Correlation/HEPData recommendations for the Yellow Report, and plotting tool

Louie Corpe (UCL)

LHCEWWG Jets and Bosons

18 May 2020

Context

- Analysis preservation is increasingly important discussion in HEP
  -->new and ambitious goals of combinations, recasting, EFT fits etc...

- Do we store enough information on HEPData to re-use LHC measurements?
  --> Not always! Areas where small policy shifts can boost impact of analyses

- Prompted by discussion on correlations at **LHCEWWG**: Dec18, Feb19, Jul19
  Attempt to **agree conventions between LHC experiments**
  - **Recommendations on what to store in HEPData** depending on level of re-interpretation needed
  - LHC Re-interpretation Forum (arXiv) thoughts align closely with this
  - presented to groups in ATLAS, CMS, LHCb and ALICE over last 6m...
  - Include our recommendations in upcoming Yellow Report

- I've now put a first draft of our recommendations document into the YR page. And used the new uncertainty breakdown format to make a handy uncertainty-band plotting tool wrapper for rivet-mkhtml

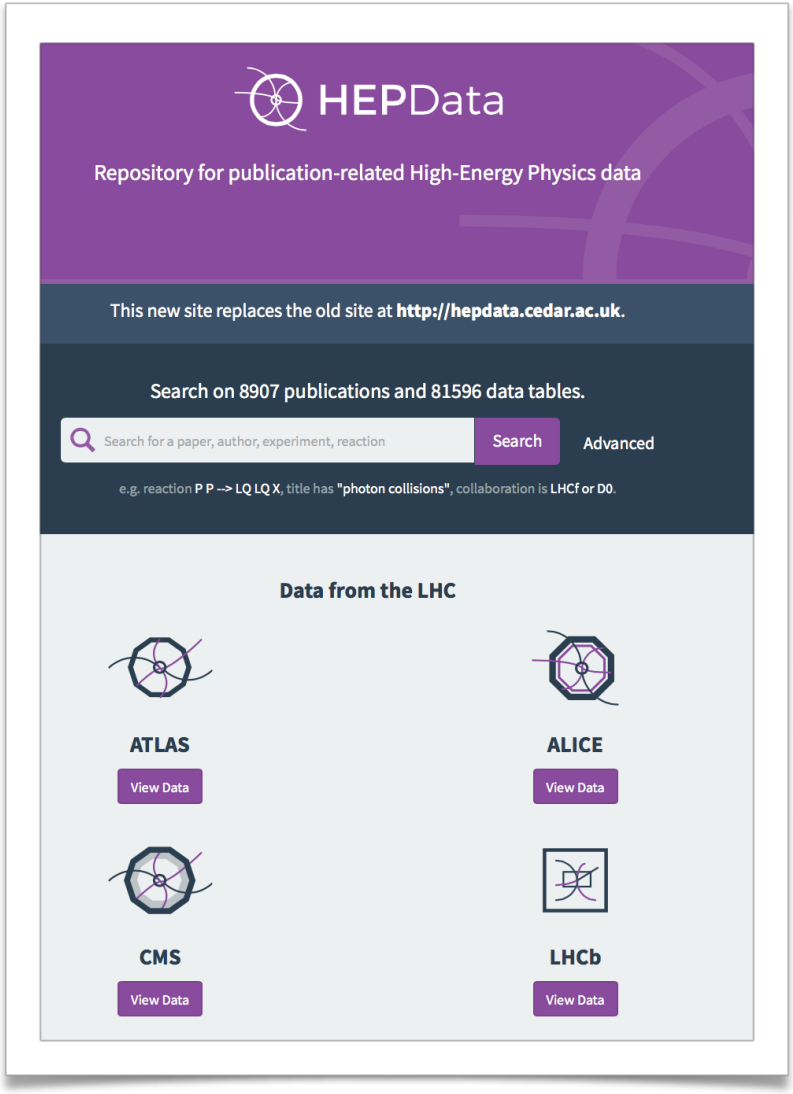
The YR section

- I've incorporated our text from our the summary of our discussions from last year into the YR draft
- Gives a short motivation, and explains the current workflow of HEPData+Rivet/ Yoda and how this is already established in the generator/tuning/re-interpretation community
- Explain the limitations of the current practices and how they could be improved
- define new recommendations, based on 3 scenarios (not intended to be strict, just a guide!)
- Specific conventions for various objects to store on HEPData
- Practical examples in the Appendix

Contents

What are the current practices?

And why do they go wrong?



Existing practices

- Good practice to define fiducial volume of measurement (eg **Rivet Routine**)
  Often comes years later, or not at all...
  Documenting analysis logic w/ code snippet is vital for preservation
  [release of Rivet 3.0.1 (many Heavy-Ion developments) will help support ALICE]

- Give results with uncertainties in each bin.
  Separate stat vs syst uncertainties at minimum.
  Stat/syst not enough to model correlations if re-interp is to be trusted!

- Exact names given by performance groups for error sources/nuisance parameters (NPs)

LHCb Forward W/Z production measurement from 2016, only stat+sys+lumi breakdown

RE	NUMU > JET X	NUMUBAR > JET X	JET X
	σ_{W^+j} [pb]	σ_{W^-j} [pb]	σ_{Zj} [pb]
	56.9 ±0.2 stat ±5.1 sys ±0.7 sys,lumi	33.1 ±0.2 stat ±3.5 sys ±0.4 sys,lumi	5.71 ±0.06 stat ±0.27 sys ±0.07 sys,lumi

LHCb Forward W/Z production measurement from 2016, but still no rivet routine!

The measured cross-sections for Wj production in four bins of η^{μ}. The uncertainties shown are sta
and due to the luminosity determination.

cmenergies 8000.0 **observables** SIG

SQRT(S)	8000 GEV	
η^{μ}	σ_{W+j} [pb]	σ_{W-j} [pb]
2.0 - 2.5	22.24 ±0.15 stat ±2.07 sys ±0.26 sys,lumi	10.16 ±0.1 stat ±1.09 sys ±0.12 sys,lumi
2.5 - 3.0	18.09 ±0.12 stat ±1.68 sys ±0.21 sys,lumi	9.53 ±0.09 stat ±1.08 sys ±0.11 sys,lumi
3.0 - 3.5	11.29 ±0.1 stat ±0.96 sys ±0.13 sys,lumi	7.39 ±0.08 stat ±0.74 sys ±0.09 sys,lumi
3.5 - 4.5	6.07 ±0.08 stat ±0.46 sys ±0.07 sys,lumi	6.29 ±0.08 stat ±0.72 sys ±0.07 sys,lumi

13 TeV jet substructure measurements in tt from CMS: with full uncertainty breakdown!

λ_0^0 (N)	incl	bottom
0 - 3.0	0.032603 ±0.00016275 stat,Stat -0.00048105 +0.00057333 sys,+jec_SubTotalPileUp_up,-jec_SubTo -0.00030539 +0.00025237 sys,+jec_SubTotalPt_up,-jec_SubTotalP -0.00053812 +0.00049466 sys,+jec_SubTotalRelative_up,-jec_Sub	0.025586 ±0.00040941 stat,Stat -0.00033805 +0.00012629 sys,+jec_SubTotalPileUp_up,-jec_SubTo -0.00040186 +0.00013767 sys,+jec_SubTotalPt_up,-jec_SubTotalP -0.0007672 -0.00018921 sys,+jec_SubTotalRelative_up,-jec_Sub

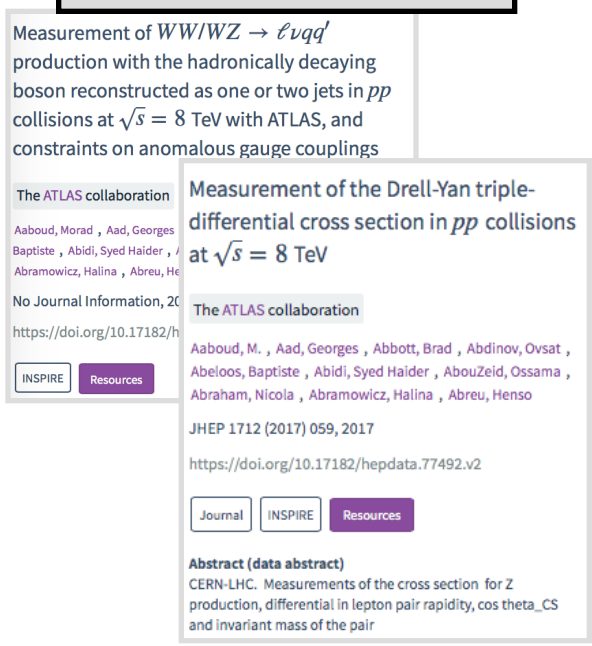
Existing practices

- Good practice to define fiducial volume of measurement (eg **Rivet Routine**)
  Often comes years later, or not at all...
  Documenting analysis logic w/ code snippet is vital for preservation
  [release of Rivet 3.0.1 (many Heavy-Ion developments) will help support ALICE]

- Give results with uncertainties in each bin. Separate stat vs syst uncertainties at minimum.
  Stat/syst not enough to model correlations if re-interp is to be trusted!

- Exact names given by performance groups for error sources/nuisance parameters (NPs)

Important ATLAS SM measurements from 2017, 2018 but still no rivet routine!



Recent ATLAS W+jets meas, only total uncert

FIRST LEADING JET PT [GEV]	D(SIG)/D(PT) [PB/GEV]
30 - 39	23.79 +2.508 -2.032
39 - 49	13.81 +1.072 -0.9109
49 - 60	8.276 +0.5578 -0.6006

13TeVATLAS ttbar + heavy jets (link)

Selection	SIG [FB]
0.5 - 1.5	181 ±5 stat ±24 sys
1.5 - 2.5	27 ±3 stat ±7 sys
2.5 - 3.5	2450 ±40 stat ±690 sys
3.5 - 4.5	359 ±11 stat ±61 sys

Stat vs syst breakdown not enough to (approximately) model correlations

Existing practices

- If strong correlations... 2 options:
  a) explicit covariance or correlation matrix
  OK if measurement not intended to be combined with other measurements
  b) breakdown of signed(!) effect of each NP.
  Can then rebuild covariance matrix

- Typically prefer b) since a) implicitly symmetrizes, and information to correlate with other measurements is insufficient.
  Alternatively, use 'extended' covariance matrix with one row per bin AND one row per NP
  The correct choice may vary depending on the intended use case!

- Statistical correlations from bootstrap method, store either replicas or correlation matrix.
  Replicas best for future combinations but need make TH*DBootstrap code public

ATLAS 13 TeV single top + W (link)

RE	P P --> W- TOP X, P P --> W+ TOPBAR X
SQRT(S)	13000 GEV
M_LEP2BJET [GEV]	DSIG(fiducial)/DM_LEP2BJET [FB/GEV]
0 - 50	1.13 ±0.34 stat ±0.036 syst,MC stat uncertainty (bootstrap) ∓0.11 syst,Unfolding non-closure ±0.017 syst,Z+jets normalisation ±0.023 syst,tW ttbar diagram subtraction remo ∓0.043 syst,tW initial final state radiation ∓0.39 syst,tW matrix element generator ∓0.22 syst,tW parton shower hadronisation ge ∓0.022 syst,Diboson normalisation ∓0.011 syst,Electron pt resolution ∓0.024 syst,Electron scale factor ID ∓0.0067 syst,Fake non-prompt background ±0.056 syst,Flavour-tagging scale factor: B 0 ±0.00039 syst,Flavour-tagging scale factor: B 1 ±0.047 syst,JES BJES Response ±0.19 syst,JES EffectiveNP 1

Nice example of full error breakdown

ATLAS 8 TeV ttbar (l+jets) (link)

RE	P P --> TOP TOPBAR X
SQRT(S) [GEV]	Cross section $\sigma_{\mathrm{inc}}(t\bar{t})$ [fb]
8000.0	248.3 ±0.7 stat ±13.4 sys_1 ±4.7 sys_2

Poor naming of uncertainty breakdown

Existing practices

- If strong correlations... 2 options:
  a) explicit covariance or correlation matrix
  OK if measurement not intended to be combined with other measurements
  b) breakdown of signed(!) effect of each NP.
  Can then rebuild covariance matrix

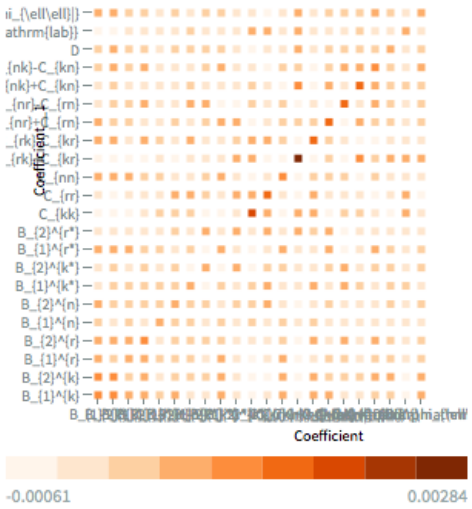
- Typically prefer b) since a) implicitly symmetrizes, and information to correlate with other measurements is insufficient.
  Alternatively, use 'extended' covariance matrix with one row per bin AND one row per NP
  The correct choice may vary depending on the intended use case!

- Statistical correlations from bootstrap method, store either replicas or correlation matrix.
  Replicas best for future combinations but need make TH*DBootstrap code public

Recent CMS tt measurement which included covariance matrix directly

Measurement of the top quark polarization and $t\bar{t}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV

Coefficient	Coefficient__1	Systematic covariance for all coefficients
B_1^k	B_1^k	0.0004253623247
B_2^k	B_1^k	0.0003758444934
B_1^r	B_1^k	4.24559268e-05
B_2^r	B_1^k	7.046833802e-05
B_1^n	B_1^k	2.113294779e-05



Existing practices

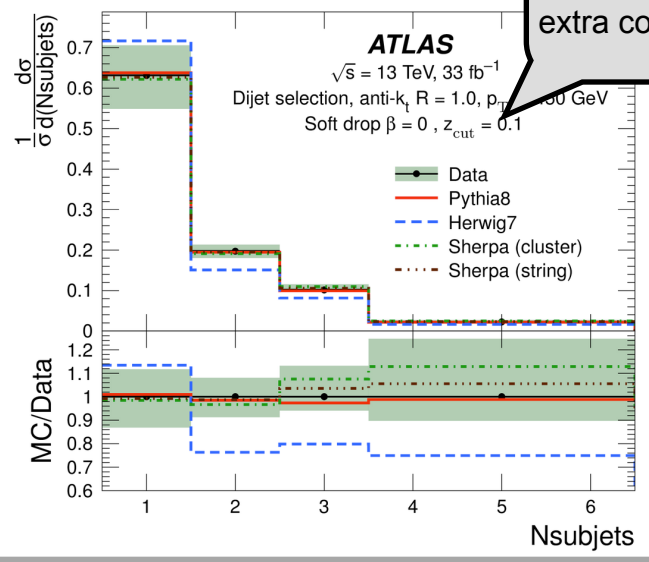
- Theory predictions not usually in HEPData...
  Exception: when theory predictions complicated/intensive to be produced
  Policy from time when SM calculations were cheap. **No longer the case!** A shift here can boost our impact...

- For searches: exclusion limits + necessary info to emulate analysis (e.g kinematic distributions, signal acceptances and selection efficiencies)
  Rivet can now preserve searches w/ custom smearing+efficiency capabilities

  See recent paper on this topic !
  https://arxiv.org/abs/1910.01637

ATLAS 13 TeV jet substructure in tWj (link)

RE	P P -> J J
SQRT(S)	13000 GEV
Nsubjets	1/SIG) * D(SIG)/Nsubjets
0.5 - 1.5	0.631596781833 +0.000124841439105 -0.000124842090418 stat +0.07408313317276 -0.07206693396204 sys +0.0 -0.0340116380789 modelling
1.5 - 2.5	0.197659658459 +9.89354611399e-05 -9.89361525685e-05 stat +0.01504682927 -0.01716718852974 sys +0.00577373999032 -0.0 modelling
2.5 - 3.5	0.102069447767 +6.09666295014e-05 -6.09660879228e-05 stat +0.00652853618392 -0.00598013153687 sys +0.0119229853657 -0.0 modelling
3.5 - 6.5	0.0221231749204 +1.57216539921e-05 -1.57214761161e-05 stat +0.00341579935785 -0.00234195137488 sys +0.0044314051288 -0.0 modelling



SM predictions in the figure but not on HEPData as extra columns?

Existing practices

- Theory predictions not usually in HEPData...
  Exception: when theory predictions complicated/intensive to be produced
  Policy from time when SM calculations were cheap. **No longer the case!** A shift here can boost our impact...

- For searches: exclusion limits + necessary info to emulate analysis (e.g kinematic distributions, signal acceptances and selection efficiencies)
  Rivet can now preserve searches w/ custom smearing+efficiency capabilities

  See recent paper on this topic !
  https://arxiv.org/abs/1910.01637

ATLAS SUSY search where SM background predictions are given

	DATA	SM BACKGROUND	SIGNAL
SQRT(S)	13000.0 GEV		
m_{eff}(incl.) [GeV]	**Events / 200 GeV**		
900.0 (bin: 800.0 - 1000.0)	53.0 ±7.28	38.76 ±1.912 stat +3.279 -3.64 sys	4.565 ±0.8682
1100.0 (bin: 1000.0 - 1200.0)	184.0 ±13.56	153.0 ±3.87 stat +23.62 -29.25 sys	34.24 ±2.322
1300.0 (bin: 1200.0 - 1400.0)	192.0 ±13.86	176.4 ±4.933 stat +16.98 -18.53 sys	47.85 ±2.673
1500.0 (bin: 1400.0 - 1600.0)	163.0 ±12.77	132.8 ±3.616 stat +19.59 -20.24 sys	38.96 ±2.668
1700.0 (bin: 1600.0 - 1800.0)	113.0 ±10.63	89.99 ±2.567 stat +19.14 -22.83 sys	27.76 ±2.622
1900.0 (bin: 1800.0 - 2000.0)	61.0 ±7.81	51.19 ±2.164 stat +3.891 -5.698 sys	11.71 ±1.205

Rivet plugin for a SUSY search

Rivet analyses reference

ATLAS_2016_I1458270

0-lepton SUSY search with 3.2/fb of 13 TeV pp data
Experiment: ATLAS (LHC)
Inspire ID: 1458270
Status: VALIDATED
Authors:

Proposed LHC-wide HEPData Recommendations

Defines 3 scenarios for levels of information to provide on HEPData

Gives concrete recommendations for the format of objects which are to be stored

3 Scenarios for re-interpretation

- Identify different levels of recommendations, depending on the analysis type and how re-interpretable it needs to be:

Scenario A - Minimum Requirements for Analysis Preservation
Scenario B - Approximate Re-interpretability
Scenario C - Maximum Re-interpretability

Best case - aims to provide maximal information for reinterpretations. Should be gold standard for precision measurements

Closest to current situation. Plenty of information published. Not necessarily enough for strict combinations... but good enough for many analyses (especially searches)

Bare minimum for a search to be re-interpretable

- The scenarios are **not intended to be "strict"**, but are more **designed to get groups thinking about what their intended level of re-interpretability** is, and **what they should preserve** as a result

A - Bare Minimum

- **Minimum amount of info** for result to be re-used meaningfully**.**
  e.g if **only rough estimate of MC/data agreement** or **sensitivity to new models** needed

- **Phase Space Definition**: Ideally, **Rivet routine...** if not...
  - **detailed description of the region of interest**
  - **per-object efficiency tables**
  - **explicit definitions of each variable used in the selection**,
  - **cutflows of the effect of each selection** on well-defined signals
- **Statistical correlations:** omitted if negligible bin migrations.
  Stat error per bin still needed (assumed uncorrelated between bins)
- **Systematic correlations: uncertainty breakdown** or explicit covariance matrices
- **Generator Prediction**: **SM prediction of MC generators**, with breakdown of theory uncertainty if possible

B - Approximate Re-interpretability

- For standard measurements or searches to be **re-interpreted approximately**. E.g **generator tuning ,** and **recasting of searches**
- **Phase space definition**: **Rivet analysis** must be provided concurrently with arxiv submission
  - If results only at detector level, Rivet analysis can still be provided, with adequate smearing and efficiency tables
- **Statistical correlations**: **correlation matrices**. *Can't infer correlations between analyses, but OK if re-interpreting result in isolation*
- **Systematic correlations: uncertainty breakdown:** effect of each major uncertainty source/NP on each bin
  - can create covariance matrix + correlate with other measurements
  - some sources may be grouped, eg JES NPs
- OR, covariance matrix for each distribution: e.g. for simplified likelihoods
- OR `pyhf` likelihood
- **Generator Prediction**: include **SM prediction from latest MC generators** with breakdown of theory uncertainty if possible

C - Maximum Re-interpretability

- **For precision analyses**: for **future combinations**, **measurements of SM parameters, PDF fitting...** Enough info for exact combination

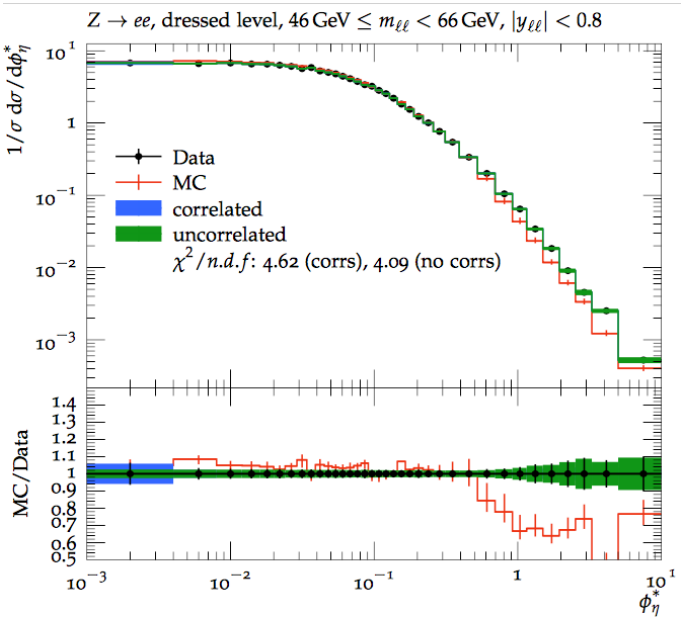
- **Phase space definition**: **Rivet analysis** must be provided concurrently with arXiv submission
- **Stat correlations**: **Bootstrap Replicas** attached to HEPData entry
  [plan for Bootstrap code to be made public]
- **Syst correlations** as detailed uncertainty breakdown, with no grouping of NPs (e.g. for JES, use full granularity of NPs)
  OR `pyhf` likelihood
  OR enlarged covariance matrix with rows/columns for each bin and each NP
- **Generator Prediction**: include **SM prediction from latest MC generators** w/ breakdown of theory uncertainty if possible
- If likelihood fit used: **post-fit values of the NPs in each bin**
- Rotation matrix if applicable

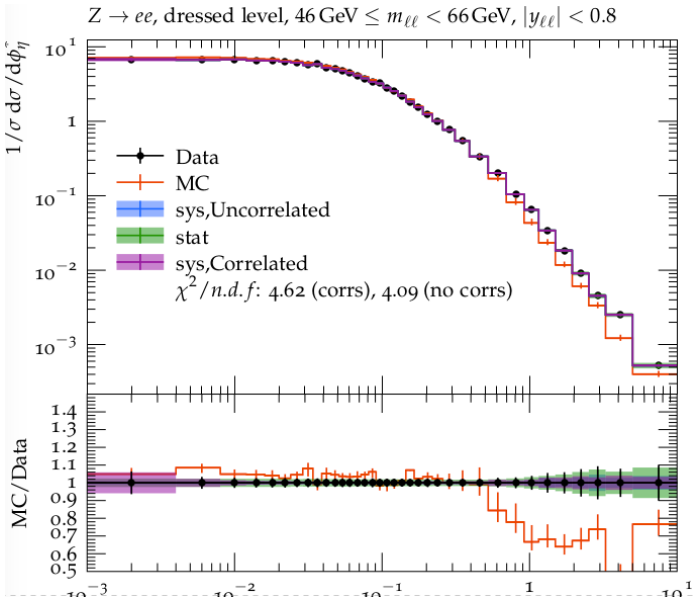
The power of storing an uncertainty breakdown

rivet-mkhtml wrapper

- To illustrate who one should go to the effort of uploading uncertainty breakdown to HEPData, I wrote a small wrapper to rivet-mkhtml which extracts the error breakdown from reference files if available, and calculates chi2 using covariance matrix
- https://gitlab.cern.ch/lhcewkwg/lhcewkwg-vjets/correlations-library/

./rivet-mkhtml-lhcewwg example.yoda:MC -o outdir -m .*dl.* --corr '.*Cor.*' --uncorr '.*stat.*,.*uncor.*,.*Uncor.*' (--applyGroups 0)





How it works

./rivet-mkhtml-lhcewwg example.yoda:MC -o outdir -m .*dl.* --corr '.*Cor.*' --uncorr '.*stat.*,.*uncor.*,.*Uncor.*' (--applyGroups 0)

- Loop through analysis objects in input file (example.yoda), and find relevant reference analysis data in rivet install dir

```
BEGIN YODA_SCATTER2D_V2 /ATLAS_2015_I1408516/d02-x01-y01
ErrorBreakdown: {0: {stat: {dn: -0.16267199999999998, up: 0.16267199999999998}, "sys,Correlated":
{dn: -0.399902, up: 0.399902}, "sys,Uncorrelated": {dn: -0.061001999999999994, up: 0.061001999999999994}},
1: {stat: {dn: -0.15322599999999997, up: 0.15322599999999997}, "sys,Correlated": {dn: -0.15322599999999997,
IsRef: 1
Path: /ATLAS_2015_I1408516/d02-x01-y01
Title: doi:10.17182/hepdata.71339.v1/t2
Type: Scatter2D
Variations: [""]
---
# xval   xerr-   xerr+   yval  yerr-   yerr+
2.000000e-03  2.000000e-03  2.000000e-03  6.778000e+00  3.999020e-01  3.999020e-01
6.000000e-03  2.000000e-03  2.000000e-03  6.662000e+00  1.532260e-01  1.532260e-01
1.000000e-02  2.000000e-03  2.000000e-03  6.781000e+00  1.017150e-01  1.017150e-01
1.400000e-02  2.000000e-03  2.000000e-03  6.561000e+00  7.217100e-02  7.217100e-02
1.800000e-02  2.000000e-03  2.000000e-03  6.540000e+00  6.540000e-02  6.540000e-02
2.200000e-02  2.000000e-03  2.000000e-03  6.327000e+00  6.959700e-02  6.959700e-02
2.650000e-02  2.500000e-03  2.500000e-03  6.102000e+00  5.491800e-02  5.491800e-02
3.150000e-02  2.500000e-03  2.500000e-03  5.682000e+00  4.545600e-02  4.545600e-02
3.650000e-02  2.500000e-03  2.500000e-03  5.868000e+00  5.281200e-02  5.281200e-02
4.200000e-02  3.000000e-03  3.000000e-03  5.263000e+00  2.631500e-02  2.631500e-02
4.800000e-02  3.000000e-03  3.000000e-03  5.032000e+00  3.019200e-02  3.019200e-02
5.400000e-02  3.000000e-03  3.000000e-03  4.796000e+00  2.877600e-02  2.877600e-02
6.050000e-02  3.500000e-03  3.500000e-03  4.443000e+00  3.110100e-02  3.110100e-02
```

How it works

./rivet-mkhtml-lhcewwg_example.yoda:MC -o outdir -m .*dl.* -corr '.*Cor.*' --uncorr '.*stat.*,.*uncor.*,.*Uncor.*' (--applyGroups 0)

- Use regexes in command line to group components into correlated and uncorrelated parts

```
BEGIN YODA_SCATTER2D_V2 /ATLAS_2015_I1408516/d02-x01-y01
ErrorBreakdown: {0: {stat: {dn: -0.16267199999999998, up: 0.16267199999999998}, sys,Correlated":
{dn: -0.399902, up: 0.399902}, "sys,Uncorrelated" {dn: -0.06100199999999994, up: 0.06100199999999994}},
1: {stat: {dn: -0.15322599999999997, up: 0.15322599999999997}, "sys,Correlated": {dn: -0.15322599999999997,
IsRef: 1
Path: /ATLAS_2015_I1408516/d02-x01-y01
Title: doi:10.17182/hepdata.71339.v1/t2
Type: Scatter2D
Variations: [""]
---
# xval   xerr-   xerr+   yval  yerr-   yerr+
2.000000e-03  2.000000e-03  2.000000e-03  6.778000e+00  3.999020e-01  3.999020e-01
6.000000e-03  2.000000e-03  2.000000e-03  6.662000e+00  1.532260e-01  1.532260e-01
1.000000e-02  2.000000e-03  2.000000e-03  6.781000e+00  1.017150e-01  1.017150e-01
1.400000e-02  2.000000e-03  2.000000e-03  6.561000e+00  7.217100e-02  7.217100e-02
1.800000e-02  2.000000e-03  2.000000e-03  6.540000e+00  6.540000e-02  6.540000e-02
2.200000e-02  2.000000e-03  2.000000e-03  6.327000e+00  6.959700e-02  6.959700e-02
2.650000e-02  2.500000e-03  2.500000e-03  6.102000e+00  5.491800e-02  5.491800e-02
3.150000e-02  2.500000e-03  2.500000e-03  5.682000e+00  4.545600e-02  4.545600e-02
3.650000e-02  2.500000e-03  2.500000e-03  5.868000e+00  5.281200e-02  5.281200e-02
4.200000e-02  3.000000e-03  3.000000e-03  5.263000e+00  2.631500e-02  2.631500e-02
4.800000e-02  3.000000e-03  3.000000e-03  5.032000e+00  3.019200e-02  3.019200e-02
5.400000e-02  3.000000e-03  3.000000e-03  4.796000e+00  2.877600e-02  2.877600e-02
6.050000e-02  3.500000e-03  3.500000e-03  4.443000e+00  3.110100e-02  3.110100e-02
```

How it works

./rivet-mkhtml-lhcewwg example.yoda:MC -o outdir -m .*d1.* --corr '.*Cor.*' --uncorr '.*stat.*,.*uncor.*,.*Uncor.*' (--applyGroups 0)

- Make dummy yoda files for corr/uncorr components
- Build covariance matrix using direct method, and evaluate chi2 between data and MC with/without the off-diagonal terms

```
Processing /ATLAS_2015_I1408516/d25-x01-y01 (8 bins)
--> $\chi^2/n$: 3.21 (corrs), 3.07 (no corrs)
```

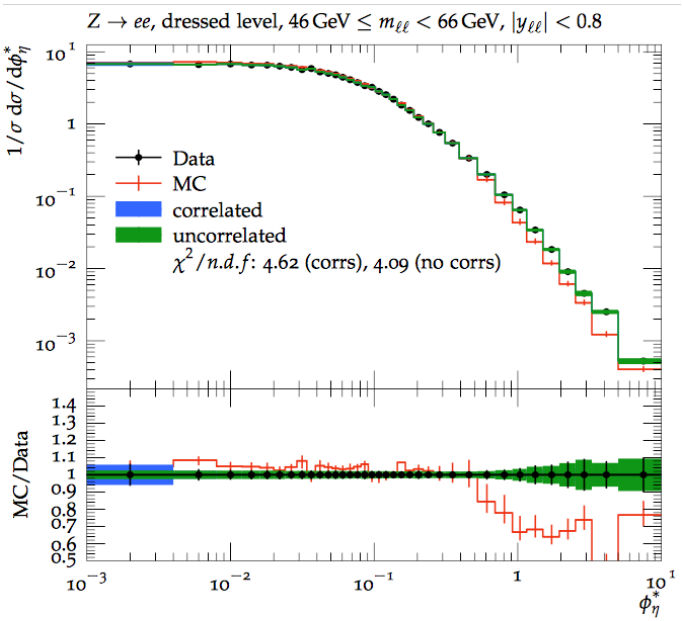
- Call rivet-mktml to do the actual plotting including chi2 in legends:

```
rivet-mkhtml example.yoda:MC
correlated.yoda:ErrorBands=1:LineColor=blue:ErrorBandColor=blue:RatioPlotSameStyle=1:ErrorBandOpacity=0.5:Title=correlated
uncorrelated.yoda:ErrorBands=1:LineColor=green:ErrorBandColor=green:RatioPlotSameStyle=1:ErrorBandOpacity=0.5:Title=uncorrelated
other.yoda:ErrorBands=1:LineColor=orange:ErrorBandColor=orange:RatioPlotSameStyle=1:ErrorBandOpacity=0.5:Title=other -o outdir -m .*d02.* -c extra.plot --errs
```

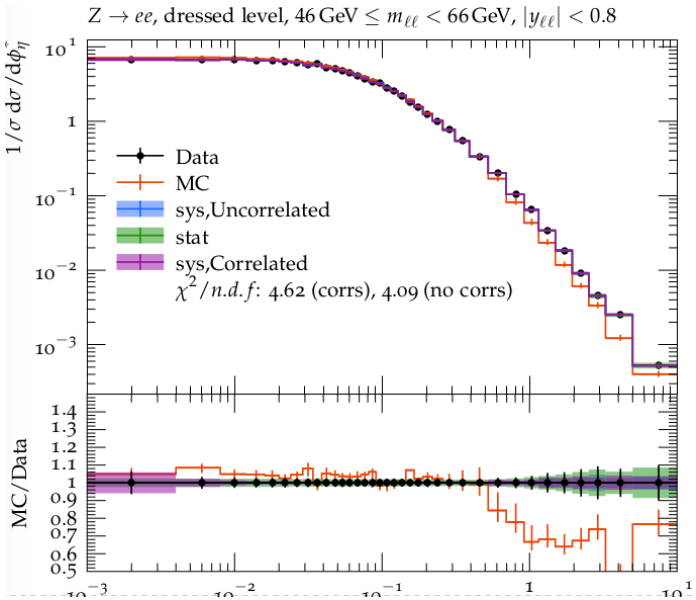
How it works

./rivet-mkhtml-lhcewwg example.yoda:MC -o outdir -m .*dl.* --corr '.*Cor.*' --uncorr '.*stat.*,.*uncor.*,.*Uncor.*' (--applyGroups 0)

- --applyGroups determines whether to plot bands for each syst component or just corr/uncorr
- All other rivet-mkhtml commands accepted as usual



--applyGroups 1



--applyGroups 0

Backup

Summary

- LHCEWWG and LHC Re-interpretation fora have prompted discussions on HEPData, and effort to synchronise better between LHC experiments
  - Proposals have been presented to ATLAS/CMS/LHCb/ALICE

- Nevertheless, recent developments are excellent opportunity to review status and see what we can do better
  --> maximise impact of our measurements and searches !

- Wrote a handy rivet-mkhtml wrapper to help use the correlation info: please give it a try!