Outline

❖ Introduction
  - Briefly Normal Mode & Expert Mode

❖ LHC recasting with MadAnalysis 5
  - Improving exclusion limits statistics with likelihood
  - Particle propagation module for long-lived particles

❖ Conclusion

❖ Hands-on session
Introduction
Why designing & recasting is important?

- Exploiting the full potential of LHC (for new physics)
  - Designing new analyses (based on MC simulations)
  - Recasting LHC analyses (The LHC legacy)
- Data preservation in HEP is mandatory
  - Going beyond raw data via analyses
- Related tools need to be supported by the entire community
  - Both theorists & experimentalists
- Universal recasting tool

Les Houches Recommendations (EPJC '12)
Reinterpretation Forum Report (SciPost '20)
MadAnalysis 5

What is MadAnalysis 5?

- A framework for **phenomenological analyses**
- **Any level of sophistication:** partonic, hadronic, detector, reconstructed
- **Several input formats:** STDHEP, HEPMC, LHE, LHCO, ROOT (from Delphes)
- **User-friendly, flexible & Fast!!!**
- Interfaces several HEP packages: MadGraph, FastJet, Delphes, pyhf

**Normal Mode**
- Intuitive commands typed in the Python interface
- Analysis performed **behind the scenes** (black box)
- Human readable output: HTML and LaTeX

**Expert Mode**
- C++ programming with the SampleAnalyzer framework
- Support for multiple sub-analyses, an efficient way for handling cuts and histograms, etc.
LHC recasting with MadAnalysis 5
Signal Events
(STDHEP or HEPMC format)

Detector Simulation

Recast

Exclusion limit calculation

Validation

Public Analysis Database++;

SFS@Ma5 or Delphes

Cut-flows uncertainties

TH uncertainties, HL extrapolations simplified and full likelihoods

Reimplementing an analysis in MadAnalysis 5

JYA, Fuks, Polykratis EPJC ‘21

De Favereau et al; JHEP ‘14

JYA, Frank, Fuks EPJC ‘20

Algusero, JYA, Fuks, Kraml; in progress

Signal Events

Detector Simulation

Recast

Exclusion limit calculation

Validation

Public Analysis Database++;

SFS@Ma5 or Delphes

TH uncertainties, HL extrapolations simplified and full likelihoods

C++

Cut-flows uncertainties

Reimplementing an analysis in MadAnalysis 5

JYA, Fuks, Polykratis EPJC ‘21

De Favereau et al; JHEP ‘14

JYA, Frank, Fuks EPJC ‘20

Algusero, JYA, Fuks, Kraml; in progress
Reimplementing an analysis in MadAnalysis 5

- Calculating **exclusion limits**, **expected** and **observed excluded cross sections** via uncorrelated signal regions.

- **Improved limits** via full statistical models constructed from HistFactory-like likelihoods (ATLAS)

- **Improved limits** via simplified likelihoods with CMS’ correlation matrices

- **Exclusion limits with theoretical uncertainties & higher luminosity extrapolations**
Reimplementing an analysis in MadAnalysis 5

Implementation

Blackbox

Detector Simulation

SFS@Ma5 or Delphes

Recast

ma5> set main.recast = on

Signal Events (STDHEP or HEPMC format)

New theory Lagrangian

Public Analysis Database

ma5> install PAD

16 ATLAS & 16 CMS analyses at $\sqrt{s} = 13$ TeV

Yes

Is excluded?

No

Yay!

ma5> submit

Stress free exclusion limits

Dumont, Fuks, Kraml, et. al. EPJC '15

Jack Y. Araz - MadAnalysis 5
More on
Theoretical uncertainties & HL extrapolations

\( \sqrt{s} = 13 \text{ TeV} \)

**pp \rightarrow gg**

- \( \text{mad\text{-}set defaultset.xsection} = 1.8524785535e-08 \)
- \( \text{mad\text{-}set defaultset.scale\_up\_variation} = 0.388 \)
- \( \text{mad\text{-}set defaultset.scale\_down\_variation} = 0.262 \)
- \( \text{mad\text{-}set defaultset.pdf\_up\_variation} = 0.131 \)
- \( \text{mad\text{-}set defaultset.pdf\_down\_variation} = 0.131 \)

**ATLAS-SUSY-2016-07**

- \( \text{Obs. Limit from } M_{H_{125}} \text{ - based searches} \)
  - @ NNLO_{approx} + NLL :
    - 95% CL_{excl} Exclusion Limit
    - 95% CL_{excl} with Scale & PDF Variation
    - @ NLO :
      - 95% CL_{excl} Exclusion Limit
      - 95% CL_{excl} with Scale Variation
      - 95% CL_{excl} with Scale & PDF Variation
More on Theoretical uncertainties & HL extrapolations

Extrapolated 2016 results

2019 results

$pp \to \tilde{g}\tilde{g}, \text{@NLO } BR(\tilde{g} \to q\bar{q}\chi^0) = 100\%$

$\sqrt{s} = 13$ TeV, 139 fb$^{-1}$

ATLAS-CONF-2019-040:
- 95\% CL_{exp} Exclusion Limit
- 95\% CL_{exp} with Scale & PDF Variation
- ATLAS-SUSY-2016-007

$pp \to \tilde{g}\tilde{g}, \text{@NLO, } BR(\tilde{g} \to q\bar{q}\chi^0) = 100\%$

$\sqrt{s} = 13$ TeV, 3000 fb$^{-1}$

ATLAS-SUSY-2016-007
- ATLAS-CONF-2019-040
Why important?

- The mathematical description of the analysis is provided within its statistical model.

- The likelihood profile enables the standard statistical approaches to extract information.

- *i.e.* how reasonably aligned the theoretical predictions with the experimental observations?

SModelS: Alguero, Kraml, Waltenberg ’20

ATLAS SUSY and Exotics workshop
S. Kraml ’20

Les Houches Recommendations (EPJC ’12)

CMS-NOTE-2017-001

Full likelihoods from ATLAS

ATL-PHYS-PUB-2019-029

Simplified likelihoods from CMS

For details see the “Publication of statistical models: hands on workshop”

ma5> install pyhf
Correlation matrix allows us to form a covariance matrix which then used in a simplified likelihood under Gaussian approximation.

\[ \mathcal{L}_S(\mu, \theta) = \prod_{i=1}^{N} \frac{(\mu s_i + b_i + \theta_i)^{n_i} e^{-(\mu s_i + b_i + \theta_i)}}{n_i!} \exp \left( -\frac{1}{2} \theta^T V^{-1} \theta \right) \]

\[ V = \mathbb{E}[\theta_i \times \theta_j] \]
Electroweakino searches

- Wino-like electroweakinos
- Choosing the most sensitive signal region with respect to expected exclusion cross section.
Electroweakino searches

- Simplified likelihoods with CMS-SUS-16-039
- MadAnalysis 5
- Observed exclusion limit with best SR

CMS Supplementary arXiv:1709.05406 35.9 fb⁻¹ (13 TeV)

44 Signal Regions!
Electroweakino searches

- Wino-like electroweakinos
- ATLAS shares HistFactory like json files to form full profile likelihoods.
- Each file includes detailed information on backgrounds and corresponding nuisance parameters.
- Simplified likelihoods are achieved by compressing all the information into a single nuisance parameter.
More on Simplified & full likelihoods in action!

Electroweakino searches

- Wino-like electroweakinos
- ATLAS shares HistFactory like json files to form full profile likelihoods.
- Each file includes detailed information on backgrounds and corresponding nuisance parameters.
- Simplified likelihoods are achieved by compressing all the information into single nuisance parameter.

ATLAS–SUSY–2019–08
\( \sqrt{s} = 13 \text{ TeV}, 139 \text{ fb}^{-1}, 95\% \text{ CL} \)

ATL-PHYS-PUB-2021-038
Simplify module has been used for orange curve:

Alguero, JYA, Fuks, Kraml; 2206.14870

Jack Y. Araz - MadAnalysis 5
Simplified likelihoods and pyhf interface has been fully integrated in MadAnalysis 5 v1.10.4

- ATLAS - SUSY - 2018 - 31
- ATLAS - SUSY - 2018 - 04
- ATLAS - SUSY - 2018 - 08
- ATLAS - SUSY - 2018 - 06
- ATLAS - SUSY - 2018 - 14
- CMS - SUS - 2016 - 39
- CMS - SUS - 2017 - 01
- CMS - SUS - 2019 - 06
- CMS - EXO - 20 - 004

By Andreas Albert; see his talk at RAMP#3

Alguero, JYA, Fuks, Kraml; 2206.14870
The goal of reinterpretation is not just reproducing the experimental results but testing the new theories we come up with in the shower!

<table>
<thead>
<tr>
<th>$1 - \text{CL}_S \text{ (obs)}$</th>
<th>$M_2 = 600$ GeV</th>
<th>$M_2 = 800$ GeV</th>
<th>$M_2 = 1$ TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS best-SR combined</td>
<td>$0.71$</td>
<td>$0.66$</td>
<td>$0.94$</td>
</tr>
<tr>
<td>ATLAS best-SR combined</td>
<td>$0.70$</td>
<td>$0.59$</td>
<td>$0.91$</td>
</tr>
<tr>
<td>ATLAS best-SR combined</td>
<td>$0.29$</td>
<td>$0.21$</td>
<td>$0.57$</td>
</tr>
<tr>
<td>CMS best-SR combined</td>
<td>$0.83$</td>
<td>$0.80$</td>
<td>$0.98$</td>
</tr>
<tr>
<td>CMS best-SR combined</td>
<td>$0.84$</td>
<td>$0.74$</td>
<td>$0.97$</td>
</tr>
<tr>
<td>CMS best-SR combined</td>
<td>$0.80$</td>
<td>$0.56$</td>
<td>$0.92$</td>
</tr>
</tbody>
</table>

Original analysis: $pp \to \tilde{t}_1 \tilde{t}_1$ and $\tilde{b}_1 \tilde{b}_1$; $m_{\tilde{t}_1} = 60$ GeV

$\sqrt{s} = 13$ TeV, $95\%$ CL

CMS-SUS-19-006:
- MadAnalysis 5—exclusion limit with best signal region
- MadAnalysis 5—exclusion limit with simplified likelihood

ATLAS-SUSY-2018-31:
- MadAnalysis 5—exclusion limit with best signal region
- MadAnalysis 5—exclusion limit with full statistical model

$M_2 :=$ Wino mass

$M_{\tilde{Q}_3} :=$ left squark soft-mass for the third generation
Particle propagation in SFS

❖ SFS module allows for simple observable smearing based on transfer functions.

❖ Transverse impact parameter \( (d_0) \) and longitudinal impact parameter \( (d_z) \) can be calculated with straight trajectory assumption (default behaviour for other recasting softwares).

❖ Modification of particle trajectories under constant magnetic field can provide relevant effects for unusual particle signatures.
Particle propagation in SFS

Decay product

Production/decay vertex

Closest approach

$d_0$

$d_z$

Beam Axis

llp

$|d_z|$
Step I) Validation of the SFS module with particle propagator

Existing recast from 2018 adapted to the SFS with particle propagation module.

Improved track based isolation cones.

Very scarce validation material.

No available statistics!
CMS-EXO-16-022: Impact of the particle propagator

\[ \delta d_0 = \frac{d_0 - d_0^{\text{approx}}}{d_0^{\text{approx}}} \]

\( m_{\tilde{t}_1} = 20 \text{ GeV} \)
More displacement on the decay vertex

\( m_{\tilde{t}_1} = 700 \text{ GeV} \)
Simplified & full likelihoods in action!

Particle propagator module and fresh LLP recasts are available in MadAnalysis 5 - v1.9_beta

- ATLAS - SUSY - 2017 - 04
- CMS - EXO - 19 - 010
- CMS - EXO - 16 - 022
- CMS - EXO - 18 - 003

For details see the MadAnalysis talk at “Tenth workshop of the LLP Community”

\[ c \tau > 1 \text{ m} \]
Conclusion
Conclusion

❖ Particle propagation can have a significant effect on analysis outcome, depending on the theory behind it.

❖ Uncorrelated signal regions do not represent the statistical model of the analysis well enough. Full or simplified statistical models are essential for better reinterpretation.

NEW MadAnalysis 5 is fully capable of using correlation matrices and full likelihood profiles to improve exclusion limits.

NEW The particle propagation module is available with MadAnalysis v1.9 alongside various LLP recasts.

WANTED: Analysis codes

Scientific reproducibility and data preservation solely depend on preserving analysis logic in a reinterpretable form. You can contribute to the HEP community by sharing the LHC recast you have implemented in the MadAnalysis 5 framework through Public Analysis Database! Please send us your analysis code, detector card, info file and validation note to be included in PAD for public use.

More information and examples can be found in the proceedings of the second MadAnalysis 5 Workshop on LHC recasting in Korea. Analysis codes have been published, documented and got a DOI so that they can now be cited.
Hands on!
MadAnalysis 5: Expert Mode

User Inputs
- Script and/or Commands for reconstruction
- Event Samples

Back-end
- Libraries
  - SampleAnalyzer
  - FastJet*
  - SFS
  - Delphes 3*
  *To be installed from the Python interface

Analysis code
- Job
- Execution
- SAF files

Reports & Output
- HTML
- LATEX

Coming soon!!

ma5-expert 2.0.4
pip install ma5-expert

Cacciari, Salam, Soyez; EPJC '11
JYA, Fuks, Polycratis; EPJC '21
De Favereau et al; JHEP '14

Jack Y. Araz - MadAnalysis 5
Boosted top tagging

- With the increased boost factor, jets (top decay products) are getting more collimated.
- Hadronic top tagging tools: Mass grooming and filtering, Pruning, Trimming, Soft Drop Tagger, Mass Drop Tagger, HEPTopTagger, Machine Learning
HEPTopTagger on $t\bar{t}$ events

Hadronic Top Reconstruction
HEPTopTagger reconstructs 3-prong jet-substructure where

$$m_{123} \approx 172 \pm 40 \text{ [GeV]}$$

**Diagram Explanation**

**Step 1:**
Mass drop decomposition

**Step 2:**
Loop over all combinations of 3 mass drop subjets

**Step 3:**
Recluster with $R_{\text{filt}} = \min(0.3, \Delta R_{\text{min}}/2)$

**Step 4:**
Filtering: keep only the 5 leading subjets

**Step 5:**
Repeat reclustering and filtering procedure for all combinations of 3 mass drop subjets

**Step 6:**
Pick the combination with filtered mass closest to the top mass. Recluster to force 3 subjets

**On-shell $t\bar{t}$ events**

**The Standard Model**

$$m_W/m_t$$

$$m_{12}/m_{123} = m_W/m_t$$

**Plehn, Spannowsky, Takeuchi; PRD ’11**

**Marzani, Gregory, Spannowsky ’19**
HEPTopTagger is looking for the resonance signature of W and top, which can be seen from the Dalitz plot.