

# A Gaussian-sum filter for vertex reconstruction

T. Speer, Physik-Institut der Universität Zürich, Zürich, Switzerland

R. Frühwirth, Institute for High-Energy Physics, Austrian Academy of Sciences, Vienna, Austria

## Abstract

A vertex reconstruction algorithm was developed based on the Gaussian-sum filter (GSF) and implemented in the framework of the CMS reconstruction program. While linear least-square estimators are optimal in case all observation errors are Gaussian distributed, the GSF offers a better treatment of non-Gaussian distributions of track parameter errors when these are modelled by Gaussian mixtures. In addition, this ensures better protection against outliers and offers some robustness.

## INTRODUCTION

The most often used algorithm for vertex reconstruction is the *Kalman filter* (KF, [1]). It is mathematically equivalent to a global least-squares minimization, which is the optimal estimator when the model is linear and all random noise is Gaussian. For non-linear models or non-Gaussian noise, it is still the optimal linear estimator.

One method that takes non-Gaussian distributions of measurement errors better into account is the *Gaussian-sum filter* (GSF). In this method, the distributions of both the measurements errors and the estimated quantities are modelled by mixtures of Gaussians. The main component of the mixtures describes the core of the distributions, and the tails are described by one or several additional Gaussians. The GSF is a non-linear estimator, as the weights of the components depend on the measurements.

This approach has first been developed and tested for track reconstruction [2], where it has been seen that in the presence of large tails, the GSF has smaller variance than the KF. It is particularly useful for electron reconstruction, as the Bethe-Heitler distribution of bremsstrahlung energy loss is highly non-Gaussian and can be modelled by a Gaussian mixture. This has been successfully implemented in the CMS reconstruction software [3], where an improvement of the track parameter resolution is seen.

## THE GAUSSIAN-SUM FILTER

For the fit, as in the KF, an iterative procedure is applied, where the estimate of the vertex is updated with one track at the time. When one track is added to the vertex, each component of the vertex state mixture is updated with each component of the track measurement mixture by a KF, effectively doing an exhaustive combination of the components of the two mixtures. In addition, the weight of each combination has to be computed. The number of components of the estimated vertex rises thus exponentially, as

at each step it is multiplied by the number of components modelling the new track.

## Validation

To validate the algorithm, a simplified simulation in a fully controlled environment has been used. A single four-track vertex is generated per event. The direction of the jet is parallel to the  $x$ -axis, and the tracks are in a cone of 0.5 rad. No track reconstruction is done, and track parameters are smeared according to a two-component Gaussian mixture model. The component modelling the non-Gaussian tails of the distribution (the *wide component*) has a standard deviation ten times larger than that of the core component (the *narrow component*). Their relative weights are 90% for the narrow component and 10% for the wide component. The standard deviations of the impact parameters are 100  $\mu\text{m}$  and 1000  $\mu\text{m}$ , respectively.

The fit with the KF uses only one component. The variance of the track parameters used is that of the dominant (narrow) component<sup>1</sup>. The distributions of the vertex-coordinate residuals and pulls (Fig. 1) have a Gaussian core with tails, and 36% of the fits have a  $\chi^2$ -probability below 0.01.

The GSF uses both components, each with the correct weight and variance. In these first tests, the number of components kept is not limited. The distributions of the residuals (Fig. 2) have fewer outliers and the core, when fitted with a Gaussian, has a smaller standard deviation (called hereafter the *resolution*). The few outliers are due to vertices with several track-outliers. The distributions of the pulls do not show tails and are nearly perfectly Gaussian with a standard deviation very close to 1. This indicates that the errors on the track-outliers are correctly taken into account and that a high weight is assigned to the correct component.

The distribution of the  $\chi^2$ -probability shows a dip at 0, large values of  $\chi^2$  occurring less frequently than expected. This can be explained by the fact that the tails of the narrow core component are well within the range of the wide component. Observations in the tails of the core are therefore interpreted as coming from the wide component, and their contribution to the  $\chi^2$  is accordingly small. Large values of  $\chi^2$  can occur only if all observations are in the tails of

<sup>1</sup>In case of a high percentage of tails, it is arguable what the best choice for that variance would be. The position of the estimated vertex, and therefore the residuals, are nevertheless independent of the scaling factor of the variance used by the fit, whereas the uncertainty of the estimated vertex, and thus the pulls and the  $\chi^2$ , are no longer independent of that scaling factor.

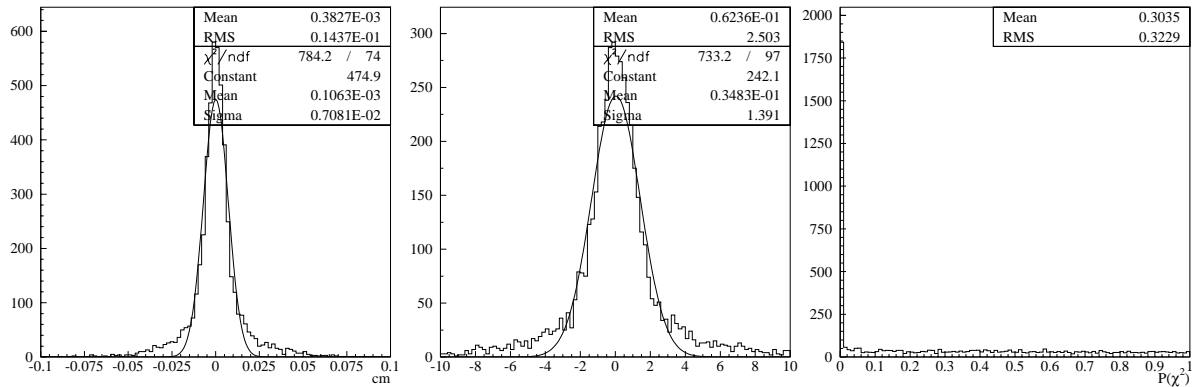


Figure 1: Residual (left) and pull (middle) of the  $y$ -coordinate of the reconstructed vertex and  $\chi^2$ -probability (right) of the vertex fit using the Kalman Filter.

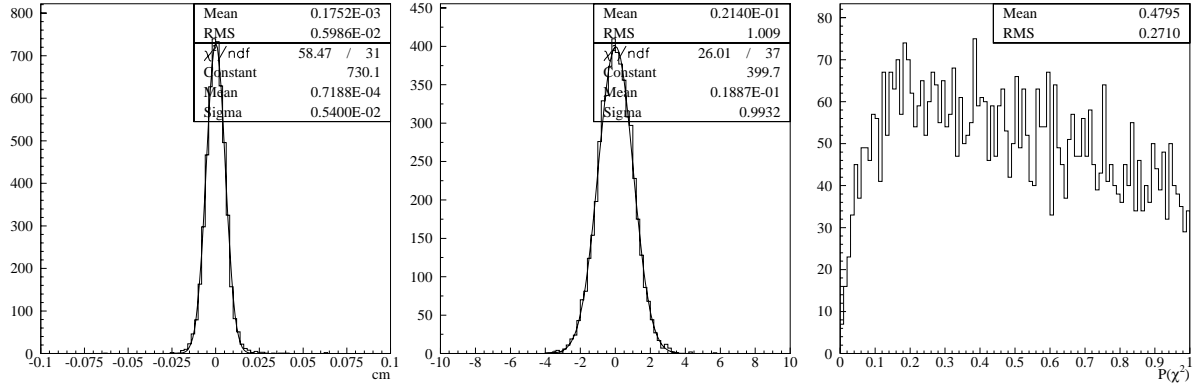


Figure 2: Residual (left) and pull (middle) of the  $y$ -coordinate of the reconstructed vertex and  $\chi^2$ -probability (right) of the vertex fit using the GSF, without limiting the number of components.

the wide component, and this has a very low probability already with as few as four tracks.

### Limitation of the number of components

As the number of components of the vertex state mixture increases exponentially, it has to be limited to an acceptable number. This is achieved by clustering (collapsing) components which are close, according to a defined distance measurement, pair-wise until the desired number of components is reached. Two distance measurements are used, the *Kullback-Leibler Distance* [4] and the *Mahalanobis Distance* [5]. No significant difference has been found when using either of the distance measurements. We conclude that the GSF shows little sensitivity to the number of components kept during the fit, and good results can be achieved with even a small number of components.

## ROBUSTNESS TESTS

The GSF is very efficient when outliers are accurately described by the *pdfs* modelling the errors of the parameters, whereas, as is well known, LS-estimators such as the KF are very sensitive to those same outliers. The robustness with respect to true outliers, i.e. tracks which are not modelled correctly, can be tested by adding mismeasured tracks (type 1 outliers) or tracks from another vertex (type 2 outliers) to the list of correct tracks. In this respect,

the GSF can be compared to another non-linear filter, the *Adaptive Filter* (AF) [6]. The AF is an iterative re-weighted KF which down-weights tracks according to their  $\chi^2$  distance to the vertex. It has been shown that this filter is very stable with a high break-down point. Both filters can actually be combined. Indeed, as in the AF the computation of the vertex position is independent of the computation of the track weights, the KF used in the default implementation can be replaced by a GSF. In this way, the complete mixture modelling the track measurements is taken into account, instead of only a single component. This filter is referred to as the *Adaptive-GSF* (A-GSF). The AF being sensitive to only a single component, the variance used for this filter is that of the narrow component.

For vertices without outliers, the distributions of the residuals and of the pulls of the vertices fitted with the AF and the A-GSF are similar to those obtained with the GSF. To assess the improvement of the filters with respect to the KF, the half-widths of the symmetric intervals covering 50% and 90% of the residual distribution (the *50%* and *90% coverages*) of the  $y$ -coordinate are used. Table 1 shows that the results of the three non-linear filters are similar, although the 90% coverage indicates that the residual distribution for the AF has slightly heavier tails. In addition, some 19% of the fits performed with the AF still result in a  $\chi^2$ -probability below 1%. The  $\chi^2$ -probability distribution of the A-GSF is very similar to the one of the GSF.

Table 1: Comparison of the average  $\chi^2$ -probability, resolution, pulls, 50% and 90% coverages of the  $y$ -coordinate of vertices estimated with the different filters.

Filter	Mean $P(\chi^2)$	Resolution [ $\mu\text{m}$ ]	Pull	50% Cov. [ $\mu\text{m}$ ]	90% Cov. [ $\mu\text{m}$ ]
KF	0.32	71	1.39	48	229
GSF	0.27	54	0.99	36	90
AF	0.30	59	1.08	40	113
A-GSF	0.27	54	0.94	36	91

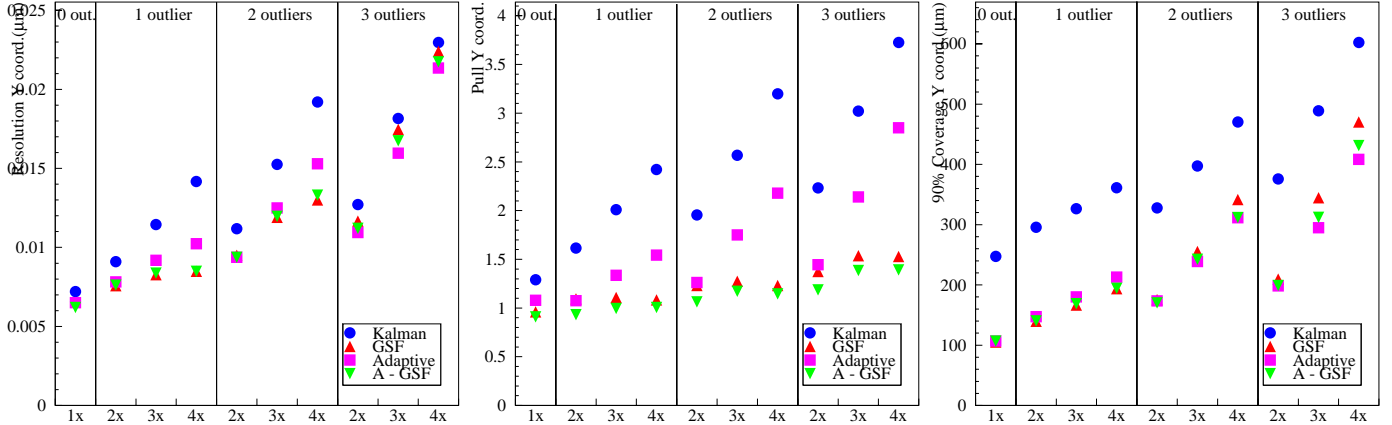


Figure 3: Resolution (left), pulls (middle) and 90% coverage of the  $y$ -coordinate of vertex fits for tracks with two components for different number of outliers among the four tracks, and different ratios of their standard deviations between the outliers and the inliers.

### Type 1 outliers

Type 1 outliers have been simulated by smearing the tracks parameters with a different mixture than the one used in the vertex fit, the latter being the default mixture. Such outliers simulate tracks with errors that have been seriously underestimated. These tests have been done with four-track vertices in which 1, 2 or 3 tracks out of the 4 are outliers. These outliers are smeared with a two-component mixture which is obtained from the default mixture by increasing the standard deviations by factors of 2, 3 or 4.

The values of the resolutions, pulls and the 50% and 90% coverages are summarized in Figure 3. Vertices fitted with the KF are significantly degraded by the track-outliers, the core of both the residual and pull distributions being significantly broader with heavier tails. The non-linear filters show a consistent improvement of the distribution of the residuals with respect to the KF, both in terms of resolution and coverages. For vertices fitted with the GSF or the A-GSF, the RMS of the cores of the residual distributions are obviously somewhat broader than those found for vertices without the outlier, with only a modest increase of the number of outliers, and the pull distributions are nearly unchanged. Vertices fitted with the AF feature broader residual and pull distributions, as can also be seen in the coverages. The pull distributions of the GSF and in particular those of the A-GSF are very stable, and the variance remains close to 1. A large number of fits performed with the KF have a  $\chi^2$ -probability below 1%, while for the AF this number is approximately half as large. For the other two

filters, the distribution has no peak, but the distributions are shifted to lower values with respect to those obtained without outliers.

### Type 2 outliers

To test the sensitivity to type 2 outliers, an outlying track originating from a second vertex is added to the tracks from the main vertex. This second vertex is displaced by distances varying between 1 and 5 mm in the direction ( $y$ ) transverse to the jet-axis of the main four-track vertex. The track parameters of all tracks, both inliers and outliers, are smeared with the default two-component mixture.

Vertices fitted with the KF are shifted towards the second vertex, as can be expected by the configuration of the tracks. The distributions of the residuals and pulls of vertices fitted with either of the non-linear filters are hardly affected by the presence of the outliers, and remain remarkably stable. Again, the residual and pull distributions for fits with the AF have somewhat heavier tails, as is confirmed by the 90% coverage. The  $\chi^2$ -probability distribution reveals a significant peak at 0. While the  $\chi^2$ -probability distribution for the GSF is shifted to lower values, the corresponding distribution for the A-GSF is nearly unchanged. The values of the resolutions, pulls and the 50% and 90% coverages are summarized in Figures 4 and 5.

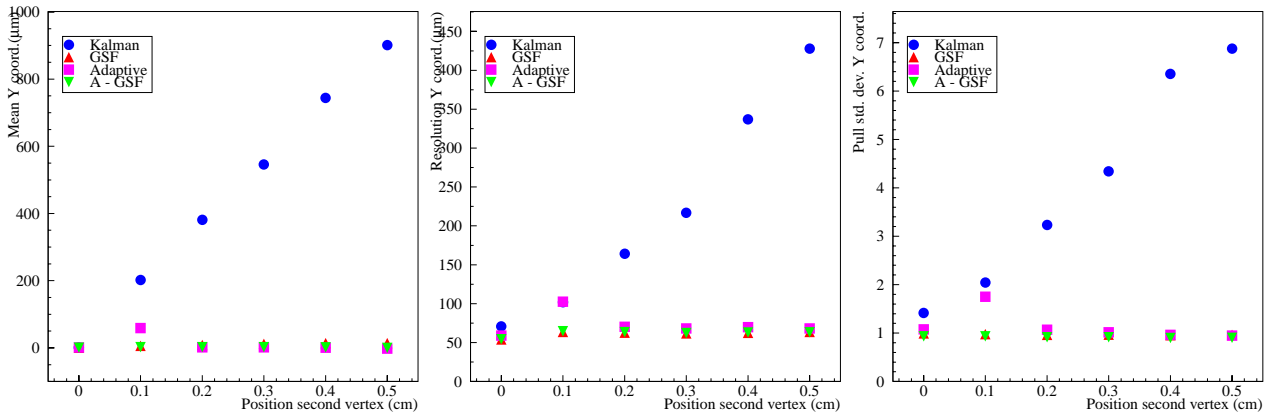


Figure 4: Means (left) and standard deviations (middle) of the residual distributions and standard deviations (right) of the pull distributions of the  $y$ -coordinate of vertex fits for tracks with two components for different positions along the  $y$ -axis of the vertex of the outlying track.

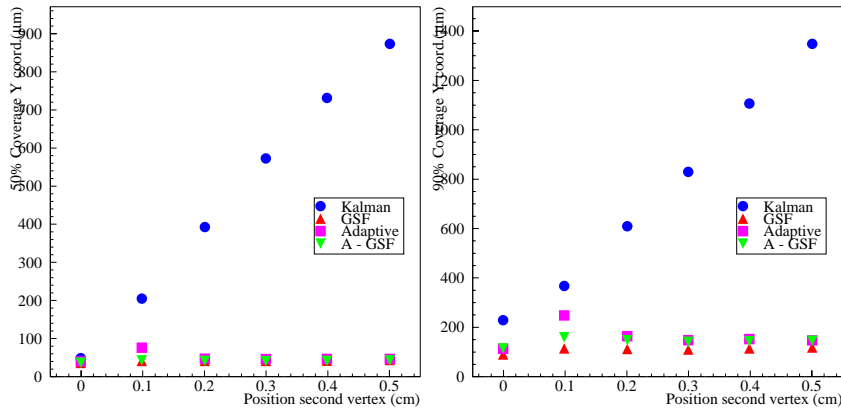


Figure 5: 50% (left) and 90% coverage (right) of the  $y$ -coordinate of vertex fits for tracks with two components for different positions along the  $y$ -axis of the vertex of the outlying track.

## CONCLUSION

A Gaussian-sum filter for vertex reconstruction has been implemented in the CMS reconstruction software. It has been validated with a simplified simulation, with different distributions of track parameter errors. While these distributions are not those obtained with a real track reconstruction, the results with the KF are very close to those observed with fully simulated data, and indicate that similar results could be obtained with the GSF.

These tests demonstrate the validity of the algorithm and its efficiency in the presence of non-Gaussian noise. An improvement of the resolution and error estimate of the fitted vertex and of the  $\chi^2$  of the fit with respect to the KF when the track parameters errors have non-Gaussian tails are obtained. In addition, when using electron tracks reconstructed with the GSF, the full mixture can be used, instead of only a single collapsed state. The GSF shows little sensitivity to the number of components kept during the fit, such that only a small number of these can be kept without degrading the fit too much. This allows the filter to be used without incurring a high performance penalty.

This filter also demonstrates a higher degree of robustness than LS-estimators such as the KF in the presence of outliers. These tests show the adaptive power of the

GSF, which, in case of an outlier, is able to assign a higher weight to the track-components with the widest standard deviation. This ensures better protection against outliers and offers some robustness. Obviously, as the model does not describe the data adequately, the  $\chi^2$ -probability will be distorted. This distortion is nevertheless less severe than for LS estimators. A more robust approach is to combine the AF with a Gaussian-sum filter, as the combined filter is able to both down-weight outliers and use the full mixture for the remaining tracks.

## REFERENCES

- [1] R. Frühwirth, R. Kubinec, W. Mitaroff and M. Regler, *Comp. Phys. Comm.* **96**, 189 (1996)
- [2] R. Frühwirth, *Comp. Phys. Comm.* **100**, 1 (1997)
- [3] W. Adam, R. Frühwirth, A. Strandlie, T. Todorov, eConf **C0303241** (2003) TULT009
- [4] G. Kitagawa, *Comput. Math. Appl.* **18**, 503 (1989)
- [5] K.V. Mardia, J.T. Kent, J.M. Bibby, *Multivariate Analysis*, Academic Press, London, UK, 1979
- [6] J. D'Hondt, P. Vanlaer, R. Frühwirth, W. Waltenberger "Sensitivity of Robust Vertex Fitting Algorithms", CMS internal note 2004/002 (2004), to appear in "IEEE Transactions on Nuclear Science"