Air Distribution in Rooms, (ROOMVENT 2000)
Editor: H.B. Awbi

STATISTICAL PREDICTION MODEL OF THE OUTDOOR/INDOOR AIR POLLUTANT TRANSFER

V. Iordache, C. Ghiaus, P. Blondeau, F. Allard

LEPTAB (Laboratoire d'Etude des Phénomènes de Transfert Appliqués au Bâtiment) – Université de La Rochelle, France

ABSTRACT

Efficient control of ventilation systems needs information on indoor air pollutant concentration. But most of the time, the pollutant concentration is not measured. However, outdoor air pollution forecast models are becoming operational and a relation between outdoor and indoor pollutant concentration may be used to predict pollutant concentration peaks and to infer recommendations to reduce its value. A prediction model of the outdoor/indoor air pollutant transfer, based on experimental data, was obtained using multiple linear regression. The measurements were achieved in a primary school classroom in the coastal region of La Rochelle, France. The measured parameters are the outdoor and indoor concentrations, the indoor temperature and humidity, window openings, number of photocopies and the number of pupils in the classroom. Models for NO, NO_2 and O_3 outdoor/indoor relation are achieved and tested.

KEYWORDS

Statistical model, Pollutant concentration, Indoor air, Outdoor/Indoor transfer

INTRODUCTION

Most of the ventilation standards consider the outdoor air as clean air, even if it has already been proved that it contains numerous pollutants: O_3, NO, NO_2, SO_2, particles, Pb and others. Studies have proved that people spend almost 90% of their time indoors, therefore the great interest in knowing the concentration of indoor air pollutants. Three major factors influence the indoor air pollutant concentration: 1) wall covering materials (Tiffonnet 2000), 2) indoor activities and occupation, 3) and the outdoor/indoor pollutant transfer. Hereafter we analyze the suitability of the statistical modeling in the evaluation of the third major factor (outdoor/indoor pollutant transfer) and the prediction of the dynamic variation of the indoor pollutant concentration. This model may be used in conjunction with a pollutant concentration forecast (Ghiaus and Allard, 1999) in order to predict the level of indoor air pollutant concentration. Two methods, physical and statistical modeling, are employed to predict the outdoor/indoor air concentration relation. Physical modeling is based on the transport equations of the pollutant into the room (Shair et al. 1974, Blondeau, 1996, Sohn et al. 1998). Statistical modeling derives a mapping between the output and input parameters based on experimental data rather than on physical phenomena. Multiple linear regression models were already used to determine the importance on indoor air pollutant concentration of indoor characteristics: chalkboard, fabrics materials, open shelves, climatic conditions, dwelling characteristics (Smedje et al. 1999, Bartlett et al. 1999).

EXPERIMENTAL STUDY

Children are the most sensible to pollution; consequently, most of the studies are oriented towards the protection of the children. Our measurements were achieved in a classroom of a primary school in La Rochelle, France, between 28 April 1999 and 13 May 1999. The dimensions of the classroom are: 5m x 4m x 3m and the measurement point is in the middle of the classroom, between the window and the door, at 0.4m below the ceiling. The copying machine is placed in an adjacent room, and not expected to have substantial influence upon the classroom O_3 concentration (Figure 1).

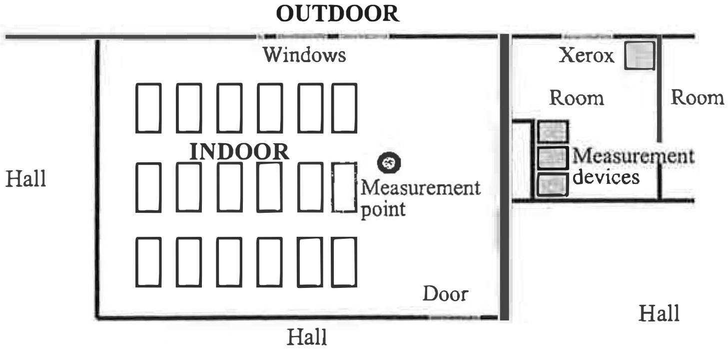

Figure 1: Classroom measurement devices location.

Outdoor and indoor NO, NO_2 and O_3 concentrations, and indoor air temperature and humidity were sampled every 15 minutes for two weeks. The window opening duration, the number of pupils in the classroom and the number of photocopies were also recorded. The indoor air NO concentration oscillated between 1 and 2 $\mu g/m^3$, so we could not infer any conclusion regarding the indoor/outdoor pollutant concentration relation.

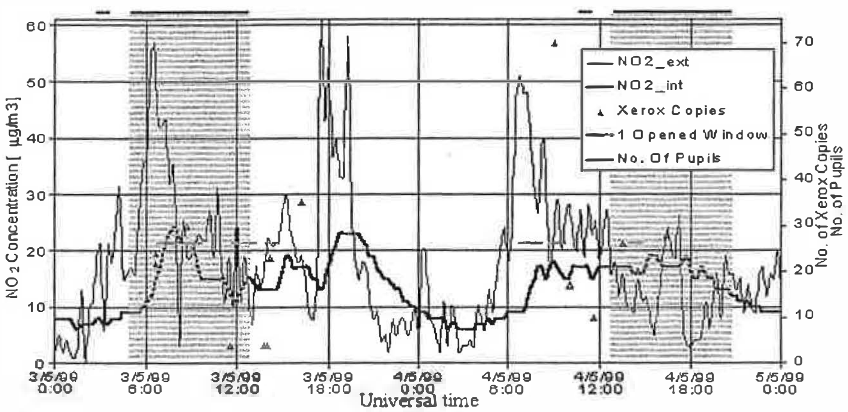

Figure 2: Variations of the outdoor and indoor NO_2 concentrations, number of pupils and photocopies.

By the contrary, the outdoor and indoor NO_2 and O_3 concentrations (Figure 2) permitted to notice the existence of a time lag between the outdoor and indoor air concentration: on 3 May, 1999, to outdoor NO_2 concentration drop at 17:00H corresponds the indoor NO_2 drop at 17:30H. During window openings, the indoor NO_2 concentration approaches the outdoor concentration (10:00–15:00H on 3/5/99, 10:00–15:00H on 4/5/99, dashed areas in Figure 2). The NO_2 outdoor concentration peaks shorter than 30 minutes do not influence the indoor concentration (19:15H on 3/5/99, 00:30H on 4/5/99).

BASIC MODEL CONSTRUCTION

Previous studies based on physical laws (Shair, 1974; Iordache, 1999) resulted in linear models of the indoor air pollutant concentration as a function of the outdoor one. Consequently, a multiple linear regression model was used to determine the indoor-outdoor relation: the present indoor air concentration C_i^{τ} $(\mu g/m^3)$, as a function of the input data: the previous indoor air pollutant concentrations $C_i^{\tau-\Delta\tau}$ $(\mu g/m^3)$, present and previous outdoor air concentrations C_o^{τ}, $C_o^{\tau-\Delta\tau}$ $(\mu g/m^3)$, present and previous values of the temperature T^{τ}, $T^{\tau-\Delta\tau}$ (°C), relative humidity H^{τ}, $H^{\tau-\Delta\tau}$ (%), and events as: window openings W^{τ}, $W^{\tau-\Delta\tau}$ (*binary value*), number of pupils in the classroom O^{τ}, $O^{\tau-\Delta\tau}$ (-), and number of photocopies X^{τ}, $X^{\tau-\Delta\tau}$ (-). Using this notation the model has the form:

$$C_i^{\tau} = a_o + a_1 C_i^{\tau-\Delta\tau} + a_2 C_o^{\tau} + a_3 C_o^{\tau-\Delta\tau} + a_4 T^{\tau} + a_5 T^{\tau-\Delta\tau} + a_6 H^{\tau} + a_7 H^{\tau-\Delta\tau} + a_8 W^{\tau} + a_9 W^{\tau-\Delta\tau} + a_{10} O^{\tau} + a_{11} O^{\tau-\Delta\tau} + a_{12} X^{\tau} + a_{13} X^{\tau-\Delta\tau}. \quad (1)$$

The input variables should have not only the present and the last previous value, but also the values at all time steps that could significantly influence the present value of the indoor air pollutant concentration. The number of significantly important time lags may be estimated using the cross-correlation coefficients (Figure 3). Cross-correlation coefficients, $c_{xy}(m)$, between the time series x and y, are given by (Bendat et al. 1971, Seber et al. 1989):

$$c_{xy}(m) = \sum_{n=0}^{N-|m|-1} x_n y_{n+m} \qquad m \geq 0. \quad (2)$$

where N is the length of the two series and m is the time lag.

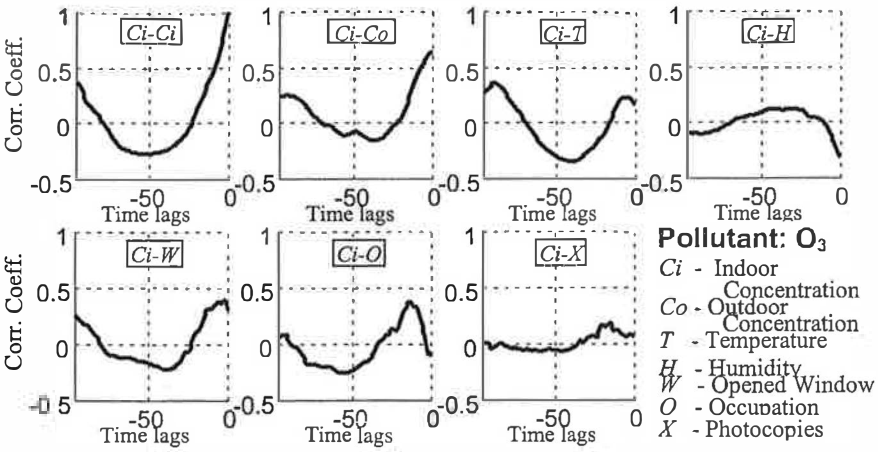

Figure 3: Cross - Correlation Coefficients for O_3

The number of important time lags is considered for the significant correlation coefficients. We note in Figure 3 that both values of the indoor and outdoor air pollutant concentration closer to the current moment have a significant influence upon its present value (Ci-Ci and Ci-Co graphs). By the contrary, neither the indoor air temperature and humidity, nor the number of photocopies influence the present value of the indoor air concentration.

In the basic model, we consider as input data only one time lag value for each parameter. The model is determined by its coefficients $a_{j\ (j\ =\ 0 \ldots n)}$. The standard modeling procedure is to find the regression coefficients using a set of data and to test the model on another time interval. The basic model is constructed using 100 time steps (96 time steps = 1 day) and the model is verified for all the measurement period (1248 time steps). The indoor air pollutant concentration calculated with the statistical model is a good approximation for the measured indoor concentration (Figure 4). The statistically obtained results (dashed curve) are close to the measured concentrations (solid curve), although their mean is lower.

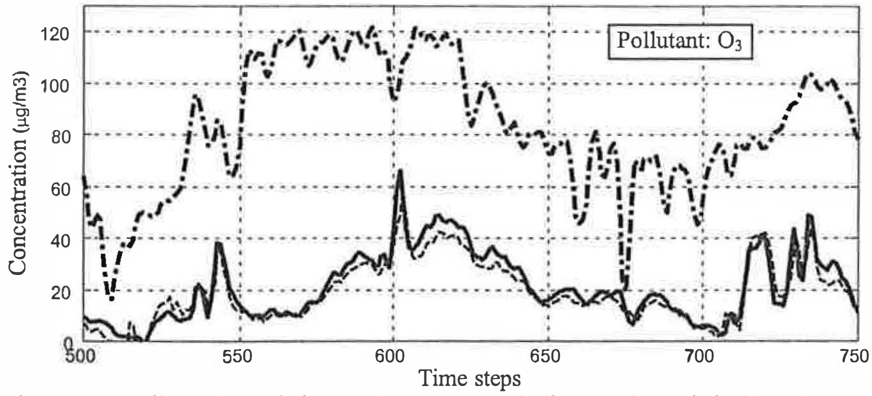

Figure 4: Detail of the variation of the measured (solid curve), statistically calculated (dashed curve) indoor, and outdoor (dot-dashed curve) air O_3 concentration.

The model (eqn. 1) for O_3 becomes:

$$\begin{aligned} C_i^{\tau} = {} & 51.4722 + 0.7764\, C_i^{\tau-\Delta\tau} + 0.1623\, C_o^{\tau} - 0.1054\, C_o^{\tau-\Delta\tau} - 2.5744\, T^{\tau} + \\ & 1.0313\, T^{\tau-\Delta\tau} - 0.6713\, H^{\tau} + 0.2881\, H^{\tau-\Delta\tau} + 2.5822\, W^{\tau} + 7.1900\, W^{\tau-\Delta\tau} - \\ & 0.0570\, O^{\tau} + 0.0764\, O^{\tau-\Delta\tau} - 0.0448\, X^{\tau} + 0.0122\, X^{\tau-\Delta\tau}. \end{aligned} \tag{3}$$

Similarly, other two models have been obtained for NO and NO_2.

The model is considered to be valid if its error are normal and independent, with mean zero (Myers, 1990). Thus, the conditions that the error should verify are: its mean value should not be significantly different to zero, the error should not be auto-correlated, and it must have a normal distribution. The basic model is valid since the error mean value is -1.36 μg/m^3, standard deviation is 3.47 μg/m^3, auto-correlation coefficients are close to 0, and its distribution approaches a normal distribution (Figure 5).

We note from the graphic of the cross-correlation coefficients of the error, that the errors closer in time have a greater influence on the present error than the further ones.

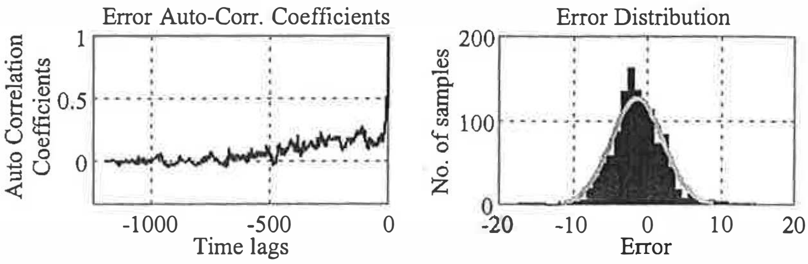

Figure 5: Validity test of the basic statistical model.

IMPROVED MODELS

In order to improve the accuracy of our model, we modified two major factors: the number of points (time steps) taken into account for the multi-linear regression model, and the number of time steps for each input parameter considered in the model. We developed two models for each pollutant. Both models are constructed using 1000 time steps instead of 100 in the basic model. The first improved model has as input data only the last value of each input parameter ($n_{ci} = n_{co} = n_t = n_h = n_x = n_o = n_w = 1$). In the case of NO_2, the error mean and the standard deviation diminish 40 times and 3 times respectively, as compared to the basic model. The second improved model uses as input data the last 8 values for each input parameter ($n_{ci} = n_{co} = n_t = n_h = n_x = n_o = n_w = 8$). This modification of the second major factor brings the error mean and the standard deviation 10 times closer to 0 than their values obtained with the first developed model (Table 1)

TABLE 1
ERROR ANALYSIS

Pollutant		n_{ci}	n_{co}	n_t	n_h	n_x	n_o	n_w	Error : ε			
									min	*max*	*mean*	σ_ε
NO	I	1	1	1	1	1	1	1	-0.994	0.988	-8.1e-4	0.202
	II	8	8	8	8	8	8	8	-0.093	0.087	8.2e-5	0.021
NO_2	I	1	1	1	1	1	1	1	-9.168	7.435	0.033	1.189
	II	8	8	8	8	8	8	8	-0.750	0.651	-0.003	0.117
O_3	I	1	1	1	1	1	1	1	-15.881	13.985	0.032	2.765
	II	8	8	8	8	8	8	8	-1.798	1.438	0.004	0.302

CONCLUSIONS

The prediction of indoor air pollutant concentration was achieved with a multi-linear regressive model. Its accuracy depends on the number of the time lags of input data and on the number of time steps used to determine the model coefficients: more time lags and time steps, better the model. Among the seven parameters, the most influential one is the outdoor air concentration due to outdoor/indoor pollutant transfer. This preliminary study proves that statistical analysis could lead to precise prediction of the indoor air pollutant concentration due to the outdoor air concentration. The perspective is to reduce the complexity of the models obtained for the three pollutants. We will also try to generalize the models as a function of room and building destination.

ACKNOWLEDGEMENTS

The study was achieved with the support of the municipality of La Rochelle, France. We appreciate the collaboration with the school Vauguin. The indoor and outdoor pollutant concentration measurements were accomplished by AREQUA (Regional Association for Air Quality in Poitou-Charantes).

REFERENCE

Barlett K.H., Kennedy S.M., Brauer M., Dill B. and van Netten C. (1999) Predictors of exposure to indoor CO_2 and bioaerosols in elementary school classrooms, Indoor Air Proceedings '99.

Bendat J.S., Piersol A.G. (1971). Random Data: Analysis and Measurement Procedures, New York: John Wiley & Sons, 332.

Blondeau P. (1996). Contribution a l'evaluation de la qualite globale des ambiances habitee – Role de la ventilation en periode estivale, thèse de doctorat, University of La Rochelle, France.

Ghiaus C., Allard F. (2000) Linear discriminant analysis applied to forecast ozone concentration classes in sea-breeze regime, Atmospheric Environment (Submitted)

Iordache V. (1999) Prediction des concentrations interieures en polluant: modele theorique et statistique, DEA Genie Civil, Physique batiment, Stage recherche, University of La Rochelle, France, III 1/43 – III 43/43.

Myers R.H. (1990). Classical and Modem Regression with applications, Duxbury Press, Blemont California, 83-94.

Seber G.A.F., Wild C.J. (1989). Nonlinear regression, John Wiley & Sons, 271-325.

Shair F.H., Heitner K.L. (1974). Theoretical model for relating indoor pollutant concentrations to those outside, Env. Sci. And Tech. 8:5.

Smedje G., Norback D. (1999). Factors affecting the concentration of pollutants in school buildings, Indoor Air Proceedings '99.

Sohn M.D., Small M.J. (1999) Parameter estimation of unknown air exchange rates and effective mixing volumes from tracer gas measurements for complex multi-zone indoor air models, Building and Environment **34:3**, 293-303.

Tiffonnet A.L. (2000) Contributions à l'analyse de la qualité de l'air interieur: influence des transports de composés organiques volatils (COV) entre les parois et l'ambiance, thèse de doctorat, LEPTAB, Université de La Rochelle, France