CHIMERA

Supported by the Engineering and Physical Sciences
Research Council (Grant number EP/T017791/1)

Integrative Analysis and Imputation of Multiple Data Streams via Deep Gaussian Process

Ali Septiandri

Background

- In ICU settings, data come from multiple sources and are inherently related

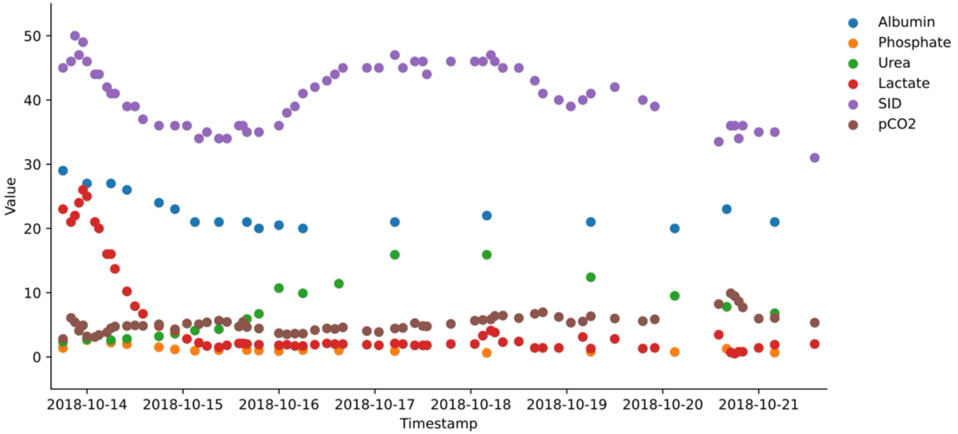
- Measurements collected at irregular intervals (informative sampling)—aligning them will result in missing values

- Cannot always get more samples! Some measurements are invasive (Siegal et al., 2023)

Photo by National Cancer Institute on Unsplash

Challenges

- We want to impute missing values…
- but traditional imputation often ignores temporal structure (e.g. MICE) & uncertainty (e.g. deep learning)
- Need for robust, uncertainty-aware imputation in critical care datasets



On uncertainty quantification

- Medical observations are inherently uncertain, coming from measurement errors or the use of surrogate markers → leading to unreliable model predictions (Cabitza et al., 2017)
- Alerts triggered by prediction tools are often not accompanied by a clinically actionable change → alarm fatigue (Embi & Leonard, 2012; Umscheid et al., 2015)

Physicochemical model

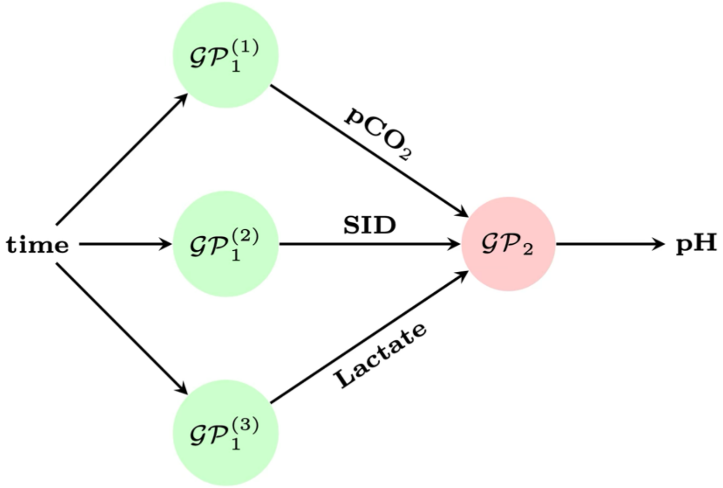
- In critical care medicine, clinicians monitor pH levels to inform them about the conditions of a patient
- While pH is the primary variable to monitor, other covariates provide information on metabolic status (Gattioni et al., 2017)
- pH can be modelled from strong ion difference (SID), total weak acid, and pCO2 by the Stewart-Fencl approach

$$[SID] + [H^+] - K_C \frac{pCO_2}{[H^+]} - \frac{K_A A_{TOT}}{K_A + [H^+]} - K_3 \frac{K_C pCO_2}{[H^+]^2} - \frac{K_W}{[H^+]} = 0$$

where SID, A_{TOT}, and pCO_2 are independent variables and K_X are constants.

Proposed solution

- GPs and Deep GPs are typically used for emulating computationally expensive numerical models
- Integrates longitudinal & cross-sectional information
- Joint modelling for all data streams
- Provides uncertainty quantification for imputed values



Deep Gaussian Process with Stochastic Imputation (Ming et al., 2023)

Gaussian processes

$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{R}(\boldsymbol{X}))$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the mean vector, σ^2 is the scale parameter, and $\boldsymbol{R}(\boldsymbol{X}) \in \mathbb{R}^{N \times N}$ is the correlation matrix

Cell ij in the matrix $\boldsymbol{R}(\boldsymbol{X})$ is specified by $k\left(\boldsymbol{X}_{i*}, \boldsymbol{X}_{j*}\right) + \eta 1_{\{\boldsymbol{X}_{i*},=\boldsymbol{X}_{j*}\}}$, where $k(\cdot,\cdot)$ is a given kernel function with η being the nugget term and $1_{\{\cdot\}}$ being the indicator function

Gaussian processes

Given a new input position $\boldsymbol{x}_0 \in \mathbb{R}^{1\times D}$, then

$\mu_0 = \boldsymbol{r}(\boldsymbol{x}_0)^T \boldsymbol{R}(\boldsymbol{X})^{-1} \boldsymbol{y}$

$\sigma_0^2 = \sigma^2(1 + \eta - \boldsymbol{r}(\boldsymbol{x}_0)^T \boldsymbol{R}(\boldsymbol{x})^{-1} \boldsymbol{r}(\boldsymbol{x}_0))$

where $\boldsymbol{r}(\boldsymbol{x}_0) = [k(\boldsymbol{x}_0, \boldsymbol{x}_{1*}), \dots, k(\boldsymbol{x}_0, \boldsymbol{x}_{N*})]^T$

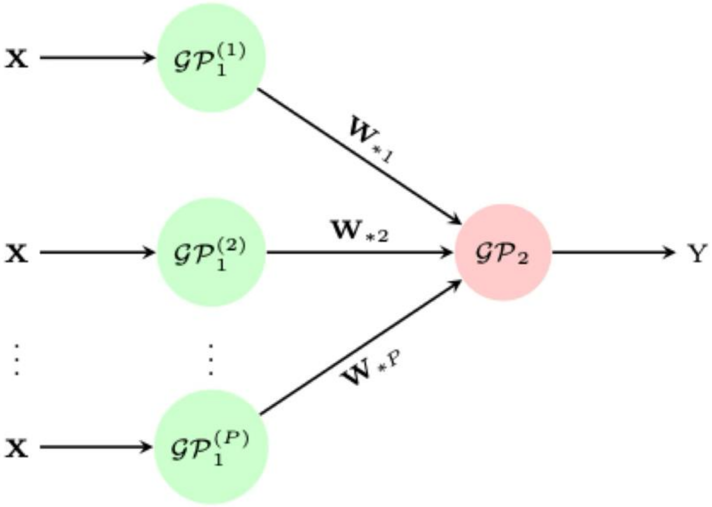
Deep GPs

- Consider a GP model with N sets of D-dimensional input ($\boldsymbol{X} \in \mathbb{R}^{N \times D}$) and produces N sets of P-dimensional output ($\boldsymbol{W} \in \mathbb{R}^{N \times D}$)
- In the Stewart–Fencl approach, this multi-output GP model can be interpreted as using time as a shared input variable and predicting covariates as outputs
- We can assume that the output $\boldsymbol{W}$ of this model, i.e. the column vectors $\boldsymbol{W}_{*p}$, is conditionally independent with respect to $\boldsymbol{X}$
- We then link the output $\boldsymbol{W}$ to a second GP model that produces $\boldsymbol{N}$ one-dimensional outputs ($\boldsymbol{Y} \in \mathbb{R}^{N}$), e.g. to predict pH

Deep GPs

We can see it as a linked GP where, for a new input position $\boldsymbol{x}_0$, the posterior predictive distribution of the output can be written as

$$\begin{aligned} p(y_0 \mid \mathbf{x}_0; \mathbf{y}, \mathbf{w}, \mathbf{x}) &= \int p(y_0 \mid \mathbf{w}_0; \mathbf{y}, \mathbf{w}, \mathbf{x}) p(\mathbf{w}_0 \mid \mathbf{x}_0; \mathbf{y}, \mathbf{w}, \mathbf{x}) d\mathbf{w}_0 \\ &= \int p(y_0 \mid \mathbf{w}_0; \mathbf{y}, \mathbf{w}) \prod_{p=1}^{P} p(w_{0p} \mid \mathbf{x}_0; \mathbf{w}_p^*, \mathbf{x}) d\mathbf{w}_0 \end{aligned}$$



Deep GPs

Then the mean and variance become

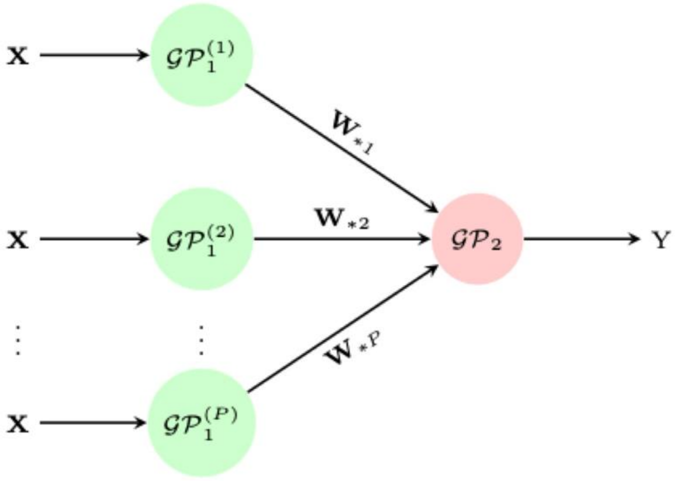
$\tilde{\mu}_0 = \boldsymbol{I}(\boldsymbol{x}_0)^T \boldsymbol{R}(\boldsymbol{w})^{-1} \boldsymbol{y}$

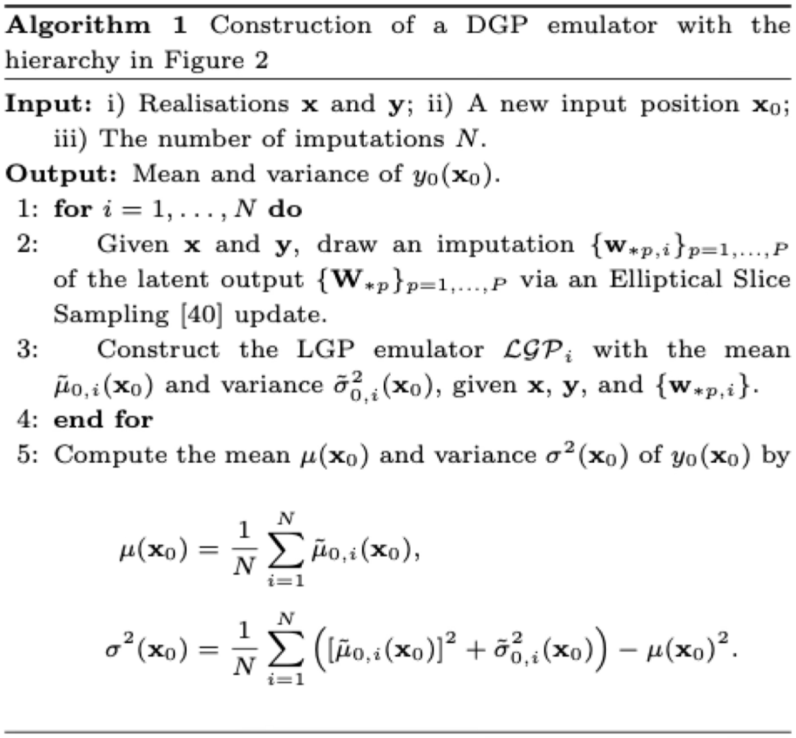
$\tilde{\sigma}_0^2 = \boldsymbol{y}^T \boldsymbol{R}(\boldsymbol{w})^{-1} \boldsymbol{J}(\boldsymbol{x}_0) \boldsymbol{R}(\boldsymbol{w})^{-1} \boldsymbol{y} \; - [\boldsymbol{I}(\boldsymbol{x}_0)^T \boldsymbol{R}(\boldsymbol{w})^{-1} \boldsymbol{y} \,]^2 + \; \sigma^2 (1 + \eta - \boldsymbol{tr}[\boldsymbol{R}(\boldsymbol{w})^{-1} \boldsymbol{J}(\boldsymbol{x}_0)])$

where $\boldsymbol{I}(\boldsymbol{x}_0) \in \mathbb{R}^{N \times 1}$ with its i-th element $\boldsymbol{I}_i = \prod_{p=1}^{P} \mathbb{E}[k_p(W_{0p}(\boldsymbol{x}_0), w_{ip})]$

and $\mathbf{J}(\boldsymbol{x}_0) \in \mathbb{R}^{N \times N}$ with its ij-th element $\boldsymbol{J}_{ij} = \prod_{p=1}^{P} \mathbb{E}[k_p(W_{0p}(\boldsymbol{x}_0), w_{ip}) k_p(W_{0p}(\boldsymbol{x}_0), w_{jp})]$

Deep GP algorithm





Algorithm 1 Construction of a DGP emulator with the hierarchy in Figure 2

```
Input: i) Realisations $\mathbf{x}$ and $\mathbf{y}$; ii) A new input position $\mathbf{x}_0$;
    iii) The number of imputations $N$.
Output: Mean and variance of $y_0(\mathbf{x}_0)$.
for $i = 1, \dots, N$ do
    Given $\mathbf{x}$ and $\mathbf{y}$, draw an imputation $\{\mathbf{w}_{*p,i}\}_{p=1,\dots,P}$
    of the latent output $\{\mathbf{W}_{*p}\}_{p=1,\dots,P}$ via an Elliptical Slice
    Sampling [40] update.
    Construct the LGP emulator $\mathcal{LGP}_i$ with the mean
    $\tilde{\mu}_{0,i}(\mathbf{x}_0)$ and variance $\tilde{\sigma}^2_{0,i}(\mathbf{x}_0)$, given $\mathbf{x}$, $\mathbf{y}$, and $\{\mathbf{w}_{*p,i}\}$.
end for
Compute the mean $\mu(\mathbf{x}_0)$ and variance $\sigma^2(\mathbf{x}_0)$ of $y_0(\mathbf{x}_0)$ by

    $\mu(\mathbf{x}_0) = \frac{1}{N} \sum_{i=1}^{N} \tilde{\mu}_{0,i}(\mathbf{x}_0),$

    $\sigma^2(\mathbf{x}_0) = \frac{1}{N} \sum_{i=1}^{N} \left( [\tilde{\mu}_{0,i}(\mathbf{x}_0)]^2 + \tilde{\sigma}^2_{0,i}(\mathbf{x}_0) \right) - \mu(\mathbf{x}_0)^2.$
```

Numerical experiment

- Data used: Paediatric ICU admissions (n=14)
- Variables: pCO2, SID (Na^+, Cl^-), lactate (weak acid), pH
- Preprocessing: Hourly discretisation, z-score normalisation, masking to simulate missingness
- Benchmarks:
  - Last observation carried forward (LOCF)
  - MICE
  - GP interpolation

Model evaluation

- Four levels of missingness: 10%, 20%, 30%, 40%
- Two evaluation metrics
  - Mean absolute error – imputation accuracy

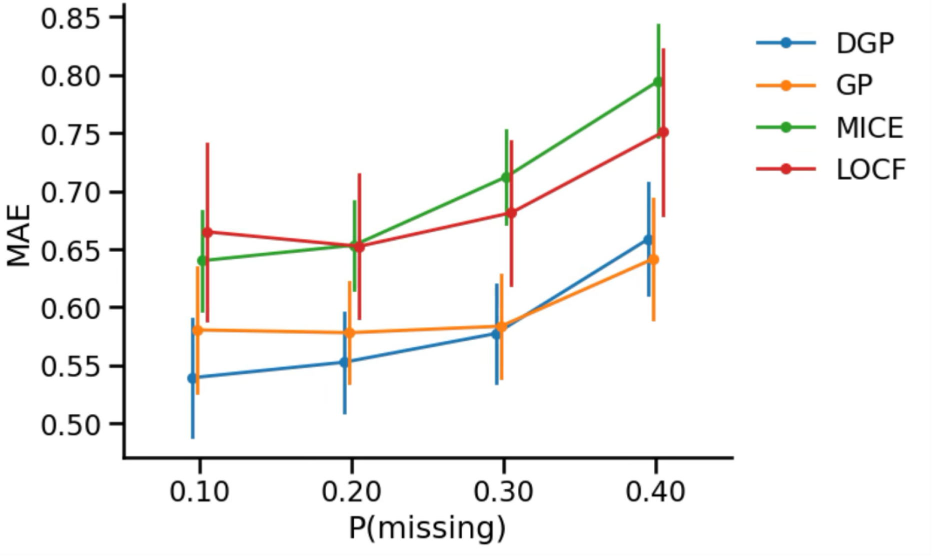
$$MAE = \frac{1}{N \times D} \sum_{i=1}^{N} \sum_{d \in D} |Y_{id} - \hat{Y}_{id}|$$

  - Negative log likelihood – uncertainty quantification

$$NLL = -\sum_{i=1}^{N} \log p(Y_i | X_i; \theta)$$

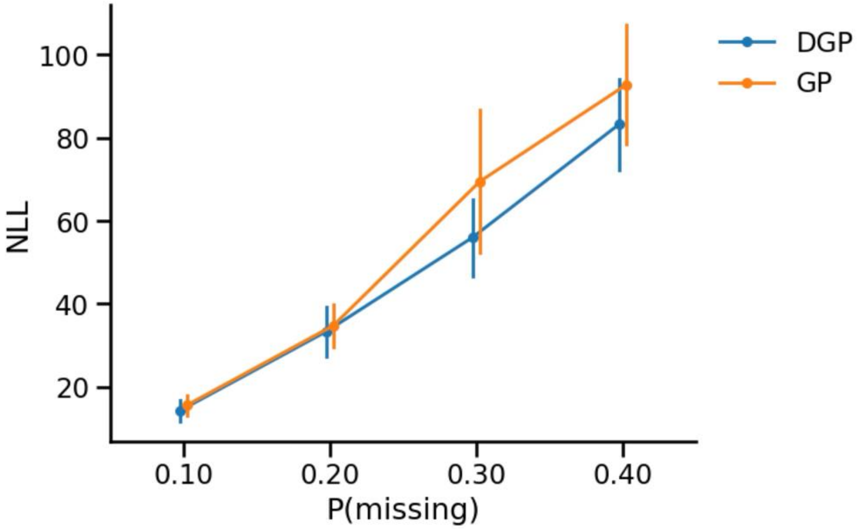
Imputing missing values

- DGP achieved the lowest error rates at 10% to 30% missing values—covering the typical 15–30% missingness in critical care data (Luo et al., 2017)
- As missingness rate increases, longitudinal information is more valuable than cross-sectional information
- DGP combines both → optimal results in lower missingness



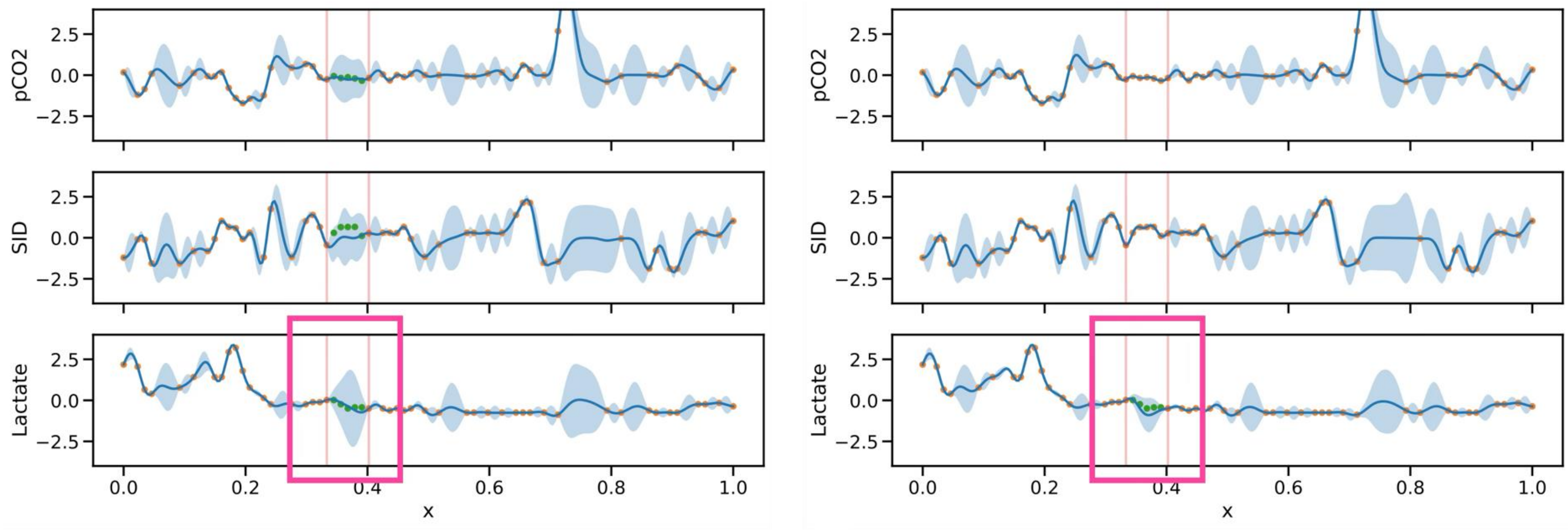
Uncertainty quantification

- DGP performs best when taking into account the uncertainty quantification
- As missingness rate increases, DGP maintains tighter uncertainty bounds than GP
- LOCF was excluded as it does not provide uncertainty quantification



Uncertainty quantification

As the covariates were connected through pH in the output layer using DGP-SI, an observation from one covariate could affect the uncertainty of another covariate where an observation was unavailable.



Implications

- DGP-SI provides reliable, uncertainty-aware imputations to aid clinical decision-making
- Insight into patient status between lab measurements
- Similar problems can be found in human activity recognition from multiple sensors, sleep disorder diagnosis using EEG, and hepatocellular carcinoma (Han et al., 2021)

Limitations & future work

- Computational expense with large datasets → sparse GP (Snelson & Ghahramani, 2007) or GPU parallelisation (Wang et al., 2019)
- Propagated uncertainty which may result in worse performance for predicting pH
- Comparison to deep learning models
- Analysis in higher missingness

Thank you

aliakbars.id/dgpsi
ali.septiandri.21@ucl.ac.uk