# A short term prediction model for surface ozone at southwest part of Mexico valley

## J. L. BRAVO, M. T. DIAZ

Instituto de Geofísica, UNAM, Circuito Exterior, CU, 04510, México, D. F., México

#### C. GAY

Centro de Ciencias de la Atmósfera, Circuito Exterior, CU, 04510, México, D. F., México

## J. FAJARDO

Instituto Nacional de Pediatría, Av. Iman #1, 04530, México, D. F., México.
(Manuscript received April 28, 1995; accepted in final form Aug. 12, 1995)

#### RESUMEN

La Ciudad de México está situada en un valle en una zona tropical con elevadas cantidades de radiación solar, tiene poca ventilación y una gran actividad industrial y de transporte, en consecuencia, presenta problemas graves de contaminación atmosférica. Se ha escogido al ozono como representativo de los oxidantes atmosféricos y de la contaminación ambiental. Se propone en este trabajo un modelo multiplicativo y se ajusta a 3 años de observaciones diarias para la estación Pedregal de San Angel, situada al Suroeste del Valle de México, se evalúa la importancia de algunos parámetros de uso general en las estaciones meteorológicas. El modelo explica 50% de la variabilidad diaria empleando la radiación solar y la concentración de ozono del día anterior y el 61% con la totalidad de las variables usadas. El modelo podría ser útil para predecir la concentración de ozono con el auxilio de una predicción de la radiación solar y emplearse en alertas ambientales.

#### ABSTRACT

Mexico City is located in a valley in the tropical zone with high intensity of total solar radiation; it has poor wind circulation and a great deal of industrial and transportation activity, as a consequence of this it has a serious problem with air pollution. Ozone is representative of total atmospheric oxidants and of air pollution. In this work a multiplicative model is proposed and adjusted to 3 years of daily observations at the Pedregal de San Angel Station, located at the Southwest part of Mexico's Valley. The importance of some common meteorological parameters in the explanation of daily variance is evaluated. 50% of the total variance is explained with total solar radiation and the previous day ozone concentration and 61% using all variables. This model could be useful in the prediction of ozone concentrations with help of a model to predict solar radiation and could be used in the establishment of criteria for environmental alerts.

## 1. Introduction

Ozone is an oxidant gas that produces irritation of the respiratory airways of laboratory animals and humans, therefore its occurrence in the atmosphere represents an important risk for health (Campos et al., 1992; Romero et al., 1994). Likewise, this gas can damage the vegetation, and react with several compounds, like plastics and rubber, debilitating them (Miller et al., 1994). Then ozone is an undesirable atmospheric gas at the surface level. Due to its high concentration in urban atmospheres and to the relatively easiness to measure it, ozone has been considered representative of oxidant gases (Bravo et al., 1991a, 1991b).

Substantial effort has been dedicated to modeling ozone in polluted atmospheres. Ozone takes part in complex photochemical and transport processes and is occurring in high concentrations in areas with industrial air pollution and abundant solar radiation, downwind of emission areas, or in places with poor wind circulation (Seinfeld, 1989, Chock and Heuss, 1987). For several years, Mexico City has had a very important problem of atmospheric pollution. Several workers have shown the general pattern of production and transportation of ozone (Miller et al., 1994, Riveros et al., 1993, Quadri and Sánchez, 1992, Bravo et al., 1991a, 1991b, 1988a, Jáuregui 1988a, 1988b). In this work a multiplicative linearized model to quantify the influence of the main factors identified as sources of ozone variation at the southwestern corner of the Mexico Basin is applied. Quantification is made by using a multivariate linear regression with the logarithm of the surface ozone concentration as the dependent variable, and total solar radiation flux, wind speed, wind direction, the previous day ozone concentration (autoregressive process), a trend term, and a set of seven "dummy" variables to quantify the effect of the day of the week. These parameters are measured in an urban climatological station of the Mexico Meteorological Service. A little correlation structure was found in residual analysis after model fitting.

# 2. Area of study and data

Mexico City is located in the Mexico Basin, an inland valley at highlands of central Mexico (19 20' N, 99 11', with an altitude of 2268 masl), that is surrounded by high mountains to the East, South and West, with poor air circulation. During the dry season anticiclone conditions prevail resulting in frequent clear skies and abundant insolation. The urban area has a surface of about 800 km. In the northern part there is a big industrial zone; in the central part, a great commercial, office, and habitational zone; the central - south area is mainly habitational, and the southern part of the metropolitan area is habitational and rural. This distribution of activities causes a very heavy transportation travelling around the valley, burning gasoline in cars, and diesel and gasoline in public transportation.

Consumption of fuel causes emission of nitrogen oxides. Due to the reduction of oxygen available for combustion by motors at Mexico City altitude, there is an increased emission of non-methane hydrocarbons (Nickerson et al., 1992). These compounds react photochemically in the presence of solar radiation producing ozone that is accumulated in the atmosphere at the surface level and dispersed to the middle troposphere.

Another source of ozone variability is the wind transportation and the presence of the "heat island" as has been discussed by Jáuregui (1988a). It has been shown that ozone precursors, that are emitted in the north industrial part of the city, are transported to the southwest by the dominant wind circulation, producing high ozone concentration at the southwest of Mexico Basin (Bravo et al., 1991a, 1991b, 1988a).

Results of in situ measurements of suspended particles, sulfur dioxide and ozone (Nickerson et al., 1992) show the presence of layers of particles and ozone during the first hours of the day

in the low troposphere; these layers are result of pollution accumulation from the previous day. Such accumulation has been demonstrated by the detection of important ozone concentrations during the first hours of the night, when there is no light to activate photochemical processes; these concentrations are measured when atmospheric conditions allow the descend of upper air.

Ozone data from the Pedregal de San Angel monitoring station were used in this work. This station is located on the southwstern suburbs of Mexico City and is part of a network of automated monitoring stations set up by the Government. The studied zone is residential and surrounded by green areas; however, there are very important avenues with heavy traffic near the station.

634 days were selected from the period 1989 - 1991. The following parameters were available:
1) mean hourly surface ozone between 12 and 13 hours (local time, usually the time of daily maximum ozone concentration); 2) average wind speed between 6 and 12 hrs.; 3) prevailing wind direction between 12 and 13 hrs.; 4) hourly average ozone concentration between 12 and 13 hours of the previous day, and 5) total solar radiation flux between 6 and 12 hours.

Solar Radiation was measured at the Solar Radiation Observatory in the Campus of the Universidad Nacional Autónoma de México (UNAM), located 3 Km to NE from the air quality monitoring station. Measurements were made with a Kipp and Zonen piranometer intercompared periodically with international standards.

## 3. The Model

Figure 1 shows the behavior, with respect to time, of the hourly mean of ozone between 12 and 13 hrs. Figure 2 shows the corresponding histogram. The graphs illustrate a no-uniform variance, an increasing trend of the ozone concentration, a probability density function skewed to the right and no significant seasonal variation. This suggests a transformation of ozone concentration to stabilize the variance and a linear term in time to consider the trend. Natural logarithm of ozone concentration was used to stabilize the variance (Holland and Fritz - Simons, 1982), giving residuals with a normal distribution, as it is discussed later.

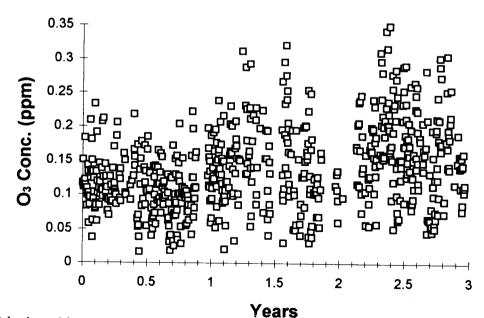


Fig. 1. Behavior, with respect to time, of the hourly mean of ozone between 12 and 13 hrs for 634 days during 1989 - 1991 period. Ozone concentration expressed in parts per million (ppm).

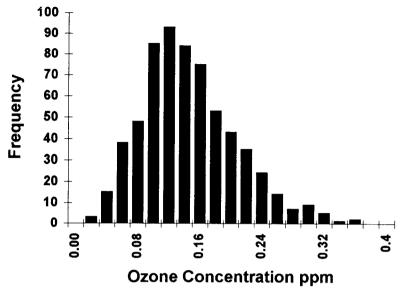


Fig. 2. Histogram of 634 ozono mean concentrations between 12 and 13 hrs during 1989 - 1991 period, expressed in parts per million (ppm).

A linear model is proposed using the natural logarithm of mean ozone concentration between 12 and 13 hrs as the dependent variable, and the following factors as independent variables:

The total solar radiation flux between 6 and 12 hrs has been chosen since radiation is the source of energy for the photochemical processes that produce ozone. Figure 3 shows a scatter diagram of ozone concentration versus solar radiation flux. The trend of the points in the graph indicates the dependence of the two variables, and it is clear that a second order term is necessary to reproduce the behavior properly.

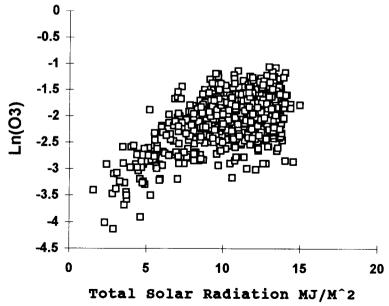


Fig. 3. Scatter diagram of ozone concentration versus solar radiation flux between 6 and 12 hrs. Solar radiation is expressed in Mega-joules/square meter.

It is known that meteorological conditions tend to remain the same for several days; this means that when conditions favorable for the formation and accumulation of ozone occur, they tend to stay for several days. This fact makes autoregressive the series of meteorological parameters and contaminants. On the other hand, permanence during the night of ozone and/or its precursors has been reported (Nickerson et al., 1992, Bravo et al., 1988b); this accumulated ozone (or its precursors) is a remanent of the previous day pollution. Then, at the beginning of the day, when solar heating causes air convection, remaining previous day ozone and/or precursors will contribute to increase the ozone concentration next day. To take into account this mechanism, hourly mean ozone concentration between 12 and 13 hours for the previous day is included in the regression formula.

Another recognized factor producing ozone variability is the transport of precursor gases (Jáuregui, 1988a); consequently wind direction is included in the model. To obtain a positive regression coefficient, the coordinate system is oriented in the direction of minimum ozone concentration, in this case the south. In other words we applied a 180° axis rotation. Figure 4 shows the scatter diagram of ozone concentration vs wind direction rotated as explained.

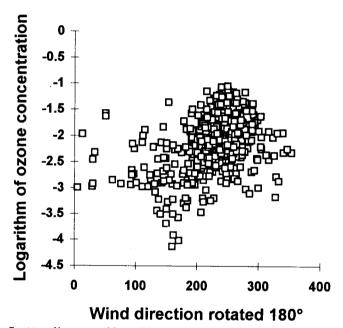


Fig. 4. Scatter diagram of logarithm of ozone concentration vs wind direction rotated 180°.

Another important phenomenon for the accumulation of precursor gases and for the ozone formation is the atmospheric stability. Conditions of atmospheric stability are favored by frequent thermal inversions (Jáuregui, 1988b). During these conditions wind speed is minimum. In the model the magnitude of wind speed was considered as a surrogate variable of atmospheric stability.

As it was mentioned earlier, data showed an increasing trend during the period of study, hence a linear term in time was considered.

Finally, to evaluate the effect of the day of the week in the variability of ozone concentration, seven "dummy" terms were included, corresponding to each day of the week; i.e. the "dummy" variable corresponding to Sunday has value 1 if the day considered is Sunday or 0 if it is any

38 J. L. BRAVO et al.

other day, the second is 1 if the day is Monday and 0 if it is not. Values are assigned in the same way for the rest of the days of the week.

Table 1. Correlation matrix of the variables in the model: LOG(O3): logarithm of ozone concentration. GR: total solar radiation. O3R: ozone concentration from the previous day. WVEL: wind speed. WDIR: wind direction rotated 180°.

	Ln (O3)	GR (	gr²	ОзВ	WVEL	WDIR	DATE
Ln (03)	1.0	0.563	0.505	0.462	-0.246	0.383	0.285
GR	0.563	1.0	0.985	0.258	-0.0012	0.315	-0.054
GR <sup>2</sup>	0.505	0.985	1.0	0.260	0.021	0.277	-0.040
ОзR	0.462	0.258	0.260	1.0	-0.103	0.159	0.337
WVEL	-0.246	-0.0012	0.021	-0.103	1.0	-0.173	-0.230
WDIR	0.383	0.315	0.277	0.159	-0.173	1.0	0.053
DATE	0.285	-0.054 -	0.040	0.337	-0.230	0.053	1.0

Table 1 shows the correlation matrix. The variables with the largest correlation coefficient with ozone concentration are the solar radiation flux between 6 and 12 hours, the previous day ozone concentration, and wind direction. It is noteworthy that the correlation between wind direction and solar radiation flux is similar in value to the correlation between wind direction and ozone concentration. The fitting was made using the least squares method obtaining the coefficients shown in the following formula:

$$LOG_e(O_3) = 0.369 \; GR - 0.015 \; GR^2 + 1.981O_3R - 0.0087WVEL$$
  $+0.0010WDIR + 0.177DATE - 4.816 \; SUN - 4.670 \; MON$   $-4.628 \; TUE - 4.632 \; WED - 4.714 \; THU - 4.572 \; FRI$   $-4.727SAT$ 

(1)

Where variables have the following meaning:

- $LOG_e(0_3)$  is the natural logarithm of the mean ozone concentration between 12 and 13 hrs expressed in parts per million.
- GR is the total solar radiation flux expressed in megajoules/square meter.
- $O_3R$  is the hourly mean ozone concentration between 12 and 13 hrs for the previous day expressed in parts per million.
- WVEL is the sum of the hourly means of the magnitude of wind speed expressed in m/sec.
- WDIR is the mean of wind direction between 12 and 13 hrs. The direction of the minimum ozone concentration (south) coincides with 0 direction.
- DATE is the time expressed in fractions of year. -
- SUN, MON, TUE, WED, THU, FRI, SAT are dummy variables. They are 1 if coincide with the corresponding day of the week, and zero otherwise. These variables are playing the role of the independent term in the equation, and allow the use of all observations in the determination of the set of regression coefficients using only one fitting.

The fitting process was made by steps. The following is the order in which the variables were introduced in the regression equation.

- Total solar radiation flux.
- Ozone concentration of the previous day.
- Squared total solar radiation flux.
- Trend.
- Wind speed.
- Rotated wind direction.

Variables are required to be independent to evaluate the regression coefficients. In this case the seven dummy variables are not linearly independent. To solve this problem the independent term was eliminated in the regression equation. Then, each of the dummy variables plays the role of the mentioned term, one for each day of the week. Coefficient values and their standard errors are shown in Table 2.

Table 2. Values of the regression coefficients for the model, mean values of logarithm of surface ozone concentration between 12 and 13 hrs. and their standard errors: GR, total solar radiation. O3R, ozone concentration for the previous day. WVEL, wind speed. WDIR, wind direction rotated 180°.

VARIABLE	GR	GR <sup>2</sup>	ОзR	WVEL	WDIR	DATE	
COEFFICIENT	0.369	-0.015	1.981	-0.0087	0.0010	0.117	
EST. ERR.	0.028	0.001	0.224	0.0018	0.0003	0.015	
VARIABLE	SUN	MON	TUE	WES	THU	FRI	SAT
COEFFICIENT	-4.816	-4.670	-4.628	-4.632	-4.714	-4.572	-4.727
EST. ERR.	0.144	0.144	0.145	0.148	0.145	0.147	0.146
MEAN OF THE	-2.216	-2.143	-2.071	-2.033	-2.160	-1.991	-2.067
LOGARITHM							
EST. ERR.	0.041	0.057	0.059	0.046	0.058	0.046	0.047

## 4. Goodness of fit for the model and residual analysis

Figure 5 shows the scatter diagram for the ozone concentration calculated vs observed. The correlation coefficient of the natural logarithm of measured ozone concentrations and the value calculated with the model is 0.78. Figure 6 shows residuals (observed value - calculated value) as a function of time. The graph does not show evidence about lack of independence or lack of constancy in the residuals; however, the residuals corresponding to the first year have smaller values than the other two years. Residuals have zero mean. Figure 7 shows the histogram of residuals and the fit of a normal curve. A chi square test was made to this fitting giving a value of 17.3 with 16 degrees of freedom. This means that the normal distribution is a good approximation for the residuals. The runs test was also applied (Conover, 1980) and did not show a significant difference with a random series. Figure 8 shows the autocorrelation function for the residuals; there is a little structure. The first values of the autocorrelation function are

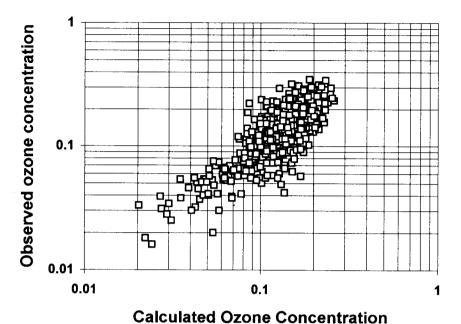


Fig. 5. Scatter diagram for the ozone concentration calculated with the model vs observed. Concentrations in parts per million and logarithmic scales; the correlation coefficient is 0.78.

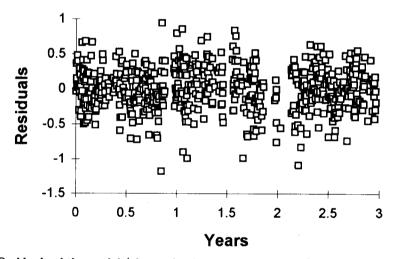


Fig. 6. Residuals of the model (observed value - calculated value) as a function of time.

greater than zero with two values greater than the 95% confidence level. This evidence suggests that residuals show a little trend during the first part of the series. As it was mentioned, Figure 6 shows a few trend to decrease in the residuals, this implies that the model overestimates slightly the ozone concentration during the first year.

It is considered that the fitting satisfies the condition that residuals are a not correlated random variable with zero mean and normal distribution. These conditions are essential for the goodness of fit (Drapper and Smith, 1981).

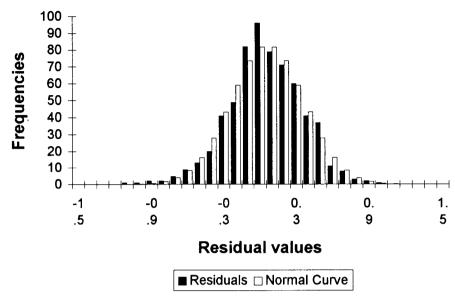


Fig. 7. Histogram of residuals and an adjusted normal curve. A chi square test was made to this fitting giving a value of 17.3 with 16 degrees of freedom.

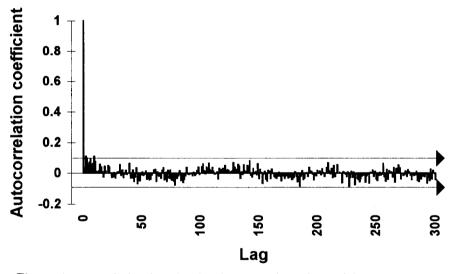


Fig. 8. Autocorrelation function for the residuals of the model.

## 5. Discussion

As it has been mentioned, Table 2 shows the values of the regression coefficients and their standard errors. All the coefficients are one order of magnitude larger than their standard errors, with the exception of the term corresponding to wind speed. In this case, two times the standard error is not larger than the coefficient.

Table 3 shows values for the correlation coefficients after introducing by steps each of the variables. The two solar radiation terms, linear and quadratic, explain 40% of the variance;

42 J. L. BRAVO et al.

after the introduction of the term corresponding to ozone concentration of the previous day, the explained variance increases to 50%; addition of the trend term increases the explained variance to 56%, and the rest of the terms to 61%.

Table 3. Model correlation coefficient after introduction of variables step by step.

Introduced Variable Correlation	GR + GR <sup>2</sup>	O3R	Trend	Wind speed
Coefficient	0.63	0.71	0.75	0.76
Introduced Variable Correlation	Wind direct	tion W	eekly varia	ation
Coefficient	0.77		0.78	

As it has been shown, the relationship between ozone concentration and solar radiation is not linear and thus it was necessary to introduce a second order term. This means that there must be a value for the total solar radiation for which ozone concentration has a maximum. This effect is probably due to days with high solar flux having little atmospheric pollution, with little ozone formation and with high atmospheric transparency. Deriving equation 1 with respect to solar radiation and equating to zero, we obtain 12.3 MJ/M, that is the solar radiation value for maximum ozone concentration.

The coefficient of regression for the previous day ozone is greater than 1, this value must produce an unstable random process. However, in this case, the process is not strictly autoregressive because the dependent variable is the logarithm of the ozone concentration and the independent variable is the ozone concentration of the previous day. This factor, in the second order for the explanation of the variance, shows the importance of the persistence of meteorological conditions in the variations of ozone concentrations and reinforces the hypothesis of the accumulation of ozone and/or its precursors in middle tropospheric layers, probably up to the top of the mountains altitude (approximately 4000 masl), where wind circulation is not restricted. This accumulated gases could stay during the night and contribute to ozone formation next day, after the initiation of mixing due to air convection. These atmospheric process contributes with 10% to the explanation of the variance, after considering the effect of solar radiation.

The trend term has a positive regression coefficient indicating an increase in ozone concentrations during the study period, this agrees with recent results (Juárez et al., 1995). Because the regression was calculated using the loge of ozone concentration, it is necessary to apply the inverse transformation to quantify the trend in the original system. Exponentiation of equation 1 results in a product of exponentials. The term corresponding to the trend has the following value for the period of study:

$$\exp(0.117 \times 3) = 1.42 \tag{2}$$

This means that there is an increment in ozone concentration of 42% during this period at the studied area. This increment is free from the influence of the trends in the parameters considered in the regression equation.

The regression coefficient corresponding to wind speed is negative because the ozone concentration has an inverse relationship with the stability of the air.

The coefficient of the wind direction term is positive because of the reason explained before. This term lost importance after the introduction in the equation of the solar radiation term. This could be explained for a correlation between wind direction and solar radiation as was mentioned earlier.

The terms that explain less variance are the wind direction and the wind speed. This could be because, although the industrial zone is situated at the north of the Valley, the main contribution to air pollution in the area of study comes from the emission of automobiles that are spread over a very extended area.

The values of the dummy coefficients are all different from zero with a minimum value for Sunday and maximum for Friday. These variations could be related to the industrial and vehicular activity during the week in the city. Table 2 also shows mean values of the logarithm of ozone concentrations for the days of the week and their standard errors; these show a similar behavior with minimum values for Sunday, Thursday and Saturday and maximum for Friday; between Friday and Sunday there is a decrease of 22%. The weekly variation in the activity of the city is larger than 22%; furthermore more variance is explained with solar radiation and with ozone concentration of the previous day than that explained with the weekly variation of emissions. This behavior suggests that the permanence of pollutants in the atmosphere during the night has an important contribution in the daily variations of the ozone concentration.

## 6. Conclusions

The variations of ozone concentration at the southwest part of the Valley of Mexico are dominated by the variation of the solar radiation and for the pollution from the previous day. Solar radiation, as has been demonstrated, is an index of the intensity for photochemical processes. These two factors explain 50% of total variance.

This model uses information currently obtained in a urban air monitoring station; then, with the help of a prediction of solar radiation, it could be used to predict high ozone concentrations and to install environmental alerts.

Ozone concentrations have increased 42% during the period from January 1989 to December 1991 at the southwest part of Mexico's Valley. Adding a trend term to the regression equation, after solar radiation and ozone concentration of the previous day, the explained variance increases up to 56%.

The weekly ozone variation has a maximum at Friday and a minimum on Sunday. This variation could be caused by the variation in emissions during the week. The decrease of ozone concentration between Friday and Sunday is 22%. However the decrease of activity in the city is larger than this percentage. This behavior suggests that the permanence of pollution during the night represents an important contribution to the daily variations of the emissions in the city.

The ozone concentration has a minimum when wind blows from the south because industrial area is in the north part of the Valley. However, this factor explains only 3% of total variance. This small contribution to the variance is because the main emission sources are the vehicles that are spread over the extended area of the Valley of Mexico.

The variance that is not explained by the phenomena considered in this work is due to the lack of consideration of the vertical structure of the atmosphere. For example: the altitude

and duration of the thermal inversion, the vertical component of the wind or the intensity and direction of the wind at several altitudes; and also because the oversimplification of complex photochemical processes by the introduction of ozone concentration from the previous day.

## Acknowledgements

We thank Rogelio Montero, Ernesto Jiménez de la Cuesta, Vidal Valderrama, Luis Galindo and Emilia Velazco for their valuable technical assistance.

#### REFERENCES

- Bravo, H., G. Roy Ocotla, P. Sánchez and Torres J., 1991a. Contaminación atmosférica por ozono en la zona metropolitana de la Ciudad de México: Evolución Histórica y perspectivas. Omnia, Coordinación General de Estudios de Posgrado año 7-23, 39-48.
- Bravo, H., R. Sosa and R. Torres, 1991b. Ozono y lluvia ácida en la Ciudad de México. Ciencias No. 22, 33 40.
- Bravo, H., F. Perrin, R. Sosa and R. Torres, 1988a. Incremento de la contaminación atmosférica por ozono en la zona metropolitana de la Ciudad de México. Ingeniería Ambiental 1, 8 14.
- Bravo, H., R. Torres and R. Sosa, 1988b. Ozone and its nightime in the southern Mexico City metropolitan area. *Geofís. Int.* 27 (1), 83 98.
- Campos, M. G., P. Segura, M. H. Vargas, B. Vanda, H. Ponce- Monter, M. Selman and L. M. Montano, 1992. Ozone induced airway hyperresponsiveness to noncholinergic system and other stimuli. J. Appl. Physiol. 73 (1): 354 361.
- Conover, W. J., 1980. Practical non parametric statistics. Second Edition. John Wiley & Sons.
- Chok, D. and J. Heuss, 1987. Urban ozone and its precursors. Environ. Sci. Tecnol., 21-12, 1147-1153.
- Drapper, N. R. and H. Smith, 1981. Applied regression analysis. Second Edition, John Wiley & Sons.
- Holland, D. and T. Fritz-Simons, 1982. Fitting statistical distributions to air quality data by the maximum likelihood method. Atmospheric Environment, 16, 1071-1076.
- Jáuregui, E., 1988a. Efectos del Clima urbano sobre los niveles de contaminantes en la Ciudad de México. Geografía y Desarrollo, 1-2, 37 44.
- Jáuregui, E., 1988b. Inversiones Térmicas y Smog. Unión Geofís. Mex. Bol. Geos, época II, 8-1, 16-19.
- Juárez, A., C. Gay and J. L. Bravo, 1995. Influence of urban ozone in the measurements of total ozone column in Mexico City. Atmósfera, 8, pp. 35-43.
- Miller, R. P., M. de Bauer, A. Quevedo and T. Hernández, 1994. Comparison of ozone exposure characteristics in forested regions near Mexico City and Los Angeles. *Atmospheric Environment*, 28, 141 148.
- Nickerson, E., G. Sosa, H. Hochtein, P. McCaslin, W. Luke and A. Schanot, 1992. Project Aguila in situ measurements of Mexico City air pollution by a research aircraft. Atmospheric Environment, 26B, 445-451.

- Quadri, G. and L. Sánchez, 1992. La Ciudad de México y la contaminación atmosférica. Ed. Limusa, México.
- Riveros H. G., J. L. Bravo, V. Páramo and J. Tejeda, 1993. El ozono y el consumo de gasolina en la zona metropolitana de la Ciudad de México. Bol. Acad. Inv. Científ. México. 13, 31 36.
- Romero T., J. Pérez, J. L. Bravo and S. Gómez, 1994. Efecto citogenético del ozono en cultivo de linfocitos humanos. Revista del Instituto Nacional de Enfermedades Respiratorias. México. Vol. 7, no. 3, julio septiembre.
- Seinfeld, H., 1989. Urban Air Pollution: State of the Science. Science, Vol. 243, February, 745 752.