

# Spatio-Temporal Analysis of Air Pollution Data in Malta

Ms Luana Chetcuti Zammit<sup>1</sup>, Dr Kenneth Scerri<sup>1</sup>, Dr Maria Attard<sup>2</sup>, Ms Thérèse Bajada<sup>3</sup>, Mr Mark Scerri<sup>4</sup>

<sup>1</sup>Systems and Control Engineering Department, Room 419, Engineering Building, University of Malta, MSD2080, Malta.  
Tel: +356 2340 2080  
E-mail: lche0003@um.edu.mt; kenneth.scerri@um.edu.mt

<sup>2</sup>Geography Department, University of Malta, OH132, University of Malta, Msida MSD2080, Malta.  
Tel: +356 2340 2147  
E-mail: maria.attard@um.edu.mt

<sup>3</sup>Institute for Sustainable Development, First Floor, Regional Building, Triq l-Imhalled Paolo Debono, University of Malta, MSIDA MSD 2033, Malta.  
Tel: +356 2340 3404  
E-mail: therese.bajada@um.edu.mt

<sup>4</sup>Malta Environment and Planning Authority, Unit D2, Environment Protection Directorate, PO Box 200, Marsa MRS1000, Malta.  
Tel: +356 2290 7203  
E-mail mark.scerri@mepa.org.mt

## 1. Introduction

Air pollution measurements display patterns over space and time allowing for spatio-temporal modelling, through which pollution concentrations and trends can be analysed. In Malta, the MEPA (Malta Environment and Planning Authority) collects monthly averaged data for various pollutants from a network of 123 diffusion tubes located around the Islands (Figure 1). This preliminary study uses data associated with traffic, that is nitrogen dioxide (NO<sub>2</sub>) and benzene, collected monthly between the period 2004 and 2010 with the objectives to i) develop a computationally efficient method that best describes the data; ii) determine the level of dependency of each site on neighbouring ones and iii) identify any factors that affect the behaviour and patterns of pollution. Results will show that generally there is a low spatial dependency between close sites, thus implying that local sources, rather than diffusion, have a predominant effect on the measurements. This analysis will prove valuable in MEPA's redistribution exercise of the diffusion tube network to determine which sites are necessary to retain and which sites can be removed without significantly affecting the information gathered.

## 2. The case study - Malta

Malta is located in the centre of the Mediterranean Sea at approximately 100 kilometres south of Sicily. The archipelago consists of three main islands with a population of over 400,000, with the major environmental concern being the air quality (Government of Malta 2002, Office of the Prime Minister 2010). The main contributors to air pollution are the high demands for energy generation and the growth in private car use. Electricity is generated from the combustion of fuel oil at two power stations. In addition, the islands were home to 229,016 private vehicles in 2009 (NSO 2009), one of the highest car ownership rates in the world. These have increased the risks of atmospheric pollution particularly by carbon monoxide, oxides of nitrogen, volatile organic compounds and particulate matter (MEPA 2010).

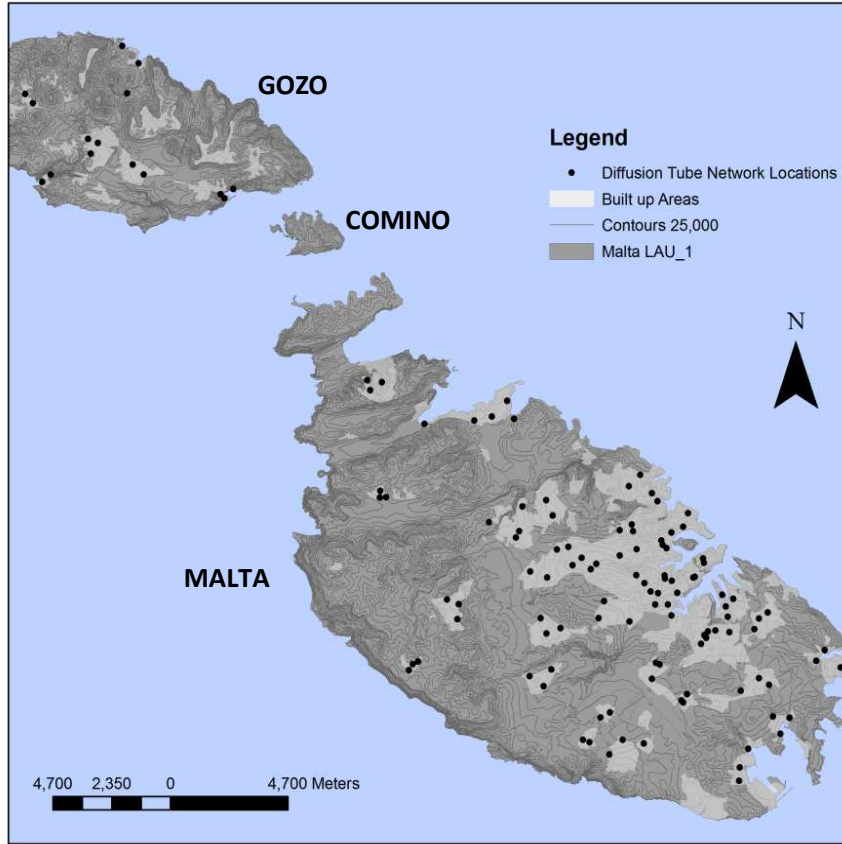


Figure 1. Location of Passive Diffusion Tubes (drawn by author)

### 3. Data Analysis and Validation

Data-driven modelling strategies (Ljung 1999) are applied in this research to study the dependencies between readings taken at different sites. In particular, statistical multivariate or Vector AutoRegressive (VAR) models are used to represent the spatio-temporal relationships in the data. These methods were first proposed for the study of such phenomena in Pfeifer and Deutsch (1980a, b).

Seasonal and temporal trends in the data were eliminated so as to identify dependencies among sites rather than correlations due to these trends. Detrending was performed by differencing (Chatfield 2004) and the stationary time series obtained were modelled by the VAR( $p, q$ ) models, given by equation 1:

$$\mathbf{z}_t = \mathbf{A}_1 \mathbf{z}_{t-1} + \mathbf{A}_2 \mathbf{z}_{t-2} + \mathbf{A}_3 \mathbf{z}_{t-3} + \mathbf{A}_4 \mathbf{z}_{t-4} + \dots + \mathbf{A}_p \mathbf{z}_{t-p} \quad (1)$$

where  $p$  denotes the temporal order,  $q$  denotes the spatial order of the system,  $\mathbf{A}_i \in \mathbb{R}^{123 \times 123}$  are the

autoregressive terms and  $\mathbf{z}_t \in \mathbb{R}^{123 \times 1}$  denotes the air pollution observations of all 123 sites at time  $t$ .

Due to the large number of parameters to estimate from the limited data, the computationally advantageous method in de Luna and Genton (2005) was adopted, summarized in Algorithm 1. Some assumptions have been made, namely that the climate remains homogenous throughout all sites under study, that the spatially closer sites have a larger probability of being correlated thus providing a natural ordering for the sites and that only temporal dependences over a monthly period can be captured due to the data's temporal resolution and thus shorter term dependences cannot be ruled out.

*Iterate for site,  $s_i=1,2,\dots,123$*   
*Order all sites in ascending order of distance relative to  $s_i$*   
*For  $p=1,2,\dots,n$*   
*For  $q=1,2,\dots,k$  (where  $n$  and  $k$  represent the maximum temporal and spatial order respectively)*  
*Estimate  $A_1, A_2, \dots, A_p$*   
*Identify the best model orders for  $s_i$  based on some comparative measures*

Algorithm 1: Iterative model building strategy (de Luna and Genton 2005)

The applied comparative measures are based on model selection criteria, which include the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC) and the Mean Squared Error (MSE). The principle of parsimony (Chatfield 2004) is applied thus aiming to identifying the simplest model that produces good prediction results. Figures 2 and 3 show MSE, AIC and BIC values for different temporal model orders for benzene and NO<sub>2</sub> respectively. Since AIC and BIC values tend to penalize models with larger spatial and temporal orders, these results are weighted more heavily when deciding on the model order.

Note that in Algorithm 1, the computational demand is significantly reduced when compared to estimating full VAR models since only the statistically significant coefficients are estimated. Note also that, to the authors' knowledge these methods have only been applied to datasets with a small number of observation sites (usually less than 10), while in this work the flexibility of this method to solve higher dimension problems (123 sites) has been tested. The predictive accuracy of the models obtained has been tested by a validation data set not used in the estimation procedure. Using this dataset, the one step ahead residues after modelling were found to be temporally white (thus void of any further linear temporal relationship) up to a confidence interval of 88% for benzene and 91% for NO<sub>2</sub>. Based on these models the spatial order for both benzene and NO<sub>2</sub> was mostly limited to the three closest neighbours.

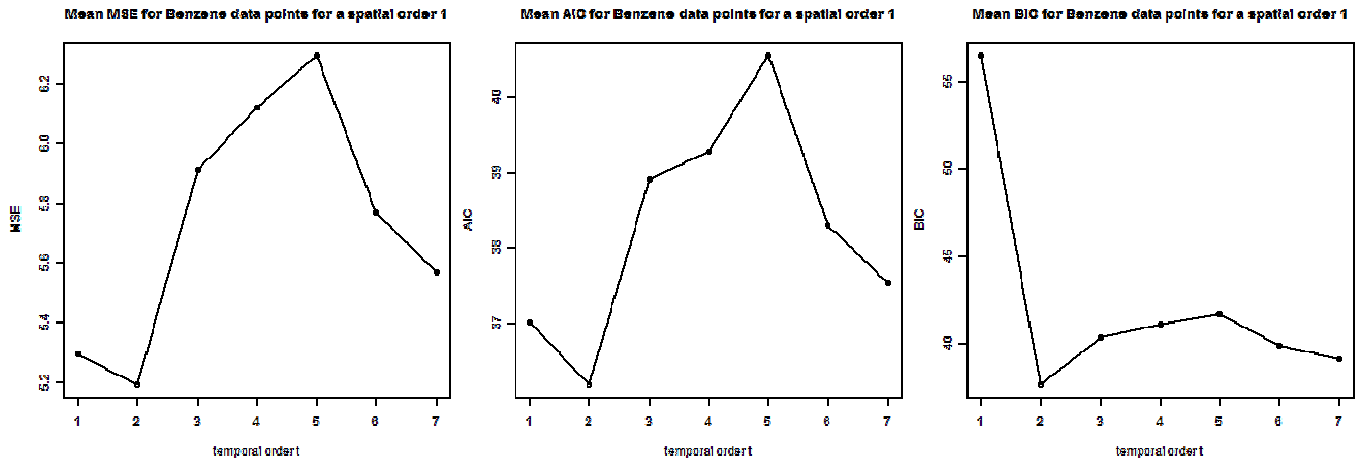


Figure 2. The mean MSE, AIC and BIC for benzene over all sites.

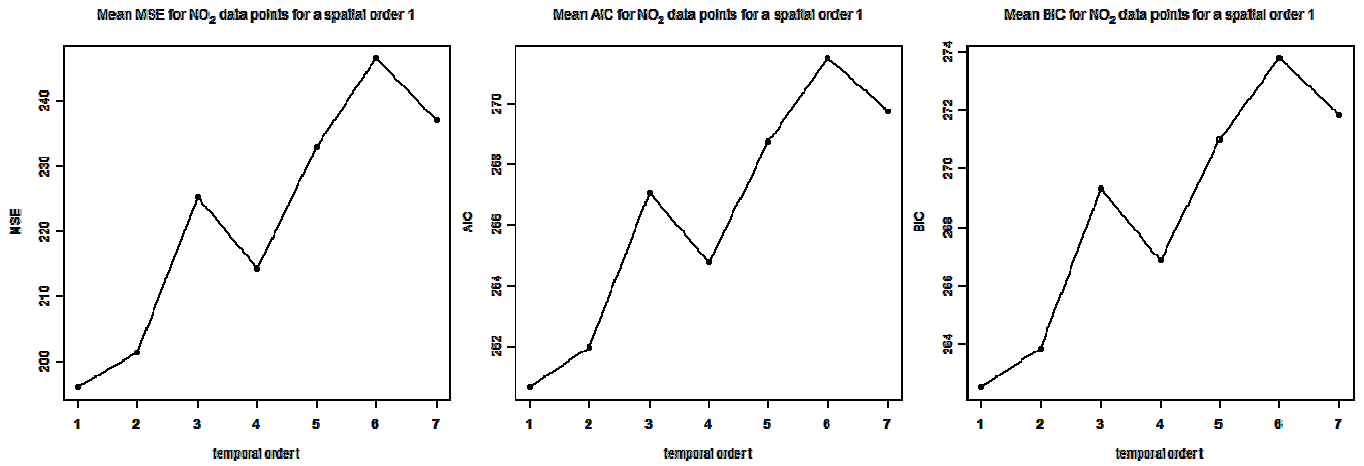


Figure 3. The mean MSE, AIC and BIC for NO<sub>2</sub> over all sites.

#### 4. Conclusions

Figures 4a and 4b give histograms for the number of dependent sites while Figures 5a and 5b show the number of dependent sites for each location for benzene and NO<sub>2</sub> respectively.

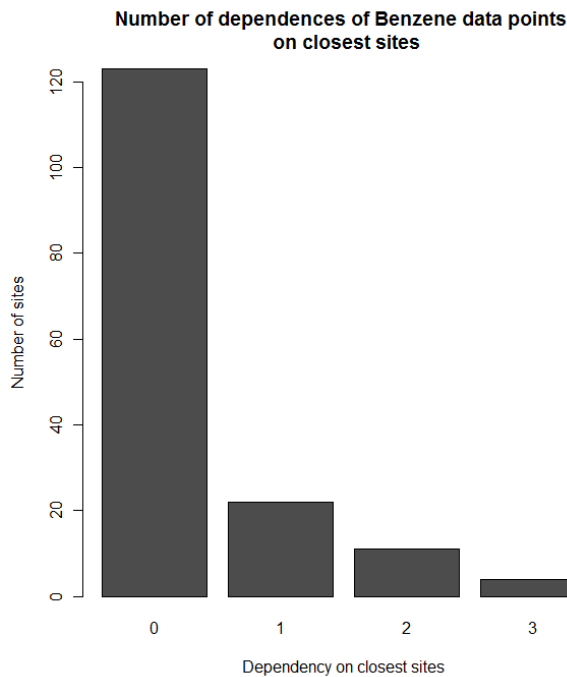


Figure 4a. Histogram showing dependency of Benzene data points on closest sites

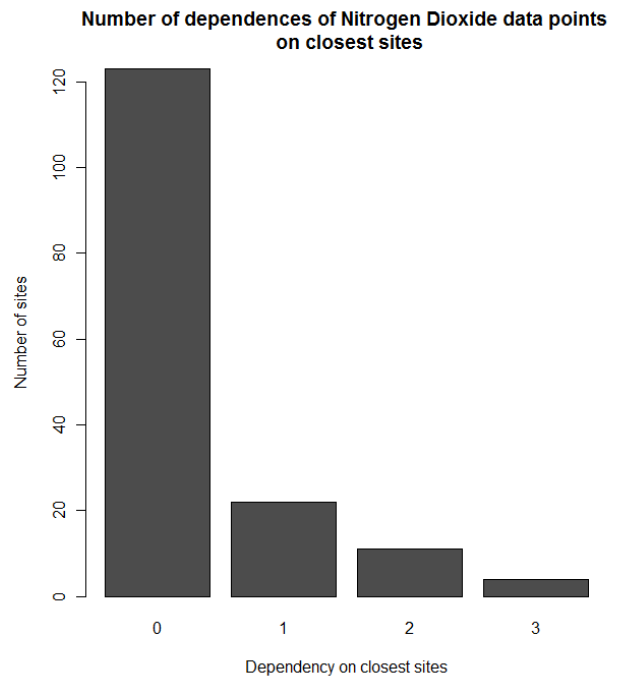


Figure 4b. Histogram showing dependency of NO<sub>2</sub> data points on closest sites

In Figures 4 and 5, a value of 0 indicates that the reading at that particular site is only dependent on previous readings at the same site, while a value of 1 indicates that the measurements are dependent on the site itself and its first closest neighbour, and so on for the other values.

The assumption that dispersal of pollutants is equidistant and therefore one source of pollution in one area has an effect on the neighbouring areas is not supported here. This is further demonstrated by the overlaying of potential sources of pollution in the main island such as traffic density, industrial estates, power stations and the airport. Notwithstanding that most of the points are located relatively close to each other and to these sources, most readings seem to be independent.

The overall spatially independent behaviour of these pollutants would suggest that there are other, more local factors that are affecting air pollution. Some possible interpretations follow.

- Since there is input from a stable source (e.g. traffic) similar temporal patterns can be observed. However, at another location, the source input levels may change (for example, less traffic) and therefore the behaviour of that point, even though it is relatively close, is independent. This is most evident in the area northwest of the Grand Harbour (marked A in Figures 5a and 5b). This is reasonable since in the Maltese urban environment the urban density, urban fabric and traffic change considerably over a relatively short distance.
- A few points experience higher spatial dependencies. These are marked with the letters B and C in Figure 5a. In these cases we note that (i) the pollution values at some of these locations are relatively low, thus affecting the accuracy of the modelling procedure (area marked B) (ii) there are very similar environmental conditions (traffic and urban density) affecting the sites (area marked C).
- These results are validated by MEPA's approach adopted for the location of diffusion tubes. MEPA selects two to three sites per locality, one of which is a traffic site and the other/others are background sites (without traffic).

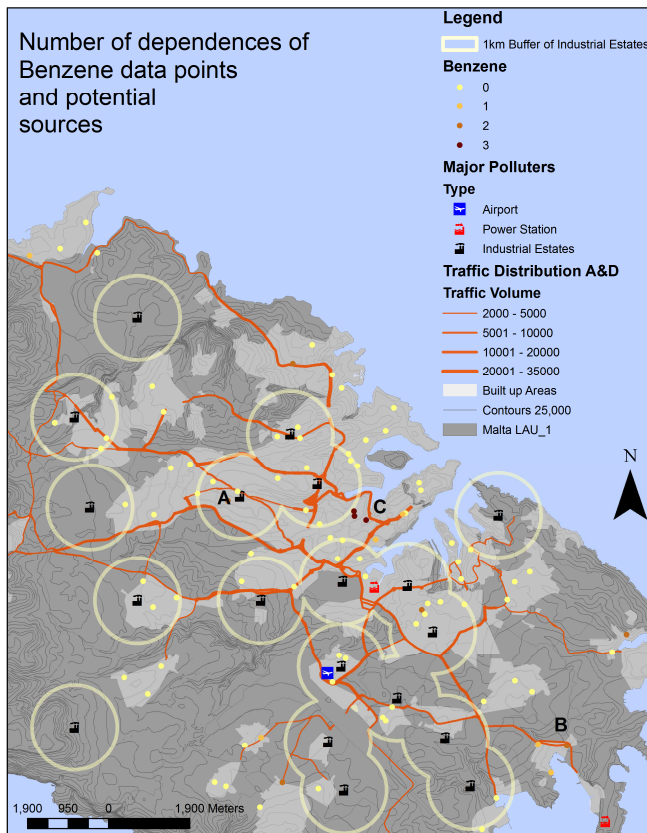


Figure 5a. The number of dependent sites for each location for Benzene.

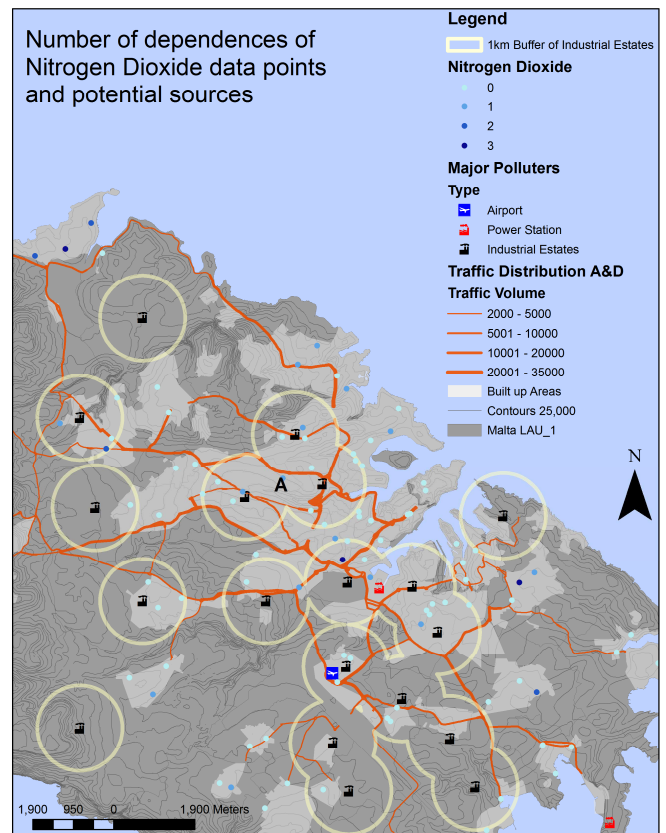


Figure 5b. The number of dependent sites for each location for NO<sub>2</sub>

Future work will focus on introducing measured pollution sources to the mathematical model to verify the dependency of the pollution readings on these sources.

## 5. Acknowledgements

The research work disclosed in this abstract is funded by the Malta Government Scholarship Scheme (MGSS).

## 6. References

- Chatfield C, 2004, *The Analysis of Time Series*. Chapman & Hall / CRC, USA.
- De Luna X and Genton MG, 2005, Predictive Spatio-Temporal Models for spatially sparse environmental data. *Statistica Sinica*, 15: 547-568. Available online at <http://www3.umu.se/stat/personal/xavier.deluna/papers/stpredict.pdf>. Last accessed 16 March 2011.
- Government of Malta, 2002, Johannesburg Summit 2002, Malta Country Profile.
- Ljung L, 1999, *System Identification – Theory For the User*. Prentice Hall, USA.
- MEPA, 2010, *Air: Sources and Effects*. Malta Environment and Planning Authority, Malta. Available online at <http://www.mepa.org.mt/air-sources>. Last accessed 15 October 2010.
- NSO, 2009, *Motor Vehicles: Q4/2009* News Release. 20 January 2010. 008/2010. Available online at <http://www.nso.gov.mt/statdoc/document file.aspx?id=2669>. Last accessed 15 January 2010.

- Office of the Prime Minister, 2010, *Air Quality Plan for the Maltese Islands*. Prepared by the Malta Environment and Planning Authority, Floriana, Malta. Available online at <http://www.mepa.org.mt/air-publications>. Last accessed 5 January 2010.
- Pfeifer PE and Deutsch SJ, 1980a, A three-stage iterative procedure for space-time modeling. *Technometrics*, 22(1): 35-47.
- Pfeifer PE and Deutsch SJ, 1980b, Identification and Interpretation of First-Order Space-Time ARMA Models. *Technometrics*, 22(3): 397-403.