Extending and Applying the Gaussian Process Autoregressive Regression Model

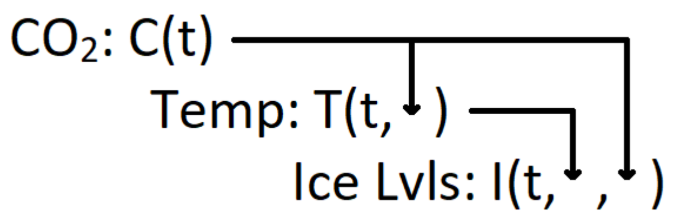
Justin Bunker, Dr. Richard E. Turner

UNIVERSITY OF CAMBRIDGE

Introduction

The process of modelling systems with multiple outputs has a lot of practical value.

Example:
climate modelling



Our goal is to incorporate the output relationships into the modelling process to improve predictions [1].

GPAR Model

Suppose we have a set of functions that have the following relationships:

$y_1(x) = f_1(x)$
$y_2(x) = f_2(x, y_1(x))$
…
$y_n(x) = f_n(x, y_1(x), \dots, y_{n-1}(x))$

$$p(y_1(x), y_2(x), \dots, y_n(x)) = \prod_{i=1}^{n} p(y_i(x) \mid y_1(x), \dots, y_{i-1}(x))$$

Training

1. Find an appropriate ordering of each of the functions using a greedy approach.
2. Train each GP with inputs composed of available observations and outputs from foregoing GPs.

Prediction

$\mathbf{y_1} \leftarrow GP_1 + \mathbf{x}$
$\mathbf{y_2} \leftarrow GP_2 + \mathbf{x} + \mathbf{y_1}$
…
$\mathbf{y_n} \leftarrow GP_n + \mathbf{x} + \mathbf{y_1} + \dots + \mathbf{y_{n-1}}$

Model for output *i*

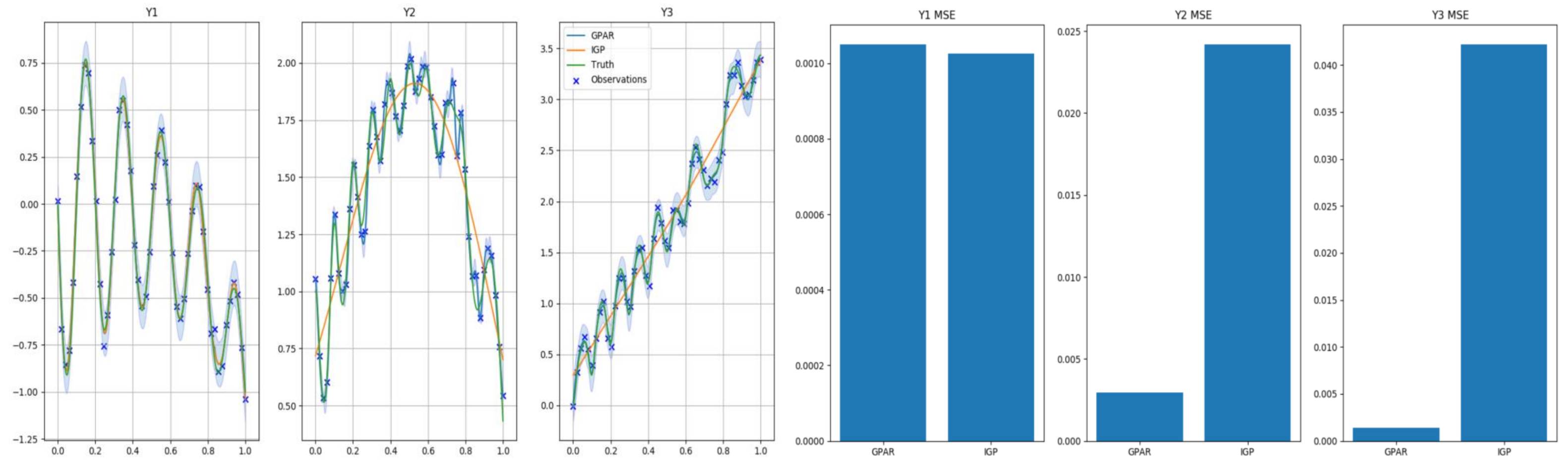
$GP_i(0, K)$ with $K: \Omega \times \Omega \rightarrow \mathbb{R}$,
where $[x, y_1, \dots, y_{i-1}]^T \in \Omega$

Challenges

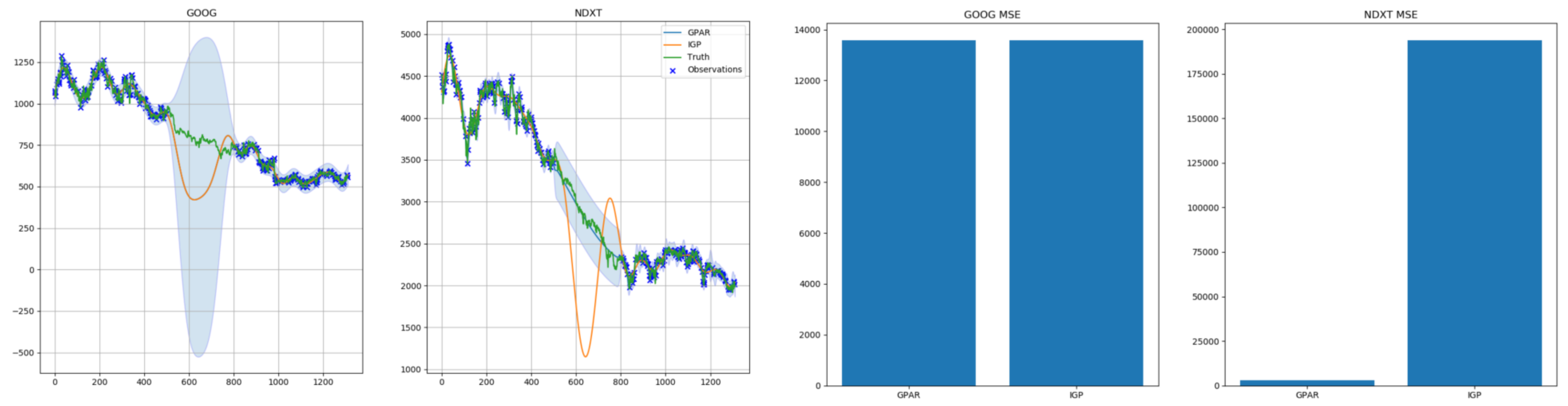
- Using an appropriate kernel.
- Finding an optimal ordering.
- Dealing with missing data and noisy inputs.
- Restricting the number of hyperparameters.

Experiments

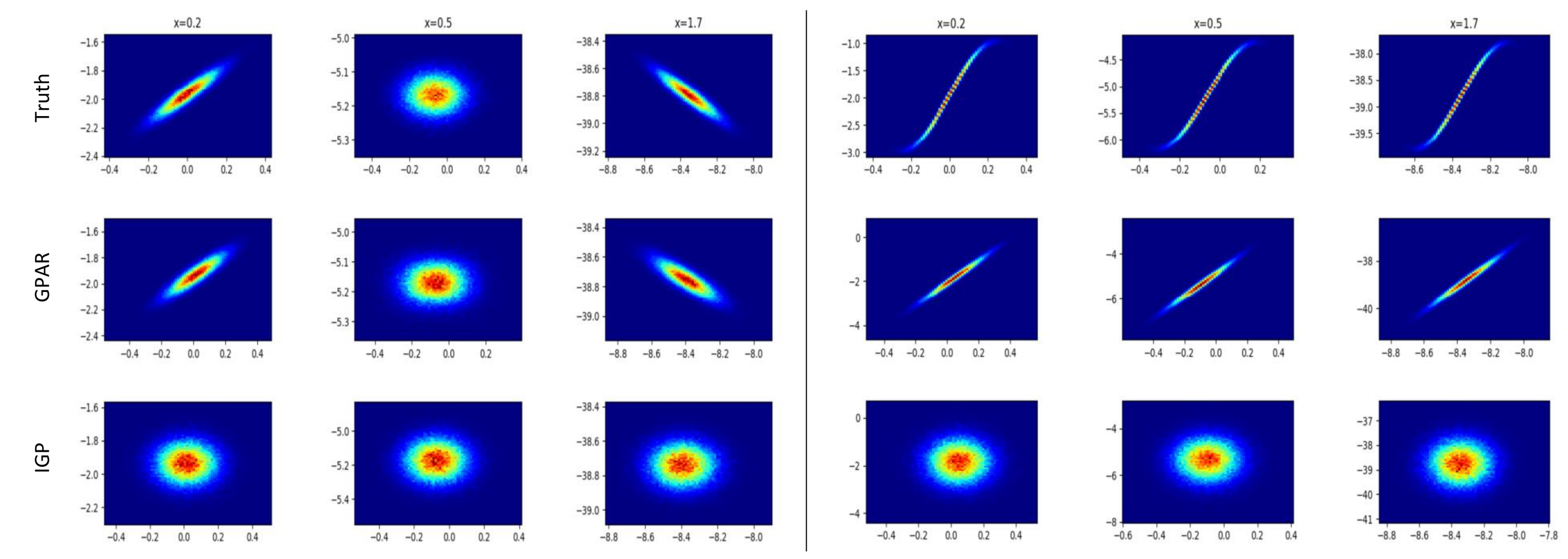
GPAR vs independent GPs (IGPs) on synthetic data:



GPAR vs IGPs using sparse GPs on real data:



Noise structure recovery on synthetic data:

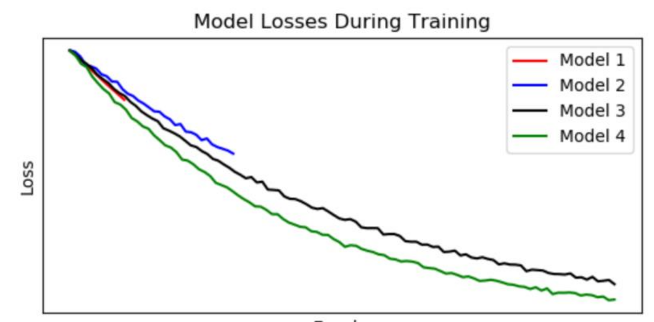


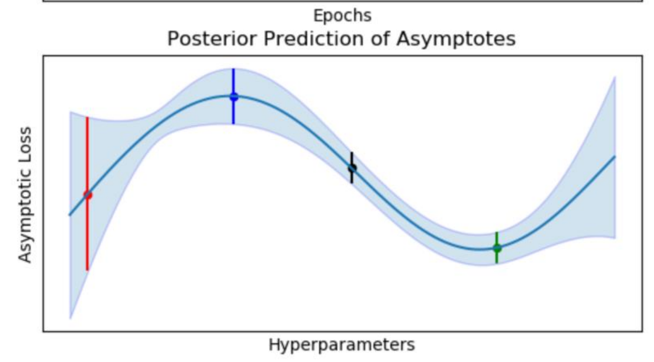
Freeze-Thaw Bayesian Optimization

Freeze-thaw is an information-theoretic approach that uses Bayesian optimization for hyperparameter tuning [2]. Our goal is to improve this with GPAR.

Approach

- Model each loss during the training process using a different set of GPs (using a custom kernel).
- Model the asymptotic cross-validation loss over a set of feasible hyperparameters using a GP.





Bayesian Quadrature

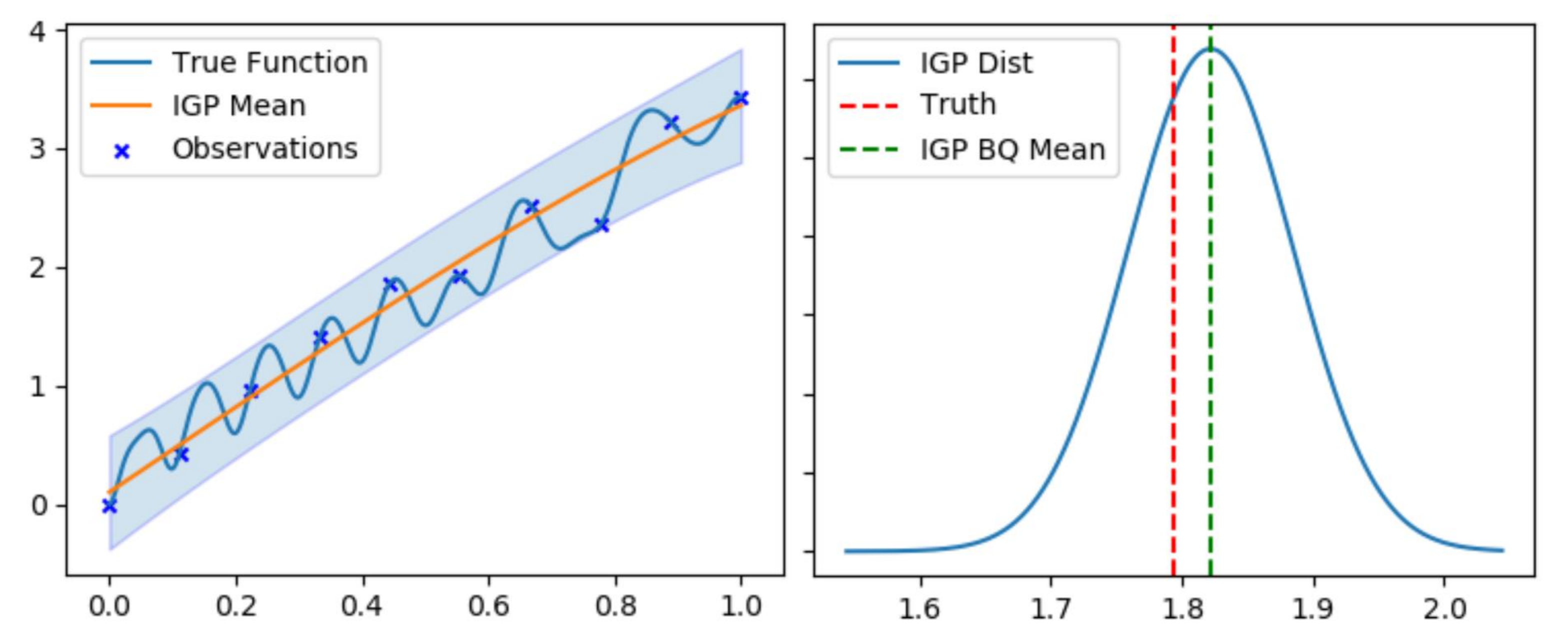
We are interested in solving a collection of integrals where the integrands, indexed by *k*, have a causal relationship [3]:

$$Z = \int_{\mathcal{X}} f_k(x) dx$$

1. Model each integrand with a GP prior.
2. Condition the priors with observed data.
3. Pass each GP through the integral operator.

$$\mu_N = \left\{ \int_{\mathcal{X}} k(\cdot, \mathbf{X}) \mathrm{d}\mathbf{x} \right\} k(\mathbf{X}, \mathbf{X})^{-1} f(\mathbf{X})$$

$$\sigma_N^2 = \left\{ \int_{\mathcal{X}} \int_{\mathcal{X}} k(\cdot, \cdot) \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{x}' \right\} - \left\{ \int_{\mathcal{X}} k(\cdot, \mathbf{X}) \mathrm{d}\mathbf{x} \right\} k(\mathbf{X}, \mathbf{X})^{-1} \left\{ \int_{\mathcal{X}} k(\mathbf{X}, \cdot) \mathrm{d}\mathbf{x} \right\}$$



Possible Extensions

Deep GPs, parameter tying, optimal conditional ordering, neural architecture search.

References

[1] James Requeima, Will Tebbutt, Wessel Bruinsma, and Richard Turner. The gaussian process autoregressive regression model (gpar). 02 2018.

[2] Kevin Swersky, Jasper Snoek, and Ryan Prescott Adams. Freeze-thaw bayesian optimization. 06 2014.

[3] Xiaoyue Xi, Francois-Xavier Briol, and Mark Girolami. Bayesian quadrature for multiple related integrals. 01 2018.