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

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

- 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

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

Some operators induce **growth with the partonic centre-of-mass energy**:

increased sensitivity in LHC cross-sections in the TeV region

\[ \sigma(E) = \sigma_{\text{SM}}(E) \left( 1 + \sum_i N_{d6} \frac{\omega_i c_i m_{\text{SM}}^2}{\Lambda^2} + \sum_i \tilde{\omega}_i \frac{c_i E^2}{\Lambda^2} + \mathcal{O} \left( \Lambda^{-4} \right) \right) \]

**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

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

Some operators induce **growth with the partonic centre-of-mass energy**:

increased sensitivity in LHC cross-sections in the TeV region

\[ \sigma(E) = \sigma_{\text{SM}}(E) \left(1 + \sum_i^{N_{d6}} \omega_i \frac{c_i m_{\text{SM}}^2}{\Lambda^2} + \sum_i^{N_{d6}} \tilde{\omega}_i \frac{c_i E^2}{\Lambda^2} + \mathcal{O}(\Lambda^{-4}) \right) \]

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  
Efficient exploration of parameter space  
Faithful uncertainty estimate (exp & th)  
Avoiding under- and over-fitting

**Methodology**

Juan Rojo
The optimisation conundrum

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, ....

Deterministic algorithms: follow the gradient of the cost function

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

Genetic Algorithms

Juan Rojo
The SMEFiT framework

An 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,
The SMEFiT method

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

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

- **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.

\[
\mathcal{O}_i^{(\text{art})(k)} = S_{i,N}^{(k)} \mathcal{O}_i^{(\text{exp})} \left( 1 + \sum_{\alpha=1}^{N_{\text{sys}}} r_{i,\alpha}^{(k)} \sigma_{i,\alpha}^{(\text{sys})} + r_{i}^{(k)} \sigma_{i}^{(\text{stat})} \right), \quad k = 1, \ldots, N_{\text{rep}}
\]

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

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

**to be determined from the data**

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

Juan Rojo

HEFT2019, CP3 Louvain-la-Neuve
The SMEFiT method

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

\[ \mathcal{O}_i^{(\text{art})(k)} = S_{i,N}^{(k)} \mathcal{O}_i^{(\text{exp})} \left( 1 + \sum_{\alpha=1}^{N_{\text{sys}}} r_{i,\alpha}^{(k)} \sigma_i^{(\text{sys})} + r_{i}^{(k)} \sigma_i^{(\text{stat})} \right), \quad k = 1, \ldots, N_{\text{rep}} \]

Construct theory calculations where the SM is extended by SMEFT corrections

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

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

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

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

\[
E(\{c_i^{(k)}\}) \equiv \frac{1}{N_{\text{dat}}} \sum_{i,j=1}^{N_{\text{dat}}} \left( \mathcal{O}_{i}^{(\text{th})}(\{c_n^{(k)}\}) - \mathcal{O}_{i}^{(\text{art})(k)} \right) (\text{cov}^{-1})_{ij} \left( \mathcal{O}_{j}^{(\text{th})}(\{c_n^{(k)}\}) - \mathcal{O}_{j}^{(\text{art})(k)} \right)
\]

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

**t_0** prescription

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

\[
\text{COV}_{ij} = \text{COV}_{ij}^{(\text{exp})} + \text{COV}_{ij}^{(\text{th})}
\]

**th uncertainties: PDFs**

\[
\text{COV}_{ij}^{(\text{th})} = \langle \mathcal{O}_{i}^{(\text{th})(r)} \mathcal{O}_{j}^{(\text{th})(r)} \rangle_{\text{rep}} - \langle \mathcal{O}_{i}^{(\text{th})(r)} \rangle_{\text{rep}} \langle \mathcal{O}_{j}^{(\text{th})(r)} \rangle_{\text{rep}},
\]

Juan Rojo

HEFT2019, CP3 Louvain-la-Neuve
The SMEFiT method

- Determine the SMEFT coefficients \textit{replica-by-replica} by minimising a cost function

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

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

\[ \text{cov}_{ij} = \text{cov}^{(\text{exp})}_{ij} + \text{cov}^{(\text{th})}_{ij} \]

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

\[ \langle c_l \rangle \equiv \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} c_l^{(k)} \quad \rho(c_i, c_j) = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} c_i^{(k)} c_j^{(k)} - \langle c_i \rangle \langle c_j \rangle }{\delta c_i \delta c_j} \]
Sampling the SMEFT probability distribution

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

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

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

\[
\left\{ c_n^{(k)} \right\}, \quad n = 1, \ldots, N_{\text{op}}, \quad k = 1, \ldots, N_{\text{rep}}
\]

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

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

\[ O_{\text{QQ8}} \]

![Distribution of c_{QQ8} (TeV^2) for h1 with 1000 entries, mean -0.5848, std dev 5.795]
The SMEFiT method

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

\[
(\delta c_n)^2 = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} (c_n^{(k)})^2 - \langle c_n \rangle^2
\]

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

\[ E(\{c_1^{(k)}\}) \]
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 $O(\Lambda^{-2})$ and $O(\Lambda^{-4})$
  - 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

Juan Rojo

HEFT2019, CP3 Louvain-la-Neuve
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,
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.
The top quark sector of the SMEFT

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

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

Juan Rojo

HEFT2019, CP3 Louvain-la-Neuve
The top quark sector of the SMEFT

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

\[
\sigma_{i}^{\text{th}}(\{c_n\}) = \sigma_{\text{SM},i} + \sum_{n=1}^{N_{\text{op}}} \sigma_{i,n} \frac{c_n}{\Lambda^2} + \sum_{n,m=1}^{N_{\text{op}}} \sigma_{i,nm} \frac{c_n c_m}{\Lambda^4}
\]
## The top quark sector of the SMEFT

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

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

### Notation and Sensitivity

<table>
<thead>
<tr>
<th>Notation</th>
<th>Sensitivity at $O(\Lambda^{-2})$</th>
<th>Sensitivity at $O(\Lambda^{-4})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$</td>
<td>single-top</td>
<td></td>
</tr>
<tr>
<td>$tW$</td>
<td>$tZ$</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}W$</td>
<td>$t\bar{t}Z$</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}t$</td>
<td>$t\bar{t}\bar{b}$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Notation</th>
<th>Sensitivity at $O(\Lambda^{-2})$</th>
<th>Sensitivity at $O(\Lambda^{-4})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_1$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_2$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_3$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_4$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_5$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_6$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_7$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_8$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_9$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td>$q_{10}$</td>
<td>$\checkmark$</td>
<td></td>
</tr>
</tbody>
</table>

Juan Rojo
HEFT2019, CP3 Louvain-la-Neuve
The top quark sector of the SMEFT

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

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

Juan Rojo

HEFT2019, CP3 Louvain-la-Neuve
## Input dataset (I)

<table>
<thead>
<tr>
<th>Process</th>
<th>Dataset</th>
<th>$\sqrt{s}$</th>
<th>Info</th>
<th>Observables</th>
<th>$N_{\text{dat}}$</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$</td>
<td>ATLAS_tt_8TeV_1jets</td>
<td>8 TeV</td>
<td>lepton+jets</td>
<td>$d\sigma/d</td>
<td>y_t</td>
<td>, d\sigma/dp_T^T, d\sigma/dm_{t\bar{t}}, d\sigma/d</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>CMS_tt_8TeV_1jets</td>
<td>8 TeV</td>
<td>lepton+jets</td>
<td>$d\sigma/dy_t, d\sigma/dp_T^T, d\sigma/dm_{t\bar{t}}, d\sigma/dy_{t\bar{t}}$</td>
<td>10, 8, 7, 10</td>
<td>[78]</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>CMS_tt2D_8TeV_dilep</td>
<td>8 TeV</td>
<td>dileptons</td>
<td>$d^2\sigma/dy_td^2p_T^T, d^2\sigma/dy_dm_{t\bar{t}}, d^2\sigma/dp_T^Tdm_{t\bar{t}}, d^2\sigma/dy_dm_{t\bar{t}}$</td>
<td>16, 16, 16, 16</td>
<td>[79]</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>CMS_tt_13TeV_1jets</td>
<td>13 TeV</td>
<td>lepton+jets</td>
<td>$d\sigma/d</td>
<td>y_t</td>
<td>, d\sigma/dp_T^T, d\sigma/dm_{t\bar{t}}, d\sigma/d</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>CMS_tt_13TeV_1jets2</td>
<td>13 TeV</td>
<td>lepton+jets</td>
<td>$d\sigma/d</td>
<td>y_t</td>
<td>, d\sigma/dp_T^T, d\sigma/dm_{t\bar{t}}, d\sigma/d</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>CMS_tt_13TeV_dilep</td>
<td>13 TeV</td>
<td>dileptons</td>
<td>$d\sigma/dy_t, d\sigma/dp_T^T, d\sigma/dm_{t\bar{t}}, d\sigma/dy_{t\bar{t}}$</td>
<td>8, 6, 6, 8</td>
<td>[86]</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>ATLASCMS_AcMtt_8TeV</td>
<td>8 TeV</td>
<td>Asymm comb</td>
<td>$A_C(m_{t\bar{t}})$, Eq. (3.1)</td>
<td>6</td>
<td>[80]</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>ATLAS_WhelF_8TeV</td>
<td>8 TeV</td>
<td>W helicity fract</td>
<td>$F_0, F_L, F_R$</td>
<td>3</td>
<td>[81]</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>CMS_WhelF_8TeV</td>
<td>8 TeV</td>
<td>W helicity fract</td>
<td>$F_0, F_L, F_R$</td>
<td>3</td>
<td>[82]</td>
</tr>
</tbody>
</table>
## Input dataset (II)

<table>
<thead>
<tr>
<th>Process</th>
<th>Dataset</th>
<th>$\sqrt{s}$</th>
<th>Info</th>
<th>Observables</th>
<th>$N_{\text{dat}}$</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single $t$</td>
<td>CMS_t_tch_8TeV_inc</td>
<td>8 TeV</td>
<td>$t$-channel</td>
<td>$\sigma_{\text{tot}}(t), \sigma_{\text{tot}}(t\bar{t})$ ($R_t$)</td>
<td>2 (1)</td>
<td>[95]</td>
</tr>
<tr>
<td>Single $t$</td>
<td>CMS_t_sch_8TeV</td>
<td>8 TeV</td>
<td>$s$-channel</td>
<td>$\sigma_{\text{tot}}(t + \bar{t})$</td>
<td>1</td>
<td>[96]</td>
</tr>
<tr>
<td>Single $t$</td>
<td>ATLAS_t_sch_8TeV</td>
<td>8 TeV</td>
<td>$s$-channel</td>
<td>$\sigma_{\text{tot}}(t + \bar{t})$</td>
<td>1</td>
<td>[97]</td>
</tr>
<tr>
<td>Single $t$</td>
<td>ATLAS_t_tch_8TeV</td>
<td>8 TeV</td>
<td>$t$-channel</td>
<td>$d\sigma(tq)/dp^t_T, d\sigma(\bar{t}q)/dp^t_T$ (d\sigma(tq)/dy_t, d\sigma(\bar{t}q)/dy_t)</td>
<td>5, 4, 4, 4</td>
<td>[98]</td>
</tr>
<tr>
<td>Single $t$</td>
<td>ATLAS_t_tch_13TeV</td>
<td>13 TeV</td>
<td>$t$-channel</td>
<td>$\sigma_{\text{tot}}(t), \sigma_{\text{tot}}(t\bar{t})$ ($R_t$)</td>
<td>2 (1)</td>
<td>[99]</td>
</tr>
<tr>
<td>Single $t$</td>
<td>CMS_t_tch_13TeV_inc</td>
<td>13 TeV</td>
<td>$t$-channel</td>
<td>$\sigma_{\text{tot}}(t + \bar{t})$ ($R_t$)</td>
<td>1 (1)</td>
<td>[100]</td>
</tr>
<tr>
<td>Single $t$</td>
<td>CMS_t_tch_8TeV_dif</td>
<td>8 TeV</td>
<td>$t$-channel</td>
<td>$d\sigma/dp^{(t+\bar{t})}_T$, $d\sigma/dy^{(t+\bar{t})}$</td>
<td>6</td>
<td>[101]</td>
</tr>
<tr>
<td>Single $t$</td>
<td>CMS_t_tch_13TeV_dif</td>
<td>13 TeV</td>
<td>$t$-channel</td>
<td>$d\sigma/dp^{(t+\bar{t})}_T$, $d\sigma/dy^{(t+\bar{t})}$</td>
<td>4</td>
<td>[102]</td>
</tr>
<tr>
<td>$tW$</td>
<td>ATLAS_tW_inc_8TeV</td>
<td>8 TeV</td>
<td>inclusive</td>
<td>$\sigma_{\text{tot}}(tW)$</td>
<td>1</td>
<td>[103]</td>
</tr>
<tr>
<td>$tW$</td>
<td>CMS_tW_inc_8TeV</td>
<td>8 TeV</td>
<td>inclusive</td>
<td>$\sigma_{\text{tot}}(tW)$</td>
<td>1</td>
<td>[104]</td>
</tr>
<tr>
<td>$tW$</td>
<td>ATLAS_tW_inc_13TeV</td>
<td>13 TeV</td>
<td>inclusive</td>
<td>$\sigma_{\text{tot}}(tW)$</td>
<td>1</td>
<td>[105]</td>
</tr>
<tr>
<td>$tW$</td>
<td>CMS_tW_inc_13TeV</td>
<td>13 TeV</td>
<td>inclusive</td>
<td>$\sigma_{\text{tot}}(tW)$</td>
<td>1</td>
<td>[106]</td>
</tr>
<tr>
<td>$tZ$</td>
<td>CMS_tZ_inc_13TeV</td>
<td>13 TeV</td>
<td>inclusive</td>
<td>$\sigma_{\text{fid}}(Wb\ell^+\ell^-q)$</td>
<td>1</td>
<td>[107]</td>
</tr>
<tr>
<td>$tZ$</td>
<td>ATLAS_tZ_inc_13TeV</td>
<td>13 TeV</td>
<td>inclusive</td>
<td>$\sigma_{\text{tot}}(tZq)$</td>
<td>1</td>
<td>[108]</td>
</tr>
</tbody>
</table>
The fit includes more than 100 cross-section measurements at 8 and 13 TeV from 10 different top-quark production processes.
# Theory calculations

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

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
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.

Figure 4.9. Same as Fig. 4.5 comparing the L2 closure tests with and without cross-validation.

Terms $O_4$ are not constrained in the fit without $O_4$ terms, and are therefore set to zero. From this comparison, we see that the bounds on the coefficients generally improve when $O_4$ corrections are included in the theoretical calculation. For example, the bound on $OtZ$ decreases from $\sigma_{6\text{ TeV}} \approx 2\text{ TeV}$ to $\sigma_{2\text{ TeV}} \approx 2\text{ TeV}$. The slight worsening observed for the bounds on some few operators when only linear terms are included is consistent with statistical fluctuations, and is therefore not significant. In any case, the fit results are qualitatively similar irrespective of the inclusion of $O_4$ corrections. Note that some of the degrees of freedom are highly correlated, therefore the interpretation of the results at the individual bound level should be taken with care.
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 $\chi^2/n_{dat}$ of 1.11 (1.06) before (after) fit

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

Juan Rojo
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 \( t\bar{t}t\bar{t} \), \( t\bar{t}b\bar{b} \), and \( t\bar{t}h \) the SMEFT shifts can be as large as 20% (reflecting the larger experimental errors).
High-energy behaviour

$t\bar{t}$ production @ 13 TeV, CMS lepton+jets L=36 fb$^{-1}$

$\Delta_i^{(\text{smeft})} \equiv \kappa_i \frac{C_i}{\Lambda^2}$

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 $O(\Lambda^{-4})$ terms strengthens bounds for several operators.
Dependence on dataset

SMEFiY analysis of LHC top quark data

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

Juan Rojo

HEFT2019, CP3 Louvain-la-Neuve
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.

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

- **weight of** \( k \)-th replica
- **number of data points in new data**
- **total \( \chi^2 \) 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

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

\[
N_{\text{eff}} = \exp \left( \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \omega_k \ln \left( \frac{N_{\text{rep}}}{\omega_k} \right) \right)
\]

5. For Bayesian reweighting to be used reliably, one requires that \( N_{\text{eff}} > 50 \), 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
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 previous studies (first-ever bounds in some cases).

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

Juan Rojo  
HEFT2019, CP3 Louvain-la-Neuve
Extra Material
SMEFT effects

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

\[ \dagger O_{uG}^{ij} = (\bar{q}_i \sigma^{\mu\nu} T^A u_j) \tilde{\phi} G^A_{\mu\nu}, \]

SM: N(NLO) QCD  interference  squared
SMEFT effects

\[ O_{uW}^{(ij)} = (\bar{q}_i \sigma^{\mu\nu} \tau^I u_j) \phi W^I_{\mu\nu} \]

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

Standard Model

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**.

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

- Weight of the k-th replica
- Number of data points in new data
- Total $\chi^2$ of new data for k-th replica
- MC replicas of a prior fit

\[
\omega_k \propto \exp\left(-\chi_k^2/2\right), \quad k = 1,\ldots, N_{\text{rep}}
\]
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!
GK and NNPDF results different even for operators unaffected by the new data!
The reweighted distributions of SMEFT degrees of freedom look peculiar with GK...
The reweighted distributions of SMEFT degrees of freedom look peculiar with GK...
NNPDF vs GK reweighting

\[ C_{13qq} \]

\[ C_{tZ} \]