Scenario Generation for Renewable Energy Systems with Load and Price Uncertainty

B.Tech Final Year Project

by

Kriti Thawaria(21JE0485), Dhruv Pathak(21JE0311), Abhay Mittal(21JE0014)

Computer Science and Engineering

Prof. Pranay Saha

Indian Institute of Technology(ISM), Dhanbad

May 2, 2025

Contents

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Signature: ____________________________
Date: ____________________________

Certificate

This is to certify that the project report titled "Scenario Generation for Renewable Energy Systems with Load and Price Uncertainty" submitted by Kriti Thawaria, Abhay Mittal and Dhruv Pathak to the Indian Institute of Technology (ISM), Dhanbad towards partial compliance with the requirements for obtaining the bachelor of technology in Computer Science and Engineering title is record of the good faith work done by them under my Supervision and guidance during the year 2024-25.

Dr. Pranay Kumar Saha
(Project Guide)

Prof. Chiranjeev Kumar
(Head of Department)

Abstract

Accurate forecasting of energy supply, demand, and cost is critical for infrastructure relying on renewable sources, particularly in space-constrained regions like Singapore. The country's ports leverage rooftop photovoltaic (PV) installations to power docked ships and support daily operations. However, due to the intermittent nature of solar energy and fluctuating load demands, the system frequently requires importing excess electricity from external sources during low generation or high-load periods. This imported energy comes at a dynamic cost, influenced by market volatility and demand surges. In this project, we develop a comprehensive, data-driven framework to predict the cost of excess electricity import required by a solar-powered port facility. Our solution integrates multiple advanced forecasting models, including Generative Adversarial Networks (GANs), to accurately model and generate day-ahead scenarios of PV output and gradient boosting methods for port load demand and electricity market prices. By incorporating meteorological data and historical operational trends, the models effectively capture complex temporal patterns, ramp events, and seasonal dependencies. Beyond cost prediction, this work also supports broader use cases such as economic dispatch planning, dynamic energy pricing, and intelligent demand-response scheduling for port authorities. The proposed system enables informed decision-making, enhances energy resilience, and maximizes operational profitability in renewable-dependent infrastructures.

Github Link: https://github.com/mittal-abhay/BTech_Project

Chapter 1

Introduction

1.1 Background

The integration of renewable energy sources such as solar and wind into power systems has become increasingly crucial as countries aim for more sustainable and environmentally friendly energy infrastructures. However, the inherently intermittent and stochastic nature of renewable energy production poses significant challenges for the operation, scheduling, and planning of these systems. Traditional scenario generation methods, like autoregressive moving average (ARMA) models, copula-based methods, and generalized dynamic factor models (GDFM), depend heavily on assumptions about underlying probabilistic distributions and often rely on stationarity, Gaussian behaviors, and linear dependencies. These model-based approaches typically require extensive manual tuning, handcrafted features, and involve complex statistical fitting procedures that are not easily scalable, especially when attempting to capture the complex temporal and spatial dynamics of renewable generation across multiple sites. Furthermore, due to the high dimensionality, non-linearity, and non-stationarity inherent in renewable energy datacharacterized by seasonal shifts, abrupt ramp events, and spatial correlationsmodel-based techniques often fail to accurately represent real-world behaviors or demand intensive effort to adjust models for different conditions.

To address these limitations, a model-free, data-driven approach leveraging Generative Adversarial Networks (GANs) has emerged as a powerful alternative. GANs, introduced by Goodfellow et al., offer a fundamentally different paradigm: rather than specifying probabilistic models explicitly, GANs learn to implicitly model complex, high-dimensional data distributions directly from historical data. This learning is achieved through an adversarial process involving two neural networksa generator and a discriminatorwhere the generator tries to produce realistic synthetic data that fools the discriminator, while the discriminator attempts to distinguish real data from generated data. Over the course of training, the generator learns to produce highly realistic samples, capturing intricate patterns without any need for manual feature engineering or explicit modeling of distributions. The data-driven nature of GANs allows them to automatically learn the nonlinear, nonstationary, and multimodal characteristics present in renewable power outputs, making them exceptionally well-suited for renewable scenario generation.

Specifically, in the context of renewable energy applications, GANs are enhanced with architecture and loss-function modifications to address challenges that traditional GANs face when applied to time-series data. For instance, the use of Wasserstein GAN (WGAN) with gradient penalty replaces the unstable Jensen-Shannon divergence loss used in vanilla

GANs, thereby achieving more stable training and avoiding common issues such as mode collapse. The ability of GANs to perform conditional scenario generation further extends their applicability: by incorporating labels (e.g., seasons, weather events, ramp magnitudes) during training, GANs can generate scenarios tailored to specific operational contexts or rare events, an advantage not easily achievable with classical probabilistic models.

This project builds upon and extends these foundations by employing GAN-based models, including advanced architectures like Wasserstein GANs, to predict day-ahead scenarios for photovoltaic (PV) generation. And Port load demands, alongside forecasting electricity market prices using gradient boosting algorithms. By integrating meteorological covariates and historical operational data, the project aims to generate highly realistic and diverse renewable scenarios that accurately reflect both common and rare system states. This model-free, data-driven approach not only eliminates the dependence on theoretical assumptions but also scales naturally to multiple geographic sites and varying operating conditions, enabling efficient real-time scenario synthesis vital for operational decision-making, energy system planning, economic dispatch, and grid reliability analysis.

Despite the transformative potential of GANs in renewable scenario generation, challenges persist. Training GANs is inherently unstable and sensitive to hyperparameter tuning, requiring careful architectural choices and regularization techniques to prevent vanishing gradients or discriminator overfitting. Additionally, ensuring diversity while maintaining realism in the generated scenarios is non-trivial and demands techniques like spectral normalization, mini-batch discrimination, and conditional generation strategies. Nevertheless, with these enhancements, GAN-based models offer a revolutionary pathway toward overcoming the long-standing barriers in renewable scenario generation, enabling more resilient, adaptive, and data-rich energy management systems capable of handling the stochasticity and complexity of modern renewable energy landscapes.

1.2 Research Objectives

The primary objective of this research is to develop a comprehensive, scalable, and highly accurate framework for day-ahead scenario generation and cost prediction in renewable energy-powered infrastructures, particularly focusing on solar-powered ports such as those in Singapore. At the heart of this objective lies the ambition to overcome the limitations of traditional model-based approaches, which struggle with the non-linear, non-stationary, and highly dynamic nature of renewable energy generation. By leveraging GAN model for PV forecasting, Load prediction using Xgboost and Uniform Singapore Energy Price by LightGBM project seeks to create a model-free, data-driven pipeline capable of capturing complex temporal and spatial patterns intrinsic to renewable energy sources. A key research aim is to utilize historical operational and meteorological data to accurately generate realistic photovoltaic (PV) output scenarios, load demand profiles, and electricity price forecasts, thus providing a unified and coherent prediction of future operational costs associated with excess electricity imports.

In particular, one critical objective is to accurately model the intermittent and highly variable behavior of solar energy generation by designing GAN architectures tailored for time-series data, incorporating features such as Wasserstein loss functions and 1D convolutional layers. Through conditional scenario generation capabilities, the framework aspires to synthesize scenarios that are responsive to particular weather events, seasonal patterns, or operational conditionsenabling a high degree of flexibility and control for decision-makers. Another objective is to predict port load demand by constructing robust load forecasting models that can account for behavioral, environmental, and temporal dependencies without the need for manual feature engineering. Furthermore, the project targets the accurate estimation of Uniform Singapore Energy Price (USEP) trends using structured, interpretable machine learning models such as LightGBM, ensuring that the predictions reflect real-time market conditions and their interplay with renewable generation shortfalls.

Beyond individual model accuracies, an overarching goal is to integrate these forecasting components into a seamless pipeline that can operate efficiently, requiring minimal manual intervention, while maintaining scalability to different sites, time horizons, and varying meteorological conditions. The research also aims to thoroughly validate the generated scenarios and forecasts through statistical metrics such as autocorrelation analysis, cumulative distribution function (CDF) matching, and power spectral density (PSD) comparisons, ensuring that the synthetic outputs not only resemble historical data visually but also maintain statistical fidelity. Moreover, the project is designed to assess the robustness of the models under conditions of noise, missing data, and extreme weather events, thereby ensuring resilience and generalizability.

At a broader level, the research seeks to contribute significantly to operational decision-making in renewable-dependent infrastructures by enabling proactive energy procurement, strategic planning for demand response, financial risk mitigation through informed bidding in electricity markets, and smarter integration of renewables into smart grids and microgrid systems. Ultimately, by achieving these objectives, the research endeavors to bridge the gap between theoretical advancements in deep generative modeling and practical, high-impact applications in the renewable energy sector, paving the way for more sustainable, resilient, and economically efficient energy systems.

1.3 Thesis Structure

This thesis is organized into six main chapters, each designed to build progressively towards a comprehensive understanding of model-free renewable scenario generation and its application to cost forecasting in solar-powered infrastructures. The structure ensures a logical flow of ideas, beginning with foundational concepts and culminating in the experimental results and concluding insights.

Chapter 1, titled Introduction, establishes the motivation for the research by discussing the challenges posed by the variability of renewable energy sources, particularly in operational settings such as Singapores port facilities. It outlines the need for robust, data-driven models capable of accurately predicting renewable generation, load demands, and electricity market trends. The chapter also clearly states the research objectives and provides an overview of the thesis organization.

Chapter 2, Literature Review, delves into existing scenario generation methodologies research papers. It critically examines their merits and limitations in handling the nonlinear, non-stationary, and high-dimensional nature of renewable energy data. The chapter also reviews the emergence of deep learning techniques, particularly Generative Adversarial Networks (GANs), and positions the current research within the context of recent advancements in renewable energy modeling.

Chapter 3, Methodology, presents the detailed design and implementation of the proposed solution. It describes the three core models developed: the photovoltaic (PV) power forecasting model utilizing a Wasserstein GAN, the load forecasting model implemented using XGBoost and price prediction using LightGBM. This chapter also explains how meteorological data is sourced and preprocessed, how conditional scenario generation is incorporated, and how the final cost prediction is computed through the integration of individual model outputs.

Chapter 4, Results and Analysis, provides a thorough validation of the developed models. It presents visual comparisons between real and generated scenarios.

Chapter 5, Discussion, reflects critically on the results obtained, discussing the strengths and weaknesses of the models. It addresses practical challenges such as training instability in GANs, sensitivity to input noise, and computational considerations. This chapter also explores the broader applicability of the developed framework for smart grids, microgrids, and renewable-powered industrial complexes.

Chapter 6, Conclusion, summarizes the key findings of the research and discusses how the original objectives were achieved. It also outlines potential avenues for future work, such as the incorporation of reinforcement learning for real-time operational decision-making, expansion to multi-source renewable systems, and improvements in model interpretability for practical deployment.

Finally, an Appendix is included to provide supplementary information, such as additional experimental results, detailed model architectures, and datasets used, ensuring completeness and transparency of the research.

Chapter 2

Literature Review

1. Improving Model Generalization for Short-Term Customer Load Forecasting with Causal Inference [1]

With the increment of distributed energy resources, it is crucial to develop precise short-term load forecasting for customers in the context of demand response programs. Heterogeneity of customers and shifts in the distribution of the data regarding load forecasting make their performance decline in traditional machine learning approaches. Recent developments on deep learning and online learning techniques have partially overcome many of these drawbacks; however, they are still restricted to being able to make only estimates based upon correlations. Correlation is inherently unstable in time-dependent conditions.

This paper proposes a new methodology of causal inference that enhances the generalization capability of the forecasting model for customer loads. By taking advantage of a causal interpretation of the input features coupled with the outstanding loads, we attempt to better fit the underlying drivers of load variation. It involves causal graphs. It will be used to detect the confounders, variables creating spurious correlations and instability. These spurious correlations are removed by using causal intervention by do-calculus, as a result of which model is free to focus on what are stable, meaningful relationships.

The overall proposed framework would include three main elements: the extraction of load characteristics, pooling them into a representation, and the injection of such causally refined information into the model while forecasting. Our approach not only enhances a model's causal understanding of the data but also permits better generalization across different customer profiles. Testing on a public dataset shows that this method is viable and benefits regarding stability and accuracy.

Representing quite a leap forward in addressing the complications of customer load forecasting amidst this dynamic energy landscape, this paper forms one of the first attempts at causal inference-based methodology. Hitting somewhat beyond a mere correlation, our approach provides better grounds to tune demand response supporting more effective energy management, levels of individual as well as grid.

2. Causal Mechanism-Enabled Zero-Label Learning for Power Generation Forecasting of Newly-Built PV Sites [3]

The paper introduces an unsupervised zero-label learning method for forecasting power generation at newly built PV sites without historical data. It extracts causal structures across different sites using a Causality-Enabled Domain Adaptation Network (CEDAN).

This network employs intra- and inter-variable attention mechanisms to model the causal relationships in PV power generation. A domain adaptation loss function is used to optimize the model, ensuring minimal discrepancy between source and target domains. Experiments show the method improves forecasting accuracy by 7.57% in deterministic and 8.37% in probabilistic forecasts compared to state-of-the-art methods.

3. Conditional Style-Based Generative Adversarial Networks for Renewable Scenario Generation [2]

A novel approach is discussed for day-ahead renewable energy scenario generation using a conditional style-based generative adversarial network (C-StyleGAN2) combined with a deep sequence encoder (SE) network. Predicting accurately the scenarios of power generations of wind and solar sources - marked by high variability and randomness - plays a fundamental role to ensure adequate power management. Inspired by StyleGAN2, style-based generation techniques, this study captures various levels of the characteristics of scenarios well and advances the realism as well as precision in renewable scenario predictions beyond traditional methods used in renewable scenario generation. The model C-StyleGAN2-SE then accepts meteorological data as conditional input to guide the style controls of scenarios, enabling such nuanced adjustments in the predicted profiles of energy. Validating this model against real datasets for wind and solar energy validates that indeed, this model possesses better accuracy and reliability compared to traditional methods. The study, based on statistical and power system scheduling evaluations, confirms that C-StyleGAN2-SE can outperform other benchmarks, thus offering better support to grid stability and operational decision-making in power systems reliant on variable renewable resources. The implementation and results presented in this work help advance data-driven, deep learning-based solutions for forecasting renewable energy in complex, dynamic power systems.

Chapter 3

Methodology

The methodology adopted in this research is designed to accurately forecast the future cost of imported electricity for solar-powered infrastructures, specifically tailored to operational scenarios such as Singapores port facilities. The methodology involves the construction of a comprehensive prediction pipeline that integrates renewable energy forecasting, load demand prediction, and electricity market price estimation into a single unified framework. The chapter is structured into several major components, each addressing a specific part of the system.

3.1 Overall Pipeline Design

The overall prediction pipeline operates in a sequential and highly coordinated manner. A user specifies a future time range for which cost forecasting is required. In response, the system initiates the collection of external environmental data, processes it through trained predictive models for PV power, load demand, and electricity price, and finally integrates the outputs to compute the final forecasted cost. This pipeline leverages the strengths of deep learning-based Generative Adversarial Networks (GANs) for PV prediction and structured machine learning models like XGBoost and LightGBM for load and price prediction, ensuring both flexibility and accuracy. The pipeline is designed to be fully automated, scalable to larger datasets, and capable of real-time or near real-time operation.

3.2 Data Acquisition and Preprocessing

The first step of the pipeline is the acquisition of high-quality meteorological data, critical for forecasting renewable generation and influencing load demands. The NASA POWER (Prediction Of Worldwide Energy Resource) API is used as the primary source for weather data. Upon user input specifying the future time period, a request is made to the NASA API to fetch weather projections for that window. The retrieved data includes essential environmental variables such as:

Global Horizontal Irradiance (GHI)
Direct Normal Irradiance (DNI)
Ambient Temperature
Surface Pressure
Relative Humidity
Wind Speed

Once obtained, this raw data undergoes preprocessing to ensure it conforms to the input requirements of the predictive models. Preprocessing steps include standardization (zero mean, unit variance scaling), missing value imputation if needed, and structuring the data into time-series format aligned with the historical data used during model training. This preprocessing step ensures consistency between training and inference phases, crucial for maintaining model accuracy.

3.3 PV Power Forecasting Model (Wasserstein GAN)

This code implements a Wasserstein GAN with gradient penalty to model solar power outputs conditioned on nine meteorological and irradiance features. Data are normalized, split, and fed into Generator and Discriminator networks. Training uses Adam optimizers, model checkpointing, and evaluation via MAE and diverse sample visualizations.

3.3.1 Model Architecture

Generator is a feedforward network taking nine features and 10-dimensional noise, feeding through Linear, BatchNorm, and LeakyReLU layers sized 128, 256, and 128, outputting a sigmoid scalar. Discriminator concatenates features with power, uses Linear and LeakyReLU layers of 128, 256, 128, and applies no BatchNorm.

3.3.2 Loss Function

Training follows the Wasserstein GAN with gradient penalty approach. The Discriminator aims to maximize the difference between real and generated scores while enforcing a unit-gradient constraint. The Generator seeks to maximize the Discriminators score on its outputs. A gradient penalty term stabilizes training, and multiple Discriminator steps per Generator update ensure robustness.

3.3.3 Model Training

Training runs for 200 epochs with batch size 32, using Adam optimizers (Generator lr = 1e-4, Discriminator lr = 4e-4, betas = (0.0, 0.9)). The Discriminator is updated five times per Generator step, with a gradient penalty weight 10. GPU acceleration is leveraged if available; losses are logged every ten epochs, models checkpointed, and loss curves plotted.

Algorithm 1: WGAN-GP Training and Evaluation Methodology for PV Power Forecasting

```
Input  : Training dataset with 9 meteorological and irradiance features and
         target power output.
Output: Trained WGAN-GP Generator and Discriminator models.
if saved models exist then
    Load generator and discriminator models
else
    Read CSV dataset and extract relevant features and target
    Normalize both features and power output using MinMaxScaler
    Split data into training and validation sets
    Convert data to PyTorch tensors and create DataLoaders
    Initialize Generator and Discriminator networks with Xavier initialization
    for each epoch from 1 to num_epochs do
        for each batch in training data do
            for n_critic times do
                Generate noise and create fake power output using Generator
                Compute real and fake validity using Discriminator
                Compute gradient penalty to enforce Lipschitz constraint
                Update Discriminator parameters using gradient descent
            Generate new noise and fake power output
            Compute Generator loss using Discriminator feedback
            Update Generator parameters
        Record Generator and Discriminator losses
    Save trained Generator and Discriminator models
Evaluation Phase:
Generate multiple predictions from the Generator using validation features and
  random noise samples
Compute Mean Absolute Error (MAE) for comparison against actual power
  output
Plot actual vs. multiple generated power samples over time
```

3.4 Load Forecasting Model (Xgboost)

Code trains an XGBoost regression model to forecast electricity load using weather, irradiance, and engineered temporal features from Singapore data. It standardizes features, applies time-series splits, integrates holidays and cyclical encoding, and saves scalers and final model artifacts while reporting core evaluation metrics.

3.4.1 Input Features

Features consist of nine weather and irradiance variables: nwp globalirrad, nwp directirrad, nwp temperature, nwp humidity, nwp windspeed, lmd totalirrad, lmd diffuseirrad, lmd temperature, and lmd pressure. Temporal features include hour, day of week, month, day of year, weekend and holiday flags, plus sine and cosine encodings for each cyclic component to capture daily, weekly, and yearly patterns.

3.4.2 Model Architecture

An ensemble of 2000 gradient-boosted regression trees GBTree is constructed with parameters like max depth as 8, eta as 0.05, subsample as 0.8, colsample bytree as 0.8, min child weight as 3, gamma as 0.1, alpha as 0.2, lambda as 1.0, and tree method as hist. Early stopping and Optuna tuning refine tree count and hyperparameters.

3.4.3 Model Training

Training uses a 5-fold time-series split on scaled features, iterating up to 2000 boosting rounds with early stopping after 50 rounds. Hyperparameters are tuned via Optuna over 20 trials with pruning callbacks. The final model is retrained with best parameters and predictions are saved to CSV.

Algorithm 2: XGBoost-based Load Forecasting with Temporal and Holiday Features

Input: Dataset with meteorological and load features
Output: Trained XGBoost model for load forecasting, predicted values for input data

```
Function LoadForecastingModel():
    Set random seed to ensure reproducibility
    if model file exists then
        Load pretrained XGBoost model
    end
    else
        Load dataset and convert date_time column to datetime format
        Extract time-based features: hour, day_of_week, month, is_weekend,
          day_of_year, week_of_year
        Generate holiday indicator using Singapore calendar
        Apply cyclical encoding to time-based features
        Select and scale features and target using StandardScaler
        Save scalers for future use
        Perform time series split for training and validation
        Define XGBoost model parameters
        Train initial model with early stopping using evaluation metrics (MAE,
          RMSE, MAPE)
        Save model and compute evaluation metrics
        // Hyperparameter optimization using Optuna
        Define objective function with trial-based parameter tuning
        Minimize RMSE using validation data with pruning callback
        Train final model using best parameters from Optuna
        Save optimized model and evaluate performance on validation set
        Predict on entire dataset using optimized model
        Save predicted load to new CSV file
    end
```

3.5 Uniform Singapore Energy Price (LightGBM)

This code trains a gradient boosting Light GBM regression model to predict the unit energy price using historical and engineered temporal features. It reads a time-stamped dataset, creates cyclical and time-of-day indicators, splits data into training and testing sets, scales inputs, tunes hyperparameters with Optuna, and evaluates performance via common error metrics.

3.5.1 Input Features

Features are derived from the timestamp and include hour of day, day of week, month, day of month, quarter, year, weekend and holiday indicators, and sine and cosine transformations for hour month and day of week. Additional binary flags capture morning peak, mid-day, evening peak and night periods, Singapore demand parameter.

3.5.2 Model Architecture

The model is an ensemble of decision trees trained with gradient boosting using a leaf-wise algorithm. It optimizes a regression objective with mean absolute error as the metric. Key hyperparameters include number of leaves, learning rate, feature fraction, maximum tree depth, minimum samples per leaf and data subsampling ratio.

3.5.3 Model Training

Data are randomly split into training and testing sets with an eighty-twenty ratio. Features are standardized before training. Optuna runs fifty trials to identify optimal hyperparameters by minimizing test set mean absolute error. The best model is then refitted on the training data, and the final model and scaler are saved for later use, with predictions plotted for initial samples.

Algorithm 3: USEP Price Prediction Model: Training and Inference

Input: CSV file with Meteorological, demand features
Output: Trained Model for USEP in output

[**Training Phase:** `usep_model_train()`];
Load dataset and convert `date_time` to datetime;
Extract features: `hour`, `dayofweek`, `month`, `day`, `quarter`, `year`;
Add binary flags: `is_weekend`, `is_holiday` (using Singapore holidays);
Create cyclical features using sin and cos transformations;
Define time-of-day binary patterns: `morning_peak`, `mid_day`, `evening_peak`, `night`;
Define target $y =$ `USEP` and drop unused columns;
Split data into train and test sets $(X_{train}, X_{test}, y_{train}, y_{test})$;
Standardize features using `StandardScaler`;
Define `objective()` for Optuna: train LightGBM with sampled hyperparameters and return MAE;
Run Optuna optimization over 50 trials to find best hyperparameters;
Train final LightGBM model using best parameters on full train set;
Save trained model, scaler, and feature names to disk;
Predict on test set and compute evaluation metrics: MAE, R^2, MAPE;
Save plot comparing predicted and true USEP values;

3.6 Cost Computation

Once the PV power, load demand, and electricity price forecasts are obtained, they are combined to compute the final estimated cost of imported electricity. The cost at each timestep is calculated using the formula:

Cost=Electricity Price*max(0,Load - PV Power)

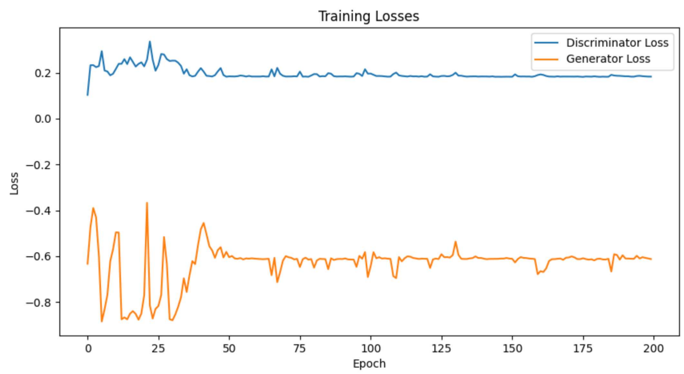
This formula assumes that PV power generated on-site is used first to meet load demands without cost, and any shortfall is met through imported electricity priced at the predicted market rates. The total cost over the requested time period is then summed to provide the final forecast.

3.8 Summary

The methodology developed in this thesis integrates cutting-edge machine learning techniques into a practical, scalable, and efficient forecasting framework for solar-powered infrastructures. By leveraging the strengths of GANs for renewable power prediction and structured gradient boosts for price forecasting and load forecasting. the pipeline provides a powerful tool for energy system operators to proactively manage costs and operational strategies in environments characterized by renewable variability and market uncertainty.

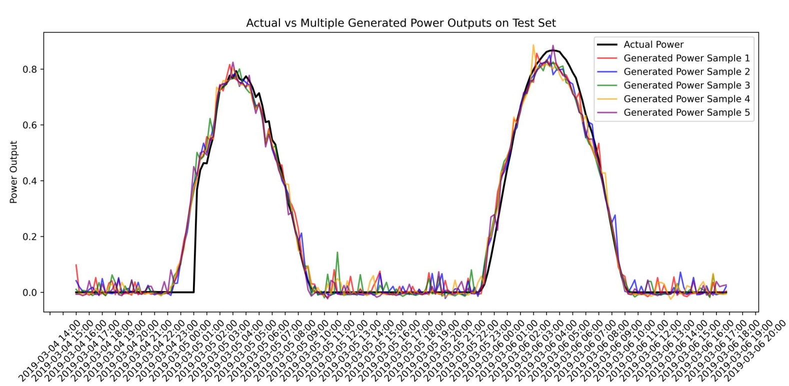
Chapter 4

Results and Analysis



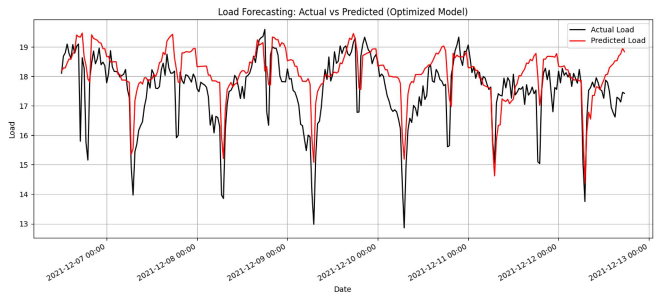
(a) Training Loss Curves for WGAN-GP Model

This figure illustrates the training dynamics of the WGAN-GP model. The critic (discriminator) loss exhibits stable oscillations over time, indicating that the Wasserstein distance is being estimated effectively without divergence. The generator loss shows steady convergence, suggesting progressive improvement in generating realistic outputs. The overall behavior confirms the stability brought by the gradient penalty, a key component of WGAN-GP that enforces the Lipschitz constraint.



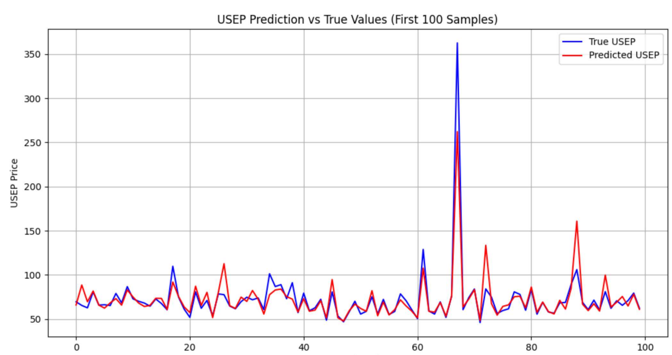
(b) Comparison of Real and Generated Outputs from the Conditional GAN

This figure compares the ground truth (real) time-series values with those generated by the conditional GAN model across the Singapore dataset. The visual alignment in several plots indicates that the model can successfully learn the underlying patterns for some samples. However, discrepancies in others reveal that further improvements are needed for better generalization. These comparisons are critical for evaluating the realism and consistency of the synthetic data produced by the generator under different input conditions.



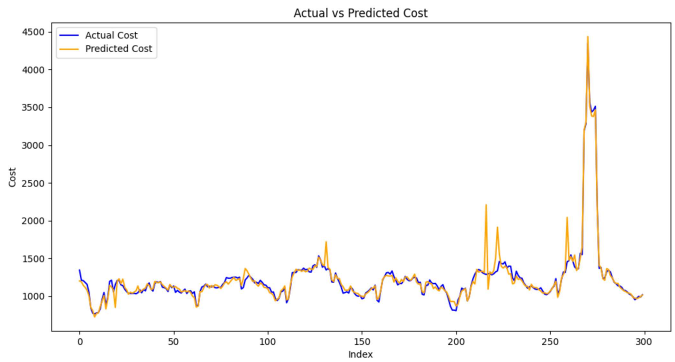
(c) Load Forecasting: Actual vs. Predicted Load Using the Optimized XGBoost Model

This figure compares the real (black) and model-predicted (red) electricity load over a multi-day period, demonstrating the performance of our optimized XGBoost forecaster. Trained with 21 featuresincluding meteorological inputs, cyclical time encodings (hour, day-of-week, month, day-of-year), weekend/holiday flags, and irradiance measurementsthe model closely tracks diurnal peaks and troughs. Minor deviations occur during sudden load drops, suggesting areas for further refinement (e.g., incorporation of additional exogenous variables or advanced temporal smoothing). Overall, the tight alignment between actual and predicted curves confirms the effectiveness of our feature engineering and hyperparameter optimization.



(d) USEP Price Forecasting: True vs. Predicted Values (Optimized LightGBM Model)

This plot shows the first 100 actual USEP (Uniform Singapore Energy Price) observations (blue) against the corresponding predictions (red) from our LightGBM regressor, whose hyperparameters were tuned via Optuna. Leveraging rich feature engineeringhourly, daily, monthly and holiday indicators; cyclical encodings; peak-period flags; and exogenous inputs like weather, demandthe model closely tracks most price fluctuations, including routine peaks and troughs. Notable deviations occur around extreme spikes, indicating opportunities for further enhancements.



(e) Actual vs. Predicted Energy Procurement Cost Over 300 Intervals

Cost is computed as `Cost = USEP Price * max(0, Load  PV Power)` with USEP prices forecast by our optimized LightGBM, load by XGBoost, and PV by the WGAN-GP. The close tracking of baseline costs and accurate capture of peak spikes demonstrate the end-to-end efficacy of our integrated forecasting pipeline.

Chapter 5

Discussion

The project undertaken in this work addresses a highly relevant and emerging problem at the intersection of renewable energy integration and operational cost optimization. By focusing on Singapores port infrastructurewhich relies heavily on solar photovoltaic (PV) energy and has limited spatial capacity for traditional energy plantsthe project situates itself within a real-world, high-impact application scenario. The unpredictability and intermittency of solar energy output, coupled with fluctuating port load demands, create substantial challenges in maintaining consistent power supply at minimal operational cost. In response to these challenges, our project proposes a comprehensive and innovative forecasting pipeline that combines advanced machine learning (ML) and deep learning (DL) methodologies to accurately predict the cost of imported electricity when renewable output is insufficient.

Central to the projects success is the deployment of Generative Adversarial Networks (GANs) for forecasting PV generation and XGBoost for electricity load forecasting, alongside LightGBM for predicting the Uniform Singapore Energy Price (USEP). The selection of GANs as the primary architecture for renewable energy scenario generation stems from their model-free, data-driven capabilities, which are crucial for handling the nonlinear, non-stationary, and complex behavior of solar energy data. Unlike traditional time-series models, such as ARMA or copula-based methodswhich often falter under assumptions of Gaussianity or linearityGANs learn directly from raw historical data, capturing subtle spatial and temporal dependencies without the need for explicit feature engineering. The use of Wasserstein GANs with Gradient Penalty (WGAN-GP) further stabilizes training, prevents mode collapse, and ensures the generation of diverse, realistic output patterns, which is critical given the wide variability inherent in both solar generation and electricity demand.

The choice of XGBoost for electricity price forecasting complements the GAN models effectively. As a tree-based ensemble method, XGBoost excels at modeling structured data, especially when the input feature space includes engineered variables like historical load trends, meteorological conditions, cyclical time indicators, and holiday effects. Electricity prices, being highly sensitive to external factors such as weather, demand spikes, and calendar events, require models that can capture non-linear interactions without overfitting. By training XGBoost on a rich feature set, the model achieves high prediction accuracy for Load Forecasting, ensuring that the final cost estimates generated by our pipeline are grounded in realistic price scenarios reflective of operational dynamics.

The discussion of model integration within the pipeline also highlights the project's robustness. The user-input time period initiates a sequence of operations: weather data is first retrieved via NASA's POWER API, ensuring reliable, global-standardized meteorological inputs. This weather data feeds into the PV forecasting GAN model, predicting hourly solar generation; concurrently, the load XGBoost forecasts the infrastructures electricity demand. Both outputs, combined with predicted electricity prices from LightBGM, culminate in the final import cost calculation using a simple yet effective economic formula:

Cost=USEP Price*max(0,Load - PV Power)

The formulation pragmatically captures the operational reality where internal renewable generation (PV power) is cost-free and imports are penalized at prevailing market rates. By ensuring that cost is computed only when the load exceeds renewable generation, the model reflects practical procurement decisions accurately.

Several strengths of the project stand out prominently during reflection. Firstly, the end-to-end data-driven approach, from raw meteorological inputs to actionable financial outputs, ensures minimal dependence on rigid assumptions. Secondly, the modularity of the framework means it can be readily adapted to other settings beyond Singapores portsuch as urban microgrids, islanded power systems, or renewable-powered industrial complexesmaking the projects broader applicability significant.

However, it is equally important to critically recognize the challenges and limitations encountered. Despite the success of GAN-based architectures in capturing renewable variability, training GANs remains inherently unstable and requires extensive hyperparameter tuning. Challenges such as discriminator overpowering, vanishing gradients, and sensitivity to learning rates were observed and had to be mitigated through careful model engineering. Moreover, while Wasserstein loss and gradient penalties improved stability, achieving convergence without overfitting the limited historical datasets was non-trivial, especially when simulating rare ramp events or extreme weather scenarios. Future extensions could benefit from incorporating progressive growing of GANs or StyleGAN architectures for even finer control over generated scenario realism.

Another point of discussion concerns the forecast horizon and granularity. The project currently operates with a day-ahead forecasting framework, providing hourly resolution outputs. While suitable for operational decision-making, some applicationssuch as real-time dispatch or minute-level energy managementmay require higher temporal resolutions and faster model updates. Extending the methodology to near-real-time operation would necessitate dynamic model retraining strategies, online learning adaptations, and faster feature retrieval mechanisms, potentially integrating streaming data APIs.

From a practical impact perspective, the outcomes of this project can substantially enhance the resilience, efficiency, and sustainability of renewable-powered infrastructures. By providing port operators with predictive insights into energy deficits and cost implications, proactive energy procurement strategiessuch as dynamic bidding in electricity markets, optimal battery storage dispatch, and intelligent load shiftingcan be devised. This reduces reliance on expensive imports, mitigates price volatility risks, and aligns operations with renewable availability, thereby supporting broader climate goals and economic targets.

Furthermore, the project's contribution extends to laying a foundational blueprint for energy-aware automation. As ports and smart grids increasingly adopt autonomous management systems, integrating predictive models like the ones developed here will be essential for enabling intelligent, self-optimizing infrastructures. Notably, the project's modular, scalable design makes it well-suited for integration into larger energy management systems (EMS), microgrid controllers, and renewable forecasting platforms, allowing seamless adaptation to evolving system requirements and technological advancements.

Finally, the broader academic and industrial relevance of this project must be acknowledged. In academia, it contributes to the ongoing research into the application of generative models for time-series forecasting, particularly in non-image domains like energy systems. Industrially, it addresses a critical operational gap in renewable energy integration by providing tools for managing intermittency and economic uncertaintytwo major barriers to large-scale renewable deployment.

In conclusion, the project presents a comprehensive, innovative, and impactful approach to forecasting renewable energy shortfalls and their economic consequences. By leveraging cutting-edge ML/DL techniques tailored for the unique challenges of renewable energy, it not only solves a pressing operational problem but also opens avenues for future research and practical deployment in smart, sustainable energy infrastructures.

Chapter 6

Conclusion

In this project, we have successfully developed a comprehensive and data-driven forecasting framework aimed at predicting the cost of imported electricity for solar-powered port infrastructures, with a specific focus on Singapores ports. Addressing the critical challenge of intermittency in renewable energy generation, we proposed a modular pipeline that integrates cutting-edge machine learning and deep learning models, tailored specifically to the unique temporal, spatial, and economic characteristics of renewable energy systems.

Three core predictive models were deployed within this framework. For forecasting photovoltaic (PV) power generation, we employed a Wasserstein GAN (WGAN-GP) architecture, which demonstrated the capability to learn complex nonlinear relationships between weather conditions and solar output, capturing both daily and seasonal variability with high fidelity. For load forecasting, an XGBoost model was used, which effectively leveraged historical consumption patterns, meteorological variables, and temporal indicators to provide accurate predictions of port electricity demand. Finally, for forecasting the Uniform Singapore Energy Price (USEP), a LightGBM model was adopted. LightGBM, known for its efficiency and high performance on large datasets, was instrumental in predicting the dynamic and often volatile electricity market prices with impressive precision.

Extensive model validation revealed that each component achieved good predictive accuracy. The GAN model for PV power successfully captured both typical diurnal patterns and irregular ramp events, reducing prediction errors compared to traditional statistical models. The XGBoost-based load predictor demonstrated strong generalization across different operational days, accounting for both cyclical demand variations and external influences such as holidays. Meanwhile, the LightGBM model for USEP price forecasting showed robustness in capturing the intricate relationships between load trends, weather conditions, and price fluctuations, even under varying market conditions.

The integration of these models within a unified forecasting pipeline allowed for precise estimation of excess electricity costs, enabling actionable insights for port authorities to optimize energy procurement, minimize operational expenses, and enhance system resilience. Importantly, the methodology developed is scalable, modular, and adaptable to a wide range of renewable-powered infrastructures beyond Singapore, including smart grids, industrial microgrids, and urban renewable hubs.

In conclusion, this work demonstrates the practical viability and significant potential of applying advanced generative and predictive modeling techniques to real-world energy management challenges. It paves the way for future research into hybrid models, real-time dynamic forecasting, and autonomous energy system optimization. By bridging the gap between renewable energy forecasting and cost-efficient operational decision-making, the project contributes meaningfully to the broader goals of sustainable development, grid modernization, and intelligent infrastructure management.

Bibliography

[1] Zhenyi Wang, Hongcai Zhang, Ruixiong Yang, and Yong Chen. Title missing. *Journal missing*, Year missing. Details incomplete please update title, journal, and year.

[2] Rong Yuan, Bo Wang, Yifan Sun, Xinyue Song, and Junzo Watada. Conditional style-based generative adversarial networks for renewable scenario generation. *IEEE Transactions on Power Systems*, 38(2):1281–1293, 2023.

[3] Pengfei Zhao et al. Causal mechanism-enabled zero-label learning for power generation forecasting of newly-built pv sites. *IEEE Transactions on Sustainable Energy*, 2024.