SMEFiT: a Monte Carlo global analysis of the SMEFT

Juan Rojo
VU Amsterdam & Theory group, Nikhef

Higgs and Effective Field Theory (HEFT2019)
CP3, Louvain-la-Neuve

Towards a global SMEFT analysis

The Standard Model EFT

Systematic parametrisation of the **theory space** in vicinity of Standard Model

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_j^{N_{d8}} \frac{b_j}{\Lambda^4} \mathcal{O}_j^{(8)} + \ldots$$

- ☑ SMEFT: **low-energy limit** of generic UV-complete theories at high energies
- ☑ Assumes **SM field content and symmetries** (except the accidental ones)
- ☑ **Complete basis** at any given mass-dimension
- ☑ **Fully renormalizable**, full-fledged QFT: can compute higher orders in QCD and EW
- ☑ Can be matched to **any BSM model** that reduces to the SM at low energies

The Standard Model EFT

Systematic parametrisation of the **theory space** in vicinity of Standard Model

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_j^{N_{d8}} \frac{b_j}{\Lambda^4}\mathcal{O}_j^{(8)} + \ldots$$

- Some operators induce **growth with the partonic centre-of-mass energy**: increased sensitivity in LHC cross-sections in the TeV region

$$\sigma(E) = \sigma_{\rm SM}(E)\left(1 + \sum_i^{N_{d6}} \omega_i \frac{c_i\, m_{\rm SM}^2}{\Lambda^2} + \sum_i^{N_{d6}} \widetilde{\omega}_i \frac{c_i\, E^2}{\Lambda^2} + \mathcal{O}\left(\Lambda^{-4}\right)\right)$$

*enhanced sensitivity from **TeV-scale processes:***
unique feature of LHC

The Standard Model EFT

Systematic parametrisation of the **theory space** in vicinity of Standard Model

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_j^{N_{d8}} \frac{b_j}{\Lambda^4}\mathcal{O}_j^{(8)} + \ldots$$

- Some operators induce **growth with the partonic centre-of-mass energy**: increased sensitivity in LHC cross-sections in the TeV region

$$\sigma(E) = \sigma_{\rm SM}(E)\left(1 + \sum_i^{N_{d6}} \omega_i \frac{c_i\, m_{\rm SM}^2}{\Lambda^2} + \sum_i^{N_{d6}} \widetilde{\omega}_i \frac{c_i\, E^2}{\Lambda^2} + \mathcal{O}\left(\Lambda^{-4}\right)\right)$$

- The number of SMEFT operators is large: **59 non-redundant operators at dimension 6** for one fermion generation, **2499 operators** without any flavour assumption

- A global SMEFT analysis needs to explore a **huge complicated parameter space**

Recipe for a global SMEFT analysis

Theory

(**N)NLO QCD + NLO EW** for SM xsecs

NLO QCD for SMEFT contributions

State-of-the-art **Parton Distributions**

Data

Higgs and **gauge boson** production

Top quark and **jet** production

Precision **LEP**, **low energy**, **flavour, ….**

Global SMEFT fit

Bounds can be compared with

specific UV completions

New data incorporated without redoing fit

Delivery

Efficient exploration of **parameter space**

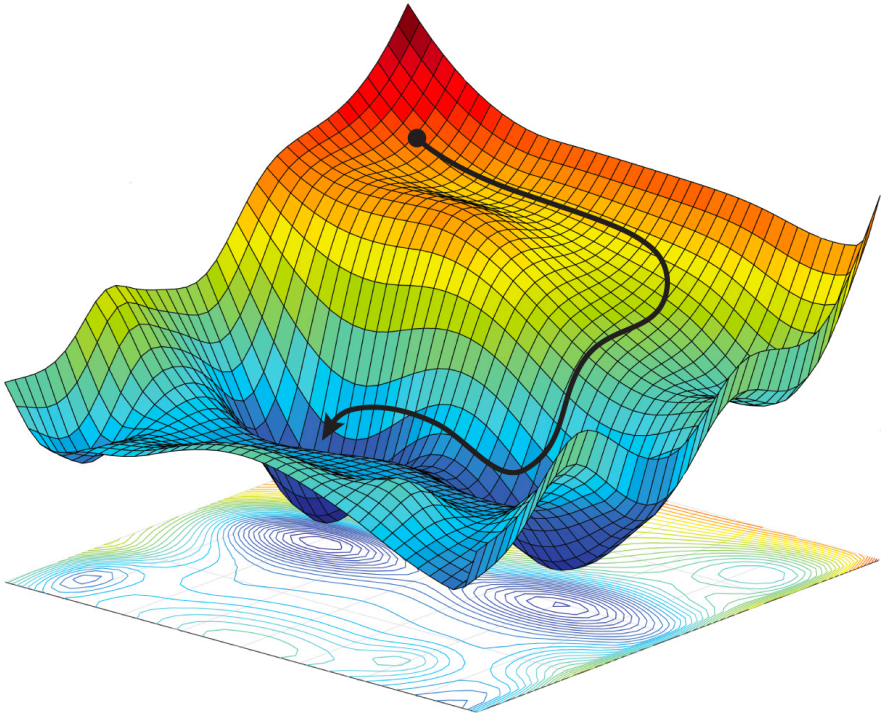
Faithful **uncertainty estimate** (exp & th)

Avoiding under- and over-fitting

Methodology

The optimisation conundrum

(Stochastic) Gradient Descent



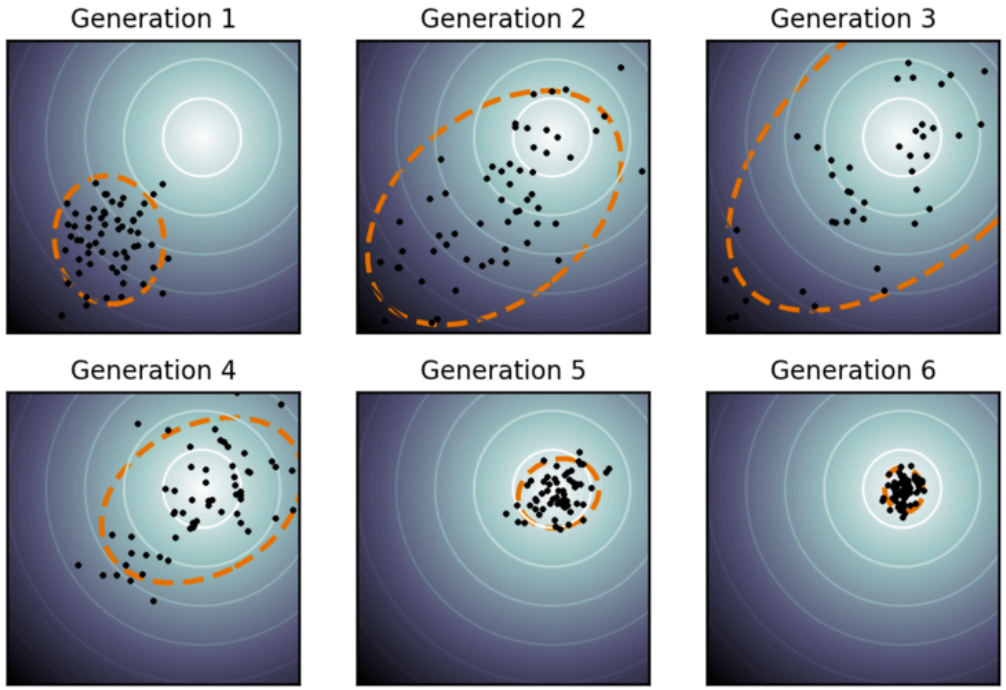
- **Deterministic algorithms**: follow the gradient of the cost function

- **Evolutionary algorithms**: act on population of solutions with random mutations and selection

- A challenge for any SMEFT global analysis is the efficient exploration of the **huge parameter space**

- Several pitfalls to be avoided: under-fitting, over-fitting, **local minima**, saddle points, ….

Genetic Algorithms



The SMEFiT framework

A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector,

Nathan P. Hartland, Fabio Maltoni, Emanuele R. Nocera, Juan Rojo,

Emma Slade, Eleni Vryonidou, Cen Zhang, arXiv:1901.05965, JHEP in press

The SMEFiT method

- Generate a large sample of **Monte Carlo replicas** to construct the **probability distribution** in the space of experimental data

$$\mathcal{O}_i^{(\mathrm{art})(k)} = S_{i,N}^{(k)} \mathcal{O}_i^{(\mathrm{exp})} \left(1 + \sum_{\alpha=1}^{N_{\mathrm{sys}}} r_{i,\alpha}^{(k)} \sigma_{i,c}^{(\mathrm{sys})} + r_i^{(k)} \sigma_i^{(\mathrm{stat})} \right) , \quad k = 1, \ldots, N_{\mathrm{rep}}$$

cross-section for k-th replica — ***central value (data)*** — ***correlated systematic uncertainties*** — ***statistical uncertainties*** — ***# MC replicas***

The SMEFiT method

- Generate a large sample of **Monte Carlo replicas** to construct the **probability distribution** in the space of experimental data

$$\mathcal{O}_i^{(\mathrm{art})(k)} = S_{i,N}^{(k)} \mathcal{O}_i^{(\mathrm{exp})} \left(1 + \sum_{\alpha=1}^{N_{\mathrm{sys}}} r_{i,\alpha}^{(k)} \sigma_{i,c}^{(\mathrm{sys})} + r_i^{(k)} \sigma_i^{(\mathrm{stat})} \right) , \quad k = 1, \ldots, N_{\mathrm{rep}}$$

- Construct theory calculations where the SM is **extended by SMEFT corrections**

to be determined from the data

$$\sigma_i^{\mathrm{th}} \left(\{c_n\} \right) = \sigma_{\mathrm{SM},i} + \sum_{n=1}^{N_{\mathrm{op}}} \widetilde{\sigma}_{i,n} \frac{c_n}{\Lambda^2} + \sum_{n,m=1}^{N_{\mathrm{op}}} \widetilde{\sigma}_{i,nm} \frac{c_n c_m}{\Lambda^4} , \quad i = 1 \ \ldots, N_{\mathrm{dat}}$$

SM: compute at (N)NLO QCD

SMEFT: compute at (N)LO QCD with aMC@NLO

The SMEFiT method

- Generate a large sample of **Monte Carlo replicas** to construct the **probability distribution** in the space of experimental data

$$\mathcal{O}_i^{(\mathrm{art})(k)} = S_{i,N}^{(k)} \mathcal{O}_i^{(\mathrm{exp})} \left(1 + \sum_{\alpha=1}^{N_{\mathrm{sys}}} r_{i,\alpha}^{(k)} \sigma_{i,c}^{(\mathrm{sys})} + r_i^{(k)} \sigma_i^{(\mathrm{stat})} \right) , \quad k = 1, \ldots, N_{\mathrm{rep}}$$

- Construct theory calculations where the SM is **extended by SMEFT corrections**

$$\mathcal{O}_i^{\mathrm{th}} \left(\{c_n\} \right) = \sigma_{\mathrm{SM},i} + \sum_{n=1}^{N_{\mathrm{op}}} \widetilde{\sigma}_{i,n} \frac{c_n}{\Lambda^2} + \sum_{n,m=1}^{N_{\mathrm{op}}} \widetilde{\sigma}_{i,nm} \frac{c_n c_m}{\Lambda^4} , \quad i = 1 \ \ldots, N_{\mathrm{dat}}$$

- Determine the SMEFT coefficients **replica-by-replica** by minimising a cost function

$$E(\{c_l^{(k)}\}) \equiv \frac{1}{N_{\mathrm{dat}}} \sum_{i,j=1}^{N_{\mathrm{dat}}} \left(\mathcal{O}_i^{(\mathrm{th})} \left(\{c_n^{(k)}\} \right) - \mathcal{O}_i^{(\mathrm{art})(k)} \right) (\mathrm{cov}^{-1})_{ij} \left(\mathcal{O}_j^{(\mathrm{th})} \left(\{c_n^{(k)}\} \right) - \mathcal{O}_j^{(\mathrm{art})(k)} \right)$$

The SMEFiT method

- Determine the SMEFT coefficients **replica-by-replica** by minimising a cost function

$$E(\{c_l^{(k)}\}) \equiv \frac{1}{N_{\rm dat}} \sum_{i,j=1}^{N_{\rm dat}} \left(\mathcal{O}_i^{(\rm th)}\left(\{c_n^{(k)}\}\right) - \mathcal{O}_i^{(\rm art)(k)} \right) (\mathrm{cov}^{-1})_{ij} \left(\mathcal{O}_j^{(\rm th)}\left(\{c_n^{(k)}\}\right) - \mathcal{O}_j^{(\rm art)(k)} \right)$$

- The covariance matrix includes **all sources of experimental errors** + some theory errors

t_0 prescription

$$(\mathrm{cov}_{t_0})_{ij}^{(\rm exp)} \equiv \left(\sigma_i^{(\rm stat)}\right)^2 \delta_{ij} + \left(\sum_{\alpha=1}^{N_{\rm sys}} \sigma_{i,\alpha}^{(\rm sys)} \sigma_{j,\alpha}^{(\rm sys)} \mathcal{O}_i^{(\rm exp)} \mathcal{O}_j^{(\rm exp)} + \sum_{\beta=1}^{N_{\rm norm}} \sigma_{i,\beta}^{(\rm norm)} \sigma_{j,\beta}^{(\rm norm)} \mathcal{O}_i^{(\rm th,0)} \mathcal{O}_j^{(\rm th,0)} \right)$$

$$\mathrm{cov}_{ij} = \mathrm{cov}_{ij}^{(\rm exp)} + \mathrm{cov}_{ij}^{(\rm th)}$$

th uncertainties: PDFs

can be extended to MHOUs

$$\mathrm{cov}_{ij}^{(\rm th)} = \left\langle \mathcal{O}_i^{(\rm th)(r)} \mathcal{O}_j^{(\rm th)(r)} \right\rangle_{\rm rep} - \left\langle \mathcal{O}_i^{(\rm th)(r)} \right\rangle_{\rm rep} \left\langle \mathcal{O}_j^{(\rm th)(r)} \right\rangle_{\rm rep} ,$$

The SMEFiT method

- Determine the SMEFT coefficients **replica-by-replica** by minimising a cost function

$$E(\{c_l^{(k)}\}) \equiv \frac{1}{N_{\rm dat}} \sum_{i,j=1}^{N_{\rm dat}} \left(\mathcal{O}_i^{(\rm th)}\left(\{c_n^{(k)}\}\right) - \mathcal{O}_i^{(\rm art)(k)} \right) (\mathrm{cov}^{-1})_{ij} \left(\mathcal{O}_j^{(\rm th)}\left(\{c_n^{(k)}\}\right) - \mathcal{O}_j^{(\rm art)(k)} \right)$$

- The covariance matrix includes **all sources of experimental errors** + some theory errors

$$\mathrm{cov}_{ij} = \mathrm{cov}_{ij}^{(\rm exp)} + \mathrm{cov}_{ij}^{(\rm th)}$$

- The ensemble of coefficients $\{c_l^{(k)}\}$ then provides a sampling of the **probability density** in the **SMEFT parameter space**

$$\left\langle c_l \right\rangle \equiv \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} c_l^{(k)} \qquad \rho\left(c_i, c_j\right) = \frac{\frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} c_i^{(k)} c_j^{(k)} - \left\langle c_i \right\rangle \left\langle c_j \right\rangle}{\delta c_i \delta c_j} .$$

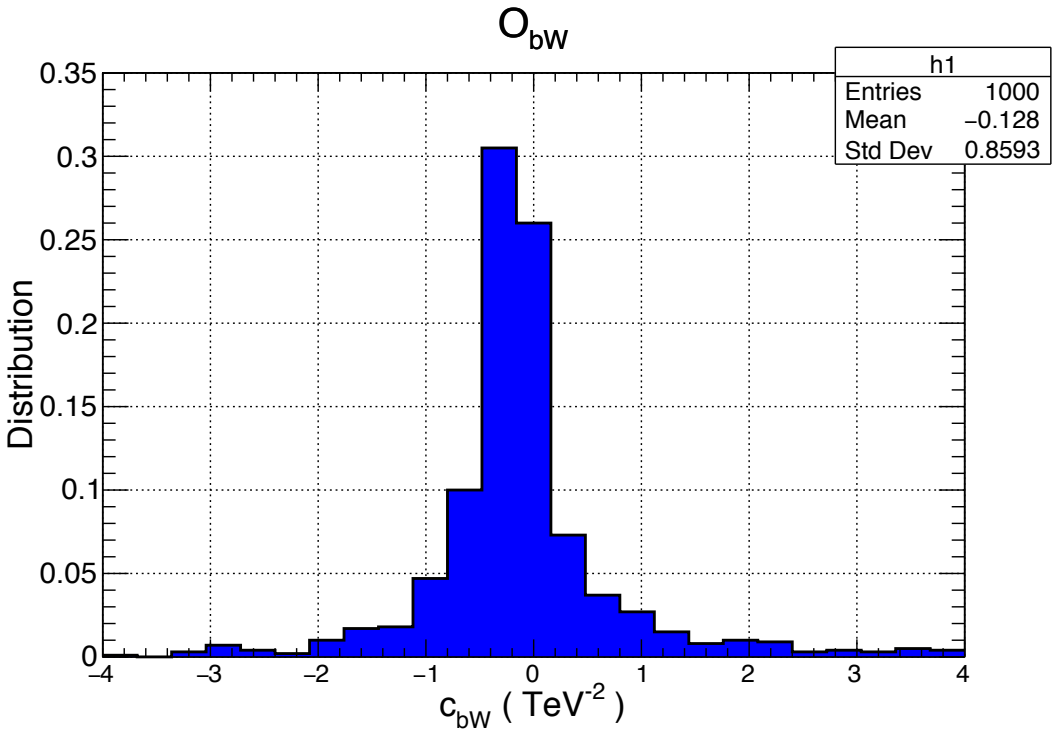
Sampling the SMEFT probability distribution

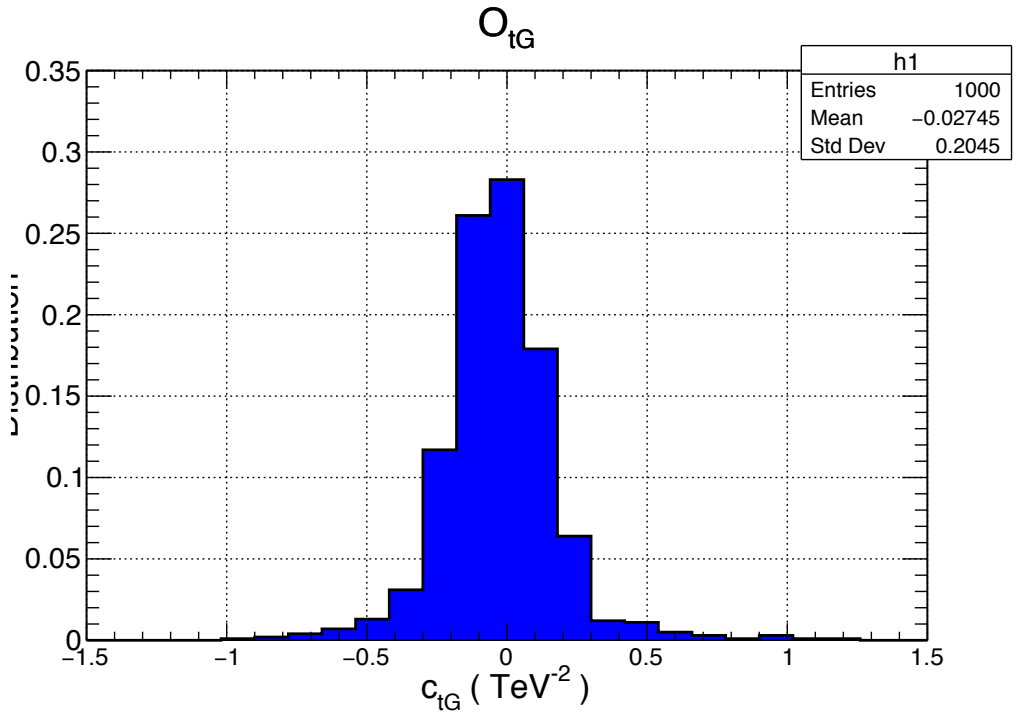
- The output of SMEFiT is a sampling of the **probability distribution** in the SMEFT space

$$\left\{ c_n^{(k)} \right\}, \quad n = 1 \ldots, N_{\rm op}, \quad k = 1 \ldots, N_{\rm rep}$$

- Used to **evaluate statistical estimators** such as variances, correlations, higher moments, …

- Distributions are **reasonably Gaussian** for well-constrained degrees of freedom ….





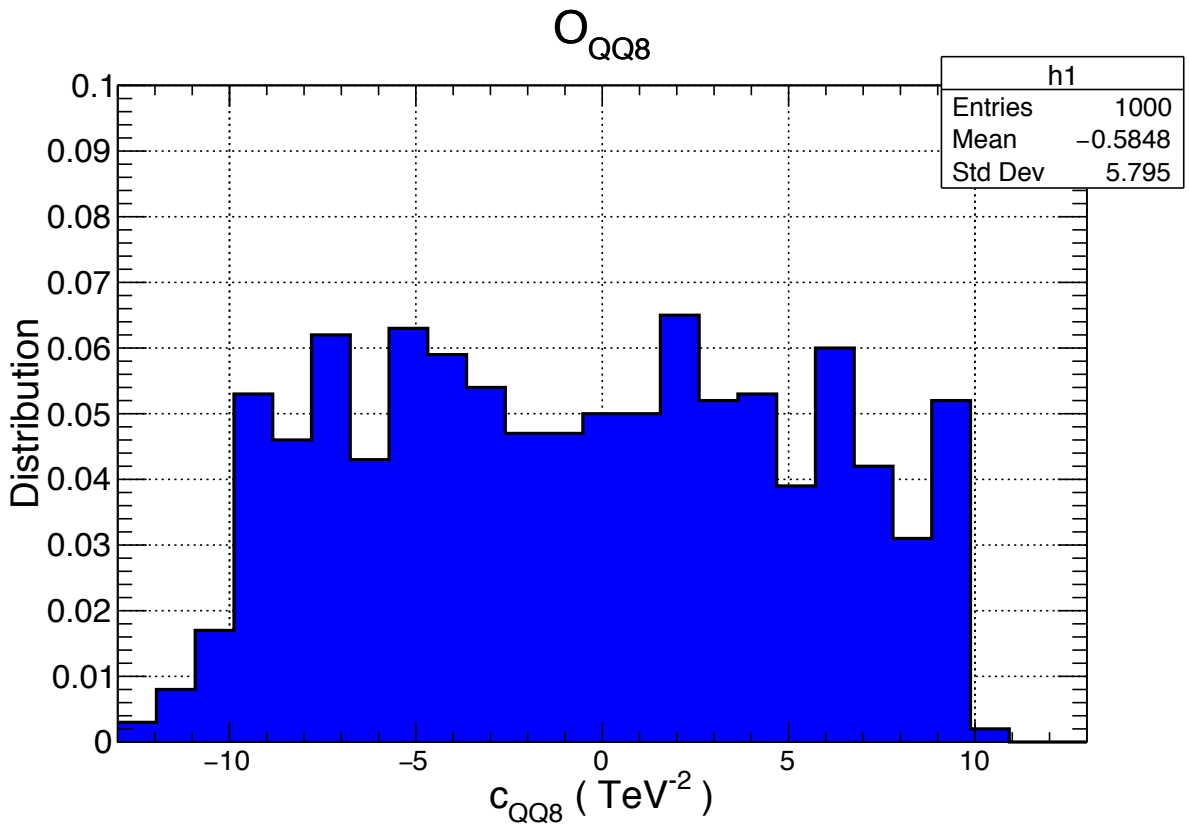
Sampling the SMEFT probability distribution

- The output of SMEFiT is a sampling of the **probability distribution** in the SMEFT space

$$\left\{ c_n^{(k)} \right\}, \quad n = 1 \ \ldots, N_{\rm op}, \quad k = 1 \ \ldots, N_{\rm rep}$$

- Used to **evaluate statistical estimators** such as variances, correlations, higher moments, …

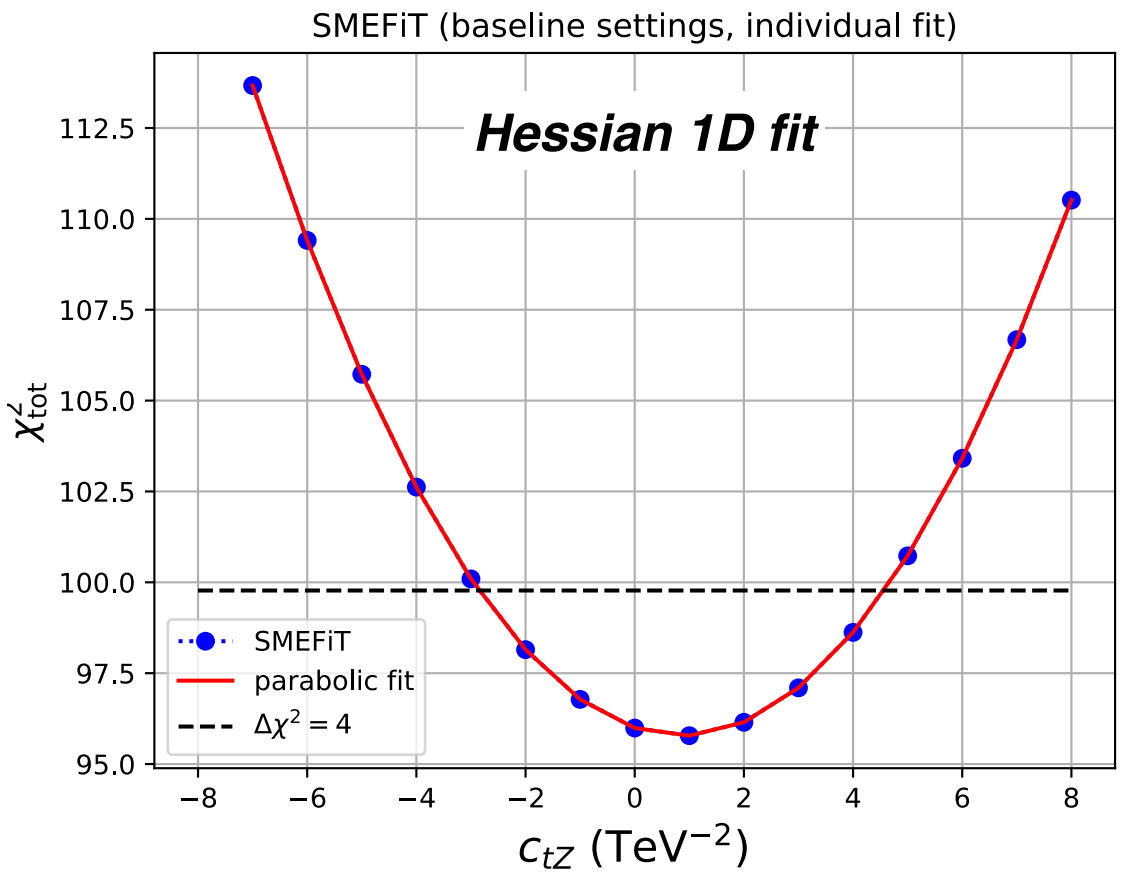
- …. but much less so for **under-constrained** or **redundant** operators



The SMEFiT method

- Uncertainties on the SMEFT degrees of freedom evaluated from **variance of MC sample**

$$\left(\delta c_n\right)^2 = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \left(c_n^{(k)}\right)^2 - \left\langle c_n \right\rangle^2$$



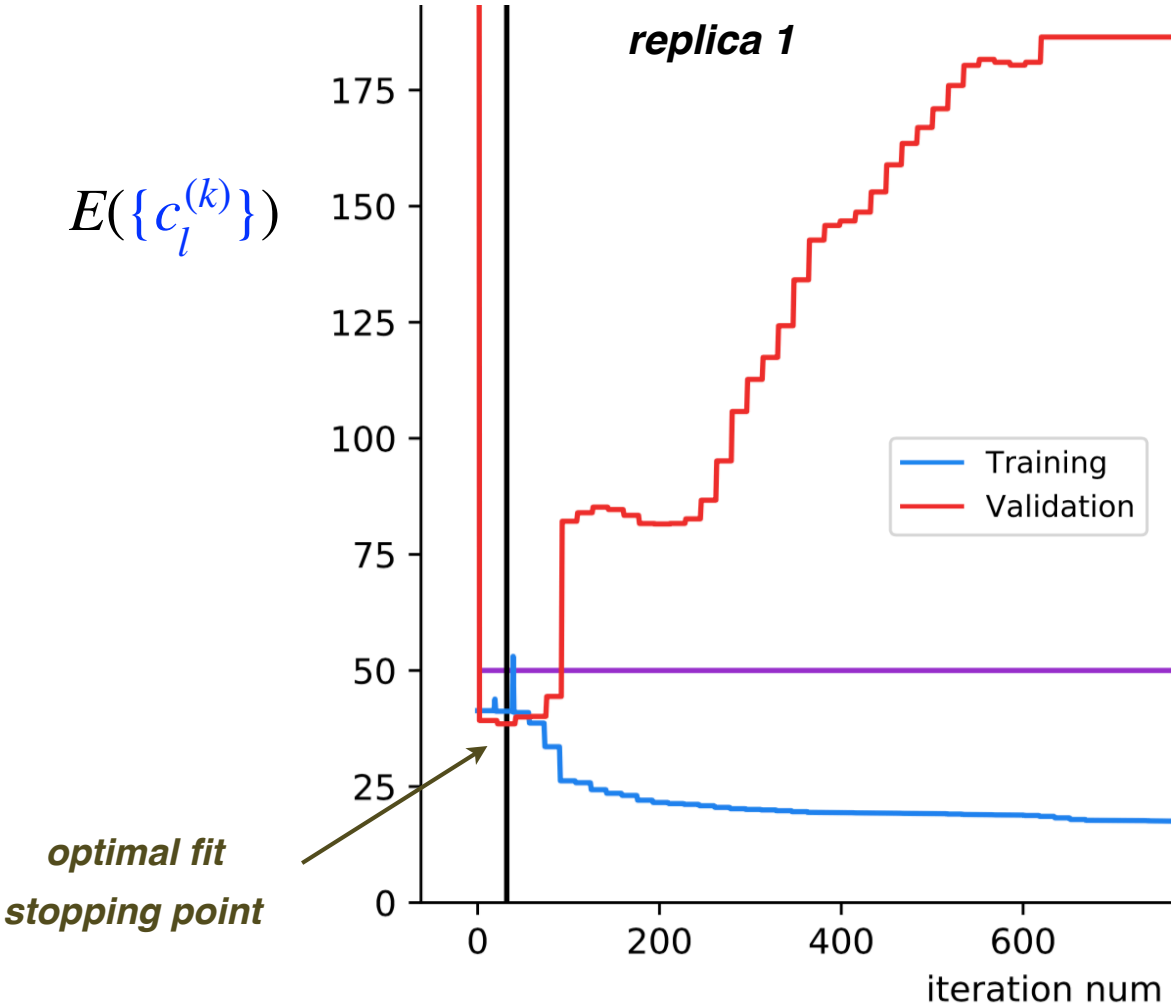
- For single-parameter fits, **Monte Carlo results** benchmarked with **Hessian method**, finding good agreement

- The Hessian method **numerically less stable** as dimensionality of parameter space increases

The SMEFiT method

- Since in general there will be **unconstrained/degenerate directions** in the parameter space, it is crucial to avoid overfitting (that is, fitting statistical fluctuations)

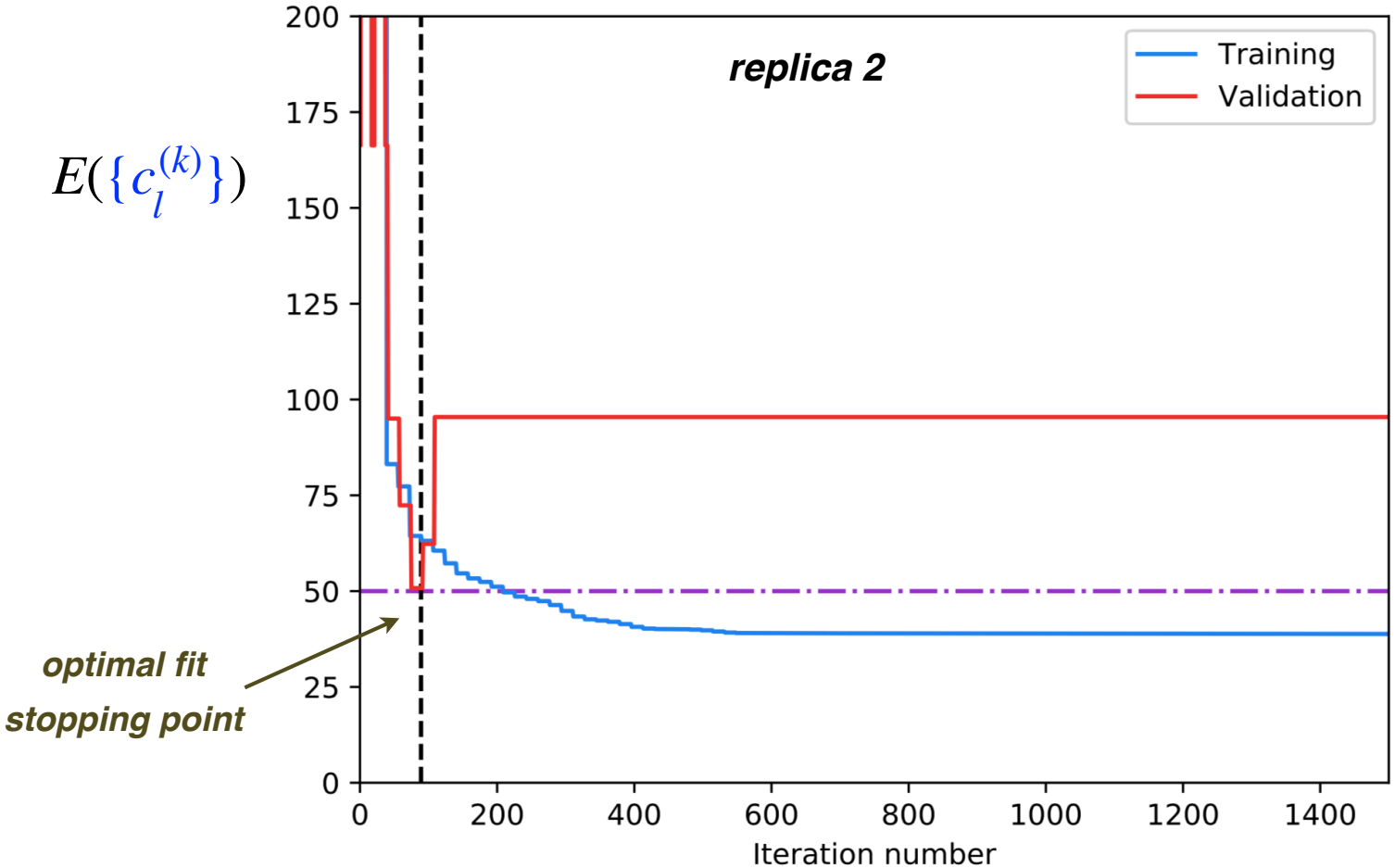
- Achieved by the **cross-validation look-back validation stopping** method



The SMEFiT method

- Since in general there will be **unconstrained/degenerate directions** in the parameter space, it is crucial to avoid overfitting (that is, fitting statistical fluctuations)

- Achieved by the **cross-validation look-back validation stopping** method



SMEFiT code structure

Stand-alone **Python code**, which exploits functionalities of the **NNPDF framework**

NNPDF code

- **Experimental data** and covariance matrices
- **NLO** APPLgrids + **NNLO** C-factors (for processes used in PDF fit)

aMC@NLO

- NLO **QCD** (benchmark)
- LO, NLO **SMEFT**
- Both **O(Λ^{-2})** and **O(Λ^{-4})** from ***d=6* operators**

MCFM

- NLO **QCD** (consistent choice of PDFs)
- Cross-checks of **aMC@NLO**

Python analysis code

- Assemble **theory predictions** for generic SMEFT Wilson coefficients
- **Optimisation** with Sequential Quadratic Programming (**SciPy**)
- Look-back **cross-validation stopping**
- **Monte Carlo replicas** for uncertainty propagation

SMEFiT:
the top quark case

A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector,

Nathan P. Hartland, Fabio Maltoni, Emanuele R. Nocera, Juan Rojo,

Emma Slade, Eleni Vryonidou, Cen Zhang, arXiv:1901.05965, JHEP in press

Operator basis

- We follow the same flavour assumptions as in the **LHC Top WG note**
- Minimal Flavour Violation (MFV), diagonal CKM, zero Yukawas for first two quark gens, CP conservation assumed
- Include those SMEFT dimension-6 operators of Warsaw basis with **at least one top quark**
- The fit includes a total of **34 independent degrees of freedom**
- Include both **interference** and **quadratic contributions** from these operators

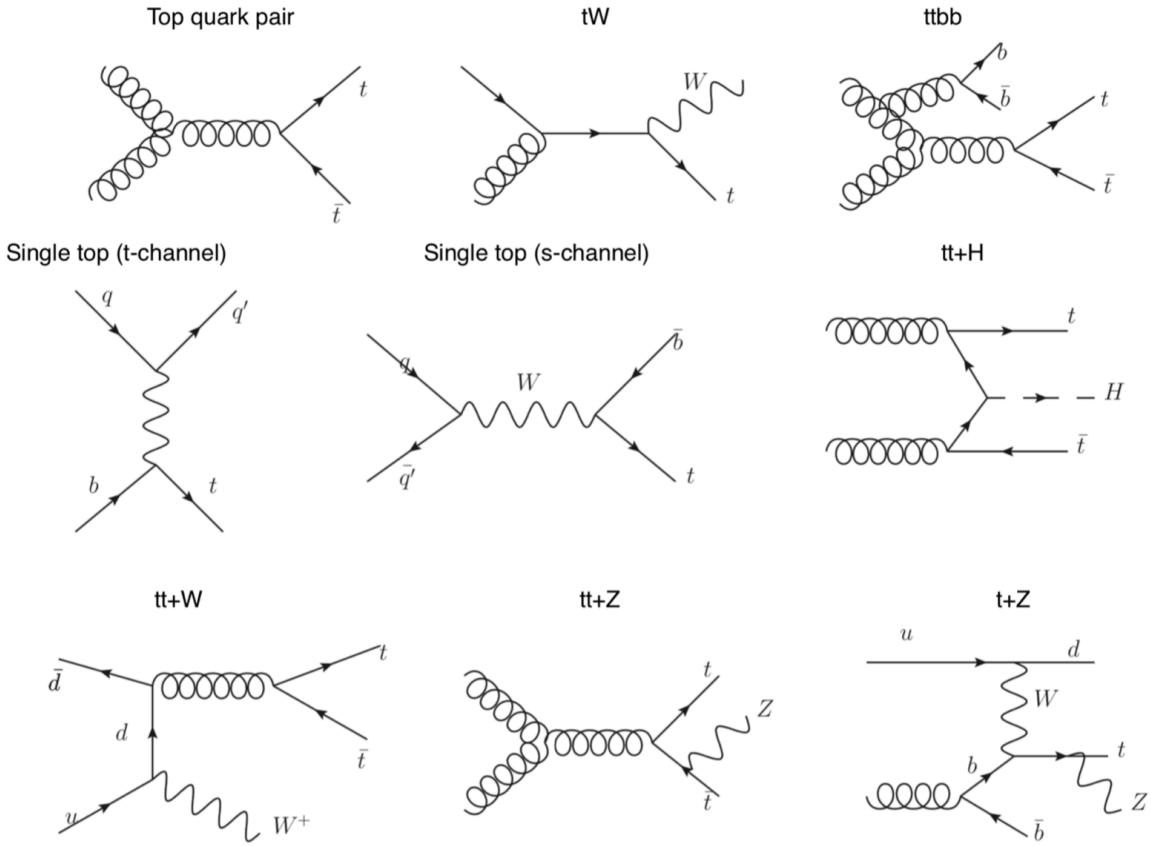
	Class	Notation	Degree of Freedom	Operator Definition
4-heavy	$QQQQ$	OQQ1	c^1_{QQ}	$2C_{qq}^{1(3333)} - \frac{2}{3}C_{qq}^{3(3333)}$
		OQQ8	c^8_{QQ}	$8C_{qq}^{3(3333)}$
		OQt1	c^1_{Qt}	$C_{qu}^{1(3333)}$
		OQt8	c^8_{Qt}	$C_{qu}^{8(3333)}$
		OQb1	c^1_{Qb}	$C_{qd}^{1(3333)}$
		OQb8	c^8_{Qb}	$C_{qd}^{8(3333)}$
		Ott1	c^1_{tt}	$C_{uu}^{(3333)}$
		Otb1	c^1_{tb}	$C_{ud}^{1(3333)}$
		Otb8	c^8_{tb}	$C_{ud}^{8(3333)}$
		OQtQb1	c^1_{QtQb}	$C_{quqd}^{1(3333)}$
		OQtQb8	c^8_{QtQb}	$C_{quqd}^{8(3333)}$
2-heavy-2-light	$QQqq$	O81qq	$c^{1,8}_{Qq}$	$C_{qq}^{1(i33i)} + 3C_{qq}^{3(i33i)}$
		O11qq	$c^{1,1}_{Qq}$	$C_{qq}^{1(ii33)} + \frac{1}{6}C_{qq}^{1(i33i)} + \frac{1}{2}C_{qq}^{3(i33i)}$
		O83qq	$c^{3,8}_{Qq}$	$C_{qq}^{1(i33i)} - C_{qq}^{3(i33i)}$
		O13qq	$c^{3,1}_{Qq}$	$C_{qq}^{3(ii33)} + \frac{1}{6}(C_{qq}^{1(i33i)} - C_{qq}^{3(i33i)})$
		O8qt	c^8_{tq}	$C_{qu}^{8(ii33)}$
		O1qt	c^1_{tq}	$C_{qu}^{1(ii33)}$
		O8ut	c^8_{tu}	$2C_{uu}^{(i33i)}$
		O1ut	c^1_{tu}	$C_{uu}^{(ii33)} + \frac{1}{3}C_{uu}^{(i33i)}$
		O8qu	c^8_{Qu}	$C_{qu}^{8(33ii)}$
		O1qu	c^1_{Qu}	$C_{qu}^{1(33ii)}$
		O8dt	c^8_{td}	$C_{ud}^{8(33ii)}$
		O1dt	c^1_{td}	$C_{ud}^{1(33ii)}$
		O8qd	c^8_{Qd}	$C_{qd}^{8(33ii)}$
		O1qd	c^1_{Qd}	$C_{qd}^{1(33ii)}$
2-heavy + V/h	$QQ + V, G, \varphi$	OtG	c_{tG}	$\mathrm{Re}\{C_{uG}^{(33)}\}$
		OtW	c_{tW}	$\mathrm{Re}\{C_{uW}^{(33)}\}$
		ObW	c_{bW}	$\mathrm{Re}\{C_{dW}^{(33)}\}$
		OtZ	c_{tZ}	$\mathrm{Re}\{-s_W C_{uB}^{(33)} + c_W C_{uW}^{(33)}\}$
		Off	$c_{\varphi tb}$	$\mathrm{Re}\{C_{\varphi ud}^{(33)}\}$
		Ofq3	$c^3_{\varphi Q}$	$C_{\varphi q}^{3(33)}$
		OpQM	$c^-_{\varphi Q}$	$C_{\varphi q}^{1(33)} - C_{\varphi q}^{3(33)}$
		Opt	$c_{\varphi t}$	$C_{\varphi u}^{(33)}$
		Otp	$c_{t\varphi}$	$\mathrm{Re}\{C_{u\varphi}^{(33)}\}$

The top quark sector of the SMEFT

Notation	Sensitivity at $\mathcal{O}(\Lambda^{-2})$ ($\mathcal{O}(\Lambda^{-4})$)								
	$t\bar{t}$	single-top	tW	tZ	$t\bar{t}W$	$t\bar{t}Z$	$t\bar{t}H$	$t\bar{t}t\bar{t}$	$t\bar{t}b\bar{b}$
OQQ1								✓	✓
OQQ8								✓	✓
OQt1								✓	✓
OQt8								✓	✓
OQb1								(✓)	✓
OQb8								(✓)	✓
Ott1								✓	✓
Otb1								(✓)	✓
Otb8								✓	✓
OQtQb1									
OQtQb8									
O81qq	✓				✓	✓	✓	✓	✓
O11qq	✓				(✓)	(✓)	(✓)	✓	✓
O83qq	✓	✓		(✓)	✓	✓	✓	✓	✓
O13qq	✓	✓		✓	(✓)	(✓)	(✓)	✓	✓
O8qt	✓				✓	✓	✓	✓	✓
O1qt	✓				(✓)	(✓)	(✓)	✓	✓
O8ut	✓					✓	✓	✓	✓
O1ut	✓					(✓)	(✓)	✓	✓
O8qu	✓					✓	✓	✓	✓
O1qu	✓					(✓)	(✓)	✓	✓
O8dt	✓					✓	✓	✓	✓
O1dt	✓					(✓)	(✓)	✓	✓
O8qd	✓					✓	✓	✓	✓
O1qd	✓					(✓)	(✓)	✓	✓
OtG	✓				✓	✓	✓	✓	✓
OtW		✓	✓	✓					
ObW		(✓)	(✓)						
OtZ				✓		✓			
Off		(✓)	(✓)	(✓)					
Ofq3		✓	✓	✓		✓			
OpQM				✓		✓			
Opt				✓		✓	✓		
Otp							✓		

A large number of different dimension-6 SMEFT operators modify **top production at LHC**

$$\sigma_i^{\rm th}\left(\{c_n\}\right) = \sigma_{{\rm SM},i} + \sum_{n=1}^{N_{\rm op}} \tilde{\sigma}_{i,n}\frac{c_n}{\Lambda^2} + \sum_{n,m=1}^{N_{\rm op}} \tilde{\sigma}_{i,nm}\frac{c_n c_m}{\Lambda^4}$$

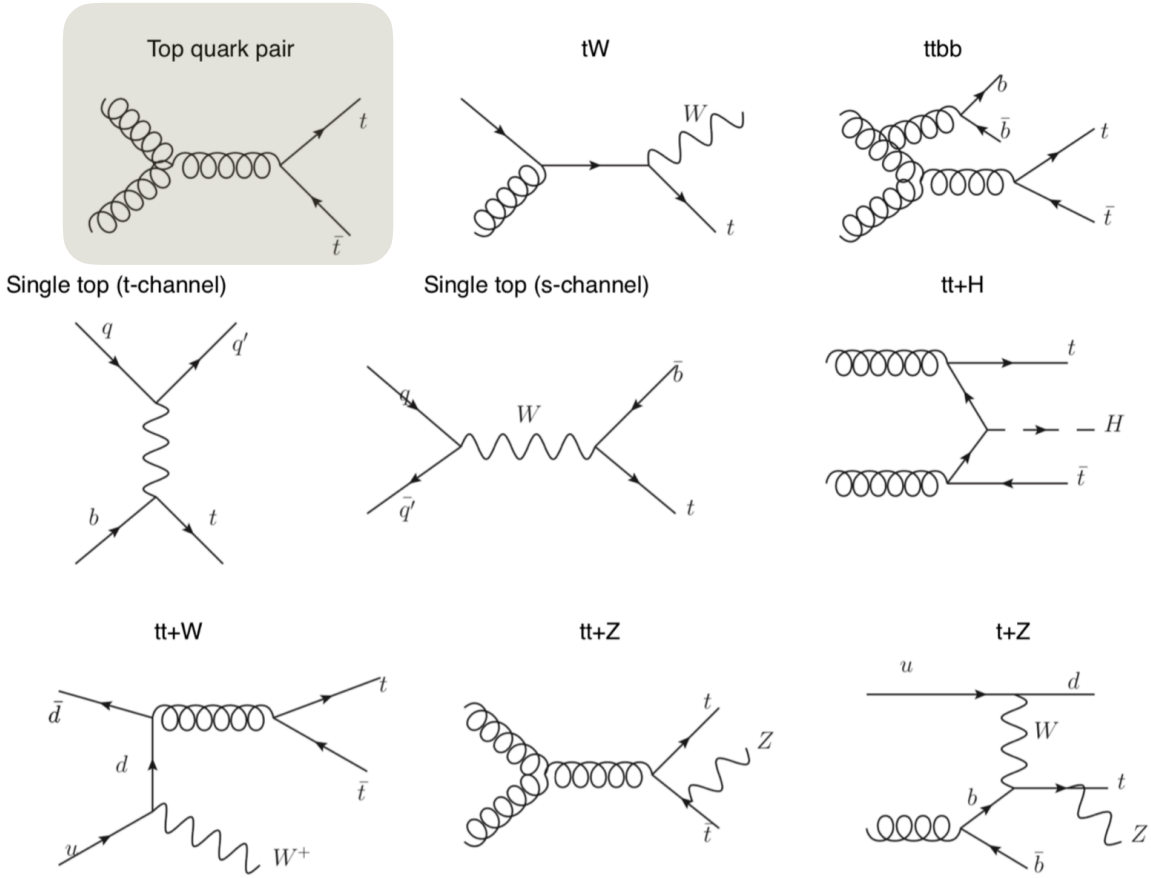


The top quark sector of the SMEFT

Notation	Sensitivity at $\mathcal{O}(\Lambda^{-2})$ ($\mathcal{O}(\Lambda^{-4})$)								
	$t\bar{t}$	single-top	tW	tZ	$t\bar{t}W$	$t\bar{t}Z$	$t\bar{t}H$	$t\bar{t}t\bar{t}$	$t\bar{t}b\bar{b}$
OQQ1								✓	✓
OQQ8								✓	✓
OQt1								✓	✓
OQt8								✓	✓
OQb1								(✓)	✓
OQb8								(✓)	✓
Ott1								✓	✓
Otb1								(✓)	✓
Otb8								✓	✓
OQtQb1									
OQtQb8									
O81qq	✓				✓	✓	✓	✓	✓
O11qq	✓				(✓)	(✓)	(✓)	✓	✓
O83qq	✓	✓		(✓)	✓	✓	✓	✓	✓
O13qq	✓	✓		✓	(✓)	(✓)	(✓)	✓	✓
O8qt	✓				✓	✓	✓	✓	✓
O1qt	✓				(✓)	(✓)	(✓)	✓	✓
O8ut	✓					✓	✓	✓	✓
O1ut	✓					(✓)	(✓)	✓	✓
O8qu	✓					✓	✓	✓	✓
O1qu	✓					(✓)	(✓)	✓	✓
O8dt	✓					✓	✓	✓	✓
O1dt	✓					(✓)	(✓)	✓	✓
O8qd	✓					✓	✓	✓	✓
O1qd	✓					(✓)	(✓)	✓	✓
OtG	✓				✓	✓	✓	✓	✓
OtW		✓	✓	✓					
ObW		(✓)	(✓)						
OtZ				✓		✓			
Off		(✓)	(✓)	(✓)					
Ofq3		✓	✓	✓		✓			
OpQM				✓		✓			
Opt				✓		✓	✓		
Otp							✓		

A large number of different dimension-6 SMEFT operators modify **top production at LHC**

$$\sigma_i^{\rm th}\left(\{c_n\}\right) = \sigma_{{\rm SM},i} + \sum_{n=1}^{N_{\rm op}} \tilde{\sigma}_{i,n}\frac{c_n}{\Lambda^2} + \sum_{n,m=1}^{N_{\rm op}} \tilde{\sigma}_{i,nm}\frac{c_n c_m}{\Lambda^4}$$

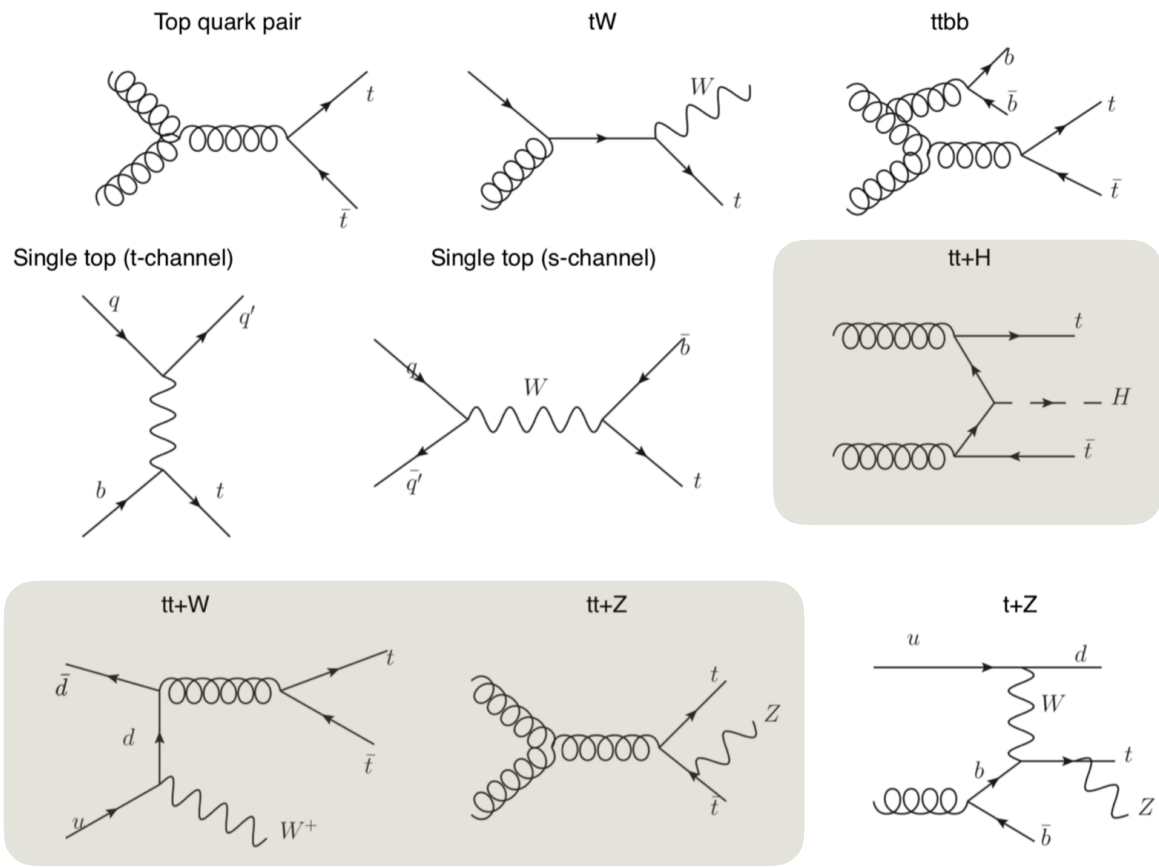


The top quark sector of the SMEFT

Notation	Sensitivity at $\mathcal{O}(\Lambda^{-2})$ ($\mathcal{O}(\Lambda^{-4})$)								
	$t\bar{t}$	single-top	tW	tZ	$t\bar{t}W$	$t\bar{t}Z$	$t\bar{t}H$	$t\bar{t}t\bar{t}$	$t\bar{t}b\bar{b}$
OQQ1								✓	✓
OQQ8								✓	✓
OQt1								✓	✓
OQt8								✓	✓
OQb1								(✓)	✓
OQb8								(✓)	✓
Ott1								✓	✓
Otb1								(✓)	✓
Otb8								✓	✓
OQtQb1									
OQtQb8									
O81qq	✓				✓	✓	✓	✓	✓
O11qq	✓				(✓)	(✓)	(✓)	✓	✓
O83qq	✓	✓		(✓)	✓	✓	✓	✓	✓
O13qq	✓	✓		✓	(✓)	(✓)	(✓)	✓	✓
O8qt	✓				✓	✓	✓	✓	✓
O1qt	✓				(✓)	(✓)	(✓)	✓	✓
O8ut	✓					✓	✓	✓	✓
O1ut	✓					(✓)	(✓)	✓	✓
O8qu	✓					✓	✓	✓	✓
O1qu	✓					(✓)	(✓)	✓	✓
O8dt	✓					✓	✓	✓	✓
O1dt	✓					(✓)	(✓)	✓	✓
O8qd	✓					✓	✓	✓	✓
O1qd	✓					(✓)	(✓)	✓	✓
OtG	✓				✓	✓	✓	✓	✓
OtW		✓	✓	✓					
ObW		(✓)	(✓)						
OtZ				✓		✓			
Off		(✓)	(✓)	(✓)					
Ofq3		✓	✓	✓		✓			
OpQM				✓		✓			
Opt				✓		✓	✓		
Otp							✓		

A large number of different dimension-6 SMEFT operators modify **top production at LHC**

$$\sigma_i^{\rm th}\left(\{c_n\}\right) = \sigma_{{\rm SM},i} + \sum_{n=1}^{N_{\rm op}} \tilde{\sigma}_{i,n}\frac{c_n}{\Lambda^2} + \sum_{n,m=1}^{N_{\rm op}} \tilde{\sigma}_{i,nm}\frac{c_n c_m}{\Lambda^4}$$

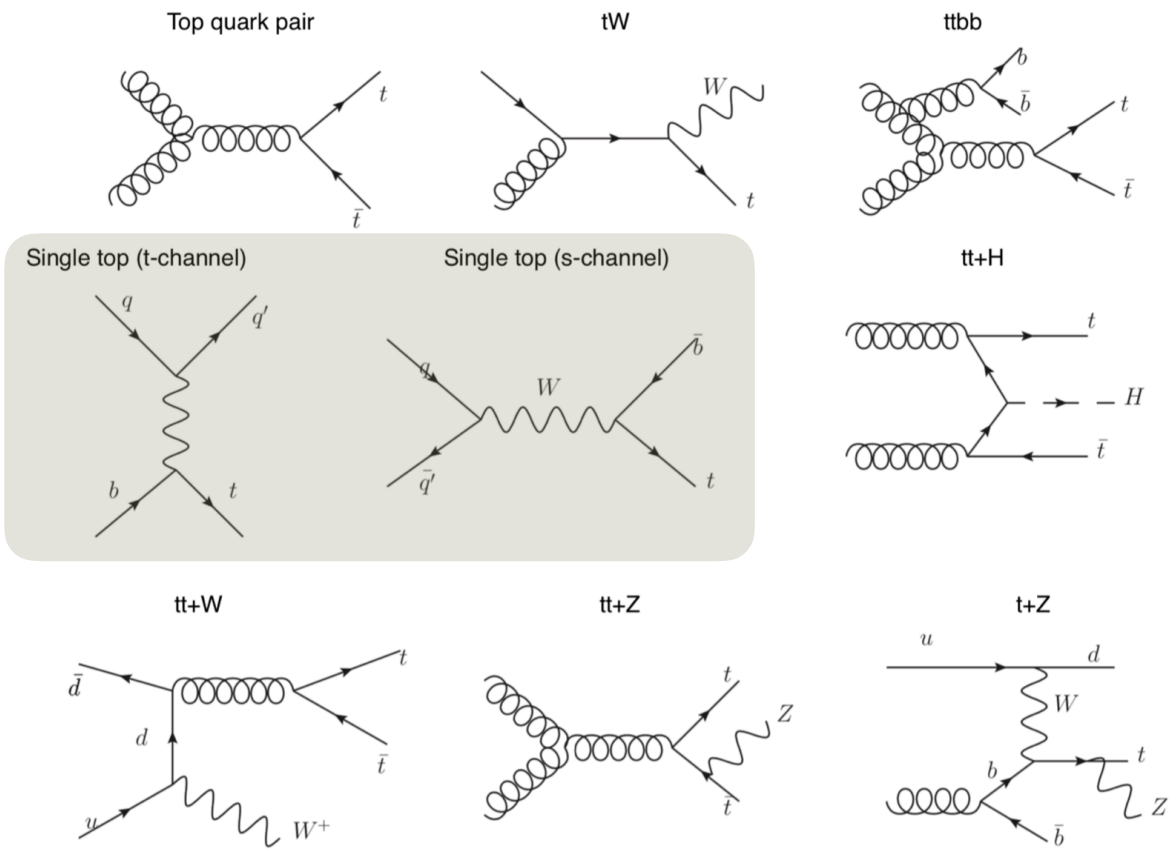


The top quark sector of the SMEFT

Notation	Sensitivity at $\mathcal{O}(\Lambda^{-2})$ ($\mathcal{O}(\Lambda^{-4})$)								
	$t\bar{t}$	single-top	tW	tZ	$t\bar{t}W$	$t\bar{t}Z$	$t\bar{t}H$	$t\bar{t}t\bar{t}$	$t\bar{t}b\bar{b}$
OQQ1								✓	✓
OQQ8								✓	✓
OQt1								✓	✓
OQt8								✓	✓
OQb1								(✓)	✓
OQb8								(✓)	✓
Ott1								✓	✓
Otb1								(✓)	✓
Otb8								✓	✓
OQtQb1									
OQtQb8									
O81qq	✓				✓	✓	✓	✓	✓
O11qq	✓				(✓)	(✓)	(✓)	✓	✓
O83qq	✓	✓		(✓)	✓	✓	✓	✓	✓
O13qq	✓	✓		✓	(✓)	(✓)	(✓)	✓	✓
O8qt	✓				✓	✓	✓	✓	✓
O1qt	✓				(✓)	(✓)	(✓)	✓	✓
O8ut	✓					✓	✓	✓	✓
O1ut	✓					(✓)	(✓)	✓	✓
O8qu	✓					✓	✓	✓	✓
O1qu	✓					(✓)	(✓)	✓	✓
O8dt	✓					✓	✓	✓	✓
O1dt	✓					(✓)	(✓)	✓	✓
O8qd	✓					✓	✓	✓	✓
O1qd	✓					(✓)	(✓)	✓	✓
OtG	✓				✓	✓	✓	✓	✓
OtW		✓	✓	✓					
ObW		(✓)	(✓)						
OtZ				✓		✓			
Off		(✓)	(✓)	(✓)					
Ofq3		✓	✓	✓		✓			
OpQM				✓		✓			
Opt				✓		✓	✓		
Otp							✓		

A large number of different dimension-6 SMEFT operators modify **top production at LHC**

$$\sigma_i^{\rm th}\left(\{c_n\}\right) = \sigma_{{\rm SM},i} + \sum_{n=1}^{N_{\rm op}} \tilde{\sigma}_{i,n}\frac{c_n}{\Lambda^2} + \sum_{n,m=1}^{N_{\rm op}} \tilde{\sigma}_{i,nm}\frac{c_n c_m}{\Lambda^4}$$



Input dataset (I)

Process	Dataset	$\sqrt{s}$	Info	Observables	$N_{\rm dat}$	Ref
$t\bar{t}$	ATLAS_tt_8TeV_ljets	**8 TeV**	lepton+jets	$d\sigma/d\lvert y_t\rvert$, $d\sigma/dp_t^T$, $d\sigma/dm_{t\bar{t}}$, $d\sigma/d\lvert y_{t\bar{t}}\rvert$	5, 8, 7, 5	[77]
$t\bar{t}$	CMS_tt_8TeV_ljets	**8 TeV**	lepton+jets	$d\sigma/dy_t$, $d\sigma/dp_t^T$, $d\sigma/dm_{t\bar{t}}$, $d\sigma/dy_{t\bar{t}}$	10, 8, 7, 10	[78]
$t\bar{t}$	CMS_tt2D_8TeV_dilep	**8 TeV**	dileptons	$d^2\sigma/dy_t dp_t^T$, $d^2\sigma/dy_t dm_{t\bar{t}}$, $d^2\sigma/dp_{t\bar{t}}^T dm_{t\bar{t}}$, $d^2\sigma/dy_{t\bar{t}} dm_{t\bar{t}}$	16, 16, 16, 16	[79]
$t\bar{t}$	CMS_tt_13TeV_ljets	**13 TeV**	lepton+jets	$d\sigma/d\lvert y_t\rvert$, $d\sigma/dp_t^T$, $d\sigma/dm_{t\bar{t}}$, $d\sigma/d\lvert y_{t\bar{t}}\rvert$	7, 9, 8, 6	[83]
$t\bar{t}$	CMS_tt_13TeV_ljets2	**13 TeV**	lepton+jets	$d\sigma/d\lvert y_t\rvert$, $d\sigma/dp_t^T$, $d\sigma/dm_{t\bar{t}}$, $d\sigma/d\lvert y_{t\bar{t}}\rvert$	11, 12, 10, 10	[85]
$t\bar{t}$	CMS_tt_13TeV_dilep	**13 TeV**	dileptons	$d\sigma/dy_t$, $d\sigma/dp_t^T$, $d\sigma/dm_{t\bar{t}}$, $d\sigma/dy_{t\bar{t}}$	8, 6, 6, 8	[86]
$t\bar{t}$	ATLASCMS_AcMtt_8TeV	**8 TeV**	Asymm comb	$A_C(m_{t\bar{t}})$, Eq. (3.1)	6	[80]
$t\bar{t}$	ATLAS_WhelF_8TeV	**8 TeV**	W helicity fract	F_0, F_L, F_R	3	[81]
$t\bar{t}$	CMS_WhelF_8TeV	**8 TeV**	W helicity fract	F_0, F_L, F_R	3	[82]

Input dataset (II)

Process	Dataset	$\sqrt{s}$	Info	Observables	$N_{\rm dat}$	Ref
Single t	CMS_t_tch_8TeV_inc	**8 TeV**	t-channel	$\sigma_{\rm tot}(t), \sigma_{\rm tot}(\bar{t})$ (R_t)	2 (1)	[95]
Single t	CMS_t_sch_8TeV	**8 TeV**	s-channel	$\sigma_{\rm tot}(t+\bar{t})$	1	[96]
Single t	ATLAS_t_sch_8TeV	**8 TeV**	s-channel	$\sigma_{\rm tot}(t+\bar{t})$	1	[97]
Single t	ATLAS_t_tch_8TeV	**8 TeV**	t-channel	$d\sigma(tq)/dp_T^t, d\sigma(\bar{t}q)/dp_T^{\bar{t}}$ $d\sigma(tq)/dy_t, d\sigma(\bar{t}q)/dy_t$	5, 4 4, 4	[98]
Single t	ATLAS_t_tch_13TeV	**13 TeV**	t-channel	$\sigma_{\rm tot}(t), \sigma_{\rm tot}(\bar{t})$ (R_t)	2 (1)	[99]
Single t	CMS_t_tch_13TeV_inc	**13 TeV**	t-channel	$\sigma_{\rm tot}(t+\bar{t})$ (R_t)	1 (1)	[100]
Single t	CMS_t_tch_8TeV_dif	**8 TeV**	t-channel	$d\sigma/dp_T^{(t+\bar{t})}$, $d\sigma/d\lvert y^{(t+\bar{t})}\rvert$	6 6	[101]
Single t	CMS_t_tch_13TeV_dif	**13 TeV**	t-channel	$d\sigma/dp_T^{(t+\bar{t})}$, $d\sigma/d\lvert y^{(t+\bar{t})}\rvert$	4 4	[102]
tW	ATLAS_tW_inc_8TeV	**8 TeV**	inclusive	$\sigma_{\rm tot}(tW)$	1	[103]
tW	CMS_tW_inc_8TeV	**8 TeV**	inclusive	$\sigma_{\rm tot}(tW)$	1	[104]
tW	ATLAS_tW_inc_13TeV	**13 TeV**	inclusive	$\sigma_{\rm tot}(tW)$	1	[105]
tW	CMS_tW_inc_13TeV	**13 TeV**	inclusive	$\sigma_{\rm tot}(tW)$	1	[106]
tZ	CMS_tZ_inc_13TeV	**13 TeV**	inclusive	$\sigma_{\rm fid}(Wbl^+l^-q)$	1	[107]
tZ	ATLAS_tZ_inc_13TeV	**13 TeV**	inclusive	$\sigma_{\rm tot}(tZq)$	1	[108]

Input dataset (III)

Process	Dataset	$\sqrt{s}$	Info	Observables	$N_{\rm dat}$	Ref
$t\bar{t}b\bar{b}$	CMS_ttbb_13TeV	**13 TeV**	total xsec	$\sigma_{\rm tot}(t\bar{t}b\bar{b})$	1	[87]
$t\bar{t}t\bar{t}$	CMS_tttt_13TeV	**13 TeV**	total xsec	$\sigma_{\rm tot}(t\bar{t}t\bar{t})$	1	[88]
$t\bar{t}Z$	CMS_ttZ_8_13TeV	**8+13 TeV**	total xsec	$\sigma_{\rm tot}(t\bar{t}Z)$	2	[89, 90]
$t\bar{t}Z$	ATLAS_ttZ_8_13TeV	**8+13 TeV**	total xsec	$\sigma_{\rm tot}(t\bar{t}Z)$	2	[91, 92]
$t\bar{t}W$	CMS_ttW_8_13TeV	**8+13 TeV**	total xsec	$\sigma_{\rm tot}(t\bar{t}W)$	2	[89, 90]
$t\bar{t}W$	ATLAS_ttW_8_13TeV	**8+13 TeV**	total xsec	$\sigma_{\rm tot}(t\bar{t}W)$	2	[91, 92]
$t\bar{t}H$	CMS_tth_13TeV	**13 TeV**	signal strength	$\mu_{t\bar{t}H}$	1	[93]
$t\bar{t}H$	ATLAS_tth_13TeV	**13 TeV**	total xsec	$\sigma_{\rm tot}(t\bar{t}H)$	1	[94]

The fit includes more than **100 cross-section measurement**s at 8 and 13 TeV from **10 different top-quark production processes**

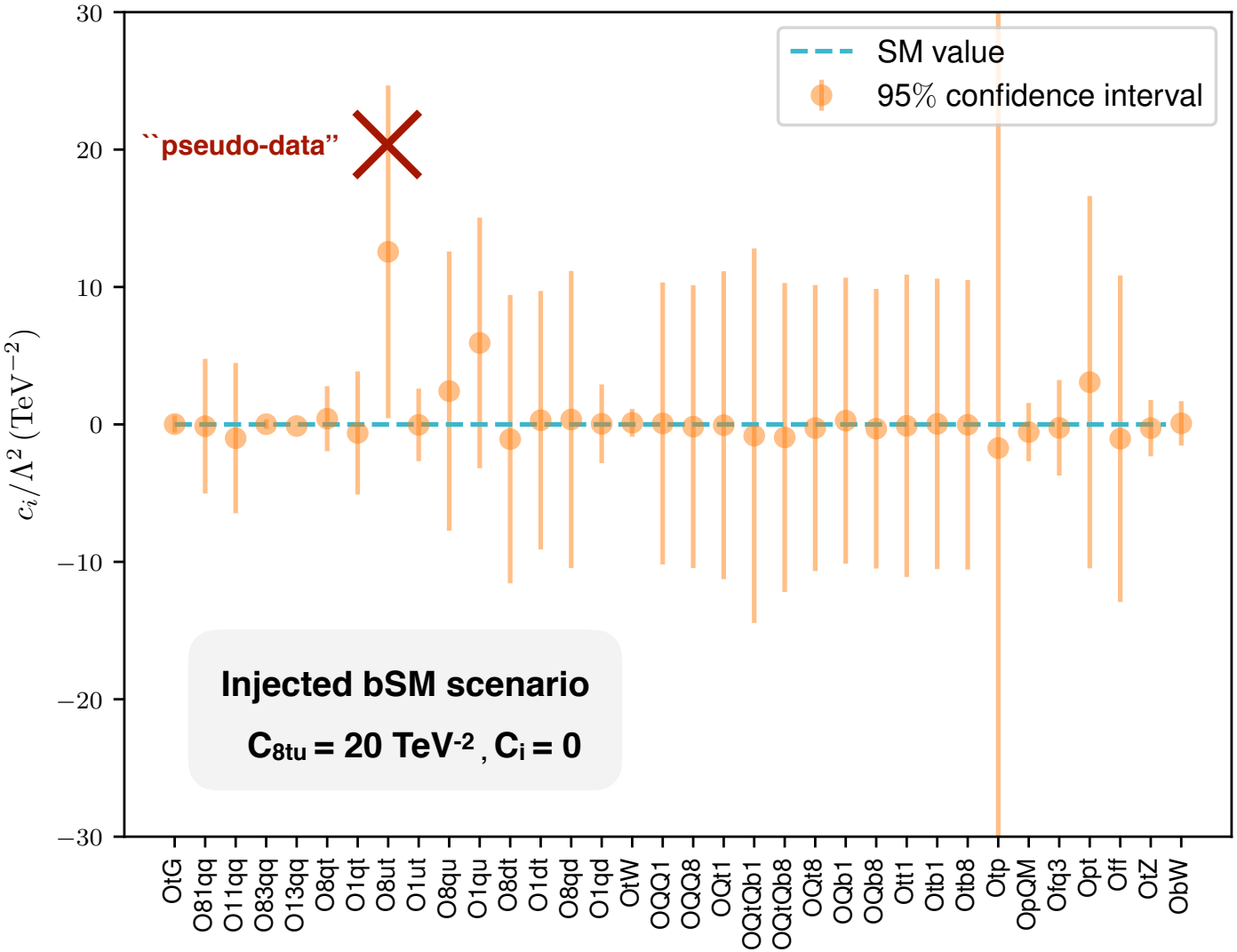
Theory calculations

Process	SM	Code	SMEFT	Code
$t\bar{t}$	NNLO QCD	MCFM/SHERPA NLO + NNLO K-factors	NLO QCD	MG5_aMC
single-t (t-ch)	NNLO QCD	MCFM NLO + NNLO K-factors	NLO QCD	MG5_aMC
single-t (s-ch)	NLO QCD	MCFM	NLO QCD	MG5_aMC
tW	NLO QCD	MG5_aMC	NLO QCD	MG5_aMC
tZ	NLO QCD	MG5_aMC	LO QCD + NLO SM K-factors	MG5_aMC
$t\bar{t}W(Z)$	NLO QCD	MG5_aMC	LO QCD + NLO SM K-factors	MG5_aMC
$t\bar{t}h$	NLO QCD	MG5_aMC	LO QCD + NLO SM K-factors	MG5_aMC
$t\bar{t}t\bar{t}$	NLO QCD	MG5_aMC	LO QCD + NLO SM K-factors	MG5_aMC
$t\bar{t}b\bar{b}$	NLO QCD	MG5_aMC	LO QCD + NLO SM K-factors	MG5_aMC

PDF set: **NNPDF3.1 NNLO no-top**

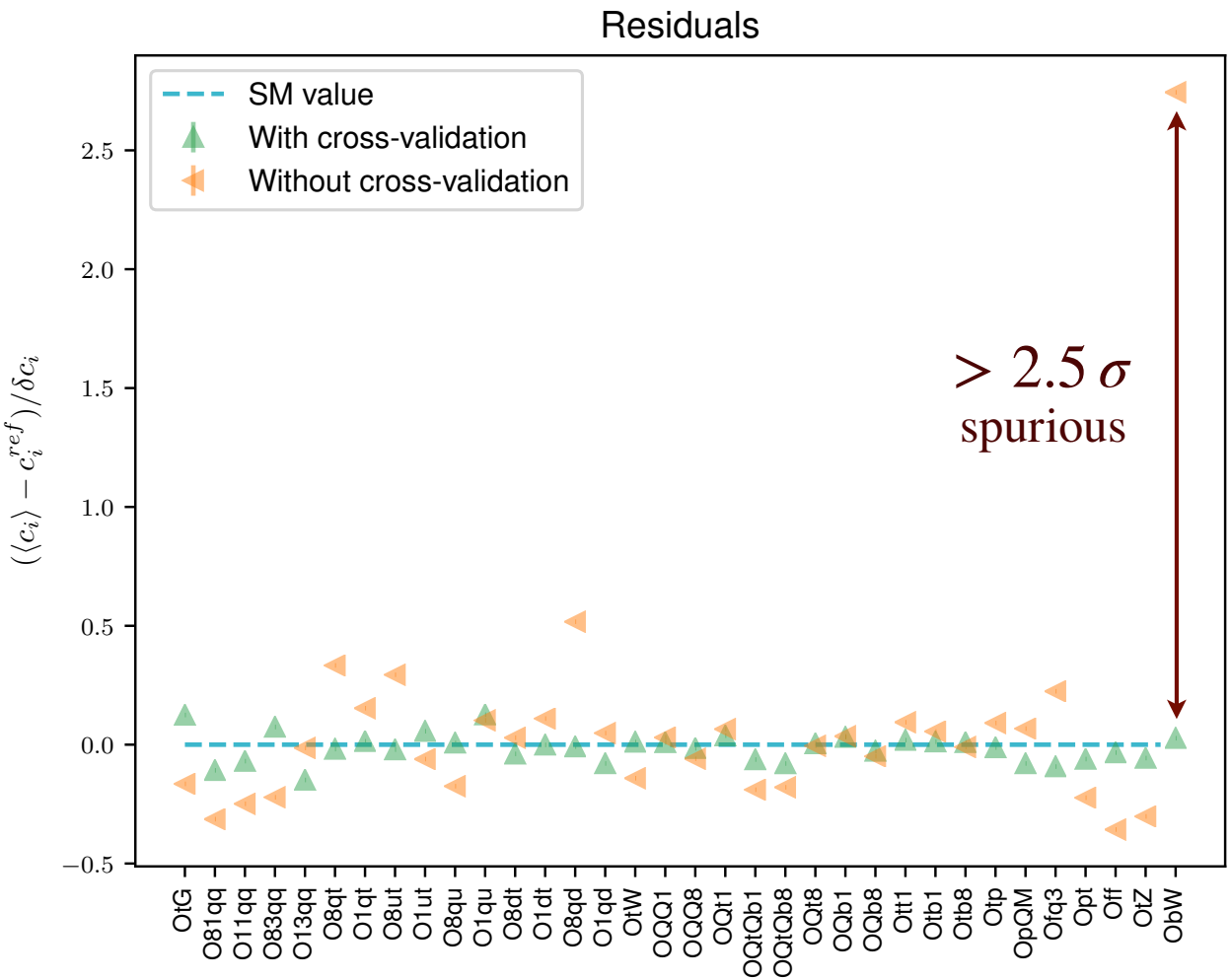
Closure Tests

- Generate **pseudo-data based** on a given scenario (SM or BSM) and check that the correct (known) results are reproduced after the fit
- Allows quantifying the **expected statistical significance** for BSM deviations



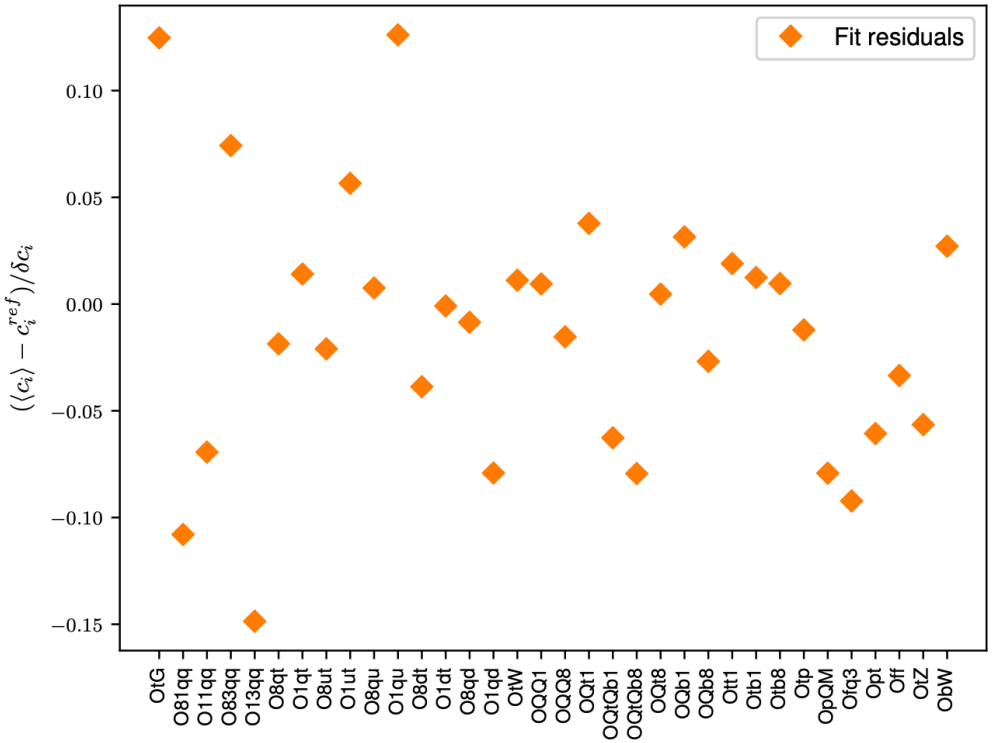
Cross-validation

- Since N_{par} is not too different from N_{dat}, overfitting will take place for an efficient optimiser
- **Artificial tensions with the SM** are likely to be generated by overfitting!
- Fit residuals **consistent with true result (SM)** only with cross-validation

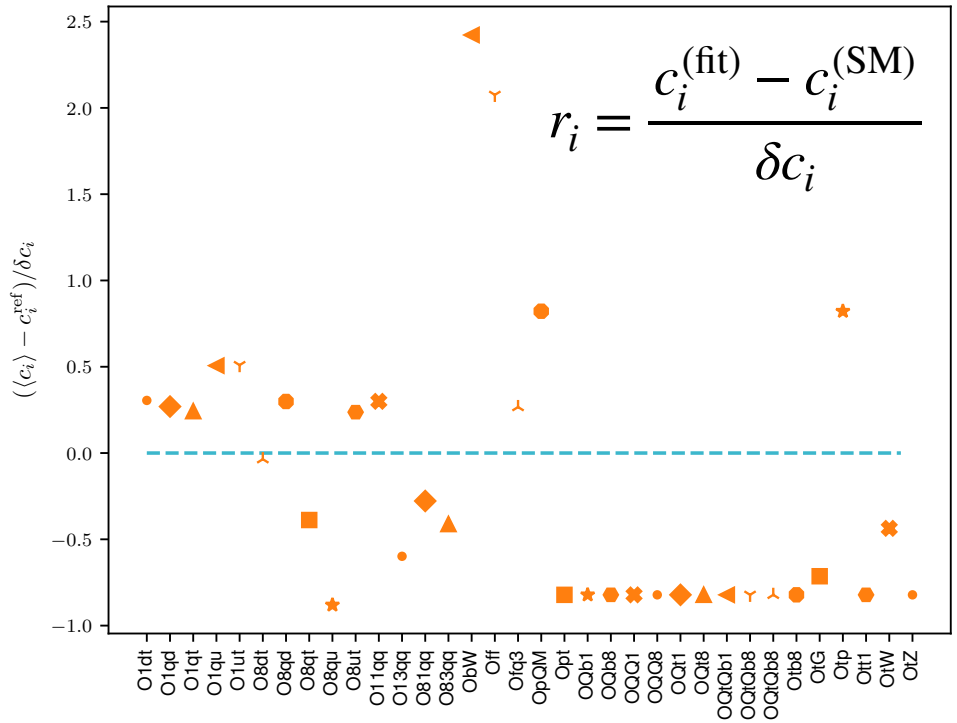


Global vs individual fits

Global fits (all params fitted)



Individual fits (one param at a time)



- If each operator was a truly **independent random variable**, we would expect that at least **2 operators** have residuals **|r| > 1** (bounds are 95% CL)
- This is far from being the case when all operators are fitted simultaneously
- Explained by **correlations between operators + degeneracies in parameter space**: much larger fluctuations if we fit one operator at a time

Fit quality

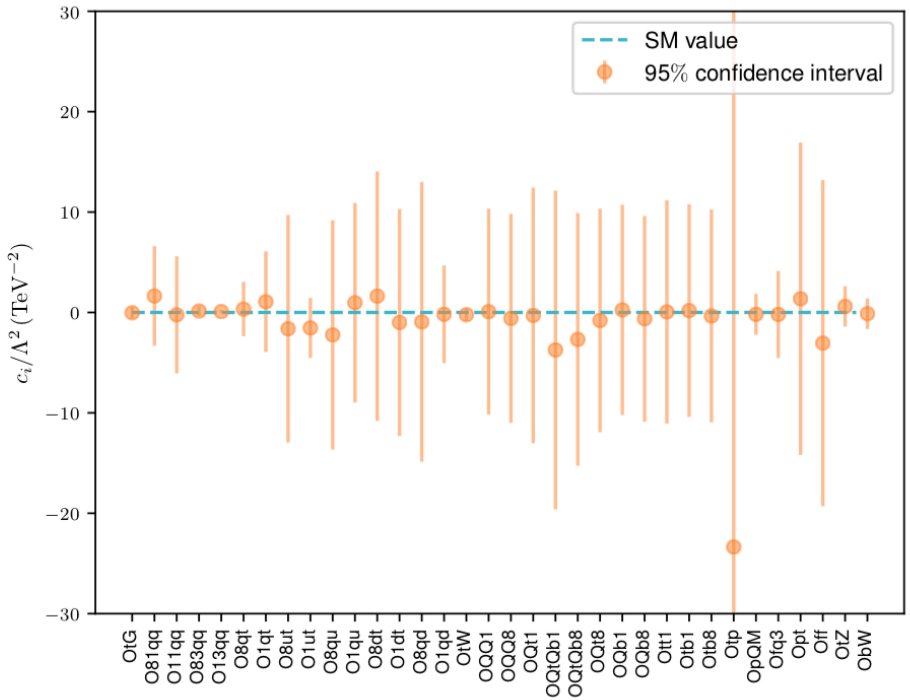
- **Good agreement** between theory (SM and SMEFT) and data for most datasets

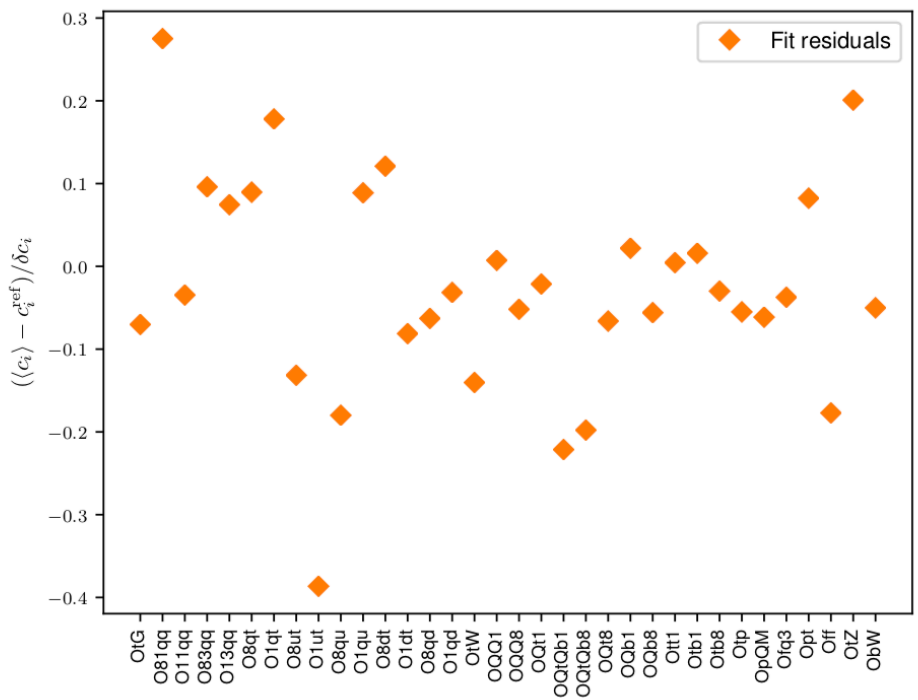
- For the **103 fitted cross-sections**, we find χ^2/n_{dat} *of* **1.11 (1.06)** before (after) fit

- Including SMEFT effects improves agreement with data: need to quantify how **significant** this improvement is

Dataset	$\chi^2/n_{\rm dat}$ (prior)	$\chi^2/n_{\rm dat}$ (fit)	$n_{\rm dat}$
ATLAS_tt_8TeV_ljets [$m_{t\bar{t}}$]	1.51	1.25	7
CMS_tt_8TeV_ljets [$y_{t\bar{t}}$]	1.17	1.17	10
CMS_tt2D_8TeV_dilep [$(m_{t\bar{t}}, y_t)$]	1.38	1.38	16
CMS_tt_13TeV_ljets2 [$m_{t\bar{t}}$]	1.09	1.28	8
CMS_tt_13TeV_dilep [$m_{t\bar{t}}$]	1.34	1.42	6
CMS_tt_13TeV_ljets_2016 [$m_{t\bar{t}}$]	1.87	1.87	10
ATLAS_WhelF_8TeV	1.98	0.27	3
CMS_WhelF_8TeV	0.31	1.18	3
CMS_ttbb_13TeV	5.00	1.29	1
CMS_tttt_13TeV	0.05	0.02	1
ATLAS_tth_13TeV	1.61	0.55	1
CMS_tth_13TeV	0.34	0.01	1
ATLAS_ttZ_8TeV	1.32	5.29	1
ATLAS_ttZ_13TeV	0.01	1.06	1
CMS_ttZ_8TeV	0.04	0.06	1
CMS_ttZ_13TeV	0.90	0.67	1
ATLAS_ttW_8TeV	1.34	0.27	1
ATLAS_ttW_13TeV	0.82	0.65	1
CMS_ttW_8TeV	1.54	0.54	1
CMS_ttW_13TeV	0.03	0.09	1
CMS_t_tch_8TeV_dif	0.11	0.32	6
ATLAS_t_tch_8TeV [y_t]	0.91	0.43	4
ATLAS_t_tch_8TeV [$y_{\bar{t}}$]	0.39	0.45	4
ATLAS_t_sch_8TeV	0.08	1.92	1
ATLAS_t_tch_13TeV	0.02	0.09	2
CMS_t_tch_13TeV_dif [y_t]	0.46	0.49	4
CMS_t_sch_8TeV	1.26	0.76	1
ATLAS_tW_inc_8TeV	0.02	0.06	1
CMS_tW_inc_8TeV	0.00	0.07	1
ATLAS_tW_inc_13TeV	0.52	0.82	1
CMS_tW_inc_13TeV	4.29	1.68	1
ATLAS_tZ_inc_13TeV	0.00	0.00	1
CMS_tZ_inc_13TeV	0.66	0.34	1
Total	**1.11**	**1.06**	**103**

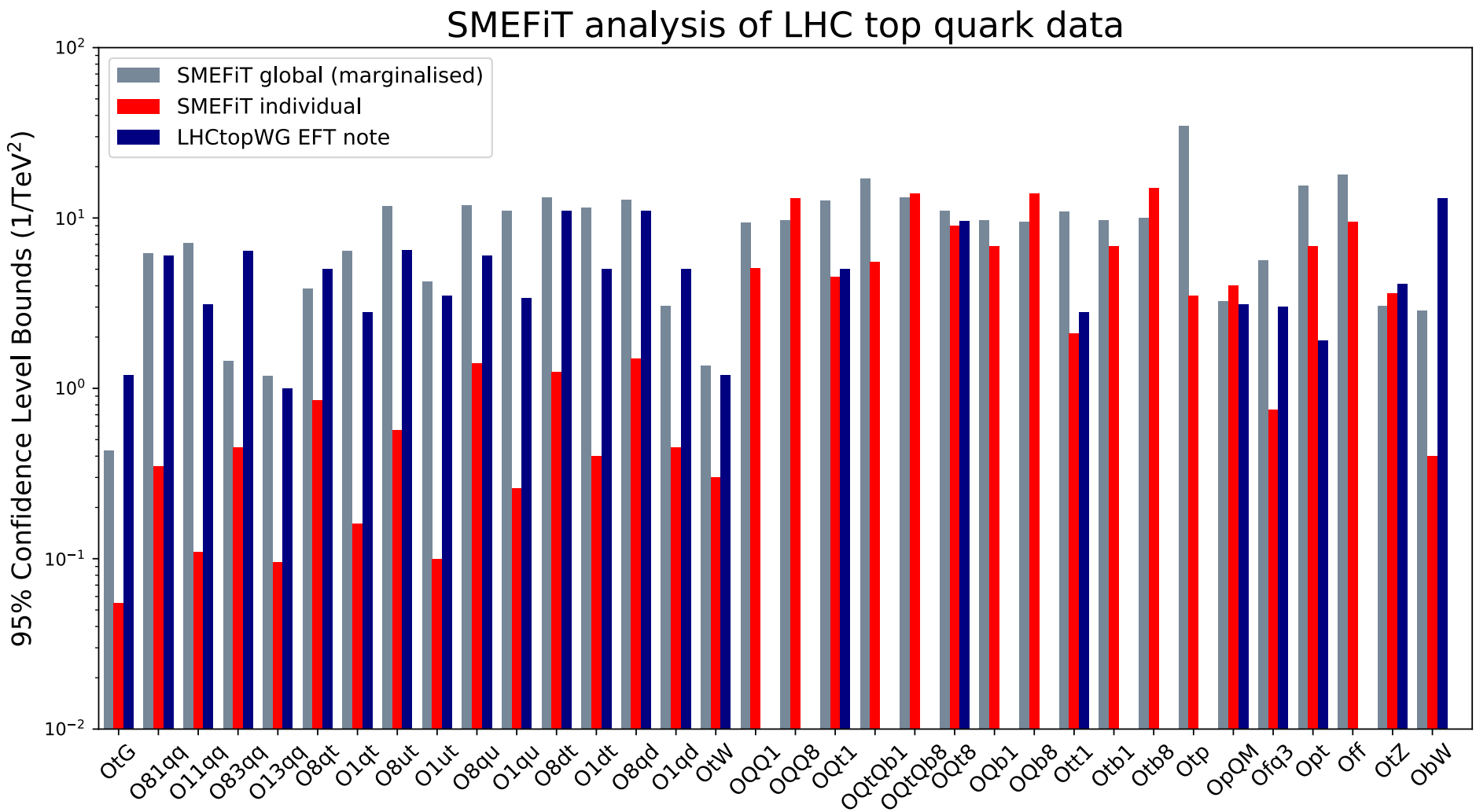
SMEFiT results





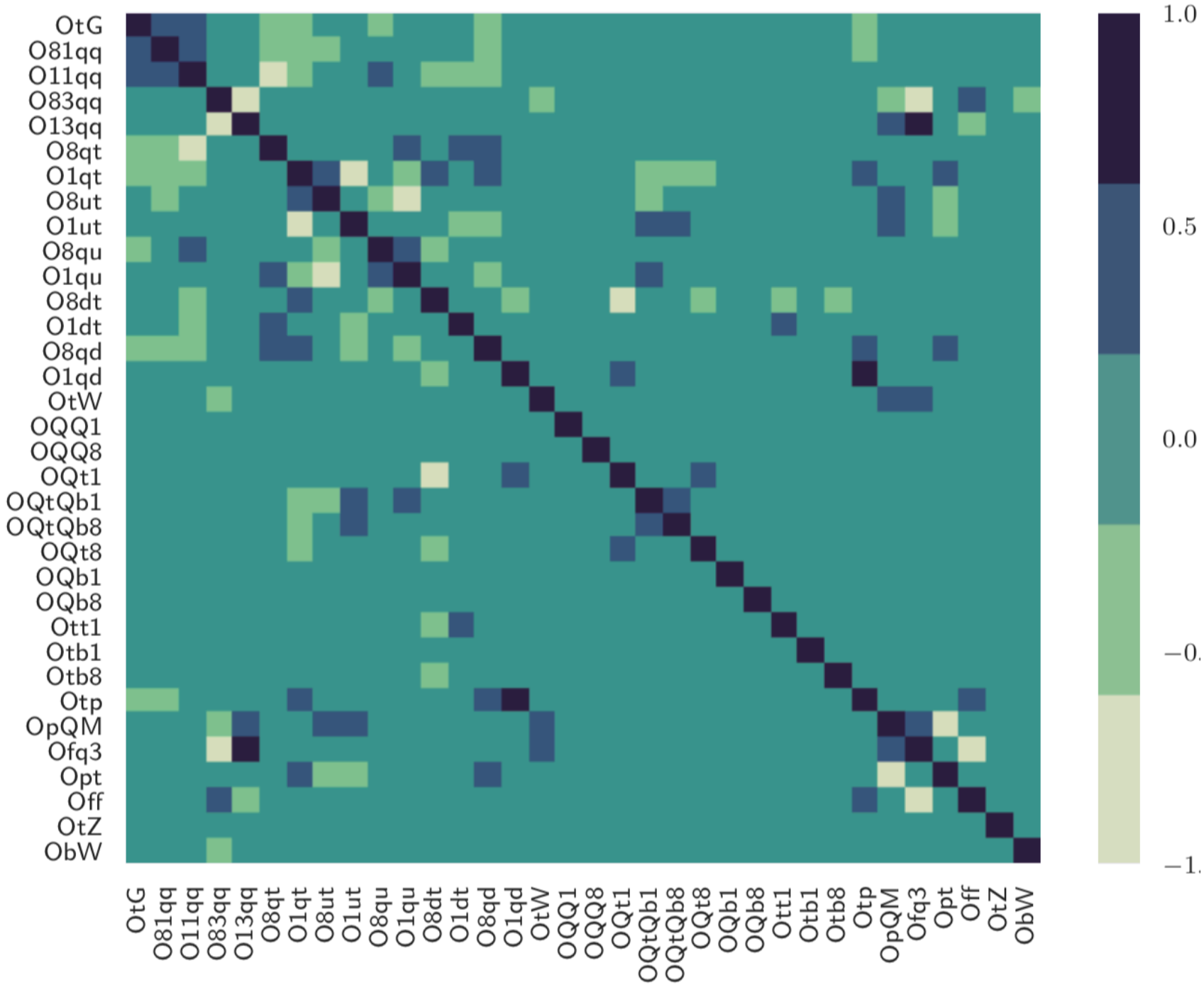
- Agreement with the SM expectation **within uncertainties**
- Bounds on individual operators are in general largely **correlated among them**
- Large differences between the bounds obtained from each operator

Comparison with 1D fits and previous bounds

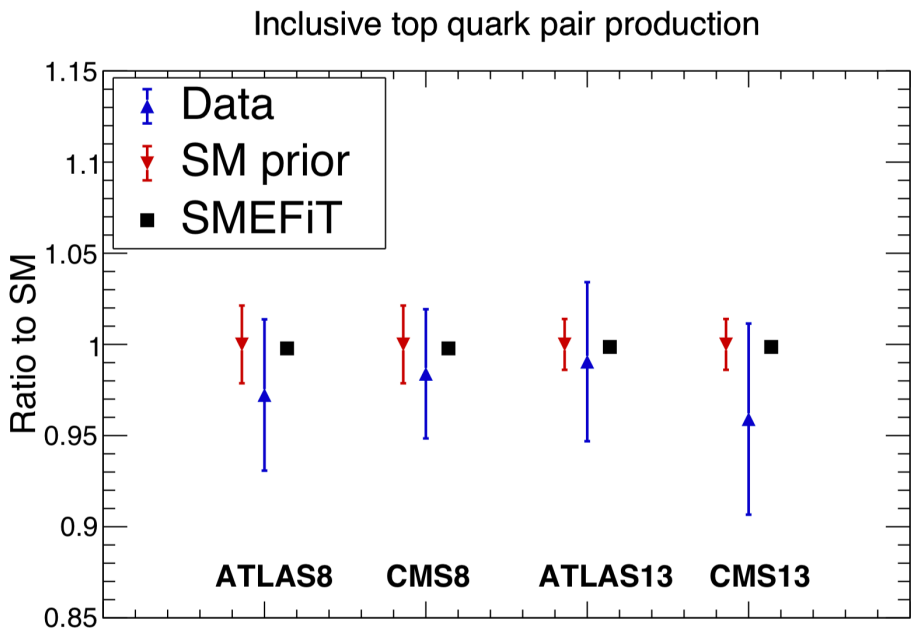


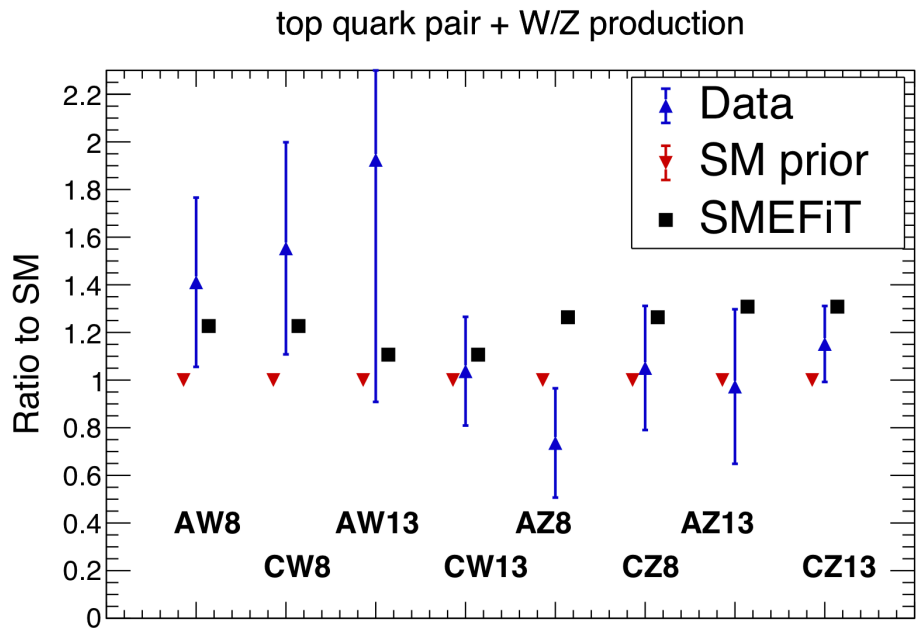
- **Improvement** found (more stringent bounds) in most fitted degrees of freedom
- For some specific operators **our bounds are the first ones** to be reported
- Individual bounds can dramatically **overestimate** the actual (marginalised) bounds

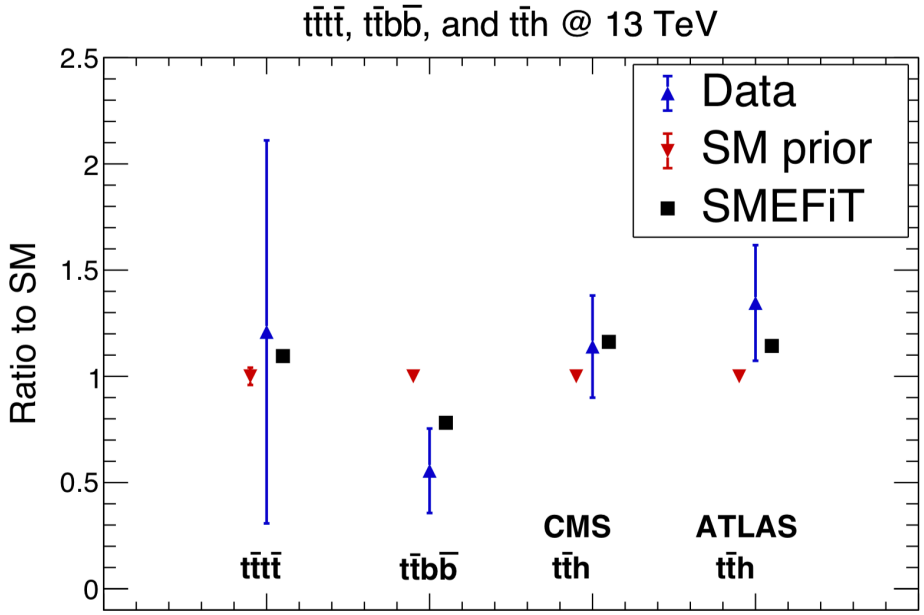
Correlation map



Comparison with data

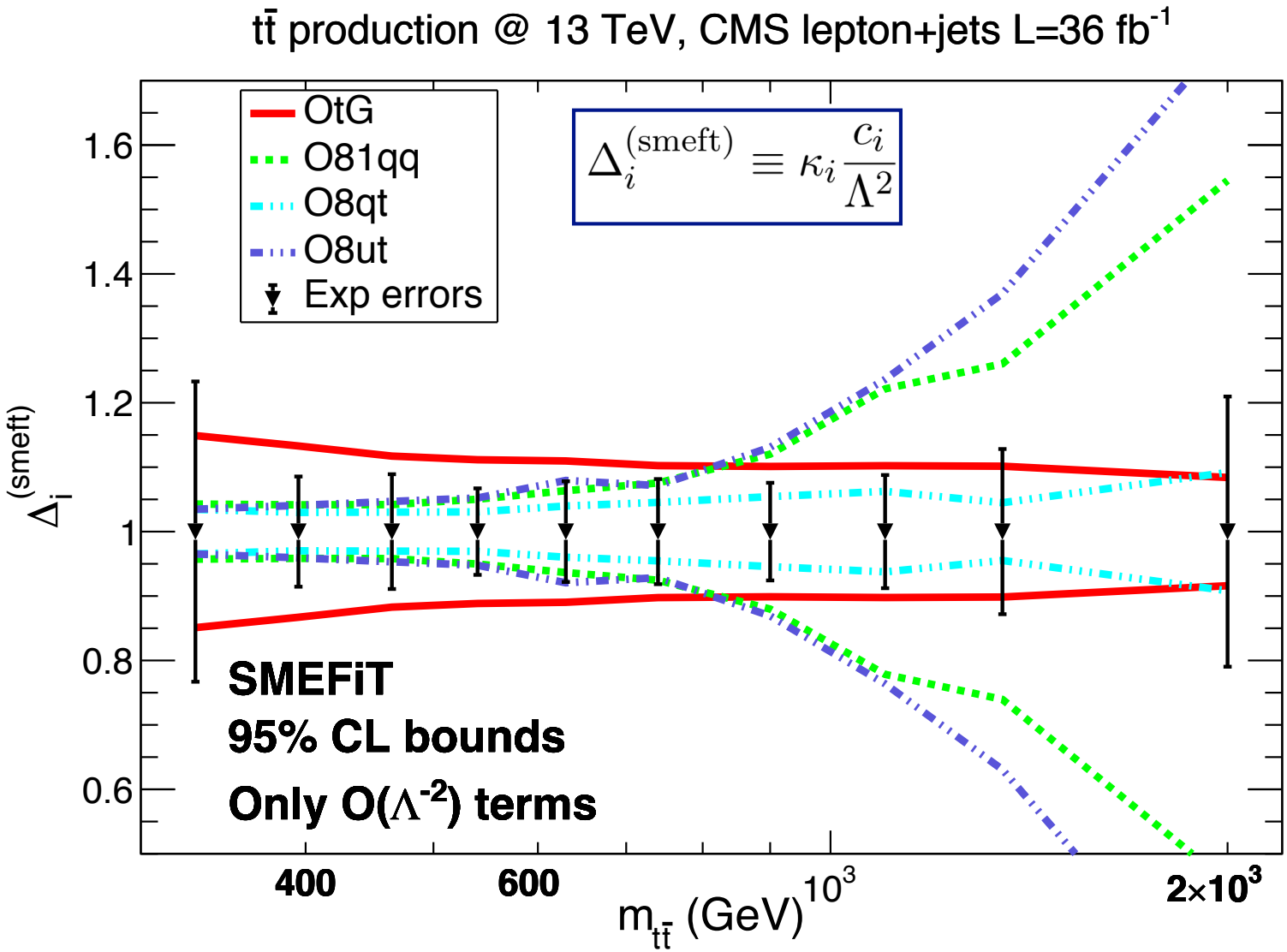






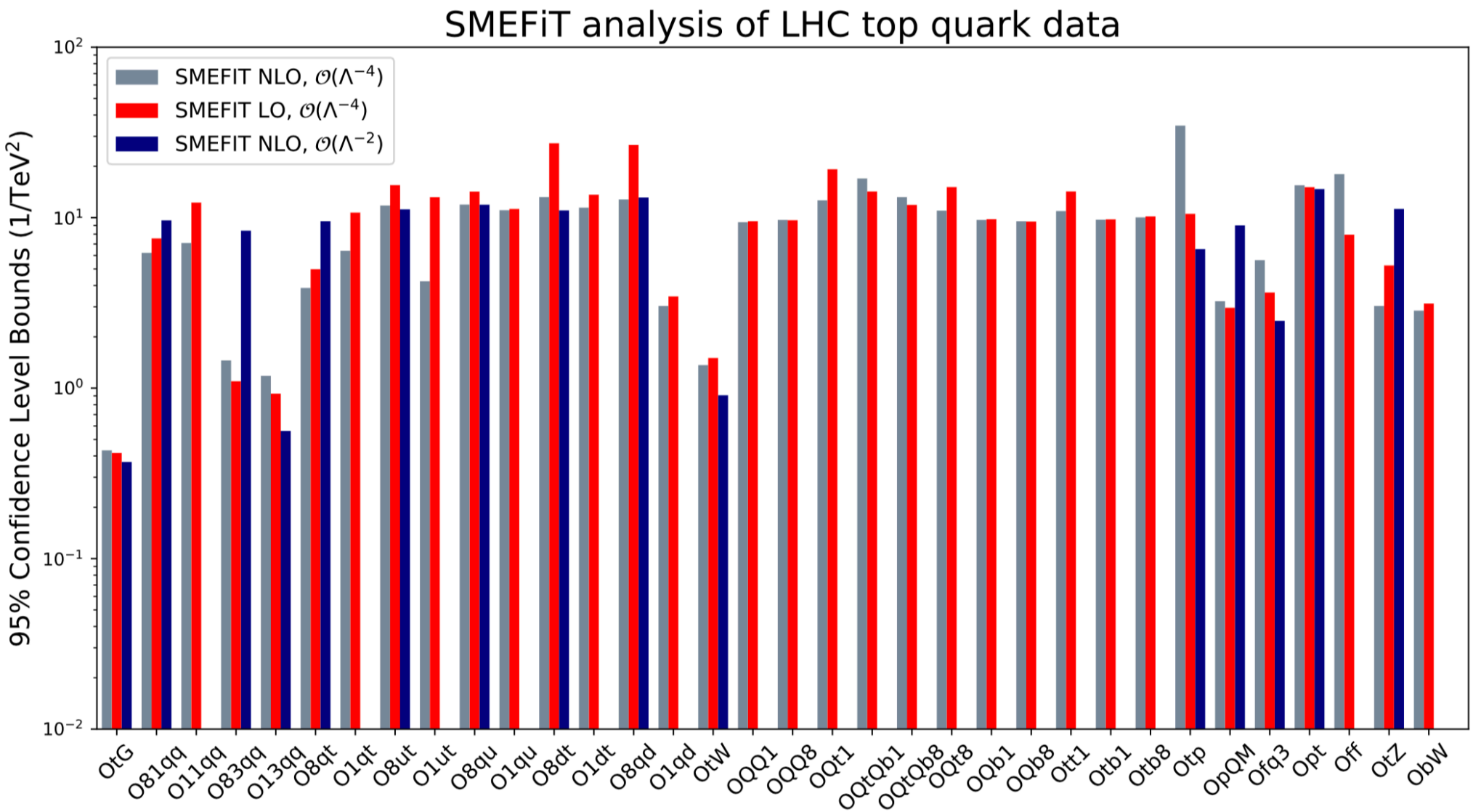
- The **best-fit SMEFT-induced shift** wrt the SM calculation depends on the process
- For inclusive top quark pair and single top, the SMEFT shifts are < 2%
- For ***tttt, ttbbb***, and ***tth*** the SMEFT shifts can be as large as 20% (reflecting the larger experimental errors)

High-energy behaviour



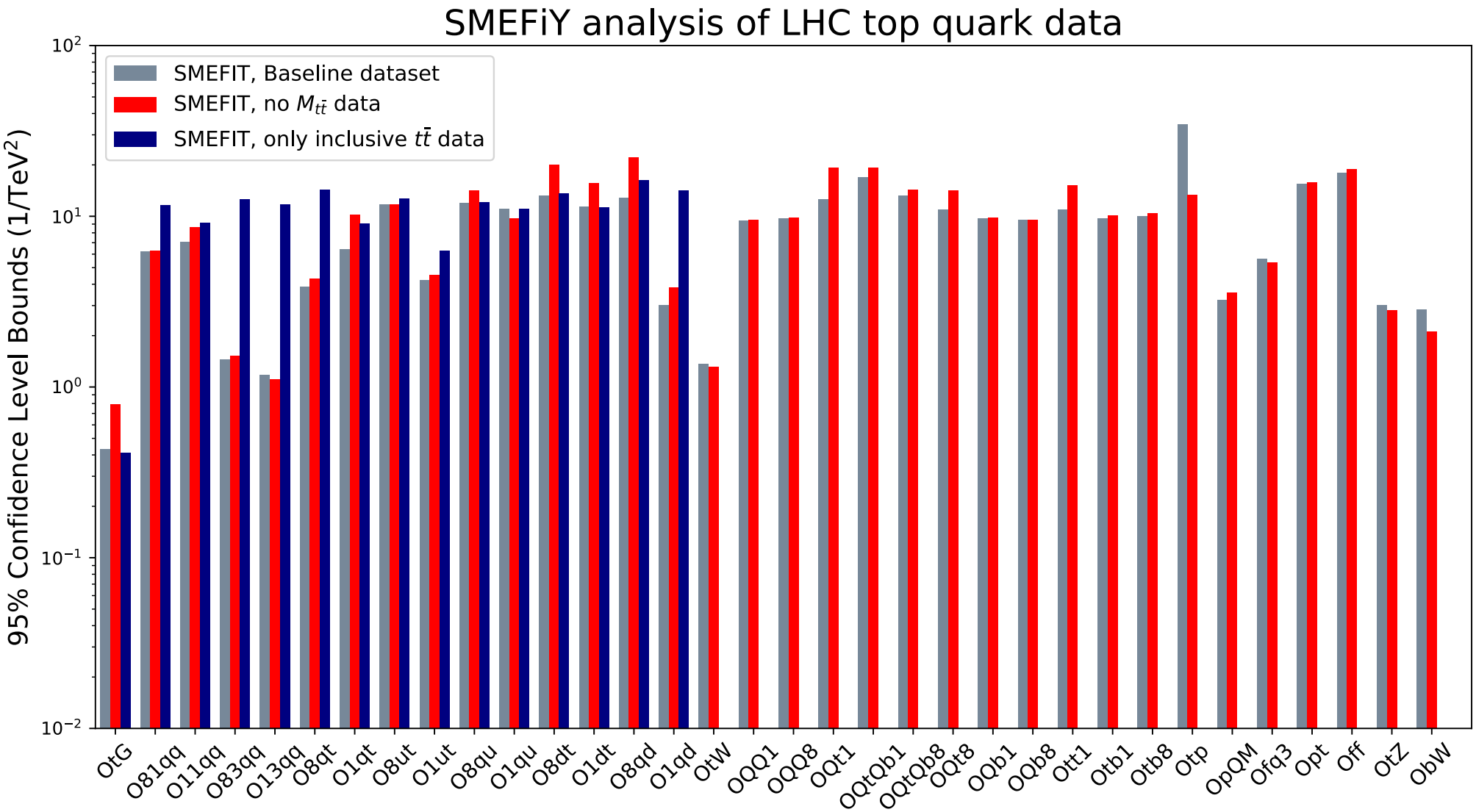
Energy-growing effects enhance sensitivity to SMEFT effects with **TeV-scale cross-sections** but need to be careful to ensure **validity of EFT description**

Dependence on theory settings



Accounting for the **quadratic O(Λ^{-4}) terms** strengthens bounds for several operators

Dependence on dataset



Reasonable stability of the fit results with respect to the **choice of dataset**

Bayesian reweighting the SMEFT parameter space

Samuel van Beek, Emanuele R. Nocera, Juan Rojo, Emma Slade, in preparation

Bayesian reweighting

- Under many circumstances, one would like to **quantify the impact of a new measurement** in the SMEFT parameter space **without having to redo the full fit**

- One would also like to quantify (and compare) the **amount of information** contained in current and (possible) future measurements

Bayesian Inference: update (``reweight") the SMEFiT probability distribution with the information provided by the **new measurements**

$$\omega_k \propto \left(\chi_k^2\right)^{(n_{\rm dat}-1)/2} \exp\left(-\chi_k^2/2\right) \ , \quad k = 1,\ldots,N_{\rm rep}$$

weight of k-th replica

number of data points in new data

total χ^2 of new data for k-th replica

MC replicas of a prior fit

- Extensive **validation** of reweighing with **direct fits** in the PDF case. What about SMEFT?

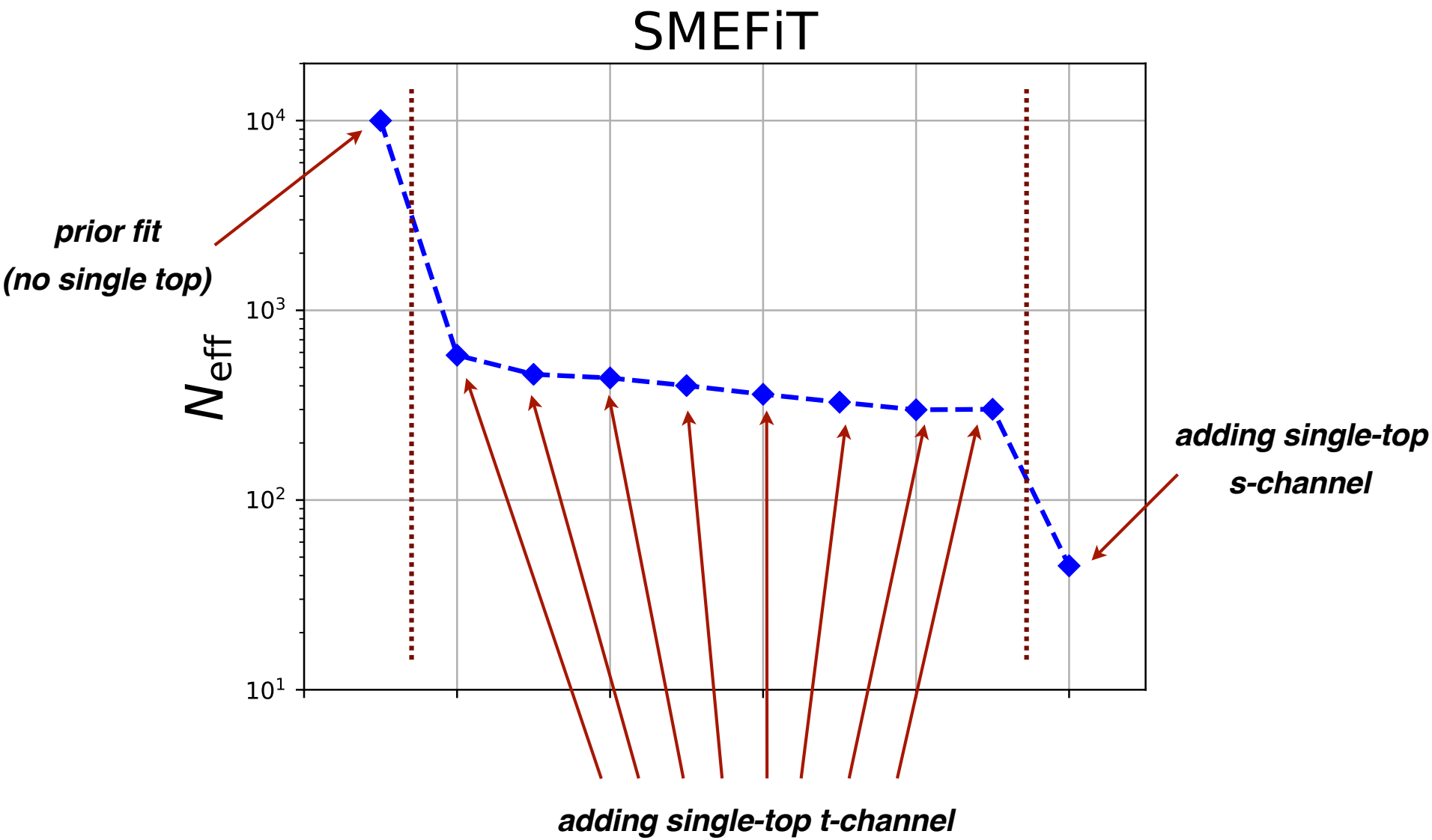
Bayesian reweighting

- Start from a variant of SMEFiT which excludes **LHC single top production data**
- To ensure sufficient statistics, this prior is constructed with N_{rep} **= 10000 MC replicas**
- Then add different combinations of single top data either by **reweighting** or by a **direct fit** and compare the results
- The amount of new information in each case is quantified by Shannon's entropy: the **effective number of replicas**

$$N_{\rm eff} = \exp\left(\frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}\omega_k \ln\frac{N_{\rm rep}}{\omega_k}\right)$$

- For Bayesian reweighting to be used reliably, one requires that $N_{eff} > 50$, else we run **out of statistics** and a direct refit is required

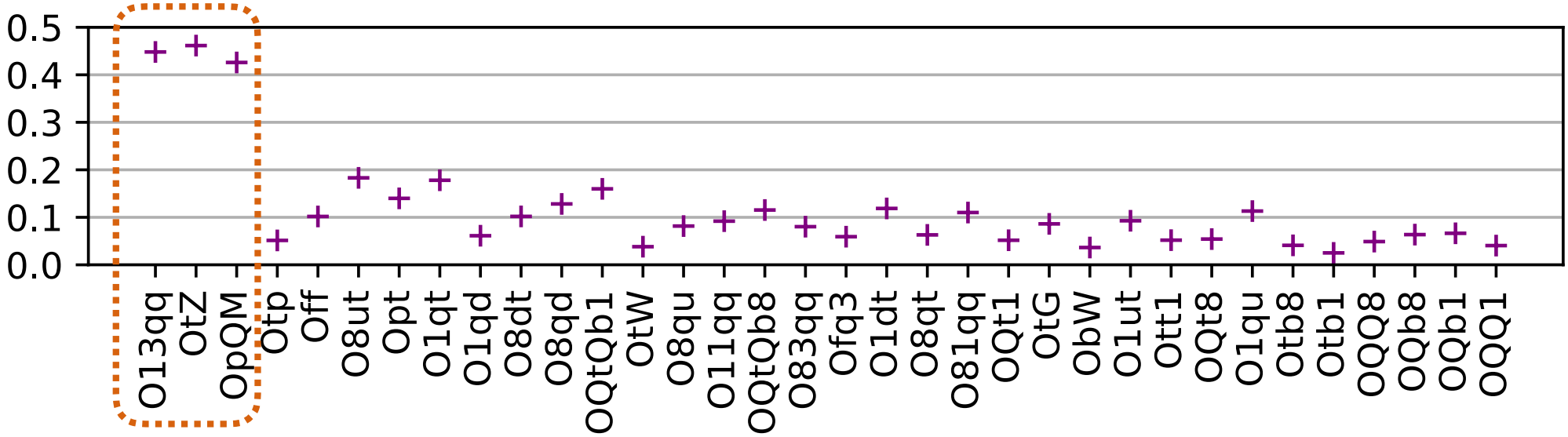
Reweighting efficiency



Significant amount of **new information** each time new process added via reweighting: marked decrease in effective number of replicas

Kolmogorov-Smirnov statistic

- To identify which SMEFT directions are more constrained by the new data, evaluate the Kolmogorov-Smirnov statistic between the **prior** and **reweighted probability distributions:** the larger the KS-statistic, the larger the effect of the new data

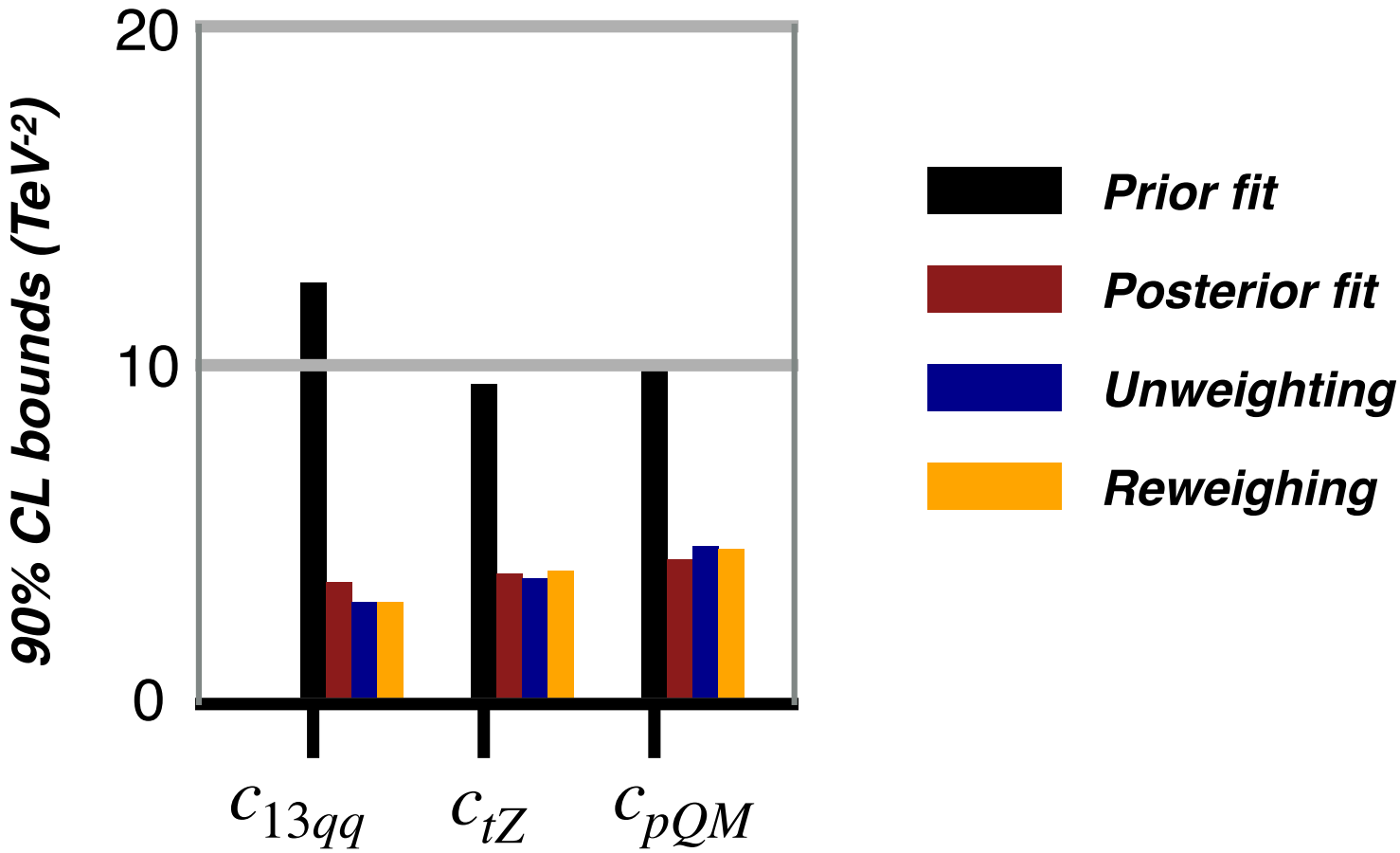


KS-statistic for the N_{op}=34 operators after adding single-top t-channel data

- Note that information can be added *(i)* due to **new direct constraints** and/or *(ii)* by **breaking degeneracies** in the parameter space

Validation

Validate the results of the reweighting (adding t-ch single top) by comparing with the corresponding fit, for those degrees of freedom with the largest KS statistic



Excellent **agreement** between the reweighted and direct fit results!

Summary and outlook

- Presented a novel framework, **SMEFiT,** suitable for global analyses of the SMEFT, which exploit expertise inherited from global PDF fits

- As a proof-of-concept, applied this framework to the exploration of the constraints in the SMEFT parameter space provided by **LHC top quark data**

- **Improved constraints** compared to previous studies (first-even bounds in some cases)

- Demonstrated **Bayesian reweighting** for the *a posteriori* inclusion of the constraints from new measurements on SMEFiT without need of redoing fit

- Next steps: enlarge the operator fitting basis and include **additional LHC cross-sections** (Higgs, electroweak, jets) as well as flavour and low-energy observables, and explore implications for specific **UV-complete models**

- Ultimately the **simultaneous determination of PDFs and SMEFT** degrees of freedom might be required to fully exploit the LHC potential: see **Maria's talk on Wed!**

Summary and outlook

- Presented a novel framework, **SMEFiT,** suitable for global analyses of the SMEFT, which exploit expertise inherited from global PDF fits

- As a proof-of-concept, applied this framework to the exploration of the constraints in the SMEFT parameter space provided by **LHC top quark data**

- **Improved constraints** compared to pre[illegible] me cases)

- Demonstrate[illegible] constraints from new m[illegible]

- Next ste[illegible] fitting basis and include **additional LHC cross-sections** (Higgs, el[illegible], jets) as well as flavour and low-energy observables, and explore implications for specific **UV-complete models**

- Ultimately the **simultaneous determination of PDFs and SMEFT** degrees of freedom might be required to fully exploit the LHC potential: see **Maria's talk on Wed!**

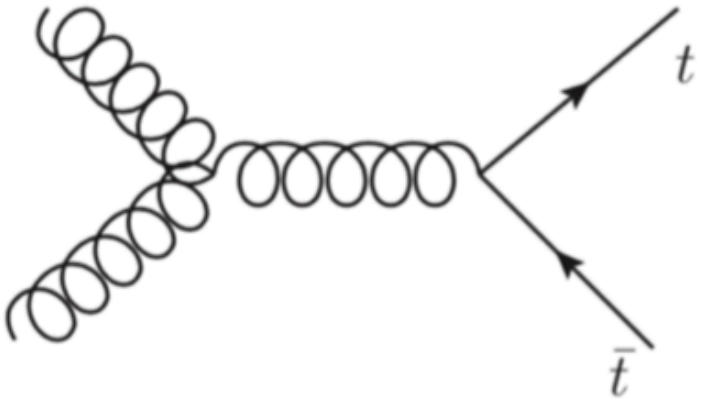
Thanks for your attention!

Extra Material

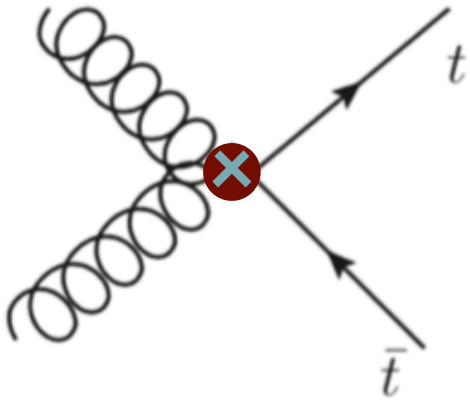
SMEFT effects

Standard Model

$$^{\ddagger}O_{uG}^{(ij)} = (\bar{q}_i \sigma^{\mu\nu} T^A u_j)\, \tilde{\varphi} G^A_{\mu\nu}\,,$$



+



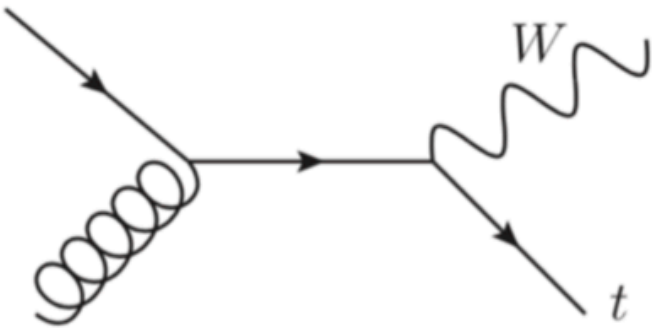
$$= \sigma_{SM} \times \left(1 + a\frac{c_{tG}}{\Lambda^2} + b\frac{c_{tG}^2}{\Lambda^4} \right)$$

SM: N(NLO) QCD **interference** **squared**

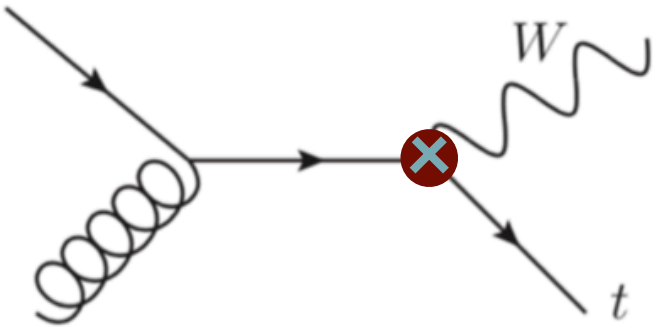
SMEFT effects

Standard Model

$$O_{uW}^{(ij)} = (\bar{q}_i \sigma^{\mu\nu} \tau^I u_j)\, \tilde{\varphi} W^I_{\mu\nu}$$



+



$$= \sigma_{SM} \times \left(1 + a \frac{c_{tW}}{\Lambda^2} + b \frac{c_{tW}^2}{\Lambda^4} \right)$$

SM: N(NLO) QCD **interference** **squared**

Bayesian reweighting

- Under many circumstances, one would like to **quantify the impact of a new measurement** in the SMEFT parameter space without having to redo the full fit
- One would also like to quantify (and compare) the **amount of information** contained in current and (possible) future measurements

Bayesian Inference: update (``reweight") the SMEFiT probability distribution with the information provided by the **new measurements**

NNPDF

$$\omega_k \propto \left(\chi_k^2\right)^{(n_{\rm dat}-1)/2} \exp\left(-\chi_k^2/2\right) \ , \quad k = 1,\ldots,N_{\rm rep}$$

weight of k-th replica

number of data points in new data

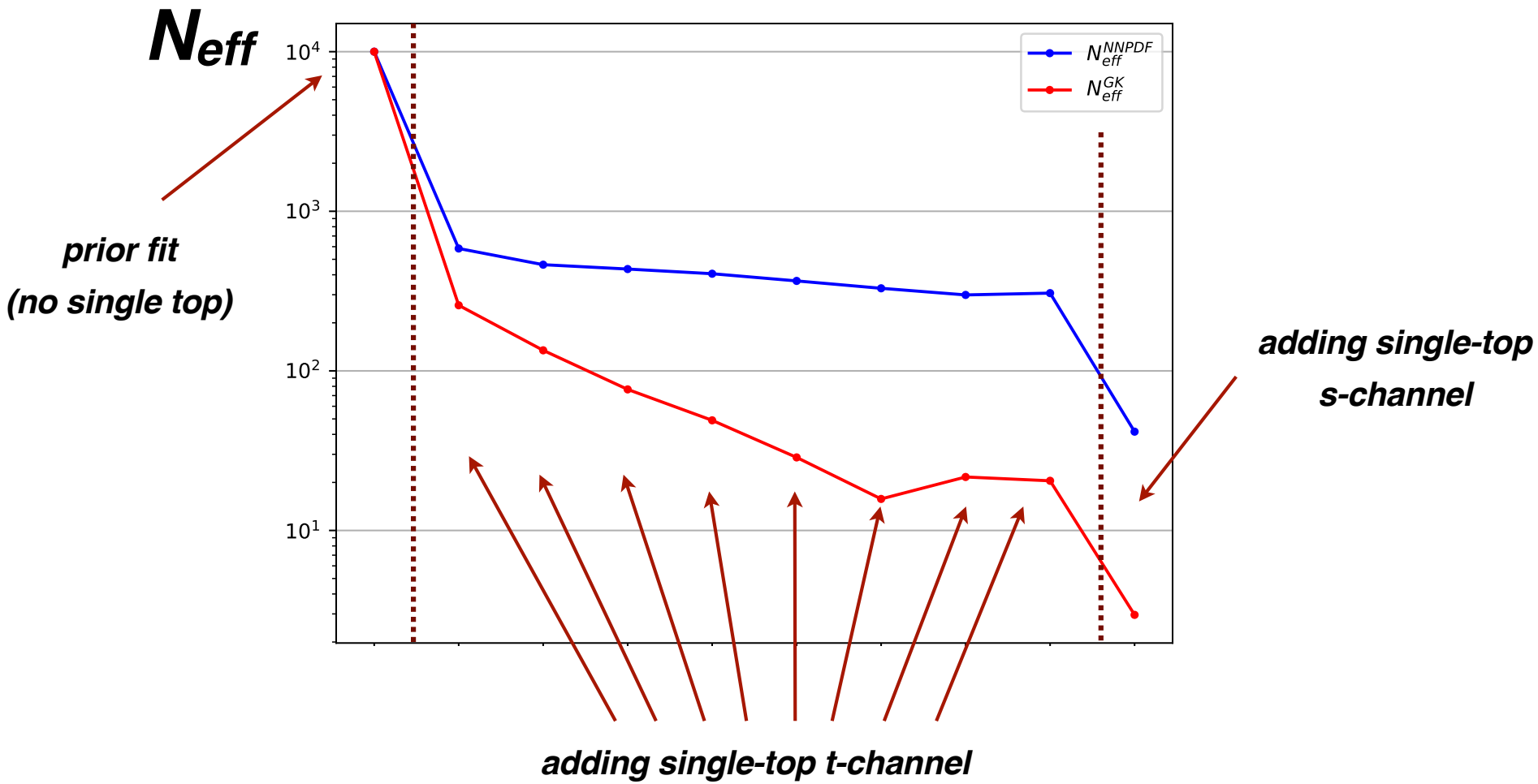
total χ^2 of new data for k-th replica

MC replicas of a prior fit

Giele-Keller

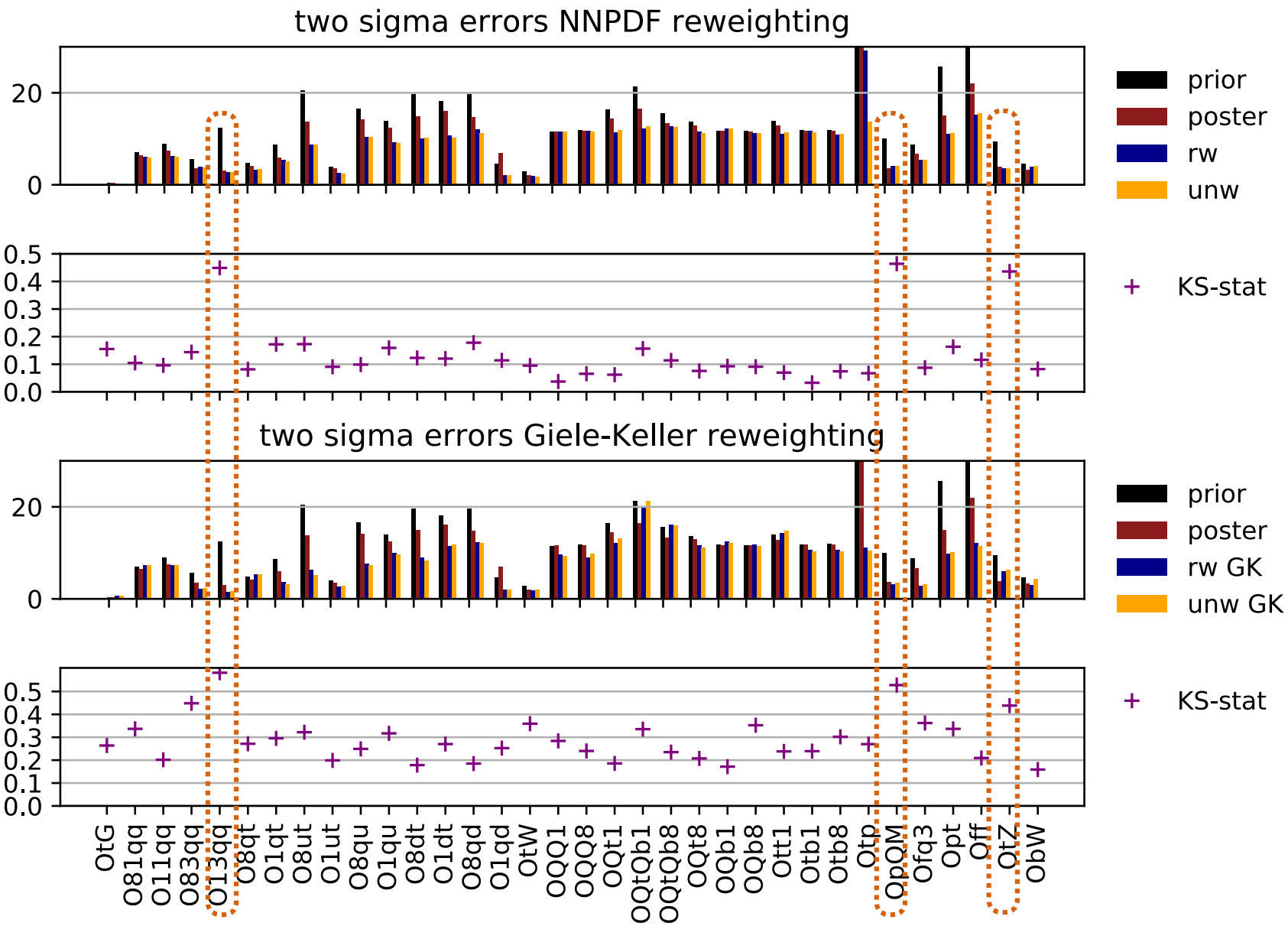
$$\omega_k \propto \exp\left(-\chi_k^2/2\right) \ , \quad k = 1,\ldots,N_{\rm rep}$$

NNPDF vs GK reweighting



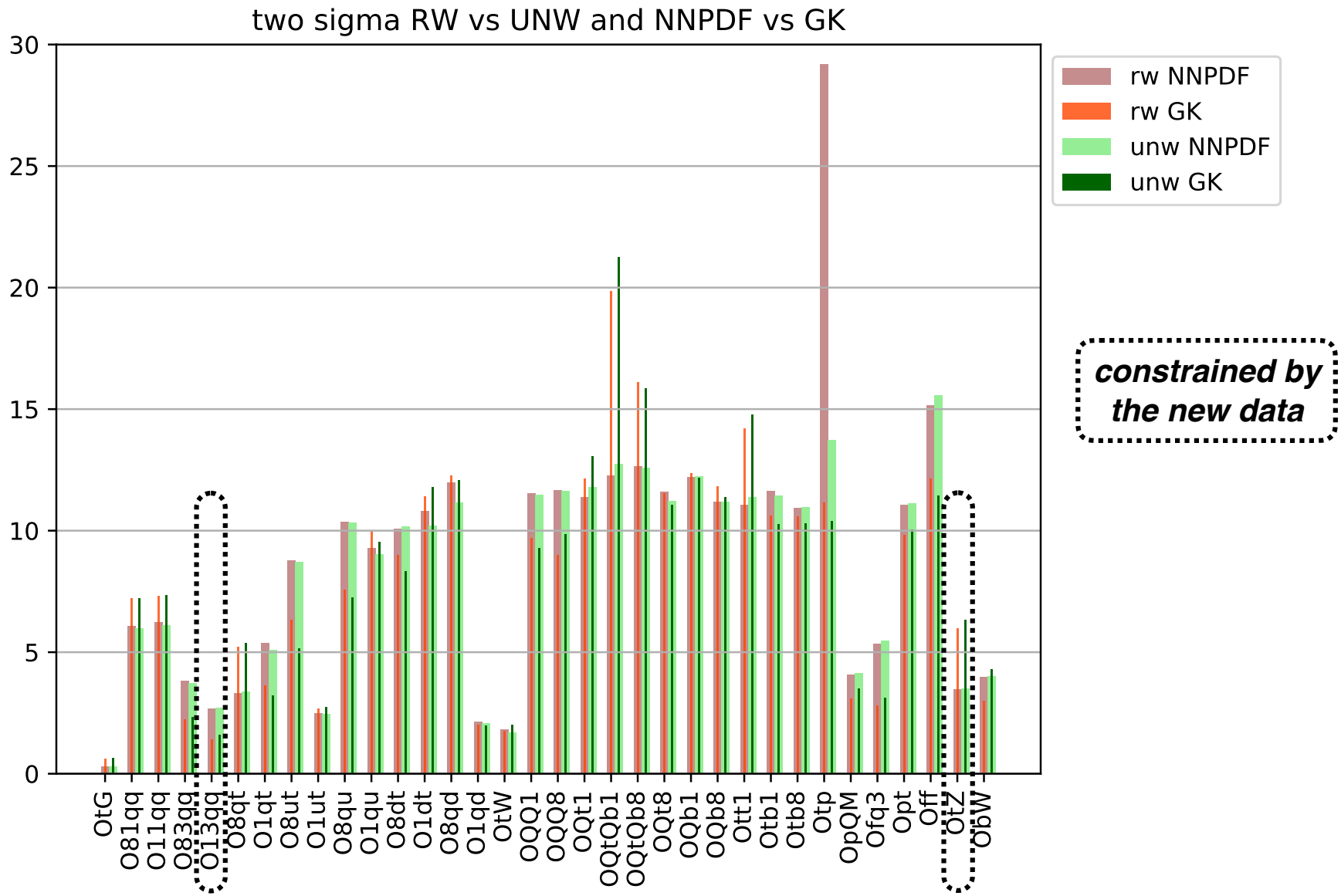
For the same experimental data added via reweighting, **GK loses efficiency** much quicker than NNPDF (runs faster out of replicas)

NNPDF vs GK reweighting



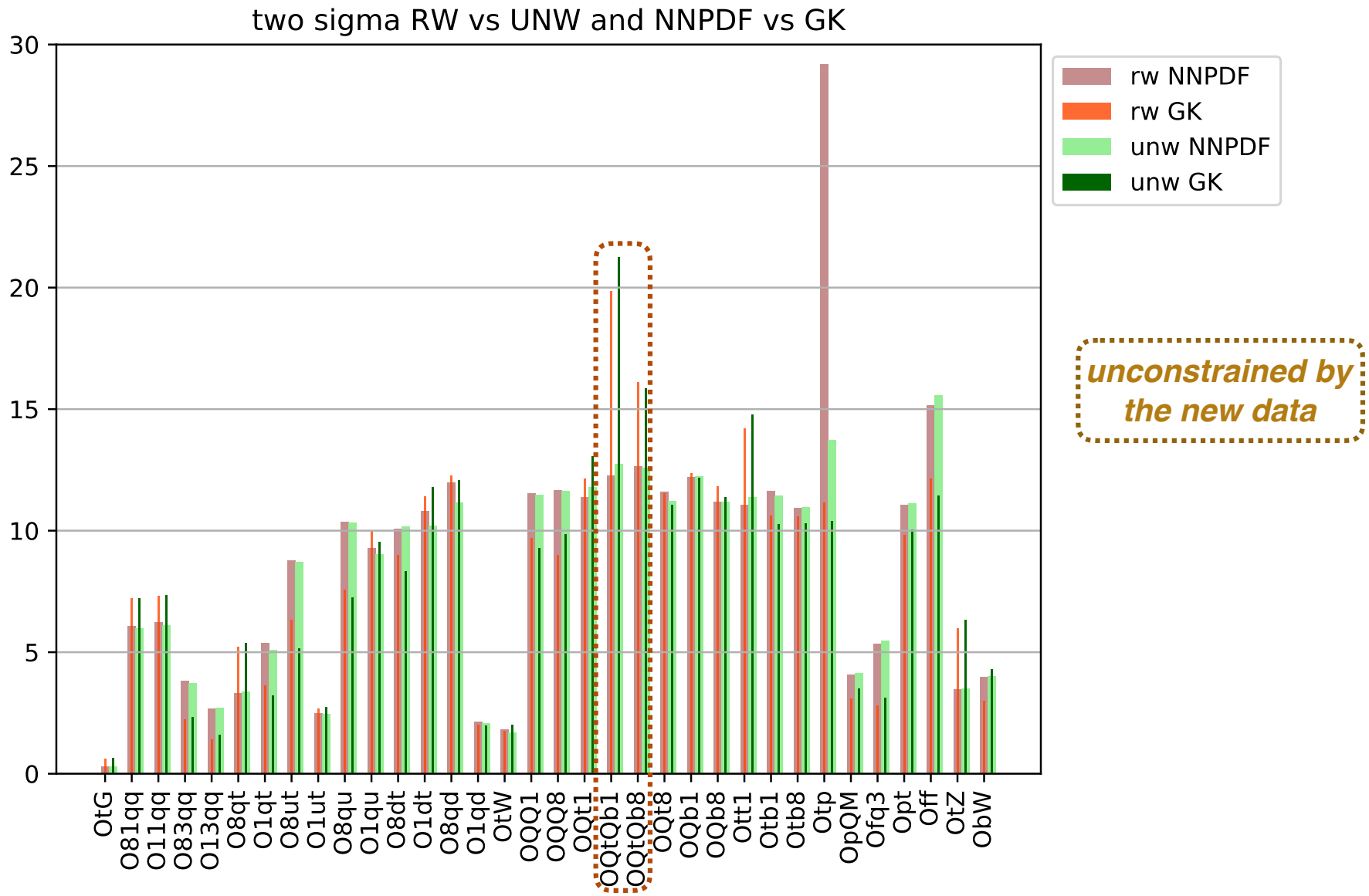
NNPDF reweighting in better **agreement with posterior fit** than GK

NNPDF vs GK reweighting



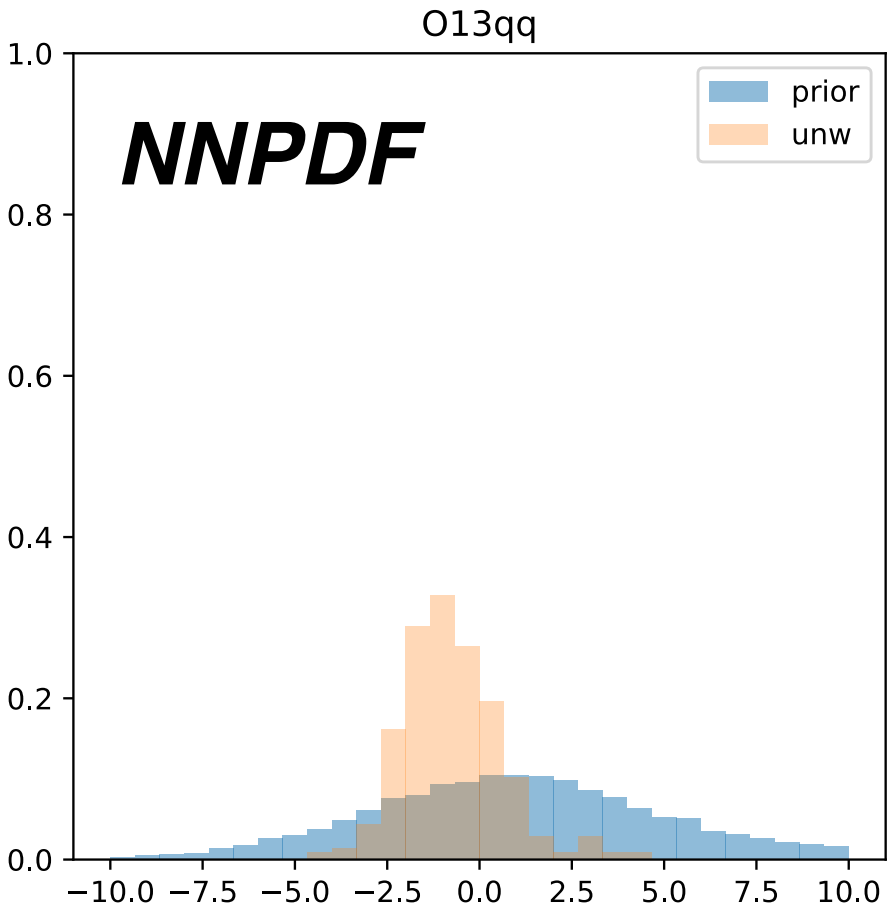
GK and NNPDF results different even for **operators unaffected** by the new data!

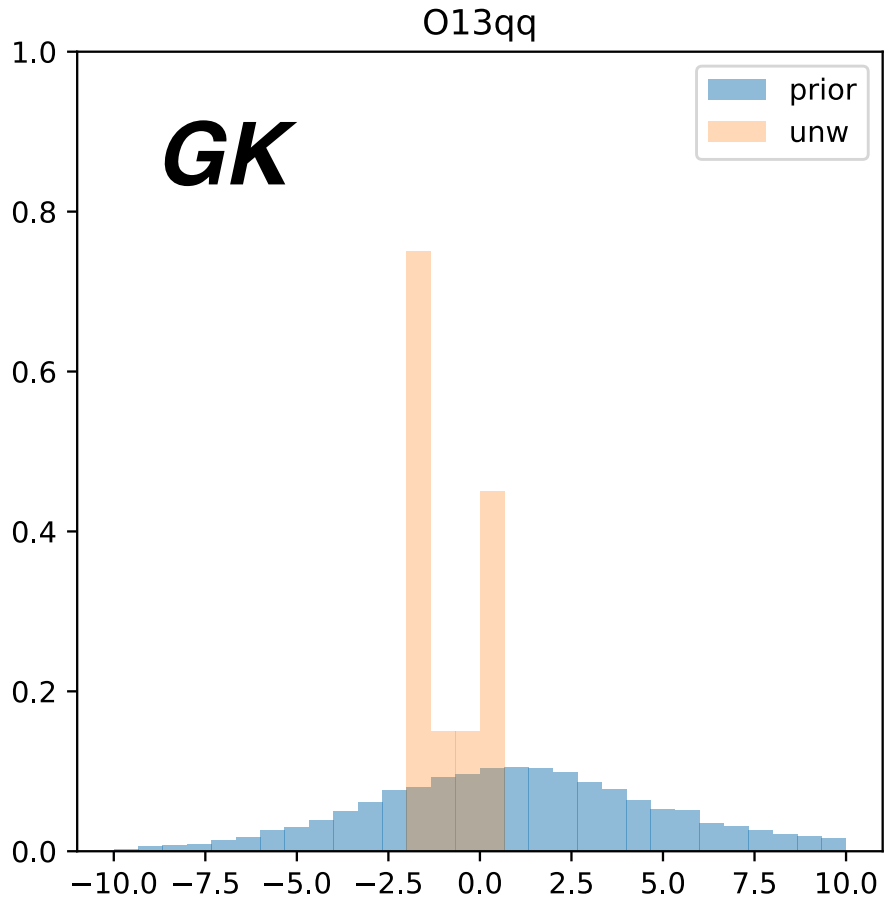
NNPDF vs GK reweighting



GK and NNPDF results different even for **operators unaffected** by the new data!

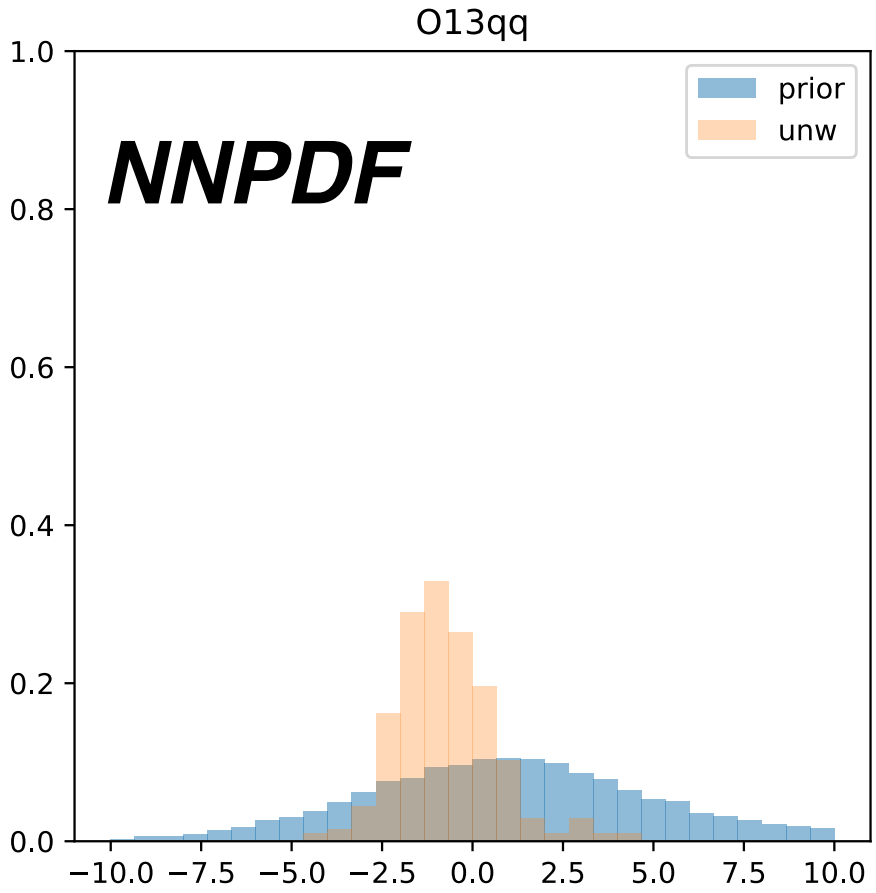
NNPDF vs GK reweighting

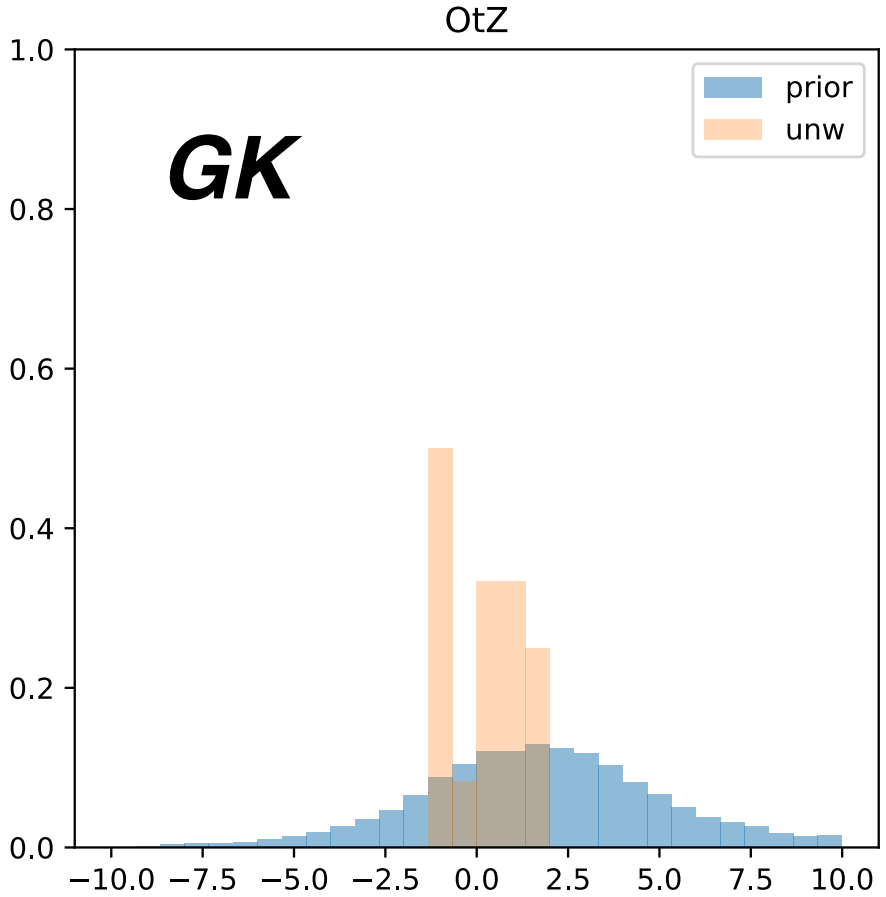




The **reweighted distributions** of SMEFT degrees of freedom look peculiar with GK…

NNPDF vs GK reweighting





The **reweighted distributions** of SMEFT degrees of freedom look peculiar with GK…

NNPDF vs GK reweighting