Machine Learning Approaches for Predictions of CO2 Emissions in the Building Sector

Spyros Giannelos
Imperial College London
London, United Kingdom
s.giannelos@imperial.ac.uk

Federica Bellizio,
Urban Energy Systems Laboratory
Swiss Federal Laboratories for Materials Science and Technology,
Dübendorf, Switzerland
federica.bellizio@empa.ch

Goran Strbac
Imperial College London
London, United Kingdom
g.strbac@imperial.ac.uk

Tai Zhang
Imperial College London
London, United Kingdom
tai.zhang22@imperial.ac.uk

***Abstract*—The building sector has historically accounted for around 50% of the energy-related carbon dioxide (CO2) emissions on a global scale. As a result, it attracts significant attention as part of the worldwide effort to decarbonize the energy system. This paper presents and compares a variety of Machine Learning (ML)-based approaches for long-term predictions of CO2 emissions from buildings until the year 2050. These approaches include Linear Regression, ARIMA (Autoregressive Integrated Moving Average), Shallow Neural Networks, and Deep Neural Networks, all conducted using both univariate and multivariate modelling and with different approaches for the Features Extraction process; namely, the lagged values approach and the polynomial transformation. The analysis is conducted for different regions of the world including Brazil, India, China, South Africa, the United States, Great Britain, the World-average and the European Union. A variety of tests are conducted to evaluate and compare the predictive performance of the different ML approaches.**

***Index Terms*—ARIMA, CO2 emissions, Linear Regression, Machine Learning, Time Series Forecasting, Neural Networks**

I. INTRODUCTION

The prospect of extreme weather events due to climate change has motivated a worldwide effort to limit greenhouse gas emissions from various energy sectors [1]-[3]. Particularly, the emissions related to the building sector - industrial, commercial, and residential - originate from burning fossil fuels (oil, coal, and natural gas) for the generation of heat and electricity consumed by this sector [4]. This process significantly contributes to global warming, according to which, when sunlight reaches the Earth, only some of it is able to escape back into space, while the rest is absorbed by the greenhouse gases (such as carbon dioxide, methane, and nitrous oxide), thereby leading to a gradual warming of the Earth's surface. Over time, this effect has the potential to cause the melting of ice caps, and the rising of sea levels, among other impacts [5].

Hence, it is imperative to reduce greenhouse gas emissions in general and specifically those related to the building sector. This may require the adoption of a wide range of measures such as increasing the buildings' energy efficiency (e.g., by increasing insulation and optimizing HVAC systems), enhancing their potential to be producers of renewable electricity (e.g., using solar Photovoltaic panels and wind turbines for electricity and heating), installing smart technologies (e.g., automation for temperature and lights control and energy storage appliances), and implementing more efficient building design (e.g., using natural ventilation and lighting, as well as sustainable building materials) [6]. New policies addressing the next frontier of energy and emission reductions are needed [7].

In this context, various countries have adopted legally binding measures for the reduction of CO2 emissions related to the building sector [8]. Brazil has adopted the AQUA green building certification that incentivizes developers to adopt environmentally friendly practices in the building construction process. India has established guidelines for energy-efficient building construction as well. China has set out the Green Building Action Plan that includes mandatory requirements for the construction of green buildings in government-funded projects. South Africa has adopted the SANS 10400-XA energy efficiency standards that all new buildings must comply with. The United States has adopted the LEED Certification (Leadership in Energy and Environmental Design), which is a green-building standard. The United Kingdom has committed to net zero by 2050, which influences building regulations to be more environmentally stringent over time. The European Union has adopted the Energy Performance of Buildings Directive (EPBD) that aims for all new buildings to achieve net zero by 2050. On a global scale, the United Nations has developed UNEP (United Nations

Environmental Program), which includes several initiatives for sustainable buildings including the Building and Construction Climate Initiative and the Global Alliance for Buildings and Construction (Global ABC) network.

All these initiatives are based on time-series forecasting of CO2 emissions [9], which can help identify which regions are expected to be the most significant contributors to CO2 emissions in the future. This information can allow policymakers to target interventions more effectively and design policies to reduce CO2 emissions, such as through carbon taxes, renewable energy mandates and emissions standards. In this context, Machine Learning (ML) can play an instrumental role in generating forecasts of CO2 emissions for the buildings sector [10]. Its ability to handle data and learn from data patterns can make it a valuable tool for predictive modelling in climate change and power systems in general [11] –[13].

In this paper, we present the application of ML algorithms such as ARIMA [14]-[16], Linear Regression [17]-[19], and Neural Networks [20]-[23] for implementing time-series forecasts of CO2 emissions related to the building sector. These algorithms all constitute autoregressive models since the future values of a time series are predicted using as inputs past values of the same time series. They are all data-driven, meaning that they rely on historical data to make future predictions. In this paper, the performance of these models is evaluated using metrics such as the Mean Absolute Percentage Error (MAPE), which gives an indication of how well the model is likely to perform on unseen datasets.

In this context, the key contributions of this research are as follows:

- For the first time in the literature, a wide variety of ML predictive models are applied to CO2 emission datasets related to the building sector.
- For the first time in the literature, forecasting analysis and insights are provided for a number of regions across the world regarding CO2 emissions from the buildings sector, which can be very beneficial to our collective effort to tackle climate change.
- A comprehensive comparison is made in terms of performance metrics for a wide range of algorithms and implementations.

In terms of structure, the first section of this paper includes the introduction that sets out the context of the problem and the overall objectives of this research. The second section presents the forecasting methodology that is used in this paper step-by-step. This methodology starts with the *Data Preprocessing* stage and concludes with *Model Selection;* the last step selects the ML models to be used for the long-term time-series forecasts of CO2 emissions related to the building sector, across various countries and until far into the future (specifically, until the year 2050). This is the final validation of the selected ML models provided in this section. Finally, this paper concludes with a discussion of the results, including possible avenues for future work.

II. Methodology

A. *Introduction*

This section presents the four-step ML-based methodology, which is illustrated in Figure 1 below. Note that this methodology is the basis for all types of ML algorithms. In this paper, we have focused on the following algorithms: Linear Regression (univariate and multivariate), ARIMA (univariate and multivariate), as well as Shallow and Deep Neural Networks (both univariate and multivariate versions). These are very high-performance algorithms for time series forecasting; however, they have never been used before in the context of forecasting CO2 emissions from the buildings sector [24]. According to this methodology, the first step includes the *Data Preprocessing* stage, with the exploration of the input data from a reliable source and their further processing. Then, the *Features Extraction* process follows, which involves the transformation of the data into input features for the ML predictive models [25]. Specifically, for Features Extraction, we applied the lagged values approach and the 3rd-degree polynomial transformation. The third step constitutes the generation of the training and test set predictions. This is done through the rolling predictions method according to which the predictions are generated through an iterative process which gradually increases the size of the training dataset. Finally, the *Model Selection* step includes the overfitting analysis, as well as the naïve model benchmark test and the determination of which models are selected and trained for the final forecasting analysis.

Figure 1. The four-step Machine Learning Methodology used in this paper for the generation of forecasts.

B. *Data Preprocessing*

The first step of the methodology is called *Data Preprocessing* and involves the exploration and processing of the data from a reliable source. To the best knowledge of the authors, data related to the CO2 emissions from the building sector are only available in the World Bank's database. This database includes the annual CO2 emissions from the building sector for different countries and regions, expressed as a percentage of the total CO2 emissions related to all economic sectors of the corresponding country/region. Note that the focus of this work is not on the CO2 emissions in general but specifically on the CO2 emissions related to the buildings sector.

Specifically, the data obtained from the World Bank's database and used in this paper cover the period between 1971 and 2014 and there is a single value for every year, which is the annual average. As such, a country-specific dataset consists of 44 observations equal to the number of years between 1971 and 2014. Note that it was not possible to access more recent data, and the World Bank was the only source for which we could find reliable data related to CO_2 emissions specifically related to the building sector.

These values have been obtained for the following regions: Brazil, India, China, South Africa, the United States, Great Britain, the world average and the European Union. The selection of these regions is based on their economic development.

Specifically, Brazil is a regional power and part of the BRICS group of emerging economies, as well as it is a country rich in natural resources, including oil, which can have implications for CO_2 emissions. Similarly, India is an emerging economy, also part of the BRICS group, and is the second most populous country in the world, which inherently makes the level of its emissions significant on a global scale. China is the world's largest emitter of CO_2 emissions and an economic powerhouse. South Africa is also an emerging economy, part of the BRICS group, and heavily reliant on coal for meeting its energy requirements, making it a significant emitter on the global stage. The United States being the largest economy in the world, can significantly influence international climate policies. Similarly, Great Britain a major global financial hub and a G7 member, has considerable influence on international climate policies. The European Union is one of the world's largest economies and emitters and has set ambitious sustainability targets. Finally, the world average can serve as a baseline for comparing individual countries' efforts to reduce the level of their emissions.

Following the input data exploration, the *Data Preprocessing* stage involves conducting analysis on the stationarity of the datasets. A stationary time series shows constant statistical properties over time. This is an important concept in time series forecasting analysis, including the modelling of CO_2 emissions, as the stationarity of the time series data ensures that the predictive models can be applied more confidently with the expectation that they will perform better on new, unseen data. Specifically, detecting stationarity is essential for accurately determining correlations between datasets of different countries, which is a crucial step in evaluating the adequacy of multivariate models. In other words, highly correlated time series input data justifies the use of multivariate modelling.

In our analysis, we identified correlations between the datasets, which warrant the use of multivariate forecasting modelling. In particular, we have used the KPSS statistical test to check for stationarity with its results indicating that the datasets for all countries, except for Brazil, South Africa and Great Britain, are non-stationary. Since only stationary data are correlated, and the correlation is crucial to use multivariate models, the non-stationary time series were differenced, with the resulting ones being stationary.

Following the stationarity analysis, correlations were detected. For example, 68.4% between the datasets of the World and the United States, 60.76% between the World and the European Union, and 38% between the European Union and the United Kingdom. As a result, the presence of correlation means that the multivariate modelling will result in accurate predictions of future values of a given time series, as both the past values of the same time series and the past values of other time series correlated to it, are relevant features.

C. Features Extraction

The next step of the methodology is the *Features Extraction* process, which is the process of transforming the input raw data into numerical input features that the ML models require so that they can produce outputs. That is, the models receive input features in order to produce output features, i.e., predictions or forecasts.

Several approaches can be used to transform the input raw data. In this work, we used the following two approaches: the lagged values approach, and the polynomial transformation. Both these models take the original data as input and generate as outputs the training and the test components of the features matrix, as well as the training component of the target variables. Note that these transformations can be used with both univariate and multivariate forecasting models. In this paper, we used a lag equal to 3 years, meaning that we used the previous 3 years to forecast the next year. This number is a hyperparameter appropriately selected by minimizing the model's error on an independent validation set.

Note that the former approach (i.e., the lagged values approach) requires the data to be stationary. For this reason, the time series of the original data that are non-stationary are differenced to become stationary and are then used in the analysis. This means that in the next stage in which predictions are performed, inversion of the differencing operation has to take place to obtain the actual predictions.

In contrast, the polynomial transformation does not involve the requirement for the time series to be stationary. For this reason, the original data were transformed using a third-degree polynomial.

D. Predictions and Errors

In this stage, the input time series data are split into training and testing sets through the process of rolling predictions. According to this process, the model is fitted iteratively to a different portion of the training data up to some year T, and then the fitted model is used to generate the predictions for the testing set for year T+1. In each of the iterations, the *Features Extraction* process is implemented either via the lagged values approach or via the polynomial transformation.

The predictive performance was evaluated using the MAPE as a metric. Such a metric expresses the average percentage difference between the actual data corresponding to the testing set (which covers the time period between 2005 and 2014 i.e., the last 10 years of the dataset) and the predicted values for the same time period. Note that the errors between actual and predicted values of the testing set, also known as test errors, constitute proxies of the overall forecasting error of the models. In other words, the test errors (MAPE) reflect the model's performance on new unseen datasets. Both the testing dataset (2005-2014) and the forecasting dataset (2015-2050) are unseen to the models since these have been trained (i.e., have "seen") using the training dataset exclusively.

Table I shows the predictive performance in terms of MAPE for the different regions when using Linear Regression models. By observing the mean, standard deviation and maximum errors in the last three rows, it can be concluded that the multivariate models performed much better than the univariate ones. Similarly, Tables II-III-IV show the same performance analysis when using ARIMA models, Shallow and Deep Neural Networks, respectively, concluding that the multivariate approach was superior to the univariate one for all the used models.

Dataset	***LR Univariate***	***LR Multivariate (Lags)***	***LR Multivariate (Polynomials)***
Brazil	93.01	6.09	21.03
India	15.68	9.69	4.15
China	74.74	15.81	17.22
South Africa	25.48	33.14	29.44
USA	8.52	5.22	5.70
Great Britain	10.53	5.77	6.77
World Average	3.79	5.06	2.63
European Union	11.02	4.73	6.33
Mean	*30.34*	*10.69*	*11.66*
Standard Deviation	*33.98*	*9.81*	*9.71*
Max	*93.01*	*33.14*	*29.44*

TABLE I. TEST SET ERRORS (MAPE) FOR LINEAR REGRESSION UNIVARIATE/MULTIVARIATE MODELS USING LAGS VALUES / POLYNOMIAL TRANSFORM (FEATURES EXTRACTION PROCESS).

Dataset	***ARIMA Univariate***	***ARIMA Multivariate (Lags)***	***ARIMA Multivariate (Polynomials)***
Brazil	2.17	6.09	8.18
India	10.07	9.69	11.17
China	89.69	15.81	20.50
South Africa	38.84	33.14	32.79
USA	8.36	5.22	5.07
Great Britain	5.73	5.77	7.76
World Average	4.20	5.06	7.33
European Union	4.36	4.73	7.76
Mean	*20.43*	*10.69*	*12.56*
Standard Deviation	*30.38*	*9.81*	*9.43*
Max	*89.69*	*33.14*	*32.79*

TABLE II. TEST SET ERRORS (MAPE) FOR ARIMA UNIVARIATE/MULTIVARIATE MODELS USING LAGS VALUES / POLYNOMIAL TRANSFORM (FEATURES EXTRACTION PROCESS).

Dataset	***SNN Univariate***	***SNN Multivariate (Lags)***	***SNN Multivariate (Polynomials)***
Brazil	4.49	10.83	10.57
India	13.72	5.10	5.61
China	44.35	4.43	28.07
South Africa	24.22	26.78	28.68
USA	8.67	5.62	6.53
Great Britain	5.50	7.85	6.49
World Average	2.38	2.02	2.44
European Union	4.87	7.19	5.02
Mean	*13.52*	*8.73*	*11.68*
Standard Deviation	*14.31*	*7.74*	*10.55*
Max	*44.35*	*26.78*	*28.68*

TABLE III. TEST SET ERRORS (MAPE) FOR SHALLOW NEURAL NETWORK (SNN) UNIVARIATE/MULTIVARIATE MODELS USING LAGS VALUES / POLYNOMIAL TRANSFORM (FEATURES EXTRACTION PROCESS).

.

Dataset	DNN Univariate	DNN Multivariate (Lags)	DNN Multivariate (Polynomials)
Brazil	5.59	12.69	6.95
India	12.75	6.58	3.96
China	28.47	3.75	11.59
South Africa	25.85	23.96	29.64
USA	8.32	5.14	5.58
Great Britain	9.38	8.29	6.75
World Average	4.33	1.65	2.70
European Union	4.45	7.14	5.10
Mean	*12.39*	*8.65*	*9.03*
Standard Deviation	*9.56*	*7.00*	*8.73*
Max	*28.47*	*23.96*	*29.64*

TABLE IV. TEST SET ERRORS (MAPE) FOR DEEP NEURAL NETWORK (DNN) UNIVARIATE/MULTIVARIATE MODELS USING LAGS VALUES / POLYNOMIAL TRANSFORM (FEATURES EXTRACTION PROCESS).

The following paragraph includes the last step of the methodology, namely the selection of the models for generating the final forecasts for the validation dataset.

E. Model Selection

This step involves the selection of ML models that can be used to generate forecasts until the year 2050. This selection will be made on the basis of whether each of the models can successfully pass the overfitting test (i.e., the models do not suffer from overfitting) and the naïve model benchmark test (i.e., the models indeed display acceptable test errors). The model which passes both tests is selected to generate forecasts of the validation dataset.

In particular, the overfitting analysis is achieved by comparing the training errors with the test errors. In this context, small differences between them indicate a lack of overfitting whereas large differences, i.e., greater than 10%, indicate that the models are fitted so well to the training datasets that they cannot generalize to unseen datasets, which indicates overfitting. Clearly, models that overfit cannot be used to generate forecasts.

In addition, the naive model benchmark test allows the detection of unacceptably high-test errors (MAPE). This is different from the overfitting test in that while the overfitting test focuses on the difference between the training and test errors, while the naïve model benchmark test focuses on the test errors themselves. It involves comparing the calculated test error (MAPE) with the test error of a naïve/simplistic model. A naive model is one whose predictions are found by a simple shift of the data by one year so that the CO2 emissions in year t+1 are equal to those in year t.

In this context, if the test error of the naïve model is smaller than the error of the ML model under study (e.g., Linear Regression), then the test error of the studied ML model is considered unacceptably high. In other words, the naïve model serves the purpose of a benchmark against which the test errors can be characterized as unacceptably high or not. After performing the aforementioned tests, the results are shown in the following tables.

Table V shows the Linear Regression models (i.e., univariate/ multivariate modelling, lagged/ polynomial feature extraction) that passed both aforementioned tests, meaning that they are the ones that are selected for the generation of forecasts, as they are considered to have the highest chance of yielding least error forecasts. Notice that none of the univariate models has passed both tests, which means that none of them can be used for predictions.

On the other hand, the multivariate models performed better with the ones fitted to the datasets of "European Union", "Great Britain", and "World Average" passing both tests.

LR Univariate	LR Multivariate (Lags)	LR Multivariate (Polynomials)
-	LR fitted to "European Union" dataset	LR fitted to "World Average" dataset
	LR fitted to "Great Britain" dataset	

TABLE V. LINEAR REGRESSION MODELS SELECTED FOR GENERATING FORECASTS DEPENDING ON THE TYPE OF MODELLING (UNIVARIATE / MULTIVARIATE) AND THE TYPE OF FEATURES EXTRACTION (LAGGED VALUES, POLYNOMIAL TRANSFORMATION)

Table VI shows the ARIMA models that passed both the overfitting test and the naïve model benchmark test. Notice that none of the multivariate models where the Polynomial transformation was used, has passed both tests. This is because multivariate ARIMA models should not use polynomials transformation as Features Extraction process since it relies on non-stationary data. Rather, such models should be used with stationary time series, which can happen only with the lagged values approach.

Therefore, only the univariate ARIMA models fitted to the datasets of "European Union", "Brazil" and "Great Britain" as well as the multivariate ARIMA models fitted to "European Union" and "Great Britain" datasets, and where the lagged values approach was used, were selected for the generation of forecasts.

ARIMA Univariate	*ARIMA Multivariate (Lags)*	*ARIMA Multivariate (Polynomials)*
LR fitted to "European Union" dataset	LR fitted to "European Union" dataset	-
LR fitted to "Brazil" dataset	LR fitted to "Great Britain" dataset	
LR fitted to "Great Britain" dataset		

TABLE VI. ARIMA MODELS SELECTED FOR GENERATING FORECASTS DEPENDING ON THE TYPE OF MODELLING (UNIVARIATE / MULTIVARIATE) AND THE TYPE OF FEATURES EXTRACTION (LAGGED VALUES, POLYNOMIAL TRANSFORMATION)

Finally, Tables VII-VIII show the Shallow and Deep Neural Network models that passed both the overfitting test and the naïve model benchmark test.

SNN Univariate	*SNN Multivariate (Lags)*	*SNN Multivariate (Polynomials)*
SNN fitted to "World Average" dataset	SNN fitted to "World Average" dataset	SNN fitted to "World Average" dataset
SNN fitted to "European Union" dataset		SNN fitted to "European Union" dataset
SNN fitted to "Brazil" dataset		
SNN fitted to "Great Britain" dataset.		

TABLE VII. SHALLOW NEURAL NETWORK MODELS SELECTED FOR GENERATING FORECASTS DEPENDING ON THE TYPE OF MODELLING (UNIVARIATE / MULTIVARIATE) AND THE TYPE OF FEATURES EXTRACTION (LAGGED VALUES, POLYNOMIAL TRANSFORMATION)

DNN Univariate	*DNN Multivariate (Lags)*	*DNN Multivariate (Polynomials)*
DNN fitted to the "European Union" dataset	DNN fitted to the "China" dataset	DNN fitted to the "World Average" dataset
	DNN fitted to the "World Average" dataset	DNN fitted to the "European Union" dataset

TABLE VIII. DEEP NEURAL NETWORK MODELS SELECTED FOR GENERATING FORECASTS DEPENDING ON THE TYPE OF MODELLING (UNIVARIATE / MULTIVARIATE) AND THE TYPE OF FEATURES EXTRACTION (LAGGED VALUES, POLYNOMIAL TRANSFORMATION)

Figure 2. and Figure 3. show examples of forecasts produced for the validation dataset using the selected ML models against non-selected models, respectively. A comparison of these figures emphasizes the improved quality of the former forecasts; observe that the latter contained negative values (which are physically infeasible for CO2 emissions) and abrupt non-realistic movements.

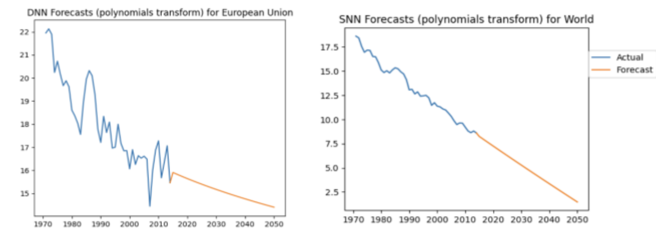

Figure 2. Forecasts of the CO2 emissions from the building sector using selected models, namely multivariate Deep Neural Network (DNN) and Shallow Neural Network (SNN) models fitted to the "European Union" and "World Average" datasets.

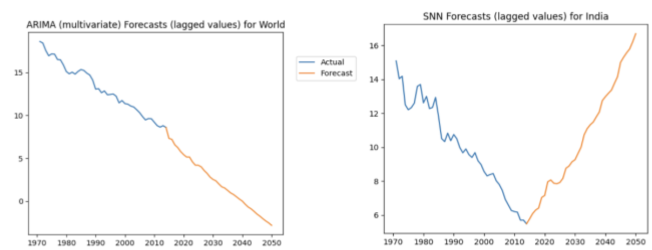

Figure 3. Forecasts of the CO2 emissions from the building sector using non-selected models.

III. CONCLUSIONS

This paper presents for the first time in the literature the application of a wide range of ML models to the forecasting analysis of CO2 emissions from the building sector across various regions in the world. The models that are used include Linear Regression, ARIMA, Shallow and Deep Neural Networks. Both univariate and multivariate modelling approaches have been used while considering different types of the *Features Extraction* process, namely the lagged values approach and the polynomial transformation.

According to the presented methodology, the analysis starts with the *Data Preprocessing* stage, followed *by Features Extraction*, the calculation of *Predictions and Errors*, and finally the *Selection* of those models that will be used for generating forecasts, based on their performance on the overfitting and the naïve model benchmark tests. The selected models are considered highly likely to generate forecasts of high accuracy given that the test errors (MAPE) are considered to be proxies of the forecasting errors.

Future work includes focusing on optimizing the value of hyperparameters using methodologies such as Backwards Induction and least-worst regret [27] as well as uncertainty analysis methods based on artificial neural networks [28]. The authors are also interested in evaluating the effect of external and contextual factors, such as the level of technological development and of GDP, on the forecasts of CO2 emissions in the various regions around the world [29]. Finally, the use of advanced decomposition methodologies applied to large-scale mathematical optimization problems [30]-[31] can offer significant insights into possible future pathways of CO2 emissions across different regions.

REFERENCES

[1] Bar Gai, D.H.; Ogunrinde, O.; Shittu, E. Self-Reporting Firms: Are Emissions Truly Declining for Improved Financial Performance? *IEEE End. Manag. Rev. 2020, 48, 163-170. https://doi.org/10.1109/emr.2020.2969405.*

[2] Available online: https://www.ipcc.ch/report/ar6/wg2/ (accessed on 1 September 2023).

[3] Rüdisüli M, Romano E, Eggimann S, and Patel MK. "Decarbonization strategies for Switzerland considering embedded greenhouse gas emissions in electricity imports". Energy Policy. 2022 Mar 1

[4] Available online: https://www.ieta.org/resources/Resources/COP/COP26-Summary-Report.pdf (accessed on 1 September 2023).

[5] Available online: https://unfccc.int/sites/default/files/english_paris_agreement.pdf (accessed on 1 September 2023).

[6] Available online: https://globalabc.org/resources/publications/2021-global-status-report-buildings-and-construction

[7] Skillington, Katie, Robert H. Crawford, Georgia Warren-Myers, and Kathryn Davidson. "A review of existing policy for reducing embodied energy and greenhouse gas emissions of buildings." *Energy Policy* 168 (2022): 112920.

[8] United Nations Department of Economic and Social Affairs, 2022. The Sustainable Development Goals Report 2022 – July 2022. New York, USA. Available online at https://unstats.un.org/sdgs/report/2022

[9] Available online: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks (accessed on 12 September 2023).

[10] Giannelos, S.; Moreira, A.: Papadaskalopoulos, D.; Borozan, S.; Pudjianto, D.; Konstantelos, I.; Sun, M.; Strbac, G. A Machine Learning Approach for Generating and Evaluating Forecasts on the Environmental Impact of the Buildings Sector. Energies 2023, 16, 2915. Available online: https://doi.org/10.3390/en16062915.

[11] F. Bellizio, J. L. Cremer and G. Strbac, "Transient Stable Corrective Control Using Neural Lyapunov Learning," in IEEE Transactions on Power Systems, vol. 38, no. 4, pp. 3245-3253, July 2023, doi: 10.1109/TPWRS.2022.3204459

[12] F. Bellizio, S. Karagiannopoulos, P. Aristidou and G. Hug, "Optimized Local Control for Active Distribution Grids using Machine Learning Techniques," 2018 IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, USA, 2018, pp. 1-5, doi: 10.1109/PESGM.2018.8586079.

[13] Federica Bellizio, Al-Amin B. Bugaje, Jochen L. Cremer, Goran Strbac,Verifying Machine Learning conclusions for securing Low Inertia systems, Sustainable Energy, Grids and Networks,Volume 30, 2022, 100656, ISSN 2352-4677,

[14] Contreras, J.; Espinolar, R.; Nogales, F.; Conejo, A. ARIMA models to predict next-day electricity prices. IEEE Trans. Power Syst. 2003, 18, 1014-1020. Available online: https://doi.org/10.1109/tpwrs.2002.804943

[15] Chen, P.; Pedersen, T.; Bak-Jensen, B.; Chen, Z. ARIMA – Based Time Series Model of Stochastic Wind Power Generation. IEEE Trans. Power Syst. 2009, 25, 667-676. Available online: https://doi.org/10.1109/tpwrs.2009.2033277

[16] Guo, J.; He, H.; Sun, C. ARIMA-Based Road Gradient and Vehicle Velocity Prediction for Hybrid Electric Vehicle Energy Management. IEEE Trans. Veh. Technol. 2019, 68, 5309-5320. Available online: https://doi.org/10.1109/tvt.2019.2912893

[17] Sarkar, M.R.; Rabbani, M.G.; Khan, A.R.; Hossain, M.M. Electricity demand forecasting of Rajshahi city in Bangladesh using fuzzy linear regression model. In Proceedings of the 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). Savar, Bangladesh, 21-23 May 2015; pp. 1-3. Available online: https://doi.org/10.1109/ICEEICT.2015.7307424.

[18] Zhao, Z.; Peng, Y.; Zhu, X.; Wei, X.; Wang, X.; Zuo, J. Research on Prediction of Electricity Consumption in Smart Parks Based on Multiple Linear Regression. In Proceedings of the 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 11-13 December 2020l pp. 812-816. https://doi.org/10.1109/ITAIC49862.2020.9338976

[19] Chen, Y.; Wu, C.; Qi, J. Data-driven Power Flow Method Based on Exact Linear Regression Equations. J. Mod. Power Syst. Clean Energy 2022, 10, 800-804, Available online: https://doi.org/10.35833/mpce.2020.000738

[20] Szkuta, B.; Sanabria, L.; Dillon, T. Electricity price short-term forecasting using artificial neural networks. IEEE Trans. Power Syst. 1999, 14, 851-857. Available online: https://doi.org/10.1109/59.780895.

[21] Alanis, A.Y. Electricity Prices Forecasting using Artificial Neural Networks. IEEE Lat. Am. Trans. 2018, 16, 105-111. Available online: https://doi.org/10.1109/tla.2018.8291461.

[22] Liu, X.; Wang. D.; Lin, S-B. Construction of Deep ReLU Nets for Spatially Sparse Learning. IEEE Trans. Neural Networks Learn. Syst. 2022. Available online: https://doi.org/10.3390/en15093334.

[23] Jena, P.R.; Managi, S.; Mahji, B. Forecasting the CO2 Emissions at the Global Level: A Multilayer Artificial Neural Network Modelling. Energies 2021, 14, 6336. Available online: htttps://doi.org/10.3390/en1496336

[24] Ahmed, Nesreen K., Amir F. Atiya, Neamat El Gayar, and Hisham El-Shishiny. "An empirical comparison of machine learning models for time series forecasting." Econometric reviews 29, no. 5-6 (2010)

[25] Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., and Saeed, J. "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction." Journal of Applied Science and Technology Trends 1, no. 2 (2020): 56-70

[26] Available online: https://databank.worldbank.org/source/world-development-indicators/EN.CO2.BLDG.ZS (accessed on 29 September 2023).

[27] Giannelos, S.; Borozan, S.; Strbac, G. A Backwards Induction Framework for Quantifying the Option Value of Smart Charging of Electric Vehicles and the Risk of Stranded Assets under Uncertainty. Energies 2022, 15, 3334. Available online: https://doi.org/10.3390/en15093334.

[28] Jimenez-Martinez, M.; Alfaro-Ponce, M. Effects of synthetic data applied to artificial neural networks for fatigue life prediction in nodular cast iron. J. Braz. Soc. Mech. Sci. Eng. 2021, 43, 1-9. Available online: https://doi.org/10.1008/s40430-020-02747-y.

[29] Faruque, O.; Rabby, A.J.; Hossain, A.; Islam, R.; Rashid, M.U.; Muyeen, S. A Comparative Analysis to forecast carbon dioxide emissions. Energy Rep. 2022, 8, 8046-8060. Available online: https://doi.org/10.1016/j.egyr.2022.06.025.

[30] Giannelos, S.; Jain, A.; Borozan, S.; Falugi, P.; Moreira, A.; Bhakar, R.; Mathur, J.; Strbac, G. Long-Term Expansion Planning of the Transmission Network in India under Multi-Dimensional Uncertainty. Energies 2021, 14, 7813. Available online: https://doi.org/10.3390/en14227813.

[31] P. Falugi, I. Konstantelos, and G. Strbac, "Application of Novel Nested Decomposition Techniques to Long-Term Planning Problems," 2016 Power Systems Computation Conference (PSCC), Genoa, Italy, 2016, pp. 1-8, doi: 10.1109/PSCC.2016.7540872