Particle pollution distribution in the Sydney airshed

By

Hiep Duc[*], Merched Azzi[**], Vo Anh[***] and Quang Tien[***]

* Environment Protection Authority NSW, PO Box 29, Lidcombe NSW 2141
** CSIRO Division of Coal and Energy Technology, PO Box 136, North Ryde, NSW 2113
*** Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001

Abstract The spatial distribution of particle pollution is strongly influenced by wind patterns. In this paper, an attempt will be made to detect the pattern of particle pollution, as measured at a number of monitoring stations in the Sydney region, for summer and winter periods. The wind field pattern at different hours of the day for different seasons will have influences on the dynamics of the particle distribution in the Sydney basin. This will explain the temporal short-term relationship between the particle distribution and meteorological conditions. However, by taking into account both the anthropogenic long-term trend and the intermittency in the time series for the classification of stations, the particle distribution will have a different pattern compared to the short-term one. This distribution pattern is not dependent on meteorological conditions.

Fine particles levels, as measured by nephelometry, at different stations, are classified using the correlation method to determine the short-term particle distribution pattern. The Hellinger-Kakutani metric of spectral distance between the spectra of air pollution time series at the stations is used to determine the long-term particle clustering pattern. The spectrum of fine particle time series will be determined using a recently developed model. This model consists of a Long Range Dependence (LRD) component that extracts the trend due to anthropogenic activities without the influence of meteorological parameters and the intermittency component which accounts for the short-term, high frequency fluctuation in the time series.

The results of the two classifications based on the short-term correlation and the spectral Hellinger-Kakutani metric incorporating both long-term particle trends and intermittency characteristics at different stations will give a complete picture of the spatial characteristics of the particle distribution in the Sydney basin.

1. Introduction

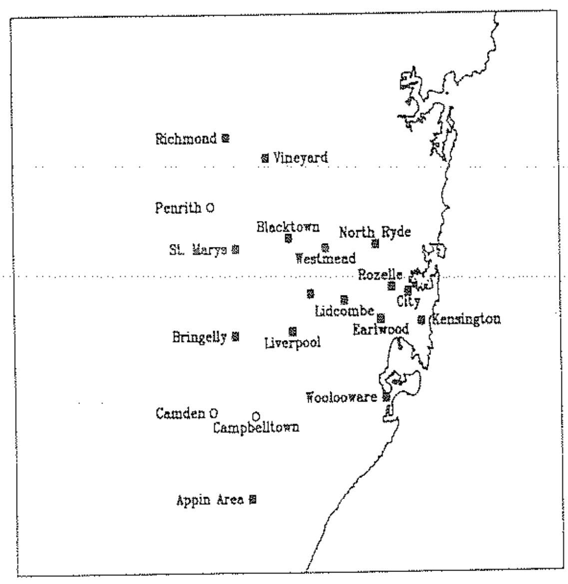
The Sydney basin currently has 18 monitoring stations scattered throughout the region. Air pollutants and meteorological parameters are measured continuously on a 2-minute basis and consolidated into hourly values that are used as the basis for all statistical analyses. The current monitoring network in the Sydney basin is depicted in Figure 1. For most of 1996, 9 stations were measuring fine particles by nephelometry.

The two major air pollution problems in the Sydney airshed are photochemical smog and fine particles. Photochemical smog (measured as ozone) is produced primarily on summer days from complex reactions of oxides of nitrogen, and reactive organic compounds (ROC) under strong sunlight. Fine particles, as measured by nephelometer for visibility range, are produced in both winter and summer periods by the emission of particles from motor vehicles, wood burning, bush fires etc. High levels of fine particles, however, occur more frequently in winter than in summer.

Photochemical smog has been studied extensively by the Metropolitan Air Quality Study (MAQS) (1). The classification of stations based on ozone characteristics also has been recently studied (2) to determine the spatial photochemical smog characteristics across the Sydney region.

In this paper, we examine some aspects of fine particle

Figure 1 Air quality monitoring stations in Sydney



spatial distribution in the Sydney basin by classifying stations according to correlation and spectral density.

Because of the different nature and characteristics of photochemical smog and fine particle pollution, the spatial distribution patterns for each will be different.

It is understood that the particle characteristics at a location are contained in the particle concentration time series at that site. These characteristics can be modelled in the spectral representation of the time series. The similarity or difference of the characteristics between the sites can be determined by their respective spectral representations. A model representing the spectrum as well as a measure of the similarity of the two specified spectral representations will be described.

The study of particle spatial distribution can provide an understanding of the dynamics of particles movement (due to meteorology) and information about sub-regions, which have different pollution characteristics. In addition, it can be used to assist in the assessment of whether the current number of monitoring stations is adequate for collecting information representative of pollution concentrations for a particular region. To adequately understand these issues, the study of spatial distribution of particles over different periods under different conditions is also required.

2. Cluster analysis

Cluster analysis is used to identify groups of objects which have similar characteristics to a degree greater than chance of occurrences. Most clustering methods are based on a similarity or dissimilarity measure.

Two commonly used measures of similarity in airshed classification are sample correlation computed from the time series at a pair of sites and Euclidean metric between two sites computed from their air pollution goal exceedences. A simple linkage algorithm by Anderberg (6) can be used for the correlation measure, but the average-linkage hierarchical measure is more appropriate for an Euclidean metric based on the goal exceedence patterns.

3 Classification of stations using spectral characteristics of time series

3.1 Methodology

The model of the particle time-series is based on the long-range dependence (LRD) and second order intermittency method as described by Anh et al (3). The method describes a stochastic process with LRD, characterised by a spectral density having a singularity of some fractional order at frequency 0, and some intermittency. The LRD component in the air pollution time series provides an estimate of the long-term trend (5), while the intermittency component characterises the high frequency fluctuation in the series.

The spectral density of the particle time series can be modelled as

$$f(\omega)=\frac{c}{\omega^{2\gamma}}\frac{1}{(1+\omega^2)^\beta},$$

$$c>0, 0<\gamma<1/2, 0<\beta\leq 1, \omega\in R \quad (1)$$

and, in discrete form, is approximated as

$$f_D(\omega)=\frac{\sigma^2}{2\pi}\frac{1}{\left|1-e^{i\omega}\right|^{2(\gamma+\beta)}}\frac{1}{\left|1-\theta_1 e^{i\omega}-..-\theta_p e^{jp\omega}\right|^2}$$

$$\sigma^2>0, \omega\in(0,\pi] \quad (2)$$

where the second component on the right hand side of equation (2) is the spectral density of a stationary AR(p) process.

The above spectrum corresponds to the model

$$(1-\theta_1 B-..-\theta_p B^p)(1-B)^{\gamma+\beta}X_t=\varepsilon_t$$

where B is the backshift operator

$$BX_t=X_{t-1}$$

and ε_t is white noise with variance σ^2.

γ is the LRD parameter while β represents the infinite variance or intermittency parameter.

Estimation of $\gamma+\beta$ can be determined by applying one of the methods for finding the LRD component (Anh et al (7) or Haslett and Raftery (8)).

The exponent γ can be found by first applying a wavelet transform, using the Haar wavelet function, to the time series then one of the above mentioned methods for finding the LRD component can be applied on the transform series to estimate γ (3).

3.2 Hellinger-Kakutani metric

A commonly used metric to measure the distance between two spectral densities, f and g, is the Euclidean metric:

$$d_1(f,g)=\left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(f(\lambda,\mu)-g(\lambda,\mu))^2 d\lambda d\mu\right)^{1/2} \quad (3)$$

In this paper, we consider the Hellinger-Kakutani metric which provides an optimum distance that can be readily computed from these spectral representations (4).

Another metric that has been used is the Kullack-Leibler information distance. This metric has been used successfully before in photochemical smog classification (2).

The Hellinger-Kakutani metric is defined as:

$$d_2(f,g) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \ln\left(\frac{f(\lambda,\mu)+g(\lambda,\mu)}{2\sqrt{f(\lambda,\mu)g(\lambda,\mu)}}\right)d\lambda d\mu \quad (4)$$

This metric is invariant with respect to linear filtering while d_1 is not. It is therefore advantageous to use d_2 (5).

3.3 Application to particle time series

There are 9 stations that have a reasonable long time series of maximum daily data for the period 1993 to 1996 (Kensington was relocated to Randwick in March 1995). These stations are:

1	Blacktown	2	Bringelly
3	Kensington	4	Liverpool
5	Richmond	6	Rozelle
7	St Mary	8	Westmead
9	Woolooware		

Missing daily data in the series were replaced by interpolated values. However, if more than 3 daily values are missing, the averages of the values for the same day in other years were used to replace the missing data.

The spectral densities of the time series are calculated using both the LRD and second-order intermittency parameters, ie. γ and β. The Hellinger-Kakutani metrics between stations are then found by using equation (4) as described above.

3.4 Classification results

The calculated Hellinger-Kakutani metrics are used to classify the stations based on spectral similarity for daily peak of fine particle concentrations. The results show that there is a strong similarity of long-term spectral characteristics between the 3 stations: Richmond, Blacktown and St Mary. These stations are located in the West and North West of Sydney in the Nepean-Hawkesbury catchment area.

Surprisingly, there are also some similarities, in terms of spectral characteristics, between Bringelly and Woolooware, and between Liverpool and Rozelle despite their physical separation (see map in Figure 1).

Figure 2 shows the dendrogram and the groups of classified stations.

4. Classification of stations using particle correlation patterns.

A conventional method for determining the similarity of particle characteristics at different stations is to find the correlation coefficients between the specified-hourly time series. Because the correlation is also influenced by meteorology, time series representing different seasons should be used.

In this way, we can determine the dynamics of the particle pattern across the Sydney basin changing over time for each season. The following 12 sites are used in the correlation and clustering analysis.

1	Kensington	2	Rozelle
3	Liverpool	4	Blacktown
5	Bringelly	6	Woolooware
7	Richmond	8	St. Marys
9	Westmead	10	Peakhurst
11	Smithfield	12	Douglas Park

4.1 Particle (Nephelometer), Winter 1993

There are only 9 sites (first 9 of the above 12 sites) that have daily 1-hour data for the winter period of 1993 (June-August). The correlation coefficients for 9am and 3pm of daily data between the sites are calculated and then classified using the usual Euclidean metrics.

The results show that at 9am and 3pm the particle distribution pattern are not very different. At 9am, the stations at St Mary, Liverpool and Bringelly are strongly related. The same can be said for the stations of Kensington, Rozelle and Westmead. At 3pm, St. Mary, Liverpool and Bringelly are still in the same group while the second group now includes Blacktown but leaves out Westmead.

The result of the classification is shown in Figure 3.

4.2 Particle (Nephelometer), Summer 1994

For the period of summer 1994, we have 12 monitoring stations that have data to be used for correlation analysis.

The result, in Figure 4, shows that the particle patterns at 9am and 3pm are very different. At 9am, there is only one group of stations showing strong correlation with each other in terms of daily values. These stations are Blacktown, Smithfield, Liverpool, Bringelly and St. Mary. However, at 3pm, there are 2 or 3 distinct groups of stations. The first cluster consists of St Marys, Blacktown and Smithfield. The second cluster includes Kensington, Woolooware and possibly Peakhurst. The third cluster links Richmond and Bringelly and also has some correlation with the first group.

The different particle patterns at 9am and 3pm shows that the summer sea breeze in the afternoon does change the particle distribution in the Sydney region. There are two sub-regions: one along the coast and one in the West.

5. Wind field description

Because of the strong seasonal effect in Sydney, the wind pattern at a particular hour of the day in the winter period is different from the pattern at the same hour in the summer period.

Typical wind patterns at 9am and 3pm for each of the summer and winter periods are shown in Figure 5. The 1996 data are used instead of 1993 and 1994 data, since more stations having meteorological measurements were added to the network and hence the resolution of the wind pattern is increased. Wind patterns, for the summer and winter periods at 9am and 3pm, using 1993 or 1994 data are similar to the corresponding patterns using 1996 data.

The wind data measured at each station are averaged for the summer period of 1996 (1/12/1995 to 01/03/1996) at 9am and 3pm. Similarly, the 9am and 3pm patterns for the winter period are averaged for the winter period of 1996 (1/6/1996 to 1/9/1996) at each of the above hours.

In the winter period, at 9am, there is a light southerly wind flow from the South-West and westerly flow from the West to the eastern part of Sydney. In the afternoon, at 3pm, the wind is moderate and on average from the South-West.

In the summer period, averaged wind data shows that light wind flows from the South in the early morning across the Sydney basin but became moderate south-easterly in the afternoon. This is consistent with the different particle distribution patterns in the Sydney region at 9am and 3pm.

6. Conclusion

Some information on the particle distribution pattern in the Sydney region can be obtained by first analysing the long-term and intermittency characteristics of the particle concentration time series at each of the stations and then classifying them into pattern clusters. The result shows that the 3 stations in the West and North-West (St Marys, Blacktown and Richmond) have similar long-range trend and high frequency fluctuation or exceedence-like behaviour.

Short-term correlation analysis (used with a clustering method) and the wind field patterns can also reveal the dynamics of the particle distribution. We have shown that, using this technique, during the winter period, there are roughly two groups of particle distribution characteristics: one in the West and South-West and one in the East and along the Parramatta river basin. The distribution pattern is approximately the same at 9am in the morning and at 3pm in the afternoon.

In the summer period, the particle distribution patterns are different in the morning and in the afternoon. In the morning, there is only one group of stations in the West and South-West of Sydney which shares similar particle distribution characteristics. In the afternoon, due to sea breezes, there are roughly two groups of particle distribution characteristics: one along the coast and one inland.

It is important to note that monitoring stations measuring fine particles can be grouped into clusters based on some measures of similar characteristics. But it does not mean simply that each clustering region can be served by only one station as the clustering pattern changes with time and season of the year.

References

(1) Metropolitan Air Quality Study Reports, 1996, NSW EPA

(2) Anh, V., M. Azzi, H. Duc and G. Johnson, Classification of air quality monitoring stations in the Sydney basin, *Proceedings of the Asia-Pacific Conference on Sustainable Energy and Environmental Technology*, 1996, pp. 266-271

(3) Anh, V. H. Duc and Q. Tieng, Modelling persistence and intermittency in air pollution, (*Proc. Air Pollution 97, 16-18 Sept 1997, Bologna, Italy*)

(4) Anh, V., Q. Tieng, T. Bui and G. Chen, The Hellinger-Kakutani metric for pattern recognition, (*Proc. IEEE International Conference on Image Processing, Santa Barbara, 1997*)

(5) Anh V., H. Duc and M. Azzi, Modelling anthropogenic trends in air quality data, *Journal of Air & Waste Management*, Jan 1997, vol 47, pp. 66-71.

(6) Anderberg, M.R, *Cluster Analysis for Applications*, Academic Press, New York, 1973.

(7) Anh, V., K. Lunney, P. Best, G. Johnson, M. Azzi, and H. Duc, A multivariate causality model for Sydney ozone prediction (edited B. Laussade, H. Power, C. Brebbia), *Proceedings of Air Pollution IV, Monitoring, Simulation and Control*, Toulouse, France, 1996, pp. 145-154

(8) Haslett, J. and A. Raftery, Space-time modelling with long-memory dependence: Accessing Ireland's wind power resource, *Applied Statistics*, 1989, 38, pp. 1-50.

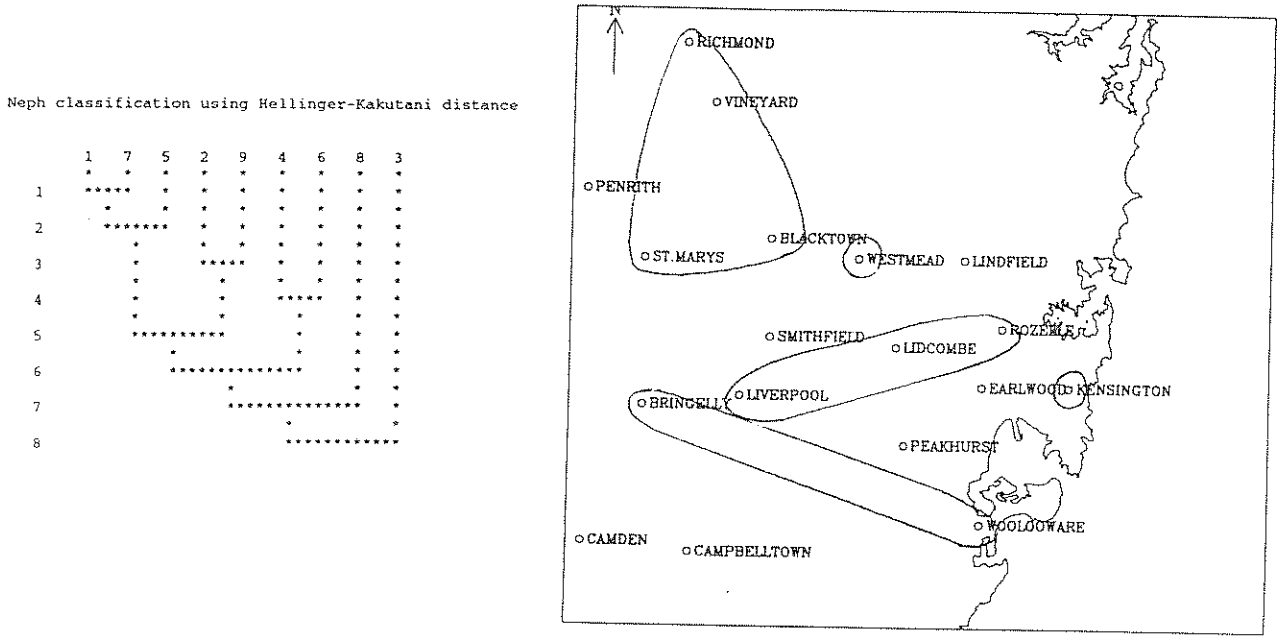


Figure 2 Clustering using Hellinger-Kakutani spectral distance

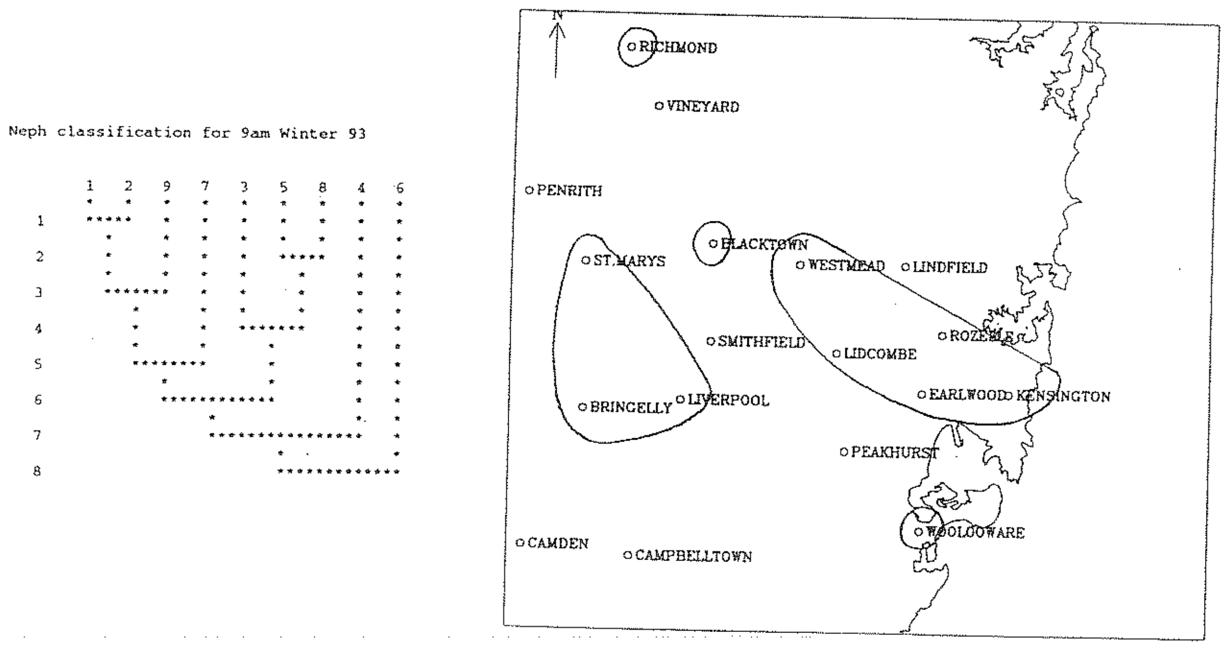
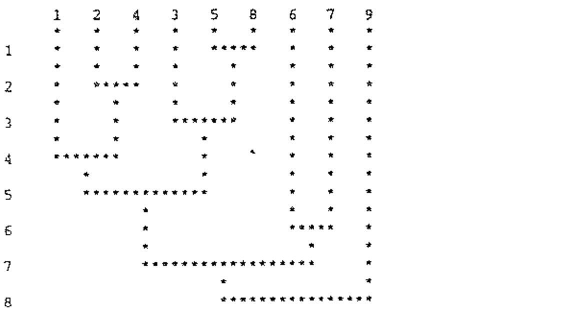


Figure 3 Clustering using particle correlation at 9am and 3pm for winter 1993.

Neph classification (using corelation matrix) for 3pm Winter 93



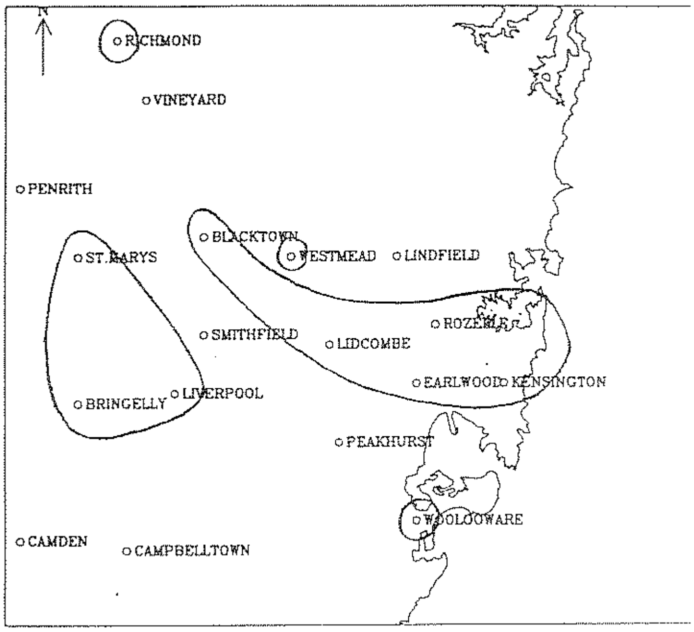
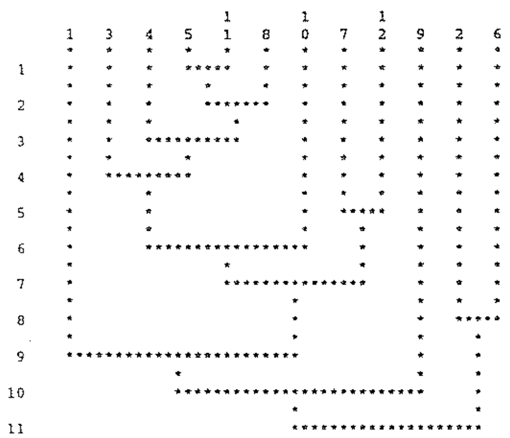


Figure 3 (continued)

Nephelometer summer 94 at 9am cluster



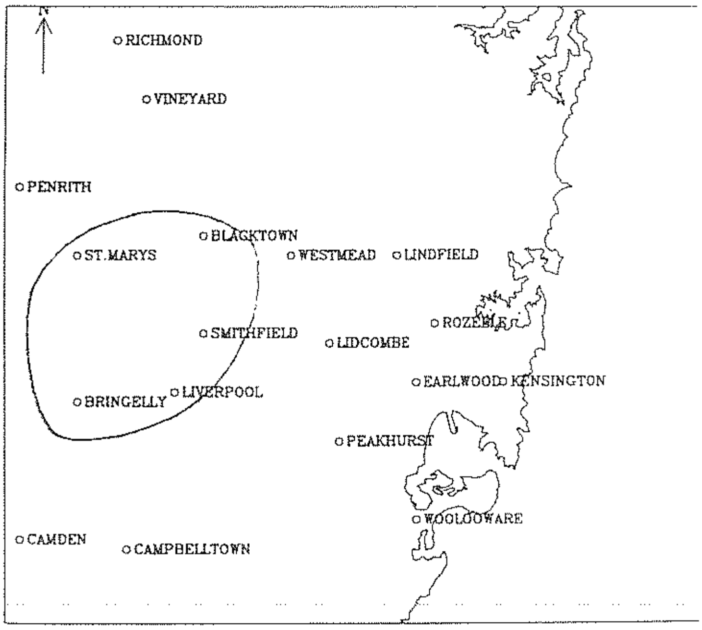


Figure 4 Clustering using particle correlation at 9am and 3pm for summer 1994.

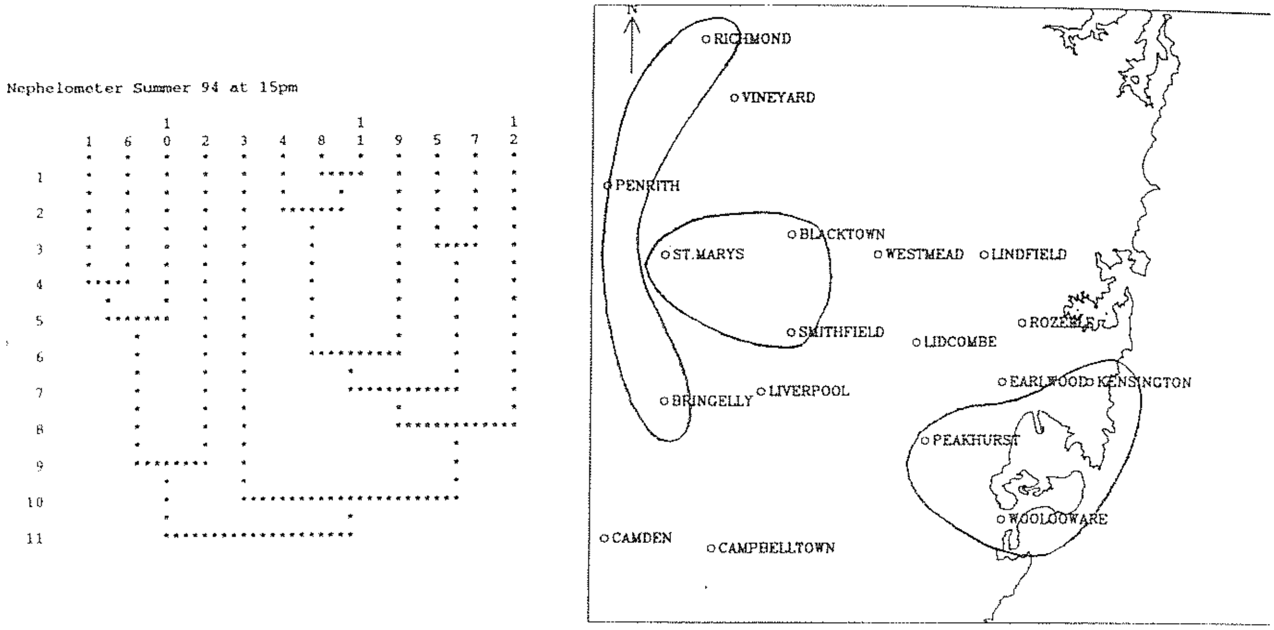


Figure 4 (continued)

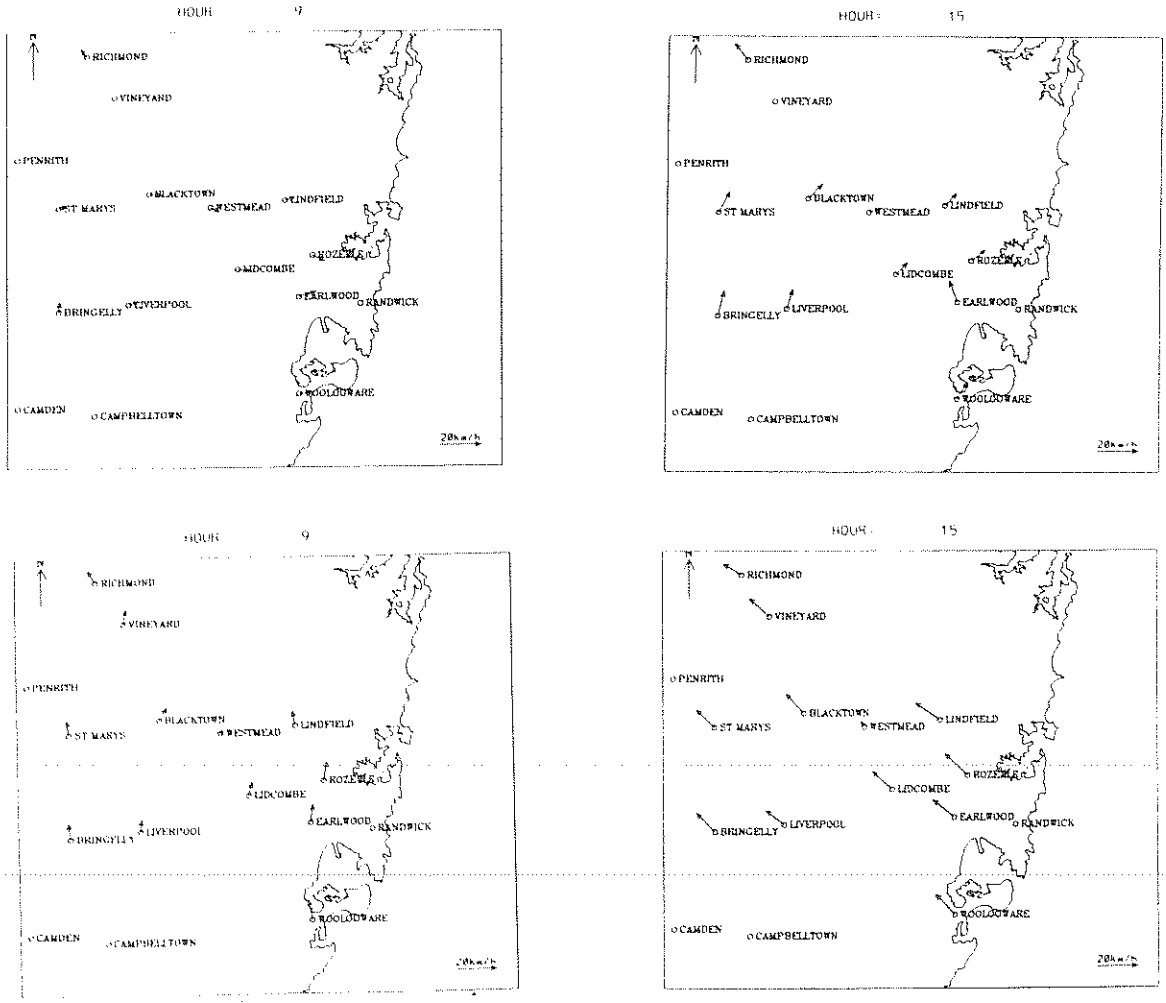


Figure 5 Wind field patterns at 9am and 3pm for winter and summer 1996.