# WHEN IS A NOISY SIGNAL OSCILLATING AND WHEN IS IT NOISY ?

SUPERVISED BY PETER SWAIN

# WHEN IS A NOISY SIGNAL OSCILLATING AND WHEN IS IT NOISY ?

## Topics to be covered

PROBLEM AT HAND

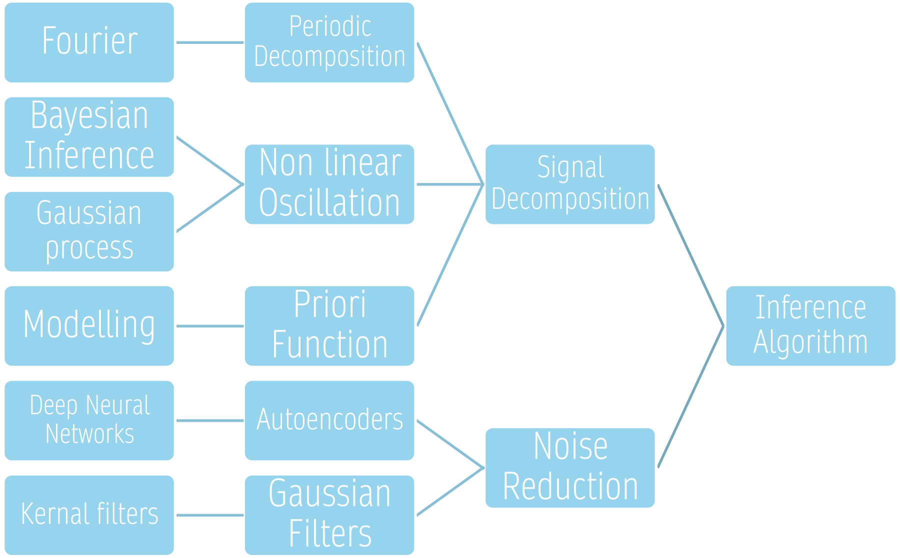
INFERENCE ALGORITHMS

BAYESIAN FRAMEWORK

GAUSSIAN PROCESS REGRESSION

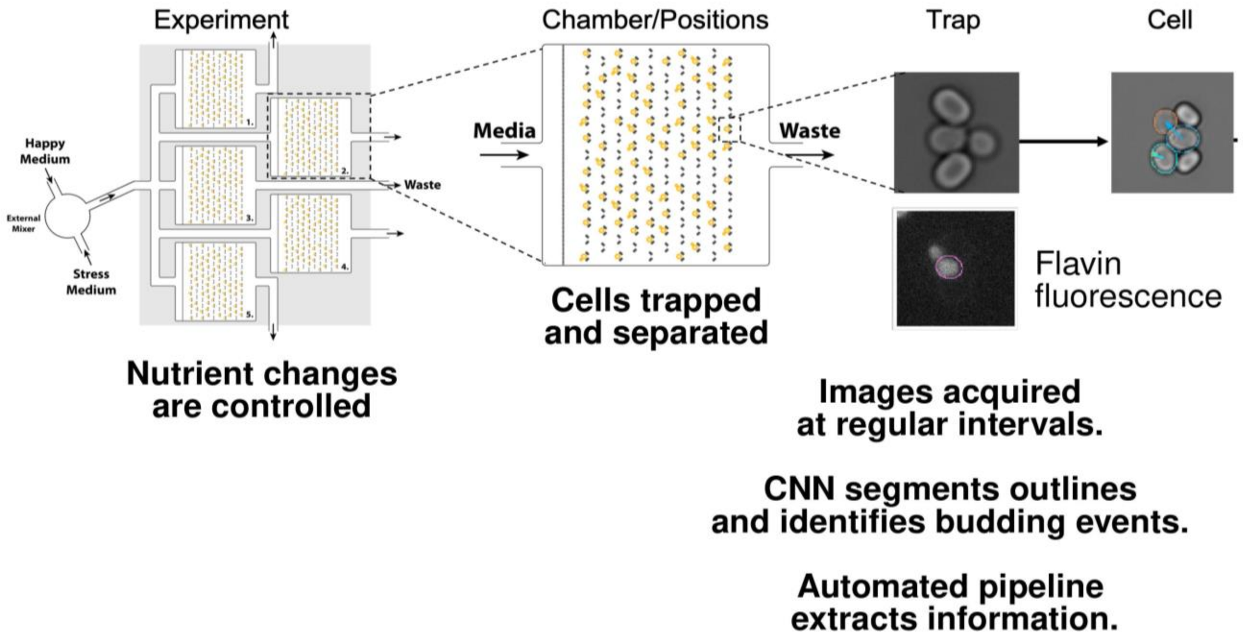
CONCLUDING

## Prior of project



# PROBLEM AT HAND

- **WE ARE ANCHORING OUR METHODS TO THIS PROBLEM AND SOLUTION**




Preprint, Astrophysical Journal in press Jan. 17, 1999

BAYESIAN PERIODIC SIGNAL DETECTION. 1.
ANALYSIS OF 20 YEARS OF RADIO FLUX MEASUREMENTS OF THE X-RAY BINARY LS I $+61^\circ 303$

P. C. Gregory
Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, V6T 1Z1

ABSTRACT

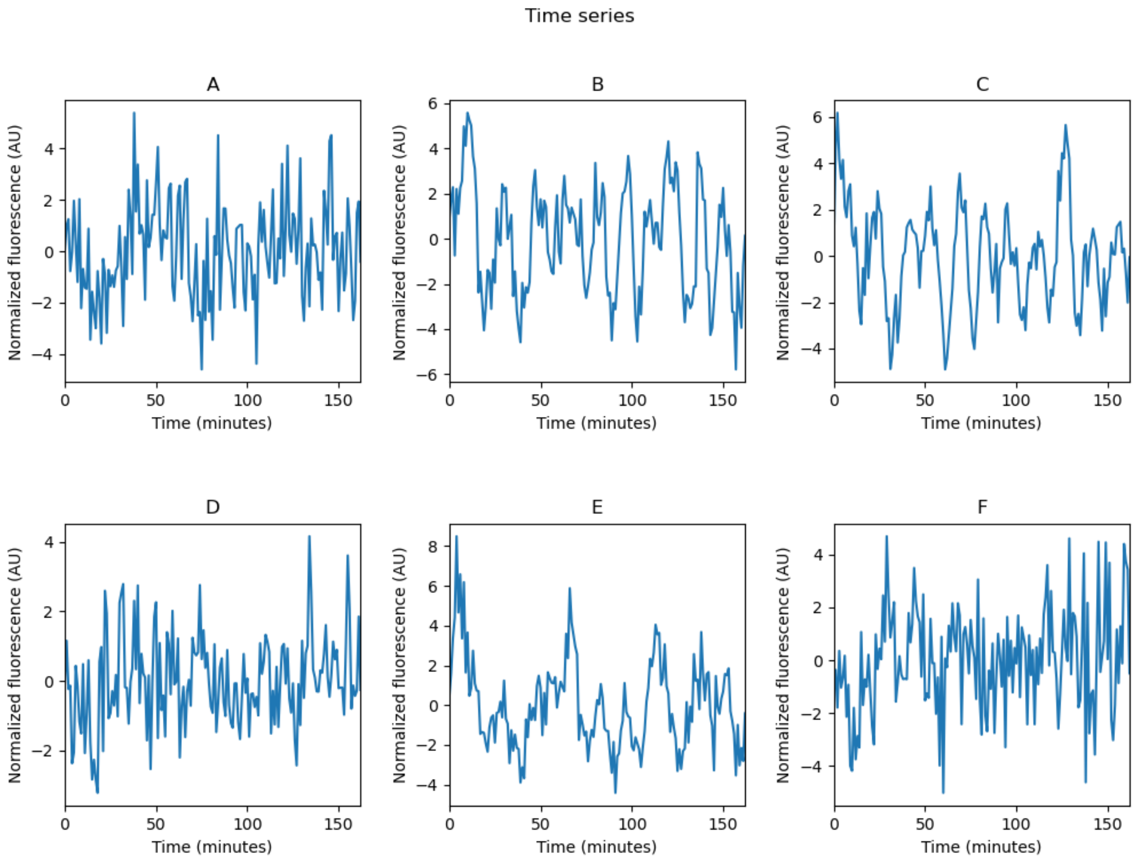
We extend the Bayesian method of Gregory and Loredo (1992) for the detection of a periodic signal of unknown shape and period, to the case where the noise sampling distribution is independent Gaussian. The analysis readily handles nonuniformily sampled data and allows for an unknown noise variance. The method is applied to the radio astronomy data for the interesting X-ray binary system LS I $+61^\circ 303$, which exhibits periodic radio outbursts with a period of 26.5 days. Several authors have suggested that the peak flux density of the outbursts exhibit a periodic or quasi-periodic modulation of approximately 1600 days. Our Bayesian analysis of the outburst peak flux densities provides strong support for such a modulation. We derive the posterior probability density function of the modulation period and the estimated mean shape of the modulation based on the available flux density data. The probability density for the modulation period exhibits a broad peak in the range 1599 to 1660 days (68 % credible region) with a mean value of 1632 days. The RMS flux density deviation from the mean shape, amounting to 45 mJy, is much larger than the measurement errors of $\approx 10$ mJy which suggest additional complexity in the source which is yet to be understood. The next maximum in the long term flux modulation is predicted to occur near July 22, 1999 (Julian day 2,451,382).

*Subject headings:* Bayesian methods, period detection, Gregory-Loredo method, LS I $+61^\circ 303$, X-ray binaries, pulsars, time series analysis

1. INTRODUCTION

LS I $+61^\circ 303$ (V615 Cas, GT 0236+610, 2CG 135+01) is particularly interesting among high-mass X-ray binaries because of its strong variable emission from radio to X-ray and probably $\gamma$-ray (Gregory and Taylor 1978, Kniffen et al. 1997). At radio wavelengths it exhibits periodic

# WHEN IS A NOISY SIGNAL NOISY AND WHEN IS IT OSCILLATING



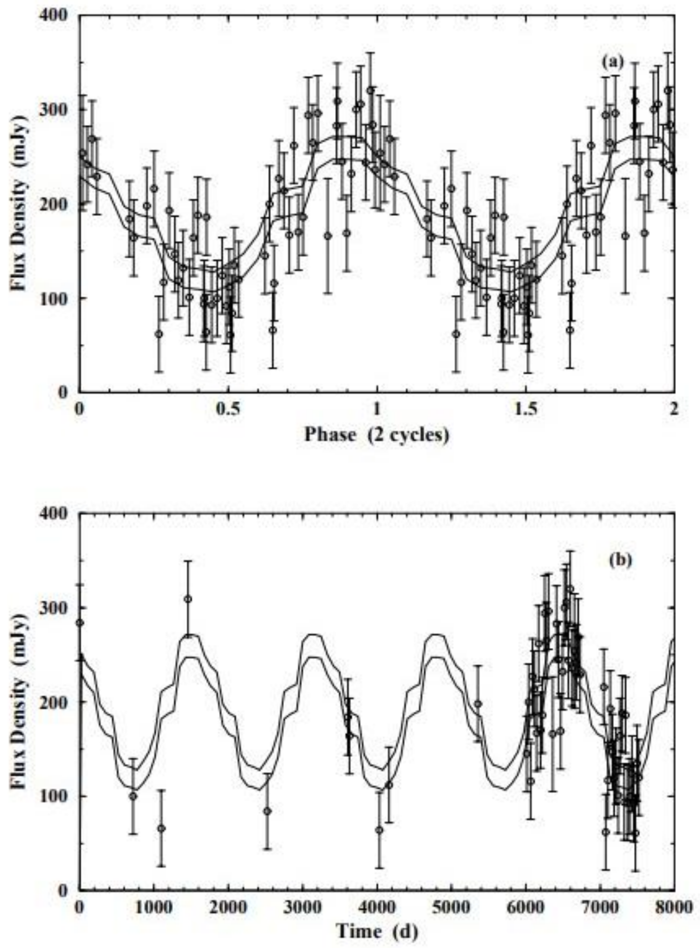




Fig. 3.— Panel (a) shows the shape estimate of the outburst peak flux modulation for LSI +61°303 and measured values plotted for two cycles of phase. The solid curves shown are the mean flux ±1 one standard deviation. Panel (b) compares the Bayesian flux modulation light curve with the data.

# FOURIER ANALYSIS IN ONE SLIDE

## FFT

DFT

$$\begin{pmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \\ \vdots \\ \hat{f}_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & & 1 \\ 1 & w_n & w_n^2 & \cdots & w_n^{n-1} \\ 1 & w_n^2 & w_n^4 & & w_n^{2(n-1)} \\ & \vdots & & \ddots & \vdots \\ 1 & w_n^{n-1} & \cdots & \cdots & w_n^{2(n-1)} \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{pmatrix}$$

$$where\ w_n = e^{-i2\pi/n}$$

Fourier series in limiting case

$$\hat{f}_k = \sum_{j=0}^{n-1} f_n e^{-i2\pi jk/n} \pi r^2$$

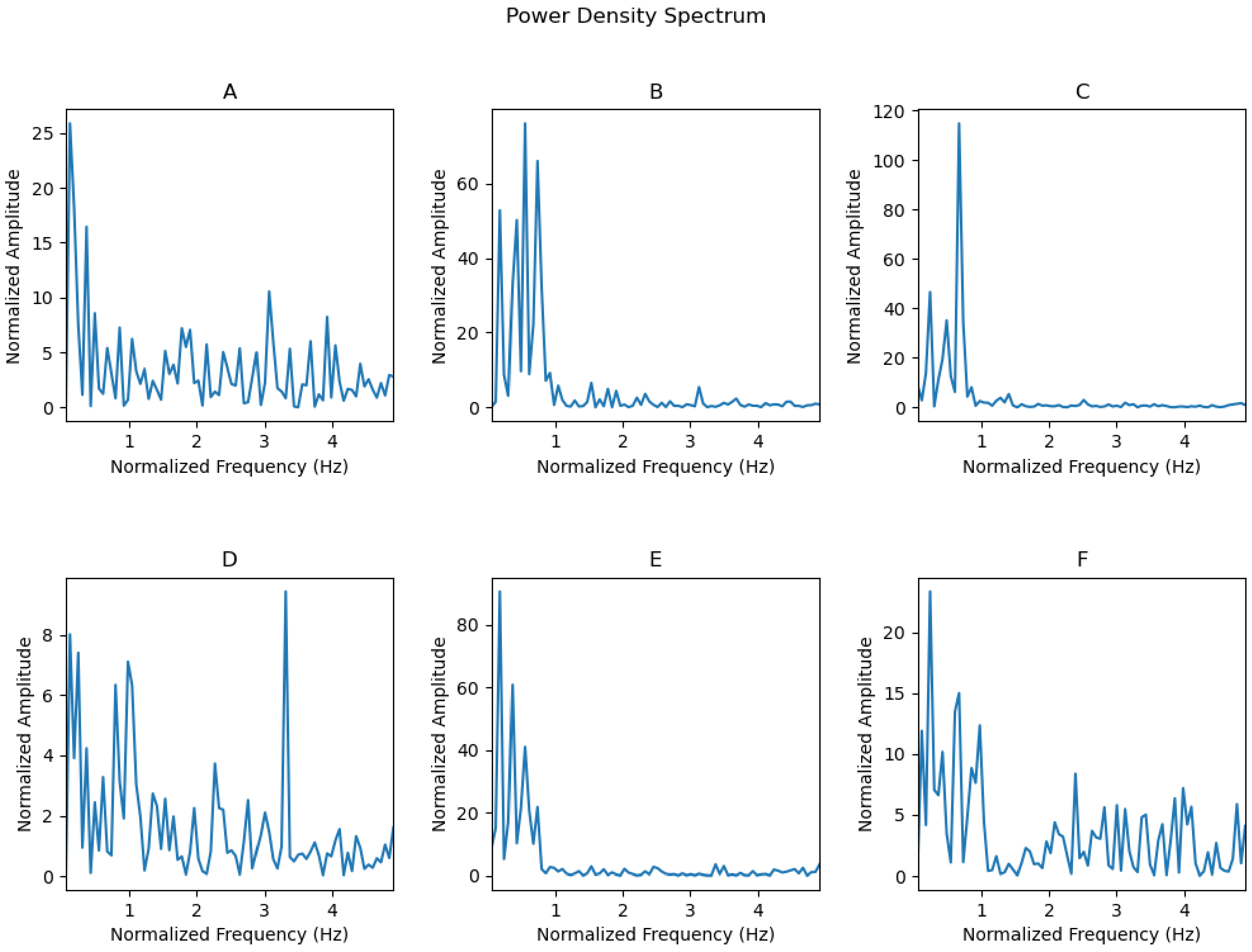
FFT

$$\hat{f}_k = \mathcal{F}_{1024} f = \begin{pmatrix} I_{512} & -D_{512} \\ I_{512} & -D_{512} \end{pmatrix} \begin{pmatrix} \mathcal{F}_{512} & 0 \\ 0 & \mathcal{F}_{512} \end{pmatrix} \begin{pmatrix} f_{even} \\ f_{odd} \end{pmatrix}$$

$$\mathcal{F}_{1024} \rightarrow \mathcal{F}_{512} \rightarrow \mathcal{F}_{256} \rightarrow \mathcal{F}_{128} \ldots \rightarrow \mathcal{F}_2$$

# PROBLEM AT HAND

**FFT**



# INTRODUCTION TO **BAYESIAN FRAMEWORK**

**FFT**

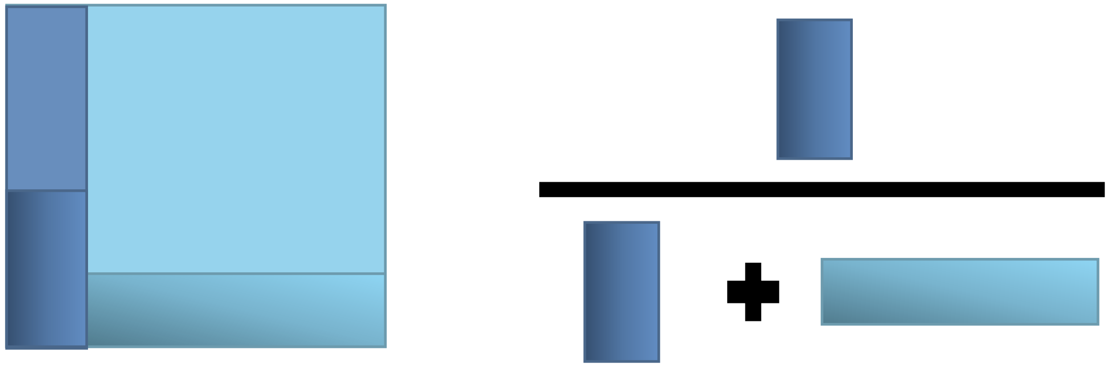
**BAYES**

$$\mathcal{P}(\mathcal{H}_i|\mathcal{E}) = \frac{\mathcal{P}(\mathcal{E}|\mathcal{H}_i)\mathcal{P}(\mathcal{H}_i)}{\mathcal{P}(\mathcal{E})}$$

$Where \quad \mathcal{P}(\mathcal{E}) = \mathcal{P}(\mathcal{E}|\mathcal{H}_i)\mathcal{P}(\mathcal{H}_i) + \mathcal{P}(\mathcal{E}|\neg\mathcal{H}_i)\mathcal{P}(\neg\mathcal{H}_i)$

$$Posterior = \frac{likelihood \times Prior}{Marginal\ likelihood}$$

## BAYES THEOREM PROOF ON FLY

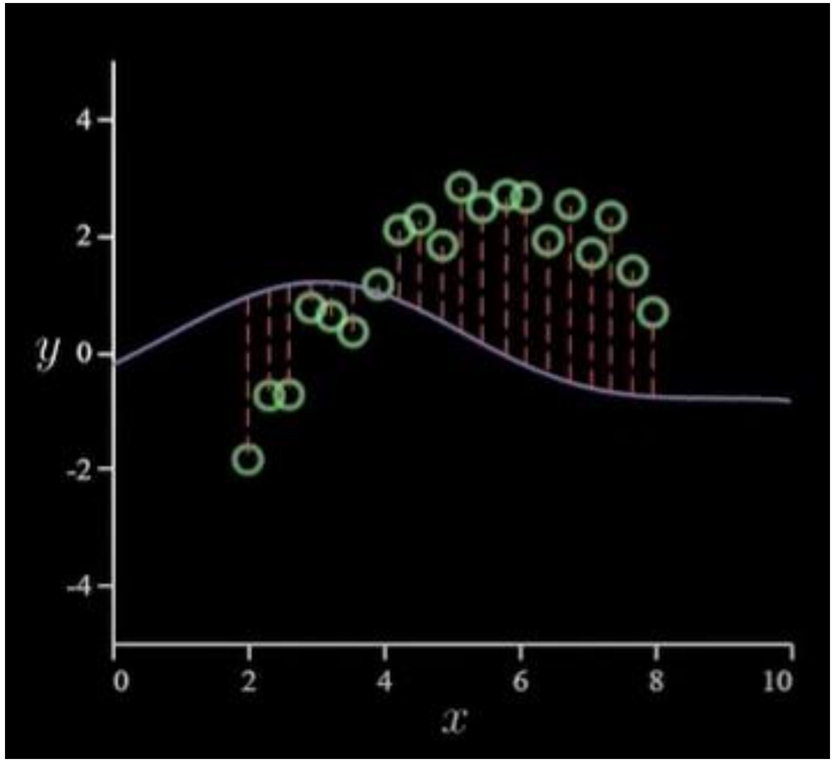


- **A FRAMEWORK**
- **ANALOGY WITH QM**
- **BAYESIAN COMMUNITY**

# WHY IS BAYESIAN FRAMEWORK HELPFUL

**FFT**

**BAYES**




## Game of minimising red

*1)* $f(\tau) \rightarrow \chi^2$ e.g. 30 oscillators

- Random attempt
- Gradient decent
- Monte Carlo

2) What if Hyper-parameter involved e.g. may be 40 oscillators involved

- Optimising parameters for each hyper parameter
- Exponential computation increase

3) Assume we have prior knowledge of hypers distribution.

- E.g. High probability of 30 score
- Lets give prior normal distribution to hypers around 30

# HIERARCHICAL BAYESIAN MODEL SELECTION

FFT

BAYES

Models

Hyper-Parameters

Parameter Value

$$\mathcal{P}(\mathcal{H}_i|\mathbf{y},X) = \frac{\mathcal{P}(\mathbf{y}|X,\mathcal{H}_i)\mathcal{P}(\mathcal{H}_i)}{\mathcal{P}(\mathbf{y}|X)}$$

Where $\mathcal{P}(\mathbf{y}|X) = \sum_i \mathcal{P}(\mathbf{y}|X,\mathcal{H}_i)\mathcal{P}(\mathcal{H}_i)$

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{y},X,\mathcal{H}_i) = \frac{\mathcal{P}(\mathbf{y}|X,\boldsymbol{\theta},\mathcal{H}_i)\mathcal{P}(\boldsymbol{\theta}|\mathcal{H}_i)}{\mathcal{P}(\mathbf{y}|X,\mathcal{H}_i)}$$

Where $\mathcal{P}(\mathbf{y}|X,\mathcal{H}_i) = \int \mathcal{P}(\mathbf{y}|X,\boldsymbol{\theta},\mathcal{H}_i)\mathcal{P}(\boldsymbol{\theta}|\mathcal{H}_i)d\boldsymbol{\theta}$

$$\mathcal{P}(\mathbf{w}|\mathbf{y},X,\boldsymbol{\theta},\mathcal{H}_i) = \frac{\mathcal{P}(\mathbf{y}|X,\mathbf{w},\mathcal{H}_i)\mathcal{P}(\mathbf{w}|\boldsymbol{\theta},\mathcal{H}_i)}{\mathcal{P}(\mathbf{y}|X,\boldsymbol{\theta},\mathcal{H}_i)}$$

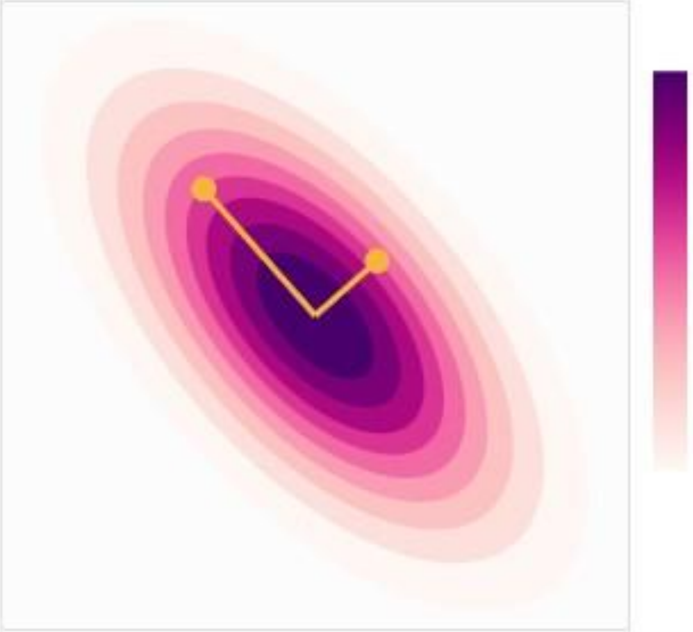
Where $\mathcal{P}(\mathbf{y}|X,\boldsymbol{\theta},\mathcal{H}_i) = \int \mathcal{P}(\mathbf{y}|X,\mathbf{w},\mathcal{H}_i)\mathcal{P}(\mathbf{w}|\boldsymbol{\theta},\mathcal{H}_i)d\mathbf{w}$

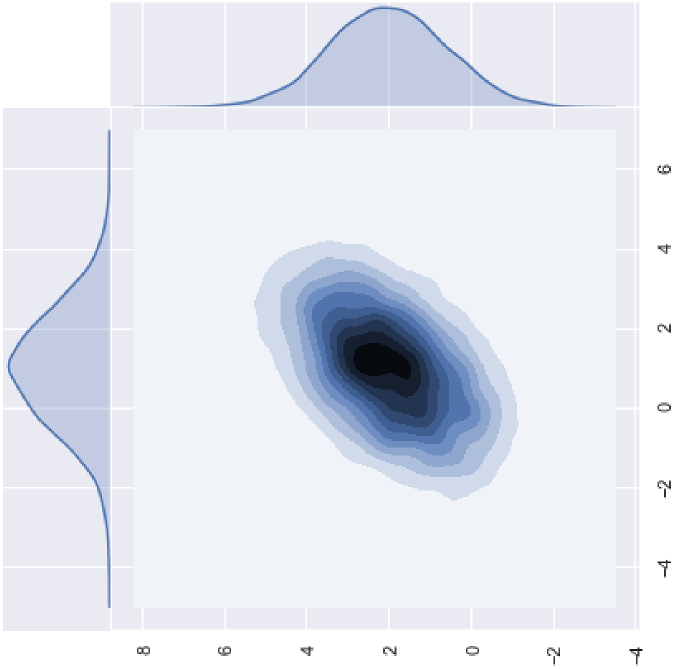
# WHAT IS MULTIVARIANT GAUSSIAN DISTRIBUTION

**FFT**

**BAYES**

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \sim N(\mu, \Sigma)$$






Covariance matrix (Σ)

| | |
|---|---|
| 0.66 | -0.44 |
| -0.44 | 0.78 |

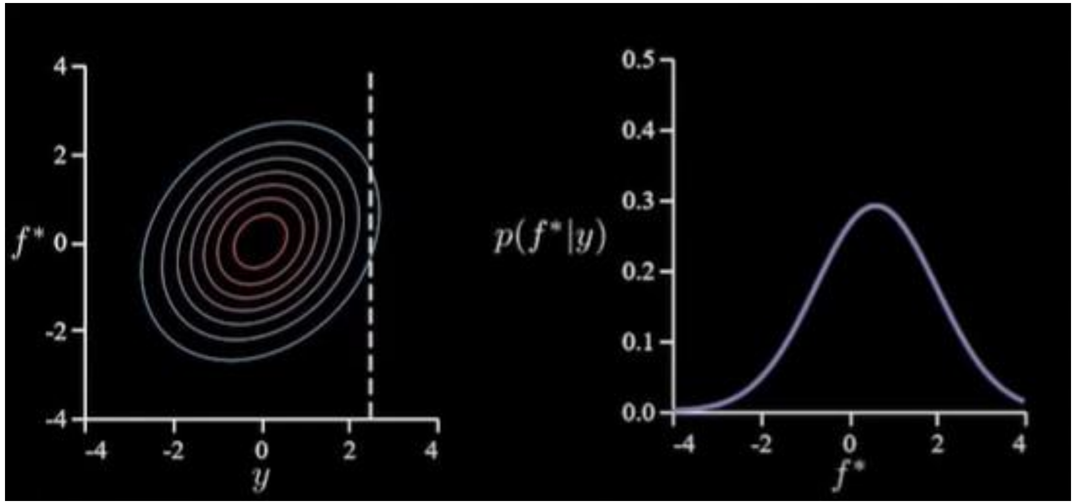
$$\Sigma = Cov(X_i, X_j) = E\left[(X_i - \mu_i)(X_j - \mu_j)^T\right]$$

# SPECIAL FEATURES OF MULTIVARIABLE GAUSSIAN

FFT

BAYES

## CONDITIONING




## MARGINALIZING

$$P_{X,Y} = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\mu, \Sigma) = N\left( \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right)$$

$$\log \mathcal{P}(\boldsymbol{y}|X, \boldsymbol{\theta}) = log \int \mathcal{P}(\boldsymbol{y}|X, \boldsymbol{w}, \mathcal{H}_i)\mathcal{P}(\boldsymbol{w}|\boldsymbol{\theta}, \mathcal{H}_i)d\boldsymbol{w} \quad = \log N\left(y|o, \acute{\Sigma}_{x,x}\right)$$

# BENEFITS OF USING GAUSSIAN DISTRIBUTION IN BI

**FFT**

**BAYES**

## OPTIMIZATION :

$$\frac{\partial}{\partial \theta_i} log(y|X,\theta) = \frac{1}{2} y^T K^{-1} \frac{\partial K}{\partial \theta_i} K^{-1} y \quad - \frac{1}{2} \mathrm{tr}\left(K^{-1} \frac{\partial K}{\partial \theta_i}\right)$$

$$= \frac{1}{2} \mathrm{tr}\left( (\alpha \alpha^T - K^{-1}) \frac{\partial K}{\partial \theta_i} \right) \quad \text{where } \alpha = K^{-1} y$$

## PREDICTION :

$$\mathrm{Pr}(\mathrm{y}_* | x_*, \mathrm{X}, \mathrm{w}) = N\left(\sigma^{-2} \phi(x_*)^T A^{-1} \phi y \,, \;\; \phi(x_*)^T A^{-1} \phi(x_*)\right)$$

Where $\overline{w} = \sigma^2 A^{-1} \phi y$ and $\mathrm{A} = \sigma^{-2} \phi \phi^T + \Sigma^{-1}$

# GAUSSIAN PROCESS REGRESSION

## SWITCHING FROM WEIGHT TO FUNCTION SPACE VIEW

**FFT**

**BAYES**

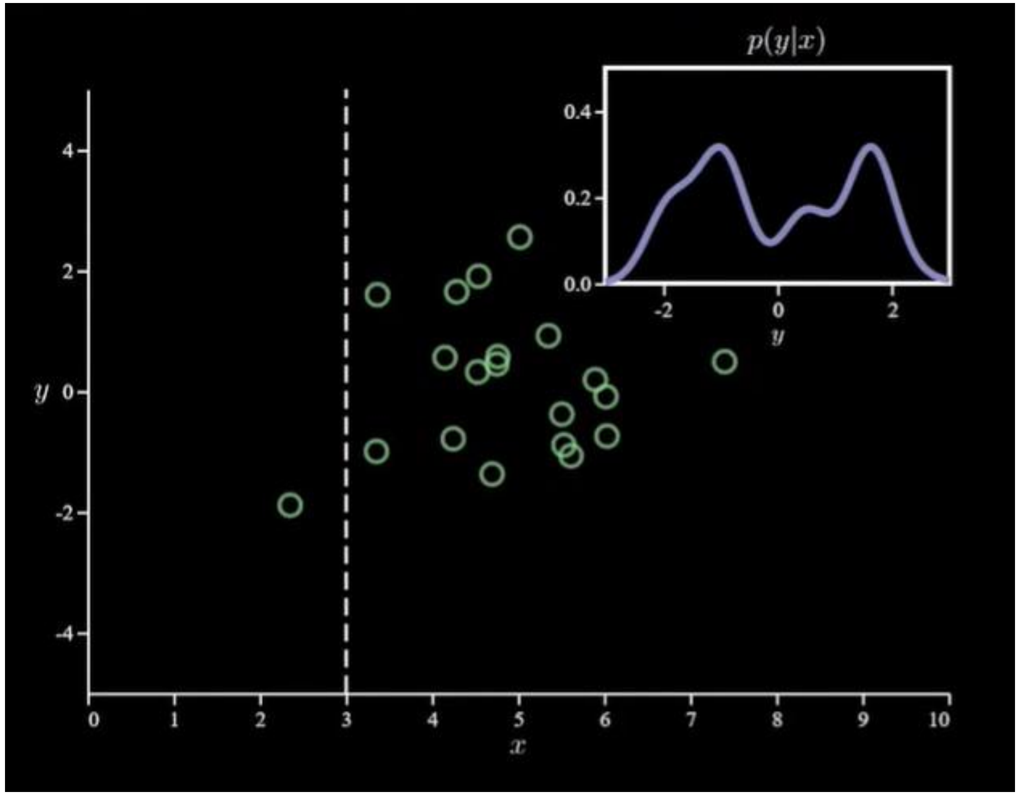
**GAUSSIAN PROCESS**

- **WHAT DOES IT MEAN ?**
- **HOW IT WILL BE HELPFUL ?**
- **WHAT ARE PARAMETERS ?**

$$y_i = f(x_i) + \varepsilon_i$$

$$f(x) \sim GP(m(x), k(x, \acute{x}))$$

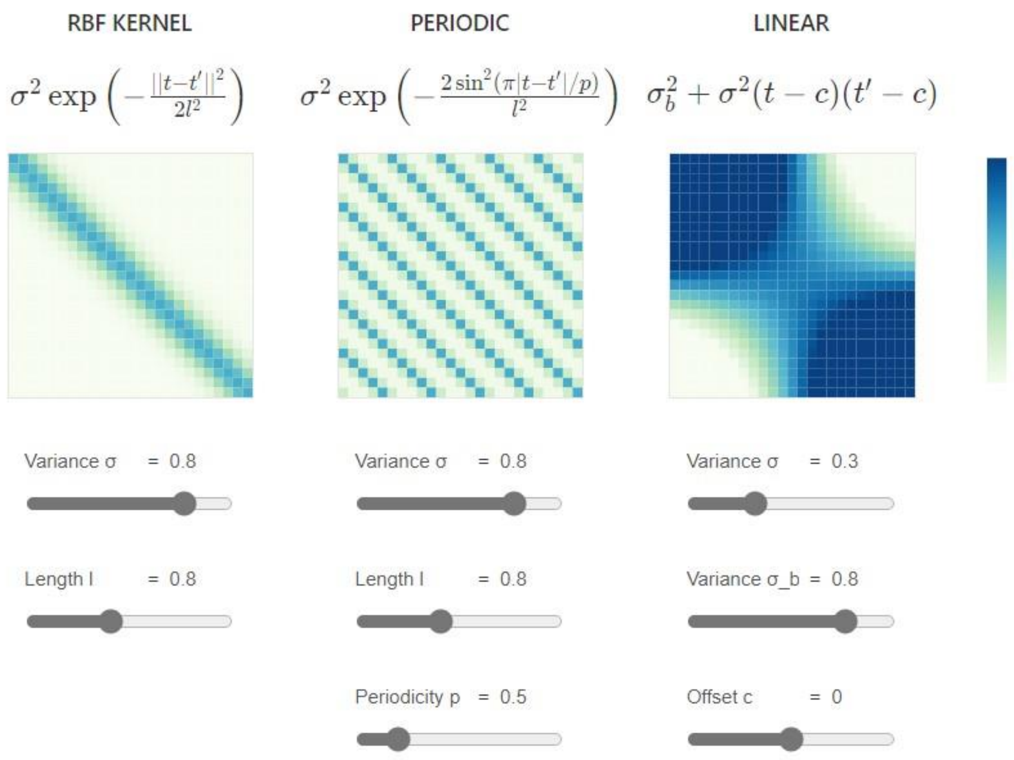
$$\kappa : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad \Sigma = Cov(X_i, X_j) = \kappa(t, \acute{t})$$



# KERNEL IN GAUSSIAN PROCESS

FFT

BAYES




- NOT THE PARAMETERS
- CAN NOT BE ANY FUNCTION
- NEW KERNELS FROM OLD
- KERNEL HYPERPARAMETERS CAN GIVE FUNCTION PARAMETERS

$$\kappa(\boldsymbol{x}, \acute{x}|\tau)$$

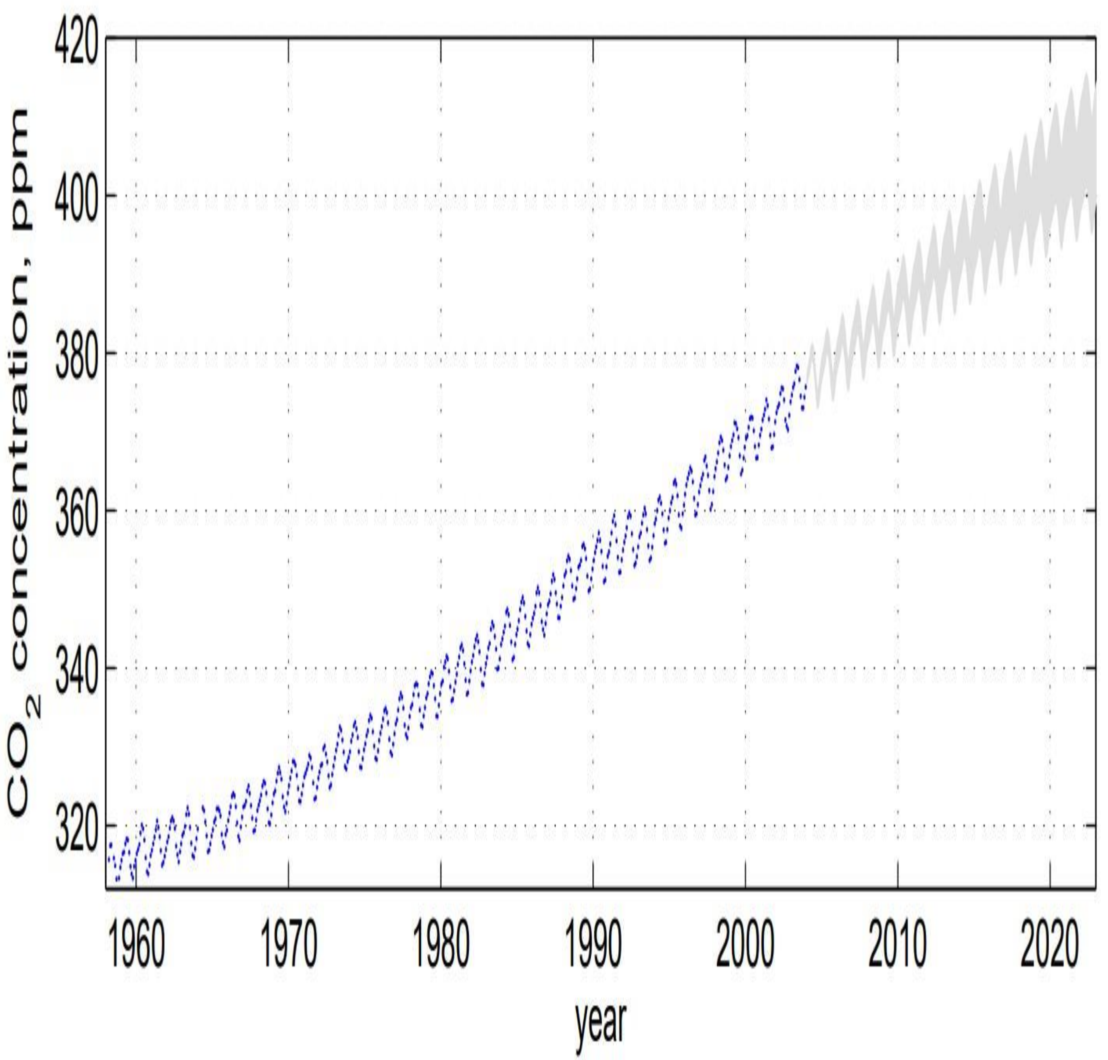
$$\kappa_c(\boldsymbol{x}, \acute{x}|\tau_c) = \kappa_a(\boldsymbol{x}, \acute{x}|\tau_a) + \kappa_b(\boldsymbol{x}, \acute{x}|\tau_b)$$

# GP REGRESSION  A WORKING EXAMPLE

**FFT**

**BAYES**

**GAUSSIAN PROCESS**




$$k_1(x,\acute{x}) = \theta_1^2 \exp\left(-\frac{(x-\acute{x})^2}{2\theta_2^2}\right)$$

$$k_2(x,\acute{x}) = \theta_3^2 \exp\left(-\frac{(x-\acute{x})^2}{2\theta_4^2} - \frac{2sin^2(\pi(x-\acute{x}))}{\theta_5^2}\right)$$

$$k_3(x,\acute{x}) = \theta_6^2 \left(1+\frac{(x-\acute{x})^2}{2\theta_8\theta_7^2}\right)^{-\theta_8}$$

$$k_4(x_p,x_q) = \theta_9^2 \exp\left(-\frac{(x_p-x_q)^2}{2\theta_{10}^2}\right) + \theta_{11}^2\delta_{pq}$$

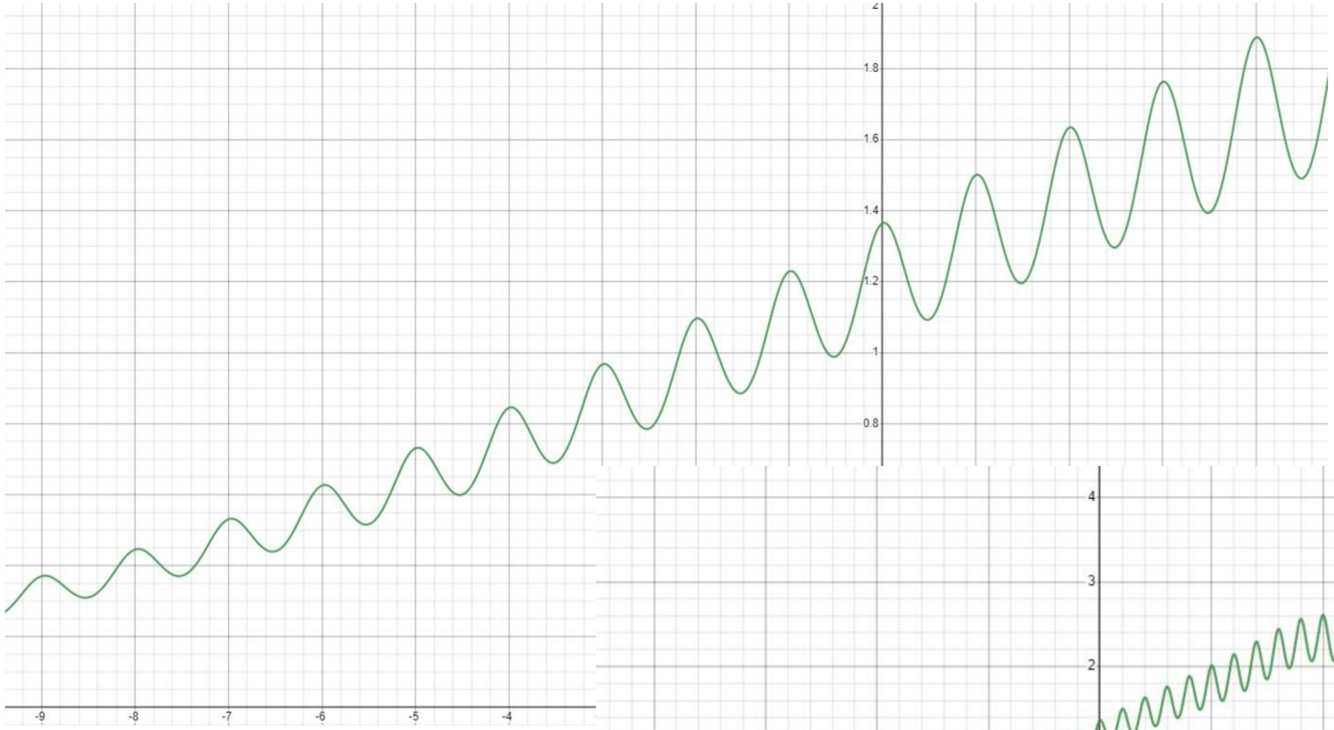
$$k(x,\acute{x}) = k_1(x,\acute{x}) + k_2(x,\acute{x}) + k_3(x,\acute{x}) + k_4(x_p,x_q)$$

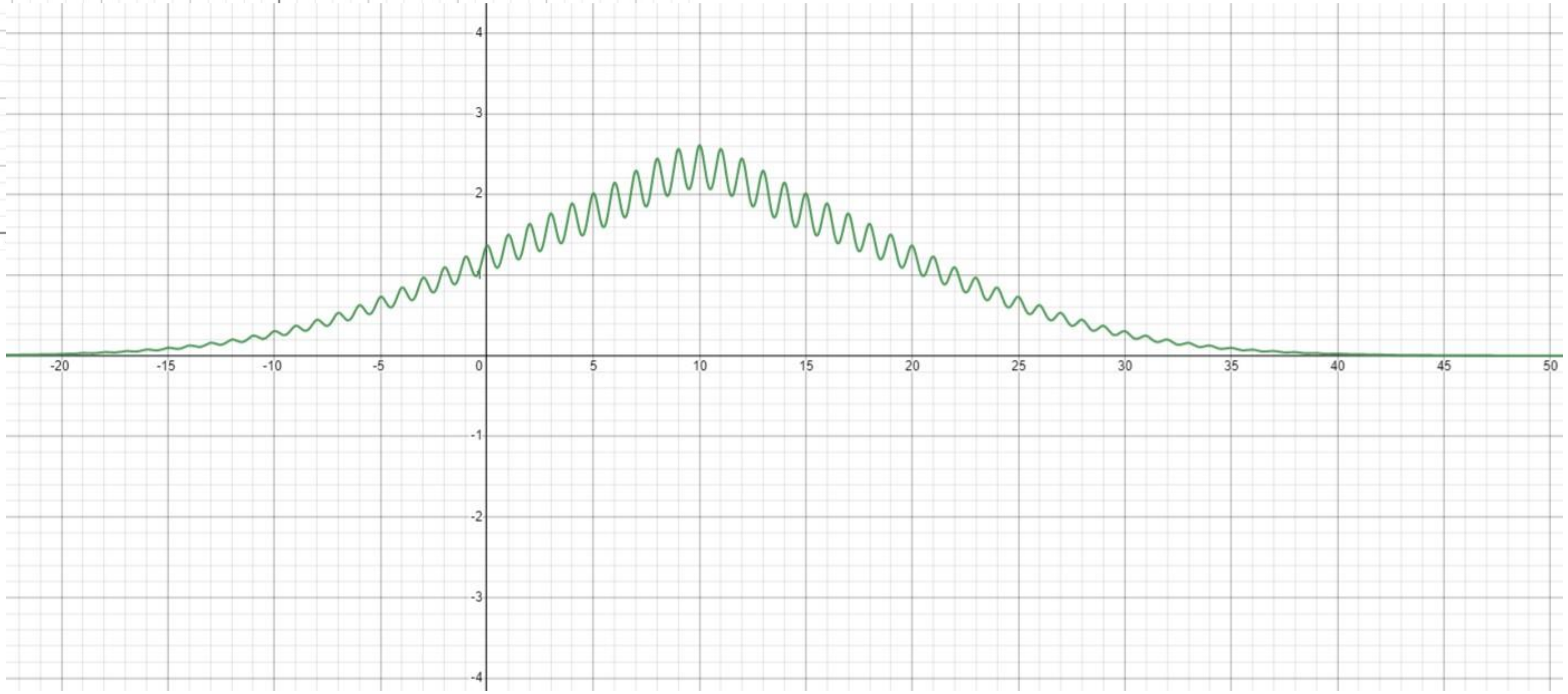
# GP REGRESSION  A WORKING EXAMPLE

**FFT**

**BAYES**

**GAUSSIAN PROCESS**



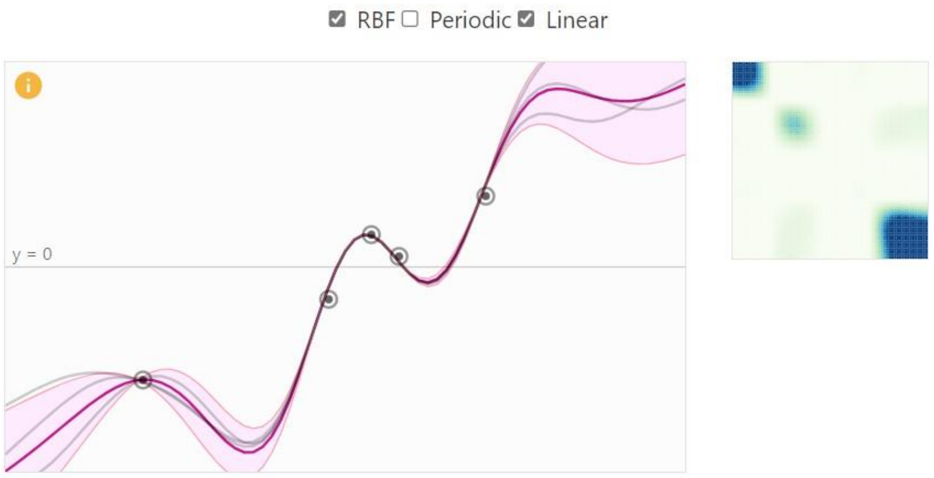



# INTERPRETATION OF GPR RESULTS

FFT

BAYES

GAUSSIAN PROCESS



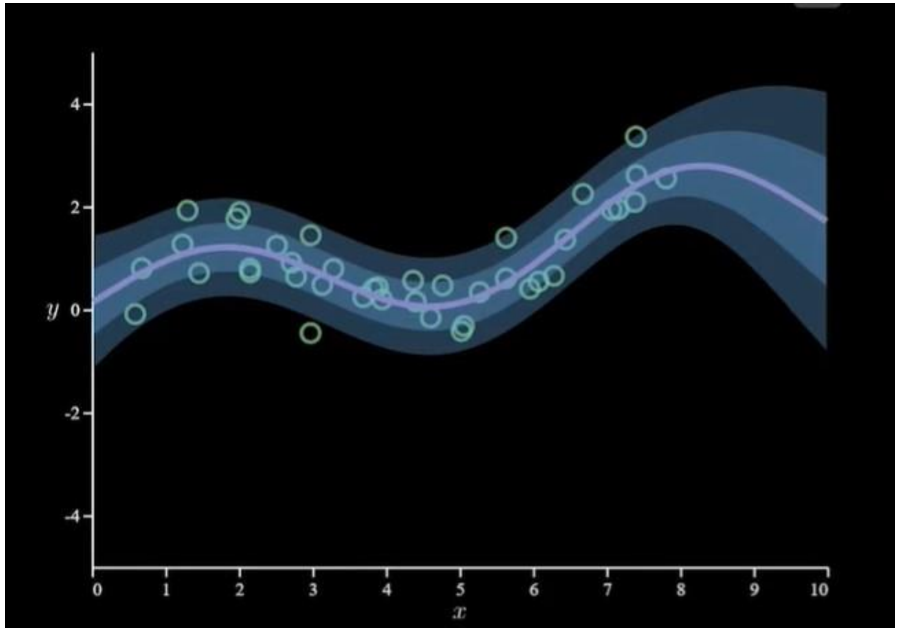

Using the checkboxes, different kernels can be combined to form a new Gaussian process. Only by using a combination of kernels, it is possible to capture the characteristics of more complex training data.

Noise variance Off

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim N(\mu, \Sigma) = N\left(0, \begin{bmatrix} K(x,x) & K(x,x_*) \\ K(x_*,x) & K(x_*,x_*) \end{bmatrix}\right)$$

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim N(\mu, \Sigma) = N\left(0, \begin{bmatrix} K(x,x) + \sigma_n^2 I & K(x,x_*) \\ K(x_*,x) & K(x_*,x_*) \end{bmatrix}\right)$$

Noise variance on



# GAUSSIAN PROCESS REGRESSION ALGORITHM

**FFT**

**BAYES**

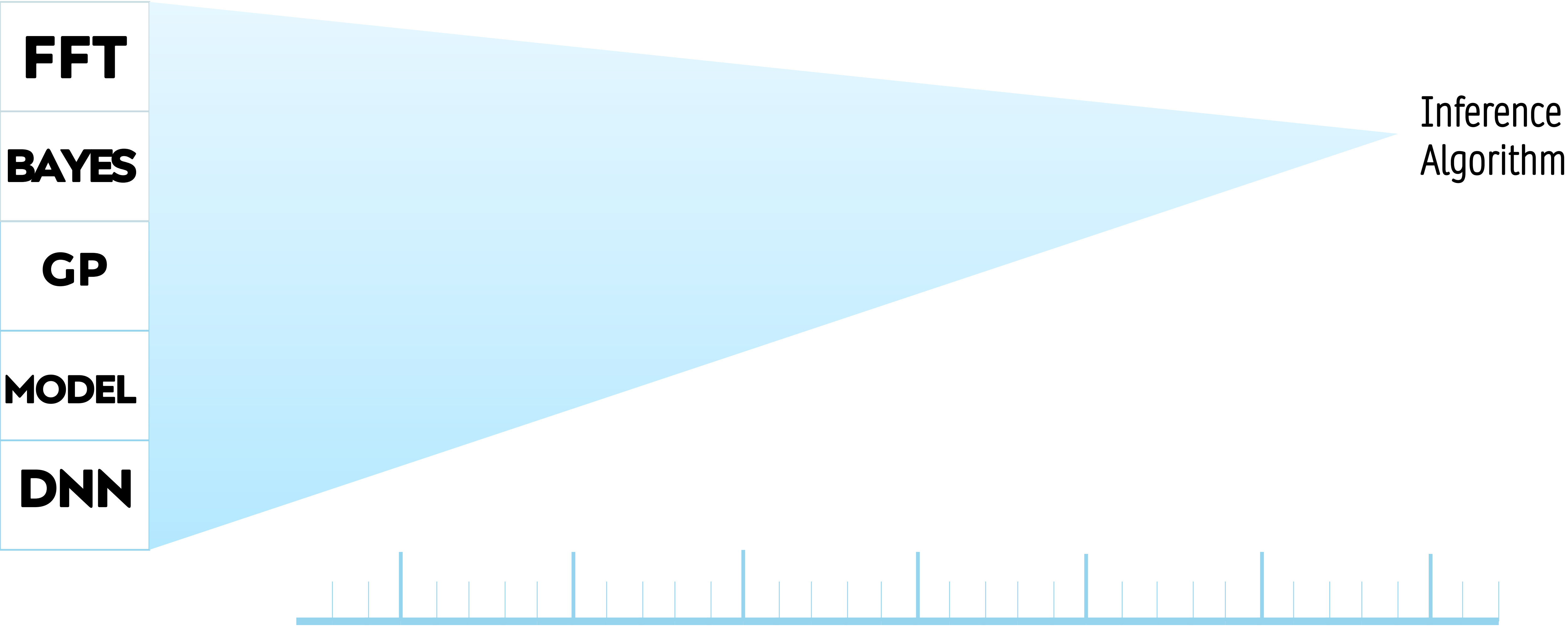
**GAUSSIAN PROCESS**

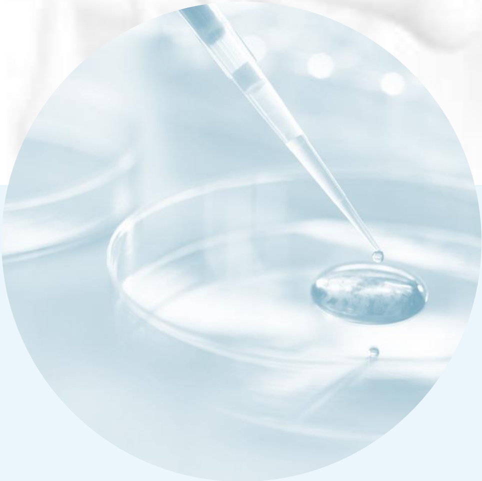
1. CHOOSE A KERNEL
2. BUILD COVARIANCE MATRIX
3. TAKE ITS INVERSE
4. MINIMIZE THE NLL $= \frac{1}{2}\mathrm{tr}\left((\alpha\alpha^T - K^{-1})\frac{\partial K}{\partial \theta_i}\right)$ where $\alpha = K^{-1}y$

Prediction : $\Pr(y_*|x_*, X, y) = N\left(\bar{f}(x_*), \acute{k}(x_*, x_*)\right)$

Where $\bar{f}(\cdot) = k(\cdot, X)(K + \sigma^2 I)^{-1} y$ *and* $\acute{k}(\cdot,\cdot) = k(\cdot, X)(K + \sigma^2 I)^{-1} k(X,\cdot)$

# WHEN IS A NOISY SIGNAL (NOISY X OSCILLATING) ?





# THANK YOU

RITESH BALAYAN
BALAYANRITESH@GMAIL.COM | S2456148@ED.AC.UK