

**Data Analysis of Heating Systems for Buildings
A Tool for Energy Planning, Policies and Systems Simulation**

Noussan, Michel; Nastasi, Benedetto

DOI

[10.3390/en11010233](https://doi.org/10.3390/en11010233)

Publication date

2018

Document Version

Final published version

Published in

Energies

Citation (APA)

Noussan, M., & Nastasi, B. (2018). Data Analysis of Heating Systems for Buildings: A Tool for Energy Planning, Policies and Systems Simulation. *Energies*, 11(1), Article 233.
<https://doi.org/10.3390/en11010233>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright


Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Article

Data Analysis of Heating Systems for Buildings—A Tool for Energy Planning, Policies and Systems Simulation

Michel Noussan ¹  and Benedetto Nastasi ^{2,*} 

¹ Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; michel.noussan@polito.it

² Department of Architectural Engineering & Technology, Environmental & Computational Design Section, TU Delft University of Technology, Julianalaan 134, 2628BL Delft, The Netherlands

* Correspondence: benedetto.nastasi@outlook.com

Received: 6 December 2017; Accepted: 16 January 2018; Published: 18 January 2018

Abstract: Heating and cooling in buildings is a central aspect for adopting energy efficiency measures and implementing local policies for energy planning. The knowledge of features and performance of those existing systems is fundamental to conceiving realistic energy savings strategies. Thanks to Information and Communication Technologies (ICT) development and energy regulations' progress, the amount of data able to be collected and processed allows detailed analyses on entire regions or even countries. However, big data need to be handled through proper analyses, to identify and highlight the main trends by selecting the most significant information. To do so, careful attention must be paid to data collection and preprocessing, for ensuring the coherence of the associated analyses and the accuracy of results and discussion. This work presents an insightful analysis on building heating systems of the most populated Italian region—Lombardy. From a dataset of almost 2.9 million of heating systems, selected reference values are presented, aiming at describing the features of current heating systems in households, offices and public buildings. Several aspects are considered, including the type of heating systems, their thermal power, fuels, age, nominal and measured efficiency. The results of this work can be a support for local energy planners and policy makers, and for a more accurate simulation of existing energy systems in buildings.

Keywords: open energy data; thermal systems; conventional and condensing boilers; natural gas; urban energy planning; building energy efficiency; space heating

1. Introduction

Buildings represent a major share of the total energy consumption around the world. Multiple drivers are influencing the energy demand of buildings, and the trends show that the total energy demand remained the same in the last few decades on a world basis, but, with a significant increase in the quality of services [1]. On the other hand, in several countries, multiple policies are moving towards the promotion of energy efficiency measures in buildings [2], thanks to the refurbishment of existing buildings [3], high requirements for new constructions [4] and operational optimization of buildings' management [5]. In this framework, the evolution of ICT allows to significantly increase the amount of data that is collected for monitoring energy systems and for assessing other building-related performance indicators. The availability of live data with high temporal resolution from smart meters allows developing advanced models for the optimization of buildings' energy performance [6]. Smart meters are also fostering an enhanced awareness of the users towards their energy consumption and the actual effect of some energy efficiency measures [7,8].

As a matter of fact, the availability of large datasets including information on a large number of buildings has proven to be a crucial support for the development of energy planning, local policies and citizens' behaviors [9]. Municipal, Regional and National databases have been developed for the classification of the information obtained through different procedures (e.g., buildings certifications, heating plant tests, etc.). However, the creation and use of large datasets require strict standards for the data collection, as well as proper algorithms for data preprocessing and analysis [10].

There is still a gap to be filled on data quality and access: different administrative levels and stakeholders increase the complexity and the dispersion of the data [11]. Additional obstacles are privacy issues and lack of standards, resulting in a need of open and shared storing procedures and exchange processes. In some contexts, also the input uncertainties need to be quantified, based on the calculation of probability distributions to overcome the scarce quantity and quality of available data [12]. Then, a huge difference is found between heating and cooling systems mainly due to a different emissions allocation to be controlled by regulation. In other words, once installed, a heating system starts emitting CO₂ and other compounds during its operation phase where it is installed. Conversely, cooling systems are electricity-driven and, since the power production does not happen near its installation, no specific check or data collection are mandatory by regulations. This crucial diversity is the cause of a different potential not only in data availability but even in data collection opportunity and procedures. Indeed, energy storage for heating purposes is further studied in terms of potential and optimization [13] as an indirect environmental mitigation strategy.

Given that the availability of information from real systems is crucial for the development of different works related to energy systems simulation, local energy planning, and subsequent policy implementation, only heating systems are considered here.

In this regard, probes, certificates for compliance with environmental and municipal regulations, and technical start-up are the firsthand information collected by Heating, Ventilation, and Air Conditioning (HVAC) technicians [14,15]. Linked to this data collection, energy simulation models' calibration with real data has been deeply addressed by literature for any single component such as heat exchangers [16], boiler for fault analysis [17], thermal machine performance assessment [18] up to the entire building [19] even coupled with other environmental data [20]. Referring to the energy systems, other kinds of data affect the models' accuracy, including the actual performance of energy conversion and distribution systems [21], fuel quality and its impact [22], outside temperature-based performance [23], and meteorological effect on renewable energy production and systems scheduling [24,25]. The use of real data is not limited to the scale of analysis, involving from building to city-level approach to support strategic plans and programs [26] and future urban energy system layouts [27], in order to create simplified grey-box model for performance prediction [28].

As aforementioned, the importance given to those datasets is due to the opportunity to monitor both energy and environmental performance of the centralized and individual energy systems [29] as well as indirect information such as the heat demand allocation for mapping urban energy use [30] and to propose innovative energy efficiency interventions by means of lowering supply temperature [31] or coupling renewable energy and storage solutions [32].

This work presents a data analysis based on an open dataset on the heating systems in Lombardy, the most populated region among in Italy. Data are available for almost 2.9 million systems, representing around 85% of the total heating systems in the region. Starting from this huge amount of information, data pre-processing is needed to ensure the quality of the data for the analysis by removing the non-significant records [33]. The results of this work provide useful insights on the actual characteristics of heating systems in Northern Italy, which can be a useful reference for different applications, including buildings' heating systems simulation, local energy planning and policies development.

2. Methodology

The main goal of this research work is to provide useful insights for energy planning and input parameters for energy simulation models, starting from the analysis of an up-to-date database of real

heating plants in operation. Insights from real data are useful to confirm standard reference values for a number of performance indicators, including the average efficiency, the share of fuels use in the heating sector, the average installed thermal power, the number of units per inhabitant, etc.

2.1. Data Analysis

This analysis is performed by using open data from a registry of the heating plants installed in Lombardy, a large region (around 10 million inhabitants) located in Northern Italy. This choice has been made both for the climate features of this location and for the availability of open data. Lombardy is characterized by a continental climate, with an average of around 2300 Degree Days measured by Eurostat in 2009 [34], which is comparable to other locations in Europe and around the world. On the other hand, the availability of open data is crucial to guarantee the replicability and future updates of the results of this study.

The data analysis presented in this paper has been completely performed in R, an open source language and environment for statistical computing [35,36].

2.2. Description of the Dataset

The dataset is based on the regional registry of heating plants in Lombardy [37]. This registry has been developed thanks to the regional energy legislation, which makes mandatory the conduction of a census for all the heating plants by 15 October 2014 and thereafter: for both new installations and maintenance operations on existing plants. Since the majority of heating plants requires a maintenance check yearly or every two years, the current update of this dataset should include the majority of the systems in the region.

Data are currently available for 2.89 million plants (as of November 2017), including:

- fossil fuels boilers (Natural Gas, diesel oil, Liquefied Petroleum Gas (LPG), other fuels);
- biomass boilers (wood and pellet);
- heat pumps with output thermal power higher than 12 kW, i.e., with an electrical load generally between 2.5 kW and 5 kW;
- solar collectors with output thermal power higher than 12 kW, i.e., with an equivalent surface of the array usually higher than 20 m² and 25 m² in the case of Combined Thermal and Photovoltaic (PV/T);
- chillers with output cooling power higher than 12 kW;
- heat exchangers for users of District Heating (DH) networks;
- Combined Heat and Power (CHP) or Combined Cooling Heat and Power (CCHP) systems.

As a matter of fact, the registry does not include:

- water heaters for single families (for domestic hot water only), such as electric boilers installed in buildings where the heating systems are often centralized and dedicated only to space heating (in this case, the database accounts only for the space heating provided by the centralized boilers);
- wood fireplaces or stoves;
- heating plants used for industrial processes;
- heat pumps and chillers with output power lower than 12 kW (considered as the threshold from the regional regulations).

Each record of the dataset includes 41 different features that can be grouped into five categories:

- Location: features related to the municipality, the address and the cadastral references (no information is available for longitude and latitude);
- Building features: the total heated volume and the cooled one, the building category, the availability of an energy certificate and its reference number;

- Heating system features: characteristics of the heating system that are not related to the heat generator itself, such as the control logic, the emission units, and the availability of heat metering;
- Heat generator nominal features: generator type, fuel, date of installation, nominal heat output, nominal efficiency, manufacturer;
- Heat generator performance test features: date of the performance test, result (passed/failed), measured combustion efficiency.

The data distribution and some statistical summary for the most relevant features provide useful insights for a preliminary description of the heating plants installed in the region. Moreover, some indicators and some potential relations are described in the following sections.

Focus on Heat Generator Performance

An important remark is about the generator efficiency, which is available both as a nominal value and as a measured value from test reports. However, while nominal value represents the total efficiency of the heat generator, the measured value from the test reports is only considering the combustion efficiency. The test procedure is described by the Italian regulation UNI 10389-1:2009, which can be applied on each boiler fueled with liquid or gaseous fuels. The test requires a measure of the air temperature, the flue gases temperature and the O₂ concentration in the flue gases. The flue gases heat losses share Q_S is calculated by Equation (1):

$$Q_S = [A_1 / (21 - O_2) + B] \times (t_f - t_a), \quad (1)$$

where O₂ is the oxygen concentration in volume fraction in the flue gases (with a precision of $\pm 0.3\%$), t_f is the flue gases temperature (± 2 °C) and t_a is the ambient temperature (± 1 °C).

The values of A_1 and B are specific constants related to the fuel and are provided by Italian Standardized Regulation UNI 10389-1:2009 (e.g., for Natural Gas $A_1 = 0.66$ and $B = 0.010$). An alternative formula allows the calculation of Q_S by using the concentration of CO₂ instead of O₂ by using the following relation, where CO_{2,th} is theoretical carbon dioxide concentration referred to the dry exhaust gas:

$$O_2 = [1 - CO_2 / CO_{2,th}] \times 20.9. \quad (2)$$

After the calculation of the heat losses, the combustion efficiency η_{comb} can be calculated for noncondensing boilers by using Equation (3):

$$\eta_{comb} = 100 - Q_S. \quad (3)$$

The efficiency for the condensing boilers is computed accounting for the increase of the thermal output obtained through the steam condensation in the flue gases, which is performed by applying a more complex procedure than the one described before in the same regulation. The reference accuracy of the combustion efficiency in the tests defined by UNI 10389-1:2009 is in the range $\pm 2.0\%$.

2.3. Data Quality and Preprocessing

The issue of data quality is particularly critical on large datasets, where data cannot be checked manually, and automated procedures or rules need to be defined. The data quality is a key aspect in energy systems, both for energy planning analyses [33] and for management, innovation and operation [38,39]. Multiple aspects affect the coherence of the data, especially when they are not recorded by the same observer, be it a human or a sensor. Even if a method is precisely defined and codified, often the handling of unexpected results leads to differences in the recorded values.

The main problems for data quality in the present study are the following ones:

- missing data: codified with Nan, empty space or codes such as "99999";
- data of the wrong type or without physical sense (e.g., negative values for energy or power);

- data with an incorrect order of magnitude (potentially caused by a wrong interpretation of the measurement unit, but difficult to be corrected);
- accuracy of the data, which are often approximated (e.g., rounded values, approximated estimations when information is not available).

While missing data can be easily ruled out, the other errors need a more careful evaluation. A first step in the analysis is the definition of a validity range to exclude the non-acceptable values. However, for some quantities, this range can be defined from literature values (e.g., conversion efficiency) while, in other cases where the expected range is not known a priori, a manual analysis of the dataset is needed. The exclusion of the outliers is generally performed by an analysis on the percentiles of the data distribution, as often wrong data that have larger orders of magnitude need to be filtered out before performing further actions.

Finally, the aspect of data accuracy is non-trivial, as particular anomalies cannot be found by simple automatic algorithms. In some cases, anomalous distributions can be the result of compilation errors or data approximations. Figure 1 clearly shows an anomalous data record: the heating generators installed in January represent one third of all the installations. This is very unlikely, especially considering the fact that a new heat generator is seldom installed during the heating season, unless there is a major incident that cannot be repaired.

The cause is due to the fact that often only the installation year is available, but, since the system requires an installation date, the operators use January, 1st as fake date. As a result, any analysis based on the month of installation would be biased, and these anomalies cannot be easily described by a common pattern. For this reason, some aspects are still requiring a human interpretation, although artificial intelligence could provide a useful support in the future.

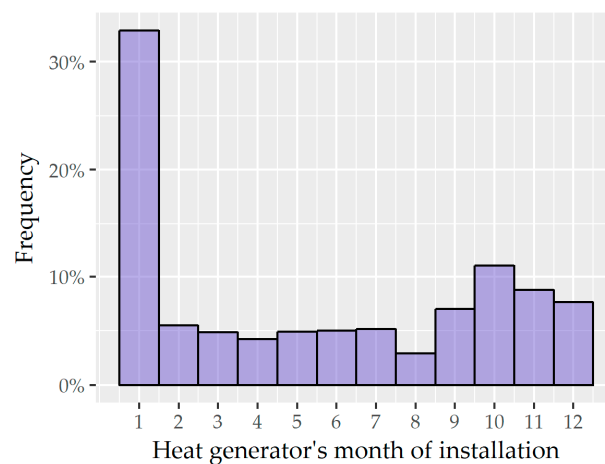


Figure 1. Example of data anomaly: number of heat generators installed per month.

Therefore, each analysis will be carried out on the largest subset with available and acceptable data for the type of interest. As a result, the analyses will be performed on different subsets of the entire dataset, depending on the available data for each analysis.

An alternative solution could be the preliminary filter on the entire dataset considering all the desired aspects, but this would result in an excessive reduction of the final dataset, as a limited amount of records has all the aspects that are correct and available. For this reason, each analysis will focus on distributions, medians, etc., in order to provide useful information that is affected by these potential errors as little as possible.

2.4. Calculations and Indicators

A large part of the calculations performed in this analysis are related to statistical evaluations on the features of the dataset entries. Data will be described by considering the distributions, medians,

percentiles, etc. The analysis of the data distributions allows describing the characteristics of the heating systems, by highlighting the main trends.

The availability of several features can be an advantage for looking at some relations among them, in order to focus on specific aspects.

Among the available features, two indicators have been considered with major detail: the ratio between installed thermal power and heated volume, and the boiler efficiency. These indicators provide useful information for the estimation of preliminary heating systems parameters both for energy simulation and for local energy planning.

3. Results and Discussion

3.1. Geographical Indicators

The first indicator of interest is related to the geographical distribution of the heating systems. The information available in the dataset includes province, municipality, zipcode, address and cadaster info. Lombardy hosts 10 million people, distributed into 1551 municipalities in 12 Provinces. The systems are well distributed among provinces, the three main provinces being Milan (21.8%), Brescia (14.5%) and Bergamo (13.8%).

The population of each municipality of the region has been compared to the number of heating systems in the same municipality, in order to calculate the number of systems per capita.

Figure 2 shows the map of the region with the administrative municipality boundaries. The majority of the values lay in the interval 0.3–0.4 systems per capita, as a result of the mix between centralized and autonomous systems. It has to be noted that, although the large majority of systems are residential, other buildings are also included into the dataset, which can lead in some particular cases to (small) municipalities with more heating systems than inhabitants (i.e., mainly mountain municipalities in the northern part of the region). The analysis of installed power per inhabitant shows a similar pattern, with median values of all provinces' distributions around 10 kW per capita.

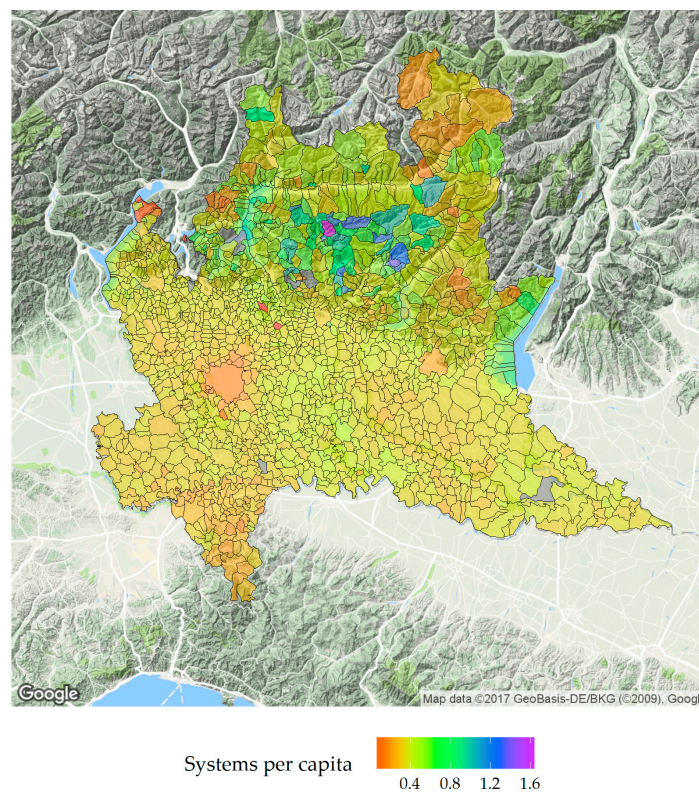


Figure 2. Number of systems per capita in each municipality.

3.2. Analysis of the Dataset

3.2.1. Features of the Building

The largest part of the systems is installed in residential buildings, accounting for a total of 89% of the records. Residential buildings include normal houses and holiday houses, but this level of detail is available for only 30% of the residential buildings (of which 2% only are recorded as holiday houses). The other buildings with a share above 1% are industrial or similar activities (4.8%, for a total of 140,000 systems), offices (2.2% of share, i.e., 64,000 units) and commercial buildings (1.4%, i.e., 39,000 systems). The other types of buildings sum up to a marginal share, and therefore they do not have a statistical relevance, but they can be useful for specific analyses. The dataset includes heating systems installed in schools (11,000), hospitals and other sanitary buildings (8000), museums, hotels, bars and restaurants, sport centers, etc.

Another piece of useful information is the heated volume, which allows for characterizing the buildings and evaluate possible indicators on the sizing of the systems (i.e., specific thermal power, see Section 3.3). The information related to the heated volume is available for a limited number of units, around 85% of the systems. The majority of systems serve small buildings or apartments, with 54% smaller than 250 m³, 36% between 251 m³ and 500 m³, and only 10% over 500 m³. However, considering the installed power instead of the number of units, the above-mentioned categories account roughly for one third each (with the middle one slightly larger). For some buildings, a reference to the energy performance certificate (called APE in the Italian regulation) is available. This reference allows for connecting the heating system to the database of the energy certifications (which is also available as open data). However, only less than 1% of the systems are currently installed in a building with a codified energy certificate, but this number is likely to increase and be useful for a deeper analysis by integrating these two datasets.

3.2.2. General Features of the Heating Systems

The first piece of information is the purpose of the heating system: the largest difference is between space heating and domestic hot water. Cooling systems are included into this dataset, but are a marginal part (less than 1%). The large majority of the heating systems, i.e., 85%, have the purpose of producing domestic hot water and providing space heating at the same time. The remaining part is mainly devoted to space heating only (10%), while other combinations including cooling sum up to the remaining 5%. Other aspects not related to the heat generator itself (which will be described in Section 3.2.3) are the features of the heat distribution and emission systems, mainly quality aspects related to design solutions and types. In particular, availability of heat metering, type of heat emission systems and heating system control logics are included. The heat metering in Italian buildings is very rare: in the entire dataset, only 2% of the heating systems are coupled to heat metering (and mainly in systems after 2005). Considering the emission systems, the large majority is represented by radiators (around 85%), while all the other systems include a variety of types, each of them lower than a two-percent share (radiant floors, air systems, fan coils, combined configurations). Finally, analyzing the control logic of the heating systems, 19% of the generators are still installed without any control logics, which remains a significant issue for the energy efficiency of the systems. The majority of the cases include an ON/OFF control, based on a dwelling thermostat in 55% of the cases and a zone thermostat in the remaining 16% of the cases. Few buildings have a proportional control. The correlation between control logic and year of installation is not significant, nor it is the one with the building category.

3.2.3. Features of the Heat Generator

This group includes the most interesting features for this analysis, representing the nominal features of the heat generators installed in Lombardy. In detail, the following aspects are of interest:

- Generator type;
- Fuel type;
- Nominal heat output;
- Year of installation;
- Number of generators;
- Manufacturer.

More than 97% of the systems in the dataset are heat generators or boilers, which is the main purpose of this cadaster. However, other systems include chillers, heat pumps, district heating exchangers, solar collectors and few CHP units. Since the cadaster has been developed focusing on the heating systems, additional information for other units is not available (e.g., surface of collectors, electric power of CHP units, etc.). For this reason, the following analyses will be focused on the heat generators. Considering heat generators' fuels, Natural Gas represents by far the fuel with the largest share, supplying almost 95% of the units. LPG, pellet, diesel oil and wood are the other fuels that are used by at least 1000 systems in the dataset. Natural Gas is very well distributed in Italy, with a network that reaches the majority of the municipalities.

Figure 3 shows that 86% of the municipalities in Lombardy have a natural gas penetration higher than 80%. The left part of Figure 3 shows that around a hundred of municipalities have no or limited access to Natural Gas network, while the range between 20% and 70% of natural gas penetration is almost empty. This distribution suggests that, where Natural Gas is available, it becomes the preferred fuel for the heating systems. Considering the heat output of the units, 87% of the plants have a capacity between 20 kW and 35 kW, which is the range related to autonomous heating systems. However, these units account for only 61% of the total cumulated installed heat output (which is around 104 GW), as larger centralized plants have a considerable weight. The most standard capacity appears to be 24 kW, which accounts alone for almost 50% of the units and roughly one third of the total cumulated thermal power.

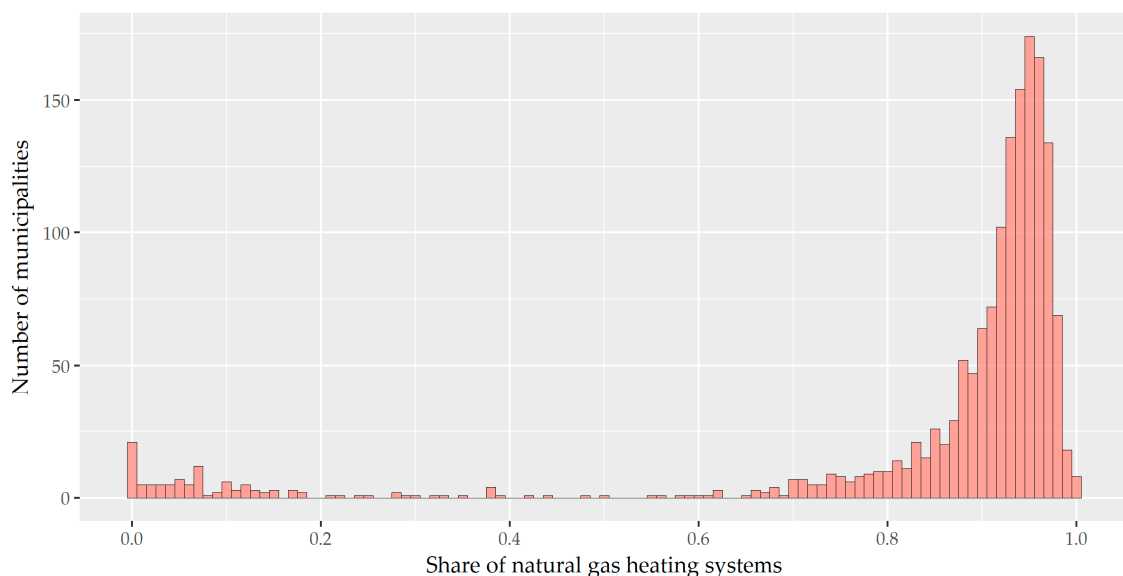


Figure 3. Distribution of municipalities per Share of Natural Gas heating systems.

Those numbers depict a reality where the majority of the users are equipped with single-dwelling boilers for space heating and domestic hot water production.

Figure 4 shows a violin plot of the five main fuels used in the boilers. As aforementioned, considering the systems with acceptable efficiency values (i.e., values that are not outside a plausible range), almost 95% of the systems are supplied by natural gas. However, each of the minor fuels has a number of units between 7000 (for the wood) to 44,000 (for LPG), thus ensuring a statistically

significant population for the analysis. Further details are reported in Table 1. The plot shows that gaseous and liquid fuels have generally higher performances than solid fuels. Natural Gas and LPG have very similar patterns, with strong peaks at 87%, 90% and 92%, which are specific values that are associated with regulation limits evolved over the years. Similar distributions can be observed for diesel oil and pellet, which show specific peaks at 90% for pellet, and at 90% and 87% for diesel oil. Wood shows generalized lower performances with a large variability, representative of less standard and automatic plants than for fossil fuels or pellet. Measured efficiency for solid fuels (wood and pellet) is not available as there is currently no regulation for the definition of a standard for the measurement. The median age of the heating systems in Lombardy is around 11 years, as it can be seen by the distribution of the installation years plotted in Figure 5. A large anomaly can be seen for the year 2000 (and a smaller one for 1990), probably due to an estimate for the systems with unknown installation year, performed by the professionals who filled out the reports.

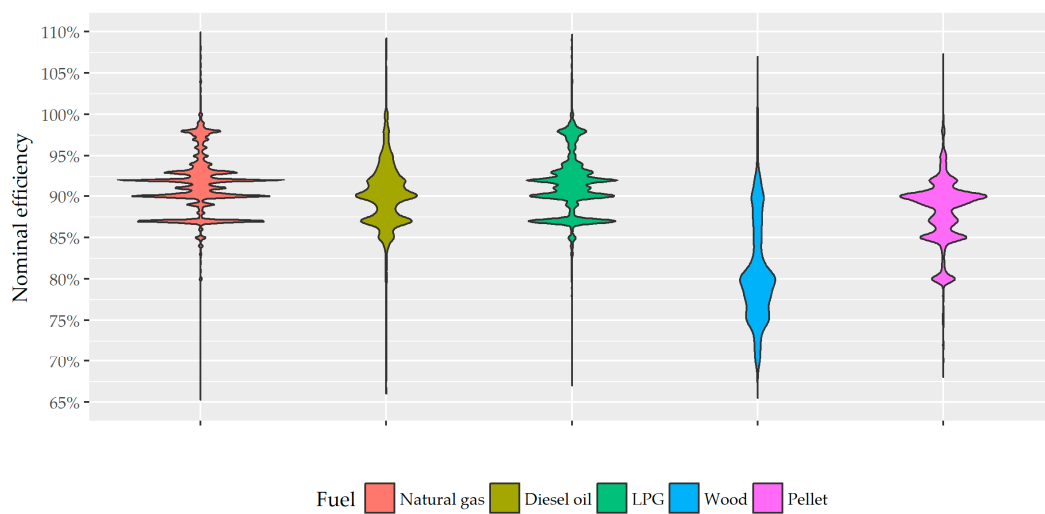


Figure 4. Nominal efficiency for boilers supplied by different fuels.

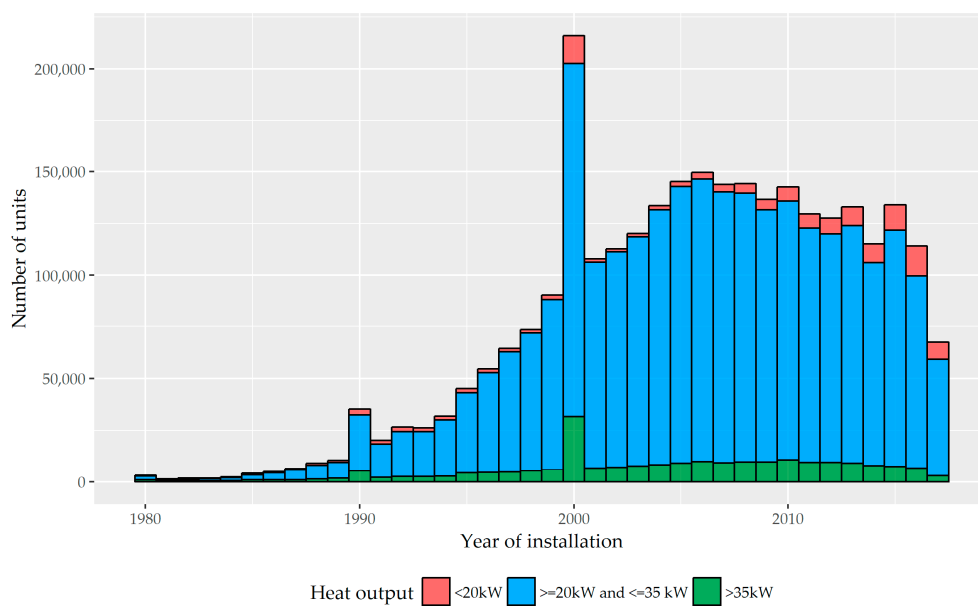


Figure 5. Distribution of heat generators per year of installation.

Table 1. Features of the heating systems by fuel type.

Fuel Type	Number of Plants	Of Which “NA”	Cumulative Power (MW)	Mean Power (kW)	Median Power (kW)
Natural Gas	2,666,028	1363	94,095	35.3	24.1
Oil	34,497	37	3035	88.0	32.0
LPG	58,628	57	1733	29.6	24.5
Pellets	43,927	0	525	12.0	10.3
Wood	12,020	3	192	16.0	11.0
Other ¹	76,756	21,534	4843	87.7	22.0
Total	2,891,856	22,994	104,424	36.4	24.1

¹ “Other” may include one of the previous fuel types, but codified in a wrong way.

However, this bias is not drastically influencing the distribution. An additional aspect noticeable in Figure 5 is the slight increase in the last decade of small units (20 kW of output thermal power), probably due to the diffusion of high-efficiency buildings that have lower heat demand. The data for the year 2017 are obviously partial and they cannot represent any significant trend. As aforementioned, a great part of the installations happen at the beginning of the winter season, so, for 2017, the period between September and November will be the crucial one. Finally, two marginal aspects are worth being mentioned: the number of generators in each system and the manufacturer.

Referring to the first aspect, the heat generators of this dataset are mainly part of a heating system with a single heat generator (93%), while minor shares of units are in a two-unit heating plants (3%), or three-unit plants (1%). Larger groups are available but with negligible share on the dataset. Then, as regards the second aspect, the top five manufacturers (Beretta: Lecco, Italy; Vaillant: Remscheid, Germany; Immergas: Reggio Emilia, Italy; Riello: Verona, Italy; Baxi: Warwick, UK) together account for 53% of the units and 43% of the total installed thermal power.

3.3. Specific Thermal Power

The ratio between the installed thermal power and the heated volume of the building or apartment is defined as specific thermal power.

This parameter is often useful for a preliminary estimation of the requested power, and it depends on multiple aspects, including the geometrical features of the buildings (the surface area to volume ratio, the share of glazed area, the surface area contiguous to another building, etc.), the insulation and other heating design parameters. For this reason, the value of specific thermal power usually shows some variability. Figure 6 reports the frequency distribution of the specific thermal power, dividing the heat generators classified as “Space heating and Domestic Hot Water (DHW)” (around 86% of the total) and the ones for “Space heating only” (around 10% of the total, the remainder being classified for other purposes, such as “Other”, “Cooling” or a combination of the previous ones).

The distribution of “Space heating only” systems has a mode of 35 W/m³, and a median of 54 W/m³. On the other hand, the distribution of “Space heating and DHW” systems has higher values (a median of 96 W/m³ and a mode of 80 W/m³). The need for producing instant DHW usually leads to a large oversizing of boilers in small dwellings: around one quarter of the total boilers are rated to a single capacity (24 kW), whereas 30% of the buildings or dwellings have a volume of 300 m³, 270 m³ or 240 m³.

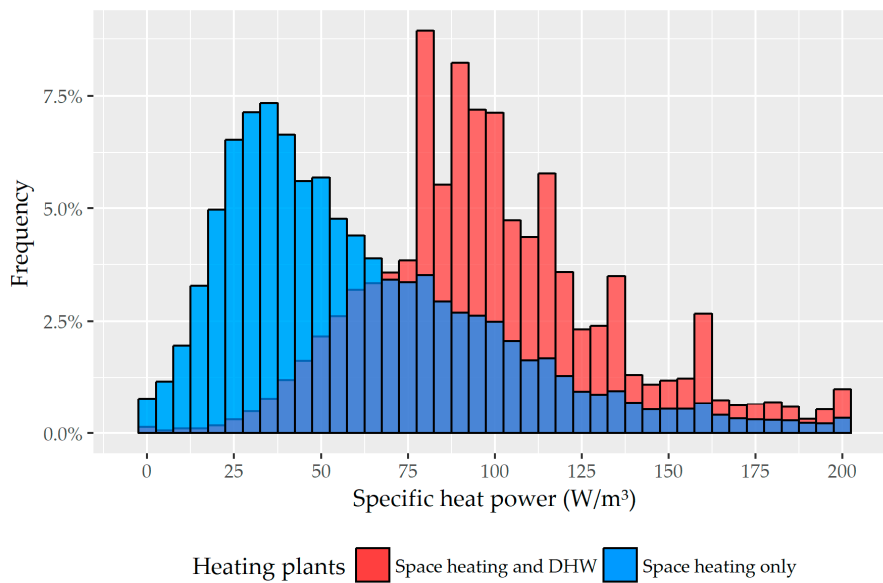


Figure 6. Distributions of nominal and measured (combustion) efficiency.

3.4. Natural Gas Boilers—Efficiency

Figure 7 shows a histogram of nominal and measured efficiency distributions, limited to natural gas boilers. Both traditional and condensing boilers have been considered, and the latter are responsible for higher efficiency, especially when observing nominal values. Natural Gas boilers in Lombardy are 2.67 million, of which only 64% have acceptable nominal and measured efficiency. The efficiency has been considered as acceptable only in the range 75% to 110%, in order to filter out the values that may lead to results with low significance.

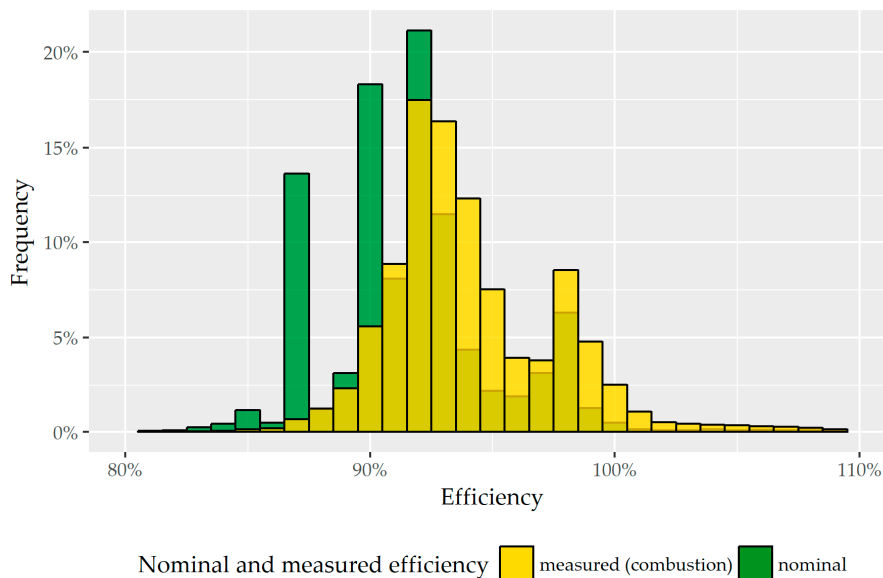


Figure 7. Distributions of nominal and measured (combustion) efficiency for Natural Gas boilers.

The largest part of unacceptable values is due to nominal efficiency (around 28.1%), while a smaller part has unacceptable measured efficiency (a share of 3.9%). Measured efficiency is related to a test required by the regional legislation, while nominal efficiency is non-compulsory information, which is therefore often ignored or reported as a wrong value (usually 0 or 1).

The bins' width in the histogram of Figure 7 has been set to 1%, as often nominal efficiencies are reported rounded to the percentage units, i.e., with no decimals. In particular, the three values of efficiency with a higher frequency are 92%, 87% and 90%, which represent, respectively, 12.9%, 11.3% and 10.5% of the total boilers. These values were associated in past years with some limits required by the regulations. Therefore, the boilers entering the market have been built in accordance with those limits.

The most noticeable aspect in Figure 7 is the fact that the distribution of measured efficiency contains higher values than the distribution of nominal efficiency. However, it has to be noted that the measured efficiency is only representing the combustion efficiency, as only the flue gas losses are accounted for during the performance field controls. The case losses are therefore not considered in the measurement.

On the other hand, when the boiler is installed in a heated room, case losses are contributing to the space heating, and therefore they should not be accounted. Other aspects that cause the differences between those efficiency distributions are the accuracy of the instruments (the accuracy of field instruments is estimated around $\pm 2\%$) and the additional measures available in laboratory (e.g., fuel composition, heating value).

For all these reasons, a comparison of nominal efficiency and measured efficiency should take into account all the aspects mentioned above.

A further analysis can be performed on the measured efficiency, as reported in Figure 8.

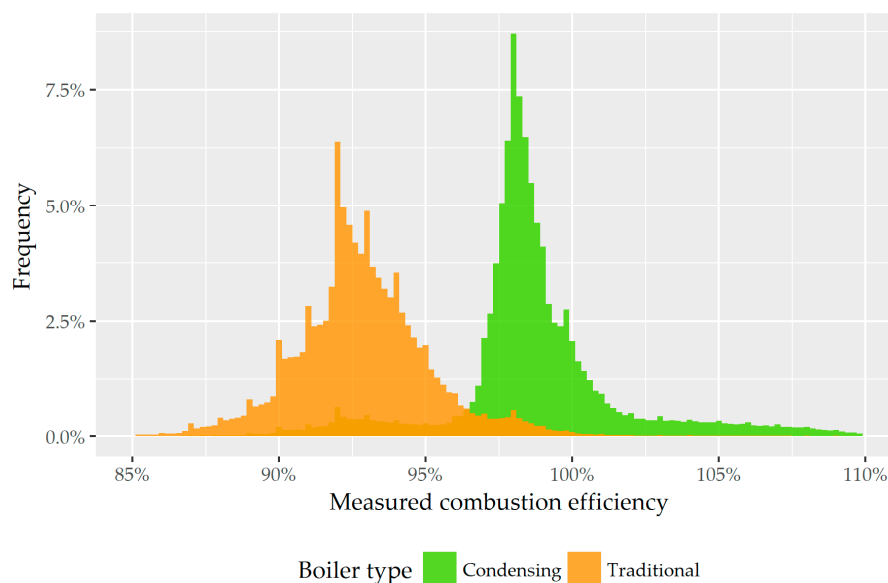


Figure 8. Distributions of nominal and measured (combustion) efficiency for Natural Gas boilers.

In this case, the dataset also includes the systems with unacceptable nominal efficiency but acceptable measured efficiency (around 2.55 M); for this reason, the frequency of Figure 8 shows different values from those previously discussed of Figure 7, although the trend is comparable. The bins' width of this chart has been set to 0.2%.

The duality of the distribution is related to the type of boiler: traditional boilers show lower efficiency, with a median of 92.7%, while condensing boilers have a median efficiency of 98.4%. Moreover, traditional boilers show a wider variability, which is probably caused by their average higher age and its wider range, and the more stringent limits set by the regulations for new and condensing boilers.

A final remark is related to the evolution of the efficiency over the last years, following the limits set by the regulations, which gradually raised the limits for a 30 kW traditional natural gas boiler from 85% before 1993 to 90% after 2005, and to 92% for a condensing boiler.

4. Conclusions

This work presents a data analysis on a large dataset of heating systems in a region in Northern Italy. The results provide useful insights for further works related to energy simulation and local energy planning. The main aspects are the following ones:

- The availability of large datasets is a precious support for the analysis of the characteristics of existing systems. However, attention must be paid to the data quality, as missing data points and errors could significantly affect the aggregated results. The dataset considered in this study shows that, while the availability of big data is a powerful resource, the data quality should be further improved. For this reason, distributions and medians provide more accurate insights than means and sums, since the former are less affected by missing data or erroneous outliers.
- The large majority of heating systems, both for number of units and heated volume (where available), is composed by residential buildings or dwellings. Around 90% of the heating systems are installed in buildings with a heated volume lower than 500 m³.
- The ratio between installed thermal power and heated volume is a useful indicator for design parameterizations. Considering heating systems used only for space heating, the distribution of the specific thermal power shows a mode of 35 W/m³ and a median of 54 W/m³. The simultaneous production of domestic hot water leads to a significant increase of the distribution. The main driver appears to be the standardization of boilers in small dwellings, which usually have a thermal power output between 24 kW and 28 kW.
- Natural Gas is the most diffused fuel for heat production in Northern Italy. The municipalities served by the network have a very high share of natural gas heating systems (usually above 85%–90%), while there are still some municipalities (mainly in mountain regions) in which natural gas is not available.
- The nominal efficiency of the heat generators shows a considerable dependence on the fuel. Another major driver appears to be the lower acceptance limits set by the regulations in recent years, which correspond to the most common nominal efficiency values (87%, 90% and 92% for Natural Gas).
- The dataset also includes information about the measured combustion efficiency of the boilers. An analysis of the natural gas heat generators shows that two separate distributions can be highlighted for traditional and condensing boilers, the former with a median of 92.7%, and the latter with a median of 98.4% and a lower variability.

These insights describe the current situation of heating systems in Lombardy, which is representative of the situation of Northern Italy. The results of this work can be the basis for further analyses on different domains: energy policies, local planning and simulation of energy systems.

In more detail, a first step can be taken by overlapping this database to the one related to the energy label to assess the quality of the heating demand and to provide a clearer picture of the Public Administration. Furthermore, heat metering and sensors as part of Building Management Systems or inclusion in the monitoring phase for accessing incentive schemes, or even energy efficiency credits, can certainly increase the requests for data analysis similar to the proposed one by the authors as well as create dedicated guidelines to collect and manage this data for codified energy strategy and associated checking procedures.

Author Contributions: The authors equally contributed to the paper. M.N. conceived the idea and analyzed the data, B.N. and M.N. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Urge-Vorsatz, D.; Cabeza, L.F.; Serrano, S.; Barreneche, C.; Petrichenko, K. Heating and cooling energy trends and drivers in buildings. *Renew. Sustain. Energy Rev.* **2015**, *41*, 85–98. [[CrossRef](#)]
2. Im, J.; Seo, Y.; Cetin, K.S.; Singh, J. Energy efficiency in U.S. residential rental housing: Adoption rates and impact on rent. *Appl. Energy* **2017**, *205*, 1021–1033. [[CrossRef](#)]
3. Collado, R.R.; Díaz, M.T.S. Analysis of energy end-use efficiency policy in Spain. *Energy Policy* **2017**, *101*, 436–446. [[CrossRef](#)]
4. Krarti, M.; Dubey, K.; Howarth, N. Evaluation of building energy efficiency investment options for the kingdom of Saudi Arabia. *Energy* **2017**, *134*, 595–610. [[CrossRef](#)]
5. Noussan, M.; Jarre, M. Multicarrier energy systems: Optimization model based on real data and application to a case study. *Int. J. Energy Res.* **2018**. [[CrossRef](#)]
6. Tronchin, L.; Manfren, M.; Tagliabue, L.C. Optimization of building energy performance by means of multi-scale analysis—Lessons learned from case studies. *Sustain. Cities Soc.* **2016**, *27*, 296–306. [[CrossRef](#)]
7. Stankovic, L.; Stankovic, V.; Liao, J.; Wilson, C. Measuring the energy intensity of domestic activities from smart meter data. *Appl. Energy* **2016**, *183*, 1565–1580. [[CrossRef](#)]
8. Chou, J.S.; Yutami, G.A.N. Smart meter adoption and deployment strategy for residential buildings in Indonesia. *Appl. Energy* **2014**, *128*, 336–349. [[CrossRef](#)]
9. Pisello, A.L.; Rosso, F.; Castaldo, V.L.; Piselli, C.; Fabiani, C.; Cotana, F. The role of building occupants' education in their resilience to climate-change related events. *Energy Build.* **2017**, *154*, 217–231. [[CrossRef](#)]
10. Miranda, M.T.; Montero, I.; Sepúlveda, F.J.; Arranz, J.I.; Rojas, C.V. Design and Implementation of a Data Acquisition System for Combustion Tests. *Energies* **2017**, *10*, 630. [[CrossRef](#)]
11. Cajot, S.; Peter, M.; Bahu, J.; Guignet, F.; Koch, A.; Maréchal, F. Obstacles in energy planning at the urban scale. *Sustain. Cities Soc.* **2017**, *30*, 223–236. [[CrossRef](#)]
12. Manfren, M.; Aste, N.; Moshksar, R. Calibration and uncertainty analysis for computer models—A meta-model based approach for integrated building energy simulation. *Appl. Energy* **2013**, *103*, 627–641. [[CrossRef](#)]
13. Cheng, H.; Wang, X.; Zhou, M. Optimized Design and Feasibility of a Heating System with Energy Storage by Pebble Bed in a Solar Attic. *Energies* **2017**, *10*, 328. [[CrossRef](#)]
14. Karagiannidis, A. Burners of domestic heating boilers: A measurement-based analysis approach aiming at quantifying correlations among the basic parameters of operation. *Energy Convers. Manag.* **1996**, *37*, 447–456. [[CrossRef](#)]
15. Noussan, M.; Jarre, M.; Poggio, A. Real operation data analysis on district heating load patterns. *Energy* **2017**, *129*, 70–78. [[CrossRef](#)]
16. Turhan, C.; Simani, S.; Zajic, I.; Akkurt, G.G. Performance Analysis of Data-Driven and Model-Based Control Strategies Applied to a Thermal Unit Model. *Energies* **2017**, *10*, 67. [[CrossRef](#)]
17. Sarkar, P.; Kortela, J.; Boriouchkine, A.; Zattoni, E.; Jämsä-Jounela, S.-L. Data-Reconciliation Based Fault-Tolerant Model Predictive Control for a Biomass Boiler. *Energies* **2017**, *10*, 194. [[CrossRef](#)]
18. Rovense, F.; Amelio, M.; Scornaienchi, N.M.; Ferraro, V. Performance analysis of a solar-only gas micro turbine, with mass flow control. *Energy Procedia* **2017**, *126*, 675–682. [[CrossRef](#)]
19. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [[CrossRef](#)]
20. Royapoor, M.; Roskilly, T. Building model calibration using energy and environmental data. *Energy Build.* **2015**, *94*, 109–120. [[CrossRef](#)]
21. Glasgo, B.; Hendrickson, C.; Azevedo, I.L. Assessing the value of information in residential building simulation: Comparing simulated and actual building loads at the circuit level. *Appl. Energy* **2017**, *203*, 348–363. [[CrossRef](#)]
22. Castellani, B.; Rinaldi, S.; Bonamente, E.; Nicolini, A.; Rossi, F.; Cotana, F. Carbon and energy footprint of the hydrate-based biogas upgrading process integrated with CO₂ valorization. *Sci. Total Environ.* **2018**, *615*, 404–411. [[CrossRef](#)] [[PubMed](#)]
23. Jarre, M.; Noussan, M.; Poggio, A.; Simonetti, M. Opportunities for heat pumps adoption in existing buildings: Real-data analysis and numerical simulation. *Energy Procedia* **2017**, *134*, 499–507. [[CrossRef](#)]

24. Liao, S.; Yao, W.; Han, X.; Wen, J.; Cheng, S. Chronological operation simulation framework for regional power system under high penetration of renewable energy using meteorological data. *Appl. Energy* **2017**, *203*, 816–828. [[CrossRef](#)]
25. Rovense, F.; Perez, M.S.; Amelio, M.; Ferraro, V.; Scornaienchi, N.M. Feasibility analysis of a solar field for a closed unfired Joule-Brayton cycle. *Int. J. Heat Technol.* **2017**, *35*, 166–171. [[CrossRef](#)]
26. Lo Basso, G.; Rosa, F.; Astiaso Garcia, D.; Cumo, F. Hybrid systems adoption for lowering historic buildings PFEC (primary fossil energy consumption)—A comparative energy analysis. *Renew. Energy* **2018**, *117*, 414–433. [[CrossRef](#)]
27. Nastasi, B.; Lo Basso, G. Power-to-gas integration in the transition towards future urban energy systems. *Int. J. Hydrogen Energy* **2017**, *42*, 23933–23951. [[CrossRef](#)]
28. Miao, Q.; You, S.; Zheng, W.; Zheng, X.; Zhang, H.; Wang, Y. A Grey-Box Dynamic Model of Plate Heat Exchangers Used in an Urban Heating System. *Energies* **2017**, *10*, 1398. [[CrossRef](#)]
29. Noussan, M.; Jarre, M.; Roberto, R.; Russolillo, D. Combined vs Separate Heat and Power Production—Primary Energy comparison in high renewable share contexts. *Appl. Energy* **2018**, *213*, 1–10. [[CrossRef](#)]
30. Wyrwa, A.; Chen, Y.-K. Mapping Urban Heat Demand with the Use of GIS-Based Tools. *Energies* **2017**, *10*, 720. [[CrossRef](#)]
31. Astiaso Garcia, D. Can radiant floor heating systems be used in removable glazed enclosed patios meeting thermal comfort standards? *Build. Environ.* **2016**, *106*, 378–388. [[CrossRef](#)]
32. Rovense, F.; Amelio, M.; Ferraro, V.; Scornaienchi, N.M. Analysis of a concentrating solar power tower operating with a closed Joule Brayton cycle and thermal storage. *Int. J. Heat Technol.* **2016**, *34*, 485–490. [[CrossRef](#)]
33. Nouvel, R.; Zirak, M.; Coors, V.; Eicker, U. The influence of data quality on urban heating demand modeling using 3D city models. *Comput. Environ. Urban Syst.* **2017**, *64*, 68–80. [[CrossRef](#)]
34. Eurostat. Heating Degree-Days by NUTS 2 Regions—Annual Data. 2013. Available online: <http://ec.europa.eu/eurostat/web/energy/data/database> (accessed on 29 November 2017).
35. R Core Team. R: A Language and Environment for Statistical Computing. 2017. Available online: <https://www.r-project.org/> (accessed on 21 November 2017).
36. Wickham, H. Tidyverse: Easily Install and Load ‘Tidyverse’ Packages. 2017. Available online: <https://cran.r-project.org/package=tidyverse> (accessed on 21 November 2017).
37. CURIT. Infrastrutture Lombarde. 2017. Available online: <http://www.curit.it> (accessed on 5 December 2017).
38. Barbeito, I.; Zaragoza, S.; Tarrío-Saavedra, J.; Naya, S. Assessing thermal comfort and energy efficiency in buildings by statistical quality control for autocorrelated data. *Appl. Energy* **2017**, *190*, 1–17. [[CrossRef](#)]
39. Cappa, F.; Del Sette, F.; Hayes, D.; Rosso, F. How to Deliver Open Sustainable Innovation: An Integrated Approach for a Sustainable Marketable Product. *Sustainability* **2016**, *8*, 1341. [[CrossRef](#)]

