

## Forecasting of Daily PM<sub>10</sub> Concentrations in Brno and Graz by Different Regression Approaches

Ernst Stadlober<sup>1</sup>, Zuzana Hübnerová<sup>2</sup>, Jaroslav Michálek<sup>2</sup>  
and Miroslav Kolář<sup>3</sup>

<sup>1</sup> Graz University of Technology, Austria

<sup>2</sup> Brno University of Technology, Czech Republic

<sup>3</sup> Masaryk University, Czech Republic

**Abstract:** Brno and Graz, the second largest cities of their countries, observe in each winter season PM<sub>10</sub> concentrations of daily means which regularly exceed the limit value of  $50 \mu\text{g}/\text{m}^3$ . This is mainly caused by unfavorable dissemination conditions of the ambient air. Hence, partial regulation measures have to be taken in Brno and Graz where specific decisions for certain regulations may be based on the average PM<sub>10</sub> concentration of the next day provided that reliable forecasts of these values are available. For several sites in the two cities we establish forecasts of daily PM<sub>10</sub> concentrations based on multiple linear regression and generalized linear models utilizing both measured covariates of the present day and meteorological forecasts of the next day. The comparisons, based on different quality measures demonstrate the usefulness of both model approaches as they yield results of similar quality. Our prediction models may support future decisions concerning possible traffic restrictions or other regulations.

**Zusammenfassung:** Brünn und Graz, die jeweils zweitgrößten Städte ihrer Länder, weisen in jeder Wintersaison PM<sub>10</sub>-Konzentrationen von Tagesmittelwerten auf, die regelmäßig den Grenzwert von  $50 \mu\text{g}/\text{m}^3$  überschreiten. Dies ist vor allem begründet durch schlechte Durchlüftungsbedingungen der Umgebungsluft. Daher sind partielle Steuerungsmechanismen in Brünn und Graz notwendig, wo spezifische Maßnahmen getroffen werden sollen, die den Tagesmittelwert der PM<sub>10</sub>-Konzentration des nächsten Tages verwenden könnten, falls zuverlässige Prognosen für diese Werte verfügbar sind. Für verschiedene Messstationen der beiden Städte werden Prognosewerte von täglichen PM<sub>10</sub>-Konzentrationen hergeleitet, die auf multipler Regression und verallgemeinerten linearen Modellen basieren, welche am laufenden Tag gemessene Kovariable sowie meteorologische Vorhersagen des nächsten Tages verwenden. Die Vergleiche, basierend auf unterschiedlichen Qualitätsmaßen, demonstrieren die Brauchbarkeit beider Modellansätze, welche Resultate ähnlicher Qualität liefern. Unsere Vorhersagemodelle könnten zukünftige Entscheidungen bezüglich möglicher Verkehrsbeschränkungen oder anderer Maßnahmen unterstützen.

**Keywords:** Prediction, Forecast, PM<sub>10</sub>, Multiple Linear Regression, Generalized Linear Model, Meteorology.

## 1 Introduction

Particularly during the winter season, the basin area of Graz is exposed to weather conditions such as stationary temperature inversions, low wind velocities and rare precipitation events. These special weather conditions cause an extensive load of particulate matter (PM) in ambient air. PM/fine dust has become a local and regional issue of agglomeration centers, large cities with heavy traffic and industrial areas. The PM<sub>10</sub> (particles with an aerodynamic diameter  $< 10\mu m$ ) concentration is measured in units of  $\mu g/m^3$ . According to regulations established by the European Union (EU) in the EC Directive (2008), the limit value for the daily PM<sub>10</sub> average at the stations which show the highest concentration of pollution is  $50 \mu g/m^3$  and must not be exceeded on more than 35 days of the calendar year (valid since January 1, 2005). Also since 2005, the annual PM<sub>10</sub> average must not exceed the limit of  $40 \mu g/m^3$ . The stricter rules in Austria allowed only 30 exceedances (2005 to 2009) and reduced this even to 25 exceedances (2010). However, at the test point Graz-Mitte (near the pedestrian zone in the center of the city) we observed merely in the eight cold periods (October to March) 2002/2003 – 2009/2010 between 36 (2008/2009) and 105 (2005/2006) exceedances of the daily PM<sub>10</sub> limit and registered high annual averages between  $41 \mu g/m^3$  and  $49 \mu g/m^3$ .

Rules in the Czech Republic are in accordance with the EU regulations. The daily PM<sub>10</sub> average at the less exposed station Arboretum in Brno exceeded the limit value in 21 days in the cold period 2007/10 – 2008/03 and in 2 days in the cold period 2008/10 – 2009/03. In contrast, the stations Židenice and Zvonařka (both close to traffic spots) exhibited a much higher number of exceedances from 55 days to 73 days in the cold period 2007/11 – 2009/03. The annual PM<sub>10</sub> average at the Brno stations varied from  $27 \mu g/m^3$  to  $45 \mu g/m^3$  in 2006, from  $17 \mu g/m^3$  to  $36 \mu g/m^3$  in 2007, from  $19 \mu g/m^3$  to  $44 \mu g/m^3$  in 2008 and from  $18 \mu g/m^3$  to  $41 \mu g/m^3$  in 2009.

It is well known that PM may cause adverse health effects which have been analyzed in many epidemiological and toxicological studies. An extensive general review can be found in Pope III and Dockery (2006). A specific Austrian study AUPHEP is described in Hauck et al. (2004), while Schwarze et al. (2006) is a thorough review with 210 references of studies from different regions throughout the world.

Due to adverse health effects caused by PM<sub>10</sub> and in order to fulfill the EU regulations, policy has to react (and take drastic measures) against the PM problem. For the authorities it may be necessary to base singular decisions on reliable forecasting models for daily PM<sub>10</sub> concentrations. Our aim is to deliver reliable PM<sub>10</sub> forecasting models for specific sites and to demonstrate their practical applicability. Our models have been developed for the cold season (October to March), being the period with particularly elevated PM<sub>10</sub> concentrations at the observed sites. There are considerably lower PM<sub>10</sub> concentrations during the warm season and hence, there is no urgent need for action in that season of the year. Models for Graz based on multiple linear regression were already presented in a previous study (Stadlober, Hörmann, and Pfeiler, 2008). In contrast to many other studies it was possible to show the performance of these models in operational mode based on a three-year trial period. In Brno, generalized linear models have been used to analyze and predict PM<sub>10</sub> levels by means of meteorological and seasonal variables (Hrdličková, Michálek, Kolář, and Veselý, 2008, Veselý, Tonner, Hrdličková, Michálek, and Kolář, 2009). Other

typical approaches for PM prediction are neuronal networks (c.f. Pérez and Reyes, 2002, Hooyberghs, Mensink, Dumont, Fierens, and Brasseur, 2005, Papanastasiou, Melas, and Kioutsioukis, 2007), discriminant analysis (c.f. Silva, Pérez, and Trier, 2001), multi-gene genetic programming (Pires, Alvim-Ferraz, Pereira, and Martins, 2010) or Kalman filtering (c.f. van der Wal and Jansen, 1999). Hybrid ARIMA and artificial neural network models have been applied to improve forecasting of extreme events (Díaz-Robles et al., 2008). Sfetsos and Vlachogiannis (2010) complements the three different approaches linear regression, nearest neighbors and artificial networks by adding hourly PM10 values of the past 48 hours to the models.

In the next section, we describe the situation in Graz and Brno according to sites, data bases and covariates (input parameters). The multiple regression methodology (LM) and the generalized linear models (GLM) approach of our study are discussed in Section 3. Results of model comparisons and of test runs in operational mode are analyzed in Section 4. In the final Section 5 we conclude our findings and reveal our recommendations.

## 2 Situation in Graz and in Brno

### 2.1 Sites and database in Graz

Our analysis in Graz is based on data from two monitoring stations, the site Graz-Mitte in a traffic area near to the pedestrian zone in the city center and the site Graz-Süd near to a commercial and industrial area in the south. In Graz-Mitte we collected data of eight cold seasons (2002/2003 – 2009/2010, 1348 observations and 55 missing values), in Graz-Süd we had access to data of seven cold seasons (2003/2004 – 2009/2010, 1094 observations and 48 missing values). Graz is located in a basin area south of the main Alpine crest leading to low precipitation and low wind velocities during the cold period. Additionally, stationary temperature inversions frequently occur at that time of year due to the basin location. Some core data characteristics, listed in Table 1, illustrate the situation at Graz-Mitte. Similar results were obtained for Graz-Süd.

Graz, the capital of Styria, has approximately 263 k residents and is located in the northern part of the Graz basin at 350 m above sea level. The tree-covered hills bordering at the north are about 400 m higher. Our two measuring stations are part of a permanent PM10 monitoring network in Graz with five test points three of which are located in traffic areas and two in residential areas. Additionally to PM10, the test points provide other pollution and meteorological data on the basis of half hour averages. From the test point Kalkleiten which is situated close to Graz 710 m above sea level we obtained temperature data to measure the effect of temperature inversion. Figure 1 shows the location of the stations on the map of Graz.

Table 2 shows the number of crucial days at cold seasons, i.e. days where PM10 exceeds the EU-limit of  $50 \mu\text{g}/\text{m}^3$ . For the site Graz-Mitte we observed between 36 (20 %) (2008/2009) and 105 (58 %) (2005/2006) exceedances per cold season. The corresponding values at Graz-Süd ranged from 49 (27 %) (2008/2009) to 103 (57 %) (2005/2006).

Subsequently, we introduce the input variables suitable for our multiple regression and generalized linear models. The focus is on forecasting daily averages of PM10 for day  $t$  in the cold season, i.e. the period from October 1 to March 31:

Table 1: Percentage of PM10 levels observed within different meteorological scenarios for the test point Graz-Mitte, where O-D covers the period October 1 to December 31 and J-M the period January 1 to March 31 (8 winter seasons 2002/2003 – 2009/2010, 1348 observations, 55 missings). Medians of covariates are calculated separately for the periods O-D and J-M. Here ‘low’  $\hat{=}$  PM10  $\leq$  30, ‘moderate’  $\hat{=}$  PM10  $\in$  (30, 50], ‘high’  $\hat{=}$  PM10  $\in$  (50, 75], ‘very high’  $\hat{=}$  PM10  $\in$  (75, 100] and ‘alert’  $\hat{=}$  PM10  $>$  100.

Precipitation	$\Delta_{temp} > 0$ (no inversion)						$\Delta_{temp} \leq 0$ (inversion)					
	wind speed > median			wind speed $\leq$ median			wind speed > median			wind speed $\leq$ median		
	yes	J-M	O-D	no	J-M	O-D	yes	J-M	O-D	no	J-M	O-D
Month	O-D	J-M	O-D	J-M	O-D	J-M	O-D	J-M	O-D	J-M	O-D	J-M
Low	57.8	53.5	32.6	30.7	39.0	20.4	16.5	9.5	16.7	33.3	2.7	17.6
Moderate	32.2	34.9	46.9	36.4	42.9	41.8	39.7	30.7	58.3	58.3	40.5	35.5
High	6.7	11.6	17.1	23.6	15.2	26.5	33.1	38.0	25.0	33.3	45.9	22.6
Very high	3.3		3.4	8.4	2.9	7.1	9.1	12.4		33.3	10.8	35.5
Alert				0.9	4.1	4.1	1.7	9.5				6.5
No. observed	90	86	175	225	105	98	121	137	12	3	37	31
AVG PM10	32.0	32.7	39.2	43.9	37.4	48.1	48.4	60.3	43.2	57.5	53.9	67.5
									57.6	106.5	57.6	106.5
									66.8	90.0	66.8	90.0

Table 2: Number of observations of PM10 above 50 and total number of observations per month and season in brackets.

	02/03	03/04	04/05	05/06	06/07	07/08	08/09	09/10
Graz-Mitte	100(178)	78(171)	86(182)	105(182)	60(182)	56(183)	36(182)	50(151)
Graz-Süd	–	91(183)	89(182)	103(180)	66(175)	79(183)	49(182)	76(151)

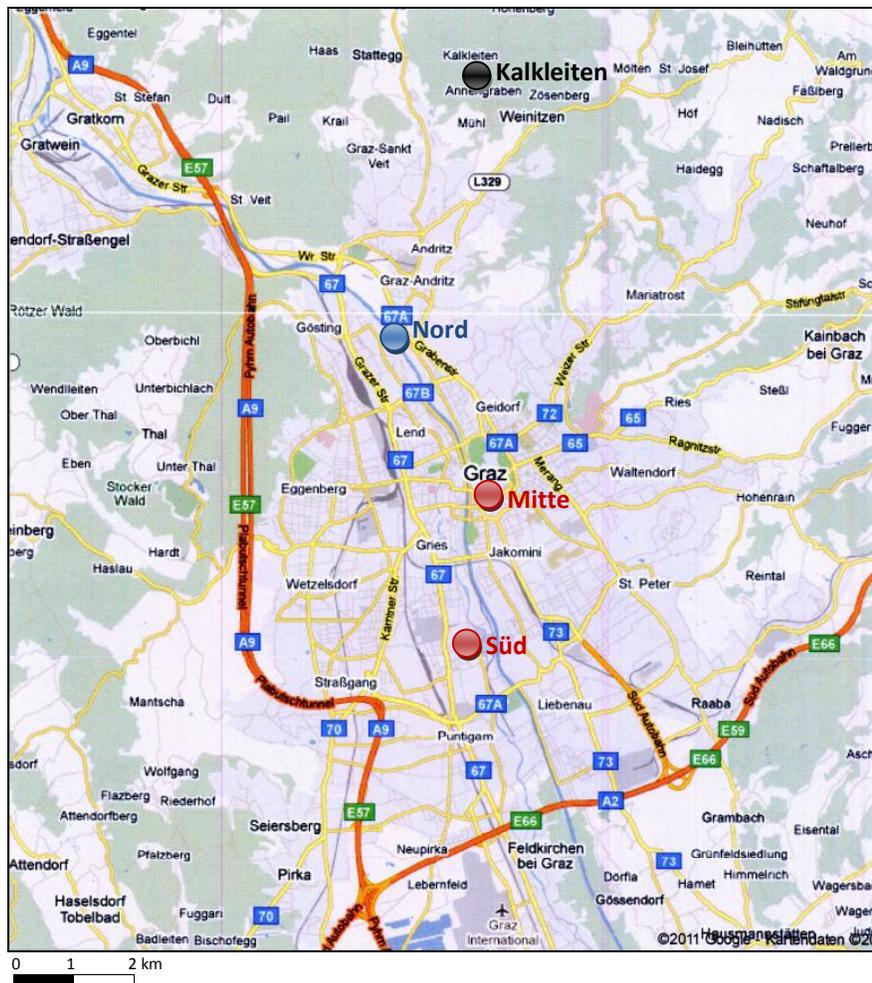


Figure 1: Map of stations in Graz.

- $A_{p_t}$ : average of 48 1/2h-measurements of PM10 on day  $t$  (0:30 to 24:00).

Temperature inversion has the most significant impact on  $A_{p_t}$ , the PM10 concentration in Graz (see also Hörmann, Pfeiler, and Stadlober, 2005). In order to measure temperature inversion for Graz we define

- $\Delta T_t$ : average temperature difference to reference test point (Kalkleiten) on day  $t$  (at Graz-Mitte for Graz-Süd).

If the daily average  $\Delta T_t$  is negative, we refer to this as an inversion day. In the cold period, the daily average of  $\Delta T_t$  is negative to a level of 13-47 % depending on specific temperature conditions in the winter season. Obviously, the emergence of wind and precipitation leads to reduced PM10 concentrations. Thus, we include the two parameters

- $V_t$ : average wind speed on day  $t$  (at Graz-Mitte for Graz-Süd),
- $Prec_t$ : 1 = precipitation, 0 = no precipitation on day  $t$  (at Graz-Nord).

Values of the three meteorological variables for day  $t$  have to be delivered at day  $t - 1$ , the day of the prediction. Hence, in operational mode they are only available as meteorological forecasts. Below, in Subsection 4.2 we show the reliability of our PM10 predictions based on these meteorological forecasts for test data from the cold season 2009/2010.

Of course, further crucial variables for our models are the so-called lag values of PM10 and the categorized temperature:

- $A_{p_{lag}}$ : average of 1/2h-measurements of PM10 from 12.30 day  $t - 2$  to 12.00 day  $t - 1$ ,
- $T_{lag}$ : categorized average of 1/2h temperature measurements at 12.30 on day  $t - 2$  to 12.00 on day  $t - 1$  (= 0 if temp > 0, = 1 if temp ≤ 0) (at Graz-Mitte for Graz-Süd).

Domestic fuel is one of the human impact factors. Naturally, during frosty periods more heating is necessary, i.e. the effect of domestic heating may be estimated partially by measures of air temperature. Because of the high correlation between temperatures on proximate days, it may be feasible to include the lag value  $T_{lag}$ . In operational mode, the predictions for day  $t$  has to be published in the afternoon (in Graz at 14.00) of day  $t - 1$ , so the observed values  $A_{lag}$  and  $T_{lag}$  are available on time.

An additional human impact is reflected to some extent by weekday/weekend differences of PM10. It is very likely that in Graz this effect arises from the reduced traffic load (see Hörmann et al., 2005, Stadlober et al., 2008). Additionally, we observed that the characteristics of PM10 changes during the winter season (higher levels than expected in February and March than expected by meteorological conditions). To model these effects we include four dummy variables indicating Saturday, Sun-/Holiday, February and March.

- saturday: 1 = Saturday, 0 = no Saturday on day  $t$ ,
- sunday: 1 = Sunday/Holiday, 0 = no Sunday/Holiday on day  $t$ ,
- febr: 1 = February, 0 = no February on day  $t$ ,
- march: 1 = March, 0 = no March on day  $t$ .

Note that all models are fitted on a training data set containing seven cold seasons 2002/2003 – 2008/2009 at Graz-Mitte (six cold seasons 2003/2004 – 2008/2009 at Graz-Süd) utilizing observed values of the covariates as in Stadlober et al. (2008). These models were employed for predictions of the test data set (days ranging from November 3, 2009 to March 31, 2010).

## 2.2 Sites and database in Brno

Brno, the second largest city of the Czech Republic with about 430 k residents is an economical, cultural and political center. The city is located in a basin at 190 m to 425 m above sea level. The basin is open only in the south and it is surrounded by arboreous hills in the other three cardinal directions. There are two rivers going through Brno and confluenting in the south of the center. The local air quality is significantly affected by the concentration of industry and motor traffic, as it lies on a road junction of highways D1 and D2.

It can be claimed that the air pollution by dust aerosol is connected especially with the winter season, when disproportionately higher emissions of the local heating plants appear. Moreover, the temperature inversions as well as adverse dispersion conditions are more frequent. On the other hand, snowfalls absorb the dust aerosol better than rainfalls. In the summer season the days with extreme values of PM<sub>10</sub> appear during the periods of higher temperatures and lower air humidity.

In the metropolitan area of Brno we investigated the three monitoring stations Arboretum, Židenice and Zvonařka (see the map of stations in Brno, Figure 2). The site Arboretum is located in the botanical garden of the Mendel University of Agriculture and Forestry (part of the inner city area). Here we had access to data from three cold seasons (2006/11 – 2007/3, 2007/10 – 2008/03, 2008/10 – 2009/03, 483 observations and 26 missing values). The site Židenice is situated near a heavy-traffic street surrounded by warehouses, manufacturing plants and a railway line. Zvonařka station is a traffic spot within an industrial area. For the last two stations data from two cold seasons were available (2007/11 – 2008/03, 2008/10 – 2009/03, 313 observations and 4 missing values).

Meteorological data together with their predictions (based on 6h-measurements) are provided by the Brno Regional Office of the Czech Hydrometeorological Institute. Note that these data are valid for the whole Brno agglomeration and therefore are not station specific.

The future development of the pollution of dust aerosol in Brno will be apparently fixed with constant year fluctuation and with maxima at the end of winter. A gradual increase of monthly PM<sub>10</sub> averages caused by more frequent extreme traffic situations in combination with adverse dispersion conditions is possible.

Similarly to Section 2.1, Table 3 and 4 provide some core data characteristics at Arboretum and Židenice respectively. Results for the station Zvonařka are analogous to those for the site Židenice.

The number of crucial days at cold seasons, i.e. days where PM<sub>10</sub> exceeds the EU-limit of  $50 \mu\text{g}/\text{m}^3$ , at the Brno stations is provided in Table 5. For the site Arboretum we observed between 2 (1 %) (2008/2009) and 29 (23 %) (2006/2007) exceedances per cold season. The corresponding values at Židenice and Zvonařka were remarkably higher.

The following covariates turned out to be important in forecasting daily averages  $Ap_t$  of PM<sub>10</sub> for day  $t$  in the period from October 1 to March 31:

- $Ap_t$ : average of 24 1h-measurements of PM<sub>10</sub> on day  $t$  (1:00 to 24:00).

The progression of the pollution at stations is significantly influenced by the wind direction. The wind speed affects the pollution differently – higher emissions come up

Table 3: The analogue of Table 1 for the test point Arboretum.

cloud cover	wind speed > median			wind speed ≤ median			wind speed > median			wind speed ≤ median						
	O-D	J-M	J-M													
Low	50.0	46.0	71.4	6.3	41.5	47.8	36.4	40.0	43.8	30.8	50.0	50.0	40.3	32.9	38.5	21.4
Moderate	32.1	49.2	14.3	93.8	48.8	50.0	54.5	60.0	47.9	56.4	50.0	33.3	46.8	48.7	53.8	50.0
High	7.1	4.8			9.8	2.2	9.1		8.3	12.8		16.7	12.9	17.1	7.7	28.6
Very high	10.7		14.3												1.3	
No. observed	28	63	7	16	41	46	11	10	48	39	2	6	62	76	13	14
AVG PM10	36.3	31.9	33.4	39.1	33.5	31.8	37.6	32.2	32.5	37.5	30.0	35.1	35.0	37.8	34.3	40.6

Table 4: The analogue of Table 1 for the test point Židenice.

cloud cover	wind speed > median			wind speed ≤ median			wind speed > median			wind speed ≤ median						
	O-D	J-M	J-M													
Low	21.7	51.4		25.0	13.6	17.5	16.7	11.1	28.1	16.0	16.7		18.8	16.3		
Moderate	52.2	29.7	50.0	25.0	40.9	42.5	16.7	22.2	40.6	32.0	33.3	25.0	18.8	26.5	33.3	
High	21.7	16.2	33.3	50.0	31.8	32.5	33.3	55.6	25.0	48.0	33.3	25.0	53.1	32.7	40.0	33.3
Very high	4.3	2.7	16.7		13.6	7.5	16.7	11.1	6.3	4.0	16.7	50.0	6.3	20.4	40.0	8.3
Alert							16.7						3.1	4.1	20.0	25.0
No. observed	23	37	6	4	22	40	6	9	32	25	6	4	32	49	5	12
AVG PM10	41.9	35.0	50.8	50.8	48.5	46.0	62.6	59.6	42.4	48.7	53.1	65.0	53.7	56.6	80.1	75.6

Table 5: Number of observations of PM10 above 50 and total number of observations per month and season in brackets.

	06/07	07/08	08/09
Arboretum	29(127)	21(183)	2(173)
Židenice	–	72(134)	73(179)
Zvonařka	–	66(134)	55(179)

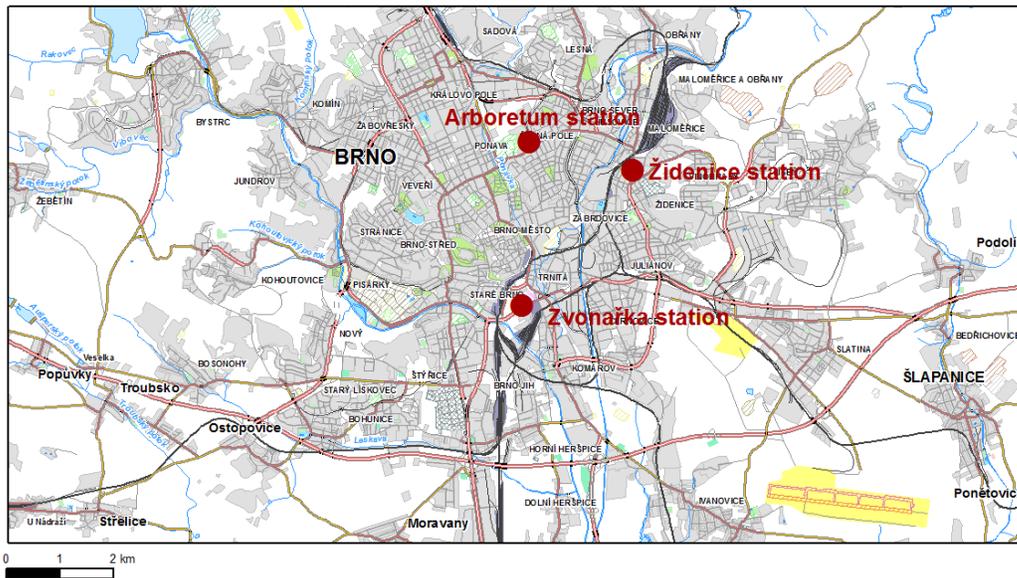


Figure 2: Map of stations in Brno.

during stronger circulation (pollutant is brought from distant sources) as well as weaker circulation (the air is ventilated worse and polluting aerosol is accumulated). Unlike in Hrdličková et al. (2008) and Veselý et al. (2009), here the predicted wind direction  $D_t$  and wind speed  $V_t$  were considered in the model. Note that wind direction is measured as an oriented angle between the vector pointing north of the station and the observed wind direction vector pointing to the station. Nevertheless, the same projections

- $V_t \sin D_t(\text{pr})$ : average of 4 predictions of projections of the wind vector at 0.00, 6.00, 12.00, 18.00 on day  $t$ ,
- $V_t \cos D_t(\text{pr})$ : average of 4 predictions of projections of the wind vector at 0.00, 6.00, 12.00, 18.00 on day  $t$

were included as in Hrdličková et al. (2008). Note that this reparametrization provides an estimate of the angle  $S$  [°] between the vector pointing north of the station and the vector pointing from the direction of maximum pollutant concentration to the station  $S$ . After estimating the parameters of the model, the corresponding component of the linear predictor can be reparametrized by the sum formula to  $\beta_S V_t \cos(D_t - S)$ . Note that because of this reparametrization after the parameter estimation, the variables  $V_t \sin D_t(\text{pr})$ ,

$V_t \cos D_t(\text{pr})$  were either both included or both excluded when searching for sufficient models in Sections 3.1 and 3.2.

Furthermore, the following lag variables of PM10 and temperature were included:

- $A_{p\text{lag}}$ : average of 1h-measurements of PM10 from 13.00 day  $t-2$  to 12.00 day  $t-1$ ,
- $T_{\text{lag}}$ : average of 4 measurements of temperature at 18.00 on day  $t-2$  and at 0.00, 6.00, 12.00 on day  $t-1$ .

In contrast to the stations in Graz, here the average temperature on day  $t$  was also considered:

- $T_t(\text{pr})$ : average of 4 predictions of temperature at 0.00, 6.00, 12.00, 18.00 on day  $t$ .

Note that data on precipitation were not available for Brno stations. Therefore, we used the predicted variable:

- $C_t(\text{pr})$ : average of 4 predictions of cloud cover at 0.00, 6.00, 12.00, 18.00 on day  $t$ .

Cloud cover is a degree of coverage of the sky by clouds and indirectly provides an information on the length of the sun shine. It ranges from 0 for completely clear sky to 10 completely overcast sky.

The effect of domestic heating was represented by the following variable:

- $HS_t$ : activity of heating plants on day  $t$ , days in same month have the same values (Oct: 0.08, Nov: 0.14, Dec: 0.17, Jan: 0.19, Feb: 0.16, Mar: 0.14).

For the same reasons as in Section 2.1, the following dummy variables indicating Weekend, Sun-/Holiday, February and March were included:

- weekend: 1 = Weekend, 0 = Working day  $t$ ,
- sunday: 1 = Sunday/Holiday, 0 = no Sunday/Holiday on day  $t$ ,
- febr: 1 = February, 0 = no February on day  $t$ ,
- march: 1 = March, 0 = no March on day  $t$ .

Note that considering the variable *weekend* instead of *saturday* (in models for Graz data) allows us to test the hypothesis, that there is no difference between the level of PM10 on Saturday and Sun-/Holiday. When testing a model without *weekend* to be sufficient, the variable *sunday* was also excluded.

Models for Arboretum are fitted on data from both two cold seasons (2006/10 – 2008/03, 310 observations and 15 missing values). For Židenice and Zvonařka the data from the season 2007/11 – 2008/03 (132 observations) were available. All models on the training data sets had access to predicted covariates (provided by the Czech Hydrometeorological Institute) which is different to the modeling of Graz data where observed values of covariates were used. These models were applied for predictions of the test data set including days of the period 2008/10 – 2009/03.

## 3 Methodology

### 3.1 Multiple linear regression

The regression models are designed for the dependent variable  $\sqrt{Ap_t}$ . To assure a constant error variance and to avoid a violation of the model assumptions we found the square root transformation to be a suitable choice.

**Model assumption:** Let  $x_t^{(i)}$  ( $i = 1, 2, \dots, m$ ) be a pool of metric input variables and  $d_t^{(j)}$  ( $j = 1, \dots, p$ ) be a pool of dummy (0/1) input variables at day  $t$ . Then we assume that the following linear model for  $\sqrt{Ap_t}$  holds:

$$\sqrt{Ap_t} = \beta_0 + \sum_{i=1}^m \beta_i x_t^{(i)} + \sum_{j=1}^p \beta_{m+j} d_t^{(j)} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} N(0, \sigma^2). \quad (1)$$

For each site the parameter estimates  $\hat{\beta}_k$ ,  $k = 0, \dots, m + p$  are chosen via a stepwise regression procedure from the candidate set of input variables (Graz, Subsection 2.1, Brno, Subsection 2.2). One remarkable feature of this simple linear model approach is that the parameters and results can still be interpreted conveniently. In contrast to other rather complex models with numerous input parameters and/or functional relationships generated by a black box mechanism, linear models of this type are still transparent for the user (see Hörmann et al., 2005, Stadlober et al., 2008).

In Graz-Mitte and Graz-Süd all variables of the specified pool were found to be significant. The corresponding results are given in Table 6 below where the covariates are ordered according to their level of significance expressed by the  $t$ -value at Graz-Mitte. Accordingly, temperature inversion  $\Delta T_t$  and the lag value  $Ap_{lag}$  of PM10 are the most important ones, followed by three covariates of similar importance: average wind speed  $V_t$ , the lag value of temperature  $T_{lag}$  and the *sunday* effect.

However, at the sites in Brno only 4 to 8 variables of the predefined pool are important (see Table 7). The most significant ones for the less exposed test point Arboretum seem to be the lag value  $Ap_{lag}$  of PM10 and the projections of the wind vector. On the other hand, at the test points Židenice and Zvonařka, which are near heavy-traffic streets, the projections of the wind vector, the lag value  $T_{lag}$  of temperature, the cloud cover  $C_t(\text{pr})$  and the *weekend* effect exhibited their significant influence. We observed that the covariate *sunday* had no relevance at all. Hence it does not appear in Table 7.

Note that, unlike to Graz-Mitte and Graz-Süd, a negative *march* effect is observed. In Brno, the heating season ends in March. In this month, the temperature inversions are shorter and less frequent, and therefore the thermic stratification, which used to be more stable during winter, is labilized. The atmosphere is not ventilated under the temperature inversion, which implies a cumulation of pollutants and an increase of their concentration. The conditions are improved after a change of the conditions, when the circulation is intensified, which is demonstrated by a sufficient spread of the pollutants or their wash off by the precipitation among others.

Table 6: Parameter estimates  $\hat{\beta}$  of the linear model for  $\sqrt{Ap_t}$  with their standard errors and  $t$  statistics for Graz stations. The table is completed by the coefficient of determination  $R^2$ , the standard deviation  $se$  of the model and number of observations  $n$ .

Variables $x$ or $d$	Graz-Mitte			Graz-Süd		
	$\hat{\beta}$	$sd(\hat{\beta})$	$t$	$\hat{\beta}$	$sd(\hat{\beta})$	$t$
const	7.412	0.120	61.5	8.317	0.143	58.0
$\Delta T_t$	-0.301	0.018	-16.6	-0.367	0.023	-16.1
$Ap_{lag}$	0.019	0.001	15.1	0.015	0.001	11.9
$V_t$	-1.591	0.132	-12.0	-2.251	0.165	-13.7
$T_{lag}$	0.954	0.084	11.4	1.185	0.106	11.2
sunday	-0.968	0.085	-11.4	-0.617	0.104	-5.9
march	0.752	0.095	7.9	0.690	0.115	6.0
saturday	-0.688	0.094	-7.3	-0.704	0.115	-6.1
$Prec_t$	-0.507	0.071	-7.1	-0.853	0.087	-9.8
febr	0.344	0.090	3.8	0.287	0.111	2.6
	$R^2$	$se$	$n$	$R^2$	$se$	$n$
	0.629	1.104	1199	0.647	1.261	1046

Table 7: Parameter estimates  $\hat{\beta}$  of the sufficient model for  $\sqrt{Ap_t}$  with their standard errors and  $t$  statistics for Brno stations. The table is completed by identified angle of highest air pollution  $S$  and corresponding regression coefficient  $\hat{\beta}_S$ . Then the coefficient of determination  $R^2$ , the standard deviation  $se$  of the model and number of observations  $n$  follow.

Variables $x$ or $d$	Arboretum			Židenice			Zvonařka		
	$\hat{\beta}$	$sd(\hat{\beta})$	$t$	$\hat{\beta}$	$sd(\hat{\beta})$	$t$	$\hat{\beta}$	$sd(\hat{\beta})$	$t$
const	4.305	0.242	17.8	7.602	0.473	16.1	8.784	0.543	16.2
$Ap_{lag}$	0.027	0.003	7.9	0.014	0.005	2.7	0.011	0.005	2.2
$T_{lag}$				-0.098	0.033	-3.0	-0.186	0.058	3.2
$V_t \sin D_t(\text{pr})$	0.153	0.020	7.6	0.268	0.054	4.9	0.356	0.059	6.0
$V_t \cos D_t(\text{pr})$	-0.006	0.021	-0.3	-0.089	0.052	-1.7	-0.083	0.056	-1.5
$T_t(\text{pr})$							0.168	0.062	2.7
$C_t(\text{pr})$				-0.109	0.049	-2.2	-0.255	0.052	-5.0
$HS_t$	5.587	1.462	3.8						
weekend				-0.701	0.238	-2.9	-0.630	0.242	-2.6
febr				0.843	0.281	3.0			
march							-1.041	0.313	-3.3
$\hat{\beta}_S$		0.153			0.282			0.366	
$S$		92°			108°			103°	
	$R^2$	$se$	$n$	$R^2$	$se$	$n$	$R^2$	$se$	$n$
	0.381	0.720	305	0.481	1.226	132	0.537	1.240	132

### 3.2 Generalized linear models with gamma distributions

Another possibility to model the response  $Ap_t$  is to assume that the measurements of  $Ap_t$  are gamma distributed. Then one can apply the generalized linear model approach with log-link function (see e.g. Fahrmeir and Tutz, 1994) as used in our previous study (Hrdličková et al., 2008).

**Model assumption:** Let  $x_t^{(i)}$  ( $i = 1, 2, \dots, m$ ) be a pool of metric input variables and  $d_t^{(j)}$  ( $j = 1, \dots, p$ ) be a pool of dummy (0/1) input variables at day  $t$ . Then we assume that  $Ap_t$  has a Gamma( $\mu_t, 1/\phi$ ) distribution and  $\log(\mu_t) = \log(E(Ap_t|X_t))$  can be described as linear model of the form

$$\log(\mu_t) = \beta_0 + \sum_{i=1}^m \beta_i x_t^{(i)} + \sum_{j=1}^p \beta_{m+j} d_t^{(j)}, \quad (2)$$

where  $\phi$  is the dispersion parameter and  $E(Ap_t|X_t)$  denotes the conditional expectation of  $Ap_t$  for the given set  $X_t$  of variables on the right-hand side of model (2). The dispersion parameter  $\phi$  can be estimated by the generalized Pearson statistic and the parameter estimates  $\beta_k$ ,  $k = 0, \dots, p + m$ , are obtained by the maximum likelihood method both implemented in the procedure *glmfit* which is part of the *Statistics Toolbox* of the MATLAB package. We selected the best model by stepwise backward selection using the Wald statistic as selection criterion, see Fahrmeir and Tutz (1994). Of course, standard residual analyses and goodness-of-fit tests were used to check the adequacy of these models.

In Graz-Mitte and Graz-Süd all variables of the specified pool were found to be significant. The corresponding results are given in Table 8 below where the covariates are ordered according their level of significance at Graz-Mitte. Note that we obtain essentially the same order as in the multiple linear regression models of Table 6 in Subsection 3.1.

For the sites in Brno a specific combination of covariates appears. The model for Arboretum contains only 6 covariates, but the two traffic spots Židenice and Zvonařka need 8 identical covariates for modeling PM10 (see Table 9). The two projections  $V_t \sin D_t(\text{pr})$  and  $V_t \cos D_t(\text{pr})$  of the wind vector and the dummy variable *march* are used in all three models, whereas the lag variable  $Ap_{\text{lag}}$  is chosen for modeling PM10 at Arboretum, and the lag variable  $T_{\text{lag}}$  is required in models of PM10 at Židenice and Zvonařka. Here, the predicted temperature  $T_t(\text{pr})$  and cloud cover  $C_t(\text{pr})$  together with the *weekend* and *february* effect also exhibited their influence. On the other hand, we observed once more that the covariate *sunday* had no relevance at all. Hence, it does not appear in Table 9.

Table 8: Parameter estimates  $\hat{\beta}$  of the sufficient GLM with log-link for  $E(Ap_t|X_t)$  with their standard errors for Graz stations. Then goodness-of-fit statistics  $D$  and  $\chi_{ans}^2$  with their degrees of freedom  $f$  and number of observations  $n$  follow.

Variables $x$ or $d$	Graz-Mitte		Graz-Süd	
	$\hat{\beta}$	$sd(\hat{\beta})$	$\hat{\beta}$	$sd(\hat{\beta})$
const	3.123	0.081	3.534	0.087
$\Delta T_t$	-0.078	0.005	-0.091	0.007
$\log(Ap_{lag})$	0.310	0.020	0.261	0.020
$V_t$	-0.498	0.038	-0.745	0.047
$T_{lag}$	0.272	0.024	0.319	0.030
sunday	-0.286	0.024	-0.174	0.030
march	0.224	0.027	0.230	0.033
saturday	-0.210	0.027	-0.215	0.033
$Prec_t$	-0.149	0.021	-0.245	0.025
febr	0.094	0.026	0.095	0.032
$D$	115.815		129.393	
$\chi_{ans}^2$	115.079		128.281	
$f$	1189		1036	
$n$	1199		1046	

Table 9: Parameter estimates  $\hat{\beta}$  of the sufficient GLM model with log-link for  $E(Ap_t|X_t)$  and their standard errors for Brno stations. The table is completed by identified angle of highest air pollution  $S$  for each station and corresponding regression coefficient  $\hat{\beta}_S$ . Then the goodness-of-fit statistics  $\chi_{ans}^2$  and Wald statistics  $W$  together with their degrees of freedom  $f$ , number of observations  $n$  and corresponding  $\chi_{0.95}^2$  quantile follow.

Variables $x$ or $d$	Arboretum		Židenice		Zvonařka	
	$\hat{\beta}$	$sd(\hat{\beta})$	$\hat{\beta}$	$sd(\hat{\beta})$	$\hat{\beta}$	$sd(\hat{\beta})$
const	2.131	0.166	4.444	0.126	4.528	0.136
$\log(Ap_{lag})$	0.360	0.043				
$T_{lag}$			-0.061	0.014	-0.067	0.015
$V_t \sin D_t(\text{pr})$	0.049	0.006	0.092	0.016	0.124	0.017
$V_t \cos D_t(\text{pr})$	-0.004	0.007	-0.019	0.015	-0.023	0.016
$T_t(\text{pr})$			0.044	0.016	0.056	0.017
$C_t(\text{pr})$			-0.047	0.015	-0.070	0.016
$HS_t$	1.617	0.474				
weekend			-0.203	0.064	-0.173	0.069
febr			0.194	0.083	0.183	0.090
march	-0.080	0.034	-0.235	0.089	-0.337	0.096
$S$	94°		102°		101°	
$\chi_{ans}^2$	15.152		15.452		17.981	
$W$	5.917		3.593		3.302	
$\chi_{0.95}^2$	12.592		7.815		7.815	
$f$	299		125		125	
$n$	305		134		134	

## 4 Comparison of Different Forecasting Models

### 4.1 Quality of forecasting

The multiple linear regression models and the generalized linear models introduced in Section 3 are used as a basis for forecasting daily means of PM10 in Graz (November 3, 2009 to March 31, 2010) and in Brno (October 1, 2008 to March 31, 2009). As a first step, we used a historical training data set to compute the estimates  $\hat{\beta}_k, k = 0, \dots, m + p$ , for the corresponding models listed in Tables 6, 7, 8, and 9 which yield predictions  $\sqrt{\widehat{Ap}_t}$  and  $\log(\widehat{Ap}_t)$ , respectively. To obtain the estimates  $\widehat{Ap}_t$  one has to take the square of the predicted  $\sqrt{\widehat{Ap}_t}$  or the exp of the predicted  $\log(\widehat{Ap}_t)$ . The resulting biases are negligible for our purposes (and thus will not be considered). See Cordeiro and McCullagh (1991) for discussion on the bias correction in GLM.

Clearly, in operational mode the covariates constituting values of the prediction day  $t$  are available as meteorological forecasts only. In Graz, this concerns the measure for temperature inversion  $\Delta T_t$ , the average wind speed  $V_t$  and the precipitation as (0/1) categorized variable  $Prec_t$ . The models for the stations in Brno need more sophisticated meteorological forecasts: average of 4 predictions of (i) the wind speed  $V_t(\text{pr})$ , of (ii) the wind direction  $D_t(\text{pr})$ , of (iii) the temperature  $T_t(\text{pr})$  and of (iv) the cloud cover  $C_t(\text{pr})$ .

The quality of the PM10 forecasts is measured and compared by a quality function introduced in Stadlober et al. (2008), which is tailored to the particular needs given by the EU-limit value of  $50 \mu\text{g}/\text{m}^3$ .

The quality function  $Q(x, y)$  assigns a value between 0 and 1 to each pair ( $x$ =observation,  $y$ =forecast). Values close to 1 signify very good forecasts, whereas small values near to zero indicate bad forecasts. Based on the value of  $Q(x, y)$  we define a scoring system:  $Q(x, y) \geq 0.8 \hat{=} 1$ : *excellent*,  $Q(x, y) \in [0.6, 0.8) \hat{=} 2$ : *good*,  $Q(x, y) \in [0.4, 0.6) \hat{=} 3$ : *satisfactory*,  $Q(x, y) \in [0.2, 0.4) \hat{=} 4$ : *bad*,  $Q(x, y) \in [0, 0.2) \hat{=} 5$ : *very bad*. The function  $Q$  is given by the relation

$$Q(x, y) = 1 - \min \left\{ \frac{|x - y|}{10d}, 1 \right\} \quad \text{with} \quad (3)$$

$$d = 1 + \frac{1}{2} \sqrt{|x - 50| + |y - 50| + 100(I_A + 10I_B)},$$

where  $A = \{x \leq 50, y \leq 50\}$ ,  $B = \{x \geq 100, y \geq 100\}$ , and  $I_E$  is the indicator function:  $I_E = 1$  if condition  $E$  is valid, and  $I_E = 0$  otherwise.

The function  $Q(x, y)$  has jumps near the threshold points ( $x = 50, y = 50$ ), ( $x = 100, y = 100$ ); it penalizes combinations  $Q(x > 50, y < 50)$  ( $Q(x > 100, y < 100)$ ), in cases if the *forecast*  $y$  is below  $50 \mu\text{g}/\text{m}^3$  (below  $100 \mu\text{g}/\text{m}^3$ ), but the *observation*  $x$  is above  $50 \mu\text{g}/\text{m}^3$  (above  $100 \mu\text{g}/\text{m}^3$ ) or vice versa. Therefore, a situation in which the observation is above a threshold and forecast is below the threshold is assessed equally to a situation in which the observation is below a threshold and forecast is above it, though the second situation might not be as crucial in reality as the first one. However, the proposed  $Q$  function has been tested for six winter seasons (2004/2005 – 2009/10) in Graz and has proven to be a suitable measure for the intended application (Stadlober et al., 2008).

## 4.2 Practical performance and comparison

In this Subsection, we compare the results of the two modelling approaches with respect to the quality function  $Q(x, y)$ , in (3). We start with the linear models and generalized linear models of Graz-Mitte and Graz-Süd fitted on training data of seven cold seasons (2002/10 to 2009/03, Graz-Mitte, Table 6) and six cold seasons (2003/10 to 2009/03, Graz-Süd, Table 8). They deliver daily forecasts  $\widehat{Ap}_t$  for test data from November 3, 2009 to March 31, 2010.

Tables 10 and 11 show a good coincidence of the quality of GLM and LM forecasts for both sites of Graz. A high correlation of the forecasts is exhibited in Figures 3 and 4 (top right panel). In Graz-Mitte it turns out that there are 70.5 % (69.1 %) very good or good forecasts. In Graz-Süd this quality is reached by both models in 66.4 % of the cases. At least satisfactory forecasts for both models can be expected with a probability of 85.2 % (Graz-Mitte) or 83.9 % (Graz-Süd).

Splitting the observations in two classes PM10(low):  $PM10 \leq 50$  and PM10(high):  $PM10 > 50$ , we get for PM10(low) at least 89 % and for PM10(high) at least 76 % satisfactory forecasts at Graz-Mitte and Graz-Süd. In Brno, the LM and GLM models are fitted to training data (October 1, 2008 to March 31, 2009, Tables 7 and 9) and they yield daily forecasts  $\widehat{Ap}_t$  for test data from October 1, 2008 to March 31, 2009. Table 12 for station Arboretum presents successful predictions of both GLM and LM models, probably caused by the low number of PM10 observations above  $50 \mu\text{g}/\text{m}^3$  as can be seen in Figure 5.

Table 13 (station Židenice) shows that both models reached a similar percentage of acceptable forecasts (GLM: 73.4 %, LM: 72.9 %), though the individual qualities of the predictions are quite different as can be observed in Figure 6, bottom left panel. The top right panel of Figure 6 reveals the fact that the GLM model tends to predict higher values than the LM model.

On the one hand, Table 14 for station Zvonařka reveals a high percentage of very good forecasts (GLM: 40.7 %, LM: 40.1 %), on the other hand a high fraction of bad and very bad forecasts (GLM: 28.8 %, LM: 27.1 %). GLM and LM forecasts agree for PM10(low) (70.8 % (GLM) and 74.0 % (LM) are at least satisfactory forecasts). However, they rather disagree for PM10(high) (76.7% (GLM) and 71.2% (LM) are at least satisfactory) with higher values obtained by the GLM model (see Figure 7, top right panel).

Table 10: Contingency table for Graz-Mitte: Quality of prediction GLM (rows) versus LM (columns), number and percentage of categories of quality 1 to 5.

GLM/LM	1	2	3	4	5	sum GLM	%	cum %
1	63	5	0	0	0	68	45.6	45.6
2	4	27	5	0	1	37	24.8	70.5
3	1	3	16	2	0	22	14.8	85.2
4	0	0	3	10	1	14	9.4	94.6
5	0	0	0	1	7	8	5.4	100.0
sum LM	68	35	24	13	9	149		
%	45.6	23.5	16.1	8.7	6.0	100.0		
cum %	45.6	69.1	85.2	94.0	100.0			

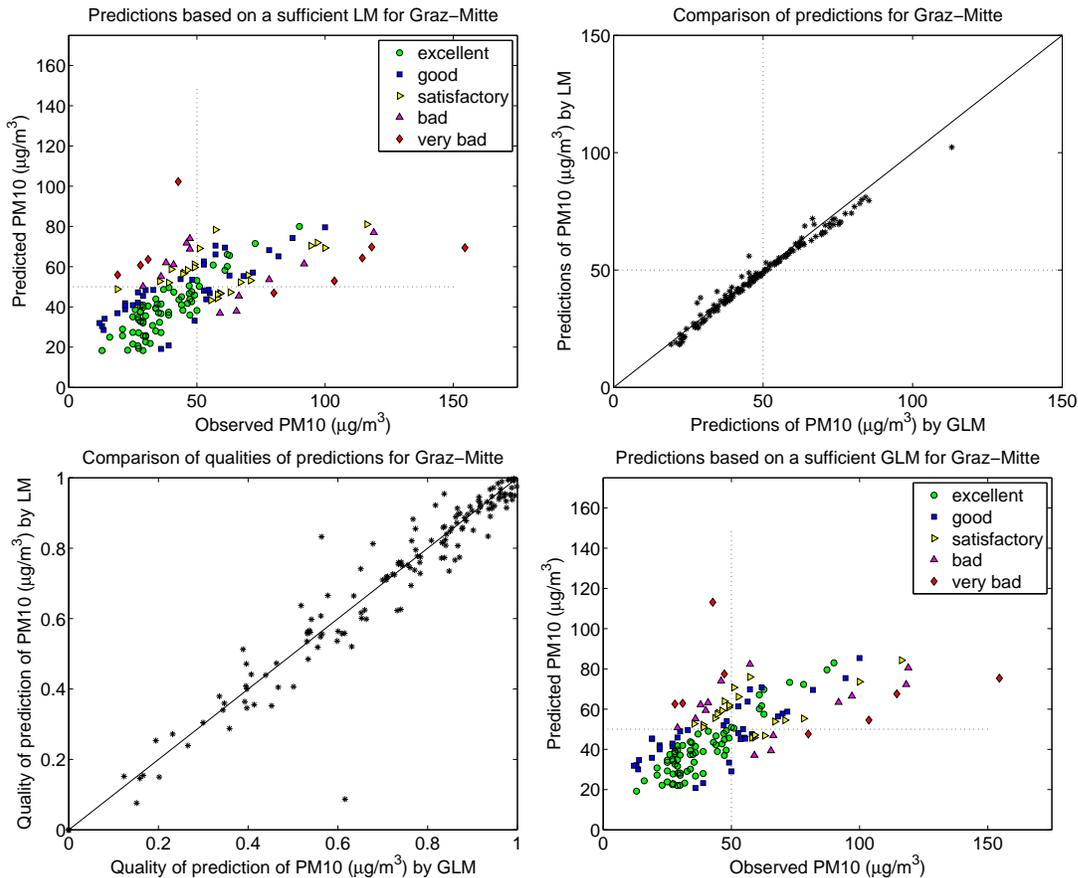


Figure 3: Graz-Mitte: scatterplot of forecasts of LM versus observations scored by  $Q$  (top left), forecasts  $\widehat{Ap}_t$  of GLM versus LM (top right),  $Q$  of GLM versus LM (bottom left), scatterplot of forecasts of GLM versus observations scored by  $Q$  (bottom right).

Table 11: Contingency table for Graz-Süd: Quality of prediction GLM (rows) versus LM (columns), number and percentage of categories of quality 1 to 5.

GLM/LM	1	2	3	4	5	sum GLM	%	cum %
1	43	9	0	0	0	52	34.9	34.9
2	8	32	7	0	1	47	31.5	66.4
3	0	6	18	2	0	26	17.5	83.9
4	0	1	1	8	3	13	8.7	92.6
5	0	0	0	1	10	11	7.4	100.0
sum LM	51	48	26	11	13	149		
%	34.2	32.2	17.5	7.4	8.7	100.0		
cum %	34.2	66.4	83.9	91.3	100.0			

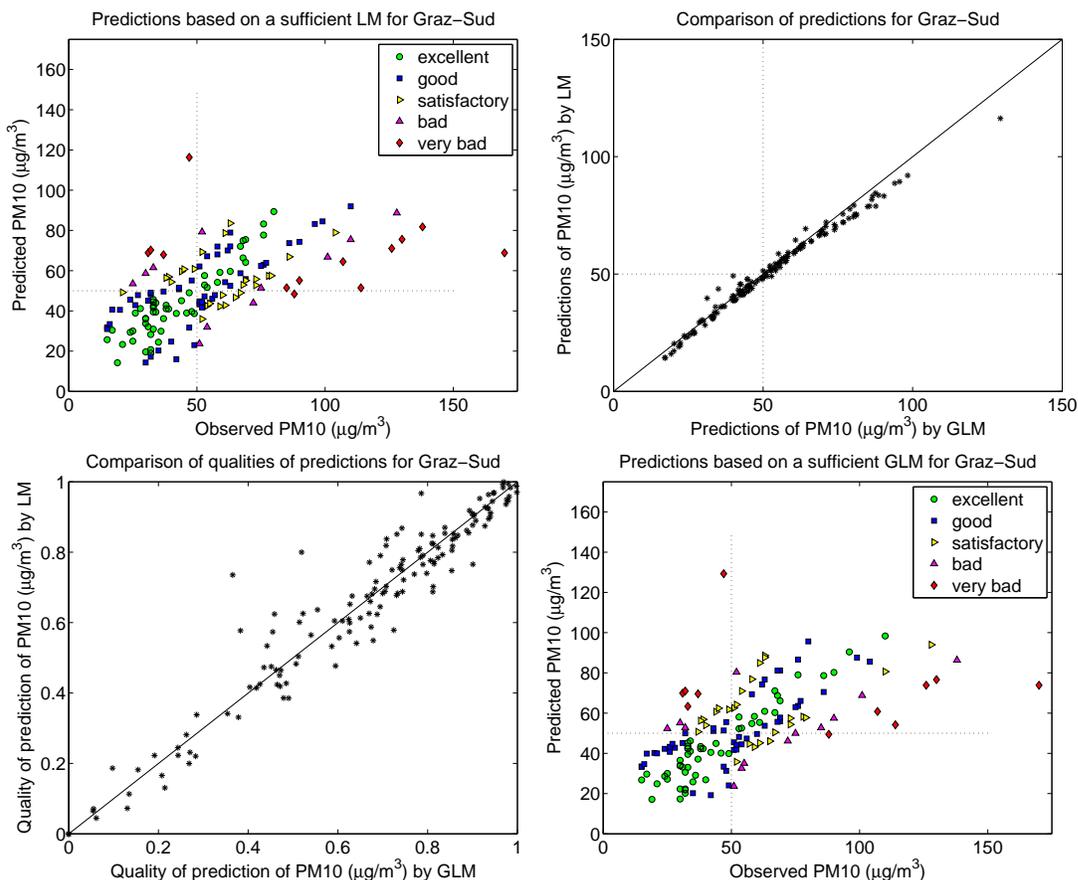


Figure 4: Graz-Süd: scatterplot of forecasts of LM versus observations scored by  $Q$  (top left), forecasts  $\widehat{Ap}_t$  of GLM versus LM (top right),  $Q$  of GLM versus LM (bottom left), scatterplot of forecasts of GLM versus observations scored by  $Q$  (bottom right).

Table 12: Contingency table for Arboretum: Quality of prediction GLM (rows) versus LM (columns), number and percentage of categories of quality 1 to 5.

GLM/LM	1	2	3	4	5	sum GLM	%	cum %
1	138	0	0	0	0	138	81.2	81.2
2	2	24	0	0	0	26	15.3	96.5
3	0	1	3	0	0	4	2.4	98.8
4	0	2	0	0	0	2	1.2	100.0
5	0	0	0	0	0	0	0.0	100.0
sum LM	140	27	3	0	0	170		
%	82.4	15.9	1.8	0.0	0.0	100.0		
cum %	82.4	98.2	100.0	100.0	100.0			

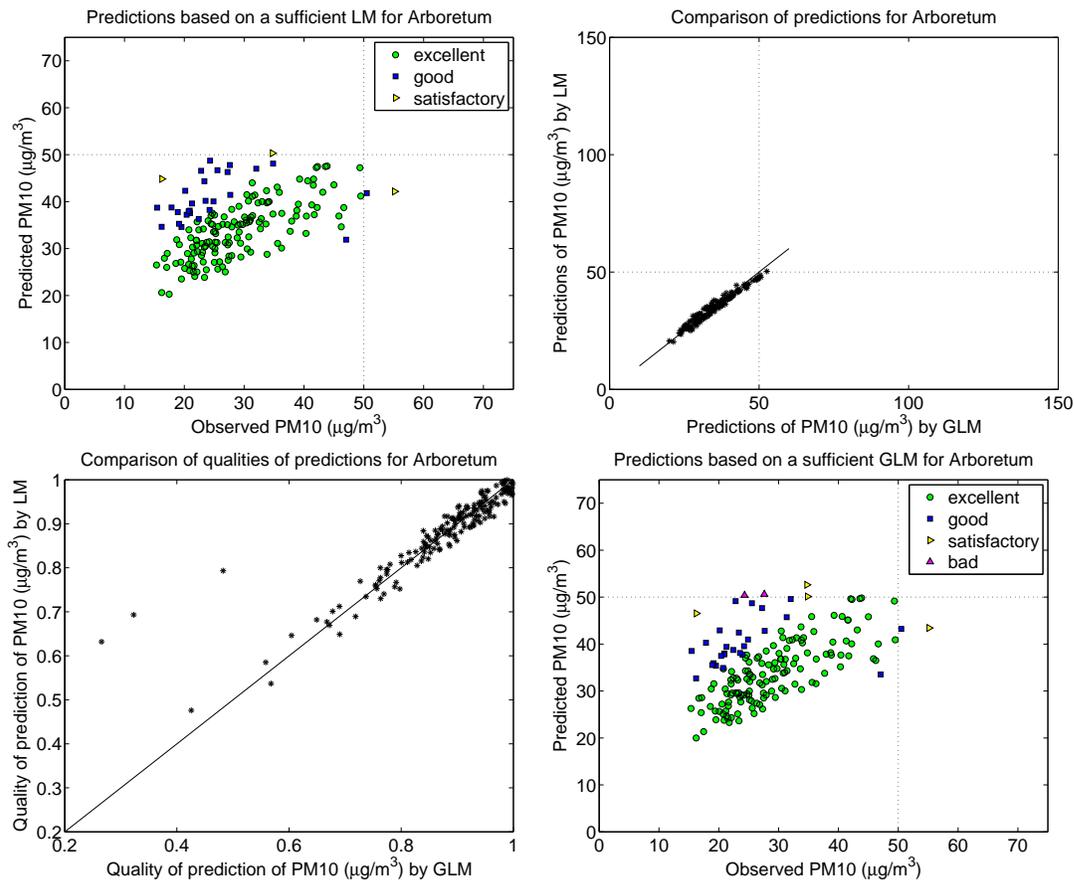


Figure 5: Arboretum: scatterplot of forecasts of LM versus observations scored by  $Q$  (top left), forecasts  $\widehat{Ap}_t$  of GLM versus LM (top right),  $Q$  of GLM versus LM (bottom left), scatterplot of forecasts of GLM versus observations scored by  $Q$  (bottom right).

Table 13: Contingency table for Židenice: Quality of prediction GLM (rows) versus LM (columns), number and percentage of categories of quality 1 to 5.

GLM/LM	1	2	3	4	5	sum GLM	%	cum %
1	54	7	4	1	0	66	37.3	37.3
2	10	18	7	3	0	38	21.5	58.8
3	6	3	7	8	2	26	14.7	73.4
4	2	0	6	10	3	21	11.9	85.3
5	0	0	5	8	13	26	14.7	100.0
sum LM	72	28	29	30	18	177		
%	40.7	15.8	16.4	16.9	10.2	100.0		
cum %	40.7	56.5	72.9	89.8	100.0			

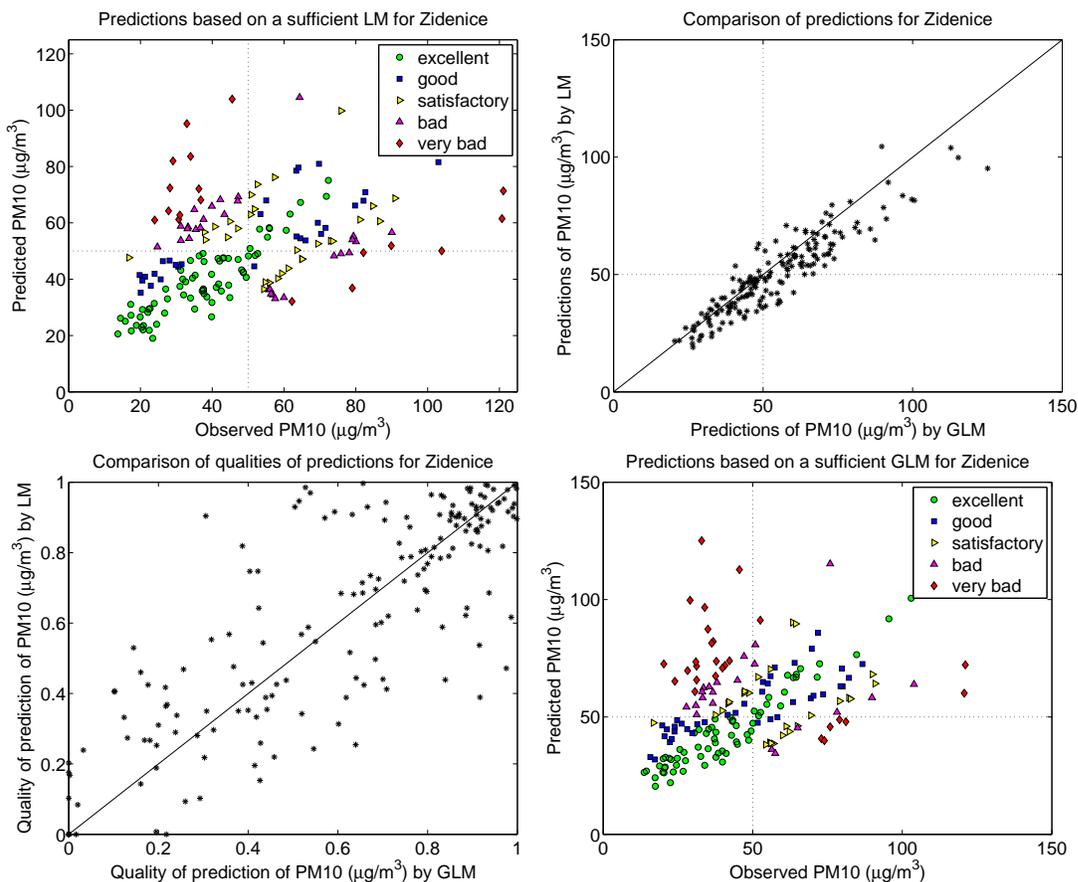


Figure 6: Židenice: scatterplot of forecasts of LM versus observations scored by  $Q$  (top left), forecasts  $\widehat{Ap}_t$  of GLM versus LM (top right),  $Q$  of GLM versus LM (bottom left), scatterplot of forecasts of GLM versus observations scored by  $Q$  (bottom right).

Table 14: Contingency table for Zvonařka: Quality of prediction GLM (rows) versus LM (columns), number and percentage of categories of quality 1 to 5.

GLM/LM	1	2	3	4	5	sum GLM	%	cum %
1	61	11	0	0	0	72	40.7	40.7
2	7	20	6	1	0	34	19.2	59.9
3	2	3	12	1	2	20	11.3	71.2
4	1	1	2	14	2	20	11.3	82.5
5	0	2	1	9	19	31	17.5	100.0
sum LM	71	37	21	25	23	177		
%	40.1	20.9	11.9	14.1	13.0	100.0		
cum %	40.1	61.0	72.9	87.0	100.0			

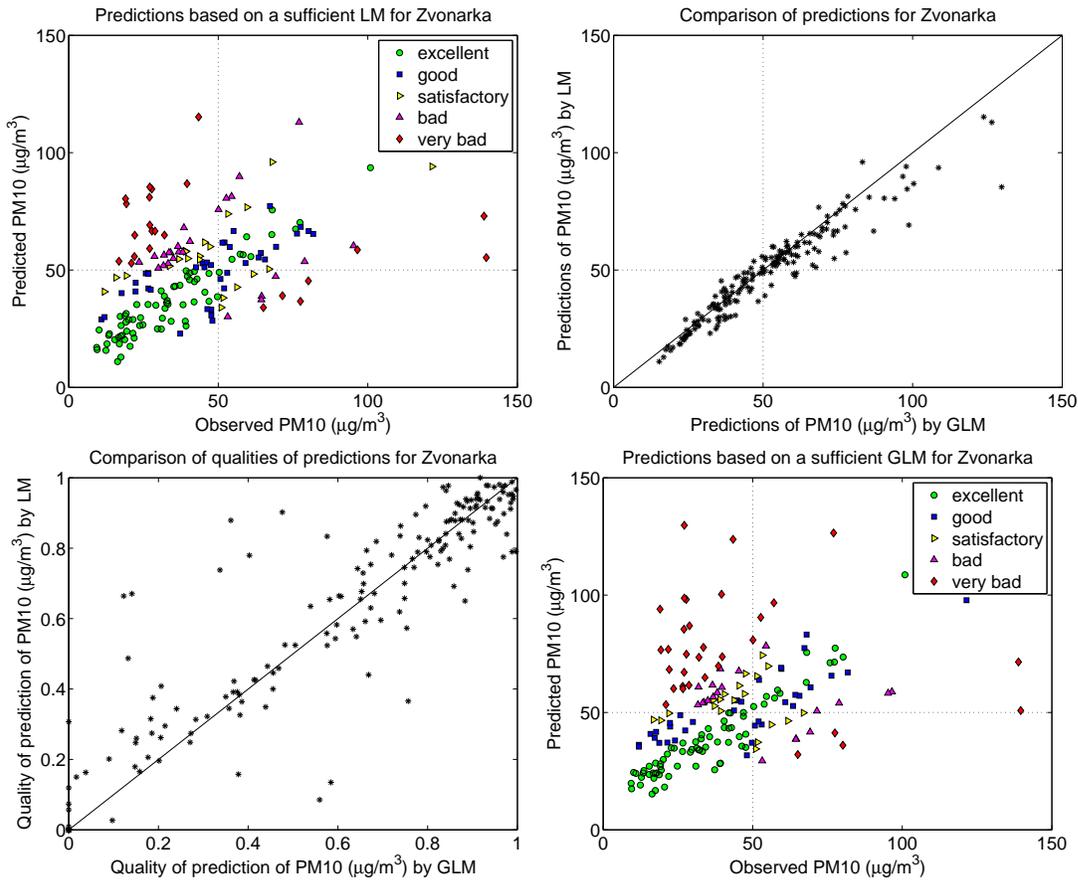


Figure 7: Zvonařka: scatterplot of forecasts of LM versus observations scored by  $Q$  (top left), forecasts  $\widehat{A}p_t$  of GLM versus LM (top right),  $Q$  of GLM versus LM (bottom left), scatterplot of forecasts of GLM versus observations scored by  $Q$  (bottom right).

## 5 Conclusion and Recommendations

In the present paper, we have shown that PM10 forecasting models for Brno and Graz based on (i) multiple linear regression with square root transformation and (ii) generalized linear modelling with gamma distribution and log-link are suitable choices for the intended goal. They yield similar and comparable results with respect to our quality measure. Due to the simple and explicit description of the model parameters we are confident that intransparent black box approaches are not preferable in this case. The covariates are selected in order to represent the meteorological and anthropogenic situation of the locations. In general, we attach importance on the practical performance and the convenient handling of the models. For an every-day-application in the field it is necessary that both the measured covariates and the covariates obtained as meteorological forecast values are available at a certain time of the previous day  $t - 1$ . Both implementation and servicing for users has to be convenient.

With respect to our quality function 71 % (traffic spots Židenice and Zvonařka in Brno), 85 % (Graz-Mitte: near to traffic spot, Graz-Süd: near to industrial area) and 98 % (Arboretum: botanical garden) of the forecasts are assigned to categories *excellent*, *good* or *satisfactory*. This performance is very promising, but the models are open for further adaptations. In general, a clever choice of additional covariates may improve the reliability of the models. On the one hand, our models may serve as regular monitoring tools, on the other they may also be established as feasible and objective base for decision making in regards to traffic regulation measures.

## Acknowledgements

The article was written in the context of the long-term research project from the Ministry of Education, Youth and Sports of the Czech Republic number MSM0021622418. The authors would like to thank Johanna Stadlober for linguistical revision of the present paper.

## References

- Cordeiro, G. M., and McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B*, 53, 629-643.
- Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Ree, G. D., Chow, J. C., Watson, J. G., et al. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment*, 42, 8331-8340.
- EC Directive. (2008). Council Directive 2008/50/EC on ambient air quality and cleaner air for Europe. *Official Journal of the European Communities, L 151*, 1-44.
- Fahrmeir, L., and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer-Verlag.
- Hauck, H., Berner, A., Frischer, T., Gomiscek, B., Kundi, M., Neuberger, M., et al. (2004). AUPHEP – Austrian project on health effects of particulates – general overview. *Atmospheric Environment*, 38, 3905-3915.

- Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., and Brasseur, O. (2005). A neuronal network forecast for daily average PM10 concentrations in Belgium. *Atmospheric Environment*, 39, 3279-3289.
- Hörmann, S., Pfeiler, B., and Stadlober, E. (2005). Analysis and prediction of particulate matter PM10 for the winter season in Graz. *Austrian Journal of Statistics*, 34, 307-326.
- Hrdličková, Z., Michálek, J., Kolář, M., and Veselý, V. (2008). Identification of factors affecting air pollution by dust aerosol PM10 in Brno City, Czech Republic. *Atmospheric Environment*, 42, 8661-8673.
- Papanastasiou, D. K., Melas, D., and Kioutsioukis, I. (2007). Development and assessment of neural network and multiple regression models in order to predict pm10 levels in a medium-sized mediterranean city. *Water Air Soil Pollution*, 182, 325-334.
- Pérez, P., and Reyes, J. (2002). Prediction of maximum om 24-h average of PM10 concentrations 30h in advance in Santiago, Chile. *Atmospheric Environment*, 36, 4555-4561.
- Pires, J. C. M., Alvim-Ferraz, M. C. M., Pereira, M. C., and Martins, F. G. (2010). Prediction of PM10 concentrations through multi-gene genetic programming. *Atmospheric Pollution Research*, 1, 305-310.
- Pope III, C. A., and Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal Air Waste Management Association*, 56, 709-742.
- Schwarze, P. E., Øvreivik, J., Låg, M., Refsnes, M., Nafstad, P., Hetland, R. B., et al. (2006). Particulate matter properties and health effects: consistency of epidemiological and toxicological studies. *Human & Experimental Toxicology*, 25, 559-579.
- Sfetsos, A., and Vlachogiannis, D. (2010). A new methodology development for the regulatory forecasting of PM10. Application in the Greater Athens Area, Greece. *Atmospheric Environment*, 44, 3159-3172.
- Silva, C., Pérez, P., and Trier, A. (2001). Statistical modelling and prediction of atmospheric pollution by particulate material: two nonparametric approaches. *Environmetrics*, 12, 147-159.
- Stadlober, E., Hörmann, S., and Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment*, 42, 1098-1109.
- van der Wal, J. T., and Jansen, L. H. J. M. (1999). Analysis of spatial and temporal variations of PM10 concentrations in the Netherlands using Kalman filtering. *Atmospheric Environment*, 34, 3675-3687.
- Veselý, V., Tonner, J., Hrdličková, Z., Michálek, J., and Kolář, M. (2009). Analysis of PM10 air pollution in Brno based on generalized linear model with strongly rank-deficient design matrix. *Environmetrics*, 20, 676-698.

Authors' addresses:

Ernst Stadlober  
Institute of Statistics  
Graz University of Technology  
Kopernikusgasse 24/III  
8010 Graz  
Austria  
E-mail: e.stadlober@tugraz.at

Zuzana Hübnerová<sup>1</sup> and Jaroslav Michálek  
Institute of Mathematics, Department of Statistics and Optimization  
Brno University of Technology  
Technická 2896/2  
616 69 Brno  
Czech Republic  
E-mail: hrdlickova.z@fme.vutbr.cz and michalek@fme.vutbr.cz

Miroslav Kolář  
Department of Geography  
Faculty of Science  
Masaryk University  
Kotlářská 2  
611 37 Brno  
Czech Republic  
E-mail: kolar@sci.muni.cz

---

<sup>1</sup>maiden name Hrdličková