



Track classification for high density beams in beam tests

Christopher Krause

christopher2.krause@tu-dortmund.de

Kevin Kröniger, Jens Weingarten

May 23. 2023

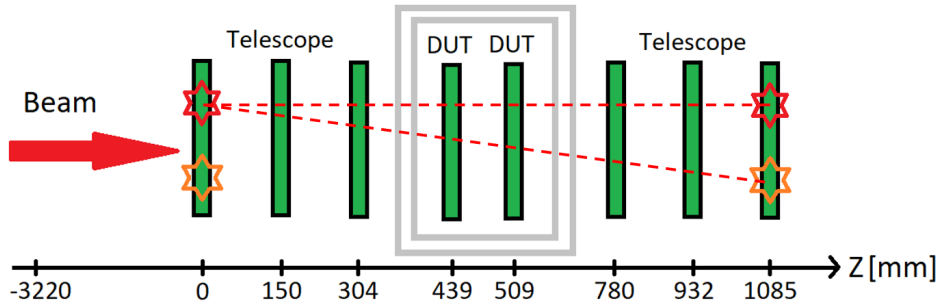
TU Dortmund

Department of Physics

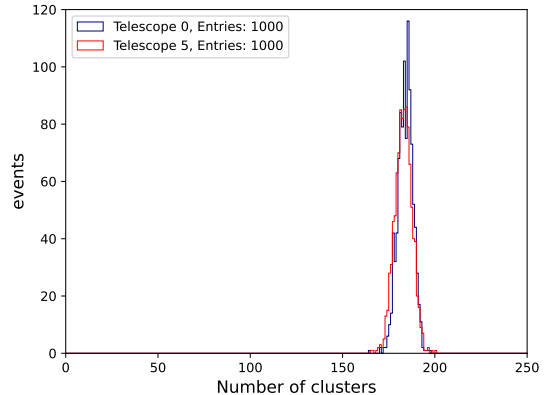
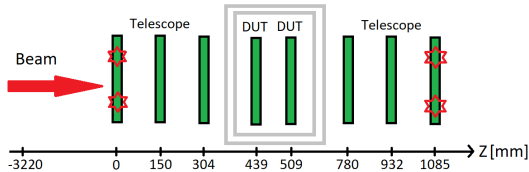
4th Allpix Squared User Workshop



- Track reconstruction of high density beam causes combinatorial problems
- All cluster combinations on first and last plane are taken into account
- Wrong combinations possible → False tracks

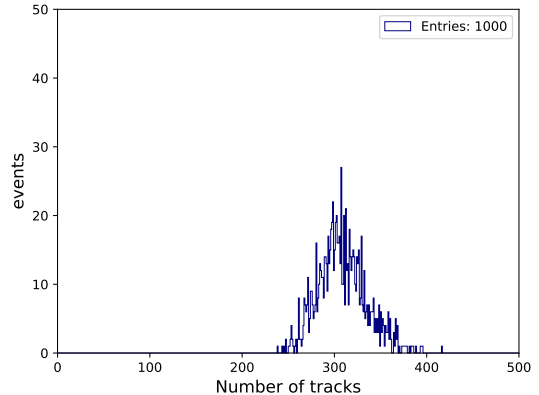


- Allpix² simulation in air, 200 pions per event (1000 events)
- Telescope consists of 6 Mimosa26 sensors
- No sensor misalignment
- Beam opening angle: 0.1mrad
- Beam energy: (120.0 ± 0.5) GeV

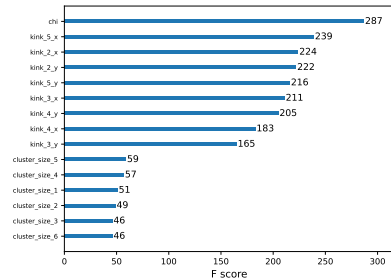
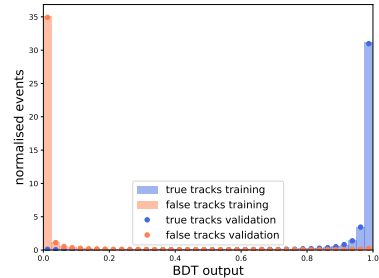


- Reconstruction configurations:
 - Demand hit on all planes for track
 - Matching Radius: $10 \times$ spatial resolution
- **364k** tracks in total, less than **200k** clusters on each plane
- Number of tracks significantly larger than expected number of true tracks

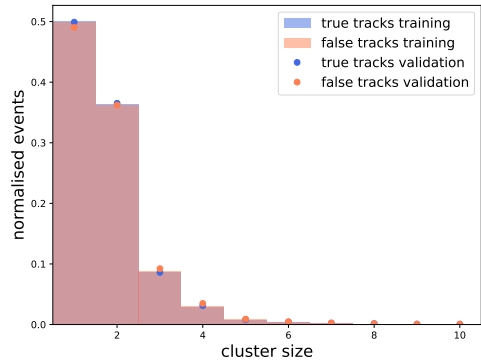
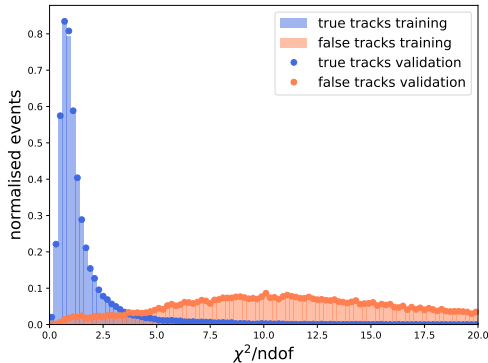
Can true and false tracks be distinguished?



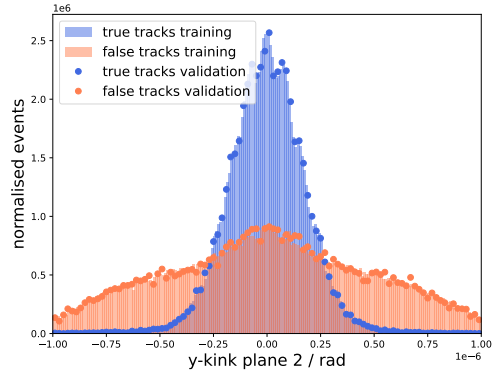
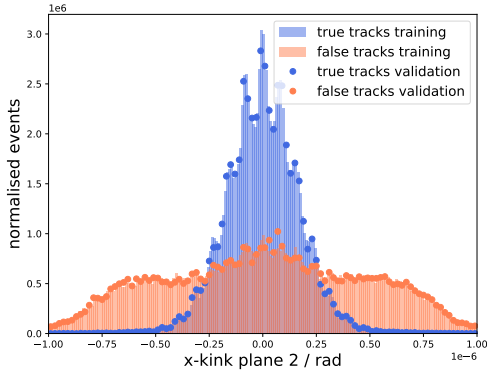
- Using machine learner (XGBoost) to determine true and false tracks
- Training data set: 200k π^+ , Validation/Test data set: 50k π^+
- Input quantities: χ^2/ndof , kink angles, cluster sizes



- True and false track distributions of the input features



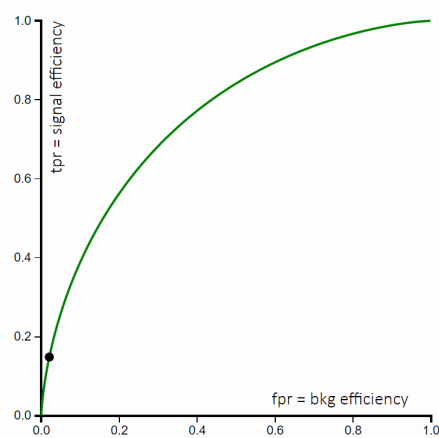
- True and false track distributions of the input features



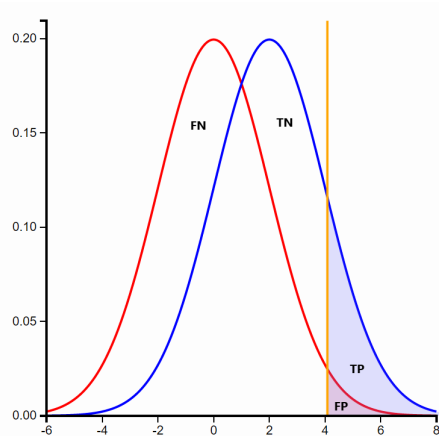
- Evaluate classification \rightarrow ROC curve
- Each cut describes one point of the curve

$$\text{TPR} = t_p / (t_p + f_n)$$

$$\text{FPR} = f_p / (f_p + t_n)$$



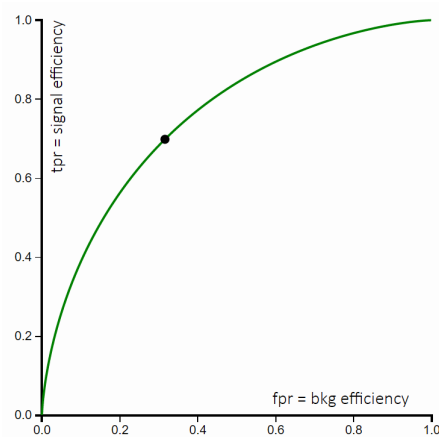
[Alex Rogozhnikov, ROC curve demonstration]



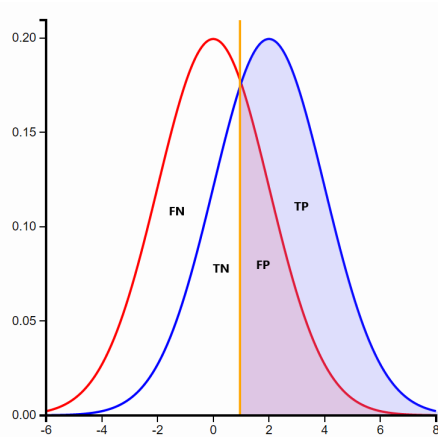
- Evaluate classification \rightarrow ROC curve
- Each cut describes one point of the curve

$$\text{TPR} = t_p / (t_p + f_n)$$

$$\text{FPR} = f_p / (f_p + t_n)$$



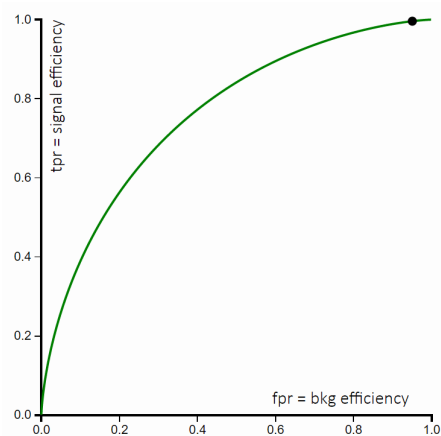
[Alex Rogozhnikov, ROC curve demonstration]



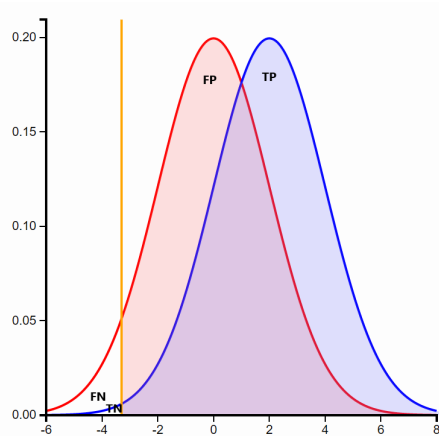
- Evaluate classification \rightarrow ROC curve
- Each cut describes one point of the curve

$$\text{TPR} = t_p / (t_p + f_n)$$

$$\text{FPR} = f_p / (f_p + t_n)$$



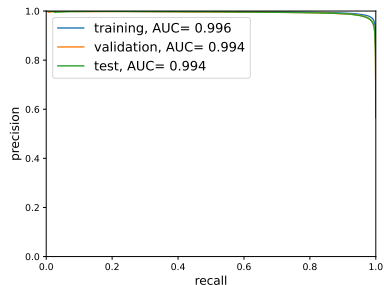
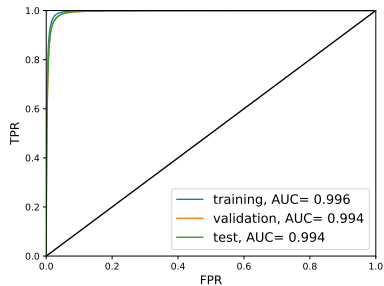
[Alex Rogozhnikov, ROC curve demonstration]



- Using ROC curve and precision-recall curve for performance evaluation
- $FPR = f_p / (f_p + t_n)$
Probability for false tracks to be labeled as true
- $TPR = \text{Recall} = t_p / (t_p + f_n)$
Ability to find all true tracks
- $\text{Precision} = t_p / (t_p + f_p)$
Ability to not label false tracks as true

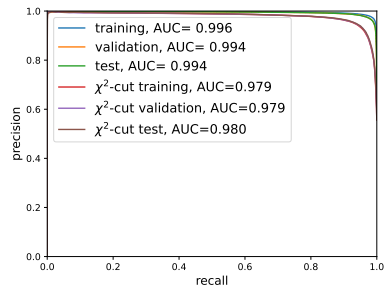
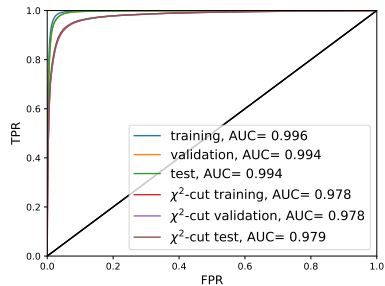
		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

[C. Dilmegani, Machine Learning Accuracy: True-False Positive/Negative, 2023]

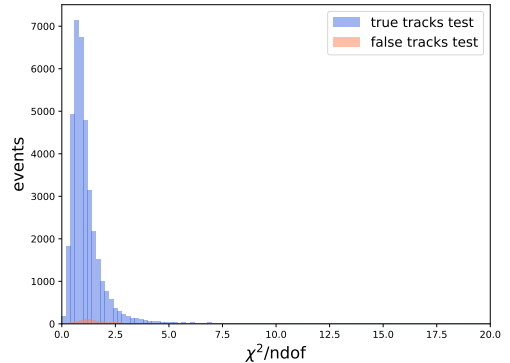
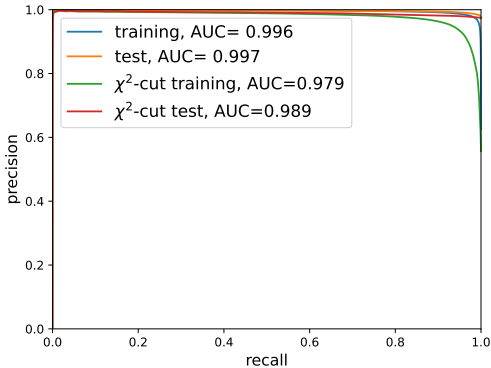


- $\text{FPR} = f_p / (f_p + t_n)$
Probability for false tracks to be labeled as true
- $\text{TPR} = \text{Recall} = t_p / (t_p + f_n)$
Ability to find all true tracks
- $\text{Precision} = t_p / (t_p + f_p)$
Ability to not label false tracks as true
- Comparison to most important quantity:
 χ^2/ndof

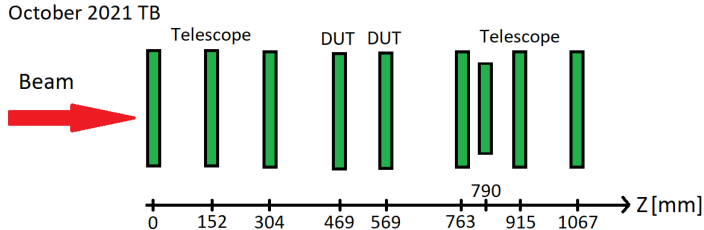
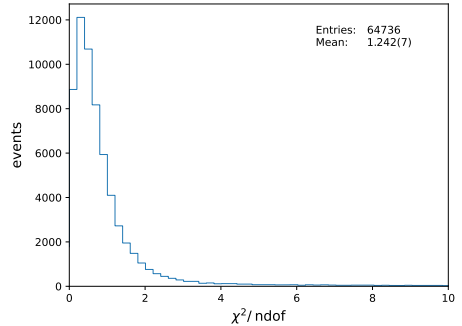
Cut on χ^2/ndof almost as good



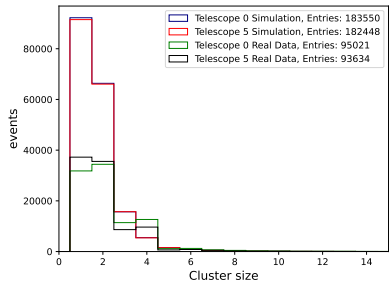
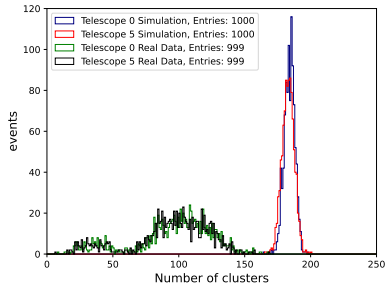
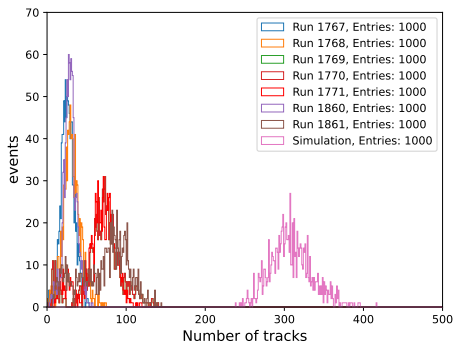
- Only using one track per cluster (keeping track with lowest χ^2/ndof)



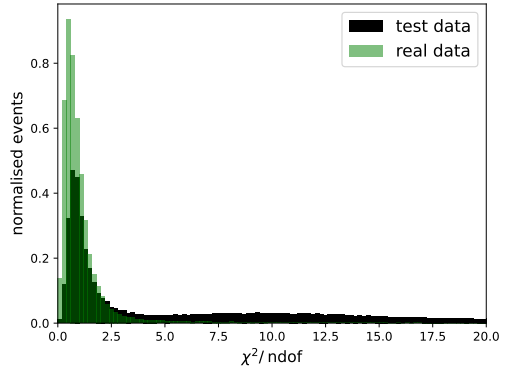
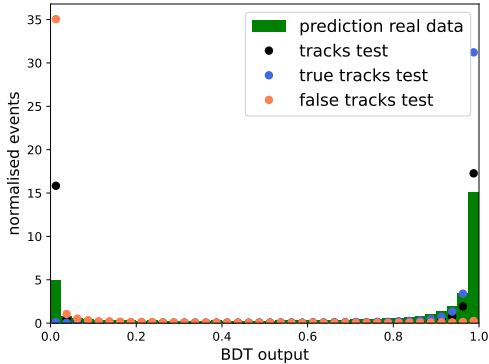
- Using Cern Testbeam data from October 2021 (High track densities)
- Telescope precisely aligned
Masking: Frequency cut = 20
- Matching Radius: $10 \times$ spatial resolution



- Using only 1k events from real data (Run 1771)
- Simulated data has significantly higher track density
 → Less false tracks in real data



- Clear distinction between tracks with low and high probabilities



- High track density beams in past campaigns had $\approx 2x$ larger track multiplicity
- Applying χ^2 -cut yields high precision and recall values
- False tracks have significantly larger χ^2 -values on average
- Using adequate χ^2 -cut/unique clusters filters out most false tracks

Track reconstruction of high density beams in beam tests possible

- High track density beams in past campaigns had $\approx 2x$ larger track multiplicity
- Applying χ^2 -cut yields high precision and recall values
- False tracks have significantly larger χ^2 -values on average
- Using adequate χ^2 -cut/unique clusters filters out most false tracks

Track reconstruction of high density beams in beam tests possible

Thank You!