Prediction and forecast of daily PM10 concentrations in Brno and Graz

Ernst Stadlober
Graz University of Technology

Zuzana Hüberová, Jaroslav Michálek
Brno University of Technology

Österr. Statistiktage 2011, Graz

9. September 2011

Contents

Situation in Graz and Brno

- Meteorological and anthropogenic factors
- Measuring stations in Graz and Brno

Statistical Prediction Models

- Multiple Linear Regression
- Generalized Linear Model

Quality of the Models

The general Situation

Graz (263k inhabitants): Basin Area

- Low wind velocities, low precipitation, temperature inversion
- Traffic, domestic fuel

Brno (430k inhabitants): Open Basin Area

- Concentration of industry, motor traffic, heating plants
- Temperature inversion

Consequence

- Treshold value daily mean PM10 $\geq 50 \mu g/m^3$ exceeded regularly
- **Graz:** 36 – 105 exceedances in winter seasons
- **Brno:** 55 – 70 exceedances

Stations in Graz

Forecast: Graz-Mitte (traffic area), Graz-Süd (industrial zone)
Precipitation: Graz-Nord
Temperature inversion: Kalkleiten

Graz-Mitte: Inversion and Weekend Effect

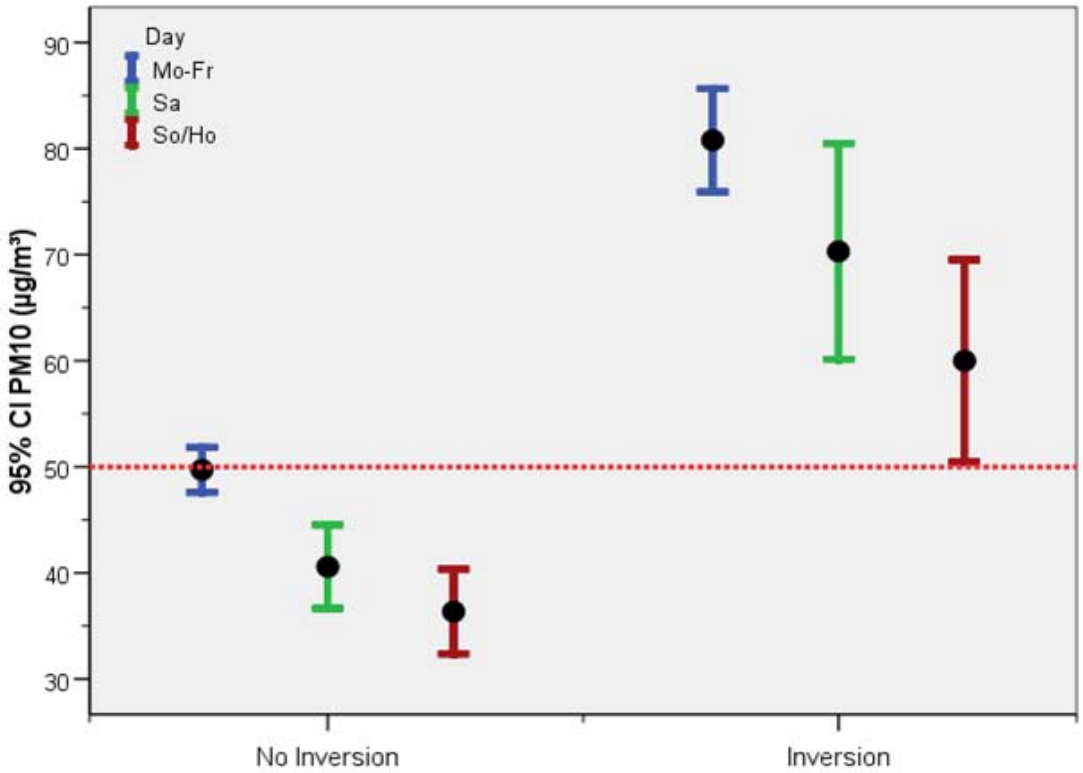


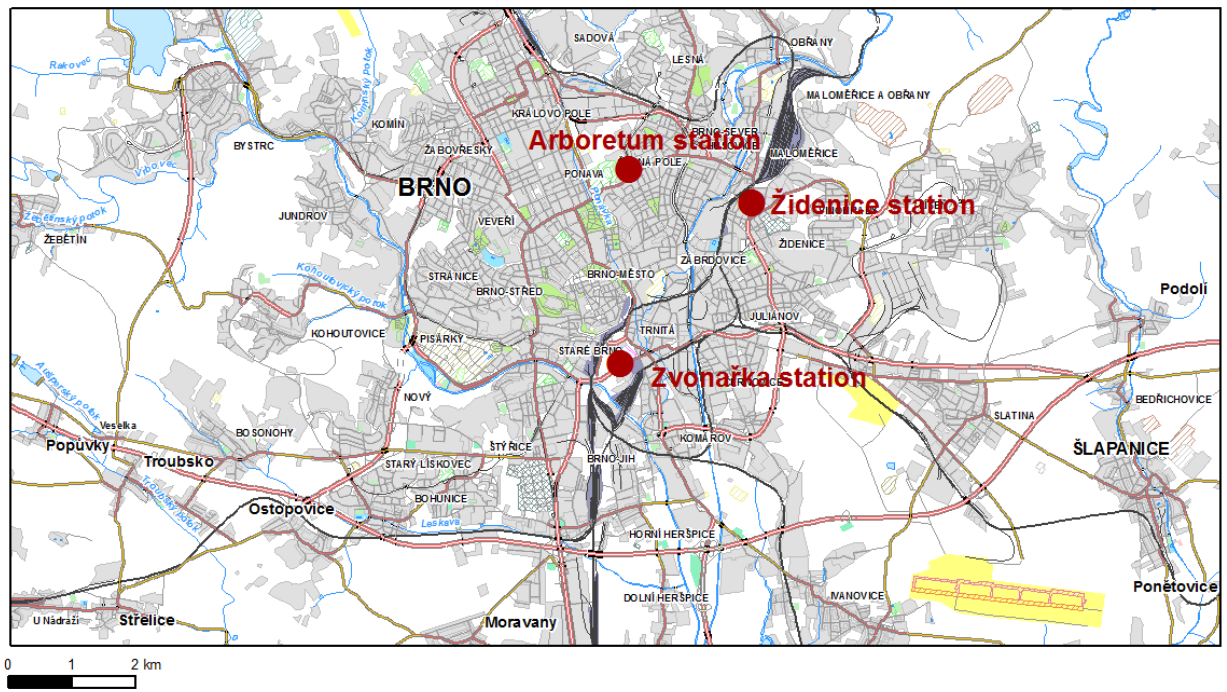
Figure: Graz-Mitte: Five winter seasons 02/03–06/07

No Inversion: 40% lower than Inversion
Sun/Holiday: 30% lower than Working Day

Stations in Brno

Forecast: Arboretum (botanical garden)

Židenice (heavy traffic street), Zvonařka (traffic spot)



Training Data and Test Data for Graz

Model fits on Training Data

- Graz-Mitte: 7 winter seasons (23.10.02–31.3.09)
- Graz-Süd: 6 winter seasons (1.10.03–31.3.09)
- All models are fitted on observed covariates

Predictions (one-day forecasts) on Test Data

- Graz-Mitte,Graz-Süd: 1 winter season (3.11.09 - 31.3.10)

Covariates for Graz

- Ap_t: daily mean PM10 day t
- ΔT_t: mean temp diff to test point Kalkleiten day t
- V_t: mean wind speed day t (Graz Mitte for Graz Süd)
- $Prec_t$: 1 = prec., 0 = no prec. on day t (Graz Nord)
- Ap_{lag}: mean of PM10 (12.30 $(t-2)$ to 12.00 $(t-1)$)
- T_{lag}: categorized mean (temp means: 12.30 $(t-2)$ to 12.00 $(t-1)$) (= 0 temp$>$0, = 1 temp$\leq$0)
  (Graz Mitte for Graz Süd)
- saturday: (0/1)
- sunday: (0/1)
- febr: (0/1)
- march: (0/1)

Training Data and Test Data for Brno

Model fits on Training Data

- Arboretum: 2 winter seasons 11/06–03/07, 10/07–03/08
- Židenice, Zvonařka: 1 winter season 11/07–03/08
- All models are fitted on predicted covariates

Predictions (one-day forecasts) on Test Data

- Arboretum, Židenice, Zvonařka:
  1 winter season (1.10.08–31.3.09)

Covariates for Brno

- Ap_t: daily mean PM10 day t
- $V_t \sin D_t(\mathrm{pred})$: mean of 4 predictions (=velocity*sin(direction)) (0.00, 6.00, 12.00, 18.00 day t)
- $V_t \cos D_t(\mathrm{pred})$: mean of 4 predictions (=velocity*cos(direction))
- $T_t(\mathrm{pred})$: mean of 4 predictions of temperature
- $Cover_t(\mathrm{pred})$: mean of 4 predictions of coverage(values 1:10)
- Ap_{lag}: mean of PM10 (13.00 $(t-2)$ to 12.00 $(t-1)$)
- T_{lag}: mean of 4 measurements of temperature (18.00 $(t-2)$, 0.00, 6.00, 12.00 $(t-1)$)

Covariates for Brno - 2

- HS_t: activity of heating plants day t
  (Oct: 0.08, Nov: 0.14, Dec: 0.17,
  Jan; 0,19, Feb: 0.16, Mar: 0.14)
- weekend: (0/1)
- sunday: (0/1)
- febr: (0/1)
- march: (0/1)

Multiple Linear Regression

Dependent variable $\sqrt{Ap_t}$

Model assumption

- $x_t^{(i)}$ *($i = 1, 2, \ldots, m$) metric input variables*
- $d_t^{(j)}$ *($j = 1, \ldots, p$) dummy (0/1) input variables at day t*
- *Linear model for $\sqrt{Ap_t}$:*

$$\sqrt{Ap_t} = \beta_0 + \sum_{i=1}^{m} \beta_i x_t^{(i)} + \sum_{j=1}^{p} \beta_{m+j} d_t^{(j)} + \epsilon_t \qquad (1)$$

$$\text{where} \quad \epsilon_t \overset{iid}{\sim} N(0, \sigma^2)\,.$$

Multiple Linear Regression: Graz-Mitte

Variables x or d	beta	sd(beta)	t-value
const	7.41	0.12	61.5
ΔT_t	-0.30	0.018	-16.6
Ap_{lag}	0.02	0.001	15.1
V_t	-1.59	0.132	-12.0
T_{lag}	0.95	0.084	11.4
sunday	-0.97	0.085	-11.4
march	0.75	0.095	7.9
saturday	-0.69	0.094	-7.3
$Prec_t$	-0.51	0.071	-7.1
febr	0.34	0.090	3.8
	R^2	se	n
	0.63	1.10	1199

Multiple Linear Regression: Zvonařka in Brno

Variables x or d	beta	sd(beta)	t-value
const	8.78	0.543	16.2
$V_t \sin D_t$	0.36	0.059	6.0
Cover_t	-0.26	0.052	-5.0
march	-1.04	0.313	-3.3
T_{lag}	-0.19	0.058	-3.2
T_t	0.17	0.062	2.7
weekend	-0.63	0.242	-2.6
Ap_{lag}	0.01	0.005	2.1
$V_t \cos D_t$	-0.08	0.056	-1.5
β_{45}		0.37	
S		103 °	
	R^2	se	n
	0.54	1.24	132

- HS_t, sunday, febr not significant

GLM model with gamma distribution and log link

Dependent variable Ap_t is gamma distributed

Model assumption

- $x_t^{(i)}$ *$(i = 1, 2, \ldots, m)$ metric input variables*
- $d_t^{(j)}$ *$(j = 1, \ldots, p)$ dummy (0/1) input variables*
- Ap_t *has Gamma$(\phi\mu_t, 1/\phi)$ distribution, ϕ dispersion*

$$\ln(\mu_t) = \ln\left(E(Ap_t|S_t)\right) = \beta_0 + \sum_{i=1}^{m} \beta_i x_t^{(i)} + \sum_{j=1}^{p} \beta_{m+j} d_t^{(j)}\,, \qquad (2)$$

$E(Ap_t|S_t)$ cond. expect. of Ap_t for given set S_t of variables

GLM with gamma log link: Graz-Mitte

Variables x or d	beta	sd(beta)
const	3.12	0.081
ΔT_t	-0.08	0.005
$\ln(Ap_{\mathrm{lag}})$	0.31	0.020
V_t	-0.50	0.038
T_{lag}	0.27	0.024
sunday	-0.29	0.024
march	0.22	0.027
saturday	-0.21	0.027
$Prec_t$	-0.15	0.021
febr	0.09	0.026
χ^2_{ans}	115.08	
f	1189	
n	1199	

GLM with gamma log link: Zvonařka in Brno

Variables x or d	beta	sd(beta)
const	4.53	0.136
$V_t \sin D_t$	0.12	0.017
T_{lag}	-0.07	0.015
Cover_t	-0.07	0.016
march	-0.34	0.096
T_t	0.06	0.017
weekend	-0.17	0.069
febr	0.18	0.090
$V_t \cos D_t$	-0.02	0.016
$\hat{\beta}_S$	0.13	
S	101°	
χ^2_{ans}	17.98	
f	125	
n	134	

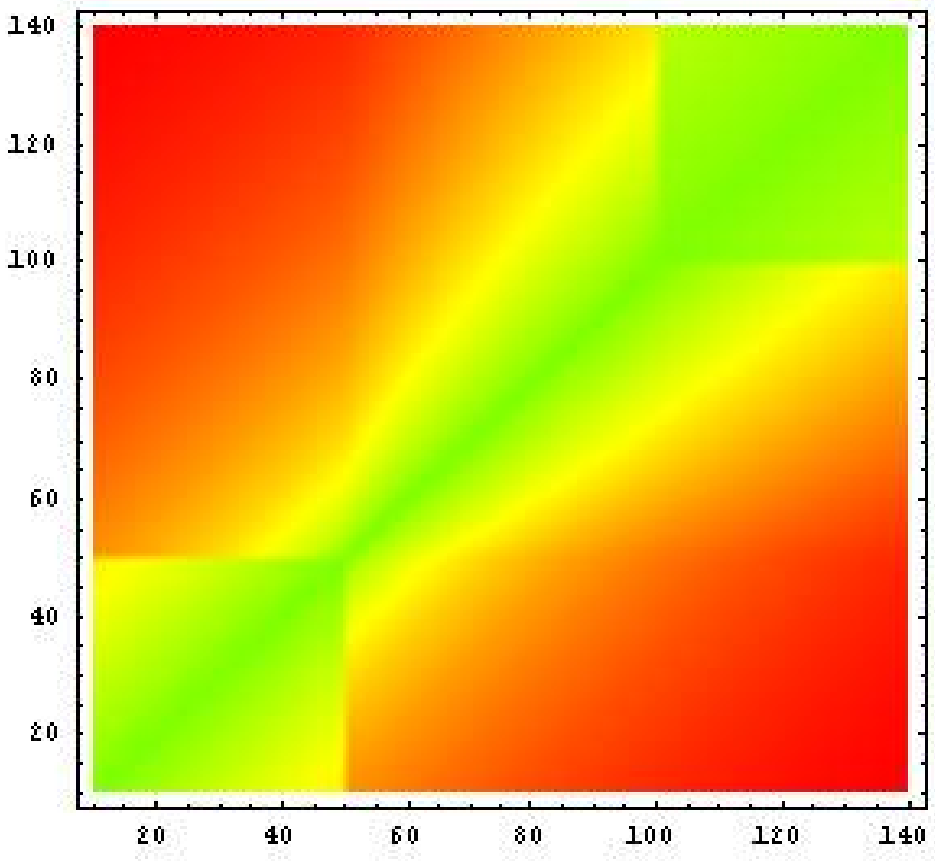
- Ap_{lag}, HS_t, sunday not significant

Quality of forecasting

- **Training data set** to compute estimates $\hat{\beta}_i$, $\hat{\beta}_{j+m}$
- Model LM applied to **test data** yield estimates $\widehat{\sqrt{Ap_t}}$
- $\widehat{Ap_t} = (\widehat{\sqrt{Ap_t}})^2$
- Model GLM applied to **test data** yield estimates $\widehat{\ln(Ap_t)}$
- $\widehat{Ap_t} = \exp(\widehat{\ln(Ap_t)})$

Quality of forecasting - 2

Quality function $Q(x, y)$: (x =observation, y =forecast)



Quality: 5 levels green (1=excellent) to red (5=very bad)

Practical performance and comparison: Graz-Mitte

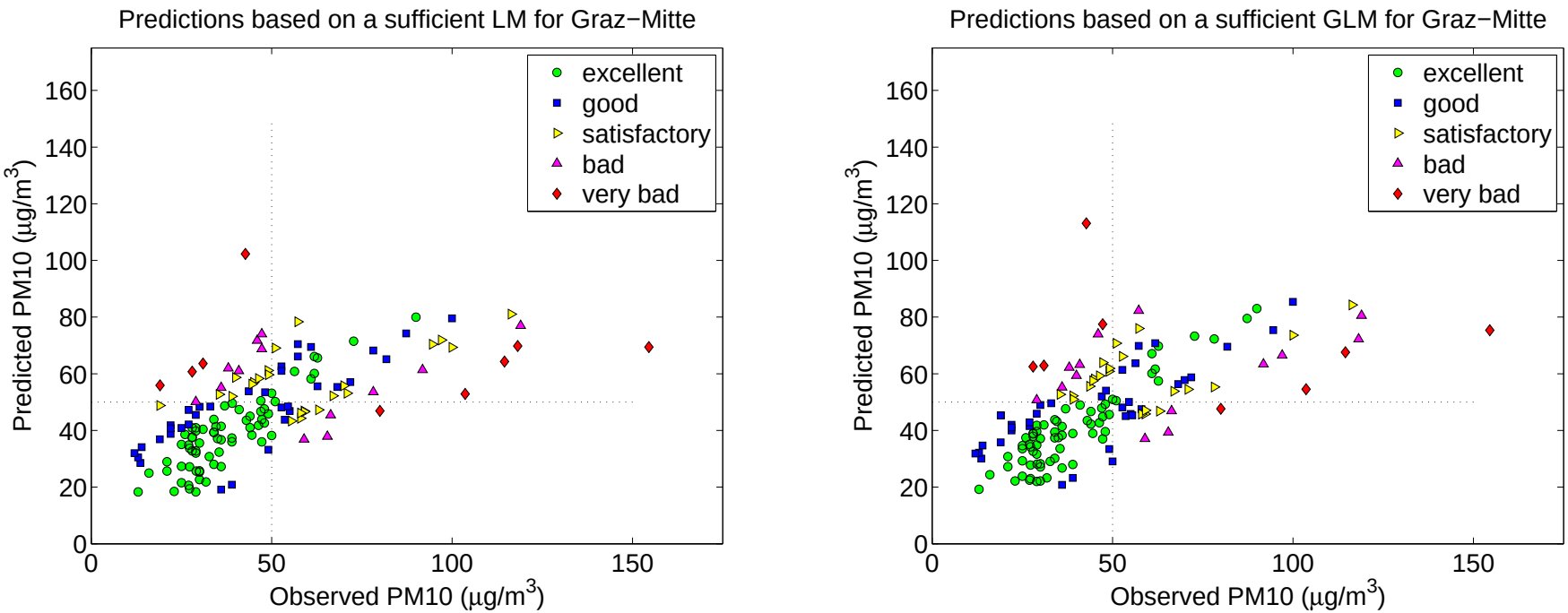


Figure: Graz-Mitte: Quality function of LM and GLM

Practical performance and comparison: Graz-Mitte - 2

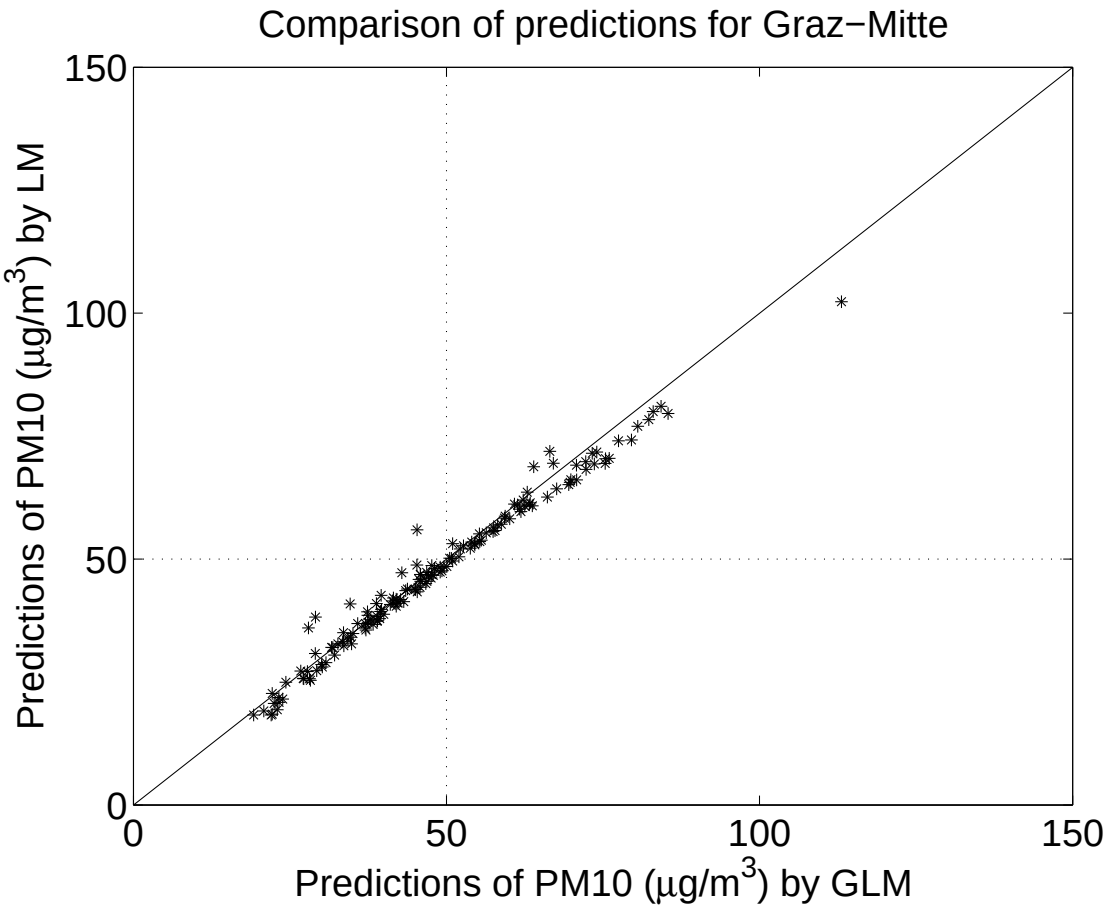


Figure: Graz-Mitte: Forecasts $\widehat{Ap_t}$ of GLM vs LM

Practical performance and comparison: Graz Mitte - 3

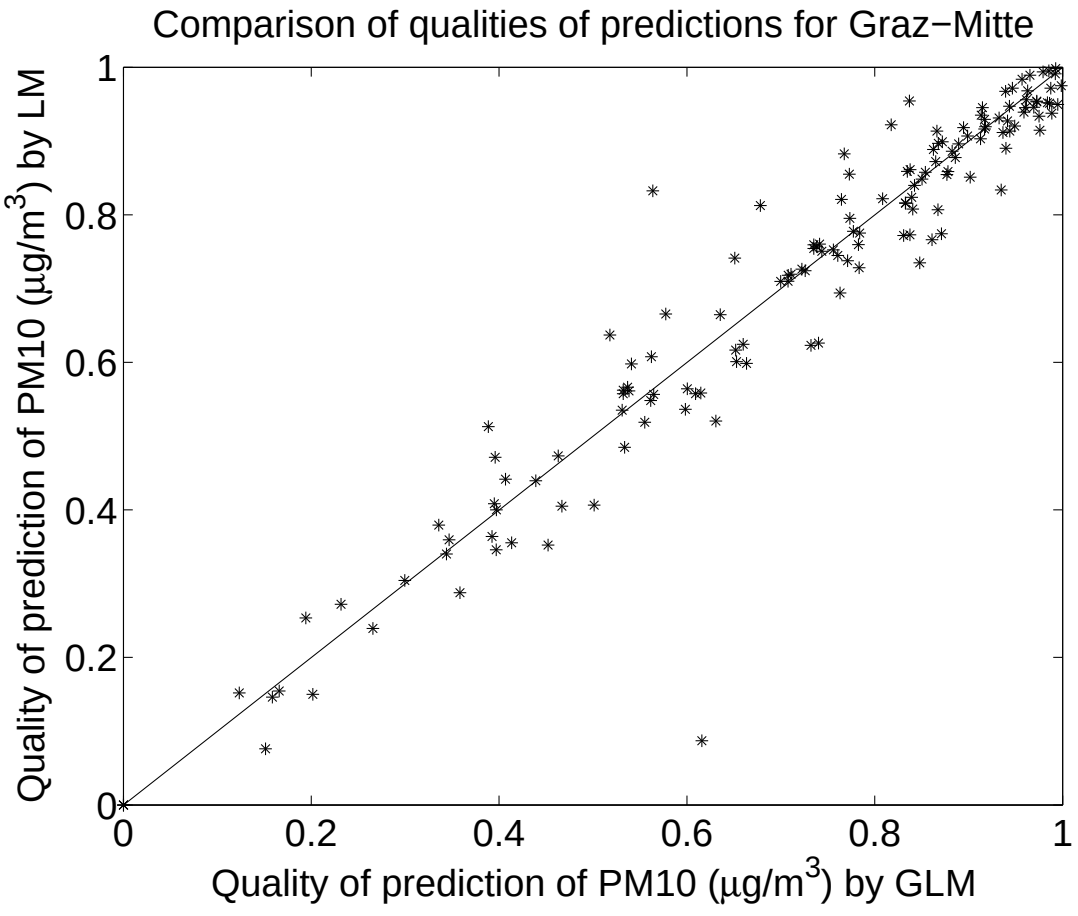


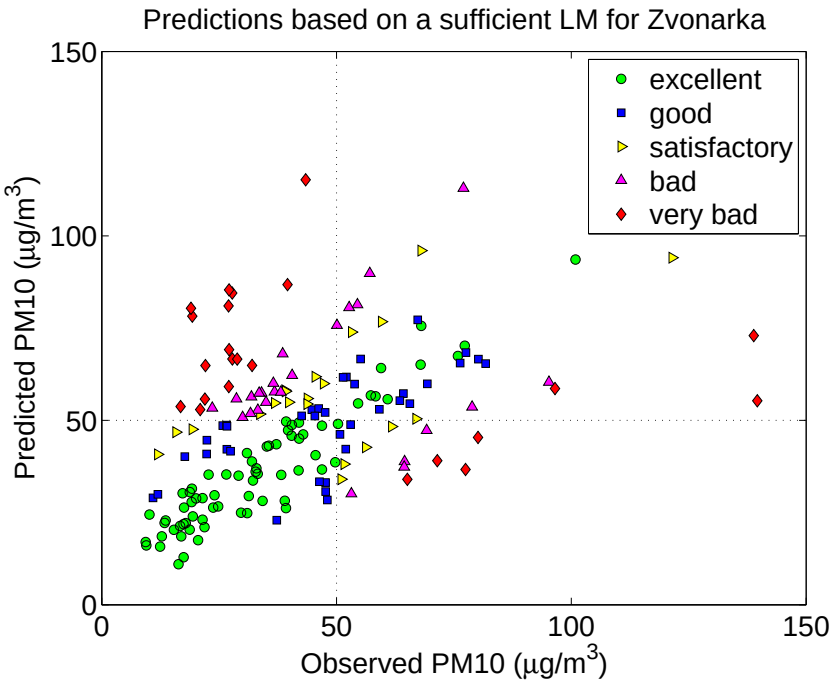
Figure: Graz-Mitte: Quality function GLM vs LM

Practical performance and comparison: Graz Mitte - 5

Table: Contingency table Graz-Mitte: Quality of GLM (rows) vs LM (columns), no. and perc. of quality 1 to 5

GLM/LM	1	2	3	4	5	sGLM	%	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
sLM	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			

Practical performance and comparison: Zvonařka



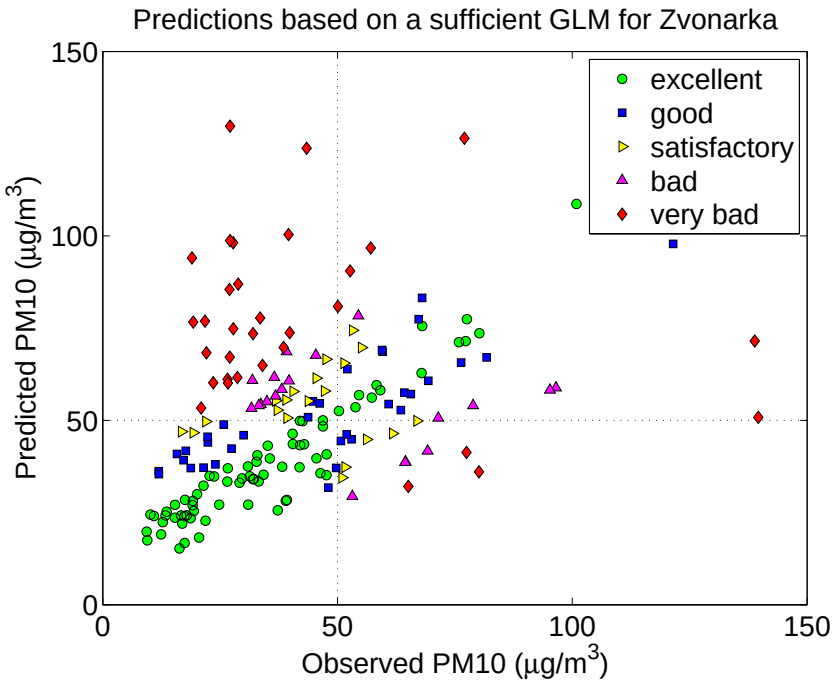


Figure: Zvonařka: Quality function of LM and GLM

Practical performance and comparison: Zvonařka - 2

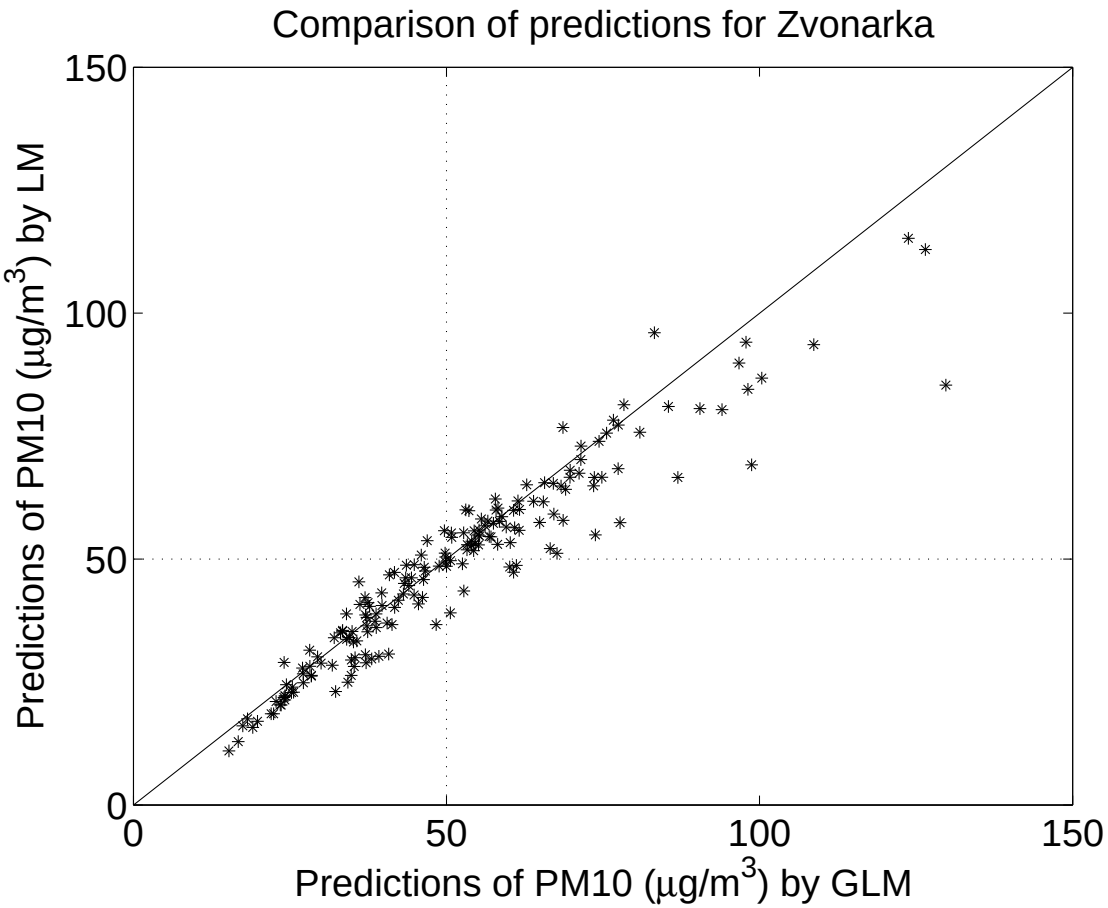


Figure: Zvonařka: Forecasts $\widehat{Ap_t}$ of GLM vs LM

Practical performance and comparison: Zvonařka - 3

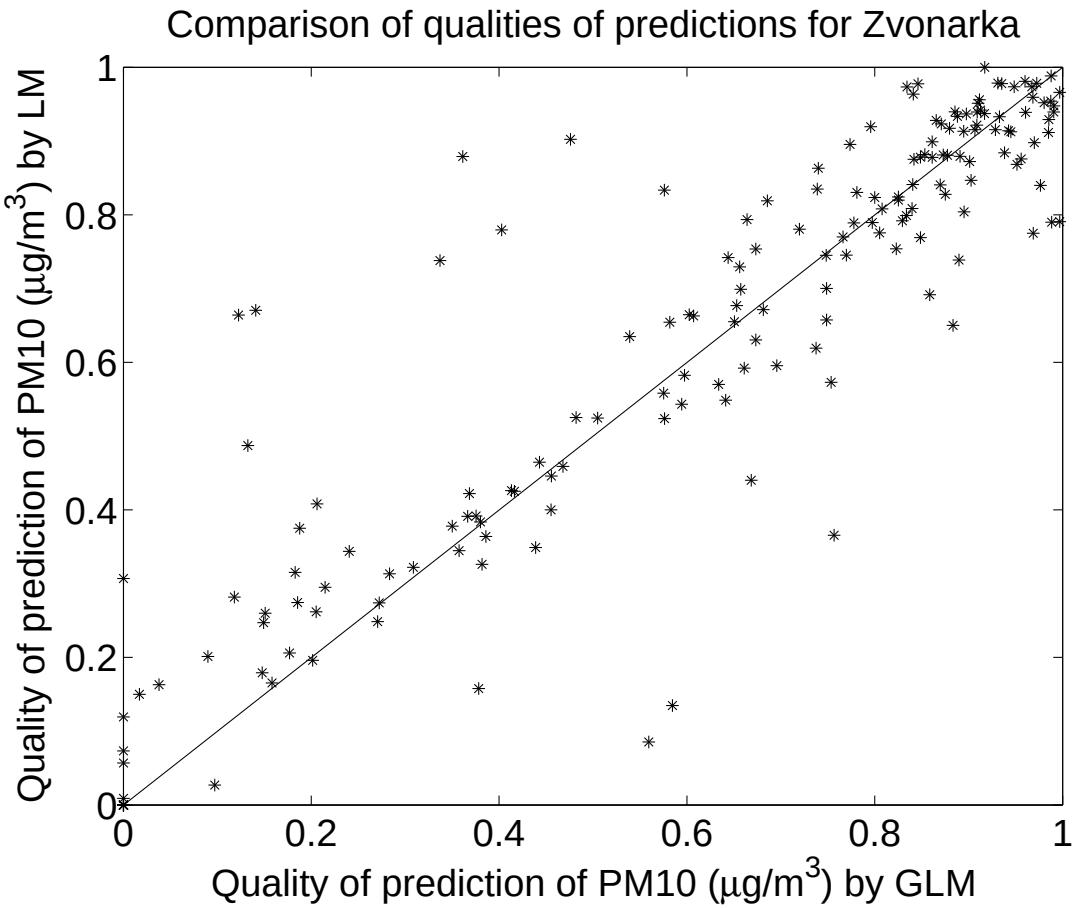


Figure: Zvonařka: Quality function GLM vs LM

Practical performance and comparison: Zvonařka - 4

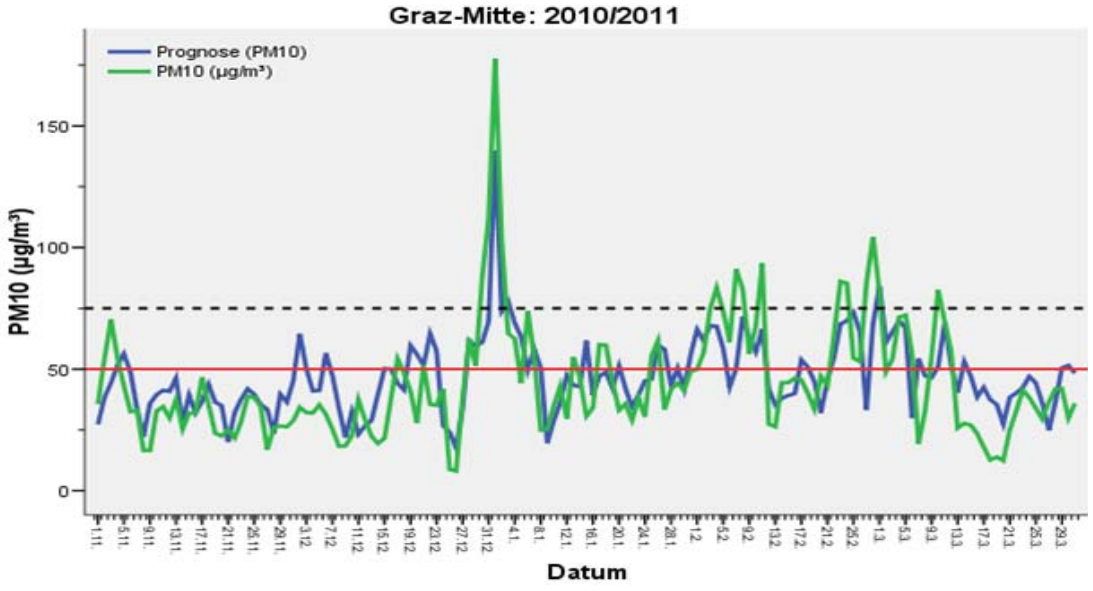
Table: Contingency table Zvonařka: Quality of GLM (rows) vs LM (columns), no. and perc. of quality 1 to 5

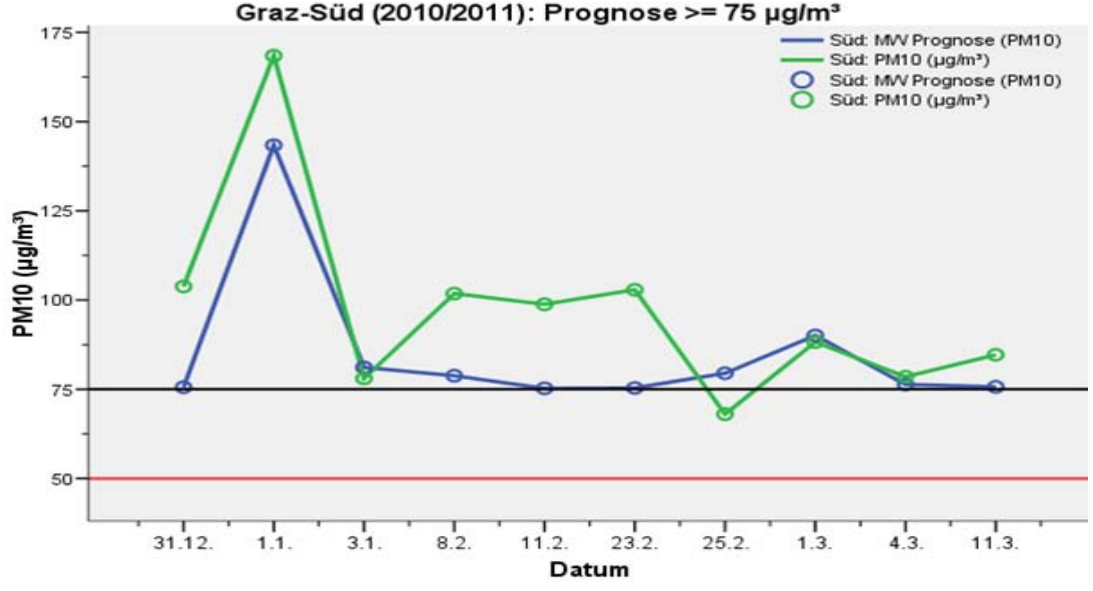
GLM/LM	1	2	3	4	5	sGLM	%	cum %
1	61	11	0	0	0	72	40.7	40.7
2	7	20	6	1	0	34	19.2	59.9
3	2	3	12	1	2	20	11.3	71.2
4	1	1	2	14	2	20	11.3	82.5
5	0	2	1	9	19	31	17.5	100.0
sLM	71	37	21	25	23	177		
%	40.1	20.9	11.9	14.1	13.0	100.0		
cum %	40.1	61.0	72.9	87.0	100.0			

Practical Experience Graz-Mitte and Graz-Süd

- Daily predictions of PM10 since winter season 04/05
- Meteorological forecasts by ZAMG Styria till 11 in the morning
- PM10-Prediction during 1 and 3 PM at www.oekostadt.graz.at
- Daily update: comparison of predictions with measured values (day $t-2$)
- e-mail with excel-file to Environmental Department of Graz
- 80 – 90% of the predictions at least satisfactory

Prediction Graz-Mitte, Graz-Süd 2010/2011





Conclusion

- LM and GLM models suitable for daily PM10 forecasts
- Results of models similar and comparable
- Brno: 1 or 2 seasons for training data, test data 2008/09
- Graz: 6 or 7 seasons for training data, test data 2009/10
- Brno: no measures for precipitation and inversion
  for each site specific selection of covariates
- Graz: all 9 covariates are significant
- Very Good quality: Arboretum (98% satisfactory)
- Good quality: Graz-Mitte, Graz-Süd (85%)
- Moderate quality: Zvonařka, Židenice (74%)