STATISTICAL PROGNOSTIC MODEL FOR DAILY MAXIMA OF CONCENTRATIONS OF URBAN AIR POLLUTANTS

E.L. Genikhovich, L.R. Sonkin, V.I. Kirillova

Main Geophysical Observatory,

St. Petersburg, Russia

CONTENT

- Major factors influencing urban air pollution;
- Deterministic and statistical air pollution forecasting;
- Operational statistical models employed in Russia;
- Description of the model presented here and some results of its validation with "conventional" pollutants;
- Prediction of daily maxima of ozone concentrations;
- Conclusion

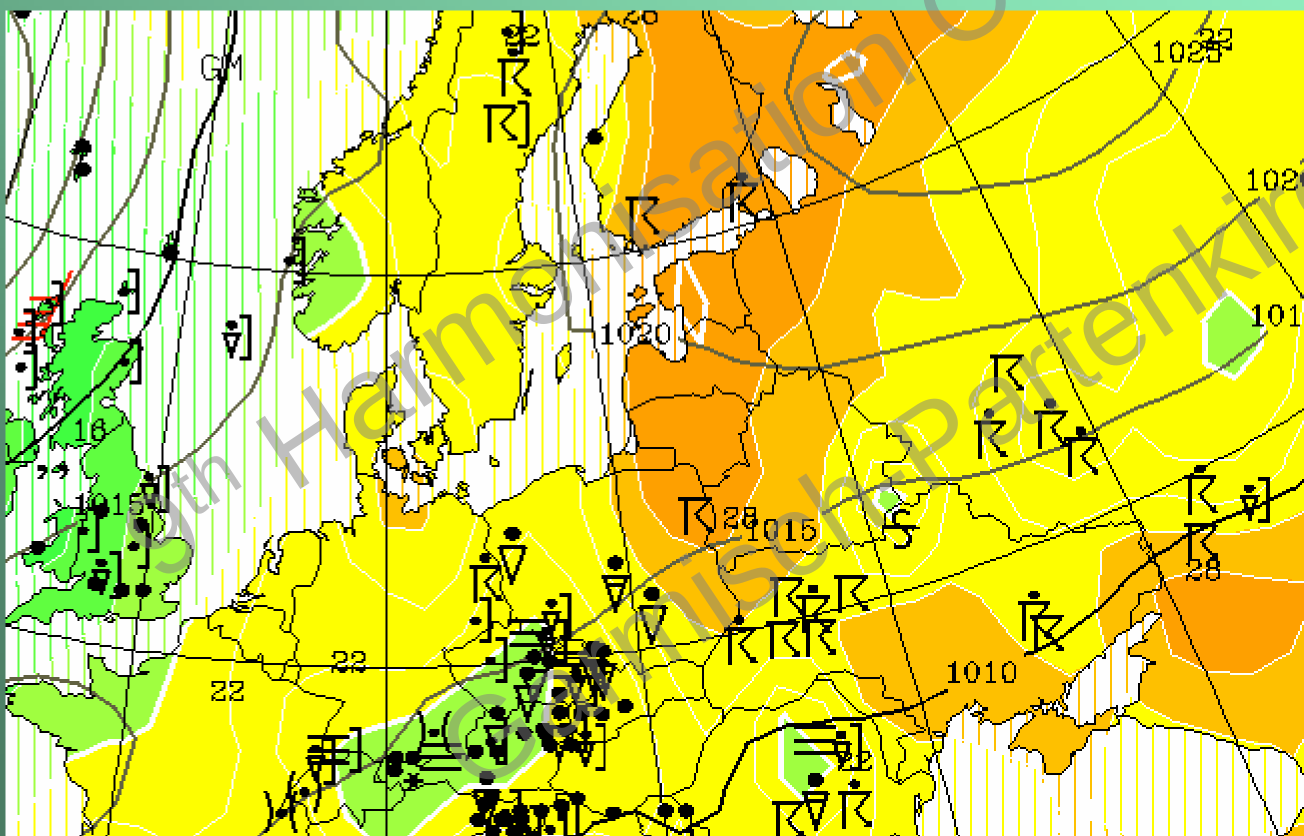
MAJOR FACTORS INFLUENCING URBAN AIR POLLUTION /1

Urban sources of emission, temporal and spatial variation of their characteristics

Urban vehicular and industrial emissions

MAJOR FACTORS INFLUENCING URBAN AIR POLLUTION /2

Meteorological and synoptic conditions



Midday surface pressure and air temperature fields corresponding to the highest ozone concentrations in St. Petersburg in 2002, July 31 (weather map from the Köln University site)

MAJOR FACTORS INFLUENCING URBAN AIR POLLUTION /3

Structure of the urban canopy, location, shape and size of buildings, roads etc

Marc Chagall

"Above the city"

URBAN AIR POLLUTION FORECASTS -- WHAT FOR?

- To warn people about possible high-pollution episodes in the city;
- To provide information for policy- and decision-makers in order to implement emission-control measures, like emission reduction at stationary sources, limitations imposed on traffic intensity and pattern etc, resulting in reduction of peak concentrations

PROBLEMS WITH PREDICTION OF URBAN AIR POLLUTION

- Emission parameters and their temporal and spatial variability are not known well;
- Fields of mean and turbulent characteristics of the wind flow transporting pollutants are non-uniform and have sharp gradients;
- Concentration fields are highly irregular in space and time;
- Concentrations are excessively noisy (include intensive stochastic component);

TWO MAIN APPROACHES TO AIR POLLUTION FORECASTING

<u>Deterministic forecast</u>

- Based on general physical principles and semi-empirical parameterizations;
- Very sensitive to errors in emission- and meteorological data;
- Cannot not reproduce the stochastic component of air pollution;
- Needs monitoring data only for validation purposes;

<u>Statistical forecast</u>

- Based on empirical relationships between characteristics of the air pollution ("predictants") and governing parameters ("predictors");
- Doesn't make use of emission data;
- In principle, can reproduce all components of the turbulent spectra;
- Uses monitoring data for constructing and teaching the model as well as for its "initialization";

STATISTICAL FORECASTS IN RUSSIA: STATUS REPORT

- Routine forecasts produced in 235 cities on daily basis in 2002 using statistic models (SMs) developed for each of these cities using the same methodology;
- Dimensionless parameter P predicted with SMs characterizes unfavorable meteorological conditions that could lead to high levels of air pollution over the whole urban area:

$$P = m/n,$$

where "n" is the total number of measurements during the day and "m" is the number of measurements in excess of 1.5 of corresponding seasonal average values;

- A quantified characteristic of the synoptic situation is used as a synoptic predictor S;
- "Previous" value of P is used as a predictor ("inertial factor");

SPECIFICS OF THE MODEL DISCUSSED IN THIS TALK

- Statistical;
- Predicts daily maxima of concentrations at monitoring stations rather than their individual values corresponding to certain moments of time or average characteristics of air pollution over the city;
- Can be efficiently used for conservative pollutants as well as ozone;
- Selects initial set of predictors from physical considerations;
- Includes stepwise transformations of predictors;
- Finally selects predictors and constructs the prognostic model using stepwise multiple regression;

MAIN STEPS OF TRANSFORMATION OF PREDICTORS

- Normalization of daily maxima (optional) :

  Cmax →B = Cmax / Cseason;

- Censuring the sample to exclude smallest concentrations (optional; uses P);
- Transformation of the predictant to the normal distribution;
- Transformation of nonlinear dependencies in linear ones: [X] = E{B|X}.

Here, X - predictor and E – math. expectation.

GENERAL FORMULATION OF THE MODEL IN USE

The prognostic equation is written as follows:

$$B_{t+1} = a_0 + \Sigma a_i[X_i] ,$$

where $[X_i]$ are transformed predictors X_i at the moment "t", and coefficients "a" are determined using the method of the stepwise regression.

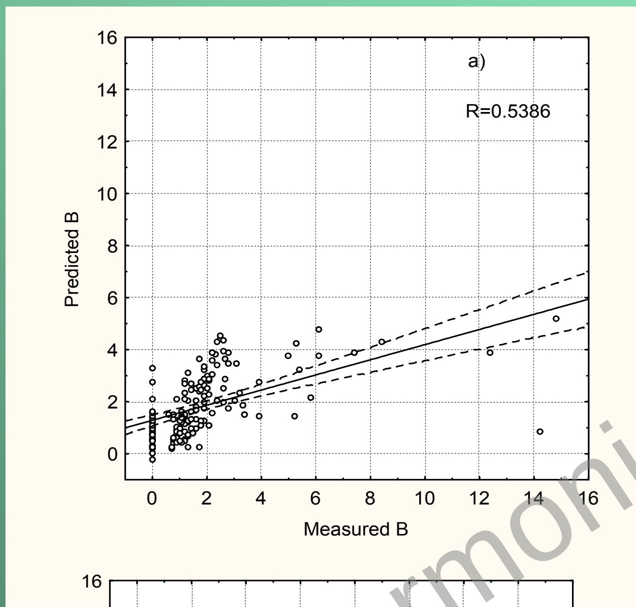
DESCRIPTION OF DATA SETS

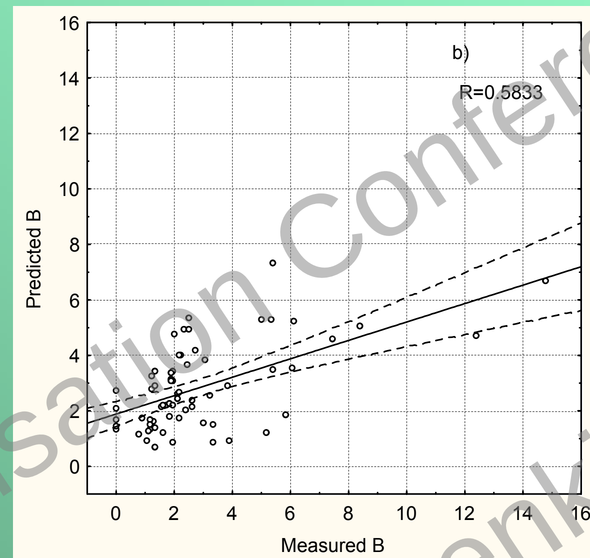
City	Pollutants	Years
Krasnoyarsk, Siberia	Carbon disulfide, hydrogen fluoride (manual sampling)	1983 - 1985
Ufa, European Russia	Ethylbenzene, benzene (manual sampling)	1988 - 1989
St. Petersburg, European Russia	Ozone, nitrogen dioxide (DOAS at two levels in a street canyon)	2002

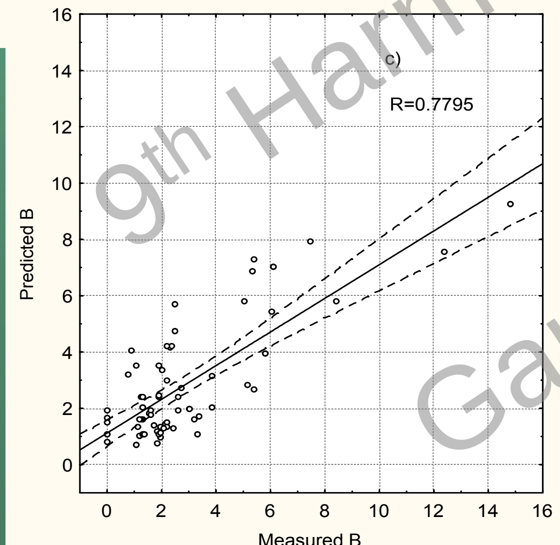
EFFICIENCY OF TRANSFORMATIONS /1

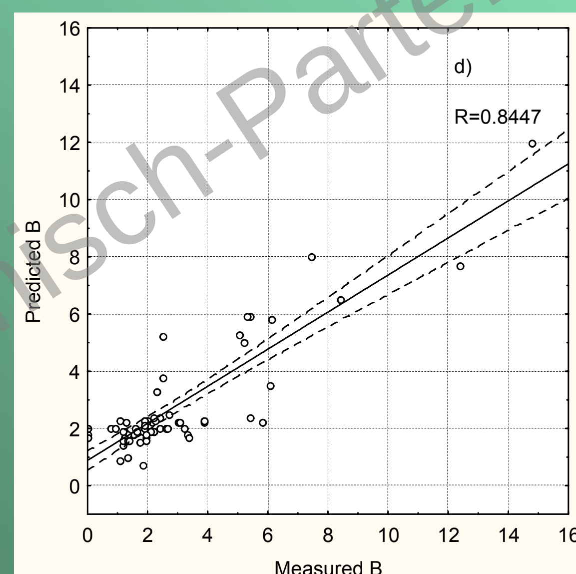
Fig	Trans-formation	Sam-ple	$\langle B_m \rangle$	$\langle B_p \rangle$	σ_m	σ_p	Corr	$\langle m/p \rangle$	$\sigma_{m/p}$
a)	No	160	1.82	1.82	2.26	1.22	0.54	1.10	1.61
b)	Censur.	70	2.82	2.82	2.62	1.51	0.58	1.10	0.91
c)	Censur. + Linear	70	2.82	2.77	2.62	1.64	0.78	1.09	0.65
d)	Censur. +Norm +Linear	70	2.82	2.82	2.62	1.96	0.84	1.05	0.57

EFFICIENCY OF TRANSFORMATIONS /2









AN EXAMPLE OF THE STATISTICAL MODEL FOR DAILY MAXIMA OF CS_2 IN KRASNOYARSK (WARM SEASON)

$$B_{t+1} = 0.45[B_t]+0.69[P]+0.52[d]+0.68[H_{INV}]-4.55$$

where B_t and B_{t+1} are daily maxima for days t and t+1, d – wind direction; H_{INV} – height of the lower boundary of the elevated inversion, P is a prognostic value for the day t+1, and other predictors in the right-hand side are determined from measurement carried out on day t+1 at 7 a.m.

Predictor	$[B_t]$	[P]	[d]	$[H_{INV}]$
Correlations with B_{t+1}	0.44	0.67	0.56	0.41
Multiple correlation	0.78			

PERFORMANCE OF STATISTICAL FORECASTS AT INDIVIDUAL MONITORING STATIONS

City	Pollutant	(R^2; F) at each station
Krasnoyarsk	CS_2	(0.34;7.8), (0.36;14.0), (0.36;10.2), (0.55;30.3), (0.61;22.4), (0.52;21.4), (0.30;7.6), (0.36;9.9)
Krasnoyarsk	HF	(0.38;11.2), (0.37;10.1), (0.42;9.6), (0.48;11.2), (0.48;10.9), (0.42;9.5), (0.36,17.2)
Ufa	Ethylbenzene	(0.61;22.4), (0.55;26.4), (0.50;20.3)
Ufa	Benzene	(0.45;14.4), (0.48;16.4), (0.37;9.8)

On average, R^2 ~ 0.4 – 0.6 with independent data sets

MEASURED IN KRASNOYARSK TO PREDICTED CS_2 CONCENTRATIONS (EPISODES WITH B>5 AT FOUR OR MORE STATIONS)

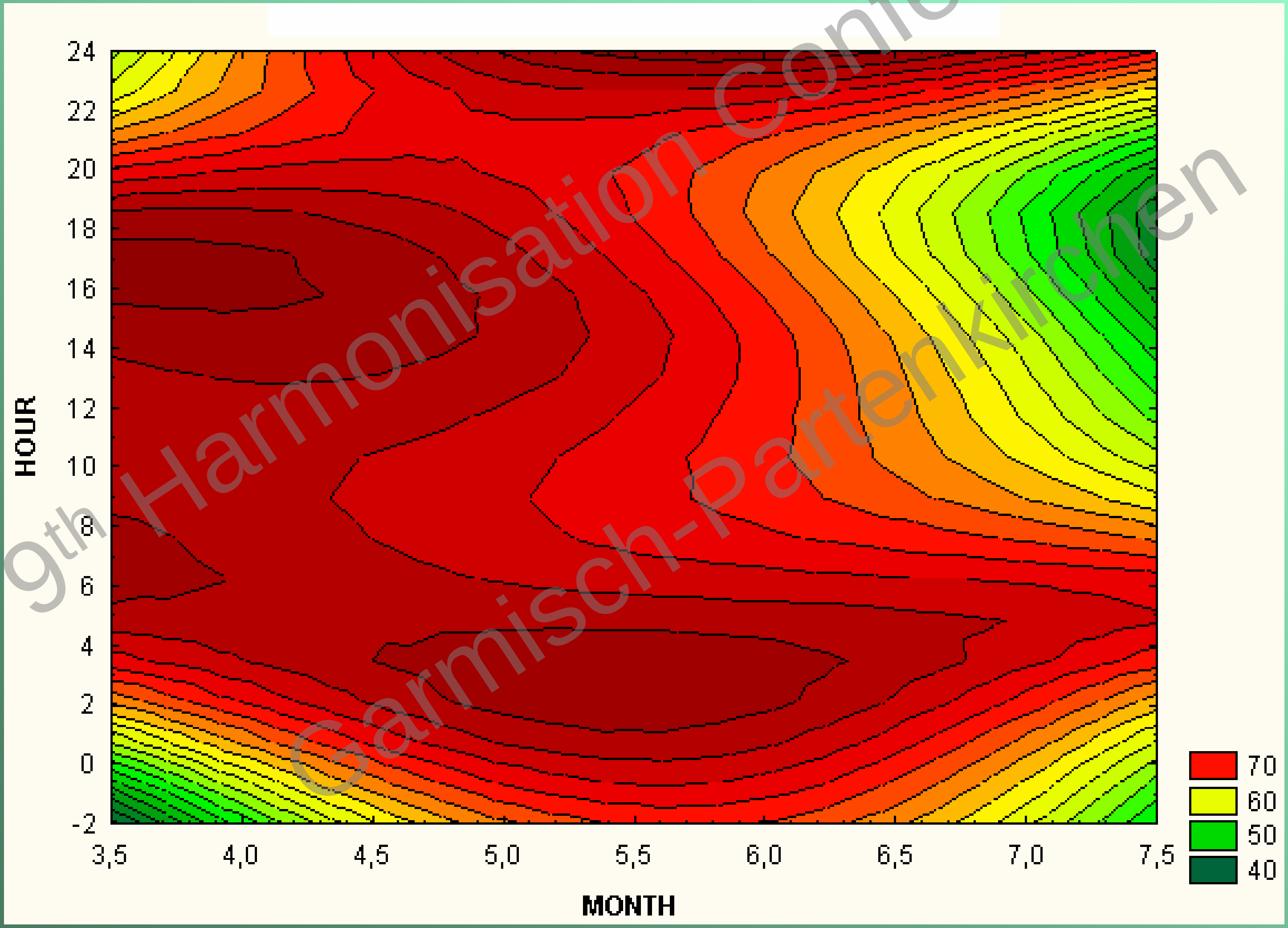
Station number	1	3	5	7	8	9	20	21
7.07.84	**6.32** **2.46**	4.11 5.11	4.32 3.42	**5.03** **5.81**	……….	**5.24** **5.20**	4.30 4.81	**8.22** **4.60**
20.07.84	0.00 1.34	0.00 1.48	0.00 1.02	**5.82** **3.82**	**5.67** **3.44**	**6.41** **3.62**	**9.19** **3.18**	0.00 2.00
21.07.84	2.35 3.48	1.57 2.60	**6.68** **4.59**	**6.10** **7.05**	4.34 4.89	**6.34** **6.44**	**9.96** **6.17**	3.17 4.16
3.08.84	**9.10** **4.88**	4.68 6.39	0.00 7.26	**7.45** **7.91**	**9.00** **6.02**	**5.58** **8.75**	**6.92** **11.62**	**7.56** **5.09**
4.08.84	4.61 6.10	**6.82** **7.03**	**6.28** **5.69**	**5.39** **7.28**	**7.95** **6.64**	4.93 6.65	4.92 8.32	3.87 6.31
2.07.85	**6.09** **4.12**	**16.11** **9.03**	**21.83** **10.60**	**8.41** **6.54**	**13.61** **9.33**	**25.46** **12.10**	**24.60** **10.96**	**9.71** **5.68**
3.07.85	**6.18** **5.37**	**15.87** **9.32**	**13.83** **10.91**	**14.86** **12.0**	4.00 4.51	**12.30** **10.13**	**16.11** **11.35**	……….

Overall statistics: $R^2 = 0.58$; $Prob\{B_{meas}/B_{predict} < 0.5\} = 0.75$

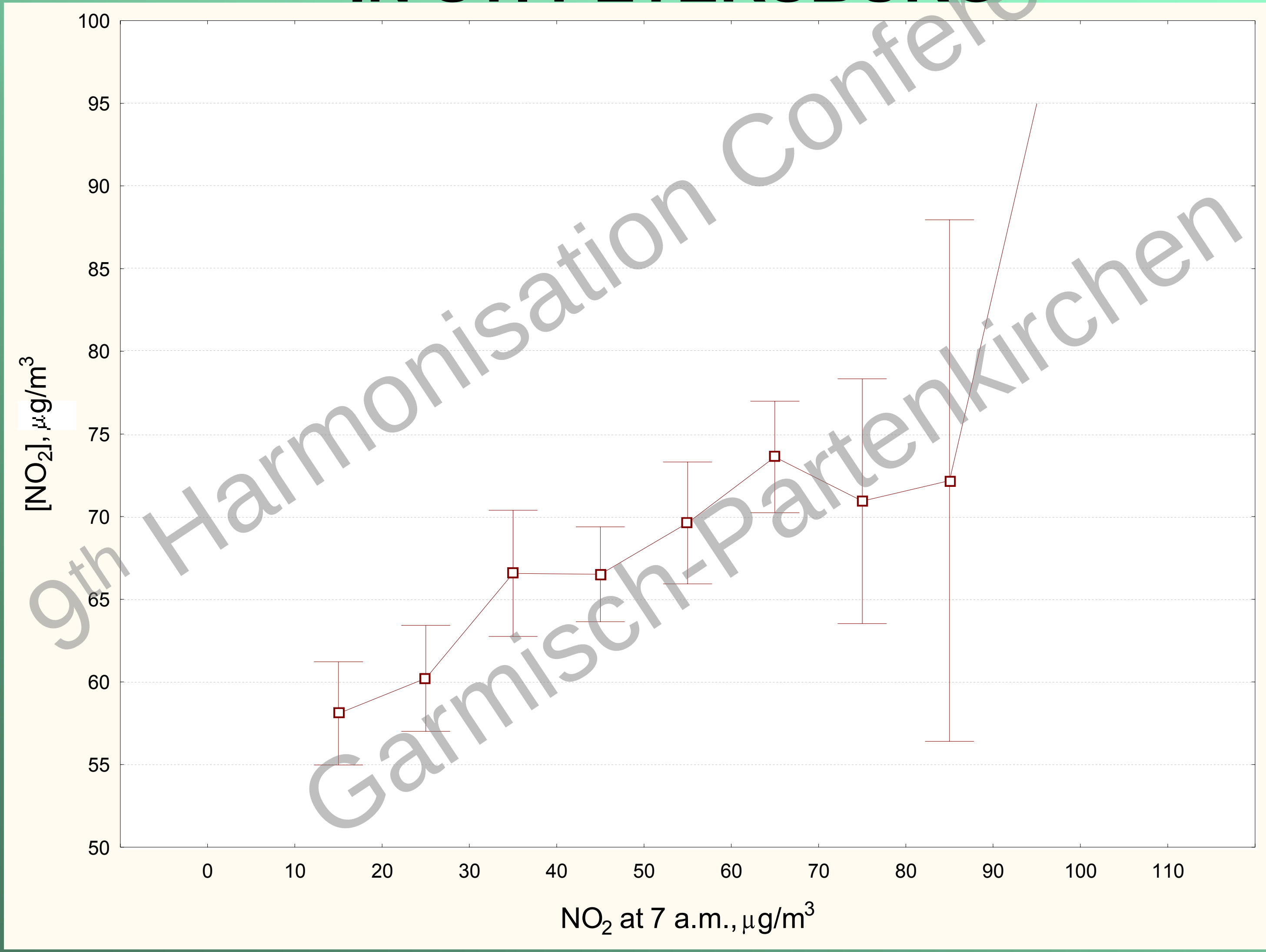
OZONE MONITORING IN ST. PETERSBURG

- Started in a street canyon in 1998 with two DOAS instruments are located at 3 m and 15 m height; several monitors were additionally put in operation last year by the city;
- The data set in the street canyon is not homogeneous because the pattern of traffic changed several times;
- Meteorological mast is mounted on the top of the building with DOAS instruments;
- Concentrations are usually rather moderate;

DAILY OZONE MAXIMA IN 2002 DEPENDING ON THE DATE AND HOUR OF OBSERVATIONS



TANSFORMATION NO_2 → $[NO_2]$ FOR PREDICTING OZONE IN ST. PETERSBURG



STATISTICAL MODEL # 1 FOR DAILY OZONE MAXIMA IN ST. PETERSBURG (WARM SEASON)

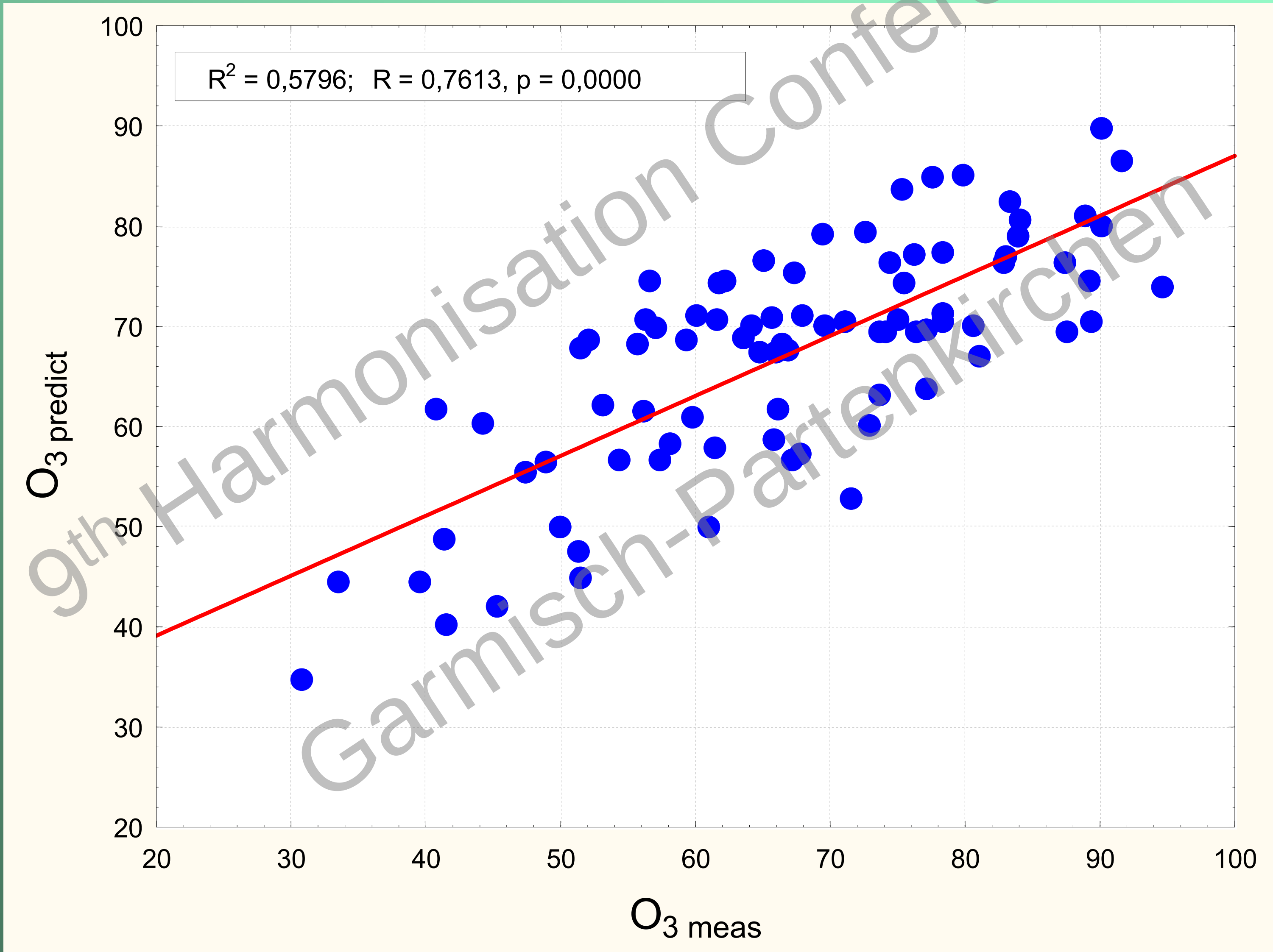
$O_{3,t+1} = 0.684[O_{3,t}]+0.705[d]+ 0.53[NO_2] + 0.525[T]+0.683[V]- 141.9$

where O_{3t} and O_{3t+1} are daily maxima for days t and t+1,
d – wind direction, T – air temperature, V – wind speed,
NO_2 – concentration of nitrogen dioxide;
all predictors in the right-hand side are determined from measurements carried out on day t+1 at 7 a.m.

Predictor	$[O_{3t}]$	[d]	$[NO_2]$	[T]	[V]
Correlations with O_{3t+1}	0.25	0.11	0.34	- 0.15	- 0.05

Coefficient of multiple correlation R = 0.76

PREDICTED VS MEASURED OZONE CONCENTRATIONS IN ST. PETERSBURG – STATISTICAL MODEL #1



STATISTICAL MODEL # 2 FOR DAILY OZONE MAXIMA IN ST. PETERSBURG (WARM SEASON)

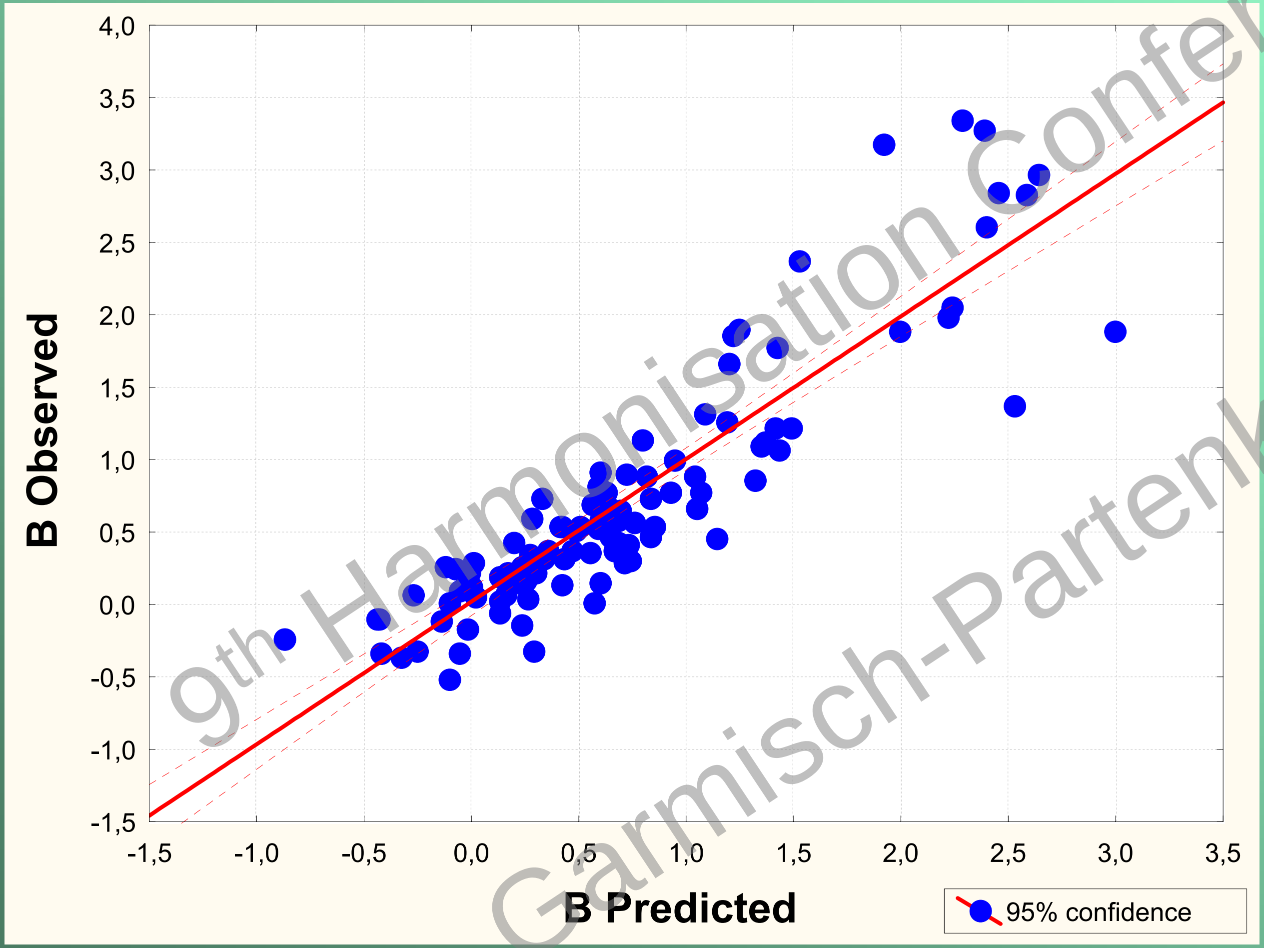
$$B_{t+1} = 0.404[B_t]+0.322[d]+0.205[NO_2]+0.186[T]+0.175[V]-141.9$$

WHERE $B = (O_{3,MAX}-O_{3,7})/O_{3,7}$

Predictor	$[B_t]$	[d]	$[NO_2]$	[T]	[V]
Correlations with B_{t+1}	0.59	0.46	0.34	0.36	0.15

Coefficient of multiple correlation $R \sim 0.9$

PREDICTED VS MEASURED OZONE CONCENTRATIONS IN ST. PETERSBURG – STATISTICAL MODEL #2



$B = (O_{3,MAX} - O_{3,7})/O_{3,7}$

CONCLUSION

- **Statistical models could be used as an efficient tool in air pollution forecasting;**
- **When predicting daily maxima, one can expect reduction in the noise influencing the performance of the model;**
- **Unlike neural networks, proposed approach does not generate "black boxes";**
- **Ozone forecasts could perform better, if efficient procedures of censuring the data are developed**

THANK YOU VERY MUCH FOR YOUR ATTENTION!