Precision measurements

PHYSTAT-Systematics, S. Glazov, 1 Nov 2021

Outline

- Propagation of uncertainties: offset and toy MC methods
- Covariance matrix and nuisance parameters representation
- Reduced uncertainty sets
- Bias estimation
- Statistical uncertainty on the systematic uncertainty
- Systematics vs cross check
- Post-fit separation of statistical and systematic uncertainties

Error propagation: offset method

Consider differential (binned) cross section measurement μ_i which depends on detector/physics modelling, background subtraction and other systematic sources s_j. As the result of calibration, best central values for s_j are determined to be s_j^0 and their uncertainties δs_j.

In the offset method, the uncertainty on μ_i is determined by varying s_j by δs_j:

$$\Gamma_{ij} = \mu_i(s_j^0 + \delta s_j) - \mu_i(s_j^0)$$

Both $\pm\delta s_j$ variations are often considered to determine Γ_{ij}^+ and Γ_{ij}^-, as well as symmetrized version $0.5(\Gamma_{ij}^+ - \Gamma_{ij}^-)$.

(there are also other prescriptions for symmetrisation, it is important to keep signs for Γ_{ij})

Error propagation: toy MC method

Consider systematic source (e.g. electron identification efficiency) which is determined using control channel in multidimensional space with N_s bins in total and a measurement with $N_b \ll N_s$ bins. In this case instead of using N_s offset sources, it is often better to use N_r toy MC replica such that $N_b < N_r < N_s$.

- Before analysing data, prepare N_r randomly varied systematic tables: $s'_j = s^0_j + r_j \delta s_j$, where r_j follows normal distribution $N(0, 1)$. Other distributions can be considered as well.
- Perform m^r_i measurements using modified systematic sources.
- Estimate uncertainty using standard deviation of the results:

$$\sigma\mu_i = \sqrt{\frac{1}{N_r - 1} \sum_{k=1}^{N_r} (\mu^r_i - \overline{\mu^k_i})^2}\,.$$

- Other estimators, such as qunantile-based σ_{68} can be used too.

Covariance matrix and nuisance parameters

Systematic uncertainties determined by the offset method can be directly used in profile likelihood function:

$$\chi^2(m_i, b_j) = \sum_i \left(\frac{m_i - (\mu_i + \sum_{j=1}^{N_s} \Gamma_{ij} b_j)}{\sigma_i} \right)^2 + \sum_j (b_j)^2 ,$$

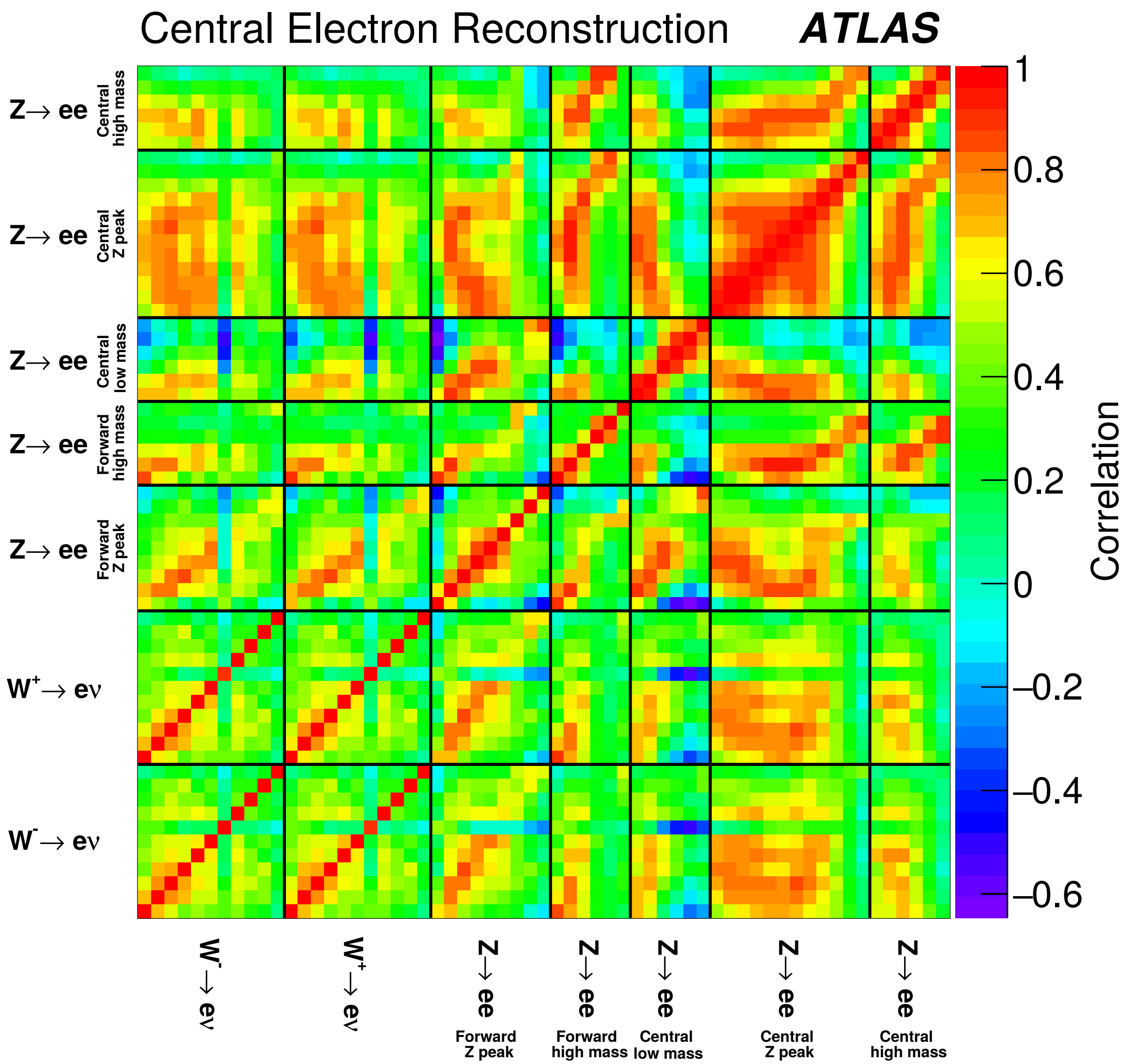
where σ_i are stat. uncertainties on μ_i, m_i are expectations and b_j are nuisance parameters for systematic sources, $b_j = (s_j - s_j^0)/\delta s_j$. They can be also converted to covariance matrix:

$$C_{il}^{\mathrm{sys}} = \sum_{j=1}^{N_s} \Gamma_{ij} \Gamma_{lj}$$

For the toy MC method, covariance matrix can be directly determined as:

$$C_{il}^{\mathrm{sys}} = \frac{1}{N_r - 1} \sum_r \sum_{k=1}^{N_r} (\mu_i^r - \overline{\mu_i^k})(\mu_l^r - \overline{\mu_l^k})$$

Precision measurement of W, Z cross sections at ATLAS

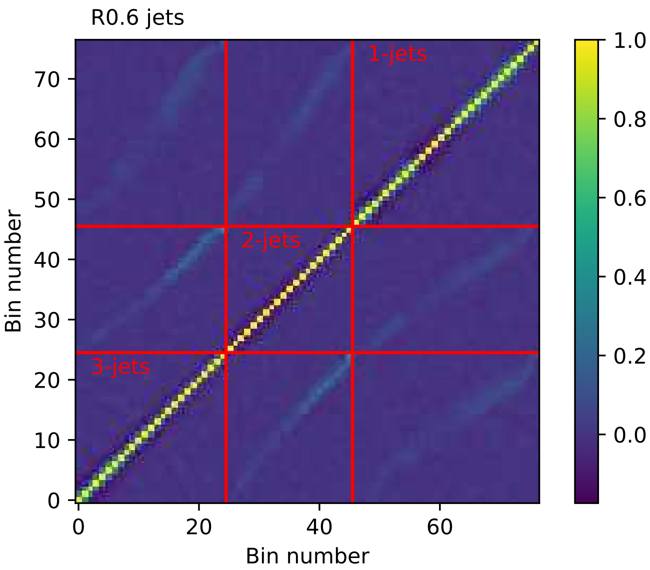


ATLAS, EPJC 77 (2017) 367

Correlation matrix for the central electron reconstruction efficiency for the correlated measurement of differential in η_ℓ W- and in $y_{\ell\ell}$ Z-boson production cross sections.

Correlation of toys across analyses is essential for the method.

Using results from different publications



Stat. correlation for $R = 0.6$ jets, first bin in y, y^* or $|Y^*|$.
ATLAS, JHEP 02 (2015) 153 (1-jet)
ATLAS, JHEP 05 (2014) 059 (2-jet)
ATLAS, EPJC (2015) 75 (3-jet)
Plot is based on the HEPDATA records of the publications

- Consistent treatment of the data in PDF fits requires proper treatment of correlations across publications
- Both statistical and systematic correlations should be taken into account
- For statistical correlations, one way to keep it is to store bootstrap replica of the measurement, with bootstrap generator seeded uniquely for each data event (ATL-PHYS-PUB-2021-011).

From covariance matrix to nuisance parameters

The covariance matrix $C_{il}^{\rm sys}$ can be decomposed to be into a product of Γ_{ij} using different methods. For example, using eigenvectors V and diagonal matrix built using eigenvalues U:

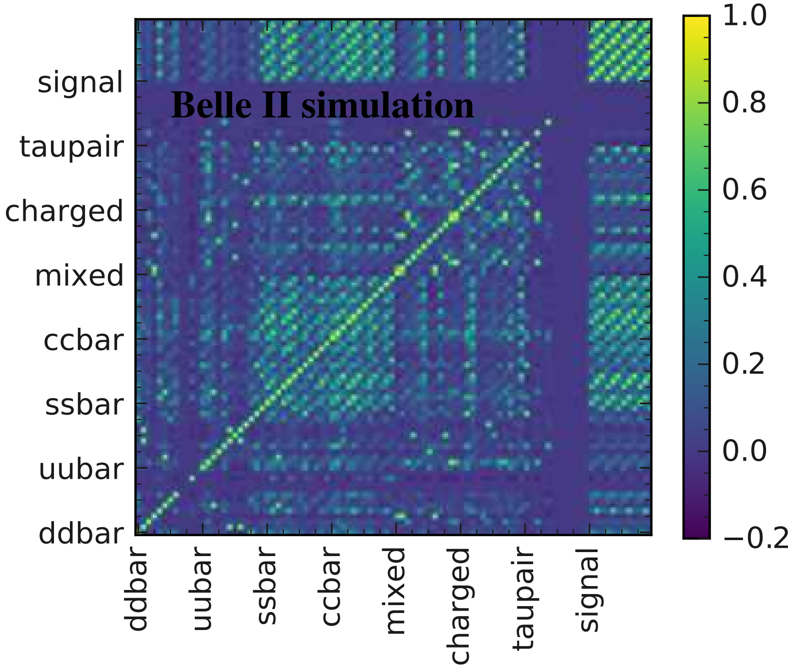
$$C = V^T U V = (V\sqrt{U})^T (V\sqrt{U}) = \Gamma^T \Gamma\,.$$

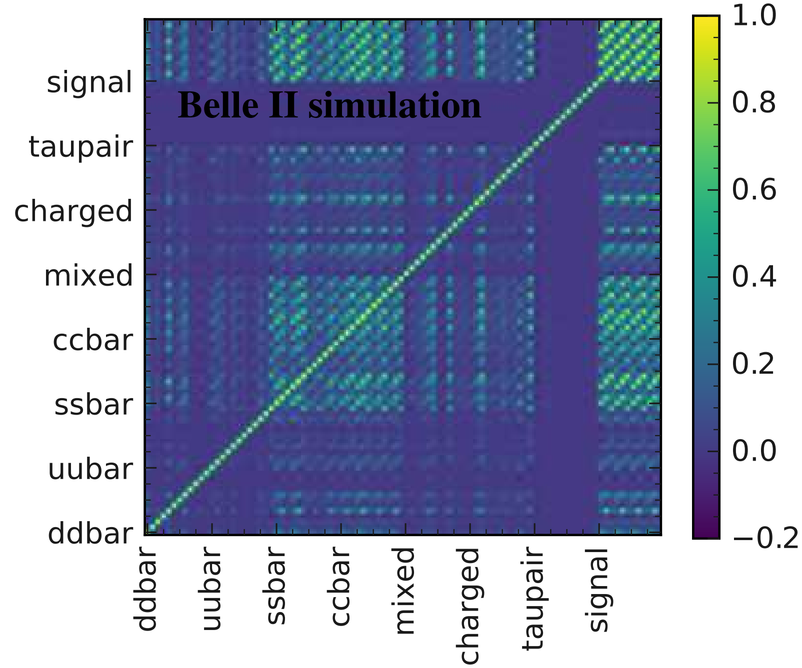
Note that if $N_m > N_s$ or $N_m > N_r$, only N_s or N_r eigenvalues of $C^{\rm sys}$ are above zero, other should be equal to zero but may numerically become below zero. The corresponding eigenvectors should be pruned from Γ_{ij}.

This procedure is useful to convert toy MC to nuisance parameter representation, and also reduce number of systematic sources.

The reduced representation can be checked by building covariance matrix and comparing it with the original.

PID systematics for $B \to K\nu\bar{\nu}$ search at Belle II

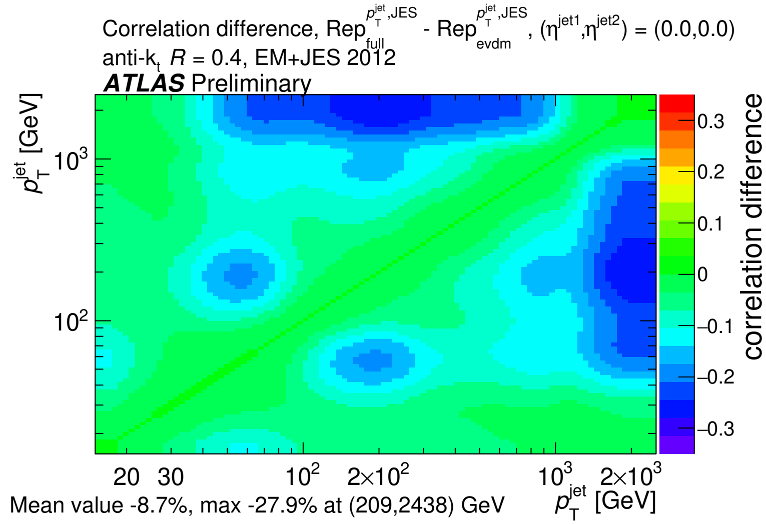




Efficiency and fake $\pi \to K$ PID weights stat. uncertainty, determined in control mode, propagated using toy MC method to $B \to K\nu\bar{\nu}$ analysis which is performed in $3 p_T^K \times 4$ BDT bins. Uncertainties affect both signal and various background contributions. Left: original correlation matrix with $N_r = 500$ toys, right: correlation matrix based on 3 leading eigenvectors.

Belle II, PRL 127, 181802 (2021)

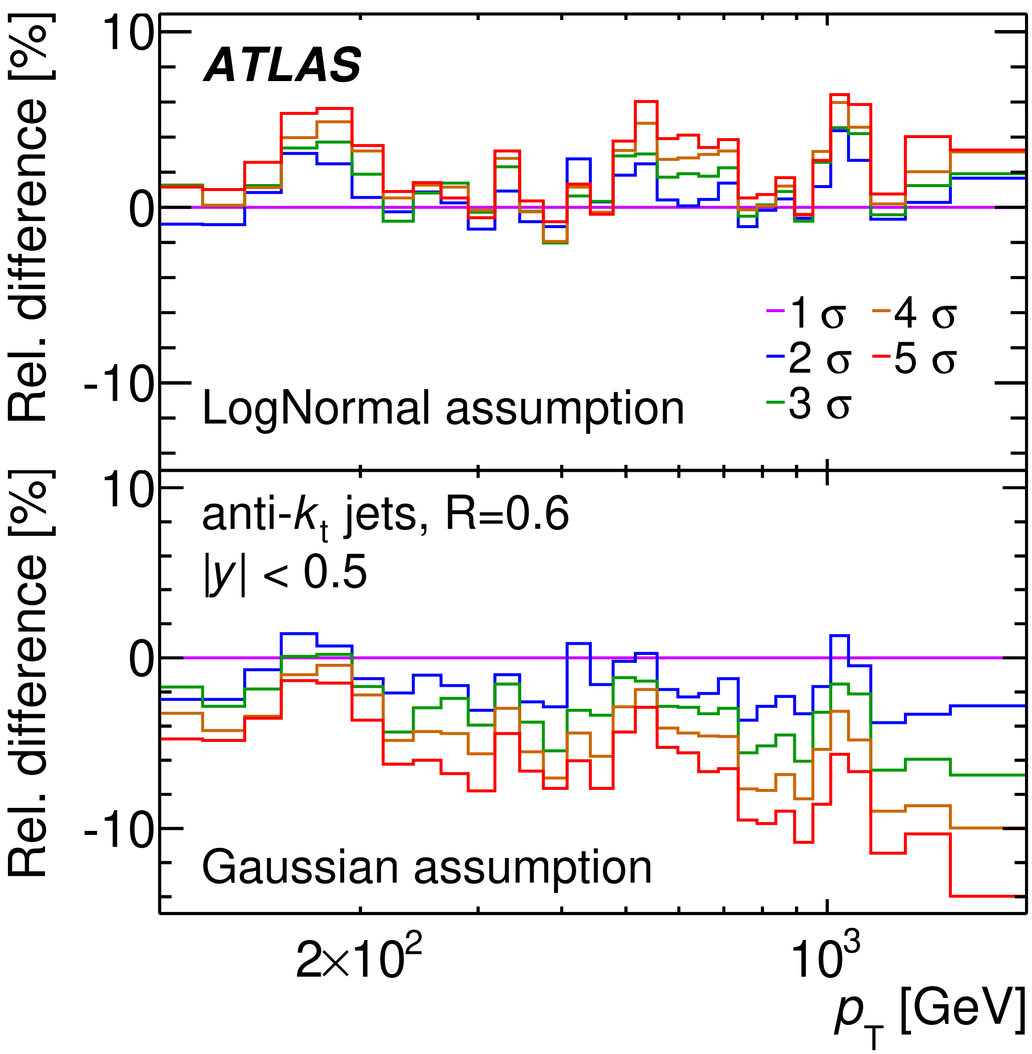
ATLAS Jet systematics pruning



- JES is described by $O(100)$ sources, may be redundant for stat. bound analyses.
- Several alternative strongly reduced sets, to bound uncertainty from this approximation
- Alternative correlation representations, to estimate uncertainty on the correlation model, used, e.g. in ATLAS, JHEP 09 (2017) 020.
- Different methods to measure correlation difference between representations.
- Ideas how to improve treatment of two-point sources

  ATL-PHYS-PUB-2015-014, EPJC 80 (2020) 12

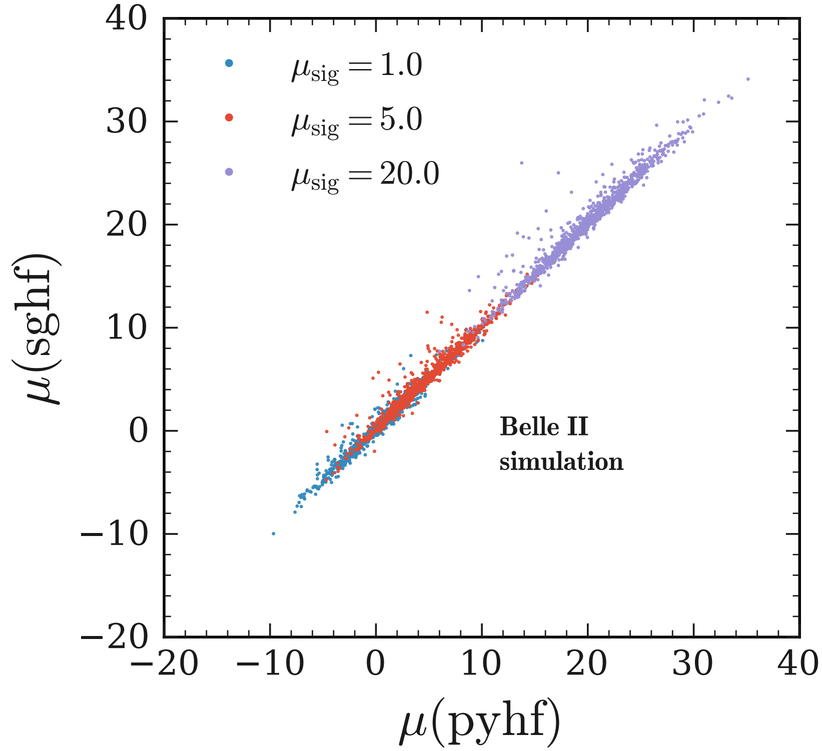
Non Gaussian effects



JHEP 02 (2015) 153

In many cases, log-normal distribution can give more adequate description of cross section variation caused by a systematic source. This is tested explicitly by ATLAS 7 TeV inclusive jet measurement where an jet energy scale component (energy scale in calorimeter) is varied by up to 5σ and the resulting cross sections are compared to the estimate based on 1σ variation plus Gaussian and log-normal assumption. Log-normal gives better descriptions, with reduced bias opposite to Gaussian.

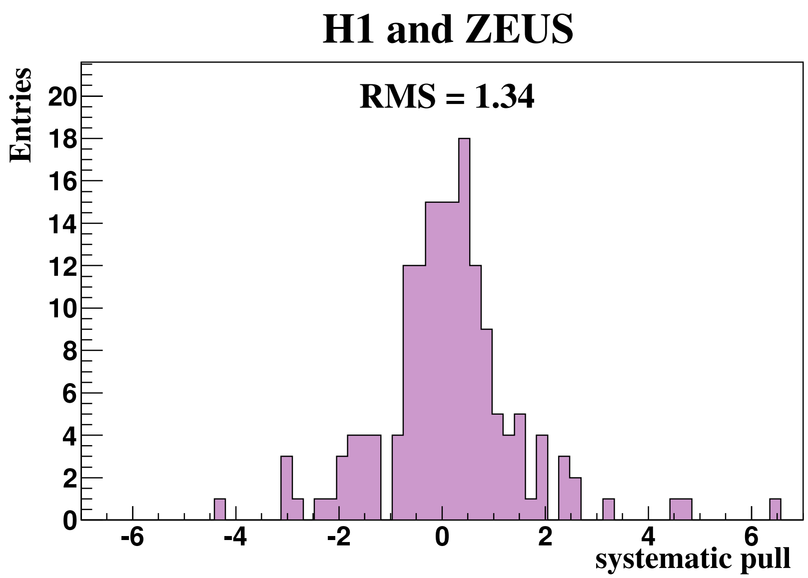
Improved χ^2 vs full model



← signal injection studies in Belle II PRL 127, 181802 (2021) analysis. `sghf` stands for "simplified Gaussian hist factory" which uses input compatible with `pyhf`

Given that PDF of systematic sources is not exactly known, there is often not much advantage using complex likelihood functions vs improved Gaussian least-square tools. E.g. HAverager and BLUE use "multiplicative" errors to approximate log-normal distribution, $\Gamma'_{ij} = \frac{m_i}{\mu_i}\Gamma_{ij}$. Detailed benchmarking results show very similar behaviour, with least-square tools being orders of magnitude faster.

Checks of data model

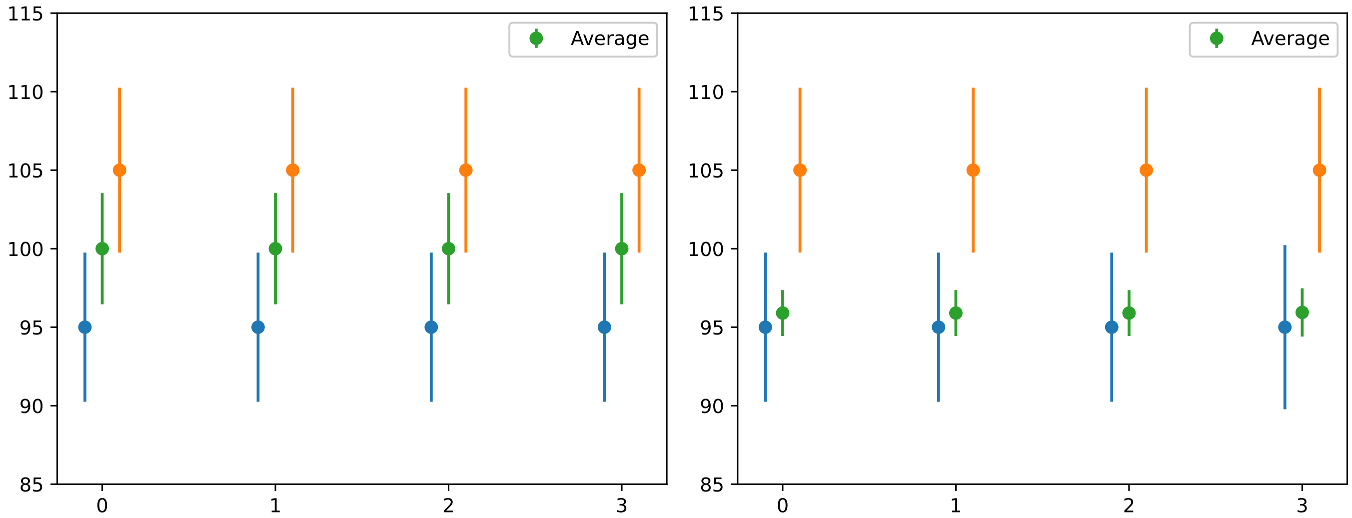


Statistical model for the data description can be studied by examining profile likelihood fit (combination) results for systematic sources:

- Large shifts $|b_j| > 2$: PDF for systematics is approximate.
- Large uncertainty $\sigma b_j \ll 1$ reduction: potentially too tight correlation, "uncertainty on uncertainty"
- Pulls $\frac{b_j}{\sqrt{1-\sigma^2 b_j}}$ should follow normal $N(0,1)$ distribution

H1 and ZEUS Eur.Phys.J.C75 (2015) 12, 580

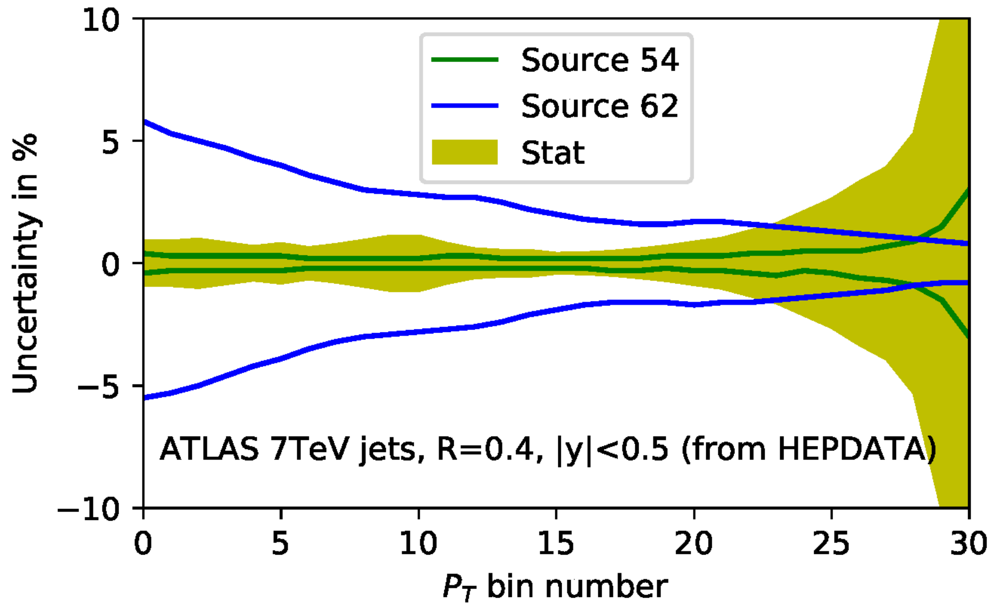
Uncertainty on uncertainty



Systematic uncertainty vectors Γ_{ij} are considered in profile likelihood as determined without uncertainty. This is not generally true and may lead to incorrect correlation pattern. Consider combination of two toy measurements in 4 bins: $\mu_i^1 = 95, \mu_i^2 = 105$, $\sigma_{\text{stat}}^{1,2} = 0.1\%$. Each measurement contains single independent "global normalisation", systematic source b_1, b_2 each of 5%. Consider also "fluctuation" of the systematic uncertainty estimate for the last bin of the first measurement to 5.5%. The combined results are pretty different.

Source	Shift	Error	Pull	Shift	Error	Pull
b_1	1.0	0.71	-1.41	0.19	0.30	0.19
b_2	-1.0	0.71	+1.41	-1.98	0.31	-1.99

Estimation of statistical uncertainty on systematics

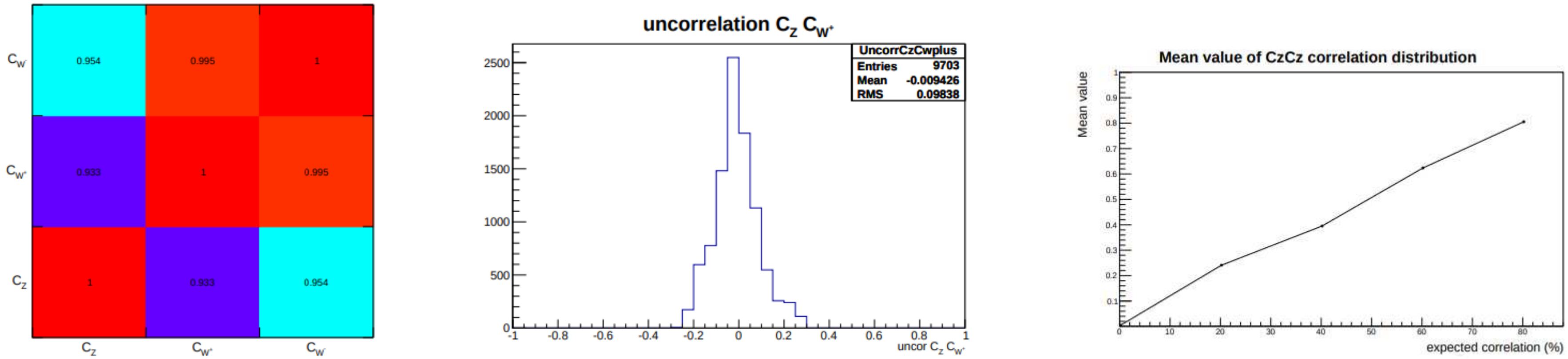


Statistical uncertainties for systematic sources arise primarily from kinematic migrations (e.g. JES for inclusive jet measurement).

- Estimate stat. uncertainties using bootstrap method (Poisson weights with $\lambda = 1$, 10K replica).
- Smooth the fluctuations based on statistical significance, by rebinning and Gaussian kernel smoothing.

ATLAS, JHEP 02 (2015) 153

Stat. uncertainty on systematics: ToyMC method



- Simultaneous measurement of W^+, W^- and Z cross section at 13 TeV
- Sizable correlation due to electron identification (left), measured with 100 replica of toys.
- Correlation vanishes for any wrong permutation of toys (center) → use to determine stat. error for zero correlation.
- Keep fraction of toys in sync: can scan different correlation points (right)

ATLAS PLB 759 (2016) 60; A. Trofymov PhD thesis

Systematics vs cross check

- Systematic uncertainties should address calibration and modelling issues for the current analysis setup.
- Difference between alternative analyses should be treated as a cross check. To make them useful, determine uncertainty on the analysis difference (e.g. use bootstrap for statistics).
- Variations of the analysis setup such as cut variation, fit model change should be avoided as systematics and treated as cross checks.
- Other two-point systematic sources such as unfolding using different MC models should be reduced in favour of using detailed (theoretical) uncertainties for the selected primary model.

See also hep-exp:0207026 for more discussions.

Decorrelation of systematics

- Formula such as $C_{il} = \sum_j \Gamma_{ij}\Gamma_{lj}$ and the penalty term $\sum(b_j)^2$ in the profile likelihood assume that input systematic sources are uncorrelated. Programs such as HAverager return orthogonal sources after the combination.
- It is not always straightforward to ensure that input uncertainties are uncorrelated.
- An example is lepton isolation requirement which is correlated to pileup and shower simulation: efficiency scale factors need to be determined for different pileup conditions to avoid double counting of effects.
- In general, control channels which are used to control systematic should be treated with the same set of systematic uncertainties as the signal.

Separation of stat. uncertainty after combination

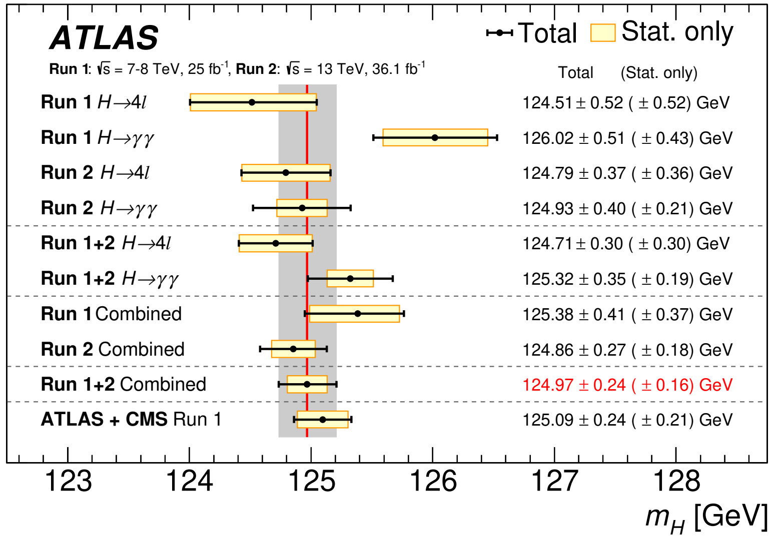
What is the separation between statistical and systematic uncertainty for the result which is obtained by a combination of different categories, channels, experiments? There are two prescriptions:

- Commonly used in profile likelihood fits: fix b_j to the best fit result, repeat the fit with statistical uncertainties only.
- Textbook prescription: propagate uncertainty according to the weight of each component using ”dispersion rule”. Implemented in BLUE, can be introduced to profile likelihood fits with e.g. toys.

Similar prescriptions exist for identifying impact of the input systematic source on the combined measurement.

The prescriptions may give significantly different results when combining measurements with similar accuracy but different statistical vs systematic uncertainty components.

ATLAS Higgs mass combination



Combination of RunI plus partial RunII ATLAS Higgs mass measurements in $H \rightarrow \gamma\gamma$ and $H \rightarrow 4\ell$ channels using profile likelihood with nuisance parameters for systematic uncertainties. First prescription gives “Stat. only” uncertainty of 0.16 GeV

BLUE combination result:

$$m_H = 124.97 \pm 0.23\text{GeV} = 124.97 \pm 0.19(\text{stat}) \pm 0.13(\text{syst})\text{GeV}$$

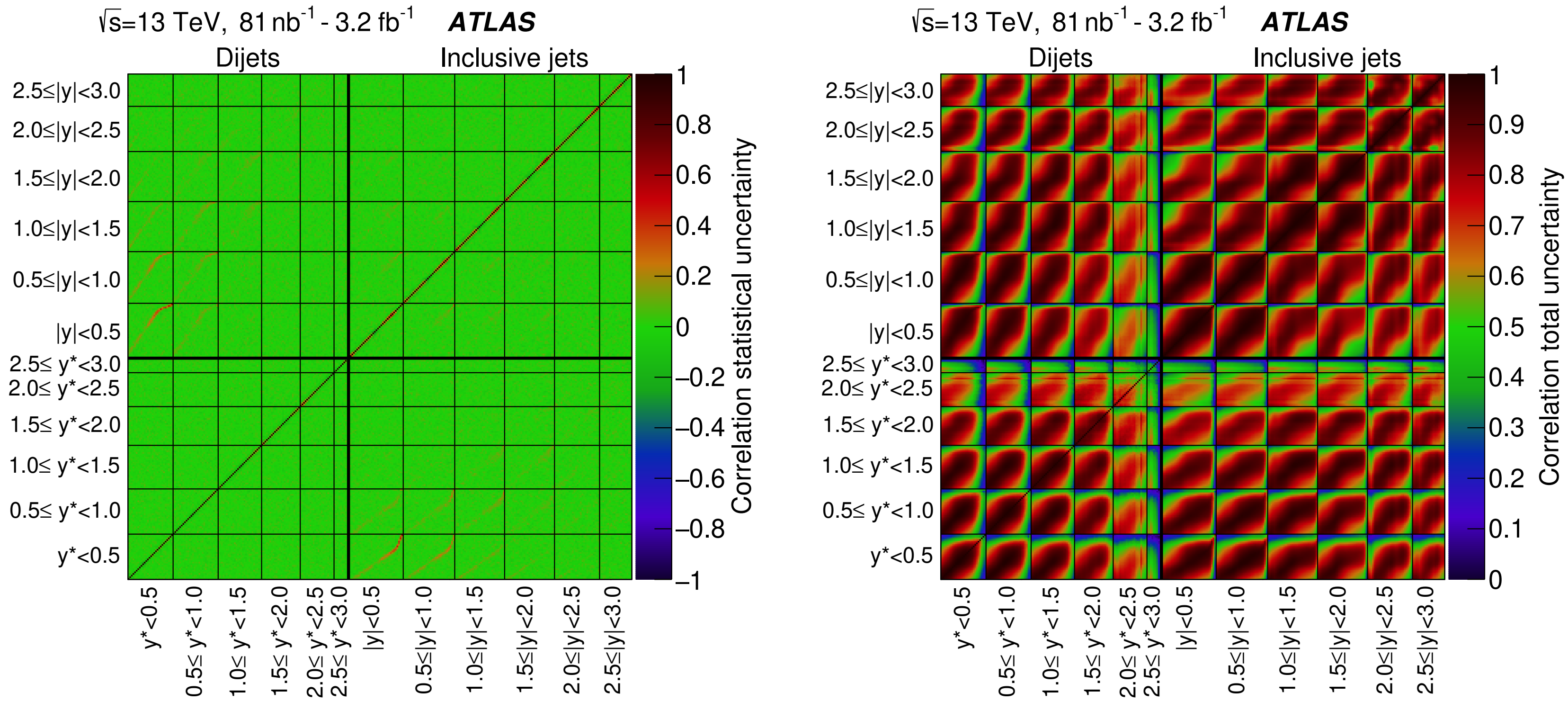
$\rightarrow$ the measurement turns from syst. to stat. uncertainty dominated, because of the larger weight of the $H \rightarrow 4\ell$ channel.

ATLAS, PLB 784 (2018) 345

Summary

- Two main methods to propagate systematic uncertainties, with their relative advantages and disadvantages.
- It is important to reduce statistical component in systematics
- Simplified least-square methods such as BLUE, HAverager, sghf can be used for fast tests instead of more complex tools; full PDF for systematics is an ultimate dream.
- “Bad” systematic sources, such as two-point systematics, should be avoided.
- Quoted statistical uncertainty may depend on prescription.
- Toy MCs/Bootstrap are useful tools for studies of systematics.

ATLAS inclusive jets and dijets at 13 TeV



- ATLAS inclusive and dijet measurement at 13 TeV.
- Statistical correlations determined using bootstrap method, all correlations provided for phenomenological analysis.

ATLAS, JHEP 05 (2018) 195