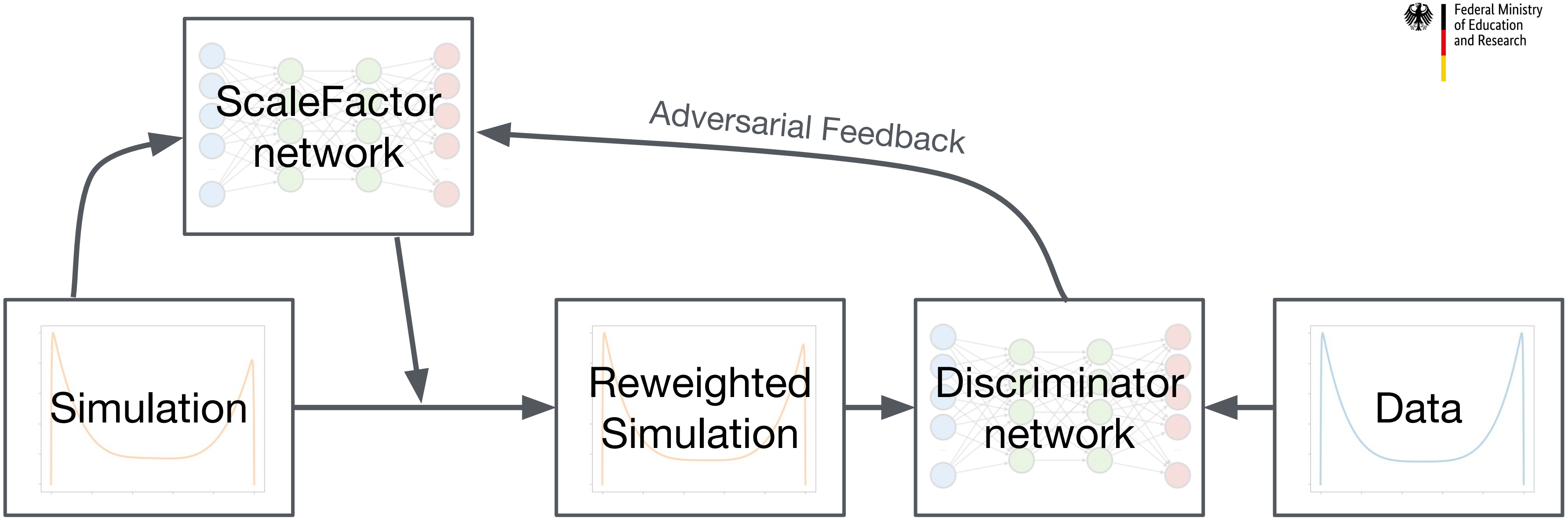


SPONSORED BY THE
Federal Ministry of Education and Research

Adversarial Neural Network-based Data-Simulation Corrections for Jet-Tagging at CMS

Martin Erdmann, Benjamin Fischer, Yannik Rath, Marcel Rieger
on behalf of the CMS collaboration

ACAT 2019 - Saas Fee

13th March 2019

“Traditional” iterative scale factors

Purpose

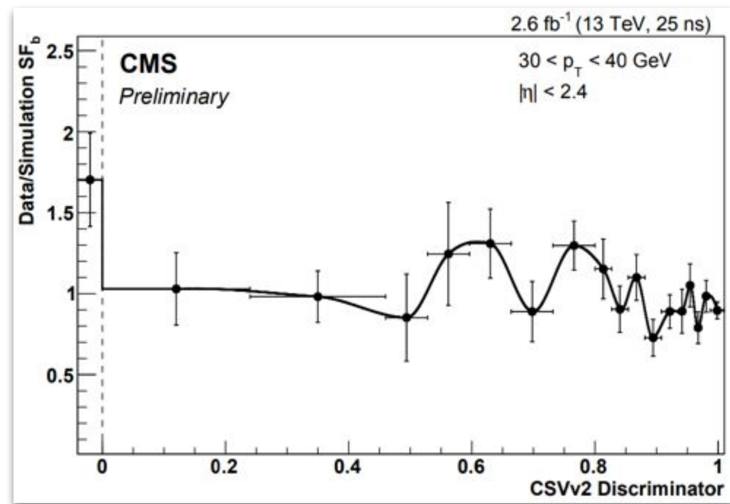
➜ Reduce data-simulation differences
in b-tag discriminator shape through event weights

Method: Tag & Probe (1712.07158)

- flavor f (heavy & light) enriched regions R
- Infer jet Scale-Factor per bin as:

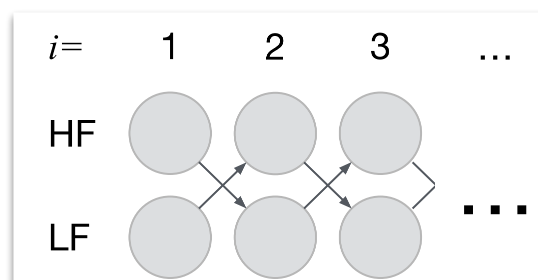
$$SF_{f,i+1} = \left.\frac{Data - \omega_i \cdot MC_{\neg f}}{MC_f}\right|_{R=R(f)}$$

$$\omega = \prod_j^{\text{jets}} SF_j$$



CMS-PAS-BTV-15-001

- Repeat iteratively (i) as SFs depend
on those of other flavor ($\neg f$)



Result $SF_f(p_T, \eta, b\text{-}tag)$

Binned in p_T, η, b-tag value of probe jet

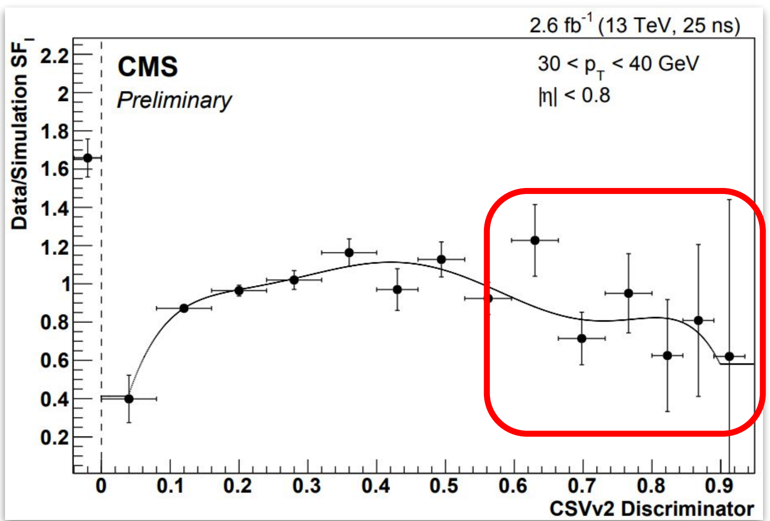
➜ bins must have sufficient statistics

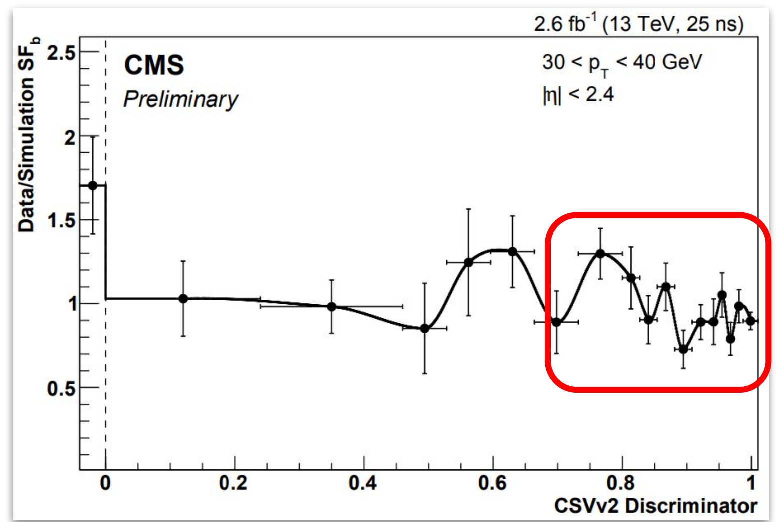
➜ number dependent variables limited

Smoothen by fitting a function to bins

➜ needs adjustment to prevent overfitting

CMS-PAS-BTV-15-001





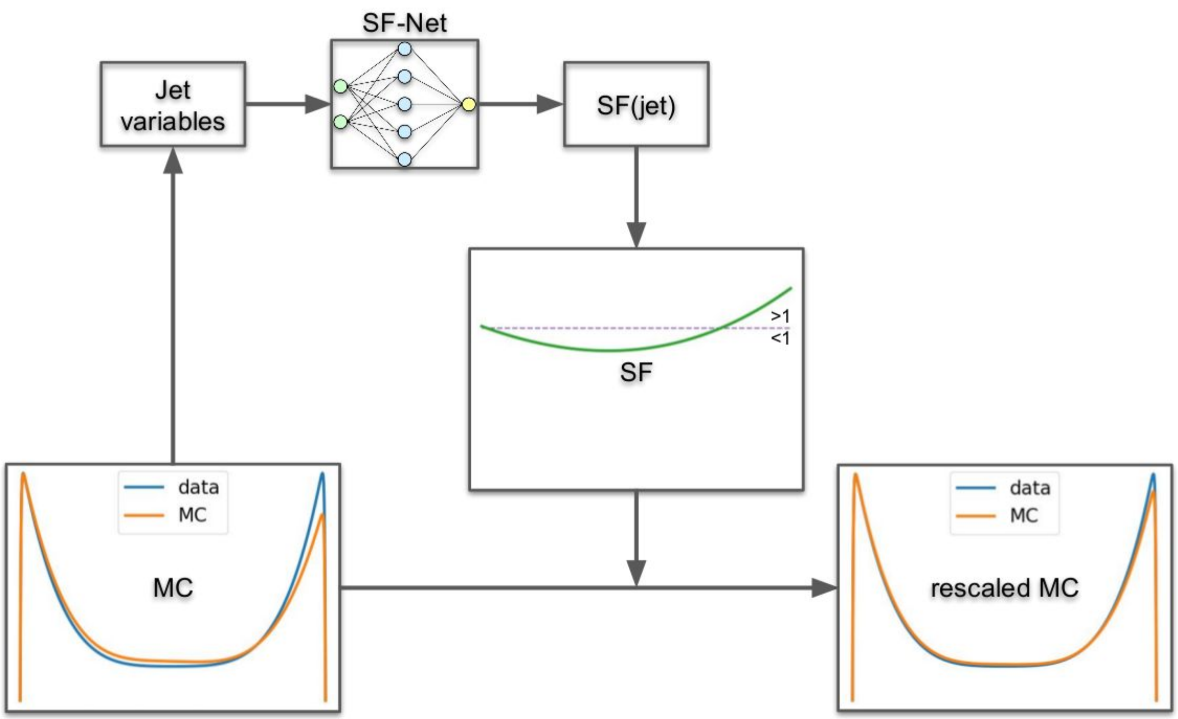
CMS-PAS-BTV-15-001

2 enriched regions R

➜ does not readily translate to multiple flavors/regions

Model entire function $SF_f(p_T, \eta, b\text{-}tag)$ by a Neural Network

- ✔ "fit function" can model anything
- fits according to available statistics (controlled by regularization)
  - ✔ no need for prior binning
  - ✔ can handle more inputs

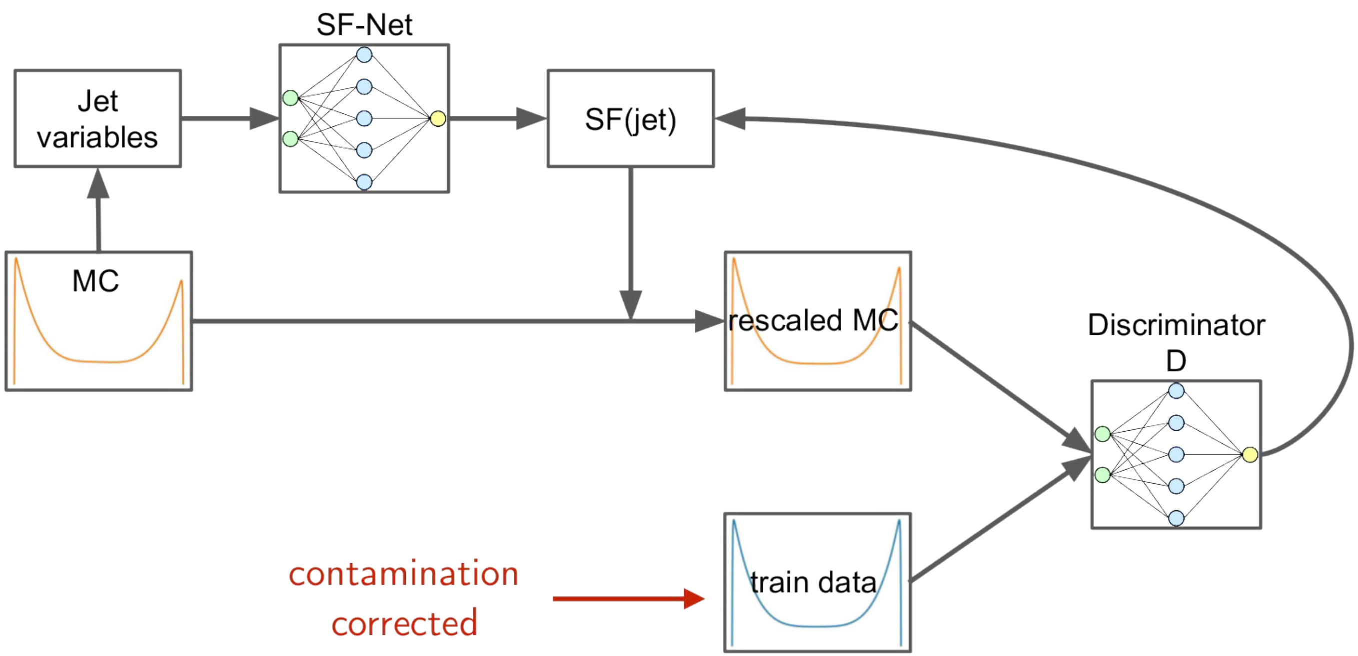


But: how to train it?

- training target is an distribution

➜ encode into discriminator (adversary)

✔ can cover multiple regions



SF-net:

- Loss: $MSE\left(SF, \frac{Data}{MC} = \frac{D}{1-D}\right)$

Discriminator D (Adversary):

- Loss: SF weighted cross-entropy between Data and MC events

Architecture and training schedule

Benjamin Fischer, 13th Mar. 2019

Architecture

- fully connected
- Same for all
  - SFnet/Discriminator
  - heavy/light flavor region

Layers	10
Units	128
Activation	LeakyRELU @ 0.01
Weight decay	3e-5
Batch size	4096
Train/Valid. split	80/20

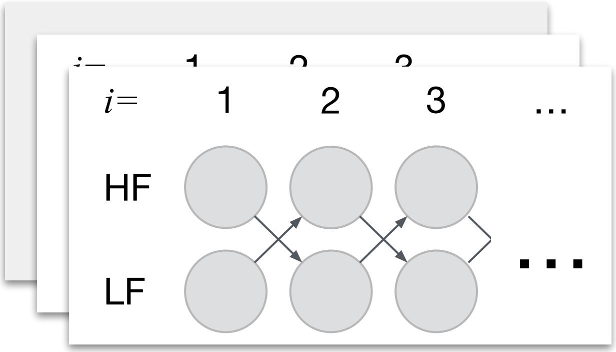
Training schedule

1. Train discriminator until converged
2. Train SFnet for 1 batch
3. Repeat until SFnet converged

- Typically finished after ~50 epochs (no limit)

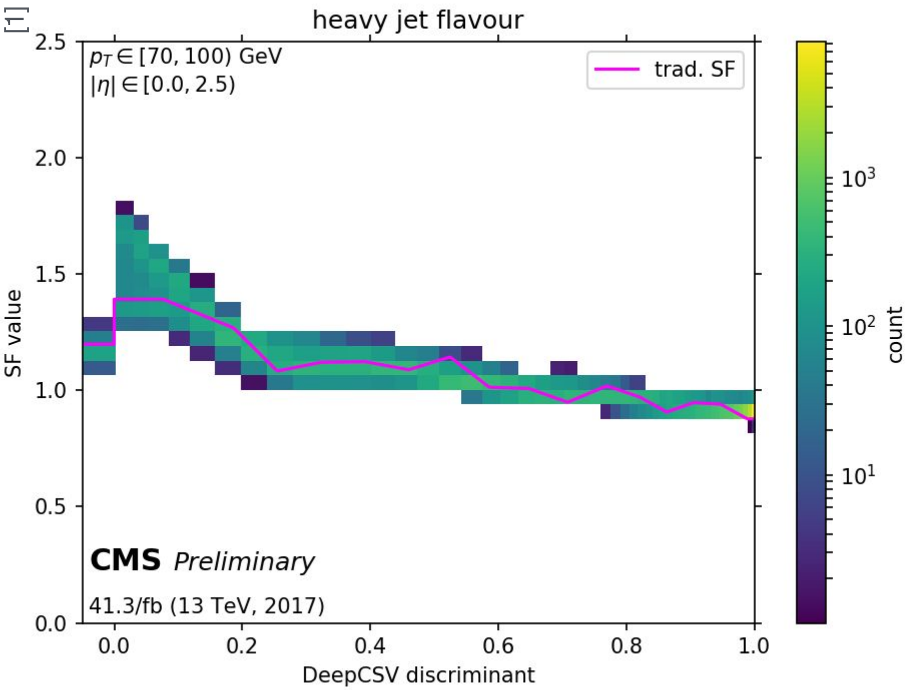
Ensembling

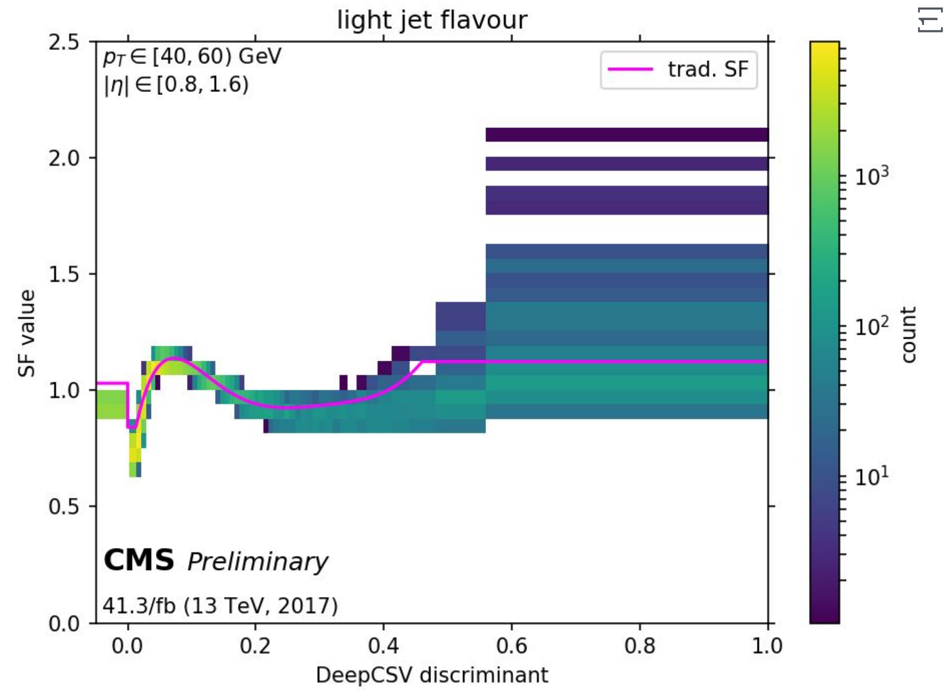
- mean output of 25 fully separate trainings
- variance indicates stability of procedure



Scale Factor Comparison - Histogram

Benjamin Fischer, 13th Mar. 2019





Shown

- $SF_f(p_T, \eta, b\text{-}tag)$vs. $b\text{-}tag$ value
- in specific (p_T, η) bins of traditional SFs
- for all probe jets of particular flavor
- using same events as traditional SFs (DiLepton of ttbar & DY, 2017)

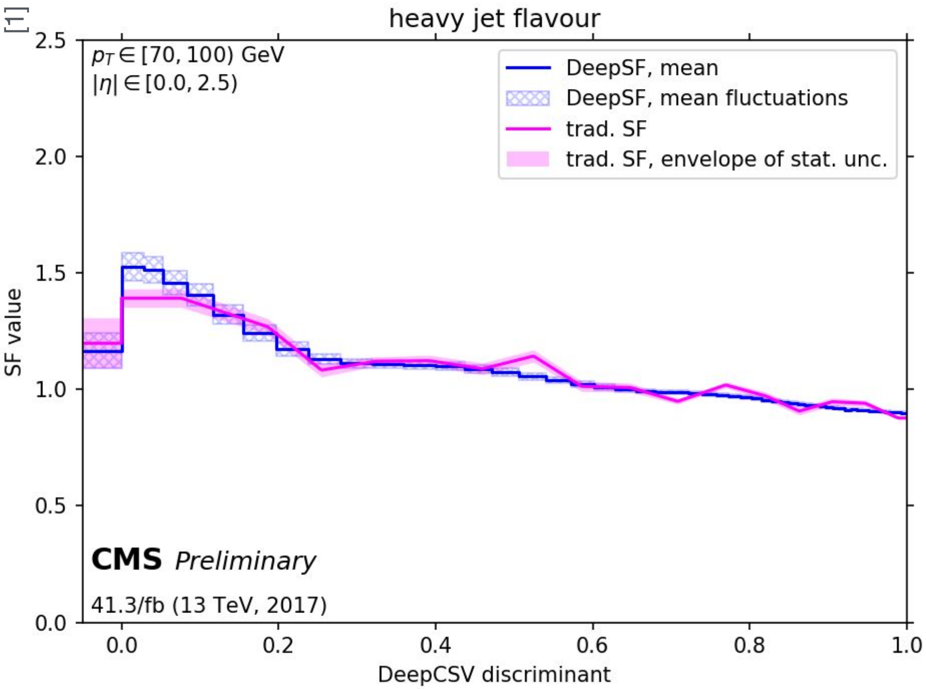
Observations

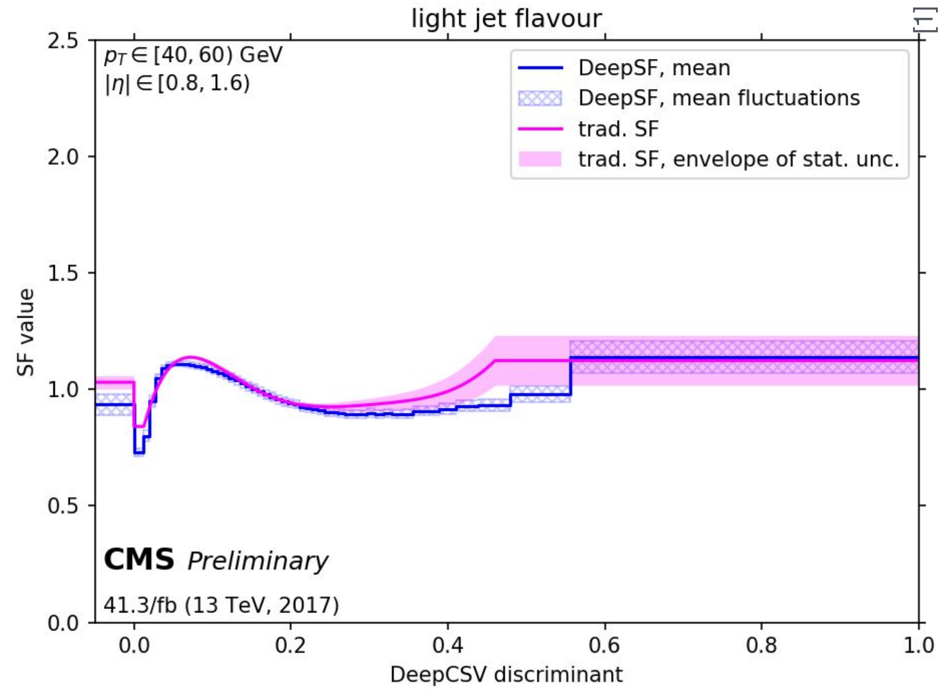
➜ **Traditional SF** have fixed value

➜ **DeepSF** one value per jet

- broader distribution where needed

Scale Factor Comparison - Mean





DeepSF

- mean of each bin's entries
- hatched: ensemble fluctuations (standard error of mean)

Traditional SF

- w/ envelope of systematics derived from (limited) statistics

Observations

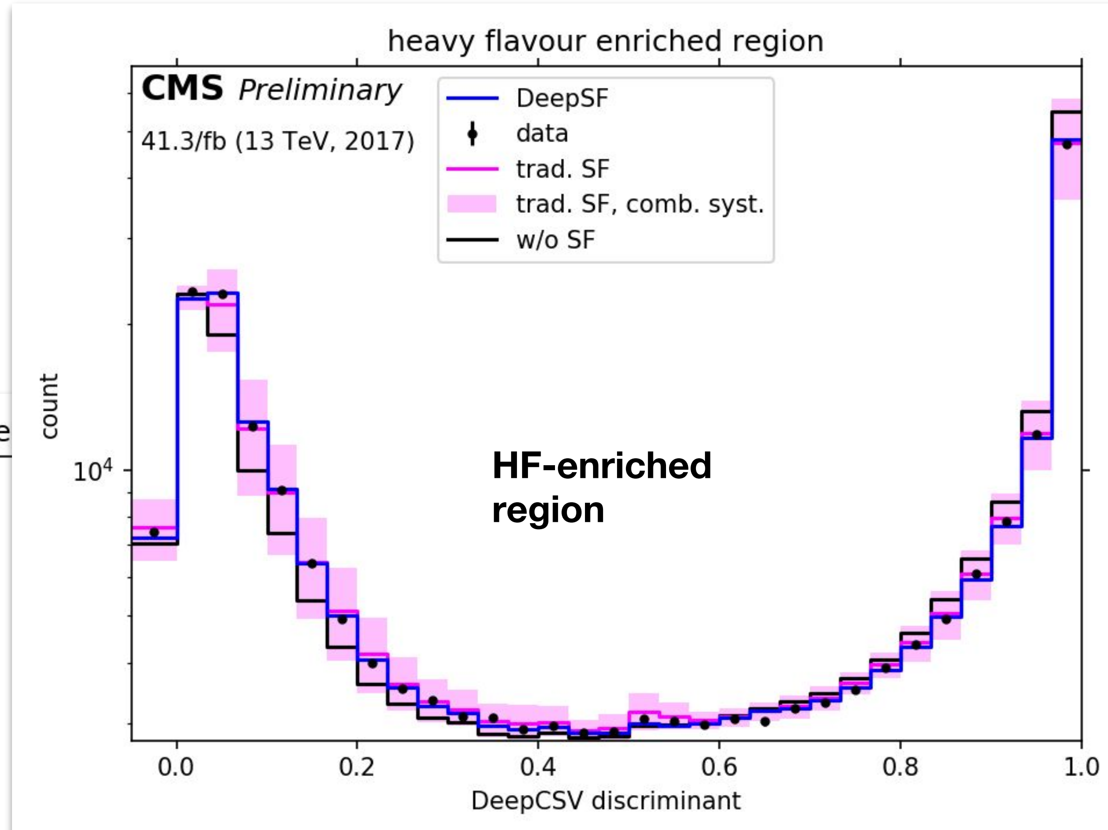
- Ensemble fluctuations:
  - ➜ encoding limited statistic
  - ➜ not the (sole) cause of broadened deepSF values

1: CMS-DP-2019-003

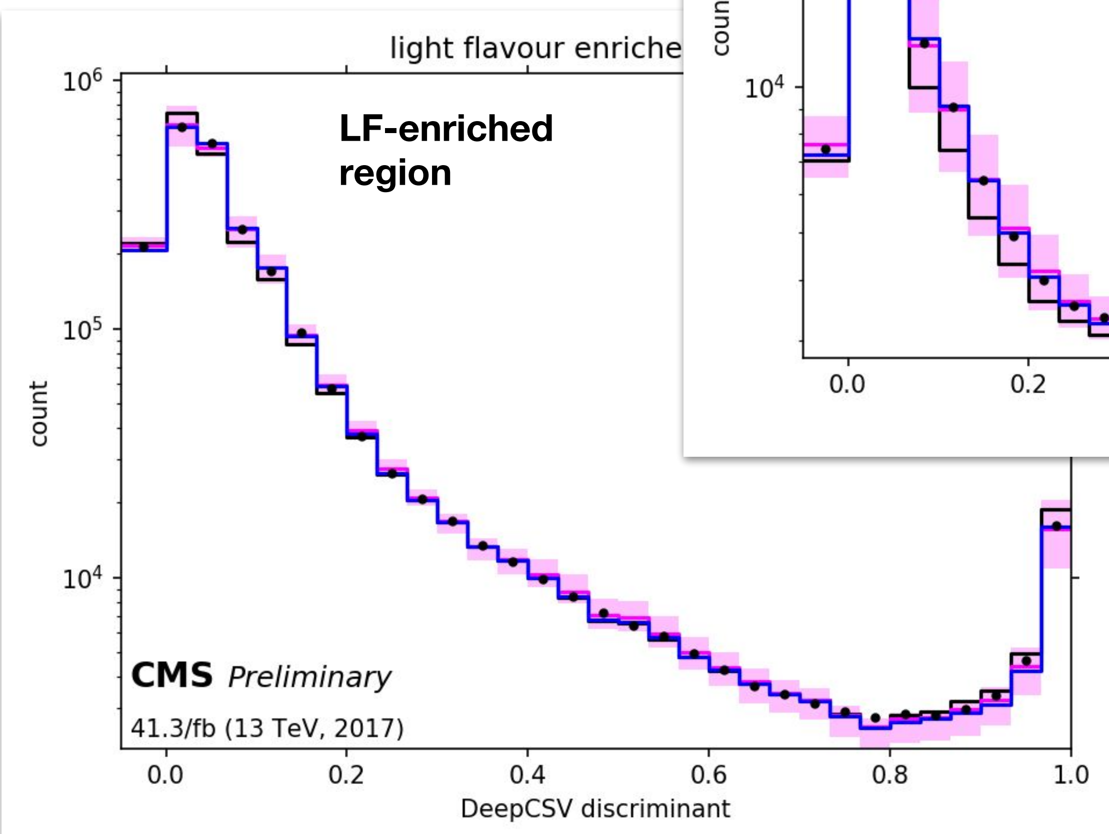
Effect on b-tag distributions

Shown

- Applied to all probe jets of each region
- vs. *b-tag* value



CMS-DP-2019-003

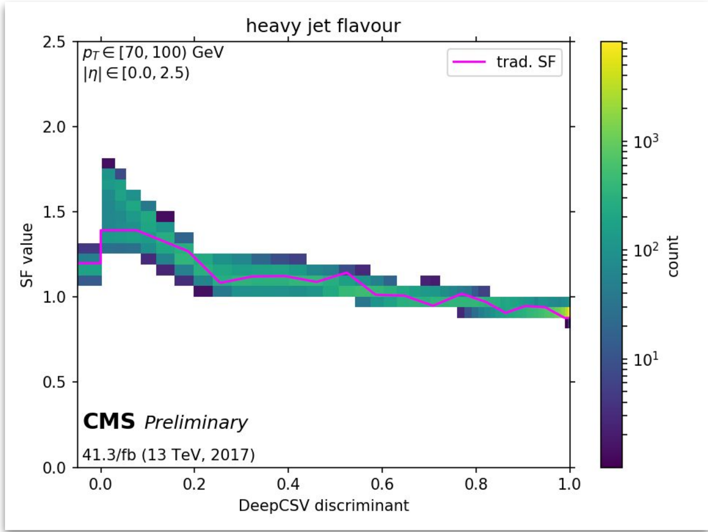


CMS-DP-2019-003

➜ Overall good agreement

Summary & Outlook

- B-tagging methods becoming very sophisticated
  - ➜ SFs may depend on many variables
  - ➜ Need to “babysit” binning & fit functions in traditional SF
- DeepSF method working well
  - ➜ Good agreement with traditional SF
  - ➜ Extensible to more
    - ■ input variables
    - ■ flavor-enriched regions
- In progress
  - ○ Processing all systematics
  - ○ Testing on real-life analyses
  - ○ Consolidating network-structure & training method



CMS-DP-2019-003