26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Clustering Irregular Cytokine Time Series Using Gaussian Process Mixture Models

Samantha Beard [a]

[a]*School of Computer and Mathematical Sciences, University of Adelaide, Adelaide, SA, Australia*
Email: samantha.beard@adelaide.edu.au

Abstract: Monitoring cytokine responses over time provides critical insight into immune activation, but statistical analysis is challenged by irregular sampling, noise, and missing data in clinical studies. Conventional clustering approaches assume synchronised measurements and complete data, raising doubts about their suitability for such settings.

This abstract presents a methodological framework for clustering irregular cytokine time series using Gaussian Process Mixture Models (GPMMs), applied to longitudinal data from the Human Integrated Sensor System (HISS) study. The work addresses a relevant challenge in biomedical data analysis—namely, the difficulty of clustering noisy, irregular, and incomplete time-series data from clinical studies. In total, 126 participants (24 Adelaide, 102 Melbourne) received the influenza vaccination and were monitored with repeated cytokine and physiological measurements (Table 1). Conventional approaches—including principal component analysis, k-means, and hierarchical clustering—were first evaluated on cytokine+physiology versus physiology-only datasets. Cluster stability was assessed using supervised classification (logistic regression, random forest), with predictive accuracy of only ≈60–70%, confirming poor reproducibility. Including cytokines yielded weak and inconsistent cluster structures, while physiology alone provided negligible discriminatory power.

To address these limitations, GPMMs were applied. Unlike vector-based methods, GPMMs explicitly model temporal structure and accommodate non-matching observations across individuals. The approach recovered three coherent immune response clusters, providing stable subgroupings that conventional methods failed to detect. In comparison, traditional clustering produced up to eight unstable groups, whereas GPMMs consistently recovered three coherent clusters with stable membership. Cluster recovery was quantitatively evaluated using a chi-squared statistic, confirming that smaller chi-squared values corresponded to more influential time windows for accurate clustering. An occlusion analysis further showed that early post-injection samples (0.25–1 hour) and the 8-hour measurement were most influential for the Adelaide cohort, whereas removal of the combined 7–8-hour window was most influential for the combined cohort, highlighting critical windows for study design and monitoring. The influence of each timepoint across cohorts is visualised in Figure 1.

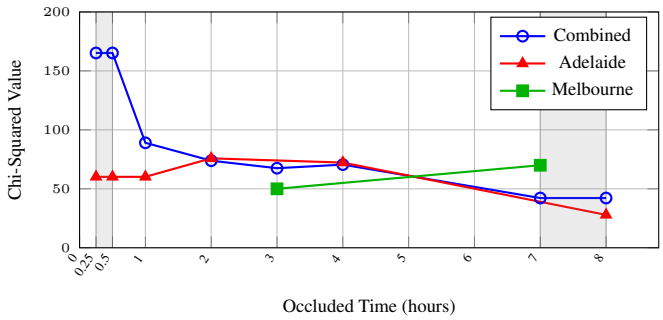

Figure 1. Chi-squared values by occluded timepoint for each cohort. Smaller values indicate more influential windows for cluster recovery.

Table 1. Cohort characteristics and clustering results. Key timepoints and stability highlighted.

Cohort	n	Timepoints (hrs)	Key Results
Melbourne (single-arm)	102	3, 7	Stable clusters at 3hr
Adelaide (crossover)	24	0.25, 0.5, 1, 2, 4, 8	Consistent; 0.25–1hr and 8hr most influential
Combined	126	0.25, 0.5, 1, 2, 3, 4, 7, 8	Stable; 7–8hr windows most influential

This work demonstrates that meaningful immune subgroups can only be recovered when temporal structure is respected, highlighting the inadequacy of static methods. The findings have direct implications for immunological research, informing both methodological strategies and the prioritisation of critical sampling intervals.

Keywords: *Cytokine clustering, Gaussian Process Mixture Models, temporal data, immunology, longitudinal analysis*

© MSSANZ and The Authors, CC-BY