Evaluating and reporting systematic uncertainties in flavour physics at the LHC

Roger Barlow
The University of Huddersfield

PHYSTAT-Flavour @ CERN (virtually)

19^{th} October 2020

Contents

- Systematics in principle
- Systematics in practice:

examples from 4 recent results from

ATLAS

CMS

LHCb

- Summary and conclusions

With acknowledgements and hearty thanks to:

Nick Wardle, Bogdan Malescu, Matt Kenzie and Hans Dembinski

and apologies for my non-expert descriptions of beautiful analyses

Systematic Uncertainties - or 'nuisance parameters'

A taxonomy

- Continuous uncertainties with explicit consequences.
  Typical examples: acceptance error, luminosity error.
- Continuous uncertainties with implicit consequences.
  Typical examples: MC model parameters
- Discrete uncertainties.
  Typical examples: choice of MC model (Herwig/Pythia)

May be frequentist (often through an ancillary experiment) or Bayesian

Correlations need careful handling (off topic but see backup slides)

Possible confusion

Not the same as checks for inconsistencies/impossibilities.

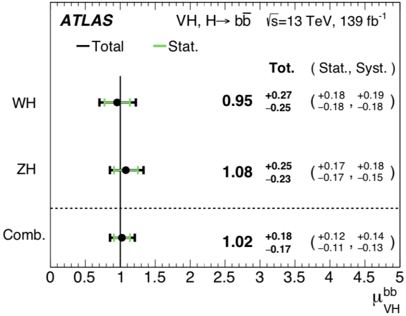
Warning!!!

Danger of inflation. For errors, "conservative" is another word for "wrong"

What happens when principle meets practice?

ATLAS1: WH and ZH production in the $H \to b\bar{b}$ channel

arXiv:2007.02873v1



Systematic Uncertainties

(1) Experimental
(2) Signal Modelling
(3) Background Modelling

BDTs used for final signal discrimination

ATLAS1: continued - Experimental uncertainties

Experimental

1. $b-$tagging data/MC correction factors, for various topologies and event properties.. From data using $t\bar{t}$ samples with $t \to W(\to \ell\nu)b$ decays. 57 separate uncertainties *after* negligible ones pruned.
2. Jet Energy Scale: 30 components. From data using P_T balance with γ or Z or multijet
3. Jet Energy Resolution: 8 components, as above. Plus separate uncertainty for b and c jet calibration.
4. Lepton id and reco uncertainties - small effect.
5. E_T^{miss} values and different trigger efficiencies
6. Luminosity uncertainty 1.7% from LUCID2
7. μ in simulation scaled by 1.03 to improve agreement: uncertainty 0.03 applied.

ATLAS1: continued - signal uncertainties

Table 8: Summary of the systematic uncertainties in the signal modelling. 'PS/UE' indicates parton shower/underlying event. An 'M+S' symbol is used when a shape uncertainty includes a migration effect that allows relative acceptance changes between regions. Instances where an uncertainty is considered independently in different regions are detailed in parenthesis. Where the size of an acceptance systematic uncertainty varies between regions, a range is displayed.

	Signal
Cross-section (scale)	0.7% (qq), 25% (gg)
$H \to b\bar{b}$ branching fraction	1.7%
Scale variations in STXS bins	3.0%–3.9% ($qq \to WH$), 6.7%–12% ($qq \to ZH$), 37%–100% ($gg \to ZH$)
PS/UE variations in STXS bins	1%–5% for $qq \to VH$, 5%–20% for $gg \to ZH$
PDF+α_S variations in STXS bins	1.8%–2.2% ($qq \to WH$), 1.4%–1.7% ($qq \to ZH$), 2.9%–3.3% ($gg \to ZH$)
m_{bb} from scale variations	M+S ($qq \to VH$, $gg \to ZH$)
m_{bb} from PS/UE variations	M+S
m_{bb} from PDF+α_S variations	M+S
p_T^V from NLO EW correction	M+S

STXS = Simplified Template Cross Section. Particular Higgs production mode in particular kinematic region

Ranges given where error varies for different kinematic regions

PS/UE = parton shower/underlying event model variations from tuning and Powheg/Herwig differences

M+S Migration+Shape uncertainty

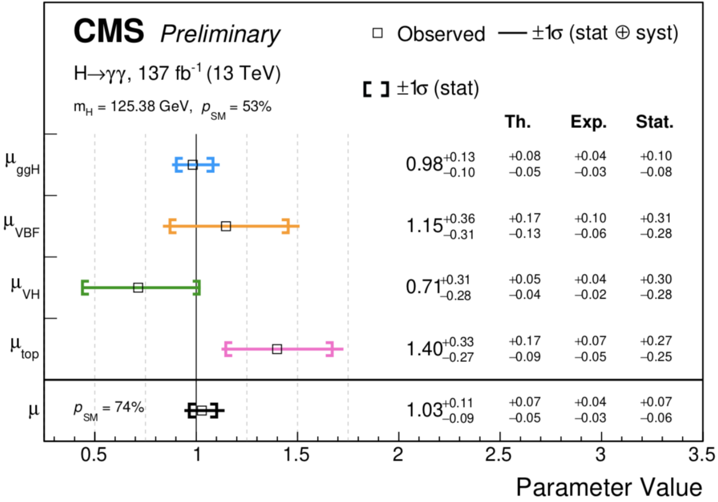
'unless stated otherwise, the uncertainty is taken from the alternative sample that differs most in shape from the nominal sample'

ATLAS1: continued - background uncertainties

Like signal uncertainties but more of them (58 listed)
Uncertainties in size, shape, and relative acceptance. Backgrounds from: $Z + jets$,W+jets, $t\bar{t}$, single t, multijets, ZZ, WZ, WW
Main ones in fit, others from calculated cross sections

R_{BDT}

Need to look at effect on shape of BDT disciminant. For small $Z + HF$, single-t, VV backgrounds just vary by uncertainty and look at input variable variation. For larger $W + HF$ and $t\bar{t}$ use likelihood-free inference (Cranmer et al. arXiv 1506.02169):
For p_T^V variation: train another BDT (BDT$_S$)to discriminate between nominal and alternative sample using all variables *except* p_T^V.
Use ratio R_{BDT} of BDT$_S/BDT$ outputs to correct between samples to see effect of change.



CMS Preliminary
Observed
±1σ (stat ⊕ syst)
[ ] ±1σ (stat)
H→γγ, 137 fb-1 (13 TeV)
mH = 125.38 GeV, pSM = 53%
Th.
Exp.
Stat.
μggH
0.98 +0.13 −0.10
+0.08 −0.05
+0.04 −0.03
+0.10 −0.08
μVBF
1.15 +0.36 −0.31
+0.17 −0.13
+0.10 −0.06
+0.31 −0.28
μVH
0.71 +0.31 −0.28
+0.05 −0.04
+0.04 −0.02
+0.30 −0.28
μtop
1.40 +0.33 −0.27
+0.17 −0.09
+0.07 −0.05
+0.27 −0.25
μ
pSM = 74%
1.03 +0.11 −0.09
+0.07 −0.05
+0.04 −0.03
+0.07 −0.06
0.5
1
1.5
2
2.5
3
3.5
Parameter Value

CMS1:continued

Consider several different categories of production (STXS)

9 theoretical uncertainties

QCD scales (multiply and divide by factor ~ 2).
ggH and qqH STXS fractions, vary parameters. Include migration effects.
PDF and α_s uncertainties use well-established prescription
$H \to \gamma\gamma$ BR 2% uncertainty

16 experimental uncertainties:

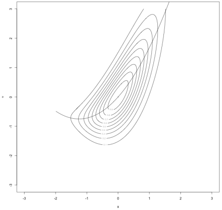
Photon energy scale and nonlinearity from $Z \to e^+e^-$, shower shape corrections, light collection nonuniformity, modelling material, modelling underlying event,luminosity, photon ID, JES and JER, photon energy resolution and effiiciency,, trigger efficiency, P_T^{miss} resolution, pileup jet ID, lepton ID and isolation, b tagging efficiency.

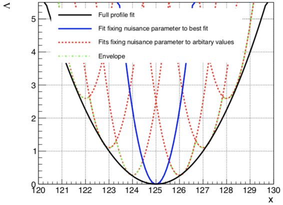
Extracted from data by many ingenious means, or from simulation comparisons.

CMS1:continued

Different background models (poly, exponential...) give different results.
Should include as systematic uncertainty as we don't know the true form
Can take rms spread of results of several models.

Smarter alternative: incorporate nature of model as (discrete) nuisance parameter and use in profile likelihood (P. D. Dauncey, M. Kenzie, N. Wardle and G. J. Davies, arXiv:1408.6865v5)





Take the envelope of the likelihood curves.
Poor models automatically disqualify themselves.
Models with more parameters need a penalty

LHCB1: Study of the $\psi_2(3823)$ and $\chi_{c1}(3872)$ states in $B^+ \to (J/\psi\pi^+\pi^-)K^+$ decays

JHEP 2008 123 (2020) arXiv:2005.13422

Measure ratios of branching fractions and mass differences

Systematics table for BFs
(similar table for mass differences)
Signal templates by varying models
Polynomial from 1st and 2rd order
(using Toy MC)
ψ_2 decay using physics-based model as opposed to phase space, varying unknown interference phase

Table 2: Relative systematic uncertainties (in %) for the ratios of branching fractions $\mathcal{R}^{\mathrm{X}}_{\mathrm{Y}}$.

Source	$\mathcal{R}^{\psi_2(3823)}_{\chi_{c1}(3872)}$	$\mathcal{R}^{\psi_2(3823)}_{\psi(2S)}$	$\mathcal{R}^{\chi_{c1}(3872)}_{\psi(2S)}$
Signal and background shapes			
B^+ signal template	0.6	0.5	0.1
$X_{c\bar{c}}$ signal template	0.3	0.2	0.2
Polynomial components	2.5	2.7	0.2
$\psi_2(3823)$ decay model	0.2	0.2	—
Efficiency corrections	< 0.1	0.2	0.2
Trigger efficiency	1.1	1.1	1.1
Data-simulation agreement	1.0	1.0	1.0
Simulation sample size	0.3	0.4	0.4
Sum in quadrature	3.0	3.2	1.6

Trigger efficiency from MC, corrected by known modes e.g. $B^+ \to J/\psi K^+$
Data-simulation agreement refers to uncertainty in tracking and PID efficiency corrections.

LHCB1 - continued

Good news

Identification of sources comprehensive and sensible

Bad news

The 'error' is taken from the range - over-conservative and inflationary.

between 270 and 302 keV [2]. The maximal deviations in the ratios $\mathcal{R}^{\mathrm{X}}_{\mathrm{Y}}$ with respect to the baseline fit model are taken as systematic uncertainties for each of the systematic signal model sources. For the systematic uncertainty related to the modelling of the smooth

---

the first to the second order, separately for each fit component and each channel. In each case the ratio $\mathcal{R}^{\mathrm{X}}_{\mathrm{Y}}$ is computed and the maximal difference with respect to the baseline fit model is taken as a corresponding systematic uncertainty. For each choice of the fit

---

is found to be stable. It varies within 0.2% with respect to the efficiency computed for the phase-space model when the unknown phase Φ varies in the range $-\pi \leq \Phi < \pi$.

---

to ±20% change in the measured efficiency. The resulting variations in the efficiency ratios do not exceed 1%, which is taken as a corresponding systematic uncertainty. The last

ATLAS2: Measurement of ϕ_2 in $B_s^0 \to J/\psi\phi$ decays

arXiv:2001.07115v3

The Results

$$\phi_s = -0.087 \pm 0.036 \text{ (stat.)} \pm 0.019 \text{ (syst.) rad}$$
$$\Delta\Gamma_s = 0.0641 \pm 0.0043 \text{ (stat.)} \pm 0.0024 \text{ (syst.) ps}^{-1}$$
$$\Gamma_s = 0.6697 \pm 0.0014 \text{ (stat.)} \pm 0.0015 \text{ (syst.) ps}^{-1}$$

The Systematic Uncertainties

Table 5: Summary of systematic uncertainties assigned to the physical parameters of interest.

	ϕ_s [10^{-3} rad]	$\Delta\Gamma_s$ [10^{-3} ps^{-1}]	Γ_s [10^{-3} ps^{-1}]	$\lvert A_\parallel(0)\rvert^2$ [10^{-3}]	$\lvert A_0(0)\rvert^2$ [10^{-3}]	$\lvert A_S(0)\rvert^2$ [10^{-3}]	$\delta_\perp$ [10^{-3} rad]	$\delta_\parallel$ [10^{-3} rad]	$\delta_\perp - \delta_S$ [10^{-3} rad]
Tagging	19	0.4	0.3	0.2	0.2	1.1	17	19	2.3
Acceptance	0.5	< 0.1	< 0.1	1.0	0.8	2.6	33	56	7.0
ID alignment	0.8	0.2	0.5	< 0.1	< 0.1	< 0.1	11	7.2	< 0.1
Best candidate selection	0.5	0.4	0.7	0.5	0.2	0.2	12	17	7.5
Background angles model:									
Choice of fit function	2.5	< 0.1	0.3	1.1	< 0.1	0.6	12	0.9	1.1
Choice of p_T bins	1.3	0.5	< 0.1	0.4	0.5	1.2	1.5	7.2	1.0
Choice of mass interval	0.4	0.1	0.1	0.3	0.3	1.3	4.4	7.4	2.3
Dedicated backgrounds:									
B_d^0	2.3	1.1	< 0.1	0.2	3.0	1.5	10	23	2.1
Λ_b	1.6	0.3	0.2	0.5	1.2	1.8	14	30	0.8
Alternate Δm_s	1.0	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	15	4.0	< 0.1
Fit model:									
Time res. sig frac	1.4	1.1	0.5	0.5	0.6	0.8	12	30	0.4
Time res. p_T bins	0.7	0.5	0.8	0.1	0.1	0.1	2.2	14	0.7
S-wave phase	0.3	< 0.1	< 0.1	< 0.1	< 0.1	0.2	8	15	37
Fit bias	5.7	1.3	1.2	1.3	0.4	1.1	3.3	19	0.3
Total	20	2.2	1.8	2.2	3.4	4.4	51	84	38

ATLAS2: Continued

1. Flavour Tagging: power-weighted cone charge. Change functions used to parametrise performance.
   Largest difference taken as uncertainty.
2. Angular Acceptance. Vary binning
3. ID misalignment. Vary alignment and study variation in fit parameters
4. Trigger efficiency decay time dependence - negligible
5. Choice of best candidate in events with more than 1. Try highest p_T rather than lowest χ^2. Small effect.
6. Background model angular form. 14 $\rightarrow$ 16 order Legendre polynomial and different p_T binning taking largest changes
7. Uncertainty in B_d and Λ_b mis-id background modelling
8. Let Δm_s float
9. Fit model mass (2nd Gaussian) and lifetime change p_T binning.
10. S-wave phase vary $\alpha = 0.51 \pm 0.08$
11. Fit bias from consistency checks fitting data from default fit model

CMS2: Search for a narrow resonance in high-mass dilepton final states

CMS PAS EXO-19-019

Z' search - peak in dilepton mass

Background from Drell-Yan, also t and W

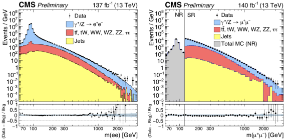


Table 2: Systematic uncertainties considered in this analysis and their magnitude.

Uncertainty source	Magnitude
Lepton selection efficiency	1–5% (two-sided), 0–6.5% (one-sided)
Mass scale uncertainty	1–3%
Mass resolution uncertainty	0–15%
Z peak normalization	1–5%
Integrated luminosity	2.3–2.5%
DY theoretical cross section	0–20%
Multijet background normalization	50–100%
Cross section for other simulated backgrounds	7%

Systematic uncertainties in efficiency and mass scale, also in background

Evaluated experimentally from Z peak

Then used as nuisance parameters in Bayesian limit setting with specific priors and quoting no robustness checks

Limits are calculated at 95% confidence level (CL) with Bayesian techniques known to have good frequentist coverage properties [52], using the framework developed for statistically combining Higgs boson searches [53], which is based on the ROOSTATS package [54]. For the signal cross section, we use a positive uniform prior, while the nuisance parameters for the uncertainties in dilepton efficiencies, resolution, and scale are modeled with log-normal priors. To allow the background yield to be constrained by its statistical uncertainties, a log-normal prior with a width of three times these uncertainties is used.

Time-dependent CP analysis of $B^0 \to D^{*\pm} D^{\mp}$

Table 4: Summary of the systematic uncertainties. The total systematic uncertainties are computed as quadratic sum of individual contributions.

Source	ΔC_{D^*D}	C_{D^*D}	ΔS_{D^*D}	S_{D^*D}
Fit bias	0.002	0.002	0.002	0.002
Mass model	0.006	0.014	0.003	0.011
$\Delta m_d, \tau_d, \Delta\Gamma_d$	0.001	0.003	0.001	0.001
Decay-time resolution	<0.001	<0.001	<0.001	<0.001
Decay-time acceptance	<0.001	<0.001	<0.001	<0.001
Flavour tagging	0.015	0.014	0.012	0.015
Total syst. uncertainty	0.016	0.020	0.012	0.019

Source	$\mathcal{A}_{\text{raw}}^{K\pi\pi\pi,\text{Run1}}$	$\mathcal{A}_{\text{raw}}^{K\pi\pi\pi,\text{Run2}}$	$\mathcal{A}_{\text{raw}}^{K\pi,\text{Run1}}$	$\mathcal{A}_{\text{raw}}^{K\pi,\text{Run2}}$
Fit bias	0.0013	0.0007	0.0008	0.0004
Mass model	0.0025	0.0024	0.0021	0.0016
$\Delta m_d, \tau_d, \Delta\Gamma_d$	0.0003	0.0002	0.0002	0.0001
Decay-time resolution	0.0002	0.0001	0.0001	0.0001
Decay-time acceptance	0.0003	0.0001	0.0002	0.0001
Flavour tagging	0.0001	0.0001	0.0001	0.0001
Total syst. uncertainty	0.0028	0.0025	0.0023	0.0016

LHCB2 - continued

Check the statistical error produced by the fit using bootstrapped samples

Fit Bias - many fits with toy MC. No bias found BUT uncertainty on mean value of bias taken as uncertainty on parameter. Cautious.

Mass Model: try different ones (order of polynomials, number of Xtal ball curves...) using toy MC pseudoexperiments, fitting nominal and other model. Then

with the alternative model. Results of the subsequent decay-time fit are compared to those obtained with the nominal fit and the distribution of their difference is built. The systematic uncertainty is defined as the sum in quadrature of the average and root mean square of the distribution.

This looks like a possible over-estimate

Many Checks by subsamples: D^0 final state, tagging algorithm, magnet polarity. Also B_s^0 contamination, spline knots, etc All within 2 sigma so nothing assigned.

Flavour tagging: uncertainties largely included in statistical error, also consider nonlinearity in SS tag probability and dependence on OS tagger on decay time, evaluated with pseudoexperiments.

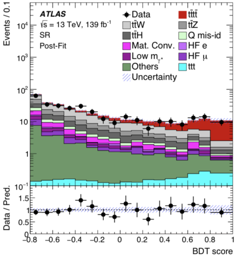
ATLAS3: Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state

arXiv:2007.14858v2

3 isolated leptons, or 2 with same charge
Irreducible background from leptons from W, Z, τ
Reducible background from fake/mis-id leptons
Analysis in terms of BDT score

$$\mu = 2.0 \pm 0.4(\text{stat})\,^{+0.7}_{-0.4}(\text{syst}) = 2.0\,^{+0.8}_{-0.6}.$$

$$\sigma_{t\bar{t}t\bar{t}} = 24 \pm 5(\text{stat})\,^{+5}_{-4}(\text{syst})\text{ fb} = 24\,^{+7}_{-6}\text{ fb}.$$



ATLAS3: continued

Experimental

- Luminosity
- Electron/muon MC correction factors
- Jet Energy Scale and resolution
- JVT (Jet Vertex Tagger - to remove pile-up events) efficiency effect from $\pm\sigma$ variation
- $b-$tagging efficiency/mistag from separate study using $t\bar{t}$ enriched sample
- E_T^{miss} uncertainty from data-MC comparisons

Modelling

- vary renormalisation/factorisation scales by factors of 2
- different model for parton shower/hadronisation (details not given)
- pdf uncertainty from rms of 100 replicas
- uncertainties on background cross sections. Informed guesses(lots of 50% figures) , or theory/expt discrepancies. But these backgrounds are small

ATLAS3: continued

Many checks performed and NOT added to systematic error

The stability of the result has been checked. The fit was repeated with the data split according to year or by splitting the signal region into two regions with either same-sign dilepton events or events with at least three leptons. Different fits were also performed by using only positively charged same-sign lepton pairs or only negatively charged same-sign lepton pairs. All these tests showed compatible μ values.

An additional test was performed by splitting the SR into five regions according to the number of leptons and b-tagged jets and by fitting the H_{T} distribution in each region. The BDT score is therefore not used in this test. The observed (expected) significance is found to be 4.3 (2.1) and the fitted signal strength is $2.2^{+0.9}_{-0.6}$. This result is consistent with the result from the default fit.

CMS3:Top quark differential cross sections

arXiv:2008.07860v1

12 Experimental uncertainties, for separate all-jet and lepton-jet channels

1. Multijet background to all-jet. Fitted. Estimate $\sim 1\%$ from simulation
2. Single t, W etc background to all-jet. Change cross sections by 50%, negligible shift. Probably overkill but contribution very small
3. Background in $\ell + jets$. Apply large uncertainties (30%-50%), get small changes (BG is fitted)
4. JES. Major factor and well studied. Vary 24 components
5. JER. Smear by JES. Small effect
6. t tag efficiency. From fit.
7. b tag efficiency (all-jet). Major factor and well studied.
8. b tag efficiency (ℓ-jet). Major factor and well studied.
9. Pileup modelling. Change cross section used in reweighting. Negligible effect
10. Trigger efficiency from orthogonal trigger.Small effect
11. Lepton ID efficiency from tag-and-probe

CMS3
continued

6 theoretical uncertainties. 3 from matrix elements studied through event weights, 3 from parton shower studied through separate samples.

1. PDFs. Standard deviation from 100 replicas of NNPDF
2. Renormalisation and factorisation scales. Take factors of 2 up or down and envelope of results
3. α_s vary by ± 0.001
4. ISR and FSR uncertainty. Change scales by (large) factors.
5. ME/PS matching. Change h_{damp} by (large) factors.
6. MC tune - vary CUETP8M2T4 parameters by $\pm 1\sigma$.

LHCB3: Measurement of the branching fractions for $B^+ \to D^{*+}D^-K^+$, $B^+ \to D^{*-}D^+K^+$, and $B^0 \to D^{*-}D^0K^+$ decays

arXiv:2005.10264v2

Table 3: Systematic uncertainties on $\mathcal{N}^{\text{corr}}$ from the signal PDF parameters (σ_{PDF}), the finite simulation samples (σ_{MC}), the PID resampling (σ_{PID}), the residual peaking background (σ_{bkg}), and the total systematic uncertainty ($\sigma_{\text{tot.}}$). All values are given as a percentage of the central value of $\mathcal{N}^{\text{corr}}$.

Decay channel	Run 1 (%)					Run 2 (%)				
	σ_{PDF}	σ_{MC}	σ_{PID}	σ_{bkg}	$\sigma_{\text{tot.}}$	σ_{PDF}	σ_{MC}	σ_{PID}	σ_{bkg}	$\sigma_{\text{tot.}}$
$B^+ \to D^{*+}_{K\pi}D^-K^+$	0.6	0.8	1.5	0.8	2.0	0.5	1.4	0.2	0.5	1.6
$B^+ \to D^{*+}_{K3\pi}D^-K^+$	1.2	1.2	0.9	1.4	2.4	1.0	2.1	0.7	0.6	2.5
$B^+ \to D^{*-}_{K\pi}D^+K^+$	0.5	1.0	0.4	0.7	1.4	0.8	1.8	0.7	0.4	2.1
$B^+ \to D^{*-}_{K3\pi}D^+K^+$	1.4	1.6	1.1	1.2	2.7	0.7	2.5	1.2	0.6	2.9
$B^0 \to D^{*-}_{K\pi}D^0_{K\pi}K^+$	0.6	0.7	0.9	0.3	1.3	0.5	1.1	0.2	0.2	1.2
$B^0 \to D^{*-}_{K3\pi}D^0_{K\pi}K^+$	0.8	1.2	0.3	0.7	1.6	0.8	1.7	0.6	0.3	2.0
$B^0 \to D^{*-}_{K\pi}D^0_{K3\pi}K^+$	0.9	1.2	0.3	0.6	1.6	0.6	2.0	0.3	0.3	2.1
$B^+ \to \overline{D}^0_{K3\pi}D^0_{K\pi}K^+$	0.6	1.1	1.0	0.9	1.8	1.1	1.8	0.5	0.4	2.2
$B^+ \to \overline{D}^0_{K\pi}D^0_{K3\pi}K^+$	0.7	1.1	0.5	0.7	1.6	0.7	1.6	0.4	0.3	1.8
$B^0 \to D^-D^0_{K\pi}K^+$	0.4	0.7	0.5	0.4	1.0	0.3	0.7	0.7	0.2	1.1
$B^0 \to D^-D^0_{K3\pi}K^+$	0.2	1.4	0.3	0.5	1.5	0.8	1.3	0.4	0.3	1.6

LHCB3: continued

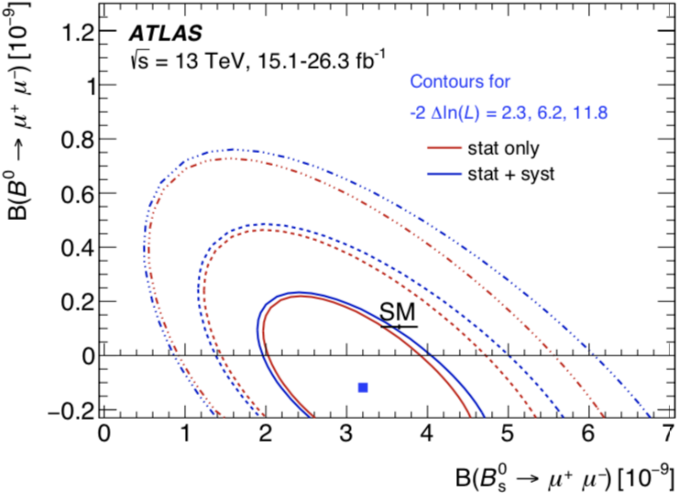
Two nice methods to note

- Signal shapes described by DSCB (Double Sided Crystal Ball) function.
  Uncertainty in tail parameters.
  Rather than take $\pm\sigma$ and refit to get σ_{BR}, do many refits
  with Gaussian random samplings of parameters, and take RMS

- Lack of MC data limits efficiency estimates.
  Repeat using bootstrapped samples from the MC data
  to get good estimate of uncertainty in efficiency

ATLAS4: Study of B^0 and B^0_s decays into muon pairs

JHEP 04 (2019) 098, arXiv:1812.03017v2



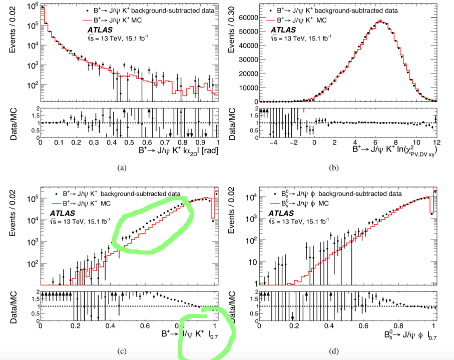
ATLAS4:continued

Measure $B^0_{(s)} \to \mu^+\mu^-$ relative to common $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$

Source	B_s^0 [%]	B^0 [%]
f_s/f_d	5.1	-
B^+ yield	4.8	4.8
R_ε	4.1	4.1
$\mathcal{B}(B^+ \to J/\psi\, K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)$	2.9	2.9
Fit systematic uncertainties	8.7	65
Stat. uncertainty (from likelihood est.)	27	150

Table 2: Summary of the uncertainties in R_ε.

Source	Contribution [%]
Statistical	0.8
Kinematic reweighting (DDW)	0.8
Muon trigger and reconstruction	1.0
BDT input variables	3.2
Kaon tracking efficiency	1.5
Pile-up reweighting	0.6



Get systematics from pseudo-experiment fits varying assumptions (signal properties, reconstruction, backgrounds, data from sidebands...)

Says 'was used to evaluate systematic uncertainties'

but don't say how... hope it was quadrature not range

CMS4: Search for dark matter in association with $Z \to e^+e^-$ or $\mu^+\mu^-$ (Z with missing P_T at some M_T)

arXiv :2008.04735v1

Write down binned likelihood with $\prod e^{(\theta_i-\hat{\theta}_i)^2/2}$ for nuisance parameters

Source of uncertainty	Impact assuming signal	Impact assuming no signal
Integrated luminosity	0.013	0.002
Lepton measurement	0.032	0.050
Jet energy scale and resolution	0.042	0.024
Pileup	0.012	0.09
b tagging efficiency	0.004	0.002
Theory	0.088	0.085
Simulation sample size	0.024	0.023
Total systematic uncertainty	0.11	0.11
Statistical uncertainty	0.089	0.073
Total uncertainty	0.14	0.13

This is an example of one particular model - Zh, where h may or may not decay to *invisible*.

Shift parameters by ± 1 sigma. Correlation included or not ,as appropriate

Biggest source 'theory' - renorm and factorisation scales, α_s, pdfs, EWK corrections

LHCB4: CP observables in $B^{\pm} \rightarrow DK^{\pm}$ and $B^{\pm} \rightarrow D\pi^{\pm}$ with $D \rightarrow K_s^0 K^{\pm}\pi^{\mp}$

JHEP 2020, 58 (2020), arXiv:2002.08858v2

D is mixture of D^0 and $\overline{D^0}$. Analysed using Dalitz plots. Separate results inside and outside K^{*+} band and for relative charges (OS,SS) of B and π in D decay.

Measure A, asymmetries between B^+ and B^-, and ratios R

Results agree with SM and will improve γ measurement in due course

Table 2: Systematic uncertainties for the K^{*+} region fit. Uncertainties are quoted as a percentage of the statistical uncertainty for a given observable, and the total uncertainty is given by the sum in quadrature of each contribution.

Observable	*Eff*	*PDF*	*Cls*	*Asym*	*PID*	Total
$A_{\rm SS}^{D\pi}$	0.0	0.5	0.4	25.6	0.8	25.6
$A_{\rm OS}^{D\pi}$	0.0	0.4	0.7	16.9	0.9	16.9
$A_{\rm SS}^{DK}$	0.0	1.7	10.1	11.9	6.3	16.9
$A_{\rm OS}^{DK}$	0.0	0.3	16.7	1.3	5.5	17.7
$R_{\rm SS/OS}$	33.6	0.5	0.2	0.1	0.5	33.6
$R_{\rm SS}^{DK/D\pi}$	29.2	3.2	31.3	0.1	8.1	43.7
$R_{\rm OS}^{DK/D\pi}$	15.5	2.7	40.9	0.1	4.9	44.1

Table 3: Systematic uncertainties for the non-K^{*+} region fit. Uncertainties are quoted as a percentage of the statistical uncertainty for a given observable, and the total uncertainty is given by the sum in quadrature of each contribution.

Observable	*Eff*	*PDF*	*Cls*	*Asym*	*PID*	Total
$A_{\rm SS}^{D\pi}$	0.0	0.4	0.6	14.3	1.0	14.4
$A_{\rm OS}^{D\pi}$	0.0	0.7	0.5	18.4	1.7	18.5
$A_{\rm SS}^{DK}$	0.1	0.5	17.8	7.1	5.8	20.0
$A_{\rm OS}^{DK}$	0.0	1.5	10.9	1.3	9.4	14.5
$R_{\rm SS/OS}$	48.6	0.6	0.5	0.1	0.4	48.6
$R_{\rm SS}^{DK/D\pi}$	14.8	3.0	44.4	0.1	4.4	47.1
$R_{\rm OS}^{DK/D\pi}$	18.6	4.0	32.7	0.1	7.3	38.5

Dominated by different sources for different observables

Efficiency corrections, Fixed shape parameters, Charmless backgrounds, $B^{\pm}$ etc asymmetry corrections, PID corrections

Avoid systematic from D decay model by using data (from CLEO-c).

Summary

Wow!

Very impressive quantity and quality of work in these analyses

Taxonomy holds up

Categorisation into Experimental and Modelling (=theory?) uncertainties also useful

Common Experimental Uncertainties well studied

Jet Energy Scale and Resolution, Trigger Efficiency, *b*-tagging, lepton and hadron ID, Luminosity

Common Theoretical Uncertainties well studied

QCD scales, α_s, pdf parametrisations

When is it correct to take the envelope?

Very seldom

- If you're an engineer. Then you quote tolerances. If you have a 99 ± 1 mm peg and a 101 ± 1 hole you want it to fit *every* time.
- If you define 'error' as '68% central confidence region' as opposed to 'rms spread'[1] and an ultrafrequentist approach to '68% confidence means θ lies in the region at least 68% of the time', i.e. for absolutely all values of ν, rather than profiling or even marginalising.
  Note that if you take this route you can no longer add in quadrature

We must not be afraid of quoting a result that may be more than 1 sigma from the true value.

[1] For Gaussian distributions it's the same

Other issues

QCD scale factors

These have big ‘uncertainty’. - factor of 2 either way - but the nature of that uncertainty is subtle

Bias in fitting

Need to think hard about whether a bias is expected (from statistical properties of estimator) or whether this is a check for some hidden mistake

Should one vary the binning?

As a check. Yes.
As an estimate of uncertainty? Questionable.

Bayesian analysis: choice of prior

There is no unique correct prior Different priors give different answers
Statisticians check for ‘robustness under choice of prior’. Physicists lazy?
Not sure if this is a ‘systematic error’, but should be reported somewhere

Conclusions

Principle confronts practice

The taxonomy holds up,
but there are new improved ways of estimating contributions to the error

- R_{BDT} where properties are used in BDT training
- Sampling from Gaussian rather than just taking $\pm\sigma$
- Bootstrap to check fit accuracy is what it says it is, and to get errors from finite samples
- Treating model choice as 'nuisance parameter'

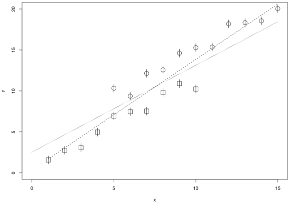
But we have work to do

There are still many 'conservative' (i.e. too large) systematic errors quoted
The statistics expert community (this means you) should take a firm line

Backup: combining experiments

arXiv: 1701.13701v2

Fitting functions to data from several experiments
Should you
(1) minimise χ^2 using inverse correlation matrix
(2) introduce extra parameters and just use the diagonal terms?



Answer (after some algebra): it doesn't matter.
The two methods are equivalent.
This applies both to additive and multiplicative errors
(2) is usually easier and more informative, and $N_p + N_e$ variable minimisation splits into $N_p \times N_e$

Multiplicative errors lead to bias if applied to the data but not if applied to the function