作業報導

# ●跨域巨量資料應用－以大臺北地區交通管理對空氣品質之影響為例

行政院環境保護署環境監測及資訊處設計師　莊惟然

## 前言

空氣污染一直是全球各地所關心的議題，針對環境科學方面，利用巨量分析來對環境空氣品質分析研究與預測是刻不容緩的工作。而交通工具一直是空氣中移動污染源的主要來源之一，尤其大臺北都會區更是如此。因此，我們以大臺北環境空氣品質為分析研究對象，整合包含交通車流、車速、氣象因子與各空氣污染物濃度等各種不同領域的巨量資料，以非線性統計分析角度建立彼此間的關聯性模型，以分析交通因子對於空氣品質的影響，提供政府有效的交通與環境管理政策方針之建議。

## 資料來源

### 行政院環境保護署各空氣品質監測站2001-2016年小時尺度各種空品污染物資料。

### 行政院環境保護署各空氣品質監測站2001-2016年小時尺度各種氣象因子資料。

### 交通部中央氣象局各測站2001-2016年小時尺度各種氣象因子資料。

### 臺北市交通局提供之2012-2016年車輛偵測器Vehicle Detector （VD）小時尺度資料。

經資料清理分析，處理總資料筆數約達一億七千萬筆資料，並從中萃取有用資訊進行後續分析。

## 實作工具／方法

為分析多個不同因子間的非線性關係，我們運用廣義可增式模型（Generalized Additive Model，GAM）建立交通與氣象因子對於空氣污染的非線性模型。有別傳統統計模型，GAM為一種強大且實用的非線性統計迴歸模型。一般可如下式所示：

$$E\left(Y,X\_{1},X\_{2},…,X\_{p}\right)=α+f\_{1}\left(X\_{1}\right)+f\_{2}\left(X\_{2}\right)+…+f\_{p}\left(X\_{p}\right)$$

式1：廣義可增式模型

實作上，我們以開源統計程式軟體R之mgcv套件進行GAM之模式建立。R為一開源自由軟體程式語言，具有眾多強大的統計分析套件、繪圖與完整的程式語言功能，因此能夠批次進行大量地分析，特別適用於統計分析建模、巨量資料分析、資料探勘與科學計算。R語言程式介面如下圖1所示。



圖1：本研究工具R語言程式介面

## 分析與探討

以大臺北都會區古亭地區實測的NOx為dependent variable為例，將古亭空品測站NOx、風速等資料以及交通局古亭車輛偵測器（VELJA00羅斯福路三段雙向各四線道）之車速車流量資料彙整，建立GAM方法以了解車流、車速對空氣品質（NOx）的非線性影響。

分析結果如圖2顯示，可看出所有解釋因子的p-value皆遠小於0.05，統計上皆為顯著。自由度皆為6.6至8.8不等，明顯皆為非線性關係，因此適宜採用本GAM模型。其模式整體解釋能力約為49.1% （圖2）。圖3結果顯示為各因子之偏平滑函數，透過此圖可分析各因子在模式擬合之下對於應變數的貢獻變化。其中X軸為各因子原始值，Y軸可表示為該因子的模式平滑函數。



圖2：以summary()函數輸出GAM計算結果

整體來說，NOx明顯隨車流量增加而遞增，從上升斜率可以看出，當車流量從0（輛／小時，以下略）至750之間與大於約1,400，隨車流增加，NOx增加的幅度最大，即在此區間內車流量增加少許即會對於NOx有較大的影響。因此，車流量若能控制在至少1,400以內，或更嚴苛控制在1,144以內，將對空品有最大的幫助（圖3）。



圖3：GAM分析結果車流對NOx關係圖

而車速分析顯示在平均車速約30.4 km／hr達到最高峰，對NOx有最大的負面影響，而後車速增加對於NOx的影響逐漸遞減，至45 km／hr後又開始升高（圖4）。故以此地市區速限50 km／hr道路而言，平均車速盡可能控制在38至50 km／hr 之間，對於空氣品質將有最好的幫助。



圖4：GAM分析結果車速對NOx關係圖

而氣象因子則相對單純，風速影響呈現接近線性遞減，風速越大則NOx濃度越低（圖5）。



圖5：GAM分析結果風速對NOx關係圖

## 結論

本研究使用GAM之分析方法，分析車流量、車速對於NOx濃度之非線性影響，並能量化車流與車速對於空氣品質正負影響的關鍵值。本研究結果顯示，NOx隨著車流遞增、隨車速遞減，除此之外關鍵數值可提供政府相關部門改善交通影響空氣品質之決策建議。

從本案例分析經驗，使用R語言從資料科學角度出發，以純粹實際資料面分析萃取有用的價值，這對環境空氣品質分析研究與預測是現今資訊科技日益重要的時代刻不容緩的工作。