

## Prediction of driving conditions in winter

Igor Grabec and Franc Švegl

Amanova Ltd., Technology Park, Ljubljana,  
& University of Ljubljana, Slovenia  
Email: [igor.grabec@amanova.si](mailto:igor.grabec@amanova.si) ; [franc.svegl@amanova.si](mailto:franc.svegl@amanova.si)

### ABSTRACT

Driving conditions on winter roads are mainly determined by slipperiness and visibility that can hardly be predicted in advance. Here a non-parametric regression is used to map data from weather and environmental conditions forecasts into variables describing driving conditions. The visibility is here characterized by the concentration of particulate matter PM10 while the slipperiness is introduced as suggested by the research project SRIS ([www.sris.nu](http://www.sris.nu)). There the estimation of slipperiness is performed by expert drivers based upon signals from ABS and ESP sensors in a car and personal feelings during driving. In order to clone performance of an expert driver we developed an intelligent computer program that learns from a statistical data basis obtained by past joint observations of environmental and driving conditions to estimate the driving conditions in a new situation. During operation the computer obtains new data about the environment and compares them with the corresponding ones in the data base. Based upon their similarity, the associated past data are accounted in the prediction of new driving conditions. Prediction is performed non-parametrically by the conditional average estimator. Performance of the corresponding method is characterized by the correlation coefficient  $r$  between predicted and really observed data. We demonstrate the performance of the method by using slipperiness data from high-ways in Sweden and data about PM10 in the Po valley in Italy. Relatively high values of correlation coefficient ( $r \sim 75\%$ ) indicate that the proposed method is applicable for prediction of hard winter driving conditions based upon weather forecasts.

**Keywords:** non-parametric modeling, prediction, driving conditions, slipperiness, PM10.

### 1. INTRODUCTION

Driving conditions on a road are influenced by weather and road surface states that are changing rather stochastically.[1,2,3] Consequently, we treat driving conditions as non-autonomous stochastic phenomena and describe their properties statistically by a general, non-parametric model.[3,4] Information for the creation of the model is extracted from joint records of variables describing driving conditions and environmental properties quantitatively. The fundamentals of the statistical non-parametric modeling are briefly described in the next section,[4] while the applicability of the method is demonstrated in the subsequent section by predicting variables representing slipperiness and concentration of particulate matter that influences visibility on a road surface. Based upon results of both examples the performance of the proposed method is quantitatively described by the correlation coefficient between predicted and original values of characteristic variables.[4] The final goal of our approach is to provide for a quantitative prediction of driving conditions from weather forecasts.[5] The predicted data could be utilized as a valuable information support to participants in road traffic and road services, especially in winter.

### 2. FUNDAMENTALS OF NON-PARAMETRIC MODELING

Our aim is to proceed to a general description of a relation between environmental and driving conditions on a particular point of a road. Both types of conditions are represented by vectors  $\mathbf{X}$ ,  $\mathbf{Y}$  which comprise components of the state vector  $\mathbf{Z}=(\mathbf{X},\mathbf{Y})$  that is further considered as a stochastic variable. Consequently, the description of the complete phenomenon requires utilization of the joint probability density function (PDF)  $f(\mathbf{Z})$ .[4,6-8] This function has to be created based upon experimental observations of the complete phenomenon. Let us suppose that a series of such observations has yielded  $N$  joint statistical samples  $\{\mathbf{Z}_n=(\mathbf{X},\mathbf{Y})_n; n = 1, \dots, N\}$  which we further consider as a statistical data base. The accuracy of experimental observation can be characterized by using the scattering function of the instruments comprising the traffic and weather observation system. It is usually determined during the system calibration. Most frequently the scattering of instrument output at a fixed

input corresponds to normal distribution that is described by the multivariate Gaussian function. Without essential loss of generality we further assume that the value of the standard deviation  $s$  is equal for all components of vector  $\mathbf{Z}$ . For the purpose of modeling it is convenient if the number of data  $N$  is selected so that the distance between samples  $\mathbf{Z}_n$  is approximately equal to the standard deviation  $s$ .

In our description we do want to avoid any a priori suppositions about the properties of the observed phenomenon as is usually done in a parametric statistical modeling of relations between stochastic variables. Consequently, we follow a non-parametric approach and first estimate the PDF by the following kernel estimator:[4,6,7]

$$f(\mathbf{Z}) = \frac{1}{N} \sum_{n=1}^N g(\mathbf{Z} - \mathbf{Z}_n) \quad (1)$$

in which we apply the scattering function  $g(\mathbf{Z} - \mathbf{Z}_n)$  as the kernel. The kernel describes the scattering of instrument output  $\mathbf{Z}$  during measurement of this variable at a fixed input with the mean value  $\mathbf{Z}_n$ . [6,7]

In order to proceed to our goal we pose the following question: "What would be the driving conditions  $\mathbf{Y}$  if the environmental conditions are given by  $\mathbf{X}$ ?" To answer it we suppose that a relation between both variables:

$$\mathbf{Y} = \mathbf{F}(\mathbf{X}) \quad (2)$$

could be estimated by using the PDF given in Eq. (1). The minimization of the mean square statistical error of estimation yields as the optimal statistical estimator or predictor the conditional average:

$$\mathbf{Y}_p = E[\mathbf{Y} | \mathbf{X}] \quad (3)$$

Here  $E[ | ]$  denotes the conditional mean value of variable  $\mathbf{Y}$  at the condition  $\mathbf{X}$ . If we calculate the conditional average by the PDF given in Eq. (1), we obtain the estimator expressed in terms of samples  $(\mathbf{Y}_n, \mathbf{X}_n)$  as:

$$\mathbf{Y}_p(\mathbf{X}) = \frac{\sum_{n=1}^N \mathbf{Y}_n g(\mathbf{X} - \mathbf{X}_n)}{\sum_{i=1}^N g(\mathbf{X} - \mathbf{X}_i)} = \sum_{n=1}^N \mathbf{Y}_n S(\mathbf{X} - \mathbf{X}_n) \quad (4)$$

Here we introduced the *measure of similarity* between  $\mathbf{X}$  and  $\mathbf{X}_n$  by the expression:

$$S(\mathbf{X} - \mathbf{X}_n) = \frac{g(\mathbf{X} - \mathbf{X}_n)}{\sum_{i=1}^N g(\mathbf{X} - \mathbf{X}_i)} \quad (5)$$

The more similar is the given condition  $\mathbf{X}$  to the  $n$ -th sample  $\mathbf{X}_n$ , the more the complement  $\mathbf{Y}_n$  associated to it contributes to the estimated value  $\mathbf{Y}_p$ . Therefore, the method of estimation specified by Eq.(4) corresponds to an associative recall of stored data from the data base, while the estimation of the function  $\mathbf{F}$  corresponds to learning from examples that is a basis of artificial intelligence.[4] Determination of function  $\mathbf{F}$  quite generally represents creation of new information about the observed phenomenon based upon measured data, and consequently, the estimator in Eq. (4) is called *general or non-parametric regression*. [4,6,7] Since it can be performed automatically, Eq. (4) provides a valuable tool that could be simply included into intelligent traffic systems.[3] In fact, it has already been utilized for modeling of very complex phenomena related to traffic, energy consumption and turbulent fields.[3-5,8-10] If we want to predict driving conditions  $\mathbf{Y}$  we just have to provide the data  $\mathbf{X}$  about environmental conditions and the data base containing joint samples obtained by previous observations of the same phenomenon. Execution of arithmetical operations indicated in Eq. (4) then yields the predicted value. The performance of the proposed estimation can be quantitatively described by the correlation coefficient  $r$  between predicted and measured values.[4] It is specified by the standard formula from statistics:

$$r = \frac{\text{Cov}(Y, Y_p)}{\sqrt{\text{Var}(Y) \text{Var}(Y_p)}} \quad (6)$$

in which  $\text{Var}(\dots)$ , and  $\text{Cov}(\dots)$  denote the variance, and covariance, respectively. In the case when multi-component variable is predicted, the correlation coefficient turns to a matrix.

For the purpose of performance estimation we usually split the data base into two portions and subsequently use the first one for modeling and the second one for testing. Such cross-validation method was applied also in cases presented in the next section.

### 3. EXAMPLES OF MODELING AND FORECASTING

#### 3.1 Prediction of road slipperiness

The surface slipperiness describes the most important road property during winter in Canada, Scandinavian and Alpine countries.[5,11] It is well known that living beings have a well developed sense for estimation of slipperiness and that slipperiness essentially depends on the past and present weather conditions. However, it is not easy to characterize this property technically since the friction coefficient is subject to large fluctuations when measured as a ratio of horizontal and vertical force acting on a moving vehicle. Therefore, the idea is to join the information about slipperiness estimated by a fleet of expert drivers driving in different conditions with the information provided by various kinds of sensors in vehicles and weather observation stations. Based upon joined information an intelligent system should be developed that could learn from the acquired data to predict slipperiness from weather data. Such projects started recently in various Scandinavian countries. In order to demonstrate applicability of conditional average estimator for this purpose we utilize here data published by Swedish Slippery Road Information System - SRIS ([www.sris.nu](http://www.sris.nu)).[11]

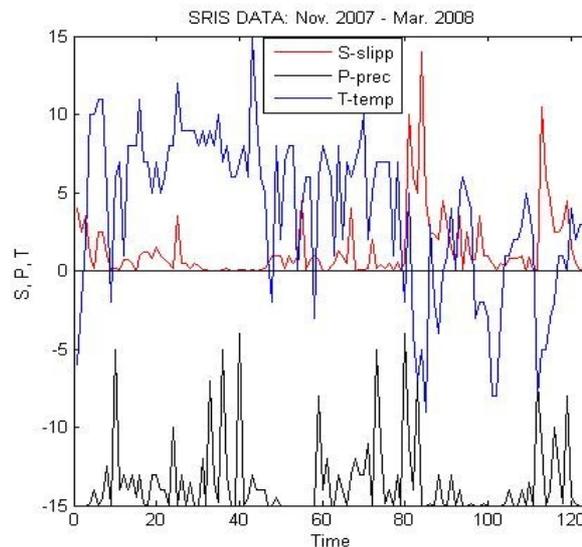


Figure 1. Record of joint weather and slipperiness data as published by [www.sris.nu](http://www.sris.nu).

Figure 1 shows a record of the following joint variables: temperature - T, precipitation - P, and slipperiness - S that was estimated by a fleet of expert drivers. Although these variables appear independent, their dependence is in fact hidden in the joint PDF. If we apply several components of past and present weather data together with the slipperiness to describe the state vector  $\mathbf{Z}$ , then the non-parametric modeling renders possible to predict the slipperiness from weather data and past records. Figure 2 shows records of predicted and actually observed slipperiness in dependence of time.

To characterize the accuracy of the slipperiness prediction we apply the correlation plot shown in Fig. 3. A point in this Figure is determined by the really measured value  $X$  and the corresponding predicted value  $X_p$ . The correlation coefficient of the predicted and actually observed slipperiness is calculated from all points in the graph and equals  $r = 0.77$ . It is instructive that this result could be essentially improved if the slipperiness variable is transformed non-linearly. For this purpose we first determine the maximal value of measured slipperiness  $X_m$  and arbitrary select its half value  $0.5 X_m$  as a critical level of slipperiness. Based on this value we then define a transformed slipperiness variable by a unit step function  $U$ :  $X_{tr} = U(X - 0.5 X_m)$ . Its value vanishes for slipperiness below the critical value and equals 1 for slipperiness above it; hence the transformed variable  $X_{tr}$  is applicable to roughly indicate critical conditions. The correlation coefficient of predicted and actually observed transformed variable  $X_{tr}$  is in then  $r \sim 1$ . This means, that a proper deterministic processing of data can significantly support and improve the statistical method.

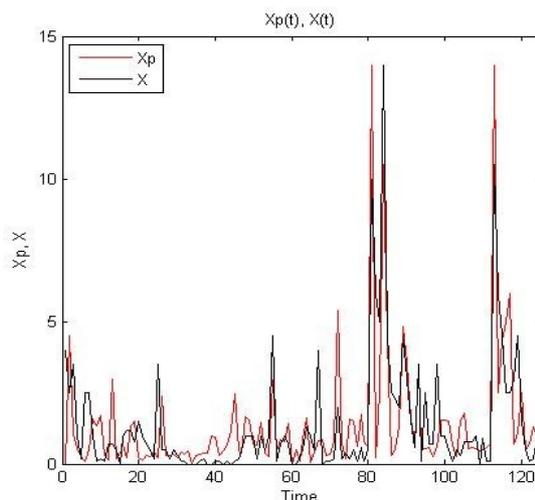


Figure 2. Slipperiness  $X$  as measured (black) or predicted (red) by the conditional average estimator from weather data.

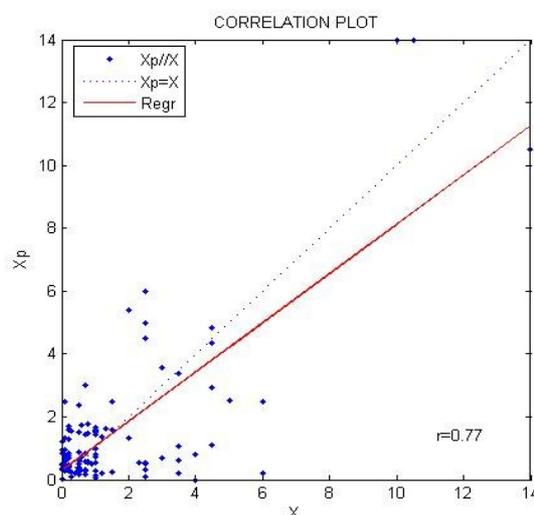


Figure 3. Correlation plot of measured and predicted slipperiness  $X$  from weather data.

It is important that forecasting of weather is well developed and that corresponding weather variables could be further utilized to forecast the slipperiness. Joining of traffic and weather data is needed for this purpose. This possibility is now a basis for development of an intelligent system for prediction of driving conditions.

The demonstrated example of slipperiness prediction indicates that non-parametric statistical modeling is applicable in the development of intelligent information processing systems that resemble properties of human experts. Such systems could substitute operators at various stations for traffic management and control. For this purpose the prediction method has to be included into more complex method of intelligent control.[4,12]

### 3.2 Prediction of PM10

Beside slipperiness the air pollution is very important for characterization of driving conditions.[5] It essentially depends on concentration of microscopic particles in the air which is denoted as PM10. This variable describes the concentration of solid particles having diameter in the range between 0 and 10  $\mu\text{m}$  and significantly influences visibility and development of fog. Our goal was to develop a method for prediction of  $P=PM10$  from given data about other environmental variables describing the average air temperature -  $T$ , humidity -  $H$ , and the average wind velocity -  $W$ . For the testing of the non-parametric modeling and forecasting we utilized the data base provided by ARPV - Centro Meteorologico, Teolo, Italy. It contains data obtained by measurements in the Po vally in Italy. A portion of the data base is presented in Fig. 4. From this data base records of the state vector components  $Z = (W,H,T,P)$  have been then extracted. They are shown in Figs. 5a,b,c,d. Here the time is next used as the sample index  $n$ .

Microsoft Excel - Arpv\_data

File Edit View Insert Format Tools Data Window Help

100% Arial 10

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		Percentage of wind calm <0.5 m/s	Average wind speed	Average wind speed (without values<0.5 m/s)	Average wind direction	Hmix rural	Hmix urban	Stanford Index	Average radiation	T minimum	Average T	Total precipitation	Equivalent potential temperature	Concentration PM10 Air Quality Station in Arcella	Concentration PM10 in Air Quality Station in Mandia	Concentration PM10 in Air Quality Station, Average Mandia and Arcella				
2	date	% wind calm	wv_m	wvc_m	dv_m	z1_r	z1_u	stan_u	rmed	tmim	tmed	ptot	thte	PM10_arc	PM10_man	PM10_tipo		aa	mm	99
3	1.1.2003	17	1	1.1	289,1	156	206	0,54	81	277	280	0	296	63	62	63,0		2003	1	1
4	2.1.2003	12	1	1.1	352,9	46	90	0,69	7	278	278	0,2	294	55	54	55,0		2003	1	2
5	3.1.2003	25	1		85,2	74	126	1,37	17	277	278	0	293	53	46	50,0		2003	1	3
6	4.1.2003	17	0,8	0,9	75,1	93	117	2,74	50	274	276	0,2	289	62	46	54,0		2003	1	4
7	5.1.2003	8	2,7	2,9	299,5	266	491	0,39	31	276	278	5,8	293	44	40	42,0		2003	1	5
8	6.1.2003	8	2,3	2,5	326	262	464	0,05	42	274	276	8	288	48	36	42,0		2003	1	6
9	7.1.2003	0	2,5	2,5	338,2	245	471	0	19	274	275	5	287	33	25	29,0		2003	1	7
10	8.1.2003	0	2,9	2,9	336,4	271	521	0	12	274	275	0,2	285		41	41,0		2003	1	8
11	9.1.2003	0	3	3	334,6	285	553	0,02	28	273	274		301		32	32,0		2003	1	9
12	10.1.2003	12	1,5	1,7	337,2	188	284	0	51	273	274	0	289					2003	1	10
13	11.1.2003	0	3,5	3,5	323,2	347	623	0,03	87	272	274	0	298					2003	1	11
14	12.1.2003	4	2,1	2,1	357,9	257	391	0,36	100	269	272	0	302					2003	1	12
15	13.1.2003	37	0,7		57,1	136	178	2,75	83	267	271	0	311		130	130,0		2003	1	13
16	14.1.2003	62	0,5		41	119	156	1,81	84	269	273	0,2	314		152	152,0		2003	1	14
17	15.1.2003	62	0,5		14,4	125	166	5,1	80	270	273	0,2	315		188	188,0		2003	1	15
18	16.1.2003	8	1	1,1	12	91	126	4,98	46	270	274	0,2	305		196	196,0		2003	1	16
19	17.1.2003	17	0,9	1,1	41,7	120	163	3,54	72	272	276	0	300		127	127,0		2003	1	17
20	18.1.2003	17	1,1	1,2	34,8	84	124	1,43	39	272	277	0	291	128	103	116,0		2003	1	18
21	19.1.2003	0	1,4	1,4	68,7	156	198	4,08	105	271	274	0,2	318	76	67	72,0		2003	1	19
22	20.1.2003	33	0,7		78,4				93	271	275	0,4	306	101	97	99,0		2003	1	20
23	21.1.2003	4	2,3	2,4	312,2				17	276	278	15	293	83	87	85,0		2003	1	21
24	22.1.2003	17	1,8	2,1	39,8				31	275	279	2	294	49	50	50,0		2003	1	22
25	23.1.2003	37	0,8		70	112	141	2,39	56	273	277	0,2	296	84	73	79,0		2003	1	23
26	24.1.2003	0	1,9	1,9	339,7	170	258	2,17	107	274	279	0	291	67	52	60,0		2003	1	24
27	25.1.2003	0	3,1	3,1	326,3	331	566	1,22	108	277	279	0,2	290	49	37	43,0		2003	1	25
28	26.1.2003	0	2,3	2,3	335,9	268	365	2,23	111	275	278	0	288	47	47	47,0		2003	1	26
29	27.1.2003	4	1,1	1,2	66,4	192	245	3,29	96	273	277	0	294		86	86,0		2003	1	27
30	28.1.2003	12	1,2	1,4	0,4	141	200	2,63	87	272	276	0	303	148	134	141,0		2003	1	28

Figure 4. Data base used in modeling and forecasting of PM10.

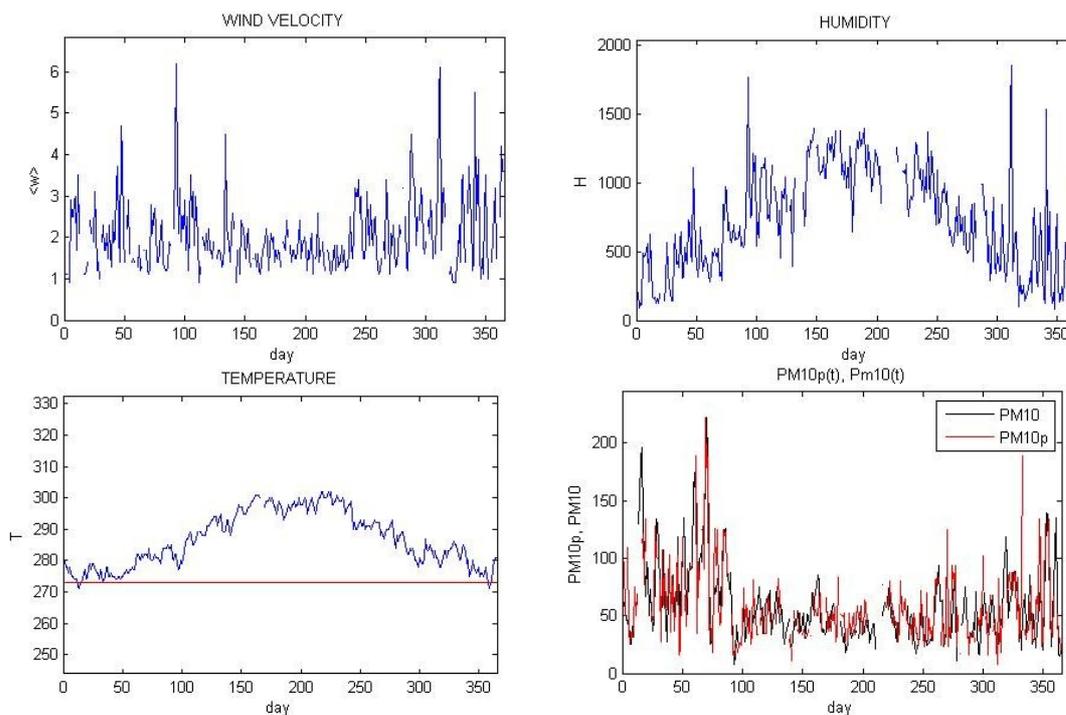


Figure 5. Records of variables used in modeling and prediction of PM10

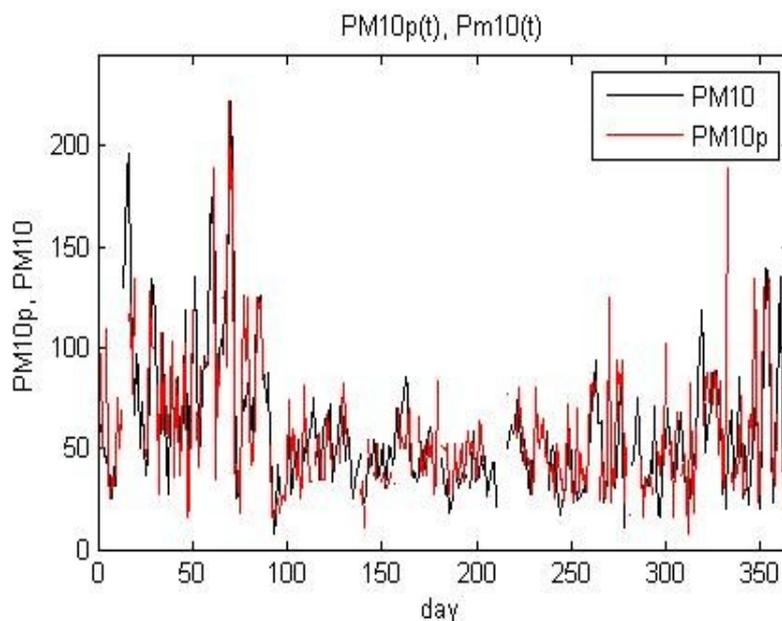


Figure 6. Records of predicted (red) and measured (black) PM10.

Based upon data from joint records shown in Fig. 5 we created statistical model of the relation  $P=F(W,H,T)$  as described by Eq. 4. By using the model we then predicted hidden concentration  $P$  from given data of  $W,H,T$ . The result of prediction is shown together with corresponding measured data by the records in Fig. 6.

Agreement between predicted and really measured data is described by the correlation coefficient  $r$  and shown in the correlation diagram in Fig. 7. Similarly as in the case of slipperiness also in this case the value of correlation coefficient  $r=0.74$  indicates an applicable prediction of variable under consideration.

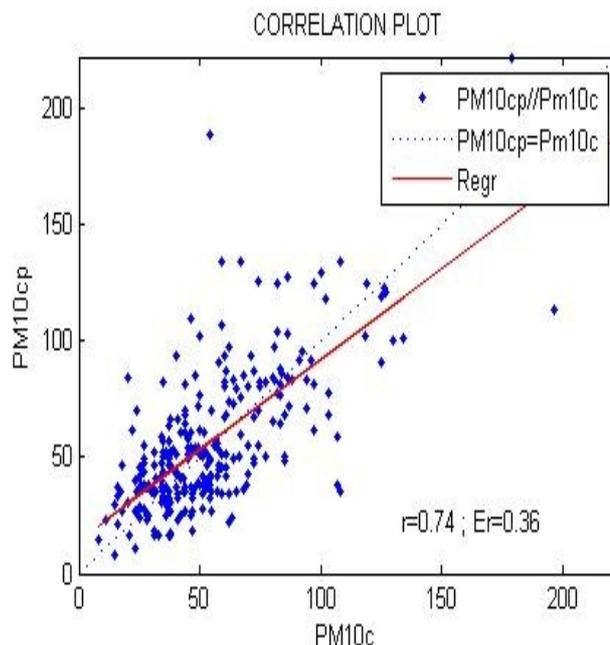


Figure 7. Correlation diagram of predicted and measured PM10. The agreement of both values is characterized by the linear regression line (red) while the diagonal line corresponds to an ideal agreement.

#### 4. CONCLUSIONS

In this article the focus is on a general statistical modeling and forecasting of winter driving conditions.[5,13] Consequently, non-parametric statistical approach is considered since it is completely based upon measured data.[4] In spite of the complexity of driving conditions on roads, the non-parametric statistical modeling of relations between characteristic variables by conditional average estimator renders possible their forecasting based on environmental data. The information generated by forecasting can be transmitted to participants in the traffic over existing communication devices such as mobile telephones or internet.[5] Beside this it is also applicable at the optimization of road service in winter.[13] An important advantage of the non-parametric statistical modeling is that the governing algorithm is formulated rather generally and can be executed practically automatically without essential modification at its adaptation to a specific case. It can be therefore easily used for ITS support.

Beside modeling of driving conditions the proposed method can be used to predict traffic flow distribution and related variables such as path integral of traffic activity and an optimal travelling time interval. It could also provide information about possibility of congestion development in a selected travelling time interval, etc.[5,14] Beside this, non-parametric approach renders possible a simple joining of weather and traffic flow data and related prediction of critical states. Related to the modeling of traffic flow is also renders possible modeling of pollution generated by the traffic and forecasting of corresponding critical states. In addition, non-parametric approach also renders possible consideration of traffic control variables in the modeling and thus provides also a basis for an intelligent control of traffic by ITS.[14]

#### 5. REFERENCES

- [1] D. Helbing, *Verkehrsdynamik*, Springer, Berlin, 1997.
- [2] B. S. Kerner, *The Physics of Traffic*, Springer, Berlin, 2004.
- [3] I. Grabec, K. Kalcher, F. Švegl, *Statistical Forecasting of Traffic Flow Rate*, Proc. of the 14th Int. Road Weather Conf. 2008<sup>®</sup>, Prague, CzR, May 14-16, 2008, ID 50, disc published by Wirelesscom, [www.sirwec.cz](http://www.sirwec.cz)
- [4] I. Grabec, W. Sachse, *Synergetics of Measurements, Prediction, and Control*, Springer, Berlin. 1997
- [5] <http://www.roadidea.eu/>; EU 7FP project ROADIDEA studies the innovation potential of the European ITS.
- [6] I. Grabec, *Experimental modeling of physical laws*, Eur. Phys. J. B, **22**, 129-135 (2001).
- [7] I. Grabec, *Extraction of physical laws from joint experimental data*, Eur. Phys. J. B, **48**, 279-286 (2005).
- [8] Several references of interest for non-parametric statistical modeling of complex physical laws are available at web page folder of I. Grabec: <http://www.fs.uni-lj.si>
- [9] B. Lunar, I. Grabec. *Forecasting Electrical Power Consumption by Normalised Radial Basis Function Neural Network*, Neural Network World, **12(3)**, 241-254 (2002)
- [10] P. Potočnik, M. Thaler, E. Govekar, I. Grabec, and A. Poredoš, *Forecasting risks of natural gas consumption in Slovenia*, Energy policy, **35(8)**:4271-4282. (2007)
- [11] Swedish Slippery Road Information System – SRIS: <http://www.sris.nu>
- [12] *Handbook of Intelligent Control*, Neural, Fuzzy and Adaptive Approaches, Eds.: D. A. White, D. A. Sofge, Van Nostrand Reinhold, New York 1992
- [13] I. Grabec, F. Švegl, *Prediction of traffic and driving conditions in winter*, Professional meeting of winter roads services, Bled, 20-22.10.2009, contribution and article in Slovenian.
- [14] I. Grabec, K. Kalcher, F. Švegl, *Modeling and Forecasting of Traffic Flow*, Nonlinear Phenomena in Complex Systems, Sept. 2009, invited article submitted.

#### ACKNOWLEDGMENT

Presented research was supported by EU 7 FP project “Roadidea”, iCAR - EU Thematic Network and Slovenian Agency for Roads. Data used in modeling were provided by collaborating companies Semcon and Klimator in Göteborg and ARPV Centro Meteorologico, Teolo, Italy.

### Biographical notes of authors

**Igor Grabec** finished study of physics at the University of Ljubljana (UL) with PhD thesis on unstable and turbulent ionization waves in weakly ionized plasma. He became professor of physics at the Faculty of Mechanical Engineering, UL where he explored acoustic emission in stressed materials and manufacturing processes. His research objective was an statistical modeling of turning, drilling, grinding, cutting and laser assisted manufacturing processes as well as at the development of intelligent analyzers of acoustic emission for non-destructive testing. His bibliography contains about 450 articles, 17 patents and four books. After retirement in 2007 he became director of a high tech company Amanova that deals with forecasting of traffic flows. He is Professor Emeritus of UL and a member of “Academy of Sciences and Arts” and "Engineering Academy” of Slovenia, Fellow of International Academy for Production Engineering – CIRP, and Honorary Member of World Innovation Foundation.

**Dr. Franc Švegl** graduated from physical chemistry at University of Ljubljana in 1990. He continued Ph.D. studies in The Laboratory for Spectroscopy of Materials at The National Institute of Chemistry in Ljubljana in the field of sol-gel science and spectroscopic characterization. After Ph.D he continued PostDoc studies at Karl Franzens University, Graz, Austria, as a Marie Curie fellow, where his main scientific work was in the field of electrochemical and optical sensors and development of in-situ spectro-electro-chemical measuring techniques. After returning from PostDoc he was leading for ten years The Laboratory for Mineral Binders and Mortars at The Slovenian National Building and Civil Engineering Institute. Currently he is a member of Amanova Ltd research team dealing with development of an information processing system capable of automatic measurement, modeling, prediction and control of complex processes or material properties that cannot be well described by analytical approach.