Measurement of W/Z production in the high p_T, boosted region at CMS

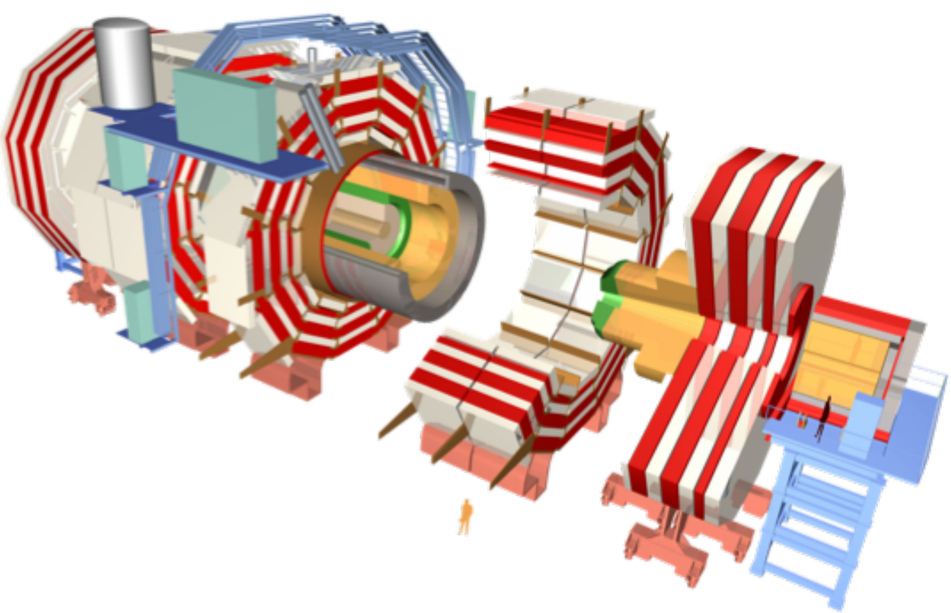
Jindřich Lidrych, Paolo Gunnellini, Hannes Jung
Deutsches Elektronen-Synchrotron (DESY)

Deutsche Physikalische Gesellschaft
Frühjahrstagung 2018, Würzburg

19th - 23rd March 2018

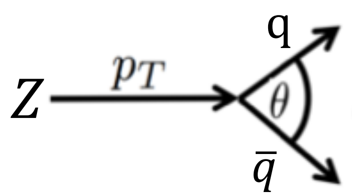
Contents

- Introduction
  - Boosted object
  - Motivation for SM measurements with boosted objects
- Analysis strategy
  - MC study
  - MC sample vs data
  - Multivariate data analysis
- Summary and Outlook



Boosted objects

- Quarks from fully hadronic decays of heavy Standard Model (SM) particles (i.e. W, Z, H) merge into single jet with characteristic substructure



$$\Delta R \approx \theta \approx 2\frac{m_Z}{p_\mathrm{T}}$$

- Jet substructure provides powerful handles to suppress many SM backgrounds

- Boosted objects can be used for SM & Beyond SM (BSM) measurements

SM measurements:

- Measurements with boosted W, Z and Higgs as the test of SM predictions at the extreme kinematic region

Beyond SM searches:

- Many BSM models predict new heavy particles that often decay to boosted heavy SM particles

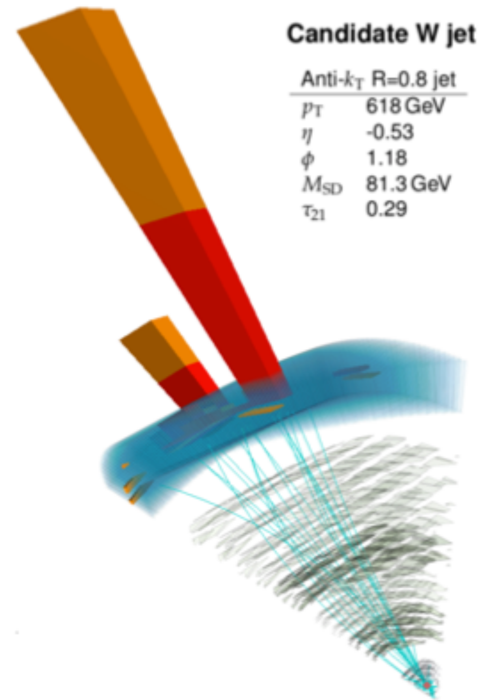
SM measurements with boosted objects

- Boosted objects can provide more stringent SM test
  - Sensitive to higher order corrections
- Boosted regime is where the background is for many searches
  - Theory understanding
- Better reconstruction precision compared to classical reconstruction method in the high p_T – region
  - Fully Hadronic vs leptonic decay of W/Z with the high p_T

Recent progress in:

- Theory behind jet substructure
- Detector performance
- Event reconstruction
- Machine learning in HEP

Precise measurements with W, Z and Higgs in boosted regime



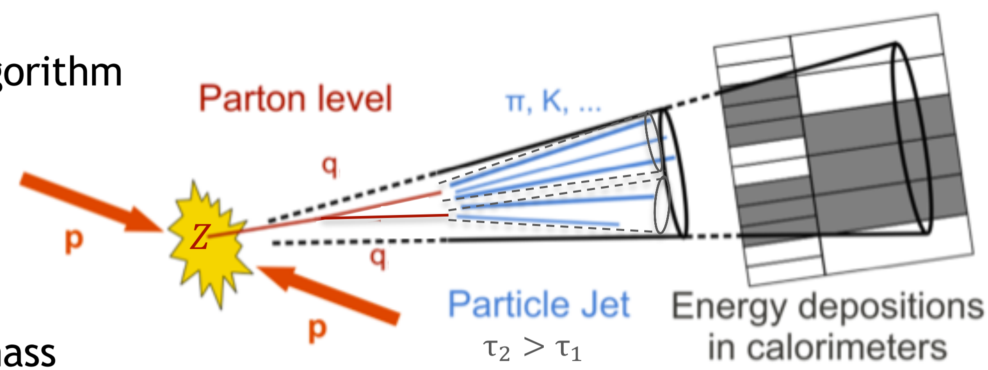
Analysis strategy

Goal:

- Inclusive cross-section of the boosted W/Z
- Azimuthal correlations of the boosted W/Z with additional jets

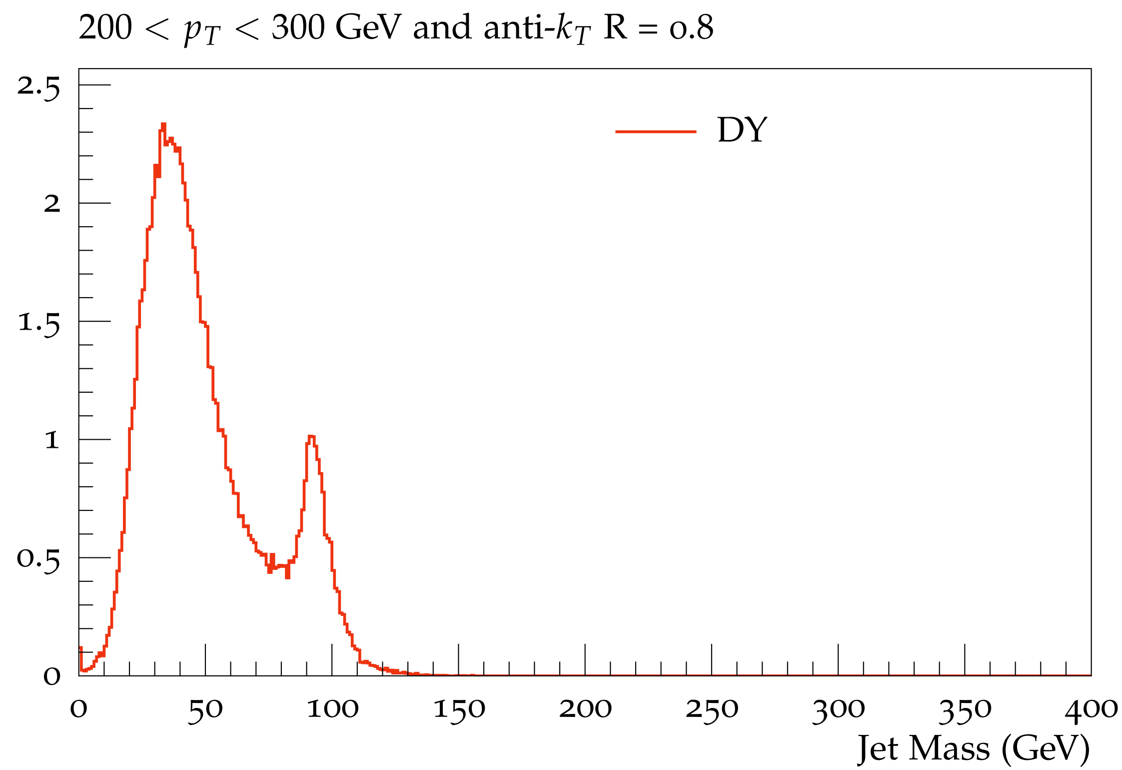
Today, we present Monte Carlo (MC) study for signal-background separation

- MC generator: Madgraph + Pythia (CMS official samples)
  - Processes: DY → qq, W → qq, Higgs → bb, QCD
- Data sample: pp collisions at $\sqrt{s}$ = 13 TeV
  - 2016, Run G
- Jet kinematics:
  - Reconstructed with anti-k_T (R=0.8) algorithm
  - p_T > 200 GeV
  - $|\eta| < 2.4$
- Considered jet variables:
  - Jet p_T, mass, energy, soft drop (SD) mass
  - N-subjettinnes: $\tau_{21} = \tau_2/\tau_1$, $\tau_{31} = \tau_3/\tau_1$ and $\tau_{32} = \tau_3/\tau_2$
  - Subjets p_T and mass, ΔR between 2 subjets, N-candidates



Signal/background cross-section: jet variables

- MC generator: Pythia
- Process:
  - $q\bar{q} \to \gamma^*/Z^0\ g$
  - $qg \to \gamma^*/Z^0\ q$
- Mass cut in MC: $\hat{m} > 4\text{GeV}$
- Jet with with anti-k_T (R=0.8) and $200 < p_T < 300$ GeV
- Peak around Z mass
- In reality, background processes are present:
  - QCD multijets
  - Hadronic W decays



Signal/background cross-section: jet variables

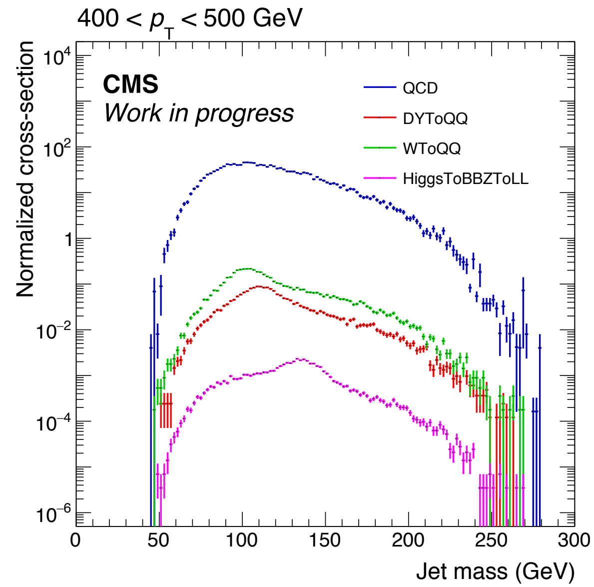
- Mass peak around Z/W/Higgs mass
- QCD – no structure

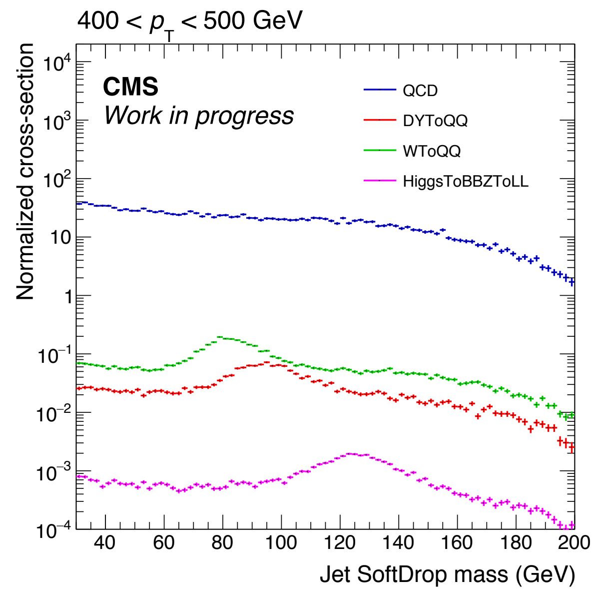
$$\frac{\text{Signal(W/Z)}}{\text{Background(QCD)}} \sim 0.01$$

- SoftDrop Mass algorithm is used to decluster the jet in two subjets removing soft radiation
- SoftDrop condition:

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} < z_{cut} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

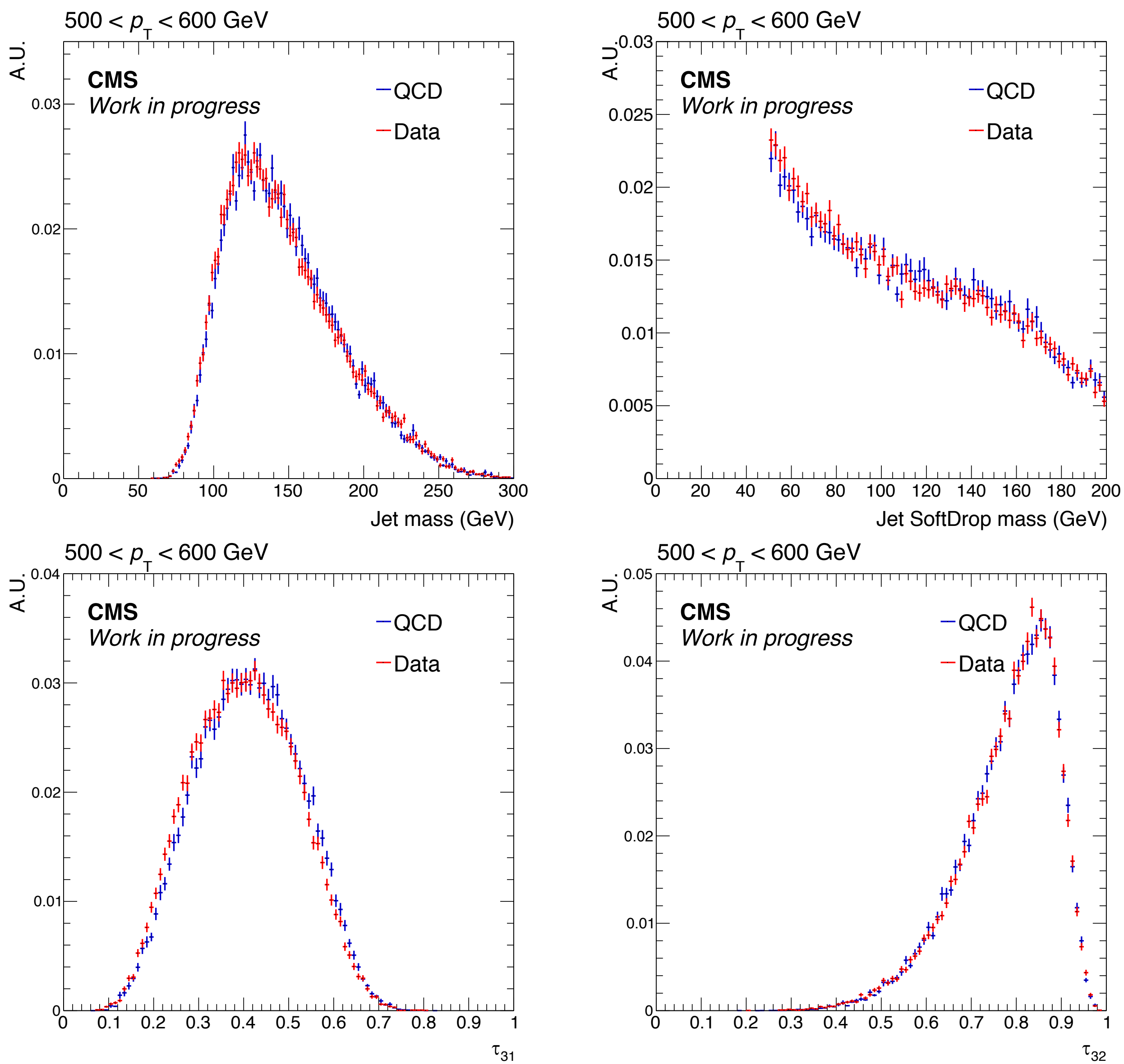
$\rightarrow$ No separation





MC sample: QCD vs Data - jet variables

- All variables are normalized to unity
- Data vs MC: good agreement
- More data is available -> reduce relatively large statistical uncertainties

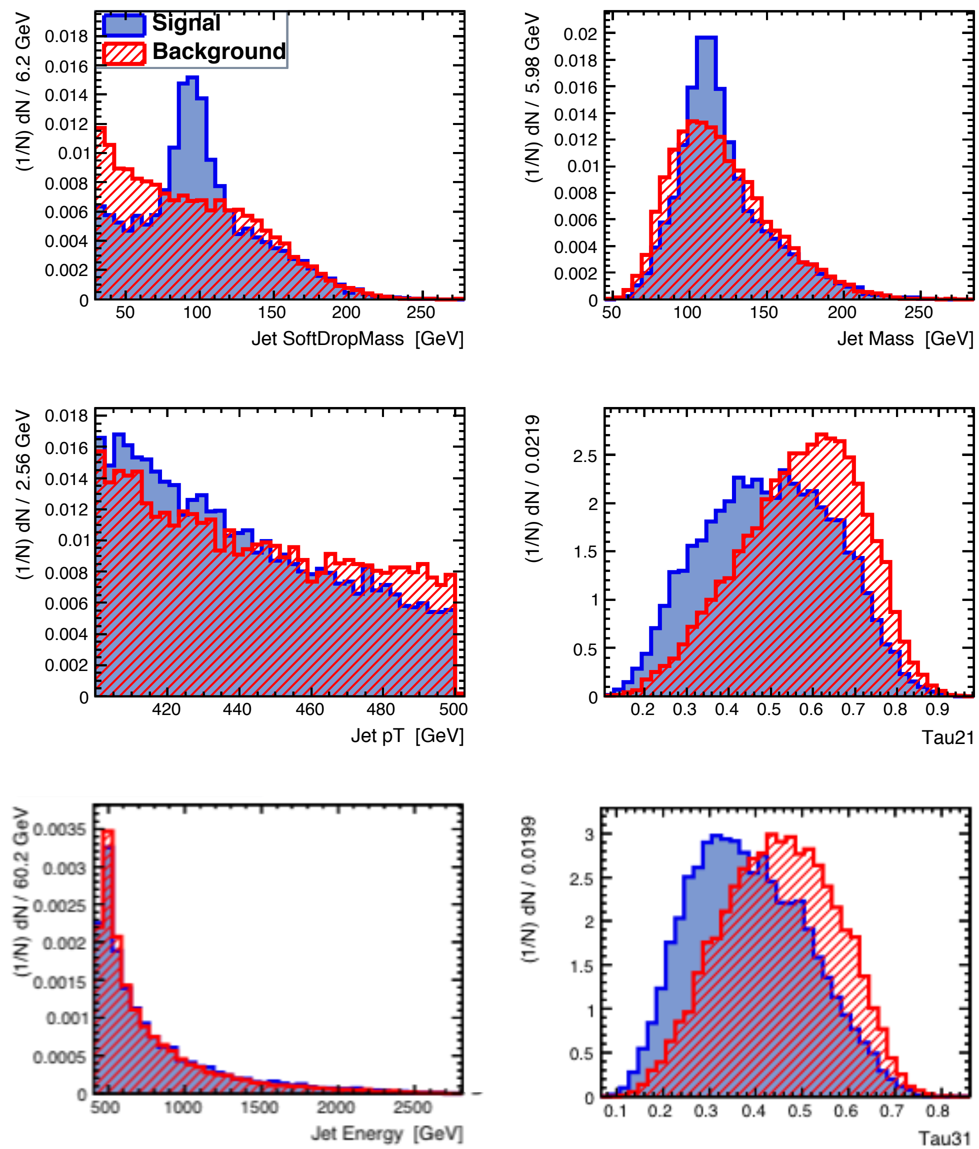


Machine learning for signal-background separation

- Toolkit for Multivariate analysis (TMVA) is used to find suitable variables and selection criteria to discriminate background

- Input: MC sample
  - Signal: $\mathrm{DY} \rightarrow qq$
  - Background: QCD

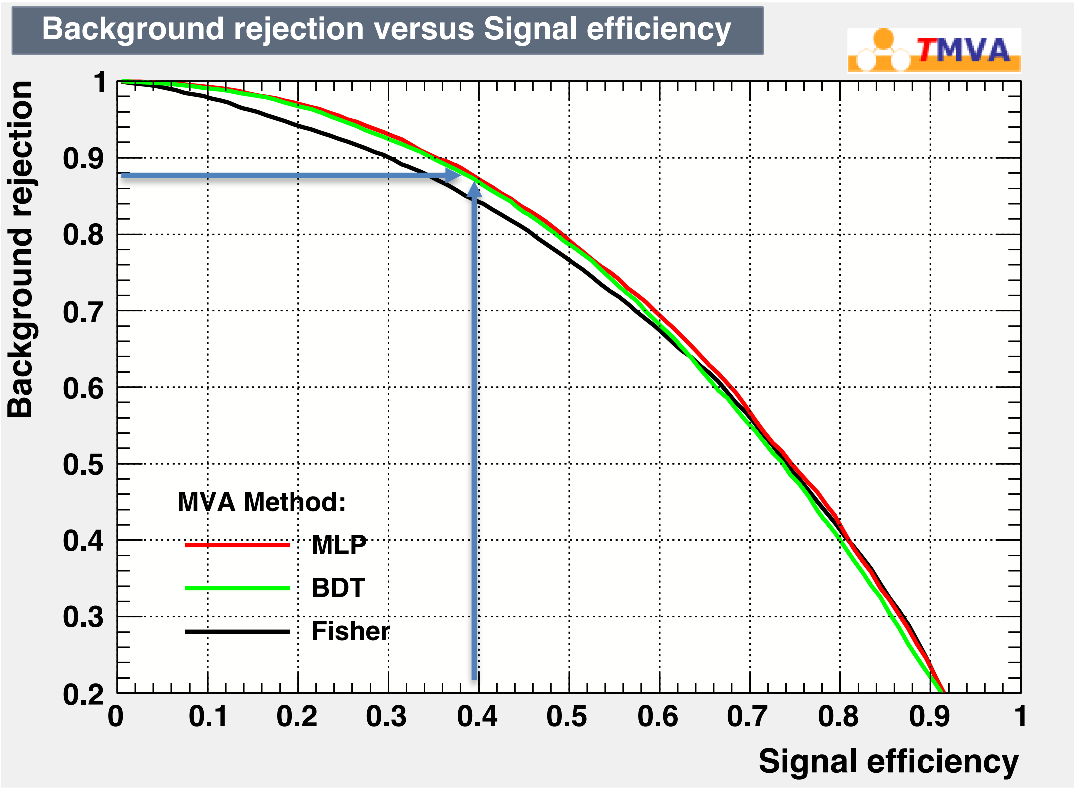
- Some of the jet variables can help to separate the signal and background



TMVA: signal-background separation

TOP 5 variable for signal-backgr. sep.

```
: --------------------------------------
: Rank : Variable        : Discr. power
: --------------------------------------
:    1 : Tau31               : 3.433e-02
:    2 : Tau21               : 2.946e-02
:    3 : Number of PF candidates :
                               2.128e-02
:    4 : jetMassSoftDrop : 2.472e-03
:    5 : jetPt               : 2.429e-03
: --------------------------------------
```



- Without any optimization, ~88% background can be rejected for ~40% signal selection with BDT, MLP
  - →better performance is needed
- Next step: use **DeepAK8 tagger** developed by CMS
  - A multi-class classifier for top, W, Z, Higgs and QCD jets based on standard anti-k_T (R=0.8) jets
  - Deep neural networks architecture using directly low-level inputs (PF candidates)
  - Exploit substructure and flavour in one go

Conclusion & Outlook

Conclusion:

- Measurements with boosted W/Z as the test of SM predictions at the extreme kinematic region was presented
- Preliminary MC study for measurement of boosted W/Z
- MC sample vs Data: good agreement of jet variable and jet substructure
- Multivariate data analysis can help to distinguish boosted W/Z from the QCD background

Outlook:

- Use DeepAK8 tagger for better signal-background separation
- Perform analysis with data: pp collisions at $\sqrt{s}$ = 13 TeV

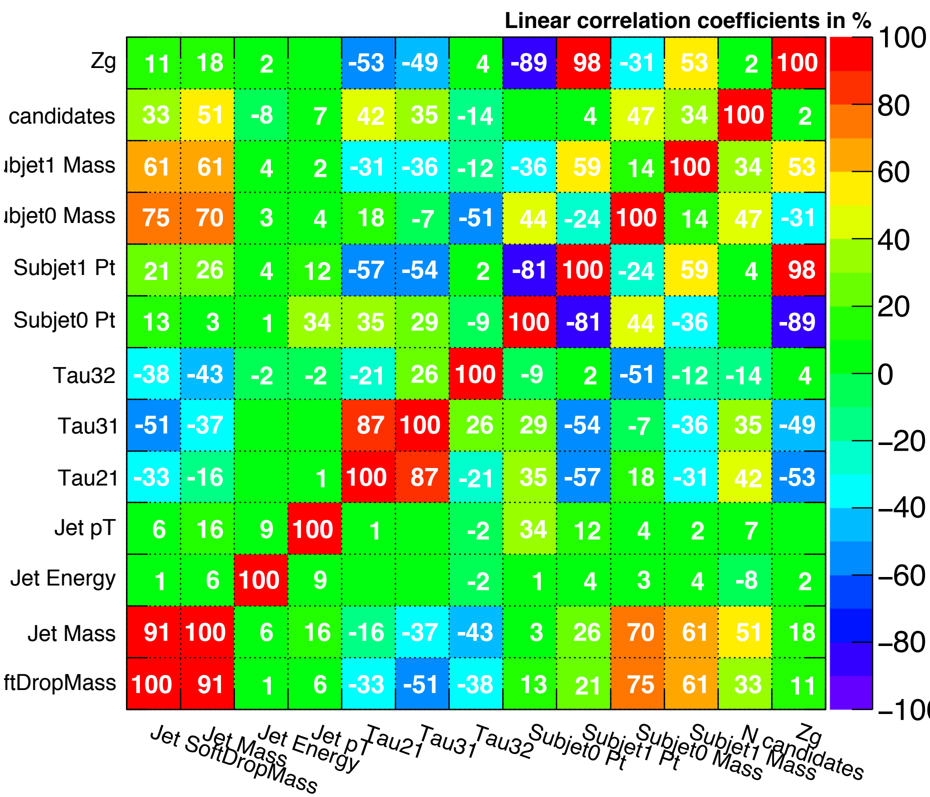
Thank you for your attention!

Back-up slides

Back-up: TMVA: signal-background separation

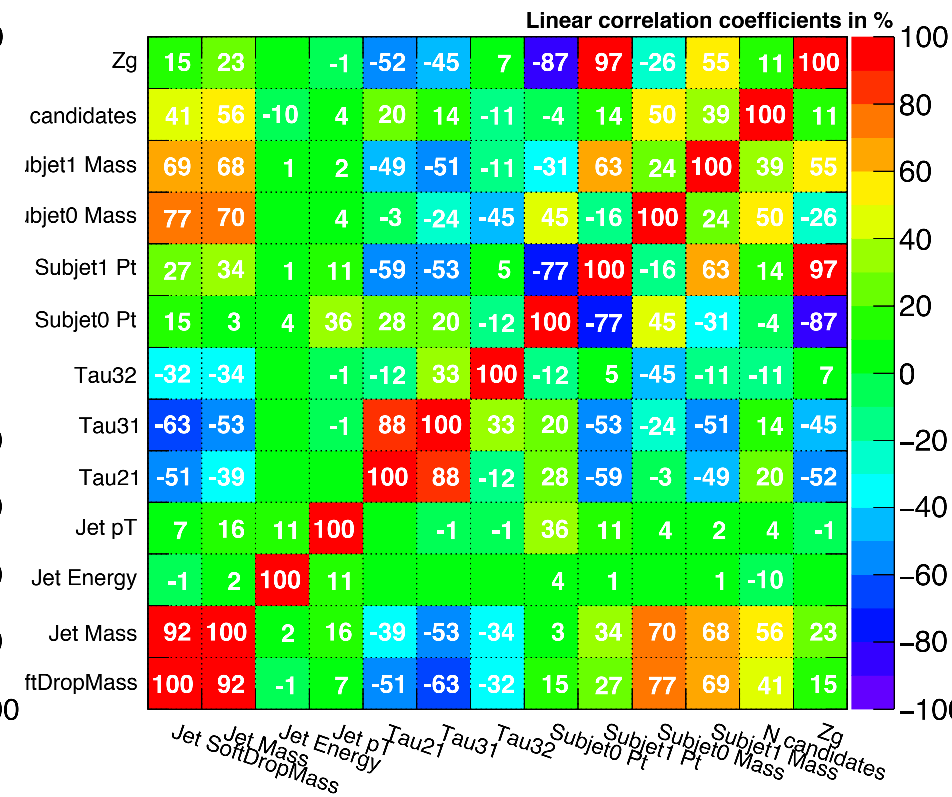
- Correlations between variables
- Different correlations between signal and background

Correlation Matrix (signal)



	Jet SoftDropMass	Jet Mass	Jet Energy	Jet pT	Tau21	Tau31	Tau32	Subjet0 Pt	Subjet1 Pt	Subjet0 Mass	Subjet1 Mass	N candidates	Zg
Zg	11	18	2		-53	-49	4	-89	98	-31	53	2	100
candidates	33	51	-8	7	42	35	-14		4	47	34	100	2
ubjet1 Mass	61	61	4	2	-31	-36	-12	-36	59	14	100	34	53
ubjet0 Mass	75	70	3	4	18	-7	-51	44	-24	100	14	47	-31
Subjet1 Pt	21	26	4	12	-57	-54	2	-81	100	-24	59	4	98
Subjet0 Pt	13	3	1	34	35	29	-9	100	-81	44	-36		-89
Tau32	-38	-43	-2	-2	-21	26	100	-9	2	-51	-12	-14	4
Tau31	-51	-37			87	100	26	29	-54	-7	-36	35	-49
Tau21	-33	-16		1	100	87	-21	35	-57	18	-31	42	-53
Jet pT	6	16	9	100	1		-2	34	12	4	2	7	
Jet Energy	1	6	100	9			-2	1	4	3	4	-8	2
Jet Mass	91	100	6	16	-16	-37	-43	3	26	70	61	51	18
ftDropMass	100	91	1	6	-33	-51	-38	13	21	75	61	33	11

Correlation Matrix (background)



	Jet SoftDropMass	Jet Mass	Jet Energy	Jet pT	Tau21	Tau31	Tau32	Subjet0 Pt	Subjet1 Pt	Subjet0 Mass	Subjet1 Mass	N candidates	Zg
Zg	15	23		-1	-52	-45	7	-87	97	-26	55	11	100
candidates	41	56	-10	4	20	14	-11	-4	14	50	39	100	11
ubjet1 Mass	69	68	1	2	-49	-51	-11	-31	63	24	100	39	55
ubjet0 Mass	77	70		4	-3	-24	-45	45	-16	100	24	50	-26
Subjet1 Pt	27	34	1	11	-59	-53	5	-77	100	-16	63	14	97
Subjet0 Pt	15	3	4	36	28	20	-12	100	-77	45	-31	-4	-87
Tau32	-32	-34		-1	-12	33	100	-12	5	-45	-11	-11	7
Tau31	-63	-53		-1	88	100	33	20	-53	-24	-51	14	-45
Tau21	-51	-39			100	88	-12	28	-59	-3	-49	20	-52
Jet pT	7	16	11	100		-1	-1	36	11	4	2	4	-1
Jet Energy	-1	2	100	11				4	1		1	-10	
Jet Mass	92	100	2	16	-39	-53	-34	3	34	70	68	56	23
ftDropMass	100	92	-1	7	-51	-63	-32	15	27	77	69	41	15