

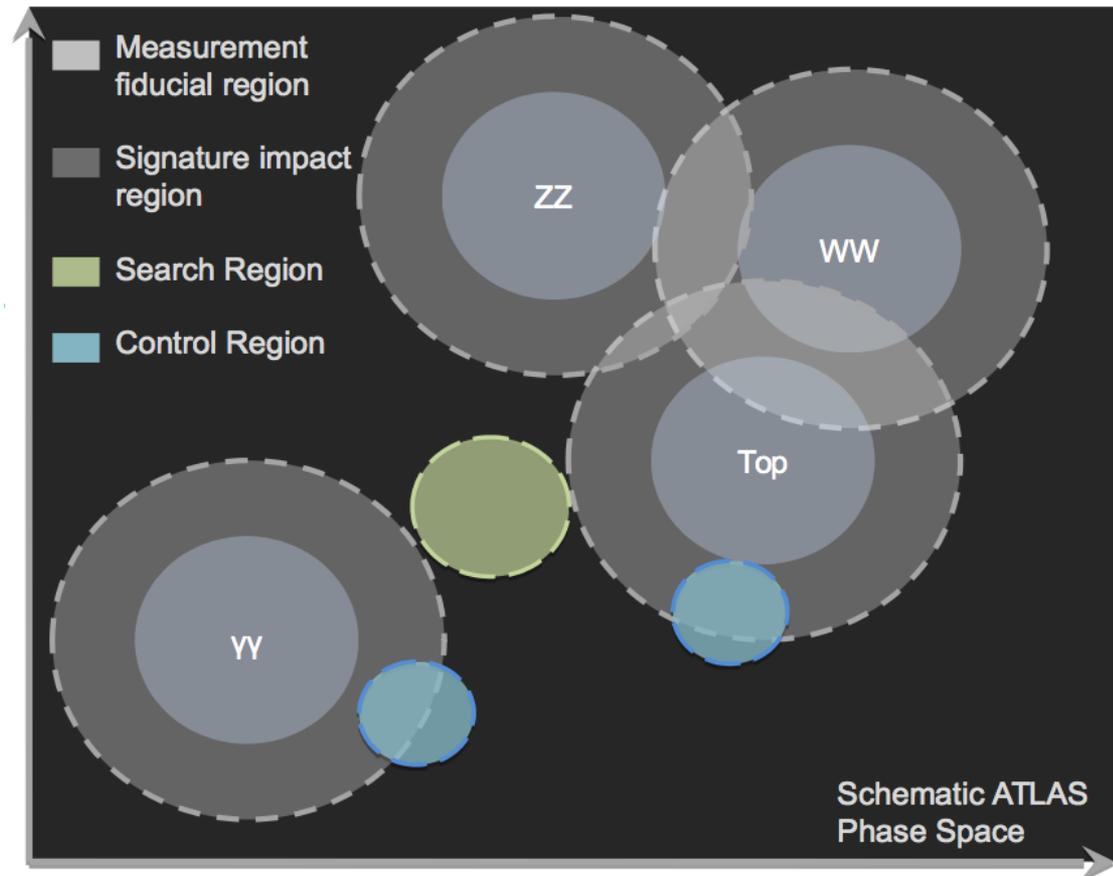
Make more from your BSM searches: making unfolded measurements in control regions for generator tuning and reinterpretation

Louie Corpe (UCL), Jon Butterworth (UCL), Gabriel Facini (UCL)

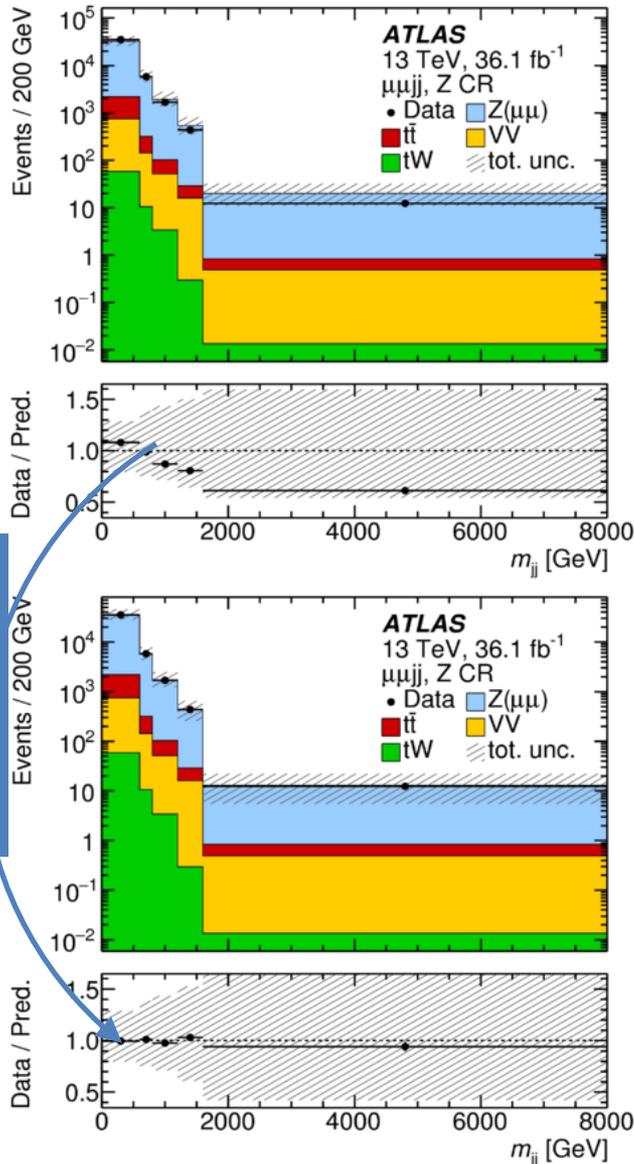
(Re)interpreting the results of new physics searches at the LHC
Imperial College, 03.04.19

Control Region Measurements... why?

- **Measurements:** fiducial volumes enhancing selected process
- **Searches** define Control Regions (CRs) to test MC in more unusual phase space
- **Turn CRs into measurement regions?** ie particle-level cross-sections (XSs)
- **Extra data for MC tuning/validation**
- Could be used to help **exclude certain models** ? (eg [CONTUR](#))



Real-life example: Bkg mis-modelling



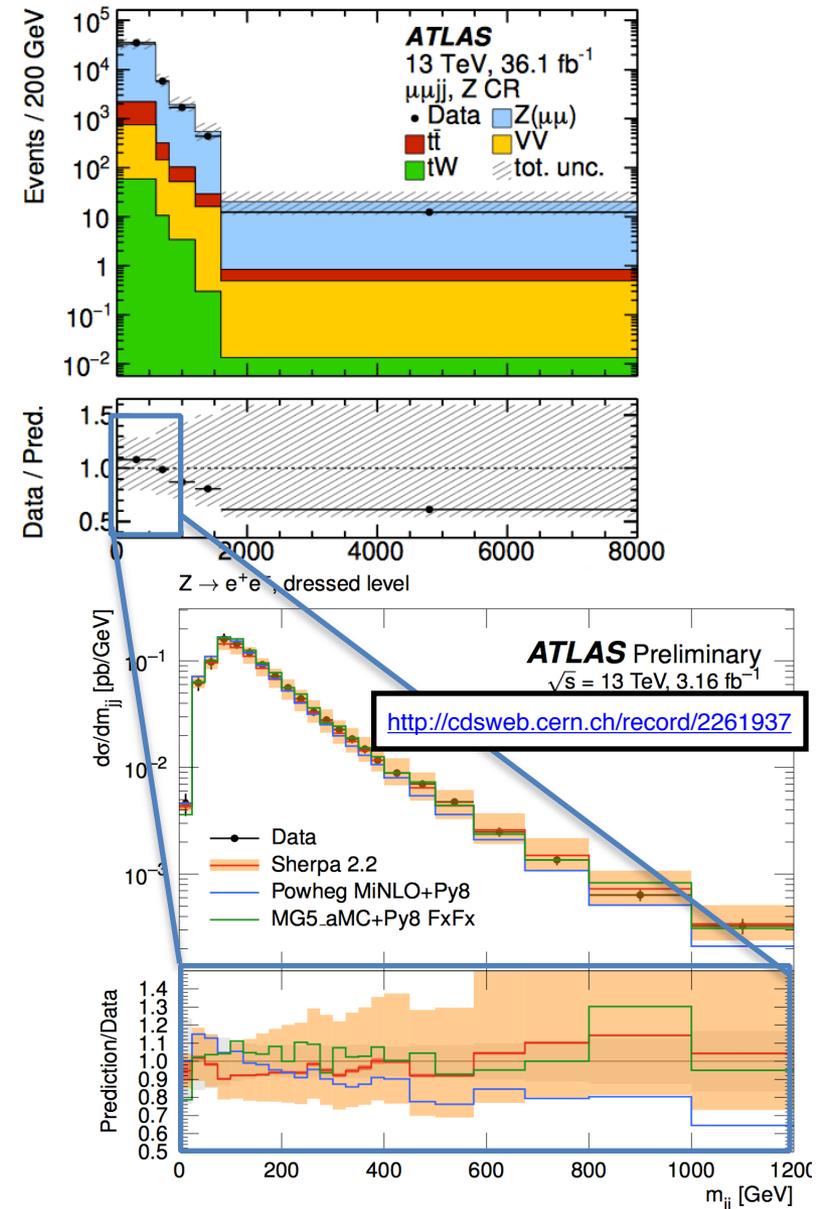
- Recent search for lepto-quarks: mismodelled Z+jets bkgd in Sherpa 2.2.1
- Resolved using ad-hoc m_{jj} re-weighting... but → additional uncertainty to search
- Similar mismodelling of jet-energy variables also in several other ATLAS exotics searches

→ What happened?

→ How can this be avoided in future?

How did this happen?

- MC tuned+validated using available Z+jets measurements (3.2/fb, 13 TeV)
 - Looser lepton selection, less “extreme” phase space...
 - Generator param choices not appropriate for this more extreme phase space
- **Can't validate for phase space where no measurements exist**
- **Let's make measurements we need**
 - improve bkg modelling!
 - avoid this situation in future!



So we made the measurements



- Paper contains 2 results
 - Usual limits for search
 - Extra **particle-level measurements** in several regions related to CRs
- One of **first of its kind in ATLAS**
- **Simple measurements..**
- **...not aiming for precision like SM group!**

arXiv:1902.00377v1 [hep-ex] 1 Feb 2019

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)



CERN-EP-2018-262
4th February 2019

Searches for scalar leptoquarks and differential cross-section measurements in dilepton–dijet events in proton–proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with the ATLAS experiment

The ATLAS Collaboration

Searches for scalar leptoquarks pair-produced in proton–proton collisions at $\sqrt{s} = 13$ TeV at the Large Hadron Collider are performed by the ATLAS experiment. A data set corresponding to an integrated luminosity of 36.1 fb^{-1} is used. Final states containing two electrons or two muons and two or more jets are studied, as are states with one electron or muon, missing transverse momentum and two or more jets. No statistically significant excess above the Standard Model expectation is observed. The observed and expected lower limits on the leptoquark mass at 95% confidence level extend up to 1.25 TeV for first- and second-generation leptoquarks, as postulated in the minimal Buchmüller–Rückl–Wyler model, assuming a branching ratio into a charged lepton and a quark of 50%. In addition, measurements of particle-level fiducial and differential cross sections are presented for the $Z \rightarrow ee$, $Z \rightarrow \mu\mu$ and $t\bar{t}$ processes in several regions related to the search control regions. Predictions from a range of generators are compared with the measurements, and good agreement is seen for many of the observables. However, the predictions for the $Z \rightarrow \ell\ell$ measurements in observables sensitive to jet energies disagree with the data.

<https://arxiv.org/abs/1902.00377>

How was it done?

Which Regions?

MR	Dominant process (purity)	Required leptons and jets	$m_{\ell\ell}$ selection	S_T selection	Remark
$eejj$	$Z \rightarrow ee$ (93%)	$= 2e ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	-	Identical to Z CR
$\mu\mu jj$	$Z \rightarrow ee$ (93%)	$= 2\mu ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	-	Identical to Z CR
$e\mu jj$	$t\bar{t} \rightarrow e\mu$ (93%)	$= 1\mu, 1e ; \geq 2\text{jets}$	-	-	-
Extreme $eejj$	$Z \rightarrow ee$ (94%)	$= 2e ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	$S_T > 600 \text{ GeV}$	-
Extreme $\mu\mu jj$	$Z \rightarrow ee$ (94%)	$= 2\mu ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	$S_T > 600 \text{ GeV}$	-
Extreme $e\mu jj$	$t\bar{t} \rightarrow e\mu$ (86%)	$= 1\mu, 1e ; \geq 2\text{jets}$	-	$S_T > 600 \text{ GeV}$	-

- **Measurements in 6 regions**
- 2 identical to Z-CRs of search
- 1 extra e- μ region without m_{ll} cut around Z-mass
- 3 identical to the above, but ‘Extreme’, ie with $S_T > 600 \text{ GeV}$ ($S_T :=$ scalar sum of 2 lepton and ≥ 2 jet p_T)
- These regions may be useful for **tuning or re-interpretation**

How was it done?

Which Processes?

MR	Dominant process (purity)	Required leptons and jets	$m_{\ell\ell}$ selection	S_T selection	Remark
$eejj$	$Z \rightarrow ee$ (93%)	$= 2e ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	-	Identical to Z CR
$\mu\mu jj$	$Z \rightarrow ee$ (93%)	$= 2\mu ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	-	Identical to Z CR
$e\mu jj$	$t\bar{t} \rightarrow e\mu$ (93%)	$= 1\mu, 1e ; \geq 2\text{jets}$	-	-	-
Extreme $eejj$	$Z \rightarrow ee$ (94%)	$= 2e ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	$S_T > 600 \text{ GeV}$	-
Extreme $\mu\mu jj$	$Z \rightarrow ee$ (94%)	$= 2\mu ; \geq 2\text{jets}$	$70 < m_{\ell\ell} < 110 \text{ GeV}$	$S_T > 600 \text{ GeV}$	-
Extreme $e\mu jj$	$t\bar{t} \rightarrow e\mu$ (86%)	$= 1\mu, 1e ; \geq 2\text{jets}$	-	$S_T > 600 \text{ GeV}$	-

- Each region dominated by single SM process
- Make **measurements for that process exclusively**, by subtracting other contributions from data

Which Variables?

- Consulted with ATLAS Physics Modelling Group to pick **variables commonly used in MC tuning**
 - Inclusive cross-section
 - Differential cross-sections for 11 variables:
 - Jet energy variables: H_T , $p_T(j_0)$, $p_T(j_1)$, M_{jj} ,
 - Jet angle variables: $\Delta\eta(jj)$, $\Delta\phi(jj)$,
 - Lepton variables: $\Delta\phi(l)$, $p_T(l)$,
 - Lepton-jet system: $\min(\Delta\phi(j_0,l))$, $\min(\Delta\phi(j_1,l))$, S_T

Selections



Reco Electrons

$p_T > 40\text{GeV}$,
 $|\eta| < 2.47$, not $[1.37, 1.52]$
 $d_0^{\text{sig}} < 5$, $|\Delta z_0 \sin\theta| < 0.5\text{mm}$
Loose ID

Truth Electrons

dressed $p_T > 40\text{GeV}$,
 $|\eta| < 2.5$
stable, not from hadron decay
IPDG IDI = 11

Reco Muons

$p_T > 40\text{ GeV}$
 $|\eta| < 2.5$
 $d_0^{\text{sig}} < 5$, $|\Delta z_0 \sin\theta| < 0.5\text{mm}$
Combined ID/MS, quality/isolation cuts

Truth Muons

dressed truth $p_T > 40\text{GeV}$,
 $|\eta| < 2.5$
stable, not from hadron decay
|pdg IDI = 13

Reco Jets

Anti-kT R=0.4
 $p_T > 60\text{ GeV}$
 $|\eta| < 2.5$
pass JVT veto

Truth Jets

Anti-kT R=0.4 (no μ or ν)
truth $p_T > 60\text{ GeV}$
 $|\eta| < 2.5$

Reco overlap removal

Remove jets within $\Delta R < 0.2$ of eles
Remove eles within ΔR_{slide} of remaining jets
Remove μ s within ΔR_{slide} of jets if > 3 tracks
 $\Delta R_{\text{slide}} = \min(0.4, 0.04 + 10/p_T^{\text{lep}})$

Truth overlap removal

Remove jets within $\Delta R < 0.2$ of μ/e
Remove μ/e within $\Delta R < 0.4$ of jets

Bin-by-bin unfolding

In this method, the particle-level cross-section for the dominant process p in bin i of a distribution can be expressed as:

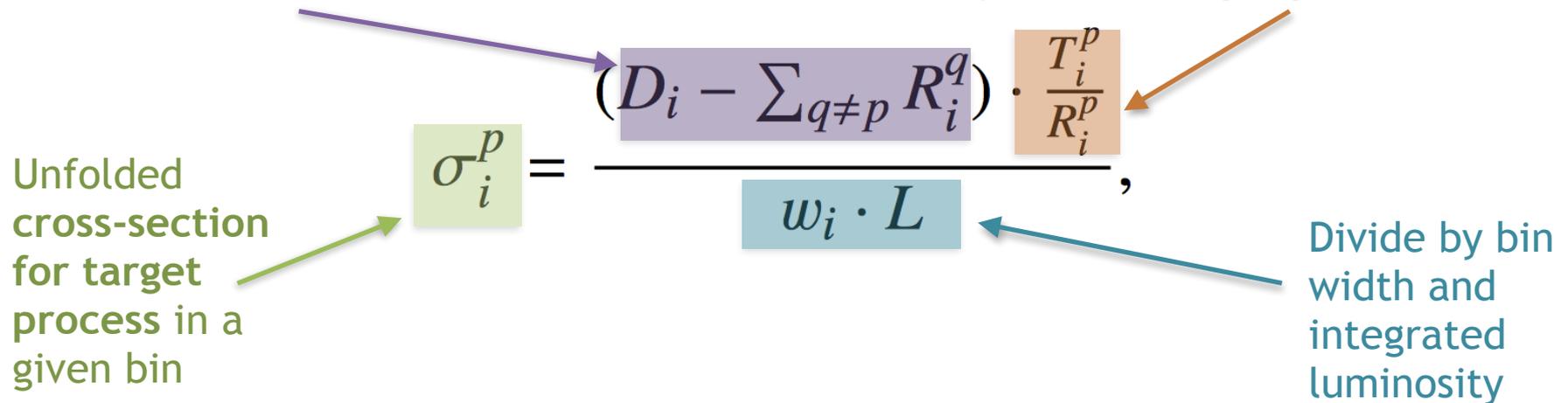
Subtract reco yield of non-target processes from the data in each bin.
eg for eejj ZCR: subtract tt, Wt, Zμμ, QCD fakes...

Multiply by bin-by-bin correction factor given by ratio of truth/reco yield for target process in each bin

Unfolded cross-section for target process in a given bin

$$\sigma_i^p = \frac{(D_i - \sum_{q \neq p} R_i^q) \cdot \frac{T_i^p}{R_i^p}}{w_i \cdot L}$$

Divide by bin width and integrated luminosity



where the sum runs over all sub-dominant SM processes which contribute to the yield, D_i is the number of events observed in data in the bin, R_i^t (T_i^t) is the yield at detector-level (particle-level) for process t , w_i is the width of the bin and L is the integrated luminosity of the data sample. The bin-by-bin correction

Method requires negligible bin migrations. Choose binning such that purity (% of events in same bin at Truth/Reco level) $\sim 90\%$ in all bins

- **Experimental uncertainties** usually same as for search. Dominant ones are JES+JER ($\sim 5\%$ on inclusive XS)
- **Theory and Modelling uncertainties** largely cancel in correction factor ratio (up to $< 2\%$ on inclusive XS)
- **Closure tests**: calculating correction factors with alt samples, and turning m_{jj} -reweighing On/Off. ($\sim 2\%$ on inclusive XS)
- **Statistical Uncertainties** evaluated with **Bootstrap replicas**. Quite small uncertainty.

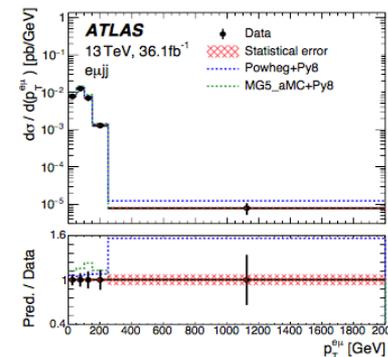
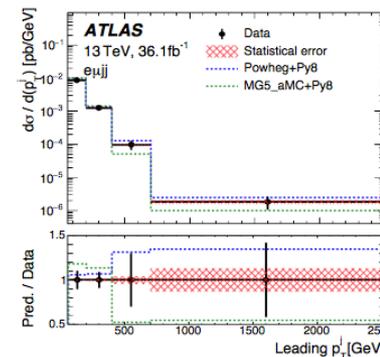
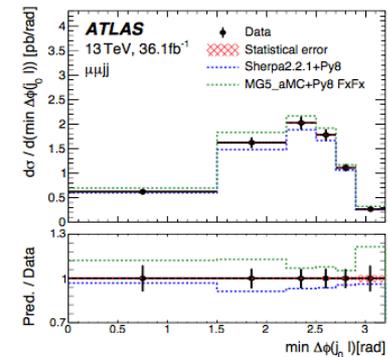
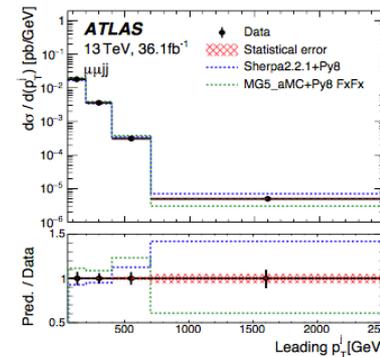
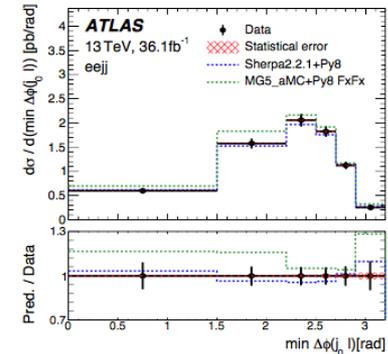
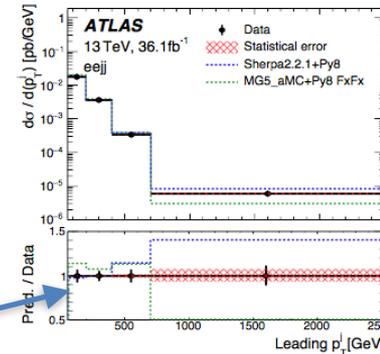
• **Total uncertainty on inclusive XS is O(5-9%) depending on the process.**

• Not aiming for precision measurements here!

A selection of results

- **Measurements in 6 regions**
~11 distributions unfolded in each
 - 66 measurements made public.
 - In particular jet variables showing the mismodelling!
- +inclusive cross-sections measurements:

Measurement region	Dominant process	Measured cross-section [pb]
$eejj$	$Z \rightarrow ee$	3.28 ± 0.02 (stat.) ± 0.19 (exp.) ± 0.09 (th.)
$\mu\mu jj$	$Z \rightarrow \mu\mu$	3.32 ± 0.02 (stat.) ± 0.21 (exp.) ± 0.09 (th.)
$e\mu jj$	$t\bar{t} \rightarrow e\mu$	1.503 ± 0.010 (stat.) ± 0.130 (exp.) ± 0.002 (th.)
Extreme $eejj$	$Z \rightarrow ee$	0.483 ± 0.005 (stat.) ± 0.024 (exp.) ± 0.014 (th.)
Extreme $\mu\mu jj$	$Z \rightarrow \mu\mu$	0.481 ± 0.004 (stat.) ± 0.022 (exp.) ± 0.014 (th.)
Extreme $e\mu jj$	$t\bar{t} \rightarrow e\mu$	0.147 ± 0.003 (stat.) ± 0.011 (exp.) ± 0.001 (th.)



- Imminently on Rivet + HEPData ... w/ full uncertainty breakdown
→ **can build covariance matrix**

- Recent work w/ HEPData + Rivet/YODA authors...
- Format for uncert. breakdown in YODA files
- Seamless propagation of uncert. breakdown from HEPdata to YODA
- [Library of fcns to manipulate YODAs](#) with uncert. breakdowns → build cov matrix
- Use for eg X^2 or simplified likelihood (**no more just looking at a ratio plot!**)

Upload New Files

Download All

Filter 71 data tables

Meas. of leading p_T^j in $eejj$

Figure 7
10.17182/hepdata.83968.v1/t2
Differential cross-section and uncertainty from each source, as a function of leading p_T^j for the dominant process in the $eejj$...

passed review

Meas. of leading p_T^j in $\mu\mu jj$

Figure 7
10.17182/hepdata.83968.v1/t3
Differential cross-section and uncertainty from each source, as a function of leading p_T^j for the dominant process in the $\mu\mu jj$...

passed review

Meas. of leading p_T^j in $e\mu jj$

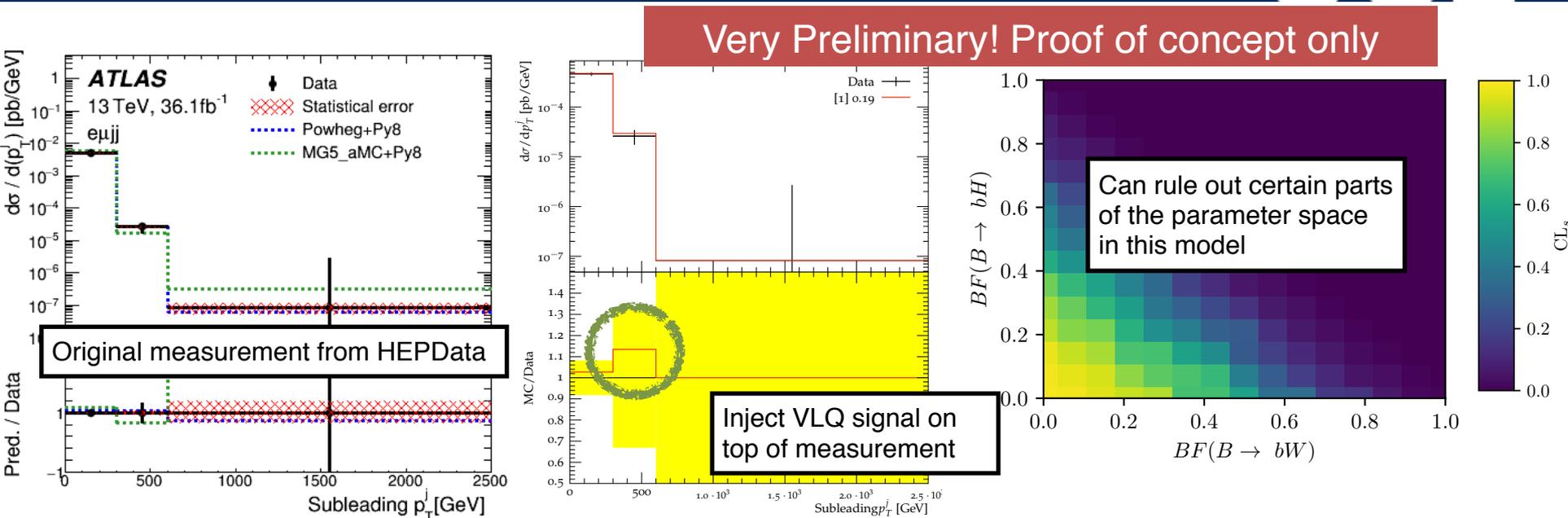
Figure 7
10.17182/hepdata.83968.v1/t4
Differential cross-section and uncertainty from each source, as a function of leading p_T^j for the dominant process in the $e\mu jj$...

passed review

leading p_T^j [GeV]	$d\sigma/d(p_T^j)$ [pb/GeV]
60.0 - 200.0	0.017507 $\pm 6.7625e-05$ sys_tbar_aMC_at_NLO ± 0.00019807 sys_tbar_TH_Scale_uncertainty ± 0.00099817 sys_JET_GroupedNP_1 $\pm 9.2897e-06$ sys_EG_SCALE_ALL $\pm 4.5866e-05$ sys_EL_EFF_Reco_TOTAL_INPCOR_PLUS $\pm 7.5596e-05$ stat_Data_statistics ± 0.00019783 sys_DY_TH_PDF_uncertainty ± 0.00037685 sys_Luminosity $\pm 6.7625e-05$ sys_tbar_dilepton $\pm 9.8648e-05$ sys_JET_JER_SINGLE_NP $\pm 7.0715e-05$ sys_EG_RESOLUTION_ALL ± 0.00013547 sys_DY_TH_ALTPDF_uncertainty ± 0.00014821 sys_DY_TH_AlphaS_uncertainty $\pm 3.4694e-18$ sys_MUON_SAGITTA_RHO $\pm 6.8427e-05$ sys_QCD_fakes ± 0.00018672 sys_EL_EFF_ID_TOTAL_INPCOR_PLUS_UN $\pm 6.7625e-05$ sys_tbar_radiation_up_up ± 0.00013381 stat_MC_statistics $\pm 2.4377e-05$ sys_tbar_TH_PDF_uncertainty $\pm 1.6749e-05$ sys_tbar_TH_ALTPDF_uncertainty $\pm 6.7625e-05$ sys_tbar_Colour_reconnection $\pm 6.7625e-05$ sys_tbar_radiation_down ± 0.00012973 sys_JET_GroupedNP_3 $\pm 8.4901e-05$ sys_JET_EtaInterCalibration_NonClosure $\pm 6.3619e-05$ sys_EL_EFF_Iso_TOTAL_INPCOR_PLUS_U ± 0.00011768 sys_DY_MJ_RW_off $\pm 3.768e-05$ sys_EL_EFF_Trigger_TOTAL_INPCOR_PLU ± 0.00033601 sys_DY_TH_Scale_uncertainty $\pm 9.3737e-05$ sys_JET_GroupedNP_2
200.0 - 400.0	0.0035965 $\pm 3.5515e-05$ sys_tbar_TH_Scale_uncertainty ± 0.00011375 sys_JET_GroupedNP_1 $\pm 2.0998e-06$ sys_EG_SCALE_ALL $\pm 1.2472e-05$ sys_EL_EFF_Reco_TOTAL_INPCOR_PLUS $\pm 2.8393e-05$ stat_Data_statistics $\pm 7.4598e-05$ sys_DY_TH_PDF_uncertainty $\pm 7.742e-05$ sys_Luminosity $\pm 1.393e-05$ sys_JET_JER_SINGLE_NP $\pm 6.2952e-06$ sys_EG_RESOLUTION_ALL $\pm 6.5584e-05$ sys_DY_TH_ALTPDF_uncertainty $\pm 3.0197e-05$ sys_DY_TH_AlphaS_uncertainty $\pm 6.505e-19$ sys_MUON_SAGITTA_RHO $\pm 2.0517e-05$ sys_QCD_fakes $\pm 5.4707e-05$ sys_EL_EFF_ID_TOTAL_INPCOR_PLUS_UN $\pm 1.0702e-05$ stat_MC_statistics $\pm 3.9663e-06$ sys_tbar_TH_PDF_uncertainty $\pm 2.7447e-06$ sys_tbar_TH_ALTPDF_uncertainty $\pm 2.5698e-05$ sys_JET_GroupedNP_3 $\pm 1.593e-05$ sys_JET_EtaInterCalibration_NonClosure $\pm 3.702e-05$ sys_EL_EFF_Iso_TOTAL_INPCOR_PLUS_U $\pm 4.5718e-06$ sys_DY_MJ_RW_off $\pm 9.5775e-06$ sys_EL_EFF_Trigger_TOTAL_INPCOR_PLU $\pm 3.6041e-05$ sys_DY_TH_Scale_uncertainty $\pm 2.4987e-05$ sys_JET_GroupedNP_2
400.0 -	0.00033471 $\pm 4.8872e-06$ sys_tbar_TH_Scale_uncertainty

- **ATLAS Physics Modelling Group** preparing next gen of background samples
 - Including new V+jets samples in latest Sherpa (3.x) version
 - Integration and development happening soon
 - When validating new generator configuration, CR measurements will be used
 - → validate new samples in more extreme phase space directly, without need for reconstruction
 - better background modelling → more sensitive searches

How to use this for re-interpretation?

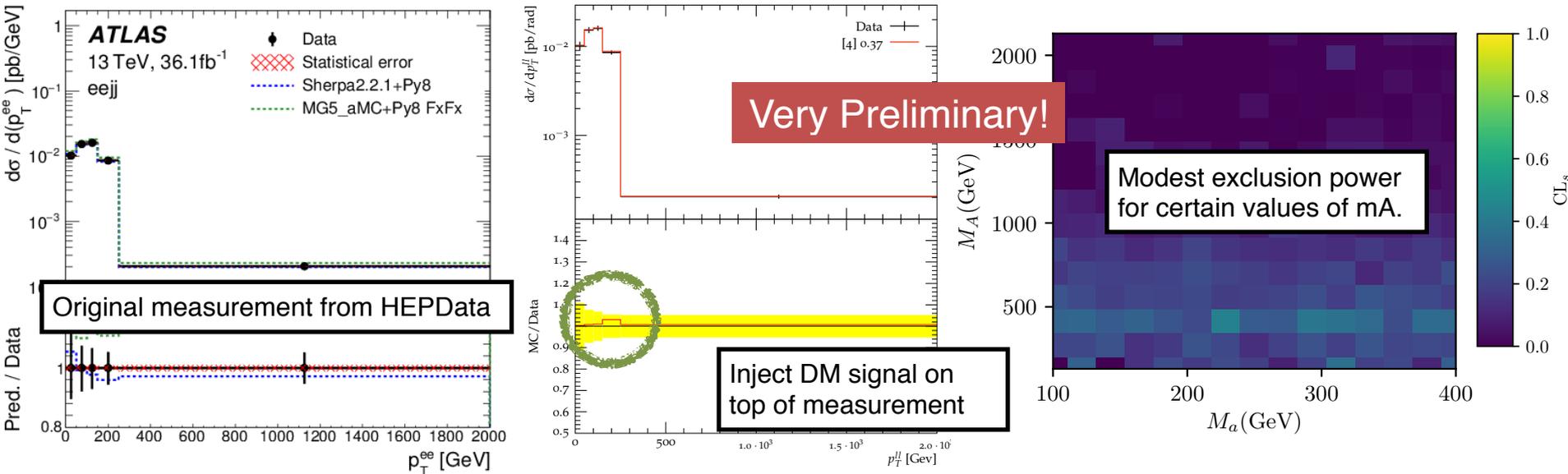


- Example: generic vector-like quark model with pair production of the B quark dominating, coupled only to the 3rd gen quarks
- CONTUR assumes data==SM and injects a signal on top
- $e\mu jj$ measurements rule out certain regions of parameter space
- Could be used in conjunction with other measurements to rule out larger areas of parameter space!

- We **can ask more of our searches**, and extract more information from the CRs for posterity
- Discussed **prototype paper** where such **measurements are made**:
 - Simple unfolding method, can be done by **1- or 2-person team without slowing down search...**
 - **Not competing with SM group precision!**
- Will be used to **validate next generation of ATLAS samples**
→better bkg modelling →better sensitivity for searches
- Measurements+CONTUR: directly **scan BSM model param spaces**
- Could be a **template for many full run 2 search papers?**
- ... why stop at control regions? If no excess, also do search region. Even more sensitive/useful for re-interpretation.
→working on it, watch this space.

[Additional Material]

How to use this for re-interpretation?



- Example: Dark Matter 2HDM model
 - CONTUR assumes data==SM and injects a signal on top.
 - Modest exclusion power in this model...
 - ... just to show that machinery works

Ok fine, this is quite neat... How could I do it in my analysis?



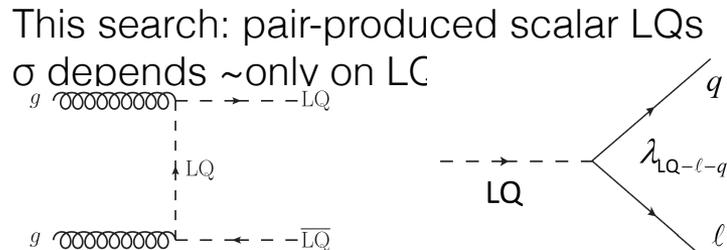
1. Decide **which regions you want to unfold** -> Measurement regions MRs
2. Determine **truth-level selection**.
3. Pick an **unfolding method**. (eg **bin-by-bin unfolding**)
4. Decide **which variables to unfold**.
5. **Fill 2D response matrices** (TH2D): x-axis=truth quantity, y-axis=reco quantity
6. For bin-by-bin unfolding, **optimise binning for negligible migrations between bins**
7. **Calculate correction factors**: x-axis underflow/ y-axis underflow for target process.
8. **Subtract non-target processes from Data...**
9. **..multiply c-factors, divide by bin width and lumi**
10. **Systematics** (theory, detector, closure tests)

--> **Enjoy your nice new unfolded CR measurements,**
and integrate them into your paper.

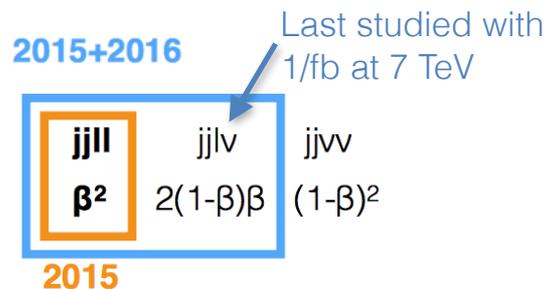


Search

- Leptoquarks (LQs) arise in various unification models, RPV SUSY ...
- LQs could help explain tensions in recent B-sector measurements



- final state: 2 jets + 2 leptons
 main bkg: V+jets , ttbar
- BR for decay to charged leptons = β



Measurement

- SM *measurements* focus on fiducial region enhancing signal process
- *Searches* define control regions (CRs) to test/correct bkg in other regions
- Extract complimentary info from CRs to test SM in more extreme regions?
 measure particle-level differential $d\sigma/dX$
- Why? Improve modelling of dominant backgrounds
- Use pair-produced LQ analysis as proof of principle.
- SM group was involved in review process and OK'd our methods

Visual representation of simple bin-by-bin unfolding

