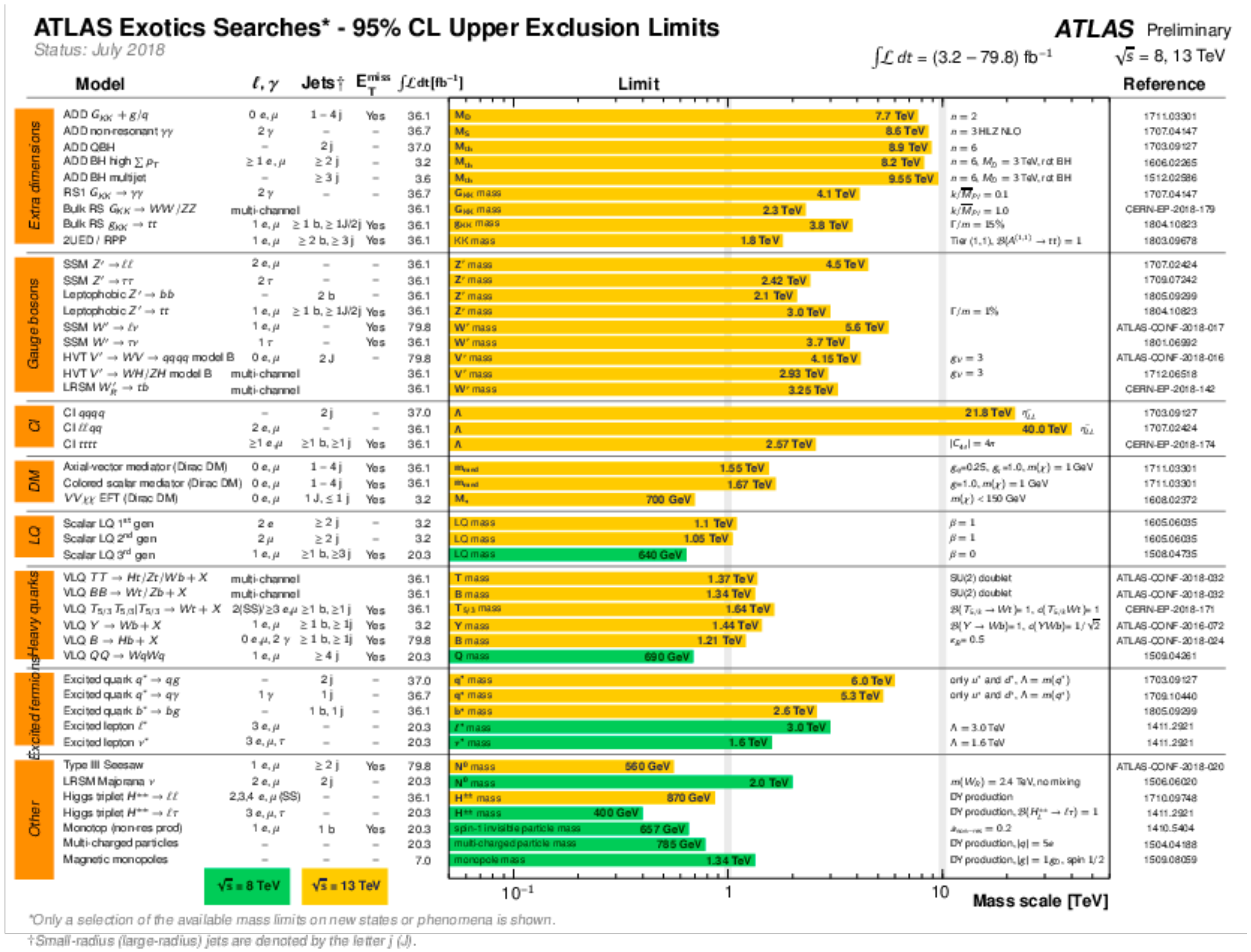
Getting serious about systematics with Machine Learning

Daniel Whiteson, UC Irvine

Oct 2022

Motivation



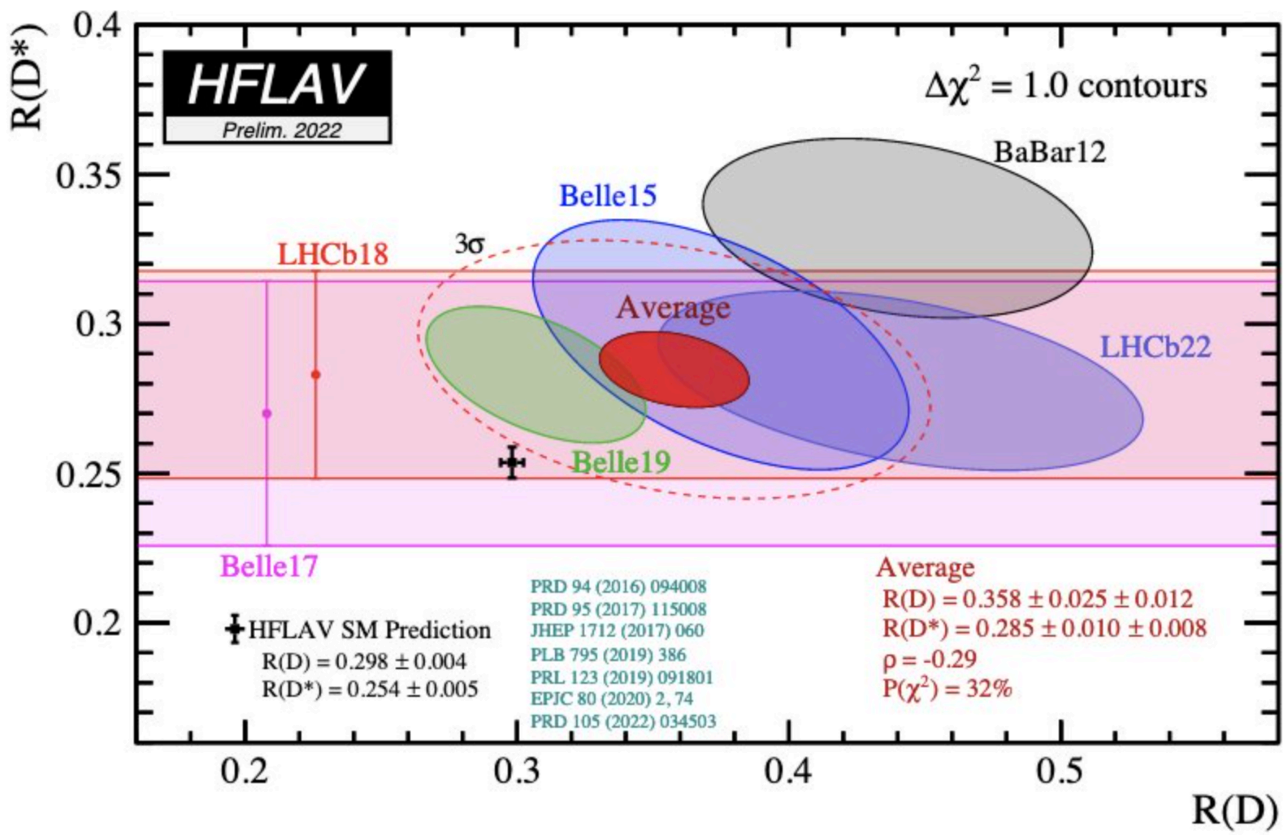
There are no clear post-Higgs discoveries

Direct searches

The path forward

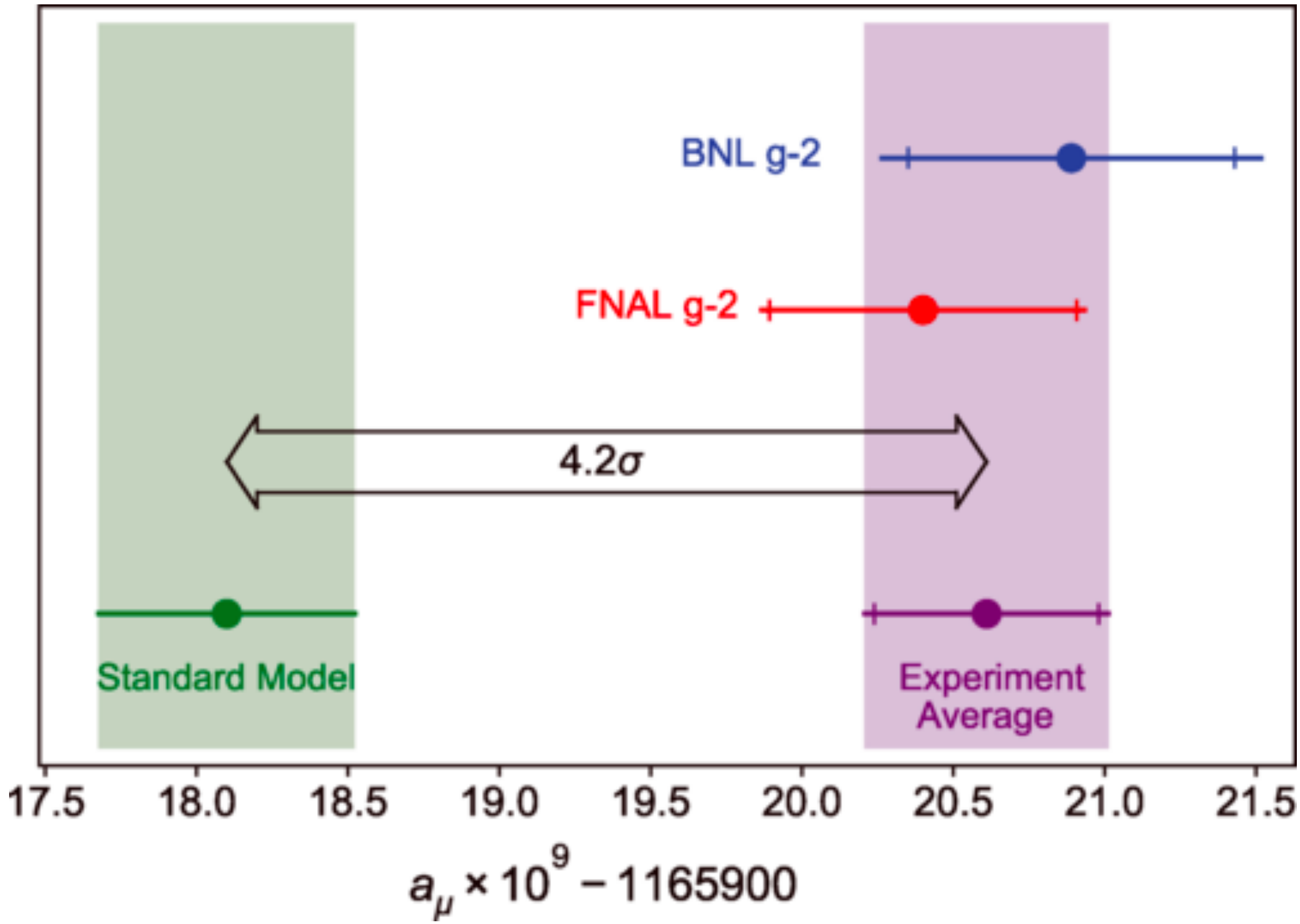
This era will be defined by searches for subtle deviations from the SM

Flavor anomalies

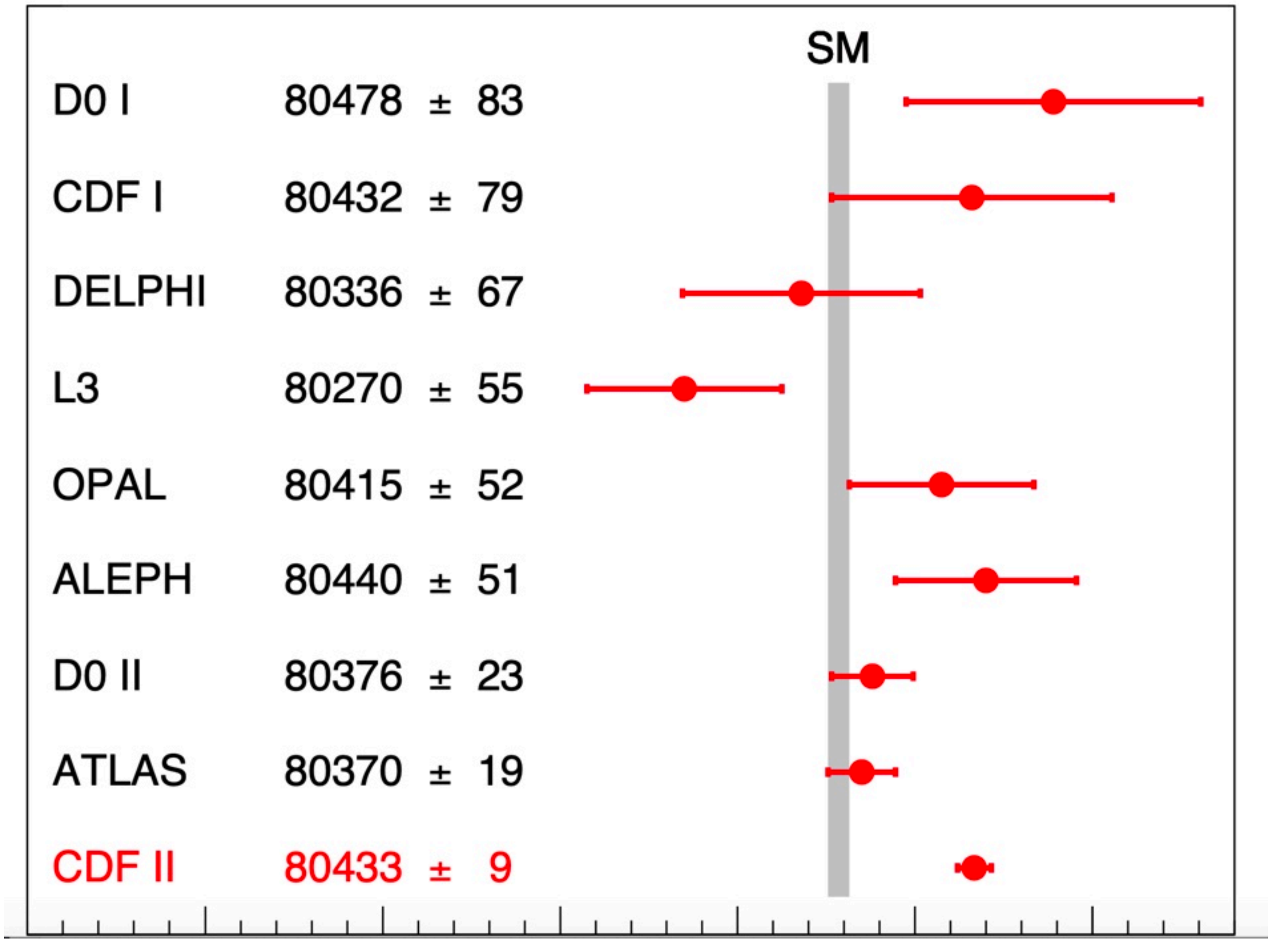


- New preliminary average: slightly lower $\mathcal{R}(D^*)$, slightly higher $\mathcal{R}(D)$, reduced correlation
  - $3.3\sigma \to 3.2\sigma$ agreement with SM
  - Excellent overall agreement between measurements

g-2

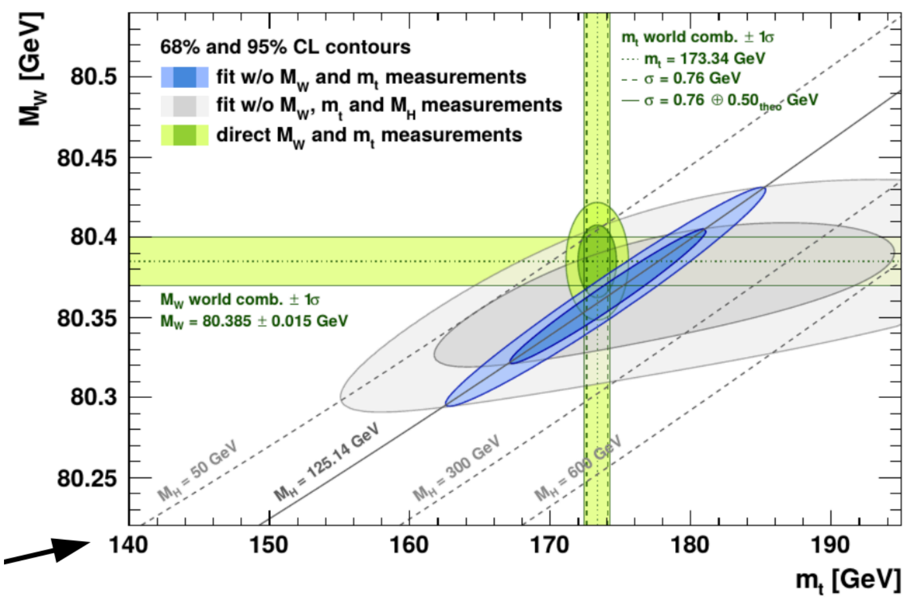

BNL g-2
FNAL g-2
4.2σ
Standard Model
Experiment Average
17.5
18.0
18.5
19.0
19.5
20.0
20.5
21.0
21.5
$a_\mu \times 10^9 - 1165900$

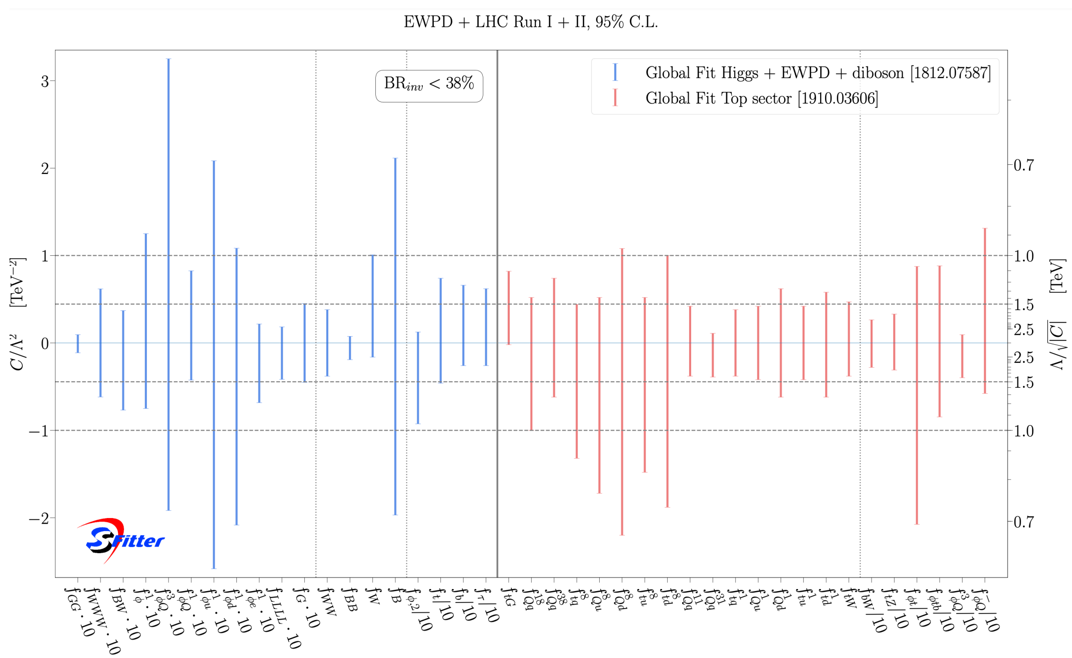
CDF W mass



"with a significance of 7.0σ" (p=2.5e-12)

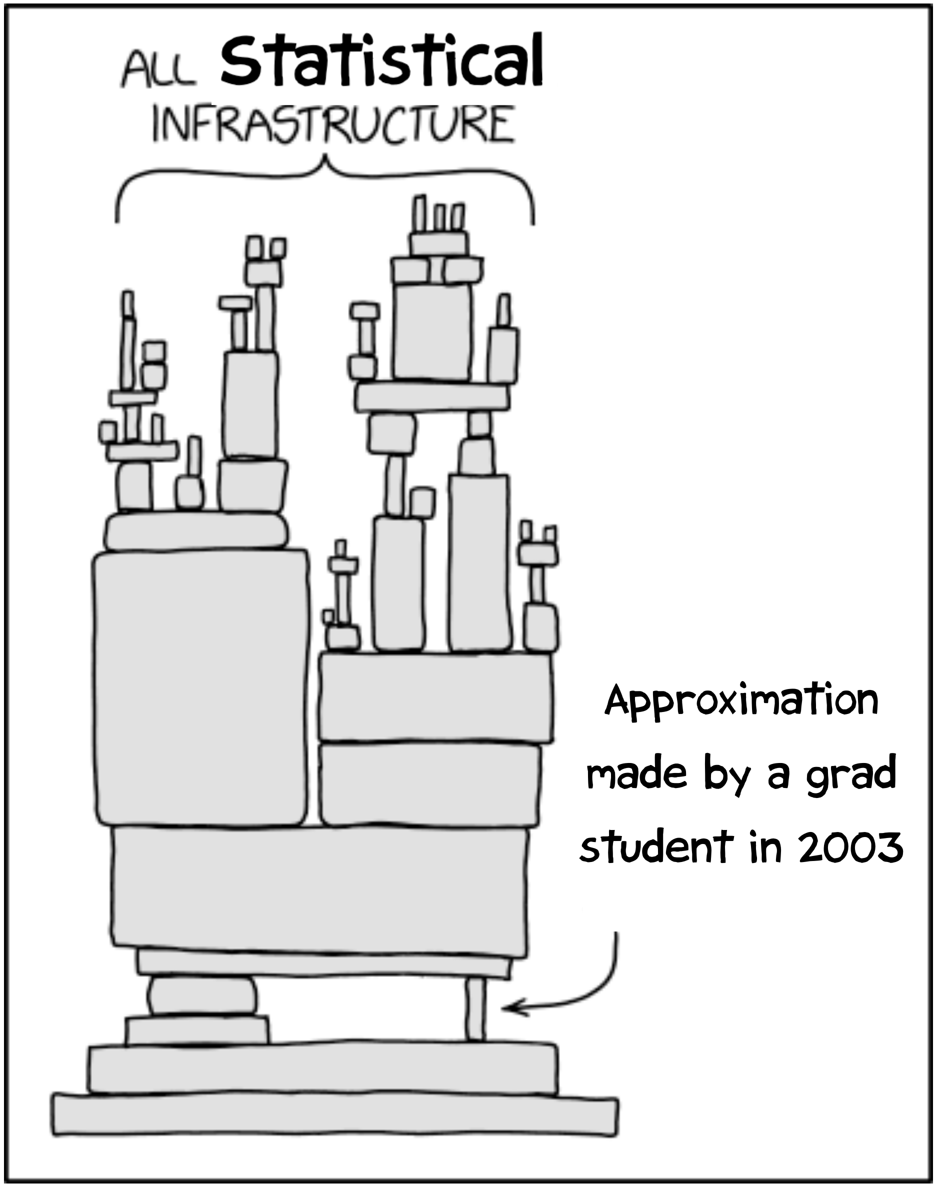
Global fits





Systematic uncertainties
will be a crucial component

NP infrastructure



Time to re-examine some of the underlying pieces

Are they up to the task of the precision era?

Outline

1. Theoretical

2. Experimental

Sources of Systematics

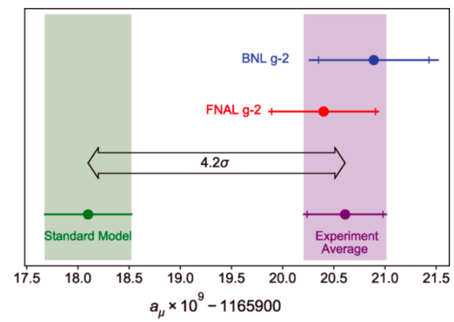
Where do systematic uncertainties come from?

Aux data

- Measurement in orthogonal dataset
- Bayesian: latent variable with a prior
- Has statistical uncertainty
- Repeated measurements give diff results
- Eg detector calibration from dedicated beam
- Sometimes theory uncertainties!
  - eg:https://indico.cern.ch/event/565930/contributions/2371310/attachments/1387348/2111830/HFSF16_Nachman.pdf)

Sources of Systematics

Where do systematic uncertainties come from?



Aux data

Theoretical uncertainties

- Lack of first principles prediction (showering)
- Inability to do infinite-order calculations
- Repeated measurements may give same results
    - If no stochastic source

How they are used

Unified treatment

Treated as random variables

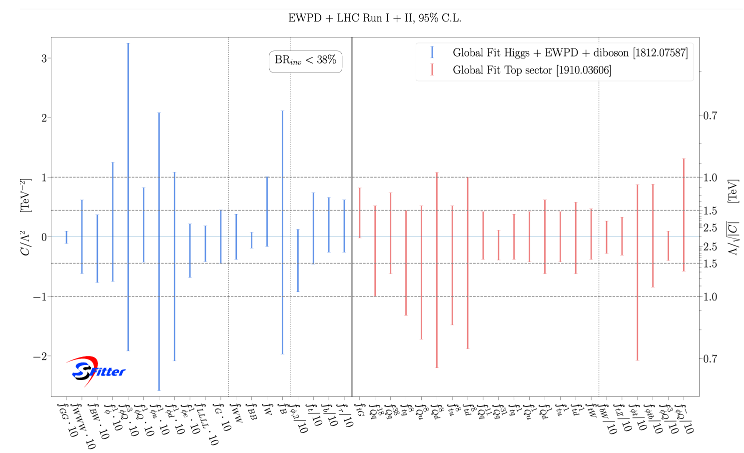
With some aux likelihood/prior

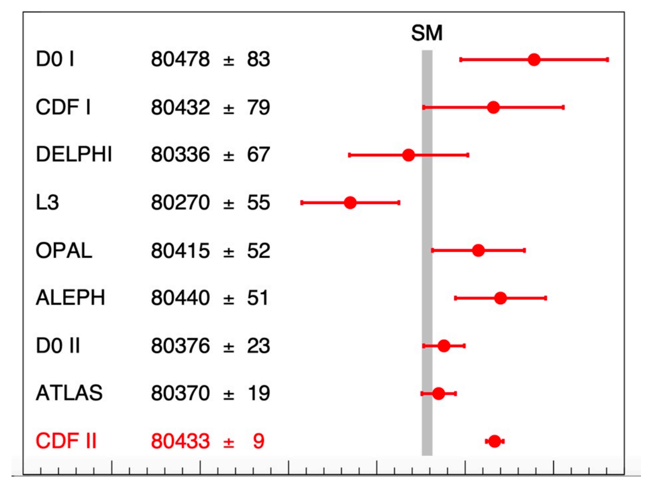
Marginalize/profile/pivot

Eg: 2105.08742

Does that work for theory?

What are the distributions?



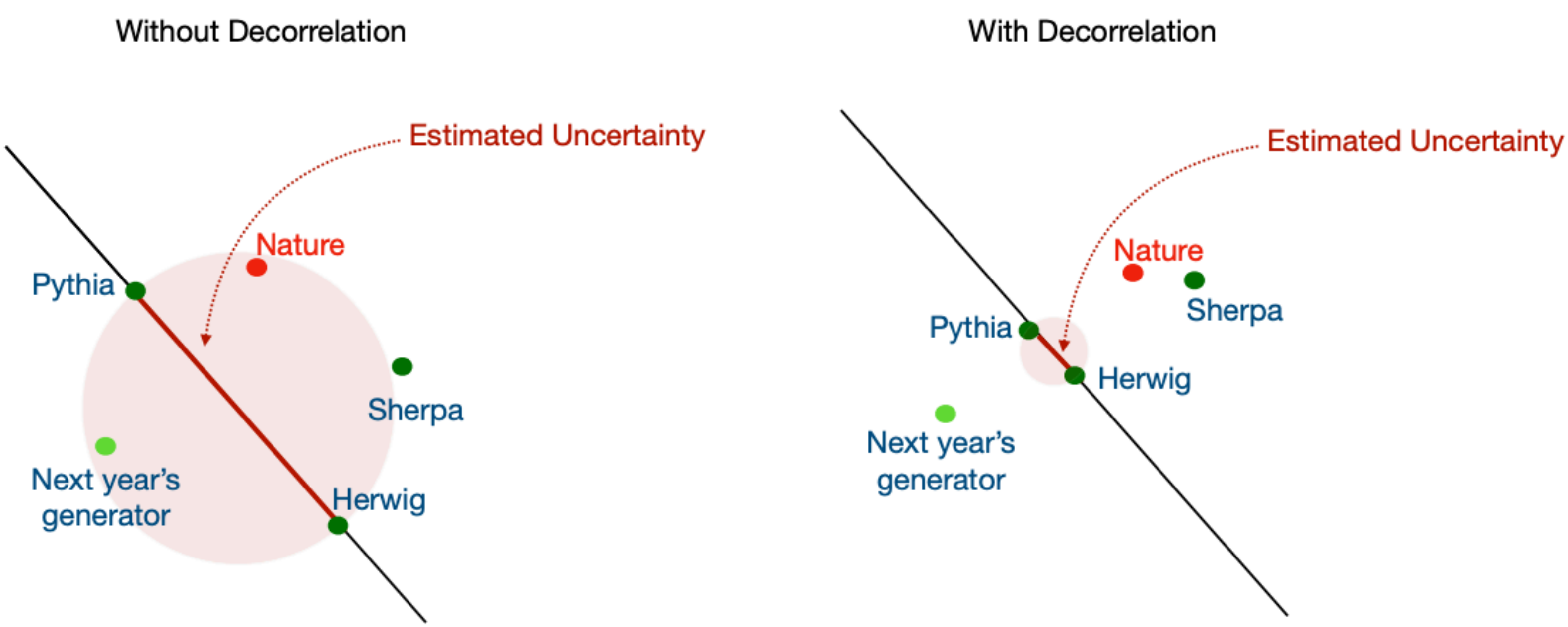


Cautionary tale

2109.08159

A Cautionary Tale of Decorrelating Theory Uncertainties

Aishik Ghosh[a,b] **and Benjamin Nachman**[b,c]



(N)LO calculations

Incomplete calculations

Statistical Patterns of Theory Uncertainties

Aishik Ghosh[1,2], Benjamin Nachman[1,3], Tilman Plehn[4], Lily Shire[2],
Tim M.P. Tait[2], and Daniel Whiteson[2]

220x.xxxxx
(Submitted to arXiv today!)

(N)LO calculations

Incomplete calculations

We want to know the rate of processes

$$\sigma(\theta) \approx \sum_{a,b} \int dx_a x_b f_a(x_a;\mu_F) f_b(x_b;\mu_F) \hat{\sigma}_{ab}(\theta;\mu_F,\mu_R),$$

But the partonic cross-section

$$\hat{\sigma}_{ab}(\theta;\mu_F,\mu_R)$$

Is expressed as a perturbative sum

Scales

At each order, there are issues

Infrared+collinear divergences

Absorbed into parton densities

Gives unphysical scale μ_F

Ultraviolet divergences

removed by renormalization to cutoff μ_R

Artifact of truncation of series—
scale dependence should vanish at all orders

Scales

What is the error in cross-section due to truncation?

$$\sigma \in [\sigma_-, \sigma_+] \equiv \left[\sigma(\mu_{R,+}), \sigma(\mu_{R,-})\right] ,$$

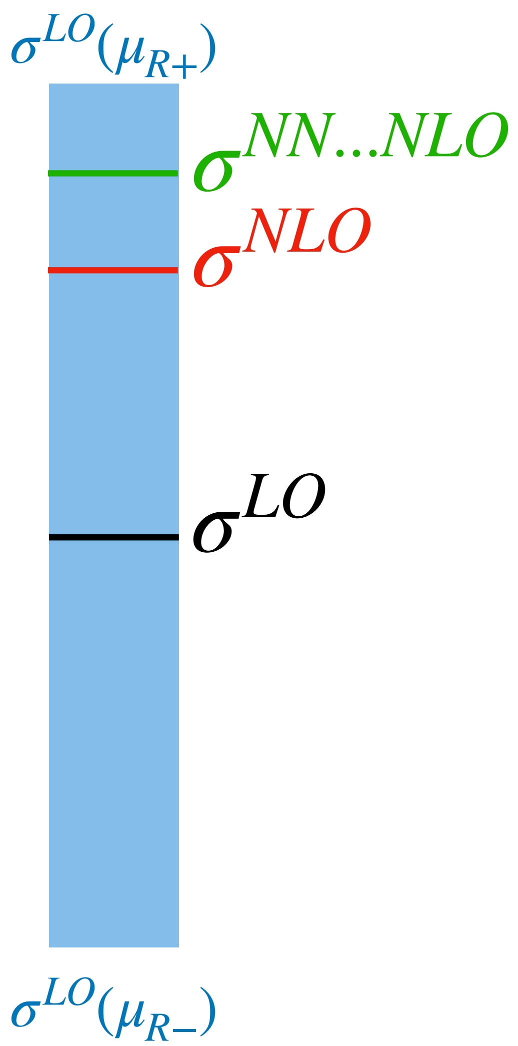
$$\mu_0 = \frac{H_\mathrm{T}}{2} = \frac{1}{2} \sum_{\text{final state}} \sqrt{p_\mathrm{T}^2 + m^2}$$

$$\mu_{R,+} = 2 \quad \mu_0$$

$$\mu_{R,-} = 1/2 \quad \mu_0$$

Use dependence on scale to estimate uncertainty

Open questions



Can you feed this into your stats package like an uncertainty?

What is distribution of LO relative to NLO?

How accurate is uncertainty?

See also: 1105.5152, 2006.16293,1409.5036

Pull

Use pull to examine

$$t_{\text{scale}} = \frac{\sigma_{\text{NLO}} - \sigma_{\text{LO}}}{\Delta\sigma_{\text{LO scale}}}$$

Critical issue:

need a large (>>10)
set of processes calculated
under identical conditions

See also: 1105.5152, 2006.16293,1409.5036

Madgraph paper

The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations

J. Alwall[a], R. Frederix[b], S. Frixione[b], V. Hirschi[c], F. Maltoni[d], O. Mattelaer[d], H.-S. Shao[e], T. Stelzer[f], P. Torrielli[g], M. Zaro[hi]

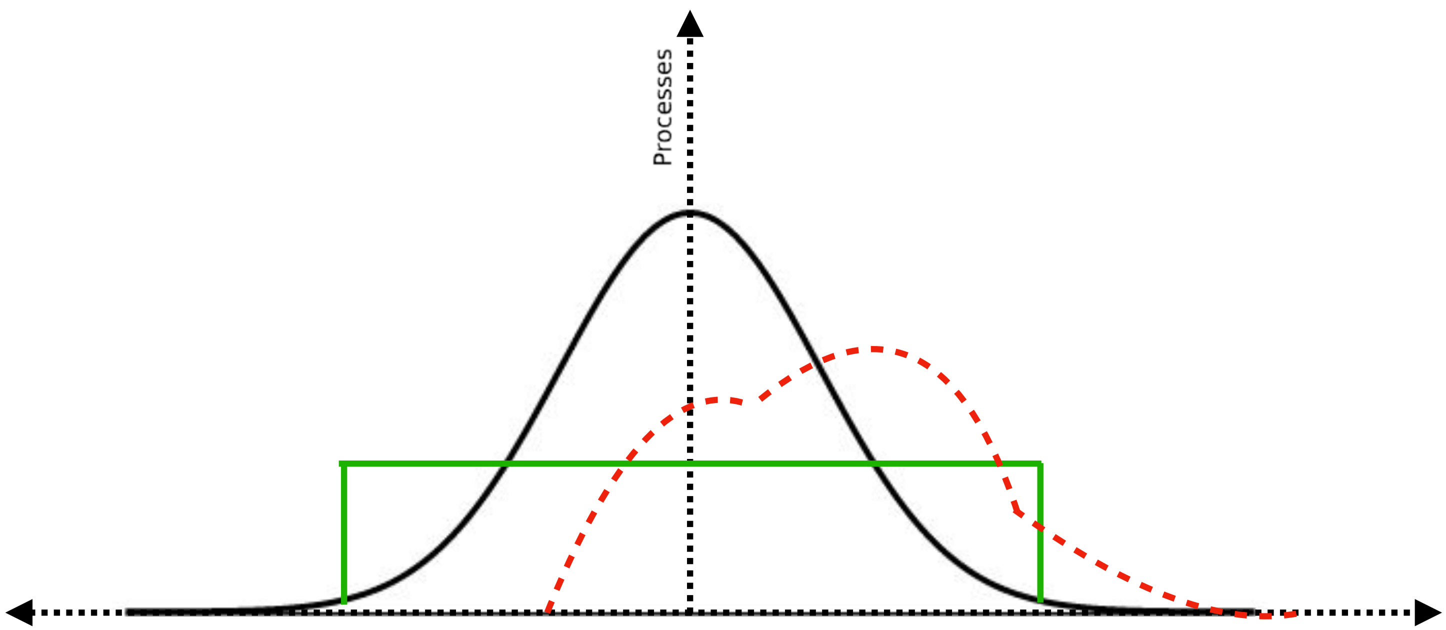
Process		Syntax	Cross section (pb)					
Vector boson +jets			LO 13 TeV			NLO 13 TeV		
a.1	$pp \to W^{\pm}$	`p p > wpm`	$1.375 \pm 0.002 \cdot 10^{5}$	$^{+15.4\%}_{-16.6\%}$	$^{+2.0\%}_{-1.6\%}$	$1.773 \pm 0.007 \cdot 10^{5}$	$^{+5.2\%}_{-9.4\%}$	$^{+1.9\%}_{-1.6\%}$
a.2	$pp \to W^{\pm} j$	`p p > wpm j`	$2.045 \pm 0.001 \cdot 10^{4}$	$^{+19.7\%}_{-17.2\%}$	$^{+1.4\%}_{-1.1\%}$	$2.843 \pm 0.010 \cdot 10^{4}$	$^{+5.9\%}_{-8.0\%}$	$^{+1.3\%}_{-1.1\%}$
a.3	$pp \to W^{\pm} jj$	`p p > wpm j j`	$6.805 \pm 0.015 \cdot 10^{3}$	$^{+24.5\%}_{-18.6\%}$	$^{+0.8\%}_{-0.7\%}$	$7.786 \pm 0.030 \cdot 10^{3}$	$^{+2.4\%}_{-6.0\%}$	$^{+0.9\%}_{-0.8\%}$
a.4	$pp \to W^{\pm} jjj$	`p p > wpm j j j`	$1.821 \pm 0.002 \cdot 10^{3}$	$^{+41.0\%}_{-27.1\%}$	$^{+0.5\%}_{-0.5\%}$	$2.005 \pm 0.008 \cdot 10^{3}$	$^{+0.9\%}_{-6.7\%}$	$^{+0.6\%}_{-0.5\%}$
a.5	$pp \to Z$	`p p > z`	$4.248 \pm 0.005 \cdot 10^{4}$	$^{+14.6\%}_{-15.8\%}$	$^{+2.0\%}_{-1.6\%}$	$5.410 \pm 0.022 \cdot 10^{4}$	$^{+4.6\%}_{-8.6\%}$	$^{+1.9\%}_{-1.5\%}$
a.6	$pp \to Zj$	`p p > z j`	$7.209 \pm 0.005 \cdot 10^{3}$	$^{+19.3\%}_{-17.0\%}$	$^{+1.2\%}_{-1.0\%}$	$9.742 \pm 0.035 \cdot 10^{3}$	$^{+5.8\%}_{-7.8\%}$	$^{+1.2\%}_{-1.0\%}$
a.7	$pp \to Zjj$	`p p > z j j`	$2.348 \pm 0.006 \cdot 10^{3}$	$^{+24.3\%}_{-18.5\%}$	$^{+0.6\%}_{-0.6\%}$	$2.665 \pm 0.010 \cdot 10^{3}$	$^{+2.5\%}_{-6.0\%}$	$^{+0.7\%}_{-0.7\%}$
a.8	$pp \to Zjjj$	`p p > z j j j`	$6.314 \pm 0.008 \cdot 10^{2}$	$^{+40.8\%}_{-27.0\%}$	$^{+0.5\%}_{-0.5\%}$	$6.996 \pm 0.028 \cdot 10^{2}$	$^{+1.1\%}_{-6.8\%}$	$^{+0.5\%}_{-0.5\%}$
a.9	$pp \to \gamma j$	`p p > a j`	$1.964 \pm 0.001 \cdot 10^{4}$	$^{+31.2\%}_{-26.0\%}$	$^{+1.7\%}_{-1.8\%}$	$5.218 \pm 0.025 \cdot 10^{4}$	$^{+24.5\%}_{-21.4\%}$	$^{+1.4\%}_{-1.6\%}$
a.10	$pp \to \gamma jj$	`p p > a j j`	$7.815 \pm 0.008 \cdot 10^{3}$	$^{+32.8\%}_{-24.2\%}$	$^{+0.9\%}_{-1.2\%}$	$1.004 \pm 0.004 \cdot 10^{4}$	$^{+5.9\%}_{-10.9\%}$	$^{+0.8\%}_{-1.2\%}$

+127 more pp processes from 1405.0301!

(Not a random sampling)

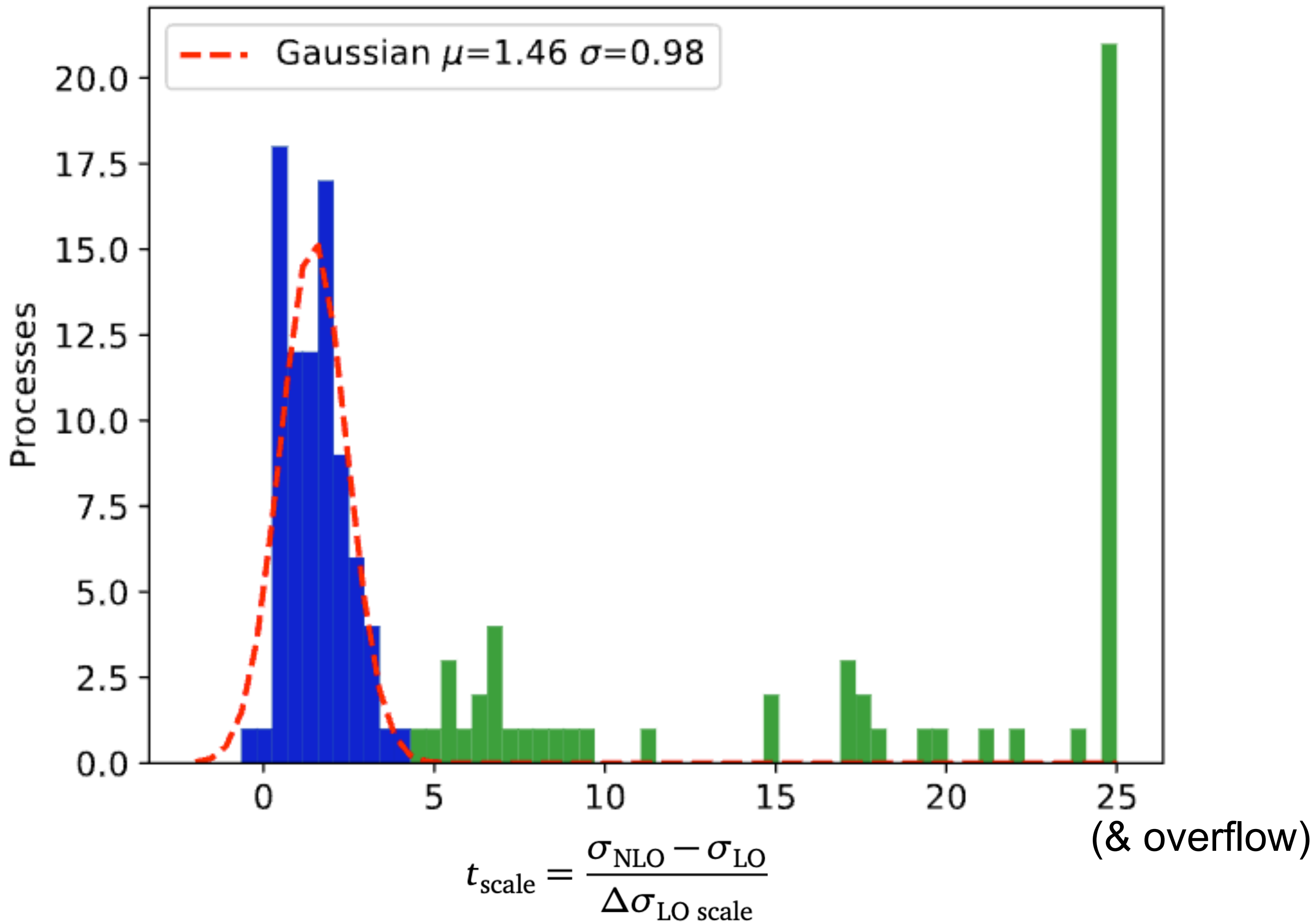
What does it look like?

Which of these distributions do you expect?

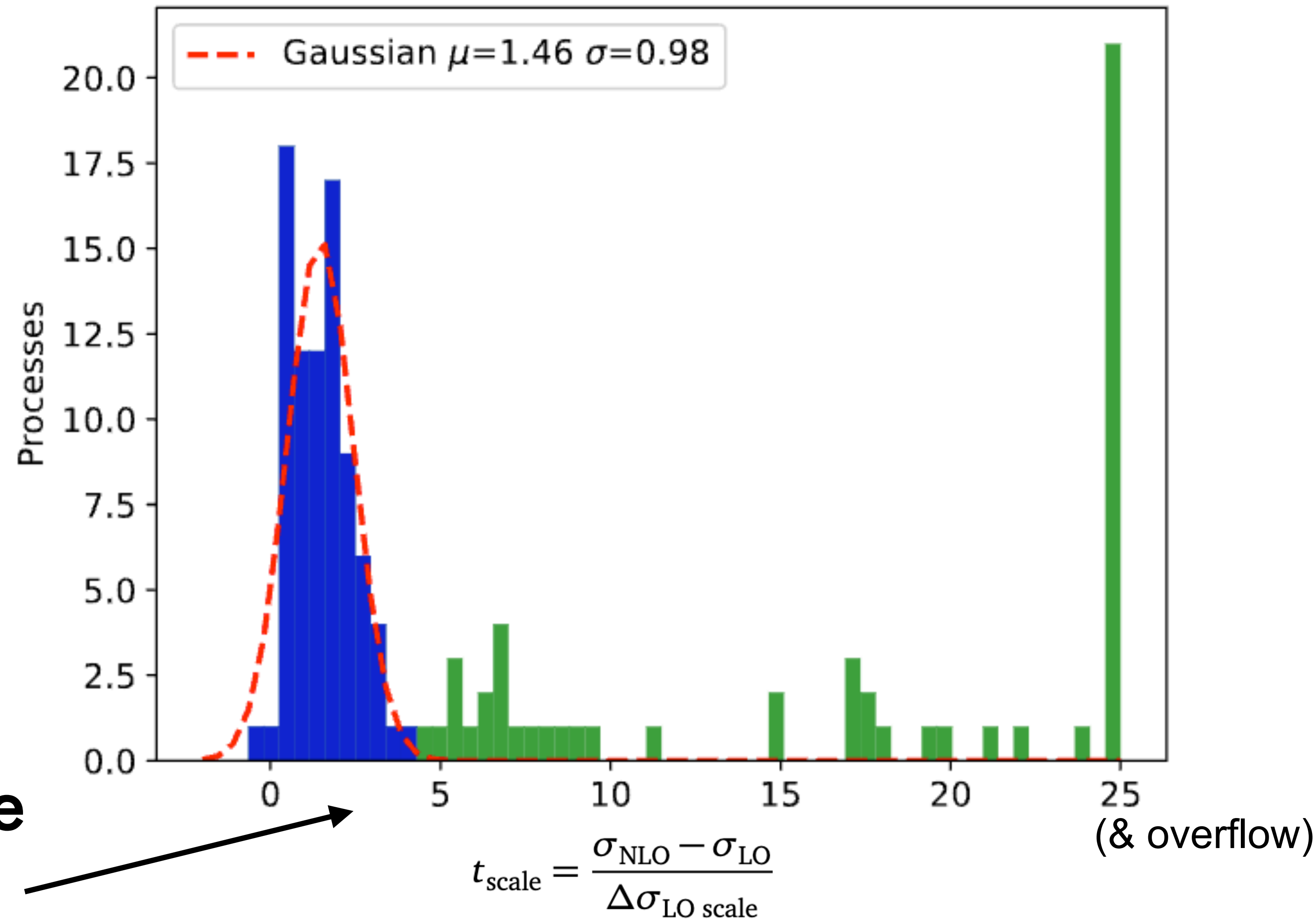


$$t_{\text{scale}} = \frac{\sigma_{\text{NLO}} - \sigma_{\text{LO}}}{\Delta\sigma_{\text{LO scale}}}$$

The pull

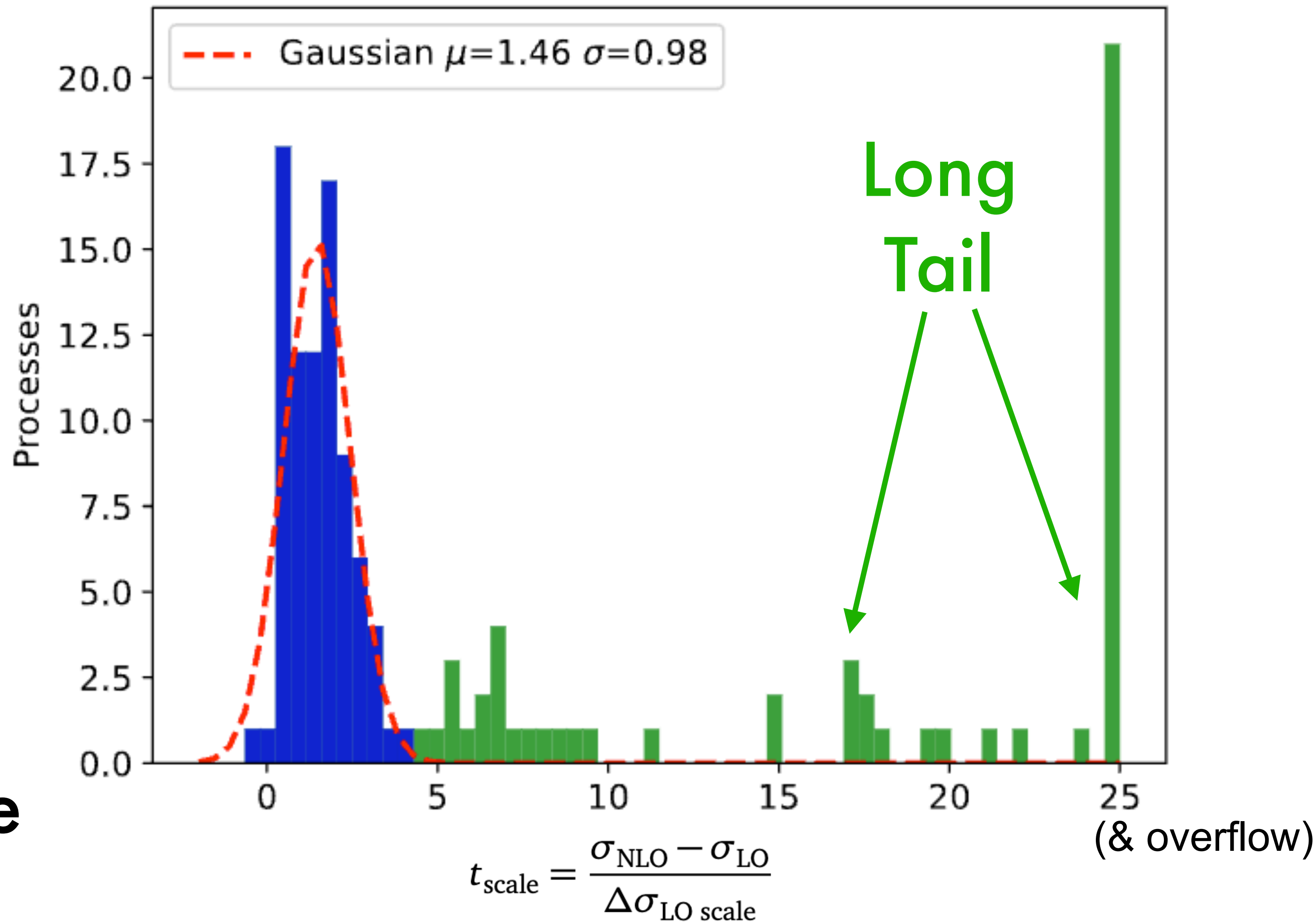

Gaussian μ=1.46 σ=0.98
Processes
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
5
10
15
20
25
(& overflow)
$t_{\mathrm{scale}} = \frac{\sigma_{\mathrm{NLO}} - \sigma_{\mathrm{LO}}}{\Delta\sigma_{\mathrm{LO\ scale}}}$

The pull



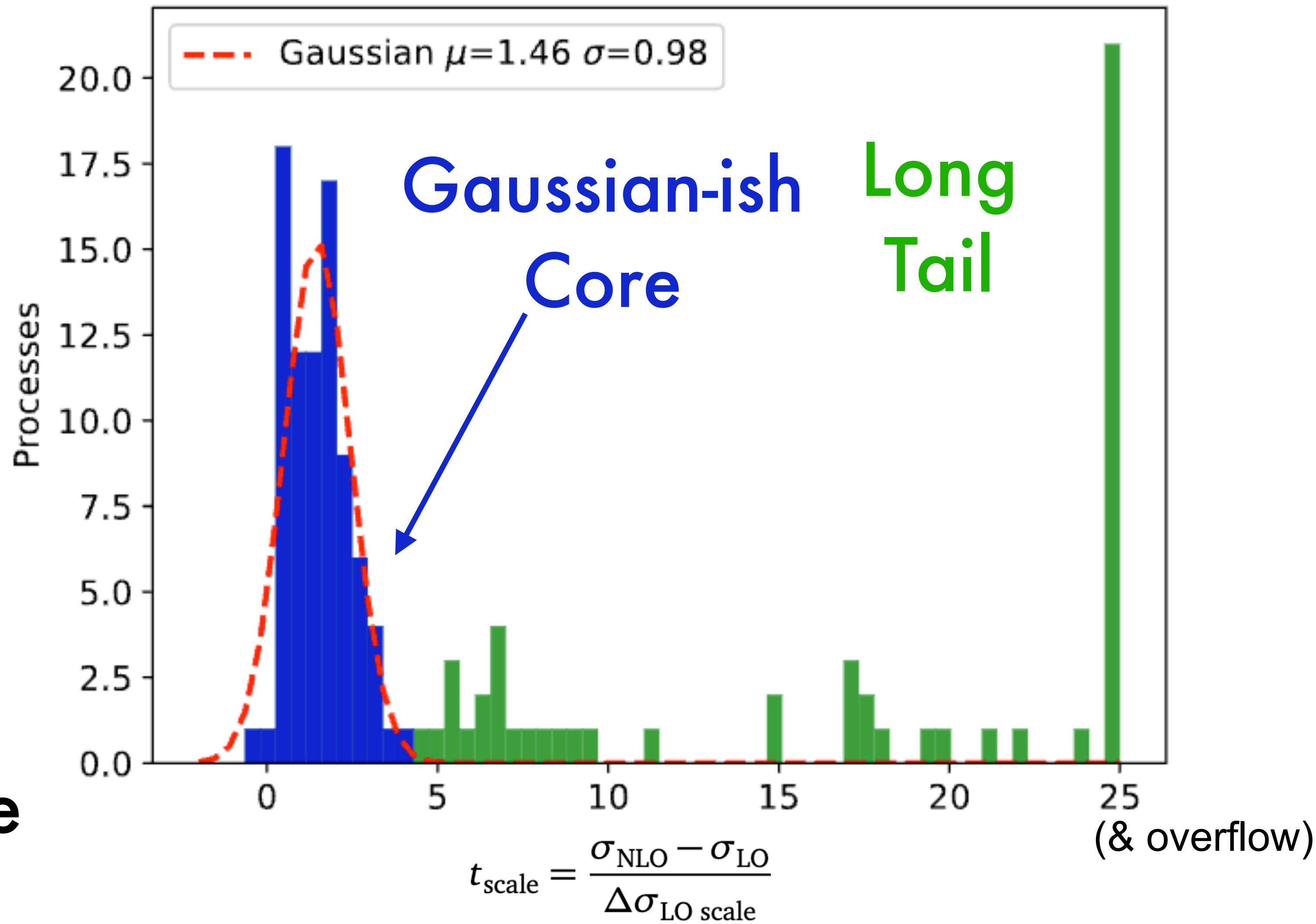
Positive Mean

The pull



Positive
Mean

The pull



Positive Mean

The tail

Process	n_{part}	$\Delta\sigma/\sigma_0$	$\frac{\sigma_{\text{NLO}}-\sigma_0}{\Delta\sigma}$
`p p > wpm`	1	1.54×10^{-1}	1.84
`p p > wpm j`	2	1.97×10^{-1}	1.96
`p p > wpm j j`	3	2.45×10^{-1}	0.59
`p p > wpm j j j`	4	4.10×10^{-1}	0.25
`p p > z`	1	1.46×10^{-1}	1.87
`p p > z j`	2	1.93×10^{-1}	1.82
`p p > z j j`	3	2.43×10^{-1}	0.56
`p p > z j j j`	4	4.08×10^{-1}	0.27
`p p > a j`	2	3.12×10^{-1}	5.33
`p p > a j j`	3	3.28×10^{-1}	0.85
`p p > w+ w- wpm`	3	1.00×10^{-3}	610.69
`p p > z w+ w-`	3	8.00×10^{-3}	92.39
`p p > z z wpm`	3	1.00×10^{-2}	85.00
`p p > z z z`	3	1.00×10^{-3}	302.75
`p p > a w+ w-`	3	1.90×10^{-2}	42.33
`p p > a a wpm`	3	4.40×10^{-2}	47.24
`p p > a z wpm`	3	1.00×10^{-3}	1244.49
`p p > a z z`	3	2.00×10^{-2}	17.24

The tail

Process	n_{part}	$\Delta\sigma/\sigma_0$	$\frac{\sigma_{\text{NLO}}-\sigma_0}{\Delta\sigma}$
p p > wpm	1	1.54×10^{-1}	1.84
p p > wpm j	2	1.97×10^{-1}	1.96
p p > wpm j j	3	2.45×10^{-1}	0.59
p p > wpm j j j	4	4.10×10^{-1}	0.25
p p > z	1	1.46×10^{-1}	1.87
p p > z j	2	1.93×10^{-1}	1.82
p p > z j j	3	2.43×10^{-1}	0.56
p p > z j j j	4	4.08×10^{-1}	0.27
p p > a j	2	3.12×10^{-1}	5.33
p p > a j j	3	3.28×10^{-1}	0.85
p p > w+ w- wpm	3	1.00×10^{-3}	610.69
p p > z w+ w-	3	8.00×10^{-3}	92.39
p p > z z wpm	3	1.00×10^{-2}	85.00
p p > z z z	3	1.00×10^{-3}	302.75
p p > a w+ w-	3	1.90×10^{-2}	42.33
p p > a a wpm	3	4.40×10^{-2}	47.24
p p > a z wpm	3	1.00×10^{-3}	1244.49
p p > a z z	3	2.00×10^{-2}	17.24

What's the pattern?

These are all electroweak

They have no QCD vertices

Scale dependence absent at LO

Scale variation is a poor scheme for these processes

Reference processes

QCD processes also have a simple pattern

Process	$\frac{\Delta\sigma}{\sigma_0}$		n	$\frac{\Delta\sigma}{n\,\sigma_0}$	
`p p > j j`	$+2.49\times10^{-1}$	-1.88×10^{-1}	2	$+1.24\times10^{-1}$	-9.40×10^{-2}
`p p > b b`	$+2.52\times10^{-1}$	-1.89×10^{-1}	2	$+1.26\times10^{-1}$	-9.45×10^{-2}
`p p > t t`	$+2.90\times10^{-1}$	-2.11×10^{-1}	2	$+1.45\times10^{-1}$	-1.06×10^{-1}
`p p > j j j`	$+4.38\times10^{-1}$	-2.84×10^{-1}	3	$+1.46\times10^{-1}$	-9.47×10^{-2}
`p p > b b j`	$+4.41\times10^{-1}$	-2.85×10^{-1}	3	$+1.47\times10^{-1}$	-9.50×10^{-2}
`p p > t t j`	$+4.51\times10^{-1}$	-2.90×10^{-1}	3	$+1.50\times10^{-1}$	-9.67×10^{-2}
`p p > b b j j`	$+6.18\times10^{-1}$	-3.56×10^{-1}	4	$+1.54\times10^{-1}$	-8.90×10^{-2}
`p p > b b b b`	$+6.17\times10^{-1}$	-3.56×10^{-1}	4	$+1.54\times10^{-1}$	-8.90×10^{-2}
`p p > t t j j`	$+6.14\times10^{-1}$	-3.56×10^{-1}	4	$+1.53\times10^{-1}$	-8.90×10^{-2}
`p p > t t t t`	$+6.38\times10^{-1}$	-3.65×10^{-1}	4	$+1.60\times10^{-1}$	-9.12×10^{-2}
`p p > t t b b`	$+6.21\times10^{-1}$	-3.57×10^{-1}	4	$+1.55\times10^{-1}$	-8.93×10^{-2}
average				$+1.47\times10^{-1}$	-9.34×10^{-2}

Table 1: Scale dependence for LHC processes with only QCD particles in the final state. For each process, we report the relative scale uncertainty, the number of final state particles, and the per-particle relative scale uncertainty.

New uncertainty

Replace process scale uncertainty

$$\frac{\Delta\sigma_{\text{ref}}}{\sigma_0} = n \times \left\langle \frac{\Delta\sigma}{n\sigma_0} \right\rangle_{\text{QCD}}$$

with uncertainty estimated by average of QCD processes scaled by number of particles

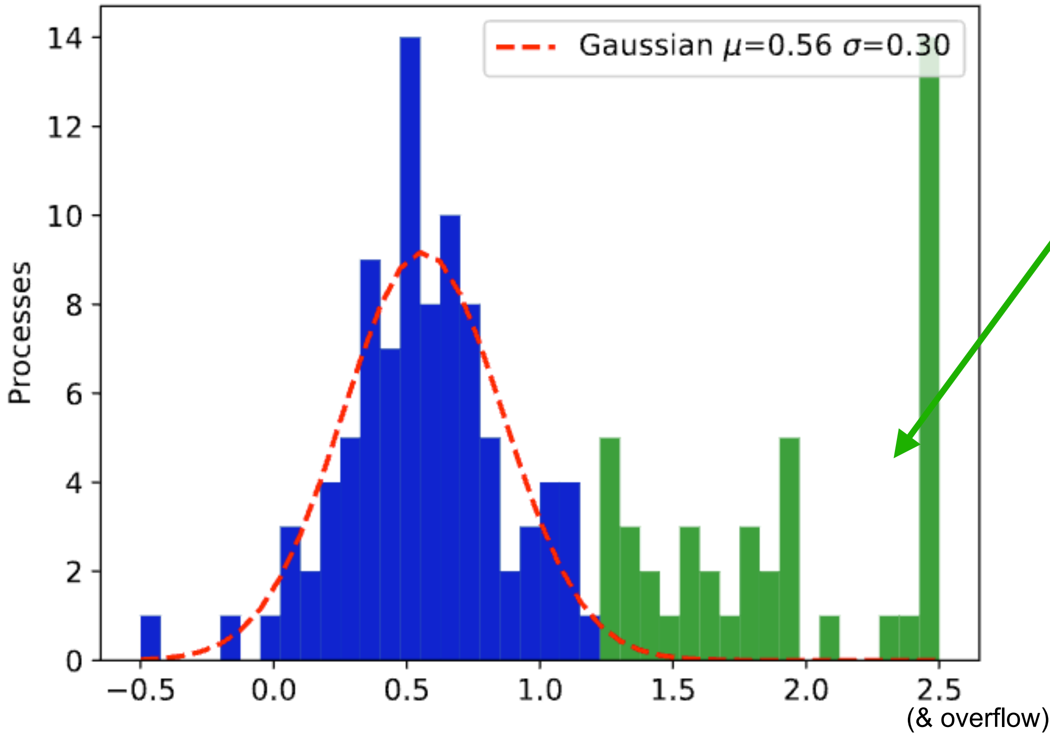
Examples

Process	n_{part}	$\Delta\sigma/\sigma_0$	$\frac{\sigma_{\text{NLO}}-\sigma_0}{\Delta\sigma}$	$\Delta\sigma_{\text{ref}}/\sigma_0$	$\frac{\sigma_{\text{NLO}}-\sigma_0}{\Delta\sigma_{\text{ref}}}$
p p > wpm	1	1.54×10^{-1}	1.84	1.47×10^{-1}	1.92
p p > wpm j	2	1.97×10^{-1}	1.96	2.94×10^{-1}	1.31
p p > wpm j j	3	2.45×10^{-1}	0.59	4.41×10^{-1}	0.33
p p > wpm j j j	4	4.10×10^{-1}	0.25	5.88×10^{-1}	0.18
p p > z	1	1.46×10^{-1}	1.87	1.47×10^{-1}	1.86
p p > z j	2	1.93×10^{-1}	1.82	2.94×10^{-1}	1.19
p p > z j j	3	2.43×10^{-1}	0.56	4.41×10^{-1}	0.31
p p > z j j j	4	4.08×10^{-1}	0.27	5.88×10^{-1}	0.19
p p > a j	2	3.12×10^{-1}	5.33	2.94×10^{-1}	5.66
p p > a j j	3	3.28×10^{-1}	0.85	4.41×10^{-1}	0.63
p p > w+ w- wpm	3	1.00×10^{-3}	610.69	4.41×10^{-1}	1.39
p p > z w+ w-	3	8.00×10^{-3}	92.39	4.41×10^{-1}	1.68
p p > z z wpm	3	1.00×10^{-2}	85.00	4.41×10^{-1}	1.93
p p > z z z	3	1.00×10^{-3}	302.75	4.41×10^{-1}	0.69
p p > a w+ w-	3	1.90×10^{-2}	42.33	4.41×10^{-1}	1.82
p p > a a wpm	3	4.40×10^{-2}	47.24	4.41×10^{-1}	4.72
p p > a z wpm	3	1.00×10^{-3}	1244.49	4.41×10^{-1}	2.82
p p > a z z	3	2.00×10^{-2}	17.24	4.41×10^{-1}	0.78

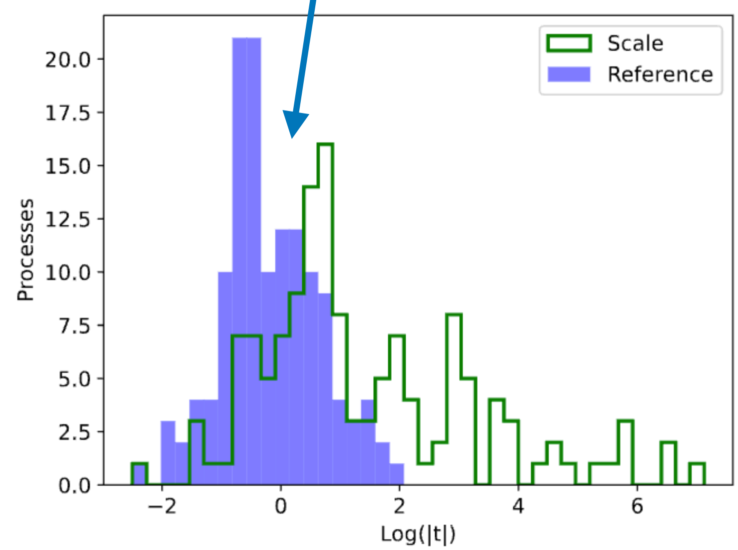
New pulls

Still tails

Much reduced



$$t_{\mathrm{ref}} = \frac{\sigma_{\mathrm{NLO}} - \sigma_0}{\Delta\sigma_{\mathrm{ref}}}$$



Remaining tails

Process	n_{part}	$\Delta\sigma/\sigma_0$	$\frac{\sigma_{\text{NLO}}-\sigma_0}{\Delta\sigma}$	$\Delta\sigma_{\text{ref}}/\sigma_0$	$\frac{\sigma_{\text{NLO}}-\sigma_0}{\Delta\sigma_{\text{ref}}}$
`p p > h`	1	3.48×10^{-1}	3.02	1.47×10^{-1}	7.15

Large corrections to the
Loop-induced 2->1 process

Not easily extracted from LO
Would be interesting to study in NLO->NNLO

Discussion

Why the Gaussian core?

No stochastic process

Repeating the same approximation

How to do for NLO, NNLO?

No similar, consistent large datasets available

ML connection

NLO and NNLO and NNNLO

Artisanal calculations

Could we use ML to learn patterns for estimates?

Would ML have noticed QCD pattern?

Hard to train if answers unknown.

Could we use ML to improve estimation?

Symbolic regression could aid interpretability.

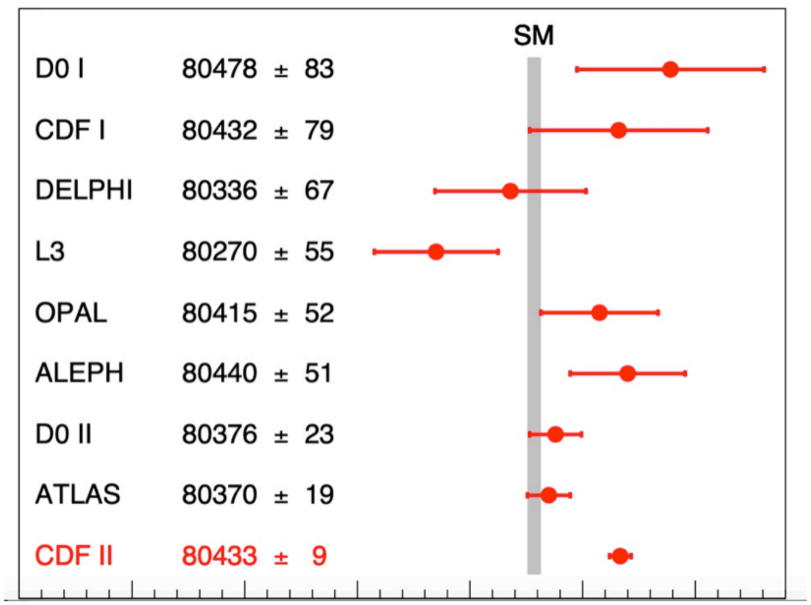
Outline

1. Theoretical

2. Experimental

Experimental systematics

Do we have confidence
in our understanding
of the SM and its uncertainties to 7σ ?



"with a significance of 7.0σ" (2.5e-12)

Experimental systematics

Do we have confidence
in our understanding
of the SM and its uncertainties to 7σ ?

We need to calculate

P(data | SM)

to one part in 1,000,000,000,000!

P(data|SM)

P(data|SM) is a distribution

Due to finite statistics

But SM also has parameters

$SM(\theta, \nu)$

Where θ are physics parameters (masses, couplings)
And ν are nuisance parameters (resolutions etc)

Nuisance parameters

Really, need to calculate

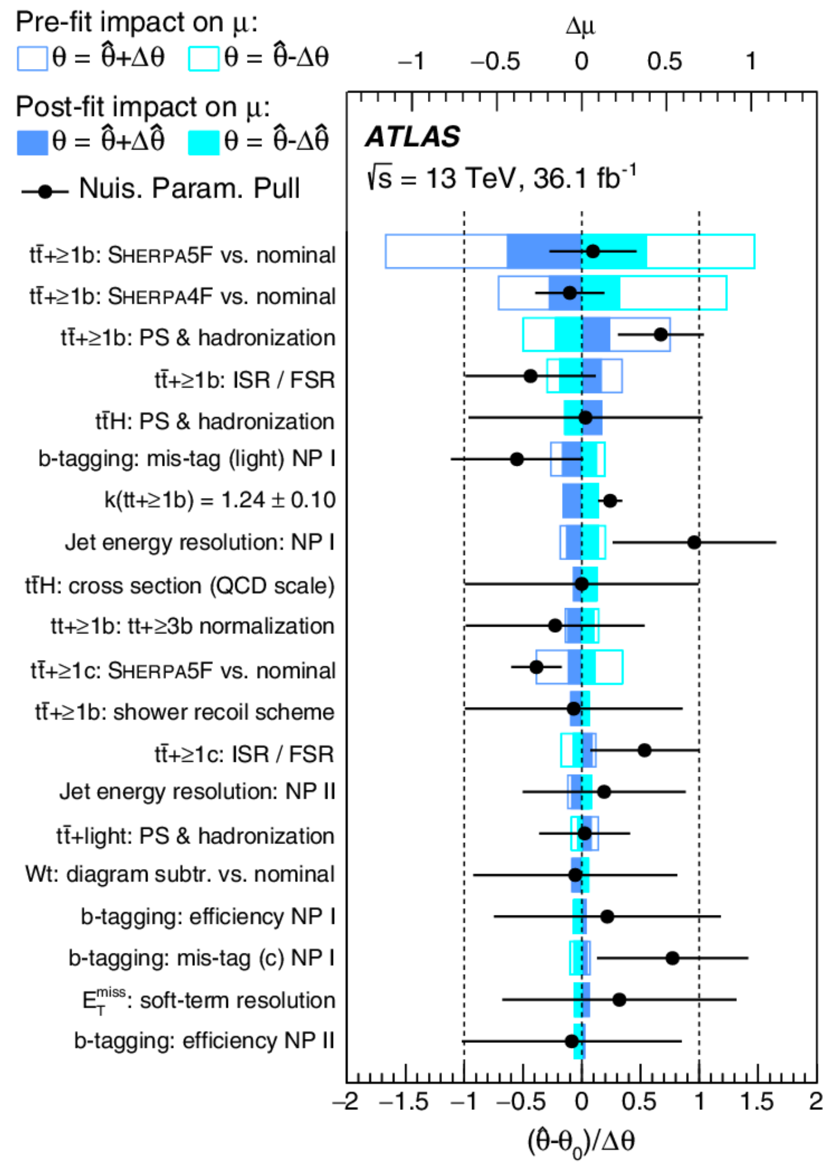
P(data | SM(θ, ν))

To calculate p-values

But ν can be high-dimensional

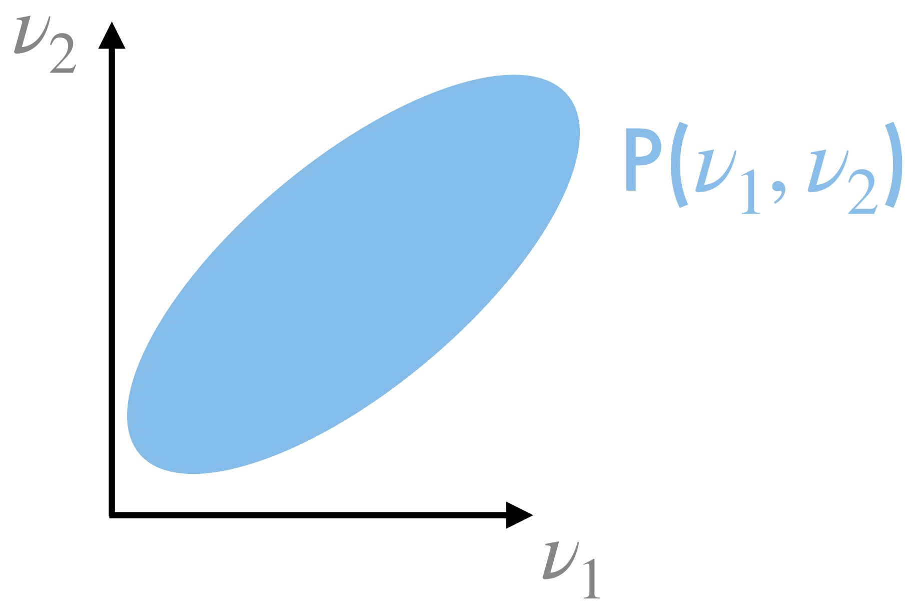
Often profiled away from nominal

Do we understand the ν space?



Multiple NP dimensions

Fairly **well studied**:

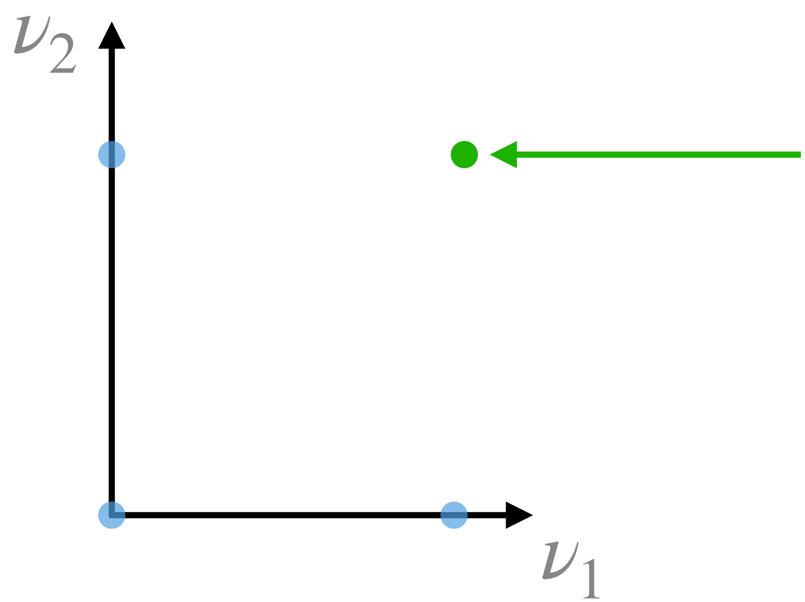


Correlated priors among NPs

Multiple NP dimensions

Less **well studied** is the crucial quantity:

$P(\text{data} \mid SM(\theta, \nu))$

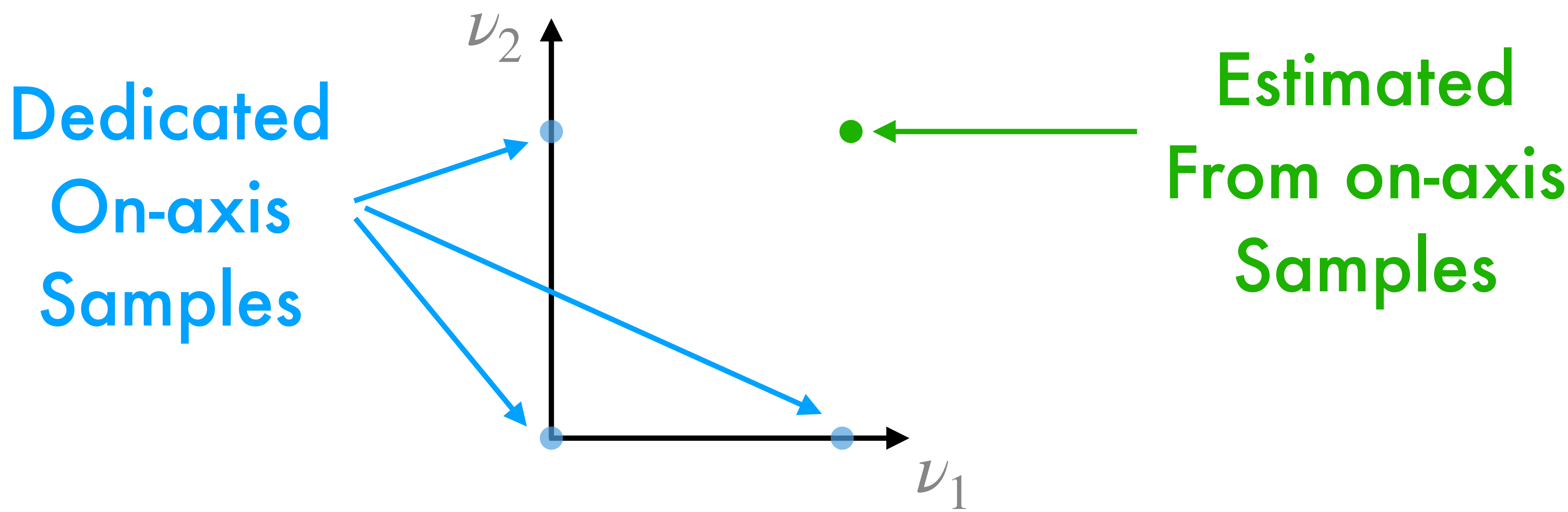


We need P(data|SM) to get p-values

Multiple NP dimensions

Less **well studied** is the crucial quantity:

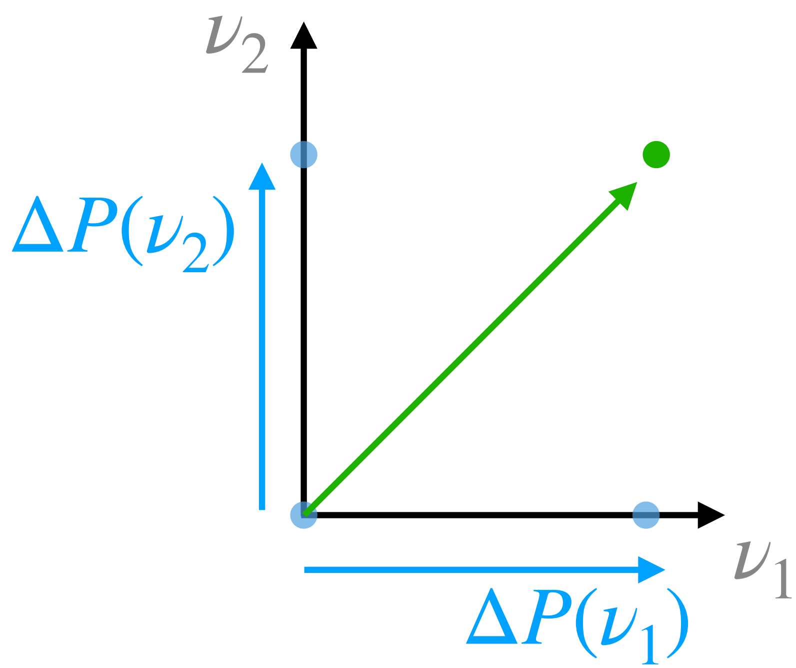
P(data | SM(θ, ν))



How do we do **off-axis modeling?**

Multiple NP dimensions

What is being assumed here?



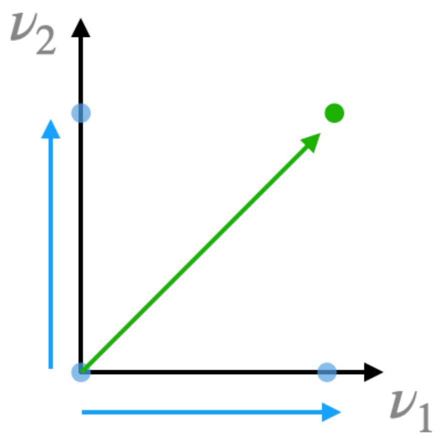
P(data | SM(θ, ν_1, ν_2)) can be approximated from

P(data | SM(θ,0,0)), $\Delta P(\nu_1), \Delta P(\nu_2)$

Multiple NP dimensions

P(data | SM(θ, ν_1, ν_2)) can be approximated from

P(data | SM(θ,0,0)), $\Delta P(\nu_1)$, $\Delta P(\nu_2)$



Several possible approaches:

Linear extrapolation

Morphing

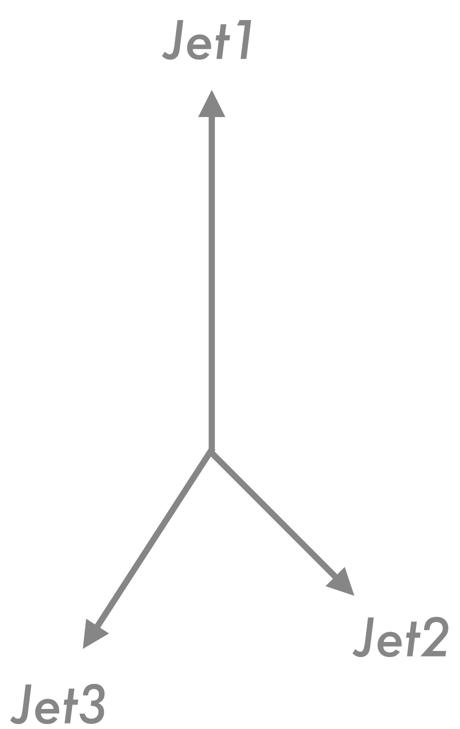
Factorizing

All assume

Impact are independent of other NPs

(Not the same as correlation of NP prior)

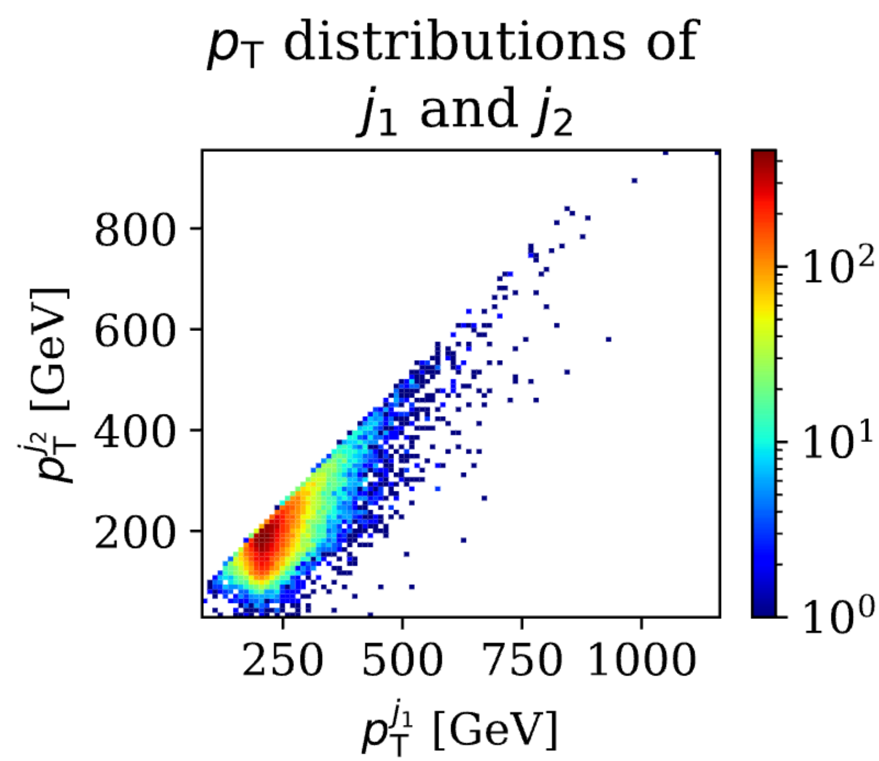
Example



3-jet event

1 high pT jet

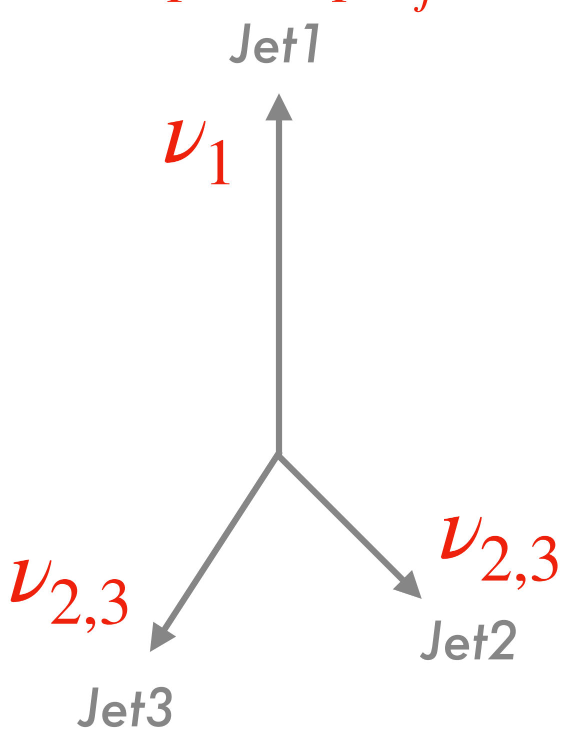
2 low pT jets



Example

Jet energy uncertainty

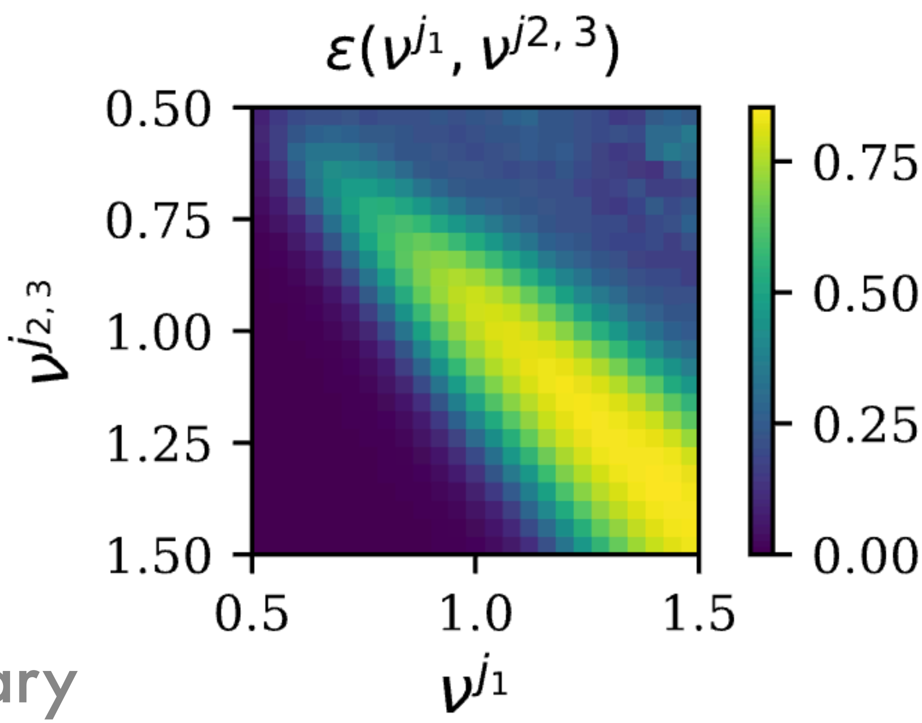
$p_T \to p_T/\nu_j$



3-jet event

Selection: MET<50 GeV

What is efficiency vs NPs?



Efficiency high when BOTH ν vary

Efficiency drops when one varies

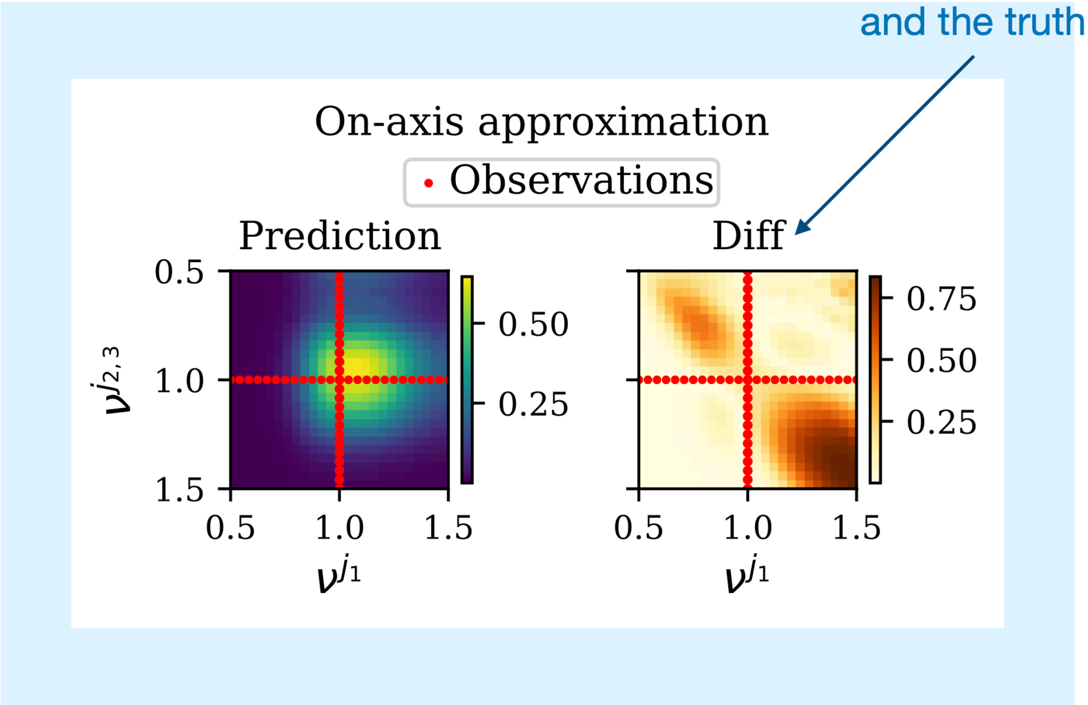
On-axis approximations

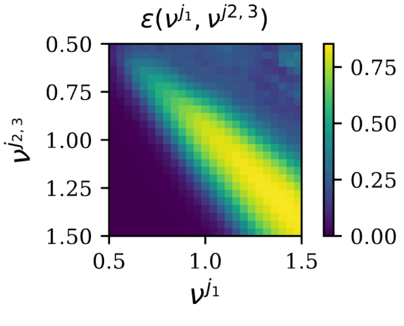
On-axis approximation

We sample the efficiency along the "central axes" and calculate the rest of the values according to $\epsilon(\nu^{j_1}, \nu^{j_{2,3}}) \approx \epsilon(1,\nu^{j_{2,3}}) \times \epsilon(\nu^{j_1},1)$

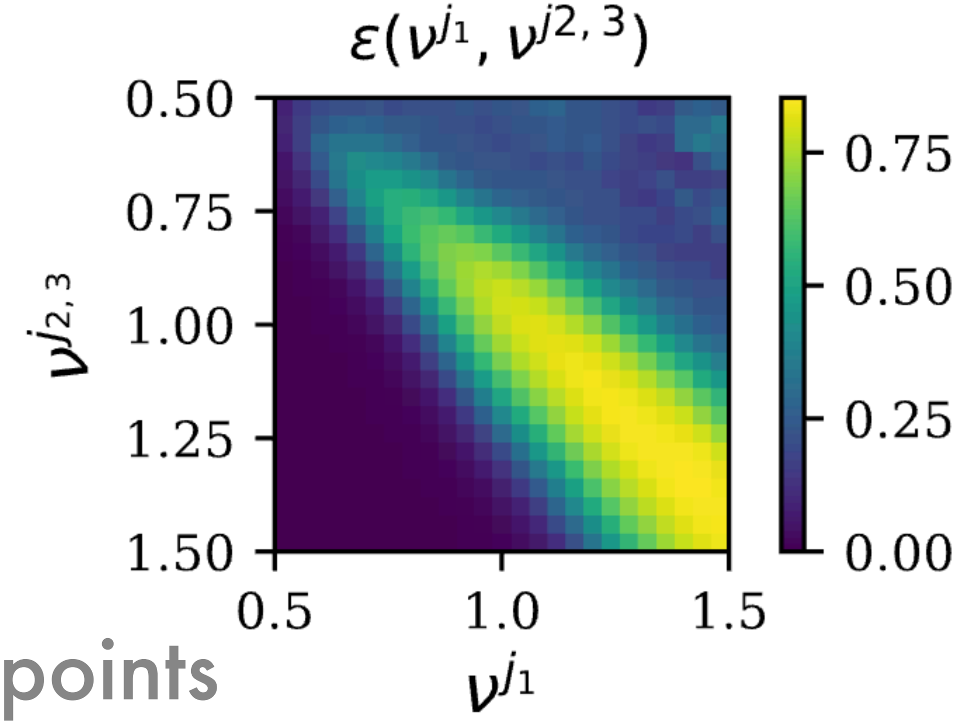
Total of 49 observations are sampled (red dots).

Absolute difference between the on-axis prediction and the truth





What to do?



<u>Other approaches</u>

Naive scan - expensive

<u>Our approach</u>

Probe space with minimal points

Estimate function with a Gaussian Process

Use Bayesian experimental design to select points

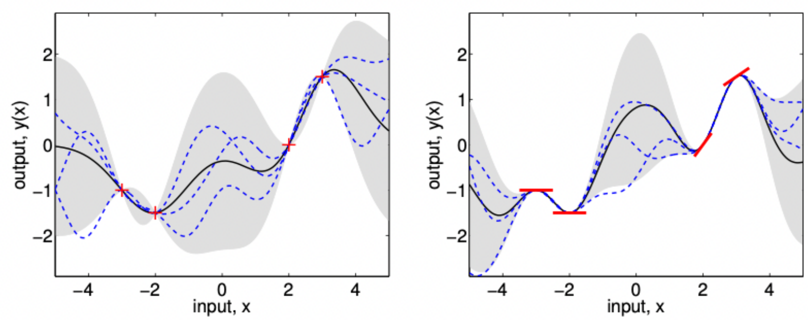
Use derivative info to speed convergence

Work in progress

Efficient Estimation of Multiple Systematic Uncertainties with Gaussian Processes and Bayesian Experimental Design

Alexis Romero,[1] Kyle Cranmer,[2] and Daniel Whiteson[1]

Basics of GPs with derivative information:



(a) (b)

Figure 9.1: In panel (a) we show four data points in a one dimensional noise-free regression problem, together with three functions sampled from the posterior and the 95% confidence region in light grey. In panel (b) the same observations have been augmented by noise-free derivative information, indicated by small tangent segments at the data points. The covariance function is the squared exponential with unit process variance and unit length-scale.

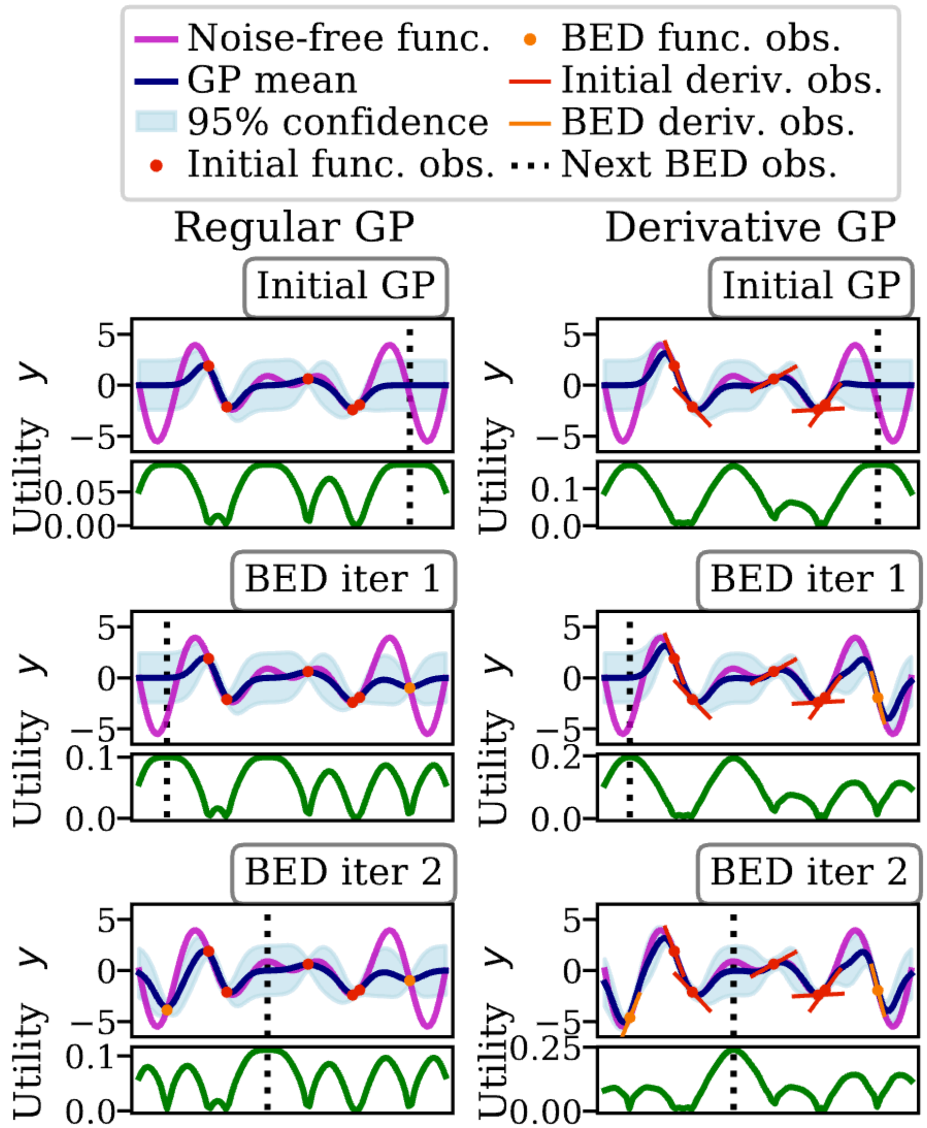
http://gaussianprocess.org/gpml/chapters/RW9.pdf

Procedure

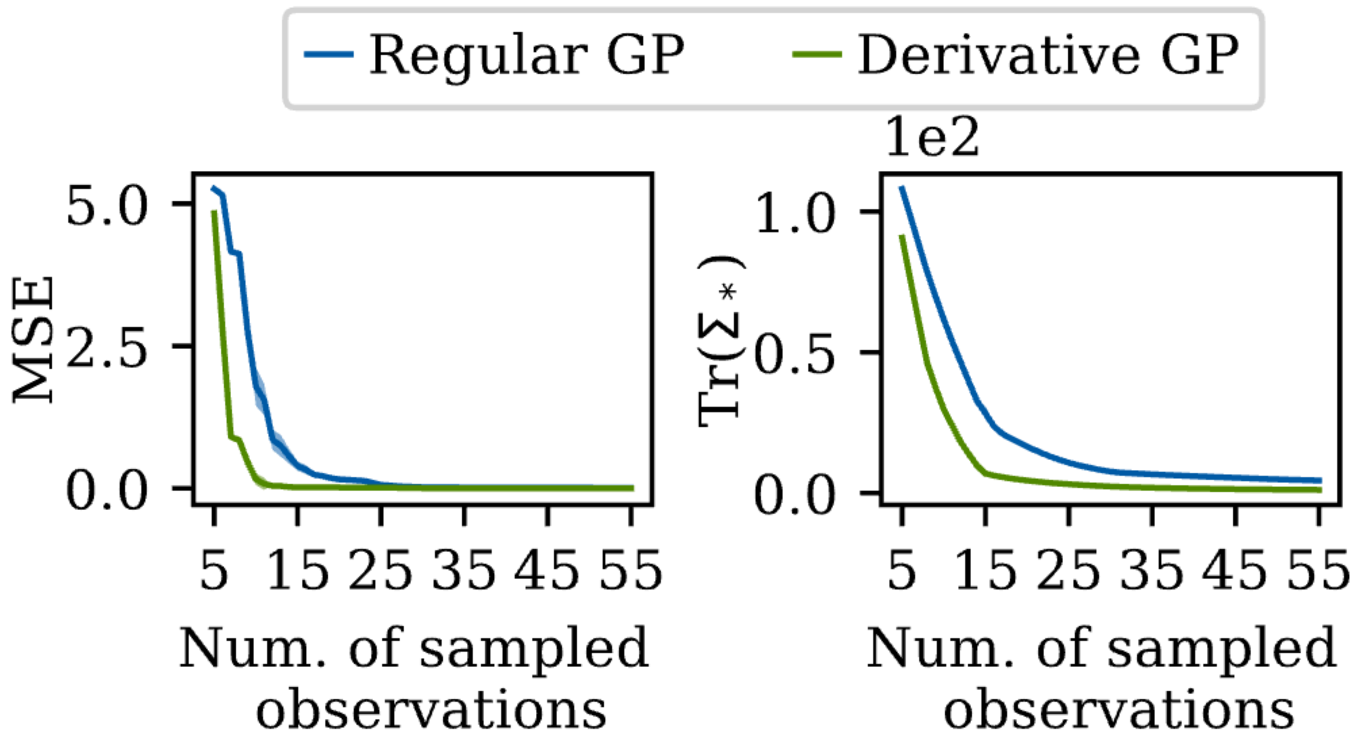
Initial coarse scan

Build GP

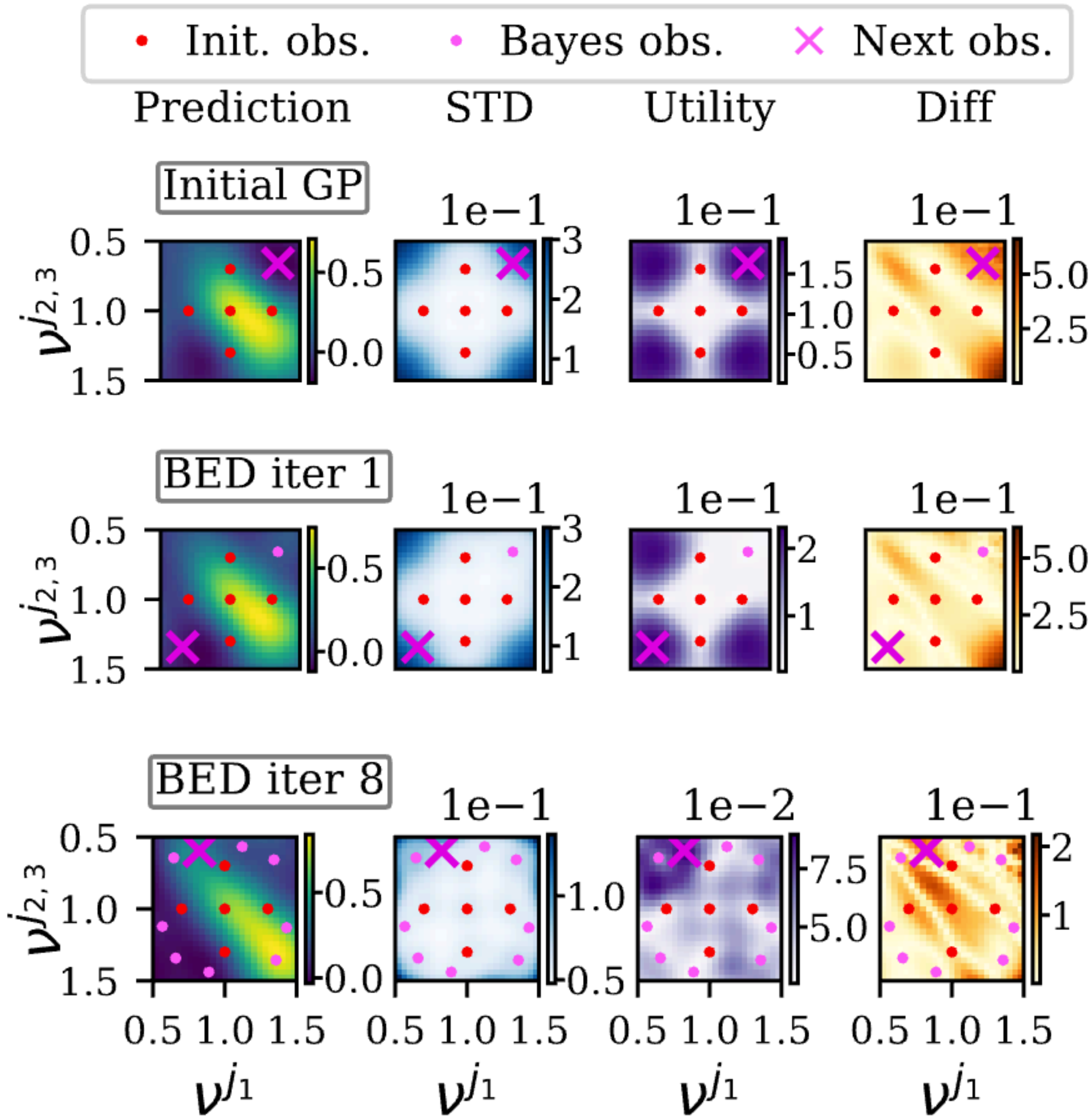
Choose sample points to minimize expected uncertainty $\mathrm{Tr}[\Sigma_Q]$



Results - Toy demo

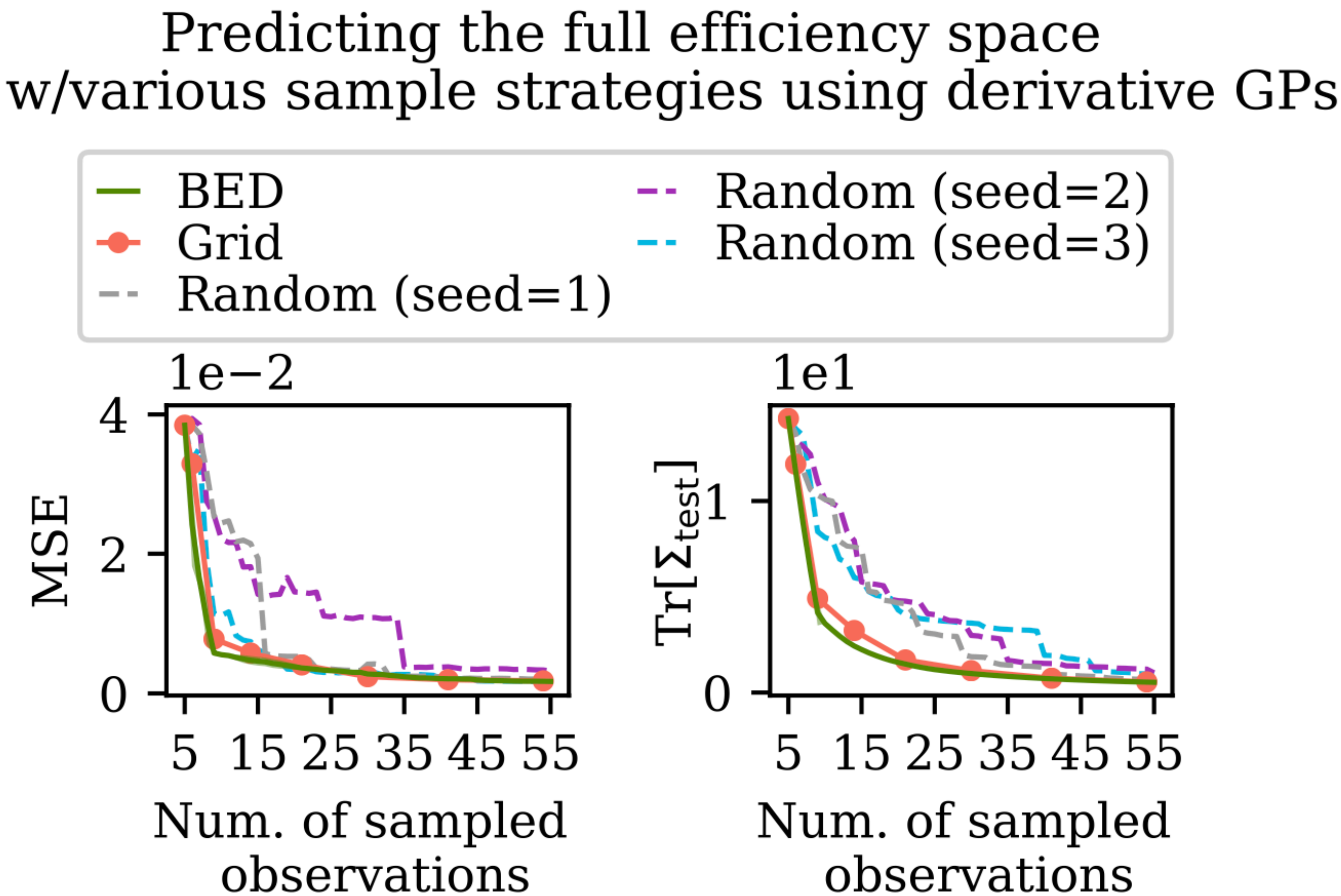


Jet problem



Init. obs.
Bayes obs.
Next obs.
Prediction
STD
Utility
Diff
Initial GP
BED iter 1
BED iter 8
1e−1
1e−2
$\nu^{j_{2,3}}$
ν^{j_1}
0.5
1.0
1.5

Performance

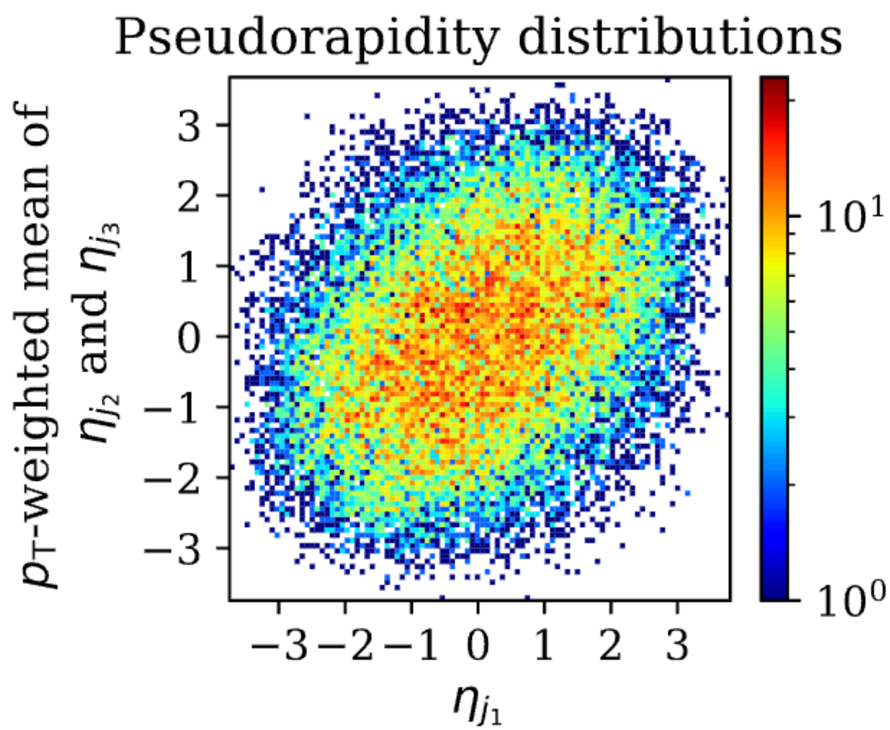


4D space

Eta separation

3 jet system (j_1, j_2, j_3):

- Jet energy scale of hardest jet:
  When $\eta_1 < 1 : \nu_{in}^{j_1}$
  When $\eta_1 \geq 1 : \nu_{out}^{j_1}$
- Jet energy scale of two softer jets:
  When $\eta_{2,3} < 1 : \nu_{in}^{j_{2,3}}$
  When $\eta_{2,3} \geq 1 : \nu_{in}^{j_{2,3}}$
  Where $\eta_{2,3}$ is the pT-weighted era average between j_2 and j_3



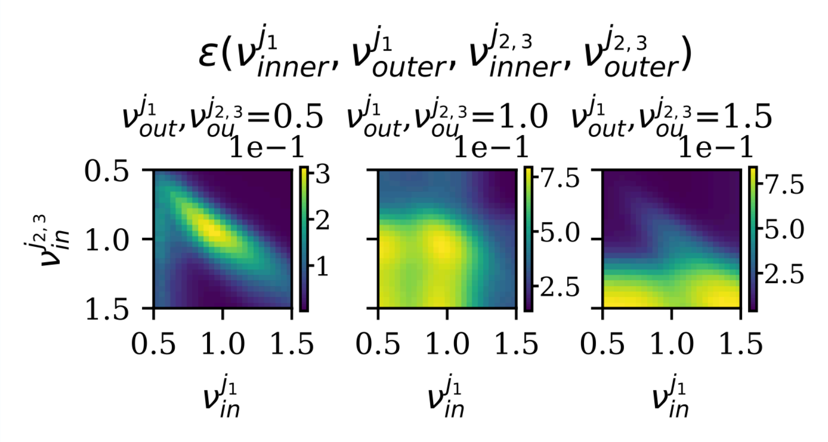
4D space

Eta separation

3 jet system (j_1, j_2, j_3):

- Jet energy scale of hardest jet:
  When $\eta_1 < 1 : \nu_{in}^{j_1}$
  When $\eta_1 \geq 1 : \nu_{out}^{j_1}$
- Jet energy scale of two softer jets:
  When $\eta_{2,3} < 1 : \nu_{in}^{j_{2,3}}$
  When $\eta_{2,3} \geq 1 : \nu_{in}^{j_{2,3}}$
  Where $\eta_{2,3}$ is the pT-weighted era average between j_2 and j_3

Truth



Since it's hard to visualize a 4D space, we hold two prams ($\nu_{out}^{j_1}, \nu_{out}^{j_{2,3}}$) to have a fixed value and we plot the efficiency as we vary the other two ($\nu_{in}^{j_1}, \nu_{in}^{j_{2,3}}$)

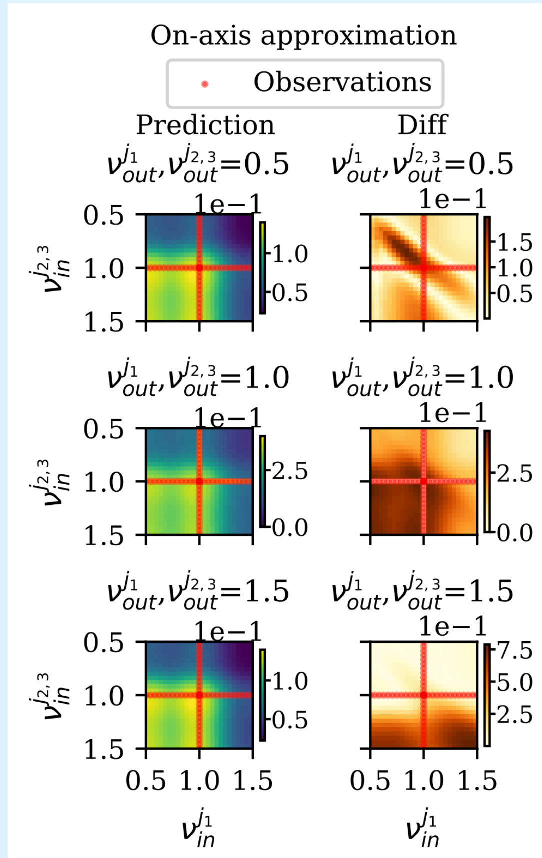
Approximations

On-axis approximation

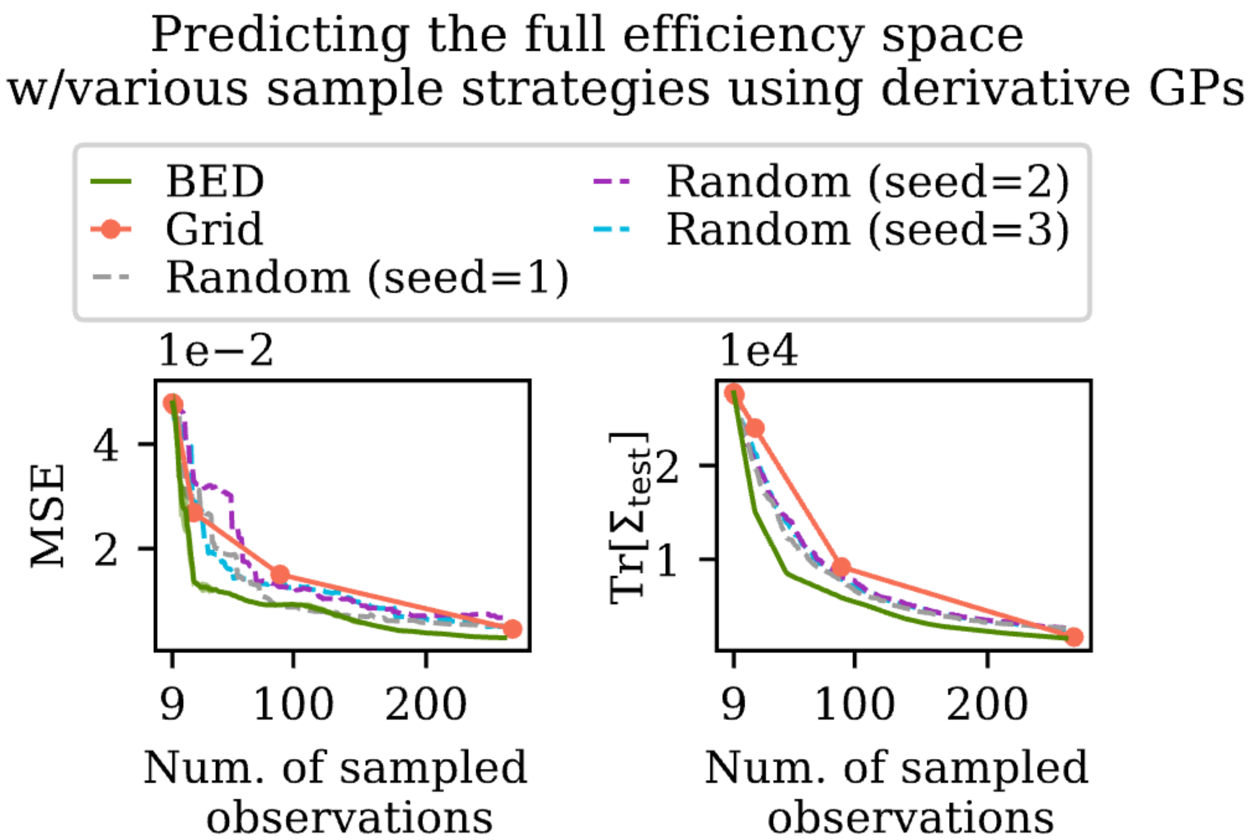
Total of 97 observations along the central axes.

The off-axis values are calculated according to

$$\epsilon(\nu_{in}^{j_1}, \nu_{in}^{j_{2,3}}, \nu_{out}^{j_1}, \nu_{out}^{j_{2,3}}) \approx \epsilon(\nu_{in}^{j_1},1,1,1) \times \epsilon(1,\nu_{in}^{j_{2,3}},1,1) \times \epsilon(1,1,\nu_{out}^{j_1},1) \times \epsilon(1,1,1,\nu_{out}^{j_{2,3}})$$

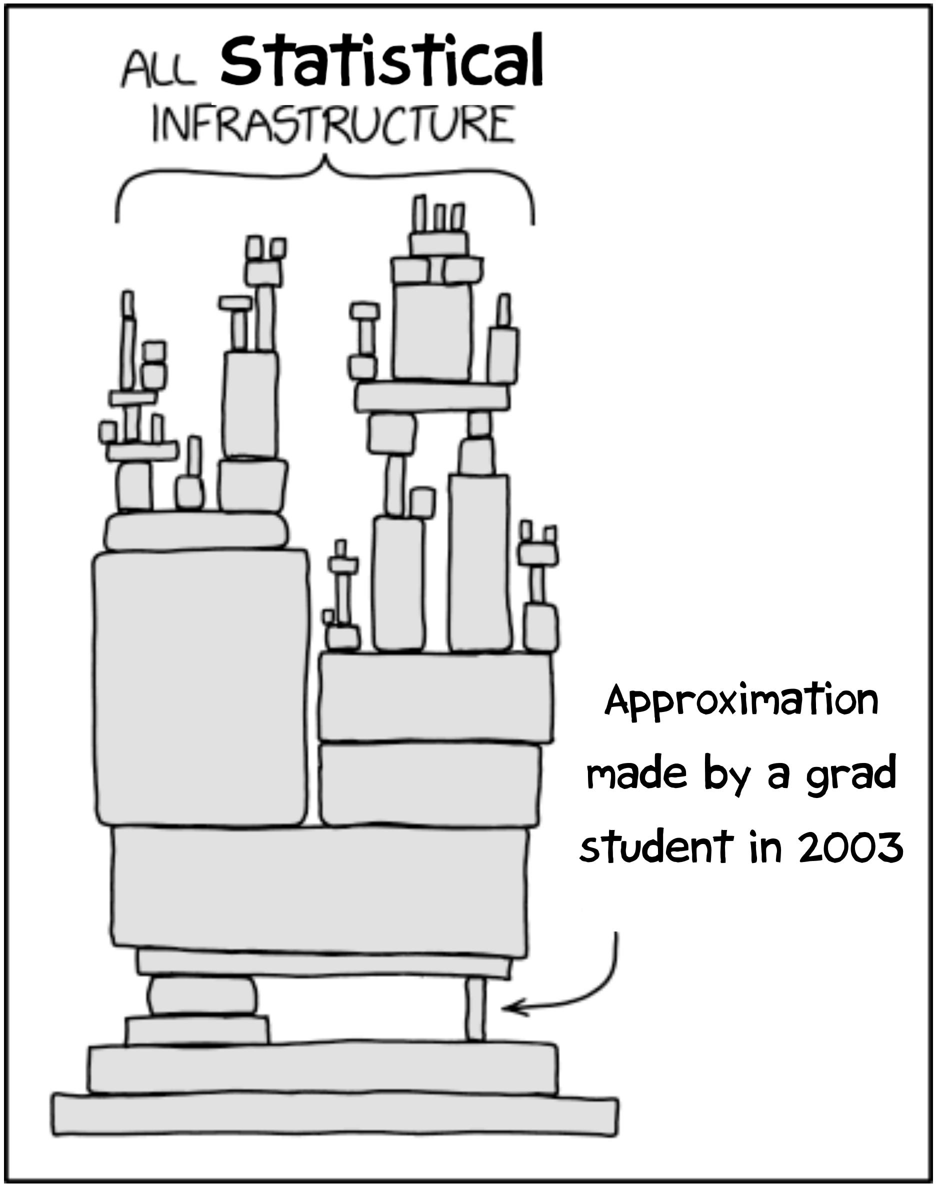


Results



Predicting the full efficiency space
w/various sample strategies using derivative GPs
BED
Grid
Random (seed=1)
Random (seed=2)
Random (seed=3)
1e−2
MSE
4
2
9
100
200
Num. of sampled observations
1e4
Tr[Σtest]
2
1
9
100
200
Num. of sampled observations

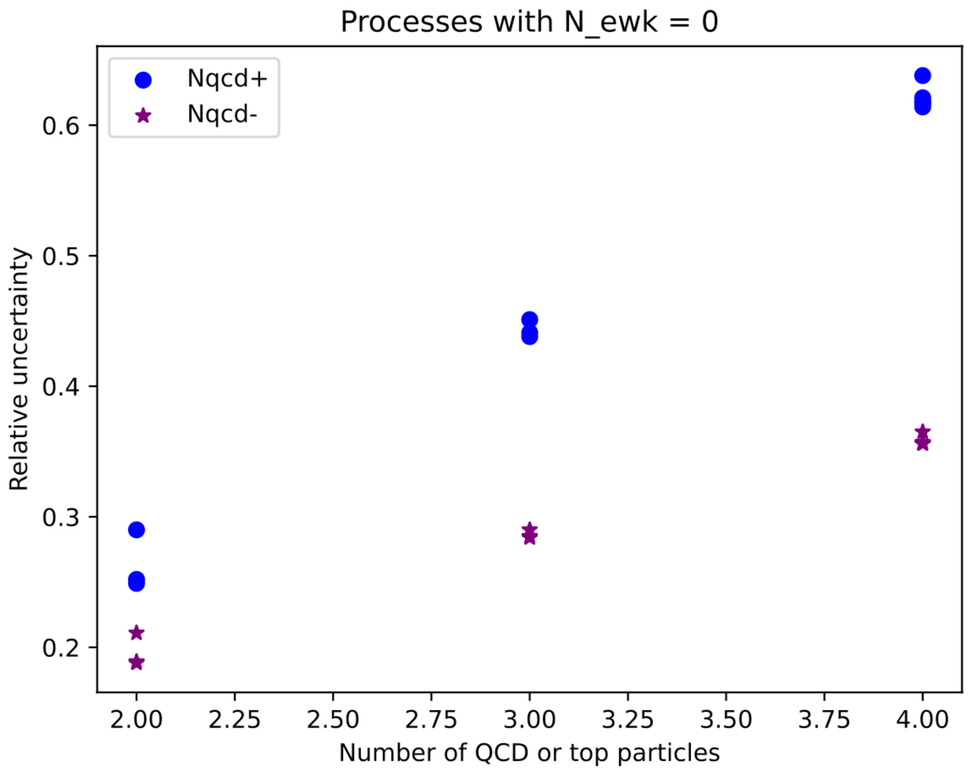
Conclusions



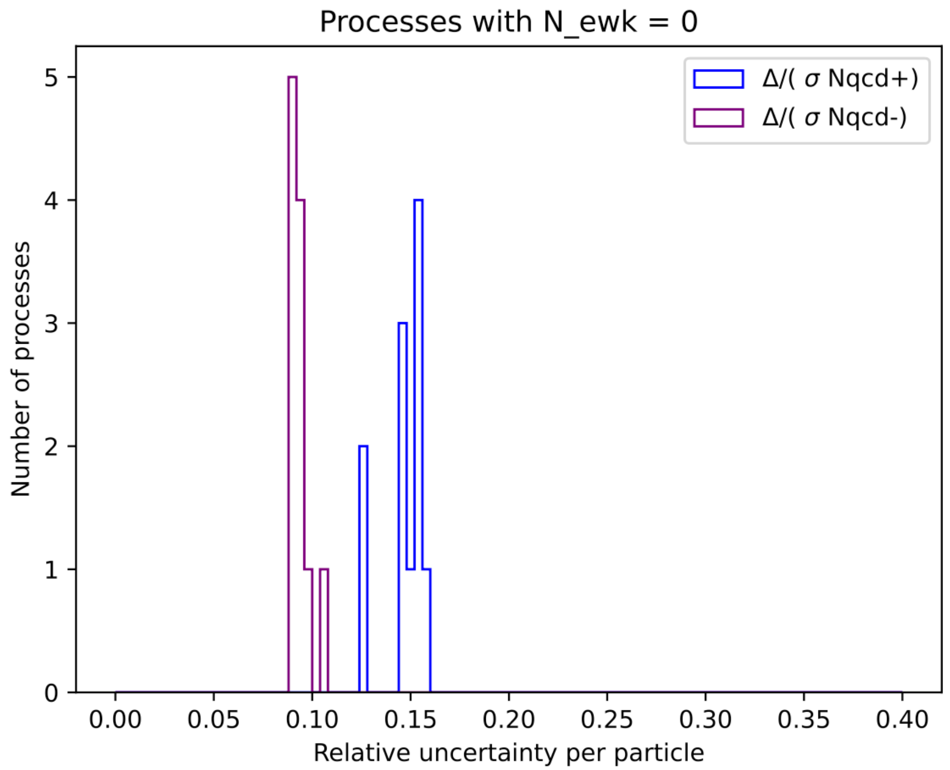
Our approach to systematics has worked well but will face new challenges

Time to use more powerful recent tools to make it more robust!

QCD scaling



Processes with N_ewk = 0
Nqcd+
Nqcd-
Relative uncertainty
Number of QCD or top particles



Processes with N_ewk = 0
Δ/(σ Nqcd+)
Δ/(σ Nqcd-)
Number of processes
Relative uncertainty per particle