Monte Carlo and Measurements

Andrea Knue
Albert-Ludwigs Universität Freiburg

– MC-Net school 2018 –

What is this lecture about?

“Issues related to modelling, interpretation and unfolding of the real world”

What is this lecture about?

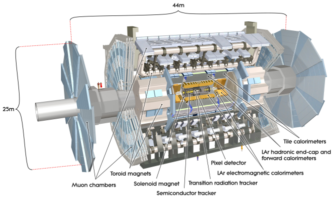
“Issues related to modelling, interpretation and unfolding of the real world”

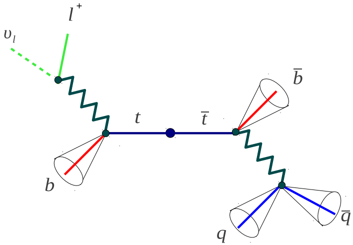
Disclaimer

Heavy bias towards examples from ATLAS and from the top world... Sorry ;)

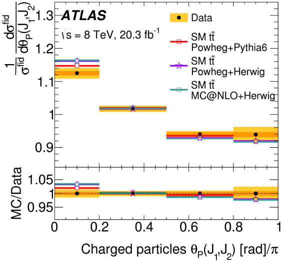
Our job in a (very) simplified way

– HEP Experiment – + **– Process of interest –**





↓



= fancy plot?

http://arxiv.org/abs/1506.05629

Not that simple!

Short Brainstorming:

1. What information do we have?
2. What information are we interested in?
3. What do we want to measure and in what way?

What do you need for data analysis?

A lot of important steps before extracting the final quantity:

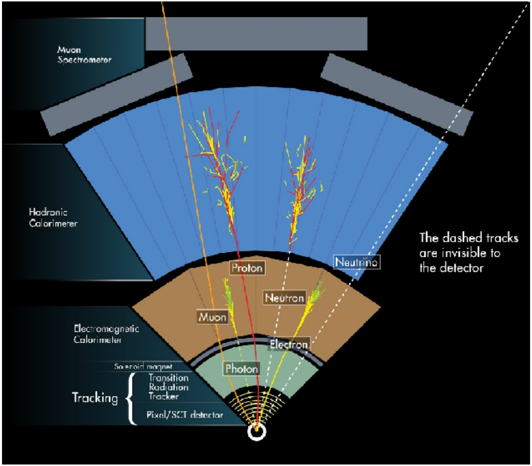
- experiment $\rightarrow$ data taking!
- detector calibration
- event selection
- but also: produce simulation (Monte Carlo generators) to compare with!
- run all steps over data and MC
- compare expectation with what data shows us

How do we get from the matrix-element to the measured object?

Sherpa,
Powheg,
MadGraph,
aMC@NLO,
Pythia,
Herwig,
...

- generate event up to the parton shower, then simulate the full detector
- in ATLAS: full detector simulation takes about 15 min per event!
- very CPU/storage intensive, would be easier if we only needed step 1!

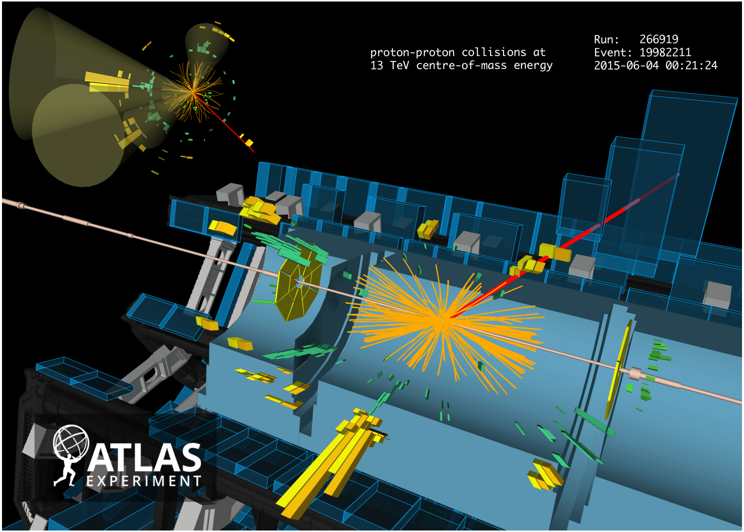
What information can we obtain?



We are interested in:

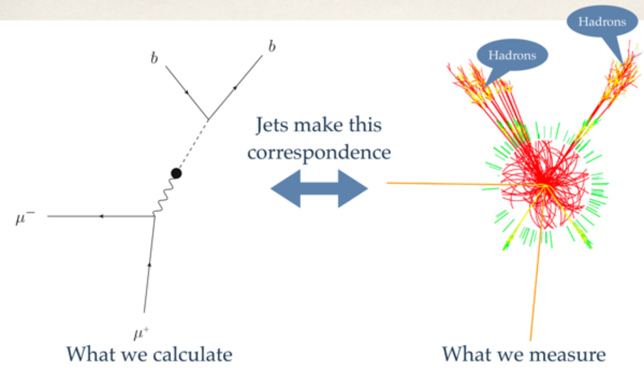
- jets/b-jets
- hadronic tau leptons
- muons/electrons
- photons
- missing E_{T}

Event display



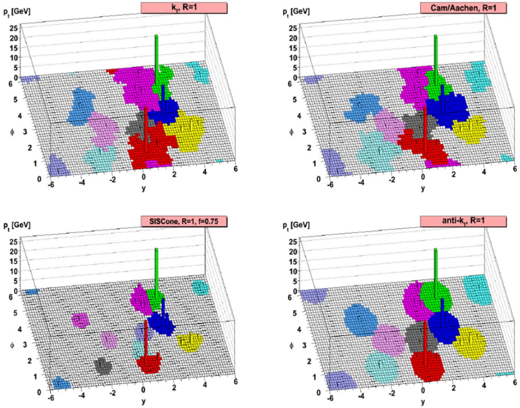
What is a jet and how to define it?

Figure stolen from Quantum-diaries



- cannot observed coloured quarks in the detector, but color-neutral hadrons
- jets are seen as spray of hadrons, but: what hadron belongs to which jet?

How to define jets? ▸ Link

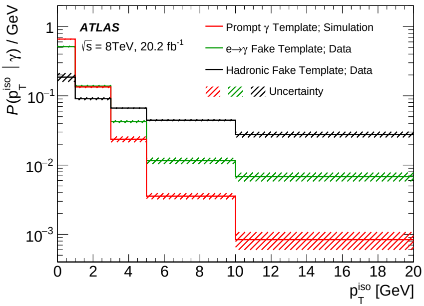


- want jet reconstruction algorithms to be infrared and collinear safe
- to resolve jets: anti-k_t algorithm is best suited (and is widely used)
- jet substructure: also Cambridge/Aachen very well suited (e.g. CMS)

Photon definition ▸ TOPQ-2015-21

- photons radiated off charged particles: mainly soft

  $\hookrightarrow$ should they be considered in the lepton definition?
- photons from π^0 decays: distinguish from photons discussed above
- at detector-level: use isolation requirements (see below)



How to define a lepton?

Detector-level: track in the inner detector, matched to either an energy cluster in the EM calorimeter (electron) or another track in the MS (muons)

Usually distinguish three definitions:

1. born leptons: before QED final state radiation

   $\hookrightarrow$ can compare to calculations where FSR is not accounted for, but larger systematics

2. bare leptons: after QED final state radiation

   $\hookrightarrow$ problem: model dependent (cut-off...), and different for electrons and muons

3. dressed leptons: include all photons around a lepton with $\Delta R < 0.1$

   $\hookrightarrow$ generator independent, allows to combine electrons and muons

   $\hookrightarrow$ cannot be directly compared to fixed-order calculation

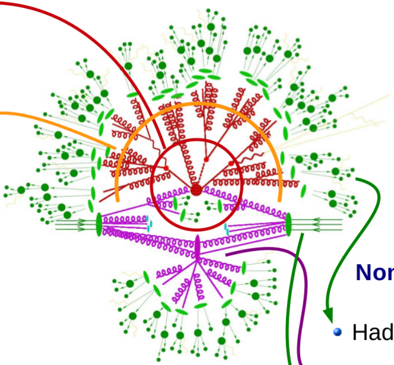
Monte Carlo generators

Perturbative QCD

- **Hard scattering**
  - Fixed Order (Powheg, aMC@NLO, etc...)
- Fragmentation
  - Parton Shower
    - Initial state
    - Final state

Factorization

$$\sigma_{p\bar{p}\to X} = \Sigma_{i,j} \int dx_1\, dx_2\, f_i^p(x_1,\mu)\, f_j^p(x_2,\mu) \times \sigma_{i,j}$$

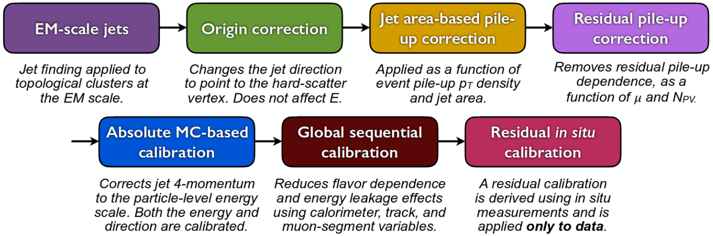


Non perturbative QCD

- Hadronization
- MPI
- Primordial k_T

Slide stolen from Stefano Camarda LINK

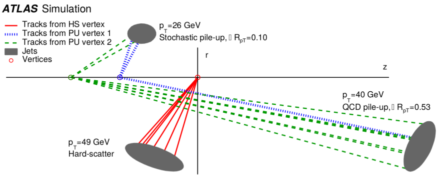
Jet calibration: use of MC information



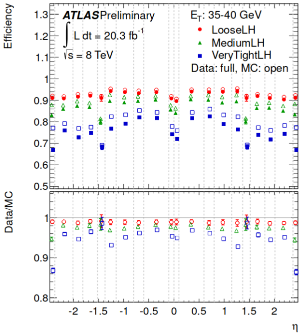
- origin correction: relies on MC and uses also truth jets
- pileup correction: has residual correction based on MC
- absolute JES calibration: dijet MC, uses also truth jets
- in-situ calibration: also relies on in-situ calibration in MC

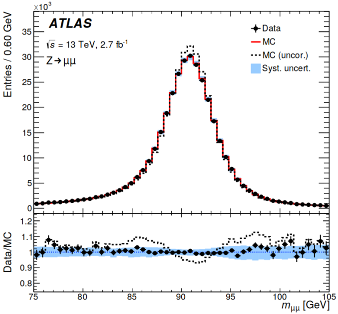
Summary: MC plays an important part in many of these stages:
$\hookrightarrow$ therefore also in the corresponding systematic uncertainty!

Where else is MC info used?

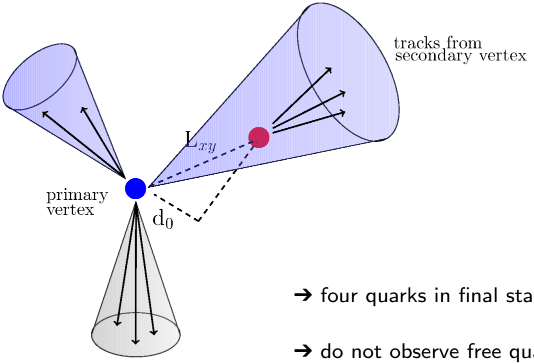


- overlay all “physics” MC simulation with corresponding pileup profile



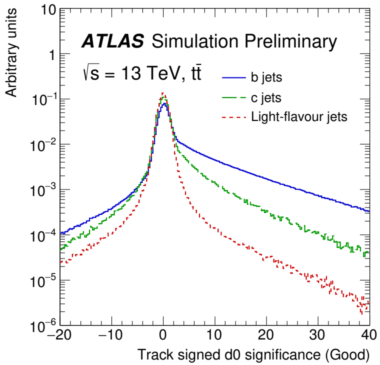


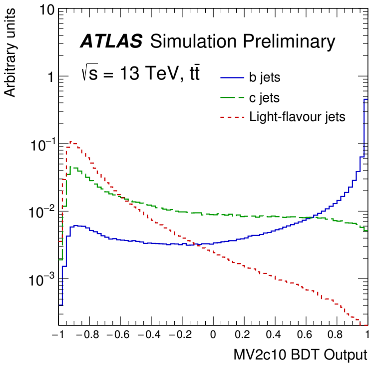
Identify jets from *b*-hadrons



- ➔ four quarks in final state
- ➔ do not observe free quarks: hadronisation to jets
- ➔ lifetime from *b*-jets: $< \tau_b >= 1.6$ ps
- ➔ identified by secondary vertex: ***b*-tag**

Impact parameter ATL-PHYS-PUB-2016-012





➔ different basic tagging algorithms are combined in Boosted Decision Tree

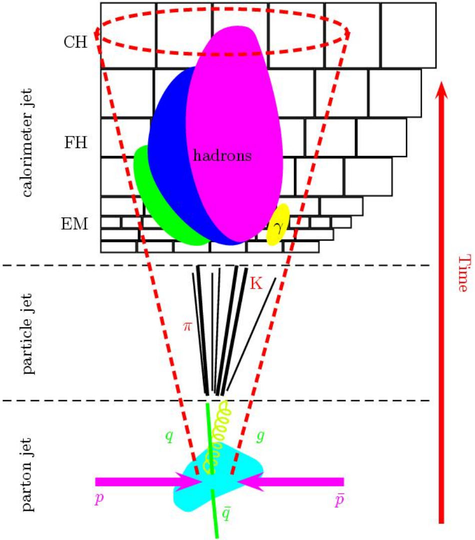
➔ for cut on BDT value: get *b*-tagging efficiency and light-/*c*-tag rejection

What kind of information do we have?

detector level or reconstruction level: after full detector simulation

stable particle level:
"stable" $\tau_{\mathrm{part.}} > 30\mathrm{ps}$

parton level



How far can/may we rely on this truth information?

Stable particle level

- using dressed leptons: generator/model independent
- used e.g. also for jet calibration: match the particle-level jet to a reconstructed jet

Parton level

- allows to compare with fixed-order calculations as for example done in MCFM
- try to remove hadronisation and MPI effects, very model dependent!
- some analyses use these spectra to reweight the MC... always dangerous

Should we use data or MC for modelling of physics backgrounds?

Might need to use data-driven estimates instead if:

- MC has large uncertainties (e.g. $tt + bb$ production as bkg to ttH)
- if data is not well described by MC
- sometimes use some kind of "hybrid" estimate:

  $\hookrightarrow$ rely on MC for background shape, but estimate normalisation from data

Many ways to estimate the background

$\hookrightarrow$ matrix-method, ABCD, template fit, fake factor, sideband fits, ...

$\hookrightarrow$ will only be able to discuss some of them here

$\hookrightarrow$ define control region for estimate (signal depleted!), define validation region

$\hookrightarrow$ how kinematically similar is CR to SR (is extrapolation safe?)

Data driven estimates: Misidentified leptons

QCD multijet production: high cross-section, difficult to simulate

Electrons:

- photon conversion
- neutral pions
- jets misidentified als electrons

Muons:

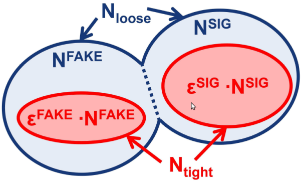
- can originate from heavy flavour jets: fake isolation
- pion and kaon decays

Different methods to estimate the contribution from data:

- talk today about matrix method (define loose and tight samples)

➔ perform estimate in **control region** (could for example invert E_T^{miss} cuts)

Data driven estimates: Misidentified leptons



$$N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}} \quad .$$

$$N^{\text{tight}} = N^{\text{tight}}_{\text{real}} + N^{\text{tight}}_{\text{fake}} \quad .$$

Calculate number of tight leptons:

$$N^{\text{tight}} = \epsilon_{\text{real}} N^{\text{loose}}_{\text{real}} + \epsilon_{\text{fake}} N^{\text{loose}}_{\text{fake}} \quad ,$$

with efficiencies:

$$\epsilon_{\text{real}} = \frac{N^{\text{tight}}_{\text{real}}}{N^{\text{loose}}_{\text{real}}} \quad , \quad \epsilon_{\text{fake}} = \frac{N^{\text{tight}}_{\text{fake}}}{N^{\text{loose}}_{\text{fake}}} \quad .$$

Data driven estimates: Matrix Method

The rate of fake leptons that survive the tighter cuts can now be derived as:

$$N^{\mathrm{tight}}_{\mathrm{fake}} = \frac{\epsilon_{\mathrm{fake}}}{\epsilon_{\mathrm{real}} - \epsilon_{\mathrm{fake}}}(N^{\mathrm{loose}}\epsilon_{\mathrm{real}} - N^{\mathrm{tight}}) \quad .$$

$$w_{\mathrm{loose}} = \frac{\epsilon_{\mathrm{fake}}\epsilon_{\mathrm{real}}}{\epsilon_{\mathrm{real}} - \epsilon_{\mathrm{fake}}} \quad ,$$

while the event weight for the tight selection is calculated as:

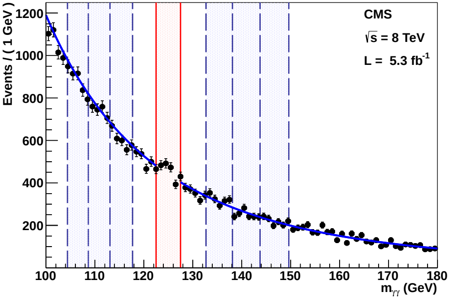
$$w_{\mathrm{tight}} = \frac{\epsilon_{\mathrm{fake}}(\epsilon_{\mathrm{real}} - 1)}{\epsilon_{\mathrm{real}} - \epsilon_{\mathrm{fake}}} \quad .$$

Careful: tight events have always negative weights here!

$\hookrightarrow$ if your loose/tight definition is too similar, you can end up with overall negative event yield, or at least with some negative bin entries!

Background from sidebands

J. High Energy Phys. 06 (2013) 081



- mass distribution with signal hypothesis in certain window
- perform likelihood fit in sideband regions
- interpolate this to signal region
- careful: are the sidebands really signal-free?

  $\hookrightarrow$ otherwise problem much more complex (correlation etc)

Advantages/disadvantages of using DD estimates

Data-driven estimate:

- allows to reduce uncertainty on background prediction
- does not depend on MC mismodellings
- but: careful with signal contamination, low statistics, extrapolation!

MC based estimate:

- often larger systematic uncertainties
- sometimes does/cannot model data well
- but: less problem with statistics

Differential cross-section measurements

- want to test predictions in different parts of phase space
- unfold to parton-level and stable particle level
- allows to compare results from different experiments
- make fiducial measurements

  ↪ allows to test different models

  ↪ helps to constrain systematic uncertainties

Distributions unfolded to:

➔ parton level: top after radiation, but before decay

➔ particle level: stable leptons and jets clustered from stable particles

Why should we make model-independent measurements?

When designing an analysis to test only one specific model:

$\hookrightarrow$ measurement has very limited lifetime and information

$\hookrightarrow$ cannot be reused in a few years, when other models come on the market!

You will never be able to verify a certain model and to exclude all others:

$\hookrightarrow$ **"All models are wrong, but some are useful" (George Box)**

Effective Field Theory

➔ There are loads and loads of theories on the market to solve the mentioned problems.

$\hookrightarrow$ but: a lot of work to test them all

$\hookrightarrow$ the new physics processes could occur at higher energies than previously studied

➔ one approach: model independent search

➔ new physics could show up in higher order terms to Lagrangian

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i$$

with f_i being Wilson coefficients and Λ being the new physics scale.

Effective Field Theory

- large number of dimension 6 and 8 operators
- can for example look for anomalous couplings
- would change production rate or show up in deviations of differential distributions
- for good limits on coefficients

  $\hookrightarrow$ need many and precise measurements
- papers: ▸ Phys.Lett.B 759 (2016) 672

Why doing fiducial measurements? ▸ Link, ▸ Link

Fiducial volume:

➔ define cross-section based on experimental selection and acceptance cuts

- less model-dependent, since extrapolation to full phase space not needed
- smaller uncertainties since acceptance corrections not needed
- best: based on simple kinematic selection cuts: easier to be re-used later
- MVA used in selection:

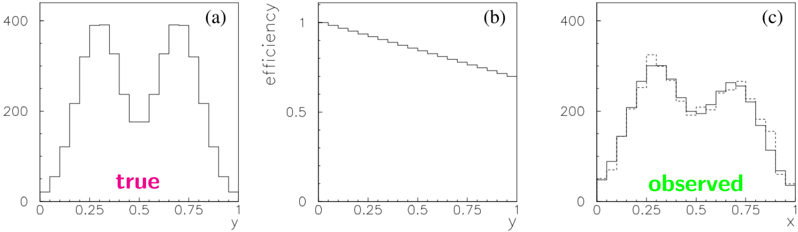
  ↪ might have smaller uncertainties, but makes re-use much more difficult

$$\sigma_i^{\mathrm{fid}} = \Sigma_j A_{ij}^{\mathrm{th}} \sigma_j^{\mathrm{tot}}$$

What exactly is unfolding?

Problem:

- want to measure a spectrum (mass, transverse momentum, rapidity...)
- detector has only finite resolution and cannot perfectly measure it
  $\hookrightarrow$ distribution will be distorted (different width, shifted)



$\boldsymbol{\mu} = (\mu_1, ..., \mu_M)$ $\boldsymbol{\nu} = (\nu_1, ..., \nu_N)$

$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j \quad i = 1, ..., N$$

Lets look into this in more detail...

R is the response matrix:

$$R_{ij} = \text{Prob(observed in bin i} \mid \text{true value in bin j)}$$

Need efficiency ϵ (middle plot):

$$\epsilon_j = \sum_{i=1}^{N} R_{ij} = \text{Prob(observed anywhere} \mid \text{true value in bin j)}$$

Now put together the full equation:

➔ with data $\boldsymbol{n} = (n_1, ..., n_N)$

➔ and background events β

$$E[\boldsymbol{n}] = \boldsymbol{\nu} = R\boldsymbol{\mu} + \beta$$

How does this help us now?

Idea:

- correct this distribution, to remove the detector effects!
- for this correction: rely on the MC "truth" information

Now get the unfolded distribution μ from:

$$\boldsymbol{\mu} = R^{-1}(\boldsymbol{\nu} - \boldsymbol{\beta}) \tag{1}$$

And the ML estimator:

$$\hat{\boldsymbol{\mu}} = R^{-1}(\boldsymbol{n} - \boldsymbol{\beta}) \tag{2}$$

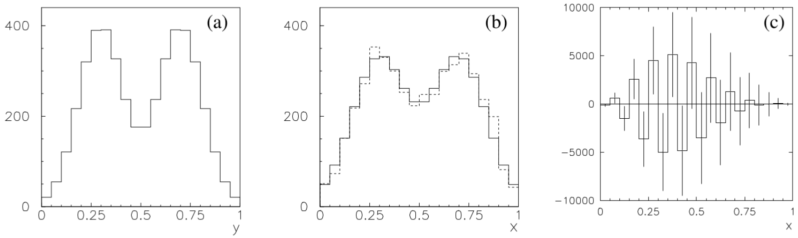
$\hookrightarrow$ with data $\boldsymbol{n}$ being Poisson distributed

Is it really that straightforward?

Problem:

- if matrix R has larger off-diagonal elements or large fluctuations in data:

  $\hookrightarrow$ $\hat{\mu}$ can have large variances (see right plot)



Unfolding methods

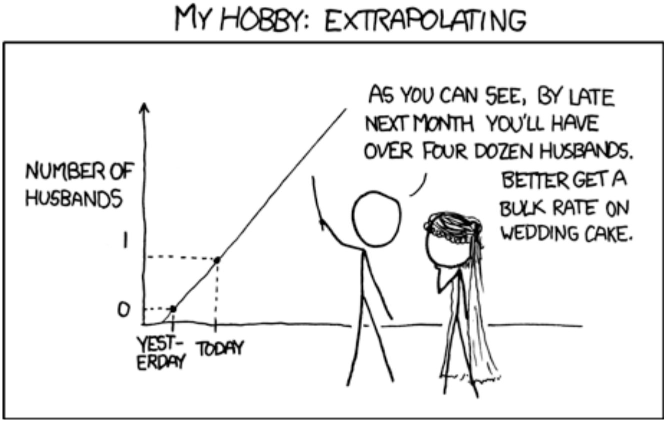
Some examples:

- bin-by-bin unfolding
- Iterative Bayesian Unfolding (IBU, D'Agostini)
- Full Bayesian Unfolding (FBU) ▸ arXiv:1201.4612
- Singular value decomposition (SVD)

➔ check that method is unbiased: do proper stress-tests!

➔ careful not to over-regularise

Extrapolation into different phase spaces



Why do we need to be careful with extrapolation of results?

Recent example: Performance of Lionel Messi

messivsronaldo.net

in spanish league:

Season	goals per game
2013/14	0.90
2014/15	1.13
2015/16	0.79
2016/17	1.09
2017/18	0.94

➔ had always $\geq$ 0.79 goals per game

Extrapolation to 4 WC matches 2018:

➔ at least 3 goals, maybe more!

Why do we need to be careful with extrapolation of results?

Recent example: Performance of Lionel Messi ▸ messivsronaldo.net

in spanish league:

Season	goals per game
2013/14	0.90
2014/15	1.13
2015/16	0.79
2016/17	1.09
2017/18	0.94

➔ had always $\geq$ 0.79 goals per game

Extrapolation to 4 WC matches 2018:

➔ at least 3 goals, maybe more!

BUT: World cup is a very different phase space...

$\hookrightarrow$ actual performance: 1 goal $\rightarrow$ 0.25 goals per game

Physics example for extrapolation

Single top uncertainties

➔ interference between $t\bar{t}$ and Wt production:

➔ try to remove interference part from Wt process using two methods: "diagram removal" and "diagram subtraction"

$\hookrightarrow$ difference between two schemes used as systematic uncertainty

Look at impact in measurements and searches:

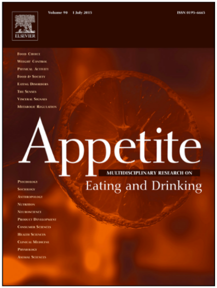
1. $t\bar{t}$ inclusive cross-section: impact $\approx$ 1-2 %, can almost be ignored?
2. SUSY searches however: up to 50% difference in normalisation!

Same applies to the everyday life

- People Who Order Coffee Black Are More Likely To Be Psychopaths

The research, which comes from the University of Innsbruck in Austria, examined the taste preferences of about 1,000 people, finding those who preferred more bitter notes in foods like black coffee scored higher on a series personality questionnaires that assessed Machiavellianism, a term used in psychology to describe personalities that are dark, psychopathic, narcissistic and sadistic.

The study also notes that participants who reported a fondness for other bitter foods like radishes, celery and tonic water were also more likely to exhibit psychopathic traits.



Christina Sagioglou, Tobias Greitemeyer

PII: S0195-6663(15)30042-8

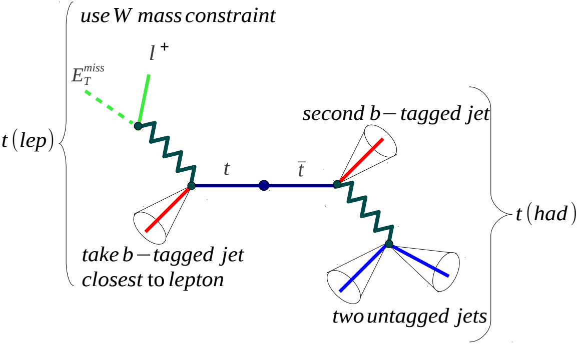
DOI: 10.1016/j.appet.2015.09.031

Frédéric Déliot, Top2016, 23-SEP-16

➔ **Should you really extrapolate this to the person next to you... ?**

Example: measurement of $t\bar{t}$ cross-section

- want to fully reconstruct $t\bar{t}$ event from final state objects
- run same algorithm on particle and detector level



➔ Unfolding to particle level: generator independent studies possible

Object definitions

"Particle-level jets are clustered using the anti-k$_t$ algorithm with radius parameter R = 0.4 or R = 1.0, starting from all stable particles, except for selected leptons (e, μ) and their radiated photons, as well as neutrinos."

"The four-momenta of leptons are modified by adding the four-momenta of all photons within ΔR=0.1 and not originating from hadron decays, to take into account final-state photon radiation"

$\rightarrow$ make your selection as close as possible to selection at detector level!

Unfolding procedure

- combination of the electron and muon channel before unfolding

  $\hookrightarrow$ have same definition of fiducial region and similar efficiencies/corrections and migration matrices

- method: Iterative Bayesian method (RooUnfold, $N_{\mathrm{iter.}} = 4$)

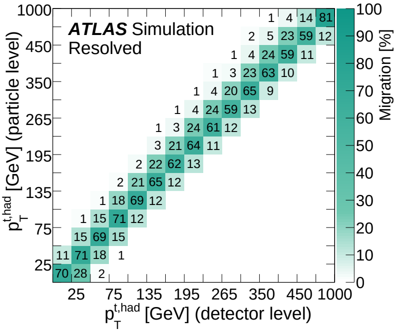
Efficiency and acceptance corrections:

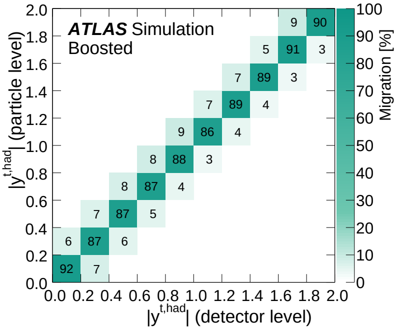
$$f_{\mathrm{eff}}^{i} = \left(\frac{N_{\mathrm{part}}}{N_{\mathrm{reco}\wedge\mathrm{part}}}\right)^{i} \quad f_{\mathrm{acc}}^{j} = \left(\frac{N_{\mathrm{reco}\wedge\mathrm{part}}}{N_{\mathrm{reco}}}\right)^{j}$$

Fiducial cross-section:

$$\frac{d\sigma^{\mathrm{fid}}}{dX^{i}} = \frac{1}{\mathcal{L}\Delta X^{i}} f_{\mathrm{eff}}^{i} \sum_{j} \mathcal{M}_{ij}^{-1} f_{\mathrm{acc}}^{j} f_{\mathrm{match}}^{j} (N_{\mathrm{reco}}^{j} - N_{\mathrm{bkg}}^{j})$$

Response matrices



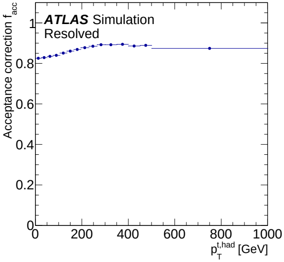


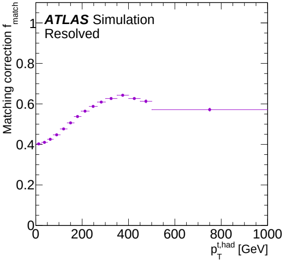
↪ elements in row add up to unity

↪ choose binning such that the diagonal entries are always > 50%

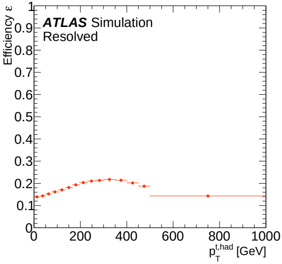
Additional correction neccessary for unfolding

generated outside fiducial phase-space, pass the detector-level selection



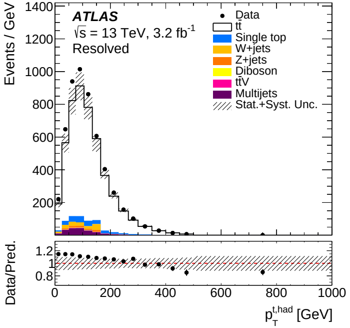


correction such that detector- and particle-level objects forming the pseudo-top quarks are well matched

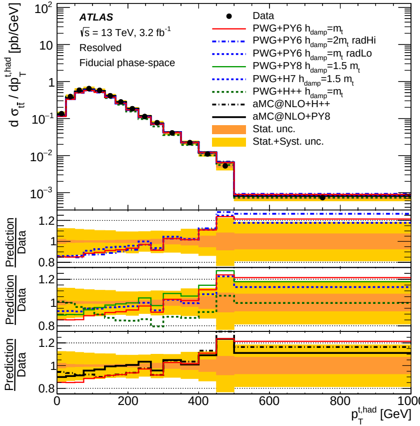


pass the particle-level selection, but not reconstructed at detector level

Detector-level vs. unfolded distribution



➔

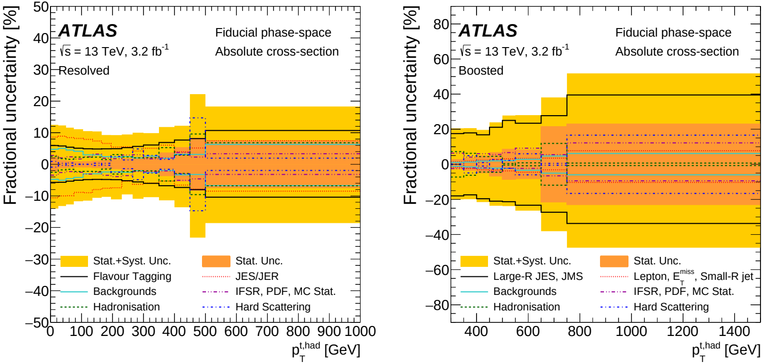


See same trend here almost everywhere: what does it mean?

↪ is there some systematic uncertainty that we are missing?

↪ is something missing in the MC event generation?

Relative cross-section uncertainties resolved/boosted



- resolved topology:

  ↪ already dominated by modelling unc. (will discuss later which ones)

- boosted topology:

  ↪ still large experimental unc., but do not cover observed difference

- here: NLO+PS (LO), what happens at NNLO or with EW corrections?

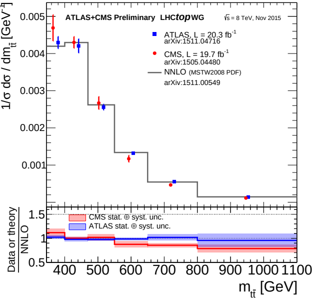
Comparison with NNLO predictions @ 8 TeV

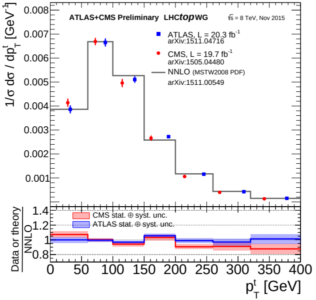
➔ compare now unfolded ATLAS and CMS data with NNLO predictions

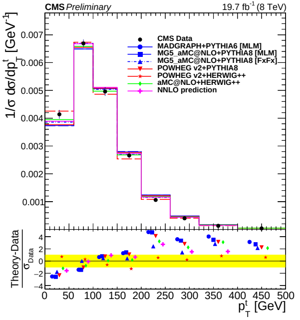
➔ $m_{t\bar{t}}$ and top p_{T} distributions (unfolded to **parton level**):

↪ NNLO prediction harder than CMS data, good agreement for ATLAS

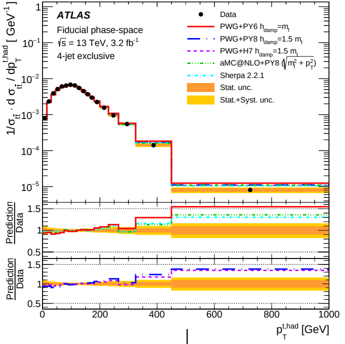
➔ large differences between generators predictions ▸ CMS-TOP-15-011



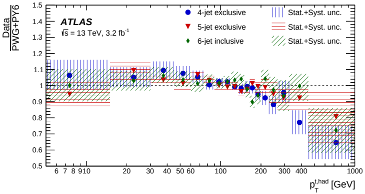


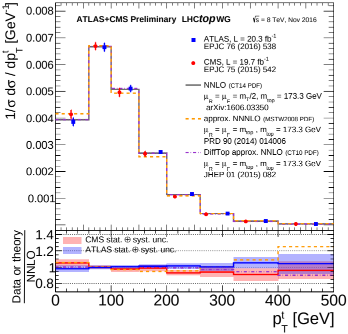


Go even more differential: Predictions as function of N_{jets} ▸ arXiv:1802.06572



↓

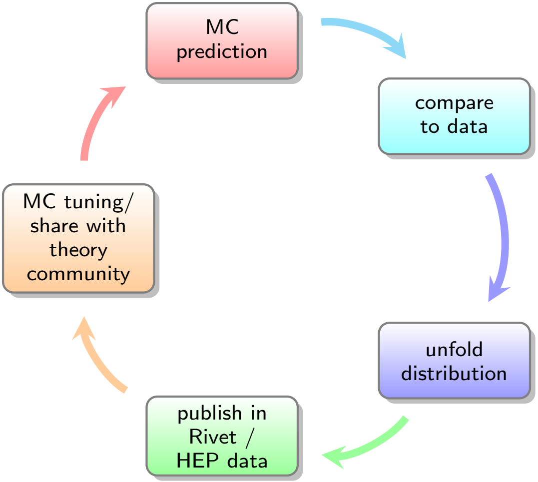




→ disagreement larger for exactly four jets

→ but agreement with NNLO much better

What do we do now with these differential measurements?



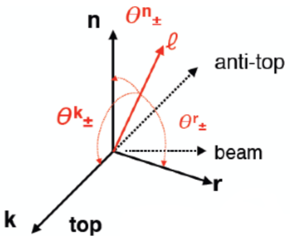
➔ important: have knowledge update here!

Measurement of top spin observables in the dilepton channel ATLAS-CONF-2018-027

- tops decay before hadronisation: spin information transferred to top decay products
- top quarks are produced unpolarised, but spins of tops are correlated:

$$\frac{1}{\sigma}\frac{d^2\sigma}{d\cos\theta_1 d\cos\theta_2} = \frac{1}{4}(1 + \alpha_1 P_1 \cos\theta_1 + \alpha_2 P_2 \cos\theta_2 - \underbrace{\alpha_1\alpha_2 A}_{\text{Correlation}} \cos\theta_1 \cos\theta_2)$$

Define quantisation axes:



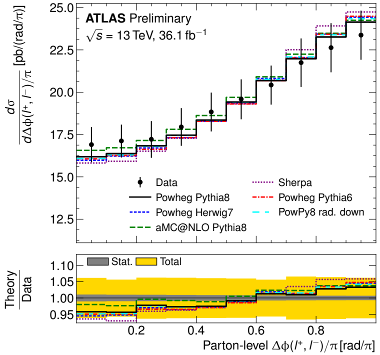
Analysis strategy:

- use angular information of charged leptons
- unfold to parton and stable particle level
- no event reconstruction neccessary

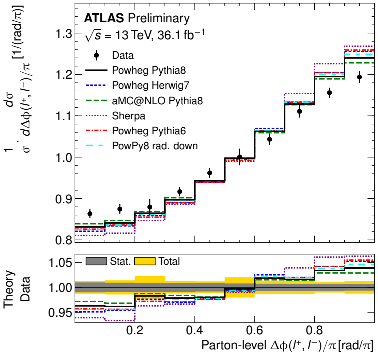
Example: spin correlation measurement (inclusive)

ATLAS-CONF-2018-027

absolute



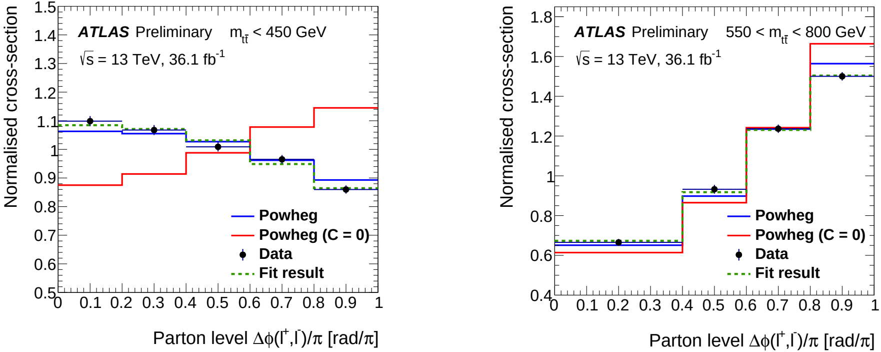
normalised



- see more spin correlation than expected in the Standard Model
- how large is the descrepancy?

Example: spin correlation measurement (differential)

▸ ATLAS-CONF-2018-027



- look at the effect as a function of the $m_{t\bar{t}}$ system
- effect seems to be different for different $m_{t\bar{t}}$ regions: how significant?

Example: spin correlation measurement: how significant? ATLAS-CONF-2018-027

Region	f_{SM}	Significance (incl. theory uncertainties)
$m_{t\bar{t}} < 450$ GeV	$1.11 \pm 0.04 \pm 0.13$	0.85 (0.84)
$450 < m_{t\bar{t}} < 550$ GeV	$1.17 \pm 0.09 \pm 0.14$	1.00 (0.91)
$550 < m_{t\bar{t}} < 800$ GeV	$1.60 \pm 0.24 \pm 0.35$	1.43 (1.37)
$m_{t\bar{t}} > 800$ GeV	$2.2 \pm 1.8 \pm 2.3$	0.41 (0.40)
inclusive	$1.250 \pm 0.026 \pm 0.063$	3.70 (3.20)

- effect has been visible since Run I:

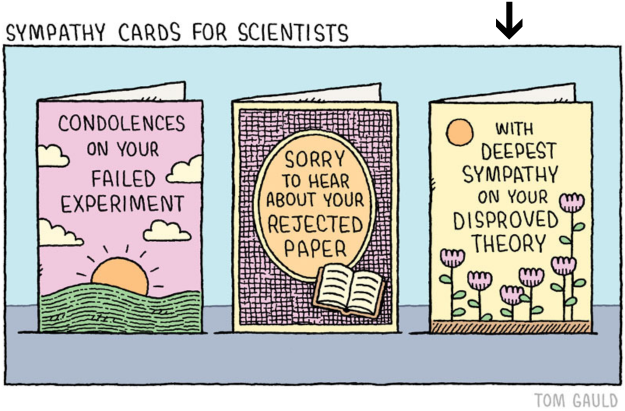
  $\hookrightarrow$ significant now due to large dataset available
- What could cause this?

  $\hookrightarrow$ New physics, or something missing in our MC events?
- have made this public now for ICHEP:

  $\hookrightarrow$ hope to get feedback from the HEP community!

Search for new physics: Is there a better model?

?



New particle found??? Not quite clear...

- as shown in Higgs discovery slides:

  $\hookrightarrow H \rightarrow \gamma\gamma$ channel has clean signature and good mass resolution

- there are several BSM particles predicted that could decay into two photons

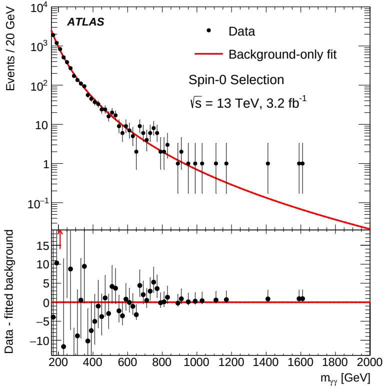
- choose benchmark models to test in data:

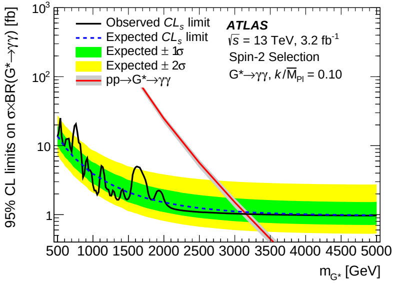
1. particle with spin 2: Randall-Sundrum graviton, would have a very narrow resonance and $m > 500$ GeV

2. particle with spin 0: $m > 200$ GeV, decay products would be isotropically distributed in detector

$H \to \gamma\gamma$ resonances in ATLAS ▸ Paper mid 2016

- ➔ Bkg: as in discovery search
- ➔ measure in CR from data
- ➔ maximum LH fit to $m_{\gamma\gamma}$ spectrum
- ➔ systematics: nuisance parameters





Significance at 750 GeV:

	local	global
spin 2	3.8σ	2.1σ
spin 0	3.9σ	2.1σ

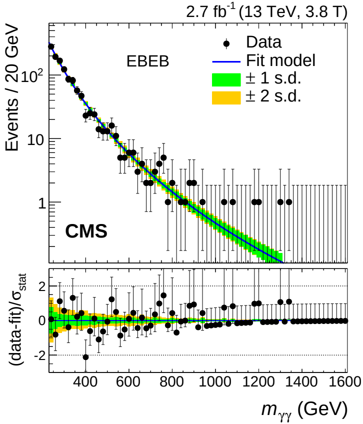
$H \to \gamma\gamma$ resonances in CMS ▸ Paper mid 2016

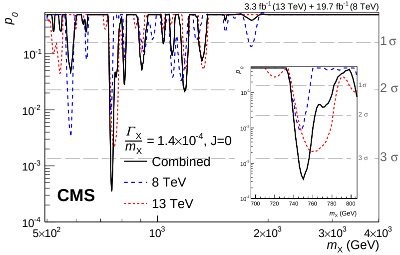
➔ Bkg: as in discovery search

➔ combine with 8 TeV

➔ look for bump in several detector regions

➔ mainly when both photons are in the ECAL barrel





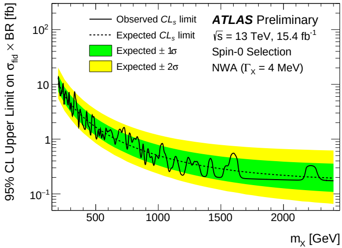
Significance at 750 GeV:

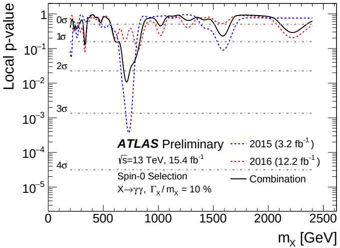
➔ local significance: 3.4σ

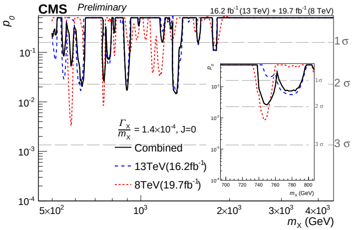
➔ global significance: 1.6σ

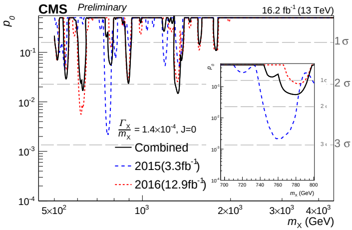
But with more data...

▸ CMS-PAS-EXO-16-027 ▸ ATLAS-CONF-2016-059









➔ no significant excess visible!

Ambulance chasing? ▸ arXiv:1603.01204v1

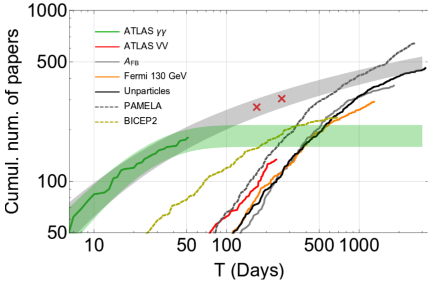
➔ when these results were published

➔ number of follow-up papers published exploded ;)

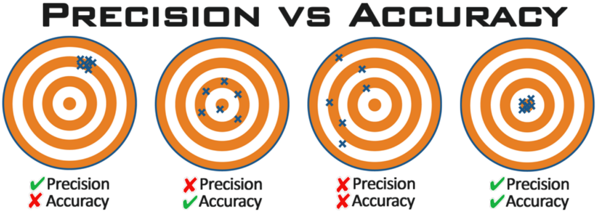
➔ is important that we work close together with theorists

➔ but: also need to be careful, often more data will kill a local excess

➔ exciting, but important to make sure all the background estimates, systematic uncertainties etc are estimated thoroughly



How to establish realistic systematic uncertainties?

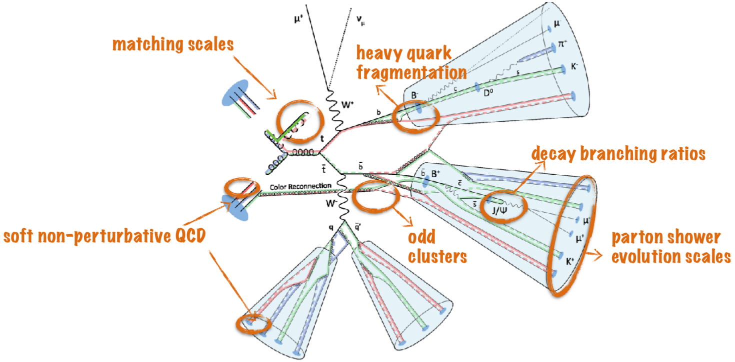


Procedure often used: factorisation approach

- change matrix-element generator (ME)
- change parton shower model (PS)
- vary scales in ME and PS and other parameters (h_{damp})

  $\hookrightarrow$ vary until the systematic samples bracket the total uncertainty

CMS: Modelling uncertainties ▸ LHC Top WG meeting



➔ is there an effect missing that needs implementation in the MCs?

➔ are we double-counting anything?

Quantitative tuning

What we usually get (best case ;)):

➔ Rivet routine (code validated against analysis code)

➔ Data with full uncertainties in form of yoda file

What we need for quantitative analysis:

➔ systematic uncertainties/covariance matrices (store in HEP data!)

➔ calculate χ^2 values

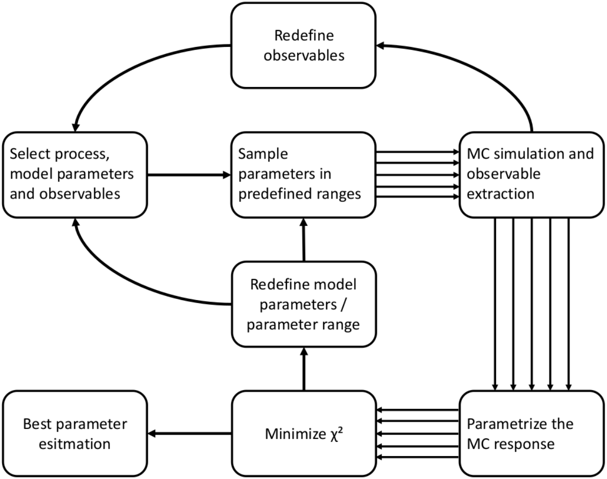
HEPData

High Energy Physics Data Repository

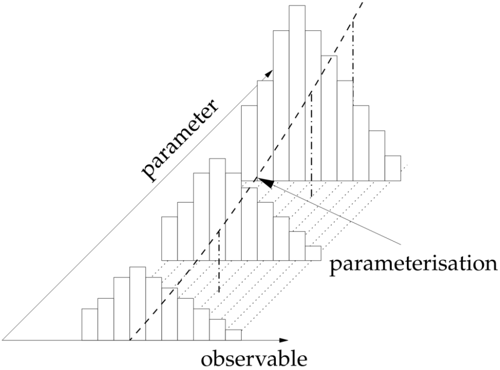
Full tuning using Professor framework

- will not go into too many details here
- want to tune parameters that cannot be obtained from first principles
- define parameters to vary and range in which to vary
- produce samples with different parameter sets
- minimise χ^2 value, can get weights to different distributions and/or bins
- exclude bins where the generator cannot model the data well by definition

Tuning scheme



Parametrisation of observables

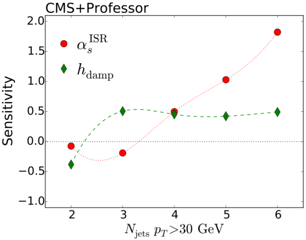


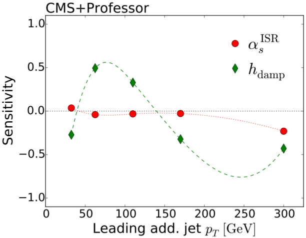
- depending on parameters of interest: can use LEP data, LHC Z+jets, dijets, minbias, etc ...

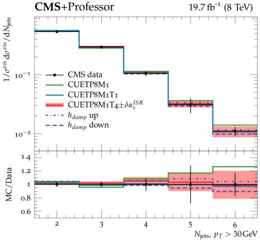
  $\hookrightarrow$ need measurements published in Rivet!

- choose appropriate set of parameters to tune and set of observables

Example: CMS CUETP8M1T4 tune ▸ TOP-16-021-PAS







Tune $h_{\rm damp}$ and α_s(ISR) to $N_{\rm jet}$ and jet $p_{\rm T}$ distributions

- best values[a]: $h_{\rm damp} = 1.581^{+0.658}_{-0.585} \cdot m_{\rm top}$, $\alpha_s^{\rm ISR} = 0.1108^{+0.0145}_{-0.0142}$
  - ↪ in good agreement with ATLAS settings
  - ↪ $h_{\rm damp}$ var. includes the previous default $m_{\rm top}$, compatible with $2 \cdot m_{\rm top}$
  - ↪ α_S variation compatible with QCD scale variation of 0.5,2.0
- much improved data/MC agreement for $N_{\rm jets}$ with new tune

[a]Sensitivity $= \frac{dMC(p)}{dp} \frac{p_C}{MC(p_C)} \longrightarrow MC(p) =$ bin-entry for parameter p

CMS CUETP8M2T4 tune ▸ TOP-16-021-PAS

- with results from CUETP8M1T4:

  $\hookrightarrow$ tune parameters sensitive to MPI and CR (CUETP8M2T4)

- new tune was applied to MG5_aMC@NLO (FxFx) + Pythia8:

  $\hookrightarrow$ gives good data description as well

- no big improvement for aMC@NLO (MLM) and aMC@NLO+Pythia8

- agreement checked with many particle and parton level distributions

From summary of CMS tuning:

Tuning a parameter in the parton shower **can potentially bias particular new physics searches,** in particular if the tune is based on the dataset where the search is performed. [...] In particular, the fact that global event variables such as missing E_T and H_T are not changed by the new tune gives us confidence that searches involving missing E_T will not be biased. However, **one should pay particular attention if the new physics search relies on $N_{\rm jets}$,$p_T(t\bar{t})$ or p_T of leading additional jets in $t\bar{t}$-like processes.**

What will we do in the future?

➔ a lot of discussion in the recent year

➔ both experiments are using the “factorisation approach”, but:

$\hookrightarrow$ changing e.g. the generator can have bigger effects than changing parameters within a single generator

➔ What to do if no generator describes a specific distribution well?

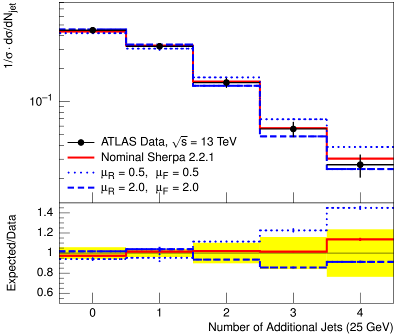
$\hookrightarrow$ e.g. CMS: top p_{T} uncertainty

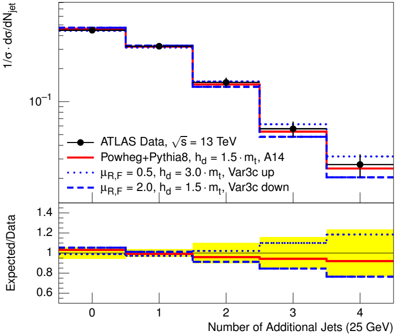
$\hookrightarrow$ have we included all sources of uncertainty?

Two main questions:

1. Should we tune parameters like h_{damp}, and what would be an appropriate uncertainty? (CMS tune: ▸ TOP-16-021-PAS)
2. Or should we take uncertainty variations suggested by theorists, and include them in a profile likelihood fit which will constrain the uncertainties?

Tuning and predictivity





➔ left: Sherpa out of the box, only ME scale variations

➔ right: Powheg+Py8, tuned so that data is described well

↪ scale, h_{damp} and tune variations such that bracket data

➔ be careful not to "overtune": can lead to loss of predictivity

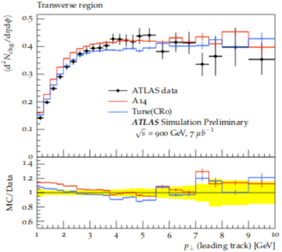
Example tuning: first studies of new CR models in Pythia8 ATL-PHYS-PUB-2017-008

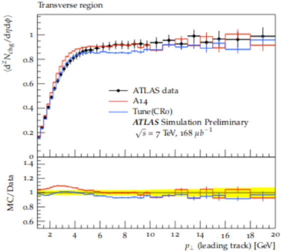
Parameter	Definition
MPI Parameters	
`MultipartonInteractions:pT0Ref`	p_{T} regularisation parameter
`MultipartonInteractions:expPow`	Exponent of matter overlap function
MPI based CR model (CR0)	
`ColourReconnection:range`	CR strength
QCD-based model (CR1)	
`ColourReconnection:m0`	Mass parameter of order λ_{QCD} used in the string length measure
`ColourReconnection:junctionCorrection`	Multiplicative correction to string length above
Gluon-move scheme (CR2)	
`ColourReconnection:m2Lambda`	Equivalent to m0 for QCD-based model
`ColourReconnection:fracGluon`	Average fraction of gluons that undergo a colour reconnection
`ColourReconnection:dLambdaCut`	Minimal value for decrease in string length

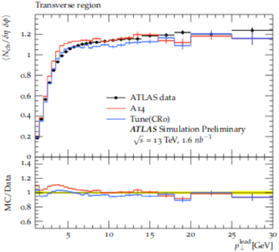
Table 1: Tuning parameters and their definitions. The `MultipartonInteractions:expPow` can only be used with an exponential MPI matter overlap function (`MultipartonInteractions:bProfile = 3`). The parameters specific to a CR model are stated together. CR1 model was used was used with `ColourReconnection:allowDoubleJunRem = off` setting, as recommended by the authors.

- first setup: use A14 set of tuned parameters, only change CR model with recommended settings
- then: make tune for each model, use UE data at all three CME

Tuning for CR0 model ▸ ATL-PHYS-PUB-2017-008







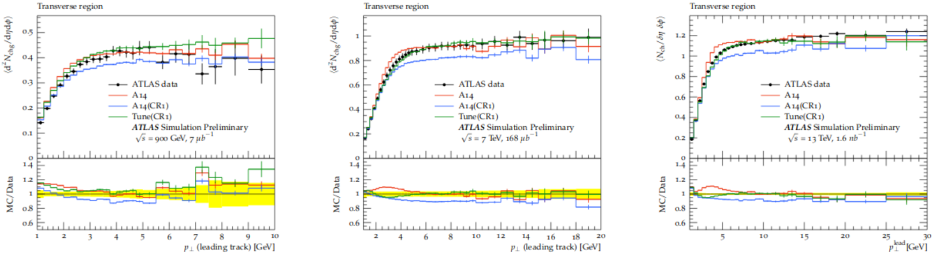
Observation:

- red line: some distributions not well described, but others look already ok
- blue line: data better described after tuning

  $\hookrightarrow$ but: shape differences are still observed after tuning

Tuning for CR1 model

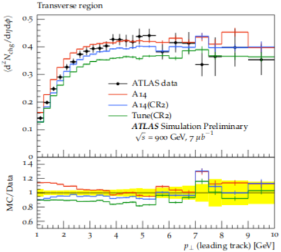
ATL-PHYS-PUB-2017-008

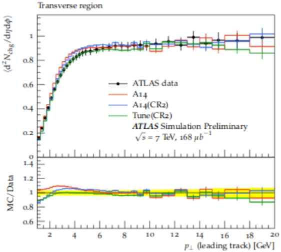


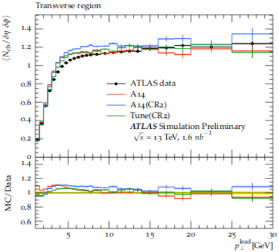
Observation:

- red line: some distributions not well described, but others look already ok
- blue line: see larger differences with CR1 simply put on top of A14 tune
- green line: after tuning, good agreement is restored (similar to previous slide)

Tuning for CR2 model ATL-PHYS-PUB-2017-008







Observation:

- behaves differently than CR0 and CR1 cases
- using just A14 + CR2 model: agreement still looks ok
- but: tuned result looks worse, one of CR parameters going to unphysical ranges

How much do the three results differ? ATL-PHYS-PUB-2017-008

Parameter	A14/ Default (range)	Tune CR0	Tune CR1	Tune CR2
`MultipartonInteractions:pT0Ref`	2.09	2.15	1.89	2.21
`MultipartonInteractions:expPow`	1.85	1.81	2.10	1.63
`ColourReconnection:range`	1.71	2.92	–	–
`ColourReconnection:m0`	0.3 (0.1 - 5)	–	2.17	–
`ColourReconnection:junctionCorrection`	1.20 (0.01 - 10)	–	9.33	–
`ColourReconnection:m2Lambda`	1.0 (0.25-16)	–	–	6.73
`ColourReconnection:fracGluon`	1.0 (0-1)	–	–	0.93
`ColourReconnection:dLambdaCut`	0 (0-10)	–	–	*0.0*
χ^2, $Ndof$		17706, 2929	18597, 2928	113814, 2928
$\chi^2/Ndof$		6.1	6.4	38.9

Table 2: Obtained tuned values for different parameters for each setup is shown, along with the goodness of fit measure. The parameters that are not defined for a particular model are left blank. In the first column (for A14), the recommended values suggested by the authors, with the allowed range is shown. The values in *italics* indicate that they were fixed during the tuning process.

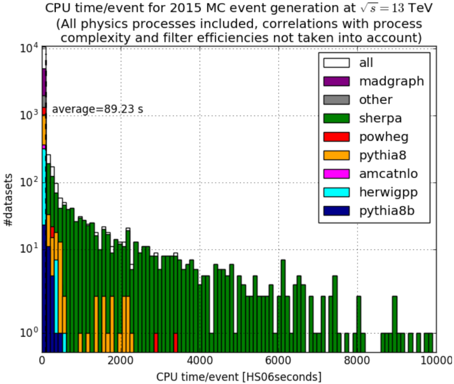
- tuned results for CR0 and CR1 look ok
- tune for CR2 has a much worse goodness-of-fit, parameter at edge of allowed range

CR models: what did we learn?

- most observables are described within 20%, but no model describes all observables well
- CR0 and CR1 do slightly better than CR2
- several models now available: can we exclude any of them in the future?

  $\hookrightarrow$ some of the new models could have large impact on precision measurements such as m_{top}
- need to think about what measurements to perform during the long shutdown, to learn more from the huge amount of data available!

Technical issues we are dealing with



- large CPU time/event for some generators (e.g. Sherpa) ▸ Talk Josh McFayden

  ↪ note: Sherpa usually produces more complex final states

- negative event weights in aMC@NLO: need **up to 9 times more statistics** than for generators without (or little) negative weights (e.g. Powheg)

Negative weights

Numbers stolen from Timothee Theveneaux-Pelzer:

→ N events, negative weights fraction $f = 1/g$: $N_{\text{effective}} = N[\frac{(g-2)}{g}]^2$

- ◇ if $g = 4$, $N_{\text{effective}} = N/4 \Rightarrow$ need to generate $4 \cdot N$ to compensate
- ◇ if $g = 3$, $N_{\text{effective}} = N/9 \Rightarrow$ need to generate $9 \cdot N$ to compensate
- ◇ if $g = 2$, $N_{\text{effective}} = 0 \Rightarrow$ need to generate ∞ to compensate

Simple studies EVNT level: $t\bar{t}$ MG5_aMC@NLO + Pythia 8, 1 lepton, 1 b-jet:

- 4 jets: 24.9 ± 0.07 % negative weights
- 5 jets: 27.5 ± 0.12 % negative weight
- 6 jets: 28.9 ± 0.18 % negative weight

Studied other phase spaces as well:

➔ in phase spaces like the $t\bar{t}H(H \to b\bar{b})$ final state: > 35 %

Summary

- HEP measurements heavily relying on MC simulation

  $\hookrightarrow$ feedback loop with theory/MC community

- should always publish results in a way that its useful later and useful for people outside our experiment

  $\hookrightarrow$ make model-independent measurements!

  $\hookrightarrow$ if possible, publish in Rivet and HEPdata!

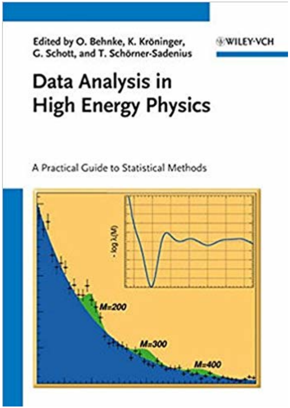
  $\hookrightarrow$ this will increase also the longevity of your result, can be compared with newer models in a few years

- many measurements at the LHC are already dominated by systematic uncertainties:

  $\hookrightarrow$ think: what measurements can be useful to exclude models and reduce systematic uncertainties?

Additional References

- Proceedings Glen Cowan
- Master thesis Fabian Klimpel Link



What to tune to what data?

➔ want to tune the soft QCD part, which we cannot derive from first principles

- use LEP data + LHC data (jet shapes, multiplicities) for hadronisation and fragmentation functions + FSR
- ISR tuned to Zp_T data (also sensitive to N_{jets} and gap fractions in $t\bar{t}$)
- tune MPI and colour reconnection using minimum bias data and UE observables
- b-fragmentation: b-jet shapes, other x_B (measured at LEP but by proxy also possible at LHC)

Colour reconnection models shown in this talk ATL-PHYS-PUB-2017-008

MPI based CR model (CR0)

In this model, the probability for coloured partons with transverse momentum p_{T} from MPI to reconnect is calculated according to:

$$P_{\mathrm{rec}}(p_T) = \frac{(R_{\mathrm{rec}} p_{T0})^2}{R_{\mathrm{rec}} p_{T0} + p_{\mathrm{T}}^2} \qquad (3)$$

where R_{rec} is the ColourReconnection:range and p_{T0} is the MultipartonInteractions:pT0Ref parameter [...]. For each MPI system that undergoes a reconnection, partons from lower p_{T} MPI systems are added to the dipoles defined by the higher p_{T} MPI system, in a way that minimizes the total string length.

Colour reconnection models shown in this talk ‣ ATL-PHYS-PUB-2017-008

QCD based model (CR1)

This model evolved from the MPI based model. The main difference with respect to CR0 and CR2 (described next) is the more complete treatment of the QCD multiplet structure and in particular that reconnections of dipoles can produce structures of three (anti-)colour indices (junctions), thereby enhancing the production of baryons. As in the CR2 model, in this model only reconnections which lower the string length are performed.

Colour reconnection models shown in this talk ▸ ATL-PHYS-PUB-2017-008

Gluon-move model (CR2)

In the gluon move model (CR2), reconnections are performed in the same way as in the CR0 model. The main differences with the default model are that in the CR2 model only gluons are considered for reconnection. For each gluon all the reconnections to all MPI systems are considered (not only the ones for softer MPIs), therefore in principle the colour flow from the hard interaction can be affected more significantly than in the default model.

CMS tuning

From summary of CMS tuning: ▸ TOP-16-021-PAS

Tuning a parameter in the parton shower **can potentially bias particular new physics searches,** in particular if the tune is based on the dataset where the search is performed. [...] In particular, the fact that global event variables such as missing E_T and H_T are not changed by the new tune gives us confidence that searches involving missing E_T will not be biased. However, **one should pay particular attention if the new physics search relies on N_{jets},$p_T(t\bar{t})$ or p_T of leading additional jets in $t\bar{t}$-like processes.**

CMS: Proposal for parton shower modelling uncertainties (Run II)

Source	Handle	Weights	Variation	Comment
Shower scales	ISR scale (SpaceShower:renormMultFac)	no	0.5–2.0	see TOP-16-021
	FSR scale (TimeShower:renormMultFac)	no	0.5–2.0	or $\frac{1}{\sqrt{2}}$, $\sqrt{2}$ from LEP
Matching	h_{damp}	no	$1.581^{+0.658}_{-0.585} \cdot m_{\mathrm{top}}$	see TOP-16-021
Soft QCD	Underlying event (MultipartonInteractions:pT0Ref MultipartonInteractions:expPow MultipartonInteractions:range)	no	up/down	MPI and CR strength (does not affect resonance decays)
Odd clusters	Colour reconnection (MPI-based + QCD inspired + gluon move)	no	different simulations	affects resonance decays
Fragmentation	$x_b = p_{\mathrm{T}}(B)/p_{\mathrm{T}}(bjet)$	yes	Bowler-Lund + Peterson parameter unc. based on LEP fits	see TOP-16-022
Flavour response/ hadronisation	Pythia vs. Herwig	no	JES flavour group for light,g,b,c	
Decay tables	semi-leptonic BR	yes	vary BR by 0.77/-0.45% or scale Pythia8 up to PDG BR where needed	see PDG