Background modeling and signal estimation using Gaussian Processes in the H→ $\gamma\gamma$ channel

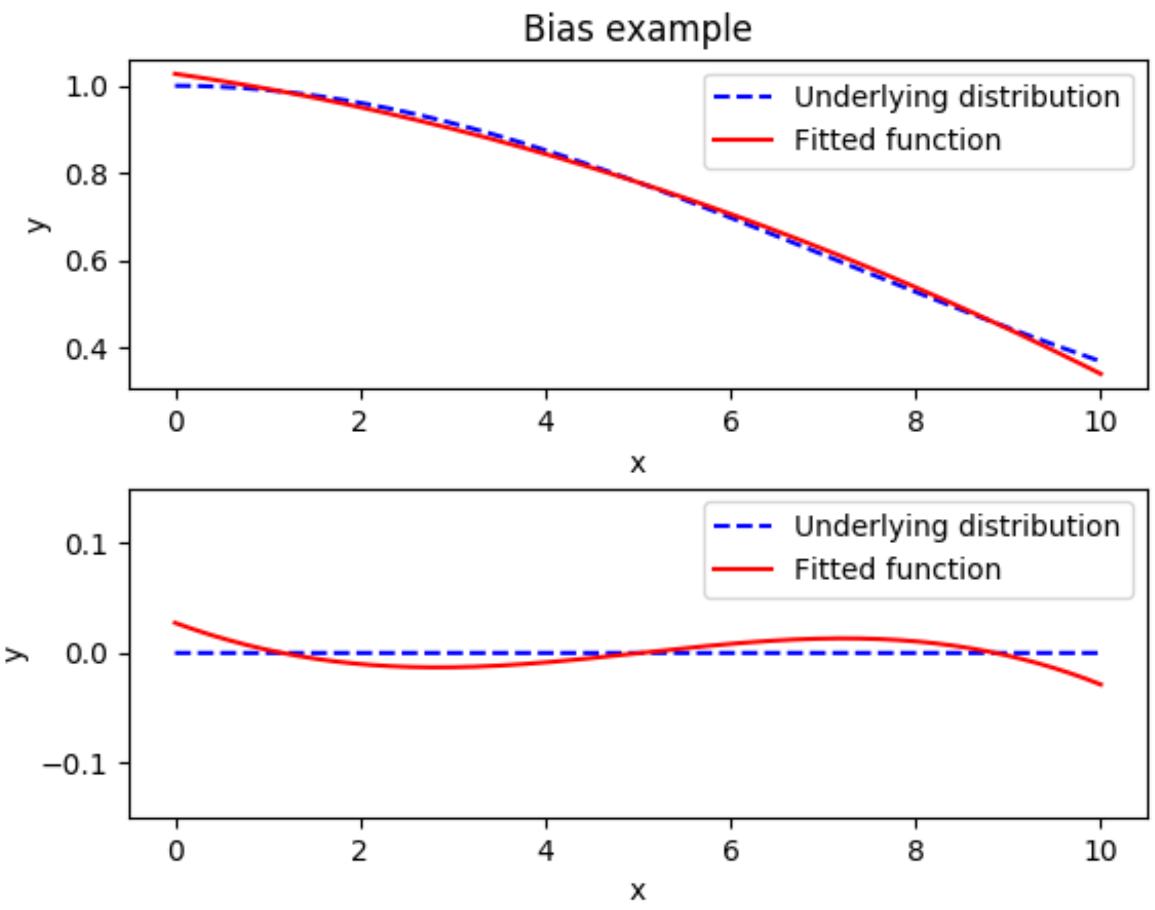
Simon Millerjord

Master student, UiO

Nordic high energy physics conference, Spåtind, 06.01.20

Motivation

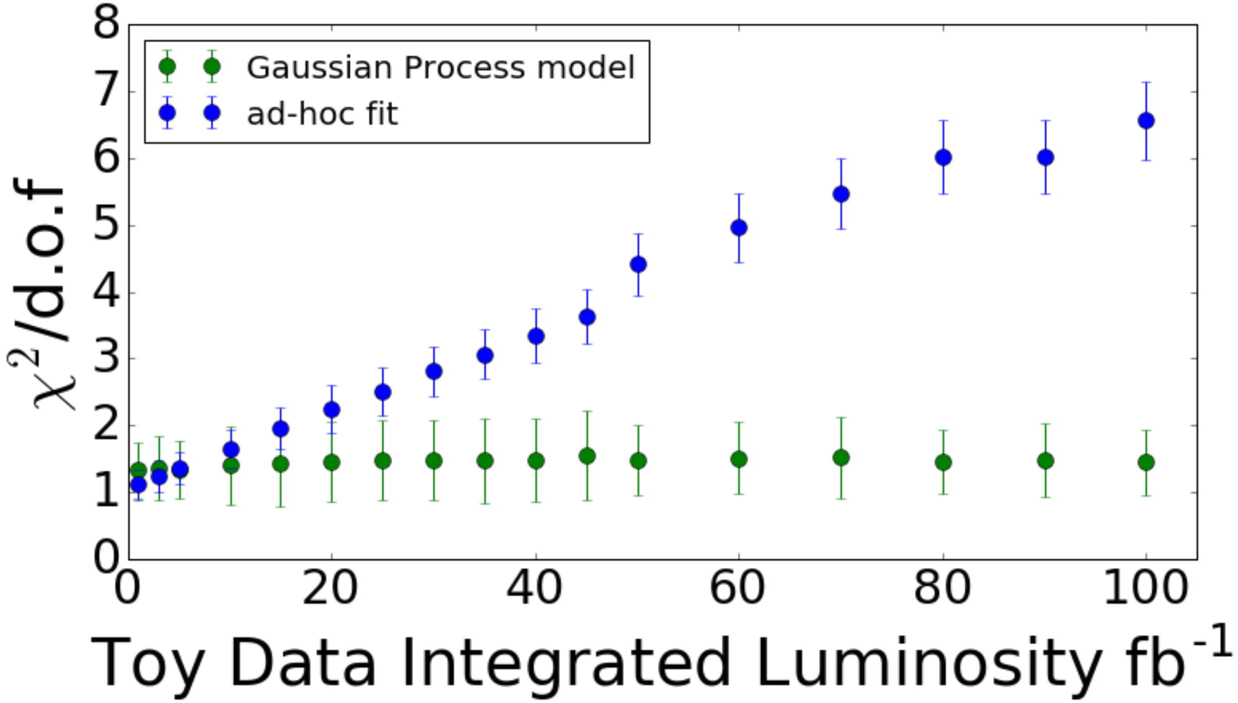
- H → yy channel
  - Parametric functions
- Background modeling
  - Exponential
  - Exponential of polynomials
  - Bernstein polynomials
- Increase of data to 3000 fb^{-1} by 2037
- Relative statistical uncertainty decreases – hidden features may arise



Motivation

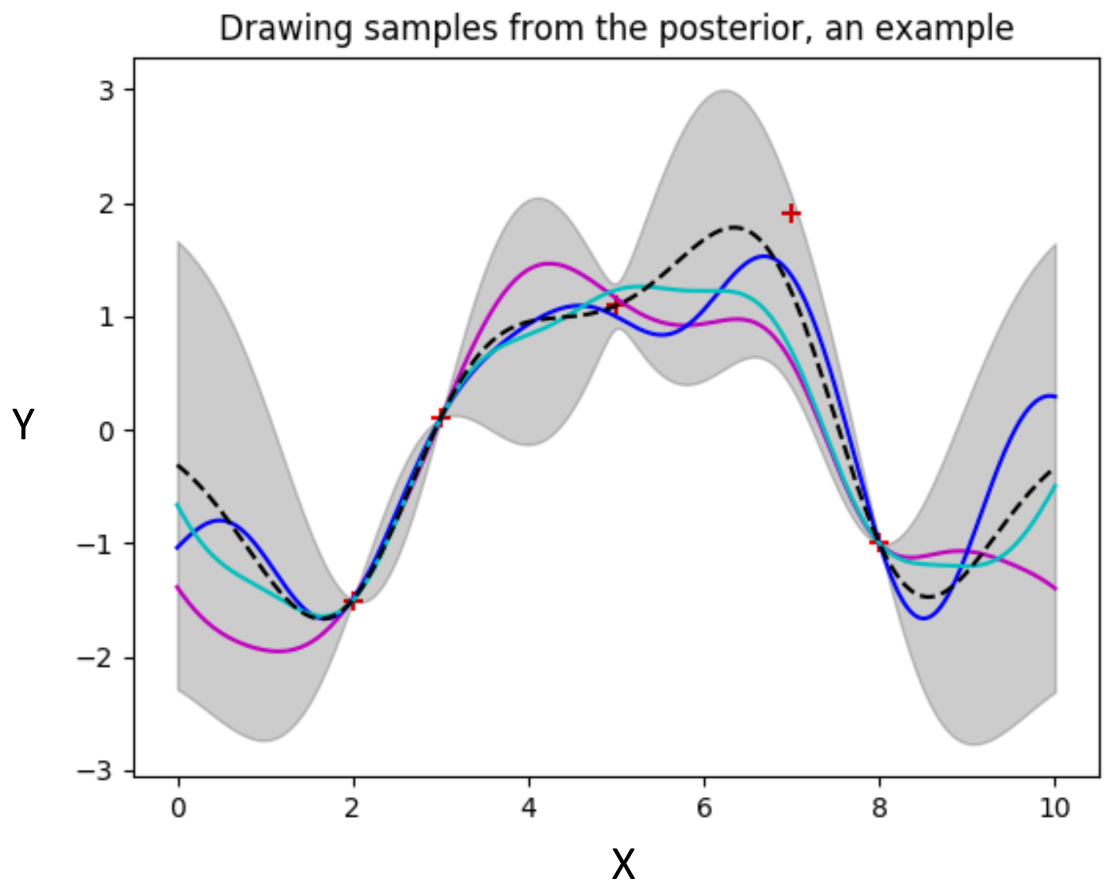
- Article by Frate et. al.
  - arxiv:1709.05681
  - Explore Gaussian Processes
    - Background modeling
    - Generic signal modeling
  - ttbar ATLAS data set with 3.6 fb^{-1}

- Luminosity independence
- What is a good model?
  - Background model modeling without bias
  - Background model does not swallow signal
  - Signal model modeling without bias
  - Minimizing the total uncertainty on signal amplitude



Gaussian Processes (GP)

- Bayesian machine learning method
  - Mean function
  - Covariance matrix – kernel
    - Squared exponential: $Ae^{-\frac{(x_1 - x_2)^2}{l^2}}$
- Non-parametric samples
- Updating the prior information



Gaussian Processes (GP)

Methods

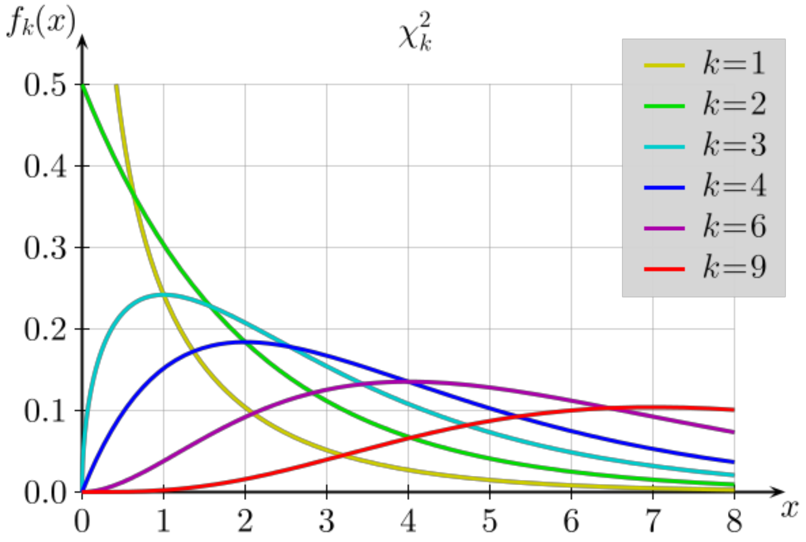
- Python
  - GEORGE
  - IMINUIT
- Optimization of hyperparameters
  - Minimization of the negative log marginal likelihood

$$-\ln L = \frac{1}{2} y^T K_y^{-1} y + \frac{1}{2} log\left|K_y\right| + \frac{n}{2} \log 2\pi$$

- Background kernel – Squared exponential
  - $Ae^{-\frac{(x_1 - x_2)^2}{l^2}}$
- Signal kernel – Local Gaussian kernel
  - $Ae^{-\frac{(x_1 - x_0)^2 + (x_2 - x_0)^2}{2w}}$

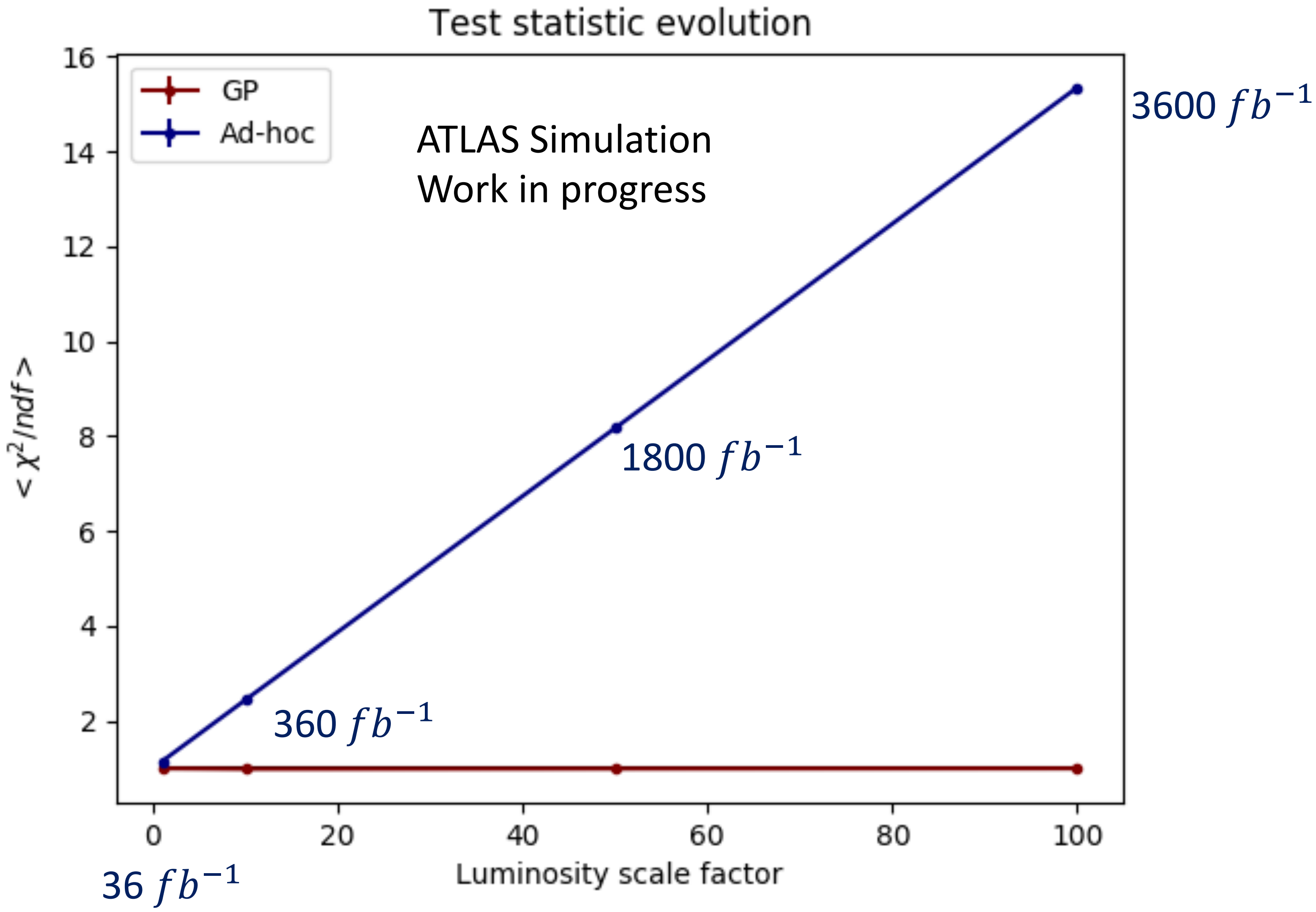
Analysis

- 20 000 toys
- 36 $\boldsymbol{fb^{-1}}$, 360 $\boldsymbol{fb^{-1}}$, 1800 $\boldsymbol{fb^{-1}}$, 3600 $\boldsymbol{fb^{-1}}$
- Parametric (ad-hoc) function for comparison: e^{ax^2+bx+c}
- Goodness of fit: $\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i;\theta))^2}{\sigma_i^2}$
- Degrees of freedom for background kernel:
  - Assumed two: Amplitude and lengthscale
  - *Effective* degrees of freedom: ~5
- Number of degrees of freedom for χ^2 distribution:
  - Ndof = 40 – 2 = 38
  - Actual ndof = 35



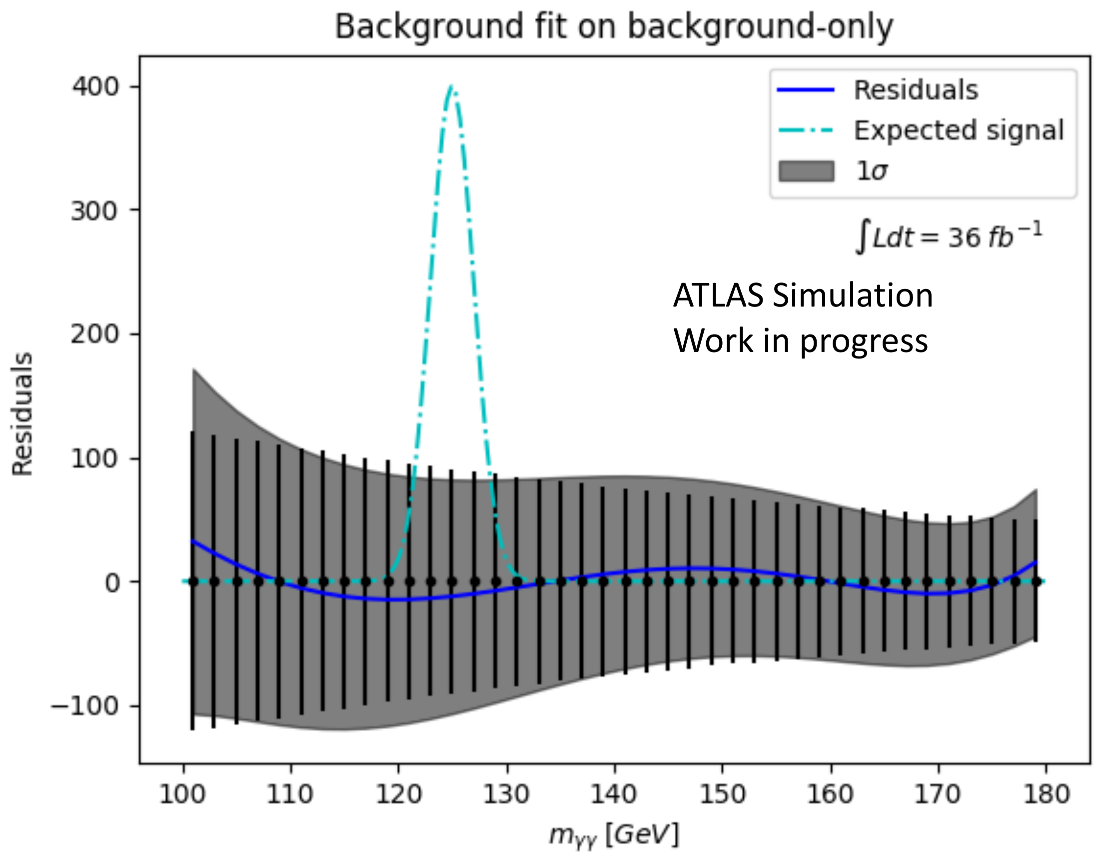
Results

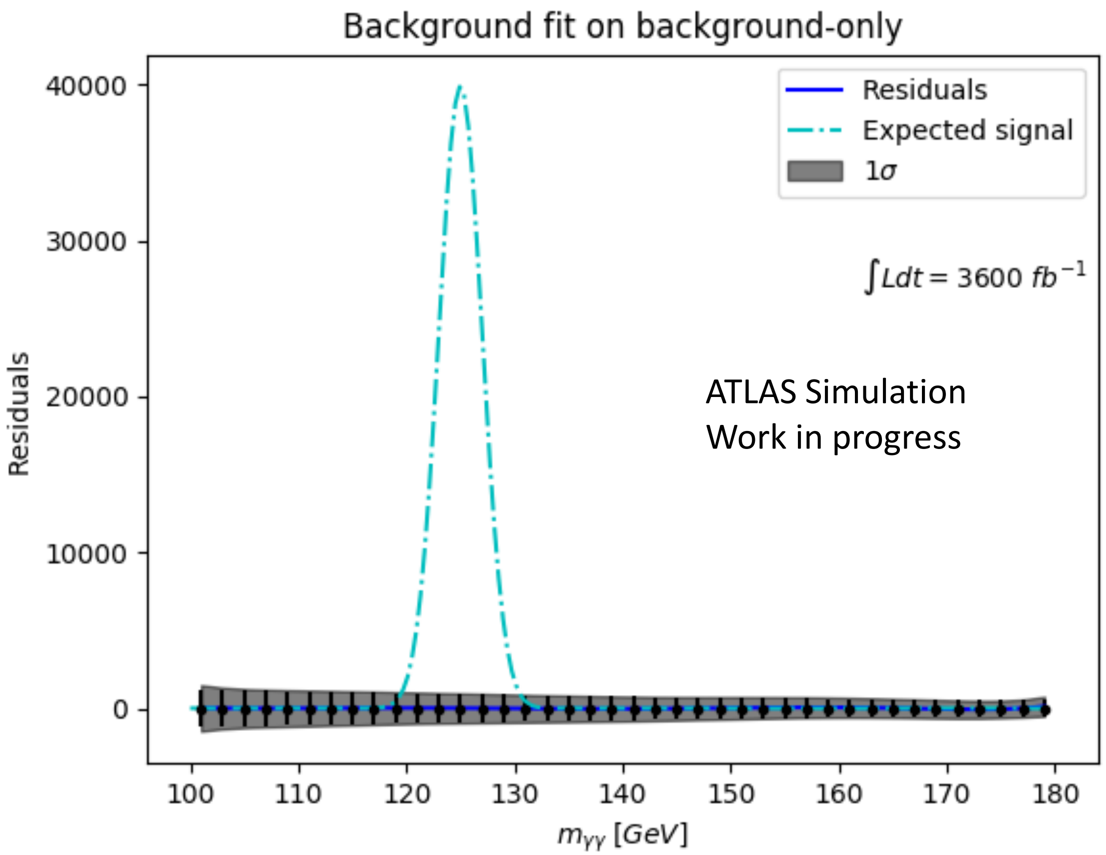
Mean test statistic evolution



Results

Residuals – Background only kernel

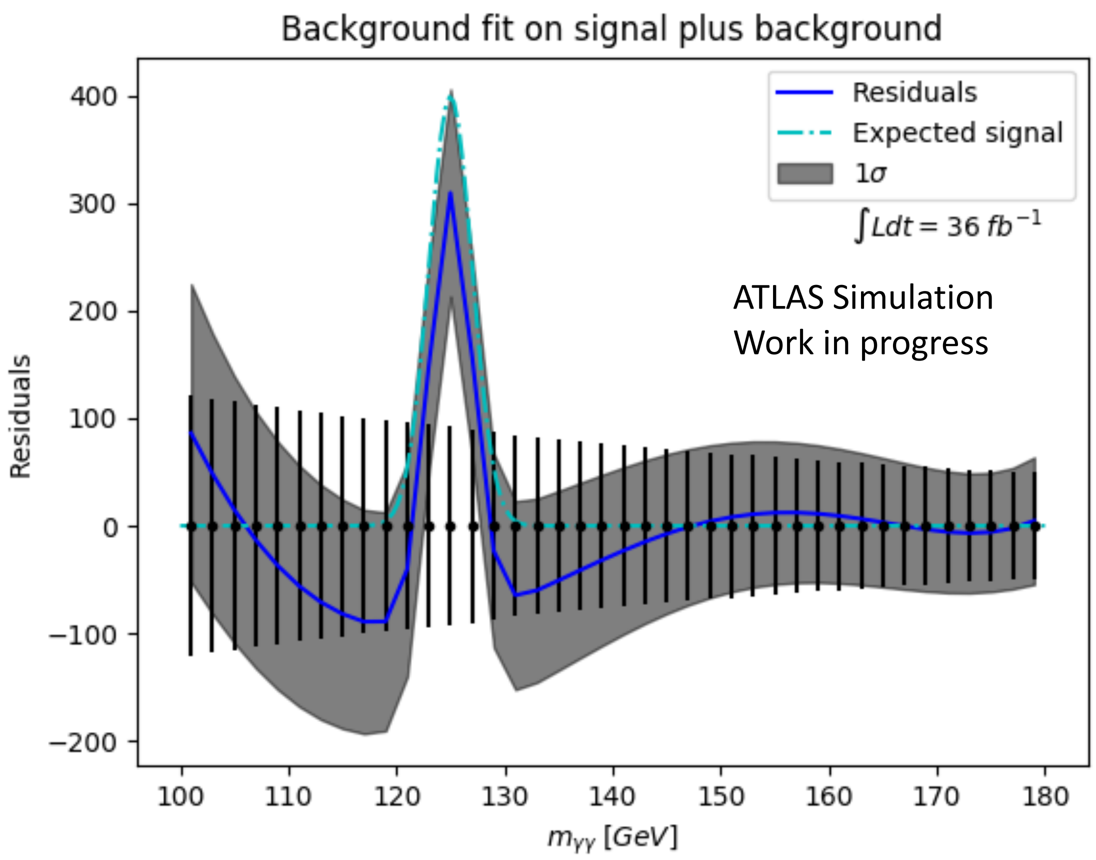


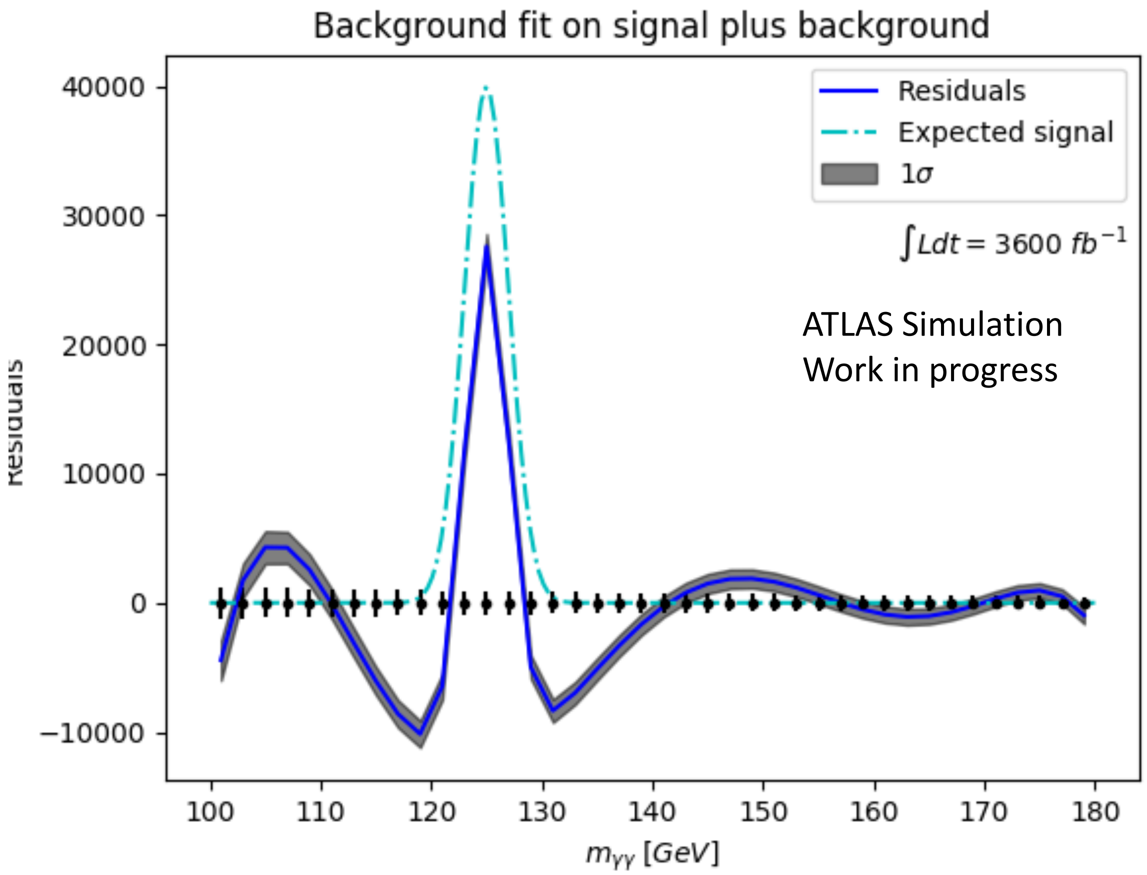


These residuals are the difference between the Asimov data set and the predicted function.

Results

Signal injection test



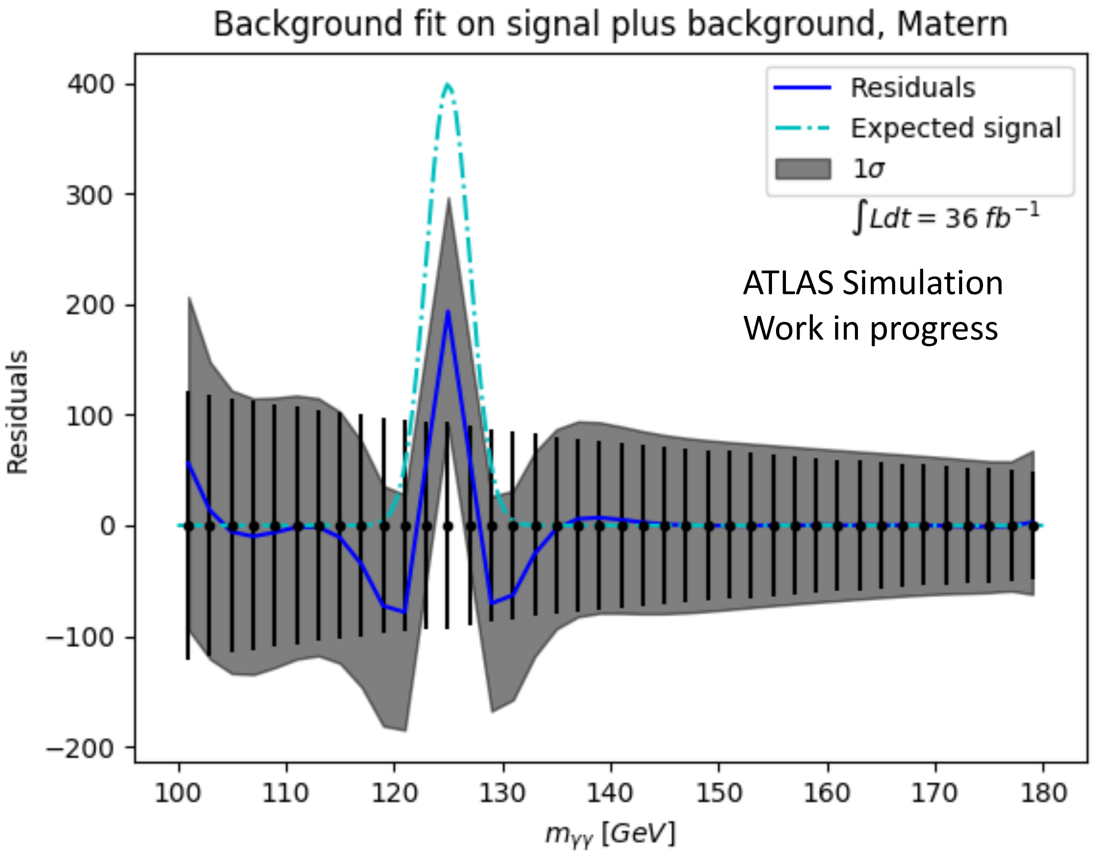


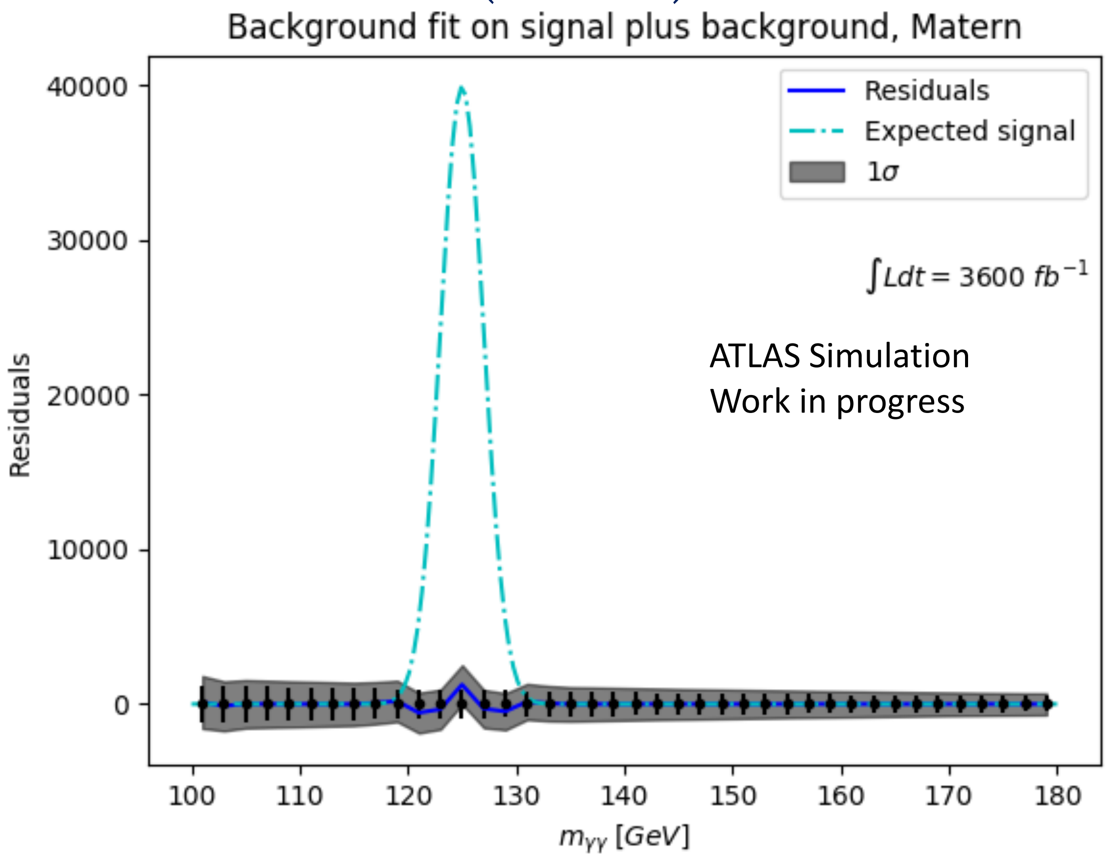
These residuals are the difference between the Asimov data set and the predicted function.

Results

Signal injection test using a Matérn kernel

$$k_{3/2} = \left(1 + \frac{\sqrt{3}r}{l}\right) e^{\left(-\frac{\sqrt{3}r}{l}\right)}$$

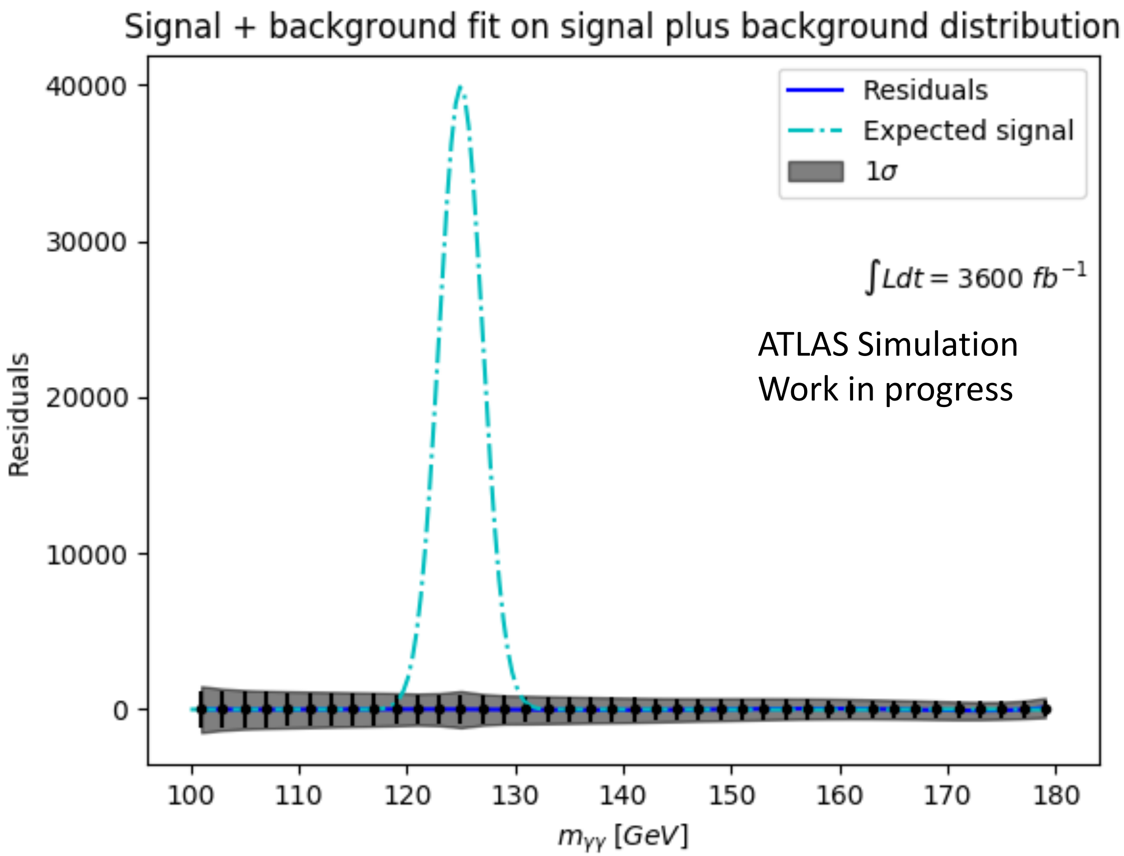




These residuals are the difference between the Asimov data set and the predicted function.

Results

Signal amplitude estimation



Luminosity $[fb^{-1}]$	Expected signal amplitude	Estimated signal amplitude	Uncertainty [%]
36	398	373 ± 271	72
360	3989	3955 ± 2778	70
1800	19947	19927 ± 14096	70
3600	39894	39901 ± 28218	70

These residuals are the difference between the Asimov data set and the predicted function.

Conclusions

The Gaussian Process:

- Using the hyperparameters for estimating the signal amplitude was unsuccessful due to the large uncertainties.
  - A different method for estimating the signal is required
- Manages to find the underlying function and is luminosity independent.
- Does not produce a large bias relative to the expected Higgs boson signal
- Does not swallow a inserted signal.