MadAnalysis5
A framework for event file analysis.

Benjamin Fuks (IPHC Strasbourg / Université de Strasbourg).

In collaboration with E. Conte & G. Serret.

MadGraph 2011 Meeting @ Academia Belgica (Roma).
September 20-23, 2011.

Outline.

1. Quick overview.
2. Initialization and displaying information.
3. Event files.
4. Performing an analysis.
5. Current to-do list.

MadAnalysis 5 in a nutshell.

- **Scope**.
  - * **Analysis of event files** produced by Monte Carlo generators.
  - * Definition of various **selection cuts** on the input samples.
  - * Production of **histograms** of different distributions.
- **Programming languages used by** MadAnalysis.
  - * User interface written in Python and ROOT.
  - * LHEAnalyzer: core program written in C++, using ROOT, Boost.
  - * Python **version**: 2.6 and higher, no 3.x.
  - * ROOT **version**: 5.28 and higher.
- **Additional features**.
  - * Output in ROOT, HTML, LaTeX.
  - * To be used **within** MadGraph **5 or as a standalone package**.
- **Wiki page**: `https://server06.fynu.ucl.ac.be/projects/madanalysis`
  - * **Demo version** related to this talk: `http://server06.fynu.ucl.ac.be/downloads/madanalysis-current.tar.gz`
  - * **Please send us your comments and suggestions** (tickets on the wiki).

Disclaimer.

- **The program is far from being achieved**.
  - Please be patient.
  - Quality before speed!
- **Is your favorite feature missing?**
  - Ask.
  - Wait (our to-do list is still huge).
  - You will be advertised when it will be implemented.
- **No manual presently**.

Outline.

1. Quick overview.
2. Initialization and displaying information.
3. Event files.
4. Performing an analysis.
5. Current to-do list.

Initialization.

- **Download the tar-ball.**

```
http://server06.fynu.ucl.ac.be/downloads/madanalysis-current.tar.gz
```

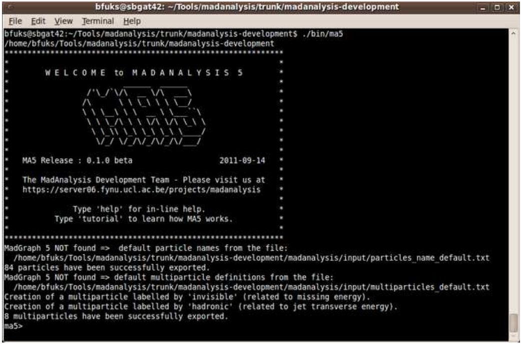
- **Two different ways of usage.**
  - * **In the context of** MadGraph.
    - ► Unpack the tar-ball within your MadGraph directory.
  - * **As a standalone** Python **module.**
    - ► Unpack the tar-ball anywhere.

- **Please check that** ROOT **is installed.**
  - * The directory `root/bin` must be included in the `$PATH` variable.
    ⇒ **Then, users do not have to worry about** ROOT **anymore.**

- **Lauching the program.**

```
./bin/ma5
```

Initialization.



bfuks@sbgat42: ~/Tools/madanalysis/trunk/madanalysis-development
File Edit View Terminal Help

```
bfuks@sbgat42:~/Tools/madanalysis/trunk/madanalysis-development$ ./bin/ma5
/home/bfuks/Tools/madanalysis/trunk/madanalysis-development
*************************************************************
*                                                           *
*        W E L C O M E  to  M A D A N A L Y S I S  5        *
*                                                           *
*                 /'\_/`\/\  ___ \/\  ____\                 *
*                /\      \ \ \_\ \ \ \___/                  *
*                \ \ \__\ \ \  __ \ \___``\                 *
*                 \ \ \_/\ \ \ \/\ \/\ \L\ \                *
*                  \ \_\\ \_\ \_\ \_\ \____/                *
*                   \/_/ \/_/\/_/\/_/\/___/                 *
*                                                           *
*   MA5 Release : 0.1.0 beta                   2011-09-14   *
*                                                           *
*   The MadAnalysis Development Team - Please visit us at   *
*   https://server06.fynu.ucl.ac.be/projects/madanalysis    *
*                                                           *
*              Type 'help' for in-line help.                *
*          Type 'tutorial' to learn how MA5 works.          *
*                                                           *
*************************************************************
MadGraph 5 NOT found =>  default particle names from the file:
  /home/bfuks/Tools/madanalysis/trunk/madanalysis-development/madanalysis/input/particles_name_default.txt
84 particles have been successfully exported.
MadGraph 5 NOT found => default multiparticle definitions from the file:
  /home/bfuks/Tools/madanalysis/trunk/madanalysis-development/madanalysis/input/multiparticles_default.txt
Creation of a multiparticle labelled by 'invisible' (related to missing energy).
Creation of a multiparticle labelled by 'hadronic' (related to jet transverse energy).
8 multiparticles have been successfully exported.
ma5>
```

1. **Detect MadGraph or not**.
2. **Default (multi)particle names**.
   - * MG5: MadGraph defaults.
   - * **Alone**: own defaults.
3. The multiparticle `hadronic`.
   - * Related to **jet energy**. (variables such as H_T).
4. The multiparticle `invisible`.
   - * Related to **missing energy**.

Typing commands in MadAnalysis.

- **Help available for any command (type `help <command>`).**
- **Command line completion (⇆).**
  - * Show **available commands/options** from what has been typed.
  - * Works with an **empty command line**.
  - * If a command has been typed ⇒ shows the **available options**.

▶help di ⇆

```
ma5>help di
display                       display_datasets
display_multiparticles        display_particles
ma5>help display
```

▶help display Enter

```
ma5>help display
   Syntax: display <object name>
   Displays the definition and the properties of an object.
ma5>
```

Typing commands in MadAnalysis.

- **Help available for any command (type `help <command>`).**
- **Command line completion (↹).**
  - * Show **available commands/options** from what has been typed.
  - * Works with an **empty command line**.
  - * If a command has been typed ⇒ shows the **available options**.

▶display ↹

```
ma5>display
a     cr~  el-        j     n2    svm   ta-    u~   x1+
b     c~   er+        l+    n3    svm~  ta1+   ve   x1-
b1    d    er-        l-    n4    svt   ta1-   ve~  x2+
b1~   dl   g          main  p     svt~  ta2+   vl   x2-
b2    dl~  go         mu+   s     s~    ta2-   vl~  z
b2~   dr   h+         mu-   sl    t     t~     vm
b~    dr~  h-         mul+  sl~   t1    u      vm~
c     d~   h2         mul-  sr    t1~   ul     vt
cl    e+   h3         mur+  sr~   t2    ul~    vt~
cl~   e-   hadronic   mur-  sve   t2~   ur     w+
cr    el+  invisible  n1    sve~  ta+   ur~    w-
ma5>
```

Typing commands in MadAnalysis.

- **Help available for any command (type `help <command>`).**
- **Command line completion (⇆).**
  - * Show **available commands/options** from what has been typed.
  - * Works with an **empty command line**.
  - * If a command has been typed ⇒ shows the **available options**.

▶`display` **main.** ⇆

```
ma5>display main.
main.SBratio  main.currentdir  main.lumi
main.normalize  main.uncertainty
ma5>
```

Displaying information.

- **The `display` command** (syntax: `display <object_name>`).
  - * Show the **definition and the properties** of an object.
- **The `display_particles` and `display_multiparticles` commands**.

  (syntax: `display_particles` and `display_multiparticles`).
  - * Show the list of **all (pre)defined (multi)particles**.
- **The `display_datasets` command** (syntax: `display_datasets`).
  - * Show the list of **all imported datasets** (*i.e.*, all event files). [see below.]

▶`display mu+` Enter

```
ma5>display mu+
 The particle 'mu+' is defined by the PDG-id -13 .
ma5>
```

▶`display l+` Enter

```
ma5>display l+
 The multiparticle 'l+' is defined by the PDG-ids -13 -11 .
ma5>
```

Creating and destroying particles and multiparticles.

- **The `define` command allows to define new (multi)particles**.
  - Example: collecting all muons (μ^+ and μ^-).
  - Works as in MadGraph5.
- **The `remove` command allows to remove existing (multi)particles**.

▶`define mu = mu+ mu-` Enter

`display mu` Enter

```
ma5>define mu = mu+ mu-
ma5>display mu
   The multiparticle 'mu' is defined by the PDG-ids -13 13 .
ma5>
```

▶`remove mu` Enter

`display mu` Enter

```
ma5>remove mu
ma5>display mu
    ** ERROR: no object called 'mu' found.
ma5>
```

Outline.

1. Quick overview.
2. Initialization and displaying information.
3. Event files.
4. Performing an analysis.
5. Current to-do list.

Importing event files.

- **This proceeds through the command** `import`.
  - * Python-like syntax.
  - * Both **.lhe and .lhe.gz** files are supported.
  - * **Wildcards and** ∼ are allowed ⇒ import of several event files in one shot.
- **Storage**.
  - * Imported files are stored as **datasets**.
  - * **Default set name**: `defaultset` ; otherwise **specified by the user**.

▶`import demo/ttbar.lhe.gz` Enter

```
ma5>import demo/ttbar.lhe.gz
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/ttbar.lhe.gz' in the dataset 'defaultset'.
ma5>
```

▶`import demo/ww.lhe as ww` Enter

```
ma5>import demo/ww.lhe.gz as ww
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/ww.lhe' in the dataset 'ww'.
ma5>
```

Importing event files.

- **This proceeds through the command** `import`.
  - * PYTHON-like syntax.
  - * Both **.lhe and .lhe.gz** files are supported.
  - * **Wildcards and** ∼ are allowed ⇒ import of several event files in one shot.
- **Storage**.
  - * Imported files are stored as **datasets**.
  - * **Default set name**: `defaultset` ; otherwise **specified by the user**.

▶`import demo/*` Enter

```
ma5>import demo/*
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/ww.lhe' in the dataset 'defaultset'.
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/ttbar.lhe.gz' in the dataset 'defaultset'.
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/dy.lhe.gz' in the dataset 'defaultset'.
ma5>
```

Signal and background events.

- **Two type of datasets: signal and background**.
  - To be developed in a not too far future.
  - In the prospects of **comparing signal and background** distributions.
  - Default: the type of a dataset is **signal-like**.
  - The type of an event sample can be modified through the **command** `set`.

►Setting the type of a dataset.

```
ma5>set defaultset.type = signal
ma5>set ww.type = background
ma5>
```

Displaying dataset information.

- **The command** `display_datasets`.
  - * Gives **the list of all imported datasets**.
- **The command** `display` **also works with datasets**.
  - * Cross sections, number of events, line color in histograms, . . .
  - * The properties can be changed through the **command** `set`.

►**New MadAnalysis session: defining three datasets.**

```
ma5>import demo/ttbar.lhe.gz as ttbar
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/ttbar.lhe.gz' in the dataset 'ttbar'.
ma5>import demo/ww.lhe as ww
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/ww.lhe' in the dataset 'ww'.
ma5>import demo/dy.lhe.gz as dy
   -> Storing the file '/home/bfuks/Tools/madanalysis/trunk/madan
alysis-development/demo/dy.lhe.gz' in the dataset 'dy'.
ma5>
```

Displaying dataset information.

- **The command** `display_datasets`.
  - * Gives **the list of all imported datasets**.
- **The command** `display` **also works with datasets**.
  - * Cross sections, number of events, line color in histograms, . . .
  - * The properties can be changed through the **command** `set`.

▶Displaying the list of imported datasets.

```
ma5>display_datasets
   ********* List of defined datasets *********
   dy (signal)
   ttbar (signal)
   ww (signal)
   ********************************************
ma5>
```

Displaying dataset information.

- **The command** `display_datasets`.
  - * Gives **the list of all imported datasets**.
- **The command** `display` **also works with datasets**.
  - * Cross sections, number of events, line color in histograms, . . .
  - * The properties can be changed through the **command** `set`.

▶Displaying information related to the (still not processed) `ww` dataset.

```
ma5>display ww
  ******************************************
  Name of the dataset = ww (signal).
  User-imposed cross section = 0.0.
  User-imposed weight of the set = 1.0.
  Line color in histograms = black.
  Background color in histograms = invisible.
  List of LHE files included in this dataset:
   - /home/bfuks/Tools/madanalysis/trunk/madanalysis-development/dem
o/ww.lhe
  ******************************************
  Cross section = (0.0 +/- 0.0) pb.
  Total number of events = 0.
  ******************************************
```

Setting dataset properties and destroying datasets.

- **The properties of a dataset can be changed through the command `set`.**
- **A dataset can be removed through the command `remove`.**

►Setting properties.

```
ma5>set ww.type = signal
ma5>set ttbar.backcolor = red
ma5>set dy.linecolor = blue
ma5>
```

Setting dataset properties and destroying datasets.

- **The properties of a dataset can be changed through the command** `set`.
- **A dataset can be removed through the command** `remove`**.**

►**remove ww** Enter .

```
ma5>display_datasets
   ********* List of defined datasets *********
   dy (signal)
   ttbar (signal)
   ww (signal)
   ********************************************
ma5>remove ww
ma5>display_datasets
   ********* List of defined datasets *********
   dy (signal)
   ttbar (signal)
   ********************************************
```

Outline.

1. Quick overview.
2. Initialization and displaying information.
3. Event files.
4. Performing an analysis.
5. Current to-do list.

Defining histograms (1).

- **The command** `plot` **allows to define the distributions to be investigated**.
  * Requires as argument **the observable to be analyzed**.
  * Requires also **binning information** (number of bins, min and max values).
  * **Future developments**: automatic binning information from the event file.
- **Cuts: being developed**.
- **Distributions without any argument (**`plot <obs> nbin xmin xmax`**)**.
  * Total and missing transverse energy (`TET` **and** `MET`).
  * Total and missing transverse hadronic energy (`THT` **and** `MHT`).
  * Final particles present in the samples (`NPID`).
  * Linked to the `hadronic` **and** `invisible` multiparticles.

Defining histograms (2).

- **Distributions with one or several argument(s).**
  (`plot <obs>[arg1 arg2 ...]  nbin xmin xmax`**)**.
  - The arguments are **(multi)particles**.
    - ◇ `plot PT[mu+]`: the p_T of the antimuon.
    - ◇ `plot PT[mu+ mu-]`: the p_T of a muon anti-muon pair.
  - Multiplicities of the final particles present in the samples (`N`).
  - Kinematical variables (energy, mass, momentum, angles, rapidity, ...).
    ⇔ `E`, `M`, `P`, `ET`, `MT`, `PT`, `PX`, `PY`, `PZ`, `RHO`, `THETA`, `PHI`, `ETA`, `Y`, `GAMMA`, `BETA`.
  - Scalar and vector sums/differences/ratios. (`sPT`, `vPT`, `dsPT`, `dvPT`, `rPT`, `...`).
  - **Remark: differences and ratios only defined for two particles**.

- **The** `or` **and** `all` **keywords can be used**.
  - Example-1: `plot M[mu+ mu+ or mu- mu-] 201 0 200`.
  - Example-2: `plot PT[all mu] 201 0 200` (several muons in the events).
  - **Remark: the keyword** `all` **is not fully implemented yet**.

Implementing an analysis.

- **The command** `display` **allows to show the analysis to be performed**.

►Displaying the list of imported datasets.

```
ma5>display_datasets
   ********* List of defined datasets *********
   dy (signal)
   ttbar (signal)
   ww (signal)
   ********************************************
ma5>
```

►Defining the analysis to be performed.

```
ma5>define mu = mu+ mu-
ma5>plot PT[mu] 501 0 500
ma5>plot N[mu] 6 0 5
ma5>plot MET 101 0 100
ma5>plot TET 501 0 500
ma5>
```

Implementing an analysis.

- **The command `display` allows to show the analysis to be performed.**

▶Displaying the analysis to be performed.

```
ma5>display selection
   *********** Selection ***********
   1. Plot : PT [ mu ]
   2. Plot : N [ mu ]
   3. Plot : MET
   4. Plot : TET
   *********************************
ma5>
```

▶Removing a selection.

```
ma5>remove selection[4]
ma5>display selection
   *********** Selection ***********
   1. Plot : PT [ mu ]
   2. Plot : N [ mu ]
   3. Plot : MET
   *********************************
ma5>
```

Implementing an analysis.

- **The command `set` allows to modify the properties of the histograms to be plotted**.

►Changing selection properties.

```
ma5>display selection
   *********** Selection ***********
   1. Plot : PT [ mu ]
   2. Plot : N [ mu ]
   3. Plot : MET
   *********************************
ma5>set selectoin[1].logY = true
ma5>set selectoin[3].logY = true
ma5>set selection[3].nbin = 100
ma5>
```

Performing the analysis.

- **The command `submit` allows to perform the selected analysis**.
    - * **Reading** of the lhe event files.
    - * **Updating** the information associated to each dataset.
    - * **Creating a ROOT file** with the analysis itself.
    - * **Syntax:** `submit <dirname>` (where the produced files are stored).

▶Submitting the analysis.

```
ma5>submit testdir
  Creating folder 'testdir'...
  Copying 'LHEanalyzer' source files...
  Inserting your selection into 'LHEanalyzer'...
  Writing the list of datasets...
  Creating a 'Makefile'...
  Compiling 'LHEanalyzer'...
  Linking 'LHEanalyzer'...
```

Performing the analysis.

- **The command** `submit` **allows to perform the selected analysis**.
  - * **Reading** of the lhe event files.
  - * **Updating** the information associated to each dataset.
  - * **Creating a** ROOT **file** with the analysis itself.
  - * **Syntax:** `submit <dirname>` (where the produced files are stored).

►Submitting the analysis (cntn)'.

```
Running 'LHEanalyzer' over dataset 'ww'...
 *******************************************************
 * LHEanalyzer 1.2 for MadAnalysis 5 - Welcome.
 * Option choices: 'keeping only final states'.
 * Extracting the following LHE files:
 * 1/1  .../demo/ww.lhe
 *   => sample produced by MadGraph + Pythia.
 *   => Number of processed events: 10000.
 * Creating a root file...
 * Total number of processed events: 10000.
 * Goodbye.
 *******************************************************
```

Performing the analysis.

- **The command** `submit` **allows to perform the selected analysis**.
  - * **Reading** of the lhe event files.
  - * **Updating** the information associated to each dataset.
  - * **Creating a ROOT file** with the analysis itself.
  - * **Syntax:** `submit <dirname>` (where the produced files are stored).

►Submitting the analysis (cntn)'.

```
Running 'LHEanalyzer' over dataset 'ttbar'...
 *******************************************************
 * LHEanalyzer 1.2 for MadAnalysis 5 - Welcome.
 * Option choices: 'keeping only final states'.
 * Extracting the following LHE files:
 * 1/1  .../demo/ttbar.lhe.gz
 *   => sample produced by MadGraph + Pythia.
 *   => Number of processed events: 10000.
 * Creating a root file...
 * Total number of processed events: 10000.
 * Goodbye.
 *******************************************************
```

Performing the analysis.

- **The command** `submit` **allows to perform the selected analysis**.
  - * **Reading** of the lhe event files.
  - * **Updating** the information associated to each dataset.
  - * **Creating a ROOT file** with the analysis itself.
  - * **Syntax:** `submit <dirname>` (where the produced files are stored).

►Submitting the analysis (cntn)'.

```
Running 'LHEanalyzer' over dataset 'dy'...
 *******************************************************
 * LHEanalyzer 1.2 for MadAnalysis 5 - Welcome.
 * Option choices: 'keeping only final states'.
 * Extracting the following LHE files:
 * 1/1  .../demo/dy.lhe.gz
 *   => sample produced by MadGraph + Pythia.
 *   => Number of processed events: 10000.
 * Creating a root file...
 * Total number of processed events: 10000.
 * Goodbye.
 *******************************************************
```

Performing the analysis.

- **The command** `submit` **allows to perform the selected analysis**.
  - * **Reading** of the lhe event files.
  - * **Updating** the information associated to each dataset.
  - * **Creating a ROOT file** with the analysis itself.
  - * **Syntax:** `submit <dirname>` (where the produced files are stored).

```
 Checking LHEanalyzer output...
 Extracting data from the output files...
ma5>
```

Displaying updated dataset information.

- **The command** `display_datasets`.
  - * The information is now **updated**.

```
ma5>display ww
   ******************************************
   Name of the dataset = ww (signal).
   User-imposed cross section = 0.0.
   User-imposed weight of the set = 1.0.
   Line color in histograms = black.
   Background color in histograms = invisible.
   List of LHE files included in this dataset:
    - .../demo/ww.lhe
   ******************************************
   Cross section = 72.6 pb.
   Total number of events = 10000.
   ******************************************
ma5>
```

Generating human-readable output.

- **A ROOT-file is already there**.
- **Three additional output format can be generated**.
  - * HTML: `generate_html <dirname>`.
  - * LaTeX: `generate_latex <dirname>` or `generate_pdflatex <dirname>`.

Let's go for a live demo.

Outline.

1 Quick overview.

2 Initialization and displaying information.

3 Event files.

4 Performing an analysis.

5 Current to-do list.

To-do list.

- **Short term plans**.
  - `update` command ('replot' in GNUPLOT).
  - `preview/open` **command.**
  - **Hardcoded analyses** (multiplicities, p_T, invariant masses, ...
  - The `all` **keyword**.
  - **Sanity check** routine of the LHE files.
  - p_T **ordering** of the particles (leading muon, . . .).
  - **Particle history** (mother, grand-mother, . . .).
  - **Restoring** MADANALYSIS **configuration** from the ROOT files.
- **Longer term plans**.
  - **Cuts** and calculation of the **efficiencies**.
  - Automated treatment for the **binning** of the histograms.
  - Basic **detector simulation**.
  - **Manual**.