ACTA CLIMATOLOGICA ET CHOROLOGICA
Universitatis Szegediensis, Tomus 42-43, 2009, 107-119

LONG-RANGE TRANSPORT OF PM_{10}, PART 2

L. MAKRA[1], T. SÁNTA[1] and I. MATYASOVSZKY[2]

[1]*Department of Climatology and Landscape Ecology, University of Szeged, P.O.Box 653, 6701 Szeged, Hungary*
E-mail: makra@geo.u-szeged.hu
[2]*Department of Meteorology, Eötvös Loránd University, Pázmány Péter st. 1/A, 1117 Budapest, Hungary*

Summary – The purpose of this study is to identify long-range transport patterns that may have an important influence on PM_{10} levels in four European cities at different latitudes, namely those in Thessaloniki (Greece), Szeged (Hungary) and Helsinki and Oulu (Finland). Calculations were performed on the methodological bases presented in Part 1 and the results are shown in this part. It was found that for Thessaloniki, Szeged and Helsinki, clusters of the trajectory positions originating from North Africa and/or the Sahara are strongly associated with the highest PM_{10} episodes, though with a different frequency. Furthermore, two statistical indices were used to estimate PM_{10} exceedances for the clusters retained for each city.

Key words: PM_{10}, long-range transport, backward trajectory positions, cluster analysis, Mahalanobis metric

4. RESULTS AND DISCUSSION

In the following sections a cluster analysis, the ANOVA results and the PM_{10} levels indices will be presented. Then a goodness of cluster analysis will be performed for the 2D and 3D clustering procedures for each of the four cities studied.

4.1. Comparison of goodness of 2D and 3D cluster analyses

In order to decide whether 2D or 3D clustering is more efficient, a comparison was made between the two procedures (Table 1a-b). Our cluster analysis resulted in a smaller ID with a bigger ED for Thessaloniki, with a smaller number of clusters corresponding to the 2D case (case a). Hence 2D clustering appears more efficient, especially when taking into account the fact that the probability of having significant differences between mean PM_{10} concentrations within these cluster pairs is larger than that among 3D clusters. (See method 1 in Section 3.3, Part 1, when comparing a goodness of cluster analysis in 2D – 3D clustering). The remaining cities belong to case c. Here, the relative differences between IDs corresponding to 2D and 3D as well as relative differences between EDs for the 2D and 3D cases are compared. The relative difference for the ID value is larger than the relative difference for the ED value for Helsinki and Oulu, while the ED-based relative difference is larger than the ID-based relative difference in Szeged. Each case suggests that 2D clustering performs better (Table 1a). The probability values listed (the last two columns of Table 1b), except those for Helsinki and Oulu, support this view.

Table 1 Comparison of goodness of cluster analysis (CA) in 2D – 3D clustering

a

City	2D Cluster no. retained	Internal distance	External distance	3D Cluster no. retained	Internal distance	External distance
Thessaloniki	10	0.2389	0.5673	11	0.3381	0.4794
Szeged	14	0.2774	1.1808	11	0.3271	0.4647
Helsinki	10	0.2329	0.5715	13	0.3443	0.6681
Oulu	10	0.2360	0.5881	13	0.3583	0.6983

b

City	Result of comparison Case	 CA is better in D below	[1]Probability 2D	 3D
Thessaloniki	a	2D	0.8200	0.7107
Szeged	c	2D	0.8776	0.8264
Helsinki	c	2D	0.7600	0.7811
Oulu	c	2D	0.8000	0.8166

When comparing the goodness of the 2D and 3D clustering procedures for a city, it can happen that

a: clustering with fewer number of clusters produces smaller ID with larger ED;

b: clustering with larger number of clusters delivers smaller ID with smaller ED;

c: every possible other cases;

[1] Probability of significant differences of the mean PM_{10} concentrations in the cluster pairs, according to method 1 (See method 1 in Section 3.3, when comparing goodness of cluster analysis in 2D – 3D clustering);

As 2D clustering has been shown to be more efficient for each of the four cities, in the following sections we will mostly discuss the results for the 2D clustering. The findings of our 3D cluster analysis will be mentioned only if they provide some additional information.

4.2. Thessaloniki

For Thessaloniki, 2D clustering produced 10 clusters according to the RMSD analysis (Fig. 2-3). An analysis of the percentage change in RMSD was restricted to those cluster numbers which were considered when selecting the proper cluster number in order to group the backward trajectory positions (Fig. 3). The easiest way to determine the proper number of clusters is to calculate the percentage change in CRMSD. In this case the retained number of clusters for a city is indicated by the highest CRMSD value (Fig. 4).

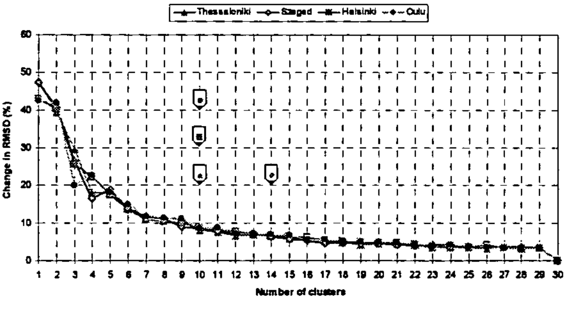

Fig. 2 Change in RMSD against number of 2D clusters. Arrows indicate the number of clusters retained.

Cluster 9 comprises most trajectory positions with around 39.9% of all the trajectory positions (Fig. 5, upper large and lower left panels) and they are often associated with low-moving air masses (Fig. 5, lower right panel). This cluster covers the regions of the Balkan Peninsula and the Near East and comprises trajectory positions of slow movement. Trajectory positions of Cluster 9 correspond to moderate movements, since this cluster covers a small region. Further relatively small regions are delineated by clusters 1, 2, 6 and 7 indicating slow or moderate movements. On the other hand, clusters 3, 4, 5, 8 and 10 cover the largest regions, referring to the fastest movements of trajectory positions (Fig. 5, upper large panel). Clusters 1, 2, 7 and 8 have relatively high ratio of the trajectory positions, while the remaining clusters indicate low frequency (Fig. 5, lower left panel). Highest ratio of air masses streaming on higher elevations is involved in cluster 1 and in all the clusters with low frequency of the trajectory positions. On the other hand, highest ratio of low moving air masses is connected with clusters 2 and 9 (Fig. 5, lower right panel).

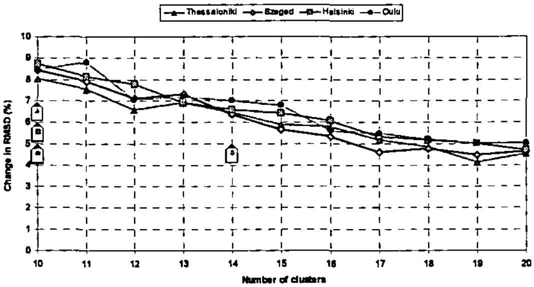

Fig. 3 Change in RMSD against number of 2D clusters *for cluster numbers taking into account when selecting the* proper cluster number. Arrows indicate the number of clusters retained.

Analysis of variance revealed a significant difference in mean PM_{10} concentrations among the clusters. Pairwise comparisons of cluster averages using Tukey test detected 41 significant differences of all the possible 45 cluster pairs (91.11%) (Fig. 6). Based on the differences of the mean PM_{10} concentrations of the clusters, the following groups of clusters were created: (1) clusters 1, 4 and 8 with low PM_{10} averages (mean concentrations: 56.43-56.97 μgm^{-3}); (2) clusters 5, 6, 9 and 10 with medium PM_{10} averages (mean concentrations: 59.65-61.32 μgm^{-3}); and (3) clusters 2, 3 and 7 with high PM_{10} averages (mean concen-trations: 62.04-63.84 μgm^{-3}). According to the results, the highest PM_{10} levels can be clearly linked to clusters 2 (North-Africa, with the main part of the Sahara) and 7 (western basin of the Mediterranean) (Fig. 5, upper large panel), which occur frequently (Fig. 5, lower left panel) with high ratios of low-moving trajectory positions (Fig. 5, lower right panel). This may provide evidence of Saharan dust intrusion over Greece, which is in accordance with previous papers (Balis et al. 2004, Koukouli et al. 2006, Kaskaoutis et al. 2008). Cluster 3 (region of the North Atlantic and North America) can also be linked to the highest PM_{10} levels. However, the frequency of this cluster is very low and thus provides only a limited contribution to overall air-quality degradation. All three furthermost source regions (clusters 3, 10, 4) have either low concentrations or frequency (see Fig. 5), which corroborates the finding of Kallos et al. (2007) and suggests that intercontinental transport of particulate matter to the greater Mediterranean region is of limited importance. As regards the EU regulations, daily mean PM_{10} levels exceeded the 24-h PM_{10} limit value (50 μgm^{-3}) on 3959 days (54.23%) in the period 2001-2005 (Fig. 13). It was found that the likelihood of the 24-h

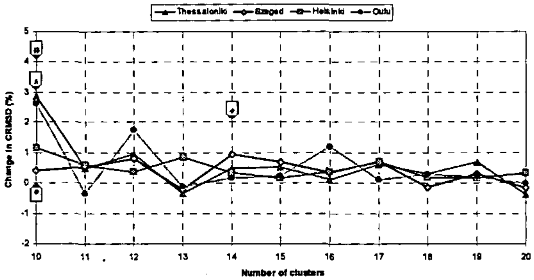

Fig. 4 Change in CRMSD against number of 2D clusters for cluster numbers taking into account when selecting the proper cluster number. [Change in $CRMSD(cluster_{i+1})$ = Change in $RMSD(cluster_i)$ – Change in $RMSD(cluster_{i+1})$] Arrows indicate the number of clusters retained.

PM_{10} exceedances (INDEX1) was uniformly high in each cluster. On the other hand, the likelihood of a given trajectory position being present on a PM_{10} exceedance day (INDEX2) demonstrates larger differences between the clusters (Fig. 13).

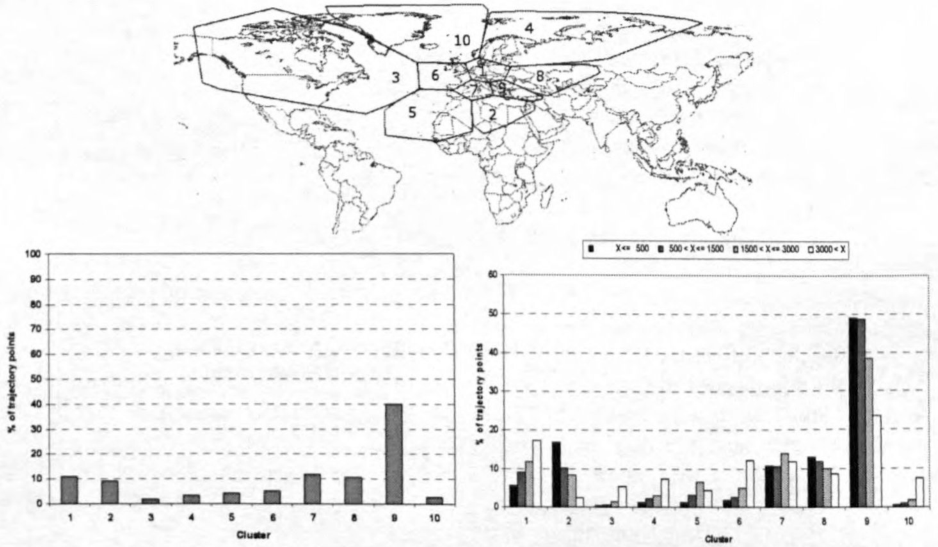

Fig. 5 2D clusters of backward trajectory positions (upper large panel), percentages of the 2D backward trajectory positions (lower left panel) and percentages of their height (lower right panel) in the clusters retained, Thessaloniki, 2001-2005

INDEX1 and INDEX2 as proba-bilities are relative and absolute categories of PM_{10} exceedances, respectively. If the values of INDEX1 are high for each cluster, this means that those of INDEX2 are small, as is the case for Thessaloniki. The values of INDEX2, besides their original interpreta-tion, refer to the frequency of the trajectory positions in the given cluster, as well. Namely, in the case of uniformly high values of INDEX1 for each cluster, higher values of INDEX2 mean a higher frequency of the positions of trajectories in the cluster and vice versa. In this way, clusters 7 (the Western Mediterranean) and 9 (the Balkan Peninsula and Turkey), with high INDEX1 and the highest INDEX2 values (Fig. 13), and with the highest frequency of the trajectory positions advancing mostly on low elevations (Fig. 5, lower left and right panels) are confirmed as important PM_{10} exceedance episodes related to trajectory positions arriving from North Africa and the Near East, respectively (Fig. 13).

	1	2	3	4	5	6	7	8	9
2	**X**								
3	**X**								
4		**X**	**X**						
5	**X**	**X**	X	**X**					
6	**X**	**X**		**X**	X				
7	**X**	**X**		**X**	**X**				
8		**X**	**X**		**X**	**X**	**X**		
9	**X**	**X**	**X**	**X**		**X**	**X**		
10	**X**	**X**	**X**	**X**		X	**X**		

Fig. 6 Significant differences in the means of PM_{10} concentrations between the clusters retained, according to Tukey's honestly significant difference test for Thessaloniki (normal-face: 95% of significance, bold-face: 99% of significance)

4.3. Szeged

For Szeged, 14 clusters were retained in 2D, based on the percentage change in RMSD (Figs. 2-3) and in CRMSD (Fig. 4).

Clusters 1, 3, 6 and 7 are the most frequent and each indicates the slow movement of trajectory positions (Fig. 7, upper large lower left panels). They cover the mainland of Europe excluding Russia, Scandinavia and the Iberian Peninsula. Trajectory positions belonging to clusters 4, 5, 8, 9 and 12, surrounding the above clusters, are relatively compact and indicate slow movement, while those of 2, 10, 11, 13 and 14 cover fast-moving trajectory positions (Fig. 7, upper large panel). Cluster 3 comprises most of the trajectory positions; 31.7% of the total. Clusters 1, 6 and 7 indicate a medium frequency, while the rest of the clusters consist of very few trajectory positions; their relative frequency does not exceed 6% in any case (Fig. 7, lower left panel). As regards the most frequent clusters, trajectory positions in cluster 1 can be associated with mostly high-moving air masses, while in clusters 3 and 6 trajectory positions are found at lower elevations (Fig. 7, lower right panel).

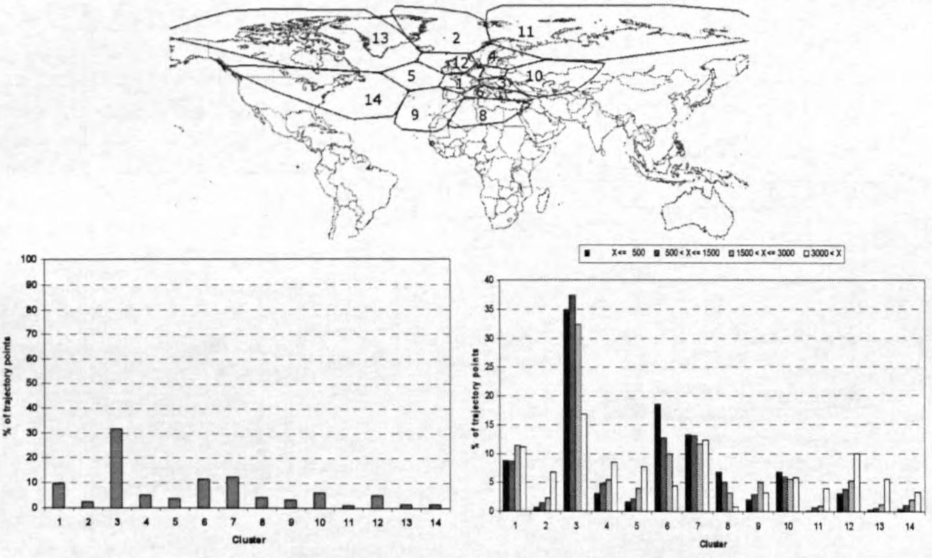

Fig. 7 2D clusters of backward trajectory positions (upper large panel), percentages of the 2D backward trajectory positions (lower left panel) and percentages of their height (lower right panel) in the clusters retained, Szeged, 2001-2005

According to the results of ANOVA, pairwise comparisons of the clusters found a significant difference between the mean PM$_{10}$ levels. When comparing pairwise cluster averages, the Tukey test revealed 86 significant differences among the 91 possible cluster pairs (94.51%) (Fig. 8). For the mean PM$_{10}$ concentrations, the following groups of clusters were determined: (1) clusters 4, 7, 10 and 12 with low PM$_{10}$ averages (mean concentrations: 43.52-45.98 μgm^{-3}); (2) clusters 1, 2, 3, 5, 6, 9 and 11 with medium PM$_{10}$ averages (mean concentrations: 48.16-50.99 μgm^{-3}); and (3) clusters 8, 13 and 14 with high

PM_{10} averages (mean concentrations: 52.25-55.96 μgm^{-3}). As a result, local PM_{10} levels tend to be diluted when air masses arrive at the Carpathian Basin from North and North-western Europe (clusters 4, 7 and 12) as well as from the region of Ukraine, the southern part of the European Russia and Kazakhstan (cluster 10). On the other hand, the highest PM_{10} levels can be related to air masses arriving at Szeged from the regions of North Africa and the Sahara (cluster 8) (Fig. 7, upper large panel). Despite the frequency of this cluster being small and covering mostly low-moving air masses, it does certify the casual appearance of North African dust over Hungary, which was confirmed by earlier studies (Fig. 7, lower left and right panels) (Borbely-Kiss et al. 1999, Koltay et al. 2006). Clusters 13 and 14 (regions of the North Atlantic and North America, northern and southern part, respectively) include very few trajectory positions with mostly high-moving air masses. Therefore, though they are both related to PM_{10} exceedances, their role is negligible (Fig. 7, lower left and right panels).

The monitoring station in Szeged recorded lower PM_{10} levels than those in Thessaloniki. The mean PM_{10} concentrations were above the 50 μgm^{-3} limit value on 51.12% of the days considering the five-year period 2001-2005 (Fig. 13). If the mean annual value of PM_{10} is near the 24-h limit value (50 μgm^{-3}) (more polluted site) and its annual variation is not large, which is the case for Szeged (Makra et al. 2009), then a very small pollution load arriving by the trajectory positions is enough to exceed the limit value. Accordingly, values of INDEX1 are high for each cluster, while those involving INDEX2 are smaller, as is the case for Szeged and also for Thessaloniki. It was found that INDEX1 values are uniformly high in each cluster; however, they are highest in clusters 8 (the Sahara), 11 (Russian Arctic region) and 14 (the Middle Atlantic – North American region), which assumes a continental transport influence though with a small frequency especially for the latter two clusters (Figs. 7 and 13). On the other hand, it should be stressed that with uniformly high INDEX1 values and with the highest INDEX2 values clusters 1 (France), 3 (Austria, Croatia, Hungary, Serbia and Romania), 6 (the Northern Mediterranean) and 7 (Northern Germany, Czech and Poland) play, in the long-range transport of particulates, the most important role in PM_{10} exceedance episodes. The relevance of these clusters is confirmed by their high frequency values and mostly low-moving trajectory positions (Figs. 7 and 13).

	1	2	3	4	5	6	7	8	9	10	11	12	13
2	**X**												
3		**X**											
4	**X**	**X**	**X**										
5	**X**		**X**	**X**									
6	**X**		**X**	**X**									
7	**X**	**X**	**X**		**X**	**X**							
8	**X**	**X**	**X**	**X**	**X**	**X**	**X**						
9	**X**	X	**X**	**X**		**X**	**X**	X					
10	**X**	**X**	**X**	X	**X**	**X**	**X**	**X**	**X**				
11				**X**			**X**	**X**	X	**X**			
12	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		
13	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**	
14	**X**	**X**	**X**	**X**	**X**	**X**	**X**	X	**X**	**X**	**X**	**X**	

Fig. 8 Significant differences in the means of PM_{10} concentrations between the clusters retained, according to Tukey's honestly significant difference test for Szeged (normal-face: 95% of significance, bold-face: 99% of significance)

4.4. Helsinki

10 clusters were identified for Helsinki, according to the percentage change in RMSD (Figs. 2-3) and in CRMSD (Fig. 4).

Cluster 10 comprises the most trajectory positions (35.69%) but those of 1, 3, 4, and 9 can also be considered relatively frequent (Fig. 9, upper large and lower left panels). On the other hand, clusters 6 and 7 comprise extremely few trajectory positions. Clusters 3, 4, 9 and 10 cover comparatively small regions; hence, trajectory positions belonging to them are more compact and, in this way, indicate their slow movement. On the other hand, the remaining clusters consist of fast moving trajectory positions (Fig. 9, upper large panel). Cluster 10, which is the most frequent, is characterised by low-moving trajectory positions. At the same time, the rarest clusters 6 and 7 generally include trajectory positions streaming on high elevations (Fig. 9, lower left and right panels).

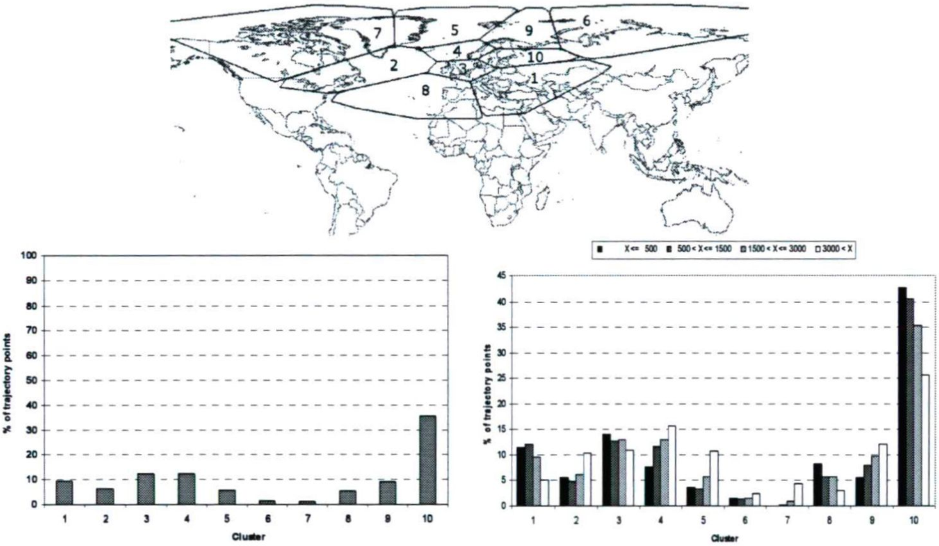

Fig. 9 2D clusters of backward trajectory positions (upper large panel), percentages of the 2D backward trajectory positions (lower left panel) and percentages of their height (lower right panel) in the clusters retained, Helsinki, 2001-2005

ANOVA identified a significant difference among the mean PM$_{10}$ levels of the clusters. Based on the Tukey test, 39 of all the 45 pairwise cluster averages differed significantly from each other (86.67%) (Fig. 4). On the basis of the mean PM$_{10}$ concentrations, the following groups of clusters were formed: (1) clusters 2, 4, 5, 7 and 9 with low PM$_{10}$ averages (mean concentrations: 15.44-15.89 μgm^{-3}); (2) clusters 3, 6, and 8 with medium PM$_{10}$ averages (mean concentrations: 16.44-16.83 μgm^{-3}); and (3) clusters 1 and 10 with high PM$_{10}$ averages (mean concentrations: 17.02-19.56 μgm^{-3}). Accordingly, air masses arriving at Helsinki from the regions of the European Arctic, the North Atlantic and Scandinavia, Greenland and Canada indicate the lowest PM$_{10}$ concentrations. On the other hand, air masses arriving from southern Finland, the mainland of Europe excluding France and the Iberian Peninsula but including the Mediterranean, Northern Africa, the Near East and Kazakhstan and the European part of Russia (clusters 1 and 10) carry the highest PM$_{10}$ levels (Fig. 9, upper large panel). The importance of these trajectory positions in PM$_{10}$ exceedance episodes is that they are the most frequent and are mostly related to low-moving air masses (Fig. 9, lower left and right panels). Evidence for recurrent long-range transport of elevated PM_{10} concentrations from the above regions over southern Finland were reported by Ansmann et al. (2003), Hongisto and Sofiev (2004), Anttila et al. (2008) and Niemi et al. (2009).

PM_{10} levels recorded at the Helsinki Kallio monitoring station are extremely low, compared to those of Thessaloniki and Szeged. This is partly due to its urban background environment, where the local particulate load is lower than that at the traffic sites of Thessaloniki and Szeged, and in part it is a demonstration of the overall lower PM concentrations in Scandinavia (see e.g. EEA, 2007). Furthermore, the 24-h PM_{10} exceedances are also very small. They make up just 1.30% of the days within the five-year period examined (Fig. 10). On the other hand, if the mean daily value of PM_{10} is very small, the values of INDEX1 are also very small for each cluster, hence those of INDEX2 are higher, which is the case for Helsinki (and Oulu, as well). The calculation of INDEX1 yielded very small values for each cluster, which is due to the very low local PM_{10} levels. In this case particulates arriving by long-range transport can only rarely contribute to exceeding the PM_{10} limit value. Since INDEX2 shows the absolute importance of the given cluster in PM_{10} exceedance episodes, clusters 1, 4, 5, 9 and 10 with the highest INDEX2 values are of particular interest. The role of cluster 10 (Baltic Sea region and Northern Russia) is the most specific as a potential source region. It has a very small INDEX1 value (Fig. 10); however, due to its highest number of PM_{10} exceedance episodes, its INDEX2 value is also the highest (Fig. 10). Moreover, it shows the highest frequency of the trajectory positions, mostly with low-moving trajectories (Fig. 9, lower left and right panels).

	1	2	3	4	5	6	7	8	9
2	**X**								
3	**X**	**X**							
4	**X**	**X**	**X**						
5	**X**	**X**	**X**						
6	**X**	**X**		**X**	**X**				
7	**X**		**X**			X			
8	**X**	**X**	**X**	**X**	**X**				
9	**X**	X	**X**			**X**		X	
10	**X**	**X**	X	**X**	**X**		**X**	**X**	

Fig. 10 Significant differences in the means of PM_{10} concentrations between the clusters retained, according to Tukey's honestly significant difference test for Helsinki (normal-face: 95% of significance, bold-face: 99% of significance)

4.5. *Oulu*

On the basis of the percentage change in RMSD (Figs. 2-3) and in CRMSD (Fig. 4) 10 clusters were retained for Oulu.

Clusters 3, 5 and 8, as the nearest ones, include the most trajectory positions and among them cluster 8 is the most frequent (35.55%). Clusters 4 and 7 contain the fewest trajectory positions (Fig. 11, upper large and lower left panels). The position and coverage of the clusters are both very similar to those of Helsinki. Namely, clusters 3, 5, 6 and 8 are characterised as relatively small regions with compact and, hence, slow-moving trajectory positions; while the rest of the clusters cover large regions with fast-moving trajectory positions (Fig. 11, upper large panel). Cluster 8 for Oulu with the most frequent trajectory positions fits the cluster 10 for Helsinki and can be described in a similar way, mostly with low-moving trajectory positions. Parallel to this, clusters 4 and 7 for Oulu match clusters 6 and 7 for Helsinki with similar characteristics. Namely, both comprise extremely few trajectory positions moving mostly on high elevations (Fig. 11, lower left and right panels).

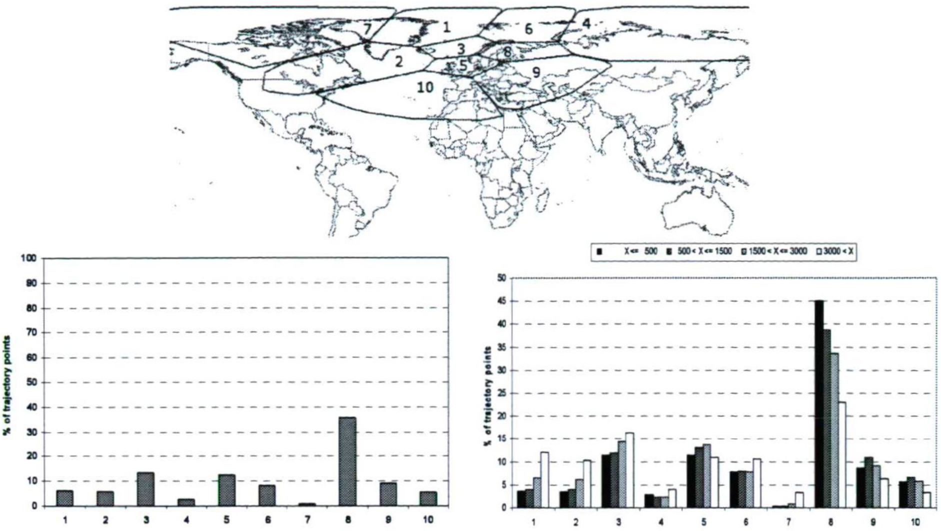

Fig. 11 2D clusters of backward trajectory positions (upper large panel), percentages of the 2D backward trajectory positions (lower left panel) and percentages of their height (lower right panel) in the clusters retained, Oulu, 2001-2005

ANOVA detected significant differences among cluster averaged PM$_{10}$ concentrations. The Tukey test revealed 41 significant pairwise differences of the cluster averages of all the possible 45 cluster pairs (91.11%) (Fig. 12). After considering the mean PM$_{10}$ levels of the individual clusters, the following cluster groups were identified: (1) clusters 2 and 4 with low PM$_{10}$ averages (mean concentrations: 18.23-18.77 μgm^{-3}); (2) clusters 1, 3, 6, 7 and 10 with medium PM$_{10}$ averages (mean concentrations: 19.35-20.27 μgm^{-3}); and (3) clusters 5, 8 and 9 with high PM$_{10}$ averages (mean concentrations: 21.23-22.21 μgm^{-3}).

Consequently, the air masses from the North Atlantic and Canada, as well as from Siberia transport the lowest PM$_{10}$ levels over the target region. The highest PM$_{10}$ levels, on the other hand, are transported from Finland itself, the mainland of Europe excluding France and the Iberian Peninsula but including the Eastern Mediterranean, the Near East, Kazakhstan and the European part of Russia (clusters 5, 8 and 9). Clusters 8 and 9 cover the very same regions that were identified for Helsinki (clusters 1 and 10) as those with high PM$_{10}$ averages, only slightly shifted northwards (Fig. 11, upper large panel). These

	1	2	3	4	5	6	7	8	9
2	**X**								
3	X	**X**							
4	**X**		**X**						
5	**X**	**X**	**X**	**X**					
6	X	**X**		**X**	**X**				
7		**X**			**X**				
8	**X**	**X**	**X**	**X**	**X**	**X**	**X**		
9	**X**	**X**	**X**	**X**	**X**	**X**	**X**		
10		**X**	**X**		**X**	**X**		**X**	**X**

Fig. 12 Significant differences in the means of PM$_{10}$ concentrations between the clusters retained, according to Tukey's honestly significant difference test for Oulu (normal-face: 95% of significance, bold-face: 99% of significance)

frequently developing low-moving air masses are supposed to affect mostly the PM_{10} related air quality of Oulu. (Fig. 11, lower left and right panels). Long-range transport of PM_{10} to northern Finland has been less intensively studied than that of the southern coast. However, countrywide coincidental PM_{10} episodes have been identified and related to PM_{10} long-range transport (Anttila and Salmi 2006). The high similarities of the cluster characteristics of Oulu to the clusters of Helsinki suggests that long range transport of PM to these sites, 600 km apart form each other, is mainly controlled by the same large scale source areas.

Helsinki and Oulu show the lowest mean PM_{10} values, while those for Thessaloniki and Szeged are some three times higher. On the other hand, both the mean PM_{10} values and the 24-h PM_{10} exceedances are higher for Oulu than those for Helsinki. At the Oulu Centre monitoring site the resuspended dust from the nearby busy road increases the PM_{10} levels, while at Helsinki Kallio the emissions from the more distant roads have a minor impact. For Oulu, the 24-h PM_{10} exceedances comprised 5.59% of the days during the period 2001-2005 (Fig. 13). The INDEX1 values, due to the higher frequency of the 24-h PM_{10} exceedances, are higher for Oulu than for Helsinki. Furthermore, the INDEX2 values are also higher for Oulu. Since the values of INDEX1 are uniformly low in each cluster, higher values of INDEX2 in clusters 3, 5 and 8 indicate their higher frequency (Fig. 13). However, cluster 8 (covering the Middle and Northern part of Scandinavia and Northern Russia) seems to be the most relevant cluster because it has the highest frequency value and has mostly low-moving trajectory positions (Fig. 11).

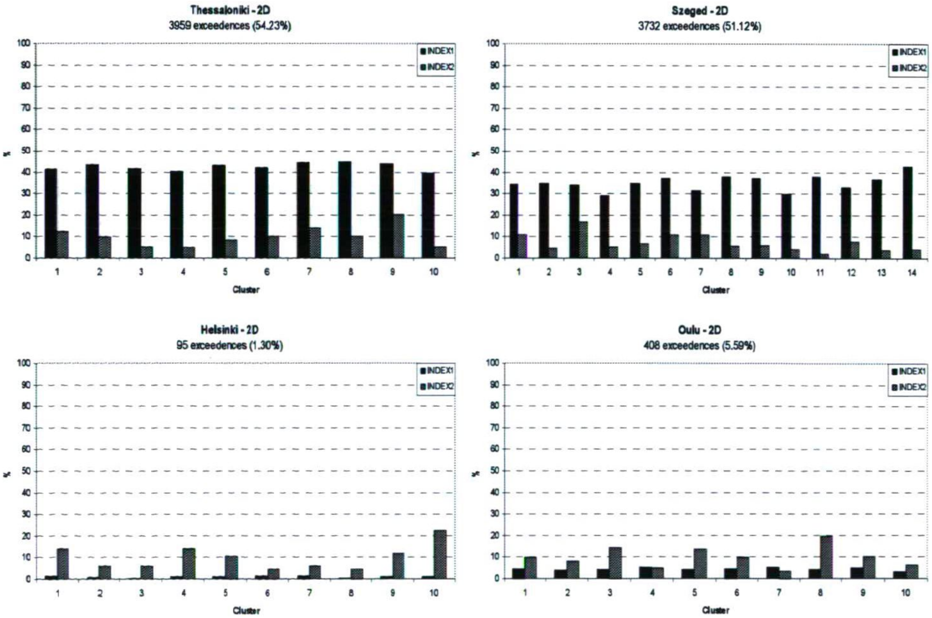

Fig. 13 Indices 1 and 2 for 2D clusters of backward trajectory positions in Thessaloniki, Szeged, Helsinki and Oulu, respectively

5. COMMENTS ON THE METHODOLOGY

A Cluster analysis was carried out on 4-day, 6-hourly positions of backward trajectories. Clustering can be shown with complete trajectories belonging to given clusters (e.g. Borge et al. 2007). We label them as "trajectory clusters" since the trajectories themselves are clustered and shown in this case. However, the individual source regions cannot be separated clearly with this kind of representation, because a great many backward trajectories are clustered and, as a result, the origin of polluted air masses looks like a large spot. This paper shows clusters of 6-h trajectory positions and not the complete trajectories. The 2D delineation of 6-h positions of clusters and preparation of figures for 2D clusters of backward trajectory positions (upper large panels of figures 5-8) were carried out using the function convhull. The algorithm (qhull procedure; www.qhull.org) gathers the extreme trajectory positions (positions farthermost from the centre) belonging to a cluster which are then enclosed. It is important to stress again that clusters retained in this study comprise positions of different trajectories. Accordingly, the clusters themselves are not trajectory clusters as they do not include complete trajectories, but consist of their 4-day, 6-hourly positions. However, these clusters including trajectory positions can also indicate long-range transport patterns that may have a substantial impact on local air quality.

Borge et al. (2007) used a two-stage clustering procedure in order to capture the influence of relatively short, slow-moving trajectories on local air quality. They observed that the original one-stage cluster analysis including all trajectories for a given city was strongly influenced by the trajectory length. Accordingly, long trajectories representing fast-moving air masses were highly disaggregated, even though they often came from the same geographical region, as the Euclidean distance between the initial trajectory points was very large. For the same reason, many short trajectories representing slow-moving air masses were grouped together, although they came from very heterogeneous regions. In order to achieve a more specific representation of slow-moving air masses Borge et al. (2007) grouped together and reanalysed with the same methodology (second-stage) the short trajectory clusters identified with unclear orientation in the initial cluster analysis (first-stage).

However, second-stage cluster analysis (Borge et al. 2007) is not necessary if the metric in the clustering procedure is not Euclidean. The problem justifying the two steps will be avoided if the Mahalanobis metric is used. We can get this metric if principal component analysis is performed on the vectors, then each of the principal component vectors are standardized and cluster analysis is performed with these new vectors using a Euclidean distance. In this way, the Mahalanobis metric takes into account different standard deviations of components of the vectors as well as correlations among the components (Mahalanobis 1936). (Using the Mahalanobis metric we can prevent a number of variables from correlating with each other, and thus less total information will be considered with unduly high weight under the classification compared to another set of variables that is virtually non-correlated and thus contains a relatively large amount of total information.) The problem of two-stage cluster analysis (Borge et al. 2007) comes from different standard deviations of the co-ordinates of the trajectory points being far and near in time (well-separated points typically have high standard deviations, while points that are close together are usually characterized by small standard deviations). In order to demonstrate the role of different standard deviations, let us take a difference of 200 km in the position of a given trajectory point. A difference like 1500 km from us almost counts for nothing, while the same difference is considered very large close to the arriving point of the trajectory.

Furthermore, it should be noted that in the paper of Borge et al. (2007), when using two-stage clustering, the selection of clusters to be included in the second-stage analysis was based on a visual identification of highly heterogeneous first-stage clusters composed of short trajectories. This subjectivity can be removed by applying the Mahalanobis metric.

6. CONCLUSIONS

Cluster analysis was applied to 4-day, 6-hourly positions of backward trajectories arriving in Thessaloniki, Szeged, Helsinki and Oulu over a 5-year period. The Mahalanobis metric was used which makes unnecessary for a two-stage cluster analysis introduced in Borge et al. (2007). A change in the change in RMSD (i.e. CRMSD) was introduced and applied instead of a change in RMSD, as an easier way to select the proper cluster number. A comparison of the efficiency of the 2D and 3D cluster analyses for each city is a novelty of the study. Furthermore, the 2D delineation of the trajectory positions of the given clusters and their presentation was carried out by a novel approach.

The analysis of variance, performed on the basis of the mean PM_{10} values, found a significant difference in the mean PM_{10} levels among clusters retained for each of the cities. The results of the cluster analysis assume that for Thessaloniki, Szeged and Helsinki, clusters of the trajectory positions originating from North Africa and/or the Sahara are strongly associated with the highest PM_{10} episodes, though with a different frequency; while for Oulu, the farthermost source regions are the Mediterranean, the Near East and Kazakhstan. On the basis of the daily PM_{10} exceedances and the frequencies of the trajectory positions in the clusters, two statistical indices were applied. They were used to determine the likelihood of the exceedances in relation to specific air masses. We should mention that after the cluster analysis, the application of ANOVA and the Tukey test on cluster averages allows us to detect source regions of transported PM_{10} significantly contributing to local mean PM_{10} concentrations. Also, the use of INDEX1 and INDEX2 values helps us to determine individual source areas which substantially influence daily PM_{10} exceedances.

Our study demonstrated that the application of the HYSPLIT model together with different statistical techniques on pollutant data might be a useful tool for detecting long-range transport patterns which influence air quality in given European cities. In order to clarify the connection between clusters and local PM_{10} exceedances more precisely, in a future study the values of the meteorological variables at the trajectory positions will also be taken into account.

Acknowledgements – The authors would like to thank Roland Draxler for his useful advice and consultations on the HYSPLIT model, version 4.8; to Gábor Motika (Environmental Protection Inspectorate of Lower-Tisza Region, Szeged), to Helsinki Metropolitan Area Council and to Environment Office of the Oulu region for providing PM_{10} data and to Zoltán Sümeghy for the digital mapping in Fig. 1. This study was supported by the EU-6 Quantify project [No. 003893 (GOCE)].

REFERENCES

Ansmann A, Bosenberg J, Chaikovsky A, Comeron A, Eckhardt S, Eixmann R, Freudenthaler V, Ginoux P, Komguem L, Linne H, Marquez MAL, Matthias V, Mattis I, Mitev V, Muller D, Music S, Nickovic S, Pelon J, Sauvage L, Sobolewsky P, Srivastava MK, Stohl A, Torres O, Vaughan G, Wandinger U, Wiegner M (2003) Long-range transport of Saharan dust to northern Europe: The 11-16 October 2001 outbreak observed with EARLINET. J Geophys Res – Atmos 108 D24: Article No: 4783

Anttila P, Salmi T (2006) Characterizing temporal and spatial patterns of urban PM_{10} using six years of Finnish monitoring data. Boreal Environ Res 11:463-479

Anttila P, Makkonen U, Hellen H, Kyllonen K, Leppanen S, Saari H, Hakola H (2008) Impact of the open biomass fires in spring and summer of 2006 on the chemical composition of background air in south-eastern Finland. Atmos Environ 42:6472-6486

Balis DS, Amiridis V, Nickovic S, Papayannis A, Zerefos C (2004) Optical properties of Saharan dust layers as detected by a Raman lidar at Thessaloniki, Greece. Geophys Res Lett 31: Article No: L13104

Borbely-Kiss I, Koltay E, Szabo GY, Bozo L, Tar K (1999) Composition and sources of urban and rural atmospheric aerosol in eastern Hungary. J Aerosol Sci 30:369-391

Borge R, Lumbreras J, Vardoulakis S, Kassomenos P, Rodriguez E (2007) Analysis of long-range transport influences on urban PM_{10} using two-stage atmospheric trajectory clusters. Atmos Environ 41:4434-4450

EEA (2007) Air pollution in Europe 1990–2004 European Environment Agency, Report No 2/2007

Hongisto M, Sofiev M (2004) Long-range transport of dust to the Baltic Sea region. Int J Environ Pollut 22:72-86

Kallos G, Astitha M, Katsafados P, Spyrou C (2007) Long-range transport of anthropogenically and naturally produced particulate matter in the Mediterranean and North Atlantic: Current state of knowledge. J Appl Meteorol Climatol 46:1230-1251

Kaskaoutis DG, Kambezidis HD, Nastos PT, Kosmopoulos PG (2008) Study on an intense dust storm over Greece. Atmos Environ 42:6884-6896

Koltay E, Borbely-Kiss I, Kertesz Z, Kiss AZ, Szabo G (2006) Assignment of Saharan dust sources to episodes in Hungarian atmosphere by PIXE and TOMS observations. J Radioanal Nucl Ch 267:449-459

Koukouli ME, Balis DS, Amiridis V, Kazadzis S, Bais A, Nickovic S, Torres O (2006) Aerosol variability over Thessaloniki using ground based remote sensing observations and the TOMS aerosol index. Atmos Environ 40:5367-5378

Mahalanobis PC (1936) On the generalized distance in statistics Proc Nat Inst Sci India 2(1):49-55

Makra L, Mayer H, Mika J, Sánta T, Holst J (2009) Variations of traffic related air pollution on different time scales in Szeged, Hungary and Freiburg, Germany. Phys Chem Earth (accepted)

Niemi JV, Saarikoski S, Aurela M, Tervahattu H, Hillamo R, Westphal DL, Aarnio P, Koskentalo T, Makkonen U, Vehkamaki H, Kulmala M (2009) Long-range transport episodes of fine particles in southern Finland during 1999-2007. Atmos Environ 43:1255-1264