

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Critical Episodes of PM10 Particulate Matter Pollution in Santiago of Chile, an Approximation Using Two Prediction Methods: MARS Models and Gamma Models

Sergio A. Alvarado O.<sup>1,2\*</sup>, Claudio Z. Silva<sup>1</sup> and Dante L. Cáceres<sup>1,2</sup>

<sup>1</sup>*Division of Epidemiology, School of Public Health,  
Faculty of Medicine, University of Chile,*

<sup>2</sup>*Grups de Recerca d'Amèrica i Àfrica Llatines (GRAAL), Unitat de Bioestadística,  
Facultat de Medicina, Universitat Autònoma de Barcelona, Barcelona,*

<sup>1</sup>*Chile*

<sup>2</sup>*Spain*

## 1. Introduction

Several studies worldwide have reported the effects of air pollution on human health, especially those related with exposure to particulate matter <sup>1,8,17,20,21,25</sup>, research it is currently focussed on studying the acute and short-term effects, especially the mortality and morbidity impact due to cardiovascular and respiratory reasons <sup>1,8,17,21,25</sup>.

This has meant that countries take a series of environmental management measures to control particulate material emissions and also to generate early prediction of high air pollution episodes. Some measures are: a systematic shift to cleaner fuels, restriction of daily circulation of a certain percentage of motor vehicles, temporary shutdown of some industries, and so on. Air pollution causes are diverse, being anthropogenic activities the major contributor to the problem. But the air pollution level is also influenced by other factors such as climate and topography. Climate has a decisive influence on the persistence of air pollutants, and the winds, temperature and solar radiation drastically alter the dispersion and the type of contaminants present at one time. Topography influences the movement of air masses and hence the persistence of pollution levels in a given geographical area. The combination of these ultimately determines the quality of air <sup>20</sup>.

Prediction of critical episodes of air pollution in large cities has become an environmental management tool aimed to protect the health of the population, allowing health authorities to know with some certainty the likely levels of air pollution border within a certain time interval. This prediction has been addressed through different models combining deterministic and probabilistic approaches using various types of information <sup>6,9,13,14,22,28</sup>. The official methodology in use by the administrative authority of the Metropolitan Region of Santiago de Chile to forecast PM10 concentrations is the Cassmassi Model, proposed in 1999

---

\*Corresponding Author

by Joseph Cassmassi<sup>2</sup>. It uses multiple linear regression to predict the maximum concentration of 24-hour average PM10 for the next day (00 to 24 hours). This model includes observed meteorological variables, observed and forecasted weather conditions rates, pollutants expected concentrations rates of expected variations in emissions and others predictors.

Different statistical methods have been used in Chile to model airborne PM10 concentrations of air pollutants including time series<sup>20</sup>, neural networks<sup>13,22</sup> and regression models based on multivariate adaptive smoothing functions (splines) MARS<sup>23</sup>. The predictive efficiency of these models is variable and is closely associated with the behavior and evolution of environmental characteristics. Models that use extreme value theory are widely used for this purpose, especially for episodes that occur over short periods of time and present extreme values or exceedances of emergency limits established by the authority<sup>3</sup>.

The aim of this study is to compare the predictive efficiency of multivariate predictive models Gamma vs MARS to predict “tomorrow” maximum concentration of PM10 in Santiago de Chile in the period between April 1 and August 31 of the years 2001, 2002, 2003 and 2004.

## 2. Methods

### 2.1 Information sources

We used the databases of PM10 collected at the Pudahuel monitoring station, that it is element of the MACAM2-RM monitoring network, for the years 2001, 2002, 2003 and 2004. For each year measurements were selected from April 1 to August 31, which correspond to the time of year with less ventilation in the Santiago basin. We worked with the moving average of 24 hours. For missing data, imputed values were generated by a double exponential smoothing with a smoothing coefficient  $\alpha = 0.7$  (see Annex 1). We selected this monitoring station because most of the year presents the highest levels of PM10 concentration. It is also the most influential in taking administrative decisions about forecasting critical episodes for the next day. Moreover, because of environmental management measures implemented when declaring a critical episode of PM10 pollution, the behavior of the time series, is affected generating lower levels of concentrations that do not reflect actual observed concentration that would have occurred without environmental management measures, so we penalize this effect by introducing a correction constant given by  $\Delta C_I = \text{mean}[CPM_{24_{I-1}} - CPM_{24_I}]$  where  $CPM_{24_{I-1}}$  and  $CPM_{24_I}$  are the average of PM10 concentration on the day before and the day of intervention, respectively, for each month of the study period; the number of episodes in excess of 240 ( $\mu\text{g}/\text{m}^3$ ) for 2001, 2002, 2003 and 2004 corresponds to 4 (2.6%), 11 (7.2%), 5 (3.3%) and 2 (1.3%) respectively. For the construction of the Gamma models use the statistical program Stata 11.0 and MARS models use a demo of the program obtained from the web page of Salford Systems.

### 2.2 Procedures

152 multi-dimensional observations were used; they consisted of a response variable PM10 and 13 predictors. The modeling includes delays of 1 and 2 days for the variables of interest, corresponding (N+1) to tomorrow, i.e. the day to be modeled. The predictor variables for today (delay 1) were defined as **pm0** the hourly average PM10 concentration at 0:00 hrs of day N, **pm6** the average hourly PM10 concentration at 6:00 hrs on day N, **pm12** the average hourly PM10 concentration at 12:00 hrs on day N, **pm18** the hourly average PM10 concentration at 18:00 hrs on day N.

Some predictor variables incorporate delays of 2 days; they are: **pm10h** maximum PM10 concentration of 24-hour moving average between 19 hrs of day N-1 and 18 hrs day N, **mtb** maximum temperature between 19 hrs of day N-1 and 18 hrs o day N, **mhrh** minimum relative humidity between 19 hrs of day N-1 and 18 hrs of day N, **dth** difference between maximum and minimum temperature between 19 hrs of day N-1 and 18 hrs of day N, **vvh**, average wind velocity 19 hrs of day N-1 and 18 hrs of day N. The predictors of tomorrow (N+1) correspond to: **mtm** maximum temperature of day N+1, **mhrm** minimum relative humidity of day N+1, **dtm** difference between the maximum and minimum temperature of day N+1 and **vvm** average speed wind of day. The response in this study (**pm10m**) is the maximum concentration of the 24 hs moving average of PM10 of a day N+1. The values of the variables of tomorrow are forecasts validated and reported by the Chilean Meteorological Office using models Mesoscale Modeling System (MM5), which is a numerical model that uses the equations of physics of the atmosphere for weather forecasting in limited areas at the regional level <sup>15</sup>,

The authorities of the National Environment Commission (CONAMA), have defined four levels of PM10, in order to make management decisions when critical events occur: **good** 0 - 193 ( $\mu\text{g}/\text{m}^3$ ); **alert** 194 - 239 ( $\mu\text{g}/\text{m}^3$ ), **pre-emergency** 240 - 329 ( $\mu\text{g}/\text{m}^3$ ) and **emergency** PM10 > 330 ( $\mu\text{g}/\text{m}^3$ ) <sup>3</sup>. For our study we dichotomized the response into two classes **I**: pm10m < 240 ( $\mu\text{g}/\text{m}^3$ ) and **II**: pm10m > 240 ( $\mu\text{g}/\text{m}^3$ ), ie, "good or alert" versus "pre-emergency or emergency." The aim of the dichotomy is to generate 2 x 2 cross-tabulations between the observed and predicted values of the response for each proposed model.

### 2.3 Construction of Gamma models and MARS models

The fit of the models of a given year was validated with the information of the following year, thus ensuring the independence of the data used for validation with respect to those used in its construction. Therefore no predictions are given for the model year built with year 2004 information since there is no information for 2005. Each model was estimated with data from the period between April 1 and August 31 of a year and applied to the following year's data for the same period, evaluating the fit of these estimates compared with actual observations for that second year.

#### Gamma regression

Gamma models are used in situations where the variable has non negative values; were originally used for continuous data, but now the family of Gamma generalized linear models is used with count data <sup>7</sup>. In general, these models consider different ways of how to work the response variable, such as exponentiation of the response using the log-gamma transformation <sup>12</sup>.

The probability density function for the generalized gamma function is given by

$$f(y; \kappa, \mu, \sigma) = \frac{\gamma^\gamma}{\sigma y \sqrt{\gamma} \Gamma(\gamma)} e^{(z\sqrt{\gamma} - u)}; y \geq 0, \text{ where } \gamma = |\kappa|^{-2}, z = \text{sign}(\kappa) \times \left\{ \frac{\ln(y) - \mu}{\sigma} \right\}$$

and  $u = \gamma \exp(|\kappa|z)$ . The parameter  $\mu$  is equal to  $x^t \beta$  where  $x$  is the matrix of predictors including the intercept and  $\beta$  is the vector of coefficients. For generalized Gamma distribution the expected value conditional on  $x$  is given by:

$$E(y|x) = \exp[x\hat{\beta} + (\hat{\sigma} / \hat{\kappa}) \ln(\hat{\kappa}^2) + \ln(\Gamma\{(1 / \hat{\kappa}^2) + (\hat{\sigma} / \hat{\kappa})\} - \ln(\Gamma(1 / \hat{\kappa}^2)))]$$

where  $\hat{\sigma} = (1/n) \sum_i \exp(\alpha_0 + \alpha_1 \ln(f(x_i)))$  and  $\ln(\sigma)$  is parameterized as  $\alpha_0 + \alpha_1 \ln(f(x))$ .

Since it is required to report estimates and results in the original measurement scale, we work with the exponentiation of the model using the log-gamma transformation  $\mu = E(Y) = \exp(x^t \beta)$  and ensure that the transformation does not affect the interpretations which refers directly to the original scale<sup>12</sup>.

### Multivariate Adaptive Regression Splines (MARS)

MARS is a methodology proposed by Jerome Friedman in August 1991, it tries to build a model of non-linear regression that is based on a product of functions called base smoothing functions (splines). These functions incorporated into its structure predictors entering the model as part of a function, not directly as in classical regression, produces a model for the response in study that may be continuous or binary that automatically selects the predictors present in the final equation, they are incorporated in the smoothing basal functions<sup>5,10,11</sup>.

#### Model for a predictor

The methodology MARS, proposed by Friedman<sup>5</sup>, selects  $K$  nodes of the predictor variable  $x$ , denoted by  $t_k, k = 1, \dots, K$ , which could correspond to each of the observations of the variable; then  $K + 1$  defined regions on the range of  $x$ , where it is associated to each node the linear smoothing function, generating a family of basis functions:

$$B_K^{(q)}(x) = \begin{cases} x^j & j = 0, \dots, q \\ (x - t_k)_+^q & k = 1, \dots, K \end{cases}$$

Where  $(x - t_k)_+^q$  it is known as a truncation function. For the approximation of order  $q$ , we

estimate the function  $\hat{f}_q(x) = \sum_{k=0}^{K+q} a_k B_K^{(q)}(x)$ , usually the order of smoothing to be taken must

be less than or equal to three, so the function and its  $q - 1$  first derivatives are continuous. This restriction and the use of polynomial functions in each subregion produce smooth and tight functions.

#### Generalization to $p$ predictors

For the vector of predictors  $x = (x_1, x_2, \dots, x_p)$  the smoothing function is defined analogously to the univariate case. In this case the space  $R^p$  is divided into a set of disjoint regions and within each region a polynomial of  $p$  variables is fitted.

For  $p > 2$  disjoint regions are considered to define the smoothing approximation as tensor products of disjoint intervals in each of the variables associated to the node location. So placing  $K_j$  nodes in each variable produces a product of  $K_j + 1$  regions,  $j = 1, \dots, p$ . A set of basis functions generating the space of smoothing functions for the entire set of regions, is the tensor product of the corresponding basal one-dimensional smoothing functions associated with the location of the nodes in each variable given by:

$$\hat{f}_q(x) = \sum_{k_1=0}^{K_1+q} \cdots \sum_{k_p=0}^{K_p+q} a_{k_1}, \dots, a_{k_p} \prod_{j=1}^p B_{k_j}^{(q)}(x_j)$$

The selection of basis functions looks for a good set of regions to define a smoothing approach adequate to the problem; MARS generates the basis functions by a stepwise process. It starts with a constant in the model and then begins the search for a variable-node combination that improves the model. The improvement is measured in part by the change in the sum of squared errors (MSE), adding basis functions is done whenever it reduces the MSE.

To evaluate this model, Friedman proposes using the *Generalized Cross Validation* statistic

$$GCV = \frac{\sum_{i=1}^N (y_i - \hat{f}_q(x_i))^2}{N} \bigg/ \left(1 - \frac{C(M)}{N}\right)^2 \quad \text{where } C(M) = 1 + \text{trace}(B(B^t B)^{-1} B^t),$$

$B$  is the design matrix, the numerator is the lack of fit on the training data set and the denominator is a penalty term that reflects the complexity of the model.

To compare the models we considered the following statistics: (a) Pearson's linear correlation between the observed and the predicted value and (b) mean absolute error ratio (mpab) between the observed (obs) and the predicted value (pred) given by

$$mpab = \sum_{i=1}^n \left[ \frac{|PM10_{obs} - PM10_{pred}|}{PM10_{obs}} \right] / n$$

that is equivalent to evaluate the average errors committed by both models in the predictions. We also considered the concordance proportion in each class.

In order to compare the settings of the models and the regions selected for predicting the response variable two MARS models were constructed to each year, one based on 20 and another using 40 base functions, allowing us to choose that pattern which best fits the expected response based on the partitions of the predictor variables<sup>11</sup>. In the case of Gamma regression we used a logarithm link function, with three sets of predictors: the first corresponds to all variables, the second set involved yesterday and today variables and the third included the variables: pm0, pm6, pm18, dtm, dth and vvm. This last set of variables would better describe the behavior of the pm10 concentrations, as has been described by other authors<sup>18</sup>.

### 3. Results

Table 1 shows the descriptive statistics of the variables incorporated in the final modeling of PM10 concentrations. We can see that the maxima for the years 2001, 2002 and 2003 exceed the value of 240 ( $\mu\text{g}/\text{m}^3$ ), such behavior is not seen for 2004.



2001				2002		
	Mean (sd)	Minimum	Maximun	Mean (sd)	Minimum	Maximun
pm10m	122,83 (58,4)	18,50	307,80	123,90 (67,2)	10,40	298,00
pm0	123,85 (106)	1,00	492,00	133,00 (126,6)	1,00	625,00
pm6	78,06 (57,9)	1,00	250,00	80,10 (56,9)	1,00	222,00
pm18	117,50 (83,5)	2,00	467,00	116,70 (83,6)	4,00	412,00
dtm	10,37 (5,3)	2,50	23,00	11,00 (5,1)	0,00	24,00
vvm	1,54 (0,48)	0,53	3,40	1,40 (0,4)	0,00	2,49
2003				2004		
	Mean (sd)	Minimum	Maximun	Mean (sd)	Minimum	Maximun
pm10m	127,50 (50,50)	28,38	260,30	101,00 (43,10)	20,60	230,00
pm0	133,60 (101,40)	1,00	450,00	105,50 (88,30)	1,00	519,00
pm6	86,50 (55,20)	1,00	263,00	68,60 (42,30)	1,00	202,00
pm18	116,10 (72,10)	6,00	385,00	91,40 (59,20)	1,00	345,00
dtm	11,90 (5,03)	2,11	22,90	11,20 (5,02)	1,53	24,00
vvm	1,42 (0,40)	0,56	2,72	missing	missing	missing

sd: standard deviation;  
pm0, pm6, pm18: PM10 concentration at 00, 06 and 18 previous day, respectively;  
dtm: temperature difference of tomorrow; vvm: wind velocity of tomorrow.

Table 1. Mean and standard deviation of predictor variables used to model PM10. Years 2001 to 2004.

Table No. 2 shows the successes of the class I and II by the Gamma models and the MARS models. The correlations are significant for the three models per year, the mpab remain high for all models, except the 40 fb MARS 2002 model where we can see a 19% mean absolute error similar to the Gamma Model. In general for all three years the Gamma models have better performance than MARS for the PM10 concentrations < 240 (µg/m³); MARS models have better performance for PM10 exceeding 240 (µg/m³).

	2001 to 2002			2002 to 2003			2003 to 2004		
	Gamma	MARS 20 bf	MARS 40 bf	Gamma	MARS 20 bf	MARS 40 bf	Gamma	MARS 20bf	MARS 40bf
<b>mpab %</b>	38,00	44,00	32,00	19,00	29,00	19,00	28,00	14,00	25,00
<b>r<sub>pearson</sub></b>	0,83	0,86	0,74	0,82	0,81	0,86	0,78	0,92	0,82
<b>Class I %</b>	98,00	67,00	44,40	99,00	0,00	100,00	99,00	99,00	99,00
<b>Class II %</b>	50,00	97,00	95,40	12,50	98,00	97,00	0,00	100,00	0,00

mpab: mean absolute proportion  
bf: basis function.

Table 2. Results of MARS and Gamma modeling.

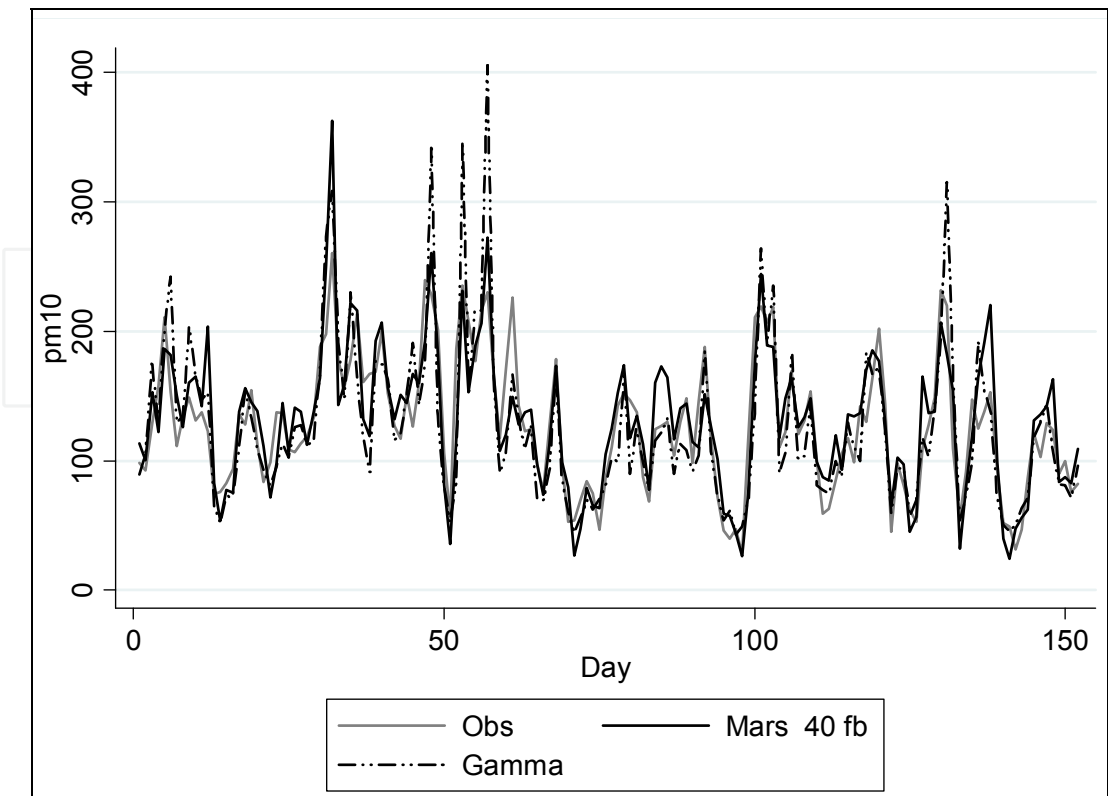


Fig. 1. Shows that the Gamma model predictions are further away from observed PM10 than MARS predictions for high PM10 concentrations for year 2003 using model of 2002, but Gamma model gives better predictions for values below 200 ( $\mu\text{g}/\text{m}^3$ ).

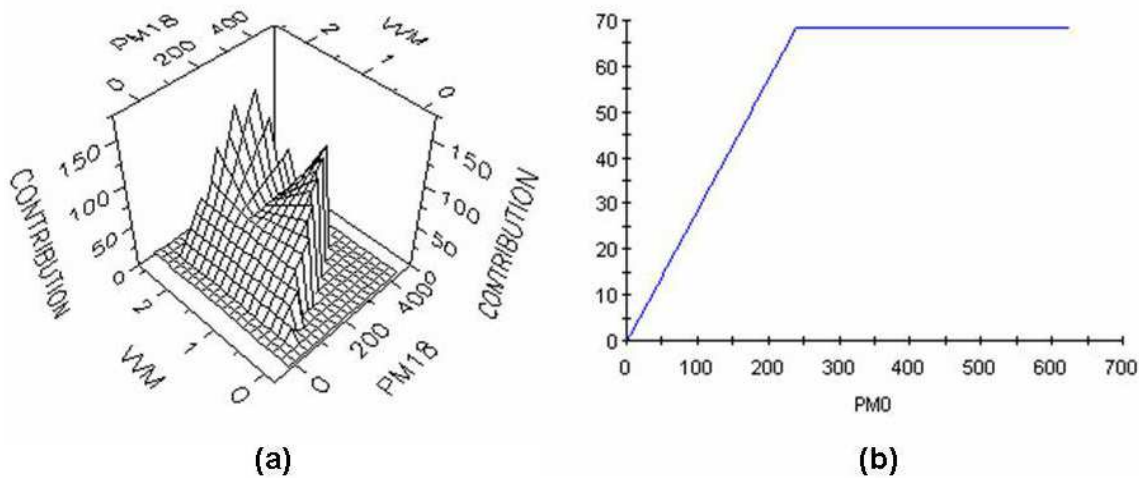


Fig. 2. Corresponds to the MARS model with 40 basis functions for 2002 and illustrates in part (a) the interaction of the basis functions max (0; pm18-137) and max (0; vvm-1, 522) that generate the basis function BF8 which represents an interaction surface (Table No. 3) with maximum value 165 ( $\mu\text{g}/\text{m}^3$ ). This implies that for PM10 concentrations at 18 pm (pm18) above 137 ( $\mu\text{g}/\text{m}^3$ ) and wind speed for tomorrow of 1,522 m/s, the contribution to the PM10 tomorrow concentration due to the interaction has a maximum of 165 ( $\mu\text{g}/\text{m}^3$ ). On the other hand, figure (b) shows that for concentrations above 240 ( $\mu\text{g}/\text{m}^3$ ), the variable pm0 has a maximum contribution to the response of 70 ( $\mu\text{g}/\text{m}^3$ ), which represents the value with which the base function BF4 contributes to the predicted value.



Table No. 3 shows the explicit model for the MARS models with 20 and 40 basis functions and the Gamma models for the years 2001, 2002 and 2003. The complexity of the model is seen in the number of basis functions incorporated into the explicit model, the model for 2002 has nine basis functions of which five are interactions of univariate basis functions, the BF9 and BF8 functions correspond to interactions of the mirror variable for vvm and the corresponding basis function associated with pm18.

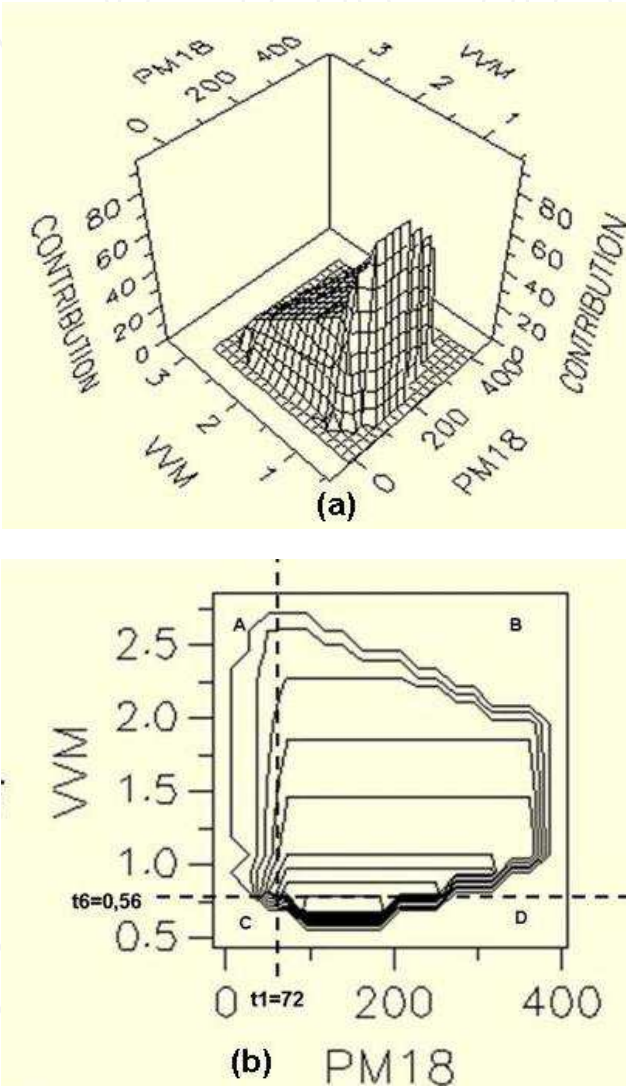


Fig. 3. Shows MARS model with 20 basis functions for 2003 and illustrates in part (a) the interaction of the basis functions max (0;72-pm18) and max (0; vvm-0,56) that generate the basis function BF10 which represents an interaction surface (Table N . 3). In the other figure (b) shows the surface contour are four regions where the letters A, B, C and D, D represents the area for values less than 0,56 meters per second (vvm) and particulate matter concentrations at 18 hours greater than 72 micrograms per cubic meter that generated by the interaction of both base functions with the maximum contribution values to the response variable of interest.

Year 2001		
Mars		
20 basis function	BF1 = max(0;pm18-180)	BF7 = max(0; 1,035-vvm)
	BF2 = max(0; 180-pm18)	BF9 = max(0;1,56-vvm)*BF1
	BF3 = max(0; pm6-13)	BF10 = max(0; pm0-241)
	BF5 = max(0; dtm-2,55)	BF15 = max(0;87-pm6)*BF1
	BF6 = max(0;vvm-1,03)	BF16 = max(0;pm18-183)*BF6
pm10m=109,5-0,48*BF2+0,24*BF3+2,98*BF5+204,5*BF7+0,61*BF9+0,26*BF10-0,008*BF15+0,68*BF16		
Gamma		
pm10m=exp(4,15+0,00076*pm0+0,00223*pm6+0,00292*pm18+0,02504*dtm-0,20295*vvm)		
Year 2002		
Mars		
40 basis function	BF1 = max(0; pm18-137)	BF7 = max(0; 1,307-vvm)*BF5
	BF2 = max(0; 137-pm18)	BF8 = max(0; vvm-1,522)*BF1
	BF4 = max(0; 240-pm0)	BF9 = max(0; 1,522-vvm)*BF1
	BF5 = max(0; dtm+ 0,00000512)	BF12 = max(0; vvm-0,998)*BF5
		BF24 = max(0; vvm-0,867)*BF5
pm10m=146,343-0,553*BF2-0,286*BF4+7,072*BF7+2,093*BF8+0,537*BF9-22,888*BF12+21,844*BF24		
Gamma		
pm10m=exp(3,818+0,00118*pm0+0,002267*pm6+0,00298*pm18+0,035488*dtm-0,149482*vvm)		
Year 2003		
Mars		
20 bases function	BF1 = max(0; pm18-72)	BF8 = max(0;212-pm0)*BF1
	BF2 = max(0; 72-pm18)	BF9 = max(0;pm0-1)*BF3
	BF3 = max(0;pm6-16)	BF10 = max(0;vvm-0,56)*BF2
	BF5 = max(0;vvm-1,07)	BF11 = max(0;dtm-8,17)*BF1
	BF6 = max(0;1,07-vvm)	BF16 = max(0;dtm-2,11)
	BF7 = max(0;pm0-212)*BF1	BF17 = max(0;pm18-79)*BF16
	pm10m=94,2-1,58*BF2-25,99*BF5+111,24*BF6-0,0022*BF7-0,002*BF8+ 0,0009*BF9+0,75*BF10-0,1*BF11+2,02*BF16+0,1*BF17	
Gamma		
pm10m=exp(3,885+0,00108*pm0+0,00107*pm6+0,0030047*pm18+0,02658*dtm)		

bf: basis functions;  
pm0, pm6, pm18: PM10 concentration at 00, 06 and 18 previous day, respectively;  
dtm: temperature difference of tomorrow; vvm: wind velocity of tomorrow.

Table 3. Explicit MARS models for 20 and 40 basis functions and Gamma models for years 2001, 2002 and 2003.

4. Discussion

The MARS modeling selects those significant predictors and detects possible interactions between them generating more flexible models from the point of view of interpretation; since interactions are always restricted to a subregion, they are expressed algebraically through the basis functions, generating a parsimonious model that represents without any further transformation the nature of the working variables. This procedure creates nodes or

cutting points that act as threshold values for each predictor variable selected, indicating the change in the contribution generated by the basis function to the response under study.

After selecting the optimal model, MARS fits the model removing one variable, in order to determine the impact on the quality of the model due to eliminate that variable and assigns a relative ranking from the variable most important to the least important. Thus, replacement or competing variables are defined in the MARS methodology allowing us to treat missing values generating a special basis function for those variables that have no information, ie, generate an allocation basis function whose purpose is to impute the average value of the predictor without information.

As could be determined in this study, the MARS models performed similarly in their predictive efficiency when changing the number of basis functions, regardless of the year for which the model was built and the year used for validation. Partitioning the range of the predictor variables did not improve the quality of predictions, showing that sometimes a less fine partition (minus subregions) methodology was more robust than a thin partition. This could be explained as the time series of PM10 over time show a downward trend and variations in lower concentrations since the intervention measures applied <sup>3</sup>.

The Pudahuel monitoring station registered changes in annual and monthly concentrations of PM10 between 1998 and 2004. These changes were due mainly to global decontamination measures, extraordinary administrative actions on critical air pollution events, use the forecast model of critical events to reduce their negative impact and meteorological factors. Such changes affect the performance of these models allowing a less complex structure without loss of efficiency because the relationship established between the predictors is not as complex, this is reflected in the basis functions constructed. In turn there are differences from one year to another, which could be influenced by weather conditions of each year, such as dry year, El Niño and other climate changes. Another factor to evaluate and that could be relevant in this behavior is the increase in car ownership in recent years.

Additionally, different methodologies have been implemented to model the concentration of particulate matter in the metropolitan region of Santiago de Chile. For example, Silva and colleagues applying time series using transfer functions involving meteorological variables, reported a 40% average ratio of absolute error in the prediction of critical events<sup>24</sup>. Moreover Perez and colleagues, using neural networks with pre-smoothing, reported approximately 30% mean absolute error in predicting the critical episodes <sup>19</sup>. Subsequent work by the same author using a neural network shows results superior to classical linear regression <sup>18</sup>. Furthermore, Silva and colleagues assessed two methodological approaches to the problem of predicting air pollution by particulate matter, reporting that MARS generated better predictions than the discriminant analysis <sup>23</sup>.

Application of these models to the monitoring station Pudahuel shows that the predictor variables that best predict the answer would be (i) some PM10 concentrations: pm0, pm6 and pm18 and (ii) meteorological variables: vvm dtm; this is consistent since MARS selected variables related to persistence of ventilation conditions, which are related to meteorology <sup>23</sup>.

Prediction methodologies give us adequate models to study particulate matter pollution. Gamma regression was generally lower in Class II hits than MARS models, except for predicting from 2003 to 2004, so MARS appears as a better prediction tool of pollution

episodes above the 240 ( $\mu\text{g}/\text{m}^3$ ). Additionally, an advantage of the Gamma model is that generally makes better predictions for the class I ( $\text{PM}_{10} < 240 \mu\text{g}/\text{m}^3$ ), ie, is sensitive to values of concentrations of particulate matter associated to air quality good or warning; this would be given by the behavior of the variable of interest. For example, for 2001 the intervention measures explain the best fit of the Gamma modeling in that year, whereas later on these measures had a moderate impact on the values of the series of PM10, making these models less efficient compared to MARS, as happens for the years 2002, 2003 and 2004.

This study basically aims to proposed models being able to detect concentrations above the threshold of 240  $\mu\text{g}/\text{m}^3$  that mandate periods of epidemiological alert in Santiago de Chile; this consideration would place the MARS modeling as a tool statistically more powerful than Gamma modeling. This last point is consistent with previous findings showing that MARS is more efficient than other techniques <sup>4,23,27</sup>. This could be explained by the smoothing approximation that uses this methodology generating breakdowns in the predictors time series and locally adjusting the basis functions in function on such nodes.

## 5. Appendix 1. – Exponential smoothing

This technique uses a smoothing constant; if the constant is close to 1 it affects very much the new forecast and conversely when this constant is close to 0, the new forecast will be very similar to the old observation . If you want a sharp response to changes in the predictor variable you must choose a large smoothing constant. The formula that relates the coefficient and the time series is given by  $S_t^{[2]} = \alpha S_t + (1 - \alpha)S_{t-1}^{[2]}$  where  $S_t = \alpha x_t + (1 - \alpha)S_{t-1}$ ; to generate the smoothing is necessary to have the values  $S_0$  and  $S_t$ , where  $x_t$  corresponds to the values of the original series <sup>16,26</sup>.

## 6. References

- [1] Biggeri A, Bellini P, Terracini B. [Meta-analysis of the Italian studies on short-term effects of air pollution--MISA 1996-2002]. *Epidemiol Prev* 2004;28(4-5 Suppl):4-100.
- [2] Cassmassi J. Improvement of the forecast of air quality and of the knowledge of the local meteorological conditions in the Metropolitan Region; 1999.
- [3] CONAMA. Comisión Nacional del Medio Ambiente. Decreto Ley N° 59, Diario Oficial de la Republica de Chile. Lunes 25 de mayo de 1998". 1998.
- [4] De Veaux R, Psichogios D, Ungar L. A comparison of two nonparametric estimation Schemes: MARS and Neural Networks. *Computers Chemical Engineering*, 1993; 17(8): 819-837.
- [5] Friedman J. Multivariate Adaptive Regression Splines *The Annals of Statistics* 1991: 19(1): 1-141.
- [6] Gokhale S, Khare M. A theoretical framework for episodic-urban air quality management plan (e-UAQMP). *Atmospheric Environment* 2007; 41(2007): 7887-7894.
- [7] Hardin J, Hilbe J. Generalized linear models and extensions. USA: College Station Texas; 2001.
- [8] Janke K, Propper C, Henderson J. Do current levels of air pollution kill? The impact of air pollution on population mortality in England. *Health Econ* 2009;18(9):1031-55.
- [9] Kurt A, Gulbagci B, Karaca F, Alagha O. An online air pollution forecasting system using neural networks. *Environ Int* 2008; 34(5): 592-598.

- [10] Lewis PAW, Stevens JG. Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). *J Am Statist Assoc.* 1991;86:416.
- [11] MARS. User guide. Salfords-Systems 2001. Multivariate Adaptive Regression Splines Vol. 2010; 2001.
- [12] McCullagh P, Nelder J. Generalized Linear Models. London, UK: Chapman Hall, 1989.
- [13] McKendry IG. Evaluation of artificial neural networks for fine particulate pollution (PM<sub>10</sub> and PM<sub>2.5</sub>) forecasting. *J Air Waste Manag Assoc* 2002; 52(9): 1096-1101.
- [14] Meenakshi P, Saseetharan MK. Urban air pollution forecasting with respect to SPM using time series neural networks modelling approach--a case study in Coimbatore City. *J Environ Sci Eng* 2004; 46(2): 92-101.
- [15] MM5 Community Model. Vol. 2010. Pennsylvania State University/National Center for Atmospheric Research Numerical Model 2008.
- [16] Montgomery DC, L. A. Johnson, Gardiner. JS. Forecasting and Time Series Analysis. , 1990.
- [17] Pascal L. [Short-term health effects of air pollution on mortality]. *Rev Mal Respir* 2009;26(2):207-19.
- [18] Perez P, Reyes J. Prediction of maximum of 24-h average of PM<sub>10</sub> concentrations 30 h in advance in Santiago, Chile. *Atmospheric Environment* 36: 2002: 36: 4555-4561.
- [19] Pérez P, Trier A, Silva C, Montaña RM. Prediction of atmospheric pollution by particulate matter using a neural network. In: Conf on Neural Inf Proc; 1998 Proc. of the 1997; Dunedin, New Zealand: Springer-Verlag, Vol 2; 1998.
- [20] Sanhueza P, Vargas C, Jimenez J. [Daily mortality in Santiago and its relationship with air pollution]. *Rev Med Chil* 1999; 127(2): 235-242.
- [21] Sanchez-Carrillo CI, Ceron-Mireles P, Rojas-Martinez MR, Mendoza-Alvarado L, Olaiz-Fernandez G, Borja-Aburto VH. Surveillance of acute health effects of air pollution in Mexico City. *Epidemiology* 2003;14(5):536-44.
- [22] Scott GM, Diab RD. Forecasting air pollution potential: a synoptic climatological approach. *J Air Waste Manag Assoc* 2000; 50(10): 1831-1842.
- [23] Silva C, Pérez P, Trier A. Statistical modelling and prediction of atmospheric pollution by particulate material: two nonparametric approaches. *Environmetrics*, 2001: 12: 147-159.
- [24] Silva C, Firinguetti L, Trier A. Contaminación ambiental por partículas en suspensión: Modelamiento Estadístico. In; 1994; Actas XXI Jornadas Nacionales de Estadística, Concepción, Noviembre; 1994.
- [25] Staniswalis JG, Yang H, Li WW, Kelly KE. Using a continuous time lag to determine the associations between ambient PM<sub>2.5</sub> hourly levels and daily mortality. *J Air Waste Manag Assoc* 2009;59(10):1173-85.
- [26] Stata. Stata: Release 11. Statistical Software. In: College Station TSLTS, ed., 2009.
- [27] Steinberg D. An Alternative to Neural Nets: Multivariate Adaptive Regression Splines (MARS). *PC AI's* 2001; 15(1): 38.
- [28] Tobias A, Scotto MG. Prediction of extreme ozone levels in Barcelona, Spain. *Environ Monit Assess* 2005; 100(1-3): 23-32.





## **Air Pollution - Monitoring, Modelling and Health**

Edited by Dr. Mukesh Khare

ISBN 978-953-51-0424-7

Hard cover, 386 pages

**Publisher** InTech

**Published online** 23, March, 2012

**Published in print edition** March, 2012

Air pollution has always been a trans-boundary environmental problem and a matter of global concern for past many years. High concentrations of air pollutants due to numerous anthropogenic activities influence the air quality. There are many books on this subject, but the one in front of you will probably help in filling the gaps existing in the area of air quality monitoring, modelling, exposure, health and control, and can be of great help to graduate students professionals and researchers. The book is divided in two volumes dealing with various monitoring techniques of air pollutants, their predictions and control. It also contains case studies describing the exposure and health implications of air pollutants on living biota in different countries across the globe.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sergio A. Alvarado O., Claudio Z. Silva and Dante L. Cáceres (2012). Critical Episodes of PM10 Particulate Matter Pollution in Santiago of Chile, an Approximation Using Two Prediction Methods: MARS Models and Gamma Models, Air Pollution - Monitoring, Modelling and Health, Dr. Mukesh Khare (Ed.), ISBN: 978-953-51-0424-7, InTech, Available from: <http://www.intechopen.com/books/air-pollution-monitoring-modelling-and-health/critical-episodes-of-pm10-pollution-in-santiago-of-chile-an-approximation-using-mars-and-gamma-model>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen