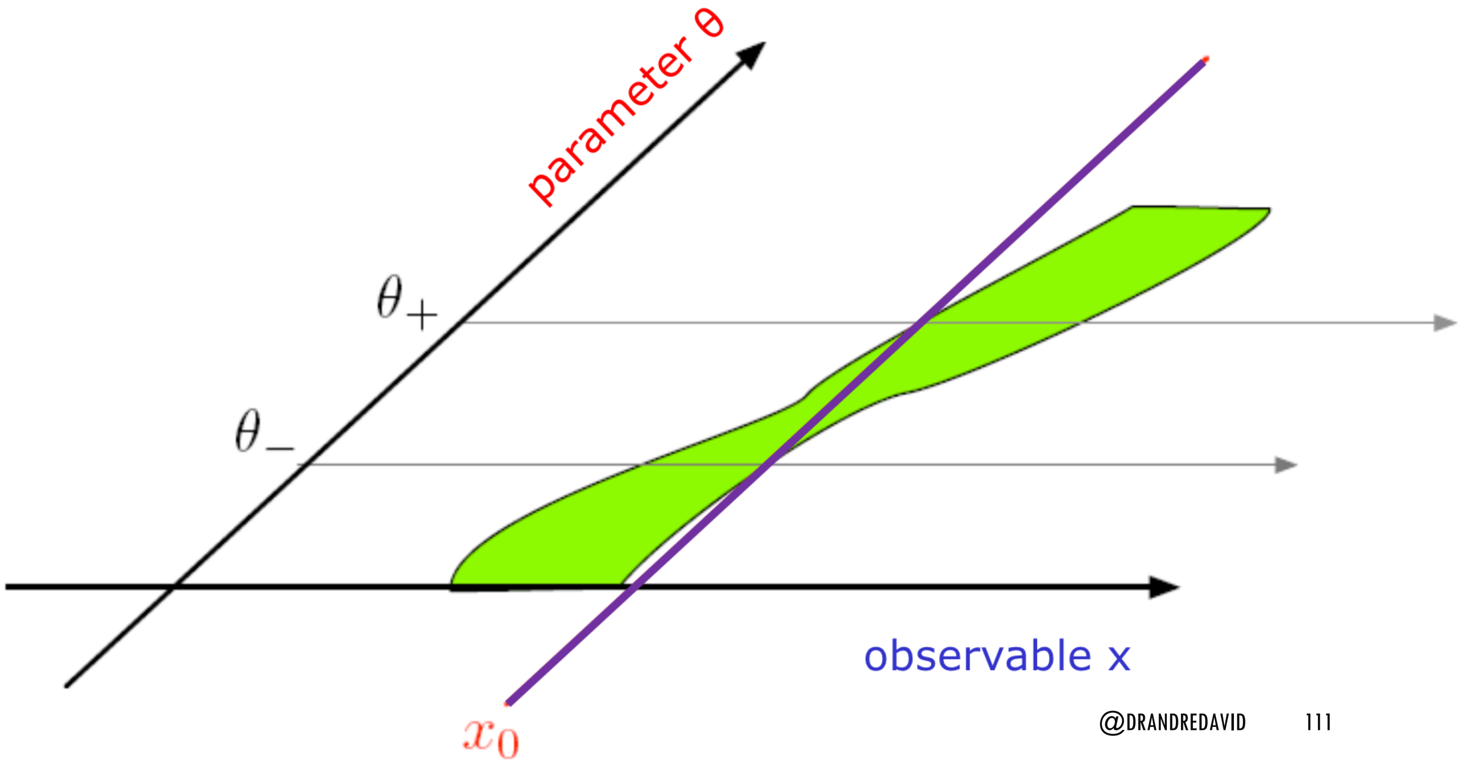
- Now make an acceptance interval in observable x for each value of parameter θ



NEYMAN CONSTRUCTION DECONSTRUCTED

ℹ️ Neyman invented this procedure as a “quality control” procedure. His goal was to guarantee that intervals from different people would be comparable.

- The confidence belt can constructed *in advance of any measurement*, it is a property of the model, not the data
- Given a measurement x_0, a confidence interval $[\theta_+,\theta_-]$ can be constructed as follows
- The interval $[\theta_-,\theta_+]$ has a 68% probability to cover the true value



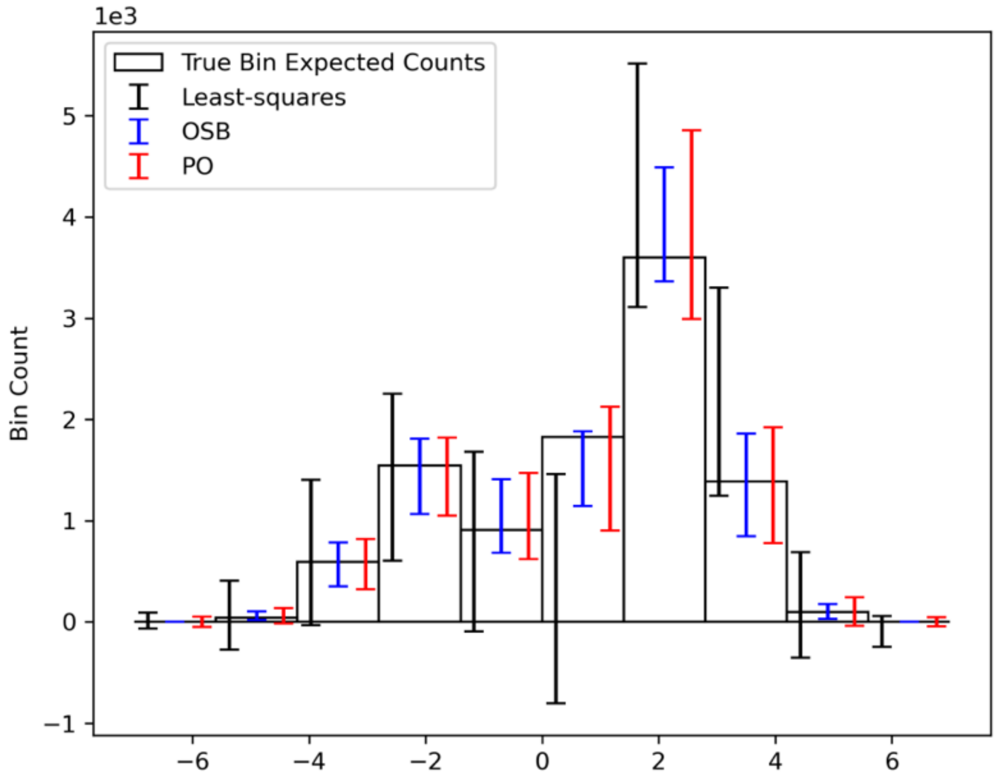
ADDITIONAL TOPICS

PROGRESS IN UNFOLDING

Important physics tool for theory-experiment communication.

- Avoids theorists having to turn their calculations into full-fledged simulations.

Exciting progress with many open questions for future work.

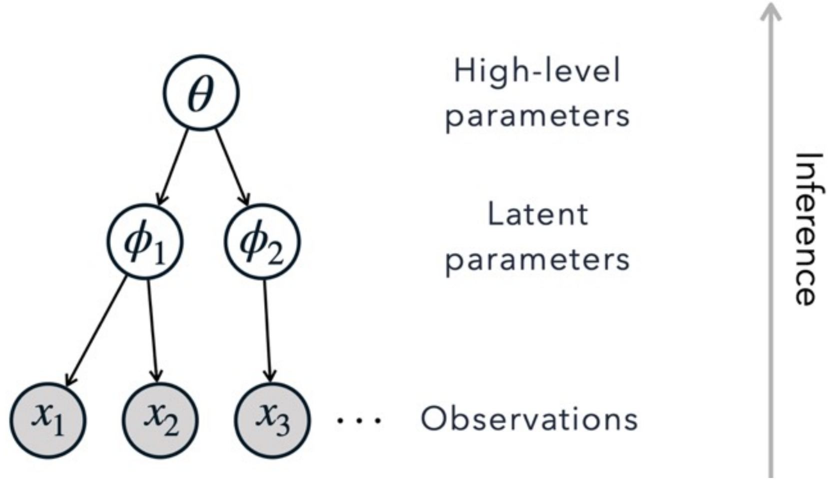


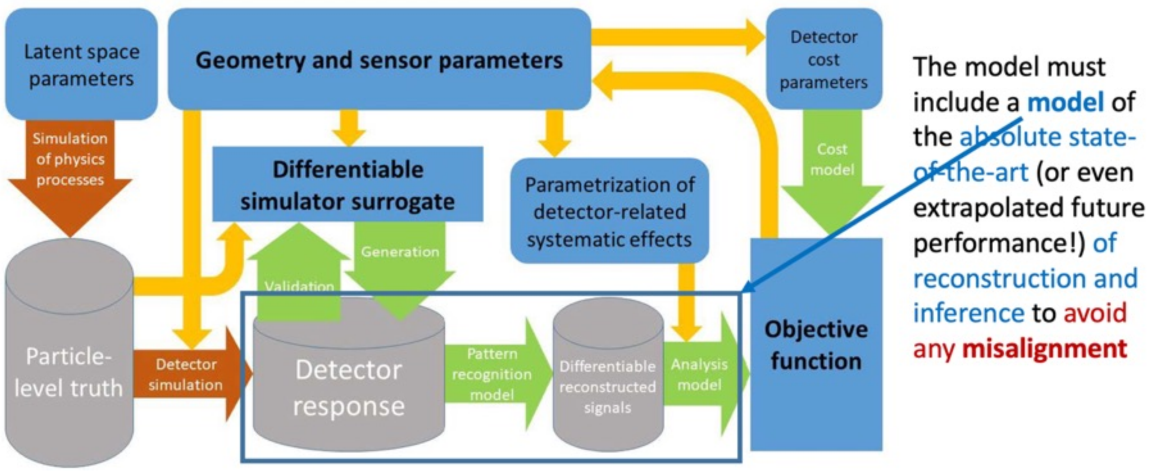
HIERARCHIES TO DIVIDE AND CONQUER

Specifying intermediate "quantities of interest" or "observables".

Not new: we calibrate energies of individual hits and reconstruct momenta of individual tracks.

Not a conclusion, just a feeling; a theme.





LHC'S BUT ONE CORNER OF PARTICLE PHYSICS

Specific issues that deserve just as much attention from statisticians.

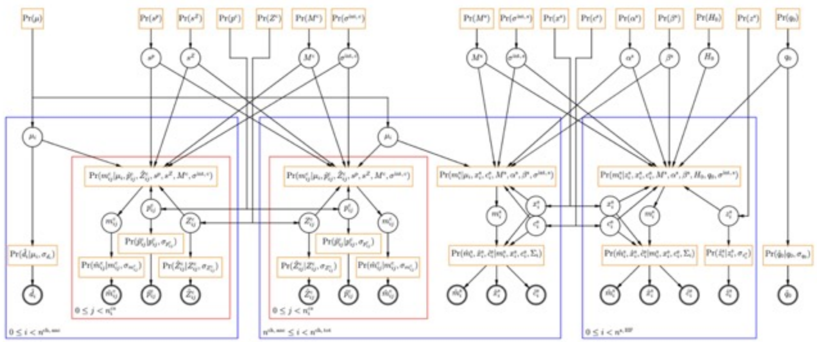
- Fertile (safe?, welcoming?) ground for Bayesian methods.

LHCb and Belle II: measurement style

For measuring $\mathscr{B}$ of rare B-decays **LHCb uses mostly relative** and **Belle II absolute approach:**

- $\mathscr{B}(B^+ \to \mu^+\mu^-\mu^+\nu) = \mathscr{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+) \times \frac{\epsilon(B^+ \to J/\psi(\to \mu^+\mu^-)K^+)}{\epsilon(B^+ \to \mu^+\mu^-\mu^+\nu)} \times \frac{N(B^+ \to \mu^+\mu^-\mu^+\nu)}{N(B^+ \to J/\psi(\to \mu^+\mu^-)K^+)}$
- $\mathscr{B}(B^+ \to K^+\nu\bar{\nu}) = \frac{N(B^+ \to K^+\nu\bar{\nu})}{\epsilon(B^+ \to K^+\nu\bar{\nu})}$

$$N^i_{pred} = \int_{E_{\min}}^{E_{\max}} \overbrace{P(\nu_\alpha \to \nu_\beta)}^{\text{signal}} \times \underbrace{\overbrace{\Phi(E_\nu)}^{\text{beam}} \times \overbrace{\sigma(E_\nu), \vec{x})}^{\text{cross section}} \times \overbrace{\epsilon(\vec{x})}^{\text{detector}}}_{\text{nuisance parameters}} dE_\nu$$



[Feeney+ MNRAS 2018]

THE BULK AND THE TAILS

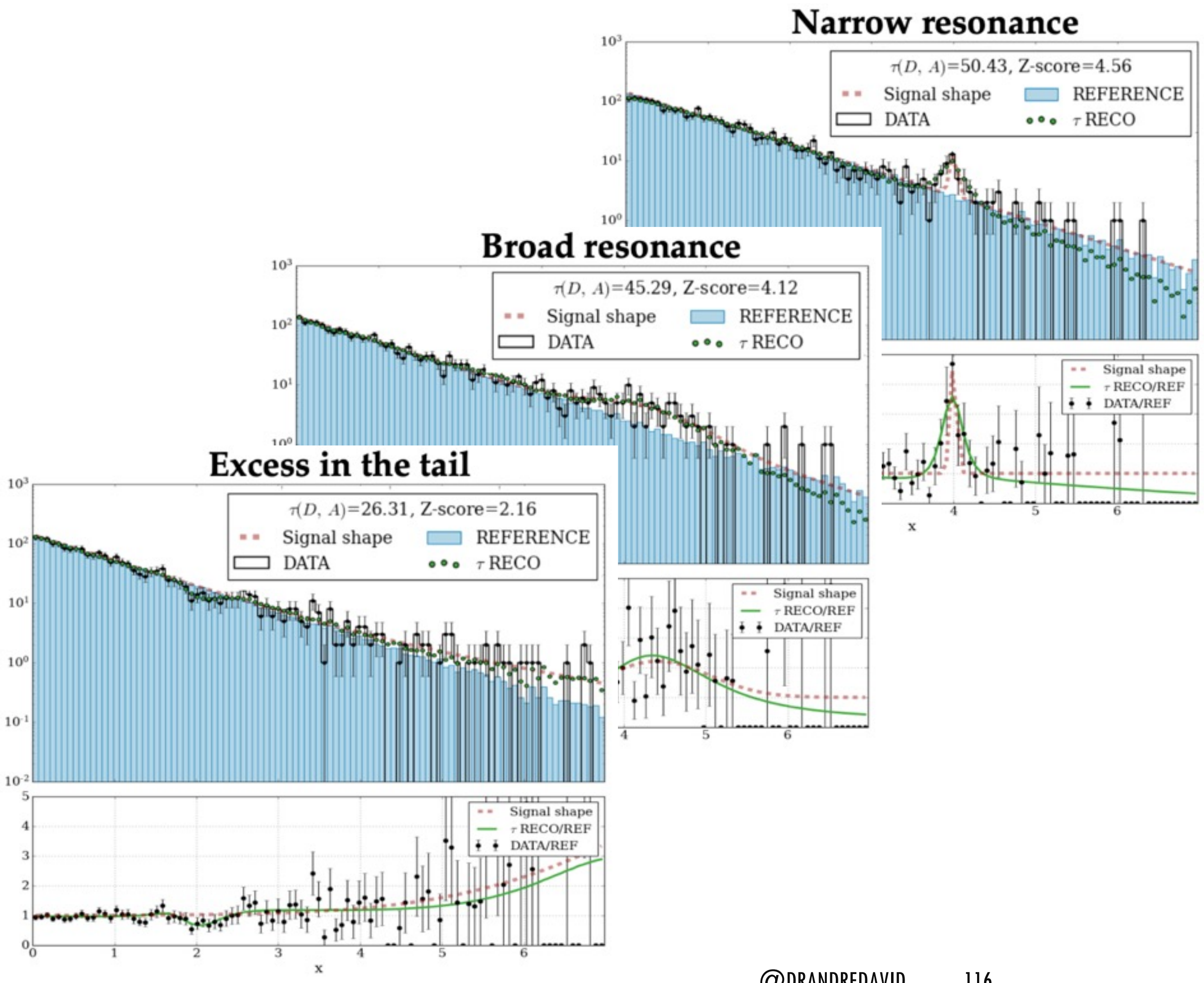
BSM physics unlikely to be the obvious stuff already looked for in the last 40 years.

- Must be within reach and be very subtle (bulk), or
- Out of reach and very energetic (tails).

Requiring same support as the SM simulation **does not cover second case above**.

- I.e. events beyond SM sim. support that could still be SM.
- Connected also to amount of SM sim. that can be afforded.

Can outlier estimation come to the rescue?



ML FOR AI – I.E. FOR ACTUAL INTELLIGENCE*

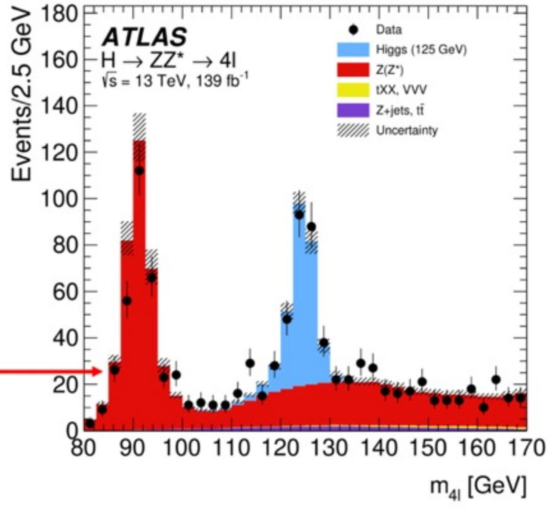
Progress: agreement that optimality and correctness are not the same.

- 90% of cases.
  - Can live with consequences.
- 10% of cases.
  - Can have dire consequences.

Question of optimality:

- Did ML get best reconstruction or event selection?
- Effects definition of discriminating variables, but doesn't affect compatibility with data

Things that affect $p(\cdot|\,\lambda(\theta))$



Questions of correctness:

- Did ML learn an accurate fast simulation?
- Did ML learn a good background estimate?
- Effects statistical model & compatibility with data!

EPJC 80 (2020) 942

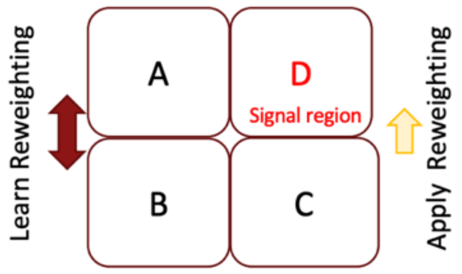
ML FOR AI – I.E. FOR ACTUAL INTELLIGENCE

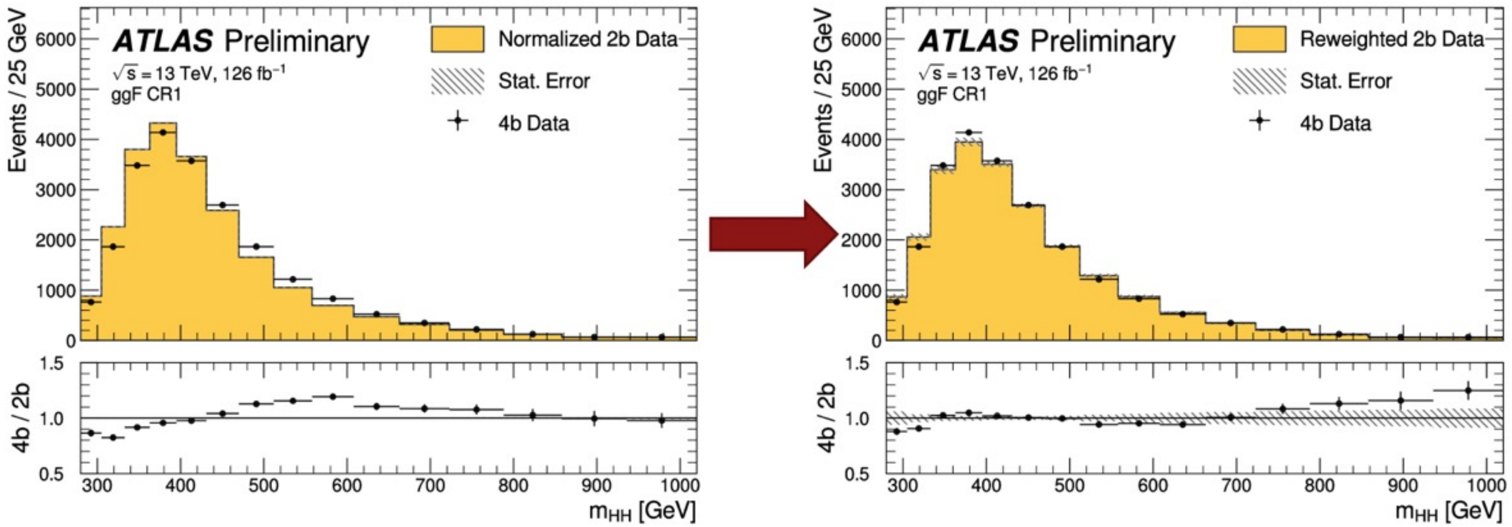
Progress: agreement that optimality and correctness are not the same.

- 90% of cases.
  - Can live with consequences.
- **10% of cases.**
  - **Can have dire consequences.**

ABCD excellent playground to test and learn.

- Also on density learning vs OT mapping.





ML FOR AI – I.E. FOR ACTUAL INTELLIGENCE

Large potential and broad applicability

- Detector operation.
- Construct observables.
- Detector designs.
- Model-independent methods vs SM sim. statistics.
- Skirt systematically-affected phase spaces.
- …

My take: algorithms can more easily explore outside the box **iff** we manage to write loss functions that can do that.
Also, ML is not yet wise.