Gaussian Processes for Model Independent Resonance Searches

Fabricio Jiménez
Machine Learning @ Cézeaux

17th of May 2018

Outline

- Gaussian process for MI Resonance Searches
- What are Gaussian Processes
- Modeling smooth bkgs and generic localized sigs using GPs 1709.05681
- Preliminary work on using the approach in MI searches:
  - Details on the GPs used
  - Preliminary tests fitting 2-jet mass spectrum
  - Injection of signal
- Future work and concluding remarks

Note: most of what's shown here is based on a work from Meghan Frate et al., published in 1709.05681, with codes here and here.

Gaussian Processes for Ml Resonance Searches

Main idea is to use Gaussian Processes (GPs) to provide a smooth model of background and signal invariant mass (and other) distributions.

Among the advantages of Gaussian Processes:

- “Non-parametric” approach
- Not dependent on a completely ad-hoc functional form as it’s the case in many analyses
- Correlation is modeled using a “kernel” function that can include physics insights
- Can be used as a complementary approach in many of the GS signatures

Gaussian Processes (GPs)

GP: associate a multivariate gaussian distribution to a set of random variables

→ The gaussian will have as many dimensions as random variables we have

A set of N values (bin counts) **y** can be associated with

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \sim \mathrm{Gaus}(\boldsymbol{\mu}, \Sigma)$$

Infer new values y_* by extending (the dim. of) the Gaussian distribution

Use a *kernel* or *measure of similarity* between points (bin centers) and a *mean function*

→ One kernel example is the exponential squared kernel:

$$k(x_i, x_j) = A \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right)$$

where A and l are (hyper)parameters to be fixed

Gaussian Processes (GPs)

Infer a new value y_* located in x_*, using the following

$$p(\boldsymbol{y}_*|\boldsymbol{x}_*, \boldsymbol{x}, \boldsymbol{y}) = \text{Gaus}(\boldsymbol{y}_*|\boldsymbol{\mu}_*, \Sigma_*)$$

$$\boldsymbol{\mu}_* = m(\boldsymbol{x}_*) + \boldsymbol{K}_*^T \Sigma^{-1}(\boldsymbol{y} - m(\boldsymbol{x}))$$

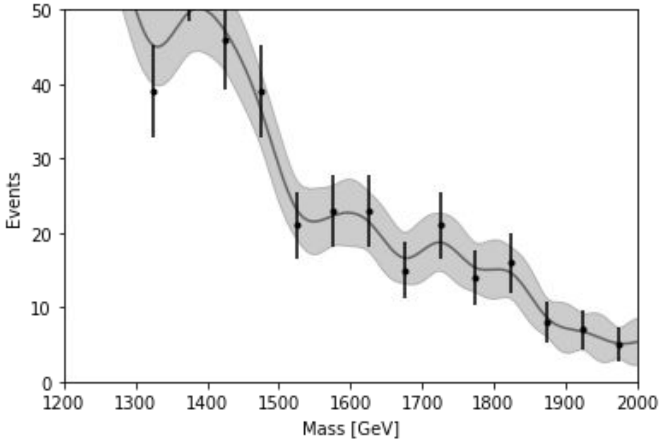
$$\Sigma_* = \boldsymbol{K}_{**} - \boldsymbol{K}_*^T \Sigma^{-1} \boldsymbol{K}_*$$

Where K is constructed from:

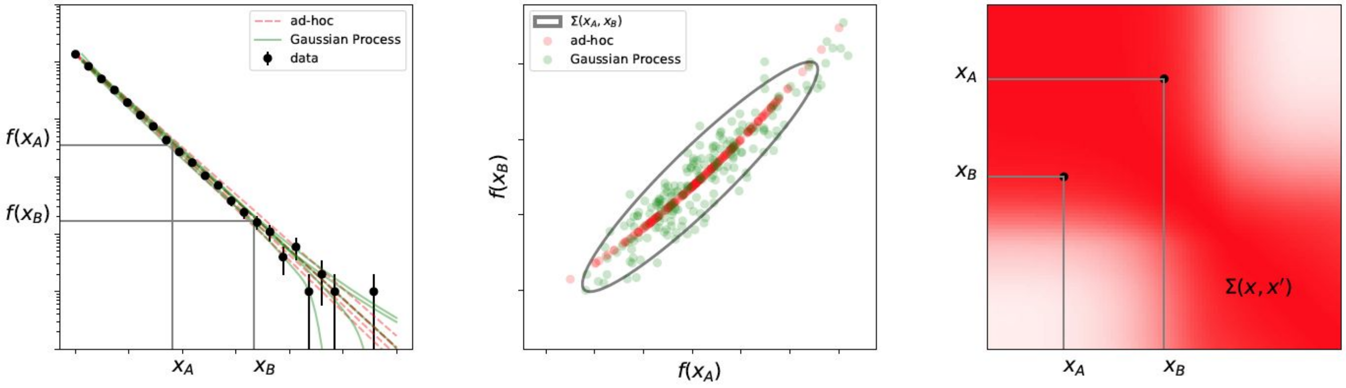
$$\boldsymbol{K}_* = k(\boldsymbol{x}, \boldsymbol{x}_*), \qquad \boldsymbol{K}_{**} = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$$

Note:

- The hyperparameters of the kernel are optimized using e.g. Maximum Likelihood
- In general, GPs are flexible enough to model the mean of the distribution having m(x) = 0
- Can be used over pdfs



Gaussian Processes (GPs)



[1709.05681](https://arxiv.org/abs/1709.05681)

Modeling smooth backgrounds and generic localized signals using GPs

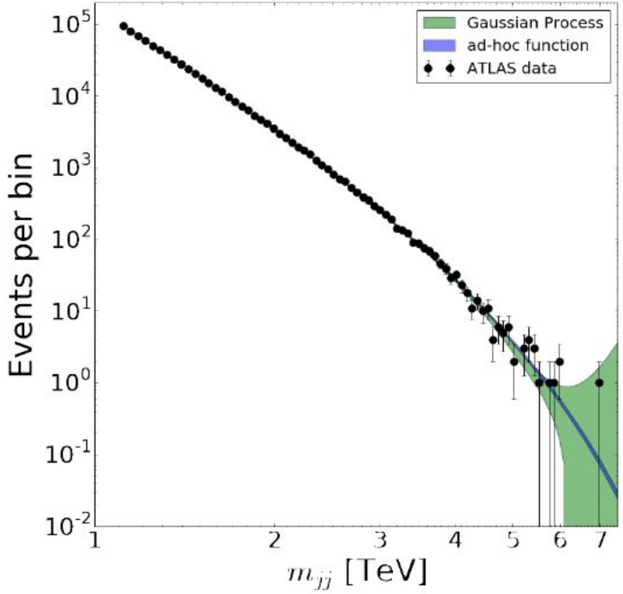
Use the dijet spectrum data as an alternative to the historic “parameterized” approach:

$$f(x|\theta) = \theta_0(1-x)^{\theta_1}x^{\theta_2}x^{\theta_3 \log(x)}$$

Use a kernel that contains physics input: describe properly sources of uncertainty from e.g. JES and PDFs

$$k(x_i, x_j) = Ae^{\frac{d-(x_i+x_j)}{2a}}\sqrt{\frac{2l(x_i)l(x_j)}{l(x_i)^2+l(x_j)^2}}e^{\frac{-(x_i-x_j)^2}{l(x_i)^2+l(x_j)^2}}$$

Use the parameterized fit as input for mean prediction



1709.05681

Modeling smooth backgrounds and generic localized signals using GPs

- In a background-only scenario: fit is better than the one in the parameterized approach
- S+B tests: combine a background GP and fit a signal model shape. Test B and S+B hypotheses
- Use a GP in the background-only data (could be MC) and then combine that bkg kernel (with its parameters already fixed) with a signal kernel, e.g.

$$k_s(x_i, x_j) = Ae^{-\frac{1}{2}(x_i - x_j)^2/l^2} e^{-\frac{1}{2}((x_i - m)^2 + (x_j - m)^2)/t^2}$$

→ Successfully extract signal bumps of different shapes on top of the background

1709.05681

Procedure

1. Take binned dataset (MC or real data) and perform a GP fit using ML
2. From the dataset, smear and add Poisson noise to each of the bins.
3. Inject some (e.g. Gaussian) signal to that resulting dataset
4. Keep bkg fit parameters frozen and perform GP fit using bkg+sig kernels to extract signal params

At the end we will have: fits for bkg and bkg+sig plus a dataset with injected signal

Preliminary work

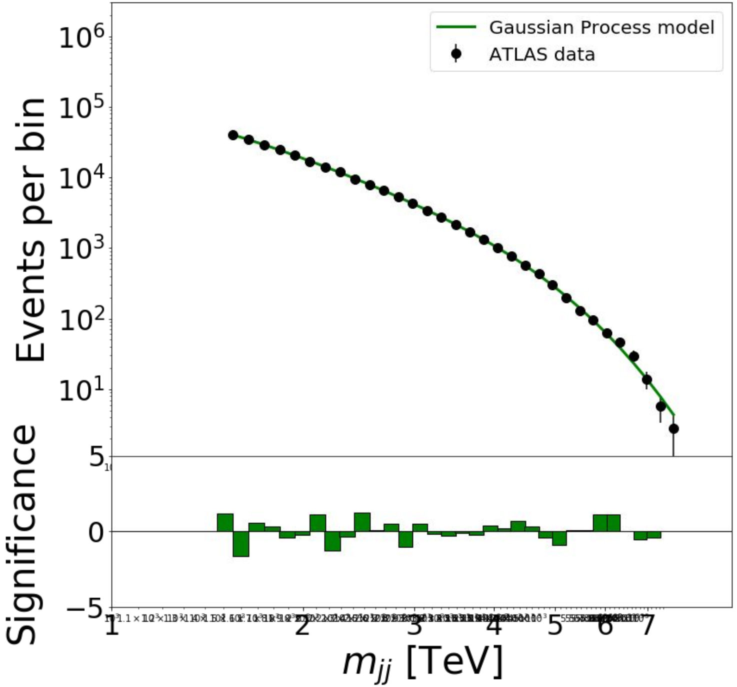
We propose to extend this to a larger set of signatures as a Model-Independent approach

Directions:

- Use an extended set of signatures (from GS)
- Get rid of the mean function

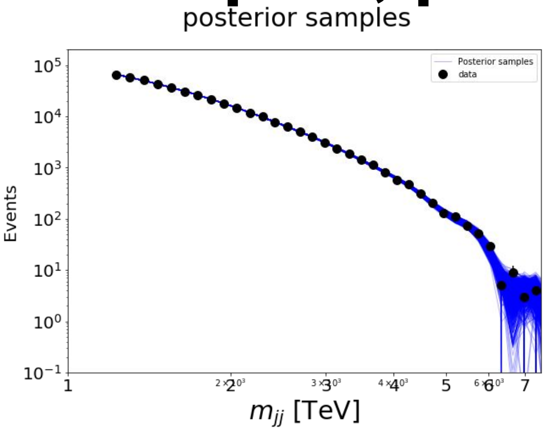
For the (background only) GPs I used:

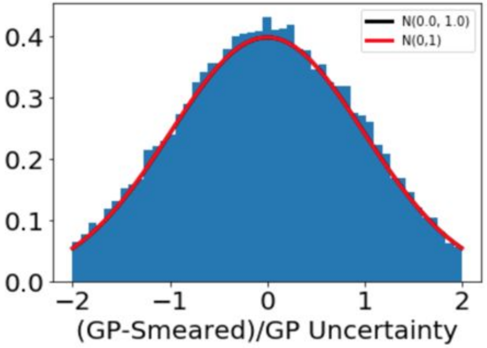
- Exponential squared kernel, “custom” kernel
- No input for the mean function (assume zero vector)



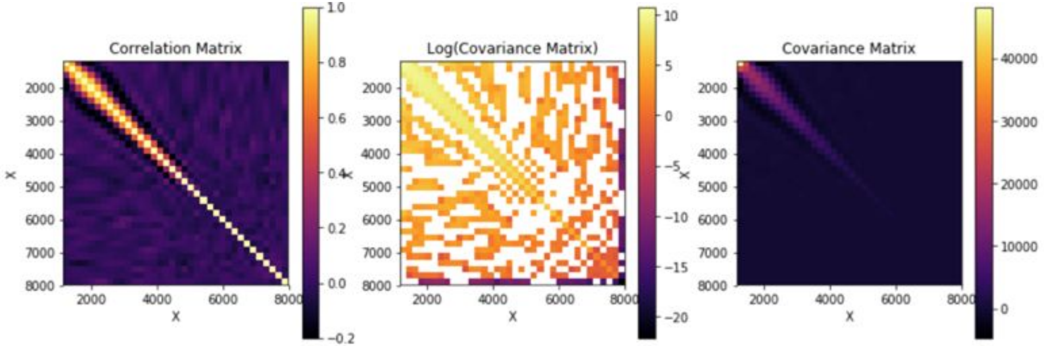
dijet data from 2015 GS analysis (3.2/fb)

Samples, pulls and covariance



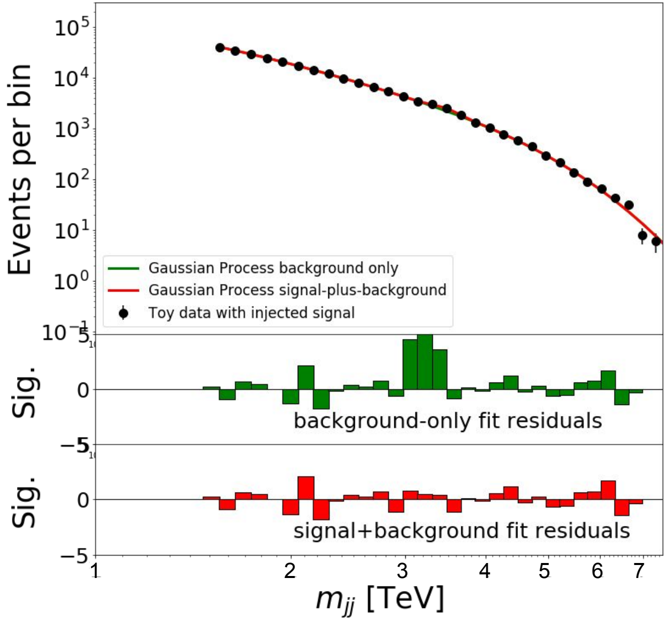


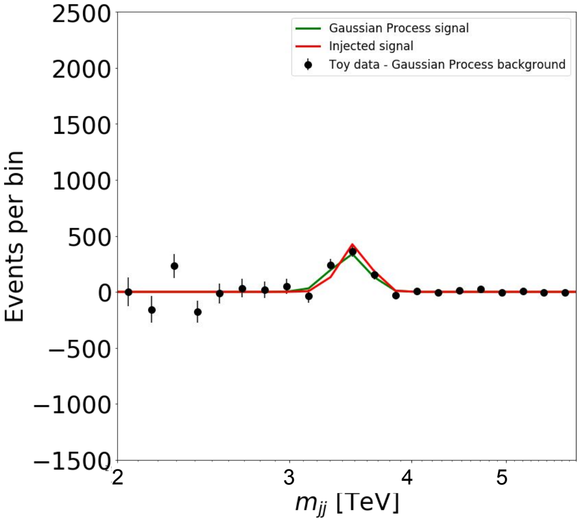
Sample many distributions from the fit and draw pull plots



Covariance+correlation structure of the GP (kernel)

2-jet mass spectrum with signal injected





Left - residuals:
Data + signal vs. bkg GP
Data vs. bkg GP

Right:
Extracted signal
→ Final sig kernel hyperparams

Further details

Gaussian Processes using george: link

ML optimization through Migrad - Minuit

Kernels as quoted in yml format available here

Current + Future work

Work in progress:

- Work ongoing with the 2-jet sample from GS
- Produce toy samples from dataset + inject an artificial gaussian signal
- Fit GP in background MC and compare to data+signal
- Use signal kernel to extract signal
- Get rid of the “help” of the parametric approach - in general not available

Directions in the near future:

- Extend this to more (n-jet) signatures
- Further investigations on already-working kernels (exp squared, “custom”, signal kernel)
- Have this as a complementary approach in the context of GS

This Report is part of a project that has received funding from the **European Union's Horizon 2020 research and innovation programme under grant agreement Nº675440**