Conditional bivariate probability function for source identification

Iratxe Uria-Tellaetxea, David C. Carslaw

Department of Chemical and Environmental Engineering, University of the Basque Country, School of Engineering of Bilbao, Alameda Urquijo, 48013 Bilbao, Spain

King’s College London, Environmental Research Group, Franklin Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

A R T I C L E   I N F O

Article history:
Received 1 November 2013
Received in revised form 24 April 2014
Accepted 2 May 2014
Available online

Keywords:
Receptor model
Source identification
Air quality data
Dispersion model
Air pollution
Openair

A B S T R A C T

In this paper a new receptor modelling method is developed to identify and characterise emission sources. The method is an extension of the commonly used conditional probability function (CPF). The CPF approach is extended to the bivariate case to produce a conditional bivariate probability function (CBPF) plot using wind speed as a third variable plotted on the radial axis. The bivariate case provides more information on the type of sources being identified by providing important dispersion characteristic information. By considering intervals of concentration, considerably more source information can be revealed that is absent in the basic CPF or CBPF. We demonstrate the application of the approach by considering an area of high source complexity, where many new sources can be identified and characterised compared with currently used techniques. Dispersion model simulations are undertaken to verify the approach. The technique has been made available through the openair R package.

2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

Software availability

The methods described in this work are available as part of software called openair. The openair software is freely available as an R package. Details on installing R and optional packages including openair can be found at R Core Team (2014) and http://www.r-project.org. R will run on Microsoft Windows, Linux and Apple Mac computers. No special hardware is required to run openair other than a standard desktop computer. Some large data sets or complex analyses may require a 64-bit platform. Ref: R Core Team (2014). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

1. Introduction

1.1. Background

Identifying local and distant emission sources through receptor modelling is an important area in the management of air pollution. Receptor modelling techniques are diverse and have been applied to a very wide range of situations. Among the more important aspects of receptor modelling is the ability to identify and characterise emission sources, which would perhaps be difficult or impossible by other means. While air quality models can be used together with emission inventories to provide such information, in practice this is difficult. It is difficult for many reasons including incomplete information of the sources and the difficulty in modelling boundary layer processes. For this reason the analysis of ambient air quality data remains a central approach used for understanding emission sources.

A commonly used method for identifying sources is the Conditional Probability Function (CPF). The CPF is a simple but effective technique for providing directional information concerning major sources (Ashbaugh et al., 1985; Vedantham et al., 2013). The CPF calculates the probability that in a particular wind sector the concentration of a species is greater than some specified value. The value specified is usually expressed as a high percentile of the species of interest e.g. the 75th or 90th percentile. It is also possible to extract and filter source information data through conditional analysis as described by Malby et al. (2013). As Malby et al. (2013) show filtering air pollution data by wind speed, direction and time of day can help isolate specific source types for further analysis e.g. the calculation of long term trends.

Bae et al. (2011) used a CPF technique to help identify the directionality of sources contributing to observed pollutant concentrations at a rural site in New York State. The species considered included hourly averaged PM$_{2.5}$ mass, Organic Mass (OM)
from Organic Carbon (OC), optical Elemental Carbon (optical EC), SO2, CO, NOx and O3 for the period of December 2004 to December 2008. In addition, Bae et al. (2011) also considered seasonal variations of these species. Bae et al. (2011) enhanced the basic CPF technique by coupling the method with back trajectory calculations to provide more information on mid to long distance sources.

More sophisticated approaches have also been used to identify dominant sources using non-parametric statistical analysis. Henry et al. (2009) developed a non-parametric wind regression approach to identify and quantify the impact of possible source regions of pollutants as defined by wind direction sectors. Using this approach Henry et al. (2009) were able to quantify the source contribution of different emission sources and demonstrate that some large sources such as a steel mill made only minor contributions to concentrations of SO2.

Kim and Hopke (2004) compared the CPF approach with the non-parametric regression approach for fine particle concentrations (PM2.5) in the USA. They found that CPF and non-parametric regression methods both worked well in identifying known local point sources. However, the CPF approach was easier to calculate compared with the non-parametric regression approach. The results from Kim and Hopke (2004) in both cases provided broad, dominant source directions such as the Port of Seattle or in the case of sea salt, the Atlantic Ocean. An advantage of the non-parametric regression approach is that it is also able to provide uncertainties in the source direction for major sources.

Most of the techniques previously described are focused on identifying and may be quantifying dominant sources affecting a receptor. However, many or most receptors are affected by a much larger number of sources — but they can be difficult to identify. These other sources could include major point sources that are too far from the receptor to be detected clearly or local minor sources that are similarly difficult to detect. There are however compelling reasons why it is useful to detect such sources at a receptor. While there may only be a minor contribution at a specific receptor, it may well be the case that at other locations (perhaps where no measurements are made), the contribution could be much larger and should therefore be investigated. It is also useful to know the extent to which sources have an influence, as this can provide a more complete picture of how sites are affected by a wide range of sources. For example, if it can be shown that a major point source can be detected much further from its location than previously thought, such information is helpful for demonstrating this to be the case. There may also be occasions where isolating particular source types is useful e.g. thermal power plants. Furthermore, there may also be opportunities for enhanced model evaluation by being able to evaluate models over a much larger spatial area.

In this paper a new technique is developed that increases the potential to both detect and characterise source contributions at receptor locations. The new method combines a conditional probability function with bivariate polar plots. The former is useful for source detection and the latter for additional source characterisation. The approach is further enhanced by considering the full distribution of concentrations rather than concentrations exceeding a particular threshold. The method is described and then applied to an area of high source complexity that is affected by both near-field and distant sources. Model simulations are performed to show that similar findings can be gained through the analysis of model predictions.

2. Method

2.1. Bivariate CPF methodology

The ordinary CPF (Ashbaugh et al., 1985) estimates the probability that the measured concentration exceeds a set threshold criterion for a given wind sector. CPF is mathematically defined as:

\[
CPF_{\text{u,v}} = \frac{m_{\text{u,v},C|x}}{n_{\text{u,v}}}
\]

(1)

where \(m_{\text{u,v},C|x}\) is the number of samples in the wind sector \(\text{\text{u,v}}\) having concentration \(C\) greater than or equal to a threshold value \(x\), and \(n_{\text{u,v}}\) is the total number of samples from wind sector \(\Delta u\). Thus, CPF indicates the potential for a source region to contribute to high air pollution concentrations. Conventionally, \(x\) represents a high percentile of concentration e.g. the 75th or 90th.

The conditional bivariate probability function (CBPF) couples ordinary CPF with wind speed as a third variable, allocating the observed pollutant concentration to cells defined by ranges of wind direction and wind speed rather than to only wind direction sectors. It can be defined as:

\[
\text{CBPF}_{\text{u,v},\Delta u} = \frac{m_{\text{u,v},\Delta u,C|x}}{n_{\text{u,v},\Delta u}}
\]

(2)

where \(m_{\text{u,v},\Delta u,C|x}\) is the number of samples in the wind sector \(\Delta u\) with wind speed interval \(\Delta v\) having concentration \(C\) greater than a threshold value \(x\), \(n_{\text{u,v},\Delta u}\) is the total number of samples in that wind direction-speed interval. The extension to the bivariate case provides more information on the nature of the sources because different source types can have different wind speed dependencies. The use of a third variable can therefore provide more information on the type of source in question. It should be noted that the third variable plotted on the radial axis does not need to be wind speed. The key issue is that the third variable allows some sort of discrimination between sources types due to the way they disperse. For example, Carslaw and Beevers (2013) show that temperature can be a useful radial variable.

Bivariate polar plots show how a concentration of a species varies jointly with wind speed and wind direction in polar coordinates. The plots have proved to be useful in a range of settings e.g. to characterise airport sources and dispersion characteristics street canyons (Carslaw et al., 2006; Tomlin et al., 2009; Carslaw and Ropkins, 2012). Wind direction together with wind speed can be highly effective at discriminating different emission sources. By using polar coordinates the plots provide a useful graphical technique which can provide directional information on sources as well as the wind speed dependence of concentrations.

Briefly, bivariate polar plots are constructed in the following way. First, wind speed, wind direction and concentration data are partitioned into wind speed-direction bins and the mean concentration calculated for each bin. The wind components \(\text{u} = \pi \sin(2\pi/\theta)\), \(\text{v} = -\pi \cos(2\pi/\theta)\), where \(\pi\) is the mean wind speed and \(\theta\) is the mean wind direction in degrees with 90° as being from the east, and concentration \(C\) provide a surface. The concentration surface produced by \(\text{u}, \text{v}\) and \(C\) is modelled using a Generalized Additive Model (GAM) (Wood, 2006). GAMs are a useful modelling framework with respect to air pollution prediction because often the relationships between variables are non-linear and variable interactions are important, both of which issues can be addressed in a GAM framework. The surface is fitted according to Equation 3:

\[
\sqrt{C_{\text{u,v}}} = D_{\text{u,v}} + s(u_{\text{v}}, v_{\text{u,v}}) + \epsilon
\]

(3)

where \(C_{\text{u,v}}\) is the ith pollutant concentration, \(D_{\text{u,v}}\) is the overall mean of the response, \(s(u_{\text{v}}, v_{\text{u,v}})\) is the isotropic smooth function of ith value of covariate \(u\) and \(v\), and \(\epsilon\) is the ith residual. A penalized regression spline is used to model the surface as described by Wood (2003). Note that \(C_{\text{u,v}}\) is square-root transformed as the transformation generally produces better model diagnostics e.g. normally distributed residuals. Moreover the smooth function used is isotropic because \(u\) and \(v\) are on the same scales. The isotropic smooth avoids the potential difficulty of smoothing two variables on different scales e.g. wind speed and direction, which introduces further complexities. When fitting the GAM, wind speed-direction bins with few data points are down-weighted such that those with 1, 2 and 3 points have weights 0.25, 0.50 and 0.75, respectively, whereas for sample sizes 3+ are given a weighting of one. This approach therefore gives less weighting to wind speed-direction intervals (and therefore conditional probability estimates) that contain very few data points.

The CBPF can be extended further to consider intervals of concentration rather than only values greater than some threshold. In this case the CBPF for concentration intervals is defined as:

\[
\text{CBPF}_{\text{u,v},\Delta u}(i) = \frac{m_{\text{u,v},\Delta u,C|x}}{n_{\text{u,v},\Delta u}}
\]

(4)

where \(m_{\text{u,v},\Delta u,C|x}\) is the number of samples in the wind sector \(\Delta u\) with wind speed interval \(\Delta v\) having a concentration \(C\) greater than a threshold value \(x\).
information that can be used to identify far more sources. In essence, the enhanced CBPF can reveal ‘hidden’ contributions from different sources. Bennett et al. (2013) provide a comprehensive framework for the evaluation of environmental models where they propose a 5-step procedure for evaluating the performance of models. The CBPF approach is a receptor modelling approach and cannot be compared with measured values in the same way as a deterministic modelling approach. However, the framework of Bennett et al. (2013) is useful in ensuring that the approach, data used and performance evaluation metrics follow a consistent approach and the approach itself is fit for purpose.

2.2. Description of study area and data used

Scunthorpe is the administrative centre of the North Lincolnshire Unitary Area, located on the southern side of the Humber estuary, England. It is characterised by rather flat landscapes and a low population density, being a mainly agricultural area with market towns surrounded by many small villages, as shown in Fig. 1. In this study, it has been possible to identify distant sources because the terrain is very flat and its roughness is very low over the study area and therefore wind flows are not affected by issues related to terrain. However, it shelters several facilities under the Integrated Pollution Prevention and Control directive (IPPC 2008/1/EC), as well as significant road, rail and naval traffic sources. Several routine air quality monitoring stations are also located within North Lincolnshire, as part of the Council’s monitoring network.

Surrounding Scunthorpe and up to 35 km, the most relevant IPPC pollutant sources are related to intensive livestock production, waste management, energy, mineral industry, chemical industry, surface treatment, food and beverage sector and production of processing metals. Among these numerous IPPC facilities, those emitting amounts >1000 t yr\(^{-1}\) of any monitored pollutants are listed in Table 1. In terms of NO\(_x\) and SO\(_2\) emissions there are a few dominant major sources. For example, source 2 is Drax power station, 28 km from the Scunthorpe Town monitoring site; one of the largest in Europe and an important source of both NO\(_x\) and SO\(_2\) (despite using flue gas desulphurisation). The integrated steelworks (source 3) is a complex and significant local source of PM\(_{10}\), SO\(_2\) and NO\(_x\). Further to the east there

**Table 1**

<table>
<thead>
<tr>
<th>ID</th>
<th>activity</th>
<th>Distance (km)(^a)</th>
<th>Emission (t yr(^{-1}))</th>
<th>NO(_x)</th>
<th>SO(_2)</th>
<th>PM(_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Coal-fired power station</td>
<td>35</td>
<td>9444</td>
<td>7527</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Coal-fired power station</td>
<td>28</td>
<td>38,422</td>
<td>27,846</td>
<td>362</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Integrated steelworks</td>
<td>&lt;0.5</td>
<td>3073</td>
<td>3442</td>
<td>2349</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Gas-fired power station</td>
<td>31</td>
<td>2012</td>
<td>121</td>
<td>57(^b)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Gas-fired power station</td>
<td>27</td>
<td>1872</td>
<td>2</td>
<td>21(^b)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Gas-fired power station</td>
<td>26</td>
<td>2983</td>
<td>5</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Oil refinery</td>
<td>26</td>
<td>1362</td>
<td>6458</td>
<td>2(^b)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Gas-fired power station</td>
<td>27</td>
<td>1322</td>
<td>3</td>
<td>47(^b)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Oil refinery</td>
<td>27</td>
<td>2549</td>
<td>5004</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Solid fuel production</td>
<td>29</td>
<td>7(^b)</td>
<td>1465(^b)</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Gas-fired power station</td>
<td>33</td>
<td>2520</td>
<td>0</td>
<td>55(^b)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Coal-fired power station</td>
<td>28</td>
<td>18,700</td>
<td>7530</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Coal-fired power station</td>
<td>32</td>
<td>28,300</td>
<td>6040</td>
<td>312</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) From Scunthorpe Town monitoring site.  
\(^b\) Modelled value.
are many relatively large sources of NOx and SO2 from power stations and an oil refinery. These sources are typically 25–30 km from the Scunthorpe Town monitoring site. The extent to which these sources affect the concentrations of the different species in Scunthorpe will of course depend on the emissions source characteristics, as well as the prevailing meteorology. While a source such as Drax is the most important source of SO2 and NOx in terms of absolute emissions, the stack is 259 m tall. In addition, emissions from sources such as Drax will be released at high temperature and therefore plume buoyancy effects will be important in controlling how the plume disperses and where it impacts ground-level receptors. The Scunthorpe Town monitoring site is housed within an enclosed air-conditioned unit located as shown in Fig. 1. Scunthorpe Town is an urban-industrial station located in a flat open field within the urban area itself, less than 500 m away from the west boundaries of the integrated steelworks and approxi- mately 10 m to the north of a minor urban road. To the west of the site lies the bulk of the urban area of Scunthorpe. The nearest busy road is the A1029, which at its closest point is 124 m to the northeast of the station. The monitoring equipment consists of a fluorescent SO2 analyser, a NOx chemiluminescence analyser and a TEOM PM10 monitor. Wind speed and direction are also measured at the Scunthorpe Town site at a height of ca. 5 m. The meteorological data from the Scunthorpe Town site were used throughout this study. In this study, 15-min SO2, P M10, NOx, wind speed and direction data measurements from 1st January 2009 to 31st December 2009 at Scunthorpe Town were analysed.

As shown in the map (Fig. 1), all main industrial sources are located further than 20 km away from both monitoring sites, except the steelworks, which is situated very close to both stations. In addition, Scunthorpe Town will also be affected by domestic heating emissions, which are dominated by NOx emissions from natural gas-fired boilers. For 2009, Scunthorpe Town had a 92.5% data capture for the year. The mean concentration of SO2 was 5.1 μg m−3 and the maximum 15-min value 194 μg m−3. Wind speed and direction measurements had a data capture of 99.2%. The mean wind speed at the site was 2.7 m s−1 and the wind directions predominately from the south-western quadrant.

Integrated steelworks such as that at Scunthorpe consist of all necessary installations for from raw material processing to steel making, rolling and shaping. The main point sources of pollutants in Scunthorpe works are: two coking plants, a power plant, four blast furnaces, a sinter plant and a (basic oxygen and electric arc) steel making and concasting plant. With respect to emissions of SO2 (the focus of the current study) the most significant sources are the sinter plant, the coke ovens, blast furnaces and power plants. Environment Agency (2004). These principal sources are shown on Fig. 1.

The meteorological conditions for 2009 are summarised in the wind rose shown in Fig. 2. This Figure shows the prevalence of winds from the south and south–west.

3. Results and discussion

3.1. CPF and bivariate polar plots

The results in this section first consider concentrations of NOx at the Scunthorpe Town site. Fig. 3a shows a conventional CPF plot for NOx, highlighting sources where the concentration is >75% percentile NOx concentration — equivalent to a NOx concentration of 36.3 μg m−3. In this Figure there is a clear indication of higher probabilities for these concentrations from the south-east i.e. in the direction of many steelworks activities. By contrast, Fig. 3b shows a bivariate polar plot for the same NOx data. In this plot several additional and interesting features can be seen. First, the highest concentrations occur under very low wind speed conditions from all wind directions, but particularly the west; corresponding to the direction where most of the urban area of Scunthorpe lies. These high concentrations occur under stable atmospheric conditions when non-buoyant ground-level sources are important such as road transport emissions and domestic heating. The bivariate polar plot also shows are area of high concentration to the south–east, corresponding to activities on the steelworks.

In the case of SO2 (Fig. 3c and d) both plots indicate the presence of a major source to the east. However, in Fig. 3c it is not clear whether there is one or several sources. By contrast 3d indicates that there could be two major sources of SO2 — one to the east and one to the south–east. Furthermore, the high concentrations present at high wind speeds are indicative of emissions from stacks rather than non-buoyant ground-level sources as in the case of NOx. Since December 2007 UK petrol and diesel have been sulphur-free (10 ppm or less) and consequently road transport is no longer a significant source of SO2. The effects of using sulphur-free fuels can be seen be comparing the NOx and SO2 bivariate polar plots where low wind speeds are clearly associated with high NOx but not high SO2. The interpretation of SO2 concentrations has therefore become more straightforward because road vehicle sources can be ruled out as a contributory factor in these types of analysis.

Fig. 3b and d therefore provides useful additional information beyond that shown in Fig. 3a or Fig. 3c. Of particular use is not only an indication of the direction in which important sources lie, but also their dispersion characteristics. From Fig. 3b and d it is possible to identify at least two source types with different dispersion characteristics. Both the CPF and bivariate polar plots tend to highlight the dominant source types affecting a monitoring site. While such information is useful, it is known from Table 1 and Fig. 1 that there are many major source types in the area, some with high emissions — albeit located up to 25–35 km distant. Fig. 3 reveals very little information about these other sources.

3.2. Selection of intervals

An important characteristic that is revealed when considering CBPF intervals is that sources tend to occupy clear concentration intervals. Indeed, it is this characteristic that is exploited in the method to reveal many more sources. Data are clearly already effectively filtered by considering wind speed and wind direction intervals, as shown by the presence of distinct source features in the basic bivariate polar plot. In many cases it may be that a plume is only mixed down to ground level within a particular wind speed range — at a particular receptor location. However, it is also found that sources are only detectable within often narrow ranges in concentration also. This characteristic of the sources is likely due to most sources having a relatively invariant emissions rate. Therefore, the data filtering comprises of three components: wind speed, wind direction and a concentration interval, which together provide a very fine-controlled method for source identification.

The question arises as to how to identify concentration or percentile intervals that best highlight a particular source. We have found that producing a large series of CBPF plots across a range of percentile intervals works well. For example, a large series of plots can be generated by considering percentile intervals from P0 to P10, P1 to P11, ..., P90 to P100. Considering a series of plots in this way...
provides an effective means of determining where particular sources have their greatest impact. This approach could show for example that a particular source is most apparent between certain percentile intervals e.g. from the 17th to the 31st. Furthermore, producing an animation of these plots further enhances the interpretation potential. In this case, sources can be seen to ‘emerge’ and then ‘disappear’ when traversing the full range of percentiles. The most appropriate sequence of intervals to use will depend on the data in question and it may not always be best to consider sequences such as P0 to P10, P1 to P11, ..., P90 to P100. For example, for some species such as SO2 there may be a large number of zero value concentrations. In this case many percentile levels can be zero. A strict sequence of P0 to P10, P1 to P11, ..., P90 to P100 will therefore result in some redundancy. In this case it can make sense to take the sequence of percentile intervals for values greater than zero. Similarly if any percentile concentrations are repeated it makes sense to use the series of unique percentile values as the basis of selecting intervals.

3.3. Exploration of SO2 sources

As an example of CBPF, concentrations of SO2 at Scunthorpe Town have been considered in more detail. Ranges of percentile intervals were selected as described in Subsection 3.2 and a sequence of CBPF plots produced. From these plots potential sources were identified that covered discrete concentration ranges. These ranges span a wide range of concentrations from 2.7 to 11 to 58–194 μg m⁻³ SO2. The most likely sources detected in each concentration range from low to high are discussed in turn.

The two most likely sources shown in Fig. 4a are major coal-fired power stations. To the north-west the only major SO2 source is Drax power station (see Table 1, ID = 2) but is 28 km from the monitoring site. The directions of the power stations are indicated by the dashed lines. Similarly, the source to the SSW is most likely the West Burton power station, also 28 km from the monitoring site and also a significant emitter of SO2 (ID = 12). Note that Drax has taller chimneys than West Burton (269 versus 200 m), which will influence the way in which the plumes disperse from each power station. These sources also have their maximum influence at very low concentrations of SO2 from 2.7 to 11 μg m⁻³ and are not apparent in Fig. 3. Despite these low concentrations and only 1-year of data these sources are clearly visible when analysed using the CBPF approach.

The ability to detect sources at specific receptor locations will depend on the prevailing meteorology. In the current case, even though the steelworks is located to the east of the monitoring site and
there are fewer wind conditions from the east (as shown in Fig. 2), it is clear that specific sources can be detected. One of the reasons is that these steelworks sources of SO$_2$ are comparatively large and nearby. In other situations such as where the sources are of a more intermittent nature, more data may be required to characterise them e.g. by considering more than one year of monitoring data.

The analysis of data in the Scunthorpe area has benefitted from flat terrain that would help ensure that plumes would tend to disperse in straight lines unaffected by hills and other features. For this reason, even though some of the sources are up to 28 km distant, the dispersion of these plumes will be unaffected by topography. For locations with hilly terrain it would be more challenging to link identified source features to specific source locations.

The sources located on the integrated steelworks are both numerous and complex. For this reason it would not be expected that it would be possible to resolve individual sources in any detail. As discussed previously, the CPF plot for SO$_2$ (Fig. 3c) only shows that concentrations are high from the east and provides no further detail. The bivariate polar plot perhaps reveals two sources (or groups of sources), shown in Fig. 4. Nevertheless, by considering intervals of CBPF more detail can be resolved.

To the ENE there are coke ovens about 2000 m from the Scunthorpe Town site, shown in Fig. 4b and labelled as ‘Coke ovens [1]’. The sources in this direction are dominant between 21 and 29 $\mu$g m$^{-3}$. There are several potentially important SO$_2$ sources to an easterly direction, including a coke oven and a power plant, shown in Fig. 4c. The coke ovens in this direction are closer to the Scunthorpe Town site (about 1200 m), which might account for the higher concentrations seen for these sources compared with the coke ovens to the ENE. It is not possible to resolve these sources in more detail but sources in that direction have their clearest impact when SO$_2$ concentrations are from 37 to 48 $\mu$g m$^{-3}$, as shown in Fig. 4c. There are also more distant sources close to the east coast about 30 km from Scunthorpe Town, including an oil refinery and power stations. These sources will be masked by the integrated steelworks but could in principle be detected in the absence of the steelworks or if the monitoring site was located to the east of the steelworks.

The source identified to the SE shown in Fig. 4d is most likely the sinter plant. There are several features of the results that lead to this conclusion. First, there is very good alignment between the feature shown in Fig. 4d. Second, it can be shown that this feature also clearly appears on plots of NO$_x$ and PM$_{10}$, which is consistent with sinter plan operations where high emissions of all three species can be expected (Environment Agency, 2004). In addition, the high NO concentrations and narrow plume points to a local emission source. These characteristics are consistent with the sinter stack, as it is very close to the Scunthorpe monitoring site (~2 km) and is a single stack source 107 m in height.

![Fig. 4. CBPF plots for different concentration intervals of SO$_2$ concentration at the Scunthorpe Town site for 2009. (a) Plot highlighting sources of SO$_2$ to the SE and NW from 2.7 to 11 $\mu$g m$^{-3}$, (b) SO$_2$ sources in the range 21–29 $\mu$g m$^{-3}$ (c) sources to the east in the range 37–48 $\mu$g m$^{-3}$, (d) source highlighted to the SE in the range 58–194 $\mu$g m$^{-3}$. The dashed lines show the direction of potential known sources (see text for details).](image-url)
Evidence of the co-emission of SO$_2$, NO$_x$, and PM$_{10}$ is better seen in Fig. 5. The source to the SE is a major source of SO$_2$, PM$_{10}$ and NO$_x$ and appears when concentrations are between: 30.0–194.2 μg m$^{-3}$ for SO$_2$ (Figure 5a), 55–150 μg m$^{-3}$ for PM$_{10}$ (Fig. 5b), and 48.0–387.7 μg m$^{-3}$ for NO$_x$ (Fig. 5c). For all species the concentration range where they are prominent is high, suggesting that the source is an important emitter of all species. Another aspect of the NO$_x$ plot that can be observed in Fig. 5c is the feature seen at very low wind speeds, which is most apparent for westerly winds. This source corresponds to diffuse emissions of NO$_x$ from the Scunthorpe urban conurbation, which will be dominated by ground-level road traffic and domestic gas boiler emissions. There are several characteristics that lead to this conclusion. First, the wind direction where the source is dominant corresponds to the Scunthorpe Town urban area. Second, high NO$_x$ concentrations under low wind speeds are indicative of surface emissions released with little or no buoyancy. The CBPF probability values in a particular wind speed-direction interval can vary widely as shown in Fig. 4. The actual probabilities will be very data and emission source dependent. For example, a single continuous dominant source in one direction that mostly accounts for concentrations in a particular range can have high probabilities. This is the situation for the two power station sources shown in Fig. 4a where the probabilities are high (up to around 0.8). On the other hand, if there is a mixture of sources that affect a particular direction (such as the steelworks) then there will tend to be some influence of these sources across different concentration intervals, which has the effect of reducing the probability value. Steelworks sources tend to operate in a more discontinuous mode compared with large power stations and this behaviour would tend to result in lower probabilities. This behaviour is also seen in Fig. 4b–d where the probabilities are lower (in the range 0.2–0.35) for the steelworks sources compared with the power station sources. In principle, the CBPF approach can help distinguish between two sources from the same wind direction. The ability of the technique to distinguish between sources in this way depends on their type and characteristics. If two sources have very different wind speed dependencies (e.g. one prominent at low wind speeds, the other at high wind speeds) then these will be shown in different wind speed intervals in the CBPF plot. For example, Fig. 5c (for NO$_x$) shows that at low wind speeds there is evidence of a source to the west (likely Scunthorpe Town road traffic and domestic gas combustion sources) and to a lesser extent to the east where there is less of the urban area of Scunthorpe. However, in Fig. 5c it is also clear that a different source to the south—east can also be detected (thought to be the sinter plant) and which is clearly separate from the urban source seen at lower wind speeds. These results provide considerably more information on the different sources affecting the Scunthorpe Town monitor compared with the basic CPF or bivariate polar plot approach. It has been possible to resolve distant but major point sources which only make contributions at low concentrations of SO$_2$ and which are therefore easily missed. The complexity of the integrated steelworks itself does however limit the extent to which a single measurement site can resolve multiple nearby sources. An obvious extension to this work would be the use of another site to the east of the steelworks that would help triangulate and resolve the sources in more detail. Knowledge of the concentration ranges where different sources have their influence is useful for several purposes. First, the concentration ranges provide the conditions where a source has most...
influence. Therefore it would be possible to subset the original data and extract those conditions; perhaps also with ranges of wind speed and wind direction to analyse the data in other ways knowing that a particular source has most influence. Second, the concentration intervals may be helpful for model evaluation purposes. In this case the ability to quantify the concentration range an emission source has most influence in could allow a comparison with equivalent results from a dispersion model. In the context of the current work, the enhanced detection of distant sources at low concentrations would allow models to be evaluated at distances further than is currently possible.

The analysis in this study has used wind speed for the radial axis of the CBPF plots. As discussed in Subsection 2.1 other variables could be used such as ambient temperature. In addition, it could also be useful to use variables derived from an air quality model or a numerical weather prediction (NWP) model. Many such models can provide estimates of boundary layer stability or other variables representing vertical mixing e.g. as in the UK Met Office Unified Model Cullen (1993). The effectiveness of these other variables for use in CBPF plots depends on the extent to which sources can be discriminated. For example, stack emission sources would be expected to respond very differently to changes in atmospheric stability compared with ground-level sources. To date there is little or no work in using outputs from NWP models as a means of enhancing receptor models. However, as advances in the spatial resolution of NWP models develop, there should be increased opportunities to use model-derived variables to support receptor modelling. In addition to the use of NWP models it would also be useful to consider the use of more detailed meteorological measurements e.g. several meteorological sites over the domain of interest to help understand the source origins of more distant sources more fully.

### 3.4. Simulations using the ADMS model

The earlier discussion focused on the analysis of ambient measurements. It is also useful to consider whether similar findings would be attained using atmospheric dispersion models. The model chosen for the simulations was the Atmospheric Dispersion Modelling System (ADMS) developed by Cambridge Environmental Research Consultants (McHugh et al., 1997). The ADMS model is an advanced Gaussian model based on boundary layer properties characterised by the boundary layer depth and the Monin-Obukhov length. For these simulations version 4.1 of ADMS was used.

Three different stack sources were modelled based on the assumptions shown in Table 2. The first source (P1) represents the emissions typical of those expected from a large power station such as Drax with a similar annual emission of SO$_2$ of 750 g s$^{-1}$ (23.7 kt yr$^{-1}$) and a stack height of 260 m ca. 28 km from the receptor. The other two sources (P2 and P3) represent smaller stacks with lower emission rates located much closer to the receptor 2.1 and 0.3 km, respectively). Hourly meteorological data were used from the UK Met Office London Heathrow site for the purposes of the simulations. The simulations were carried out assuming flat terrain i.e. similar to the situation in North Lincolnshire.

#### Table 2

Source details used for ADMS modelling. The receptor is assumed to be located at (0, 0).

<table>
<thead>
<tr>
<th>Source</th>
<th>Location</th>
<th>Source details</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x (m)</td>
<td>y (m)</td>
</tr>
<tr>
<td>P1</td>
<td>-20000</td>
<td>-20000</td>
</tr>
<tr>
<td>P2</td>
<td>1500</td>
<td>1500</td>
</tr>
<tr>
<td>P3</td>
<td>300</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 6. Bivariate polar plot of three point sources modelled using the ADMS model based on the source assumptions shown in Table 2. The model used hourly meteorological data from the London Heathrow site for 2005. The units are μg m$^{-3}$.

Fig. 7. CBPF plot of SO$_2$ concentrations extracted from the ADMS modelling simulation of three different point sources described in Table 2. The data were analysed as described in Subsection 3.2. (a) The lowest emitting source P2, shown for the concentration interval 0.11–0.26 μg m$^{-3}$, (b) source P3 shown for the concentration interval 0.35–1.6 μg m$^{-3}$, (c) source P1 shown for the concentration interval 2.2–37 μg m$^{-3}$.  

The bivariate polar plot for the model-simulated concentrations in shown in Fig. 6. There is limited evidence of stack P3 to the east and no evidence of stack P2. The results in Fig. 6 also show that it is reasonable to expect to detect a large power station source such as Drax at a distance of ≈28 km where concentrations of up to 6 µg m⁻³ can be expected. Using the same methods as described in Subsection 3.2 the hourly ADMS model results were analysed for the presence of separate sources across the full concentration range. The result of this analysis is shown in Fig. 7, showing three clearly distinguishable sources that are not apparent when considering only the bivariate polar plot (Fig. 6). These results support the validity of the method as a way in which to extract information on sources that have a relatively minor impact on concentrations at a receptor. The most obvious feature of Fig. 6 is the P1 stack to the south-west.

4. Conclusions

This study has combined two approaches commonly used for source identification in receptor modelling. The Conditional Probability Function provides a simple but effective way in identifying major source directions and the bivariate polar plot provides additional information on how sources disperse. The latter can therefore help to discriminate different source types through their wind speed dependence. Combining these two techniques provides a new method termed the Conditional Bivariate Probability Function (CBPF).

The CBPF has been developed further to consider intervals in the concentration of a species. It is found that sources tend to occupy clear concentration intervals, which allows many more sources to be identified than was previously possible. For example, a distant but large point source might not be identified using a basic CPF or bivariate plot but can be identified in a lower discrete concentration range of the CBPF. The approach can therefore provide a more comprehensive understanding of a very wide range of sources affecting a particular monitoring site. In the application used in this study, distant point sources of SO₂ could be identified which were not apparent previously; and nearby sources from an integrated steelworks could be better disaggregated. The identification of an increased number of sources was also shown using an advanced Gaussian plume model.

Acknowledgements

We are very grateful to Ayan Chakravarty from North Lincolnshire Council for providing the air quality and meteorological data used in this study. Iratxe Uria-Tellaetxe a would like to thank the University of the Basque Country for funding her PhD.

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. Atmos. Environ. 40 (28), 5424–5434.
McIntyre, B.R., Whyatt, J.D., Timmis, R.J., 2013. Conditional extraction of air-pollutant source signals from air-quality monitoring. Atmos. Environ. 74 (0), 112–122.