Evaluation of Machine Learning techniques to enact energy consumption control of Compressed Air Generation in production plants

Santolamazza A.*, Cesarotti V.*, Introna V.*

** Department of Enterprise Engineering, "Tor Vergata" University of Rome, Via del Politecnico, 1, 00133, Rome – Italy (annalisa.santolamazza@uniroma2.it, cesarotti@uniroma2.it, vito.introna@uniroma2.it)*

Abstract: Industrial energy management is an important topic of discussion nowadays for both economic and sustainability reasons. A monitoring and control system able to guarantee the practice of a real-time control is a key point in enacting an effective management of energy consumption in a complex organization. In this context, the ISO 50000 family of standards suggest the application of different types of energy performance indicators (EnPIs), in a range of varying complexity: from simple absolute values of energy consumption, to statistical models, to engineering models. The evolution of machine learning techniques falls between the statistical and the engineering models, depending on the volume of data and the human involvement required for building a model. Therefore, the value of the present work is to explore the use of these tools, already consolidated in other fields, but not yet adequately assessed for energy performance control. In particular, the generation and distribution of compressed air is among the biggest uses of energy in production plants. This work starts with the application of the classical statistical approach and then proceeds to compare two different machine learning techniques, artificial neural networks and support vector machines, for the creation of energy performance indicators. The analysis begins comparing the feasibility of application, implementation complexity, data and level of human interaction required, making use of the results of a real application to a compressed air generation unit in a production plant. The comparison was then carried out using various performance indicators (R-squared, Mean Squared Error, Mean Absolute Percentage Error) as well as a graphical inspection of the resulting control charts produced with the different models. The work demonstrates the applicability of machine learning techniques in this specific context, proving them as an efficient compromise between the complexity and accuracy of statistical and engineering models.

Keywords: Machine Learning, Compressed Air Generation Unit, Energy Management, Artificial Neural Networks, Support Vector Machines

1. Introduction

In the wake of the outbreak of the energy crisis of 1973 and the consequent sudden and drastic increase in the price of energy (Geller et al., 2006; Di Silvio et al., 2007), public opinion was struck by the awareness that its energy resources may not be able to sustain the pace of increasing human consumption. On this occasion, the countries affected by the embargo resorted to two different strategies: the search for new energy sources and the development and implementation of energy saving policies.

In time, attention to climate change and more generally to the sustainability of human development has only strengthened the importance of the issue. Thus, legislators decided to act on several levels, establishing international agreements as well as national laws and specific sectorial standards. Since the industry sector is considered a major energy consumer and consequently producer of greenhouse gas (GHG) emissions (Edelenbosch et al., 2017), the necessity to intervene in this sector to change the current situation is clear.

In the recent market report "Energy Efficiency 2O17" the International Energy Agency (IEA) reported an improvement of energy efficiency in the industrial sector with the energy use per unit of economic output in the industrial sector falling by nearly 20% between 2000 and 2016. Moreover, the report states that "The application of energy management systems, which provide a structure to monitor energy consumption and identify opportunities to improve efficiency, is growing, driven by policy and financial incentives. The number of certifications for ISO 50001 (…) grew to nearly 12000 in 2015, 85% of which were in Europe." Indeed, the aim of an energy management system is to establish an energy policy and energy objectives and implement processes and procedures to achieve those objectives (ISO 50001:2011).

The main goal of energy management is to ensure the most efficient provision of energy to all users, at the lowest cost possible and without negative effects on the other aspects of the organization (production quality and environmental sustainability among the most important) (Petrecca, 2014). According to Turner and Doty (2004), energy management can be an effective tool in order to guarantee the economic competitiveness of an organization in the global marketplace, fostering the reduction of industrial energy intensiveness while upholding customer service’s needs for quality and delivery times.

A monitoring and control system able to guarantee the practice of a real-time control is of great importance in enacting an effective management of energy consumption in a complex organization.

Furthermore, a key factor in achieving a continuous improvement of the energy performances of the organization is the implementation of energy performance indicators (EnPIs) (ISO 50006:2014; Benedetti et al., 2017a).

On a different note, the rapid innovations regarding sensor technology, wireless transmission, network communication and cloud computing are producing an exponential growth in the volume and complexity of data acquired (Zhou et al., 2016). The diffusion of connected devices is growing, enhancing monitoring of production processes. This innovation generates several opportunities in the monitoring of energy consumption, providing a better insight on the real energy behaviour of the systems monitored (Shrouf and Miragliotta, 2015).

Among the primary challenges deriving from the use of energy data there are the issues of how to efficiently analyse and mine the data, how to use them in the decision-making process and how to obtain values from the data. In this context, many data analysis techniques can be applied from clustering to optimization in order to gain specific information from the data (Zhou et al., 2016).

When focusing on industrial energy efficiency, one aspect of great interest is the energy performance of the compressed air generation system. In fact, compressed air is widely used in industry as demonstrated by the number of companies that use compressed air in their operations (approximately 70%)(Morvay, 2008) and by the great availability of related services (Benedetti et al., 2015). The energy consumption resulting its generation represents approximately 10% of the plant electricity consumption (Morvay, 2008). Results from analyses conducted on a data collection acquired from the mandatory energy audits conducted in Italy in accordance with the European Directive 2012/27/EU seem to further confirm this fact. Compressed air systems (CAS) appear to be a significant energy use in most industries, with variable ranges from 4% to 12%, with a mean value across different industries of 7% and national industry consumption of 5% considering all analysed industries (Benedetti et al., 2017b). Moreover, the cost of a compressor over a ten-year life consists for the 73% of energy cost, while the maintenance cost is approximately 7% and capital cost and installation together are 20%. Finally, only the 8-10% of the total energy supplied to a compressor is converted into usable energy at the point of use (Carbon Trust, 2012).

The cost of generation of compressed air is influenced by the efficiency of singles compressors and by different factors (Morvay, 2008): operational procedures, compressor configuration, individual compressor control system, number of compressors used to meet the demand, overall control system, location of compressor room, temperature of the inlet air, quality of the cooling systems, and quality of maintenance.

Indeed, many are the factors that should be taken into account when monitoring the performance of compressed air systems because their variable characteristics can make the monitoring of reliable key performance indicators very difficult (Salvatori et al., 2018).

However, as already stated, the definition of clear and effective energy performance indicators is a key point in the process of performance control of energy systems.

In literature, the range and complexity of possible indicators is huge. In particular, with the purpose of providing guidance and clarification in the difficult process of establishing energy performance indicators (EnPIs) to measure energy performance and energy performance changes, the International Organization for Standardization (ISO) has published another specific standard, ISO 50006:2014. In this standard there are proposed four types of different EnPIs, presented in increasing degree of complexity: measured energy value, ratio of measured values, statistical model or engineering model.

The first two types of EnPIs are simple and widely diffused. Their drawback is that they do not take into account the influence of different conditions and variables that can impact the energy consumption. Instead, the other two types of EnPIs are more complex and take into consideration changes in relevant variables and their interactions. Various statistical models use historical data to individuate the relation between energy consumption and its drivers with different complexity (Cesarotti et al., 2009): from simple univariate linear regression to quadratic or cubic regressions. Engineering models use physical and thermodynamic functions to exactly derive the value of the theoretical energy consumption. They can be divided into simplified and elaborate models according to the number of equations and variables considered in the model (Magoulès, and Zhao, 2016). However, the energy performance of a system can be also modelled with other more advanced techniques. In fact, many different artificial intelligence techniques, such as artificial neural networks, decision trees and support vectors machines, could be used for creating different energy performance indicators (Benedetti et al., 2016; Magoulès, and Zhao, 2016).

Figure 1 shows all the possible types of EnPIs, highlighting the ones cited in the ISO 50006:2014.

The evolution of machine learning techniques falls between the statistical and the engineering models, depending on the volume of data and the human involvement required for building a model.

This work describes the application of the classical statistical approach and then proceeds to compare two different machine learning techniques, artificial neural networks and support vector machines, for the creation of energy performance indicators. Therefore, the value of the present work is to explore the use of these tools, already consolidated in other fields, but not yet adequately assessed for energy performance control.

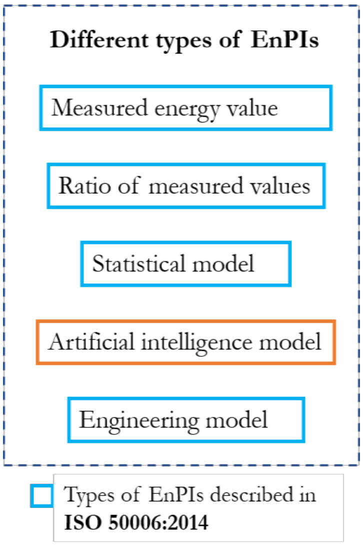

Figure 1: types of different EnPIs

Moreover, to the best of our knowledge we believe that this is the first paper that focus directly to the use of these three specific techniques in the specific domain of energy consumption control of a compressed air generation system.

2. Methodology

With the aim of highlighting the differences and the similarities in the use of the three techniques, building on the characteristics recognised by Magoulès, and Zhao (2016), five categories of drivers have been identified: model complexity, data requirements, accuracy, control issues, hardware and software requirements.

Figure 2 shows the five drivers identified and their sub-categories of comparison.

2.1 Model complexity

This first category is related to the capabilities for the human resources to have to effectively create the model and use it in time.

First of all, machine learning models are usually described as "black-box" models (Benedetti et al., 2016) while the statistical models are called "white-box". The reason behind this different description is that in contrast to a statistical model, whose equation can be explicitly known, for machine learning models it is not possible to know what relation has been learned by the algorithm during the training (Model comprehensibility).

Moreover, machine learning techniques require a different body of knowledge than the one used to create the statistical models both for the development of the model itself and for its maintenance in time (Knowledge required).

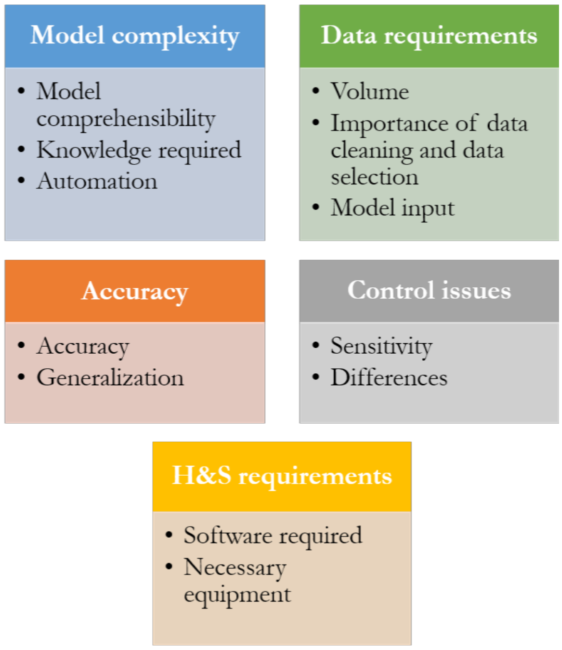

Figure 2: drivers for the comparison of the different types of models presented

Another aspect to consider is the feasibility of the automation of the model execution. This aspect is connected to the realization and execution of the model as well as the infrastructure used for the implementation of the control (Automation). Since the implementation of machine learning models already requires the existence of a certain infrastructure, the development of advanced real-time control procedures could be proved easier. On the other hand, the issues always at the basis of machine learning, concerning the cleaning and processing procedures of input data, complicate the automatization of the procedure, which would therefore also involve the automation of these steps.

2.2 Data requirements

The amount of data necessary to create a machine learning model is very huge in comparison to the data necessary to the creation of a statistical model. The lack of an adequate amount of data will result in a poorly generalized model, unable to predict accurately the energy consumption expected with different conditions (Volume). This will also mean that according to the volume and nature of data available (e.g. monthly or weekly data for a limited period), certain types of model could be impossible to create.

Furthermore, the choice of the right set of data is of the upmost importance. Since in machine learning the relation described by the model is not known, it is possible to go into overfitting of the model when choosing an inadequate data set and thus losing the capacity of generalization of the model (Importance of data cleaning and data selection).

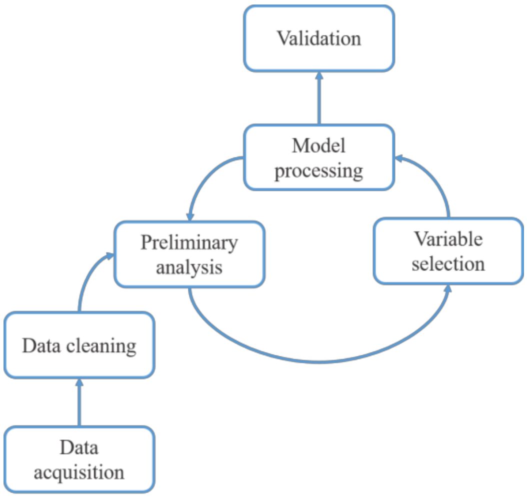

Figure 3: simplified process of creation for the model, from data acquisition to model validation

Lastly, the choice of model inputs is affected by the type of model used. First of all, machine learning models can be applied to nonlinear problems, whereas a statistical linear regression would not work efficiently. This means that in machine learning models, in contrast with the statistical linear regression one, even variables that do not have an approximately linear relation with the output can be added. Besides, with artificial intelligence models it is possible to use not only continuous variable but also categorical ones (Magoulès and Zhao, 2016).

However, it is important to remember the aim in the use of an energy performance model: the model must take into account only non-controllable factors, thus highlighting the real inefficiencies in the energy performance when compared with the actual energy consumption and distinguishing from the differences in behaviour that are inherent of the system and therefore are not the object of the control system (Model input).

In Figure 3 it is shown the simplified process required for the creation of the model, from data acquisition to model validation.

2.3 Accuracy

The accuracy of the model represents its capacity of correctly predict the energy baseline performance of the system when presented with new data. However, it is important for the model to not learn the "faulty behaviour of the system" otherwise it would become ineffective in the detection of behavioural anomalies.

The comparison is carried out using three performance indicators: R-squared, Mean Squared Error, Mean Absolute Percentage Error (Accuracy).

Furthermore, the comparison includes a graphical inspection of the resulting residuals control charts produced, showing the difference between the actual value of energy consumption and the one predicted by the model, in order to detect out of range points or non-random patterns like shifts, trends or mixtures for example (Montgomery, 2009).

Maintaining the same trend and behaviour shows a capacity of generalization from the model (Generalization).

2.4 Control issues

The residuals control charts are made using control limits ensued by the different models. The width of the limits reflects the sensitivity of the control system: the larger they are, more probable is the possibility of not detecting an anomaly in the energy performance of the system (Sensitivity).

Since different input variables can be used in the models and different relationships can be described it is important to analyze the differences in the control charts and the anomalies detected by the different models (Differences).

2.5 Hardware and software requirements

A smart object, according to Shrouf and Miragliotta (2015) is an object which possesses some functionalities such as self-diagnosis and location awareness, communication with other smart objects and/or with the central acquisition system, optional interaction with the surrounding environment. It could also possess the capability of data processing, i.e. elaboration of the data collected. It is obvious that the feasibility of the model will be firmly connected to the presence of an adequate measurement system (Necessary equipment).

While the elaboration of a statistical model could be usually achieved through a simple software or even by mathematical calculation, machine learning models requires different kind of infrastructures to enable the processing of the models, even more if the architecture and desired characteristics are personally designed by the user (Software required).

3. Case study

The comparative analysis has been applied to the case study of a compressed air generation system in an Italian manufacturing company.

The system is composed of four air compressors with similar capacity, with one of them equipped with variable-speed drive. All compressors produce a single compressed air flow at a pressure of 8 bar and part of the flow is reduced immediately to the pressure of 3 bar.

Main variables monitored by the monitoring system are: low pressure flow rate (3 bar) and medium pressure flow rate (8 bar), energy consumption of singles compressors, external air temperature and humidity. All data is collected with a fifteen minutes time step. The original dataset available is from the 1st September 2017 till the 14th January 2018.

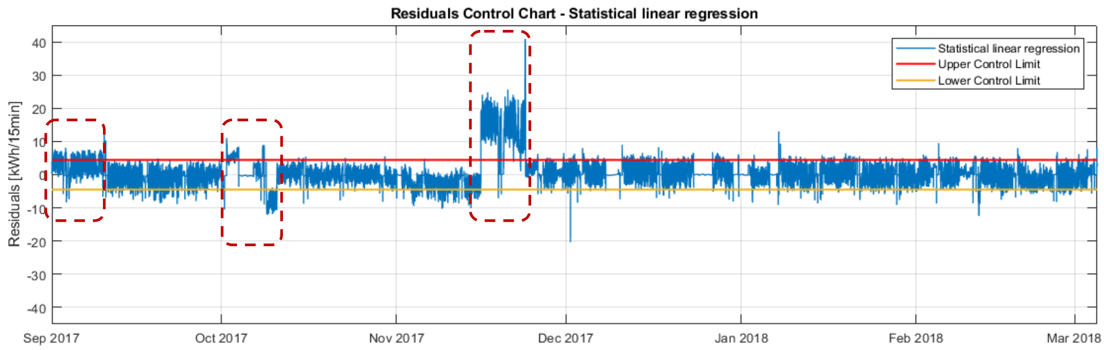

Figure 4: Residuals Control Chart created using the statistical linear regression model – Complete dataset to allow the identification of anomalous periods, excluded in the training (1st September 2017 - 4th March 2018)

During the data cleaning process and preliminary analysis, it was detected a measurement error in the second and third weeks of November. Thus, data used for model processing excluded the period 15 - 23 November. This process is quite critical when working with machine learning techniques because differences in the variability of the dataset could result in poorly generalized models.

Furthermore, it was discovered that the humidity sensor goes frequently in error, signalling zero instead of the actual value. It was so decided to exclude this variable because deemed unreliable.

To take into consideration the variability between the two flow rates profiles, a new variable was created: the ratio between low pressure flow rate and total flow rate. The inputs of the three regression models are global flow rate, ratio flow rate and external air temperature. The output of all models is the global energy consumption (sum of the single values of energy consumption for all the compressors).

Three different typologies of models have been therefore realized: artificial neural network, support vector machine and statistical linear regression.

For the artificial neural network was chosen a Multilayer Feedforward Perceptron structure with only one hidden layer. The selection of the ideal number of neurons was made through different experimentations, changing both the random sampling selection and the number of neurons in the hidden layer (from 1 to 27) to determine the structure with the best performance. Before training the network, the training data was divided into three groups: data used to actually train the network, data used for validation, and data used to test the performance, with a percentage respectively of 60%, 20% and 20%.

For the support vector machines model instead, different kernel functions were tested, resulting in the selection of a linear kernel.

After having created the three models, the process of validation via observation of the residuals control charts highlighted anomalies in the behaviour of the system examined. The control limits were estimated as multiples of the standard deviation of the residuals' distribution in the period, assessed via moving range. Further analysis confirmed the occurrence of anomalies in the systems. Consequently, new models were created, eliminating the two anomalous time periods: 1-10 September 2017 and 8-11 October 2017.

The three new models were then applied to a new data set: from 15th January to 4th March 2018.

As an example, in Figure 4 is depicted the application of the recalculated statistical linear regression model to the complete dataset (1st September 2017 - 4th March 2018). It is possible to observe the anomalous periods excluded in both training procedures.

The performance indicators of the resulting models on the new training period are shown in Table 1. Overall the three models seem to have similar performances, with the ANN model showing slightly more precision. Moreover, the control limits, were narrower for the machine learning models: respectively 3.56 kWh/15min for the ANN, 3.94 kWh/15min for the SVM and 4.47 kWh/15min for the statistical linear regression.

Table 1: Performance indicators on the training period

	R^2	MSE $(kWh/15min)^2$	MAPE
Statistical linear regression	0.9698	6.28	66.1%
Artificial Neural Network	0.9734	5.54	17.7%
Support Vector Machines	0.9697	6.31	126.7%

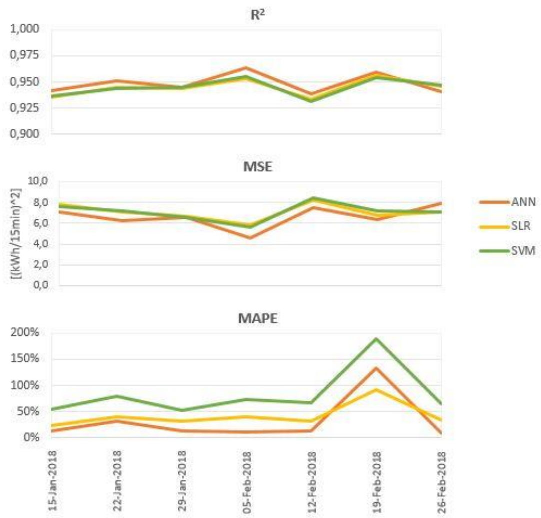

Figure 5: weekly estimation of the performance indicators (R^2, MSE, MAPE) in the test period (15 January-4 March)

Figure 5 shows the indicators estimated weekly in the test period: it appears that except for the week starting the 26th February, the ANN model retained a slightly better accuracy. Moreover, for a performance comparison, the models were tested simultaneously for real-time monitoring by assessing the reliability of the reported anomalies and the timeliness of their detections.

Indeed, the control charts in Figure 6 show similar profiles but with a different sensitivity. Furthermore, it can be observed that anomalies superior to the single point are always reported by all three models.

While the use of machine learning techniques enables the modelling of complex non-linear behaviours, probably due to the simplicity of the system, in this specific case, using only the three variables selected, the application of machine learning models does not hugely improve the control system performance. Nevertheless, the strength of these models is their ability to include a lot more variables, even categorical ones, to characterize the system's behaviour more accurately. It is possible, for example, to include information about the state of single air compressors, by using binary variables. Indeed, the inclusion of new variables made the models acquire more accuracy (see the new performance indicators in Appendix A). However, it is important to point out that the inclusion of new variables would change the aim of the control system: the new model would be normalized in reference to the new variables included. In this context, it could be interesting to use both types of models (with and without the new variables) to build a control system able to detect the probable reason of the anomalous behaviour by the comparison between the different residuals' profiles. Thus, the complexity of the model would be justified by the different objective of the control system.

In reference to the hardware and software requirements, in this case for the company to continuously enact a real-time control through machine learning techniques it would be necessary to install new equipment as well as new software to be able to modify the model in case of need.

4. Conclusions

In this paper a comparative analysis has been presented to investigate the multiple implications of the use of different types of modelling techniques in the specific case of a compressed air system examined for energy control purposes. The case highlighted how the choice of a technique is connected to the specific application and its context.

While continuously striving towards the paradigm of

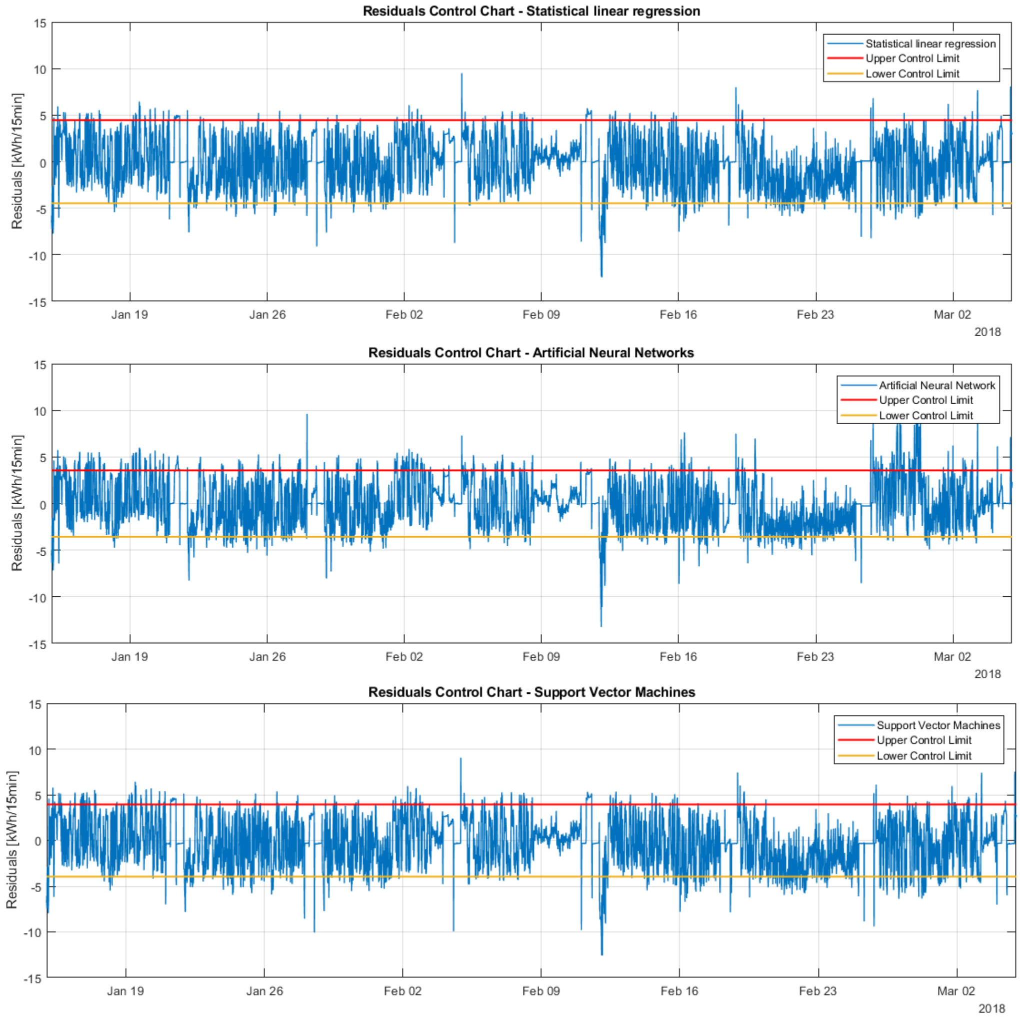

Figure 6: Residuals Control Chart created using the three models – Test period (15th January - 4th March 2018)

Industry 4.0, in fact, organizations are now trying to orient themselves and build their own path towards this goal. In this context, different new tools are made available everyday by the innovations in digitalization and machine learning and the resulting versatility of applications provides many possibilities of improvement in various aspects of industrial plants. According to the result of this research, simpler tools, based on statistical regression can successfully be implemented for the detection of main anomalies in common systems, whereas machine learning techniques as artificial neural networks or support vector machines can enable the implementation of additional functions such as failures analysis or even prescriptive maintenance methods. Furthermore, the development of sensors can foster low-cost implementations, bringing significant benefits to systems in industrial plant not yet affected by strategies for preventing inefficiencies and failures because deemed economically unfeasible. On the other hand, the need to adapt to different data availability lends itself to different levels of introduction in industrial plants.

Indeed, the proposed framework can and will be applied to further studies to investigate the application of these techniques in other contexts (e.g. chillers, boilers, HVACs, characterized by more complexity). The choice of tools will have to consider data availability, the specific systems' characteristics and measurement systems available. Analysing possible implementations by means of these categories can help determine easily specific drawbacks and advantages could emerge from their use, thus making a more conscious choice.

References

Benedetti, M., Cesarotti, V., Holgado, M., Introna, V., & Macchi, M. (2015). A Proposal for Energy Services' Classification Including a Product Service Systems Perspective. *Procedia CIRP*, 30, 251-256.

Benedetti, M., Cesarotti, V., Introna, V. and Serranti, J. (2016). Energy consumption control automation using Artificial Neural Networks and adaptive algorithms: Proposal of a new methodology and case study. *Applied Energy*, 165, 60-71.

Benedetti, M., Cesarotti, V. and Introna, V. (2017a). From energy targets setting to energy-aware operations control and back: An advanced methodology for energy efficient manufacturing. *Journal of Cleaner Production*, 167, 1518-1533

Benedetti, M., Bonfa', F., Bertini, I., Introna, V. and Ubertini, S. (2017b). Explorative study on Compressed Air Systems' energy efficiency in production and use: First steps towards the creation of a benchmarking system for large and energy-intensive industrial firms. *Applied Energy*.

Carbon Trust, 2012. CTV050 – Compressed air – Opportunities for businesses. Great Britain.

Cesarotti, V., Di Silvio, B., Introna, V. (2009). Energy budgeting and control: a new approach for an industrial plant. *International Journal of Energy Sector Management*, 3(2), 131-156.

Di Silvio, B., Cesarotti, V., Introna, V. (2007). Evaluation of electricity rates through characterization and forecasting of energy consumption: A case study of an Italian industrial eligible customer. *International Journal of energy sector management*, 1(4), 390-412.

Doty, S. and Turner, W. (2004). *Energy management handbook*. The Fairmont Press, Inc.

Edelenbosch, O. Y., Kermeli, K., Crijns-graus, W., Worrell, E., Bibas, R. and Fais, B. (2017). Comparing projections of industrial energy demand and greenhouse gas emissions in long-term energy models, *Energy*, 122, 701–710.

Geller, H., Harrington, P., Rosenfeld, A., Tanishima, S. and Unander, F. (2006). Polices for increasing energy efficiency: Thirty years of experience in OECD countries. *Energy Policy*, 34(5), 556-573.

IEA International Energy Agency, (2018). Market Report Series - Energy efficiency 2O17.

ISO International Standard Organization (2011). ISO 50001 Energy Management Systems - Requirements with Guidance for Use.

ISO International Standard Organization (2014). ISO 50006 Energy Management Systems - Measuring Energy Performance Using Energy Baselines (EnB) and Energy Performance Indicators (EnPI) - General Principles and Guidance.

Magoulès, F., & Zhao, H. X. (2016). *Data Mining and Machine Learning in Building Energy Analysis: Towards High Performance Computing*. John Wiley & Sons.

Montgomery, D. C. (2009). *Introduction to statistical quality control*. John Wiley & Sons, New York.

Morvay, Z. and Gvozdenac, D. (2008). *Applied industrial energy and environmental management*. Management. JohnWiley & Sons Ltd, United Kingdom.

Petrecca, G. (2014). *Energy Conversion and Management*. Springer, New York.

Salvatori, S., Benedetti, M., Bonfà, F., Introna, V. and Ubertini, S. (2018). Inter-sectorial benchmarking of compressed air generation energy performance: Methodology based on real data gathering in large and energy- intensive industrial firms. *Applied Energy*. Elsevier, 217, 266–280.

Shrouf, F. and Miragliotta, G. (2015). Energy management based on Internet of Things: practices and framework for adoption in production management. *Journal of Cleaner Production*. Elsevier Ltd, 100, 235–246.

Zhou, K., Fu, C. and Yang, S. (2016) 'Big data driven smart energy management: From big data to big insights'. Elsevier Ltd, 56, 215–225.

Appendix A. Performance indicators with categorical variables included in the machine learning models

Week	R^2		
	ANN	Stat. linear regression	SVM
15-Jan-2018	0.975	0.936	0.972
22-Jan-2018	0.979	0.945	0.970
29-Jan-2018	0.979	0.944	0.975
05-Feb-2018	0.989	0.954	0.981
12-Feb-2018	0.979	0.933	0.962
19-Feb-2018	0.984	0.957	0.983
26-Feb-2018	0.974	0.946	0.963

Week	MSE $(kWh/15min)^2$		
	ANN	Stat. linear regression	SVM
15-Jan-2018	3.0	7.8	3.4
22-Jan-2018	2.7	7.0	3.8
29-Jan-2018	2.5	6.6	2.9
05-Feb-2018	1.4	5.8	2.4
12-Feb-2018	2.6	8.2	4.7
19-Feb-2018	2.5	6.8	2.7
26-Feb-2018	3.4	7.1	4.9

Week	MAPE		
	ANN	Stat. linear regression	SVM
15-Jan-2018	9%	23%	59%
22-Jan-2018	34%	39%	77%
29-Jan-2018	11%	31%	53%
05-Feb-2018	9%	40%	76%
12-Feb-2018	11%	32%	70%
19-Feb-2018	77%	93%	229%
26-Feb-2018	13%	33%	71%