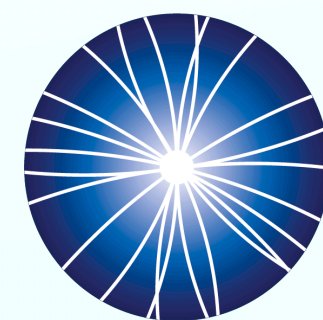


A Quick Intro to Snakemake

Miguel Fernández Gómez
LHCb Starterkit 2024



IGFAE

Instituto Galego de Física de Altas Enerxías



**XUNTA
DE GALICIA**



Here is a very real situation...

- Let us say I want to study $K_S \rightarrow \mu\mu$ ()
- I need a normalization channel, usual choice: $K_S \rightarrow \pi\pi$

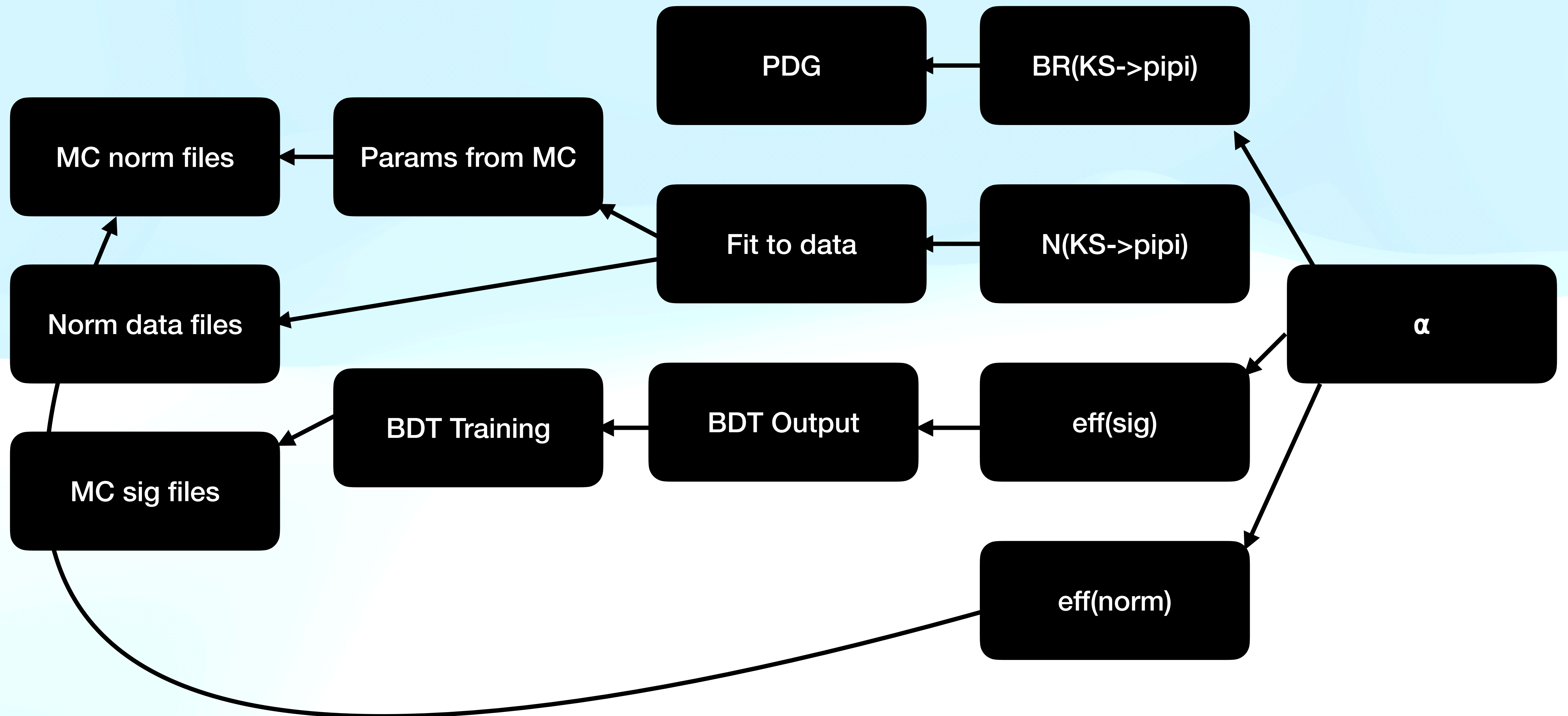
- $$N(K_S^0 \rightarrow \mu\mu) = N(K_S^0) \times BR(K_S^0 \rightarrow \mu\mu) \times \mathcal{E}_{sig}$$

- $$N(K_S^0 \rightarrow \pi\pi) = N(K_S^0) \times BR(K_S^0 \rightarrow \pi\pi) \times \mathcal{E}_{norm}$$

- $$BR(K_S^0 \rightarrow \mu\mu) = \alpha \times N(K_S^0 \rightarrow \mu\mu)$$

- $$\alpha = \frac{BR(K_S^0 \rightarrow \pi\pi)}{N(K_S^0 \rightarrow \pi\pi)} \times \frac{\mathcal{E}_{norm}}{\mathcal{E}_{sig}}$$

I need to compute α



Introducing Snakemake

- This is an example of an analysis workflow
- We have tools to automate this: Snakemake
- Pros:
 - We can keep track of the whole flow of the analysis in one place
 - This is useful for us but also other people who may want to re-run it in the future
 - It forces us to structure the code
- Useful links:
 - <https://hsf-training.github.io/analysis-essentials/snakemake/README.html>
 - <https://snakemake.readthedocs.io/en/stable/index.html>

Introducing Snakemake

- A workflow management system allows you to:
 - Keep a record of how your scripts are used and what their input dependencies are
 - Run multiple steps in sequence, parallelising where possible
 - Automatically detect if something changes and then reprocess data if needed
- Using a workflow management system forces you to:
 - Keep your code and your locations in order
 - Structure your code so that it is user-independent
 - Standardise your scripts
 - Bonus: Standardised scripts can sometimes be used across analyses!

Introducing Snakemake

- A workflow management system allows you to:
 - Keep a record of how your scripts are used and what their input dependencies are
 - Run multiple steps in sequence, parallelising where possible
 - Automatically detect if something changes and then reprocess data if needed
- Using a workflow management system forces you to:
 - Keep your code and your locations in order
 - Structure your code so that it is user-independent
 - Standardise your scripts
 - Bonus: Standardised scripts can sometimes be used across analyses!

The basics of Snakemake

- Snakemake operates with rules
- Each rule is given:
 - A name
 - An input
 - An output
 - A shell command to generate the output from the input
 - There are more optional parameters you can check in the documentation
- At this point we'll move on to the online lesson

The basics of Snakemake