UNIVERSITY OF CAMBRIDGE
Institute of Computing for Climate Science

Using Gaussian Processes to Uncover the Secrets of the Universe

Henry Moss @ MASCOTNUM 2024

UNIVERSITY OF CAMBRIDGE

Institute of Computing for Climate Science

Using Gaussian Processes to Uncover the Secrets of the Universe

Stochastic Equation Discovery via Interpretable Additive Models

Henry Moss @ MASCOTNUM 2024

Why do we want to learn symbolic equations?

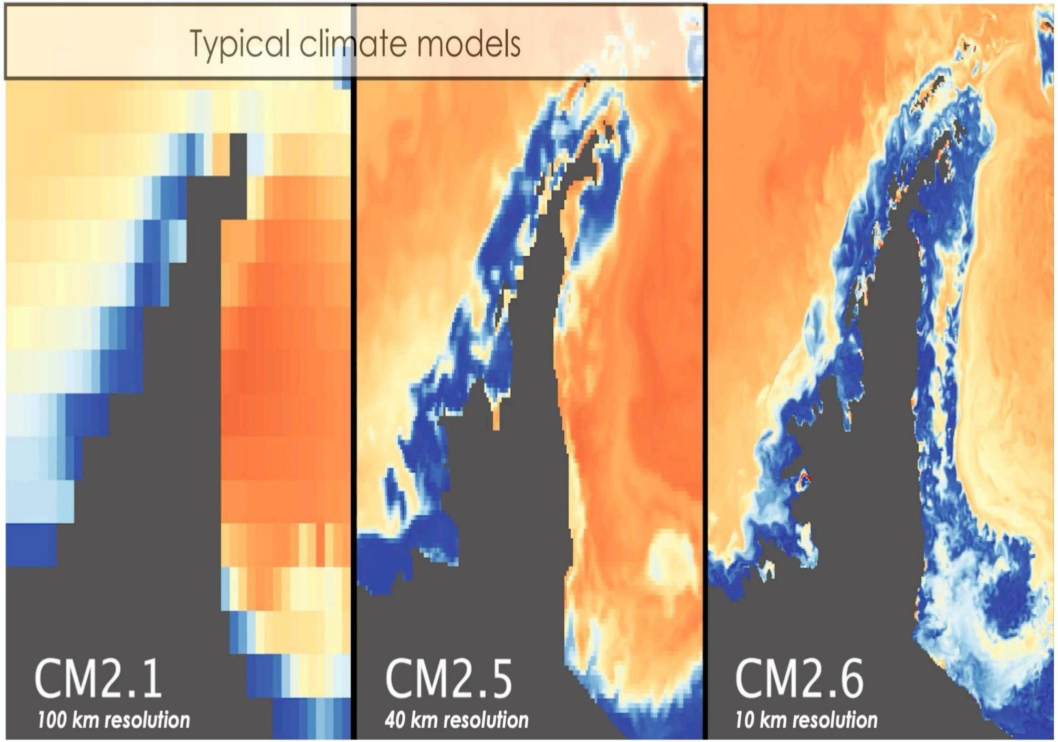
Why do we want to learn symbolic equations?

Why do we want to learn symbolic equations?

Why do we want to learn symbolic equations?

Why do we want to learn symbolic equations?

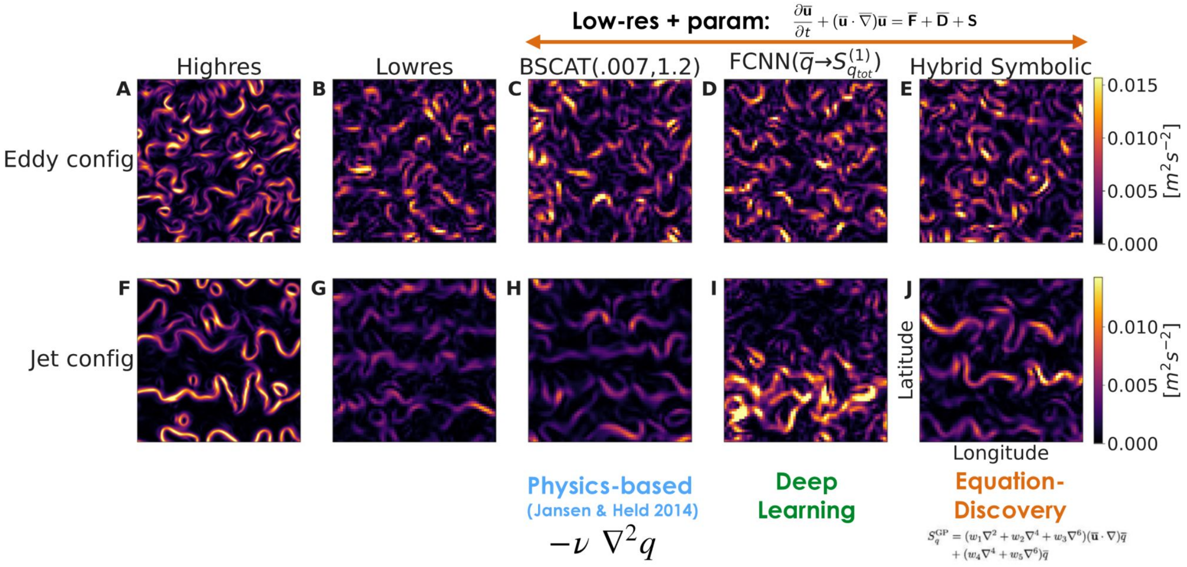
To learn parameterisations



NOAA GFDL CM2 Suite; Animation from J. Busecke

Why do we want to learn symbolic equations?

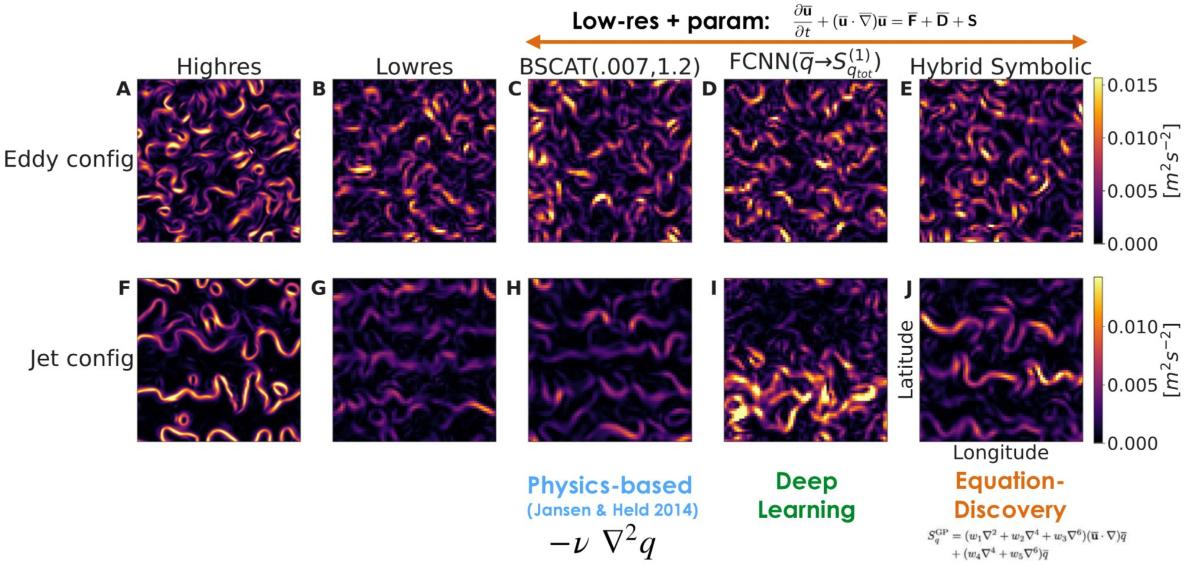
To learn parameterisations



Zanna et al 2020

Why do we want to learn symbolic equations?

To learn parameterisations



Zanna et al 2020

$$\mathcal{C}_{\text{Sundqvist}} \stackrel{\text{def}}{=} 1 - \sqrt{\frac{\min\{\text{RH}, \text{RH}_{\text{sat}}\} - \text{RH}_{\text{sat}}}{\text{RH}_0 - \text{RH}_{\text{sat}}}} \qquad \mathcal{C}_{\text{Teixeira}} \stackrel{\text{def}}{=} \frac{D q_c}{2 q_s (1 - \widehat{\text{RH}}) K} \left(-1 + \sqrt{1 + \frac{4 q_s (1 - \widehat{\text{RH}}) K}{D q_c}} \right)$$

$$f(\text{RH}, T, \partial_z \text{RH}, q_c, q_i) = I_1(\text{RH}, T) + I_2(\partial_z \text{RH}) + I_3(q_c, q_i),$$

Grundner et al. 2023

How can we learn symbolic equations?

Brunton et al. 2016

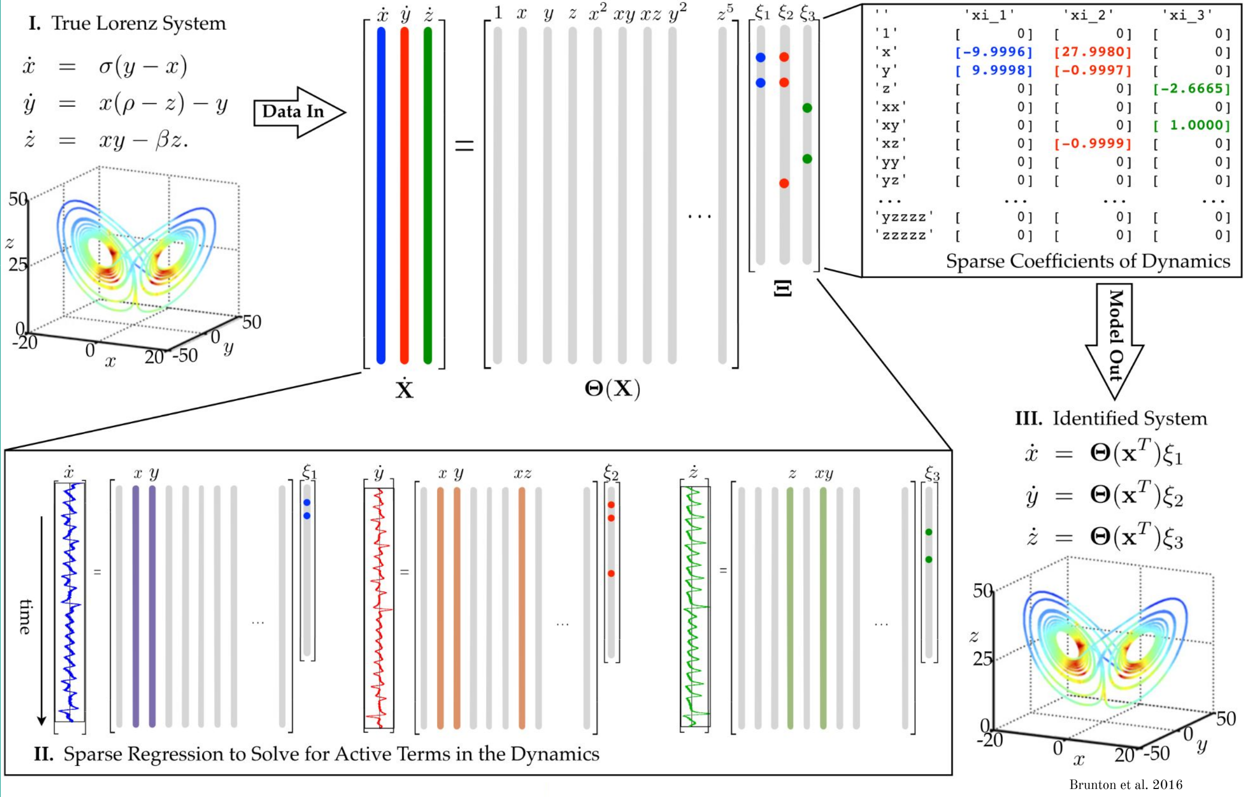
How can we learn symbolic equations?

E.g. Sparse Identification of Nonlinear Dynamics

Brunton et al. 2016

How can we learn symbolic equations?

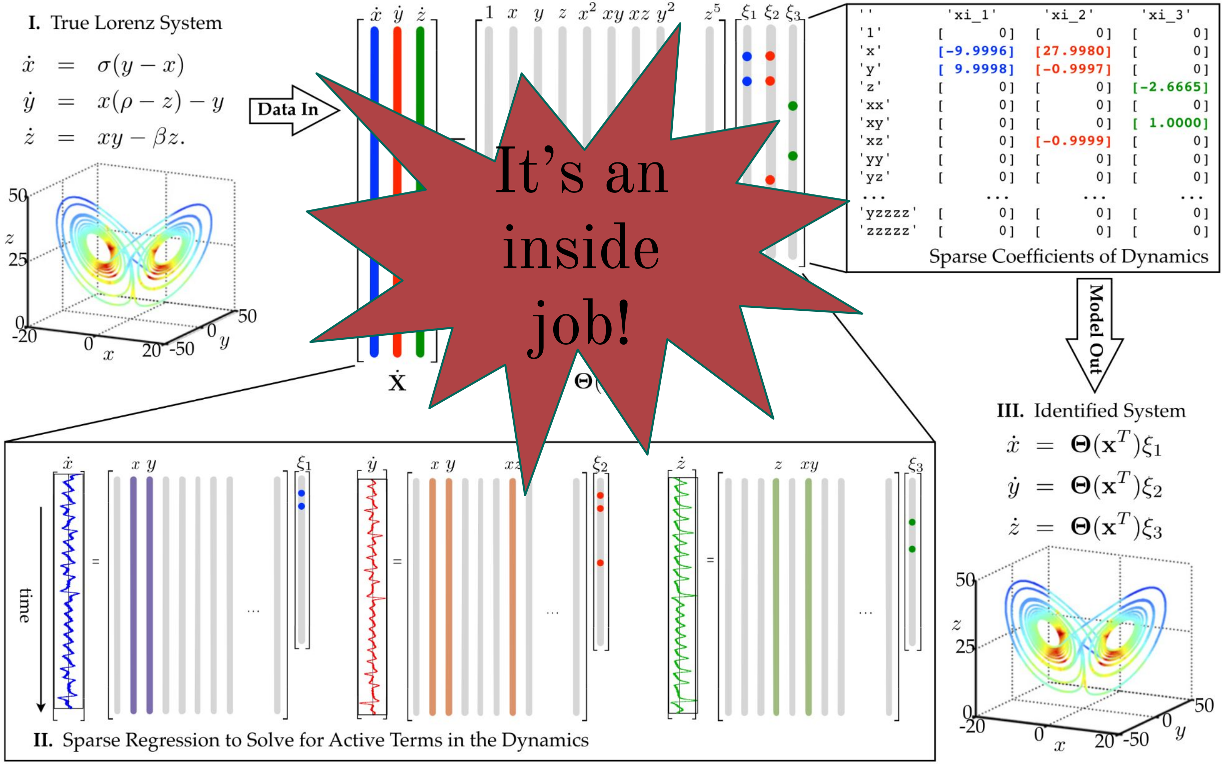
E.g. Sparse Identification of Nonlinear Dynamics



How can we learn symbolic equations?

- Functional form in advance
- Correlated inputs
- Only uncertainty over params

E.g. Sparse Identification of Nonlinear Dynamics



What do we want?

- ML to **HELP** scientists discovery equations
- Learn **STOCHASTIC** equations

Lets use Gaussian processes!

Lets use Gaussian processes!

They can’t handle large data volumes

They can’t handle large data volumes

Only for Gaussian data.......

Gaussian processes!

They can't handle high-dimensional data

They can't handle large data volumes

Gaussian processes!

Only for Gaussian data.......

They can't handle large data volumes

They can't handle high-dimensional data

Only for Gaussian data.......

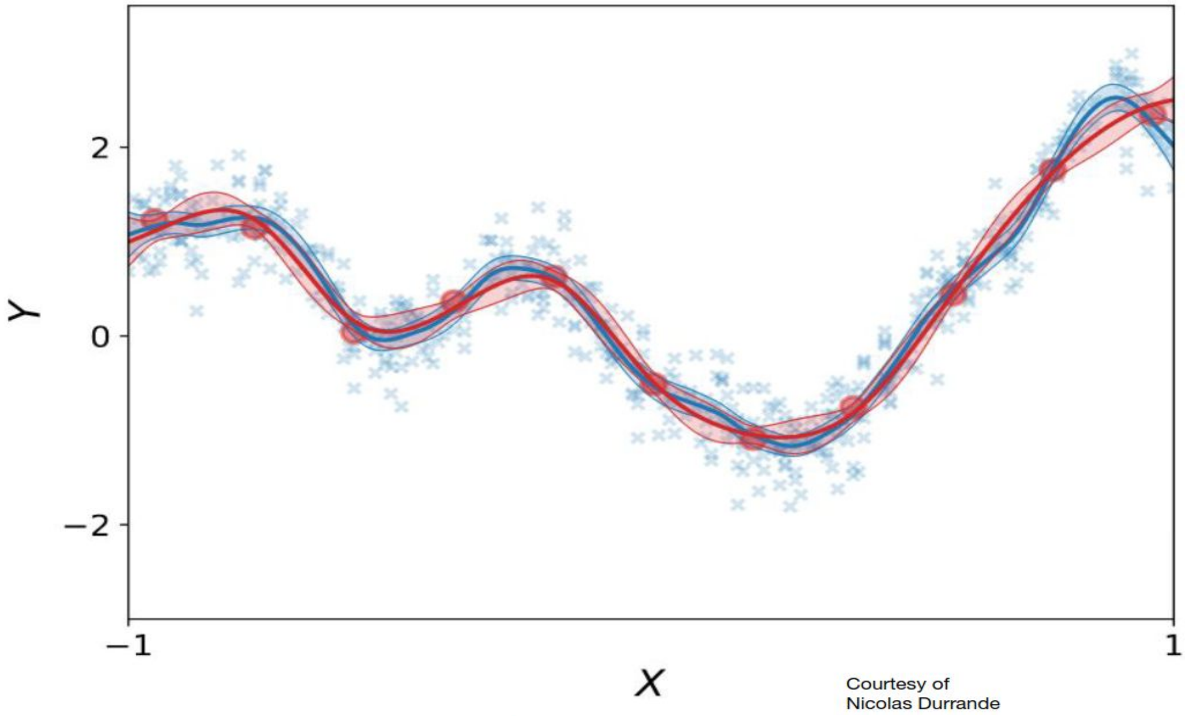
They are not interpretable (symbolic)

B … Gaus … is … esses!

GPs for big data?

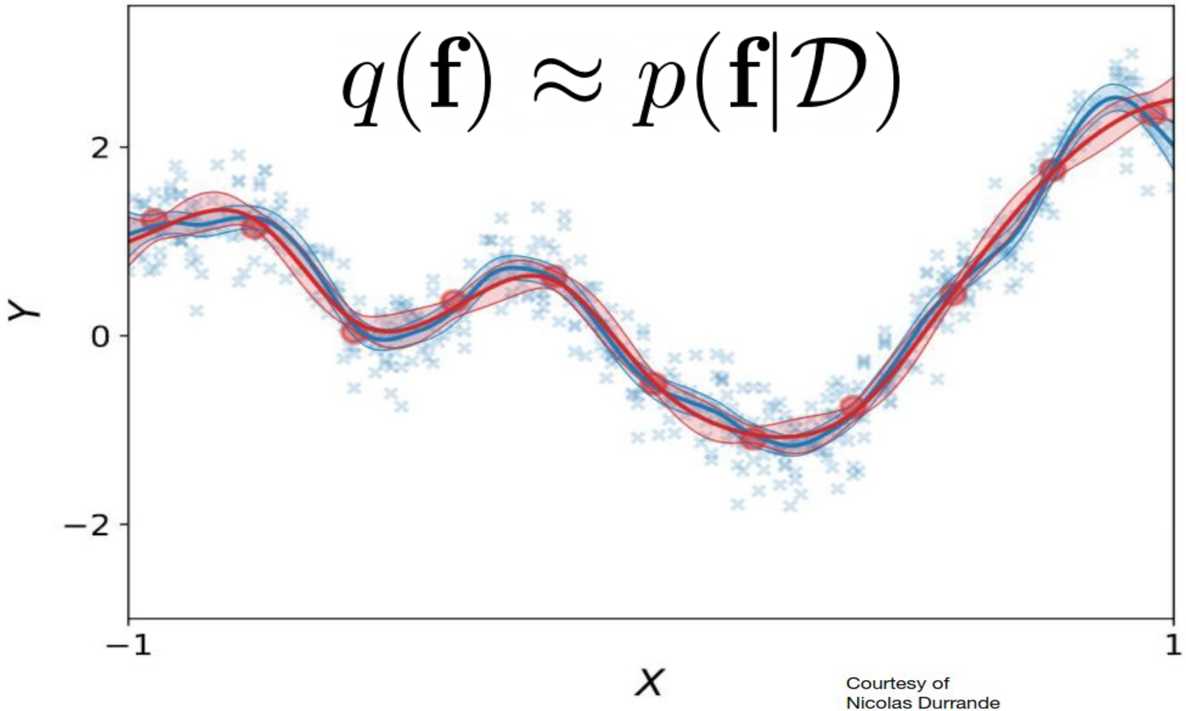
- Use Sparse variational GP
- Replace with M (<<N)

  representative points



GPs for big data?

- Use Sparse variational GP
- Replace with M (<<N) representative points



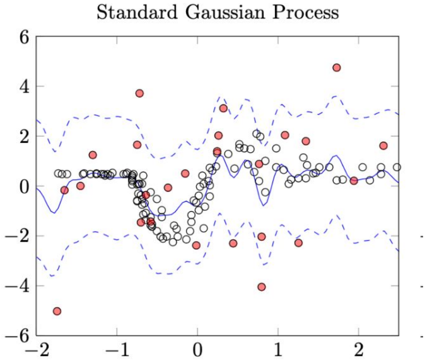
Courtesy of
Nicolas Durrande

$$\mathrm{ELBO}(q(\mathbf{f})) = \int q(\mathbf{f}) \log p(\mathbf{y}|\mathbf{f}) d\mathbf{f} - \mathcal{KL}(q(\mathbf{f}), p(\mathbf{f}))$$
$$= \sum_{i=1}^{N} \int q(f_i) \log p(y_i|f_i) d\mathbf{f} - \mathcal{KL}(q(\mathbf{f}), p(\mathbf{f}))$$

SVGPs for non-Gaussian data?

(Hensman et al. 2015, Saul et al. 2016)

$$y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2)$$

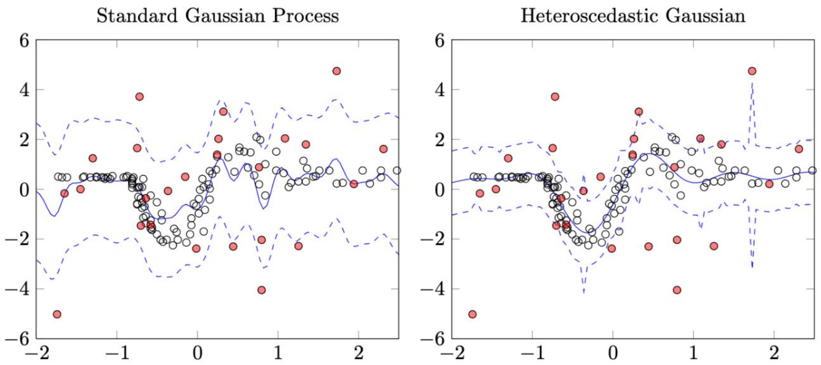


SVGPs for non-Gaussian data?

(Hensman et al. 2015, Saul et al. 2016)

$$\cancel{y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2)}$$

$$y_i \sim \mathcal{N}(f_0(\mathbf{x}_i), e^{f_1(\mathbf{x}_i)})$$



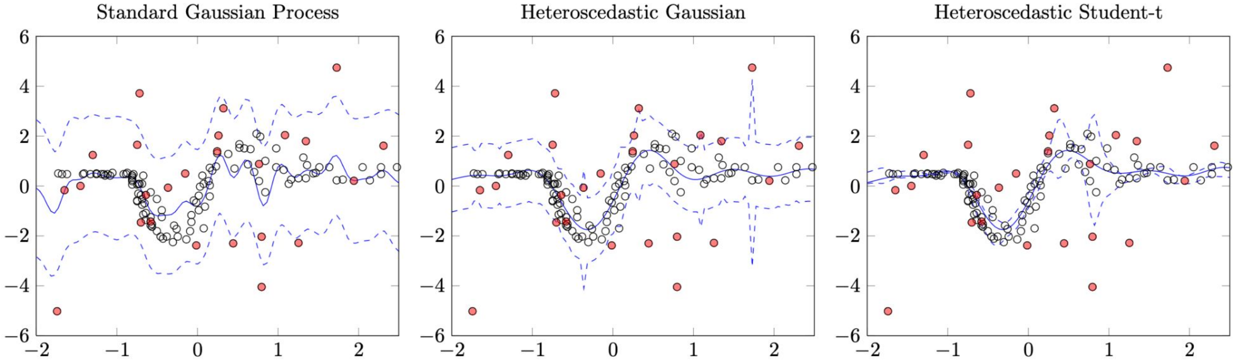
SVGPs for non-Gaussian data?

(Hensman et al. 2015, Saul et al. 2016)

$$\cancel{y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2)}$$

$$\cancel{y_i \sim \mathcal{N}(f_0(\mathbf{x}_i), e^{f_1(\mathbf{x}_i)})}$$

$$y_i \sim \mathcal{S}t(f_0(\mathbf{x}_i), e^{f_1(\mathbf{x}_i)}, \nu)$$



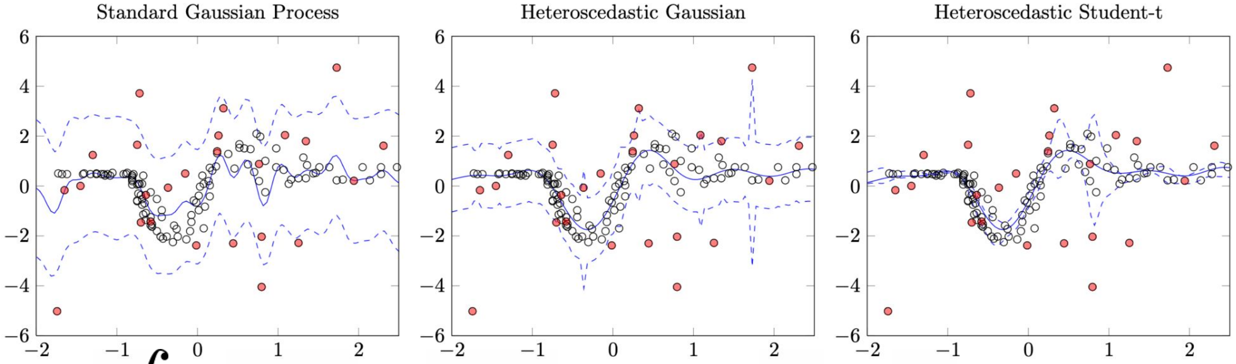
SVGPs for non-Gaussian data?

(Hensman et al. 2015, Saul et al. 2016)

$$\cancel{y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2)}$$

$$\cancel{y_i \sim \mathcal{N}(f_0(\mathbf{x}_i), e^{f_1(\mathbf{x}_i)})}$$

$$y_i \sim \mathcal{S}t(f_0(\mathbf{x}_i), e^{f_1(\mathbf{x}_i)}, \nu)$$



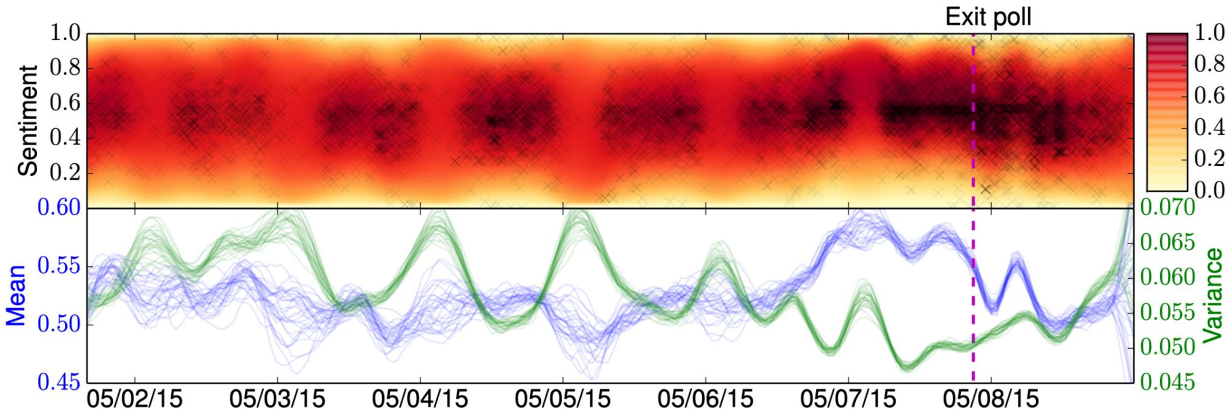
$$\mathrm{ELBO}(q(\mathbf{f}_0), q(\mathbf{f}_1)) = \int q(\mathbf{f}_0)q(\mathbf{f}_1)\log p(\mathbf{y}|\mathbf{f}_0, \mathbf{f}_1)d\mathbf{f}_0 d\mathbf{f}_1$$

$$-\mathcal{KL}(q(\mathbf{f}_0), p(\mathbf{f})) - \mathcal{KL}(q(\mathbf{f}_1), p(\mathbf{f}))$$

SVGPs for non-Gaussian data?

(Hensman et al. 2015, Saul et al. 2016)

$$y_i \sim \mathcal{B}(\alpha = f_0(\mathbf{x}_i), \beta = e^{f_1(\mathbf{x}_i)})$$



GPs for high-dim data?


Beware the curse of dimensionality ....


GPs for high-dim data?

- GPs are great in high-dim
- RBF kernels are not……
- $l_i \propto \sqrt{D}$

$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2l^2}}$$

GPs for high-dim data?

$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2l^2}}$$

$$= \prod_{i}^{d} k_i(x_i, y_i)$$

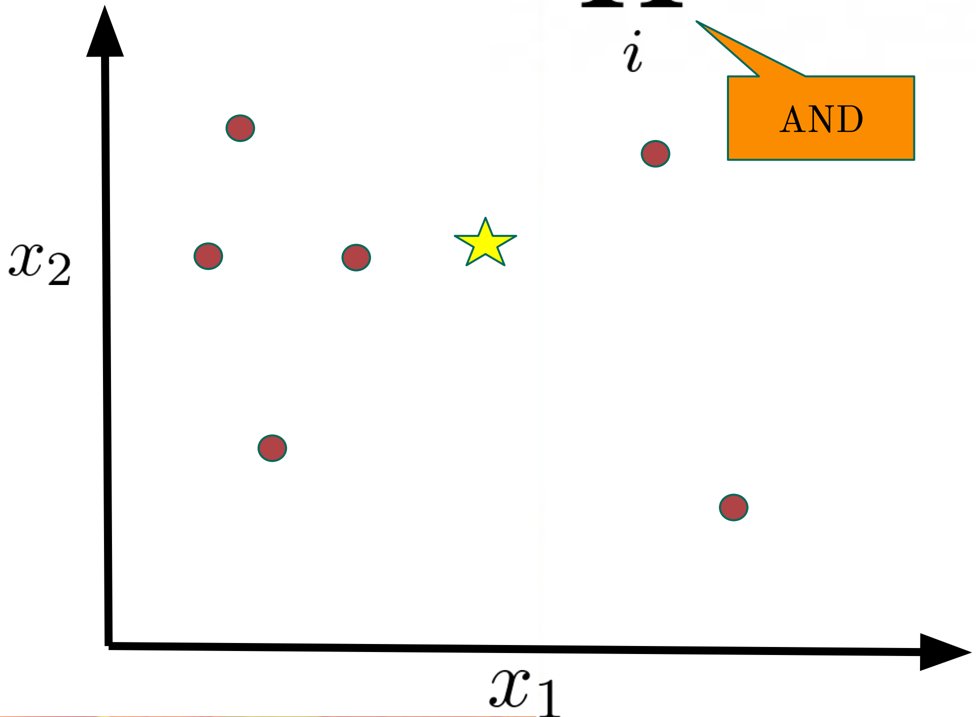
AND

GPs for high-dim data?

$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2l^2}}$$

$$= \prod_{i}^{d} k_i(x_i, y_i)$$

AND

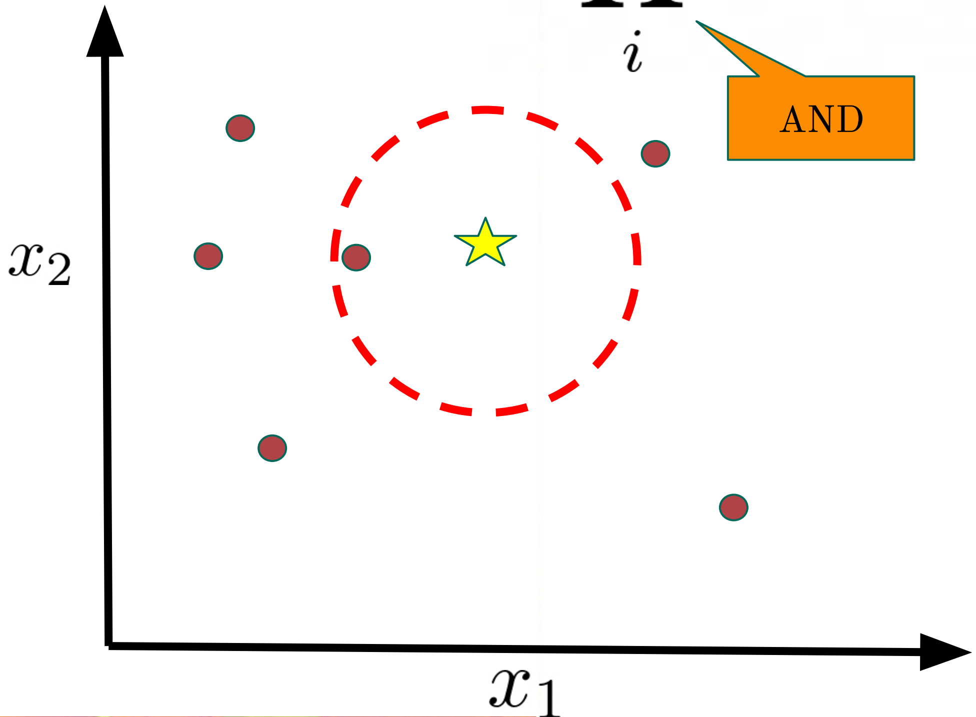


GPs for high-dim data?

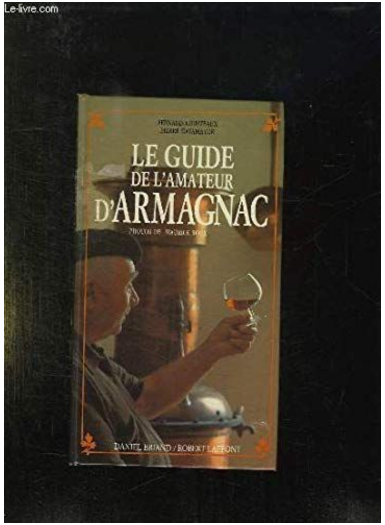
$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2l^2}}$$

$$= \prod_i^d k_i(x_i, y_i)$$

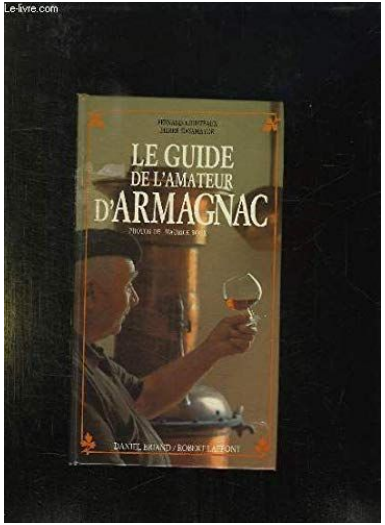
AND



GPs for high-dim data?

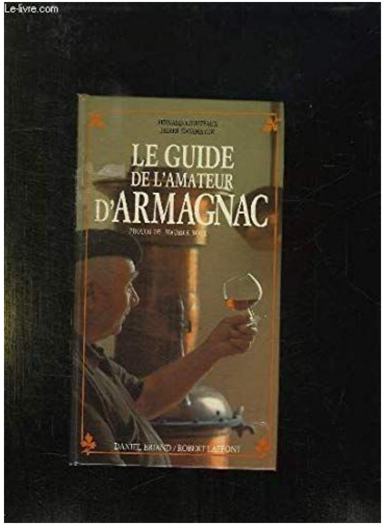


GPs for high-dim data?



GPs for high-dim data?

- Type of still (column/pot?)
- Type of grape (Ugni Blanc?)
- Wood for the barrel
- Location (Armagnac-Ténarèze, Bas-Armagnac ,Haut-Armagnac?)
- Blend
- Age



GPs for high-dim data?

$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2l^2}}$$

$$= \prod_{i}^{d} k_i(x_i, y_i)$$

AND

GPs for high-dim data?

$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2l^2}}$$

$$= \prod_i^d k_i(x_i, y_i)$$

AND

$$k_1(\mathbf{x}, \mathbf{y}) = \sum_i^d k_i(x_i, y_i)$$

OR

$$k_2(\mathbf{x}, \mathbf{y}) = \sum_{i<j}^d k_i(x_i, y_i) k_j(x_j, y_j)$$

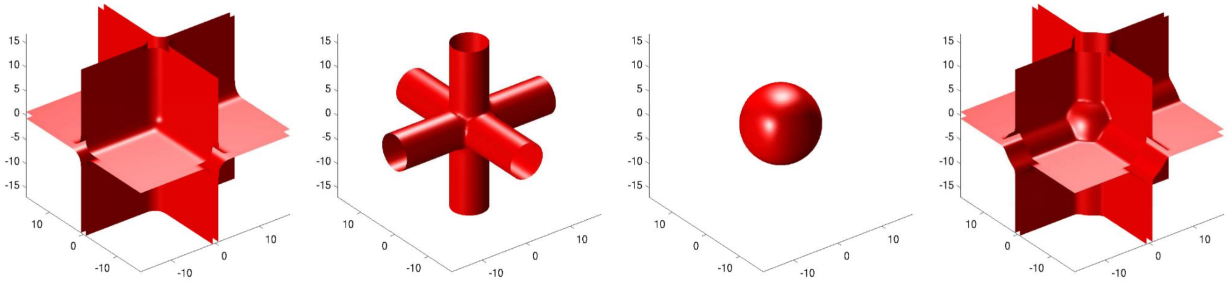
Additive Gaussian Processes

$$k(x, y) = \overset{0}{k_0} + \overset{1}{\sum k_i(x_i, y_i)} + \overset{2}{\sum_{i<j} k_i(x_i, y_i) k_j(x_j . y_j)}$$

Additive Gaussian Processes

$$k(x, y) = \overset{0}{k_0} + \overset{1}{\sum k_i(x_i, y_i)} + \overset{2}{\sum_{i<j} k_i(x_i, y_i) k_j(x_j . y_j)}$$

(Duvenaud et al 2011)



1st order interactions
$k_1 + k_2 + k_3$

2nd order interactions
$k_1 k_2 + k_2 k_3 + k_1 k_3$

3rd order interactions
$k_1 k_2 k_3$
(Squared-exp kernel)

All interactions

(Additive kernel)

Additive Gaussian Processes

$$k(x,y) = \underbrace{k_0}_{0} + \underbrace{\sum k_i(x_i, y_i)}_{1} + \underbrace{\sum_{i<j} k_i(x_i, y_i) k_j(x_j . y_j)}_{2}$$

$$\Updownarrow$$

Ginsbourger et al. (2016)

$$f(\mathbf{x}) = f_0 + \sum f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j)$$

Additive Gaussian Processes

0 1 2

$$k(x,y) = k_0 + \sum k_i(x_i, y_i) + \sum_{i<j} k_i(x_i, y_i) k_j(x_j . y_j)$$

$$\Updownarrow$$

$$f(\mathbf{x}) = f_0 + \sum f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j)$$

- Standard RBF -> $O(d(N^2 + NM))$
- d additive RBF -> $O(2^d(N^2 + NM))$

Additive Gaussian Processes

- Newton Girard (Duvenaud et al 2011)

$$k(x,y) = \overset{0}{k_0} + \overset{1}{\sum k_i(x_i,y_i)} + \overset{2}{\sum_{i<j} k_i(x_i,y_i)k_j(x_j.y_j)}$$

$$\Updownarrow$$

$$f(\mathbf{x}) = f_0 + \sum f_i(x_i) + \sum_{i<j} f_{ij}(x_i,x_j)$$

- Standard RBF -> $O(d(N^2 + NM))$
- d additive RBF -> $O(2^d(N^2 + NM))$
- d additive BBF (NG) -> $O(d^2(N^2 + NM))$

Additive Gaussian Processes

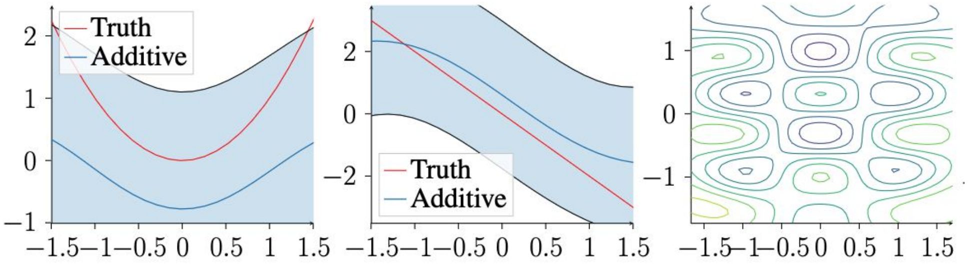
$$k(x, y) = k_0 + \sum k_i(x_i, y_i) + \sum_{i<j} k_i(x_i, y_i) k_j(x_j . y_j)$$

$$\Updownarrow$$

Ginsbourger et al. (2016)

$$f(\mathbf{x}) = f_0 + \sum f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j)$$

$$f(x_1, x_2) = x_1^2 - 2x_2 + cos(3x_1)sin(5x_2)$$



(a) f_1 (b) f_2 (c) Interaction

$$E[f_i(x_i)|\mathcal{D}] = k_i(x_i, X)K(X, X)^{-1}\mathbf{y}$$

Lu et al. 2022

$$k(x,y) = \overset{0}{k_0} + \overset{1}{\sum k_i(x_i, y_i)} + \overset{2}{\sum_{i<j} k_i(x_i, y_i) k_j(x_j . y_j)}$$

$$\Updownarrow$$

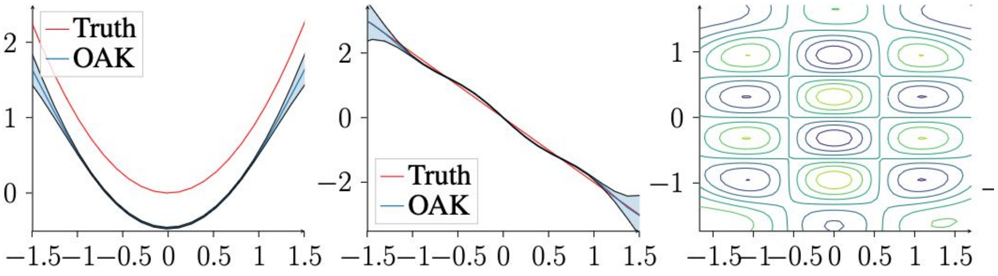
$$f(\mathbf{x}) = f_0 + \sum f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j)$$

Additive Gaussian Processes

- Orthogonalise (Durrande et al 2012)

$$f(x_1, x_2) = (f_1(x_1) + \delta) + (f_2(x_2) - \delta)$$

$$f(x_1, x_2) = x_1^2 - 2x_2 + cos(3x_1)sin(5x_2)$$



(f) f_1 (g) f_2 (h) Interaction

Lu et al. 2022

Additive Gaussian Processes

- Orthogonalise (Durrande et al 2012)

$$f(x_1, x_2) = (f_1(x_1) + \delta) + (f_2(x_2) - \delta)$$

- By conditioning

$$f_i(x_i) \Big| \int f_i(x_i) p(x_i) dx_i = 0$$

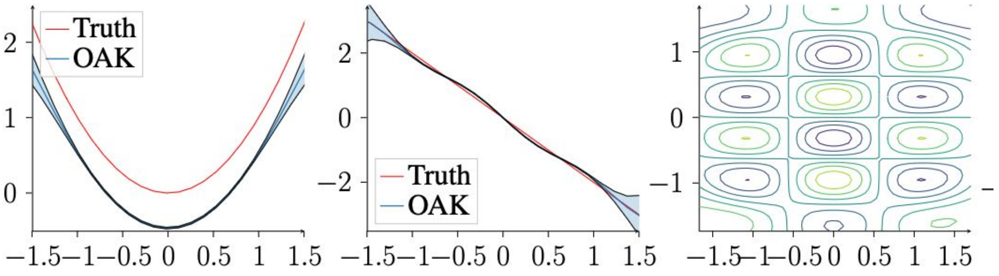
0 1 2

$$k(x, y) = k_0 + \sum k_i(x_i, y_i) + \sum_{i<j} k_i(x_i, y_i) k_j(x_j . y_j)$$

$$\Updownarrow$$

$$f(\mathbf{x}) = f_0 + \sum f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j)$$

$$f(x_1, x_2) = x_1^2 - 2x_2 + cos(3x_1) sin(5x_2)$$



(f) f_1 (g) f_2 (h) Interaction

Lu et al. 2022

Additive Gaussian Processes

- Orthogonalise (Durrande et al 2012)

$$f(x_1, x_2) = (f_1(x_1) + \delta) + (f_2(x_2) - \delta)$$

- By conditioning

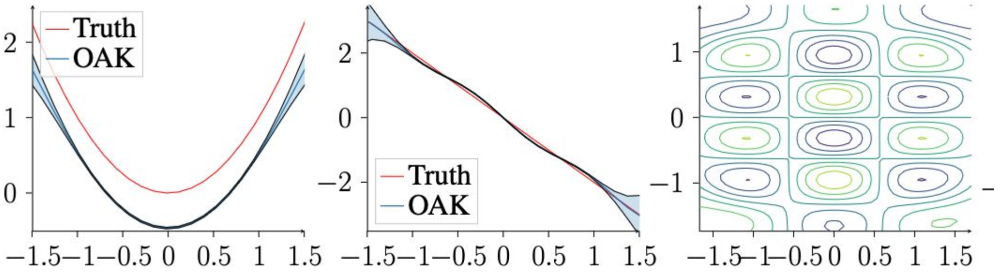
$$f_i(x_i) \Big| \int f_i(x_i) p(x_i) dx_i = 0$$

$k(x, y$ ... $\cdot y_j)$

This model is quite interpretable......

$$f(\mathbf{x}) = f_0 + \sum f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j)$$

$$f(x_1, x_2) = x_1^2 - 2x_2 + cos(3x_1)sin(5x_2)$$



(f) f_1 (g) f_2 (h) Interaction

Lu et al. 2022

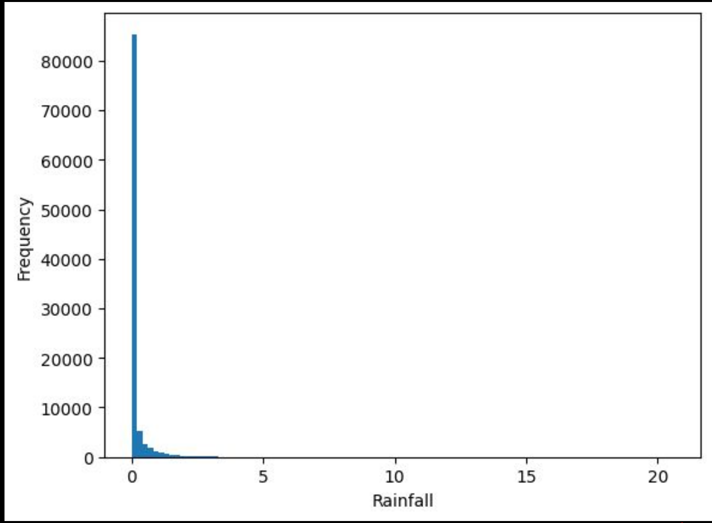
So, lets learn an equation

Predicting rainfall

- >100 climate variables –> rainfall

Predicting rainfall

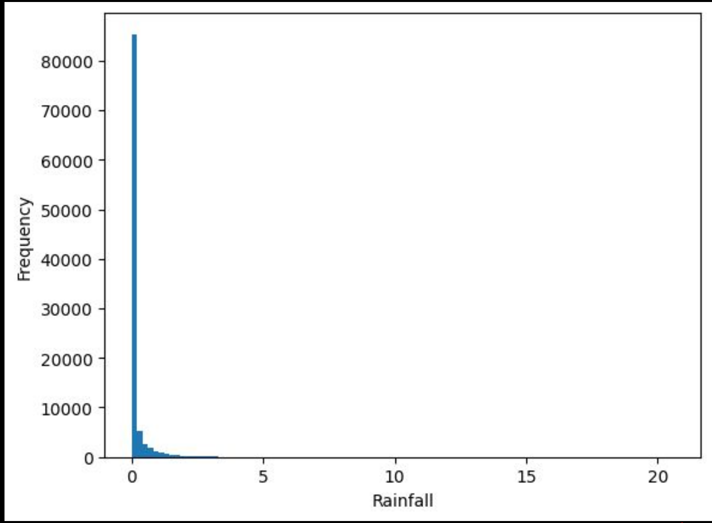
- >100 climate variables –> rainfall
- Non-Gaussian (Bernoulli-gamma)



$$\cancel{p(y|f) = \mathcal{N}(f, \sigma^2)}$$

Predicting rainfall

- >100 climate variables –> rainfall
- Non-Gaussian (Bernoulli-gamma)



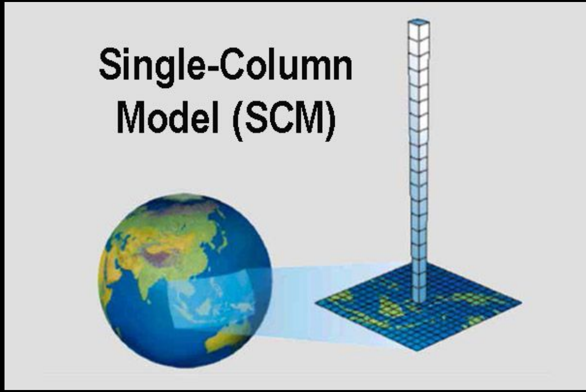
$$\cancel{p(y|f) = \mathcal{N}(f, \sigma^2)}$$

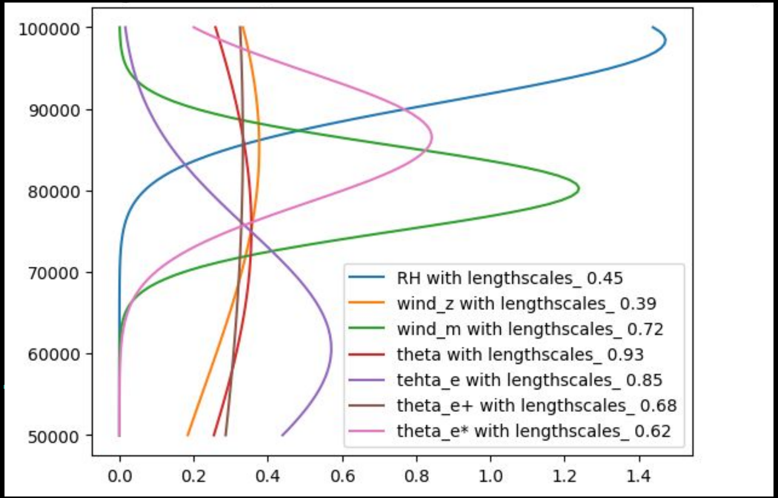
$$p(\underline{y|f_1, f_2}, f_3) = \begin{cases} 1 - f_1 & if \quad y = 0 \\ f_1 \Gamma(f_2, f_3) & o.w. \end{cases}$$

Predicting rainfall

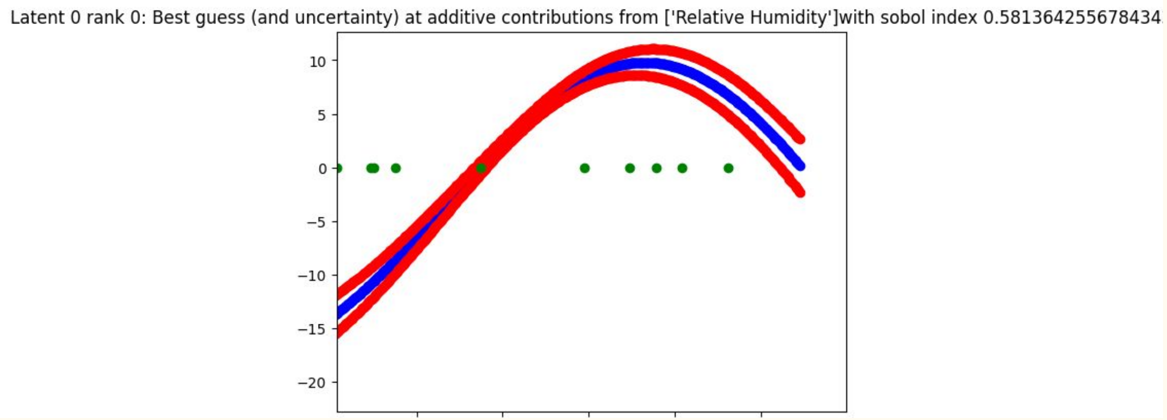
- >100 climate variables –> rainfall
- Non-Gaussian (Bernoulli-gamma)
- Data-driven vertical integration

https://e3sm.org/single-column-model-intercomparison-of-diurnal-cycle-of-precipitation/

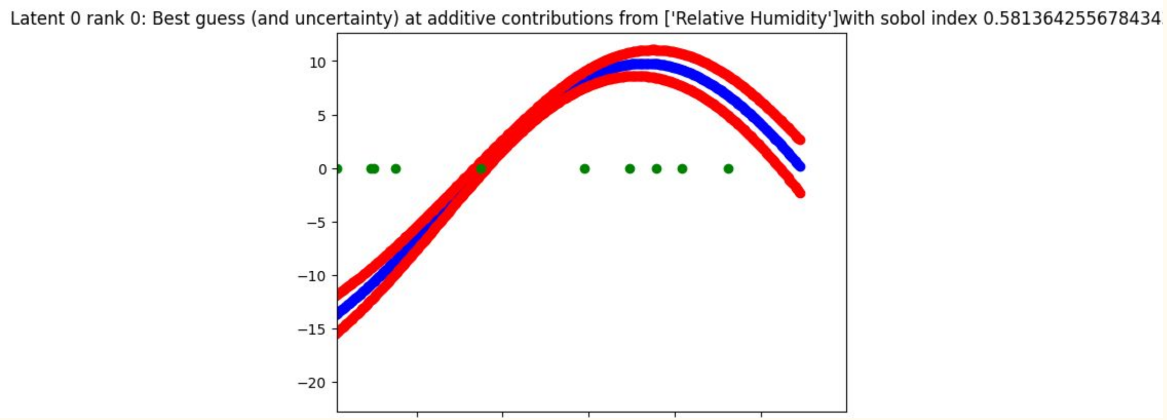


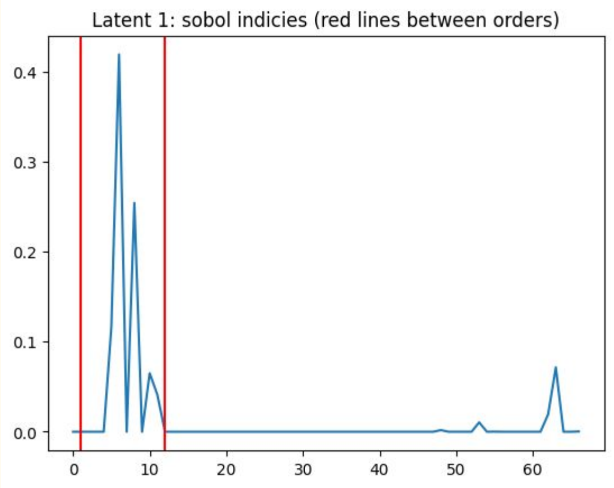


Additive GP model output

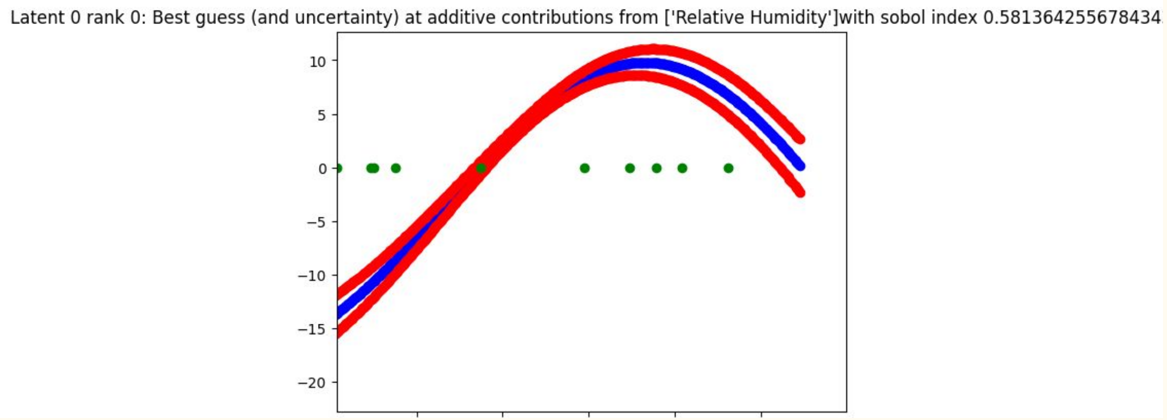


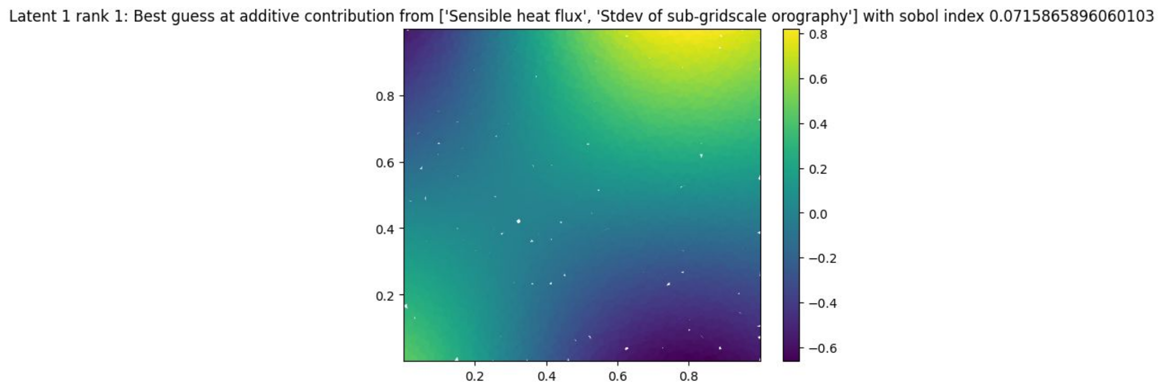
Additive GP model output

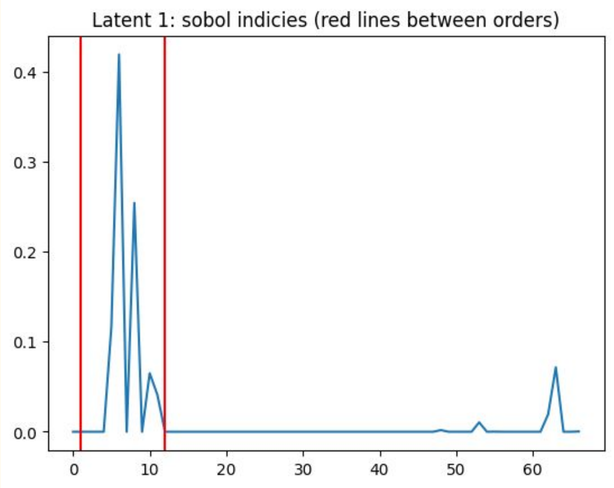




Additive GP model output







Learn a Stochastic Eq (via lots of easy SRs)

$$p(y|f_1, f_2, f_3) = \begin{cases} 1 - f_1 & if \quad y = 0 \\ f_1 \Gamma(f_2, f_3) & o.w. \end{cases}$$

- $f_1 = e^{\lambda_0 + \lambda_1 RH - \lambda_2 RH \sigma_0}$
- $f_2 = \lambda_3 + \lambda_4 (SHF - \lambda_5)^2$
- $f_3 = \lambda_6 + \lambda_7 \theta_+$

Just illustrative results

Learn a Stochastic Eq (via lots of easy SRs)

$$p(y|f_1, f_2, f_3) = \begin{cases} 1 - f_1 & if \quad y = 0 \\ f_1 \Gamma(f_2, f_3) & o.w. \end{cases}$$

$$E[y] = e^{\lambda_0 + \lambda_1 RH - \lambda_2 RH * \sigma_0} \frac{(SHF - \lambda_3)^2}{\lambda_4 + \lambda_5 \theta_+}$$

Just illustrative results

Whats next .....

- Gravity waves / cloud cover
- **Learn the likelihood** structure (another layer of symbolic regression)
- Improve “orthogonality” for correlated inputs
- **Sample** multiple candidate equations (Pareto front?)
- More user interaction
- Encode **known physics** (symmetries, invariances, conservation laws e.t.c)

La **presqu'île de Blackpool**

Thanks for your time!

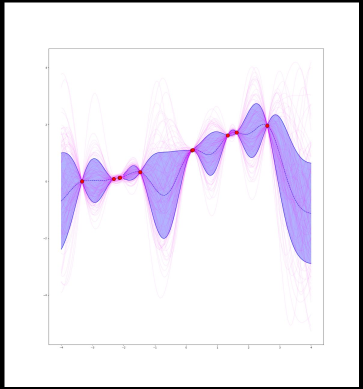
Extra slides

Scientific priors via conditioning

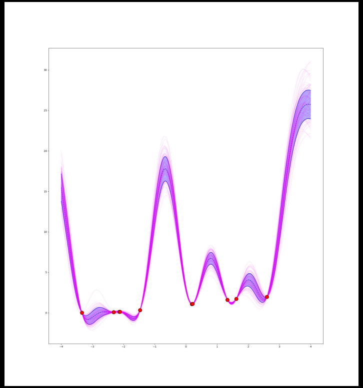
Condition on an integral

$$\mathcal{O}(f) = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

$\mathbb{P}(f)$



$\mathbb{P}(\mathbf{f}|\mathcal{O})$

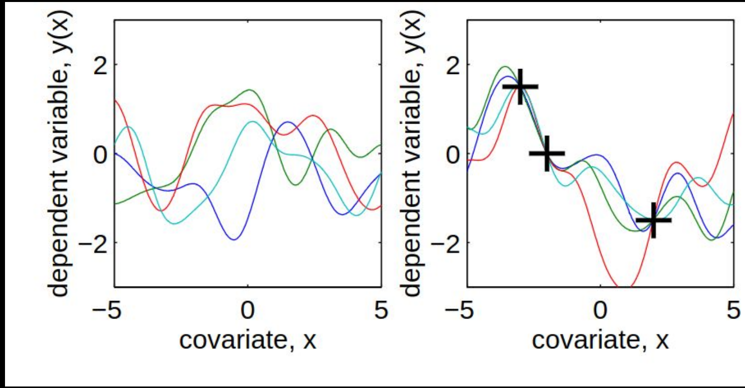


Scientific priors via conditioning

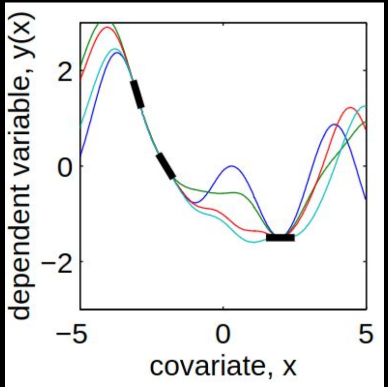
Condition on an derivative

$$\mathcal{O}(\mathbf{f}) = \frac{\partial f}{\partial \mathbf{x}}$$

$\mathbb{P}(f)$



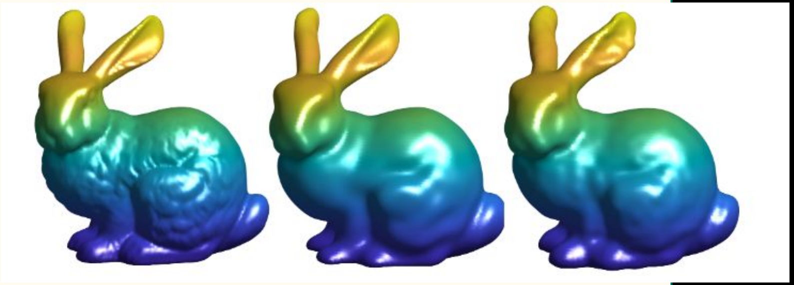
$\mathbb{P}(\mathbf{f}|\mathcal{O})$



Solak et al. (2002)

Scientific priors via conditioning

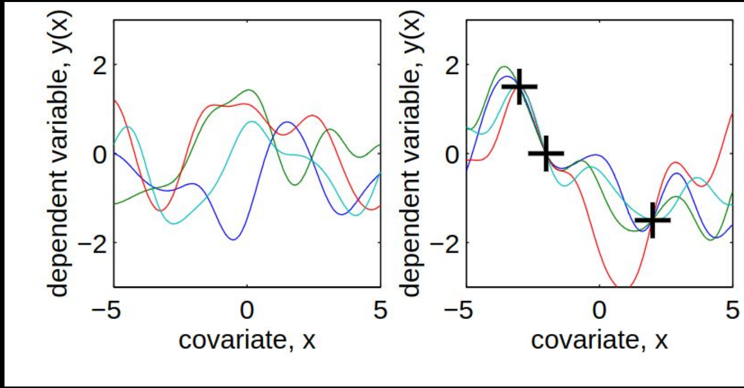
Condition on an derivative



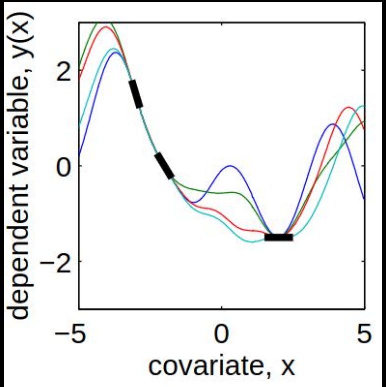
Padidar et al. (2021)

$$O(\mathbf{f}) = \frac{\partial f}{\partial \mathbf{x}}$$

$\mathbb{P}(f)$



$\mathbb{P}(\mathbf{f}|O)$



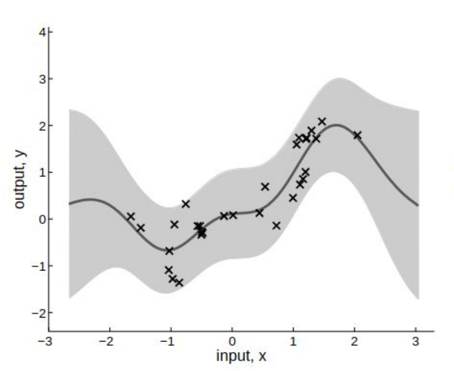
Solak et al. (2002)

Scientific priors via conditioning

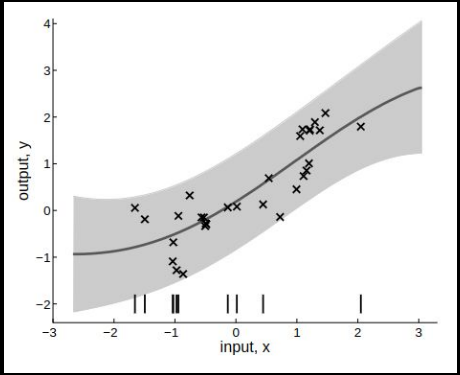
Condition on monotonicity

$$O(\mathbf{f}) = \left(\frac{\partial f}{\partial \mathbf{x}} > 0\right)$$

$\mathbb{P}(f)$



$\mathbb{P}(\mathbf{f} | O)$



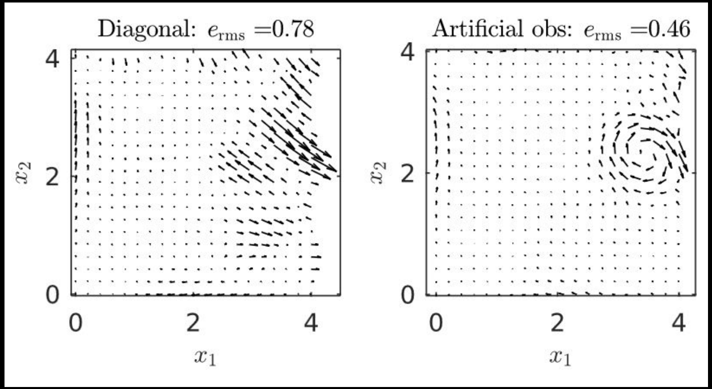
Solak et al. (2002)

Scientific priors via conditioning

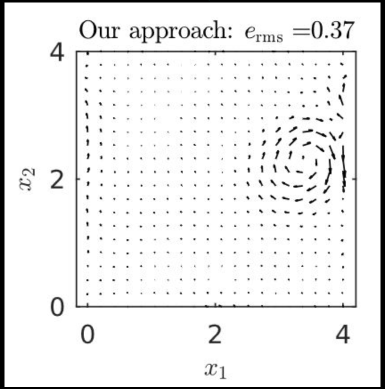
Condition on linear operator

$$O(f) = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}$$

$\mathbb{P}(f)$



$\mathbb{P}(\mathbf{f}|O)$



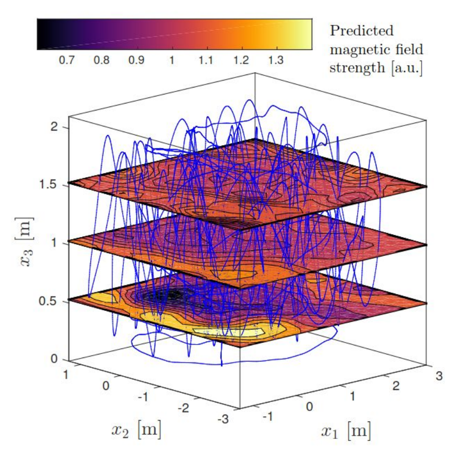
Jidling et al. (2019)

Scientific priors via conditioning

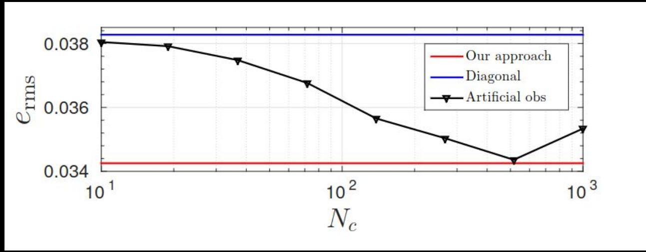
Condition on linear operator

$$O(\mathbf{f}) = \nabla \times \mathbf{f}$$

$$\mathbb{P}(f)$$



$$\mathbb{P}(\mathbf{f}|O)$$



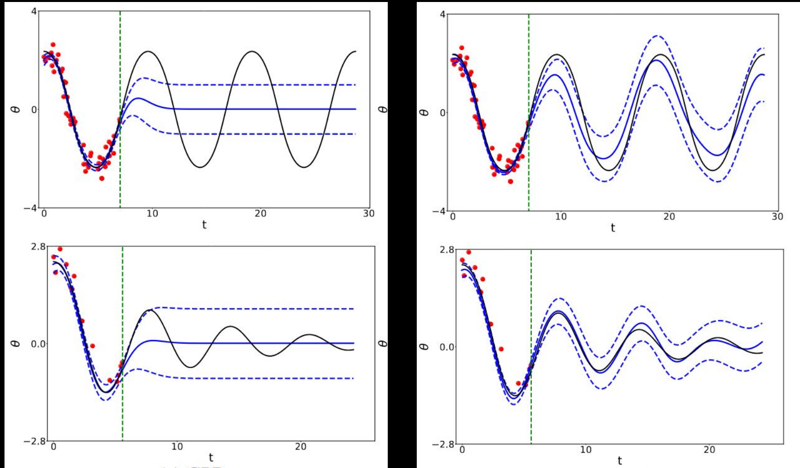
Jidling et al. (2019)

Scientific priors via conditioning

$$\frac{df^2}{dt} + \sin(t) + \beta \frac{df}{dt} = 0$$

Condition on whatever you want and pretend its Gaussian

$$\mathbb{P}(f|D) \propto \mathbb{P}(D|f)\mathbb{P}(f)\mathbb{P}(O(f))$$



(a) GPR (c) AutoIP-C

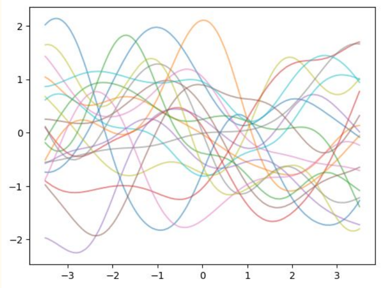
Long et al. 2022

Two ways to be encode info into GPs

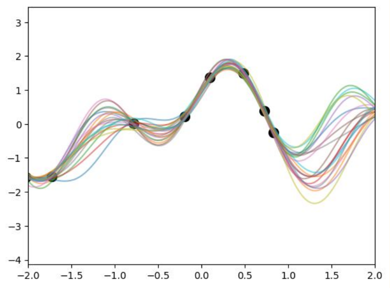
See "The kernel cookbook"
https://www.cs.toronto.edu/~duvenaud/cookbook/

1) Additional conditioning

$\mathbb{P}(f)$



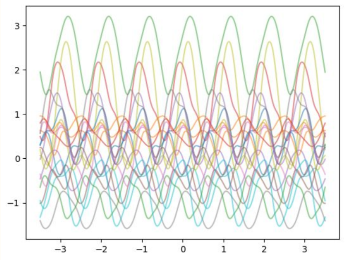
$\mathbb{P}(f|\mathcal{D})$



2) Thinking hard ....

- I want f to be periodic
- So choose a periodic kernel

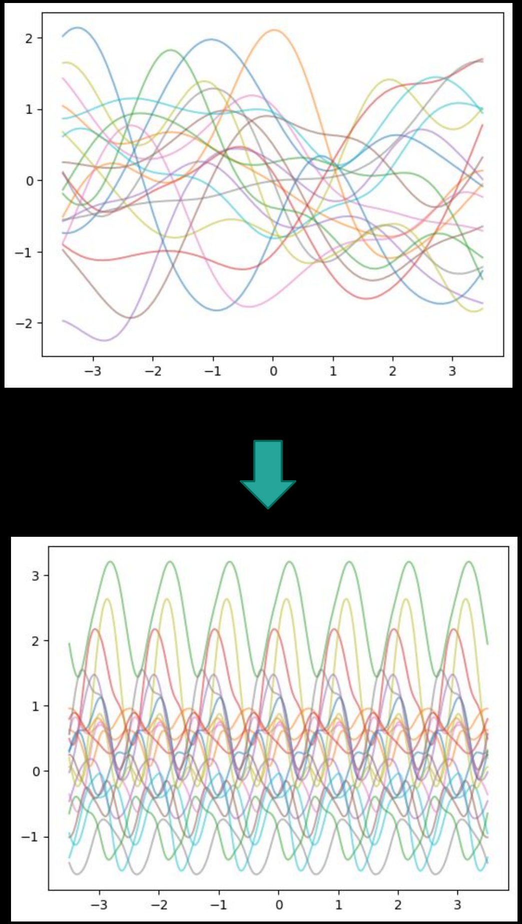
$$k_{per}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(\frac{-2\sin^2(\pi|\mathbf{x}-\mathbf{x}'|/p)}{l^2}\right)$$



Scientific priors via kernel design

Fiddle with the kernel to get periodicity

$$k_{per}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(\frac{-2\sin^2(\pi|\mathbf{x} - \mathbf{x}'|/p)}{l^2}\right)$$



Scientific priors via kernel design

General idea:

$$T(f) = f \Leftrightarrow T(k(\mathbf{x}, .)) = k(\mathbf{x}, .)$$

$$\hat{k}(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') + k(T(\mathbf{x}), \mathbf{x}')$$

Ginsbourger et al. 2013
Van der Wilk et al. 2018