Evolving Dynamic Bayesian Networks for CO_2 Emissions Forecasting in Multi-Source Power Generation Systems

Talysson M. O. Santos and Michel Bessani and Ivan N. Da Silva

***Abstract*—Global warming is a significant challenge. Among the contributors, CO_2 emission is the foremost, and almost 40% of global emissions come from electricity generation. In this sense, an accurate prediction of CO_2 emissions in a multi-source system combining traditional and renewable sources can be used to support the reduction of carbon emissions without affecting the energy demand-supply. Despite the several relevant research in this topic, because of higher uncertainty and variability caused mainly by the intermittent nature of renewable energy, CO_2 emissions forecasting in multi-source power generation systems is a current challenge. This paper presents CO_2 emissions forecasting for multi-source power generation systems using evolving discrete Dynamic Bayesian Networks. Our proposal uses an analytical threshold for selecting directed edges by the occurrence frequency as data arrives, allowing a constant adaptation to smoothly converges into a robust forecast model. It was tested using real data from multi-source power generation systems of Belgium, Germany, Portugal, and Spain. Its performance was compared with other forecasting methods. Comparing the results against a traditional DBN that not evolves the structure over time, our proposal was superior highlighting a contribution of performance improvement. The proposed method was better when compared against ANN and XgBoost, with the difference in performance statistically significant.**

***Index Terms*—CO_2 emissions forecasting, Energy Management and Sustainability, Dynamic Bayesian Networks, Multi-Source power generation system.**

I. Introduction

Global warming and climate change are one of the main discussions around the worldwide community to propose alternatives that make sustainable development possible [1]. Regarding the consequences of global warming, the world faces extreme climate events such as heating up of the atmospheric temperature, glacier melting, tsunamis, and rising sea levels, highlighting the necessity to make efforts to mitigate environmental pollution [2]. Among the set of greenhouse gases (GHG) that are contributors factors to global warming, carbon dioxide (CO_2) emission is the major contributor [3], [4] and has increased by 47% over the past 170 years due to human activities [1], [5].

Among human activities, economic development increases industrialisation and urbanisation which causes excessive consumption of natural resources and also increases the energy demand [6], [7]. Almost 40% of global CO_2 emissions come from using fossil fuels to generate electricity [1]. In Europe, the energy sector is responsible for roughly 66.67% of all GHG emissions [3], [8] and other economies like China, the USA, and India also presented higher CO_2 emissions coming from the energy sector [9]. In Latin America, buildings are responsible for 22% of the total energy demand, and the forecasts indicate that energy demand will increase by at least 80% by 2040 due to the expansion of the middle class [10]. These facts highlight an enormous potential of actuation in the energy sector to achieve the goal of reducing GHG significantly.

During energy generation, the total of CO_2 emitted varies as a function of the sources used to generate it [3]. In other words, each source has its CO_2-equivalent intensity factor associated with one kWh of energy produced. One possibility to reduce the CO_2 emissions without affecting the energy demand-supply, is the use of alternative green energy sources such as solar and wind combined with other traditional sources that do not have the intermittent nature of renewable energy [11], [12]. In this context, efficient rescheduling of energy generation integrating renewable energy sources can reduce up to 40% emissions [13].

Recent efforts have been made to forecast the environmental impact during energy generation to manage the production coming from heterogeneous supplies in order to regulate and reduce pollutant emissions [1]–[3], [11]. With an accurate prediction of CO_2 emissions in multi-source systems, it is possible to act in architecture design, capacity planning, and energy management strategies to achieve the goals regarding the management and reduction of carbon emissions [11], [14].

For such a purpose, Qader *et al.* applied multiple methods such as neural network time series nonlinear auto-regressive, Gaussian Process Regression, and Holt's methods for forecasting CO_2 emission of Bahrain, concluding that the neural network time series nonlinear auto-regressive model has performed better [1]. Bokde *et al.* used decomposition approaches to short-term CO_2 emissions forecasting and its impact on electricity market scheduling of five European countries [9]. In [15], the authors proposed a combination of artificial neural networks (ANN) model with an agent-based architecture to forecast the hourly gas consumption and electrical production and then calculate the equivalent amount of emitted CO2 for both energy sources. Xu *et al.* proposed the use of non-equigap grey model with conformable fractional accumulation to inves-

Talysson M. O. Santos is with the Department of Electrical and Computing Engineering, University of São Paulo - USP, Brazil (e-mail:talyssonsantos@usp.br)

Michel Bessani is with the Department of Electrical Engineering, Federal University of Minas Gerais - UFMG, Brazil (e-mail:mbessani@ufmg.br)

Ivan N. Da Silva is with the Department of Electrical and Computing Engineering, University of São Paulo - USP, Brazil (e-mail:insilva@sc.usp.br)

tigate the relationship between energy consumption and CO_2 emissions. Using consumption as input and CO_2 emissions as output, CO_2 emissions of 53 countries and regions in North America, South America, Europe, the Commonwealth of Independent States, the Middle East, Africa, and Asia Pacific are predicted [16].

A comparative analysis to forecast CO_2 emissions was presented in Faruque *et al.* [17]. The paper examines the relationships between CO_2 emissions, electrical energy consumption, and gross domestic product (GDP) in Bangladesh from 1972 to 2019. Long short-term memory (LSTM) neural networks, Convolution neural networks (CNN), CNN-Long short-term memory networks (CNN-LSTM), and ANN with more than one layer (Deep Neural Networks DNN) were used. The authors highlighted that the number of neuron layers in all deep learning models affects predicting accuracy and all hyper-parameters are manually adjusted through trial and error. The best performance comes from the use of the DNN technique. Javanmard *et al.* applied machine learning algorithms and optimisation models to forecast CO_2 emissions with energy market data from Iran. Among the nine machine learning algorithms used, results indicate that auto-regressive-based models algorithms are higher than other algorithms, followed by ANN. The worst forecast accuracy is related to LSTM and Support Vector Regression (SVR) [18].

Despite the relevant achievements presented in the studies aforementioned, it is interesting to investigate new approaches in search of performance improvements. Most methods presented are based on ANN or deep learning [2] that give a good fit and forecasting performance for specific systems, but are not generalist and is hard to interpret the results to track and correct erroneous behaviour [3], [19]. An interesting option that has been applied in the energy sector is probabilistic models [20]–[24], especially Bayesian Networks (BNs), which account for uncertainty in a rigorous and transparent manner and allow an easy interpretation of the results [25]. For properly handling time series, Dynamic Bayesian Networks (DBN) are probabilistic graphical expressing qualitative and quantitative relationships among variables over adjacent time steps [26] and can be applied for stationary and non-stationary time series [3]. In [3], the authors used DBN to CO_2 emissions forecasting on the multi-source power generation system of Germany and found good results. Although the high prediction accuracy, Santos *et al.*, Wang *et al.*, and Meng *et al.* pointed out as crucial future research the development of a method to evolve the entire structure model as data is arriving to be a robust forecasting model and properly fit the data without manual interventions [3], [27], [28].

In this sense, we propose a methodology to make CO_2 emissions forecast in multi-source power generation systems using Evolving discrete Dynamic Bayesian Networks (EDBN) by an analytical threshold for selecting directed edges by the occurrence frequency as data is arriving. Using the datasets collected each day, the algorithm learns the partial structure of the DBN using the Akaike Information Criteria (AIC) score metric in combination with the hill-climbing method. Then, using each partial structure obtained along the days, our approach actualises the occurrence frequency of each edge and uses the analytical threshold for selecting the directed edges by the occurrence frequency. Our method constantly adapts to the arrival of new datasets to properly fit the data and smoothly converges to a robust forecast model. We choose the discrete model with an uninformative prior to having a non-parametric model that can be adjusted due to changes in behaviour changes without manual modifications [3], [29], [30], i.e., a data-driven proposal.

In summary, the main contributions of this article are as follows:

- We propose the use of an Evolving discrete Dynamic Bayesian Network by an analytical threshold for selecting directed edges by the occurrence frequency as data is arriving for handling CO_2 emissions forecast in multi-source power generation systems.
- Besides being data-driven, our approach smoothly converges and adapts the structure as data arrives, forming a robust method for dealing with forecasting of time series data.
- We evaluate the capability of our method using real datasets of multi-source power generation systems of four countries: Belgium, Germany, Spain, and Portugal. Moreover, it can be applied in Latin America's energy systems without manual modifications.
- We benchmark the performance against a set of widely used forecasting methods.

The remainder of this paper is organised as follows. Section II formally introduces the fundamental concepts, as CO_2 emissions in generation systems, followed by the theory of DBN, the analytical threshold for evolving DBN, and information theory. Section III presents the materials and methods, describing the multi-source power generation systems data, the data pre-processing, the CO_2 emissions forecasting through EDBN, the performance evaluation approach, and the computational resources utilised. Section IV presents and discusses the results comparing our method against other methods. Section V concludes this manuscript.

II. Theoretical Background

A. CO_2 Emissions in Multi-Source Power Generation Systems

Due to the efforts to track and reduce GHGs emissions, the emissions resulting from using a particular energy source need to be quantified in the function of the total kWh produced. This section explains the concept of dynamic CO_2-equivalent intensity factor and how it can be used to compute the emissions in a multi-source power generation system.

Given the amount of energy generated per type of source, it is possible to quantify the total emissions using the emission factors of each source [13]. These factors express the dynamic CO_2-equivalent intensity factor associated with one kWh of energy produced. In [13], [31], several emission factors for different sources are given. Table I shows the emissions factors used in this study.

Using the factors expressed in Table I, the joint dynamic CO_2 emissions intensity of a multi-source power generation system can be calculated as Eq. 1

$$EF_{j,t} = \sum_{s} E_{t,s}.EF_s, \tag{1}$$

TABLE I
EMISSION FACTORS FOR DIFFERENT SOURCES EXPRESSED IN gCO_2eq/kWh. EMISSION FACTORS OF SOURCES WITH "*" ARE CALCULATED AS THE MEAN OF ALL OTHER FACTORS (OF THE SAME CLASS - RENEWABLE OR NOT) EXPRESSED IN THIS TABLE.

Biomass	Solar PV	Wind onshore	Wind offshore
71	43	8	9
Geothermal	Pumped-Storage	Run-of-the-river	Reservoir
45	34	4	9
Nuclear	Lignite	Coal	Coal-derived gas
11	820	800	800
Gas	Oil	Waste	Other renewable* 33
400	520	690	Other* 376

where $E_{t,s}$ is the energy produced by source s in time step t, and EF_s is the emission factor of source s.

B. Dynamic Bayesian Network

A directed graph G is a pair (V, E), where V is a set of vertices or nodes. E is the set of edges that form pairs of distinct elements of V. If $(v_i, v_j) \in E$, v_i is a parent of v_j and there is a directed edge from v_i to v_j. A *directed acyclic graph* (DAG) is a directed graph G without cycles [32].

A Bayesian Network (BN) is a probabilistic graphical model [29] composed of a qualitative (structure) and a quantitative part (parameters) [33]. Given a set of n random variables $V = \{v_1, v_2, \ldots, v_n\}$ under analysis, the structure is a DAG G and represents the conditional dependencies among the variables of V. Considering the edges of G, the quantitative part is the set of conditional probability distributions $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ [32]. The pair $B = (G, \Theta)$ is a BN and according to the Markov condition, each vertex $v_i \in V$ is conditionally independent of all its non-descendants given all its parents in G. As a consequence, the state of a variable v_i can be computed by a conditional probability $\theta_i = P(v_i|pa_{v_i})$, where pa_{v_i} are the parents of v_i in the structure G. Using this, the joint probability distribution (JPD) encoded by B can be computed directly from the chain rule as [32]

$$P(v_1, v_2, \ldots, v_n) = \prod_{i=1}^{n} P(v_i|pa_{v_i}). \tag{2}$$

DBN is a BN with an additional ability to relate variables to each other over adjacent time steps [3]. Given a set of n random variables, a DBN of *k*th-order estimates the probability distribution using the information of the k previous time window $(\tau - k + 1 : \tau)$ to forecasting the variables in the next time slice $(\tau + 1)$ [3]. This ability to find and represent temporal connections improves the performance in applications of multivariate time series in stationary and non-stationary cases [28].

A DBN model is formed by the pair (G, Θ) and both can be learned completely data-driven [3]. To simplify model learning and reduce computational complexity [27], it is common assumpts that there is only a limited time slice that influences the future state of the process. This assumption results in a DBN of order 1, which is a 2-slice temporal Bayesian network (2-DBN). The future states are conditionally dependent just on the observations of the actual time slice, i.e, $P(V^{\tau+1}|V^{\tau}) \equiv P(V^{t:t+\Delta_p}|V^{1:t})$ where $\Delta_p > 0$ is how far into the future we want to predict.

Structural learning usually can be score-based, constraint-based, and hybrid [34]. Score-based uses heuristics to search in the space of DAGs for structures and then use some score metrics to evaluate the structure [35]. Constraint-based uses independence tests to evaluate and select edges to form G and hybrid methods combine the other two approaches [29]. Using a score-based approach, structural learning can be posed as an optimisation problem: given a dataset D with n random variables, the scoring metric can be maximised by finding a pair $B = (G, \Theta)$ [33].

Among the possibilities of score metrics to evaluate the structure during structural learning, AIC tends to result in models with a good predictive performance [29].

After obtain the DAG G, it is necessary to learn the parameters of the DBN, i.e., learn the quantitative part (Θ). The quantitative part depends on the edges in G and also on the dataset D being modelled [26]. For the discrete model, the quantitative part is formed by conditional probability tables (CPTs), where each one describes the probabilities of each state of a variable given the relations of the structure G [32]. This learning stage can be performed by maximum likelihood or also a Bayesian estimation [29], [36].

As the discrete model require CPTs for each variable, learning the quantitative part including all variables and edges of a complete DAG can demand a high computational cost [26]. When the goal is to forecast future observations of a random variable, usually a subset is enough. This subset that contains all the useful information is called a Markov blanket [37] and includes all parents of the variable to be predicted, children and children's parents.

With a complete DBN, it is possible realise Bayesian inference using maximum a posteriori estimation (MAP) [32].

C. An Analythical Threshold for Evolving Dynamic Bayesian Networks

As illustrated in the previous section, using the score function option, structural learning can be posed as an optimisation problem. Incomplete or noisy data can provide a partially spurious structure [33], [38]. Moreover, a method to evolve the entire model as new data is coming in can smoothly converge into a robust forecast model, properly fit the new coming data to improve the forecast performance [3], [28]. In this sense, we used an approach based on the averaging strategy with an analytical threshold to select the edges by the occurrence frequency as new datasets arrive.

In the context of bootstrap resampling instead of new datasets arriving, this type of model learning technique selecting the edges by the occurrence frequency was investigated in [33], [38], [39]. To select a coherent threshold value, [33] made a deduction using an analogy with an adapted one-dimensional random-walk and evaluated this threshold via data perturbation by dataset bootstrap replicas using Matthews

Correlation Coefficient (MCC) as the performance metric. The authors presented the following closed-form expression:

$$f_{th} = \frac{1}{3} + \sqrt{\frac{2}{R}}, \tag{3}$$

where R is the number of bootstrap resamplings.

Considering scenarios of evolving the model as new data is arriving, [26] proposed an adaptation regarding the threshold proposed by [33]. The authors pointed out that without modifications, f_{th} is close to one at the beginning of the process and then can reject all edges principally in the presence of data problems as missing values. The closed-form expression used was:

$$f_{th} = \begin{cases} 0.6, & \text{if } W < 28 \\ \frac{1}{3} + \sqrt{\frac{2}{W}}, & \text{otherwise} \end{cases}, \tag{4}$$

where W is the total of datasets collected along each day w = 1, . . . , W.

Using real and simulated datasets, [26] showed that using the adapted threshold the model converges to a robust data imputation model in the presence of missing completely at random and not at random. Using the adapted threshold proposed and evaluated in [26], our proposal applies the following steps to select the edges as new data arrives:

For w = 1, . . . , W days:

1) get the dataset collected on day w;
2) organise the dataset D_w formed by the collected data;
3) learn the structure $G_w = (V, E_w)$ from D_w using an algorithm that search in the DAG space with local-search combined with score metric;
4) get the Makov Blanket of the target variable, i.e., G_w in this step reduces to a subset that contains just the useful information.
5) Estimate the probability that each connection v_i - v_j is present in true network $G^* = (V, E^*)$ as

$$\bar{e}_{ij} = \bar{e}_{ji} = \frac{1}{W} \sum_{w=1}^{W} (e_{ij}^w + e_{ji}^w), \tag{5}$$

where i, $j \in \{1, \ldots, n\}$, e_{ij}^w and $e_{ji}^k \in E_w$ and the superscript w is just an index and does not mean a potentiation.

6) Update the threshold f_{th}.
7) The link v_i - v_j exists (is true) if $\bar{e}_{ji}$ overcomes the threshold f_{th}.
8) For every link judged as significant ($\bar{e}_{ji} > f_{th}$), choose as the edge orientation the direction with higher frequency observed along the W learned structures:

$$e_{ij}^* = \begin{cases} 0 \text{ and } e_{ji}^* = 1, & \text{if } (f_{eij} < f_{eji}) \\ 1 \text{ and } e_{ji}^* = 0, & \text{otherwise} \end{cases}, \tag{6}$$

where i, $j \in \{1, ..., n\}$, $f_{e_{ij}} = \frac{1}{W} \sum_{w=1}^{W} e_{ij}^w$ and $f_{e_{ji}} = \frac{1}{W} \sum_{w=1}^{W} e_{ji}^w$.

D. Optimal Bin Size Selection

The approach used in this investigation is based on the discrete DBN model. When the dataset under analysis is formed by time series variables, the data must be quantised in order to limit the number of states and make the application computationally feasible [3]. In this sense, an important task is to determine the optimal quantisation level taking care of this process can result in the mischaracterization of the data [3], [40].

For an optimal quantisation of each variable, a good option is the method for selecting the bin size of a time histogram proposed by [41]. This method selects the bin size from the spike counts statistics alone, so that the resulting bar or line graph time histogram best represents the signal. The following steps describe the process for data quantisation:

1) define the min number of bins (N_{min}) and the max number of bins (N_{max}) to be tested;
2) for N ranging from N_{min} to N_{max}:
   a) divide the observations of period T of a variable into N bins of width Δ;
   b) count the number of spikes h_i from all n sequences that enter the ith bin;
   c) calculate the mean ($\overline{h}$) and variance (var) of the number of events h_i;
   d) compute the cost function: $C_n(\Delta) = \frac{2\overline{h} - var}{(n\Delta)^2}$;
3) $N_{optimum}$ and $\Delta_{optimum}$ is when N minimises $C_n(\Delta)$;
4) divide the observations of the period T into $N_{optimum}$ bins of width $\Delta_{optimum}$. At this point, all observations were compressed in $N_{optimum}$ bins.

E. Information Theory Concepts: Mutual Information

Based on the information theory, the concept of Mutual Information (MI) arises as an indication of mutual dependency between variables [42] and can be used for non-linear relationships [43]. MI provides a measure of the amount of information discovered about a random variable through knowledge of other variables [43]. Given two variables v_1 and v_2, the MI measures the reduction of uncertainty about v_1 by knowing v_2, and vice versa.

During applications to forecasting, the future states are predicted based on information from the past. Due to this, MI has been applied in different situations [3], [29], [44] to evaluate the shared information between the original time series and its lagged version. With the use of MI, it is possible to perform feature selection, select how many lagged variables should be selected as a new feature [29], and select a reasonable forecast horizon (Δ_p) given the available information [3].

The MI between two random variables v_1 and v_2 is defined in [43] as

$$MI(v_1; v_2) = \sum_{x \in \Re} \sum_{y \in \Re} p(x, y) \log \frac{p(x, y)}{p(x)\, p(y)} \tag{7}$$

where $p(x, y)$ is the joint probability mass function of v_1 and v_2, $p(x)$ is the marginal probability mass function of v_1 and $p(y)$ is the marginal probability mass function of v_2. The higher the MI value, more information can be obtained about v_1 from v_2, i.e, the uncertainty of v_1 reduces [43].

To scale the measure between 0 (no mutual information) and 1 (perfect correlation), the MI can be normalised by minimal entropy $\min[\mathrm{H}(x_{i,t}), \mathrm{H}(x_{i,t-k})]$ [3], resulting in the Normalised Mutual Information (NMI).

Entropy is a concept used in various fields, including physics, information theory, thermodynamics, and mathematics. In [45], the authors presented a historical background on the evolution of the term "entropy", and provides mathematical evidence and logical arguments regarding its interconnection in various scientific areas.

In [46], the concept of information theory with the concept of entropy was presented. Entropy is a measure of the average amount of information required to represent or transmit a message from a given set of possible messages. It quantifies the uncertainty associated with a random variable or a probability distribution. The higher the entropy, the more uncertain or unpredictable the information is. Considering a discrete random variable v with probability distribution $p(x)$ and n states, the average information content about v is given by the Shannon entropy:

$$H(v) = -\sum_{i=1}^{n} p_i(x) \log p_i(x). \tag{8}$$

III. Materials and Methods

A. Multi-Source Power Generation Systems Data

In this investigation, it was used the electricity grid data of Belgium, Germany, Spain, and Portugal. All data is publicly available by the European Network of Transmission System Operators for Electricity (ENTSO-E) transparency platform [47]. ENTSO-E is a central collection and publication of electricity generation, transportation, consumption data, and information about energy prices for different European countries.

The requested data for each country comprises records from January 1, 2019 to December 31, 2021 with a one-hour sampling rate. The dataset contains temporal data from different energy sources, consumption, hour and date. Using the collected data and the concepts presented in Section II-A, the variable Emissions ($E_{t,s}$) was added to the dataset.

B. CO_2 Emissions Forecasting trough Evolving Dynamic Bayesian Networks

This section describes the steps of the proposed method to perform CO_2 emissions forecast. As previously mentioned, the model utilised in this investigation is a discrete approach. Due to this, in the first step, each variable of all datasets was quantised using the optimal bin size selection approach described in Section II-D. The min and the max number of bins tested were 4 and 40 respectively. After data quantisation, NMI was used to select the forecast horizon (Δ_p). Using Δ_p, the stage of data pre-processing organize the dataset for the 2-DBN model. The variables at this point are doubled: the variables of τ are the original time series delayed by Δ_p and the original information is stored as $\tau + 1$. As structural learning of BN is an NP-hard problem [33], to reduce the complexity the step of data pre-processing is finalised by performing NMI among the target variable (emissions) and the other variables of the dataset to eliminate variables that are not relevant. All variables that the NMI with the target variable is less than the median are considered irrelevant and discarded.

In the second stage (structural learning), a new sub-dataset D_w collected on the day w already pre-processed is used to perform structural learning. For this purpose, the DBN structure G_w is learned using the Hill Climbing (HC) algorithm combined with the AIC score to traverse the search space visiting neighbour DAGs by deleting, adding, or reversing an arc, and the algorithm advances to the one that provides the highest improvement to the AIC score. The algorithm stops the search when no operation yields improvement to the score function, i.e. when finding a local maximum. The structure learned is reduced to the Markov Blanket of the target variable (emissions), the edges frequencies are updated and G^* is obtained by selecting the edges using the threshold f_{th}. If the threshold rejects all edges, is adopted that G^* is formed by a single directed edge from the target variable to the target variable in the next time slice (emissions (τ), emissions ($\tau + 1$)). It is important to highlight that the study in [29] empirically verified the convergence of the combination of a local search algorithm and AIC metric, resulting in satisfactory networks.

Stage 3 refers to parameter learning. Despite the possibility of cycles in G^* reducing due to the cutoff frequency, it theoretically does not ensure the absolute absence of one or more cycles. To deal with cycles, our proposal checks the presence of cycles in G^* and tries to eliminate them by reversing a single edge. If reversing a single edge does not eliminate the cycles, our approach reverses two edges and in the last case reverses one edge and eliminates one edge. After ensuring that there are no cycles in G^*, using G^* and the last seven days of pre-processed and prepared data, the CPT (Θ) is obtained. Θ describes the probabilities for each state conditioned to its parents' states.

The last step is responsible for the CO_2 emissions forecasting. For this purpose, our approach uses (G^*, Θ) and the last information as evidence to forecast the observation Δ_p hours ahead using MAP estimation [3], [48]. Then, the values predicted are transformed to continuous, go through a smoothing filter to mitigate noisy values, and are stored. After obtaining knowledge about the real data of $\tau + 1$, they are used to update the CPDs.

Fig. 1 shows a summary of the steps described above represented as a flowchart.

C. Performance Evaluation

The proposed method was evaluated by comparing the performance against widely used time-series forecasting methods. The competitor methods are: feedforwarding Multilayer Perceptron (MLP), a fully connected class of feedforward ANN [49], [50]; XGBoost, a decision tree-based machine learning algorithm that uses a Gradient boosting structure [50], [51]; traditional DBN with structure learned in one step. The traditional DBN was used in [3] to perform CO_2 in multi-source power generation systems. The authors pointed out in

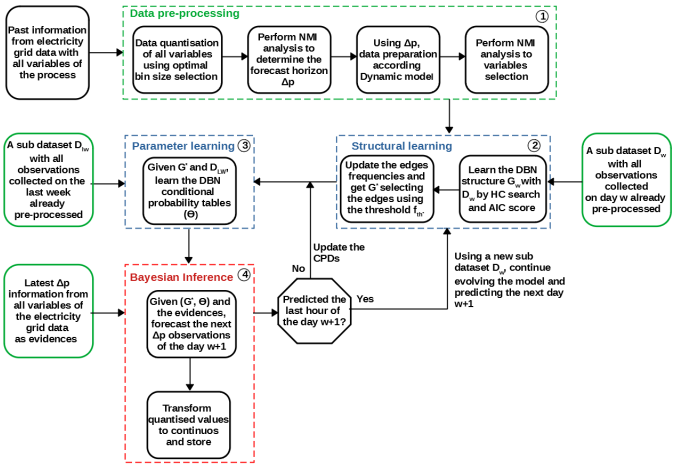

Fig. 1. Flowchart for forecasting CO_2 emissions using the EDBN proposed in this paper. The process is organised into 4 parts: data pre-processing, structural learning, parameter learning, and Bayesian inference.

future work the importance of approaches that allow evolving the entire structure over time. The comparison is interesting because it is a measurement against previous work and also a measure of performance improvement regarding the use of DBNs. All these models used the same interval of data used by the proposed EDBN model for training and fitting the model.

For performance evaluation, were used Normalised Root Mean Squared Error (NRMSE), Mean Absolute Error (MAE), and Median Absolute Error (MedAE). These are computed as follows [3], [26], [29].:

$$\text{NRMSE} = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{y_{\max} - y_{\min}}}, \tag{9}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} \mid y_i - \hat{y}_i \mid, \tag{10}$$

$$\text{MedAE} = \text{median}\left(\mid y_1 - \hat{y}_1 \mid, \ldots, \mid y_n - \hat{y}_n \mid\right), \tag{11}$$

where $\hat{y}_i$ is the forecasting value, y_i is the real value, N is the number of forecasts performed, $y_{\min}$ and $y_{\max}$ are the minimum and maximum values observed in the test set.

After calculating the performance metrics for all methods in all scenarios, the one-way ANOVA test is used to verify the null hypothesis that the methods have the same population mean (same performance) [52]. If the methods presented a statistically significant difference between the means, Tukey's post hoc test can be used to make pairwise comparisons between the means of each group to find out exactly which groups are different from each other [53].

D. Computational Resources

For all implementations, we used a laptop computer with an Intel(R) core i5 8th Gen processor, 16 GB of RAM, and Linux Mint 20.1 Ulyssa operating system. The algorithms were implemented in Python 3 language.

The datasets used in this manuscript, all dependencies of scientific packages used during the implementation and evaluation of this investigation and the scripts developed are publicly available on GitHub to ensure full reproducibility of this paper.

IV. Results and Discussion

Following the steps of the flowchart in Fig. 1, first, our approach realises data quantisation of all variables using optimal bin size selection. With a small number of bins, the data conversion results in the mischaracterization of the signal. Using a large number of bins, the high number of states increases the computational demand of the DBN method. On the other hand, using the optimal number of bins the number of states is reduced without inserting major errors.

After data quantisation, Fig. 2 illustrated the NMI among all variables for different lag values to select the forecast horizon. The first heatmap is using a delay of three hours and the second frame for twelve hours. Note that for three hours the NMI is higher than for twelve hours. For twelve hours of delay, even between variables and their lagged versions (main diagonal on the heat map), the NMI is close to 0, i.e., the variables no longer share information. The third frame shows the average NMI of the variables with their delayed versions for different delays. For three hours the average NMI is 0.34 and decreases rapidly as the lag increase, highlighting the difficulty of making long-term forecasts with the available information in the dataset. Important to mention that the NMI increases in a periodic way for cycles of 24 lags and this pattern was observed in other contexts of electric systems. [29], [54]–[56].

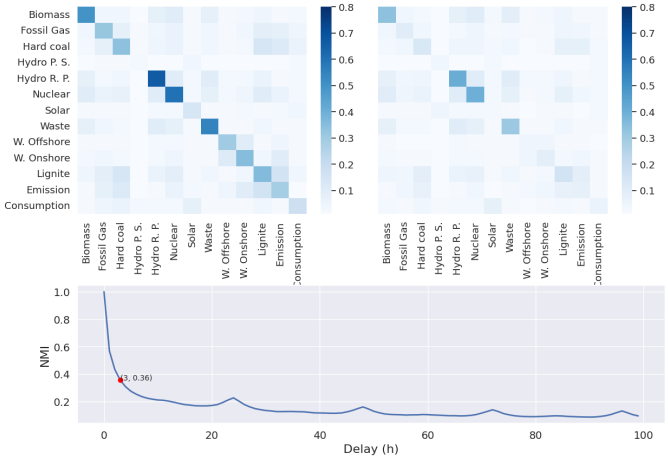

Fig. 2. NMI for the discrete dataset of Germany's power generation system at different lag values. First heatmap with a lag of three hours and the second used a delay of twelve hours. At the bottom of the figure is the average NMI of the variables with the lagged versions for different delays. In the heatmaps, darker colours represent more considerable NMI.

From the results of Fig. 2, we choose the forecast horizon and time window period of three hours ($\Delta_p = 3$). Using $\Delta_p = 3$, the datasets were prepared according to the dynamic model. The variables of $\tau + 1$ are the original ones and the variables of τ are the variables of $\tau+1$ delayed by three hours. With the dataset prepared for the dynamic model, we perform NMI analysis in relation to emissions variable to eliminate irrelevant variables to the forecast. All features with NMI less than the median were eliminated. Fig. 3 shows the results for each country. Important to mention that the set of relevant features varies greatly from country to country, reflecting the diversity of power generation profiles and the capability of our

selection proposal fully data-driven with no need for manual adjustments.

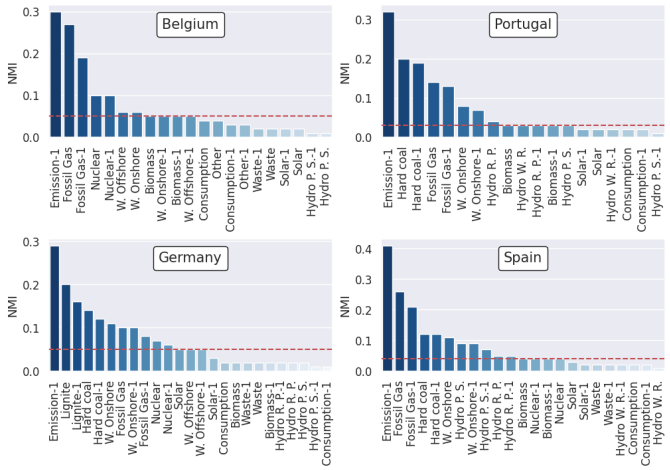

Fig. 3. Features selection using NMI in relation to emissions variable. For each country, all variables with NMI bigger than the median were selected. The horizontal dashed line represents the threshold of selection.

With the datasets already pre-processed, our method constantly adapts to the arrival of new datasets learning the partial structure of the DBN combining the AIC score metric with the hill-climbing search method and, then selecting the directed edges by the occurrence frequency using the analytical threshold. With $G*$ fitted ever using the past week of data to parameter learning and using the last observation available, the emissions forecast with a horizon of three hours ahead was carried out. The competitor methods used the same data interval and conditions as the proposed EDBN. Fig. 4 presents an illustration of one day of forecast performed by them using the dataset of Germany.

As the plots highlight, all methods were able to predict the behaviour without errors of great magnitude, evidencing that the forecast horizon selected is adequate for the dataset used. The proposed EDBN presented the best performance followed by the traditional DBN, XgBoost, and ANN. Our proposal showed better accuracy and forecast capability where forecast values accompanied the real values from the beginning to the end with smaller error values.

The example illustrated in Fig. 4 represents one day of the process. The dataset of each country comprises records from January 1, 2019 to December 31, 2021 with a one-hour sampling rate. The forecast was carryout from January 8, 2019 to the end. In Table II, a summary of the values observed for the forecasting performance metrics NRMSE, MAE and MedAE are presented for the EDBN and the other three methods used for comparison. The bold values indicate which method resulted in the lowest average value. The proposed EDBN presented the smallest NRMSE, MAE and MedAE for Germany and Spain. ANN was the best for Belgium and Portugal. The forecasting using conventional DBN results in all of the highest values for all metrics. In general, EDBN and ANN have better performance, followed by XgBoost which presented a similar performance.

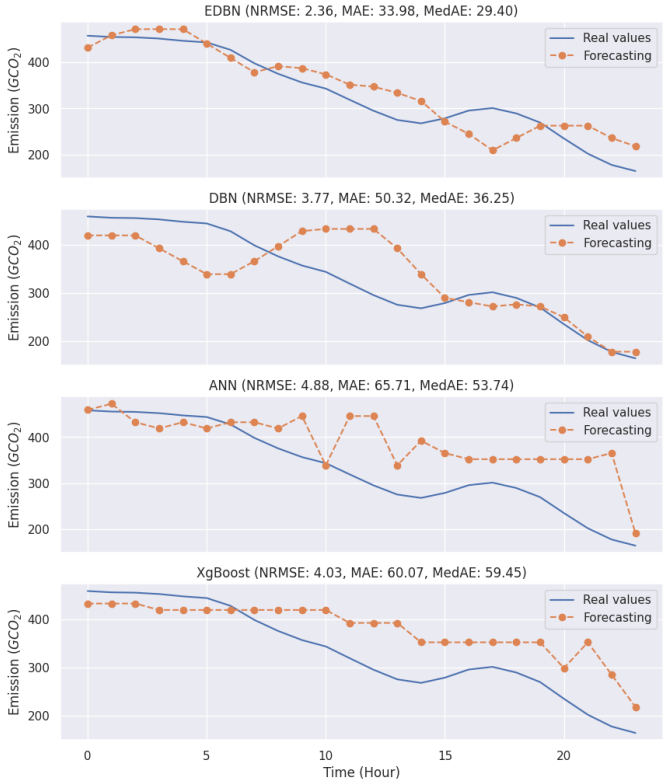

Fig. 4. Emissions forecasting for the proposed EDBN method and DBN, ANN, and XgBoost for one day. The solid line represents real values and the dashed line with markers illustrates the forecasting.

TABLE II
PERFORMANCE METRICS CALCULATED USING THE CO_2 EMISSIONS FORECASTING FOR THE INTERVAL FROM 8TH JANUARY 2019 TO 31ST DECEMBER 2021 IN BELGIUM, GERMANY, PORTUGAL AND SPAIN. THE VALUES PRESENTED ARE THE AVERAGE ± STANDARD DEVIATION.

Methods	NRMSE			
	Belgium	**Germany**	**Portugal**	**Spain**
EDBN	2.54 ± 0.72	**3.92 ± 1.12**	3.83 ± 1.40	**2.41 ± 0.74**
DBN	4.26 ± 2.64	4.94 ± 1.90	5.71 ± 3.65	3.26 ± 1.93
ANN	**2.50 ± 0.63**	4.21 ± 1.03	**3.76 ± 1.15**	2.80 ± 0.78
XgBoost	2.60 ± 0.65	4.43 ± 1.44	4.08 ± 1.44	2.77 ± 0.86
Methods	**MAE**			
	Belgium	**Germany**	**Portugal**	**Spain**
EDBN	14.28 ± 5.55	**34.82 ± 12.95**	30.90 ± 15.86	**13.03 ± 5.48**
DBN	25.92 ± 18.02	43.33 ± 19.30	46.92 ± 29.36	16.93 ± 10.24
ANN	**13.81 ± 4.98**	37.10 ± 13.40	**30.12 ± 14.02**	14.86 ± 5.46
XgBoost	14.28 ± 4.97	37.93 ± 14.03	31.83 ± 15.20	14.41 ± 5.45
Methods	**MedAE**			
	Belgium	**Germany**	**Portugal**	**Spain**
EDBN	12.44 ± 5.53	**30.20 ± 13.34**	26.82 ± 15.74	**11.42 ± 5.45**
DBN	25.10 ± 19.66	36.89 ± 18.93	43.82 ± 31.38	14.05 ± 9.10
ANN	**11.62± 4.96**	31.36 ± 13.55	**25.83 ± 13.95**	12.94 ± 5.62
XgBoost	12.01 ± 4.97	31.75 ± 14.60	26.53 ± 14.91	12.06 ± 5.21

Analysing Table II, our proposal was superior to the DBN for handling CO_2 emissions forecasting in multi-source power generation systems of Belgium, Germany, Portugal and Spain, i.e., a contribution of performance improvement in relation to DBN approach. For the Belgium generation system, in relation to ANN, the EDBN increases the average NRMSE, MAE, and MedAE at 1.60%, 3.40%, and 7.06% respectively. Regarding Germany, EDBN reduces in relation to the second

better (ANN) the average NRMSE, MAE and MedAE at 6.89%, 6.55%, and 3.84% respectively. ANN was the best for Portugal, where EDBN increases average NRMSE, MAE, and MedAE at 1.86%, 2.59%, and 3.83%. For the Spanish system, it was the scenario with the greatest difference. EDBN was the best and XgBoost the second better, with a reduction of the average NRMSE, MAE, and MedAE at 13.00%, 9.58%, and 5.31%.

Using all NRMSE calculated in each day of CO_2 emissions forecasting on Belgium, Germany, Portugal, and Spain, the one-way ANOVA test was carried out to verify the null hypothesis that the methods have the same performance and Tukey's post hoc test was used to make pairwise comparisons between which method. Table III presents the results of the comparison. The performance difference between the methods is statistically significant and the EDBN was the best method for the set of data used in this investigation. ANN presented the second-best performance followed by XgBoost and DBN.

TABLE III
MULTIPLE COMPARISONS OF MEANS USING TUKEY HSD WITH ALPHA 0.05. THE TEST INVESTIGATES IF EXISTS A DIFFERENCE BETWEEN THE NRMSE PRESENTED FOR THE METHODS DURING CO_2 EMISSIONS FORECASTING.

Method 1	Method 2	Mean Diff	p-value adj	Lower Diff	Upper Diff
EDBN	DBN	1.3657	0.0	1.2674	1.4640
EDBN	ANN	0.1442	0.001	0.0458	0.2425
EDBN	XgBoost	0.2949	0.0	0.1965	0.3932
DBN	ANN	-1.2215	0.0	-1.3199	-1.1235
DBN	XgBoost	-1.0708	0.0	-1.1692	-0.9725
ANN	XgBoost	0.1507	0.0005	0.0524	0.2491

In addition to the performance analysis, the time that the methods take during emissions forecasting was investigated. The proposed EDBN and the other methods are computationally efficient and could be used in online applications. All methods spend an average time in a matter of seconds, while data is sampled only every hour. It is important to highlight that the time spent by the EDBN to learn the parameters and make the prediction is smaller than by the DBN due to the fact that the selection of edges by frequency results in smaller structures.

V. CONCLUSION

This study investigated the performance and viability of performing CO_2 emissions forecasting in multi-source power generation systems using evolving discrete Dynamic Bayesian Networks. The use of an analytical threshold for selecting directed edges by the occurrence frequency as data is arriving allows the method constantly adapts and converge to a robust forecast model. The performance was evaluated against a set of widely used forecasting methods. For this purpose, we used real datasets of multi-source power generation systems of Belgium, Germany, Spain, and Portugal with records from January 1, 2019, to December 31, 2021.

The proposed approach showed to be capable of dealing with CO_2 emissions forecasting in the systems evaluated in this study. Comparing the results against a traditional DBN that not evolves the structure over time, our proposal was superior highlighting a contribution of performance improvement. It is important to highlight that the traditional DBN used for comparison was used in previous work to perform CO_2 emissions forecast in multi-source power generation systems. The proposed method was better when compared against ANN and XgBoost, with the difference in performance statistically significant. Moreover, the model also is computationally efficient with forecasting run-time in order of seconds. All these findings made our proposal a good option for embedding such an approach in CO_2 emissions forecasting fully data-drive and with real-time forecasting.

Future research includes the investigation of the proposed EDBN in other time horizons and energy aggregation levels and also in other power systems forecasting problems.

ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

REFERENCES

[1] M. R. Qader, S. Khan, M. Kamal, M. Usman, and M. Haseeb, "Forecasting carbon emissions due to electricity power generation in bahrain," *Environmental Science and Pollution Research*, vol. 29, no. 12, pp. 17 346–17 357, 2022.

[2] P. R. Jena, S. Managi, and B. Majhi, "Forecasting the co2 emissions at the global level: A multilayer artificial neural network modelling," *Energies*, vol. 14, no. 19, 2021.

[3] T. M. O. Santos, J. N. O. Júnior, M. Bessani, and C. D. Maciel, "Co2 emissions forecasting in multi-source power generation systems using dynamic bayesian network," in *2021 IEEE International Systems Conference (SysCon)*, 2021, pp. 1–8.

[4] H. HUANG and F. LI, "Bidding strategy for wind generation considering conventional generation and transmission constraints," *Journal of Modern Power Systems and Clean Energy*, vol. 3, pp. 51–62, 2015.

[5] NASA. (2020) Vital signs - carbon dioxide. [Online]. Available: https://climate.nasa.gov/vital-signs/carbon-dioxide/

[6] A. Jahanger, M. Usman, and P. Ahmad, "A step towards sustainable path: The effect of globalization on china's carbon productivity from panel threshold approach," *Environmental Science and Pollution Research*, vol. 29, pp. 8353–8368, 2021.

[7] R. Amna Intisar, M. R. Yaseen, R. Kousar, M. Usman, and M. S. A. Makhdum, "Impact of trade openness and human capital on economic growth: A comparative investigation of asian countries," *Sustainability*, vol. 12, no. 7, 2020.

[8] L. Fiorini and M. Aiello, "Energy management for user's thermal and power needs: A survey," *Energy Reports*, vol. 5, pp. 1048 – 1076, 2019.

[9] N. D. Bokde, B. Tranberg, and G. B. Andresen, "Short-term co2 emissions forecasting based on decomposition approaches and its impact on electricity market scheduling," *Applied Energy*, vol. 281, p. 116061, 2021.

[10] M. Panait, L. R. Janjua, S. A. Apostu, and C. Mihăescu, "Impact factors to reduce carbon emissions. evidences from latin america," *Kybernetes*, 2022.

[11] X. Hu, P. Li, and Y. Sun, "Minimizing energy cost for green data center by exploring heterogeneous energy resource," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 1, pp. 148–159, 2021.

[12] C. ZHANG, Y. DING, Q. WANG, Y. XUE, and J. OSTERGAARD, "Uncertainty-averse transco planning for accommodating renewable energy in co2 reduction environment," *Journal of Modern Power Systems and Clean Energy*, vol. 3, pp. 24–32, 2015.

[13] L. Fiorini and M. Aiello, "Household co2-efficient energy management," *Energy Informatics*, vol. 1, pp. 22–34, 2018.

[14] L. Liu, H. Sun, C. Li, Y. Hu, T. Li, and N. Zheng, "Exploring customizable heterogeneous power distribution and management for datacenter," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 12, pp. 2798–2813, 2018.

[15] S. Bouziane and M. Khadir, "Predictive agents for the forecast of co2 emissions issued from electrical energy production and gas consumption," *Advances in Intelligent Systems and Computing*, vol. 1076, pp. 183–191, 2020.

[16] Z. Xu, L. Liu, and L. Wu, "Forecasting the carbon dioxide emissions in 53 countries and regions using a non-equigap grey model," *Environmental Science and Pollution Research*, vol. 28, p. 15659–15672, 2021.

[17] M. O. Faruque, M. A. J. Rabby, M. A. Hossain, M. R. Islam, M. M. U. Rashid, and S. Muyeen, "A comparative analysis to forecast carbon dioxide emissions," *Energy Reports*, vol. 8, pp. 8046–8060, 2022.

[18] M. Emami Javanmard and S. Ghaderi, "A hybrid model with applying machine learning algorithms and optimization model to forecast greenhouse gas emissions with energy market data," *Sustainable Cities and Society*, vol. 82, p. 103886, 2022.

[19] F. Sun and T. Jin, "A hybrid approach to multi-step, short-term wind speed forecasting using correlated features," *Renewable Energy*, vol. 186, pp. 742–754, 2022.

[20] S. Zhang and J. J. Yu, "Bayesian deep learning for dynamic power system state prediction considering renewable energy uncertainty," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 913–922, 2022.

[21] X. Liu, D. Tang, and Z. Dai, "A bayesian game approach for demand response management considering incomplete information," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 492–501, 2022.

[22] Y. Zhao, W. Zhu, M. Yang, and M. Wang, "Bayesian network based imprecise probability estimation method for wind power ramp events," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 6, pp. 1510–1519, 2021.

[23] J. Chen, E. Chu, Y. Li, B. Yun, H. Dang, and Y. Yang, "Faulty feeder identification and fault area localization in resonant grounding system based on wavelet packet and bayesian classifier," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 760–767, 2020.

[24] J. QIU, Z. DONG, J. ZHAO, K. MENG, F. LUO, K. P. WONG, and C. LU, "A low-carbon oriented probabilistic approach for transmission expansion planning," *Journal of Modern Power Systems and Clean Energy*, vol. 3, pp. 14–23, 2015.

[25] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*, 1st ed. New York, NY, USA: Basic Books, Inc., 2018.

[26] T. M. de Oliveira Santos, I. Nunes da Silva, and M. Bessani, "Evolving dynamic bayesian networks by an analytical threshold for dealing with data imputation in time series dataset," *Big Data Research*, vol. 28, p. 100316, 2022.

[27] H. Wang, L. Wang, Q. Yu, Z. Zheng, A. Bouguettaya, and M. R. Lyu, "Online reliability prediction via motifs-based dynamic bayesian networks for service-oriented systems," *IEEE Transactions on Software Engineering*, vol. 43, no. 6, pp. 556–579, 2017.

[28] Q. Meng, Y. Wang, J. An, Z. Wang, B. Zhang, and L. Liu, "Learning non-stationary dynamic bayesian network structure from data stream," in *2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*, 2019, pp. 128–134.

[29] M. Bessani, J. A. Massignan, T. M. Santos, J. B. London, and C. D. Maciel, "Multiple households very short-term load forecasting using bayesian networks," *Electric Power Systems Research*, vol. 189, p. 106733, 2020.

[30] N. Bassamzadeh and R. Ghanem, "Multiscale stochastic prediction of electricity demand in smart grids using bayesian networks," *Applied Energy*, vol. 193, pp. 369 – 380, 2017.

[31] D. Weisser, "A guide to life-cycle greenhouse gas (ghg) emissions from electric supply technologies," *Energy*, vol. 32, no. 9, pp. 1543 – 1559, 2007.

[32] R. E. Neapolitan, *Learning Bayesian Networks*. Pearson Prentice Hall Upper Saddle River, 2004.

[33] T. J. Gross, M. Bessani, W. D. Junior, R. B. Araújo, F. A. C. Vale, and C. D. Maciel, "An analytical threshold for combining bayesian networks," *Knowledge-Based Systems*, vol. 175, pp. 36 – 49, 2019.

[34] M. Scutari, "An empirical-bayes score for discrete bayesian networks," in *Conference on Probabilistic Graphical Models*, 2016, pp. 438–448.

[35] Scutari, "Dirichlet bayesian network scores and the maximum relative entropy principle," *Behaviormetrika*, vol. 45, pp. 337 – 362, 2018.

[36] D. Koller, N. Friedman, and F. Bach, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[37] J. Pearl, "Chapter 3 - markov and bayesian networks: Two graphical representations of probabilistic knowledge," in *Probabilistic Reasoning in Intelligent Systems*, J. Pearl, Ed. San Francisco (CA): Morgan Kaufmann, 1988, pp. 77–141.

[38] M. Scutari and R. Nagarajan, "Identifying significant edges in graphical models of molecular networks," *Artificial Intelligence in Medicine*, vol. 57, no. 3, pp. 207 – 217, 2013.

[39] N. Friedman, M. Goldszmidt, and A. Wyner, "Data analysis with bayesian networks: A bootstrap approach," *Proc Fifteenth Conf on Uncertainty in Artificial Intelligence (UAI)*, 01 2013.

[40] N. Peker and C. Kubat, "Application of chi-square discretization algorithms to ensemble classification methods," *Expert Systems with Applications*, vol. 185, p. 115540, 2021.

[41] H. Shimazaki and S. Shinomoto, "A method for selecting the bin size of a time histogram," *Neural computation*, vol. 19, pp. 1503–27, 2007.

[42] X. Gu, J. Guo, L. Xiao, and C. Li, "Conditional mutual information-based feature selection algorithm for maximal relevance minimal redundancy," *Applied Intelligence*, vol. 52, no. 2, p. 1436 – 1447, 2022.

[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[44] N. Chakraborty and P. J. Van Leeuwen, "Using mutual information to measure time lags from nonlinear processes in astronomy," *Physical Review Research*, vol. 4, no. 1, 2022.

[45] J. Natal, I. Avila, V. Tsukahara, M. Pinheiro, and C. Maciel, "Entropy: From thermodynamics to information processing," *Entropy*, vol. 23, p. 1340, 10 2021.

[46] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 10 1948.

[47] ENTSO-E. (2021) Transparency platform restful api - user guide. [Online]. Available: https://www.entsoe.eu/data/transparency-platform/

[48] J. Su, H. Fang, W. Bao, H. Sun, J. Gao, and L. Zhao, "A maximum a posteriori estimation based method for estimating pulse time delay," *Advances in Space Research*, vol. 69, no. 11, p. 3966 – 3982, 2022.

[49] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Comparative evaluation of machine learning models and input feature space for non-intrusive load monitoring," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 5, pp. 1161–1171, 2021.

[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[51] J. Han, N. Liu, and J. Shi, "Optimal scheduling of distribution system with edge computing and data-driven modeling of demand response," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 989–999, 2022.

[52] Z. Liu, Y. Zhang, X. Wang, and D. Rodrigue, "Reinforcement of lignin-based phenol-formaldehyde adhesive with nano-crystalline cellulose (ncc): Curing behavior and bonding property of plywood," *Materials Sciences and Applications*, vol. 6, pp. 567–575, 2015.

[53] S. P and D. U., "Statistics in experimental cerebrovascular research: comparison of more than two groups with a continuous outcome variable," *Journal of Cerebral Blood Flow and Metabolism*, vol. 30, no. 9, pp. 1558–1563, 2010.

[54] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Applied Soft Computing*, vol. 54, pp. 246–255, 2017.

[55] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowledge-Based Systems*, vol. 82, pp. 29–40, 2015.

[56] A. Lahouar and J. B. H. Slama, "Day-ahead load forecast using random forest and expert input selection," *Energy Conversion and Management*, vol. 103, pp. 1040–1051, 2015.

Talysson M. O. Santos received the B. S. degree in electrical engineering in 2016 from the Federal University of Ouro Preto, at João Monlevade, Brazil. In 2018 received the M. S. degree in electrical engineering from the Federal University of São João del-Rei, Brazil. He is a current Ph.D. candidate in Electrical Engineering at the University of São Paulo, in São Carlos, Brazil. His research interests include data science, probabilistic models, signal processing, state estimation, and applications related to smart grids.

Michel Bessani received the B. S., M.Sc., and Ph.D. degrees, all in Electrical Engineering, respectively, in 2012, 2015, and 2018 from the University of São Paulo, at São Carlos, Brazil. He is now an assistant teacher at the Department of Electrical Engineering, of the Federal University of Minas Gerais, Belo Horizonte, Brazil. His research areas include systems reliability and resilience, statistical modeling, stochastic simulation, and computational intelligence.

Ivan N. Da Silva received a B.S. degree in Computer Science and an Electrical Engineering degree, both from the Federal University of Uberlândia, MG, Brazil, respectively, in 1991 and 1992. He received M.Sc. and Ph. D. degrees, both in Electrical Engineering, from the University of Campinas, Brazil, respectively, in 1995 and 1997. He is now a Full Professor at the Department of Electrical Engineering, of the University of São Paulo, in São Carlos, Brazil. He has been a researcher at the CNPq since 2000. His research areas are related to intelligent automation, including electric power systems, intelligent control of machines and equipment, design of intelligent systems architecture, and identification and systems optimisation.