LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Jet Observables - Exploratory Survey

Strong 2020 Jet Observables Workshop
20th November 2020

Miguel Crispim Romão
mcromao@lip.pt

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824093."

Fundação para a Ciência e a Tecnologia

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

Outline

- Simulation Setup and Details
- The Observables
- Exploratory Data Analysis
- Conclusions

Simulation Setup and Details

Simulation
Setup

- Jewel 2.2.0
- LHAPDF 5.9.1
- Docker image publicly available at
  - https://hub.docker.com/r/mcromao/jewel
  - Dockerfile: https://github.com/Strong2020-JetQGP/dockerfiles

- Usage examples:
  - `docker run --rm -v $PWD:$PWD -w $PWD --user $(id -u):$(id -g) mcromao/jewel:latest jewel-2.2.0-simple`
  - `docker run --rm -v $PWD:$PWD -w $PWD --user $(id -u):$(id -g) mcromao/jewel:latest jewel-2.2.0-simple my_card.dat`

Simulation
Details

- 320k events of both vacuum and simple (medium) simulation
- Kinematics:
  - CM Energy = 5020 GeV
  - PTMIN = 40 GeV
  - PTMAX = 250 GeV
  - ETAMAX = 2.5
  - No recoils
- Medium:
  - TAUI = 0.4 fm/c
  - TI = 440 MeV
  - TC = 170 MeV
  - Centrality = 0-10%

The Observables

The Observables

Jet 4-momentum

- Computed from the reconstructed anti-kt jets
  - η (`eta`)
  - ϕ (`phi`)
  - p_T (`pt`)
  - Mass (`mass`)
- And the number of constituents `nconst`

The Observables

Angularities 1408.3122

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \left(\frac{\Delta R_{i,\text{jet}}}{R_0} \right)^\beta$$

$$\Delta R_{i,j} = \sqrt{(\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2}$$

κ	β	Expression	In Code	Comments
0	1	$\frac{1}{N_{const}} \sum_i \Delta R_i$	`mr`	$\frac{1}{N_{const}}$ for mean
0	2	$\frac{1}{N_{const}} \sum_i \Delta R_i^2$	`mr2`	$\frac{1}{N_{const}}$ for mean
1	1	$\sum_i z_i \Delta R_i$	`rz`	Also known as g
1	2	$\sum_i z_i \Delta R_i^2$	`r2z`	
2	0	$\frac{1}{N_{const}} \sum_i z_i^2$	`mz2`	$p_T D = N_{const} \sqrt{mz2}$

The Observables

Subjettiness 1011.2268

$$\tau_N = \frac{\sum_{i=1}^{N} p_T^i \min(\Delta R_{1,i}, \ldots, \ \Delta R_{N,i})}{R_0 \sum_{i=1}^{N} p_T^i}$$

$$\tau_{N,N-1} = \frac{\tau_N}{\tau_{N-1}}$$

- As implemented by `fastjet-contrib Nsubjettiness`
- N=1,...5 for τ (`tau1`,... , `tau5`)
- $\tau_{2,1}$, $\tau_{3,2}$ (`tau2tau1`, `tau3tau2`)

The Observables

Jet Charge 1209.2421

$$Q_\kappa = \sum_{i \in \text{jet}} z_i^\kappa Q_i$$

- Following CMS we used kappa = 0.3, 0.5, 0.7, 1.0
  - `jetcharge03`, `jetcharge05`, `jetcharge07`, `jetcharge10`

The Observables
Soft-Drop quantities 1402.2657

Implemented by `fastjet-contrib RecursiveTools`
w/ zcut = 0.1 and beta = 0 (mMDT)

- Recursively declusters the Jet branching history and discards the resulting sub-jets until the current splitting fulfills the SD condition

$$\frac{\min[p_{T,i}, p_{T,j}]}{p_{T,i} + p_{T,j}} > z_{cut} \Delta R_{i,j}$$

- This defines three observables
  - n_{SD} (nSD) number of splits until condition is met
  - z_g (zg) fraction of the momentum of least energetic subject at splitting where condition is met
  - R_g (Rg) the radial separation at the splitting where the conditions is met

The Observables

Dynamical Grooming quantities 1911.00375

- Implementation adapted from Alba Soto-Ontoso's code

$$\kappa^{(a)} = \frac{1}{p_{T,jet}} \max_{i \in \mathrm{C/A\ seq}} \left[z_i (1 - z_i) p_{T,i} \left(\frac{\Delta R_i}{R_0} \right) \right]$$

- Three interesting possibilities for a
  - a=2 TimeDrop
  - a=1 kTDrop
  - a=0 zDrop (a=0.1)
- Just like SD, this defines observables z_g, R_g for each possibility and the value of κ
  - `deltaR_TD/ktD/zD`, `kappa_TD/ktD/zD`, `zg_TD/ktD/zD`

The Observables

Summing up

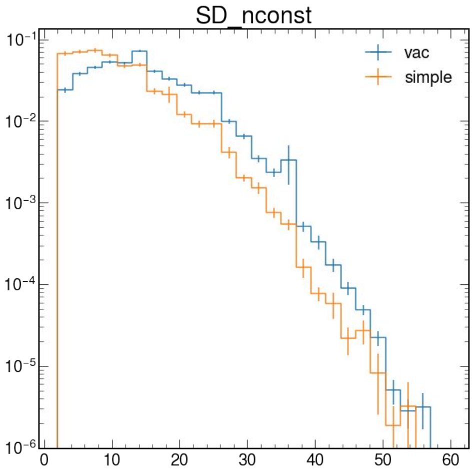
- 4-momenta
  - eta, phi, pt, mass, nconst
- Angularities
  - mr, mr2, rz, r2z, mz2, ptd
- Subjetiness
  - tau1, …, tau5, tau3tau2, tau2tau1
- Jet Charges
  - jetcharge03, jetcharge05, jetcharge07, jetcharge10
- Soft-Drop Quantities
  - nSD, zg, Rg
- Dynamical Grooming Quantities
  - kappa, zg, Rg (for a=2,1,0)

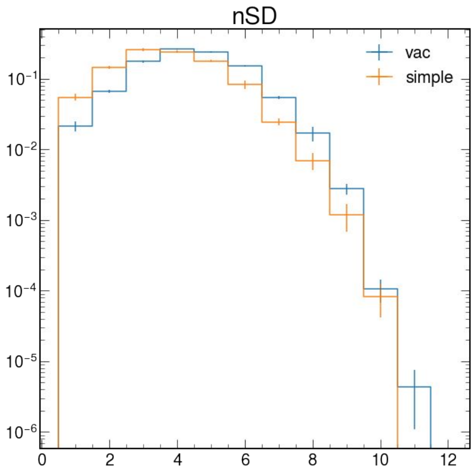
Exploratory Data Analysis

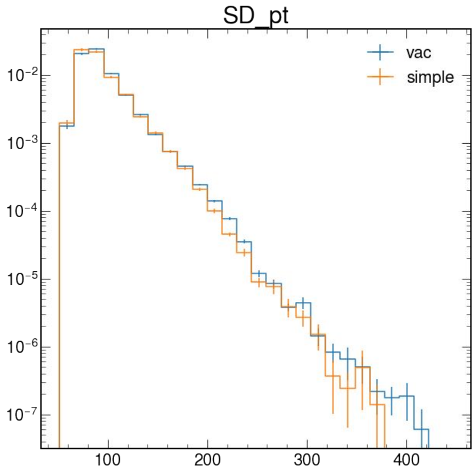
Exploratory Data Analysis
Preliminaries

- Jets were reconstructed with anti-kt (R=0.4) using `fastjet 3.3.4`
  - Full jets without subtraction
- A docker image with all required dependencies is available
  - https://hub.docker.com/r/mcromao/processors
- Per Jet observables saved to a TTree
- Analysis is done after we process the HEPMC into the ROOT TTree
- For the analysis:
  - Quantities computed with the SD groomed version of jet
  - $p_T >$ 80 GeV, and Rg, zg $>$ 0 (dropped SD untagged jets)

Exploratory Data Analysis
Some distributions

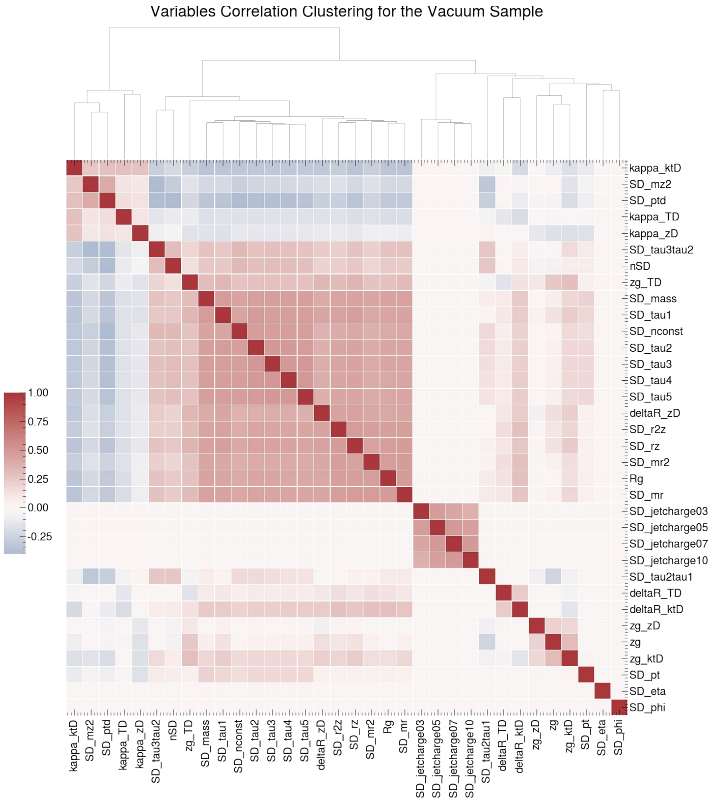


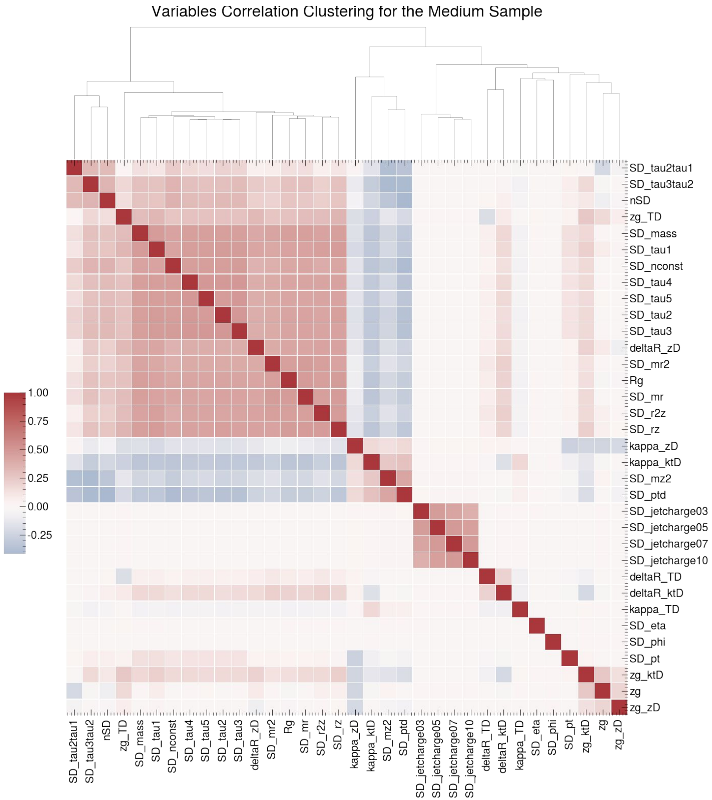




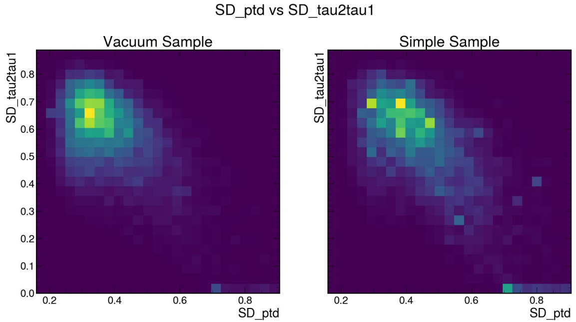
33 variables => Need more systematic way of studying them and their relations

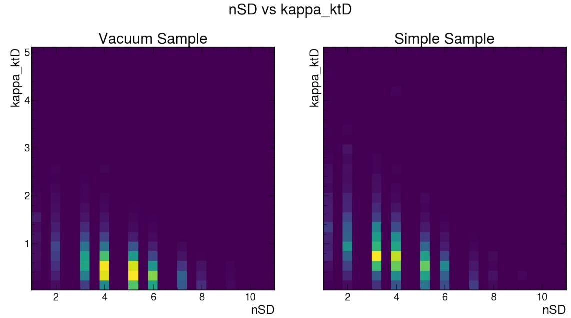
Exploratory Data Analysis: Correlations

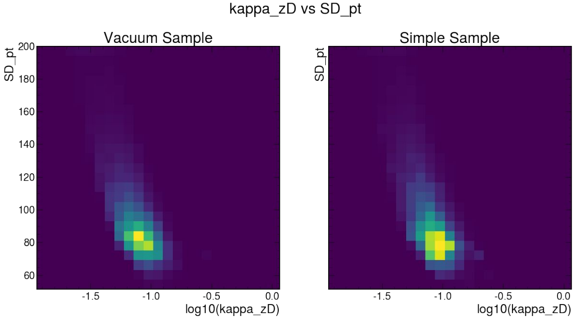


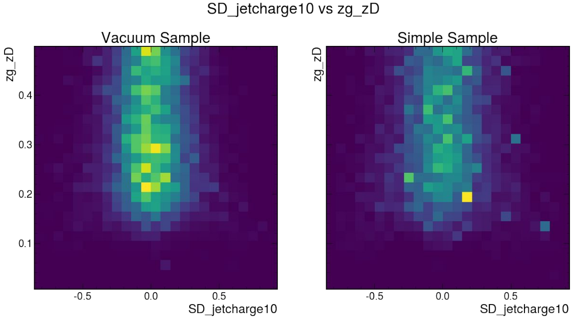


Exploratory Data Analysis: Correlations
Some 2d histograms



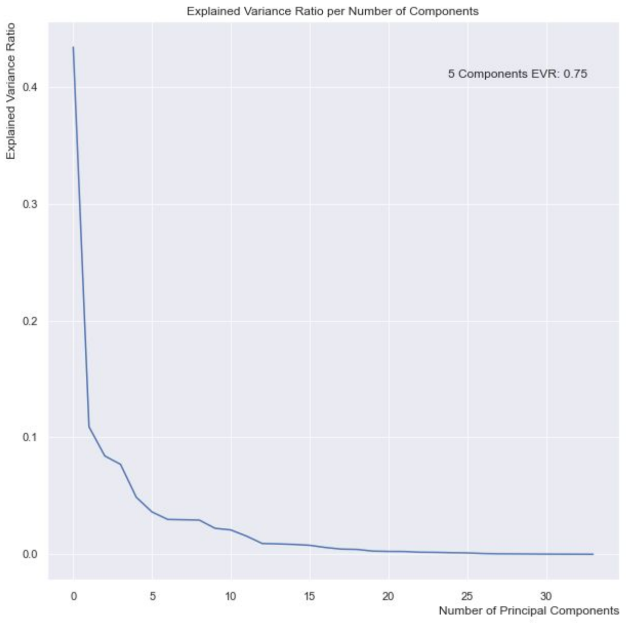






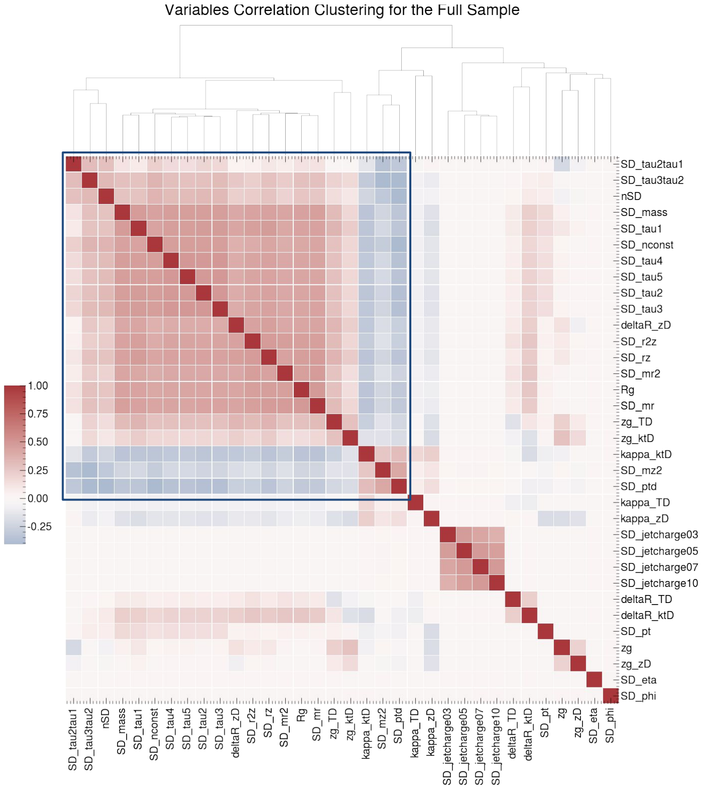
Exploratory Data Analysis: Correlations
Principal Component Analysis

- Clustering over the covariance matrix suggests most of the variables are fairly collinear
  - Perform Principal Component Analysis: disentangle linear correlations
  - How many Principal Components are there?
  - 75% of the covariance is explained by only 5 components

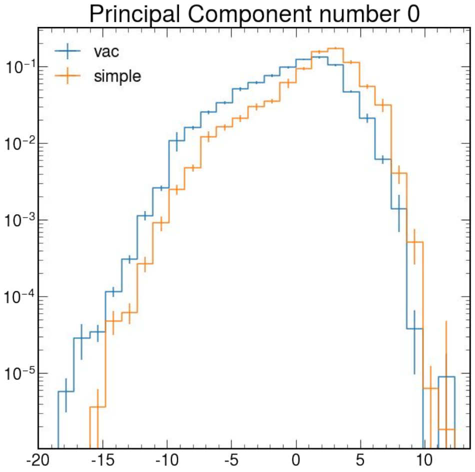


Exploratory Data Analysis: Correlations

Principal Component Analysis

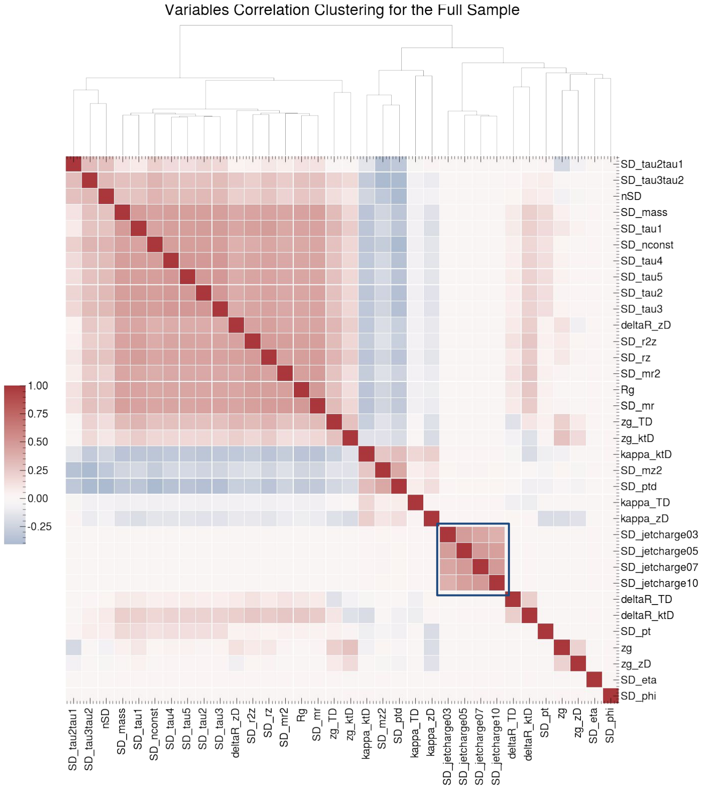


Component 0	
SD_ptd	0.202301
kappa_ktD	0.194061
SD_mz2	0.144330
zg_ktD	-0.124228
nSD	-0.169223
SD_tau3tau2	-0.172789
zg_TD	-0.186161
SD_mr2	-0.224263
deltaR_zD	-0.225578
SD_r2z	-0.233269
Rg	-0.237735
SD_mr	-0.241182
SD_tau5	-0.245185
SD_nconst	-0.245295
SD_tau1	-0.245703
SD_mass	-0.246122
SD_tau2	-0.246489
SD_rz	-0.246701
SD_tau4	-0.247284
SD_tau3	-0.248432

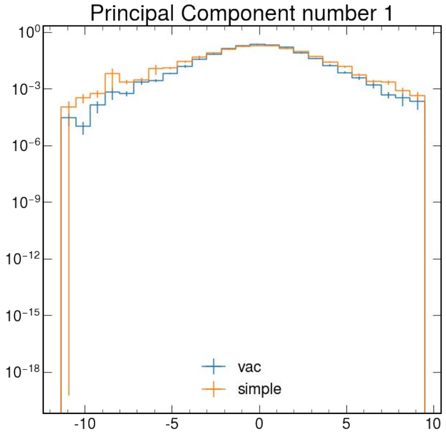


Exploratory Data Analysis: Correlations

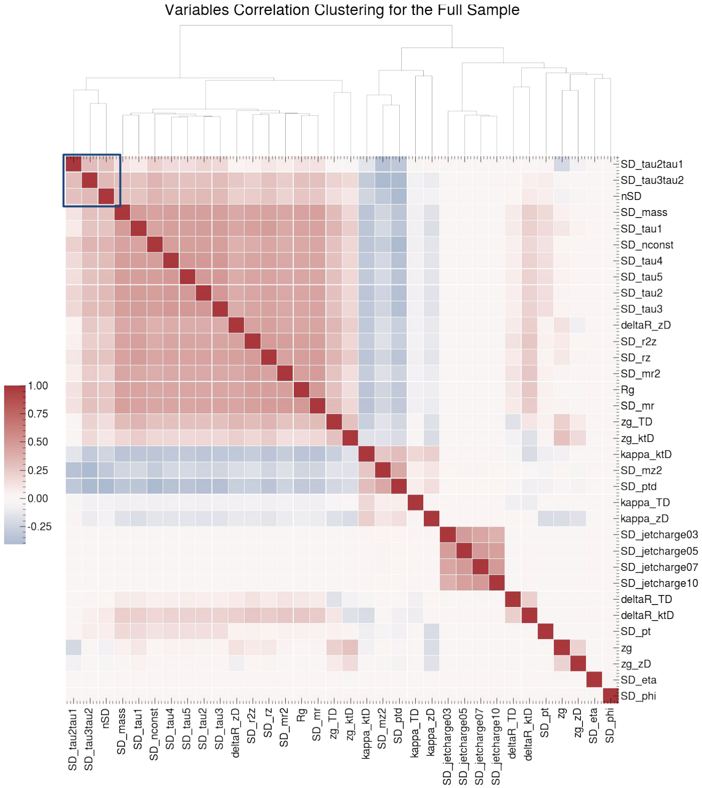
Principal Component Analysis



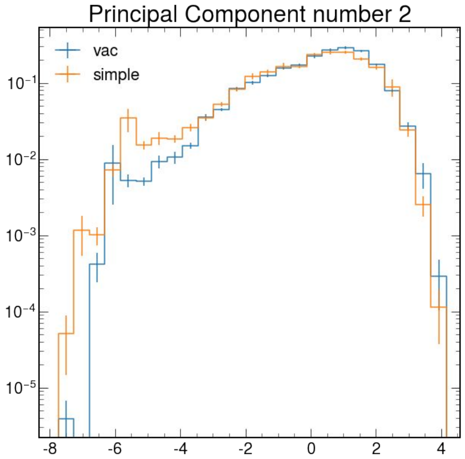

```
Component 1
SD_jetcharge03    -0.474703
SD_jetcharge10    -0.490372
SD_jetcharge07    -0.512734
SD_jetcharge05    -0.516057
```



Exploratory Data Analysis: Correlations
Principal Component Analysis

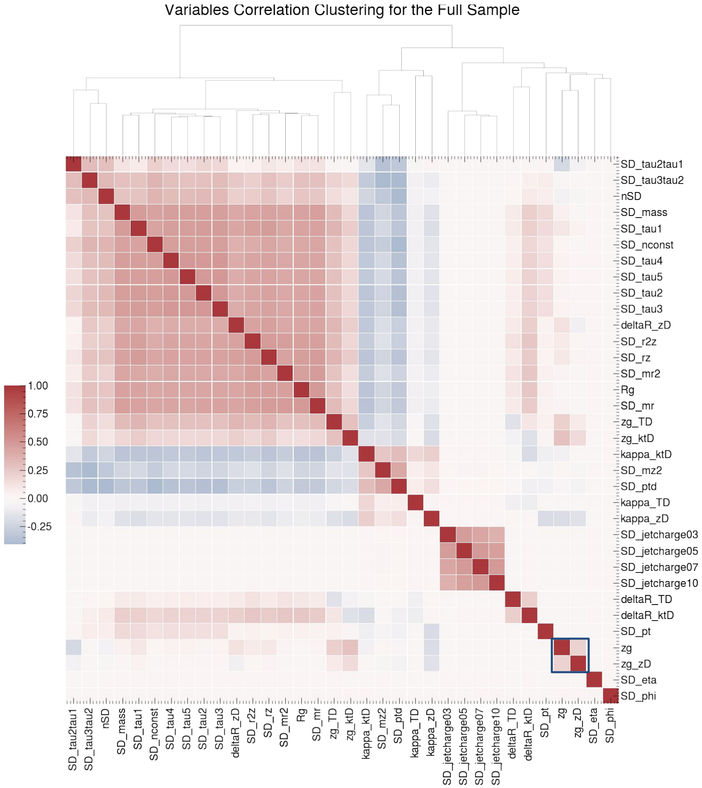


Component 2	
SD_tau2tau1	0.513019
nSD	0.296634
SD_tau3tau2	0.274460
SD_mz2	-0.318791
zg	-0.410864

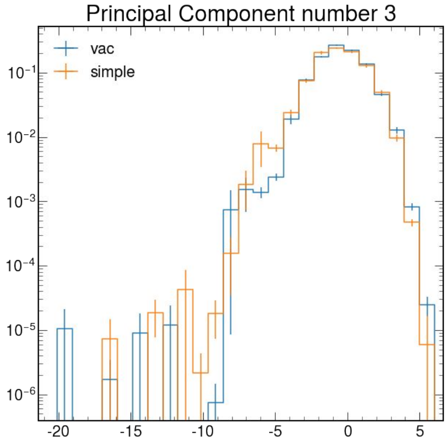


Exploratory Data Analysis: Correlations

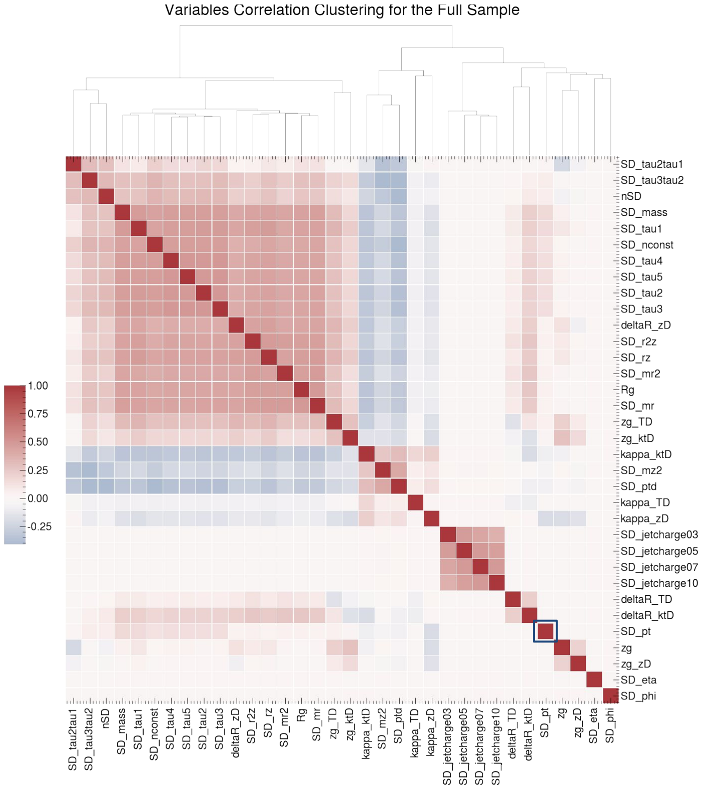
Principal Component Analysis



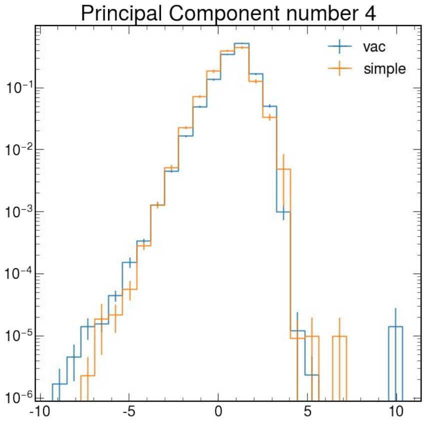

```
Component 3
zg_ktD          0.324438
zg_zD           0.255778
SD_pt           0.244787
zg              0.234625
SD_mr2         -0.207843
SD_ptd         -0.210595
SD_mz2         -0.239358
deltaR_TD      -0.316737
kappa_zD       -0.370820
deltaR_ktD     -0.385327
```



Exploratory Data Analysis: Correlations
Principal Component Analysis




```
Component 4
SD_pt    -0.648165
```

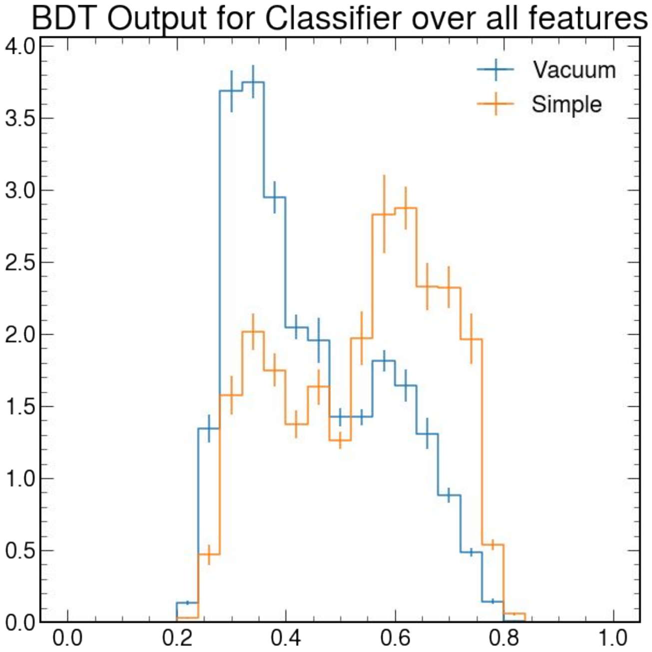


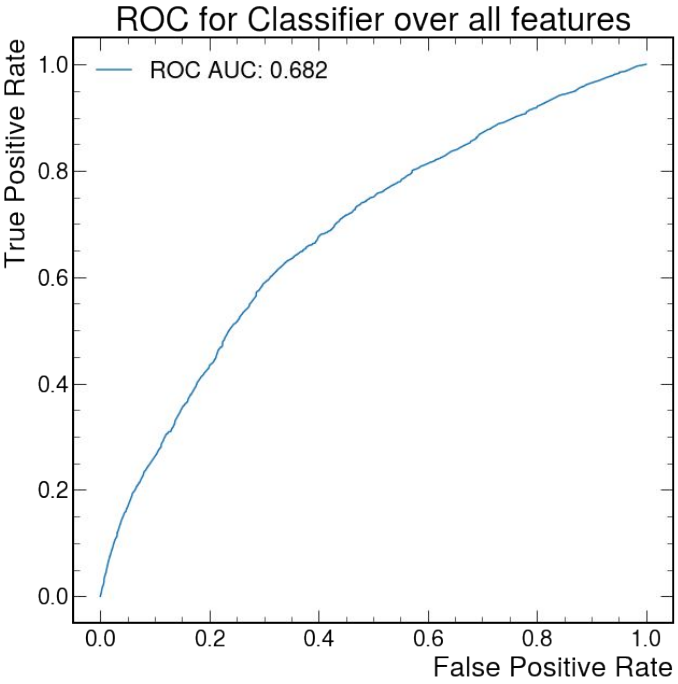
Exploratory Data Analysis
Machine Learning Study on Discriminative Power

- Correlations only highlight pairwise linear relations
- How does each variable help to discriminate between vacuum and medium jets
  - By themselves
  - In combination with other variables
- Train BDT on the whole set to assess maximum discriminative power (quantified by area of ROC). Then:
  - Train BDT for each variable in isolation
  - Train BDT for each pair of variables

Exploratory Data Analysis

Machine Learning Study on Discriminative Power



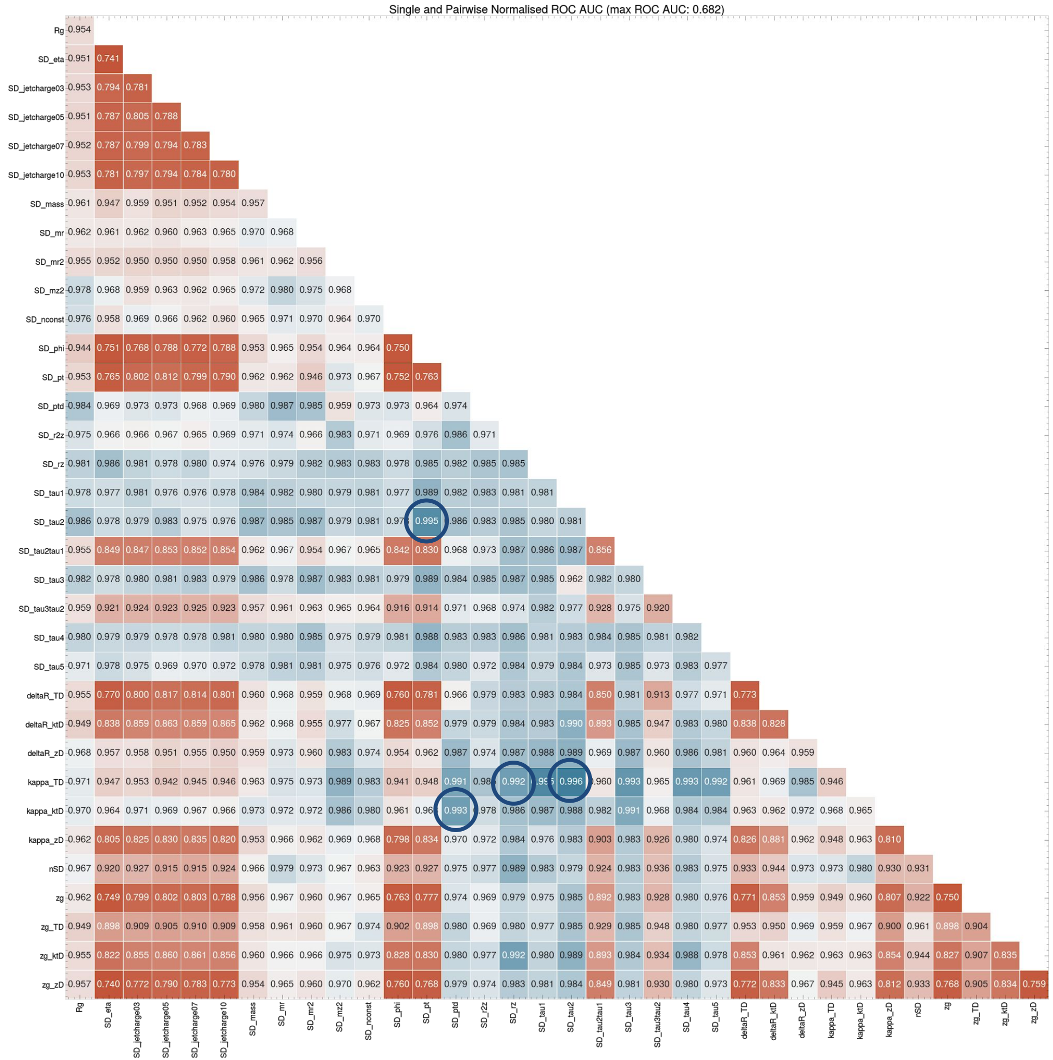


Exploratory Data Analysis

Machine Learning Study on Discriminative Power

Audience question time:

What pair of variables has the highest vacuum-medium discriminative power?



Bingo for:

- (tau2, kappaTD)
- (tau2, pt)
- (rz, kappaTD)
- (ptD, kappaktD)

$$\kappa^{(a)} = \frac{1}{p_{T,jet}} \max_{i \in \text{C/A seq}} \left[z_i (1 - z_i) p_{T,i} \left(\frac{\Delta R_i}{R_0} \right) \right]$$

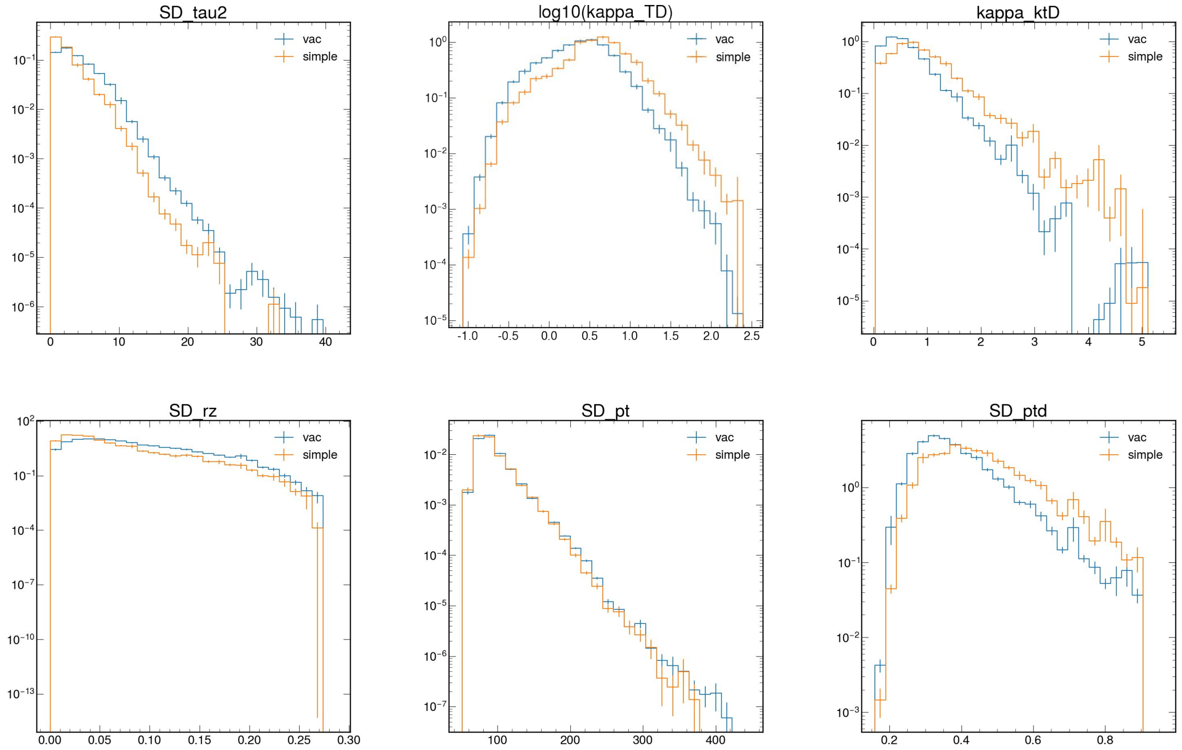
$$rz = \sum_{i \in \text{jet}} z_i \Delta R_{i,jet}$$

$$\tau_N = \frac{\sum_{i=1}^{N} p_T^i \min(\Delta R_{1,i}, \ldots, \Delta R_{N,i})}{R_0 \sum_{i=1}^{N} p_T^i}$$

$$p_T D = \frac{\sqrt{\sum_{i \in jet} p_{T,i}^2}}{\sum_{i \in jet} p_{T,i}}$$

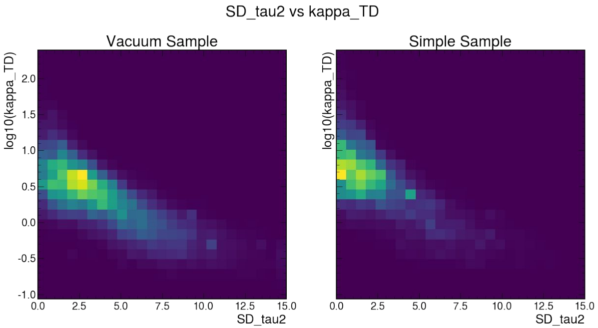
Exploratory Data Analysis

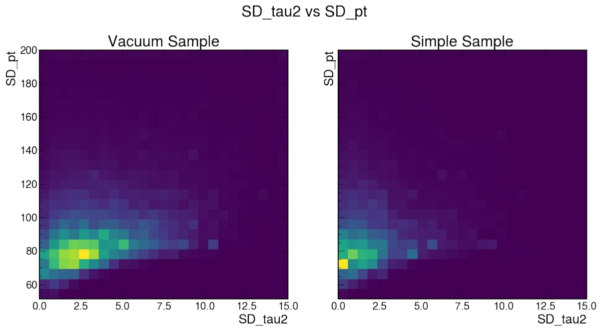
Machine Learning Study on Discriminative Power

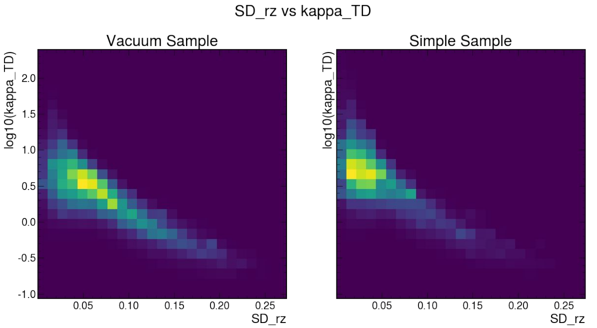


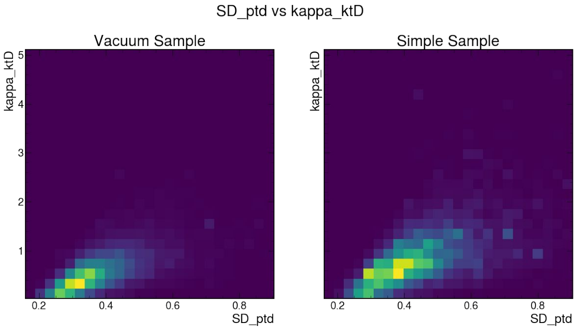
Exploratory Data Analysis

Machine Learning Study on Discriminative Power









Conclusions

Conclusions

- We studied an ensemble of jet observables on Jewel samples
- A correlation studied, in addition to a PCA, was carried out to disentangle the linear relations amongst the observables
  - Many observables appear to be highly correlated, with evidence for considerably lower intrinsic dimensionality
- A ML classifier was used to pair up observables in terms of their vacuum-medium discriminative power
  - One can saturate the maximal performance with a handful of pairs (notably involving tau2, kappa_TD, ptD)

Future Work and Directions

- Replicate these steps with other generators
  - Prepare Docker images with other generators
  - Settle on the observables list
- Check if the discrimination power and the correlations are robust, i.e. generator independent
  - Could we teach a classifier to guess from which generator a sample comes from?
  - Identify specific features of generators to understand if they correspond to specific behaviours that can distinguish physical models

Thanks!

Follow up:

mcromao@lip.pt