GEOFIZIKA VOL. 32 2015

DOI: 10.15233/gfz.2015.32.5
Original scientific paper
UDC: 551.588.74

Using a generalized additive model to quantify the influence of local meteorology on air quality in Zagreb

Andreina Belušić, Ivana Herceg-Bulić and Zvjezdana Bencetić Klaić

University of Zagreb, Faculty of Science, Department of Geophysics,
Andrija Mohorovičić Geophysical Institute, Zagreb, Croatia

Received 9 January 2015, in final form 16 June 2015

This paper reports the estimated response of hourly mean concentrations of selected air pollutants, namely carbon monoxide (CO), sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and particulate matter with an aerodynamic diameter of up to 10 μm (PM_{10}), to local scale meteorology in Zagreb, Croatia for the period 2006–2012. A new method is applied here for the urban area of Zagreb. In a general model, a logarithm of hourly mean air pollutant concentrations is expressed as the sum of the nonlinear functions of meteorological and several time variables, with the latter included accounting for temporal variation in emissions. The model can be formulated within the framework of generalized additive models (GAMs) and is additive on the logarithmic scale, which results in multiplicative effects on the original scale. Although the model is nonlinear, it is simple and easy to interpret. It quantifies the impact of meteorological conditions and emissions on air pollution. A measure of the relative importance of each predictor, partial effects and statistical evaluation of the model are also presented. Overall, the results show that the most important predictors are those related to emissions. The aggregate impact of meteorological variables in the model explained 45% of variance in CO, 14% in SO_2, 25% in NO_2 and 24% in PM_{10}. This indicates that meteorology, at least on a local scale, is a noticeable driver of air quality in Zagreb. Stable atmospheric conditions in the urban area favour the occurrence of higher concentrations of air pollutants. Convection processes dominate under unstable conditions, resulting in the dilution of pollutant concentrations within the boundary layer.

Keywords: atmospheric boundary layer, statistical modelling, urban air quality

1. Introduction

Nowadays, it is well known that concentrations of airborne pollutants are significantly affected by meteorological conditions (e.g. Finlayson-Pitts and Pitts, 1986; Levy et al., 2003; Aldrin and Haff, 2005; Hussein et al., 2006; Cheng et al.,

2007; Prtenjak et al., 2009, 2013; Klaić et al., 2012, 2015). Past investigations focusing on the interplay between meteorology and air quality in Croatia can be grouped as follows:

1) *Modelling of the long-range transport of airborne pollutants*. These early studies describe the development and application of a simple Lagrangian numerical box-model simulating the long-range transport of sulphur compounds towards Croatia (Klaić, 1990, 1996; Klaić and Beširević, 1998).

2) *Modelling of regional pollutant transport.* These studies focus on the further development of the European mesoscale Eulerian chemical transport modelling system (EMEP) (Simpson et al., 2003) and its adaptation to domains over Croatia (Jeričević et al., 2007, 2010). Additionally, some effort was also put into development of the atmospheric Lagrangian particle stochastic model (Kos et al., 2004).

3) *Case studies of observed pollution episodes*. These studies employ either existing Eulerian mesoscale meteorological models (e.g. Klaić et al., 2003; Prtenjak et al., 2012) or Eulerian mesoscale meteorology models coupled with atmospheric chemistry models (e.g. Prtenjak et al., 2009, 2013) in order to detect meteorological conditions responsible for the establishment of investigated pollution episodes.

4) *Simple statistical analyses of relationships between observed air quality data and concurrent meteorological data.* For example, Bešlić et al. (2007) analysed observed daily mean airborne particulate matter with an aerodynamic diameter of up to 10 μm (PM_{10}) concentrations with respect to six synoptic weather types typical for continental Croatia. The authors found that whereas elevated particle concentrations were associated with the radiation weather type and south-eastern advection, lower concentrations were most often found together with the wind weather type. In order to identify wind directions associated with the transport of pollutants towards selected measuring sites, some of authors have used EMEP daily sector analysis (Klaić, 1988; Bešlić et al., 2008), a technique based on backward two-dimensional trajectories (available at *http://www.emep.int/Traj_data/traj2D.html*). Gvozdić et al. (2011) applied principal component analysis in order to investigate the relationship between pollutant (NO_2, H_2S, SO_2, $PM_{2.5}$ and O_3) concentrations and meteorological variables (temperature, relative humidity, wind speed and direction and precipitation) in Slavonski Brod (eastern Croatia). The authors concluded that winter pollution episodes are associated with low air temperatures, weak winds and high relative humidity. Bralić et al. (2012) reported on Spearman correlations between daily mean pollutant concentrations (SO_2, NO_2 and black-smoke) and meteorological parameters (daily mean temperature, wind speed, cloudiness and total precipitation) recorded in the coastal city of Split (Middle Adriatic). These authors found statistically significant increases in SO_2 and NO_2 concentrations occurring together with increasing daily mean temperature and cloudiness, respectively. In contrast, a sta-

tistically significant decrease in daily mean SO_2 was associated with an increase in precipitation amount, wind speed and cloudiness. Recently, 1-min mean outdoor (Klaić et al., 2012) and indoor (Klaić et al., 2015) PM_1 (particles aerodynamic diameters up to 1 μm) mass concentrations recorded at measuring sites in residential Zagreb were statistically analysed with respect to short-term outdoor atmospheric conditions (specifically, 1-min mean relative humidity, wind speed and direction, sea level pressure, temperature and global radiation). Results suggested the influence of regional and/or long-range pollutant transport on outdoor PM_1 levels. An increase in both outdoor and indoor PM_1 levels with increasing mean sea level pressure and outdoor relative humidity was also recorded.

5) *Neural network modelling*. Hrust et al. (2009) developed a new method with which to forecast hourly concentrations of air pollutants in Zagreb, Croatia. The authors found the temporal variables which imply human activities (such as traffic, heating etc.) to be the most important for CO. Furthermore, the agreement between the modelled and measured NO_2 and CO was better than that for PM_{10}, with this poorer model performance for PM_{10} suggested as being due to the inherent irregularities associated with processes affecting particle production (such as traffic emissions and particle resuspension) and/or the omission of relevant input variables (e.g. boundary layer height). Recently, Grgurić et al. (2014) tested four different empirical models (specifically, a univariate linear regression model, a multivariate linear regression model with first order effects, a multivariate linear regression model with first and second order effects, and an artificial neural network model) in order to establish the relationship between daily mean ground-based PM_{10} mass concentrations and aerosol optical depth (AOD) data. Whereas for the univariate linear regression model the independent variable was AOD, for the multivariate models and neural network model the meteorological parameters taken as independent predictors were AOD, boundary layer height, surface relative humidity, wind speed and direction, air temperature at 2 m above ground level, and surface pressure. The authors concluded that the univariate linear regression model failed to explain data variability (suggesting nonlinearity in the PM_{10}-AOD relationship), with the multivariate models and neural network model performing better.

As the above review suggests, no study has as yet dealt with the relationship between pollutant concentrations and meteorological conditions in Croatia based on the use of generalized additive models (GAMs). As statistical models able to simulate nonlinear relationships by smoothing concurrent input variables (Hastie and Tibshirani, 1990; Hastie et al., 2009), GAMs are suitable for a wide range of environmental issues and have thus been used in many recent studies (e.g. Schlink et al., 2003; Aldrin and Haff, 2005; Ito et al., 2007; Zhang and Batterman, 2010; Pearce et al., 2011; Chen et al., 2012; Gonzales et al., 2012; Munir et al., 2013; Quisthoudt, 2013; Otto et al., 2014; Wanka et al., 2014). According to Schlink et al. (2003), GAMs performance is comparable to those of neural network models, and, GAMs are applicable for ozone forecasting, for which purpose they (as well

as the neural network approach) outperform linear forecasting techniques. Aldrin and Haff (2005) employed a GAM in order to model PM_{10}, $PM_{2.5}$ and $PM_{10} - PM_{2.5}$ (particle sizes between 2.5 and 10 μm) concentrations based on meteorological predictors. For temperatures below 0 °C the authors found that both PM_{10} and $PM_{2.5}$ increased with decreasing temperature. In contrast, the effects on relative humidity were not so straightforward; whereas concentrations of larger fractions (i.e. PM_{10} and $PM_{10} - PM_{2.5}$) decreased with an increase in relative humidity, concentrations of $PM_{2.5}$ increased. Additionally, concentrations of large fractions were also reduced when the ground surface was covered with snow or ice. Again based on GAM analysis, Pearce et al. (2011) concluded that local-scale meteorological conditions have the largest impact on air quality in Melbourne, Australia. The most significant variables for PM_{10} were found to be temperature, wind, water vapour pressure and boundary layer height, with temperature, followed by wind and water vapour pressure, substantially affecting NO_2.

The present study similarly employs the novel GAM approach, this time with the aim of developing models for the estimation of the relative importance and partial effects of input variables (predictors) on pollutant (CO, SO_2, NO_2 and PM_{10}) concentrations for locations across the urban area of Zagreb, Croatia. A GAM was selected as the basic model because a simple predictor-response formulation was preferred. Separate models were formulated for hourly mean CO, SO_2, NO_2 and PM_{10} concentrations for three sites in Zagreb, with the overall objective being to investigate pollutant response to local scale meteorology.

2. Measuring sites and measurements

2.1. Local meteorological data

Links between air pollutants and local weather conditions were determined based on hourly observations made during a seven-year period (1 January 2006–31 December 2012). Meteorological variables measured at Zagreb-Maksimir Observatory (45° 49' 15.25" N, 16° 2' 5.58" E) (Fig. 1 and Tab. 1) and provided by the Croatian Meteorological and Hydrological Service included: temperature (°C), mean sea-level pressure (hPa), relative humidity (%), wind direction (°), wind speed ($m\,s^{-1}$) and precipitation ($mm\,h^{-1}$). Whereas wind speed and direction were measured at 10 m above the ground, the remaining meteorological parameters were measured in a standard meteorological shelter placed 2 m above a grassy surface.

Two precipitation variables also included in the analysis (weighted mean precipitation corresponding to the preceding four hours, and weighted mean precipitation corresponding to the preceding week) deserve a more detailed explanation. As per previous studies (e.g. Aldrin and Haff, 2005), average precipitation for the last four hours was employed in order to account for the removal of pollutants from the atmosphere by precipitation, as well as the effect of dry/

Table 1. Summary statistics of meteorological variables used as predictors in the GAM. Variables (hourly means) are recorded at Zagreb-Maksimir Observatory during the period 1 January 2006–31 December 2012.

Variable	Minimum	Mean	Median	Maximum	Standard deviation
Temperature (Temp.) (°C)	–16.70	12.22	12.40	38.30	9.28
Pressure (Press.) (hPa)	980.81	1016.33	1016.08	1046.73	7.93
Relative humidity (Rel. hum.) (%)	15.00	71.98	77.00	100.00	19.90
Wind direction (dir.) (°)	0.00	18.62	18.00	36.00	12.02
Wind speed (m s^{-1})	0.00	1.32	1.00	9.40	1.10
Precipitation last four h (Prec. 4h) (mm h^{-1})	0.00	0.03	0.00	5.60	0.12
Precipitation last week (Prec. week) (mm h^{-1})	0.00	0.03	0.02	2.00	0.04

wet roads on pollutant resuspension. The weighted average precipitation corresponding to the preceding four hours is calculated as follows:

$$\frac{1}{10}\left(4P_t + 3P_{t-1} + 2P_{t-2} + P_{t-3}\right), \tag{1}$$

where P_t denotes precipitation corresponding to hour t. The weighted average precipitation corresponding to the preceding week was employed in order to account for the effect of abundant precipitation, assuming that such an event may wash pollutants away from the road. Weighted precipitation corresponding to the preceding week is calculated as follows:

$$\frac{1}{\sum_{j=1}^{168} w_j} \sum_{j=1}^{168} w_j P_{t-3-j}\,, \tag{2}$$

where the weights $w_j = 169 - j$ vary from 1 to 168, and 168 is the aggregated number of hours in one week. P_{t-3-j} denotes the amount of precipitation recorded starting from the preceding four hours. The weights decrease linearly in order to ensure that hours closest to the selected hour are the most influential. Table 1 contains a basic statistical overview of the meteorological data used as predictor variables in the model.

2.2. Air pollution data

Local air pollution data were provided by the Air Quality Department of the Croatian Meteorological and Hydrological Service. Hourly mean concentrations measured at three sites across Zagreb, namely Zagreb-1 (45° 48' 1.28" N, 15° 58' 27.62" E), Zagreb-2 (45° 49' 25.4" N, 16° 2' 9.87" E) and Zagreb-3

(a)

(b)

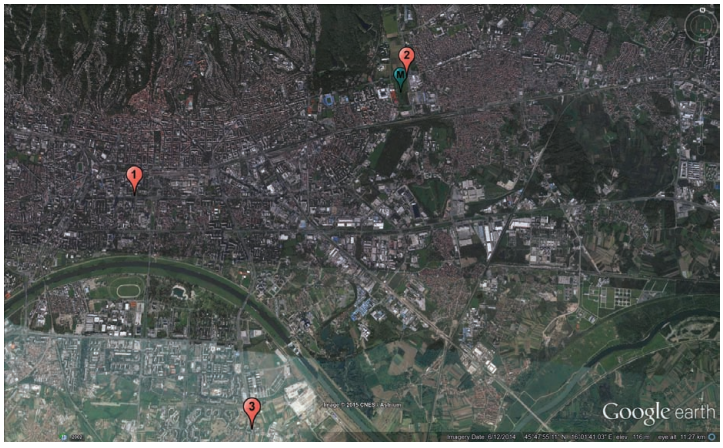

Figure 1. Satellite view showing larger area including Croatia and Balcan area (*a*). Satellite view of Zagreb (*b*) (source: Google Earth). Positions of the measuring sites are indicated by bubbles, with red and blue bubbles corresponding to the air pollution measuring sites and meteorological measuring site, respectively. Zagreb-1 is in the urban centre of town, near a road subject to high traffic densities. Zagreb-2 and Zagreb-3 are located in suburban areas of eastern and southern Zagreb, respectively. The Zagreb-Maksimir Observatory is approximately 200 m from Zagreb-2.

(45° 45' 53.85" N, 16° 0' 24.31" E), were analysed (Figs. 1a and 1b and Tab. 2) for pollutants including CO, SO_2, NO_2 and PM_{10}. At each site, air pollution is measured using an intake collector for air sampling, with the latter then transported towards the measuring equipment placed inside a measuring container. Collector entries are placed at heights ranging from 1.5 to 4 m above the ground surface. Therefore, all parameters are measured within the lowermost portion of the atmospheric boundary layer, which is governed by turbulent processes.

Zagreb-1 is located in the urban centre of Zagreb, near a road subject to heavy traffic levels. The representativeness of a station can be expressed in terms of a radius (km) around the measuring site for which the concentration of a particular pollutant is approximately constant (Mihajlović et al., 2012). In the present study, the radius for Zagreb-1 is less than 1 km. Zagreb-2, located in a suburban area of eastern Zagreb, has a representativeness radius of 1–10 km. Finally, Zagreb-3 is located in suburban southern Zagreb, approximately 1.5 km west of the Jakuševac refuse dump. The representativeness radius of Zagreb-3 is 1–10 km. At all three monitoring sites, traffic is the main source of pollution. The approximate distances between Zagreb-1, -2, -3 and Zagreb-Maksimir are 5 km, 200 m and 7 km, respectively. We note that the surface meteorological station and the air quality stations are not collocated. Nevertheless, meteorological site is the main (synoptic) meteorological station of Meteorological and Hydrological Service of Croatia and thus, it is considered representative for the greater Zagreb area.

The collected air pollution data were subject to minor processing (quality control) in order to remove any inconsistencies. Due to measurement errors, a few negative pollutant concentration values appeared occasionally in the raw data. All such values were removed from further analysis (i.e. replaced with a "no data" annotation). Since the pollution data were to be modelled on a log scale, zero values would also have caused problems. Indeed, it is very unlikely that urban air in Zagreb would be absolutely clean (a pollutant concentration equal to zero). Any observed zero values, which were presumably caused by the limited sensitivity of the measurement instruments to low pollutant levels rather than reflecting actual nil concentrations, were here thus also considered erroneous. Finally, the raw data series also contained some periods during which no

Table 2. Summary of pollutant concentration statistics for Zagreb-1, -2 and -3, respectively.

		Minimum	Mean	Median	Maximum	Standard deviation
CO ($mg\,m^{-3}$)	ZG1	0.01	0.59	0.48	2.32	0.39
	ZG2	0.01	0.52	0.42	2.07	0.35
	ZG3	0.07	0.47	0.38	1.75	0.30
SO_2 ($\mu g\,m^{-3}$)	ZG1	0.01	5.06	4.35	28.42	5.06
	ZG2	0.01	8.50	7.47	36.94	6.04
	ZG3	0.01	7.34	4.72	36.92	7.46
NO_2 ($\mu g\,m^{-3}$)	ZG1	0.01	39.17	35.75	146.50	23.16
	ZG2	0.01	34.36	31.43	132.40	21.09
	ZG3	0.01	28.70	24.23	115.10	19.39
PM_{10} ($\mu g\,m^{-3}$)	ZG1	0.01	31.18	26.14	122.60	21.14
	ZG2	0.01	30.77	26.41	111.90	19.20
	ZG3	0.01	30.84	25.81	120.60	20.82

observations were made, which were caused by instrument malfunction and/or repair. For each pollutant only data available simultaneously for all three sites were considered for further analysis. After omitting all incorrect data, between 54% and 67% of initial data was available for each pollutant.

Table 2 presents a basic statistical overview of air pollution values after the application of the data quality control process. Corresponding time series of hourly concentrations are not shown. Overall, correlation coefficients between concentrations of the same pollutant recorded at two different measuring sites varied from 0.7 to over 0.9 (depending on the pollutant and measurement site pair).

3. Methodology

3.1. Generalized additive models

Generalized additive models are regression models in which smoothing splines are used instead of linear coefficients for covariates (Hastie and Tibshirani, 1990; Hastie et al., 2009). This approach has been found particularly useful for handling the complex nonlinearity associated with air pollution research (e.g. Dominici et al., 2002; Schlink et al., 2006; Carslaw et al., 2007). Additive models employed in the context of concentration time series can be written in the following form (Hastie and Tibshirani, 1990):

$$log\left(y_i\right) = s_0 + \sum_{j=1}^{p} s_j\left(x_{ji}\right) + \varepsilon_i, \tag{3}$$

where i varies from 1 to n and n is the number of observations, j varies from 1 to p, where p is the number of predictor variables in the model, y_i is the i-th air pollution concentration, s_0 is the overall mean of the response, $s_j(x_{ji})$ is the smoothing function of the i-th value of covariate j, and ε_i is the i-th residual. The right-hand side of Eq. 3 (excluding residuals) describes the fitted values obtained by the model. The residuals are assumed to be normally distributed with $mean(\varepsilon_i) = 0$, variance $var(\varepsilon_i) = \sigma^2$ and are independent of the fitted values. (We note that exact value of variance is needed only for forecast purposes, which is not the case in the present study. Therefore, we will not calculate it.)

Described in Hastie and Tibshirani (1990) and Hastie et al. (2009), the employed logarithmic transformation has also been used elsewhere (e.g. Schlink et al., 2003) to ensure both homoscedasticity of data (in the present case homogeneity of variance for all four pollutants and all three sites) and that all concentrations are positive on the original scale. The model given by Eq. 3 is additive on the log scale and can be transformed back to the original scale, as given by Eq. 4:

$$y_i = S_{0i}S_{1i}\left(x_{1i}\right)...S_{pi}\left(x_{pi}\right)E_i, \tag{4}$$

where $S(*) = exp(s(*))$ and $E_i = exp(\varepsilon_i)$.

This model structure is easy to interpret, since each predictor enters the model separately. However, potentially important interactions between two predictors (e.g. those between wind speed and wind direction) are not handled in the present study.

3.2. Model development

Each of the four pollutants were modelled separately using the model given by Eq. 3, with seven meteorological variables and three time variables applied via the *gam* modelling function in the R environment (R Development Core, 2009) for statistical computing inside the *mgcv* package (Wood, 2006). As mentioned above, traffic density data for Zagreb were unavailable and thus three temporal variables were included to roughly account for traffic density and industrial emissions. The predictor *hour of the day* (hereafter *HD*) was used to account for diurnal variation not taken into account by other predictors, with values ranging from 1 (corresponding to the time interval between 00 and 01 local standard time, LST) to 24 (corresponding to the time interval between 23 and 24 LST). Similarly, the variable *day of the week* (*DW*) was employed to account for weekly variations and varied from 1 (Monday) to 7 (Sunday). The variable *day number* (*DN)* varied from 1 (1 Jan 2006) to 2557 (31 Dec 2012) and was included to account for seasonal variations. If correlated at all, the available predictors were at most moderately correlated (absolute values of correlation coefficients ranging from 0.00 to 0.58, not shown here). Although Aldrin and Haff (2005) inspected colinearity based on correlation coefficients solely, in the present study we additionally calculated variance inflation factor (VIF) (Zuur et al., 2009). For all variables VIF values were lower than 3, and they were ranging from about 1.0 for the day of the week (*DW*) to 2.4 for the relative humidity. Thus, we assumed that variables are not collinear, and that a regression method could be applied.

The first step in the selection of the individual model was to fit a preliminary model comprising time variables only (Eq. 5):

$$log\left(y_i\right) = s_0 + s\left(HD, k = 12\right) + s\left(DW, k = 7\right) + s\left(DN, k = 28\right) + \varepsilon_i, \qquad (5)$$

where k is the maximum number of knots used by the smoother. Since the employed air pollution data were seasonal, a predetermined smoothing parameter k was used for smoother construction. The smoothing spline for *HD* had 12 knots and was employed to account for processes on time scales larger than two hours, where the selected time scale threshold of 2 hours was optional. The variables *DW* and *DN* had 7 and 28 knots respectively, one for each day and each of the four seasons in the seven-year study period. To check the adequacy of these assumptions, residual histograms and scatterplots were examined (not shown here). Both the histograms and scatterplots confirmed that the determined numbers of knots were appropriate. That is, the majority of residuals grouped around zero, while scatterplots of fitted values *vs*. residuals did not suggest any functional dependency (see Section 4.4.).

Apart from Akaike's Information Criterion (AIC), other methods for selecting final model are also applicable (e.g. Zuur et al., 2009). We decided to select the simplest and the least computational time consuming approach (Wood, 2001). All predictor variables of interest were added to the preliminary model as follows (Eq. 6):

$$log\left(y_i\right) = s_0 + s\left(HD, k=12\right) + s\left(DW, k=7\right) + s\left(DN, k=28\right) + s\left(temp., k\right) + s\ \ press\ \ k\quad s\ \ re$$
$$+ s\left(press., k\right) + s\left(rel.hum., k\right) + s\left(speed, k\right) + s\left(dir., k\right) + s\left(prec.4h, k\right) + s\ \ prec.week, k$$
$$+ s\left(prec.week, k\right) + \varepsilon_i \,. \tag{6}$$

The unwanted terms were thereafter removed using three Wood's criteria, that is, based on the answers to the following three questions within the frame of the *mgcv* package:

1) Are the estimated degrees of freedom (EDF) of a particular term close to 1?
2) Does the plotted 95% confidence band of a particular term include zero everywhere?
3) Does the GCV score drop when the term is omitted?

If the answer to all three questions is positive then the term should be omitted. However, in the present study, the answers to all three questions were negative, that is, all predictors were associated with new statistically significant information and thus were kept in the model.

Additionally, we used a hypothesis testing procedure in order to confirm the selection procedure. The significance of each variable in GAM was determined by means of analysis of variance (one-sided F-test). The test confirmed that all terms were significant at the 95% level (not shown here). This is further corroborated by narrow confidence bands depicted in Figs. 3–11.

The smoothness of each function s_j in the model determined by Eq. 6 is controlled by a smoothness parameter, here expressed by the maximum number of knots for each smoother (k). This value must be chosen before the smoothing function is estimated. In the present study, the maximum number of knots for the time variables had already been selected, as discussed above. However, the maximum number of knots for the meteorological variables was still to be determined. This particular parameter should be large enough so that the main processes which govern concentration values are included in the model. As the number of knots increases, the function becomes less smooth but better fitted to data. However, if the number of knots is too large, the results are not improved and computational time is prolonged. The easiest method with which to choose the optimal knot number is forward validation, which is a special form of cross validation. In the present study, forward validation for each pollutant was based on hourly predictions of concentrations for Zagreb-1, one day in advance. For each day and for a given maximum number of knots, the model was re-estimated using the data up to the day before the day in question. The logarithm of hourly concentrations for

the next day was then predicted, assuming that the predictors for that day were known. The prediction was afterwards compared to the logarithm of the actual (observed) value and the hourly prediction errors calculated. This procedure was repeated for each day of the chosen period and for each pollutant. Finally, the root mean squared error (RMSE) of the prediction was calculated. The entire procedure was then repeated for various choices of maximum knot number. Since it would have been too time consuming to vary the number of knots for each meteorological variable separately, we assumed that all variables had the same maximum number of knots. Within the framework of the *mgcv* package, after model computation, one can determine the actual number of knots used for each smoothing function. If the actual used number is smaller than $k-1$, k can be considered sufficiently large (Wood, 2001). The minimum RMSE for each pollutant corresponded to $k=9$. Insight into the actual number of knots for all predictors showed that a value of $k=5$ was large enough only for the two precipitation variables. Accordingly, the maximum number of knots for these two variables was reduced. Further, it was assumed that the maximum number of knots for the other two sites was equal to that obtained for Zagreb-1.

The employed residuals ε_i were in practice autocorrelated and as such could be described by an autoregressive model (AR). Employment of an AR is important for prediction and has little effect on the estimation of s functions. According to Liang and Zeger (1986), the produced estimates are consistent even though the autocorrelation is ignored. As a result, we chose to omit a model for the residuals. However, for the purpose of concentration forecasting, appropriate modelling of residuals would be necessary.

Finally, we constrained the GAM by assuming a gamma distribution. A gamma distribution was employed because the concentration values were > 0 and their distribution skew positive. The gamma distribution considered appropriate for the observed concentration data (not shown here) generally reduces model errors for low concentration values.

We then calculated the squared correlation coefficient (explained variance; R^2) for each model on the log scale. The aggregated impacts of local meteorological parameters on each pollutant were assessed in terms of the difference in R^2 values between the model determined by Eq. 5 and the model determined by Eq. 6. Furthermore, a measure of the relative importance of the predictor variables was calculated, with the partial response graphs plotted as percentage values. The employed measure of relative importance is expressed by Eq. 7 as follows:

$$100\frac{\hat{\sigma}_{-j}^2-\hat{\sigma}^2}{\sum_{j=1}^{p}\hat{\sigma}_{-j}^2-p\hat{\sigma}^2}\,, \tag{7}$$

where $\hat{\sigma}_{-j}^2$ is the unexplained variation (i.e. that not explained by the model when the effect of the j-th predictor variable is ignored; a minus sign in front of the subscript j is written in order to emphasise that j-th variable is omitted) and

$\hat{\sigma}^2$ is the unexplained variation for the model given by Eq. 6. Summing the individual differences of all predictors gives the denominator. The partial effect is expressed by Eq. 8 as follows:

$$100\frac{S_j\left(x_j\right)}{S_j\left(x_{j,ref}\right)}\ , \tag{8}$$

where x_{ref} is a reference value. Since reference values do not affect results associated with partial effects qualitatively, they were chosen subjectively for every predictor. Namely, values 0 °C (temperature), 1000 hPa (pressure), 97% (relative humidity), 0 deg (wind direction) and 0 m s^{-1} (wind speed), 0 mm (precipitation in the last four hours), 0 mm (precipitation during the last week), 1 (hour of the day), 1 (day number), and 1 (the day of the week) were selected. A more detailed description of relative importance and partial response can be found in Aldrin and Haff (2005). In the final section we present a statistical assessment of the model.

4. Results and discussion

4.1. The explained variation

Based on the data described in Section 2, the model given by Eq. 6 was employed to estimate the levels of four pollutants recorded at three different locations in Zagreb. Table 3 displays the explained variation for the entire model (R^2) and for meteorological variables only (R^2(met. var.)), where R^2(met. var.) corre-

Table 3. Explained variation R^2 for each pollutant concentration modelled on the log scale (third column). The fourth column (R^2(met. var.)) shows the explained variation for meteorological variables only, that is, for the model defined by the difference between the models given by Eq. 6 and Eq. 5, respectively. The percentage value displayed in parentheses is the ratio of R^2(met. var.) and R^2. As confirmed by a t-test, all correlation coefficients (R) are significant at the 0.05 significance level.

Pollutant	Measuring site	R^2	R^2 (met. var.)
CO (mg m^{-3})	ZG1	0.55	0.35 (63.64%)
	ZG2	0.57	0.33 (58.89%)
	ZG3	0.60	0.45 (75.00%)
SO_2 (μg m^{-3})	ZG1	0.50	0.20 (40.00%)
	ZG2	0.54	0.15 (27.78%)
	ZG3	0.51	0.14 (27.45%)
NO_2 (μg m^{-3})	ZG1	0.45	0.15 (33.33%)
	ZG2	0.48	0.14 (29.17%)
	ZG3	0.43	0.25 (58.14%)
PM_{10} (μg m^{-3})	ZG1	0.43	0.25 (58.14%)
	ZG2	0.42	0.19 (45.24%)
	ZG3	0.41	0.24 (58.54%)

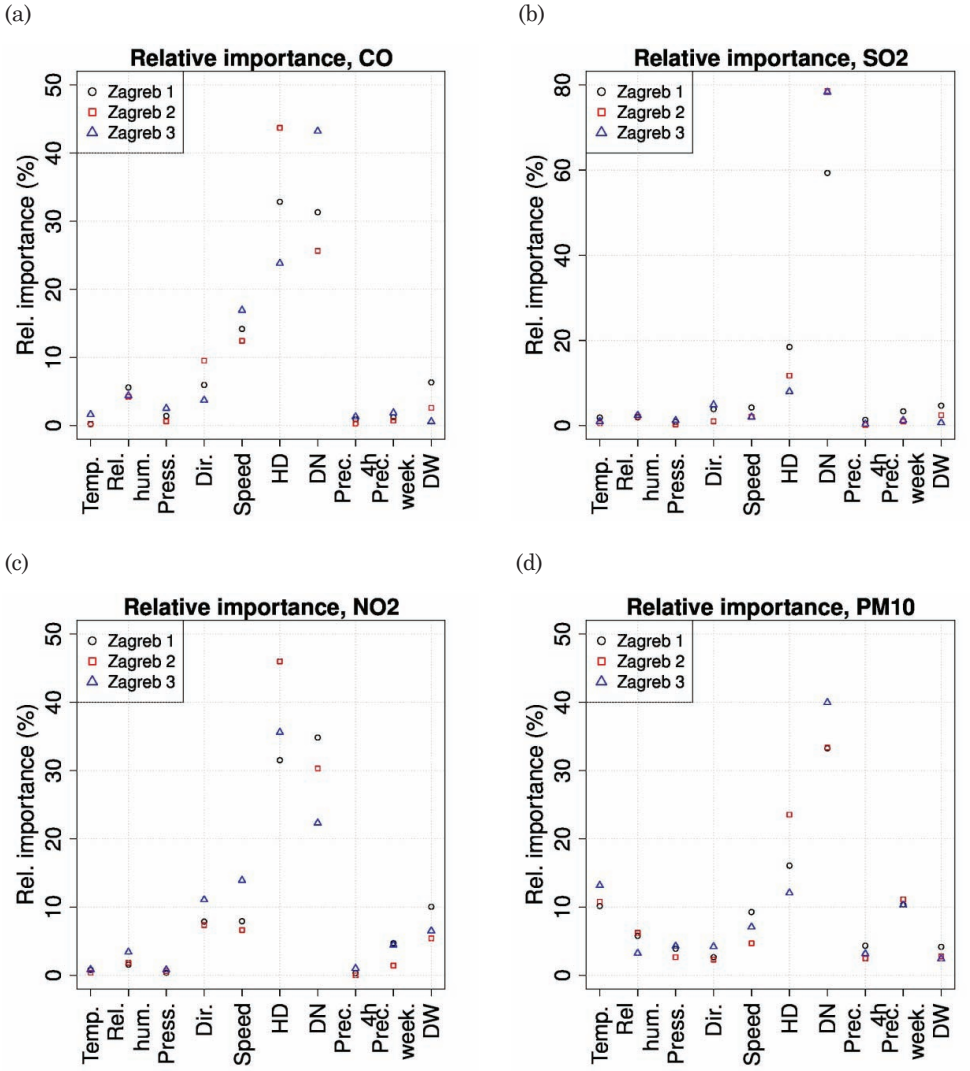

Figure 2. The relative importance of predictors for each of the four pollutants and all three measuring sites. Predictor relative importance is the proportion (in %) of the variation explained by the j-th predictor variable in the model.

sponds to the explained variation of the new model, defined as the difference between the models given by Eq. 6 and Eq. 5, respectively. The percentage in the fourth column (shown in parentheses) is the ratio of R^2(met. var.) and R^2. Values between 0.41 and 0.60 indicate that the model explains most of the variation, although considerable variation remains unaccounted for. The percentage in the

fourth column reveals that meteorology, at the local scale at least, is a relatively strong driver of air quality in Zagreb. The highest values of R^2 were obtained for CO.

4.2. Relative importance

Figure 2 illustrates the relative importance of each predictor, as expressed in terms of the percentage of the variation explained by the *j*-th model predictor. According to these results, variables *HD* and *DN* are the most important predictors for all three measuring sites and for all four pollutants, and in comparison

Table 4. Summary of variable relative importance for CO, where 1 and 10 correspond to the most and the least important variable, respectively.

CO	*Temp.*	*Press.*	*Rel. hum.*	*Dir.*	*Speed*	*Prec. 4 h*	*Prec. Week*	*HD*	*DN*	*DW*
Zagreb-1	9	7	6	5	3	8	10	1	2	4
Zagreb-2	10	8	5	4	3	9	7	1	2	6
Zagreb-3	7	6	4	5	3	9	8	2	1	10

Table 5. Same as Tab. 4 but for SO_2.

SO_2	*Temp.*	*Press.*	*Rel. hum.*	*Dir.*	*Speed*	*Prec. 4 h*	*Prec. Week*	*HD*	*DN*	*DW*
Zagreb-1	8	10	7	5	4	9	6	2	1	3
Zagreb-2	8	9	4	6	5	10	7	2	1	3
Zagreb-3	8	7	5	3	4	10	6	2	1	9

Table 6. Same as Tab. 4 but for NO_2.

NO_2	*Temp.*	*Press.*	*Rel. hum.*	*Dir.*	*Speed*	*Prec. 4 h*	*Prec. Week*	*HD*	*DN*	*DW*
Zagreb-1	8	9	7	5	4	10	6	2	1	3
Zagreb-2	9	8	6	3	4	10	7	1	2	5
Zagreb-3	8	10	7	4	3	8	6	1	2	5

Table 7. Same as Tab. 4 but for PM_{10}.

PM_{10}	*Temp.*	*Press.*	*Rel. hum.*	*Dir.*	*Speed*	*Prec. 4 h*	*Prec. Week*	*HD*	*DN*	*DW*
Zagreb-1	4	9	6	10	5	7	3	2	1	8
Zagreb-2	4	8	5	10	6	9	3	2	1	7
Zagreb-3	2	6	8	7	5	9	4	3	1	10

with meteorological variables, they are much more important (since they implicitly account for emissions). Meteorological variables have the largest impact at Zagreb-3. Although temperature has a strong influence on PM_{10}, its influence on other pollutants is much weaker. As expected, based on a number of previous studies (e.g. Aldrin and Haff, 2005), relative humidity and air pressure have some effect on all pollutants. Wind speed and direction have a large effect on CO, NO_2 and PM_{10}, a pattern also found elsewhere (e.g. Levy et al., 2003). Both precipitation variables have some importance for PM_{10} but have little effect on the other three pollutants. In summary, meteorological variables seem to have the strongest and weakest impact on CO and SO_2, respectively. Tables 4–7 display the order of predictor relative importance, where 1 and 10 correspond to the most important and least important variable, respectively.

4.3. Partial effects

The actual estimated nonlinear smooth curves for Zagreb-3 are shown in Figs. 3–11, with the displayed curves illustrating the partial effects of each predictor on pollutant concentration. Curves were set to 100 at the chosen reference value of predictor *x*. The dashes adjacent to the *x*-axis indicate whether the value of the variable *x* appears in the analysis. Dashed lines show 95% confidence intervals (that is, they indicate how frequently the value of interest is repeated).

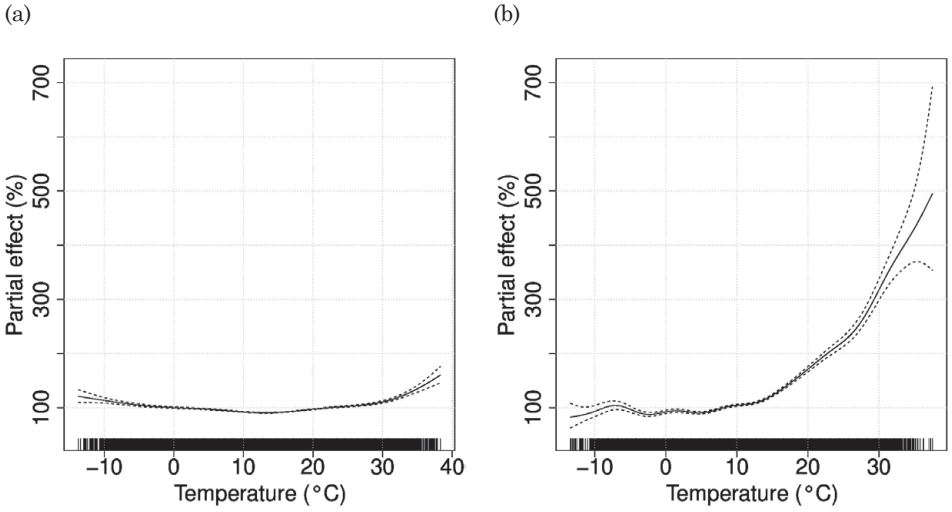

Figure 3. Actual estimated non-linear smooth curves for the Zagreb-3 measuring site, depicting the partial effects of each predictor on pollutant concentrations. The curve is set to 100 at the chosen reference value of the predictor *x*. Dashes adjacent to the x-axis appear only if variable *x* appears in the analysis. Dashed lines show 95% confidence intervals. Partial effects of temperature on CO and PM_{10} are shown in panels (*a*) and (*b*), respectively. The results for SO_2 and NO_2 are similar to those obtained for CO.

Curve interpretation (e.g. temperature and PM_{10}) can be expressed as follows: When the temperature increases from 10 °C to 30 °C and all other predictors are constant, PM_{10} concentration increases from 100 to 300. Similar curves were obtained for Zagreb-1 and -2 (not shown).

Figures 3–11 are organised in the following manner: If only one panel is shown, the results for all four pollutants are similar. In this case, the results for only one pollutant are depicted for Zagreb-3. If two or three panels are shown, the results for the different pollutants vary. The effect of temperature is similar for CO, SO_2 and NO_2 (Fig. 3a displays only CO data). Temperatures below 0 °C are associated with the occurrence of higher pollutant concentrations, which is consistent with increased fuel consumption in winter months. Temperatures above 10 °C occur together with a slight increase in CO, SO_2 and NO_2 concentrations, but a sudden increase in PM_{10} concentrations (Fig. 3b). This phenomenon is likely caused by rising SO_2 levels, since the oxidation of SO_2 is known to produce PM_{10} (e.g. Ryaboshapko et al., 1996). We note however, that the increase in PM_{10} levels at higher temperatures observed in the present study is in contrast to the result found in Norway by Aldrin and Haff (2005).

Generally, the results obtained regarding the impact of mean sea level pressure between 990 and 1030 hPa on pollutant concentrations were as expected. Figure 4a displays the results for CO, with Fig. 4b showing those for SO_2; the data shown in these two figures are similar to those obtained for NO_2 and PM_{10}, respectively. During high pressure events, the atmosphere is, in general, statically stable. As such atmospheric stability limits vertical fluxes, near-ground

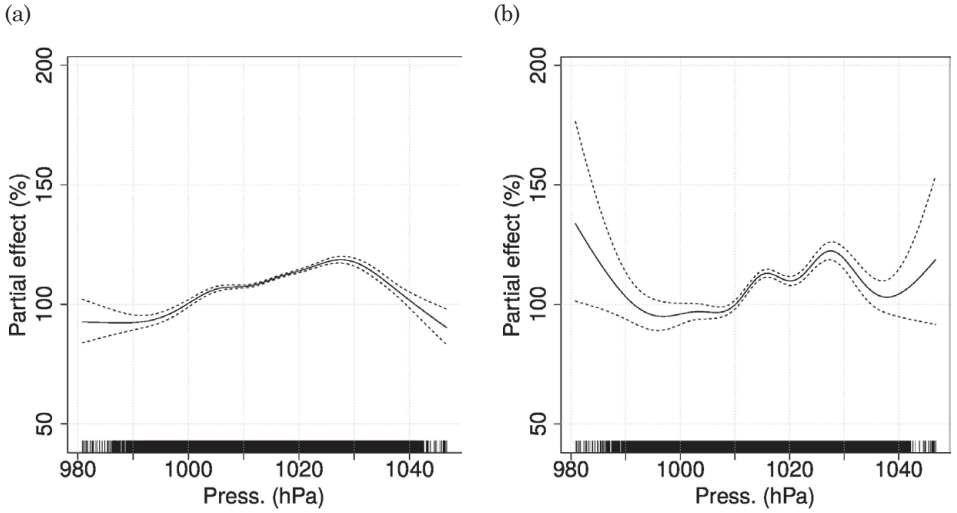

Figure 4. Partial effect of mean sea-level pressure on CO (*a*) and SO_2 (*b*). The results for NO_2 and PM_{10} are similar to those obtained for CO and SO_2, respectively.

pollutant concentrations are higher. During low pressure events, the atmosphere is unstable and pollutants can be transported by vertical fluxes, which are thus associated with pollution dilution (e.g. Prtenjak et al., 2009; Pearce et al., 2011). However, as illustrated in Fig. 4b, the impact of pressure is not clear for extremely low and extremely high values, especially for SO_2 and PM_{10}. In any case, the obtained result cannot be considered reliable since only a few episodes of extremely low or high pressure were recorded in Zagreb during the study period.

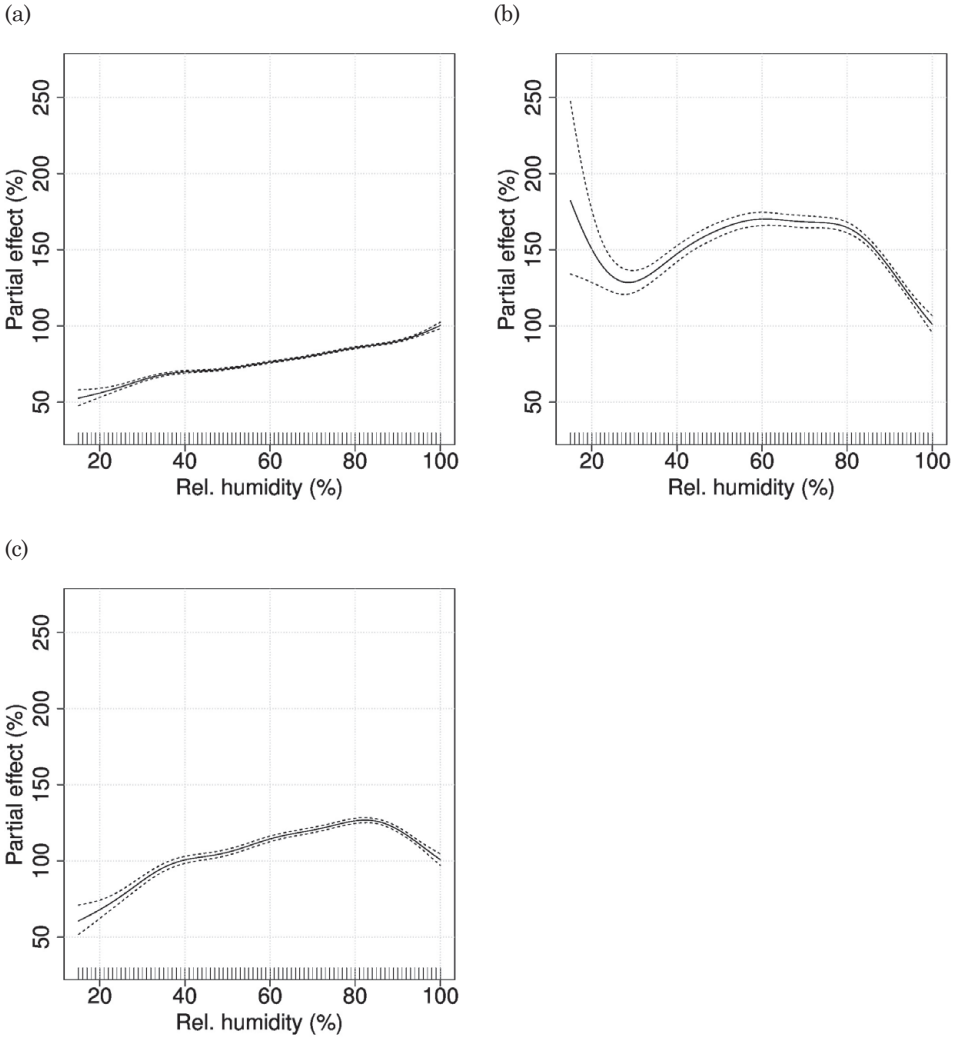

Figure 5. Partial effect of relative humidity on CO (*a*), SO_2 (*b*) and NO_2 (*c*). The results for PM_{10} are similar to those obtained for NO_2.

The effect on SO_2 and NO_2 (Figs. 5b and 5c) is obvious for high values of relative humidity. The effect on NO_2 in particular is similar to that on PM_{10}, with concentrations initially increasing for rising relative humidity levels of up to about 80%, followed by a concentration decrease beyond this threshold. This drop in concentration is likely the result of pollutant wet deposition, in which SO_2 and NO_2 react with water vapour (H_2O(g)) in the atmosphere to produce acid rain. A similar decrease in PM_1 concentrations with high relative humidity was also found by Klaić et al. (2012) for residential Zagreb. However, the increase in SO_2 concentrations for low values of relative humidity is not clear. The effect of high relative humidity values is the opposite for CO. CO reacts with H_2O to produce CO_2, a reaction thus representing a CO sink. In contrast, the oxidation of methane (CH_4) with OH radicals is the main source of atmospheric CO (e.g. Levy et al., 1971). As the oxidation of CH_4 in the atmosphere is faster than the reaction of CO with water vapour and as there are more OH radicals available at higher relative humidity, CO concentrations increase.

Figure 6 shows the results obtained regarding the impact of wind direction. All pollutants exhibit maximum concentrations for directions around 330° and minimum concentrations at around 240°. CO (Fig. 6a), NO_2 and PM_{10} all show secondary maxima for E–NE wind (70–100°), whereas SO_2 (Fig. 6b) has a secondary maximum for wind directions of around 30°. It should be noted that this latter secondary maximum coincides with the secondary minima for CO, NO_2 and PM_{10}. As Zagreb-3 is located south-west of the city's industrial zone, this particular measuring site is subject to pollutant transport by north-easterly

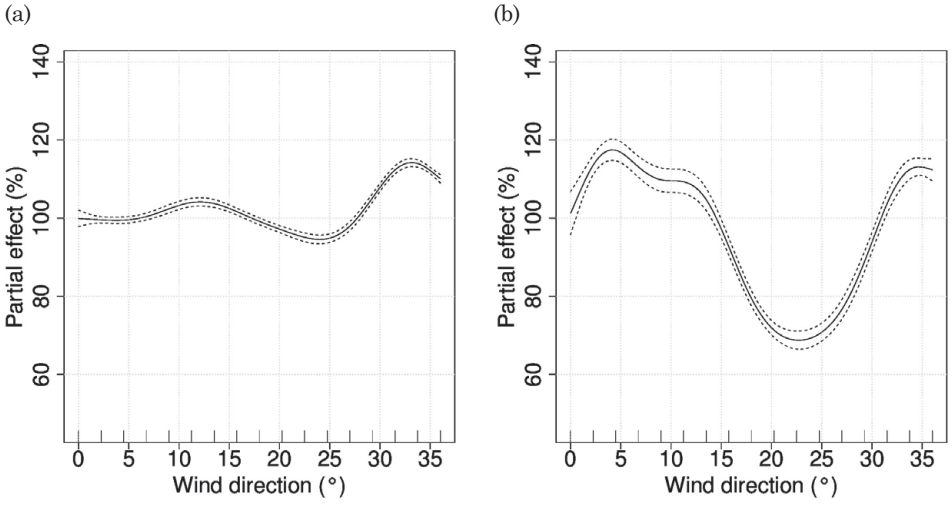

Figure 6. Partial effect of wind direction on CO (*a*) and SO_2 (*b*). The results for NO_2 and PM_{10} are similar to those obtained for CO.

winds. The effect of north-north-easterly flows on CO (as well as NO_2 and PM_{10}) and SO_2 concentrations are opposite (compare Figs. 6a and 6b), with the impact on CO being minimal and on SO_2 maximal. Furthermore, we believe that the pattern found for W–NW flows, which have a large influence on the concentrations of all four pollutants, might be due to urban sources found in western Zagreb (the Špansko industrial zone). Another possible explanation for the influence of NW flows is perhaps the presence of older particles transported toward the measuring site via long-range transport (Klaić et al., 2012). However, it remains rather difficult to distinguish the relative impacts of these processes. Considering the position of Zagreb-3 with respect to the city itself, one would expect N flows to be associated with maximum pollutant concentrations due to the nighttime down-slope winds established on southern slopes of Mount Medvednica (e.g. Klaić et al., 2002, 2003), which transport urban polluted air toward the measuring site. Similarly, daytime up-slope southerly winds should therefore contribute to a decrease in concentrations. The results, particularly for SO_2 (Fig. 6b), confirm such a pattern. It should be noted that the impact of the above-mentioned wind directions on pollutant concentrations is the opposite to that found by Klaić et al. (2012) due to the different location of the selected measuring site.

Wind speeds at 2–4 m above the ground are presumably weaker than at 10 m due to stronger surface friction. Pollutant concentrations and wind speeds recorded in the present study exhibit a mainly negative correlation (Fig. 7), clearly reflecting the effect of local ventilation. For weaker winds, the decrease in concentration with increasing wind speed suggests that ventilation predominates over pollutant advection. Low wind speeds and calm periods may indicate stable atmospheric conditions and thus pollutant concentrations are higher during such episodes. The results obtained for CO (Fig. 7a) and PM_{10} (not shown) exhibit similar patterns. Maximum SO_2 concentrations occur at wind speeds of around 6 m s^{-1} (Fig. 7b) and maximum NO_2 concentrations (Fig. 7c) at 8 m s^{-1}. Therefore, we believe that for stronger winds the wind speed effect is the opposite of that recorded for weaker winds; that is, pollution transport to the site via advection (regional and/or long-range transport of pollutants) predominates over local ventilation, an effect also found for western winds over Zagreb by Klaić et al. (2012). In order to corroborate this hypothesis, the relationships between pollutant concentrations and both wind speed and direction must be determined. Since in the present study a certain amount of measurement data was missing, future work will aim to improve the model by including the relationship between wind speed and wind direction.

As the obtained results regarding precipitation in the preceding four hours and preceding week are similar, only curves for the former variable are shown for CO and SO_2 (Figs. 8a and 8b). At weaker precipitation intensities, the presence of precipitation decreases concentrations of all four studied pollutants, likely due to pollutant scavenging via rain drops or snow. Conversely, large precipitation intensities seem to correspond to higher pollutant concentrations,

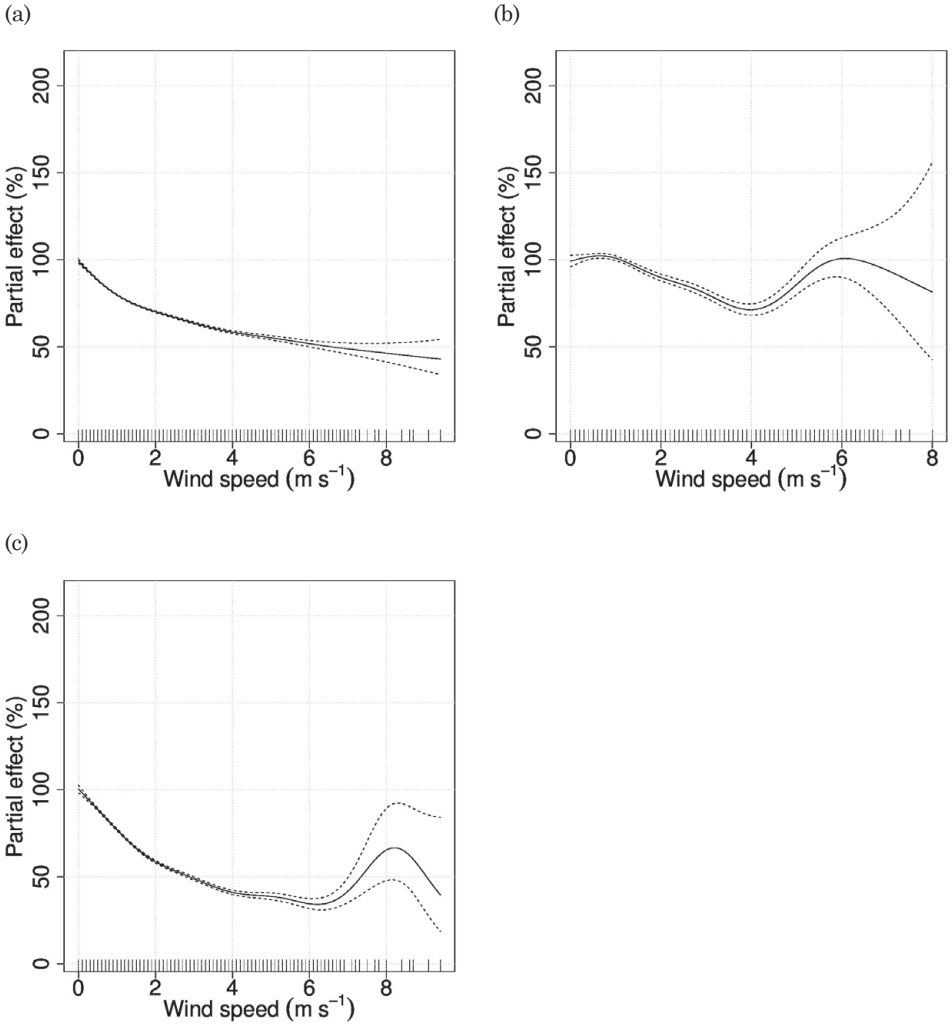

Figure 7. Partial effect of wind speed on CO (*a*), SO_2 (*b*) and NO_2 (*c*). The effect on PM_{10} is similar to that on CO.

a pattern also found by Aldrin and Haff (2005). It should be noted however, that higher intensity precipitation events (> 3 mm h^{-1}) occurred much more rarely than those of weaker intensity during the study period.

The time variable *HD* roughly accounts for the impact of traffic (Fig. 9). At the study sites, pollutant concentrations fall to a minimum at around 5 LST and increase until 10 LST, which corresponds to the morning rush hour. A secondary minimum can be observed at around 17 LST and a secondary maximum at

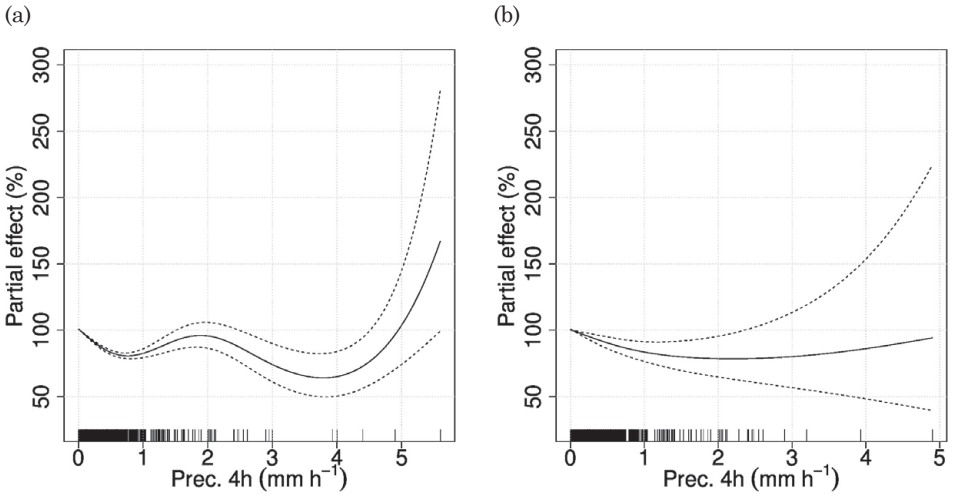

Figure 8. Partial effect of precipitation averaged over the preceding four hours on CO (*a*) and SO_2 (*b*). The results for NO_2 and PM_{10} are similar to those for CO.

around 22 LST. Generally, concentrations are higher during morning hours than during the afternoon. *HD* could also represent the effects of diurnal variation in boundary layer height (BLH). The fact that the BLH is shallower in the morning than the afternoon could thus also explain the recording of higher concentrations during morning hours.

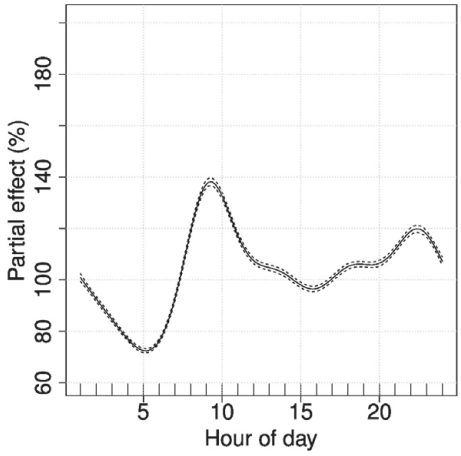

Figure 9. Partial effect of hour of day (*HD*) on CO. The results for NO_2, SO_2 and PM_{10} are similar to those obtained for CO.

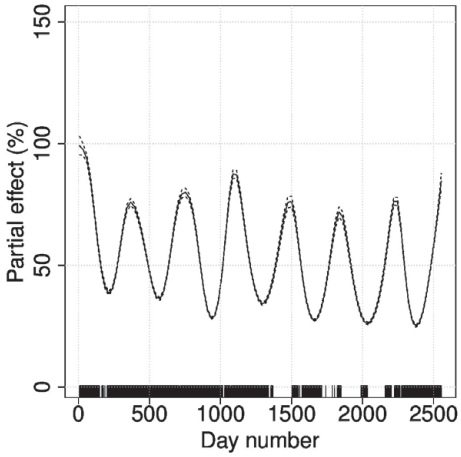

Figure 10. Partial effect of day number (*DN*) on CO. The results for NO_2, SO_2 and PM_{10} are similar to those obtained for CO.

Analysis of the day number (*DN*) curve reveals a maximum in winter and a minimum in summer for every year during the study period (Fig. 10), suggesting the occurrence of a wintertime phenomenon which results in elevated pollutant concentrations. One possible reason for these higher concentrations is an increase in emissions associated with fuel consumption (Chaloulakou et al., 2003).

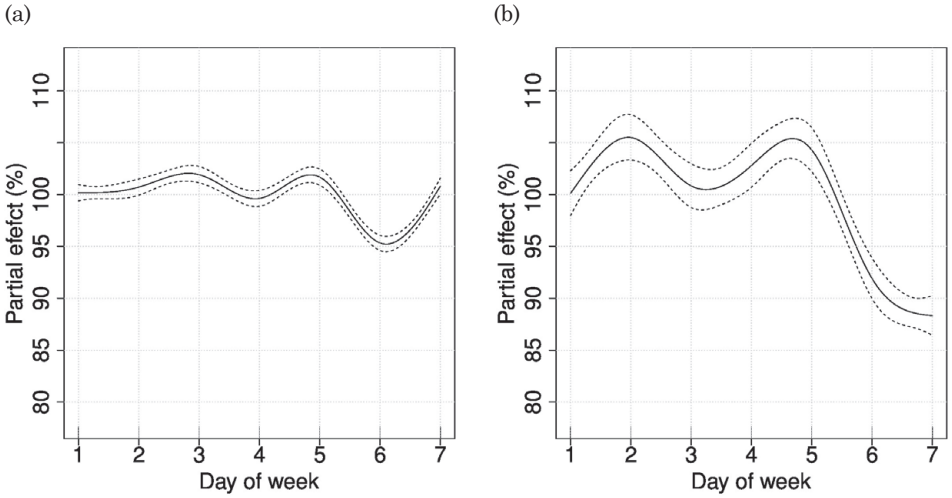

Figure 11. Partial effect of day of the week (*DW*) on CO (*a*) and SO_2 (*b*). The results for NO_2 and PM_{10} are similar to those obtained for SO_2.

Another possible reason could be related to the fact that the atmosphere is generally less stable in summer than in winter. Accordingly, pollution is transported via vertical fluxes to higher levels in summer and thus surface concentrations should generally be lower (Klać et al., 2012).

From Monday to Thursday (inclusive), pollutant concentrations remain more or less unchanged (Fig. 11). On Friday, concentrations increase, probably due to people travelling away from Zagreb (by car, bus, etc.) (Figs. 11a and 11b). During the weekend, concentrations decrease, a finding in accordance with those of Chaloulakou et al. (2003), with only CO levels increasing on Sunday (Fig. 11a).

Analysis of the results reveals a clear distinction between the effects of human activity (anthropogenic emissions) and those of weather conditions, with the former roughly incorporated in the model via the use of temporal variables. Whereas the temporal variables data (Figs. 9–11) exhibit a noticeable dependence on anthropogenic activity, the impact of meteorological variables varies under different synoptic conditions. Stable atmospheric conditions, such as those associated with high pressure, low winds and poor vertical mixing, favour the occurrence of higher concentrations of air pollutants (Prtenjak et al., 2009; Klaić et al., 2012) which in turn can affect human health. Conversely, unstable atmospheric conditions (i.e., low pressure, high wind speeds and precipitation) tend to decrease pollutant concentrations, with the dominant convection processes diluting levels within the boundary layer.

4.4. Assessing the fit of the GAM

The use of a GAM in combination with partial effect plots proved an efficient method with which to characterise the relationship between individual meteorological and time variables (the former representing local weather conditions and the latter emissions) with well as air pollution.

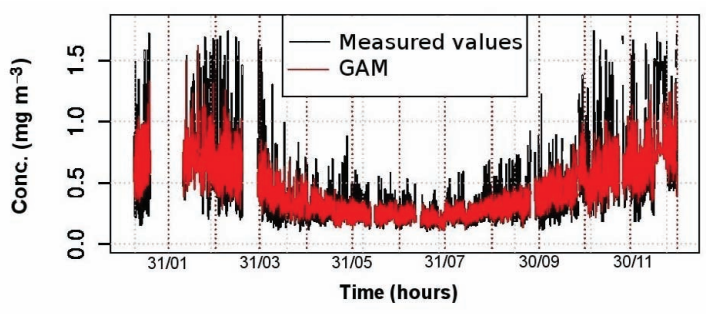

Figure 12. Difference between measured (*black*) and modelled concentrations (*red*) of CO at Zagreb-3 during 2012. Modelled values were obtained as a sum of the nonlinear functions of the meteorological and several time variables. Comparison is presented at the original scale. Dashed, red vertical lines indicate the end of each month.

Figure 12 illustrates the difference between measured and modelled concentrations of CO at Zagreb-3; other pollutant and measuring site data are not shown as the results are similar to those presented in this figure. The comparison is presented at the original scale. As mentioned above, no traffic data were included in the model and thus one would expect the measured values to be slightly underestimated. Although the presented model underestimated the amplitude of pollutant concentration changes, it was able to describe temporal variation rather well.

The final step is to estimate the model's efficiency in describing the obtained data. Figure 13 illustrates the relationship between modelled and measured values of CO concentrations at Zagreb-3; again, other pollutant and measuring site data are not shown as they are similar to those presented in this figure. The red line represents the theoretical linear relationship (modelled=observed). For low concentrations, values follow a linear relationship, whereas for higher concentrations the scattering is more evident.

Residual plots also provide an insightful method which characterises model efficacy (for residuals see Section 3). Theoretically, mean residual values should be equal to 0. In the present study, those for CO, SO_2, NO_2 and PM_{10} were 0.00, –0.04, –0.14 and 0.02, respectively. Similarly, the correlation coefficients between residuals and fitted values should theoretically also be 0, with values of 0.00, –0.03, –0.04 and 0.00 for the above four pollutants recorded here. Given these figures, the employed model can be considered successful. Figure 14 clearly shows that the majority of residuals group around zero, as expected. Whereas

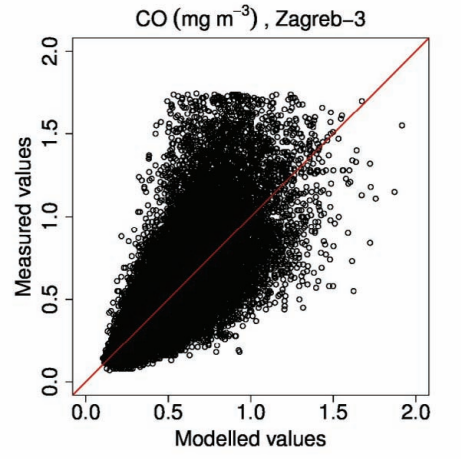

Figure 13. Relationship between modelled and measured CO concentrations at Zagreb-3 for the period 2006–2012. Correlation coefficient (r) equal to 0.66. The red line represents the theoretical linear relationship.

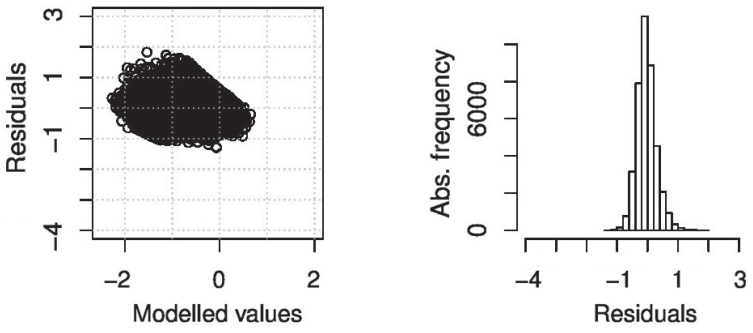

Figure 14. Residual plots for CO at Zagreb-3 for the period 2006–2012. *Left*: relationship between residuals and fitted values; *Right*: histogram of residuals, exhibiting a normal distribution. The majority of residuals group around zero, as expected. For ease of comparison, the y-axis range on the left-hand plot and the *x*-axis range on the right-hand plot are the same.

the left-hand scatter plot, which describes the relationship between residuals and fitted values, exhibits a random pattern, the right-hand plot, the residual histogram, exhibits a normal distribution for CO at Zagreb-3.

A statistical evaluation of the model on the original scale for all pollutants at Zagreb-3 is presented in Tab. 8. Three main criteria should be employed for model evaluation: Measurement standard deviation must be similar to model standard deviation, and both the RMSE and the modified RMSE must be less than measurement standard deviation. In the present study only the first of these three criteria was not fulfilled, with the model generally underestimating values. However, the model was able to describe mean values rather well. The root mean squared error (RMSE) can be calculated via Eq. 9 as follows:

Table 8. Statistical evaluation of the model on the original scale for all pollutants at Zagreb-3 for the entire 2006–2012 study period. The three main criteria employed for model evaluation are as follows: Measurement standard deviation must be similar to model standard deviation, and both the RMSE and modified RMSE must be less than measurement standard deviation.

	CO (mg m^{-3})	SO_2 (μg m^{-3})	NO_2 (μg m^{-3})	PM_{10} (μg m^{-3})
RMSE	0.19	5.16	13.12	16.00
IOA	0.86	0.76	0.80	0.75
Measurement standard deviation	0.30	7.46	19.39	20.82
Model standard deviation	0.23	6.25	14.51	13.36
Measurement mean	0.47	7.34	28.70	30.83
Model mean	0.47	7.63	29.42	30.81
Modified RMSE	0.19	5.16	13.12	16.00

$$RMSE = \sqrt{\frac{1}{n}\sum_i \left(O_i - M_i\right)^2} \ , \tag{9}$$

where O and M correspond to observed and modelled values, respectively, and n is the total number of measurements.

The index of agreement (IOA) is given by Eq. 10 (e.g. Hrust et al., 2009):

$$IOA = 1 - \frac{\sum_i \left(O_i - M_i\right)^2}{\sum_i \left(\left|O_i - \overline{O}\right| + \left|M_i - \overline{O}\right|\right)^2} \ , \tag{10}$$

where IOA=1 represents full agreement and IOA=0 corresponds to no agreement between modelled and measured values.

The modified RMSE is given by Eq. 11 as follows:

$$RMSE_{modified} = \sqrt{\frac{1}{n}\sum_i \left(O_i - \overline{O} - M_i + \overline{M}\right)^2} \ . \tag{11}$$

5. Summary and conclusions

This paper presents a new method with which to estimate the relationship between each of four pollution variables – CO, SO_2, NO_2 and PM_{10} – and meteorological as well as several time variables for the urban area of Zagreb, Croatia. The model is additive on the log scale, with estimation made using hourly data collected during a seven year period at three different locations. Estimates for the different locations share many common traits and exhibit a number of general trends. Hence, similar models could be used in the future for other locations in Zagreb.

The model provides a reasonably good fit in terms of the explained variance. At all three measuring sites, the best model performance was observed for CO. Explanation for this finding remains unclear and deserves further study. The time variables (hour of the day, day of the week and day number), which are supposed to roughly account for the time-dependence of anthropogenic emissions (i.e. traffic density and fuel burning), appear have the largest and a rather clear impact on air quality. In terms of the selected meteorological variables, the impact of wind direction and speed is the strongest, followed by relative humidity and temperature. Meteorological variables were found to have the greatest impact at the site that is in comparison with other two sites the most distant from urban pollution sources (Zagreb-3), except for SO_2. Overall, stable atmospheric conditions, often accompanied by weak winds, calms and high mean sea level

pressure, increase pollutant concentrations, which can be harmful for human health (e.g. Pope and Dockery, 2006). Conversely, unstable atmospheric conditions, often accompanied by strong winds, low mean sea level pressure and precipitation, tend to reduce pollutant concentrations.

In summary, the relationship between air pollution (especially regarding changes in concentration) and meteorological variables was fairly well estimated by the employed GAM, with the primary aim of the present study thus achieved. However, the model is not suitable for prediction purposes.

Future work should therefore preferably focus on model improvement. Firstly, including traffic density (i.e. as a relevant input variable) as an additional predictor should lead to better data fitting in terms of amplitude. Also, the addition of boundary layer height as a predictor could improve model assessment. As the vertical structure of the atmosphere also affects pollution levels, the incorporation of radiosonde data may also contribute to a better understanding and modelling of the impact of atmospheric conditions on pollution levels. Furthermore, in the present study we neglected possible interactions and/or relationships between several predictors (Hastie and Tibshirani, 1990). However, it should be noted that any interpretation of the results obtained by such a complex, nonlinear model would be rather difficult.

Finally, although we focused only on quantifying the influence of meteorology and time variables on pollution levels, the model may also be useful for forecasting pollutant concentrations if residual data are included. Certainly, in that case, reliable forecasts of all predictors would be required. Nevertheless, we can here conclude that temporal changes in CO, SO_2, NO_2 and PM_{10} concentrations in Zagreb are substantially affected by local meteorological conditions.

Acknowledgments – Useful comments of the two reviewers are greatly appreciated. We are grateful to the Meteorological and Hydrological Service of the Republic of Croatia for providing the meteorological and air pollution data. AB is grateful to Prof. Branko Grisogono of the Department of Geophysics for useful discussions. This work has been supported in part by Croatian Science Foundation under projects #09/151 (CATURBO) and #2831 (CARE).

References

Aldrin, M. and Haff, I. H. (2005): Generalized additive modelling of air pollution, traffic volume and meteorology, *Atmos. Environ.*, **39**, 2145–2155, DOI: 10.1016/j.atmosenv.2004.12.020.

Bešlić, I., Šega, K., Čačković, M., Klaić, Z. B. and Vučetić, V. (2007): Influence of weather types on concentrations of metallic components in airborne PM_{10} in Zagreb, Croatia, *Geofizika*, **24**, 93–107.

Bešlić, I., Šega, K., Čačković, M., Klaić, Z. B. and Bajić, A. (2008): Relationship between 4-day air mass back trajectories and metallic components in PM_{10} and $PM_{2.5}$ particle fractions in Zagreb air, Croatia, *Bull. Environ. Contam. Toxicol*, **80**, 270–273, DOI: 10.1007/s00128-008-9360-6.

Bralić, M., Buljac, M., Periš, N., Buzuk, M., Dabić, P. and Brinić, S. (2012): Monthly and seasonal variations of NO_2, SO_2 and black-smoke located within the sport district in urban area, city of Split, Croatia, *Croat. Chem. Acta*, **85**, 139–145, DOI: 10.5562/cca1966.

Carslaw, D. C., Beevers, S. D. and Tate, J. E. (2007): Modelling and assessing trends in traffic-related emissions using a generalized additive modelling approach, *Atmos. Environ.*, **41**, 5289–5299, DOI: 10.1016/j.atmosenv.2007.02.032.

Chaloulakou, A., Kassomenos, P., Spyrelli, N., Demokritou, P. and Koutrakis, P. (2003): Measurements of PM_{10} and $PM_{2.5}$ particle concentrations in Athens, Greece, *Atmos. Environ.*, **37**, 649–660, DOI: 10.1016/S1352-2310(02)00898-1.

Chen, C.-C., Wu, C.-F., Yu, H.-L., Chan, C.-C. and Cheng, T.-S. (2012): Spatiotemporal modeling with temporal-invariant variogram subgroups to estimate fine particulate matter $PM_{2.5}$ concentrations, *Atmos. Environ.*, **54**, 1–8, DOI:10.1016/j.atmosenv.2012.02.015.

Cheng, C. S. Q. and Campbell, M. (2007): A synoptic climatological approach to assess impact on air quality in South-central Canada. Part I: Historical analysis, *Water Air Soil poll.*, **182**, 131–148, DOI: 10.1007/s11270-006-9327-3.

Gonzalez, M., Myers, O., Smith, L., Olvera, H. A., Mukerjee, S., Li, W.-W., Pingitore, N., Amaya, M., Burchiel, S., Berwick, M. and ARCH Study Team (2012): Evaluation of land use regression models for NO_2 in El Paso, Texas, USA, *Sci. Total. Environ.*, **432**, 135–142, DOI: 10.1016/j.scitotenv.2012.05.062.

Grgurić, S., Križan, J., Gašparac, G., Antonić, O., Špirić, Z., Mamouri, R. E., Christodolou, A., Nisantzi, A., Agapiou, A., Themistocleous, K., Fedra, K., Panayiotou, C. and Hadjimitsis, D. (2014): Relationship between MODIS based aerosol optical depth and PM_{10} over Croatia, *Centr. Eur. J. Geosci.*, **6**, 2–16, DOI: 10.2478/s13533-012-0135-6.

Gvozdić, V., Brana, J., Malatesti, N., Puntarić, D., Vidosavljević, D. and Roland, D. (2011): An analysis of the pollution problem in Slavonski Brod (eastern Croatia), *Coll. Antropol.*, **35**, 1135–1141.

Dominici, F., McDermott, A., Zeger S. L. and Samet, J. M. (2002): On the use of generalized additive models in time series studies of air pollution and health, *Am. J. Epidemiol.*, **156**, 193–203, DOI: 10.1093/aje/kwf062.

Finlayson-Pitts, B. J. and Pitts, J. N. (1986): *Atmospheric chemistry: Fundamentals and experimental techniques*. Willey & Sons, New York, 1098 pp.

Hastie, T. and Tibshirani, R. (1990): *Generalized additive models*. Monographs on statistics and applied probability, Vol. 43, 1st ed., Chapman and Hall, London, 335 pp. 335 pp.

Hastie, T., Tibshirani, R. and Friedman, J. (2009): *The elements of statistical learning. Data mining, inference, and prediction*. 2nd ed., Springer, 763 pp.

Hrust, L., Klaić, Z. B., Križan, J., Antonić, O. and Hercog, P. (2009): Neural network forecasting of air pollutants hourly concentrations using optimized temporal averages of meteorological variables and pollutant concentrations, *Atmos. Environ.*, **43**, 5588–5596, DOI: 10.1016/j.atmosenv.2009.07.048.

Hussein, T., Karppinen, A., Kukkonen, J., Härkönen, J., Aalto, P. P., Hämeri, K., Kerminen, V.-M. and Kulmala, M. (2006): Meteorological dependence of size-fractionated number concentrations of urban aerosol particles, *Atmos. Environ.*, **40**, 1427–1440, DOI: 10.1016/j.atmosenv.2005.10.061.

Ito, K., Thurston, G. D. and Silverman, R. A. (2007): Characterization of $PM_{2.5}$, gaseous pollutants and meteorological interactions in the context of time-series health effects models, *J. Expo. Sci. Env. Epid.*, **17**, S45–S60, 10.1038/sj.jes.7500627.

Jeričević, A., Kraljević, L., Vidič, S. and Tarrason, L. (2007): Project description: High resolution environmental modelling and evaluation programme for Croatia (EMEP4HR), *Geofizika*, **24**, 137–143.

Jeričević, A., Kraljević, L., Grisogono, B., Fagerli, H. and Večenaj, Ž. (2010): Parameterization of vertical diffusion and the atmospheric boundary layer height determination in the EMEP model, *Atmos. Chem. Phys.*, **10**, 34 –364, DOI: 10.5194/acp-10-341-2010.

Klaić, Z. (1988): Utjecaj dalekih izvora na zagađenje zraka i oborine procijenjen metodom analize sektora, *Zašt. atm.*, **16**, 114–117.

Klaić, Z. (1990): A Lagrangian one-layer model of long-range transport of SO_2, *Atmos. Envrion.*, **24A**, 1861–1867, DOI: 10.1016/0960-1686(90)90518-R.

Klaić, Z. (1996): A Lagrangian model of long-range transport of sulphur with the diurnal variations of some model parameters, *J. Appl. Meteorol.*, **35**, 574–586, DOI: 10.1175/1520-0450(1996)035<0574:ALMOLR>2.0.CO;2.

Klaić, Z. B. and Beširević, S. (1998): Modelled sulphur depositions over Croatia, *Meteorol. Atmos. Phys.*, **65**, 133–138, DOI: 10.1007/BF01030273.

Klaić, Z. B., Nitis, T., Kos, I. and Moussiopoulos, N. (2002): Modificatiom of the local winds due to hypothetical urbanization of the Zagreb surroundings, *Meteorol. Atmos. Phys.*, **79**, 1–12, DOI: 10.1007/s703-002-8225-z.

Klaić, Z. B., Belušić, D., Bulić, I. H. and Hrust, L. (2003): Mesoscale modeling of meteoro-logycal conditions in the lower troposphere during a winter stratospheric ozone intrusion over Zagreb, Croatia, *J. Geophys. Res.*, **108**, D23, 4720, DOI: 10.1029/2003JD003878.

Klaić, Z. B., Ružman, K., Smiljanić, I., Stojnić, M. and Erbe, P. (2012): The influence of meteorological conditions and weak to moderate traffic density on PM_1 levels in a residential area of Zagreb, *Geofizika*, **29**, 69–102.

Klaić, Z. B., Ollier, S. J., Babić, K. and Bešlić, I. (2015): Influence of outdoor meteorological conditions on indoor wintertime short-term PM_1 levels. *J. Appl. Meteorol. Climatol.*, under review.

Kos, I., Belušić, D., Jeričević, A., Horvath, K., Koračin, D. and Prtenjak, M. T. (2004): Education and research: Initial development of the Atmospheric Lagrangian Particle Stochastic (ALPS) Dispersion Model, *Geofizika*, **21**, 37–52.

Levy, J. I., Bennett, D. H., Melly, S. J. and Spengler, J. D. (2003): Influence of traffic patterns on particulate matter and polycyclic aromatic hydrocarbon concentrations in Roxbury, Massachussetts, *J. Expo. Sci. Environ. Epidemiol.*, **13**, 364–371, DOI: 10.1038/sj.jea.7500289.

Liang, K. and Zeger, S. L. (1986): Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

Mihajlović, D., Ružman K., Kraljević, L. and Jeričević, A. (2012): *Analiza mikrolokacija i mjerne opreme mjernih postaja za mjerenje kvalitete zraka.* Državni hidrometeorološki zavod, 225 pp.

Munir, S., Habeebullah, T. M., Seroji, A. R., Morsy, E. A., Mohammed, A. M. F., Abu Saud, W., Abdou, A. E. A. and Awad, A. H. (2013): Modeling particulate matter concentrations in Makkah, applying a statistical modelling approach, *Aerosol Air Qual. Res.*, **13**, 901–910, DOI: 10.4209/aaqr.2012.11.0314.

Otto, S. A., Kornilovs, G., Llope, M. and Möllmann, C. (2014): Interactions amongdensity, climate, and food web effects determine long-term life cycle dynamics of a key copepod, *Mar. Ecol. Prog. Ser.*, **498**, 73–84, DOI: 10.3354/meps10613.

Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J. and Tapper, N. J. (2011): Quantifying the influence of local meteorology on air quality using generalized additive models, *Atmos. Environ.*, **45**, 1328–1336, DOI: 10.1016/j.atmosenv.2010.11.051.

Pope, C. A. III and Dockery, D. W. (2006): Health effects of fine particulate air pollution: Lines that connect, *J. Air Waste Manage. Assoc.*, **56**, 709–742, DOI: 10.1080/10473289.2006.10464485.

Prtenjak, M. T., Jeričević, A., Kraljević, L., Bulić, I. H., Nitis, T. and Klaić, Z. B. (2009): Exploring atmospheric boundary layer characteristics in a severe SO_2 episode in the north-eastern Adriatic, *Atmos. Chem. Phys.*, **9**, 4467–4483, DOI: 10.5194/acp-9-4467-2009.

Prtenjak, M. T., Srnec, L., Peternel, R., Madžarević, V., Hrga, I. and Stjepanović, B. (2012): Atmospheric conditions during high ragweed pollen concentrations in Zagreb, Croatia, *Int. J. Biometeorol.*, **56**, 1145–1158, DOI: 10.1007/s00484-012-0520-3.

Prtenjak, M. T., Jeričević, A., Klaić, Z. B., Alebić-Juretić, A. and Bulić, I. H. (2013): Atmospheric dynamics and elevated ozone concentrations in the northern Adriatic, *Meteorol. Appl.*, **20**, 482–496, DOI: 10.1002/met.1312.

Quisthoudt, K., Adams, J., Rajkaran, A., Dahdouh-Guebas, F., Koedam, N. and Randin, C. F. (2013): Disentangling the effects of global climate and regional land-use change on the current and future distribution of mangroves in South Africa, *Biodivers. Conserv.*, **22**, 1369–1390, DOI: 10.1007/s10531-013-0478-4.

R. Development Core Team (2009): *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna, Austria. The R manual, available at http://www.r-project.org

Ryaboshapko, A., Bazhanov, V. and Oyola, P. (1996): Ozone, sulphur dioxide, condensation nuclei and aerosol composition over the eastern part of the North Atlantic basin, *Atmos. Environ*., **30**, 1305–1315, DOI: 10.1016/1352-2310(95)00345-2.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M. and Doyle, M. (2003): A rigorous intercomparison of ground-level ozone prediction, *Atmos. Environ.*, **37**, 3237–3253, DOI: 10.1016/S1352-2310(03)00330-3.

Simpson, D., Fagerly, H., Jonson, J. E., Tsyro, S., Wind, P., and Tuovinen, J.-P. (2003): *Transboundary acidification, eutrophication and ground level ozone in Europe. Part I: Unified EMEP model description.* EMEP Status Report 1/2003, Norwegian Meteorological Institute, Oslo, Norway, 74 pp.

Wanka, E. R., Bayerstadler, A., Heumann, C., Nowak, D., Jörres, R. A., and Fischer, R. (2014): Weather and air pollutants have an impact on patients with respiratory diseases and breathing difficulties in Munich, Germany, *Int. J. Biometeorol.*, **58**, 249–262, DOI: 10.1007/s00484-013-0730-3.

Wood, S. N. (2001): Mgcv: GAMs and generalized ridge regression fo R, *R News,* **1/2**, 20–25, available at http://www.utstat.utoronto.ca/reid/sta450/Rgam.pdf

Wood, S. N. (2006): *Generalized additive models – An introduction with R.* Chapman and Hall, London. 410 pp.

Zhang, K., and Batterman, S. (2010): Near-road air pollutant concentrations of CO and $PM_{2.5}$: A comparison of MOBILE6.2/CALINE4 and generalized additive models, *Atmos. Environ*., **44**, 1740–1748, DOI: 10.1016/j.atmosenv.2010.02.008.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev A. A. and Smith, G. M. (2009): *Mixed effects models and extensions in ecology with R.* Statistics for biology and health. Springer, New York, 574 pp.

SAŽETAK

Određivanje utjecaja lokalnih meteoroloških uvjeta na kvalitetu zraka u Zagrebu primjenom generaliziranih aditivnih modela

Andreina Belušić, Ivana Herceg-Bulić i Zvjezdana Bencetić Klaić

U ovom se radu određuje utjecaj lokalnih meteoroloških uvjeta u Zagrebu na satnu koncentraciju odabranih polutanata: ugljikovog monoksida (CO), sumporovog dioksida (SO_2), dušikovog dioksida (NO_2) i lebdećih čestica aerodinamičkog polumjera do 10 μm (PM_{10}) za razdoblje od 2006. do 2012. godine. Primijenjena je nova metodologija za urbano područje Zagreba. U modelu je logaritam satne koncentracije polutanata određen pomoću sume nelinearnih funkcija meteoroloških i nekoliko vremenskih varijabli. Uključene vremenske varijable opisuju vremenske promijene u emisijama. Takav model pripada generaliziranim aditivnim modelima (GAM) i aditivan je na logaritamskoj skali, što rezultira umnošcima na originalnoj skali. Iako je model nelinearan, rezultati se vrlo lako interpretiraju, opisujući utjecaj meteoroloških uvjeta i emisija na kvalitetu zraka pomoću relativne važnosti i parcijalnih utjecaja. Provedena je i statistička procjena uspješnosti modela. U konačnici, dobiveni rezultati su pokazali da su najvažnije varijable one koje opisuju emisije. Ukupni utjecaj meteoroloških varijabli u modelu objasnio je 45% varijance za CO, 14% za SO_2, 25% za NO_2 i 24% za PM_{10}. Time je pokazano da meteorološki

uvjeti, barem lokalno, osjetno utječu na kvalitetu zraka u Zagrebu. Stabilni atmosferski uvjeti u urbanom okruženju pogoduju većim koncentracijama navedenih polutanata. Pri nestabilnim atmosferskim uvjetima dominira konvekcija, koja razrjeđuje koncentracije polutanata unutar graničnog sloja.

Ključne riječi: atmosferski granični sloj, kvaliteta zraka urbanog područja, statističko modeliranje

Corresponding author's address: Andreina Belušić, University of Zagreb, Faculty of Science, Department of Geophysics, Andrija Mohorovičić Geophysical Institute, Horvatovac 95, HR-10000 Zagreb, Croatia, tel: +385 1 46 05 900, e-mail: andreina.belusic@gfz.hr