

Something's Missing!

Data Imputation in Critical Care Medicine

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

Chemistry Refresher

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 $pH = 7.0$

Water & Alkalinity

- At 0°C: pH = 7.5 (alkaline)
- At 100°C: 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.

Motivation

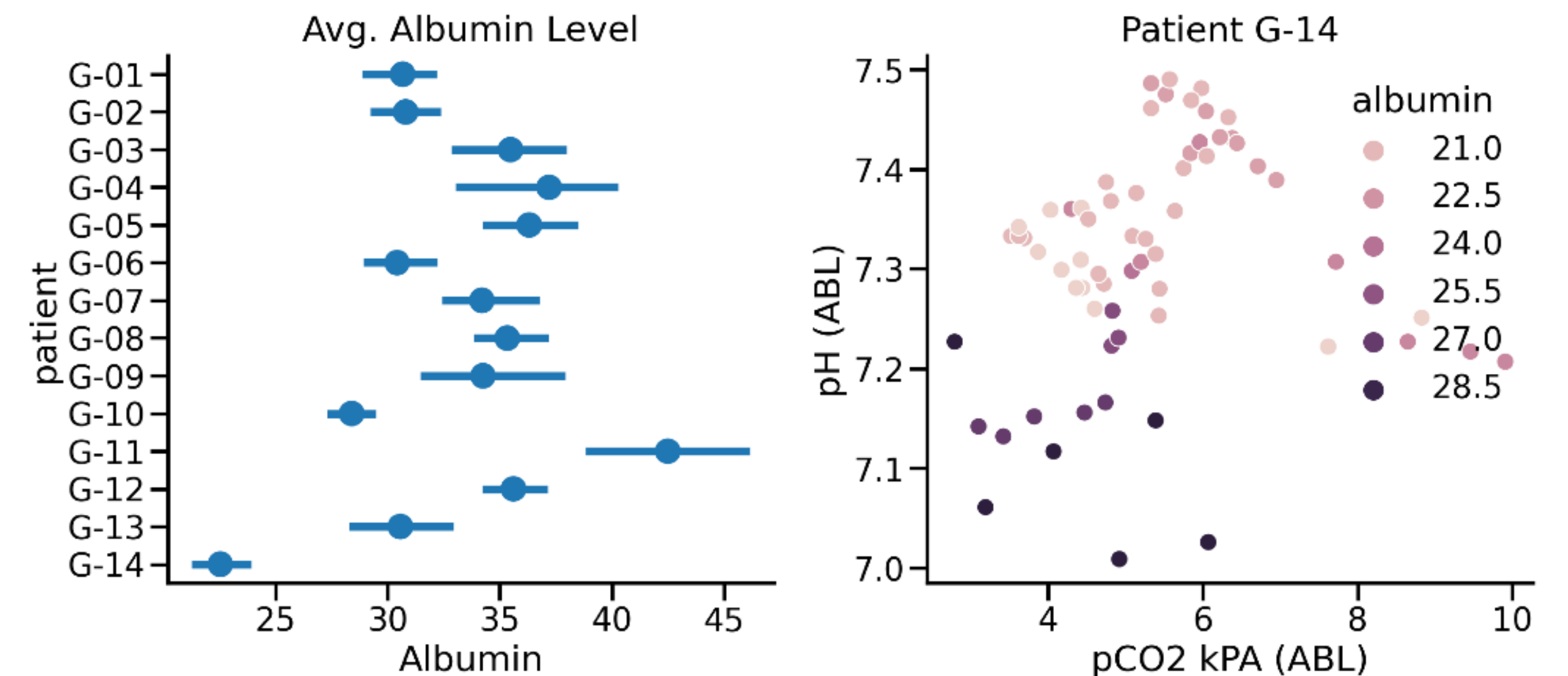
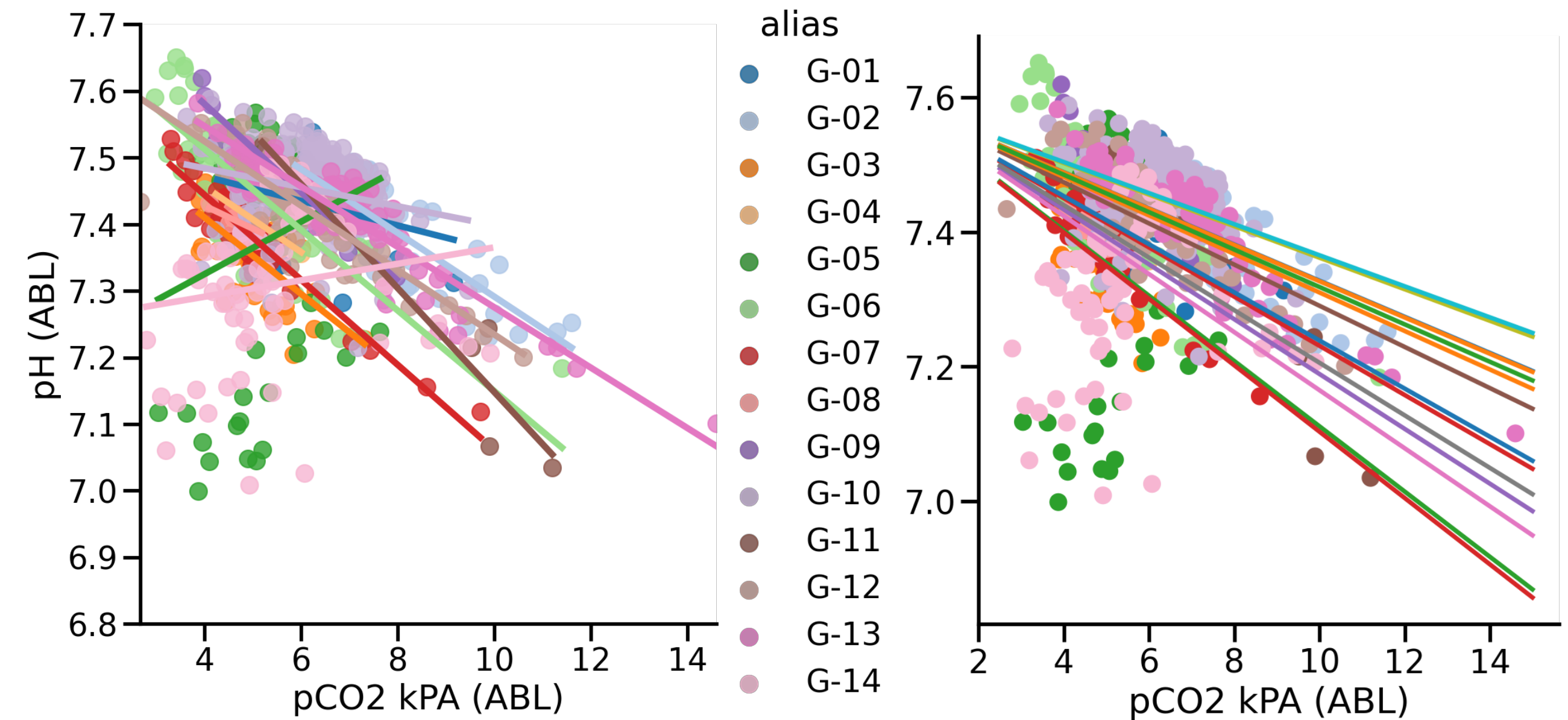
Issues

- Concentration of CO₂ can be easily monitored
- It is not enough to tell the whole story
- Other variables are collected in different frequencies

Issue

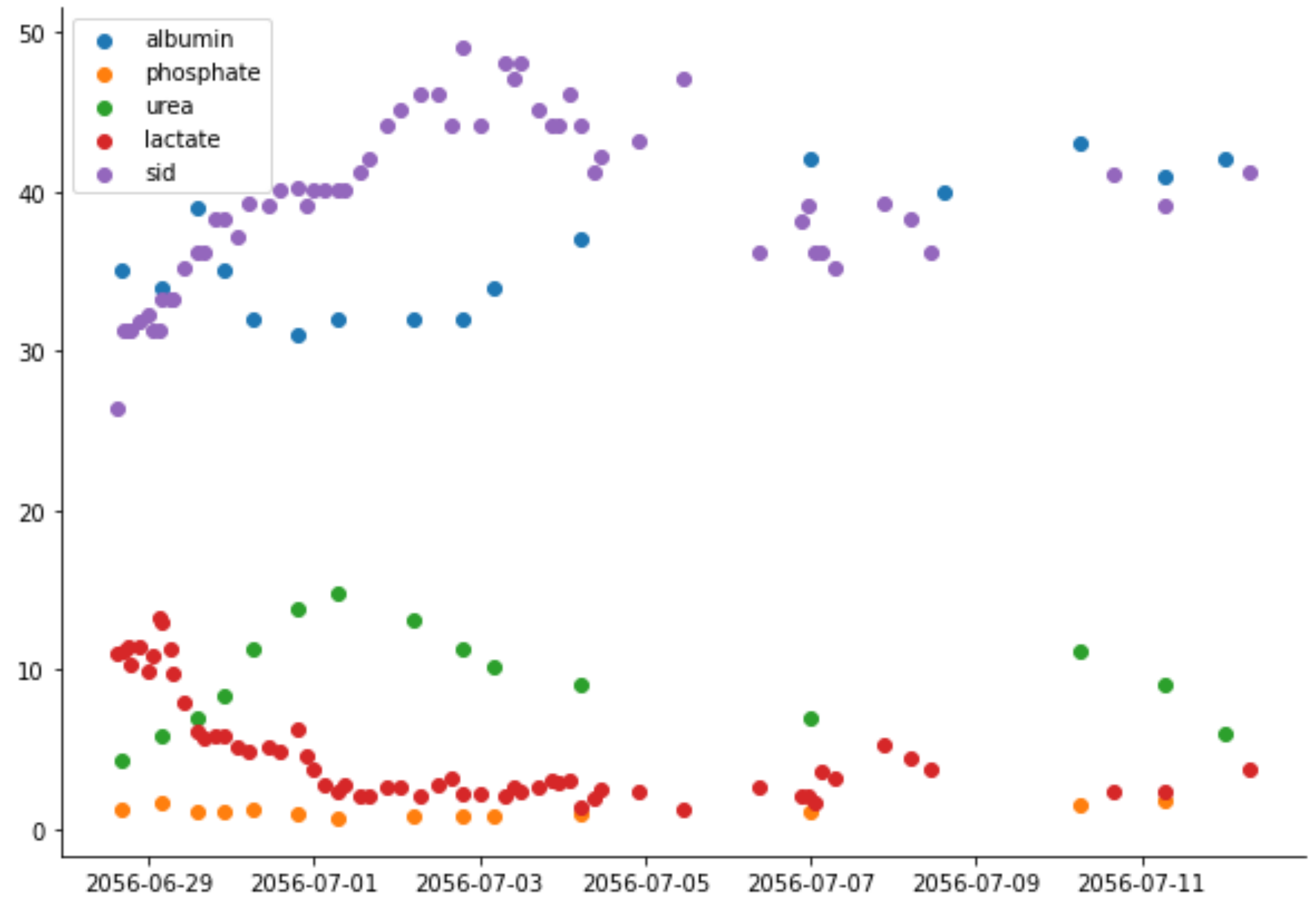
CO₂ is not enough to tell the whole story

- Model: $\text{pH} \sim \text{pCO}_2$
 - 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 animals occurs



Issue

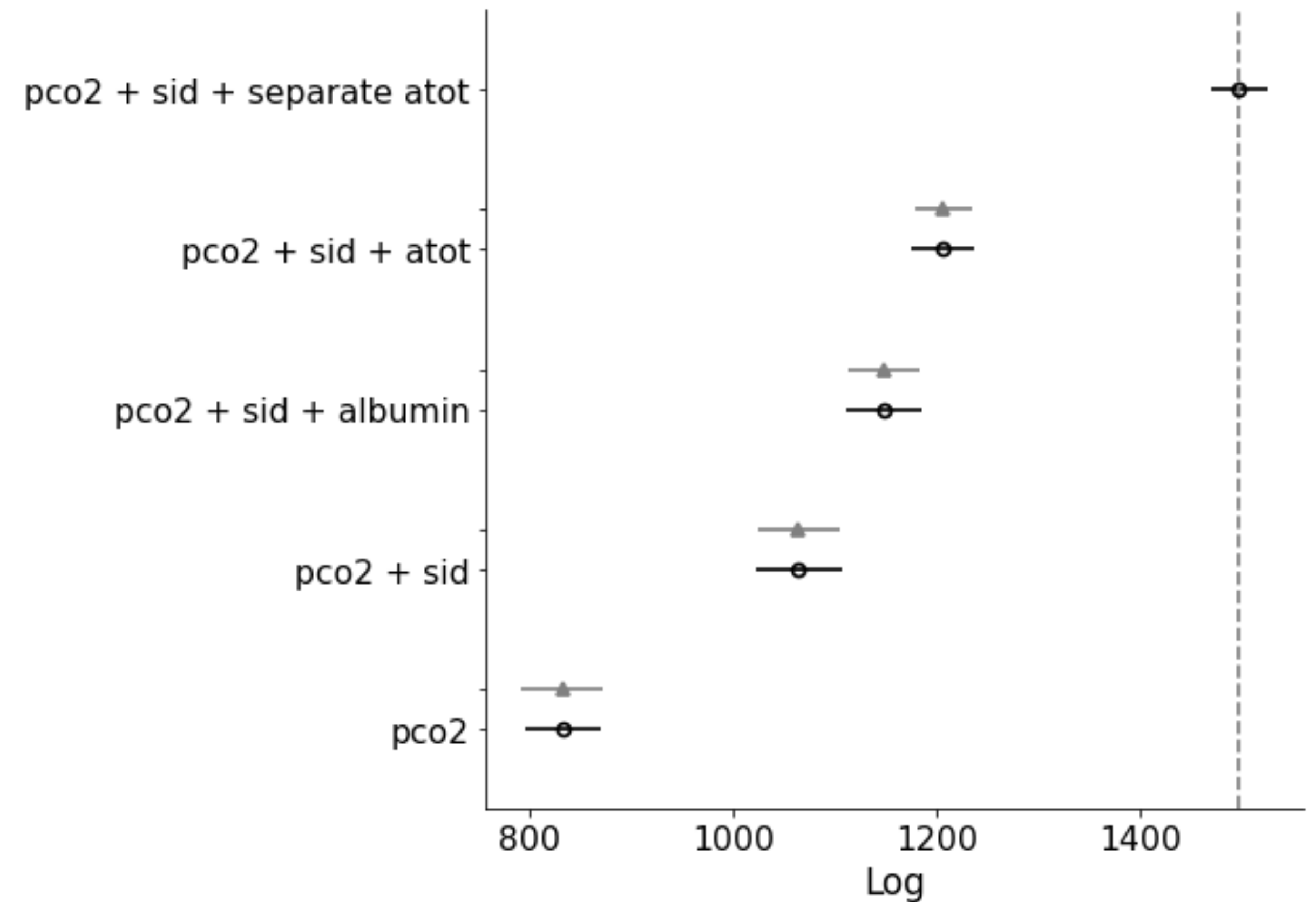
Data comes in different frequencies



The healthier the patient, the less data you would see

Adding Covariates

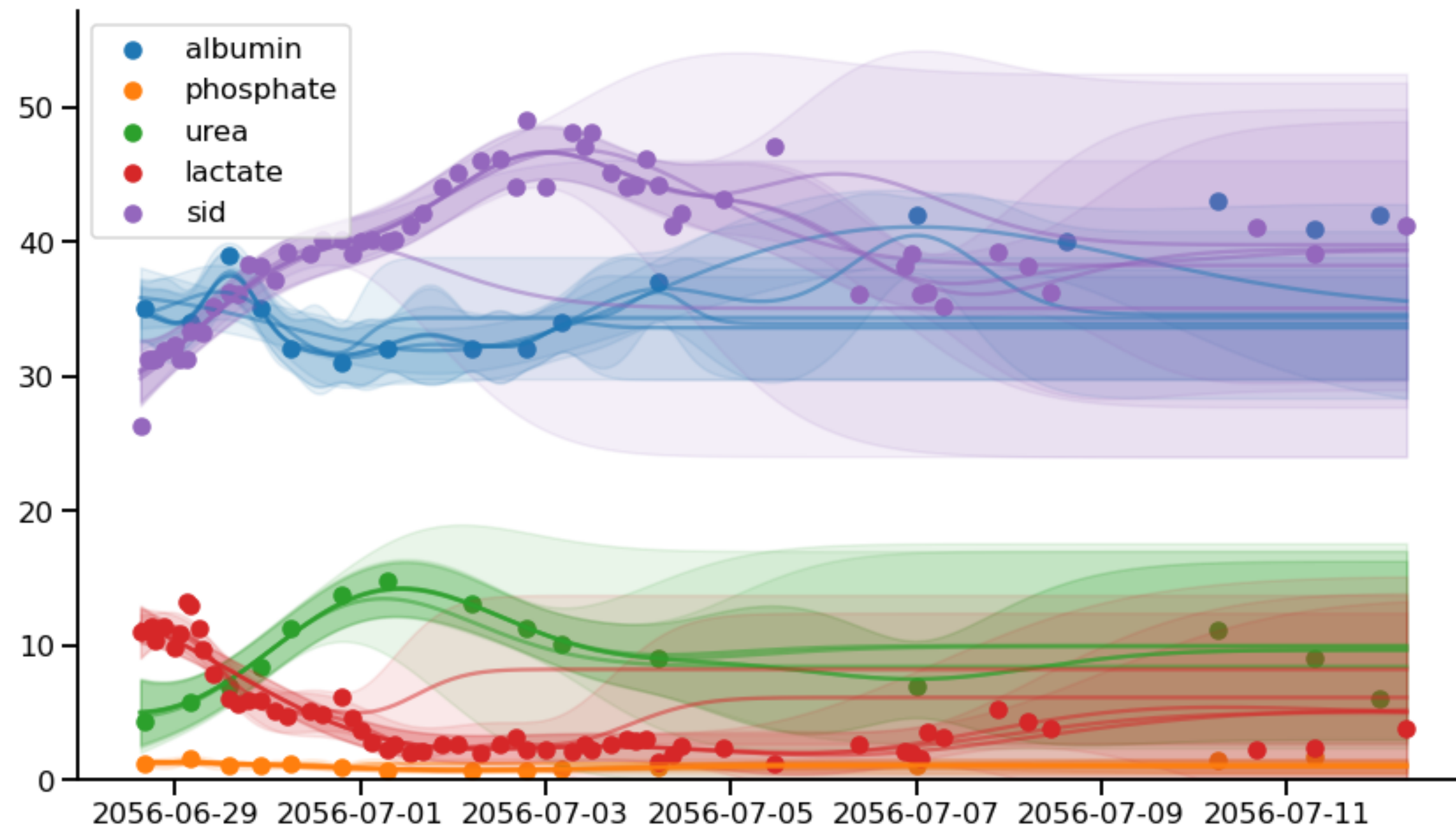
Last known value imputation



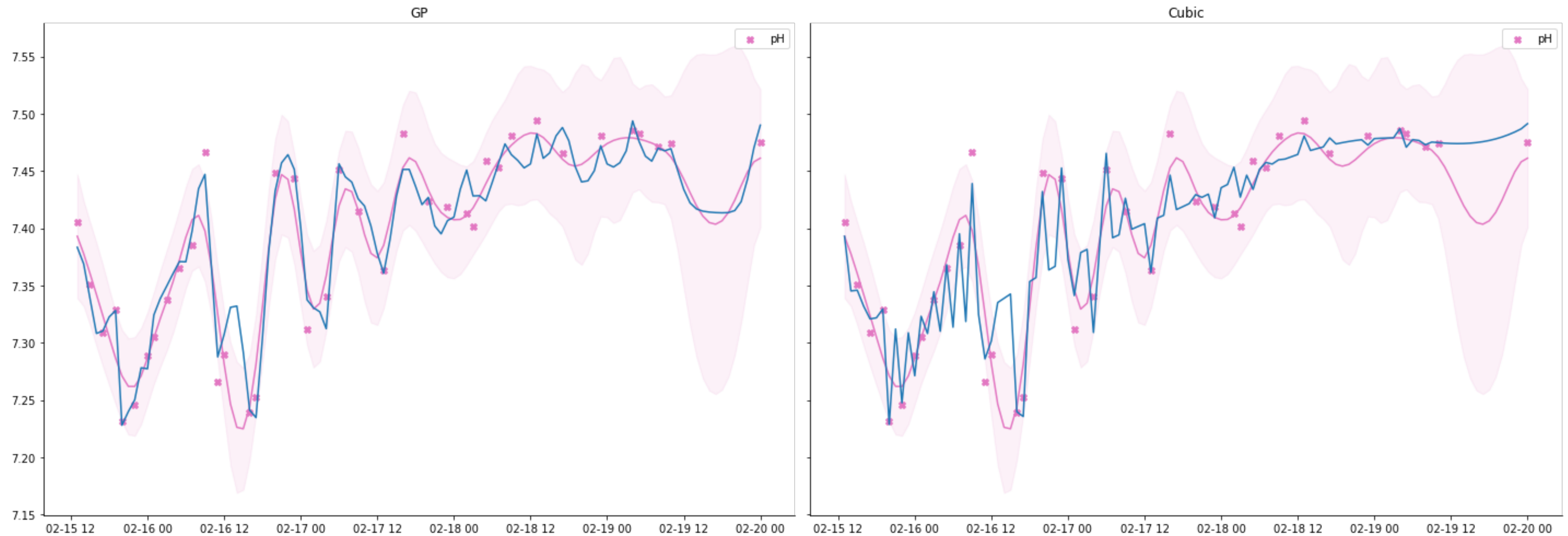
Log pointwise predictive density (Vehtari et al., 2017)

Proposal

Time Series Cross-Validation with GP



GP vs Cubic Spline Interpolation

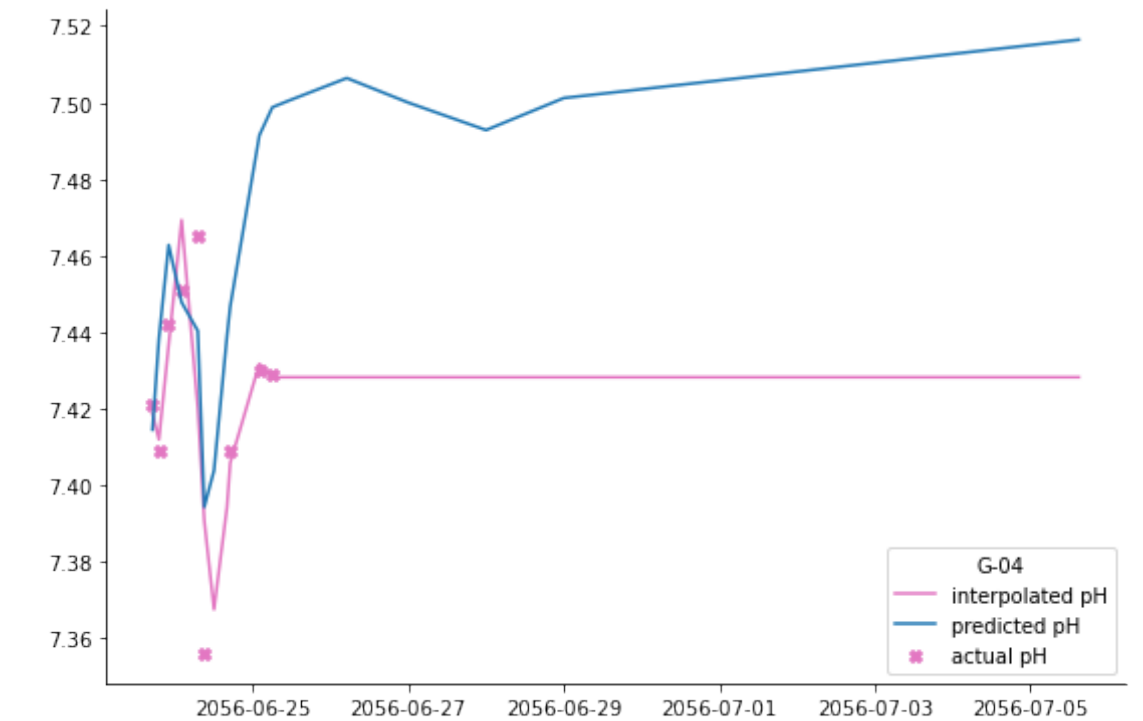
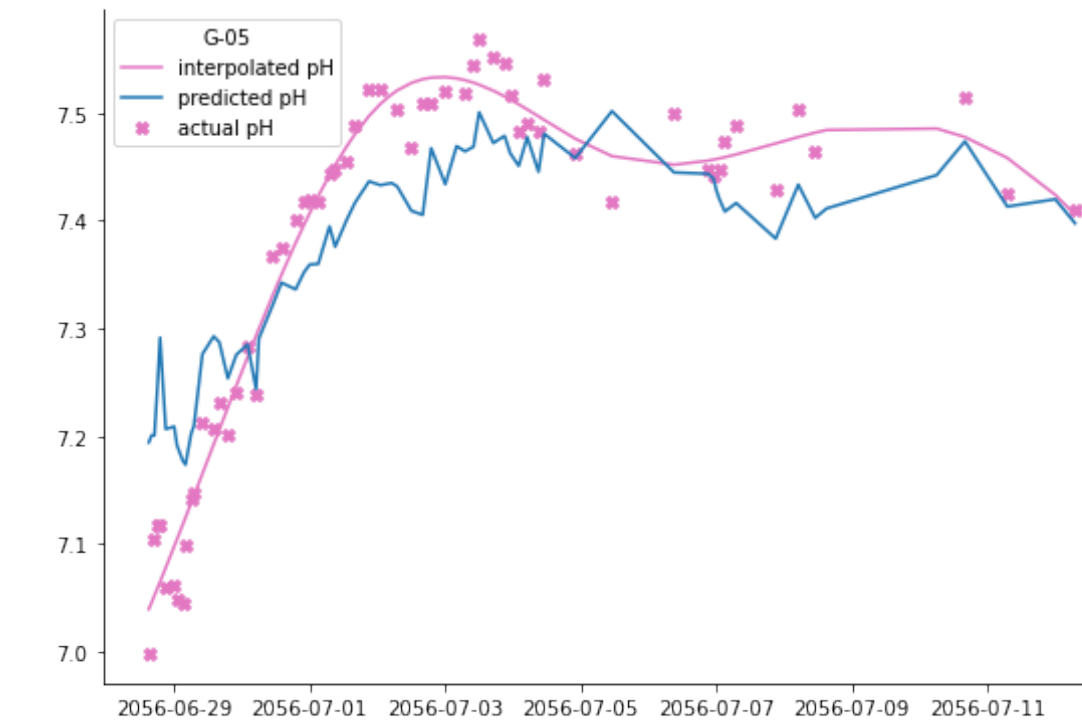
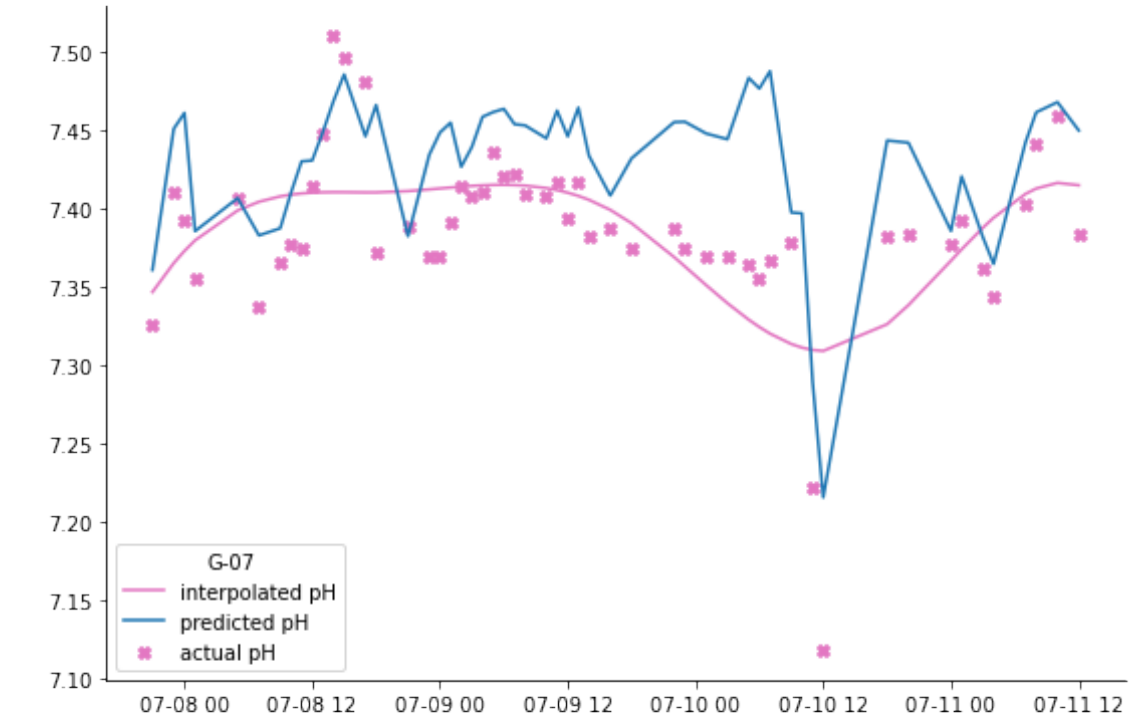
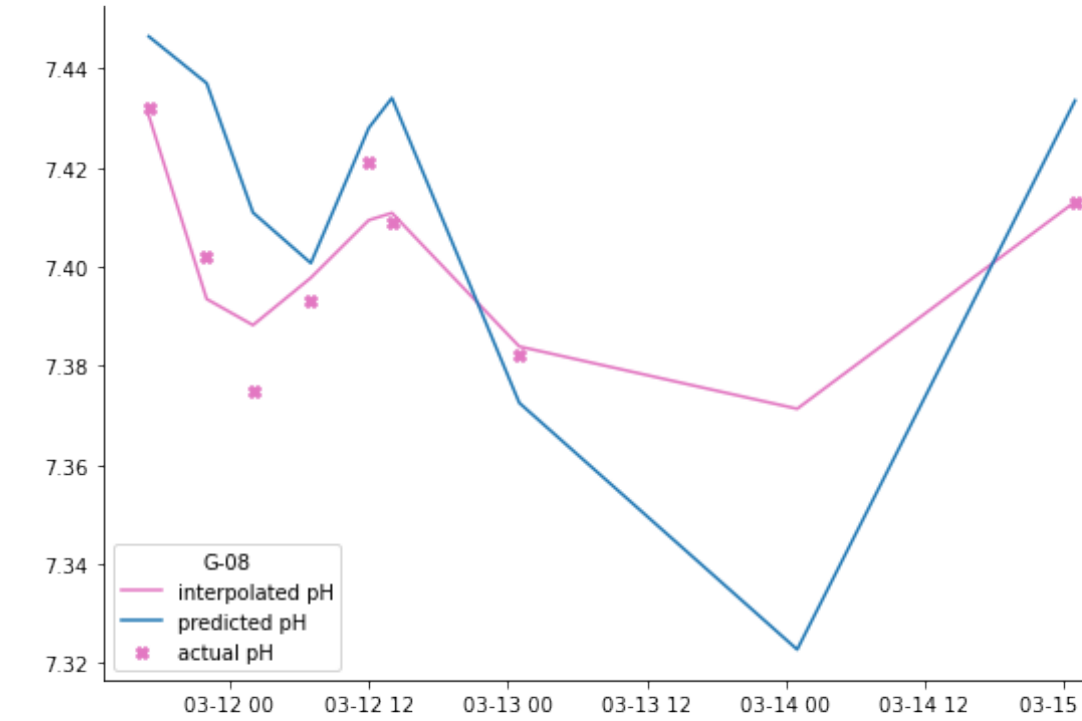
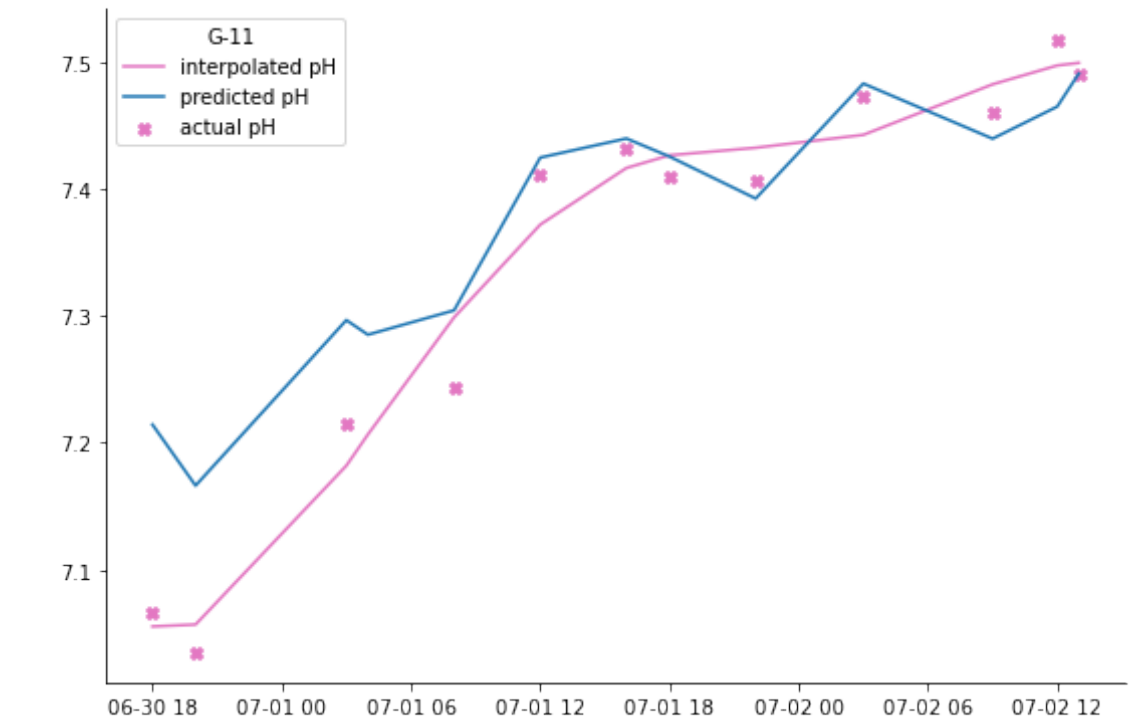
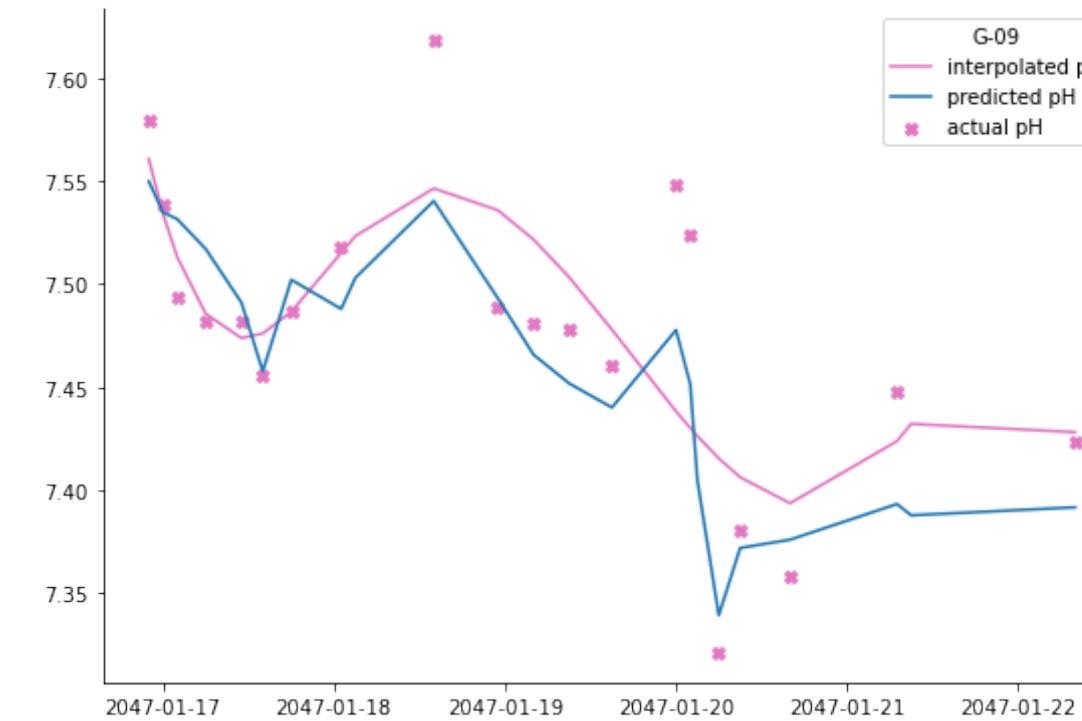


Interpolate the covariates and predict the pH

Cross-Validation

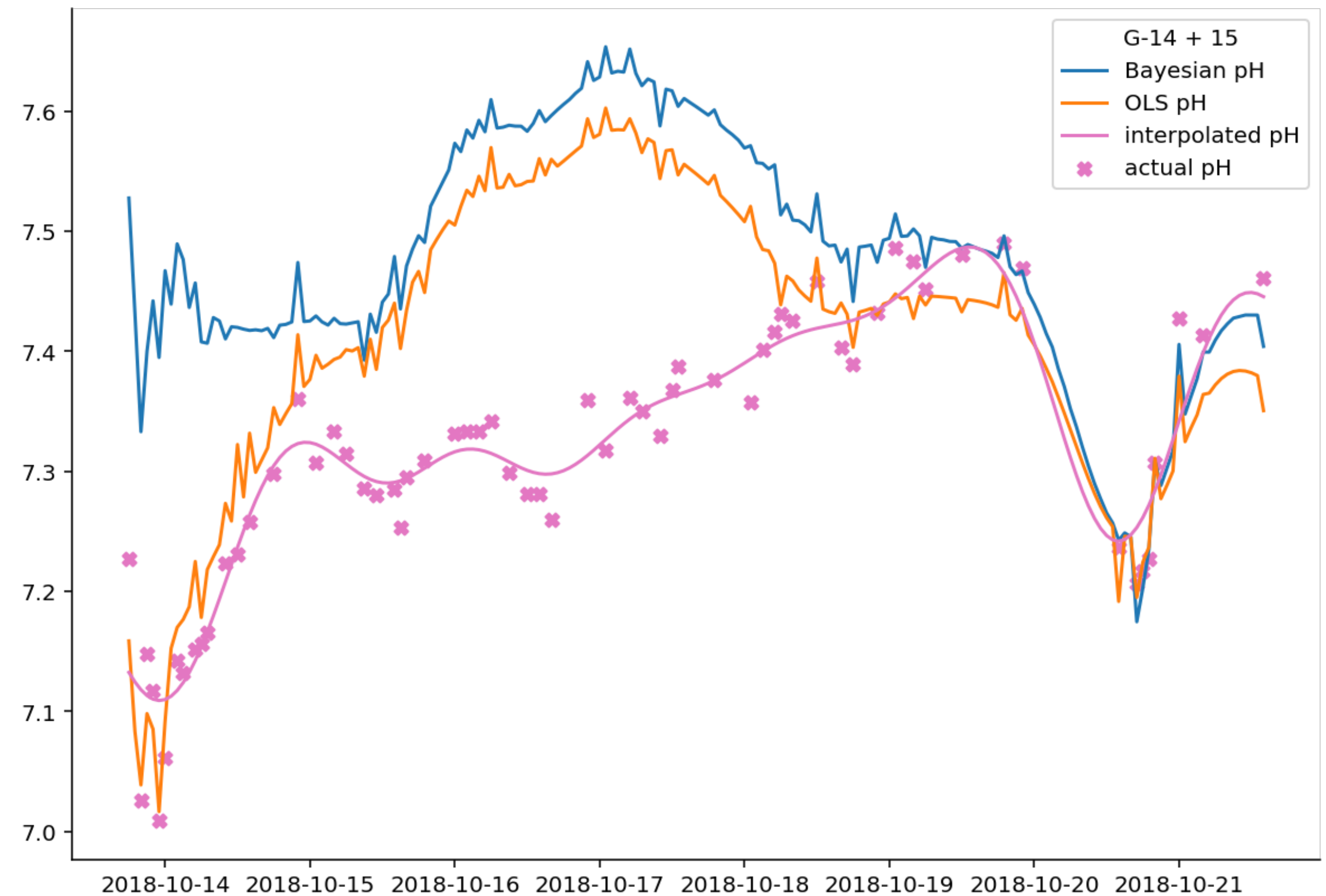
Leave-One-Patient-Out

- Pink line is GP-interpolated pH
- Blue line is OLS on interpolated covariates
- The difference is not significant



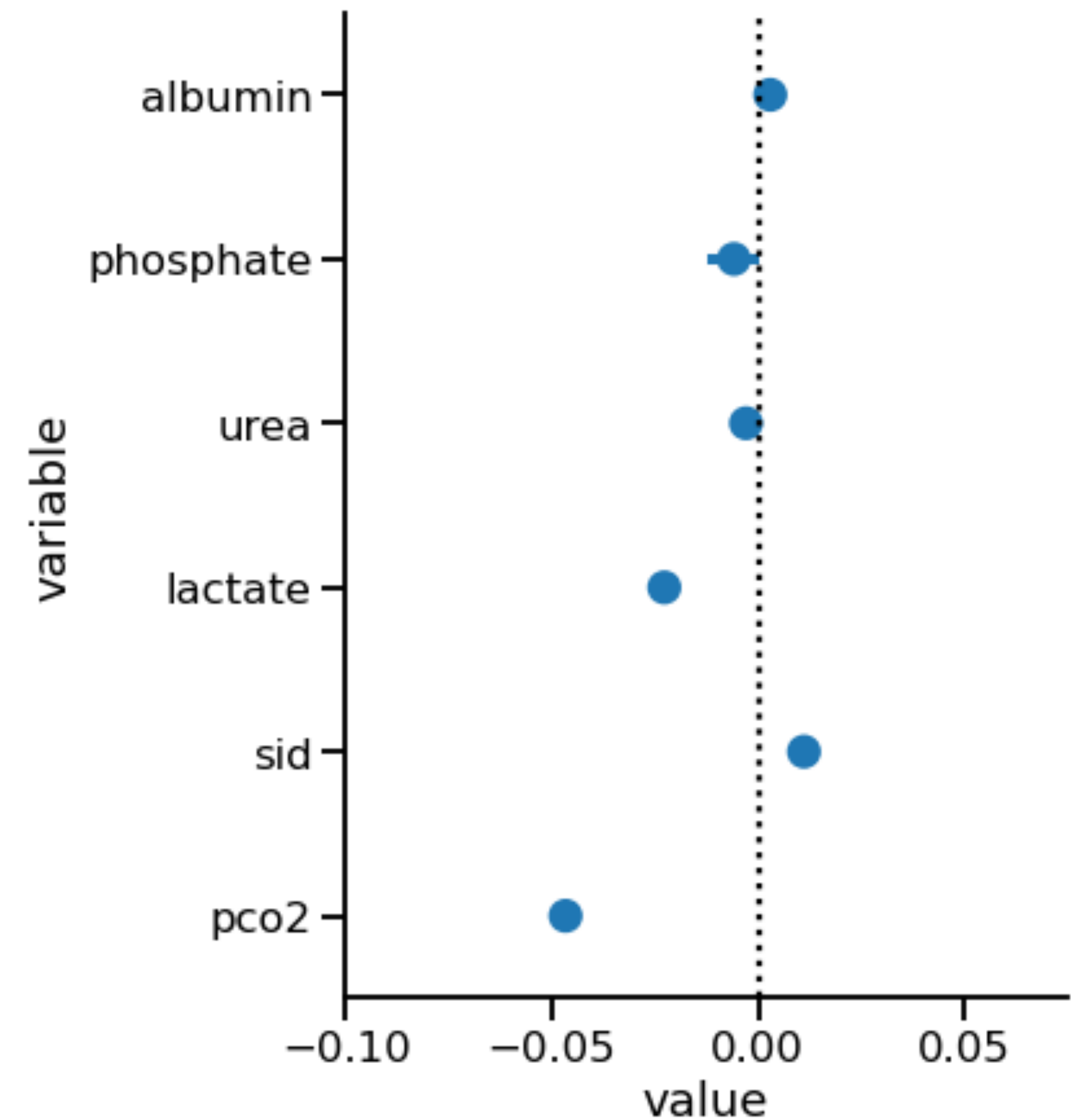
Systematic Issue

We might still be missing something



Systematic Issue

Albumin should not have a positive slope



Future Work

- Find a way to incorporate uncertainties in the covariates into the final model
- Explore some more interesting methods: Sparse Gaussian Process, Hierarchical model with AR(1)
- Confirm with the physiochemical approach

Thank you