ATLAS MC strategies and work for Run II
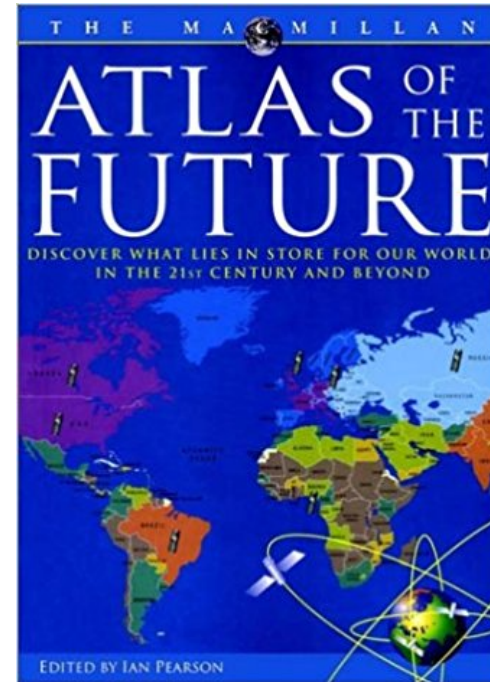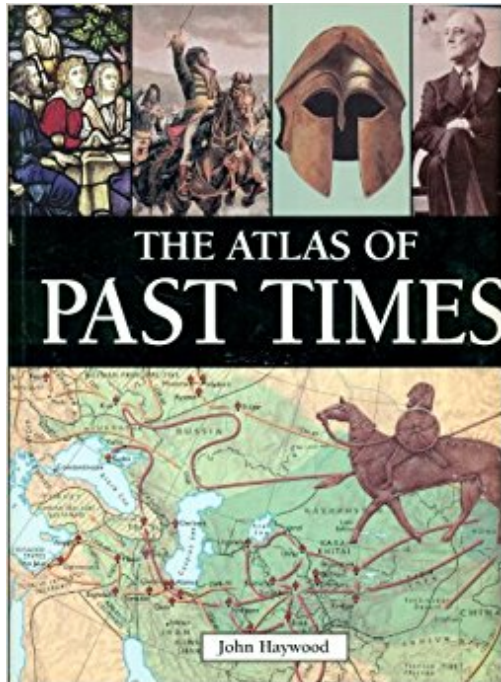
Zach Marshall (LBNL) on behalf of the ATLAS Collaboration

LHC Theory ERC Meeting

22 March 2017

Overview

- Some demographics
- How we use MG5_aMC
  - In searches and for the Standard Model
- Technical implementation and issues
- Future plans and wish list

Demographics

- ~15,000 samples made with MG5_aMC just in the current production campaign!
- 1,125 NLO samples with 761M events
  - 2:1 SM:BSM in terms of number of events
  - 1:6 SM:BSM in terms of number of datasets
- 13,609 LO samples with 1647M events
  - 3:2 SM:BSM in terms of number of events
  - 1:30 SM:BSM in terms of number of datasets!
- ~17% of the total ATLAS MC for the current campaign
- Very helpful interactions with the authors have made much of this production possible; thank you all for that!
  - Stefano Frixione is included in the ACE program, which allows generator authors more direct access to internal information.
 - We are always happy to talk about other projects that might benefit from such an arrangement!

Searches with MG5_aMC

- MG5_aMC entered heavy use in ATLAS for signal production, first for SUSY and later for exotics models
  - Exotics now making heavy use of (FeynRules) UFO models
- Almost every simplified model we use is made with MadGraph5
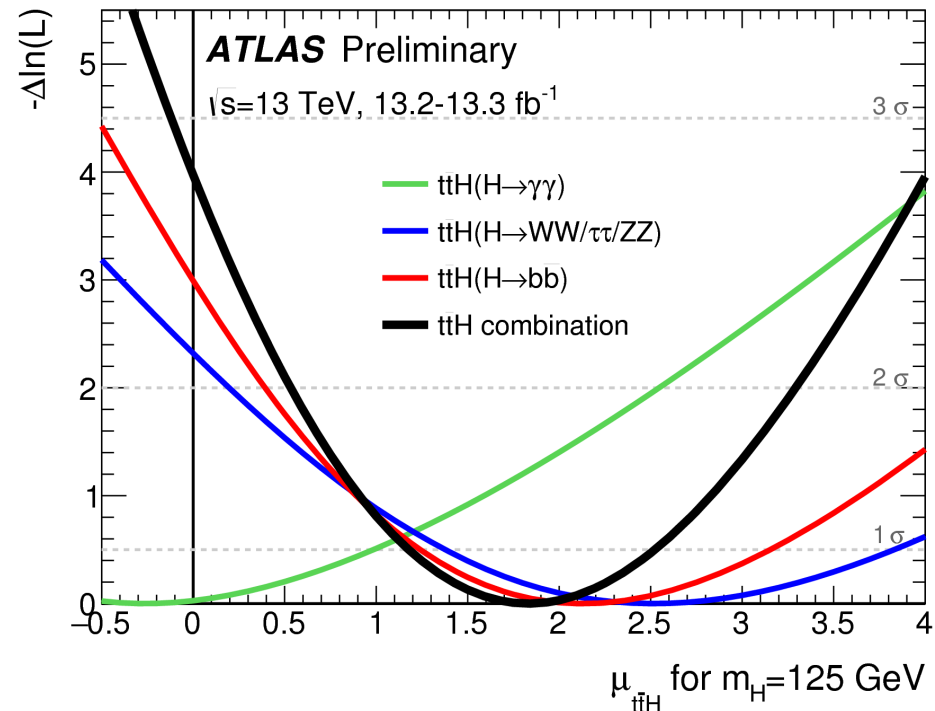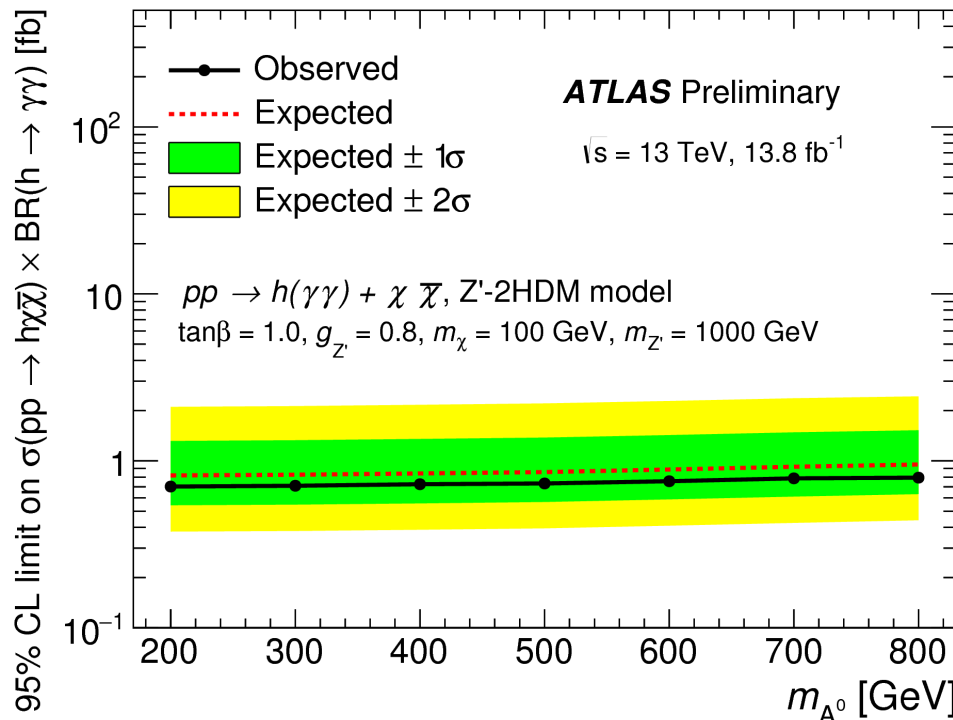 - You can almost pick *any* ATLAS search paper and find a MadGraph5 model!
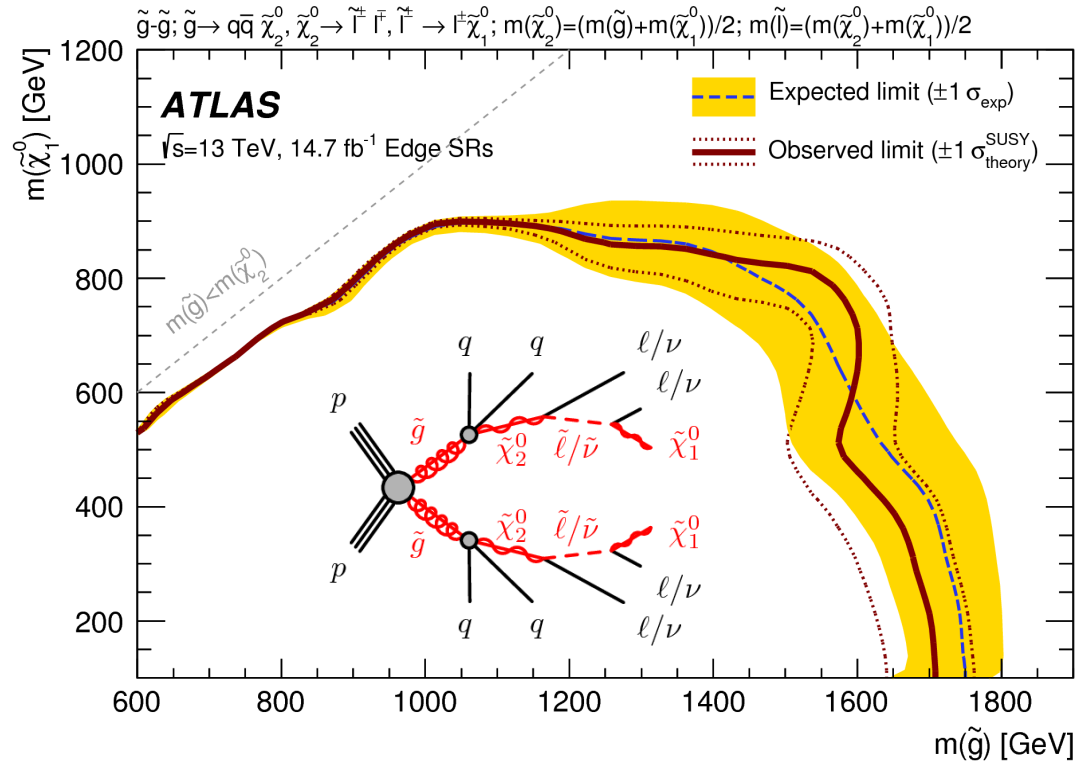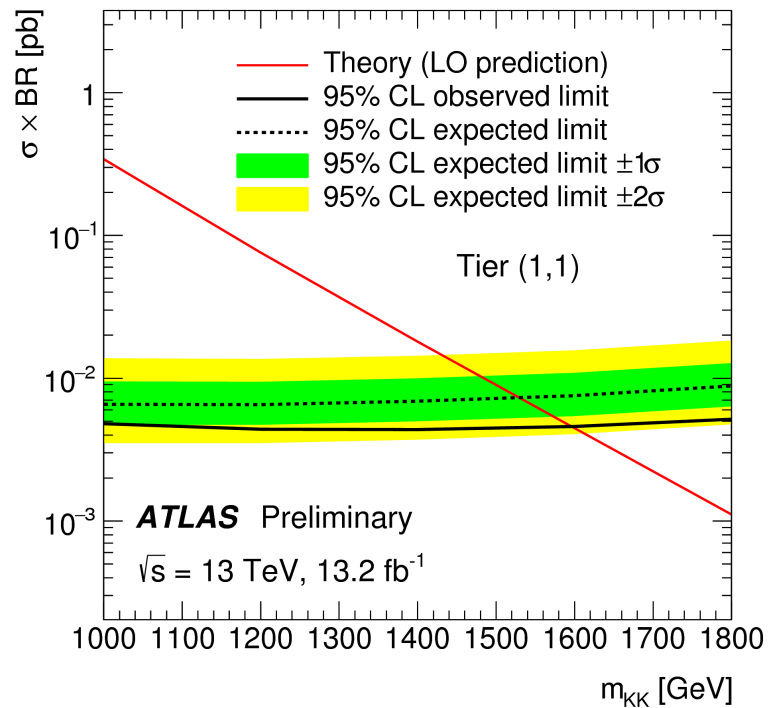
Searches Pushing Further

- A number of simplified models are adding MadSpin now
  - Dealing with top and EWK spin correlations and compressed regions
- Two major papers using MG5_aMC for re-interpretation
 - Huge numbers of processes automatically generated to re-create O(300k) complete pMSSM model points
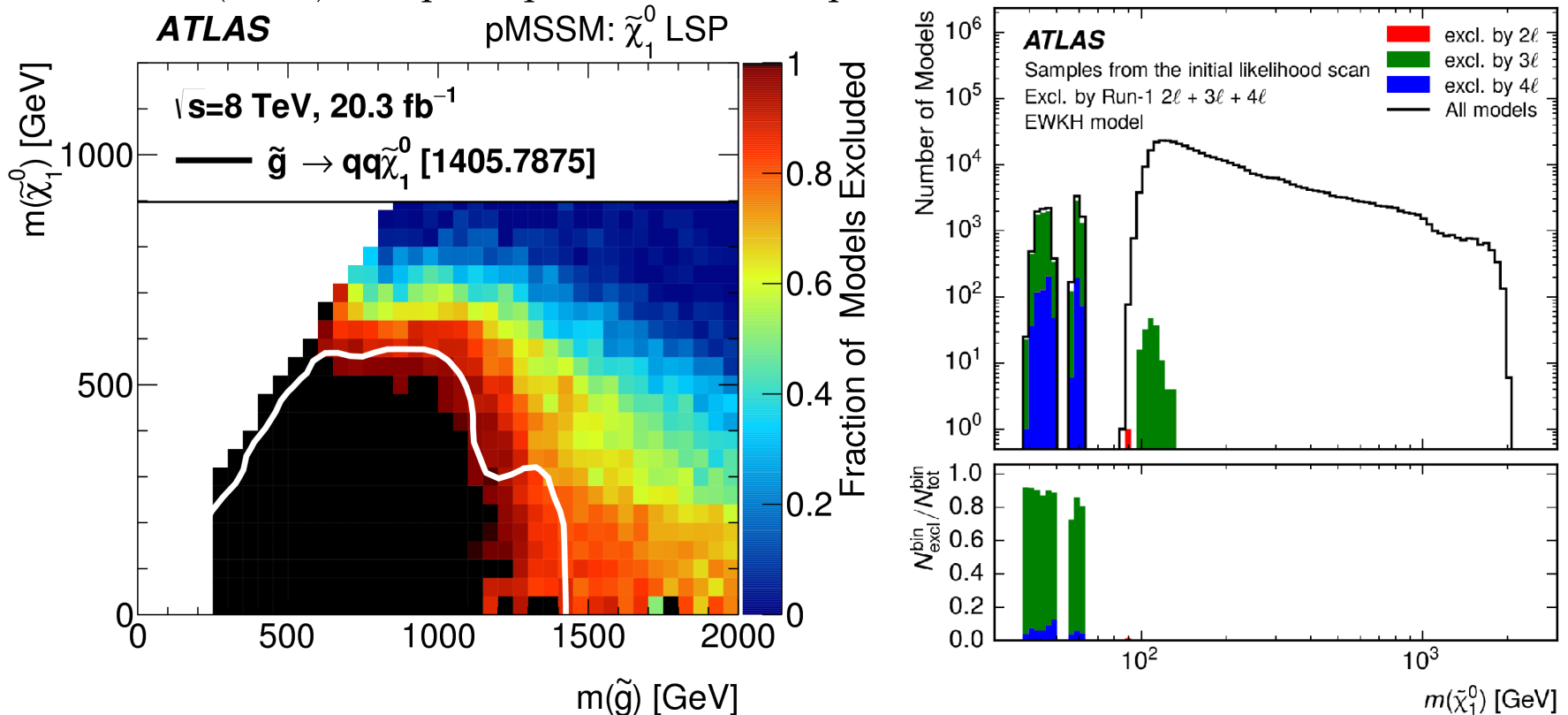
Searches with MG5_aMC (II)

- The Higgs group is using MadGraph5_aMC@NLO for:
  - Exotics Higgs searches (e.g. 2HDM), primarily with models from FeynRules, at both LO and NLO
  - Standard Model searches in new channels, like ttH
  - Standard Model measurements (more in a moment)

Left plot: 95% CL limit on $\sigma(pp \to h\chi\bar{\chi}) \times BR(h \to \gamma\gamma)$ [fb] vs m_{A^0} [GeV]. Legend: Observed, Expected, Expected ± 1σ, Expected ± 2σ. **ATLAS** Preliminary, $\sqrt{s} = 13$ TeV, 13.8 fb^{-1}. $pp \to h(\gamma\gamma) + \chi\,\bar{\chi}$, Z'-2HDM model, $\tan\beta = 1.0$, $g_{Z'} = 0.8$, $m_\chi = 100$ GeV, $m_{Z'} = 1000$ GeV.

Right plot: $-\Delta\ln(L)$ vs $\mu_{t\bar{t}H}$ for $m_H = 125$ GeV. **ATLAS** Preliminary, $\sqrt{s} = 13$ TeV, 13.2-13.3 fb^{-1}. Legend: ttH(H→γγ), ttH(H→WW/ττ/ZZ), ttH(H→bb), ttH combination. Dashed lines at 1σ, 2σ, 3σ.
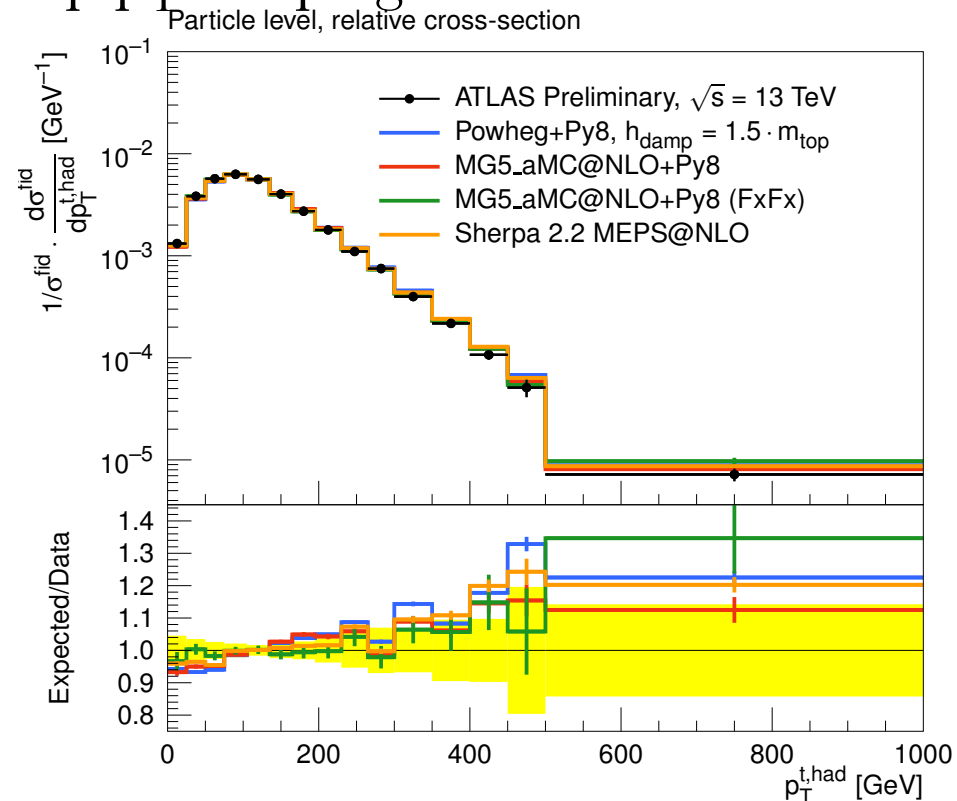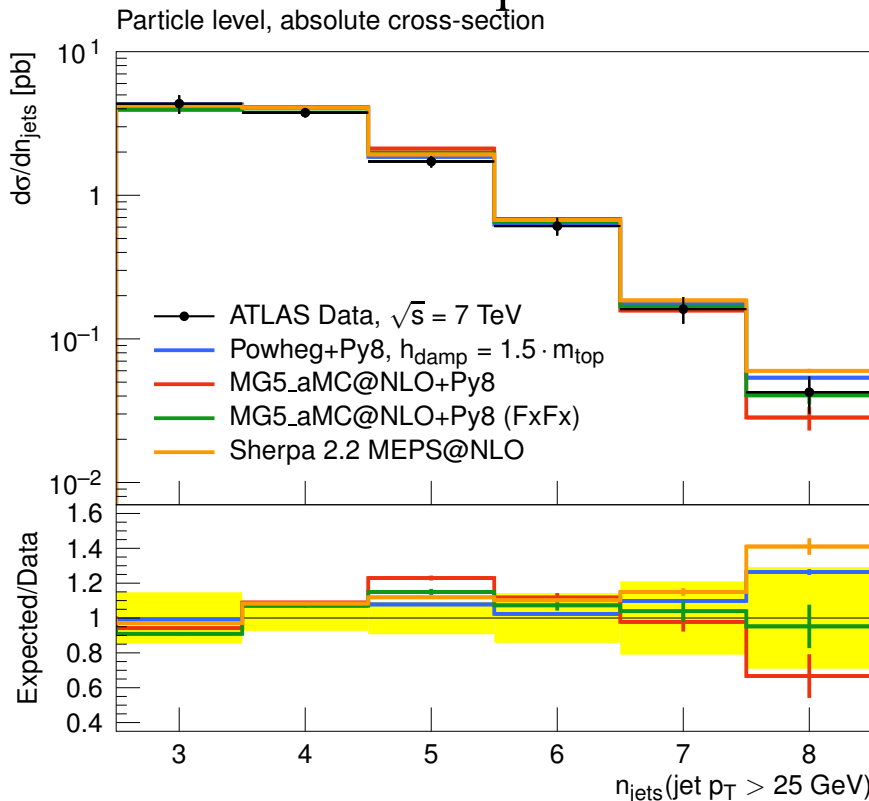
Z+Jets at 13 TeV

- We rely heavily on *slicing* of phase space to populate tails
- Still these samples are massive − O(1B) events for V+jets
 - Being used more often now in ATLAS; agreement with data looks good
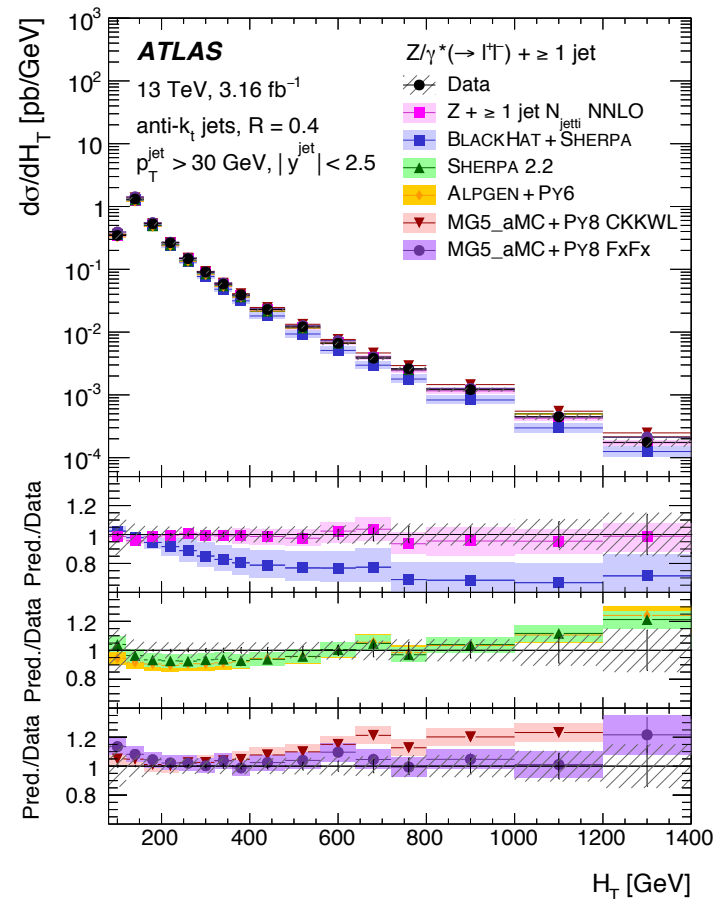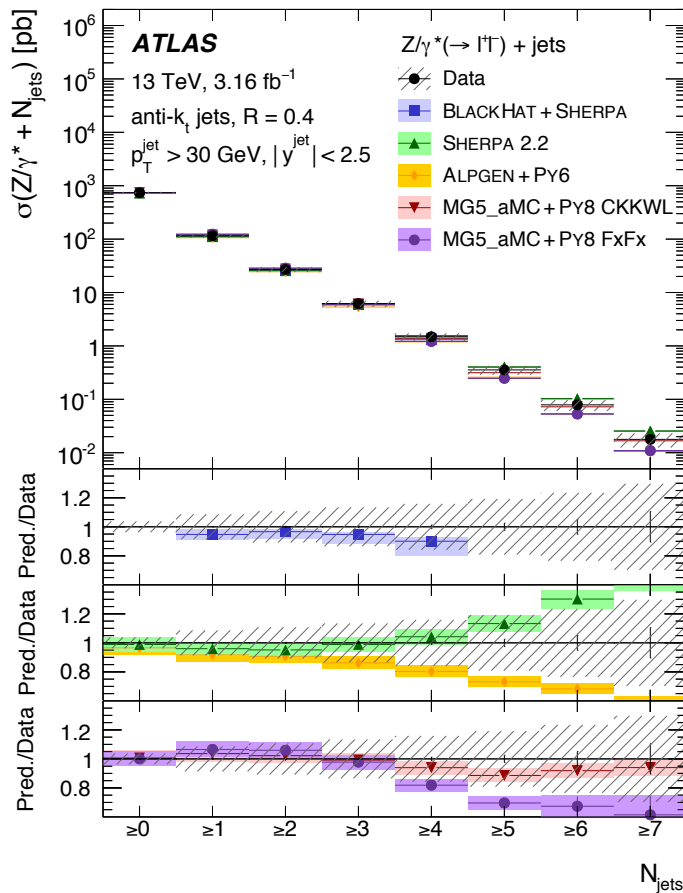
ttbar Modeling

- Working hard to develop a MG5_aMC ttbar sample that could become the baseline sample, especially for analyses that are probing higher jet multiplicity/H_T regions

  – Interplay with showering and tune here still need understanding

- Same notorious problem with the top p_T has plagued us for some time

Particle level, absolute cross-section

Particle level, relative cross-section

ATLAS Data, $\sqrt{s}$ = 7 TeV
Powheg+Py8, h_{damp} = 1.5 · m_{top}
MG5_aMC@NLO+Py8
MG5_aMC@NLO+Py8 (FxFx)
Sherpa 2.2 MEPS@NLO

ATLAS Preliminary, $\sqrt{s}$ = 13 TeV
Powheg+Py8, h_{damp} = 1.5 · m_{top}
MG5_aMC@NLO+Py8
MG5_aMC@NLO+Py8 (FxFx)
Sherpa 2.2 MEPS@NLO

ttbar Modeling

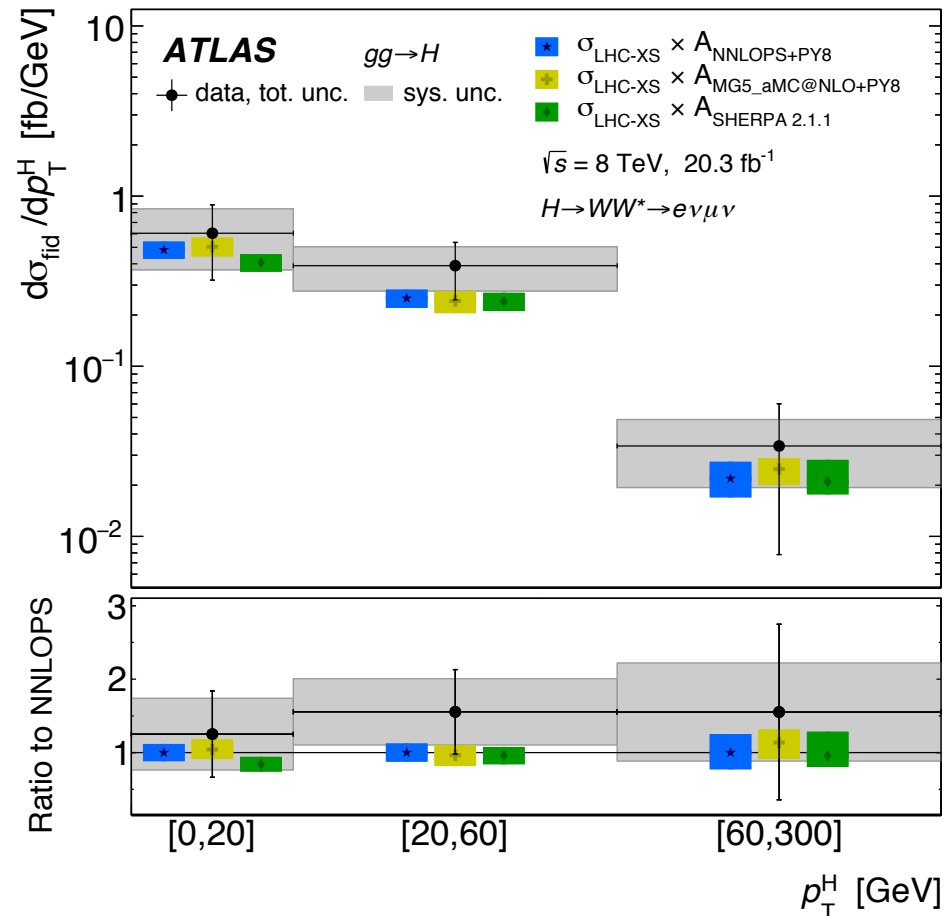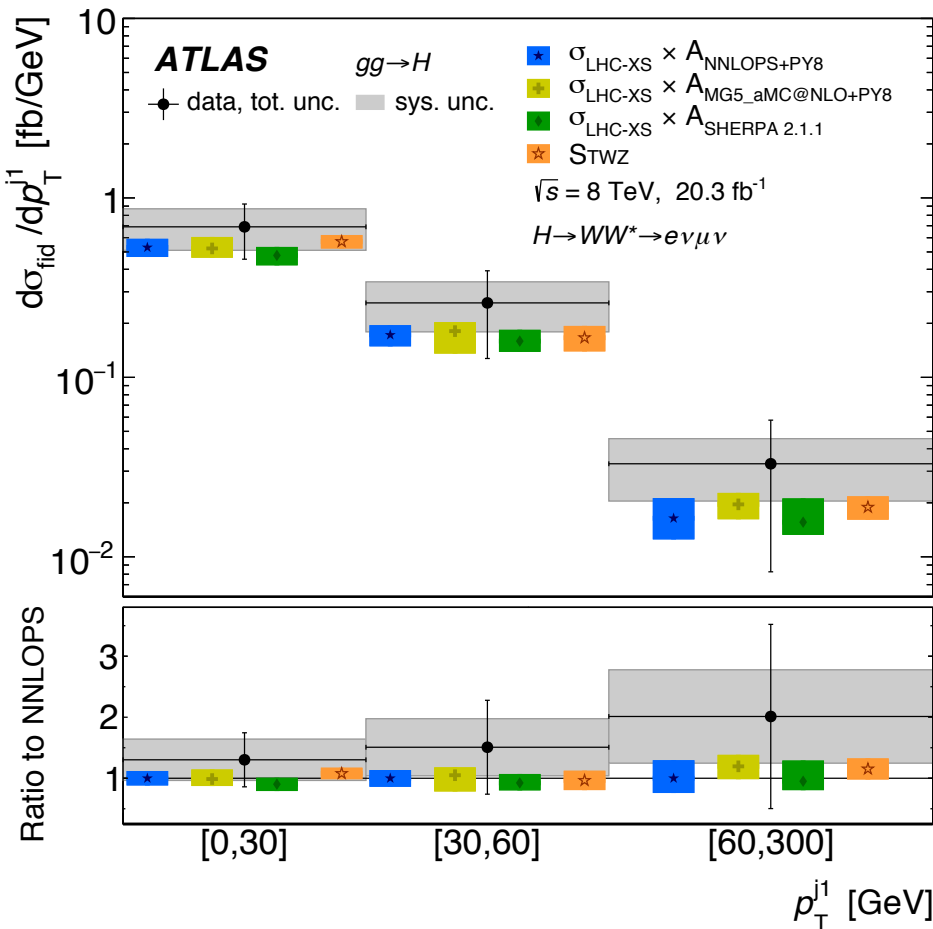- Starting work on WbWb at NLO (feedback will be soon)
- Also looking into how the functional for the scale affects the top p_T
  - Does appear to have significant impact

Particle level, absolute cross-section

Particle level, relative cross-section
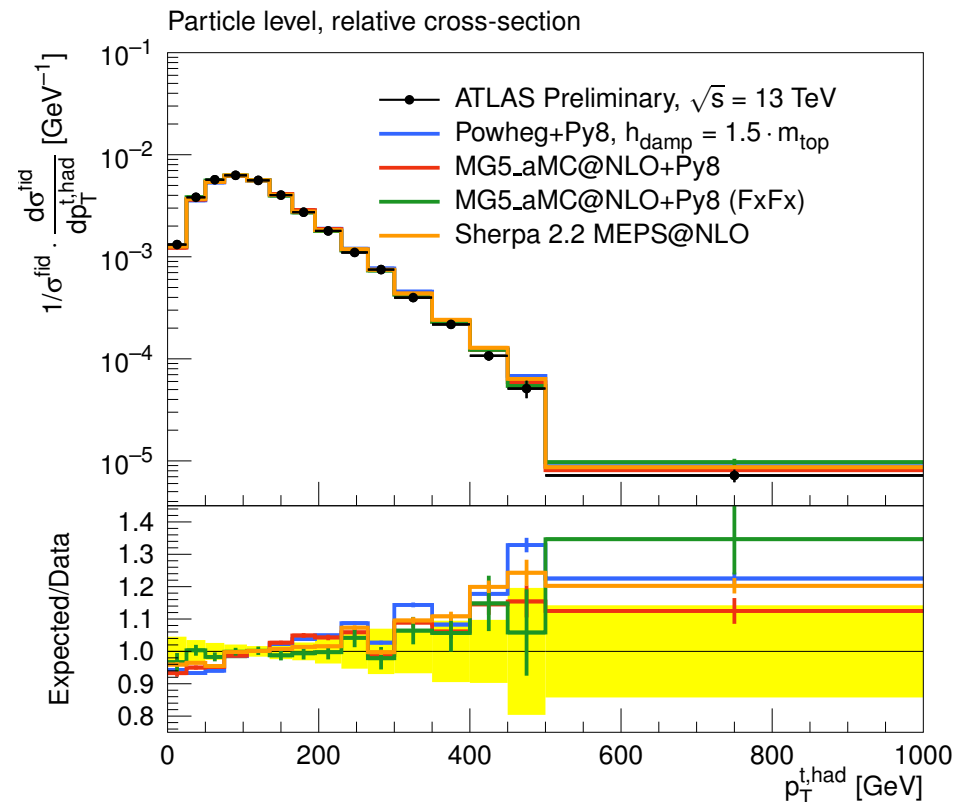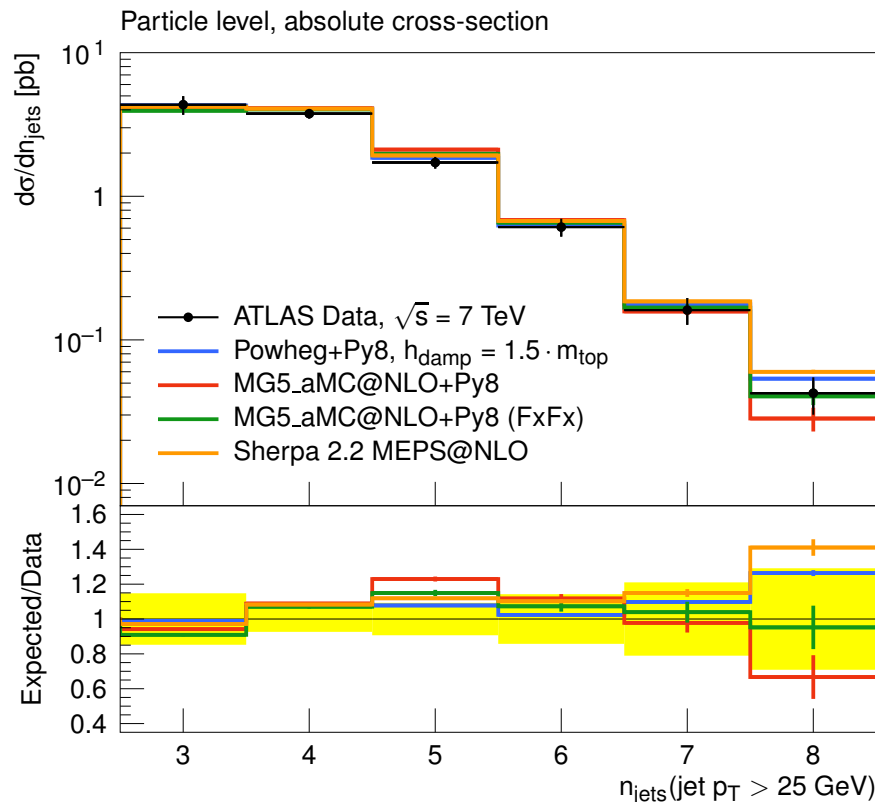
Higgs Modeling (Differential)

- No clear conclusions yet, but expect much higher statistics tests to come soon!
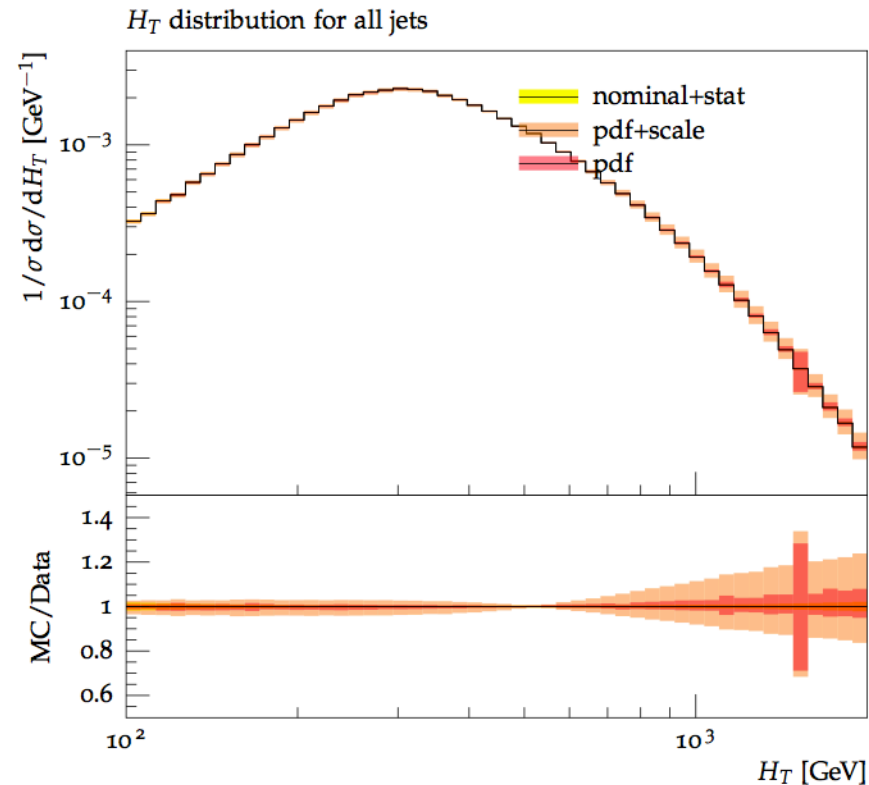
Standard Model

- Scale and PDF uncertainties are taken as the envelope of of the varied samples (e.g. for inclusive NLO ttW)

- LHE3 weights are a huge breakthrough for evaluating uncertainties!
  - Starting to investigate reweighting module for better phase-space population and model parameter scans
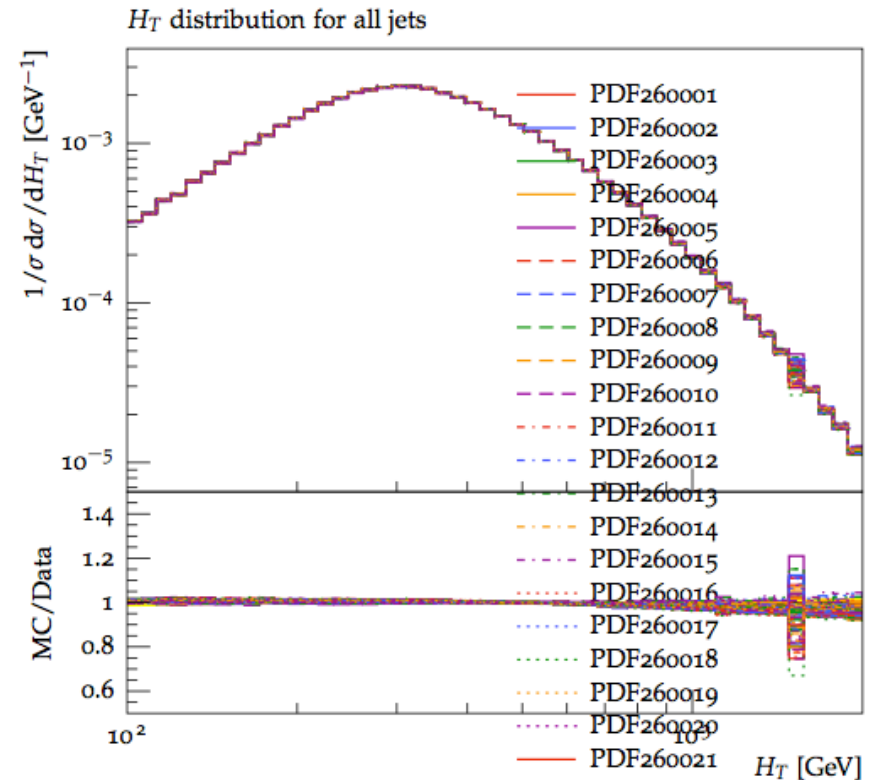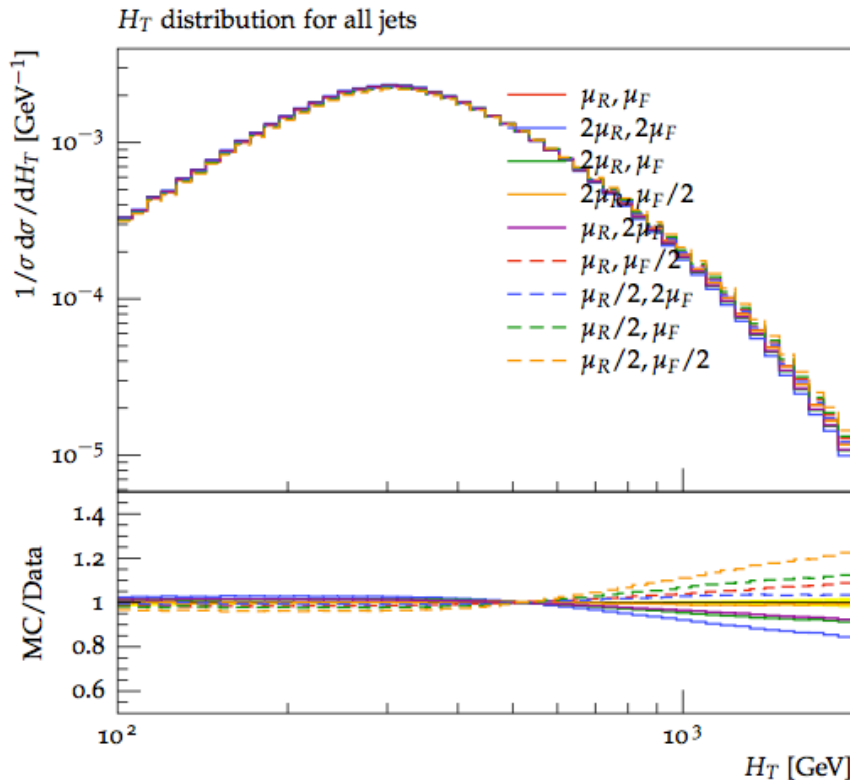
Standard Model (II)

- Now beginning to integrate MG5_aMC into parton shower tunes directly
  - Currently a very heavy process; looking into ways to make it more efficient
  - Need to understand what parameters of MG5_aMC should be tuned per process, and where generic lessons can be used
  - Matching and merging scales and systematics are one example – should we tune these, use 70/30 GeV, evaluate systematics simply by varying up and down by a factor of two or….

- Learning some lessons about improper tune settings
  - Rapidity order off tunes are important!
  - ME does matter to tuning

H_T distribution for all jets

nominal+stat
pdf+scale
pdf

$1/\sigma \, d\sigma/dH_T \; [\mathrm{GeV}^{-1}]$

MC/Data

$H_T \; [\mathrm{GeV}]$

Technical Configuration

- Our primary mode of operation is *generation on the fly*
- This means running MG5_aMC on the Grid
  - More efficient for generation with many model variations – we don't burn time locally generating LHE files and uploading them
  - Also improves reproducibility – the entire job runs in common software with a simple command, so anyone can reproduce the results
  - Can run multi-core jobs in the production system when efficient
- We also run integration for heavy jobs
  - Standard Model processes where integration takes significant time
  - Increasingly, searches with heavy final states (e.g. gluino + 2 jets)
  - As much as possible, this is running the same job as runs in the production system with a single flag set
  - These integrations can be run on local clusters or – soon? – HPC systems, as we do for other generators

Technical Configuration (II)

- Running MG5_aMC in the production system means using read-only, distributed installations of MG5_aMC
  - Lots of help with ensuring these central installations are as useful as possible – thank you for this!
  - Read-only continues to cause some problems, so it might be useful to ensure that this is tested thoroughly
  - These installations shouldn't "phone home"; >10GB/day of syslog messages from mounting `/cvmfs/cp3.uclouvain.be/madgraph`
  - Our installations are done by Genser and can be used by ***anyone***
- Using cvmfs for distribution of external models
  - The use of `PYTHONPATH` has made this *very* straightforward!!
  - Anticipating similar ways to use extensions like plugins in the future
- Running O(100) event weights per job (thanks to LHE3)
  - This has massively reduced our production load and made analyses easier and more thorough
  - Using these also for many signal models

Technical Hurdles

- The thing that has caused the most technical frustration is the use of `lhapdf-config` for configuration of the LHAPDF libraries to use in compilation
  - We relocate LHAPDF after installation to cvmfs (installed on afs), so this sometimes points back to afs
  - This also manifested itself differently between LO and NLO jobs
  - Our understanding from the LHAPDF authors is that this could be avoided on the MG5_aMC side – worth discussing with them?
- In the beginning, this caused some grid jobs to rely on afs connections back to CERN (**VERY SLOW**)

Our To Do List

- We are still working on advances in our setup

- Moving more searches to NLO generation, using MadSpin more

- Ability to save and re-shower LHE files directly on the grid

- Wider usage of FxFx and more exploration of UNLOPS

- Integration of a 'top up service' to our production system
  - Generate 100k events, shower with Pythia8, if we have not reached a sufficient number of events call back to MG5_aMC and generate more events
  - Would help with low filter efficiency samples, particularly to ensure job completion and reduce transient disk space use

Our Wish List

- There are some things that would make our lives a bit easier

- **Production-style workflow improvements**.
  - We don't need to generate web pages most of the time. Skipping the generation of webpages, diagram pictures, and so on, saves time and energy.

- **Configuration sanity checking**.
  - We were bitten recently by some inconsistent configurations that could have been caught (by us as well!)

- **Clear warnings when changing defaults and interfaces**.
  - That previous issue was due to a change in a default parameter in the run card that we had not noticed. Switching to process-driven defaults in the cards was great, but caused some teething problems as we adapted.

- **More thorough technical validation.**

- **More efficient integration with H7 and Py8**
  - Particularly as we explore VINCIA, DIRE, etc and showering systematics