# Evolving MC tuning methodology: robust estimation of tune uncertainties

**Andy Buckley**, University of Glasgow
Holger Schulz, FNAL / University of Cincinnati

QCD@LHC 2019, Buffalo, 15 July 2019

# Introduction

**MC tuning: a necessary evil!**

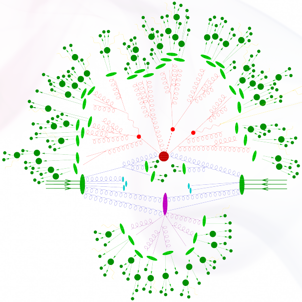
- Experiments need good data-description
- And *ab initio* theory needs to be comparable with data

⇒ **fitting pheno models to data** (cf. PDFs!)

**Professor** is numerical machinery frequently used to aid MC generator tuning. Used by Sherpa, Herwig, ATLAS, CMS. . .

**Data and models aren't perfect:** need to estimate tune systematics. Methods exist, but large arbitrariness

**In this talk: overview of tuning methodology, and putting tune systematics on a statistically sound footing**
And re-learning basic statistics! (cf. PDFs!)



# Context

**Think back to 2008-2012:** new collider, very new energy regime, even $\sim 100\%$ uncertainty on $\sigma_{\mathrm{tot}}^{pp}$!

$\Rightarrow$ Flurry of new tunes & methods. First PYTHIA6, then Py8 and other C++ gens. Eventually Monash and ATLAS/CMS tunes for Py8, author tunes for Herwig and Sherpa

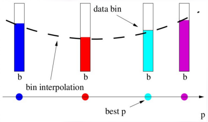
**First tuning heyday has passed!** Core tunes largely sufficient, except:

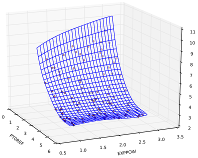
- MB/UE model tensions – e.g. pile-up modelling
- Strange and heavy-flavour production / challenges to hadronisation universality
- Perturbative tuning, e.g. Powheg HDAMP, specialist DY tunes, matched tunes

For most purposes, SHG default tunes are decent data proxies. **But Run 3 & HL-LHC $\Rightarrow$ new pressure on MC**

# The Professor method

- **MC is slow: $\gtrsim$ 1 CPU-day per run**
  $\Rightarrow$ can't use in serial optimisation.
- Simple solution: trivially parallelise MC runs through ranges of parameter space, and use sampled points to interpolate each bin's param dependence. Up to $\mathcal{O}(15)$ params.
- Usually use SVD polynomial fits – requires that values vary in a polynomial fashion *or are transformed to do so.* Not fundamental.
- **Fast analytic interpolations**
  **$\Rightarrow$ serial minimisation of an objective function.** Typically pseudo-$\chi^2$
- Available as public C++/Python code



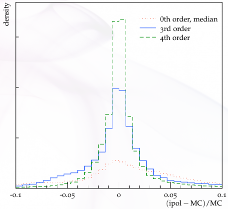



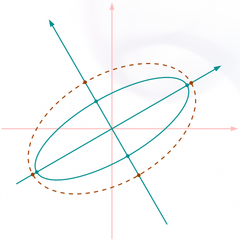
# Goodness-of-fit and systematics

- Usually optimise a simple pseudo-$\chi^2$:

$$\chi^2(\vec{p}) = \sum_b w_b^2 \frac{(f_b(\vec{p}) - \mathrm{ref}_b)^2}{\Delta\mathrm{ref}_b^2 + \Delta f_b^2(\vec{p}) + \epsilon^2}$$

- Note weights $w_b$ and regularising $\epsilon$. Correlations possible, but rarely available. Parametrisation error $\Delta f_b(\vec{p})$ probably an overestimate
- GoF defines the best fit: can we get systematics from its shape? Yes: *eigentunes*, cf. PDF eigenvectors.
  - Maximally orthogonal error sources
  - Reasonable number (ish)
  - **But: in practice, $\Delta\chi^2$ tolerance rules don't work...**

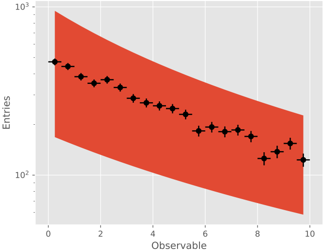


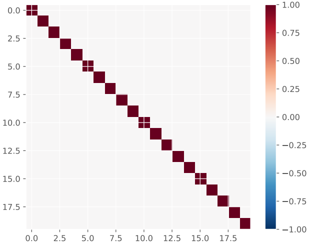



# Toy model

Let's explore the basic statistics a bit, so we know what we're doing.

**Toy model to both generate pseudodata and tune:**
$y_b = p_0/(p_1 + x_b)^2$



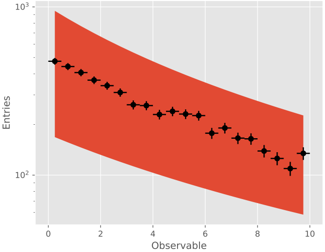



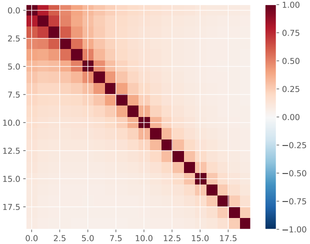

**Correlation modes from "none" to "mad"**

# Toy model

Let's explore the basic statistics a bit, so we know what we're doing.

**Toy model to both generate pseudodata and tune:**
$y_b = p_0/(p_1 + x_b)^2$



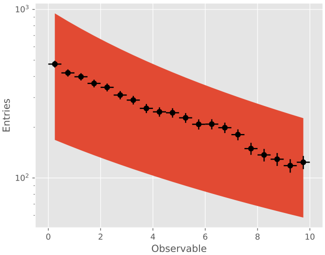



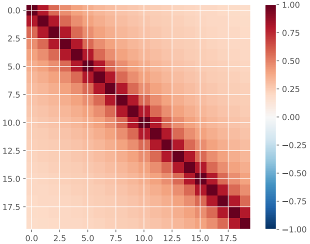

**Correlation modes from "none" to "mad"**

# Toy model

Let's explore the basic statistics a bit, so we know what we're doing.

**Toy model to both generate pseudodata and tune:**
$y_b = p_0/(p_1 + x_b)^2$

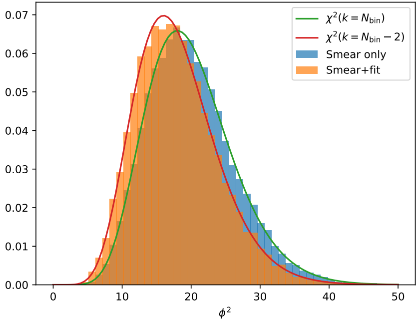



**Correlation modes from "none" to "mad"**

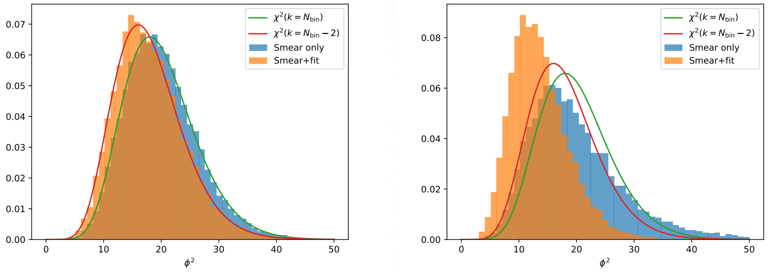
# Failure of the $\chi^2$ construction

- **A true $\chi^2$ statistic will be described by a chi-square distribution with an appropriate number of degrees of freedom.**
- If each of $N_b$ bins fluctuates independently, that's $N_b$ degrees of freedom. The Professor fitting to noisy data using $N_p$ params reduces it to $k = N_b - N_p$, e.g. in this 2-param fit:



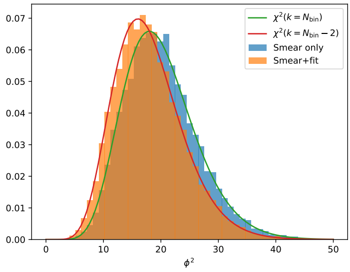
# Failure of the $\chi^2$ construction

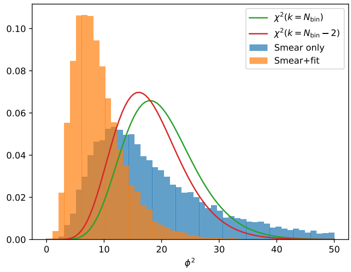
- **A true $\chi^2$ statistic will be described by a chi-square distribution with an appropriate number of degrees of freedom.**
- This gets broken by correlations – from shared kinematics, normalisations, and experimental systematics:



# Failure of the $\chi^2$ construction

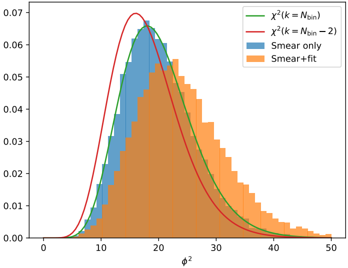
- **A true $\chi^2$ statistic will be described by a chi-square distribution with an appropriate number of degrees of freedom.**
- And bigger correlations…note $\chi^2$ *reducing* since fewer true d.o.f as correlations get stronger:

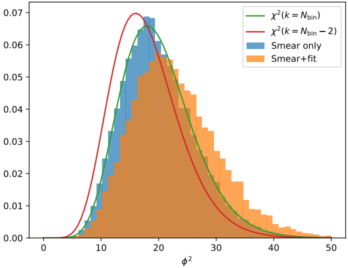


# Failure of the $\chi^2$ construction

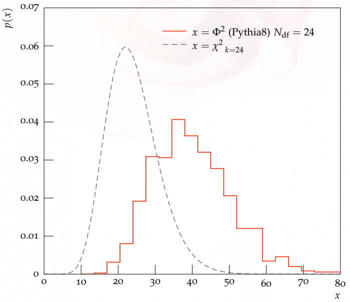
- **A true $\chi^2$ statistic will be described by a chi-square distribution with an appropriate number of degrees of freedom.**
- Perhaps more importantly, $\chi^2$ scaling is only true if the model *can* describe all the data – what if we break it? (fit wrong exponent)



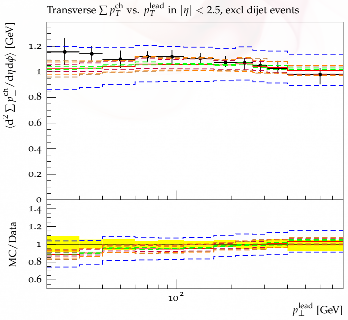
# Failure of the $\chi^2$ construction

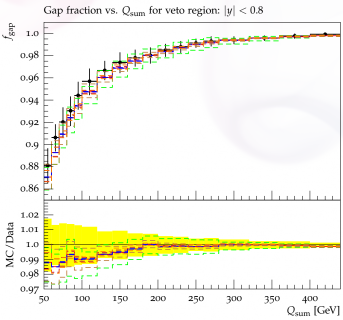
- **A true $\chi^2$ statistic will be described by a chi-square distribution with an appropriate number of degrees of freedom.**
- Real generator scaling looks more like model or data tension:



# Failure of the $\chi^2$ construction

- **A true $\chi^2$ statistic will be described by a chi-square distribution with an appropriate number of degrees of freedom.**
- And $\Delta\chi^2$ also fails: idea is that bin fluctuations cancel, so $k \sim N_p$, but much larger. $\Rightarrow$ ATLAS A14 eigentunes done by eye:



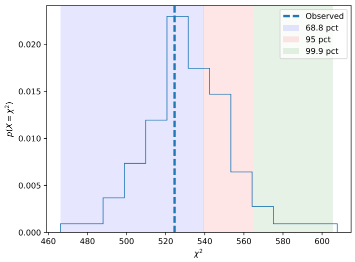
# Bootstrap to victory

**Recap:**

- real MC generator $\chi^2$ distributions don't scale as chi-squared
- due to a mix of unknown correlations, and incomplete models (and data tensions…)
- $\Rightarrow$ **usual recipe fails**

But we don't need chi-squared scaling: how about using empirical test-stat intervals?

Introduce bootstrap smearing (again cf. PDFs!): re-sample many replicas from distribution bins, and find best $\phi^2 = \chi^2/2$




**But: best-fit will also be outside the CL some fraction of the time**

# Better tuning errors from the bootstrap

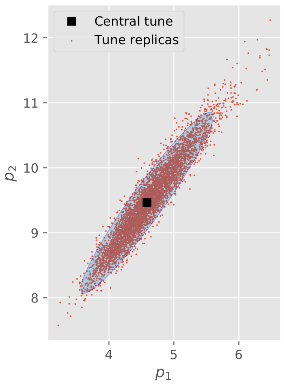
**So *ignore* the $\phi^2$ after fitting: just work with the replica distribution**

Can be the end of the story for PDFs, but not for tunes: can't cheaply reweight an MPI or hadronisation tune

Need to reduce, maybe cf. mc2hessian [arXiv:1505.06736] but needs a *basis*: ok for 1-variable PDFs, not for general MCs

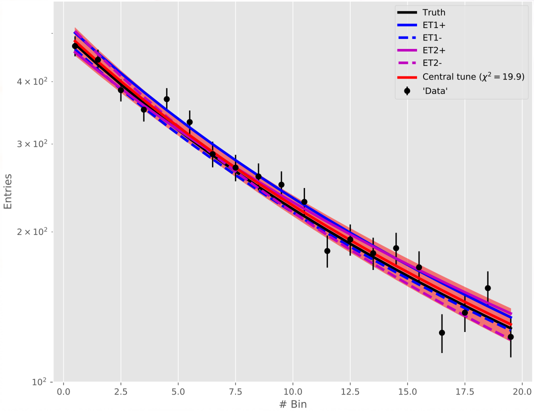
**Instead, construct Hessian ellipsoid to give CL coverage of replicas:**

- Centre from nominal best-fit or mean of replicas
- Orientation and aspect from minimiser covariance or replica covariance
- **Take intersections of principle axes with ellipse as $2N_p$ error tunes**



# Applications

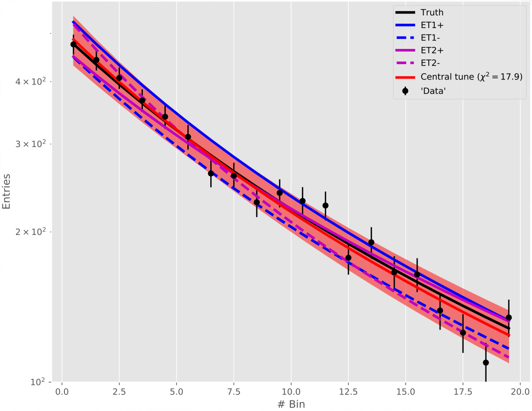
**First the toy model:**




Lack of correlations in $\phi^2$ may bias the tune params to non-ideal places, but naïve data coverage is good…note correlation effect

# Applications

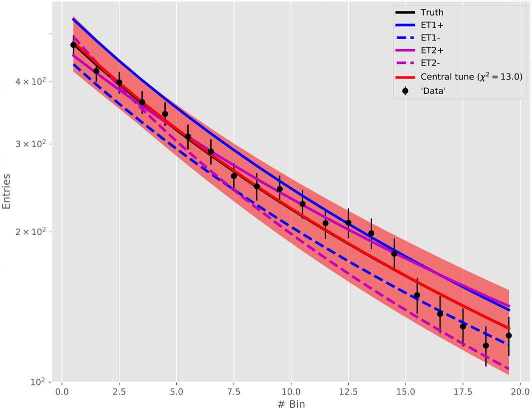
**First the toy model:**




Lack of correlations in $\phi^2$ may bias the tune params to non-ideal places, but naïve data coverage is good…note correlation effect

# Applications

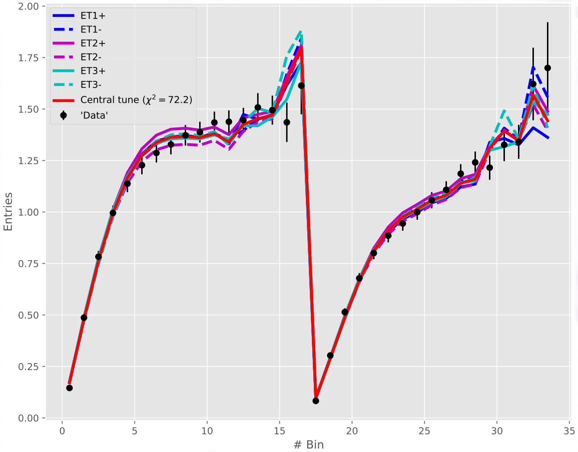
**First the toy model:**




Lack of correlations in $\phi^2$ may bias the tune params to non-ideal places, but naïve data coverage is good…note correlation effect

# Applications

**And a real application** (3-param Pythia 8 UE tune, concatenated vars):




Model limitations more important than correlations here

# More methodology improvements

**Several other developments ongoing, interesting for "next round" tunes:**

- **Padé rational approximants:** for non-polynomial parameter-dependence (of $y_i(\vec{p})$, not $y(x_i)$)
- **Auto-tuning / portfolio "metatuning":** attempt to reduce arbitrariness of parameter weight choices. Really possible, or at risk of being driven by latent biases?
- **Error-tune dimensional reduction:** more robust wanted / needed? cf. ATLAS A14 procedure
- **Correlations:** can try post-hoc estimation of correlations by MC Poisson bootstrap – but *far* better that this comes from the experiments

# Summary & outlook

- **MC tuning not very active right now, but precision data challenges MC in new areas: it will return!**
- Professor is a well-established tool to aid in many-parameter MC tuning. *Not a replacement for physics awareness.*
- Also uses in BSM fitting and model exploration: it's all fitting! cf. unfolding, PDFs, …
- Eigentunes also quite established, but dirty secret of arbitrary $\Delta\chi^2$ tolerance
- Simple statistical toys show the issues, and lead to a way forward through empirical $\phi^2$ bootstrapping, and a new, coverage-based eigentune construction
- Looks good on toy model, needs some debugging in real-data case, but should be complete soon
- Other methodology developments, and experimental correlation-culture $\Rightarrow$ ready for the next phase