

# Predicting air quality by integrating a mesoscopic traffic simulation model and simplified air pollutant estimation models

Adriana Simona Mihăiță<sup>1\*</sup>, Mirian Benavides Ortiz<sup>2</sup>, Mauricio Camargo<sup>2</sup>, Chen Cai<sup>1</sup>  
1. DATA61, 13 Garden Str., Eveleigh, 2015, Australia, [simona.mihaita@data61.csiro.com.au](mailto:simona.mihaita@data61.csiro.com.au)  
2. ERPI laboratory EA6737, 8 Rue Bastien Lepage, 54010, Nancy, France

Continuous growth in traffic demand has led to a decrease in the air quality in various urban areas. More than ever, local authorities for environmental protection and urban planners are interested in performing detailed investigations using traffic and air pollution simulations for testing various urban scenarios and raising citizen awareness where necessary. This article is focused on the traffic and air pollution in the eco-neighbourhood “Nancy Grand Cœur”, located in a medium-size city from north-eastern France. The main objective of this work is to build an integrated simulation model which would predict and visualize various environmental changes inside the neighbourhood such as: air pollution, traffic flow or meteorological information. Firstly, we conduct a data profiling analysis on the received data sets together with a discussion on the daily and hourly traffic patterns, average nitrogen dioxide concentrations and the regional background concentrations recorded in the eco-neighbourhood for the study period. Secondly, we build the 3D mesoscopic traffic simulation model using real data sets from the local traffic management centre. Thirdly, by using reliable data sets from the local air-quality management centre, we build a regression model to predict the evolution of nitrogen dioxide concentrations, as a function of the simulated traffic flow and meteorological data. We then validate the estimated results through comparisons with real data sets, with the purpose of supporting the traffic engineering decision-making and the smart city sustainability. The last section of the paper is reserved for further regression studies applied to other air pollutants monitored in the eco-neighbourhood, such as sulphur dioxide and particulate matter and a detailed discussion on benefit and challenges to conduct such studies.

**Keywords:** *Mesoscopic traffic simulation, air pollution, concentration estimation, eco-neighbourhood*

## 1. Introduction

**Context:** As many urban areas have known a rapid increase, it is expected that by 2050, almost 70% of the population will live in cities [1]. According to TomTom [2], some of the most congested cities around the globe are: Mexico City (66%), followed by Bangkok (61%), Jakarta (58%), Chongqing (52%) and Bucharest (50%). The large number of vehicles in densely populated areas contributes to an increasing deterioration of the air quality due to motor vehicle emissions. According to 2012 report of the U.S. Environmental Protection Agency (US EPA) 61% of the total carbon monoxide concentrations (CO) and 35% of total nitrogen oxide concentrations were produced by highways [3]. More recently, in 2015, the organisation has further organised the International Decontamination Research and Development Conference for exploring current issues and future directions in detection, environmental emergency response, risk communication, sampling, treatment, decontamination and decision support tool development [4].

**Health impact:** Various studies which have been recently published showcase the severe health impact that air pollution can have on public health. In 2017, the European Environment Agency (EEA) had released a detailed analysis based on official data from more than 2500

monitoring stations across Europe [5]. Although the data shows that air quality in Europe is slowly improving due to past/current policies and advanced technological developments, the PM<sub>2.5</sub> concentrations were responsible for an estimated 428000 premature deaths annually. The report indicates that in 2015, about 19 % of the EU-28 urban population was exposed to PM<sub>10</sub> above the EU daily limit value, 53% was exposed to concentrations exceeding the stricter WHO AQG value for PM<sub>10</sub>, while 30% was exposed to higher O<sub>3</sub> concentrations above the EU target value threshold. The report also outlines that the road transport sector was the largest contributor to the total NO<sub>x</sub> emissions, as 89% of all values above the annual limit value were observed around traffic stations. Traffic is a major source of NO which reacts with O<sub>3</sub> to form NO<sub>2</sub>. Furthermore, 98% of the stations with values above the annual limit were urban or suburban. Therefore, reductions in NO<sub>2</sub> concentrations need to be focused on traffic and urban locations because its emissions are close to the ground and are distributed over densely populated areas.

Taking it to an international level, the 2017 official report of OMS (Organisation Mondiale de la Santé) has estimated to 7 million per year the number of total deaths caused by bad air quality, especially in Asian countries [6]. The causes listed can vary from high energy consumption,

intense industrial manufacturing, deforestation, sand storms, with a deep emphasis on road traffic congestion.

**Economic impact:** This health scourge causes a huge loss of revenue for the global economy: 225 billion dollars loss of revenue per year according to the 2016 report of the World Bank, which wanted to raise high awareness on the financial burden of poor air quality around the globe [7]. Global welfare losses from exposure to ambient PM<sub>2.5</sub> rose by 63% from 1990 to 2013, reaching \$3.55 trillion—a reflection of worsening exposure in many fast-growing countries together with a higher marginal cost for fatality risks which is normally associated with rising incomes.

Without further sound policies and annual strict measures, air pollution could cost the global economy 2,600 billion dollars by 2060. These staggering figures come from a study led by the Organization for Economic Cooperation and Development (OECD) on “The Economic Consequences of Outdoor Air Pollution” published in [8]. Over the next 50 years, with rising living standards and rising demand for energy, emissions of most pollutants will increase significantly. In particular, emissions of nitrogen oxides and ammonia, which contribute to the formation of PM<sub>2.5</sub> could almost double. The increasing concentrations of particulate matter and ozone in the air will also lead to an increase in cases of illness requiring hospitalization, and implicitly in health costs and sick days remuneration. OCDE sets the alarm that from 21 billion dollars in 2015, worldwide health expenditure could reach 176 billion dollars by 2060.

Therefore, we believe that by predicting the impact of air quality one would gain quicker insights on the pollution evolution in cities, allowing councils and regional air quality centres to trigger alarms and adopt specific measures to reduce pollution concentrations.

**Proposed solution for air quality prediction:** The complexity of the air pollution lies in its extent and the large number of factors changing its behaviour, making it even more difficult to implement measures for protecting the citizens. Therefore, a systematic evaluation of efficient traffic management strategies to reduce traffic congestion requires effective transportation simulation modelling in order to efficiently assess operational traffic performance and pollution impact at different spatial and temporal resolutions (e.g. network, corridor and segment levels, second-by-second, peak hours, 24-h, multiple days) [9]. Mesoscopic and microscopic traffic simulation tools are widely used for obtaining very detailed traffic analysis and provide traffic dynamic insights second-by-second. Although they require a high amount of data and can be computationally intensive, they can be efficiently combined with various tools for air pollution monitoring; a recent comparison study of various integrated traffic simulation and air pollution tools is provided in [10]. For detailed insights regarding all traffic and air quality

modelling approaches, the reader can refer to [11]. While the construction of the traffic simulation or air pollution models can follow standard modelling techniques, integrating both modules together requires an increased complexity and data analysis.

In this article we build a simplified nitrogen dioxide (NO<sub>2</sub>) estimation model from traffic and meteorological data which validates the concept and efficacy of an integrated mesoscopic traffic simulation model. In the next section we introduce the context of our project and the organisational framework of our work. We further conduct a data profiling analysis in Section 2.3 and provide more information on background concentrations registered in the regional area of our study. Section 2.4 details the steps for building the 3D mesoscopic traffic simulation model using real data sets received from the Grand Nancy traffic management centre. Although various research studies consider other factors, such as speed and acceleration to model and predict the air pollution concentrations, these data sets are not always available in many urban areas with limited traffic detectors or monitoring capabilities, this being the case for Nancy Grand Cœur as well. The results of the simulation are used to implement the simplified estimation model for NO<sub>2</sub> (detailed in Section 3), by considering as well meteorological data received from the local air monitoring association Air Lorraine. In Section 3.2 we further conduct a seasonality study over the traffic volume and validate the estimation results over real data. Section 4 of the paper is used for further investigations on other air pollutants available for analysis during the same study period, such as SO<sub>2</sub> and PM<sub>10</sub>. Lastly, Section 5 presents the conclusions and future perspectives are closing the paper.

## 2. Integrated traffic and air pollution simulation model

### 2.1 Context

Maintaining a sustainable development in congested cities has become a priority for local authorities, in order to provide flexibility, intermodal transportation systems and green mobility for its citizens. As a response to these problems, the “eco-neighbourhoods” have become the perfect test-bed for new technologies in the context of a smart city. With the urban project “Nancy Grand Cœur” (NGC) [12], the Grand Nancy council wants to rehabilitate the boundaries of the historic train station of Nancy (hosting almost 9 million passengers every year), and the surrounding belonging to the city centre, as represented in Figure 1a).

This ecological urban project is intended to be delivered by 2020, and the objectives for the central train station of the city are manifold: new green mobility, traffic

regulation, reconciliation between historical and modern neighbourhoods of the city, environment quality and green public spaces, reduced energy consumption, comfortable homes and offices. As the neighbourhood is currently suffering structural reconfigurations that are meant to reduce congestion and increase fluidity, no study has been undertaken so far to analyse the impact of these traffic reconfigurations on the air pollution inside the eco-neighbourhood.

The traffic analysis in this paper is a continuation of our previous study of the neighbourhood presented in [13] and currently under review in [14], in which we build the 3D mesoscopic traffic simulation model of the eco-neighbourhood in FlexSim, and propose an evolutionary algorithm for optimising the traffic plan in the most congested crossroads (as marked in Figure 1b)). In this paper, we only focus on the Viaduct Kennedy, located near the train station area, which contains the fixed air pollution station of the local air monitoring association Air Lorraine (Figure 1c)).

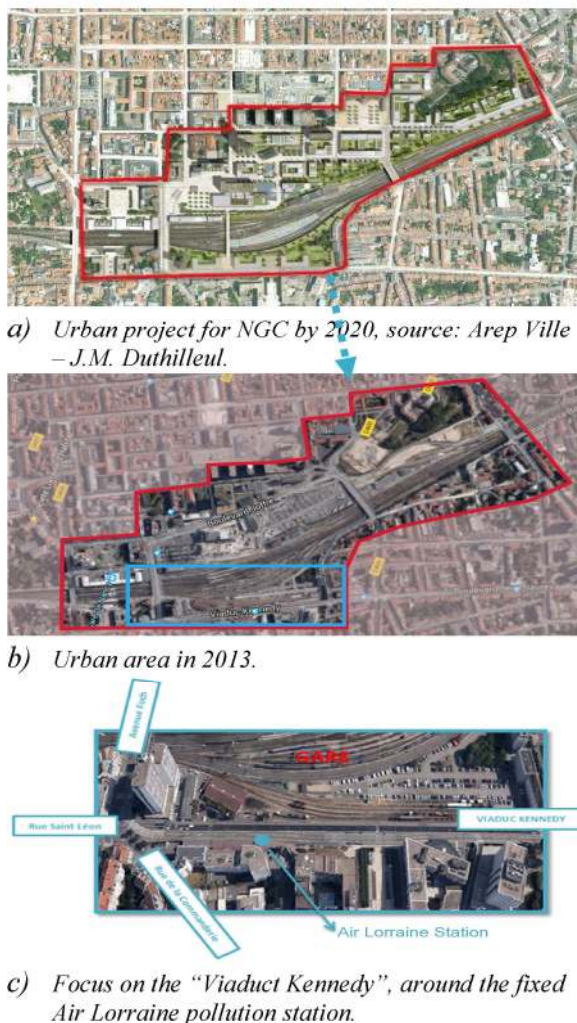


Figure 1 Case study of the eco-neighbourhood "Nancy Grand Cœur".

The main objective is to conduct an initial study in order to simulate, predict and analyse simultaneously the traffic flow and the pollution concentration in this highly congested area. The choice of focusing on this area is also related to the air quality workstation of Air Lorraine, which provides accurate and real-time information for model validation and testing.

## 2.2 Organisational framework

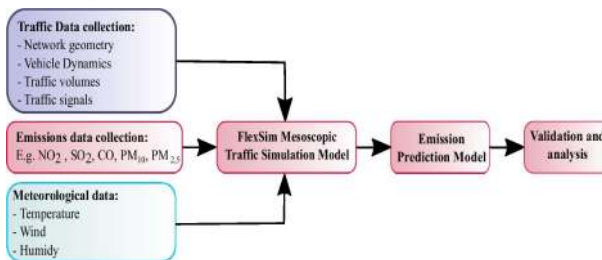


Figure 2 Organisational framework of the article.

Figure 2 summarises the organisational framework of the work proposed in this paper, which contains the traffic simulation model and the emission prediction model. We start by collecting and analysing the real data sets available for our study. For building the 3D traffic simulation model in FlexSim, Grand Nancy provided: a) the network geometry of the Viaduct Kennedy, b) the hourly traffic volumes during one month period (January 2015) and c) the traffic signal control plan of the crossroads. For analysing the air pollution registered in the neighbourhood, the local air quality association Air Lorraine provided the hourly concentrations registered by the fixed air monitoring station (as marked on Figure 1c)), as well as the meteorological indicators during the chosen study period such as wind, temperature and humidity. As 56% of the nitrogen dioxide in the air is known to be caused by road transportation [15], for this study we mainly focus on NO<sub>2</sub> concentrations, which are not as sparse and volatile as other pollutants such as SO<sub>2</sub> and PM<sub>10</sub>, which are discussed in last section of the paper. By using the provided real data sets we build the 3D mesoscopic traffic simulation model, which allows us to obtain statistical measures of the traffic model, such as mean number of cars/average stay-time (average travel-time on each road section), and also graphically visualise the evolution of meteorological conditions and NO<sub>2</sub> concentrations. Using the outputs of the simulation model, we construct a simplified prediction model for the NO<sub>2</sub>, which is later validated through analysis and comparison tests with real-data sets.

## 2.3 Data profiling

### 2.3.1 Background concentrations in Lorraine

Before entering the air pollution analysis, we provide some insights over the background concentration and

extended air quality measures taken in the Nancy region in 2015 by the Air Lorraine agency which was officially published one year later in [16]. Their main motivation that year was to evaluate the air quality in areas exposed to traffic circulation by using: a) a mobile station measuring NO<sub>2</sub>, PM<sub>10</sub> and PM<sub>2.5</sub> and b) fixed passive diffusion tubes for measuring volatile organic compounds (specifically benzene, toluene, ethylbenzene, m+p-xylene and o-xylene), and NO<sub>2</sub> in the ambient air. Their cylindrical tubes contain a reactive agent for trapping the pollutant to be studied. The study extended over 4 different periods and for a total of 103 days in 2015, and had various observations concerning the background concentrations, especially when compared to previous years. Main findings of the report are the following:

- a) Air Lorraine maintains an inventory of atmospheric emissions of pollutants and greenhouse gases resulting from the different sectors of activity in France. The dominant sectors in terms of polluting emissions in our study area are the manufacturing industry-waste-construction treatment, road transport, and the residential sector. According to [16], sulphur dioxide (SO<sub>2</sub>) is almost exclusively issued by the manufacturing-waste-construction industry (98%). Same for CO, which goes up to 77%. Most of the nitrogen oxides come from road transport (56%) and the manufacturing-waste-construction industry (29%). Similar for PM<sub>10</sub> but in different proportions: 33% comes from the manufacturing industry and 24% from road transportation. PM<sub>2.5</sub> comes mainly from the residential sector (32%), road transport (30%) and the manufacturing industry-waste-construction (18%).
- b) The annual maximal limits for background concentrations in 2015 were settled around: 40 µg/m<sup>3</sup> for NO<sub>2</sub>, 40 µg/m<sup>3</sup> for PM<sub>10</sub> and 10 µg/m<sup>3</sup> for PM<sub>2.5</sub>.
- c) The mobile station circulating in daily traffic indicated low annual means of 26 µg/m<sup>3</sup> for NO<sub>2</sub> and 23 µg/m<sup>3</sup> for PM<sub>10</sub> but higher values of 15 µg/m<sup>3</sup> for PM<sub>2.5</sub>. However, the regional fixed air pollution stations indicated average annual values of 45 µg/m<sup>3</sup> for NO<sub>2</sub> which are higher than the accepted annual limit. According to Air Lorraine, these limits are slightly higher than measures taken in the previous years, especially in the city centre, where the traffic flow has increased consistently in latest years.
- d) The passive tubes monitoring NO<sub>2</sub> revealed that 8 out of 21 measuring points exceeded the annual limit value of 40 µg/m<sup>3</sup> and that most of them have significant spatial disparities ranging between 13 and 63 µg/m<sup>3</sup>. The highest concentrations were generally observed in narrow canyon-type streets circulated by a large number of vehicles. The measurements near our case study location have revealed mean values of 51 µg/m<sup>3</sup> in February 2015, which is a period following immediately our case study evaluation for January 2015.

- e) For PM<sub>10</sub>, the threshold of 50 µg/m<sup>3</sup> was exceeded for 2 days (16<sup>th</sup> and 17<sup>th</sup> of February 2015). These exceedances were also observed at other fixed stations in the Lorraine region, which triggered the regional procedure for information and recommendations related to pollutant emissions, combined with favourable weather conditions for the accumulation of this pollutant in the ambient air.

These findings have been further used to analyse and evaluate the results obtained in our case study, which are detailed in the next section.

### 2.3.2 Data profiling for our case study

The first step in understanding the traffic and air pollution behaviour is to analyse the initial data we have received from Grand Nancy and Air Lorraine, during the month of January 2015.

Table 1 provides a summary of the meteorological conditions registered during our case study analysis. This period was characterized by north-easterly winds blowing in early January and almost daily snowy showers with a cold regime (touching minimal temperatures of -4.7°C), followed by an increase in temperature to a maximum of 13.1°C in mid-January, only to drop again to the end of the month. Overall temperatures registered an average of 2.98°C while wind reached in average 11 km/h with a maximum of 37 km/h on 13<sup>th</sup> of January 2015. The wind rose from Météo-France Nancy-Essey station (30 years of measurements) indicates that most of winds in this region are weaker but colder in winter. The wind rose from the fixed measurement point of Air Lorraine located in Nancy-West (Brabois) for the period 2005-2015 indicates the same trends in the provenance of the winds.

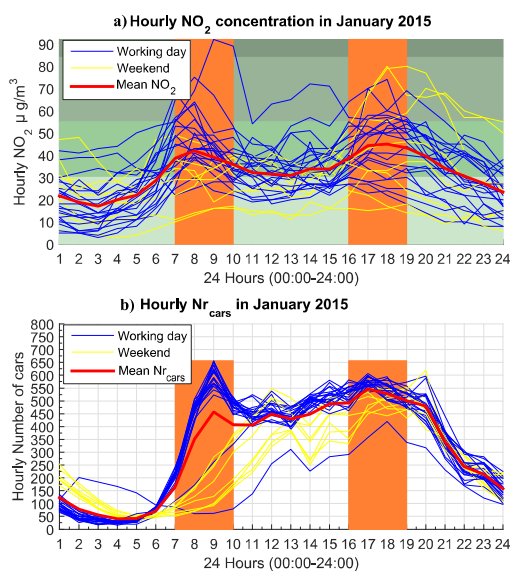
The meteorological conditions seem to be favourable to an accumulation of pollutants in the ambient air, especially as January is a month of return-to-work activity after long holidays which registers normally an increased level of domestic heating, road traffic and industrial activity.

	Min	Max	Average
Temperature [°C]	-4.7	13.1	2.98
Wind [km/h]	0	37	11
Humidity [%]	0.38	0.98	0.85
Wind direction	North-East		

**Table 1 Summary of meteorological conditions during our study period.**

Figure 3a) presents the evolution of NO<sub>2</sub> concentrations for 24 hours; we represent every working day of the month in blue, weekends in yellow, and in red the evolution of the daily average NO<sub>2</sub> concentrations. As the behaviour of the pollutant is highly influenced by the traffic in the city, in Figure 3b) we represent as well the

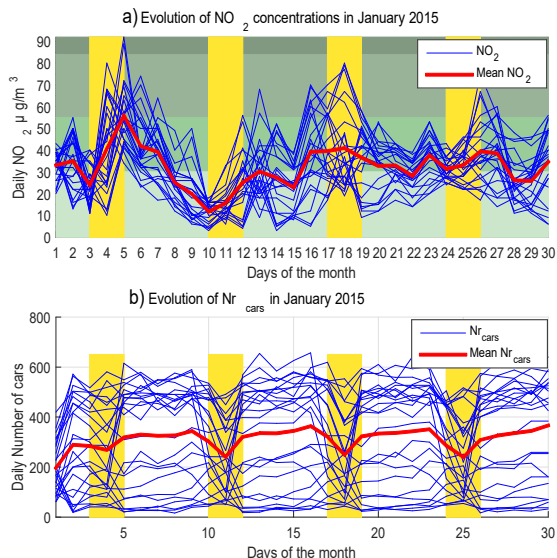
hourly evolution of the traffic volume, which we denote  $Nr_{cars}$ . As the two graphics show, during the morning and afternoon traffic rush hours (07:00-09:00 and 16:00-19:00), high levels of  $NO_2$  concentrations are dispersed in the air as registered by the air pollution station. An interesting remark is that in two of the  $NO_2$  series corresponding to weekend days, we observe higher  $NO_2$  levels ( $80 \mu g/m^3$ ) than in working days ( $70 \mu g/m^3$ ). Overall the mean hourly concentrations for this period seemed to reach a maximum peak of  $45.6 \mu g/m^3$  during the evening rush hour around 5 pm. This level exceeds the annual average limits of  $40 \mu g/m^3$  and by taking a closer look to the hourly registered values once can notice that these records sometimes exceeds even twice the maximal approved limit; this finding reflects a high situational awareness that needs to be triggered in advance for the current regional area transiting towards a more ecological and sustainable environment.



**Figure 3 a) Hourly  $NO_2$  concentrations and b) hourly traffic volume from January 2015.**

Figure 4a) and 4b) presents the evolution of traffic and  $NO_2$  concentrations over the whole month. The maximal  $NO_2$  concentration took place on the 5<sup>th</sup> of January 2015, when it reached  $90 \mu g/m^3$ , an acceptable value according to the air quality index of the European Union [17], but bigger than regional annual accepted limits according to Air Lorraine. By looking at the weekend 10<sup>th</sup> to 12<sup>th</sup> of January 2015, one would remark a dramatically  $NO_2$  reduction compared to the previous weekend, and initially conclude to a similar trend during other weekends; the trend does not seem to repeat itself, as the weekend 17<sup>th</sup> to 19<sup>th</sup> of January 2015 signalled another increase in  $NO_2$ , although the traffic volume was similar to the previous weekend. Although in most days the recorded average  $NO_2$  concentrations fall under  $40 \mu g/m^3$ , various weekends and

beginnings of working week registered records which sometimes reached  $60 \mu g/m^3$ , possibly due to return-to-work activity in the city.

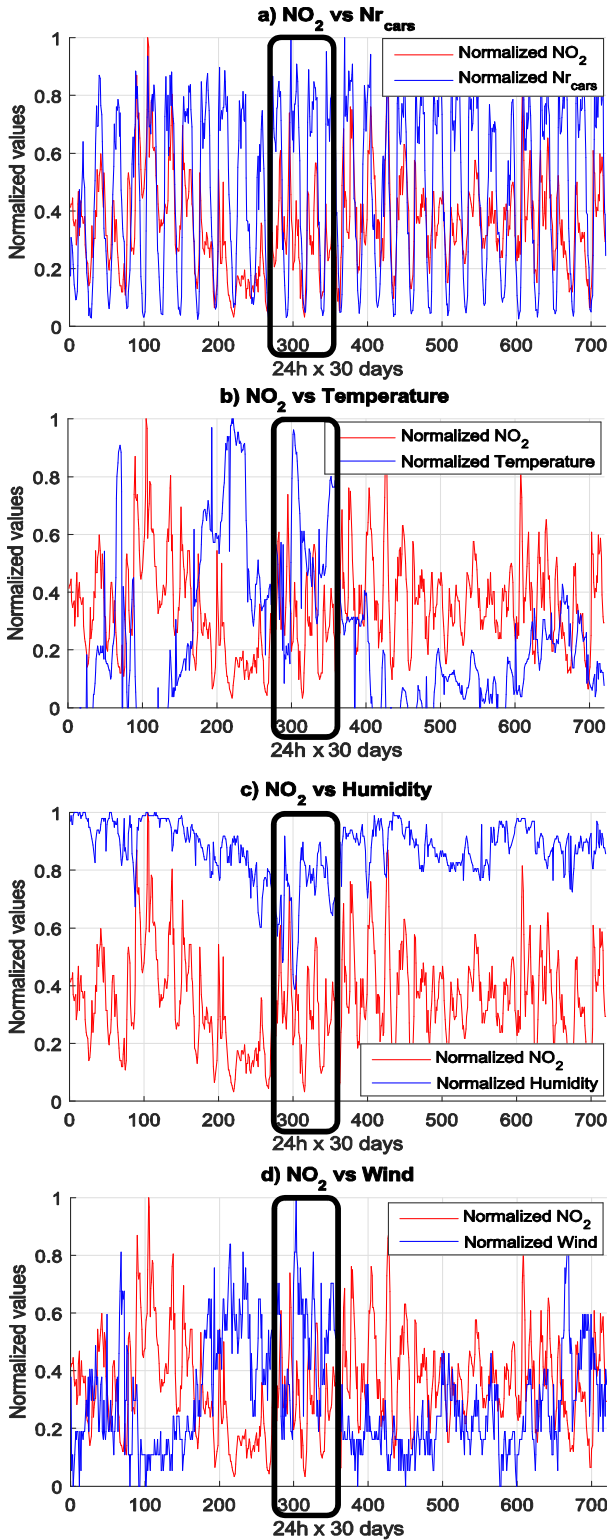


**Figure 4 a) Monthly evolution of  $NO_2$  concentrations and b) Monthly traffic counts.**

The evolution of  $NO_2$  [ $\mu g/m^3$ ] seems to be as well influenced by other external factors, such as wind [ $km/h$ ], humidity [%] and temperature [ $^{\circ}C$ ], which we consider for the current analysis, as shown in Figure 5. For easing the visualisation of variables with different unit measures, we plot only the normalized values (ex. for temperature:  $(T - T_{min}) / (T_{max} - T_{min})$ ).

By closely looking at the highlighted region in Figure 5d) we observe that low  $NO_2$  concentrations usually occur when either the wind is strong, humidity is high or there are few cars on the streets. Conversely, high  $NO_2$  concentrations appear when either wind and humidity are low, or there is a higher number of cars on the streets (compared data fields 300 marked in each subfigure in Figure 5). Figure 5b) showcases the antagonistic evolution between temperature and  $NO_2$  which can be observed especially in the second part of the study period: when the temperature was low, the  $NO_2$  concentrations were higher. The phenomena is more obvious in the beginning of our study period (5<sup>th</sup> of January being the first day back to work after winter holidays), when temperatures reached their minimum of  $-4.7^{\circ}C$  (this means normalized valued around  $0^{\circ}C$ ) while  $NO_2$  concentrations reached their highest peak touching a maximum of  $90 \mu g/m^3$ . This behaviour has further triggered the correlation analysis presented in the following section.





**Figure 5** Normalized concentrations of  $\text{NO}_2$  versus  $Nr_{cars}$ , Temperature, Humidity and Wind.

To further investigate the stochastic influence of the climate and human factors (domestic heating, road traffic, industrial activity, etc.) over the air pollution, we conduct a correlation analysis between these four parameters (temperature  $T$ , Wind  $W$ , humidity  $H$  and number of cars  $Nr_{cars}$ ) and the  $\text{NO}_2$  concentration, for January 2015, as presented in Table 2. The loose Spearman correlation (0.509) between the  $\text{NO}_2$  concentrations and the number of cars, indicates that other factors (apart from the number of cars) seem to influence pollution, as previously mentioned. Some of the factors we have considered to adopt the Spearman correlation coefficient are the following: a) the variables we have studied are not normally distributed and the relationship between them is not linear b) the data sets seemed to present high outliers which are hard to interpret especially under the given meteorological circumstances and c) the data we have received is continuous and provided records every hour.

Although Wind and Temperature seem to be highly correlated (0.767) for our analysis (probably due to winter conditions: colder winds can cause the temperatures to drop even more), they both have a negative correlation with nitrogen dioxide. This confirms the complexity of the model and the need to further build a more complex regression model which would help predict the changes in the air pollution and help traffic planners and local authorities to test various reconfiguration traffic scenarios.

Indicator	$Nr_{cars}$	$T$	$W$	$H$	$\text{NO}_2$
$Nr_{cars}$	1	0.12	0.194	-0.185	0.509
$T$	-	1	0.767	-0.55	-0.441
$W$	-	-	1	-0.549	-0.492
$H$	-	-	-	1	0.281
$\text{NO}_2$	-	-	-	-	1

**Table 2.** Spearman correlations between traffic and various air quality indicators.

## 2.4 Simulation model

At the intersection between macroscopic and microscopic models, the mesoscopic modelling applies principles of non-equilibrium statistical mechanics to kinetic theory to model the traffic flow. Considering the amount of literature that has been generated during the last few decades, it seems to us that there is not a unanimous consensus as to what exactly constitutes mesoscopic traffic flow models [18]. In general, the most popular approach when it comes to mesoscopic modelling is to group nearby vehicles together with respect to one of their traffic flow characteristics, e.g., their space-mean speed. Instead of having to perform detailed updates of all vehicles' speeds and positions, the cluster approach

allows to treat these vehicles as a set of groups (called either clusters, cells, packets, or macro-particles), which are then propagated downstream without the need for explicit lane-changing manoeuvres (leading to the coalescing and splitting of various colliding and separating groups). Other approaches are also available based on either headway distribution models or gas-kinetic model as detailed in [19], while a comprehensive state of art for various traffic flow models can be found in [20]. As well, for details on available traffic simulation tools which can be used at various modelling levels, the reader can redirect to the work of Barcelo [21].

There are also some traffic simulation models which are at the boarder of different modelling levels and which integrate various phenomena non-existent in most similar models, as detailed in [22]. For example, KRONOS, often classified as a macroscopic model, has the possibility to simulate lane changing, merging and diverging behaviour, which could be also considered a mesoscopic model. INTEGRATION which is normally a microscopic model, does not explicitly consider all the details of lane-changing and car-following behaviour, but relies on the aggregated speed-volume traffic interaction which is the attribute of macro models. The AVENUE simulation model is another example which represents traffic as a continuum flow, but at the same time it moves discrete vehicle images for the convenience of route choice calculation and for handling conflicts of vehicles in crossroads. We consider that the exploration of different traffic simulation models at the mesoscopic level is still open for new debates with various integrations between different modelling layers, based on the needs and specification of the modelling project.

Due to the size and nature of the simulation requirement, the current traffic simulation model has been constructed at a mesoscopic level in FlexSim, a discrete-event simulator in which groups of vehicles have been assigned proportionally to a set of paths depending on different time intervals during the simulation period. This is equivalent to the MSA (Method of Successive Averages) procedure of the Dynamic User Equilibrium in Aimsun [23], which redistributes the flows among the available paths in an iterative procedure. However, this requires the number of iterations to be defined by the O-D pairs and the time interval of the simulation.

Traffic flow inside the NGC neighbourhood is highly impacted by the central train station of the city and its particular suspended structure as previously shown in Figure 1c). The study zone we are considering for the traffic simulation contains the junction made by *Viaduct Kennedy* with *Rue Saint Léon*, *Avenue Foch* and *Rue de la Commanderie*.

The data used to create vehicles was received from the Grand Nancy traffic management centre and represents a time-sliced profile demand for each 15-min time interval

during morning peak hours; from this data we have further computed dynamic turning rates between all road sections during the simulation time interval. The simulator does not distinguish nor trace the individual behaviour of each vehicle in the system but can specify the probabilistic turning rates of small groups of vehicles moving together in the network. Groups of vehicles will travel on each road section with a random speed following a uniform distribution which is bounded by the speed limit of each network link. Every connection inside the model influences the travel time computation for the next road section, as the average travel time is influenced by the flow, the occupation of the segment or its capacity.

The traffic simulation modelling in FlexSim comprises several steps such as: network geometry importation (AutoCAD files), importation of the 3D environment built on this geometry (see a 3D visual representation of the model in Figure 6), 15-min time-sliced profile demand generation and path assignment, traffic signals groups creation which includes a priority feature for the tramway (see Figure 7b), as well as traffic control plans for pedestrian crossing synchronisation (see Figure 7a).



Figure 6 FlexSim 3D simulation model of the Viaduct Kennedy in FlexSim.

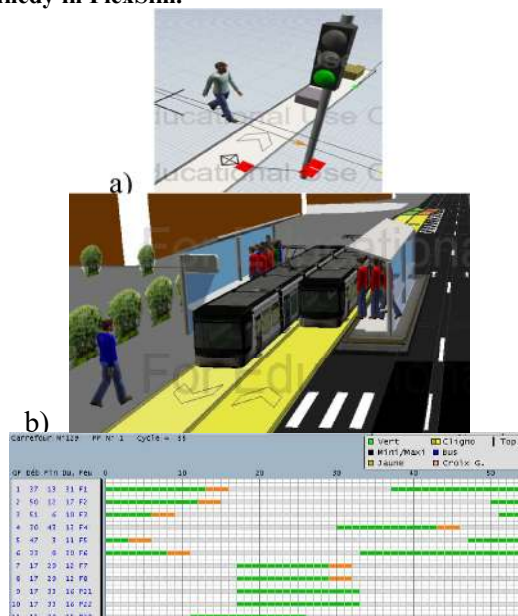


Figure 7 a) Traffic light plans representation b) Tramway stop in FlexSim c) traffic control plan example used for red-yellow-green light.

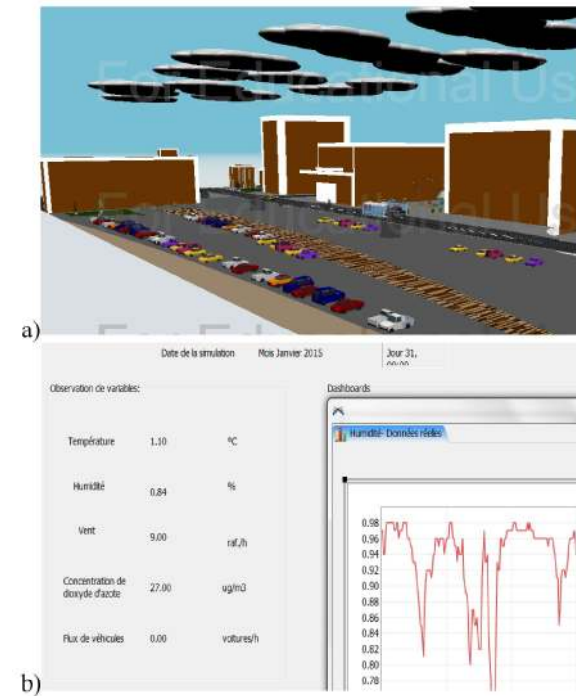
By traffic control plan we denote the red-yellow-green cycles of all the traffic lights inside the simulation mode, as shown in Figure 7c). For example, the traffic light F<sub>1</sub> has a starting green-light at second 37 and ends at second 13 in the 55-second cycle length, therefore lasting for 31 seconds. The yellow light lasts for only 3 seconds while the red-light will start at second 17 and end at second 37 after which the green light will be activated again. As an observation, a traffic control plans is conceived for all normal traffic lights and also for all pedestrian lights inside the considered intersection.

In our simulation model, the vehicles circulate on road sections which have predefined speed limits. For all vehicles circulating together on a specific road section, the group follows a variable speed allocation, which can vary between the minimum of the vehicle speed and the maximum speed limit of each path. If the path that the vehicle is on is a “Passing connection” in the model, then the vehicles will accelerate or decelerate to the appropriate speed once they get on this path. If it is non-passing, however, then vehicles will immediately change their speed to the appropriate speed, without using acceleration or deceleration.

Due to the stochastic nature of the mesoscopic model, several iterations need to be run in order to obtain accurate results. The method suggested by [24] is to run successive simulations until the average mean and standard variation of the average travel-time (or the average number of cars) fall within an acceptable confidence interval calculated in relation to the standard t-distribution. Using this procedure in accordance with a confidence interval of 95%, the number of runs indicated approximately 13 runs per scenario. Given the importance of the result accuracy, we decided to conduct 15 simulation runs for each time-period scenario. The final simulation outputs have been used for comparison with real-traffic flows and recalibration when needed. FlexSim can run parallel replications of the simulation model according to the number of available processors. The simulations have been made using an Intel Quad Core i7 (2.4 GHz) computer having 8 GB DDR3 SDRAM memory.

The traffic simulation outputs have been validated through manual field measurements taken by volunteers during the simulation times for morning peak (07:00 AM to 09:00 AM). The R<sup>2</sup> value between simulated and real traffic volumes was 0.8254 which we believe sufficient enough due to possible human errors in counting. As well, we compared our experienced travel-times on 4 main routes during the morning peak with the average travel time of the simulation which indicated a percentage error of 16.7%; this aspect gave confidence in the accuracy of the simulation outputs. Further on, the average traffic volumes have been used in the following step of the integrated simulation and air pollution model.

Apart from constructing a clear insight of the traffic conditions in this area during congested periods, our efforts have been also oriented towards the integration of pollution data and meteorological information in the simulation model, which can be a complementary source of information in the process of traffic and air quality dynamic monitoring. Various air pollution studies use dedicated tools for representing the concentrations of pollutants in a specific area, such as MISKAM (Microscale Flow and Dispersion Model), which has been used for the air pollution study in the eco-neighbourhood Danube, from Strasbourg, France [25]. MISKAM is a three-dimensional microscopic simulation tool which integrates fluid dynamics equations to simulate pollutant concentration as a 3D mesh. Although it can offer precise estimations of various pollutant concentrations, it needs large amount of input data, such as the topography and height of the buildings, the annual traffic emissions computed with Circul'Air [26], meteorological data including the direction and wind speed, etc. The main limitation we encountered for integrating MISKAM and FlexSim outputs was mainly related to the differences in the static and dynamic behaviour of the tools. As our initial objective was to build a rapid and dynamic simulation model which offers an immediate visual insight on the hourly variation of the NO<sub>2</sub> concentrations during January 2015, we integrated the pollution concentrations as entities circulating in the model (as shown in Figure 8a), while the wind, temperature and humidity as visual indicators (Figure 8b).



**Figure 8 a) Visual representation of NO<sub>2</sub> b) Climate indicators dashboard.**



### 3. Regression model of the NO<sub>2</sub>

After validating and building the traffic simulation model, we use the simulated average traffic volumes to build the estimation model of the NO<sub>2</sub> concentrations. We make the hypothesis that all vehicles considered here are contributing to NO<sub>2</sub> concentrations, as no data regarding different types of vehicles circulating in the area is currently available. We start by training a least squares multiple regression model (LSMRM) on the first 20 days of our data set from January 2015, and we further test the results on the temporally holdout 10 days. In order for the prediction model to work accurately, one needs to use continuous data sets for training the model. All variations which are corresponding to weekends and week day have been initially incorporated in the overall model evolution and it will showcase on the final accuracy of the model as represented in Figure 9 and later discussed. As an observation, we have chosen LSMRM after first fitting a simple linear regression model that only slightly improved the performance over a baseline. Using the statistical tool Minitab, we obtain the following regression equation, showing the evolution of the NO<sub>2</sub> concentration as a function of the average number of cars on the Viaduct Kennedy, the temperature, the wind and the humidity, which we denote as predictors in this section:

$$NO_2 = 28.43 + 0.05 Nr_{cars} - 1.04T - 0.91W + 2.04H \quad (1)$$

As an example from our winter data set, if we consider that no cars are on the road ( $Nr_{cars} = 0$ ), and the Temperature is 2.9°C, Wind = 11.28 km/h and Humidity = 85.47%, then we obtain an NO<sub>2</sub> concentration of 29.74 [ $\mu\text{g}/\text{m}^3$ ]; this falls under the overall average concentration which was 32.17 [ $\mu\text{g}/\text{m}^3$ ]. Table 3 indicates whether the predictors we have used are independent variables, as shown by the variance inflation factor (*VIF*) [27]. A predictor with a *VIF* indicator superior to 5 would indicate that it is highly correlated to another predictor. For our study the current results show that the considered predictors are non-correlated, as their *VIF* values are inferior to 5. In order to analyse the weight and influence of each predictor over the NO<sub>2</sub> concentration, we analyse the *P-value* which is based on the *T-test*. *P* determines the appropriateness of rejecting the null hypothesis in a hypothesis test and ranges from 0 to 1. According to [28], a predictor with a *P*-value superior to 0.05 has a high chance of not influencing the estimated parameter, which is the case of the Humidity variable, with a *P*-value of 0.68. This result indicates that we can stop considering the Humidity variable for further studies related to NO<sub>2</sub>. Other pollutants have different behaviour and similar analysis need to be conducted before evaluating the importance that each predictor/feature/variable plays in the overall model accuracy. If only after running a similar analysis one

discovers a predictor with a “*p*-value” higher than 0.05 then this predictor can be dropped from the analysis.

The intercept value (28.43) in equation (1) indicates the presence of other external factors which we haven't included in this study and which might have a strong influence over the NO<sub>2</sub> variation, such as topography, height of the buildings, wind direction, etc. The regression test also returns a value of  $R^2 = 0.67$  for the coefficient of determination, which indicates that the data fits in a good proportion to the regression model.

PREDICTOR	COEF F.	T	P	VIF	[MIN; MAX]
Constant	28.43	5.63	0.00	-	-
$Nr_{cars}$	0.05	21.96	0.00	1.06	[0;657]
T[°C ]	-1.04	-5.88	0.00	3.19	[-4.7;13.1]
W [km/h]	-0.91	-8.82	0.00	3.30	[0;37]
H[%]	2.04	0.40	0.68	1.71	[38;98]

**Table 3 NO<sub>2</sub> Regression analysis results for the first 20 days of January.**

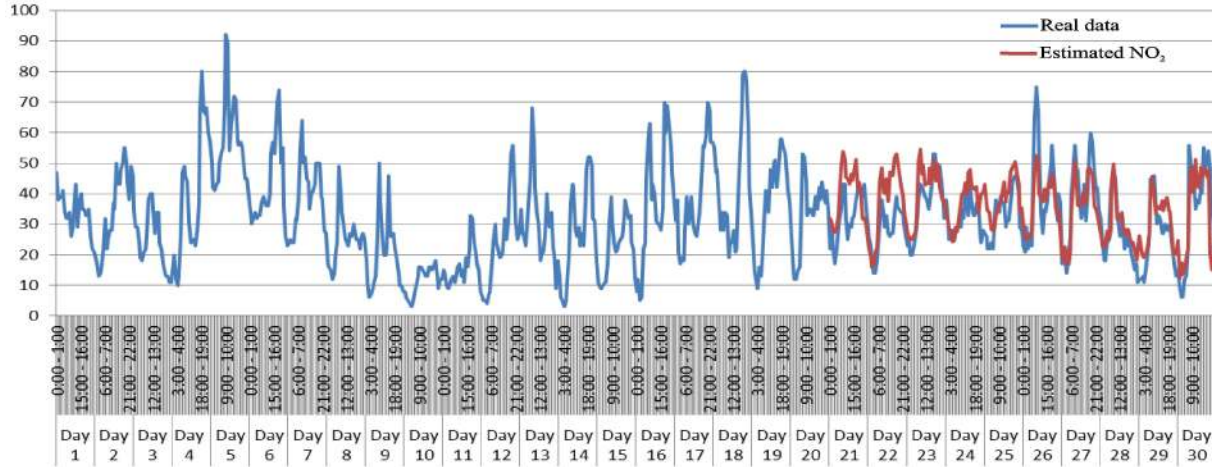
#### 3.1 Validation of the prediction model

Using equation (1) we forecast the NO<sub>2</sub> concentrations over the last 10 days of the data set, and compare them to the initial real data set, as shown in Figure 9. In this figure, the evolution of the NO<sub>2</sub> concentrations is presented in a time series manner where each data point represents the hourly NO<sub>2</sub> concentration, starting from the beginning of the data records, at midnight 1<sup>st</sup> of January 2015. Due to the small scale of the figure on the page, the labels on the Ox axis are summarized by day, although the series contains all hourly records. The last ten days contain both the real time series and the predicted one, which shows that our model is capable of accurately predict the evolution in time of the NO<sub>2</sub> concentrations, regardless of the type of day (weekend or weekday). A more detailed seasonality analysis is further provided in Section 3.2.

According to [29], there are various indicators for measuring the accuracy of a forecast, such as the Mean Absolute Percentage Error (MAPE), Mean absolute Deviation (MAD), Mean squared deviation (MSD) or Mean Squared Errors (MSE). For the error accuracy and ease of calculation we use the MAPE, which is defined as:

$$MAPE = \frac{1}{n} \sum_{k=1}^n \left| \frac{F_k - Y_k}{Y_k} \right| \quad (2)$$

where  $n$  is the number of observation points,  $Y_k$  is the actual observation of the studied variable and  $F_k$  is the forecasted value.



**Figure 9 Real versus estimated NO<sub>2</sub> concentrations during the last 10 days of January 2015.**

For our study we obtained a MAPE of 23.61%, which validates the forecast model. According to [30] the MAPE needs to be inferior to 30% for an accurate model validation.

### 3.2 Seasonality study

By analysing the current results, we observe that one of the independent predictors (traffic volume) is sensitive to time changes during the whole data set we are testing. As previously shown in Figure 3b), the traffic volume has a daily seasonality, with increasingly number of vehicles during the morning and afternoon rush hours, and low volumes at night. By conducting a trend analysis and seasonal decomposition on the first 20 days of the data set, we predict as well the number of cars during the last 10 days of January 2015. The best estimations we have obtained is by using a multiplicative model (trend and seasonal components are multiplied and then added to the error component), which returned a MAPE of 21.15%. This was better when compared to the additive model (the effects of individuals factors are differentiated and added together to model the data) which returned a MAPE of 43.6%. After integrating the fitted trend equation of the multiplicative model in equation (1), we obtain a new regression equation:

$$NO_2 = 28.43 + 0.05((337 + 0.0026 \cdot H_d) \cdot i) - 1.04T - 0.91W + 2.04H \quad (3),$$

where  $H_d$  represents the daily hours and  $i$  the seasonality index per hour. When validating again the NO<sub>2</sub> concentrations over the last 10 days of the data set using equation (3), we obtain a MAPE of 25.39%, which is still below the 30% validation threshold.

Although less accurate than the initial estimation, this result could provide good NO<sub>2</sub> estimations when access to traffic data is limited. The MAPE can be further improved

if longer data sets would be available for future studies and comparison. Validating the model on longer periods needs a more detailed analysis and verification, especially if other predictors would be available for the study.

## 4 Limitations and further studies on SO<sub>2</sub> and PM<sub>10</sub>

The main limitations of this work are: a) restricted data sets concerning the air pollutant concentrations and traffic counts which have been received from the traffic management centre; having access to longer periods of data would help to build more traffic demand scenarios and analyse for example how traffic control measures would improve the air quality in the neighbourhood, b) integrating the emissions from the train station would also improve the current model.

After the encouraging results obtained for the predictability of NO<sub>2</sub> using an integrated traffic and air pollution simulation model, we extended our analysis to other pollutants which were monitored by the Air Quality Monitoring station in Viaduct Kennedy. More specifically, we investigated the evolution of sulphur dioxide (SO<sub>2</sub>) and fine Particulate matter (PM<sub>10</sub>) during the same time interval (January 2015), and applied the same regression methodology as previously described.

### 4.1 SO<sub>2</sub> findings

SO<sub>2</sub> can be present in the air in areas with high industrial activity (that uses materials containing sulphur) but can also be present in motor vehicle emissions as the result of fuel combustion. Although several desulphurization procedures and technical improvements have been adopted to prevent this gas from being released in the atmosphere, its presence is still detectable in the air and can have a high impact on citizen's health. Figure 10a) presents the hourly SO<sub>2</sub> concentration over the study period, which shows alarming peaks registered mostly

during the morning and afternoon traffic rush hours (7:00-10:00 and 16:00-19:00). Compared to hourly NO<sub>2</sub> concentrations, which were more compact and followed almost the same trend evolution during working days or weekends, the hourly SO<sub>2</sub> concentrations are far sparser from one hour to another and do not follow a specific daily evolutionary trend. This behaviour is as well observed by studying the monthly evolution of SO<sub>2</sub> presented in Figure 10b), which suffered a high peak in concentration mostly over the third week of January 2015.

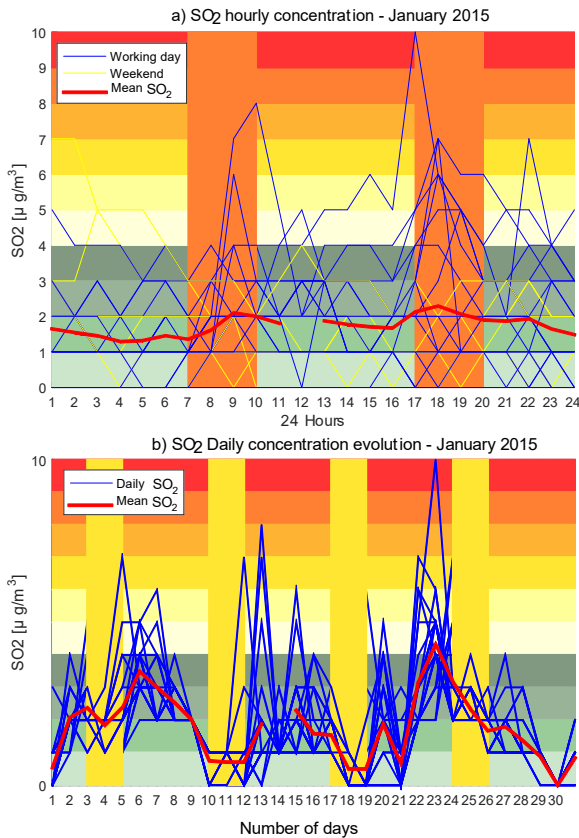


Figure 10 a) Hourly and b) daily SO<sub>2</sub> concentrations.

	SO <sub>2</sub>	Nr <sub>cars</sub>	T	W	H
SO <sub>2</sub>	1	-	-	-	-
Nr <sub>cars</sub>	0.207	1	-	-	-
T	-0.074	0.192	1	-	-
W	-0.155	0.194	0.767	1	-
H	-0.037	-0.185	-0.555	-0.549	1

Table 4 Spearman correlations between SO<sub>2</sub>, number of cars, wind, temperature and humidity.

By also conducting a Spearman correlation analysis between SO<sub>2</sub> and the rest of predictors we notice very loose correlation levels (0.207) with the traffic counts and small negative correlation indexes with temperature, wind and humidity, which confirm that other factors might be more important in influencing the SO<sub>2</sub> concentrations besides the current investigated ones (see Table 4). As

previously mentioned, in [16] the authors enhance on the fact that sulphur dioxide is almost exclusively issued by the manufacturing-waste-construction industry to an extent of almost 98%, which makes any prediction or environmental analysis very challenging as the data sets will be highly influenced by the industrial activity in the region where the air pollution monitoring takes place and the season when the data has been collected. For a consistent analysis, bigger data sets on long period of times (continuous monitoring over several months/years) would be necessary to be able to accurately predict the evolution of SO<sub>2</sub> and identify the most important predictor factor that can be easily integrated in all future predictions.

### 4.2 PM<sub>10</sub> findings

The presence in the air of Particulate Matters with a diameter smaller than 10 microns can highly affect the air quality and: a) produce respiratory problems for citizens, b) change the nutrient and chemical balance in water bodies, or c) reduce traffic visibility. As many of the PM<sub>10</sub> are produced by Diesel engines, we investigated as well the impact of the number of cars on the evolution of PM<sub>10</sub> registered in the current study area.

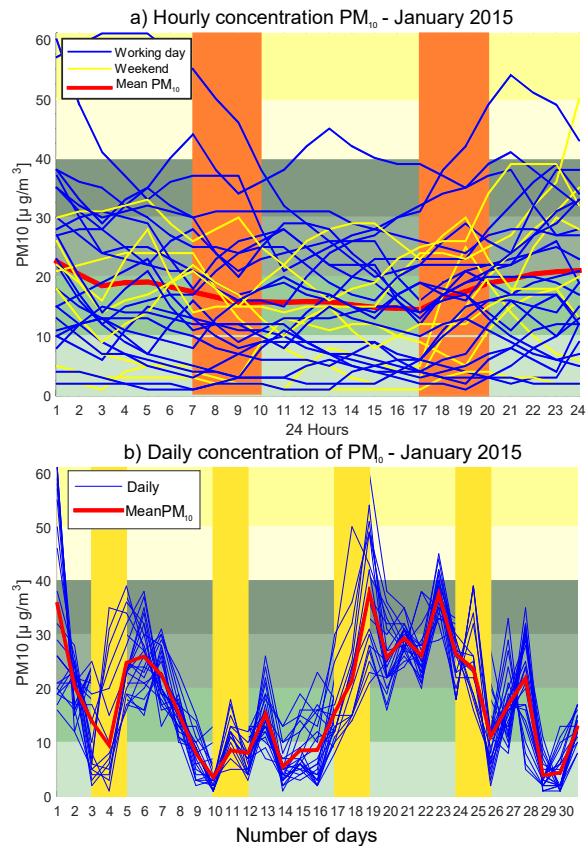


Figure 11 PM<sub>10</sub> daily and hourly concentration

Figure 11a) shows the daily evolution of PM<sub>10</sub> concentrations, which present sparse levels throughout the day, without being directly influenced by the morning or afternoon traffic peak hours. On the other hand, by observing the monthly evolution of PM<sub>10</sub> in Figure 11b, one would notice a smooth pollutant transition from one day to another without extreme outliers from the average; the highest peak was registered around 60 µg/m<sup>3</sup>, a low concentration level for PM<sub>10</sub> according to the air quality index of the European Union [17] but much higher than the annual limit threshold of 40 µg/m<sup>3</sup> identified by Air Lorraine in [16].

As many studies have indicated that humidity [31], wind speed and temperature [32] have different effects on particle number concentrations in various global regions, meteorological conditions are essential factors in analysing the levels and variations of particulate matter concentration. By further conducting the correlation analysis between PM<sub>10</sub> and the number of cars registered on the roads, as well as meteorological conditions, one can observe a negative and small correlation (-0.144) between traffic and fine particles for the time period investigated (see Table 5). This is mainly due to high level of humidity and precipitations during winter when the data has been collected, as shown by the negative and higher correlation factors between PM<sub>10</sub> and T (-0.547) or PM<sub>10</sub> and W (-0.494). Therefore, one would assume that fine particle concentration registered in humid weather conditions are not directly linked with the number of cars on the road.

	PM <sub>10</sub>	Nr <sub>cars</sub>	T	W	H
PM <sub>10</sub>	1	-	-	-	-
Nr <sub>cars</sub>	-0.144	1	-	-	-
T	-0.547	0.192	1	-	-
W	-0.494	0.194	0.767	1	-
H	-0.315	-0.185	-0.555	-0.549	1

**Table 5 Spearman correlations between PM<sub>10</sub>, number of cars, wind, temperature and humidity.**

This finding influences as well the LSMRM regression analysis on PM<sub>10</sub> applied for the first 20 days of the data set, which leads to the following equation:

$$PM_{10} = 9.14 - 0.00417 \cdot Nr_{cars} - 1.466 \cdot T + 0.0603 \cdot W + 15.75 \cdot H, \quad (4)$$

and the predictor analysis presented in Table 6. The VIF indicators remain below 5 which indicate that the chosen predictors are not correlated one to each other. On the other hand, the *p-values* for the number of cars and the wind are bigger than 0.05, which indicate that these predictors will not influence the initial regression equation, therefore their role will not be significant in the prediction step. If we remove these 2 predictors, one would only have to use the temperature and the humidity

to predict the evolution of fine particles in the neighbourhood. The regression test also returned a value of  $R^2 = 0.44$  for the coefficient of determination, which indicates that the data does not fit in a good proportion to the regression model. These findings put the prediction process for PM<sub>10</sub> on hold until further data sets will be available for use.

PREDICTOR	COEFF.	T	P	VIF
Constant	9.14	1.91	0.056	
Nr <sub>cars</sub>	-0.0014	-1.85	0.065	1.07
T[°C]	-1.466	-8.77	0	3.19
W [km/h]	0.0603	0.62	0.539	3.31
H[%]	15.75	3.26	0.001	1.71

**Table 6 Predictor analysis for PM<sub>10</sub>.**

## 5 Conclusions and future perspectives

In this paper, we proposed an integrated air pollution and traffic simulation model for building a simplified NO<sub>2</sub> estimation model, inside an eco-neighbourhood from north-eastern part of France. The first step was to conduct a detailed data profiling analysis, by taking into consideration the latest available information on the background concentration measured in the regional area which has been used for initial data assessment. The second step was to construct the 3D mesoscopic traffic simulation model of the study area using real data sets provided by the Grand Nancy council consisting in profiled traffic counts (every 15 minutes), traffic light signal controls and plans, pedestrian information and manual counts verifications. The third step consisted in using the simulated traffic volumes to build the regression model and estimate the NO<sub>2</sub> concentrations over the last 10 days of the model. The last step included a seasonality study in the prediction model based on a daily variation of number of cars. All results show a good forecasting for NO<sub>2</sub> concentrations which is subject to improve if more data would be available.

The same procedure was applied on SO<sub>2</sub> and PM<sub>10</sub> but the data sets have shown different behaviour, especially for SO<sub>2</sub> whose evolution is very hard to analyse and predict by using only traffic and meteorological data. One would need further information about the manufacturing industry in the area, the operating hours and the products being manufactured, which might be very difficult to obtain. The PM<sub>10</sub> data analysis and prediction indicated that more data sets would be required to build the prediction model and that one month of observations is not enough. This would imply further calibration and validation of the simulation model on extended time periods, with further manual counts and verifications which are the subject of approval from the Grand Nancy council.



The integrated platform represents a good support for testing various traffic scenarios, such as the limitation of traffic access in the neighbourhood during rush hours for regular cars, in order to encourage the use of electric vehicles. The simulation platform is a perfect tool for testing as well traffic light plan optimization methods or reconfiguration scenarios that would help improve the air quality in this highly-circulated neighbourhood. We are currently working on extending the hybrid model to the whole NGC eco-neighbourhood and build more detailed air pollution estimation models which would take into consideration all the types of vehicles on the roads. We are also testing wireless air pollution sensors in the neighbourhood, which offer real-time information of the air pollution at the human level. By human level we refer to an average height of around 1.5 meters, as the mobile stations were carried by hand by various volunteers when walking inside the eco-neighbourhood during rush hours. These sensors can be used on a daily basis for home-to-work journeys and represent an accurate supplementary source of information regarding the air pollution in the neighbourhood. The latest findings have been recently submitted for publication in [33] and [34].

## 6 Acknowledgements

This work has been developed in the ERPI laboratory, from Nancy France, under the Chaire REVES project funding. The final writing and submission of the paper has been done in the DATA61|CSIRO research laboratory from Sydney, Australia, with further work on the analysis of SO<sub>2</sub> and PM<sub>10</sub> pollutants. The authors of this work are grateful for the data and support provided by Grand Nancy, Air Lorraine and FlexSim Conseil.

## 7 References

- [1] UN, "World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)," Department of Economic and Social Affairs, Population Division, New York: United Nations, 2014.
- [2] TomTom, 2018. [Online]. Available: <https://bit.ly/1RxyKAI>. [Accessed 4 4 2018].
- [3] U. EPA, "What Are the Six Common Air Pollutants?," 2016. [Online]. Available: <http://www3.epa.gov/airquality/urbanair/>. [Accessed 30 03 2017].
- [4] U. EPA, "International Decontamination Research and Development Conference," National Homeland Security Research Center, Durham, NC, 2015.
- [5] EEA, "Air quality in Europe — 2017 report," European Environment Agency, Copenhagen, Denmark, 2017.
- [6] OMS, "Plus sain, plus juste, plus sur: l'itinéraire de la santé dans le monde 2007-2017," Organisation mondiale de la Santé ; Licence : CC BY-NC-SA, Genève , 2017.
- [7] W. B. Group and IHME, "The Cost of Air Pollution, Strengthening the Economic Case for Action," The World Bank and Institute for Health Metrics and Evaluation, University of Washington, Seattle, 2016.
- [8] OCDE, "Les conséquences économiques de la pollution de l'air extérieur," Éditions OCDE , Paris, 2016.
- [9] Z. Xuesong, T. Shams, L. Hao, J. Taylor, L. Bin and R. Nagui M., "Integrating a simplified emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of traffic management strategies," vol. 37, pp. 123-136.
- [10] T. P. S. F. P. B. J. & C. M. Fontes, "How to combine different microsimulation tools to assess the environmental impacts of road traffic? Lessons and directions," vol. 34, pp. 293-306, 2015.
- [11] M. F. Shorshani, M. André, C. Bonhomme and C. Seigneur, "Modelling chain for the effect of road traffic on air and water quality: Techniques, current status and future prospects," vol. 64, pp. 102-123, 2015.
- [12] G. Nancy. [Online]. Available: [www.grand-nancy.org/grands-projets/nancy-grand-coeur/](http://www.grand-nancy.org/grands-projets/nancy-grand-coeur/) . [Accessed 30 3 2017].
- [13] A. Mihaita, M. Camargo and P. Lhoste, "Optimization of a complex urban intersection using discrete event simulation and evolutionary algorithms.," Cape Town, South Africa, 2014.
- [14] A. Mihaita, L. Dupont and M. Camargo, "An urban traffic signal optimization using a 3D mesoscopic simulation approach and evolutionary algorithms (submitted)," 2016.
- [15] MEDE, "Bilan de la qualité de l'air en France en 2012," Ministère de l'écologie et du développement durable, Direction Générale de l'Energie et du Climat, Paris, 2012.
- [16] A. Lorraine, "Caractérisation de la qualité de l'air ambiant à Nancy en 2015 en contexte de proximité trafic," Air Lorraine, Nancy, 2016.
- [17] S. van den Elshout, "Citeair II, Common Information to European Air," European Union, Bruxelles, 2012.
- [18] D. Ni, "Multiscale Modeling of Traffic Flow," *Mathematica Aeterna*, vol. 1, no. 2011, pp. 27 - 54, 2011.
- [19] S. Maerivoet and B. De Moor, "Transportation Planning and Traffic Flow Models," Katholieke Universiteit Leuven, 2005.

- [20] S. P. Hoogendoorn and P. H. L. Bovy, "State-of-the-art of Vehicular Traffic Flow Modelling," *Journal of Systems and Control Engineering*, vol. 215, no. 4, 2001.
- [21] J. Barceló, *Fundamentals of Traffic Simulation*, Springer-Verlag New York, 2010.
- [22] Y. Wang and P. D. Prevedouros, "Synopsis of Traffic Simulation Models," University of Hawaii, Manoa, 1996.
- [23] Aimsun, TSS Barcelona, 2017. [Online]. Available: <https://bit.ly/2G174KH>. [Accessed 24 03 2018].
- [24] J. Archer and K. Hogskolan, "Indicators for Traffic Safety Assessment and Prediction and Their Application in microsimulation modelling: A study of urban and suburban intersections," KTH, Stockholm, 2005.
- [25] ASPA, "Modélisation de la qualité de l'air sur le futur éco-quartier Danube," Association pour la surveillance et l'étude de la Pollution Atmosphérique en Alsace., Alsace, 2012.
- [26] J. Galineau, "Bilan des émissions atmosphériques du transport routier en Lorraine," Air Lorraine, Nancy, 2012.
- [27] D. Liao and R. Valliant, "Variance inflation factors in the analysis of complex survey data.," vol. 38, no. 1, pp. 53-62, 2012.
- [28] S. D. Schlotzhauer, "Elementary Statistics Using JMP," Cary: NC: SAS Institute Inc., 2007.
- [29] S. Makridakis, S. C. Wheelwright and R. J. Hyndman, "Forecasting Methods and Applications.," 1998.
- [30] Y. Barlas, "Model Validation in System Dynamics," pp. 1-10, 1994.
- [31] H. A., K. H., S. S.L., J. J., J. H., P. T., A. F., N. T., K. M. and S. J.N., "The role of relative humidity in continental new particle formation," *J. Geophys. Res. Atmos*, vol. 116, p. 909–926, 2011.
- [32] T. Hussein, A. Karppinen, j. KUKkonen, j. Härkönen, P. Aalto, K. Hämeri, V.-M. Kerminen and M. Kulmala, "Meteorological dependence of size-fractionated number concentrations of urban aerosol particles," *Atmos. Environ.*, vol. 40, no. 1427–1440, 2006.
- [33] A.-S. Mihăiță, L. Dupont, O. Chery, M. Camaego and C. Cai, "Air quality monitoring using stationary versus mobile sensing units: a case study from Lorraine, France," vol. Special Number of ITS World COngress 2018, no. (submitted), 2018.
- [34] A.-S. Mihăiță, L. Dupont, O. Chery, M. Camaego and C. Cai, "Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling," vol. (submitted), 2018.



**Adriana-Simona Mihăiță** is a Research Scientist working in the Advanced Data Analytics in Transport (ADAIT) team from DATA61|CSIRO, Australia. She holds a PhD in Automatic Control obtained from the "Institut National Polytechnique de Grenoble", France in 2012. In 2016 she was the project leader for investigating the positioning accuracy from connected vehicles using DSRC in CITI (project funded by New South Wales government, Australia). Her research interest expands to: large scale traffic simulations using machine learning technologies in order to assist traffic planners to predict the congestion impact on the Sydney traffic network, air quality monitoring and prediction, hybrid systems, Markov Chains and event-based control.



**Mirian-Jannet Benavides** is an engineer graduated from University of Santiago de Chile, which enrolled in a double master program with the University of Lorraine, France in 2015. She was previously a research associate teaching in the Department of Technology and Information Communication from Santiago University. She worked under the supervision of Dr. MIHAITA during her master license.



**Mauricio Camargo** is a full Professor currently working in the domain of Technology Management and Innovation at the School of Industrial Engineering in the Lorraine University (ENSGSI) from Nancy, France. Prof. Camargo is involved in the development of Lorraine Smart Cities Living Lab, an innovation laboratory with the main aim of establishing a "user-driven" based development model to enhance the citizen's quality of life and to support the local economic and urban development. His main research studies are new product development, decision making in product-process design and cost estimation models.



**Chen Cai** is a senior researcher of the Analytics Research Group at Data61. He obtained his PhD on computational transportation science from University College London in 2010. His expertise lies in transport modelling and machine learning techniques. He is currently the deputy leader of the Advanced Data Analytics in Transport team at Data61. He is also a conjoint senior lecturer at UNSW.