Energy & Buildings 309 (2024) 114039

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Energy consumption prediction of industrial HVAC systems using Bayesian Networks

Francesco Giuseppe Ciampi [a], Andrea Rega [a,*], Thierno M.L. Diallo [b], Francesco Pelella [a], Jean-Yves Choley [b], Stanislao Patalano [a]

[a] Department of Industrial Engineering, University of Naples Federico II, 80125 Naples, Italy
[b] Quartz Laboratory (EA7393), ISAE-Supméca, 93400 Saint-Ouen, France

ARTICLE INFO

Keywords:
Industrial engineering
Energy consumption prediction
Machine learning
Bayesian networks
HVAC systems

ABSTRACT

Predicting energy consumption has become a critical issue for energy-intensive industrial contexts. A significant contribution to their overall energy load is due to the Heating Ventilation and Air Conditioning (HVAC) systems. This work, therefore, aims to validate the applicability of a probabilistic graphical approach, the Bayesian Network, in predicting the HVAC systems' energy consumption. As a data-driven approach, it is compared with more common AI-based models like Support Vector Machine, Artificial Neural Networks and Random Forest. The graphical approach ensures a better interpretation of the main factors determining the energy consumption and the relationships underlying these dependences. After an initial contextualisation and an analysis of the state of the art, the design methodology of a Bayesian network is investigated in detail, deepening in the various solutions for each step and evaluating their performance through the application on two industrial case studies. The results show that Bayesian networks, despite not always providing the best results, are a valid solution, trading off between simplicity, flexibility, and performance. Moreover, the possibility to provide a physical interpretation of the results is one of its main strengths. The critical aspect encountered, instead, is the need for discretisation, which strongly influences the quality of the results.

1. Introduction

The growing interest in the efficient management of energy resources in industry can be associated to several aspects, such as new trends in environmental policies, a strong fluctuation in prices or the increasing importance of energy in the overall production costs. Whatever the real causes, there is no doubt that a correct forecast of consumption is required to better plan the supply and usage of energy [1]. This study, therefore, aims to focus, within the industrial context, on the Heating, Ventilation and Air Condition (HVAC) systems that are, in many cases, among the most energy-intensive and hardest-to-optimise systems. Industrial HVAC systems, in fact, must ensure that two main requirements are met: the comfort of the personnel working in the facility and the environmental conditions required in specific processes. Although there are some differences between the two use cases, mainly due to the tolerance ranges, in both, the HVAC system must guarantee that temperature, humidity and air quality are maintained around specific setpoints [2]. Currently, this is mostly done through non-optimised control, often based on expert knowledge. The consumption prediction could then represent a relevant aspect to achieve optimization and savings. In the industrial context, it has to be considered that HVAC systems have a different energy consumption behaviour compared to traditional machines or processes. In fact, the use of HVAC systems is highly dependent on external factors such as seasonality and climatic conditions. This leads to greater difficulty in forecasting.

Looking into the models available for this goal, a first distinction should be made between prediction and forecasting. In the first one, the output is estimated by knowing the inputs in the same time step, whereas in the second case, inputs in previous time steps are used [3]. In this case, prediction has been chosen because it offers the most accurate estimate of energy consumption. Indeed, energy consumption at a given time depends largely on the values of certain input parameters (weather, occupancy of buildings, occupant needs, etc.) at that time. The HVAC system's consumption will therefore vary according to these parameters. To obtain the best estimate of consumption at a given time, it is therefore better to predict it on the basis of the input values at that time. Some of these inputs, such as outdoor conditions, may be forecasts. Among

* Corresponding author.
E-mail address: andrea.rega@unina.it (A. Rega).

https://doi.org/10.1016/j.enbuild.2024.114039
Received 16 November 2023; Received in revised form 29 January 2024; Accepted 24 February 2024
Available online 25 February 2024

Abbreviations

ANN	Artificial Neural Network
AR	Autoregressive model
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BDeu	Bayesian-Dirichlet equivalent uniform
BIC	Bayesian Information Criterion
BN	Bayesian Network
CNN	Convolutional Neural Network
CPD	Conditional Probability Distribution
CPT	Conditional Probability Table
CV-RMSE	Coefficient of Variation of the Root Mean Square Error
DAG	Directed Acyclic Graph
DL	Deep Learning
EFD	Equal Frequency Discretization
EWD	Equal Width Discretization
HVAC	Heating, Ventilation and Air Conditioning
KPI	Key Performance Indicator
LR	Linear Regression
LSTM	Long Short-Term Memory
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MLP	Multiple Layer Perceptron
MLR	Multiple Linear Regression
MMHC	Max-Min Hill Climb
NLR	Non-Linear Regression
NP	Non-deterministic Polynomial-time (hardness)
PC	Peter-Clark (algorithm)
PSO	Particle Swarm Optimization
ReLU	Rectified Linear Unit (activation function)
RF	Random Forest
RM	Regression Model
RNN	Recurrent Neural Networks
SVM	Support Vector Machine

prediction models, therefore, 3 main types can be considered: physical, data-driven and hybrid models [4], corresponding, respectively, to the white box [5], black box, and grey box designations [6]. Black box models allow predictions to be made only based on historical data, without deepening the physics of the system and this enables solutions to be easily generalised. However, it is hard to investigate the reasons of the prediction results [6]. A partial solution may be the use of graphical probabilistic models, such as Bayesian networks, which, compared with classical black box models, are easier to interpret, due to their graphical structure [7].

This paper, therefore, presents the application of Bayesian Networks in the prediction of HVAC energy consumption to analyse the methodology and to validate the model applicability as a general data-driven solution and, on the other hand, make some observations on the physical interpretation of the prediction model.

This paper is organised as follows: Section 2 presents a summary of the state of the art in the field of interest. Section 3 describes Bayesian Networks and their design process. Section 4 introduces the case studies, then the results obtained in the applications are presented in section 5. Finally, section 6 shows the conclusions and lays the foundations for future works.

2. Background and related work

Research in the field of energy consumption prediction models is very prolific and providing a clear but brief overview implies the need to select a limited number of sources. Specifically, several reviews have been analysed, some more general [6,8–10] and others specific on data-driven models [11–14]. Based on them, a simple classification is proposed, considering 3 categories: conventional, AI-based and hybrid models.

2.1. Conventional models

Two main types, among all, are labelled as conventional: Time Series (TS) and Regression Models (RM). Among TS models, Auto-Regressive (AR), Moving Average (MA), Auto-Regressive Moving Average (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA) worth to be mentioned [13]. The work of Penya et Al. (2011) [15] compares AR and ARIMA with AI-based models in non-residential building energy load prediction, experiencing good results and verifying the already established viability of TS approaches. For RM, Linear Regression (LR), Multiple Linear Regression (MLR) and Non-Linear Regression (NLR) can be included [13]. Lei et Al. (2009) [16] evaluates these approaches in office building energy consumption forecasting, comparing single and multiple variable models. Bracale et Al. (2019) [17] applies MLR to industrial reactive power forecasting and verifies that the choice of the most suitable model and the obtained performance varies depending on the case study.

2.2. AI-based models

This category is extremely wide and exhaustive reviews can be found in literature [8,18]. In this paper, among others, only 3 sub-categories are presented: Neural Networks, Support Vector Machine and Bayesian Models [8]. The reason of this tight selection is that the most used approaches fall into the first two groups, while the third includes the model of interest: the Bayesian Networks.

Neural networks, developed as emulation of the human neural system, are the most common solution for non-linear problems. They consist of at least three layers: input, hidden and output layers. With a single hidden layer, the models are simply called Artificial Neural Networks (ANN), otherwise they are considered Deep Learning (DL) models. Among ANNs, the simplest approach is Multiple Layer Perceptron (MLP) whereas, within DL models, can be mentioned Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) as increasingly popular approaches [11]. Ding et Al. (2018) [19] summarizes the results obtained by MLP in commercial building heating load prediction, finding good performances in the ANN approach. Somu et Al. (2021) [20] presents the application of Neural Networks models in building energy consumption forecasting, also considering combinations of multiple DL approaches with excellent results, despite the complexity.

Support Vector Machine (SVM) is a kernel-based model suitable for non-linear problems in the absence of large amounts of data [14]. Applied to regression problems, it becomes Support Vector Regression (SVR). Li et Al. (2009) [21] proposed SVM for cooling load prediction, achieving greater accuracy than the Neural Network models used for comparison.

The Bayesian Models use Bayes' conditional probability theory to solve classification or regression problems. These approaches include Naïve Bayes, Gaussian Naïve Bayes, Bayesian Networks (BNs), Bayesian Belief Networks [8]. Although commonly used in other fields of predictive analysis, they are less common in energy consumption prediction [7]. Some applications of BNs are the works of:

Huang et Al. (2018) [22] that applies BN on cooling load prediction, experiencing similar performance and lower computational times than SVM or ANN.

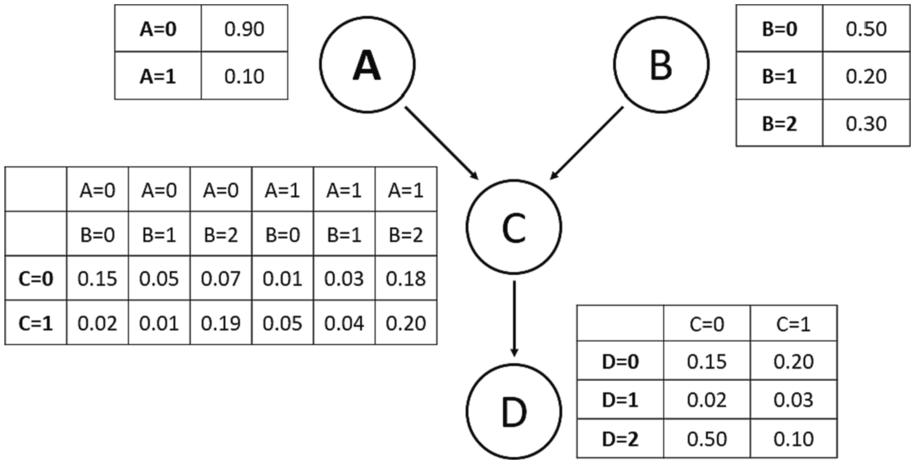

Fig. 1. Example of Bayesian Network DAG and CPTs.

Soares Geraldi et Al. (2019) [23] that studies energy consumption prediction in schools and deepens the design process.

O'Neill et Al. (2016) [24] that predicts building energy performance, analysing critical aspects, such as the discretization and the uncertainty assessment.

2.3. Hybrid models

Hybrid models group ensemble and improved approaches. The former involves the application of a single prediction technique multiple times (homogeneous ensemble) or the use of several prediction models in combination (heterogeneous ensemble). An example of homogeneous ensemble is Random Forest (RF). The work of Ahmad et Al. (2017) [25] proposes RF as an alternative solution to ANN for building energy consumption prediction. The improved prediction models, instead, are the combination of a single prediction model and an optimization technique, such as Particle Swarm Optimization (PSO) algorithm [14]. The work of Zhang et Al. (2016) [26] compares different optimization algorithms on SVR for building energy consumption forecast, emphasizing the importance of parameter optimization. The work of Tran et Al. (2020) [27] presents, instead, a metaheuristic ensemble heterogeneous model to forecast energy consumption in residential buildings.

2.4. Remarks on the state of the art

The three categories analysed allow some preliminary considerations

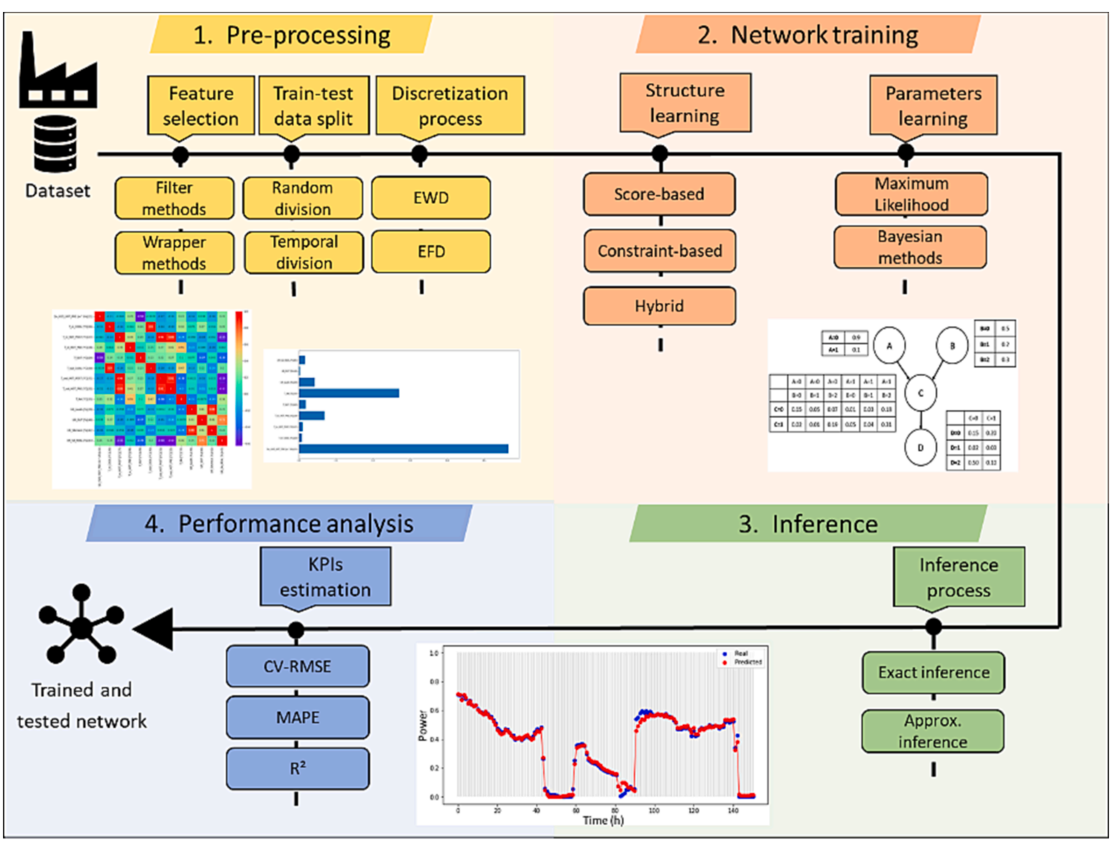

Fig. 2. Design process outline.

to be made. Conventional models, although simple to apply, may perform poorly in the presence of strong non-linearities. AI-based models, on the other hand, overcome this limitation but at the expense of greater complexity and more data required. Similarly, the use of hybrid models can result in even better performance but requiring more sophisticated solutions [12]. A result to highlight is, therefore, that the suitability and the performance of the model depends on the application. For this reason, MLP, RF and SVR have been selected as benchmarking models. The choice has been based mainly on the work of [28].

2.5. Innovative contributions

This work sets three main goals in terms of innovation: (i) application in energy consumption prediction, (ii) systematic definition of a design methodology, and (iii) physical interpretation of the predictions. Regarding the application, it aims to evaluate the performance of Bayesian networks in predicting energy consumption, specifically for industrial applications, whose literature is still sparse. In terms of methodology, there is a lack of works that examine the design phases of Bayesian networks and evaluate strengths or weaknesses of the available solutions. Therefore, this work aims at filling the gap. Finally, the physical interpretation of the results is examined, demonstrating that this represents an advantage of using probabilistic graphical models.

3. Methods and tools

This section presents the Bayesian networks and the design methodology that, as mentioned in Section 2.5, is analysed with a systematic approach, describing all the required steps and the existing techniques for each one. In a final point, the benchmarking models and the software tools for the implementation are mentioned.

3.1. BN definition

A Bayesian network (BN) is a graphical model that builds relationships among variables that are identified as nodes [29]. These relationships are quantified by the conditional probability P(Y|X) where Y is the effect and X the cause, determined through the Bayes theorem [23] with Eq. (1):

$$P(Y|X) = \frac{P(X|Y)^{*}P(Y)}{P(X)} = \frac{P(X \cap Y)}{P(X)} \; with \; P(X) \neq 0 \tag{1}$$

The BN can be defined formally as $\mathscr{B}(S, \theta)$ and consists of two components:

A structure S in the form of a Directed Acyclic Graph (DAG), which contains the nodes and links.

A set of parameters θ representing the Conditional Probability Distributions (CPDs), based on the structure. These parameters are placed into Conditional Probability Tables (CPTs) [30].

An example is shown in Fig. 1.

The Bayesian Networks are, in their simplest form, classification models. Therefore, to be applied to regression problems, they require that continuous variables are discretised into classes, to reduce the amount of probability functions to be assigned [23].

3.2. Design process

The design process is summarised in Fig. 2. The proposed method is divided into four steps: pre-processing, network building, inference, and performance analysis. It has been outlined from the work of [23] and subsequently expanded. It should also be noted that, from a general point of view, the approach is common in the design of prediction models, such as in [31].

3.2.1. Pre-processing

The pre-processing phase gathers all the operations that allow the raw data to be processed into a dataset suitable for the subsequent phases. Assuming that the raw data are already cleansed of outliers and missing values, the study of the pre-processing phase focuses on 3 aspects: feature selection, training-test data split and discretization.

The common methods for feature selection are distinguished in:

Domain knowledge: literature information, engineering experience and physics knowledge are some potential references that can be used in feature selection.

Filter methods: they evaluate the redundancy of features by assigning a score, then eliminate the ones that have the lowest scores. Pearson correlation represent an example.

Wrapper methods: they search, in the field of all the possible dataset, the one that can provide the best accuracy and generalization, then assign a score on goodness-of-fit and number of variables [18,32]. Since, in general, any modelling algorithm combined with a search strategy can be used as wrapper [33], in this case Random Forest is proposed for the feature importance assessment.

Regarding the dataset split for the training and testing phase, two approaches are considered: one based on a random selection [34] and another based on a temporal partition. For performance evaluations, random selection was used, which is less dependent on the time coverage provided by historical data. Temporal partition is only used for the purpose of simulating a prediction process.

For the data discretization, the two proposed approaches are:

Equal Width Discretization (EWD): this method divides the number of observations into k intervals of equal width, where k corresponds to the number of classes.

Equal Frequency Discretization (EFD): this method divides the dataset into k intervals where each one contains approximately the same number of training cases [23].

In accordance with [35] the discretisation process is a key factor for a good prediction and needs to be thoroughly investigated. The reason why other approaches such as those described in [36] have not been proposed is mainly due to requirements of simplicity and generality which led to the selection of the simplest techniques.

3.2.2. Network building

The building of a Bayesian network is done in 2 steps: the structure learning and the parameters learning.

The structure learning consists in finding the best graph that explains all the cause-effect relationships among the variables. The available approaches are based on expert knowledge or data. The first ones, however, are restricted to cases in which the cause-effect relationships are known or easily detectable and the number of variables is limited. Regarding the data-based structure learning, instead, the techniques are usually classified into 3 categories [30,37,38]:

Constraint-based methods: they take a set of conditional independence relations between the variables and build a BN that takes the structure that best approximates these relationships (e.g., PC algorithm).

Score-based methods: they use a search algorithm to browse smartly in the space of possible BN networks (e.g., Hill Climb) and a scoring function that provides a score for every candidate BN to choose the best (e.g., BDeu, BIC, K2).

Hybrid structure learning: it aggregates constrained-based and score-based structure learning algorithms to obtain the advantages of both (e.g., MMHC).

Once the structure of the Bayesian network is specified, the Conditional Probability Tables (CPTs) need to be derived through the parameter learning. The algorithms available for this task have been analysed and presented in the work of [39]. Examples are the Maximum Likelihood Estimator, the Bayesian Estimator, and the Expectation-Maximization Algorithm.

3.2.3. Inference

After the network building, the model can be used to compute the probability distribution for a query variable, given a set of evidence

Table 1
Details of the parameters used in the prediction models.

Model	Parameter	Value/Range of search
RF	Max depth	100
MLP	Hidden layer sizes	[50/50/50, 50/100/50, 100/1]
	Activation function	[ReLU, Tanh, Logistic]
	Solver	Adam
SVR	C	[1e-3, 1e3]
	Epsilon	[1e-8,1e-1]
	Gamma	[1e-3, 1e3]

variables. Depending on the method used to make the inference, two cases can be distinguished:

Exact inference: it finds the exact solution to the inference problem. As an NP-hard problem, some efficient algorithms exist to solve the exact inference by trying to reduce the time of execution (e.g., Variable elimination and Belief propagation).

Approximate inference: all exact Bayesian network inference algorithms have an exponential computational time related to the dimensions of the graph, such as number of nodes, number of classes etc. For this reason, many approximate inference algorithms have been developed. They include stochastic simulation algorithms (e.g., sampling algorithms), model simplification methods, search-based methods and loopy belief propagation [40].

The results provided by the inference are in the form of a CPT and, for the purposes of this study, must subsequently be processed to obtain, from a classification result, a predicted numerical value. In this case, it has been done calculating the weighted average of the mean values of each discretization interval based on the probability distribution.

3.2.4. Performance analysis

The performance analysis measures the capacity of each network to predict reliable results. It is possible to consider different types of analysis such as: sensitivity analysis, influence analysis, model complexity and performance analysis [41]. In this research work, the focus is on the performance analysis. In order to facilitate a comparative assessment with literature results, the most common KPIs, used in regression problems, are proposed [11,13,42]. The selected KPIs (Eqs. (2), (3) and (4)) are dimensionless for a better comparison among different case studies:

$$\textit{Coefficient of Determination } R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i^* - y_i)^2}{\sum_{i=1}^{N}(\overline{y_i} - y_i)^2} \tag{2}$$

$$\textit{Mean Absolute Percentage Error MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i^* - y_i|}{y_i}*100\% \tag{3}$$

$$\textit{Coefficient of Variation of the Root Mean Square Error CV} - \textit{RMSE} = \frac{1}{\overline{y}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^* - y_i)^2}*100\% \tag{4}$$

More specifically, the R square represents the goodness of fit. The higher the accuracy of the model, the more its value will tend to one. The MAPE, on the other hand, provides a dimensionless measure that is easy to interpret due to its linear formulation. However, it is not applicable if there are zero values and becomes less reliable for values near to zero. Finally, CV-RMSE provides a dimensionless measure not subject to compensation, being derived from the RMSE [11].

3.3. Benchmarking models and software tools

As mentioned in Section 2, three models are applied for benchmarking: MLP, SVR and RF. In the first two, hyperparameter optimisation is performed respectively with Grid Search and Particle Swarm Optimization (PSO) algorithm. For these two cases, to have an unbiased evaluation both during hyperparameters' tuning and at the final stage, the samples are divided into three datasets: training, validation, and testing. In Table 1, some details related to the parameters of these models are presented.

Python 3.0 pre-existing libraries have been employed for the implementation of all the models, specifically Pgmpy library [43] for BN and Scikit-learn library [44] for MLP, RF and SVR. The simulations have been carried out using a laptop with Intel i7 8750H, Nvidia GeForce GTX

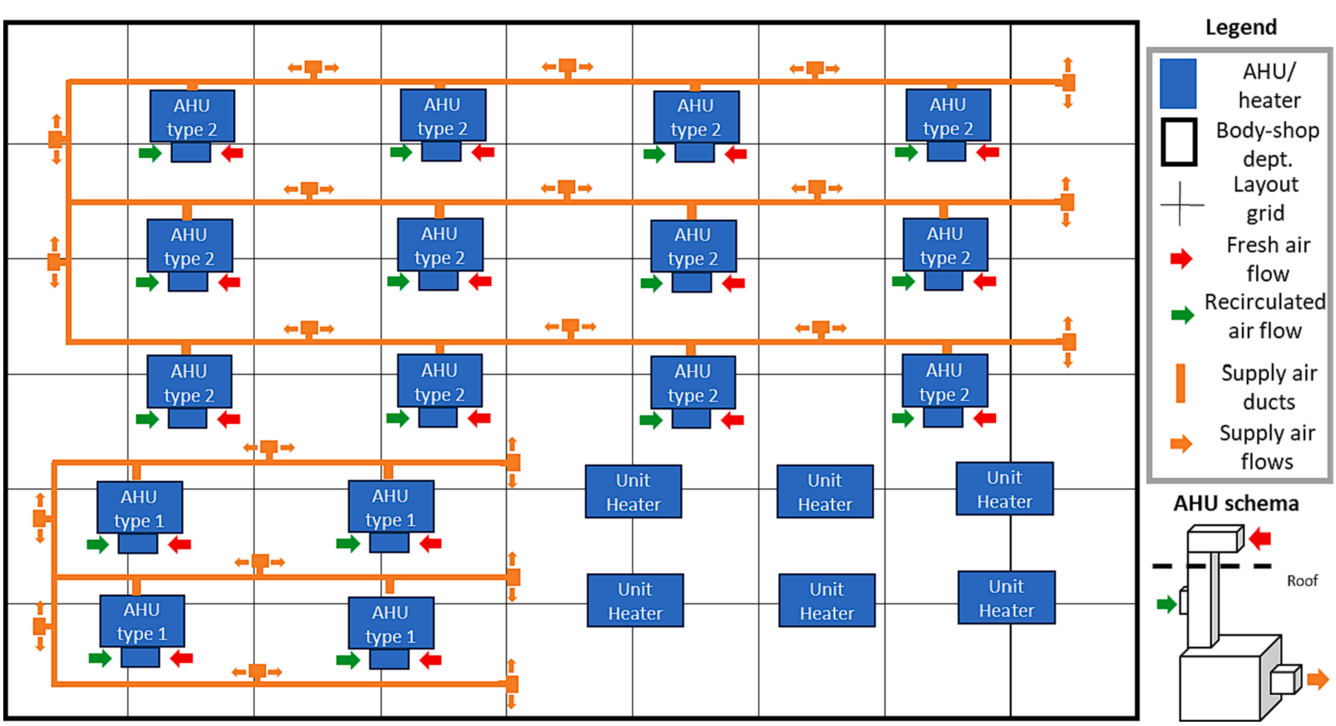

Fig. 3. General schema of the Body-shop Department case study.

Table 2
Body-shop Department variables.

Variable	Unit
Outside dry-bulb temperature	°C
Outside dew-point temperature	°C
Wind speed	m/s
Wind direction	°
Direct normal irradiance	W/mˆ2
Diffuse horizontal irradiance	W/mˆ2
Atmospheric pressure	Pa
Solar altitude	°
Solar azimuth	°
Indoor average temperature	°C
Production scheduling	–
Mechanical ventilation, natural ventilation, and infiltrations	Ac/h
Occupancy	kW
Air temperature	°C
Gas consumption	kWh

1050 Ti 4 Gb and Ram 16 Gb DDR4.

4. Case studies

In this section, two industrial case studies are presented to evaluate the proposed model in real contexts [45]. The first case study focuses on the gas consumption of an HVAC system located in a body-shop department, referred as Body-shop Department, while the other regards the thermal power consumption of an AHU deployed in a topcoat process in a paint-shop department, referred as Paint-shop Department. The analysed case studies are representative of the two mentioned requirements: the regulation of thermal conditions for personnel comfort and the control of the environmental conditions of an industrial process. However, despite the diversity of their implementation and use, industrial HVAC systems share the same main parameters that influence their consumption, including outdoor conditions and production environmental conditions. Such an approach, tested and validated on the presented use cases, can therefore be generally adapted to other HVAC systems.

4.1. Case study 1: Body-shop Department

The Body-shop department is a large building where there are AHUs and unit heaters of diverse types. Fig. 3 shows an illustrative schema of the Body-shop heating system. The AHUs can operate with partial or full air recirculation by adjusting the percentage of fresh air they use. The position of the machines is associated with the nearest pillar, which can be identified through a grid location system where a distinct position within the building layout is defined by a pair of values, namely the row and column coordinates. The objective is to predict the gas consumption of the HVAC system. Data are obtained from the AHUs' probes, the production management system, and the nearby weather station. The available data are listed in Table 2. As a starting dataset, 1541 samples without missing values have been used with an hourly frequency.

4.2. Case study 2: Paint-shop Department

The second case study concerns an AHU of a topcoat process. The AHU ensures that temperature and humidity conditions remain in the optimal ranges during the process, which is schematised in Fig. 4. Table 3 resumes the variables considered in the model design. In this case, the data present a 15-minute sampling for a total of 3000 values covering approximately two months.

5. Results and discussion

This section presents the results obtained in the applications of the proposed method to the case studies, deepening the design steps with a comparison of alternative solutions. Specifically, the aspects analysed are: feature selection, discretization, network building and inference. In

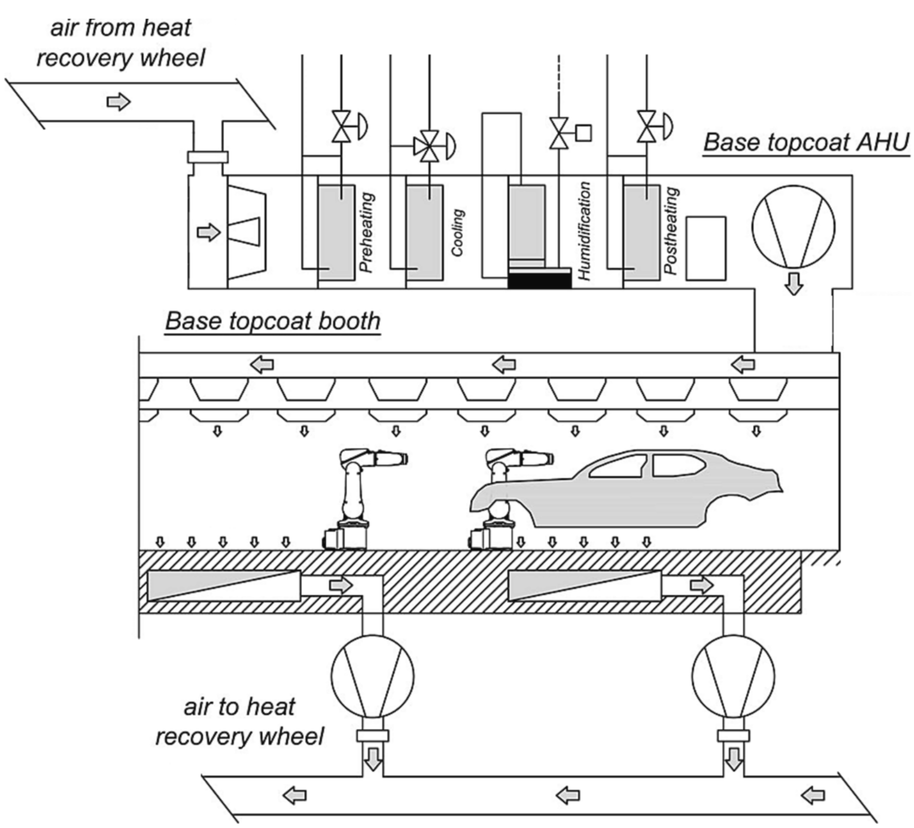

Fig. 4. Process diagram of the Paint-shop Department case study.

Table 3
Paint-shop Department variables.

Variable	Unit	Description
T_RA	°C	return air temperature from the heat wheels
UR_RA	%	return air relative humidity from the heat wheels
T_OUT	°C	external air temperature
UR_OUT	%	external air relative humidity
Qv_H2O_HOT_PRE	m^3/h	superheated water flow rate - return from pre-heating exchanger
T_in_HOT_PRE	°C	delivery superheated water temperature towards the pre-heating exchanger
T_out_HOT_PRE	°C	return water temperature from the pre-heating exchanger
Qv_H2O_HOT_POST	m^3/h	superheated water flow rate - return from the post-heating exchanger
T_in_HOT_POST	°C	delivery superheated water temperature towards the post-heating exchanger
T_out_HOT_POST	°C	return superheated water temperature from the post-heating exchanger
T_in_COOL	°C	delivery chilled water temperature towards the cooling exchanger
T_out_COOL	°C	return chilled water temperature from the cooling exchanger
UR_SAcheck	%	treated air RH for variable control
T_SA_REAL	°C	air temperature from the heat wheels
UR_SA_REAL	%	RH of the air from the heat wheels
UR_booth	%	booth internal RH
P_AHU_REAL_HOT	kW	real thermal heating power of the AHU (pre-heating + post-heating)
P_AHU_REAL_COOL	kW	real thermal cooling power of the AHU
P_AHU_REAL	kW	real thermal power of the AHU (heating + cooling)

the end, Bayesian networks are compared with the benchmarking models, investigating the results both with test-train random split and with a time horizon division, as mentioned in Section 3.2.1. Specifically, the datasets have been divided into 90 % training and 10 % testing for both case studies. Moreover, for the hyperparameters tuning and data preprocessing, the training dataset has been further divided with the same criteria to obtain a validation dataset.

5.1. Feature selection

The first aspect analysed concerns the choice of variables to be considered within the model. Four alternatives are evaluated: no selection, filter method only, wrapper method only, filter and wrapper. An opposite situation arises between the two case studies, as shown in Table 4. In Body-shop Department, in fact, it occurs that the removal of correlated variables significantly improves performance, while, instead, is a negative factor in the Paint-shop Department, where features that have a non-negligible weight on the prediction result are removed. On the other hand, the use of a wrapper method based on feature importance is useful, in both cases, to simplify the model and reduce the computational burden, without affecting the result. That said, in general, it might be worth to conduct the feature selection process even at the cost of a reduction in performance to prevent the invalidation of the model due to overly correlated variables. Based on this principle, the subsequent analysis has been performed considering full feature selection for both case studies. Only in the comparison with other prediction models it has been chosen the process with the best performance. The variables considered after the feature selection process are summarised in Table 5.

5.2. Discretization

For the discretization process, EFD and EWD techniques are evaluated with different numbers of bins. Table 6 shows how EFD is better for both case studies. Another reason that penalises the EWD method is the risk of having empty bins leading to an error in the training phase, which is why in the Paint-shop Department there is no valid solution for K > 10. Analysing the number of bins, instead, it shows how a low K decreases performance due to overly large ranges in which the approximation in considering the class mean value is significant, as shown in Fig. 5. High numbers, on the other hand, increase the size of the CPTs and consequently the computational burden and may complicate the pattern detection due to the lack of data [36]. An acceptable number in relation to the amount of data used is 25. With more data, it is expected that this number will increase.

5.3. Network building

The construction of the network is carried out in the two phases of structure learning and parameter estimation. In the first phase, Hill Climb search with BDEU score, Chow-Liu algorithm, PC algorithm, Max-Min Hill Climb algorithm and Exhaustive search are analysed. For the

Table 5
Variables considered after the feature selection process.

Bodyshop		Paintshop	
Feature	**Unit**	**Feature**	**Unit**
Outside dry-bulb temperature	°C	T_RA	°C
Solar altitude	°	T_OUT	°C
Solar azimuth	°	Qv_H2O_HOT_PRE	m^3/h
Indoor average temperature	°C	T_in_HOT_PRE	°C
Occupancy	kW	T_in_HOT_POST	°C
Air temperature	°C	T_in_COOL	°C
Gas consumption	kWh	UR_SA_REAL	%
		UR_booth	%
		P_AHU_REAL	kW

Table 6
Comparison of discretization techniques.

Model	Body-shop Department			Paint-shop Department		
	CV-RMSE (%)	R^2 (-)	MAPE (%)	CV-RMSE (%)	R^2 (-)	MAPE (%)
EFD with K = 10	24.1	0.959	1.03	27.7	0.938	5.78
EFD with K = 20	23.3	0.963	0.856	26.0	0.970	1.59
EFD with K = 25	19.0	0.975	1.03	25.2	0.974	2.73
EWD with K = 10	34.1	0.914	5.76	75.4	0.511	71.9
EWD with K = 20	35.8	0.909	3.44	–	–	–
EWD with K = 25	34.6	0.931	2.28	–	–	–

Table 4
Comparison of feature selection approaches.

Model	Body-shop Department				Paint-shop Department			
	CV-RMSE (%)	R^2(-)	MAPE(%)	Time of ex. (s)	CV-RMSE (%)	R^2(-)	MAPE (%)	Time of ex. (s)
No features' selection	61.9	− 0.330	23.1	5.97	17.4	0.987	5.42	4.53
Pearson Correlation filter	19.0	0.975	1.03	4.88	25.2	0.974	2.73	2.78
Random Forest wrapper	61.9	− 0.330	22.9	1.83	20.9	0.981	1.93	3.00
Filter + wrapper	19.0	0.975	1.03	2.58	25.2	0.974	2.73	2.34

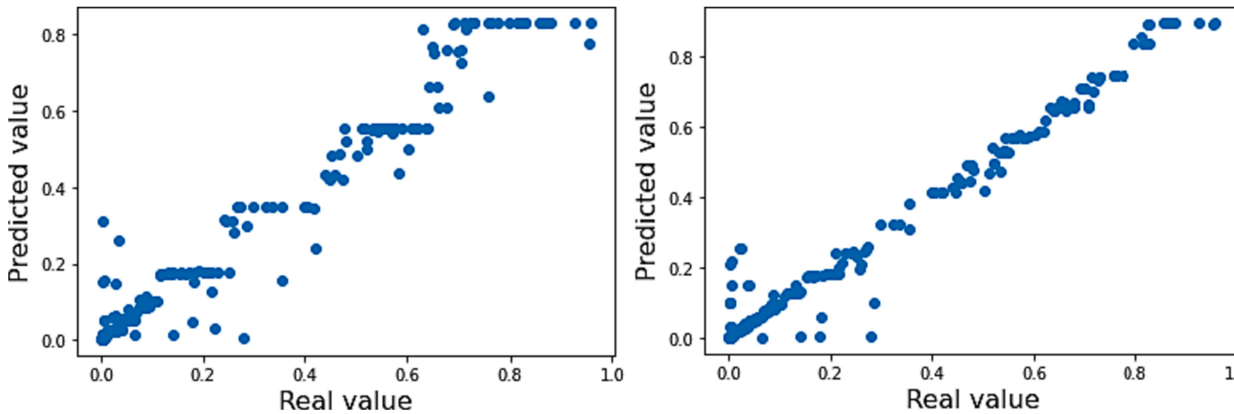

Fig. 5. Prediction results distributions of EWD cases with K = 10 and K = 25 in Body-shop Department.

Table 7
Comparison of structure learning algorithms.

Model	Body-shop Department				Paint-shop Department			
	CV-RMSE(%)	R^2 (-)	MAPE (%)	Time of ex.(s)	CV-RMSE (%)	R^2 (-)	MAPE (%)	Time of ex. (s)
Hill Climb	19.0	0.975	1.03	2.58	25.2	0.974	2.73	2.34
Chow-Liu	20.9	0.971	0.731	2.36	30.2	0.958	3.94	1.96
PC	30.3	0.938	1.46	4.37	62.4	0.774	83.6	16.3
MMHC	19.0	0.975	1.03	5.89	37.9	0.939	4.38	11.3
Exhaustive search	20.7	0.970	1.10	1.63	36.2	0.945	3.73	32.5

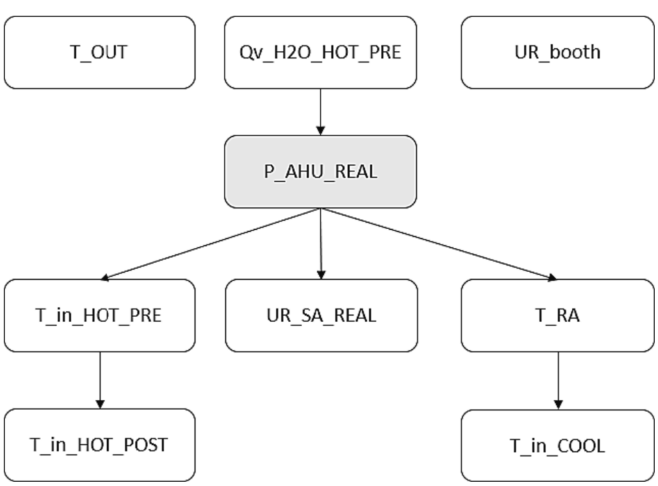

Fig. 6. Paint-shop structure obtained by Hill Climb algorithm.

last three, the number of variables has been reduced because simulation times resulted several orders of magnitude greater compared with the other techniques. The results are shown in Table 7.

It is therefore verified that the score-based technique (Hill Climb) certainly represents the best compromise between speed of learning and performance. Indeed, it must be considered that the Chow-Liu algorithm, although performant, limits the search to tree structures, while the other algorithms are constrained by a higher computational burden. The choice, hence, falls on Hill Climb both for its applicability and performances.

Assessing the physical interpretation instead, we refer to the Paint-shop case study. Looking at the structure obtained through Hill Climb search in Fig. 6, it can be observed that the relationships between the variables are justifiable from a physical point of view. Temperature and humidity depend on the power as it is plausible to assume considering that the AHU is designed to keep these parameters in a certain range, while the power depends on the flow rate of superheated water as amply demonstrated by thermodynamics. As regards the two unconnected variables, they are irrelevant, plausible for the outside temperature, less so for the booth humidity.

As for the parameter estimation, Maximum Likelihood Estimator, Bayesian Estimator and Expectation-Maximization Algorithm have been considered. The results are not reported as, except for Bayesian Estimator, the others led in some cases to errors due to the existence of divisions by zero. For this reason, the choice has fallen on the only generally valid approach.

5.4. Inference

Causal inference, variable elimination, and belief propagation are considered among the existing inference approaches. Table 8 shows that the first two algorithms have the same performance except for execution time. It was expected to find this result because they are exact inference algorithms. As for Belief propagation, on the other hand, the model did not provide a solution because the algorithm, to be applied, needs all

Table 8
Comparison of inference algorithms.

Model	Body-shop Department				Paint-shop Department			
	CV-RMSE (%)	R^2 (-)	MAPE (%)	Time of ex.(s)	CV-RMSE (%)	R^2 (-)	MAPE (%)	Time of ex. (s)
Causal inference	19.0	0.975	1.03	2.58	25.2	0.974	2.73	2.34
Variable Elimination	19.0	0.975	1.03	12.6	25.2	0.974	2.73	7.67
Belief propagation	–	–	–	–	–	–	–	–

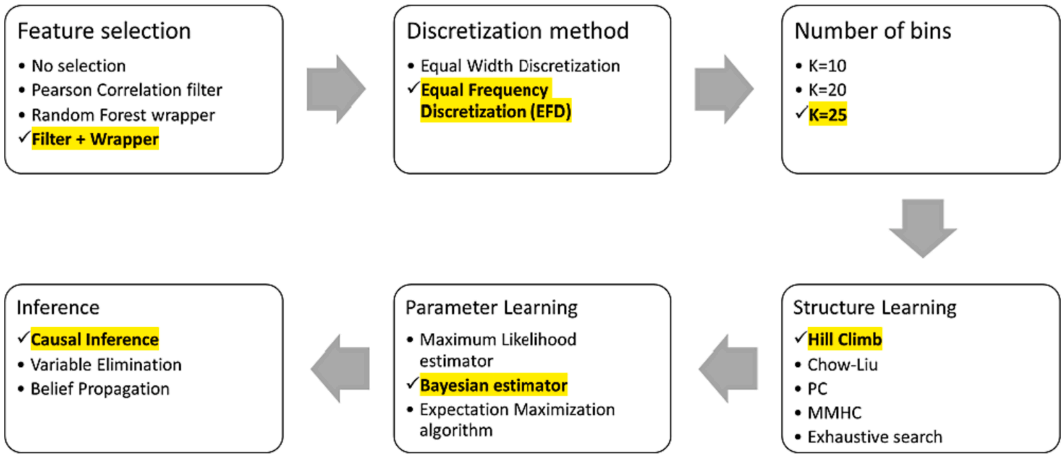

Fig. 7. Best solutions for each step of the prediction model designing process.

Table 9
Comparison of prediction models on a randomly train-test split.

Prediction Models	Body-shop Department				Paint-shop Department			
	CV-RMSE (%)	R^2 (-)	MAPE (%)	Time of ex.(s)	CV-RMSE (%)	R^2 (-)	MAPE (%)	Time of ex. (s)
Bayesian Network	19.0	0.975	1.03	2.58	17.4	0.987	5.42	4.53
Random Forest	21.5	0.970	0.904	0.696	16.5	0.989	4.43	2.42
Multiple Layer Perceptron	19.0	0.975	1.03	35.0	4.50	0.999	0.479	46.3
SVR + PSO	24.4	0.961	1.62	7.06	20.2	0.981	8.06	8.15

Table 10
Comparison of prediction models on a specific time-horizon.

Prediction Models	Body-shop Department				Paint-shop Department			
	CV-RMSE (%)	R^2 (-)	MAPE (%)	Time of ex.(s)	CV-RMSE (%)	R^2 (-)	MAPE (%)	Time of ex. (s)
Bayesian Network	39.9	0.944	2.39	2.32	19.5	0.883	0.705	4.14
Random Forest	50.3	0.906	3.50	0.856	16.4	0.927	0.315	2.55
Multiple Layer Perceptron	32.8	0.970	0.837	42.0	5.11	0.993	0.259	52.9
SVR + PSO	59.0	0.905	1.58	6.73	21.1	0.883	29.6	6.93

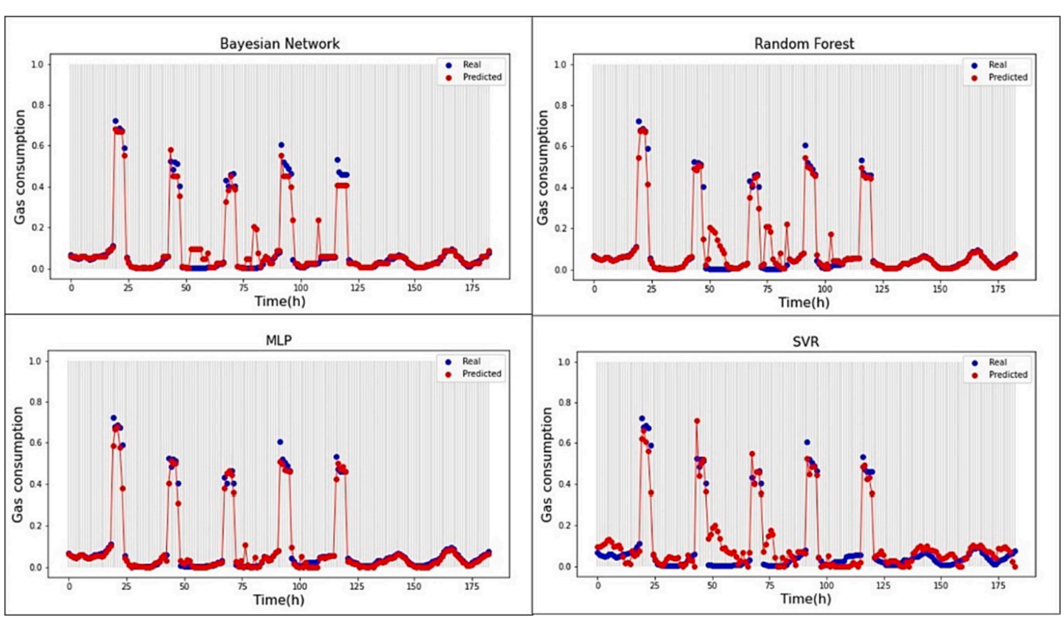

Fig. 8. Graphical results from Body-shop Department.

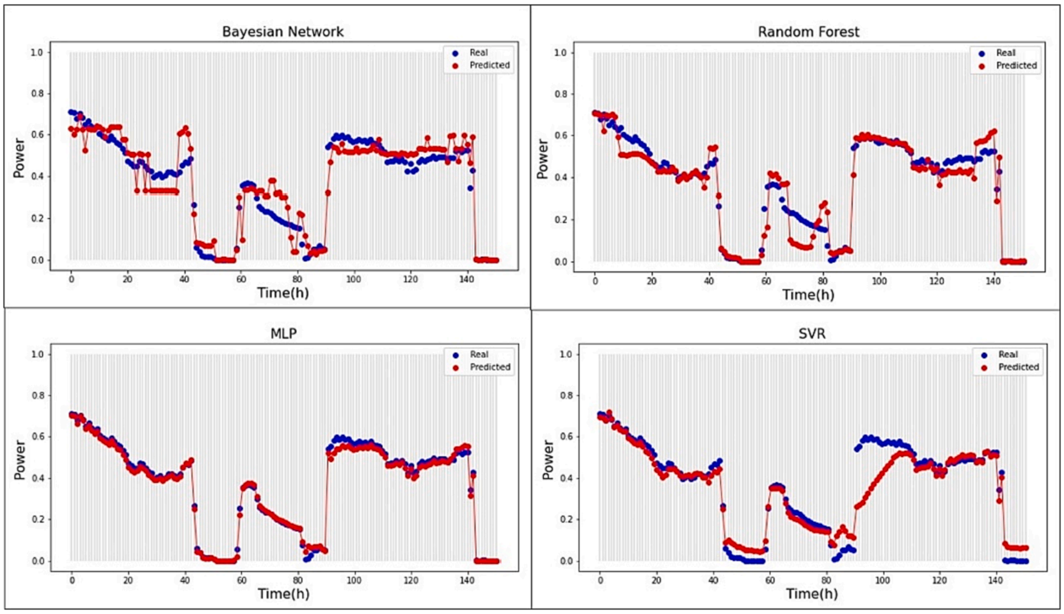

Fig. 9. Graphical results from Paint-shop Department.

nodes to be connected in a single structure, a condition that is difficult to control in an automated learning process. Therefore, Causal inference has been selected as inference algorithm. Approximate inference is not considered as the computational burden of the proposed algorithms is already adequate and therefore there is no need.

5.5. Performance analysis

Fig. 7 resumes all the techniques analysed for each step and highlights the ones that provided the best results. The best-performing BN model has therefore been compared with the benchmarking models.

In the first case, the test dataset contains randomly selected data. Table 9 shows that the performances of BN are comparable with all the other models for Body-shop, whereas for Paint-shop there is a clear prevalence of MLP, despite longer run times. The obtained results refer to the best hyperparameter configuration for MLP and SVR. In the case of MLP, it presents an hidden layer size of 50/50/50 for the Body-shop department and 50/100/50 for the Paint-shop department, ReLU activation function and Adam solver for both cases. For SVR, instead, three hyperparameters are defined as C, Epsilon and Gamma which, for the Body-shop department, are respectively 1000/0.003/32.7 and for the Paint-shop department are 332/0.02/44.9. RF has not been subject to optimisation as the only hyper-parameter considered was the maximum depth set at 100.

The second case, shown in Table 10, considers instead a test dataset containing the last provided data, simulating a specific time horizon prediction. In this case, considering the Body-shop Department, the high RMSEs can be justified by the preponderance of close-to-zero values that strongly conditions the validity of a relative error. In general, instead, what was seen for the previous case is equally valid, also with reference to the Paint-shop Department.

To conclude this overview, Fig. 8 and Fig. 9 present the same results of Table 10 in a graphical way. From a qualitative analysis, it can be highlighted that, for the Body-shop Department, the graphs reflect the KPIs without showing specific trends. For the Paint-shop Department, instead, it can be noticed how SVR seems to be accurate but only when it identifies the correct trend while BN and RF, although recognising the trend, still show considerable noise.

6. Conclusions and future works

This work focused on the implementation of a probabilistic graphical model, the Bayesian Network, for predicting the energy consumption of industrial HVAC systems. The objectives were the validation of BN for this type of applications, the systematic definition of the design methodology, and the evaluation of the physical interpretability of the model. With reference to the methodology, this work included an in-depth study of the prediction model building, achieved by analysing the design process and detailing the available solutions for each step. These solutions have been evaluated by implementing the models on two case studies and selecting the techniques that provided the best results (see Fig. 7). To assess the model validation in predicting energy consumption, instead, BN has been compared with benchmarking models, specifically MLP, RF and SVR. The obtained results, although not showing the best prediction results among the implemented models, highlight the potential of this methodology in terms of physical interpretation of the network (see Fig. 6), simplicity of implementation through pre-existing python libraries and the flexibility of application, typical of data-driven models. However, some weaknesses have also been analysed, the main one is the need to discretise the data which strongly influences the quality of the results. The discretization process is, indeed, difficult to optimise, since parameters such as the optimal number of classes depend on the quantity of data. This aspect projects towards some possible continuations of this work, such as the implementation of Bayesian Network's typologies that support continuous data or time series, or the use of supervised discretization algorithms to optimize the process. Other aspects for further studies concern the behaviour of the model when varying the amount of provided data or the investigation of correlation between the various design phases.

CRediT authorship contribution statement

Francesco Giuseppe Ciampi: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Andrea Rega:** Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Thierno M.L. Diallo:** Conceptualization, Supervision, Writing – review & editing. **Francesco Pelella:** Conceptualization, Writing – review & editing. **Jean-Yves Choley:** Conceptualization, Supervision, Writing – review & editing. **Stanislao Patalano:** Conceptualization, Supervision, Writing – review

& editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This research work was carried out as part of the ENERMAN (ENERgy-efficient manufacturing system MANagement) project founded by the European Union's Horizon 2020 Program under Grant Agreement No. 958478.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enbuild.2024.114039.

References

[1] V.J. Mawson, B.R. Hughes, Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector, Energ. Build. 217 (2020) 109966, https://doi.org/10.1016/j.enbuild.2020.109966.
[2] V.J. Mawson, B.R. Hughes, Coupling simulation with artificial neural networks for the optimisation of HVAC controls in manufacturing environments, Optim. Eng. 22 (1) (2021) 103–119, https://doi.org/10.1007/s11081-020-09567-y.
[3] J. Runge, E. Saloux, A comparison of prediction and forecasting artificial intelligence models to estimate the future energy demand in a district heating system, Energy 269 (2023) 126661, https://doi.org/10.1016/j.energy.2023.126661.
[4] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: a review, Renew. Sustain. Energy Rev. 23 (2013) 272–288, https://doi.org/10.1016/j.rser.2013.03.004.
[5] N. Wei, L. Yin, C. Li, C. Chan, F. Zeng, Forecasting the daily natural gas consumption with an accurate white-box model, Energy 232 (2021) 121036, https://doi.org/10.1016/j.energy.2021.121036.
[6] Y. Chen, M. Guo, Z. Chen, Z. Chen, Y. Ji, Physical energy and data-driven models in building energy prediction: a review, Energy Rep. 8 (2022) 2656–2671, https://doi.org/10.1016/j.egyr.2022.01.162.
[7] T. Li, Y. Zhao, K. Yan, K. Zhou, C. Zhang, X. Zhang, Probabilistic graphical models in energy systems: a review, Build. Simul. 15 (5) (2022) 699–728, https://doi.org/10.1007/s12273-021-0849-9.
[8] Q. Qiao, A. Yunusa-Kaltungo, R.E. Edwards, Towards developing a systematic knowledge trend for building energy consumption prediction, J. Build. Eng. 35 (2021) 101967, https://doi.org/10.1016/j.jobe.2020.101967.
[9] J. Walther, M. Weigold, A systematic review on predicting and forecasting the electrical energy consumption in the manufacturing industry, Energies 14 (4) (2021) 968, https://doi.org/10.3390/en14040968.
[10] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592, https://doi.org/10.1016/j.rser.2012.02.049.
[11] Y. Sun, F. Haghighat, B.C.M. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, Energ. Buildings 221 (2020) 110022, https://doi.org/10.1016/j.enbuild.2020.110022.
[12] M. A. Mat Daut, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, F. Hussin, 'Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review, *Renewable Sustainable Energy Rev.*, 70, 1108–1118, 2017, doi: 10.1016/j.rser.2016.12.015.
[13] N. Wei, C. Li, X. Peng, F. Zeng, X. Lu, Conventional models and artificial intelligence-based models for energy consumption forecasting: a review, J. Pet. Sci. Eng. 181 (2019) 106187, https://doi.org/10.1016/j.petrol.2019.106187.
[14] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, Modeling and forecasting building energy consumption: A review of data-driven techniques, *Sustainable Cities and Society*, 48, p. 101533, 2019, doi: 10.1016/j.scs.2019.101533.
[15] Y. K. Penya, C. E. Borges, D. Agote, and I. Fernandez, Short-term load forecasting in air-conditioned non-residential Buildings, in *2011 IEEE International Symposium on Industrial Electronics*, Gdansk, Poland: IEEE, 2011, pp. 1359–1364. doi: 10.1109/ISIE.2011.5984356.
[16] F. Lei, P. Hu, A Baseline Model for Office Building Energy Consumption in Hot Summer and Cold Winter Region, in *2009 International Conference on Management and Service Science*, Beijing, China: IEEE, Sep. 2009, pp. 1–4. doi: 10.1109/ICMSS.2009.5301031.
[17] A. Bracale, G. Carpinelli, P. De Falco, T. Hong, Short-term industrial reactive power forecasting, Int. J. Electr. Power Energy Syst. 107 (2019) 177–185, https://doi.org/10.1016/j.ijepes.2018.11.022.
[18] L. Zhang, et al., A review of machine learning in building load prediction, Appl. Energy 285 (2021) 116452, https://doi.org/10.1016/j.apenergy.2021.116452.
[19] Y. Ding, Q. Zhang, T. Yuan, K. Yang, Model input selection for building heating load prediction: a case study for an office building in Tianjin, Energ. Buildings 159 (2018) 254–270, https://doi.org/10.1016/j.enbuild.2017.11.002.
[20] N. Somu, G. Raman M R, K. Ramamritham, A deep learning framework for building energy consumption forecast, *Renewable Sustainable Energy Rev.*, 137, p. 110591, 2021, doi: 10.1016/j.rser.2020.110591.
[21] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks, Energ. Conver. Manage. 50 (1) (2009) 90–96, https://doi.org/10.1016/j.enconman.2008.08.033.
[22] S. Huang, W. Zuo, M.D. Sohn, A bayesian network model for predicting cooling load of commercial buildings, Build. Simul. 11 (1) (2018) 87–101, https://doi.org/10.1007/s12273-017-0382-z.
[23] M. Soares Geraldi, M. V. Bavaresco, E. Ghisi, Bayesian Network for Predicting Energy Consumption in Schools in Florianópolis – Brazil, presented at the Building Simulation 2019, Rome, Italy, pp. 4188–4195. doi: 10.26868/25222708.2019.210484.
[24] Z. O'Neill, C. O'Neill, Development of a probabilistic graphical model for predicting building energy performance, Appl. Energy 164 (2016) 650–658, https://doi.org/10.1016/j.apenergy.2015.12.015.
[25] M.W. Ahmad, M. Mourshed, Y. Rezgui, Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption, Energ. Build. 147 (Jul. 2017) 77–89, https://doi.org/10.1016/j.enbuild.2017.04.038.
[26] F. Zhang, C. Deb, S.E. Lee, J. Yang, K.W. Shah, Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique, Energ. Buildings 126 (2016) 94–103, https://doi.org/10.1016/j.enbuild.2016.05.028.
[27] D.-H. Tran, D.-L. Luong, J.-S. Chou, Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings, Energy 191 (2020) 116552, https://doi.org/10.1016/j.energy.2019.116552.
[28] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Appl. Energy 127 (2014) 1–10, https://doi.org/10.1016/j.apenergy.2014.04.016.
[29] L. Hawarah, S. Ploix, M. Jacomino, User Behavior Prediction in Energy Consumption in Housing Using Bayesian Networks, in *Artificial Intelligence and Soft Computing*, vol. 6113, L. Rutkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, Eds., in Lecture Notes in Computer Science, vol. 6113. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 372–379. doi: 10.1007/978-3-642-13208-7_47.
[30] P. Larrañaga, H. Karshenas, C. Bielza, R. Santana, A review on evolutionary algorithms in bayesian network learning and inference tasks, Inf. Sci. 233 (2013) 109–125, https://doi.org/10.1016/j.ins.2012.12.051.
[31] Z. Dong, J. Liu, B. Liu, K. Li, X. Li, Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification, Energ. Build. 241 (2021) 110929, https://doi.org/10.1016/j.enbuild.2021.110929.
[32] L. Zhang, J. Wen, A systematic feature selection procedure for short-term data-driven building energy forecasting model development, Energ. Buildings 183 (2019) 428–442, https://doi.org/10.1016/j.enbuild.2018.11.010.
[33] A. Jovic, K. Brkic, N. Bogunovic, A review of feature selection methods with applications, in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia: IEEE, 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.
[34] A. Kusiak, G. Xu, Modeling and optimization of HVAC systems using a dynamic neural network, Energy 42 (1) (2012) 241–250, https://doi.org/10.1016/j.energy.2012.03.063.
[35] F. Nojavan A., S. S. Qian, C. A. Stow, Comparative analysis of discretization methods in Bayesian networks, *Environ. Modell. Software*, 87, 64–71, 2017, doi: 10.1016/j.envsoft.2016.10.007.
[36] T. Beuzen, L. Marshall, K.D. Splinter, A comparison of methods for discretizing continuous variables in bayesian networks, Environ. Model. Softw. 108 (Oct. 2018) 61–66, https://doi.org/10.1016/j.envsoft.2018.07.007.
[37] M. Scanagatta, A. Salmerón, F. Stella, A survey on bayesian network structure learning from data, Prog Artif Intell 8 (4) (2019) 425–439, https://doi.org/10.1007/s13748-019-00194-y.
[38] N.K. Kitson, A.C. Constantinou, Z. Guo, Y. Liu, K. Chobtham, A survey of bayesian network structure learning, Artif. Intell. Rev. (2023), https://doi.org/10.1007/s10462-022-10351-w.
[39] Z. Ji, Q. Xia, G. Meng, A Review of Parameter Learning Methods in Bayesian Network, in *Advanced Intelligent Computing Theories and Applications*, vol. 9227, D.-S. Huang and K. Han, Eds., in Lecture Notes in Computer Science, vol. 9227., Cham: Springer International Publishing, 2015, pp. 3–12. doi: 10.1007/978-3-319-22053-6_1.
[40] H. Guo and W. Hsu, A Survey of Algorithms for Real-Time Bayesian Network Inference, p. 12.
[41] B.G. Marcot, Metrics for evaluating performance and uncertainty of bayesian network models, Ecol. Model. 230 (2012) 50–62, https://doi.org/10.1016/j.ecolmodel.2012.01.013.

[42] L. Ouazzani Chahidi, M. Fossa, A. Priarone, and A. Mechaqrane, Evaluation of Supervised learning models in predicting greenhouse energy demand and production for intelligent and sustainable operations, *Energies*, vol. 14, no. 19, p. 6297, Oct. 2021, doi: 10.3390/en14196297.

[43] A. Ankan, A. Panda, 'pgmpy: Probabilistic Graphical Models using Python', presented at the Python in Science Conference, Austin, Texas, 2015, pp. 6–11. doi: 10.25080/Majora-7b98e3ed-001.

[44] F. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python, *MACHINE LEARNING IN PYTHON*.

[45] Enerman, EnerMan H2020 - Energy Efficient Manufacturing System Management, EnerMan. Accessed: Nov. 12, 2022. [Online]. Available: https://enerman-h2020.eu/.