UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

STAYING ON TOP OF LIKELIHOOD ANALYSES

Using pyhf for global SMEFT analyses with SFitter

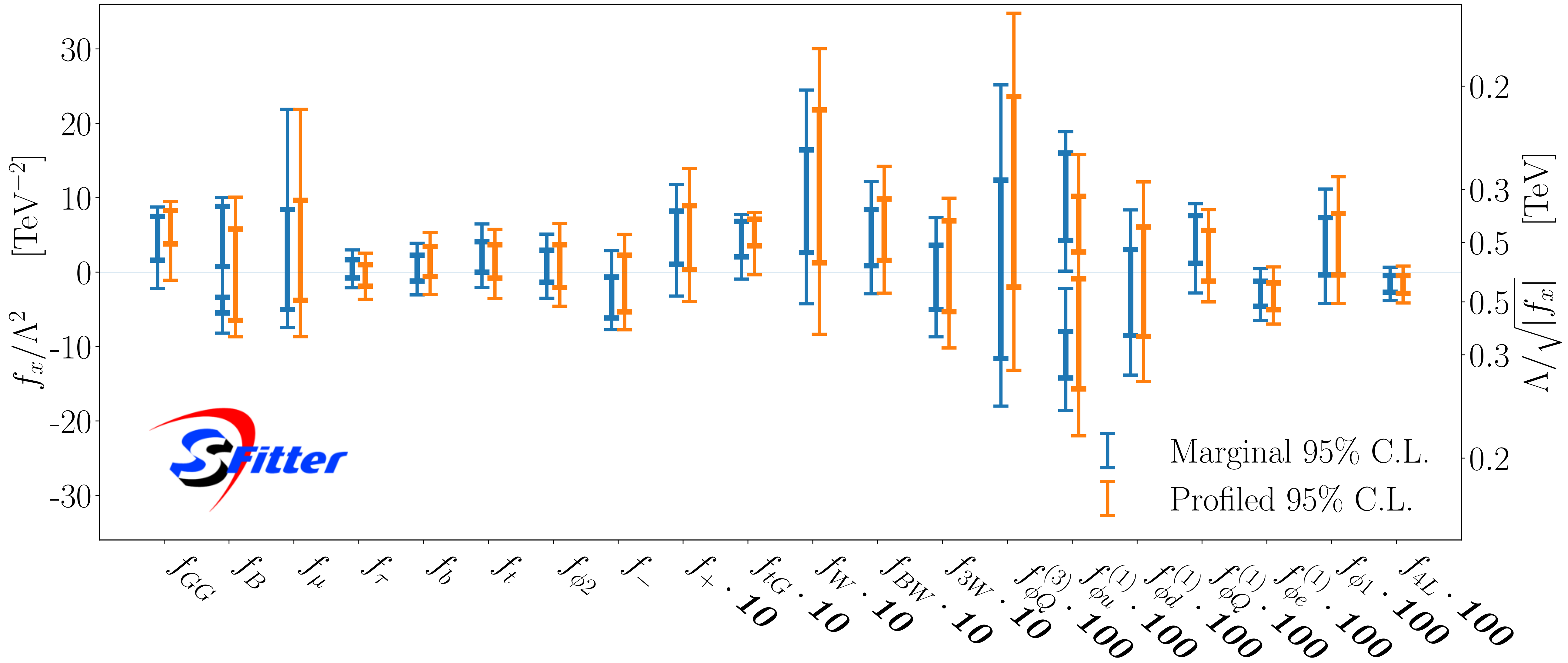
Nikita Schmal

Collaborators: Nina Elmer, Emma Geoffray, Michel Luchmann, Maeve Madigan, Tilman Plehn

What’s the purpose of this talk?

- **Problem:** Large number of observations cannot be explained by the SM alone
- **What we do:** Global SMEFT analyses using **SFitter**
- **Goal:** Put constraints on physics beyond the Standard Model

arXiv:2208.08454 [hep-ph]



Outline

- **Intro:** Standard Model Effective Field Theory
- **Part I:** Statistical analysis using SFitter
- **Part II:** SFitter analyses with pyhf
- **Conclusion**

Standard Model Effective Field Theory

SMEFT

- Well established model agnostic approach in searches for BSM physics

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_{d=5}^{n} \frac{C_i^{(d)}}{\Lambda^{d-4}} O_i^{(d)}$$

- Up to **quadratic order** SMEFT contributions included i.e.

$$\sigma = \sigma_{SM} + \frac{c_6}{\Lambda^2}\sigma_6 + \frac{c_6^2}{\Lambda^4}\sigma_{6\times 6} + \frac{c_8}{\Lambda^4}\sigma_8 + \mathcal{O}(\Lambda^5)$$

Standard Model Effective Field Theory

SMEFT

- Well established model agnostic approach in searches for BSM physics

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_{d=5}^{n} \frac{C_i^{(d)}}{\Lambda^{d-4}} O_i^{(d)}$$

- Up to **quadratic order** SMEFT contributions included i.e.

$$\sigma = \sigma_{SM} + \frac{c_6}{\Lambda^2}\sigma_6 + \frac{c_6^2}{\Lambda^4}\sigma_{6\times 6} + \frac{c_8}{\Lambda^4}\sigma_8 + \mathcal{O}(\Lambda^5)$$

- Restrict ourselves to operators of **dimension 6**

Standard Model Effective Field Theory

Model and dataset

arXiv:1910.03606 [hep-ph]

- Restrict ourselves to the Top sector
  - Include $t\bar{t}$, $t\bar{t}Z$, $t\bar{t}W$ and single top data
    - Total **~116 datapoints**
- Impose $U(2)_q \times U(2)_u \times U(2)_d$ symmetry
  - Consider a total of **22 Operators**

Operator	Definition	Operator	Definition
$O_{Qq}^{3,8}$	$(\bar{Q}\gamma_\mu T^A \tau^I Q)(\bar{q}_i \gamma^\mu T^A \tau^I q_i)$	$O_{Qq}^{3,1}$	$(\bar{Q}\gamma_\mu \tau^I Q)(\bar{q}_i \gamma^\mu \tau^I q_i)$
$O_{Qq}^{1,8}$	$(\bar{Q}\gamma_\mu T^A Q)(\bar{q}_i \gamma^\mu T^A q_i)$	$O_{Qq}^{1,1}$	$(\bar{Q}\gamma_\mu Q)(\bar{q}_i \gamma^\mu q_i)$
O_{tu}^{8}	$(\bar{t}\gamma_\mu T^A t)(\bar{u}_i \gamma^\mu T^A u_i)$	O_{tu}^{1}	$(\bar{t}\gamma_\mu t)(\bar{u}_i \gamma^\mu u_i)$
O_{td}^{8}	$(\bar{t}\gamma_\mu T^A t)(\bar{d}_i \gamma^\mu T^A d_i)$	O_{td}^{1}	$(\bar{t}\gamma_\mu t)(\bar{d}_i \gamma^\mu d_i)$
O_{Qu}^{8}	$(\bar{Q}\gamma^\mu T^A Q)(\bar{u}_i \gamma_\mu T^A u_i)$	O_{Qu}^{1}	$(\bar{Q}\gamma^\mu Q)(\bar{u}_i \gamma_\mu u_i)$
O_{Qd}^{8}	$(\bar{Q}\gamma^\mu T^A Q)(\bar{d}_i \gamma_\mu T^A d_i)$	O_{Qd}^{1}	$(\bar{Q}\gamma^\mu Q)(\bar{d}_i \gamma_\mu d_i)$
O_{tq}^{8}	$(\bar{q}_i \gamma^\mu T^A q_i)(\bar{t}\gamma_\mu T^A t)$	O_{tq}^{1}	$(\bar{q}_i \gamma^\mu q_i)(\bar{t}\gamma_\mu t)$
$O_{\phi Q}^{3}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu \phi)(\bar{Q}\gamma^\mu \tau^I Q)$	$O_{\phi Q}^{1}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu \phi)(\bar{Q}\gamma^\mu Q)$
$O_{\phi t}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu \phi)(\bar{t}\gamma^\mu t)$	$O_{\phi tb}$	$(\tilde{\phi}^\dagger i D_\mu \phi)(\bar{t}\gamma^\mu b)$
O_{tB}	$(\bar{Q}\sigma^{\mu\nu} t)\tilde{\phi} B_{\mu\nu}$	O_{tW}	$(\bar{Q}\sigma^{\mu\nu} t)\tau^I \tilde{\phi} W_{\mu\nu}^I$
O_{bW}	$(\bar{Q}\sigma^{\mu\nu} b)\tau^I \phi W_{\mu\nu}^I$	O_{tG}	$(\bar{Q}\sigma_{\mu\nu} T^A t)\tilde{\phi} G_{\mu\nu}^A$

PART I

Statistical analysis with SFitter

What is our tool of choice?

SFitter

- Used for various global SMEFT analyses (Higgs, Di-Boson, EWPO, **Top**)
- Comprehensive treatment of **uncertainties**
- **Fully correlated** systematic uncertainties within experiments
- Allows for both **profiling and marginalization** methods
- Mapping of likelihood using **MCMC**

- **Goal of this part:** Explain what we mean with all of this

What is SFitter?

The exclusive likelihood

- Likelihood for a **single measurements** modelled as

$$\mathcal{L}_{excl} = \text{Pois}(d|p(\alpha_n, \theta_i, b))\text{Pois}(b_{CR}|b\, k)\prod_i \mathcal{C}(\theta_i, \sigma_i)$$

- SMEFT contributions are incorporated into model parameters α_n
- Uncertainties included via nuisance parameters (NP) θ_i
- Constraint term $\mathcal{C}(\theta_i, \sigma_i)$ depends on uncertainty considered

What is SFitter?

Uncertainty constraints

$\mathcal{C}(\theta_i, \sigma_i)$

- Systematic unc. $\mathcal{N}(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
- Statistical unc. $\mathrm{Pois}(n|\nu) = \frac{\nu^n e^{-\nu}}{n!}, \quad \nu > 0$
- Theory unc. $\mathcal{F}(x|\mu,\sigma) = \frac{1}{2\sigma}\Theta\left[x-(\mu-\sigma)\right]\Theta\left[(\mu+\sigma)-x\right]$

- Choice of constraint is motivated by physical intuition
  - **However:** They are a choice and could technically be chosen differently

What is SFitter?

Generalization to multiple measurements

- Global analyses study numerous different processes

$$\mathcal{L}_{\text{excl,full}} = \prod_c \text{Pois}(d_c|p_c)\text{Pois}(b_{CR_c}|b_c\, k_c) \prod_i \mathcal{C}(\theta_{i,c}, \sigma_{i,c})$$

- Take into consideration correlations between these measurements

$$\mathcal{N}(\theta_{syst,i}|0, \sigma_i) \longrightarrow \mathcal{N}(\vec{\theta}_{syst,i}|\vec{0}, \Sigma_i)$$

- **Assumption:** Systematics are **fully correlated** between measurements

What is SFitter?

Systematic uncertainties

- Each category of systematic is fully correlated within CMS and ATLAS
- Luminosity correlated between both experiments

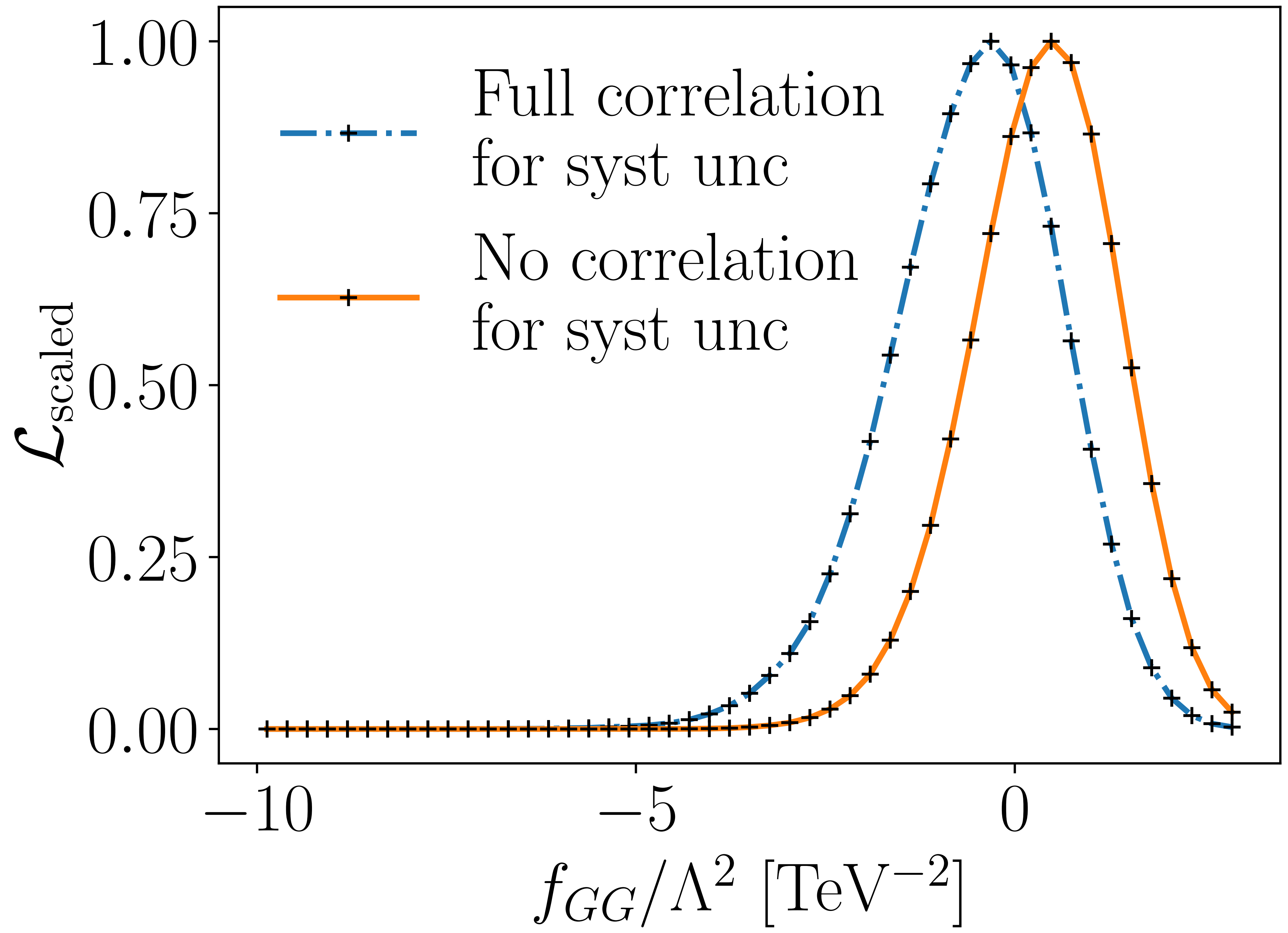
Systematic uncertainties
Beam
Background (Separate for each channel)
ETmis
Jets
Leptons
LightTagging
Luminosity
Pileup
Trigger
Tune
bTagging
partonShower
tTagging
tauTagging

What is SFitter?

Systematic uncertainties

arXiv:2208.08454 [hep-ph]

- ➤ Each category of systematic is fully correlated within CMS and ATLAS
- ➤ Luminosity correlated between both experiments
- ➤ Clear shift in the likelihoods **due to correlations** between systematics



What is SFitter?

To profile or to marginalize

- Common exclusive likelihood constructed

$$\mathcal{L}_{excl} = \text{Pois}(d|p(\alpha_n, \theta_i, b))\text{Pois}(b_{CR}|b\,k)\prod_i \mathcal{C}(\theta_i, \sigma_i)$$

- The NPs θ_i are not physically interesting

What is SFitter?

To profile or to marginalize

- Common exclusive likelihood constructed

$$\mathcal{L}_{excl} = \mathrm{Pois}(d|p(\alpha_n, \theta_i, b))\mathrm{Pois}(b_{CR}|b\,k) \prod_i \mathcal{C}(\theta_i, \sigma_i)$$

- The NPs θ_i are not physically interesting
  - **Decision:** How do we handle the NPs?

Profiling:

$$\mathcal{L}_{\mathrm{prof}}(\alpha) = \max_\theta \mathcal{L}_{\mathrm{excl}}(\alpha, \theta)$$

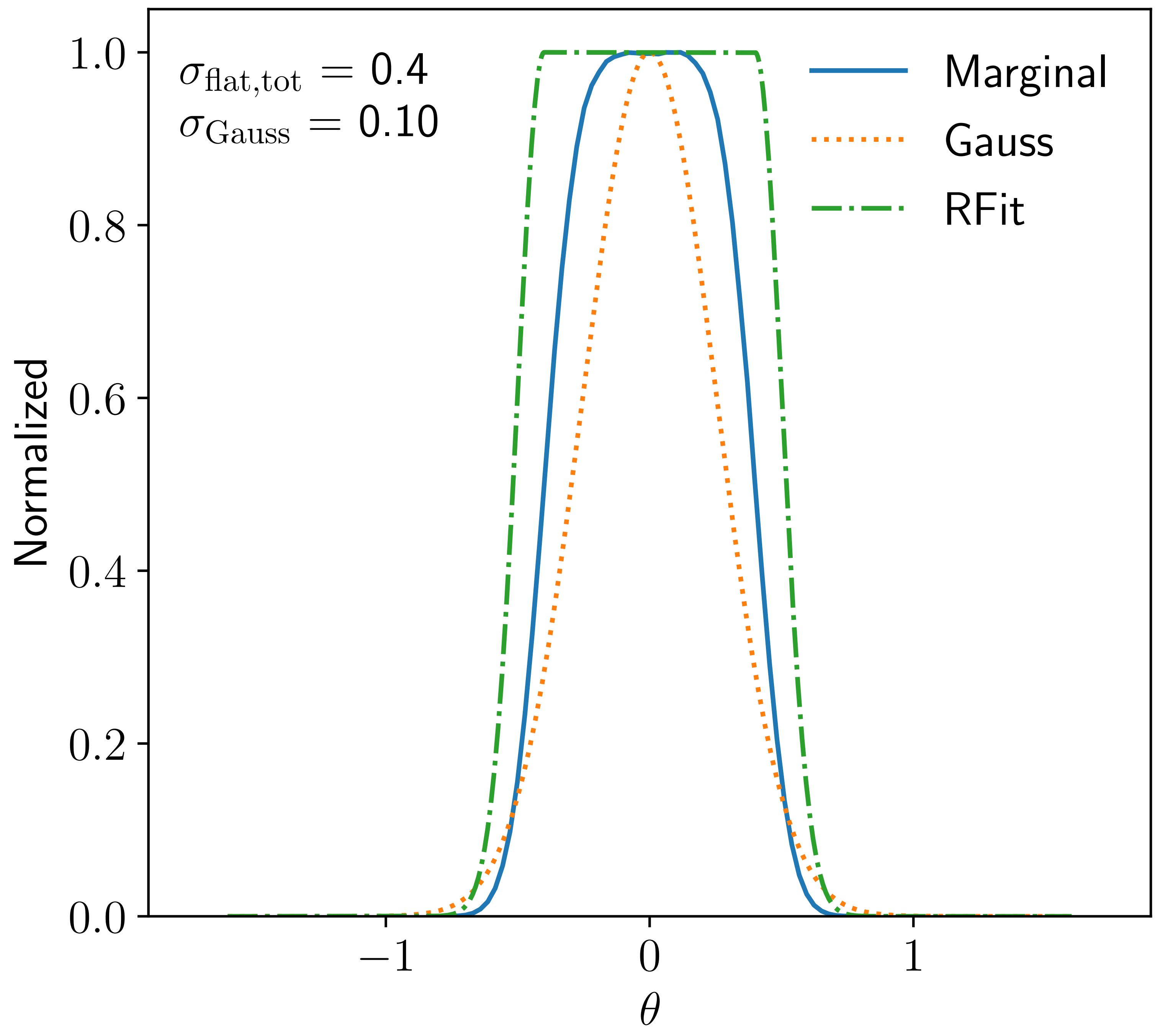
Marginalization:

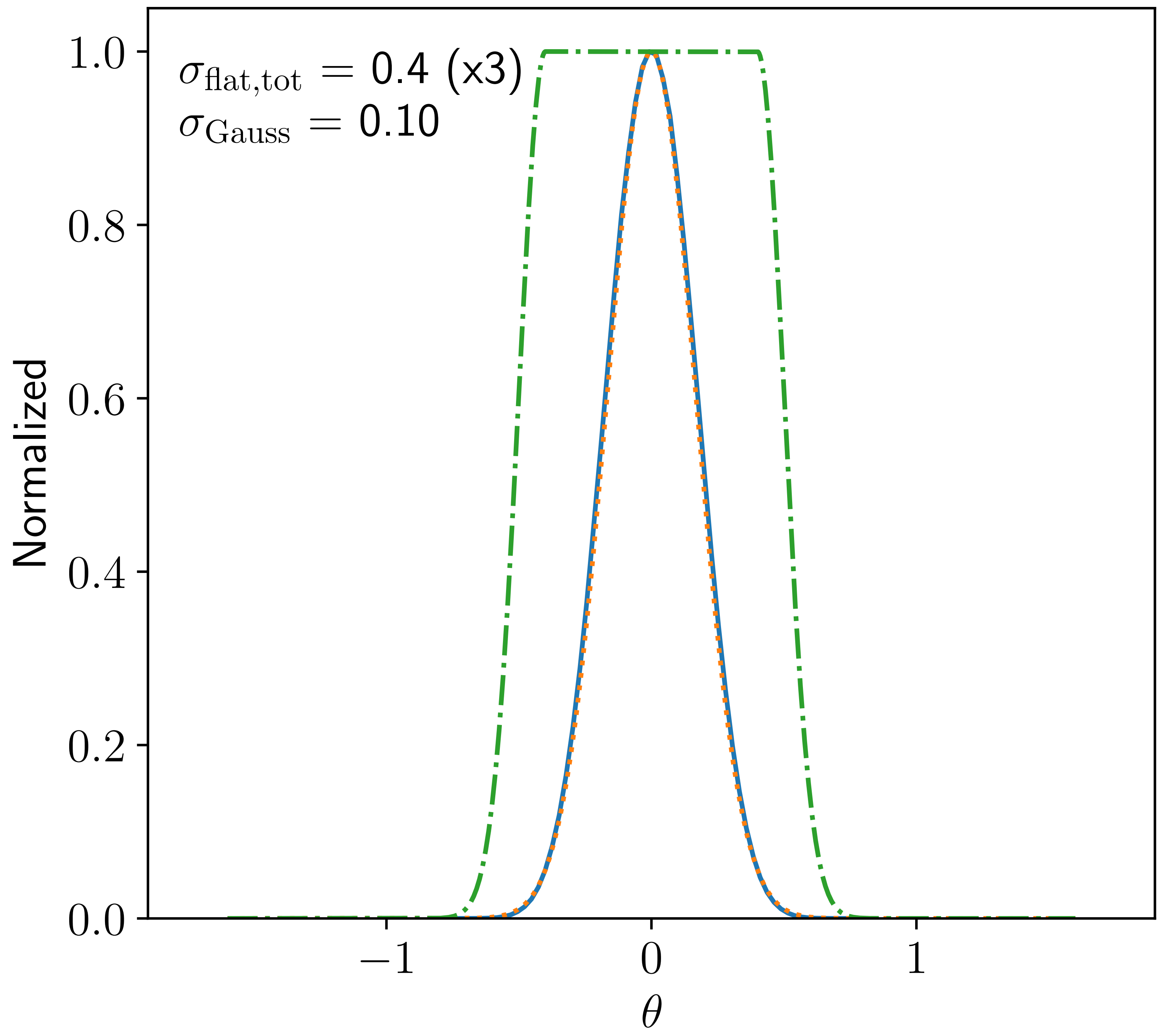
$$\mathcal{L}_{marg}(\alpha) = \int d\theta \mathcal{L}_{excl}(\alpha, \theta)$$

What is SFitter?

To profile or to marginalize

arXiv:2208.08454 [hep-ph]

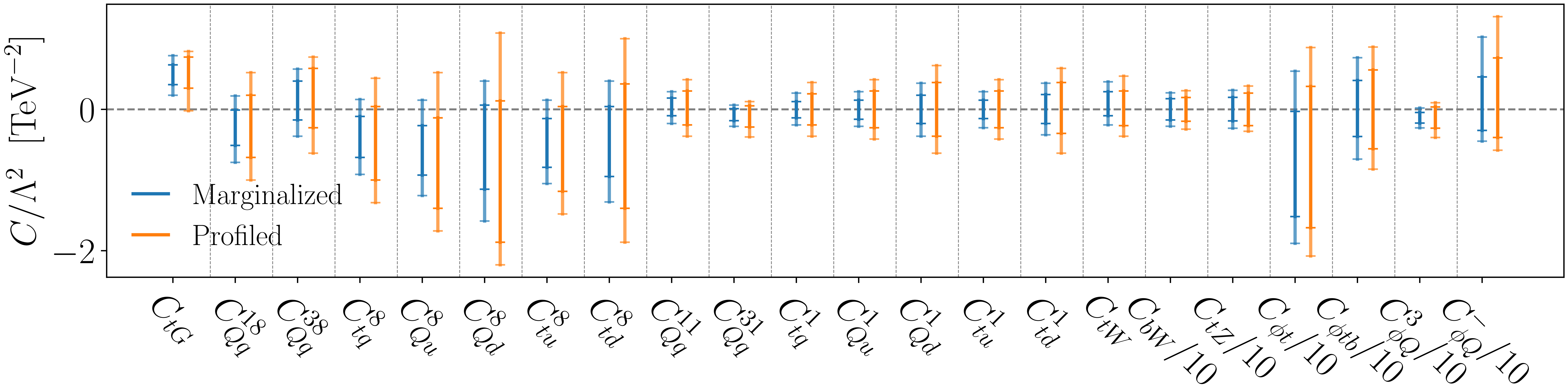




- Comparison for the product of Gaussian and uniform distributions
  - Marginalization over multiple flat unc. gives **Gaussian** results

Some Results

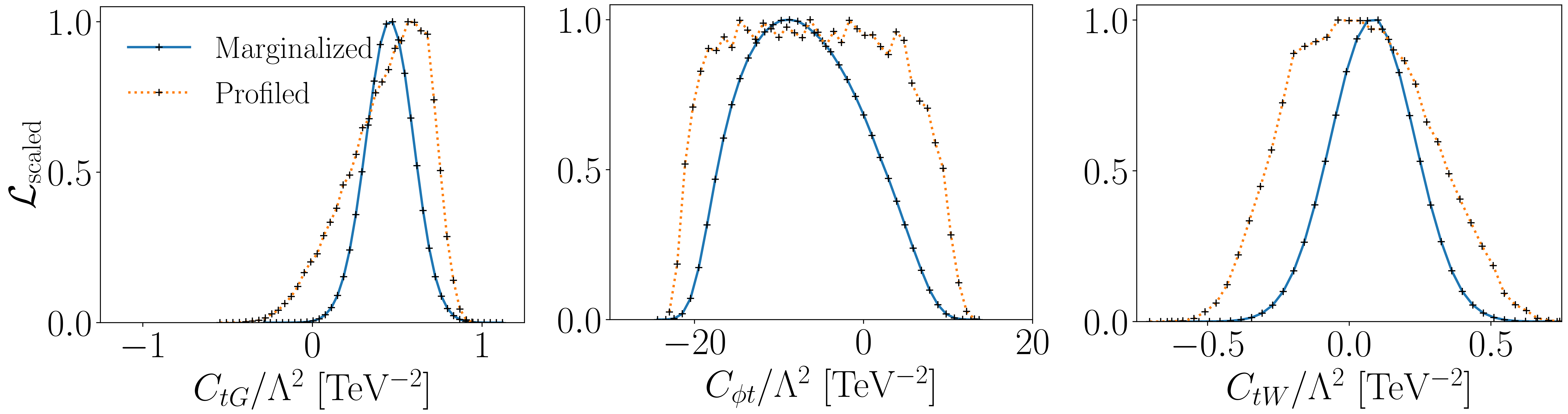
Profiling vs Marginalization



➤ Stronger constraints for marginalized likelihood as a result of **large theory uncertainties**

Some Results

Profiling vs Marginalization



- ➤ Expected behaviour due to marginalization of **flat theory uncertainties**

What is SFitter?

Takeaway

- **Uncertainty treatment** essential to our SFitter analysis
  - Implementation of theory, statistical and systematic uncertainties
  - Furthermore: **Correlated systematics** of the same type

- Theory prediction and uncertainties done by us
  - However: The systematics have to be provided **by experiment**
  - How is this data provided and how can we use it?

PART II

SFitter analyses with pyhf

∂ifferentiable
ℒikelihoods

Published Likelihoods

Quick overview

- Likelihoods published in the **HistFactory** format

$$\mathcal{L}(n_{cb}, a_\chi | \eta, \chi) = \prod_{c \in \text{channels}} \prod_{b \in \text{bins}} \text{Pois}(n_{cb} | \nu_{cb}(\eta, \chi)) \prod_{\chi \in \vec{\chi}} c_\chi(a_\chi | \chi)$$

- Provides effect of large number of **individual NPs**
- Analysed using dedicated python libraries such as **pyhf** and **cabinetry**
  - **Question:** How to make use of this in SFitter analyses?

Likelihoods published by ATLAS

arXiv:2006.13076 [hep-ex]

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Phys. Lett. B 810 (2020) 135797
DOI: 10.1016/j.physletb.2020.135797

CERN-EP-2020-096
10th November 2020

Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s}$ = 13 TeV with the ATLAS experiment

The ATLAS Collaboration

arXiv:2103.12603 [hep-ex]

Eur. Phys. J. C (2021) 81:737
https://doi.org/10.1140/epjc/s10052-021-09439-4

THE EUROPEAN PHYSICAL JOURNAL C

Check for updates

Regular Article - Experimental Physics

Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a Z boson at $\sqrt{s}$ = 13 TeV with the ATLAS detector

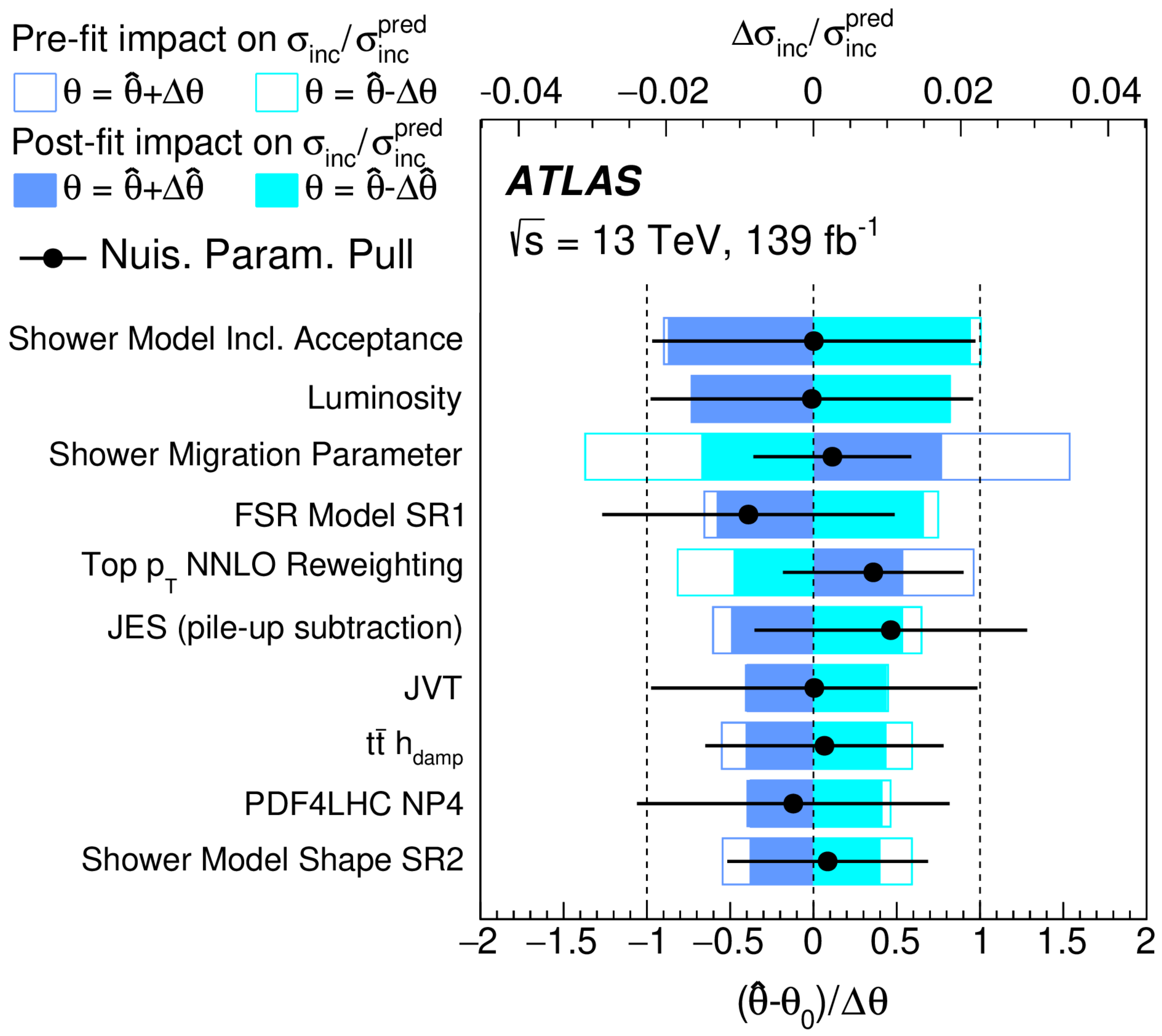
ATLAS Collaboration*
CERN, 1211 Geneva 23, Switzerland

Received: 24 March 2021 / Accepted: 10 July 2021 / Published online: 16 August 2021

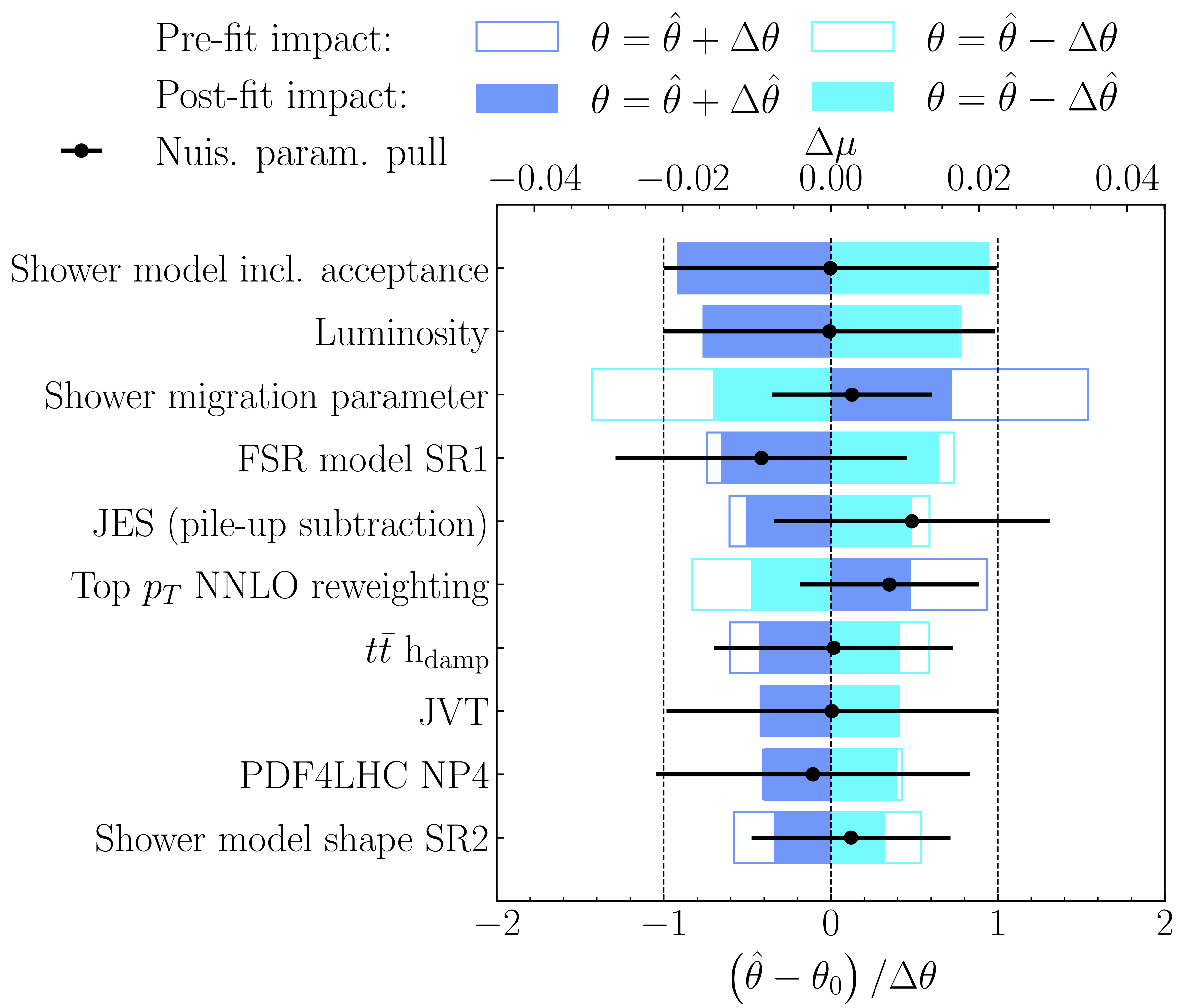
➤ Full likelihoods **publicly available** on HEPData

Published Likelihoods

Quick overview (Reproduction)



From [arXiv:2006.13076](https://arxiv.org/abs/2006.13076) [hep-ex]



cabinetry

pyhf ∂ifferentiable ℒikelihoods

Published Likelihoods

Uncertainties

- **Previously:** Uncertainties taken as given in the paper
- **Now:** Uncertainties extracted from profiling fit via pyhf
  - Implemented into SFitter using the constraints terms $\mathcal{C}(\theta_i, \sigma_i)$
- **Problem:** Difficult to automate due to inconsistent naming conventions

Uncertainty	Reproduced $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]	Paper $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]
ttZ parton shower	3.1	3.1
tWZ modeling	2.9	2.9
b-tagging	2.9	2.9
WZ/ZZ + jets modeling	2.7	2.8
tZq modeling	2.6	2.6
Lepton	2.3	2.3
Luminosity	2.2	2.2
Jets + E_T^{miss}	2.1	2.1
Fake leptons	2.1	2.1
$t\bar{t}Z$ ISR	1.7	1.6
$t\bar{t}Z \mu_F$ and μ_r scales	0.9	0.9
Other backgrounds	0.8	0.7
Pile-up	0.7	0.7
$t\bar{t}Z$ PDF	0.2	0.2
Stat	5.2	5.2

Published Likelihoods

Uncertainties

Uncertainty	Reproduced $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]	Paper $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]
ttZ parton shower	3.1	3.1
tWZ modeling	2.9	2.9
b-tagging	2.9	2.9
WZ/ZZ + jets modeling	2.7	2.8
tZq modeling	2.6	2.6
Lepton	2.3	2.3
Luminosity	2.2	2.2
Jets + E_T^{miss}	2.1	2.1
Fake leptons	2.1	2.1
$t\bar{t}Z$ ISR	1.7	1.6
$t\bar{t}Z\mu_F$ and μ_r scales	0.9	0.9
Other backgrounds	0.8	0.7
Pile-up	0.7	0.7
$t\bar{t}Z$ PDF	0.2	0.2
Stat	5.2	5.2

Assign →

Systematic uncertainties
Beam
Background (Separate for each channel)
ETmis
Jets
Leptons
LightTagging
Luminosity
Pileup
Trigger
Tune
bTagging
partonShower
tTagging
tauTagging

Published Likelihoods

Uncertainties

Uncertainty	Reproduced $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]	Paper $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]
ttZ parton shower	3.1	3.1
tWZ modeling	2.9	2.9
b-tagging	2.9	2.9
WZ/ZZ + jets modeling	2.7	2.8
tZq modeling	2.6	2.6
Lepton	2.3	2.3
Luminosity	2.2	2.2
Jets + E_T^{miss}	2.1	2.1
Fake leptons	2.1	2.1
$t\bar{t}Z$ ISR	1.7	1.6
$t\bar{t}Z \mu_F$ and μ_r scales	0.9	0.9
Other backgrounds	0.8	0.7
Pile-up	0.7	0.7
$t\bar{t}Z$ PDF	0.2	0.2
Stat	5.2	5.2

Assign

Systematic uncertainties
Beam
Background (Separate for each channel)
ETmis
Jets
Leptons
LightTagging
Luminosity
Pileup
Trigger
Tune
bTagging
partonShower
tTagging
tauTagging

Published Likelihoods

Uncertainties

Uncertainty	**Reproduced** $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]	**Paper** $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]
ttZ parton shower	3.1	3.1
tWZ modeling	2.9	2.9
b-tagging	2.9	2.9
WZ/ZZ + jets modeling	2.7	2.8
tZq modeling	2.6	2.6
Lepton	2.3	2.3
Luminosity	2.2	2.2
Jets + E_T^{miss}	2.1	2.1
Fake leptons	2.1	2.1
$t\bar{t}Z$ ISR	1.7	1.6
$t\bar{t}Z\mu_F$ and μ_r scales	0.9	0.9
Other backgrounds	0.8	0.7
Pile-up	0.7	0.7
$t\bar{t}Z$ PDF	0.2	0.2
Stat	5.2	5.2

Assign

Systematic uncertainties
Beam
Background (Separate for each channel)
ETmis
Jets
Leptons
LightTagging
Luminosity
Pileup
Trigger
Tune
bTagging
partonShower
tTagging
tauTagging

Published Likelihoods

Uncertainties

Uncertainty	Reproduced $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]	Paper $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]
ttZ parton shower	3.1	3.1
tWZ modeling	2.9	2.9
b-tagging	2.9	2.9
WZ/ZZ + jets modeling	2.7	2.8
tZq modeling	2.6	2.6
Lepton	2.3	2.3
Luminosity	2.2	2.2
Jets + E_T^{miss}	2.1	2.1
Fake leptons	2.1	2.1
$t\bar{t}Z$ ISR	1.7	1.6
$t\bar{t}Z \mu_F$ and μ_r scales	0.9	0.9
Other backgrounds	0.8	0.7
Pile-up	0.7	0.7
$t\bar{t}Z$ PDF	0.2	0.2
Stat	5.2	5.2

Assign

Systematic uncertainties
Beam
Background (Separate for each channel)
ETmis
Jets
Leptons
LightTagging
Luminosity
Pileup
Trigger
Tune
bTagging
partonShower
tTagging
tauTagging

?

➤ **Previously:** Possibly incompatible groups, how to correlate?

Published Likelihoods

Uncertainties

Uncertainty	Reproduced $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]	Paper $\frac{\Delta\sigma_{t\bar{t}Z}}{\sigma_{t\bar{t}Z}}$ [%]
ttZ parton shower	3.1	3.1
tWZ modeling	2.9	2.9
b-tagging	2.9	2.9
WZ/ZZ + jets modeling	2.7	2.8
tZq modeling	2.6	2.6
Lepton	2.3	2.3
Luminosity	2.2	2.2
Jets + E_T^{miss}	2.1	2.1
Fake leptons	2.1	2.1
$t\bar{t}Z$ ISR	1.7	1.6
$t\bar{t}Z \mu_F$ and μ_r scales	0.9	0.9
Other backgrounds	0.8	0.7
Pile-up	0.7	0.7
$t\bar{t}Z$ PDF	0.2	0.2
Stat	5.2	5.2

Assign →

Systematic uncertainties
Beam
Background (Separate for each channel)
ETmis
Jets
Leptons
LightTagging
Luminosity
Pileup
Trigger
Tune
bTagging
partonShower
tTagging
tauTagging

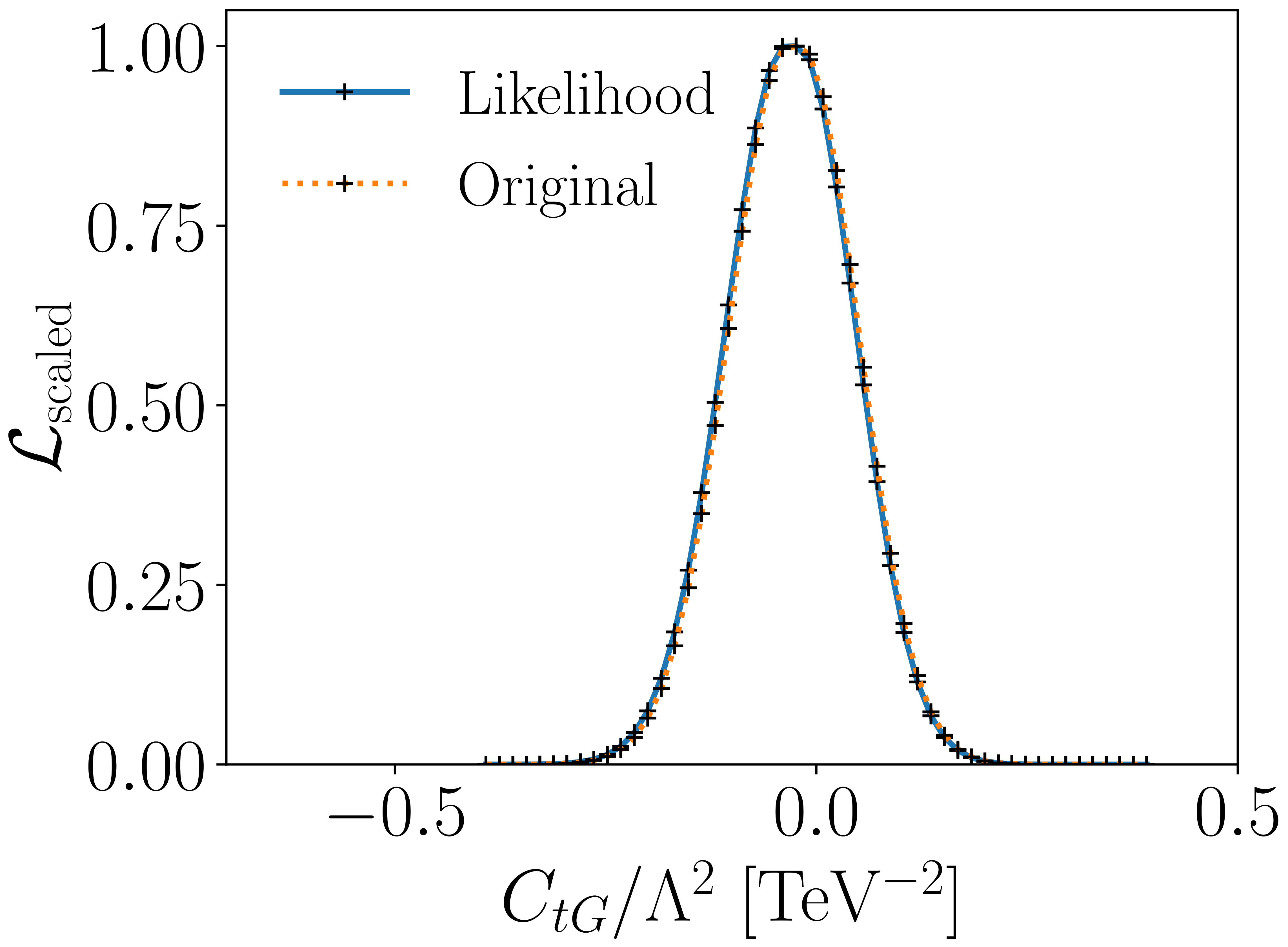
?

➤ **Now:** Simply separate the NPs in profile likelihood fit

Published Likelihoods

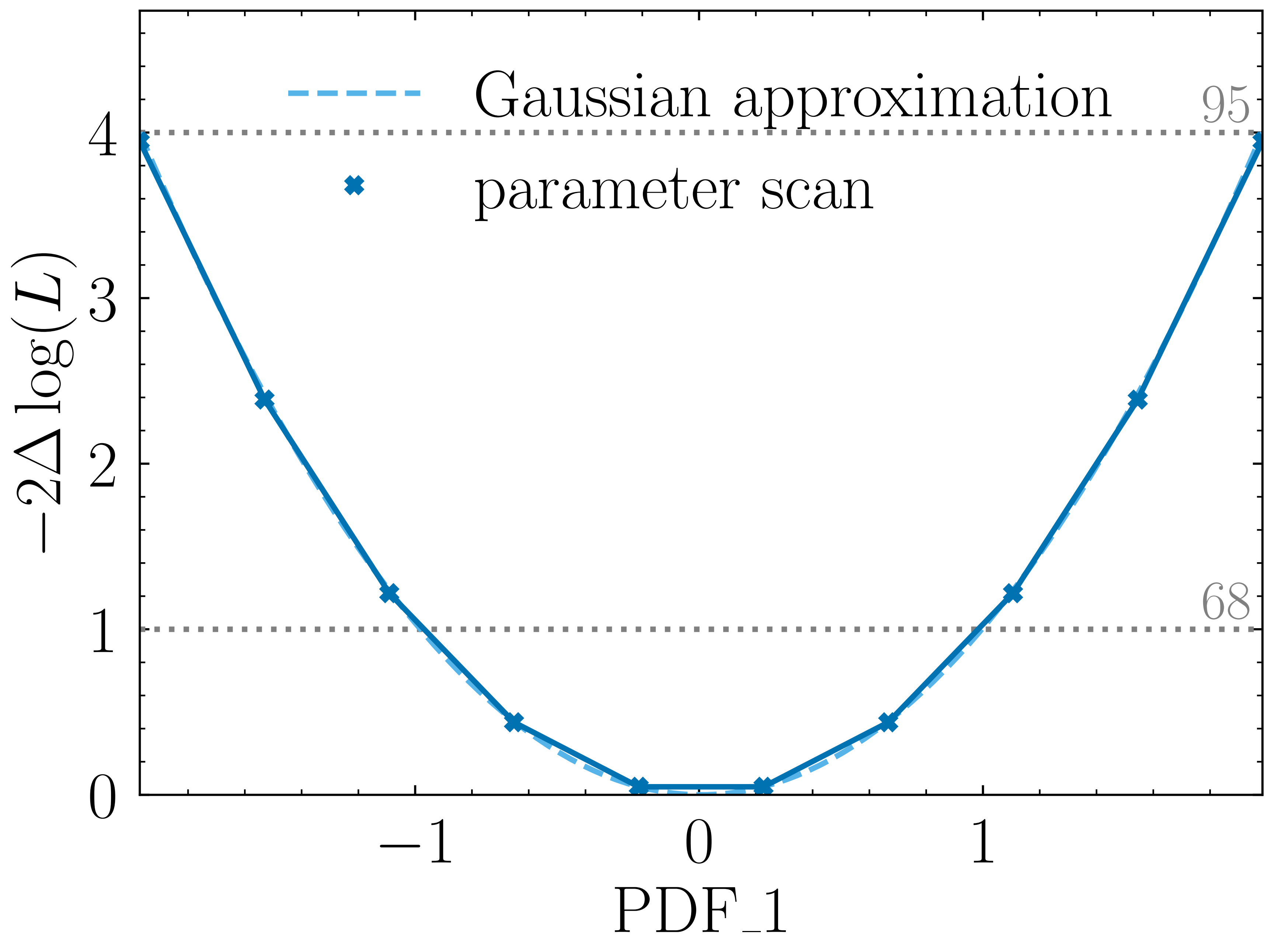
Implementation

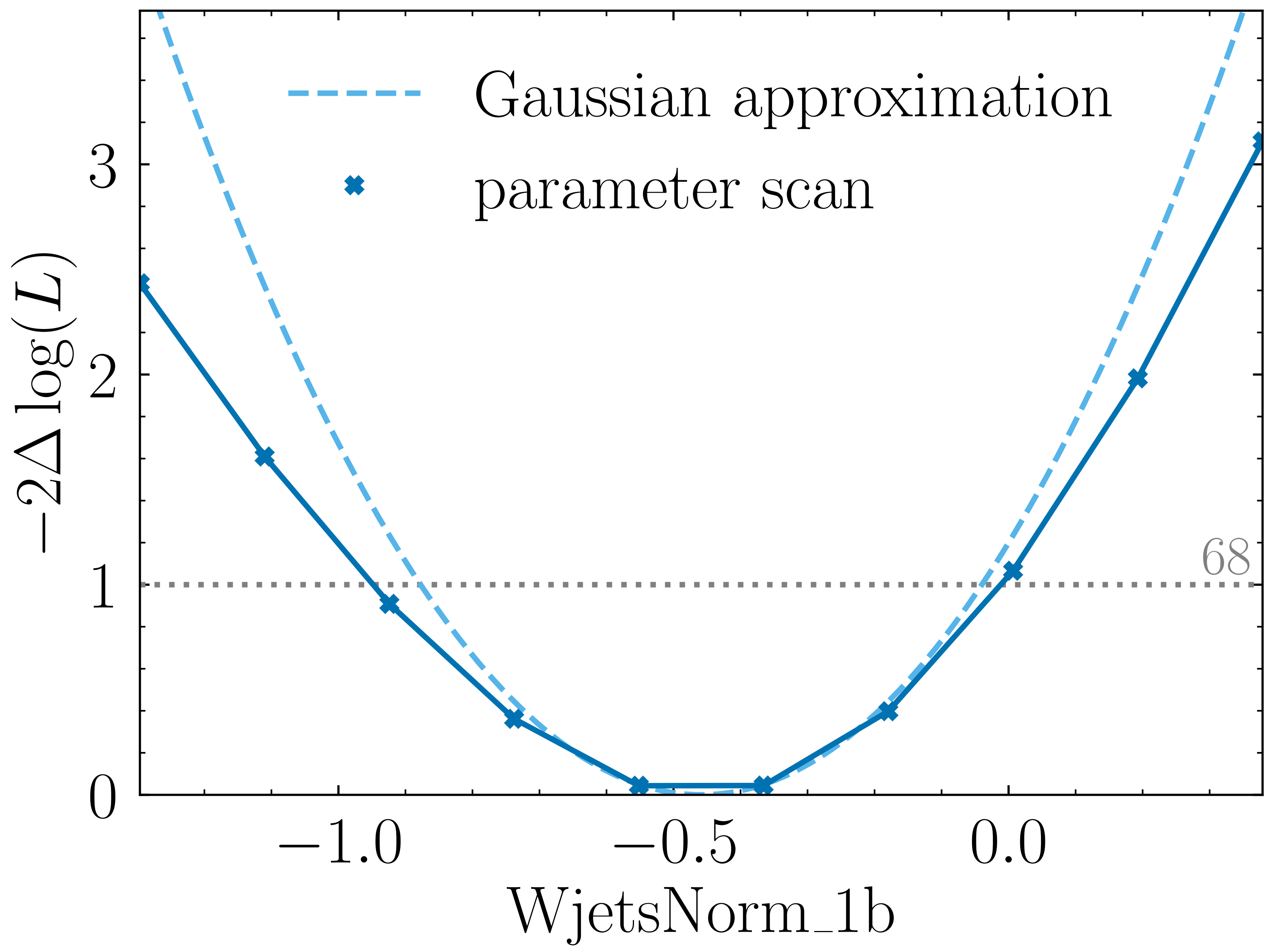
- Low dimensional fit to **only** C_{tG} and total cross section measurements
- Neglect theory uncertainties
- **Excellent agreement** between both methods of implementation



Published Likelihoods

Parameter scans with cabinetry

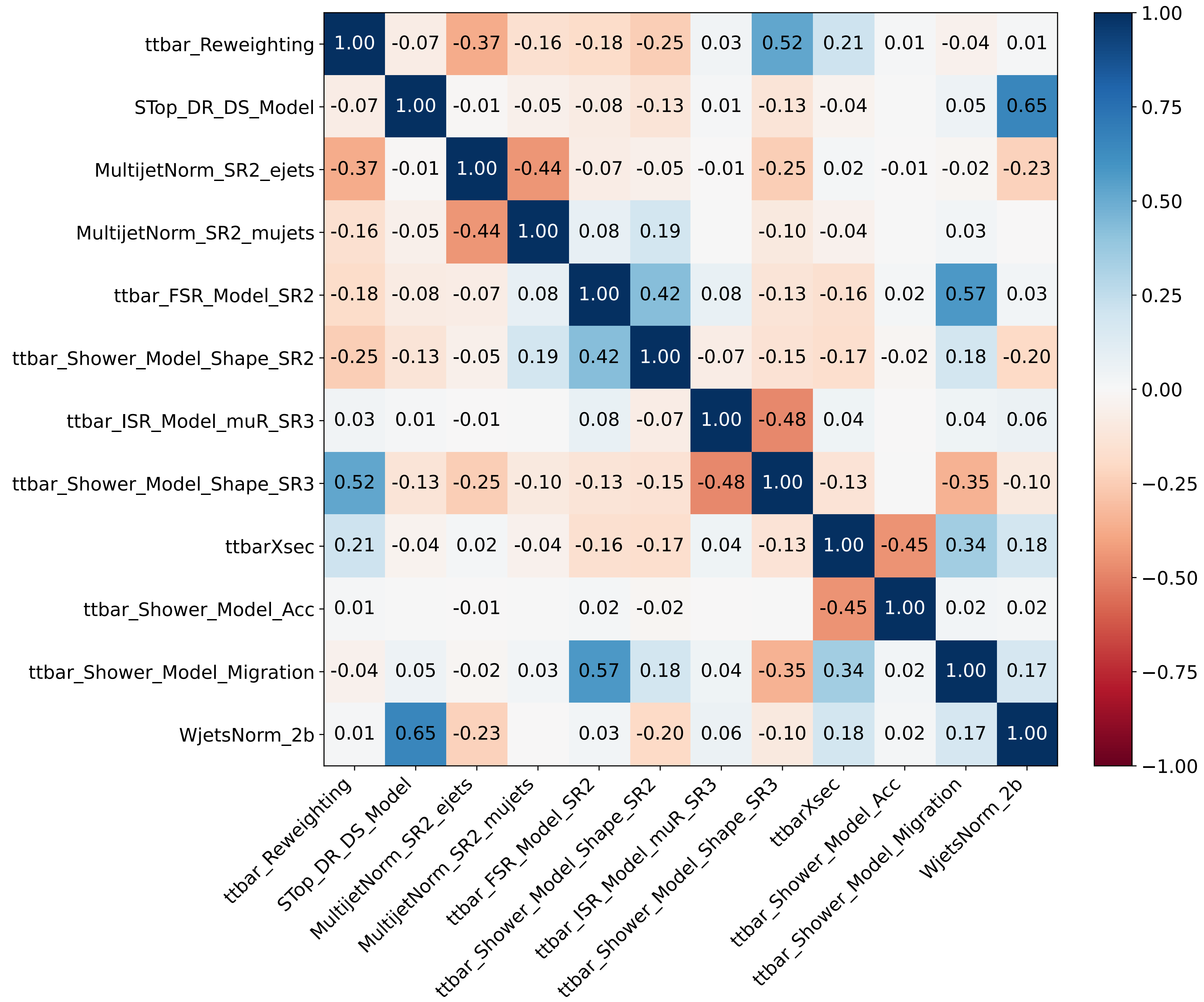




- NPs are all very Gaussian, only small number of exceptions
  - Validates **Gaussian** constraint term $\mathcal{C}(\theta_i, \sigma_i)$ for systematics

Published Likelihoods

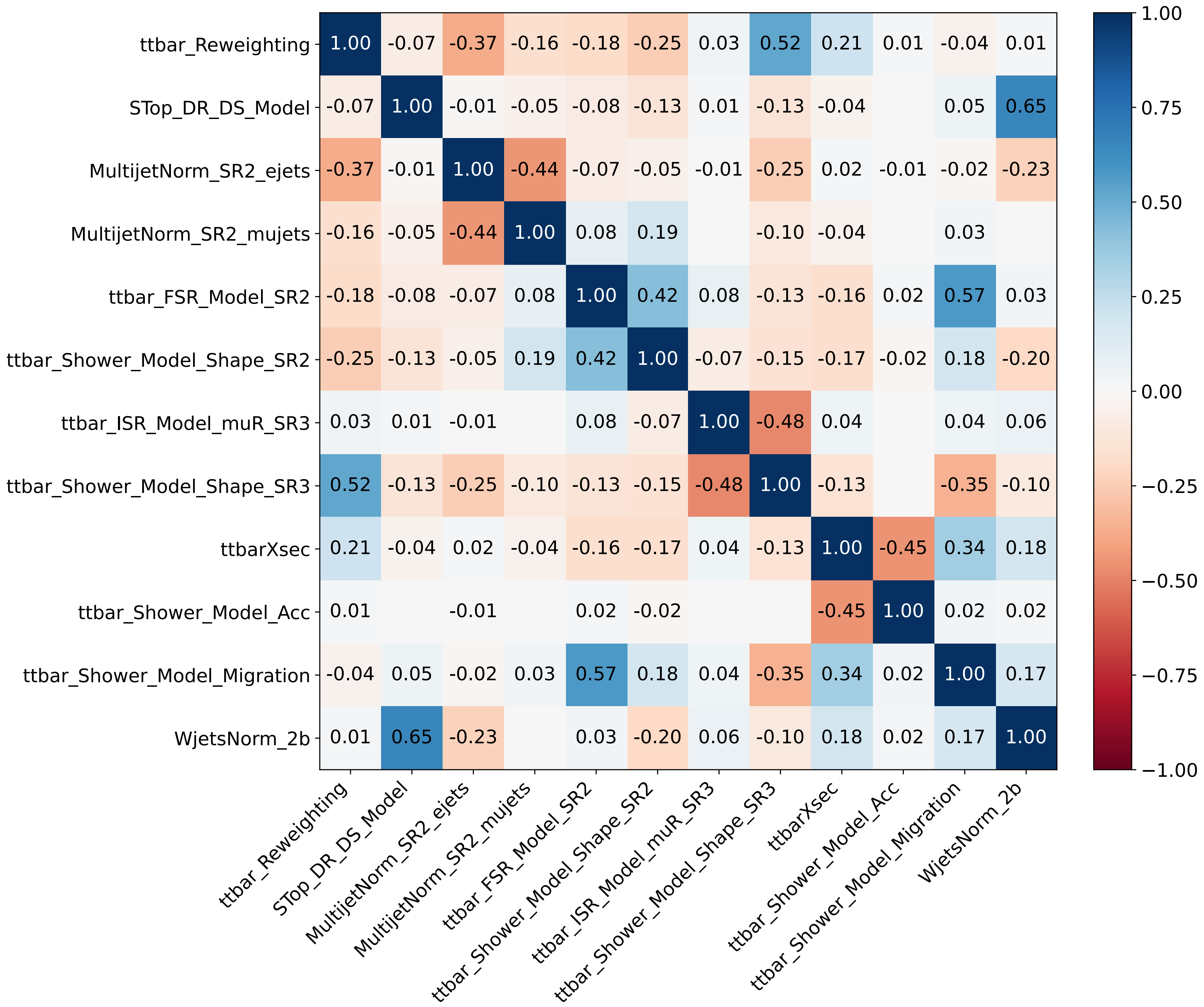
Concerning Correlations

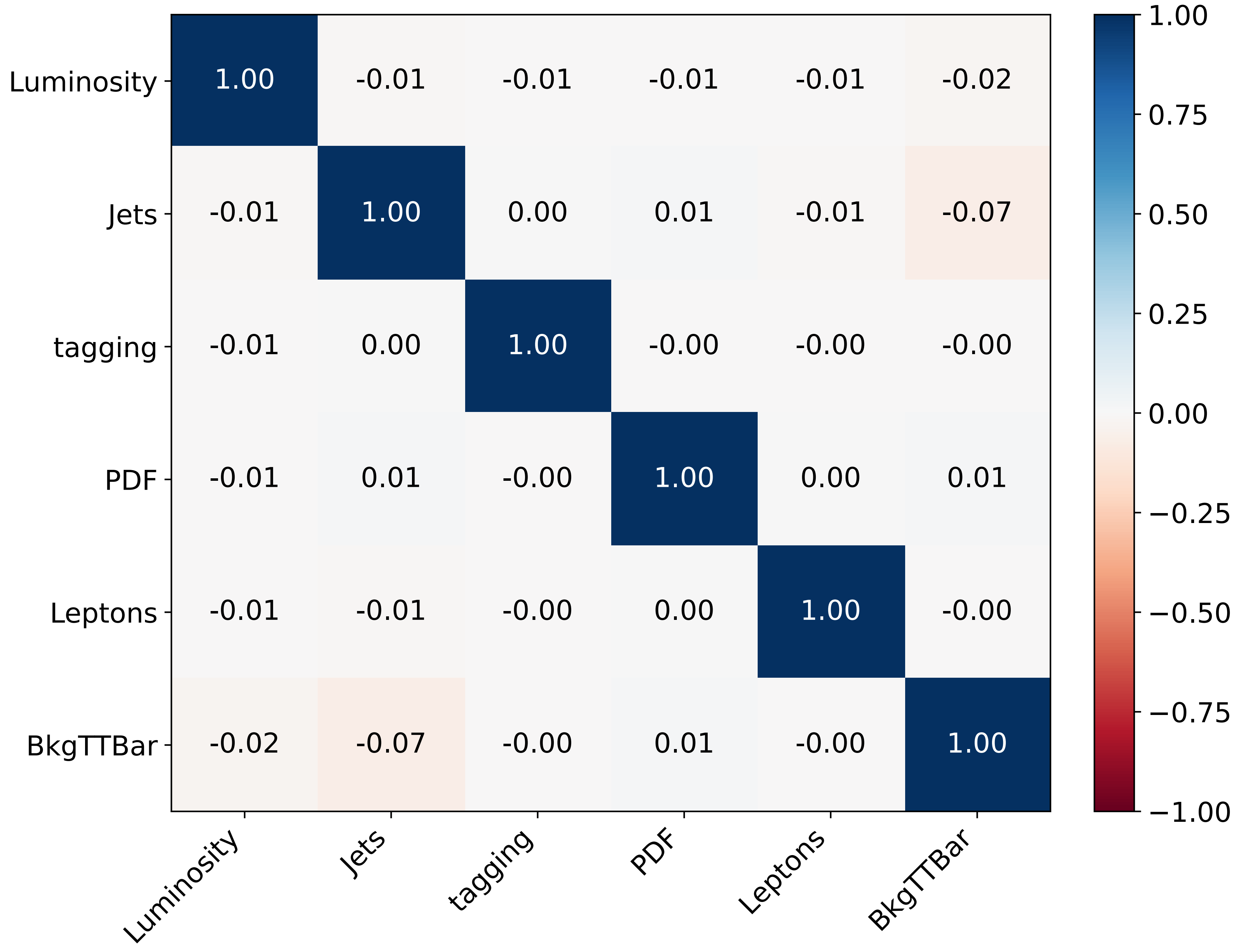


➤ **Currently:** No correlations between uncertainties in SFitter

Published Likelihoods

Concerning Correlations





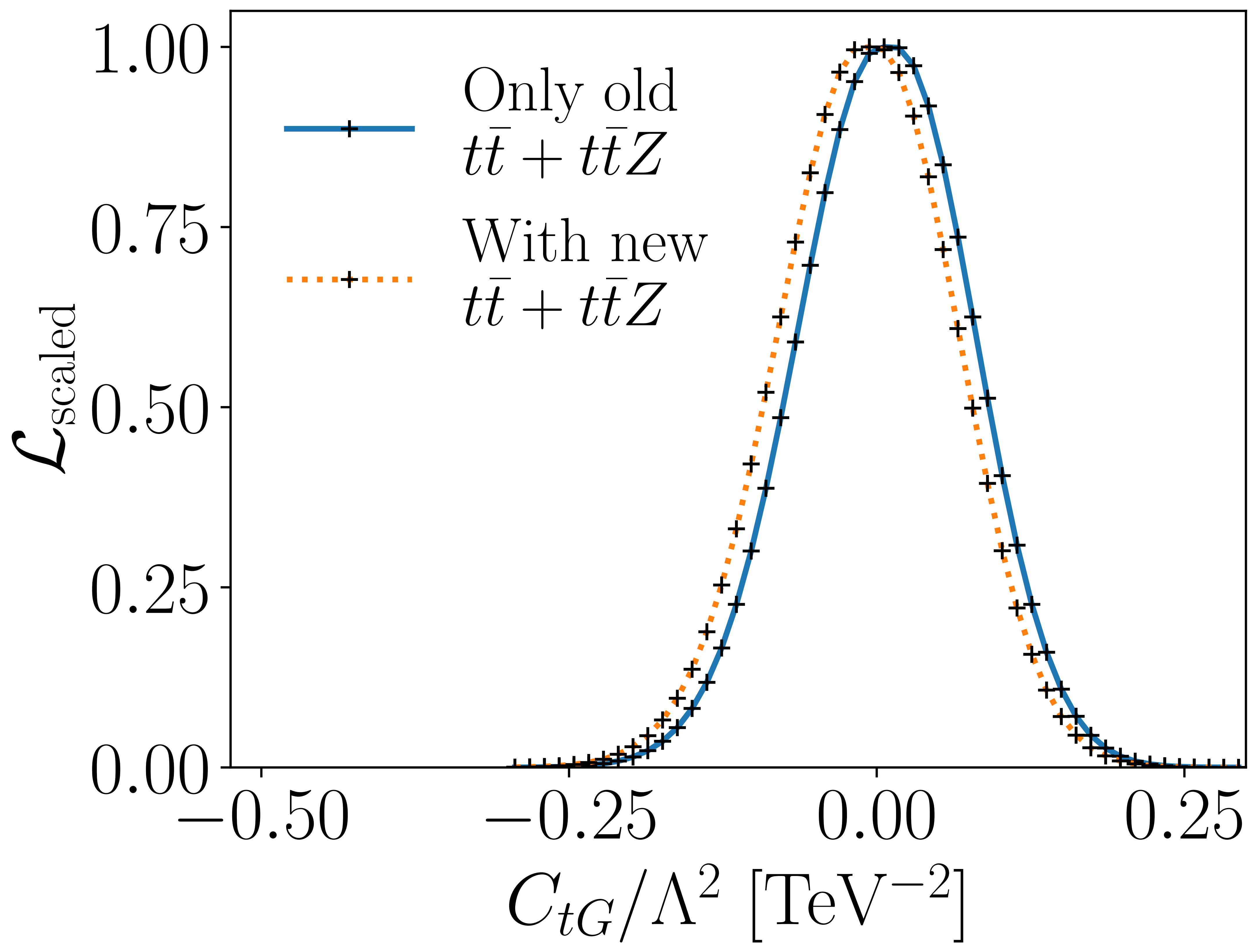
➤ **Currently:** No correlations between uncertainties in SFitter

  ➤ Correlations of systematics included in SFitter are **negligibly small**

Published Likelihoods

Constraints

- Visible shift from new measurements
- Constraints **shift slightly** after including both new measurements
- Measurements of **total cross sections** barely affect constraints



Concluding

- **Summary:** Uncertainties and correlations are essential to SFitter constraints
  - Large effect of **theory uncertainties** in the top sector
  - Published likelihoods provide an alternative way to use experimental data
    - **Validates assumptions** made in previous analyses

- **However:** Currently available likelihoods not particularly SMEFT sensitive
  - Publication of more **differential measurements** would be useful
  - **Global SMEFT analysis** requires data from all kinds of processes