CMS Policies and Practises on providing information for reinterpretations

José Ruiz
on behalf of CMS Collaboration
This talk will focus on BSM searches at CMS
How to get away with a search

An analysis team

1,000 papers
An analysis team

Data analysis

Background modelling

Statistical analysis

Private documentation

Analysis validation

Physics group
EXO-SUS-B2G

Collaboration (as a whole)

Public paper

Results review

Rigorous and long process

Simplistic description
Depending on many factors it can take roughly one or more years to make and publish and analysis. Analysis teams as well as the coordination areas are doing a continuous effort to get results in shorter timescales as well as in the most usable way by the community.
Communicating two worlds

Theory

Experiment
Communicating two worlds

Theory

Experiment
CMS policies

What is a CMS result (paper):

1. Strategy and procedures description.
2. Figures, tables, numbers:
   a. Histograms
   b. Estimations (e.g. the trigger efficiency is at least 99% for signal like events)
   c. Total counts
   d. Efficiencies.
   e. Covariance matrix.
   f. Acceptance maps.
   g. Limits
3. Statistical interpretation

What is not (usually) in a CMS public result:

1. Montecarlo samples
2. Data
3. Data analysis code
4. Technical recipes/specifications
5. Internal review details
From a CMS paper (experiments in general)

- Difficult to reproduce. Some important details are often private or it is not feasible to redo them outside the collaboration.
- Numbers are fetchable, histograms not so much: using histograms for reinterpretations imply to postprocess them and convert them to numbers, if they are delivered only as figures.
- Statistical analysis and interpretation: typically the statistical procedure can have details that are very technical and the statistic routines could be difficult to use without guidance.
- For new signals a whole set of approximations have to be done as detector effects are not really possible to calculate from outside the collaboration.
Improving communication of results

Different approaches:

1. Detector unfolding.
2. Supplemental material.
3. Open data.
4. Providing likelihoods.
Detector unfolding

Events \[\rightarrow\] Unfolded events
RIVET toolkit (Robust Independent Validation of Experiment and Theory)

Rivet analysis coverage

Rivet analyses exist for 976/5812 papers = 17%. 183 priority analyses required.

Total number of Inspire papers scanned = 10551, at 2020-11-26

Breakdown by identified experiment (in development):

<table>
<thead>
<tr>
<th>Key</th>
<th>ALICE</th>
<th>ATLAS</th>
<th>CMS</th>
<th>LHCb</th>
<th>Forward</th>
<th>HERA</th>
<th>e⁺e⁻ (≥ 12 GeV)</th>
<th>e⁺e⁻ (≤ 12 GeV)</th>
<th>Tevatron</th>
<th>RHIC</th>
<th>SPS</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rivet wanted</td>
<td>236</td>
<td>287</td>
<td>435</td>
<td>202</td>
<td>43</td>
<td>519</td>
<td>733</td>
<td>617</td>
<td>1238</td>
<td>464</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>(total)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rivet REALLY</td>
<td>36</td>
<td>38</td>
<td>79</td>
<td>8</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>wanted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rivet provided</td>
<td>26/262</td>
<td>167/454</td>
<td>89/524</td>
<td>16/218</td>
<td>8/51</td>
<td>9/528</td>
<td>193/926</td>
<td>381/998</td>
<td>59/1297</td>
<td>8/472</td>
<td>5/56</td>
<td>18/19</td>
</tr>
<tr>
<td></td>
<td>= 10%</td>
<td>= 37%</td>
<td>= 17%</td>
<td>= 7%</td>
<td>= 16%</td>
<td>= 2%</td>
<td>= 21%</td>
<td>= 38%</td>
<td>= 5%</td>
<td>= 2%</td>
<td>= 8%</td>
<td>= 95%</td>
</tr>
</tbody>
</table>
Provides:
- Analysis code.
- Validation framework.
- Analysis results.
CMS and RIVET

- Typically an effort pursued mainly by generators group.
- Specially interesting for the collaboration in order to validate Montecarlo generators, as it provides unfolded event.
- It would require some non-negligible amount of work to implement each analysis and making it available in RIVET.
- It has been difficult to establish a policy to enforce such implementation for every CMS analysis. Coordination has recommended it to the collaboration and the analysis teams are aware of the importance and very often try to make the effort to implement it. Technical difficulties usually overcome this effort.
Measurement of the Z boson differential production cross section using its invisible decay mode (Z→ν̅ν) in proton-proton collisions at √s = 13 TeV

The CMS collaboration: Sirunyan, Albert M ; Tumasyan, Armen ; Adam, Wolfgang ; et al.


Measurements of the total and differential fiducial cross sections for the Z boson decaying into two neutrinos are presented at the LHC in proton-proton collisions at a center-of-mass energy of 13 TeV. The data were collected by the CMS detector in 2016 and correspond to an integrated luminosity of 35.9 fb$^{-1}$. In these measurements, events are selected containing an imbalanced...
Provides:
Analysis results in a more suitable format for recasting. Common framework for different sources.
CMS and HEPData

- Has become an official and necessary step in order to make an analysis public. No analysis should be made public, at least in BSM, without the accompanying HEPData entry.
- All official plots and tables in the paper have to be added to the HEPData entry.
- Lightweight work, compared to other tools: It usually takes a reasonable amount of work to “translate” results into HEPData format. Additionally there is always an expert which helps preparing the HEPData entry and validating the process.
- Beyond tables and figures: CMS also usually provides efficiencies, acceptance maps, covariance/correlation matrices, and simplified likelihoods.
- Automatization code: [Link]
Getting Started with CMS 2010 Open Data

To analyse CMS data collected in 2010, you need version 4.2.8 of CMSSW, supported only on Scientific Linux 5. If you are unfamiliar with Linux, take a look at this short introduction to Linux or try this...

Getting Started with CMS 2011 Open Data

To analyse CMS data collected in 2011 and 2012, you need version 5.3.32 of CMSSW, supported only on Scientific Linux 6. If you are unfamiliar with Linux, take a look at this short introduction to Linux or try this...

CMS Guide to generate configuration files for event production

Guide to produce CMSSW python configuration files to simulate, generate and/or reconstruct collision events.
Provides:
Full reproducibility.
Datasets.
Montecarlo events.
Code.
Providing public likelihoods

- Since 2017 a proposal for simplified likelihoods for public release has been made by the collaboration: Link.
- So far there is no official requirement to publish full likelihoods for CMS analyses.
- However, there is a motivation to move along that direction and use the simplified likelihoods approach.
- Some steps have been taken towards having a common recipe to publish likelihoods and some internal discussion has taken place moreover the last year.
- More work is needed to overcome the technical difficulties to get CMS analyses up to this point.
Beyond publishing

CMS is also interested to have a fluent and continuous communication with the theory community:

1. **Theory talks at working groups:** Invited talks given by a theorist to the plenary meeting of the different physics groups.
2. **CMS theorists:** Theorists officially linked with CMS that can look at private information and advice analysis teams.
3. **Working groups:** Special working groups for specific topics which are thought to need a more intricate development with theory colleagues, e.g. EFT forum.
Conclusions

1. Several policies have been implemented within CMS to assure a better communication of analysis results, mainly thinking towards an easier and more precise way reusing CMS results.
2. Also practices have been set and tools have been put in place in order to have a more broad application within the analysis teams.
3. A culture of sharing the analysis results and “making the effort” has been spread within the collaborations in the last years.
4. More works need to be done, specially to achieve full likelihood publishing.
Thank you!