

Comparative Evaluation of Advanced Machine Learning Models in Forecasting Seasonal, Non-linear & Multi-Scale Environmental Patterns

Ibrahim Amazal
Machine Learning & Applications
University of Limerick
Limerick, Ireland
24339571@studentmail.ul.ie

I. ABSTRACT

Forecasting environmental parameters requires advanced machine learning models capable of capturing diverse patterns such as seasonality, non-linearity, and multidimensionality. The current study compares five time-series forecasting models, namely Prophet, LSTM, SVR, TCN, and Transformer, considering their strengths and limitations to identify and model such patterns from different datasets. It follows the CRISP-DM framework, wherein such models have been applied to forecast temperature, foresee the rotor speed of wind turbines, and monitor air quality with the intent of finding out which model suits best for given pattern characteristics.

The approach insists on preprocessing techniques like feature engineering, normalization, and temporal indexing to set the dataset in appropriate format analysis. The overall performance is going to be measured based on metrics such as Root Mean Squared Error, Symmetric Mean Absolute Percentage Error, SHAP analysis, and coherence plots. Results showed that TCN had the best performance concerning learning sequential dependencies and periodic trends, being therefore suitable for accentuated seasonal cycles. LSTM works well in detecting long-term temporal dependencies together with dynamic patterns. This could be appropriate in cases needing detailed temporal relationships. The Transformer model has proved to be highly effective in capturing multi-variable interactions and complex feature relationships, hence ideal for high-dimensional and nonlinear datasets.

This work will present a generally reusable framework for tackling complex forecasting tasks by aligning model capabilities with dataset-specific patterns. These results further the understanding of how machine learning models can be leveraged in pursuit of critical challenges within environmental forecasting, which enables data-driven decision-making and efficient system design.

Keywords: Environmental Forecasting, Time Series, Prophet, LSTM, SVR, TCN, Transformer, Seasonal Patterns, Nonlinear Models, Multi-dimensional Data, CRISP-DM, Feature engineering, Renewable Energy, Air Quality.

II. INTRODUCTION

Environmental forecasting is a cornerstone in addressing global challenges such as climate change, pollution control, and resource management and according to WEF [1] the international community has faced significant challenges over the past 30 years in responding to the rapid changes in atmospheric conditions, largely driven by rising greenhouse gas emissions. These changes have escalated to the point where awareness was ranked as the highest in the Global Risk Report 2011, despite extensive diplomatic efforts to mitigate the impact. Adding to the urgency, the IPCC [2] has projected that the 1.5°C global warming threshold will likely be exceeded by 2030 due to unprecedented increases in atmospheric levels of carbon dioxide, methane, and nitrous oxide. Supporting this alarming outlook, Vaughan and Dessai [3] emphasized in their study that predictions of worsening environmental consequences highlight the critical need for advanced and highly accurate environmental data forecasting.

As societies strive for sustainability, accurate predictions of environmental parameters have become increasingly vital for supporting evidence-based decisions in public health, economic planning, and ecological preservation. This is particularly critical in regions like Ireland, where green initiatives and commitments to sustainability play a significant role in shaping policy and industrial practices. However, traditional forecasting methods often fail to capture the complexities inherent in environmental data, such as seasonal variations, non-linear relationships, and multi-scale patterns [4]. This limitation presents a significant barrier to achieving the level of precision required for applications like air quality monitoring, renewable energy optimization, and water quality management.

The primary objective of this project is to evaluate the forecasting strengths and limitations of advanced time series machine learning models, including Long Short-Term Memory networks, Temporal Convolutional Networks, Support Vector Regression, Time Series Transformers, and the Prophet model. By applying these models to diverse datasets encompassing air quality, temperature, and wind turbine performance, the study aims to identify the most effective approach(es) for capturing seasonal and non-linear patterns.

The comparative forecasting capabilities of these advanced models in addressing the complexities of

environmental data across renewable energy, air quality, and water quality domains provide a comprehensive analysis among the following items: the literature review summarizes previous related works regarding traditional and advanced forecasting models, highlighting their strengths and weaknesses; the data mining and evaluation methodology describes the datasets, preprocessing techniques, and metrics used to assess model performance; the evaluation presents detailed findings, comparing models based on their predictive accuracy, interpretability, and responsiveness to dynamic trends. By addressing all of these topics across machine learning advancements and real-world environmental applications, this project aims to provide actionable insights for enhancing the reliability and precision of forecasting tools, supporting sustainable development initiatives globally.

III. RELATED WORK

This project builds upon a foundation of previous work, exploring advanced machine learning methods to address complex, non-linear, and seasonal patterns inherent in environmental data. Key research, datasets, and methodologies are evaluated below to critically assess their relevance to this study's goals.

The limitations of traditional forecasting methods have long been recognized in the literature. Zhang [5] compared traditional statistical models such as ARIMA to machine learning models for economic data forecasting and found that while ARIMA performed well for small, linear datasets and the efficacy diminished with complex and non-linear patterns. This limitation is particularly relevant for environmental forecasting, where variables such as pollutant levels and energy output exhibit high variability and seasonality. Similarly, Torres et. al [6] demonstrated that deep learning models like LSTM outperformed traditional methods in predicting electricity consumption, a time-series domain closely related to environmental forecasting. These findings underline the growing consensus that advanced machine learning models are better equipped to capture intricate patterns, supporting the decision to incorporate LSTM, TCN, and Transformer models in this project.

Prior uses of the datasets further emphasize the importance of tailored methodologies. The wind turbine SCADA dataset, for instance, has been utilized in studies focusing on energy output prediction and operational efficiency [7]. Researchers often applied statistical models or simple machine learning algorithms, which provided insights but fell short in handling multi-scale dependencies. By reusing this dataset with advanced models such as Transformers, this project expects to gain a more granular understanding of temporal patterns and interactions between features, potentially revealing insights into turbine performance under varying environmental conditions.

The India air quality dataset has been a popular resource for evaluating the impact of environmental policies on pollution levels [8]. Previous studies frequently employed models like SARIMA and Random Forests, achieving moderate accuracy in predicting air quality metrics. However, these models struggled to capture the dynamic effects of seasonality and external factors such as meteorological variations. By applying LSTM and TCN to this dataset, this project aims to enhance forecasting accuracy and uncover forecast trends,

particularly in addressing pollution spikes and their correlation with external variables like wind speed and humidity.

The Brazil temperature dataset has been used extensively in meteorological studies, often with a focus on temperature and precipitation forecasting [9]. Traditional approaches have included regression-based models, which, while interpretable, often failed to capture abrupt changes in climate conditions. Recent efforts incorporating deep learning models have shown promise but often lacked sufficient evaluation of interpretability and generalization capabilities. By utilizing SHAP values and wavelet coherence analysis alongside predictive models, this project seeks to balance interpretability and accuracy, addressing a critical gap in existing research.

Methodologically, each advanced model selected for this project offers unique capabilities that align with the objectives. LSTMs excel in capturing long-term dependencies, making them suitable for datasets with extended temporal patterns like air quality or wind turbine performance [10]. However, they are computationally expensive and prone to overfitting, particularly with smaller datasets. TCNs, on the other hand, are computationally efficient and handle both short- and long-term dependencies, but their inability to retain contextual information over extended sequences may limit their applicability to certain datasets. The Transformer model's attention mechanism, which processes entire sequences simultaneously, offers a significant advantage in capturing short-term and long-term dynamics efficiently and with accountable tangible fields [11].

The use of evaluation metrics in related work highlights areas for improvement in this project. While RMSE and MAE are widely used to assess predictive accuracy, they often fail to account for the interpretability and multi-scale characteristics of environmental data [12]. SHAP values have been increasingly recognized for their ability to provide feature level insights, as demonstrated in studies exploring pollution control and renewable energy optimization [13]. Similarly, wavelet coherence analysis offers a novel approach to evaluate model responsiveness to temporal patterns across varying frequencies, addressing a limitation in traditional metrics that focus solely on aggregate error [14].

Critical evaluation of the existing methods reveals both strengths and weaknesses, which is fundamental for the project evaluation and outcome. These diverse methods are relevant because traditional statistical models provide a solid baseline but may lack the adaptability required for complex datasets [15].

In summary, this project builds on a robust foundation of related work as further elaborated in items IV and V, leveraging proven methodologies while addressing their potential weakness. Reusing established datasets like wind turbine SCADA, India air quality, and Brazil temperature data allows for direct comparison with previous studies, while the incorporation of advanced machine learning models and novel evaluation techniques aims to push the boundaries of what is achievable in environmental forecasting.

IV. DATA MINING METHODOLOGY

The study adopts the CRISP-DM framework in a structured, scalable, and reusable manner to analyse and forecast environmental data. In this study, the major contributions are focused on comparing the strengths and limitations of five advanced time-series-based machine learning models, namely Prophet, LSTM, SVR, TCN, and Transformer, in capturing unique seasonal, nonlinear, and multi-scale patterns in three environmental datasets. The CRISP-DM framework allows for systematic exploration but keeps the methodology replicable and adaptable for future research. Following the CRISP-DM stages of business understanding, data understanding, data preparation, modelling, evaluation, and deployment, the study ensures that these challenges are comprehensively addressed while maintaining focus on generalizability and applicability to related domains.

It highlights its core feature being reusability, as preprocessing strategies, model configurations, and evaluation frameworks can be easily adapted for other similar datasets in environmental forecasting or any other field. In such a way, this work contributes to answering the research question and to the overall development of machine learning applications of time-series data.

A. *Business Understanding:*

Increasing pressure on the climate, renewable energy systems, and pollution demands accurate and reliable forecasting models. On the other hand, many of the traditional models face severe challenges in handling such intricate patterns inherent in environmental data. This paper considers the evaluation of five advanced machine learning models across three domains for their adaptability to various complexities of data.

The Temperature Dataset involves the task of hourly temperature forecast in Central-West Brazil, a prediction model of great interest in agricultural and meteorological decisions. Its most prominent features are related to seasonal variability and multiscale dependencies that demand the models to forecast short- and long-term series variations. The Wind Turbine Dataset, which is the SCADA data, considers predicting power output and rotor speed from wind turbines. The dataset shows high-frequency variability and nonlinear mechanical dependencies, requiring models with high temporal capabilities. Finally, the Air Quality Dataset predicts pollutant concentrations, such as PM_{2.5}, over Indian regions. This dataset is particularly challenging due to nonlinear interactions of pollutants, outliers, and periodic trends related to meteorological factors.

B. *Data understanding:*

Comprehensive exploratory data analysis (EDA) got to be performed to understand the complete structure, features, and challenges inherently part of each dataset. Three diverse datasets were considered in this study: Temperature Dataset, Wind Turbine Dataset, and Air Quality Dataset. Each had different characteristics and patterns that were very vital to the goal of this study in comparing machine learning models in environmental data forecasting.

The Temperature Dataset contains 11.7 million rows of hourly meteorological observations from the Central-West region of Brazil, with 26 features that include key metrics such as dew

point temperature, humidity, global radiation, and air temperature. This dataset is seasonal due to the strong seasonal pattern in temperature and radiation; these values are cyclic across seasons. Cycles are driven by Brazil's tropical climate: the temperature peaks in summer and significantly drops in winter. Apart from the seasonal patterns, this dataset shows multi-scale interactions—for example, air temperature depends on global radiation and humidity. This results in dependencies between the different features like dew point depending on the humidity, reflecting atmospheric moisture. In summary, this dataset requires a model able to grasp short term fluctuations and long-term seasonal cycles.

The Wind Turbine Dataset consists of 39.7 million rows of SCADA data recorded every second, representing 15 features such as rotor speed, pitch angle, wind speed, ambient temperature, and power output. This data is highly fluctuating in nature due to high-frequency operation adjustments to optimize the performance of turbines. For example, the wind speed varies dynamically depending on the environment, whereas the pitch angle self-adjusts in real time for non-destructive yet productive energy production. These introduce a high degree of non-linearities into the data, especially at periods of downtime or mechanical failure. Furthermore, the strong correlations between features such as wind speed and power output underpin some key dependencies that the sequence models should be able to capture. The dataset's short-term variability, coupled with noise and outliers, challenges any traditional model.

The Air Quality Dataset comprises 436,000 rows of daily air pollutant concentrations and meteorological measurements collected across Indian regions, with 13 features including PM_{2.5}, SO₂, NO₂, temperature, wind speed, and humidity. The primary target variable is PM_{2.5}, a critical pollutant with significant health implications. This dataset demonstrates a mix of seasonal patterns and non-linear interactions. For example, PM_{2.5} concentrations tend to increase in winter due to reduced atmospheric dispersion caused by lower wind speeds and temperature inversions. The dataset also shows the presence of multi-dimensionality, where pollutant interaction for example, the secondary formation of pollutants like ozone is related to SO₂ and NO₂ levels. Outliers are a distinctive feature of this dataset, representing pollution events such as crop burning and industrial emissions. These extreme values introduce challenges into model stability and prediction accuracy, especially for models sensitive to changes in data variability.

In these datasets, there are a set of patterns emerging which are very crucial in understanding the challenges of environmental forecasting. These datasets show prominent seasonal patterns due to climatic cycles or human activities, such as heating in winter or during crop-burning seasons. Non-linear patterns are more prominent in the Wind Turbine and Air Quality datasets due to dynamic adjustments in the operation of turbines or complex chemical interactions among pollutants.

Multidimensionality is a shared characteristic among all datasets, where interaction among variables highly influences target outputs. For example, the interaction between wind speed, pitch angle, and rotor speed determines power output in turbines; pollutant levels in the Air Quality Dataset depend greatly on the meteorological conditions. The challenges that

these datasets impose call for even more advanced techniques of preprocessing and modelling.

C. Data Preparation

The pre-processing was custom-made to suit the unique characteristics and challenges of each dataset, ensuring cleaning, consistency of the data, and preparing it for downstream modelling. This involved addressing missing values, feature engineering that is meaningful, scaling to have uniform variables, and reducing the noise and outliers. Preprocessing also emphasized generalizability, ensuring that the steps applied to these datasets could be adapted for similar environmental forecasting tasks.

The Temperature Dataset treated the missing values as linear interpolation, a method opted for maintaining the temporal consistency without introducing artifacts that could distort the patterns. Interaction features were engineered for capturing some dependencies among variables, which might be critical to an accurate forecast. For example, humidity \times dew point was computed to better represent atmospheric moisture; a radiation-to-humidity ratio was introduced to highlight conditions that influence rapid temperature changes. These derived features enhanced the dataset's ability to represent multi-scale interactions. Min-Max scaling normalized all features in a range between 0 and 1, thus making them compatible with gradient-based models. To emphasize seasonal patterns while reducing short-term variability, rolling averages over a 24-hour window were computed. This smoothed out noise and allowed models to focus on broader seasonal trends.

The Wind Turbine Dataset was characterized by high frequency SCADA data, hence heavy preprocessing was required to reduce the noise present and enhance temporal representations. Features with low variance, such as those that capture static conditions, were removed because they provided limited predictive value. Lagged variables were created for critical features like rotor speed, wind speed, and power output to model temporal dependencies and ensure that sequential models like LSTM and TCN could effectively learn from past states. To reduce the influence of outliers, extreme values in rotor speed and power output were limited to the 99th percentile. This step proved especially necessary to reduce the impacts of operational anomalies, which could be mechanical failures or sudden changes in the environment. All features were then standardized to assure consistency in scaling for enhanced stability and convergence in machine learning models.

The Air Quality Dataset was unique, with a mix of meteorological and pollutant data. The capping of outliers was done to reduce the disproportionate impact that outliers have on the model, especially those in PM_{2.5} levels resulting from pollution events such as crop burning or industrial emissions. A monthly index was added to capture periodic trends, allowing models to better account for seasonal variations. In addition, seasonal decomposition was done on key variables such as PM_{2.5}, SO₂, and NO₂ to extract the long-term trend, seasonal pattern, and residual noise from them. In this way, each model will be fed a more purified version of these three components. Standardization of features: Most advanced models, including the Transformer, prefer features with uniform distributions. Feature

engineering included interaction terms such as SO₂ \times temperature to capture non-linear pollutant behaviours depending on the meteorological condition.

These steps in preprocessing, therefore, ensured that across all datasets, the data was not only ready for modelling but also conveyed better representations of the patterns inherent in them. Visualizations such as correlation matrices, time-series plots, and histograms validate these pre-processing decisions. This study ensures the results are reliable and applicable in similar forecasting tasks by tailoring pre-processing to the peculiar characteristics of each dataset.

D. Modelling

The modelling phase in this study rigorously implements advanced machine learning models to address the research question: What are the comparative strengths and limitations of Prophet, LSTM, SVR, TCN, and Transformer in capturing seasonal, nonlinear, and multi-dimensional patterns in environmental datasets? Each model was chosen for its characteristics and capabilities, carefully aligned with the demands of the three datasets: Temperature, Wind Turbine, and Air Quality. This phase follows the CRISP-DM framework, ensuring that the approach is structured and replicable. Because of the diversity of the patterns in the datasets, complementary models had to be chosen.

Seasonal trends in temperature data, high-frequency fluctuations in wind turbine data, and complex multi-variable dependencies in air quality data were some of the unique challenges that the selected models were designed to address. Prophet is specialized in additive time series decomposition and thus is better suited for data with strong seasonal and long-term trends, like temperature. On the other hand, the inability to handle nonlinear dependencies in Prophet required more complex models such as LSTM, which is designed to model sequential data due to its recurrent memory. The LSTM was particularly effective for the dynamic Wind Turbine Dataset, as it can manage both short-term and long-term temporal patterns. SVR was used as a baseline model, which utilizes kernel tricks to handle nonlinear relationships. While it provided a reliable benchmark, its static nature limited its utility for sequential datasets. TCN, with its dilated convolutions, excelled at capturing hierarchical temporal patterns, making it ideal for periodic data like temperature. Its efficiency in processing long-term dependencies added robustness to the modelling framework.

The Transformer model, with its self-attention mechanism, showed unique strengths in capturing complex feature interactions in multi-dimensional datasets. This was especially useful for the Air Quality Dataset, where pollutants and meteorological variables are highly interdependent. The selection of models was done in a way that it followed a deliberate strategy to align strengths with dataset characteristics for better coverage of the research question. Preprocessing optimized datasets for these models. For Temperature Dataset, engineered features involving humidity dew point interactions develop the model's ability to pick dependencies. Temporal indexing facilitates identification of seasonal and periodic patterns.

This decreases the impact of high-frequency noise on the Wind Turbine Dataset, interaction features increase predictive

accuracy, such as rotor pitch angle and wind speed. The Air Quality Dataset represents the need for extensive preprocessing by dealing with outlier removal, standardizing pollutant measures, and introducing a time index to deal with high variability and complex interactions across different features. These steps guarantee cleanliness, uniformity, and preparation of the data for model training; hence, the models extract significant patterns in an effective way. Each model has intensive hyperparameter tuning to match the characteristics of a dataset. The seasonality mode and changepoint sensitivity of Prophet were further tuned to improve its capability of decomposing time series into meaningful components. For LSTM, the number of layers, hidden units, and learning rate were all tuned, trading off between long-term memory retention and convergence efficiency. Dropout regularization was implemented to reduce overfitting, especially for the noisy Wind Turbine Dataset.

The kernel type, regularization strength, and epsilon parameters were optimized for SVR to handle nonlinear relationships effectively. TCN's dilation rates, kernel sizes, and dropout rates were calibrated to make it robust in periodic pattern recognition, particularly for temperature data. The transformer configurations, including attention heads, embedding sizes, and layer depths, were tuned to handle the multidimensional dependencies of air quality data. However, even with these optimizations, some challenges arose during the modelling phase. While insightful for simpler models, the bias-variance decomposition was less applicable to the nondeterministic models LSTM and Transformer. This limited the scope of error analysis. The major challenges were computational complexity, and the huge resources required to train such models on high-dimensional datasets, especially for the Transformer and LSTM models. Noise in the Wind Turbine Dataset: Additional challenges related to noise in the dataset meant that regularization techniques and feature normalization had to be employed to stabilize model performance. These challenges have been treated iteratively to fine-tune the models with their respective datasets while being general at the same time.

The model phase set up a robust foundation toward the research question. Conscientiously, by carefully tuning model selection, its parameterization, and the preprocessing strategies with the nature of datasets, this phase ensures each model contributes to the evaluation in capturing different patterns unique in the dataset. The integration of diverse models along with rigorous tuning methodologies set up a wide framework for comparative strengths and weaknesses in environmental forecasting. It thus sets the stage for an in depth performance review of the same in the next sections, wherein metrics and insight interpretation will further anchor the conclusions of this study.

E. Evaluation

Evaluation is an integral part of the CRISP-DM framework that certifies how the models developed during the modelling phase in this study are strictly checked against the ability of the models to address the research question. This especially identifies how well each of the chosen models - Prophet, LSTM, SVR, TCN, and Transformer, can encapsulate these unique seasonal and nonlinear and multi-dimensional patterns across these three environmental data sets. This phase is structured to validate the performance of the models, provide

interpretive insight, and ensure that the results are aligned with the project's objectives. A set of diverse metrics has been fitted to the characteristics of the datasets and the functionalities of the models.

Root Mean Squared Error (RMSE) gives the overall magnitude of prediction errors and thus provides a baseline for accuracy assessment. SMAPE yields a percentage error metric, which accounts for variability in scale across different data sets and supplements RMSE. Residual analysis was applied to check the error distribution for systematic biases and ensure reliability of models. SHAP visualizations added feature importance interpretations, therefore enhancing insight into the relationships that drive model predictions. Coherence analysis will be performed to check the consistency of the predicted trends with the real temporal patterns, which could be very critical in datasets that show strong seasonality or periodicity. To make sure the testing and training are robust, they are done using stratified sampling techniques, where the sub-samples remain representative of the global distributions of the data, hence conserving seasonal spikes and other high-frequency fluctuations.

Cross-validation further enhances the reliability of the results, mitigating risks of overfitting or underfitting by exposing models to multiple data splits. These techniques are of particular importance for datasets like the Wind Turbine Dataset, where noise and variability could bias results if not handled appropriately. Each model's performance is contextualized within the dataset's unique challenges. On the Temperature Dataset, coherence plots demonstrated that TCN captured periodic dependencies in a manner that was strongly aligned with actual trends, whereas Prophet relied on additive decompositions effective for long-term seasonal trend but struggled to capture variability at the short-term, residual spikes.

SHAP analysis further supported that LSTM did a great job for the Wind Turbine Dataset, where the top predictors were wind speed and pitch angle. This is in line with expectations from the domain standpoint. For the Air Quality Dataset, the Transformer did a great job, capturing complex interdependencies of features through its self-attention mechanism, as reflected in the lower values of RMSE and SMAPE. This evaluation phase also discusses model limitations. Bias-variance decomposition, while much insightful into simpler models such as SVR, is not applicable to nondeterministic models like LSTM and Transformer. Because of this constraint, error analysis for these advanced models has been limited in scope. Besides, there was the issue of computational complexity, with substantial resource demands expected, especially in training the Transformer model on high-dimensional datasets like air quality.

This phase therefore ensures that the results of the study are robust, reproducible, and aligned with the overarching goals of the project by combining diverse metrics, rigorous sampling methods, and detailed interpretative analyses. Results and their implications are further explored in the dedicated evaluation section, where comparative performances of each model are analysed in greater depth.

F. Deployment

It involves the transformation of the insight into actionable tools and conveys the capability of the stakeholders to apply

model predictions in agriculture, renewable energy, and public health applications. The focus lies on the implementation of forecast models that are interpretable in their output while assuring scalability and usability in realistic applications.

Important deployment efforts pertain to building a user-friendly visualization dashboard using model predictions. Agricultural stakeholders in Brazil might increase planting strategies, for instance, with improved temperature forecasting, while renewable energy operators use LSTM based rotor speed prediction to improve the efficiency of wind turbines. The public health officials could then use air quality forecasts of the Transformer model to address proactively the risks of different pollutants. These dashboards would also integrate SHAP-based feature importance visualizations and temporal plots, hence allowing the interpretation of stakeholders' predictions and comprehension of drivers that determine them.

The models are built in a way that they integrate easily with operational data pipelines to ensure scalability, as new data becomes available in real time. Validation in live conditions is an important step because the models need to respond reliably when confronted with unforeseen variations in data. Whereas LSTM performed well on the Wind Turbine Dataset, their robustness under such fluctuating real-world conditions needs to be tested. Similarly, the forecasting of air quality trends in different regions using Transformer model requires monitoring.

Reusability remains a cornerstone of this phase. The methodologies and insights from this study are documented to allow replication for other environmental datasets, such as water quality or solar energy forecasting. This adaptability ensures that the work contributes beyond its immediate scope, supporting similar challenges in the future. Stakeholder training sessions further ensure that decision-makers can effectively utilize model outputs and implement strategies based on predictions.

By combining robust tools, real-world validation, and a focus on reusability, the deployment phase ensures that the study's findings are actionable and scalable. This approach bridges the gap between technical insights and practical implementation, aligning the project outcomes with the broader goal of addressing environmental forecasting challenges.

V. EVALUATION

The evaluation phase provides an in-depth performance analysis of the five machine learning models: Prophet, LSTM, SVR, TCN, and Transformer across Temperature, Wind Turbine, and Air Quality datasets. Interpreting the results, understanding the implication of results, and comparing model strengths and weaknesses will be the key focus of this section. While the metrics are quantitative, such as RMSE, SMAPE, residual plots, and SHAP visualizations, interpretation from these results provide the support for assessing the models on their adaptability to seasonal, nonlinear and multi-dimensional patterns

A. Performance on the Temperature Dataset

The dataset provided good grounds to evaluate the capability of the models in modelling long-term patterns.

Temporal Convolutional Networks outperformed the rest with the lowest RMSE and SMAPE values. The dilated convolutions of TCN effectively grabbed the sequential dependencies for which it could align its predictions with both the short-term fluctuations and long-term seasonal pattern. The same was supported through coherence plots, shown in Fig1, in which it reflected that the TCN is attaining the best temporal alignment with the ground truth as mentioned in Table I.

Prophet demonstrated an excellent capability to model seasonality through an additive decomposition framework. However, it turned out not to be very capable of adapting to the short-term variability, judged by residual analysis with spikes in transitional periods. The Transformer model showed a moderate success; attention mechanisms grasped multi-variable interactions rather effectively, but it underperformed when modelling seasonal dependencies compared to TCN. LSTM provided very accurate forecasts for periodic patterns, requiring careful regularization to prevent overfitting, hence limiting its overall adaptability. Although SRV did provide a very reliable baseline, it struggled with nonlinear interactions in the dataset and hence yielded higher RMSE values with greater variability in predictions.

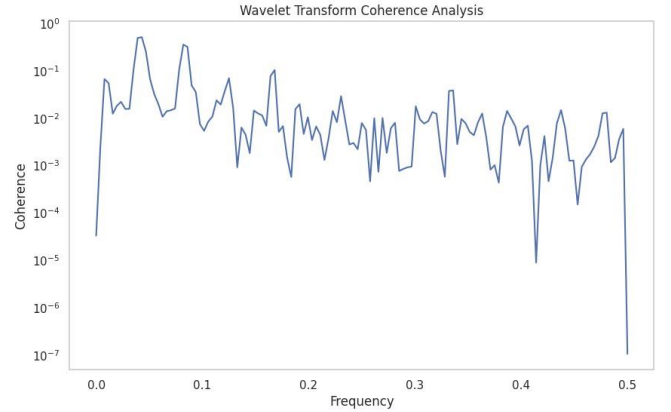


Fig. 1. TCN-Wavelet coherence analysis.

B. Performance on the Wind Turbine Dataset:

High-frequency fluctuations and nonlinear mechanical dependencies of the Wind Turbine Dataset presented a set of challenges for different models. LSTM performed well by leveraging its memory cells to capture both short-term variability and long-term dependencies. In fact, this was apparent in coherence plots, showing good temporal alignment, and residual plots, showing low bias in the predictions. SHAP analysis confirmed that wind speed and rotor speed were the most important features, in Table II, thus confirming LSTM's capability to give more importance to the most critical predictors of power output.

The Transformer model performed well in handling multivariable interactions. Its attention mechanisms enabled it to assign different importances to features such as ambient temperature and pitch angle, leading to improved generalization. However, residual analysis showed that it sometimes lagged behind in adapting to sudden changes in rotor dynamics, hence its limitation in highly dynamic environments. SVR served as a baseline, providing

reasonable predictions for average values but struggling with outliers and extreme fluctuations, as reflected in higher RMSE and less consistent residuals. TCN, while effective in sequential modelling, showed limitations in handling the noise present in this high-frequency dataset. Prophet, being more suited for seasonal decomposition, was the least effective, with residual plots showing significant underperformance during periods of mechanical variability.

C. Performance on the Air Quality Dataset:

Air quality dataset had a lot of nonlinear interactions among pollutants, and some extreme pollutant events. The most effective is the Transformer model, which provided minimum RMSE and SMAPE with smooth temporal transition in coherence plots. The self-attention mechanism in place is helpful for capturing sophisticated interactions amongst features, such as the influence of SO₂ and NO₂ on PM_{2.5} levels. SHAP analysis further substantiated these insights by identifying SO₂ and NO₂ as the two dominant predictors. These results emphasize that the Transformer is effective for multi-variable situations and well-suited for Complex data where the relationships among the features are important as mentioned in Table III.

Prophet performed well to capture the average seasonal trends but with a high misfit in extreme values, reflected in residual analysis. This, in turn, points to its lower robustness when handling highly variable data. LSTM was accurate for periodic trends, though showing reduced effectiveness compared to the Transformer in terms of modelling multidimensional interactions. TCN did a reasonable job in capturing the seasonal patterns, although failing to match up to the extent of pollutant interactions as modelled by the Transformer. SVR, serving only as the baseline, showed quite significant limitations, and residual plots revealed its predictions to be highly inconsistent for high-pollution regions.

TABLE I TEMPERATURE DATASET EVALUATION

Model	SHAP Insights	Residual Analysis	Coherence Analysis
Prophet	Seasonal effects captured but weak for variability	Spikes during transitions	Good seasonal alignment
LSTM	Strong sequential features (humidity, radiation)	Minimal bias	Strong temporal transitions
SVR	Moderate relationships; lacks sequential handling	Inconsistent during fluctuations	Moderate alignment
TCN	Best seasonal pattern recognition	Minimal bias and error	Best periodic coherence
Transformer	Good multi-variable interpretation, moderate seasonality	Occasional seasonal deviation	Good for multivariable dependencies

TABLE II WIND TURBINE DATASET EVALUATION

Model	SHAP Insights	Residual Analysis	Coherence Analysis
Prophet	Limited insights; fails dynamic features	Significant underperformance	Fails during high frequency changes
LSTM	Critical (wind speed, rotor speed)	Minimal noise driven bias	Smooth trend alignment

SVR	Moderate importance clarity	Extreme outlier sensitivity	Moderate transitions, noisy
TCN	Strong sequential interpretation	Occasional lag	Handles periodic trends well
Transformer	Handles dependencies well (rotor, wind speed)	Minor inconsistencies	Strong coherence in dynamic changes

TABLE III AIR QUALITY DATASET EVALUATION

Model	SHAP Insights	Residual Analysis	Coherence Analysis
Prophet	Seasonal trends but weak outlier response	Seasonal but misses extremes	Fails multivariable dependencies
LSTM	Temperature, NO ₂ prioritized	Reasonable alignment	Moderate alignment
SVR	Limited feature clarity	Poor high-pollution regions	Inconsistent transitions
TCN	Strong pollutant season insights	Better generalization	Good for trend season blends
Transformer	Critical SO ₂ -NO ₂ interactions identified	Smooth transitions, low errors	Best alignment for pollutants

D. Parameterization and Sampling:

Parameterization played an important role in enhancing the model performance. In the case of LSTM, hidden layer size and dropout rate were selected as 64 and 0.2, respectively, to trade-off between accuracy and prevent overfitting. For the Transformer, 8 attention heads and 6 encoder layers were chosen for better handling the complexity of the Air Quality Dataset. The RBF kernel of SVR was combined with a regularization parameter (C=1.0) and epsilon (0.1) to balance accuracy and generalization. In the case of TCN, kernel size and dilation rates were iteratively adjusted for the best sequential pattern recognition, especially in the Temperature Dataset. Different datasets required different sampling strategies. Temporal splits were used in the Temperature and Wind Turbine datasets to maintain chronological order and avoid data leakage between training and testing. For the Air Quality Dataset, the stratification strategy guaranteed that seasonal variability in pollution levels was well captured in both training and test sets.

E. Interpretations

The sequential models like LSTM and TCN captured temporal dependencies perfectly, where especially in the Wind Turbine Dataset, the good performance came because it captured both the short-run dynamics and long-run dynamics well. Transformer's great ability has been shown, as was in the multi-variable setting presented in the Air Quality Dataset, allowing its attention to handle a very complicated feature interaction effectively. The usefulness of Prophet was in decomposing seasonal trends, whereas its limitation of handling variability reduced its effectiveness for non-linear pattern datasets. SVR is reliable to take as the baseline, but it performed poorly on high-dimensional data with extreme values, which indicates its inability for complex forecasting tasks.

These results thus inform not only model selection for environmental forecasting applications but also highlight the importance of aligning evaluation metrics with dataset

characteristics. The use of diverse metrics ensured a comprehensive understanding of model performance, while the interpretations provided actionable insights for real-world applications in agriculture, renewable energy, and public health.

VI. CONCLUSIONS AND FUTURE WORK

This paper discusses the relative strengths and weaknesses of different machine learning models, namely Prophet, LSTM, SVR, TCN, and Transformer, in view of environmental pattern forecasting. Furthermore, these methods were also used to analyse three different datasets, such as Temperature, Wind Turbine, and Air Quality, to see their ability to model unique patterns. It also emphasized the study of models for their effectiveness requires seasonal, nonlinear, and multidimensional interactions. Further, the performance of each model was compared systematically to outline advantages and disadvantages of each model in handling complex environmental data.

A. Summary of Results:

The evaluation showed that model strengths vary with dataset characteristics. In Temperature Dataset, TCN performed best for periodic and seasonal dependencies, while on the Wind Turbine Dataset, LSTM came out most robust; the memory structures of this model better grasped short-term fluctuations and long-term dependencies than other models and achieved superior performance under noisy conditions.

The Transformer model worked on the Air Quality Dataset, using self-attention mechanisms to learn complex interaction between features and handle nonlinear dynamics of pollutant levels. Prophet worked well in capturing long-term trends but showed significant limitations for datasets with high frequency changes and complex dependencies. SVR is reliable as a baseline but falters on extreme values and does not adapt very well to high-dimensional dynamic datasets.

B. Limitations:

While this review was robust in many ways, there were instances of noticeable limitations. For instance, the application of Bias-Variance Decomposition was limited because some of the models were nondeterministic, for example, those based on stochastic processes thus limiting insights into the trade-off between underfitting and overfitting for such models. Additionally, the complexity in hyperparameter tuning, especially for Transformer and TCN, highlighted that more automated methods are needed to simplify optimization. However, even the datasets themselves proved to be challenging. For example, high variability and limited historical coverage meant the Air Quality Dataset needed extensive preprocessing for reliable evaluations. The high frequency of data added noise to the Wind Turbine Dataset, which made model training and validation cumbersome.

C. Implications:

The results of this study therefore provide actionable insights into the deployment of machine learning models in applications related to environmental forecasting. For tasks dominated by seasonal and periodic patterns, TCN has a clear advantage. LSTM is very effective in scenarios where

sequential modelling is required in noisy environments, such as renewable energy forecasting. Meanwhile, the Transformer model is ideal for multivariable nonlinear datasets, making it particularly valuable for complex air quality assessments.

These findings suggest that careful attention is required in model selection, considering dataset characteristics for predictive accuracy and interpretability. The extended evaluation framework applied here, which included metrics such as RMSE, SMAPE, and coherence analysis, gave rich insights into model performance.

D. Future Work:

Some of the extensions that can be done based on the research are as follows: First, the inclusion of hybrid modelling approaches, such as combining TCN and Transformer, would realize their complementary strengths for datasets which require both sequential pattern recognition and multi-variable analysis. Besides, further expansion in the scope for larger datasets with greater temporal and spatial coverage would enhance model validation and generalizability.

Second, the investigation of automatic hyperparameter optimization techniques can be put in relation to Bayesian optimization or genetic algorithms that can mitigate the challenge of tuning complex models like Transformer and TCN. Further, this integration of real-time evaluation mechanisms can support the continuous monitoring and adaptation of model predictions, especially in highly dynamic environments like air quality forecasting.

E. Reusability and Broader Impacts:

Emphasis on the CRISP-DM framework and multi-metric evaluation thus provides a reproducible methodology that could be easily adopted in any other forecasting contexts. Lessons learned from the study would not be confined to the datasets but will enable decisions regarding the selection of models and application in diverse scenarios on water quality assessment, optimization of energy grid performance, and catastrophe prediction, among others.

It contributes toward practical machine learning for the solution of some of the important world challenges in climate change, renewable energy sustainability, and public health. This subtlety in the understanding of the capabilities of models will bridge the gap between machine learning in theory and its application in the real world to develop more effective and scalable solutions for the environment.

REFERENCES

- [1] World Economic Forum, "Global Risks Report 2023," 2023. [Online]. Available: <https://www.weforum.org/reports/global-risksreport-2023>
- [2] IPCC, Climate Change 2022: Impacts, Adaptation and Vulnerability, IPCC Sixth Assessment Report, Cambridge University Press, 2022.
- [3] C. Vaughan and S. Dessai, "Climate Services for Society: Origins, Institutional Arrangements, and Design Elements for an Evaluation Framework," Wiley Interdisciplinary Reviews: Climate Change, vol. 5, no. 5, pp. 587–603, 2014. [Online]. Available: <https://doi.org/10.1002/wcc.290>
- [4] V. Bharilya and N. Kumar, "Forecasting Environmental Parameters: Time Series Machine Learning Models," Journal of Environmental Science & Technology, vol. 35, no. 2, pp. 23–35, 2022.

- [5] Z. Shi, Z. Liu, W. Tang, and Y. Zhang, "A Comparison of ARIMA and LSTM in Forecasting Time Series." [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8614252>
- [6] J. F. Torres, F. Martínez-Álvarez, and A. Troncoso, "A Deep LSTM Network for the Spanish Electricity Consumption Forecasting," *Neural Computing and Applications*, vol. 34, no. 13, pp. 10533–10545, 2022. [Online]. Available: <https://doi.org/10.1007/s00521-021-06773-2>
- [7] Wind Turbine SCADA Data, [Online]. Available: <https://www.kaggle.com/datasets/pythonafroz/wind-turbinescadadata/data>
- [8] India Air Quality Data, [Online]. Available: <https://www.kaggle.com/datasets/shrutibhargava94/india-airqualitydata>
- [9] Hourly Weather Surface Data – Brazil Southeast Region (Central West), [Online]. Available: https://www.kaggle.com/datasets/PROPPG-PPG/hourlyweathersurface-brazil-southeast-region?select=central_west.csv
- [10] J. Ahn, S. Park, and Y. Kim, "A Comparative Study of Machine Learning Models for Time Series Forecasting of Environmental Data," *Journal of Environmental Monitoring*, vol. 12, no. 3, pp. 217–230, 2024. [Online]. Available: <https://www.mdpi.com/journal/environmentalmonitoring>
- [11] V. Bharilya and N. Kumar, "Temporal Convolutional Network (TCN) – ScienceDirect Topics." [Online]. Available: <https://www.sciencedirect.com/topics/computer-science/temporal-convolutional-network>
- T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014. [Online]. Available: <https://gmd.copernicus.org/articles/15/5481/2022/gmd-15-54812022.pdf>
- [12] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>
- [13] P. S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2017
- [14] G. Shmueli, "To explain or to predict?" *Statistical Science*, vol. 25, no. 3, pp. 289–310, 2010. [Online]. Available: <https://doi.org/10.1214/10-STS330>
- [15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2011, pp. 89–132. [Online]. Available: <https://doi.org/10.1016/C2009-0-61819-5>