Single Gaussian Process Method for Arbitrary Tokamak Regimes

Jarrod Leddy[1], Sandeep Madireddy[2],
Eric Howell[1], Scott Kruger[1]

[1]Tech-X Corporation
[2]Argonne National Lab

funded by DOE Award DE-SC0021380

Motivation

- Equilibrium reconstruction is a process that uses experimental measurements to calculate flux surface location, current profile, q-profile, pressure profile, etc.

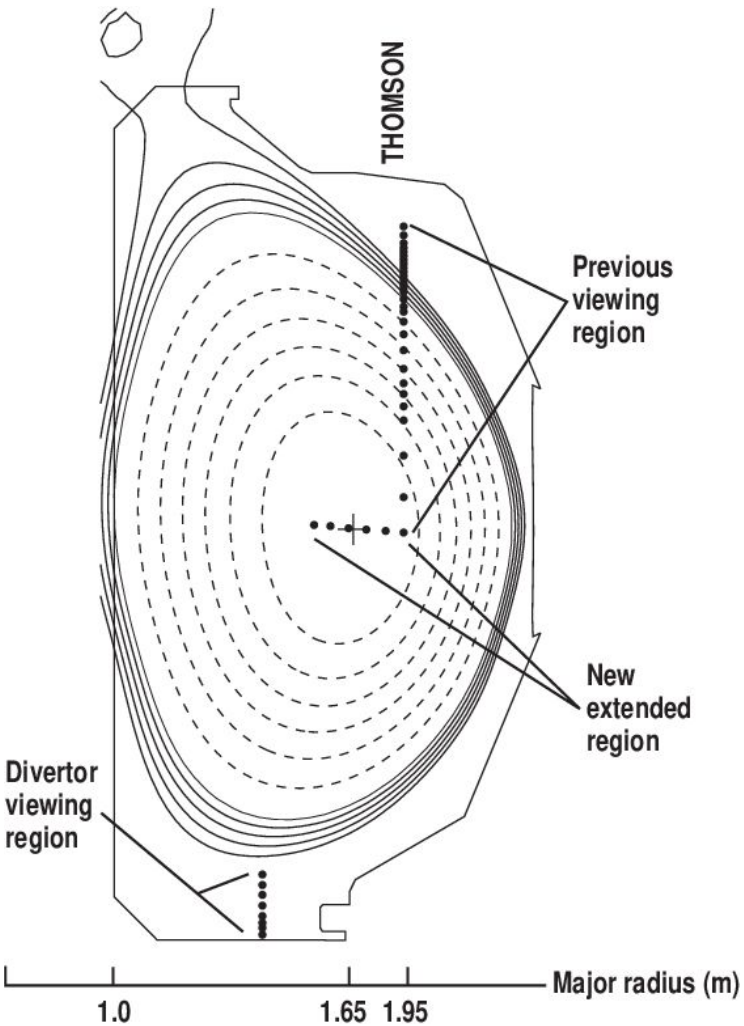
- These values help us understand the state of the plasma and can be used in calculations and simulations regarding transport, MHD stability, and more

- The experimental input for equilibrium reconstruction consists of various data measured by diagnostics in the system
  - These can be noisy, contain outliers, and the qualitative behavior can vary significantly depending on the tokamak regime

We present a method for representing (fitting) these data using Gaussian Process Regression

Experimental profiles

- The Thomson scattering diagnostic provides data used to calculate density and temperature profiles for the plasma
- These profiles vary significantly depending on the tokamak regime (L-mode/H-mode/other)
- Spatial resolution varies across the domain with lower resolution in the core and higher resolution in the edge/pedestal region
- Raw data are spectra that are fit to obtain n_e and T_e, which can therefore be noisy and fits can go poorly leading to outliers



Fitting Experimental Data

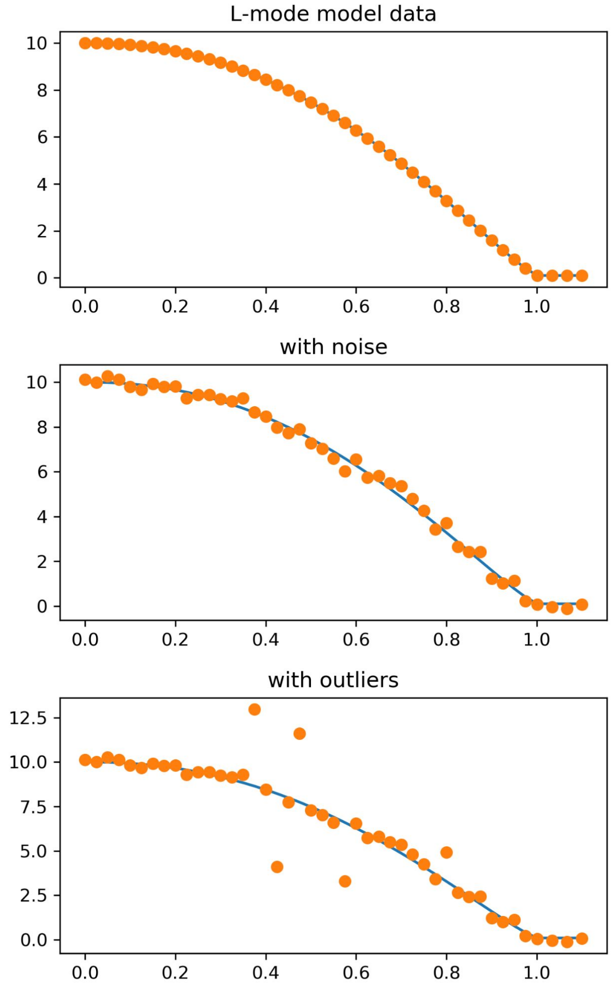
An ideal fitting technique:

- provides good fits for the data even with missing/bad data points
- is non-parametric thus requires minimal assumptions about the data
- prevents overfitting
- accounts for uncertainties in the data and provides proper errors on the fit

Gaussian Process Regression (GPR) is able to achieve all of these objectives with the right setup

To set it up correctly, it is useful to test with synthetic data

Synthetic Data



- To test fitting techniques as we experimented with various settings, we developed a python module to generate synthetic data
- This includes methods to:
  - add gaussian noise of a specified sigma
  - add a number of outliers with a specified offset
  - add pedestals in specified locations and sizes to create H-mode and ITBs
  - Keep a function of the underlying model for error calculation

Overview of GPR

- GPR is nonparametric regression method based on Bayes' theorem

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}} \longrightarrow p(\mathbf{w}|\mathbf{y}, X) = \frac{p(\mathbf{y}|X, \mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|X)}$$

where **w** represents the hyperparameters/model, **y** are the observed data, *X* contains the locations of the points **y**

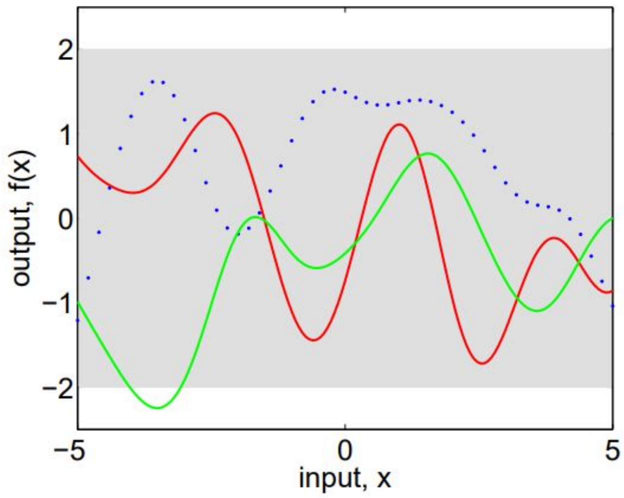
- The **predictive distribution** can be calculated by weighting all possible predictions by their calculated posterior distribution

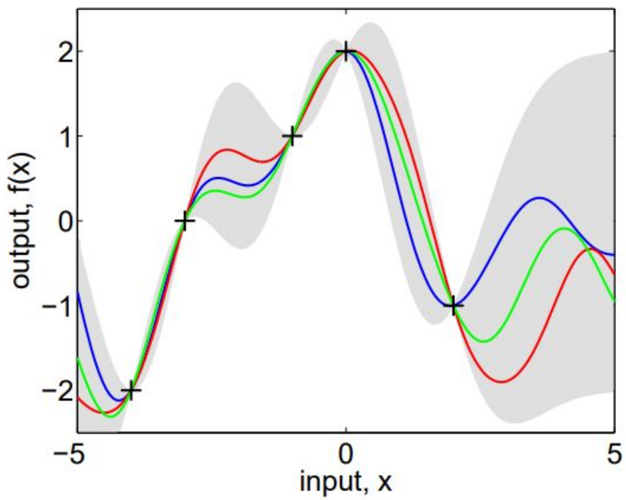
$$p(f^*|x^*, y, X) = \int_w p(f^*|x^*, w)p(w|y, X)dw$$

where *f** and *x** are the value and location of the fit

Overview of GPR

- The fit is generated by sampling from a multivariate normal distribution with a specified covariance matrix
- It is possible to generate random, smooth data sets (like in the upper right) with GPR fitting no data points
- We can constrain the prediction by adding data points (with errors) reducing the variance near to these points (lower right)

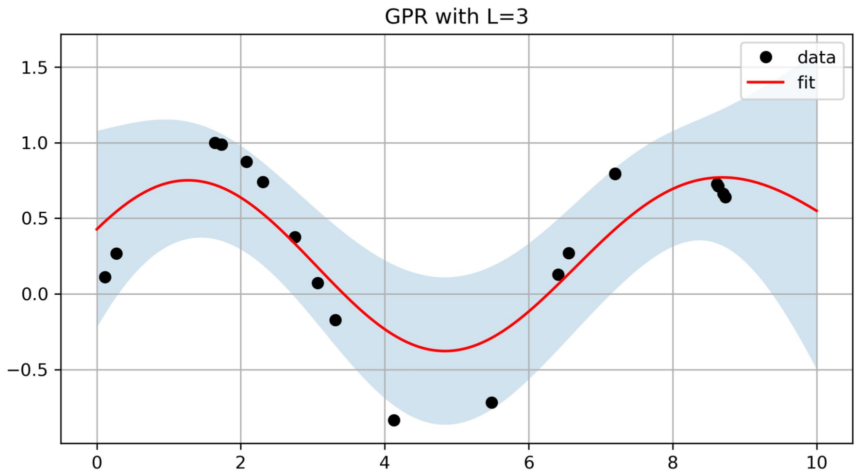


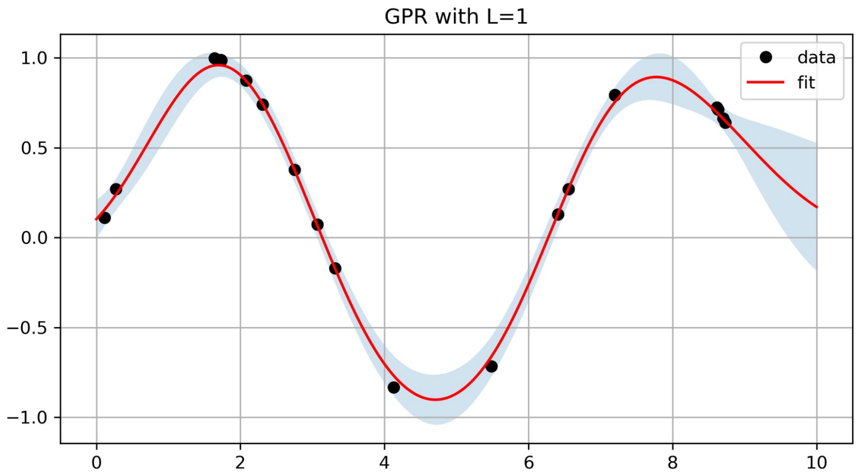


So how is the covariance matrix calculated?

GPR Kernel

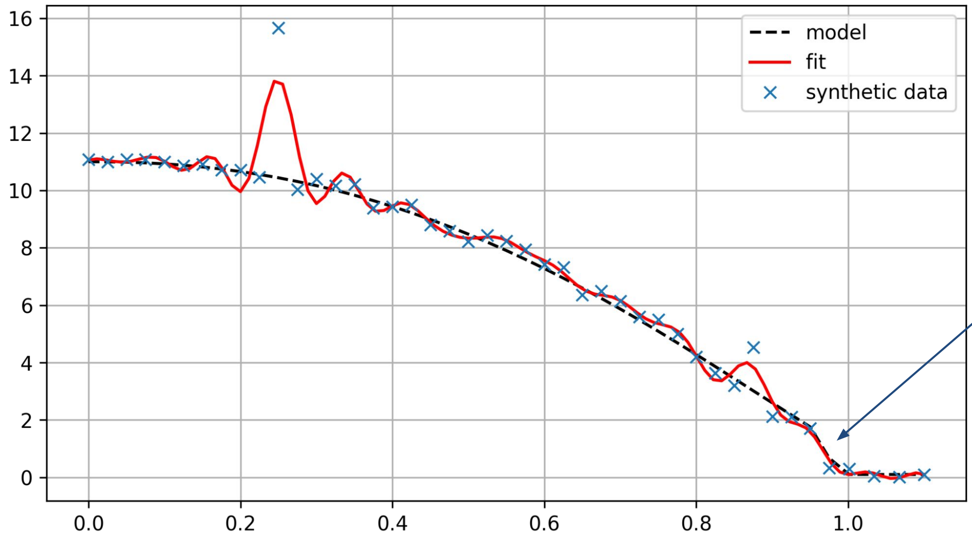
- The **kernel** is a function used to calculate the covariance matrix using the desired output axis as well as the location of the data points
- It is a function of the distance between points, and for GPR it is usually a peaked function at d=0, such as a squared exponential or Matern function
- This means points close to each other are highly correlated, and points far from each other are nearly uncorrelated
- These kernels have a hyperparameter defining this correlation length scale, L, and its value can significantly change the resulting fit





GPR Kernel choice

- The value for the length scale is important for obtaining a good fit - how do we choose this?
- Additionally, does a single length scale satisfactorily provide fits for tokamak profiles?

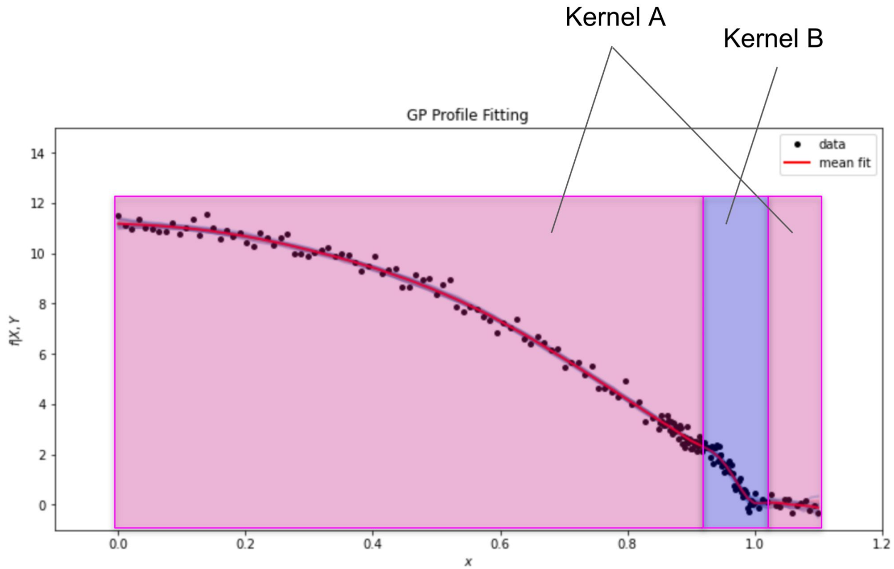


Small length scale fits pedestal well - overfits the rest of the profile

- For H-mode, we need a **variable length scale** so that the pedestal can be fit without overfitting the rest of the profile

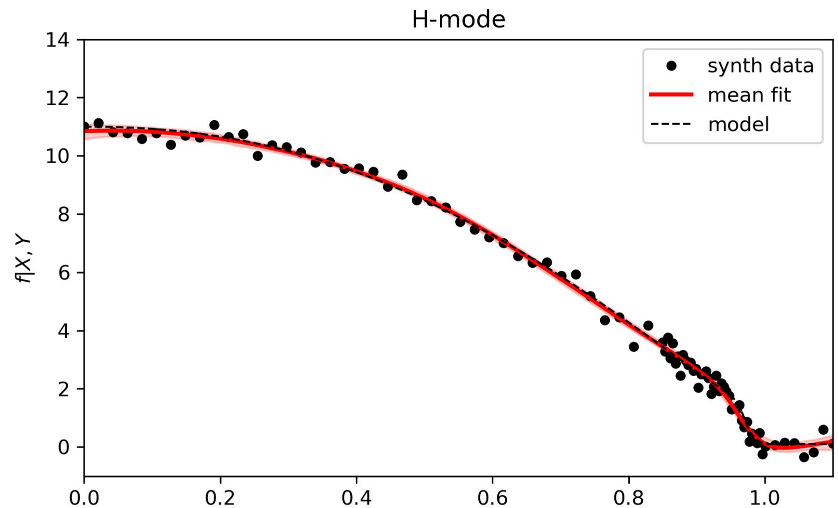
GPR Kernel choice (cont)

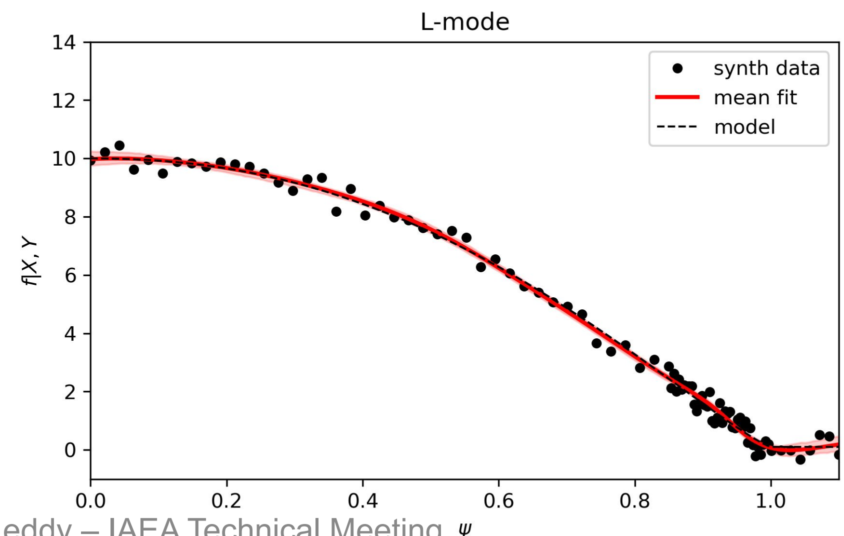
- This usually requires a non-stationary kernel, such as the Gibbs kernel, adding many more hyperparameters to characterize the spatial function of the length scale
- We instead use a change point kernel (GPFlow feature) that switches between stationary kernels at certain radial locations
- We chose to use two Matern kernels, each with independent hyperparameters
  - Matern kernel is a standard covariance function that is a generalization of a gaussian function
- The combination of these two kernels allows us to fit both H-mode and L-mode with the same setup

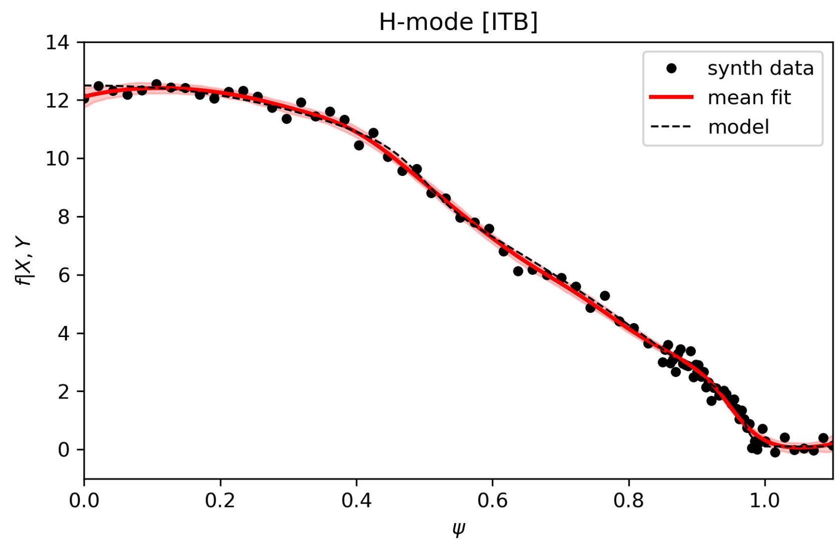


Fitting Noisy Data in Arbitrary Regimes

- Noisy data are fit with smooth curves without overfitting
- The same settings generate fits for various regimes without having to identify them before fitting
- But **how** are the length scales determined?

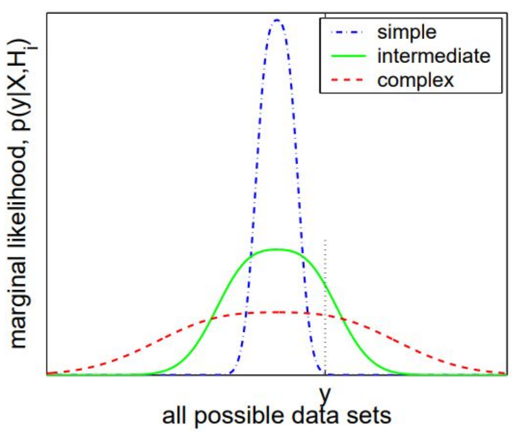






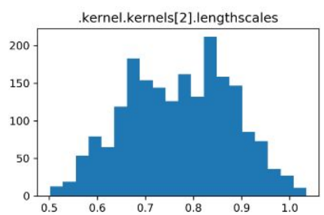
Hyperparameter Optimization

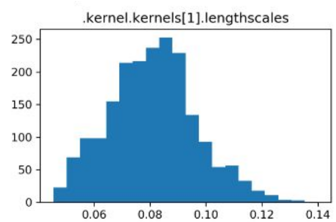
- A **marginal likelihood function** can be defined that is used as a sort of cost function for hyperparameter optimization
- This function calculates the probability of the data given the fit - p(y|X,H)
- The marginal likelihood is a *normalized* probability distribution, so maximizing it will prefer intermediate complexity instead of too simple (bad fit) or too complex (overfit)
- This ensures a good fit without overfitting overfitting is prevented.
- We use a full Bayesian approach to sample the hyperparameter space instead of simply finding optimal values
  - This is opposed to “empirical Bayes” where gaussian distributions are assumed for the hyperparameters

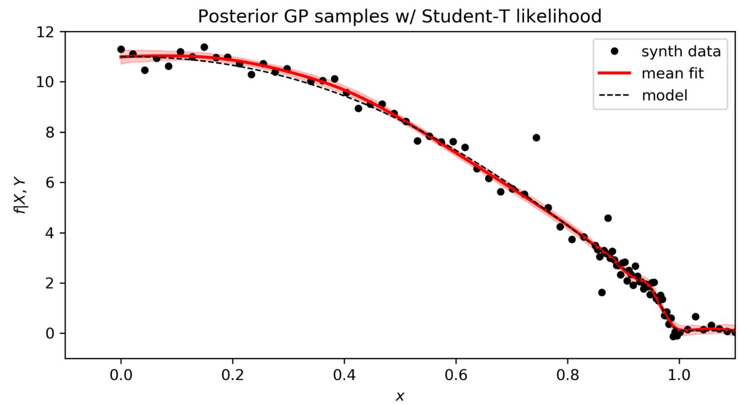


Hyperparameter Optimization

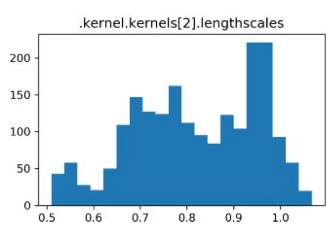
- H-mode - we see large length scale for most of the profile, and an order of magnitude smaller length scale in the pedestal region

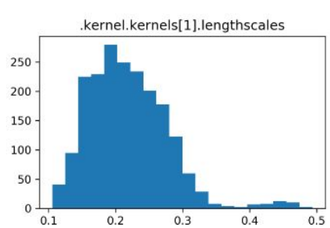


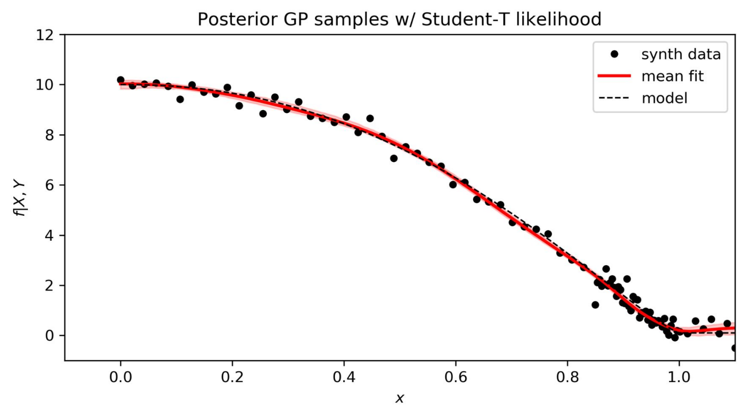




- L-mode - only a small difference between length scales in different regions

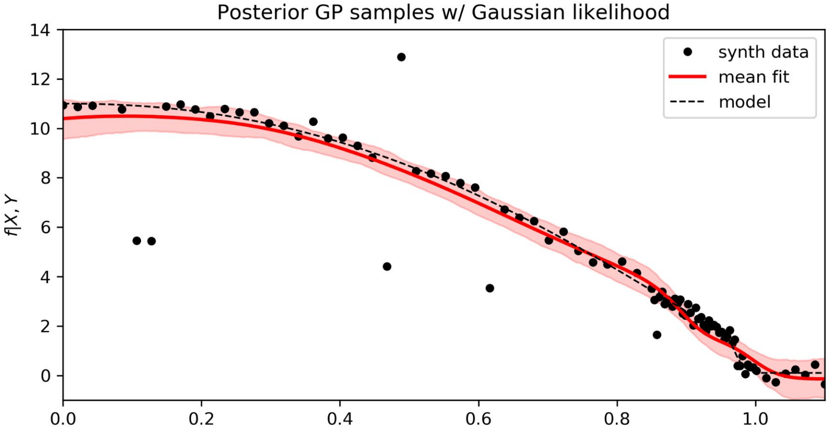


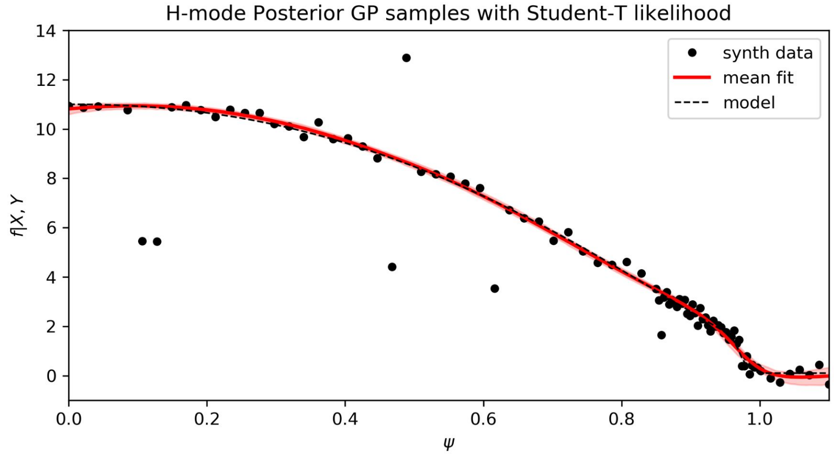




Accounting for Outliers

- The standard likelihood function is a gaussian when errors on the data are thought to be gaussian
- When a data set has outliers in addition to noisy data, the assumed gaussian error would have to be much larger to account for the outliers
- We can instead use a heavy-tailed likelihood function, specifically **student-t** that approaches gaussian at high degrees-of-freedom and is heavy tailed at low degrees-of-freedom
- No need for prior knowledge of which points are outliers





Summary

- Fitting plasma profiles is important for equilibrium reconstruction
- GPR provides a robust and accurate algorithm for fitting profiles (and can be readily extended for multidimensional fitting)
- A linear combination of stationary kernels can be used with hyperparameter optimization to fit arbitrary tokamak regimes
- A student-t likelihood function allows GPR to fit data sets containing outliers without compromising the quality of the fit

Next Steps:

- Expand use of fitting to experimental Thomson scattering data
- Comparisons with parameterized methods
- Application of GPR to magnetics data including inferring missing data

References

1. Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian processes for machine learning*. MIT Press.
2. Matthews, A. G., van der Wilk, M., et al (2017). GPflow: A gaussian process library using TensorFlow. Journal of Machine Learning Research.
3. Chilenski, M. A., Greenwald, M., et al (2015). *Improved profile fitting and quantification of uncertainty in experimental measurements of impurity transport coefficients using Gaussian process regression*. Nuclear Fusion.