A data driven search for non-resonant features in the dilepton spectra at ATLAS

Aaron White

University of Michigan

US LHC Users Meeting

October 26, 2018

Motivation

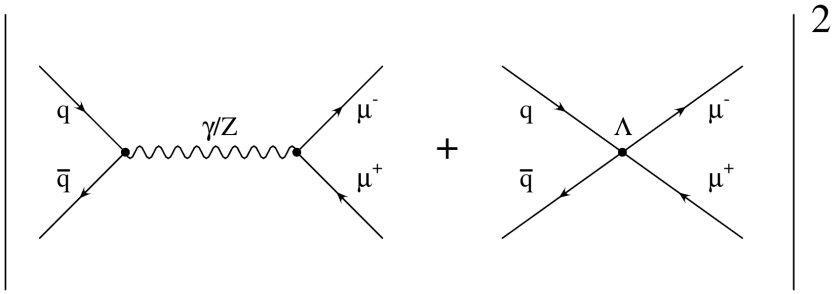
Object:

- Study the high mass ee, $\mu\mu$ invariant mass spectrum

Physics goals:

- Search generic **non-resonant** features:
  - contact interactions
  - lepton compositeness

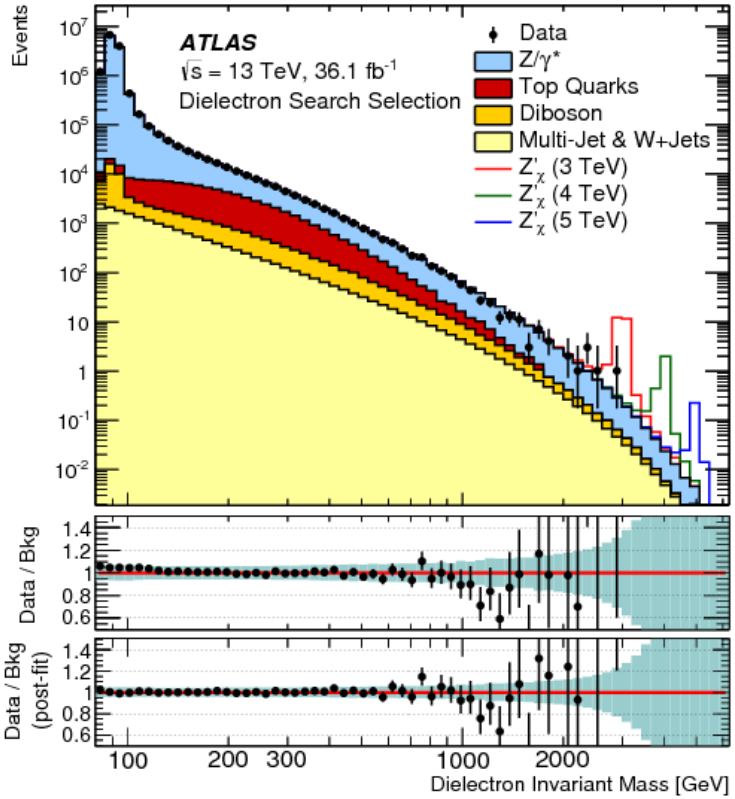
Contact Interaction



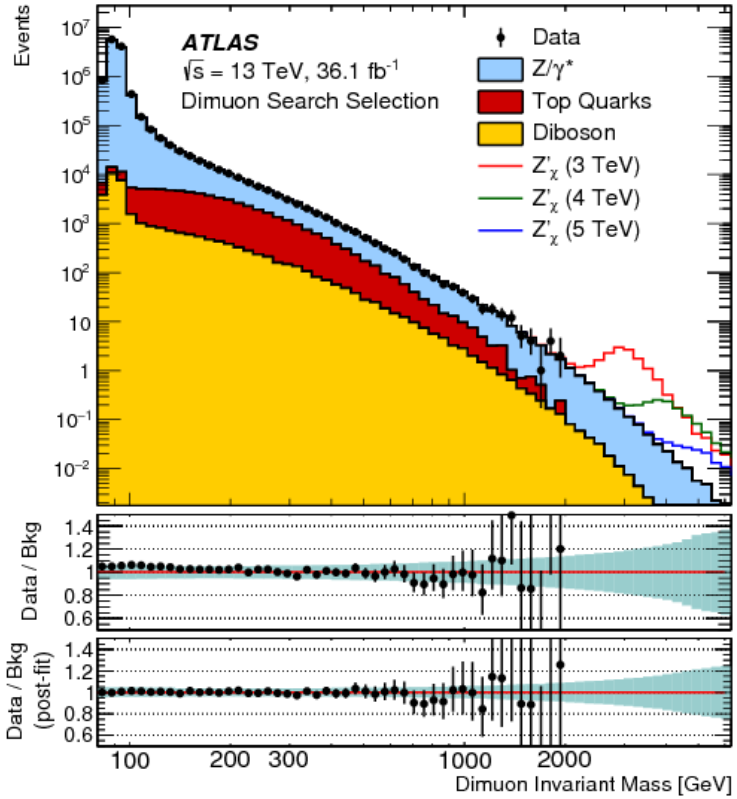
$$\sigma_{\text{tot}}(m_{ll}) = \sigma_{\text{DY}}(m_{ll}) \boxed{\pm \frac{F_I}{\Lambda^2} + \frac{F_C}{\Lambda^4}}$$

Previous Result

Dielectron



Dimuon

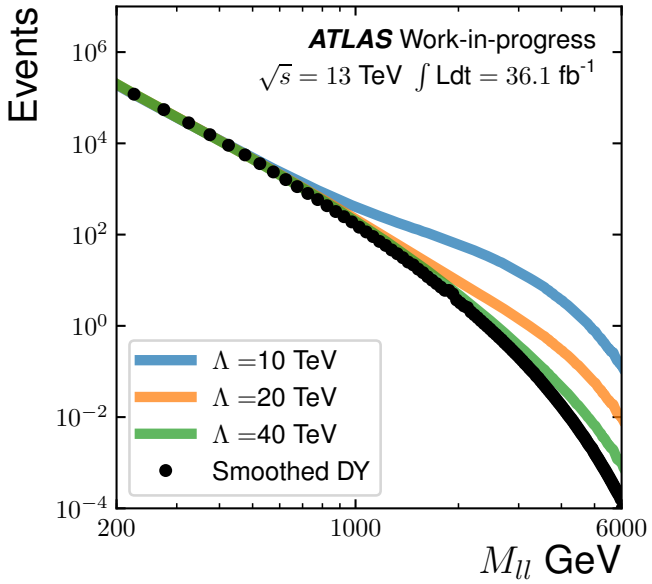


- Previous result: simulation (MC) driven background model
- Not enough simulated background to keep up with new luminosity
  $\rightarrow$ data-driven analysis strategy is needed

Previous result: arXiv:1707.02424

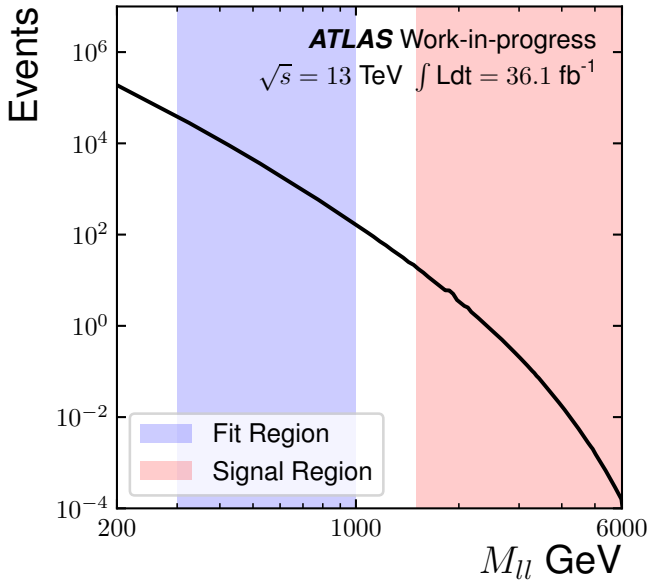
Data-driven approach

Challenge



- Non-resonant features are expected as broad excesses in the M_{ll} spectrum
- A fit to the full spectrum could swallow non-resonant features.

Strategy



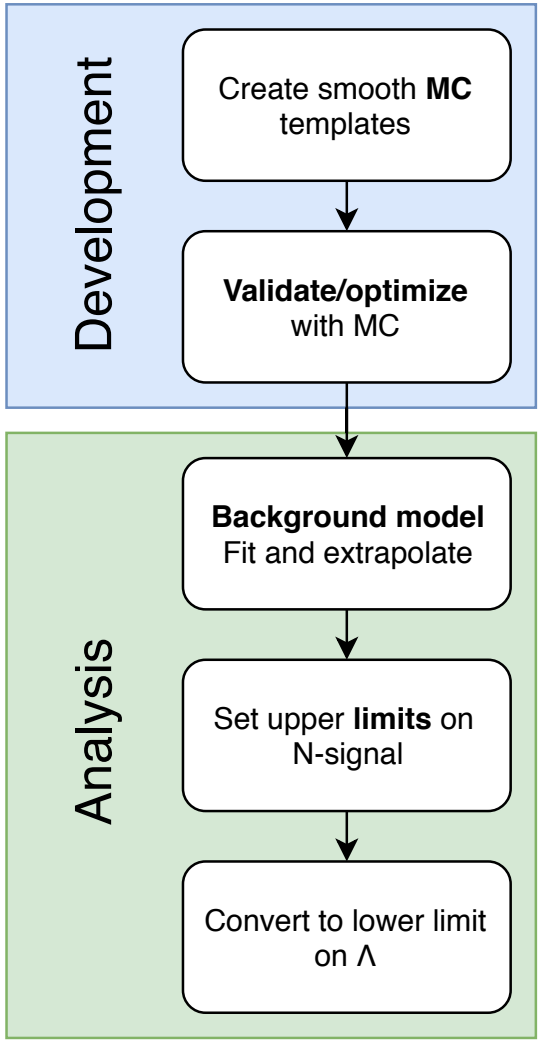
Fit in a low mass region, and extrapolate the fit to a high mass region

- Fit region constrain parameters of background estimation
- Signal region integrate fit and data yields as a single bin

Strategy

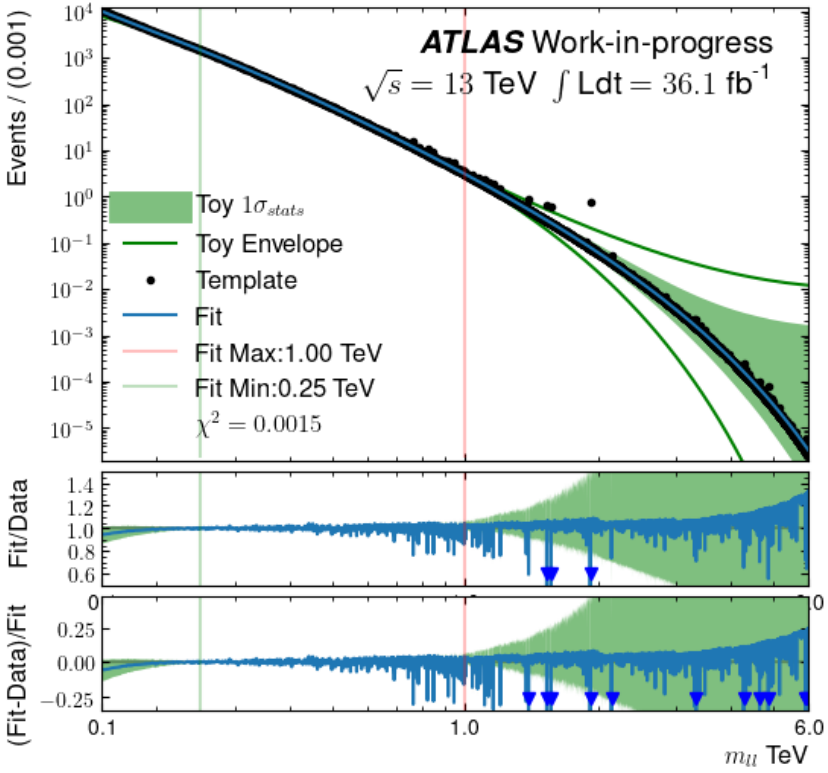
1. Simulation (MC) produced to understand background shape
2. Choices like mass ranges are optimized using MC
3. Fit the data and extrapolate to signal region
   - $f(x, \theta) = (1 - \sqrt{x})^b * \mathrm{pow}(x, \theta_0 + \theta_1 * \log(x) + \theta_2 * \log(x)^2 + \theta_3 * \log(x)^3$
   - x=$m_{ll,TeV}/13$
   - $b = 1/3$ for the dimuon fits
   - $b = 1$ for the dielectron fits
   - θ_i are nuisance parameters of the model
4. Integrate, set upper limits on signal events in the "signal region"
5. Convert signal event limits to model dependant lower limits on CI scale, Λ

Analysis Layout



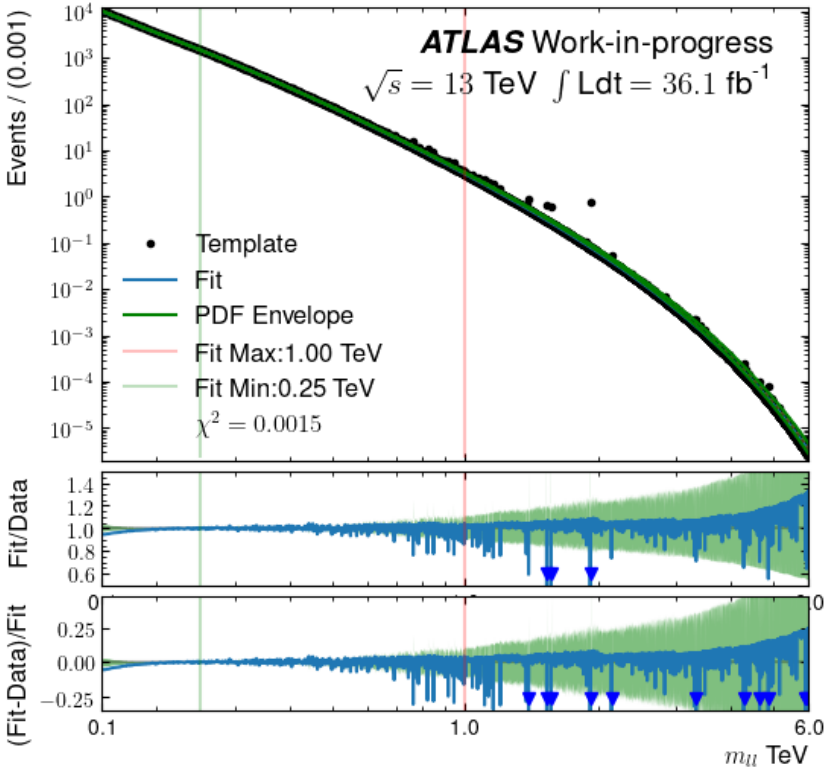
Uncertainties

Toy uncertainty band



- Understand impact of statistical fluctuations in the fit region on the background model
- Ensemble of 400 statistical toys is drawn from the MC
- These toys are fit, and the error band contains $\pm 1\sigma$ (calculated separately)

MC uncertainty band

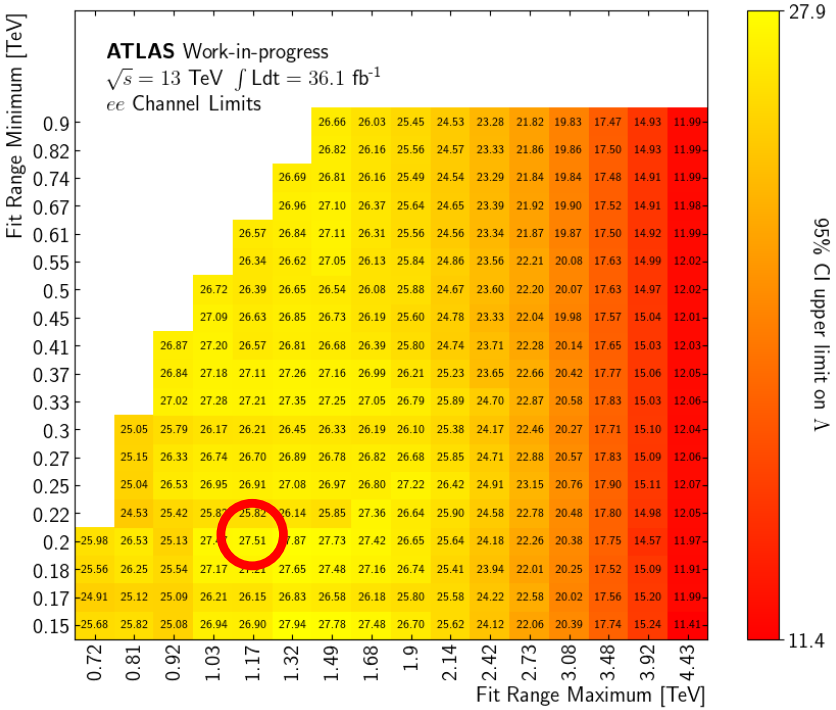


- Understand how adaptable our background model is to variations in the underlying data distribution
- Several systematics influence the background MC, these are fit
- These are fit, and their envelope is shown

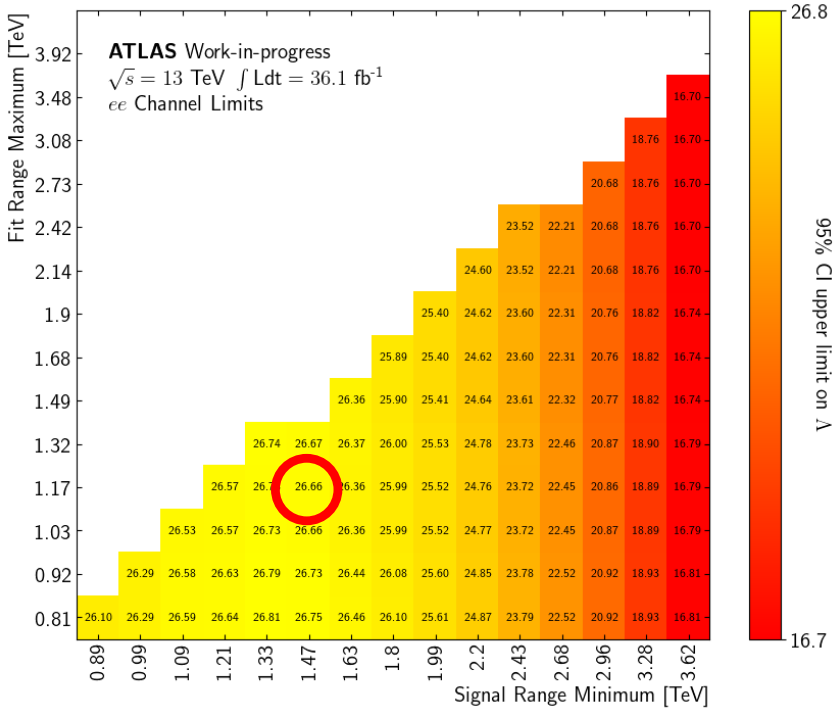
Optimization

The Fit Region and Signal Region need to be picked

Choose fit region:



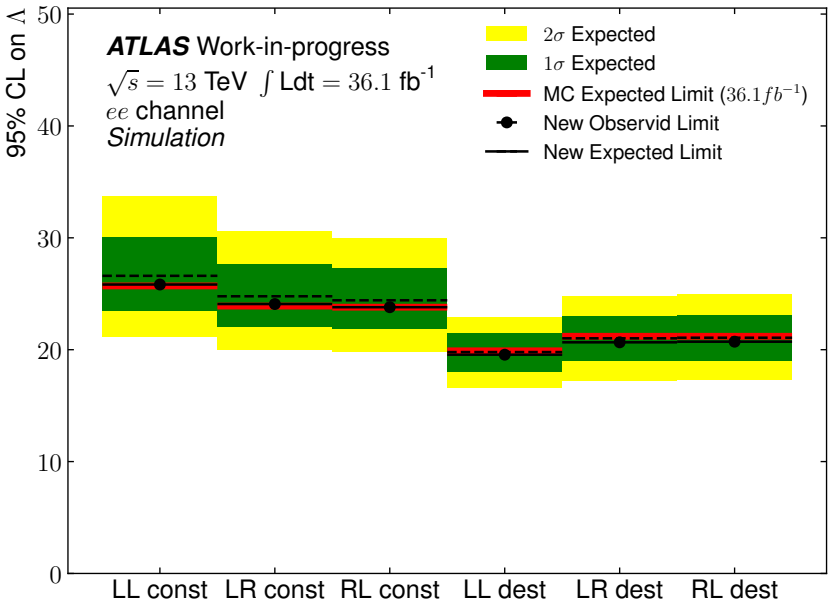
Choose signal region:



- Fit Region: influences the quality of the fit
- Signal region: influences signal yield
- Criteria for selection: maximum expected limit on Λ

Sensitivity Comparison

- Comparison between
  - MC driven limits
  - **data-driven** limits
- This shows preliminary agreement between the limits from the previous result, and the limits with the **new data-driven strategy**.
- The new limits exclude MC systematics.



"Observed" is from a MC template

Summary

Conclusion

- **Data-driven driven** limits are competitive with MC limits at luminosity of $36.1 fb^{-1}$
  - Expected to beat MC with more luminosity
- Further optimization for the destructive signal limits looks promising
- The analysis aims for a full Run-2 publication

Thanks to Deshan Abhayasinghe who also works on the non-resonant search.

Backup

Analysis Details

Analysis Channels

Dielectron

- $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$
- $p_{T,e} > 30$ GeV
- $M_{ee} > 130$ GeV
- Trigger: two e with p_T 12, 17, 24 GeV for years 2015, 16, 17
- Highest E_T dielectron pair selected

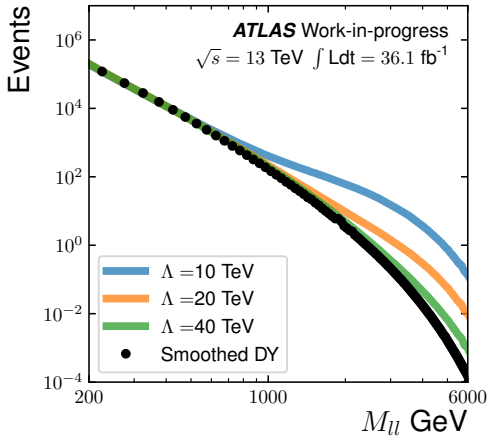
Dimuon

- Not qualifying for dielectron
- $|\eta| < 2.5$
- $p_{T,e} > 30$ GeV
- $M_{\mu\mu} > 130$ GeV
- Trigger: $p_T = 26$ GeV isolated μ or $p_T = 50$ GeV μ
- Highest p_T dimuon pair selected

Data/MC

- The analysis will use the full Run 2 dataset
- MC is used as a guide for the development of the analysis strategy:
  - Powheg+Pythia8 $Z \to ll$ main background
  - Powheg+Pythia8 $t\bar{t} \to ll$, single t secondary background
  - Truth-only DY MC and transfer functions
  - Pythia8 DY reweighted to mass Λ signal

Signal Model



Object selection/MC

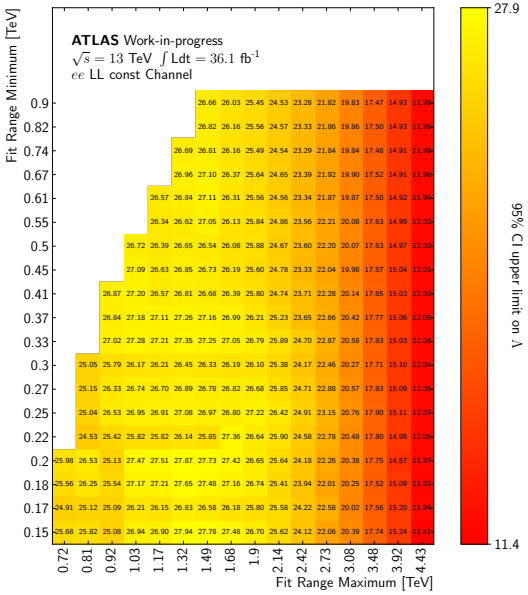
Muon selection	
Criteria	Value
Selection Working Point	High-pT
Isolation Working Point	Loose Track Only
Momentum Calibration	Sagitta Correction Not Used
p_T Cut	$p_T > 30$ GeV
η Cut	$\lvert\eta\rvert < 2.5$
d0 Significance Cut	$3\,\sigma$
z0 Cut	0.5 mm

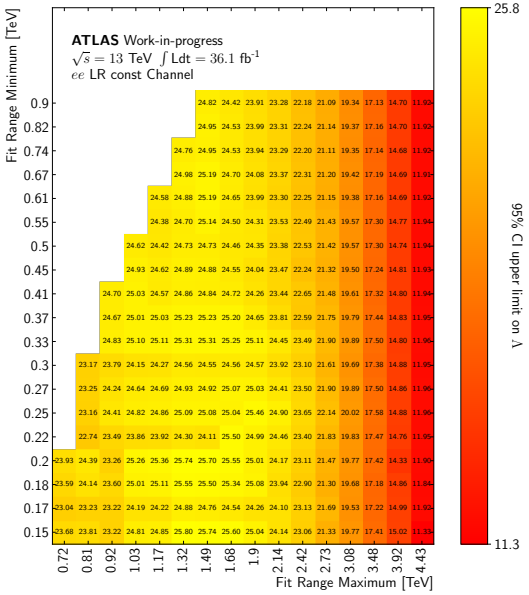
Electron selection	
Feature	*Criteria*
Pseudorapidity range	$(\lvert\eta\rvert < 1.37)$ || $(1.52 < \lvert\eta\rvert < 2.47)$
Energy calibration	es2017_R21_v0 (ESModel)
Transverse momentum	$p_T > 30$ GeV
Object quality	Not from a bad calorimeter cluster (BADCLUSELECTRON)
	Remove clusters from regions with EMEC bad HV (2016 data only)
Track to vertex association	$\lvert d_0^{BL}(\sigma)\rvert < 5$
	$\lvert\Delta z_0^{BL} \sin\theta\rvert < 0.5$ mm
Identification	LooseAndBLayerLLH (QCD fakes), MediumLLH (main selection)
Isolation	Loose

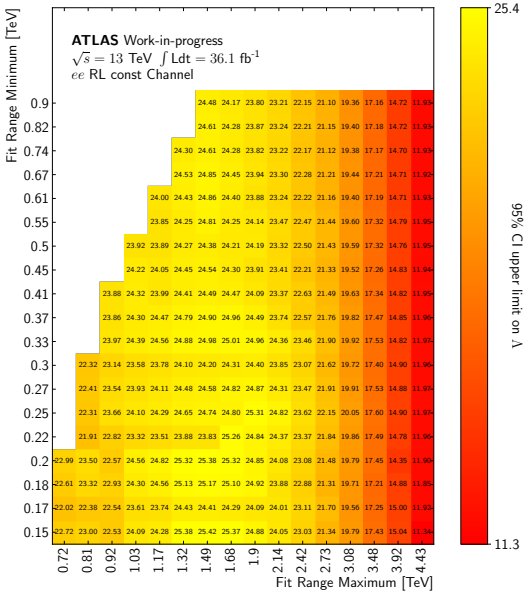
MC used

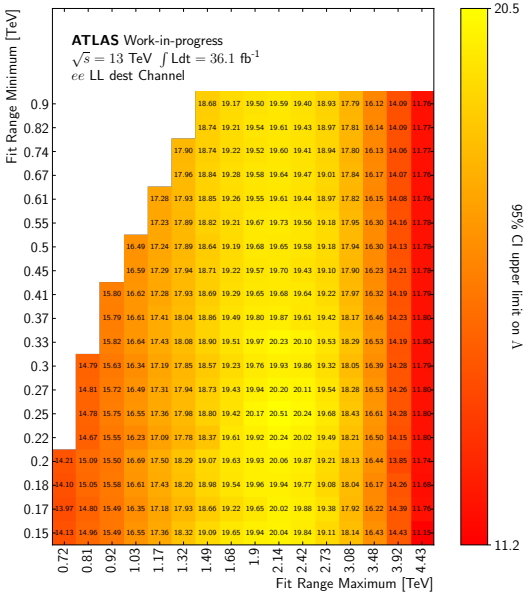
- Powheg+Pythia8 $Z \to ll$ main background
- Powheg+Pythia8 $t\bar{t} \to ll$, single t secondary background
- Pythia8 DY reweighted to mass Λ signal

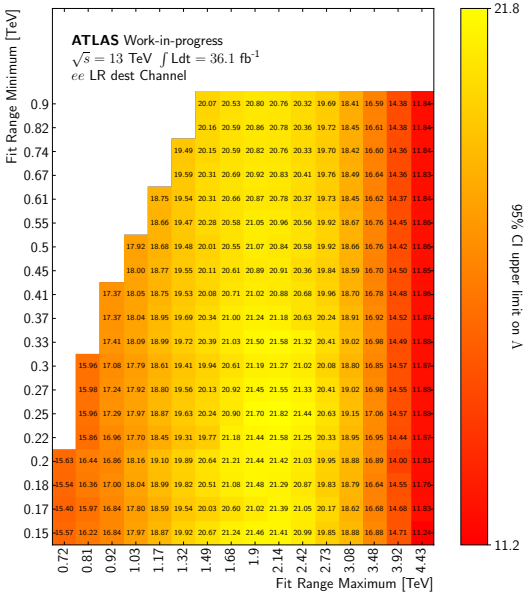
Optimizations

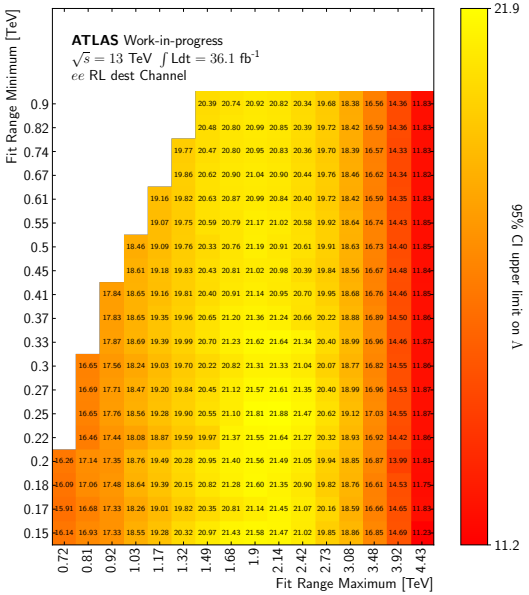




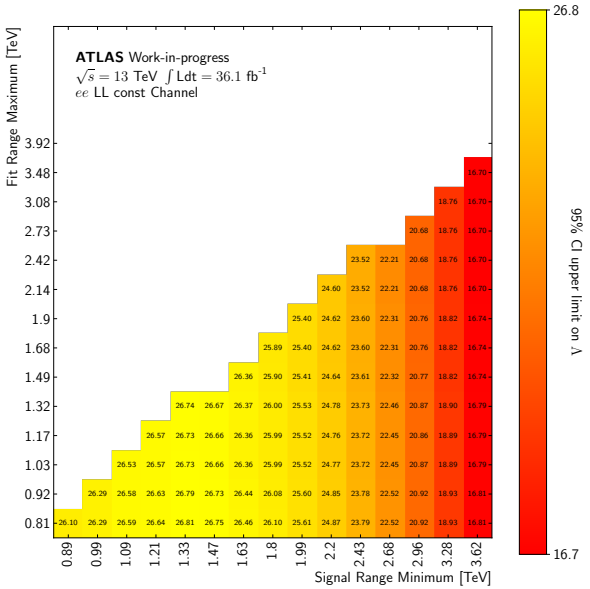


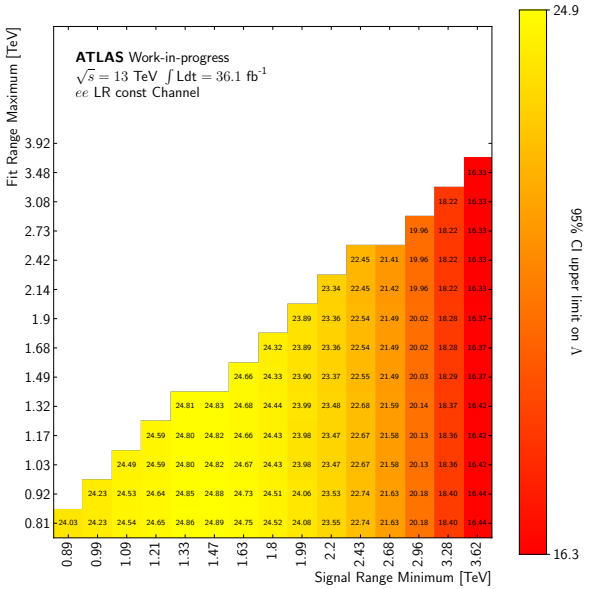


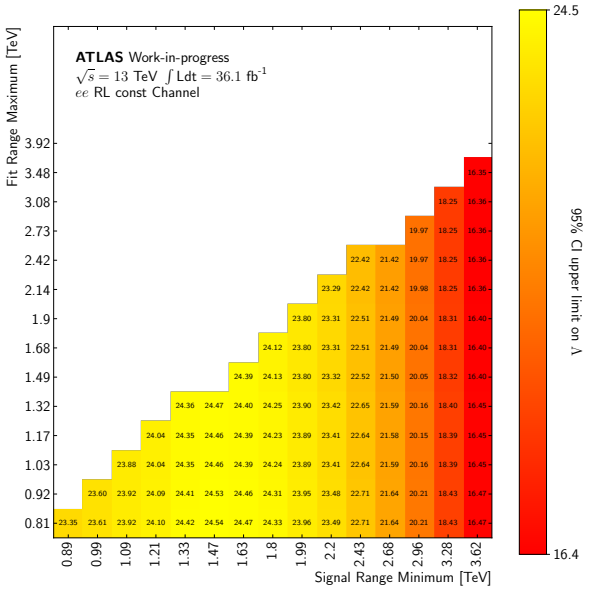


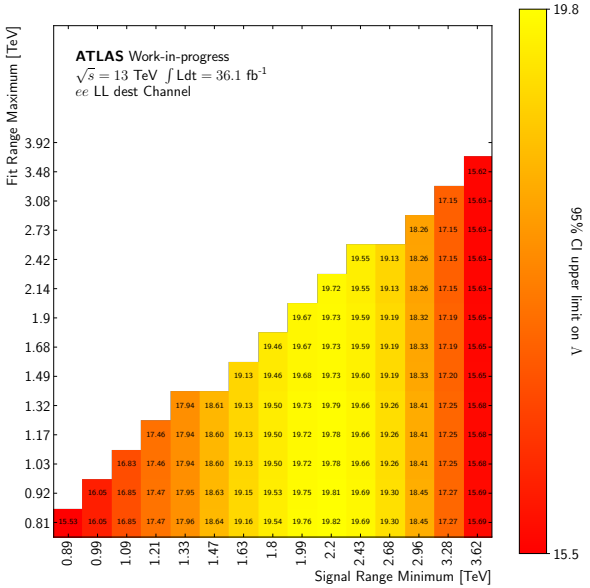


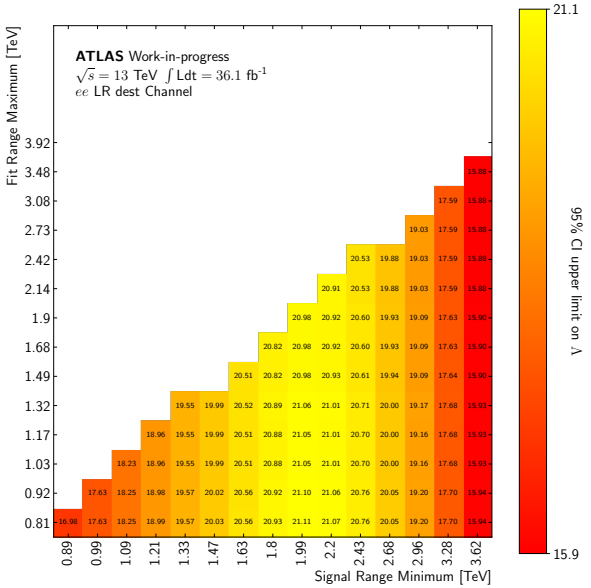
Optimizations

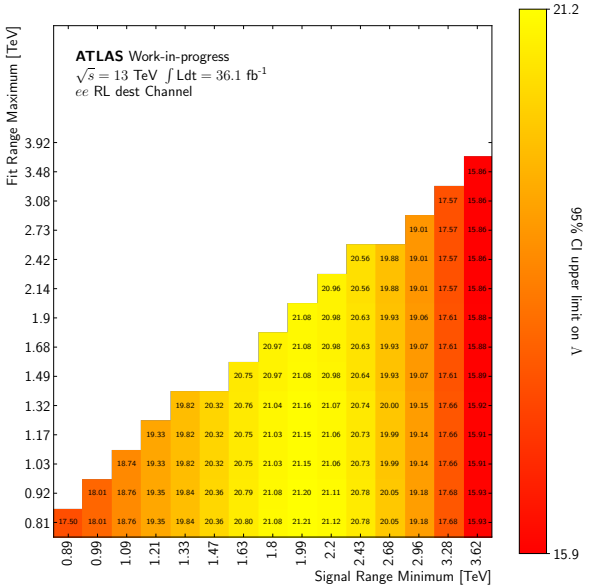




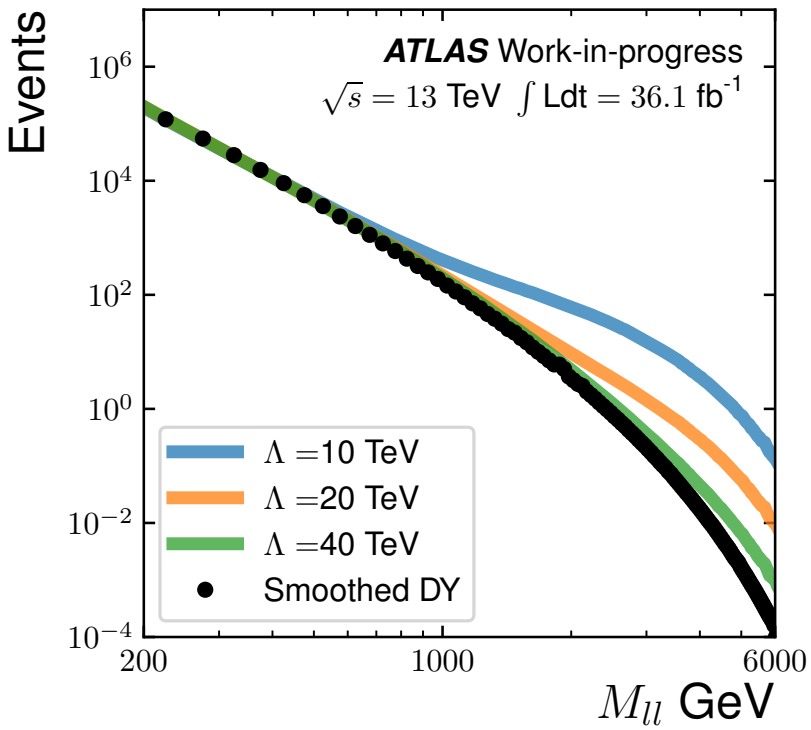








Signal Models

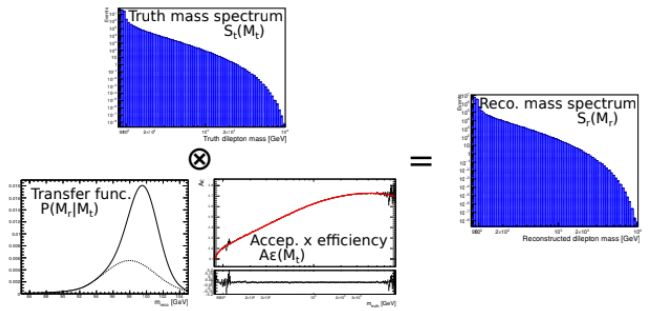


Lepton channel	Interference	Chirality
ee	constructive	LL, RL, LR, RR
ee	destructive	LL, RL, LR, RR
$\mu\mu$	constructive	LL, RL, LR, RR
$\mu\mu$	destructive	LL, RL, LR, RR

Background Model

Smoothed MC

- While the background model will be data driven, MC is used to optimize the analysis strategy
- Not enough fully reconstructed MC events
- Truth MC events used instead
- The truth MC is smeared is smoothed using transfer functions trained on fully reconstructed MC



PDF Model

- The PDF is fit to data, and then extrapolated to a "signal region"
- The selection of the function was optimized with the smoothed MCs
- PDF model:

  $f(x,\theta) =$
  $(1-\sqrt{x})^b * \mathrm{pow}(x, \theta_0 + \theta_1 * \log(x) + \theta_2 * \log(x)^2 + \theta_3 * \log(x)^3$

  Where
  - x=$m_{ll,TeV}/13$
  - $b = 1/3$ for the dimuon fits
  - $b = 1$ for the dielectron fits
  - θ_i are nuisance parameters of the model