Environmental
Research Group

Kernel density approach for PM_{10} exceedance characterisation

King's College London Environmental Research Group, University of Leeds and AEA

Prepared for Defra
version 2nd August 2010

David Carslaw, Karl Ropkins and Andrew Kent

Contents

1 Introduction

This report outlines the development of a method to better understand and characterise daily mean PM_{10} exceedances at air pollution monitoring sites across the UK. This work forms part of a wider work programme between Defra, AEA, King's College London and the University of Leeds. In particular, this work builds on a set of freely available open-source air quality data analysis tools known as **openair** (http://www.openair-project.org, Carslaw and Ropkins (2010)).

This work is concerned with developing a better understanding of the causes and characteristics of exceedances of the daily mean PM_{10} limit value of 50 $\mu g\ m^{-3}$, which must not be exceeded more than 35 days per year.

While is easy to identify which days exceedances occur on, it is much more difficult to understand the conditions under which they occur. For example, compared with days where the limit value is met, are there different conditions that lead to exceedances? Is it possible to better identify the source(s) of exceedances and how do they vary by day of the week or with other variables of interest? By looking at the problem rather more widely, the hope is that better information can be gained into the factors that lead to high concentrations of PM_{10}. This document focuses on the development of the method, with a few brief examples at different monitoring sites.

The technique developed is general enough to consider threshold concentrations of other pollutants e.g. daily maximum 8-hour mean O_3 concentrations or hourly concentrations of NO_2 >200 $\mu g\ m^{-3}$. However, the focus in this document is exceedances of the daily mean PM_{10} limit value.

This work is also a response to a Defra and local government need for policy applicable tools, which can be used in a wide range of situations e.g. understanding exposure reduction. Furthermore, this work is not only useful for identifying sources but prioritising them to find the sources that matter with respect to limit value exceedances. As such they may be useful to better understand the impact of abatement measures for effectiveness and cost efficiency.

The main purpose of this document is to introduce the technique and briefly demonstrate its potential uses. Part of the aim is to develop a function that can be used in **openair** and can be made available to the air quality community.

1.1 Outline of approach

The basic approach is concerned with better understanding the conditions under which exceedances of PM_{10} occur. It has been shown that considering the wind speed and direction dependence of pollutant concentrations can yield lots of information concerning dominant source types (Carslaw et al., 2006). In particular, the development of 'polar plots' has proved to be particularly useful in understanding source characteristics. However, the approaches so far have considered all hours in a year and not specifically those hours most associated with exceedances – which are often of most interest.

Figure 1 shows a polar plot for PM_{10} concentrations at the Santon site to the north-east of the main Scunthorpe steel works and other industrial activities. The plot shows two distinct high PM_{10} concentration regions to the west. However, this analysis does not provide

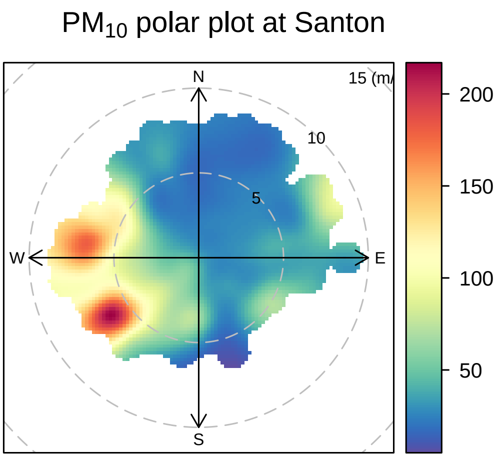

Figure 1: Polar plot showing concentrations of PM_{10} at Santon, as a function of wind speed and direction.

any information as to whether these two distinct conditions are also responsible for PM_{10} daily exceedances. It may be for example, that only a very few conditions make up these features and that overall they do not make an important contribution to daily exceedances. By extension, it may also mean that controlling the sources shown in Figure 1 may not help in meeting daily mean exceedances – even if the concentrations are high for individual hours.[1]

The approach used here therefore focuses on exceedance days.[2] The days above the threshold concentration of 50 $\mu g\ m^{-3}$ are first extracted from the data set in question. Each of these days is associated with meteorological conditions and concentrations of other pollutants, which are of interest because it is these conditions that are most strongly associated with exceedances. However, not every hour of every day will necessarily be strongly associated with exceedances, due for example to variations in wind direction or speed. Nevertheless, this subset of conditions will consist of hours that do contribute to exceedances. By accumulating all of these hours, patterns begin to emerge because certain conditions tend to keep 'cropping up'. The frequency of occurrence of particular conditions therefore provides an indication of the strength that a particular set of conditions are important for exceedances.

What is required is a way of representing the frequency with which these important hours occur. One approach would be to 'bin' the data into different wind speed-direction intervals and count the frequency of occurrence. The problem with this approach is that the bin size tends to be completely arbitrary: a bin size that is too coarse will tend to result in a loss of information and conversely, a bin size too small will result in too much noise. A better approach is to apply what is known as a *kernel density*.

[1] It should always be remembered however there may be other locations where monitoring is not undertaken where these sources could be important to daily mean exceedances.
[2] Throughout TEOM PM_{10} data have been used; multiplied by a factor of 1.3.

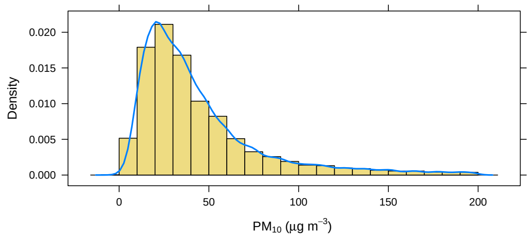

Figure 2: Illustration of a kernel density estimate for PM_{10} concentrations at Santon, Scunthorpe.

To illustrate what is meant by a kernel density plot, Figure 2 shows a 'traditional' histogram of PM_{10} concentrations at Santon where the data are binned into regular intervals. Overlaid on the plot is a kernel density estimate of the same data, shown by the smooth line. It is much easier to see, for example, from the kernel density estimate that the most frequent PM_{10} concentrations are around 20 µg m^{-3}. By contrast the Histogram masks this peak due to the partitioning of the data into different bins. Kernel density estimates can also be made in two dimensions, which is the approach used in here. There are issues related to the 'bandwidth' used for kernel density estimates, which is analogous to the bin size for histograms. However, we find that for air pollution data of the type of interest here, that the choice of bandwidth does not strongly affect the results. In fact, automatic methods for bandwidth selection work well. The latter point is important because it means it is easier to implement these methods and for the non-specialist end-user to use them.

2 Examples of use

In this section we do not aim to give a comprehensive analysis of the orginins of PM_{10} exceedances, but more to highlight the different ways in which the method can be used.

2.1 Application to Scunthorpe

The approach is first applied to sites close to Scunthorpe steelworks that are known to exceed the daily mean PM_{10} limit value. Consideration is given to Santon and Scunthorpe Town sites for approximately three years of data.

Figure 3 shows an approach based on 'binning' the data into wind speed/direction intervals at the Santon site (to the north-east of the steelworks site). In this particular plot all the days where PM_{10} was >50 µg m^{-3} were extracted together with all the hours of meteorology. It is apparent from Figure 3 that the conditions where exceedances mostly occur are for winds from the south-west and wind speeds between 1–4 m s^{-1}.

While Figure 3 is useful, it lacks some important details that are better revealed with a kernel density approach. Figure 4 shows the equivalent plot to Figure 3 but using kernel density estimates to show the surface. The plot shows the 'density' of points associated with

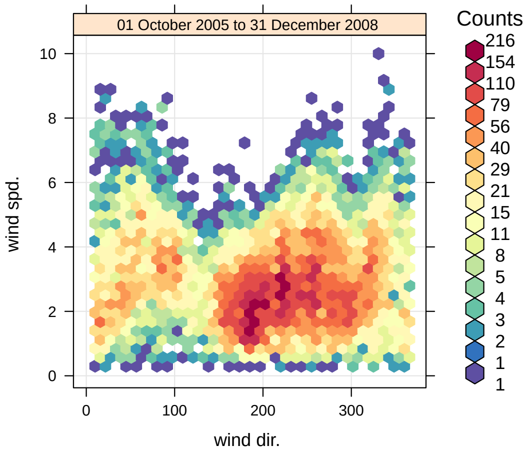

Figure 3: Frequency of wind speed and wind direction occurrences when daily mean PM_{10} concentrations exceed 50 μg m^{-3} at Santon, Scunthorpe. This plot is analogous to a 2D histogram.

exceedances: a discrete dark red colour for example showing a very distinct area associated with exceedances. Note, however, that the colours are not scaled to show exceedance levels and the reader should note the *estimated* number of exceedances shown by the number in the top-left of the panel (331 in this case). Further, if the colours are fairly evenly spread, it suggests that exceedances do not tend to be strongly associated with specific conditions of wind speed/direction. Already, there is rather more detail shown that highlights the conditions under which PM_{10} is exceeded e.g. regions around 180 degrees and wind speeds of about 1 m s^{-1} and another main region for south-westerly winds with higher wind speeds of ~3 m s^{-1}.

Compared with Figure 1, Figure 4 shows that the conditions that account mostly for exceedances of PM_{10} are not actually from the two locations to the west shown on Figure 1. In fact, most of the exceedances of the daily mean PM_{10} standard are from the south west and south and for lower wind speeds. This difference shows the benefit of specifically considering the exceedance conditions. While high wind speed conditions from the west (> 5 m s^{-1}) may result in high concentrations of PM_{10} (> 150 μg m^{-3}), these conditions do not appear to be important for exceedances of the limit value.

While the information shown in Figure 4 is useful it can be enhanced significantly by considering how these conditions depend on the value of other variables. These other approaches are aimed at answering questions such as: do the exceedances depend on season, day of the week, the magnitude of concentration of another pollutant? By considering the data in this way it is hoped that new information concerning the exceedances can be gained that will help better understand their origin.

First, consider the seasonal dependence of PM_{10} exceedances at Santon shown in Figure 5. There are several points to be noted here:

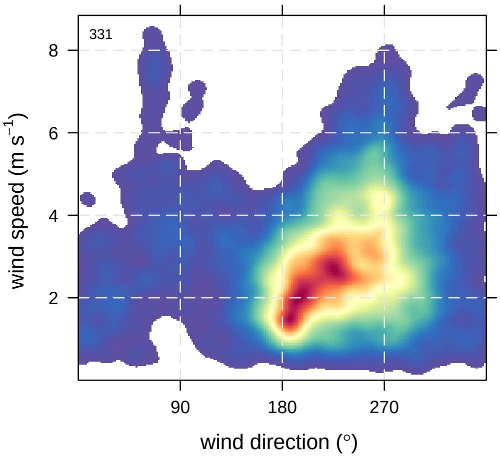

Figure 4: Kernel density plot of wind speed and wind direction occurrences when daily mean PM_{10} concentrations exceed 50 µg m^{-3} at Santon, Scunthorpe. The '331' shown on the plot indicates a total of 331 days when PM_{10} was > 50 µg m^{-3}.

1. Most exceedances are associated with autumn (134 days) and the least in winter (28 days).
2. In the spring and summer there is some evidence to suggest that the exceedances are mostly associated with winds from the SW and wind speeds of ~3 m s^{-1}.
3. In autumn most of the daily PM_{10} exceedances are associated with winds from the south and wind speeds of ~2 m s^{-1}.

The points above do not prove that the character of PM_{10} exceedances is different by season for an underlying reason other than by chance – but they do point to some evidence of maybe different processes or emissions by time of the year.

Another insightful way of considering the dependence of PM_{10} exceedances on other variables is to plot the kernel densities by different levels of another pollutant. In this case, we consider a pollutant that acts as an excellent tracer of combustion processes — SO_2. However, SO_2 is not a tracer for all combustion processes because, for example, it is no longer a good tracer for transport sources because fuel sulphur contents of petrol and diesel are very low. We split the SO_2 concentrations into four *quantiles* i.e. roughly equal numbers of points for four levels of concentration.[3]

Figure 6 shows the kernel density estimates as a function of the concentration of SO_2. Some points to note are:

1. Most exceedances of PM_{10} are associated with high and low concentrations of SO_2. There are about 130 exceedance days for low SO_2 concentrations (between 0–2.7

[3]Depending on the data, it is not always possible to have four levels, as in the case of Figure 6.

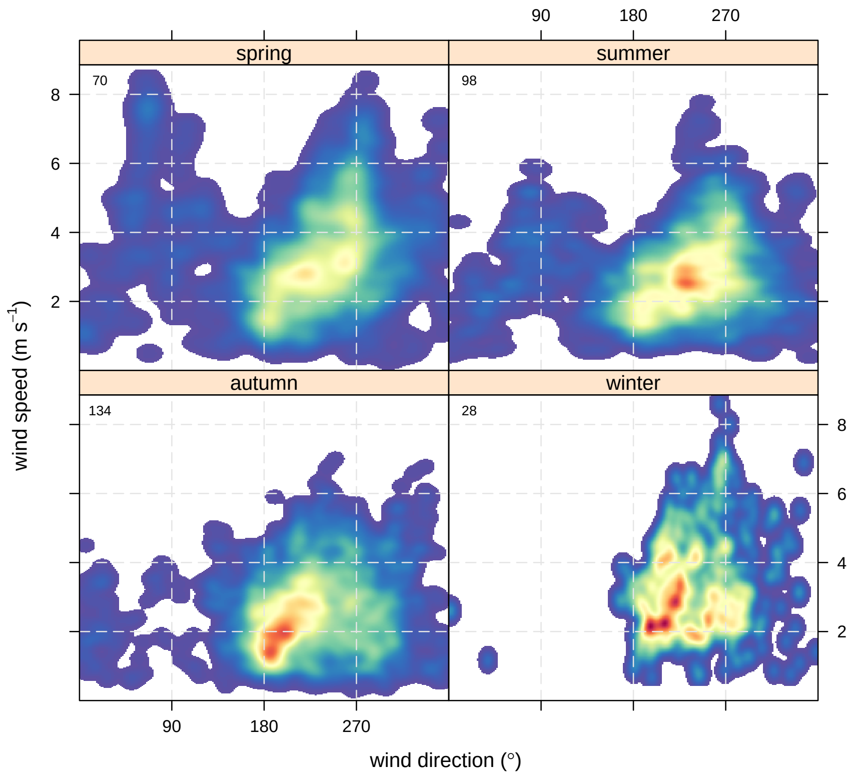

Figure 5: Kernel density plot of wind speed and wind direction occurrences when daily mean PM_{10} concentrations exceed 50 µg m^{-3} at Santon, Scunthorpe. Each panel shows the origin of exceedances by season, with the *approximate* number of exceedance days associated with each season.

µg m^{-3}) and 126 days for higher concentrations of SO_2 (8–114 µg m^{-3}).

2. There is a very different set of conditions that lead to PM_{10} exceedances for high and low concentrations of SO_2. For low concentrations of SO_2, the exceedances are strongly clustered around southerly winds of ∼1 m s^{-1}. Conversely, at high concentrations of SO_2, the exceedances are dominated by south-westerly winds of around 2–3 m s^{-1}.

It is worth noting that if there was *absolutely no relationship* between exceedances and other variables, the kernel density plot would look like Figure 4.

In the case of SO_2 (Figure 6) there is a clear difference in the distribution of exceedances by level of SO_2 concentration, which does highlight that the types of exceedances do seem to differ depending on the concentration of SO_2. *It should be stressed, however, that just because there is a difference in the conditions contributing to exceedances by level of SO_2 concentration, it does not mean the relationship is causal.* Indeed, the same could be said for any relationship between variables: just because it is good, it does not signify cause-effect. It is always

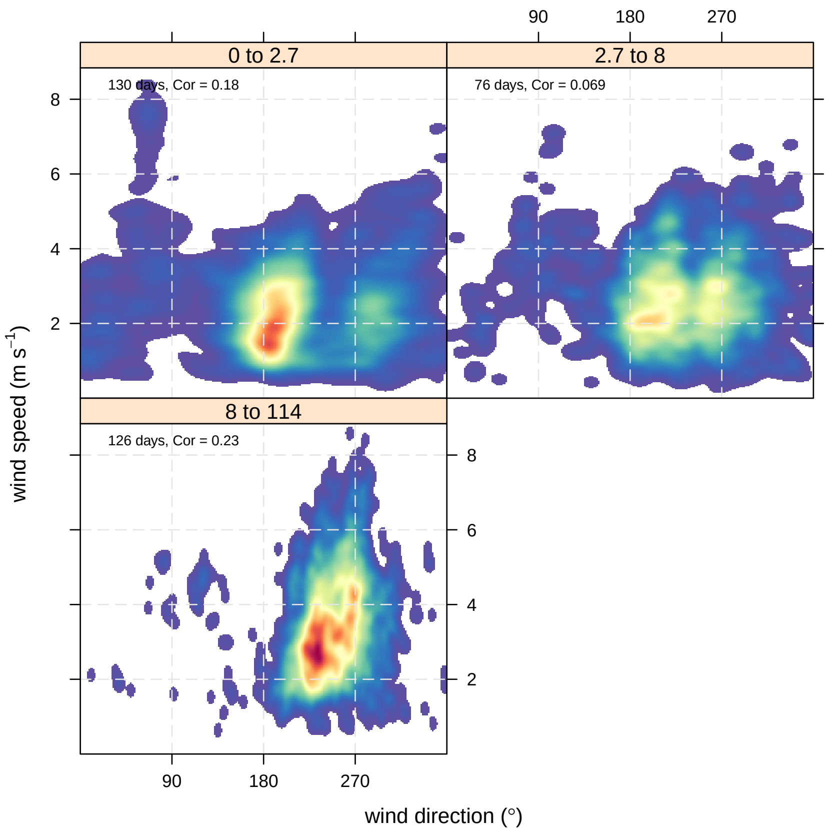

Figure 6: Kernel density plot of wind speed and wind direction occurrences when daily mean PM_{10} concentrations exceed 50 µg m^{-3} at Santon, Scunthorpe. Each panel shows the origin of exceedances by different levels of the concentration of SO_2, with the *approximate* number of exceedance days associated with each season. For example, the first panel considers SO_2 concentrations from 0–2.7 µg m^{-3}.

important therefore to understand the processes involved when making such judgements.

However, by understanding how concentrations depend on other variables does begin to build a picture of the factors that may be important in governing exceedance levels. Also shown in Figure 6 is the correlation coefficient between hourly PM_{10} and SO_2 concentration, which does suggest a slightly higher correlation between the two for high concentrations of SO_2. In this particular case there does seem to be two different source types leading to exceedances. The high number of exceedances with high SO_2 concentrations may not mean that the PM_{10} is combustion-related, but more that the source is in a similar location. Clearly, a knowledge of the site and the processes would help understand the origins of these exceedances more.

2.2 Application to Marylebone Road

PM_{10} concentrations at Marylebone Road are dominated by road vehicle sources. It is therefore useful to apply the technique to this location because it differs greatly from locations that are dominated by industrial sources such as Scunthorpe and Port Talbot. In this section we look a little more closely at the conditions that lead to $PM_{10} > 50\ \mu g\ m^{-3}$ as hourly values and conditions where daily means are greater than $50\ \mu g\ m^{-3}$.

Figure 7 shows the kernel density plots for Marylebone Road PM_{10} concentrations. The plots show the conditions where PM_{10} is $> 50\ \mu g\ m^{-3}$ for all hourly concentrations (left panel) and daily means $> 50\ \mu g\ m^{-3}$ (middle panel) and days where PM_{10} is $> 50\ \mu g\ m^{-3}$ which additionally filters for hours on those days where PM_{10} is $> 50\ \mu g\ m^{-3}$. There is clearly a difference between the three plots, with the most frequent occurrence of hourly values $> 50\ \mu g\ m^{-3}$ being dominated by winds from 180–270 degrees and wind speeds from about 1–7 m s^{-1}. By contrast the conditions where the daily mean PM_{10} concentrations is $> 50\ \mu g\ m^{-3}$ are dominated by lower wind speeds and (typically $\approx$1 m s^{-1}). These results suggest that while quite high wind speed conditions can result in hourly concentrations of PM_{10} being $> 50\ \mu g\ m^{-3}$, these conditions are not as important for daily exceedances.

Because not all hours are equally important to exceedances when the daily mean PM_{10} is $> 50\ \mu g\ m^{-3}$, it might be suspected that there could be some bias in the results. For example, some days when PM_{10} is $> 50\ \mu g\ m^{-3}$ will include some hours e.g. at nighttime when wind speeds are low and PM_{10} concentrations may also be low. However, in the case of Marylebone Road, filtering also for hours $> 50\ \mu g\ m^{-3}$ (right hand panel in Figure 7) shows a plot very similar to the middle panel that just consider all hours on days that exceed.

The fact that low wind speed conditions are more important for daily PM_{10} exceedances suggests that it is stable atmospheric conditions that are most important. It seems therefore that exceedances of PM_{10} at Marylebone Road comprise of two types of condition: street canyon recirculation at high wind speeds and low wind speed conditions where the road and background sources are important.

To further investigate the factors that lead to PM_{10} exceedances at Marylebone Road, Figure 8 shows how the daily PM_{10} exceedances depend on the level of NO_X (an excellent tracer for combustion sources). There are several points to note here. First, as the concentration of NO_X increases, so to does the number of exceedance days e.g. from 46 days at “low” NO_X to 160 days at “high” NO_X. There is also a very different pattern of conditions that lead to exceedances. For lower NO_X concentrations the exceedances are dominated by north-easterly winds. Conversely, at higher concentrations of NO_X it is clear that the road itself is important by the strong signal for low wind speeds and wind directions from 180–270 degrees.

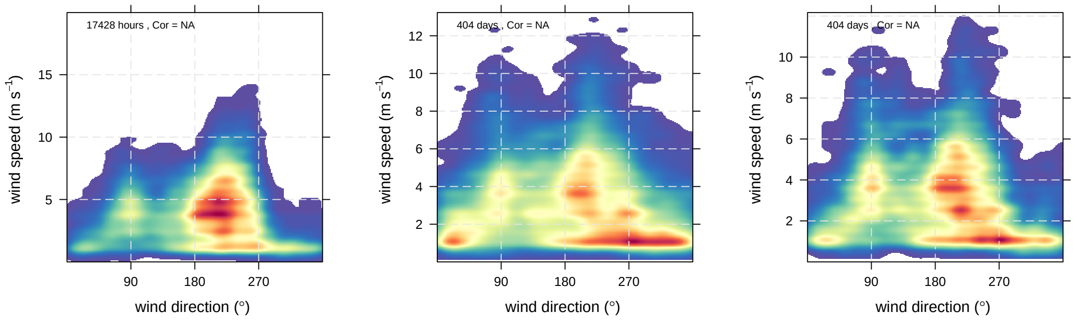

Figure 7: Kernel density plot of wind speed and wind direction occurrences when daily mean PM_{10} concentrations exceed 50 μg m^{-3} at Marylebone Road (1997–2009). The left panel shows the kernel density plot considering all hours where PM_{10} > 50 μg m^{-3}. The middle plot shows the conditions under which daily mean PM_{10} concentrations exceed 50 μg m^{-3}, and the right panel extracts the hours > 50 μg m^{-3} *on days when the mean is > 50 μg* m^{-3} .

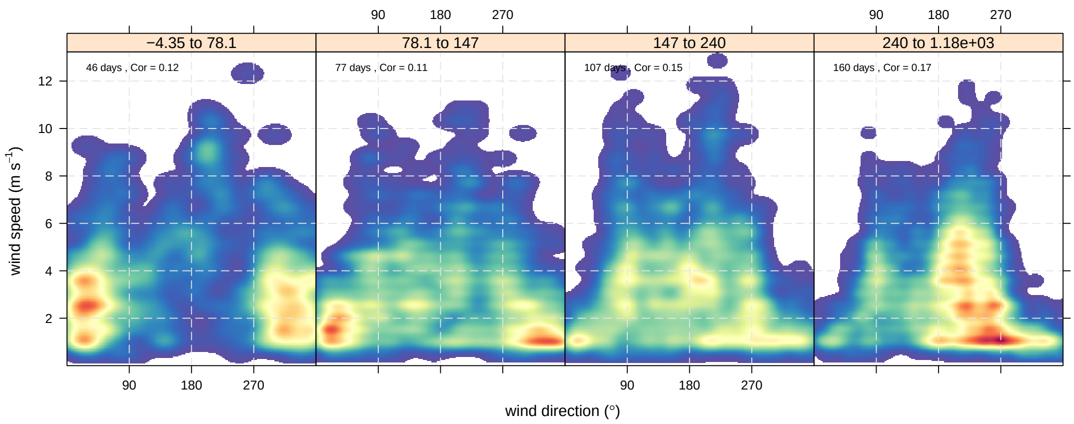

Figure 8: Kernel density plot of wind speed and wind direction occurrences when daily mean PM_{10} concentrations exceed 50 μg m^{-3} at Marylebone Road (1997–2009). The different panels show different levels of NO_X concentration; increasing from left to right.

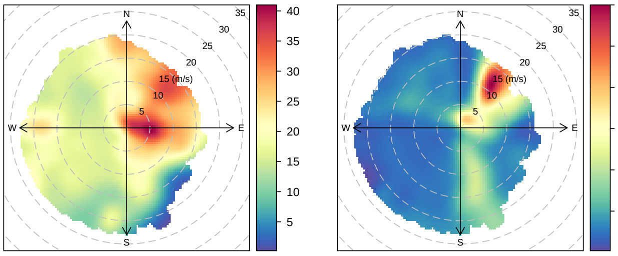

Figure 9: Left panel: polar plot of PM_{10} concentrations at Middlesbrough AURN site ($\mu g\ m^{-3}$). Right panel: polar plot of SO_2 concentrations at Middlesbrough AURN site ($\mu g\ m^{-3}$). Both plots show the presence of a source to the north-east.

2.3 Application to Middlesbrough

This section considers the application of the technique to the Middlesbrough AURN site. The Middlesbrough site is classified as industrial. The docks are about 1.5 km north of the monitoring station and the closest refinery is about 4 km to the north east. The ICI Billingham factory is approximately 5 km to the north west. We have considered data from 1999–2008 using the meteorological data from Loftus – approximately 21 km east of Middlesbrough.

Figure 9 shows the bi-variate polar plot of PM_{10} and SO_2 concentrations at Middlesbrough. This plot indicates sources of PM_{10} to the east and to the north-east. In the latter case the high concentrations of PM_{10} at high wind speeds may suggest the influence of stack emissions from industrial activities. It is also noted that there is strong evidence of stack emissions of SO_2 to the north-east. The most likely origin of these sources is the refinery about 4 km to the north-east of the site given. Of interest is knowing whether these sources are important for exceedances of the daily mean limit. To explore this question, the kernel density approach is applied to the data.

The application of the basic technique is shown in Figure 10. This Figure shows that PM_{10} exceedances are dominated by winds from the ESE (about 100 degrees) for a reasonably wide range of wind speeds from about 2–7 m s^{-1}. The location (in terms of wind speed/direction) of the feature shown in Figure 10 is different from the polar plot shown in Figure 9. Sources to the north-east do not seem to be important for exceedances of the daily limit. Further, the kernel density plot suggests sources to the ESE are more important than easterly sources, although the two directions are similar to one another.

Considering the daily exceedances by year shows that 2003 was clearly the most important year having 33 days >50 $\mu g\ m^{-3}$, whereas a 'typical' year would have around 10 days >50 $\mu g\ m^{-3}$. Given that 2003 was a year dominated by the impacts of secondary aerosol and long-range transport, it appears that this site is also dominated by such events. Furthermore, the kernel density estimates suggest that sources to the north-east do not make a significant contribution to daily exceedances, but can sometimes result in high PM_{10} concentrations where the hourly value is >50 $\mu g\ m^{-3}$.

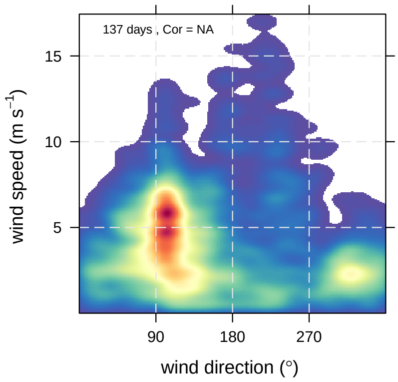

Figure 10: Kernel density plot of wind speed and wind direction occurrences when daily mean PM_{10} concentrations exceed 50 µg m^{-3} at Middlesbrough (1999–2008).

3 Conclusions

This work has developed and applied a new technique for the characterisation of PM_{10} exceedances from monitoring data. The main purpose of the approach is to better characterise those conditions under which exceedances occur – chiefly wind speed and wind direction. Together, wind speed and wind direction provide a lot of useful information on likely source types. This is because different source types can lead to very different wind speed dependencies due to their specific characteristics. For example, emissions from tall stacks often have their maximum impact at *high* wind speeds, whereas many other sources tend to have reduced impacts at higher wind speeds.

The initial application of the kernel density approach has revealed some interesting insights into the factors that are important for daily exceedances. At several sites it has been noted that while certain wind conditions can result in high PM_{10} concentrations (>50 µg m^{-3}) it does not necessarily follow that daily mean exceedances follow similar patterns. In Middlesbrough for example, winds from the north-east are shown to result in high PM_{10} (and SO_2) concentrations, but in terms of daily mean exceedances, these sources do not appear to be important. Similar findings were observed for other sites — even Marylebone Road where PM_{10} concentrations are dominated by a single source type.

A novel development in the current work is to consider how exceedances vary by wind speed, wind direction *and* by levels of other variables. By careful selection of the third variable, it is possible to derive more insight from the approach. For example, in the case of the Santon site at Scunthorpe it was shown that daily mean PM_{10} exceedances derive from two source types: high and low SO_2 concentration regimes. While the approach used in isolation cannot determine causal relationship, it does provide additional information on the characteristics of exceedances that would be difficult to obtain by other means.

It should be noted that the approach is general enough not only to apply to PM_{10} or exceedances or the 50 µg m^{-3} limit. It is possible (and often useful), for example, to consider high percentiles values, which could be higher or lower than the 50 $\mu g\ m^{-3}$ threshold to explore how these concentrations depend on other factors. Similarly, the approach could equally be applied to other pollutants such as O_3. Such an approach is relevant to exposure reduction where there is a need to reduce concentrations overall and specifically high concentrations of pollutants.

Work will continue to develop this approach and make it available in **openair**.

References

Carslaw, D. C., Beevers, S. D., Ropkins, K., Bell, M. C., 2006. Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. Atmospheric Environment 40 (28), 5424–5434. 3

Carslaw, D. C., Ropkins, K., 2010. Open-source tools for analysing air pollution monitoring data. King's College London, Environmental Research Group, version 26th May 2010. 3