UNIVERSITY OF CAMBRIDGE

Machine Learning and the Physical World

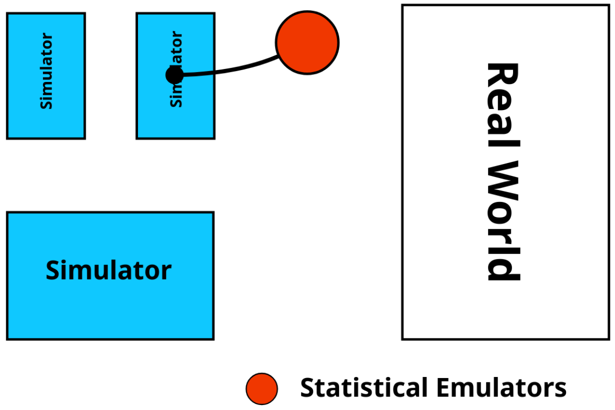
Lecture 12 : Multifidelity Modelling

Carl Henrik Ek - che29@cam.ac.uk

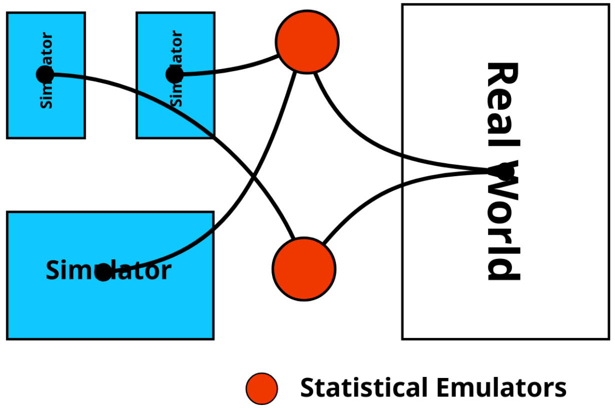
19th of November, 2025

http://carlhenrik.com

Introduction

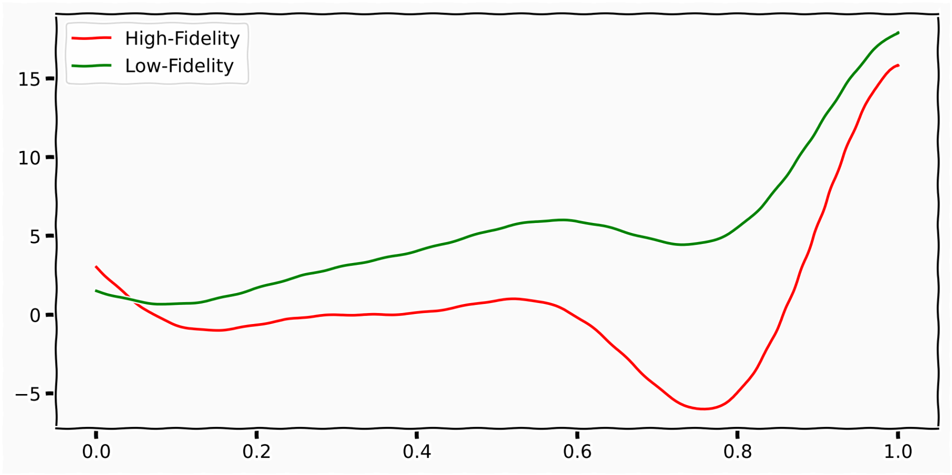


Simulator
Simulator
Simulator
Real World
Statistical Emulators


Simulator
Simulator
Simulator
Real World
Statistical Emulators

Linear Multifidelity Models

Forrester Function



Linear Multifidelity Model

$$f_{\mathsf{high}} = f_{\mathsf{err}} + \rho f_{\mathsf{low}}$$

GP Multifidelity Model

$$\begin{bmatrix} f_{\text{low}} \\ f_{\text{high}} \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{\text{low}} & ? \\ ?^{\mathrm{T}} & ?? \end{bmatrix} \right)$$

Multifidelity GP ?

$$\mathrm{Var}(f_{\mathrm{high}}, f_{\mathrm{low}}) = \mathbb{E}\left[(f_{\mathrm{err}} + \rho f_{\mathrm{low}} - 0)(f_{\mathrm{low}} - 0)\right]$$

$$\begin{aligned}\mathrm{Var}(f_{\mathrm{high}}, f_{\mathrm{low}}) &= \mathbb{E}\left[(f_{\mathrm{err}} + \rho f_{\mathrm{low}} - 0)(f_{\mathrm{low}} - 0)\right] \\ &= \mathbb{E}\left[f_{\mathrm{err}} f_{\mathrm{low}}\right] + \rho \mathbb{E}\left[f_{\mathrm{low}}^2\right]\end{aligned}$$

Multifidelity GP ?

$$\begin{aligned}\mathrm{Var}(f_{\text{high}}, f_{\text{low}}) &= \mathbb{E}\left[(f_{\text{err}} + \rho f_{\text{low}} - 0)(f_{\text{low}} - 0)\right] \\ &= \mathbb{E}\left[f_{\text{err}} f_{\text{low}}\right] + \rho \mathbb{E}\left[f_{\text{low}}^2\right] \\ &= \rho k_{\text{low}}\end{aligned}$$

Multifidelity GP ??

$$\mathsf{Var}(f_{\mathsf{high}}) = \mathbb{E}\left[(f_{\mathsf{err}} + \rho f_{\mathsf{low}} - 0)(f_{\mathsf{err}} + \rho f_{\mathsf{low}} - 0)\right]$$

Multifidelity GP ??

$$\begin{aligned}\mathsf{Var}(f_{\mathsf{high}}) &= \mathbb{E}\left[(f_{\mathsf{err}} + \rho f_{\mathsf{low}} - 0)(f_{\mathsf{err}} + \rho f_{\mathsf{low}} - 0)\right] \\ &= \mathbb{E}\left[(f_{\mathsf{err}} + \rho f_{\mathsf{low}} - 0)^2\right] = \mathbb{E}\left[f_{\mathsf{err}}^2 + \rho^2 f_{\mathsf{low}}^2 + 2\rho f_{\mathsf{low}} f_{\mathsf{err}}\right]\end{aligned}$$

Multifidelity GP ??

$$
\begin{aligned}
\mathrm{Var}(f_{\mathrm{high}}) &= \mathbb{E}\left[(f_{\mathrm{err}} + \rho f_{\mathrm{low}} - 0)(f_{\mathrm{err}} + \rho f_{\mathrm{low}} - 0)\right] \\
&= \mathbb{E}\left[(f_{\mathrm{err}} + \rho f_{\mathrm{low}} - 0)^2\right] = \mathbb{E}\left[f_{\mathrm{err}}^2 + \rho^2 f_{\mathrm{low}}^2 + 2\rho f_{\mathrm{low}} f_{\mathrm{err}}\right] \\
&= \mathbb{E}\left[f_{\mathrm{err}}^2\right] + \rho^2 \mathbb{E}\left[f_{\mathrm{low}}^2\right] + 2\rho \mathbb{E}\left[f_{\mathrm{low}} f_{\mathrm{err}}\right]
\end{aligned}
$$

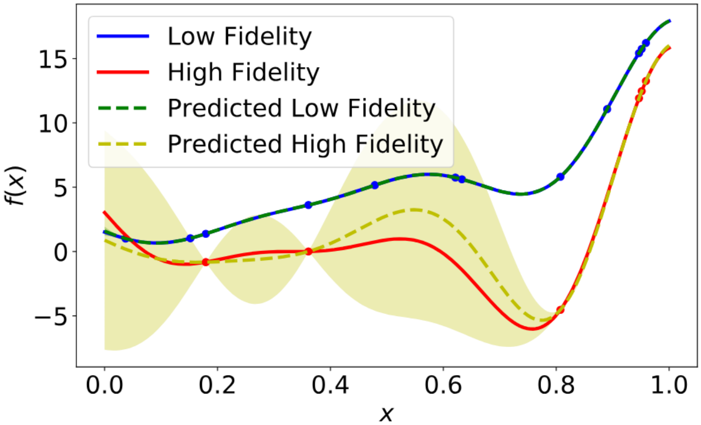
Multifidelity GP ??

$$
\begin{aligned}
\mathrm{Var}(f_{\text{high}}) &= \mathbb{E}\left[(f_{\text{err}} + \rho f_{\text{low}} - 0)(f_{\text{err}} + \rho f_{\text{low}} - 0)\right] \\
&= \mathbb{E}\left[(f_{\text{err}} + \rho f_{\text{low}} - 0)^2\right] = \mathbb{E}\left[f_{\text{err}}^2 + \rho^2 f_{\text{low}}^2 + 2\rho f_{\text{low}} f_{\text{err}}\right] \\
&= \mathbb{E}\left[f_{\text{err}}^2\right] + \rho^2 \mathbb{E}\left[f_{\text{low}}^2\right] + 2\rho \mathbb{E}\left[f_{\text{low}} f_{\text{err}}\right] \\
&= k_{\text{err}} + \rho^2 k_{\text{low}}
\end{aligned}
$$

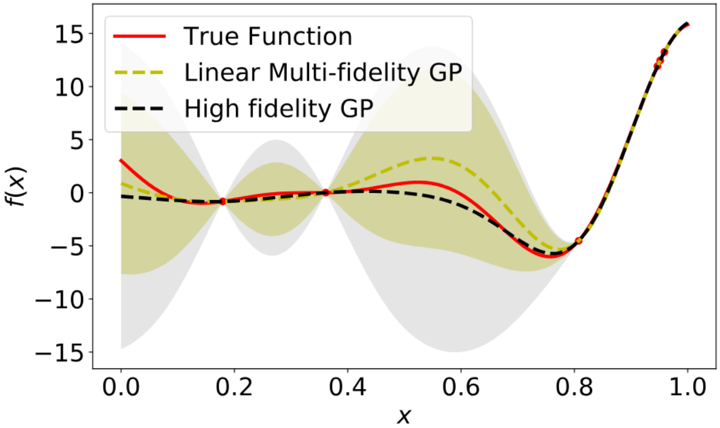
GP Multifidelity Model

$$\begin{bmatrix} f_{\text{low}} \\ f_{\text{high}} \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{\text{low}} & \rho k_{\text{low}} \\ \rho k_{\text{low}}^{\mathrm{T}} & k_{\text{err}} + \rho^2 k_{\text{low}} \end{bmatrix} \right)$$

Results



Low Fidelity
High Fidelity
Predicted Low Fidelity
Predicted High Fidelity
15
10
5
0
−5
f(x)
0.0
0.2
0.4
0.6
0.8
1.0
x


True Function
Linear Multi-fidelity GP
High fidelity GP
15
10
5
0
−5
−10
−15
f(x)
0.0
0.2
0.4
0.6
0.8
1.0
x

Linear Model of Co-regionalisation

$$\mathbf{y}_i = \mathbf{W}\mathbf{f}_i + \epsilon$$

$$f_l \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_l)$$

$$\mathrm{cov}(y_{iq}, y_{iq'}) = \sum_{l=1}^{L} k_l(\mathbf{x}_i, \mathbf{x}_j) w_{ql} w_{ql'}$$

- L latent functions $f_1, \ldots, f_l$
- $\mathbf{W} \in \mathbb{R}^{DL}$ loading matrix
- Gaussians are closed under linear operations

Linear Model of Co-regionalisation

$$\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{nd} \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} + \sigma^2 \mathbf{I} nd \right)$$

- General case

$$\mathbf{K} = \sum_{i=1}^{L} \mathbf{w}_l \mathbf{w}_l^{\mathrm{T}} \otimes \mathbf{K}_l$$

- All processes share the same kernel

$$\mathbf{K} = \mathbf{W}\mathbf{W}^{\mathrm{T}} \otimes \mathbf{K}_{nn}$$

Multi-output Gaussian Processes (Álvarez et al.)

$$\mathbf{K} = \mathbf{B}_{dd} \otimes \mathbf{K}_{nn}$$

- Inverse

$$\mathbf{K}^{-1} = (\mathbf{B}_{dd})^{-1} \otimes (\mathbf{K}_{nn})^{-1}$$

- Determinant

$$\det(\mathbf{K}) = \det(\mathbf{B}_{dd})^{n} \det(\mathbf{K}_{nn})^{d}$$

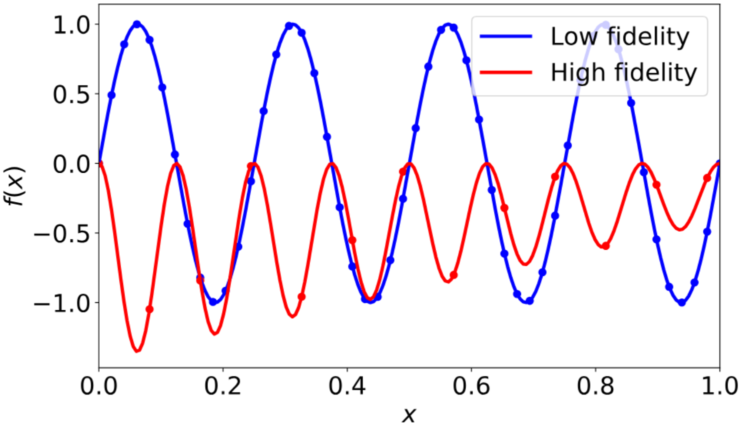
Mauricio A Álvarez et al. (2012). "Kernels for Vector-Valued Functions: A Review". In: *Foundations and Trends® in Machine Learning* 4.3, pp. 195–266

Non-Linear Multifidelity Models

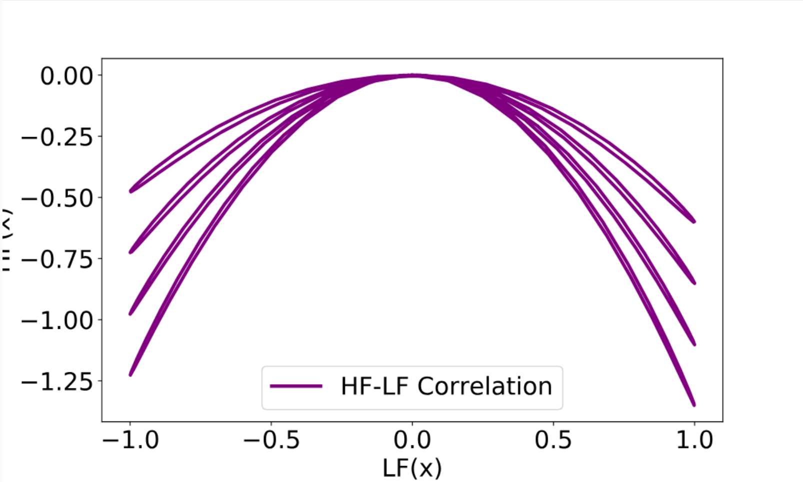
Non-Linear Multifidelity

$$
\begin{aligned}
f_{\text{low}}(x) &= \sin(8\pi x) \\
f_{\text{high}}(x) &= (x - \sqrt{2}) f_{\text{low}}^2(x)
\end{aligned}
$$

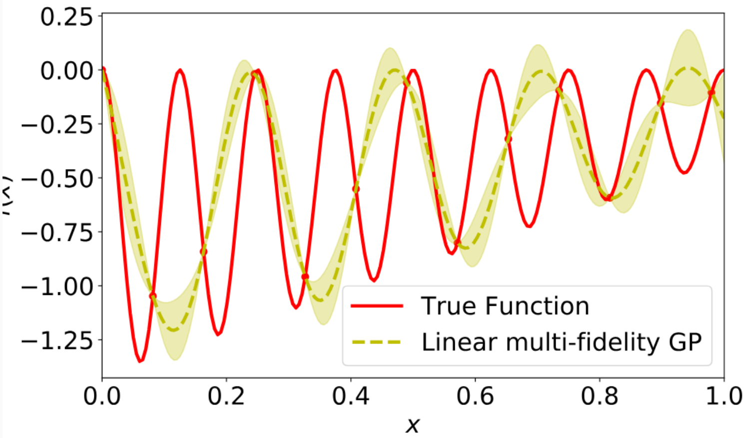
Non-Linear Multifidelity



Non-Linear Multifidelity: Correlation



Non-Linear Multifidelity: Linear Model



Non-Linear Multifidelity

$$f_{\text{high}}(x) = \rho(f_{\text{low}}(x)) + f_{\text{err}(x)}$$

Composite Functions

$$p(y \mid x) = \int p(y \mid \rho) p(\rho \mid f_{\text{low}}) p(f_{\text{low}} \mid x) \mathrm{d}\rho \mathrm{d}f_{\text{low}}$$

Composite Gaussian Processes

$$p(y) = \int p(y \mid f_2) p(f_2 \mid f_1) p(f_1) \mathrm{d}f_2, \mathrm{d}f_2$$

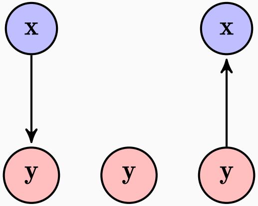
- GP Prior

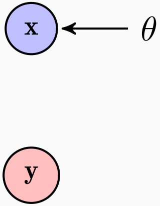
$$p(f_2 \mid f_1) \sim \mathcal{N}(f_2; 0, K(f_1, f_1)) \propto e^{\frac{1}{2}(f_2^{\mathrm{T}} K(f_1, f_1)^{-1} f_2)}$$

- Likelihood

$$p(y \mid f_2) \sim \mathcal{N}(y; f_2, \beta^{-1}) \propto e^{-\frac{1}{2}(y - f_2)^{\mathrm{T}}(y - f_2)}$$

Variational Inference





$$p(y) = \int_x p(y|x)p(x) = \frac{p(y|x)p(x)}{p(x|y)}$$

$$q_\theta(x) \approx p(x|y)$$

Lower Bound (Titsias et al., 2010)

$$p(f, u \mid x, z)$$

- Add another set of samples from the same prior
- Conditional distribution

Lower Bound (Titsias et al., 2010)

$$p(f, u \mid x, z) = p(f \mid u, x, z)p(u \mid z)$$

- Add another set of samples from the same prior
- Conditional distribution

Lower Bound (Titsias et al., 2010)

$$
\begin{aligned}
p(f, u \mid x, z) &= p(f \mid u, x, z) p(u \mid z) \\
&= \mathcal{N}(f \mid K_{fu} K_{uu}^{-1} u, K_{ff} - K_{fu} K_{uu}^{-1} K_{uf}) \mathcal{N}(u \mid \mathbf{0}, K_{uu})
\end{aligned}
$$

- Add another set of samples from the same prior
- Conditional distribution

Lower Bound

$$p(y, f, u, x \mid z) = p(y \mid f)p(f \mid u, x)p(u \mid z)p(x)$$

- we have done nothing to the model, just project an additional set of marginals from the GP
- *however* we will now interpret u and z not as random variables but variational parameters
- i.e. the variational distribution $q(\cdot)$ is parametrised by these

Lower Bound

- Variational distributions are approximations to intractable posteriors,

$$q(u) \approx p(u \mid y, x, z, f)$$
$$q(f) \approx p(f \mid u, x, z, y)$$
$$q(x) \approx p(x \mid y)$$

Lower Bound

- Variational distributions are approximations to intractable posteriors,

$$q(u) \approx p(u \mid y, x, z, f)$$
$$q(f) \approx p(f \mid u, x, z, y)$$
$$q(x) \approx p(x \mid y)$$

- Bound is tight if u completely represents f i.e. u is sufficient statistics for f

$$q(f) \approx p(f \mid u, x, z, y) = p(f \mid u, x, z)$$

Lower Bound

$$\mathcal{L} = \int_{x,f,u} q(f)\,q(u)\,q(x) \log \frac{p(y \mid f)\,p(f \mid u, x, z)\,p(u \mid z)}{q(f)\,q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

Lower Bound

$$\mathcal{L} = \int_{x,f,u} q(f)q(u)q(x) \log \frac{p(y \mid f)p(f \mid u, x, z)p(u \mid z)}{q(f)q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

$$= \int_{x,f,u} p(f \mid u, x, z)q(u)q(x) \log \frac{p(y \mid f){\color{red}p(f \mid u, x, z)}p(u \mid z)}{{\color{red}p(f \mid u, x, z)}q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

Lower Bound

$$\mathcal{L} = \int_{x,f,u} q(f)q(u)q(x)\log\frac{p(y \mid f)p(f \mid u, x, z)p(u \mid z)}{q(f)q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

$$= \int_{x,f,u} p(f \mid u, x, z)q(u)q(x)\log\frac{p(y \mid f){\color{red}p(f \mid u, x, z)}p(u \mid z)}{{\color{red}p(f \mid u, x, z)}q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

$$= \int_{x,f,u} p(f \mid u, x, z)q(u)q(x)\log\frac{p(y \mid f)p(u \mid z)}{q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

Lower Bound

$$\mathcal{L} = \int_{x,f,u} q(f)q(u)q(x)\log\frac{p(y \mid f)p(f \mid u,x,z)p(u \mid z)}{q(f)q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

$$= \int_{x,f,u} p(f \mid u,x,z)q(u)q(x)\log\frac{p(y \mid f)\color{red}{p(f \mid u,x,z)}p(u \mid z)}{\color{red}{p(f \mid u,x,z)}q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

$$= \int_{x,f,u} p(f \mid u,x,z)q(u)q(x)\log\frac{p(y \mid f)p(u \mid z)}{q(u)} - \mathrm{KL}(q(x) \parallel p(x))$$

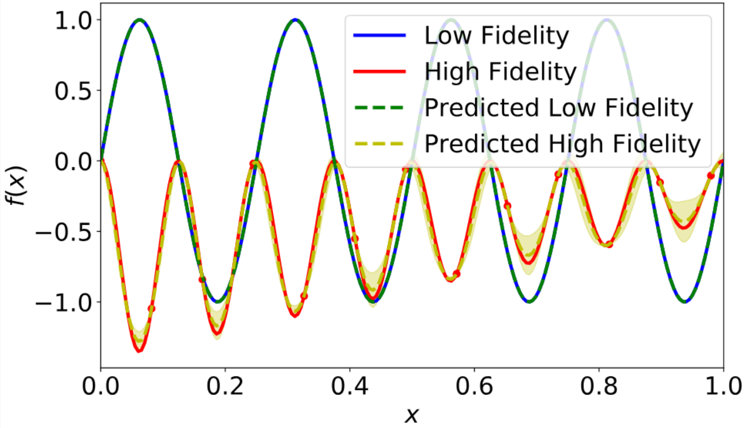
$$= \mathbb{E}_{p(f \mid u,x,z)}[p(y \mid f)] - \mathrm{KL}(q(u) \parallel p(u \mid z)) - \mathrm{KL}(q(x) \parallel p(x))$$

Lower Bound

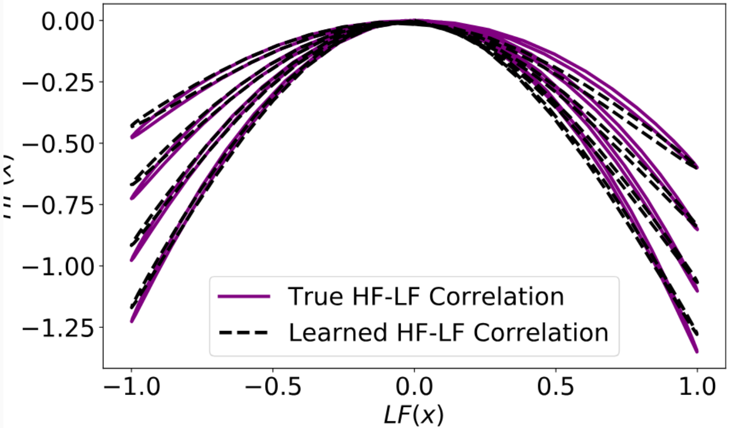
$$\mathcal{L} = \mathbb{E}_{p(f|u,x,z)}[p(y \mid f)] - \mathrm{KL}(q(u) \parallel p(u \mid z)) - \mathrm{KL}(q(x) \parallel p(x))$$

- Expectation tractable (for some co-variances)
- Allows us to place priors and not "regularisers" over the latent representation
- Stochastic inference Hensman et al., 2013
- Importantly $p(x)$ only appears in $\mathrm{KL}(\cdot \parallel \cdot)$ term!

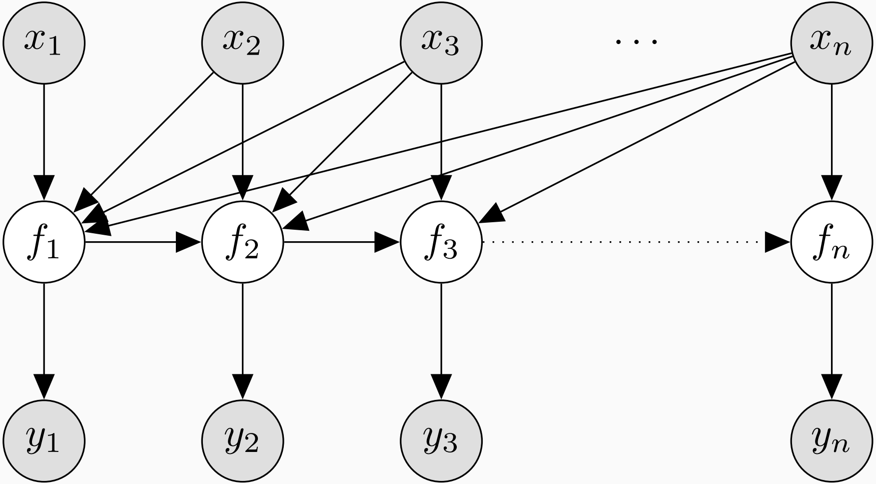
Non-Linear Multifidelity



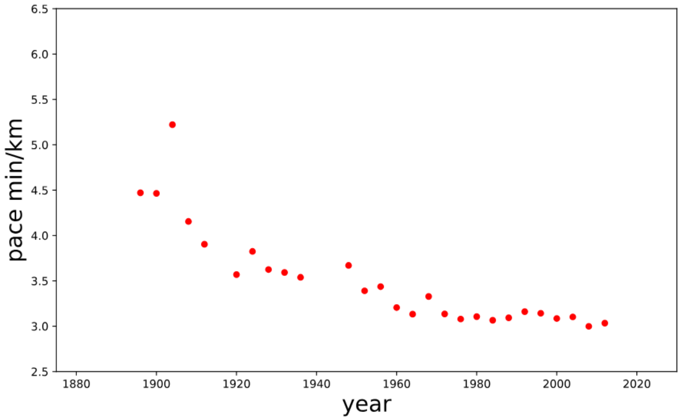
Non-Linear Multifidelity



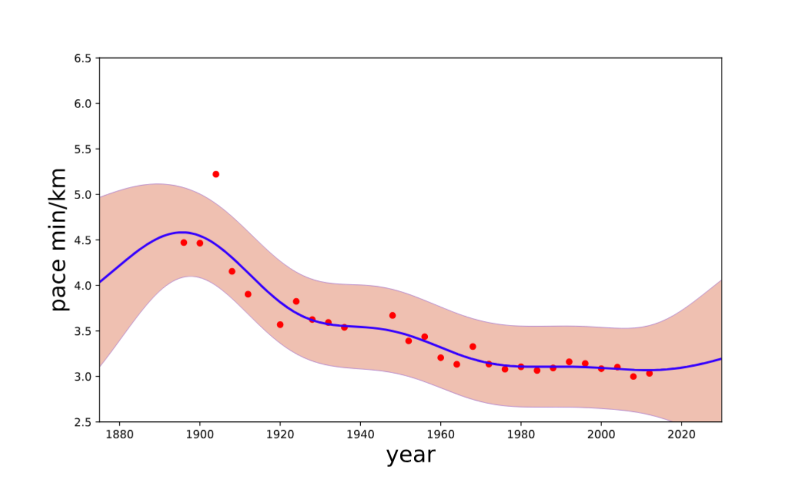
Non-Linear Multifidelity (Cutajar et al.)



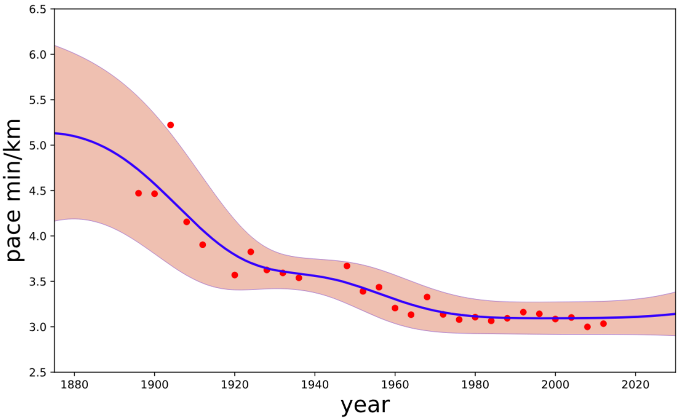
Olympic Marathon



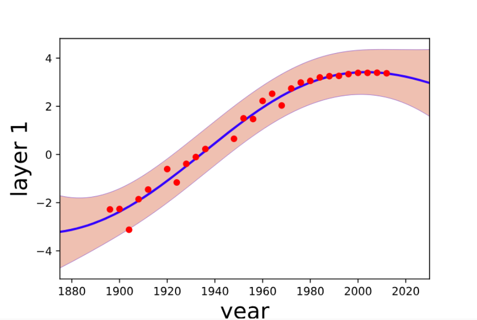
Olympic Marathon: GP

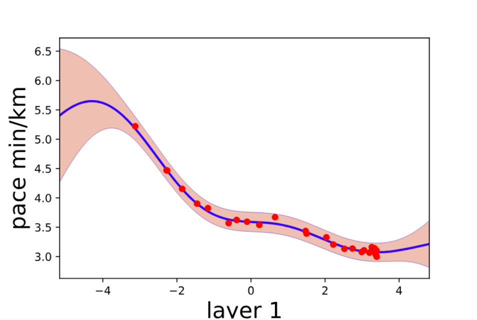


Olympic Marathon: Composite GP

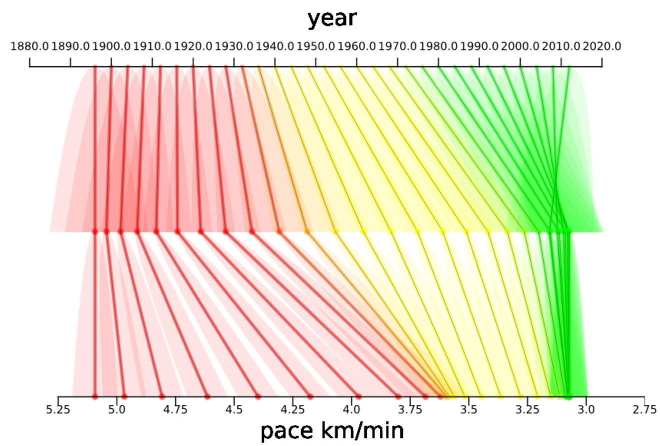


Olympic Marathon: Composite GP Layers

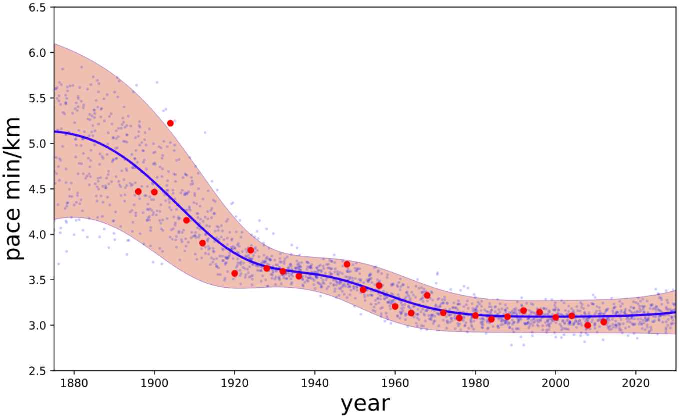




Olympic Marathon: Pinball Plot



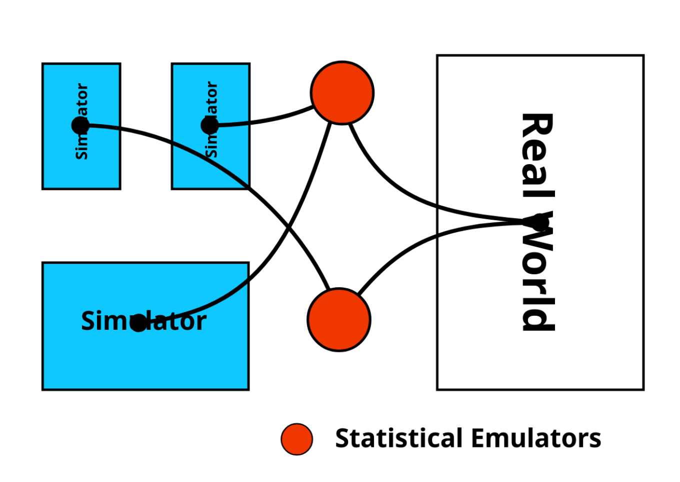
Olympic Marathon: Composite GP Samples



Summary

Summary

- We can often acquire much more data if we acknowledge their fidelity
- Linear Model of Co-regionalisation
- Non-linear Multifidelity Models


Simulator
Simulator
Simulator
Real World
Statistical Emulators

Remaining Lectures

Monday Stirred Tank Reactor Design - *Bethany Conroy*

Wednesday Further Generative Models - *Carl Henrik*

Monday Potential Research Talk - *Carl Henrik*

Wednesday Q/A

eof

References

- Álvarez, Mauricio A, L. Rosasco, and Neil D Lawrence (2012). **“Kernels for Vector-Valued Functions: A Review”.** In: *Foundations and Trends® in Machine Learning* 4.3, pp. 195–266.
- Cutajar, Kurt et al. (2019). **“Deep Gaussian Processes for Multi-Fidelity Modeling”.** In: *CoRR*.
- Hensman, James, N Fusi, and Neil D Lawrence (2013). **“Gaussian Processes for Big Data”.** In: *Uncertainty in Artificial Intelligence*.

Titsias, Michalis and Neil D Lawrence (2010). “Bayesian Gaussian Process Latent Variable Model”. In: *International Conference on Airtificial Inteligence and Statistical Learning*, pp. 844–851.