# Getting serious about systematics with Machine Learning



#### Daniel Whiteson, UC Irvine Oct 2022

### Motivation

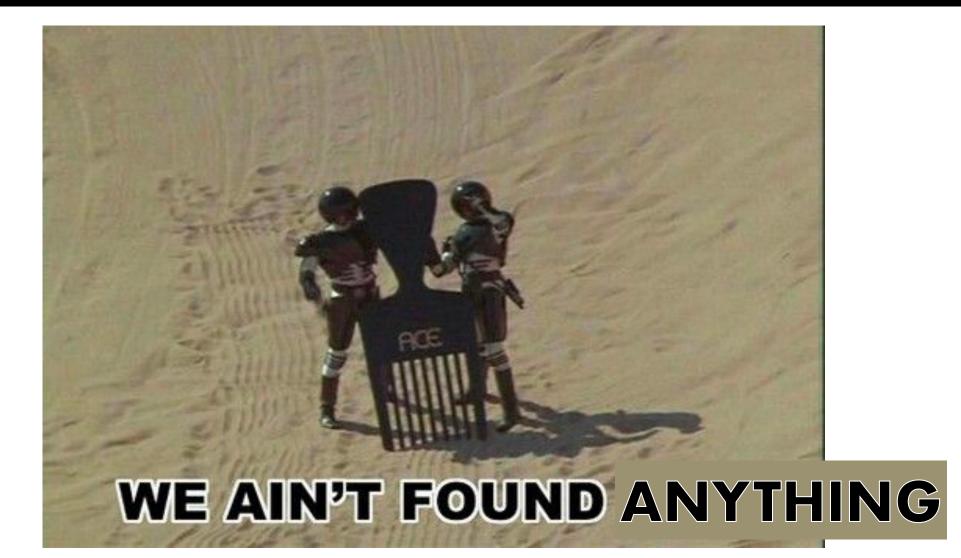
#### ATLAS Exotics Searches\* - 95% CL Upper Exclusion Limits ATLAS Preliminary Status: July 2018 $\sqrt{s} = 8, 13 \text{ TeV}$ $\int \mathcal{L} dt = (3.2 - 79.8) \text{ fb}^{-1}$ Jets † E\_\_\_\_\_ {Ldt[fb-1] Model $l, \gamma$ Limit Reference ADD $G_{KK} + g/q$ 0 e.u 1-4jYes 36.1 7 TeV 1711.03301 n = 2ADD non-resonant yy 2γ 36.7 8.6 TeV n = 3 HLZ NLO 1707.04147 ADD OBH 2 i 37.0 8.9 TeV n = 61703/09/27 ADD BH high 5: pr $\geq 1 e, \mu$ ≥2 i -3.2 8.2 TeV n = 6, Mn = 3 TeV, rot BH 1606.02265 ADD BH multilet ≥3 j 36 9.55 TeV n = 6. Mo = 3 TeV, rot BH 1512.02586 RS1 $G_{KK} \rightarrow \gamma\gamma$ 2γ 4.1 TeV $k(\overline{M}_{Pl} = 0.1)$ 36.7 1707.04147 Bulk RS $G_{KK} \rightarrow WW/ZZ$ $k/\overline{M}_{\rm PM} = 1.0$ CERN-EP-2018-179 multi-channel 36.1 2.3 TeV Bulk RS $g_{KK} \rightarrow t\bar{t}$ 1 e,µ ≥1 b,≥1J/2j Yes 1804.10823 3.8 TeV 36.1 $\Gamma/m = 15\%$ 2UED / RPP 10.0 > 2 b. > 3 i Yes 36.1 1.8 Te V Tig: (1,1), $\mathcal{B}(A^{(1,1)} \rightarrow tt) = 1$ 1803.09678 SSM $7' \rightarrow ll$ 2 e. µ 36.1 4.5 Te V 1707/02424 SSM $Z' \rightarrow \tau \tau$ $2\tau$ 36.1 2.42 TeV 1709/07242 Leptophobic $Z' \rightarrow bb$ 2 b 36.1 2.1 TeV 1805.09299 Leptophobic $Z' \rightarrow tt$ 36.1 1 e, μ ≥ 1 b, ≥ 1J/2j Yes 3.0 TeV $\Gamma/m = 1\%$ 1804.10823 SSM $W' \rightarrow \ell r$ 79.8 5.6 TeV ATLAS-CONF-2018-017 1 e. µ Yes $SSM \ W' \to \pi'$ 3.7 TeV $1\tau$ Yes 36.1 1801.06992 $HVT V' \rightarrow WV \rightarrow qqqq \mod B$ 0 e, µ ATLAS-CONF-2018-016 2 J 79.8 4,15 TeV $\kappa v = 3$ HVT V' → WH/ZH model B multi-channel $g_V = 3$ 2.93 TeV 1712.06518 36.1 LRSM $W'_{6} \rightarrow tb$ multi-channel 36.1 3.25 TeV CERN-EP-2018-142 37.0 CLagga 21.8 TeV 10 2 j 1703.09127 0 CI // qq 2 e, µ 40.0 TeV 1707.02424 36.1 $\overline{\eta_{12}}$ ≥1*e*,µ ≥1b,≥1j Yes $|C_{44}| = 4\pi$ CI rrrr 36.1 2.57 TeV CERN-EP-2018-174 Axial-vector mediator (Dirac DM) 0 e. u 1 - 4iYes 36.1 1.55 TeV g<sub>0</sub>=0.25, g<sub>i</sub>=1.0, m(\chi) = 1 GeV 1711.03301 Colored scalar mediator (Dirac DM) 0 e, µ 1 - 4j36.1 $g=1.0, m(\chi) = 1 \text{ GeV}$ 1711.03301 Yes 1.67 TeV VV vv EFT (Dirac DM) 0 e, µ 1 J, ≤ 1 j Yes 3.2 $m(\chi) < 150 \text{ GeV}$ 1608.02372 ≥2 j Scalar LQ 1st gen 2 e 3.2 1.1 TeV $\beta = 1$ 1605.06035 Scalar LQ 2nd gen $\beta = 1$ 1605.06035 $2\mu$ ≥2 j 3.2 .05 TeV $\beta = 0$ 1508.04735 Scalar LQ 3rd gen 1 e.µ ≥1 b,≥3j Yes 20.3 $VLQ TT \rightarrow Ht/Zt/Wb + X$ 1.37 TeV ATLAS-CONF-2018-032 multi-channel 36.1 SU(2) doublet $VLQ BB \rightarrow Wt/Zb + X$ multi-channel 36.1 1.34 TeV SU(2) doublet ATLAS-CONF-2018-032 2(SS)/≥3 e,µ ≥1 b, ≥1 j Yes 1.64 TeV VLQ $T_{5/3}T_{5/3}|T_{5/3} \rightarrow Wt + X$ 36.1 $\mathcal{B}(T_{5:0} \rightarrow Wt) = 1, d(T_{5:0}Wt) = 1$ CERN-EP-2018-171 $VLQ Y \rightarrow Wb + X$ 1*e*,µ ≥1b,≥1j 1.44 TeV $\mathcal{B}(Y \rightarrow Wb)=1$ , $d(YWb)=1/\sqrt{2}$ ATLAS-CONF-2016-072 Yes 3.2 $VLQ B \rightarrow Hb + X$ r\_p= 0.5 0 e.µ, 2 γ ≥1 b, ≥1j Yes 79.8 1.21 TeV ATLAS-CONF-2018-024 $VLQ QQ \rightarrow WqWq$ 1 e, µ ≥4 j Yes 20.3 1509.04261 2 j Excited quark $q^* \rightarrow qq$ 6 0 TeV 1703.09127 37.0 only $u^{i}$ and $d^{i}$ , $\Lambda = m(a^{i})$ 1 y Excited quark $q^* \rightarrow qy$ only $u^i$ and $d^i$ , $\Lambda = m(q^i)$ 1 j 36.7 5.3 TeV 1709.10440 2.6 Te V Excited quark $b^* \rightarrow bg$ 1 b, 1 j -36.1 1805.09299 Excited lepton *l*\* 3 e, µ 1411.2921 20.3 $\Lambda = 3.0 \text{ TeV}$ Excited lepton v\* 3 e. u. T 20.3 1.6 TeV $\Lambda = 1.6 \text{ TeV}$ 1411.2921 Type III Seesaw 1 e. u >2 i Yes 79.8 ATLAS.CONE.2018.020 LBSM Majorana y 2 e. µ 2 j 20.3 $m(W_R) = 2.4$ TeV, no mixing 1506.06020 2,3,4 e, µ (SS) Higgs triplet $H^{++} \rightarrow \ell \ell$ 36.1 DY production 1710.09748 Higgs triplet $H^{++} \rightarrow \ell \tau$ DY production, \$(H)\*\* 20.3 30.4.7 1411.2921 Monotop (non-res prod) $a_{mn-res} = 0.2$ 1410.5404 1 e, µ 1 b Yes 20.3 Multi-charged particles DY production, |q| = 54 20.3 1504.04188 -Magnetic monopoles DY production, $|g| = 1_{(0)}$ , spin 1/2 1509.08059 7.0 . . . . . . . $\sqrt{s} = 13 \text{ TeV}$ $\sqrt{s} = 8 \text{ TeV}$ $10^{-1}$ 10 Mass scale [TeV]

"Only a selection of the available mass limits on new states or phenomena is shown

†Small-radius (large-radius) jets are denoted by the letter j (J).

#### There are no clear post-Higgs discoveries

### Direct searches

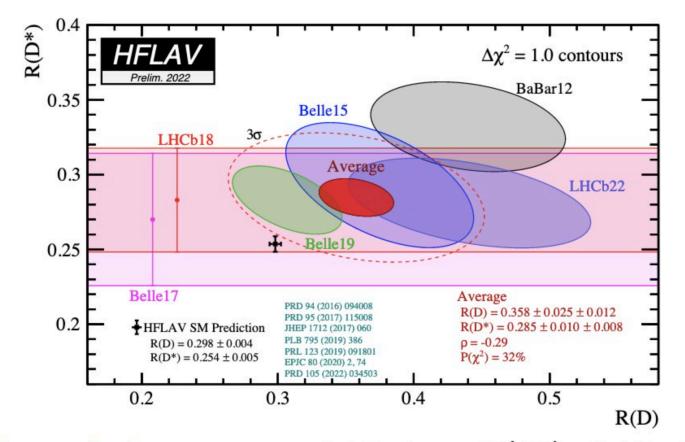


# The path forward



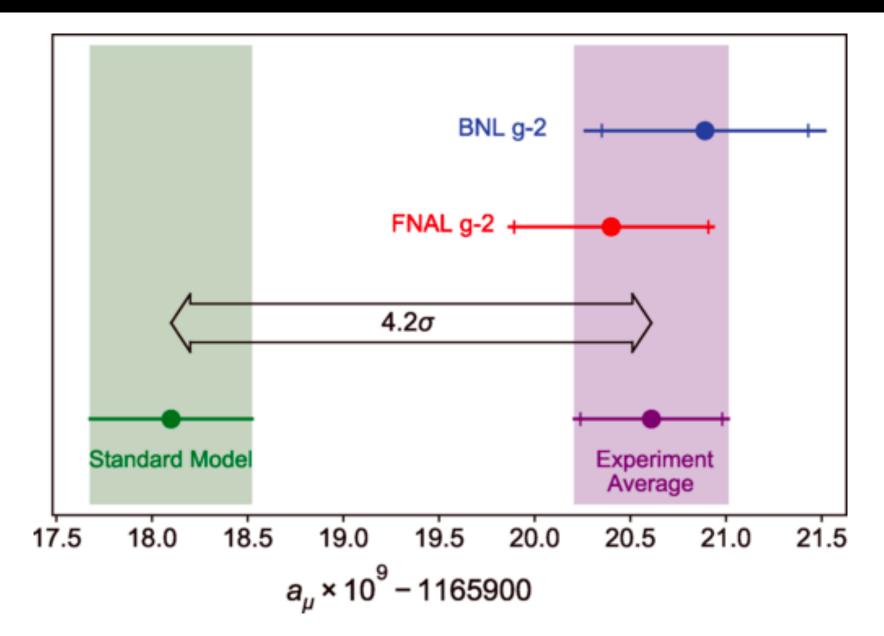
# This era will be defined by searches for subtle deviations from the SM

### **Flavor anomalies**

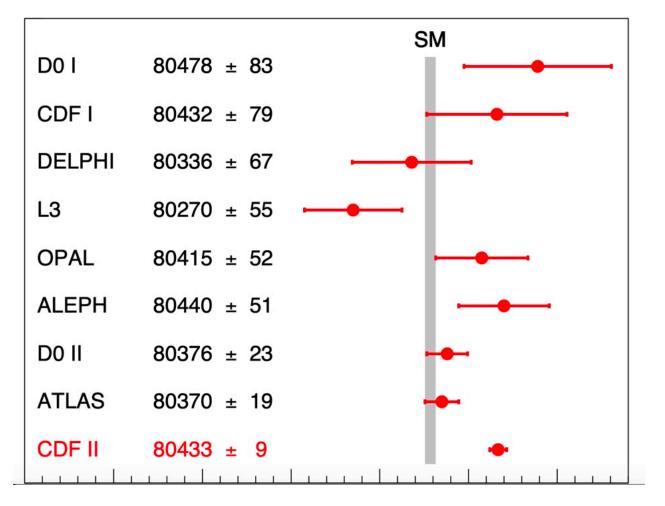


- New preliminary average: slightly lower R(D\*), slightly higher R(D), reduced correlation
  - $3.3\sigma \rightarrow 3.2\sigma$  agreement with SM
  - Excellent overall agreement between measurements



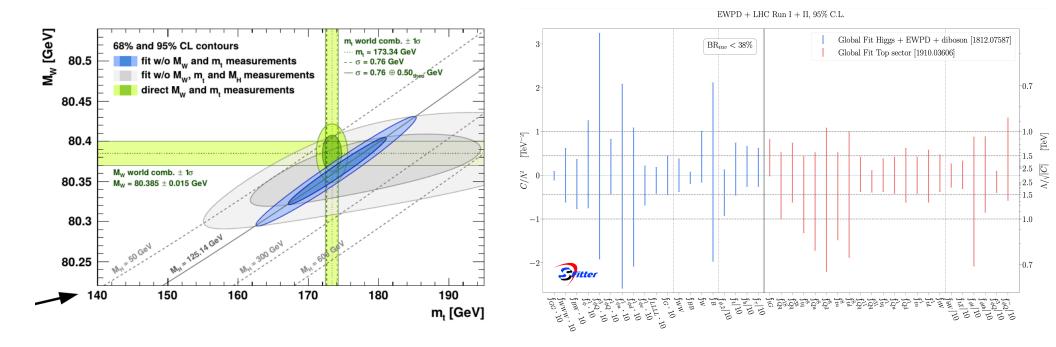


### CDF W mass



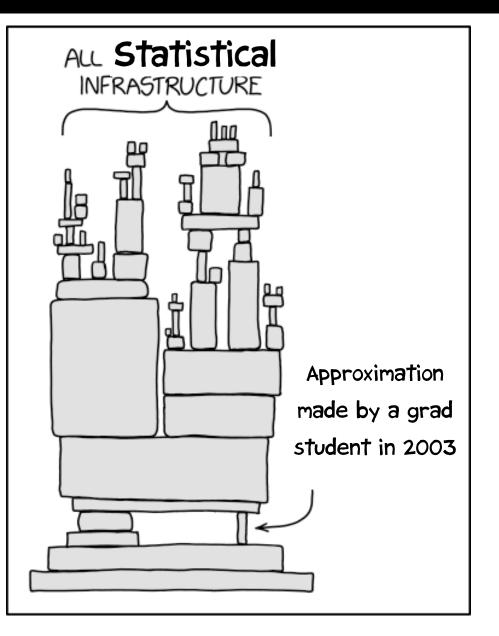
"with a significance of 7.0 $\sigma$ " (p=2.5e-12)

### 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)

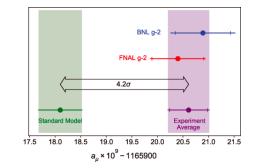
# Sources of Systematics

#### 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

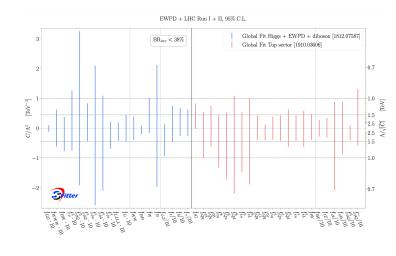
#### Unified treatment

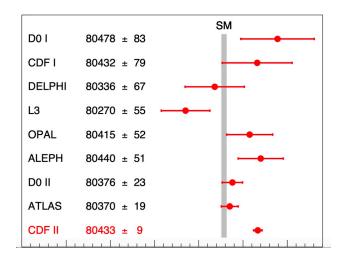
Treated as random variables

With some aux likelihood/prior

Marginalize/profile/pivot Eg: 2105.08742

Does that work for theory? What are the distributions?





### Cautionary tale

#### 2109.08159

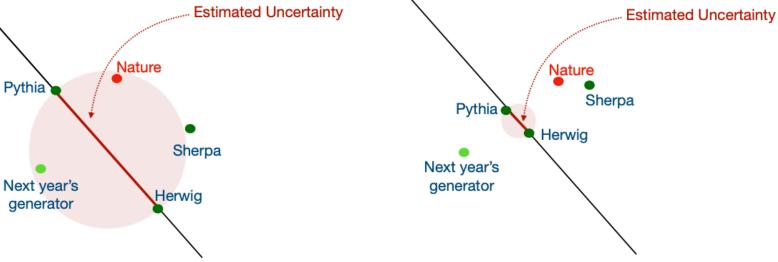
#### A Cautionary Tale of Decorrelating **Theory Uncertainties**

Pythia

Aishik Ghosh<sup>*a,b*</sup> and Benjamin Nachman<sup>*b,c*</sup>

Without Decorrelation





With Decorrelation

# (N)LO calculations

#### Incomplete calculations

#### **Statistical Patterns of Theory Uncertainties**

Aishik Ghosh<sup>1,2</sup>, Benjamin Nachman<sup>1,3</sup>, Tilman Plehn<sup>4</sup>, Lily Shire<sup>2</sup>, Tim M.P. Tait<sup>2</sup>, and Daniel Whiteson<sup>2</sup>

220x.xxxx (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 $\mu_F$

#### <u>Ultraviolet divergences</u>

removed by renormalization to cutoff  $\mu_R$ 

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

### Scales

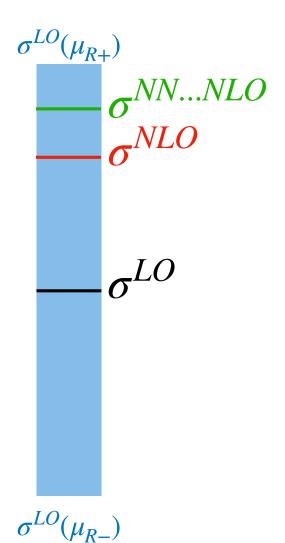
What is the error in cross-section due to truncation?

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

$$\mu_0 = \frac{H_T}{2} = \frac{1}{2} \sum_{\text{final state}} \sqrt{p_T^2 + m^2} \qquad \mu_{R,+} = 2 \qquad \mu_0$$
$$\mu_{R,-} = \frac{1}{2} \qquad \mu_0$$

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

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

#### <u>Critical issue:</u>

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<sup>*a*</sup>, R. Frederix<sup>*b*</sup>, S. Frixione<sup>*b*</sup>, V. Hirschi<sup>*c*</sup>, F. Maltoni<sup>*d*</sup>, O. Mattelaer<sup>*d*</sup>, H.-S. Shao<sup>*e*</sup>, T. Stelzer<sup>*f*</sup>, P. Torrielli<sup>*g*</sup>, M. Zaro<sup>*hi*</sup>

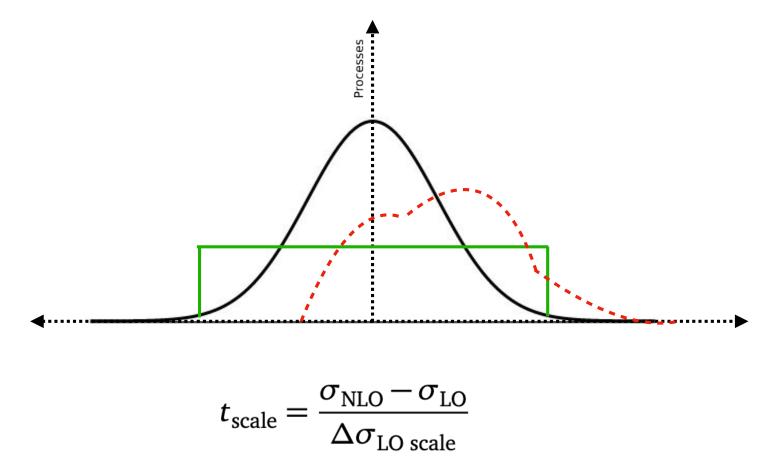
Process		Syntax	Cross section (pb)				
Vector boson +jets			LO 13 Te	eV	NLO 13 $TeV$		
a.1 a.2 a.3 a.4	$pp  ightarrow W^{\pm}$ $pp  ightarrow W^{\pm}jj$ $pp  ightarrow W^{\pm}jjj$ $pp  ightarrow W^{\pm}jjj$	pp>wpm pp>wpmj pp>wpmjj pp>wpmjjj	$\begin{array}{c} 1.375 \pm 0.002\cdot 10^5 \\ 2.045 \pm 0.001\cdot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

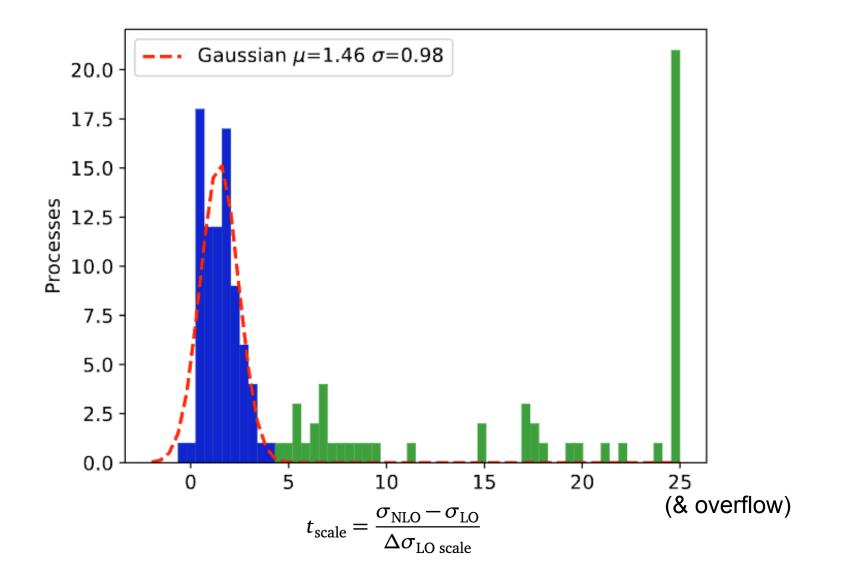
+127 more pp processes from 1405.0301!

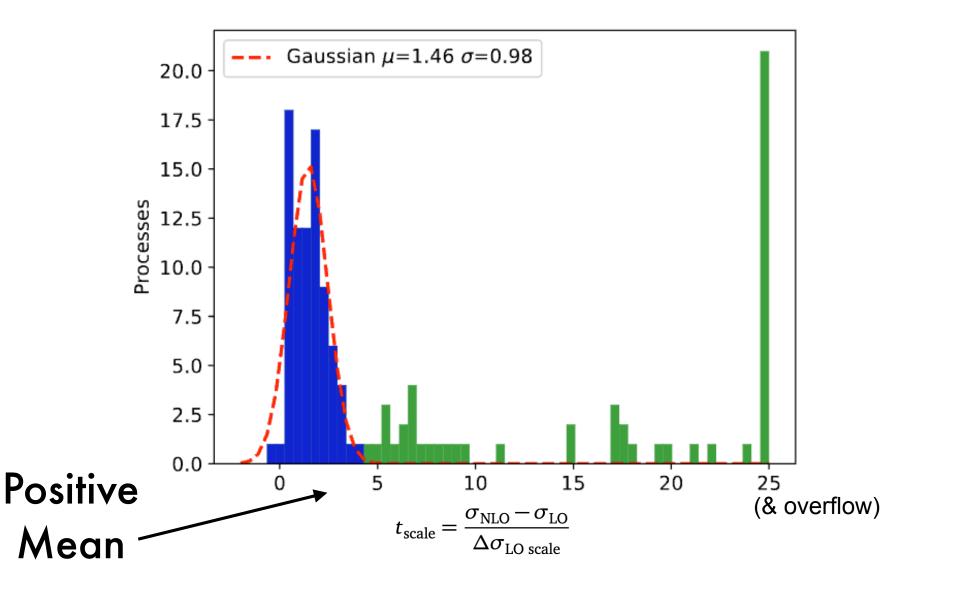
(Not a random sampling)

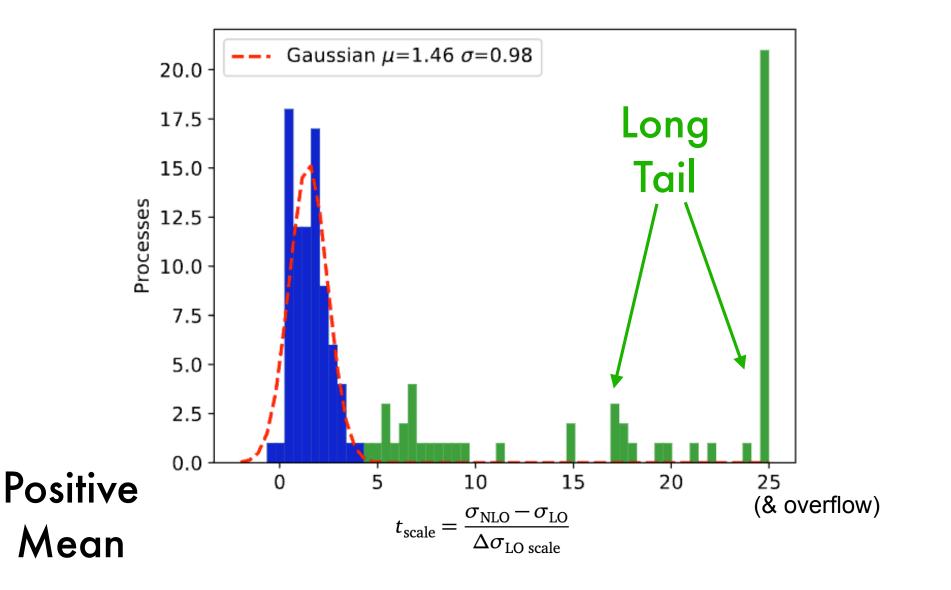
## What does it look like?

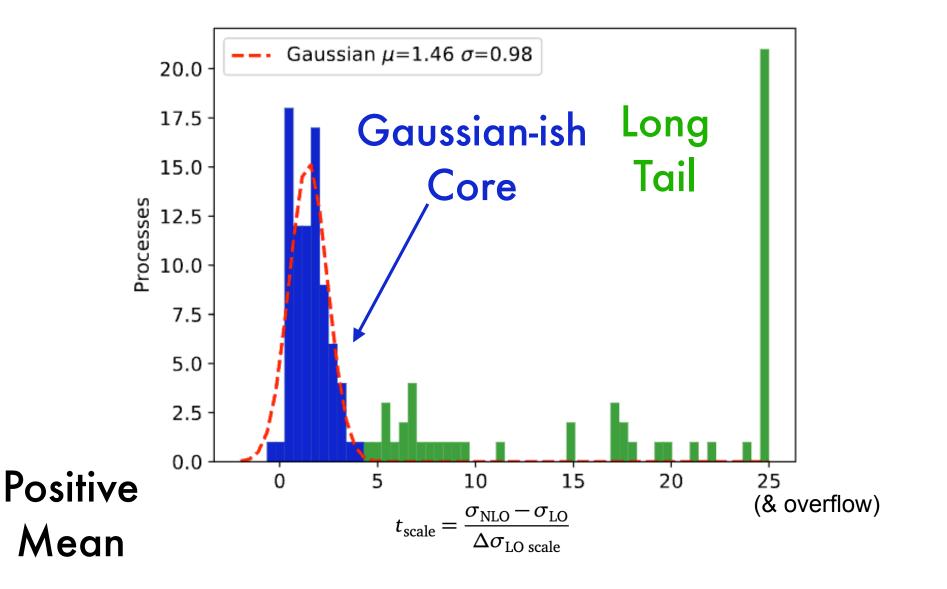
Which of these distributions do you expect?











### The tail

Pr	oc	ess		n <sub>part</sub>	$\Delta\sigma/\sigma_0$	$rac{\sigma_{ m NLO}-\sigma_0}{\Delta\sigma}$
р	р	>	wpm	1	$1.54 \times 10^{-1}$	1.84
р	р	>	wpm j	2	$1.97 \times 10^{-1}$	1.96
р	р	>	wpm j j	3	$2.45 \times 10^{-1}$	0.59
р	р	>	wpm j j j	4	$4.10 \times 10^{-1}$	0.25
р	р	>		1	$1.46 \times 10^{-1}$	1.87
р	р	>	z j	2	$1.93 \times 10^{-1}$	1.82
р	р	>	zjj	3	$2.43 \times 10^{-1}$	0.56
р	р	>	zjjj	4	$4.08 \times 10^{-1}$	0.27
р	р	>	a j	2	$3.12 \times 10^{-1}$	5.33
р	р	>	ajj	3	$3.28 \times 10^{-1}$	0.85
р	р	>	w+ w- wpm	3	$1.00 \times 10^{-3}$	610.69
р	р	>	z w+ w-	3	$8.00 \times 10^{-3}$	92.39
р	р	>	z z wpm	3	$1.00 \times 10^{-2}$	85.00
р	р	>	ZZZ	3	$1.00 \times 10^{-3}$	302.75
р	р	>	a w+ w-	3	$1.90 \times 10^{-2}$	42.33
р	р	>	a a wpm	3	$4.40 \times 10^{-2}$	47.24
р	р	>	a z wpm	3	$1.00 \times 10^{-3}$	1244.49
р	р	>	a z z	3	$2.00 \times 10^{-2}$	17.24

### The tail

Process	$n_{\rm part}$	$\Delta\sigma/\sigma_0$	$\frac{\sigma_{\rm NLO}-\sigma_0}{\Delta\sigma}$
p p > wpm	1	$1.54 \times 10^{-1}$	1.84
pp>wpmj	2	$1.97 \times 10^{-1}$	1.96
pp>wpmjj	3	$2.45  imes 10^{-1}$	0.59
pp>wpmjjj	4	$4.10 \times 10^{-1}$	0.25
p	1	$1.46 \times 10^{-1}$	1.87
pp>zj	2	$1.93 \times 10^{-1}$	1.82
pp>zjj	3	$2.43 \times 10^{-1}$	0.56
pp>zjjj	4	$4.08  imes 10^{-1}$	0.27
рр>ај	2	$3.12 \times 10^{-1}$	5.33
	2	$2.90 \times 10^{-1}$	0 QE
p p > w + w - wpm	3	$1.00 \times 10^{-3}$	610.69
p p > z w+ w-	3	$8.00 \times 10^{-3}$	92.39
p p > z z wpm	3	$1.00 \times 10^{-2}$	85.00
p p > z z z	3	$1.00 \times 10^{-3}$	302.75
pp>aw+w-	3	$1.90 \times 10^{-2}$	42.33
pp>aawpm	3	$4.40 \times 10^{-2}$	47.24
r rr	-	1 00 10-3	1044 40
pp>azwpm	3	$1.00 \times 10^{-3}$	1244.49

#### <u>What's the pattern?</u>

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	$\left  \frac{\Delta\sigma}{\sigma_0} \right $		$\frac{\Delta\sigma}{n\sigma_0}$
p p > j j	$ +2.49 \times 10^{-1} - 1.88 \times 10^{-1} $ $ +2.52 \times 10^{-1} - 1.89 \times 10^{-1} $	2 2	$+1.24 \times 10^{-1} -9.40 \times 10^{-2} +1.26 \times 10^{-1} -9.45 \times 10^{-2}$
p	$+2.90 \times 10^{-1} -2.11 \times 10^{-1}$	2	$+1.26 \times 10^{-1} - 9.43 \times 10^{-1}$ $+1.45 \times 10^{-1} - 1.06 \times 10^{-1}$
p p > j j j	$+4.38 \times 10^{-1}$ $-2.84 \times 10^{-1}$	3	$+1.46 \times 10^{-1} -9.47 \times 10^{-2}$
p	$+4.41 \times 10^{-1} -2.85 \times 10^{-1}$	3	$+1.47 \times 10^{-1} -9.50 \times 10^{-2}$
pp>ttj	$+4.51 \times 10^{-1} -2.90 \times 10^{-1}$	3	$+1.50 \times 10^{-1} - 9.67 \times 10^{-2}$
ppゝbbjj	$+6.18 \times 10^{-1}$ $-3.56 \times 10^{-1}$	4	$+1.54 \times 10^{-1}$ $-8.90 \times 10^{-2}$
p	$ +6.17 \times 10^{-1} - 3.56 \times 10^{-1} $	4	$+1.54 \times 10^{-1} - 8.90 \times 10^{-2}$
pp>ttjj	$+6.14 \times 10^{-1} - 3.56 \times 10^{-1}$	4	$+1.53 \times 10^{-1} - 8.90 \times 10^{-2}$
pp>tttt	$+6.38 \times 10^{-1} - 3.65 \times 10^{-1}$	4	$+1.60 \times 10^{-1} - 9.12 \times 10^{-2}$
p	$+6.21 \times 10^{-1} -3.57 \times 10^{-1}$	4	$+1.55 \times 10^{-1}$ $-8.93 \times 10^{-2}$
average			$+1.47 \times 10^{-1} - 9.34 \times 10^{-2}$

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

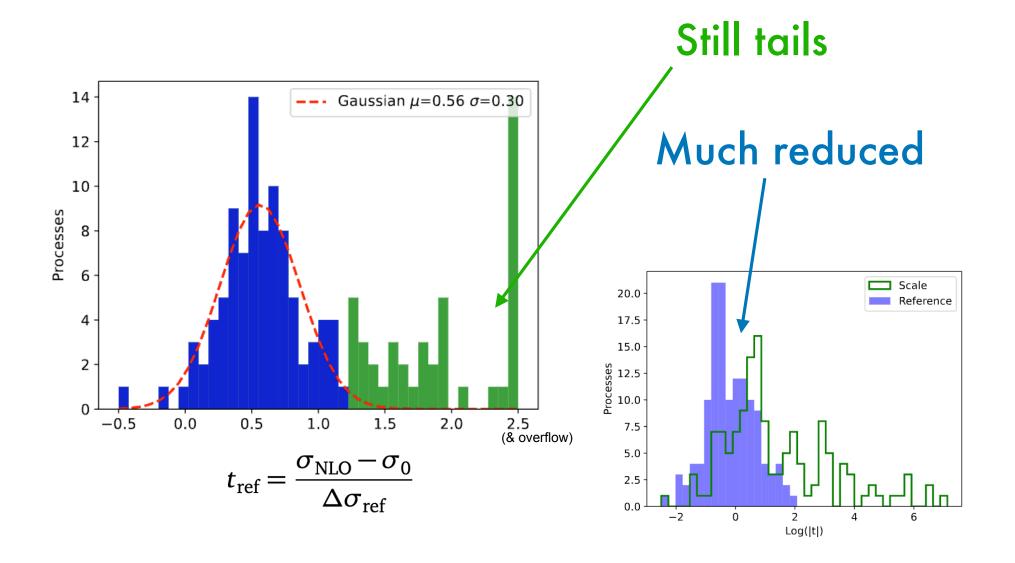
$$\frac{\Delta\sigma_{\rm ref}}{\sigma_0} = n \times \left\langle \frac{\Delta\sigma}{n\sigma_0} \right\rangle_{\rm QCD}$$

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

# Examples

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

## New pulls



# Remaining tails

Process	$n_{\rm part}$	$\Delta\sigma/\sigma_{0}$	$\frac{\sigma_{\rm NLO}-\sigma_0}{\Delta\sigma}\Big $	$\Delta\sigma_{ m ref}/\sigma_0$	$rac{\sigma_{ m NLO}-\sigma_0}{\Delta\sigma_{ m ref}}$
p p > h	1	$3.48 \times 10^{-1}$	3.02	$.47 \times 10^{-1}$	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

#### <u>How to do for NLO, NNLO?</u>

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.

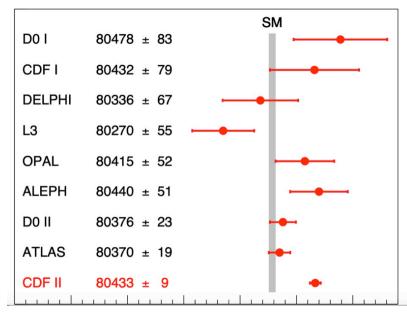
### Outline

### 1. Theoretical

# 2.Experimental

## Experimental systematics

## Do we have confidence in our understanding of the SM and its uncertainties to $7\sigma$ ?



"with a significance of 7.0 $\sigma$ " (2.5e-12)

## Experimental systematics

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

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  $\theta$  are physics parameters (masses, couplings) And  $\nu$  are nuisance parameters (resolutions etc)

## Nuisance parameters

Really, need to calculate

P(data |  $SM(\theta, \nu)$ )

To calculate p-values

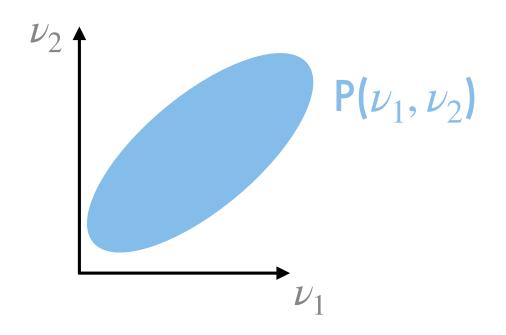
But  $\nu$  can be high-dimensional

Often profiled away from nominal

Do we understand the  $\nu$  space?

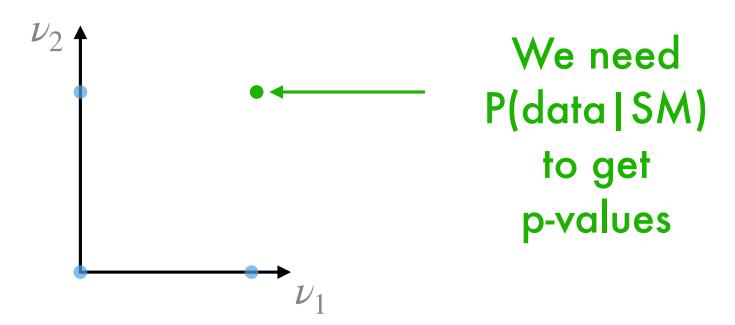
Pre-fit impact on  $\mu$ :  $\Box \theta = \hat{\theta} + \Delta \theta \qquad \Box \theta = \hat{\theta} - \Delta \theta$ -1 -0.50 0.5 ........................ Post-fit impact on µ: ATLAS  $\theta = \hat{\theta} + \Delta \hat{\theta} \qquad \theta = \hat{\theta} - \Delta \hat{\theta}$  $\sqrt{s} = 13 \text{ TeV}, 36.1 \text{ fb}^{-1}$ - Nuis, Param, Pull tt+≥1b: SHERPA5F vs. nominal tt+≥1b: SHERPA4F vs. nominal tt+≥1b: PS & hadronization tt+≥1b: ISR / FSR tTH: PS & hadronization b-tagging: mis-tag (light) NP I  $k(tt+\geq 1b) = 1.24 \pm 0.10$ Jet energy resolution: NP I tTH: cross section (QCD scale) tt+≥1b: tt+≥3b normalization tt+≥1c: SHERPA5F vs. nominal tt+≥1b: shower recoil scheme tt+≥1c: ISR / FSR Jet energy resolution: NP II tt+light: PS & hadronization Wt: diagram subtr. vs. nominal b-tagging: efficiency NP I b-tagging: mis-tag (c) NP I  $E_{\tau}^{miss}$ : soft-term resolution b-tagging: efficiency NP II -2 -1.5 -1 -0.5 0 0.5 1.5 1  $(\hat{\theta} - \theta_0) / \Delta \theta$ 

Fairly well studied:

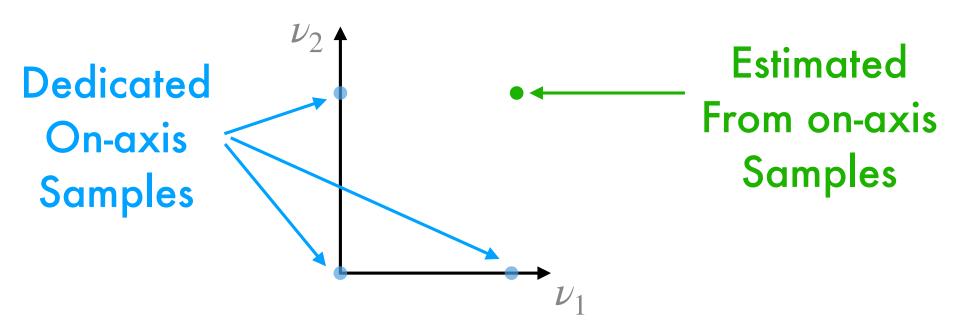


**Correlated priors** among NPs

#### Less well studied is the crucial quantity: P(data | $SM(\theta, \nu)$ )

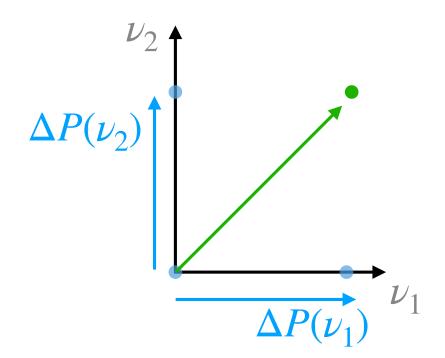


#### Less well studied is the crucial quantity: P(data | $SM(\theta, \nu)$ )



How do we do off-axis modeling?

What is being assumed here?



P(data | SM( $\theta, \nu_1, \nu_2$ )) can be approximated from P(data | SM( $\theta, 0, 0$ )),  $\Delta P(\nu_1), \Delta P(\nu_2)$ 

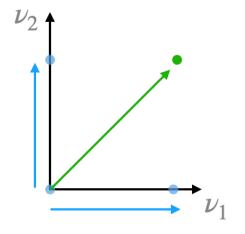
#### P(data | SM( $\theta, \nu_1, \nu_2$ )) can be approximated from P(data | SM( $\theta, 0, 0$ )), $\Delta P(\nu_1), \Delta P(\nu_2)$

#### Several possible approaches:

Linear extrapolation Morphing Factorizing

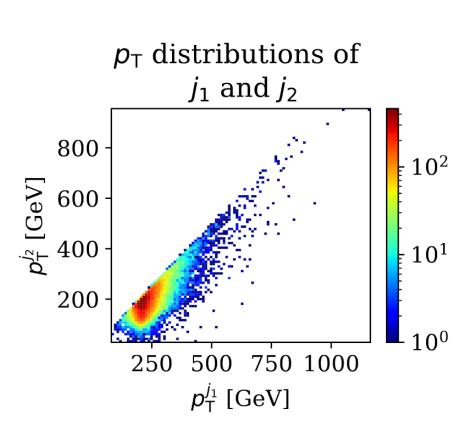
#### All assume

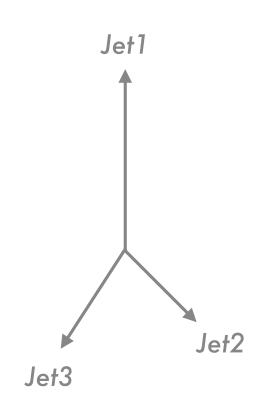
Impact are independent of other NPs (Not the same as correlation of NP prior)



#### Example

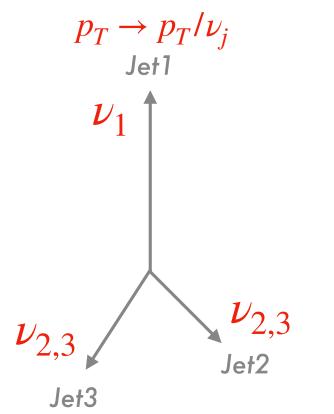
#### <u>3-jet event</u> 1 high pT jet 2 low pT jets





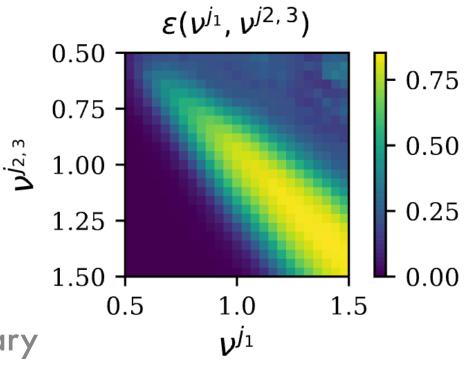
## Example

#### Jet energy uncertainty



#### <u>3-jet event</u>

Selection: MET<50 GeV What is efficiency vs NPs?

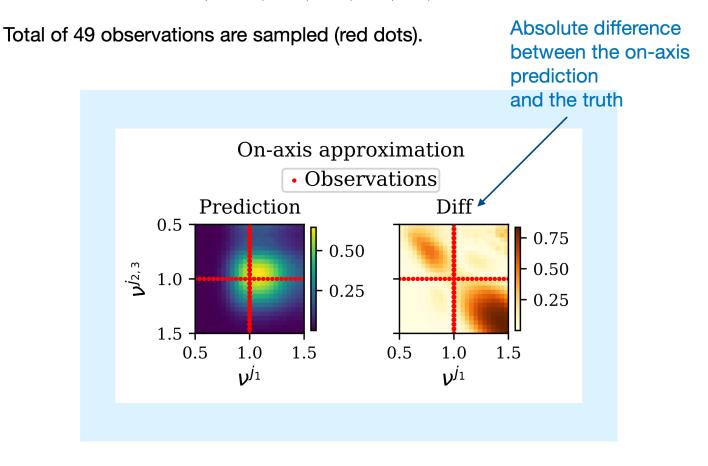


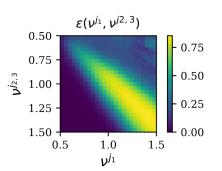
Efficiency high when BOTH  $\nu$  vary Efficiency drops when one varies

## On-axis approximations

#### **On-axis approximation**

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





## What to do?

0.75

0.50

0.25

1.5

# Other approaches0.50Naive scan - expensive0.50Our approach1.25Our approach1.50Probe space with minimal points $y^{j_1}$

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,<sup>1</sup> Kyle Cranmer,<sup>2</sup> and Daniel Whiteson<sup>1</sup>

Basics of GPs with derivative information:

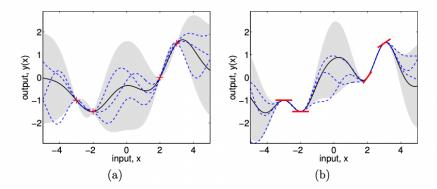


Figure 9.1: In panel (a) we show four data points in a one dimensional noise-free regression problem, together with three functions sampled from the posterior and the 95% confidence region in light grey. In panel (b) the same observations have been augmented by noise-free derivative information, indicated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

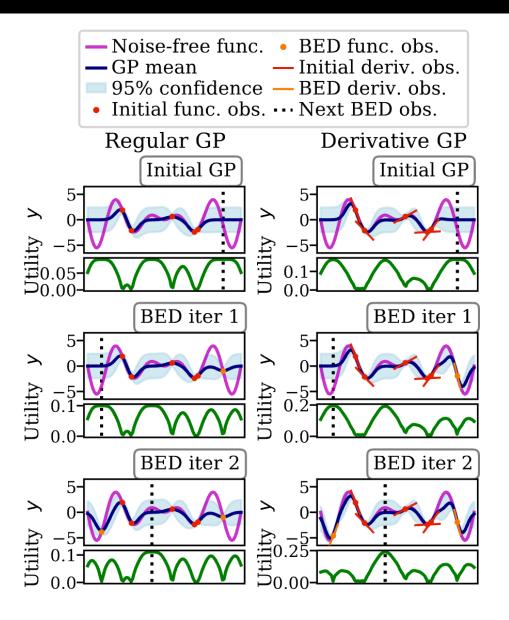
http://gaussianprocess.org/gpml/chapters/RW9.pdf

#### Procedure

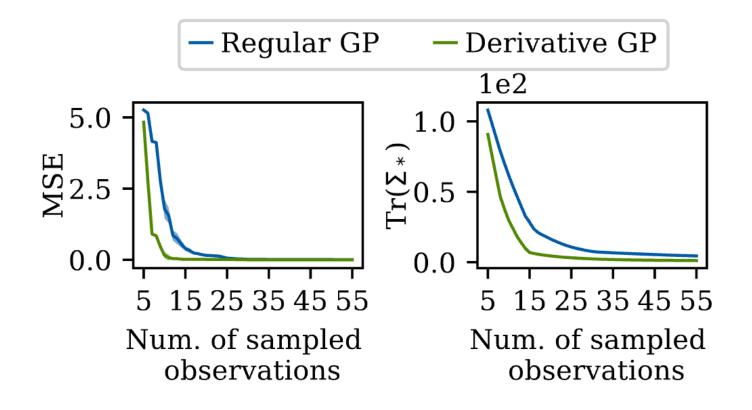
#### Initial coarse scan

Build GP

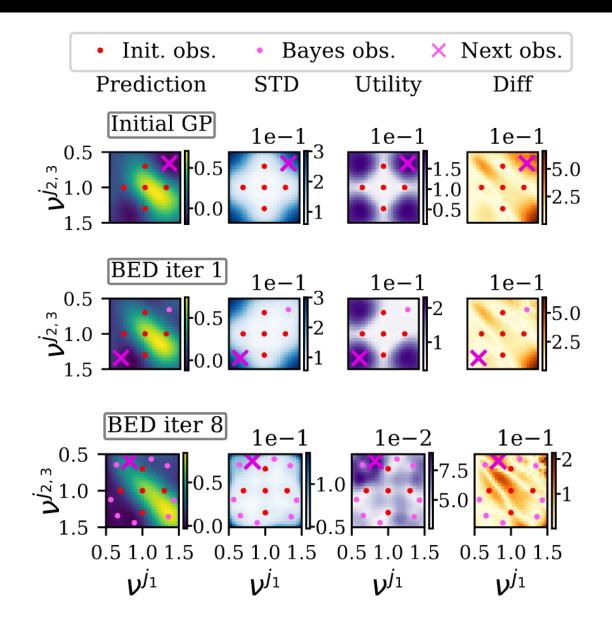
Choose sample points to minimize expected uncertainty  $Tr[\Sigma_Q]$ 



## Results - Toy demo

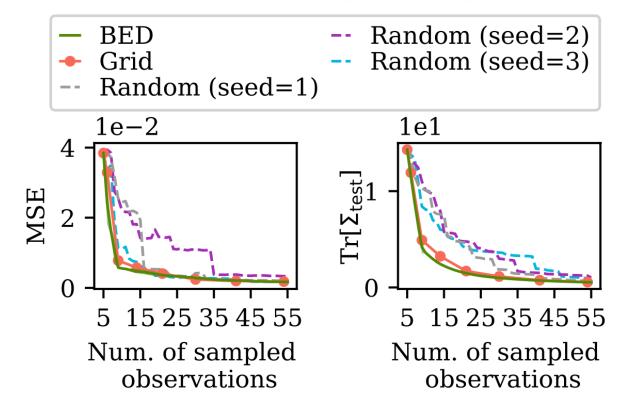


## Jet problem



#### Performance

Predicting the full efficiency space w/various sample strategies using derivative GPs



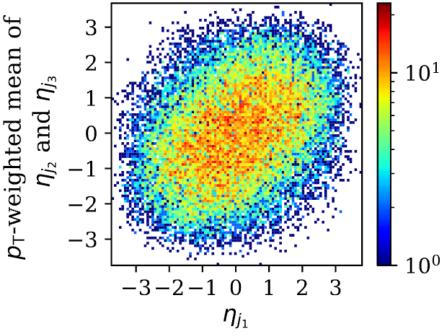
## 4D space

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

#### Pseudorapidity distributions

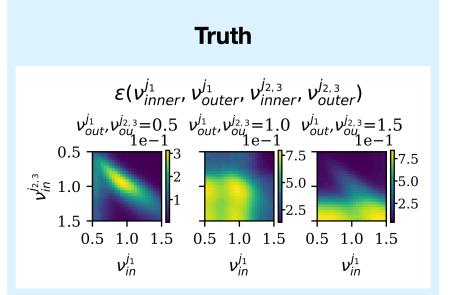


## 4D space

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



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 ( $\nu_{in}^{j_1}$ ,  $\nu_{in}^{j_{2,3}}$ )

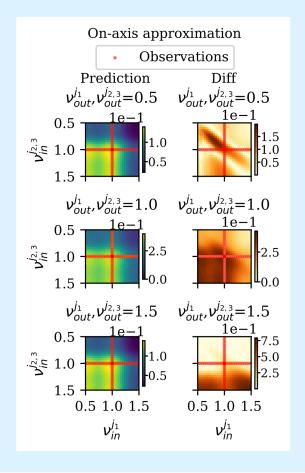
## Approximations

#### **On-axis approximation**

Total of 97 observations along the central axes.

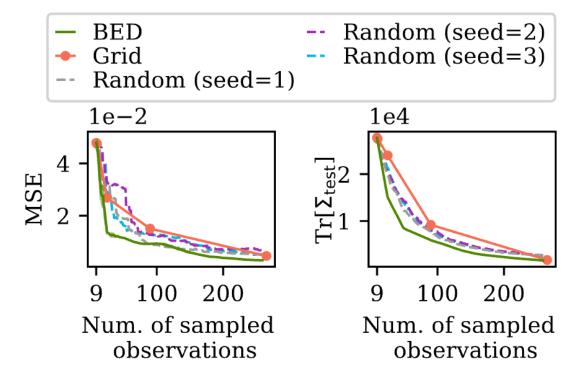
The off-axis values are calculated according to

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

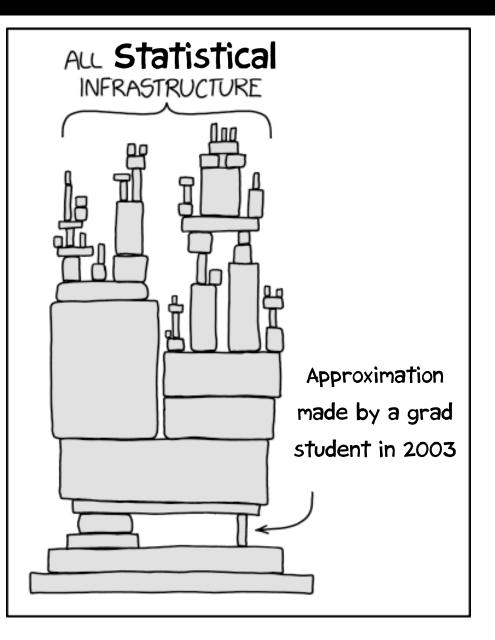


#### Results

Predicting the full efficiency space w/various sample strategies using derivative GPs



#### 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

