Performance analysis of electron tracking with the CMS experiment at the LHC

Summer Student Report

Supervisors:
Antonio Tropiano, Giacomo Sguazzoni

Student:
Francesco Peri

20 August 2013

1 Abstract

This report provides a brief description of the tracking algorithm developed by the CMS collaboration for the reconstruction of charged particles in the CMS tracker. In particular it will concentrate on the electron tracking algorithm, using samples of simulated events of particle gun and proton-proton collisions, analyzing the performance and suggesting possible improvements.

2 Introduction

At the LHC design luminosity of $10^{34} cm^{-2} s^{-1}$, the CMS tracker is traversed at each bunch crossing by about 1000 charged particles produced by an average of more than 30 overlapping interactions (pile-up). In this challenging environment the main aim is to obtain a high track efficiency and resolution, and at the same time to keep the fraction of fake tracks low. In addition calculation time must be as low as possible. The goal of this paper is to describe the algorithms used to achieve this and show the performance obtained.

3 CMS structure

CMS (Compact Muon Solenoid) is one of the two biggest experiments in LHC. It is a cylindrical detector of length equal to 21m, with a diameter of 15m and weight of 12500 tons. It is situated in Cessy, France, one of the four interaction points of the beamlines. It is a general purpose detector; it is able to study many aspect of high energy proton-proton collision. It is geometrically divided in a cylindrical central part ("barrel") closed at the two ends by two "endcaps". The inner part of the detector is the silicon tracker. This part is surrounded by the electromagnetic calorimeter (ECAL), followed by the hadronic calorimeter (HCAL). All this material is situated inside a superconductive solenoid magnet which generates a magnetic field of 4 Tesla. Outside the magnet there is the muon revelation system which is placed in the iron yoke that closes the force lines of the magnetic field.

4 The tracker

The tracker is formed by three layers of silicon pixels, in the inner part, and up to ten layers of silicon strips in the outer part. When particles go trough these sensors, they produce a small electric signal that can be revealed. The pixels and strips have different orientations depending on the area they occupy in the tracker in order to reveal the largest number of tracks. The tracker is able to reconstruct the trajectory of electrons, muons, hadrons and can also recognize tracks due to the decay of particles with very low average life such as quark b. The purpose of the tracker is to to trace the trajectory of the particles in the most accurate way, trying not to influence it. Thanks to the presence of a magnetic field coaxial with the detector (approximately of 3.8 T), which deflects the trajectory of the charged particle with radii inversely proportional to their transverse momentum (P_t), it is possible to calculate the momentum of the particle itself.

5 General procedure for tracking reconstruction

The first step of the reconstruction is known as local reconstruction. It consists of the estimating the hits position and uncertainty, measured in the local coordinate frame of the sensor using an orthogonal coordinate system (u, v) in the plane of each sensor. Once that the exact position of the hit has been found, it is possible to continue with the real reconstruction. The CMS tracking software is known as the Combinatorial Track Finder (CTF). The collection of reconstructed tracks is produced by multiple steps (iterations) of the CTF track reconstruction sequence, in a process called iterative tracking. The basic philosophy of iterative tracking is that the early iterations search for the tracks that are easiest to find (high Pt tracks). After each iteration, the collection of hits associated with tracks already found is removed, and the algorithm proceeds with more difficult tracks. Six iterations are used and each iteration is divided in 4 more steps:

- Seed generation: the seed generation initializes the track parameters and uncertainties using only a few hits. In the magnetic field present in the tracker, charged particles follow helices and therefore five parameters are needed to define a starting trajectory. To obtain these five parameters requires at least 3 hits, or 2 hits and a beam constraint.
- Track finding: it is based on a Kalman filter. It extrapolates the seed trajectories, adding hits to the track candidate. The information needed at each layer includes the location and uncertainty of any hits found, as well as the amount of material crossed, which is used to estimate the effects of multiple Coulomb scattering and energy loss. At each layer the χ^2 of the track is recalculated adding the hit found in a layer. Only the tracks whose χ^2 is below a certain configurable value are used.
- Track fitting: it provides the best possible estimation of the parameters of each trajectory. The full information about the trajectory is only available at the last hit of the trajectory and the estimate can be biased by constraints applied during the seeding stage.
- Track selection: it sets quality flags and discards tracks that fail certain criteria. The track finding procedure produces also a significant fraction of fake tracks, where a fake track is defined as one not associated with a charged particle. During this stage quality requirements are applied to reduce the number of fake tracks.

The main differences between the 6 iterations lie in the configuration of the seed generation and final track selection steps.

6 Electron track reconstruction

Electrons, being charged particles, can be reconstructed by the standard tracking. However, due to their low mass, electrons often emit a huge quantity of energy due to bremsstrahlung emission in the detector material. (About 35% of electrons radiate more than 70% of their initial energy before reaching the electromagnetic calorimeter (ECAL) that surrounds the tracker). Bremstrahlung emission is inversely proportional to the mass of the particle. For this reason electron are affected by this loss in a stronger way then the other particles. The energy loss distribution is highly non-Gaussian, which means that it is not well modeled by the standard Kalman filter track reconstruction, which is optimal if all variables have Gaussian uncertainties. As a result, the resolution of the standard tracking is not particularly good for electrons. In consequence CMS opts to reconstruct electron candidates modifying the previous procedure using a particle pre-selection to find electrons. This starts by searching energy deposit in ECAL and matching the tracker seeds with these deposits. Subsequently the algorithm verifies that the ratio between HCAL deposit and ECAL deposit is lower than the 15%. In this way it is sure that the track is not an hadron.

After this procedure a first track reconstruction is made using the Kalman filter, and then another reconstruction is performed using the GSF (Gaussian Sum Filter). This filter approximates the probability distribution of parameters and the energy loss using a gaussian mixture, instead of only 1 gaussian as the KF. In this way the filter can manage in a better way the not-gaussian uncertainties of bremsstrahlung. Another difference between the KF and the GSF is that the χ^2 threshold for new candidates is incremented to 2000 (it is around 100 for the KF). Moreover the GSF uses a seeding method that takes in consideration both the tracker and ECAL seeds, merging them together. In the end, information from GSF are processed by a Boosted Decision Tree which finalizes the pre-selection, finally determining if the track is an electron. Once this pre-selection is ended, parameters are refitted and new hits are added to the tracks using a new version of the GSF (slower and more precise). The process described above could be used for all the particles, but the calculation time necessary for estimating the trajectories in this way would be too high. Therefore is only used for the electrons that are the most problematic particles.

7 Performance of electron tracking

CTF algorithm performances can be analyzed using simulated events. Simulated tracks (SimTracks) are reconstructed using the techniques described in the previous part. In this way it is possible to calculate all the particle parameters(Pt, mass, etc.) from the reconstructed tracks (recoTracks). Simulated tracks and reconstructed tracks are associated to valuate efficiency and fake rate. Generally a SimTrack is considered associated to a recoTrack if at least the 75% of its hits were produced by the SimTracks. The efficiency is defined as the ratio of simulated track that can be associated to a recoTrack. The fake rate is defined as the ratio of reconstructed tracks not associated to any simTrack. In this case a track is considered a fake one if the number of its hits associated to a simulated track is lower than 50%.

$$efficiency = \frac{n^o\, associated\, tracks}{n^o\, simulated\, tracks} \quad (1) \qquad fake\, rate = \frac{n^o\, not\, associated\, tracks}{n^o\, reconstructed\, tracks} \quad (2)$$

As already said the GSF algorithm takes into account both the tracker and the ECAL seeds to increase the total precision. In the Figure1 it is possible to notice how the merged seeding method of GSF algorithm gives the best results in terms of efficiency and fake rate. For this reason the GSF used in the next analysis always uses the merged method.

Despite the best modeling of energy loss of GSF, comparing the two algorithms it is possible to notice how the Gaussian Sum Filter has advantages but disadvantages too. In the Figure2 there is the comparison between efficiency and fake rate of three kind of tracks: red points are the pixel tracks, the blue points are the general tracks (KF) and the green ones the GSF tracks. The event sample used is formed by electron-positron pairs of Pt=10GeV. It is possible to notice that the general tracks are better in terms of efficiency and fake rate. This is because of the more restrictive conditions during the track reconstruction. In this way it is easier to avoid the introduction of fake hits of other particles in the wrong trajectory. Moreover the GSF gets worse in the transition area between the barrel and endcaps. This is due to the fact that the GSF uses ECAL data and the electromagnetic calorimeter has a dead zone in that area (because of cables etc.). Furthermore the GSF algorithm has a worse efficiency for tracks with a low or very high number of hits,due to the little amount of data in the first case and due to the direction of the particle in the second case (it necessarily passes in the transition area if its tracks has a big amount of hits). Finally the fake rate of Gsf is better for high Pt because of the better energy loss modeling.

So which is the advantage of using the GSF? It's the better resolution of parameters. The Figure3 shows the distributions of the recoTrack parameters in comparison with the parameters of the simulated track. This is made by defining the value $p = \frac{x_r - x_s}{\sigma_r}$. x_r and x_s are the values of the reconstructed parameter and the simulated one, instead σ_r is the standard deviation associated to x_r. The perfect distribution of the pulls is a gaussian with 0 as central value and $\sigma = 1$. It is also reported the RMS value for the plots, which has an expectation value of 1. It is simple to notice that the resolution of all the parameters is better using the GSF algorithm. In particular the RMS value indicates that the resolution of the GSF can be also twice better than the KF one.

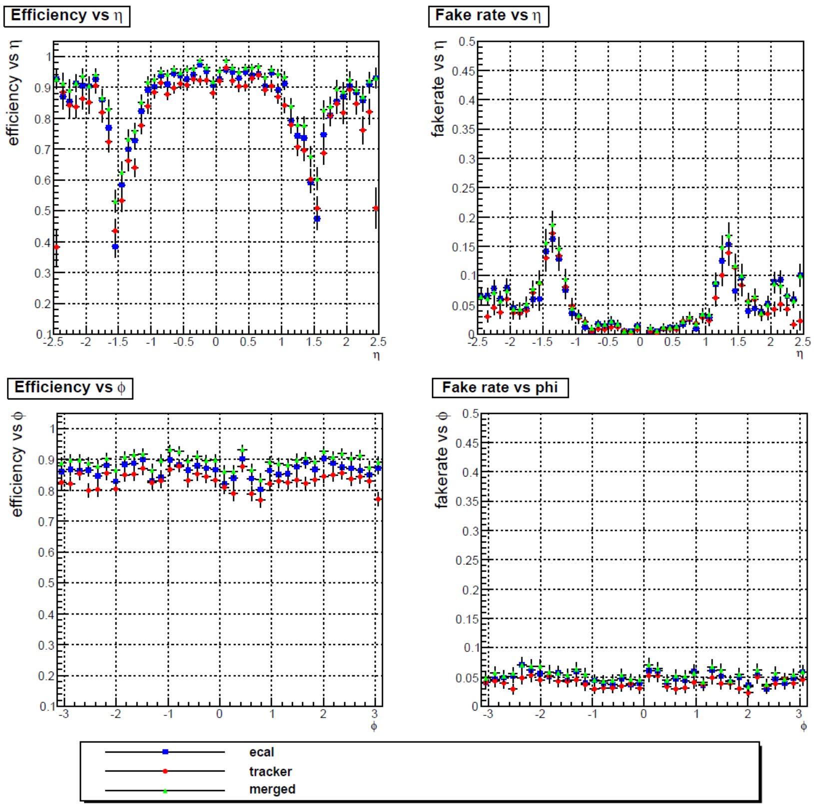

Figure 1: GSF efficiency and fake rate with different seeding methods

8 Ideas for improvements

One of the the most noticeable things is that the GSF has a huge drop in efficiency in the transition area and also a huge fake tracks production in the same zone. A possible method to fix this problems is to use different algorithms for different track's η. In particular it could be possible to use the KF for $1 < |\eta| < 2$ and the GSF in the other areas. To estimate the meaningfulness of this hypothesis it was decided to compare efficiency, fake rate and resolution of the two tracking methods in the transition area (Figure4). It is immediately clear that there is an improvement in terms of efficiency and fake rate. The fraction of associated tracks for the KF is over the 90%, instead of the 60% of the GSF, and also the fake rate is definitely better. The meaning of all this is that the whole recoTracks of GSF algorithm are not sufficient to cover the entire simulated tracks family. Nevertheless, observing the pull plots, it is possible to notice that the resolution of the parameters is still better for the GSF algorithm. For example the RMS and mean parameters for KF are respectively 2.978 e -2.254, instead the GSF parameters are 1.363 and 0.3572. The gain is a factor of 2 for the first parameter and higher for the second. For this reason the idea of using different algorithms is not useful. It is preferable to modify and tune the parameters of the GSF, trying to estimate better values.

For the next analysis the simulated events that have been used are gluon-gluon collisions, with the production of a Z-boson which decays in an electron-positron pair. This samples are more realistic then the single particle's one and they simulate also the pile-up of the events. There are a lot GSF tunable parameters but it was decided to modify only the most relevant of them:

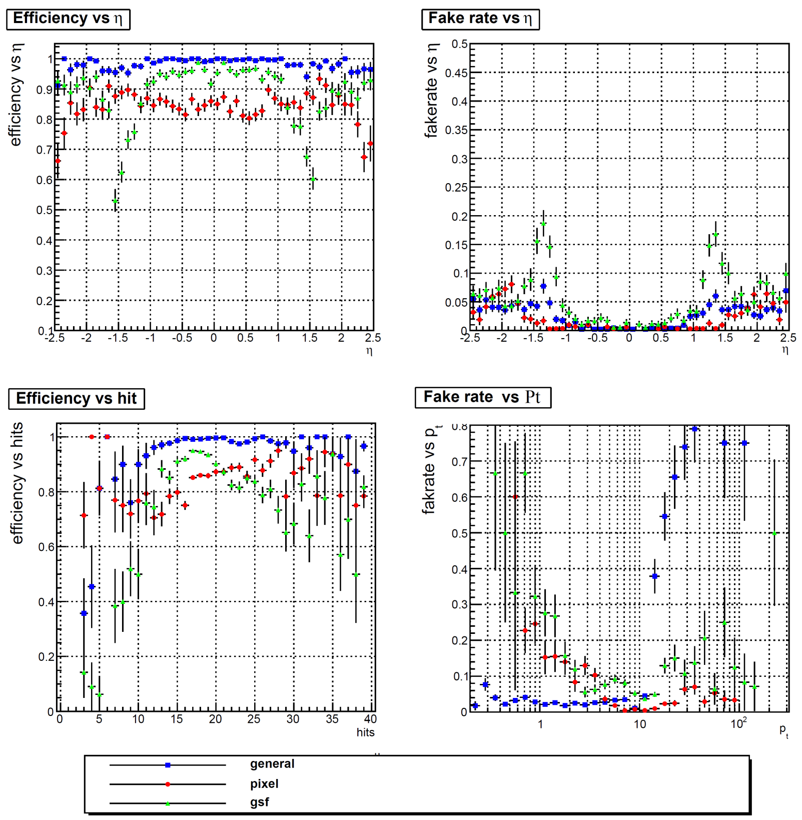

Figure 2: Efficiency and fake rate of electron reconstructed tracks using different algorithms. Pt=10 GeV

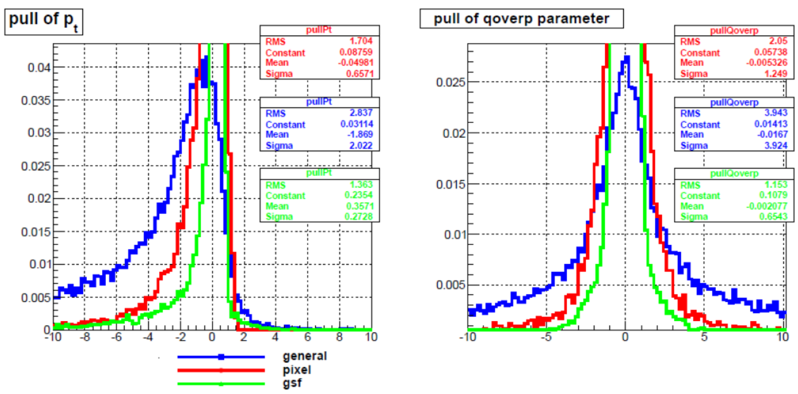

Figure 3: Pulls of electron reconstructed tracks using different algorithms. Pt=10 GeV

- maximum number of candidates: it is the number of hits that are considered "good candidates" to be associated to the tracks at each iteration.
- minimum number of hits: it is the minimum number of hits for a track to be added in the recoTrack family.

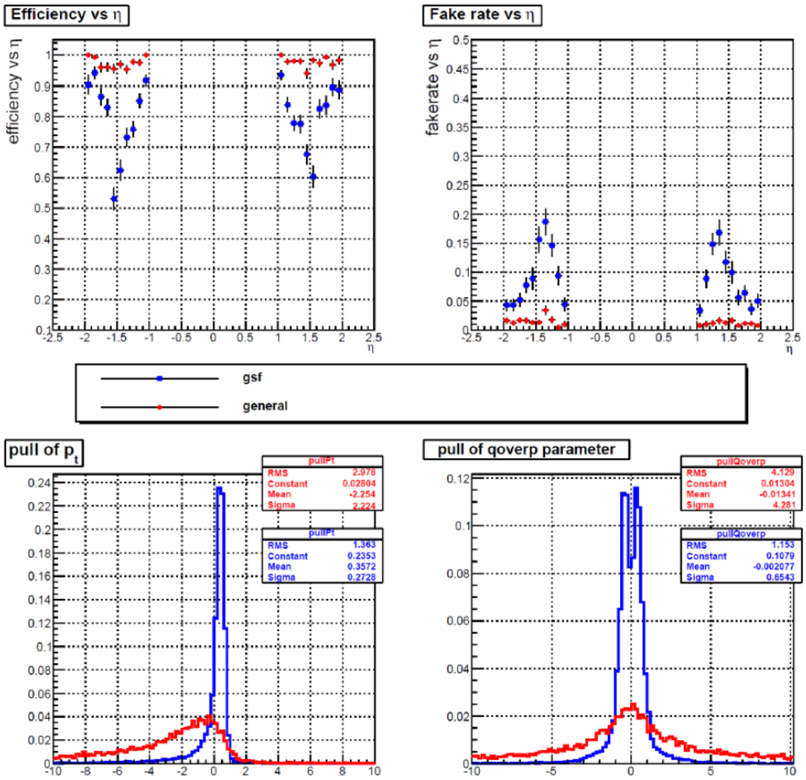

Figure 4: Efficiency, fake rate and pulls of electron recoTracks reconstructed using GSF and KF algorithms in the transition area barrel - endcap $(1 < |\eta| < 2)$. Pt=10 GeV

- χ^2 : it is the χ^2 of the hit position compared to the fitted track. During every iteration of the filter the hits that are added to the tracks must have a χ^2 below a certain tunable value (default 2000)
- maximum number of lost hits: this is the maximum number of layer that the particle can pass through without leaving a hit.

The modification of the minimum number of hits and the maximum number of lost hits does not lead to substantial improvements. The question changes when the other two parameters are modified. If the maximum number of candidates is decreased it is possible to notice an important improvement in terms of fake rate. This is due to the fact that ,decreasing the number of candidates, the chance of adding a wrong hit to the track decreases too. Modifying the χ^2 leads to a similar behavior. When the χ^2 is decreased the acceptance window of new candidates decreases too. As consequence the number of fake tracks is smaller and the efficiency is better. This phenomenon suggests that the estimate of the loss of energy by the electrons is erroneous. The default value, equal to 2000, allows the GSF to consider as candidates a very high number of hits, to assess the possibility that the electron sharply modifies its trajectory. The fact that decreasing this parameter leads to a better efficiency could indicate that on average the radiation emitted is less than what is currently estimated by the GSF.

The Figure5 shows the efficiency and the number of fake rate of electron tracks reconstructed using $\chi^2 = 100$ and the maximum number of candidates equal to 2. There is an overall gain in terms of efficiency of $3 - 4\%$ in the entire tracker and the results in terms of fake rate are still better (also 10%!). Analog improvements are present for plots over Pt and number of hits too.

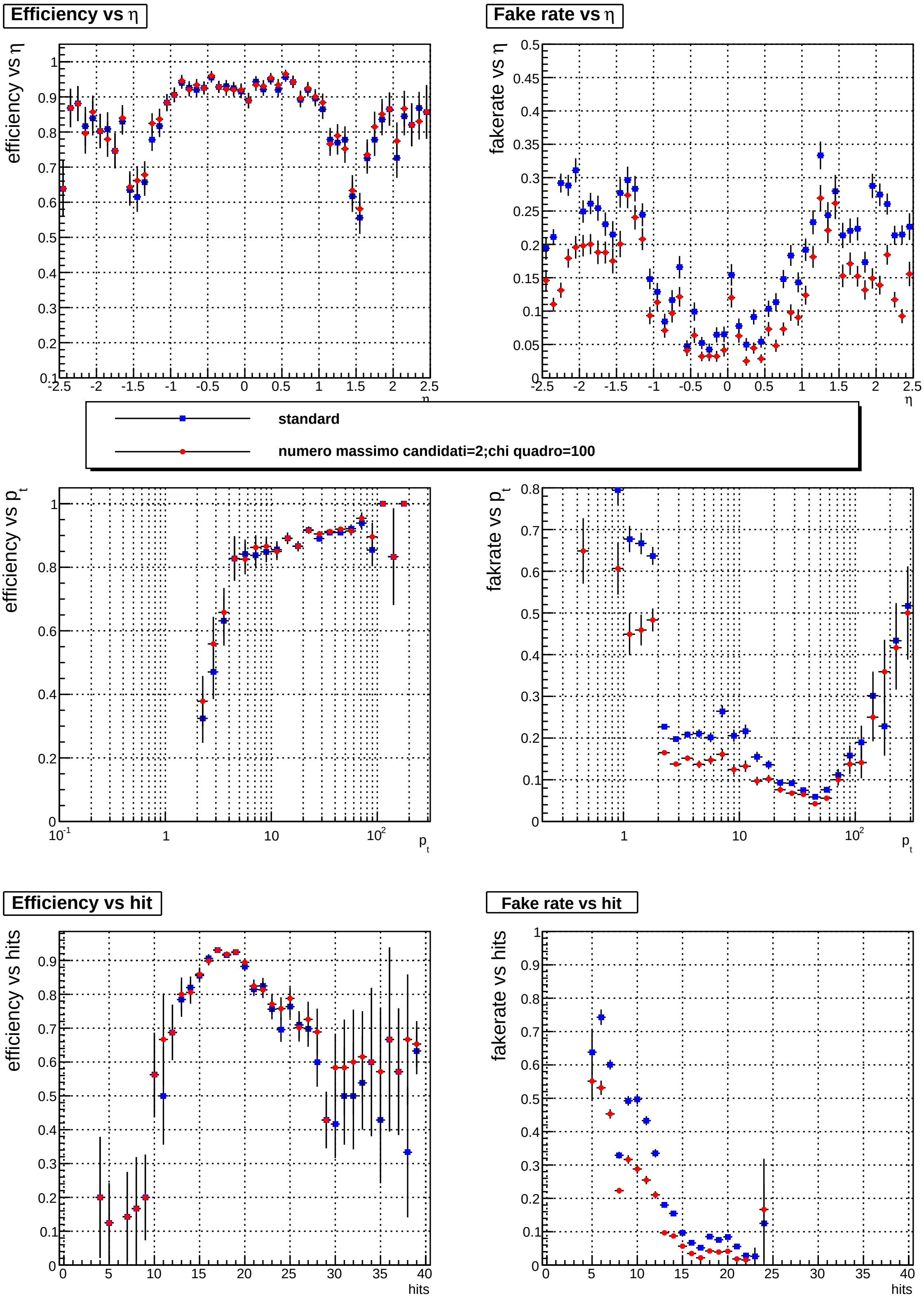

Figure 5: Efficiency and fake rate of electron reconstructed tracks whit modified GSF parameters: χ^2=100 maximum number of candidates=2. $Z- > e^- e^+ < pileup >= 30$

9 A question mark

During the performance analysis it was possible to notice another problem which may lie in the way the GSF algorithm estimates the Pt of tracks. Because of the huge radiation emitted by electrons it seems that the GSF overestimates the energy loss of the particles associating to them an energy greater than necessary.

The first clue of this possible error was highlighted in the previous section: if the χ^2 maximum values is decreased there is an efficiency and fake rate improvement and this is and indication of a potential overestimation. This thesis is supported by another factor: observing the Figure6 it's simple to see that the peak of equivalent gaussian of GSF tracks is drift on the right of 0 while the peak of the general track gaussian is on the left. This last behavior is easier to justify because it is due to the large amount of track which are lost immediately by the filter, and for this reason they are associated to low Pt tracks. The fact that the GSF adds energy is present also for higher Pt an must be investigated. there is a feedback of this problem also in the q/p pull: the plot should be a gaussian centered over the 0. Instead it is possible to see two distinct peaks. They could be produced by an overestimation of Pt both for e^+ and $e-$.

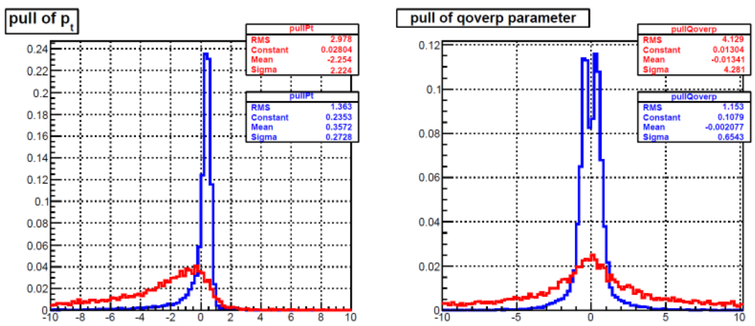

Figure 6: Pulls of electron recoTracks reconstructed using GSF and KF algorithms. Pt=10 GeV

The presence of this problem is noticeable also looking at a plot of reconstructed tracks purity over the Pt. The Figure7 represents this plot. There is an abundance of high Pt tracks (also over 35 GeV) between the ones with purity equal to 1 (perfectly associated tracks). The red line represents the medium Pt of the tracks which should be 10 GeV.

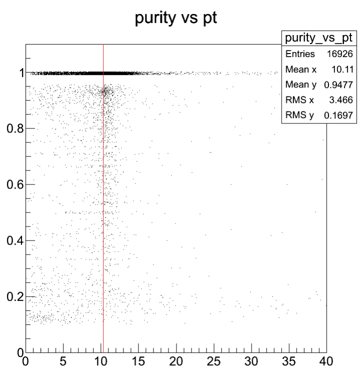

Figure 7: Purity of electron tracks over Pt. Pt=10 GeV