Analysis and clustering of natural gas consumption data for thermal energy use forecasting

Alessandro Franco, Fabio Fantozzi

Department of Energy, Systems, Territory and Constructions Engineering University of Pisa, Largo Lucio Lazzarino, 56126 PISA

E-mail: alessandro.franco@ing.unipi.it

Abstract. In this paper, after a brief analysis of the connections between the uses of natural gas and thermal energy use, the natural gas consumption data related to Italian market are analyzed and opportunely clustered in order to compute the typical consumption profile in different days of the week in different seasons and for the different class of users: residential, tertiary and industrial. The analysis of the data shows that natural gas consumption profile is related in an important way to seasonality pattern and mainly to the outside temperature, but there is also an important daily pattern both related to industrial and civil energy consumption that, at a lower degree than the previous one, does affect the consumption profile and have to be taken into account for defining an effective short and mid term thermal energy forecasting method. This information extracted appears to be interesting not only in the perspective of defining a model of energy use based on both physical and statistical elements but also to support energy providers with some information that can be used to tailor bills according to actual energy consumption profiles.

1. Introduction

The development of provisional instruments for the short and mid term prevision of the patterns that govern the energy demand can permit significant energy savings [1]. Accurate models for electric and thermal energy short and mid term forecasting are essential to the operation and planning of a utility company and to increase the efficiency in the energy production and distribution, [2].

Forecasting of the thermal energy use is a quite complex task: an important amount of the thermal energy consumed is caused by the heating of civil/tertiary sector, but the part relating to industrial uses, such as can be for example the process steam, is important too. While the first component can be directly correlated to climatic variables (mainly temperature, humidity and wind speed) and seasonality pattern, the second component is more directly related to the planning of production (day of week) and less dependent on meteo data. The geographic aggregation of the data appears to be important too. The development of forecasting method for electricity can be based on available data of consumption and electricity load, [3-4] the same cannot be so simple for the thermal energy. Peak energy demand is surely a critical element in the short, mid and long term which for a country like Italy, consists of electricity (the mix and spatial distribution of generation technologies, transmission, and distribution) and natural gas (production, transmission, distribution, and storage), [5-6].

Some important drivers like the use of natural gas for thermal energy production and the diffused use of combined cycle power plants using natural gas have altered the traditional cyclical demand of

natural gas. Another important factor is the cost and availability of natural gas, which is itself related to many political and economical factors, many times not fully predictable. The economic factor has an influential impact on the demand for natural gas in both the residential and in the industrial sector, and in both cases it is not simple to quantify the short, mid and long term impact in gas consumption. The analysis of the drivers of natural gas consumption is currently object of research, as suggested by a recent paper, [7], but it is clear that in a particular country like Italy, the use of thermal energy use is the most important driver. Furthermore forecasting of natural gas consumption has been already object of analysis in the literature, for example by [8] and [9] with a general perspective. The residential and non residential consumption of natural gas in Italy has been recently object of analysis in the literature. In particular Bianco et al. in [10-11] proposed a model to forecast residential and non-residential consumption of natural gas. Consumption drivers are identified and discussed and a single equation demand model is identified. Anyway the object of the two papers is the obtainment of long term scenario and the effect of economic drivers is considered of primary importance. It is important to remark that the renovation of the natural gas wholesale and retail markets together with a technological review of current smart energy practices is a key step to accomplish an optimal energy consumption pattern, to avoid energy wastes and to implement a sustainable energy system. In this vision smart meters installed at the utility endpoints to monitor usage could be an important component of the smart grid. As discussed in the previous section, high resolutions automated meter reading system for residential gas meters, which can be used to record gas consumption for each appliance have been already experimentally tested and applied, even if in most cases they are not in the commercial development phase yet ([12]). Research on smart meters appears to be interesting from a technological point of view, but the development and application of smart metering is a particularly complex task mainly for aspects related to data treatment and communications. The true challenge and difficult task in the field of smart meters' development, which also occurs in the use of Automatic Meter Reading (AMR) devices, appears to be the fact that monitoring energy usage data poses privacy risks that might not be easy to solve, [13-14]. On the other hand, fine-grained energy consumption data collected by AMR could reveal sensitive information coming from home (number of occupants, vacancies, typical energy usage, etc.). While there is a general consensus that energy monitoring devices can be an opportunity to "provide consumers with a more comprehensive and nuanced understanding of their usage patterns", the only real perspective of considering demand side management is the gas flow monitoring, [15]. In this case even if there has recently been a lot of research in this topic, [12] and the italian distributors were required to install SGMs at 100% of all non-domestic consumers' premises by the end of 2012 and at 80% of domestic consumers' premises by the end of 2016, [16], the availability of data is limited to integral data. This work represent the first step of the construction of the model for the definition of thermal energy use consumption in complex and largely extended energy systems. Results can be extended in some directions that are currently under investigation by the authors; for example to other countries to check if the same features that appeared to be particularly informative in the Italian case are equally important in other countries.

2. The link between natural gas consumption and thermal energy use: the Italian case

Natural gas has been used since the '70s to satisfy the needs of many commercial and residential users throughout the world through a huge and complex network. Demand for natural gas has traditionally been cyclical, simply following the characteristic seasonal patterns. As a consequence, the cyclical nature of the demand was used in the past to accurately predict the required natural gas: demand was highest during the coldest months in winter and lowest during the warmest months of the year in summer; also, a direct correlation with average medium environmental temperature could be observed. In addition, until a few years ago, the connection between the gas sector and the electricity sector was marginal. Things began to change at the beginning of the '90s, when the commercial development of combined cycle power plants led to a shift towards the use of natural gas for the generation of electricity as well. This had a strong impact on the traditional cyclical gas demand. Obviously, while demand of natural gas heating decreases during the summer months, the demand for air

conditioning/cooling is relevant during the same period. Nowadays electricity provides the primary source of energy for residential and commercial cooling, thus causing a significant increase in electricity demand especially in summer months. As a further complication to this scenario, in the last few years there has been a consistent increase in the penetration of intermittent renewable energy production, particularly in countries like Italy, Denmark and Germany. In turn, this has implied that gas is being used as a back-up for electricity production when renewables are not available (e.g., to back-up solar plants at night time), thus contributing to tighten the connections between gas and electricity. Due to the particular development in the 80 and 90s of thermal energy systems based on the use of natural gas, Italy represents a particular case where the use of thermal energy could be correlated with the consumption of natural gas. Only in recent years, however, the gas sector was heavily influenced by new constraints emerged with this new integrated vision system of the energy systems. The significant penetration of RES in electricity production has imposed specific logic to the consumption of natural gas in thermoelectric plants, mainly natural gas combined cycle power plants. The topic is discussed by one of the author in [17].

2.1. Database definition

For the Italian case, daily gas flows of 2012 are available from the database of SNAM network national splitted to intended use (Fig. 1), [18]; the raw data set, containing data as-is from the Snam Rete Gas database has been used. The weather is represented by the average temperature data set, obtained from a specific database, [19]. Fig. 1 provides the volume of gas natural consumption in Italy in 2012, to support the previous discussion regarding the difficulty of predicting natural gas demand. The natural gas consumption data are represented in aggregated form and in the three main customer components (gas for thermoelectric plants, gas for industrial sector and gas for civilian/urbane sector). The effect of season is represented by an average daily value of the temperature registered in the observed day in a particular town of Central Italy. Fig. 1 shows the strong fluctuation of the natural gas demand, expressed in Mm^3 (1 $Mm^3 = 10^6 m^3$), taking into account the main sectors of use. As can be seen in Fig. 1, gas for heating varies in a significant way depending on the time of the year (cold vs. warm seasons); but gas for thermoelectricity remains of the same order of magnitude over the year, and fluctuates according to other dynamics (e.g., weekly patterns). Overall, it is not trivial to predict accurately the natural gas demand. While natural gas consumption in power plants is linked to the electricity market, because it is directly referred to the electricity generated in gas fuelled power plants and is not dependent on thermal energy use, the other components have major correlation.



Fig. 1. Consumption of natural gas in Italy in 2012



Fig. 2. Consumption of natural gas in civil sector in 2012

Residential consumption of natural gas is determined by the demand for heating, sanitary wter and cooking facilities in residential buildings. In this case the main consumption drivers are represented by external temperature of the heating season, population and buildings characteristics (i.e. insulation, facilities, etc.). Nonresidential natural gas consumption represents the usage of natural gas related to economic activities, particularly industry (i.e. manufacturing, food, construction, etc.) and services (i.e. offices, shops, etc.). Concerning natural gas consumed in the industrial sector is mainly used in production processes, even though a substantial share difficult to estimate is also utilized for heating purposes (i.e. heating of very large industrial buildings), whereas the consumption in the service sector is substantially due to heating demand. A clear link between natural gas consumption and average outside temperature can be remarked. But the influence of other elements, like the day of the week (in particular the difference between working days and holidays) cannot be neglected. The civilian/urbane consumption is the share of gas flowing in the network that is used to heat the civil sector (residential and tertiary). By making an initial clustering of the data between the various sector, analyzing the trend of the civilian consumption during the year (Fig. 2) it is possible to clearly identify the dependence of the structure of the demand not only on the level of the outside temperature.

2.2. Preliminary analysis of the data

Fig. 1 shows clearly that the natural gas consumption, directly related to heat load and thermal energy consumption is clearly correlated to seasonality patterns. By visual analysis of the data if is possible to propose two possible clusters related to two different meteorological seasons: the "cold" season (starting from the middle of October to the end of April) in which the natural gas consumption can be clearly linked to the average external temperature and the "warm" season (from May to middle of October), in which the consumption is not particularly influenced by the outdoor temperature. Analyzing the aggregated data of natural gas consumption it can be also evidenced that the gas consumption data, both in the "cold" season and in the "warm" season rely on the specific type of the day, e.g., working days, vs. holidays, vs. pre-holidays, two weeks of August. The gas consumption is smaller on Saturdays (some industries and some activities are closed) and even smaller on Sundays.

3. Data analysis and clustering of natural gas consumption data: research of peculiar patterns

Classification and clustering of time series date is recognized as an important area of research for energy data both considering demand and production. Clustering refers to the ability to aggregate similar data together and the basic clustering operation corresponds to take a set of a given number n of objects and/or data and group them into k clusters. A good clustering method has predictive power;

in this case: clusters can be used to predict future trend of the data, for example concerning energy consumption. Moreover clusters allow to compress a great number of information into a single or a reduced number of information. In addition clusters permits to identify the "outliers", i.e., the cases in which clusters fail to accurately represent particular data. The objective of this section is to propose possible clustering of natural gas consumption data. Analyzing natural gas consumption data, the seasonality effects connected with the average temperature can be clearly observed with sensibly higher consumption in winter, lower consumption values in autumn and spring and the lowest values in summer. But analyzing the data it is also interesting to distinguish the various components of natural gas consumption, in particular referred to industrial load, civil/residential load and thermoelectric load and the correlation with the temperature is different in the three cases. Also, three main types of energy consumption users can be identified. The first type (majority of users) is characterised by the domestic consumers who consume a significant amount of energy also during the night time. The second type of consumers is characterised by the tertiary activities, having a quasi-flat consumption profile during the day. Finally, the third type involves industrial use of natural gas, both considering thermal energy and other industrial uses.

3.1. Detailed analysis of the data

Concerning the residential usage of natural gas, the analysis of the consumption data shows that the consumption can be particularly referred to seasonality pattern: this as can be evidenced by the strong link with the average external temperature, mainly when the temperature is below 15 °C. However different profiles in different days of the week can be also evidenced. It can be easily seen by visual inspection that natural gas consumption data rely on the specific type of the day, e.g., working days, vs. holidays, vs. preholidays. Clearly, the consumption is smaller on Saturdays, due to the fact that some activities (like offices and schools) are closed and even smaller on Sundays (a lot of commercial activities are closed too). This suggests that the load can be splitted into three clusters corresponding to the category of the day of the week (e.g., working days, Saturdays and Sundays). In case of industrial users, by classifying load consumption into two categories (working days and holidays), it is possible to identify two clusters while the natural gas consumption appears to be less connected with the outside temperature. On the other hand, visual inspection of the aggregated natural gas consumption data concerning the civil sector, reported in Fig. 2, suggests that the consumption, for a well defined value of the outside temperate load in this case too is different depending on whether the day is a working day or not. But the difference between the two days is less remarkable than for industrial sector. More generally, it is possible to recognise whether a given datum belongs to a holiday (i.e., Sundays, or other Italian holidays, e.g., Christmas day), to a working day, or to a preholiday. With the term preholidays it is defined the set of Saturdays, the days within two holidays when most offices are closed, and the working days during conventional Italian summer and Christmas holidays (i.e., the two weeks around August 15 and the two weeks between Christmas and January 6). Fig. 3 and 4 represents the link of daily consumption of natural gas and average outside temperature in the weekdays and during the other days, grouping preholidays, holidays and special days, according to the detailed description reported in Fig. 5. The data on the horizontal axis represents the progressive number of the day in accordance with the description of Fig. 5, including all holidays, pre-holidays, working days and special days in Italy in the 2012 calendar. Considering the aggregated data analyzed, a multiplicative model seems to be the useful in order to predict the natural gas consumption in the civil sector in Italy. The mathematical formulation will be the following:

$$C(t) = F(d(t) \cdot f(w(t)) + R(t)$$
(1)

where C(t) is the actual consumption at time t, d(t) is the day of the week, F(d) is the daily component, w(t) is a function of the weather data that include the temperature, humidity and wind chill, f(w) is the weather function, and R(t) is a term of correction (random). This last term reflects the fact that thermal energy uses depends not only on the current weather conditions but also on the weather during the

previous days. In particular special conditions occurs when the cold weather continues for several days or when a cold day arrives after some "hot" days. Analysing the consumption of natural gas referred to industrial sector (Fig. 6), it is possible to evidence the reduced dependence on the temperature and a most important influence of the day of the week: a remarkable difference between working days and holidays can be clearly evidenced. Elasticity values with respect to the temperature have been estimated, for the natural gas consumption in industrial sector. Fig. 7 reports the link between the daily consumption and the temperature in the weekdays, while Fig. 8, 9 and 10 provides the same data for the three different types of days: pre-holidays, holidays and special days of the year 2012. The days are defined using the reference data of Fig. 5. As it can be evidenced from the various figures, there is a basic consumption level and an additional amount directly linked to the temperature. It is possible to conclude this analysis putting in evidence that if in the case of natural gas consumption in the civil/residential sector, the link with weather conditions and temperature is clearly evident (the difference between the minimum level is remarkable, going from 20-30 Mm³ for each day up to 300 Mm³ for each day in the coldest period of the year), the industrial sector appears to be different.



Figure 3. Natural gas consumption data during the days of the week



Figure 4. Natural gas consumption data during preholidays, holidays and special days

!!N po!	2!	3!	4!	5!	6!	7!	8!	9!	:!	21	22!	23!	24!	25!	26!	27!	28!	29!	2:!	31!	32	33	34!	35!	36!	37!	38!	39!	3:!	41!	42!
Kbo!									•	•	•	•	•	•		•	•	•!	•	•			•	•	•	•	•			•	•
Gfc!	•	•	•	•!		•	▲!	▲!		•	•					•		•!		•	•	٠	•	•			•	•		!!	!
N bs!	•	•	•!		•	•	•!	•	•	٠		•!					•		•	•	•	٠	•	•		•	•	•			•!
Bqs!		•	•	•	•	•!		▲!		•	•	•	•	•		•	•	•!	•	•	•		•	•		•	•	•		•	!
N bz!		٠	•!	•	•!		•!	•!	•	•	•	•!			•	•	•	•	•		•	٠	•	•	•	•		•	•		•!
Kvo!	•			•	•	•	•!	•	•!		•	•!				•		•!	•	•	•	٠	•		•	•	•	•		•	! !
Kvnh		•	•	•	•	•	•!		•	•	•	•!		•				•!	•	•	•			•	•	•	•	•			•!
Bvh!	•	٠	•!	•!		•!	•!	•!	•	•								▲!		•	•	٠	•	•	•		•	•	•	•	•!
Tfq!	•		•	•	•	•	•!	•!		•	•	•!			•			•!	•	•	•	٠		•	•	•	•	•	•		!
P du!	•	•	•	•	•	•!		•	•	•	•	•!	•					•!	•	٠		•		•	•	•	•				•!
Opw!		٠	•!		•	•	•!	•!	•	•		•					•		•	•	•	•		•		•	•	•			!
Efd!	•		•	•	•	•	•!			•	•	•			•			•!	•	•	•	•									. ▲ !

Figure 5. Working days (circle), preholidays (romb), holidays (square), special days (triangle) during 2012 in Italy.

As it can be clearly evidenced in Figs 7-9, a minimum level of 20 Mm^3 for each day is consumed in each day of the year, while an additional amount of maximum 30 Mm^3 for each day can be related to the temperature and to the day of the week. Analyzing the data of natural gas consumption in the industrial sector, an additive model can be proposed for the detailed description: this takes the following form:

$$C = C_n + C_w + C_d + C_R \tag{2}$$

where C is the total consumption, C_n represents the "baseload" component, which is a set of standardized load shapes for each "type" of day that has been identified as occurring throughout the year, C_w represents the weather sensitive part, C_d is a special day component that create a substantial deviation from the usual load pattern and C_R is a completely random term, the typical noise term.

3.2. Discussion, use of the present analysis and possible structure of thermal energy use forecasting The analysis of the aggregated natural gas data consumption permits some interesting considerations on the use of thermal energy and for the definition of models for short and mid term forecasting.



Figure 6. Natural gas consumption in the industrial sector



Figure 7. Natural gas consumption trend in industrial sector during the weekdays



Figure 8. Natural gas consumption trend in industrial sector during the holidays



Figure 9. Natural gas consumption trend in industrial sector in the pre-holiday days



Figure 10. Natural gas consumption trend in industrial sector in the pre-holiday days

In principle it seems that both for natural gas consumption in Italy (strongly linked to the thermal energy end uses) and for thermal energy use in general, a combination of end-use model and econometric model can be proposed as the more appropriate one. The end-use approach directly estimates energy consumption by using extensive information on end use and end users. Concerning the civil/residential sectors information about sizes of houses, appliances, the customer use, their age, and so on, are required. Statistical information about customers along with dynamics of change is another additional element for the forecast. In general, weather conditions influence in maximum part both the load and the natural gas consumption. It is clear that forecasted weather parameters are the most important factors for load forecasting. So, the accuracy of load forecasting depends on the accuracy of forecasted weather scenarios. But the time factors including the time of the year and the day of the week (mainly the day type) and in principle the hour of the day are quite important too. While it is really difficult to elaborate hourly models, as observed analyzing the natural gas consumption data, some remarkable differences can be evidenced between weekdays and weekends, mainly in the industrial sector. For this reason, a combination of an additive and multiplicative models can be used in order to develop a model for thermal energy use and heat load forecasting. In addition, economic factors such as per capita incomes can influence the thermal energy use and natural gas consumption. For example natural gas prices can be included in econometric models.

4. Conclusions

This paper reports the first step of the development of a method for the forecast of thermal energy use of a complex system. After a preliminary analysis of the forecasting of energy uses and the discussion of some of the proposal obtained by the technical literature. Considering the particular link of thermal energy and natural gas consumption of the work is focused on a specific objective: to analyze and describe the structure of natural gas consumption and the correlation of those data with physical and statistical elements like a given period of the year, a given day in the week and a given customer category. Then the consumption data of one year, in particular 2012, are completely analyzed. The research of some peculiar patterns in natural gas consumption data, are organized and clustered; the analysis provides interesting information regarding a typical energy consumption in a given day of the week or in a given season. It also shows the particular link of the natural gas consumption in civil/residential sector represents the main component during cold season and is directly connected with the temperature, while the natural gas consumption in the industrial sector is only partially connected with the temperature and part with the day of the week. The authors believe that this

information can be used both to summarise the thousands of profile data in a single daily data (e.g., the centroid of the corresponding cluster), and also to support energy providers with some information that can be used to tailor bills according to actual energy consumption. This work represents just the first step of the construction of a model, conceived for the definition of thermal energy use consumption in complex and largely extended energy systems. It can be extended in some directions that are currently under investigation by the authors. Results will be extended to other countries to check if the same features that appeared to be particularly informative in the Italian case are equally important in other countries. The results of the present analysis will be useful for the development of a model for thermal energy forecasting, whose basic elements are briefly introduced in section 4.2.

References

- [1] Suganthi L and Samuel Anand A 2012 Energy models for demand forecasting—A review. Renewable and Sustainable Energy Reviews 16, 1223–1240.
- [2] Huang Z Yu H Peng Z and Zhao M., 2015. Methods and tools for community energy planning: A review. Renewable and Sustainable Energy Reviews 42, 1335–1348.
- [3] Sandels C Widén J and Nordström L 2014 Forecasting household consumer electricity load profiles with a combined physical and behavioral approach. Applied Energy 131, 267–278.
- [4] Ardakani F J and Ardehali M M 2014 Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. Energy 65, 452-461
- [5] Schellong W 2011, *Chapter 5: "Energy Demand Analysis and Forecast*", Energy Management Systems, Dr Giridhar Kini (Ed.), ISBN: 978-953-307-579-2.
- [6] Hernandez L, Baladron C, Aguiar J M, Carro B, Sanchez-Esguevillas A and Lloret J, 2014 Artificial neural networks for short-term load forecasting in microgrids environment. Energy 75, 252-264.
- [7] Dilaver O, Dilaver Z and Hunt L C 2014 What drives natural gas consumption in Europe? Analysis and projections. Journal of Natural Gas Science and Engineering 19, 125-136.
- [8] Sabo K, Scitovski R, Vazler I and Zekic´-Sušac M., 2011. Mathematical models of natural gas consumption. Energy Conversion and Management 52, 1721–1727.
- [9] Soldo B, 2012. Forecasting natural gas consumption. Applied Energy 92, 26–37.
- [10] Bianco V, Scarpa F and Tagliafico L A, 2014 Analysis and future outlook of natural gas consumption in the Italian residential sector. Energy Conversion and Management 87, 754–764.
- [11] Bianco V, Scarpa F and Tagliafico L A, 2014 Scenario analysis of nonresidential natural gas consumption in Italy. Applied Energy 113, 392–403.
- [12] Tewolde M, Longtin J P, Das S R and Sharma S, 2013 Determining appliance energy usage with a high-resolution metering system for residential natural gas meters. Appl. Energy 108, 363-372.
- [13] NIST U.S., 2010 Guidelines for smart cyber security: vol. 2, privacy and the Smart Grid The Smart Grid, NISTIR 7628.
- [14] Rouf I Mustafa H Xu M Xu W Miller R and Gruteser M 2012. Neighborhood watch: security and privacy analysis of automatic meter reading systems, In: Proceedings of the 2012 ACM conference on computer and communications security. Raleigh, North Carolina, USA: 462–473.
- [15] Kerr R and Tondro M 2012 Residential feedback devices and programs: opportunities for natural gas. US Department of Energy, Energy Efficiency & Renewable Energy – Report 2012.
- [16] Di Castelnuovo M and Fumagalli E 2013 An assessment of the Italian smart gas metering program. Energy Policy 60, 714-721.
- [17] Franco A and Salza P 2011 Strategies for optimal penetration of intermittent renewables in complex energy systems based on techno-operational objectives. Renewable Energy 36, 743-753.
- [18] Snam Rete Gas 2015 The gas transportation in Italy: day ahead capacity booking, http://www.snamretegas.it/en/services/Thermal_Year_2012_2013/Transportation_capacity/
- [19] Il Meteo. 2015 http://www.ilmeteo.it/portale/archivio-meteo/ [in Italian].