

Vier Physizierende mit Schluckauf...

Sommerakademie Annecy 2024

Theo Lequy^{1,5}, Jakov Wallbrecher^{1,6}, Erik Haufs^{2,3}, Linn Fischbach¹,
ChatGPT⁴

¹ETH Zürich, ²Ruhr-Universität Bochum, ³Heinrich-Heine-Universität Düsseldorf, ⁴OpenAI San Francisco, ⁵Otto-von-Guericke-Universität Magdeburg, ⁶Ludwig-Maximilians-Universität München



große Bins, Datensatz A-D

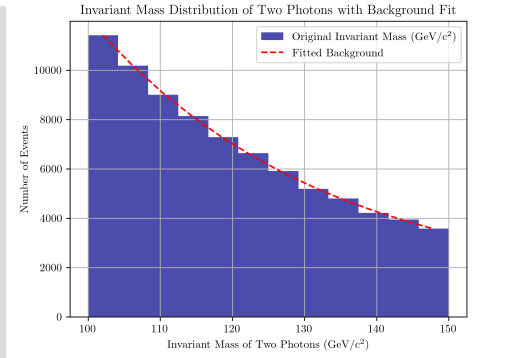


Abb. Der gesamte Datensatz wurde mit `photon_isTightID == True & photon_ptcone30 < 5` gefiltert und mit einer Background Funktion von $a * e^{b*x} + c * x + d$ modelliert.

große Bins, Datensatz A-D

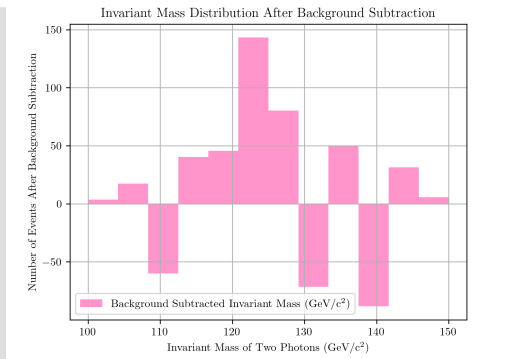


Abb. Wenn der Background abgezogen wird entsteht eine grobe Gauss Verteilung um $125 \text{ GeV}/c^2$ zentriert.

Manual Filters

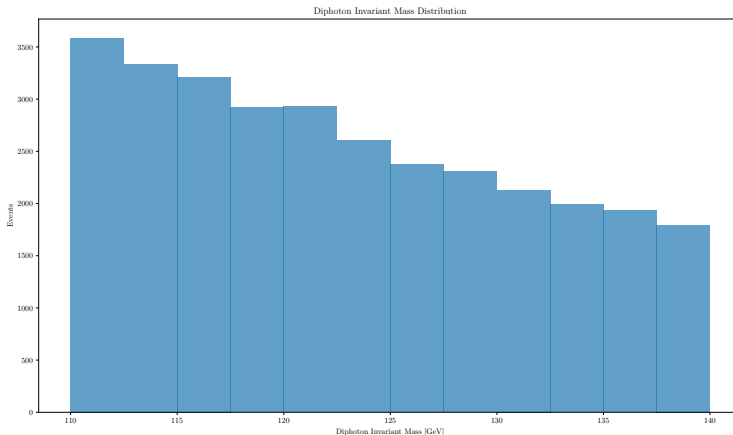


Abb. gefilterte Datensätze A-D mit `photon_isTightID == True & jet_n == 0 & met_et/1000 < 25`

Manual Filters

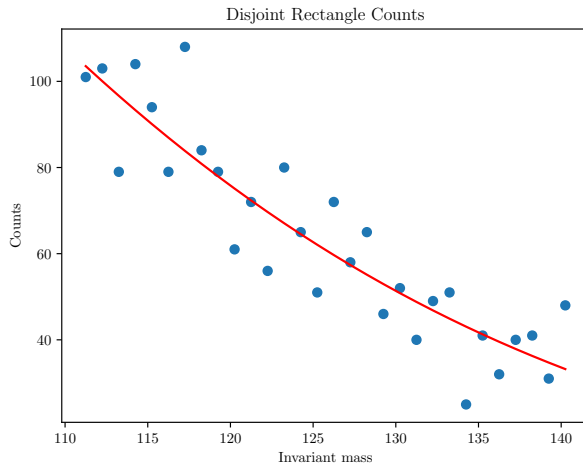


Abb. applied moving average with disjoint rectangles, trend fit with $a \exp(bx + cx^2)$



Manual Filters: Sliding

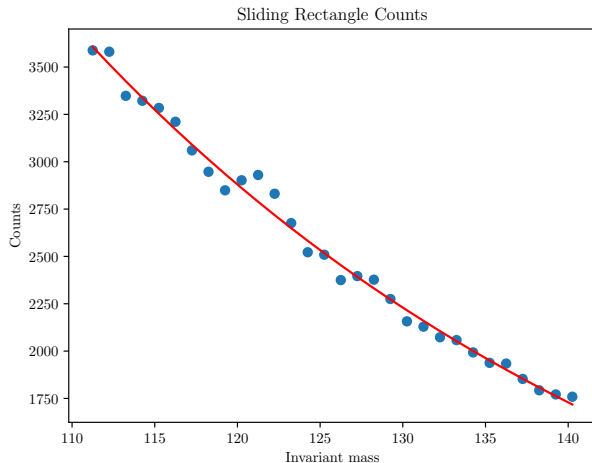


Abb. applied moving average with rectangular kernel, trend fit with $a \exp(bx + cx^2)$



Manual Filters: Sliding & Epanechnikov detrended

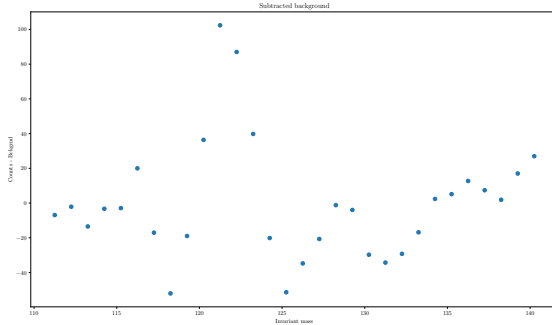


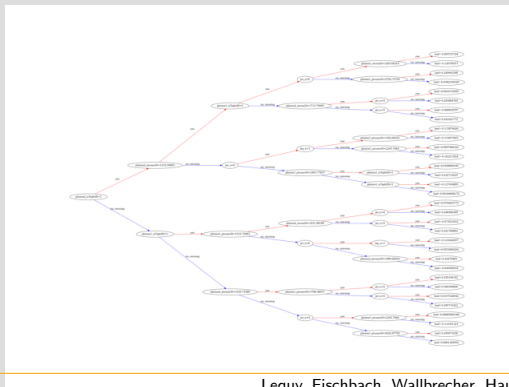
Abb. manually filtered, applied weighted moving average with Epanechnikov kernel, detrended with exp-polynomial of degree 2

XGB decision tree



Training eines Decision-trees zur Unterscheidung zwischen Background und Higgs-Event.

Parameter: photon_trigMatched, jet_n, lep_n, met_et, photon_isTightID, photon_ptcone30, photon_etcon20



The plot shows the density of predicted probabilities for two classes: background (blue line) and MC signal (red line). The background distribution is extremely narrow and centered near 0.0, while the MC signal distribution is broader and shifted towards higher predicted probabilities, with a notable peak near 1.0.

Decision Tree, Data Separation

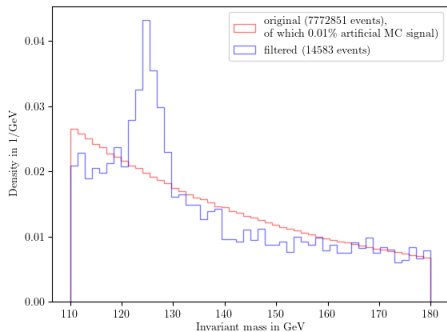


Abb. with 0.01% artificial signal

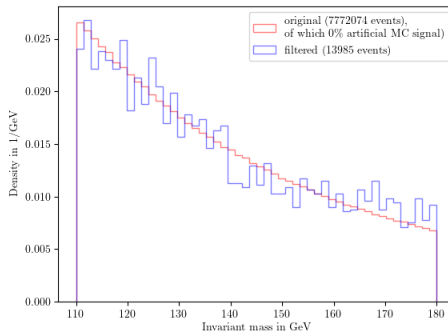


Abb. without artificial signal



XGB decision tree & Sliding Epanechnikov

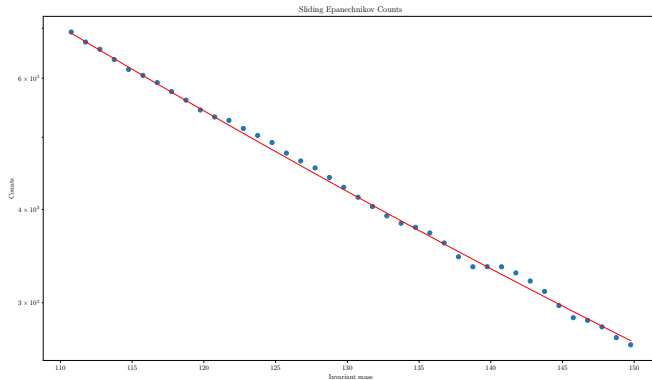


Abb. filtered data with XGB decision tree, applied weighted moving average with Epanechnikov kernel, trend fit with $a \exp(bx + cx^2)$

The plot shows the distribution of 'Counts' (y-axis, ranging from -150 to 150) versus 'Iron-rich mass' (x-axis, ranging from 110 to 150). The data points are represented by blue dots with vertical error bars. A red curve represents a Gaussian fit to the data, which peaks at approximately x=125 and y=100.

- Über Poisson-Näherung: im Bereich [120,124]
- $p = 0.008, Z = 2.38$
- Über simulierte Verteilung:
- $p = 1.018 \cdot 10^{-5}, Z = 4.26$
- **Mit Vorsicht zu genießen!**
- Besser: Nun auf unbekannten Daten testen!

Einsatz von ChatGPT



- Erklärung der verschiedenen Parameter der Diphoton-Events
- Finden relevanter Parameter zur Higgs-Event-Filterung
- Generieren von Code-Logik bis zu einem gewissen Grad hilfreich, oft Debugging nötig (verwendet zB falsche Variablen zum Filtern)
 - falls genaue Vorstellung von Codeaufbau, ist Eigenentwicklung oft schneller
- Generieren von Plotting-Code