Scale factors implementation and event yields crosscheck

Huiling Hua [1] **Fabio Iemmi** [1]
Hongbo Liao [1] Hideki Okawa [2] Yu Zhang [2]

[1]Institute of High Energy Physics (IHEP), Beijing

[2]Fudan University, Shanghai

June 23, 2021

b tagging scale factors

- There are **some important items to be discussed** regarding b tagging scale factors
  - How to evaluate jet flavor
  - Changes to the scale factor evaluation function that I shared with you
  - Phase space extrapolation factors

How to evaluate jet flavor

- b tagging **scale factor computation** in the reshaping framework **needs the jet flavor**
- Following the instructions in Recommendation for Using b-tag Objects in Physics Analyses:
  - **Must use** the so called **hadron-based** definition
  - It is `Jet_hadronFlavour` in BSM code
  - Not OK to use parton-based definition: there may be rare but possible mismatches wrt hadron-based
- **Let's make sure we are doing thing correctly here**

Changes to the scale factor evaluation function

- While setting up macro for SF implementation, reviewed the code
- The **code** was good but **not perfect**
- I believe we have to change a couple of points:
  - Switch to `eval_auto_bounds` method (instead of simple `eval`)
  - Remove some sources of systematic uncertainty where they're not needed
  - Add JES uncertainty for c-flavored jets
  - Understand the new sources of uncertainties
- **Details in the coming slides**

`eval_auto_bounds` method

- SF computation relies on the values of jet p_T $|\eta|$, flavor, b-tagging discriminator
- If you pass **values outside of the bounds**, `BTagCalibrationReader::eval` will **return 0**
- p_T and discriminator have variable bounds, hard to deal with them by hand
- **`BTagCalibrationReader::eval_auto_bounds()`** does that for you
- Following the recommendations, this method increases the systematic uncertainty when inputs are out of bounds

Removing some sources of systematics

- In my first implementation, for each flavor of jet I **implemented all the systematic sources**
- **Assumed** that some of them would be **==1 when not involved**
  - E.g., there's no HF contamination uncertainty when dealing with HF jets (see AN2013_130)
- Realized that there are (rare) cases in which this is not true
- **Removed all the unneeded systematics**: safer!

Adding JES uncertainty to c-flaovred jets

- About this, I simply forgot: my understanding is that **all flavors need the JES uncertainty** to be applied
- Now I'm implementing it

Data type of the b tagging discriminant

- I found a **potential bug in my code**, worth discussing it
- b tagging discriminator was saved as a `std::vector<double>`, but I was reading it as a `std::vector<float>`
- **`SetBranchAdderss` was not complaining** but...
- b tagging discriminator values were often out of the expected [0,1] range!
- This can cause problems with `eval_auto_bounds`

New sources of uncertainty

- Recently found that in DeepJet .csv file there are more sources of uncertainty wrt to old CSVv2 .csv file
- They come from JES uncertainty, see here BTagShapeCalibration Twiki
- There's a Twiki called Jet energy scale uncertainty sources
- It may explain the meaning of each source, **plan to read it** (didn't do it yet)

Phase Space Extrapolation Factors (PSEFs)

- From BTagShapeCalibration Twiki
- **Before** applying any **b-tag selection** requirement, **yields should be preserved** before and after SF application
- Analyses phase spaces are various and arbitrary, so no reason for this to happen in a particular analysis
- **Analysts should compute** $\boxed{\sum \omega_{\mathrm{bef}}/\omega_{\mathrm{aft}}}$ and use this as a **phase space extrapolation factor**
- If analysis has many jets, consider doing it per jet multiplicity
- We have b tagging requirements since the very beginning (preselection), so **maybe we don't need this** (?)

My status

- **Implemented** all these **changes** in my macro
- I **need to implement the remaining SFs** (this should require less work)
- **Should have scaled yields very soon**

Analysis timeline

- I presented a progress report during last IHEP CMS group meeting
- **I was asked by Joshuha** Thomas-Wilsker (tt+X convener) **about the timeline** of the analysis
- I believe we should clarifiy it a bit
- I chatted with Josh after the meeting. He suggests to have
  - either a chat between us, him and Olaf (2nd tt+X convener)
  - or a short presentation in tt+X to ring the bell and discuss about plans
- **Other 4tops analyses are running towards unblinding**
- There's lots to do!