“Shape-based” Scale Factors for Electrons

Martín Alcalde

September 2022

1 Introduction

Whenever we compare the data obtained from a real physics experiment with a simulation, we may see that the simulation doesn’t match the data to perfection. Simulations aren’t a perfect replica of reality, they are the best approximation we have to understanding the underlying processes that occur at a scale we can’t see or understand.

2 Motivation

In high energy physics, a huge amount of data is produced in order to have high enough statistics to allow us to carry out physical analyses. To filter the data produced in order to discard the uninteresting data, a series of cuts are applied to different values and magnitudes derived from the events. When the resulting data is compared with the simulation, we may see disagreements between them and in order to reconcile data and simulation a series of weights called *scale factors* (SFs) are calculated as the ratio between the efficiency of the data and the efficiency of the simulation.

For the purpose of extracting the values of the scale factors, we will use a framework known as *spark_tnp*. So far, this framework has only been used in the case of pairs of muons; in this project, we will focus on the viability of the application of this framework for the extraction of scale factors for electron pairs produced in Z-boson resonances from 2018.

3 Effinciency. The *Tag And Probe* method

As previously stated, the scale factors are calculated using a rather simple formula:

$$\mathrm{SF} = \frac{\varepsilon_{data}}{\varepsilon_{MC}} \tag{3.1}$$

Nevertheless, extracting the values is no trivial matter. In order to calculate the value of the scale factors, one first needs to know the value of the efficiencies of the reconstructed lepton pairs. In order to achieve an unbiased pair selection, a method known as *Tag And Probe* is applied.

The first step of this method consists in applying a very restrictive (tight) cut to the set of events. The ones that manage to pass this filter are then categorized as **tags**. This step allows us to make sure that one lepton of the pair is going to be a real lepton.

To find the other lepton for the pair, a loose cut is applied to the original set of events. The passing events are then known as **probes**. Note that a lepton can be both a tag and a probe.

In order to reconstruct the initial state, we pair tags and probes (passing probes and failing probes, separately) so that we have a set of passing pairs and a set of failing pairs. With each of these sets the resonance is reconstructed and the yields are calculated through a fitting.

Finally, to calculate the efficiency we just need to apply this next formula:

$$\varepsilon = \frac{P_{pass}}{P_{pass} + P_{fail}} \tag{3.2}$$

where P_{pass} and P_{fail} are the yields for the passing and failing pairs sets, respectively. Using this method, we can extract the values of the scale factors applying the formula (3.1).

4 Scale Factors

Scale factors are extracted in kinematic regions (or bins) of the transverse momentum and pseudorapidity. There are two types of scale factors: cut-based or MVA-based; in this project, we will use the latter. Additionally, in the case of the MVA-based scale factors, a discriminant known as MVA (with a value range from -1 to 1) is used to set regions or *working points* (WPs) for every which scale factors are extracted. The MVA discriminant is the output of a neural network used to identify whether a lepton is a real lepton or a false lepton; the more confident the neural network is that the input is a real lepton, the closer the MVA value gets to one.

In this case, we aim to produce "shape-based" SFs, this means that the values extracted for the WPs will be interpolated to obtain a function that given the value fo p_T, η and MVA will output the value of the corresponding SF. These continuous SFs allow us to use the MVA values for physical analysis; we can only use the values once they have been corrected using the SFs.

5 The framework: *spark_tnp*

This is a framework that is currently being used for extracting the values of the SFs for muons using the Tag And Probe method, it has never been used for electrons before.

The framework uses very few commands. In the scope of this project, the use of this framework can be summed up in four steps:

1. **Flatten:** This is the first step. It will allow us to convert the histograms saved in the parquet files to a format that can be used in the next step.

2. **Fit (I):** In this step, the framework takes the flattened histograms and proceeds to fit the passing and failing pairs. This step is run in a computation cluster.

3. **Fit (II):** Once all the jobs sent to the cluster in the previous step have finished, we run a command to recover the output and save it locally.

4. **Prepare:** In this final step, the framework produces a series of graphs plotting the efficiencies, the scale factors and their uncertainties.

We need to run these four steps for every working point, but the bins for the kinematic regions are set in a configuration file so the SFs are calculated for all of them in the given WP. The higher the number of kinematic regions, the more time the framework will take to finish the process.

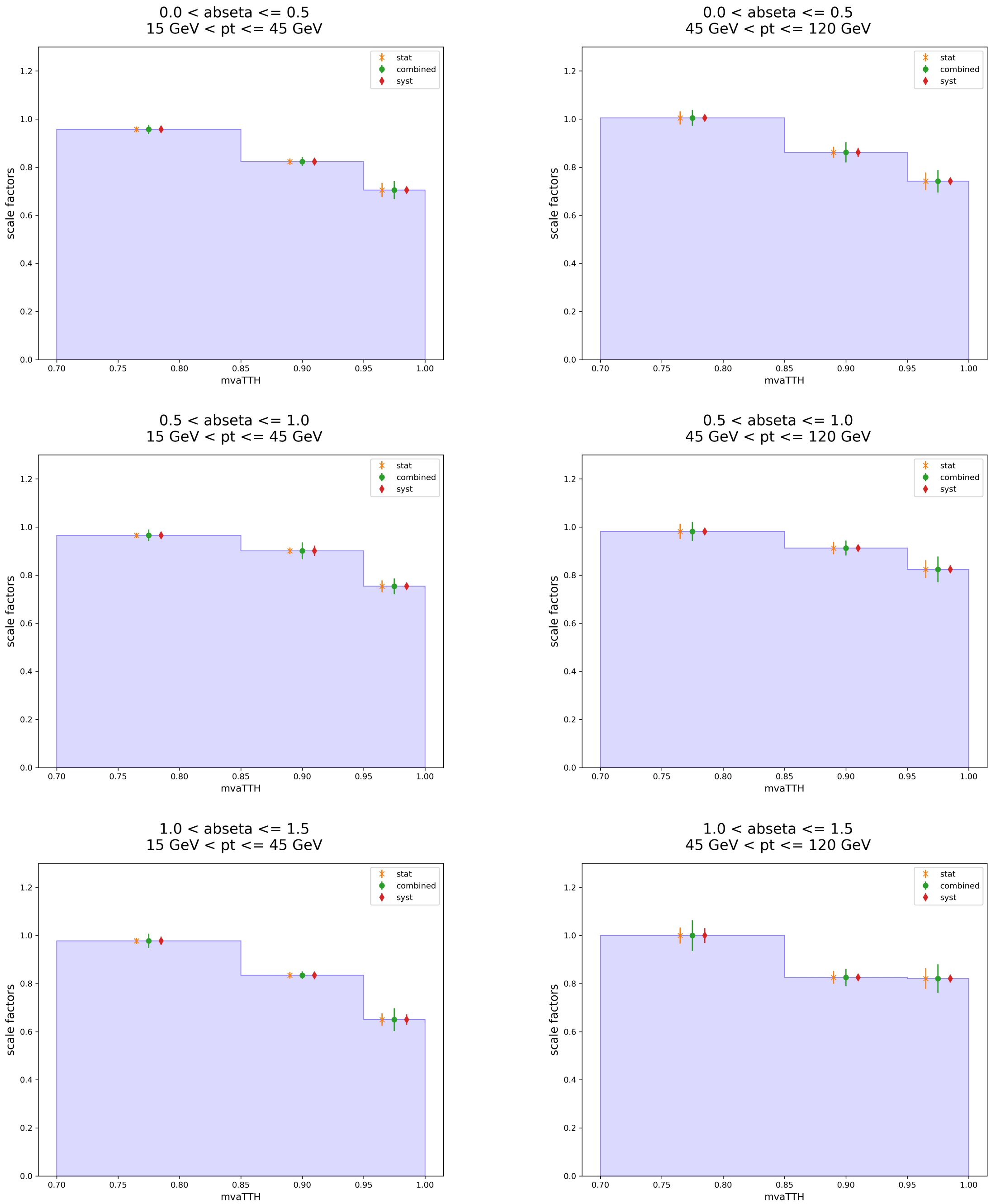
6 Results

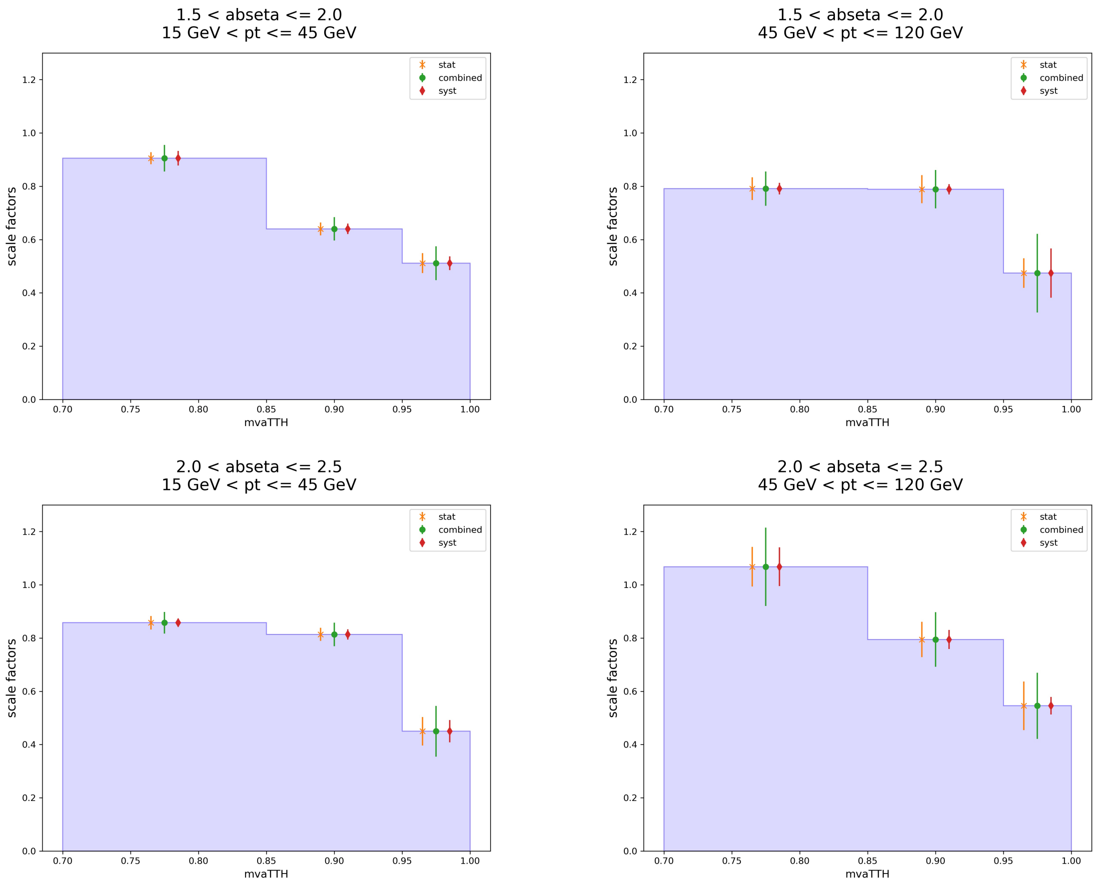
6.1 Scale Factors for electrons

The *spark_tnp* framework has proved to work excellently with electrons, being able to produce the SFs with no problem whatsoever. The figures containing the scale factor values and their uncertainties for every working point and kinematic region combination can be seen in the Appendix A Electron SFs.

7 Summary and outlook

A Electron SFs




1.5 < abseta <= 2.0
15 GeV < pt <= 45 GeV
1.5 < abseta <= 2.0
45 GeV < pt <= 120 GeV
2.0 < abseta <= 2.5
15 GeV < pt <= 45 GeV
2.0 < abseta <= 2.5
45 GeV < pt <= 120 GeV
stat
combined
syst
scale factors
mvaTTH