

Validation of ADL/CutLang analyses with SModelS EM-creator for reinterpretation studies

**Analysis Description Language workshop
[November 21-22, 2022]**

**Sezen Sekmen¹, Changgi Huh¹, [Junghyun Lee¹](#), Gokhan Unel²,
Wolfgang Waltenberger³, Jan Mrozek⁴**

¹Department of Physics, Kyungpook National University

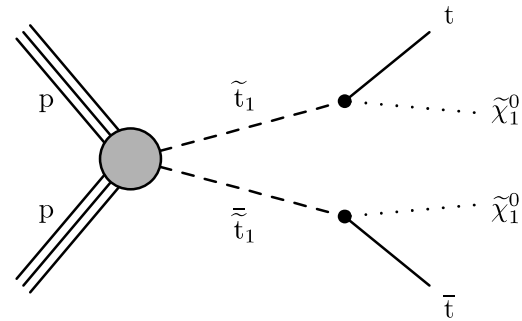
²The University of California, Irvine

³Research institute of the Austrian Academy of Sciences in Vienna (HEPHY)

⁴Red Hat Software

Simplified model spectra (SMS)

- **Simplified model spectra (SMS) are defined by a set of particles and their production & decays.**
 - Particle masses, cross sections, branching ratios are free parameters.
- **SMSs are practical for comparing sensitivities of different analyses.**
- **SMSs can be combined to form full models.**

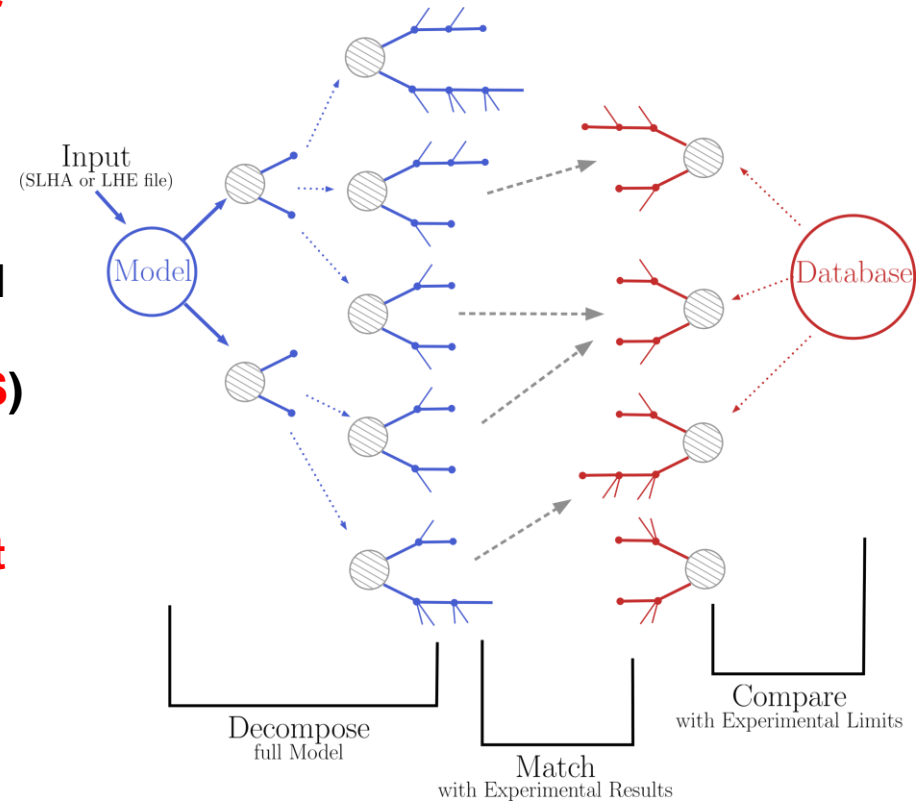


So, How do you interpret the LHC result of the Simplified model?

SModels

- **SModels** is an automatic, **public tool for interpreting simplified-model results** from the LHC.
- It is based on a general procedure to **decompose** Beyond the Standard Model (**BSM**) **collider signatures** presenting a Z_2 **into Simplified Model Spectrum (SMS) topologies**.
- **SModels** method **provides a way to cast BSM predictions for the LHC** in a model independent framework, which can be directly confronted with the relevant experimental constraints.

• Refer to official documentation for SModels :
<https://smodels.readthedocs.io/en/stable/>



• **Smodels** github :
<https://github.com/SModels/smodels>

Motivation of automation system

- SModelS requires **SMS efficiency maps**: event selection efficiencies versus SMS mass parameters.
 - It **calculates limits** from these efficiency maps.
- We need **to generate many mass points** to draw the limit plots for validation.
- That would be a **very repetitive work**

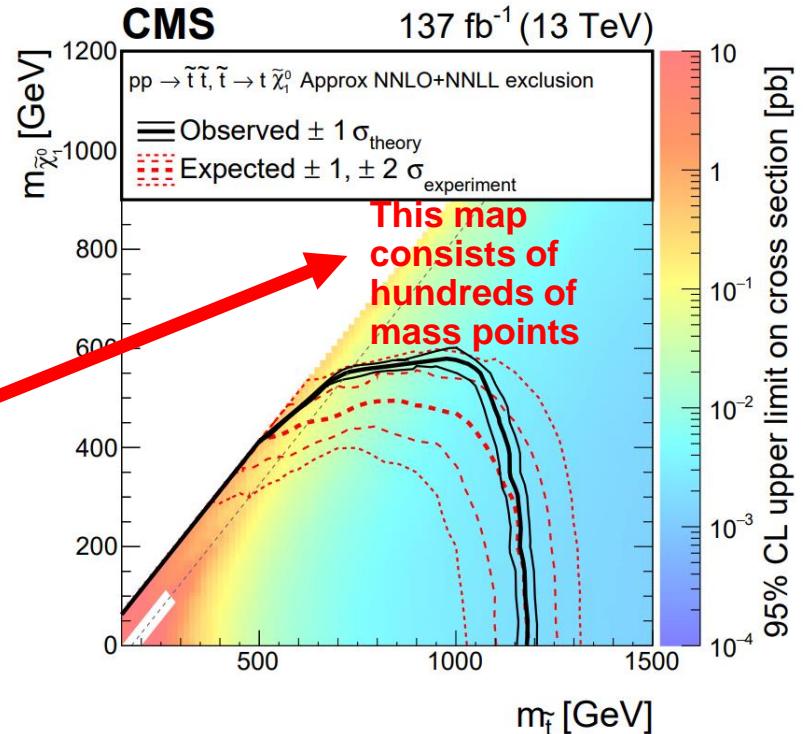
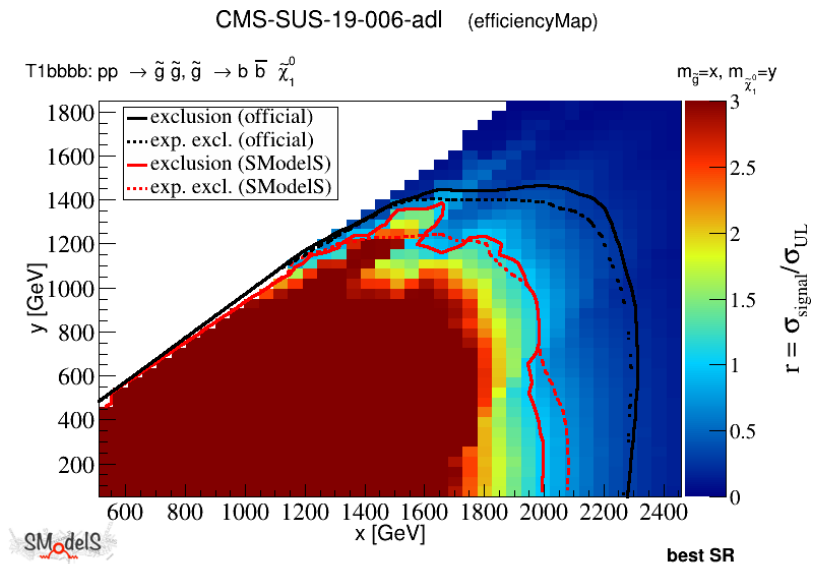
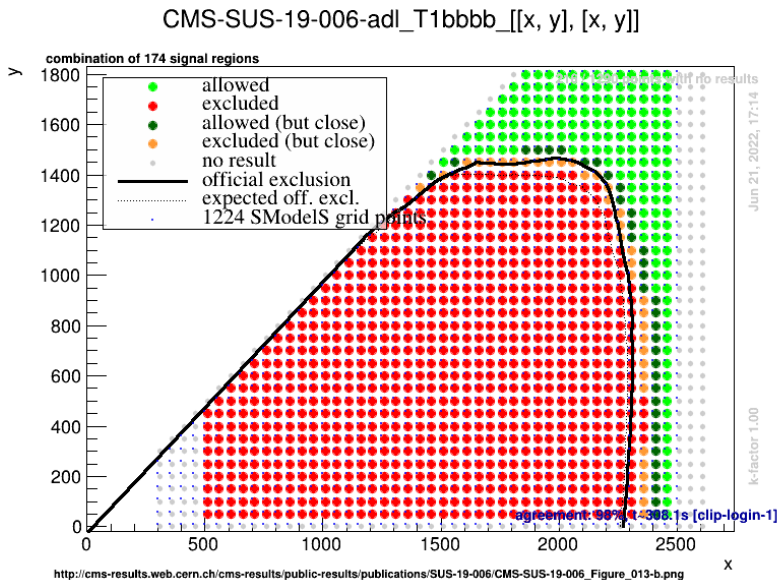


Figure : The 95% CL upper limit on the production cross sections of the T2tt
Ref : [CMS-SUS-19-006 \(cern.ch\)](https://arxiv.org/abs/1906.06622)

**WE NEED AN
“ AUTOMATION SYSTEM ”**

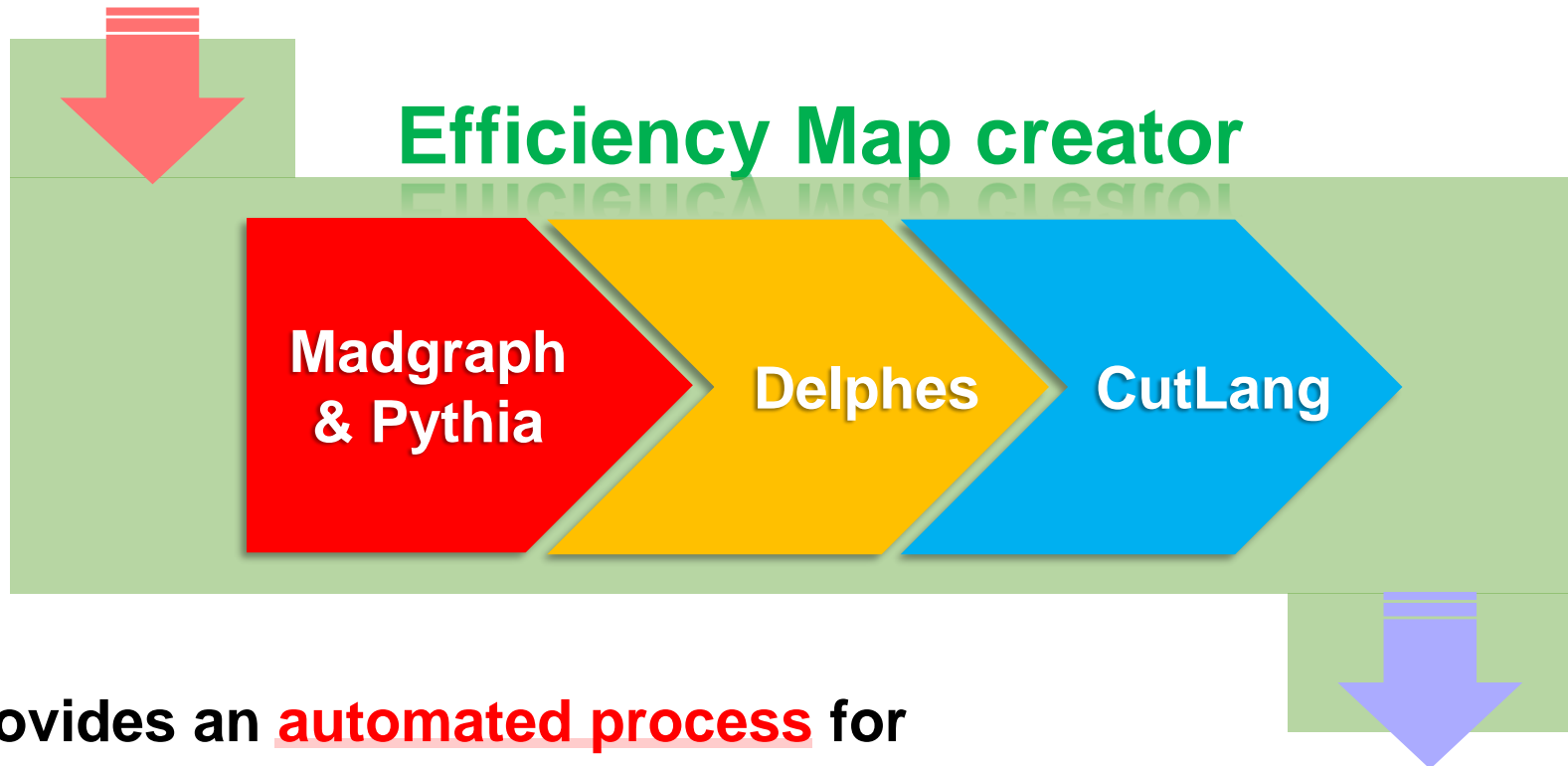
EM-creator

- **EM-creator** (Efficiency Map creator) was developed by Wolfgang W. to produce efficiency maps that can be input to SModelS.
 - It can be used to validate analysis implementations by comparing to experimental results.
- **EM-creator** was adapted to work with ADL/CutLang by Jan M., Gökhan Ü. and Wolfgang W.
 - **Efficiencies** are calculated with CutLang using analyses written in ADL.



EM-creator workflow

Number of events, ADL analysis file,
mass range, SMS topology, ...



Provides an **automated process** for
analysis through **simple command line**

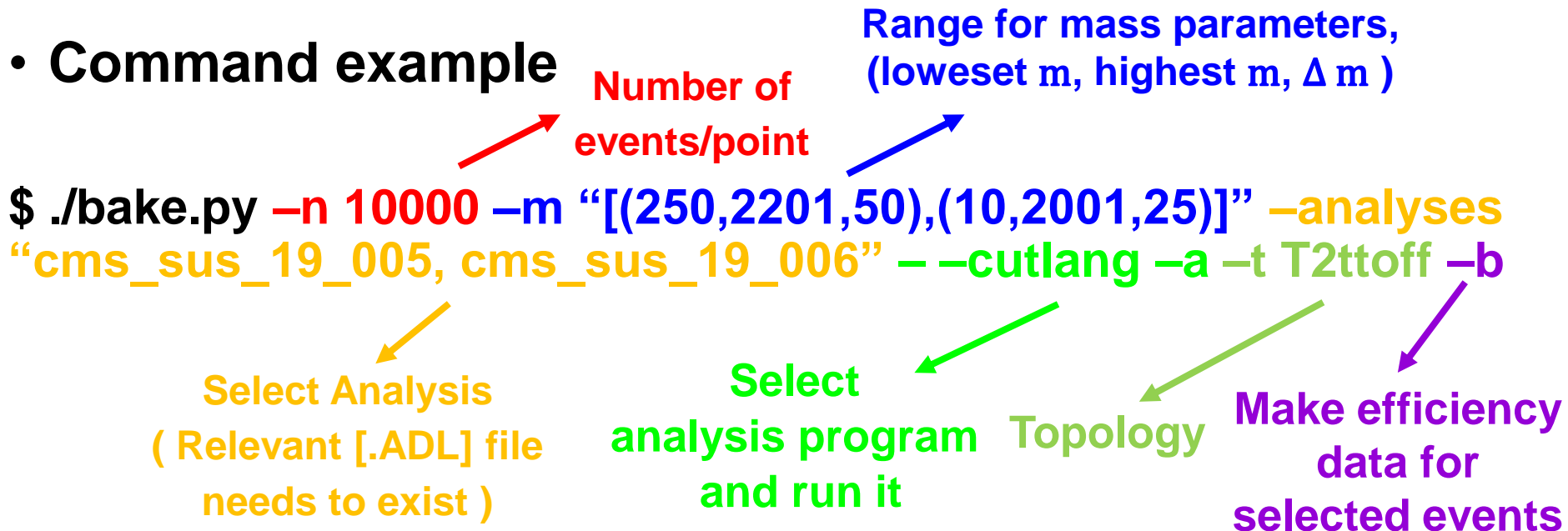
Selected number of
events / efficiencies in all
regions & bins

Run EM-creator

- Working environment

CMSSW_12_3_0_pre6 in KNU Tier3 system

- Command example



Operating log (1250, 1085) mass point

```

[CutLangWrapper] for temp T2ttoff we set mtag1 to 1.
[mg5Wrapper] this is MG5 v3_4_1
[mg5Wrapper] initialised
[mg5Wrapper] starting MG5 on (1250, 1085)[T2ttoff] at Fri Nov 24 11:57:08 UTC 2022
[mg5Wrapper] wrote run card /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS.root
[mg5Wrapper] run mg5 for (1250, 1085)[T2ttoff]. /home/condor/dir_28426/CMSSW_12_3_0_pre6/mg5proc9x59vgr4
[mg5Wrapper] now execute for T2ttoff[(1250, 1085)]: python3 /home/condor/dir_28426/CMSSW_12_3_0_pre6/bin/mg5_aMC T2ttoff_1jet.1250_1085/mg5proc 2>&1 | tee /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS.log
[mg5Wrapper] - ran compiler gfortran
INFO: Use c++ compiler g++
INFO: Generate web pages
Output to directory /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS
Type "launch" to ...
[mg5Wrapper] now execute for T2ttoff[(1250, 1085)]: python3 /home/condor/dir_28426/CMSSW_12_3_0_pre6/bin/mg5_aMC T2ttoff_1jet.1250_1085/mg5cmd 2>&1 | tee /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS.log
[mg5Wrapper] - h
  
```

1. Madgraph operates

Initialisation complete. Writing output into directory /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS. Masses are (1250, 1085). Decompressing /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS. Found hepmpfile at cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet.1250_1085.13.hepmp.gz to cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet/temp/T2ttoff_1250_1085.13.hepmp.gz

```

files: 86% (40/46)
      : 95% (44/46)
Updating files: 100% (46/46), done.

[CutLangWrapper] ADLLHC Analyses initialisation finished.
[CutLangWrapper] Delphes initialised.
[CutLangWrapper] Initialisation complete.
[CutLangWrapper] Writing output into directory /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS.
[CutLangWrapper] Masses are (1250, 1085)
[CutLangWrapper] Decompressing /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS.
[CutLangWrapper] Found hepmpfile at cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet.1250_1085.13.hepmp.gz to cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet/temp/T2ttoff_1250_1085.13.hepmp.gz
[CutLangWrapper] exec: /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS/src/emcreator02/delphes/DelphesHepMC2 /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS/cards/delphes_card_CMS.tcl cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet/temp/T2ttoff_1250_1085.13.hepmp.gz --delphes --port 2555
** INFO: adding module Efficiency
** INFO: adding module ChargedHadron
** INFO: adding module Efficiency
  
```

Madgraph result

2. Start delphes with madgraph result

```

[CutLangWrapper] exec: cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet/temp/T2ttoff_1250_1085.root DELPHES -i /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS/CMS-SUS-19-006_CutLang.adl
/home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS/src/emcreator02/delphes/DelphesHepMC2 /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS/cards/delphes_card_CMS.tcl cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet/temp/T2ttoff_1250_1085.13.hepmp.gz --delphes --port 2555
/home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS/src/emcreator02/delphes/DelphesHepMC2 /home/condor/dir_28426/CMSSW_12_3_0_pre6/ANALYSIS/ANALYSIS/cards/delphes_card_CMS.tcl cutlang_results/CMS-SUS-19-006/ANA_T2ttoff_1jet/temp/T2ttoff_1250_1085.13.hepmp.gz --delphes --port 2555
Analysis with Multiple Regions
A region, ID: 0 named: [prese1]
A region, ID: 1 named: [searchbins]
  this region depends on: prese1
A region, ID: 2 named: [ASR1]
  
```

3. Using the delphes MC results, CutLang starts running the analysis

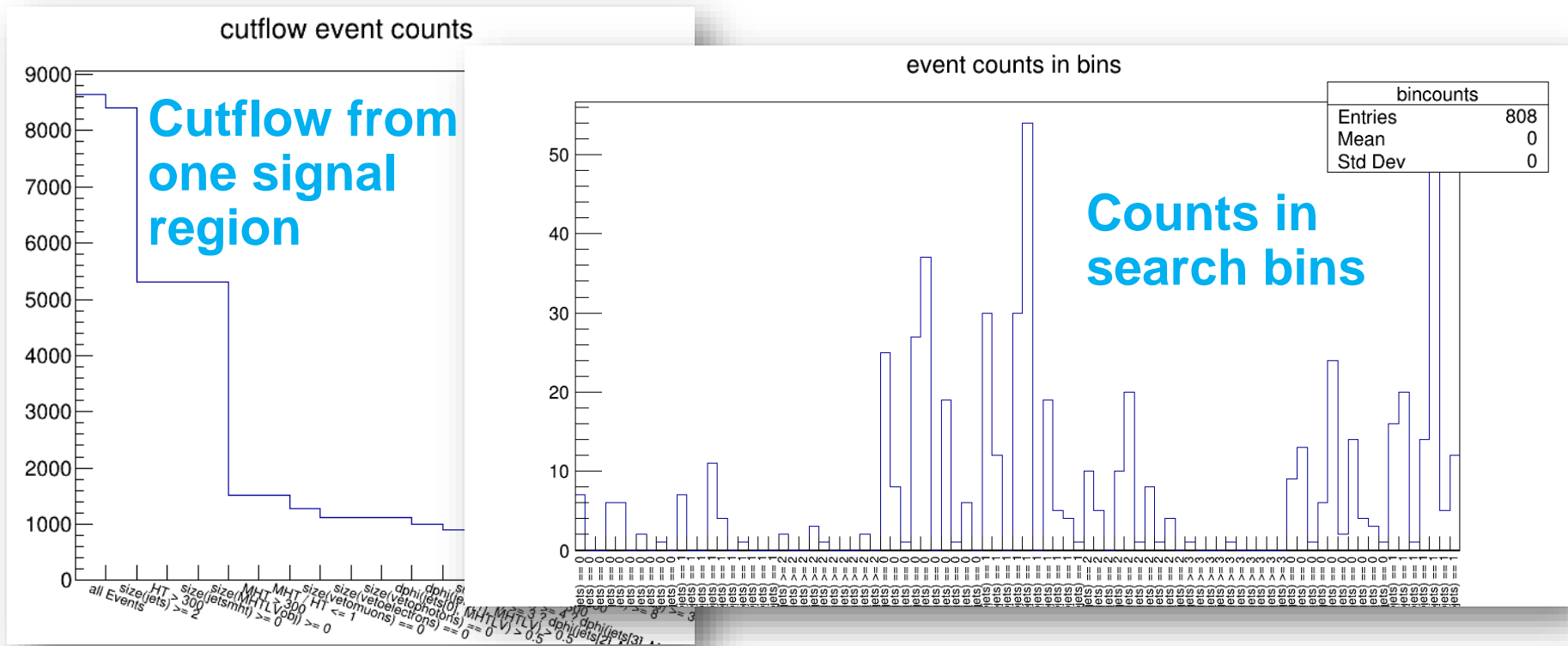
```

ASR12 Based on 8410 events:
size(jets) >= 2 : 0.9738 +- 0.00172 evt: 8410
HT > 300 : 0.6303 +- 0.00526 evt: 5301
size(jetsmht) >= 0 : 1 +- 0 evt: 5301
size(MHTLVobj) >= 0 : 1 +- 0 evt: 5301
MHT > 300 : 0.2877 +- 0.00622 evt: 1525
MHT / HT <= 0.888 : 0.0091 evt: 1507
size(vetomuons) == 0 : 0.854 +- 0.0091 evt: 1287
size(vetoelectrons) == 0 : 0.8772 +- 0.00915 evt: 1129
size(vetophotons) == 0 : 0.9056 +- 0.0091 evt: 1124
s[0], MHTLV > 0.5 : 0.9991 +- 0.000889 evt: 1123
s[1], MHTLV > 0.5 : 0.8878 +- 0.00942 evt: 997
MHTLV > 0.3 : ALL : 0.9137 +- 0.00889 evt: 911
MHTLV > 0.3 : ALL : 0.8902 +- 0.0104 evt: 811
MHT >= 850 : 0.05179 +- 0.00778 evt: 42
HT >= 850 : 1 +- 0 evt: 42
  
```

Selection cutflow by CutLang

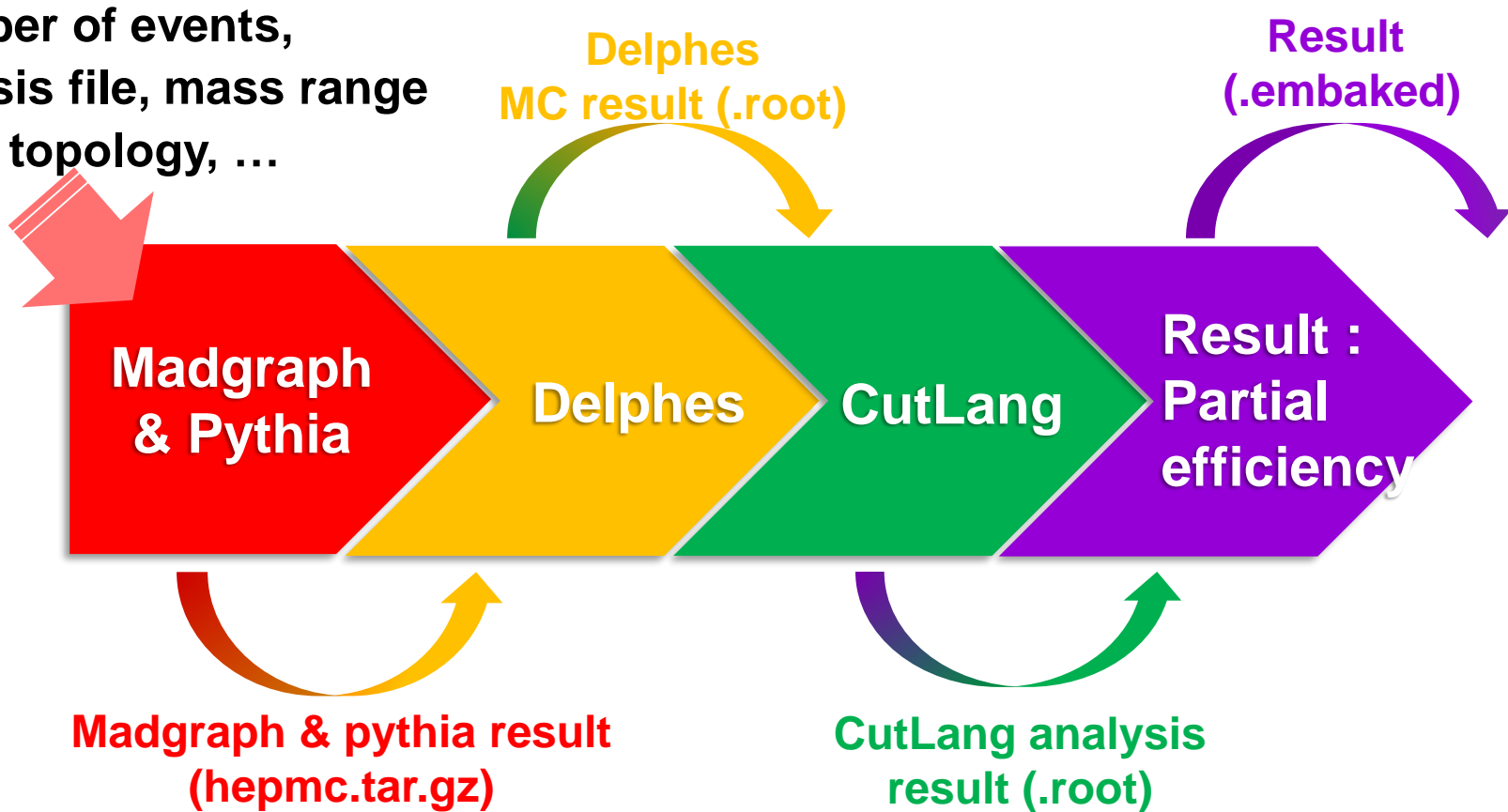
Results for (1250, 1085) mass point

- Mass point (1250, 1085)'s **CutLang** result
 - histoOut-CMS-SUS-19-006_CutLang.root



EM-creator workflow

Number of events,
ADL analysis file, mass range
SMS topology, ...



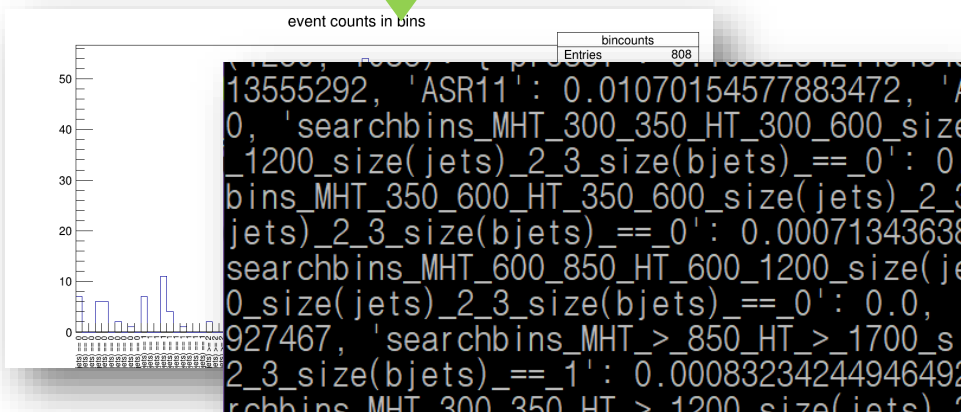
- This setup is highly object-oriented and factorizable. Each component can be run individually.

EM-creator operation using HTCondor

- The em-creator is very **easy to divide** and execute works **for each mass point**
- Suitable for use with HTCondor.
We set up EM-creator in the KNU T3 HTCondor system.
- Example)
 - Apply a mass range like **-m “[(250,2201,50),(10,2001,25)]”** to em-creator **261 mass points** come out
 - One can **submit 261 jobs** for **each mass point**.

Result of em-creator with HTCondor

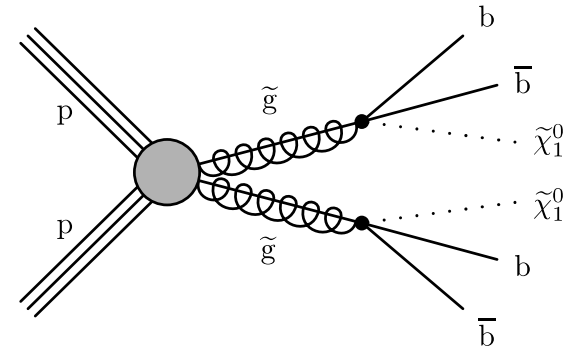
```
knu 4096 Nov 18 10:44 100Mass(950-810)
knu 4096 Nov 18 10:28 101Mass(950-835)
knu 4096 Nov 18 10:27 102Mass(950-860)
knu 4096 Nov 18 10:30 106Mass(1000-835)
knu 4096 Nov 18 10:25 107Mass(1000-860)
knu 4096 Nov 18 10:14 108Mass(1000-885)
knu 4096 Nov 18 10:14 109Mass(1000-910)
knu 4096 Nov 18 10:35 113Mass(1050-885)
knu 4096 Nov 18 10:24 114Mass(1050-910)
knu 4096 Nov 18 10:22 115Mass(1050-935)
knu 4096 Nov 18 10:36 116Mass(1050-960)
knu 4096 Nov 18 10:38 120Mass(1100-935)
knu 4096 Nov 18 10:44 121Mass(1100-960)
knu 4096 Nov 18 10:48 122Mass(1100-985)
```



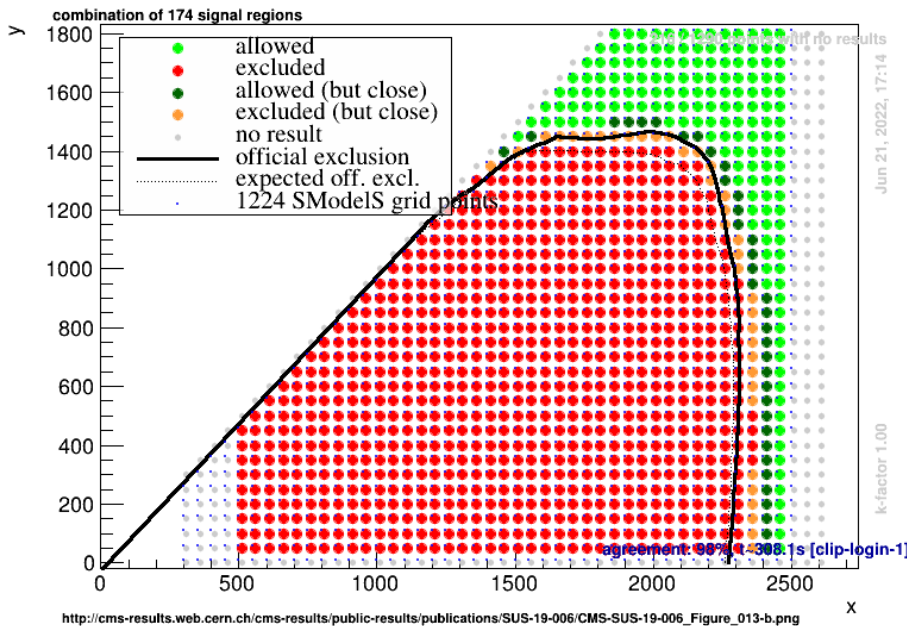
- The directory corresponding to each mass point has **CutLang result** and **efficiency result** files.
- We can **validate numerous models very effectively and conveniently through EM-creator!**

Final goal

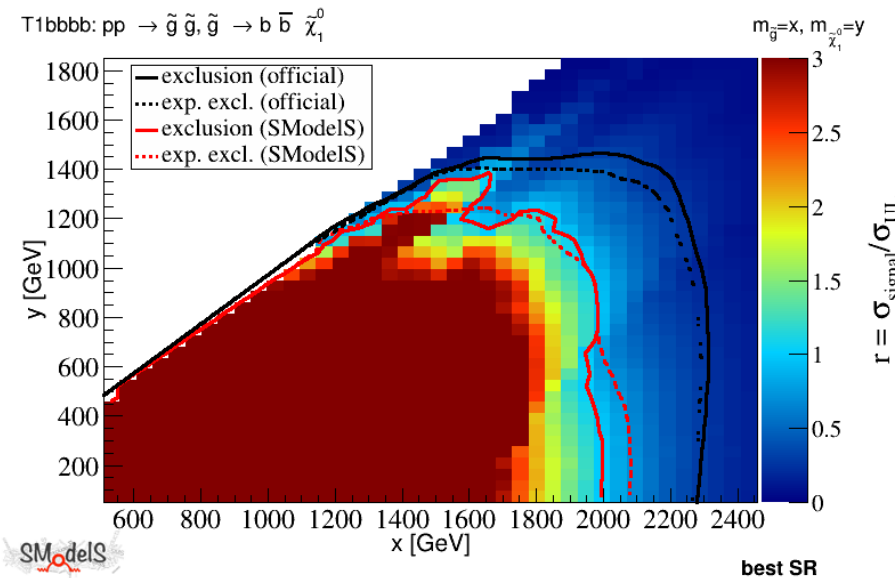
- **Efficiency maps created for T1bbbb for CMS-SUS-19-006 by Wolfgang at the Vienna system. We will achieve this at KNU.**



CMS-SUS-19-006-adl_T1bbbb_[[x, y], [x, y]]



CMS-SUS-19-006-adl (efficiencyMap)



Summary

- **SmodelS EM-creator was adapted to ADL/CutLang** by Jan, Gökhan and Wolfgang.
- We **successfully installed** and **tested** the **EM-creator** in KNU T3 HTCondor system.
- We set up **each job to run a single point**.
(One job for 10K evts from MG to CL took ~40 minutes)
- We **adapted the scripts to start from any stage in the chain**, e.g. only run the CutLang and efficiency bake.
- We will perform one complete run using CMS-SUS-19-006 to produce efficiencies.

Next steps

- **Coordinate with the SModelS team to set up a system to calculate limits from the em-baked efficiencies.**
- **Perform more robustness checks in our setup.**
- **Start an organized large scale validation of several analyses.**
- **Move towards publishing the EM-creator for public use.**