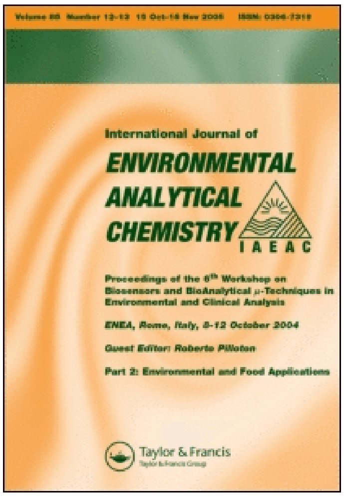

International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

Using principal component analysis to detect outliers in ambient air monitoring studies

Richard J. C. Brown[a]; Sharon L. Goddard[a]; Andrew S. Brown[a]
[a] Analytical Science Team, National Physical Laboratory, Teddington, Middlesex, TW11 0LW, United Kingdom

Online publication date: 05 July 2010

To cite this Article Brown, Richard J. C. , Goddard, Sharon L. and Brown, Andrew S.(2010) 'Using principal component analysis to detect outliers in ambient air monitoring studies', International Journal of Environmental Analytical Chemistry, 90: 10, 761 — 772

To link to this Article: DOI: 10.1080/03067310903094545
URL: http://dx.doi.org/10.1080/03067310903094545

Intern. J. Environ. Anal. Chem.
Vol. 90, No. 10, 20 August 2010, 761–772

Using principal component analysis to detect outliers in ambient air monitoring studies

Richard J.C. Brown*, Sharon L. Goddard and Andrew S. Brown

Analytical Science Team, National Physical Laboratory, Teddington, Middlesex, TW11 0LW, United Kingdom

(*Received 24 March 2009; final version received 28 May 2009*)

The need to determine outliers in analytical data sets is important to ensure data quality. More sophisticated techniques are required when the checking of individual results is not possible, for instance with very large data sets. This paper outlines a novel method for the detection of possible outliers in multivariate sets of air quality monitoring data, here the metals content of ambient particulate matter. Principal component analysis has been used to take advantage of the expected correlation between metals concentrations at individual monitoring sites to produce a summary statistic based on the deviation of each observation from the expected pattern, which can then be interrogated using one-dimensional robust statistical techniques to identify possible outliers. The sensitivity of this statistic to the number of principal components included in the summary statistic has been examined, and the method has been demonstrated on exemplar data from the UK Heavy Metals Monitoring Network where it has produced very accurate predictions of outlying data.

Keywords: principal component analysis; analytical chemistry; data screening; ambient air; air quality networks

1. Introduction

The need to ensure the robustness of very large data sets produced by analytical measurement processes is ever present. This requires data screening techniques that can identify outliers in large analytical data sets that have undergone multiple data-handling and manipulation procedures as well as being affected by errors introduced by sample collection, preparation and analysis procedures. An example of the requirement for the screening of complicated analytical data sets is an air quality network measuring multiple pollutants at a number of monitoring sites. Outliers in such measured mass concentration data could be caused by a number of factors, including sample contamination, incorrect transcription or manipulation of sampling or analytical data, or an analytical error. Outliers may also be introduced by monitoring at local ‘hot-spots’ or by medium or long-range transport events. It is important to remove such outliers as they may bias summary air quality statistics such as time-averaged concentration values. Once a possible outlier is identified, it is important to investigate fully the suspect observation in order to determine

*Corresponding author. Email: richard.brown@npl.co.uk

ISSN 0306–7319 print/ISSN 1029–0397 online

DOI: 10.1080/03067310903094545
http://www.informaworld.com

whether there is an identifiable cause for the unexpected value. If such a reason is found then these values may be removed from the data set. Outlying values occurring from 'hot spot' events or long range transport episodes should be retained for separate analysis of the event in question. The main difficulty with spotting outliers in the concentration values of pollutants in ambient air is that one expects the true concentrations of these pollutants at a specific monitoring location to vary in a very complex manner which will depend on changes in emission rates from local and remote, diffuse and point sources, and variations in meteorological conditions.

It is very difficult to judge whether any given value in a series of individual measurements of a single pollutant constitutes an outlier, if a detailed study of emissions and meteorological parameters is not carried out. Fortunately several ambient pollutants are usually measured at once at air quality monitoring stations–for instance, this paper will discuss the concentration of various metals in ambient particulate matter, 12 of which are measured at each location on the UK Heavy Metals Monitoring Network [1]. The expectation is that the concentrations of many of these metals will show some correlation with each other because they originate from the same sources. Deviation from this expected correlation, rather than changes in the absolute measured values is therefore proposed here as a method to identify possible outliers requiring further examination.

The analysis of correlation in multivariate data is the basis of principal component analysis (PCA). However, other simpler techniques have been previously used to detect outliers in ambient pollutant data. It has been shown previously [2] that Benford's law is useful for screening measurements of one pollutant at many measurements sites or of many pollutants at one measurements site provided the measured quantities span a large range. However, Benford's law cannot effectively screen data sets which are small or do not span large ranges. Moreover, Benford's law does not specifically use the correlations between the multiple components measured at one site, which can provide added value to the screening process. Zipf's law [3] has also been proposed as a data screening technique for analytical data. Zipf's law is able to take account of correlations between pollutants at one site, and is also effective for small data sets, which show a small range of values, and for mishandled data that cannot be effectively detected using Benford's law. However, Zipf's law requires prior 'calibration' with an unbiased data set before it can be used effectively in this regard. For data sets with a large number of outliers, or whose characteristics are not well known, this is not possible. The use of PCA as an outlier detection tool requires no prior 'calibration' with a known data set. PCA has previously been used directly as an outlier detection technique using complex statistical and mathematical methods [4–6], and in air quality studies to optimise the design of air quality networks [7], and for source apportionment studies [8]. However, to the best of the authors' knowledge no significant studies have used PCA to identify outliers in data from air quality monitoring networks. Moreover, the motivation in this work is to provide a simple and user-tuneable technique using PCA to identify potential outliers in air quality studies, particularly for large data sets produced by air quality networks, where over-complicated and time consuming mathematical techniques are not appropriate.

Therefore, in this paper PCA is used in a two part process with simple robust statistics to identify possible outliers in sets of ambient pollutant data from an air quality network. It is shown that often several principal components (PCs) must be considered in order to ensure that possible outliers are properly captured.

2. Experimental

Individual UK air quality monitoring networks are operated on behalf of the Department of the Environment, Food and Rural Affairs (Defra) by scientific contractors who are responsible for reporting measured pollutant concentrations to Defra and the public Air Quality Archive [9]. It is advantageous, both to Defra, and to contractors of the monitoring networks, to have a simple and robust procedure to screen data sets. In 2006 the UK Heavy Metals Monitoring Network consisted of 17 sites situated around the UK at roadside, industrial, rural and urban background locations [1] (this has since risen to 24 sites following a reorganisation and expansion in 2007–2008 in order to ensure compliance with the EC Fourth Air Quality Daughter Directive [10]). Particulate samples are taken at all sites using Partisol 2000 instruments, fitted with PM_{10} (defined as: air pollutants consisting of particles with an aerodynamic diameter less than or equal to 10 micrometers) size selective heads, operating at a calibrated flow rate of approximately $1 m^3 h^{-1}$. Samples are collected for a period of one week onto 0.8 μm pore size mixed cellulose ester membrane filters (Pall, GN Metricel). After sampling, the filters are digested in acid and analysed for their As, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Pt, V, Zn and Hg content using a PerkinElmer Elan DRC II ICP-MS and the standard procedure detailed in EN 14902 [11] and previously described [12]. (The analysis of Pt and Hg requires a slightly different digestion procedure.) This produces a maximum of 52 sets of results for 12 metals at each site every year. Usually fewer than 52 sets of data are produced because of instrument servicing, or breakdown. Therefore the data matrices analysed by PCA for each of the 17 monitoring sites consisted of between 52 and 44 observations (depending on location), with each observation comprising measurements of 10 analytes. For the purposes of this data analysis, results for Pt and Hg at all sites have been discarded since these are commonly below the detection limit and would bias the outcomes of this study. 'Concentration' is used to refer to 'mass concentration' throughout this study.

Data analysis was rapid and straightforward, and used widely available PC software such as Microsoft Excel, with the PCA performed using commercially available XLSTAT (Addinsoft) add-in software. Weekly concentration data (expressed as $ng\,m^{-3}$, in ambient air) acquired by the UK Heavy Metals Monitoring Network in 2006 for 10 elements at 17 monitoring sites was used to demonstrate the applicability of detecting outliers. Before performing PCA the data were pre-processed using variance scaling and mean centring–known as auto-scaling when combined. The auto-scaling pre-processing equalised the importance of each metal (as a result of the variance scaling), and ensured that the sum of the scores for each PC is zero (as a result of the mean centring). In the analysis each network site was considered separately. The method described in this manuscript is based on the first order approximation that all measured values at an individual site are expected to be influenced mainly by the local sources and meteorology, so that significant correlation is expected. Furthermore it is assumed for the purposes of this demonstration that there is little or no correlation expected between different sites, since metals are primary pollutants with localised sources and the overlap of sources for any two sites is small compared to the contributions from the local source. This approximation is supported by modelling studies of heavy metals in the UK [13], although it is acknowledged that the complete picture is significantly more complex and contributions from medium and long-range transport may be significant [14,15].

3. Results and discussion

PCA allows the quantity of data considered to be reduced when correlation is present such that the data is re-plotted with respect to new axes which best describe the variability of the data set. This notwithstanding, PCA does not in itself identify possible outliers, but may be used to display the data in a way that standard quantitative outlier tests may be used to identify possible outliers.

In this way the process aims to identify individual sample filters or 'observations' that are outlying, because the relative proportions of metals measured on them is not consistent with the expected relative proportions. Once a filter is identified as a potential outlier, it is then necessary to review the data for that filter to identify which individual metals may have the problem associated with them. However, between filters, the metals identified as outlying may change. These changes may affect the position of the individual filters with respect to their scores on the first two PCs in different ways and so it is often necessary to examine more than just the first two PCs to ensure potential outliers are captured, as shall be demonstrated. The requirement therefore is to provide a single statistic following PCA analysis that combines the information from several PCs.

The results of the PCA for the observations (each comprising the concentrations of 10 elements) at each monitoring site on the UK Heavy Metals Monitoring Network in 2006 are displayed as a plot of eigenvalue against principal component (scree plot) for each monitoring site in Figure 1. As can be seen the monitoring sites show similar relative profiles. Often in PCA the number of useful PCs is determined from such a plot either qualitatively, by examining the point at which the curve, when plotted on linear axes, shallows out (the point at which the steep slope of 'mountain' becomes the shallow 'scree' at the base of the mountain), or quantitatively by using all PCs with eigenvalues of 1 or greater. Observing the data in Figure 1, this latter criterion would imply the use of only the

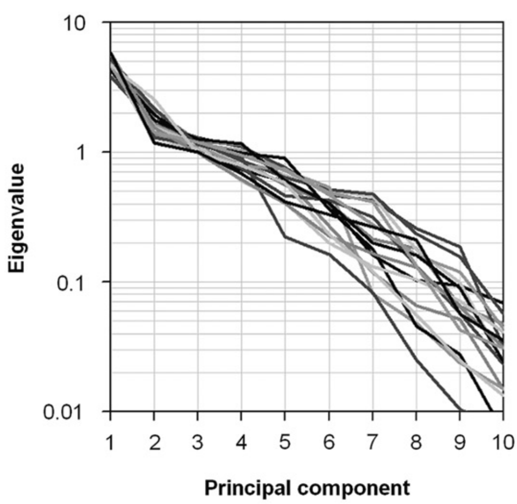

Figure 1. 'Scree plot' of eigenvalues (on a log scale) against principal components for the weekly observations of the concentration of 10 metals in ambient particulate matter in 2006 at 17 monitoring sites on the UK Heavy Metals Monitoring Network, each site being represented by a different line on the plot.

first three or four PCs. However, this may not be sufficient when attempting to minimise false negatives (i.e. outliers not picked up by the process).

The results of the PCA have been expressed as the vector distance of the individual observations from the origin of the principal component space, similarly to a technique previously outlined [16], such that:

$$D_{i,m} = \sqrt{\sum_{k=1}^{m} u_{i,k}^2} \tag{1}$$

where $D_{i,m}$ is the vector distance of observation i from the origin of the principal component space over the first m PCs, and $u_{i,k}$ is the principal component score of observation i on principal component k. As a result of the mean centring pre-processing $\sum_{i=1}^{n} u_{i,k} = 0$. This vector distance can then be expressed in relative terms as the proportion of the total deviation expressed by all observations attributable to one observation. This is given by:

$$D_{rel,i,m} = \frac{D_{i,m}}{\sum_{i=1}^{n} D_{i,m}} \tag{2}$$

where $\sum_{i=1}^{n} D_{i,m}$ is the sum of the vector distances from the origin of the principal component space of all n observations, over the first m PCs.

Since the data set ensemble considered using this technique is from a single monitoring site we would not expect several discreet clusters of observations, but a single cluster around the origin of the principal component space. Outlying observations would sit further away from this cluster and thus the proposed statistic in Equation (2) is a useful measure to detect such observations. This gives a simple summary statistic encompassing information from the first m PCs that can then be used to determine outlying observations (here representing suspect sampled filters) using traditional one-dimensional outlier tests.

This procedure has been undertaken for the data at the London Cromwell Road site in 2006 and this data is displayed in Figure 2. (There are fewer than 52 observations because of incomplete time coverage during the year.) As can be seen, as the number of PCs included in the analysis increases, the profile of the plot changes consistently up until about four or five PCs are used, after which point the profile of the plot remains fairly constant, but the relative magnitudes decrease. The important point to notice from this is that it is essential to include more than just the first two PCs to ensure that all significant outliers have been captured. Including extra PCs after the profile of the plot remains invariant and serves only to decrease the signal to noise ratio of outliers to the other observations.

In simple terms we want to include enough PCs to ensure that all outliers have been captured, but not so many that the prominence of the outliers becomes diluted by the random variability encompassed by the higher PCs. Deciding on the number of PCs that should be retained for outlier analysis is important. Retention of too few components can distort the output of the analysis giving both false positives and negatives, whereas retention of too many PCs diminishes the signal-to-noise giving false negatives. We will work on the general principle in this study that the optimum number of PCs will identify the maximum number of outliers and is likely to be greater than one. In the context of the analysis being considered here it is preferable to identify false positives, and discount these during later investigation, rather than allow false negatives to escape investigation. Since the potential discrepant values when they occur are generally extreme compared to the

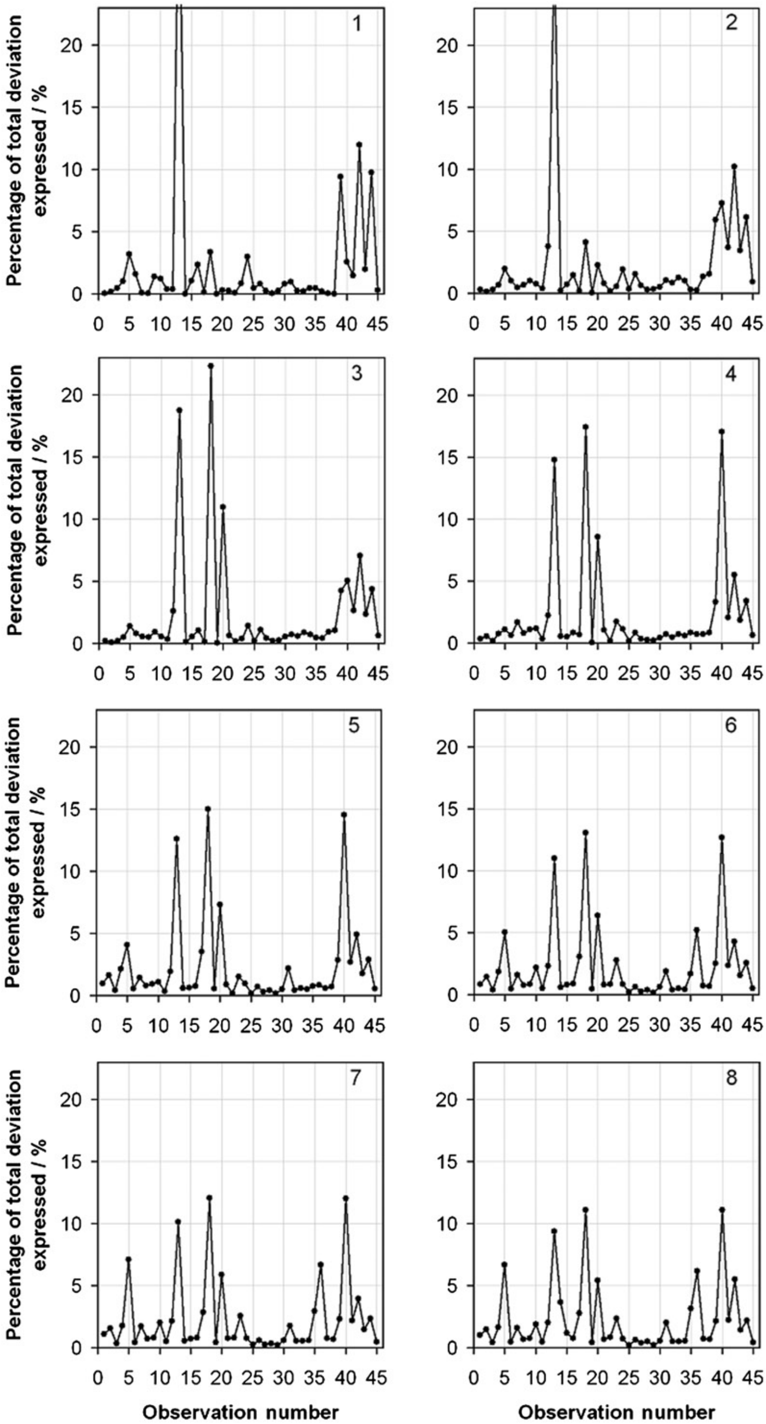

Figure 2. Percentage of the total deviation, $D_{rel,i,m}$, expressed by each observation (sampled filter) at Cromwell Road in 2006 (in chronological order), calculated with Equation (2) using a varying number of principal components from 1 to 8 (as displayed in the top right of each graph). The off-scale values for the plots using 1 and 2 principal components are 36 and 27, respectively.

mean value, the most appropriate method for outlier detection from the summary statistics will involve robust statistical methods that are not biased by extreme outliers [17].

In this way, for each plot in Figure 2 (including the analogous ones for 9 and 10 PCs not shown in Figure 2) the median and median absolute deviation of the set of $D_{rel,i,m}$ values have been calculated. The robust standard deviation has then been estimated (assuming the underlying concentration distributions to be normal) as 1.5 times the median absolute deviation. A possible outlier is then indicated if for observation *i*:

$$\frac{\left|D_{rel,i,m} - \hat{\mu}\right|}{\hat{\sigma}} \geq z \tag{3}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the median and robust standard deviations, respectively, of the set of n values of $D_{rel,i,m}$, and z is the outlier criterion. The results of this analysis for observations at Cromwell Road, Sheffield and Walsall, respectively, in 2006 are displayed in Figures 3(a), 3(b) and 3(c). As expected the number of possible outliers detected decreases as the value of z increases, however there is clearly some additional structure associated with the number of PCs included. In general the number of possible outliers detected is high when one PC only is included, but then falls initially as multiple PCs are included, before increasing again and reaching a maximum for the inclusion of about five PCs. After this point the number of possible outliers detected falls once more as the model moves towards including all the PCs. The results in Figures 2, 3(a), 3(b) and 3(c) support the earlier assertion that the optimum number of PCs to include is likely to maximise the predicted potential outliers, and be greater than one. At low numbers of PCs, high numbers of predicted outliers are observed but this is expected to include many false positives and false negatives. As more PCs are included, up to the number giving the most possible outliers, the number of false positives is minimised, whilst the number of false negatives falls further. As the number of PCs used increases further and the signal to noise decreases there is a large increase in false positives, but the number of false negatives does not show such a dramatic increase. Hence it seems sensible to err on the side of including slightly too many PCs, rather than too few. The effectiveness of the technique may additionally be tuned by variation of the outlier criterion, z. Unlike changing the number of PCs used in the analysis, the effect on the number of outliers predicted as z changes is monotonic. When z is too low too many false positives will be detected, and if z is too high there will be too many false negatives. False positives may be tolerated, as they may be re-examined and then discarded, but they act to increase the quantity of work required to be performed on the data set. False negatives, however, are more serious as they allow discrepant data to pass unnoticed. The sensitivity of the technique to the number of PCs included in the analysis is highlighted in Figures 4(a), 4(b) and 4(c), which display the observations identified as possible outliers for $m = 4$, 5 and 6, when $z = 3$, for observations at Cromwell Road, Sheffield and Walsall, respectively, in 2006. Considering the Cromwell Road data, as we move from 4 to 5 PCs an extra observation is identified as a potential outlier (which turns out to be a genuine outlier, see Table 1), and then as we move from 5 to 6 PCs a further extra observation is included (which turns out to be a false positive upon investigation), but two observations are dropped (which are genuine outliers – see Table 1 – thereby becoming false negatives). Similar trends are observed for the Sheffield and Walsall data, with 5 PCs predicting the maximum number of outliers in all cases, and non-identical sets of observations being predicted for the different PCs. This strongly supports the hypothesis that under these conditions the use of the first 5 PCs is the optimum configuration.

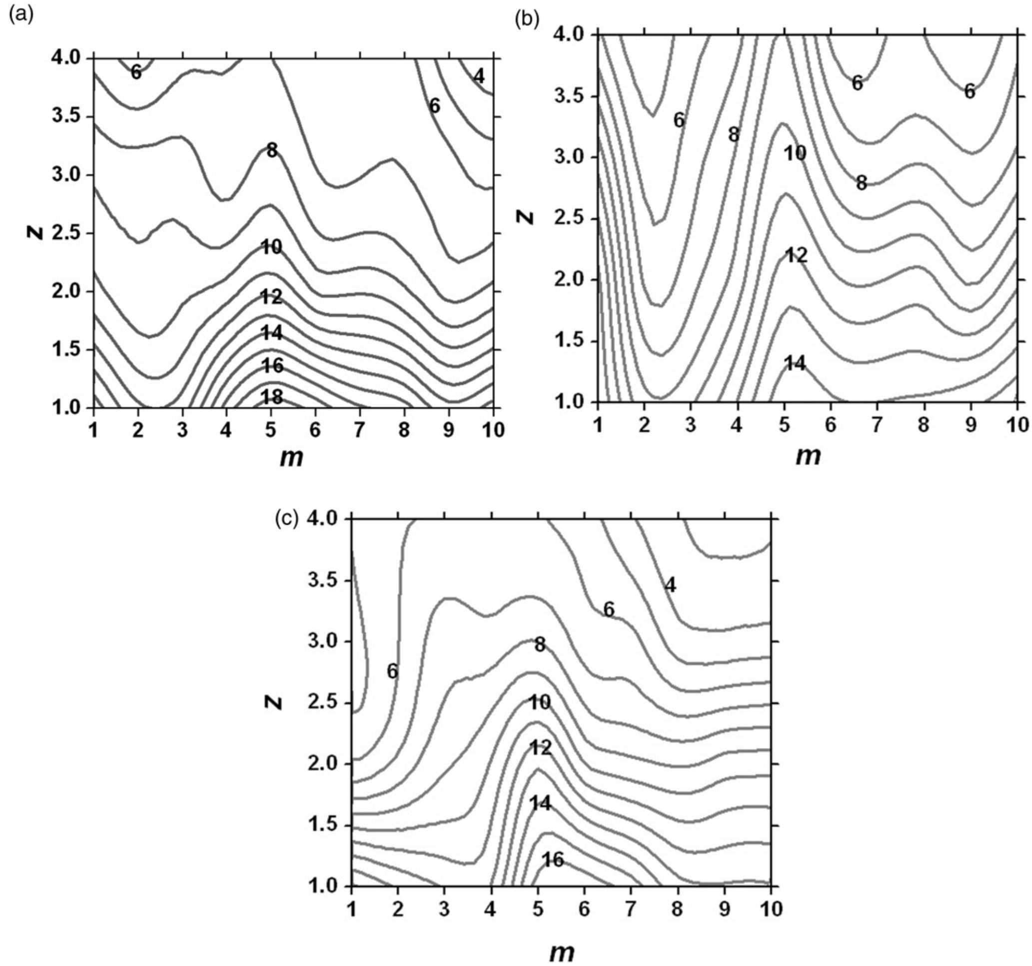

Figure 3. (a) The number of possible outliers predicted for the observations at Cromwell Road during 2006 using the method described in the text, as a function of the number of principal components, *m*, included and the outlier criterion, *z*. The outliers predicted decrease from 18 (bottom centre) to 3 (top right) with each contour line representing a change of 1 in this value – as additionally indicated by the contour markers. The contour lines have been smoothed for clarity. (b) The number of possible outliers predicted for the observations at Sheffield during 2006 using the method described in the text, as a function of the number of principal components, *m*, included and the outlier criterion, *z*. The outliers predicted decrease from 14 (bottom centre) to 5 (top left) with each contour line representing a change of 1 in this value – as additionally indicated by the contour markers. The contour lines have been smoothed for clarity. (c) The number of possible outliers predicted for the observations at Walsall during 2006 using the method described in the text, as a function of the number of principal components, *m*, included and the outlier criterion, *z*. The outliers predicted decrease from 16 (bottom centre) to 2 (top right) with each contour line representing a change of 1 in this value – as additionally indicated by the contour markers. The contour lines have been smoothed for clarity.

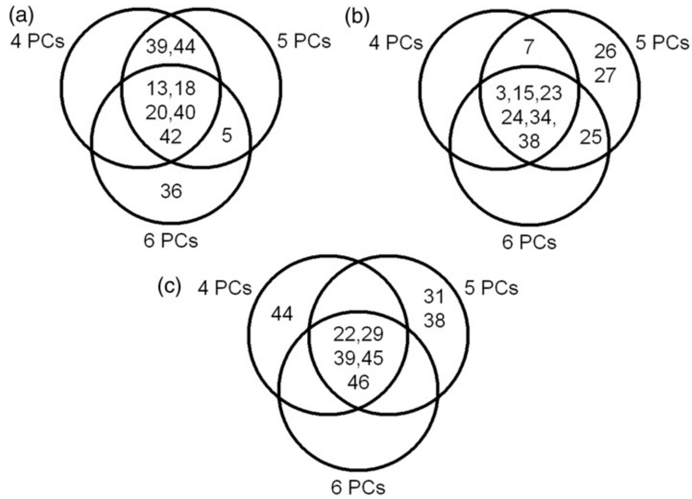

Figure 4. (a) Venn diagram displaying the observation numbers predicted as possible outliers at Cromwell Road during 2006 by the procedure described in the text using 4, 5 and 6 PCs when $z = 3$. (b) Venn diagram displaying the observation numbers predicted as possible outliers at Sheffield during 2006 by the procedure described in the text using 4, 5 and 6 PCs when $z = 3$. (c) Venn diagram displaying the observation numbers predicted as possible outliers at Walsall during 2006 by the procedure described in the text using 4, 5 and 6 PCs when $z = 3$.

Table 1. The possible outlying observations at Cromwell Road in 2006 identified using the procedure described in the text for $m = 5$ and $z = 3$, their outlying characteristics, the results of an investigation into the observation, and the action taken as a result of this investigation.

Observation number	Outlying characteristics	Results of investigation	Action taken
5	Cd (high)	High standard deviation on analytical measurement	Uncertainty of value increased accordingly
13	All elements (high)	Exposure volume under reported	All values corrected
18	Ni (high)	Likely filter contamination	Ni value removed
20	Ni (high)	Likely filter contamination	Ni value removed
39	All elements (low)	Sampler failure resulted in very low exposure volume	Observation removed
40	Cr, Mn (high)	No obvious problem	Values retained
42	As, Pb (high)	High standard deviation on analytical measurements	Uncertainty of values increased accordingly
44	All elements (low)	Sampler failure resulted in very low exposure volume	Observation removed

(This may not necessarily be exactly the same for other pollutant data sets, but the similarities of the plots in Figure 1 suggest that this conclusion is transferable for other sites on the UK Heavy Metals Network.)

From the data in Figures 2, 3(a), 3(b) and 3(c) values of z and m of 3 and 5, respectively, have been chosen as appropriate criteria for the test data sets presented here. Table 1 shows the results of this analysis for the Cromwell Road monitoring site. The procedure

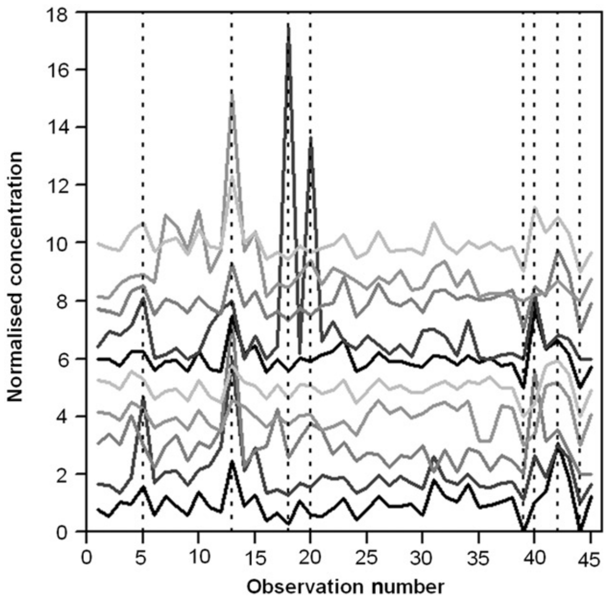

Figure 5. Concentrations for the 10 metals measured at Cromwell Road in 2006, normalised to the relevant 2006 average at Cromwell Road for each metal, in the order: As (bottom), Cd, Cr, Cu, Fe, Mn, Ni, Pb, V and Zn (top) with each plot offset by +1 from the plot beneath it. The possible outliers identified using the procedure described in the text for $m = 5$ and $z = 3$ are indicated by the vertical dotted lines.

suggested eight possible outliers. These are plotted in Figure 5 along with the normalised concentration data for the 10 elements measured at the monitoring site, for comparison. All the observations in Figure 5 have been investigated fully to determine whether there were problems with the sampling or analysis procedure. This identified seven values in the year, which were subject to likely filter contamination, sampling problems, data reporting errors, or analytical problems. The results of the novel method described above identified all seven of these observations, and additionally one false positive result for which no problem could be found. Figure 5 also gives a good indication of the value added to the outlier determination by using PCA. Robust outlier tests on the individual element concentrations alone would have produced many different possible sets of outliers, ultimately resulting in increased numbers of false positive and false negatives. The use of PCA prior to the use of robust statistics enables the expected correlation of the concentrations of individual elements to be taken into consideration.

4. Conclusions

The requirement to identify outliers in analytical data sets is always present. The need to do this for large and multivariate data sets requires more sophisticated methods. This paper has presented a novel method for the determination of outliers in multivariate air quality data sets. This technique involves the use of PCA to provide a summary parameter equal to the vector distance of the observation from the origin of principal component space. Robust outlier tests have then been applied to the values produced for each observation in order to identify possible outliers. It has been shown that the number of possible outliers identified is a function of the number of PCs included in the summary statistics for each observation and the number of robust standards deviations chosen as the outlier criterion. The number of PCs is chosen so as to minimise the number of false positive and false negatives (with false negatives being more serious). It has been observed that the specific observations identified as outliers change initially as extra PCs are included, but once the outliers identified are stable the addition of extra PCs acts only to decrease the signal to noise ratio of the summary statistics for a set of observations – ultimately decreasing the number of observations identified as outliers. This method has been demonstrated on exemplar data from the UK Heavy Metals Monitoring Network where it successfully predicted all the known outlying observations with only a small false positive rate. It has been shown that the proposed method is more effective and faster than the examination of trends of each of the individual metal concentrations for a set of observations, as it uses the expected correlation between the concentrations of different metals as a basis against which possible outlying data is judged, rather than the absolute deviation from the mean or median used by traditional outlier tests. It is proposed that this technique could be very useful in air quality networks that measure multiple pollutant species at each site, and where the concentrations of these pollutants are expected to show correlation. In particular it will be most useful at industrial and urban sites where correlations between primary pollutants from common sources are expected to be strong. The technique will be less useful at background and rural sites where secondary pollutants are being measured, or where long-range transport is significant, so strong correlation between observed values would not be expected. Further testing of this method on other monitoring sites, and for other pollutants, is essential to confirm the promising potential of this method demonstrated in the study presented here.

Acknowledgements

The UK Department for Innovation, University and Skills' funding of the National Measurement System Chemical & Biological Metrology Programme, and the UK Department for Environment, Food and Rural Affairs' funding of NPL's operation and management of the UK Heavy Metals Monitoring Network, are both gratefully acknowledged.

References

[1] R.J.C. Brown, R.E. Yardley, D. Muhunthan, D.M. Butterfield, M. Williams, P.T. Woods, A.S. Brown, and S.L. Goddard, Environ. Monit. Assess. **142**, 127 (2008).
[2] R.J.C. Brown, Analyst **130**, 1280 (2005).
[3] R.J.C. Brown, Analyst **132**, 344 (2007).
[4] C. Caroni, Commun. Statist. Simul. **29**, 139 (2000).
[5] W.J. Egan and S.L. Morgan, Anal. Chem. **70**, 2372 (1998).
[6] M. Hubert, P.J. Rousseeuw, and K. Vander Branden, Technometrics **47**, 64 (2005).
[7] J.C.M. Pires, S.I.V. Sousa, M.C. Pereira, M.C.M. Alvim-Ferraz, and F.G. Martins, Atmos. Env. **42**, 1249 (2008).
[8] Y. Song, S.D. Xie, Y.H. Zhang, L.M. Zeng, L.G. Salmon, and M. Zheng, Sci. Tot. Env. **372**, 278 (2006).
[9] UK Air Quality Archive. <www.airquality.co.uk> (accessed February 2009).

[10] Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air, Official J., L 023, 26/01/2005 P. 0003–0016.
[11] CEN, EN 14902:2005, Ambient air quality – Standard method for the measurement of Pb,Cd, As and Ni in the PM10 fraction of suspended particulate matter (CEN, Brussels, 2005).
[12] R.J.C. Brown, R.E. Yardley, A.S. Brown, and M.J.T. Milton, J. Anal. At. Spectrom. **19**, 703 (2004).
[13] K. Vincent and N. Passant, Assessment of Heavy Metal Concentrations in the United Kingdom: Report AEAT/ENV/R/2013 to the Department for Environment, Food and Rural Affairs, Welsh Assembly Government, the Scottish Executive and the Department of the Environment for Northern Ireland, 2006.
[14] S. Canepari, C. Perrino, F. Olivieri, and M.L. Astolfi, Atmospheric Environ. **42**, 8161 (2008).
[15] S. Canepari, A. Pietrodangelo, C. Perrino, M.L. Astolfi, and M.L. Marzo, Atmospheric Environ. (2009) doi: 10.1016/j.atmosenv.2008.09.059, in press.
[16] B. Mertens, M. Thompson, and T. Fearn, Analyst **119**, 2777 (1994).
[17] Analytical Methods Committee, Technical Brief No. 6, ISSN 1757–5958 (2001).