Supplementary material: Temporal and spatial analysis of ozone concentrations in Europe based on time scale decomposition and a multi-clustering approach

Eirini Boleti[1,2], Christoph Hueglin[1], Stuart K. Grange[1,4], André S. H. Prévôt[3], and Satoshi Takahama[2]

[1]Empa, Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, Dübendorf, Switzerland
[2]EPFL,École Polytechnique Fédérale de Lausanne, Route Cantonale, 1015 Lausanne, Switzerland
[3]PSI, Paul Scherrer Institute, 5232 Villigen, Switzerland
[4]Wolfson Atmospheric Chemistry Laboratories, University of York, York, YO10 5DD, UK

Correspondence to: Christoph Hueglin (Christoph.Hueglin@empa.ch)

Contents

S1   Choice of number of clusters

The analysis of the appropriate number of clusters is shown in Figs. S1 and S2. The sum of squared distance error within a cluster is ploted against the number of clusters. The number of clusters where the sum of squared distance error is mostly minimized is considered to be the most appropriate choice.

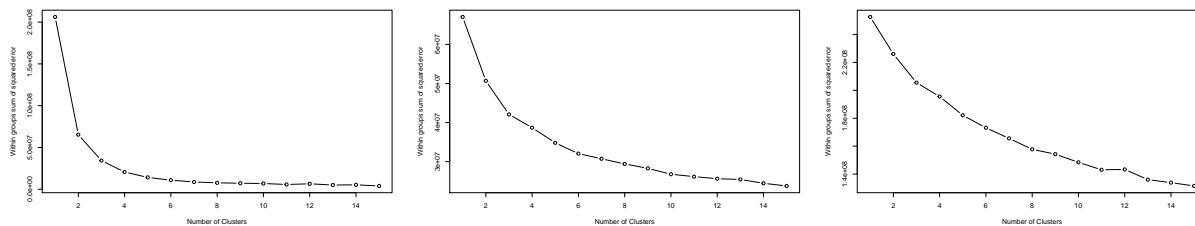



Figure S1. Sum of squared distance error within a cluster against number of clusters for daily mean O_3 LT(t), S(t) and W(t).

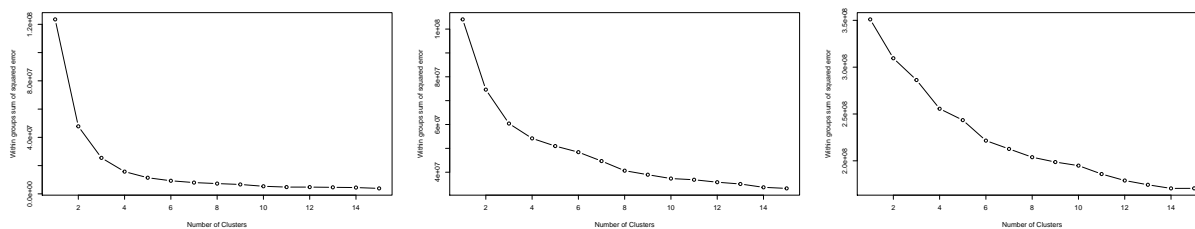



Figure S2. Same as Fig. S1 but for MDA8 O_3.

S2   Assessment of clusters

Silhouette width (S_w) is considered a good metric to assess the goodness of a clustering. Positive Silhouette width values mean that objects within the cluster have high similarity while negative ones mean that the objects have a low similarity with the rest objects in the cluster. Here, a S_w is assigned to each member of the cluster (here each measurement site) to assess the level of

5   similarity in the cluster. It is defined as:

$$S_w = \frac{b(i) - a(i)}{max[a(i) - b(i)]} \tag{S1}$$

where $a(i)$ the average distance (dissimilarity) of i with all other data points within the same cluster and $b(i)$ the lowest average distance of i to all points in any other cluster, of which i is not a member. Positive S_W value for a specific site means that similarity to other sites in the cluster is high. A negative S_w indicates that a site is more similar to a site from another

10   cluster than to its own cluster centroid, and is an indication of lower similarity to other sites in its own cluster.

S_w plots for daily mean O_3 are shown in Figs. S3-S5 and for the MDA8 O_3 in Figs. S6-S8.

Figure S3. Silhouette width plots for daily mean O_3 LT(t).

Silhouette plot of pam(x = data.matrix, k = 5, metric = "euclidean")

n = 291

5 clusters C_j

$j : n_j \mid \text{ave}_{i \in C_j} \ s_i$

1 : 94 | 0.20

2 : 92 | 0.13

3 : 77 | 0.08

4 : 4 | 0.57

5 : 24 | 0.13

Silhouette width s_i

Average silhouette width : 0.14

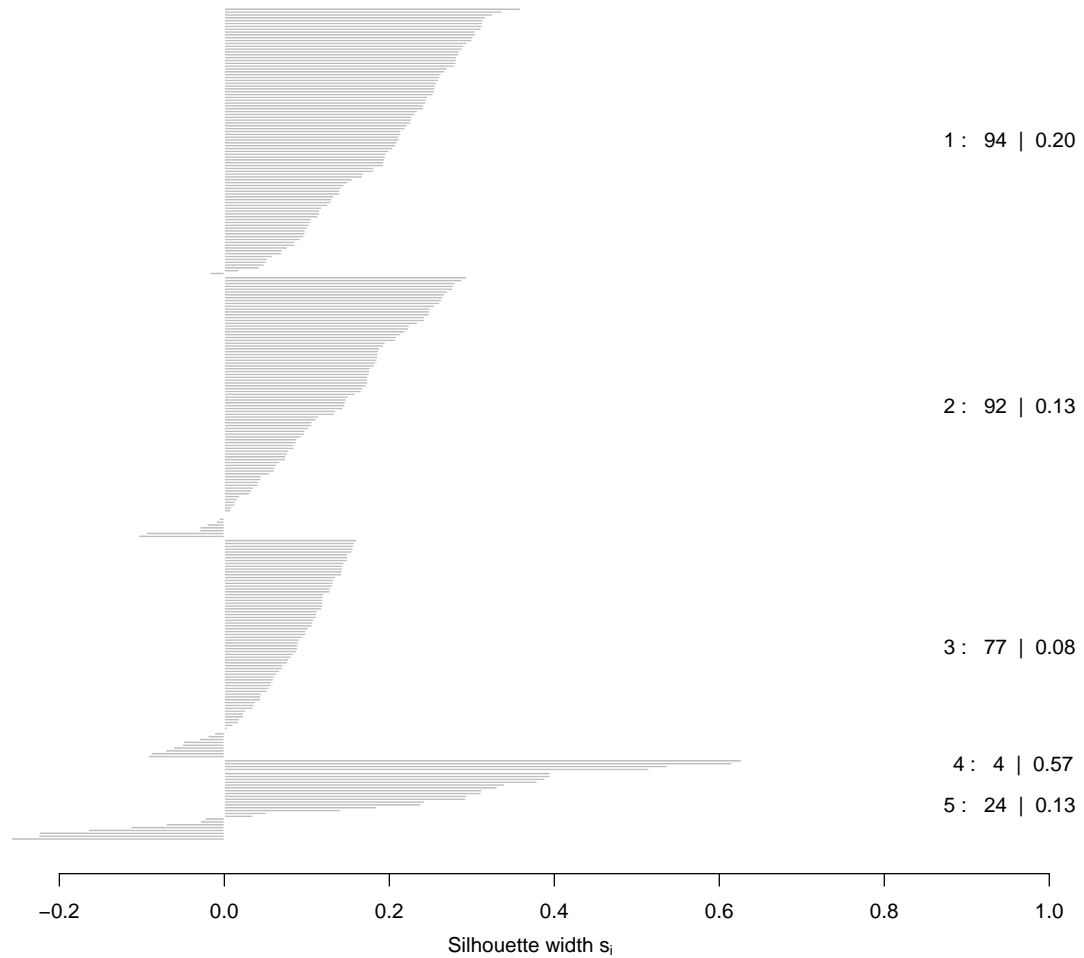



Figure S4. Same as in Fig. S3 but for S(t).

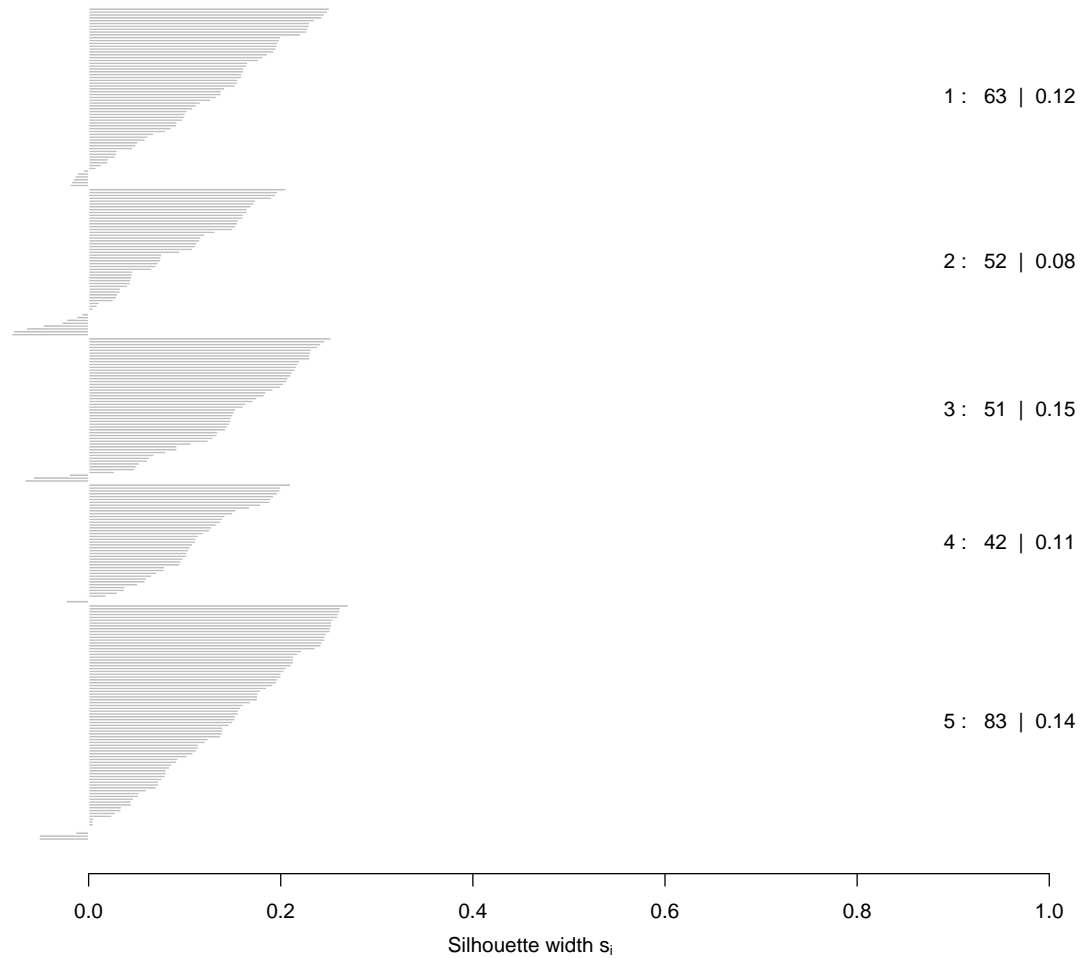



Figure S5. Same as in Fig. S3 but for S(t).

Silhouette plot of pam(x = data.matrix, k = 4, metric = "euclidean")

n = 291

4 clusters C_j
j : n_j | $ave_{i \in C_j}$ s_i

1 : 26 | 0.47

2 : 84 | 0.43

3 : 101 | 0.50

4 : 80 | 0.43

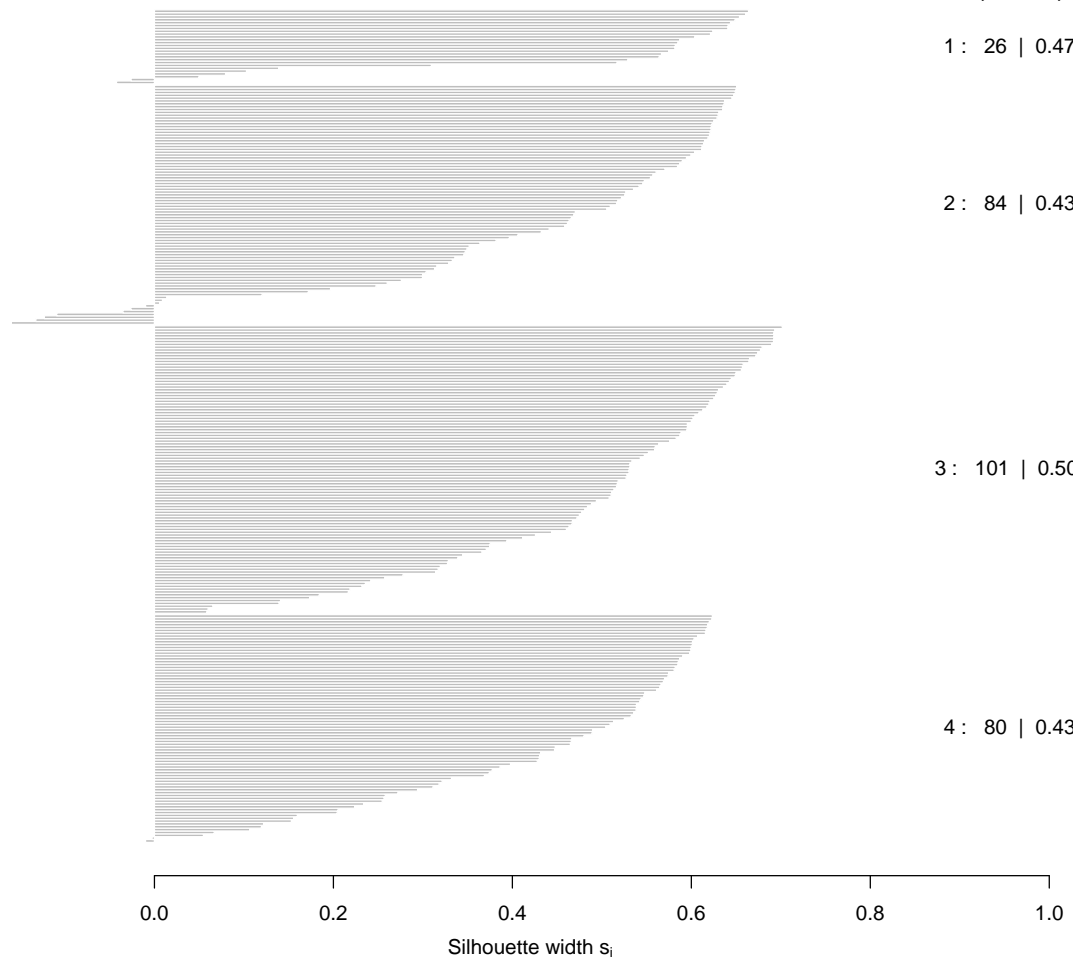
Silhouette width s_i

Average silhouette width : 0.46

Figure S6. Same as in Fig. S3 but for MDA8 LT(t).

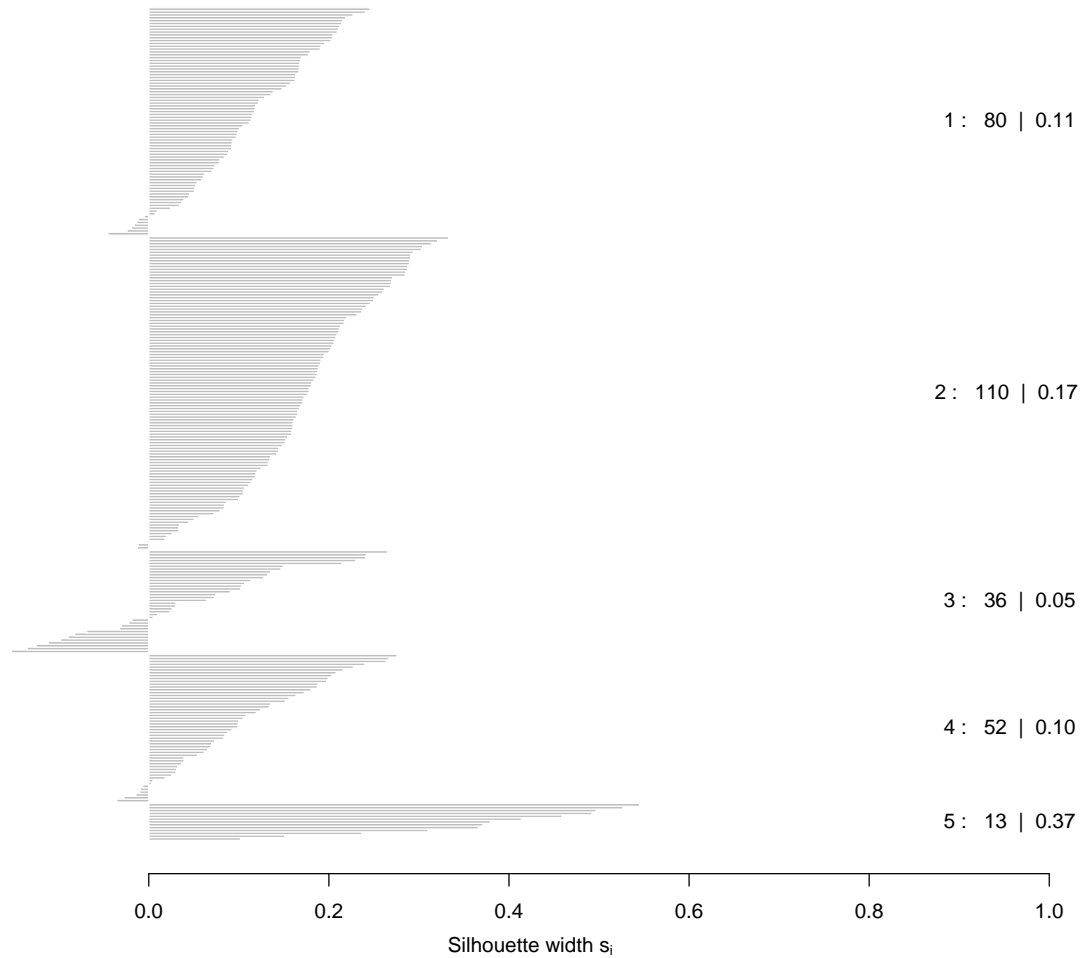



Figure S7. Same as in Fig. S6 but for S(t).

Figure S8. Same as in Fig. S6 but for W(t).

S3   Additional information on clusters

In this section we present more detailed information about the clusters extracted from the daily mean and MDA8 O_3 LT(t), S(t) and W(t). From the MDA8 O_3 S(t) the following five clusters were identified: (1) "West" contains biggest part of France, Belgium and Spain, (2) "East" includes eastern and central Germany, eastern Austria and Czech Republic, (3) "Central" covers

5   most part of Switzerland, some sites in central Germany and eastern Austria, (4) "North" is the cluster covering sites in northern Germany (North Sea region) and Netherlands, (5) "Atlant-Infl" covers most of the UK, some sites in the West coast and Scandinavia influenced by background O_3.

The five clusters obtained from the MDA8 O_3 W(t) are the following: (1) "CentralEast" covers the region of eastern Austria, Czech Republic and Poland, (2) "Central" for central and southern Germany and northern Switzerland, (3) "NorthWest" is the

10   cluster in northern France, Belgium and south UK (4) "SouthWest" in central and southern France, Spain and Italy, and (5) "NorthEast" includes the region of northern Germany, Netherlands and southern Scandinavia.

The following five clusters for daily mean O_3 W(t) were obtained (Fig. S12): (1) "CentralEast" that includes eastern Austria, Czech Republic, Poland and stations in Baltic and northern Scandinavia, (2) "Central" with stations in south-western Germany and Switzerland, (3) "NorthWest" covering the UK, northern France and Belgium, (4) "SouthWest" with southern France and

15   Spain, i.e. mostly the Mediterranean region, and (5) "NorthEast" for northern Germany, Netherlands and southern Scandinavia.

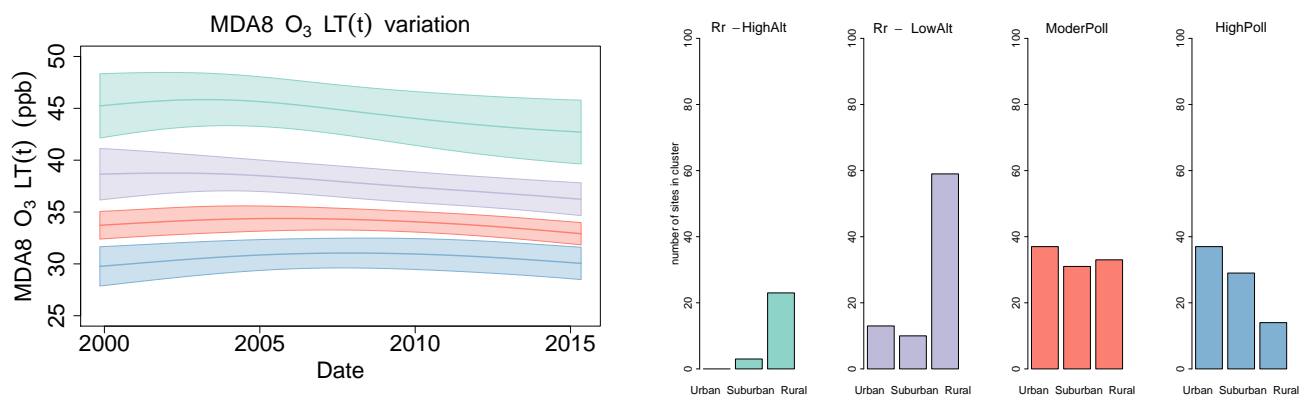



Figure S9. Clusters based on MDA8 O_3 LT(t). Map indicating the sites that belong in each cluster and average LT(t) in each cluster with $\pm$ the standard deviation of the sites that have SW>0.

The daily mean and MDA8 O_3 W(t)-clustering resulted in clusters with a regional dependence, because it is mainly driven by short-term local meteorological conditions. Comparing the clustering based on S(t) and W(t), we conclude that both lead to a similar regional site classification. However, there are some differences: (a) the "Central" obtained by the daily mean S(t)-clustering is divided in the W(t)-clustering in two clusters the "Central" (south-western Germany and Switzerland without the

20   Austrian sites of the S(t) cluster "Central") and the "CentralEast" (eastern Austria, Czech Republic, Poland and Baltic region, in S(t) those sites fall into the cluster "NorthEast"), (b) the "West" in the S(t)-clusters appears in the W(t)-clusters as two

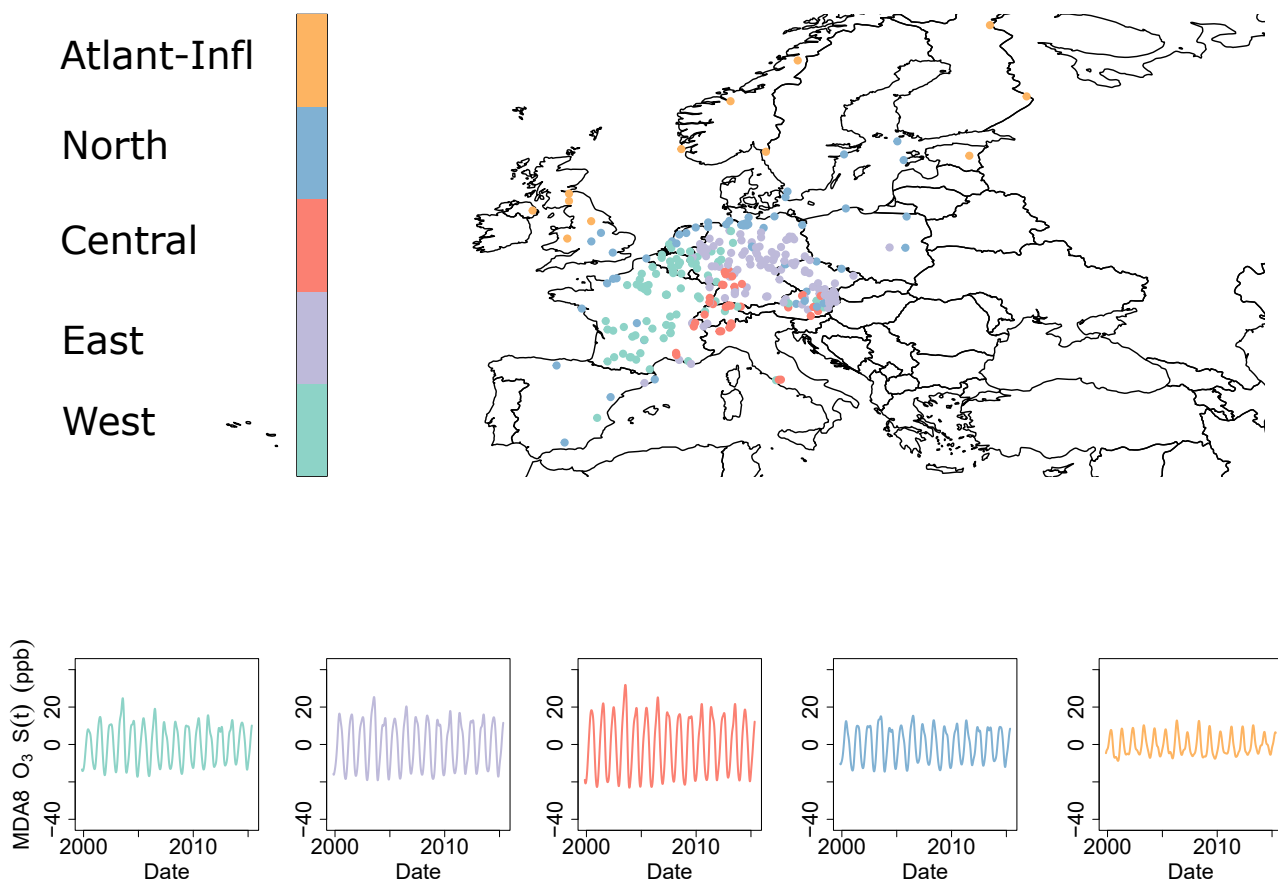



Figure S10. Clusters derived from MDA8 O$_3$ S(t). Map indicating the sites that belong in each cluster and average S(t) (standard deviation) in each cluster of the sites that have SW>0.

separate clusters the "NorthWest" (UK, northern France and Belgium) and "SouthWest" (southern France and Spain), which in the S(t)-clusters belong together in the "West" cluster.

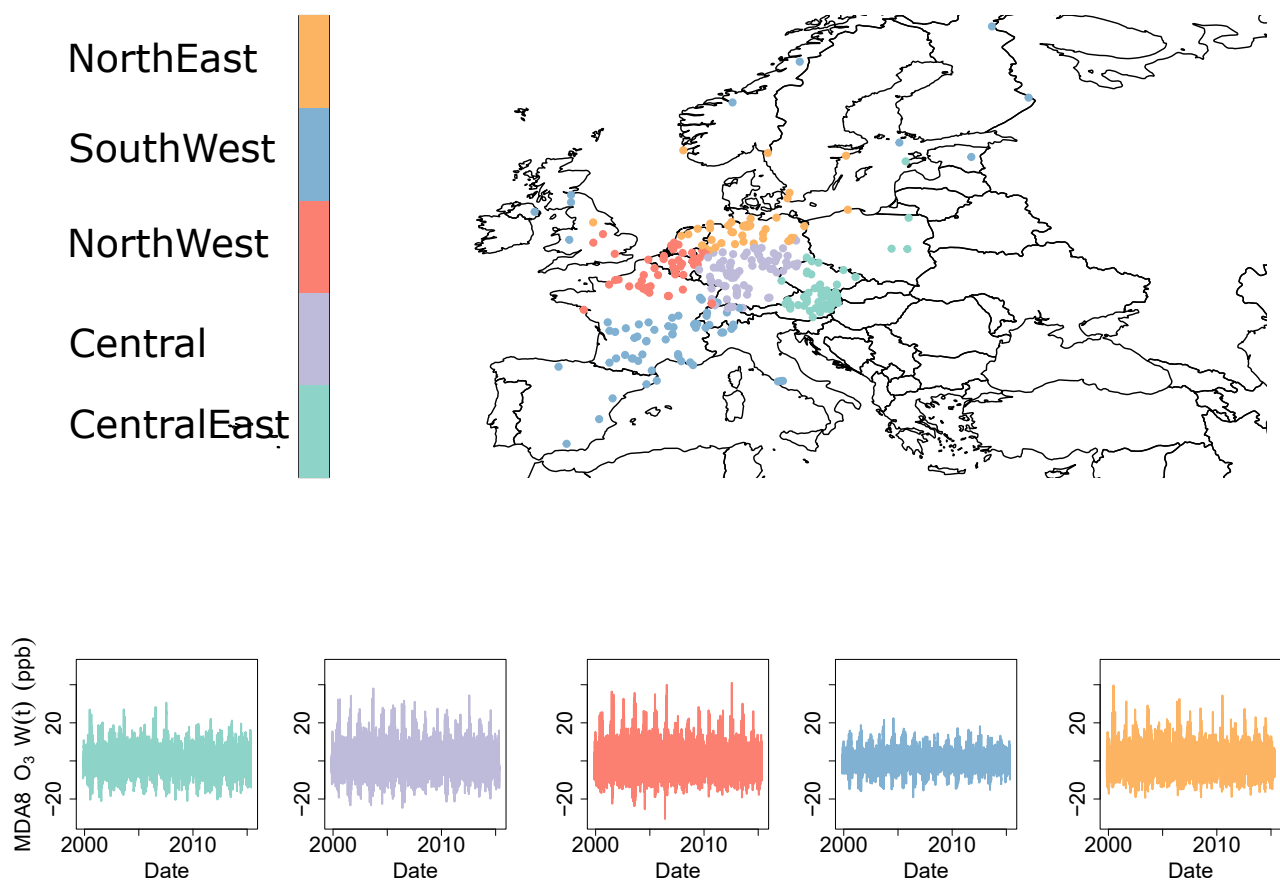



Figure S11. Clusters derived from MDA8 O₃ W(t). Map indicating the sites that belong in each cluster and average W(t) in each cluster of the sites that have SW>0.

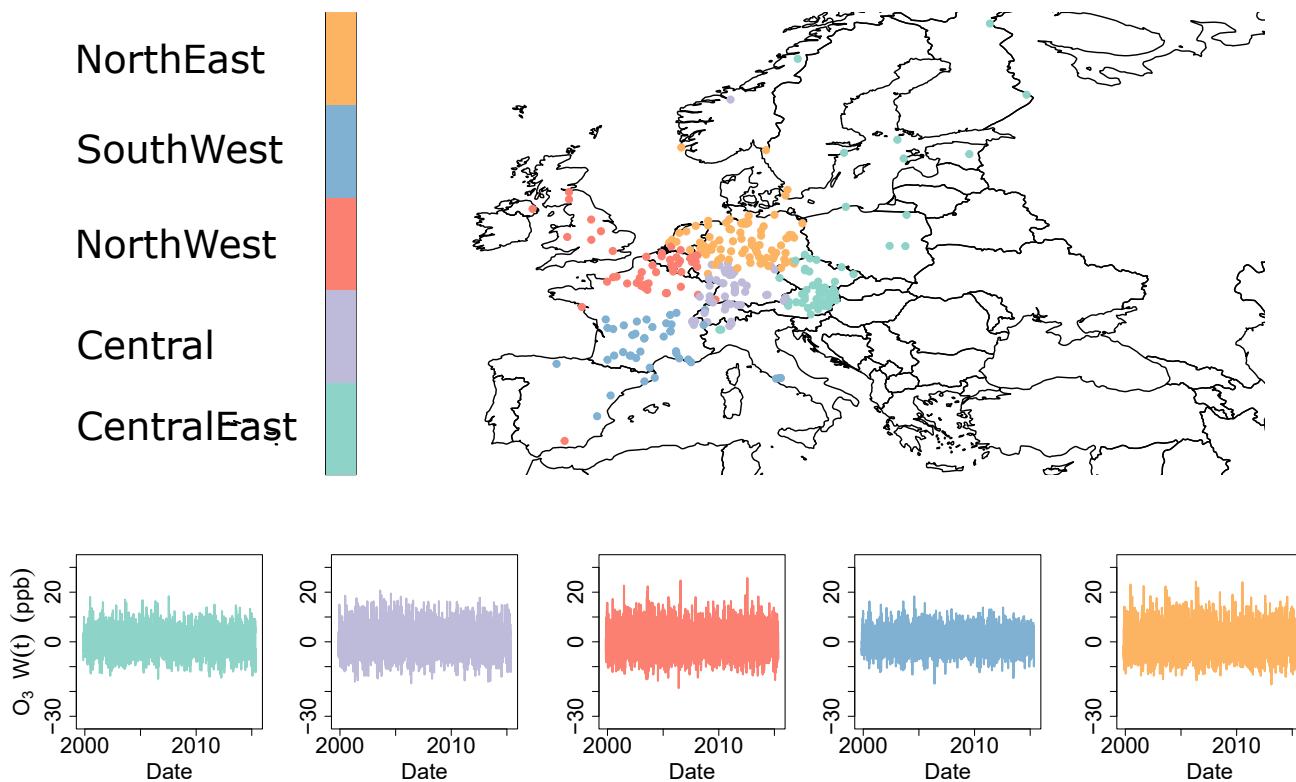



Figure S12. Clusters derived from daily mean O_3 W(t).(a) Map indicating the sites that belong in each cluster and average W(t) in each cluster of the sites that have SW>0.

S4   MDA8 and 4-MDA8 trends

Trends of MDA8 O_3 concentrations are negative in most sites. (Fig. S13). For the MDA8 O_3 in 69% of all sites significant trends were found (80% among the significant trends were negative and 19% positive). The most pronounced decrease is observed in rural sites ("Rural-HighAlt" and "Rural-LowAlt"). The level off or small increase in the "HighPoll" stations can be attributed to the smaller rate of reduction of VOCs, which resulted to reduced titration of O_3 by NO.

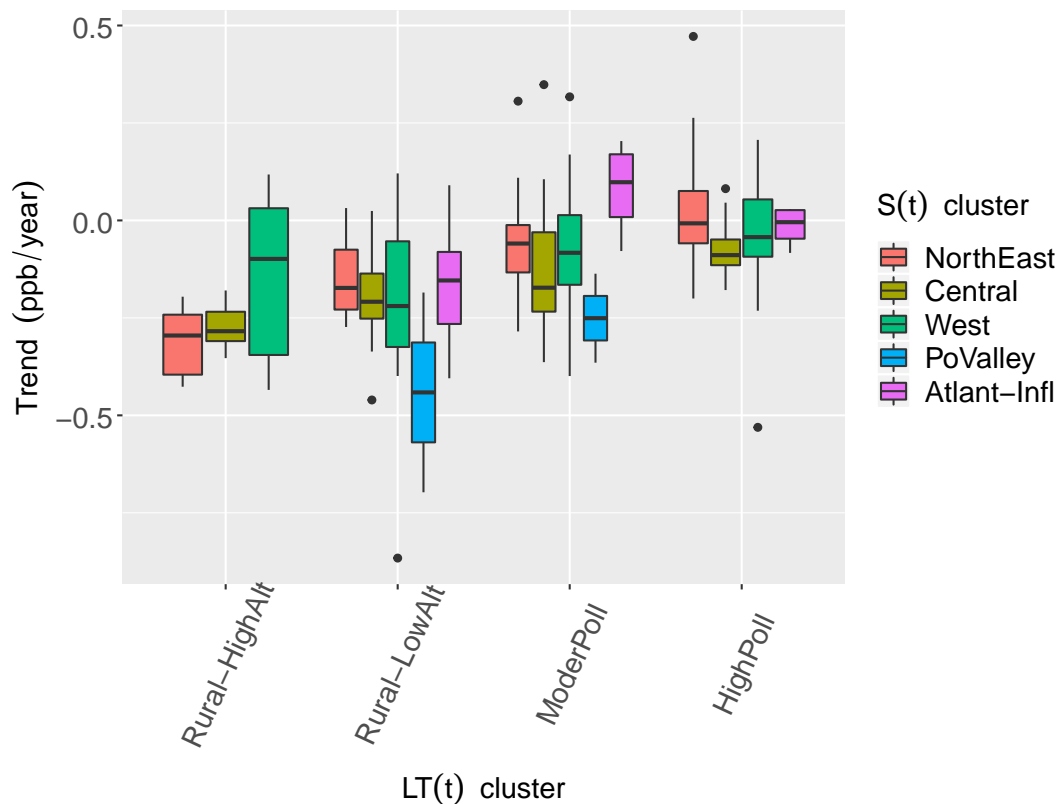



Figure S13. De-seasonalized MDA8 O_3 trends for the daily mean O_3 LT(t)- and S(t)-clusters.

Significant negative meteo-adjusted 4-MDA8 trends (Fig. S14 were observed in 47% of the sites respectively, while without meteo-adjustment significant negative trends were identified in only 18% of the sites. Overall, in 96% of all sites negative 4-MDA8 trends respectively were calculated.

S5   Sites with negative Sihlouette width

Here, the sites with negative SW that were not considered in the discussion of the trends are presented. In the LT(t)-clustering four sites with negative SW were identified (Fig. S15), in the S(t) 26 sites (Fig. S16) and in the W(t) 24 sites. This is probably

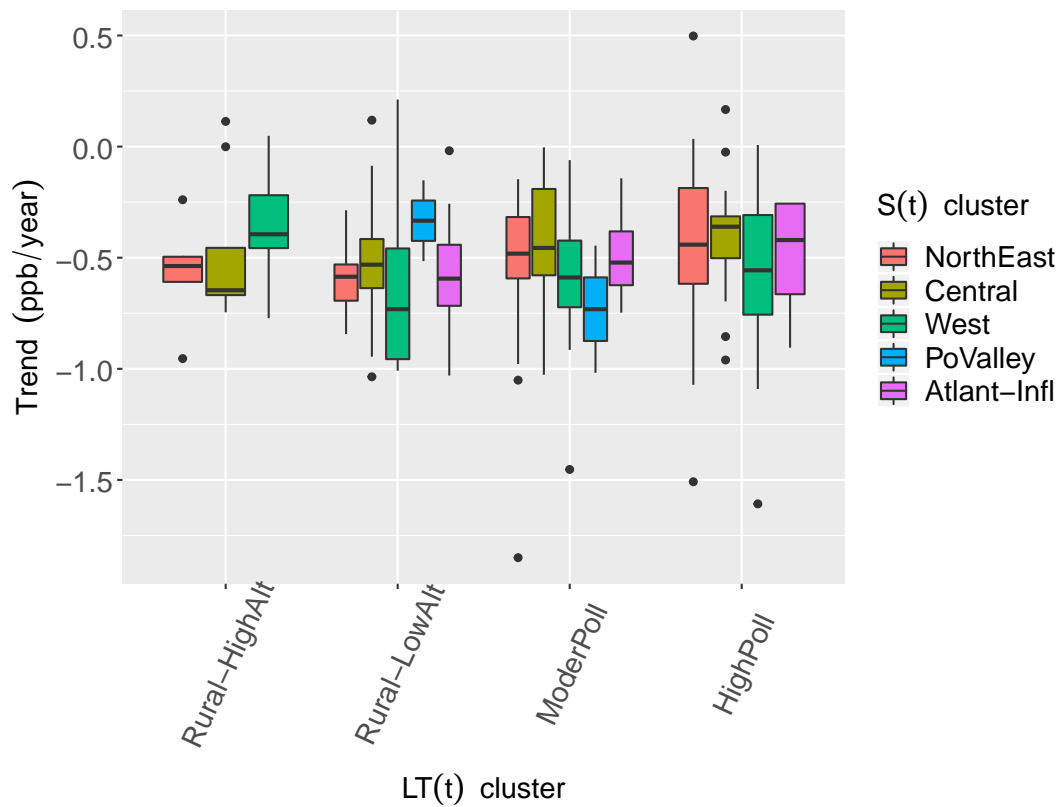



Figure S14. Trends of meteo-adjusted 4-MDA8 for the daily mean O_3 LT(t)- and S(t)-clusters.

due to some distinct differences in the LT(t), S(t) and W(t) compared to the rest sites in the cluster. By excluding these sites from the discussion of the long-term trends, we avoid strong deflections from the average behavior in the cluster.

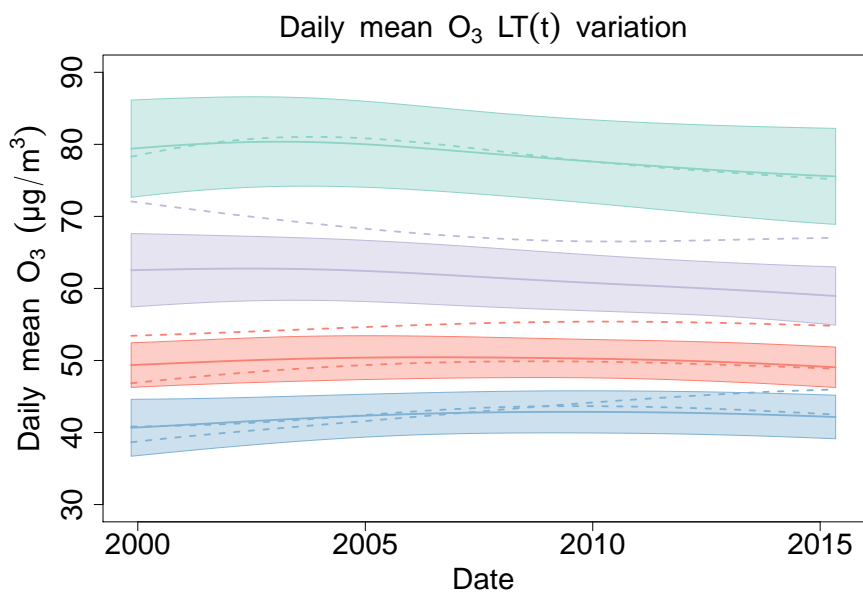



Figure S15. Sites with negative SW in the LT(t)-clustering, in comparison with the clusters average LT(t).

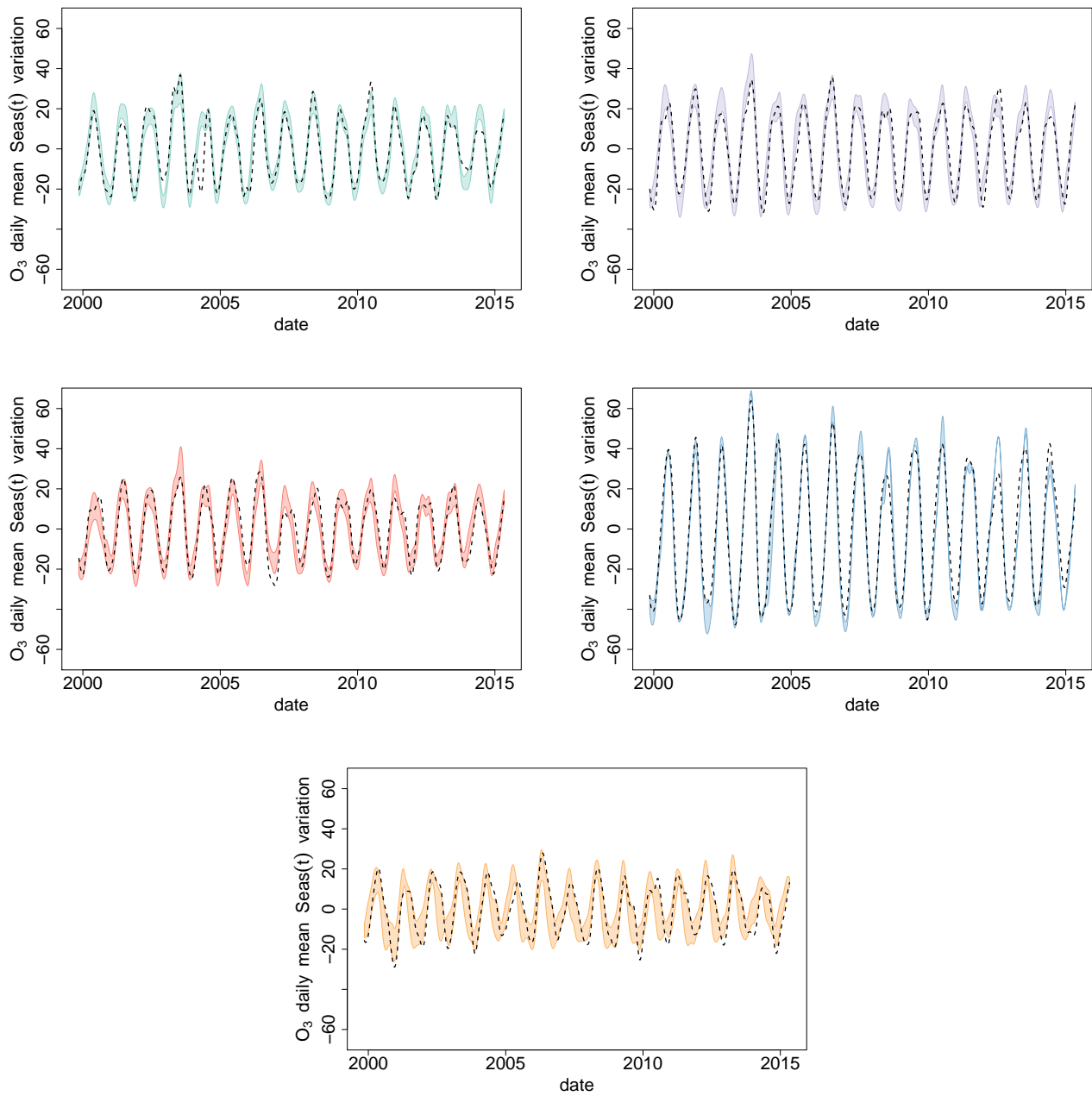



Figure S16. Example cases of sites with negative SW in the S(t)-clustering (black dashed line) in comparison with the clusters average S(t).