Abstract. Since the 60’s, there has been a strong industrial development in the Sines area, on the southern Atlantic coast of Portugal, including the construction of an important industrial harbour and of, mainly, petrochemical and energy-related industries. These industries are, nowadays, responsible for substantial emissions of SO₂, NOₓ, particles, VOCs and part of the ozone polluting the atmosphere. The major industries are spatially concentrated in a restricted area, very close to populated areas and natural resources. Their emissions are very similar, making the identification of individual pollutant sources and of their contributions to air pollution difficult.

In this study, the regional spatial dispersion of sulphur dioxide (SO₂) is characterized, through the combined use of diffusive tubes (Radiello Passive Samplers) and classical monitoring stations’ air quality data.

The objective of this study is to create a regional predictive model of the contribution of different emission sources to the pollutant concentrations captured at each monitoring station.

A two-steps methodology was used in this study. First, the time series of each data pair – industrial emission and monitoring station records – was filtered, in order to obtain contiguous time periods with a high contribution of that specific industrial emission to the equivalent monitoring-station measurements. For this purpose, an iterative optimisation process was developed, using a variogram between industrial emissions and monitoring-station time series as the objective function. Afterwards, probability neural networks (PNN) were applied to achieve an automatic classification of the time series into two classes: a class of (emission/monitoring station) pairs of highly correlated points and a class of pairs of points without correlation.

In a second step, the relationship between emissions and air quality (AQ) monitoring station records – time model – is validated for the entire area for a given period of time, using for that purpose the diffusive samplers measurements. A spatial stochastic simulation is applied to generate a set of equi-probable images of the pollutant, which relationship with the different emissions is calculated using the PNN predictive model.

1 Introduction

It is well known that air pollutants at ground level can be harmful to human health, if their concentrations exceed certain limits. As pollutants accumulate in, or near, large
metropolitan areas, populations are typically more exposed to unhealthy pollutant concentrations (Russo et al., 2004; Cobourn et al., 2000; Kolehmainen et al., 2000). Considering the effect of air pollutant concentrations on human health, a study that allows the identification of regional pollutant emission-receptions patterns and the quantification of the contribution of local industrial units is of great interest for the health system and environmental policy making (Russo et al., 2004; Cobourn et al., 2000; Kolehmainen et al., 2000).

Predictive modelling of the different emissions’ contribution to the pollutant concentrations captured at a certain monitoring station, will allow an analysis of the impact caused in the monitoring station’s area and its translation into an air quality index (Russo et al., 2004).

Nevertheless, in order to develop robust predictive air quality (AQ) models, wide-range monitoring systems are necessary. Modelling therefore often needs to be used in conjunction with other objective assessment techniques, including monitoring, emission measurement and inventories, interpolation and mapping (WHO, 1999).

Air quality monitoring can essentially be accomplished by the use of continuous automatic sensors, passive samplers, active samplers and remote sensors. The use of passive samplers offers a simple and cost-effective method of screening air quality in an area (Cruz and Campos, 2002; Mukerjee et al., 2004). The low unit costs permit sampling at numerous points in the area of interest, what is useful in highlighting “hot-spots” of high pollutant concentrations. Combined with automatic sensors, that can provide high-resolution measurements (typically hourly averages) at few point for most of the criteria pollutants (SO_2, NO_2, O_3), a spatial-temporal monitoring net may be accomplished.

Briefly, the purpose of this study is to analyse possible relations between sulphur dioxide (SO_2) emissions, generated by one industrial complex (Petrogal) located in the Sines area (Portugal), and AQ data collected by three air quality monitoring stations (Sonega, Monte Chãos, Monte Velho) and also by Radiello diffusive tubes covering the Sines area, by means of linear and non-linear modelling, as described in section 4.

2 Objectives

The objective of this study is to create a predictive model of the contribution of different emission sources to the pollutant concentrations captured at each monitoring station. The combined use of spatial information (captured by passive monitors (diffusive tubes)) and temporal information (captured by monitoring stations) for the same pollutants will allow to create a model of the contribution of different emission sources to pollutant concentrations in the area.

A two-steps methodology was used in this study: i) First, the time series of each data pair – industrial emission and monitoring station records – was filtered, in order to obtain contiguous time periods with a high contribution of that specific industrial emission to the equivalent monitoring-station measurements. For this purpose, an iterative optimisation process was developed, using the variogram between industrial emissions and monitoring-station time series as the objective function. Afterwards, probability neural networks (PNN’s) were applied to predict the probability of pollutant emissions causing the pollutant concentrations measured at the monitoring stations;
ii) In a second step, the relationship between emissions and AQ monitoring station records – time model – is validated for the entire area for a given period of time, using the diffusive samplers measurements. A spatial stochastic simulation (direct sequential simulation) is applied to generate a set of equi-probable images of the pollutant and the relationship of different emissions with local simulated values is evaluated for the entire area.

3 Case Study

The main objective of this study consists in developing and implementing a methodology that allows classifying the contribution of different emission sources to air quality (AQ) in the region of Sines, Portugal (Fig. 1). Automatic sensors and passive samplers (Radiello diffusive tubes) were used in order to collected AQ data in the Sines area.

The case study covers an area with very different land occupations: industrial, urban, rural and leisure. Although the urban occupation is very small, compared with the rural occupation, the industrial sources are of great importance and make an important contribution to long-term or peak concentrations of criteria pollutants (SO$_2$, NO$_2$, O$_3$, etc.). This kind of areas, with mixed occupation, should be continuously monitored and controlled, in order to prevent future critical situations in terms of air quality.

![Figure 1. An overview of the Sines Peninsula (Petrogal, Borealis and CPPE industrial complexes in light gray; AQ monitoring stations in Santiago do Cacém, Sonega, Monte Chãos, Monte Velho in dark gray).](image)

The sulphur dioxide concentration (mg/m$^3$) is daily measured in three main industrial emissions (Borealis, Petrogal and CPPE) and four monitoring stations – Sonega, Monte Chãos, Monte Velho and Santiago do Cacém. For the purpose of this study, those
measures were converted into daily averages for a period of 12 months (from 1/1/2002 to 31/12/2002) (Figs. 2 3). Meteorological data – wind speed and direction on an hourly basis, for the same period – was also collected and analysed (Fig. 4).

![Figure 2](image1.png)

**Figure 2.** SO$_2$ emitted by the three industrial complexes.

![Figure 3](image2.png)

**Figure 3.** SO$_2$ measured by the monitoring stations (SO - Sonega, MC - Monte Chãos, MV - Monte Velho).

![Figure 4](image3.png)

**Figure 4.** Wind speed and modal wind direction registered.

The available data was standardized in order to minimize the effect of different local means and variances in the evaluation of the emissions/AQ measurements relationships. Afterwards, those days, which did not have any register of data in at least one of the emission-reception stations, were deleted from the file. Diffusive tubes measurements of SO$_2$ were available for a period of 11 consecutive days (from 31/3 to 10/4/2003) (Fig. 5). The sampling period was characterized by dominant
winds from north/northwest with average speeds of 10-17 km/h. The humidity levels varied between 80% and 100%. The air temperature had a typical spring variation, with an average temperature of about 15 °C.

![Diffusive tubes spatial SO2 dispersion](image)

**Figure 5.** Diffusive tubes spatial SO2 dispersion.

4 Methodology

The two steps methodological approach proposed for this study can be summarized as follows: i) A predictive model of the different emission sources contributions to the pollutant concentrations, captured at each monitoring station, divides the time series into two classes: pairs of highly correlated points and pairs of points with poor correlation; ii) Validation of a time model for the entire area.

4.1 CLASSIFICATION OF TIME PERIODS WITH HIGH CORRELATION BETWEEN EMISSION AND MONITORING STATION RECORDS

After the first attempts of including meteorological variables into the prediction models, we concluded that the available data of wind speed and direction wasn’t responsible for the observed dynamics of the different pollutant plumes; the main reason being that the meteorological data was often collected at an altitude and locations inadequate to capture emissions from the Petrogal chimneys.

In a first step, the time series of each data pair – industrial emissions and monitoring stations records – was filtered with the purpose of obtaining contiguous time periods with high correlation between that specific industrial emission and the equivalent
monitoring station measurements. This process consists in the implementation of a simple iterative procedure. The variogram of each pair of emissions-monitoring station AQ measurements during a period T (365 days – N error values) is:

\[
\gamma(z_1, \Psi_1) = \frac{1}{T} \sum_{i=1}^{T} \left( z_1(i) - \Psi_1(i) \right)^2,
\]

where \( z_1(i) \) and \( \Psi_1(i) \) are the measurements of the emission source \( z_1 \) and of the monitoring station \( \Psi_1 \) for the instant \( i \) after standardization. This variogram was assumed as an objective function that tends to decrease (increasing the correlation between \( z_1 \) and \( \Psi_1 \)) as pairs of points with less contribution are iteratively removed.

A probabilistic neural network (PNN) was used to automatically classify data into the two classes described above. PNN’s can be useful for classification problems and have a straightforward design. A PNN is guaranteed to converge to a Bayesian classifier providing it is given enough training data. And, these networks generalize well (Haykin 1994, Beale and Demuth 1998).

4.2 VALIDATION OF TIME MODEL FOR THE ENTIRE AREA

The obtained PNN is a predictive (classification) model valid for a period with statistical characteristics identical to the past and for the emissions-AQ monitoring station records pairs. The objective of this second step of the proposed methodology, is to validate and generalize this classification model for the entire area. In other words, to analyse the spatial extension of the classification model, calculated and tested for the AQ monitoring stations.

Hence the following geostatistical methodology is applied:

i) First, diffusive tubes measurements are used to determine a local trend of the SO\(_2\) concentration corresponding to the 11 days period, through ordinary kriging;

ii) Based on the diffusive tubes variograms (spatial pattern) and the monitoring stations AQ values, a set of simulated images of SO\(_2\) is obtained for the 11 days period, using direct sequential simulation (Soares, 2000) with local means, i.e., the local trend previously calculated;

iii) To validate the classification model for the entire area, the individual contributions of different emissions were mapped as follows: After averaging the simulated images for each day, the resulting most probable image was classified with a PNN (cf. Section 4.1), resulting in areas with high and areas with low correlation with the different emissions.
5 Results and Discussion

5.1 CLASSIFICATION OF TIME PERIODS WITH HIGH CORRELATION BETWEEN EMISSION AND MONITORING STATION RECORDS

With the purpose of obtaining contiguous time periods with high correlation between each pair of industrial emission and monitoring station measurements, the time series of each data pair was previously filtered using the methodology described in Section 4.1. An example of scatter plot of the standardized values of the original data series, for the Petrogal (emission) and Sonega (monitoring station) pair, is shown in figure 6 (a). A scatter plot of Class 1 data points (high correlation between emission and AQ monitoring station) is shown in figure 6 (b). Figure 7 represents the time series of these Class 1 values, showing contiguous periods of time, i.e., time periods where, in principle, the meteorological conditions are in accordance with the direction emission/AQ monitoring station.

Figure 6. (a) Petrogal and Sonega’s data sets before being filtered; (b) Petrogal and Sonega’s Class 1 data points.

Figure 7. Petrogal and Sonega’s Class 1 data points.
The ability of the PNN to classify, correctly, time series into Class 1 (highly correlated pairs of emission and AQ monitoring stations) and Class 2 (pairs of emission and AQ monitoring stations with poor correlation) was superior to 90%, for the three monitoring stations.

5.2 VALIDATION OF TIME MODEL FOR THE ENTIRE AREA

Validation of the PNN predictive model is necessary to evaluate the probability of other areas around the monitoring stations to belong to either of the classes of correlation with the emissions or, in other words, to evaluate the probability that the pollutant concentration in other regions is caused by the industrial emissions.

First, the diffusive tubes measurements (Fig. 5) were used to calculate (through ordinary kriging) a local trend for the pollutant concentration for the 11 days period (Fig. 8).

![Spatial trend of the SO2 dispersion measured in the diffusive tubes.](image)

As the diffusive tubes are the only available spatial data, it is assumed that the variogram calculated with this data reflects the spatial pattern of the average behaviour for the 11 days period. Hence, the variogram model for the diffusive tubes measurements – following an isotropic spherical model with one structure of range \( a = 20,000 \) m – was considered for the subsequent steps.

Direct sequential simulation was applied to generate a set of 30 images. The local trend of Figure 8 was assumed as local mean. AQ monitoring stations values of those 11 days were taken as conditioning data.

In Figure 9 examples of SO2 maps simulated for three consecutive days are shown. Average and variance maps for the first and last day of the 11 days period are shown in Figure 10 a) and b), respectively.
Figure 9. Examples of SO$_2$ maps simulated for three consecutive days.

Figure 10. Examples of SO$_2$ average (a) and variance (b) maps for the first and last day of the 11 days period.

With the simulated spatial images, we attempted to calculate the correlation coefficient between each set of simulated images and each emission for the 11 days. But, given the very homogeneous time period in terms of emissions, the resulting correlation coefficients were, most of the times, rather spurious statistics. Hence, after averaging the simulated images for each day, the resulting most probable image was classified with a PNN (Fig. 11). Figure 11 shows the areas with highly correlated points and areas without correlation, with the different emissions.
Figure 11. Areas with highly correlated points and areas without correlation, with the different emissions.

Note that the PNN determines the probability of a given pair of points belonging to a linear relationship between emissions and monitoring stations. Hence, all PNN for the three industrial emissions and AQ monitoring stations are very similar, producing similar final maps for the entire region.

As all of the emissions, coincidentally, show intermediate values for the 11 days period, one can see that:

i) The areas, which have a high probability of being related with the emissions, are the ones with intermediate values of pollutant concentration.

ii) The PNN determines the probability of a given pair of points belonging to a linear relationship between emissions and monitoring station records. Hence, all PNN for the three industrial emissions and AQ monitoring stations are very similar. As the emissions are similar for the 11 days period of time, the final maps of each emission contribution to the pollution of entire region are also similar.

iii) The areas affected by high pollutant concentrations do not show any correlation with any of the industrial emissions. In fact, both high-value plumes are located in the two main villages of the region, suggesting other pollutant source than the industrial emissions.

5.3 DISCUSSION

Combining two AQ sampling systems – classical monitoring stations and diffusive tubes – we succeeded in showing an approach that allows for an impact evaluation of the different emissions for the entire Sines area.
The predictive time model is strictly valid for the spatial location of the emissions/monitoring stations pairs. With a more spatially representative monitoring – using diffusive tubes – and with a spatial geostatistical model – through stochastic simulation – the model is successfully generalized for the entire area.

Inferences for the entire area are obviously just valid for the period of the diffusive tubes exposure. The more yearly campaigns of diffusive tubes one has, the more representative (in terms of space and time) the conclusions become.

In this case study, the conclusions taken with the eleven days of the first campaign are just illustrative of the potentiality of the two steps approach. Although the results are coherent, it does not validate or the model for the entire space-time domain of the study.

6 Conclusions

This study deals with a well-known characteristic common to most AQ monitoring networks: high density of sample values in time collected at just few spatial locations. This can be a serious limitation if one wishes to evaluate impact costs or carry out an environmental risk analysis of the emissions for the different land uses, eco-systems and natural resources of a region.

The presented approach, based on the use of two different monitoring systems – AQ monitoring stations, with an high density sampling rate in time, and diffusive tubes, that cover the entire space for a limited period of time – shows to be an alternative for the impact study of the air quality in the entire region.

In spite of the illustrative purpose of this paper, it is worth mentioning that the model should be validated for the entire area with more diffusive tubes campaigns. It is important to acknowledge that the model’s performance could also be improved using longer AQ data series and another kind of meteorological data.

References


