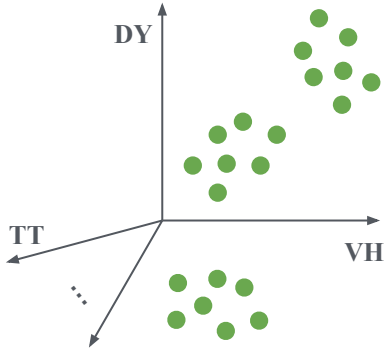
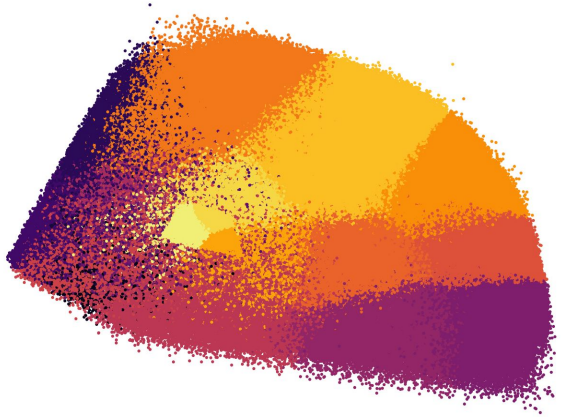
Federal Ministry
of Education
and Research

Funded by
DFG Deutsche
Forschungsgemeinschaft
German Research Foundation





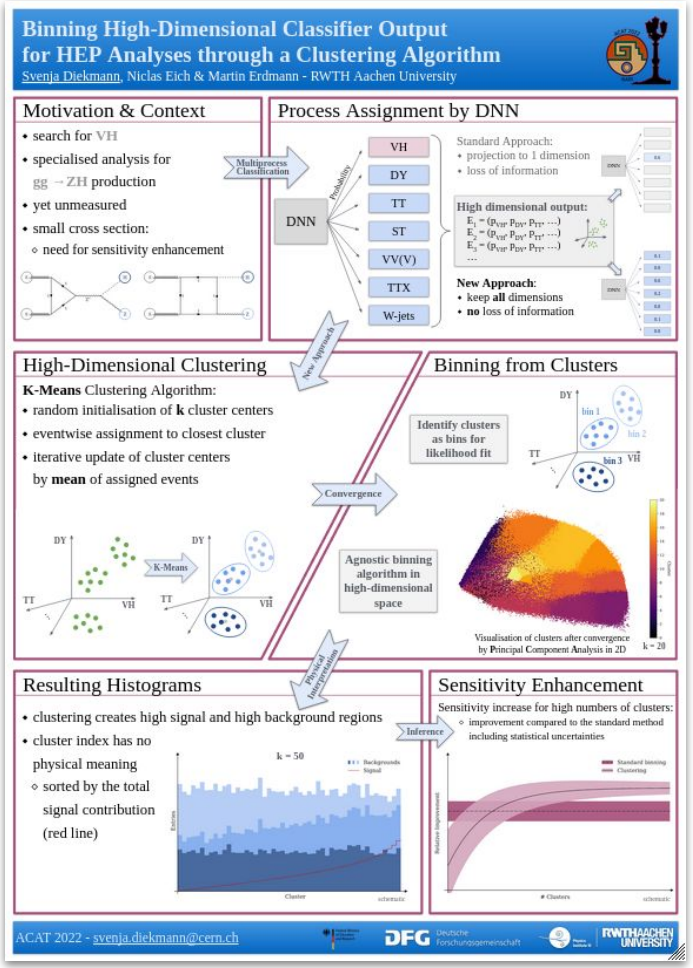
Binning High-Dimensional Classifier Output for HEP Analyses through a Clustering Algorithm

Svenja Diekmann, Niclas Eich & Martin Erdmann

ACAT 2022
28.10.22

RWTH AACHEN UNIVERSITY

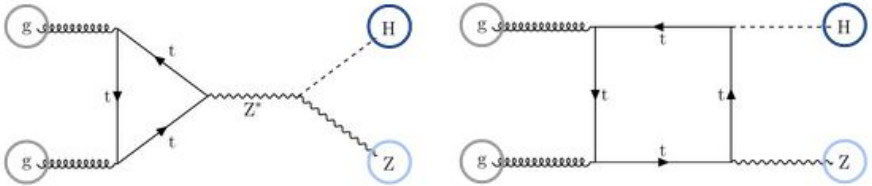
Thank you for voting



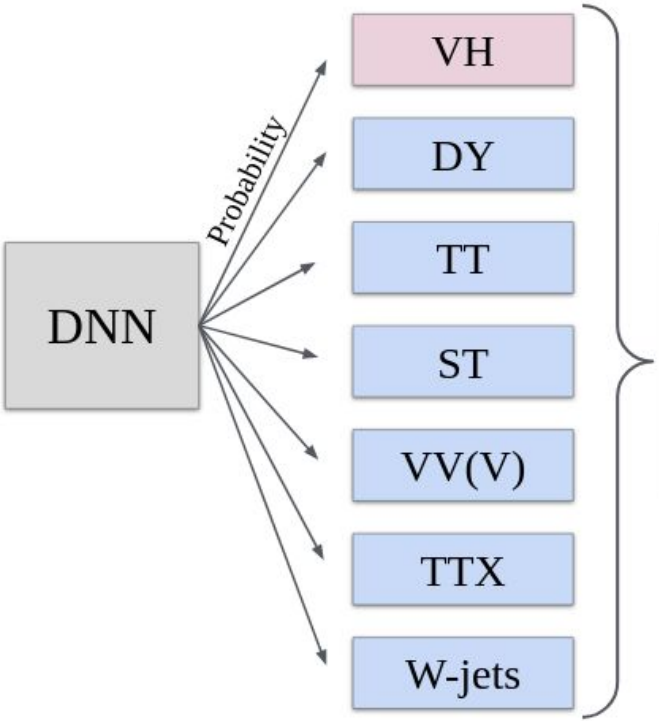
Motivation & Context

- search for VH
- specialised analysis for

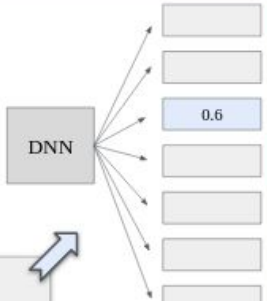
  gg → ZH production
- yet unmeasured
- small cross section:
  - need for sensitivity enhancement



Process Assignment by DNN



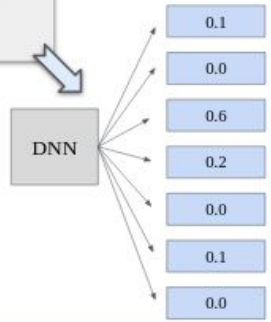
Standard Approach:
- projection to 1 dimension
- loss of information



High dimensional output:

$E_1 = (p_{VH}, p_{DY}, p_{TT}, \ldots)$
$E_2 = (p_{VH}, p_{DY}, p_{TT}, \ldots)$
$E_3 = (p_{VH}, p_{DY}, p_{TT}, \ldots)$
$\ldots$

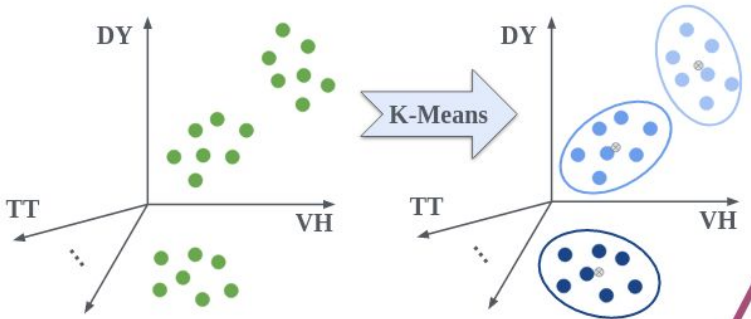
New Approach:
- keep **all** dimensions
- **no** loss of information



High-Dimensional Clustering

K-Means Clustering Algorithm:

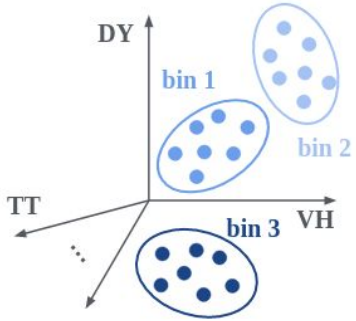
- random initialisation of **k** cluster centers
- eventwise assignment to closest cluster
- iterative update of cluster centers by **mean** of assigned events



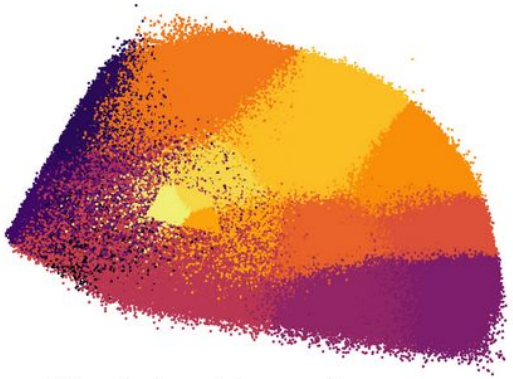
RWTH AACHEN UNIVERSITY
Svenja Diekmann 28.10.2022

Binning from Clusters

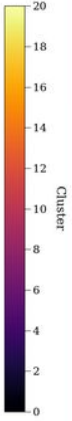
Identify clusters as bins for likelihood fit



Agnostic binning algorithm in high-dimensional space



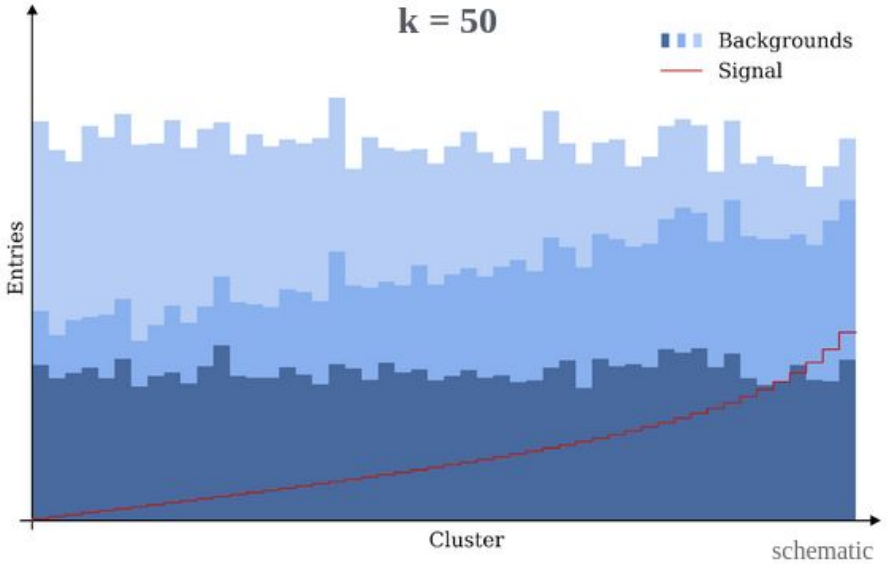
Visualisation of clusters after convergence by **P**rincipal **C**omponent **A**nalysis in 2D



k = 20

Resulting Histograms

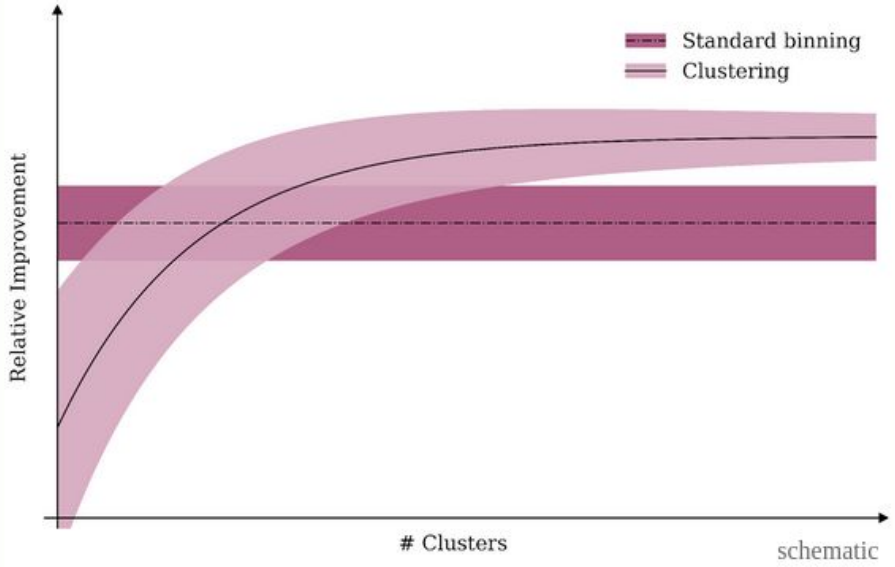
- clustering creates high signal and high background regions
- cluster index has no physical meaning
  - sorted by the total signal contribution (red line)

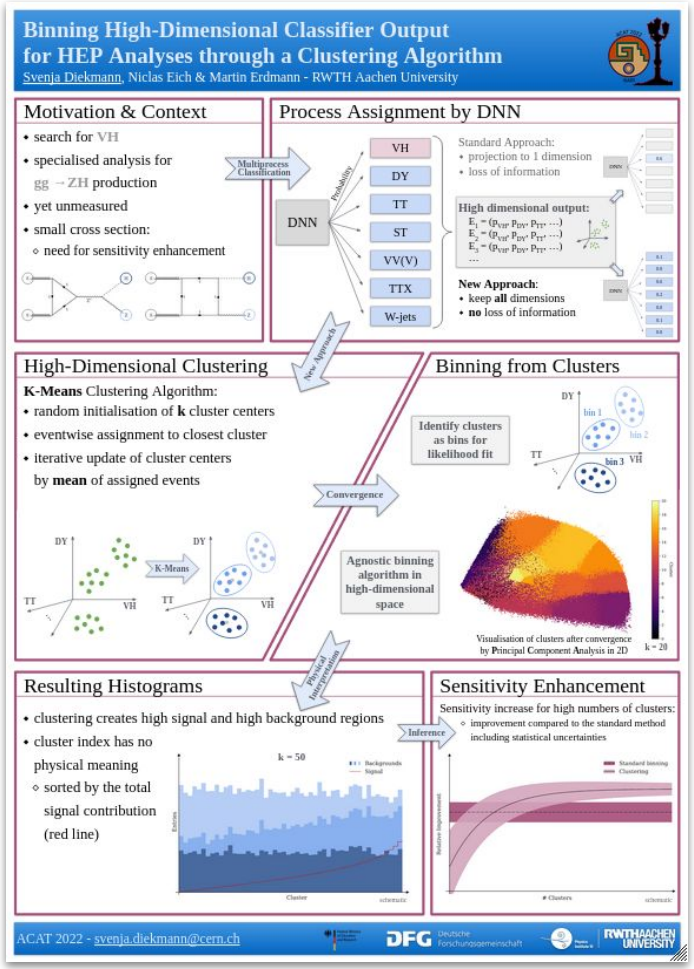


Sensitivity Enhancement

Sensitivity increase for high numbers of clusters:

- improvement compared to the standard method including statistical uncertainties





Thank you
for your attention