Research article
Modeling tropospheric ozone and particulate matter in Tunis, Tunisia using generalized additive model

Zouhour Hammouda¹,², Leila Hedhili Zaier¹ and Nadège Blond³

¹Institut Supérieur de Gestion, Tunis, Tunisia
²Esprit School of Engineering, Tunis, Tunisia
³Laboratoire Image Ville Environnement, UMR 7362 CNRS/UDS, 3, rue de l’Argonne, 67000 Strasbourg, France

Received: 8 October 2020 - Reviewed: 23 November 2020 - Accepted: 10 September 2021
https://doi.org/10.17159/caj/2021/31/2.8880

Abstract
The main purpose of this paper is to analyze the sensitivity of tropospheric ozone and particulate matter concentrations to changes in local scale meteorology with the aid of meteorological variables (wind speed, wind direction, relative humidity, solar radiation and temperature) and intensity of traffic using hourly concentration of NOₓ, which are measured in three different locations in Tunis, (i.e. Gazela, Mannouba and Bab Aliwa). In order to quantify the impact of meteorological conditions and precursor concentrations on air pollution, a general model was developed where the logarithm of the hourly concentrations of O₃ and PM₁₀ were modeled as a sum of non-linear functions using the framework of Generalized Additive Models (GAMs). Partial effects of each predictor are presented. We obtain a good fit with R² = 85% for the response variable O₃ at Bab Aliwa station. Results show the aggregate impact of meteorological variables in the models explained 29% of the variance in PM₁₀ and 41% in O₃. This indicates that local meteorological condition is an active driver of air quality in Tunis. The time variables (hour of the day, day of the week and month) also have an effect. This is especially true for the time variable “month” that contributes significantly to the description of the study area.

Keywords
Air pollution, Particulate Matter, Tropospheric Ozone, GAM, meteorology, traffic

Introduction
Nowadays, it is well known that air pollution and its impact on human health have become a primary topic in atmosphere research. A good number of epidemiological studies have demonstrated the strong link between atmospheric pollution and daily deaths and hospitalizations of pulmonary and cardiac diseases (Sinhay et al., 2017; Bourdrel et al., 2017). Tunisia is a beautiful country with diverse, complex geography and is located between the Mediterranean coast and the Saharan region. This location together with a diversity of air pollution sources (e.g. traffic, industrial, dust) leads to exceedances of air quality guideline values recommended by the World Health Organization (WHO, 2016). Tunisia reports high annual mean concentrations of PM₂.₅ and PM₁₀, which should not exceed 10 and 20 μg.m⁻³, respectively (WHO, 2016). Accelerated growth in emission sources of air pollutants in most important Tunisian cities like Tunis, Sfax and Gabes (Melki, 2007; Bouchlaghem and Nsom, 2012) now cause an urgent need to adopt specific policies in managing air pollution.

Air pollution modeling is an integral part of air pollution management and policy (Karaca et al., 2006; Saffarini and Odat, 2008). Previous air quality studies conducted in Tunisia mainly focused on the physical characteristics, correlations between pollutants, the sources of PM₁₀ and forecasting air quality (Melki, 2007; Bouchlaghem et al., 2009; Ayari, Nouira and Tabelsi, 2012; Calzolai et al., 2015). A few investigations focusing on the interplay between meteorology and air quality has been done in Tunisia. The study conducted in Tunis (Melki, 2007) presents the role of the temperature inversions, which determine the majority of the highest pollution levels in the north of the country. They used multiple linear regressions to evaluate the statistic dependence between the ozone concentrations and the weather conditions. According to Bouchlaghem et al. (2009), some sea breeze events are responsible for air quality. Their result shows that under these circumstances, the nearby power plant is responsible for air quality degradation in the region of Sousse (the East central part of Tunisia). Bouchlaghem and Nsom (2012) highlighted the influence of the Saharan dust on PM₁₀ concentrations. They concluded that PM₁₀ concentrations on days with Saharan dust contributions are higher than the average daily value with the absence of this phenomenon. In sum, no study has as yet dealt with the relationship between particulate matter and ozone concentrations and meteorological conditions in Tunisia based on the use of a non-linear statistical approach.
Generalized Additive Model, as an extension of Generalized Linear Model, has been employed in few studies for modeling pollutant concentrations, especially PM\(_{10}\) (Taheri Shahraiyin et al., 2015) and O\(_3\) (Ma et al., 2020). As a statistical tool that is able to simulate non-linear relationships by smoothing input variables (Hastie and Tibshirani, 1990), Generalized Additive Models (GAM) have been used in many environmental issues and recent studies (Ma et al., 2020; Yang et al., 2020). In the last two decades, this statistical approach has been used as a standard analytic tool in time-series studies of air pollution and human health (He, Mazumdar and Arena, 2005; Dehghan et al., 2018; Ravindra et al., 2019).

GAM models delivered good performance and can be equivalent to those of other methods such as neural networks (Schlink et al., 2003). Aldrin and Haff (2005) used meteorological predictors in order to model PM\(_{10}\), PM\(_{2.5}\) and the difference between PM\(_{10}\) and PM\(_{2.5}\) mass concentrations, and their models gave a reasonably good fit in terms of the squared correlation coefficient with 72% and 80% for PM\(_{10}\) and NO\(_x\), respectively. Pearce et al. (2011) noted the influence of local-scale meteorological conditions on air quality in Melbourne (Australia). Munir et al. (2013) offered a new GAM to predict daily concentrations of PM\(_{10}\) in Makkah using lag PM\(_{10}\) concentrations. This model showed the vital role of meteorological variables and traffic related air pollutants in describing the variations of the PM\(_{10}\) concentrations. Again based on GAM analysis, Belušić, Herceg-Bulić and Bencetić Klaić (2015) employed the novel GAM approach to quantify the influence of local meteorology on air quality in Zagreb, Croatia. This study confirmed the well-known impact of wind direction and speed in variations of air pollution.

The objective of this study is to investigate the magnitude in which pollutant concentrations respond to measures of local meteorology and temporal variables in Tunis. Statistical models were developed for hourly mean PM\(_{10}\) and O\(_3\) concentrations for three sites of Tunis in order to quantify the impact of meteorology on PM\(_{10}\) and O\(_3\) levels. The paper is organized as follows: The Materials and methods section provides information on our data collection and meteorological variable values after the application of the data quality control process. Fig. 3 shows the average seasonal evolution of O\(_3\) in the studied regions. We note different behavior at the various sites with very high levels compared to the PM\(_{10}\) annual limit of the 2008 EU Air Quality Directive (40 \(\mu\)g.m\(^{-3}\)). The right-hand plot indicates that average seasonal evolution of O\(_3\) is around the O\(_3\) maximum daily 8-hour mean limit (60 ppb) of the 2008 EU Air Quality Directive (Directive, 2008), except for Gazela site, an overshoot was observed. So, pollution levels can be

Table 1 presents a basic statistical overview of air pollution and meteorological variable values after the application of the data quality control process. Fig. 3 shows the average seasonal evolution of PM\(_{10}\) (from January 2008 to December 2009) in the studied regions. We note different behavior at the various sites with very high levels compared to the PM\(_{10}\) annual limit of the 2008 EU Air Quality Directive (40 \(\mu\)g.m\(^{-3}\)). The right-hand plot indicates that average seasonal evolution of O\(_3\) is around the O\(_3\) maximum daily 8-hour mean limit (60 ppb) of the 2008 EU Air Quality Directive (Directive, 2008), except for Gazela site, an overshoot was observed. So, pollution levels can be
Table 1: Summary statistics of data used for model development, showing the mean, median, standard deviation, minimal and maximal values of the data collected over the 3 studied stations (01/01/2008 to 31/12/2009).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_3$</td>
<td>ppb</td>
<td>54.25</td>
<td>60</td>
<td>1</td>
<td>257</td>
<td>23.38</td>
</tr>
<tr>
<td>PM$_{10}$</td>
<td>μg.m$^{-3}$</td>
<td>68.26</td>
<td>52</td>
<td>1</td>
<td>801</td>
<td>59.92</td>
</tr>
<tr>
<td>NO$_X$</td>
<td>ppb</td>
<td>25.96</td>
<td>15</td>
<td>1</td>
<td>395</td>
<td>28.15</td>
</tr>
<tr>
<td>Temperature (TT)</td>
<td>ºC</td>
<td>18</td>
<td>18</td>
<td>3</td>
<td>43</td>
<td>6.99</td>
</tr>
<tr>
<td>Wind speed (WS)</td>
<td>m.s$^{-1}$</td>
<td>1.70</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>1.19</td>
</tr>
<tr>
<td>Wind direction (WD)</td>
<td>deg</td>
<td>201.8</td>
<td>249</td>
<td>0</td>
<td>360</td>
<td>115.5</td>
</tr>
<tr>
<td>Solar radiation (SR)</td>
<td>W.m$^{-2}$</td>
<td>177.8</td>
<td>44</td>
<td>0</td>
<td>927</td>
<td>235.5</td>
</tr>
<tr>
<td>Relative humidity (RH)</td>
<td>%</td>
<td>61.83</td>
<td>63</td>
<td>11</td>
<td>100</td>
<td>16.83</td>
</tr>
<tr>
<td>Day of the week (DW)</td>
<td>Days</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>Hour of the day (HD)</td>
<td>Hours</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>24</td>
<td>-</td>
</tr>
<tr>
<td>Month (Month)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>12</td>
<td>-</td>
</tr>
</tbody>
</table>
differentiated by geographical area. In Algeria, the north African country like Tunisia, the modeling results of Belhout et al. (2018) show that the Algerian annual average limit for PM$_{10}$ (80 μg.m$^{-3}$) has been exceeded in some Algiers areas; by consequence, air quality guidelines fixed by the WHO (20 μg.m$^{-3}$), (WHO, 2006) and the European Union (EU) (40 μg.m$^{-3}$) for PM$_{10}$ are also exceeded. Rahal et al. (2014) found that significant pollutant releases in the study area are located at hyper-centre and at centre of the Wilaya of Algiers. Many sites in Greater Agadir Area, Morocco, have high levels of ozone and other pollutants that meet national air quality standards. The annual average of PM$_{10}$ is largely below the limit value on Agadir city (Chirmata, Leghrib and Ichou, 2017). All countries of the North Africa sub-region do not have specific legislation on air quality.

**Generalized additive models**

Generalized Additive Models (Hastie and Tibshirani, 1990) are used to assess the relationship between air pollution concentrations and different factors. GAMs are regression models in which linear predictor $\sum \beta_j x_j$ is replaced by a sum of smooth functions of covariates $\sum s_k(x_k)$. Additive models are considered as a semi-parametric extension of the generalized linear model (GLM) which automatically estimate the optimal degree of non-linearity of the model. The additive model in general form can be written as:

$$ g(E(y_i)) = g(\mu_i) = s_0 + \sum_{k=1}^{p} s_k(x_{ki}) + \epsilon_i $$

(1)

where $g$ is a link function that links the expected value to the predictor variables, $\mu_i$ is the expectation of the response variable $y_i$, $s_0$ is the overall means of the response, $s_k(x_{ki})$ is the smooth function of $i^{th}$ value of covariate $k$, $p$ is the total number of covariates, and $\epsilon_i$ is the $i^{th}$ residual which is assumed to be normally distributed: $\epsilon_i \sim N(0, \sigma^2)$. The smooth function was used to minimize the penalized residual sum of squares (shown in equation 2):

$$ RSS(f, \lambda) = \sum_{j=1}^{n} (y_j - s(x_j))^2 + \lambda \int s'(t)^2 \, dt $$

(2)

The term $\int s'(t)^2 \, dt$ evaluates the closeness to the data and $\lambda$ penalizes curvature in the function. $\lambda$ is a fixed smoothing parameter. The increase of the value of $\lambda$ provides a smoother function. The choice of this parameter becomes
The variable day of the week (DW) was used to account for weekly variations. Also, the predictor hour of the day (HD) was employed with values ranging from 1 to 24. This variable is meant to take care of diurnal variation that is not explained by the other variables. Additionally, since air pollution data are known to be seasonal, a variable representing the month of the year was introduced into the model. The variable day of the week (DW) was used to account for weekly variations. Also, the predictor hour of the day (HD) was employed with values ranging from 1 to 24. This variable is meant to take care of diurnal variation that is not explained by the other variables. Additionally, since air pollution data are known to be seasonal, a variable representing the month of the year was introduced into the model.

The maximum number of knots for each smoother \( k \) must be chosen before the smoothing function is estimated. It controlled the smoothness of each function \( s_k(x_k) \) in the final model. This particular parameter should be large enough so that the main process which governs concentrations values are included in the model. Many studies were employed forward validation which is a special form of cross-validation and is considered as the easiest method to choose optimal knots (Aldrin and Haff, 2005; Belušić, Herceg-Bulić and Benetić Kliač, 2015). So, in this work, forward validation for each pollutant was based on hourly predictions of concentrations for Tunis, one day in advance. For each day and for the maximum number of knots, the model was re-estimated using the data up to the day before. Then, the hourly log PM\(_{2.5}\) and log \( O_3 \) concentrations for the next day are predicted. The prediction is compared to the logarithm of the observed value and the hourly prediction errors calculated. For each day and for each of the two pollutants, this procedure was repeated.

**Results and discussion**

Based on the data described in Section “Site description and sample collection”, the additive model with all variables was estimated for the two pollution variables PM\(_{2.5}\) and \( O_3 \) recorded at three different stations in Tunis.

The first two columns of Table 2 show the explained variation (squared correlation coefficients \( R^2 \)) for the entire model (equation 4). The second part of the table presents the explained variation for meteorological variables only \( (R^2_{m.v}) \) which measured the aggregate impacts of local meteorology on each pollutant. \( R^2_{m.v} \) corresponds to the explained variation of a new model given by the difference of the models with only time variables and with all variables. The highest values of \( R^2 \) were obtained for \( O_3 \) at Bab Aliwa station. We found that the explained variance for the entire model is between 0.56 and 0.85, indicating that the models explain most of the variation in pollutant concentrations, but a considerable amount of variation is still unexplained. The aggregate impact of meteorological variables was measured between 0.21 and 0.42.
Tropospheric ozone is considered a secondary pollutant which is formed by photochemical reactions involving the oxides of nitrogen NO and NO$_2$ (summed as NO$_X$), hydrocarbons and sunlight, particularly ultraviolet light. In urban areas, high ozone levels are observed during warm summer months when the temperature is high and the wind velocity is low. In Tunis, we found that the final model explained 85% (site of Bab Aliwa) of the variance of log-transformed O$_3$ concentrations (Table 2). The aggregate impact of meteorological variables explained 41% of the variance in O$_3$ for the same site (Bab Aliwa). The estimated effects of meteorological and temporal variables on O$_3$ are shown in Fig. 5 (a), (b) and (c) for three stations in Tunis. Most meteorological, traffic and temporal factors were statistically significant in a highly non-linear way.

**Ozone**

Tropospheric ozone is considered a secondary pollutant which is formed by photochemical reactions involving the oxides of nitrogen NO and NO$_2$ (summed as NO$_X$), hydrocarbons and sunlight, particularly ultraviolet light. In urban areas, high ozone levels are observed during warm summer months when the temperature is high and the wind velocity is low. In Tunis, we found that the final model explained 85% (site of Bab Aliwa) of the variance of log-transformed O$_3$ concentrations (Table 2). The aggregate impact of meteorological variables explained 41% of the variance in O$_3$ for the same site (Bab Aliwa). The estimated effects of meteorological and temporal variables on O$_3$ are shown in Fig. 5 (a), (b) and (c) for three stations in Tunis. Most meteorological, traffic and temporal factors were statistically significant in a highly non-linear way.

### The influence of local meteorology on O$_3$

**Temperature effect**

For all three measurement stations, temperature (TT) was an important meteorological variable for O$_3$. The effect of temperature on O$_3$ is similar at Gazela and Bab Aliwa sites. A positive effect is seen for temperatures ranging between 5°C-20°C across only these two sites. A negative effect is noted...
for temperatures ranging between 20°C and 40°C for all three sites. So, if temperature increases, ozone concentrations are seen to decrease. This disagrees with common understanding of this relationship (Cheng et al., 2007; Polinsky and Shavell, 2010; Pearce et al., 2011; Ma et al., 2020), but can due to correlations of temperature with other variables like wind direction. The formation and concentration of ground level \( O_3 \) depends on the concentrations of \( NO_x \) and VOCs, and the ratio of \( NO_x \) and VOCs. Ozone levels do not always increase with increases in temperature, such as when the ratio of VOCs to \( NO_x \) is low. As study area was surrounded by reliefs, the speeds of surface winds are low. It may be more thermal breezes than synoptic-scale winds (Melki, 2007). The high frequency of thermal breezes and calm periods may indicate stable atmospheric conditions and thus \( O_3 \) concentrations are higher during such episodes.

**Wind effect**

The curves in the center of Fig. 5 (a), (b) and (c) show the results obtained regarding the impact of wind direction. The estimated response for the wind direction is different for the various locations. This is as normal, since the effect of wind direction is strongly correlated on the emission locations. A non-linear relationship is observed for all stations: edf=6.51, edf=6.22 and edf=6.15 at Gazela, Mannouba and Bab Aliwa, respectively (Table 3). At the first site, \( O_3 \) exhibits maximum concentration for E-NE wind (70°-100°) and minimum concentration at around 200°. However, by examining the wind speed-direction frequencies graph of this site (Fig. 6), there is a very remarkable effect of this variable on ozone concentration. A possible explanation is the location of this measuring site which is subject to northern European pollution (i.e. \( O_3 \) is transported from Italy to Tunis). While crossing the city towards Mannouba site, the effect of wind decreases. In this station, \( O_3 \) shows secondary maxima for S-W wind (250°). The wind direction at the Bab Aliwa site seems to have a different effect on \( O_3 \) concentration. Wind direction has a positive effect on \( O_3 \) concentration for directions between 100° and 250°. This is probably associated with the cemetery effect which promotes ozone’s transport. A light minimum is then observed at 270°. The effect of road traffic can explain this. In this study, increasing wind speed was found to correspond to increasing \( O_3 \) concentrations. This tendency is particularly marked for the Bab Aliwa station (Figure 5c). This agrees with previous findings of Melki, (2007). At the Gazela site, the effect of this variable is very local, so, difficult to explain. It may be possible to understand this effect on a scale larger than a city.

**Solar radiation and relative humidity effects**

Solar radiation had a non-linear association: edf=6.47, edf=2.75
Figure 5c: GAM estimated relationships for temporal, meteorological and traffic variables on $O_3$ concentration for Bab Aliwa. The x-axis represents increasing variations. The y-axis indicates the contribution of the smoother to the fitted values. The region between the dashed lines represents the 95% confidence interval.

Table 3: Model estimates of the effects of predictors on $O_3$ (all sites). edf=effective degrees of freedom of the smooth function terms (edf>1 indicate non-linear relationships); $F$ value is an approximate $F$-test, SE=asymptotic standard error. *** Significant at the 0.000 level

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Gazela Site</th>
<th>Mannouba Site</th>
<th>Bab Aliwa Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>s(Hour of the Day)</td>
<td>11.5</td>
<td>9.43</td>
<td>10.91</td>
</tr>
<tr>
<td>s(Day of the Week)</td>
<td>5.27</td>
<td>2.97</td>
<td>5.68</td>
</tr>
<tr>
<td>s(Temperature)</td>
<td>10.75</td>
<td>13.74</td>
<td>7.01</td>
</tr>
<tr>
<td>s(Wind Speed)</td>
<td>4.42</td>
<td>2.66</td>
<td>4.45</td>
</tr>
<tr>
<td>s(Wind Direction)</td>
<td>6.51</td>
<td>6.22</td>
<td>6.15</td>
</tr>
<tr>
<td>s(Relative Humidity)</td>
<td>8.72</td>
<td>8.00</td>
<td>8.02</td>
</tr>
<tr>
<td>s(Month)</td>
<td>4.87</td>
<td>3.98</td>
<td>4.70</td>
</tr>
<tr>
<td>s(NO$_X$)</td>
<td>7.40</td>
<td>13.63</td>
<td>8.94</td>
</tr>
<tr>
<td>s(Solar Radiation)</td>
<td>6.47</td>
<td>2.75</td>
<td>6.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linear terms</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.2</td>
<td>0.001</td>
<td>3.94</td>
<td>0.004</td>
<td>2.95</td>
<td>0.004</td>
</tr>
<tr>
<td>Explained Deviance</td>
<td>73%</td>
<td>-</td>
<td>61.5%</td>
<td>-</td>
<td>85.6%</td>
<td>-</td>
</tr>
<tr>
<td>GCV score</td>
<td>0.01</td>
<td>-</td>
<td>0.12</td>
<td>-</td>
<td>0.08</td>
<td>-</td>
</tr>
</tbody>
</table>
and edf= 6.25 at Gazela, Mannouba and Bab Aliwa, respectively, (Table 3) with $O_3$ concentrations. These results are very clear, higher solar radiation corresponds to higher concentrations of $O_3$. This positive effect was found to be strongest after values surpassed 400 W.m$^{-2}$ (Gazela and Bab Aliwa station). This relationship is consistent with the literature (Pearce et al., 2011) as radiation plays a significant role in photochemistry of ozone production (Dawson, Adams and Pandis, 2007). The nature of response of $O_3$ to the RH showed a 10% under low RH, and then exhibited a modest negative relationship where high levels resulted in a regional decrease of up to 10% for Gazela and Mannouba, and 5% for Bab Aliwa. So, the curves go downward for increasing humidity. Generally, the results obtained in this analysis of meteorological parameters were expected, i.e. that higher ozone concentrations were associated with high temperature, low relative humidity and prolonged sunshine (Lacour et al., 2006). In this coastal region of the northern Mediterranean, at night the relative humidity of the air is important (96% on average), combined with a decline in temperature (18°C on average). This conjunction will reduce $O_3$ concentrations.

The impact of time and traffic variables on $O_3$

The upper left panel of Fig. 5 (a), (b) and (c) show how the concentrations of $O_3$ varies as the hour of the day (HD) changes. Each curve corresponds to one of the measurements stations. Since this variable describes the diurnal variation of $O_3$ in three locations, different curves are observed. The diurnal variation for Mannouba and Bab Aliwa sites show a similar pattern with $O_3$ concentrations reaching the peak at around 9:00 at the Mannouba site and at around 14:00 at the Bab Aliwa site. The increase in $O_3$ concentrations during day time is due to the increase in solar radiation, which powers the photochemical reactions and consequently $O_3$ concentration (Khoder, 2009). The hour’s period of negative effect is presumably due to high emissions of NOx caused by the intensity of traffic. Monks et al. (2015) highlighted the non-linearity of the O$_3$-VOC-NOx system. VOC-limited refers to the fact that the production of $O_3$ is limited by the input of VOC. Indeed, high NOx lead to lower $O_3$ because NOx directly react with NO. The local production of ozone is less reduced because the NOx react with hydroxyl radical species formed in the atmosphere. When these hydroxyl radicals do not react with NOx (example: low emission of NOx), they contribute to the VOC degradation and the ozone production. Unlike other sites, Gazela is considered a residential site which is characterized by the domination of NOx emissions (at this site, VOCs are only due to traffic, and not as much emitted as by factories like the other sites). In fact, a minimum of $O_3$ concentration is observed at around 8:00 and a maximum at around 19:00 when traffic is an important source of emissions and the vertical mixing is reduced. Influenced by transport of $O_3$ from other regions and local NOx concentrations at night, the increase of the surface $O_3$ concentration during the night time was larger than that during the daytime (Lei and Wang, 2014).

Day of the week at Gazela and Mannouba (Table 3) was found to have little influence on ozone, ($F=3.81$, $F=3.27$. respectively). For Monday to Wednesday the ozone concentrations remain more or less unchanged (Figure 5). The rise in ozone concentrations is observed on Thursday and Friday but is followed by a drop as of Saturday. This continues on Sunday when the levels of ozone then join those on Monday. This result was also found by Pont and Fontan (2000) for five large French cities: This study does not show any significant variation in ozone concentrations between weekend and week except for the strongest values where a 40% reduction in precursors would lead to a 20% increase in ozone. The weekend effect would be reversed. Due to constant of road traffic during all the days of the week in Bab Aliwa, no effect of the variable DW was observed. NOx also has a non-linear association with $O_3$ concentration, with edf=7.40 and edf=8.94 at Gazela and Bab Aliwa, respectively (Table 3). Increased NOx for these two sites was found to have a negative effect on $O_3$. This finding is in agreement with other work since the chemical coupling of $O_3$ and NOx make levels of $O_3$ inextricably linked: Ozone production is dependent on the state of NOx, as NOx and NO increase the production and dissociation of $O_3$, respectively. Consequently, an increased NOx/NOy ratio reduces the ozone concentration (Melkonyan and Kuttler, 2012). Analysis the results of Mannouba station reveals a different NOx effect, when the NOx concentrations is over 200 ppb, an increase of NOx concentrations leads to a lower decrease of $O_3$ concentrations than at the other stations. An increase in $O_3$ concentrations is seen above 280 ppb of NOx concentrations. This is presumably due to the location of this station, which includes small forests in the west and chemical plants in the south which promote VOCs emissions, then the increase of both $O_3$ and NOx concentrations. A positive effect is detected for the

![Figure 6: Wind speed-direction frequencies for three Meteorological Stations (from left to the right) Gazela, Mannouba and Bab Aliwa. Each cell gives the total number of hours the wind was from that wind speed/direction (period of 2008-2009). The number of hours is coded as a color scale shown to the right. The dashed circular grey lines show the wind speed scale.](image-url)
variable Month on $O_3$ concentration in warm months (spring and summer). In this period, there is an increase in temperatures and in the intensity of solar radiation. These meteorological conditions promote the mixing process of pollutants and $O_3$ formation. The ozone evolution is controlled not only by the influence of climate but also by the movement of pollutants. In fact, the same result was found in two regions: Spain and Italy which belong to the Mediterranean climate (Domínguez-López et al., 2014; Myriokefalitakis et al., 2016).

PM$_{10}$

The impact of traffic and site location on PM$_{10}$

Atmospheric PM$_{10}$ are multicomponent aerosols. They originate from a variety of mobile, stationary and other natural sources, and are also formed in the atmosphere through chemical and physical processes. $SO_2$ (mainly issued from industrial sector) and NOX (mainly issued from transport sector) are two precursors of secondary particulate matter (Harrison, Jones and Lawrence, 2004). Their chemical and physical compositions vary widely. Many studies showed that the PM$_{10}$ yearly, daily and hourly average concentration exceeds the Tunisian and the European standard limits at all the sampling stations (Bouchlaghem et al., 2009). A significant proportion of PM$_{10}$ in Tunis has many sources like sea salt, mineral dust (Calzolai et al., 2015). In the Mediterranean Tunisian regions, the average seasonal evolution of PM$_{10}$ is characterized by a winter maximum (November and December) (Bouchlaghem and Nsom, 2012). On the other hand, ozone concentration reaches its maximum values during summer period under the great photochemical activity and the effect of land-sea breeze. This difference has been highlighted in many studies and has been explained by the formation of PM$_{10}$ as a complex mixture of many chemical species. Indeed, both the proximity to traffic sources and the different types of air mass scenarios make PM$_{10}$ formation rather complex and associated with geographic, temporal and meteorological conditions. In Tunis, we found that the final model explained between 56% and 59% of the variance of log-transformed PM$_{10}$. The highest value of $R^2$ was found at Bab Aliwa station and the aggregate impact of meteorological variables accounting for 29%. The estimated effects of independent variables of the model are shown in Fig. 7 (a), (b) and (c) for three stations in Tunis. The model shows how the association of PM$_{10}$ concentrations varies with the levels of other variables. The association between NOX concentrations and PM$_{10}$ concentrations was non-linear with edf=8.41, edf=8.48 and edf=8.12 at Gazela, Mannouba and Bab Aliwa respectively (Table 4) and is characterized by a general positive effect. It is reasonable and also found in Munir et al. (2013). Actually both NOX and PM$_{10}$ are largely issued from road traffic. The curve for Bab Aliwa is the one going farthest to the
were observed for different sites. In the first station, Gazela, (center of Fig. 7 (a)), PM$_{10}$ exhibit a first maximum concentration for wind direction around 170°. This can be explained by localized effect of the road. The secondary maximum is observed around 320°, clearly reflecting the effect the small factory situated north of the study area. As Bab Aliwa is based next to taxi and bus stations, this particular measuring site is subject to PM$_{10}$ transport by southeast winds. For relative humidity, the results are very clear especially for Gazela and Mannouba sites, which find that high humidity was associated to low PM$_{10}$ concentration. So, the curves go downward for humidity better than 80%. This agrees with previous findings of Aldrin and Haff (2005) and Belušić, Herceg-Bulić and Bencetić Klaić (2015).

Particles are then removed from contaminated surface air by wet deposition in precipitation added to dry deposition (Giri, Murthy and Adhikary, 2008). The estimate curves of temperature have the same slope for the various locations. Temperature was named as the most significant meteorological variable for Bab Aliwa (F=179.51, p-value <0.001) and Gazela (F=175.75, p-value <0.001) sites. Interpretation of the curves (lower left of Fig. 7 (a), (b) and (c)) can be expressed as follows: increasing temperature corresponds with increasing PM$_{10}$ concentration. So, the curves go downward for humidity better than 80%.This agrees with previous findings of Aldrin and Haff (2005) and Belušić, Herceg-Bulić and Bencetić Klaić (2015).

The influence of local meteorology on PM$_{10}$
A non-linear association was observed between PM$_{10}$ and wind speed. This variable has a positive effect on PM$_{10}$ concentration from 4 m.s$^{-1}$ to 8 m.s$^{-1}$ at Gazela site. The curves for Mannouba and Bab Aliwa (Fig. 7 (b) and Fig. 7 (c)) reached the peak at 5 m.s$^{-1}$ then decrease. The same wind behavior was observed in three sites and was found in Belušić, Herceg-Bulić and Bencetić Klaić (2015): For large wind speeds, PM$_{10}$ concentration decrease. This result was as expected as low wind and stable atmospheric conditions support higher concentrations of PM$_{10}$. We note however that the decrease in PM$_{10}$ levels at higher winds observed in the present study is in contrast to the result found in Makkah by Munir et al., (2013) and in Maribor by Lešnik, Mongus and Jesenko (2019). Wind direction had variable association with PM$_{10}$: edf=6.88 at Bab Aliwa site (Table 4). Several curves were observed for different sites. In the first station, Gazela, (center of Fig. 7 (a)), PM$_{10}$ exhibit a first maximum concentration for wind direction around 170°. This can be explained by localized effect of the road. The secondary maximum is observed around 320°, clearly reflecting the effect the small factory situated north of the study area. As Bab Aliwa is based next to taxi and bus stations, this particular measuring site is subject to PM$_{10}$ transport by southeast winds. For relative humidity, the results are very clear especially for Gazela and Mannouba sites, which find that high humidity was associated to low PM$_{10}$ concentration. So, the curves go downward for humidity better than 80%.This agrees with previous findings of Aldrin and Haff (2005) and Belušić, Herceg-Bulić and Bencetić Klaić (2015).

Particles are then removed from contaminated surface air by wet deposition in precipitation added to dry deposition (Giri, Murthy and Adhikary, 2008). The estimate curves of temperature have the same slope for the various locations. Temperature was named as the most significant meteorological variable for Bab Aliwa (F=179.51, p-value <0.001) and Gazela (F=175.75, p-value <0.001) sites. Interpretation of the curves (lower left of Fig. 7 (a), (b) and (c)) can be expressed as follows: increasing temperature corresponds with increasing PM$_{10}$ concentration. So, the curves go downward for humidity better than 80%.This agrees with previous findings of Aldrin and Haff (2005) and Belušić, Herceg-Bulić and Bencetić Klaić (2015).

The influence of local meteorology on PM$_{10}$
A non-linear association was observed between PM$_{10}$ and wind speed. This variable has a positive effect on PM$_{10}$ concentration from 4 m.s$^{-1}$ to 8 m.s$^{-1}$ at Gazela site. The curves for Mannouba and Bab Aliwa (Fig. 7 (b) and Fig. 7 (c)) reached the peak at 5 m.s$^{-1}$ then decrease. The same wind behavior was observed in three sites and was found in Belušić, Herceg-Bulić and Bencetić Klaić (2015): For large wind speeds, PM$_{10}$ concentration decrease. This result was as expected as low wind and stable atmospheric conditions support higher concentrations of PM$_{10}$. We note however that the decrease in PM$_{10}$ levels at higher winds observed in the present study is in contrast to the result found in Makkah by Munir et al., (2013) and in Maribor by Lešnik, Mongus and Jesenko (2019). Wind direction had variable association with PM$_{10}$: edf=6.88 at Bab Aliwa site (Table 4). Several curves were observed for different sites. In the first station, Gazela, (center of Fig. 7 (a)), PM$_{10}$ exhibit a first maximum concentration for wind direction around 170°. This can be explained by localized effect of the road. The secondary maximum is observed around 320°, clearly reflecting the effect the small factory situated north of the study area. As Bab Aliwa is based next to taxi and bus stations, this particular measuring site is subject to PM$_{10}$ transport by southeast winds. For relative humidity, the results are very clear especially for Gazela and Mannouba sites, which find that high humidity was associated to low PM$_{10}$ concentration. So, the curves go downward for humidity better than 80%.This agrees with previous findings of Aldrin and Haff (2005) and Belušić, Herceg-Bulić and Bencetić Klaić (2015).
Figure 7c: GAM estimated relationships for temporal, meteorological and traffic variables on PM$_{10}$ concentration for Bab Aliwa. The x-axis represents increasing variations. The y-axis indicates the contribution of the smoother to the fitted values. The region between the dashed lines represents the 95% confidence interval.

Table 4: Model estimates of the effects of predictors on PM$_{10}$ (all sites). edf = effective degrees of freedom of the smooth function terms (edf>1 indicate non-linear relationships); F value is an approximate F-test; SE=asymptotic standard error. *** Significant at the 0.000 level

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Gazela Site</th>
<th>Mannouba Site</th>
<th>Bab Aliwa Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>s(Hour of the Day)</td>
<td>10.29</td>
<td>12.75</td>
<td>13.10</td>
</tr>
<tr>
<td>s(Day of the Week)</td>
<td>4.68</td>
<td>2.03</td>
<td>2.53</td>
</tr>
<tr>
<td>s(Temperature)</td>
<td>9.64</td>
<td>10.98</td>
<td>4.09</td>
</tr>
<tr>
<td>s(Wind Speed)</td>
<td>4.93</td>
<td>6.78</td>
<td>5.66</td>
</tr>
<tr>
<td>s(Wind Direction)</td>
<td>6.93</td>
<td>6.56</td>
<td>6.88</td>
</tr>
<tr>
<td>s(Relative Humidity)</td>
<td>6.44</td>
<td>3.84</td>
<td>7.66</td>
</tr>
<tr>
<td>s(Month)</td>
<td>4.95</td>
<td>3.89</td>
<td>4.70</td>
</tr>
<tr>
<td>s(NO$_X$)</td>
<td>8.41</td>
<td>8.48</td>
<td>8.12</td>
</tr>
<tr>
<td>s(Solar Radiation)</td>
<td>5.00</td>
<td>7.64</td>
<td>1.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linear terms</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.72</td>
<td>0.005</td>
<td>3.9</td>
<td>0.007</td>
<td>4.26</td>
<td>0.006</td>
</tr>
<tr>
<td>Explained Deviance</td>
<td>58.5%</td>
<td>-</td>
<td>54.5%</td>
<td>-</td>
<td>60.1%</td>
<td>-</td>
</tr>
<tr>
<td>GCV score</td>
<td>0.27</td>
<td>-</td>
<td>0.36</td>
<td>-</td>
<td>0.18</td>
<td>-</td>
</tr>
</tbody>
</table>
and Nsom, 2012). However, the positive relationship between temperature and PM$_{10}$ is probably explained by the dust layer created over three sites especially during peak hours.

The impact of time variables on PM$_{10}$

The time variable hour of the day (HD) has a non-linear association with PM$_{10}$ concentration. It was mainly used to account the effect of traffic. At the study stations, PM$_{10}$ concentration fall to a minimum between 7:00-8:00 and increase until 10:00, this corresponds to the morning peak traffic flow. In Bab Aliwa site, an evening peak traffic flow was noted at around 21:00. This second peak is probably due to people’s daily commuting between the capital and the suburbs. Curves of partial effect of the variable Month pointed out that in all measuring sites, PM$_{10}$ is characterized by a winter maximum (December-January-February). This result is consistent with the data of Bouchlaghem and Nsom (2012), who found a winter PM$_{10}$ peak in five different stations (traffic, industrial and residential) in Tunisia. This is presumably due to the influence of low mixing in the atmosphere and the advection of Saharan plumes. We note the absence of the second peak observed during the summer in the previous works (Bouchlaghem and Nsom, 2012). The slight effect of Saharan dust can be explained by the temporal difference between the South and the North of Tunisia and the geographical locations of the monitoring stations far from the southwest origin of the Saharan event. Since the Mannouba station is placed close to agriculture fields, plowing during the autumn season (September-October) promotes increasing PM$_{10}$ concentrations.

Assessment of the model performance

Table 5: Statistical evaluation of the model for all pollutants at Gazela site for the entire study period

<table>
<thead>
<tr>
<th></th>
<th>O$_3$ (ppb)</th>
<th>PM$_{10}$ (μg.m$^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOA</td>
<td>0.91</td>
<td>0.77</td>
</tr>
<tr>
<td>RMSE</td>
<td>6.46</td>
<td>38.21</td>
</tr>
<tr>
<td>Modified RMSE</td>
<td>6.46</td>
<td>38.21</td>
</tr>
<tr>
<td>Measurement standard deviation</td>
<td>12.11</td>
<td>51.81</td>
</tr>
<tr>
<td>Model standard deviation</td>
<td>10.21</td>
<td>34.82</td>
</tr>
<tr>
<td>Measurement mean</td>
<td>67.89</td>
<td>56.50</td>
</tr>
<tr>
<td>Model mean</td>
<td>67.89</td>
<td>56.50</td>
</tr>
</tbody>
</table>

Various metrics (RMSE, modified RMSE, measurement standard deviation, model standard deviation and IOA (see Appendix A)) were used to assess the model performance. This statistical evaluation of the model on the original scale is presented in Table 5.

Table 5 is for all variables at Gazela site; other pollutants and measuring site data are not shown here as the results are similar to these. The first criterion for model evaluation was checked and both RMSE and Modified RMSE are less than measurement standard deviation. In addition, the index of agreement is 0.91 and 0.77 for O$_3$ and PM$_{10}$, respectively, which corresponds to a good compromise between modeled and measured values. Fig. 8 shows the relationship between the response and fitted values of O$_3$ concentration at Gazela site. PM$_{10}$ and other measuring site data are not shown as they are similar to those presented in this figure. This figure shows a positive linear relationship with a good deal of scatter. Residual plots are also used to characterize model efficacy. Fig. 9 clearly shows that the majority of residuals group around zero, as expected. The x-axis range on the left-hand plot and the y-axis range on the right-hand plot are the same.

Conclusions

The objective of this work was to estimate the relationship between each of two pollution variables, namely concentrations of PM$_{10}$ and tropospheric ozone O$_3$ and NO$_x$ concentrations (taking as a proxy of traffic) as well as a set of meteorological variables for the urban area of Tunis. To achieve this objective,
a statistical methodology is used based on the Generalized Additive Model (GAM). We have shown that the GAM can model the non-linear effect of the covariates. The model is additive on the log scale and the estimates were made on hourly data collected during two years at three different locations in Tunis. The model provides a reasonably good fit in terms of the explained variance. For all stations, $O_3$ was easier to model (i.e. with more explanatory power and higher values of $R^2$). The most significant important variables for $O_3$ are $NO_2$, wind direction and relative humidity. The impact of temperature and $NO_3$ is the strongest for $PM_{2.5}$, followed by relative humidity and wind variables. The time variables (hour of the day, day of the week and month) appear to have a particular impact on air quality. In this study, the variable Month plays a significant role in the characterization of the study area as a function of time. In fact, we note the seasonal behavior of $O_3$ and $PM_{2.5}$ pollutants, with the highest concentrations in summer and winter, respectively. These results allow a first and fast analysis of the air pollution due to $O_3$ and $PM_{2.5}$ in 3 locations in Tunis. It emphasizes the critical role of the local conditions on the air pollution, and especially the emissions and the weather as two main drivers of urban air pollution. Our findings suggest focusing on model improvement as future work. The addition of precipitation and traffic density (number of vehicles) variables could help to improve the model assessment. So, it is necessary to take into account all the sources of emissions exhaustively. In summary, the use of GAM in combination with partial residual plots offered an effective way to outline the relationships between temporal, meteorological and traffic variables and air pollution. Although our study did not detail chemical and physical aspects of air pollution, the results produced were reasonable and comparable to other studies. Furthermore, the results may be considered as relevant because research work on air pollution is insufficient in Tunisia. To this end, after quantifying the influence of all used variables, we plan to use GAM and GAMM (Hastie and Tibshirani, 1990; Wood, 2006) models to forecast pollutant concentrations.

References


Domínguez-López, D., Adame, J.A., Hernández-Ceballos, M.A., Vaca, F., De la Morena, B.A. and Bolivar, J.P. (2014) ’Spatial and temporal variation of surface ozone, NO and NO$_2$ at urban, suburban, rural and industrial sites in the southwest of the...
Iberian Peninsula', *Environ Monit Assess*. https://doi.org/10.1007/s10661-014-3783-9


**Appendix A**

**edf**: The effective degrees of freedom (edf) estimated from generalized additive models were used as a proxy for the degree of non-linearity in stressor-response relationships. An edf of 1 is equivalent to a linear relationship, an edf > 1 and ≤ 2 is a weakly non-linear relationship, and an edf > 2 indicates a highly non-linear relationship.

**GCV**: generalized cross validation score can be taken as an estimate of the mean square prediction error based on a leave-one-out cross validation estimation process. We estimate the model for all observations except $i$, then note the squared residual predicting observation $i$ from the model. Then we do this for all observations. GCV criteria is numerically stable and efficient, but its computation become extensive especially when several smoothing parameters have to be estimated.

**F-statistic**: An F statistic is a value you get when you run an ANOVA test or a regression analysis to find out if the means between two populations are significantly different. In regression case, the F value is the result of a test where the null hypothesis is that *all of the regression coefficients are equal to zero*. In other words, the model has no predictive capability. Basically, the F-test compares your model with zero predictor variables (the intercept only model), and decides whether your added coefficients improved the model.

**Asymptotic Standard Error**: Asymptotic standard error is an approximation to the standard error, based upon some mathematical simplification. In regression analysis, the term "standard error" refers either to the square root of the reduced chi-squared statistic, or the standard error for a particular regression coefficient (as used in, say, confidence intervals).

**VIF**: Variance Inflation Factor detects multicollinearity in regression analysis. For an independent variable $X_i$, it can be calculated by the formula below using R-squared values:

$$VIF_i = \frac{1}{1 - R^2}$$

**IOA**: Index of Agreement is a standardized measure of the degree of model prediction error which varies between 0 and 1. IOA=1 represents full agreement and IOA=0 indicates no agreement at all.

$$IOA = 1 - \frac{1}{2} \sum_{i=1}^{n} \left( O_i - P_i \right)^2$$

**RMSE**: The Root Mean Square Error is used to measure the difference between values predicted and values observed.

$$RMSE = \sqrt{ \frac{1}{n} \sum_{i=1}^{n} \left( O_i - P_i \right)^2 }$$

$$Modified \ RMSE = \sqrt{ \frac{1}{n} \sum_{i=1}^{n} \left( O_i - \bar{O} - P_i + \bar{P} \right)^2 }$$