Higgs-to-four-lepton analysis example using the ReANA platform: Demo

Diyaselis Delgado López
University of Puerto Rico Mayagüez

UPR
Recinto Universitario de Mayagüez

reana

1 Input data

What is your input data?

- input files
- input parameters

2 Analysis code

Which code analyses it?

- software frameworks
- user code

3 Compute environment

What is your environment?

- operating system
- database calls

4 Analysis workflow

Which steps did you take?

- single command
- complex workflows

Analysis structure

- This example studies the Higgs-to-four-lepton decay channel that led to the Higgs boson experimental discovery.
- Uses CMS Open Data data released in 2011 and 2012 and available CMS Software environment

What is our input data?

- CMS collision data: /DoubleMuParked/Run2012C-22Jan2013-v1/AOD
- CMS simulated data: /SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUgenV3-pythia6/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM

Which code analyses it?

- Process the original collision data using demoanalyzer_cfg_level3data.py

```
import FWCore.ParameterSet.Config as cms
from RecoMuon.TrackingTools.MuonServiceProxy_cff import *
import FWCore.PythonUtilities.LumiList as LumiList
import FWCore.ParameterSet.Types as CfgTypes
process = cms.Process("Demo")

# intialize MessageLogger and output report
process.load("FWCore.MessageLogger.MessageLogger_cfi")
process.MessageLogger.cerr.threshold = 'INFO'
process.MessageLogger.categories.append('Demo')
process.MessageLogger.cerr.INFO = cms.untracked.PSet(
        limit = cms.untracked.int32(-1)
        )
process.options = cms.untracked.PSet( wantSummary = cms.untracked.bool(True) )
process.maxEvents = cms.untracked.PSet( input = cms.untracked.int32(-1) )

# define JSON file for 2012 data
goodJSON = '../../../inputs/Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt'

myLumis = LumiList.LumiList(filename = goodJSON).getCMSSWString().split(',')

# define the input data set here by inserting the appropriate .txt file list *
import FWCore.Utilities.FileUtils as FileUtils
process.source = cms.Source("PoolSource",
    fileNames = cms.untracked.vstring(
         'root://eospublic.cern.ch//eos/opendata/cms/Run2012C/DoubleMuParked/AOD/22Jan2013-v1/10000/F2878994-766C-E211-8693-E0CB4
    )
)

# apply JSON file
#   (needs to be placed *after* the process.source input file definition!)
process.source.lumisToProcess = CfgTypes.untracked(CfgTypes.VLuminosityBlockRange())
process.source.lumisToProcess.extend(myLumis)

process.source.skipEvents = cms.untracked.uint32(0)

process.demo = cms.EDAnalyzer('HiggsDemoAnalyzerGit'
)

process.TFileService = cms.Service("TFileService",
```

Which code analyses it?

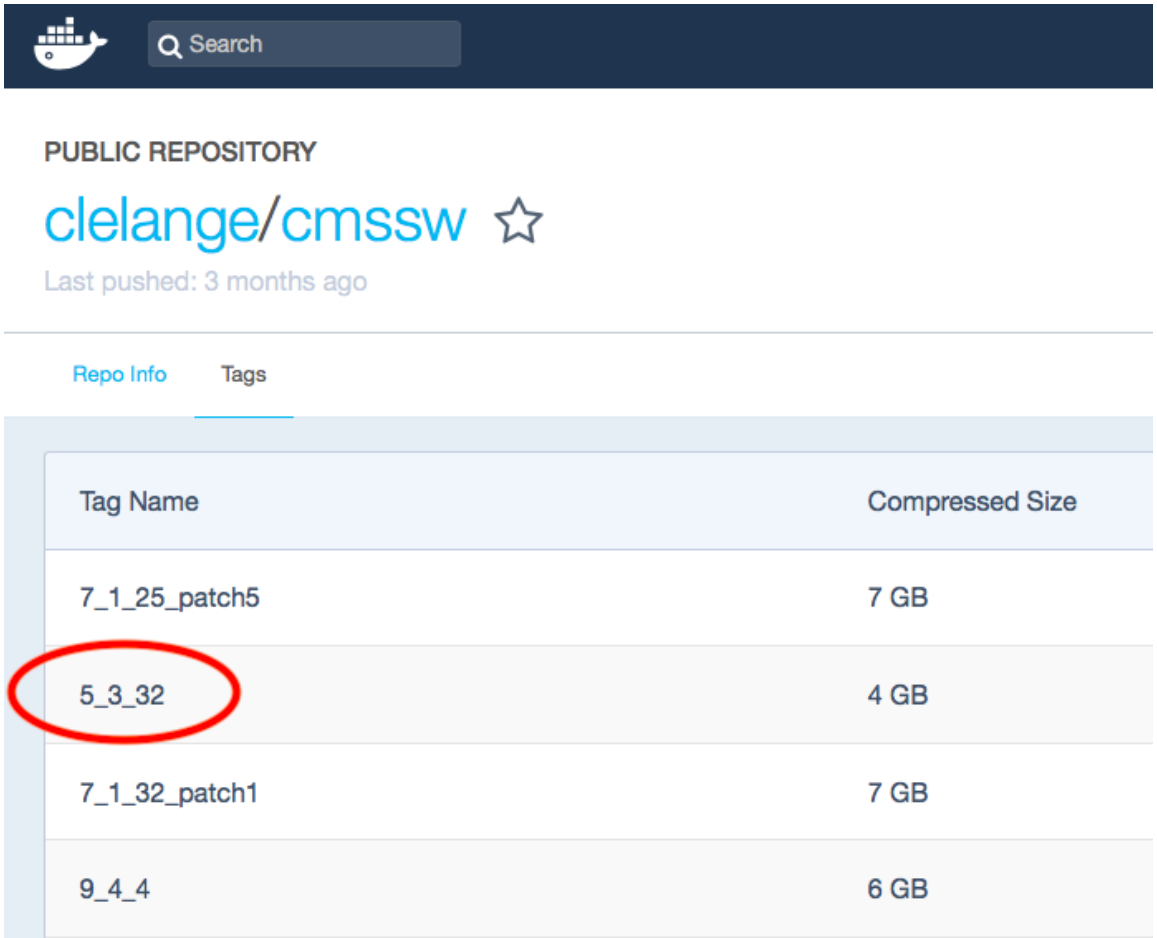
- Process the simulated data using demoanalyzer_cfg_level3MC.py

```
import FWCore.ParameterSet.Config as cms
from RecoMuon.TrackingTools.MuonServiceProxy_cff import *
import FWCore.PythonUtilities.LumiList as LumiList
import FWCore.ParameterSet.Types as CfgTypes
process = cms.Process("Demo")


# intialize MessageLogger and output report
process.load("FWCore.MessageLogger.MessageLogger_cfi")
process.MessageLogger.cerr.threshold = 'INFO'
process.MessageLogger.categories.append('Demo')
process.MessageLogger.cerr.INFO = cms.untracked.PSet(
        limit = cms.untracked.int32(-1)
        )
process.options = cms.untracked.PSet( wantSummary = cms.untracked.bool(True) )
# ----------------------------------------------------------------------
# set the maximum number of events to be processed                     *
#    this number (argument of int32) is to be modified by the user     *
#    according to need and wish                                        *
#    default is preset to -1 (all events)                              *
# ----------------------------------------------------------------------
process.maxEvents = cms.untracked.PSet( input = cms.untracked.int32(-1) )

# set the number of events to be skipped (if any) at end of file below

# define JSON file for 2012 data (not needed for MC)
#goodJSON = '../../../inputs/Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt'

#myLumis = LumiList.LumiList(filename = goodJSON).getCMSSWString().split(',')

# ----------------------------------------------------------------------------
# define the input data set here by inserting the appropriate .txt file list *
# ----------------------------------------------------------------------------
import FWCore.Utilities.FileUtils as FileUtils
```

```
# ----------------------------------------------------------------------
# load the data set                                                    *
# this example uses the 2012 Higgs->4lepton simulated dataset          *
# ----------------------------------------------------------------------
#
# *** 2012 Higgs->4lepton simulated data set (299973 events) ***
#files2012data = FileUtils.loadListFromFile ('/home/cms-opendata/CMSSW_5_3_32/src/Demo/DemoAnalyzer/MCsets/CMS_M
#process.source = cms.Source("PoolSource",
#    fileNames = cms.untracked.vstring(*files2012data
#    )
#)
#
# to speed up, read only first file with 7499 events
process.source = cms.Source("PoolSource",
    fileNames = cms.untracked.vstring(
        'root://eospublic.cern.ch//eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/SMHiggsToZZTo4L_M-125_8TeV-pow
    )
)

# apply JSON file (not for MC)
#   (needs to be placed *after* the process.source input file definition!)
#process.source.lumisToProcess = CfgTypes.untracked(CfgTypes.VLuminosityBlockRange())
#process.source.lumisToProcess.extend(myLumis)

# ------------------------------------------------
# number of events to be skipped (0 by default)  *
# ------------------------------------------------
process.source.skipEvents = cms.untracked.uint32(0)


process.demo = cms.EDAnalyzer('HiggsDemoAnalyzerGit'
)
# ------------------------------------------------------------
# output file name                                           *
# default is Higgs4L1file.root                               *
# ------------------------------------------------------------
process.TFileService = cms.Service("TFileService",
      fileName = cms.string('../../../outputs/Higgs4L1file.root')
                                  )

process.p = cms.Path(process.demo)
```

What is the environment?

- Encapsulate the current compute environment preparing a Docker container image for our analysis steps



Which steps were taken?

- Workflow #1: input.yaml

```yaml
inputs:
  class: Directory
  path: ../inputs
code:
  class: Directory
  path: ../code
```

- Workflow #2: step1data.cwl

```yaml
cwlVersion: v1.0
class: CommandLineTool

baseCommand: /bin/zsh

requirements:
  DockerRequirement:
    dockerPull:
      clelange/cmssw:5_3_32
  InitialWorkDirRequirement:
    listing:
      - $(inputs.code)
      - $(inputs.inputs)

inputs:
  inputs:
    type: Directory
  code:
    type: Directory

stdout: step1data.log

outputs:
  step1data.log:
    type: stdout
  DoubleMuParked2012C_10000_Higgs.root:
    type: File
    outputBinding:
      glob: "outputs/DoubleMuParked2012C_10000_Higgs.root"

arguments:
  - prefix: -c
    valueFrom: |
      cp -r ../../code/HiggsExample20112012 .; \
      scram b; \
      cd ../../code/HiggsExample20112012/Level3; \
      mkdir -p ../../../outputs; \
      cmsRun demoanalyzer_cfg_level3data.py
```

Which steps were taken?

- Workflow #3: step1mc.cwl

```
cwlVersion: v1.0
class: CommandLineTool

baseCommand: /bin/zsh

requirements:
  DockerRequirement:
    dockerPull:
      clelange/cmssw:
  InitialWorkDirRequi
    listing:
      - $(inputs.code
      - $(inputs.inpu

inputs:
  inputs:
    type: Directory
  code:
    type: Directory

stdout: step1mc.log
```

```
outputs:
  step1mc.log:
    type: stdout
  Higgs4L1file.root:
    type: File
    outputBinding:
      glob: "outputs/Higgs4L1file.root"

arguments:
  - prefix: -c
    valueFrom: |
      cp -r ../../code/HiggsExample20112012 .; \
      scram b; \
      cd ../../code/HiggsExample20112012/Level3; \
      mkdir -p ../../../outputs; \
      cmsRun demoanalyzer_cfg_level3MC.py
```

- Workflow #4: step2.cwl

```
cwlVersion: v1.0
class: CommandLineTool

baseCommand: /bin/zsh

requirements:
  DockerRequirement:
    dockerPull:
      clelange/cmssw:5_3_32
  InitialWorkDirRequirement:
    listing:
      - $(inputs.code)
      - $(inputs.inputs)

inputs:
  inputs:
    type: Directory
  code:
    type: Directory
  DoubleMuParked2012C_10000_Higgs:
    type: File
  Higgs4L1file:
    type: File

stdout: step2.log
```

```
outputs:
  step2.log:
    type: stdout
  mass4l_combine_userlvl3.pdf:
    type: File
    outputBinding:
      glob: "outputs/mass4l_combine_userlvl3.pd

arguments:
  - prefix: -c
    valueFrom: |
      cp -r ../../code/HiggsExample20112012 .;
      cd ../../code/HiggsExample20112012/Level3
      mkdir -p ../../../outputs; \
      cp $(inputs.DoubleMuParked2012C_10000_Hig
      cp $(inputs.Higgs4L1file.path) ../../../o
      root -b -l -q ./M4Lnormdatall_lvl3.cc
```

Which steps were taken?

- Workflow #5: workflow.cwl

```
#!/usr/bin/env cwl-runner

cwlVersion: v1.0
class: Workflow

requirements:
  InitialWorkDirRequirement:
    listing:
      - $(inputs.code)
      - $(inputs.inputs)

inputs:
  inputs:
    type: Directory
  code:
    type: Directory

outputs:
  mass4l_combine_userlvl3.pdf:
    type: File
    outputSource:
      step2/mass4l_combine_userlvl3.pdf

steps:
  step1data:
    run: step1data.cwl
    in:
      code: code
      inputs: inputs
    out: [DoubleMuParked2012C_10000_Higgs.root
  step1mc:
    run: step1mc.cwl
    in:
      code: code
      inputs: inputs
    out: [Higgs4L1file.root, step1mc.log]
  step2:
    run: step2.cwl
    in:
      code: code
      inputs: inputs
      DoubleMuParked2012C_10000_Higgs: step1data/DoubleMuParked2012C_10000_Higgs.root
      Higgs4L1file: step1mc/Higgs4L1file.root
    out: [mass4l_combine_userlvl3.pdf, step2.log]
```

- Workflow #6: reana.yaml

```
version: 0.3.0
inputs:
  files:
  - code/HiggsExample20112012/HiggsDemoAnalyzer/src/HiggsDemoAnalyzerGit
  - code/HiggsExample20112012/Level3/demoanalyzer_cfg_level3data.py
  - code/HiggsExample20112012/Level3/demoanalyzer_cfg_level3MC.py
  - code/HiggsExample20112012/Level3/M4Lnormdatall_lvl3.cc
  parameters:
     input: workflow/input.yaml
workflow:
  type: cwl
  file: workflow/workflow.cwl
environments:
  - type: docker
  image: clelange/cmssw:5_3_32
outputs:
  files:
   - results/mass4l_combine_userlvl3.pdf
```

Testing CWL workflows

```
bash-3.2$ cd
bash-3.2$ ls
CMS             Downloads       Music           VirtualBox VMs
Desktop         Library         Pictures        cwl
Documents       Movies          Public
bash-3.2$ cd CMS/reana-demo-cms-h4l/
bash-3.2$ mkdir cwl-local-run
bash-3.2$ cd cwl-local-run/
bash-3.2$ cp -a ../code/ ../workflow/input.yaml .
bash-3.2$ source ~/cwl/bin/activate
(cwl) bash-3.2$ cwltool --quiet --outdir="../results" ../workflow/workflow.cwl input.yaml
WARNING: In non-interactive mode release checks e.g. deprecated releases, production architectures are disabled.
181025 08:56:34 001 Xrd: XrdClientConn: Error resolving this host's domain name.
25-Oct-2018 08:56:34 CEST  Initiating request to open file root://eospublic.cern.ch//eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/SMHiggsToZZ
53_V19-v1/10000/029D759D-6CD9-E211-B3E2-1CC1DE041FD8.root
%MSG-i XrdFileInfo:  file_open 25-Oct-2018 08:56:34 CEST pre-events
Opened root://eospublic.cern.ch//eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUgenV3-pythia6/AODSIM/PU_
.root
%MSG
%MSG-i XrdFileInfo:  file_open 25-Oct-2018 08:56:34 CEST pre-events
Connection URL root://p05798818w87152.cern.ch:1095//eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUgenV3
1-B3E2-1CC1DE041FD8.root
%MSG
25-Oct-2018 08:56:37 CEST  Successfully opened file root://eospublic.cern.ch//eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/SMHiggsToZZTo4L_M-
v1/10000/029D759D-6CD9-E211-B3E2-1CC1DE041FD8.root
%MSG-i Root_Information:  AfterFile TClass::TClass() 25-Oct-2018 08:56:38 CEST  pre-events
no dictionary for class pair<edm::IndexIntoFile::IndexRunKey,Long64_t> is available
%MSG
%MSG-i Root_Information:  AfterFile TClass::TClass() 25-Oct-2018 08:56:38 CEST  pre-events
no dictionary for class pair<edm::IndexIntoFile::IndexRunLumiKey,Long64_t> is available
%MSG
%MSG-i Root_Information:  AfterFile TClass::TClass() 25-Oct-2018 08:56:38 CEST  pre-events
no dictionary for class pair<edm::BranchKey,edm::ConstBranchDescription> is available
%MSG
%MSG-i Root_Information:  AfterFile TClass::TClass() 25-Oct-2018 08:56:38 CEST  pre-events
no dictionary for class pair<edm::BranchID,unsigned int> is available
%MSG
Begin processing the 1st record. Run 1, Event 180901, LumiSection 604 at 25-Oct-2018 08:56:45.944 CEST
%MSG-i Demo:  HiggsDemoAnalyzerGit:demo 25-Oct-2018 08:56:45 CEST  Run: 1 Event: 180901
Starting to analyze
Event number: 180901, Run number: 1, Lumisection: 604
```

Testing CWL workflows

```
=============================================

MessageLogger Summary

 type     category        sev    module        subroutine        count    total
 ---- -------------------- -- ---------------- ----------------  -----    -----
    1 fileAction           -s file_close                             1        1
    2 fileAction           -s file_open                              2        2

 type    category    Examples: run/evt        run/evt          run/evt
 ---- -------------------- ---------------- ---------------- ----------------
    1 fileAction           PostEndRun
    2 fileAction           pre-events       pre-events

Severity    # Occurrences   Total Occurrences
--------    -------------   -----------------
System                  3                   3
WARNING: In non-interactive mode release checks e.g. deprecated releases, production architectures are disabled.
Info in <TCanvas::Print>: pdf file ../../../outputs/mass4l_combine_userlvl3.pdf has been created
{
    "mass4l_combine_userlvl3.pdf": {
        "location": "file:///Users/diyaselis/CMS/reana-demo-cms-h4l/results/mass4l_combine_userlvl3.pdf",
        "basename": "mass4l_combine_userlvl3.pdf",
        "class": "File",
        "checksum": "sha1$9d47e21cf57842346580add4deeed0a0bc4727f7",
        "size": 18138,
        "path": "/Users/diyaselis/CMS/reana-demo-cms-h4l/results/mass4l_combine_userlvl3.pdf"
    }
}
(cwl) bash-3.2$ ls -lh ../results/
total 40
-rw-r--r--  1 diyaselis  wheel    18K Oct 25 09:06 mass4l_combine_userlvl3.pdf
(cwl) bash-3.2$ open ../results/*
```

Run analysis within the ReANA platform → *Instructions*

```
diyaselis:~ diyaselis$ pip3 install export virtualenv
Collecting export
  Downloading https://files.pythonhosted.org/packages/39/ab/83cf
Collecting virtualenv
  Downloading https://files.pythonhosted.org/packages/b6/30/96a0
    100% |████████████████████████████████| 1.9MB 5.5MB/s
Installing collected packages: export, virtualenv
Successfully installed export-0.1.2 virtualenv-16.0.0
```

```
diyaselis:~ diyaselis$ virtualenv ~/.virtualenvs/myreana
Using base prefix '/usr/local/Cellar/python/3.7.0/Frameworks/Python.framework/Versions/3.7'
New python executable in /Users/diyaselis/.virtualenvs/myreana/bin/python3.7
Also creating executable in /Users/diyaselis/.virtualenvs/myreana/bin/python
Installing setuptools, pip, wheel...done.
diyaselis:~ diyaselis$ source ~/.virtualenvs/myreana/bin/activate
(myreana) diyaselis:~ diyaselis$ pip install reana-client
Collecting reana-client
  Downloading https://files.pythonhosted.org/packages/8c/b4/1fd9ef50219b73d278870ed6c862fd78
Collecting cwltool==1.0.20180912090223 (from reana-client)
  Downloading https://files.pythonhosted.org/packages/af/ac/669b251930e0cc8d728fdfa79b7b3d4b
    100% |████████████████████████████████| 3.2MB 3.8MB/s
Collecting pyOpenSSL==17.3.0 (from reana-client)
```

```
(myreana) diyaselis:~ diyaselis$ export REANA_SERVER_URL=http://reana-qa.cern.ch/
(myreana) diyaselis:~ diyaselis$ export REANA_ACCESS_TOKEN=lYamW6S1rX5wbrP2zFo-SY6
(myreana) diyaselis:~ diyaselis$
(myreana) diyaselis:~ diyaselis$ reana-client ping
Connected to http://reana-qa.cern.ch/ - Server is running.
```

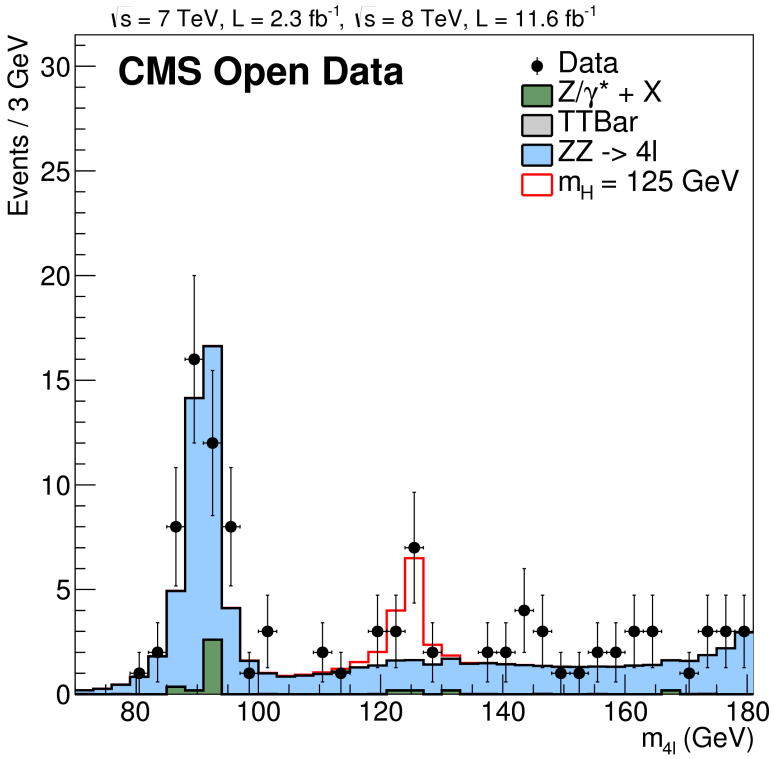
```
(myreana) diyaselis:~ diyaselis$ cd CMS/reana-demo-cms-h4l/
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ cd ..
(myreana) diyaselis:CMS diyaselis$ rm -rf reana-demo-cms-h4l/
(myreana) diyaselis:CMS diyaselis$ git clone git@github.com:diyaselis/reana-demo-cms-h4l.git
Cloning into 'reana-demo-cms-h4l'...
Enter passphrase for key '/Users/diyaselis/.ssh/id_rsa':
remote: Enumerating objects: 72, done.
remote: Counting objects: 100% (72/72), done.
remote: Compressing objects: 100% (69/69), done.
remote: Total 142 (delta 39), reused 3 (delta 1), pack-reused 70
Receiving objects: 100% (142/142), 1.33 MiB | 349.00 KiB/s, done.
Resolving deltas: 100% (50/50), done.
(myreana) diyaselis:CMS diyaselis$ cd reana-demo-cms-h4l/
```

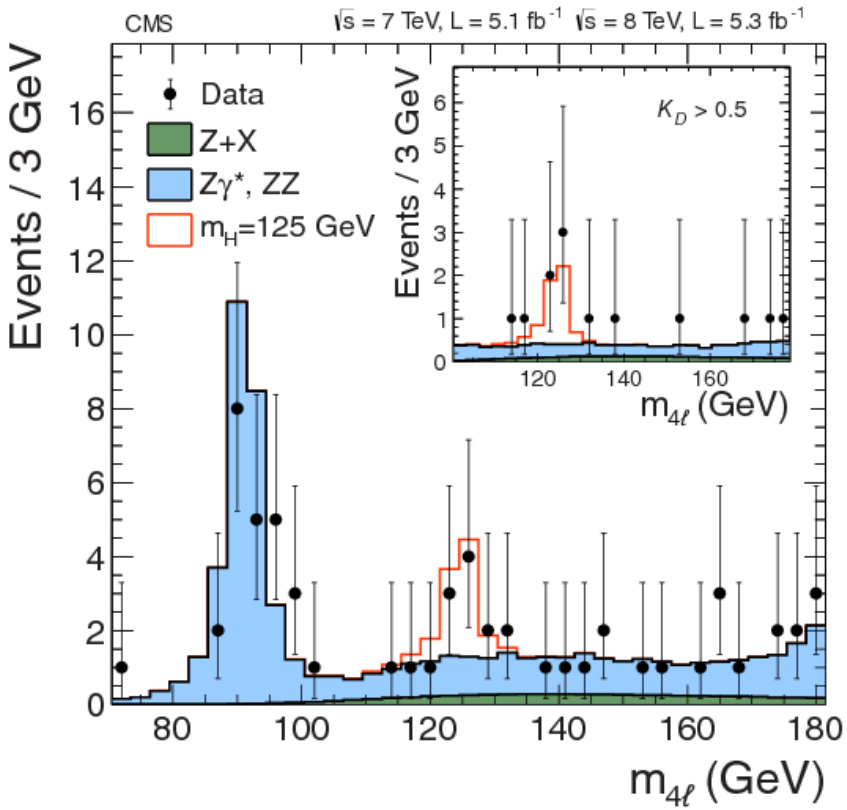
```
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ vi reana.yaml
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ reana-client create -f reana.yaml
workflow.1
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ reana-client status -w workflow.1
NAME       RUN_NUMBER   CREATED              STATUS    PROGRESS
workflow   1            2018-10-25T08:02:24  created   -/-
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ export REANA_WORKON=workflow.1
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ reana-client upload ./code
File code/HiggsExample20112012/Level3/demoanalyzer_cfg_level3data.py was successfully uploaded.
File code/HiggsExample20112012/Level3/M4Lnormdatall_lvl3.cc was successfully uploaded.
File code/HiggsExample20112012/Level3/demoanalyzer_cfg_level3MC.py was successfully uploaded.
File code/HiggsExample20112012/HiggsDemoAnalyzer/BuildFile.xml was successfully uploaded.
File code/HiggsExample20112012/HiggsDemoAnalyzer/src/HiggsDemoAnalyzerGit.cc was successfully uploaded.
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ reana-client list
NAME                                                                SIZE    LAST-MODIFIED
code/HiggsExample20112012/Level3/demoanalyzer_cfg_level3data.py     3649    2018-10-25T08:06:14
code/HiggsExample20112012/Level3/M4Lnormdatall_lvl3.cc              15913   2018-10-25T08:06:14
code/HiggsExample20112012/Level3/demoanalyzer_cfg_level3MC.py       3705    2018-10-25T08:06:14
code/HiggsExample20112012/HiggsDemoAnalyzer/BuildFile.xml           305     2018-10-25T08:06:14
code/HiggsExample20112012/HiggsDemoAnalyzer/src/HiggsDemoAnalyzerGit.cc  83797  2018-10-25T08:06:15
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ reana-client upload ./data
File data/DY1011.root was successfully uploaded.
File data/DoubleMu12.root was successfully uploaded.
File data/ZZ2mu2e12.root was successfully uploaded.
File data/TTBar12.root was successfully uploaded.
File data/TTJets11.root was successfully uploaded.
File data/DY101Jets12.root was successfully uploaded.
File data/ZZ4e12.root was successfully uploaded.
File data/DoubleE12.root was successfully uploaded.
File data/DY50TuneZ11.root was successfully uploaded.
File data/Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt was successfully uploaded.
```

```
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ reana-client start
workflow.1 has been started.
(myreana) diyaselis:reana-demo-cms-h4l diyaselis$ reana-client status
NAME       RUN_NUMBER   CREATED              STATUS     PROGRESS
workflow   1            2018-10-25T08:02:24  finished   0/3
```

WAITING FOR UPDATE

Results

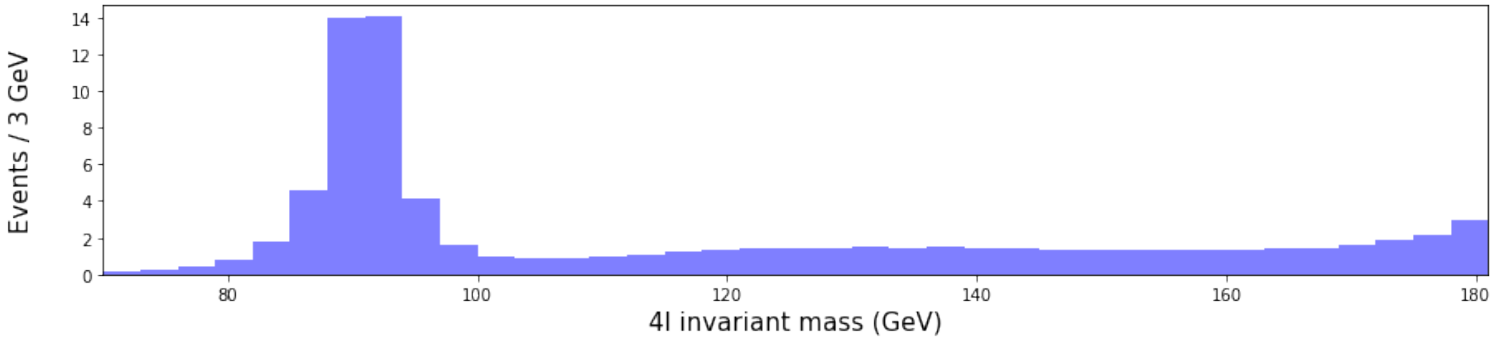

√s = 7 TeV, L = 2.3 fb-1, √s = 8 TeV, L = 11.6 fb-1
CMS Open Data
Data
Z/γ* + X
TTBar
ZZ -> 4l
mH = 125 GeV
Events / 3 GeV
m4l (GeV)


CMS
√s = 7 TeV, L = 5.1 fb-1 √s = 8 TeV, L = 5.3 fb-1
Data
Z+X
Zγ*, ZZ
mH=125 GeV
KD > 0.5
Events / 3 GeV
m4ℓ (GeV)

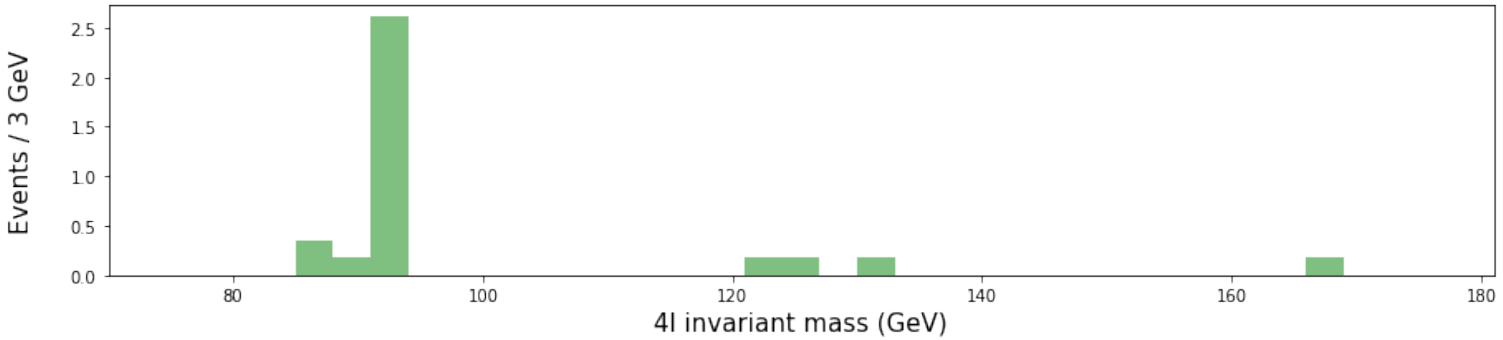
Let's look at some Monte Carlo -simulated values that have already been weighted by luminosity, cross-section and number of events.

What we'll measure in the accelerator:

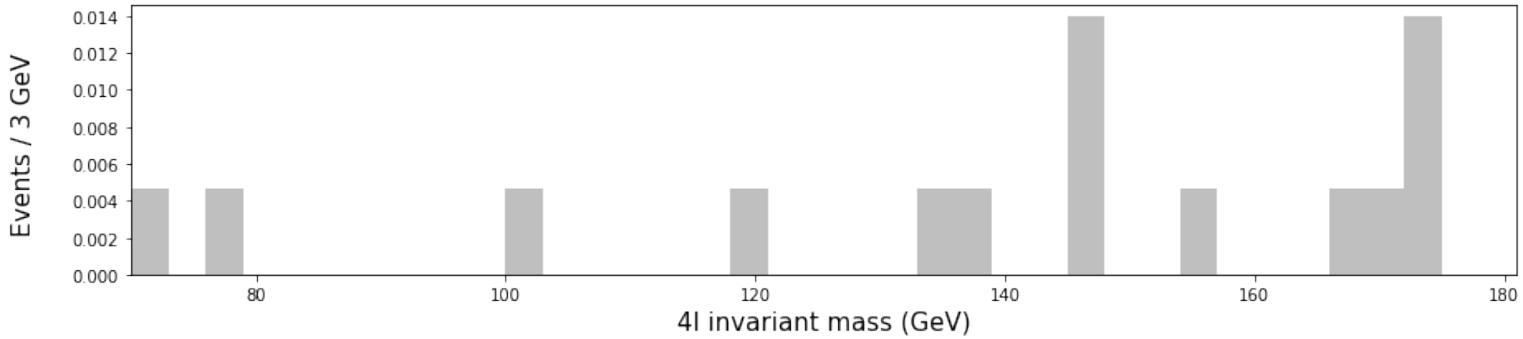
- ZZ, a pair of heavier bosons

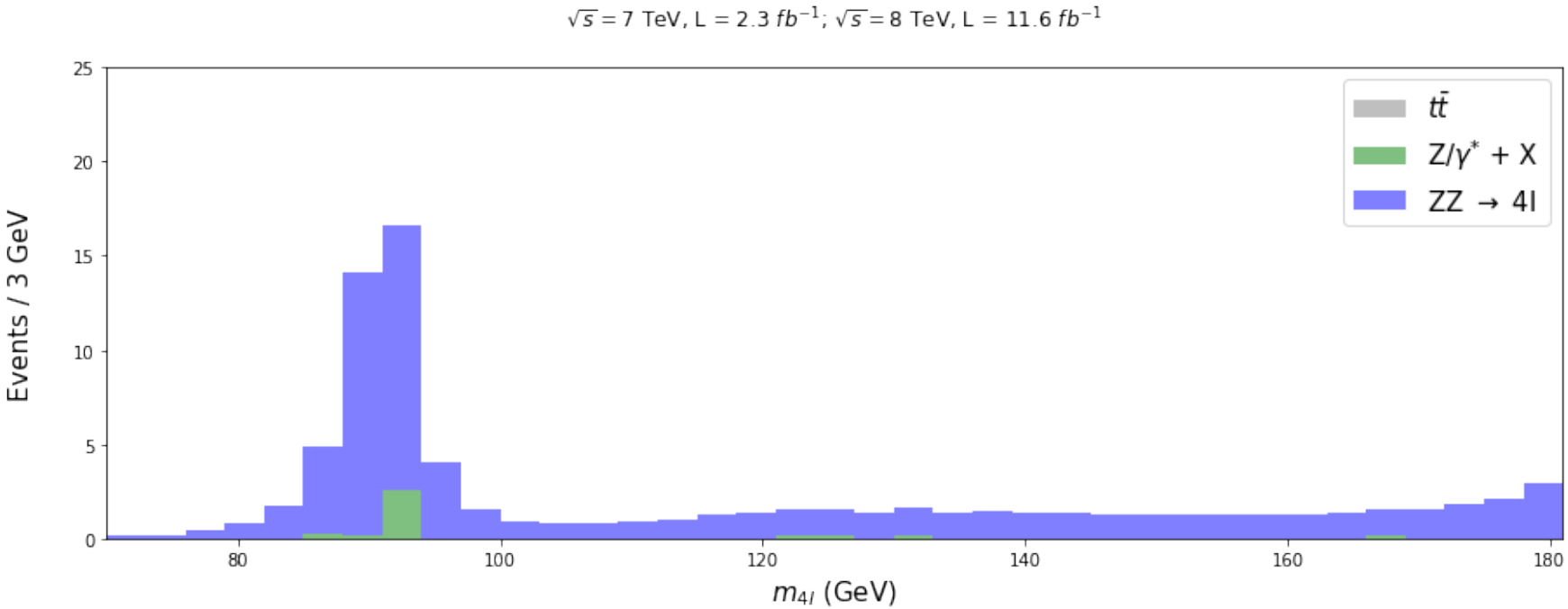


- DY, some irreducible background from singular Z bosons

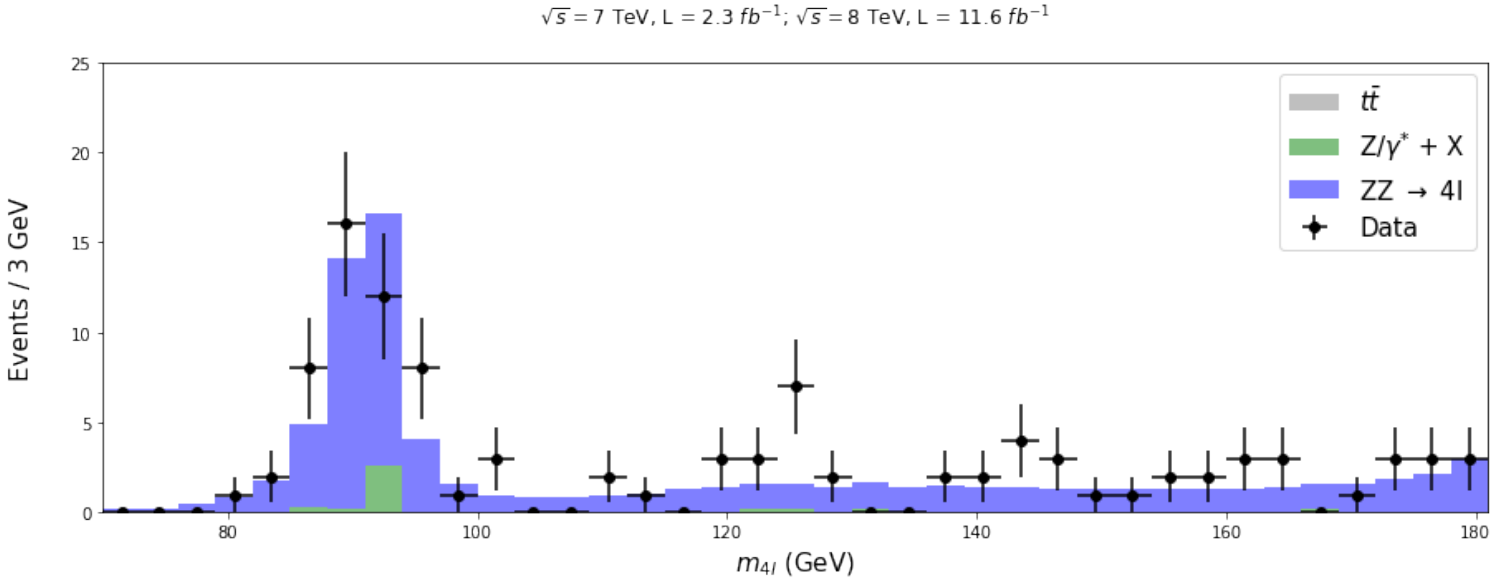


- ttbar, a pair of top and anti-top quarks



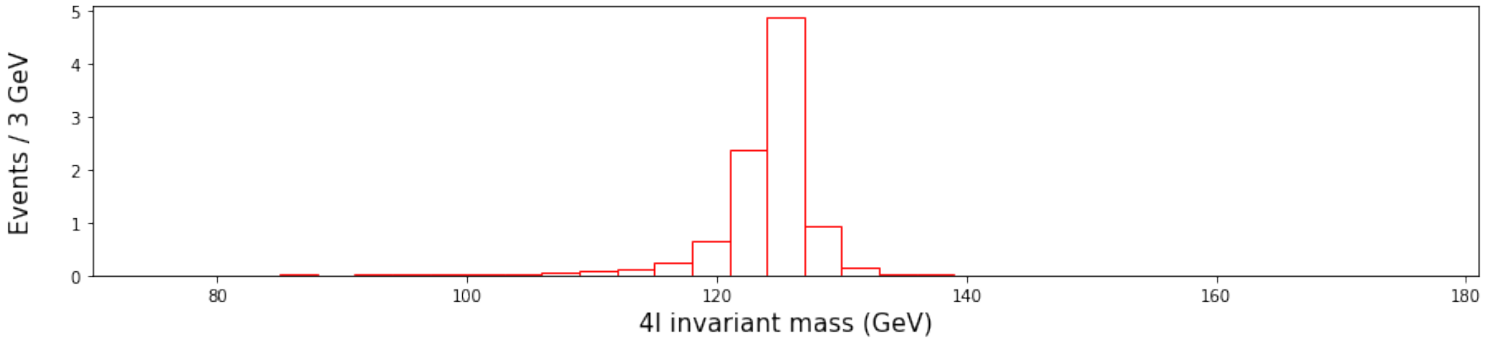


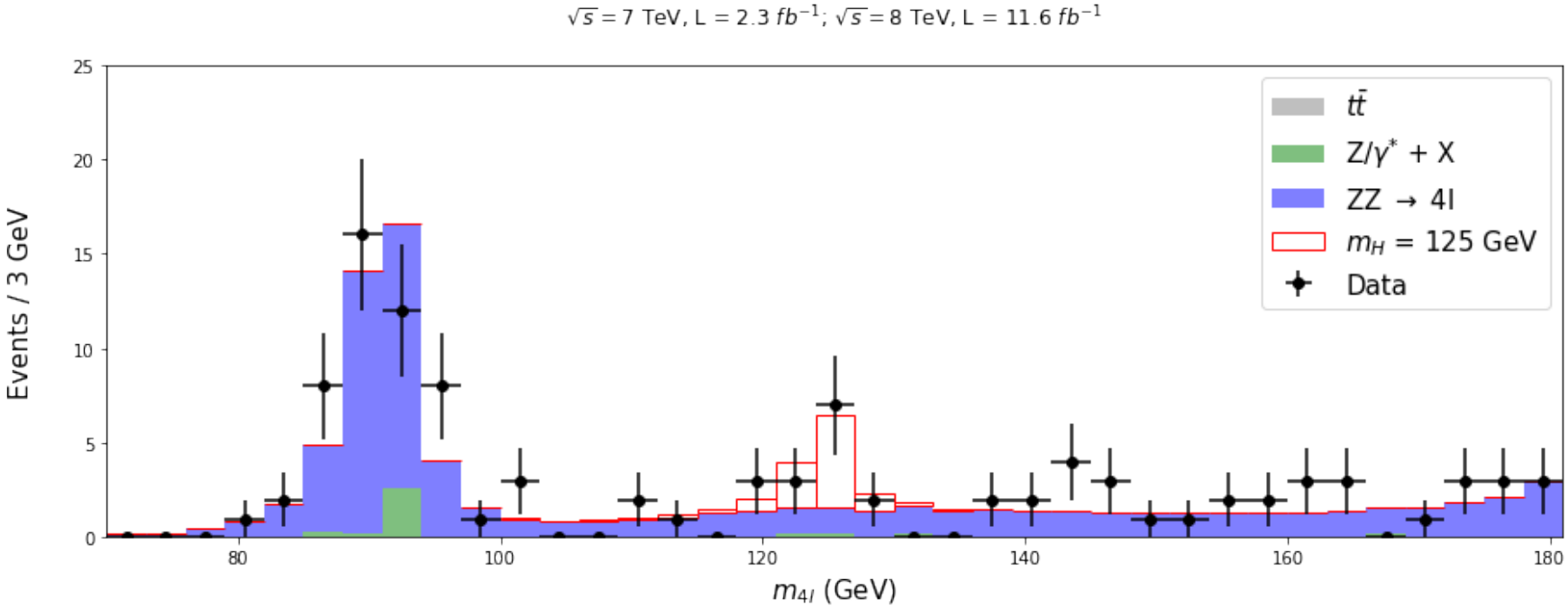
Let's add our measured data on top of that



This graph shows what the Higgs boson should look like, if it had a mass of 125 GeV.

- HZZ, our theoretical assumption of a Higgs via two Z bosons





This data set is too small to say anything for certain, but it isn't too far off from actual analysis results.