Measurement of q/g-jet fractions and related issues in the CMS experiment

S. Shulha

JINR (RU)
F. Skorina GSU (BY)

XXVth International Baldin Seminar on High Energy Physics Problems
September 18-23, 2023
JINR, Dubna

- $h_{\mathrm{MC}}(V)$ - NON-NORMALIZED jet histogram for jet macro parameter V - consists of q/g-histograms $h^{g}_{\mathrm{MC}}(V)$ and $h^{q}_{\mathrm{MC}}(V)$:

$$h_{\mathrm{MC}} = h^{g}_{\mathrm{MC}} + h^{q}_{\mathrm{MC}}$$

$$N_{entries} = N^{g}_{entries} + N^{q}_{entries}$$

- $H_{\mathrm{MC}}(V)$ - NORMALIZED histogram – linear combination of MC "q/g-templates" $H^{g}_{\mathrm{MC}}(V)$ and $H^{q}_{\mathrm{MC}}(V)$:

$$H_{\mathrm{MC}} = \alpha^{g}_{\mathrm{MC}} \cdot H^{g}_{\mathrm{MC}} + (1 - \alpha^{g}_{\mathrm{MC}}) \cdot H^{q}_{\mathrm{MC}}$$

g-fraction

$$\alpha^{g}_{\mathrm{MC}} = \frac{N^{g}_{entries}}{N_{entries}}$$

- In DATA, the same equation has the form :

$$H_{\mathrm{DAT}} = \alpha^{g}_{\mathrm{DAT}} \cdot H^{g}_{\mathrm{DAT}} + (1 - \alpha^{g}_{\mathrm{DAT}}) \cdot H^{q}_{\mathrm{DAT}}$$

Measured histogram

Three unknown "quantities"

- To find three unknown "quantities" you need to write three equations for three jet samples. However, three jet samples have five unknowns – three "g-fractions" and two "q/g-templates...

- To find out g-jet fraction in the data, we need to accurately determine the objects (q/g-jets) that we recognize in the data

- So, in DATA, equation for measuring g-fraction must be written with MC q/g-templates:

$$H_{\mathrm{DAT}} = \alpha^{g}_{\mathrm{DAT}} \cdot H^{g}_{\mathrm{MC}} + (1 - \alpha^{g}_{\mathrm{DAT}}) \cdot H^{q}_{\mathrm{MC}}$$

- $\Longrightarrow$ measured g-fraction for $\boldsymbol{V}$-bin:

$$\alpha^{g}_{\mathrm{DAT}}(V) = \frac{H_{\mathrm{DAT}}(V) - H^{q}_{\mathrm{MC}}(V)}{H^{g}_{\mathrm{MC}}(V) - H^{q}_{\mathrm{MC}}(V)}$$

- Definition: measured g-fraction in a sample = average of $\boldsymbol{V}$-bins:

$$\alpha^{g} \equiv \langle \alpha^{g}_{V} \rangle = \frac{\sum_{V=1}^{N_V} \alpha^{g}_{V}}{N_V} \quad \text{with uncertainty} \quad \Delta\alpha^{g} \equiv \frac{\sqrt{\langle {\alpha^{g}_{V}}^{2} \rangle - \langle \alpha^{g}_{V} \rangle^{2}}}{\sqrt{N_V}}$$

(June, 2023 г)

- Two sources of uncertainty $\Delta\alpha^{g}$: statistical fluctuations and systematic deviation of true q/g-templates from model ones:

$$H^{f}_{\mathrm{DAT}}(V) \neq H^{f}_{\mathrm{MC}}(V)$$

- To find true q/g-templates, H^q_{DAT}, H^g_{DAT}, we need two jet samples with the same kinematics:

$$H_{1,\mathrm{DAT}} = \alpha^g_{1,\mathrm{DAT}} \cdot H^g_{\mathrm{DAT}} + (1 - \alpha^g_{1,\mathrm{DAT}}) \cdot H^q_{\mathrm{DAT}}$$
$$H_{2,\mathrm{DAT}} = \alpha^g_{2,\mathrm{DAT}} \cdot H^g_{\mathrm{DAT}} + (1 - \alpha^g_{2,\mathrm{DAT}}) \cdot H^q_{\mathrm{DAT}}$$

$$\alpha^g_{2,\mathrm{DAT}} \neq \alpha^g_{1,\mathrm{DAT}}$$

- $\Longrightarrow$ find true q/g-templates:

$$H^q_{\mathrm{DAT}} = \frac{\alpha^g_{2,\mathrm{DAT}} H_{1,\mathrm{DAT}} - \alpha^g_{1,\mathrm{DAT}} H_{2,\mathrm{DAT}}}{\alpha^g_{2,\mathrm{DAT}} - \alpha^g_{1,\mathrm{DAT}}}, \qquad H^g_{\mathrm{DAT}} = (g \to q, 1 \leftrightarrow 2)$$

- Ratio $S^f(V) \equiv H^f_{\mathrm{DAT}} / H^f_{\mathrm{MC}}$ is called "**data-driven Scale Factor**"

- **Proposition**: Deviations of $S^f(V)$ from 1 are within uncertainties of measured $\alpha^g_{1,\mathrm{DAT}}$ и $\alpha^g_{2,\mathrm{DAT}}$

Sketch of proof

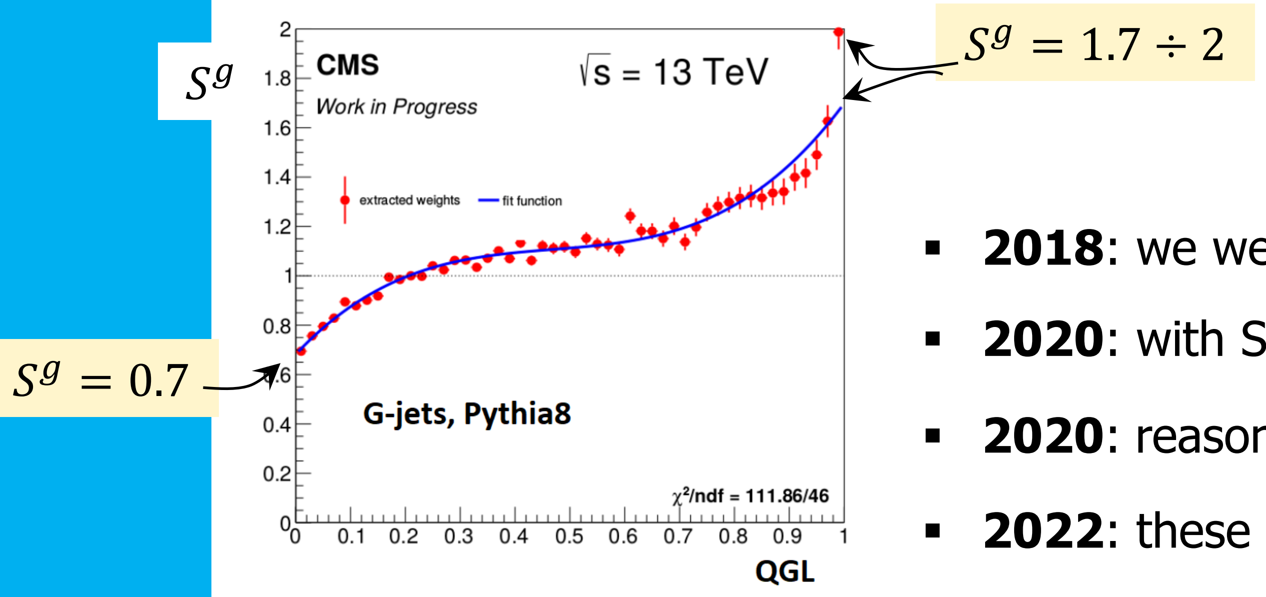
$$H_{1,\mathrm{DAT}} \approx \alpha^g_{1,\mathrm{DAT}} \cdot H^g_{\mathrm{MC}} + (1 - \alpha^g_{1,\mathrm{DAT}}) \cdot H^q_{\mathrm{MC}}$$
$$H_{1,\mathrm{DAT}} = \alpha^g_{1,\mathrm{DAT}} \cdot H^g_{\mathrm{DAT}} + (1 - \alpha^g_{1,\mathrm{DAT}}) \cdot H^q_{\mathrm{DAT}}$$
$$\Longrightarrow \alpha^g_{1,\mathrm{DAT}} \cdot (H^g_{\mathrm{MC}} - H^g_{\mathrm{DAT}}) + (1 - \alpha^g_{1,\mathrm{DAT}}) \cdot (H^q_{\mathrm{MC}} - H^q_{\mathrm{DAT}}) \approx 0$$
$$\Longrightarrow H^g_{\mathrm{MC}} - H^g_{\mathrm{DAT}} \approx 0, \quad H^q_{\mathrm{MC}} - H^q_{\mathrm{DAT}} \approx 0$$

- **Corollary:** There is no SF task

- In CMS, task and method of finding "data-driven SF's" were defined **11 years ago** and are still an official recommendation
- Why is that? →
  - 1st reason: only in 2018 we first showed possibility to measure g-fractions
  - We found a strong g-jet suppression (30-50%) in data relative to MC
  - 2nd reason: equations for SF were written in UNNORMALIZED form

$$h_{1,\mathrm{DAT}}(V) = h^{g}_{1,\mathrm{MC}}(V)\cdot S^{g}(V) + h^{q}_{1,\mathrm{MC}}(V)\cdot S^{q}(V)$$
$$h_{2,\mathrm{DAT}}(V) = h^{g}_{2,\mathrm{MC}}(V)\cdot S^{g}(V) + h^{q}_{2,\mathrm{MC}}(V)\cdot S^{q}(V)$$

$\Longrightarrow$ MC g-fractions are used but **hidden**! $\Longrightarrow$ SF is very large!



- **2018**: we were asked to apply this SF
- **2020**: with SF $\alpha^{g}_{\mathrm{DAT}} \approx \alpha^{g}_{\mathrm{MC}}$
- **2020**: reason – official SF uses α^{g}_{MC}
- **2022**: these conclusions are accepted in CMS...

- The model determines the g-fraction unambiguously and does not allow data-driven corrections
- However, the model q/g-templates and q/g-templates in DATA are very different

$$H^f_{\mathrm{DAT}} \neq H^f_{\mathrm{MC}}$$

- To verify this, you need to measure g-fractions α^g_{DAT} with different jet Macro Parameters (MP) → α^g_{DAT} will be different
- The reason for these differences is that the true MPs of the jets differ from the model ones
- Variation of α^g_{DAT} for "a complete set of independent jet MPs" gives "MODEL UNCERTAINTY" (M.U.)
- M.U. is for one model
- M.U. – lower limit of THEORETICAL UNCERTAINTY

- To determine M.U., we take the "complete set of independent jet MPs" $V_{1,2,,\dots,M}$ and average over all $V_{1,2,,\dots,M}$-bins:

$$\alpha^g \equiv \langle \alpha_V^g \rangle = \frac{\sum_{k=1}^{M} \sum_{V_k=1}^{N_{V_k}} \alpha_{V_k}^g}{N_V}$$

$$N_V \equiv \sum_{k=1}^{M} N_{V_k}$$

$$\text{Uncertainty} \quad \Delta\alpha^g \equiv \frac{\sqrt{\langle {\alpha_V^g}^2 \rangle - \langle \alpha_V^g \rangle^2}}{\sqrt{N_V}}$$

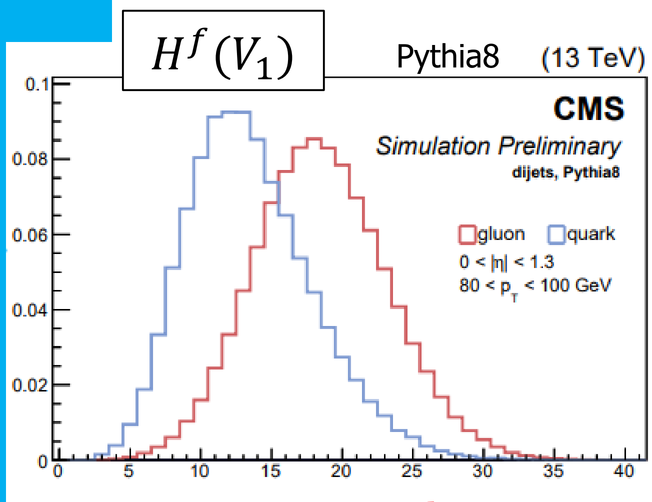
$\Delta\alpha^g$ includes stat. uncert. and syst. deviations of model q/g-templates for all MPs $V_{1,2,,\dots,M}$ from unknown q/g-templates in DATA

- MPs that are most sensitive to jet flavours [1]
  - Total multiplicity inside jet ($mult$) $= V_1$
  - Minor axis of jet ellipse in (η, φ)-space a_2 $= V_2$
  - "Fragmentation function" $p_T D = \frac{\sqrt{\sum_i p_{T\,i}^2}}{\sum_i p_{T\,i}} \in [0,1]$ $= V_3$

[1] CMS PAS JME-13-002
CMS PAS JME-16-003

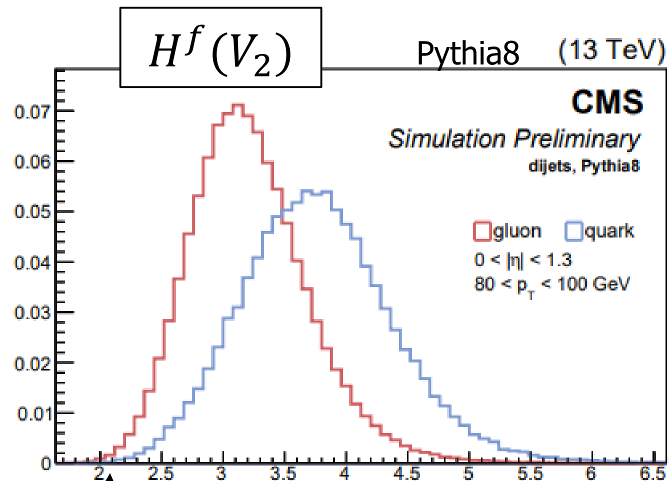
- We use these three MPs to measure q/g fractions

Fig. 1: q/g-templates



$V_1 = mult$

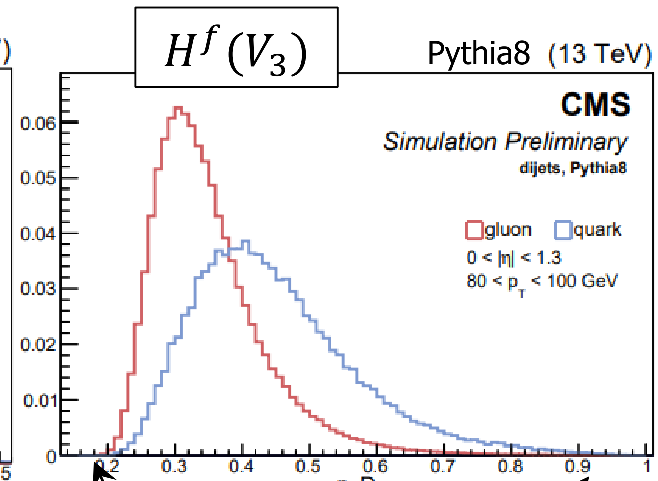
$\frac{n^g}{n^q} \approx 1.5$



$V_2 = -\log(a_2)$

Wide jets

Narrow jets



$V_3 = p_T D$

P_T^{jet} uniformly distributed among constituents

P_T^{jet} concentrated in a limited number of constituents

- Combined jet MP: discriminator Quark-Gluon Likelihood (QGL)[1] :

$$V_4 \equiv QGL = \frac{Q(\vec{V})}{Q(\vec{V}) + G(\vec{V})}$$

$$(V_1 = mult, V_2 = a_2, V_3 = p_T D) \equiv \vec{V}$$

$$Q(\vec{V}) = \prod_{i=1}^{3} H^q(V_i), \qquad G(\vec{V}) = \prod_{i=1}^{3} H^g(V_i)$$

[1] CMS PAS JME-13-002
CMS PAS JME-16-003

- The sensitivity of QGL to jet flavour is much stronger:

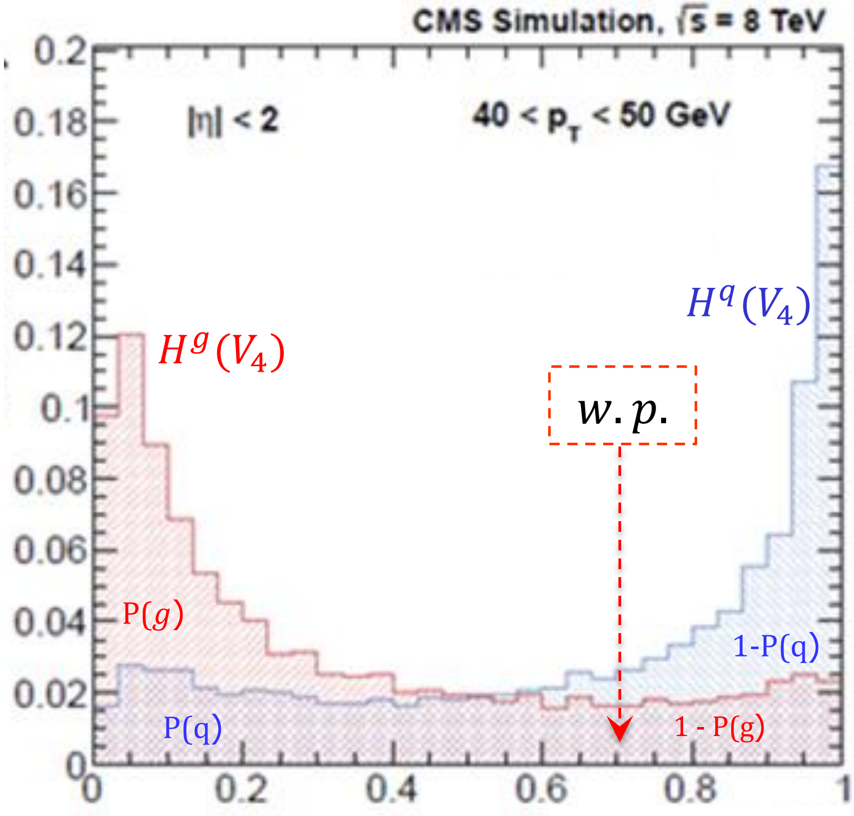


Fig.2: QGL-templates

- QGL-templates are used to mark q/g jets
- Measurement of g-fractions is good test for QGL:
  - We measured g-fractions with QGL templates and got incorrect g-fractions
  - We prepared new QGLs for CMS Run-2 and tested them by measuring g-fractions

Introduction
Scale Factor
Model Uncertainty (M.U.)
Jet macro parameters (MP)
QGL
CMS results
Gluon jet suppression
Summary

- Here are the results of measuring α^g using MP's: $V_{1,2,3,4}=mult, a_2, p_T D, QGL$

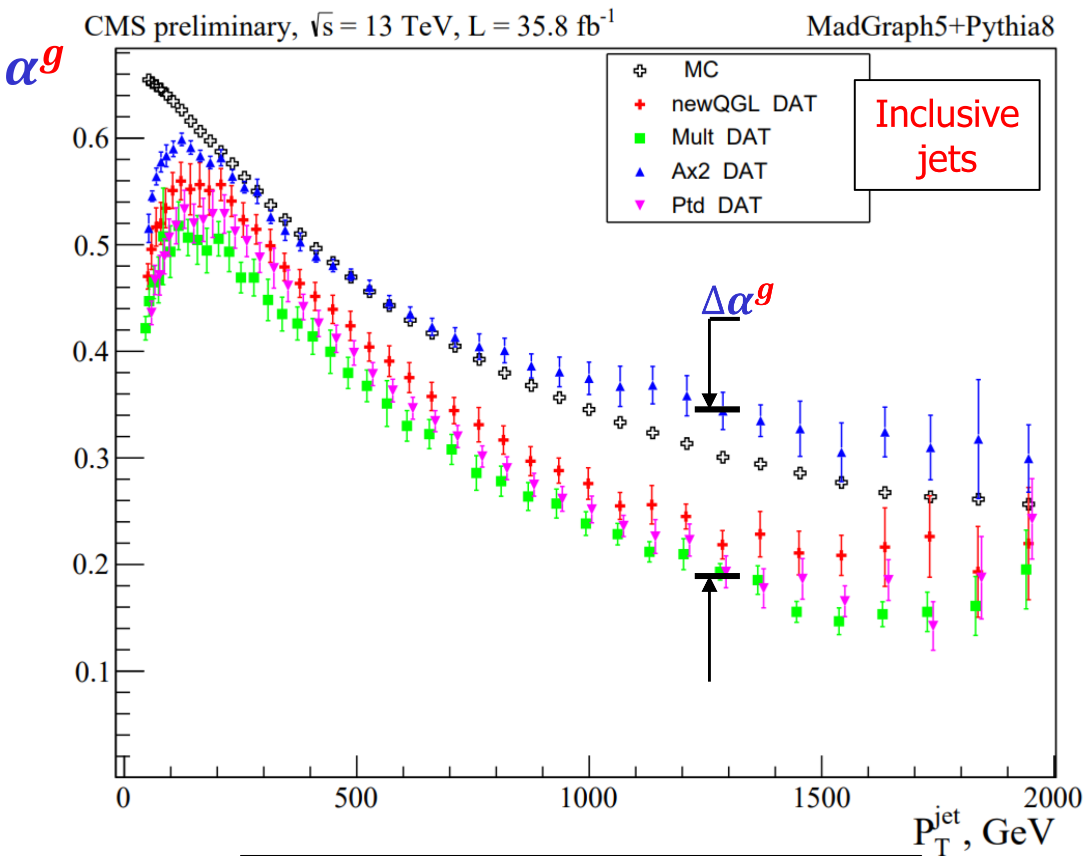


Fig. 3: Demonstration of M.U.

- This preliminary results were obtained in CMS group "Gluon-jet/Quark-jet analyses" [1]:

  S.S., D.Budkouski(JINR), J.Strologas (GR), O.Atakisi(TR)

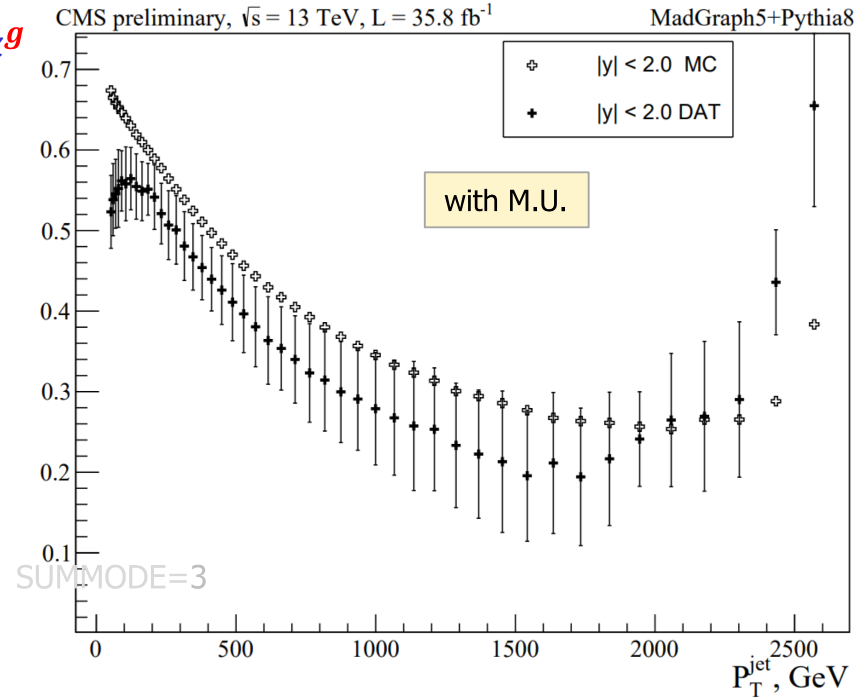
- This group was created in April 2021 purposefully to measure g-fractions in inclusive jet channel with Run-II data

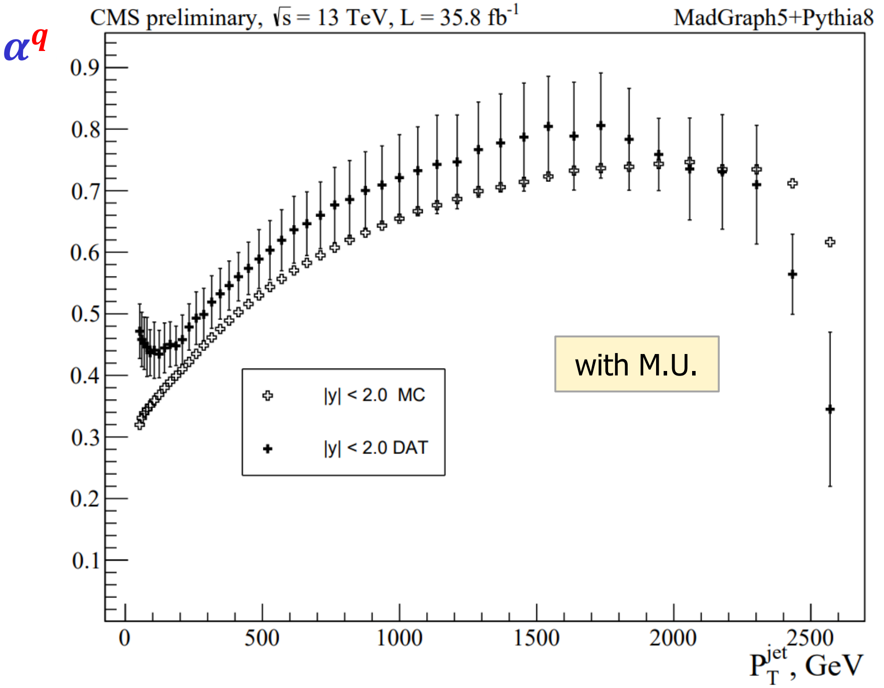
[1]https://indico.cern.ch/category/12755/

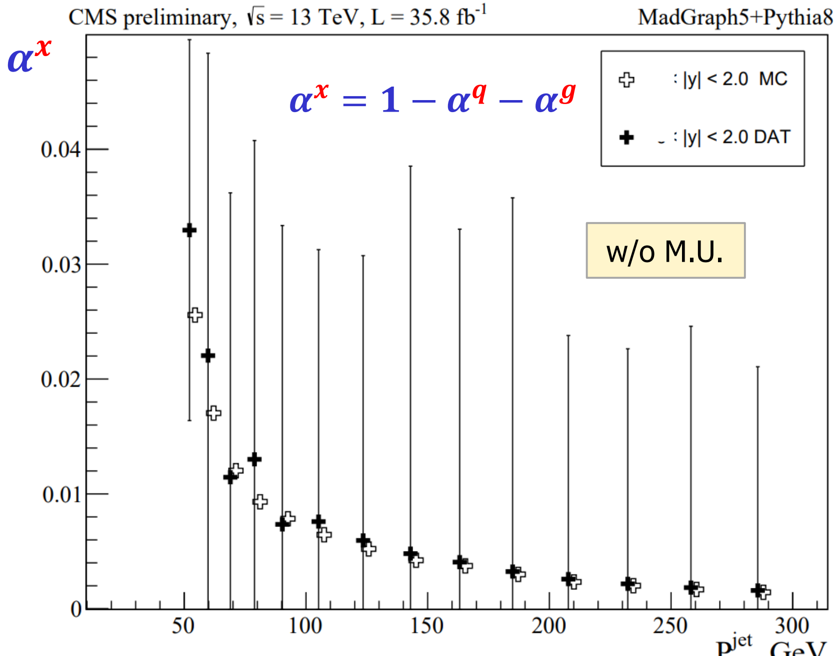
- Measurement of g-fraction demonstrates indirectly large deviation of true unknown DATA q/g-templates from Pythia8 ones

q/g/x-jet fractions

Inclusive jets







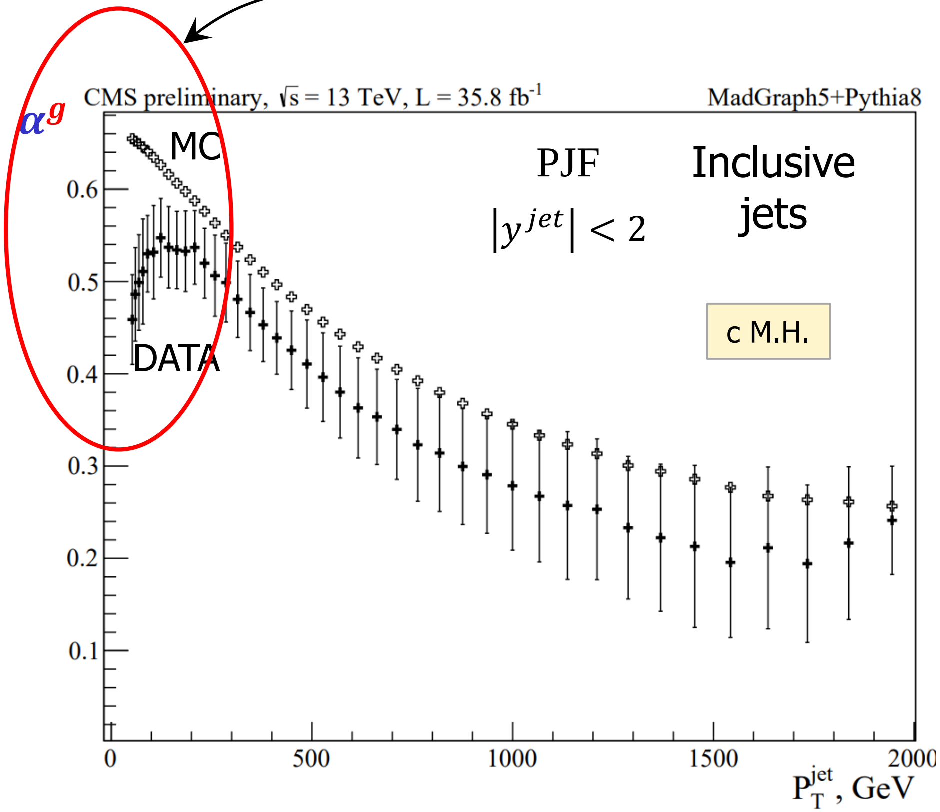
- This measurement results were obtained in CMS group "Gluon-jet/Quark-jet analyses"[1] :

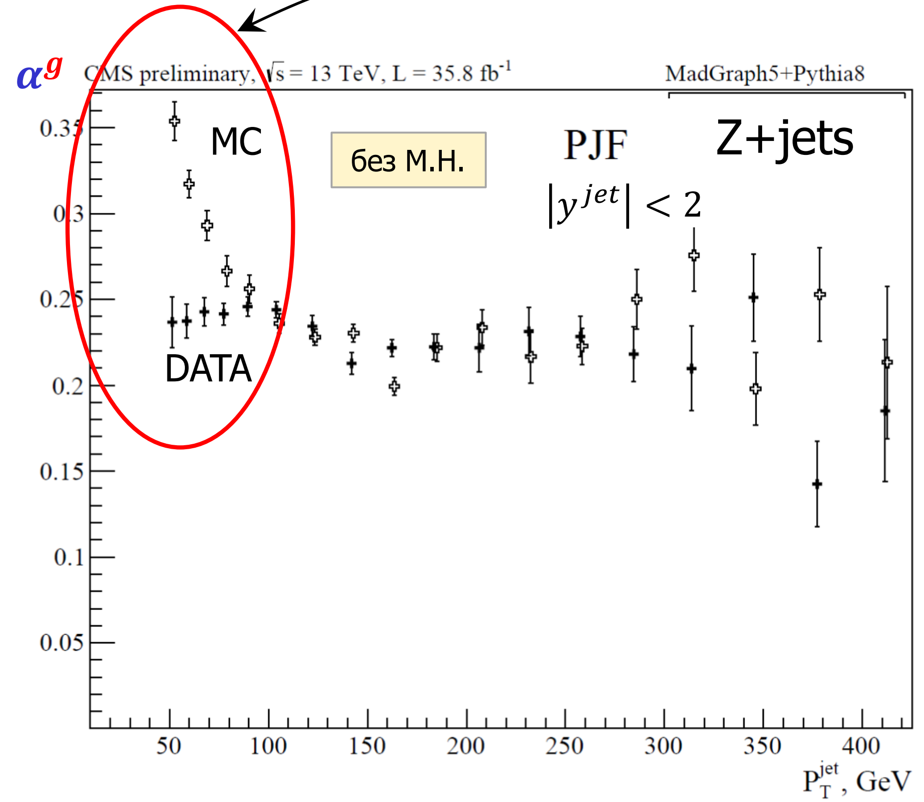
  S.S., D.Budkouski(JINR), J.Strologas (GR), O.Atakisi(TR)

The Figs. are taken from my talk SMP-HAD (June 2023)

Run-II(2016)

- **g**-jet suppression is visible at low P_T^{jet} in "**Inclusive jets**" and in "**Z+jets**"



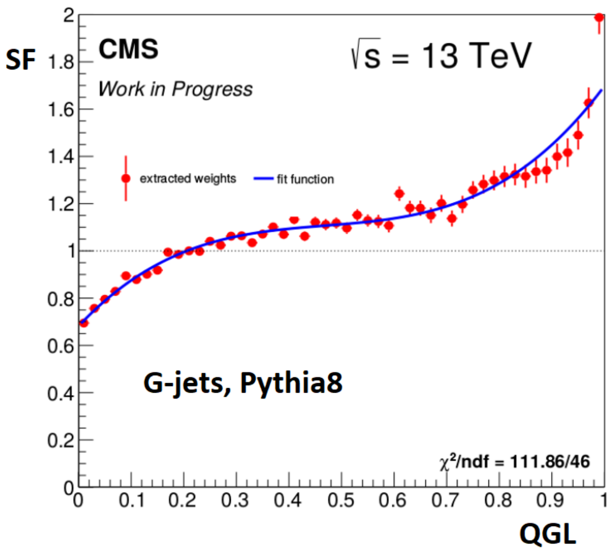


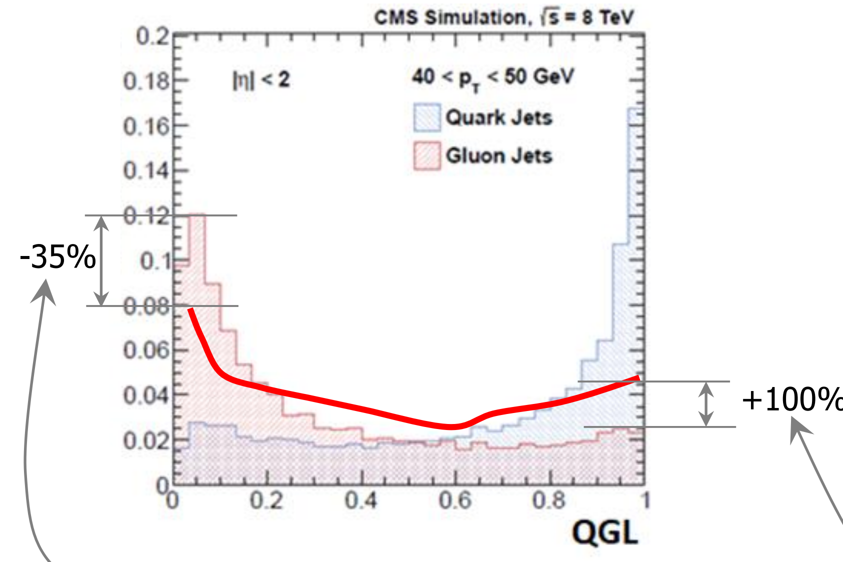
MadGraph5+Pythia8

JetFinder Anti-kT, R = 0.4

- We must attribute the 1st observation of g-jet suppression to those who calculated SF using the official method**:** Run-1(2013) and Run-2(2016) :

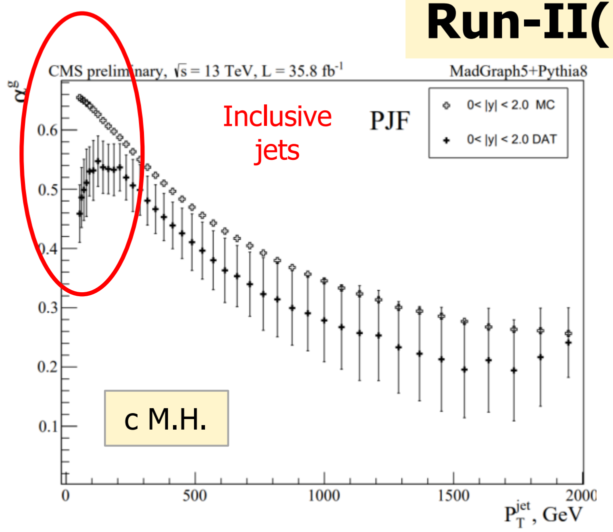
However, only after measuring g-fractions, we understand why gluon SF **was so large** - the reason is used MC g-fractions and that $\alpha^g_{\mathrm{DAT}} \ll \alpha^g_{\mathrm{MC}}$

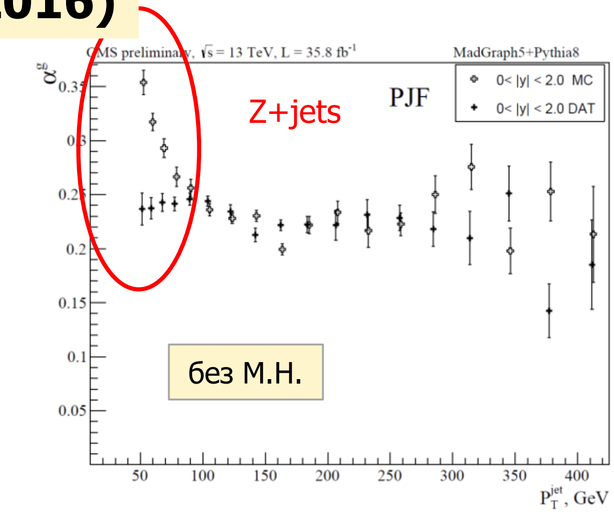




- The SF-modified g-template has a left peak 35% lower and a right peak 100% higher than the original MC g-template

Run-II(2016)

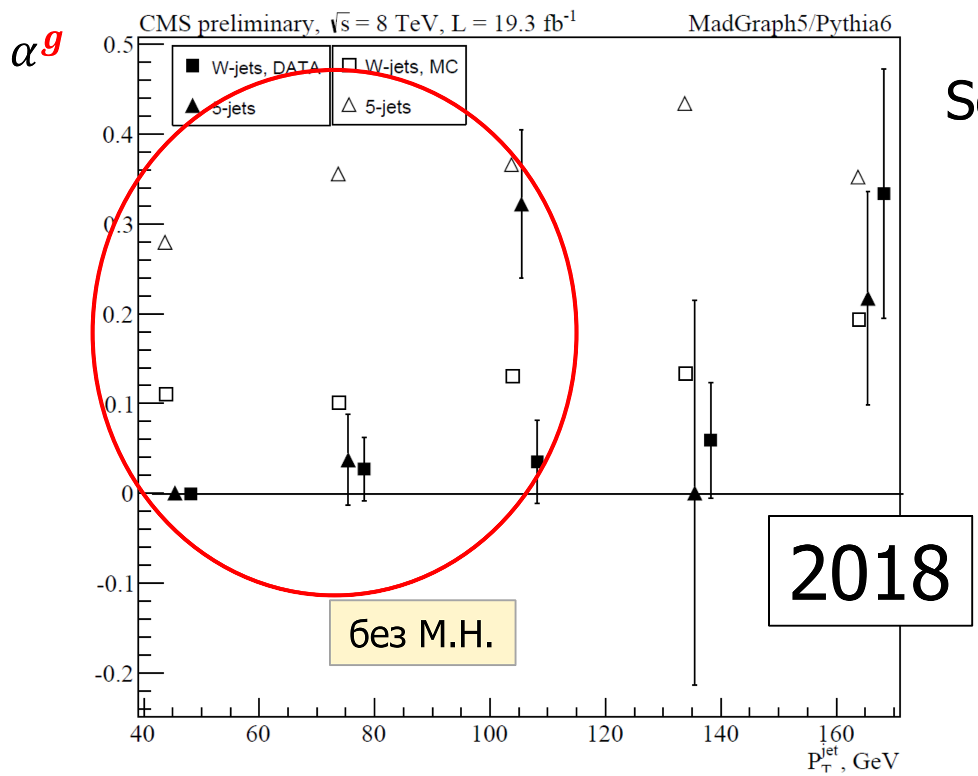


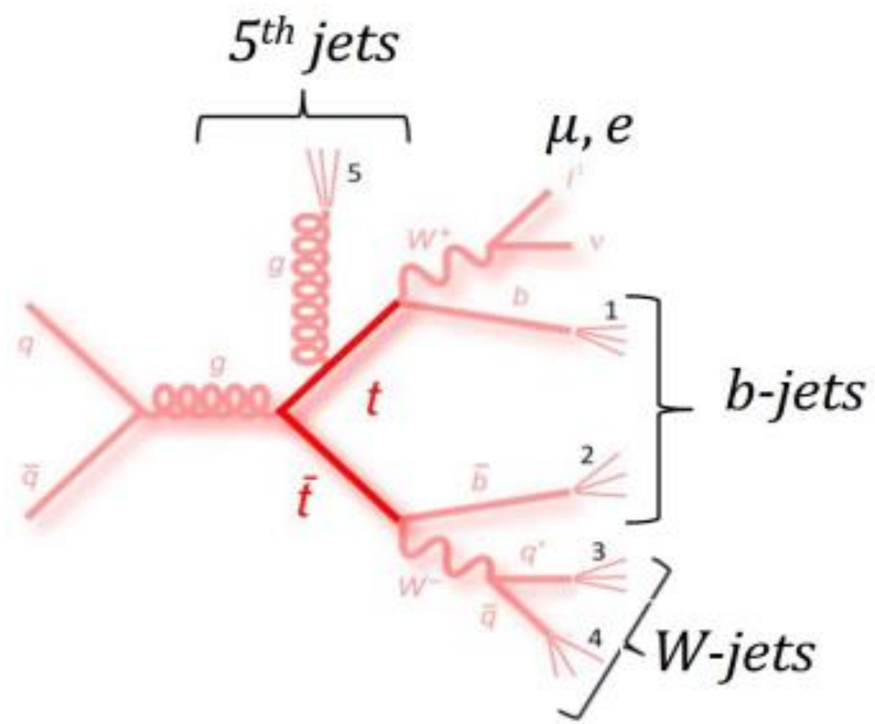


- **S**imilar results we obtained earlier for **Run-I (2012)**
- **R**un-I results are well documented:

S.S., S.Shmatov, A.Zarubin: CMS AN-2018-131, **2018**
S.S. D.Budkouski, CMS AN-2020-143, **2020**
S.S. D.Budkouski, CMS AN-2021-024, **2021**
S.S. SMP-HAD Workshop, 11 Feb **2020**, https://indico.cern.ch/event/861896/
S.S. SMP-HAD Meeting, 1 June **2018**, https://indico.cern.ch/event/732652/

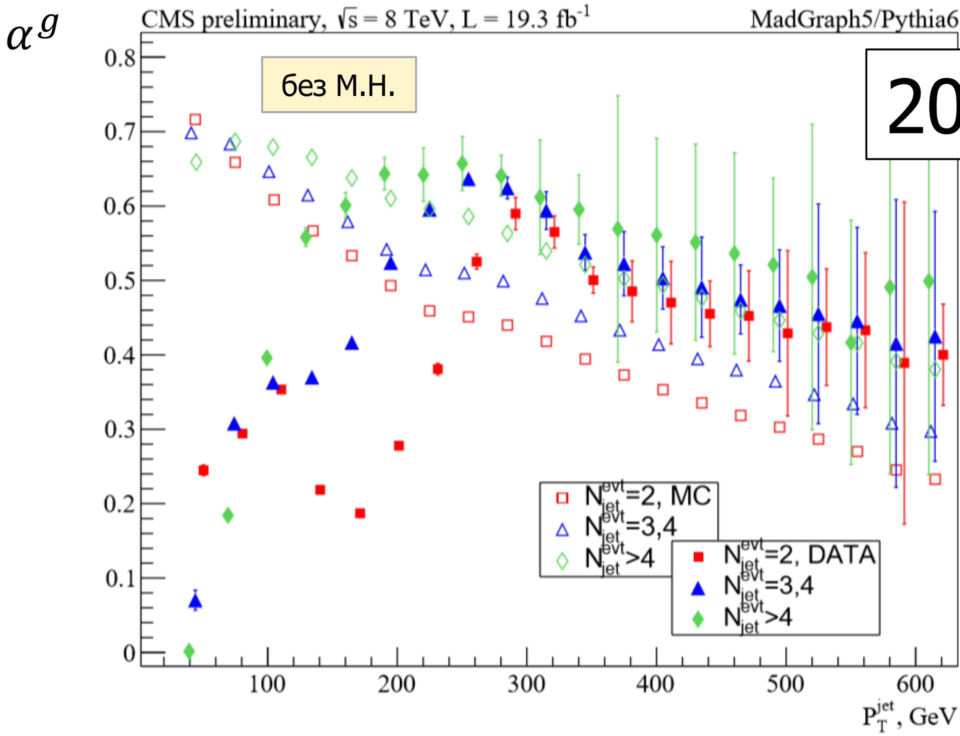
Semileptonic $t\bar{t}$ channel, Run-1(2012)





JetFinder Anti-kT, R = 0.5

N_{jets}^{evt}	Jet name	P_T^{jet}, GeV	$\alpha_k^{g,DAT}$, %	$\alpha_k^{g,MC}$, %
4	W-jets	30÷150	0÷5 (±5)	10÷11
≥ 5	5th-jets	30÷90	0÷3 (±5)	28÷34



2020

ak5-jets: R = 0.5

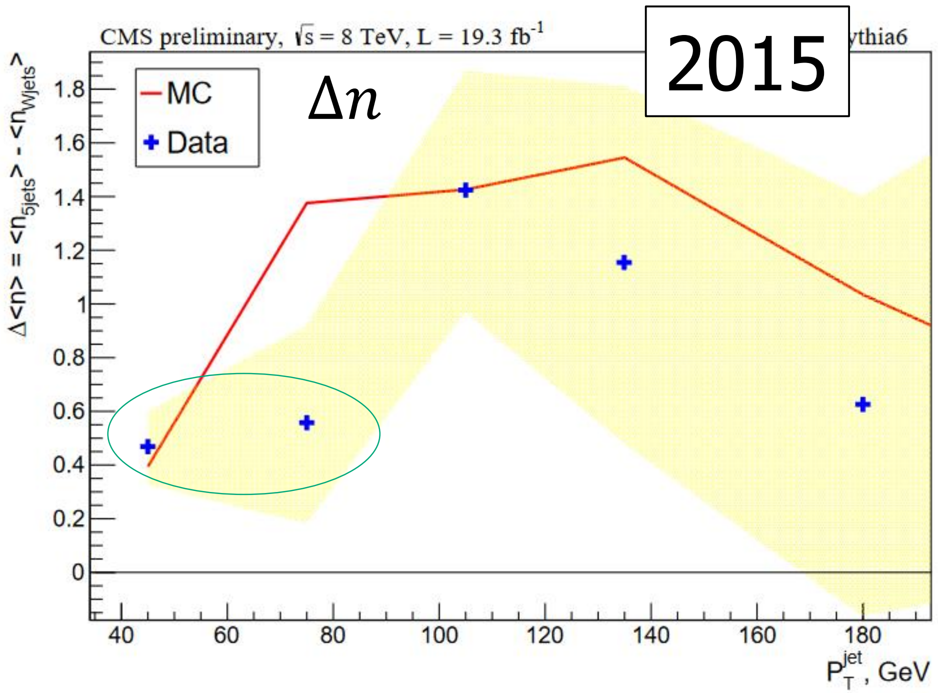
- Dijet, Run-I(2012)
- HLT prescaling is not taken into account

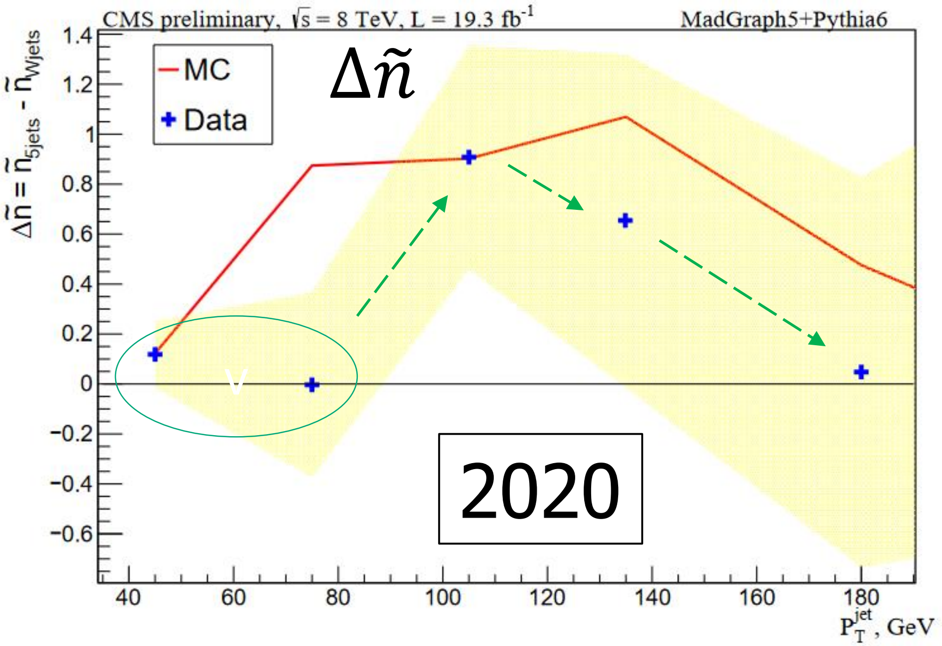
N_{jets}^{evt}	P_T^{jet}, GeV	$\alpha_k^{g,DAT}$, %	$\alpha_k^{g,MC}$, %	Sample name
2	30÷210	16÷35	72÷50	"dijet-1" (red)
3,4	30÷180	6÷40	70÷60	"dijet-2" (blue)
≥5	30÷120	0÷40	65÷69	"dijet-3" (green)
4	30÷150	0÷5 (±5)	10÷11	W-jets
≥ 5	30÷90	0÷3 (±5)	28÷34	5th-jets

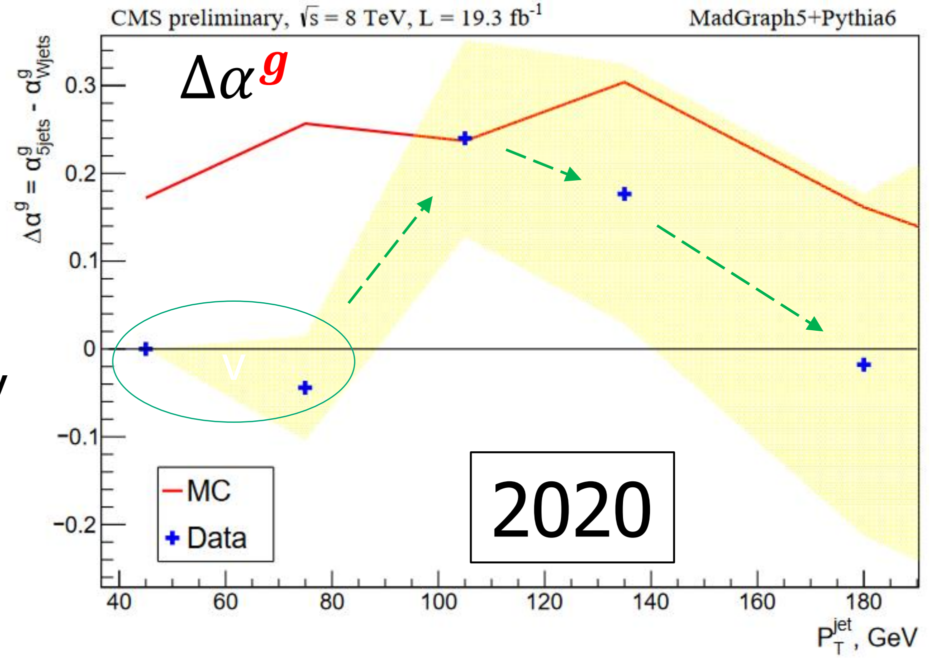
Semi-leptonic $t\bar{t}$

Run-I(2012) semileptonic $t\bar{t}$

$$A \cdot \Delta\tilde{n} = \Delta\alpha^{g}$$







- $\Delta\tilde{n}$ and $\Delta\alpha^{g}$ are similar:

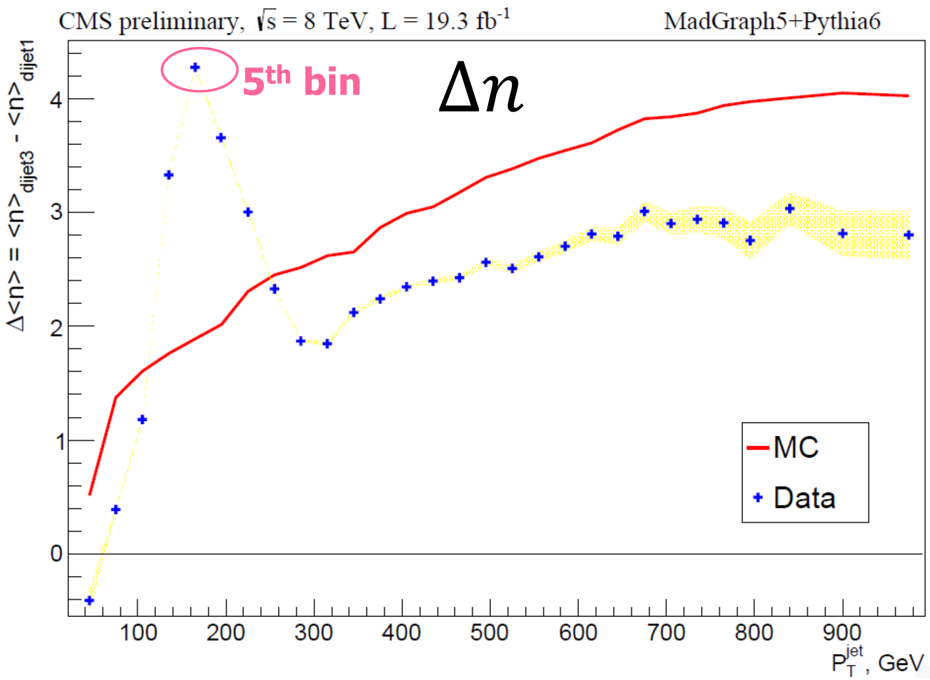
$\Delta\tilde{n} = A\,\Delta\alpha^{g} \approx 0$ in 1st and 2nd bins **!**

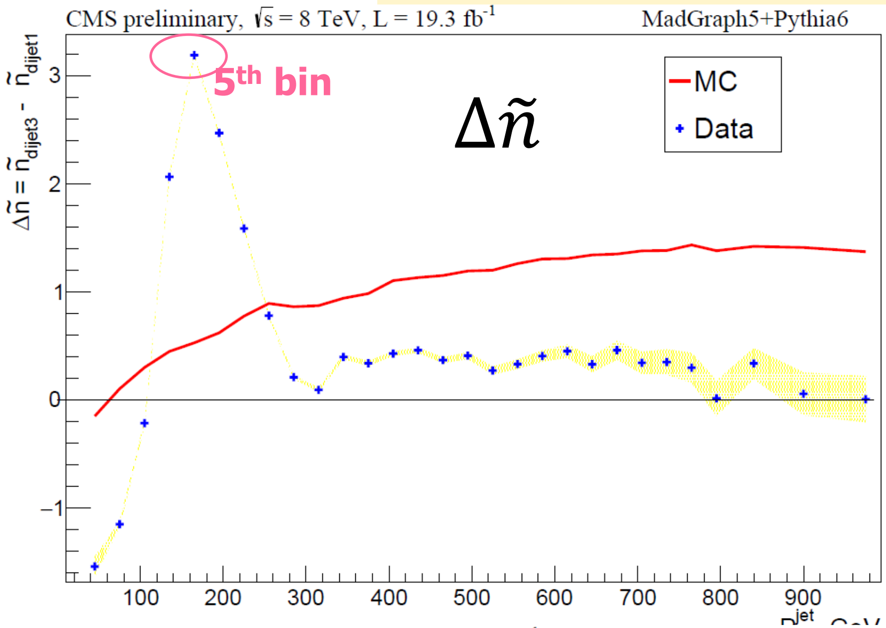
- Measurement of mean jet charged-particle multiplicity inside jet indirectly **confirms g-jet suppression**

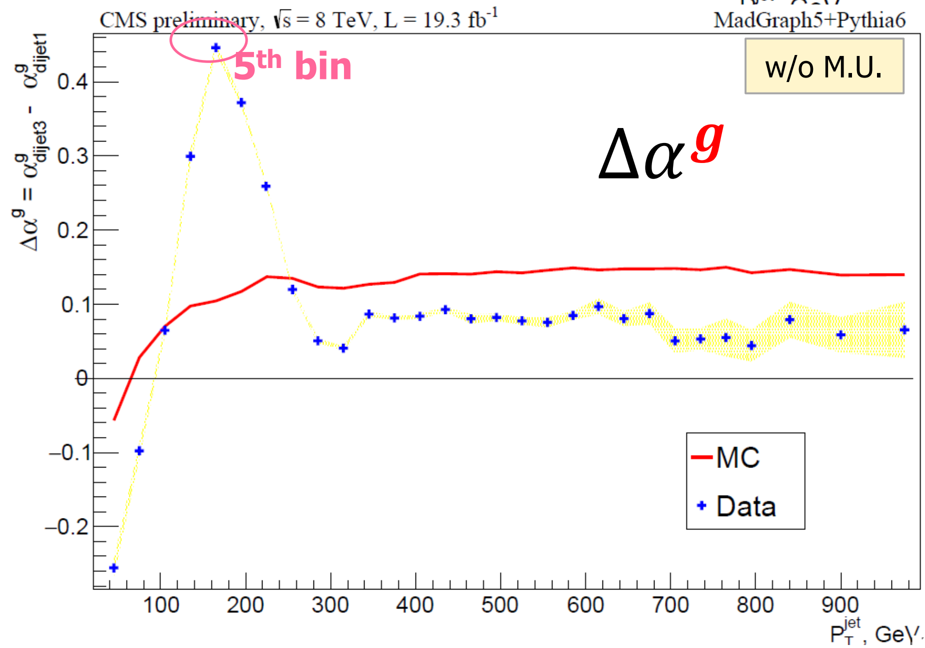
Run-I(2012) Dijet

2020

$$A \cdot \Delta\tilde{n} = \Delta\alpha^{g}$$







- $\Delta\tilde{n}$ and $\Delta\alpha^{g}$ are **similar** in all bins:

$$A \cdot \Delta\tilde{n} = \Delta\alpha^{g} \quad \textbf{!}$$

- Measurement of mean jet charged-particle multiplicity inside jet indirectly **confirms *g*-jet suppression**

- Measurement of g-fractions was proposed, developed and implemented for many channels in CMS (Run-1 and Run-2)
- "Data-driven SF's" ≈ 1: there is no task to find SFs for MC q/g-templates
- Model uncertainty: a large discrepancy between the Pythia8 hadronization model and the real picture of the process
- Suppression of g-jets in the region of low P_T^{jet} is observed (not approved in CMS yet, but work for inclusive jets is in final stage)
- Possible reason of g-jet suppression: gluon splitting at the beginning of parton branching, $g \to gg$, $g \to q\bar{q}$ – need to be investigated **experimentally**

Thank you very much for your attention!