Improving Air Shower Simulations by Tuning Pythia 8/Angantyr with Accelerator Data

Michael Windau, Chloé Gaudu,
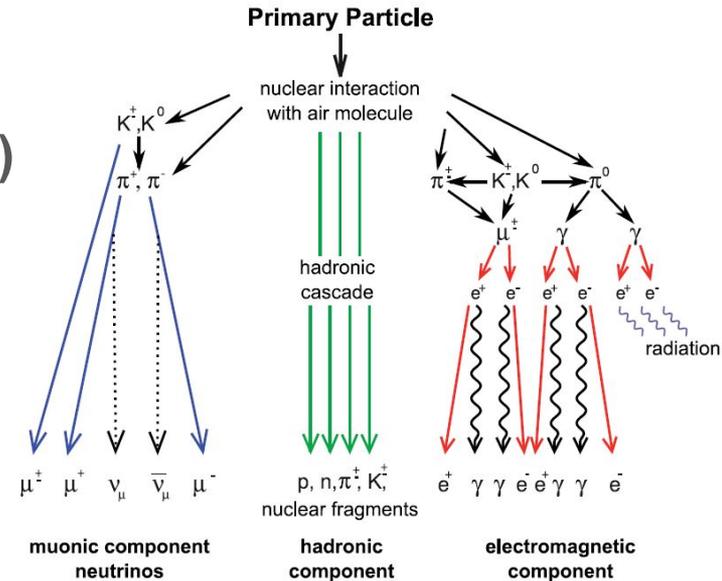Karl-Heinz Kampert, Kevin Kröninger

ICRC 2025

TU Dortmund, AG Kröninger
17.07.2025

The need for MC generator tuning

Hadronization cannot be calculated from first principle
→ Need for **hadronic models (MC generators)**

Hadronization has **large impact** on air shower features

Tuning **essential** to achieve **high-quality** simulations

Coherent Tunes with Different Experiments

MC generator tunes are problem-specific
→ Tunes usually based on data sets from intended applications

Idea: explore possibility of tuning using
data from various experiments simultaneously
→ Investigate interplay between different datasets

Long run goal: improve description of air showers by including data from collider and fixed-target experiments in tune

Tuning a MC Generator

MC generator: **PYTHIA** – widely used in high-energy physics
arXiv:2203.11601   and general-purpose generator

Adjust **free parameters** to achieve a **good** description of data

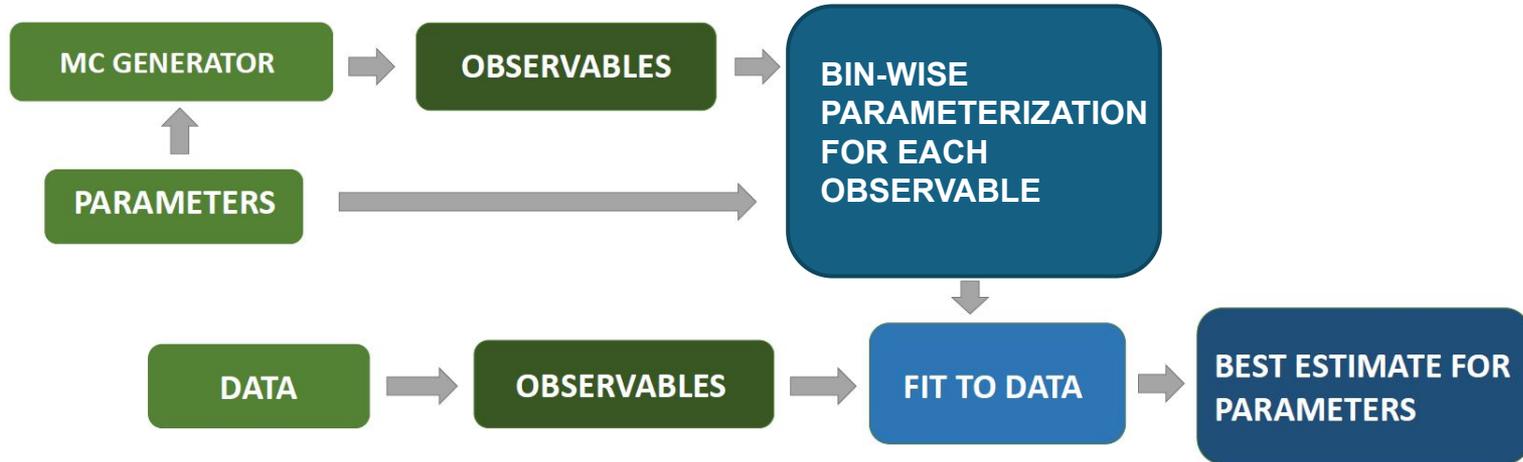Manual or brute-force tuning difficult due to **high computing cost**

Need for systematic tuning workflow
→ **Bayesian Analysis** MC tuning

Parameter based generator tuning

Optimize free parameters of MC generator using experimental data and Bayesian Methods

La Cagnina; JINST 18 (2023) P10033

Choice of Parameters

QCD differential cross sections

→ Each additional parameter adds a dimension to the tune
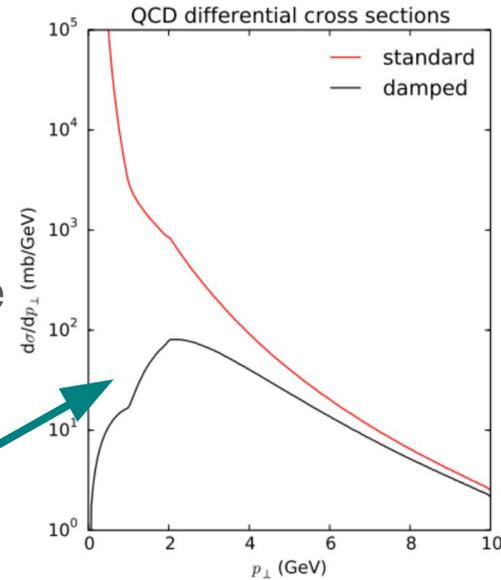 → Focus on three parameters for now

MultipartonInteractions:pT0Ref - reference transverse
 momentum cutoff for multiple parton interactions
→ no impact on LEP observables

StringZ:aLund - parameter a
in the Lund fragmentation model

Lund fragmentation function

$$f(z) = \left(\frac{1}{z}\right)(1-z)^{a} \times \exp\left(-\frac{bm_T^2}{z}\right)$$

StringPT:sigma - spread of transverse momentum of hadrons during string fragmentation

Choice of Observables

→ Focus on few **selected observables** which are meaningful and show high sensitivity to parameter variations in Pythia

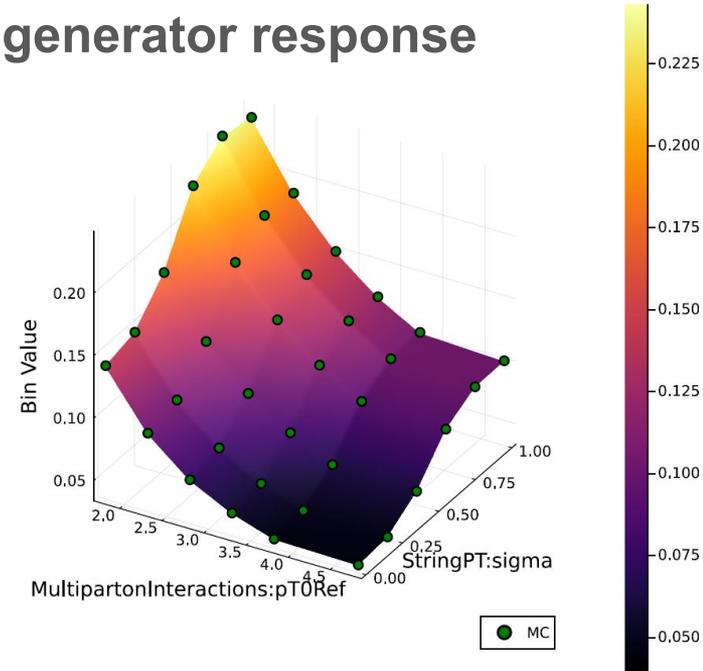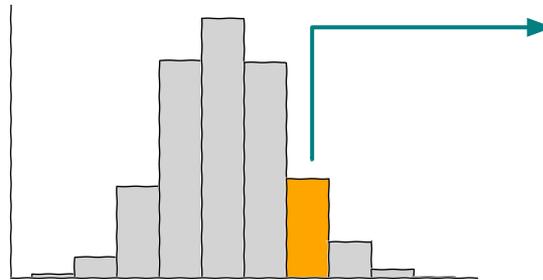→ Choose **Multiplicity** & **Momentum of charged particles** as **observables**

Experiments	RIVET plugin	Observables
LEP	L3_2004_I652683	Multiplicity
		Momentum
	ALEPH_1996_I428072	Momentum
LHC	ATLAS_2010_I882098	Multiplicity
		Transverse Momentum
fixed-target	EHS_1988_I265504	Transverse Momentum

Parameterization

The **bin-wise parameterization** needs to describe **generator response sufficiently** to ensure successful tune

Parameter dependent MC generator response often **quite complex**

→ Not approx. by **low dimensional** function
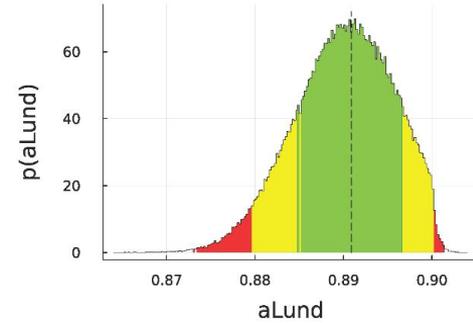→ Choose **linear interpolation between neighboring grid points**
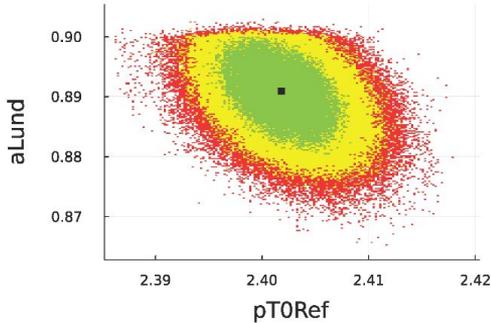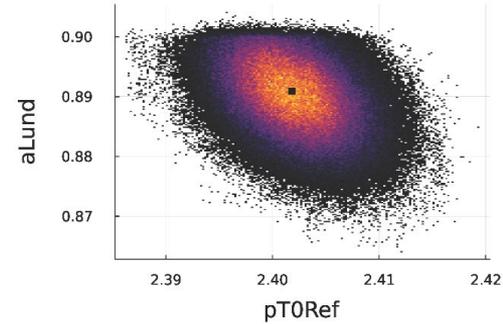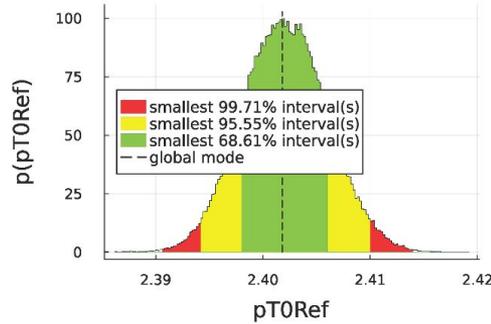
Tuning Step

Use parameterization and data to get **Likelihood**

Choose **uniform distributions** in the **allowed parameter ranges** as **prior**

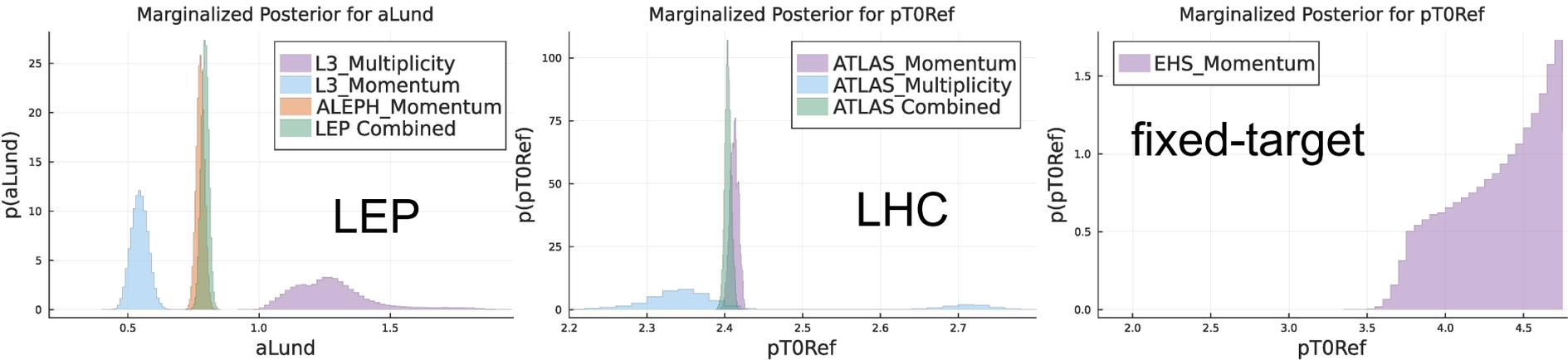Apply **Markov-Chain-Monte-Carlo** to **sample** the **posterior distribution**

The **global mode** is used as **tuned parameter values**

Sample posterior for uncertainty propagation to observable

Marginalized 1D Posteriors

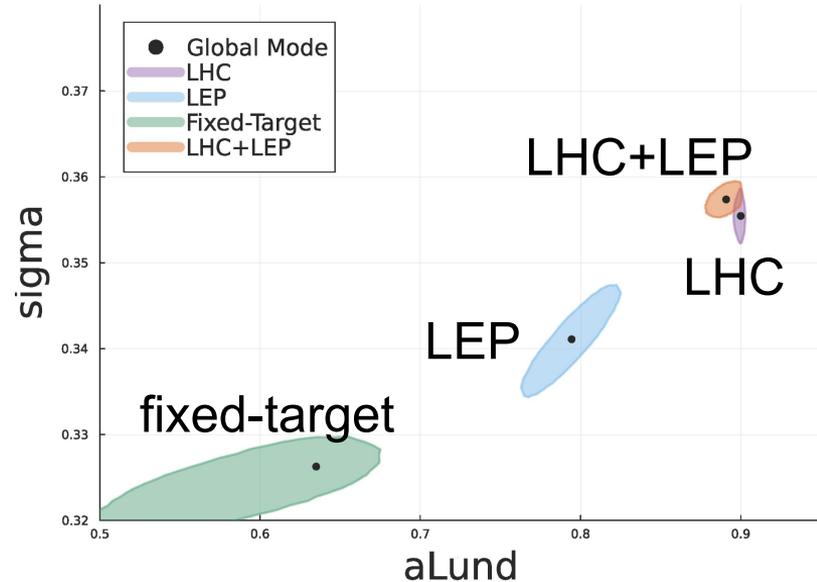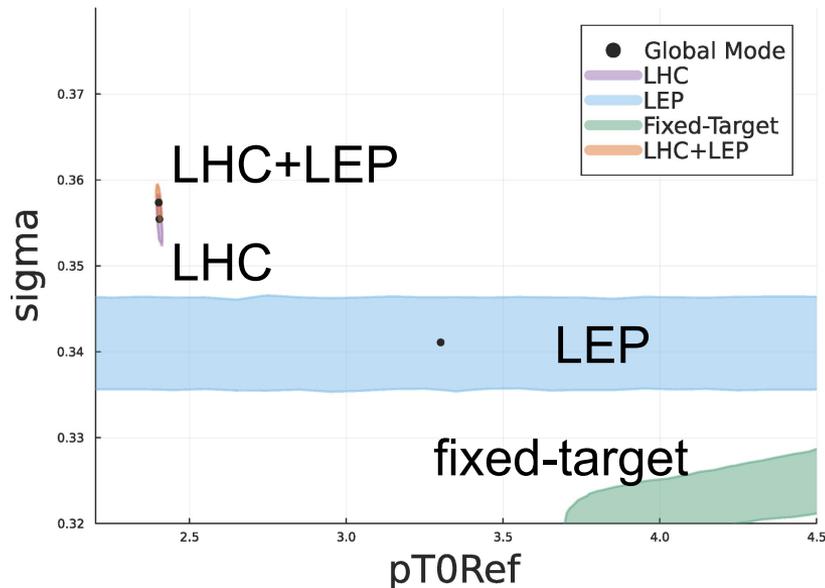**Fitting observables individually** result in **different** posterior distributions
Examples:

Observation: Fit to **fixed-target** data tends to **large pT0Ref** values

Marginalized 2D Posteriors

Due to pT0Ref, fixed-target is hard to combine with LEP and LHC
→ For now: try **two different tunes**, **LEP+LHC** and only **fixed-target**
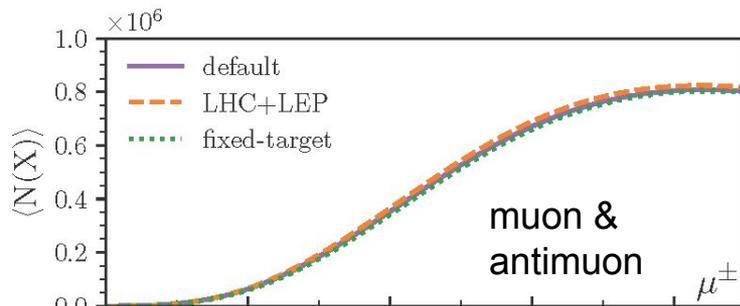Examples:
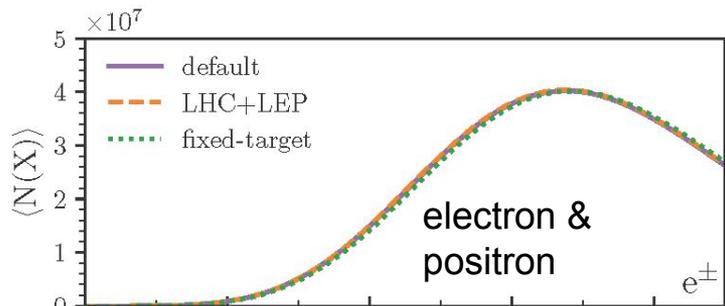
Tuned Parameter Values

	pT0Ref	aLund	sigma
chosen min.	1.78	0.2	0.0
default	2.28	0.68	0.335
chosen max.	4.78	2.0	1.0
LHC+LEP	2.404 ± 0.004	0.890 ± 0.005	0.357 ± 0.001
fixed-target	4.780 ± 0.574	0.640 ± 0.090	0.326 ± 0.004

pT0Ref was limited to a maximum of 4.78 for the tune

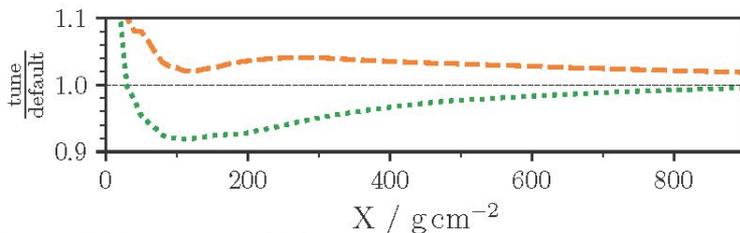Tunes are only close in StringPT:sigma
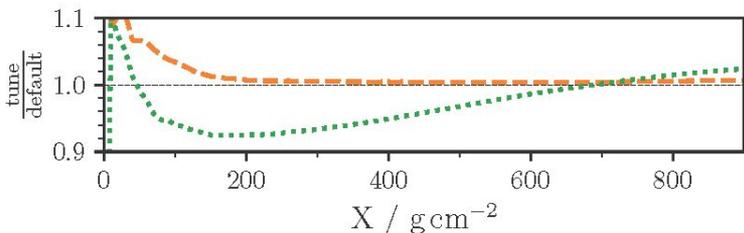
Impact on CORSIKA 8 Air Shower

Simulate **1000 proton** air showers at **1e17 eV**, zenith angle: **30°**
→ For LHC+LEP and Fixed-Target tune and default Pythia 8
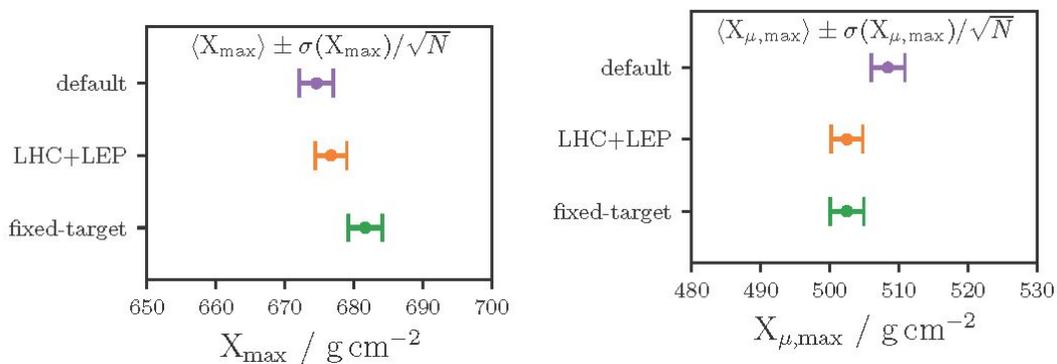**Average longitudinal** shower profile:

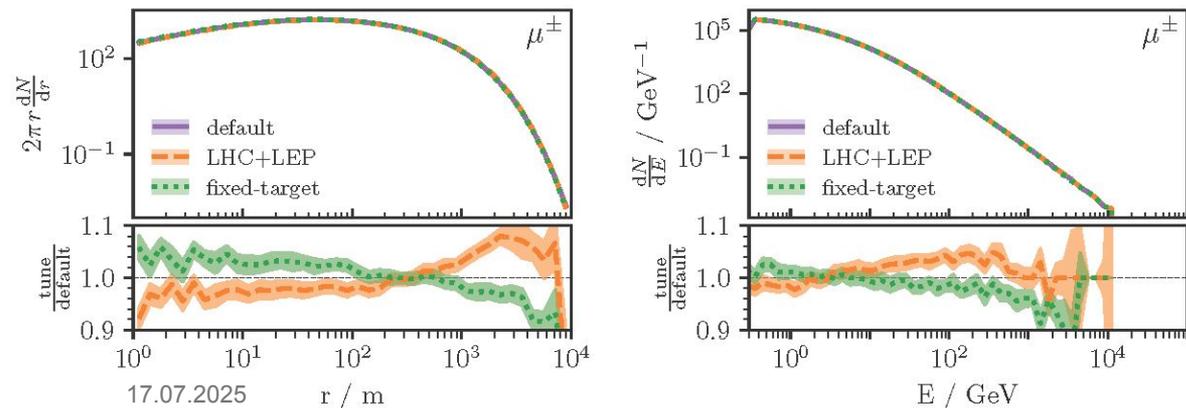LHC+LEP:

close to default for electron/positron

fixed-target:

fewer leptons in the beginning

Depth, Lateral and Energy

Average **depth** of **shower** and **muonic shower maximum**

Median **lateral distributions** & median **energy spectra** of **muons** at **Auger height** → **5-10% difference in both distributions**
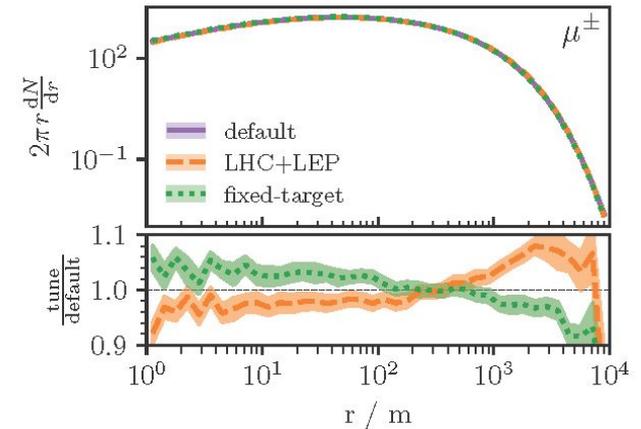
Conclusion

2D Marginalized Posterior: pT0Ref and sigma
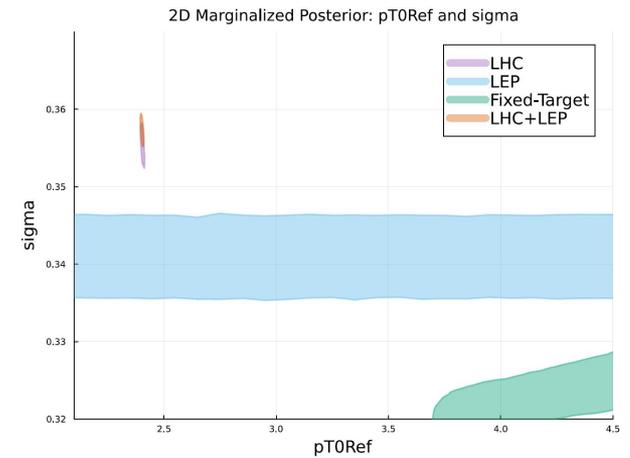
We investigated the interplay of collider and fixed-target data in the context of tuning
→ Hard to combine with chosen observables/parameters

Our tunes affect CORSIKA 8 simulations slightly, changing some observables by up to 10%

⇒ Studies with additional observables and parameters are needed, but easily doable with the developed tuning framework
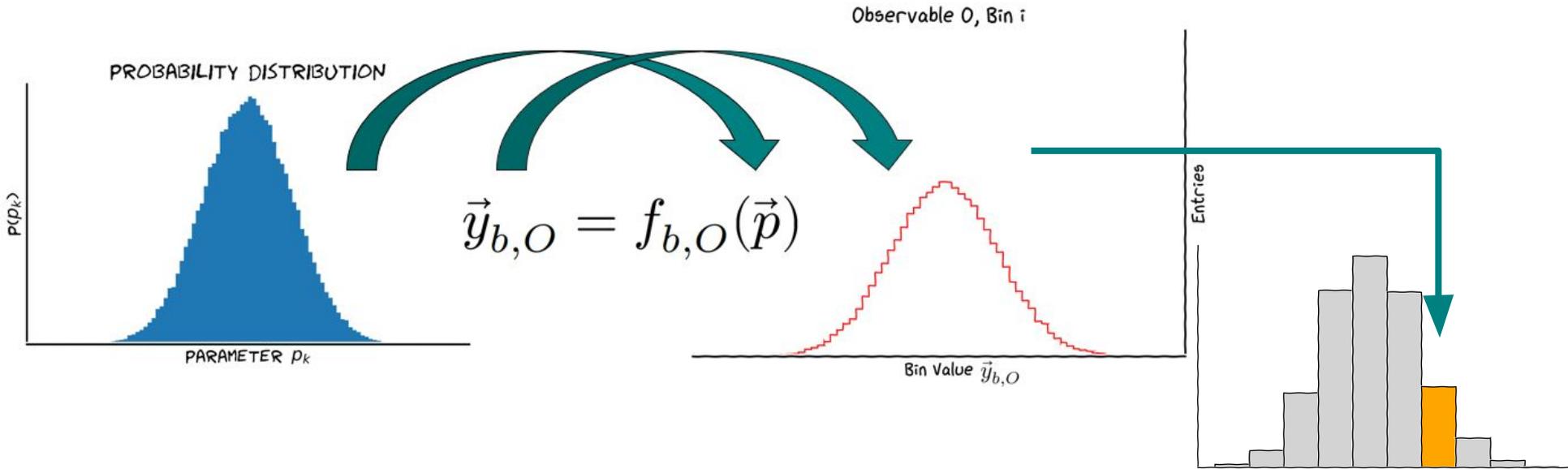

We are not presenting a global tune - but a first step to do it!
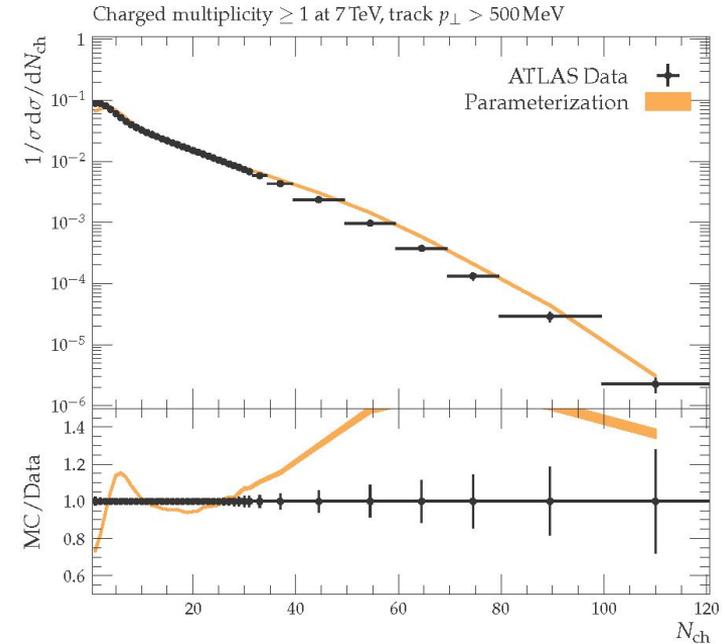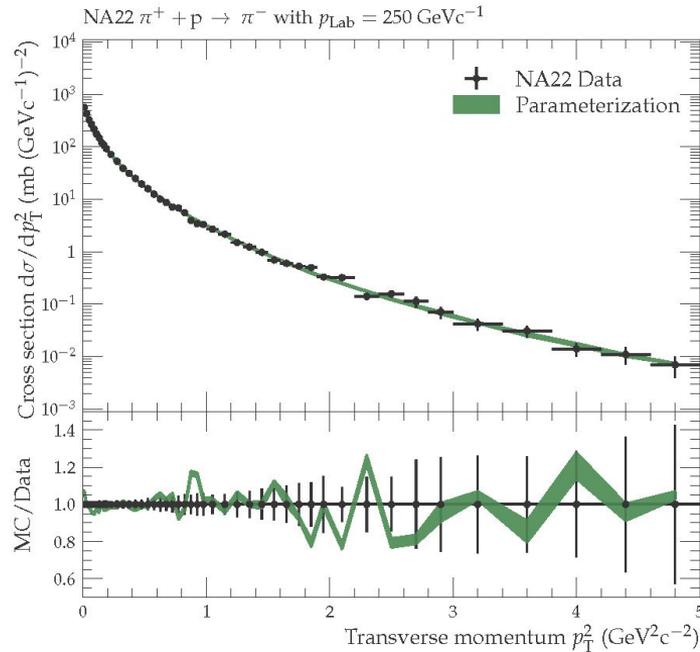
BackUp

Uncertainty Propagation

Sample posterior of the parameter space and use parametrization to get uncertainties on observable

$$\vec{y}_{b,O} = f_{b,O}(\vec{p})$$

Uncertainty Propagation

Green band width - uncertainty of tune for parameterization
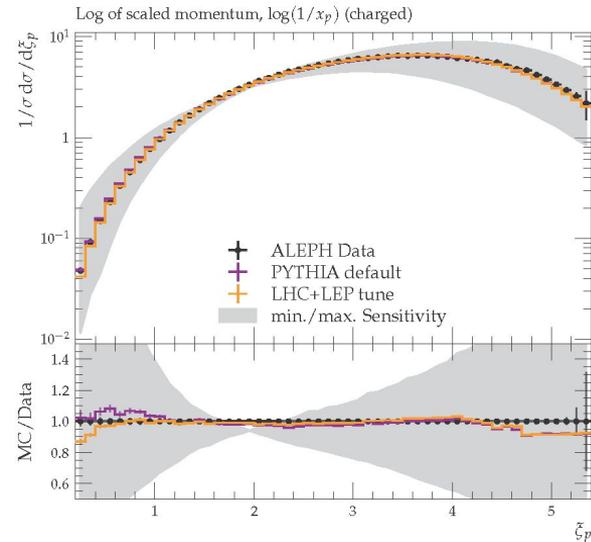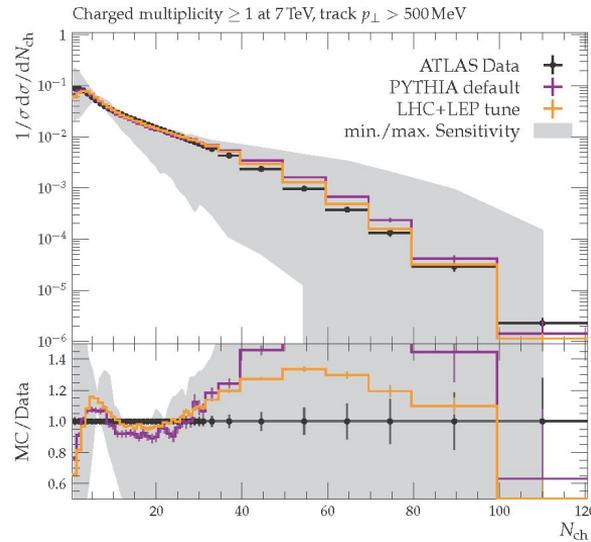
Examples:

Tuning Result on Observables

In most cases, the corresponding tune **improves** the **MC simulation**

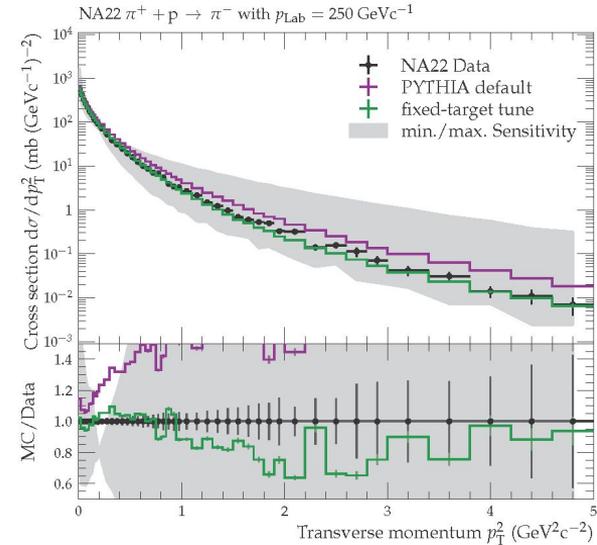$\rightarrow$ lower reduced χ^2 - value, compared to default Pythia 8 value
Examples:

Likelihood

Likelihood is built with the EFTfitter.jl Julia package

$$\ln L(\vec{D}|\vec{\lambda}) = -\frac{1}{2}[\vec{D} - \vec{f}(\vec{\lambda})]^T M^{-1}[\vec{D} - \vec{f}(\vec{\lambda})]$$