Defense Scientific Information & Documentation Centre,

Defense Research & Development Organization

Delhi, India

Cluster Analysis of Air Pollution in Delhi

December 26, 2018 - January 20, 2019

Introduction and Aim

Air pollution monitoring in India began in 1978 when the National Environmental Engineering Research Institute (NEERI) established a network in 10 cities including Delhi. In each city, at least three sampling sites representing residential, commercial and industrial zones were selected (NEERI). Many years later the Central Pollution Control Board (CPCB) also began a monitoring programme. The main function of the CPCB under the Air (Prevention & Control of Pollution) Act 1981 is to improve the quality of air and to prevent, control and abate air pollution in the country. In order to assess the baseline situation, measure trends and evaluate the impact of interventions, it is essential to collect air quality data. Therefore, a National Ambient Air Quality Monitoring (NAAQM) programme was initiated in 1984 with 7 stations at Agra and Annapara. By the end of March 1993, 290 monitoring stations covering 92 cities/towns were operational. The monitoring stations have been located in different areas, viz., Residential/Rural (R), Industrial (I), and Sensitive (S) to cover the spatial variation in different cities. The last category includes monuments, hill stations, and sanctuaries. Currently, only the criteria pollutants, i.e. Sulphur dioxide (SO_2), Nitrogen dioxide (NO_2) and Suspended Particulate Matter (SPM) are regularly measured. As per the programme, monitoring for SO_2 and NO_2 should be conducted at 4-hourly intervals for 24 hours and for SPM at 8-hourly intervals for 24 hours on a bi-weekly basis (CPCB). SO_2 is measured by the Improved West & Gaeke method. NO_2 by the Jacob and Hocheiser modified method and SPM by the Gravimetric method. (Earlier, SPM was collected continuously on a 24hr basis, and gases on an 8 hourly batch basis, sequentially for 24 hours.). In addition, meteorological

parameters like wind speed, wind direction, temperature, and relative humidity are also measured at each station. Advanced statistical tests have been used to analyze the data (NEERI) for example, Daniel=s test based on Spearman's rank correlation coefficients for studying yearly variations.

In Delhi, there are 9 stations, shown in Table 1 operated by CPCB and 3 by the NEERI. Land-use patterns have changed significantly since CPCB adopted the land-use based classification system for its monitoring stations. More importantly, heavy traffic - a ubiquitous phenomenon - would contribute to air pollution even in non-industrial areas. Therefore, it is not safe to assume that ambient levels in industrial areas would necessarily always be higher than in commercial areas, with residential areas being the least polluted. The objective of our study was to determine if Delhi's sampling sites can be grouped together in any manner. Cluster analysis was used for sophisticated space-time modeling. Also, for a city of its size, the sampling network is not dense enough for space-time modeling. This study will provide a simple methodology for Indian researchers, practitioners, and regulatory authorities to do an exploratory study of spatial patterns and/or data quality issues of air pollution in Indian cities using the National Ambient Air Quality Monitoring System data. The results have serious implications in terms of improving the monitoring network, sampling strategies, and data quality. The results also have a regulatory significance in terms of moving towards a uniform ambient air quality standard for cities such as Delhi.

Table 1. Location of ambient air quality monitoring stations in Delhi

Station Code	Location	Operated by	Zone	Land-use type
C1	Ashok Vihar	CPCB	West	Industrial
C2	Janak Puri	CPCB	West	Residential
C3	Nizamuddin	CPCB	South	Industrial
C4	Shahdra	CPCB	East	Industrial
C5	Shahzada Bag	CPCB	East	Industrial
C6	Siri Fort	CPCB	South	Residential
N1	Town Hall	NEERI	Central	Residential
N2	Najafgarh Road	NEERI	West	Industrial
N3	Netaji Nagar	NEERI	South	Residential

Description of Study Region

Delhi is the most polluted mega-city of the world vis-a-vis air quality. The city is spread over 1483 km2. The Delhi Metropolitan Area that included satellite towns covers an area of 3182 km^2. More than three-quarters of the emissions of air pollutants are caused by vehicles. The total number of vehicles in 2016 was 9.71 million. Delhi has three big thermal power stations - all coal-based. While two of these are located within the city, the third is on the outskirts.

Factories are mainly located in west, north-west or east Delhi. There are certain non-polluting industrial estates in south Delhi. Industrial air pollution in Delhi is mainly due to the very large number of small-scale industries. Delhi has many satellite towns which are industrialized to a large extent. Major infrastructural units and large scale industries are located in this region. Delhi and its satellite towns form the National Capital Region - a large area source of pollution.

Methodology

Preparing for the Data Analysis

I used the monthly mean concentration of three criteria pollutants SO_2, NO_2, and SPM. The data were obtained from published sources (CPCB). Missing values were substituted. If, for example, for a particular month the value was missing, then it was substituted by taking the average of the preceding and succeeding months. This was done to preserve the seasonal patterns (as opposed to the effect of the procedure of substituting by the annual average.). In some cases where no value was reported for many consecutive months, the data corresponding to these months for all nine stations was dropped for that year's analysis.

Cluster Analysis

Cluster analysis involves splitting a data set into a number of groups of observations which are distinct in terms of typical group values of the variables. The aim is to maximize between-group variance and to minimize within-group variance. Cluster analysis is a classification technique where any number of variables may be used to classify members of the sample. The classification was done separately for each pollutant, viz., SPM, SO_2, and NO_2, though classification based on combinations of two or three pollutants at a time is also possible. Using monthly means for the nine stations meant that I worked on a 9 x 12 matrix

for each pollutant for each year. I used a hierarchical agglomeration algorithm for clustering. Two distance measures (Euclidean and Squared Euclidean) and four methods for combining clusters (average linkage between groups, single linkage, complete linkage, and centroid method) have been used.

Results

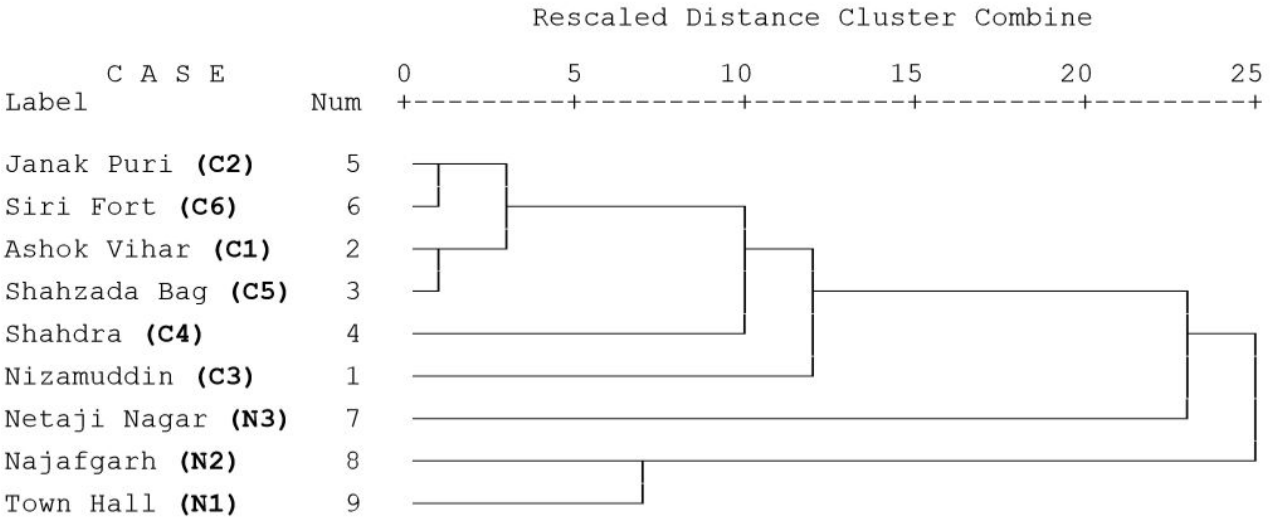
With SPM as the criterion, it was observed that two distinct clusters existed in the years 2012 – 2014 (Figure1). Najafgarh and Town Hall were members of a common cluster in all these years. Netaji Nagar was a member in 2014 but was the nearest neighboring cluster in 2012 & 2013. I observed that Janak Puri and Siri Fort – the two residential areas - were clustered together, and they're combined annual mean was higher than that of the combined mean of the four industrial area stations. We speculate that this may be due to their being downwind of sources of SPM.

It is evident from Figure 2 that with respect to SO_2 two clusters existed in all years. Netaji Nagar was always in the distinct cluster – on its own in 2012 and 2014 and with Siri Fort in 2013. However, in the initial years (2006 -2009) three stations (Town Hall, Najafgarh, and Netaji Nagar - all operated by NEERI) were clustered together. In the later years Town Hall (an area characterized by high traffic density) emerged as a distinct cluster in 2010 and in 2011 it was clustered with Najafgarh. Also, two clusters were observed, with Shahzada Bag forming a group of its own.

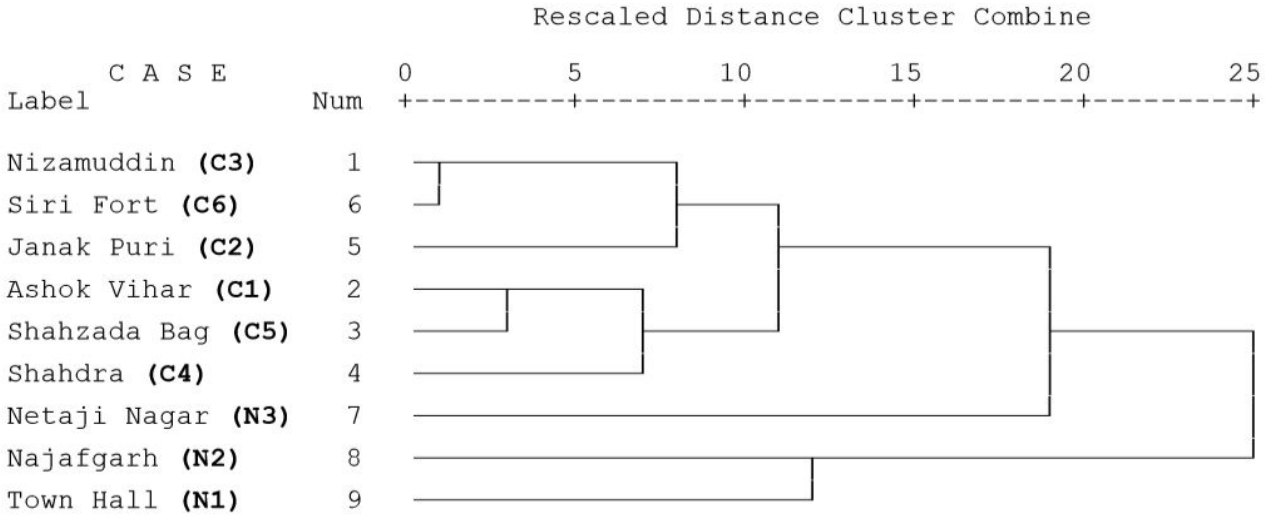
Using NO_2 as a clustering criterion we observed that in all years two distinct clusters were formed (Figure 3). The three NEERI stations were clustered together in 2014 and 2013. Before 2013 Town Hall formed a cluster of its own.

Dendrogram for SPM data

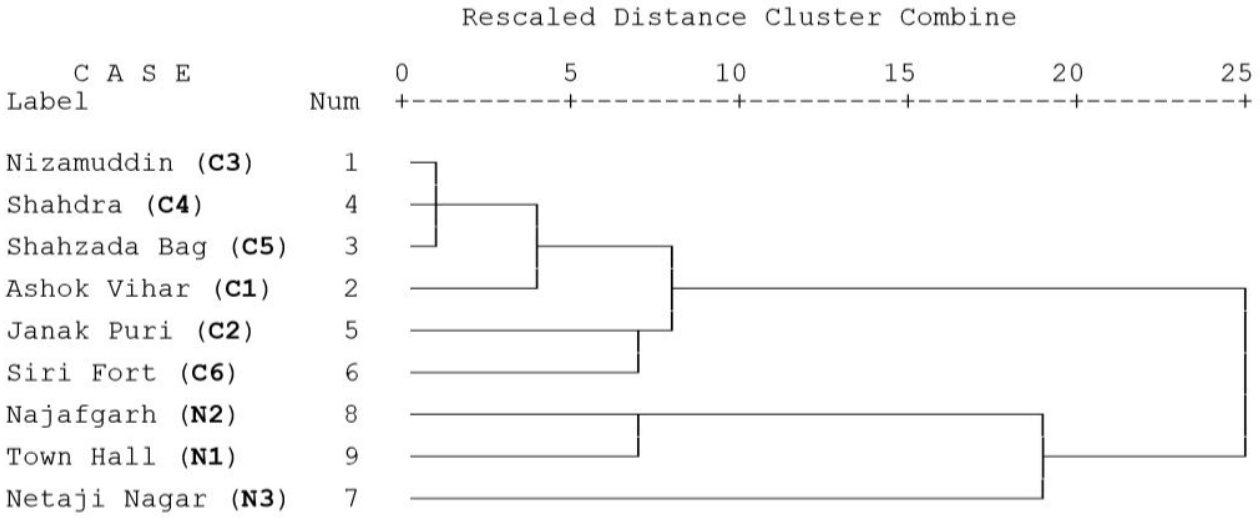
a) 2012



b) 2013

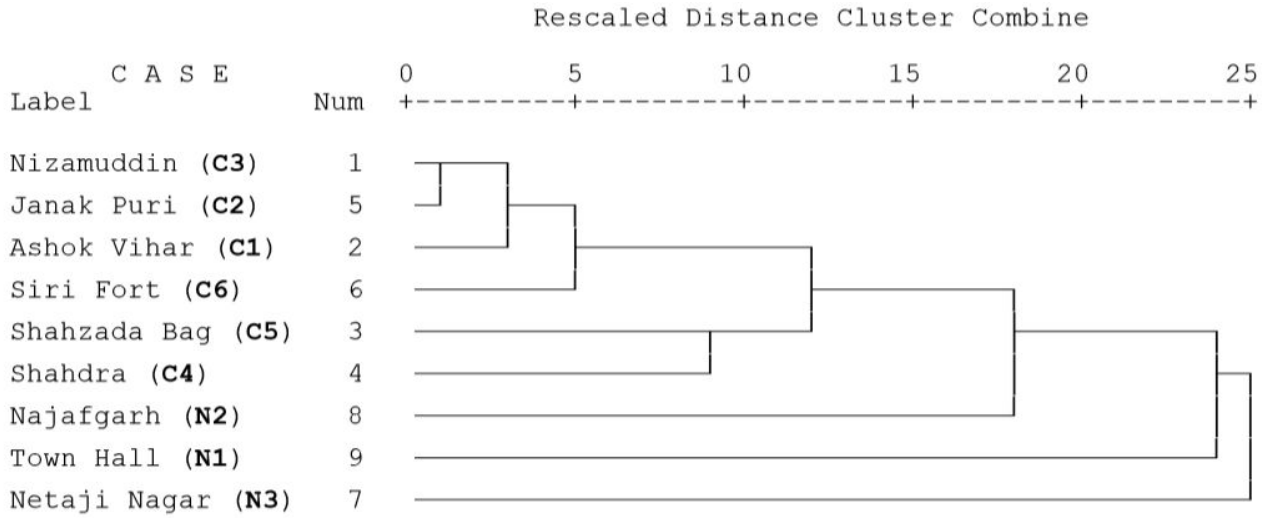


c) 2014

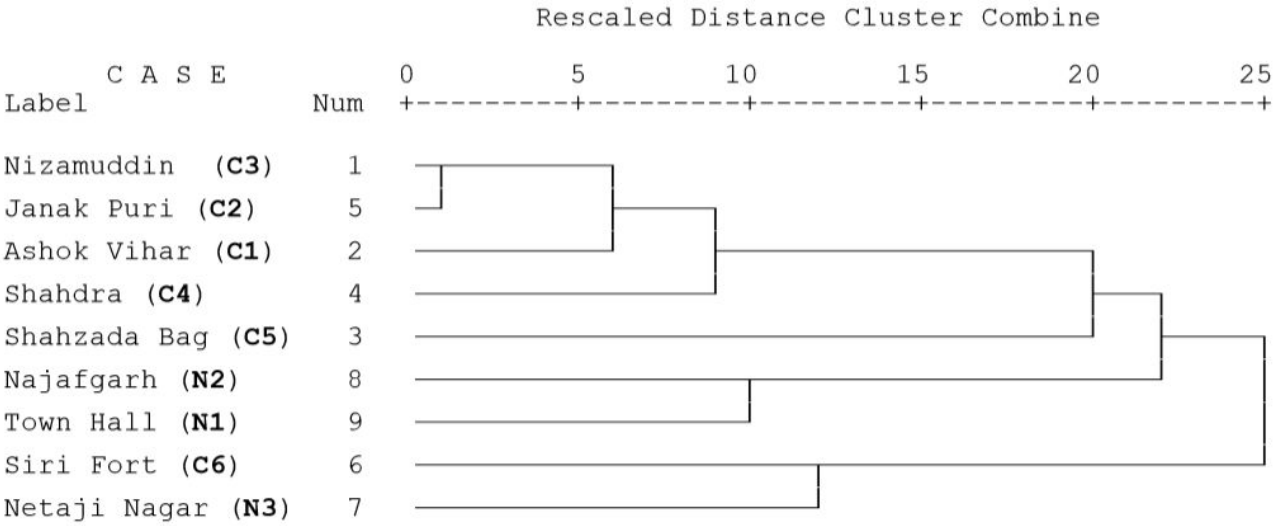


Dendrogram for SO_2 data

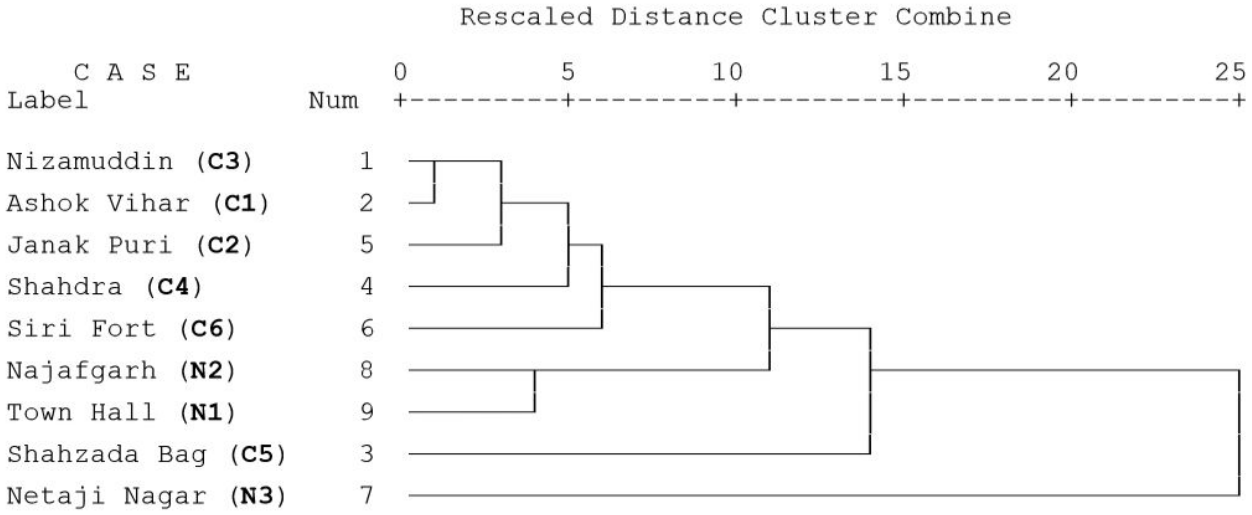
a) 2012



b) 2013

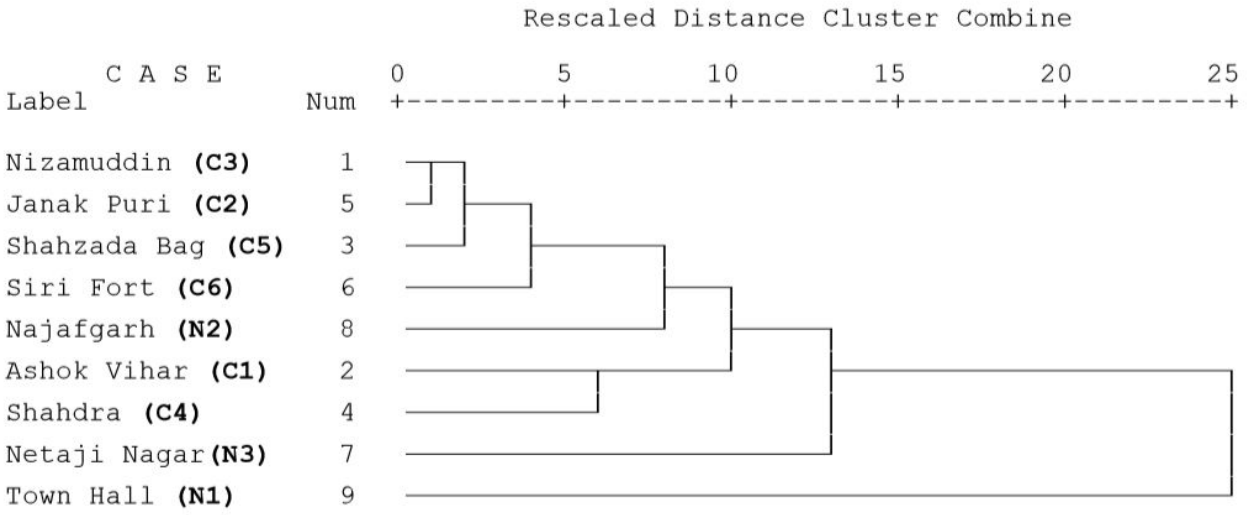

Rescaled Distance Cluster Combine
C A S E
Label Num
0 5 10 15 20 25
Nizamuddin (C3) 1
Janak Puri (C2) 5
Ashok Vihar (C1) 2
Shahdra (C4) 4
Shahzada Bag (C5) 3
Najafgarh (N2) 8
Town Hall (N1) 9
Siri Fort (C6) 6
Netaji Nagar (N3) 7

c) 2014

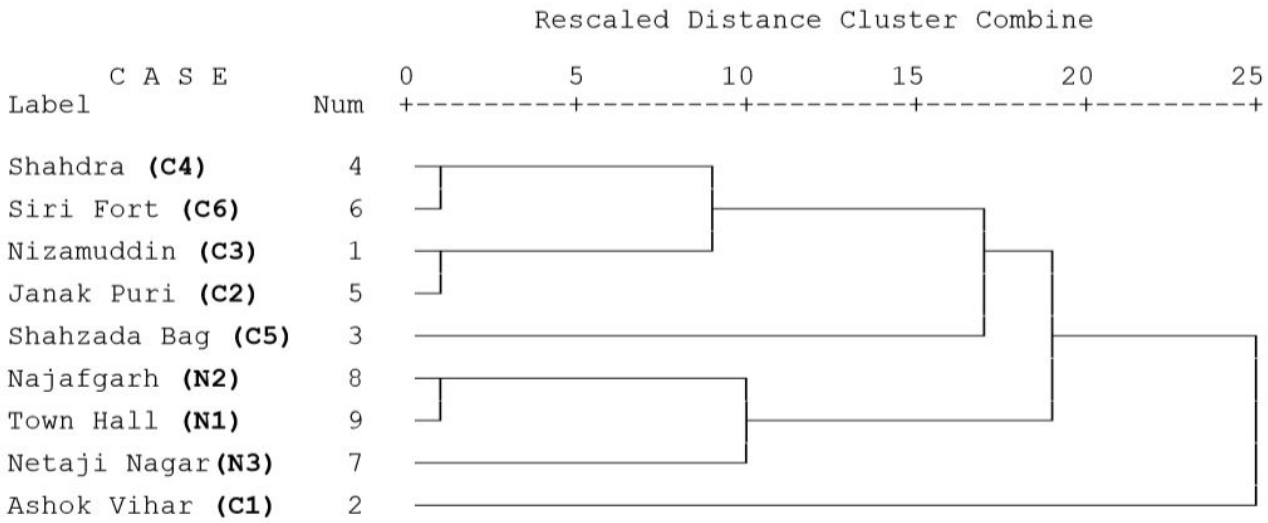

Rescaled Distance Cluster Combine
C A S E
Label Num
0 5 10 15 20 25
Nizamuddin (C3) 1
Ashok Vihar (C1) 2
Janak Puri (C2) 5
Shahdra (C4) 4
Siri Fort (C6) 6
Najafgarh (N2) 8
Town Hall (N1) 9
Shahzada Bag (C5) 3
Netaji Nagar (N3) 7

Dendrogram for NO_2 data

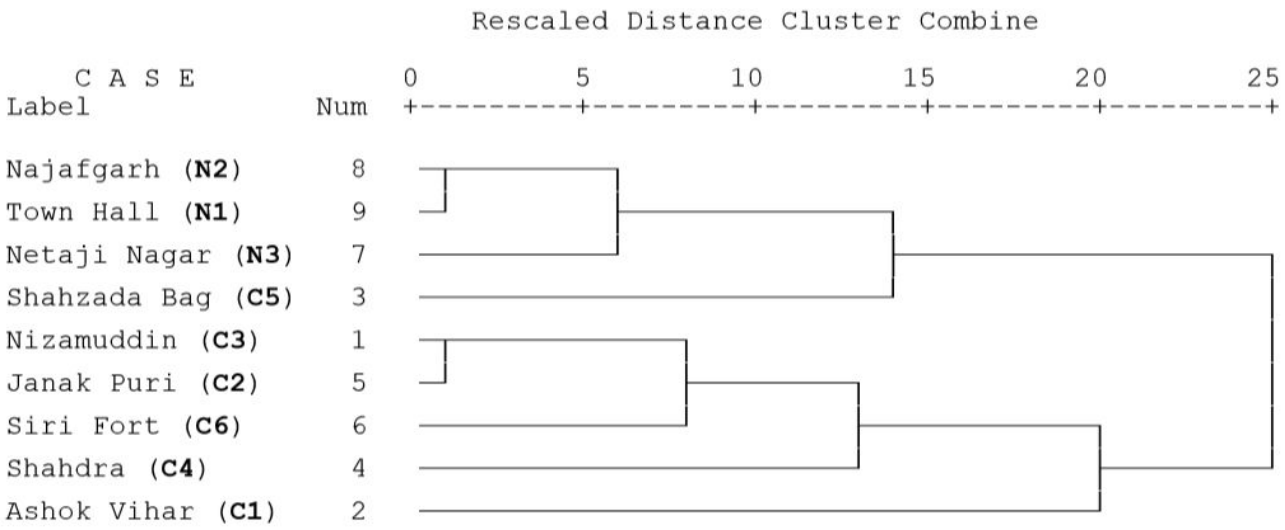
a) 2012



b) 2013



c) 2014



Discussion

Results of cluster analysis suggest that in Delhi there is generally no difference in pollutant levels across land-use types. We speculate that this may be due to the fact that industrial activity has gradually moved out of the city owing to government directives and the build-up of commercial areas near and within residential areas (thereby increasing vehicular emissions in residential areas). This fact would be useful, if and when the authorities consider modifying the network or expanding it in Delhi. It is recommended that a similar analysis is done using PM10 data when it becomes available. Owing to possibly large background levels of SPM spatial patterns may not get easily revealed. The results also provide some support for the recommendation that for Indian cities like Delhi there must be a single ambient air quality standard, rather than standards classified based on land-use. In the future researchers may also wish to do a cluster analysis based on daily data (not yet available in the public domain in India) and to do this seasonally. It is very likely that spatial patterns of peak levels are quite different from the spatial patterns of grossly averaged levels.

The purpose of this study was to study spatial patterns of air pollution in Delhi over a ten year period using three criteria pollutants. Cluster analysis indicated that until 2014, by and large, two distinct classes existed. This study will provide a methodology for Indian researchers, practitioners, and regulatory authorities to do

an exploratory study of spatial patterns and/or data quality issues of air pollution in Indian cities using the National Ambient Air Quality Monitoring System data.

References

Central Pollution Control Board (CPCB).Guidelines for National ambient air quality monitoring, Series:NAAQM/25/2003- 04, Parivesh Bhavan, Delhi, 2009.

ISI, Methods for measurement of air quality: Part 14 Guidelines for Planning the sampling of atmosphere and location of monitoring stations. IS 5182 (Part 14), Indian Standards Institute, New Delhi.

The central pollution control board (CPCB). National air quality index, Series CUPS/82/2014-15.

Analitis A., K. Katsouyanni, E. Dimakopoulou, A. K. Samoli, Y. Nikolouopoulos, G. Petasakis, J. Touloumi, H. Schwartz, H. R. Anderson, K. Cambra, F. Forastiere, D. Zmirou, J. M. Vonk, L.clancy, B. Kriz, J. Bobvos and J. Pekkanen. Short-term effects of ambient particles on cardiovascular and respiratory mortality. Epidemiol 17; 2006: 230-233.

Kaushik. C. P., Ravindra K., Yadav K., Mehta S., and Haritash A. K. Assessment of ambient air quality in urban centers of Haryana(India) in relation to different anthropogenic activities and health risks. Environment Monitoring and Assessment. 122; 2006: 27-40.

Pope C. A. III and D. W.Dockery Health effects of fine particulate air pollution; lines that connect. Journal of Air and Waste Management Association. 56; 2006: 709-742.