UNCOVERING THE DEPENDENCE OF MAXIMUM PARTICULATE AND GASEOUS AIR POLLUTANT CONCENTRATIONS ON WIND DIRECTIONS BY NONPARAMETRIC REGRESSION

Ch. HUEGLIN[1], A. MARTI[2] and A. RUCKSTUHL[2]

[1]Swiss Federal Laboratories for Materials Testing and Research (EMPA), 8600-Duebendorf, Switzerland

[2]Institute for Data Analysis and Process Design, Zurich University of Applied Sciences (ZHW), 8401-Winterthur, Switzerland

Keywords: PM10, Air quality monitoring, Source identification, Gaussian kernel estimator.

INTRODUCTION

In a recent study by Henry *et al.* (2002), nonparametric regression was successfully applied to determine the relationship between the concentration of air pollutants and wind direction. A similar but modified method is applied here to identify wind directions with maximum average concentrations of particulate (PM10, particle number concentration) and gaseous air pollutants (e.g. NO_x, CO, O_3, NMHC). The outputs of the described method are smooth curves and associated confidence bands. In this study, nonparametric regression models are used to verify the site characterisation scheme within the Swiss national air pollution monitoring network (NABEL) and to study the impact of nearby sources.

METHODS

Figure 1 shows a scatter plot of 10 minutes mean values of PM10 versus wind direction at a rural site during 2002. From this representation, it seems impossible to recognise wind directions associated with elevated average air pollutant levels. Therefore, data of the kind shown in Figure 1 are typically grouped into bins of a certain width and displayed as bar charts. However, such bar charts have major limitations: They depend on the selected bin size and on the location of the bin boundaries. Kernel smoothing methods can be used to overcome these limitations (Härdle, 1990). As with moving averages, averages are taken over sliding windows of width λ. However, a weight scheme is incorporated to down-weight the contributions of data points away from the centre x_0 of the window. Reasonable weights are selected e.g.

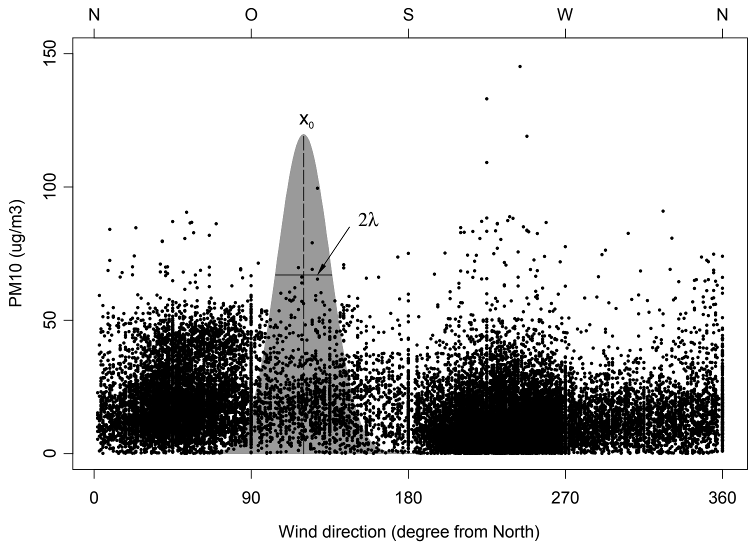

Figure 1. Ten minutes means of PM10 versus wind direction during 2002 at the rural site in Taenikon, Switzerland. Measurements with wind speed < 0.5 m/s are excluded from the above figure, because the wind direction is not well defined at low wind speeds. The grey area indicates a Gaussian kernel at target point x_0 with bandwidth λ.

according to a normal distribution (so-called Gaussian kernel, see Figure 1). The bandwidth λ is the only adjustable parameter. The optimum value for λ is determined by minimising the mean squared error (MSE) of the estimated concentration as a function of λ. The MSE is estimated by splitting the data set in a training and test set. Edge effects with the considered circular data are prevented by gluing the edges at 0° and 360° together.

RESULTS

PM10 as well as concentrations of other air pollutants are typically not normally distributed (visible in Figure 1) but may rather follow a lognormal distribution. Consequently, a logarithmic transformation of the data is employed. The 95% confidence bands are estimated according to Henry *et al.* (2002).
Figure 2 gives an example of nonparametric regression of PM10 and NO_x on wind direction for data from Taenikon. Taenikon is a rural site in an agricultural area in north eastern Switzerland. However, from the northern to eastern direction, a motorway passes the site in a distance of four to five kilometres. The question is, whether an impact of vehicle emissions from the motorway can be detected at the site. From the analysis it is very likely that the site is influenced by emissions from the nearby motorway. Smoothed PM10 and NO_x concentrations are highest for wind directions from north to east (the effect is approximately significant when confidence intervals do not overlap). Note that the confidence intervals are wider for less frequent wind directions (cf. sector from 90-180°).
The described method can also be useful for a variety of other applications in atmospheric sciences, e.g. for the investigation of the temporal behaviour of air pollutants.

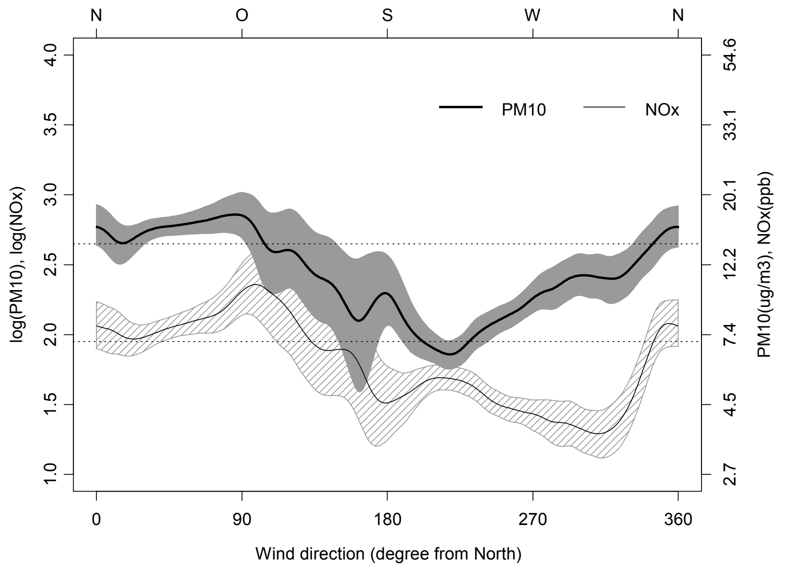

Figure 2. Uncovering the dependence on wind directions. Nonparametric regression of PM10 and NO_x on wind direction for ten minutes mean values during 2002 at the rural site in Taenikon, Switzerland. A Gaussian kernel with λ=18° (PM10) and λ=19° (NO_x) was used. The grey and shaded regions are the associated 95% confidence bands. Data with wind speeds < 0.5 m/s are excluded.

REFERENCES

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press, Cambridge.
Henry, R. C., Yu-Shuo, C., Spiegelmann C. H. (2002). Locating nearby sources of air pollutants by nonparametric regression of atmospheric concentrations on wind direction. *Atmos. Environ.*, 36, 2237-2244.