

## Reviewer 1:

*This was a very comprehensive review of European NO<sub>2</sub> measurements 2010-2014. Just some minor corrections for readers not so familiar with cluster analysis and other concepts. A few general comments may help the authors to explain some of the concepts a bit more clearly:*

**Response:** We thank the reviewer for their time spent reviewing the manuscript, for their positive comments on our work, and for the minor suggestions, which have improved the paper. We respond to their general and specific suggestions below.

*Cluster analysis using hourly and monthly averages: Section 2.1 line 27: mention the hourly and monthly average contribution to the annual average NO<sub>2</sub>.*

**Response:** We agree that specifically referring to the statistics for which increasing variance was explained is useful, and have therefore made the following amendment.

**Original Text P7 L5:** ‘The aim was to maximise the explained variance between sites with as few clusters as possible, and therefore, for each 10 µg m<sup>-3</sup> 2010-2014 NO<sub>2AA</sub> bin, the point at which increasing the number of clusters resulted in relatively small increases in the proportion of variance explained was identified.’

**Amended Text P7 L5:** ‘The aim in selecting the number of clusters of sites was to maximise the explained variance in the percentage contribution of each month, hour of day, and hourly NO<sub>2</sub> concentrations in 5 µg m<sup>-3</sup> bins to NO<sub>2AA</sub> between sites with as few clusters as possible. Therefore, for each 10 µg m<sup>-3</sup> 2010-2014 NO<sub>2AA</sub> bin, the point at which increasing the number of clusters resulted in relatively small increases in the proportion of variance explained was identified.’

*This is how the cluster analysis is calculated. Could you explain how sometimes an hourly contribution might be more or less important? Does it pull out the sites that have large diurnal variability and those that stay the same.*

**Response:** In the cluster analysis, the grouping of sites is based on the similarity across 77 variables, the percentage contribution from 12 months of the year, 24 hours of the day, and 41 x 5 µg m<sup>-3</sup> hourly NO<sub>2</sub> concentration bins. At the start of the clustering, all sites are contained in individual clusters (each containing 1 site). At each step, the clusters are merged that result in the smallest increase in within cluster variance, i.e. for this first step, the two sites for which the 77 variables are most similar are grouped into one cluster. Hence the clustering algorithm is identifying groups of sites for which the pattern of percent contributions to NO<sub>2AA</sub> across the day, month and across the hourly NO<sub>2</sub> concentration distribution are similar. As the reviewer suggests, this can result in sites with large diurnal variability being grouped in one cluster, and those with less variability grouped in another cluster (Figure 5 shows this clearly for sites with 2010-2014 NO<sub>2AA</sub>

between 50 and 60  $\mu\text{g m}^{-3}$ , with sites in Cluster 4 having large diurnal variation (shown by the larger difference in contribution to  $\text{NO}_{2\text{AA}}$  from the top and bottom 6 hours), and Clusters 1 and 5 having a smaller difference). In addition to grouping sites with relatively large and small differences in diurnal variation, it also highlights sites with different patterns in diurnal contribution, as shown in Figure 6, which shows a differences in the proportion of hours with large contribution to  $\text{NO}_{2\text{AA}}$  occurring in the morning/evening for different types of sites.

However, the key advantage of the cluster analysis undertaken in this work is the grouping of sites based on the similarity of monthly, hour of day, and hourly  $\text{NO}_2$  contribution to  $\text{NO}_{2\text{AA}}$  together, rather than evaluating these contributions separately. This facilitated the identification of sites where, e.g. the contribution to  $\text{NO}_{2\text{AA}}$  from hours of the day was similar, but monthly contributions, or contributions from 5  $\mu\text{g m}^{-3}$  hourly  $\text{NO}_2$  concentration bins differed. The major distinctions and similarities between clusters are highlighted in Section 3.2 (e.g. P10 L17 highlights the similarity in monthly contribution to  $\text{NO}_{2\text{AA}}$  for the two major clusters for sites with 2010-2014  $\text{NO}_{2\text{AA}}$  between 60 and 70  $\mu\text{g m}^{-3}$ , but differences in diurnal contribution, and in contribution from across the diurnal  $\text{NO}_2$  concentration distribution). The identification of similarity across clusters in some contributions to  $\text{NO}_{2\text{AA}}$  (e.g. across the year), but differences in others (e.g. across the day or hourly  $\text{NO}_2$  concentration distribution) underpins aspects of the discussion, e.g. P13 L26: ‘However, the extent of the reduction from reducing emissions during peak hours of the day, and from reducing the short-term peak hourly  $\text{NO}_2$  concentrations, varies between sites, which may reflect variability between sites in meteorological conditions or in traffic flows (i.e. strength of the emission source in close proximity to the monitoring sites).’

We feel it is important for the reader to understand the user-defined decisions made in the clustering algorithm. These user choices include the clustering algorithm (hierarchical or non-hierarchical methods), the choice of distance measure (i.e. how the similarity across sites is quantified), and, for hierarchical methods, the method by which sites are grouped (Ward’s method was used here, alternatives include Average, Minimum or Maximum linkage (Kaufman and Rousseeuw, 1990)). We have therefore acknowledged that there are various types of cluster algorithms, and that the clustering result may be affected by these choices. We have referred to a previous study that analysed different clustering algorithms, and found that Ward’s method hierarchical cluster analysis performed effectively compared to alternative methods (Mangiameli et al., 1996). We have also added more text to highlight the key outputs from the cluster analysis, and directed the reader to where these are discussed.

**Original Text P6 L12:** ‘The data capture for each hour of the day, and each month of the year was also calculated for each year at each site. For those sites in each 2010-2014  $\text{NO}_{2\text{AA}}$  bins with >75% data capture in each month and each hour, the contributions of

each month, hour of day, and hourly NO<sub>2</sub> concentrations in 5 µg m<sup>-3</sup> bins to NO<sub>2AA</sub> were calculated. These sites were further analysed to identify the major differences in the ‘state’ statistics across sites with similar 2010-2014 NO<sub>2AA</sub> values. The relatively low number of sites in the 70-80 µg m<sup>-3</sup> and >80 µg m<sup>-3</sup> bins meant that variation in these statistics was examined across the sites individually. However, for the lower NO<sub>2AA</sub> concentration bins, which contained up to several hundred sites, Ward’s hierarchical cluster analysis was used to group sites according to similarity in the contribution of different months, hours, and hourly NO<sub>2</sub> concentration bins to NO<sub>2AA</sub>. In this process, each monitoring site initially constituted its own cluster. Clusters were then merged based on the similarity of the 2010-2014 average values of the percent contribution to NO<sub>2AA</sub> from January-December, from 00:00 to 23:00, and from each 5 µg m<sup>-3</sup> hourly NO<sub>2</sub> concentration bin. At each step, the two clusters were merged which results in the smallest increase in within-cluster variance, and this was repeated until all sites are contained within a single cluster (Kaufman and Rousseeuw, 1990; Ward, 1963). This process is represented by a dendrogram, which was then ‘cut’ at an appropriate level to determine the number of clusters, and the grouping of sites. The cut-off point was determined from the visual identification of ‘elbows’ in the plots of proportion of variance explained against the number of clusters into which sites were divided. The aim was to maximise the explained variance between sites with as few clusters as possible, and therefore, for each 10 µg m<sup>-3</sup> 2010-2014 NO<sub>2AA</sub> bin, the point at which increasing the number of clusters resulted in relatively small increases in the proportion of variance explained was identified. The cut-off points selected grouped sites into between 5 and 8 clusters, which explained between 57 and 75% of total between-site variance, as shown in Figures S1-S7.’

**Amended text P6 L12:** ‘The data capture for each hour of the day, and each month of the year was also calculated for each year at each site. For those sites in each 2010-2014 NO<sub>2AA</sub> bins with >75% data capture in each month and each hour, the percent contribution of hourly NO<sub>2</sub> concentrations in each of the 12 months, in each of the 24 hours of the day, and from hourly NO<sub>2</sub> concentrations across the year in 5 µg m<sup>-3</sup> bins (i.e. from hourly NO<sub>2</sub> concentrations between 0-5 µg m<sup>-3</sup>, 5-10 µg m<sup>-3</sup> ... 195-200 µg m<sup>-3</sup>, >200 µg m<sup>-3</sup>) to NO<sub>2AA</sub> were calculated. These sites were further analysed to identify the major differences in the percentage contribution to NO<sub>2AA</sub> from each month of the year, hour of the day, and hourly NO<sub>2</sub> concentrations in 5 µg m<sup>-3</sup> bins across sites with similar 2010-2014 NO<sub>2AA</sub> values. The relatively low number of sites in the 70-80 µg m<sup>-3</sup> and >80 µg m<sup>-3</sup> bins meant that variation in these statistics was examined across the sites individually. However, for the lower NO<sub>2AA</sub> concentration bins, which contained up to several hundred sites, Ward’s hierarchical cluster analysis was used to group sites according to similarity in 77 variables, i.e. the percent contribution to 2010-2014 NO<sub>2AA</sub> of different months (12 values), hours of the day (24 values), and 5 µg m<sup>-3</sup> hourly NO<sub>2</sub> concentration bins (41 values, percent contribution from hourly NO<sub>2</sub> concentrations 0-5 µg m<sup>-3</sup>, 5-10 µg m<sup>-3</sup> ... 195-200 µg m<sup>-3</sup>, >200 µg m<sup>-3</sup>). In this process, each monitoring site initially constituted its own cluster. Clusters were then merged based on the similarity

of the 2010-2014 average values of the percent contribution to  $\text{NO}_{2\text{AA}}$  from January-December, from 00:00 to 23:00, and from each  $5 \mu\text{g m}^{-3}$  hourly  $\text{NO}_2$  concentration bin (i.e. the two most 'similar' clusters were merged). Euclidean distance was used as the measure of the similarity between these variables between clusters. At each step, the two clusters were merged which results in the smallest increase in within-cluster variance, and this was repeated until all sites are contained within a single cluster (Kaufman and Rousseeuw, 1990; Ward, 1963). Other hierarchical and non-hierarchical clustering algorithms and distance measures could have been used in place of the Ward's hierarchical algorithm, and the Euclidean distance, respectively, that were applied in this work. Selection of alternative methods could influence the clustering result (Kaufman and Rousseeuw, 1990), but the Ward's hierarchical cluster algorithm applied here has been shown here to perform effectively compared with alternative approaches, including with respect to the dispersion of observations, the presence of outliers, and irrelevant variables (Mangiameli et al., 1996).

The clustering of monitoring sites in each  $10 \mu\text{g m}^{-3}$  2010-2014  $\text{NO}_{2\text{AA}}$  bin process is represented by a dendrogram, which was then 'cut' at an appropriate level to determine the number of clusters, and the grouping of sites. The cut-off point was determined from the visual identification of 'elbows' in the plots of the proportion of variance explained against the number of clusters into which sites were divided. The aim in selecting the number of clusters of sites was to maximise the explained variance in the percentage contribution of each month, hour of day, and hourly  $\text{NO}_2$  concentrations in  $5 \mu\text{g m}^{-3}$  bins to  $\text{NO}_{2\text{AA}}$  between sites with as few clusters as possible. Therefore, for each  $10 \mu\text{g m}^{-3}$  2010-2014  $\text{NO}_{2\text{AA}}$  bin, the point at which increasing the number of clusters resulted in relatively small increases in the proportion of variance explained was identified. The cut-off points selected grouped sites into between 5 and 8 clusters. A key feature of the clustering result is the proportion of variance across sites in the percent contribution to  $\text{NO}_{2\text{AA}}$  from months, hours of the day and hourly  $\text{NO}_2$  concentration bins that is explained by dividing sites into a certain number of clusters. The clusters produced in this work explained between 57 and 75% of the variance between sites in the  $10 \mu\text{g m}^{-3}$  2010-2014  $\text{NO}_{2\text{AA}}$  bin ranges, as shown in Figure S2. This means that there remains between 43% and 25% variability in these variables between sites with similar 2010-2014  $\text{NO}_{2\text{AA}}$  that is not explained by dividing sites into the clusters produced in this analysis. Other key aspects of the clustering are the location of sites included in each cluster, and the similarities and differences in the value of each variable (i.e. the percent contribution from each month, hour of day and  $5 \mu\text{g m}^{-3}$  hourly  $\text{NO}_2$  concentration bins) between clusters. These aspects are described in Section 3.2.'

*What factor is fed into the cluster analysis? Difference between hourly concentration and annual average and difference between monthly and annual?*

**Response:** As outlined above, the 77 variables used to group sites encapsulated variation in the contribution to  $\text{NO}_{2\text{AA}}$  across the year, day and hourly  $\text{NO}_2$  concentration distribution. These are the percentage contribution to the annual average from hourly

NO<sub>2</sub> concentrations occurring in each month of the year, hour of the day, and in 5 µg m<sup>-3</sup> bins. This has been made clearer in the text (see additions in response to the comment directly above).

*Section 2.1 Line 12 – make clearer how the 5 ugm-3 bins are different to the 10 ugm-3 bins. I will come back to this in Fig 6.*

**Response:** We have amended this text in response to a comment below to make clear that the 5 µg m<sup>-3</sup> bins relate to *hourly* NO<sub>2</sub> concentrations (and the percent contribution to the annual average from hourly NO<sub>2</sub> concentrations in these ranges), as opposed to the 10 µg m<sup>-3</sup> bins, that disaggregate sites based on annual NO<sub>2</sub> concentrations.

*Section 2.1 Line 14-can you explain what “state” statistics is?*

**Response:** This reference to a state statistic has been replaced with an explicit explanation of the statistics that are being referenced here.

**Original Text P6 L16:** ‘These sites were further analysed to identify the major differences in the ‘state’ statistics across sites with similar 2010-2014 NO<sub>2AA</sub> values.’

**Amended Text P6 L16:** ‘These sites were further analysed to identify the major differences in the percentage contribution to NO<sub>2AA</sub> from each month of the year, hour of the day, and hourly NO<sub>2</sub> concentrations in 5 µg m<sup>-3</sup> bins across sites with similar 2010-2014 NO<sub>2AA</sub> values.’

*Would a Figure in the appendix showing what a cluster analysis looks like, with the branches shown and the regions in S8 put in blocks along the x axis and coloured even if we cant see the individual station names be useful?*

**Response:** The reviewer is correct that one of the advantages of the hierarchical clustering approaches (as opposed to non-hierarchical methods such as k-means) is that observations (in this case sites) can be summarised as a dendrogram. This allows the linkages between observations and clusters to be viewed, as well as the ‘level’ at which the dendrogram is ‘cut’ to produce the set of clusters. However, for this application there are a few factors that we think reduce the utility of this visualisation. Most importantly, as identified by the reviewer, there are too many sites to be individually identified on a normal sized figure, and hence it would not be possible to use this type of plot to view the linkage between individual sites. Secondly, a key conclusion of our paper is that sites with similar variation in contributions to NO<sub>2AA</sub> were largely not grouped into specific regions. Hence colouring the x-axis according to region would have multiple colours across each cluster, which repeats the information already included in Figure 3.

With the limitations of summarising the cluster results using a dendrogram, it was our intention in the original manuscript to use a series of figures to extract and describe the most important aspects of the cluster analysis. Firstly, it is key to understand the extent to which the number of clusters selected explains the variation between sites. We show this clearly in Figures S1-S7 for sites with 2010-2014 NO<sub>2AA</sub> in each 10 µg m<sup>-3</sup> concentration bin (now Figure S2), with the number of cluster selected highlighted in red. Next, it is important that the reader understands the spatial distribution of sites assigned to each cluster, and this is shown in Figure 3 in the main manuscript, and then in supplemental maps in more detail. Finally, it is then important to understand how the variables used to group sites (i.e. percentage contribution NO<sub>2AA</sub> in each month, hour of day and hourly NO<sub>2</sub> distribution) vary between clusters, which are summarised in Figures 4-7.

We do not think that including dendrograms would increase the explanation of the clustering results that is already included in the current figures in the main manuscript and supplement. We do agree that it is important to clearly explain the concepts of cluster analysis, and its limitations. To this end we have expanded the description in the text of the cluster analysis methodology, the assumptions made, what is produced from the cluster analysis, and explanation as to what the different figures in the paper show in terms of the clustering results (see amended text in response to a previous comment (starting at P6 L16 in the main manuscript)).

*Referring to Figure 1 doesn't really help the reader so much in explaining what a cluster analysis does and looks like?*

**Response:** The cluster analysis methodology is described on P6 (L16-P7 L23). We do not refer to Figure 1 in this description.

*In Fig 4, 5 and 6 with the explanations of what type of sites generally each cluster is made of, it is not easy to work out what the difference is. Maybe you justify in the end that there is no difference between them ins most of these figures, so the cluster analysis is not specific enough to understand the characteristics of any one station?*

**Response:** As outlined above in response to a previous comment, the key advantage of the methodology adopted here is that the cluster analysis identified similarity across sites in terms of monthly, hour of day and hourly NO<sub>2</sub> concentration bin contributions to NO<sub>2AA</sub>. This means that it was possible to determine some clusters of sites for which, e.g. the contribution of different months was similar, but for which the contribution from across the day and hourly NO<sub>2</sub> concentration distribution, varied. We synthesise the key differences between clusters (informed by Figures 4, 5, 6 and 7) within the results text of the main manuscript. Key results from the interpretation of these figures, and the clustering result in general are the larger monthly variation in contribution to NO<sub>2AA</sub> at sites in northern Italy, the similar contribution from months of the year at sites in other regions with high NO<sub>2AA</sub>, the increasing winter contribution as NO<sub>2AA</sub> decreases, and the

distinction between sites with similar monthly contribution, but differences in hourly contribution (larger/smaller diurnal variation), and hourly NO<sub>2</sub> concentration distribution contribution (broader/narrower set of hourly NO<sub>2</sub> concentrations determining NO<sub>2AA</sub>).

In terms of the relevance of these results to any one particular site, it is not the aim of this paper to exhaustively describe the differences between individual sites, but to explore the main differences across the European monitoring network. However, for sites with 2010-2014 NO<sub>2AA</sub> <70 µg m<sup>-3</sup>, where the cluster analysis was used to group sites, we do quantify the extent to which the main differences between groups of sites explain the variation between all sites in the analysis. Figures S1-S7 of the original submission (now Figure S2) show the proportion of the variation between all sites in a 10 µg m<sup>-3</sup> 2010-2014 NO<sub>2AA</sub> range that is explained by the number of clusters into which sites are grouped. This gives an indication of the remaining variability between sites within clusters that is not explored within this analysis. We have now added additional text acknowledging that there is variability between individual sites within clusters that is not considered in this analysis (see amended text in response to a comment above (P7 L18)).

*% contribution to NO<sub>2AA</sub> 2.1, p.5 Line 26: This is used throughout and is not entirely clear. % contribution to NO<sub>2AA</sub> from each month of the year and hour of the day. Somehow I am left not entirely understanding how this was done – could it be explained better?*

**Response:** We have added additional explanation as to how these statistics were calculated on first usage of this term (P6 L17).

**Original Text P6 L12:** ‘The data capture for each hour of the day, and each month of the year was also calculated for each year at each site. For those sites in each 2010-2014 NO<sub>2AA</sub> bins with >75% data capture in each month and each hour, the contributions of each month, hour of day, and hourly NO<sub>2</sub> concentrations in 5 µg m<sup>-3</sup> bins to NO<sub>2AA</sub> were calculated.’

**Amended Text P6 L12:** ‘The data capture for each hour of the day, and each month of the year was also calculated for each year at each site. For those sites in each 2010-2014 NO<sub>2AA</sub> bins with >75% data capture in each month and each hour, the percent contribution of hourly NO<sub>2</sub> concentrations in each of the 12 months, in each of the 24 hours of the day, and from hourly NO<sub>2</sub> concentrations across the year in 5 µg m<sup>-3</sup> bins (i.e. from hourly NO<sub>2</sub> concentrations between 0-5 µg m<sup>-3</sup>, 5-10 µg m<sup>-3</sup> ... 195-200 µg m<sup>-3</sup>, >200 µg m<sup>-3</sup>) to NO<sub>2AA</sub> were calculated.’

*Five , ten and fifteen 5 ugm-3 hourly NO2 Please explain this on p.8 line 19 and p. 9 line24 and in Figure 6. Has this analysis added a lot to the analysis. If you keep in, please explain how this is done a bit more clearly.*



**Response:** We have added additional explanation as to how these statistics were calculated, and have kept them in the analysis (see amended text in response to a comment above (P6 L19)). Their importance is in highlighting groups of sites where a relatively large, or relatively narrow range of hourly NO<sub>2</sub> concentrations contribute to annual NO<sub>2</sub>. The EU, as well as promulgating an annual limit value, also has an hourly NO<sub>2</sub> limit value. In calculating the contribution from hourly NO<sub>2</sub> concentrations in 5 µg m<sup>-3</sup> bins, it is identified in this analysis that there is a contrast between sites with highest NO<sub>2AA</sub> at which reducing short-term peak concentrations in excess of the hourly EU limit value would also have an appreciable reduction in annual average concentrations, and those sites where short-term peak hourly NO<sub>2</sub> concentrations are less extreme, but where annual NO<sub>2</sub> concentrations are nevertheless substantially above the EU limit value.

*Trend analysis You look at 2000-2014 trends and compare how most recent (2010-2014) levels can see different trends from the past to the present. Have you explained how inferring a trend from the past against a level at the end of the trend may not be directly correlated to a trend forwards from now? It seems like it may be more useful to look at the 2000-2004 NO<sub>2</sub> levels and then show whether they increased or decreased. We don't want you to do that again but just justify why the end period is useful in understanding the trend up to that point.*

**Response:** We agree that discussion of the relevance for assessing trends at sites grouped by their annual NO<sub>2</sub> concentration at the end of the period should be added. The reason for grouping site based on the NO<sub>2AA</sub> level in 2010-2014 is to provide a link to the previous section, which explores the contributions to 2010-2014 NO<sub>2AA</sub>. The trend analysis aims to show whether these contributions, and resulting annual NO<sub>2</sub> concentrations have changed since 2000, or whether they have remained the same. This is important because over the 2000-2014 period, there have been substantial NO<sub>x</sub> emission reductions across the EU-28 Member States (EEA, 2016). The categorisation of sites based on 2010-2014 NO<sub>2AA</sub> shows whether NO<sub>2AA</sub> at sites that exceed the EU annual NO<sub>2</sub> limit value during the most recent period has decreased, or whether the changes in NO<sub>x</sub> emissions across the EU have not resulted in a decrease at these sites that currently exceed the EU standard.

We have added text to highlight these points, and to acknowledge, as the reviewer suggests, that by quantifying 2000-2014 trends at these sites, we are not making any prediction as to how trends are likely to progress into the future.

**Additional text P8 L4:** ‘Trends between 2000 and 2014 were assessed by grouping all qualifying sites based on the 2010-2014 NO<sub>2AA</sub> value at each site. These groupings were used to investigate whether changes in drivers of NO<sub>2</sub> variability (e.g. European-wide NO<sub>x</sub> emission reductions) had had different effects at sites where the magnitude of annual NO<sub>2</sub> concentrations during the most recent period (2010-2014) differed.’



**Additional text P18 L7:** ‘Finally, assessment of trends between 2000 and 2014 does not indicate how trends at each site are likely to progress into the future, which will be determined by future changes in NO<sub>x</sub> emissions, and meteorology.’

*Figures 8 and 12 are very different. Are they at odds with each other? Have you compared them? Fig 12 shows a strong linkage between NO<sub>2</sub> level and increases or decreases and Fig 8 seems to say that there are no significant trends*

**Response:** Figures 8 and 12 are not at odds with one another. Figure 8 shows trends in *annual* NO<sub>2</sub> concentrations, while Figure 12 focusses on changes in the contribution of *hourly* NO<sub>2</sub> concentration to NO<sub>2AA</sub>. Figure 8 does not show ‘that there are no significant trends’, it shows that at a substantial number of sites there has been a significant decreasing trend in annual NO<sub>2</sub> concentrations between 2000 and 2014. Figure 12 shows that at a substantial number of sites, the contribution to NO<sub>2AA</sub> from relatively high hourly NO<sub>2</sub> concentrations has decreased, with a corresponding increase in the contribution from relatively moderate and low hourly NO<sub>2</sub> concentrations. This general pattern of change in the distribution of hourly NO<sub>2</sub> concentrations at sites across Europe is consistent with decreasing trends in annual NO<sub>2</sub> concentrations.

*Specifics: 1. Intro p.4 Line 20: New sentence: This work*

**Response:** ‘This work’ is the start of the new sentence.

*p.4 Line 22: I would have liked more of an explanation for Fig 1 but maybe not all readers would*

**Response:** We have added at this point reference to Section 2.1, which describes Figure 1, and the statistics that underpin this analysis in more detail.

*p.6 Line 12. . . . in each of the 201-2014 NO<sub>2aa</sub> bins. . .*

**Response:** This sentence has been revised in line with the recommendation.

*p. 9 Line 28. . . between 60 and 70 ugm-3 (gap)*

**Response:** This sentence has been revised in line with the recommendation.

*4 Discussion: p.12, line 5 and 6. Cant this be summarised into “there is a not a regional correlation in the clustering”?*

**Response:** Yes, this is broadly what we are saying with these sentences. However, we have retained the original wording rather than that suggested by the reviewer because in some instances there were regional groupings (e.g. in northern Italy).

### *Figures*

*Fig 2 - > 80 is clearer than 80+*

**Response:** We have revised all figures in line with this recommendation.

*Fig 4 a – >80: would be nice to label the site name in the legend/ caption?!*

**Response:** The site id for those sites in the 70-80 and >80  $\mu\text{g m}^{-3}$  is shown in the figure. For sites contained within the clusters, there are simply too many to list them in the paper.

*Fig 4b line 5 caption: . . .the sites shown as..*

**Response:** This caption has been revised in line with the recommendation.

*Fig 6- comments earlier*

**Response:** We have revised this figure to change 80+ to >80.

*Figure 9- Zoom in (as there is nothing in N Scandinavia , N Africa, Atlantic, Eastern Europe). >80 is better than 80+. Either there are too many colour bins or it is too difficult to see the difference between a green colour and blue. It looks like green is inside but there is no green in the legend. Very difficult to see beyond the orange or green outer circle*

**Response:** We have revised this figure to change 80+ to >80. We have zoomed in, and have changed how the significance of the trend is signified. Rather than the outer circle colour showing the significance of the trend, it is now the shape of each point (circle,  $p < 0.05$ , triangle,  $p > 0.05$ ). Zooming in on the area of interest in the map, and having only one colour scale on the plot, shows more clearly the magnitude and significance of the trend at each site.

*Figure 10,11 and 12- a,b,c- h is a bit hidden on Figure move above*

**Response:** We have revised these figures in line with this recommendation.

*Appendix Figs S1-S7 could be smaller and put into one figure a- g?*

**Response:** We have combined these figures in line with this recommendation to form a new Figure S1.

## References

EEA: European Union emission inventory report 1990–2014 under the UNECE Convention on Long-range Transboundary Air Pollution (LRTAP). European Environment Agency Report No. 16/2016., 2016.

Kaufman, L. and Rousseeuw, P. J.: Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York., 1990.

Mangiameli, P., Chen, S. K. and West, D.: A comparison of SOM neural network and hierarchical clustering methods, *Eur. J. Oper. Res.*, 93(2), 402–417 [online] Available from: %3CGo, 1996.

Ward, J.: Hierarchical Grouping to Optimize an Objective Function, *J. Am. Stat. Assoc.*, 58, 236–244 [online] Available from: <http://www.jstor.org/stable/2282967>, 1963.