

PM2.5 時空資料之異常檢測

王舜弘

國立中山大學應用數學所

摘要

台灣過去偵測細懸浮微粒 (PM2.5) 主要是環保署觀測站的工作，但因其儀器昂貴，全台僅有七十餘站無法涵蓋太多地區。近年來成本低廉的微型感測器這類新型設施加入後，涵蓋站點成長百倍以上。但微型感測器的記錄不穩定，時有異常狀況出現。本研究希望藉助時空預測模型提供一種偵測微型感測器異常數據的方法。空間資料各位置間通常具有空間相關性，無法直接當作獨立樣本分析，本研究的手段是透過時空預測模型並計算殘差，然後利用這些殘差找出可能的異常時間段或可能發生故障之微型感測器。

本研究按月分析 PM2.5 濃度資料各別進行時空模型配適，此資料極為龐大且雜亂，因此分成數個步驟處理。我們分別使用三種時空預測模型 (Weighting by Inverse Distance with Adaptive Least Squares、Spatio Temporal Bayesian Modelling、Fixed Rank Kriging) 進行交叉驗證，得出預測的均方誤差作為評估模型的準則，選擇最佳模型用以計算其殘差。最後，我們使用最佳模型進行函數型主成分分析 (Functional Principal Component Analysis) 並使用 EM 演算法 (Expectation-Maximization) 對其結果進一步分群，再於各群分別使用時間序列相似性聯接 (Time Series All Pairs Similarity Search) 找出可能發生異常的時間段及測站。

關鍵詞：PM2.5、時空預測模型、函數型主成分分析、時間序列相似性聯接