Multiple Linear Regression (MLR) model simulation of hourly PM10 concentrations during sea breeze events in the Split area

Simulacija satnih koncentracija lebdećih čestica PM10 modelom multiple linearne regresije u uvjetima obalne cirkulacije na splitskom području

Tanja Trošić Lesar
Meteorological and Hydrological Service
Zagreb, Croatia
e-mail: trosic@cirius.dhz.hr

Anita Filipčić
Department of Geography
Faculty of Science
University of Zagreb, Croatia
e-mail: filipcic@geog.pmf.hr

Summary
The main objective of this study is to simulate the daily change of the particulate matter (PM10) concentrations using a multiple linear regression (MLR) model for the selected sea breeze cases in Split and Kaštel-Sućurac, which are situated in the central part of the Eastern Adriatic coast, for the period from 2007 to 2009. A new predictor characteristic to the daily and nightly part of the coastal circulation is included in the MLR model. The MLR model results well match the PM10 concentration measurements during the selected sea breeze cases in Split (with the correlation coefficient $R^2=0.77$ and the index of the agreement $IA = 0.89$) and Kaštel-Sućurac ($R^2=0.65$, $IA =0.81$), and they also match the measurements during land breeze (for Split $R^2=0.66$, $IA = 0.86$ and for Kaštel-Sućurac $R^2=0.66$, $IA = 0.70$).

1. INTRODUCTION / Uvod
Particulate matter (PM) pollution in urban areas has a significant impact on human health (e.g. Thurston 1996, Alebić-Juretić et al. 2007). Generally, the causes of high PM values can be (a) local pollution sources, such as intensive traffic and small-scale combustion, (b) natural sources of particles (e.g. dust, sea salt and wild-land fires), (c) inefficient local atmospheric dispersion conditions (calm conditions, temperature inversions), and (d) synoptic weather conditions that favour the long-range transport of pollutants (e.g. Bešić et al. 2007, Vardoulakis and Kassomenos 2008).

The multivariate statistical models for predicting the daily concentrations of NOx and PM10 (particulate matter of the size 10 μm or less) based on the meteorological parameters were applied in urban areas in several studies, however their focus is mainly on prediction of mean daily concentration values (e.g. Chaloulakou et al. 2003, 2005, Kukkonen et al. 2003, Paschalidou and Kassomenos 2004, Paschalidou et al. 2011, Vlachogianni et al. 2011). The MLR models have shown to predict air pollutant concentrations with remarkable success (Cordelino et al., 2001; Paschalidou et al., 2009). However, due to the nature of linear relationship in the parameters, regression models may not provide accurate predictions in some complex situations such as non-linear data and extreme values data.

Further, the studies have shown that pollutants, upon entering a coastal circulation, subsequently circulate in a vertical plane (Lyons and Olsson 1973, Keen and Lyons 1978, Simpson 1994). Apart from the daytime pollution, another large problem is caused by the wind direction change which occurs in the evening, when high concentrations of pollutants return to the same area. The occurrence of the coastal air circulation is a result of a differential heating of the land and sea, as one of the most important climatological factors (e.g. Filipčić 1994). A layer of dry and warm air in the inner convective boundary layer over the land is important in cases of air pollution in
coastal areas if there are sources of pollution (Grisogono et al. 1998, Plant and Atkinson 2002). As for the distribution of pollutants several factors are of considerable importance: the topography of coastal areas, the coastal shape and the position of industrial plants (e.g. Pielke et al. 1983, Grossi et al. 2000). Clappier et al. (2000) showed that the pollution in Athens is much higher than in Los Angeles because of the location of the industrial plants in the wider area and local orography. Evtyugina et al. (2006) explored the sea breeze at the coast of Portugal and they concluded that the inland areas show higher pollution levels than the highly populated industrialized areas of towns and villages near the coast, due to the transfer of pollutants during the day. Some studies indicate a significant effect of the local orography and the position of the coast and islands on the coastal air-circulation in the observed areas (e.g. Estoque 1962, Arritt 1989, Arritt 1993, Grisogono et al. 1998, Nitis et al. 2005). Similar results were obtained for the Croatian coast. For instance, Trošić et al. (2006) investigated the sea breeze energetics in Zadar, Croatia and they determined the influence of local orography and friction on the sea breeze development. Telišman Prtenjak et al. (2008) concluded that the topography can significantly affect the shape of the hodograph and the rotation of the wind vectors in the Northern Adriatic. Bencetić Klaić et al. (2009) concluded that the sea and the land breeze along the eastern Adriatic coast are increased by the local channelling, which further affects the penetration of the sea breeze into the land during the day. Similar results were obtained by Telišman Prtenjak et al. (2010a) for the Southern Adriatic using the surface, sodar and satellite measurements. The bora was also found to affect the local sea breeze development (Telišman Prtenjak et al. 2010b), which was strongly affected by the local orography and the bora temporal evolution (e.g. Belušić and Klaić 2006, Trošić and Trošić 2010, Trošić 2015).

The main objective of this paper is to introduce a multiple linear regression (MLR) model to simulate hourly PM10 concentrations based on hourly meteorological and air quality data during sea breeze/land breeze events in Split and Kaštel-Sućurac (see Fig. 1). A new predictor, the air temperature change, will be included. This predictor has been chosen to account for daytime and nighttime effects on local circulation. The results will be validated using the PM10 concentrations measurements at these stations, and the comparison of the obtained results will be made. In the end, a conclusion will be based on the obtained results.

2. THE METHODS OF ANALYSIS / Metode analize

2.1. Measurement data / Podaci

The measurements used in the analysis are hourly concentrations of PM10, as well as the hourly meteorological measurements of the wind and air temperatures, which are measured at the

Figure 1 The position of the measurement stations in Split (43° 30’ 39”, 16° 26’ 56”) and Kaštel-Sućurac (43° 32’ 52”, 16° 26’ 23”), and the positions of the mountains Kozjak and Mosor, as well as of the Bay of Kaštela. Elevation of the stations is 46 m for the Split station and 19 m for the Kaštel-Sućurac station.

Slika 1. Položaj mjerne postaje u Splitu (43° 30’ 39”, 16° 26’ 56”) i Kaštel-Sućurcu (43° 32’ 52”, 16° 26’ 23”), te položaj planina Kozjak i Mosor i kaštelanskog zaljeva. Nadmorska visina postaje u Splitu je 46 m, a postaje u Kaštel-Sućurcu 19 m.
Table 1 The number of selected sea-breeze days at Split and Kaštel-Sućurac from April to September in the period from 2007-2009, with all the PM10 concentrations measurement data available. The sum presents the total yearly and monthly number of days Tablica 1. Broj odabranih dana sa zmorcem u Splitu i Kaštel-Sućurcu od travnja do rujna u periodu od 2007.-2009., sa svim raspoloživim koncentracijama čestic PM10. U zbroju je prikazan ukupan godišnji i mjesecni broj dana

<table>
<thead>
<tr>
<th></th>
<th>meas. station</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Split</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Kaštel-Sućurac</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2008</td>
<td>Split</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Kaštel-Sućurac</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2009</td>
<td>Split</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Kaštel-Sućurac</td>
<td>1</td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>10</td>
<td>48</td>
</tr>
<tr>
<td>Sum</td>
<td>Split</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>4</td>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Kaštel-Sućurac</td>
<td>1</td>
<td>16</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>58</td>
</tr>
</tbody>
</table>

same site at the Split and Kaštel-Sućurac stations, in the period from 2007 to 2009. The Split measurement station Split-1 is placed in the Split town centre and it is an urban background station. The position of the stations is shown in Fig. 1. Split is situated in the central part of the Eastern Adriatic coast (see Fig. 1), separated from the inland area by mountain Mosor (1339 m height) from the NE-E and hill Kozjak (779 m) from the N-NW. Also, Split is surrounded by the three large islands, which can affect the sea breeze circulation development. The Split industrial zone, including the cement industry, is located in Kaštel-Sućurac, which is situated in the coastal area of the Bay of Kaštela. The cement industry is one of the largest sources of industrial pollution in the larger Split area. The main purpose of the station is the monitoring of pollution, as a result of traffic and industry. Kaštel-Sućurac is situated in the central part of the coastal area of the Bay of Kaštela. The measurement station in Kaštel-Sućurac is situated approximately 400 m northwest of the cement factory. In the vicinity of this station, the houses and small industrial plants can be found. The main purpose of the station is to monitor the pollution that is a result of the local industry, but this does not preclude the monitoring of other sources of pollution. The total suspended particles are measured using a sampling method of the beta-dust monitor Verewa F701, with recordings every 15 min, and the hourly measurements obtained were used. The measuring principle of the determination of dust mass is based on the fact that the beta radiation is attenuated by the transmission through matter and the radiometric measurement is achieved using a Betaemitter (C-14) and a Geiger-Müller counter.

2.2. The criteria for selection of sea breeze days with an undisturbed coastal circulation / Kriteriji izbora dana sa zmorcem s neporemećenom obalnom cirkulacijom

By the use of a several criteria the days with a notable change in wind speed and direction at the time of the morning and evening sea breeze lulls have been selected. A sea to land wind direction ranging from 130 to 310 degrees has been chosen to signify cases relevant for the analysis of sea breeze patterns. Criteria for the wind speed at the time of the sea breeze lulls was not allowed to exceed 0.9 m s⁻¹ (Marić, 1998). Also, differences in air temperature at the time of the morning and evening sea breeze lulls were set not to exceed 1 °C (Trošić, 2002). The upper and surface synoptic charts at 1200 UTC have shown low synoptic pressure gradients for the selected days over the Central Adriatic (according to the Deutsche Wetterbericht, 2007-2009). In addition, the upper level analysis charts analysis showed that the weak synoptic wind favours the sea breeze development. The cloudiness in the climatological terms was not allowed to exceed 4/10 (e.g. Prtenjak et al. 2008). Table 1 shows annual and monthly number of days for Split and Kaštel-Sućurac. Based on the criteria, 41 sea breeze cases were selected for the Split station and 38 cases for the Kaštel-Sućurac station.

2.3. The characteristics of the selected sea breeze days / Karakteristike dana sa zmorcem

Figure 2 and Figure 3 show the mean daily change of the air temperature, wind speed and wind direction from April to September and for the period from 2007 to 2009, for Split and Kaštel-Sućurac, respectively. They show an increase in the mean wind speed during the sea breeze with the rising air temperatures. It is obvious that the changes in the wind direction, from the sea breeze to land breeze and vice versa, occur at the time of the sea breeze lulls. A similar daily change of the sea breeze wind speed and wind direction is visible in each month, the sea breeze starts around 0600 UTC and lasts until 1900 to 2000 UTC.

2.4. The MLR model / MLR model

In the MLR model it is essential to determine how the meteorological factors affect the air pollutant concentrations. Thus, the PM10 concentrations can be treated as a response to the meteorological variables as predictors.

The model is given by:

\[
X_t = a_0 + a_1 M_{1,t} + a_2 M_{2,t} + \ldots + a_k M_{k,t} + \epsilon_t, \quad i = 0, \ldots, k\;
\]

where \(X_t\) is the pollutant concentration at time \(t\), \(M_{i,t}\) stand for the meteorological variables at time \(t\), \(a_i\) are regression coefficients, and \(\epsilon_t\) is the random error at time \(t\).

Ideally, all the explanatory variables in a MLR model should be independent from one another, and in some cases certain variables may be exactly, or very nearly, linear combinations of other variables, which is called multicollinearity.

Further, to avoid the discontinuity that would be caused by using the wind direction (WD), expressed in degrees, the Wind Direction Index (WDI) is used (Ordieres et al., 2005):

\[
WDI = I + \sin(WD + \pi/4).
\]

WDI is not uniquely defined in the whole 0-360 deg range, but the sea breeze and land breeze are opposite winds, so this has no influence on the final results. It is known that the sea breeze depends on the temperature differences between the land and sea, and thus the differences of the air temperatures during the sea breeze and the air temperature at the time of
the morning sea breeze lull will be used ($dT=T_i-T_lull$). The air temperature during sea breeze lulls corresponds well with the sea temperature (Trošić 2002, Trošić et al. 2006).

There are two methods that can be applied to select efficient regression models. The simplest approach is called the forward selection where the variables are added to the model one at a time. The complexity of the model should depend on its purpose and available data (e.g. Hilborn and Mangel 1997). In order to analyse multicollinearity, a variance inflation factor (VIF) is calculated for meteorological parameters in the model. It is found that VIF ranges between 1 and 2 and the correlation coefficient between the selected predictors is in all cases below 0.6, therefore it is assumed that multicollinearity between selected predictors is not present (Zuur et al. 2009). The most significant of these variables, fulfilling the criteria of $P$-value is equal or below 10% are added to the model.

Figure 2 The mean daily change of the air temperature (°C), wind speed (ms⁻¹) and wind direction (deg) at the Split station from April to September, 2007-2009

Slika 2 Srednja dnevna promjena temperature zraka (°C), brzina vjetra (ms⁻¹) i smjer vjetra (stupnjevi) na postaji Split od travnja do rujna, 2007.-2009.
3. DISCUSSION OF THE RESULTS / Rasprava

Figure 4 shows the dependence of the PM10 concentrations in Split and Kaštel-Sućurac for the dT, wind speed and WDI during the sea breeze. For Split the concentrations show a decrease with the wind speed increase, and an increase with the dT increase. On the other hand, it first shows an increase for Kaštel-Sućurac and then a decrease for both the dT and wind speed increase. In Split the concentrations were very similar regarding the air temperature changes, but still the largest for the largest dT values. The decrease of PM10 with WDI is very similar at both stations. Table 2 shows the minimum, maximum and mean PM10 concentrations for Split and Kaštel-Sućurac for the selected days, separately for sea breeze and land breeze. It shows that the largest PM10 concentrations were recorded in Kaštel-Sućurac, particularly during the sea breeze (139.29 µg m⁻³).

The MLR analysis was made for hours between morning and evening sea breeze lull. At the time of the sea breeze lull...
The wind speed is below 1 m s\(^{-1}\) and notable change of wind direction from land to sea and vice versa. Furthermore, H1, H2 and H3 are the PM10 concentrations at 0600, 0700 and 0800 UTC, respectively.

The MLR model PM10 concentrations for Split and Kaštel-Sućurac for the daytime and night-time coastal circulation are given in expressions (3)-(6):

\[
\begin{align*}
PM10_{\text{Split}_\text{day}} &= 3.2762 - 0.7501WDI + 0.2238dT + 0.3863H1 - 0.2391H2 + 0.8981H3 - 0.1666\text{Time} \quad (3) \\
PM10_{\text{Split}_\text{night}} &= 8.1814 - 3.3942WSPD + 0.6783H1 + 0.3921H2 \quad (4) \\
PM10_{\text{Kaštel-Sućurac}_\text{day}} &= 26.5521 - 3.5763WDI - 3.1274WSPD + 0.3813H1 + 0.2074H2 - 0.5434H3 \quad (5) \\
PM10_{\text{Kaštel-Sućurac}_\text{night}} &= 5.2496 + 0.4352dT + 0.2249H1 + 0.1665H2 + 0.3003\text{Time} \quad (6)
\end{align*}
\]

where WDI is the wind direction, WSPD represents the wind speed, dT is the difference of air temperature at some hour and the air temperature at the time of the sea breeze lull. Furthermore, H1, H2 and H3 are the PM10 concentrations at 0600, 0700 and 0800 UTC, respectively.

The MLR model is applied also to the night-time part of the coastal circulation. For these calculations, dT is also calculated as a difference between the air temperature at some hour during the land breeze (in the morning and evening) and the air temperature at the time of the morning sea breeze lull. The same pollutant concentrations H1, H2 and H3 are used as for the sea breeze, at 0600, 0700 and 0800 UTC, respectively.

The forecasting performance and accuracy of the PM10 MLR model are evaluated using the Root Mean Square error (RMSE), Mean Absolute Error (MAE), correlation coefficient (R) and Index of Agreement (IA) (see Table 3). The R\(^2\) is the actual variance explained (see e.g. Belušić et al. 2015). The equations for RMSE, MAE, R\(^2\) and IA, as well as VIF are shown from (7)-(11):
Figure 6. Usporedba mjesečnih srednjaka satnih simuliranih i mjerenih koncentracija PM10 čestica (μg m⁻³) za period od travnja do rujna za Split i Kaštel-Sućurac od 2007. do 2009.

\[
\text{MAE} = \frac{\sum_{i=1}^{N} |p_i - O_i|}{N}
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (O_i - P_i)^2}{N}}
\]

\[
\text{IA} = 1 - \frac{\sum_{i=1}^{N} (p_i - O_i)^2}{\sum_{i=1}^{N} [p_i - \overline{O}]^2 + [O_i - \overline{O}]^2}
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (O_i - \overline{O}) (p_i - \overline{P})^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2}
\]

\[
\text{VIF}_i = \frac{1}{1 - R_i^2}
\]

The results of the MLR model for the hours of the sea breeze show very good agreement with the measurements, especially for the Split station with R²=0.77, IA=0.89. The results for the
nightly part of the coastal circulation are also better for the Split station, with IA=0.86 and R²=0.66. For the Kaštel-Sućurac station the statistics for the daily and the nightly part of the coastal circulation give similar statistics, with R²=0.66 and IA around 0.70. The results show almost equal or even better agreement with the hourly measurements than obtained by e.g. Vlachogianni et al. (2011) with only the mean daily values. The simulated hourly PM10 concentrations for the selected day with sea breeze in Kaštel-Sućurac well matches the measurements (see Fig. 5), but it cannot simulate the hourly PM10 concentration variations observed.

Figure 6 shows the comparison of the monthly means of the hourly simulated and measured PM10 concentrations for the selected sea breeze cases from April to September, from 2007-2009 in Split and Kaštel-Sućurac.

The mean hourly fluctuations of PM10 concentrations are captured by the MLR model very well, especially for Kaštel-Sućurac. The largest PM10 concentrations at the Kaštel-Sućurac station are recorded at the time of the sea breeze lulls, and at the Split station during the night. Moreover, at the time of the morning sea breeze lull, morning traffic peak occurs which can lead to higher concentration observed in Kaštel-Sućurac. In addition, PM10 concentrations are well simulated also seasonally, especially at Kaštel-Sućurac station where PM10 concentrations were larger than in Split in all months. In Split there was an increase of concentration values around mid-day in April, May, June and July, that was not observed in August and September. Due to partial overlapping of Kaštel-Sućurac and Split coastal circulations the PM10 from Kaštel-Sućurac can be transported towards Split causing the PM10 concentrations increase around mid-day. Also, this increase can be partly influenced by the rush-hour since the station is positioned near the road with moderate traffic.

4. CONCLUSION / Zaključak

The main objective of this study was to simulate the daily change of PM10 concentrations during the sea breeze events at Split town and Kaštel-Sućurac with inclusion of a new predictor characteristic for the coastal circulation. Due to a large influence of the local orography and islands, the results for both stations could be related to the ones obtained for Athens from e.g. Clappier et al. (2000) and Vlachogianni et al. (2011), but the pollutant concentrations in Athens were significantly higher. For Split, the overlapping of the coastal circulations from Kaštel-Sućurac and Split in the morning hours can cause the pollutant concentration increase around mid-day which is visible especially during the summer months. Also, this increase can be influenced by the rush-hour, since the station is positioned near a public road.

For the MLR model a new predictor relating to the daily air temperature differences, which are the main reason for the coastal circulation development, is used for the sea breeze cases in Split and Kaštel-Sućurac. Vlachogianni et al. (2011) conclude that the accuracy of the MLR model is crucially dependent on the selection and preparation of the input variables, and they were carefully selected in the present analysis. The simplest approach called the forward selection is used for the available data. In order to analyse multicolinearity, a variance inflation factor (VIF) is calculated for all meteorological variables and time in the model. It is found that VIF ranges between 1 and 2 and the correlation coefficient is in all cases below 0.6, therefore it is assumed that multicolinearity between selected predictors is not present (Zuur et al. 2009).

The MLR model results match well the measured PM10 concentrations during the sea breeze for the Split station (R²=0.77, IA = 0.89), as well as for the Kaštel-Sućurac station (R²=0.66, IA =0.70). Almost similar results are obtained for the nightly part of the coastal circulation. Kassomenos (2005) concluded that the higher population density combined with specific socioeconomic conditions cause uncertainties to the pollutant predictions, what well corresponds with our results and this can be the reason for the MLR model discrepancies with the measurements at both stations.

The comparison of the mean monthly simulated concentrations with the measurements indicates a very good match. Also, the MLR model simulated PM10 concentrations for the selected sea breeze case well match the hourly measurements for Kaštel-Sućurac.

It is known that the sea breeze has a large impact on the pollution, especially during the summer months. The MLR model with the specific predictor characteristic for the daily and nightly part of the coastal circulation, which is used in this paper, can be applied to determine the average daily PM10 variations during the sea breeze and land breeze in coastal areas.

ACKNOWLEDGEMENTS / Zahvale

The authors are grateful to the Croatian environment agency for provision of meteorological and PM10 measurement data used in this paper and Cemex Croatia d.d. for additional relevant information regarding measurements.

REFERENCES / Literatura


