Comparison of Genetic and Reinforcement Learning Algorithms for Energy Cogeneration Optimization

Comparison of Genetic and Reinforcement Learning Algorithms for Energy Cogeneration Optimization

Giorgia Ghione
DET
Politecnico di Torino
Turin, Italy
giorgia.ghione@polito.it

Vincenzo Randazzo
DET
Politecnico di Torino
Turin, Italy
vincenzo.randazzo@polito.it

Alessandra Recchia
DENERG
Politecnico di Torino
Turin, Italy
alessandra.recchia@studenti.polito.it

Eros Pasero
DET
Politecnico di Torino
Turin, Italy
eros.pasero@polito.it

Marco Badami
DENERG
Politecnico di Torino
Turin, Italy
marco.badami@polito.it

***Abstract*—Large process plants generally require energy in different forms: mechanical, electrical, or thermal (in the form of steam or hot water). A commonly used source of energy is cogeneration, also defined as Combined Heat and Power (CHP). Cogeneration can offer substantial economic as well as energy savings; however, its real-time operation scheduling is still a challenge today. Multiple algorithms have been proposed for the CHP control problem in the literature, such as genetic algorithms (GAs), particle swarm optimization algorithms, artificial neural networks, fuzzy decision making systems and, most recently, reinforcement learning (RL) algorithms.**

This paper presents the comparison of a RL approach and a GA for the control of a cogenerator, using as a case study a thermal power plant serving a factory during the year 2021. The two methods were compared based on an earnings before interest, taxes, depreciation, and amortization (EBITDA) metric. The EBITDA that could be obtained using the RL algorithm, exceeds both the EBITDA that could be generated using a per-week genetic algorithm and the one from the manual scheduling of the CHP. Thus, the RL algorithm proves to be the most cost-effective strategy for the control of a CHP.

***Keywords*—CHP, cogeneration optimization, EBITDA, energy cogeneration, genetic algorithm, neural networks, reinforcement learning.**

I. Introduction

Large process plants generally require energy in different forms: mechanical, electrical, or thermal (in the form of steam or hot water). These energies very often come from various types of energy sources such as gas turbine generators, steam turbine generators, boilers and internal combustion engines. A commonly used source of energy in both industrial sectors and residential applications [1] is cogeneration, also defined as combined heat and power (CHP), which is the combined production of thermal and electrical/mechanical energy in a single process from a single primary energy source [2]. For electricity, the national grid acts as a source of supplementary electricity when the production facilities available to the industry do not produce enough electricity and, if necessary, as a sink when excess electricity is produced. The variety of energy sources and their interdependence, as well as their changing technical and economic conditions over time, make energy cost reduction strategies non-trivial, as they are one of the main factors in the total cost of ownership of a process plant. Cogeneration can offer substantial economic as well as energy savings [3]. However, its real-time control is still a challenge today, as explained in [4] and [5].

At the state of the art, different approaches are used to control energy systems, which can be classified into three classes: white box, data-driven (black box), and grey box models. White box models, also known as model-based control strategies, apply physical principles to represent the relationship between model inputs and outputs during the control process [6]. However, the lack of good quality data and the complexity of energy systems is an obstacle to the adoption of such strategies. Data-driven models, also known as black box control methods, use knowledge derived from online or offline data processing instead of depending on the explicit or implicit information of the mathematical model [7]: as an example, Reinforcement Learning (RL) [8] belongs to this category. Finally, grey box models are those found between white and black box models: models based on fuzzy logic [9] are among those classified in this category. Multiple algorithms have been proposed for the CHP control problem in the literature, which are generally non-linear, with multi-modal objective functions and can contain both discrete and continuous variables. Some of the most commonly used methods are the following: genetic algorithms (GAs), such as the self adaptive real-coded GA proposed in [10] and the real-coded genetic algorithm using an improved Mühlenbein mutation in [11]; Particle Swarm Optimization, such as the multi-objective PSO model in [12] and the set of improved PSO algorithms in [13]; Artificial Neural Networks such as the multi-layer feedforward neural network for the simulation and optimization of cogeneration in [14], the ANN for the re-modeling and optimization of the

cogeneration system in [15]and an feedforward neural network based on the output of a physics-based model in [16]; fuzzy decision making systems such as [17] and, most recently, RL algorithms such as the offline RL-based optimization model proposed in [18], the RL algorithm for the control of a district heating network powered by a CHP [19] and an improved Distributed Proximal Policy Optimization in [20].

This work presents a comparison between a RL approach and a GA optimization approach for the control of a CHP based on the energy demands of an industrial plant. The considered case study is a cogenerator located in a thermal power plant in Italy. The combined production of electricity and heat takes place via a reciprocating engine, combined with heat recovery circuits for the production of hot water, saturated steam and low-temperature. The objective of the task under analysis is to identify the most cost-effective CHP operation scheduling approach.

The first strategy used is the off-line computation of the best CHP operation scheduling of a day or a week using a GA. The second strategy used is an on-line RL-based controller, which learns a control strategy to handle the hourly load factor of the CHP in real time.

II. System description

The case study for this work is a cogenerator plant with an internal combustion engine, located in a thermal power plant serving a factory in the Lombardy region in Italy, which produces adhesive systems for industry and consumer goods. The cogeneration plant consists of one ECOMAX 12 NGS cogeneration module, derived from GE JENBACHER JGS 416 GS-N.L. The thermal power plant serving the industry consists, as far as the production part is concerned, of the reciprocating engine and three boilers. The combined production of electricity and heat takes place via a reciprocating engine with a nominal full-load power output of 1203 electric kW, with a nominal thermal energy recovery of 1385 kW. The cogeneration unit operates based on the four-stroke Otto cycle, consumes natural gas and it is connected to the power supply network of the national power grid in parallel with a 15000V voltage.

In the ECOMAX 12 NGS cogeneration module, the first stage of heat recovery takes place within the engine block (lubricating oil circuit, engine jacket water circuit, intercooler first stage circuit), from which it was planned at the design stage to recover about 774 kW. Such power recovery, when added to the 97 kW offered by the preheating coil, enables the production of hot water at about 85 °C for the plant utilities. Combustion fumes exiting the engine block are sent to a shell-and-tube heat exchanger capable of producing 396 kW (including 41 kW offered by the economizer) in the form of saturated steam at about 175°C, which is entirely self-consumed by the plant. Finally, by means of an additional exchanger, the thermal energy of the second intercooler stage, amounting to 118 kW, is recovered in the form of low-temperature hot water at about 29.5 °C, which is self-consumed by the plant too.

Currently, the CHP is manually controlled by plant technicians based on their past experience. The heat and power produced by the CHP are approximately entirely self-consumed, therefore it is assumed that the CHP plant is a high-efficiency cogeneration plant in accordance with the directive 2012/27/EU of the European Parliament [21]. Based on this assumption, it is possible to further assume that the CHP plant has dispatching priority on the national electric power grid with respect to traditional sources of power [21].

III. Methodology

A. Dataset

The energy demands data of the considered plant refer to the whole 2021 calendar year and have an hourly granularity, for a total of 8760 data points. Each data point comprises five features:

- the electrical power demand of the plant;
- the heat demand in the form of steam;
- the heat demand in the form of high-temperature hot water;
- the heat demand in the form of low-temperature hot water;
- the heat demand of the degasser.

The real manually-scheduled load factor of the CHP during year 2021 was used to compute a benchmark for the evaluation of the results, as will be described in Sec. IV.

To achieve good performance under various scenarios of energy demand from the plant in the RL algorithm, in order to simulate all possible operating states of the CHP system, the agent was not trained on the real dataset but the trends of electrical and thermal energy demand were modified to generate a new training dataset. The electricity demand of the odd months was reduced to 50% and then 20% in 2000-hour intervals. In the remaining hours it was finally reduced to 10%. The thermal demand of the odd months was kept at 100% for the first 500 hours, then reduced to 50%, 25%, 5% at 500-hour intervals. This process was repeated on the rest of the dataset. As a test dataset, the whole original dataset was used.

B. Objective function

The objective function to be maximized is the earnings before interest, taxes, depreciation, and amortization (EBITDA) metric [22]. The revenues of the plant included in the formulation of the EBITDA are those due to Energy Efficiency Credits (EECs, i.e. documents that attest that a certain decrease of energy consumption has been achieved), the sale of exported electricity to the grid, as well as the avoided cost of natural gas for the production of thermal energy and the avoided cost of electricity purchased from the grid. The costs included in the EBITDA are those due to the purchase of natural gas feeding the cogenerator, and maintenance and operating costs of using the cogeneration plant. The self-consumption of heat was considered in the calculation of the EBITDA with respect to the directive 2012/27/EU of the European Parliament [21] on high efficiency cogeneration, relating to the calculation of

the primary energy saving. The detailed formulation of the EBITDA is the following:

$$\begin{aligned} EBITDA = & R_{EEC} * EEC + \\ & + R_{kWh} * E_{el_{sold\ to\ network}} + \\ & + C_{GN} * \frac{E_{th_{CHP}}}{9.8 * \eta_{IB}} + \\ & + C_{kWh} * (E_{el_{CHP}} - E_{el_{sold\ to\ network}}) + \\ & - C_{GN} * V_{NG_{CHP}} + \\ & - C_{O\&M} * h \end{aligned} \quad (1)$$

where EEC is the number of EECs for the performance of the CHP; R_{EEC} is the price of EECs; R_{kWh} is the per-kWh revenue from selling the electric energy produced by the CHP to the network; $E_{el_{sold\ to\ network}}$ is the electric energy produced by the CHP and sold to the network; C_{NG} is the cost of natural gas; $E_{th_{CHP}}$ is the thermal energy produced by the CHP; η_{IB} is the efficiency of integration boilers; C_{kWh} is the cost of electric energy purchased from the network; $E_{el_{CHP}}$ is the electric energy produced by the CHP during its operating time; $V_{NG_{CHP}}$ is the volume of natural gas consumed by the CHP (computed as the ratio between the electric energy produced by the CHP and the product of the electric efficiency of the CHP and the lower calorific value of natural gas); $C_{O\&M}$ is the cost for operations and maintenance of the CHP; h is the number of operating hours of the CHP (defined as the hours where the load factor of the CHP is greater than zero). In the calculation of EBITDA in the experiments reported below, all incentive aspects of the Italian legislation related to the primary energy saving directive were taken into account. Regarding the costs of natural gas, self-consumed electricity, electricity sold to the grid, and operation and maintenance, point values for the year 2021 were used, in accordance with the dataset collection period. The parameters affected by the CHP scheduling are $E_{el_{CHP}}$, $E_{el_{sold\ to\ network}}$, $V_{NG_{CHP}}$, $E_{th_{CHP}}$, h, and EEC.

C. Genetic algorithm

GAs are meta-heuristic methods that apply the principles of biological evolution processes to solve optimization problems. [23] By representing a set of points in the solution space in the form of a population of chromosomes, GAs allow iteratively obtaining solutions which evolve toward a point of optimum. A fitness function is used to assign a value to chromosomes and select them for further processing, and recombination strategies, namely crossover and mutation, are applied in the creation off-springs.

A brief description of the employed GA is provided in the following:

- An individual is a vector of 24 numbers belonging to the set α = {0, 0.50, 0.55, 0.60, ..., 0.95, 1}, representing the load factor of the cogenerator for all the 24 hours of a day.
- The fitness function is the EBITDA and is calculated over a whole year.
- A new generation is created by applying one-point crossover to the two fittest individuals, followed by random mutations on the newly generated individuals.
- At each generation, the population is updated by removing the two least fit individuals from the previous population and adding the new generation of individuals.
- The initial population is composed of randomly initialized individuals.
- The maximum number of generations used is 5000.
- The mutation rate is 0.04.
- The number of individuals in the population is 50.

The GA was fed with the energy demands of a previous time interval (day or week) and the optimal solution found was applied in the subsequent time interval for running the CHP.

D. Reinforcement Learning algorithm

Reinforcement learning is a machine learning paradigm aimed at finding learning strategies that maximize a numerical reward while an agent explores and interacts with an environment [24]. The main elements that constitute a RL algorithm are the following: a policy, defining the computations which determine the action the agent performs at each time step; a reward signal, i.e. the goal of the optimization problem; a value function, indicating the amount of reward that can be obtained from a certain state; and, optionally, a model of the environment to make inferences on future scenarios.

In the context of CHP scheduling, the environment is the mathematical model of the cogeneration system, the agent is the control system, its actions are the variations of the load factor of the CHP, and the reward is the EBITDA value at each time step based on the action of the agent in the environment. The RL model which was selected for this analysis is a Deep Q Network (DQN) [25], available in the Keras-RL library [26]. DQN integrates Q-learning [27] with neural networks and experience replay to maximize the reward of actions performed by an agent in an environment. Q learning is an off-policy model-free algorithm, which applies to the Reinforcement Learning problems that do not require any model of the environment [27].

In this work, the environment is defined by a state variable for each hour, composed by six elements:

- the electrical power demand;
- the heat demand in the form of steam;
- the heat demand in the form of high-temperature hot water;
- heat demand in the form of low-temperature hot water;
- the heat demand of the degasser;
- the load factor of the CHP.

The first five elements of the state variables are the input dataset features, while the load factor is initially set as a random value belonging to the set α = {0, 0.50, 0.55, 0.60, ..., 0.95,1} (the value 0 indicates that the CHP is switched off). The set of actions the agent can perform is defined as the set of possible variations of the load coefficient δ_α = {-0.5, -0.45, -0.40, ..., 0.40, 0.45, 0.50} and their application is

constrained to the limits defined by α. The reward function to be maximized is the EBITDA. The type of neural network implemented within the model presented in this section is a multi-layer perceptron [28] with two hidden layers composed of 128 and 64 units respectively and with ReLU activation [29]. The optimization function used is Adam and the learning rate is $5 * 10^{-}4$. The agent uses an ϵ-greedy Q exploration policy associated with a linear annealing strategy [30] to choose actions. The value of epsilon decreases linearly over 170 episodes from a value of 1 to a value of 0.01. The memory of experience replay was set to 1000. The target Q model is updated every 400000 steps. The RL algorithm receives as input the data of the current moment and is able to determine the best variation of the CHP load factor for the next hour to maximize the EBITDA. Then, due to the use of the Q policy, it continuously (each hour) updates the neural network weights during training to correct the load factor scheduling.

IV. Results

The two approaches were compared on a real EBITDA maximization dataset. In both cases, the benchmark for the evalutation of the results is the real EBITDA that the plant obtained by manually schedulng the load factor of the CHP during the year 2021, which amounted to 630702€. This value is, of course, not the highest obtainable one, due to the manual control of the CHP; however, since it was achieved via a traditional human-based approach, it can be considered as a minimum boundary for the EBITDA.

The results of the application of the GA and the DQN are reported in the following two subsections, respectively. The comparison of the results is provided in Sec. IV-C.

A. Genetic algorithm

The GA has been applied to three different scenarios:

1) Whole dataset: all available data were fed to the GA; the system was tested on the same dataset to obtain the yearly EBITDA.
2) Only data of the previous day: data of the previous day were fed to the GA; the system was tested on the data of the subsequent day; such a process was repeated over the whole dataset to obtain the yearly EBITDA.
3) Only data of the previous week: data of the previous week were fed to the GA; the system was tested on the data of the subsequent week; similarly also in this case such a process was repeated over the whole dataset to obtain the yearly EBITDA.

Table I shows the yearly EBITDA values obtained in each scenario. It may seem that the first approach performs better than the benchmark (660798€ vs 630702€); however, it represents an ideal scenario where the whole energy demands are known in advance for the whole year. In this sense, it was considered in order to have a benchmark for comparison that is close to the maximum EBITDA that could be obtained for the year 2021.

On the other hand, it is possible to observe that both the computations of the GA using data from the previous day or the previous week do not improve the obtained EBITDA of the year, causing a 32% and 20% loss respectively: such a result is significantly lower than the benchmark EBITDA obtained with CHP manual scheduling.

TABLE I
GA EBITDA results on different inputs

Data used to compute the GA	EBITDA (€)
Whole dataset	660798
Previous day	451388
Previous week	528911

In the third case (only data of the previous week), a greater EBITDA can be related to the fact that the energy demands of the plant have a weekly periodicity.

Figs. 1(a) and 1(b) show the scheduling of the CHP (purple graph) computed using data of the previous day (a) and data of the previous week (b), with respect to the electrical power demand of the plant (green graph). As expected, the larger backward window of the latter case enables a more precise scheduling of the CHP. A situation where the daily scheduling of the CHP provides a better result is highlighted in the zoomed-in parts of the plots: in Fig. (a) it is possible to observe that the one-week delay in the application of the GA causes significant losses in the exceptional case when there was no power demand in the previous week; on the contrary, in picture (b) the scheduling of the CHP is overall more coherent with respect to the electrical power demand. However, this situation is a rare occurrence in the energy demands of the factory over a whole year.

In conclusion, it can be stated that the GA is not able to reach a satisfactory optimization performance without knowing the future energy demands. In a real case scenario, an accurate forecast system of factory demands is not available at the beginning of the operating period; however, if such a tool was in place, the GA could be a significant cost-effective tachnique for maximizing the EBITDA.

B. Reinforcement Learning algorithm

As explained in Sec. III-A, to simulate all possible operating states of the CHP system, the RL agent was trained on the modified dataset and tested on the whole real dataset.

The RL algorithm was trained and tested in six separate experiments to ensure result consistency; this was possible due to the faster convergence time of the RL algorithm with respect to the genetic one. The mean obtained EBITDA is equal to 640040€, with a standard deviation equal to 1181€: it is only, on average, 3% lower than the maximum EBITDA which could be obtained in the ideal condition of computing the GA on the whole dataset. Additionally, the mean EBITDA is higher than the manually optimized benchmark by 1.5%. Furthermore, the standard deviation is very low compared to the mean EBITDA, which indicates that the RL algorithm has a consistent optimization performance even if it works in a real case scenario with random initial conditions.

Fig. 2 shows the scheduling of the CHP (purple plot), computed using the DQN algorithm, compared to the electrical

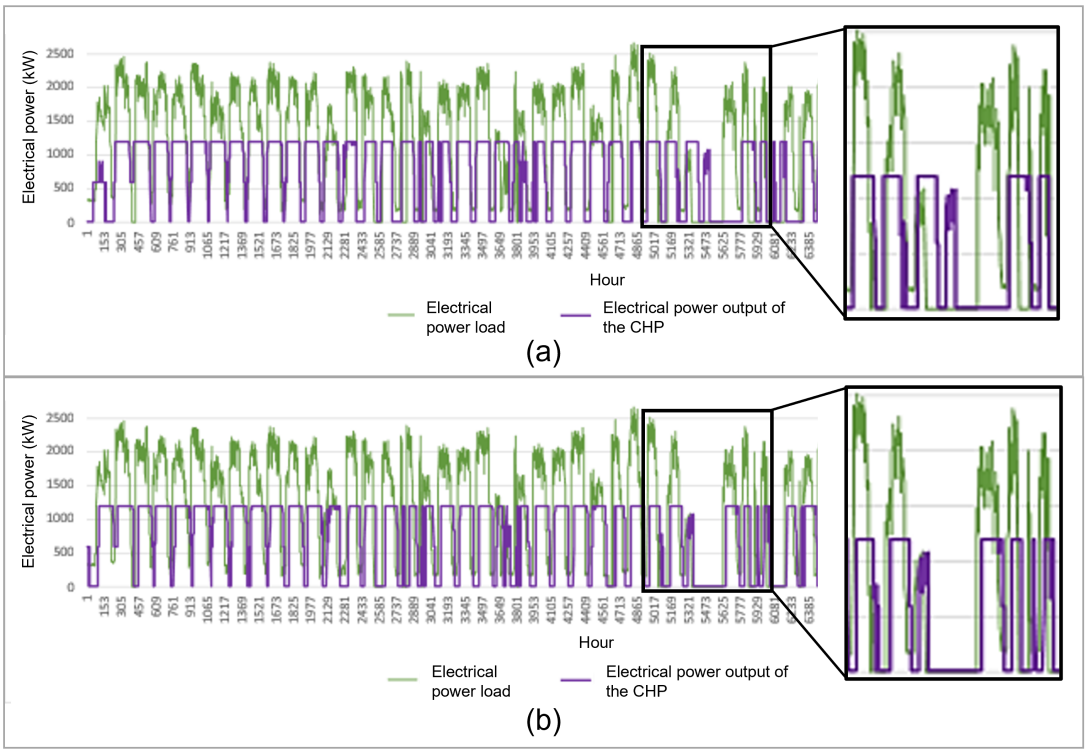

Fig. 1. Figs. (a) and (b) show the scheduling of the CHP (purple plot) with respect to a single power demand, i.e. the electrical power demand (green plot), for a clearer visualization. In particular, graph (a) shows the scheduling of the CHP (purple plot) computed with the GA using the data of the previous week as input, with respect to the electrical power demand of the plant (green plot). Graph (b) shows the scheduling of the CHP (purple plot) computed with the GA using the data of the previous day as input, with respect to the electrical power demand of the plant (green plot). The zoomed-in portion of graph (a) shows that the one-week delay in the application of the GA causes significant losses in the exceptional case when there was no power demand in the previous week. On the contrary, the zoomed-in portion of graph (b) it is possible to observe that the scheduling of the CHP is more coherent with respect to the electrical power demand in that situation. However, the weekly scheduling of the CHP enables a more precise scheduling of the CHP overall, due to the weekly periodicity of energy demands.

power demand of the plant (green plot). In this case, the scheduling of the CHP follows the electrical power demand better. Such an improvement can derive from the use of continuous feedback during training, which led to a better approximation of the lowest points of the power demand curve.

C. Comparison

Table II summarizes the EBITDAs obtained by applying the GA using, as input, the whole dataset, data of the previous day or data of the previous week, as well as EBITDA of the test of the DQN model. In addition, the benchmark value, i.e. the real EBITDA obtained by manually scheduling the CHP in the year 2021, is reported in the first row. As stated in Sec. IV-A, the case of GA fed with the whole dataset represents an ideal scenario where the whole energy demands are known in advance for the entire year. Thus, it has been considered only as a further benchmark that is close to the maximum EBITDA that could be obtained for the year 2021.

The scheduling of the CHP which provided the best performance in terms of greater EBITDA is the one computed with DQN: if this method had been used to optimize the scheduling of the CHP in the year 2021, the total EBITDA would have amounted to 640040€, which exceeds both the EBITDA of the per-week GA and the one of the manual scheduling of the CHP. Conversely, the per-day and per-week GAs resulted in a lower EBITDA compared to both the manual scheduling and the DQN scheduling of the CHP: therefore, they are not cost-effective strategies for the control of the CHP.

TABLE II
COMPARISON OF EBITDA RESULTS.

Algorithm	**EBITDA (€)**
Manual scheduling	630702
GA on whole dataset	660798
GA on previous day	451388
GA on previous week	528911
DQN	640040

V. CONCLUSION

This paper presents the comparison of a RL and a GA optimization approach for the control of a cogenerator, using a thermal power plant serving a factory during the year 2021 as a case study . The first strategy used is the off-line computation of the best hourly CHP operation scheduling of a day or a week using a GA. The EBITDA obtained from both the per-day and the per-week scheduling was significantly lower than

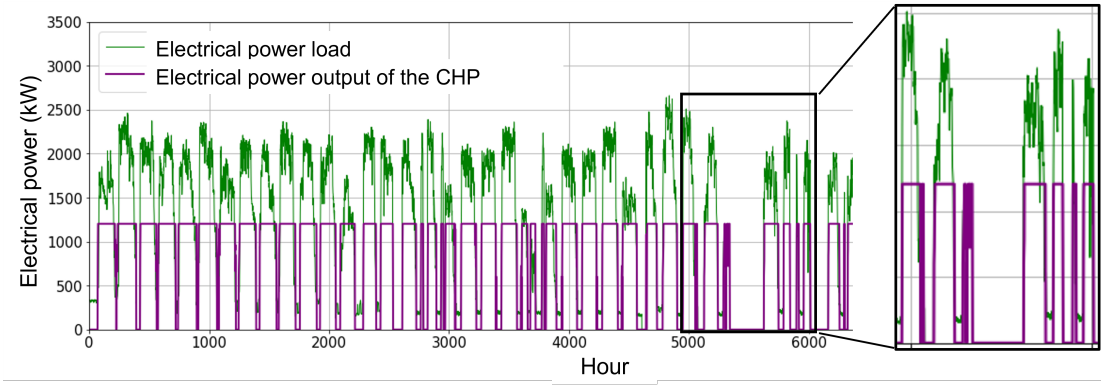

Fig. 2. The scheduling of the CHP (purple plot) computed using the RL algorithm, compared to the electrical power load of the plant (green plot). The zoomed-in portion of the graphs shows that the scheduling of the CHP follows better the electrical power demand.

the real EBITDA generated by manually scheduling the CHP during 2021. The second strategy used is an on-line RL-based controller, which learns a control strategy to handle the hourly load factor of the CHP based on the data of the previous hour. The RL algorithm that was chosen is a DQN. In this case, the mean EBITDA that could be obtained amounted to 640040€, which exceeds both the EBITDA from the per-week GA and the EBITDA from the manual scheduling of the CHP. Thus, the DQN proves to be the most cost-effective strategy for the control of a CHP, compared to the use of a GA or manual scheduling.

Future works will involve comparing different energy demand forecasting strategies to be integrated in the two proposed techniques of this paper. Additionally, a bigger real dataset will be collected to further test the two optimization approaches.

ACKNOWLEDGMENT

Dr. Randazzo acknowledges funding from the research contract no. 32-G-13427-2 (DM 1062/2021) funded within the Programma Operativo Nazionale (PON) Ricerca e Innovazione of the Italian Ministry of University and Research. Giorgia Ghione acknowledges funding from the European Union – NextGenerationEU and Trigenia S.r.l. Finally, a special thanks to Valeria Pellerey and Gioele Porro of Trigenia s.r.l. for providing data and supporting this work.

REFERENCES

[1] M. Badami, G. Chicco, A. Portoraro, and M. Romaniello, "Micro-multigeneration prospects for residential applications in italy," *Energy Conversion and Management*, vol. 166, pp. 23–36, 2018.

[2] S. D. Hu, *Cogeneration*. Prentice Hall Inc., Old Tappan, NJ, 1985.

[3] J. H. Horlock, *Combined heat and power*. Pergamon Books Inc., Elmsford, NY, 1987.

[4] M. A. Bagherian, K. Mehranzamir, A. B. Pour, S. Rezania, E. Taghavi, H. Nabipour-Afrouzi, M. Dalvi-Esfahani, and S. M. Alizadeh, "Classification and analysis of optimization techniques for integrated energy systems utilizing renewable energy sources: A review for chp and cchp systems," *Processes*, vol. 9, no. 2, 2021.

[5] E. Abdollahi, H. Wang, and R. Lahdelma, "An optimization method for multi-area combined heat and power production with power transmission network," *Applied Energy*, vol. 168, pp. 248–256, 2016.

[6] J. DeCarolis, H. Daly, P. Dodds, I. Keppo, F. Li, W. McDowall, S. Pye, N. Strachan, E. Trutnevyte, W. Usher, M. Winning, S. Yeh, and M. Zeyringer, "Formalizing best practice for energy system optimization modelling," *Applied Energy*, vol. 194, pp. 184–198, 2017.

[7] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, aug 2018.

[8] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[9] L. A. Zadeh, "Fuzzy logic," *Computer*, vol. 21, no. 4, pp. 83–93, 1988.

[10] P. Subbaraj, R. Rengaraj, and S. Salivahanan, "Enhancement of combined heat and power economic dispatch using self adaptive real-coded genetic algorithm," *Applied Energy*, vol. 86, no. 6, pp. 915–921, 2009.

[11] A. Haghrah, M. Nazari-Heris, and B. Mohammadi-Ivatloo, "Solving combined heat and power economic dispatch problem using real coded genetic algorithm with improved mühlenbein mutation," *Applied Thermal Engineering*, vol. 99, pp. 465–475, 2016.

[12] G. Piperagkas, A. Anastasiadis, and N. Hatziargyriou, "Stochastic pso-based heat and power dispatch under environmental constraints incorporating chp and wind power units," *Electric Power Systems Research*, vol. 81, no. 1, pp. 209–218, 2011.

[13] T. Nguyen Trung and D. Vo Ngoc, "Improved particle swarm optimization for combined heat and power economic dispatch," *Scientia Iranica*, vol. 23, no. 3, pp. 1318–1334, 2016.

[14] R. Zomorodian, M. Rezasoltani, and M. B. Ghofrani, "Static and dynamic neural networks for simulation and optimization of cogeneration systems," *International Journal of Energy and Environmental Engineering*, vol. 2, no. 1, pp. 51–61, 2011.

[15] A. Jamali, P. Ahmadi, and M. N. M. Jaafar, "Optimization of a novel carbon dioxide cogeneration system using artificial neural network and multi-objective genetic algorithm," *Applied Thermal Engineering*, vol. 64, no. 1-2, pp. 293–306, 2014.

[16] M. J. Kim, T. S. Kim, R. J. Flores, and J. Brouwer, "Neural-network-based optimization for economic dispatch of combined heat and power systems," *Applied Energy*, vol. 265, p. 114785, 2020.

[17] H. Sayyaadi, M. Babaie, and M. R. Farmani, "Implementing of the multi-objective particle swarm optimizer and fuzzy decision-maker in exergetic, exergoeconomic and environmental optimization of a benchmark cogeneration system," *Energy*, vol. 36, no. 8, pp. 4777–4789, 2011.

[18] G. Zhang, C. Zhang, W. Wang, H. Cao, Z. Chen, and Y. Niu, "Offline reinforcement learning control for electricity and heat coordination in a supercritical chp unit," *Energy*, vol. 266, p. 126485, 2023.

[19] A. Mugnini, F. Ferracuti, M. Lorenzetti, G. Comodi, and A. Arteconi, "Advanced control techniques for chp-dh systems: A critical comparison of model predictive control and reinforcement learning," *Energy Conversion and Management: X*, vol. 15, p. 100264, 2022.

[20] S. Zhou, Z. Hu, W. Gu, M. Jiang, M. Chen, Q. Hong, and C. Booth, "Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach," *International Journal of Electrical Power & Energy Systems*, vol. 120, p. 106016, 2020.

[21] "Directive 2012/27/eu of the european parliament and of the council of 25 october 2012 on energy efficiency, amending directives 2009/125/ec and 2010/30/eu and repealing directives 2004/8/ec and 2006/32/ec text with eea relevance," *OJ*, vol. L 315, pp. 1–56, 212-11-14.
[22] J. Bouwens, T. De Kok, and A. Verriest, "The prevalence and validity of ebitda as a performance measure," *Comptabilité-Contrôle-Audit*, vol. 25, no. 1, pp. 55–105, 2019.
[23] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, pp. 8091–8126, 2021.
[24] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
[25] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[26] M. Plappert, "keras-rl." https://github.com/keras-rl/keras-rl, 2016.
[27] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, pp. 279–292, 1992.
[28] S. Haykin, *Neural Networks and Learning Machines*. No. v. 10 in Neural networks and learning machines, Prentice Hall, 2009.
[29] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.
[30] F. Leibfried and P. Vrancx, "Model-based regularization for deep reinforcement learning with transcoder networks," 2018.