A method for the construction of strongly reduced representations of ATLAS experimental uncertainties and the application thereof to the jet energy scale

Daniela Börner[1,a] on behalf of the ATLAS collaboration

[1]*Bergische Universität Wuppertal*

Abstract. A method is presented for the reduction of large sets of related uncertainty sources into strongly reduced representations which retain a suitable level of correlation information for use in many cases. The method provides a self-consistent means of determining whether a given analysis is sensitive to the loss of correlation information arising from the reduction procedure. The method is applied to the ATLAS Jet Energy Scale (JES) uncertainty, demonstrating that the set of 60+ independent sources can be reduced to form a representation constructed of 3 nuisance parameters. By forming a set of four such representations, it is shown that JES correlation information is retained or probed over the full parameter space to within an average of 1%. This procedure is expected to significantly reduce the computational requirements placed upon early ATLAS searches in the upcoming 2015 dataset while still providing sufficient performance and correlation structure to avoid changing the analysis results.

1 Introduction

Precision measurements of particle physics phenomena, as performed with the ATLAS detector [1] at the LHC [2], require accurate calibrations and a detailed understanding of physics objects. Improvements in the calibrations result in smaller uncertainties and better understanding of correlations between different phase space regions. This on the other hand also results in a larger set of uncertainty sources, none of which is clearly dominant compared to the others. The price of this precision is the increased computational complexity. Several types of analyses, e.g. searches for physics beyond the Standard Model, do not profit from the increased precision, but are limited by the additional time and resources needed to incorporate all uncertainty sources.

To reduce the complexity of analyses, methods were developed to reduce large sets of related uncertainty sources into strongly reduced representations with only few parameters, without affecting the total uncertainty. It is important that a suitable level of correlation information is retained after the reduction. Otherwise unphysical constraints on the total uncertainty could be derived from phase-space comparisons, for which the correlations differ between the original and the reduced set of parameters. These methods were developed and probed using the ATLAS jet energy scale (JES) uncertainty [3]. Additionally, the procedure how analyses can test the influence of the loss of correlation information on their results is explained, by comparing multiple reduced representations of the

---

[a]e-mail: daniela.boerner@cern.ch

uncertainties. Searches and measurements which are not sensitive to the details of the correlation loss can then use one of the reduced representation. This procedure allows a large fraction of analyses to use a simplified approach of determining the systematic uncertainties, without necessarily providing the optimal set of parameters for each individual analysis.

1.1 ATLAS JES uncertainty

In typical ATLAS analyses, jets are reconstructed using the anti-k_t algorithm with a radius of 0.4. The associated JES uncertainties are constructed from $N_{JES}^{full} = 67$ independent sources [4]:

- 56 components from *in-situ* balance techniques covering low, medium and high p_T^{jet} regimes, depend only on p_T^{jet}

- 2 components from *in-situ* dijet-intercalibration methods, depend on both p_T^{jet} and η^{jet}

- 1 component from single-hadron response measurements in the very high p_T^{jet} regime, depends solely on p_T^{jet}

- 4 components from pileup sensitivity, depend on a combinations of p_T^{jet}, η^{jet}, N_{PV} and $\langle\mu\rangle$

- 3 components cover flavour uncertainties, depend on p_T^{jet}, η^{jet} and the flavour composition being investigated

- 1 component applied to jets which are not fully contained in the calorimeters (punch-through), which depends on p_T^{jet}, η^{jet} and the number of segments in the muon detectors aligned with the jet
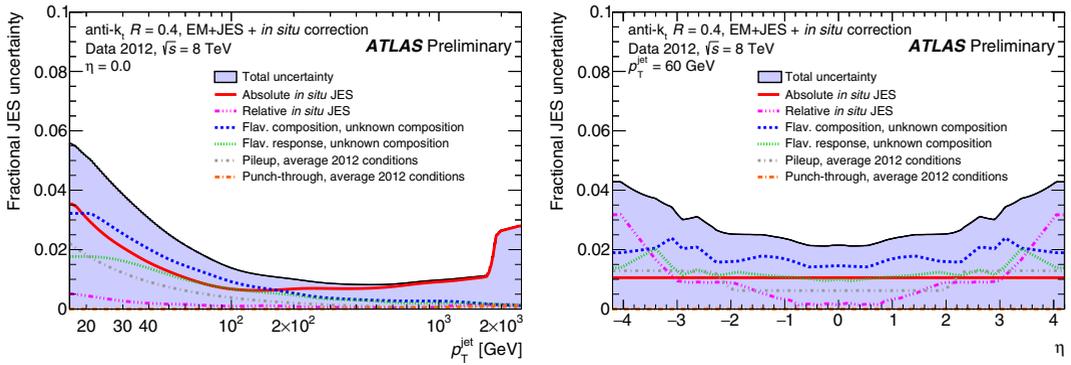

Figure 1. The full set of all 67 JES uncertainty sources are shown as a function of p_T^{jet} for $\eta^{jet} = 0$ (left) and η^{jet} for $p_T^{jet} = 60$ GeV (right). They are grouped by the general type of the component [3].

The total JES uncertainty, derived using the full representation as detailed above, is shown versus p_T^{jet} and η^{jet} in figure 1. The JES uncertainty has a remarkable precision of almost 1% for medium p_T^{jet}.

2 Reduction methods

The methods which can be used to reduce the number of components, also called nuisance parameters (NPs), will be explained in this section. For the ATLAS JES uncertainty, the number of NPs was reduced from 67 to 3, which was the lowest number of NPs that does not lead to a strong loss of information.

2.1 Correlation defintion

Reducing the number of components can lead to a loss of information about the correlation of the uncertainty in different regions of the parameter space. The correlation is defined as

$$\text{Corr}(j_1, j_2) = \frac{\text{Cov}(j_1, j_2)}{\sqrt{\text{Cov}(j_1, j_1) \cdot \text{Cov}(j_2, j_2)}} \tag{1}$$

The j_i stand for all quantities, on which JES depend on, e.g. p_T^jet, and the covariance is defined as

$$\text{Cov}(j_1, j_2) = \sum_{i=0}^{N_\text{set}} \Gamma(j_1) \cdot \Gamma(j_2) \tag{2}$$

The number of NPs in a specific set is given by N_set and Γ is the systematic shift. The correlation loss between a reduced and the original set of NPs is checked by taking the difference of the correlations between the two sets

$$\text{Corr}_\text{full} - \text{Corr}_\text{red. set.} \tag{3}$$

In the case of a significant loss of correlation information, the total uncertainty could be wrongly constrained by comparing different parameter space regions. Many analyses use so-called control regions in order to determine their backgrounds in a data-driven manner. The estimate is then propagated to the signal region. If the parameter space regimes differ between these regions, the correlation of the uncertainties can significantly affect the total uncertainty.

2.2 Eigenvector decomposition and merging

A large number of uncertainty sources depend on the same parameters. For the ATLAS JES uncertainty, e.g. the 56 *in situ* NPs depend all on p_T^jet. For NPs which depend on the same parameter, the covariance matrix can be build and diagonalized. The resulting eigenvectors do not have any physical meaning. In order to reduce the NPs, the dominant eigenvectors are kept. For this, the eigenvectors are ordered depending on their eigenvalue. All sub-dominant eigenvectors are merged using a quadrature combination to build a so-called residual term.

During LHC Run 1, ATLAS used to combine the 56 *in situ* NPs to 6 NPs. This reduced the number of NPs to 17 in total with only percent-level correlation loss. As the eigenvector decomposition and merging showed previously good reduction potential, it was also used as starting point for the strong reduction.

Compared to Run 1, this procedure is tightened: only the two dominant eigenvectors are kept and all sub-dominant ones are merged. This results in a reduction to 14 NPs in total.

2.3 Quadrature combination

The reduced set of NPs, after the application of the eigenvector decomposition and merging of the sub-dominant components, is still larger than the target of 3 NPs. As the NPs now only depend on different parameters, a different method than the eigenvector decomposition is required. For a further reduction, a good knowledge of the underlying uncertainty distribution is required. One nuisance parameter is selected to represent each region of interest. After identifying regions of interest in the parameter space, e.g. the low, medium and high p_T^{jet} region for the ATLAS JES, the components are grouped into sets, that are dominant in the same region. For each of these sets, the individual components are combined by taking the quadrature sum. This procedure results in as many NPs as dedicated regions of interest are defined, e.g. 3 NPs in the case of the ATLAS JES. This reduces the correlation loss and minimize, but does not forbid, the creation of new artificial correlation structures. This may not be the optimal solution. There could be parameters which do not depend on the parameter used to define the region of interest. If this is the case, this parameter should not be combined with the others. An additional NP should be introduced for such parameters.

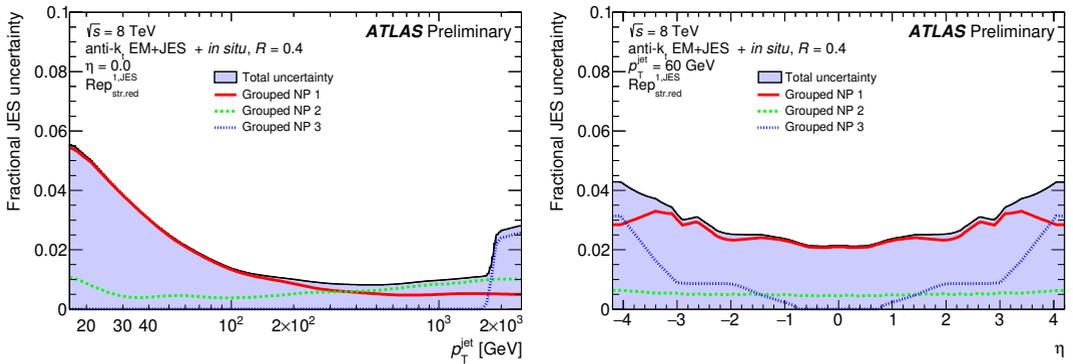### 2.4 Strongly reduced representation

Figure 2. The first reduced set of 3 JES uncertainty sources are shown as a function of p_T^{jet} for $\eta^{jet} = 0$ (left) and η^{jet} for $p_T^{jet} = 60\,\text{GeV}$ (right) [3].

Using the eigenvector decomposition and the quadrature combination, a first strongly reduced set is produced with only 3 NPs. In figure 2, the JES uncertainty is shown for the reduced set versus p_T^{jet} and η^{jet}. The different NPs – "Grouped NP x" – target different kinematic regimes in p_T^{jet} and η^{jet}. Especially in the distribution of the JES uncertainty versus p_T^{jet} (figure 2 left), the dominance of each NP is clearly visible in their dedicated region: low, medium and high p_T^{jet}.

The correlation difference for this reduced set is shown in figure 3 for several bins in η^{jet}. Every sub-plot is shown as a function of p_T^{jet} in the interval 15 GeV to 2.5 TeV using a logarithmic axis. All the white regions in the different sub-plots are beyond the kinematic limit for $\sqrt{s} = 8\,\text{TeV}$. The top (bottom) number in each sub-plot is the maximum (mean) observed correlation loss. For the mean correlation loss, the kinematic forbidden regions are excluded. Striking, especially in the higher η^{jet} bins, is the large correlation difference up to 85%. It is very important to improve this reduction method and provide possibilities to check whether this correlation loss is feasible for the different analyses.
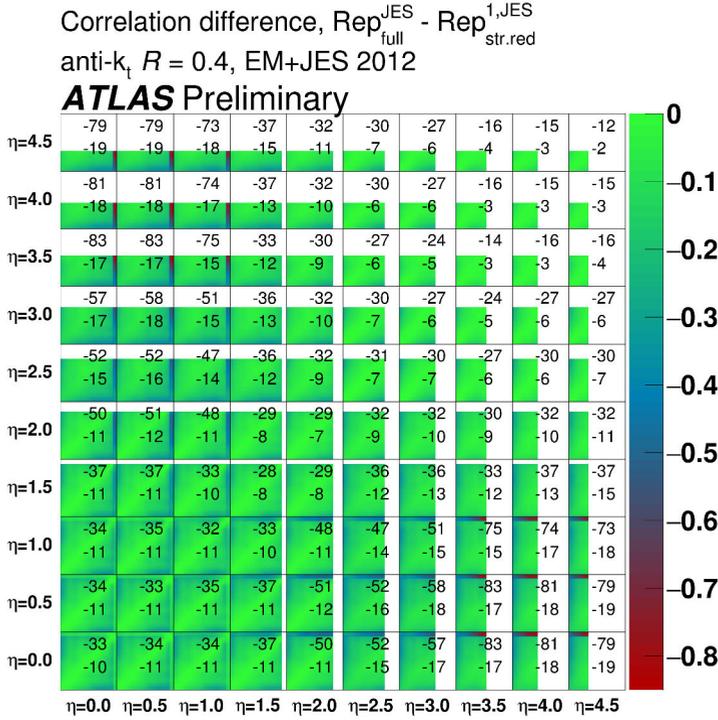
Figure 3. The correlation loss between the full set of 67 JES NPs and the first reduced set of only 3 NPs, calculated as described in formula 3. Each sub-plot is shown as function of p_T^{jet} in the interval 15 GeV to 2.5 TeV using a logarithmic axis. The white regions in the different sub-plots are beyond the kinematic limit for $\sqrt{s} = 8$ TeV. The top (bottom) number in each sub-plot is the maximum (mean) observed correlation loss. For the mean correlation loss, the kinematic forbidden regions are excluded [3].

3 Production of multiple representations

A way to check whether a large correlation loss is feasible for the different analyses is to provide several reduced representations. These multiple representations should target different phase-spaces such that the corresponding correlation loss targets different kinematic regions. For this, the reduction procedure needs to be adapted. Several different methods are explained in this section which vary in what is used to combine the final 3 NPs or are specific to reach a better performance in one kinematic phase-space.

3.1 Two-point systematics

Some uncertainties are derived by taking the full difference between two simulated samples. Normally, these two-point systematics are used as one component for the full uncertainty. An alternative method is to split this systematic in two components. These two components then have to be convoluted with complementary weights such that they add up to the originally uncertainty. The weight for the first component f_1 is selected to be in the interval 0 to 1 and chosen to follow a linear increase across $|\eta|$ or $\log p_T$. The weight of the second component calculates to: $f_2 = \sqrt{1 - f_1^2}$.

Using this method, a NP which depend on e.g. η is split such that the two sub-components also depend on p_T^{jet}. Then, these two sub-components can be combined with other p_T^{jet} dependent components. This split can add correlation information.

In figure 4, the usage of the procedure is shown for the ATLAS JES η-intercalibration. In the plot, the two sub-components are shown as well as their sum – the original uncertainty.
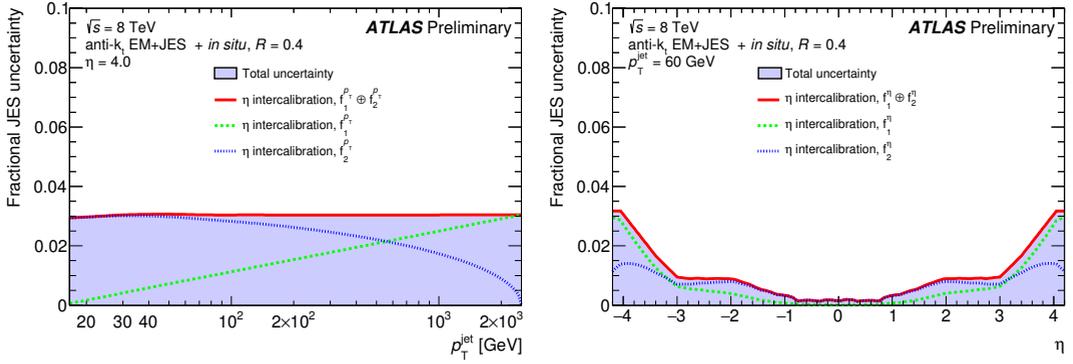

Figure 4. The ATLAS JES η intercalibration uncertainty is shown as a function of p_T^{jet} for $\eta^{jet} = 0$ (left) and η^{jet} for $p_T^{jet} = 60$ GeV (right). The different components are shown for the full uncertainty and the two sub-components [3].

3.2 NPs with large anti-correlations

A NP can have a change in the sign of the uncertainty, e.g. a large positive uncertainty at low p_T^{jet} and a negative uncertainty at high p_T^{jet}. If the magnitude of this component is non-negligible, a dedicated NP should be introduced in the strong reduction. Using a dedicated NP for these cases, no combination with other NPs is performed, as every combination would be performed as quadratic combination and thus result in absolute values. Using one NP for such NPs preserves the kinematic anti-correlation information.

3.3 Splitting the residual term

Another possibility is to weaken the eigenvector decomposition. In some kinematic regions, some eigenvectors can be more important even if their eigenvalue is overall smaller. A small eigenvector can be left out of the combination for the residual term. This does not provide a good global correlation structure, but it can be very helpful to target specific regions and therefore decrease the correlation loss in this region.

4 Multiple representations and usage for analyses

Using the ATLAS JES uncertainty, all the described methods were used in order to produce multiple strongly reduced representations of the JES uncertainty. In the end, four different sets are provided, which all target different correlation structures:

1. general representation of the full parameter space, with low, medium and high p_T^{jet} regimes (one NP each)

2. preserve the low vs medium p_T^{jet} correlation structure and $|\eta^{jet}|$ dependencies

3. preserve the medium vs high p_T^{jet} correlation structure

4. preserve the very high p_T^{jet} correlation structure

For these sets, 2 NPs are dedicated for the general representation of the correlation structure and the remaining NP is used in order to improve the correlation within a specific region. The correlation structures of the different strongly reduced representations cover complementary regions.

The performance of these four reduced sets need to be quantified. For this, two metrics were defined:

$$\text{Metric 1} = \text{sign}[\text{Corr}_{\text{full}} - \text{Corr}^i_{\text{str. red.}}] \times \min_{\text{Rep}^i_{\text{str. red.}} \in \{\text{Rep}_{\text{str. red.}}\}} |\text{Corr}_{\text{full}} - \text{Corr}^i_{\text{str. red.}}| \qquad (4)$$

$$\text{Metric 2} = \max(|\text{Metric 1}| - [\max_{\substack{\text{Rep}^i_{\text{str. red.}}, \text{Rep}^j_{\text{str. red.}} \in \{\text{Rep}_{\text{str. red.}}\}}}^{\text{Rep}^i_{\text{str. red.}} \neq \text{Rep}^j_{\text{str. red.}}} |\text{Corr}^i_{\text{str. red.}} - \text{Corr}^j_{\text{str. red.}}|]) \qquad (5)$$

The first metric gives the minimal correlation difference of one of the reduced sets compared to the full set of NPs. This allows one to see whether there is a region in which all metrics show a large correlation loss. If there are such uncovered region, this is acceptable, as long as the maximal difference between the reduced sets is larger than the minimal correlation loss between the reduced sets and the original full set. This definition was used for the second metric which checks the uncovered correlation loss.

In figure 5, the uncovered correlation loss (using Metric 2) is shown for the four reduced sets of the ATLAS JES uncertainty. The correlation loss between the four sets is as large as the correlation loss to the full set (up to 1% on average). This means that analyses which are unaffected by the difference between the scenarios are insensitive to this level of correlation loss.

In order to use the reduced set in any analyses, it has to be checked whether the specific analysis is sensitive to the correlation loss from the reduced set. For this, the results have to be calculated with all reduced sets. If the result is independent from the reduced set of NPs which was used, the full analysis is insensitive to the correlation loss and one of the reduced sets can be used. If the results differ using the reduced sets, the analysis is sensitive to the correlation loss and has to use the full set with all NPs.
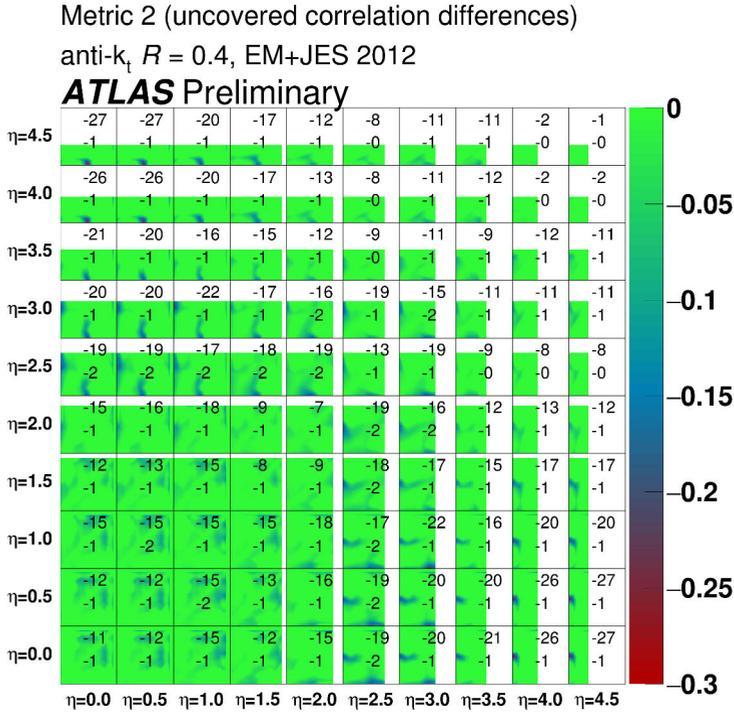
Figure 5. The uncovered correlation loss between the four reduced sets with only 3 NPs, calculated as described in formula 5. Each sub-plot is shown as function of p_T^{jet} in the interval 15 GeV to 2.5 TeV using a logarithmic. The white regions in the different sub-plots are beyond the kinematic limit for $\sqrt{s}$ = 8 TeV. The top (bottom) number in each sub-plot is the maximum (mean) observed correlation loss. For the mean correlation loss, the kinematic forbidden regions are excluded [3].

5 Conclusions

A procedure for the reduction of a large set of uncertainty sources related to a single experimental uncertainty was presented. As this reduction reduces the correlation information, different methods were shown in order to provide multiple strongly reduced representation. These methods were developed and tested for the specific use case of the jet energy scale uncertainty in the ATLAS experiment. These different representations can be checked by the analysis teams in order to test whether they are sensitive to the correlation loss. This procedure is the most commonly used approach for ATLAS searches for new physics in 2015 and 2016 as these analyses are not sensitive to the correlation loss. It was also discussed to apply similar methods to other experimental uncertainties, but all have already fewer components than the JES uncertainty.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008 JINST 3 S08003

[2] Lyndon Evans and Philip Bryant, *LHC Machine*, 2008 JINST 3 S08001

[3] ATLAS Collaboration, *A method for the construction of strongly reduced representations of ATLAS experimental uncertainties and the application thereof to the jet energy scale*, ATL-PHYS-PUB-2015-014, https://cds.cern.ch/record/2037436

[4] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur.Phys.J. C75.1 (2015) 17, arXiv: 1406.0076