Generalized model of prediction of natural gas consumption

Submitted to: Journal of Energy Resources Technology (JERT- ASME)

S.Gil[‡] *and J. Deferrari*
Distribution Division of ENARGAS.
ENARGAS – *Natural Gas Regulatory Agency of Argentina.*
Suipacha 636 - 4p. (1008) Cap. Fed. - Argentina.
FAX: (5411) 4348 -0532/0535-e-mail: sgil@df.uba.ar

Abstract: We herein present the basic characteristics of a model developed at the Distribution Division of ENARGAS, intended to predict the natural gas consumption of the major cities of Argentina. This model is an extension and generalization of the original model we developed to predict the daily gas consumption in the short range (1 to 5 days). In its original form, the model has been used successfully to predict the daily gas consumption of the Greater Buenos Aires (GBA) area and most of the major cities of Argentina. The model is able to predict the consumption 1 to 5 days in advance with 10% of uncertainty. The new version presented here, extends its applicability to the intermediate range (1 to 5 years). It allows us to estimate the annual peak consumption, load factors and the optimal transportation capacity for a given region of interest. We also present a novel procedure to obtain the distribution of daily consumption from the monthly consumption. This procedure can be used to obtain the parameters of the model, the peak consumption and load factor of the different types of gas consumers for a given region, using the information of consumption that can be obtained from the monthly billing.

I.- Introduction

The prediction of natural gas consumption is crucial for the gas distribution and transportation companies as well as for the government agencies associated to the natural gas sector. In particular, the short range prediction, 1 to 5 days, is important to ensure the normal supply of natural gas to a given city or community. This type of prediction is particularly important for countries like Argentina, where the production centers are far from the major centers of consumption. In the case of Argentina these distances are of about 2000 km. This circumstance and the lack of large reservoirs make it necessary to develop reliable models to predict the gas consumption a few days in advance.

There is also a need to predict consumption in the intermediate range of time, within 1 to 4 years, in order to adapt and upgrade the infrastructure of transportation and distribution. This type of prediction is also useful for all the sectors of the gas industries that need to plan their production and optimize their anticipated purchase.

We herein present the basic characteristics of a model to perform short range prediction and its generalization for intermediate range predictions. We also present a novel formalism to obtain the daily distribution of consumption, based on the monthly consumption. This association is important, not only because the monthly consumption is often a known piece of information, available from billing, but also because it allows us to better define the parameters of the model and to obtain the load factors for different segments of consumers. Furthermore, since the information from billing is often known for different types of consumers, i.e. residential, commercial, industrial, etc., it is possible to obtain the daily distribution of consumption for each type of consumer. The load factors are important parameters to determine the transportation cost of natural gas for different types of consumers. Often the transportation cost[1] is calculated for a given segment or type of consumer C_i (i= Residential, commercial, etc.) as:

$$C_i = C_0 + \frac{A}{LF_i} \quad ; \quad i = \textit{residential, commercial, etc.} \tag{1}$$

Here, C_0 and A are two constants that are characteristic of the system under consideration, and LF_i is the Load Charge Factor associated to the particular consumption segment. It is defined as the ratio of the daily annual average of consumption $<Q(i)>$ to the maximum daily consumption in that year $Q_{max}(i)$, i.e.

$$LF_i = \frac{<Q(i)>}{Q_{max}(i)} \quad . \tag{2}$$

Therefore, the determination of these factors is crucial to several sectors of the gas industry.

Thus, in this study we present the basic characteristics of the model we have developed for the case of the major cities of Argentina. First we describe the short range version of the model to predict the firm component of the consumption and compare the result of the model with the observed values. We then present the extension of the model to the intermediate range. We apply this model to predict the maximum annual consumption. We also discuss a novel procedure to obtain the daily distribution of consumption using the monthly consumption obtained from billing. This procedure is quite general and can be used to obtain the distribution of daily consumption for the different type of consumers. We used the firm part of the consumption as a bench mark to validate the procedure. Finally , in the appendix we present a mathematical justification of the procedure proposed.

II.- **Habits of consumption**

In this section we present the result of our observation on the habits of gas consumption for the case of Argentina. It is possible that some of these characteristics may be valid for other regions. However, it is necessary to check the validity of these habits prior to the application of the model to a new region of interest.

Throughout this work, we will concentrate mainly on the firm component of consumption. The cost of natural gas in Argentina, in the period of study (1994 to 2001), has been very stable, with variations in the order of or less than 1% per year. The average annual consumption per user, for different consumer types, during the period of time studied here, has been almost constant, with a slight tendency to decrease. Figure 1 illustrates this situation for the case region of GBA, supplied by MetroGas, for both residential (R) and commercial (C) users. The situation for all the regions analyzed and other types of users shows a similar trend. This behavior can be described by the following linear expression:

$$Q_{usr}^{(i)}(t) = Q_{usr\_0}^{(i)} \cdot \left(1 + f_{usr}^{(i)} \cdot (t - t_0)\right), \tag{3}$$

where $Q_{usr}^{(i)}(t)$ represents the average annual consumption per user, corresponding to the component i (Residential, Commercial, etc.) at the time t. $Q_{usr\_0}^{(i)}$ is the corresponding value at the reference time t_0 and $f_{usr}^{(i)}$ is the parameter that determines the annual variation of $Q_{usr}^{(i)}$.

Figure 2 shows the annual variation in the number of consumers for the GBA, supplied by MetroGas. Here again we find a linear variation in the number of users with time, i.e.:

$$N^{(i)}(t) = N_0^{(i)} \cdot \left(1 + f_N^{(i)} \cdot \Delta t\right) \tag{4}$$

where $N^{(i)}(t)$ represents the number of users type (i) for the year indicated by t, $N_0^{(i)}$ is the number of users type i for the year t_0 taken as reference, $\Delta t(=t-t_0)$ is the number of years between t and t_0. $f_N^{(i)}$ is the factor that describes the increase in the number of users type (i) for the region of interest. The parameters $N_0^{(i)}$ and $f_N^{(i)}$, as well as $Q_{usr\_0}^{(i)}$ and $f_{usr}^{(i)}$ can be obtained from least square fit of the data.

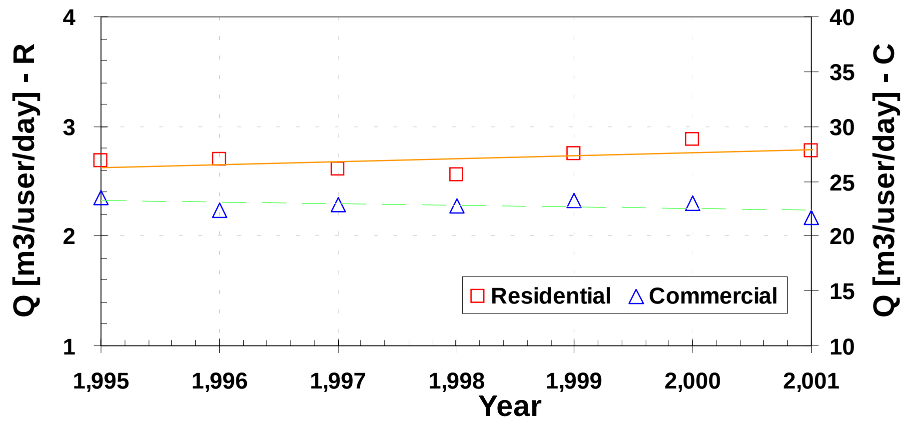


Figure 1: Average annual consumption per user, for residential (R) and commercial (C) users, for the region of GBA supplied by the distribution company MetroGas. The values presented here are the average daily consumption for each year. We observe an almost constant behavior, with a slight decreasing trend. The lines are fits to the data using expression (3).

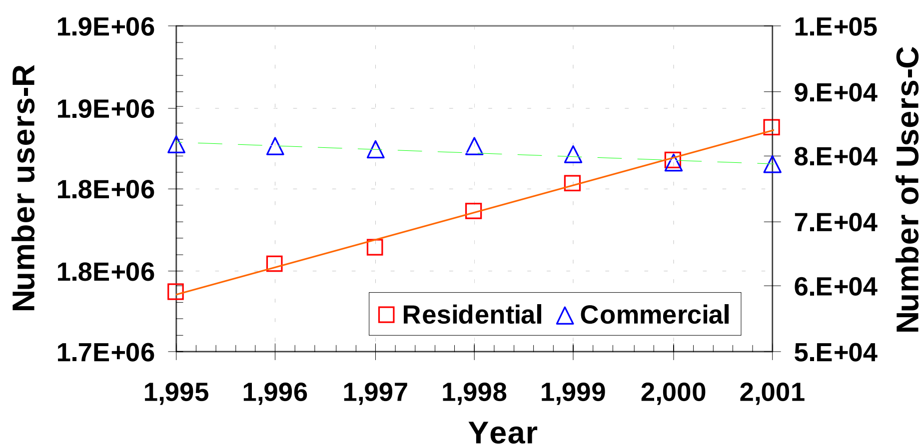


Figure 2: Variation in the number of users, Commercial (C) and Residential (R), as a function of time. These data correspond to the region of GBA supplied by MetroGas. The lines are fits to the data using expression (4).

Combining expressions (3) and (4) it is possible to obtain the total average variation of consumption for each region and for the different types of users. i.e.:

$$Q_{annual}{}^{(i)}(t) = Q_{usr\_0}^{(i)} \cdot N_0^{(i)} \cdot \left(1 + f_{usr}^{(i)} \cdot \Delta t\right) \cdot \left(1 + f_N^{(i)} \cdot \Delta t\right) \approx Q_{annual\_0}^{(i)} \cdot \left(1 + \left[f_{usr}^{(i)} + f_N^{(i)}\right] \cdot \Delta t\right), \quad (5)$$

where $Q_{annual\_0}{}^{(i)} = Q_{usr\_0}^{(i)} \cdot N_0^{(i)}$. We have also defined $f_Q^{(i)} = f_{usr}^{(i)} + f_N^{(i)}$. This parameter characterizes the total variation in consumption, in a given region and for the particular segment of users indicated by i. Therefore it is possible to write the total average variation of consumption for each region as:

$$Q_{annual}{}^{(i)}(t) = Q^{(i)}_{0\_annual} \cdot \left(1 + f_Q^{(i)} \cdot \Delta t\right) \qquad (6)$$

This last expression will be useful to make projections of consumption in the intermediate time range. In the cases analyzed here, the time dependency of the consumption per user and the number of users turned out to be linear. This particular dependence is not a requirement for the validity of the model, However, these relations must be known.

III.- Model of prediction – Short range

The prediction of the firm components (residential, commercial, etc.) of natural gas consumption is made profiting from the correlation that usually exists between the daily consumption and the temperature. Traditionally, the concept of Heating Degree Day (*HDD*) has been used to establish such a connection. This parameter is defined as the sum of the difference of the average hourly temperatures T_i and a temperature taken as reference T_{ref}. This temperature varies for each region but is often in the neighborhood of 18^0C (65^0F). More specifically, the *HDD* is defined as:

$$HDD = \sum_i \left(T_{ref} - T_i\right) \qquad (7)$$

where the sum is carried out for all the hours of the day for which $T_{ref} > T_i$. It must be pointed out that the calculation of the *HDD* requires the knowledge of the temperatures for every hour of the day. This information may be available *a posteriori* but is not simple to predict in advance. Using a data base of temperatures corresponding to several years for several cities of Argentina, we have found that the HDD is highly correlated with the mean daily temperature, $\langle T \rangle$, for the same day. $\langle T \rangle$ is defined as the average value between the maximum, T_{max}, and the minimum daily temperature, T_{min}. Figure 3 illustrates this situation for the case of the city of Buenos Aires. This figure also shows that for $\langle T \rangle$ less than T_{ref}, the value of HDD is inversely proportional to $\langle T \rangle$, which has been the case for all the cases analyzed. This observation is very significant, because it indicates that we can replace the parameter HDD with $\langle T \rangle$, which is not only much less costly to obtain and use, but also has the great advantage that it can be reliably predicted a few days in advance. This fact also justifies the approximation of taking $HDD \approx T_{ref} - \langle T \rangle$.

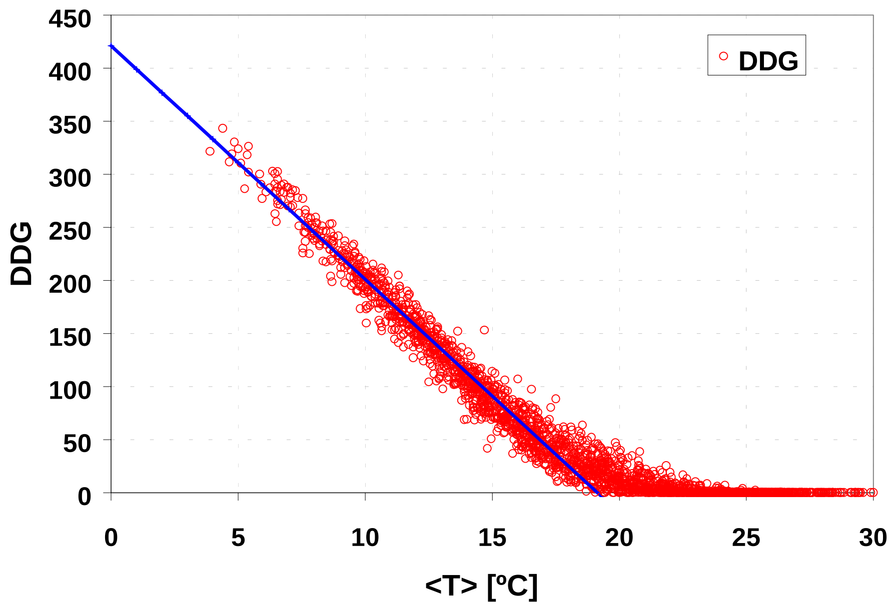


Figure 3: Variation in the number of users as a function of time. These data correspond to the region of GBA supplied by MetroGas. The continuous line is a fit to the data using expression (4).

The model we have developed[2,3] is based on the observation that the users do not respond only to the current temperature but also to the preexisting thermal scenario. This statement can be confirmed by the observation that usually the consumption of gas is greater, for the same mean temperature, on a day in Winter than on a day in Summer or Spring. This inertia is evidenced by the fact that the users usually do not turn on their heating until the temperature has remained low for a few days. Similarly, users do not turn their heating off until the increase in temperature has persisted for a few days. This effect of inertia in the consumption can be taken into account by introducing the concept of effective temperature, T_{eff}, that is defined as the linear combination between the mean $<T>$ of the day under consideration and the moving average of the mean temperature, $<T\_n>$, of the n previous days. Usually n varies between 3 and 5. More specifically:

$$<T\_n> = \frac{1}{n} \cdot \sum_{i=-n-1}^{i=-1} <T>_i \,, \qquad (8)$$

where i represents the different previous days: $i=0$ refers to today, $i=-1$ refers to yesterday, $i=-2$ the day before yesterday, etc.

$$T_{eff} = w \cdot <T> + (1-w) \cdot <T\_n> . \qquad (9)$$

Here w is a weighting factor between 1 and 0, that is obtained from the best fit of the data. $\langle T \rangle$ represents the mean temperature of the day of interest.

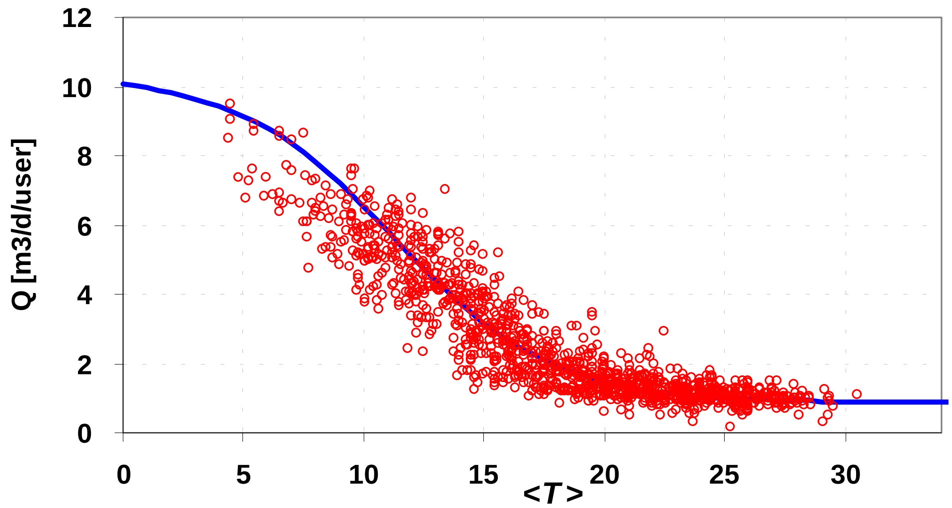


Figure 4: Variation of the daily total firm consumption (residential, commercial, etc.), represented by Q_i , as a function of the mean temperature, $\langle T \rangle$, for the region of GBA supplied by MetroGas, for all the working days over the years 1996 to 2000. The circles are the measured data and the continuous curve is a fit to the data using expression (10).

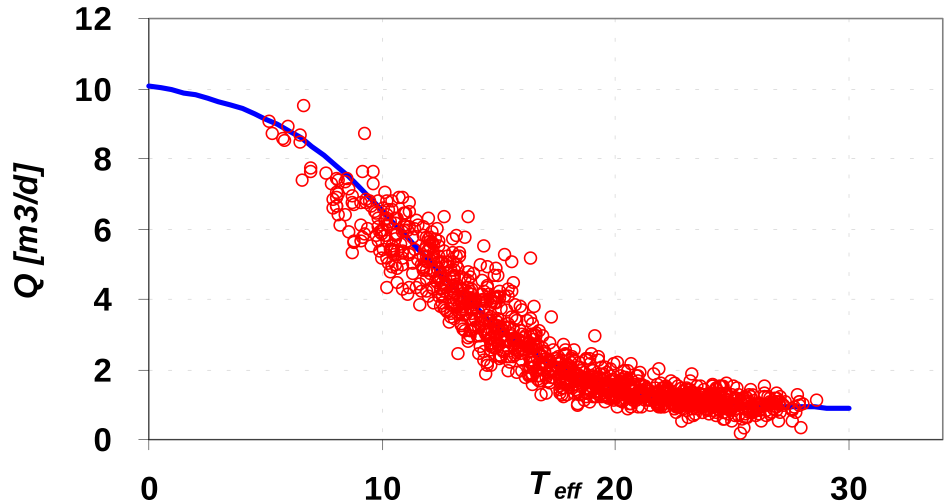


Figure 5: Variation of the daily total firm consumption, Q_i , as a function of the effective temperature, T_{eff}, for the region of GBA supplied by MetroGas, for all the working days over the years 1996 to 2000. The symbols (circles) are the measured data and the curve is a fit to the data using expression (10).

The importance of the effective temperature, is that it automatically incorporates the effects of inertia or hysterisis that is characteristic of the consumption of gas. Using this parameter it is possible to describe the relation of the consumption $Q^{(i)}$ versus T_{eff} by the function:

$$Q^{(i)}(T_{eff}) = Q_0^{(i)} \cdot \left(1 + f_c^{(i)} \cdot tanh\left(\frac{T_{eff} - T_0}{\Delta T}\right)\right). \tag{10}$$

In this expression , all the parameters T_0, ΔT, $f_c^{(i)}$ and $Q_0^{(i)}$ depend on the particular region under consideration and the type of user under consideration, the superindex *(i)* indicates these two situations. For simplicity we have not included the superindex in the parameters T_0, ΔT, but they are implicit. As can be seen in figures (4) and (5), expression (6) does provide an adequate description of the data. Also the dispersion of the data points representing the daily consumption per user show less dispersion arround the model, when we plot $Q^{(i)}$ as a function of T_{eff}. The meaning of the four parameters appearing in expression (6) can be easily understood. The inflection point of the curve characterized by the coordinates (T_0 , $Q_0^{(i)}$), $Q_0^{(i)}$, represents the consumption at $T_{eff}=T_0$ and is related to the average annual daily consumption, discussed previously. The product $Q_0^{(i)} \cdot f_c^{(i)}$ represents the asymptotic difference in consumption between the lowest (hottest) days and the highest (coldest) days. Our study reveals, as expected, that the parameters T_0, ΔT and $f_c^{(i)}$, do not change in time. They are in a way characteristic of the inhabitants of the region. On the other hand, the time dependence shown by expressions (3) and (6) can be directly applied to $Q_0^{(i)}$, i.e.

$$Q_0^{(i)}(t) = Q_{00}^{(i)} \cdot \left(1 + f^{(i)} \cdot (t - t_0)\right) \tag{11}$$

The parameter $f^{(i)}$ represents $f_{usr}^{(i)}$ or $f_Q^{(i)}$ depending on whether we are considering the consumption per user , eq. (3) or the total consumption of the region, eq.(6). In this manner, we see that the present model can incorporate intermediate range variations of the consumption with time. In fact, our model can reproduce the consumption of a given region for a period of time spanning 4 to 7 years with the same parameters. A final consideration for predicting the short range daily consumption is the consideration of the weekends and holidays. From the analysis of the data we were able to reproduce the data for most of the regions of Argentina by using a factor that reduces the prediction of the model by about 10% on Saturdays and by about 15% on Sundays and holidays. Figure 6 shows the comparison between the prediction of our model and the observed consumption for the total firm consumption of gas for the region of the GBA supplied by MetroGas. This type of agreement is representative of the quality of fit obtained for most of the regions studied. We observed that on 95% of the days the agreement between the model and the real data is of the order of 10%. It is interesting to note that the prediction of the model is more robust than the prediction based solely on the temperature. This is because the effective temperature depends only partially on forecasted temperatures, which involve uncertainties, and partially on temperatures of previous days that have already been measured.

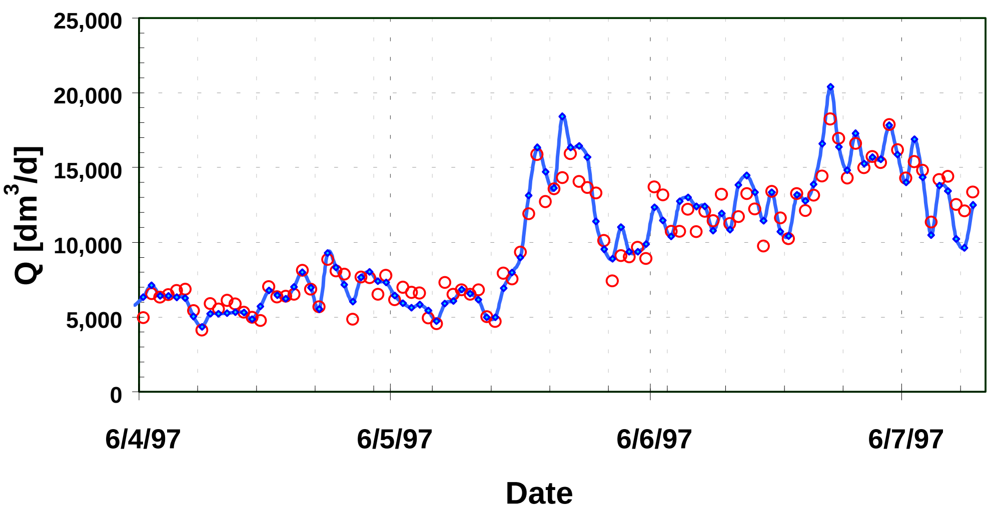


Figure 6: Comparison of the observed total firm consumption of natural gas in (open circles) for the region of GBA supplied by MetroGas and the prediction of the model (curve). The vertical grid corresponds to the beginning of the week (Sundays). The interval of time presented in this plot (from April to July, 1997) spans the Autumn and the beginning of Winter.

IV.- Model of prediction – Intermediate range

In principle the model described in the previous section can be used to make intermediate range predictions of consumption. However, due to the basic impossibility of making long term temperature forecasts, in practice, the usefulness of the model as described this far is limited to1 to 5 days.

In the case of the natural gas industry, what we usually are interested in is not necessarily predicting the consumption for a given day one or two years from now, but rather the probability of occurrence of a given scenario of consumption. The probability distribution of temperatures for a given region can be obtained from the historical data base of temperatures. In particular if we have access to a data base of daily maximum and minimum temperatures for 10 or more years, it is possible to obtain a reliable distribution of future temperatures for that region. The rationale of this statement is that the temperatures that have occured in the recent past can also occur in the future with the same probability. Here we assume that possible global temperature changes have negligible effect in the time ranges studied here, 1 to 10 years. This type of global changes are expected to be of the order of 1^0C per century [4]. Therefore, using the historical data base of temperatures, we can obtain the effective temperature for the whole period of time. From a histogram of frequency of occurrence of temperature, we can easily obtain the probability of occurrence of temperature in the future. Combining these probabilities with the model of prediction, with parameters adjusted as discussed in the previous section, we can obtain the probability of occurrence of a given scenario of consumption for any year in the near future. The evolution of consumption with time is already contained in the parameter indicated by expressions (3),(4), (6) and (7). Therefore, we are able to predict, for any region and for any year in the intermediate range, the probability distribution of consumption and most probable magnitude of the uncertainty of this distribution. In Fig. 7 we present the corresponding distribution for the case of MetroGas, taking as reference the year 2000 and the corresponding projection for the year 2001. In this figure we present the probability of consumption, expressed in thousands of standard cubic meters per day (Dm^3/day), the probability ais expressed in terms of the number of days in the year that the consumption indicated in the horizontal axis is most likely to occur. This type of information allows estimate the number of days in a given year in the future for which, the consumption may exceed a given value. The model therefore allows us to determine optimum values of gas that may be adequate to celebrate contracts (reserved capacity) with producers and transporters. This type of distribution is also useful to estimate the Load Factor (*LF*) defined by expression (2), since the model allows to obtain both the average consumption $<Q>$ and the maximum consumption, Q_{max}, for a given year. In Fig.8 we present the results of the optimum reserved capacity of gas "predicted" by the model for past years for the case of Greater Buenos Aires supplied by MetroGas, under the imposed condition that the consumption does not exceed this value on more than one day in twenty years. The maximum volumes actually used also have been plotted. We clearly see that the predictions of the model are never exceeded. The fact that for some years the real maximum consumption are below the prediction of the model, is associated with the fact that not all the years have extremely cool winters. The important thing is that for those years that did have extreme cool winters, the actual data are close to the prediction of the model but do not exceed it.

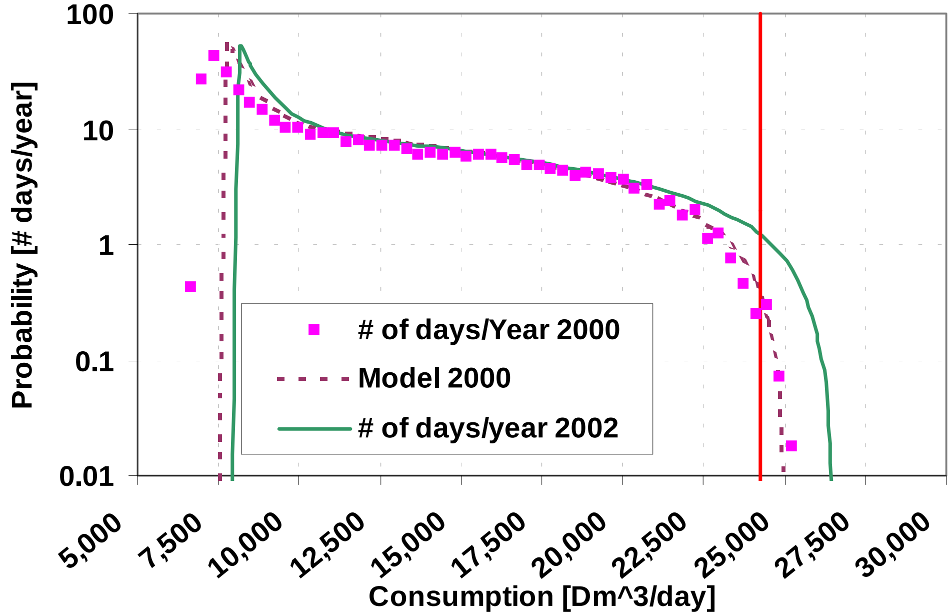


Figure 7: Probability of occurrence of firm consumption for a given year for the region of GBA supplied by MetroGas. The horizontal axis represents the total consumption in Dm^3/day and the vertical axis is the probability expressed in number of days per year that a given scenario of consumption can occur in a year. The dashed curve is the prediction of the model for the year taken as reference (2000), the square symbols are the observed data for that year. The heavy continuous curve is the prediction of the model for a future year. In particular, the area between the heavy curve and the vertical line at ordinate Q_0, represents the probability that the consumption in this region exceeds Q_0 for the year under consideration.

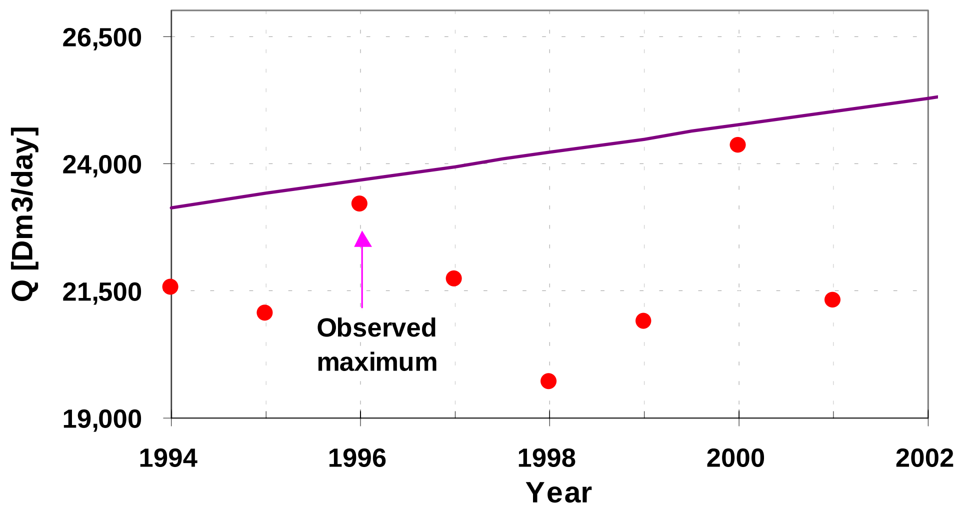
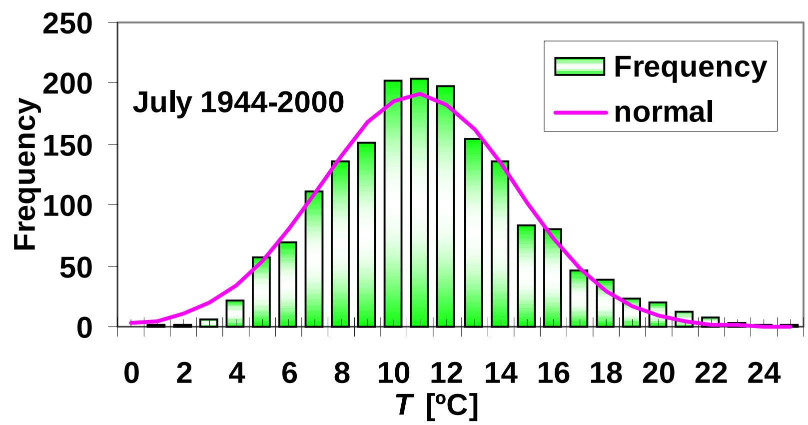


Figure 8: Estimated optimum reserved capacity for the Greater Buenos Aires supplied by MetroGas predicted by the model (line), obtained by requiring that the peak consumption does not exceed this value more that one day in twenty years. The symbols (circles) represent the actual observed maximum consumption for those years.

IV.- **Relation between daily and monthly consumption distribution – Load Factors**

In this section we present a novel procedure that stablishes the relation between the daily and monthly consumption distribution. This association allows us to estimate the load factors, from the information obtained from the monthly billing. Furthermore, this procedure allows us to obtain load factors for different segments of users in a given region. The formalism proposed here is based on the assumption that the daily consumption of natural gas can be described by a model that depends on the temperature as described by the expression (10). We will also assume that for each month of the year, the daily mean temperatures have probability distribution that is well approximated by a normal distribution, with a mean value, T_{month}, that depends on the month of the year and a standard deviation, σ_{month}, that can also depend on the month in discussion. For the major cities of Argentina that we have studied, this assumption in well justified. In figure 10 we present the histogram of daily mean temperatures for the month of July for the Greater Buenos Aires, taking the data of the observed temperatures for this month from 1944 to the year 2000. In this figure we also plotted the corresponding normal distribution. We see that the normal distribution closely approximates actual data. This agreement hlds true for all the months of the year. In Figure 10 we present the distribution of the mean value, T_{month}, as a function of the month for the Greater Buenos Aires using the same data base as in figure 9. The error bars in this plot represent the value of the standard deviation, σ_{month}. In this case, σ_{month} was almost constant (σ_{month}=3.1 ºC ± 0.6 ºC).

Figure 9: Distribution of daily mean temperatures for the month of July for the



GBA. This histogram was obtained using all the observed temperatures for this month for the years 1944 to 2000. The continuous curve is a fit to the data using a normal distribution.

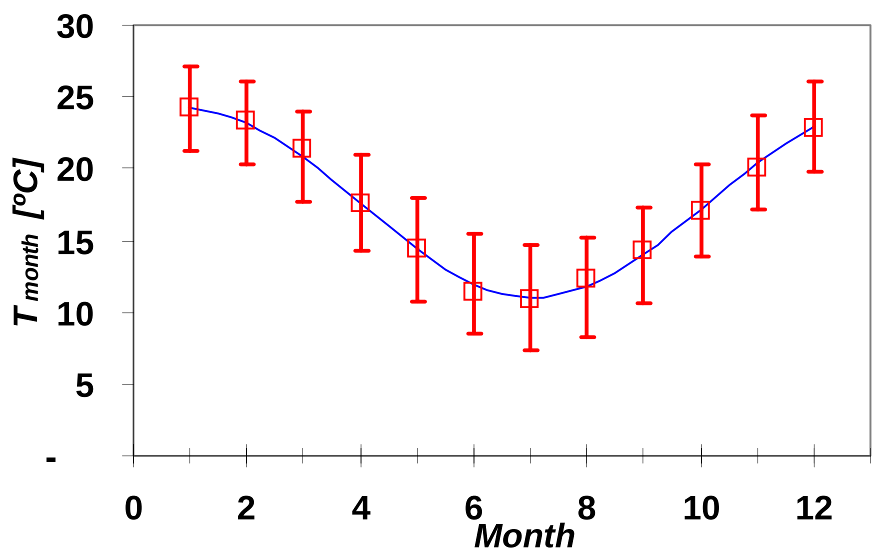


Figure 10: Distribution of the monthly mean temperatures for the GBA, the vertical error bars represent the values of the corresponding standard deviation, σ_{month}. The curve is a fit of the data using the function: $T_{month}=a+b.\cos(c.month+d)$.

If we use the effective temperature instead of the mean temperatures, we obtain the same result. Thus, all the results discussed here can be applied to both distributions: daily mean and effective temperature.

Algorithm: If the daily consumption of gas is described by a model as represented by expression (10) and the daily mean temperatures are well approximated by a normal distribution, with a mean value and a standard deviation given by (T_{month} , σ_{month}), then it is possible to obtain the monthly consumption, represented by the variable Q_{month}. This quantity is the average daily consumption for the month under consideration. Clearly this magnitude can be obtained by taking the weighted average of the daily consumption $Q(T)$, with the corresponding normal weight, *i.e.*

$$Q_{month}(T_{month}) = Q_0 \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_{month}} \cdot \int_{-\infty}^{\infty} \left[1 - f \cdot tanh\left(\frac{T - T_0}{\Delta T}\right)\right] \cdot \left[\exp - \left(\frac{T - T_{month}}{\sqrt{2} \cdot \sigma_{month}}\right)^2\right] \cdot dT \quad (12)$$

The integral in expression (12) can be solved easily using the convolution theorem of the Fourier Transform, see appendix for details. The result is given by:

$$Q_{month}(T_{month}) \cong Q_0 \cdot \left[1 - f \cdot tanh\left(\frac{T_{month} - T_0}{\Delta T_{month}}\right)\right] \quad (13)$$

Here T_{month} represents the average monthly temperature and the parameter ΔT_{month}, that determines the width of the distribution is:

$$\Delta T_{month} = \sqrt{\Delta T^2 + 1.382 \cdot \sigma_{month}^2} \quad (14)$$

The value of Q_{month} is the average daily consumption for the month. The total consumption for the whole month will be the product of Q_{month} by the number of days in the corresponding month. The expressions (13) and (14) indicate that thereis a connection between the daily distribution of consumption and the monthly consumption. This last information can be obtained from the monthly billing.

From the plot of Q_{month} versus T_{month} , it is possible to test the validity of expression (13) and to obtain the parameters of the model (Q_0, f, T_0, and ΔT_{month}). Once the monthly distribution is well characterized, using equation (14) and the known daily temperatures for each day in the month, we can use expression (10) for estimating the daily consumption distribution.

In particular, the daily distribution is useful for estimating the maximum daily consumption for a given year, Q_{max}, which is the crucial parameter for determining the load factor for each segment of users, see expression (2), and optimizing the purchase of reserved capacity. To carry out this estimation, the procedure is the following: from the data of billing we can obtain the parameters that characterized expression (13), then employing equation (14) we obtain the daily consumption distribution. The value of Q_{max} can be calculated from the minimum effective temperature, T_{min}, in the period being studied, using the expression:

$$Q_{max} \cong Q(T_{min}) = Q_0 \cdot \left[1 - f \cdot tanh\left(\frac{T_{min} - T_0}{\Delta T}\right)\right]. \quad (15)$$

From the billing information for the period of time in consideration, it is also possible to obtain the daily average consumption, <Q>, therefore, using equation (2) the load factor for each segment of user can be obtained.

In order to test the validity of the present formalism, we have applied the model to the case of the firm consumption of the region of GBA supplied by MetroGas. For this case the total daily consumption if known form direct measurements. The monthly consumption and the daily temperature are also known. The data of the daily consumption as a function of the effective temperature are shown in Figure 5, together with the fit obtained using expression (10). The data plotted here are the values of daily consumption per user, expressed in term of standard m^3/day, using the information for the years 1996 to 2000. In Figure 11, we present the corresponding monthly consumption, expressed in term of the average daily consumption for the month in consideration (working days) as a function of the average effective temperature for the month. It is useful to point out that when we take the average monthly temperature, the result is basically the same weather if we use the average daily or the effective temperature. In figure 11, we have also include the model prediction obtained using expressions (13) and (14). As can be seen, the agreement between the model and the data in figure 12 is excellent. This type of agreement was also found in the other cities of Argentina where the model was tested. This agreement give support to the formalism presented here. Furthermore, it is clear from figure 12, that the dispersion of the data around the model, is much smaller than in the case of the daily consumption (Fig. 12), therefore it more convenient and robust way of extracting the parameter of the model. The load factor obtained using figure 11 is for this segment of consumers LF=0.43(±0.03). The same load factor, obtained using the monthly distribution is LF=0.45(±0.02), so both results are in agreement within the uncertainties of the calculations.

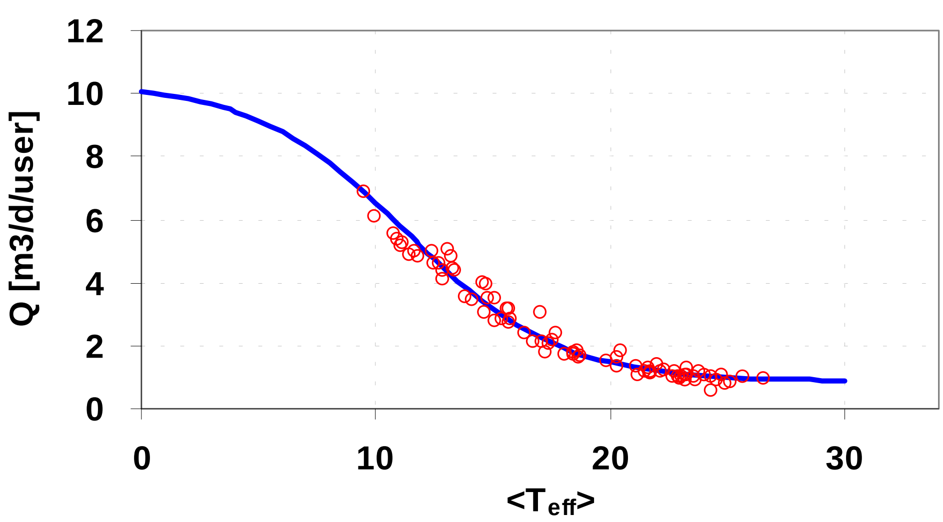


Figure 11. Monthly distribution of consumption (firm component) for the case of GBA supplied by MetroGas (symbols) as a function of the monthly average effective temperature, for all the working days over the years 1996 to 2000 The continuos curve is a fit obtained using expression (13).

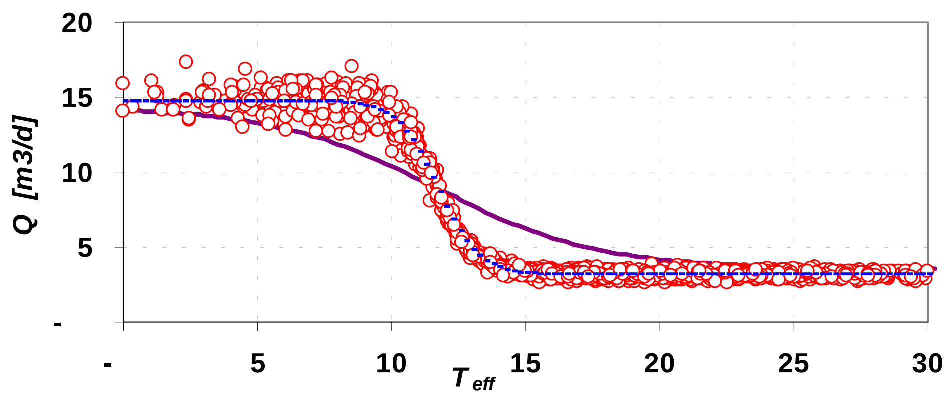


Figure 12. Comparison of daily consumption, obtained by simulation using Monte Carlo Technique, for the case of GBA, as a function of the average effective temperature. The dashed blue curve is the fit obtained using expression (10). The prediction of the monthly distribution is indicated by the continuous curve obtained using expressions (12) and (13).

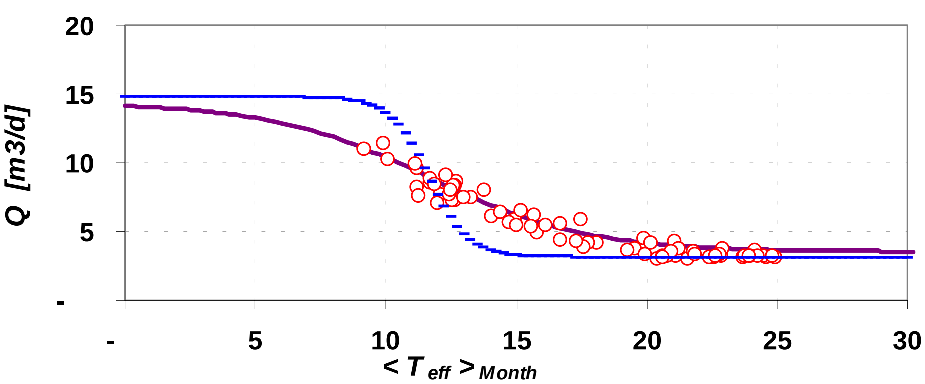


Figure 13. Comparison of monthly consumption (circles), obtained by simulation using Monte Carlo Technique, for the case of GBA, as a function of the average effective temperature. The dashed blue curve id the fit obtained using expression (10). The prediction of the monthly distribution is indicated by the continuos curve obtained using expressions (12) and (13).

We performed a second test of the present formalism to establish the relation between the daily and monthly distribution of consumption using the Method of Monte Carlo[5]. This is a very powerful technique to simulate data with a given probability distribution. Using the observed distribution of consumption for the GBA and the observed temperature distribution, we are able to generate "artifical" daily consumption data with a given degree of dispersion, 15% in our case. Furthermore, this technique allows us to explore unusual scenarios of consumption for which we may not have real data, but which we are interested in exploring. The results of this simulation are presented in figure 12 (symbols), together with the result of the fit obtained using expression (10) (dashed curve). We also show the corresponding monthly distribution obtained using expressions (12) and (13). If we take the monthly average of the data of consumption and plot them as a function of the average effective distribution, we obtain the results presented in Figure 13. We see that the data of the monthly consumption now agree with the prediction of the monthly distribution, thus confirming the validity of the model.

In summary, the validity of the model to relate the daily and monthly distribution of consumption has been established by three procedures. a) By mathematical manipulation of the analytical distribution using the convolution theorem of Fourier transform, b) by direct comparison of real data of consumption, using the firm consumption for the case of GBA and c) by a Monte Carlo simulation of data of consumption applied to the case of GBA. In all the tests performed, we have confirmed the validity of the proposed formalism for establishing the realtion between the daily and monthly distribution.

VI.- **Conclusions**

In this work we have presented a model for the prediction of natural gas consumption at short (2 to 5 days) and intermediate range (1 to 5 years). The model can be applied to predict the consumption of different segments of consumption and is also useful to predict the maximum consumption in the intermediate range. This information is useful to adapt the infrastructure of transportation as well to estimate optimum reserved capacity. The predictions of the model, have been successfully applied to the all the major cities of Argentina. The short range prediction agrees with the observed consumption within 10% on 90% of the days. The uncertainties of the intermediate range prediction are in the order of 12%.

The model presented here, reveals a useful association between the daily and monthly distribution of consumption. This relation is useful to obtain, very economically, the daily distribution of consumption for the different segments of users in a given region. It also allows us to estimate the peak consumption in the intermediate range and to extract the load factors from the monthly billing information. This alternative is very attractive economically, since it use does not require costly additional measurements.

The views and opinions of authors expressed herein do not necessarily state or reflect those of their employers or the ENARGAS.

We would like to acknowledge the useful comments and suggestions of Ing. L. Pomerantz, Ing. L. Duperron and Dr. M. Schwint.

Appendix

In this appendix we present a mathematical justification of the relation between the daily and monthly consumption distribution, i.e. we derive expression (13) and (14) from (12).

As we have discussed in this work, the distribution of consumption can be modeled using the hyperbolic tangent function(expression (10)), i.e.

$$y_1(T,T_0,\Delta T) = tanh\left[\frac{T-T_0}{\Delta T}\right] \qquad \text{(A1)}$$

The behavior of this function is similar to that of the Error Function *Fer(t)* defined as [6,7]:

$$Fer(t) = \frac{2}{\sqrt{\pi}}\cdot\int_0^t e^{-u^2}\cdot du \qquad \text{(A2)}$$

In fact the function:

$$y_2(T,T_0,\sigma_T) = Fer\left[\frac{T-T_0}{\sigma_T}\right] \qquad \text{(A3)}$$

with

$$\sigma_T = \delta\cdot\Delta T \qquad \text{(A4)}$$

and

$$\delta = 1.203\,, \qquad \text{(A5)}$$

has very similar characteristics to the function described by expression (A3) as can be seen in figure A1. In this figure, we have plotted both function referred to the left axis, and their derivatives, referred to right vertical axis. The maximum difference between these functions is in the order of 0.04%. Therefore, the function $y_2(T)$ is a good approximation to $y_1(T)$. In fact, either of these two functions could have been used to model the consumption distribution of natural gas. Our choice of expression (A1) was adopted on a discretional basis and for calculation convenience. Nonetheless, for some applications it may be more convenient to use expression (A2) to describe the consumption distribution. In fact if the distribution of consumption can be described equivalently by expression (10) or by the function:

$$Q_2(T) = Q_0\cdot\left[1 - f\cdot Fer\left(\frac{T-T_0}{\sigma_T}\right)\right] \qquad \text{(A6)}$$

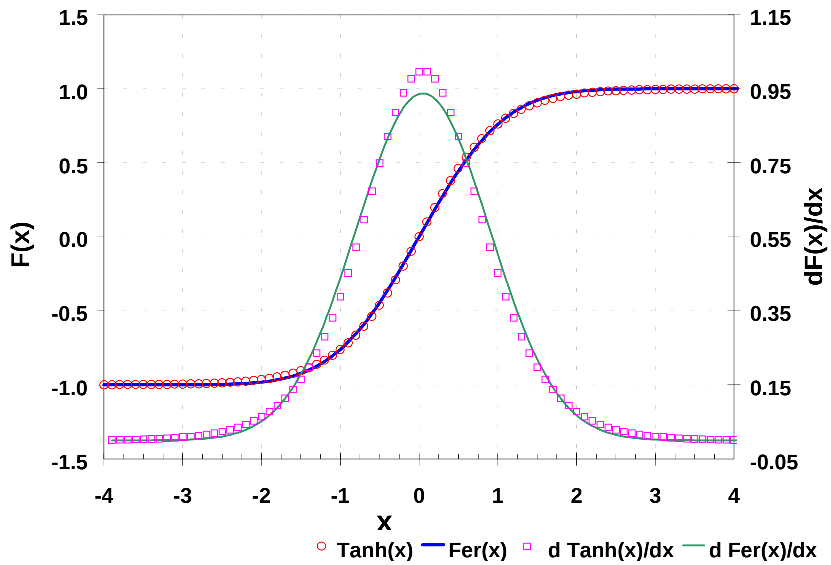


Figure A1. Comparison of distribution (A3) (circles) and (A4) continuous curve, both referred to the left vertical axis. In this figure we also show their respective derivatives referred to the right vertical axis.

Replacing this distribution in expression (12), making the following change of variable:

$$t = \frac{T - T_0}{\sigma_{Ts}} = \frac{T - T_0}{\delta \cdot \Delta T} \qquad \text{(A7)}$$

expression (12) becomes:

$$Q_{month}(T_{month}) = Q_0 - Q_0 \cdot f \cdot \frac{\sigma_T}{\sqrt{2\pi} \cdot \sigma_{month}} \int_{-\infty}^{\infty} \left(e^{-\left(\frac{t-x}{\sqrt{2}\cdot\sigma}\right)^2} \right) \cdot [Fer(t)] \cdot dt \qquad \text{(A8)}$$

Where

$$x = \frac{T_{month} - T_0}{\sigma_{Ts}} \qquad \text{(A9)}$$

and

$$\sigma = \frac{\sigma_{month}}{\sigma_T} \qquad \text{(A10)}$$

The integral of the second term of (A8) can be calculated using the techniques of Fourier transform, since it is the convolution of a *Fer(t)* function with a nomal distribution[6,7]. If we take the Fourier transform on both sides, using the known properties of the Fourier Transform [6,7], we have:

$$\Im\{I(x)\} = \Im\left\{\frac{1}{\sqrt{2\pi}\cdot\sigma_s}\int_{-\infty}^{\infty}\left(e^{-\left(\frac{t-x}{\sqrt{2}\cdot\sigma}\right)^2}\right)\cdot\left[Fer(t)\right]\cdot dt\right\} =$$

$$= \Im\{Fer(t)\}\cdot\Im\left\{\frac{1}{\sqrt{2\pi}\cdot\sigma_s}\cdot e^{-\left(\frac{t}{\sqrt{2}\cdot\sigma}\right)^2}\right\} = \frac{2}{i\cdot\omega}\cdot e^{-\frac{\omega^2}{4}}\cdot e^{-\frac{\omega^2\cdot\sigma^2}{2}} =$$

$$= \frac{2}{i\cdot\omega}\cdot e^{-\frac{\omega^2}{4}\cdot\left(2\sigma^2+1\right)}.$$

(A11)

If we anti-transform this expression, we obtain:

$$I(x) = Fer\left(\frac{x}{\sqrt{1+2\sigma^2}}\right) = Fer\left(\frac{T_{month}-T_0}{\sigma_T\cdot\sqrt{1+2\sigma^2}}\right). \quad \text{(A11)}$$

If we define the parameter ΔT_{ef} as:

$$\Delta T_{ef} = \frac{1}{\delta}\cdot\sqrt{\sigma_T^2+\sigma_{month}^2} = \sqrt{\Delta T^2+\frac{2}{\delta^2}\sigma_{month}^2}\,, \quad \text{(A12)}$$

the monthly distribution (12) can be written as:

$$Q(T_{month}) \cong Q_0\cdot\left[1-f\cdot tanh\left(\frac{T_{months}-T_0}{\Delta T_{eff}}\right)\right]. \quad \text{(A13)}$$

This last two results are the expression we wanted to prove. Furthermore, if we define the parameter ϕ as:

$$\phi = \frac{T_{month}-T_0}{\Delta T_{eff}}. \quad \text{(A14)}$$

The expression $Q_{month}(T_{month})$ as a function of ϕ becomes:

$$Q_{month}(\phi) = Q_0\cdot\left[1-f\cdot\phi\right]. \quad \text{(A15)}$$

With this substitution it is straightforward to extract the parameter from the model Q_0 and f for any segment of consumers by simply plotting Q_{month} as a function of ϕ_{month}. Here we assume that the other parameters of the model σ_{month}, ΔT and T_0 for a given region of study, can be known from a preliminary analysis of the data. Furthermore, they are expected to be the same for all the segments of consumers in a given region. This has been the case for all the cities of Argentina that we have studied.

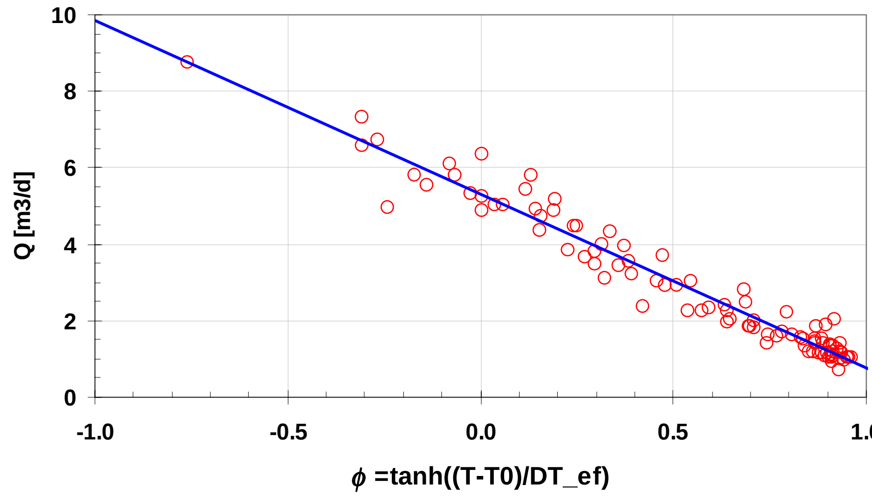


Figure A2. Representation of monthly residential consumption (MetroGas 1996-2000) as a function of the parameter ϕ. The line is a linear fit to the data.

‡ *Escuela de Ciencia y Tecnología - Universidad Nacional de San Martín Buenos Aires y Departamento de Física de la F.C.E. y N. de la Universidad de Buenos Aires - Argentina.* - *e-mail:* sgil@df.uba.ar

References

[1] Naturaql Gas Regulatory Framework (Marco Regulatorio del Gas Ley 24.076 de la Nación Argentina) - *www.enargas.gov.ar*

[2] *Modelo de Predicción de Consumo de gas natural en la República Argentina*. S.Gil y J. Deferrari. **Petrotecnia** (Revista del Instituto Argentino del Petróleo y del Gas) **XL**, N° 3, Sup. Tecn. 1,1 - Junio(1999).

[3] *Latin American and Caribbean Gas and Electricity Congress* - Punta del Este - Uruguay - March 2001. *www.iapg.org.ar*

[4] *Climate Change projections hinge on global CO_2, temperature data* – T. H. Standing – Oil & Gas Journal Nov. 12, 20, 2001

[5] *Física* re-Creativa – S. Gil y E. Rodríguez – Prentice Hall – Buenos Aires 2001

[6] *Applied Matematics for engineers and Scientists* – M. R. Spiegel – McGraw Hill – NY 1958

[7] *Applied Fourier Analysis* –Hwe P. Hsu – Int. Thompson Pub. Co. NY 1984