Search for emerging jets in pp collisions at $\sqrt{s} = 13.6$ TeV with the ATLAS experiment

[arXiv:2505.02429]

IRN Terascale 2025

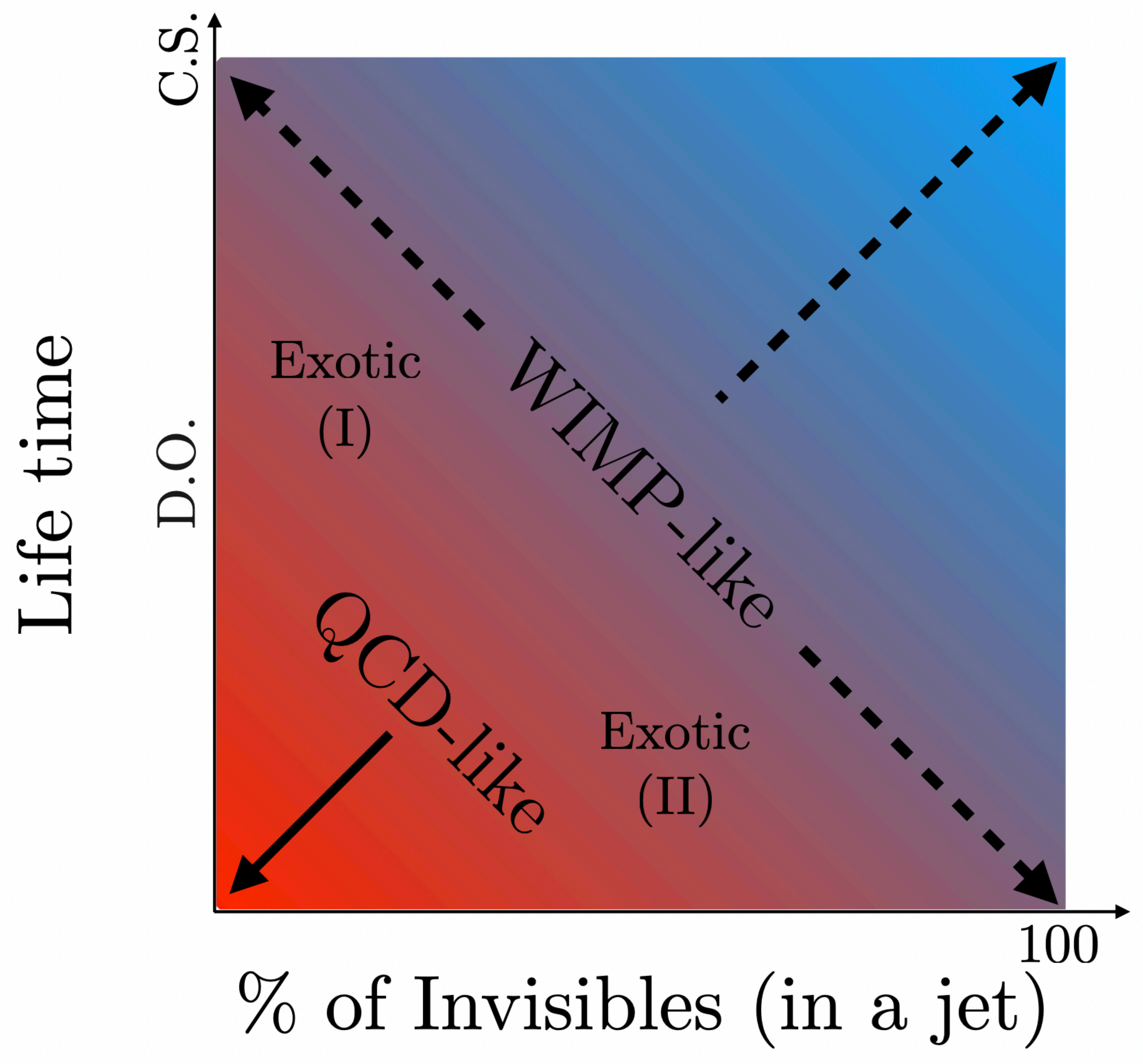
Thomas Wojtkowski, LPSC Grenoble

Dark QCD

- **Dark QCD** :
  - extension of the Standard Model (SM), provides Dark Matter candidates
  - gauge structure and particle content similar to QCD
  i.e N_f flavors of **dark quarks** q_D charged under SU(N_{C_D}), confinement at Λ_D
  - dark quarks can undergo parton shower, hadronize and form **dark hadrons**

- In dark QCD models, **interaction between dark QCD and SM particles** via new mediators :
  - especially, interaction between SM and dark quarks
  - through pp collisions, possible production of q_D leading to **dark jets**

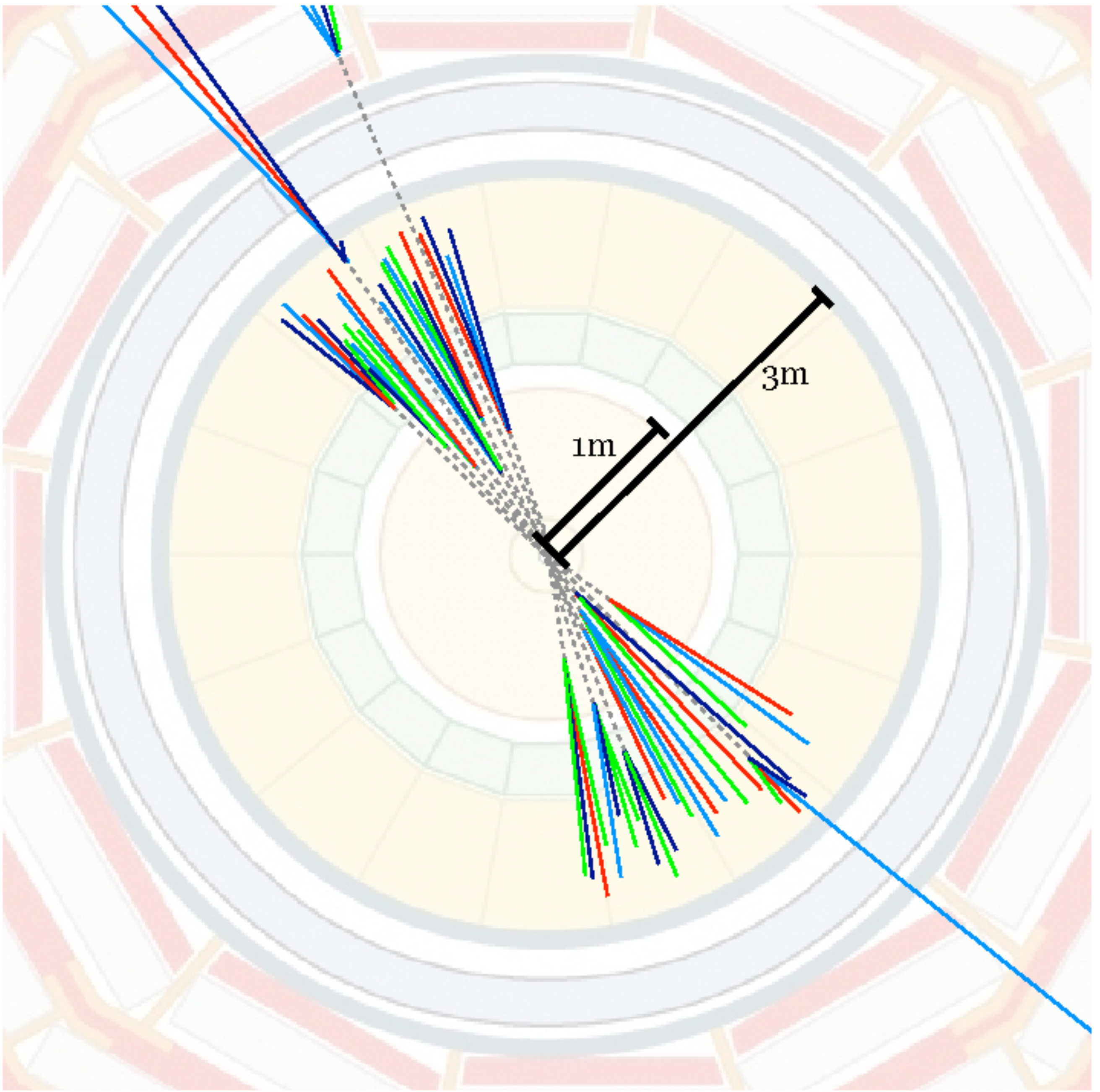
- **Several topologies for dark jets** depending on the fraction of stable dark hadrons and on the life-time of the unstable ones :
  « Dark jets » (QCD-like), « Semi-visible jets » (Exotic II)
  and « **Emerging jets** » (Exotic I) (or fully invisible : « WIMP-like »)



[arXiv:1712.09279]

Emerging jets

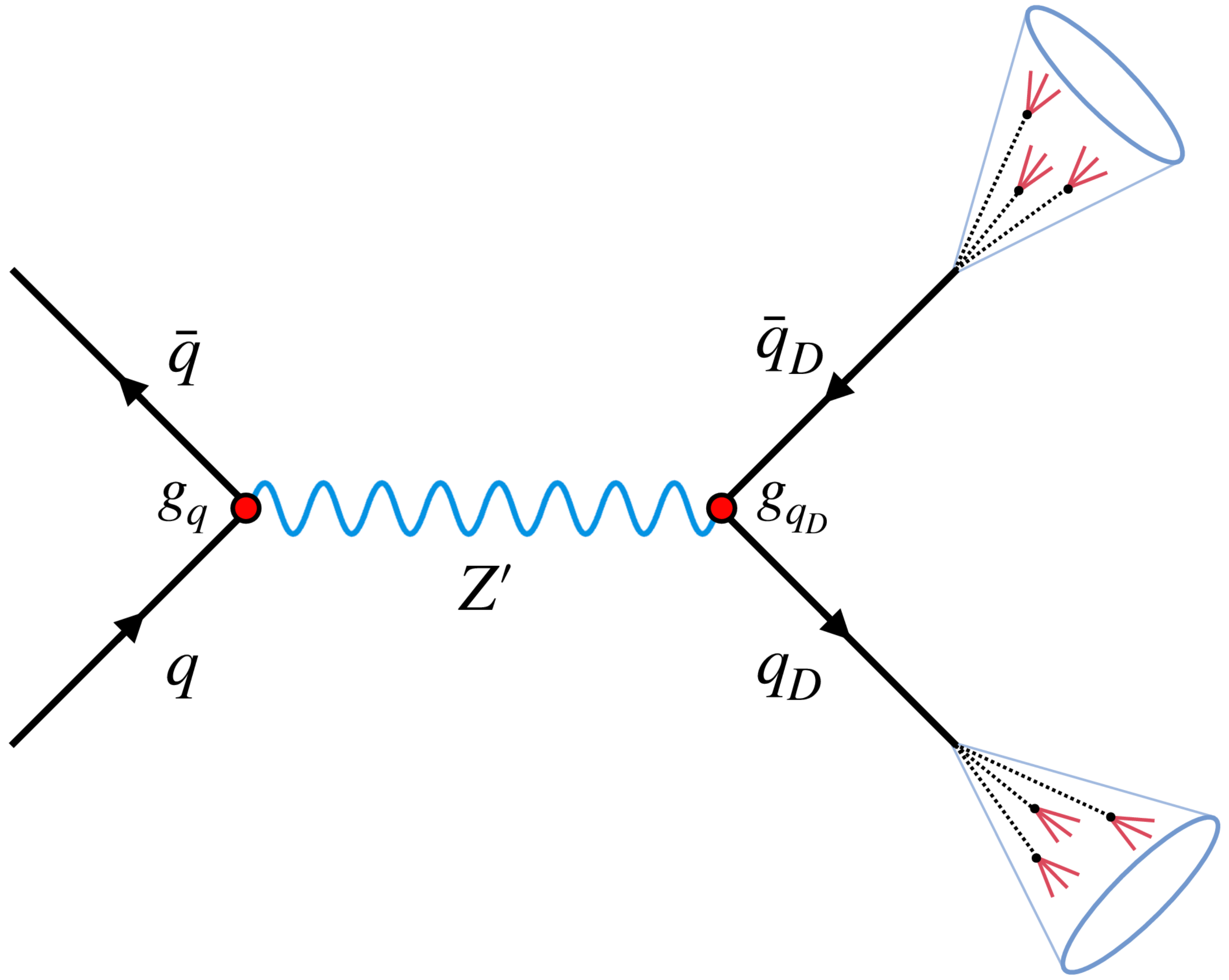
[arXiv:1502.05409]



- **Emerging jets** :
  - **life-time** of unstable dark hadrons **non negligible regarding the detector size**, low fraction of invisibility
  - jets containing multiple **displaced vertices**
  - **double hadronization** : jet sub-structure (i.e internal energy repartition) highly different from QCD jets

- In this search :
  - q and q_D coupled to a Z' **mediator**, pair production of q_D via s-channel process : unexplored mechanism
  - production of unstable dark mesons ρ_D and π_D :

  $\rho_D \rightarrow \pi_D \pi_D$ (prompt) and $\pi_D \rightarrow \bar{q}q$ with $c\tau_{\pi_D} \sim O(1-100)$ mm



$(m_{\pi_D}, \Lambda_D, m_{\rho_D})$ [GeV]	(5, 10, 20), (10, 20, 40), (20, 40, 80)
$m_{Z'}$ [TeV]	0.6, 0.8, 1.0, 1.2, 1.5, 1.8, 2.2, 2.6, 3.0, 3.5
m_Φ [TeV]	0.6, 1.0, 1.5, 2.0
$c\tau_{\pi_D}$ [mm]	1, **5**, 10, **50**, 100, **500**, 1000

Analysis strategy

- **Two complementary strategies** :
  - **cut-based** approach : selections applied to jet observables (track, vertex and jet-substructure based)
  - **machine-learning (ML) based** approach : Graph Neural Network (GNN) algorithm trained to differentiate emerging jets from QCD jets

- ML approach expected to be more sensitive to the benchmark models, but cut-based one possibly less model dependent and easier to reinterpret with alternative models

- To maximize the sensitivity, each strategy is divided into **two orthogonal regions**, each one using **distinct triggers** :
  - « low-m_{jj} » with $m_{jj} \leq 1$ TeV : to target specifically the lowest $m_{Z'}$
  - « high-m_{jj} » with $m_{jj} > 1$ TeV

Triggers

- « **Emerging jet trigger** », used in low-m_{jj} :

  at least one jet $(R = 1.0)$ with $p_T > 200$ GeV, $|\eta| < 1.8$ and PTF < 0.08

  where $\text{PTF} = \dfrac{\sum p_T^{\text{trk}}}{p_T^{\text{jet}}}$ (**Prompt-Track Fraction** of a jet)

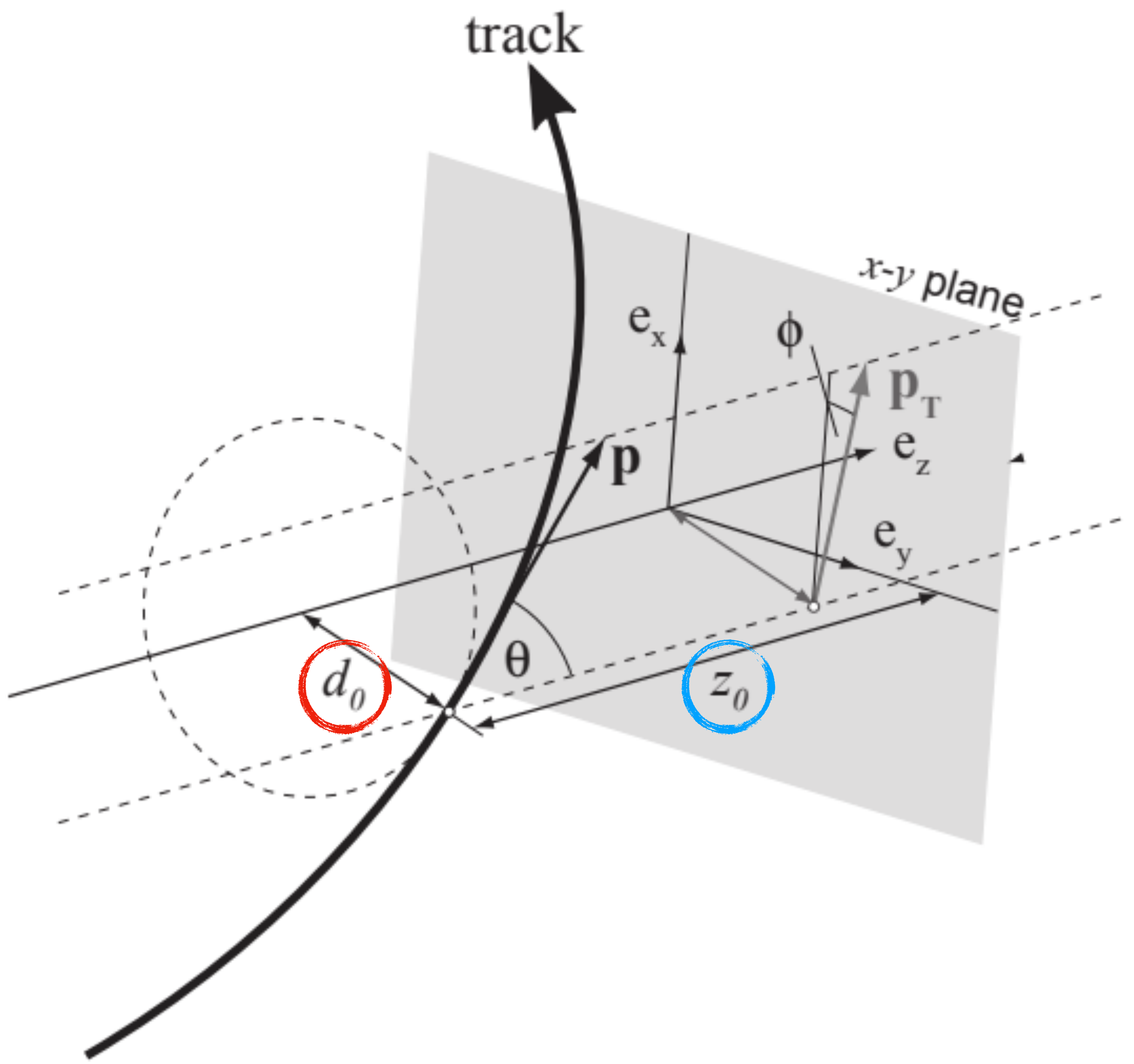
  with the sum on the tracks that are within $\Delta R < 1.2$ of the jet such as :

  $p_T^{\text{trk}} > 1$ GeV, $|d_0|/\sigma(d_0) < 2.5$ and $\Delta z = |z_{\text{PV}} - z_0| < 10$ mm

  (z_{PV} : coordinate of the hard-scattering primary vertex along the beam axis)

- « **High-p_T jet trigger** », used in high-m_{jj} :

  at least one jet $(R = 1.0)$ with $p_T > 460$ GeV



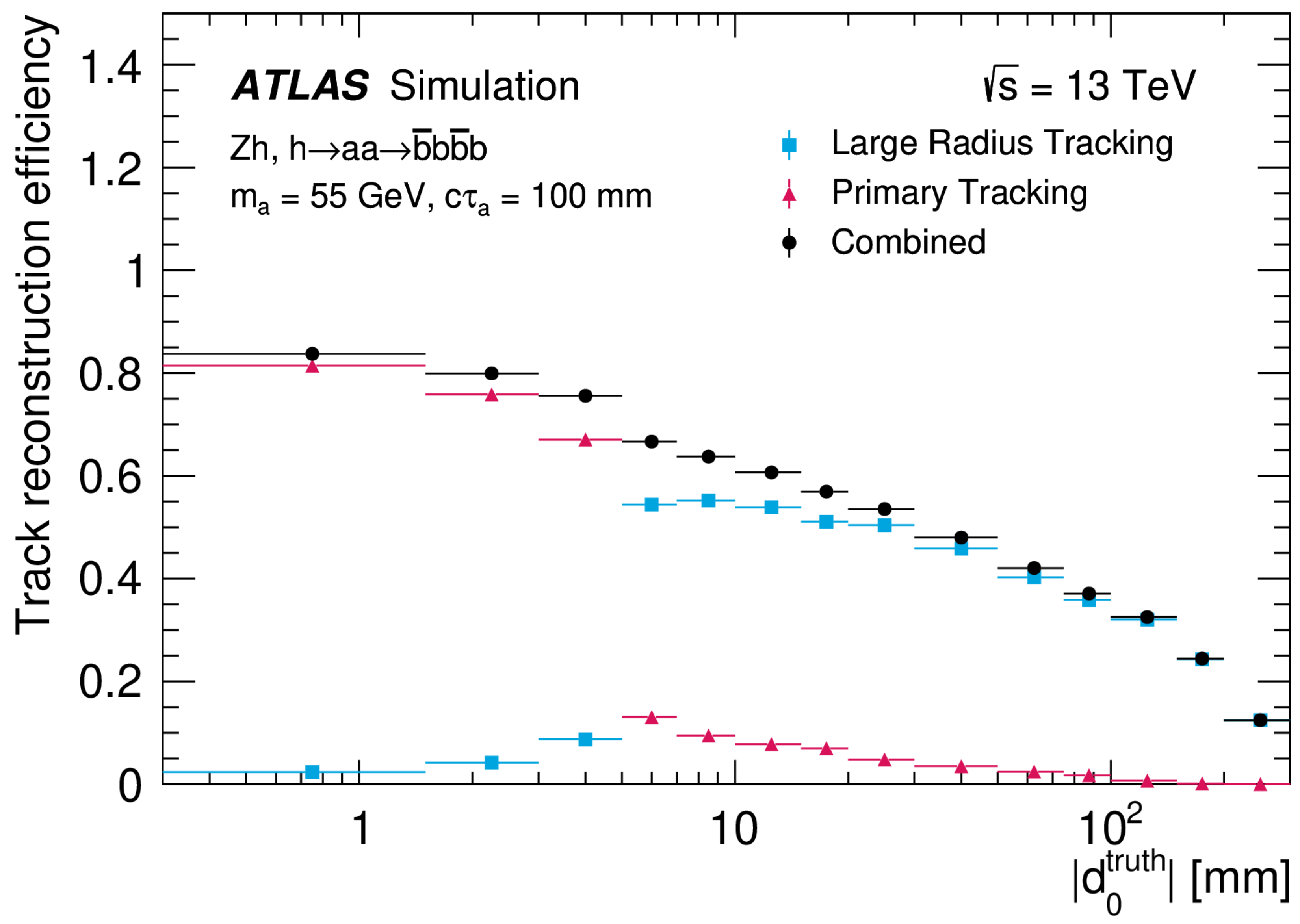
Global track parameters e.g. wrt. perigee

$$\left(d_0, z_0, \phi, \theta, \frac{q}{p}\right)$$

Event reconstruction

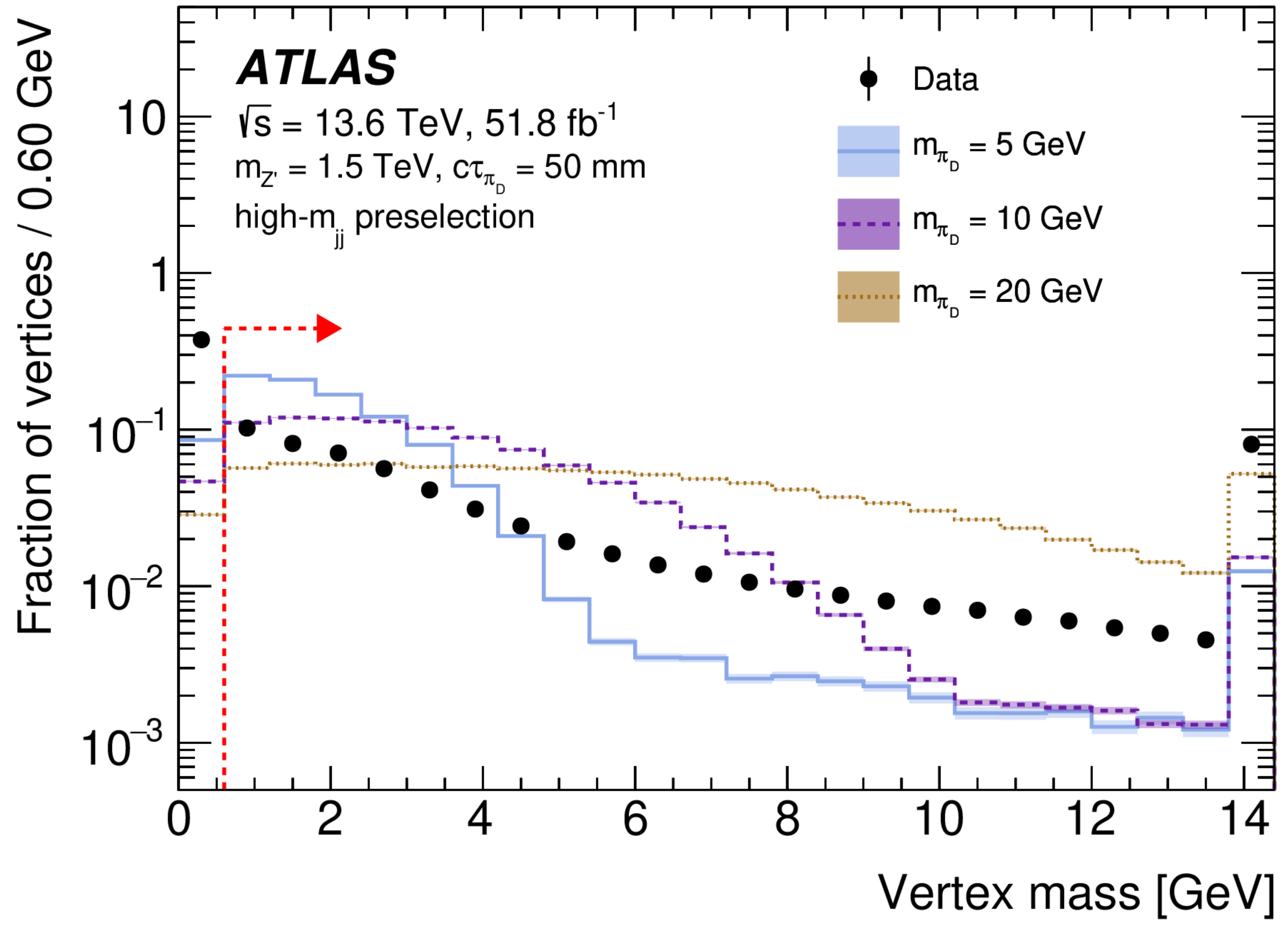
- **Tracks** :
  - « prompt tracks » : low impact parameters d_0 and z_0
  - « large radius tracks » : higher impact parameters i.e displaced from the interaction point : essential in searches for long-lived particles

- **Jets** : energy deposits in the calorimeter as constituents
  - first, clustering into **small-R sub-jets** ($R = 0.4$)
  - secondly, **sub-jets clustered to form large-R jets** ($R = 1.0$) : both with the anti-k_t algorithm
  - tracks-to-jets correspondance to compute track-based jet observables

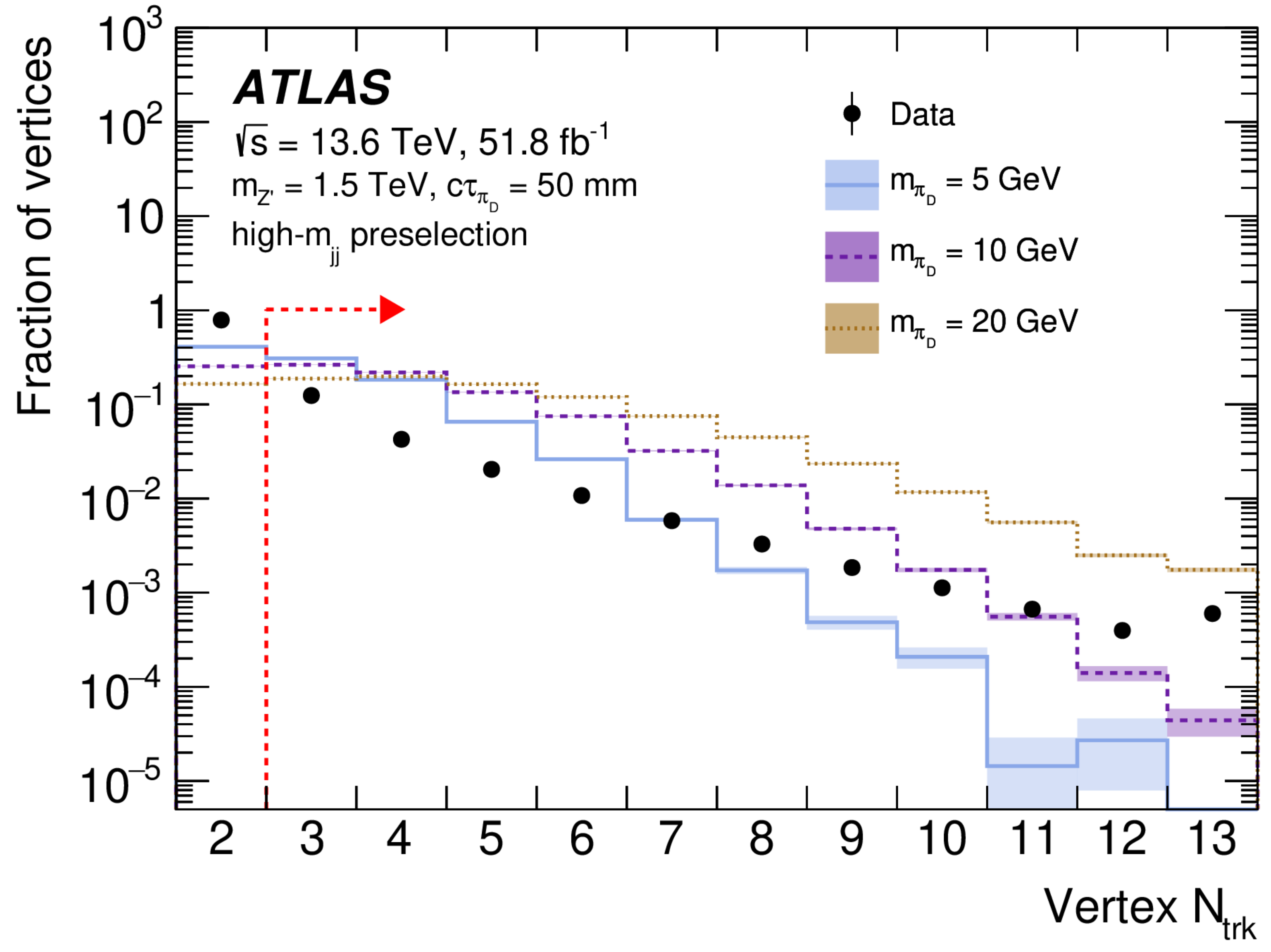


[arXiv: 2304.12867]

Event reconstruction



- **Displaced vertices :** reconstructed using all the reconstructed tracks
  - vertex 4-vector computed as the sum of the associated track 4-vectors



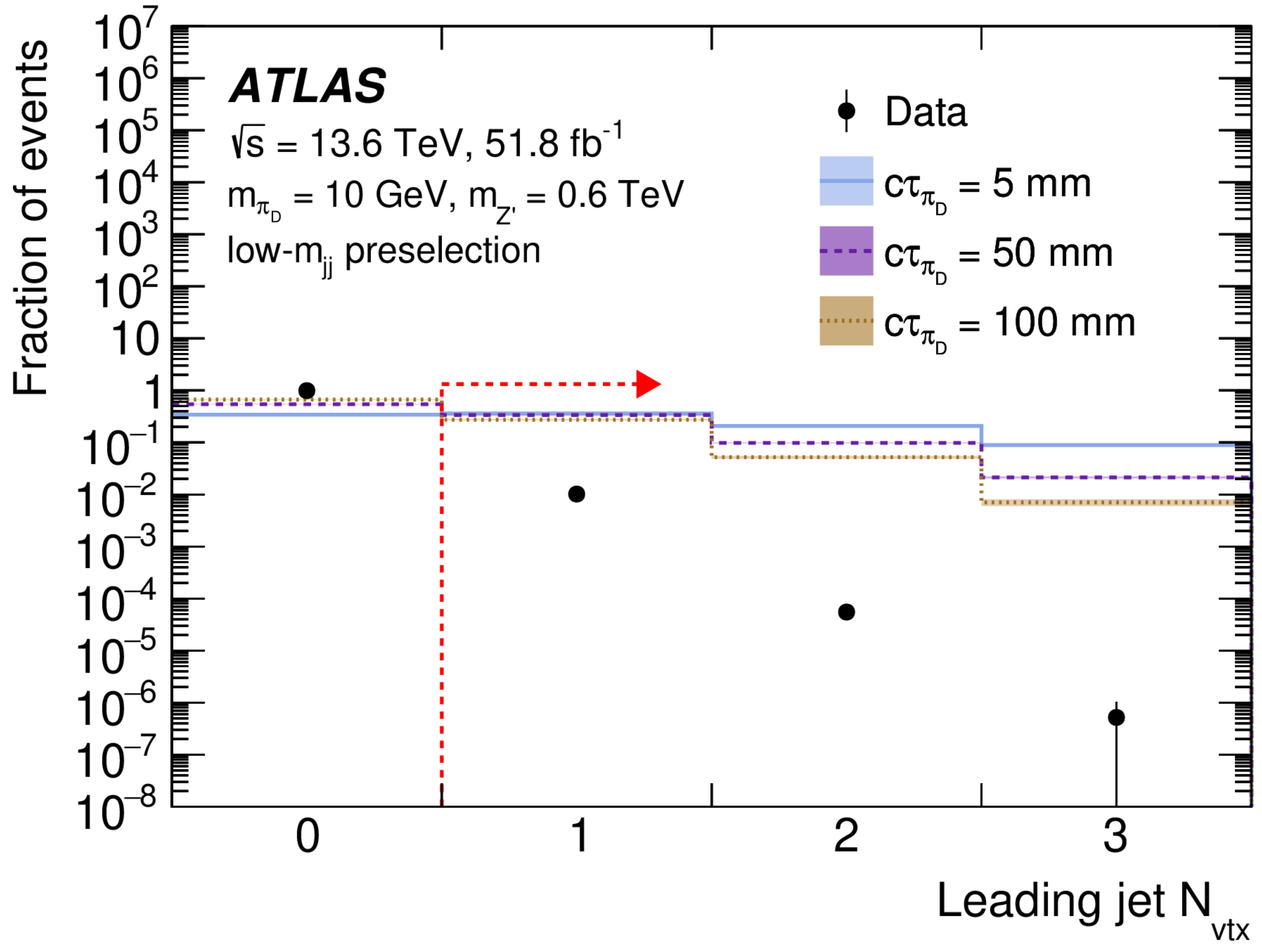
- **Additional requirements** to reduce background contributions :
  - $m_{vtx} > 0.6$ GeV to remove vertices from Kaons ($m_K \sim 0.5$ GeV)
  - $N_{trk} \geq 3$ to remove vertices from γ producing e^+e^-
  - veto to remove vertices located in detector material
  - $\Delta R(\text{jet, vertex}) < 1.0$ to be associated to a jet

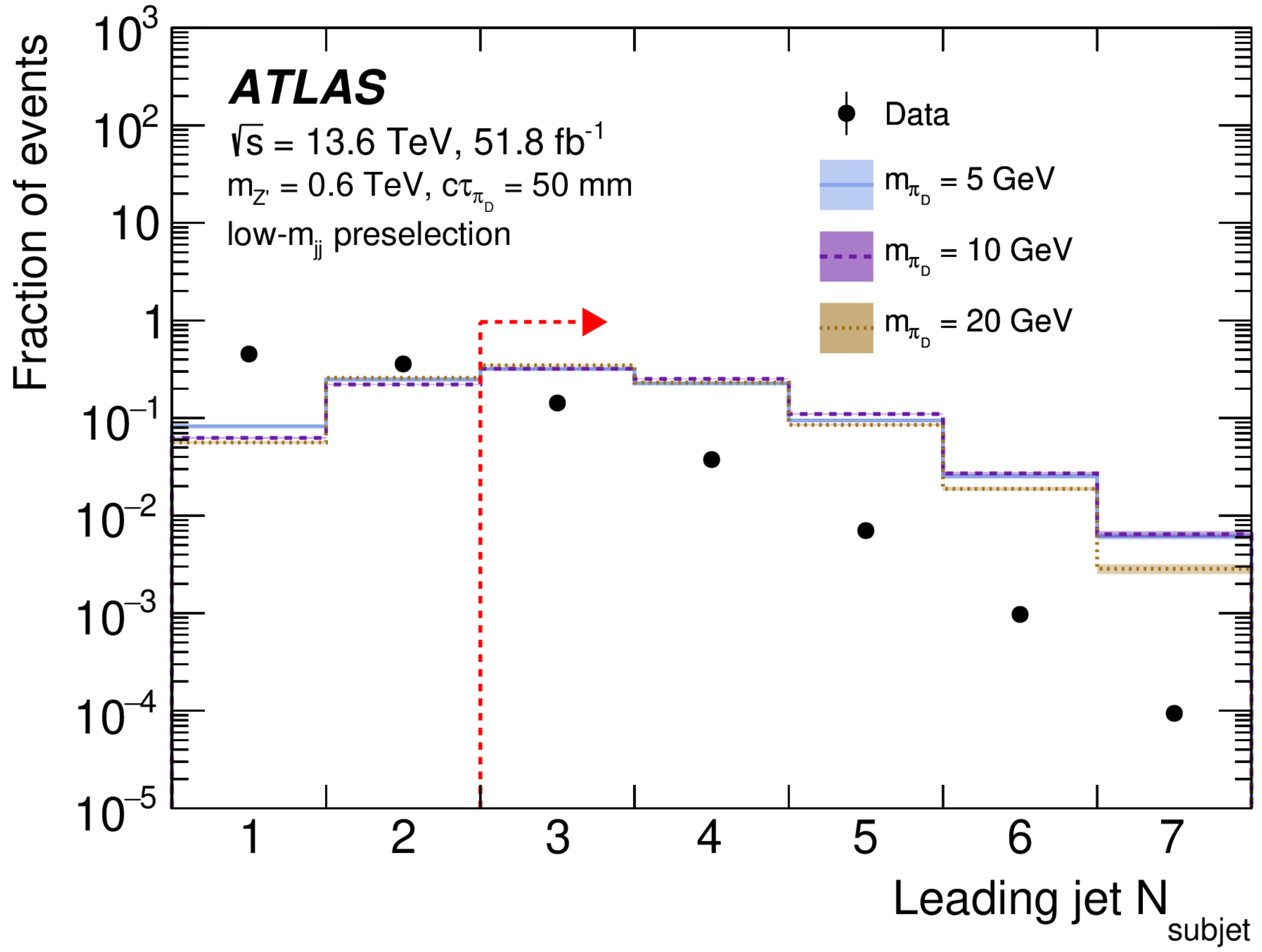
Cut-based approach
(ML-based in back-up)

Selections (low-m_{jj})

- **Pre-selections** :
  - Emerging jet trigger
  - At least two reconstructed large-R jets satisfying :
  $p_T > 200$ GeV, $|\eta| < 1.5$ and passing overlap removal with γ (i.e no reconstructed photons within $\Delta R < 1.0$ of the jet) :
  γ can produce jets with PTF ~ 0
  - $p_T > 250$ GeV & PTF < 0.04 for the jet that activate the trigger
  - $m_{jj} \leq 1$ TeV

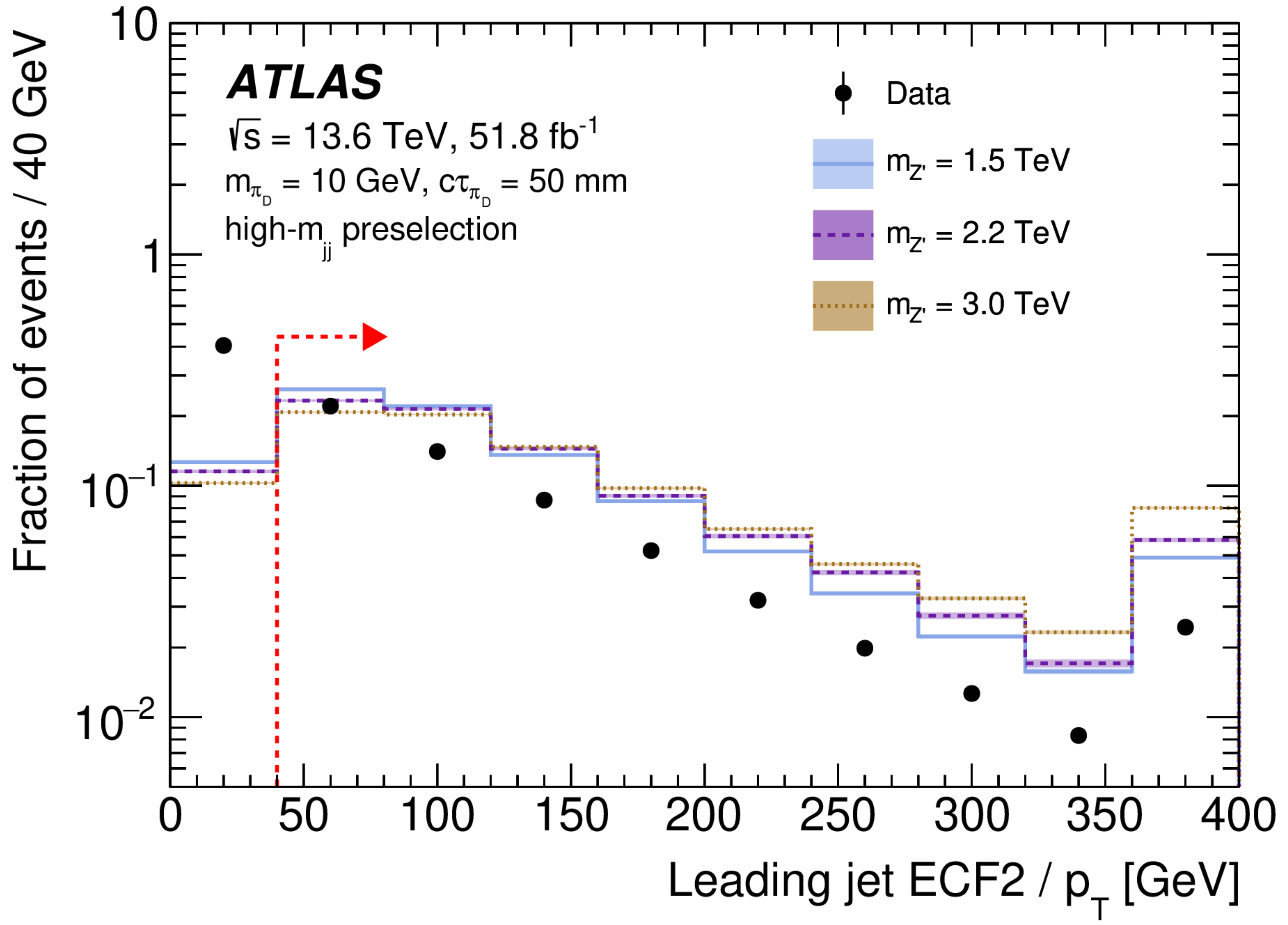
- **Selections** :
  - On the number of displaced vertices :
  $N_{\text{vtx}} \geq 1$ for the leading and sub-leading jets
  - Signal Region (SR) defined with $N_{\text{subjet}} \geq 3$ for leading and sub-leading jets, with N_{subjet} the number of small-R jets that serve as constituents to the large-R jets



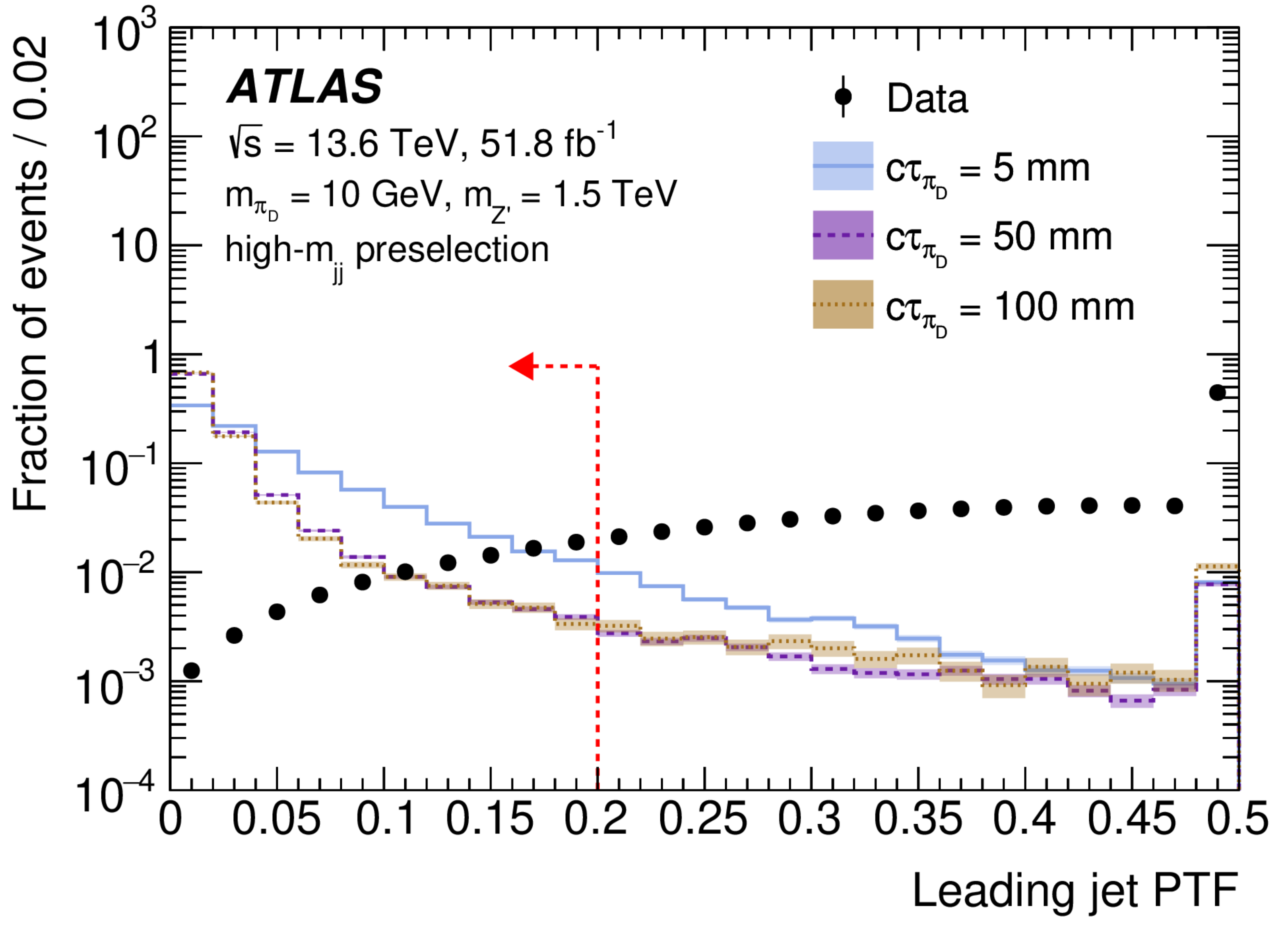


Selections (high-m_{jj})

- **Pre-selections** :
  - High-p_T jet trigger
  - At least two reconstructed large-R jets satisfying :
  $p_T > 200$ GeV, $|\eta| < 1.5$ and passing overlap removal with γ
  - $p_T > 520\ (300)$ GeV for the (sub-)leading jet
  - $m_{jj} > 1$ TeV

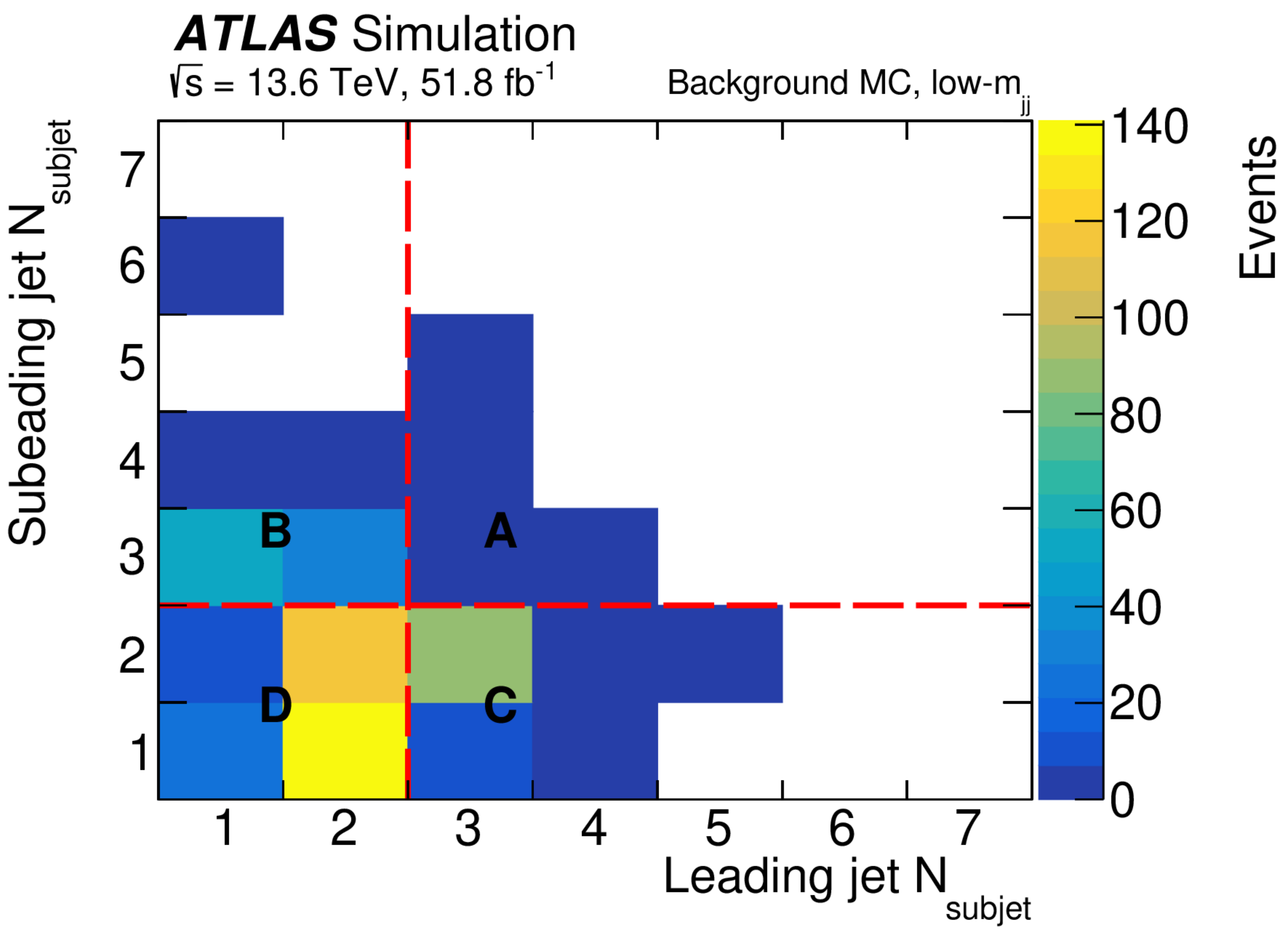


- **Selections** :
  - On the number of displaced vertices :
  $N_{\text{vtx}} \geq 1$ for the leading and sub-leading jets
  - On a jet sub-structure variable ECF2 $= \sum_{i<j\in\text{trk}} p_T^i \times p_T^j \times \Delta R_{ij}$:
  $\text{ECF2}/p_T > 40$ GeV for the leading and sub-leading jets
  - SR defined with PTF < 0.2 for leading and sub-leading jets



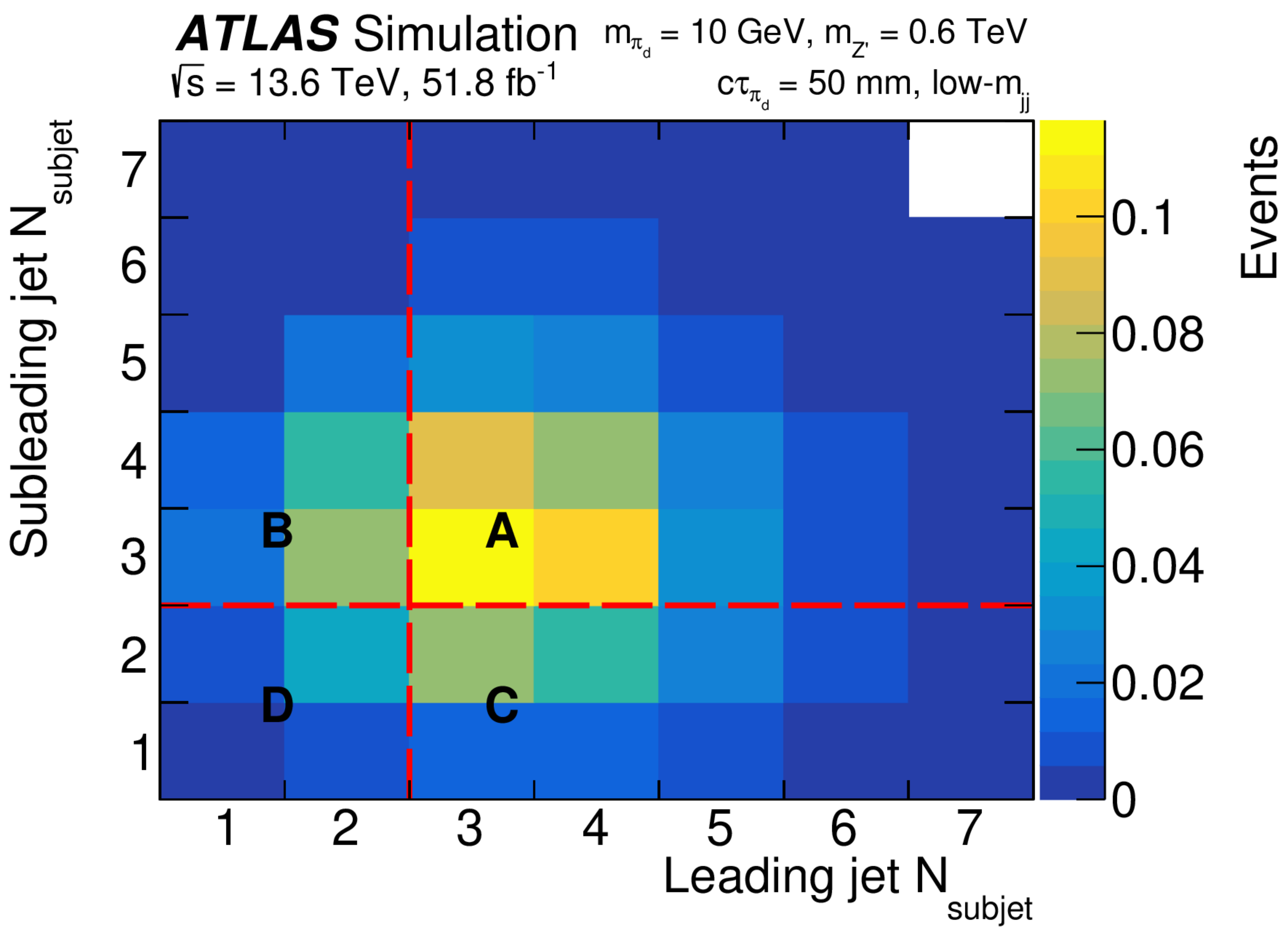
ABCD planes (low-m_{jj})

A : Signal Region
B, C & D : Control Regions



Background MC

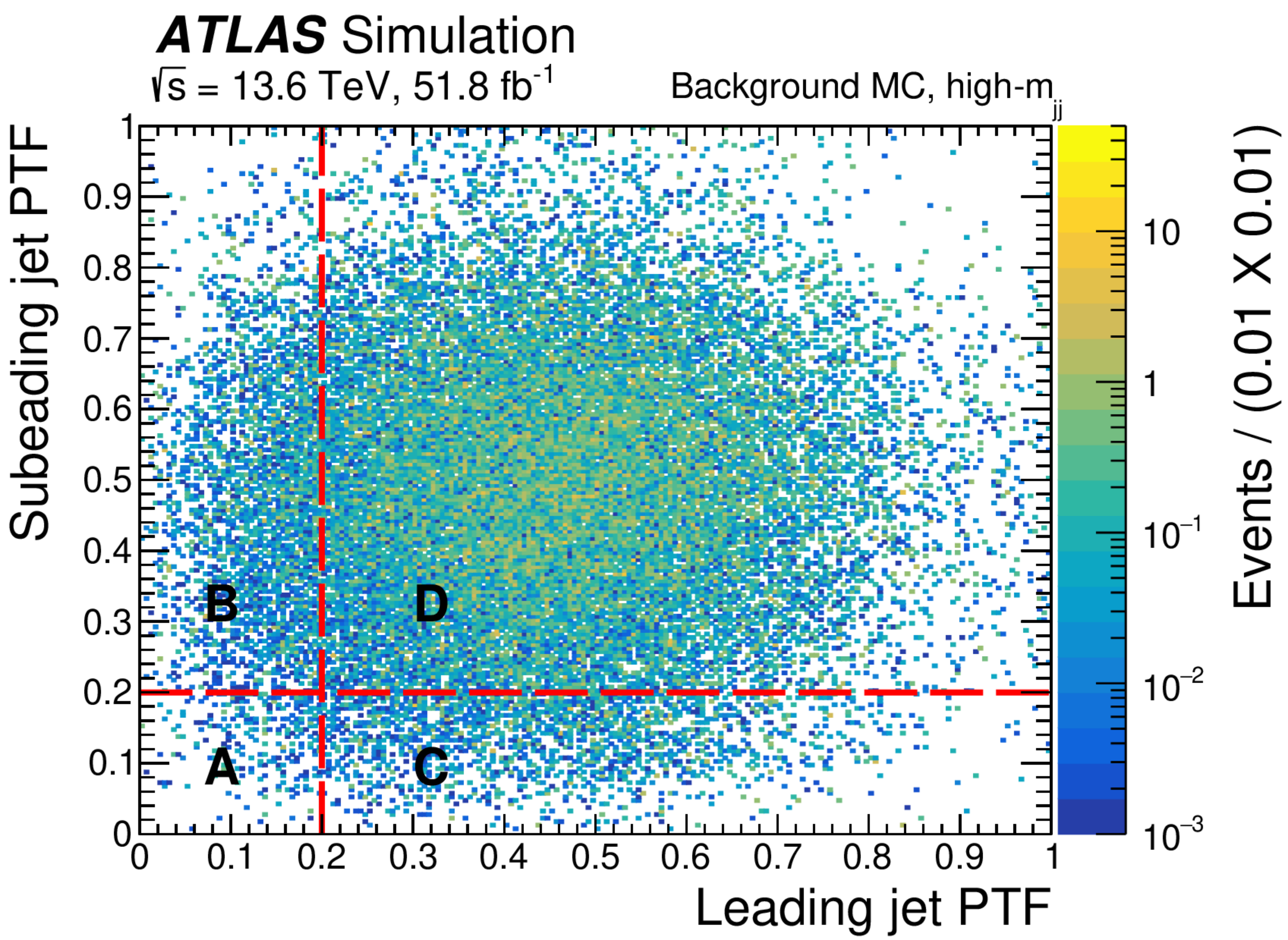
(QCD di-jet & $t\bar{t}$ events)



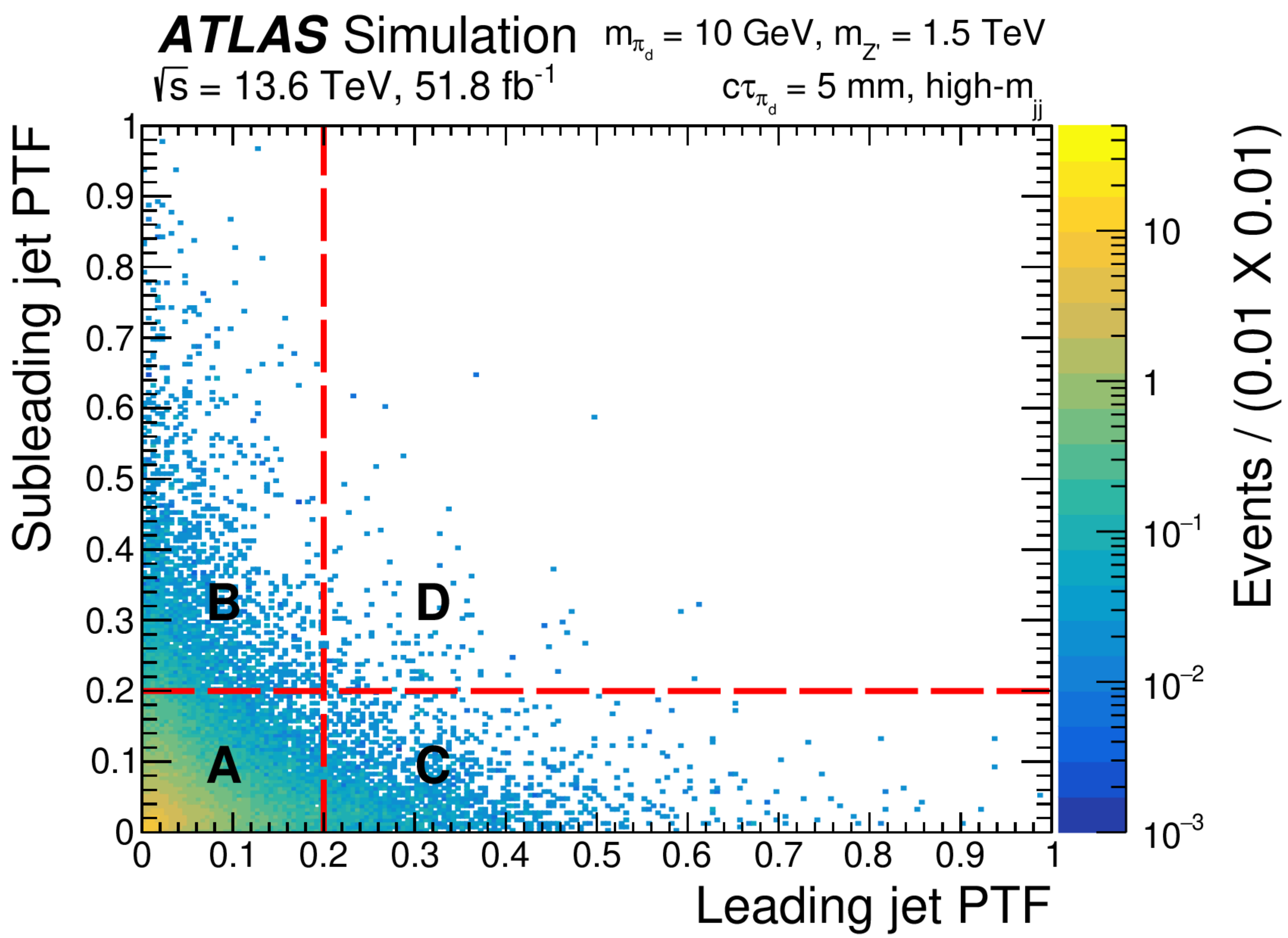
Signal

ABCD planes (high-m_{jj})

A : Signal Region
B, C & D : Control Regions



Background MC
(QCD di-jet & $t\bar{t}$ events)



Signal

Background estimation

- No search for resonance :
  - m_{jj} signal distributions too large and not enough statistics in the SRs (especially in low-m_{jj})
  - « cut-and-count » strategy instead

- **Data driven ABCD method** :
  - expected background in A : $N_A^{exp} = N_B^{bkg} \times N_C^{bkg} / N_D^{bkg}$, assuming X and Y axis variables independent for background events and A containing most of the signal
  - $N_{B,\ C,\ D}^{bkg}$ obtained from a simultaneous fit in A, B, C and D, taking into account signal presence in B, C and D

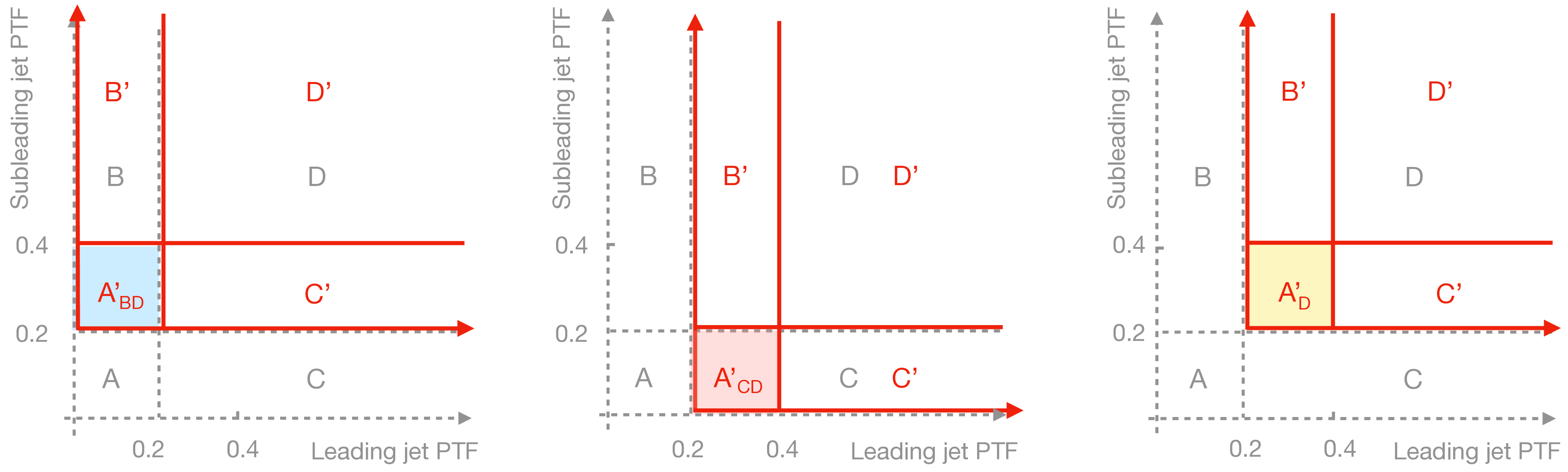
- **Validation of the ABCD method** (i.e verification of the validity of the formula) done in data in signal free ABCD plane :
  - same selections as the nominal ABCD plane plus requiring a jet classification score (from ML-based) less than 0.95 for both leading and sub-leading jets
  - signal events removed, MC background nearly unchanged

ABCD method uncertainty

- In both low and high-m_{jj} channel :
  derivation of a **non-closure systematic uncertainty on** N_A^{exp} as $f \times \sigma_{\text{stat.}}(N_{exp}^A)$

- f estimated in data ; for example in the high-m_{jj} channel :
  - in sub-regions of B, C and D named A'_{BD}, A'_{CD} and A'_D : N_{obs} and N_{exp} computed
  - $f = \text{average}[(N_{obs} - N_{exp})/\sigma_{\text{stat.}}(N_{exp})]$

- $f = 1.0$ obtained for both low and high-m_{jj}



Results

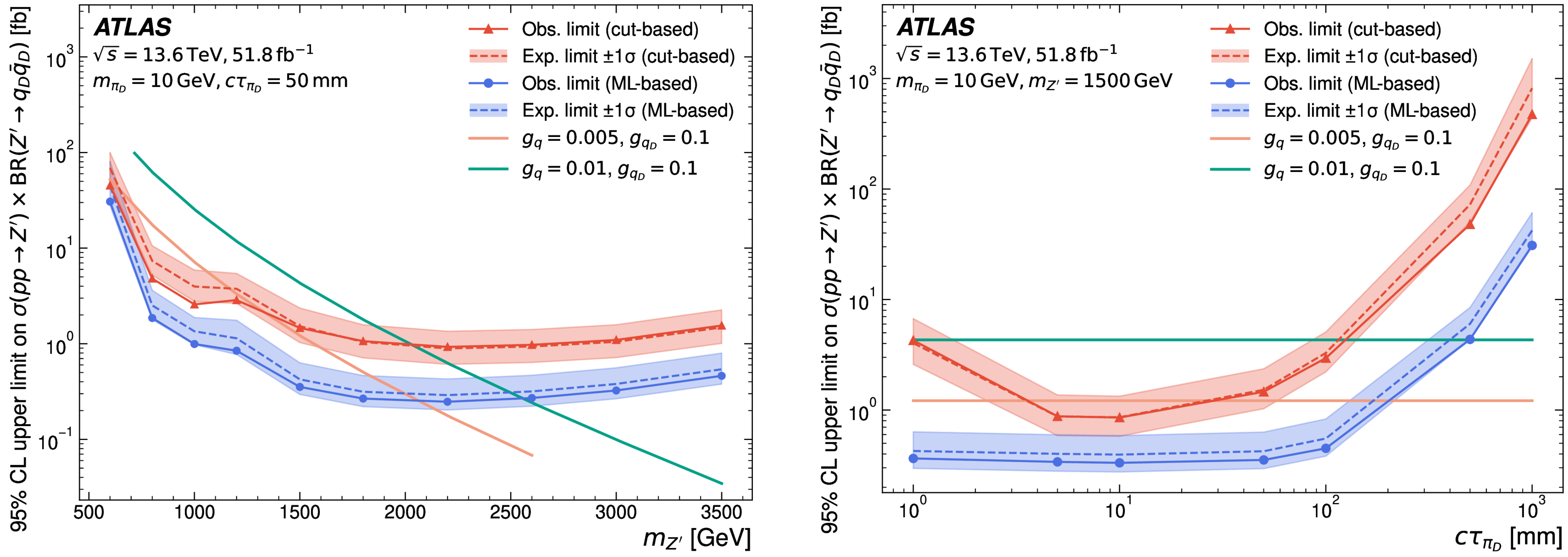
- **Observed yields in agreement with background expectations in all the SRs**

Strategy	Region	Prediction (± stat ± syst)			Observed yield
Cut-based	High-m_{jj}	7.5	±1.1	±1.1	8
	Low-m_{jj}	17.4	±5.1	±5.1	10
ML-based	High-m_{jj}	4.5	±0.3	±2.8	3
	Low-m_{jj}	31.8	±0.8	±7.5	24

- Separate statistical interpretation for the two approaches :
  - for each approach, **simultaneous likelihood-fit combining low and high-m_{jj} regions**
  - systematic uncertainties on the signal and background predictions as nuisance parameters

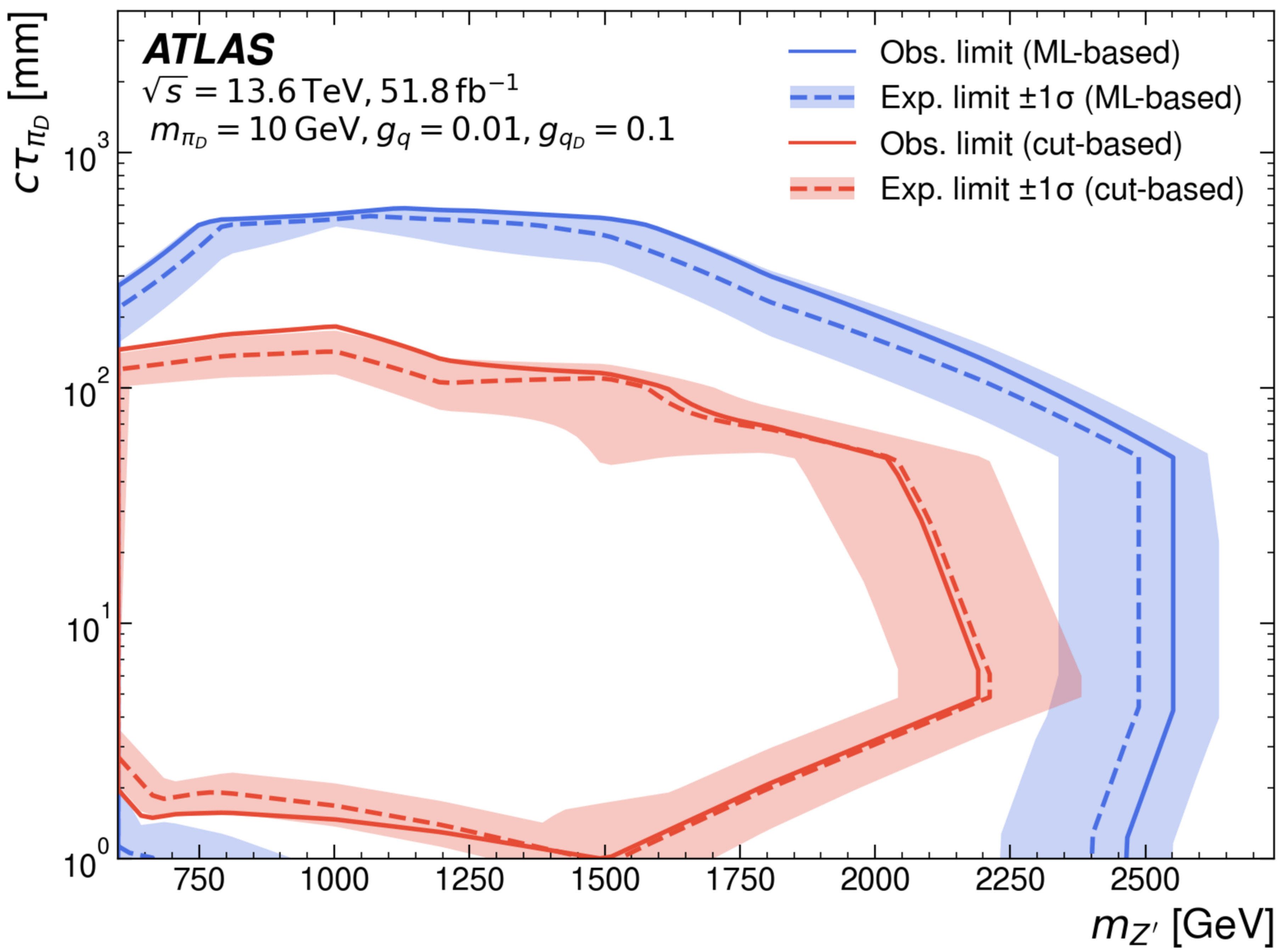
- **Upper limits at 95% CL** set on $\sigma(pp \to Z') \times \mathrm{BR}(Z' \to \bar{q}_D q_D)$ using the CLs method

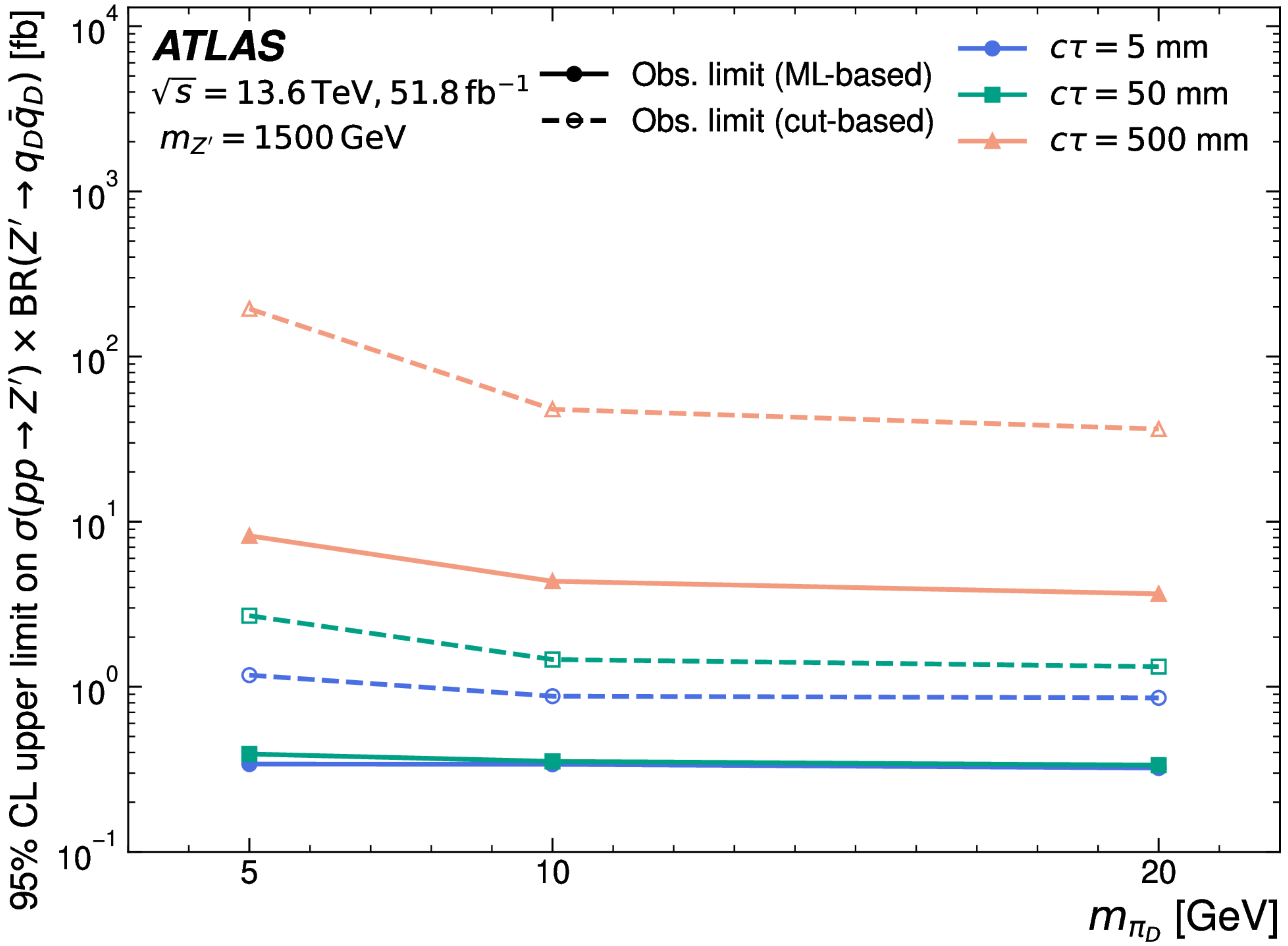
Results



- As expected, the ML-based approach sets the strongest exclusion limits
- In the cut-based approach, limits weaker at lower $c\tau_{\pi_D}$ due to requirements on PTF and N_{vtx}
- In both approaches, limits weaker at $c\tau_{\pi_D} > 100$ mm due to reduced track reconstruction efficiency

Results

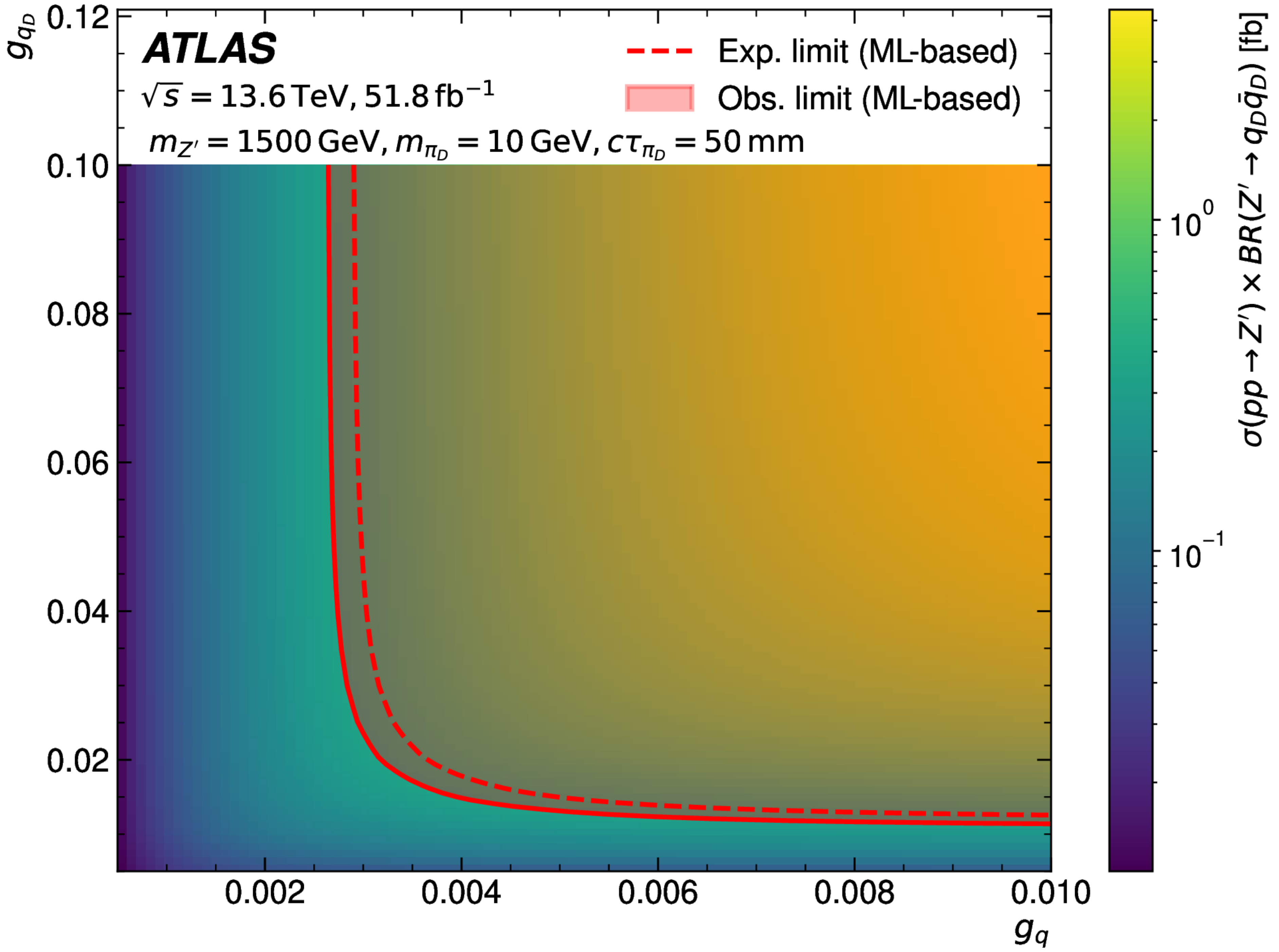




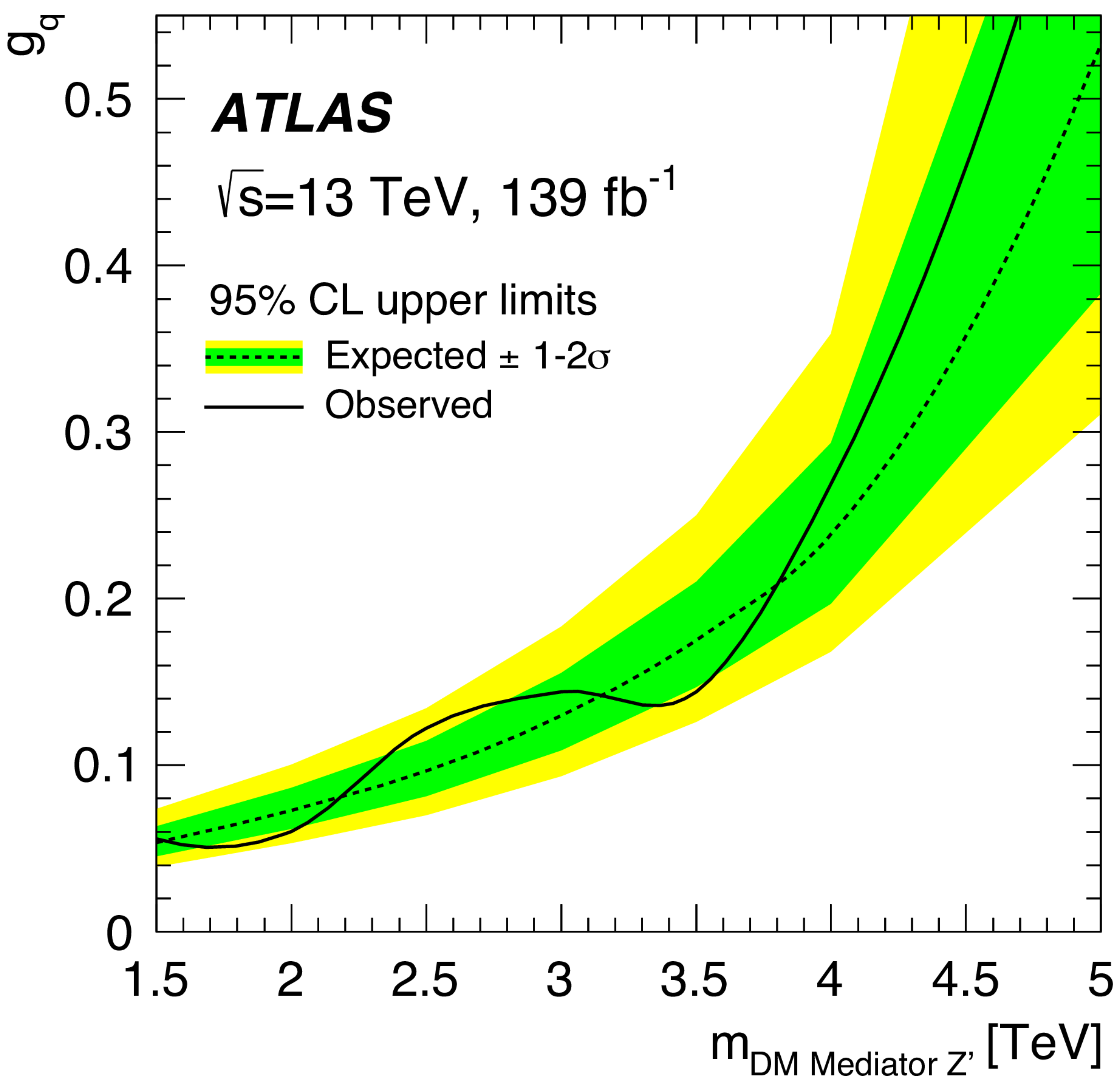
- Assuming $g_q = 0.01$ and $g_{q_D} = 0.1$:
  - ML-based (cut-based) excludes Z' masses up to 2550 (2150) GeV for $c\tau_{\pi_D} = 10$ mm, and $c\tau_{\pi_D}$ in the range 1-500 (1.5-200) mm for $m_{Z'} = 1000$ GeV

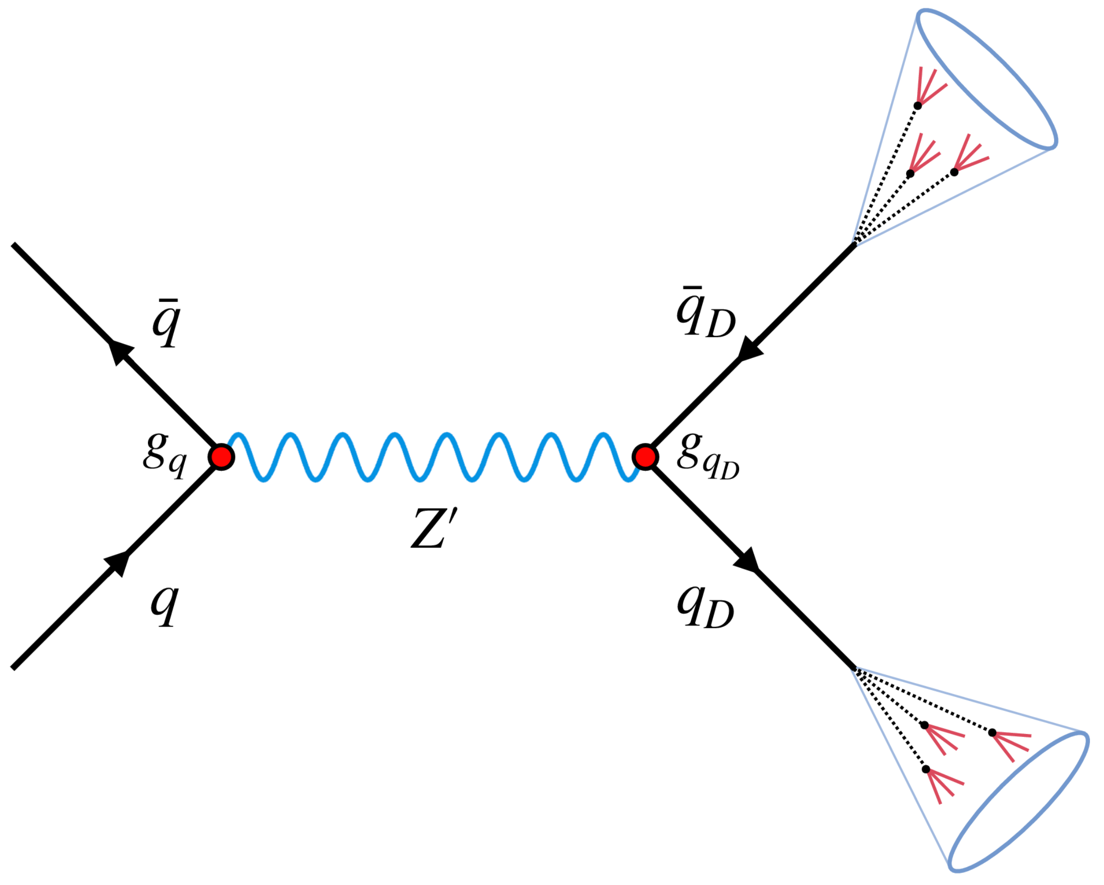
- Minimal dependence on the dark pion mass

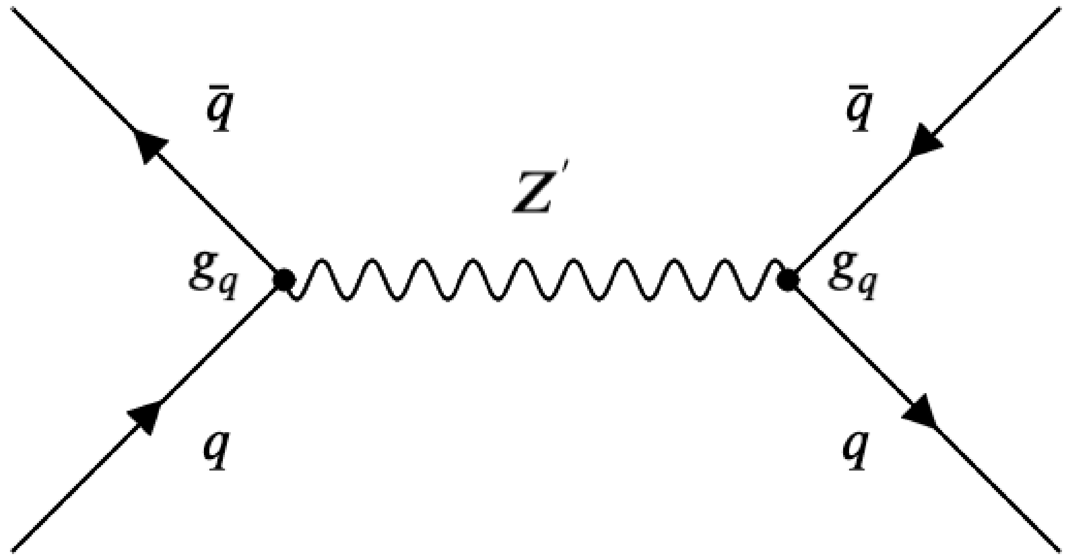
Results



[arXiv:1910.08447]







- For $m_{Z'} = 1500$ GeV, $c\tau_{\pi_D} = 50$ mm and $m_{\pi_D} = 10$ GeV :
assuming $g_{q_D} > 0.03$, ML-based excludes values of $g_q > 0.003$ (~ 20 times lower than the limit set by dijet resonance searches)

Conclusion

- **Search for a pair of emerging jets** with ATLAS using 51.8 fb^{-1} of Run-3 pp collisions data :
  - the first one considering an s-channel mediator
- **Two complementary analysis strategies** :
  - one based on event selections considering jet observables
  - one utilizing an emerging jet tagging algorithm
- Each strategy divided between low and high-m_{jj} region, each employing distinct triggers
- **No significant excess** is observed in data above a background contribution estimated with data driven techniques
- **Exclusion limits at 95% CL on** $\sigma(pp \rightarrow Z') \times \mathrm{BR}(Z' \rightarrow \bar{q}_D q_D)$

Thank you for you attention

Back-up

ML-based approach

ML-based strategy

- Use a **transformer jet tagging algorithm** based on ATLAS flavor tagging algorithm

- Input consists of jet features concatenated with feature vectors of up to 200 associated tracks

- Main task of the algorithm : **jet classification**
  - **outputs the probability that a given jet is an emerging jet** p_{EJ} (jet classification score)

- Model trained with millions of jets from MC simulations : equally from QCD di-jet and $\bar{q}q \to Z' \to \bar{q}_D q_D$ events from samples with $m_{Z'} \in \{0.6,\ 1.5,\ 3\}$ TeV and $c\tau_{\pi_d} \in \{5,\ 50\}$mm

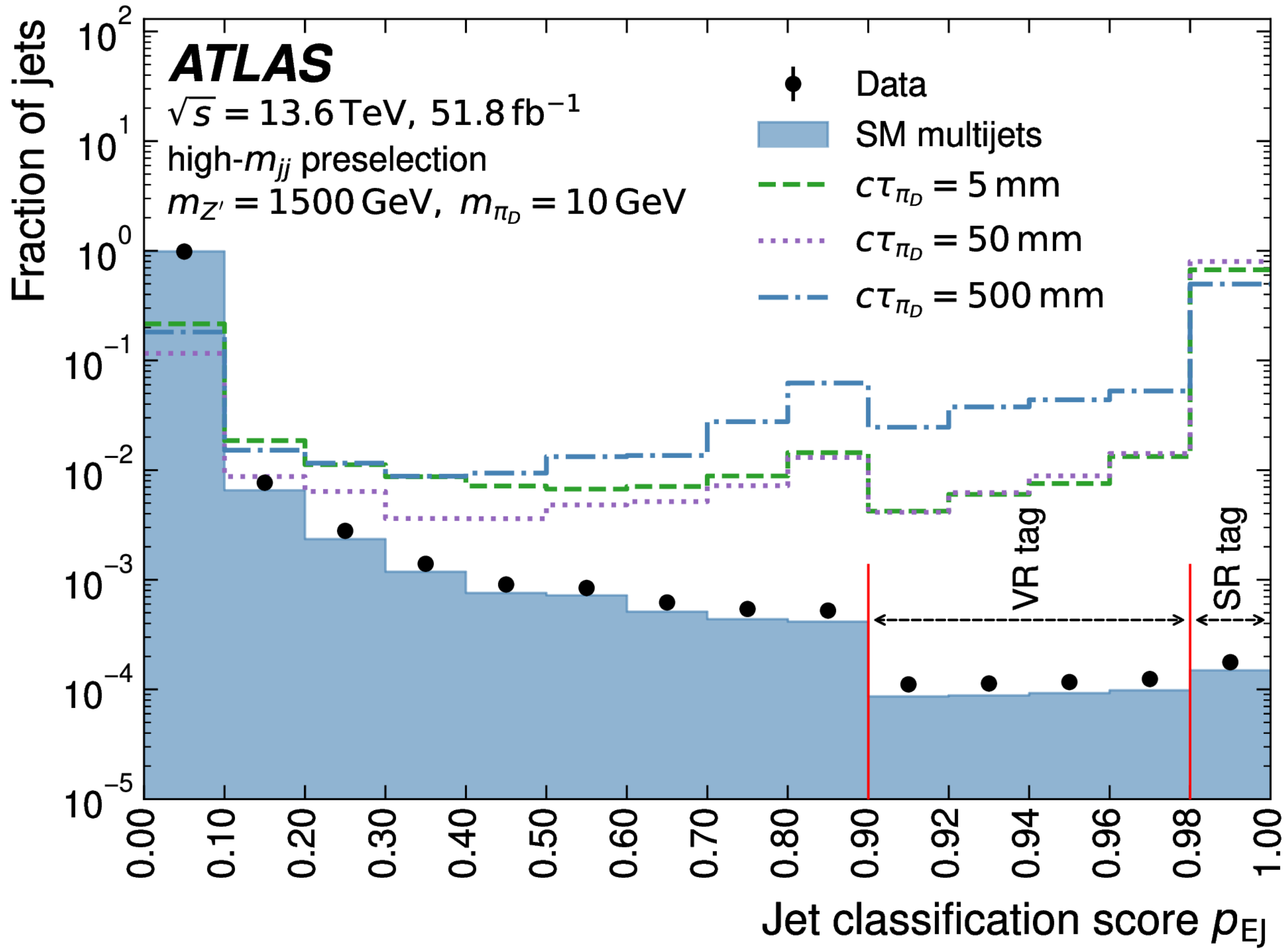
Input	Description
Jet η	Jet pseudorapidity
d_0	Track closest distance to PV in transverse plane
$z_0\sin(\theta)$	Track closest distance to PV in longitudinal plane
$\Delta\phi$	Azimuthal angle of the track, relative to the jet ϕ
$\Delta\eta$	Track pseudorapidity, relative to jet η
q/p	Track charge over momentum
$\sigma(\phi)$	Uncertainty in track ϕ
$\sigma(\theta)$	Uncertainty in track θ
$\sigma(q/p)$	Uncertainty in track q/p
$d_0/\sigma(d_0)$	signed d_0 significance
$z_0/\sigma(z_0)$	signed z_0 significance
$N_{\text{PIX hits}}$	Number of Pixel hits per track
$N_{\text{SCT hits}}$	Number of SCT hits per track
$N_{\text{IBL hits}}$	Number of innermost pixel layer hits
$N_{\text{PIX shared}}$	Number of Pixel shared hits
$N_{\text{SCT shared}}$	Number of SCT shared hits

List of track and jet features used in the tagging algorithm

Selections

- **Pre-selections** :
  - Similar as cut-based except :
  $p_T > 300$ GeV instead of 250 in the low-m_{jj} channel
  (classification task degraded at low p_T)

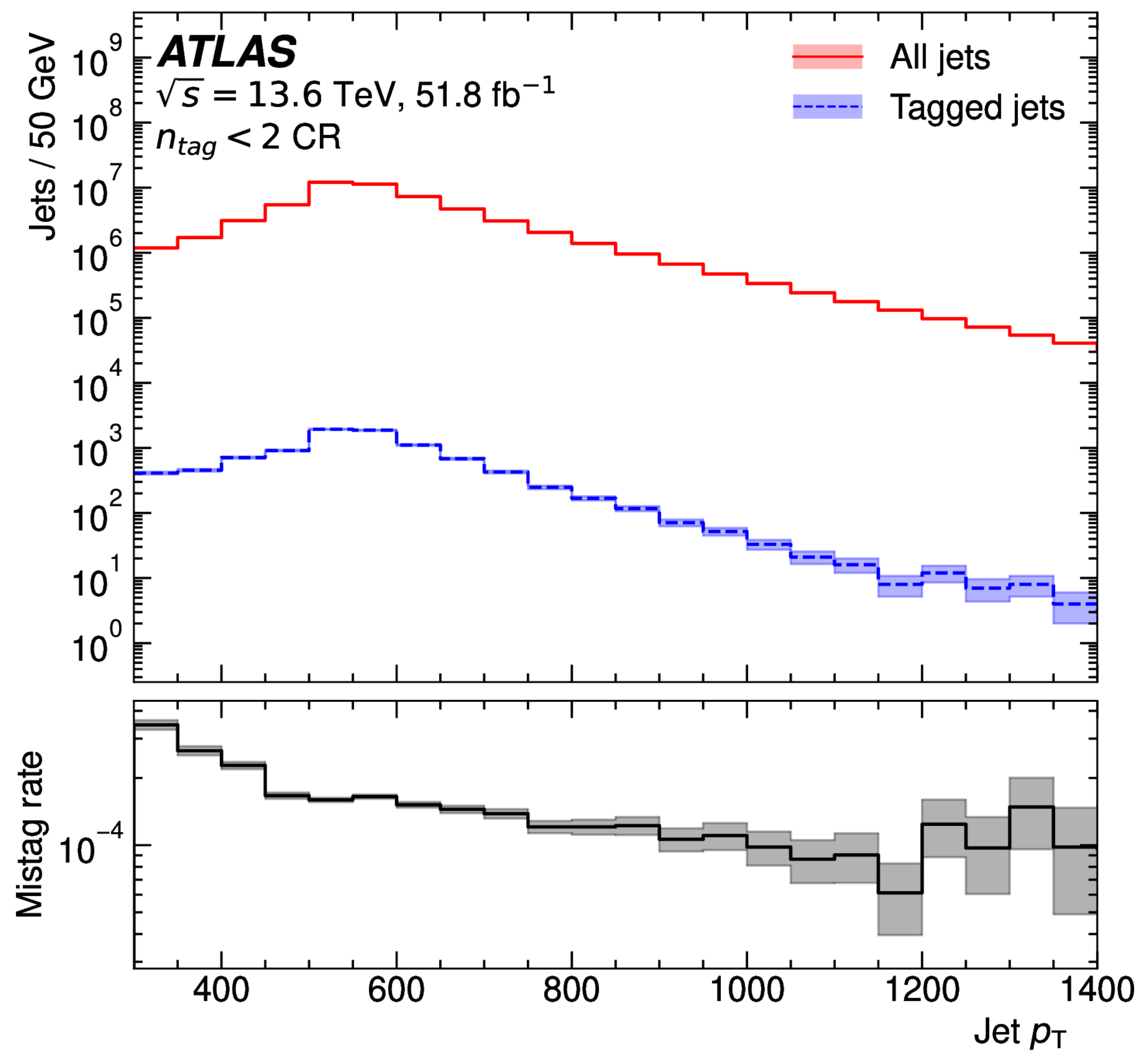
- **Selections** :
  - SR defined with $n_{\text{tags}} \geq 2$: at least two jets tagged as emerging jet i.e passing $p_{\text{EJ}} > 0.98$
  - CRs defined with $n_{\text{tags}} = 0$ or $n_{\text{tags}} = 1$
  - Threshold chosen at 0.98 to optimize both background rejection and signal acceptance

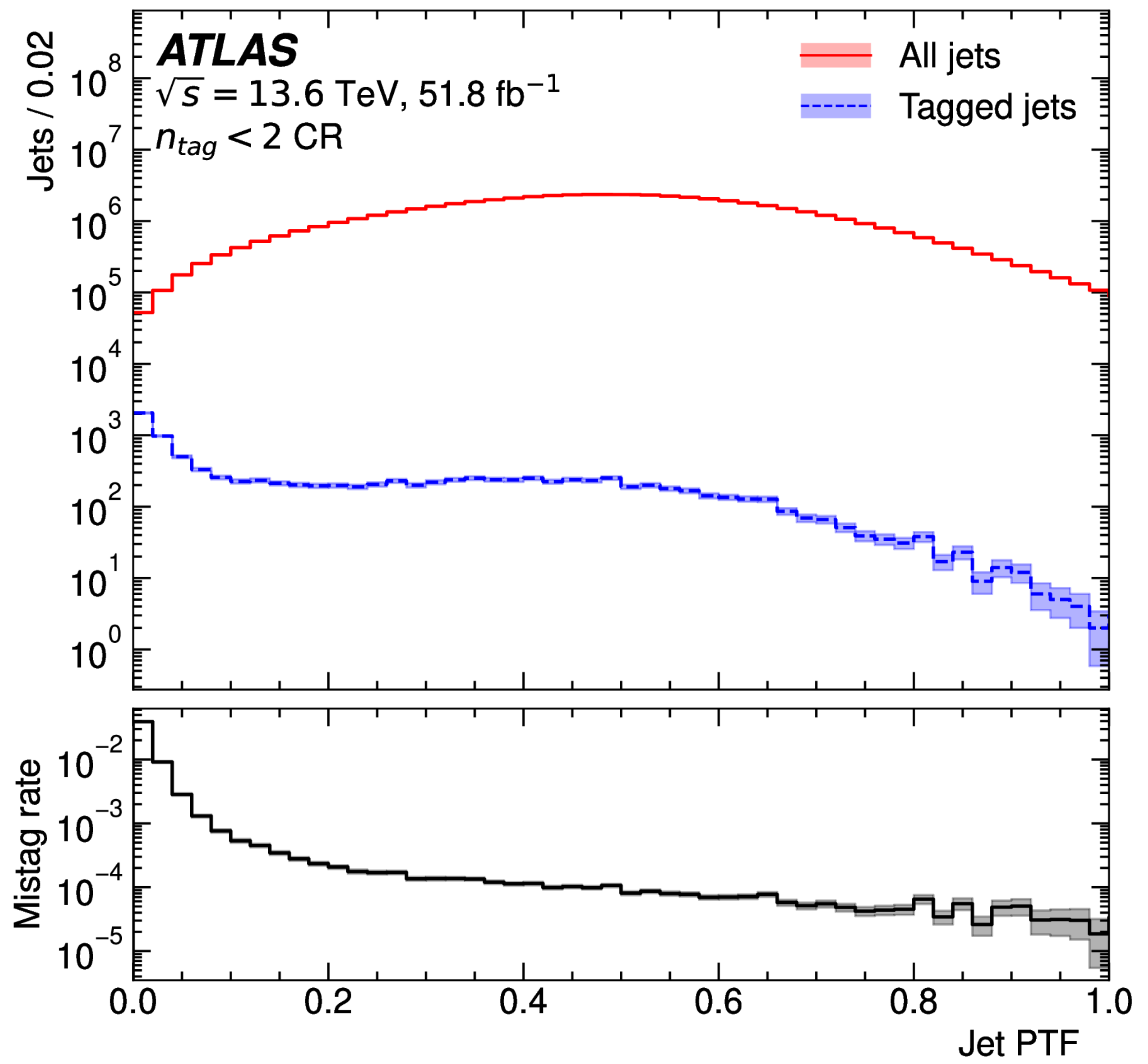


Background estimation

- **Data driven method** based on **mistag rate** : determination of the probability that a given background jet will be mistagged as an emerging jet

- Mistag rates determined directly in data in $n_{\text{tags}} < 2$ CRs :
  - correspond to the ratio of tagged jets to total number of jets
  - calculated in bins of jet p_T and PTF (highly correlated with mistag rate)





Background estimation

- Once mistag rates evaluated, computation of the probabilities to tag exactly zero, one or at least two background jets for a given event :

  - $P(0 \text{ tag}|\text{event}) = \prod_{i=1}^{n_{jet}} (1 - P(\text{tag}|j_i))$,

  with $P(\text{tag}|j_i)$ the mistag rate for the jet i

  - $P(1 \text{ tag}|\text{event}) = \sum_{i=1}^{n_{jet}} P(\text{tag}|j_i) \times \prod_{k \neq i} (1 - P(\text{tag}|j_k))$

  - $P(\geq 2 \text{ tag}|\text{event}) = 1 - P(0 \text{ tag}|\text{event}) - P(1 \text{ tag}|\text{event})$

- **Background prediction** :

  - **in the SR** : $\sum_{\text{event}} P(\geq 2 \text{ tag}|\text{event})$

  - in 1-tag region : $\sum_{\text{event}} P(1 \text{ tag}|\text{event})$

  with the sum on all the pre-selected events

Background uncertainties

- **Statistical** due to the **finite number of events in CRs** used to compute mistag rates :
  - mistagging efficiency ϵ_i in a bin i with $n_{jet,\ i}$ jets has a statistical uncertainty given by :
  $\sigma(\epsilon_i) = \sqrt{\epsilon_i(1+\epsilon_i)}/\sqrt{n_{jet,\ i}}$
  - nominal efficiencies varied with Gaussian PDF with $\sigma(\epsilon_i)$ as width
  - based on these variations, 100 alternative predictions for the number of events in the SR are computed : standard deviation of the distribution as a statistical uncertainty

- **Systematic** related to the **choice of the mistag rate parametrization** using p_T and PTF :
  - other jet observables could have been considered :
  number of b-tagged sub-jets, number of tracks and secondary vertex associated to a jet
  - alternative mistag rate parametrization considered, and background estimation computed for each parametrization : largest variation as a systematic uncertainty

Mistag rate method validation

- Validation region defined as an alternative 2-tag region :
  - to be tagged : classification score between 0.9 and 0.98
- Mistag rates computed according to this tag definition and systematic uncertainty evaluated as described previously (statistical uncertainty negligible)
- Observed yields in agreement with the prediction in the validation regions :
  no additional non closure systematic uncertainty required

	Low-m_{jj} VR tag	High-m_{jj} VR tag
Pred.	174 ± 42 (syst.)	29 ± 16 (syst.)
Obs	185	31