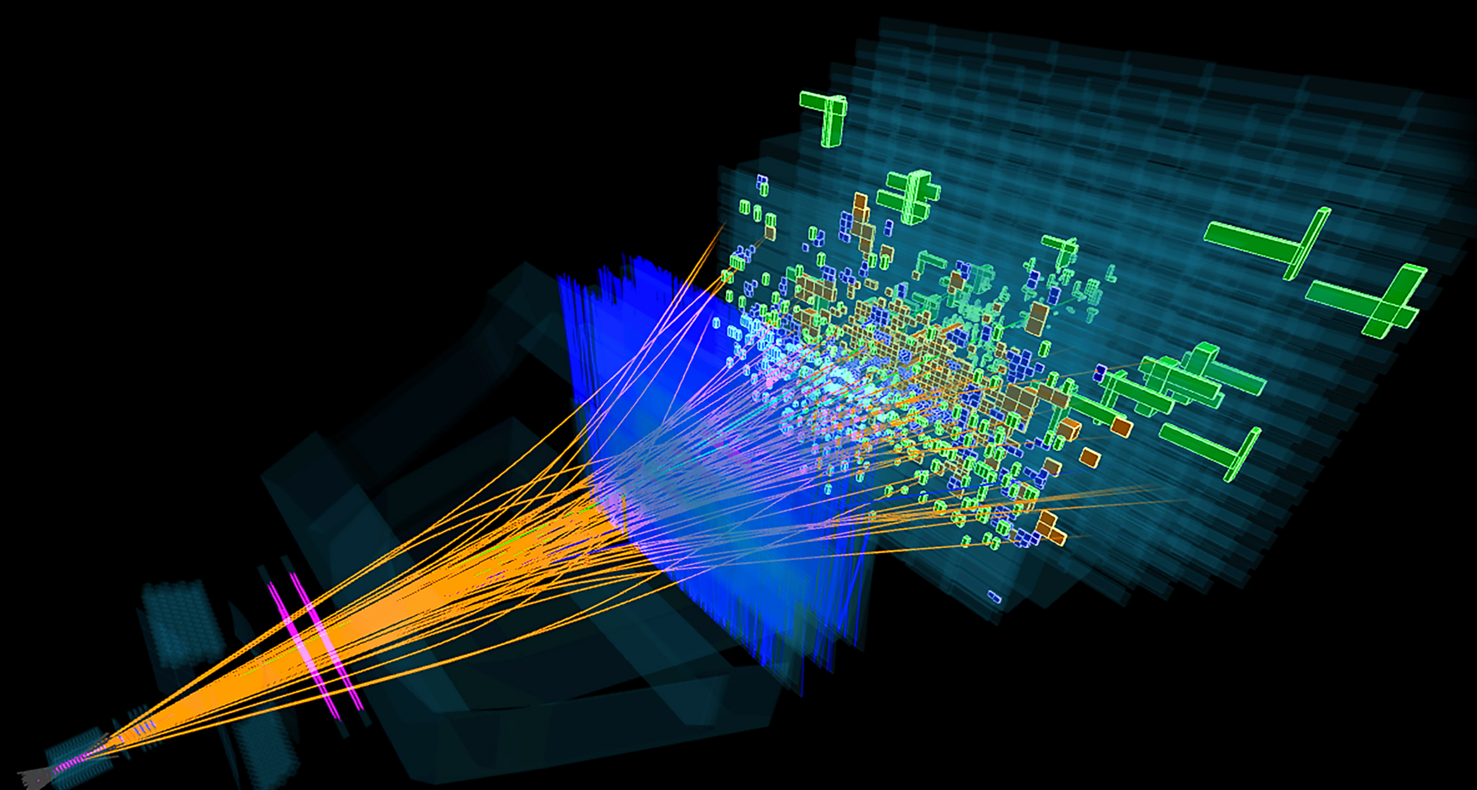
SUMMER PROJECT 2022

Nuclear & Hadron Physics Group



Davide Fontana
Supervised by Dr Derek Glazier
Co-supervisor Dr Mark Whitehead

University of Glasgow

PROJECT OUTLINE AND TIMETABLE

The project aimed at employing the fast simulation algorithm developed in [1] to reproduce LHCb data. The algorithm is written in C++ and Python.

The project lasted for a nominal period of 8 weeks.
Here is an approximate presentation of the planning and timetable.

- Week 1-2: intro to ROOT, familiarising with the software using [2].
- Week 3: plotting LHCb data, comparing datasets and plot resolutions of relevant variables.
- Week 4-5: study and understand the fast simulation, run toy model.
- Week 6-7-8: modify the algorithm so that it accommodates the structure of LHCb data. Run simulation for different variables and changing hyperparameters.

INTRODUCTION

Machine Learning algorithms provide an efficient way to produce fast detector simulations that would otherwise be computationally expensive using traditional Monte Carlo techniques.

Throughout the project the algorithm presented in [1] was used to simulate a $B^0 \to \bar{D}^0\pi^+\pi^-$ decay where $\bar{D}^0$ further decays as $\bar{D}^0 \to K^+\pi^-$.
Fig. 1 shows a scheme of the algorithm.

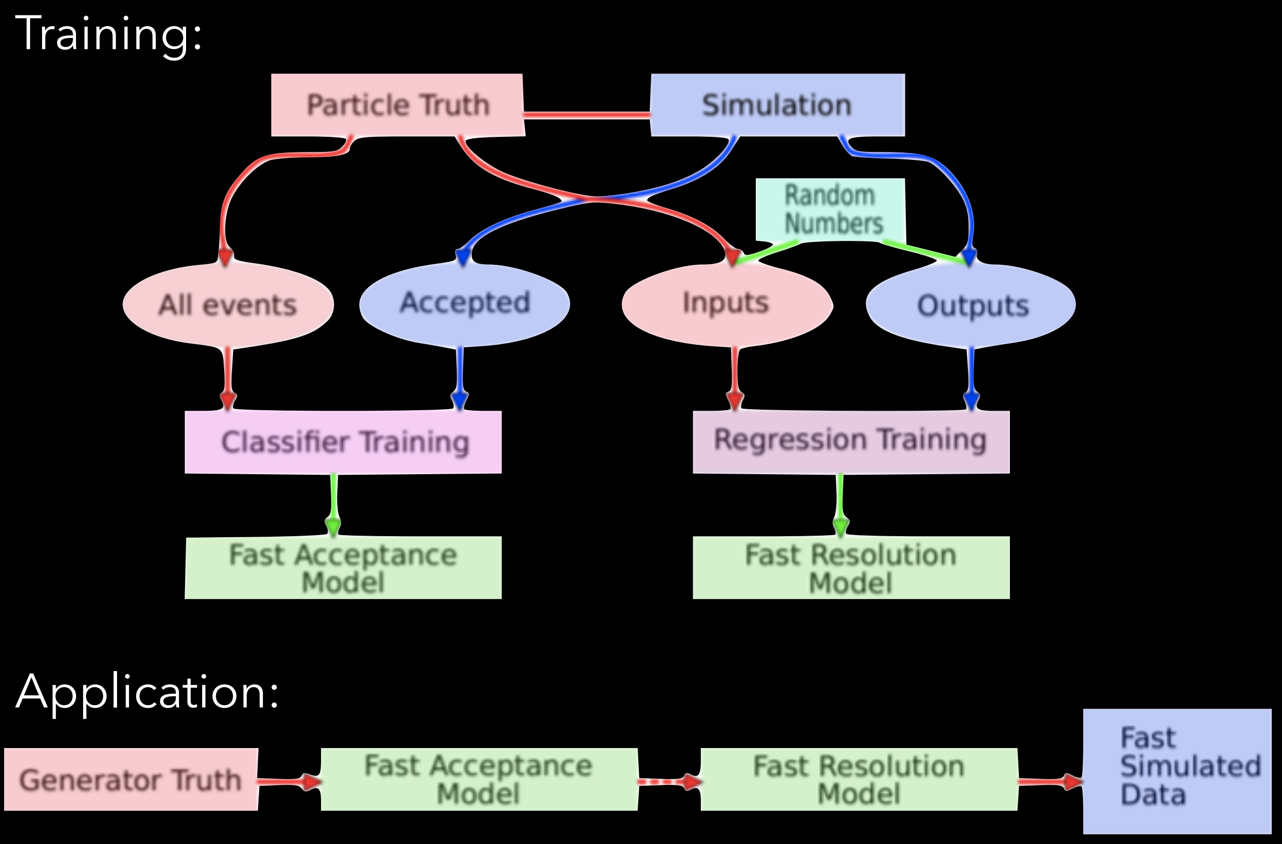


Figure 1: the diagram explains how the training and application stages of the fast simulation operate. [1]

ACCEPTANCE TRAINING

As shown in Fig. 1, the training is divided in two parts, acceptance and resolution.

<u>ACCEPTANCE:</u> Starting from the set of generated data, only particles that are detected and successfully reconstructed are selected.

In Figure 2 the results of acceptance training are showed. The simulation is carried out for variables ϕ , θ and $|\bar{p}|$.

Training datasets of size

- 7E5 for generated data
- 5E5 for reconstructed data

were used.

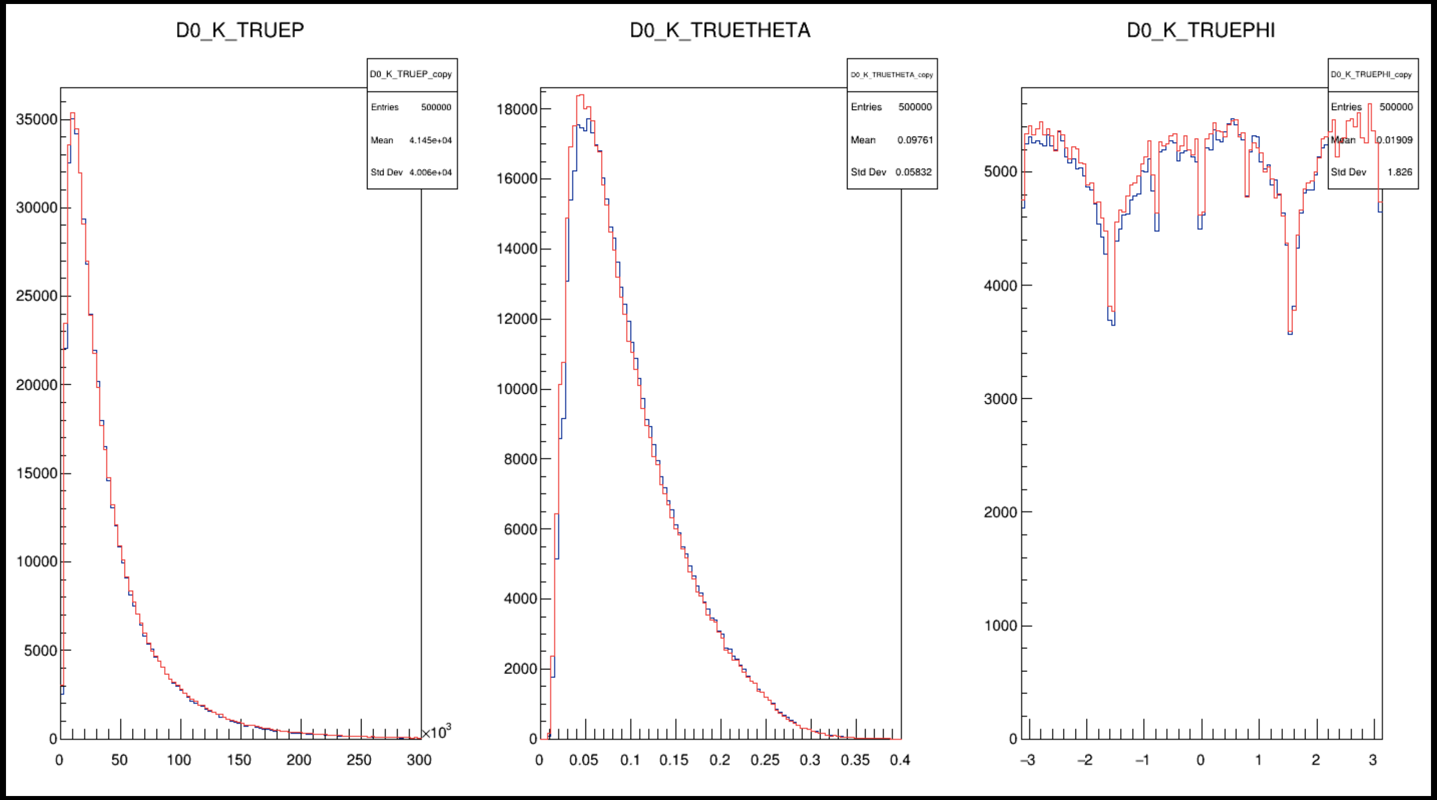


Figure 2: the plots show good results for the acceptance training. The 4 particles were trained separately, here we only show the plots for the kaon. Similar results were obtained for the other 3 particles.

RESOLUTION TRAINING

Once particles that have been detected are selected, detector effects need to be taken into account. To do so, the particle's 4-momentum is distorted. Only then the particles are fully reconstructed. This distortion is quantified by the RESOLUTION.

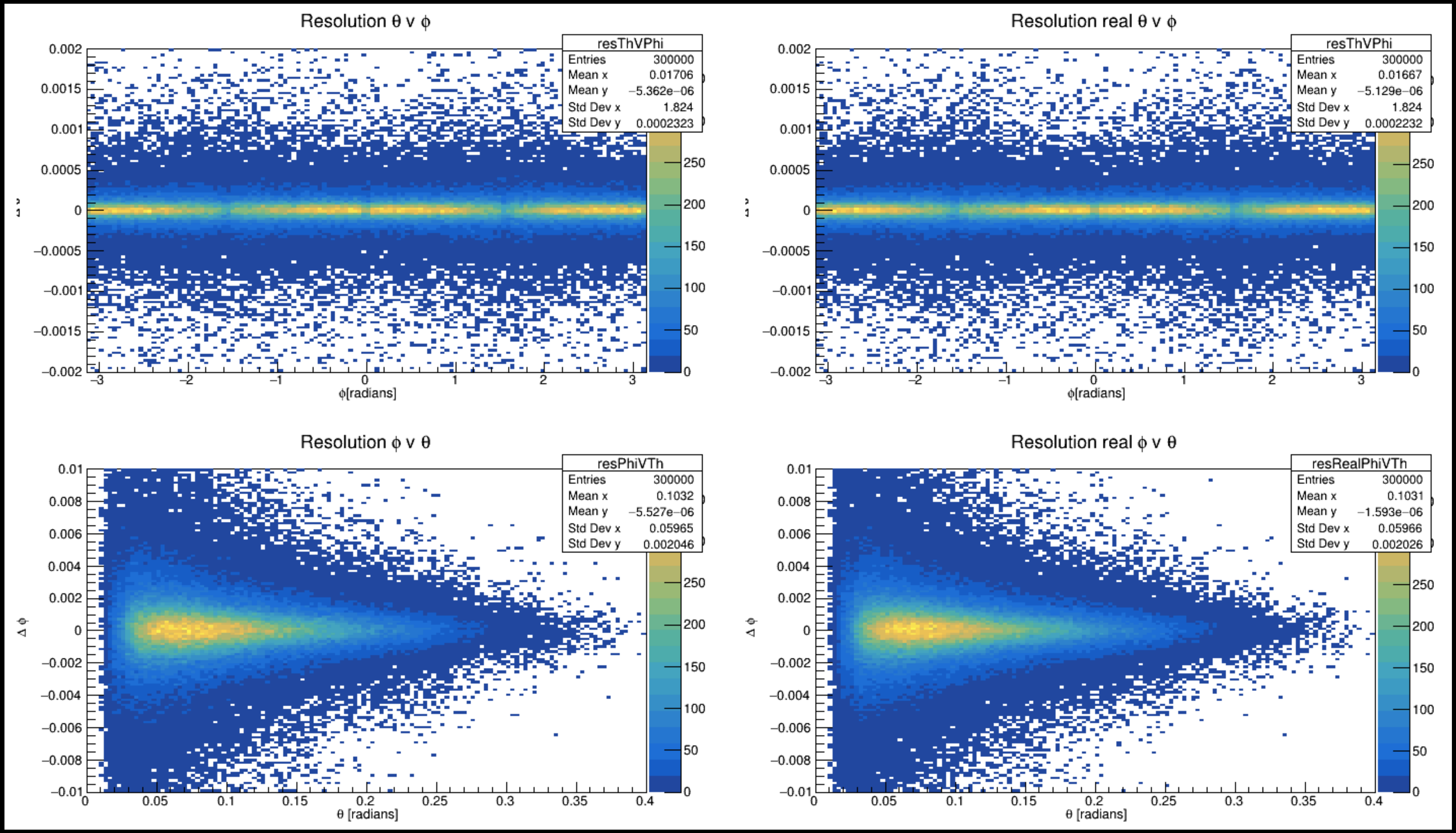
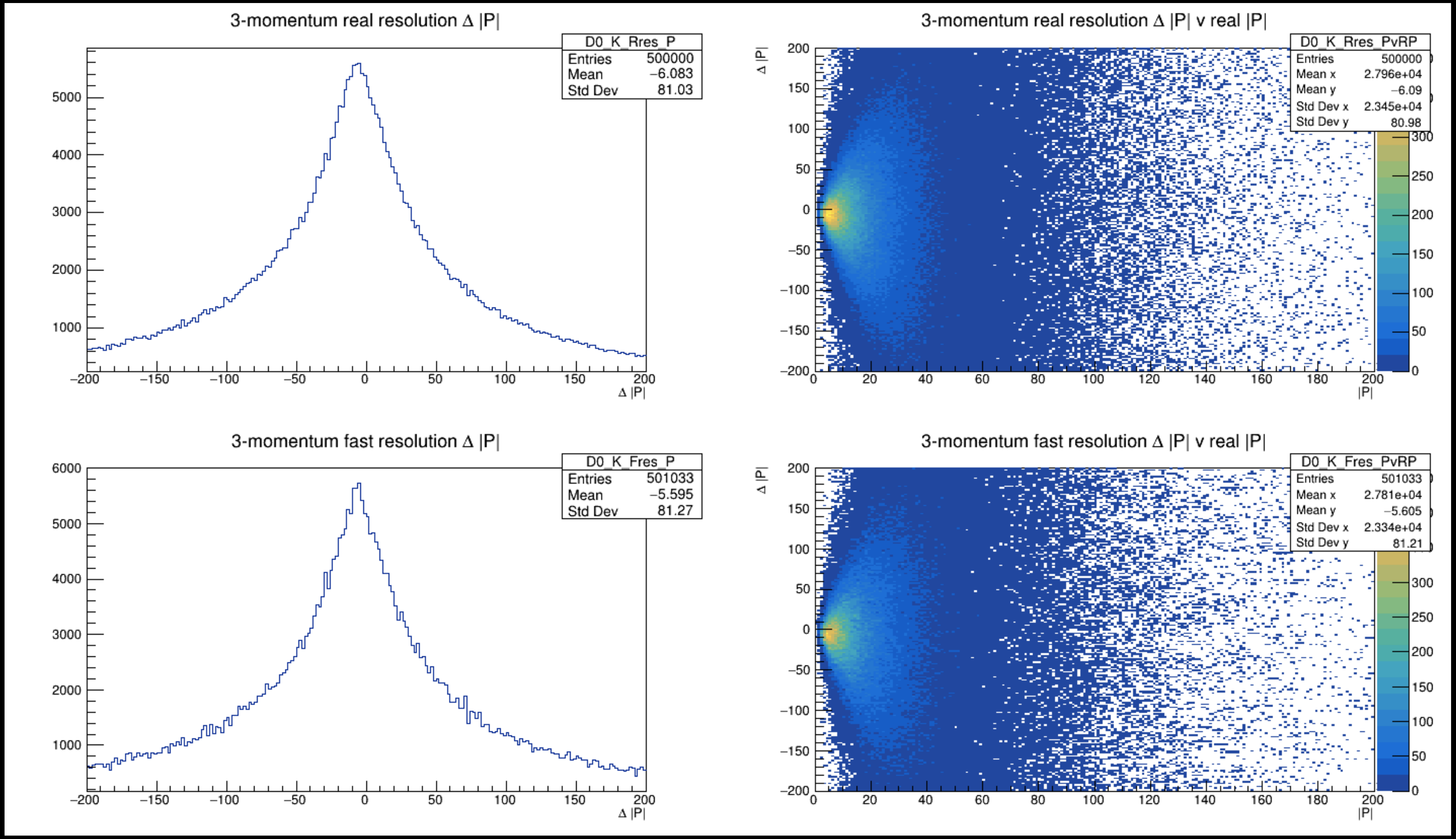


Figure 3: These plots are 2d histograms showing the resolution of θ versus ϕ (upper row) and the resolution of ϕ versus θ (lower row). The data used on the left hand-side are taken from the original dataset, whereas the ones used on the right hand-side are a result of resolution training. Plots show excellent agreement. A sample of 3E5 events was used for the training.

RESULTS

Having trained the algorithm, the simulation can be run for each particle. This process has to be repeated for each particle at present, though there is room to implement the code such that all particles are trained and simulated at the same time. Subsequently, kinematically relevant quantities can be reconstructed and compared to the original dataset.

Figure 4: As an example of the fast simulated reconstructed quantities, here the 3-momentum resolution of the kaon is plotted in 1d (left hand-side) and again versus the 3-momentum itself in 2d on the right hand-side. Note that the plots of the upper row are taken from the original dataset, whereas the lower row shows the fast simulated plots. Once again, they are in good agreement.



RECONSTRUCTED MASS PLOTS

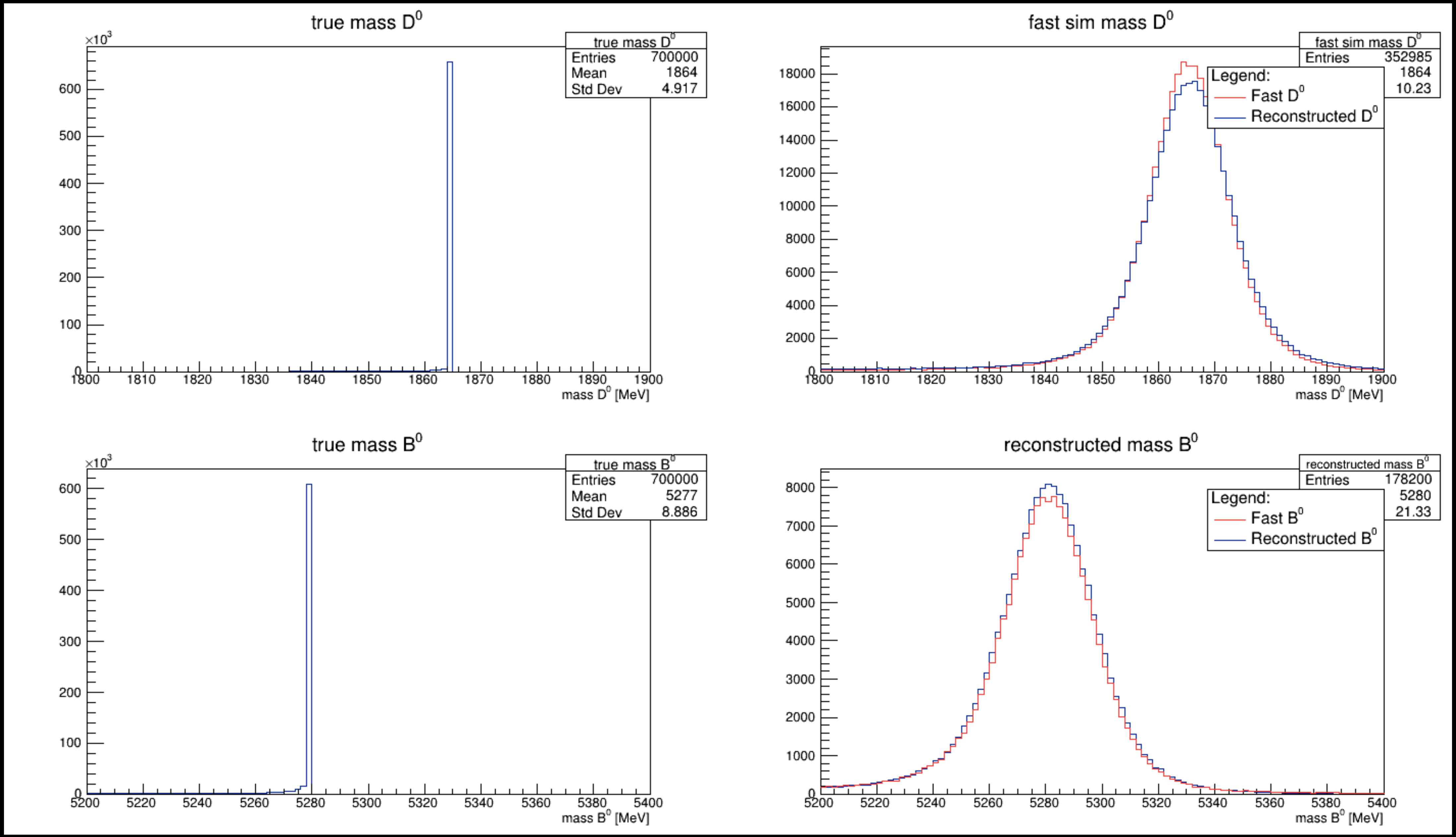


Figure 5: These plots were produced by using the original dataset as well as the data produced by fast simulation. On the left hand-side the masses of $\bar{D}^0$ and B^0 are plotted as from the generated dataset. On the right hand-side, the data from the reconstructed dataset provided for training (in blue) are plotted together with the masses reconstructed from fast simulation results (in red). The plots are normalised and they show good agreement.

DALITZ PLOT

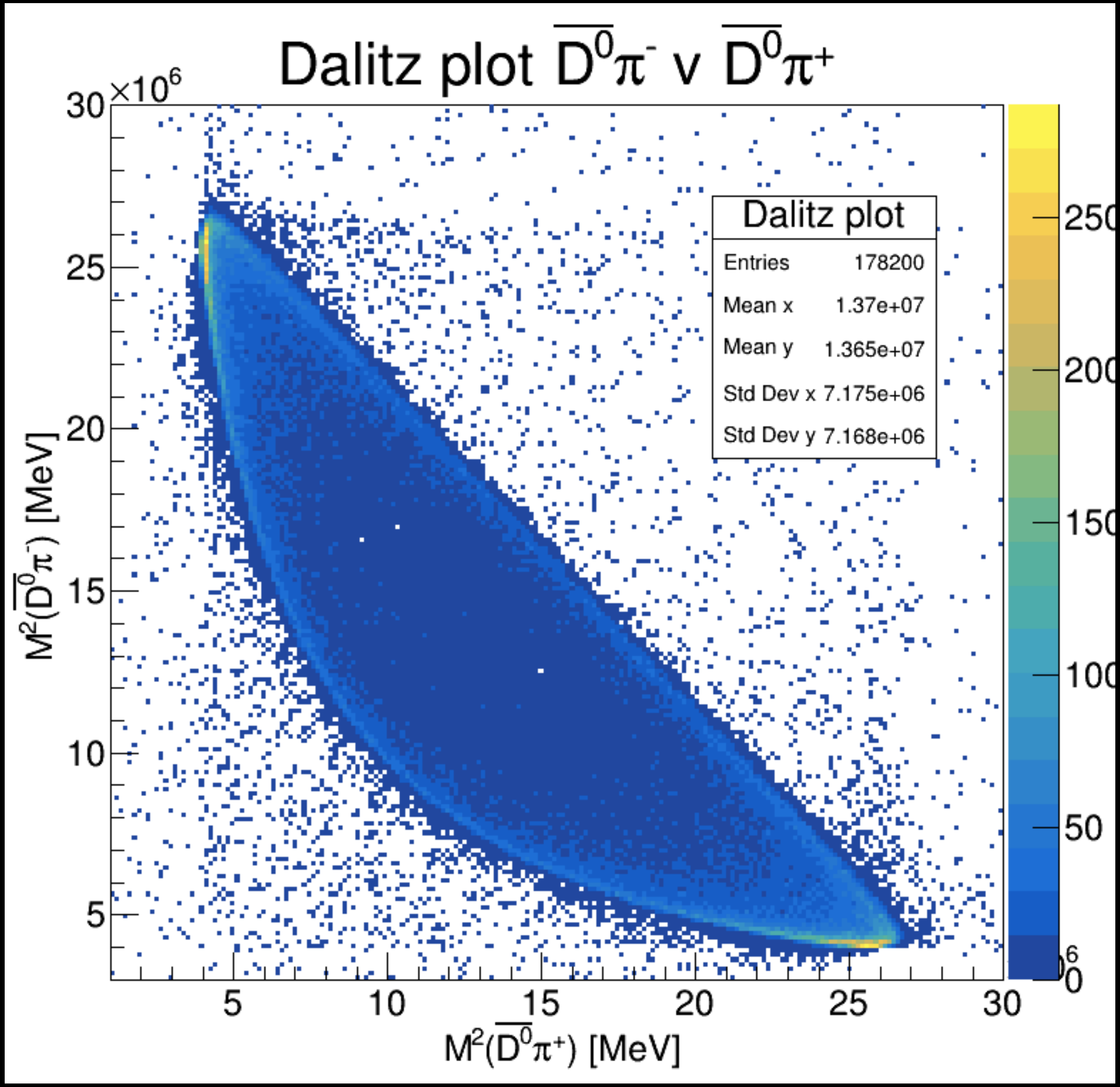


Figure 6: The Dalitz plot for this $B^0 \to \bar{D}^0\pi^+\pi^-$ decay shows no sign of resonant frequencies or spins.

Dalitz plots provide a way of visualising a three body decay. Through these plots it is easy to visualise resonances or spins in particle decays. For the decay considered in this project, $M^2(\bar{D}^0\pi^+)$ was plotted against $M^2(\bar{D}^0\pi^-)$. More on Dalitz plots can be found in [3].

STATISTICAL ANALYSIS

For statistical analysis, variables resolutions were compared. From that it was possible to plot residual as well as pull plots. Finally, χ^2 test and Kolmogorov test were performed. This was done for all 4 particles and for each of the variables used in the simulation, namely, ϕ, θ and $|\bar{p}|$.

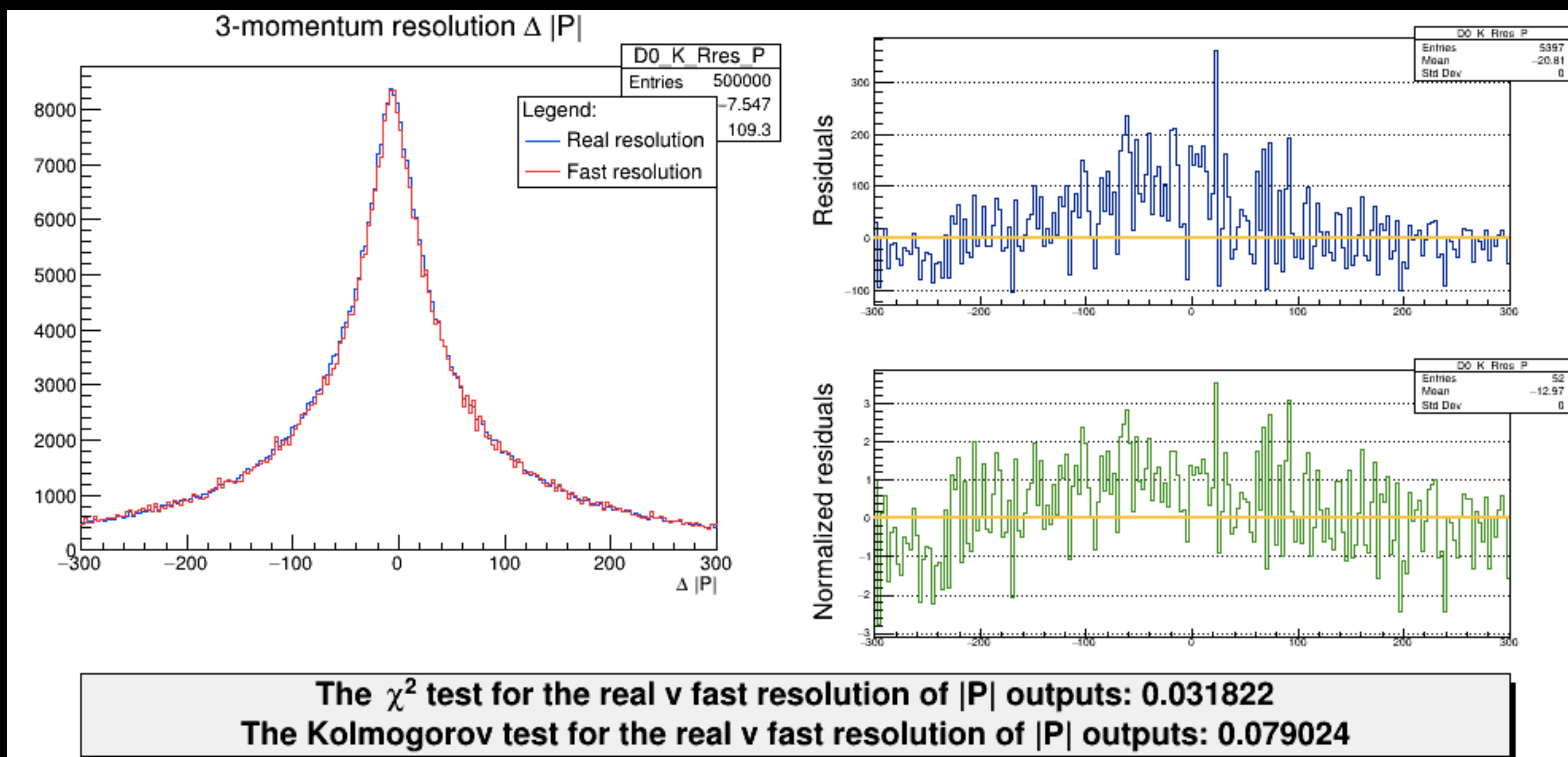


Figure 7: the statistical analysis for the modulus of the 3-momentum of the kaon is shown here. From the pull plot (lower right) it is clear that the original resolution and the simulated one are in good agreement. Both Kolmogorov and χ^2 test are of difficult interpretation in this context.

LIMITATIONS

As shown so far in this presentation, the LHCb data is successfully reproduced by the fast algorithm and the kinematical variables reconstructed. However, this comes with a number of constraints that were enforced in the code:

- Each particle needs to be trained separately.
- Acceptance training seems to work if all particles are trained together (meaning a 12 dimensional training), but the resolution training fails.
- The ratio between the number of events in the generated dataset and the reconstructed dataset used to train the algorithm needs to be finely-tuned.
- In the original dataset from LHCb cuts had been pre-applied to the phase space.
- When training a dataset without the cuts mentioned above, the training fails to give good results.

CONCLUSIONS & FUTURE WORK

In this project, a Machine Learning algorithm was employed to simulate a particle decay. For training, an original dataset from LHCb was used:

- The decay was successfully reproduced for a careful choice of certain parameters, such as the ratio between generated and reconstructed data.
- The reconstructed mass plots had to be normalised.
- The code had to be widely changed to accommodate for the structure of LHCb data.

There are many ways in which the current work can be improved:

- Generalise the algorithm such that all particles can be correctly trained all at once. This should fix the mass problem.
- Understand why the training for the dataset with no pre-applied cuts was not successful and fix it.
- Carry out a more careful error analysis to better assess the reliability of the results.

BIBLIOGRAPHY

[1] https://doi.org/10.48550/arXiv.2207.11254

[2] https://root.cern/primer/

[3] Brian Lindquist, Dalitz Plots SASS Talk, 2010