The Compact Muon Solenoid Experiment
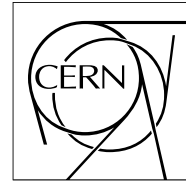
CMS Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

Lepton energy scale and resolution corrections based on an analytical likelihood approach

Fabrice Couderc, Paul Gaigne, Ozgur Sahin

Abstract

In this document, we describe a new approach to obtain the energy scale and resolution correction based on the minimization of an analytical likelihood. The energy resolution effect is incorporated with an analytical approach. This allows to carry out the minimization using the Tensorflow library and in particular its automatic differentiation capabilities. We review the description of the analytical likelihood, the validation of the tool, the potential biases and solutions when the scale and resolution are measured versus the energy itself.

In this note, we describe a novel approach to extract Data/MC relative lepton energy scale and resolution corrections based on Drell-Yann (DY) $Z \to \ell\ell$ production. The method is based on the maximization of a likelihood sensitive to the energy scale and resolution which is totally analytical with respect to the resolution and thus does no require any random number trials to predict smeared di-lepton mass distribution. Thanks to this, the likelihood is differentiable and its maximization is based on the automatic gradient capabilities of the TensorFlow library [1]. This method is implemented in a software tool named IJAZZ2.0 (I Just AnalyZe the Z).

1 Photon energy scale & smearing method

The aim of IJAZZ2.0 is to measure the differences between MC simulation and data in term of lepton energy response and resolution. We note $r_\ell(\vec{X})$ and $\sigma_\ell(\vec{X})$, respectively the data/MC relative energy scale and data/MC energy smearing, where $\vec{X}$ is a list of photon input variables (*i.e.* R_9, η_{SC}, E_T, ...). To be more specific, in this method, we correct the energy from the simulation to match the one in data. Therefore, $r_\ell(\vec{X})$ is a correction to be applied to the lepton energy:

$$E_{\text{corrS}}^{\text{mc}} = r_\ell(\vec{X}) \times E_{\text{raw}}^{\text{mc}} \tag{1}$$

where $E_{\text{corrS}}^{\text{mc}}$ is the corrected-scale energy while $E_{\text{raw}}^{\text{mc}}$ is the original lepton energy. This correction brings the lepton-energy scale in the simulation to its corresponding level in the data. Conversely, one can correct afterward the lepton energy in the data back to the simulation level (which is usually properly tuned) with the formula:

$$E_{\text{corr}}^{\text{data}} = E_{\text{raw}}^{\text{data}} / r_\ell(\vec{X}). \tag{2}$$

Concerning the resolution, it is assumed that the lepton-energy resolution is always better (*i.e.* smaller) in the simulation, thus the energy resolution in the simulation needs to be degraded (smeared) to its corresponding level in data. Because the simulation already includes most of the effects due to the detector response (energy loss, material...), it is assumed that the modest degradation due to the imperfect modelling of the simulation follows a normal distribution. The energy smearing in the simulation is done by random number trials from a normal distribution.

$$E_{\text{corr2}}^{\text{mc}} = E_{\text{raw}}^{\text{mc}} \times \mathcal{N}(1, \sigma_\ell(\vec{X})), \tag{3}$$

where $E_{\text{corr2}}^{\text{mc}}$ is the correted-resolution energy of the lepton. As a summary the lepton energy in the simulation follows the probabilistic law:

$$E_{\text{corr}}^{\text{mc}} = E_{\text{raw}}^{\text{mc}} \times r_\ell(\vec{X}) \times \mathcal{N}(1, \sigma_\ell(\vec{X})) \tag{4}$$

1.1 Definition of the lepton categories and DY regions

In order to extract the values of $r_\ell(\vec{X})$ and $\sigma_\ell(\vec{X})$, we first bin the $\vec{X}$ variables, each bin will be named b. The response and smearing will be measured in each of these N_B bins (we consider here a uni-dimensional binning even if $\vec{X}$ can be multidmensional). Thus, it boils down to measuring N_B scale and smearing parameters: $\vec{r}_\ell$ and $\vec{\sigma}_\ell$. Since the Z boson decays to 2 leptons, we need to adjust the di-lepton invariant mass scale and smearing that will be noted respectively $r_{\ell\ell}$ and $\sigma_{\ell\ell}$. The values of $r_{\ell\ell}$ and $\sigma_{\ell\ell}$ are infered from the values of r_ℓ and σ_ℓ of the 2 leptons forming the Z-boson candidate. We therefore categorize the DY events in N_C categories corresponding to all possible couples (b_1, b_2) for the 2 leptons, b_i being the scale-response bin of lepton i in the event. Because the couple (b_2, b_1) are in the same category as the couple (b_2, b_1) (the $r_{\ell\ell}$ and $\sigma_{\ell\ell}$ parameters are identical), the total number of DY categories

is thus $N_C \equiv \frac{N_B \times (N_B+1)}{2}$. As a consequence, the per-event DY parameter $r_{\ell\ell c}$ and $\sigma_{\ell\ell c}$ in each category $c = (b_1, b_2)$ are given by:

$$r_{\ell\ell c} = \sqrt{r_{\ell b_1} \times r_{\ell b_2}}$$
$$\sigma_{\ell\ell c} = 0.5 * \sqrt{\sigma_{\ell b_1}^2 + \sigma_{\ell b_2}^2} \tag{5}$$

1.2 Definition of the likelihood

In each DY category c, the values $r_{\ell\ell c}$ and $\sigma_{\ell\ell c}$ can be measured by comparing the expected smeared and scaled di-lepton mass distribution in the MC with the corresponding on in data. This is done by binning the di-lepton mass distribution with N_I bins, note that N_I can depend on the category c (for sake of simplicity we will just use the index I). This binning will be referred to as b_I in the following. The comparison is done by assuming that the predicted distribution from the scaled and smeared MC follows a multinomial law, therefore in each di-lepton mass bin i of a given category c, we use the multinomial-law based likelihood.

$$\mathcal{L}(n_{ic}; r_{\ell b}, \sigma_{\ell b}) = \prod_{c=0}^{N_C} \prod_{i=0}^{N_I} \frac{(\sum_i n_{ic})!}{\prod_i n_{ic}!} \, p_{ic}^{n_{ic}} \,. \tag{6}$$

where p_{ic} is the expected probability (depending on $r_{\ell\ell c}$ and $\sigma_{\ell\ell c}$ and consequently on $r_{\ell b}$ and $\sigma_{\ell b}$) for a di-lepton event to fall in bin i when belonging to the category c and n_{ic} is the observed number of events in data corresponding to the same bin i. By definition we have $\sum_i p_{ic} = 1$. The likelihood from Eq. 6 can be maximized with respect to the parameters $r_{\ell b}$ and $\sigma_{\ell b}$. In practice, we minimize the negative log-likelihood, defined by:

$$nll(n_{ic}; r_{\ell b}, \sigma_{\ell b}) = -\sum_{c=0}^{N_C} \sum_{i=0}^{N_I} n_{ic} \log{(p_{ic})}. \tag{7}$$

1.3 Analytical scaling and smearing

As a consequence of Eqs. 4 and 5, each di-lepton event with mass $m_{\ell\ell}{}^{\text{mc}}$ generates a probabilistic scaled and smeared mass distribution $M_{\ell\ell}^{pred}$ according to:

$$M_{\ell\ell}(r_{\ell\ell}, \sigma_{\ell\ell}) = m_{\ell\ell}{}^{\text{mc}} \times r_{\ell\ell} \times \mathcal{N}(1, \sigma_{\ell\ell}) \tag{8}$$

Because we compare the distribution of $M_{\ell\ell}(r_{\ell\ell}, \sigma_{\ell\ell})$ to the data in a binned distribution, we can therefore predict for each event in the simulation its probability to fall in bin i of the di-lepton invariant mass.

$$M_i(r_{\ell\ell}, \sigma_{\ell\ell}) = \frac{1}{2} \left[\text{Erf}\left(\frac{b_i^{\text{u}}/r_{\ell\ell} - m_{\ell\ell}{}^{\text{mc}}}{\sqrt{2}\,\sigma_{\ell\ell}\,m_{\ell\ell}{}^{\text{mc}}} \right) - \text{Erf}\left(\frac{b_i^{\text{d}}/r_{\ell\ell} - m_{\ell\ell}{}^{\text{mc}}}{\sqrt{2}\,\sigma_{\ell\ell}\,m_{\ell\ell}{}^{\text{mc}}} \right) \right] \tag{9}$$

where Erf is the error function, b_i^{u} and b_i^{d} are respectively the upper and lower bound of bin i of the di-lepton invariant mass distribution ($i \in [0, N_I]$). Finally, the invariant mass distribution in the simulation is binned with a very fine grain, this considerably fasten the computation of the probability p_{ic}, In order to avoid any bias, this binning, named b^{mc}, has to be smaller than the expected $\sigma_{\ell\ell}$, in practive a bin width below 0.2 GeV should be considered. The numbering of b^{mc} will be denoted as j and the number of events in the simulation falling in bin j of category c is noted m_{jc}. Considering Eq. 9, we can use the tensor α_{ijc} to predict the contribution from bin

j of the MC simulation to the bin i used in the computation of the aforementioned likelihood. The tensor α_{ijc} is given by:

$$\alpha_{ijc}(r_{\ell\ell c}, \sigma_{\ell\ell c}) = \frac{1}{2}\left[\mathrm{Erf}\left(\frac{b_i^{\mathrm{u}}/r_{\ell\ell c} - b_j^{\mathrm{mc}}}{\sqrt{2}\,\sigma_{\ell\ell c}\, b_j^{\mathrm{mc}}}\right) - \mathrm{Erf}\left(\frac{b_i^{\mathrm{d}}/r_{\ell\ell c} - b_j^{\mathrm{mc}}}{\sqrt{2}\,\sigma_{\ell\ell c}\, b_j^{\mathrm{mc}}}\right)\right] \tag{10}$$

where b_j^{mc} is the center of bin j of the binning b^{mc}. Thus, we can predict the p_{ic} probabilities with:

$$p_{ic} = \frac{\sum_j \alpha_{ijc}\, m_{jc}}{\sum_{i,j} \alpha_{ijc}\, m_{jc}}\,. \tag{11}$$

In Eq. 11, one can note the normalisation term at the denominator which corresponds to the fact that the binning b^{mc} can, and should, span a larger di-lepton invariant mass than the binning b_I in order to take into account event migration in and out of b_I due to the scaling and smearing of the simulation.

The correctness of the method is demonstrated on Fig. 1 which compares a Breit-Wigned distribution of 10000 events which is scaled and smeared with $r_{\ell\ell} = 0.98$ and $\sigma_{\ell\ell} = 0.005$. A different number of random trials was generated per simulation event and then it is be compare to the analytical prediction, with a large number of trials $n_{\mathrm{smear}} = 10000$ the analytical prediction and the random one perfectly match.
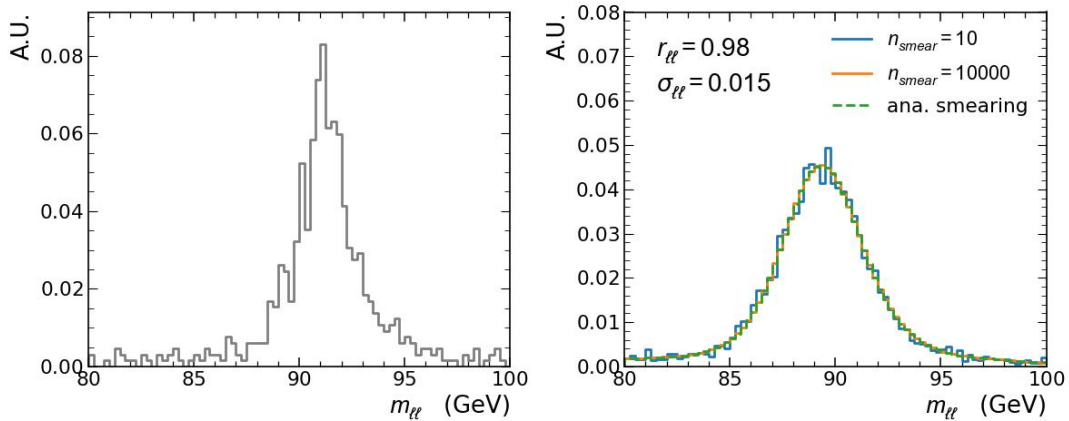

Figure 1: On the left the original MC distribution (Breit-Wigner) from 10000 generated events. On the right, the smeared MC simulation obtained with a random smearing technique, n_{smear} corresponding to the number of trials per original MC event, and with the analytical smearing (dashed line). One can see that the analytical smearing reproduces the expected smeared distribution, approximated by $n_{\mathrm{smear}} = 10000$ (the 2 distributions are indisguishable).

1.4 Qualitative comparison: analytical vs random smearing techniques

In order to compute the p_{ic} probabilities, a random method can be used with a typical number of trial per simulated event of $n_{smear} = 10$, *cf.* the CMS collaboration scale and smearing method [2]. The gain in terms of CPU time (assessed on a laptop with 6 CPU cores) to compute the p_{ic} probabilities between the analytical smearing method presented here with respect to a random smearing method, based on random number generation, ranges from 500 from a small MC sample with a size of 10^6 events to 5000 for a typical MC sample with a size of 10^7 events using $n_{smear} = 10$.

This gain of performance has to convoluted with the automatic-gradient capabilities of Tensor-Flow. With a random smearing technique, the likelihood gradient has to be computed numerically, which requires a typical time of $2 \times N_{\text{par}} \times t_{nll}^{\text{smear}}$ where N_{par} is the number of parameters to be fitted while t_{nll}^{smear} is the time to compute the likelihood with the random smearing technique. For the analytical method, empirically it has been observed that the gradient computation is about $2 \times t_{nll}^{\text{ana}}$ with t_{nll}^{ana} the time to evaluate the nll with the analytical smearing method. Therefore the total gain in term of CPU time for the maximization of the likelihood is of several orders of magnitude on recent laptop. It can be further accelerated using GPUs. Typically, a minimisation with 100 parameters and 20.10^6 evenements (data and MC) ranges from 20-30′ of CPU time on a laptop to less than a minutes on a competitive GPU. Such a minimisation would take several days with a random smearing approch.

1.5 Statistical uncertainties

The uncertainties due to the limited amount of data are computed via the covariance matrix (Σ) defined as the inverse of the Hessian matrix of nll H^{nll}, defined as:

$$\Sigma_{kp}^{-1} \equiv H_{kp}^{\text{nll}} = \frac{\partial^2 nll(n_{ic}; \vec{\theta})}{\partial \theta_k \, \partial \theta_p} \tag{12}$$

where $\vec{\theta}$ generically refers to the list of parameters in the nll function, *i.e.* r_{ℓ_b} and σ_{ℓ_b}. Note that this uncertainty does not take into account the statistical fluctuations related to the limited size of the MC simulation (which might the same size as the data sample in some cases).

1.6 Validation with a naive MC approach

We first validate the method by generating 25×10^6 MC events according to a Cauchy distribution with mean $\mu = m_Z$ and width $\gamma = \Gamma_Z/2$ where m_Z and Γ_Z are respectively the mass and the natural width of the Z boson taken from Ref. [3]. Thus in this first validation, the kinematic of the two leptons is not generated, solely a naive Z-boson line shape according to a classical Breit-Wigner distribution. For each event, we generate 2 random numbers mimicking a property X of the 2 leptons from the Z-boson decay, X is generated according to a uniform distribution other the range $[0, 100]$. We also simulate the energy-resolution of the detector by generating a random number for each event according to a normal distribution with a mean of 1 and a standard deviation of 0.015. For a fraction of this sample, 5×10^6, we further decalibrate and smear each event according to known functions depending on X, this is emulating the difference between the actual detector response (data) and the simulated one (MC). Eventually, we retrieve this decalibration and smearing parameters using the method described in this document assuming as data the latter part of the sample (which includes the additional decalibration), and as MC the rest.

The results of this validation are presented in Fig. 2 which demonstrates that the fitted parameters are in agreement with the injected ones within the statistical precision.

2 Additional specifities fo the method

In this section we will present additional specifities of the method which concern the computation of the uncertainties due to the limited size of the MC simulation, and the tuning of the binning b_I depending on the available statistics in the category c.
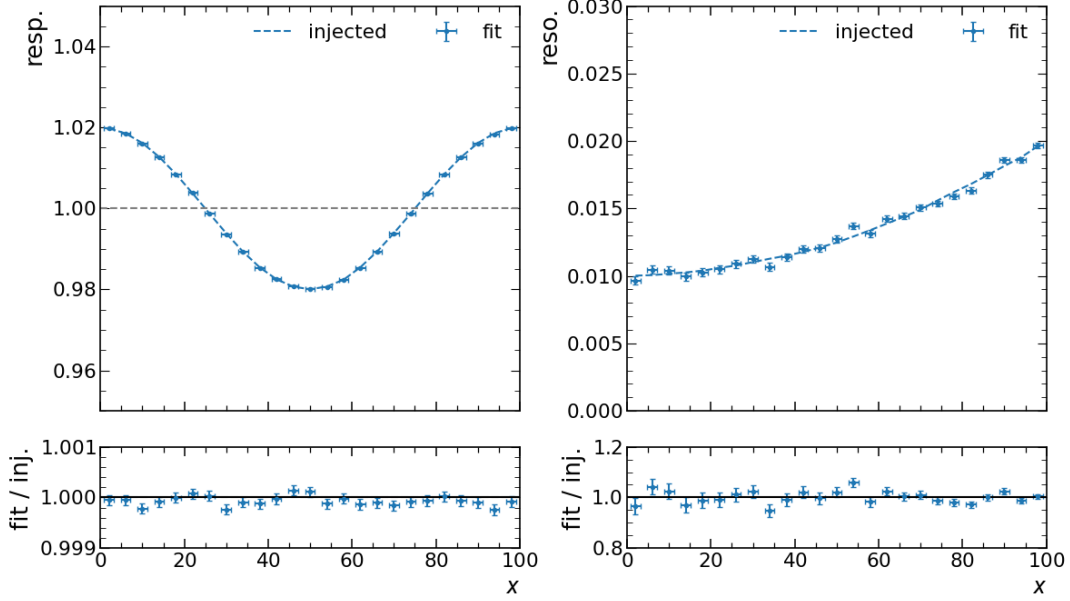
Figure 2: Validation of the method. A naive MC simulation is decalibrated and smeared according to known functions (injected curves). These parameters are properly fitted (points) by the method. The top panels show the absolute parameters while the bottom ones present the ratio of the fitted values to the injected ones. The left (resp. right) plots refer to the response (resp. resolution) parameters r_{ℓ_b} (resp. σ_{ℓ_b}).

2.1   Uncertainties due to the limited statistics of the MC simulation

As mentioned in section 1.5, the covariance matrix computed from the method only takes into account the statistical fluctuations due to limited size of the data sample, assuming that the predicted probabilities p_{ic} are computed with an infinite precision, while they are limited by the statistical size of the simulation sample. To take this effect into account, we can compute the variation of the *nll* minimum related to fluctuations of the m_{jc} numbers. If the m_{jc} varies by δ_{jc}, then the negative log-likelihood becomes nll_δ, that can be computed:

$$nll_\delta(\vec{\theta}, \delta_{jc}) = nll(\vec{\theta}) + \delta nll(\vec{\theta}, \delta_{jc}) \tag{13}$$

where $\vec{\theta}$ collectively denotes the response and resolution parameters of the *nll* function. To first order in δ_{jc}, $\delta nll(\vec{\theta}, \delta_{jc})$ is given by:

$$\delta nll(\vec{\theta}, \delta_{jc}) = \frac{\partial nll}{\partial m_{jc}}(\vec{\theta}) \times \delta_{jc}$$

$$= \left[\sum_i \frac{n_c}{p_c} \alpha_{ijc} - \sum_i \frac{n_{ic}}{p_{ic}} \alpha_{ijc} \right] \times \delta_{jc} , \tag{14}$$

with $n_c \equiv \sum_i n_{ic}$ and $p_c \equiv \sum_i p_{ic}$, the second line is obtained by derivating Eq. 7 with respect to m_{jc}. Therefore to second order, nll_δ can be written as:

$$nll_\delta(\vec{\theta}, \delta_{jc}) = nll(\vec{\theta}^*) + \frac{\partial nll}{\partial \vec{\theta}}(\vec{\theta}^*)\, \delta\vec{\theta} + \frac{1}{2}\, \delta\vec{\theta}^T\, H_{nll}\, \delta\vec{\theta} + \frac{\partial nll}{\partial m_{jc}}(\vec{\theta}^*)\, \delta_{jc}, + \frac{\partial^2 nll}{\partial\vec{\theta}\partial m_{jc}}(\vec{\theta}^*)\, \delta\vec{\theta}\, \delta_{jc} , \tag{15}$$

where $\vec{\theta}^*$ denotes the minimum of the *nll* function and $\delta\vec{\theta} \equiv \vec{\theta} - \vec{\theta}^*$. Deriving Eq. 15 with respect to $\vec{\theta}$, the minimum of the function nll_δ statisties:

$$\frac{\partial nll}{\partial \vec{\theta}}(\vec{\theta}^*) \; + \; H_{nll}\,\delta\vec{\theta} \; + \; \frac{\partial^2 nll}{\partial \vec{\theta}\,\partial m_{jc}}(\vec{\theta}^*)\,\delta_{jc} = 0 \tag{16}$$

Since $\frac{\partial nll}{\partial \vec{\theta}}\left(\vec{\theta}^*\right) = 0$ by definition of $\vec{\theta}^*$, the minimum $\vec{\theta}_{jc} \equiv \vec{\theta}^* + \delta\vec{\theta}_{jc}$ of nll_δ is thus given by:

$$\delta\vec{\theta}_{jc} = H_{nll}^{-1} \times \frac{\partial^2 nll}{\partial \vec{\theta}\,\partial m_{jc}}(\vec{\theta}^*)\delta_{jc} \; . \tag{17}$$

Since all the δ_{jc} fluctuations are independent, one should sum in quadrature to obtain the final uncertainties on $\vec{\theta}^*$ due to the limited MC statistics.

$$\left(\delta\vec{\theta}^*\right)^2 = \sum_{i,c} \left(\delta\vec{\theta}_{jc}\right)^2 = \sum_{i,c} \left[H_{nll}^{-1} \times \frac{\partial^2 nll}{\partial \vec{\theta}\,\partial m_{jc}}(\vec{\theta}^*) \right]^2 \delta_{jc}^2 \; . \tag{18}$$

One can note that for a weighted simulation $m_{jc} = \sum_q w_{jcq}$, where w_{jcq} is the weight of event q and $\delta_{jc} = \sqrt{\sum_q w_{jcq}^2}$. In order to validate this formula, we use the simple simulation described in section 1.6: 2×10^6 events are used as data sample. We measure the scale and smearing parameters for a set of 100 different simulations of 2×10^6 events each and obtained for each parameter a set of 100 measurements. The standard deviation of this set thus represents the statistical uncertainty due to the fluctuations of the m_{jc}'s. We then compare these standard deviations to the predictions of Eq. 18 for each parameter. The results are presented in Fig. 3

2.2 Adaptative binning b_I

Because the predicted p_{ic} appear in the *nll* function in a logarithm, it is mandatory to avoid null values for these probabilities. In addition, because the invariant mass distribution is very peaked, the tail of the distribution may suffer from low statistics potentially biasing the fit (increasing artificially the resolution to fill some empty p_{ic} if there is a data event in bin b_i for category c). To avoid this, we use an adaptative binning and in addition a minimum number of MC events in each category is required (100 events by default), categories not satisfying this condition are discarded. For each category the width of each bin is determined with a quantile technique so that the number of events is the same in each bin b_I, thus the binning b_I depends on the category. The total number of bins is inspired by the Freedman–Diaconis rule and is taken to be the minimum between $(\sum_i n_{ic})^{\frac{1}{3}}$ and $\Delta M_{\ell\ell}/(0.5\text{ GeV})$ with $\Delta M_{\ell\ell}$ the size of the di-lepton mass window over which the fit is performed, typically $\Delta M_{\ell\ell} = (100 - 80)$ GeV, the limit on the total number of bins in b_I comes from the fact that the typical energy-resolution is larger than 0.5 GeV at the Z peak, this default behaviour is configurable.

A demonstration of this technique is presented on Fig. 4. A similar simulation as the one described in section 1.6 is used but we reduce the total number of events to 5×10^4 for both data and simulation and the property X is generated according to a normal distribution so have some categories with lower statistics, the number of scale and smearing parameters is $N_B = 3$, consequently giving raise to $N_C = 6$ categories.
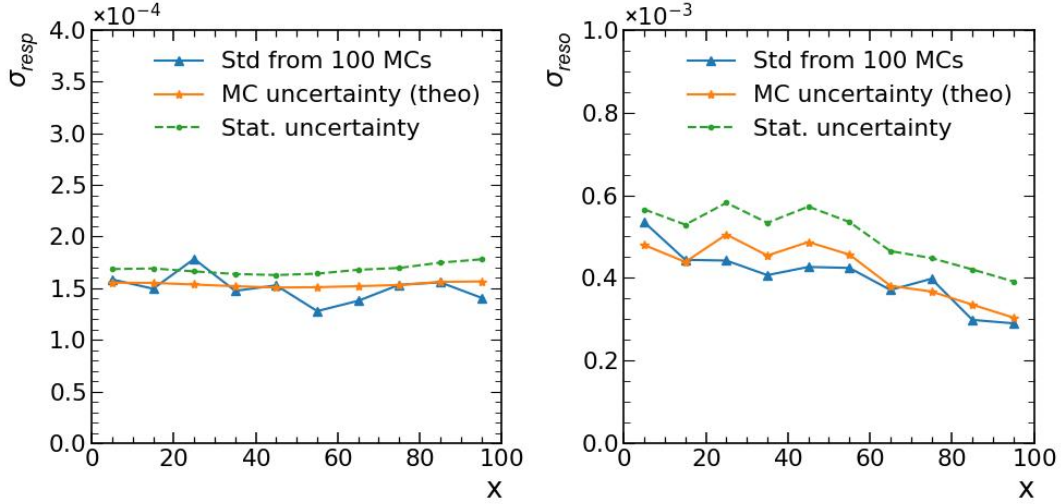
Figure 3: Validation of the statistical uncertainties due to the limited MC statistics for both the response parameters (left) and the smearing parameters (right). The triangles represent the standard deviation of 100 measurements performed with the same data but different simulations, the stars represent the MC uncertainty predictions from Eq. 18, the two are in very good agreement validating the predictions. In addition the statistical uncertainties for each parameters are also presented (dashed line), since the data sample and simulation samples are of the same size, the MC uncertainties and the statistical uncertainties are of the same order.

3 Categorisation vs p_T and validation with a Pythia-based MC simulation

The method presented in this document is designed to properly work when the set of properties $\vec{X}$ used for the lepton categorisation is not affected the measured energy-scale and smearing parameters. When the transverse momentum p_T or the energy E of the leptons enter the categorisation, the measured parameters may suffer from biases due to category migration and/or correlation between the categorisation and the reconstructed di-lepton invariant mass.

Since these biases are strongly correlated with the p_T spectrum of the lepton, we use a realistic DY simulation based on the Pythia MC event generator [4]. For each of two leptons from the Z-boson decay, its energy is smeared a first time according to a normal distribution with mean $\mu_{\rm sim} = 1$ and standard deviation $\sigma_{\rm sim} = 1.5$ %, the event properties in this simulation are referred to as *sim* (e.g. $m_{\ell\ell\rm sim}$, $p_{T\rm sim}$) . A total number of 30×10^6 events are generated this way. Half of this sample is used as reference simulation. For the other half, we introduce an additional decalibration and smearing of the lepton energies (referred to as injected parameters) to emulate the difference between the real detector and the simulated one, . We further proceed in measuring the scale and smearing parameters with a lepton-p_T dependent categorisation. The results are presented on Fig. 5. We observe a small bias introduced in the measured scale parameters (below 0.1 %) and a larger bias of the measured smearing parameters.

In order to comprehend these biases, we select events with one lepton with $45 < p_{T\ell} < 50$ GeV and the other with $50 < p_{T\ell} < 60$ GeV. These events exhibit a double-peak structure reconstructed di-lepton mass $m_{\ell\ell}$ as shown on Fig. 6a, this is due to a threshold effect related to the p_T requirements. In addition, the two peaks have different correlation to the original $m_{\ell\ell\rm sim}$, the peak at the Z-boson mass is not correlated while the secondary peak is strongly correlated. These correlations are presented on Fig. 6b. These effects introduce the biases
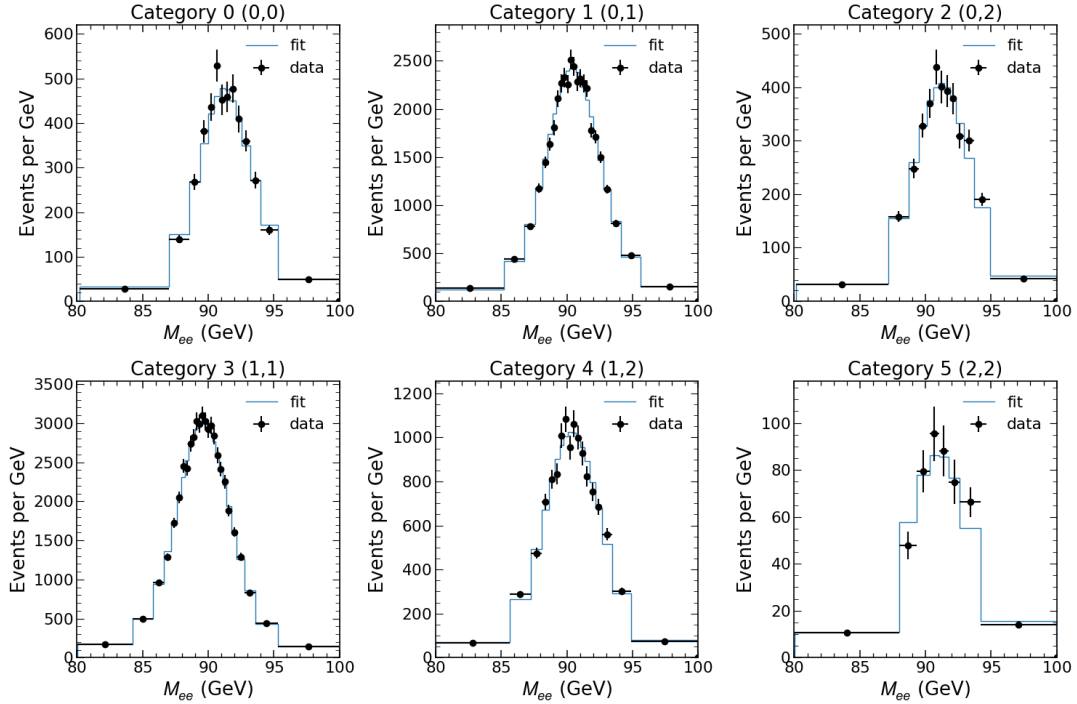
Figure 4: Adaptative binning: The bin width is adapted so that the total number of events in the simulation is the same in each bin, the total number of bins in each category depends on the available MC statistics in this category. The points corresponds to the fitted data (in this case a toy MC dataset) while the histogram refers to the fitted prediction from the smeared simulation.

described above, especially when the two peaks tend to mix (for instance for symmetrical p_T requirements). These biases can be removed by selecting events based on the relative p_T defined as $r_{p_T} = p_{T\ell}/m_{\ell\ell}$. As a comparison to the absolute p_T selection, we apply the relative p_T criteria: one lepton with $45/m_Z < p_{T\ell}/m_{\ell\ell} < 50/m_Z$ GeV and the other with $50/m_Z < p_{T\ell}/m_{\ell\ell} < 60/m_Z$ GeV where m_Z is the Z-boson mass [3]. The di-lepton mass $m_{\ell\ell}$ is shown on Fig. 6c and the correlation to the simulated mass on on Fig. 6d, demostrating that the double-peak structure in the $m_{\ell\ell}$ distribution as well as the related correlations to the simulated mass disappear with the relative p_T selection.

In light of the properties of the relative p_T categorisation, we adopt the following procedure to measure the scale and smearing parameters. The lepton-p_T categorisation is replaced by a relative p_T categorisation using, in place of p_T, the variable $p_T/m_{\ell\ell}$, the relative p_T binning is defined as the original p_T binning ($[p_{T1}, p_{T2}, \ldots, p_{Tn}]$) divided by m_Z. A first fit is performed to extract the energy-scale parameters, and the lepton p_T and di-lepton mass $m_{\ell\ell}$ are corrected. A second fit is performed to measure the energy-smearing parameter, indeed the relative p_T categorisation introduces a correlation between the two leptons and thus the scale must be 1 to properly extract the smearing parameter. Nevertheless, in order to correct the scale and/or smear the lepton energy in the simulation, one needs an absolute lepton p_T criterium. Therefore the relative p_T binning is transformed back into an absolute p_T binning. In each relative p_T bin b, the average absolute p_T, $\bar{p}_{Tb}$, is computed and used to define an absolute p_T binning: $[p_{T1}, 0.5 \times (\bar{p}_{T1} + \bar{p}_{T2}), \ldots, 0.5 \times (\bar{p}_{Tbk} + \bar{p}_{Tk+1}), \ldots, p_{Tn}]$. It is observed that this p_T-binning recasting is very close to the original p_T categories used as input. The results from this procedure are presented on Fig. 7, the x axis is obtained from the recasted absolute p_T binning aforementioned. In addition, the injected parameters presented in the figure, are measured in
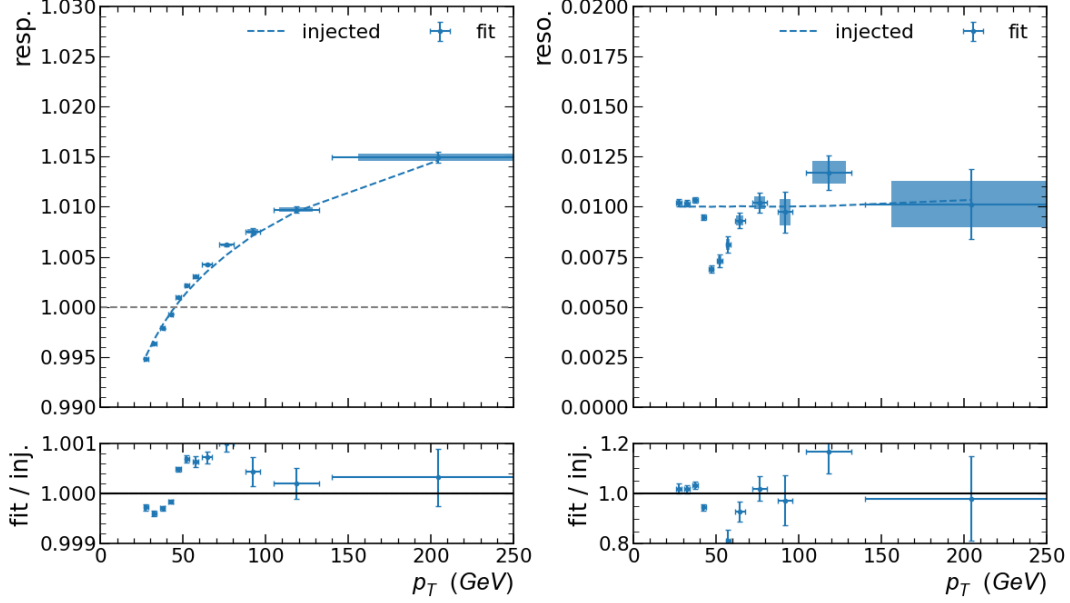
Figure 5: Measured scale (left) and smearing (right) parameters with the Pythia-based DY simulation described in the text. The dashed line corresponds to the injected parameters, the shaded uncertainties represent the contribution due to the limited MC statistics to the total uncertainty. One can observe that a p_T dependent categorisation introduces a bias both in the measured scale and in the measured smearing parameters.

the recasted-p_T categories meaning that they actually correspond to the scale and smearing parameters that would be apply when correcting the energy-scale and/or smearing the energy in the simulation. Note that this 2-steps fit procedure does not improve the smearing parameters in the situation of Fig. 5, since for the absolute p_T binning case the second fit converges to the exact same smearing parameters as the first one (this is expected as there are no correlation introduced between the two electrons as opposite to the relative p_T binning case).

From Fig. 7, one can observe that both the scale and smearing parameters are properly retrieved within the uncertainties, with the exception of the first p_T bin for the scale parameter which is slightly biased, this is due to the absolute p_T cut: $p_T > 25$ GeV.

Conclusion

In this document, we propose a novel method to measure the energy-scale and smearing corrections so that the lepton-energy in the simulation as similar properties as the one in data. The method is based on the minimizarion of negative log-likelihood comparing the di-leptong mass distribution between data and simulation to fit the scale and smearing parameters. The computation of the likelihood relies on an analytical smearing method which allows to use the automatic gradient capabilities of the TensorFlow library which fastens the minimization by several orders of magntitude compare to method such as the one used in Ref. [2]. In addition, because no random numbers are used to compute the likelihood, the likelihood dependence on the scale and smearing parameters are continuous which garanties a robust minimization. Eventually, the method is further adapted to the lepton-p_T categorisation case.
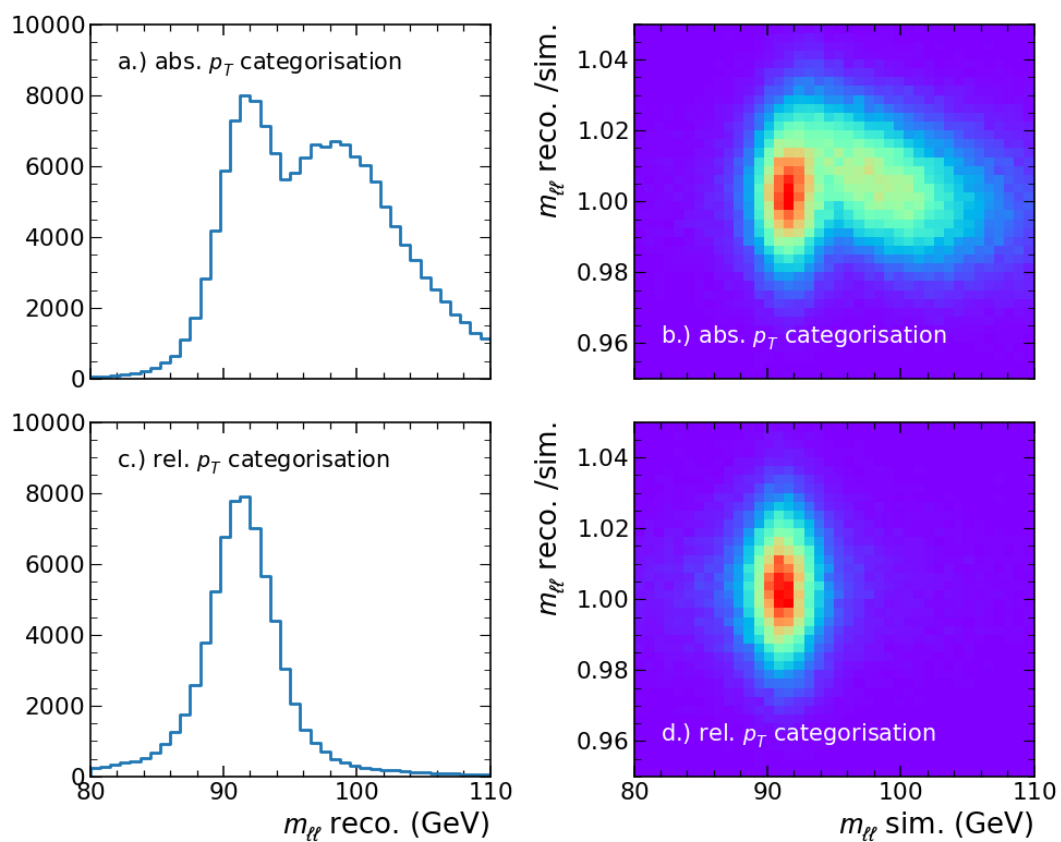
Figure 6: Left: reconstructed di-lepton mass for the absolute p_T categorisation (top) and relative p_T categorisation (bottom). Right: Correlation between the di-lepton mass smearing ($m_{\ell\ell}/m_{\ell\ell\,\text{sim}}$) from the detector level simulation and the simulated di-lepton mass.
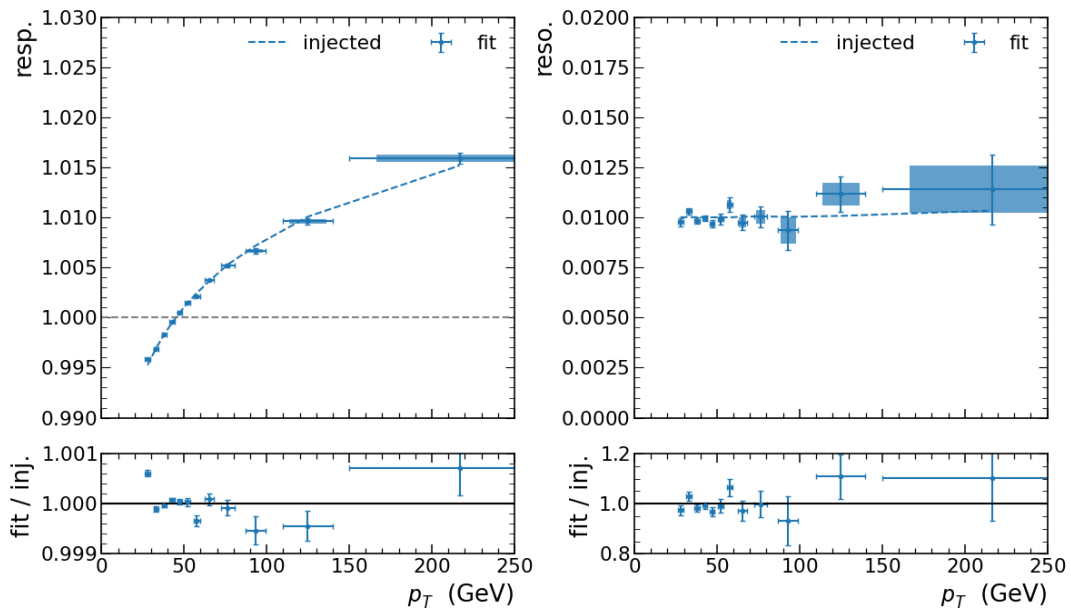
Figure 7: Measured scale (left) and smearing (right) parameters with the Pythia-based DY simulation described in the text using the relative p_T categorisation. The dashed line corresponds to the injected parameters, the shaded uncertainties represent the contribution due to the limited MC statistics to the total uncertainty.

References

[1] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems", 2015. Software available from tensorflow.org. `https://www.tensorflow.org/`.

[2] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *JINST* **16** (2021), no. 05, P05014, `doi:10.1088/1748-0221/16/05/P05014`, `arXiv:2012.06888`.

[3] Particle Data Group Collaboration, "Review of Particle Physics", *PTEP* **2022** (2022) 083C01, `doi:10.1093/ptep/ptac097`.

[4] C. Bierlich et al., "A comprehensive guide to the physics and usage of pythia 8.3", 2022. `https://arxiv.org/abs/2203.11601`.