Department of Statistical Science, Department of Mathematics
UCL Great Ormond Street Institute of Child Health
Bloomsbury Institute of Intensive Care Medicine

Handling Missing Values in Critical Care Medicine

Upgrade Report

Ali Septiandri, Takoua Jendoubi, Alejandro Díaz, Samiran Ray, Edward Palmer

CHIMERA
Engineering and Physical Sciences Research Council

Introduction

- In critical care medicine, clinicians monitor pH levels to inform them about the conditions of a patient
- Could be derived from easily monitored CO_2 concentration, but we need other variables
- Aligning other variables collected in different frequencies results in missing values

Background

Acid-Base Balance

- Human body is composed principally of water
- Water is highly ionising: $H^+ + OH^-$
- In pure water at 25°C, the $[H^+]$ and $[OH^-]$ are $1.0 \times 10^{-7} mEq/L$
- Sorenson negative logarithmic pH = log($[H^+]$) = 7.0

Water & Alkalinity

- At 0ºC: pH = 7.5 (alkaline)
- At 100ºC: pH = 6.1 (acidic)
- Arterial pH = 7.4
  - Acidosis pH < 7.3
  - Alkalosis pH > 7.5

What determines pH?

- Water dissociation equilibrium
- Weak acid dissociation equilibrium
- Conservation of mass for weak acids
- Bicarbonate ion formation equilibrium
- Carbonate ion formation equilibrium
- Electrical neutrality

What determines pH?

- $[H^+] \times [OH^-] = K_W$
- $[H^+] \times [A^-] = K_A \times [HA]$
- $[HA] + [A^-] = A_{TOT}$
- $[H^+] \times [HCO_3^-] = K_C \times pCO_2$
- $[H^+] \times [CO_3^{2-}] = K_3 \times [HCO_3^-]$
- $[SID] + [H^+] - [HCO_3^-] - [A^-] - [CO_3^{2-}] - [OH^-] = 0$

What determines pH?

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

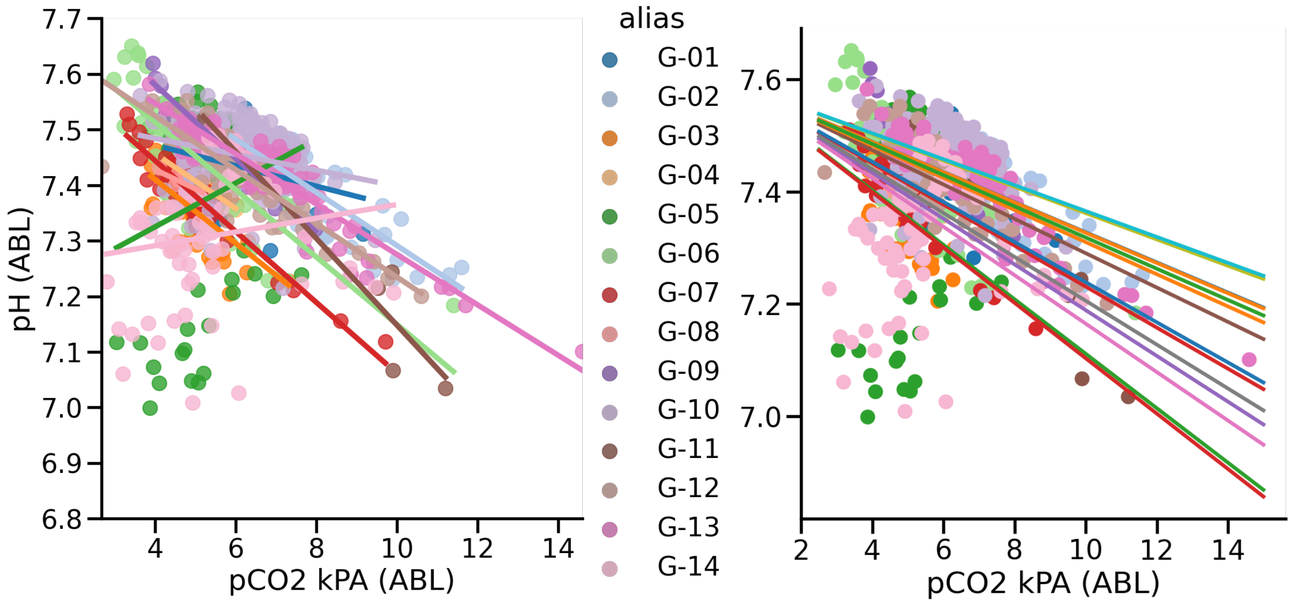
where SID, A_{TOT}, and pCO_2 are independent variables and K_X are constants.

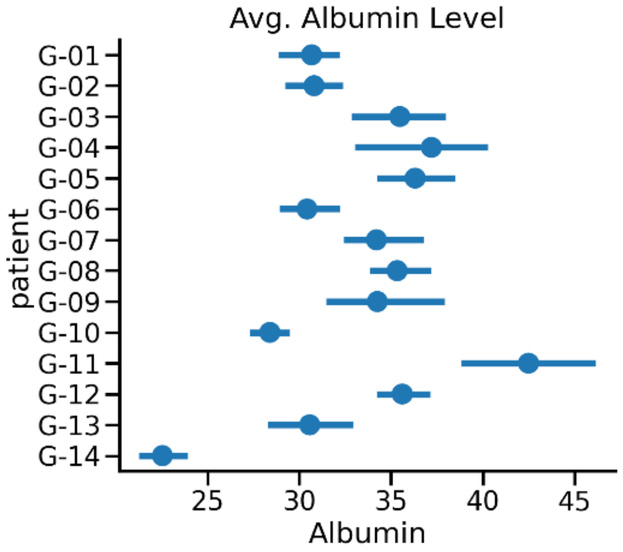
Preliminary Analysis

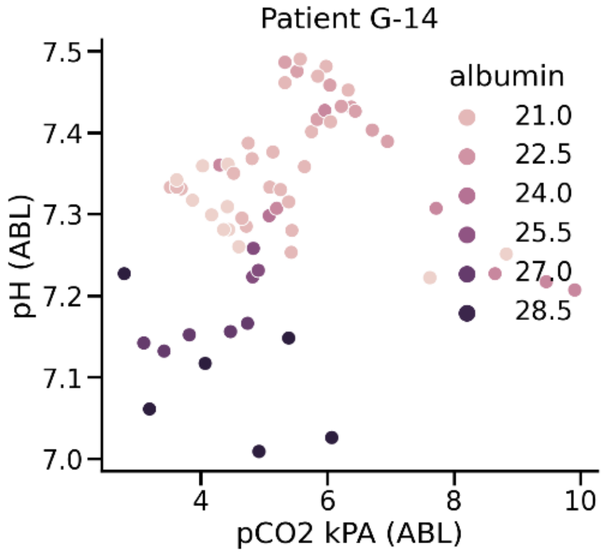
Issue

CO_2 is not enough to tell the whole story

- Model: pH ~ pCO_2
  - Left: One OLS model for each
  - Right: Hierarchical with shared intercept
- For patient G-14, albumin level is low most of the time
- It spikes when the anomaly occurs

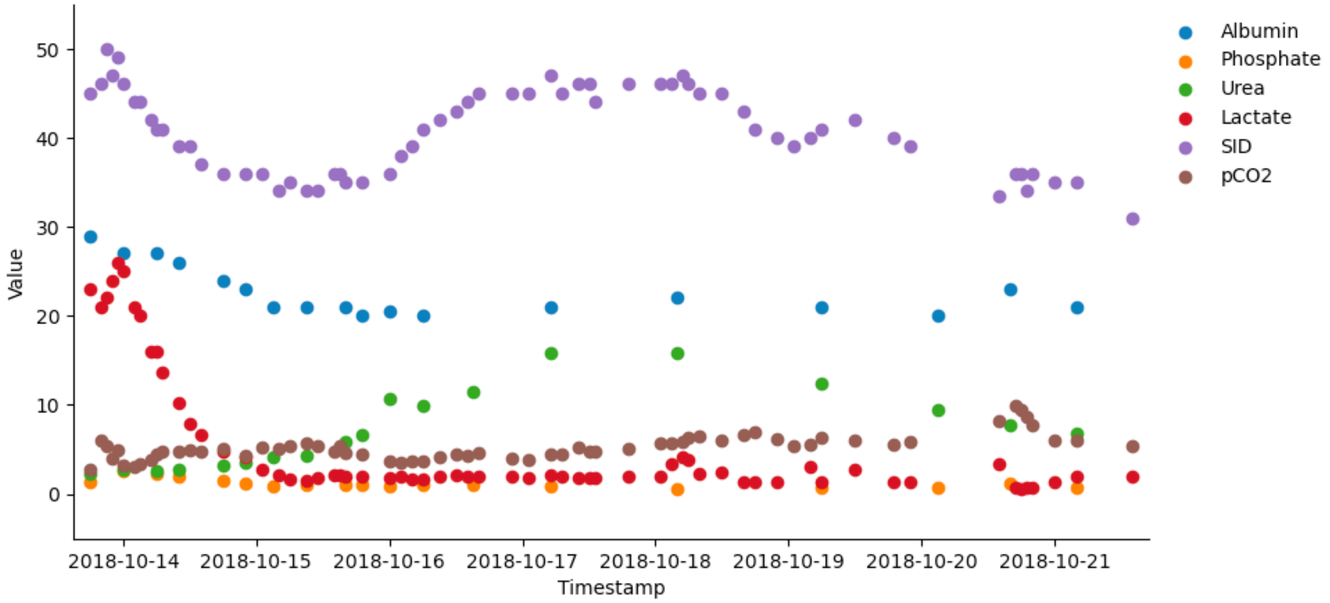






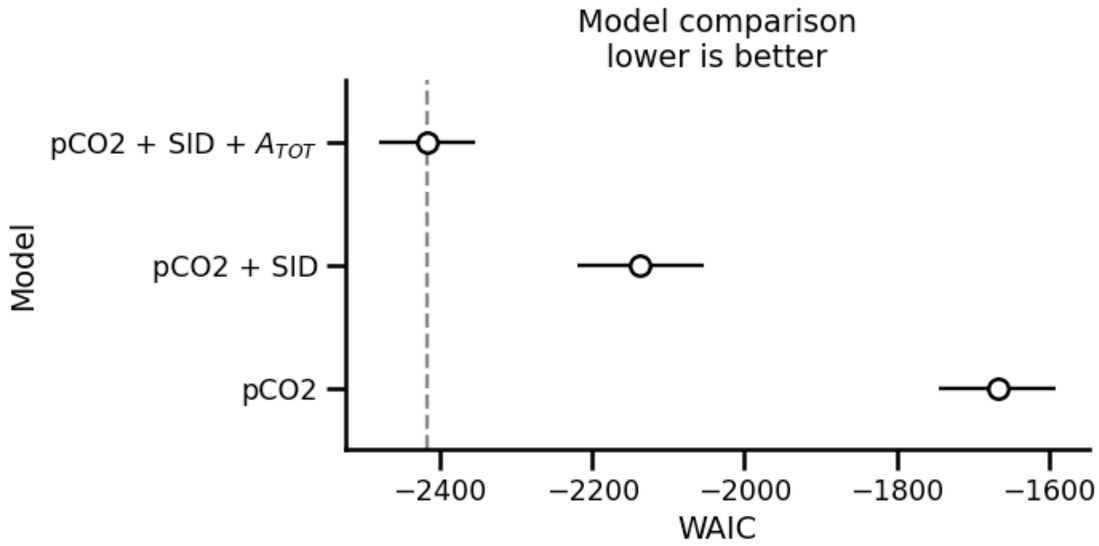
Issue

Irregular sampling of measured variables



Adding Covariates

Last known value imputation



Imputing Missing Values

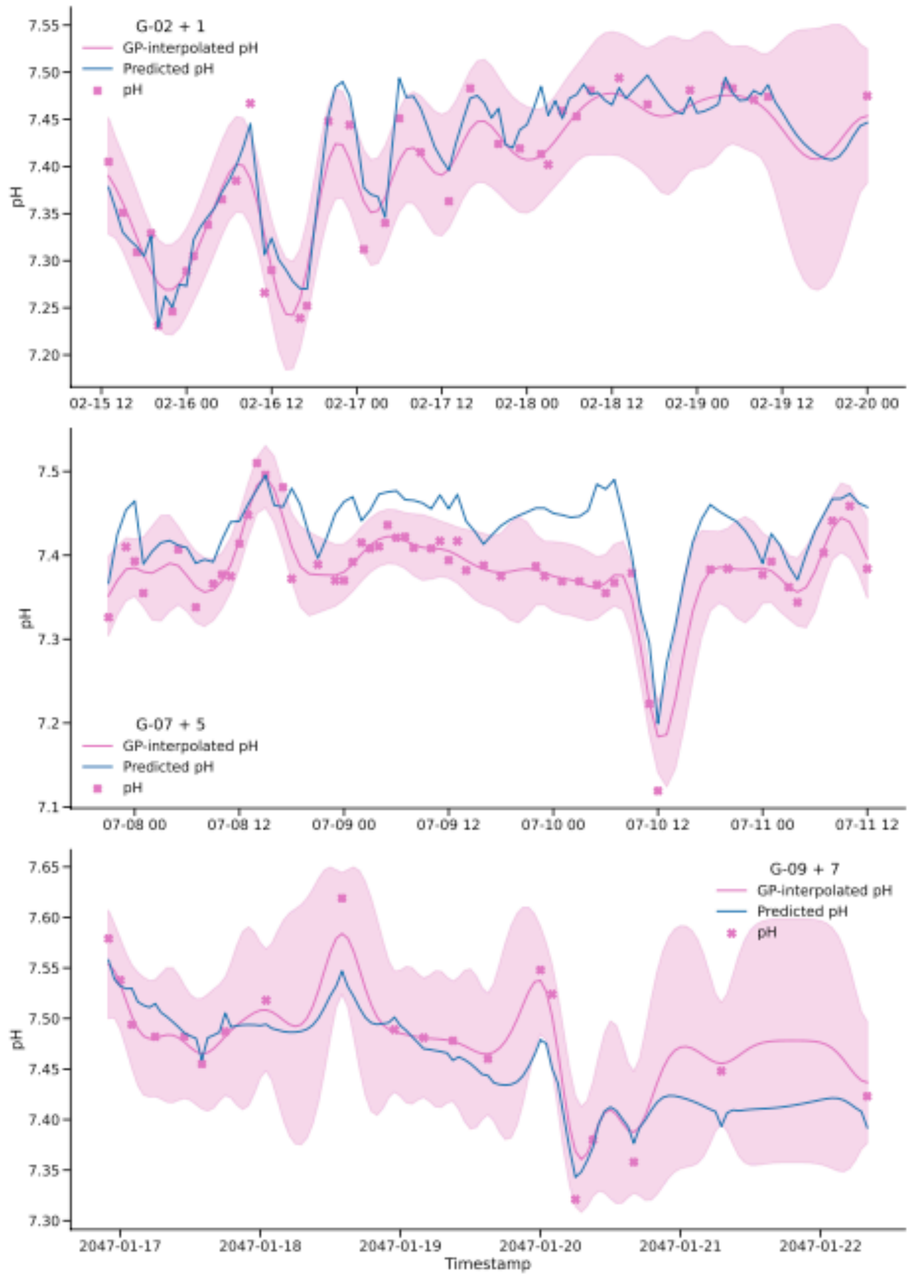
Benchmarking Imputation Methods

- Last-known value imputation
- Multiple imputation using chained equations (MICE)
- Gaussian process (GP) interpolation
- 3D-MICE: Combining MICE and GP

Cross-Validation

Leave-One-Patient-Out

- Pink line is GP-interpolated pH
- Blue line is OLS on interpolated covariates



Impact of imputation methods on MAE when predicting pH

With the toy dataset, it is still too early to say which model is the best

Window	n	Last known	GP	MICE	3D-MICE
G-01 + 0	55	**0.0322**	0.0325	0.0345	0.0345
G-02 + 1	41	0.0225	**0.0217**	0.0237	0.0281
G-03 + 2	43	0.0338	**0.0333**	0.0375	0.0389
G-04 + 3	17	0.0331	0.0359	**0.0247**	0.0373
G-05 + 3	67	0.0625	0.0616	0.0576	**0.0540**
G-06 + 4	56	0.0738	0.0741	**0.0672**	0.0708
G-07 + 5	54	0.0497	0.0525	0.0430	**0.0429**
G-08 + 6	9	0.0215	0.0203	0.0121	**0.0115**
G-09 + 7	23	0.0300	**0.0276**	0.0356	0.0342
G-10 + 8	124	**0.0346**	0.0385	0.0453	0.0456
G-10 + 9	84	**0.0249**	0.0275	0.0274	0.0355
G-11 + 10	13	0.0462	0.0507	**0.0442**	0.0557
G-12 + 11	50	0.0278	**0.0254**	0.0391	0.0415
G-12 + 12	30	**0.0232**	0.0232	0.0299	0.0303
G-12 + 13	23	0.0317	0.0340	**0.0298**	0.0314
G-13 + 14	74	0.0551	0.0582	**0.0374**	0.0440
G-14 + 15	69	**0.0957**	0.0976	0.1012	0.1018
Average	-	0.0411	0.0420	**0.0406**	0.0434

Discussion

1. The statistical approach aligned well with the physicochemical one
2. It remains inconclusive to state one imputation method is superior to the others
3. We did not propagate the uncertainties stemming from the imputation methods when predicting pH
4. No autoregressive components in the pH model

Future Work: Benchmarking

1. Developing a synthetic dataset
2. Generalisability on publicly available datasets
3. Using different downstream tasks

Future Work: Uncertainty Estimation

1. Extending Gaussian processes (GPs)
    a. Multi-task GPs (MTGPs)
    b. Deep GPs
    c. Computationally aware: Sparse GPs

Future Work: Uncertainty Estimation

2. Bayesian deep learning
    a. M-RNN (Yoon et al., 2017) & GRU-D (Che et al., 2018)
    b. Post-hoc:
        a. Last Layer Laplace Approximation
        b. Epistemic Neural Networks

Thank you