ENERGY 2024

The Fourteenth International Conference on Smart Grids, Green Communications
and IT Energy-aware Technologies

March 10th –14th, 2024

Athens, Greece

ENERGY 2024 Editors

Vivian Sultan, California State University, Los Angeles, USA

Eric MSP Veith, Carl von Ossietzky University Oldenburg, Germany

ENERGY 2024

Foreword

The Fourteenth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies (ENERGY 2024), held between March 10 – 14, 2024, continued the event considering Green approaches for Smart Grids and IT-aware technologies. It addressed fundamentals, technologies, hardware and software needed support, and applications and challenges.

There is a perceived need for a fundamental transformation in IP communications, energy-aware technologies and the way all energy sources are integrated. This is accelerated by the complexity of smart devices, the need for special interfaces for an easy and remote access, and the new achievements in energy production. Smart Grid technologies promote ways to enhance efficiency and reliability of the electric grid, while addressing increasing demand and incorporating more renewable and distributed electricity generation. The adoption of data centers, penetration of new energy resources, large dissemination of smart sensing and control devices, including smart home, and new vehicular energy approaches demand a new position for distributed communications, energy storage, and integration of various sources of energy.

We take here the opportunity to warmly thank all the members of the ENERGY 2024 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ENERGY 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ENERGY 2024 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ENERGY 2024 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the fields of smart grids, green communications and IT energy-aware technologies.

We are convinced that the participants found the event useful and communications very open. We also hope that Athens provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

ENERGY 2024 Chairs:

ENERGY 2024 Steering Committee
Eric MSP Veith, Carl von Ossietzky University – Oldenburg, Germany
Dragan Obradovic, Siemens - Corporate Technology, Munich, Germany
Mark Apperley, University of Waikato, New Zealand
Michael Negnevitsky, University of Tasmania, Australia
Vivian Sultan, California State University Los Angeles, USA
Steffen Fries, Siemens, Germany

ENERGY 2024 Publicity Chairs
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

ENERGY 2024

Committee

ENERGY 2024 Steering Committee
Eric MSP Veith, OFFIS e.V. – Oldenburg, Germany
Dragan Obradovic, Siemens - Corporate Technology, Munich, Germany
Mark Apperley, University of Waikato, New Zealand
Michael Negnevitsky, University of Tasmania, Australia
Vivian Sultan, California State University Los Angeles, USA
Steffen Fries, Siemens, Germany

ENERGY 2024 Publicity Chairs
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

ENERGY 2024 Technical Program Committee

Roozbeh Abolpour, Shiraz University, Iran
Santosh Aditham, University of South Florida, USA
Kodjo Agbossou, Université du Québec à Trois-Rivières, Canada
Awadelrahman M. A. Ahmed, University of Oslo, Norway
Miltos Alamaniotis, University of Texas at San Antonio, USA
Kamal Al-Haddad, École de technologie supérieure, Montreal, Canada
Ahmed Al-Salaymeh, The University of Jordan, Amman, Jordan
Amjad Anvari-Moghaddam, AAU Energy, Denmark
Mark Apperley, University of Waikato, New Zealand
Parimal Acharjee, NIT Durgapur, India
Paranietharan Arunagirinathan, Clemson University, USA
Ashrant Aryal, Texas A&M University, USA
Adela Bara, Bucharest University of Economic Studies, Department of Economic Informatics and Cybernetics, Bucharest, Romania
Rico Berner, Ambrosys GmbH, Potsdam, Germany
Lasse Berntzen, University of South-Eastern Norway, Norway
Vito Calderaro, University of Salerno, Italy
M. Girish Chandra, TCS Research & Innovation, India
Fathia Chekired, CDER-UDES, Algeria
Dana-Alexandra Ciupageanu, University Politehnica of Bucharest, Romania
Daniele Codetta, University of Piemonte Orientale, Italy
Luigi Costanzo, Università degli Studi della Campania Luigi Vanvitelli, Italy
Fabio D'Agostino, University of Genova, Italy
Thusitha Dayaratne, Monash University, Australia
Payman Dehghanian, The George Washington University, USA
Margot Deruyck, Universiteit Gent - IMEC - WAVES, Belgium
Giovanna Dondossola, RSE, Italy
Virgil Dumbrava, University POLITEHNICA of Bucharest, Romania

Angela Russo, Politecnico di Torino, Italy
Eckehard Schöll, Technische Universität Berlin | Institut für Theoretische Physik, Germany
Tomonobu Senjyu, University of the Ryukyus, Japan
S. Senthilraja, SRM Institute of Science and Technology, India
Farhad Shahnia, Murdoch University, Australia
Hussain Shareef, United Arab Emirates University, UAE
Bhim Singh, Indian Institute of Technology Delhi, India
Vijay Sood, Ontario Tech University, Canada
Vivian Sultan, California State University, Los Angeles, USA
Hongbo Sun, Mitsubishi Electric Research Laboratories, USA
Masoud Taghavi, Technical and Vocational University (TVU), Faculty of Noshahr, Iran
Mehrdad Tahmasebi, Islamic Azad University - Ilam Branch, Iran
Jay Taneja, University of Massachusetts, Amherst, USA
Saeed Teimourzadeh, EPRA - Engineering Procurement Research Analysis, Ankara, Turkey
Philipp Thies, University of Exeter, UK
Mihai Tirsu, Institute of Power Engineering, Moldova
Tek Tjing Lie, Auckland University Of Technology, New Zealand
Santiago Torres Contreras, Universidad de Cuenca, Ecuador
Graham Town, Macquarie University, Australia
Quoc Tuan Tran, Paris Saclay University / CEA / INES, France
Navid Vafamand, Shiraz University, Iran
François Vallee, University of Mons, Belgium
Eric MSP Veith, Carl von Ossietzky University, Oldenburg, Germany
Alekhya Velagapudi, University of Pittsburgh's School of Computing and Information, USA
Alexander Wallis, University of Applied Sciences Landshut, Germany
Daniel Wilson, Boston University, USA
Jian Xu, Texas Reliability Entity (Texas RE), USA
Weiwei Yang, Microsoft Research, USA
Sean Yaw, Montana State University, USA
Roberto Yus, University of Maryland, Baltimore County, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Day-ahead Electricity Price Forecasting of Elspot Markets in Norway

1st Markus Wiik Jensen
School of Economics,
Innovation and Technology
Kristiania University College
Oslo, Norway
email:maje012@student.kristiania.no

2nd Huamin Ren
School of Economics,
Innovation and Technology
Kristiania University College
Oslo, Norway
email:huamin.ren@kristiania.no

3rd Andrii Shalaginov
School of Economics,
Innovation and Technology
Kristiania University College
Oslo, Norway
email:andrii.shalaginov@kristiania.no

Abstract—Forecasting day-ahead electricity prices from the Elspot market holds essential importance for various stakeholders, primarily electricity producers. These producers depend on precise price forecasts when placing supply bids and fine-tuning their dispatch schedules. This paper delves into this vital area, emphasizing day-ahead Electricity Price Forecasting (EPF). Following a comprehensive assessment of EPF techniques, we have experimented with three methods: a heuristic approach, Extreme Gradient Boosting (XGBoost), and the Long Short-Term Memory (LSTM) network. We have carried out unified comparisons among these three approaches across all six Elspot markets of Norway. Our results indicate that the LSTM outperform the other models in three of the six zones, which indicates the superior efficacy of the LSTM model. We have also noticed the impact of data variance on model performance, and hence improving model generalization will be our subsequent research endeavors.

Index Terms— *Electricity Price Forecasting, Elspot prices, XGBoost, LSTM.*

I. INTRODUCTION

Accurately forecasting market trends and price fluctuations is of paramount significance for a diverse range of stakeholders, including investors, businesses, and policymakers. This importance is particularly pronounced in the context of electricity markets, which serve an integral role in modern society and have experienced substantial transformation through deregulation and the integration of renewable energy sources [8] [12] [14] [29]. Recent disruptions in European electricity markets further underscore the growing imperative for precise Electricity Price Forecasting (EPF). Such predictions are crucial for electricity producers, consumers, and market operators to effectively plan their production, consumption and trading activities [1].

The NordPool spot (Elspot) market is a day-ahead market, where the price of power is determined by supply and demand. Such spot prices are actual price for electricity the next day, which will be set at Nordpool Elspot. Our primary focus is on day-ahead price forecasting using known spot prices. This forecasting directly informs bidding strategies for the upcoming day [17]. Due to the distinct characteristics of electricity markets, each forecasting challenge is unique across different markets and necessitates bespoke model developments [22]. We propose a framework for evaluating forecasting methods for all six Elspot markets of Norway while comparing three

different numerical approaches to the problem of extrapolating prices in both univariate and multivariate configurations, facilitating the identification of region-specific models and model configurations.

In essence, this paper addresses the pressing need for accurate EPF in an evolving energy landscape, providing a systematic framework for evaluating and comparing forecasting methods across multiple markets in Norway. In Section 2 we dive into electricity markets and existing literature on EPF, Section 3 present the methodologies employed, Section 4 discuss the conducted experiments and in Section 5 we conclude in an analysis of the obtained results.

II. BACKGROUND AND LITERATURE STUDIES

In this section we review the various market mechanics characterizing electricity markets and existing literature concerning EPF.

A. Background

Electricity is produced only moments before consumption, so unlike other commodities, electricity must be balanced between production and consumption at all times [15]. In a deregulated market environment, determining the unconstrained Market Clearing Price (MCP), commonly referred to as the spot price of an electricity pool typically involves the following steps:

Figure 1: MCP bidding process.

- Generating companies bid prices for supplying energy, creating a supply curve.
- The demand curve may be set at a value derived from a forecast of the load due to short-term inelasticity for demand of electricity, resulting in a vertical line at the forecasted load value.
- Spot price is found where supply and demand curves intersect, signifying the market equilibrium.
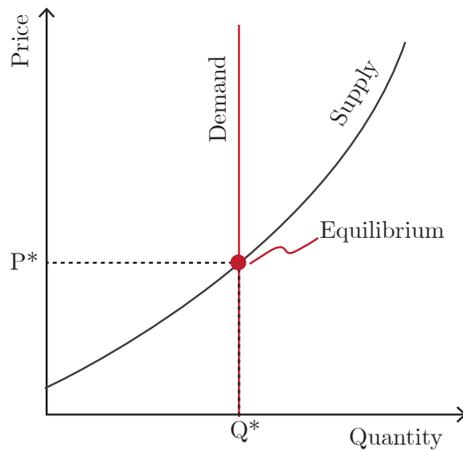
Figure 2: Equilibrium curve to determine the MCP of a bidding-pool.
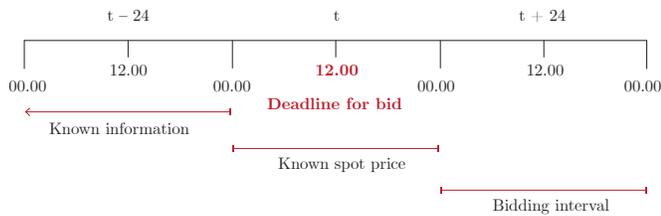
Figure 3: Deadline for bids in the Elspot markets.

The spot price is set at the equilibrium between supply and demand as seen in Figure 2 for each hour of the following day after accounting for the bids received within the deadline as illustrated in Figure 3 [12].

A time series is defined as a series of data points indexed in time order [32]. Commonly expressed as:

$$X = X_{t_{t=1}}^{\infty} = (X_1, X_2, \ldots) \tag{1}$$

where X_t denotes the observation at time t, and the sequence of observations is indexed by t ranging from 1 to infinity.

Accurately extrapolating the future poses unique challenges due to several constraints imposed by time order. Some of these constraints include look-ahead bias, stationarity, autocorrelation, seasonality, trend and noise. Time-series data, characterized by sequential observations over time, requires specialized methodologies that can capture temporal dependencies and patterns. Additionally, the electricity market is influenced by a multitude of factors, including supply and demand dynamics, changing industrial and household consumption, multiple seasonality, weather conditions, regulatory policies, fuel prices, the integration of renewable energy sources, and the rapid diffusion of price-anomalies [1] [8] [12] [14] [29]. Understanding the key drivers of price movements aids in feature selection for predictive models. For instance, if weather patterns or economic variables significantly affect prices, incorporating these into a model may improve accuracy.

The choice of methodology should also consider the nature of price drivers, as incorporating these considerations guides model selection. Furthermore, accurate price forecasts coupled with an understanding of their drivers provide valuable market insights.

B. Literature Studies

In the domain of EPF, selecting appropriate input variables, historical data duration, and modelling techniques is crucial. Most efforts that focus on forecasting day-ahead prices typically include an inference horizon of 1-4 weeks [2] [3] [10] [11] [13] [17] [18] [22] [23] [26] [30] [33]. Historical data spanning at least a year is commonly employed to capture yearly seasonality [2] [13] [18] [24] [33]. Input variables encompass a range of factors, including past prices [2] [3] [5] [6] [9] [10] [11] [13] [16] [17] [18] [21]- [27] [30] [33], system loads [13] [17] [21] [23]- [26] [30], weather variables [5] [13] [18] [24] [31], fuel costs [3] [5] [19] and sector indices [28]. Preprocessing and data transformations are essential to handle missing values and outliers, which can affect model performance. Techniques like normalization [5] [6] [30], decomposition [6] [10] [18] [23] [25] [33], and differentiation [11] are used to improve data quality and model accuracy. Statistical models, such as econometric methods, like Linear Regression [13] [21] [23] [31] and Auto-Regressive models [3] [10] [11] [13] [16] [18] [30] [33], offer interpretability and insights into correlations. Algorithmic models like Deep Learning (DL) [6] [13] [16] [17] [19] [21]- [25] and Ensemble models [3] capture complex and nonlinear patterns. Overall, the process of building a forecasting model involves decisions on input selection, preprocessing, model choice, parameter estimation, and accuracy evaluation. However, guidelines for navigating these complexities are limited, with much variation in reported approaches. Given the specific nature of EPF, establishing baselines and ensuring rigorous reporting is critical for advancing research in this field.

III. PROPOSED METHODOLOGY

In this research, the methodology centres around two key aspects of EPF: input variables and forecasting methods. The approach begins with selecting a baseline method that is heuristic-based. Building upon this baseline, the study conducts an empirical-driven progression to develop previously proven forecasting models in both univariate and multivariate configurations. Three distinct approaches are explored: a heuristic method, an algorithmic ensemble approach, and a DL approach. This methodology is designed to ensure objectivity and standardization in the evaluation process. Given the unique and inconsistent nature of electricity markets, EPF challenges vary significantly across locations and time frames, rendering cross-study evaluations potentially misleading and universal benchmarks logically unsound for this domain. Therefore the methodology involves systematic steps, including literature review of related work, data collection and preparation, model development and rigorous testing against real-world

outcomes. All the data-handling, -visualization and model-implementation and -evaluation was done using Python software.

A. Heuristic Baseline

The persistence forecast is utilized as a baseline for this study. This approach involves using the last observed value of the time series as the forecast for the corresponding day-ahead time step. In the context of day-ahead electricity price forecasting, this would mean using the most recent price value as its prediction for the same hour the next day. Assuming we have a time series of electricity prices $p_t, p_{t+1}, p_{t+2}, ..., p_{t+n}$ where t is the current time step, the persistence model predicts the current price 24 hours ahead for each time step. In the context of day-ahead EPF, the persistence model serves as a sensible baseline. While more complex modelling methods may exhibit reasonable accuracy, they must be able to generalize beyond the explicit information provided in the input data. As a baseline the heuristic provides a reference point against which more advanced models can be evaluated, ensuring that they genuinely contribute to improved forecasting performance. We can express the persistence model in mathematical notation as follows:

$$\hat{P}_t = P_{t-24} \qquad (2)$$

where $\hat{P}_t$ denotes the predicted electricity price at time t and P_{t-24} is the observed value of the electricity price 24-time steps earlier.

B. Algorithmic Ensemble

Extreme Gradient Boosting (XGBoost) is a popular gradient-boosting algorithm that is commonly used in machine-learning applications for both classification and regression tasks. It is an ensemble algorithm that combines multiple weak models (decision trees) to make a strong prediction. XGBoost learns from examples by building a series of decision trees. Each tree tries to correct the mistakes made by the previous trees reducing the risk of overfitting, and leading to a more accurate prediction [7]. The objective function for XGBoost can be written as:

$$\mathcal{L}(\Theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (3)$$

where Θ represents the set of model parameters, n is the number of training examples, y_i is the true value of the i-th example, $\hat{y}_i$ is the predicted value, $l(y_i, \hat{y}_i)$ is the loss function, K is the number of weak models, f_k represents the k-th weak model, and $\Omega(f_k)$ is the regularization term.

The weak models used in XGBoost are decision trees, which can be expressed as:

$$f(x) = \sum_{t=1}^{T} w_t q_t(x), \quad w \in \mathbb{R}^T, \quad q : \mathbb{R}^d \to \{1, 2, \dots, T\} \qquad (4)$$

where x is the input features, w is the vector of weights associated with each leaf node of the tree, T is the number

of leaf nodes, and $q(x)$ is the function that maps the input features to the index of the corresponding leaf node.

C. Deep Learning

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that is commonly used for time-series forecasting. Unlike traditional RNNs, LSTM networks are designed to overcome the problem of vanishing gradients, which make it difficult for the network to learn and remember long-term dependencies in the data. In simple terms, the LSTM network is like a specialized memory unit that can selectively remember important information from the past and use it to make predictions about the future. It achieves this by using a system of gates, which are like control switches, to control the flow of information within the network.
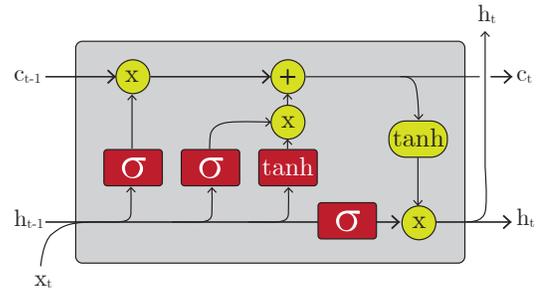
Figure 4: Long Short-Term Memory (LSTM) Network Diagram.

The LSTM network has three main types of gates as visualized in Figure 4: input gates, forget gates, and output gates. These gates allow the network to decide which information is important to keep, which information to forget, and when to output its predictions [20].

IV. EXPERIMENTS AND DISCUSSION

This section covers the datasets used, the experimental setup, and the ensuing presentation and discussion of results.

A. Dataset and Description

The data, including unit measures, granularity and data sources are described in Table I. A total of six data-sets were created, each comprising time series data from one of the six bidding zones. The data-sets consist of 14-16 variables each, with the amounts of variables varying depending on the number of exchange connections to neighbouring zones.

Missing values occurred due to multiple reasons, such as changing time zones, observations at a lower frequency than
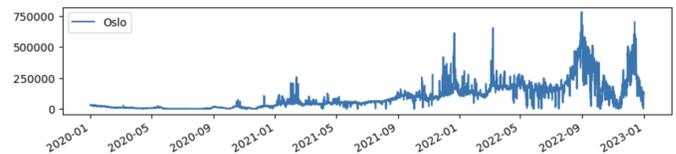
Figure 5: Historical Elspot prices for Oslo (NO1).

TABLE I: DESCRIPTION OF DATA (TARGET[*]).

Variable (units) [granularity]	Source
Elspot price (NOK/MWh) [h]	Nord Pool
Day-ahead Elspot price (NOK/MWh)[h][*]	Nord Pool
Power production (MWh) [h]	Nord Pool
Power production prognosis (MWh) [h]	Nord Pool
Power exchange (MWh) [h]	Nord Pool
Power consumption (MWh) [h]	Nord Pool
Reservoir levels (GWh) [w]	Nord Pool
Reservoir capacity (GWh) [w]	Nord Pool
Gas price (NOK/mmbtu) [d]	Yahoo-finance
Oil price (NOK/barrel) [d]	Yahoo-finance
OSEBX price (NOK/OSEBX) [d]	Yahoo-finance
Air temperature (mean/degC) [d]	MET
Wind speed (mean/ms) [d]	MET
Percipitation (sum/mm) [d]	MET

the target values and stock exchanges being closed during weekends. Missing values due to these occurrences were appropriately imputed using interpolation, backward-fill or forward-fill.

The data is split into two sections, the first contains three years of data with 26 000+ price-observations and is allocated for training and validation, the second is separated from the first and contains 4 months of recent and unseen data allocated for testing and evaluation. The date ranges are the following, 01.01.2020 00:00 - 29.12.2022 23:00 for train and validation, and 01.01.2023 00:00 - 30.03.2023 23:00 for the hold-out test set. Essentially, the train-test split contains the original time order and is not shuffled or re-ordered. Data is normalized using min-max scaling, this is done separately for the two sections in order to prevent introducing look-ahead-biases encoded in the scaling.

B. Experiments

The experiments include a heuristic baseline and are compared against each other as opposed to previous experiments from related work. First, the models are validated in the task of predicting the day-ahead hourly elspot prices on the validation set using a rolling forecast cross-validation (RFCV) scheme presented in Table II.

TABLE II: RFCV SCHEME (YYYY-MM-dd hh).

| Fold	Train Start	Train End	Val Start	Val End
1	2020-01-01 00	2021-12-31 23	2022-01-01 00	2022-01-01 23
2	2020-01-01 00	2022-01-01 23	2022-01-02 00	2022-01-02 23
3	2020-01-01 00	2022-01-02 23	2022-01-03 00	2022-01-03 23
...	...	...	...	...
365	2020-01-01 00	2022-12-28 23	2022-12-29 00	2022-12-29 23

These experiments provide information about the models' performance on a full year of daily-predictions with daily re-training. During validation, the error of the models is measured using Root Mean Squared Error (RMSE). The errors are averaged by time of day; mornings (hours 6-12), mid-days (hours 12-15), evenings (hours 15-21) and nights (hours 21-6). An example of results from rolling forecasts origin validation with visualization from a sample period of 1 week including bar charts of aggregated time-of-day scores from the entire year are presented in Figure 6 (baseline results of aggregated RMSE are marked with red dashed lines for comparisons).
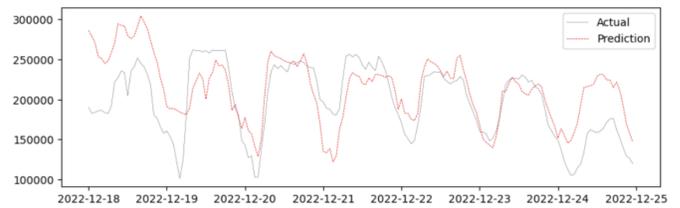
(a) Actual vs. prediction (18.12.2022 00:00 - 24.12.2022 23:00).
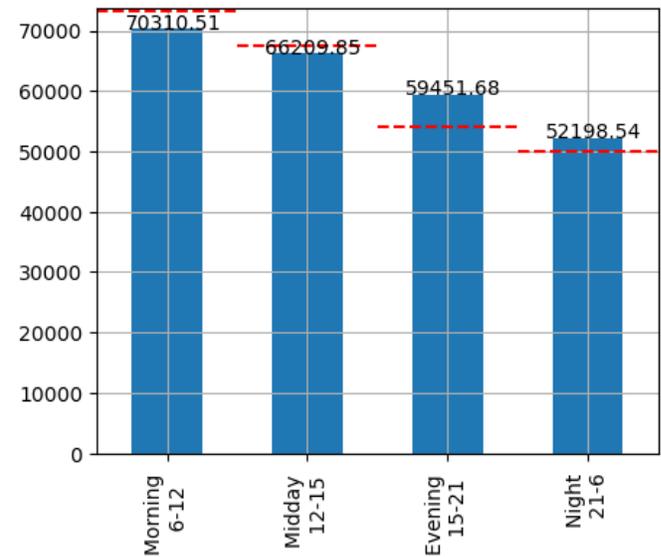
(b) Aggregated RMSE (01.01.2022 00:00 - 29.12.2022 23:00).

Figure 6: Rolling Forecast Origin Cross-validation of multivariate LSTM for Kristiansand (NO2).

After validating the models on the last year of the train-set, they are then are evaluated in their ability to extrapolate 24 time-steps ahead from the known spot-price during a 4-month out-of-sample period on a recent hold-out test-set from all the bidding-zones, with their weights and hyperparameters determined from training and tuning on the previous 3 years of data. The results of these experiments are presented in Table III, allowing for comprehensive analysis and review of the different modelling approaches in relation to the bidding zones and the addition of exogenous variables. The evaluation scheme of model performance consists of four different error terms; Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Average Percentage Error (MAPE) and Residual Sum of Squares (RSS). To gain a comprehensive understanding of the models' capacity for generalization and their ability to navigate the bias-variance trade-off, we seek to offer diverse viewpoints on the models' performance.

C. Discussion

Predictions from validation seem to be more accurate during mornings (6-12) and middays (12-15) as illustrated by the RMSE scores in Figure 6. However, none of the models consistently outperform the heuristic baseline across bidding

TABLE III: RESULTS FROM OUT-OF-SAMPLE EVALUATION (01.01.2023 00:00 - 30.03.2023 23:00).

		RMSE		MAE		MAPE		RSS	
	Model	endog	w/ exog	endog	w/ exog	endog	w/ exog	endog	w/ exog
NO1	Heuristic	29469	/	19946	/	**25.19%**	/	$18.7e^{11}$	/
	XGBoost	29156	27052	20268	**18838**	26.22%	26.92%	$17.9e^{11}$	**$15.5e^{11}$**
	LSTM	29174	**21109**	21035	20134	29.21%	26.69%	$17.9e^{11}$	$17.9e^{11}$
NO2	Heuristic	29474	/	19943	/	25.19%	/	$18.7e^{11}$	/
	XGBoost	29545	27259	20715	19266	26.77%	26.19%	$18.4e^{11}$	$15.7e^{11}$
	LSTM	28354	**26431**	20317	**18173**	29.09%	**24.81%**	$16.9e^{11}$	**$14.9e^{11}$**
NO3	Heuristic	30448	/	21069	/	37.58%	/	$20.0e^{11}$	/
	XGBoost	28469	29069	19687	19666	35.79%	31.79%	$17.1e^{11}$	$17.8e^{11}$
	LSTM	28438	**28381**	20462	**19228**	40.91%	**31.29%**	$17.1e^{11}$	**$16.6e^{11}$**
NO4	Heuristic	21456	/	11705	/	25.28%	/	$99.4e^{10}$	/
	XGBoost	20592	23149	11424	12507	25.43%	29.35%	$89.6e^{10}$	$11.3e^{11}$
	LSTM	**19448**	21675	**10519**	13155	**22.76%**	28.05%	**$79.9e^{10}$**	$96.1e^{10}$
NO5	Heuristic	25240	/	**16953**	/	**15.75%**	/	**$13.7e^{10}$**	/
	XGBoost	24950	**24156**	17137	17018	16.27%	15.94%	$13.1e^{11}$	$12.3e^{11}$
	LSTM	25427	24584	18391	18189	17.80%	17.49%	$13.6e^{11}$	$12.9e^{11}$
NO6	Heuristic	30448	/	21069	/	37.58%	/	$20.0e^{11}$	/
	XGBoost	28469	**28326**	19687	**19532**	35.79%	**31.70%**	$17.1e^{11}$	**$16.9e^{11}$**
	LSTM	28438	30100	20462	22870	40.91%	48.58%	$17.1e^{11}$	$19.1e^{11}$

zones and time-of-day during these experiments. This could be attributed to the highly volatile and disruptive prices witnessed in the year 2022 as illustrated in the historical elspot prices in Figure 5, which is the year allocated for validation, making it difficult for the models to fit the data comprehensively.

Results from the out-of-sample evaluation exhibit more promising improvements over the baseline. As seen in Table III, the LSTM and XGBoost models outperform the baseline across all evaluation criteria for most of the bidding-zones, meaning that they are able to balance between capturing price nuances while maintaining robustness to outliers. These results ultimately emphasize the potential of DL and ensemble ML techniques for capturing the complexities of EPF. Specifically, the LSTM model in its multivariate configuration achieves this feat during out-of-sample evaluation on the data-sets for bidding-zone NO2 and NO3. Surprisingly, the univariate LSTM outperforms the other models in all aspects of error for the bidding-zone NO4. The final model to outperform all other models for all aspects of error is the multivariate XGBoost model for the bidding-zone NO6. For the remaining bidding-zones NO1 and NO5 there is no clear contender for best model performance. The variability in results across the different data-sets highlights the presence of unique characteristics for the distinct bidding-zones, with varying predictability, model-performances and optimal model-configurations.

V. CONCLUSION AND FUTURE WORK

Forecasting day-ahead electricity prices plays a pivotal role in strategizing and balancing the supply and demand for the subsequent day, making it an essential area to delve into. In this paper, we introduce a framework to assess forecasting techniques across all six Elspot markets in Norway, intimi-dating heuristic methods with advanced XGBoost and LSTM deep learning networks. Results consistently showcase the superiority of LSTM model over its counterparts in out-of-sample evaluations across most bidding zones. Specifically,

the LSTM model in its multivariate configuration outperforms all other models for all aspects of error in bidding-zone NO2 and NO3. The univariate LSTM outperforms the other models in all aspects of error for the bidding-zone NO4. This is notable due to capability to capture price variance temporally, as well as a notable merits of robustness against outliers. Concurrently, the XGBoost model has also marked its presence by performing admirably in the bidding-zone NO6, also outperforming the other models for all aspects of error. This feat could be attributed to the simplicity of the XGBoost model as opposed to the more complex LSTM, making it less likely to overfit on data that exhibit less intricate and complex patterns. The variability in results across the different data-sets highlights the presence of unique characteristics for the distinct bidding-zones, therefore, model generalization will be the focal point of our future research endeavors.

REFERENCES

[1] N. Amjady and M. Hemmati, "Energy price forecasting - problems and proposals for such predictions", IEEE Power and Energy Magazine, vol. 4, pp. 20–29, 2006.

[2] R. Beigaite, T. Krilavicius and K. L. Man, "Electricity price forecasting for nord pool data", International Conference on Platform Technology and Service (PlatCon), pp. 1–6, 2018.

[3] K. Bitirgen and Basaran Filik, "Electricity price forecasting based on xgboost and arima algorithms", BSEU Journal of engineering research and technology, vol. 1, pp. 7–13, December 2020.

[4] L. Breiman, Classification and regression trees, Routledge, 2017.

[5] M. Castelli, A. Groznik, and A. Popovic, "Forecasting electricity prices: A machine learning approach", Algorithms, vol. 13, pp. 119, May 2020.

[6] Z. Chang, Y. Zhang and W. Chen, "Electricity price prediction based on hybrid model of adam optimized lstm neural network and wavelet transform", Energy, pp. 187, 2019.

[7] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, August 2016.

[8] Y.-k. Chen, A. Hexeberg, K. E. Rosendahl and T. F. Bolkesjø, "Long-term trends of nordic power market: A review", WIREs Energy and Environment, vol. 10, pp. 413, 2021.

[9] A. Ciarreta, M. P. Espinosa and C. Pizarro-Irizar, "Is green energy expensive? empirical evidence from the spanish electricity market", Energy Policy, vol. 69, pp. 205–215, 2014.

[10] A. Conejo, M. Plazas, R. Espinola and A. Molina, "Day-ahead electricity price forecasting using the wavelet transform and arima models", IEEE Transactions on Power Systems, vol. 20, pp. 1035–1042, 2005.

[11] J. Contreras, R. Espinola, F. Nogales and A. Conejo, "Arima models to predict next-day electricity prices", IEEE Transactions on Power Systems, vol. 18, pp. 1014–1020, September 2003.

[12] A. Creti and F. Fontini, Economics of electricity, Cambridge University Press, 2019.

[13] A. Cruz, A. Munoz, J. L. Zamora and R. Espınola, "The effect of wind generation and weekday on spanish electricity spot price forecasting", Electric Power Systems Research, vol. 81, pp. 1924–1935, 2011.

[14] C. Defeuilley, "Retail competition in electricity markets", Energy Policy, vol. 37, pp. 377–386, 2009.

[15] G. Erdmann, "Economics of electricity", EPJ Web of Conferences, vol. 98:06001, January 2015.

[16] G. Gao, K. Lo and F. Fan, "Comparison of arima and ann models used in electricity price forecasting for power market", Energy and Power Engineering, vol. 09, pp. 120–126, January 2017.

[17] P. S. Georgilakis, "Market clearing price forecasting in deregulated electricity markets using adaptively trained neural networks", SETN: Hellenic Conference on Artificial Intelligence, vol. 3955, pp. 56–66, 2006.

[18] L. Grossi and F. Nan, "Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources", Technological Forecasting and Social Change, vol. 141, pp. 305–318, 2019.

[19] J.-J. Guo and P. Luh, "Selecting input factors for clusters of gaussian radial basis function networks to improve market clearing price prediction", IEEE Transactions on Power Systems, vol. 18, pp. 665–672, June 2003.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural computation, vol .9, pp. 1735–1780, 1997.

[21] Z. Hu, L. Yang, Z. Wang, D. Gan, W. Sun and K. Wang, "A game-theoretic model for electricity markets with tight capacity constraints", International Journal of Electrical Power & Energy Systems, vol. 30, pp. 207–215, 2008.

[22] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark", Applied Energy, vol. 293, pp. 116983, 2021.

[23] C. Li and S. Wang, "Next-day power market clearing price forecasting using artificial fish-swarm based neural network", ISNN: Advances in Neural Networks, pp. 1290–1295, 2006.

[24] P. Mandal, T. Senjyu and T. Funabashi, "Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market", Energy Conversion and Management, vol. 47, pp. 2128–2142, 2006.

[25] M. S. Nazar, A. E. Fard, A. Heidari, M. Shafie-khah and J. P. Catalao, "Hybrid model using three-stage algorithm for simultaneous load and price forecasting", Electric Power Systems Research, vol. 165, pp. 214–228, 2018.

[26] E. Raviv, K. E. Bouwman and D. Van Dijk, "Forecasting day-ahead electricity prices: Utilizing hourly prices", Energy Economics, vol. 50, pp. 227–239, 2015.

[27] K. Skytte, "The regulating power market on the nordic power exchange nord pool: an econometric analysis", Energy Economics, vol. 21, pp. 295–308, 1999.

[28] B. A. Souhir, B. Heni and B. Lotfi, "Price risk and hedging strategies in nord pool electricity market evidence with sector indexes", Energy Economics, vol. 80, pp. 635–655, 2019.

[29] R. Weron, Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach, John Wiley Sons, 2006.

[30] R. Weron and A. Misiorek, "Forecasting spot electricity prices with time series models", Proceedings of the European Electricity Market EEM-05 Conference, pp. 133–141, May 2005.

[31] R. Weron and M. Zator, "Revisiting the relationship between spot and futures prices in the nord pool electricity market", Energy Economics, vol. 44, pp. 178–190, 2014.

[32] Wikipedia, "Time series," *Wikipedia, The Free Encyclopedia*, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Time_series&oldid=1193976963. [Accessed: 2024-02-09].

[33] Z. Yang, L. Ce, and L. Lian, "Electricity price forecasting by a hybrid model, combining wavelet transform, arma and kernel-based extreme learning machine methods", Applied Energy, vol. 190, pp. 291–305, 2017.

Analysis of California Fire-Perimeter Data Using Geographic Information Systems to Examine the Correlation Between Population Density and Acres Burned

Vivian Sultan

College of Business and Economics
California State University
Los Angeles, USA
vsultan3@calstatela.edu

Hind Bitar

Department of Information Systems
King Abdulaziz University
Jeddah, Saudi Arabia
hbitar@kau.edu.sa

Abstract—**The number of California wildfires has increased in the past two decades. This change has increased the authors' and policymakers' attention to the factors that affect this phenomenon and how to manage it. Wildfires wreak havoc on the environment by burning large sections of land, housing, animals, and people alike. Wildfires degrade air quality while hindering transportation and communication. They also present a serious threat to the power grid. This study aims to examine the correlation between population density and acres burned which may help understand and manage wildfires in the state. This research study uses California fire-perimeter data, population data, and fire-severity zones extracted from the ArcGIS hub and ScienceBase. In particular, we analyzed five years of fire-perimeter data using a geographic information system ordinary least squares analysis, attributes, and summary statistics to create new layers representing selected features involved in the process. The results show no correlation between the dependent and the explanatory variables. Further analysis suggests that wildfires may be reduced if more awareness campaigns are designed and presented to the public.**

Keywords-Fire perimeter; Wildfires; GIS; ArcGIS.

I. INTRODUCTION

California's hot climate and flammable plants give the state a high wildfire risk [1]. Millions of acres are burned yearly in California wildfires, rendering land unusable for agriculture while destroying habitats and property [2]. California's state government has massive historical data on wildfires, in cooperation with other western US states [2]. For example, in 1910, the historical event "Big Blowup" occurred and affected the Northwest states. As a result, fire-suppression policies were established [3]. In 1889, large parts of Orange County, California, were burned by a great wildfire; the Santiago Canyon Fire [2], estimated at 300,000 acres. More recently, the Matilija (1932, 220,000 acres) and Laguna (1970, 175,000 acres) fires were recorded as the largest and second largest fires in California's history until 2020 [4]. In 2003, a complex fire occurred in Southern California that destroyed 3,719 homes and killed 24 [5]. The 2018 Camp Fire in Butte County, California, damaged 18,804 structures and caused 88 mortalities [6]. In 2020, a collection of large fires broke the state's records [1], burning 4,304,379 acres, destroying 11,116 structures, and killing 33 [7].

Several reasons or factors play critical roles in wildfires' occurrence and severity. Drought increased the chances of the large fires California faced in the last decade. Related to this cause is another critical factor, climate change [2][8][9]. Another natural factor is lightning [10]. Another important factor is humans. As reported by the U.S. Forest Service research data archive, 85% of wildfires in the states are caused by human actions: discarded cigarettes, improperly tended campfires, intentional arson [10], population density [11], and other factors.

Wildfires are a serious threat to the power grid, as they can damage or destroy power lines, transformers, substations, and other infrastructure. Utilities also must sometimes shut off power to prevent sparks from igniting new fires, causing widespread blackouts affecting millions.

At least three smart-grid technologies can help enhance the power grid's wildfire resilience, reliability, and safety. Microgrids—small, localized grids that can operate independently from the main grid—can provide backup power to critical facilities: hospitals, fire stations, and water-treatment plants. Underground power lines, being unexposed to the air, are less vulnerable to damage from wind, trees, animals, and fire and reduce the visual impact and electromagnetic interference of power lines. Sensors and automation, small devices that monitor the condition and performance of the power grid, can automatically detect and isolate faults, such as downed power lines or broken equipment and can also communicate with each other and with the control center to optimize the operation and coordination of the grid. All of these technologies can help prevent or reduce the severity of power outages and restore power more quickly and efficiently after a disruption.

It is essential to analyze all aspects of these fires. Thus, many valuable research studies have been conducted on many aspects of wildfires in California including understanding fire trends [12] and analyzing data to help decision makers [13]. Some researchers have developed simulations to predict where and how fast fire will travel [14]. Researchers have studied previous fires to develop response plans [13], and ways to manage fires [15]. Some of their approaches involve using geographic information systems (GIS) to create simulations with different elements that can impact a fire's spread and severity [16][17]. Researchers have even investigated the root factors of some fires, such as the powerlines [10], to solve that problem and prevent future fires.

However, to our knowledge, this is the first study that examines the correlation between population density and acres burned using GIS. Thus, the main aim of this research is to analyze that correlation. We hypothesize a strong correlation between population density and acres burned. Correlation analysis is used to test the hypothesis through GIS to analyze several scenarios and testing the hypothesis by applying ordinary least squares (OLS). This research study answers only one question: What is the correlation between population density and acres burned?

The remainder of the paper is organized as follows. Section II presents the literature review focused on the recent studies conducted in California state. Section III describes the methodology, which includes a clear description of the data and the analytical techniques used. The remaining two sections illustrate the results, discussion, conclusion, limitations, and suggestions for future work.

II. LITERATURE REVIEW

Several research studies have used GIS science and tools to analyze, understand, manage, and develop solutions for many issues including natural disasters, including earthquakes [18] and flood risk [19]. Also, GIS has been used to analyze California wildfire data, as is the focus of the current research study, to develop useful solutions that may help decision makers and communities or develop a deeper understanding of the problem, such as knowing the associated factors [16][17].

For example, [5] answered three main questions focusing on evacuation orders during wildfires to enhance community safety: "Who is at risk?" "How long will it take to evacuate?" "How much time is available?" The authors used fire-spread modeling with GIS to answer these questions, to determine the trigger point, and to recommend evacuation if the fire is nearby a certain landmark (point) [5]. The authors' techniques were based on a buffer to determine the evacuation trigger, so they applied three steps: 1) modeling the fire spread, 2) generating a fire-spread network, and 3) originating a wildfire-evacuation-trigger buffer. They argued that there is no need to have any particular information about the fire's location and their techniques could be used in a long-term strategic or short-term operational plan [5]. Reference [13] enhanced the wildfire-trigger modeling that used a buffer by combining traffic- and fire-simulation models to set a trigger. The authors proposed a three-step method with spatiotemporal GIS. The framework helps evaluate the generated trigger. The results from the framework showed that the dynamic representation of the evacuation traffic during the wildfire is improved and linked to a better understanding of the decision making and evacuation time [13].

Reference [12] used data from the NASA RECOVER Historic Fires Database (HFD) and GIS to support decision making related to fire trends in the western United States. ArcGIS Pro helped analyze wildfires' spatiotemporal patterns, characterizing changes in fire size, severity, and frequency over time. The results showed that the mean size of a fire that occurred in 1950 was less than the mean size of fires occurring more recently in 2010 and 2019. Fire frequency showed a slight increase, and fire severity was stable [12].

Responding to the destructive Camp Fire wildfire in Butte County, California, [6] created pre- and postwildfire maps representing elementary evacuation data and mitigation plans. This study used GIS and machine learning techniques. To map the pre- and postwildfire conditions, the authors applied Landsat-8 and Sentinel-2 imagery. To classify the pre- and postwildfire map, the authors compared a hybrid model, a support vector machine (SVM) optimized by the imperialist competitive algorithm with the unoptimized SVM algorithm. The hybrid model produced better accuracy compared with the unoptimized SVM. A total of eight pre–postwildfire burn-area maps could be used to assess the area affected by the Camp Fire wildfire to develop a well-established mitigation plan in the future [6].

Reference [20] examined the statistical correlation between weather and wildfire in 2020 while mining the available online California climate and historical spatial data from 1992 until 2011 using GIS. The authors investigated the correlation between drought conditions and wildfire number per forest unit area in California and visualized the results using GIS computing technology. They found such a correlation where no correlation between the wildfire and wind existed [20].

In 2021, [21] conducted case-study research in Southern California to prioritize wildfire restoration by applying GIS-based ordered-weight averaging (OWA). They assessed the efficacy of OWA and GIS-based multicriteria decision analysis (MCDA) techniques in determining the wildfire areas that need restoration priority. The combination of the GIS–MCDA process and the OWA technique helps develop and compare several decision maps. These maps show the restoration with prioritization and different spatial distributions. The authors concluded their research by highlighting the power of the GIS-MCDA technique as a core tool in spatial decision making [21].

Another research study was conducted to investigate the association between California's extreme wildfire events that happened in the last three decades and the socioeconomic characteristics focusing on the census tracts and county levels [15]. The authors used two secondary data types, wildfire geospatial data and the sociodemographic characteristics collected from the Bureau of Census that include, for example, ethnicity and educational level. They employed GIS-based spatial analysis to create a map representing the wildfire geolocations for several geographic levels with the socioeconomic and demographic factors that affected the potential wildfire risk. The results showed that more-educated people and people who have higher to median income prefer to live in a community with low crime levels and fewer natural hazards. Also, census tracts with more Native American citizens are more exposed to wildfire compared to other census tracts [15].

III. METHODS

A. Data Selection and Acquisition

Three datasets were used in this current research study. The main dataset had information about fire perimeters in California [22]. These data came from the ArcGIS hub and
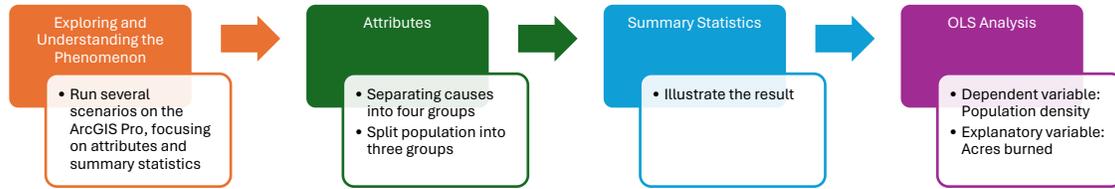
Figure 1. Used techniques and tools during the analysis phase.

were uploaded into ArcGIS Pro. The population data came from ScienceBase [23], representing the total number of people in each county. This dataset helped analyze the correlation using population information and perimeter size information for each fire. The final dataset contains California's fire-severity zones with three levels of hazard in the responsibility areas: moderate, high, and very high. The zones were developed by assigning a hazard score based on such factors as fire history, natural vegetation, terrain, blowing embers, predicted flame length, and typical fire weather in the area [22]. We used a total of 2,418 records in this study.

B. Analysis Phase

We used ArcGIS Pro, a full-featured professional desktop GIS application from the Environmental Systems Research Institute, Inc. (Esri), to run the analyses in a Windows environment. The OLS linear regression tool was used in the analysis after entering all data in one layer. However, before that, to explore and understand the phenomenon, several investigations and analyses were performed using ArcGIS Pro software focusing on attributes and summary statistics (see Figure 1).

To conduct the OLS analysis between population density and acres burned, we first split the population density into three groups, signified by color. Next, we ran a one-to-many spatial-join operation to combine the fire-perimeter and population-density layers to enable the extraction of the county where the fires took place. With all data in one layer, we conducted the analysis using OLS with the population as the dependent variable and the acres burned as the explanatory field. The attributes tool enabled the selection of some rows from the data that contained a certain number or a certain

string in a column. After selecting the required data, we created a new map layer. This tool helped separate the causes into four groups: 1) unknown/other causes, 2) human causes, 3) natural causes, and 4) industrial causes. The attributes tool also split the population density into 1) high, 2) medium, and 3) low. This splitting facilitated the analysis to understand the correlation as discussed in the results and discussion section.

After splitting the data causes and severity zones, the summary-statistics tool was used to illustrate the number of entries in the data, the sum of all the acres burned, and the average acres burned. The summary-statistics tool presented the data in a table which helps understand and make comparisons.

IV. RESULTS AND DISCUSSION

Interesting findings have been discovered after analyzing the data. First, the analysis of the population data and the fire-severity zones helped find zones with higher or lower populations. The results from the analysis show few patterns emerge between population and severity zones (see Figure 2): High and moderate fire-severity zones appear within medium- and low-population areas.

Another analysis was conducted to further investigate the proposed correlation: Only larger fire data were included, those that burned at least 5,000 acres, and compared with population. Figures 3 and 4 illustrate where most of the huge fires took place from 2016 to 2021, mostly in Northern California (see Figure 3). Figure 4 shows no patterns exist between the bigger fires and the population as the most huge fires occurred in medium- or even low-population zones.

Another analysis focused on the four main causes of fires and acres burned (see Figure 5). The most apparent causes on the map are the natural and unknown/other causes. To get a deeper understanding, we used the summary statistics tool to return the frequency of the fires per cause, the sum of all the acres burned for that cause, and the average in each fire with that cause (Table 1).

As illustrated in Table 1, the unknown/other cause is responsible for most fires. However, the unknown/other cause is just a third of the mean of acres burned. Also, interesting was that natural causes were first for the average acres burned, which means that the natural causes created fewer but more destructive fires.

Following the exact technique above, we broke the human-causes attribute into five unique associated types (Figure 6). As shown in Table 2, equipment use is a more frequent cause of fire with 298, but its mean acres is just 481.3095. Campfire burned the most acre (180,044.4) and the highest mean (3,830.73). The least cause is smoking where the frequency is 14 and the mean of acres is 78.8814.

Figure 2. Fire-severity zones with the population data.
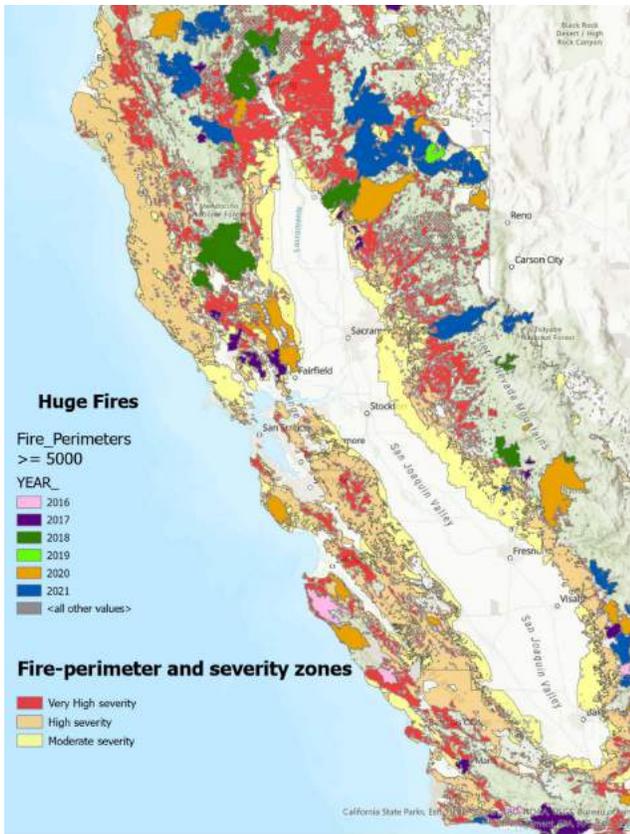
Figure 3.   Huge fire locations.
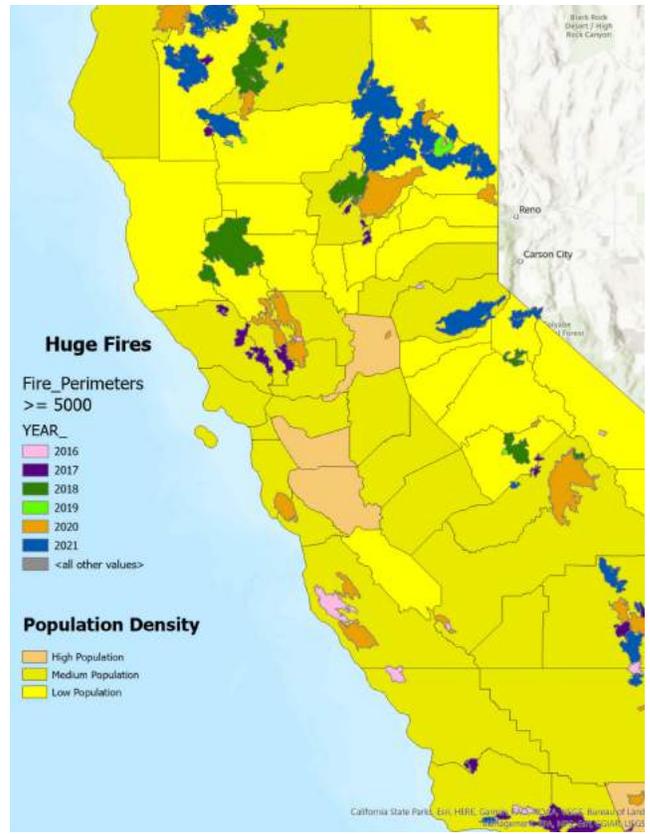
Figure 4.   Huge fire with the population.

Following the exact technique above, we broke the human-causes attribute into five unique associated types (Figure 6). As shown in Table 2, equipment use is a more frequent cause of fire with 298, but its mean acres is just 481.3095. Campfire burned the most acre (180,044.4) and the highest mean (3,830.73). The least cause is smoking where the frequency is 14 and the mean of acres is 78.8814.

Two subcauses of the natural causes were debris and lighting. We created a new layer for each cause based on the selection in ArcGIS Pro. The separate layers help visualize fires based on their specific causes. Lighting is the most common natural cause with also higher means of acres compared to debris (Figure 7 and Table 3).

Using the same techniques for the industrial revealed five unique cause types (Figure 8 and Table 4). The most frequent cause is vehicle while the higher mean of acres is powerlines. The least frequent are both railroad and aircraft causes while the acres mean for the railroad is higher than aircraft.

TABLE I.   CAUSES AND BURNING ACRES

Cause	Frequency	Sum of acres	Mean of acres
Unknown/other	1,189	3,023,862.61	2,543.20
Human	525	464,800.54	885.33
Natural	495	2,845,918.36	5,749.33
Industrial	355	1,609,638.03	4,534.19
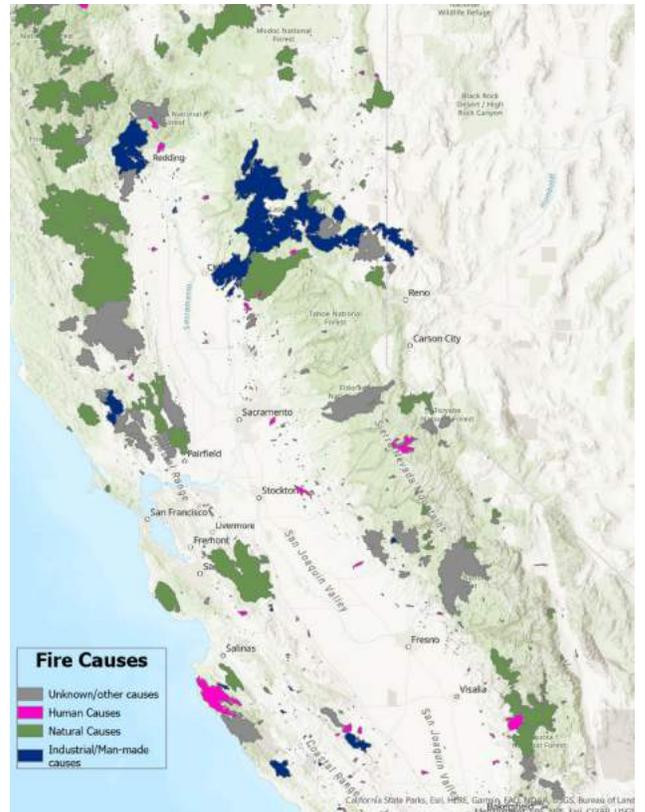
Figure 5.   Causes and acres burned.

Figure 6.   Human causes.
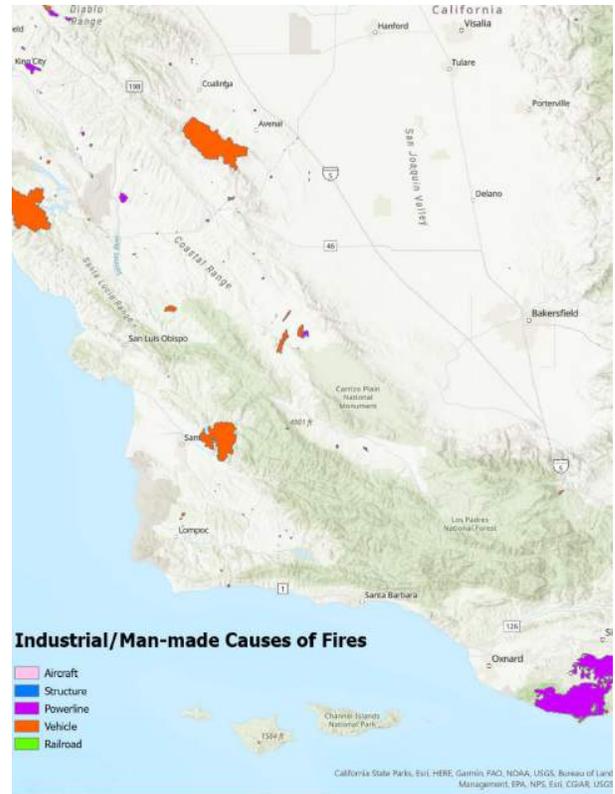
Figure 8.   Industrial causes.

TABLE II.   HUMAN CAUSES

Cause	Frequency	Sum of acres	Mean of acres
Equipment Use	298	143,430.2	481.3095
Smoking	14	1,104.34	78.8814
Campfire	47	180,044.4	3,830.73
Arson	125	135,449.9	1,083.6
Playing with fire	26	2,325.25	89.4327
Escaped prescribed burn	15	2,446.46	163.0973

TABLE III.   NATURAL CAUSE TYPES

Cause	Frequency	Sum of acres	Mean of acres
Lightning	393	4,620,219.4	11,756.28
Debris	102	62,338	611.157

TABLE IV.   INDUSTRIAL CAUSE TYPES

Cause	Frequency	Sum of acres	Mean of acres
Railroad	3	427	142.3333
Vehicle	198	368,135.4	1,859.27
Powerline	144	1,240,273.5	8,613.01
Structure	7	502.15	71.7357
Aircraft	3	300	100

Figure 7.   Natural causes.

Industrial/Man-made Causes of Fires

- Aircraft
- Structure
- Powerline
- Vehicle
- Railroad
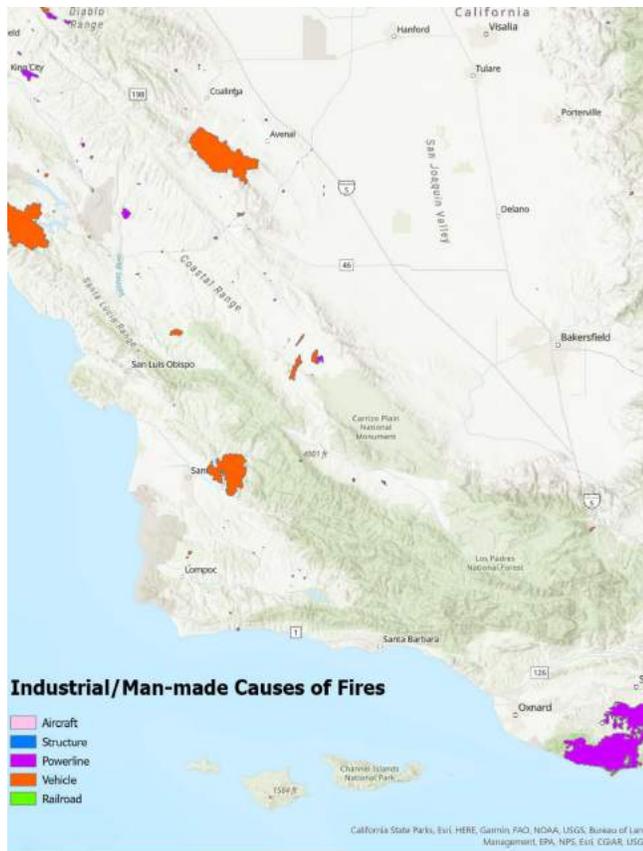
Figure 9.   Fire-perimeter and severity zones.

TABLE V.   FIRE-PERIMETER AND SEVERITY ZONES

Severity	Frequency	Sum of acres	Mean of acres
Moderate	633	149,997.29	236.96
High	608	207,325.72	340.99
Very High	1,323	7,586,896.53	5,734.62

Another investigation joined the fire-perimeter and severity zones (Figure 9). After the joining, the summary-statistics tool was used to learn how frequently the fires occurred in the area, the sum of acres burned in the severity zone, and the average acres burned per zone (Table 5).

The results showed that the very-high-severity zone accounted for over half of the fires that had occurred with a mean of acres of 5,734.62. It also had a very high sum of acres burned compared to the moderate- and high-severity zones.

To test the hypothesis, correlation analysis through OLS was conducted between population density and the number of acres burned. The results showed that the multiple R^2, which reflects the model performance, is 0.000009, a significantly low value. The adjusted R^2, which reflects the model significance, is −0.000405. This means that no correlation exists between the population density and the number of acres burned, rejecting the research hypothesis.

Multiple studies have been conducted related to wildfires in California [4][21][24]. However, most of these studies used different techniques and methods and had different aims

compared to this study. For example, in 2022, [25] performed a study to mainly evaluate the grazing effect on burn probability in California. The authors combined fire time series from 2001 through 2017 with environmental and socioeconomic covariate and grazing data. To analyze the data, they applied preregression matching and mixed-effects regression. The results show that a decrease in annual burn properties is linked to livestock grazing [25].

Since the risk associated with wildfires caused by smoke is a major concern across the United States including the Wildland–Urban Interfaces (WUI), [26] focused on the fire-danger trends over time. The authors used ArcGIS to perform their study including data from 1990 to 2010 [26] and concluded that the fire danger has increased over time during the peak season in the United States. This growth affects the WUI area as well as the people who live there. The authors also examined the relationship between fire danger and population density, finding that fire danger is increased in all medium and high densities whereas a decrease in fire danger occurred in the lowest population density [26]. This is in contrast to one of our analysis results focused on the population data and fire-severity zones. In this current study, the results indicate no pattern or relation between these two variables. One reason could be that in [26], the authors included all populations from the WUI area and others, which may affect their results. Also, different data in their study and this current study may also contribute to different results.

There is a need to develop proactive measures to prevent wildfires in California and elsewhere, which have increased in the last two decades. Reference [27] conducted a detailed analysis across California of the spatiotemporal distribution of the larger wildfires, concentrating on the human causes and others [27] using CAL FIRE data for the past two decades (2000–2019). The study showed that even though the total burned area increased, the mean burned area was still stable. Most of the wildfires were caused by humans. Natural factors were also common causes of wildfires in California. Far away from the human community, climate, and vegetation cover were the most important factors, especially the areas with heavy grass coverage, high temperature, and high vapor-pressure deficit [27]. The results are aligned with the findings in this study: Human, industrial, and natural are major causes. Hopefully, that can be managed by increasing public awareness campaigns.

V.   CONCLUSION, LIMITATIONS, AND FUTURE WORK

The increased number of wildfires in California and the consequences of these fires have caused several researchers to study the phenomena for better understanding or to find useful solutions. In this research, the main aim is to examine the correlation between population density and the number of acres burned in an area. We conducted analysis in ArcGIS Pro and with OLS spatial Statistics tool. Fire-perimeter cause data were further analyzed in multilabel scenarios at different layers. The main findings are that there is no correlation between population density and acres burned ($R^2 = 0.000009$) where more wildfires are caused by humans than by nature. However, population density does not affect fire severity. More awareness campaigns must be conducted at the state

level and might help reduce the number of acres burned. Thus, policy and decision makers must focus more on that activity.

One of the research limitations is the data used for the analysis since some causes were classified as unknown/other. If the causes were known, the results may change. Thus, as a future direction, interviews may be conducted to classify those causes, or other data may be used. Another future direction is to change the analysis techniques by, for example, using the Kernel density estimation (KDE) Spatial Analyst tool within ArcPro and/or other techniques to create a map of statistically significant hot spots for wildfires.

REFERENCES

[1] A. Wigglesworth, "Climate Change Boosts Risk of Explosive Wildfire Growth in California by 25%, Study Says," Los Angeles Times, September 04, 2023. [Online]. Available from: https://www.latimes.com/california/story/2023-09-04/climate-change-boosts-california-wildfire-risk-by-25

[2] J. E. Keeley and A. D. Syphard, "Large California Wildfires: 2020 Fires in Historical Context," Fire Ecology, vol. 17, no. 1, art. 22, 2021, doi:10.1186/s42408-021-00110-7.

[3] H. F. Diaz and T. W. Swetnam, "The Wildfires of 1910: Climatology of an Extreme Early Twentieth-Century Event and Comparison with More Recent Extremes," Bulletin of the American Meteorological Society, vol. 94, no. 9, pp. 1361–1370, 2013, doi10.1175/BAMS-D-12-00150.1.

[4] J. E. Keeley and P. H. Zedler, "Large, High-Intensity Fire Events in Southern California Shrublands: Debunking the Fine-Grain Age Patch Model," Ecological Applications, vol. 19, no. 1, pp. 69–94, 2009, doi:10.1890/08-0281.1.

[5] T. J.Cova, P. E. Dennison, T. H. Kim, and M. A. Moritz, "Setting Wildfire Evacuation Trigger Points Using Fire Spread Modeling and GIS," Transactions in GIS, vol. 9, no. 4, pp. 603–617, 2005, doi: 10.1111/j.1467-9671.2005.00237.x.

[6] M. Syifa, M. Panahi, and C.-W. Lee, "Mapping of Post-Wildfire Burned Area Using a Hybrid Algorithm and Satellite Data: The Case of the Camp Fire Wildfire in California, USA," Remote Sensing, vol. 12, no. 4, art. 623, 2020, doi:10.3390/rs12040623.

[7] CalFire, "2020 Incident Archive." [Online]. Available from https://www.fire.ca.gov/incidents/2020

[8] A. P. Williams et al., "Observed Impacts of Anthropogenic Climate Change on Wildfire in California," Earth's Future, vol. 7, no. 8, pp. 892–910, 2019, doi:10.1029/2019EF001210.

[9] K. Sweeney, R. Dittrich, S. Moffat, C. Power, and J. D. Kline, "Estimating the Economic Value of Carbon Losses from Wildfires Using Publicly Available Data Sources: Eagle Creek Fire, Oregon 2017," Fire Ecology, vol. 19, no. 1, art. 55, doi:10.1186/s42408-023-00206-2.

[10] National Park Service, "Wildfire Causes and Evaluations." [Online]. Available from: https://www.nps.gov/articles/wildfire-causes-and-evaluation.htm.

[11] F. Huot et al., "Next Day Wildfire Spread: A Machine Learning Dataset to Predict Wildfire Spreading from Remote-Sensing Data," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022, doi:10.1109/TGRS.2022.3192974.

[12] K. T. Weber and R. Yadav, "Spatiotemporal Trends in Wildfires across the Western United States (1950–2019)," Remote Sensing, vol. 12, no. 18, art. 2959, 2020, doi:10.3390/rs12182959.

[13] D. Li, T. J. Cova, and P. E. Dennison, "Setting Wildfire Evacuation Triggers by Coupling Fire and Traffic Simulation Models: A Spatiotemporal GIS Approach," Fire Technology, vol. 55, no. 2, pp. 617–642, 2019, doi:10.1007/s10694-018-0771-6.

[14] R. R. Linn et al., "QUIC-fire: A Fast-Running Simulation Tool for Prescribed Fire Planning," Environmental Modelling & Software, vol. 125, art. 104616, 2020, doi:10.1016/j.envsoft.2019.104616.

[15] S. N. Hwang and K. Meier, "Associations between Wildfire Risk and Socio-Economic-Demographic Characteristics Using GIS Technology," Journal of Geographic Information System, vol. 14, no. 05, pp. 365–388, 2022, doi:10.4236/jgis.2022.145020.

[16] M. C. Iban and A. Sekertekin, "Machine Learning Based Wildfire Susceptibility Mapping Using Remotely Sensed Fire Data and GIS: A Case Study of Adana and Mersin Provinces, Turkey," Ecological Informatics, vol. 69, art. 101647, 2022, doi:10.1016/j.ecoinf.2022.101647.

[17] V. Nasiri et al., "Modeling Wildfire Risk in Western Iran Based on the Integration of AHP and GIS," Environmental Monitoring and Assessment, vol. 194, no. 9, art. 644, 2022, doi:10.1007/s10661-022-10318-y.

[18] D. Singh, D. N. Pandey, and U. Mina, "Earthquake—A Natural Disaster, Prediction, Mitigation, Laws and Government Policies, Impact on Biogeochemistry of Earth Crust, Role of Remote Sensing and GIS in Management in India—An Overview," Journal of Geosciences and Geomatics, vol. 7, no. 2, pp. 88–96, 2019, doi:10.12691/jgg-7-2-5.

[19] Z. Zhu and Y. Zhang, "Flood Disaster Risk Assessment Based on Random Forest Algorithm," Neural Computing and Applications, vol. 34, no. 5, pp. 3443–3455, 2022, doi10.1007/s00521-021-05757-6.

[20] W. Xie, M. He, and B. Tang, "Data-Enabled Correlation Analysis between Wildfire and Climate using GIS," The Third International Conference on Information and Computer Technologies, IEEE, 2020, pp. 31–35, doi:10.1109/ICICT50521.2020.00013.

[21] T. Noth and C. Rinner, "Prioritization in Wildfire Restoration Using GIS-Based Ordered Weighted Averaging (OWA): A Case Study in Southern California," AIMS Environmental Science, vol. 8, no. 5, pp. 481–497, 2021, doi:10.3934/environsci.2021031.

[22] California Department of Forestry and Fire Protection, "ArcGIS Hub," Dec. 19, 2019. [Online]. Available from: https://hub.arcgis.com/datasets/CALFIRE-Forestry::california-fire-perimeters-all-1/explore?location=37.332052%2C-118.992700%2C7.00&showTable=true.

[23] C. Price and R. Clawges, "Digital Data Sets Describing Population Density in the Conterminous US," ScienceBase. [Online]. Available from: https://www.sciencebase.gov/catalog/item/631405b9d34e36012efa3002.

[24] A. Rosenthal, E. Stover, and R. J. Haar, "Health and Social Impacts of California Wildfires and the Deficiencies in Current Recovery Resources: An Exploratory Qualitative Study of Systems-Level Issues," PLoS ONE, vol. 16, no. 3, art. e0248617, 2021, doi:10.1371/journal.pone.0248617.

[25] K. J. Siegel et al., "Impacts of Livestock Grazing on the Probability of Burning in Wildfires Vary by Region and Vegetation Type in California," Journal of Environmental Management, vol. 322, art. 116092, 2022, doi:10.1016/j.jenvman.2022.116092.

[26] G. C. L. Peterson, S. E. Prince, and A. G. Rappold, "Trends in Fire Danger and Population Exposure along the Wildland–Urban Interface," Environmental Science & Technology, vol. 55, no. 23, pp. 16257–16265, 2021, doi:10.1021/acs.est.1c03835.

[27] S. Li and T. Banerjee, "Spatial and Temporal Pattern of Wildfires in California from 2000 to 2019," Scientific Reports, vol. 11, no. 1, art. 8779, 2021, doi:10.1038/s41598-021-88131-9.

On Sound Experiment Execution
with Learning Agents in Cyber-Physical Energy Systems

Eric MSP Veith[1,2]   Stephan Balduin[2]   Arlena Wellßow[1]   Torben Logemann[1]

[1]Carl von Ossietzky University Oldenburg
Research Group Adversarial Resilience Learning
Oldenburg, Germany
Email: `firstname.lastname@uol.de`

[2]OFFIS – Institute for Information Technology
Research Group Power Systems Intelligence
Oldenburg, Germany
Email: `firstname.lastname@offis.de`

Abstract—Autonomous and learning systems, such as Multi-Agent Systems or learning agents based on Deep Reinforcment Learning, has firmly established themselves as a foundation for approaches to create resilient and efficient Cyber-Physical Energy Systems. A substantial amount of research into different aspects of these systems is backed by simulation. However, the presentation of the simulation setup, experiment design, and experiment results evaluation often lacks crucial information, making it hard to reproduce or compare to other researcher's results. In this paper, we present the experiment design tooling of *arsenAI*, a part of the *palaestrAI* software ecosystem. We describe the work in progress on the experiment definition and mechanisms in place to aid in sound and reproducible experimentation with learning agents in co-simulated Cyber-Physical Energy Systems. We provide a document schema, name necessary tools, and also describe the relevant building blocks that enable reproducible experimentation with learning agent systems in co-simulated Cyber-Physical Energy Systems.

Keywords—*agent systems; learning agents; reinforcement learning; complex co-simulation; cyber-physical systems; modelling and simulation*

I. Introduction and Related Work

Agent systems are well-known for many aspects of research into Cyber-Physical Energy Systems (CPESs), and learning agents—e. g., such based on Deep Reinforcment Learning (DRL)—have established a firm foothold in the domain as well. From the hallmark paper that introduced DRL [1] to the development of MuZero [2] and AlphaStar [3], learning agents research has inspired many applications in the energy domain, including real power management [4], reactive power management and voltage control [5], [6], black start [7], anomaly detection [8], or analysis of potential attack vectors [9], [10].

It is in the very nature of especially learning agent systems that the most common way to demonstrate advances in the domain is through experimentation. Published papers present key performance indicators from the domain, such as a plot of voltage over time. To show that an agent learns, a plot of the reward function or utility function over time (or aggregated over training episodes) is depicted, alongside tables with cumulative values.

However, in many cases, crucial features are missing that would allow to reproduce these experiments. Basic values such as the initial seed for random number generators or the software packages, together with their version numbers, are usually not specified. In order to be meaningful, experiments would be constructed starting from a hypothesis, with invariants to validate or refute it. A Design of Experiments (DoE) [11] approach would allow to specify parameters and factors; the latter one to be varied in order to gauge the influence of particular input variables on the presented agent's or algorithm's general performance. In the context of CPESs, there is usually a number of such factors that can be considered, such as weather data, time of the year, node placement, or inverter capabilities.

In general, the goal for sound experimentation would be both, reproducibility, and a concise way to specify an experiment. CPES simulation often happens as co-simulation, using frameworks such as *mosaik* [12] or *CPSWT/C2WT-TE* [13]. Of those, some allow scenario-based modelling [14], but usually, the co-simulation itself does not consider the experiment stage, only the simulation execution stage, which follows a separation-of-concerns idea. As such, the concern for sound experimentation falls to other software packages. Surprisingly, even though there are libraries that provide the basic facilities for DoE, there is no tool suite available that would ensure sound experimentation with learning agents in such a co-simulation.

In this paper, we present the work-in-progress state of the *palaestrAI* software suite. This suite comprises a number of packages whose goal is to enable researchers to conduct reproducible and reliable experiments with (learning) agents in complex, co-simulated Cyber-Physical Systems (CPSs). The software suite consists of four major parts: *arsenAI*, *palaestrAI* itself, *hARL*, and *palaestrAI Environments*. In this stack, arsenAI is responsible to read experiment definitions, evaluate DoE statements, and create the concrete experiment run definitions. These experiment run definitions are instantiations of an experiment design, i. e., all factors are set to concrete values. *palaestrAI* itself takes care of proper execution. The *hARL* package offers implementation of learning agents, e. g., DRL algorithms such as Proximal Policy Gradient (PPO) or Soft Actor Critic (SAC), or the Adversarial Resilience Learning (ARL) agent reference implementation. *palaestrAI Environments* finally provides a unified interface to co-simulated environments, e. g., the MIDAS power grid reference scenario [15] through mosaik [12]. The complete software stack, organized in terms of a typical experimentation workflow, is depicted in Figure 1.
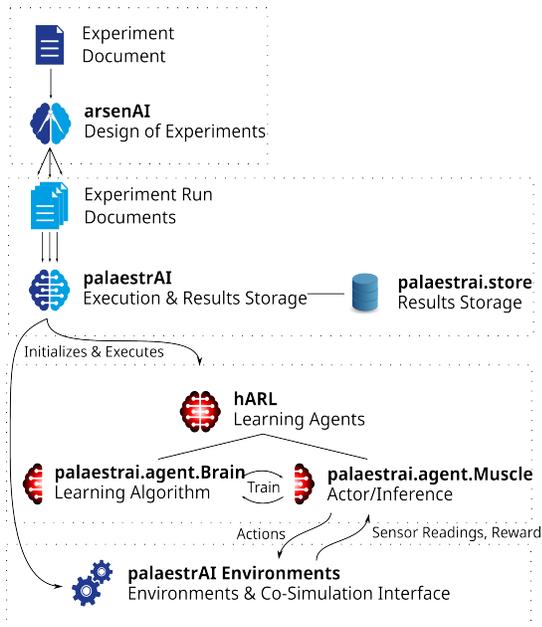
Figure 1.   The palaestrAI software stack [16]

The remainder of this paper is structured as follows: In Section II, we present the experiment definition document proper and describe its schema and design goals. Section III describes all features that aid in the reproducibility of experiments conducted with palaestrAI/arsenAI. Section IV presents methods built into the software stack to evaluate results from conducted experiments. We conclude in Section V with an outlook on our current work.

II. EXPERIMENT DEFINITION DOCUMENT

An experiment definition document serves as an intermediary data format, which needs to be both human-readable/-writable as well as machine-readable. Well-established formats at this interface are eXtensible Markup Language (XML), JavaScript Object Notation (JSON), and YAML Ain't A Markup Language (YAML). Since the barrier for humans should be as low as possible, YAML was chosen for arsenAI. YAML can be structurally validated using *YAML Schema*, which ensures the syntactic correctness of the document.

Experiment design for co-simulation setups is the major premiss for this document. This means that it specifies software modules, such as particular agent implementations or environments, which are used for the co-simulation. Co-simulation setups are not monolithic and no assumptions can be made with regards to the simulators that will be used. Thus, the experiment document itself cannot specify behaviors; it can only specify software modules and parameters to it. In this sense, parameters are what defines their behavior and not a runtime configuration, such as logging. This also means that the experimenter has to trust the software to behave correctly.

Scripted "events" can be part of an experiment design (sometimes called *incursions*). Those are then software modules, too (albeit provided by the experimenter), which use the general co-simulation Application Programming Interface (API) and are

written with the particular data exchange of concrete simulators in mind.

An important part of experiment design is the ability to reason about interactions between different parts of the setup and gauging the influence of certain factors in order to validate a hypothesis. For example, an experiment would try to answer the question "how does the choice of a particular agent training algorithm influence the observed performance of a variable over time," and explicitly not "run a simulation with agent *A* in environment *E* with configuration *C* and observe state variable *X*." As such, an experiment introduces a meta-layer. The DoE paradigm introduces *factors*, whose influence should be observed, and *parameters*, which represent values that are fixed for an experiment. The variation of factors and gauging of their influence is the core idea of DoE.

A concrete instantiation of all factors in an experiment definition creates an *experiment run definition*, which can then be picked up by a software and executed. Each experiment consists of at least one *phase*: A phase designates a particular stage of execution within an experiment, with concrete agents, each having a concrete set of sensors and actuators assigned, acting in co-simulated environments. Each phase has its own *termination condition*, each of which is an invariant against which the current state of agents and environments is checked during runtime to determine whether the phase has ended. Termination conditions can be based on the state of an environment (e.g., a blackout can be a termination condition for a power grid environment) or the state of an agent (e.g., its performance metric can hit a certain threshold). Phases can be repeated several times (called an *episode*), which is useful for learning agents. For example, an untrained agent may trigger an environment termination condition several times before developing a sensible strategy; thus, automatic restarts are necessary.

Subsequent phases can switch parts of the configuration, such as the agents that participate in it, their hyperparameters, termination conditions, and so an. A typical use case for several phases would be a training phase followed by a subsequent testing phase.

Here, the DoE approach allows to experimentally investigate hypothesis that are typical for learning agents in CPESs. For example, in order to evaluate whether an agent would learn a robust voltage control scheme, it would have to be tested in different power grid layouts, in different usage scenarios, even under adverse conditions. Also, learning agents need to be compared in terms of the algorithms they use and their respective hyperparameters. The search grid of combinations can be automatically generated through the DoE services that software packages such as arsenAI offer; even autocurriculum setups with different agents training each other or competition-style comparison setups can easily be realized with way.

An experiment document first independently provides a number of definitions: *environments*, *agents*, named sets of *sensors* and *actuators*, *simulation execution strategies* including termination conditions, and *phase* configurations (e.g., training or testing mode, or the number of episodes). An example for

these definitions can be found in Figure 2. Afterwards, concrete phases are listed in terms of a phase schedule. Each phase lists the desired factors under investigation and, thus, designs the search grid. Figure 3 is an example of such a schedule (and a continuation of Figure 2).

Finally, each experiment document requires a user-defined name. In addition, it allows to specify a random seed for reproducibility.

A design goal for the experiment document format described here is the ease-of-use for humans. Specifically, verbosity should be reduced in order to increase conciseness and, therefore, the expressiveness of an experiment document. The schedule configuration is the target for this goal, as it usually contains a number of repetitions. For example, the agents participating a specific phase may change, but the configuration of the environment might stay the same. In this case, simple repeating the environment configuration will make it difficult to discern actual changes from definitions copied over from the previous phase. Here, a cascading definition—similar to Cascading Style Sheets (CSS)—is employed: Any definition (agents, environments, or phase configuration) that is not explicitly given is considered to be equal to the one from the previous phase. Continuing with the example given, an environment has then to be defined only once (in the first phase it is used), and omitting it afterwards implicitly means that it is re-used as-is in any following phase that does not redefine the *environments* definition part of the schedule. Figure 4 lists the algorithm that is used to implement this cascading configuration scheme.

The example shown in Figure 2 defines the major components of an autocurriculum-type training and test of agents. Here, four agents should be trained, two power grid operator agents and two adversary ("attacker") agents. For easier reference, they are dubbed "Gandalf" and "Sauron" respectively. Of the two pairs, one instance should be trained in autocurriculum fashion (i. e., yielding one "Gandalf vs. Sauron" training), and two without adversary. The hypothesis that should be verified with this setup is: Autocurriculum-trained agents perform better (at the task of voltage control) than those that train alone. The task (i. e., voltage control through reactive power) is learned; the agent's utility functions are given as their *objective*. Agents are comprised of a learning module (nicknamed *Brain* in palaestrAI parley) and rollout workers (called *Muscles*). Each con receive their own set of parameters. In Figure 2, only a configuration for SAC is shown due to space constraints.

Figure 3 then defines the training and subsequent testing phases.

III. REPRODUCIBILITY

Reproducibility is an important aspect of simulation: It allows other researchers to reproduce the same experiment, arrive at the same results data and, therefore, verify the conclusions that were drawn from it. Especially Artificial Intelligence (AI) researchers have been under scrutiny because of reproducibility issues of many publications in the past [17]. Ideally, one would only need to distribute an experiment

```
uid: Classic ARL
seed: 2022
version: 3.5.0
output: palaestrai-runfiles
repetitions: 1
max_runs: 300
definitions:
  environments:
    midasmv_tar_ms:
      environment:
        name: MosaikEnvironment
        uid: midas_powergrid
        params: {}
      reward:
        name: ExtendedGridHealthReward
  agents:
    gandalf_sac_single:
      name: Gandalf SAC (single-agent-training)
      brain: &sac_brain
        name: harl:SACBrain
        params:
          fc_dims: [48, 48]
          update_after: 1000
          batch_size: 500
          update_every: 200
      muscle: &sac_muscle
        name: harl:SACMuscle
        params: {}
      objective: &defender_objective
        name: ArlDefenderObjective
        params: {}
    gandalf_sac_ac:
      name: Gandalf SAC (autocurriculum-training)
      brain: *sac_brain
      muscle: *sac_muscle
      objective: *defender_objective
    sauron_sac_single:
      name: Sauron SAC (single-agent-training)
      brain: *sac_brain
      muscle: *sac_muscle
      objective: &attacker_objective
        name: ArlAttackerObjective
        params: {}
    sauron_sac_ac:
      name: Sauron SAC (autocurriculum-training)
      brain: *sac_brain
      muscle: *sac_muscle
      objective: *attacker_objective
  sensors:
    all_sensors:
      midas_powergrid: [s1, s2]
  actuators:
    attacker_actuators:
      midas_powergrid: [a1, a2]
    defender_actuators:
      midas_powergrid: [a3, a4]
  simulation:
    vanilla_sim:
      name: TakingTurns
      conditions:
        - name: EnvTerminates
          params: {}
  phase_config:
    train: {mode: train, worker: 1, episodes: 10}
    test: {mode: test, worker: 1, episodes: 3}
```

Figure 2. Example of the definitions part of an experiment document (identifiers shortened for readability)

```yaml
schedule:
  - Adversary Single Training:
      environments: [[midasmv_tar_ms]]
      agents: [[sauron_sac_single]]
      simulation: [vanilla_sim]
      phase_config: [training]
      sensors: &sensors_single
        sauron_sac_single: [all_sensors]
        gandalf_ddpg_single: [all_sensors]
        gandalf_sac_single: [all_sensors]
      actuators: &actuators_single
        sauron_sac_single: [attacker_actuators]
        gandalf_sac_single: [defender_actuators]
  - Operator Single Training:
      environments: [[midasmv_tar_ms]]
      agents: [[gandalf_sac_single]]
      simulation: [vanilla_sim]
      phase_config: [training]
      sensors: *sensors_single
      actuators: *actuators_single
  - Autocurriculum Training:
      environments: [[midasmv_tar_ms]]
      agents: [[sauron_sac_ac, gandalf_sac_ac]]
      simulation: [vanilla_sim]
      phase_config: [training]
      sensors: &sensors_ac
        sauron_sac_ac: [all_sensors]
        gandalf_sac_ac: [all_sensors]
      actuators: &actuators_ac
        sauron_sac_ac: [attacker_actuators]
        gandalf_sac_ac: [defender_actuators]
  - Adversary (S) vs. Operator (AC) Test:
      environments: [[midasmv_tar_ms]]
      agents: [
        [sauron_sac_single, gandalf_sac_ac]]
      simulation: [vanilla_sim]
      phase_config: [test]
      sensors:
        <<: [*sensors_ac, *sensors_single]
      actuators:
        <<: [*actuators_ac, *actuators_single]
  - Adversary (AC) vs. Operator (S) Test:
      environments: [[midasmv_tar_ms]]
      agents: [
        [sauron_sac_ac, gandalf_sac_single]]
      simulation: [vanilla_sim]
      phase_config: [test]
      sensors:
        <<: [*sensors_ac, *sensors_single]
      actuators:
        <<: [*actuators_ac, *actuators_single]
  - Adversary (AC) vs. Operator (AC) Test:
      environments: [[midasmv_tar_ms]]
      agents: [[sauron_sac_ac, gandalf_sac_ac]]
      simulation: [vanilla_sim]
      phase_config: [test]
      sensors: *sensors_ac
      actuators: *actuators_ac
```

Figure 3. Experiment schedule with factors definition (continued from Figure 2)

document and all non-public data in order to allow others to recreate an experiment; distributing full setups, up to fully-configured virtual machine images, should not be necessary. Since co-simulation setups with learning agents are even more complex, ensuring reproducibility requires a number of additional precautions.

The previous section already introduced the notion of a specific seed; seeding Pseudo Random-Number Generator (PRNG) with a known number is a common practice. Then,

function EXPAND-SCHEDULE(experimentrun)
  $schedule =$ EMPTY-LIST
  for $phase \in experimentrun.phases$ **do**
    $schedule \leftarrow schedule \cup$ DEEP-COPY(
        UPDATE-MAPPING($schedule, phase$))
  end for
  return $schedule$
end function
function UPDATE-MAPPING(src, upd)
  for $key, value \in upd$ **do**
    if val isa $Mapping$ **then**
      $entry \leftarrow val_{key} \vee$ EMPTY-MAPPING()
      $src_{key} \leftarrow$ UPDATE-MAPPING($entry, value$)
    else
      $src_{key} \leftarrow value$
    end if
  end for
  return src
end function

Figure 4. Phase configuration cascade algorithm

the PRNGs will still emit random numbers, but their sequence will stay the same. This is an obviously important feature for reproducibility. Software packages usually allow this seeding with custom values, e. g., *NumPy*, a commonly used library in scientific software, has a chapter in their documentation about seeding; *PyTorch*, a popular deep learning library, allows this equally.

Since the premiss of co-simulation DoE is that software packages are used transparently, at least ensuring that the same software versions are used across devices is another important part of reproducibility. Almost all programming languages allow to query the software packages that are currently installed. For example, in Python, the command `pip freeze` outputs a list that can later be used to re-install the same versions. Such a `requirements.txt`-style software package versions list can be embedded in the experiment (run) document. Especially noteworthy in this regard are source code management systems like *git* that allow to unambiguously identify each state of a software repository through hashed commits. An additional section `software` (not shown in the abbreviated Figures 2 and 3 due to space constraints) contains a mapping of package management identifier to software package specfications list. For example, one package management identifier would be `pip3` to denote Python pacakges; it maps to a list of `requirements.txt`-style software package specifications that would be passed verbatim to `pip3 install`.

Finally, the experiment and experiment run documents themselves need to be checked for their identity: Changes to such a document should be detected by the software without requiring the user to change its name or any other unique means of identification: A human can easily forget to change an identifier when a factor is changed. Moreover, when convenience features are being used (e. g., the cascading property of arsenAI's experiment (run) documents, or YAML anchors), then the document becomes semantically ambiguous.

For example, consider two phases of one experiment, one called "training" and the other called "testing," in which an agent is first trained and then evaluated in another environment configuration. In arsenAI, there are three ways to define the second phase: (1) Not mentioning the agent and using the cascading configuration feature of arsenAI; (2) using a YAML anchor, or (3) simply copying the definition of the previous phase.

Thus, a hash value needs to be computed of an experiment document in order to allow unique identification of documents. To compute a hash value, all idiosyncrasies must be removed and the resulting document then hashed. For this, the experiment document must first be fully expanded, i.e., the cascading configuration explicitly spelled out (cf. Figure 4. Afterwards, we convert the YAML document into JSON in the most compact form possible, since the YAML format allows—by design—for a number of ambiguities. Finally, we hash the resulting character stream. The conversation is easily possible because YAML is a superset of JSON [18]; thus, YAML documents can always be reduced to JSON. The stricter syntax of JSON ensures the deterministic, non-ambiguous minimal form (i.e., with all unnecessary whitespace removed), if only commas and colons are accepted as separators and keys to objects are lexicographically sorted. Hashing the result of `to_json(expanded_experiment_document, separators=(',', ':'), sort_keys=True)` will be unambiguous across different machines.

IV. EVALUATING RESULTS

Usually, results evaluation means reporting created by a human for a particular experiment, such as a Jupyter notebook with custom-made plots or other means of (statistical) analysis. While this is the usual way for an analysis, e.g., prior to publication, it most commonly implies manual verification of numbers and plotted graphs. However, when learning agent systems are an integral part of a co-simulated CPES experiment, there are many metrics to consider, as well as their interdepedence and the factors the influence them.

Usually, presentations begin with the graph of the reward or utility function, plotted by time or episode. For the example given, Figure 5 depicts an initial plot and analysis of the agent's utility functions that seem to indicate that the autocurriculum-training does indeed lead to a better agent policy. However, additional analysis is warranted, giving a box or violin plot. Also, key metrics from the environment are presented, such as voltage magnitudes in the case of power grid simulations. As the reward/utility function of an agent is a user-defined piece of code, the correlation between observed values and the reward should be inspected. Moreover, an agent's apparent success could be due to an advantageous initial state, so these calculations should be performed given several experiment runs with different random seeds. If time series data (e.g., weather data) are part of the simulation, the starting date must be varied, too. For a sound analysis of the experiment, there are usually more metrics that can be considered, depending on the actual subject of the investigation. For DRL agents, the entropy value during training is often of interest to investigate the exploration-exploitation trade-off.

Most often, these particular values under investigation have a close relationship to each other; whether an experiment is supporting the formulated hypothesis or not then hinges on thresholds either for individual values or for their correlation factor. I. e., an experimenter can—or rather, should—usually formulate beforehand whether a hypothesis is validated or refuted. This can be done through invariants, or, more broadly, user-supplied code that, in addition to the usual analysis, returns a boolean value that indicates the state of the hypothesis.

palaestrAI stores results in a database and provides a convenience query interface that allows to quickly retrieve the most common values from it. This way, custom results validation functions can be provided as part of an experiment. The `invariants` key of an experiment (run) definition document maps to a list of names that reference classes. At the end of an experiment run, each is instantiated and their `check()` method called. This method receives the reference to the current experiment run's data in the results database. If all check functions return `true`, the hypothesis is validated.

The "experimentation pipeline" described allows not only to create a hypothesis validator for a particular experiment, but also to provide numerous building blocks for hypothesis validation that can be used in a number of experiment. For example, there are definite boundaries for voltage magnitudes. One particular check, called `VoltageHealth`, would then check that no bus would ever see a voltage magnitude outside of the range $0.85 \leq V \leq 1.15$ p.u.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a way to conduct sound and reliable experimentation with learning agents in co-simulated CPES. This includes the presentation of a experiment definition document, as well as motivating the most important design as aspects of this approach.

The approach of the experiment definition document as well as the tool suite offer a number of benefits. First, the format is both easy for humans to write, yet also easy to process by the framework. This notion not just includes the text format (YAML), but also its structure, which allows easy instantiation of software components on-the-fly to configure the simulation. Together with arsenAI's DoE features, even large studies can be formulated in one file and the executed also at large scale, with palaestrAI allowing to utilize multiple Graphics Processing Units (GPUs) as well as a fleet of containers.

However, the instantiation also gives rise to a practical challenge. Software modules are created at the start of the simulation, their parameters are—depending in the particular developer's coding style—evaluated when first used. Thus, errors in the code are caught at a later point, compared to a monolithic setup or other approach that allows static analysis. Therefore, finding and fixing bugs in one's experiment setup can be prone to longer startup cycles. Furthermore, even though basic syntax checking through YAML schema exists, there is no
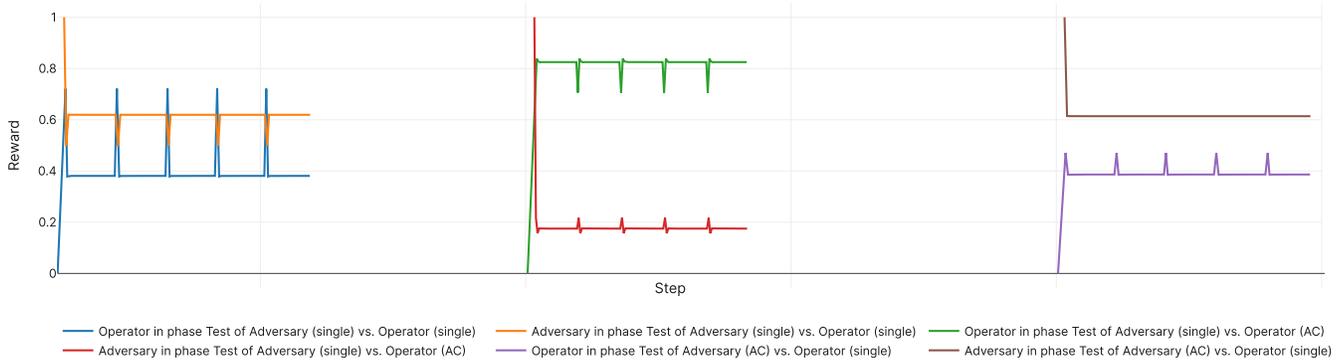
Figure 5. Preliminary results from the autocurriculum experiment (cf. Figures 2 and 3)

semantic help or specific syntax highlighting as fully integrated simulation environments can provide.

In the future, we will showcase an extensive experiment that investigates the benefits of autocurriculum learning with the ARL methodology and demonstrate how the particular elements of our approach tie into a validation of a scientific hypothesis in a largely automated fashion, driven by an experiment definition document and the palaestrAI software suite.

REFERENCES

[1] V. Mnih *et al.*, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[2] J. Schrittwieser *et al.*, *Mastering Atari, Go, Chess and Shogi by planning with a learned model*, 2019. arXiv: 1911.08265.

[3] O. Vinyals *et al.*, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, Nov. 2019, Number: 7782 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/s41586-019-1724-z. [Online]. Available: https://www.nature.com/articles/s41586-019-1724-z (visited on 01/29/2024).

[4] E. M. Veith, *Universal Smart Grid Agent for Distributed Power Generation Management*. Logos Verlag Berlin GmbH, 2017.

[5] R. Diao *et al.*, "Autonomous voltage control for grid operation using deep reinforcement learning," *IEEE Power and Energy Society General Meeting*, vol. 2019-Augus, 2019, ISSN: 19449933. DOI: 10.1109/PESGM40551.2019.8973924. arXiv: 1904.10597.

[6] B. L. Thayer and T. J. Overbye, "Deep reinforcement learning for electric transmission voltage control," in *2020 IEEE Electric Power and Energy Conference (EPEC)*, IEEE, 2020, pp. 1–8.

[7] Z. Wu, C. Li, and L. He, "A novel reinforcement learning method for the plan of generator start-up after blackout," *Electric Power Systems Research*, vol. 228, p. 110 068, Mar. 1, 2024, ISSN: 0378-7796. DOI: 10.1016/j.epsr.2023.110068. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779623009550 (visited on 01/29/2024).

[8] K. Arshad *et al.*, "Deep reinforcement learning for anomaly detection: A systematic review," *IEEE Access*, vol. 10, pp. 124 017–124 035, 2022. DOI: 10.1109/ACCESS.2022.3224023.

[9] E. Veith, A. Wellßow, and M. Uslar, "Learning new attack vectors from misuse cases with deep reinforcement learning," *Frontiers in Energy Research*, 2023.

[10] T. Wolgast *et al.*, "Analyse–learning to attack cyber-physical energy systems with intelligent agents," *SoftwareX*, Apr. 2023. DOI: 10.1016/j.softx.2023.101484. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2023SoftX..2301484W/abstract.

[11] R. A. Fisher, "Design of experiments," *British Medical Journal*, vol. 1, no. 3923, p. 554, Mar. 14, 1936, ISSN: 0007-1447. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2458144/ (visited on 01/29/2024).

[12] A. Ofenloch *et al.*, "Mosaik 3.0: Combining time-stepped and discrete event simulation," in *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*, 2022, pp. 1–5. DOI: 10.1109/OSMSES54027.2022.9769116.

[13] H. Neema, J. Sztipanovits, M. Burns, and E. Griffor, "C2wt-te: A model-based open platform for integrated simulations of transactive smart grids," 2016. DOI: 10.1109/MSCPES.2016.7480218.

[14] H. Neema *et al.*, "Simulation integration platforms for cyber-physical systems," in *Proceedings of the Workshop on Design Automation for CPS and IoT*, ser. DESTION '19, New York, NY, USA: Association for Computing Machinery, Apr. 15, 2019, pp. 10–19, ISBN: 978-1-4503-6699-1. DOI: 10.1145/3313151.3313169. [Online]. Available: https://dl.acm.org/doi/10.1145/3313151.3313169 (visited on 01/28/2024).

[15] S. Balduin, E. M. S. P. Veith, and S. Lehnhoff, "Midas: An open-source framework for simulation-based analysis of energy systems," in *Simulation and Modeling Methodologies, Technologies and Applications*, G. Wagner, F. Werner, and F. De Rango, Eds., Cham: Springer International Publishing, 2023, pp. 177–194, ISBN: 978-3-031-43824-0.

[16] E. Veith *et al.*, "palaestrAI: A training ground for autonomous agents," in *Proceedings of the 37th annual European Simulation and Modelling Conference*, EUROSIS, Oct. 2023.

[17] M. Hutson, "Artificial intelligence faces reproducibility crisis," *Science*, vol. 359, no. 6377, pp. 725–726, Feb. 16, 2018, Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.359.6377.725. [Online]. Available: https://www.science.org/doi/full/10.1126/science.359.6377.725 (visited on 01/29/2024).

[18] O. Ben-Kiki, C. Evans, and I. döt Net, *YAML ain't markup language (YAML™) version 1.2*, [Retrieved: 2024-02-09]. [Online]. Available: https://yaml.org/spec/1.2.2/#102-json-schema.

Engaging SMEs and Public Institutions in Renewable Energy and Energy Efficiency

A Case Study from Sibiu County, Romania

Mihaela Rotaru, Claudiu Isarie
Faculty of Engineering
Lucian Blaga University of Sibiu
Sibiu, Romania
e-mail: {mihaela.rotaru, claudiu.isarie}@ulbsibiu.ro

Lasse Berntzen
School of Business
University of South-Eastern Norway
Horten, Norway
e-mail: lasse.berntzen@usn.no

Abstract—The paper presents a collaborative project on renewable energy and energy efficiency in Sibiu County, Romania. Through interactive stakeholder workshops, several strengths and weaknesses were identified. Through an extensive training program for small and medium-sized enterprises and public institutions, the project has led to measurable impacts, including increased public awareness, adoption of sustainable technologies, and significant energy and monetary savings, demonstrating the effectiveness of collaborative initiatives in driving the renewable energy agenda and contributing to global efforts against climate change. The project also produced a guide to help Small and Medium-sized Enterprises (SMEs) and public institutions adopt renewable energy and energy efficiency measures.

Keywords-renewable energy; energy efficiency; training; strategy.

I. INTRODUCTION

This paper presents an overview of the collaborative project "Renewable Energy and Energy Efficiency for Sustainable Development in Sibiu County," a joint effort between Sibiu County, Romania, Lucian Blaga University of Sibiu (LBUS), Romania, and the University of South-Eastern Norway (USN). Initiated in April 2023 and slated for completion in March 2024, the project aims to foster sustainable development in Sibiu County through targeted training programs and awareness campaigns focused on renewable energy and energy efficiency.

The project's core objective is to develop training materials and courses to assist Small and Medium-sized Enterprises (SMEs) and public institutions in adopting renewable energy solutions and reducing energy consumption. This initiative seeks to train representatives from these entities and raise renewable energy awareness among the citizens of Sibiu County. While this paper concentrates on the interactions with SMEs and public institutions, a subsequent article will detail the public awareness campaign. More information about the project is available on the website [1].

In the face of projections indicating a sharp increase in electricity consumption within the residential and commercial sectors, the urgency to transition towards more sustainable energy practices becomes paramount. With non-OECD countries expected to see electricity consumption rise significantly by 2050 [2] and the European Union (EU) setting ambitious greenhouse gas reduction targets for 2030 [3], the project aligns with broader efforts to address climate change and promote energy efficiency.

The methodology for creating the project's training materials and courses involved reviewing existing strategies and qualitative research, including semi-structured interviews with local officials, focus group discussions with various stakeholders and stakeholder consultations. These approaches provided comprehensive insights into the region's challenges and opportunities for implementing renewable energy and energy efficiency measures. The outcomes of these efforts are encapsulated in a guide on renewable energy, serving as a key deliverable of the project and a resource for promoting sustainable energy practices in Sibiu County and beyond.

While many renewable energy training programs exist, the authors have searched for and found no similar programs where a county has set out to train SMBs and public institutions on renewable energy.

Section 2 describes the methodology, Section 3 discusses implementation, Section 4 provides results, Section 5 discusses the results, and Section 6 provides a conclusion and plans for future work.

II. Methodology

To create training materials and courses, the project group used a combination of reviewing current strategies and plans [4][5] combined with qualitative information gathering through the following methods:

A. Semi-Structured Interviews

Semi-structured interviews were conducted with local officials representing each Administrative Territorial Unit (ATU) within Sibiu County, where lecturers from LBUS prepared training for citizens and workshops for specialists.

These interviews provided a unique opportunity to engage with influential decision-makers and acquire detailed insights into the specific circumstances, challenges, and prospects of implementing renewable energy and energy efficiency measures in the region.

B. Focus Group Discussions

Focus group discussions were organized with representatives from diverse stakeholder groups, including local government agencies, academic institutions, energy utilities, non-governmental organizations (NGOs), and community groups.

These sessions facilitated candid dialogue, idea sharing, and collaborative ideation on sustainable energy initiatives, comprehensively examining varied viewpoints and priorities.

C. Stakeholder Consultations

Stakeholder consultations were conducted to engage a broader spectrum of stakeholders and solicit feedback on strategic planning, policy formulation, and project implementation.

These consultations comprised interactive sessions wherein stakeholders had the opportunity to articulate their perspectives, express concerns, and actively contribute to shaping the trajectory of sustainable energy projects in the ATU.

The key deliverable from the project, a guide on renewable energy and energy efficiency, was created based on the collected data.

III. Implementation

Workshops were conducted in more than 30 ATUs across Sibiu County, selected based on their designation as development poles. Development poles, characterized by concentrated economic activities and strategic importance for regional development, were identified as ideal locations for delivering training initiatives. When selecting the ATUs for citizens' training, several vital characteristics were considered:

A. Population Density

ATUs with higher population densities were prioritized to ensure broader outreach and maximize the dissemination of knowledge on renewable energy and energy efficiency practices. Also, there is the possibility that some citizens from nearby villages will attend the training.
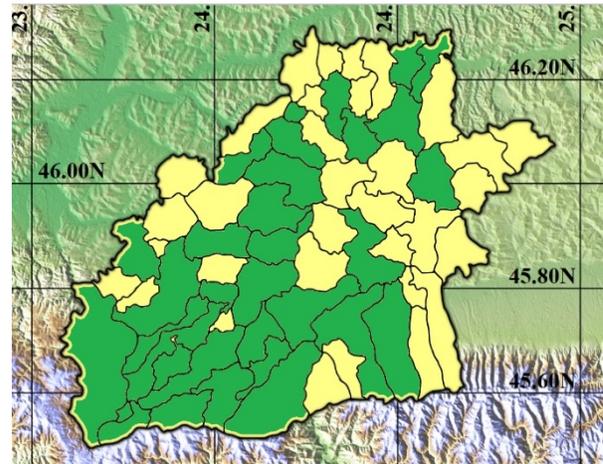

Figure 1. Sibiu County map with selected ATU's
Source: Sibiu County Council

The selected ATUs comprise roughly 81% of Sibiu County's population, which is 388,325 residents. This demonstrates substantial coverage of the county's population.

Fig. 1 shows the 64 administrative-territorial units of Sibiu County: 2 municipalities, nine towns, and 53 communes. The ATUs selected for citizens' training are highlighted in green.

Sibiu County is situated in the center of Romania, covers an area of 5433 km² and shares borders with Mures County to the north, Arges and Valcea Counties to the south, Alba County to the west, and Brasov and Mures Counties to the east.

Sibiu County's terrain comprises three main relief areas: the mountainous area, the highland area, and the depression contact area. The

mountainous region, covering approximately 30% of the county, includes the Făgăraş Mountains, with peaks exceeding 2500 m. The highland area constitutes around 50% of the county, with heights ranging from 490 m to 749 m. The depression contact area, occupying about 20% of the county, runs continuously between the mountainous and highland regions.

The decision to prioritize workshops in development poles with higher population densities was strategic, as it allowed for the efficient allocation of resources and ensured a more significant impact on a larger population segment. Moreover, conducting workshops in these areas aligns with broader regional development objectives, leveraging existing economic activities and infrastructure to promote sustainable energy practices and drive positive change within the community.

B. Energy Consumption Patterns

ATUs exhibiting significant energy consumption rates, particularly in the residential sector, were targeted to address specific energy-related challenges and opportunities relevant to the community.

Analyzing energy consumption patterns in Sibiu County involves examining various energy usage factors across different sectors, such as residential industrial and commercial energy consumption, transportation energy use, renewable energy integration, non-electrified households located more than 2 km from the distribution network, energy efficiency measures, and environmental impacts.

Sibiu County has a significant residential sector contributing to energy consumption, with a diverse range of housing types, including urban apartments, suburban houses, and rural dwellings. Household energy consumption is influenced by heating, cooling, lighting, and household appliances. Seasonal variations in energy demand are likely, with higher consumption during winter months for heating purposes. Adoption of energy-efficient technologies and renewable energy systems in residential buildings may vary across different areas of the county.

Sibiu County hosts various industries and commercial establishments, including manufacturing plants, businesses, and retail outlets. Production processes, machinery operation, and facility heating/cooling needs drive industrial energy consumption. Lighting, heating/cooling systems, refrigeration, and equipment operation influence commercial energy usage. Energy-intensive industries, such as manufacturing, significantly impact overall energy consumption patterns in the county.

Transportation significantly contributes to energy consumption in Sibiu County, combining private vehicles, public transportation, and freight transport. Commuting patterns, travel distances, and vehicle types (including conventional and electric vehicles) affect transportation energy consumption. Investments in public transportation infrastructure, promotion of sustainable mobility options, and improvements in fuel efficiency have influenced energy usage in the transportation sector.

In recent years, Sibiu County has undergone significant developments in its public transportation infrastructure, marking a notable transition towards enhanced mobility and connectivity for its populace. Positioned centrally within Romania, this region has adopted progressive measures to address transportation demands while emphasizing sustainability and environmental conservation.

Central to this transformation has been the expansion and optimization of bus networks throughout the county. Prioritizing improved linkages between urban centers, suburban locales, and rural areas, introducing new bus routes has sought to augment service frequency and reliability. Concurrently, integrating modern, low-emission buses into the fleet attempted to minimize the ecological footprint while ensuring passenger comfort and operational efficiency.

Complementing the expansion initiatives, considerable endeavors have been directed toward modernizing Sibiu County's bus fleet. Replacing older, less environmentally friendly vehicles with contemporary counterparts adhering to stringent emission standards has markedly curtailed environmental pollution, enhancing air quality within the region. Such transitions underscore an ecological prerogative and a commitment to

elevating public transit systems' overall efficacy and sustainability.

In tandem with these efforts, Sibiu County has embraced a spectrum of sustainable mobility solutions aimed at diminishing reliance on private automobiles and promoting alternative modes of transportation. Initiatives encompassing bike-sharing programs, pedestrian-oriented infrastructure enhancements, and the establishment of electric vehicle charging infrastructure have been pivotal in fostering environmentally conscious travel behaviors among residents.

Sibiu County meets the conditions to exploit various renewable energy resources, including solar, biomass, and hydroelectric potential. Integrating renewable energy sources into the county's energy mix can reduce dependence on fossil fuels, mitigate greenhouse gas emissions, and promote energy independence in isolated localities. Adoption of renewable energy technologies in residential, commercial, and industrial sectors can vary based on factors such as resource availability, economic feasibility, and policy support.

Implementing energy efficiency measures is crucial in reducing overall energy consumption and improving sustainability in Sibiu County. Initiatives, such as energy audits, building insulation, efficient lighting systems, and appliance standards, can help optimize energy use. Public awareness campaigns and incentives for energy-efficient practices may encourage individuals, businesses, and organizations to adopt energy-saving measures.

In addition, the evaluation procedure was significantly influenced by the pre-existing infrastructure within the ATUs. ATUs with infrastructure that facilitated the adoption of renewable energy sources were prioritized. The objective of this decision was to optimize the utilization of existing resources and support the successful execution of sustainable energy initiatives, thus enhancing the effectiveness and efficiency of interventions.

Community participation was another crucial aspect of ATU selection. It was determined that ATUs that actively engaged the community in sustainability initiatives were the most suitable ones for intervention. It was acknowledged that the active engagement of community members was critical for the effective adoption and execution of sustainable practices, thereby guaranteeing the initiatives' enduring viability and community ownership of initiatives.

Furthermore, the strategic significance of ATUs in the county of Sibiu was considered throughout the selection procedure. Priority was given to intervening in ATUs identified as vital economic centers or strategically positioned to stimulate broader regional development and economic expansion. Utilizing the economic potential of these ATUs, this strategic approach is intended to propel sustainability initiatives throughout the county.

In conclusion, policy support was indispensable in the ATU selection process. Local governments that exhibited a steadfast dedication to advancing renewable energy and energy efficiency were given precedence over ATUs lacking such frameworks. Policy environments of this nature were crucial in establishing a conducive atmosphere for effective training initiatives and guaranteeing the enduring viability of interventions. The workshops aimed to empower citizens with knowledge and skills related to renewable energy resources, energy efficiency measures, and available financial opportunities. Participants engaged with specialists from LBUS to gain insights into sustainable energy practices and explore potential pathways for implementation.

During two research trips to Norway, the project team gathered inspiration for innovative ideas to be implemented in Sibiu County. The trips included visits to Horten High School with BREEAM NOR Outstanding certification [6], a regional hospital with an extensive energy efficiency program, a local airport with an energy efficiency plan, a biogas production facility, and meetings with climate and energy advisors in Vestfold County Council.

IV. RESULTS

Technical specialists from six pivotal communities within Sibiu County participated in three-day training sessions focused on renewable energy, energy efficiency, and energy security.

Each session, conducted in collaboration with experts from the University of Southeast Norway (USN), addressed topics such as energy savings technology, alternative energy systems, and environmental protection. The training aimed to enhance the capacity of technical experts to implement sustainable energy practices within SMEs and public institutions.

A. Strengths

During the sessions with technical experts, the following strengths of the Sibiu County strategy were identified [17]:

1) Medium and long-term vision for reducing greenhouse gas emissions

Sibiu County demonstrates a forward-thinking approach with a clear vision for reducing greenhouse gas emissions, complemented by expert technical training that enhances the capacity of local stakeholders to implement renewable energy and energy efficiency initiatives aligned with county objectives.

2) Stable international relations

The county's stable international relations facilitate collaboration opportunities, knowledge exchange, and access to best practices from global partners in renewable energy and energy efficiency, further augmented by expert technical training that fosters international partnerships and expertise sharing.

3) Programs in place for renewable energy

Existing programs and initiatives focused on renewable energy, and expert technical training provides a robust foundation for sustainable energy transition, offering practical skills and knowledge to stakeholders involved in implementing renewable energy projects.

4) Increased capacity for external funding

Sibiu County's increased capacity to attract and manage externally funded projects is bolstered by expert technical training, which equips local stakeholders with the necessary expertise to navigate funding opportunities and effectively utilize resources for renewable energy and energy efficiency initiatives.

5) Specialized human resources

The county benefits from specialized human resources with expertise in renewable energy technologies and energy efficiency measures, augmented by expert technical training that enhances skills development and knowledge transfer among local stakeholders, ensuring the successful implementation of renewable energy projects.

6) Functional infrastructure and logistics

Sibiu County's functional infrastructure and logistics, including modern material bases and established mechanisms for project implementation, are complemented by expert technical training, which provides practical insights and guidance on effectively utilizing infrastructure and resources for renewable energy initiatives.

7) Strong partnerships with stakeholders

Stable partnerships with stakeholders are reinforced by expert technical training, which promotes collaboration and coordination among government agencies, academic institutions, and private sector entities involved in renewable energy and energy efficiency projects, fostering a supportive ecosystem for sustainable development.

8) Participatory decision-making processes

Sibiu County's participatory decision-making processes are enhanced by expert technical training, which empowers stakeholders to actively contribute to renewable energy and energy efficiency policies and initiatives, ensuring that decisions are informed by local expertise and needs.

9) Access to information sources

Easy access to information sources at national, European, and international levels is complemented by expert technical training, which provides stakeholders with up-to-date knowledge and resources to support informed decision-making and knowledge sharing in the renewable energy and energy efficiency sectors.

10) Political support

Political support within the County Council is reinforced by expert technical training, which cultivates a conducive environment for advancing renewable energy and energy efficiency agendas, strengthening the implementation of strategic initiatives and projects through collaborative efforts between policymakers and trained technical experts.

B. Weaknesses

The sessions also identified several weaknesses:

1) Insufficiently developed strategic framework for climate change

A comprehensive strategic framework for combating climate change and adapting to its effects is necessary for effective coordinated efforts and initiatives to address climate-related challenges.

2) Lack of integrated public policies

The absence of integrated public policies at the county level on climate change exacerbates the challenge of implementing cohesive strategies and actions to mitigate greenhouse gas emissions and enhance resilience to climate impacts, as well as inhibiting the effectiveness of training programs in fostering widespread adoption of sustainable practices.

3) Inadequate database management

Facing challenges in maintaining integrated and updated databases of implemented projects related to climate change, limiting the ability to track progress and assess impact effectively and hindering the training program's ability to tailor interventions to specific needs and measure outcomes accurately.

4) Uncertain legal status of assets

The uncertain legal status of some assets owned by Sibiu County, such as buildings and land, poses obstacles to implementing energy efficiency measures and sustainable development initiatives on these properties while potentially limiting the effectiveness of training programs in addressing these challenges comprehensively.

5) Lack of organized working groups

The absence of organized working groups at the county level dedicated to addressing climate change and facilitating technical training restricts collaboration and coordination among stakeholders, impeding progress in climate action planning and implementation and limiting the effectiveness of training programs in fostering knowledge sharing and capacity building.

6) Low cooperation with local authorities

A low level of cooperation with local authorities within the county regarding climate change issues hinders coordinated efforts and synergy in addressing shared challenges, impeding the training program's ability to mobilize local resources and support.

7) Lack of awareness and education initiatives

Insufficient mechanisms for raising awareness and educating citizens, decision-makers, and business representatives about the impacts of climate change in Sibiu County contribute to a limited understanding of climate-related risks and opportunities while also potentially hindering the effectiveness of technical training initiatives in fostering behavior change and adopting sustainable practices.

8) Limited disaster response capacity

Sibiu County's capacity to respond to climate-related disasters is constrained by insufficient resources and preparedness measures, increasing vulnerability to extreme weather events and other climate-induced hazards while also potentially limiting the effectiveness of technical training programs in building resilience and emergency response capabilities.

9) Poor rural infrastructure

Some rural areas in Sibiu County suffer from inadequate infrastructure, including roads, utilities, and services, which not only impairs the quality of life for residents but also poses challenges for implementing renewable energy and energy efficiency projects, as well as hindering access to and participation in technical training initiatives.

10) Skilled labor migration

The outmigration of skilled labor from Sibiu County to other regions or countries reduces the availability of qualified personnel for implementing climate change mitigation and adaptation measures. It limits the pool of potential technical training program participants, undermining efforts to build local capacity and expertise in sustainable practices.

11) Uncertain legal status of land

Ambiguities in the legal status of land ownership and usage rights in Sibiu County create uncertainty and barriers to investment and development in renewable energy and energy efficiency projects, as well as complicating efforts to implement land-based climate change mitigation and adaptation strategies, further exacerbating vulnerabilities to climate impacts.

12) Economic and political interests

Conflicting economic and political interests may prioritize short-term gains over long-term sustainability, leading to unsustainable practices in forest conservation and management, which

undermines efforts to mitigate climate change and exacerbates the county's environmental degradation and biodiversity loss.

C. *Elaboration of a Guide*

A comprehensive guide [7] was developed to enhance awareness of the benefits of renewable energy and energy efficiency in Sibiu County. It provides insights, practical information, case studies, and recommendations for adopting sustainable energy practices, targeting the public, technical specialists, policymakers, and businesses. By compiling expert and stakeholder inputs, the guide promotes sustainable development, energy conservation, and the use of renewable resources. It also offers guidance on policy frameworks, funding opportunities, and best practices for renewable energy projects and energy efficiency measures, serving as a tool for capacity building. The guide's effectiveness is demonstrated by its impact on increasing public awareness, comprehension, and motivation towards sustainable energy practices, with its reach measured by the distribution of physical copies and online views/downloads.

Surveys and interviews are conducted to assess changes in knowledge and attitudes before and after exposure to the guide. Statistical information includes:

- The percentage of respondents who demonstrate increased knowledge and understanding of renewable energy and energy efficiency concepts.
- The percentage of households or businesses that have adopted renewable energy technologies or implemented energy efficiency measures recommended in the guide.
- Quantifying the amount of energy saved because of implementing guide-recommended measures.
- Calculate the monetary savings achieved through reduced energy consumption or improved energy efficiency. This includes savings on utility bills or other related expenses.
- Assessing the environmental benefits of adopting renewable energy and energy efficiency measures, such as reductions in greenhouse gas emissions, air pollution, and water usage.
- Qualitative feedback from guide users to understand their perceptions, experiences, and suggestions for improvement.
- Track the distribution and dissemination of the guide to quantify its reach and impact across different audiences.

The success of behavior change efforts will be evaluated based on observable shifts in consumer behavior. The statistical indicator used the number of households/businesses implementing energy-saving measures recommended in the guide and the amount of renewable energy capacity installed due to guide recommendations.

The guide's contribution to sustainable development will be measured in quantifiable reductions in greenhouse gas emissions and improvements in energy efficiency. This includes percentages of carbon emissions and energy consumption reductions achieved through guide-recommended actions.

V. DISCUSSION

Supporting SMEs promotes economic and environmental sustainability. These interventions benefit individual businesses and contribute to the county's overall well-being by offering training sessions and guidance on implementing sustainable energy practices.
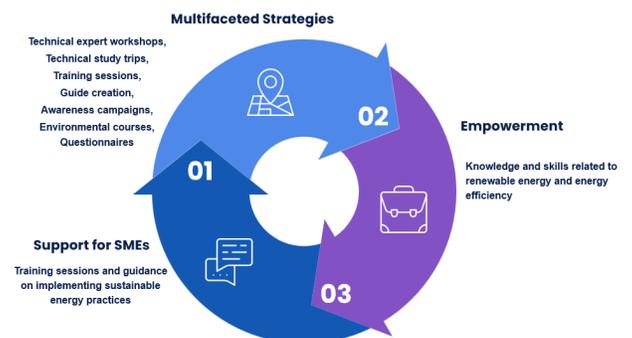

Figure 2. Successes of the project implementation

By supporting SMEs in adopting more sustainable practices, the strategy helps to reduce carbon emissions, mitigate environmental impact, and create a more resilient and prosperous local economy.

Fig. 2 illustrates the multifaceted approach. Overall, this approach, with support for SMEs and public institutions combined with a general awareness campaign, represents a promising strategy for promoting sustainable energy practices. By combining various tactics and focusing on empowerment and support for SMEs, the strategy has the potential to make a meaningful impact on both the environment and the local economy.

VI. CONCLUSION AND FUTURE WORK

In conclusion, the collaboration project in Sibiu County has created a comprehensive guide to foster sustainable development by promoting renewable energy and energy efficiency. This guide is a testament to the concerted efforts of experts, stakeholders, and the community in driving the renewable energy agenda forward. It serves as an educational resource and a catalyst for action, equipping a wide range of audiences with the knowledge and tools necessary to embark on a path toward a more sustainable and energy-efficient future.

The guide's impact is measurable and significant, as evidenced by the public's increased awareness and understanding of renewable energy and energy efficiency concepts, adoption of sustainable energy technologies, and implementation of energy-saving measures. The quantitative and qualitative outcomes— from energy and monetary savings to environmental benefits—highlight the guide's role in effecting tangible changes and contributing to the broader objectives of reducing greenhouse gas emissions and enhancing energy efficiency in Sibiu County.

Furthermore, using a linear model to assess the guide's impact on public awareness has provided valuable insights into the guide's effectiveness in promoting renewable energy and energy efficiency. This model underscores the multifaceted benefits of the guide, from increasing knowledge and understanding to encouraging the adoption of sustainable practices and measuring the subsequent environmental and economic advantages.

As we progress, the guide's success in Sibiu County is a blueprint for future initiatives to advance sustainable energy development. The project's achievements underscore the potential for collaborative efforts to significantly impact community awareness and behavior towards renewable energy and energy efficiency, laying a solid foundation for continued progress in achieving sustainability goals. This endeavor not only enhances the quality of life for the residents of Sibiu County but also contributes to the global effort to combat climate change, marking a pivotal step towards a greener, more resilient future.

REFERENCES

[1] Energie Verde, Project website, [Online]. Available from: https://energieverde.cjsibiu.ro/. 2024.02.15

[2] P. Shrestha, "Global energy use projected to increase by 50% by 2050" [Online]. Available from: https://www.energylivenews.com/2021/10/08/global-energy-use-projected-to-increase-by-50-by-2050/ (2021) 2024.02.15

[3] European Commission, "Stepping up Europe's 2030 climate ambition Investing in a climate-neutral future for the benefit of our people" [Online]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0562 2024.02.15

[4] European Commission. "Romania's 2021-2030 Integrated National Energy and Climate Plan," [Online]. Available from: https://energy.ec.europa.eu/system/files/2020-06/ro_final_necp_main_en_0.pdf 2024.02.15

[5] Consiliul Judetean Sibiu, "Strategia Judetului Sibiu de reducere a emisiilor CO2 pentru perioada 2016-2023," [Online]. Available from: https://www.cjsibiu.ro/wp-content/uploads/2017/02/2017.02.14-STRATEGIA-CJ-SIBIU-CO21.pdf 2024.02.15

[6] BRE Group, "Horten VGS is designed as an energy-positive building that brings BREEAM Outstanding to Norway [Online]. Available from: https://bregroup.com/case-studies/breeam-new-construction/horten-vgs-is-designed-as-an-energy-positive-building-that-brings-breeam-outstanding-to-norway/ 2024.02.15

[7] C. Isarie, F. Ciofu, A. Fălădău, A. Stoica, L. Prodea, "Ghid despre beneficiile utilizării energiei regenerabile și eficienței energetice în județul Sibiu," [Online]. Available from: https://energieverde.cjsibiu.ro/wp-content/uploads/2023/12/brosura_finala.pdf 2024.02.15

LiDAR Data Processing for Utility Asset Management and Fire Risk Assessment

Vivian Sultan

College of Business and Economics
California State University
Los Angeles, USA
vsultan3@calstatela.edu

Benjamin Pezzillo

College of Business and Economics
California State University
Los Angeles, USA
bpezzil@calstatela.edu

Abstract—**California requires utility companies to implement wildfire-mitigation plans to prevent and reduce the risk of catastrophic fires. Manually tracking and evaluating widely distributed equipment, often in very rural and rough terrain, is expensive and labor intensive. This paper demonstrates proof of concept for a light detection and ranging (LiDAR) point-cloud data-processing tool and explores the potential benefits associated with such a tool. LiDAR is widely used for various applications, including mass asset surveys, vegetation management, and structural-load analysis. The authors explored various ArcGIS geoprocessing tools as part of this study. In summary, this paper provides valuable insights into using ArcGIS tools for LiDAR processing and highlights the potential benefits of accurate geolocation data extraction from LiDAR point clouds within utility service territories.**

Keywords- LiDAR Data Processing; Utility Asset Management; Fire Risk Assessment; ArcGIS Tools.

I. INTRODUCTION

Light detection and ranging (LiDAR) can offer potential benefits to California utilities struggling to reduce the cost of tracking their widely distributed assets.

A. Problem Statement

In 2018, California enacted Senate Bill 901 requiring utility companies to implement wildfire-mitigation plans to prevent and reduce the risk of catastrophic fires caused by their equipment. As part of these plans, utility companies must conduct regular visual inspections of their assets, such as power lines, poles, transformers, and substations, according to their type and rate of service. These inspections range from 12 to 24 months for routine maintenance, to 3 to 5 years for comprehensive examination. However, inspecting these assets is not an easy task, as they are often located in remote and rugged areas, where access is limited, and terrain is challenging. Sending crews of inspectors to these locations is time consuming and costly and may not capture all the relevant information needed to assess the condition and performance of the assets.

To overcome these challenges, utility companies use LiDAR technology, which uses laser pulses to measure the location and reflectivity of objects in three dimensions. LiDAR can capture high-resolution point-cloud data of the utility assets and their surroundings, which can be used by utility companies to identify, locate, and monitor their assets more accurately and more efficiently. However, processing and analyzing LiDAR data is not a trivial task; it requires specialized software and expertise. Many utility companies currently outsource this task to third-party vendors, which adds to their expenses and reduces their control over data quality and security.

To address this issue, utility companies can benefit from developing their own LiDAR data-processing tools, which would allow them to bring the processing in-house and save on vendor fees. A LiDAR data-processing tool would help utility companies automate the extraction and classification of assets from point-cloud data and improve their geolocation accuracy and reliability. This would result in better wildfire-mitigation plans, as utility companies would have more up-to-date and detailed information on their assets and their potential fire hazards.

B. Objectives

This paper demonstrates a proof of concept for a LiDAR data-processing tool that would allow utility companies to process and analyze their own LiDAR point-cloud data. The paper also investigates the potential benefits of this tool for improving utility asset management and fire-risk assessment.

Utility companies use LiDAR for various purposes:

- Mass asset surveys: LiDAR can help identify the components and configurations of each pole in the service territory, such as wires, cross-arms, insulators, and transformers.
- Vegetation management: LiDAR can help survey the trees and vegetation that may infringe on the distribution lines and pose a fire hazard or a reliability issue.
- Structural load analysis: LiDAR can help determine the number and condition of poles in high-fire-risk areas and help assess their structural integrity and load capacity.

Utility companies are constantly improving the quality of their asset data by identifying and resolving data-quality issues, such as missing, inaccurate, or outdated information. They are also working on forecasting fire hazards using data-driven models and methods to estimate the probability and severity of fires caused by their equipment.

LiDAR offers the opportunity to extract more accurate geolocation data for utility assets, which can enhance the quality and reliability of the asset data and improve the accuracy and efficiency of the fire-hazard models.

This paper offers a first use case in studying the potential benefits of developing a LiDAR data-processing tool for utility companies. The paper will also evaluate the feasibility and scalability of the tool and identify the challenges and opportunities for future development.

II. BACKGROUND AND OVERVIEW OF LiDAR DATA PROCESSING TOOLS

LiDAR, a remote-sensing technology, uses pulsed lasers to measure and record distances, heights, and depths of objects and areas. It accurately, precisely, and flexibly examines natural and artificial environments. LiDAR data are collected aerially or terrestrially using an unmanned aerial vehicle (UAVs) or unmanned ground vehicles (UGVs). Technicians remotely operate UAVs to scan areas of interest from altitudes greater than ten meters. At a minimum, this process requires a two-person team to remotely operate the UAV and verify the data is correct [1]. Software can read these point-cloud data for further processing. In contrast, UGVs' detection distances range below ten meters to perform precise geometric measurements. UAVs and LiDAR data provide several benefits over sending people to physically inspect all assets of interest. For instance, a UAV can easily scan large areas without regard to terrain (steep slopes, dense forests, etc.).

Several studies have examined the extraction of objects from point-cloud data. For instance, Van Leeuwen and Nieuwenhuis [2] examined the current and future potential for leveraging LiDAR data to assess and manage forest structures, specifically how remote sensing and classification can identify specific trees in clusters and more closely identify species. The article is relevant to this use case because this study examines whether LiDAR can be used to identify power poles and structures, which may be imbedded in forests or other rural areas. Van Leeuwen and Nieuwenhuis demonstrate that remote sensing techniques may help identify objects in a forest (in their case, individual trees) and conclude that further research is needed to assess remote sensing and forest management, as well as using models to recognize objects within point-cloud data [2]. Power poles and towers may blend into a forest canopy, as do to individual trees.

In 2009, Prokhorov [3] examined how 3D LiDAR imaging could be used in conjunction with a recurring neural network (RNN) to identify different objects. With the progression of scanners, 3D LiDAR images provide enhanced measurement data [3]. Prokhorov investigated how the space of points between various objects could be leveraged to create a model to recognize objects [3]. This research concluded that the RNN model showed promise, and that further research into training RNN models is warranted, as is pursuing better 3D data.

Maggiori et al. [4] created an end-to-end framework to classify satellite imagery using convolutional neural networks (CNNs). In their study, they observed how a CNN has significant advantages when classifying satellite imagery data to identify objects and produce quality imagery. However, they also noticed that untrained models did not perform as well. They leveraged an existing model and constructed a set of manually classified data and saw significant improvement in the model. Therefore, they propose a two-step approach leveraging a small set of manually classified data to train a model to classify a large set of unclassified data.

Kudinov [5], working with ESRI and AAM Group, used the point-convolution neural-network (PointCNN) framework to automatically identify power lines and poles. The group used artificial intelligence for the labor-intensive task of manually labeling the point cloud. Their study area was a city in Australia, and their dataset contained around 540 million points. They trained their PointCNN model using four classes: other, wires, stray wires, and utility poles to successfully identify power poles.

Fan et al. [6] studied the you-only-look-once (YOLO) deep-learning algorithm to detect objects in point-cloud datasets. The focus of their research was object detection for self-driving vehicles. These vehicles need real-time information to make decisions and avoid collisions. Consequently, the researchers propose an alternative computationally efficient algorithm dubbed LS-R-YOLOv4 using color images and point-cloud data to precisely segment and detect objects. Borcs et al. [7] proposed a pipeline that quickly classifies point clouds. One component of this pipeline is a CNN trained to classify objects. The model supports the identification of vehicles and pedestrians in urban settings.

Brubaker et al. [8] showed that LiDAR data can be used to accurately pinpoint micromorphology of a large area and compared their results to field-surveyed plots to determine their accuracy. They compared a digital-elevation model (DEM) generated from LiDAR data to the surveyed plots. Their research model was accurate to within 0.3–0.4 m based on manual surveys, which is accurate up to a single point in the point cloud. Their data allowed them to generate the surface constraint of the surveyed area faster and from a greater distance compared to a traditional survey. The DEM is important as it allows LiDAR data to be accurately separated into ground, water, and any surface constraints based on elevation.

Azevedo et al. [9] showcased the use of UAVs to replace helicopters to reduce risks and associated costs. UAVs and LiDAR have lower equipment costs over time, as a team of just a few people can ensure that the data is correct and control the UAV. Equipped with the proper sensors, the UAV is able to quickly scan a large area and send data back to the controller. From there, the LiDAR data can be converted to point-cloud data and fed through an algorithm and software to help identify and sort items in the LiDAR data. They argue that, while the algorithm they used failed to correctly identify possible points, those points were classified as unidentified due to the difficulty of differentiating between vegetation and other objects. They conclude that a more powerful algorithm may correctly identify the points of interest and that graphics processing units can be used to reduce the time required to process the raw data.

Nahhas et al. [10] proposed machine learning with LiDAR data and orthophotos. They showed that the CNN algorithm was able to transform, organize, and label the data. With the orthophotos and LiDAR data, they created a digital surface model, DEM, and shapes. They also input other data to detect buildings. From their findings and experiments, the CNN and machine-learning model accurately classified

background and buildings up to a single data point and drew the geometry and shapes of the building from the LiDAR and orthophotos. Using this model, they were able to transform low-level detail into highly detailed, classified features.

Sultan et al. [11] empirically focused on machine learning to identify power poles and towers from point-cloud data. This study sought to demonstrate the use of a deep-learning model developed by Azevedo et al. [9] to determine whether deep learning is a viable solution for identifying power assets in three California areas. This study instantiated an existing trained model to determine whether deep learning is an effective solution for extracting desired objects from point-cloud data. The deep-learning model successfully identified power poles in both rural and urban areas. Although the model performance was better in urban areas than in rural areas, this study supports the literature that deep learning can successfully classify point clouds. To improve the model performance and to ensure optimal results when training the model, the authors suggest more accurately labeled data representing the objects of interest.

LiDAR data serves as a cost-efficient alternative for surveying large areas of land and generating real-time images of objects on the ground. The point-cloud data generated by scans can be analyzed to identify assets in need of maintenance. In addition to the efficiency afforded by LiDAR, utility companies can potentially lower labor and transportation costs by not sending maintenance crews into the field unnecessarily. The cost of LiDAR depends on the type of equipment purchased and the range and scope of work [12]. LiDAR drones can potentially be cost effective in difficult-to-reach forested areas, rural towns, or high elevations. LiDAR can also be used in densely populated areas such as urban or suburban areas [13]. The high upfront cost leaves just maintenance of the equipment, future upgrades, and pilot licensing as needed [14]. These costs can be calculated in advance, while the ongoing costs of dispatching workers depend on the scope of work and may not be easily estimated due to fluctuating rates of pay [15]. In many cases, contractors may need to be hired in areas that are difficult to reach and may not have the exact quality control utility companies need. On the other hand, manually assessing and inspecting equipment is beneficial as the information about them can be updated in real time, whereas LiDAR data must be processed and analyzed to ensure the data are error free [9]. A high-scale scan must be performed of target areas to produce error-free point-cloud data and these data must be processed to ensure assets are correctly identified [10]. LiDAR technology provides several benefits when surveying objects. Therefore, this study sought to answer the following question. Can a utility company process LiDAR point-cloud data to accurately define asset locations?

The literature suggests deep learning can be used to classify objects of interest. Therefore, this study will instantiate the deep-learning model deployed by Sultan et al. [11] to determine its effectiveness at processing sample point-cloud data. In addition, other ArcGIS Pro classification tools will be studied and tested to gauge their effectiveness at classifying poles and towers. This study may be of interest to executive teams of utility companies, as it can help them decide whether to bring the LiDAR data-processing in-house and the potential benefits of doing so. For example, by processing and analyzing their own LiDAR data, utility companies may be able to improve the accuracy of their asset location data, which can enhance their asset management and fire-risk assessments.

III. METHODOLOGY

For this project, the authors will explore ArcGIS geoprocessing tools, including the deep-learning model deployed, image analytics, and additional tools that complement ArcGIS. Sultan et al. [11] classifies the tools used into three categories: (a) data conversion, (b) deep learning, and (c) LiDAR Aerial Survey (LAS) conversion.

ArcGIS Pro software from the Environmental Systems Research Institute (ESRI) provides three tools to classify data, train a model, and use a model for point-cloud data classification. The following ArcGIS Pro classification tools will be explored and tested by the project team:

- Classify LAS Ground
- Classify LAS Building
- Classify LAS by Height
- Classify LAS Noise
- Change LAS Classification Codes

Phase 1 of this project will include exploration of the tools to evaluate whether ArcGIS Pro LAS-classification tools will support the classification of power poles and towers. A preliminary recommendation should follow Phase 1. Given a positive recommendation, Phase 2 of the project may start to train the model on some sample point-cloud data to give it the best chance of correctly identifying buildings and electrical-system assets in the service territory.

- **Goal:** Classify LiDAR points as wire conductors, transmission towers, and high vegetation.
- **Software:** ArcGIS Pro 3.1.2 with Advanced functionality *(e.g. 3D Analyst Tools)* and ArcGIS Pro 3.1 Deep Learning Frameworks.

A. System Preparation

To prepare a conventional personal computer to run ArcGIS Pro and the other software used in this work, update the computer system and software with the latest versions and drivers. After updating ArcGIS Pro to version 3.1.2 or higher, install the Deep Learning Library downloaded from the ESRI website.

B. Data Preparation

1) Training the Classification Model

Preparation work not covered in this guide involved the classification model and training dataset. Training data were validated, and a classification model was trained using the validated training data.

That work was done using the step-by-step instructions in "Learn ArcGIS tutorial" [16] with some modifications for Classes 05, 14, and 15, and the maximum number of epochs (50) was set in the Train Point Cloud Classification Model

tool in "Train point cloud classification model" [17] and "Classify powerlines from lidar point clouds" [18]:

The LAS dataset had to be converted into smaller training blocks using the Prepare Point Cloud Training Data geoprocessing tool in ArcGIS Pro. Ground (Class 2) and noise (Class 7) points were excluded from the training data. As Ground points typically account for a large portion of the total points, excluding ground points made the training process quicker. Block Size and Block Point limits were determined by the training and validation dataset.

Next, the Train Point Cloud Classification Model geoprocessing tool was used to train a model for classification. The focus of the model training was on three specific classes:

- 05 – High Vegetation
- 14 – Wire Conductor
- 15 – Transmission Tower

That meant in addition to 14 – Wire Conductor, the settings were adjusted in Class Remapping to include the those shown in Fig. 1. Those determinations were made after reviewing the diversity of classes in the point cloud data in the Layer Properties (Fig. 2).
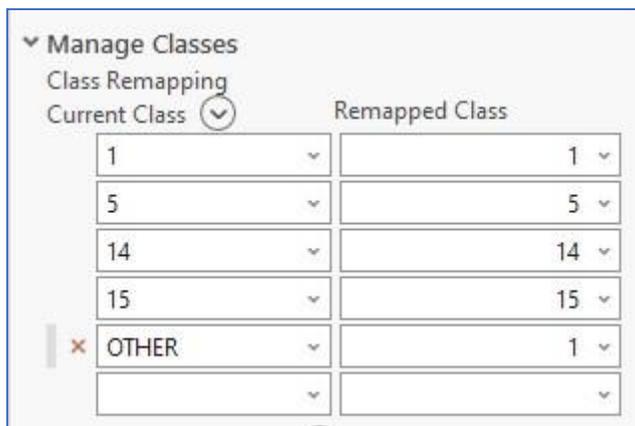

Figure 1.   Class remapping settings.
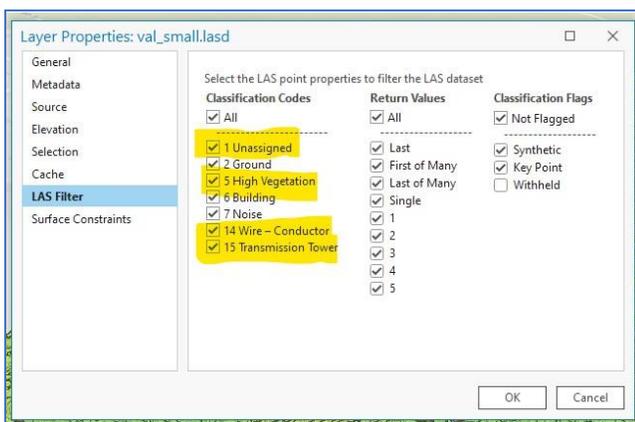
Figure 2.   Layer properties dialogue.

Existing 01 – Unassigned were remapped as unassigned. Since 02 – Ground and 07 – Noise were excluded from the training data in an earlier step, they are subsequently ignored

in this model. Any points classified as 06 – Building are remapped as "OTHER" into 01 – Unassigned.

The validated training data (Fig. 3) depicts 14 – Wire Conductor in yellow, 15 – Transmission Tower in blue, and 05 – High Vegetation in green. Data classified as 07 – Noise appears in red, 02 – Ground appears as brown, and 01 – Unassigned as gray. A fully rendered detail image (Fig. 4) shows 14 – Wire Conductor (yellow) among areas where 05 – High Vegetation is taller than the transmission lines.

Training Loss and Validation Loss values (Figs. 5, 6, and 7) generally decreased, indicating the model learned from the process. After 50 epochs, the highest recall is over .93.
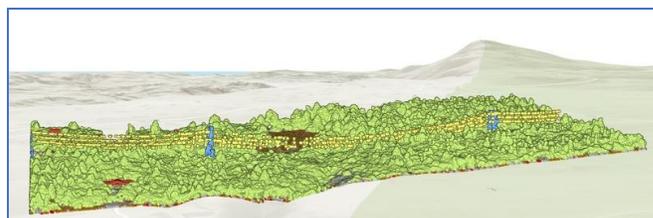

Figure 3.   Validated training data.
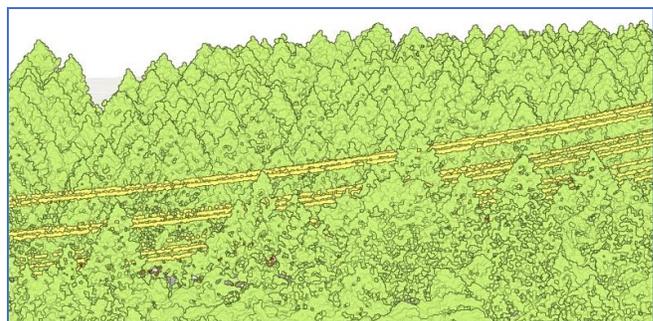
Figure 4.   Fully rendered detail image.
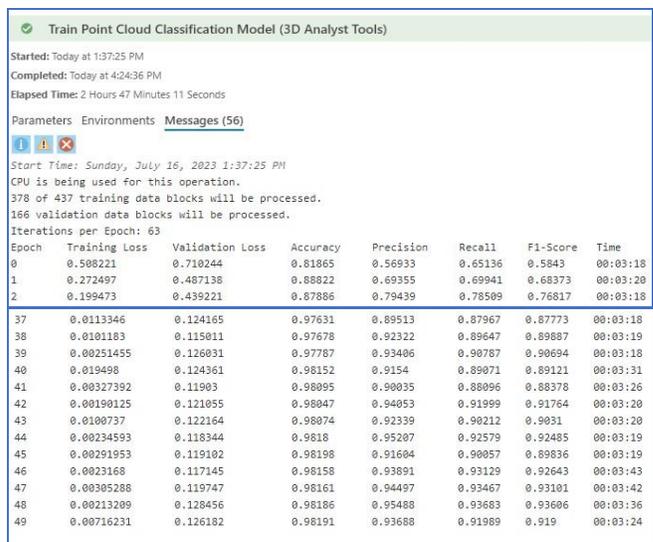
Figure 5.   Training loss and validation loss progression.
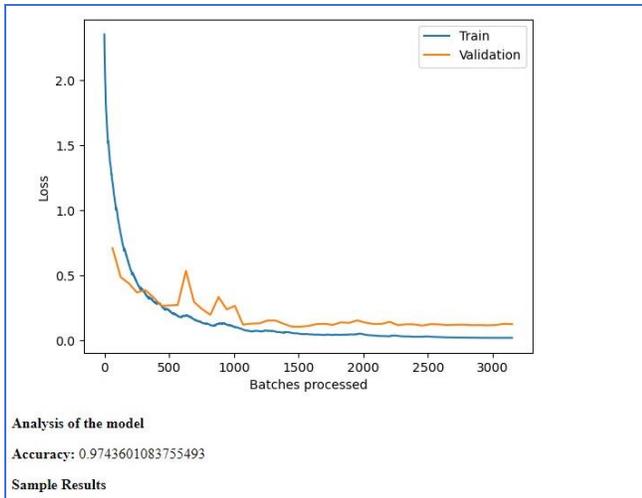
Figure 6.   Ground truth / predictions: Loss versus batches processed.

Analysis of the model

Accuracy: 0.9743601083755493

Sample Results
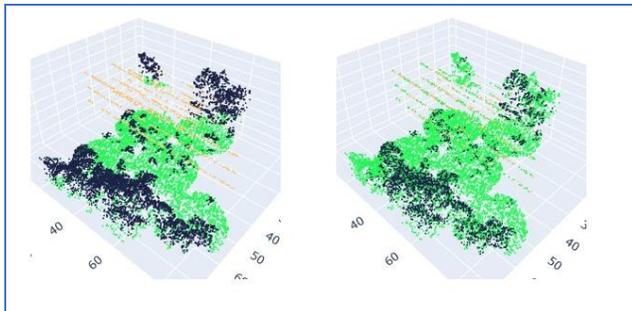
Figure 7.   Ground truth / predictions: 3D graphs.

	A	B	C	D	E	F
1	EPOCH	CLASS_CODE	CLASS_DESCRIPTION	PRECISION	RECALL	F1_SCORE
2	31	1	Unassigned	0.972693283	0.994497942	0.983444022
3	38	1	Unassigned	0.974883226	0.994331473	0.984460706
4	30	1	Unassigned	0.952662858	0.993794284	0.972646059
5	37	1	Unassigned	0.974308914	0.992460495	0.983230258
6	32	14	Wire Conductor	0.926027353	0.992211066	0.941156998
7	34	1	Unassigned	0.957427664	0.992102668	0.974271345
8	15	14	Wire Conductor	0.903094753	0.992043357	0.925726761
9	32	1	Unassigned	0.973381023	0.991933699	0.982511344
10	28	14	Wire Conductor	0.903420316	0.991588214	0.925962781
11	18	14	Wire Conductor	0.9188984	0.991575781	0.936937564
12	19	14	Wire Conductor	0.917575678	0.991234439	0.935845536
13	17	14	Wire Conductor	0.932062469	0.990922666	0.945371814

Figure 8.   Highest recall value for 14 – Wire Conductor.
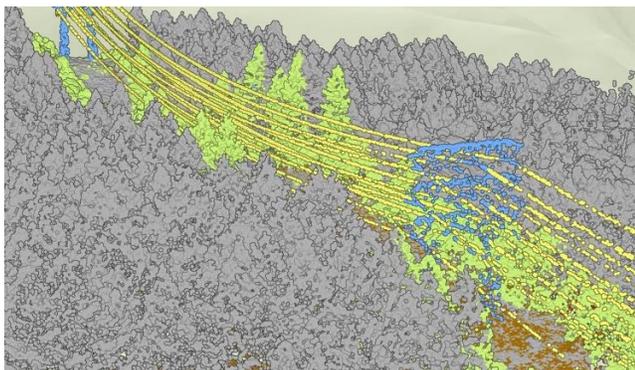
Figure 9.   Test run.

The best epoch was chosen based upon the highest recall value for 14 – Wire Conductor. As shown in Fig. 8, that was Epoch 32 with a recall value of 0.992211066 for 14 – Wire Conductor.

A test run used the trained model to classify 3.2 million cloud points previously comprised of ground points (Class 2), low-noise points (Class 7) and unassigned points (Class 1) and classified them into Wire Conductor (Class 14) in yellow, Transmission Tower (Class 15) in blue, High Vegetation (Class 5) in green, and Unassigned (Class 1) in gray (Fig. 9).

IV.   CONCLUSION AND FUTURE WORK

The aim of this paper was to demonstrate a proof of concept for a LiDAR point-cloud data-processing tool and explore the potential benefits associated with such a tool. LiDAR is widely used for various applications, including mass asset surveys, vegetation management, and structural-load analysis. The authors explored various ArcGIS geoprocessing tools as part of their study:

- Classify LAS Ground: This tool identifies ground points in LiDAR data.
- Classify LAS Building: This tool is used to classify building points.
- Classify LAS by Height: This tool segments points based on height.
- Classify LAS Noise: This tool identifies noise points.
- Change LAS Classification Codes: This tool allows modification of classification codes.

Next steps and future work include importing LiDAR data, converting LAS to LASD, and offering a step-by-step guide to classifying the converted LAS point-cloud data using the trained model. In summary, this paper provides valuable insights into using ArcGIS tools for LiDAR processing and highlights the potential benefits of accurate geolocation data extraction from LiDAR point clouds within utility service territories.

REFERENCES

[1]   National Ocean Service, "What is LiDAR," US Department of Commerce. [Online]. Available from: https://oceanservice .noaa.gov/facts/lidar.html

[2]   M. Van Leeuwen, and M. Nieuwenhuis, "Retrieval of forest structural parameters using LiDAR remote sensing," Eur. J. For. Res., vol. 129, pp. 749–770, 2010, doi:10.1007/s10342 -010-0381-4.

[3]   D. V. Prokhorov, "Object recognition in 3D LiDAR data with recurrent neural network," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 9–15, 2009, doi:10.1109/CVPRW.2009 .5204114.

[4]   E. Maggiori, Y. Tarabaka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification. [Online]. Available from: https:// doi.org/10.1109/TGRS.2016.2612821.

[5]   D. Kudinov, "PointCNN: Replacing 50,000 man hours with AI," Medium. [Online]. Available from: https://medium.com /geoai/pointcnn-replacing-50-000-man-hours-with-ai -d7397c1e7ffe.

[6] Y.-C. Fan, C. M. Yelamandala, T.-W. Chen, and C.-J. Huang, "Real-time object detection for LiDAR based on LS-R-YOLOv4 neural network," J. Sensors, vol. 2021, art. 5576262, 2021, doi:10.1155/2021/5576262.

[7] A. Borcs, N. Balazs, and C. Benedek, "Instant object detection in LiDAR point clouds," IEEE Geosci. Remote. Sens. Lett., vol. 14, no. 7, pp. 992–996, March 2017, doi:10.1109/LGRS.2017.2674799.

[8] K. M. Brubaker, W. L. Myers, P. J. Drohan, D. A. Miller, and E. W. Boyer, "The use of LiDAR terrain data in characterizing surface roughness and microtopography," Appl. Environ. Soil Sci., vol. 2013, art. 891534, 2013, doi:10.1155/2013/891534.

[9] F. Azevedo, A. Dias, J. Almeida, A. Oliveira, A. Ferreira, T. Santos, A. Martins, and E. Silva, "LiDAR-based real-time detection and modeling of power lines for unmanned aerial vehicles," MDPI, vol. 19, no. 8, art. 1812, 2019, doi:10.3390/s19081812.

[10] F. H. Nahhas, H. Z. M. Shafri, M. I. Sameen, B. Pradhan, and S. Mansor, "Deep learning approach for building detection using LiDAR–Orthophoto fusion," J. Sensors, vol. 2018, art. 7212307, 2018, doi:10.1155/2018/7212307.

[11] V. Sultan, J. Ramirez, J. Peabody, and M. Bautista, "Managing power infrastructure using LiDAR," Int. J. Adv. App. Sci., vol. 9, no. 10, pp. 106–115, 2022, doi:10.21833/ijaas.2022.10.014.

[12] J. Antunes, "Should you choose LiDAR or photogrammetry for aerial drone surveys?" Commercial UAV News. [Online]. Available from: https://www.commercialuavnews.com/construction/choose-lidar-photogrammetry-aerial-drone-surveys

[13] K. K. Singh, G. Chen, J. B. McCarter, and R. K. Meentemeyer, "Effects of LiDAR point density and landscape context on estimates of urban forest biomass," ISPRS J. Photogramm. Remote Sens., vol. 101, pp. 310–322, 2015, doi:10.1016/j.isprsjprs.2014.12.021.

[14] C. Van Tassel, "The true cost—Implementing LiDAR into your business," LiDAR News. [Online]. Available from: https://lidarnews.com/articles/the-true-cost-implementing-lidar-into-your-business

[15] T. E. Glavinich and A. L. Chichester, "Pricing service work for profit," Electrical Contractor Magazine. [Online]. Available from: https://www.ecmag.com/section/your-business/pricing-service-work-profit

[16] "Learn ArcGIS tutorial." [Online]. Available from https://learn.arcgis.com/en/projects/classify-powerlines-from-lidar-point-clouds

[17] "Train point cloud classification model." [Online}. Available from https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/train-point-cloud-classification-model.htm

[18] "Classify powerlines from lidar point clouds." [Online}. Available from https://learn.arcgis.com/en/projects/classify-powerlines-from-lidar-point-clouds/