

# 國立中山大學應用數學系

## 學術演講

講者：張源俊 教授（中央研究院統計科學研究所）

講題：It's All Samples' Fault: PCA-Quantile Sampling for Smarter Data Reduction

時間：2026/05/22 (Friday) 14:10 ~ 15:00

地點：理 SC4011 教室

茶會：13:30

### Abstract

This seminar presents PCA-Quantile Sampling (PCA-QS), a data reduction approach designed to address the computational challenges of large-scale data analysis while preserving the geometric and statistical structure of the original data. Unlike simple random sampling, PCA-QS uses principal component analysis to identify dominant directions of variation and then applies quantile-based stratification to select representative observations across the induced low-dimensional structure.

Through comparative studies using distributional and structural metrics such as Energy Distance, Mahalanobis distance, Maximum Mean Discrepancy, and Kullback–Leibler divergence, the talk illustrates how PCA-QS can better preserve the internal topology and feature distribution of the full data set than conventional random sampling. We also examine its impact on downstream machine-learning tasks, including Random Forest and XGBoost classification, where structure-preserving subsamples may improve predictive stability and reduce computational burden.

The talk emphasizes PCA-QS not merely as a sampling shortcut, but as a statistically guided strategy for retaining essential information in modern data analysis.

This is joint work with Dr. Hui-Mean Foo.

## 敬請公告！歡迎參加！

應用數學系：<http://math.nsysu.edu.tw>

校園地圖：[http://math.nsysu.edu.tw/var/file/183/1183/img/779/nsysu\\_math\\_map.jpg](http://math.nsysu.edu.tw/var/file/183/1183/img/779/nsysu_math_map.jpg)

交通資訊：<https://www.nsysu.edu.tw/p/412-1000-4132.php?Lang=zh-tw>



用數學系



校園地圖



交通資訊