Forecasting of Renewable Energy Using Statistics and Machine Learning

THESIS

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

SARITA
(2018PHXF0439P)

Under the Supervision of

DR. SUMANTA PASARI

and Co-Supervision of

PROF. RAKHEE

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
JANUARY, 2024

"*In a gentle way, you can shake the world.* "

–MAHATMA GANDHI

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis titled **"Forecasting of Renewable Energy Using Statistics and Machine Learning"** submitted by **Ms. Sarita**, ID No. **2018PHXF0439P** for the award of Ph.D. of the Institute embodies original work done by her under our supervision.

Signature of the Supervisor	Signature of the Co-supervisor
Name: **DR. SUMANTA PASARI**	Name: **PROF. RAKHEE**
Designation: **Associate Professor**	Designation: **Professor**
Department of Mathematics	Department of Mathematics
BITS Pilani, Pilani Campus	BITS Pilani, Pilani Campus

Date: **December 11 , 2023**

Acknowledgments

First and foremost, I thank God for giving me all sorts of strength needed to pursue this journey. I express my sincere gratitude and indebtedness to my Ph.D. supervisor, **Dr. Sumanta Pasari**, for giving me a wonderful opportunity to work on this topic. I thank him for continuous motivation, valuable guidance, and constant support during my research work. His invariable encouragement, prolific discussions, and valuable suggestions at different stages were really a great inspiration for me.

I also express my sincere gratitude to my co-supervisor **Prof. Rakhee** for her guidance and encouragements. Apart from her excellent supervision, her moral support made the task enjoyable and rewarding.

I feel fortunate to have my Doctoral Advisory Committee (DAC) members as **Prof. Rajesh Kumar** and **Prof. Anirudh Singh Rana** for always giving me time for all the official presentations and their valuable feedback and suggestions.

It is an honor for me to be a doctoral student in the Department of Mathematics, BITS Pilani. I owe my sincere gratitude to **Prof. Devendra Kumar**, Head, Department of Mathematics and Ex-Head, **Prof. B. K. Sharma** who provided the opportunity to work in the Department of Mathematics and in finalizing this work within time. I would like to acknowledge all the faculty members and staff of the Mathematics Department for demonstrating genuine interest and enthusiasm in their teaching and relentless support.

With great reverence, I express my gratitude to **Prof. V. Ramgopal Rao** (Vice Chancellor), **Prof. Sudhirkumar Barai** (Director), **Prof. M.B. Srinivas** (Dean, AGSRD), and **Prof. Shamik Chakraborty** (Associate Dean, AGSRD) for providing necessary facilities.

I thank **Prof. Gordan Reikard** from U.S. Cellular for providing research ideas and fruitful discussions on dynamic models for renewable energy forecasting. I would also like to thank the anonymous reviewers for taking their time and efforts to provide valuable comments and suggestions on my manuscripts.

I am grateful to all my family members who have constantly provided me moral and emotional support throughout this journey. I extend my sincere gratitude to my parents **(Mr. Hari Prakash and Mrs. Indrawati Devi)** for believing in me more than myself. Their constant support and tremendous understanding kept me motivated throughout my Ph.D. and my life, in general. I think the words are not enough to describe the quantum of their support. I am immensely grateful to my husband **Dr. Ranbir Singh** who has been supporting me along the way. His constant love and care empowered me to accomplish my thesis and also ignited me to take more challenges in life. I am so lucky to have him by my side and I cannot thank him enough for his time and attention. My special love goes to my son **Aarush** for being the joy

of my heart and light of my world. He has been my constant source of affection, happiness, and motivation. I am also grateful to my friends who have supported me during the journey. I would like to thank my batchmates **Mr. Chandan**, **Ms. Riya Jain**, **Mr. Sajan**, **Ms. Shilpa**, **Dr. Neha**, and my other colleagues of the Department for supporting and encouraging me during my Ph.D. tenure in direct or indirect ways. I also thank my juniors **Mr. Himanshu Verma**, **Ms. Sakshi Shukla**, **Ms. Sonu Devi**, and **Ms. Sharmila Devi** for assistance I received from them on a number of occasions.

I am thankful to the **University Grant Commission (UGC)**, New Delhi for providing me financial assistance as Junior and Senior Research Fellowships during my tenure at BITS Pilani as a Ph.D. research scholar.

Place: BITS Pilani
Date: August 18, 2023

Sarita
(Department of Mathematics)

Abstract

Even today, most countries across the world heavily depend on fossil fuels (oil, coal, and natural gas) as primary sources of energy to power their economies. However, in the era of rapid urbanization and industrial revolution, it would be folly to place total reliance on these resources as they are not only exhaustive but also cause astounding amount of environmental pollution. Unlike fossil energy sources, renewable energy comes from renewable resources which are naturally replenished, abundant, and environment friendly. Among all types of renewable energy sources, wind and solar energy are considered to be the fastest growing and major contributors. Despite their high potential, variable and intermittent behavior of their resources is a crucial issue in energy harvesting, energy economics, sustainable resource management, supply-demand analysis, and electric grid operations. Therefore, a reliable forecasting of renewable resources, such as wind speed and solar irradiance is the need of the hour.

An effective forecasting of wind speed and global horizontal irradiance (GHI) depends on two important factors: (i) the time horizon and (ii) the technique of forecasting. With respect to the time horizon, immediate short term, short term, medium term, and long term forecasting are the four general categories. Similarly, based on the adopted methodology, forecasting techniques are often classified into four types: (i) physical methods, (ii) time series methods, (iii) machine learning methods, and (iv) hybrid techniques. These approaches of forecasting may further be divided into two broader groups: theory-driven physical models and data-driven stochastic models. The physical models aim to describe the dynamics of the underlying renewable energy process, whereas the stochastic models aim to extract useful information in terms of observed variables and hidden variables. Due to the involvement of stringent boundary conditions, subjective knowledge of the hyper-parameters, and an ideal physical setup, the applications of theory-driven physical models are quite limited. On the other hand, data-intensive stochastic models, purely based on historical data, are capable of dealing with simultaneous linear and non-linear data patterns and inherent diversity, resulting in operational data analysis and subsequent forecasting in a rapid manner.

In light of the above scenario, the present study concentrates on hourly, daily, weekly, and monthly forecasting of wind speed and GHI using time series, machine learning, and hybrid methods. For illustration, the experimental data (2000–2014) of wind speed and solar irradiation are obtained from four different locations in India, one each from the state of Rajasthan, Gujarat, Karnataka, and Telangana. The consequential research design in the thesis is as follows. Chapter 1 provides an overview and rationale of the thesis along with the principal objective and scope of the thesis. Chapter 2 carries out data preprocessing in terms of preliminary data exploration and distribution fitting. Chapter 3, Chapter 4, and Chapter 5 are in a sense the

core of the thesis, for which the obtained results are directly relevant to the renewable energy community. Chapter 3 addresses the applicability and efficacy of various time series models in renewable energy forecasting. Chapter 4 deals with the implementation of several popular machine learning methods. Based on the complementary benefits of standalone time series and standalone machine learning models, Chapter 5 proposes a few hybrid models and compares their effectiveness in renewable energy forecasting. Finally, Chapter 6 summarizes the major contributions of the thesis along with some future research directions.

To achieve the research goal, the first task is to perform data exploration and distribution fitting to wind speed and GHI data. In data exploration, the associated time series decomposition reveals a long term seasonal pattern in the data with no prominent trend. The stationarity test at $\alpha = 5\%$ suggests overall stationarity of data. In distribution fitting, five reference probability distributions, namely exponential, gamma, lognormal, Weibull, and exponentiated Weibull are studied. Based on statistical inference using the maximum likelihood estimation and K-S goodness of fit, the exponentiated Weibull turns out to be the most suitable distribution for both wind speed and GHI data across time and space. These results of data analysis bear significant importance in determining optimal parameters of time series and machine learning models.

After performing the preliminary data exploration, the next task is to assess the efficacy of several time series methods in renewable energy forecasting. As the data patterns exhibit seasonality, stationarity, and randomness, five time-series models, namely autoregressive (AR), moving average (MA), autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), and window-sliding ARIMA (WS-ARIMA) are implemented. A grid search method is adopted to obtain the optimum values of model parameters. A comparative performance of the studied models is assessed through RMSE values. In addition, residual analysis is performed as a post-processing step to examine any systematic bias in the implemented models. The experimental results reveal that (i) for monthly forecasting, the SARIMA model has the best performance; (ii) for daily and weekly forecasting, the WS-ARIMA method consistently outperforms the conventional time series methods with significant improvement in the forecasts across time and space; and (iii) for hourly forecasting, the WS-ARIMA and ARIMA show comparable performance based on three years (2012–2014) of data. These results emphasize that the inclusion of sliding windows in conventional ARIMA model significantly improves forecasting performance.

In the next step, various machine learning methods, such as support vector regression (SVR), artificial neural network (ANN), long short term memory (LSTM), bidirectional LSTM (BiLSTM), encoder-decoder LSTM, attention layer LSTM, and convolutional neural network (CNN) are implemented for wind speed and GHI forecasting. Each of these machine learning

models is controlled by a variety of variables, including the type and number of hidden layers, the activation function, the optimization technique, the loss function, the epochs, and the learning rate. From a range of potential parameter values, the best parameters are identified in a way that the overall error is minimized. Based on the RMSE values, it was observed that no single model yields consistently best performance across time and space. This behavior may be attributed to the model sensitivity in learning from varying data size, inherent stochasticity, seasonal and non-linear variability in the dataset.

After implementing both time series and machine learning models, the next task is to assess the effectiveness of hybrid models in renewable energy forecasting. For this, three hybrid models, namely ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM are developed. The associated methodology has two connected steps: first, implementing an ARIMA model to analyze the linear part of the data and then, implementing a neural network (ANN, CNN, and LSTM) to model the residuals of the ARIMA. Based on the RMSE values corresponding to daily, weekly and monthly forecasting, the ARIMA-ANN model has the best performance in wind speed forecasting across time and space, whereas in case of GHI forecasting, the ARIMA-ANN has the best performance except one location. Furthermore, the performance of hybrid models and corresponding standalone models is compared through RMSE values and residual characteristics. Except a few instances, none of the hybrid models enables lesser RMSE than that of the standalone models. Nevertheless, the hybrid models offer a significant improvement towards the Gaussian behavior of residuals. Finally, some comprehensive performance statistics of all fifteen studied models are provided in the thesis.

In summary, the present study has significantly improved our understanding of renewable energy process, purely dictated by the observed wind speed and GHI data characteristics. The analysis presented here inevitably encourages scientists and engineers from government and industry to join hands for enhanced renewable energy forecasting, prominent site identification, operational management, and social policymaking towards the noble endeavor of "Sustainable Development Goals".

Contents

List of Figures

List of Tables

Dedicated To

My Father, for encouraging me to commence this journey

My Husband, for making possible to complete it

My Son, for his unconditional love throughout

Table 0.1: List of abbreviations

Sr. No.	Abbrevation	Stands for
1	ACF	Auto Correlation Function
2	AIC	Akaike Information Criterion
3	ANFIS	Adaptive Network based Fuzzy Inference System
4	AR	Autoregressive
5	ARMA	Autoregressive Moving Average
6	ARIMA	Autoregressive Integrated Moving Average
7	ARIMAX	ARIMA Model with Exogenous Variable
8	ARFIMA	Autoregressive Fractionally Integrated Moving Average
9	ANN	Artificial Neural Network
10	BIC	Bayesian Information Criterion
11	CNN	Convolutional Neural Network
12	DHI	Diffuse Horizontal Irradiance
13	DLM	Dynamic Linear Model
14	DNI	Direct Normal Irradiance
15	ECDF	Empirical Cumulative Distribution Function
16	EMD	Empirical Mode Decomposition
17	EOF	Empirical Orthogonal Function
18	GHI	Global Horizontal Irradiance
19	k-NN	k-Nearest Neighbors
20	LSTM	Long Short Term Memory
21	MA	Moving Average
22	MAE	Mean Absolute Error
23	MAPE	Mean Absolute Percentage Error
24	MSE	Mean Square Error
25	NWP	Numerical Weather Predictions
26	PACF	Partial Auto Correlation Function
27	RMSE	Root Mean Square Error
28	RNN	Recurrent Neural Network
29	SARIMA	Seasonal ARIMA
30	SDD	Similar Day Detection
31	SSA	Singular Spectral Analysis
32	SVM	Support Vector Machines
33	SVR	Support Vector Regression
34	TFN	Triangular Fuzzy Numbers
35	WS-ARIMA	Window-Sliding ARIMA

Chapter 1

Introduction

"*The most reliable way to forecast the future is to try to understand the present.*"

– JOHN NAISBITT

This chapter presents a general overview and motivation of the thesis work. It discusses the current status of renewable energy worldwide and particularly for the Indian region. It also includes the main research objective and the scope of the present work. A chapter-wise road map of the thesis is presented towards the end of this chapter.

Contents

1.1 Overview and Motivation

In the era of industrial revolution, the conventional resources, such as oil, coal, and natural gas have proven to be highly effective drivers of economic progression (Figure 1.1). However, these resources are not only limited but also create pollution. Emissions from such resources have been damaging our environment, contributing to the global warming and consequent climate changes. The Inter-Government Panel on Climate Change (IGPCC) has been raising the issue of gaseous emissions from fossil fuels and increasing CO_2 levels. On the other hand, the demand for energy is increasing rapidly across the world due to urbanization, industrial revolution, and increasing population. As a consequence, the world is currently experiencing its biggest energy crisis on record [40]. Prices for fossil fuels are already spiking, leading to the threat of energy poverty for billions of people. This also poses a challenge for the environment and the climate. In light of this, increased usage of carbon sequestration measures and clean and green renewable energy has been persistently suggested [156].

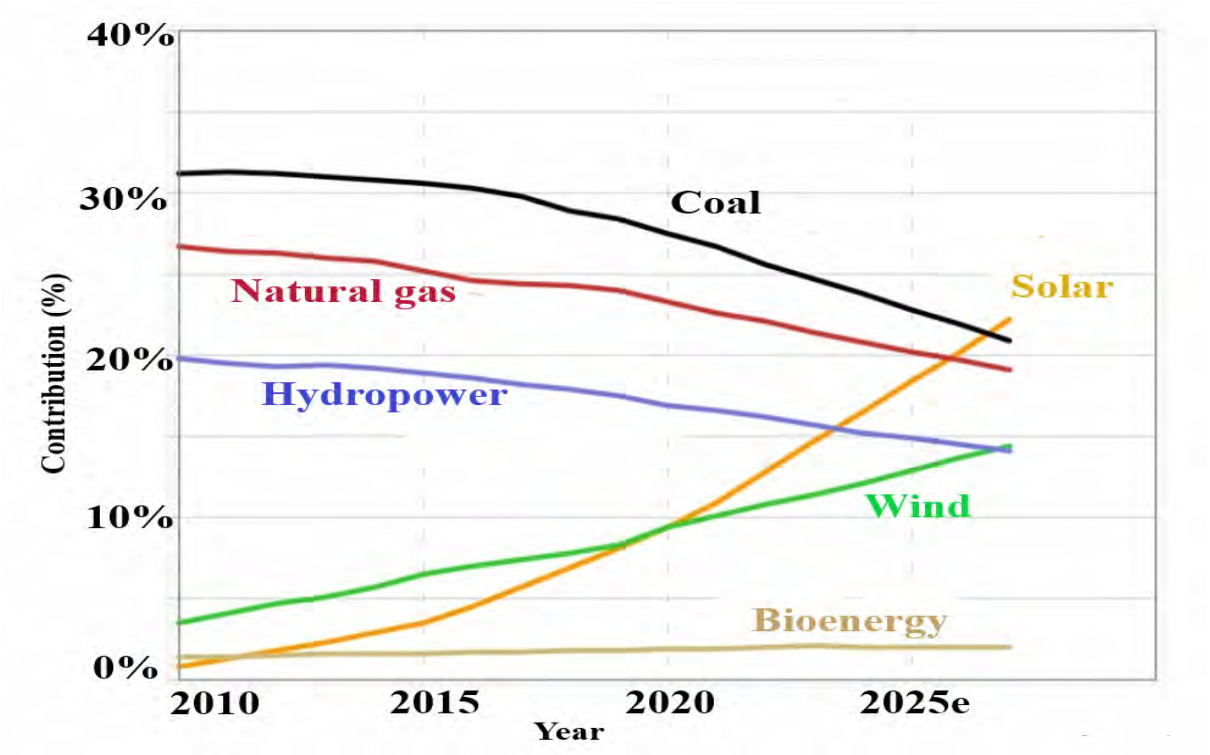


Fig. 1.1: Share of different resources in world energy consumption (values corresponding to 2023–2027 are IEA projections) [27].

Unlike fossil energy sources, renewable energy is the energy collected from renewable resources which are naturally replenished. Some major resources of renewable energy include

sunlight, wind, rain, tides, waves, geothermal heat, and biomass. The most popular renewable energy sources are of following types.

- **Solar energy**– energy harnessed from the solar irradiance using solar panels.
- **Wind energy**– energy harnessed from the wind using wind turbines.
- **Hydro-energy**– energy harnessed from moving water, such as rivers or tides, using hydroelectric turbines.
- **Geothermal energy**– energy harnessed from the heat within the Earth using geothermal power plants.
- **Biomass energy**– energy harnessed from organic matter, such as wood, crops, or waste, using biomass power plants.

Renewable energy is one of the major contributors to overall power generation worldwide. Most of the renewable energy sources are sustainable while some are not, e.g., biomass resources. Renewable energy has multiple advantages over the traditional fossil fuels as they are inexhaustible, abundant, and safer to use. Therefore, investing in renewable energy can help mitigate the energy crisis. For instance, a recent report [155] demonstrated that investing in renewable infrastructure makes more sense from both climate and financial perspective. In addition, the Renewables 2022 Global Status Report highlights the importance of renewable based economy and society, including the ability to achieve more diversified and inclusive energy governance through localised energy generation and value chains. Thus, renewables are the most affordable energy sources to improve resilience and to support decarbonisation. It can help reduce greenhouse gas emissions, improve air quality, diversify energy supply, and create jobs and economic opportunities in the industrial revolution and modern economic growth.

Though the usage of renewable energy is much older than a century, last few decades have witnessed a tremendous shift of attention. From 2011 to 2021, the share of renewables in global electricity generation grew from 20% to 29%, in which solar and wind power contribute two-thirds of the overall renewables' growth. Therefore, wind and solar power have the major contribution in overall renewable energy generation. The top five countries with the highest solar and wind electricity generation in 2019 were: China (721 TWh), United States (433 TWh), Germany (228 TWh), India (156 TWh), and Brazil (126 TWh). According to the International Energy Agency (IEA), under a sustainable development scenario, solar and wind could provide more than 40% of global electricity generation by 2040. This would require significant investments, policy changes, and innovations to overcome the current challenges of grid integration, storage, transmission, and reliability. In view of this, the present study concentrates on wind and solar energy among other renewable sources.

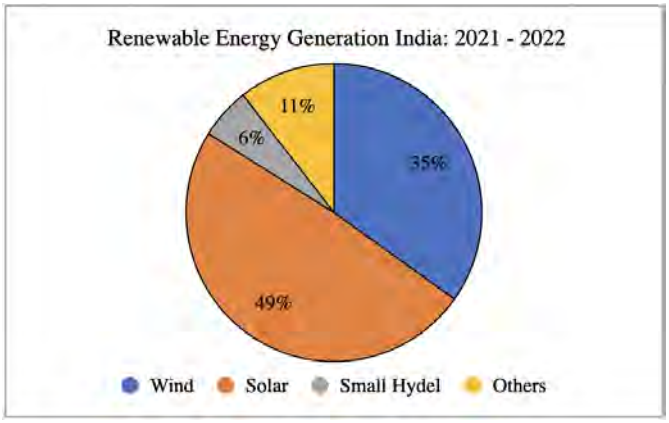


Fig. 1.2: Share of different types of sources in overall renewable energy generation in India [120].

Looking at the Indian scenario, India is the world's third largest energy consumer and the third largest renewable energy producer. The share of wind and solar energy in India is considerably high, as shown in Figure 1.2. The contribution of each state in wind and solar energy generation from December, 2021 to November, 2022 is shown in Figure 1.3. To note, the resources of wind and solar energy are wind speed and solar irradiance, respectively. The availability of average global horizontal irradiance (GHI) and wind speed in India are shown in Figure 1.4 and Figure 1.5, respectively. From these figures, we observe that the Indian region receives high amount of solar irradiance as well as wind speed. The Indian government is investing heavily in the renewable energy sector, with a primary focus on solar and wind power generation. From 94.4 *GW* in 2021, the renewable energy generation capacity of India has gone up to 119.1 *GW* in 2023. The government of India has set an ambitious target to reach 450 *GW* of renewable capacity by 2030. The country has already achieved its target of 40% installed electric capacity from non-fossil fuels. There is a steady upward trend in the installed renewable energy capacity. As a consequence, several wind and solar energy plants are currently being set up throughout the country under the purview of the Ministry of New and Renewable Energy (MNRE), Government of India. The government has launched many initiatives to promote the use of solar and wind energy across the country. Few of them are listed below.

- Rooftop Solar Scheme (2022 onward)– aims to provide subsidies to consumers who install rooftop solar systems.
- Solar Park Scheme (2014 onward)– aims to make land available to solar plant investors and developers by establishing solar parks across the country.
- UDAY (Ujjwal DISCOM Assurance Yojana) Scheme (2015 onward)– aims to improve the financial and operational efficiency of state-owned power distribution companies.

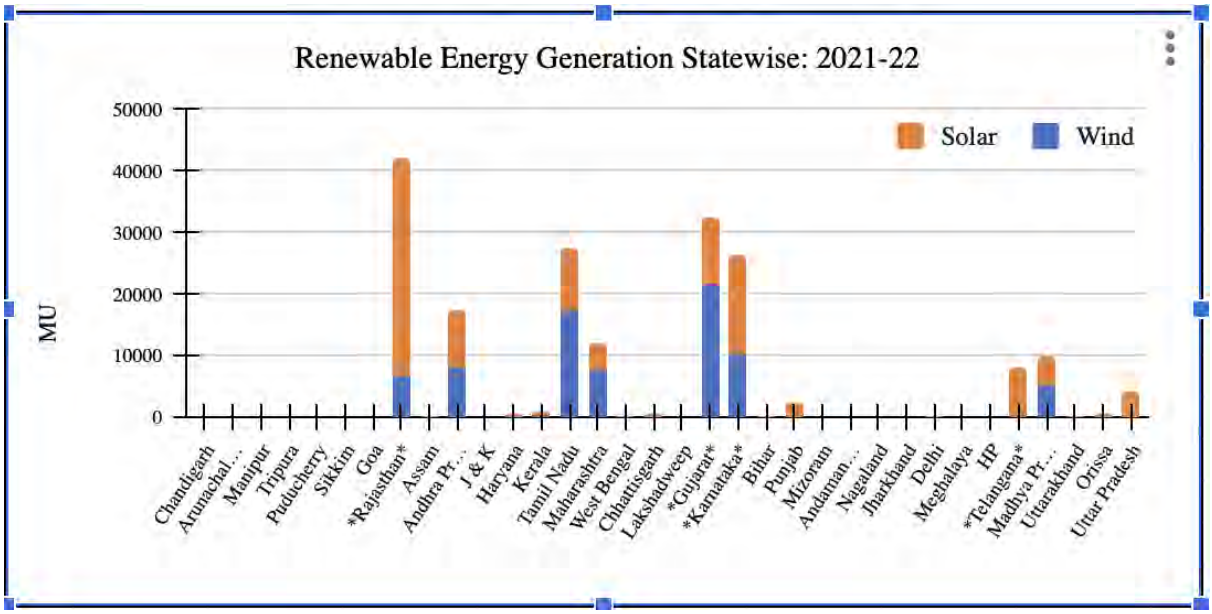


Fig. 1.3: Wind and solar energy generation in different states of India [120].

- PM KUSUM (Pradhan Mantri Kisan Urja Suraksha evam Utthan Mahabhiyaan) Scheme (2019 onward)– aims to promote the use of solar power for irrigation.
- National Wind-Solar Hybrid Policy (2018 onward)– aims to optimize the utilization of land and grid infrastructure by integrating wind and solar power generation.
- National Offshore Wind Energy Policy (2015 onward)– aims to harness the potential of offshore wind energy in India's coastal regions.

In view of the above discussion, the present study focuses on wind and solar energy forecasting for the Indian region.

1.2 Need of Renewable Energy Forecasting

When traditional sources like fossil fuels are used for power generation, the output is predominantly controlled by the plant's generation capacity. However, when renewable sources like solar and wind energy are used for power generation, the generation also depends on weather conditions apart from the machines' capacity. In this regard, two main categories of studies have evolved in the domain of renewable energy, one focusing on the smart grid or grid energy storage technology and another aiming at forecasting of renewable energy. Out of these two categories, we concentrate on the forecasting of renewable energy and highlight its need through a few prominent applications below.

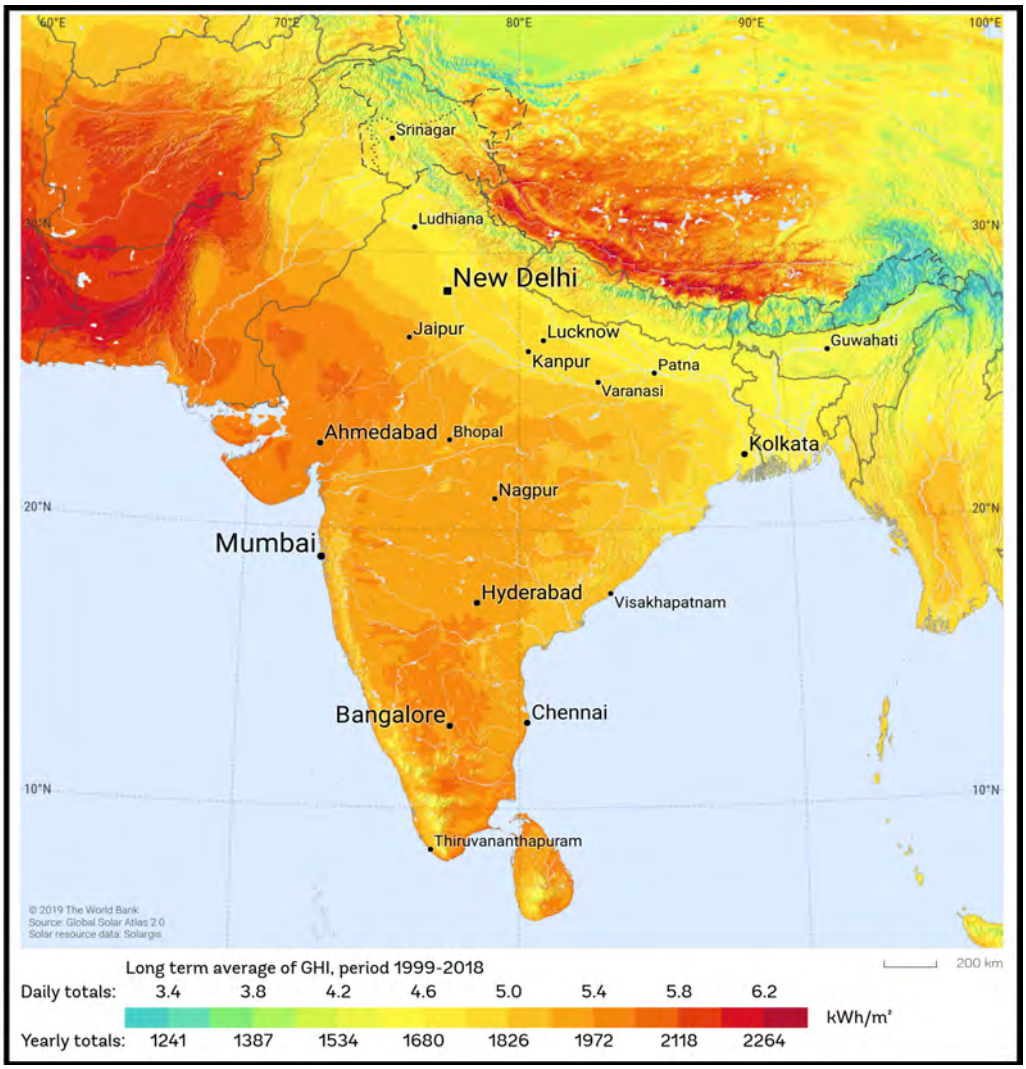


Fig. 1.4: Average global horizontal irradiance in India [134].

- **Ensuring the availability and optimal use of renewable energy**: Forecasting techniques strive to reduce the variability of production levels at any interval/point of time. Thus, accurate characterization and prediction ensure the availability and optimal use of renewable energy towards a proper management of power generation.

- **Power integration in grid system**: The unpredictability of renewable energy resources makes the power integration to existing grid system challenging [145]. Electricity suppliers are often interested in various time-horizons, such as the short term, intermediate term, and long term prediction not only to better estimate the fossil fuel saving but also to increase the integration limit of renewable energy systems. Thus, an accurate forecasting of solar irradiation and wind speed is recognized to be a major contribution for reliable large-scale solar and wind power integration [169].

- **Reliable power system operation**: Higher penetration of wind and solar energy significantly increases the uncertainties in wind and solar power systems, leading to complications in system operations and planning [53]. Therefore, enhanced forecasting of wind speed and solar irradiation enables independent system operators to function more efficiently and to run the power systems reliably.

- **Energy transactions**: As solar irradiation and wind speed are variable and intermittent over various timescales due to their dependency on several environmental conditions, the fundamental basis of managing existing and newly constructed power systems is the power generation forecasting, failing to do so may lead to inappropriate operational practices and inadequate energy transactions [146].

Having realized the necessity of renewable energy forecasting, the next section discusses various timescales relevant to energy forecasting.

1.3 Timescale of Forecasting

Prediction time horizon and forecasting approaches are the two main criteria used to classify renewable energy forecasting models. Immediate short term (few seconds to one hour), short term (few hours to few days), medium term (weeks to months), and long term (up to few years) forecasting are the four categories with respect to the time horizon. In electricity markets, short term forecasting is critical for unit commitment and operational security. Shorter forecasting time horizon, on the other hand, can produce more detailed and precise predictions, though it allows less time for power generation installations. Long term scheduling and planning in electricity market necessitate medium and long term forecasts. However, increasing the forecasting horizon unavoidably reduces forecasting accuracy [140, 148]; a few minutes to one hour ahead forecasting is required for re-dispatching imbalances rather than employing the limited and expensive frequency regulation control [41]. Longer predicting time horizons provide more long term knowledge on future energy. Therefore, temporal time horizons play a crucial role for energy operations in integration of intermittent solar and wind energy into current and future energy structure. Renewable energy forecasting may be classified into four major categories as summarized below in Table 1.1. In view of the above discussion, the present study considers the forecasting of wind speed and GHI at hourly, daily, weekly, and monthly timescales for the Indian region.

Table 1.1: Classification of renewable energy forecasting based on time horizon

Type	Time Horizon	Purpose
Immediate short term	Few seconds to one hour	Grid stability operation, voltage regulation action, and economic load dispatch planning
Short term	Few hours to few days	Load increment or decrement decisions, power scheduling, operational security in the electricity market, management of reserve power, and generator online/offline decisions
Intermediate term	Weekly to monthly	Maintenance scheduling and unit commitment decisions
Long term	Up to few years	Power plant and wind farm optimal design and restructured electricity markets

1.4 Methods of Renewable Energy Forecasting

Based on the adopted methodology, renewable energy forecasting techniques have four broad categories: (i) physical methods, (ii) time series (TS) methods, (iii) machine learning (ML) methods, and (iv) hybrid techniques.

1.4.1 Physical Methods

The physical methods, such as numerical weather prediction (NWP) models, primarily use highly non-linear equations that evolve from the underlying dynamics, thermodynamics, and radiative processes for the entire Earth involving several atmospheric factors. Due to the involvement of stringent boundary conditions, subjective knowledge of the hyper-parameters, and an ideal physical setup, the physical methods models have limited applicability in short term forecasting; they are mainly used for long term forecasting [112, 142, 164].

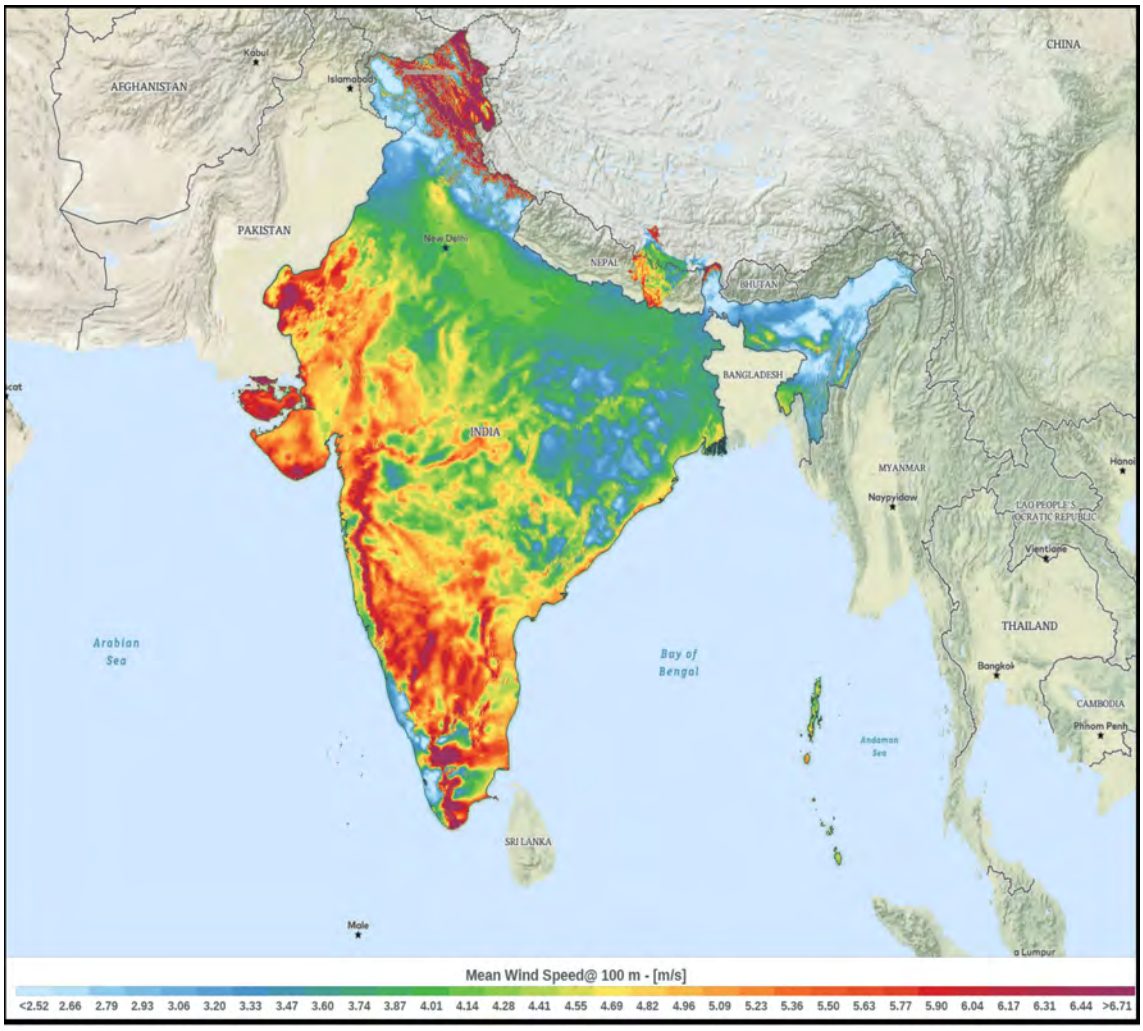


Fig. 1.5: Average wind speed in India [154].

1.4.2 Time Series Methods

Time series forecasting models are statistical models used to make predictions about future values based on historical data points arranged in a chronological order. These models analyze trends and patterns in the data and extrapolate them to make predictions about future values. The most popular time series based forecasting approaches are auto-regressive integrated moving average (ARIMA), seasonal-ARIMA (SARIMA), fractional-ARIMA (f-ARIMA), ARMA with exogenous input (ARMAX or ARX) [39], and window-sliding based ARIMA (WS-ARIMA) [122]. Time series models are comparatively easy to understand and implement for prediction purposes. Since these methods forecast the behavior of a process based on its own previously observed values, the dependency among interrelated variables can not be much taken into account [150]. In addition, the time series models with fixed coefficients can only consider linear relationships among data variables.

1.4.3 Machine Learning Methods

Artificial intelligence or machine learning methods are often employed to deal with linear as well as non-linear relationships among data variables. Machine learning models can learn from data and make predictions or decisions based on the learned patterns. Popular machine learning algorithms for wind speed and GHI prediction include support vector machines (SVMs), artificial neural networks (ANNs), recurrent neural networks (RNNs), long short term memory (LSTM), and convolutional neural networks (CNNs). A few drawbacks or the problems that creep in the machine learning techniques include over-fitting, usage of high computational power of computer systems, and extensive need of hyperparameters tuning and high dimensional data [165]. As a result, machine learning models are comparatively difficult to deal with due to their inherent complexity [94].

1.4.4 Hybrid Methods

The hybrid methods integrate the intersectional complimentary benefits of many standalone models. To develop hybrid setups, heterogeneous models, such as linear and non-linear models [138] or homogeneous models, such as neural networks with various configurations [36, 55, 76] may be utilized. Researchers have regularly employed the linear statistical time series ARIMA model from a plethora of linear techniques and artificial neural networks from the pool of non-linear approaches to build hybrid models for forecasting of solar irradiance and wind speed [70]. The popular hybrid models in wind speed and GHI forecasting include LSTM-CNN, ARIMA-ANN, light gradient boosting machine and Gaussian process regression (LGB-GPR), and CNN-LSTM-SVR [141].

In view of the above complementary benefits of statistics and machine learning, the present study aims to implement and integrate both of these techniques for an effective analysis of renewable energy resources and their consequent forecasting. Furthermore, a comparative analysis of the results, accuracy measurements through different error metrics (e.g., root mean square error and mean absolute percentage error), and sensitivity testing of the model parameters will provide more insight to the temporal variation of the underlying process as discussed in the following sections.

1.5 Measures of Forecast Accuracy

Having discussed several terminologies and methods of renewable energy forecasting in the previous sections, this section provides an overview of various performance measures to evaluate the efficacy of forecasting techniques. In literature, performance of forecasting models has been assessed using mean error, mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), histograms of the frequency distribution of the error, the correlation coefficient, mean absolute percentage error (MAPE), and the coefficient of determination. Model performances are also compared using Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. In the present study, accuracy of the forecasting models is evaluated and compared on the basis of the RMSE and MAPE error metrics. Lower the error values, better is the forecast. The formulae for RMSE and MAPE are given as:

$$\begin{aligned} RMSE &= \sqrt{\sum_{t=1}^{N} \frac{(X_t - \hat{X}_t)^2}{N}} \\ MAPE &= \sum_{t=1}^{N} \left| \frac{(X_t - \hat{X}_t)}{X_t} \right| \times \frac{1}{N} \end{aligned} \tag{1.1}$$

where, X_t is the forecast variable; $\hat{X}_t$ is the corresponding forecast value, and N denotes the length of time series of the forecast variable.

1.6 Residual Analysis

It may be noted that the accuracy measures discussed in the previous section provide information on the overall collective error rather than the instantaneous error of the forecast. The information of instantaneous error may be more relevant from an operator point of view. In this regard, there are explicit post-processing techniques which consider bias, accuracy, uncertainty, reliability, and resolution to validate the results of the forecasting models (Chapter 9 in [112]). Among these, the present study concentrates on forecast bias, a tendency for a model to consistently produce higher or lower forecast values than the actual values. Therefore, analysis of bias is an important post-processing step. The primary characteristic of the unbiased models is the normality of the residuals. Therefore, if the residuals do not exhibit a normal distribution, their randomness is lost, violating the fundamental assumption of a forecast model. We perform the residual analysis to check whether there is any systematic bias in the implemented models. As the residuals of a forecast model should exhibit normal distribution with zero mean and a constant variance, we analyze residual plots, histogram plots, and P-P plots (Section 14.8 and Section 14.9 in [8]) of the standardized residuals corresponding to the implemented models.

Moreover, we also apply Jarque-Bera test to statistically check for Gaussian behavior of the residuals.

1.7 Thesis Objective

The main objective of the thesis work is to develop renewable energy forecasting techniques in the Indian region using time series, machine learning, and hybrid techniques in order to provide a generic guideline on the choice of appropriate model for forecasting at desired time horizon. To achieve this primary objective, four sub-objectives are framed as follows:

1. To carry out preliminary data analysis and to explore the best fit probability distribution(s) for wind speed and GHI data
2. To implement various statistical time series methods for renewable energy forecasting
3. To implement several machine learning techniques for short term, intermediate term, and long term renewable energy forecasting
4. To explore hybrid setups for renewable energy forecasting and to compare their efficacy with the standalone time series and machine learning models.

1.8 Scope of the Thesis

This section explains various scopes and key work elements to accomplish the above mentioned research goal. In the first sub-objective, we study statistical characteristics and implement popular probability distributions for wind speed and GHI modeling in the Indian region. For this, we carry out the following subsequent tasks:

1. Collection of wind speed and GHI data (2000–2014) at four selected study sites in India, one each from the state of Rajasthan, Gujarat, Karnataka, and Telangana.
2. Analysis of data information using descriptive statistics and time series plots.
3. Decomposition of time series data to explore trend, seasonal, and irregular components.
4. Performing stationarity test to assess the compatibility of data to the forecasting models.
5. Implementation of five popular probability distributions, namely exponential, gamma, lognormal, Weibull, and exponentiated Weibull.

6. Comparison of model performance to identify the best fit distribution for data pattern recognition.

To accomplish the second objective, that is, to implement various statistical time series methods for renewable energy forecasting, we utilize wind speed and GHI data in the following manner.

1. Split the dataset into training (80%) and testing (20%) parts.
2. Find the optimum values of model parameters through grid search method on training data.
3. Obtain forecast values and assess the performance of the studied models (AR, MA, ARMA, ARIMA, SARIMA, and WS-ARIMA) through RMSE and MAPE.
4. Perform residual analysis as a post-processing step to examine any systematic bias in the implemented models.

The third objective focuses on the implementation of machine learning techniques (SVR, ANN, LSTM, BiLSTM, encoder-decoder LSTM, attention layer LSTM, and CNN) for short term, intermediate term, and long term renewable energy forecasting. The subsequent tasks are as follows.

1. Prepare data in the form of input vectors and matrices.
2. Obtain the optimal parameters (e.g., type and number of hidden layers, activation function, optimization algorithm, loss function, epochs, and the learning rate) in such a way that the training error values are minimized.
3. Run the model for the test data and calculate the prediction errors.
4. Perform residual analysis to validate the implemented forecasting models.
5. Compare performance of the studied machine learning models and list out the best models across time and space.

The fourth and last sub-objective focuses on the implementation of a few hybrid models and comparison of their efficacy with the standalone time series and machine learning models. The subsequent tasks are as follows.

1. Implement ARIMA model to analyze the linear part of the data.
2. Implement a neural network to model the residuals from the ARIMA model.

3. Calculate final forecasts through a combination of outputs from both ARIMA and neural network.

4. Assess efficacy of the hybrid models in comparison to the corresponding standalone models through RMSE values and residual characteristics.

5. Provide a comprehensive summary of all fifteen implemented models and list the best performing models in wind speed and GHI forecasting at a desired time horizon.

To illustrate the above steps, a simple flowchart is provided in Figure 1.6.

1.9 Structure of the Thesis

Having discussed the main objective and scope of the thesis, this section provides a brief thematic overview of the chapter-wise road map.

Chapter 1 explains the current status of renewable resources worldwide and particularly for the Indian region. It also includes the industrial application of forecasting at different timescales. This chapter covers the classification of renewable energy forecasting techniques based on their methodology. In addition, it includes the thesis objective, scope of the thesis, and a brief summary of each chapter.

Chapter 2 provides details of statistical characteristics of wind speed and solar irradiance data collected at four locations from the Indian region. It includes time series analysis through sampling, decomposition, stationarity test, and fitting of probability distributions for data pattern recognition.

Chapter 3 presents a comprehensive analysis of time series models in renewable energy forecasting. It includes the list of optimum parameters and corresponding forecasting errors of the studied models across time and space. The comparative performance of different models at different timescales is presented through relevant figures. The residual plots, histograms, and the P-P plots of the residuals corresponding to the best fit models are also included. In this way, the chapter provides a generic guideline for the applicability of different statistical models for wind speed and GHI forecasting at desired time horizon.

Chapter 4 focuses on hourly, daily, weekly, and monthly forecasting of wind speed and GHI using various machine learning methods, namely SVR, ANN, four variants of LSTM, and CNN. The optimal parameters and configurations are noted for each model. The chapter provides relevant tables and figures to summarize the comparative performance of the studied machine learning models. The graphs representing unbiased nature (through residuals) of the best fit models are also included.

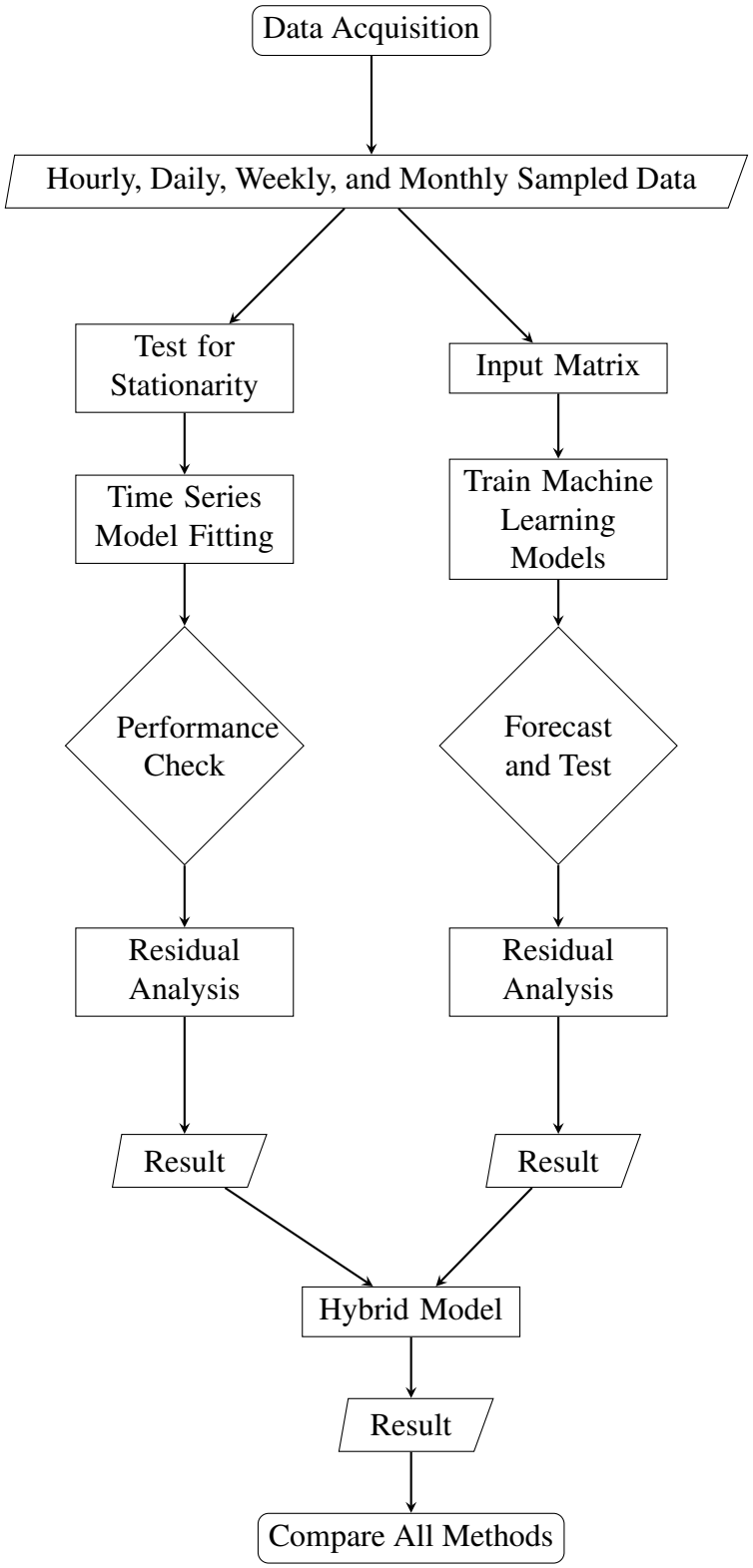


Fig. 1.6: Flowchart of the adopted methodology.

Chapter 5 discusses the methodology of three implemented hybrid models. An ARIMA model is first implemented to analyze the linear part of the data and then a neural network model is developed to model the residuals from the ARIMA model. This chapter essentially provides a comprehensive summary of all studied models and highlight the best three models in wind speed and GHI forecasting at different timescales.

Finally, **Chapter 6** summarizes the thesis work with an emphasis to the major contributions and future recommendations.

Chapter 2

Study Region and Dataset

"*The more that you read, the more things you will know. The more that you learn, the more places you'll go.*"

– DR. SEUSS

This chapter presents a statistical assessment of wind speed and solar irradiance data collected at four locations from the Indian region. For this, we carry out time series analysis through sampling, decomposition, stationarity test, and distribution fitting on the datasets. We consider several probability distributions, namely exponential, gamma, lognormal, and Weibull distribution to analyze the wind speed and GHI characteristics at the selected locations. In addition, we implement the exponentiated Weibull distribution which shares many physical properties with gamma, lognormal, and Weibull distributions. We estimate model parameters through the maximum-likelihood estimation and compare model suitability through the Kolmogorov–Smirnov (K-S test) goodness of fit test. The results reveal that the exponentiated Weibull distribution has the best representation of both wind speed and GHI data.

Contents

2.1 Introduction

To obtain energy in efficient and affordable ways, there are two crucial elements to be considered: (i) the location where the energy system will be installed, since the energy data obtained at various sites is subject to many climate fluctuations and (ii) the statistical measures and distributions to identify the characteristics of the renewable resources, mainly wind speed and solar irradiance. In view of this, the present chapter considers the statistical assessment of wind speed and solar irradiance collected at four locations in India. The associated data come from the National Renewable Energy Laboratory's (NREL) National Solar Radiation Database (NSRDB) maintained by the US Department of Energy. The detailed description of the dataset is available in [118]. For the present analysis, we use 15 years (January 1, 2000 to December 31, 2014) of hourly wind speed (m/s) and GHI (W/m^2) data from four selected locations in India, one each from Rajasthan, Gujarat, Karnataka, and Telangana as shown in Figure 2.1. It may be noted that these states belong to the major contributing states in renewable energy production (refer to Figure 1.3). As wind and solar energy potential and their feasibility of utilization depend on the characteristics of their resources [5], we concentrate on wind speed and GHI characteristics at these selected locations. We carry out the time series analysis through sampling at different timescales, decomposition into the trend, seasonal and residual components, test for stationarity, and fitting of probability distributions to the data. All these steps are inevitable in preprocessing for the implementation of forecasting models. For distribution fitting, we consider five distribution models, namely exponential, gamma, lognormal, Weibull, and exponentated Weibull. For statistical inference of these distributions, we perform parameter estimation through the maximum likelihood method and model selection using the K-S test. The emanated results and associated discussions of each step are provided in subsequent subsections.

2.2 Data Variables

Along with hourly wind speed and GHI values, the features in the dataset include direct horizontal irradiance (DHI), direct normal irradiance (DNI), and other environmental variables, such as temperature, pressure, relative humidity, and solar zenith angle. The major components of the dataset are explained below.

1. **Direct Normal Irradiance (DNI)**: It is the amount of solar irradiation received per unit area by a surface that is always held perpendicular (or normal) to the rays coming from the direction of the Sun at its current position in the sky as shown in Figure 2.2. It excludes diffuse solar radiation by atmospheric losses due to absorption and scattering.

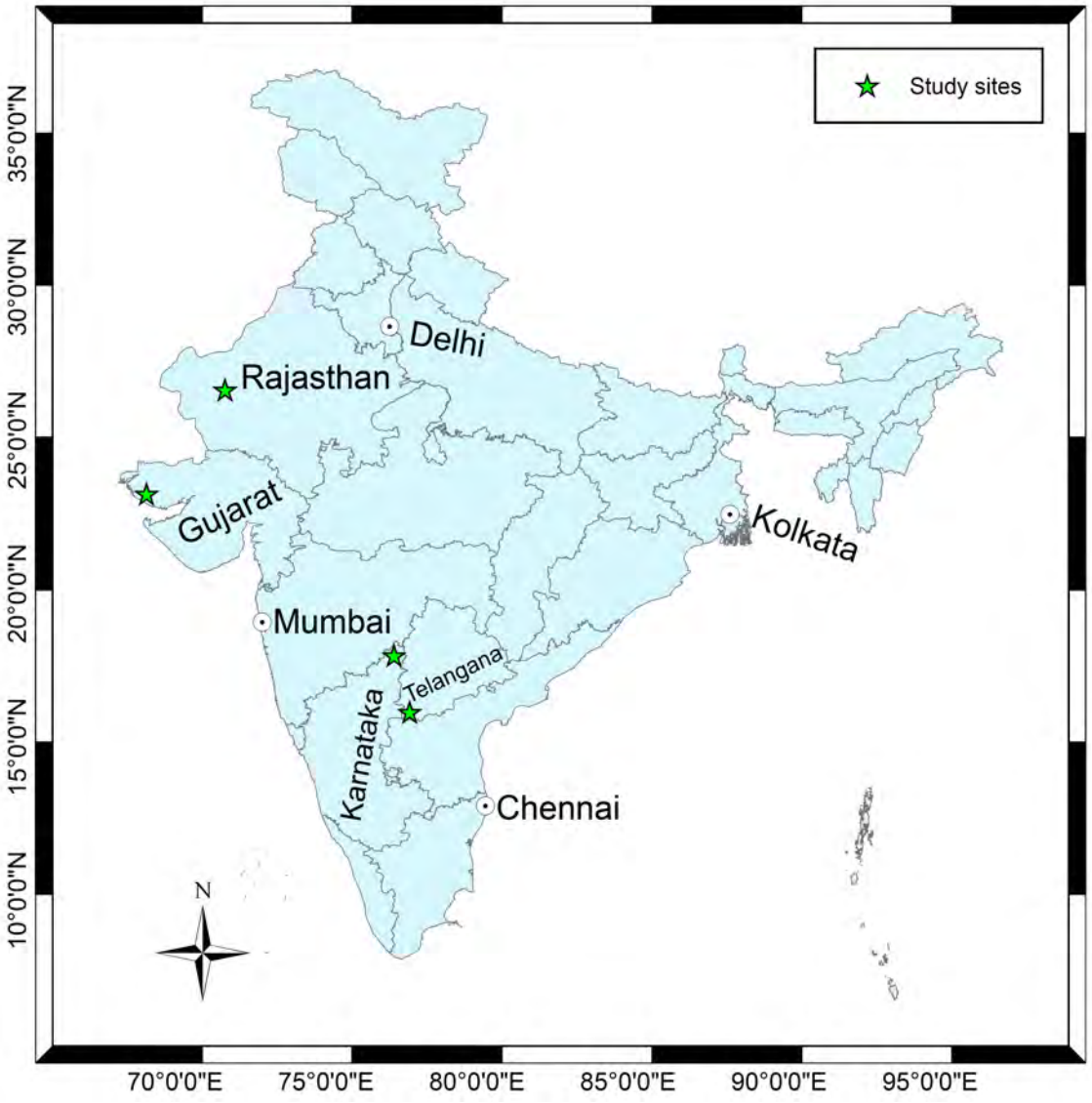


Fig. 2.1: Locations for the four selected study sites, highlighted by star.

Such losses depend on the time of day, cloud cover, moisture content, and other atmospheric variables. The irradiance above the atmosphere also varies with the time of year due to the varying distance of the sun. It is measured by an instrument called a pyrheliometer installed on a sun tracker. The DNI is measured in watt per square meter (W/m^2) and its typical value, for example, in the mid-noon of June in Rajasthan is $\sim 550\ W/m^2$.

2. **Diffuse Horizontal Irradiance (DHI)**: It is the amount of solar irradiation received per unit area by a surface that does not arrive on a direct path from the Sun, but has been scattered by molecules and particles in the atmosphere and comes equally from all directions. It is measured by pyranometer with a shadow ball or shadow ring, installed on a sun-tracker. The SI unit for the DHI is W/m^2 and its typical value, for example, in the

mid-noon of June in Rajasthan is $\sim 300\ W/m^2$.

3. **Global Horizontal Irradiance (GHI)**: The GHI is the amount of solar irradiation received from the Sun by a horizontal surface. This value includes both DNI and DHI. A pyranometer is used to measure GHI. With a solar zenith angle θ, the GHI is related to DNI and DHI as

$$GHI = DHI + DNI \times \cos(\theta) \tag{2.1}$$

The SI unit of irradiance is watt per square meter (W/m^2) and its typical value, for example, in the mid-noon of June in Rajasthan is $\sim 850\ W/m^2$. The solar energy industry uses watt-hour per square meter (Wh/m^2) per unit time. The relation to the SI unit is as follows.

$$1kW/m^2 \times (24h/day) = (24kWh/m^2)/day \tag{2.2}$$

4. **Solar Zenith Angle**: The solar zenith angle is the angle generated by the Sun's beams and the vertical direction. It is closely related to the solar altitude angle, which is the angle created by the Sun's beams and a horizontal plane. As these angles are complementary, the cosine of one equals to the sine of the other.

5. **Wind Speed**: It is the speed of wind at a point at a specific time. It is clearly related to wind energy. Higher the wind speed, higher is the wind energy output. Wind speed also depends on pressure gradient, which in turn is related to solar radiation intensity at two different places. Hence, in that sense, it can also be related to solar energy. Wind speed is measured in meter per second (m/s) and its typical value, for example, in the mid-noon of August in Rajasthan is $\sim 3\ m/s$.

6. **Atmospheric Pressure**: It is the pressure within the atmosphere of Earth, also known as the barometric pressure (after the barometer). Atmospheric pressure is caused by the gravitational attraction of the planet on the atmospheric gases above the surface. Atmospheric pressure has a significant influence on the efficiency of solar and wind energy forecast modules. It is measured in millibar ($mbar$).

7. **Temperature**: It is the temperature that persists throughout the day. It is a strong measure of the solar irradiation. The intensity of radiations directly influences the amount of generated solar energy. It has significant correlation with wind speed. The measurement unit is degree Celsius ($^\circ C$).

8. **Relative Humidity**: The amount of water vapors present in air is expressed as a percentage of the amount needed for saturation at the same temperature. This term is related

Table 2.1: Correlation matrix for the dataset from Rajasthan

	DHI	**DNI**	**GHI**	**Temp-erature**	**Zenith Angle**	**Pressure**	**Relative Humidity**	**Wind Speed**
DHI	1.00							
DNI	0.81	1.00						
GHI	0.93	0.94	1.00					
Temperature	0.61	0.47	0.59	1.00				
Zenith Angle	−0.89	−0.80	−0.87	−0.62	1.00			
Pressure	−0.17	0.08	−0.05	−0.64	0.19	1.00		
Relative Humidity	−0.20	−0.32	−0.30	−0.04	0.16	−0.53	1.00	
Wind Speed	−0.08	−0.20	−0.12	0.11	0.09	−0.40	0.22	1.00

to both wind and solar energy. Note that the higher solar radiation leads to more evaporation of water and hence, water in the atmosphere will be more. Thus, more relative humidity means more solar radiation intensity. Also, if relative humidity is higher, the air will become heavier and hence, wind speed will decrease. Thus, wind energy output will decrease.

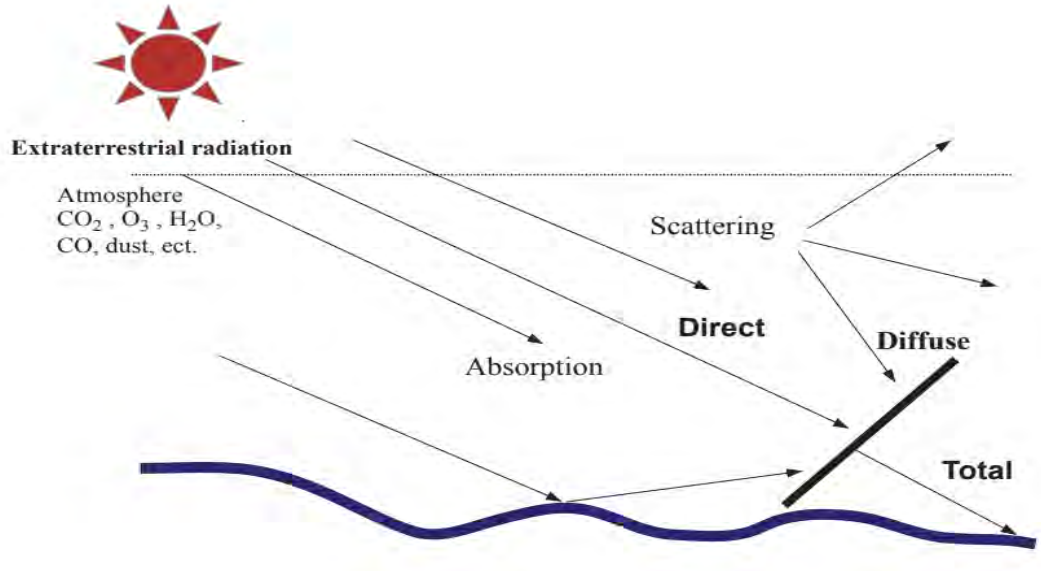


Fig. 2.2: Types of solar irradiance received at a location [139].

To study the linear relationship among these variables and how the changes in one variable affect others, we have studied the correlation among these variables. The correlation matrices for the data variables at four selected locations are presented in Tables 2.1, 2.2, 2.3, and 2.4.

We observe that the GHI is highly correlated to DHI (≥ 0.83), DNI (≥ 0.64), zenith angle (≤ -0.84), and temperature (≥ 0.59) for the four study sites, whereas the wind speed is highly

Table 2.2: Correlation matrix for the dataset from Gujarat

	DHI	**DNI**	**GHI**	**Temp-erature**	**Zenith Angle**	**Pressure**	**Relative Humidity**	**Wind Speed**
DHI	1.00							
DNI	0.74	1.00						
GHI	0.90	0.93	1.00					
Temperature	0.64	0.47	0.73	1.00				
Zenith Angle	−0.88	−0.76	−0.86	−0.70	1.00			
Pressure	−0.17	0.12	−0.04	−0.36	0.17	1.00		
Relative Humidity	−0.39	−0.65	−0.54	−0.30	0.42	−0.49	1.00	
Wind Speed	−0.05	−0.23	−0.09	0.02	0.01	−0.45	0.32	1.00

Table 2.3: Correlation matrix for the dataset from Karnataka

	DHI	**DNI**	**GHI**	**Temp-erature**	**Zenith Angle**	**Pressure**	**Relative Humidity**	**Wind Speed**
DHI	1.00							
DNI	0.64	1.00						
GHI	0.83	0.93	1.00					
Temperature	0.66	0.64	0.73	1.00				
Zenith Angle	−0.86	−0.76	−0.86	−0.70	1.00			
Pressure	−0.17	0.01	−0.09	−0.23	0.16	1.00		
Relative Humidity	−0.53	−0.61	−0.64	−0.83	0.57	−0.05	1.00	
Wind Speed	0.05	−0.18	−0.09	0.03	0.01	−0.41	0.06	1.00

correlated with pressure (≤ -0.40) for all study locations. This information is useful at several points of time in the current study.

As mentioned earlier, the collected data are recorded at hourly intervals, though other timescales such as daily, weekly, and monthly are important for the forecasting at these scales for different purposes (refer to Chapter 1). Thus, we resample the data at several timescales as explained in the next section.

2.3 Data Sampling

We have resampled the hourly wind speed and GHI data according to the desired time horizon of forecasting. We obtain daily, weekly, and monthly data from the hourly data through "resampling time series" based on their mean values [7]. For this, we use Python's inbuilt resample function 'df.resample('D',on='Date').mean()' for the daily, 'df.resample('W').mean()' for the weekly, and 'df.resample('M').mean()' for the monthly dataset. We calculate descriptive measures, such as sample mean, standard deviation, and quartiles of the hourly, daily, weekly, and

Table 2.4: Correlation matrix for the dataset from Telangana

	DHI	**DNI**	**GHI**	**Temp-erature**	**Zenith Angle**	**Pressure**	**Relative Humidity**	**Wind Speed**
DHI	1.00							
DNI	0.72	1.00						
GHI	0.88	0.94	1.00					
Temperature	0.57	0.53	0.61	1.00				
Zenith Angle	−0.88	−0.73	−0.84	−0.58	1.00			
Pressure	−0.09	0.13	−0.01	−0.32	0.12	1.00		
Relative Humidity	−0.35	−0.44	−0.44	−0.73	0.34	−0.05	1.00	
Wind Speed	−0.07	−0.24	−0.15	0.08	0.07	−0.47	0.04	1.00

monthly data (Table 2.5 and Table 2.6). For visualization, it is important to plot the data to examine inherent features of the datasets, such as (a) trend, (b) seasonal component, (c) any apparent sharp changes in behavior, and (d) any outlying observations. For this, we provide time series plots corresponding to hourly, daily, weekly, and monthly wind speed data of the four locations in Figure 2.3, Figure 2.4, Figure 2.5, and Figure 2.6, respectively. The time series plots in Figure 2.7, Figure 2.8, Figure 2.9, and Figure 2.10 represent the GHI datasets from the studied four locations at different timescales. These time series plots reveal yearly seasonal pattern in the datasets. We specifically observe seasonality without trend across all timescales and locations. Further investigation of these components (seasonality, trend, and residuals) has been carried out through the time series decomposition in the next section.

2.4 Time Series Decomposition

Time series decomposition is a popular approach to separate or decompose a time series into seasonal, trend, and irregular components. While this method can be used for forecasting, its primary applicability is to enable a better understanding of the time series. Time series decomposition methods assume that the actual time series value (Y_t) at period t is a function of three components: a trend component ($Trend_t$), a seasonal component ($Seasonal_t$), and residuals ($Irregular_t$) or error component. These three components are combined to generate the observed value by an additive or a multiplicative model. An additive model is appropriate in situations where the seasonal fluctuations do not depend upon the level of the time series [8]. An additive decomposition model takes the following form.

$$Y_t = Trend_t + Seasonal_t + Irregular_t \tag{2.3}$$

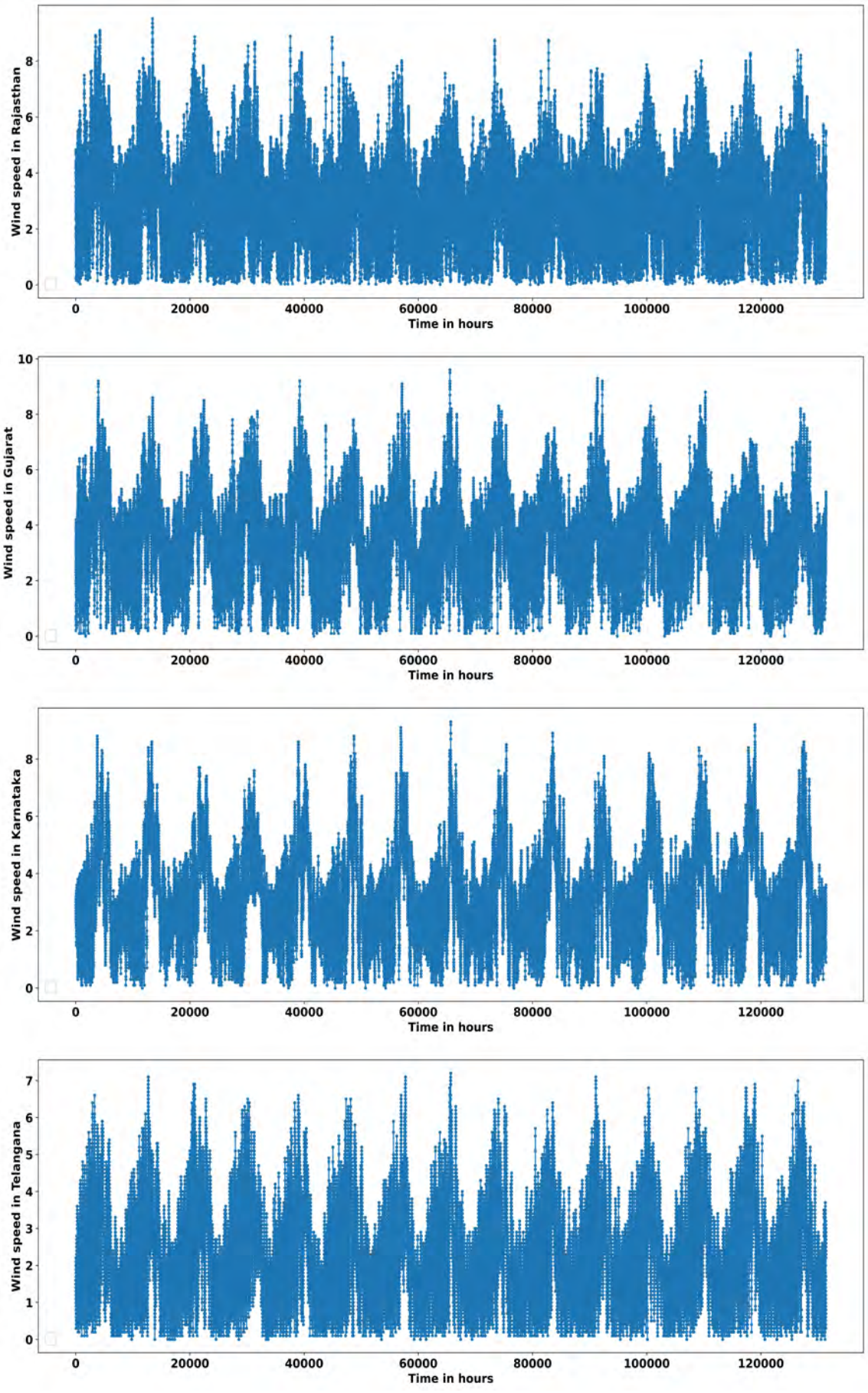


Fig. 2.3: Hourly average wind speed (m/s) at the four study sites.

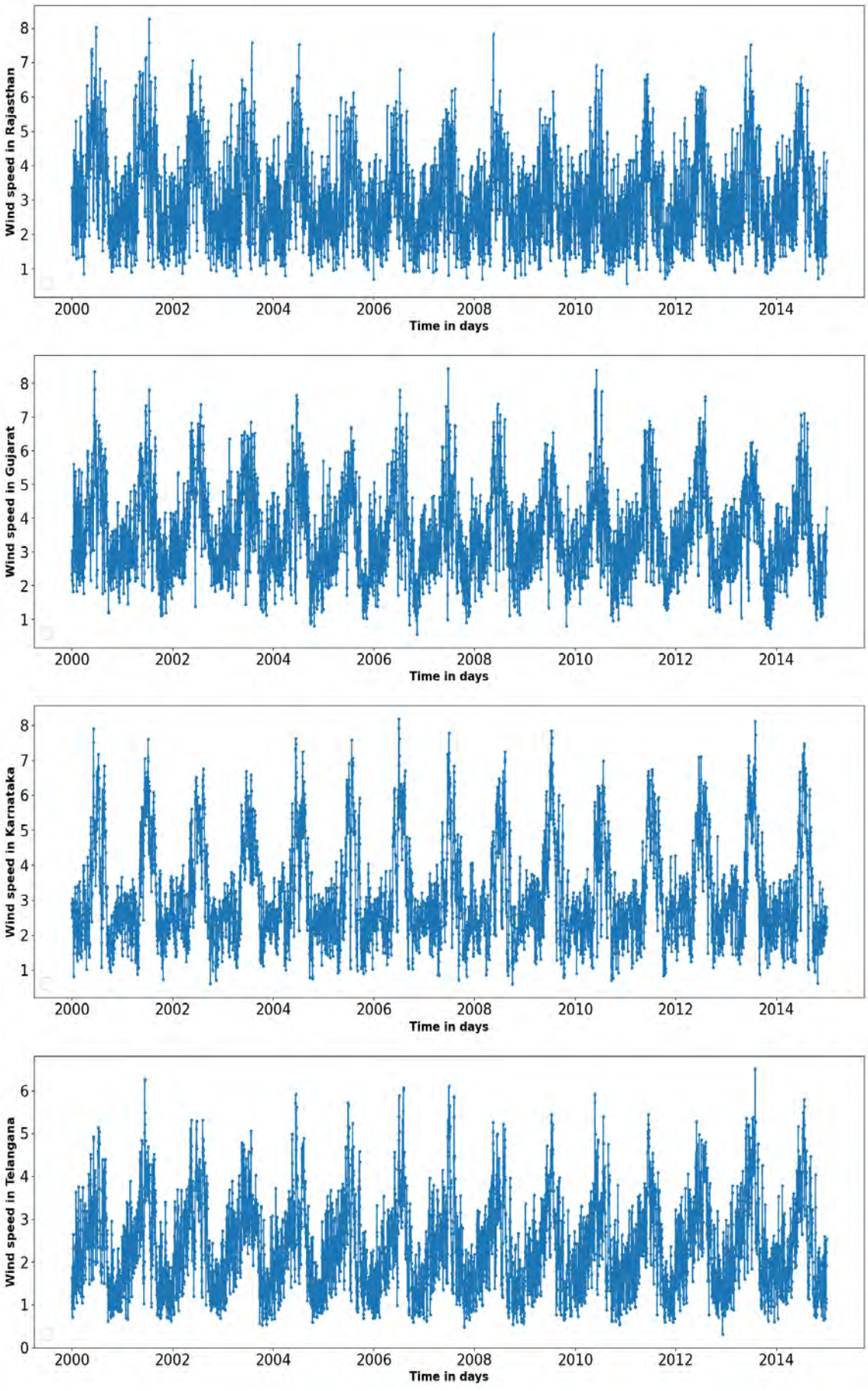


Fig. 2.4: Daily average wind speed (m/s) at the four study sites.

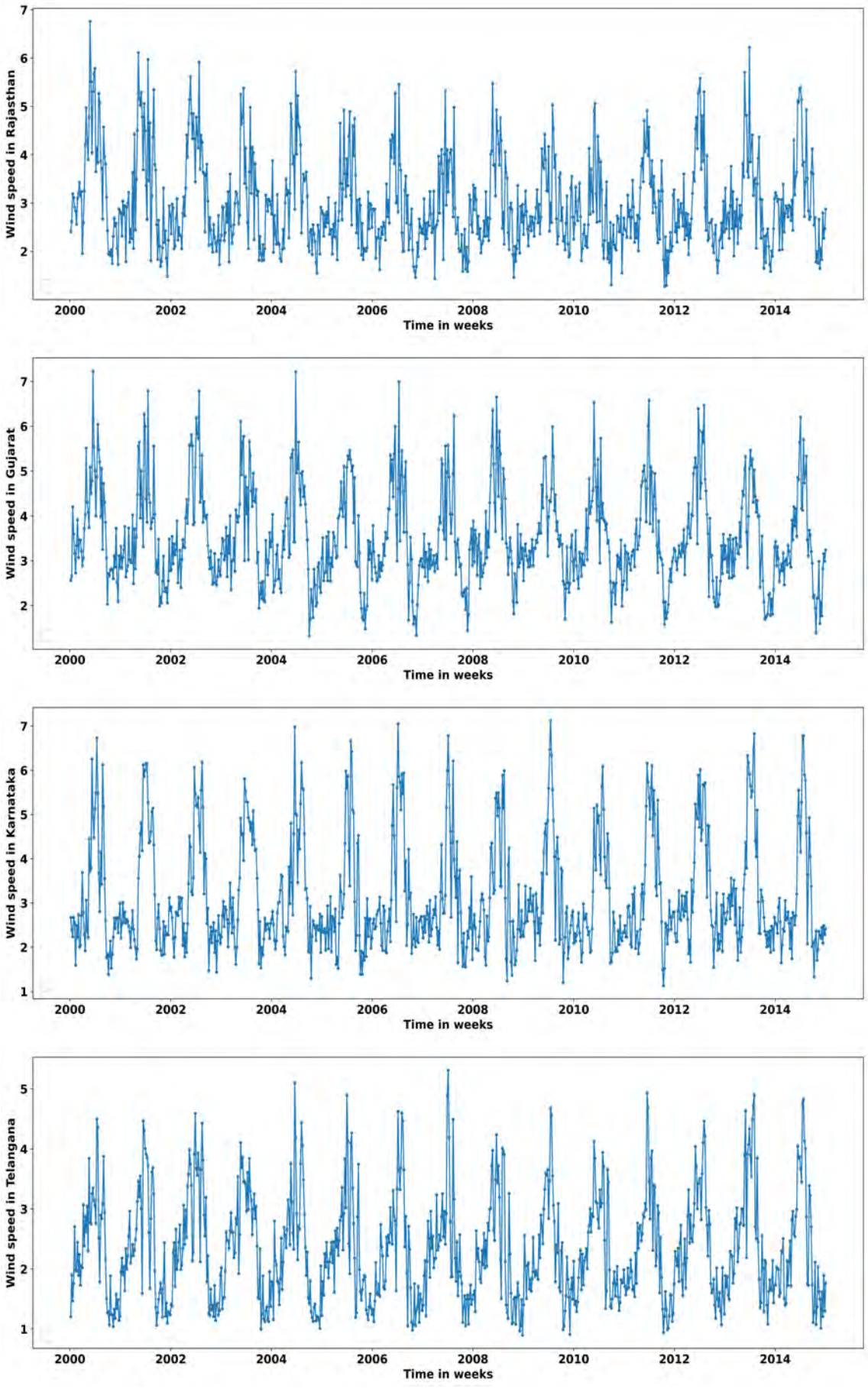


Fig. 2.5: Weekly average wind speed (m/s) at the four study sites.

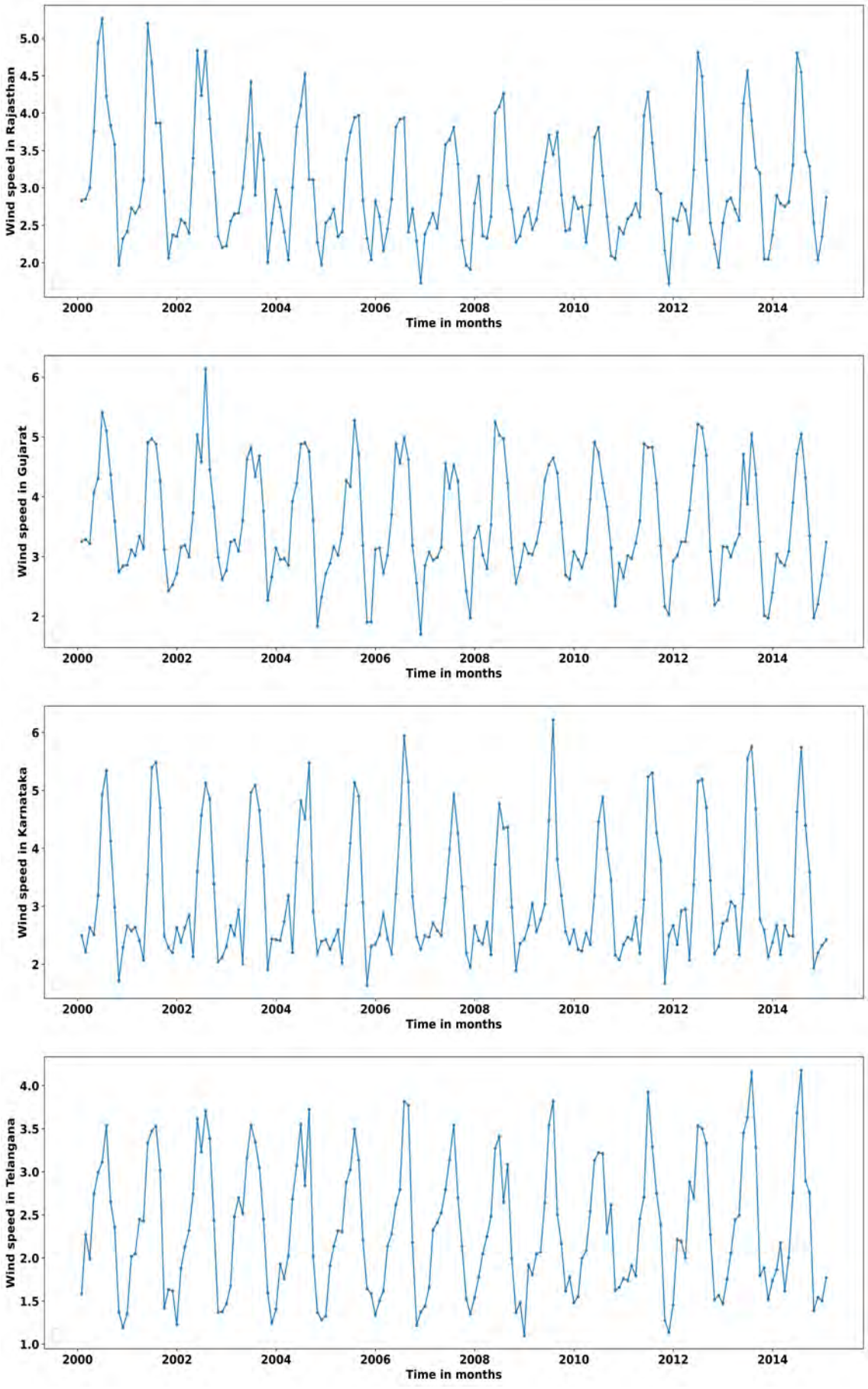


Fig. 2.6: Monthly average wind speed (m/s) at the four study sites.

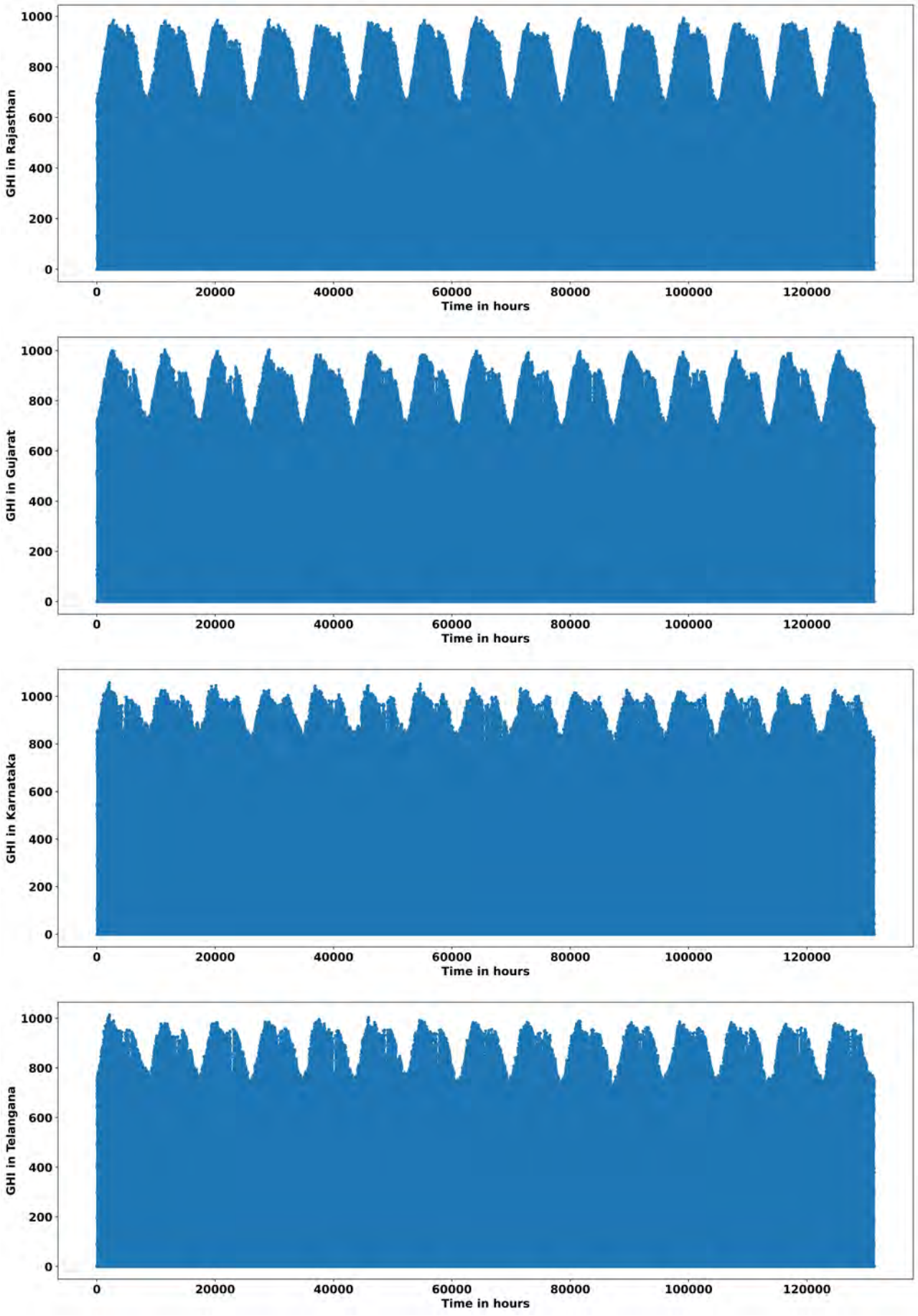


Fig. 2.7: Hourly average GHI (W/m^2) at the four study sites.

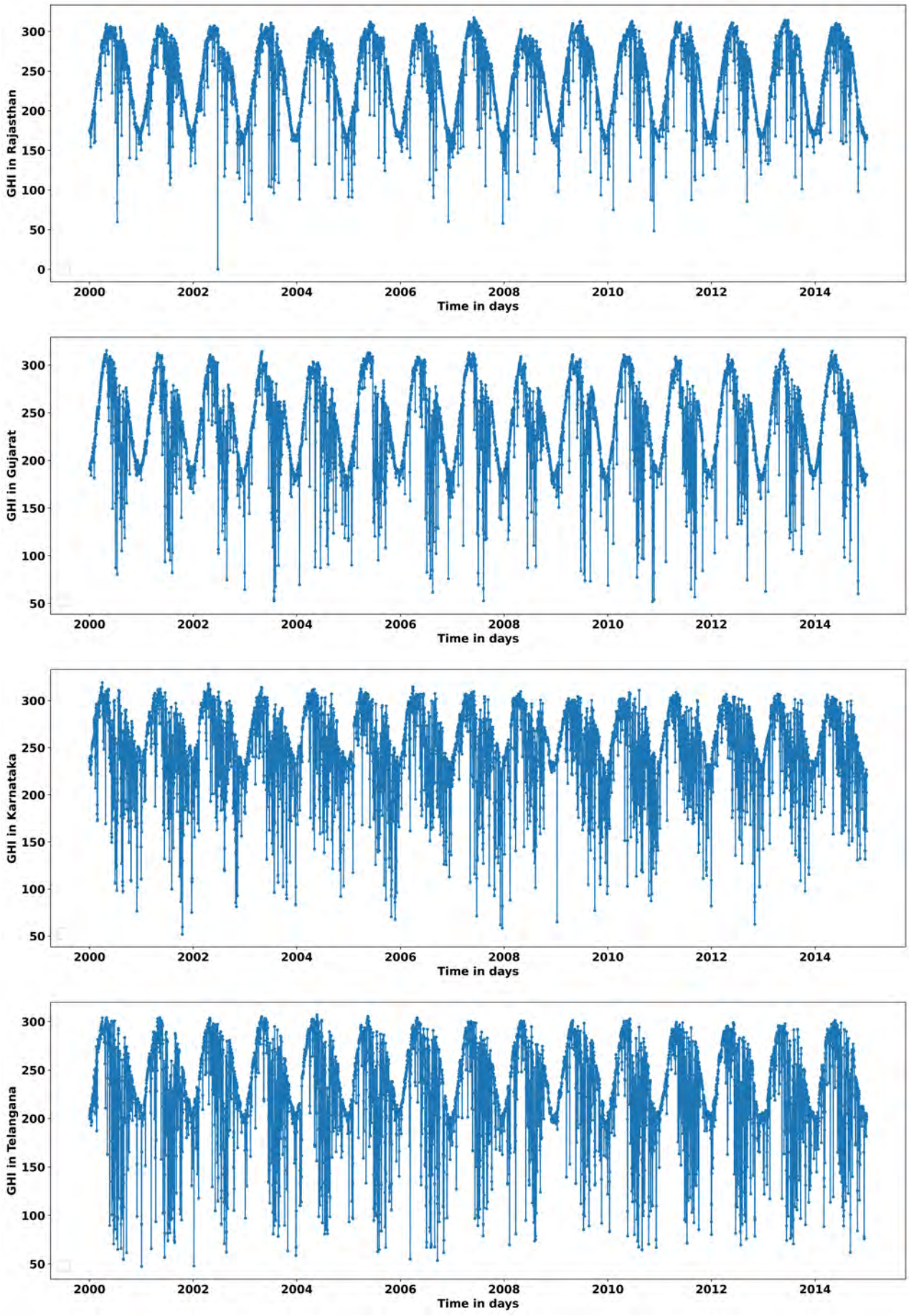


Fig. 2.8: Daily average GHI (W/m^2) at the four study sites.

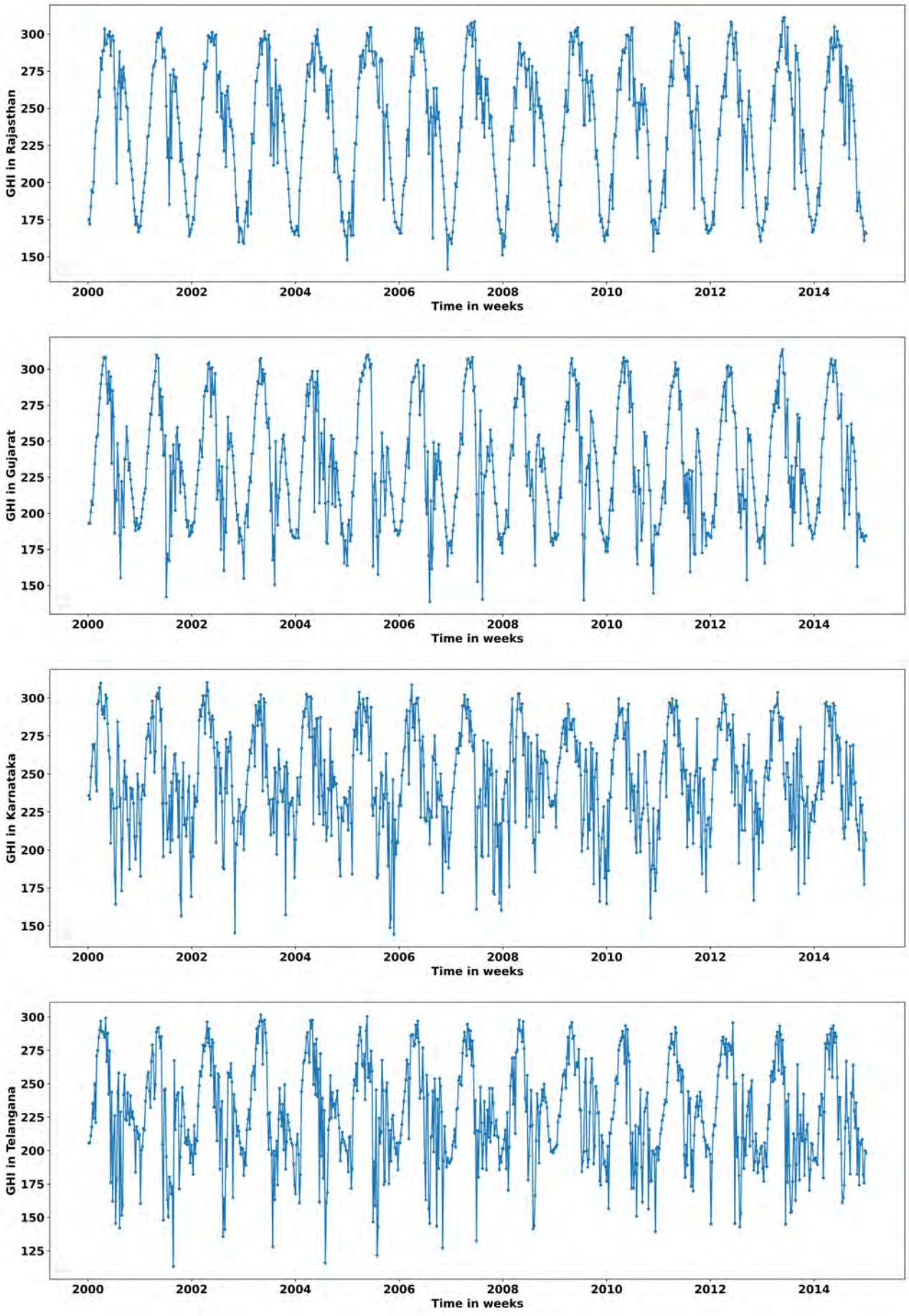


Fig. 2.9: Weekly average GHI (W/m^2) at the four study sites.

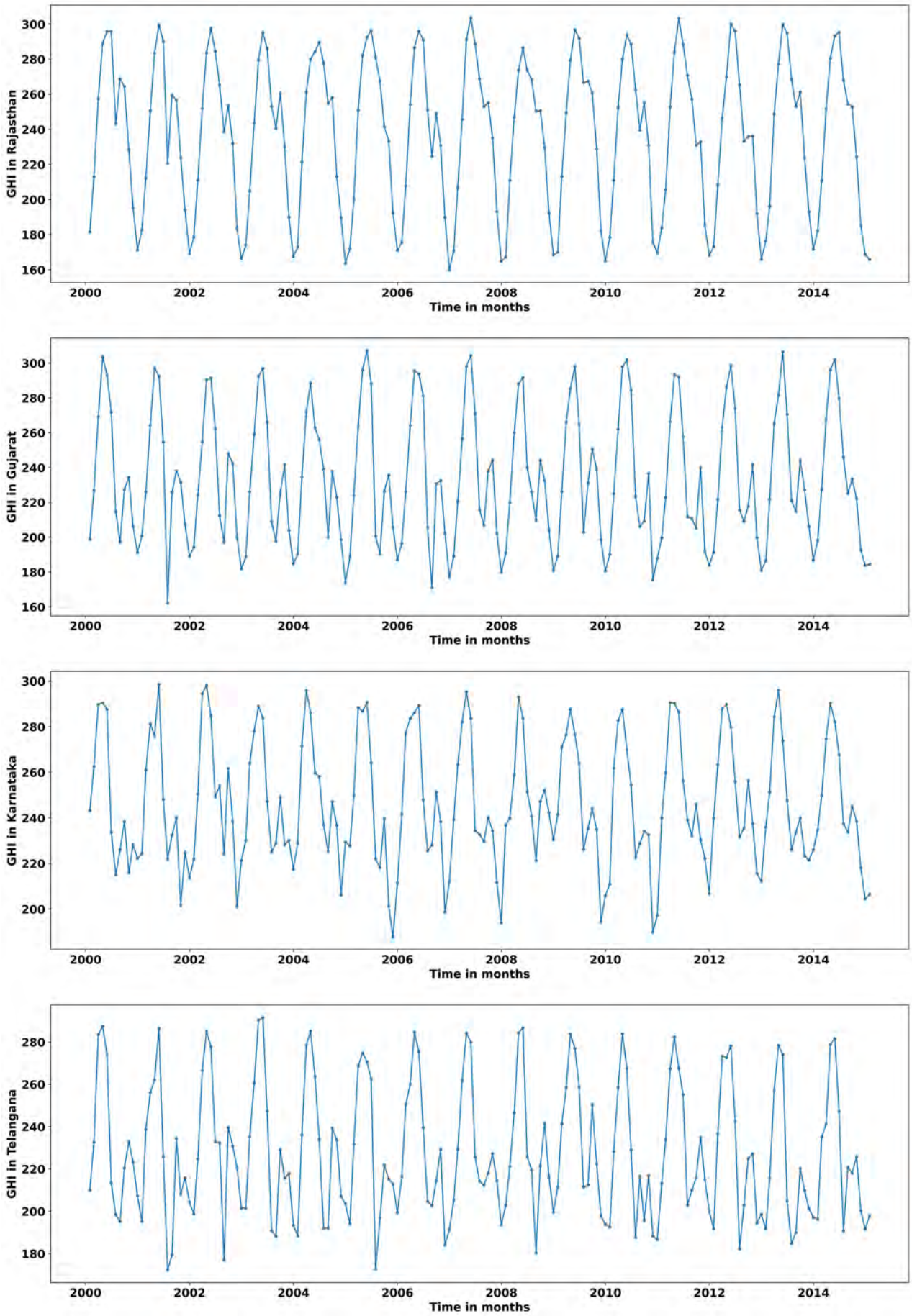


Fig. 2.10: Monthly average GHI (W/m^2) at the four study sites.

Table 2.5: Descriptive statistics of wind speed at four study sites

Study Site	Timescale	Data Count	Mean	Standard Deviation	Minimum	Maximum
Pokhran, Rajasthan (26.65°N, 71.65°E)	Hourly	131400	3.01	1.45	0.00	9.52
	Daily	5475	3.01	1.24	0.56	8.27
	Weekly	783	3.01	0.96	1.27	6.75
	Monthly	181	3.01	0.77	1.72	5.27
Bitta, Gujarat (23.25°N, 69.05°E)	Hourly	131400	3.54	1.42	0.00	9.60
	Daily	5475	3.53	1.27	0.56	8.43
	Weekly	783	3.53	1.08	1.22	7.22
	Monthly	181	3.53	0.92	1.70	6.13
Pavagada, Karnataka (14.15°N, 77.25°E)	Hourly	131400	3.17	1.50	0.00	9.30
	Daily	5475	3.17	1.38	0.59	8.18
	Weekly	783	3.17	1.27	1.13	7.12
	Monthly	181	3.16	1.09	1.60	6.20
Ramagundam, Telangana (18.75°N, 77.25°E)	Hourly	131400	2.34	1.24	0.00	7.20
	Daily	5475	2.34	1.04	0.30	6.50
	Weekly	783	2.33	0.91	0.89	5.30
	Monthly	181	2.33	0.76	1.09	4.18

Table 2.6: Descriptive statistics of GHI at four study sites

Study Site	Timescale	Data Count	Mean	Standard Deviation	Minimum	Maximum
Pokhran, Rajasthan (26.65°N, 71.65°E)	Hourly	54750	557.18	249.52	0.00	995.00
	Daily	5475	557.18	109.84	0.00	723.70
	Weekly	783	557.13	98.53	339.42	712.44
	Monthly	181	556.21	93.47	383.23	698.67
Bitta, Gujarat (23.25°N, 69.05°E)	Hourly	54750	550.05	243.96	0.00	1003.00
	Daily	5475	550.05	109.23	119.50	729.90
	Weekly	783	550.04	93.82	320.60	722.02
	Monthly	181	549.68	85.10	374.59	709.03
Pavagada, Karnataka (14.15°N, 77.25°E)	Hourly	54750	579.33	257.64	0.00	1058.00
	Daily	5475	579.33	103.38	120.40	746.50
	Weekly	783	579.35	78.16	346.14	726.24
	Monthly	181	579.32	62.69	449.13	697.86
Ramagundam, Telangana (18.75°N, 77.25°E)	Hourly	54750	537.40	255.91	0.00	1014.00
	Daily	5475	537.40	120.07	113.10	714.20
	Weekly	783	537.44	90.83	261.54	703.92
	Monthly	181	537.62	72.87	397.83	681.70

However, if the seasonal fluctuations change over time, growing larger because of a long term linear trend, then a multiplicative model is recommended [8]. Many business and economic time series follow this pattern. A multiplicative decomposition model takes the following form.

$$Y_t = Trend_t \times Seasonal_t \times Irregular_t \tag{2.4}$$

Since the seasonal fluctuations in time series plots (Figure 2.3 – Figure 2.10) do not vary with the level of the series, we prefer the additive model for the current study. To implement, we have used the naive or classical decomposition method in a function called $seasonal\_decompose()$ as provided in 'statsmodels' library. We plot the trend, seasonal and residual components of wind speed (Figures 2.11, 2.12, 2.13, and 2.14) and GHI (Figures 2.15, 2.16, 2.17, and 2.18) from Pokhran, Rajasthan. Similar analysis has also been performed for the other locations. However, the graphs are not included here intentionally due to consistent observations. This decomposition provides a structured way of thinking about the current time series forecasting problem, both in terms of modeling complexity and specifically in terms of how to best capture each of these components in a given forecasting model. The following observations are made from the time series decomposition plots.

1. There is no major upward or downward trend in the datasets at the four studied timescales.
2. The long term yearly seasonal pattern accounts for the seasonal lag at high frequency (12 for monthly, 56 for weekly, 365 for daily and 365×24 for hourly data according to number of months, weeks, days and hours in a year).
3. The residual plots suggest that there is not much noise in the data.
4. The data are distributed around the mean values.

Since trend and seasonality may affect the value of the time series at different times, causing non-stationarity of the time series, the above observations bring our attention towards a test for the stationarity of the considered datasets as described in the next section.

2.5 Test for Stationarity

The concept of stationarity of a time series can be visualized as a form of statistical equilibrium [47]. This means that the statistical properties such as mean and variance of a stationary process do not depend upon time. The stationarity is a necessary condition for building a time series model for forecasting. In addition, the mathematical complexity of the fitted models significantly reduces with this property. There are two types of stationary processes as defined below.

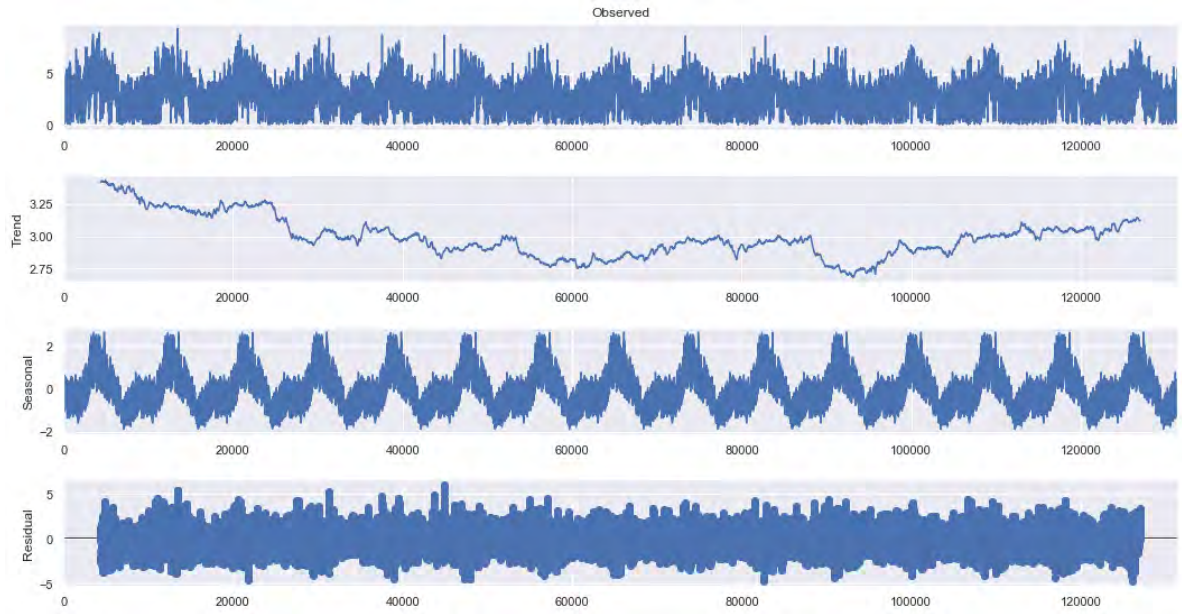


Fig. 2.11: Additive time series decomposition of hourly wind speed data from Pokhran, Rajasthan.

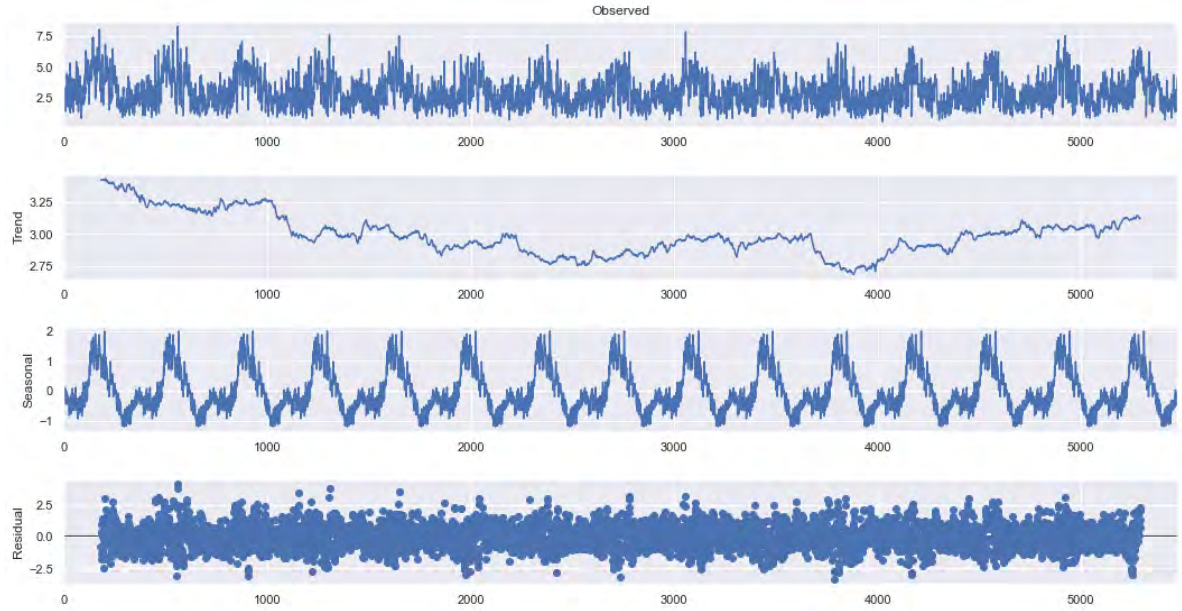


Fig. 2.12: Additive time series decomposition of daily wind speed data from Pokhran, Rajasthan.

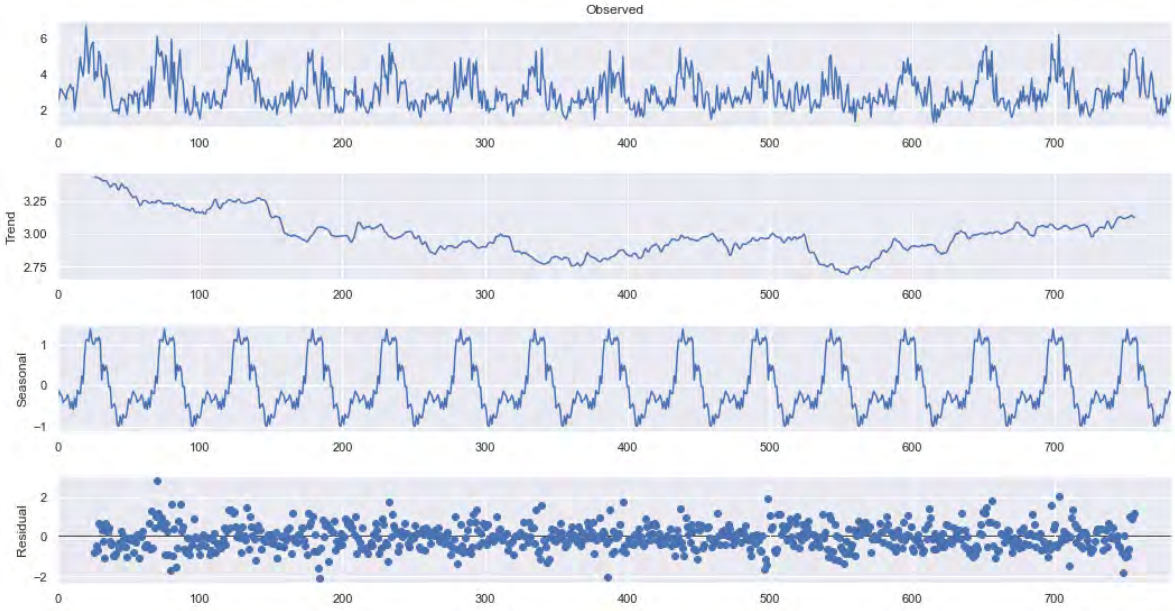


Fig. 2.13: Additive time series decomposition of weekly wind speed data from Pokhran, Rajasthan.

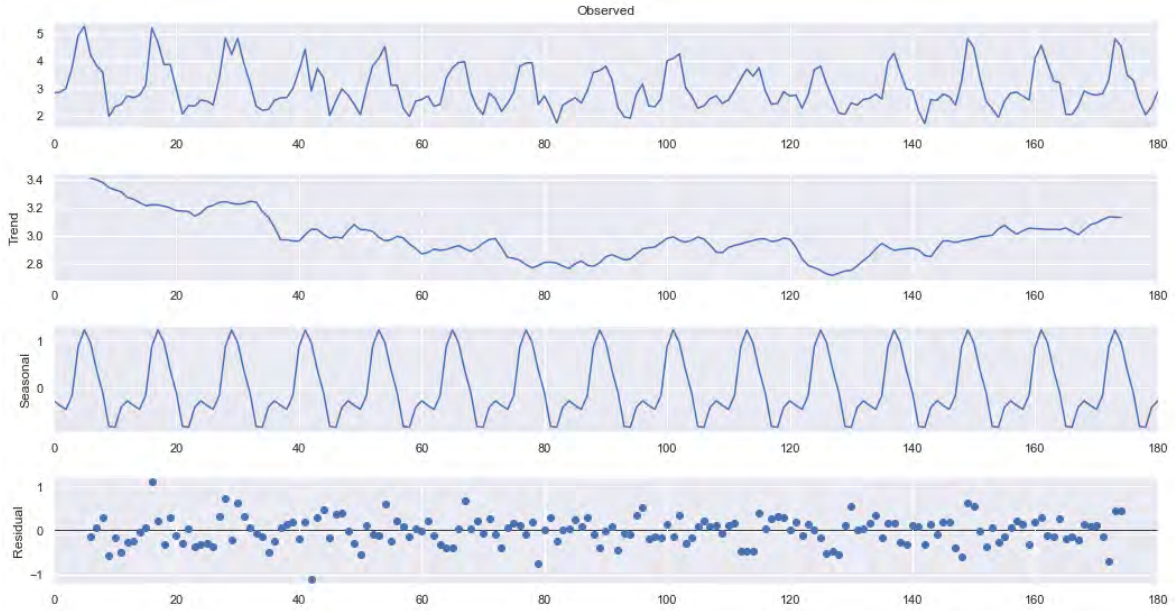


Fig. 2.14: Additive time series decomposition of monthly wind speed data from Pokhran, Rajasthan.

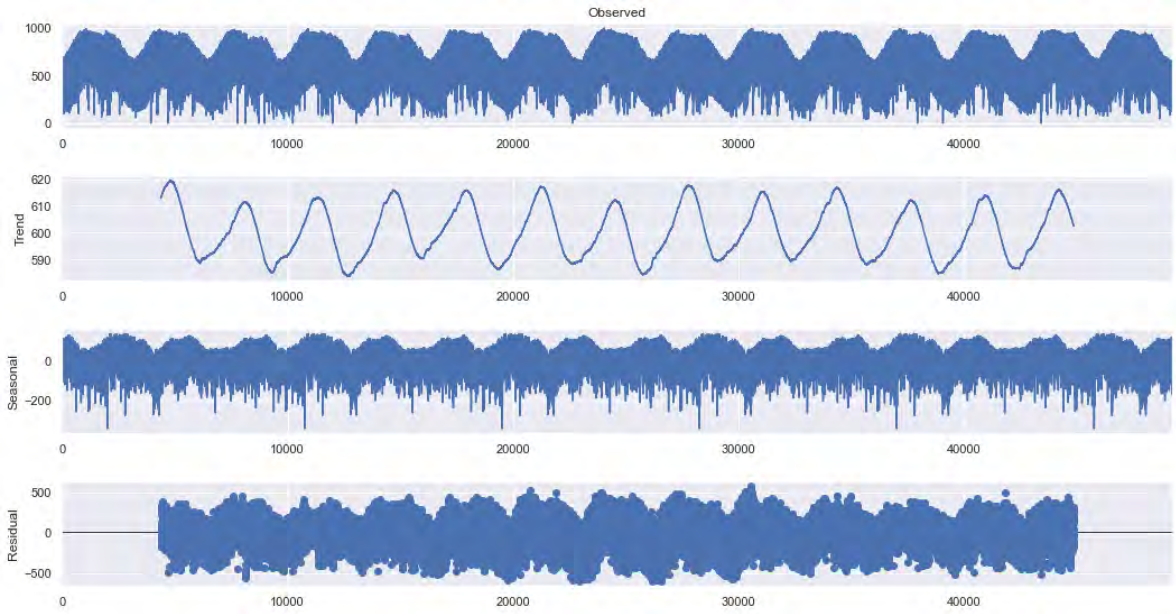


Fig. 2.15: Additive time series decomposition of hourly GHI data from Pokhran, Rajasthan.

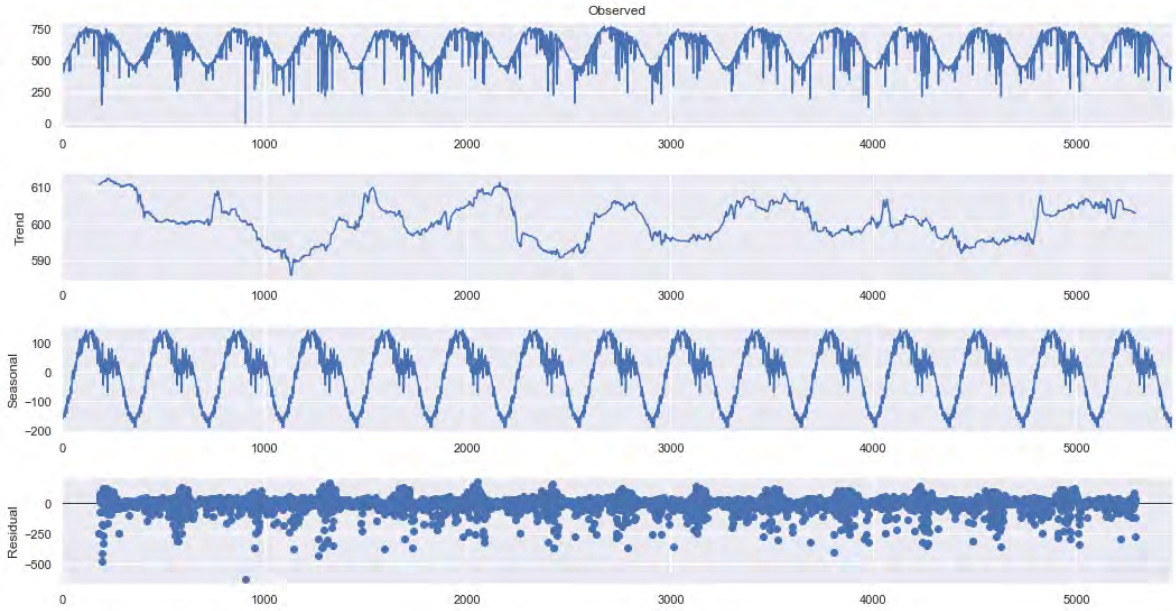


Fig. 2.16: Additive time series decomposition of daily GHI data from Pokhran, Rajasthan.

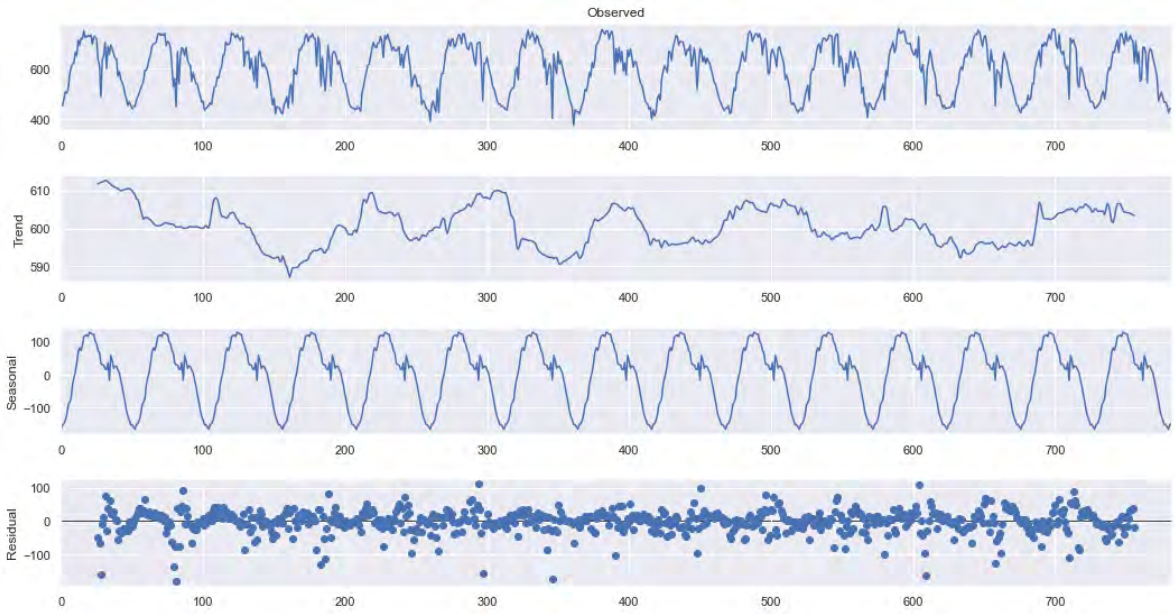


Fig. 2.17: Additive time series decomposition of weekly GHI data from Pokhran, Rajasthan.

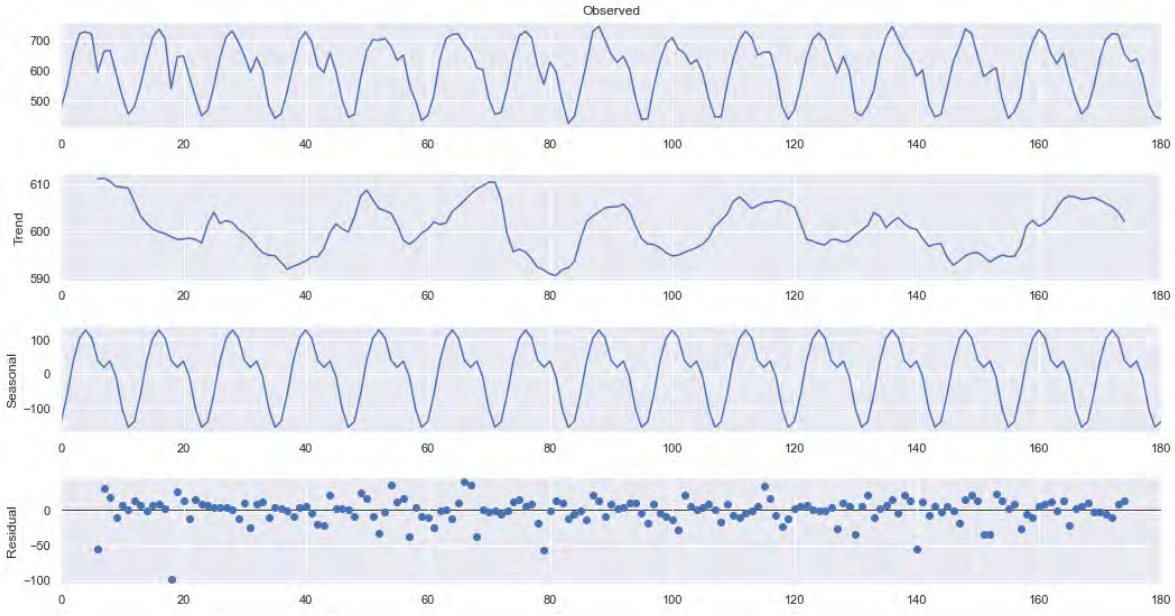


Fig. 2.18: Additive time series decomposition of monthly GHI data from Pokhran, Rajasthan.

1. A process $X(t)$, $t = 0, 1, 2, ...$ is strongly stationary or strictly stationary if the joint probability distribution function of $X(t-s)$, $X(t-s+1)$, ..., $X(t)$, ... $X(t+s-1)$, $X(t+s)$ is independent of t for all s.

2. The time series, $X(t)$ is weakly stationary if
   (i) $\mu(X(t))$ is independent of t, and
   (ii) $\gamma(X(t+h), X(t))$ is independent of t for each h.

Here, $X(t)$ has $E(X^2(t)) < \infty$. The mean function $\mu(X(t))$ and covariance function $\gamma(X(t+h),(t))$ of $X(t)$ are as follows.

$$\mu(X(t)) = E(X(t)) \tag{2.5}$$

$$\gamma(X(r,s)) = Cov(X(r), X(s)) = E\Big[\big(X(r) - \mu X(r)\big)\big(X(s) - \mu X(s)\big)\Big] \tag{2.6}$$

For practical applications, as the assumption of strong stationarity is not always needed, the weak stationary time series is referred as stationary in our study [121].

On visualization of the datasets and decomposition plots in the previous section, we have inferred that there is no significant trend in the datasets. Also, the data points are distributed around the mean value. However, there is a clear yearly seasonality on observing the whole data. These observations provide us an indication for stationarity in the data. For a stringent conclusion, we perform the Augmented Dickey Fuller (ADF) test for both wind speed and GHI values. The hypotheses are as follows:

H_0: The series has a unit root.
H_1: The series has no unit root.

If the null hypothesis (H_0) is rejected, it implies that the time series does not have a unit root, indicating its stationary behavior.

Table 2.7: Results of the ADF test for wind speed datasets

Study Site		**Hourly**	**Daily**	**Weekly**	**Monthly**
Rajasthan	ADF Statistic	−22.4877	−6.3609	−7.8903	−2.6286
	$p-$Value	0.0	2.4753e−08	4.4541e−12	0.0872
Gujarat	ADF Statistic	−18.4697	−5.9226	−8.6577	−2.2205
	$p-$Value	2.1407e−30	2.4853e−07	4.9156e−14	0.1988
Karnataka	ADF Statistic	−13.6337	−7.2096	−9.6200	−2.3644
	$p-$Value	1.7040e−25	2.2527e−10	1.7190e−16	0.1520
Telangana	ADF Statistic	−16.5376	−6.1220	−8.9989	−2.3251
	$p-$Value	1.9969e−29	8.8069e−08	6.5754e−15	0.1639

Table 2.8: Results of the ADF test for GHI datasets

Study Site		Hourly	Daily	Weekly	Monthly
Rajasthan	ADF Statistic	−7.5940	−5.5024	−9.0502	−4.1462
	$p-$Value	2.4868e−11	2.0565e−06	4.8607e−15	0.0008
Gujarat	ADF Statistic	−9.9647	−6.6878	−8.2157	−3.1746
	$p-$Value	2.3234e−17	4.1830e−09	6.6314e−13	0.0215
Karnataka	ADF Statistic	−14.8208	−6.9665	−8.2521	−3.8308
	$p-$Value	1.9644e−27	8.8847e−10	5.3542e−13	0.0026
Telangana	ADF Statistic	−14.4409	−9.2684	−10.2436	−3.4147
	$p-$Value	7.3600e−27	1.3471e−15	4.6794e−18	0.0104

The results of the ADF test are summarized in Table 2.7 and Table 2.8. It is observed that the values of ADF statistic are significantly lesser than the critical values -2.8615, -2.8620, -2.8653, and -2.6286 for hourly, daily, weekly, and monthly wind speed data, respectively. Consequently, the $p-$value in each case is much lesser than the considered significance level, $\alpha = 5\%$. Therefore, we reject the null hypothesis at $\alpha = 5\%$ and confirm stationarity of wind speed across all four timescales. Like wind speed, the hourly, daily, and weekly GHI datasets are also noted to be stationary. However, in the monthly GHI data, there is a weak evidence against the null hypothesis (the time series has a unit root), indicating its non-stationary behavior. This information is useful for deciding the parameters for the implementation of time series models in Chapter 3.

To further analyze wind speed and GHI characteristics at the selected study sites, we fit probability distributions to the datasets, as discussed in the next section.

2.6 Distribution Fitting

As the renewable energy data at various sites is subject to varying climate fluctuations, use of statistical distributions is crucial to study the characteristics of wind speed and solar irradiance. In this direction, some well-known distributions, including exponential, Weibull, Rayleigh, generalized extreme value, gamma, normal, lognormal, logistic, log-logistic, and inverse Gaussian have been used to model wind speed and power density distributions. For instance, Alyat et al. [5] applied 10 distribution functions to explore wind speed characteristics for eight locations distributed over the Northern part of Cyprus. Ouarda et al. [95] investigated wind speed characteristics of nine stations in United Arab Emirates using 11 distribution functions. They used the methods of maximum likelihood, moments, and L-moment to estimate the parameters

Table 2.9: List of reference probability distributions with their pdf and domain information

	Density Function		Parameters	
Distribution	**PDF**	**Domain**	**Role**	**Domain**
Exponential	$\frac{1}{\alpha}e^{-\frac{t}{\alpha}}$	$t>0$	α - scale,	$\alpha>0$
Gamma	$\frac{1}{\Gamma(\beta)}\frac{t^{\beta-1}}{\alpha^{\beta}}e^{-\frac{t}{\alpha}}$	$t>0$	α-scale, β-shape	$\alpha>0$, $\beta>0$
Lognormal	$\frac{1}{t\beta\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{lnt-\alpha}{\beta}\right)^2\right]$	$t>0$	α - log-scale, β-shape	$-\infty<\alpha<\infty$, $\beta>0$
Weibull	$\frac{\beta}{\alpha^{\beta}}t^{\beta-1}e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$	$t>0$	α-scale, β-shape	$\alpha>0$, $\beta>0$
Exponentiated Weibull	$\frac{\beta\gamma}{\alpha}\left(\frac{t}{a}\right)^{\beta-1}e^{-\left(\frac{t}{\alpha}\right)^{\beta}}\left(1-e^{-\left(\frac{t}{\alpha}\right)^{\beta-1}}\right)^{\gamma-1}$	$t>0$	α-scale, β-shape, γ-shape	$\alpha>0$, $\beta>0$, $\gamma>0$

of these distribution functions. The results indicated that two-parameter Weibull, Kappa distribution, and generalized Gamma distribution generally provide the best fit to the wind speed data across all heights and locations. Masseran [78] studied the distribution of wind power density at six stations in Malaysia using Weibull, gamma, and inverse gamma density functions. They used the maximum likelihood method to estimate the parameters of the models. It was found that the Weibull and gamma distributions are able to provide a good approximation of the observed wind speed data for each station. Thus, these distributions are useful for estimating wind energy potential in the studied sites. Pasari et al. [98] carried out the best fit probability model using eight popular probability distributions of solar irradiation data from Charanka Solar Park, Gujarat. Recently, Khamees et al. [61] tried to model wind speed and solar irradiance using mixture of distributions generated from the integration of the original Weibull, lognormal, gamma, and inverse Gaussian distributions. The results show that the mixture of probability density functions provides better fitting criteria for wind speed and solar irradiance than the original distribution functions. In view of the above discussed studies, we have implemented five probability distributions, namely exponential, gamma, lognormal, Weibull, and exponentiated Weibull for the GHI and wind speed datasets collected at four locations in India. The data are further sampled at hourly, daily, weekly, and monthly timescales. The reference probability density functions and associated domain information are listed in Table 2.9. The detailed description along with mathematical expressions for these distributions is provided in the next section.

2.6.1 Description of Probability Models

In this section, we provide a detailed description of each of the studied probability distributions. Also we briefly discuss the genesis of these distributions, their model properties, and interrelationships among themselves. The usual domains for all distributions are the whole positive real line. It is observed that a number of distributions (e.g., gamma, Weibull, and exponentiated Weibull) coincide with exponential distribution, when the shape parameters $\beta = 1$ and $\gamma = 1$. This implies that these distributions are generalizations or extensions of classical one-parameter exponential distribution. Renewable energy resources, especially wind speed, have been modeled using the Weibull distribution throughout the globe. This distribution fit relates to how often winds of different speeds will be observed at a location with a certain average (mean) value. Knowing this helps to choose a wind turbine with the optimal cut-in speed (the wind speed at which the turbine starts to generate usable power) and cut-out speed (the speed at which the turbine hits the limit of its alternator and can no longer put out increased power output with further increase in wind speed). In addition, it is observed [46] that the exponentiated Weibull distribution shares many physical properties (e.g., shapes of density function) with gamma, lognormal, and Weibull models. Thus, the exponentiated Weibull could be a potential model to represent the renewable energy datasets. Besides, the heavy-tailed (tail is thicker than that of exponential model) distributions, namely lognormal and Weibull ($\beta < 1$) are also implemented. The class of heavy-tailed distributions offers insights in appraising the characteristics of highly variable stochastic processes, such as the process of renewable energy [58, 111]. These distributions, though in some sense replicate the exponential distribution, are more explanatory in various applications, such as medical statistics, finance, and natural disasters.

Any probability distribution, in general, is controlled by the location, scale, and shape parameters. Each of these parameters plays different roles in modeling. For instance, the location parameter is primarily used to define the location or shift of the distribution. It first determines the range of the distribution and then accordingly shifts the probability graph to the left or to the right along the abscissa. On the other hand, the scale parameter determines the spread of the distribution and the shape parameter determines the appearance or shape of the distribution [88]. Among all these three parameters, the shape parameter usually serves as the most important tool in depicting the future behavior and monotonic property of a distribution [58]. The details and impact of the parameters of considered distributions are explained one by one in the below subsections. At this point, it may be noted that the use of the location parameter is restricted in the present study, as it often hampers the regularity conditions, existence, and uniqueness in the maximum likelihood estimation (MLE) method [58]. The root cause for such circumstances is the domain dependency of the underlined process on the location parameter. Nevertheless, to alleviate these difficulties, the modified MLE (MMLE) method, in which the

location parameter is estimated beforehand, may be performed [58, 88, 97].

2.6.1.1 Exponential Distribution

The exponential distribution has widespread applicability in the field of reliability engineering, survival analysis, chemical engineering, hydrology, psychology, finance, and other areas [58, 96]. The exponential distribution is the only continuous distribution that enjoys memoryless property. A random variable T has an exponential distribution if T has the following distribution function.

$$F(t;\alpha) = 1 - e^{-\frac{t}{\alpha}} \quad (t > 0, \alpha > 0) \tag{2.7}$$

If T has a distribution function given as in Equation (2.7), then the corresponding density function is obtained as

$$f(t;\alpha) = \frac{1}{\alpha} e^{-\frac{t}{\alpha}} \quad (t > 0, \alpha > 0) \tag{2.8}$$

The mean of an exponential distribution can be completely explained by a single parameter α (known or estimated); this fact also illustrates the simplicity of the exponential distribution.

2.6.1.2 Gamma Distribution

The gamma distribution belongs to the family of the two-parameter continuous probability distributions [58]. This distribution is very popular in modeling waiting times (e.g., occurrence of an earthquake and death of a patient). This distribution is controlled by two parameters: $\alpha-$ the scale parameter and $\beta-$ the shape parameter; $(\alpha,\beta) \in \Theta \subset {R_+}^2$, where Θ is an open set and $R_+ = (0,\infty)$.

If a random variable T follows gamma distribution, then the pdf and the cdf are given as

$$\begin{aligned} f(t;\alpha,\beta) &= \frac{1}{\Gamma(\beta)} \frac{t^{\beta-1}}{\alpha^{\beta}} e^{\frac{-t}{\alpha}} \quad (t > 0, \alpha > 0, \beta > 0) \\ F(t;\alpha,\beta) &= \frac{\gamma(\beta, \frac{t}{\alpha})}{\Gamma(\beta)} \quad (t > 0, \alpha > 0, \beta > 0) \end{aligned} \tag{2.9}$$

Here $\Gamma(.)$ is the gamma function and $\gamma(.,.)$ is the lower incomplete gamma function. These are defined [58] as

$$\begin{aligned} \Gamma(x) &= \int_0^{\infty} z^{x-1} e^{-z} dz \quad (x > 0) \\ \gamma(x,s) &= \int_0^{s} z^{x-1} e^{-z} dz \quad (x > 0) \end{aligned} \tag{2.10}$$

The gamma distribution assumes different shapes of density function for different values of β. The mean, mode, variance, and skewness of a gamma distribution are $\alpha\beta$, $(\beta-1)\alpha$, $\beta\alpha^2$, and $\frac{2}{\sqrt{\beta}}$, respectively. There is no closed form for the median. In addition, it is easy to

realize from Equation (2.10) that the distribution function of the gamma distribution is difficult to calculate for non-integer shape parameter [58].

2.6.1.3 Lognormal Distribution

The lognormal distribution is a continuous probability distribution of a random variable (X) whose logarithm (lnX) is normally (Gaussian) distributed. The lognormal distribution is sometimes known as anti-lognormal distribution, as this distribution is actually obtained by taking the exponential (anti-log) of a normal random variable. It is a popular member of the class of heavy-tailed distributions. This distribution has drawn considerable attention in agricultural sciences, biological sciences, reliability and lifetime data analysis, quality control, seismology, meteorology, and many other streams of science [58]. The lognormal probability plot (LPP), which often serves as a powerful tool to visualize observed data or simulated data, gives rise to the popularity of lognormal model. Besides, due to the standardization of the lognormal distribution to the normal (Gaussian) distribution, the whole Gaussian theory holds for the lognormal distribution. Let Z be a random variable that follows a normal distribution, then $T = e^Z$ follows a lognormal distribution. In notation, if $Z \sim N(\alpha, \beta)$ then $T = e^Z \sim LN(\alpha, \beta)$. The pdf and the cdf of a random variable $T \sim LN(\alpha, \beta)$ is

$$\begin{aligned} f(t;\alpha,\beta) &= \frac{1}{t\beta\sqrt{(2\pi)}} exp\left(-\frac{1}{2}\left(\frac{lnt-\alpha^2}{\beta}\right)\right) \quad (t>0, \beta>0) \\ F(t;\alpha,\beta) &= \Phi\left(\frac{lnt-\alpha}{\beta}\right) \quad (t>0, \beta>0) \end{aligned} \tag{2.11}$$

$\Phi(.)$ in the above equations denotes the cdf of a standard normal distribution and is defined as

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{\left(-\frac{x^2}{2}\right)} dx \tag{2.12}$$

The lognormal distribution has two parameters: α $(-\infty < \alpha < \infty)$ is the log-scale parameter and $\beta(\beta > 0)$ is the shape parameter. The parameters α and β are also the mean and the standard deviation of the underlying normal variable. All lognormal density functions are skewed (in the right) and unimodal [58]. The mean, mode, variance, and skewness of lognormal distribution are $exp\left(\alpha + \frac{\beta^2}{2}\right)$, $exp\left(\alpha - \beta^2\right)$, $\left(exp(\beta^2) - 1\right)\left(exp(2\alpha + \beta^2)\right)$, and $\left(exp(\beta^2) + 2\right)\left(\sqrt{exp(\beta^2) - 1}\right)$, respectively.

2.6.1.4 Weibull Distribution

The Weibull distribution belongs to the family of continuous probability distributions. The Swedish physicist Wallodi Weibull in 1939 first proposed this famous distribution while seeking a formula for the failure rate of welds [151, 152]. Since its inception, Weibull distribution has been the most popular and versatile probability distribution in the field of reliability and survival analysis. Besides, it has been widely used in numerous fields, such as engineering psychology, seismology, hydrology, business, economics, renewable energy, and ecology [88]. The prime advantage of Weibull model is the ability to provide a reasonable prediction even with an extremely small sample size. Moreover, the pdf can assume a wide variety of shapes. This, in turn, has made this distribution very popular to the scientific community. A two-parameter Weibull model is more advantageous over the classical two-parameter normal distribution mainly for two reasons. First, computation of cumulative Weibull distribution function is very straight forward unlike cumulative normal distribution function which requires integral approximations. Second, the shape of normal distribution, for any choice of parameter values, always is of similar type (bell shape), whereas, just by tweaking the shape parameter (β), Weibull can provide distinctive shapes as required for an analyst. In particular, the skewness index for Weibull distribution equals zero for $\beta \approx 3.6$, and thus in the vicinity of $\beta = 3.6$, this looks similar in shape to a normal distribution [88]. The Weibull distribution is quite generic in a sense that it includes exponential $(\beta = 1)$ and Rayleigh distributions $(\beta = 2)$ as special cases. The basic model properties of a random variable T drawn from a two-parameter Weibull model are given below.

$$\begin{aligned} f(t;\alpha,\beta) &= \frac{\beta}{\alpha^{\beta}} t^{(\beta-1)} exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right) \quad (t>0, \alpha>0, \beta>0) \\ F(t;\alpha,\beta) &= 1 - exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right) \quad (t>0, \alpha>0, \beta>0) \end{aligned} \tag{2.13}$$

The Weibull distribution is controlled by a shape parameter (β) and a scale parameter (α). The density function of the model is monotonically decreasing for $\beta \leq 1$. For $\beta > 1$, it becomes unimodal with mode at $\alpha\left(1-\frac{1}{\beta}\right)^{\frac{1}{\beta}}$. The mean, median, and variance are given by $\alpha\Gamma\left(1+\frac{1}{\beta}\right)$, $\alpha(ln\ 2)^{\frac{1}{\beta}}$, and $\alpha^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \left\{\Gamma\left(1+\frac{1}{\beta}\right)\right\}^2\right]$, respectively.

2.6.1.5 Exponentiated Weibull Distribution

Mudholkar and Srivastava [84] first proposed exponentiated Weibull distribution by introducing an additional shape parameter to the standard two-parameter Weibull distribution. Their concept led to the discovery of the general class of exponentiated distributions proposed by

Lehmann et al. (1955) [45] as $F(t) = [G(t)]^{\beta}$, where $G(t)$ is the base distribution and $\beta > 0$ is a shape parameter. The beauty and importance of the exponentiated Weibull family lie in its capability to accommodate monotone as well as non-monotone hazard functions, such as the unimodal shaped and bathtub shaped [84, 85]. Since its inception, the exponentiated Weibull and its variants have been applied to a wide range of practical applications, such as flood-data analysis, bus-motor failure [85], human mortality testing[13], survival analysis of head and neck cancer patients [85], excess-of-loss insurance data analysis [23], wind speed analysis, and tree-diameter prediction [149]. However, it appears that the suitability of exponentiated Weibull has not much been explored yet in the field of renewable energy modeling. The pdf and the cdf of the exponentiated Weibull random variable T are given as

$$\begin{aligned} f(t;\alpha,\beta,\gamma) &= \frac{\beta\gamma}{\alpha}\left(\frac{t}{\alpha}\right)^{(\beta-1)} e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \left(1-e^{-\left(\frac{t}{\alpha}\right)^{\beta}}\right)^{\gamma-1} (t>0,\alpha>0,\beta>0,\gamma>0) \\ F(t;\alpha,\beta,\gamma) &= \left(1-e^{-\left(\frac{t}{\alpha}\right)^{\beta}}\right)^{\gamma} (t>0,\alpha>0,\beta>0,\gamma>0) \end{aligned} \tag{2.14}$$

Here α is a scale parameter, and β, γ are the shape parameters. If γ is a positive integer, then the exponentiated Weibull family becomes a particular member of the Lehmann alternatives. From Equation (2.14), a number of special distributions can be obtained, such as exponential ($\beta = 1, \gamma = 1$), Rayleigh ($\beta = 2$, $\gamma = 1$), Weibull ($\gamma = 1$), and exponentiated exponential ($\beta = 1$).

All these parametric models described here are solely controlled by their parameter values. The next section discusses the statistical inference that comprises parameter estimation and model validation using goodness of fit tests. Some popular methods of estimation are the maximum likelihood estimation (MLE) method, expectation-maximization algorithm, and the Bayesian estimations. Similarly, the model validation step includes parametric and non-parametric approaches such as the Akaike information criterion (AIC), Anderson-Darling (A-D) test, Kolmogorov-Smirnov (K-S) test, and the chi-square criterion. In the current study, we adopt MLE for parameter estimation and K-S test for the test of goodness of fit of the implemented distributions. We describe the details of parameter estimation and goodness of fit in the following two subsections.

2.6.2 Parameter Estimation

All parametric models are solely governed by their parameter values. Reliable estimation of model parameters thus becomes very crucial to illustrate the model characteristics. The estimated parameter values depend on the data type (complete or censored) and the estimation

strategy. In general, there are two estimation approaches, namely the point estimation and the interval estimation. Further, for each of these approaches, there are a number of estimation methods which may broadly be categorized into three types – graphical methods, statistical methods, and the combination of the first two methods. In graphical methods, estimates are obtained from the plotting of data, whereas the statistical methods utilize intrinsic data characteristics for estimations. The graphical methods are useful in providing an initial estimate of the parameters which may be refined later using the rigorous statistical estimation procedures. The performance of an estimator is usually realized from a number of favorable properties of estimators, such as unbiasedness, consistency, efficiency, and sufficiency [48]. Statistical methods, in contrast to the graphical methods, are more robust and thus applicable to a wider class of models and data types [58, 88]. The main advantage of statistical methods lies in the fact that it not only estimates the model parameters but also provides an uncertainty measure for the estimated parameters. Nevertheless, the statistical methods are often complicated and encounter rigorous understanding of Linear Algebra, Optimization, and Statistics. From time to time, many different statistical estimation methods have been developed in estimating model parameters. Some of those are maximum likelihood estimation (MLE), method of moments (MoM), percentile estimation, and Bayesian estimations. In this thesis, however, we focus on the MLE method due to its consistency, compactness, and rigorous asymptotic normality results. In 1922, R. A. Fisher formally introduced the method of MLE. The MLE is considered to be the most powerful parameter estimation technique as on today. The basic principle is to maximize the likelihood of the parameters, denoted by $L(\theta|t)$, as a function of the model parameter θ. It may be noted that θ could be a single parameter or a vector of parameters like $\theta = (\theta_1, \theta_2, ..., \theta_p)$, for some integer p. The likelihood function $L(\theta|t)$ is defined as

$$L(\theta|t) = \prod_{i=1}^{n} f(t_i, \theta) \tag{2.15}$$

As logarithm is a one-to-one function, maximization of log-likelihood $ln(L(\theta|t))$ is often preferred for computational ease [58]. Further, it appears that the closed form solution of the maximum likelihood equations is not always available. In such cases, the non-linear equation solver packages of R, Python, MATLAB, and MATHEMATICA may be utilized to obtain the MLE of θ , denoted by $\hat{\theta}$. The mathematical representations for parameter estimation of the considered five probability models are provided below.

2.6.2.1 Exponential Distribution

The log-likelihood function for the exponential distribution is given as

$$lnL(\theta|t) = lnL(\alpha;t_1,t_2,t_3,...,t_n) = -n\ ln\ \alpha - \sum_{i=1}^{n} \frac{t_i}{\alpha} \tag{2.16}$$

The corresponding estimate from the log-likelihood equation is obtained as

$$\frac{\partial}{\partial \alpha} \ln L(\alpha;t_1,t_2,t_3,...,t_n) = 0 \implies \hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{2.17}$$

2.6.2.2 Gamma Distribution

The log-likelihood function of the gamma distribution is given as

$$lnL(\alpha;t_1,t_2,t_3,...,t_n) = -n\beta ln\alpha - nln\Gamma(\beta) - \sum_{i=1}^{n} \frac{t_i}{\alpha} + (\beta-1)\sum_{i=1}^{n} ln(t_i) \tag{2.18}$$

The likelihood equations are

$$\begin{aligned} \frac{\partial}{\partial \alpha} lnL(\alpha;t_1,t_2,t_3,...,t_n) = 0 &\implies \alpha\beta = \frac{1}{n}\sum_{i=1}^{n} t_i \\ \frac{\partial}{\partial \beta} lnL(\alpha;t_1,t_2,t_3,...,t_n) = 0 &\implies ln(\alpha) + \psi(\beta) = \frac{1}{n}\sum_{i=1}^{n} ln(t_i) \end{aligned} \tag{2.19}$$

$\psi(\beta)$ in the above equation denotes the digamma function defined as

$$\psi(\beta) = \frac{d}{dx} ln\Gamma(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)} \tag{2.20}$$

The explicit solution of the above set of equations given in Equations (2.19) and (2.20) is not possible. Numerical approaches (e.g., bracketing method, Newton Raphson method, and fixed point iteration) may be employed to solve it.

2.6.2.3 Lognormal Distribution

The log-likelihood function for the lognormal distribution is given as

$$lnL(\alpha;t_1,t_2,t_3,...,t_n) = -\frac{1}{2\beta^2}\sum_{i=1}^{n} (lnt_i - \alpha)^2 - nln\left(\beta\sqrt{2\pi}\right) - \sum_{i=1}^{n} ln(t_i) \tag{2.21}$$

The associated log likelihood equations are

$$\begin{aligned} \frac{\partial}{\partial \alpha} lnL(\alpha; t_1, t_2, t_3, ..., t_n) = 0 \implies \hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} ln(t_i) \\ \frac{\partial}{\partial \beta} lnL(\alpha; t_1, t_2, t_3, ..., t_n) = 0 \implies (\hat{\beta})^2 = \frac{1}{n}\sum_{i=1}^{n} ln(t_i - \hat{\alpha})^2 \end{aligned} \tag{2.22}$$

For lognormal distribution, an explicit solution for the model parameters can be obtained.

2.6.2.4 Weibull Distribution

The log-likelihood function for the Weibull distribution is given as

$$lnL(\theta|t) = lnL(\alpha, \beta; t_1, t_2, t_3, ..., t_n) = nln\beta - n\beta ln\alpha + (\beta - 1)\sum_{i=1}^{n}(t_i) - \sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} \tag{2.23}$$

The associated log likelihood equations are

$$\begin{aligned} \frac{\partial}{\partial \alpha} lnL(\alpha; t_1, t_2, t_3, ..., t_n) = 0 \implies \alpha^{\beta} - \frac{1}{n}\sum_{i=1}^{n}(t_i)^{\beta} = 0 \\ \frac{\partial}{\partial \beta} lnL(\alpha; t_1, t_2, t_3, ..., t_n) = 0 \implies \frac{n}{\beta} + \frac{1}{n}\sum_{i=1}^{n}\left[1 - \left(\frac{t_i}{\alpha}\right)^{\beta}\right] ln\left(\frac{t_i}{\alpha}\right) = 0 \end{aligned} \tag{2.24}$$

The above set of expressions in Equation (2.24) can be rewritten as

$$\begin{aligned} \frac{1}{\beta} + \frac{1}{n}\sum_{i=1}^{n} ln(t_i) - \frac{\sum_{i=1}^{n}(t_i)^{\beta} ln(t_i)}{\sum_{i=1}^{n}(t_i)^{\beta}} = 0 \\ \alpha = \left(\frac{1}{n}\sum_{i=1}^{n}(t_i)^{\beta}\right)^{\frac{1}{\beta}} \end{aligned} \tag{2.25}$$

Explicit solution of Equation (2.25) cannot be obtained.

2.6.2.5 Exponentiated Weibull Distribution

The log-likelihood function of the exponentiated Weibull distribution is

$$lnL(\alpha, \beta, \gamma; t_1, t_2, ..., t_n) = nln\left(\frac{\beta\gamma}{\alpha^{\beta}}\right) + (\beta - 1)\sum_{i=1}^{n} ln(t_i) - \sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} + (\gamma - 1)\sum_{i=1}^{n} ln\left(1 - e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}\right) \tag{2.26}$$

The MLEs of (α,β,γ), say $(\hat{\alpha},\hat{\beta},\hat{\gamma})$, are the solutions of the log-likelihood equations given as

$$\frac{\partial}{\partial\alpha}lnL=-\frac{\beta n}{\alpha}+\frac{\beta}{\alpha}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta}-(\gamma-1)\frac{\beta}{\alpha}\sum_{i=1}^{n}\frac{\left(\frac{t_i}{\alpha}\right)^{\beta}e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}}{1-e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}}=0 \tag{2.27}$$

$$\frac{\partial}{\partial\beta}lnL=\frac{n}{\beta}+\sum_{i=1}^{n}ln\left(\frac{t_i}{\alpha}\right)-\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta}ln\left(\frac{t_i}{\alpha}\right)+(\gamma-1)\sum_{i=1}^{n}\frac{ln\left(\frac{t_i}{\alpha}\right)\left(\frac{t_i}{\alpha}\right)^{\beta}e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}}{1-e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}} \tag{2.28}$$

$$\frac{\partial}{\partial\gamma}lnL=\frac{n}{\gamma}+\sum_{i=1}^{n}ln\left(1-e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}\right)=0 \tag{2.29}$$

From Equation (2.29), $\hat{\gamma}$ may be obtained as a function of (α,β), as

$$\hat{\gamma}(\alpha,\beta)=-\frac{n}{\sum_{i=1}^{n}ln(1-e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}})} \tag{2.30}$$

subtracting $\frac{\alpha ln\alpha}{\beta}$ times of Equation (2.27) from Equation (2.28) gives the following equation

$$\frac{n}{\beta}+\sum_{i=1}^{n}ln(t_i)-\beta\left[(\gamma-1)\sum_{i=1}^{n}\frac{(t_i)^{\beta}e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}}{1-e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}}-\sum_{i=1}^{n}(t_i)^{\beta}ln(t_i)\right]=0 \tag{2.31}$$

Using Equation (2.30), γ is eliminated from Equations (2.27) and (2.31). This essentially gives two equations in two unknowns as given in Equation (2.32). Besides, for simplicity, let us assume $\delta=\alpha^{\beta}$. Then the above equation reduces to

$$\begin{aligned} n\delta+\beta\delta\sum_{i=1}^{n}ln(t_i)-\beta\left[\left(\frac{n}{k_1}+1\right)k_2+k_3\right]&=0\\ n\delta-k_4-\left(\frac{n}{k_1}+1\right)k_5&=0\\ \gamma+\frac{n}{k_1}&=0 \end{aligned} \tag{2.32}$$

Expressions for $k_i,\ i = 1,\ 2,\ ...,5$ in the above equations are as follows

$$\begin{aligned} k_1 &= \sum_{i=1}^{n} ln\left(1 - e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}\right) \\ k_2 &= \sum_{i=1}^{n} \frac{e^{-\frac{t_i^{\beta}}{\delta}} t_i^{\beta} ln(t_i)}{1 - e^{-\frac{t_i^{\beta}}{\delta}}} \\ k_3 &= \sum_{i=1}^{n} t_i^{\beta} ln(t_i) \\ k_4 &= \sum_{i=1}^{n} t_i^{\beta} \\ k_5 &= \sum_{i=1}^{n} \frac{e^{-\frac{t_i^{\beta}}{\delta}} t_i^{\beta}}{1 - e^{-\frac{t_i^{\beta}}{\delta}}} \end{aligned} \tag{2.33}$$

Closed form solutions of Equation (2.32) are not available. Nevertheless, non-linear equation solver packages of R, Python, MATLAB, MATHEMATICA, NAG routines, C05NCF and C05PBF, and IMSL routine DNEQNF provide direct solution of the above set of transcendental equations.

2.6.3 Goodness of Fit

Among several competing models, it is obvious to choose the model(s) that represents the given data more meaningfully. In the present study, a well established model selection criterion, namely the Kolmogorov-Smirnov (K-S) minimum distance criterion is adopted to appraise the comparative fitness of the studied distributions. For notational simplicity, the basic theory is described with the help of two competing models. However, the same concept can be easily extended to an arbitrary number of competing models. Suppose there are two families, say, $F = \{f(t;\hat{\theta}); \hat{\theta} \in R^p\}$ and $G = \{g(t;\hat{\psi}); \hat{\psi} \in R^q\}$, where $\hat{\theta}$ is the estimated (not necessarily by MLE) value of a $p-$dimensional vector of real-valued parameter $\theta = (\theta_1, \theta_2, ..., \theta_p)$ and similarly q is the estimated value of a $q-$dimensional vector of real-valued parameter $\psi = (\psi_1, \psi_2, ..., \psi_q)$. The aim is to choose the best family for given dataset $\{t_1, t_2, ..., t_n\}$.

The Kolmogorov-Smirnov (K-S) minimum distance criterion prioritizes the competing models based on their closeness to the empirical cumulative distribution function (ECDF) of the sample data $\{t_1, t_2, ..., t_n\}$. The empirical distribution function H for n i.i.d. random variables $T_1, T_2, ..., T_n$ is calculated as

$$H_n(t) = \frac{1}{n} \sum_{i=1}^{n} I_{T_i \leq t} \tag{2.34}$$

Here, $I_{T_i \leq t}$ is the indicator function that equals to 1, if $T_i \leq t$ and otherwise, it equals to 0. This makes $H_n(t)$ a step function. For two competitive models, say F and G, the corresponding K-S distances are calculated as

$$\begin{aligned} D_1 &= \sup_{-\infty<t<\infty} |H_n(t) - F(t)| \\ D_2 &= \sup_{-\infty<t<\infty} |H_n(t) - G(t)| \end{aligned} \tag{2.35}$$

In the above expression, $\sup_t$ denotes the supremum of the set of distances. If $D_1 < D_2$, model F is chosen, otherwise model G. Any available statistical packages could be employed to calculate K-S values. It appears that the K-S test has several advantages than the other model selection criteria – (i) no restriction on the selection of parameter estimation methods, unlike maximum likelihood criterion, where model parameters necessarily need to be estimated from MLE, (ii) unlike chi-square criterion, this method does not require any user inputs and thus yields unbiased decision, and (iii) the K-S test is distribution free, so it avoids the use of special tables for each distribution [58, 88].

2.6.4 Results

In this section, we present the results in terms of estimated parameters of considered five probability distributions and the associated K-S statistic values for model prioritization. The estimated parameters of the employed distributions for four study sites of wind speed data are tabulated in Tables 2.10, 2.11, 2.12, and 2.13, whereas the estimated parameters corresponding to GHI datasets are tabulated in Tables 2.14, 2.15, 2.16, and 2.17.

Regarding the wind speed, Table 2.10 shows that the exponentiated Weibull produces the minimum K-S value for hourly (0.0043) and daily (0.0143), whereas lognormal distribution has the minimum K-S value for weekly (0.0327) and monthly (0.1250) data in Rajasthan. In Gujarat (Table 2.11), the exponentiated Weibull and gamma distributions produce the minimum K-S values for hourly (0.0319) and daily (0.0162) wind speed. However, the K-S value of exponentaited Weibull (0.0177) is close to that produced by gamma distribution in daily timescale; lognormal distribution has the least K-S values (0.0327 and 0.1126) for weekly and monthly datasets, respectively. For the wind speed in Karnataka (Table 2.12), the minimum K-S values and the associated distributions are gamma for hourly (0.0401), lognormal for daily (0.0421), and exponentiated Weibull for both weekly (0.0359) and monthly. In Telangana (Table 2.13), the exponentiated Weibull produces the minimum K-S values 0.0330, 0.0252, and 0.0329 for hourly, daily, and weekly data, respectively. For the monthly data, the lognormal distribution has the least K-S value (0.0701) which is also very close to the K-S value of the exponentiated Weibull (0.0705) distribution. However, to ascertain the overall fitness of these distributions to

the empirical distribution functions (both pdf and cdf), a number of plots are generated (Figures 2.19, 2.20, 2.21, and 2.22). These figures clearly depict that the exponentiated Weibull distribution has very close association with the original wind speeds across time and location.

In case of GHI, the exponentiated Weibull distribution produces the minimum K-S values 0.0187, 0.0338, 0.0388, and 0.0388 at hourly timescale across all the four locations. For the daily GHI, the exponentiated Weibull has the minimum K-S value (0.0737) in Rajasthan, whereas the Weibull distribution has the least K-S values 0.0548, 0.0524, and 0.0514 in other three locations. In the weekly GHI, the exponentiated Weibull has the minimum K-S values 0.0758, 0.050, and 0.0482 in Rajasthan, Karnataka, and Telangana, respectively. For Gujarat, the lognormal has the least K-S value for weekly GHI. In case of monthly timescale, the exponentiated Weibull distribution produces the minimum K-S values 0.0902, 0.0898, and 0.0871 in Rajasthan, Gujarat, and Karnataka, respectively, whereas, in Gujarat, the gamma distribution produces the minimum K-S value (0.0761). In order to assess the overall fitness of these competitive distributions to the empirical distribution functions (pdfs and cdfs), we provide comparative plots in Figures 2.23, 2.24, 2.25, and 2.26. Overall, we notice that the exponentiated Weibull distribution has comparatively better fit in GHI data, though none of them has a satisfactory representation to the original values.

In addition to the above description, we also provide a ranking strategy for wind speed and GHI datasets in Tables 2.18 and 2.19. We note that the exponentiated Weibull has been ranked 'first' for the maximum number of times for both wind speed and GHI values. This highlights its efficacy in pattern recognition of renewable energy datasets.

Table 2.10: Distributions fitted to wind speed dataset from Rajasthan at four different timescales

Distribution	**Timescale**	**Parameters (Shape, Scale)**	**K-S Statistic**	$p-$**Value**
Exponential	Hourly	3.0109	0.2365	0
	Daily	3.0109	0.3099	0
	Weekly	3.0104	0.4098	$1.9017e{-}119 \sim 0$
	Monthy	3.0070	0.4594	$1.8231e{-}35 \sim 0$
Gamma	Hourly	3.5456, 0.8492	0.0417	$4.4304e{-}199 \sim 0$
	Daily	5.9434, 0.5066	0.0204	0.0204
	Weekly	10.7201, 0.2808	0.0644	0.0028
	Monthy	16.4616, 0.1826	0.1103	0.0225
Lognormal	Hourly	0.6039, 2.5978	0.0781	0
	Daily	0.4237, 2.7615	0.0250	0.0020
	Weekly	0.3045, 2.8712	0.0457	0.0726

	Monthy	0.2444, 2.9161	0.0937	0.0776
Weibull	Hourly	2.1687, 3.3978	0.0058	0.0002
	Daily	2.5856, 3.3976	0.0467	7.6460e−11 ~ 0
	Weekly	3.2490, 3.3550	0.0953	1.1978e−06 ~ 0
	Monthly	4.0083, 3.3085	0.1434	0.0010
Exponentiated Weibull	Hourly	0.8954, 2.3123, 3.5500	0.0043	0.0137
	Daily	4.9052, 1.2652, 1.6116	0.0143	0.2088
	Weekly	85.9253, 0.7895, 0.3843	0.0267	0.6215
	Monthly	864.5458, 0.6773, 0.1567	0.0651	0.4090

Table 2.11: Distributions fitted to wind speed dataset from Gujarat at four different timescales

Distribution	**Timescale**	**Parameters (Shape, Scale)**	**K-S Statistic**	$p-$**Value**
Exponential	Hourly	3.5386	0.3017	0
	Daily	3.5384	0.3451	0
	Weekly	3.5379	0.3855	3.0860e−105 ~ 0
	Monthy	3.5334	0.4081	8.9441e−28 ~ 0
Gamma	Hourly	5.0901, 0.6952	0.0691	0
	Daily	5.9434, 0.5066	0.0162	0.1106
	Weekly	10.9495, 0.3231	0.0502	0.0367
	Monthy	14.6894, 0.2405	0.1293	0.0042
Lognormal	Hourly	0.4975, 3.1973	0.1026	0
	Daily	0.3750, 3.3087	0.0174	0.0714
	Weekly	0.3063, 3.3776	0.0327	0.3635
	Monthy	0.2646, 3.4138	0.1126	0.0186
Weibull	Hourly	2.6576, 3.9759	0.0321	1.4451e−118~ 0
	Daily	2.9628, 3.9674	0.0509	8.8744e−13 ~ 0
	Weekly	3.4292, 3.9322	0.0882	9.4101e−06 ~ 0
	Monthly	4.1696, 3.8893	0.1625	0.0001
Exponentiated Weibull	Hourly	0.9735, 2.6981, 4.0113	0.0319	4.4288e−117~ 0
	Daily	3.7678, 1.6044, 2.3622	0.0177	0.0626
	Weekly	8.8810, 1.4077,1.7267	0.0359	0.2572
	Monthly	6.3336, 1.8347, 2.2034	0.1250	0.0063

Table 2.12: Distributions fitted to wind speed dataset from Karnataka at four different timescales

Distribution	Timescale	Parameters (Shape, Scale)	K-S Statistic	$p-$Value
Exponential	Hourly	3.1730	0.2775	0
	Daily	3.1727	0.3186	0
	Weekly	3.1723	0.3722	7.0589e$-$98 ~ 0
	Monthy	3.1641	0.4335	1.8894e$-$31 ~ 0
Gamma	Hourly	4.0580, 0.7819	0.0401	1.1225e$-$184~ 0
	Daily	5.9434, 0.5066	0.0660	3.2274e$-$21~ 0
	Weekly	7.0192, 0.4519	0.1214	1.6038e$-$10~ 0
	Monthy	9.5662, 0.3307	0.1641	9.9677e$-$05~ 0
Lognormal	Hourly	0.5464, 2.7911	0.0701	0
	Daily	0.4307, 2.8940	0.0421	6.5932e$-$09~ 0
	Weekly	0.3744, 2.9492	0.0980	5.2080e$-$07~ 0
	Monthy	0.3178, 3.0002	0.1481	0.0006
Weibull	Hourly	2.2380, 3.5863	0.0564	0
	Daily	2.4369, 3.5896	0.0976	6.7530e$-$46 ~ 0
	Weekly	2.6255, 0, 3.5801	0.1387	1.2769e$-$13 ~ 0
	Monthly	3.0466, 0, 3.5450	0.1723	3.5842e$-$05
Exponentiated Weibull	Hourly	1.5183, 1.7934, 2.9881	0.0424	7.2747e$-$206~ 0
	Daily	12.9265, 0.8876, 0.8539	0.0432	2.4179e$-$09~ 0
	Weekly	497.1053, 0.4818, 0.0571	0.0359	0.2572
	Monthly	3089.5834, 0.4633, 0.02934	0.1178	0.0119

Table 2.13: Distributions fitted to wind speed dataset from Telangana at four different timescales

Distribution	Timescale	Parameters (Shape, Scale)	K-S Statistic	$p-$Value
Exponential	Hourly	2.3409	0.2290	0
	Daily	2.3403	0.2833	0
	Weekly	2.3397	0.3454	5.7803e$-$84 ~ 0
	Monthy	2.3343	0.3897	2.82410e$-$25 ~ 0
Gamma	Hourly	3.1350, 0.7467	0.0363	4.5980e$-$151~ 0
	Daily	4.9608, 0.4717	0.0252	0.0018
	Weekly	6.8739, 0.3403	0.0463	0.0672
	Monthy	9.3193, 0.2504	0.0735	0.2673
Lognormal	Hourly	0.6320, 1.9791	0.0669	0.0

	Daily	0.4685, 2.1088	0.0400	4.5863e−08∼ 0
	Weekly	0.3879, 2.1717	0.0338	0.3226
	Monthy	0.3329, 2.2102	0.0701	0.3208
Weibull	Hourly	1.9829, 2.6441	0.0346	1.3186e−137∼ 0
	Daily	2.3947, 2.6480	0.0384	1.8141e−07 ∼ 0
	Weekly	2.7490, 0, 2.6353	0.0606	0.0059
	Monthly	3.3234, 2.6068	0.0753	0.2431
Exponentiated Weibull	Hourly	1.0671, 1.9106, 2.5669	0.0330	2.5538e−125∼ 0
	Daily	3.5293, 1.3135, 1.4374	0.0252	0.0018
	Weekly	14.4331, 0.9502, 0.6674	0.0329	0.3538
	Monthly	9.9036, 1.2436, 0.9986	0.0705	0.3129

Table 2.14: Distributions fitted to GHI dataset from Rajasthan at four different timescales

Distribution	**Timescale**	**Parameters (Shape, Scale)**	**K-S Statistic**	$p-$**Value**
Exponential	Hourly	557.4270	0.2310	0
	Daily	557.2844	0.4633	0
	Weekly	557.1347	0.4881	1.6330e−172∼ 0
	Monthy	556.2161	0.5005	1.8077e−42
Gamma	Hourly	2.9605, 188.2818	0.1152	0
	Daily	23.0590, 0, 24.1677	0.1059	6.1559e−54∼ 0
	Weekly	30.2105, 18.4417	0.1018	1.5969e−07∼ 0
	Monthy	33.5891, 16.5594	0.1329	0.0029
Lognormal	Hourly	0.7530, 466.3995	0.1637	0
	Daily	0.2168, 545.2452	0.1074	1.6153e−55 ∼ 0
	Weekly	0.1854, 547.9398	0.1055	4.8163e−08∼ 0
	Monthy	0.1758, 547.9573	0.1414	0.0012
Weibull	Hourly	2.2922, 622.8662	0.0903	0
	Daily	6.2302, 601.3256	0.0908	8.7351e−40∼ 0
	Weekly	6.8359, 598.0218	0.0856	1.8951e−05
	Monthly	7.2551, 595.1489	0.1030	0.0396
Exponentiated Weibull	Hourly	0.0347, 40.1982, 946.1783	0.0187	3.4878e−17∼ 0
	Daily	0.0466, 82.8844, 705.8376	0.0737	2.3079e−26∼ 0
	Weekly	0.0799, 52.8682, 693.5414	0.0758	0.0002
	Monthly	0.0707, 62.7199, 685.8490	0.0902	0.0983

Table 2.15: Distributions fitted to GHI dataset from Gujarat at four different timescales

Distribution	Timescale	Parameters (Shape, Scale)	K-S Statistic	$p-$Value
Exponential	Hourly	550.2106	0.2352	0
	Daily	550.0511	0.4529	0
	Weekly	550.0452	0.4838	2.8018e$-$169$\sim$ 0
	Monthy	549.6809	0.5207	3.4899e$-$46$\sim$ 0
Gamma	Hourly	3.6377, 151.2494	0.1004	0
	Daily	21.3525, 25.7604	0.0787	5.9037e$-$30$\sim$ 0
	Weekly	33.8833, 16.2334	0.0642	0.0029
	Monthy	42.4827, 12.9389	0.1043	0.359
Lognormal	Hourly	0.6088, 476.5644	0.1286	0
	Daily	0.2303, 537.2224	0.0957	4.1834e$-$44$\sim$ 0
	Weekly	0.1734, 541.9487	0.0624	0
	Monthy	0.1535, 543.2243	0.1003	0.0486
Weibull	Hourly	2.42186, 619.1530	0.0835	0
	Daily	6.03917, 593.2640	0.0548	9.2674e$-$15$\sim$ 0
	Weekly	6.5433, 589.9772	0.0730	0.0004
	Monthly	6.9994, 586.8407	0.1093	0.0243
Exponentiated Weibull	Hourly	0.0903, 16.6957, 920.7074	0.0338	7.4930e$-$55$\sim$ 0
	Daily	0.0974, 36.9999, 704.2958	0.0835	1.0433 e$-$33$\sim$ 0
	Weekly	4.1824, 3.2919, 451.7205	0.0661	0.0020
	Monthly	59.8431, 1.6805, 221.1725	0.0898	0.1013

Table 2.16: Distributions fitted to GHI dataset from Karnataka at four different timescales

Distribution	Timescale	Parameters (Shape, Scale)	K-S Statistic	$p-$Value
Exponential	Hourly	579.4207	0.2253	0
	Daily	579.3360	0.4359	0
	Weekly	579.3504	0.4870	1.2094e$-$171$\sim$ 0
	Monthy	579.3215	0.5398	6.5865e$-$50$\sim$ 0
Gamma	Hourly	3.7630, 153.9771	0.1058	0
	Daily	25.3619, 22.8427	0.1140	1.7646e$-$62$\sim$ 0
	Weekly	52.0885, 11.1224	0.0508	0.0334
	Monthy	86.5800, 6.6911	0.0907	0.0955
Lognormal	Hourly	0.5860, 504.3712	0.1245	0
	Daily	0.2127, 567.9529	0.1312	1.2120e$-$82$\sim$ 0

	Weekly	0.1410, 573.7981	0.0576	0.0105
	Monthy	0.1074, 575.9791	0.0892	0.1051
Weibull	Hourly	2.4332, 653.0119	0.0786	3.1387e−295∼ 0
	Daily	7.2918, 619.4771	0.0524	1.5535e−13∼ 0
	Weekly	8.7119, 612.8780	0.0602	0.0065
	Monthly	9.9215, 607.9942	0.1292	0.0042
Exponentiated Weibull	Hourly	0.0532, 28.0172, 980.1245	0.0388	3.5649e−72∼ 0
	Daily	0.1140, 39.8521, 706.3768	0.0611	3.1963 e−18∼ 0
	Weekly	1.0811, 8.3257, 607.8909	0.0582	0.0094
	Monthly	32.0552, 2.7738, 353.0611	0.0871	0.1204

Table 2.17: Distributions fitted to GHI dataset from Telangana at four different timescales

Distribution	**Timescale**	**Parameters (Shape, Scale)**	**K-S Statistic**	$p-$**Value**
Exponential	Hourly	537.5415	0.1992	0
	Daily	537.4041	0.4066	0
	Weekly	537.4453	0.4587	3.1279e−151∼ 0
	Monthy	537.6291	0.5232	1.1329e−46∼ 0
Gamma	Hourly	3.2841, 163.6778	0.1037	0
	Daily	15.3228, 35.0720	0.1290	5.4001e−80∼ 0
	Weekly	32.7721, 16.3994	0.0507	0.0344
	Monthy	55.6717, 9.6571	0.0761	0.2326
Lognormal	Hourly	0.6305, 458.1065	0.1196	0
	Daily	0.2790, 519.9666	0.1531	1.5654e−112∼ 0
	Weekly	0.1789, 529.2667	0.0592	0.0078
	Monthy	0.1338, 532.8078	0.0770	0.2211
Weibull	Hourly	2.2330, 606.2969	0.0891	0
	Daily	5.6828, 582.1590	0.0514	4.7256e−13∼ 0
	Weekly	6.8782, 575.3147	0.0538	0.0204
	Monthly	7.8472, 570.1801	0.1146	0.0157
Exponentiated Weibull	Hourly	0.0400, 33.9195, 950.1251	0.0388	4.4653e−99 ∼ 0
	Daily	0.1248, 27.4709, 685.5259	0.0710	1.6946e−24∼ 0
	Weekly	1.2199, 6.1169, 559.6544	0.0482	0.0505
	Monthly	82.5102, 1.8059, 222.1626	0.0782	0.2060

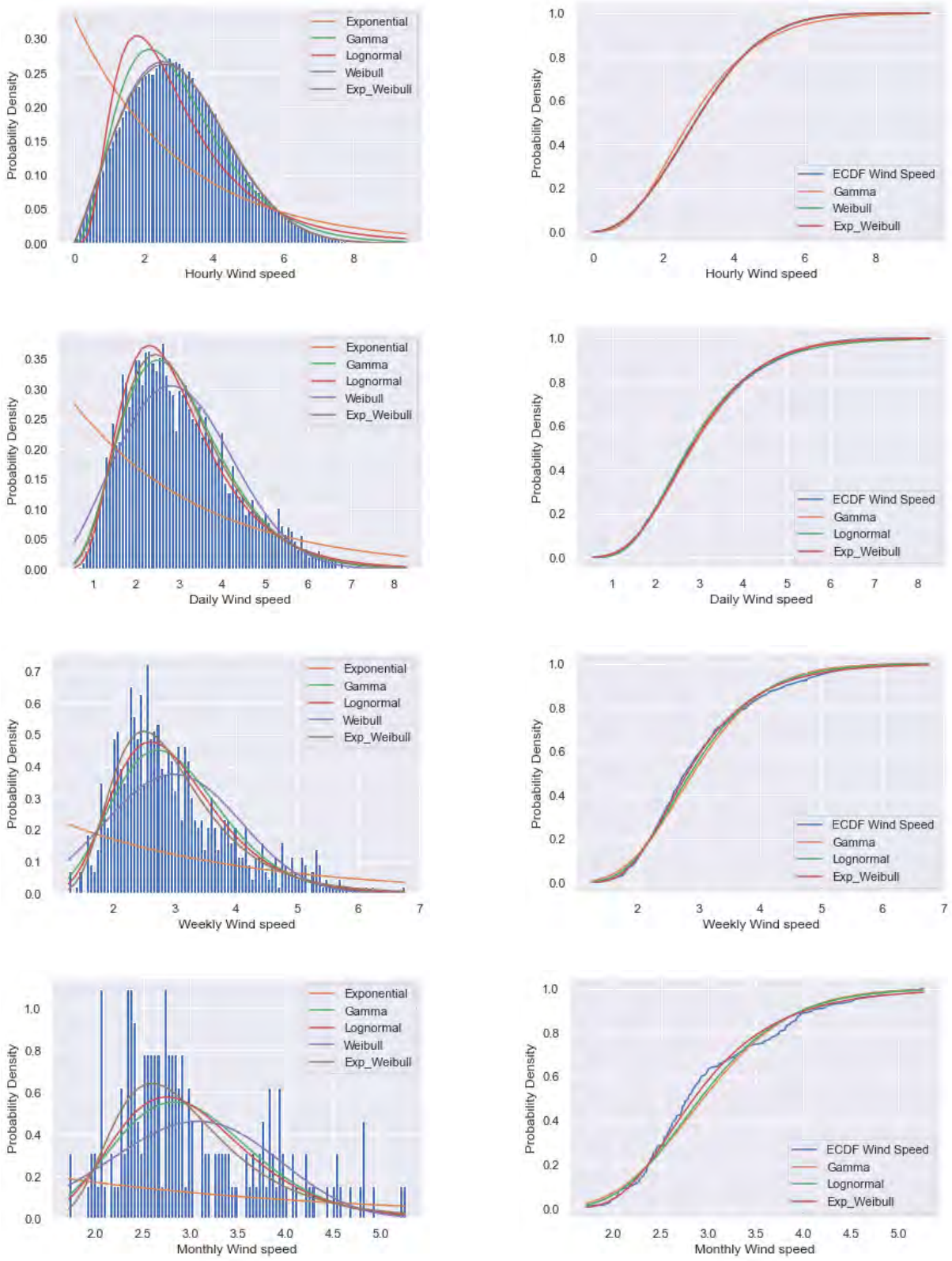


Fig. 2.19: Fitting of the pdfs and the cdfs of five distributions at different scales for the wind speed data from Rajasthan.

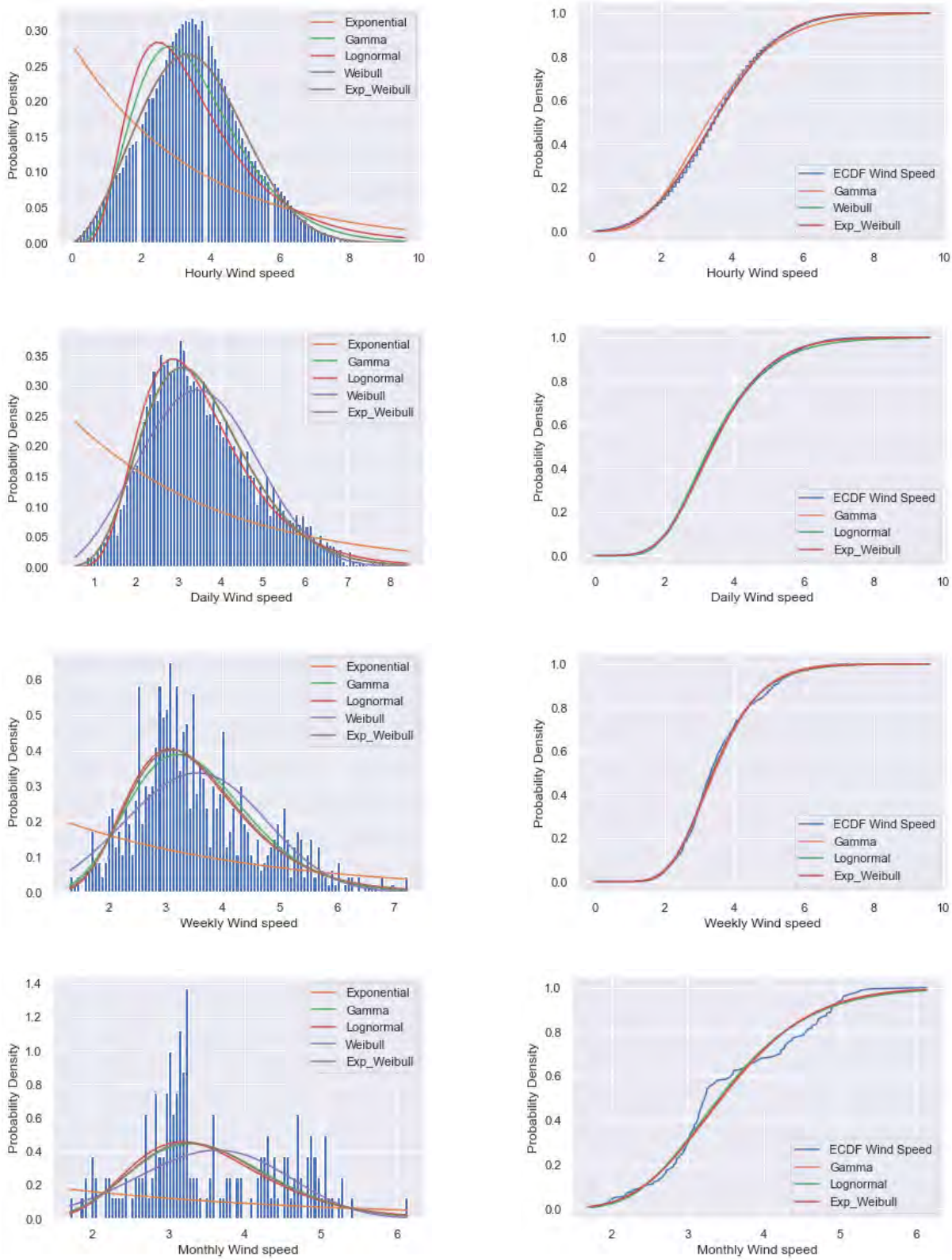


Fig. 2.20: Fitting of the pdfs and the cdfs of five distributions at different scales for the wind speed data from Gujarat.

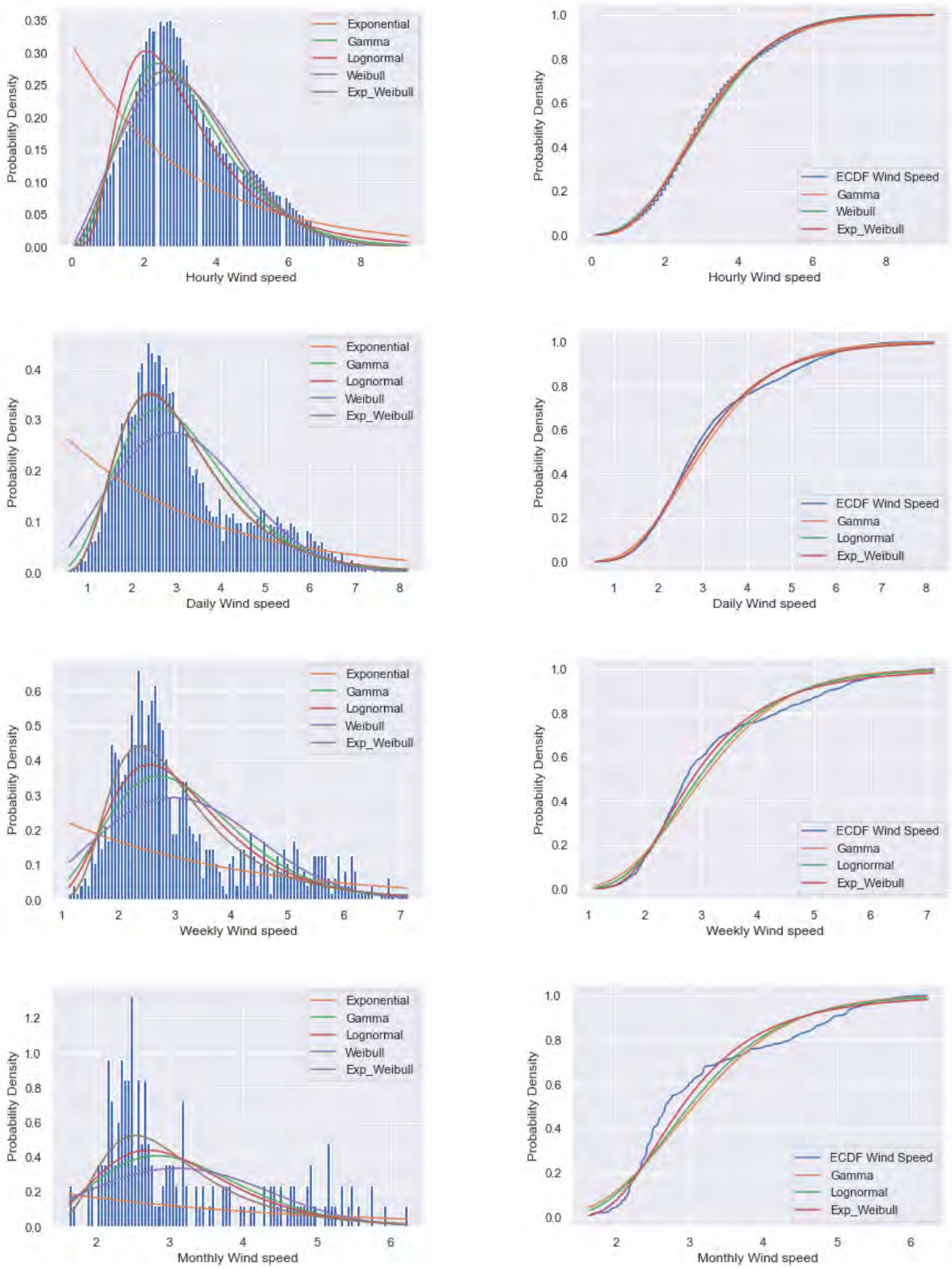


Fig. 2.21: Fitting of the pdfs and the cdfs of five distributions at different scales for the wind speed data from Karnataka.

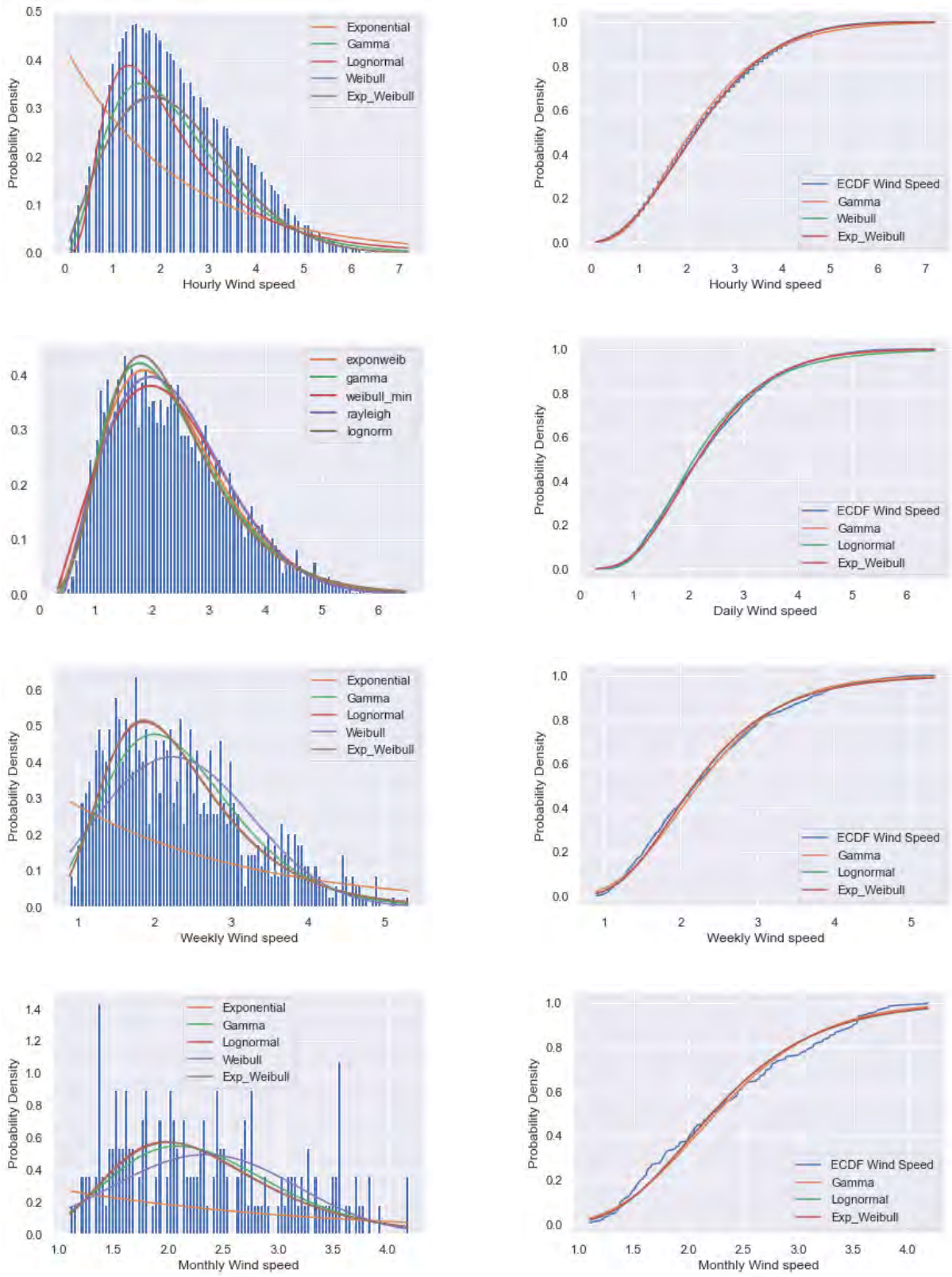


Fig. 2.22: Fitting of the pdfs and the cdfs of five distributions at different scales for the wind speed data from Telangana.

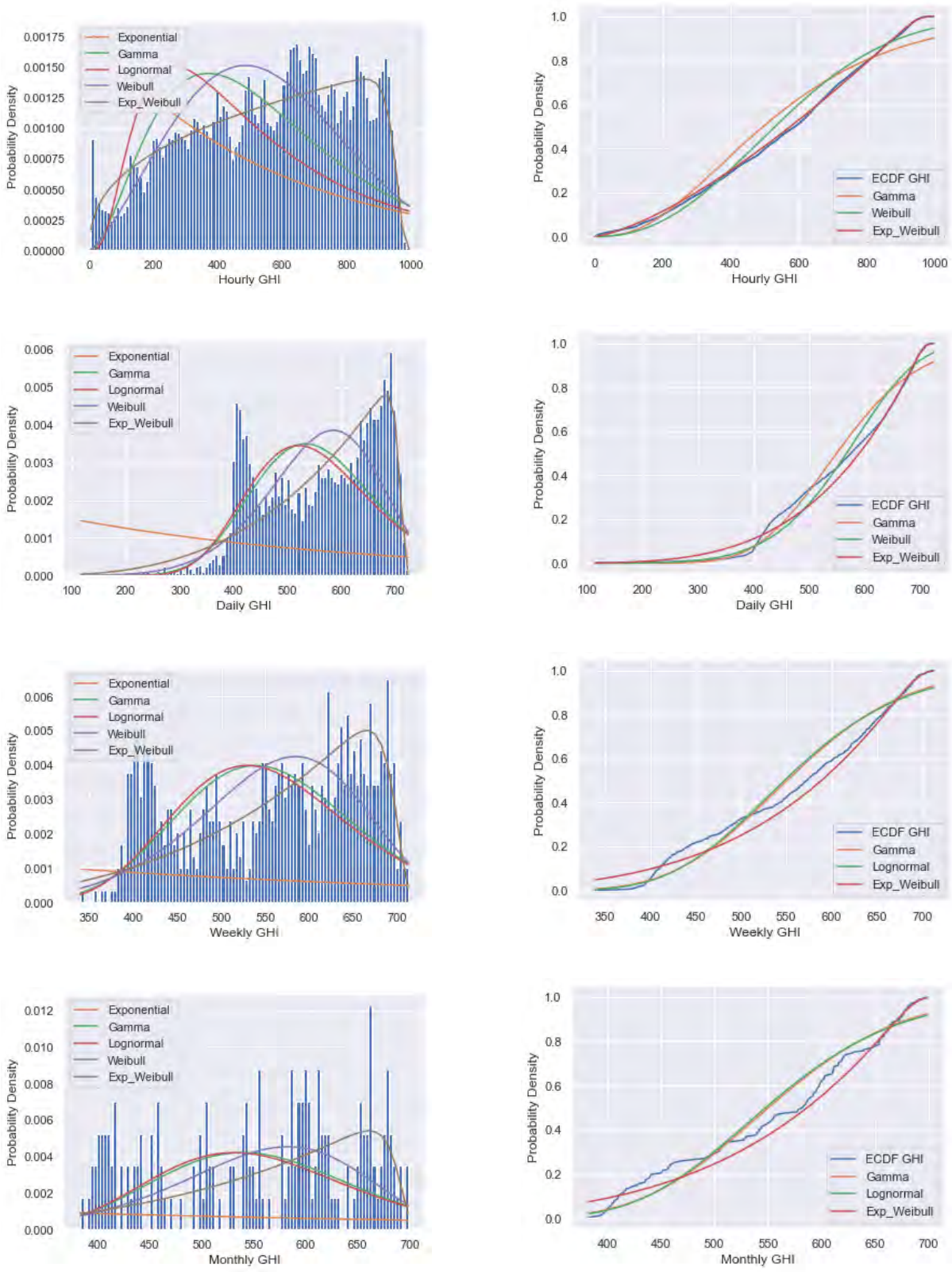


Fig. 2.23: Fitting of the pdfs and the cdfs of five distributions at different scales for the GHI data from Rajasthan.

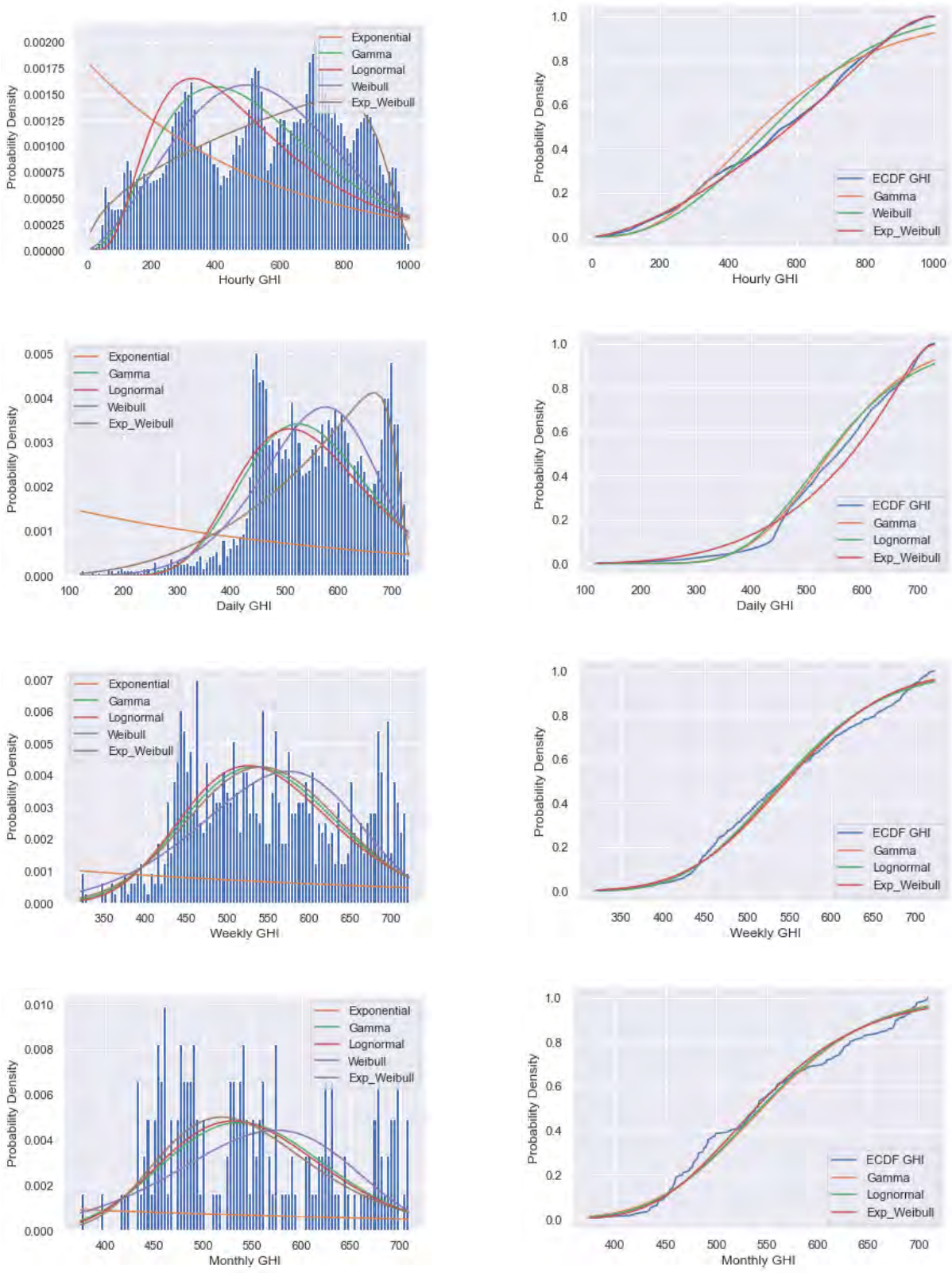


Fig. 2.24: Fitting of the pdfs and the cdfs of five distributions at different scales for the GHI data from Gujarat.

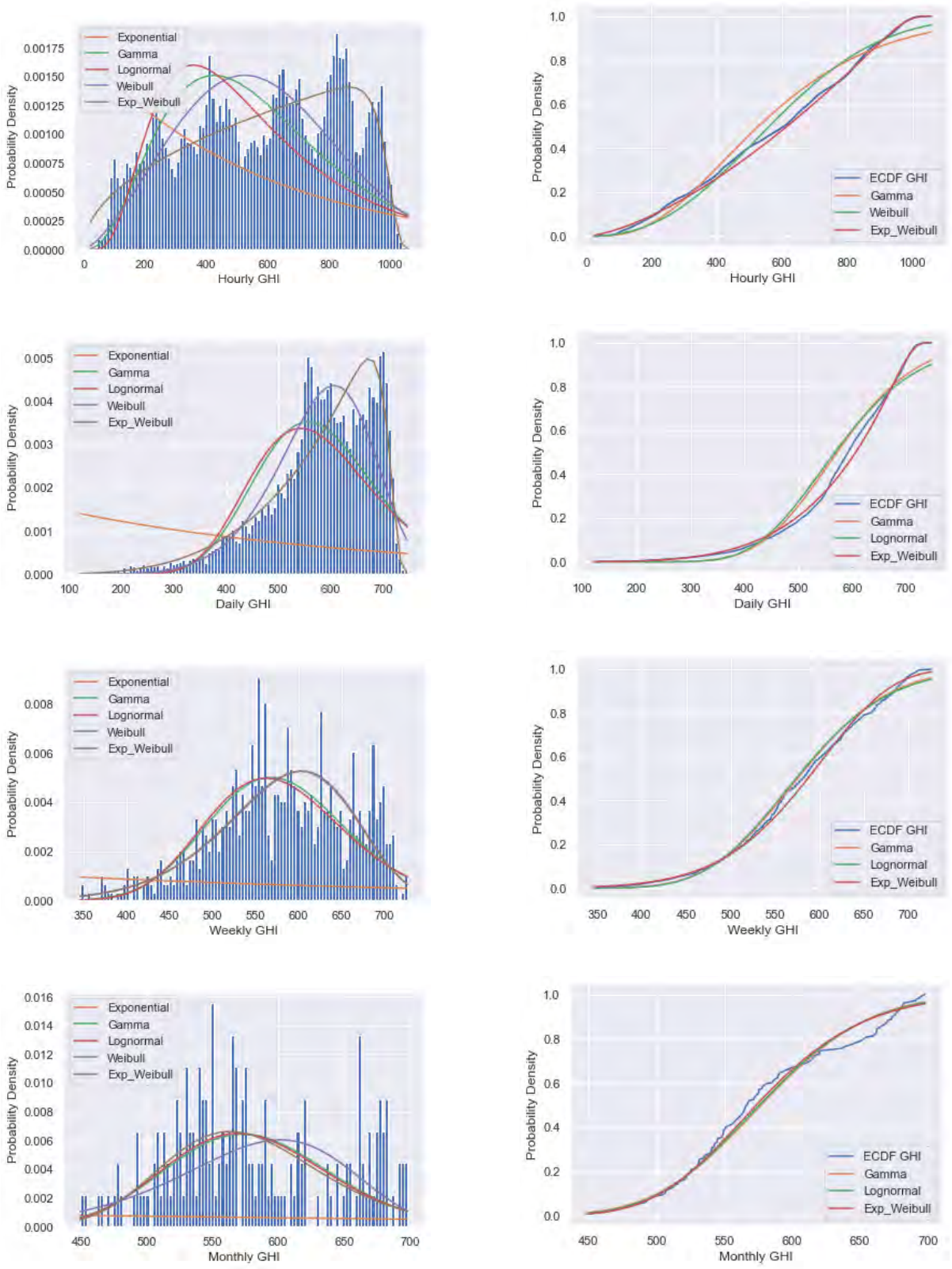


Fig. 2.25: Fitting of the pdfs and the cdfs of five distributions at different scales for the GHI data from Karnataka.

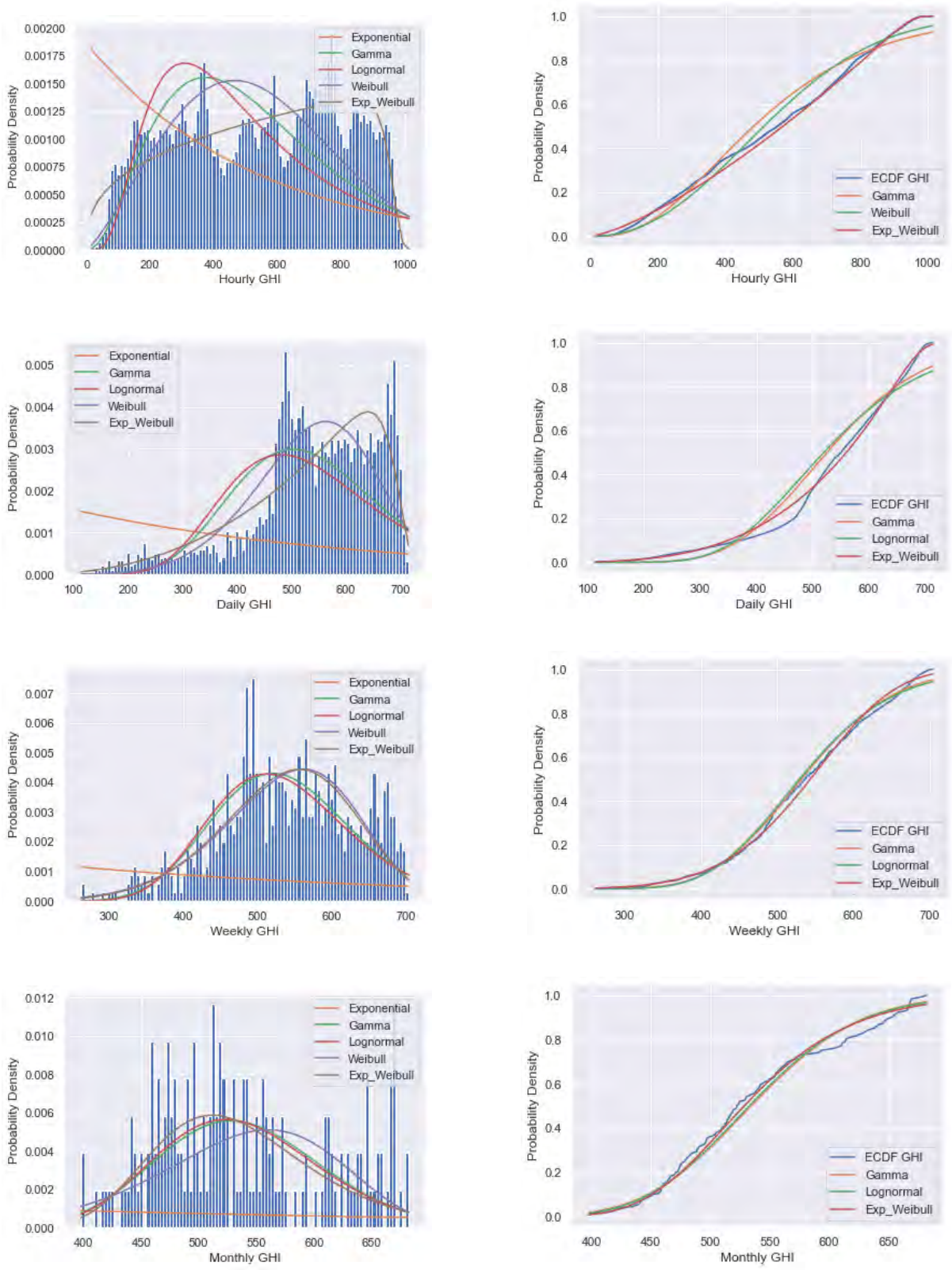


Fig. 2.26: Fitting of the pdfs and the cdfs of five distributions at different scales for the GHI data from Telangana.

Table 2.18: Ranking of the probability distributions fitted to wind speed datasets

Timescale	Location	Exponential	Gamma	Lognormal	Weibull	Exponentiated Weibull
Hourly	1	5	3	4	2	1
	2	5	3	4	2	1
	3	5	1	4	3	2
	4	5	3	4	2	1
Daily	1	5	2	3	4	1
	2	5	1	2	4	3
	3	5	3	1	4	2
	4	5	1	4	3	2
Weekly	1	5	3	2	4	1
	2	5	3	1	4	2
	3	5	3	2	4	1
	4	5	3	2	4	1
Monthly	1	5	3	2	4	1
	2	5	3	1	4	2
	3	5	3	2	4	1
	4	5	3	1	4	2

Table 2.19: Ranking of probability distributions for GHI datasets

Timescale	Location	Exponential	Gamma	Lognormal	Weibull	Exponentiated Weibull
Hourly	1	5	3	4	2	1
	2	5	3	4	2	1
	3	5	3	4	2	1
	4	5	3	4	2	1
Daily	1	5	3	4	2	1
	2	5	2	4	1	3
	3	5	3	4	1	2
	4	5	3	4	2	1
Weekly	1	5	3	4	2	1
	2	5	2	1	4	3
	3	5	3	2	4	1
	4	5	2	4	3	1
Monthly	1	5	3	4	2	1
	2	5	3	2	4	1
	3	5	3	2	4	1
	4	5	1	2	4	3

2.7 Summary

This chapter has provided a statistical analysis of wind speed and solar irradiance data at four selected locations in India. For this, (i) we first resample the hourly datasets at daily, weekly, and monthly timescales. We analyze the time series plots and study descriptive measures of the data; Gujarat has the highest mean wind speed $(3.53\ m/s)$, whereas Telangana has the lowest mean wind speed $(2.34\ m/s)$ among the four study sites; the highest mean GHI value $(579.35\ W/m^2)$ comes from Karnataka, whereas the lowest mean GHI value $(537.40\ W/m^2)$ is recorded in Telangana. (ii) Then, we decompose the time series into seasonal, trend, and irregular components. This decomposition exhibits long term yearly seasonal pattern with no major upward or downward trend. Also, data values are distributed around the mean values. (iii) After this, we check data stationarity through the ADF test. Results show that the value of the ADF statistic for each timescale is significantly lesser than the corresponding critical value at $\alpha = 5\%$, indicating stationarity of the data. (iv) In addition, to determine the best fit probability distribution functions for modeling wind speed and GHI data, we have implemented five probability distributions, namely exponential, gamma, lognormal, Weibull, and exponentiated Weibull. We have adopted the maximum likelihood estimation for parameter estimation and the K-S test for goodness of fit. Based on the least K–S value, we provide a ranking of the studied models. It is observed that the exponentiated Weibull has the best representation most of the times, suggesting it as a highly useful model for renewable energy data. Thus, the present exploration of wind speed and solar irradiance can help renewable energy community in India to achieve many of its environmental and energy policy targets.

Chapter 3

Time Series Models for Renewable Energy Forecasting

"*As for the future, your task is not to foresee it, but to enable it.*"

– ANTONIE DE SAINT EXUPERY

This chapter focuses on the forecasting of wind speed and GHI data at hourly, daily, weekly, and monthly timescales using various statistical methods, such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and seasonal autoregressive integrated moving average (SARIMA). We choose these time series models since the datasets exhibit seasonality, stationarity, and randomness. We adopt a grid search method to find optimum values of model parameters and use root mean square error (RMSE) to assess the performance of the studied models. Since the above models fail to fully capture the high amount of fluctuation (mostly, seasonal fluctuation) in the hourly, daily, and weekly observations, we additionally implement an ARIMA model with sliding windows (WS-ARIMA) to improve the modeling efficacy. The WS-ARIMA technique with a fixed or variable window length belongs to the class of adaptive models. Particularly, the sliding windows of fixed length are popular in the areas of finance, energy, and traffic management, where the dataset of necessity exhibits a seasonal pattern. Finally, we perform a residual analysis as a post-processing step to examine any systematic bias in the implemented models. The experimental results based on 15 years (2000–2014) of data reveal that (i) for monthly forecasting, the SARIMA model has the best performance and (ii) for daily and weekly wind speed and GHI data, the WS-ARIMA method consistently outperforms the conventional time series methods with significant improvement in the forecasts in both time and space. For hourly forecasting, the WS-ARIMA and ARIMA model have comparable performance based on three years (2012–2014) of data. Therefore, in summary, the present chapter provides a generic guideline for the applicability of different statistical models of wind speed and GHI forecasting at desired time horizon. Moreover, the emanated results strongly suggest the inclusion of

the WS-ARIMA model as one of the potential statistical techniques in wind speed and GHI forecasting.

Contents

Parts of this chapter have been published in the following refereed publications:

S. Sheoran, R. S. Singh, S. Pasari, R. Kulshrestha, "Forecasting of solar irradiances using time series and machine learning models: A case study from India," *Applied Solar Energy*, vol. 58, pp. 137–151, 2022.

S. Sheoran, S. Pasari S, "Efficacy and application of the window-sliding ARIMA for daily and weekly wind speed forecasting," *Journal of Renewable and Sustainable Energy*, vol. 14, p. 053305, 2022.

S. Sheoran, S. Shukla, S. Pasari, R. S. Singh, R. Kulshrestha, "Wind speed forecasting at different timescales using time series and machine learning models," *Applied Solar Energy*, vol.

58, pp. 708–721, 2022.

S. Sheoran, R. Bavdekar, S. Pasari, R. Kulshrestha, "Wind speed forecasting using time series methods: A case study," *In Chamola BP, Kumari P, Kaur L (eds) Emerging Advancements in Mathematical Sciences*, pp. 125–133, 2022.

3.1 Introduction

Reliable forecasting of renewable energy helps in planning and estimating the energy output on a short term to a long term basis [35]. Among different renewable energy resources, the contribution of wind and solar energy is remarkable. The wind and solar energy forecasting is useful for several practical purposes, such as estimation of energy outputs of plants, marketing of renewable energy, and maintenance planning of wind and solar power plants. Hourly predictions of renewable energy can be used for prompt and immediate planning by knowing the energy productions over the next few days; daily forecasting can help decide the best months for solar and wind energy productions, whereas monthly weather forecasting can be used for long term planning of power plants [123, 137]. The renewable energy forecasting techniques described in the literature are physical [161], statistical [127], and artificial intelligence [65] methods. The most popular statistical time series based forecasting approach is the autoregressive moving average (ARMA) model. Some variants of ARMA include autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), fractional ARIMA (f-ARIMA), and ARMA with exogenous input (ARMAX or ARX) [14, 77, 110, 135]. In 1991, the ARMA method was first used for hourly averaged wind speed forecasting in Jamaica [29]. Karakus et al. [60] highlighted the efficacy of polynomial autoregressive (PAR) model and compared the results with several other time series models for daily wind speed and wind power predictions in Turkey and USA. Shukur and Lee [129] developed hybrid Kalman filter-artificial neural network (KF-ANN) based ARIMA for daily wind speed data from Iraq and Malaysia. The SARIMA and Adaline neural network models were used in [21] to forecast wind speed in Mexico to demonstrate that SARIMA closely follows the actual wind speed pattern. Cadenas et al. [22] compared a univariate ARIMA model and a multivariate non-linear ARX model for wind speed prediction. Pasari and Shah [99] considered daily and monthly wind speed forecasting using univariate ARIMA (2,1,2) model based on Akaike and Bayesian information criteria. Sheoran et al. [121] demonstrated the application of several statistical methods, such as AR, MA, ARMA, ARIMA, SARIMA, and Holt Winter's technique for wind speed forecasting at hourly, daily, and monthly time horizons at a location in Madhya Pradesh, India. They have also presented a generic guideline for the applicability of different statistical models of wind speed forecasting at desired time horizon. In a similar manner, Saima et al. [114] summarized the strengths and drawbacks of statistical, hybrid, and machine learning models in the area of weather forecasting. In 2019, Alsharif et al. [7] implemented ARIMA, SARIMA, and Monte-Carlo prediction models for hourly solar radiation data of Seoul, South Korea over 37 years $(1981-2017)$. The results demonstrated that the ARIMA $(1,1,2)$ model can be used to represent daily solar irradiation, while the SARIMA $(4,1,1)$ with 12 lags can

be used to represent monthly solar irradiation. In 2019, Sharadga et al. [119] implemented ARMA, ARIMA, SARIMA, and six neural network models for three different type of days, namely sunny, cloudy, and rainy day. The results indicate that time series models of ARMA $(3,4)$, ARIMA $(2,1,3)$, and SARIMA $(2,1,3)(2,0,1)_{14}$ provide the best fit representation. A comprehensive review on statistical modeling of renewable energy data is available in [15, 90, 116, 135].

In addition to the conventional time series models, the present chapter highlights the efficacy of the WS-ARIMA method, which is nothing but a variant of the ARMA model. Based on certain pre-requisites related to data, the WS-ARIMA method utilizes fixed or variable window length to find sub-arrays within an array. Specifically, the sliding windows of fixed length are widely used in the areas of finance, energy, and traffic management, where the dataset of necessity exhibits a seasonal pattern [49]. In fact, the order of seasonality helps to determine the appropriate length of the sliding window. Reikard in [108] and Reikard and Hansen in [109] presented the WS-ARIMA method as an efficient technique for solar energy prediction. Alberg and Last in 2018 [6] developed the WS-ARIMA model for load forecasting in smart meters to balance the demand and supply of electricity. The predicting power of the WS-ARIMA model in forecasting equity returns was highlighted in [32]. Yu et al. [167] showed that the inclusion of sliding windows has improved the efficacy of the ARIMA model in traffic anomaly detection. Recently, in 2022, Medhi et al. [79] mentioned that sliding windows on fuzzy ARIMA provide the best accuracy in cloud traffic prediction. Similarly, Sheoran et al. [124] highlighted the effectiveness of the WS-ARIMA model for daily and weekly solar irradiance prediction. Based on the results of two selected study locations from Gujarat and Rajasthan, India, they have shown that the WS-ARIMA model has the best performance in comparison to the SARIMA and machine learning methods for daily and weekly data.

In the present chapter, we demonstrate the implementation of timeseries models for forecasting of wind speed and solar irradiance. We expect that based on the selected time horizon from hourly to daily, weekly, and to monthly averaged dataset, the statistical timeseries models would be capable of capturing the trend, seasonality, and randomness resulting to provide reliable results. Therefore, the proposed analysis could help to choose a suitable forecasting model for forecasting of wind speed and GHI based on the selected time horizon depending on the requirement of power stations. The emanated results from time series models can be compared with the other available techniques such as artificial intelligence methods and hybrid models (discussed in Chapter 4 and Chapter 5) to provide a guideline to choose the most suitable forecasting model.

3.2 Mathematical Description

Below, we describe the studied time series models.

3.2.1 Autoregressive (AR) Model

The autoregression in a time series considers that the output variable linearly depends on its own prior values as well as a stochastic component (an unpredictable term). An AR model [51, 83] of order p is represented by the following form.

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + Z_t \tag{3.1}$$

Here, X_t denotes the values of time series; Z_t denotes noise; $\phi = (\phi_1, \phi_2, .., \phi_p)$ is the model coefficient vector and p is a positive integer.

3.2.2 Moving Average (MA) Model

In contrast to an AR model which applies a weighted total of previous values to determine a statistical illustration, the moving average (MA) process [121, 128] considers that the output variable is linearly dependent on the present and numerous previous values of random terms. The MA process of order q is as follows.

$$X_t = \sum_{j=1}^{q} \theta_j Z_{t-j} + Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + ... + \theta_q Z_{t-q} + Z_t \tag{3.2}$$

Here, $\theta = (\theta_1, \theta_2, .., \theta_q)$ is the model coefficient vector and q is a positive integer.

3.2.3 Autoregressive Moving Average (ARMA) Model

An ARMA [24, 104] process is integrated by autoregressive (AR) and moving average (MA) methods to output a process with a minimal parametrization. An ARMA process of order (p, q) is expressed as follows.

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j Z_{t-j} \tag{3.3}$$

Here, ϕ_i and θ_j are the coefficients of AR and MA part of an ARMA model.

3.2.4 Autoregressive Integrated Moving Average (ARIMA) Model

An ARIMA [24, 33, 133] process is preferred when the data exhibit some evidences of non-stationarity. Predictions are based on past values of time-series data in AR models, whereas prior residuals are used for forecasting future values in MA models. An ARIMA (p,d,q) model is created by combining a stationary ARMA (p,q) process with the d^{th} difference of a time series. Thus, as an extension of the ARMA model, the expression for the ARIMA model is given as:

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t \tag{3.4}$$

At $d = 0$, the above formulation represents an ARMA (p,q) model, as mentioned below:

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}, \tag{3.5}$$

$$\phi(B)X_t = \theta(B)Z_t$$

Here, $\{X_t\}$ is a time series of forecast variable and Z_t denotes the random noise; B is the backshift operator. The parameters p, d and q represent the number of autoregressive terms, the order of differencing that must be performed to stationarize the time series, and the number of terms in moving average, respectively. Since the present dataset is observed to be stationary (details are provided in Section 2.5), no differencing was applied, resulting $d = 0$. Thus, the final model turns out to be an ARMA (p,q) model.

3.2.5 Seasonal ARIMA (SARIMA) Model

The expression for the SARIMA $(p,d,q)(P,D,Q)_s$ model [3, 17] is given as:

$$\phi(B)(1-B)^d \Phi(B^s)(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t \tag{3.6}$$

where, $\{X_t\}$ is a time series of forecast variable and Z_t denotes the random noise; B is the backshift operator. $\Phi(B^s)$ and $\Theta(B^s)$ are defined in a similar way of $\phi(B)$ and $\theta(B)$ in ARIMA model. The parameters for these type of models are as follows.

- p and seasonal P indicate number of autoregressive terms
- d and seasonal D indicate order of differencing that must be performed to stationarize the time series
- q and seasonal Q represent number of terms in moving average
- s indicates seasonal length.

The orders of seasonality for the SARIMA models are 12 for monthly forecasting, 52 for weekly forecasting, and 73×5 for daily forecasting. Note that we have prepared five different sets each with 73 days of data so as to reduce the high amount of fluctuation (mostly, seasonal fluctuation) in the daily observations. After optimization, the predictions of these five individual sets are combined to reduce the computational resources.

3.2.6 Window-Sliding ARIMA (WS-ARIMA) Model

From the formulations of ARMA and ARIMA model, it is evident that these models are inadequate to deal with seasonal variations effectively. In such cases, the seasonal ARIMA (SARIMA) model is often recommended. However, as the SARIMA $(p,d,q)(P,D,Q,S)$ model includes seven parameters, the model is quite complex to deal with longer time series data. The complexity of such models can be significantly reduced by the inclusion of sliding windows, where the size of the window is fixed through a complete nested loop [108, 124]. The general procedure of the sliding window technique is as follows. (i) Find the size of the required window based on data characteristics; (ii) obtain the result corresponding to the 1^{st} window, and (iii) use a loop to slide the window to compute results window by window. A simple demonstration of the process is provided in Figure 3.1. As the present wind speed and GHI data show

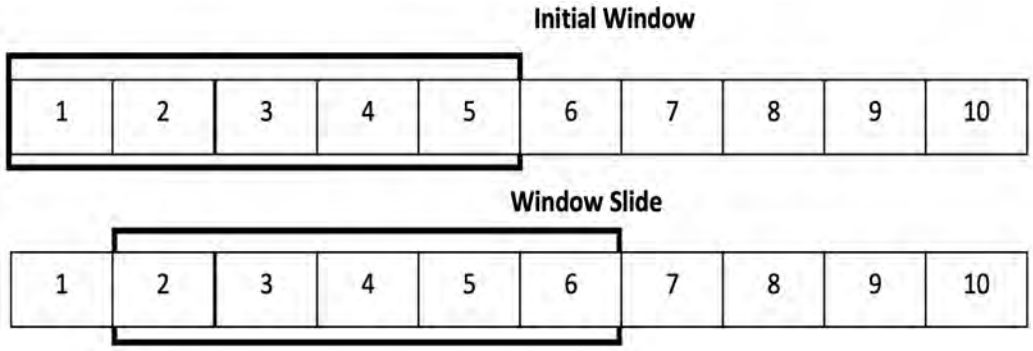


Fig. 3.1: A visual representation of the window sliding process [49].

a high positive autocorrelation peak at an interval of one year (yearly seasonal pattern as viewed from Figures 2.3– 2.10, a sliding window of 365 days is applied to make predictions for the next day. The results are consequently accumulated for three years of test dataset. Similarly, a sliding window of 52 weeks and 12 months is considered for the weekly and monthly dataset.

3.3 Methodology

The methodology for implementation of time series models is presented in a three-fold manner as detailed below.

3.3.1 Data Preparation

The data are initially split into training and testing sets, with a training period from January 1, 2000 to December 31, 2011 and the testing period from January 1, 2012 to December 31, 2014. Then, the hourly wind speed data are sampled according to the desired time horizon. For instance, to carry out daily and weekly forecast, the dataset is sampled daily and weekly. To note, for hourly forecasting, we consider only three years (2012–2014) of wind speed and GHI data due to high amount of variability in the longer time series data. Various descriptive measures, such as sample mean, standard deviation, and quartiles are computed based on the training data (Tables 2.5, 2.6). We then perform the Augmented Dickey Fuller (ADF) test to check for stationarity in the time series. The results of the ADF test for both wind speed and GHI datasets are provided in Section 2.5.

3.3.2 Model Selection and Validation

We implement the considered AR, MA, ARMA, ARIMA, SARIMA, and WS-ARIMA models on the test data. For the WS-ARIMA model, we use a sliding window of 365, 52, and 12 for daily, weekly, and monthly data, respectively. For hourly data, we consider window size of 24 (numbers of hours in a day) and 10 (number of sunny hours considered) for wind speed and GHI, respectively. The minimum values of Akaike information criterion (AIC) and root mean squared error (RMSE) are considered while performing a grid search procedure to obtain optimal parameters of the studied models. For computation, we have used ARIMA and autoarima functions from the open-source "statsmodels" and "pmdarima" python libraries. The accuracy of the forecasting models is evaluated and compared on the basis of the RMSE and MAPE error metrics. Lower the error values, better is the forecast.

To summarize, for the hourly, daily, weekly, and monthly forecasting, the wind speed and GHI data are split into training and testing data. Consequently, the best fit model (with the minimum AIC) for the training data is applied for testing data. For WS-ARIMA, the best fit ARIMA model for the training data is applied onto the sliding window model to forecast the results for testing data [108].

3.3.3 Residual Analysis

Carrying out residual analysis is a standard practice to check for any systematic bias in the implemented models. As the residuals of a forecast model should exhibit Gaussian distribution with zero mean and a constant variance, we analyze residual plots, histograms, and corresponding P-P plots (Section 14.8 and Section 14.9 [8]) of the standardized residuals. The results are discussed at a later section.

Table 3.1: Monthly wind speed forecasting at four locations

Monthly Forecasting				
Location	**Model**	**Parameters**	**RMSE**	**MAPE**
Pokhran (Rajasthan)	AR	(24)	0.488	0.116
	MA	(8)	0.518	0.112
	ARIMA	(2,0,1)	0.363	0.103
	SARIMA	(1,0,1)(2,0,1,12)	0.428	0.105
	WS-ARIMA	(2,0,1)	0.457	0.111
Bitta (Gujarat)	ARIMA	(2,0,1)	0.442	0.067
	SARIMA	(1,1,1)(1,0,1,12)	0.319	0.075
	WS-ARIMA	(2,0,1)	0.557	0.111
Pavagada (Karnataka)	ARIMA	(2,0,1)	0.468	0.140
	SARIMA	(1,1,0)(1,0,1,12)	0.383	0.110
	WS-ARIMA	(2,0,1)	0.746	0.168
Ramagundam (Telangana)	ARIMA	(2,0,3)	0.442	0.170
	SARIMA	(1,1,1)(1,0,2,12)	0.354	0.102
	WS-ARIMA	(2,0,3)	0.384	0.159

3.4 Results

The results in terms of RMSE and MAPE values in wind speed and GHI forecasting at different time horizons for the four study sites are summarized in the following subsections.

3.4.1 Results of Monthly Forecasting

Table 3.1 and Table 3.2 represent the RMSE and MAPE values of the implemented time series models in four study sites. In Pokhran, Rajasthan, the AR and the MA models have the highest error values, indicating their worst performance. Thus, for the other three study sites, we present the monthly forecasts from three best fit models, namely ARIMA, SARIMA, and WS-ARIMA. For monthly forecasting, the SARIMA model outperforms the other two models due to its ability to precisely capture the yearly seasonality of order 12 in the dataset. The comparative performance of the the three best fit models is pictorially shown in Figure 3.2 and Figure 3.3 for wind speed and GHI forecasting in Pokhran, Rajasthan.

Table 3.2: Monthly GHI forecasting at four locations

Monthly Forecasting				
Location	**Model**	**Parameters**	**RMSE**	**MAPE**
Pokhran (Rajasthan)	AR	(12)	26.996	0.094
	MA	(8)	32.514	0.101
	ARIMA	(2,0,1)	20.012	0.084
	SARIMA	(1,0,2)(1,0,0,12)	15.437	0.020
	WS-ARIMA	(2,0,1)	49.156	0.069
Bitta (Gujarat)	ARIMA	(1,0,2)	39.506	0.060
	SARIMA	(1,1,0)(1,0,0,12)	22.765	0.030
	WS-ARIMA	(1,0,2)	54.000	0.079
Pavagada (Karnataka)	ARIMA	(3,0,2)	20.107	0.034
	SARIMA	(1,0,0)(1,0,0,12)	15.821	0.027
	WS-ARIMA	(3,0,2)	39.689	0.058
Ramagundam (Telangana)	ARIMA	(3,0,1)	22.704	0.041
	SARIMA	(1,1,0)(1,0,0,12)	19.384	0.029
	WS-ARIMA	(3,0,1)	34.079	0.060

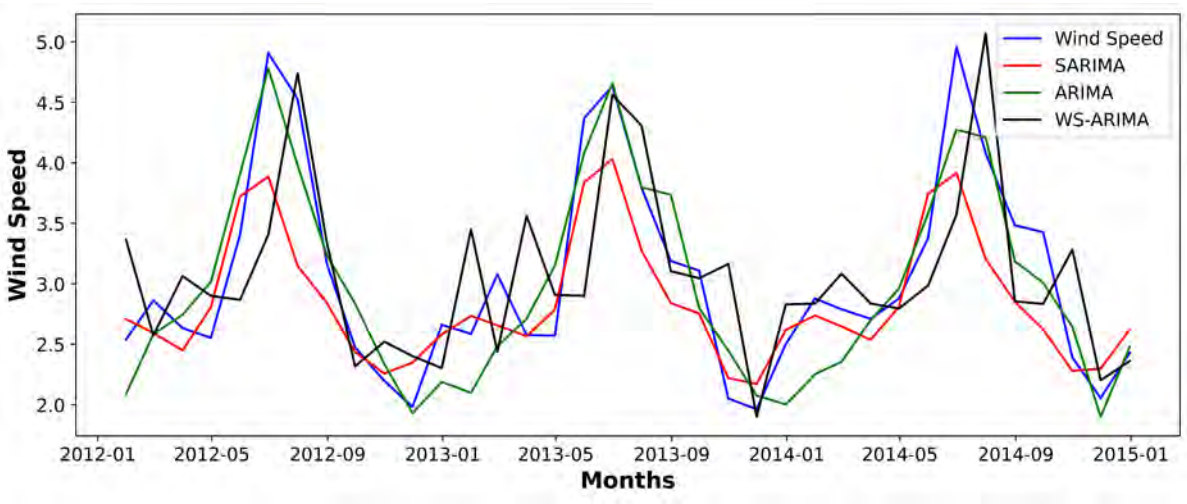


Fig. 3.2: Monthly wind speed (m/s) forecast from the best three time series models implemented in Pokhran, Rajasthan.

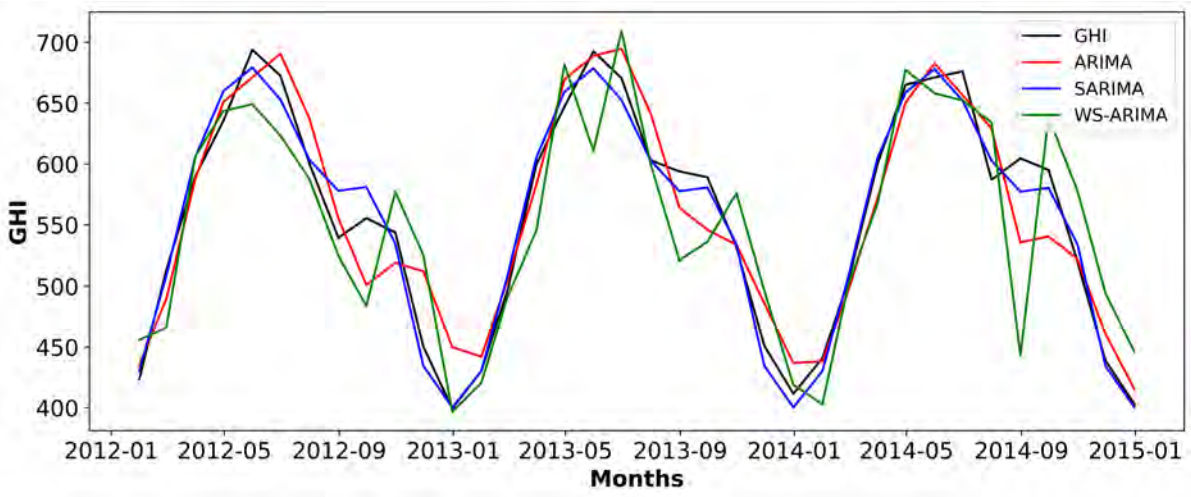


Fig. 3.3: Monthly GHI (W/m^2) forecast from the best three time series models implemented in Pokhran, Rajasthan.

3.4.2 Results of Weekly Forecasting

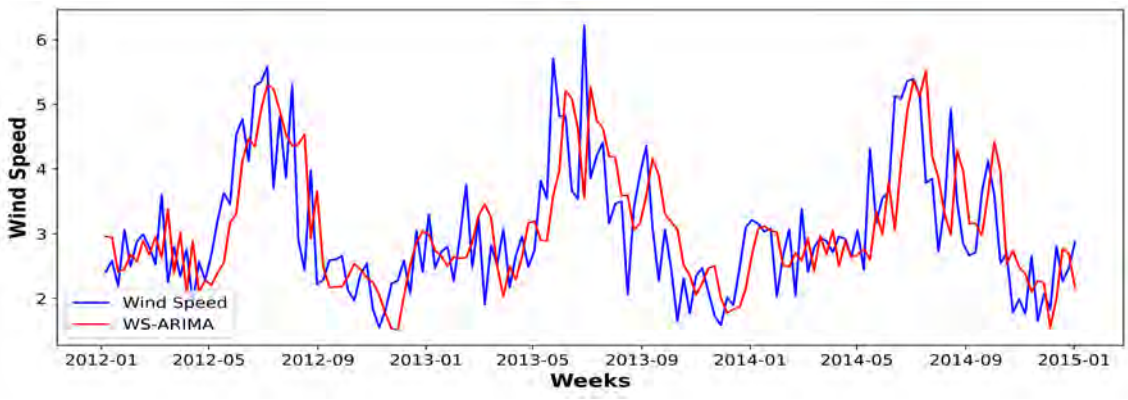
The MAPE and RMSE values of the implemented models are mentioned in Table 3.3 and Table 3.4. We observe that the SARIMA model fails to capture the high amount of fluctuation (mostly, seasonal fluctuation) in the weekly dataset. Here, the WS-ARIMA model with sliding windows of length 52 (corresponding to number of weeks in a year) reflects the least error values in both wind speed and GHI forecasting. Overall, for weekly forecasting, the WS-ARIMA model reflects the least error values followed by ARIMA and SARIMA in wind speed, whereas the SARIMA model outperforms ARIMA model in GHI. It is important to note that in the WS-ARIMA, the error values are reduced upto 50% in comparison to the conventional ARIMA model. The weekly forecasts from the best fit WS-ARIMA model at the four study sites are pictorially shown in Figure 3.4 and Figure 3.5 for wind speed and GHI datasets, respectively.

Table 3.4: Weekly GHI forecasting at four locations

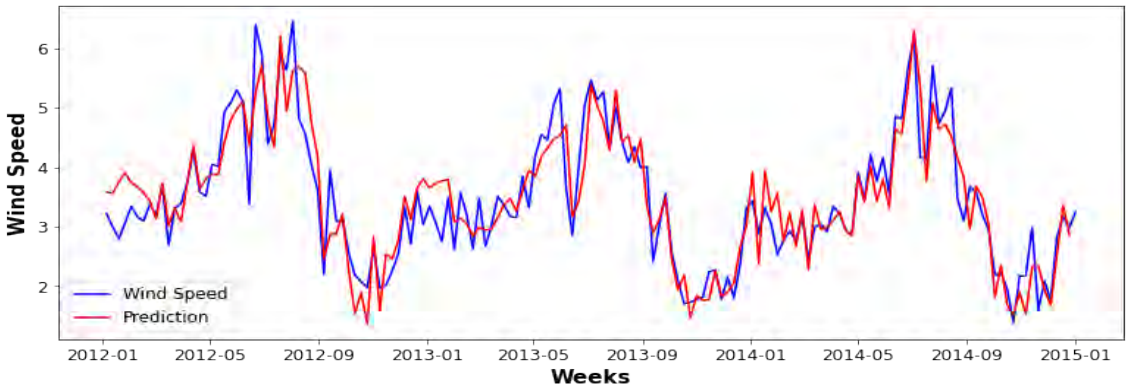
Weekly Forecasting				
Location	**Model**	**Parameters**	**RMSE**	**MAPE**
Pokhran (Rajasthan)	AR	(2)	55.509	0.160
	MA	(3)	68.961	0.195
	ARIMA	(2,0,3)	49.931	0.082
	SARIMA	(1,0,0)(1,0,1,52)	45.934	0.063
	WS-ARIMA	(2,0,3)	28.127	0.031
Bitta (Gujarat)	ARIMA	(2,0,5)	48.657	0.194
	SARIMA	(1,1,0)(1,1,0,52)	48.321	0.062
	WS-ARIMA	(2,0,5)	30.848	0.041
Pavagada (Karnataka)	ARIMA	(5,0,4)	45.912	0.124
	SARIMA	(1,0,0)(2,0,0,52)	43.178	0.112
	WS-ARIMA	(5,0,4)	32.037	0.056
Ramagundam (Telangana)	ARIMA	(1,0,1)	40.974	0.175
	SARIMA	(1,1,0)(1,1,0,52)	38.877	0.172
	WS-ARIMA	(1,0,1)	19.785	0.061

Table 3.3: Weekly wind speed forecasting at four locations

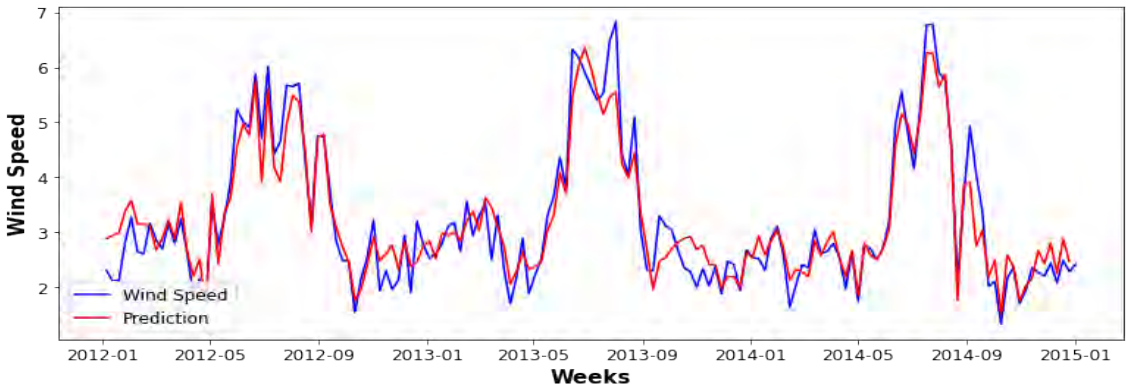
Weekly Forecasting				
Location	**Model**	**Parameters**	**RMSE**	**MAPE**
Pokhran (Rajasthan)	AR	(4)	1.288	0.323
	MA	(8)	1.617	0.412
	ARIMA	(3,0,2)	0.801	0.237
	SARIMA	(1,0,0)(1,0,1,52)	1.281	0.424
	WS-ARIMA	(3,0,2)	0.483	0.129
Bitta (Gujarat)	ARIMA	(4,0,5)	0.657	0.194
	SARIMA	(2,0,2)(1,0,0,52)	1.302	0.376
	WS-ARIMA	(4,0,5)	0.453	0.109
Pavagada (Karnataka)	ARIMA	(2,0,4)	0.868	0.238
	SARIMA	(0,0,2)(2,0,1,52)	1.072	0.331
	WS-ARIMA	(2,0,4)	0.393	0.101
Ramagundam (Telangana)	ARIMA	(2,0,5)	0.637	0.271
	SARIMA	(1,0,0)(1,0,2,52)	0.940	0.319
	WS-ARIMA	(2,0,5)	0.377	0.128



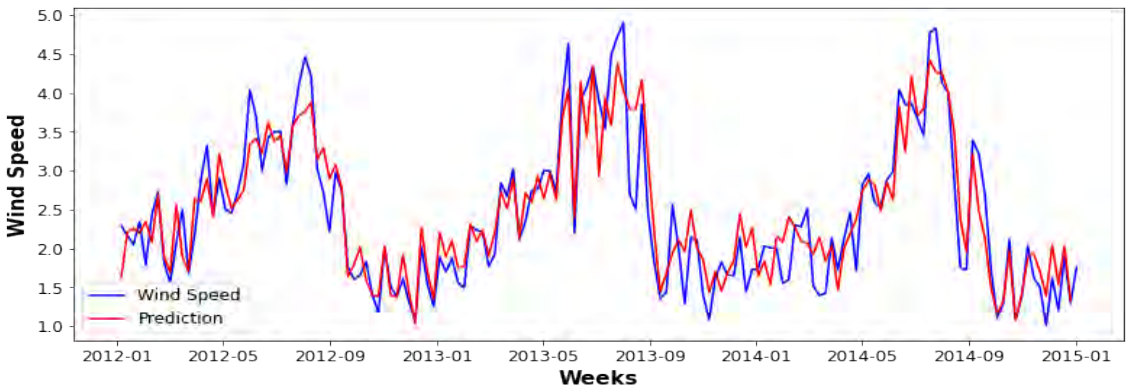
(a)



(b)

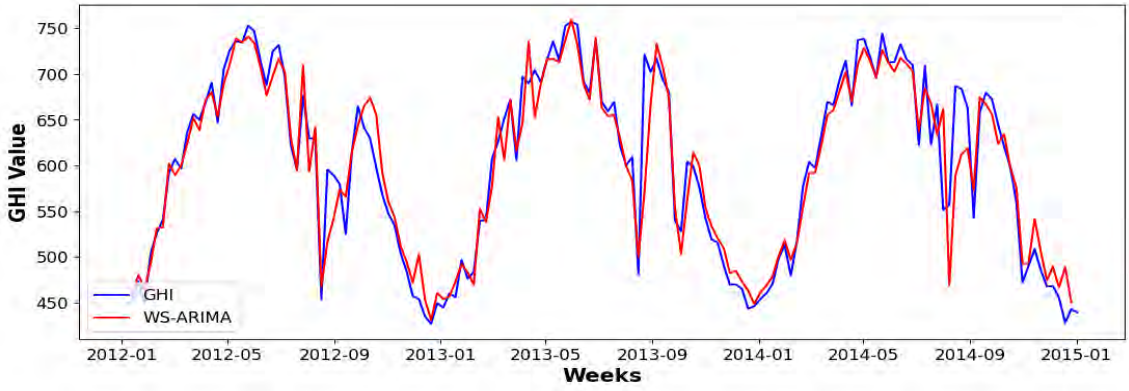


(c)

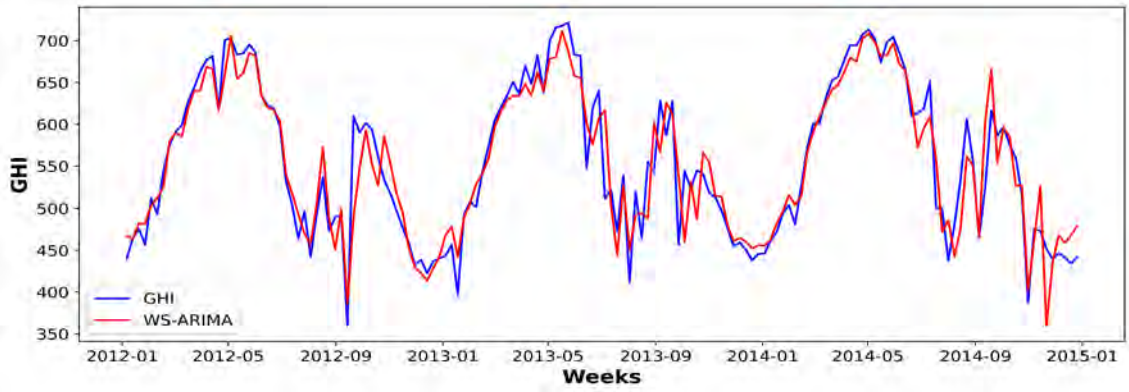


(d)

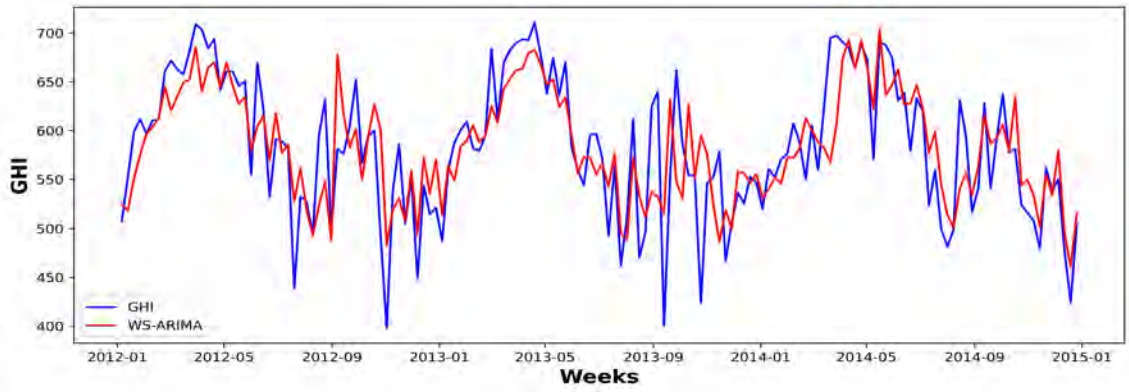
Fig. 3.4: Weekly wind speed forecasting from the best fit WS-ARIMA model for (a) Pokhran (Rajasthan), (b) Bitta (Gujarat), (c) Pavagada (Karnataka) and, (d) Ramagundam (Telangana).



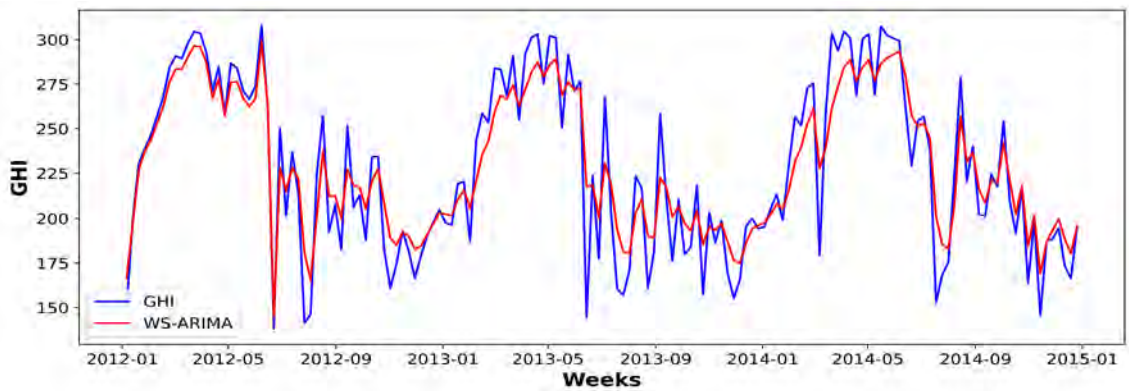
(a)



(b)



(c)



(d)

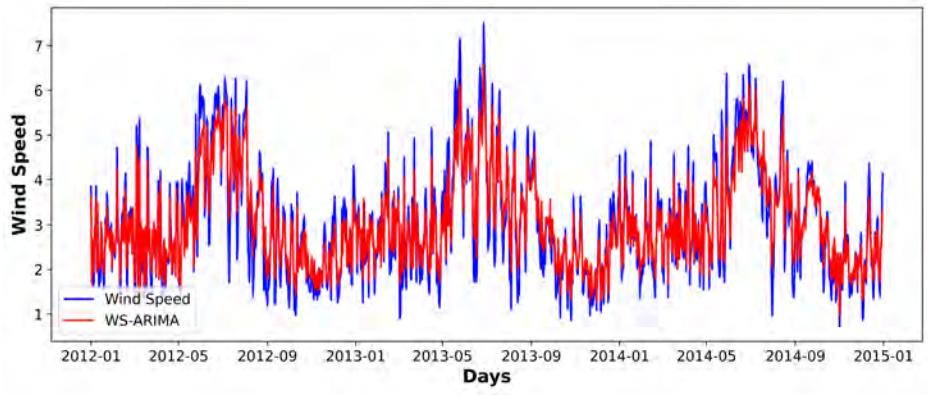
Fig. 3.5: Weekly GHI forecasting from the best fit WS-ARIMA model for (a) Pokhran (Rajasthan), (b) Bitta (Gujarat), (c) Pavagada (Karnataka) and, (d) Ramagundam (Telangana).

3.4.3 Results of Daily Forecasting

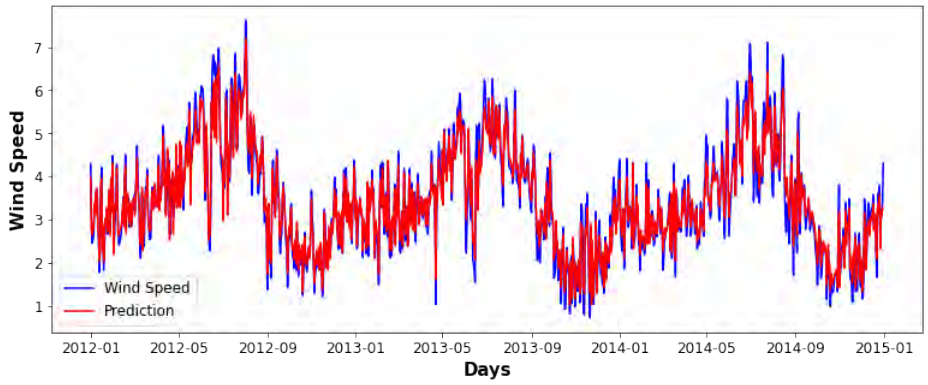
The RMSE and MAPE values of the implemented models are tabulated in Table 3.5 and Table 3.6 for wind speed and GHI datasets, respectively. Here, the WS-ARIMA model has the best performance followed by ARIMA and SARIMA models. The SARIMA model has the highest error values at daily resolutions, since there is seasonality of high order and more non-linear variability due to cloud coverage and precipitation in the environment. The WS-ARIMA model, in comparison to the conventional ARIMA method, yields RMSE reduction up to 75% in daily wind speed data. In GHI, the RMSE is reduced upto 50%. This strongly suggests that the inclusion of sliding windows with appropriate window lengths has improved capacity to deal with seasonality in a dataset. The daily forecasts from the best fit WS-ARIMA model for the four study sites are shown in Figure 3.6 and Figure 3.7.

Table 3.5: Daily wind speed forecasting at four locations

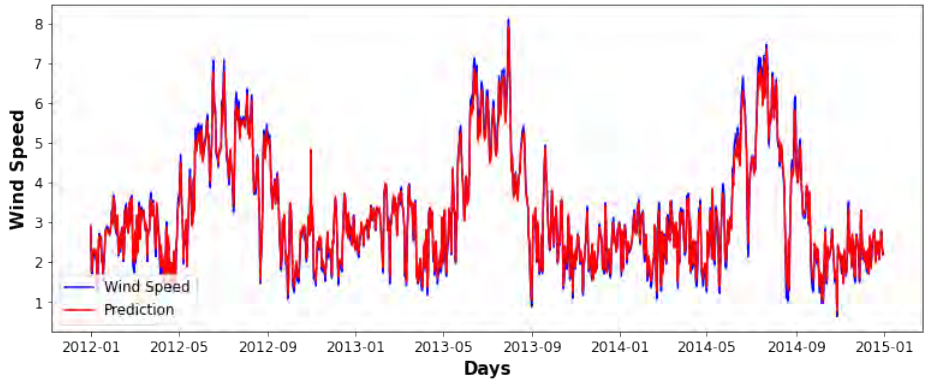
Daily Forecasting				
Location	**Model**	**Parameters**	**RMSE**	**MAPE**
Pokhran (Rajasthan)	AR	(40)	1.383	0.329
	MA	(2)	1.521	0.357
	ARIMA	(1,0,3)	1.317	0.349
	SARIMA	(1,1,0)(1,1,0,73)	2.067	0.762
	WS-ARIMA	(1,0,3)	0.372	0.113
Bitta (Gujarat)	ARIMA	(2,0,3)	1.126	0.227
	SARIMA	(1,1,0)(1,1,0,73)	2.212	0.701
	WS-ARIMA	(2,0,3)	0.282	0.073
Pavagada (Karnataka)	ARIMA	(2,0,2)	1.119	0.309
	SARIMA	(1,1,0)(1,1,0,73)	2.162	0.664
	WS-ARIMA	(2,0,2)	0.127	0.036
Ramagundam (Telangana)	ARIMA	(2,0,2)	0.926	0.375
	SARIMA	(1,1,1)(1,1,1,73)	1.352	0.504
	WS-ARIMA	(2,0,2)	0.232	0.093



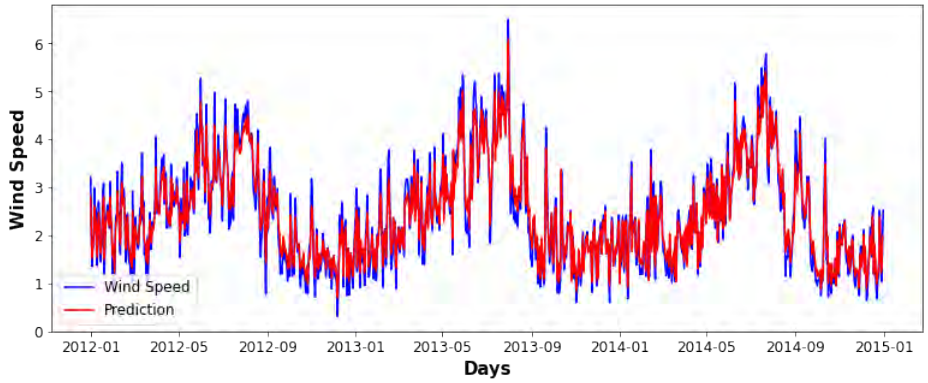
(a)



(b)

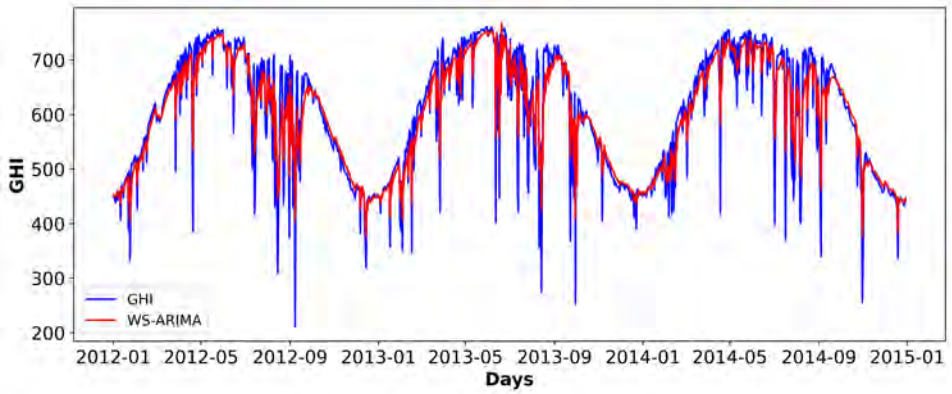


(c)

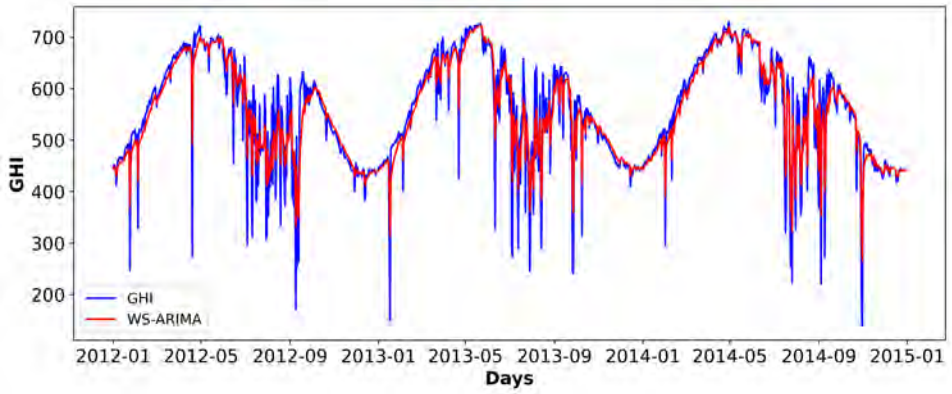


(d)

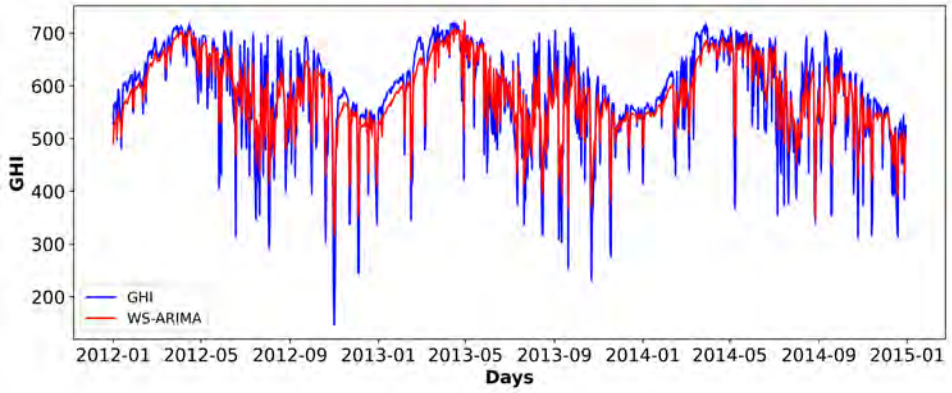
Fig. 3.6: Daily wind speed forecasting from the best fit WS-ARIMA model for (a) Pokhran (Rajasthan), (b) Bitta (Gujarat), (c) Pavagada (Karnataka), and (d) Ramagundam (Telangana).



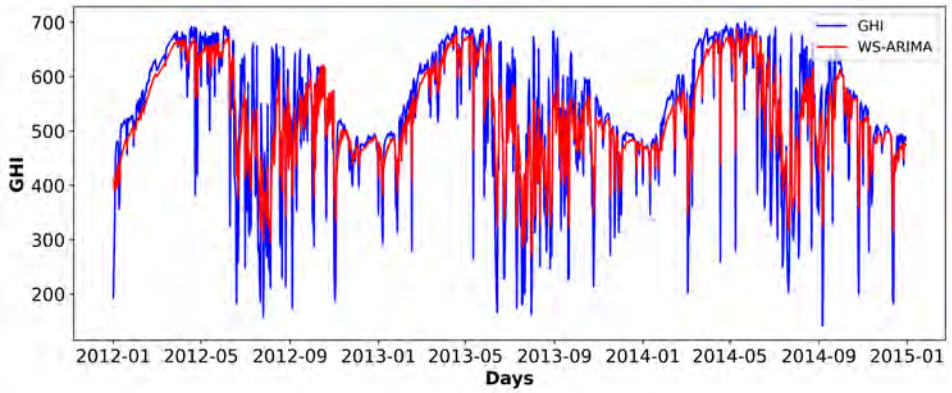
(a)



(b)



(c)



(d)

Fig. 3.7: Daily GHI forecasting from the best fit WS-ARIMA model for (a) Pokhran (Rajasthan), (b) Bitta (Gujarat), (c) Pavagada (Karnataka), and (d) Ramagundam (Telangana).

Table 3.6: Daily GHI forecasting at four locations

Daily Forecasting				
Location	**Model**	**Parameters**	**RMSE**	**MAPE**
Pokhran (Rajasthan)	AR	(49)	109.118	0.086
	MA	(13)	89.580	0.118
	ARIMA	(1,0,3)	73.330	0.069
	SARIMA	(2,1,2)(1,1,0,73)	81.677	0.093
	WS-ARIMA	(1,0,3)	34.804	0.041
Bitta (Gujarat)	ARIMA	(2,0,3)	77.952	0.092
	SARIMA	(2,1,0)(1,1,0,73)	84.718	0.106
	WS-ARIMA	(2,0,3)	33.671	0.045
Pavagada (Karnataka)	ARIMA	(2,0,2)	87.392	0.116
	SARIMA	(1,1,0)(1,1,0,73)	92.064	0.240
	WS-ARIMA	(2,0,2)	33.776	0.061
Ramagundam (Telangana)	ARIMA	(2,0,2)	75.660	0.102
	SARIMA	(2,1,0)(1,1,0,73)	100.941	0.215
	WS-ARIMA	(2,0,2)	52.633	0.085

3.4.4 Results of Hourly Forecasting

To carry out hourly forecasting using the WS-ARIMA, an extremely high seasonal order (i.e., $365 \times 24 = 8760$ hours in a year) needs to be considered. This turns out to be impractical for the present analysis. Nonetheless, one way to address this issue could be by considering several hit-and-trials of window lengths (much lesser order than 8760). The study by Reikard and Hansen [109] has used similar idea for forecasting solar irradiance at high resolution. Therefore, we note that the implementation of the WS-ARIMA or any other time series model for hourly wind speed and GHI forecasting is much more difficult than forecasting at other timescales. For the demonstration purpose, we have compared the performance of ARIMA and WS-ARIMA model using three years of data (2012–2013 for training and 2014 for testing). The error values and model parameters for hourly datasets are tabulated in Table 3.7 and Table 3.8.

Table 3.7: Hourly wind speed forecasting at four locations

Location	Model	Parameters	RMSE	MAPE
Pokhran	ARIMA	(4,1,4)	1.358	0.719
(Rajasthan)	WS-ARIMA	(4,1,4)	1.138	0.481
Bitta	ARIMA	(4,1,3)	1.247	0.472
(Gujarat)	WS-ARIMA	(4,1,3)	1.026	0.394
Pavagada	ARIMA	(5,1,1)	1.192	0.592
(Karnataka)	WS-ARIMA	(5,1,1)	1.230	0.615
Ramagundam	ARIMA	(2,1,5)	1.176	0.493
(Telangana)	WS-ARIMA	(2,1,5)	1.654	0.714

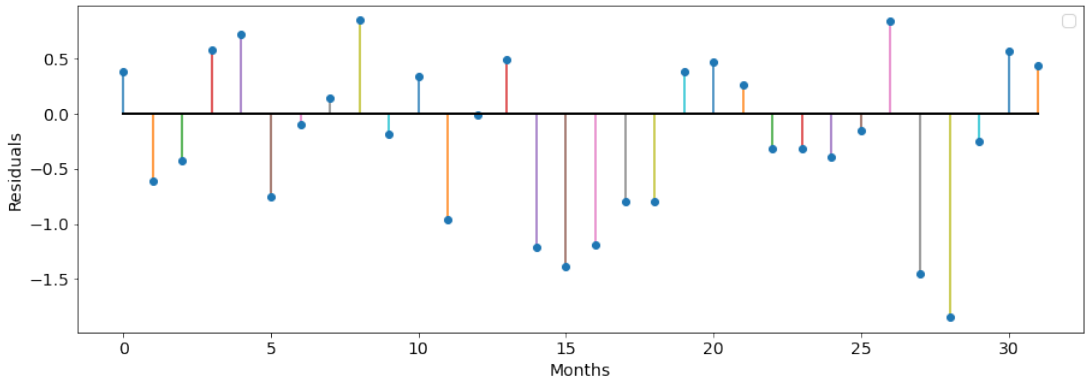
Table 3.8: Hourly GHI forecasting at four locations

Location	Model	Parameters	RMSE	MAPE
Pokhran	ARIMA	(2,0,5)	94.821	0.426
(Rajasthan)	WS-ARIMA	(2,0,5)	83.148	0.410
Bitta	ARIMA	(4,1,3)	91.247	0.500
(Gujarat)	WS-ARIMA	(4,1,3)	98.026	0.486
Pavagada	ARIMA	(4,1,3)	104.192	0.522
(Karnataka)	WS-ARIMA	(4,1,3)	71.230	0.407
Ramagundam	ARIMA	(3,0,1)	83.176	0.434
(Telangana)	WS-ARIMA	(3,0,1)	64.654	0.411

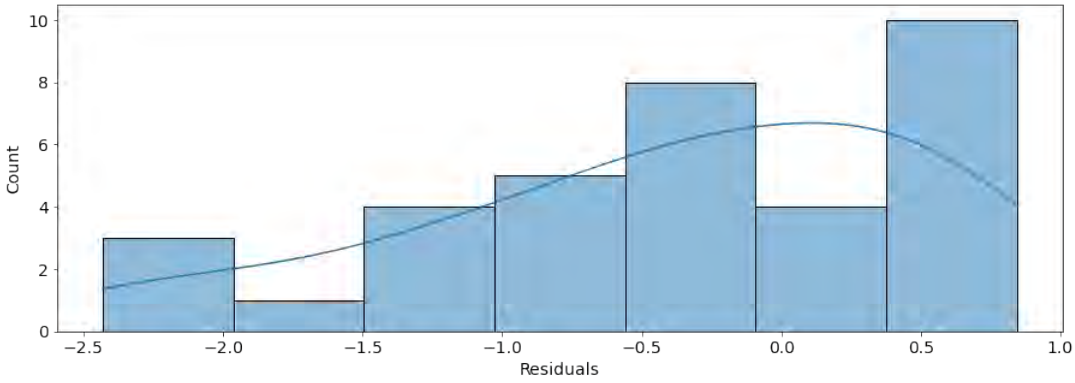
3.4.5 Results of Residual Analysis

As mentioned earlier, we perform residual analysis mainly to check whether there is any systematic bias in the implemented models. A forecast bias is a tendency for a model to consistently produce higher or lower forecast values than their actual values. Therefore, analysis of bias is an important post-processing step. The main characteristic of the unbiased models is the normality of the residuals. Therefore, if the residuals are not approximately of Gaussian shape, their randomness is lost, violating the fundamental assumption of a forecast model [8]. Here, we perform residual analysis through standardized residual plots and P-P plots corresponding to the best fit models at three chosen time horizons (Figure 3.8–3.13). From these figures, it appears that the histogram plots are bell shaped and the residual plots are symmetric about zero, exhibiting a normal distribution in each of these cases. However, the P-P plots have slight variation from the normal distribution. The deviation of residuals from normal distribution (in

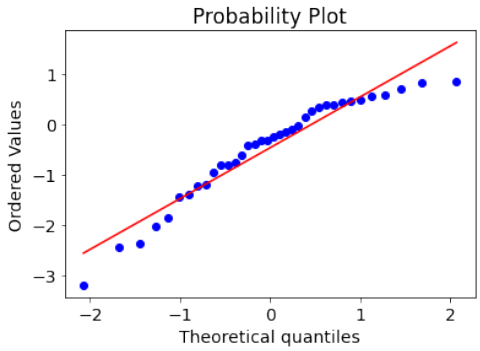
P-P plots) in daily and weekly data is more as compared to monthly data. This is possibly due to much more non-linear variability in weather patterns, causing presence of more outliers at shorter time horizons. Another reason could be the high order of seasonal fluctuations that involve differencing at the 52-week horizon and 365-day horizon, which typically requires a much longer time series to settle down [124].



(a)

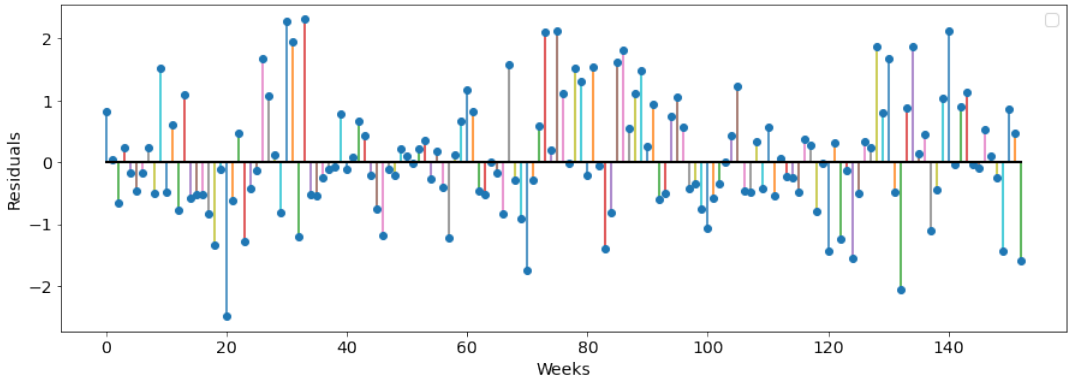


(b)

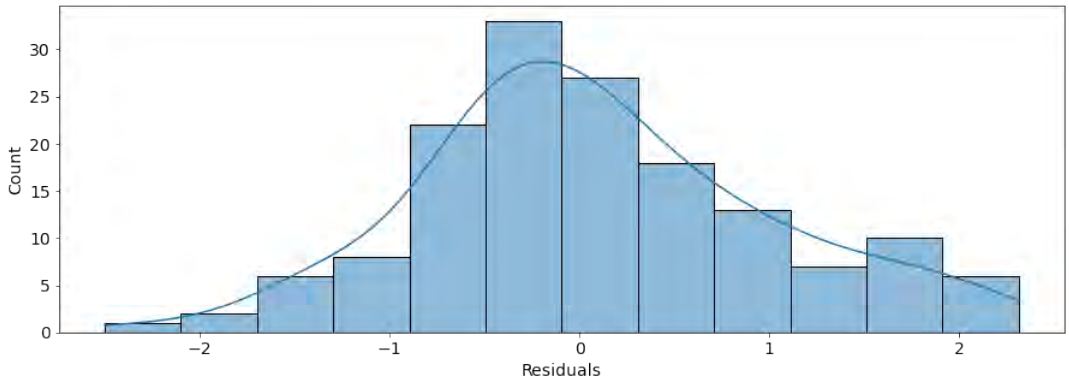


(c)

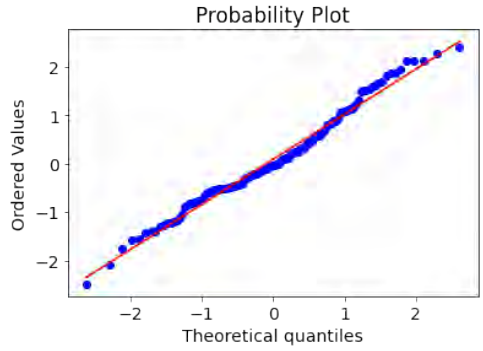
Fig. 3.8: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the WS-ARIMA model in monthly wind speed forecasting at Pokhran, Rajasthan.



(a)

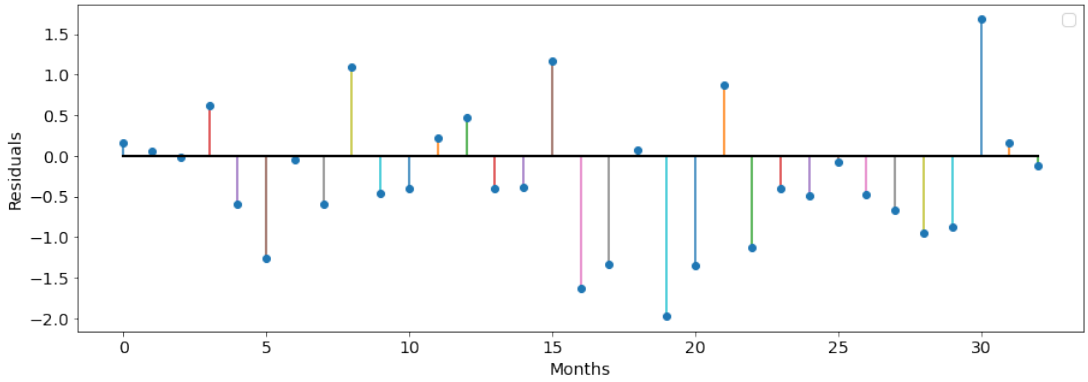


(b)

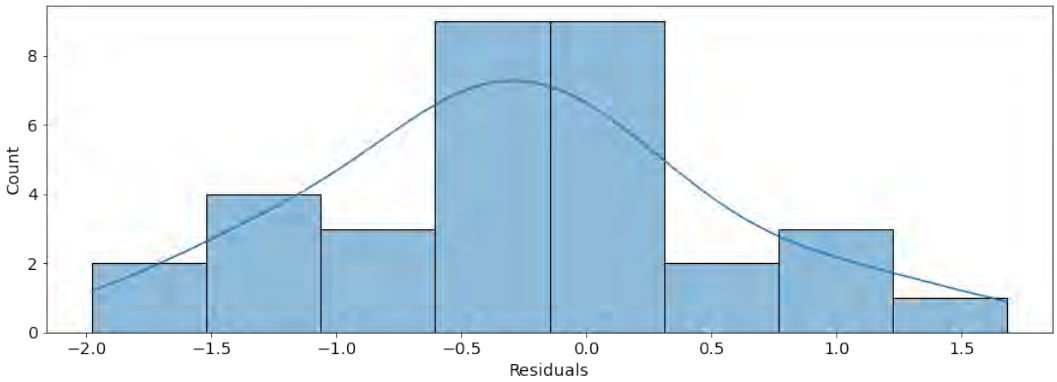


(c)

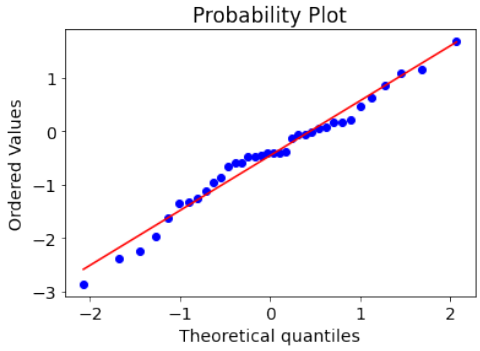
Fig. 3.9: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the WS-ARIMA model in weekly wind speed forecasting at Pokhran, Rajasthan.



(a)

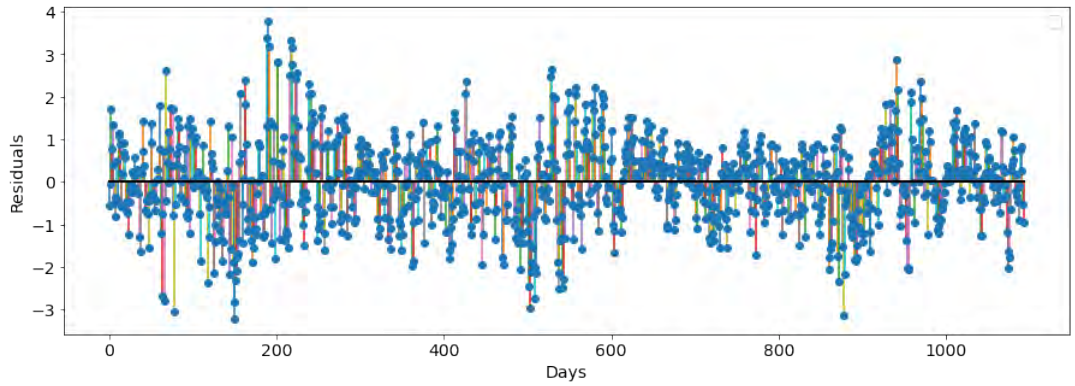


(b)

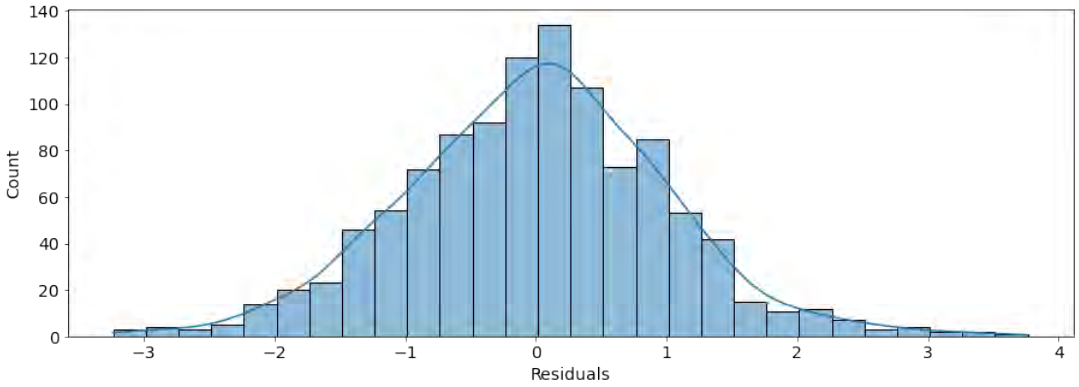


(c)

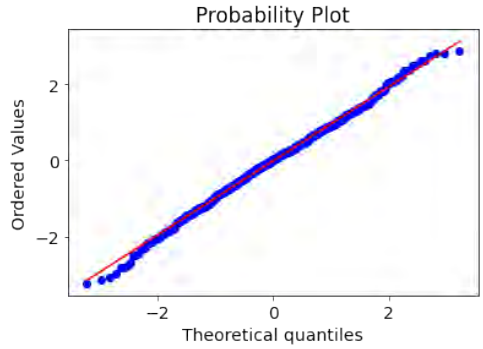
Fig. 3.11: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the SARIMA model in monthly GHI forecasting at Pokhran, Rajasthan.



(a)

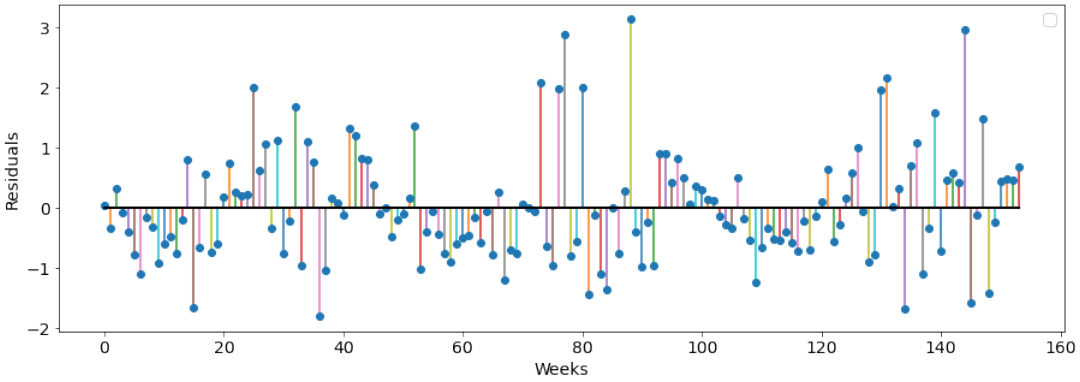


(b)

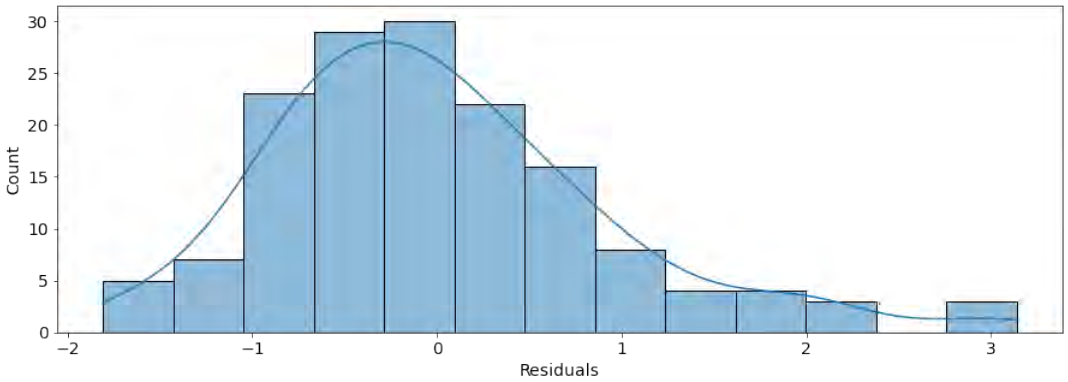


(c)

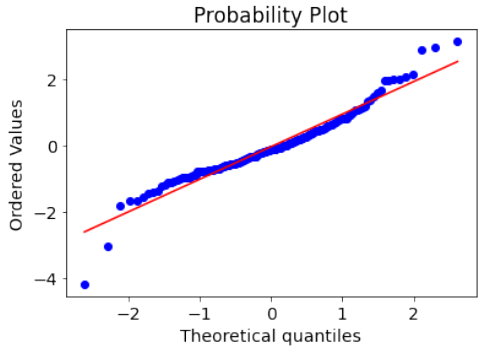
Fig. 3.10: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the WS-ARIMA model in daily wind speed forecasting at Pokhran, Rajasthan.



(a)

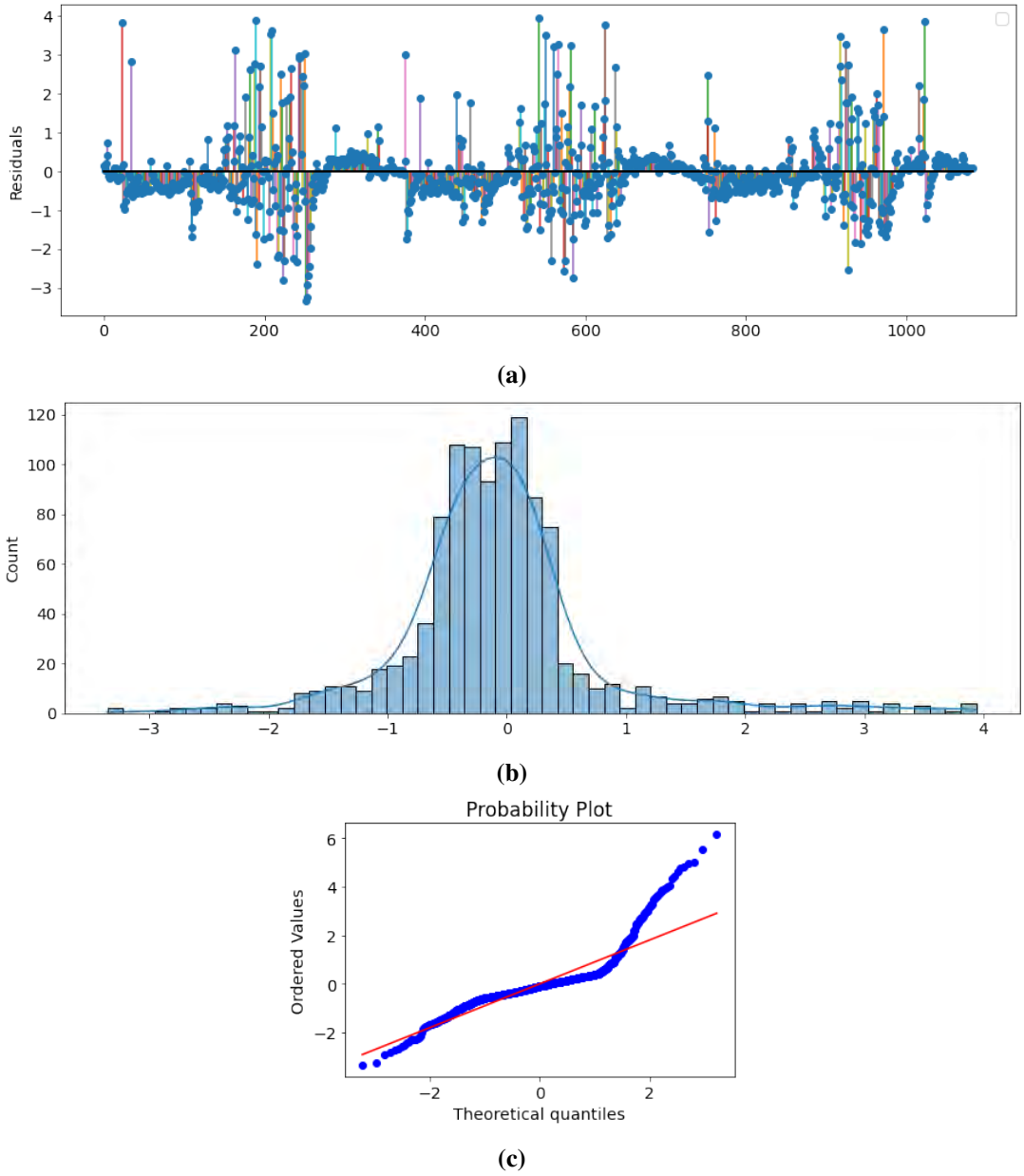


(b)



(c)

Fig. 3.12: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the WS-ARIMA model in weekly GHI forecasting at Pokhran, Rajasthan.



(a)

(b)

(c)

Fig. 3.13: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the WS-ARIMA model in daily GHI forecasting at Pokhran, Rajasthan.

3.5 Summary

In this chapter, we have implemented five time-series models for four selected locations in the Indian region to study their efficacy in forecasting of wind speed and solar irradiance. We demonstrate the applicability of the WS-ARIMA model for daily and weekly wind speed and

GHI prediction. Based on the obtained results, the present research leads to the following conclusions regarding wind speed and GHI forecasting.

1. The wind speed and GHI data at four selected sites exhibit stationary behavior. Therefore, neither differencing nor detrending was required for daily, weekly, and monthly forecasting. As a consequence, the ARIMA model turns out to be a simple ARMA model in each case.

2. For the monthly forecasting, the SARIMA model consistently produces the best results across all locations. This is because the SARIMA model can precisely capture the yearly seasonality (at 12 months) in the data. However, for daily and weekly forecasts, it is hard to exactly recognize the seasonality pattern in the data through the SARIMA model.

3. In the weekly forecasting, the WS-ARIMA model, in comparison to the conventional ARIMA method, yields RMSE reduction up to 50%.

4. In the daily wind speed and GHI forecasting, the WS-ARIMA model, in comparison to the conventional ARIMA method, yields RMSE reduction up to 75% and 50%, respectively. This strongly suggests that the inclusion of sliding windows with appropriate window lengths improves capacity to deal with seasonality in the data.

5. For hourly forecasting, the WS-ARIMA and ARIMA model have comparable performance based on three years (2012–2014) of data. In fact, we have noted that at hourly timescale, the implementation of the WS-ARIMA is comparatively difficult than daily, weekly, or monthly timescales due to high non-linear and seasonal fluctuations. Thus, additional works are necessary to figure out the optimal window length for hourly forecasting.

The emanated results from the time series models in this chapter will be further compared with the machine learning and hybrid models in Chapter 5.

Chapter 4

Machine Learning Models for Renewable Energy Forecasting

"*Any time you try to do something innovative, you should expect that there's always going to be people who doubt it, who suggest that perhaps you would be better off doing something else.*"

– ELON MUSK

This chapter focuses on hourly, daily, weekly, and monthly forecasting of wind speed and GHI using various machine learning methods, such as support vector regression (SVR), artificial neural network (ANN), long short term memory (LSTM), and convolutional neural network (CNN). We choose these models since they allow for non-linear associations and enable learning about the relationship among variables from data. The SVR is a supervised-learning approach which equally penalizes high and low mis-estimates and thereby, overcome the over fitting. The ANN model performs a non-linear functional mapping from the past observations to the future value, being equivalent to a non-linear autoregressive model. The LSTMs have an internal memory in the form of gates. This allows the LSTM to accumulate important information which may be required in future. Thus, the LSTM, a subset of recurrent neural network (RNN), is able to control the process of remembering information for noticeably longer periods of time. The CNNs are well-known as a reliable tool for extracting hidden features and creating filters based on data patterns through its three mapping layers, namely the convolutional layer, pooling layer, and the fully connected layer. All these machine learning models are governed by several number of parameters, such as the type and number of hidden layers, activation function, optimization algorithm, loss function, epochs, and the learning rate. From a spectrum of possible parameter values, we obtain the optimal parameters such that the error values are minimized. The emanated results are finally validated through residual analysis as a post-processing step of the implemented models.

Contents

Parts of this chapter have been published in the following refereed publications:

S. Sheoran, R. S. Singh, S. Pasari, R. Kulshrestha, "Forecasting of solar irradiances using time series and machine learning models: A case study from India," *Applied Solar Energy*, vol.58, pp. 137–151, 2022.
S. Sheoran, S. Shukla, S. Pasari, R.S. Singh, R. Kulshrestha (2022), "Wind speed forecasting at different timescales using time series and machine learning models," *Applied Solar Energy*, vol. 58, pp. 708–721, 2022.

4.1 Introduction

Machine learning models are computational methods that can learn from data and make predictions or decisions based on the learned patterns. Machine learning models have been very useful for forecasting of renewable energy, such as wind and solar power, which are highly dependent on weather conditions and exhibit stochastic behavior. Forecasting of renewable energy can help optimize the management and integration of renewable energy sources into the power grid, as well as can improve the reliability and stability of electricity supply. There have been several studies on the forecasting of wind speed and solar irradiance using machine learning techniques due to their intrinsic flexibility and robustness in data analysis. In forecasting, the advantage of using machine learning models is that they allow for non-linear associations and enable learning from data without making assumptions about the relationship among variables [30]. Such a facility is not available in several explicit algorithms. A few drawbacks or the problems that creep in the machine learning techniques include over-fitting (which basically means that it fits the training set perfectly but is unable to predict future outcomes, and thereby, it does not accurately predict the test set), usage of high computational power of computer systems, extensive need of hyper-parameters tuning, and high dimensional data. As a result, machine learning models are very difficult to handle due to their inherent complexity [94].

The most common machine learning models in renewable energy forecasting are SVR [125], SVM [125], ANN [54], k-NN [54], MLP [38], and LSTM [38, 74]. These models have provided satisfactory results for the wind speed and solar irradiance forecasting at different time horizons, specially in short term forecasting. A study on solar potential of Himachal Pradesh (India) was performed by Yadav and Chandal [160]. They used an ANN based global solar radiation model to assess the solar potential. Gensler et al. [38] compared performances of MLP, LSTM, DBN (Dynamic Bayesian Network), and auto-LSTM machine learning models on a German solar power dataset. The best performing model is the auto-LSTM, closely followed by the DBN model. In 2018, Qing et al. [105] proposed hourly day-ahead solar energy prediction using weather data. They compared the persistence model with ANN based models and found that the proposed algorithm outperforms the ANN based models. Dubey et al. [34] in 2021 compared the results of ARIMA, SARIMA, and LSTM on influence of energy consumption. They considered the power consumption related to the features' relations based on temperature, humidity, cloud cover, visibility, and wind index. The results revealed that the LSTM outperforms other models with an average MAE of 0.23. In 2018, Yang et al. [162] provided a comprehensive review on history and trends on solar irradiance prediction. The article comprises techniques ranging from time series and regression to artificial intelligence. Aimeur et al. [4] proposed octonion neural network to investigate the short term

forecast of solar irradiance. In comparison to the real-valued neural networks, the implemented octonion training algorithm contains eight dimensions and it provides eight values ahead forecast of solar irradiance. In 2021, Rabehi et al. [107] provided a comparative study of several machine learning techniques for solar radiation assessment in semi-arid region. The analysis demonstrates that when compared to other models, Gaussian process regression (GPR) and least-squares support vector machine (LS-SVM) models offer a high performance. Guariso et al. [43] implemented feed-forward neural network and LSTM models and discussed conceptual and computational differences between the network architectures for forecasting of solar irradiance in Italy. Mukhoty et al. [87] studied two variants of LSTM, namely encoder-decoder networks of LSTM and bidirectional LSTM or BiLTSM. Brahma and Wadhvani [18] studied several variants of LSTMs and found that the BiLSTM and attention-based LSTM models have the best representation for daily solar irradiance data in the Indian region.

In a similar manner, several machine learning models have been studied in forecasting of wind speed. Pasari et al. [100] carried out wind speed prediction using the ANN techniques. They implemented single step and multistep neural networks and concluded that a univariate single layer architecture provides better accuracy for wind speed prediction. Wu et al. [157] built a two layer neural network for wind speed forecasting. Masqood et al. [75] performed a comparative study of multi layer perceptrons (MLP), Elman recurrent neural networks, radial basis function networks, and Hopfield model for hourly weather prediction, including wind speed. Liu et al. [68, 69] performed wind speed forecasting using various machine learning methods, such as ANNs, SVMs, RNNs, auto-encoders, and LSTMs. In 2020, Choe et al. [25] highlighted the efficacy of both LSTM and bidirectional LSTM, known to be effective on capturing long term time dependency in order to predict the wind speed and in turn the wind energy. Binsu et al. [56] also implemented a LSTM based hybrid model along with the integration of an autoencoder (stacked) to forecast the wind speed. In 2021, Shobanadevi et al. [126] proposed a R-LSTM (rolling-LSTM) for the prediction of wind power in the state of Gujarat. Based on the recursive algorithm, the model provided accurate results. In 2021, Lin et al. [67] also applied the LSTM model with other deep learning algorithms including the TCN (temporal convolution network) to predict wind power. Some comprehensive reviews of time series and machine learning approaches in wind speed prediction are also available in literature [12, 15, 90, 114, 116, 135].

In view of the above discussed literature survey, the current chapter focuses on hourly, daily, weekly, and monthly forecasting of wind speed and GHI using various machine learning methods, namely the SVR, ANN, LSTM, and the CNN. The SVR, a supervised-learning approach, equally penalizes high and low mis-estimates and hence can overcome the over fitting. The

ANN model performs a non-linear functional mapping from the past observations to the future value being equivalent to a non-linear autoregressive model. The LSTMs have an internal memory to enhance the process of remembering information regarding the pattern of underlying process. The CNNs are well-known as a reliable tool for extracting hidden features through its three mapping layers, namely convolutional layer, pooling layer, and fully connected layer. We have computed the forecast iteratively, that is, only one data point is calculated at once. We finally compare the results obtained from the implemented models at different timescales corresponding to four study sites. The emanated results are finally validated through residual analysis as a post-processing step of the implemented models.

4.2 Implemented Models

A brief description of the implemented models is provided in following subsections.

4.2.1 Support Vector Regression (SVR)

The support vector regression was first introduced by V. N. Vapnik [144] who along with his collaborators also introduced support vector machines (SVM) [16, 28]. The SVR is a generalization to the SVM and is accomplished by introducing an ε-insensitive (using Vapnik's ε-insensitive approach [166]) region around the base hyperplane, called the ε-tube as shown in Figure 4.1. As a supervised learning approach, the SVR trains using a symmetrical loss function (ε-insensitive loss function) which equally penalizes high and low mis-estimates. Hence, the SVR can overcome the issue of over fitting [89, 117].

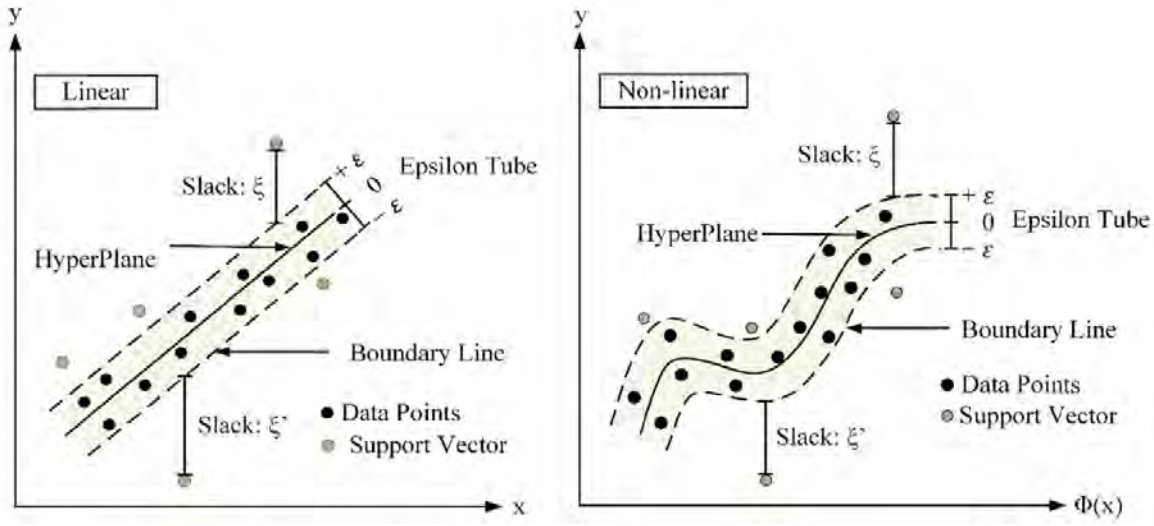


Fig. 4.1: Illustration of linear and non-linear SVR model with $\varepsilon-$insensitive loss function [31].

From the above figure, it is observed that there is ε distance between the hyperplane (predictor) and the two boundary lines (decision boundaries). There are several data points which become potential support vectors, meaning that these data points can become potential boundaries. The purpose of the SVR is to find a function as a hyperplane in the form of a regression function that matches all data input with the smallest possible error. The purpose of this function is to map the input vector into a higher dimensional feature space in which the training data may exhibit linearity, and then to perform linear regression in this feature space. Thus, the SVR is more precisely defined as an optimization problem by (i) constructing a convex ε-insensitive loss function to be reduced and (ii) locating the latest tube that includes the majority of the training points while balancing model complexity and prediction error. Following is the formulation of non-linear SVR. The continuous-valued function y being approximated can be written as in Equation 4.1. For multidimensional data, we augment x by one and include b in the w vector to simplify the mathematical notation, and obtain the multivariate regression as shown in Equation 4.2.

$$y = f(x) = < w, x > + b = \sum_{j=1}^{M} w_j x_j + b; \qquad y, b \in R, x, w \in R^M \tag{4.1}$$

$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} x \\ 1 \end{bmatrix} = w^T x + b \tag{4.2}$$

This regression problem can be expressed as an optimization problem as shown in Equation 4.3. Using a soft-margin approach similar to that employed in SVM, slack variables ξ_i and ξ_i^* are added with a regularization constant C which decides the error value whenever the data point is out of the tube (outliers). Thus, C is a tuneable parameter that provides more weight to minimizing the flatness, or the error, for this multi-objective optimization problem. Note that the problem formulated is a convex optimization problem and is further solved by the use of Lagrangian multipliers (λ, λ^*, α, α^*) as in Equation 4.4. The numbers λ, λ^*, α, and α^* are non-negative real numbers.

$$min \ \ \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{4.3}$$

subject to

$$y_i - w^T x_i \leq \varepsilon + \xi_i^* \quad i = 1, ..., N$$

$$w^T x_i - y_i \leq \varepsilon + \xi_i \quad i = 1, ..., N$$

$$\xi_i, \xi_i^* \geq 0 \quad i = 1, ..., N$$

$$\mathscr{L}(w,\xi_i^*,\xi_i,\lambda,\lambda^*,\alpha,\alpha^*) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi_i+\xi_i^*) + \sum_{i=1}^{N}\alpha_i^*(y_i - w^T x_i - \varepsilon - \xi_i^*) + \sum_{i=1}^{N}\alpha_i(-y_i + w^T x_i - \varepsilon - \xi_i) - \sum_{i=1}^{N}(\lambda_i\xi_i + \lambda_i^*\xi_i^*) \tag{4.4}$$

The minimum value of the function in Equation 4.4 is found by taking its partial derivatives with respect to the variables and setting them equal to zero. The final solution is written in terms of the number of support vectors (N_{SV}). The function approximation is represented in Equations 4.5, 4.6, and 4.7. The dual form of the optimization problem can be written as shown in Equation 4.7.

$$\omega = \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)x_i \tag{4.5}$$

$$f(x) = \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i){x_i}^T x_i; \qquad \alpha_i, \alpha_i^* \in [0,C] \tag{4.6}$$

$$max \quad \varepsilon\sum_{i=1}^{N_{SV}}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)y_i - \frac{1}{2}\sum_{j=1}^{N_{SV}}\sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)x_i^T x_j, \tag{4.7}$$

subject to,

$$\sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i) = 0; \qquad \alpha_i, \alpha_i^* \in [0,C] \tag{4.8}$$

The above problem can be further solved using quadratic programming.

In addition, kernels are often used when given the data points are not linear. It maps the data into higher dimensional space called the kernel space to obtain a higher accuracy. Therefore, replacing all instances of x in Equations 4.1–4.8 with $k(x_i,x_j)$ yields the primal formulation shown as in Equation 4.9, where $\phi(.)$ is the transformation from feature to kernel space. Equation 4.10 describes the new weight vector in terms of the transformed input. The dual problem formed is represented in Equation 4.11.

$$min \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{4.9}$$

subject to

$$y_i - w^T\phi(x_i) \leq \varepsilon + \xi_i^*; \quad i = 1,...,N$$

$$w^T\phi(x_i) - y_i \leq \varepsilon + \xi_i; \quad i = 1,...,N$$

$$\xi_i, \xi_i^* \geq 0; \quad i = 1,...,N$$

$$\omega = \sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i)\phi(x_i) \tag{4.10}$$

$$max \quad \varepsilon \sum_{i=1}^{N_{SV}} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i) y_i - \frac{1}{2} \sum_{j=1}^{N_{SV}} \sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j) \tag{4.11}$$

$$\sum_{i=1}^{N_{SV}} (\alpha_i^* - \alpha_i) = 0; \quad i = 1, ..., N_{SV}, (\alpha_i, \alpha_i^*) \in [0, C]$$

$$k(x_i, x) = \phi(x_i).\phi(x)$$

For implementation, we carry out the univariate study for GHI and wind speed datasets separately. For this, (i) we perform data preprocessing to get the resampled values at hourly, daily, weekly, and monthly basis; (ii) then, we choose (through several hit and trials) the lag value that specifies the number of previous data points on which the forecast value depends; (iii) after this, we arrange the data in form of input vectors in such a way that each tuple's (lag size) entries represent a feature vector, and (iv) these input vectors are fed to the non-linear SVR regressor with radial basis function (RBF) kernel for training. It may be noted that the RBF kernel is chosen so as to recognize the maximum non-linearity in the data [26]. The parameters C and ε are chosen optimally by a grid based search for each of the cases. Finally, the model is run for the test data and the prediction errors are calculated.

4.2.1.1 Results

The results in terms of RMSE, MAPE, and model parameters (C, ε, and lag value) for wind speed and GHI datasets corresponding to four selected study sites are provided in Table 4.1. We observe that in GHI, the RMSE values increase if we move from longer to shorter timescale, i.e., from monthly to weekly to daily, and to hourly timescale. In wind speed, the results do not reflect such pattern. For illustration, we have plotted the actual versus predicted values of wind speed and GHI data for Pokhran, Rajasthan in Figure 4.2 and Figure 4.3, respectively.

4.2.2 Artificial Neural Network (ANN)

An artificial neuron system is a data processing system that is constructed by imitating human neurons [81]. It is a computational model made up of a vast number of interconnected neurons. The neuron can be thought of storage and calculating unit. In most of the cases, a single neuron has several inputs. Every link between two neurons in a neural network is represented by a weight that reflects the strength of the connection. Finally, layers of neurons are combined

Table 4.1: Details of the SVR model for wind speed and GHI forecasting at different timescales at four study sites (L1: Pokhran, Rajasthan, L2: Bitta, Gujarat, L3: Pavagada, Karnataka, and L4: Ramagundam, Telangana)

Wind Speed						
Timescale	**Location**	**MAPE**	**RMSE**	**C**	ε	**Lag Value**
Monthly	L1	0.087	0.417	1.000	0.001	20
	L2	0.087	0.328			
	L3	0.092	0.422			
	L4	0.104	0.359			
Weekly	L1	0.174	0.663	0.100	0.100	16
	L2	0.132	0.571			
	L3	0.170	0.709			
	L4	0.174	0.542			
Daily	L1	0.202	0.709	1.000	0.010	5
	L2	0.177	0.642			
	L3	0.151	0.526			
	L4	0.209	0.578			
Hourly	L1	0.063	0.174	default	default	4
	L2	0.059	0.186			
	L3	0.057	0.186			
	L4	0.086	0.209			
GHI						
Monthly	L1	2.094	13.193	1.000	0.100	12
	L2	3.584	24.945			
	L3	2.853	18.756			
	L4	4.825	30.638			
Weekly	L1	0.050	38.353	1.000	0.001	11
	L2	0.053	42.852			
	L3	0.076	55.551			
	L4	0.092	61.179			
Daily	L1	0.063	57.515	0.100	0.010	2
	L2	0.075	59.343			
	L3	0.095	71.482			
	L4	0.151	93.687			
Hourly	L1	0.178	63.591	default	default	24
	L2	0.169	61.382			
	L3	0.289	83.040			
	L4	0.323	81.592			

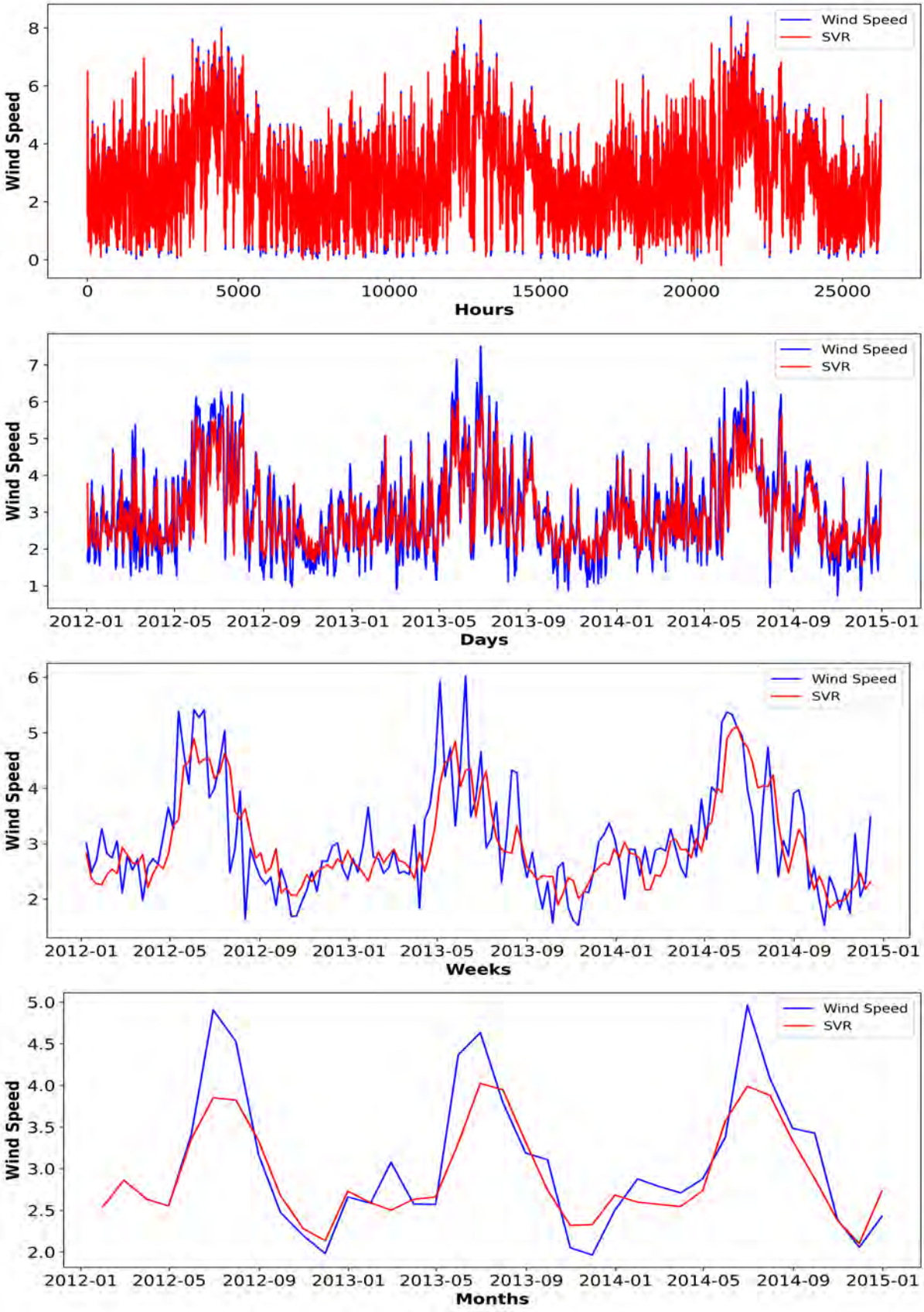


Fig. 4.2: Wind speed forecasting through the SVR model for (a) hourly, (b) daily, (c) weekly, and (d) monthly data of Pokhran, Rajasthan.

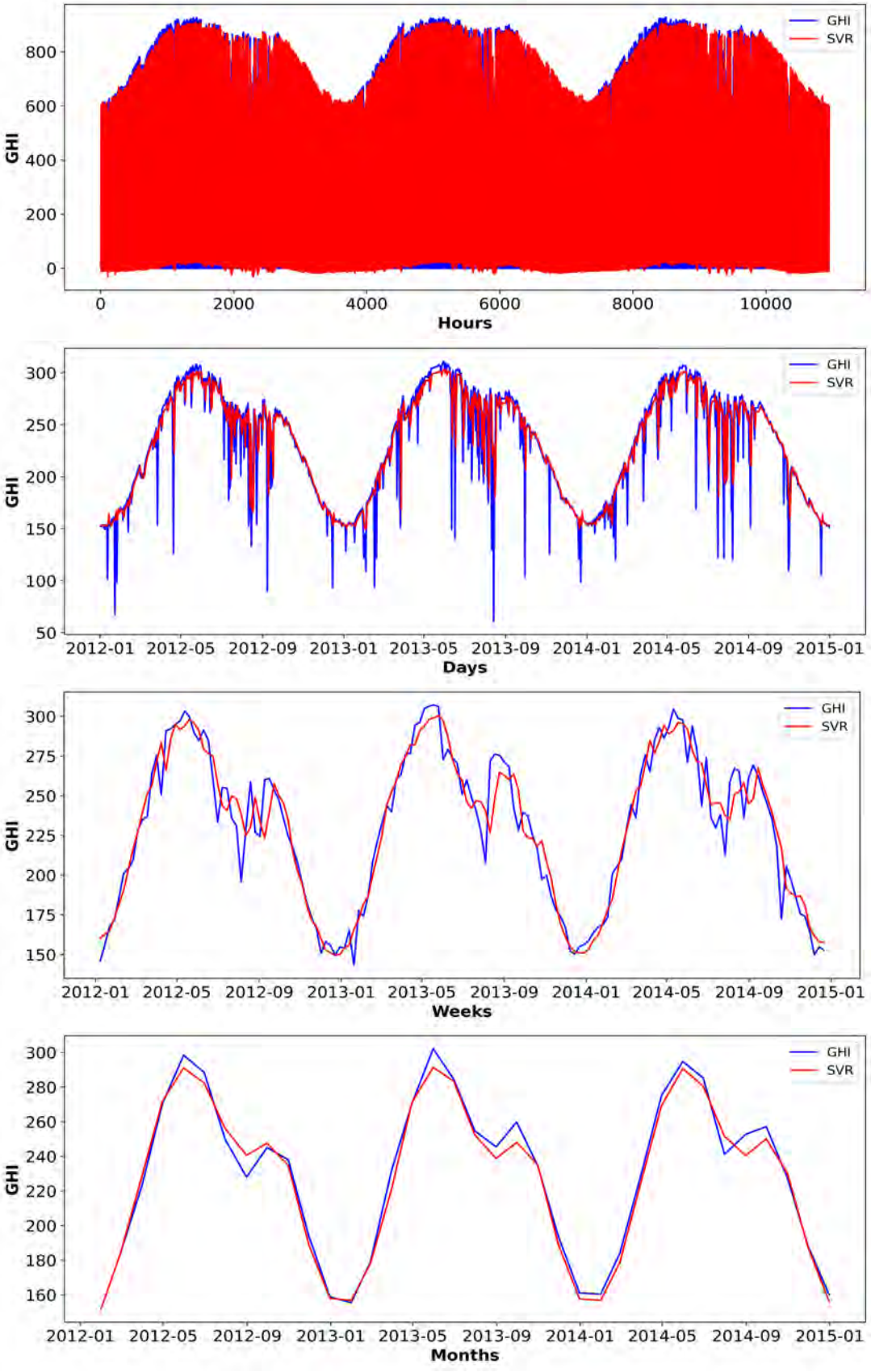


Fig. 4.3: GHI forecasting through the SVR model for Pokhran, Rajasthan at (a) hourly, (b) daily, (c) weekly, and (d) monthly timescales.

to form a neural network as shown in Figure 4.4. The hyper-parameters of ANN include the learning rate, number of hidden layers, and the batch size [80]. The specific values of these parameters have high impact on the model output.

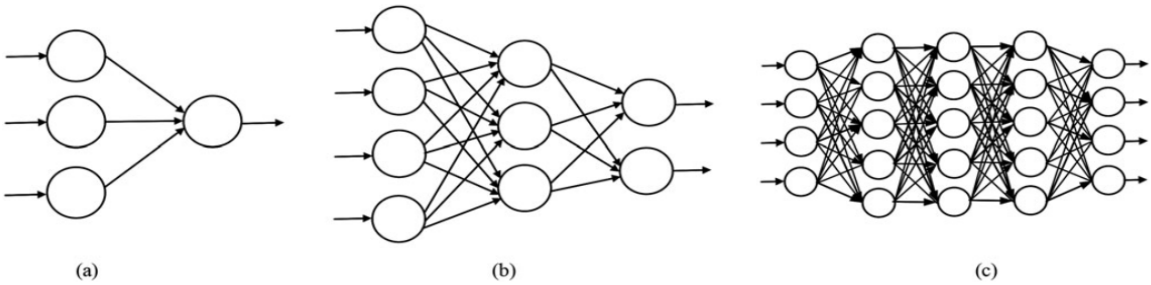


Fig. 4.4: Illustrative figure of ANN highlighting the development of neural network structure: (a) perceptron, (b) multi layer perceptron (MLP), and (c) deep learning [82].

When the input vector $[x_0, x_1, x_2, x_3, ..., x_i]$ is entered to the ANN, the output at the j^{th} hidden layer neuron is generated as below:

$$Y_j = \sum_{i=1}^{N_i} \left(w_{ij} x_i + b_j \right) \tag{4.12}$$

where, N_i is the number of neurons in the input layer; $[w_{1j}, w_{2j}, w_{3j}, ..., w_{ij}]$ is the connection weight vector of the i^{th} input layer neuron to the j^{th} hidden layer neuron, and b_j is the bias value connected to the j^{th} hidden layer neuron. Then, Y_j is processed by transfer function $f(.)$ into Z_k, the output at k^{th} output layer neuron, as shown below:

$$Z_k = f\left[\sum_{k=1}^{N_h} \left(w_{jk} f_1(y_j) + b_k \right) \right] \tag{4.13}$$

where, N_h refers to the hidden layer neurons; $(w_{1k}, w_{2k}, w_{3k}, ..., w_{jk})$ is the connection weight vector from the j^{th} hidden layer neuron to the k^{th} output layer neuron, and b_k is the bias value connected to the k^{th} output layer neuron.

For time series modeling and forecasting, single hidden layer feedforward network is the most widely used model. Hence, the ANN model in fact performs a non-linear functional mapping from the past observations (x_{t-1}, x_{t-2},...,x_{t-p}) to the future value y_t, as in Equation 4.14.

$$y_t = f(x_{t-1}, x_{t-2}, ..., x_{t-p}; w) + \varepsilon_t \tag{4.14}$$

Here, w is a vector of all parameters and f is a function determined by the network structure and connection weights. Thus, the neural network is equivalent to a non-linear autoregressive

model [115]. Note that the expression in Equation 4.14 represents one output node in the output layer which is typically used for one-step-ahead forecasting.

4.2.2.1 Results

The optimal parameters in terms of the activation function, number of hidden layers, optimizer, loss function, and the number of epochs of the ANN model for wind speed and GHI forecasting are noted in Table 4.2 and Table 4.3, respectively. The forecasting errors in terms of RMSE and MAPE values are also included in these tables. We observe that in both wind speed and GHI datasets, the RMSE values increase as we move from monthly to weekly, to daily, and to hourly data. Although most of the times, the single hidden layer network provides the best results, there are few instances where two hidden layers enable better results. The pictorial representation of original values versus forecasted values of both wind speed and GHI are provided in Section 4.3.

Table 4.2: Results of the ANN model for wind speed forecasting at different timescales and locations

Location	Time-scale	Activation Function	Hidden Layers	Optimizer	Loss Function	Epochs	RMSE	MAPE
Pokhran, Rajasthan	Monthly	RELU	1	ADAM	MSE	2000	0.114	0.032
	Weekly	RELU	1	ADAM	MSE	2000	0.449	0.121
	Daily	RELU	2	ADAM	MSE	200	0.743	0.247
	Hourly	RELU	1	ADAM	MSE	50	1.002	0.391
Bitta, Gujarat	Monthly	RELU	2	ADAM	MSE	2000	0.104	0.028
	Weekly	RELU	1	ADAM	MSE	2000	0.338	0.080
	Daily	RELU	1	ADAM	MSE	200	0.642	0.176
	Hourly	RELU	2	ADAM	MSE	50	0.930	0.328
Pavagada, Karnataka	Monthly	RELU	1	ADAM	MSE	2000	0.102	0.026
	Weekly	RELU	1	ADAM	MSE	2000	0.405	0.109
	Daily	RELU	2	ADAM	MSE	200	0.523	0.155
	Hourly	RELU	1	ADAM	MSE	50	0.917	0.287
Ramagundam, Telangana	Monthly	RELU	2	ADAM	MSE	2000	0.118	0.039
	Weekly	RELU	1	ADAM	MSE	2000	0.342	0.119
	Daily	RELU	1	ADAM	MSE	200	0.583	0.219
	Hourly	RELU	2	ADAM	MSE	50	1.031	0.369

Table 4.3: Results of the ANN model for GHI forecasting at different timescales and locations

Location	Time-scale	Activation Function	Hidden Layers	Optimizer	Loss Function	Epochs	MAPE	RMSE
Pokhran, Rajasthan	Monthly	RELU	1	ADAM	MSE	2000	0.010	9.267
	Weekly	RELU	1	ADAM	MSE	2000	0.030	27.449
	Daily	RELU	2	ADAM	MSE	200	0.319	83.030
	Hourly	RELU	1	ADAM	MSE	50	0.340	89.904
Bitta, Gujarat	Monthly	RELU	2	ADAM	MSE	2000	0.013	11.448
	Weekly	RELU	1	ADAM	MSE	2000	0.031	25.374
	Daily	RELU	1	ADAM	MSE	200	0.399	97.730
	Hourly	RELU	2	ADAM	MSE	50	0.316	84.017
Pavagada, Karnataka	Monthly	RELU	1	ADAM	MSE	2000	0.018	13.998
	Weekly	RELU	1	ADAM	MSE	2000	0.047	34.038
	Daily	RELU	2	ADAM	MSE	200	0.327	92.132
	Hourly	RELU	1	ADAM	MSE	50	0.397	94.893
Ramagundam, Telangana	Monthly	RELU	2	ADAM	MSE	2000	0.025	17.932
	Weekly	RELU	1	ADAM	MSE	2000	0.067	44.849
	Daily	RELU	1	ADAM	MSE	200	0.416	101.844
	Hourly	RELU	2	ADAM	MSE	50	0.403	97.800

4.2.3 Long Short Term Memory (LSTM)

The LSTM [74] is a special kind of RNN designed to learn long term dependencies and it is used widely in sequence models. It has applications in several kinds of sequence models, such as handwriting recognition, speech recognition, and text classification [146]. The LSTMs have an internal memory in the form of gates. This allows the LSTM to accumulate important information which may be required in future. The LSTM is also able to control the process of remembering information for noticeably longer periods of time. It has a special structure called a cell, consisting of four components, namely an input gate, a forget gate, an output gate, and a cell state. These components work together to regulate the flow of information into and out of the cell, and to preserve the relevant information over longer time period. The mathematical formulation of LSTM model is provided below.

Let x_t be the input vector at time step t, h_{t-1} be the hidden state vector at time step $t-1$, and c_{t-1} be the cell state vector at time step $t-1$. The input gate i_t decides how much of the new input x_t will be stored in the cell state. It is computed by a sigmoid (σ) function that takes

x_t and hidden state h_{t-1} as inputs.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4.15}$$

$$C_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \tag{4.16}$$

C_t is the candidate cell state; W_i, U_i, b_i, W_C, U_C, and b_C are learnable parameters, and σ and tanh are activation functions. The forget gate decides which parts of the previous cell state should be retained or discarded. It is computed as follows.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{4.17}$$

Here, f_t is the forget gate activation, and W_f, U_f, and b_f are learnable parameters. The output gate decides which parts of the current cell state should be output for the hidden state. It is computed as follows.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4.18}$$

$$h_t = o_t \odot \tanh(C_t) \tag{4.19}$$

Here, o_t is the output gate activation; h_t is the hidden state at time step t; C_t is the cell state at time step t; $\odot$ is element-wise multiplication, and W_o, U_o, and b_o are learnable parameters. The cell state is updated by combining the input gate, the forget gate, and the candidate cell state.

$$C_t = f_t \odot C_{t-1} + i_t \odot C_{t-1} \tag{4.20}$$

In literature, various variants of LSTMs are built that have different architecture and capabilities depending upon the types of dataset and purpose of forecasting [9, 102]. Following four types of LSTM models are implemented in the current study:

1. Regular LSTM
2. Bidirectional LSTM
3. Encoder-Decoder LSTM
4. Attention Layer LSTM

Brief details of these variants of LSTMs are provided in the following subsections.

4.2.3.1 Regular LSTM

Regular LSTM [113] is the simplest form of LSTM, where the hidden state is updated by a linear combination of the current input and the previous hidden state, followed by a non-linear

activation function. Regular LSTMs can suffer from the problem of vanishing or exploding gradients making them difficult to train for long sequences.

4.2.3.2 Bidirectional LSTM

As the name suggests, this type of LSTM [131] consists of two LSTMs: one taking the input in a forward direction, and the other in a backward direction. Therefore, in this model, the signal propagates both backward and forward. It can simply be considered as two LSTM layers working in opposite directions. The BiLSTMs effectively increase the amount of information available to the network, improving the context available to the algorithm.

4.2.3.3 Encoder-Decoder LSTM

The encoder-decoder LSTM [19] model consists of three parts, namely encoder, encoder vector, and decoder. The encoder reads the input sequence and encodes it into a fixed-length vector, whereas the decoder decodes the vector providing outputs in terms of the predicted sequence. The encoder vector is a hidden layer which acts as the final layer of encoder unit and the initial layer of decoder unit. Due to several layers of LSTM in its architecture, the encoder-decoder LSTM is a more complex model as compared to the regular LSTM. Unlike regular LSTMs, it can overcome the limitation of fixed-length internal representation.

4.2.3.4 Attention Layer LSTM

This model integrates an "attention layer" with the traditional LSTM layer [1]. The attention layer has been proved to be quite successful in improving the performance of the deep learning models. The idea behind it is to utilize only the important parts of the information by taking weighted combination of the encoded information. It was first created to enhance the performance of the encoder-decoder LSTM model, by carrying weighted information from encoder to decoder. This technique allows the network to learn where to pay attention in the input sequence for each item in the output sequence. Thus, the attention layer provides weights to the intermediate outputs from the encoder LSTM so that the decoder LSTM can focus on the relevant parts of the input sequence to generate better predictions.

To implement the above four LSTM models, we require to optimize the following parameters:

- Number of layers of LSTM
- Activation function
- Optimization algorithm

- Loss function
- Epochs
- Learning rate

Several variations of the above parameters are tested in order to obtain the best parameters of the studied models. The data are re-scaled to values in between 0 and 1 to normalize the input values. It not only helps to simplify the complex mathematical calculations in the model implementation but also helps to improve final results. It may be noted that in addition to the error metrics, the emphasis should also be given to the residual analysis as a part of model validation.

4.2.3.5 Results

The optimal parameters for the implemented variants of the LSTMs for wind speed and GHI data are summarized in Table 4.4 and Table 4.5, whereas the comparative performance in terms of RMSE and MAPE values are tabulated in Table 4.6 and Table 4.7. The pictorial representation of original values versus forecasted values of both wind speed and GHI data through the best fit LSTM model at each timescale are provided in Section 4.3. Regarding wind speed, we note that in monthly forecasting, the encoder-decoder LSTM has the best performance in comparison to other models. In weekly forecasting, the regular LSTM has best results across all the study sites. In daily forecasting, each of the regular LSTM and BiLSTM has the least error values in two study sites each. In hourly forecasting, the BiLSTM has the best performance in Rajasthan and Karnataka, whereas encoder-decoder LSTM has the best results in Gujarat and Telangana. It may be noted that the attention layer LSTM has the least performance against our expectation. In GHI forecasting, the regular LSTM has the best performance in monthly, weekly, and hourly data. The BiLSTM has the least RMSE value in Rajasthan and Karnataka, whereas the encoder-decoder LSTM and attention layer LSTM provide the best results in daily forecasting in Gujarat and Telangana, respectively.

4.2.4 Convolutional Neural Network (CNN)

The CNN is a deep learning model inspired by the human visual system. Since the inception of CNNs in 1998 by Yann et al. [64], they have become a fundamental architecture in computer vision applications. To improve the ability to model complex data, the CNN method consists of three mapping layers, namely the convolutional layer, pooling layer, and the fully-connected layer. The convolutional layer, core building block of CNNs, uses mathematical convolutional operations to learn spatial features from the input data. The pooling layer reduces the spatial

Table 4.4: Best parameters for different LSTMs for wind speed data

Model	Data	Activation Function	Layers	Optimizer	Loss Function	Epochs	Learning Rate
LSTM	Monthly	SIGMOID	1	RMSprop	MSE	2000	0.001
	Weekly	RELU	1	SGD	MSE	2000	0.001
	Daily	RELU	1	ADAM	MSE	200	0.010
	Hourly	RELU	1	ADAM	MSE	50	0.001
BiLSTM	Monthly	SIGMOID	1	RMSprop	MSE	2000	0.001
	Weekly	SIGMOID	1	ADAM	MSE	2000	0.001
	Daily	RELU	1	SGD	MSE	800	0.010
	Hourly	RELU	2	ADAM	MSE	50	0.001
Encoder-Decoder	Monthly	RELU	1	RMSprop	MSE	600	0.001
	Weekly	RELU	1	ADAM	MSE	1000	0.010
	Daily	RELU	1	SGD	MSE	800	0.010
	Hourly	RELU	1	ADAM	MSE	50	0.001
Attention Layer	Monthly	RELU	1	RMSprop	MSE	600	0.001
	Weekly	RELU	2	RMSprop	MAE	1600	0.010
	Daily	SIGMOID	1	ADAM	MSE	500	0.010
	Hourly	SIGMOID	1	ADAM	MSE	50	0.010

Table 4.5: Best parameters for different LSTMs for GHI data

Model	Data	Activation Function	Layers	Optimizer	Loss Function	Epochs	Learning Rate
LSTM	Monthly	RELU	1	RMSprop	MSE	600	0.001
	Weekly	RELU	1	RMSprop	MSE	1500	0.010
	Daily	RELU	2	RMSprop	MAE	1500	0.010
	Hourly	RELU	1	ADAM	MSE	50	0.001
BiLSTM	Monthly	SIGMOID	2	ADAM	MAE	600	0.001
	Weekly	SIGMOID	1	ADAM	MSE	2000	0.001
	Daily	RELU	1	SGD	MSE	800	0.010
	Hourly	RELU	2	ADAM	MSE	50	0.001
Encoder-Decoder	Monthly	RELU	1	RMSprop	MSE	600	0.001
	Weekly	RELU	1	ADAM	MSE	1000	0.010
	Daily	RELU	1	SGD	MSE	800	0.010
	Hourly	RELU	1	ADAM	MSE	50	0.001
Attention Layer	Monthly	RELU	1	RMSprop	MSE	600	0.001
	Weekly	RELU	1	ADAM	MSE	2000	0.001
	Daily	RELU	1	ADAM	MSE	200	0.010
	Hourly	RELU	1	ADAM	MSE	50	0.001

Table 4.6: Results of different LSTMs for wind speed dataset at four different study sites (L1: Pokhran, Rajasthan, L2: Bitta, Gujarat, L3: Pavagada, Karnataka, and L4: Ramagundam, Telangana)

Model	**Data**	**L1**		**L2**		**L3**		**L4**	
		RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
LSTM	Monthly	0.549	0.122	0.488	0.119	0.697	0.167	0.429	0.146
	Weekly	**0.485**	0.125	**0.390**	0.091	**0.424**	0.114	**0.370**	0.124
	Daily	**0.720**	0.241	**0.662**	0.175	0.534	0.163	0.570	0.217
	Hourly	0.164	0.075	0.174	0.087	0.158	0.082	0.174	0.086
BiLSTM	Monthly	0.552	0.134	0.750	0.172	0.603	0.134	**0.363**	0.126
	Weekly	0.669	0.171	0.592	0.131	0.697	0.166	0.528	0.192
	Daily	0.783	0.219	0.699	0.157	**0.524**	0.156	**0.543**	0.190
	Hourly	**0.154**	0.057	0.153	0.049	**0.150**	0.082	0.154	0.072
Encoder-Decoder	Monthly	**0.383**	0.086	**0.362**	0.085	**0.428**	0.110	0.369	0.111
	Weekly	0.957	0.242	0.802	0.187	0.970	0.226	0.727	0.257
	Daily	0.832	0.257	0.671	0.184	0.557	0.165	0.592	0.233
	Hourly	0.155	0.055	**0.149**	0.052	0.152	0.048	**0.153**	0.064
Attention Layer	Monthly	1.516	0.380	0.441	0.112	1.973	0.436	1.549	0.488
	Weekly	0.831	0.236	0.732	0.169	0.893	o.213	1.809	0.574
	Daily	1.141	0.360	1.028	0.295	1.082	0.295	0.940	0.362
	Hourly	0.177	0.063	0.161	0.052	0.153	0.041	0.160	0.054

Table 4.7: Results of different LSTMs for GHI dataset at four different study sites (L1: Pokhran, Rajasthan, L2: Bitta, Gujarat, L3: Pavagada, Karnataka, and L4: Ramagundam, Telangana)

Model	**Data**	L1		L2		L3		L4	
		RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
LSTM	Monthly	**17.848**	0.023	27.892	0.039	**22.097**	0.031	**32.679**	0.050
	Weekly	**31.728**	0.063	**21.992**	0.068	**30.438**	0.098	**34.594**	0.125
	Daily	68.804	0.113	74.006	0.159	74.043	0.103	88.996	0.123
	Hourly	**65.731**	0.187	**64.407**	0.144	**84.161**	0.185	**82.772**	0.147
BiLSTM	Monthly	18.353	0.026	29.498	0.039	21.492	0.028	35.492	0.059
	Weekly	40.913	0.047	44.761	0.058	52.521	0.074	54.602	0.083
	Daily	**59.451**	0.055	64.720	0.068	**69.806**	0.087	85.456	0.118
	Hourly	67.974	0.162	67.901	0.168	89.582	0.187	88.167	0.190
Encoder-Decoder	Monthly	19.319	0.024	25.884	0.032	25.594	0.026	35.449	0.051
	Weekly	57.162	0.066	63.107	0.071	66.412	0.088	76.699	0.106
	Daily	61.237	0.070	**64.263**	0.077	74.143	0.099	96.029	0.155
	Hourly	68.773	0.184	67.782	0.144	91.693	0.183	89.052	0.174
Attention Layer	Monthly	33.109	0.128	**25.612**	0.051	38.532	0.098	36.795	0.076
	Weekly	48.590	0.059	52.177	0.153	47.752	0.139	83.107	0.149
	Daily	79.095	0.101	79.052	0.113	82.962	0.186	**83.987**	0.179
	Hourly	91.141	0.214	104.065	0.286	86.993	0.290	98.005	0.276

dimensions of the feature maps and aids in translational invariance. Finally, a fully-connected layer at the conclusion of the CNN predicts the output based on retrieved features [42]. The CNNs are well-known as a reliable tool for extracting hidden features and creating filters based on data patterns. They include two primary features, namely weight sharing and local connections [170]. Each convolutional layer is designed to extract spatial patterns from the target variable (e.g., wind speed or GHI) and its related input variables (in this case, previous GHI or wind speed values). For the CNN models, the training dataset is converted to an input matrix and an output vector. For a monthly model, the input matrix contains wind speed or GHI values for all days in that month. Thus, the input dimension is 28, 30, or 31. The input dimension is 7 for a weekly model, whereas it is 24 (number of hours in a day) and 9 (number of sunny hours in a day) for daily and hourly models for wind speed and GHI data, respectively. As the final step of the preprocessing, the data are normalized so that all values lie in the interval $[0, 1]$. The 'max normalization' method is applied on the dataset for scaling the variables in this range before preparing the input matrix. The scaling of data is important because most of the activation functions for neural networks do not perform well with large positive values. Hence, due to the error back-propagation, the training is less effective when the input value has a larger range. This is because of the fact that the back-propagated error is multiplied with the derivative of the activation function. A neural network that uses data without normalization should eventually converge as well, but the training process may take longer duration.

4.2.4.1 Results

The optimal parameters of the CNN model and corresponding error values in wind speed and GHI forecasting are tabulated in Table 4.8 and Table 4.9, respectively. We observe that the RMSE values generally increase as we move from longer to shorter timescale, i.e., from monthly to weekly, to daily, and to hourly data across all locations. A pictorial representation of original values versus forecasted values of both wind speed and GHI data from Pokhran, Rajasthan is provided in Section 4.3. We have also performed a comparison of the obtained results with the other models in Section 4.3.

4.3 Comparison of Results

In this section, we compare the previously obtained error values from seven machine learning models at four different timescales and four selected locations. Table 4.10 and Table 4.11 present the error values of the studied machine learning techniques in wind speed and GHI forecasting, respectively. The minimum RMSE values for each timescale and location are highlighted in bold. For visualization of model fit, we have plotted actual versus forecasted

Table 4.8: The optimal values of parameters and corresponding error values of CNN model for wind speed data

Location	Timescale	Activation Function	Optimizer	Loss Function	Epochs	RMSE	MAPE
Pokhran, Rajasthan	Monthly	RELU	ADAM	MSE	2000	0.122	0.032
	Weekly	RELU	ADAM	MSE	200	0.442	0.117
	Daily	SIGMOID	ADAM	MSE	100	0.775	0.225
	Hourly	RELU	ADAM	MSE	50	0.701	0.240
Bitta, Gujarat	Monthly	SIGMOID	ADAM	MAE	2000	0.114	0.029
	Weekly	SIGMOID	ADAM	MSE	200	0.348	0.083
	Daily	RELU	SGD	MSE	100	0.629	0.151
	Hourly	RELU	ADAM	MSE	50	0.653	0.201
Pavagada, Karnataka	Monthly	RELU	ADAM	MSE	2000	0.100	0.026
	Weekly	RELU	ADAM	MSE	200	0.413	0.107
	Daily	RELU	ADAM	MAE	100	0.533	0.144
	Hourly	RELU	ADAM	MSE	50	0.700	0.281
Ramagundam, Telangana	Monthly	SIGMOID	ADAM	MAE	2000	0.107	0.039
	Weekly	SIGMOID	ADAM	MSE	200	0.345	0.126
	Daily	RELU	ADAM	MSE	100	0.548	0.192
	Hourly	RELU	ADAM	MSE	50	0.494	0.192

Table 4.9: The optimal values of parameters and corresponding error values of CNN model for GHI data

Location	Timescale	Activation Function	Optimizer	Loss Function	Epochs	MAPE	RMSE
Pokhran, Rajasthan	Monthly	RELU	ADAM	MSE	2000	0.017	18.488
	Weekly	RELU	ADAM	MSE	2000	0.030	27.560
	Daily	RELU	ADAM	MAE	200	0.281	73.054
	Hourly	RELU	ADAM	MSE	50	0.157	76.034
Bitta, Gujarat	Monthly	SIGMOID	ADAM	MAE	2000	0.191	15.720
	Weekly	SIGMOID	ADAM	MSE	2000	0.040	19.232
	Daily	RELU	ADAM	MSE	200	0.300	77.406
	Hourly	RELU	ADAM	MSE	50	0.185	81.911
Pavagada, Karnataka	Monthly	RELU	ADAM	MSE	2000	0.015	10.849
	Weekly	RELU	ADAM	MSE	2000	0.047	36.762
	Daily	RELU	ADAM	MAE	200	0.315	75.904
	Hourly	RELU	ADAM	MSE	50	0.184	84.621
Ramagundam, Telangana	Monthly	SIGMOID	ADAM	MAE	2000	0.022	19.113
	Weekly	SIGMOID	ADAM	MSE	2000	0.066	45.374
	Daily	RELU	ADAM	MSE	200	0.366	91.404
	Hourly	RELU	ADAM	MSE	50	0.200	92.015

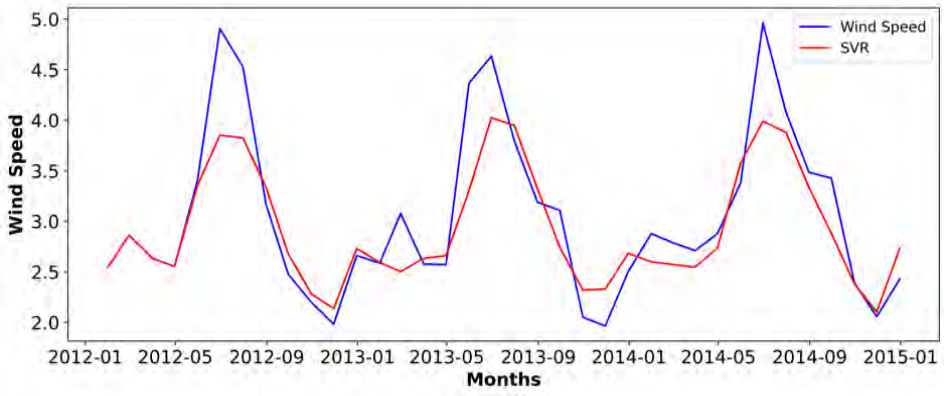
values of the SVR, ANN, CNN, and the best fit LSTM model at different timescales in Pokhran, Rajasthan in Figure 4.5– 4.10. Figure 4.11–4.15 represent the corresponding residual plots of the best fit models at different timescales. The residual plots exhibit that the residual values are generally scattered around zero, corresponding histograms are bell shaped, and the P-P plots show close association with the normal distribution. These behaviors indicate unbiasedness of the best fit models. It may be noted that we have not included the plots for hourly data due to huge cluster of data as observed in Figure 4.2 and Figure 4.3. From Table 4.10 regarding wind speed forecasting, we observe that each of the CNN model and the ANN model has the least error values in monthly forecasting for two locations out of four sites. The ANN model has the least RMSE values for three out of four study sites in weekly wind speed forecasting. In daily wind speed, the SVR, CNN, ANN, and the BiLSTM have the best representation in Rajasthan, Gujarat, Karnataka, and Telangana, respectively. In hourly wind speed forecasting, each of biLSTM and encoder-decoder LSTM reveals the best performance across two study sites. In GHI, Table 4.11 shows that the ANN model has the least error values in monthly forecasting across all four study sites. In weekly GHI forecasting, the best model turns out to be the ANN in Rajasthan, CNN in Gujarat, and the LSTM in both Karnataka and Telangana. In daily GHI forecasting, the SVR and the BiLSTM models are the best fit models in two locations each. On the other hand, the SVR is deemed to be the best model across all study sites in hourly GHI forecasting.

Table 4.10: Results obtained from different machine learning models on wind speed data at four study sites (L1: Pokhran, Rajasthan, L2: Bitta, Gujarat, L3: Pavagada, Karnataka, L4: Ramagundam, Telangana)

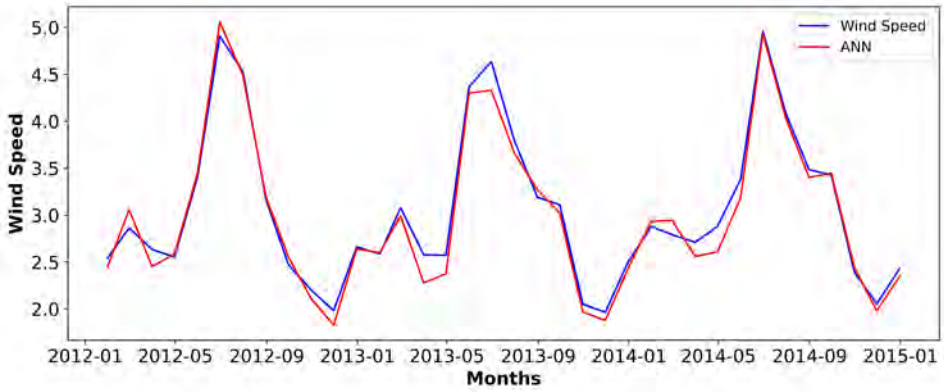
Model	Data	L1		L2		L3		L4	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
SVR	Monthly	0.417	0.087	0.328	0.087	0.422	0.092	0.359	0.104
	Weekly	0.663	0.174	0.571	0.132	0.709	0.170	0.542	0.174
	Daily	**0.709**	0.202	0.642	0.177	0.526	0.151	0.578	0.209
	Hourly	0.174	0.063	0.186	0.059	0.186	0.057	0.209	0.086
ANN	Monthly	**0.114**	0.032	**0.104**	0.028	0.102	0.026	0.118	0.039
	Weekly	0.449	0.121	**0.338**	0.080	**0.405**	0.109	**0.342**	0.119
	Daily	0.743	0.247	0.642	0.176	**0.523**	0.155	0.583	0.219
	Hourly	1.002	0.391	0.930	0.328	0.917	0.287	1.031	0.369
CNN	Monthly	0.122	0.032	0.114	0.029	**0.100**	0.026	**0.107**	0.039
	Weekly	**0.442**	0.117	0.348	0.083	0.413	0.107	0345	0.126
	Daily	0.775	0.225	**0.629**	0.151	0.533	0.144	0.548	0.192
	Hourly	0.701	0.240	0.653	0.201	0.700	0.281	0.494	0.192
LSTM	Monthly	0.549	0.122	0.488	0.119	0.697	0.167	0.429	0.146
	Weekly	0.485	0.125	0.390	0.091	0.424	0.114	0.370	0.124
	Daily	0.720	0.241	0.662	0.175	0.534	0.163	0.570	0.217
	Hourly	0.164	0.075	0.174	0.087	0.158	0.082	0.174	0.086
BiLSTM	Monthly	0.552	0.134	0.750	0.172	0.603	0.134	0.363	0.126
	Weekly	0.669	0.171	0.592	0.131	0.697	0.166	0.528	0.192
	Daily	0.783	0.219	0.699	0.157	0.524	0.156	**0.543**	0.190
	Hourly	**0.154**	0.057	0.153	0.049	**0.150**	0.082	0.154	0.072
Encoder-Decoder	Monthly	0.383	0.086	0.362	0.085	0.428	0.110	0.369	0.111
	Weekly	0.957	0.242	0.802	0.187	0.970	0.226	0.727	0.257
	Daily	0.832	0.257	0.671	0.184	0.557	0.165	0.592	0.233
	Hourly	0.155	0.055	**0.149**	0.052	0.152	0.048	**0.153**	0.064
Attention Layer	Monthly	1.516	0.380	0.441	0.112	1.973	0.436	1.549	0.488
	Weekly	0.831	0.236	0.732	0.169	0.893	o.213	1.809	0.574
	Daily	1.141	0.360	1.028	0.295	1.082	0.295	0.940	0.362
	Hourly	0.177	0.063	0.161	0.052	0.153	0.041	0.160	0.054

Table 4.11: Results obtained from different machine learning models on GHI data at four study sites (L1: Pokhran, Rajasthan, L2: Bitta, Gujarat, L3: Pavagada, Karnataka, L4: Ramagundam, Telangana)

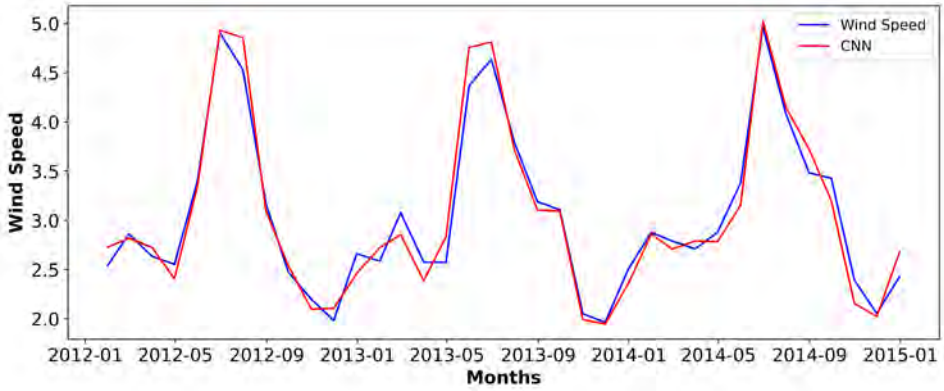
Model	Data	L1		L2		L3		L4	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
SVR	Monthly	13.193	0.020	24.945	0.035	18.756	0.028	30.638	0.048
	Weekly	38.353	0.050	42.852	0.053	55.551	0.076	61.179	0.092
	Daily	**57.515**	0.063	**59.343**	0.075	71.482	0.095	93.687	0.151
	Hourly	**63.591**	0.178	**61.382**	0.169	**83.040**	0.289	**81.592**	0.323
ANN	Monthly	**9.444**	0.010	**9.876**	0.012	**7.608**	0.012	**14.648**	0.022
	Weekly	**27.449**	0.030	25.374	0.031	34.038	0.047	44.849	0.067
	Daily	83.030	0.319	97.730	0.399	92.132	0.327	101.844	0.416
	Hourly	89.904	0.340	84.017	0.316	94.893	0.397	97.800	0.403
CNN	Monthly	18.488	0.017	15.720	0.191	10.849	0.015	19.113	0.022
	Weekly	27.560	0.030	**19.232**	0.040	36.762	0.047	45.374	0.066
	Daily	73.054	0.281	77.406	0.300	75.904	0.315	91.404	0.386
	Hourly	76.034	0.157	81.911	0.185	84.621	0.1849	92.015	0.200
LSTM	Monthly	16.838	0.022	27.366	0.038	21.073	0.029	35.058	0.055
	Weekly	31.728	0.063	21.992	0.068	**30.438**	0.098	**34.594**	0.125
	Daily	68.804	0.113	74.006	0.159	74.043	0.103	88.996	0.123
	Hourly	65.731	0.187	64.407	0.144	84.161	0.185	82.772	0.147
BiLSTM	Monthly	18.353	0.026	29.498	0.039	21.492	0.028	35.492	0.059
	Weekly	40.913	0.047	44.761	0.058	52.521	0.074	54.602	0.083
	Daily	59.451	0.055	64.720	0.068	**69.806**	0.087	**85.456**	0.118
	Hourly	67.974	0.162	67.901	0.168	89.582	0.187	88.167	0.190
Encoder-Decoder	Monthly	19.319	0.024	25.884	0.032	25.594	0.026	35.449	0.051
	Weekly	57.162	0.066	63.107	0.071	66.412	0.088	76.699	0.106
	Daily	61.237	0.070	64.263	0.077	74.143	0.099	96.029	0.155
	Hourly	68.773	0.184	67.782	0.144	91.693	0.183	89.052	0.174
Attention Layer	Monthly	33.109	0.128	25.612	0.051	38.532	0.098	36.795	0.076
	Weekly	48.590	0.059	52.177	0.153	47.752	0.139	83.107	0.149
	Daily	79.095	0.101	79.052	0.113	82.962	0.186	83.987	0.179
	Hourly	91.141	0.214	104.065	0.286	86.993	0.290	98.005	0.276



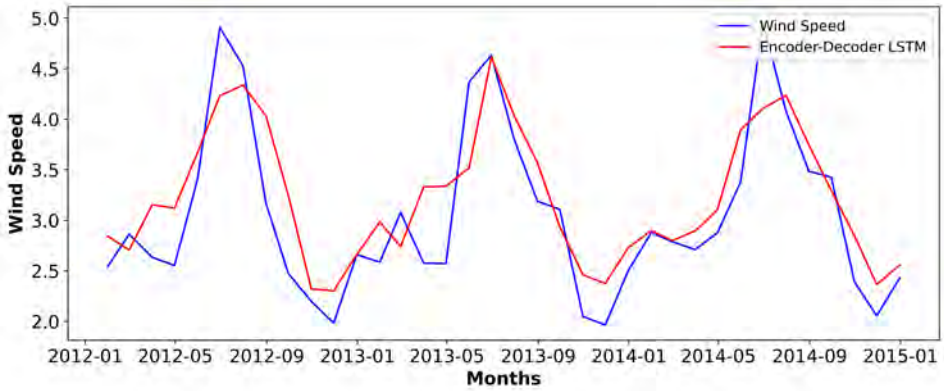
(a)



(b)

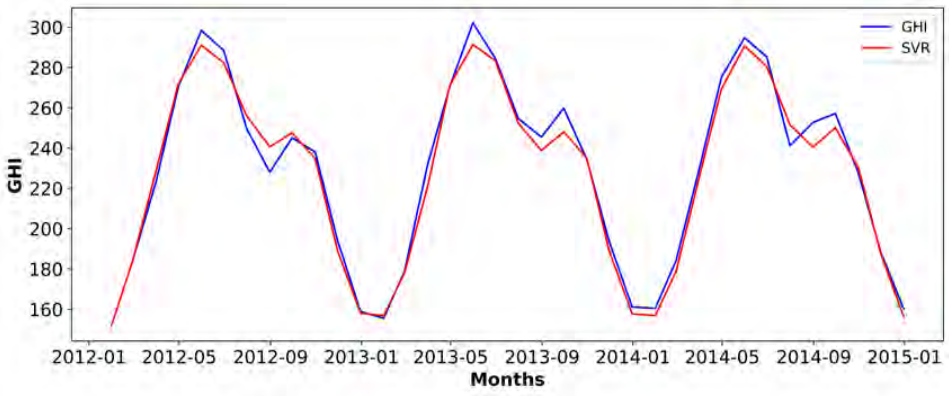


(c)

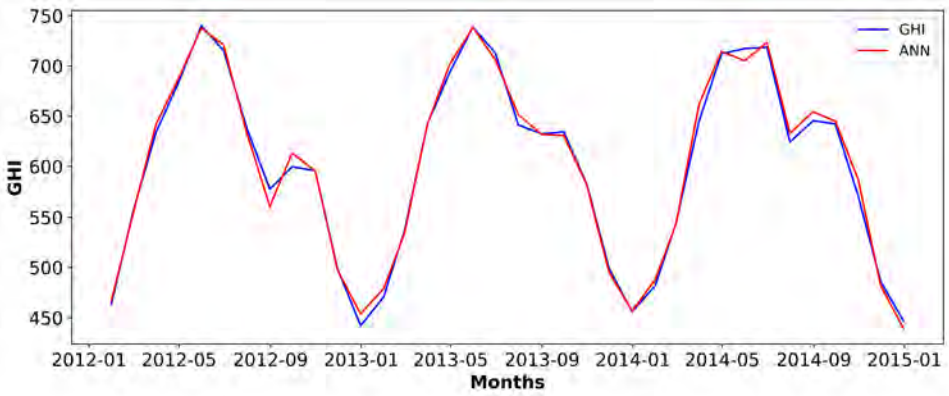


(d)

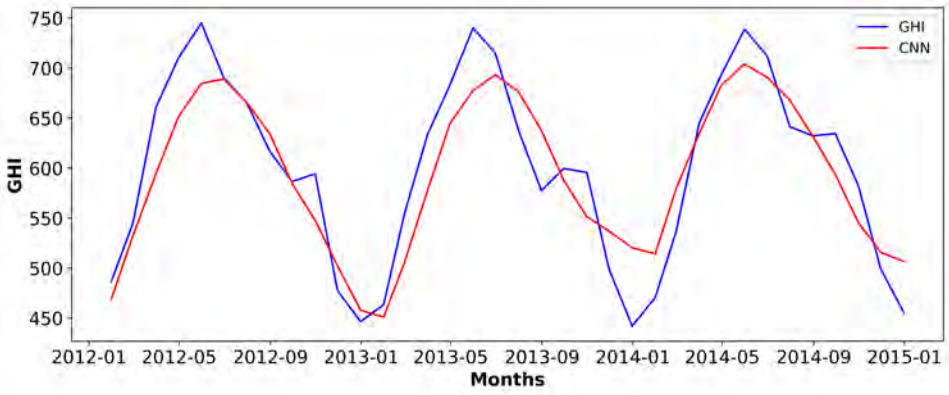
Fig. 4.5: Actual versus predicted wind speed from the (a) SVR model, (b) ANN model (c) CNN model, and (d) encoder-decoder LSTM model for monthly data at Pokhran, Rajasthan.



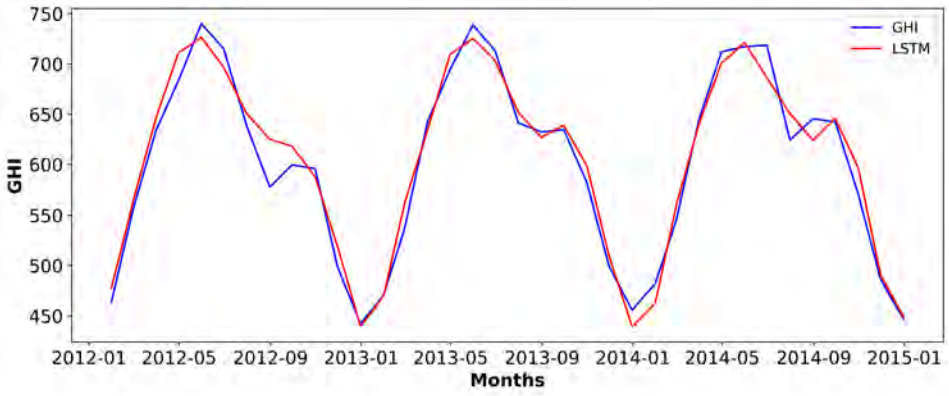
(a)



(b)

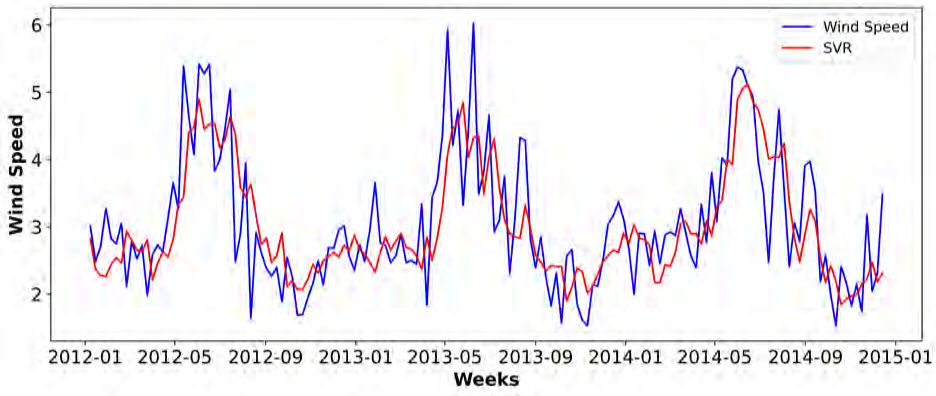


(c)

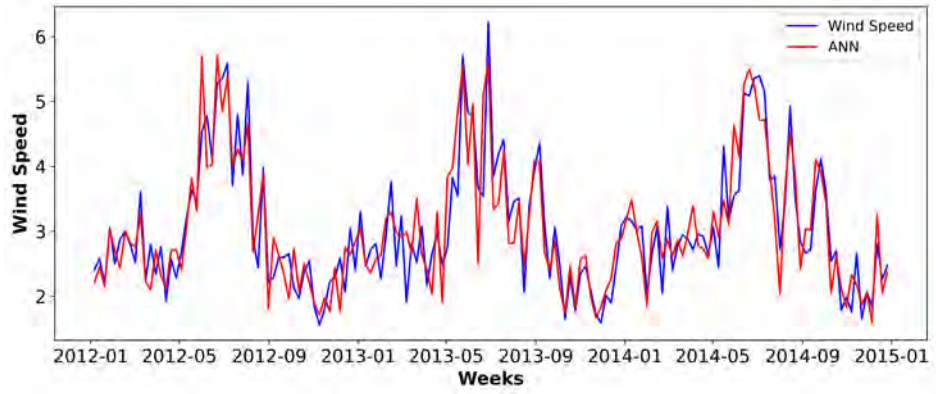


(d)

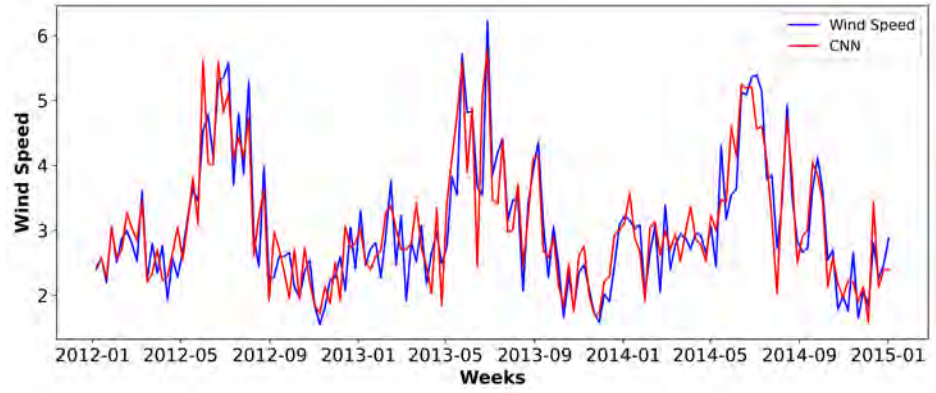
Fig. 4.6: Actual versus predicted GHI from the (a) SVR model, (b) ANN model (c) CNN model, and (d) LSTM model for monthly data at Pokhran, Rajasthan.



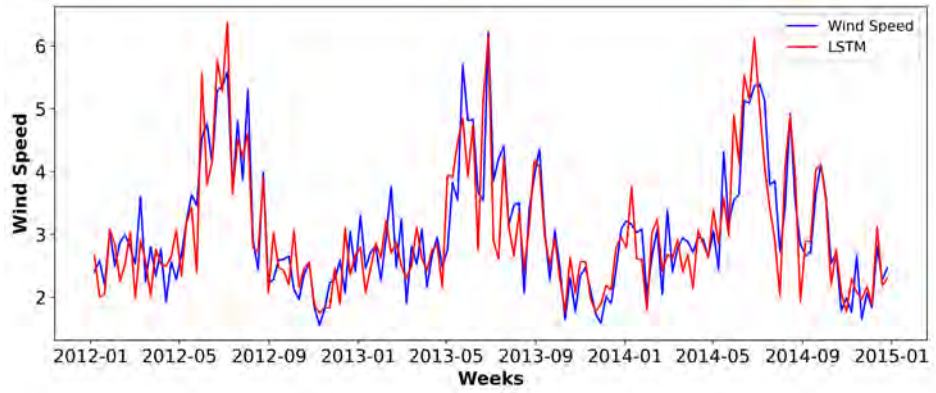
(a)



(b)

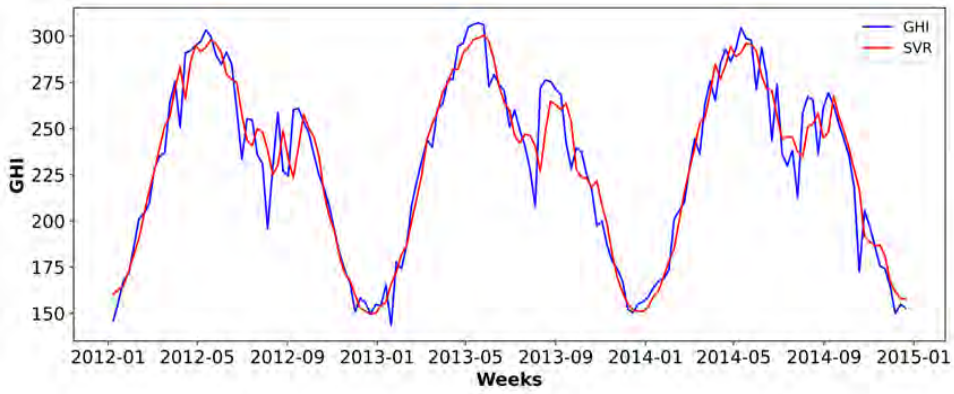


(c)

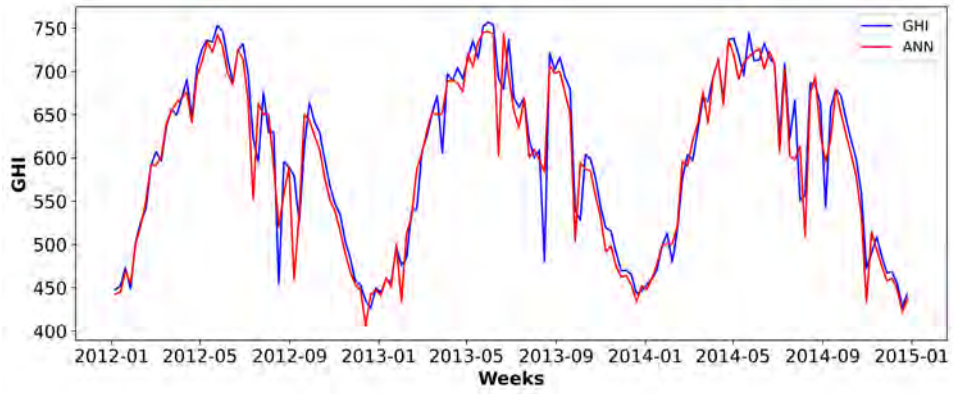


(d)

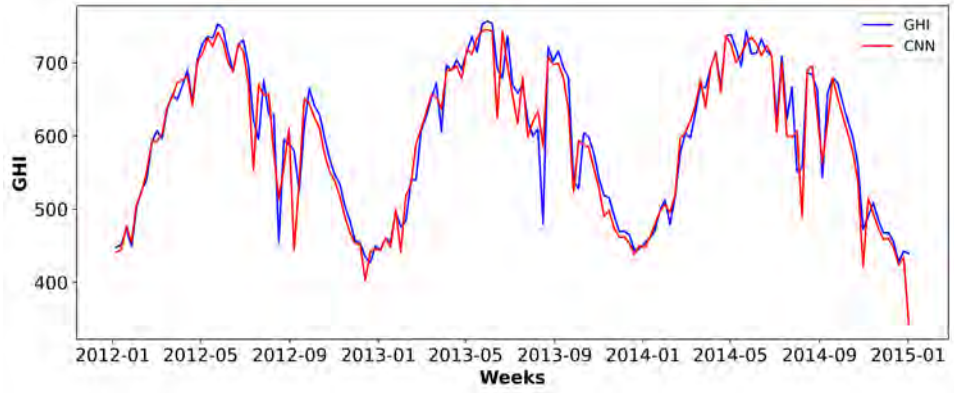
Fig. 4.7: Actual versus predicted wind speed from the (a) SVR model, (b) ANN model (c) CNN model, and (d) encoder-decoder LSTM model for weekly data at Pokhran, Rajasthan.



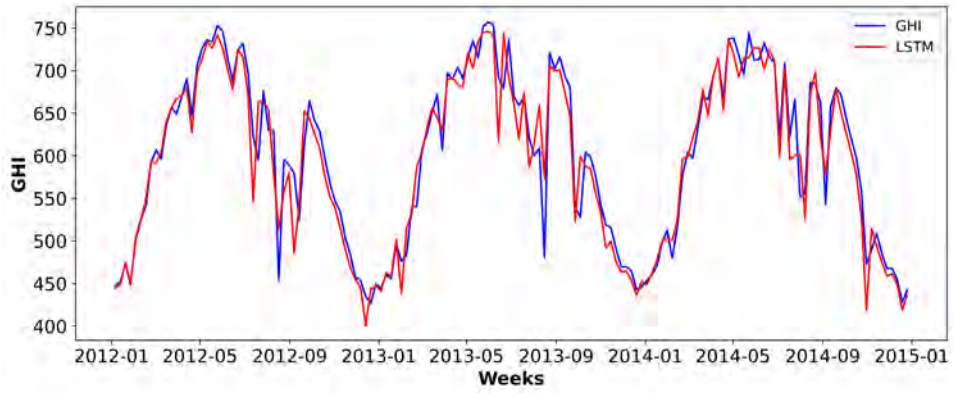
(a)



(b)

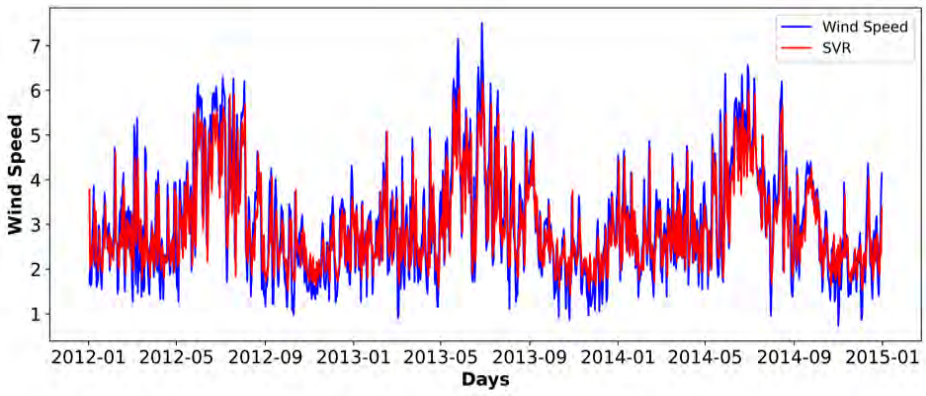


(c)

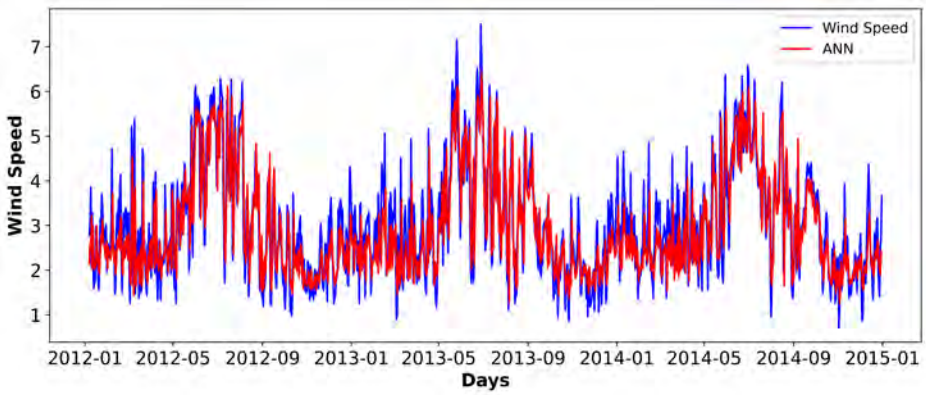


(d)

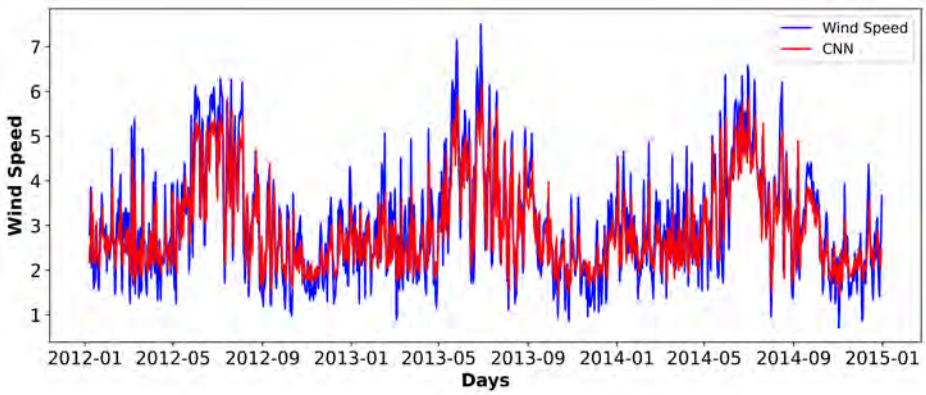
Fig. 4.8: Actual versus predicted GHI from the (a) SVR model, (b) ANN model (c) CNN model, and (d) LSTM model for weekly data at Pokhran, Rajasthan.



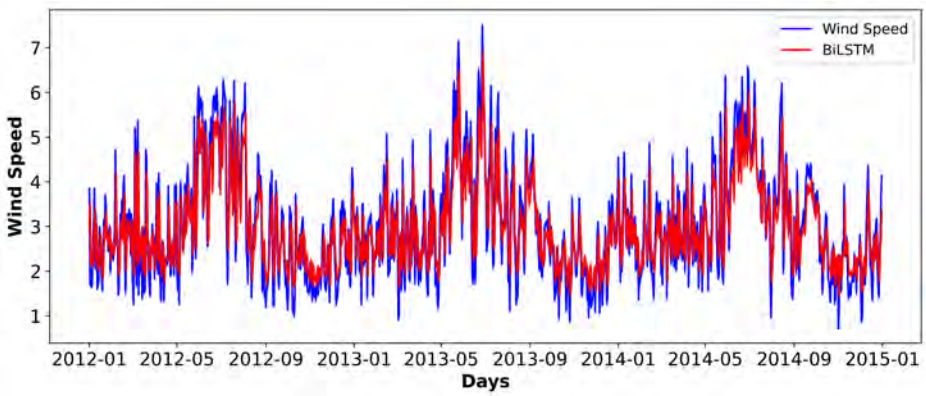
(a)



(b)

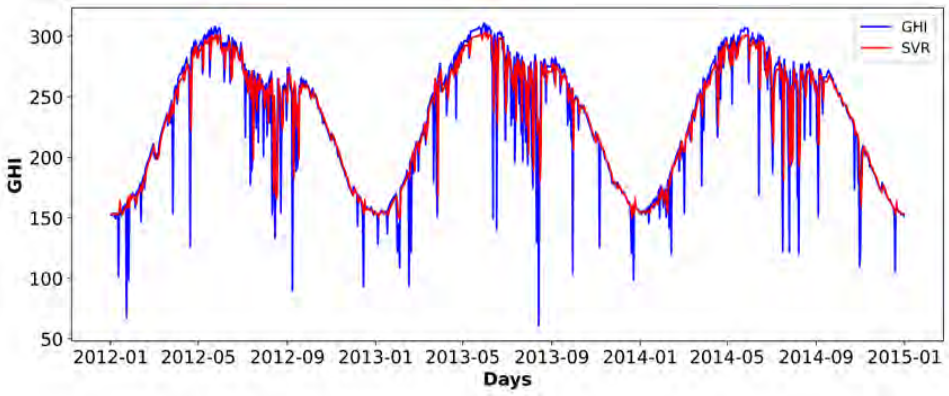


(c)

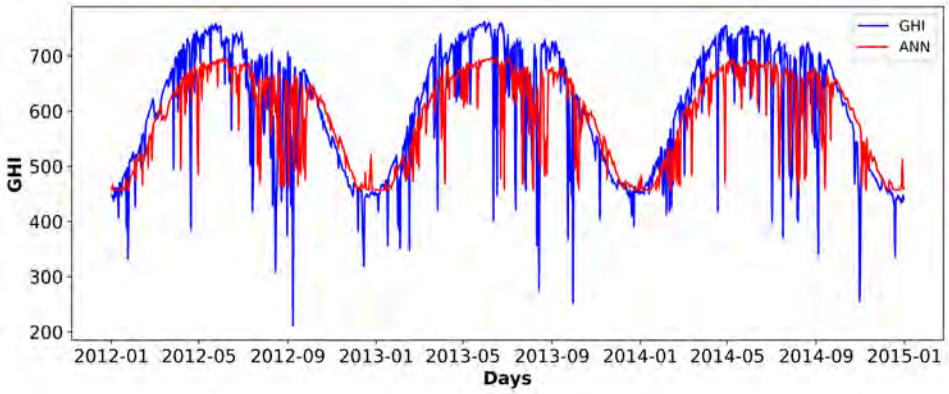


(d)

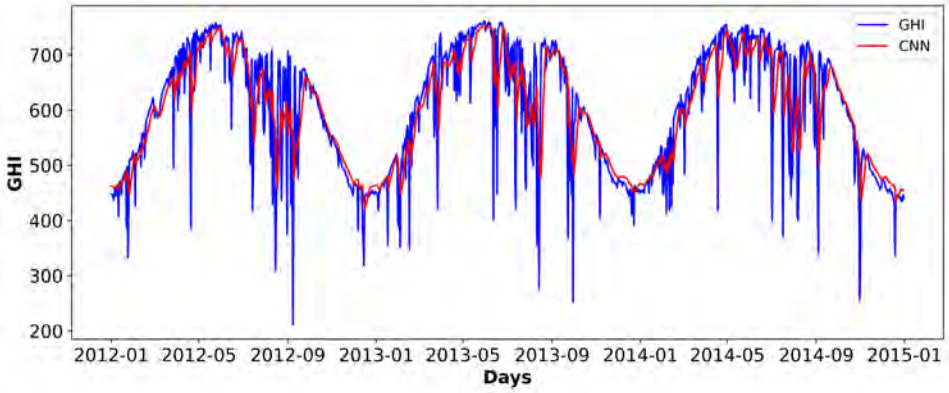
Fig. 4.9: Actual versus predicted wind speed from the (a) SVR model, (b) ANN model (c) CNN model, and (d) BiLSTM model at daily data from Pokhran, Rajasthan.



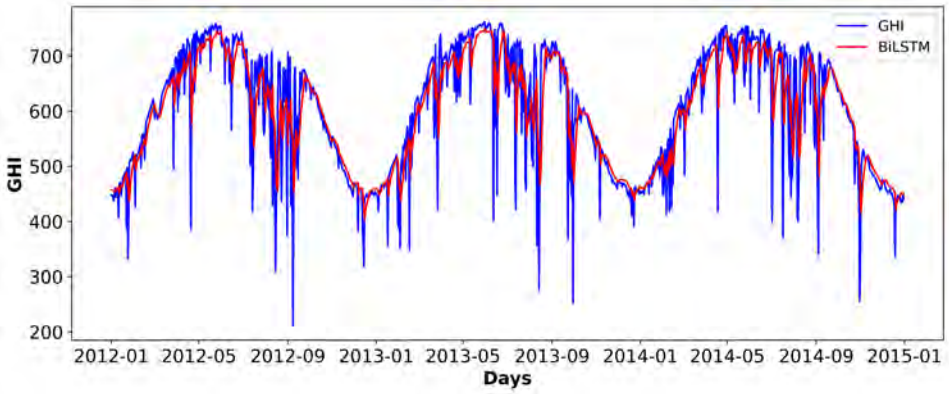
(a)



(b)

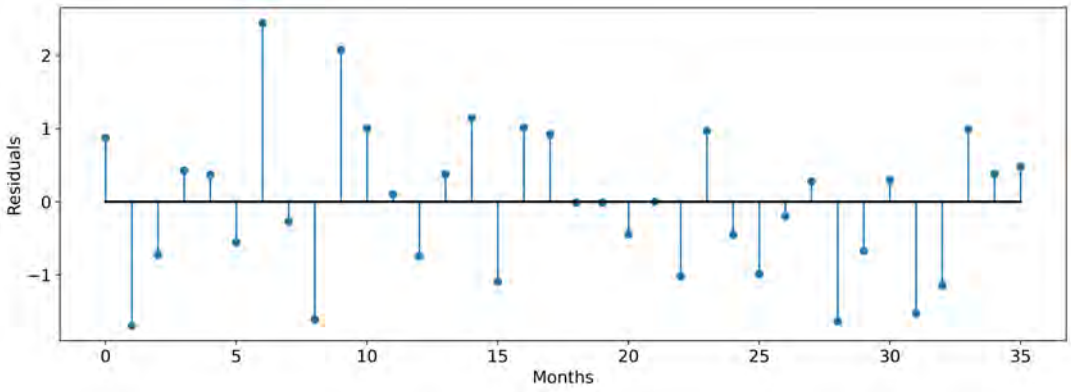


(c)

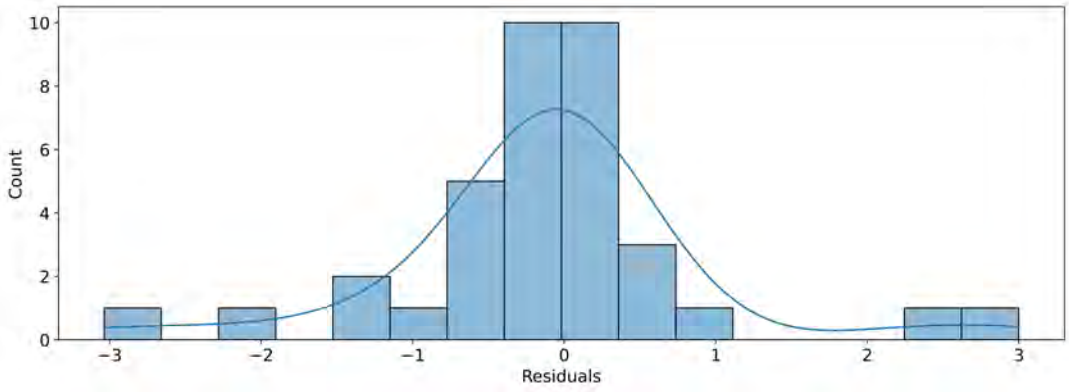


(d)

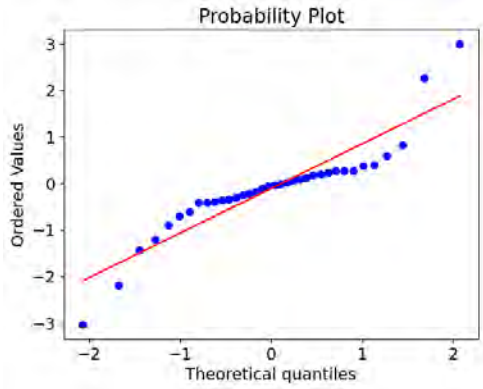
Fig. 4.10: Actual versus predicted GHI from the (a) SVR model, (b) ANN model (c) CNN model, and (d) BiLSTM model for daily data at Pokhran, Rajasthan.



(a)

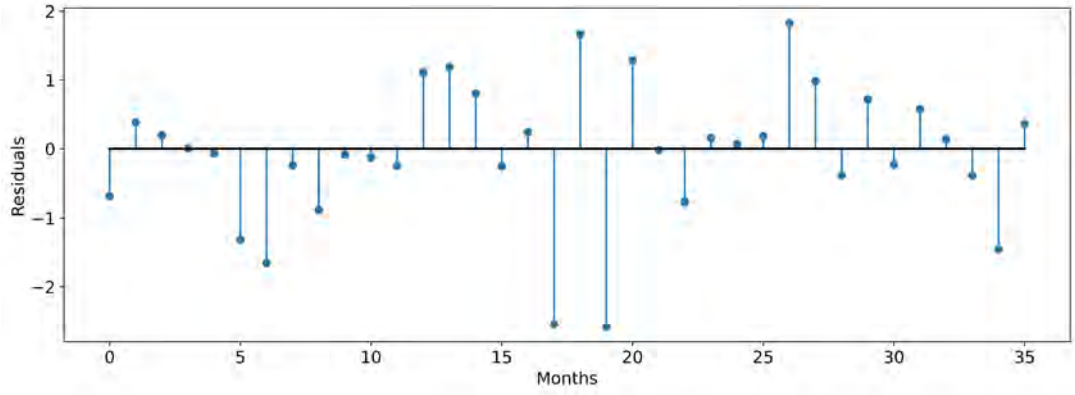


(b)

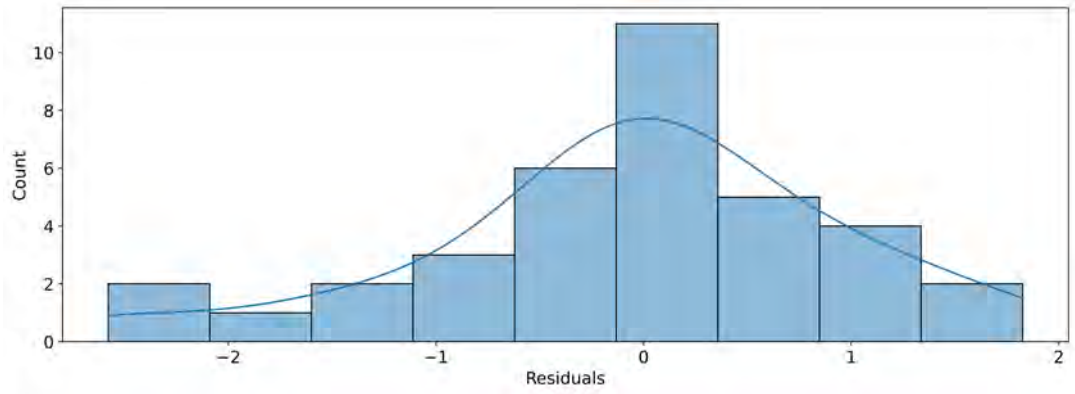


(c)

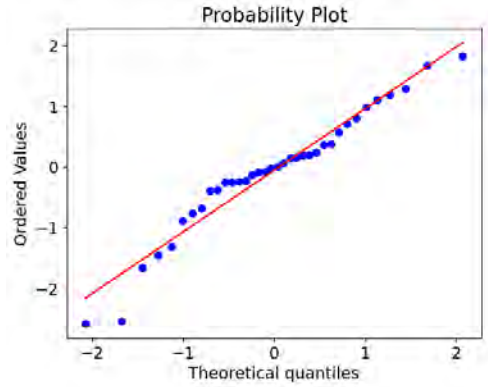
Fig. 4.11: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the ANN model in monthly wind speed forecasting at Pokhran, Rajasthan.



(a)

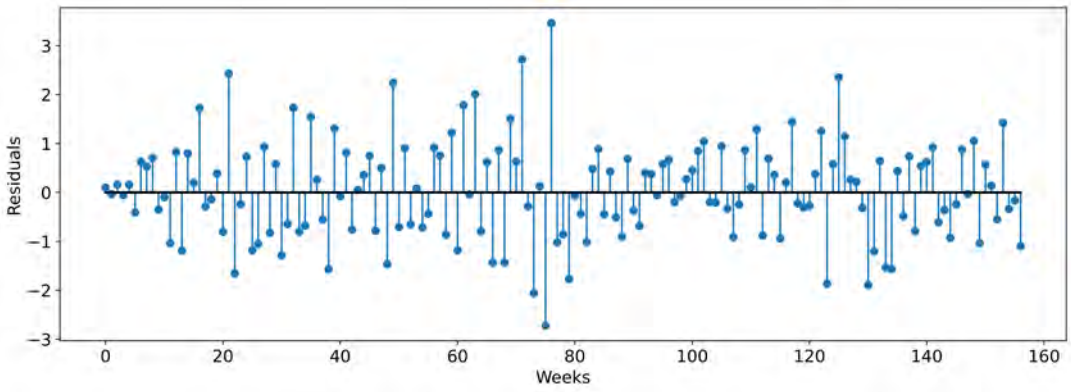


(b)

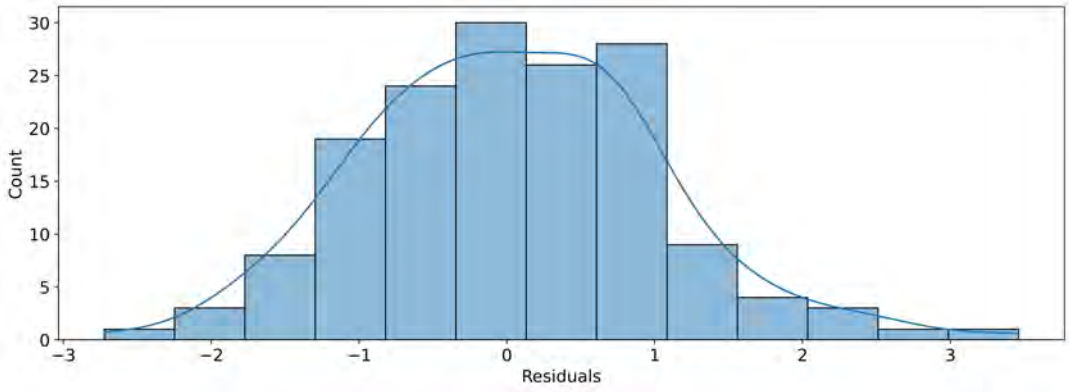


(c)

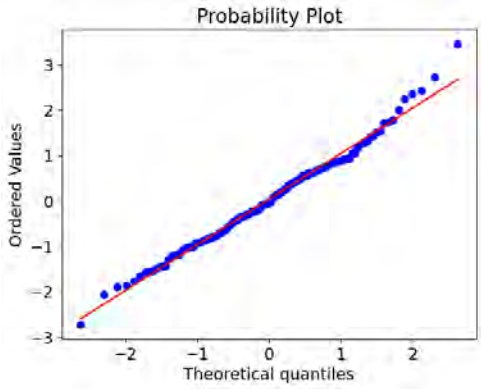
Fig. 4.12: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the ANN model in monthly GHI forecasting at Pokhran, Rajasthan.



(a)

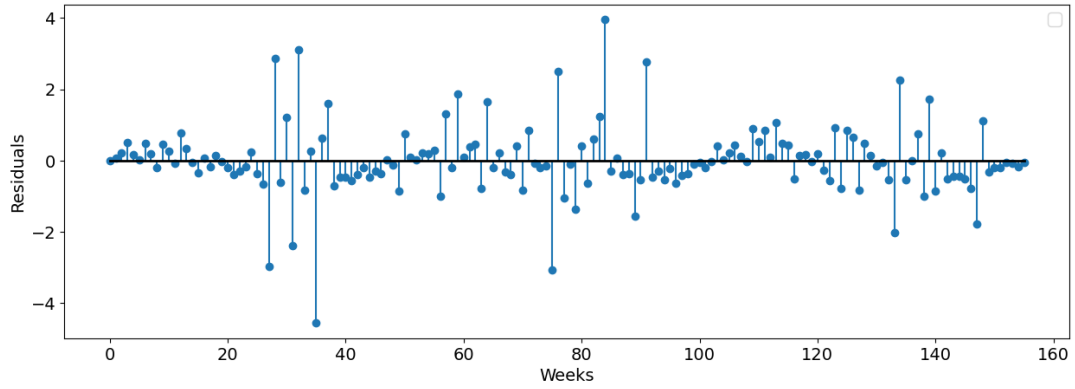


(b)

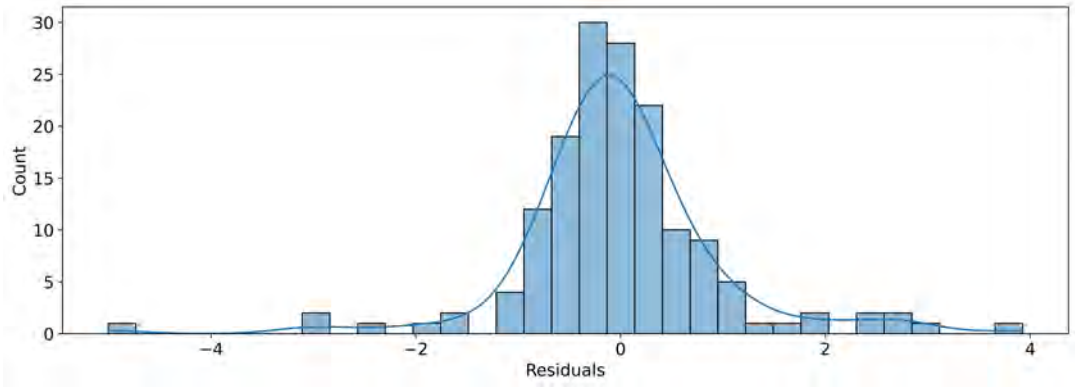


(c)

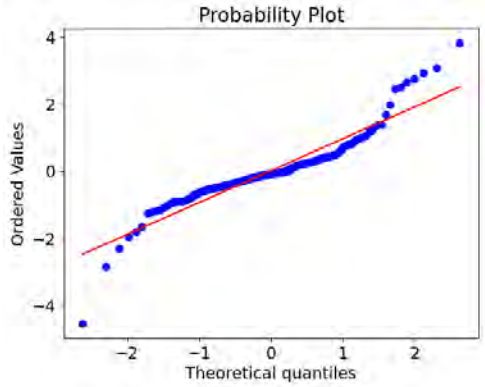
Fig. 4.13: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the ANN model in weekly wind speed forecasting at Pokhran, Rajasthan.



(a)

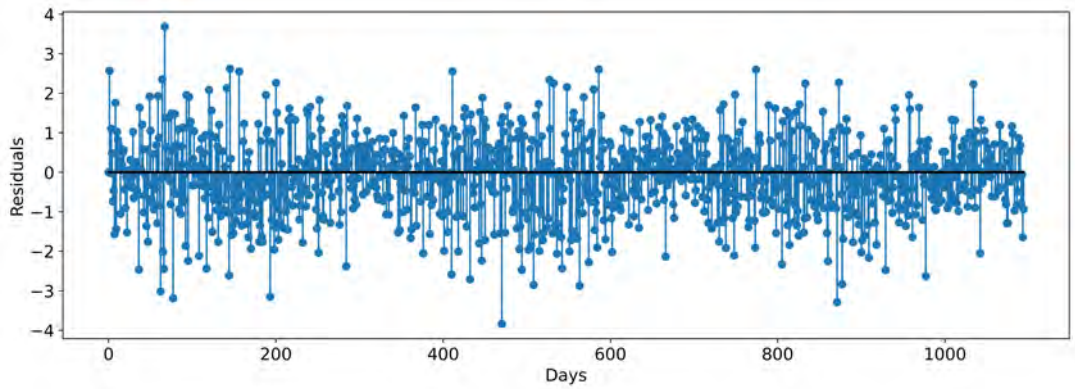


(b)

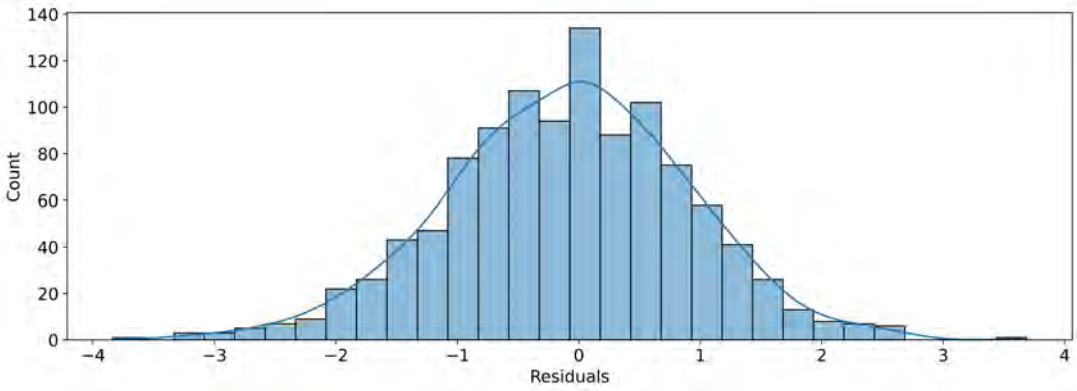


(c)

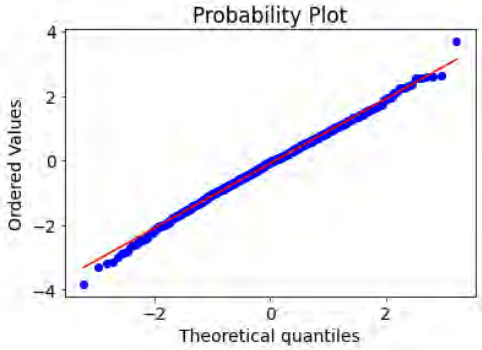
Fig. 4.14: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the ANN model in weekly GHI forecasting at Pokhran, Rajasthan.



(a)

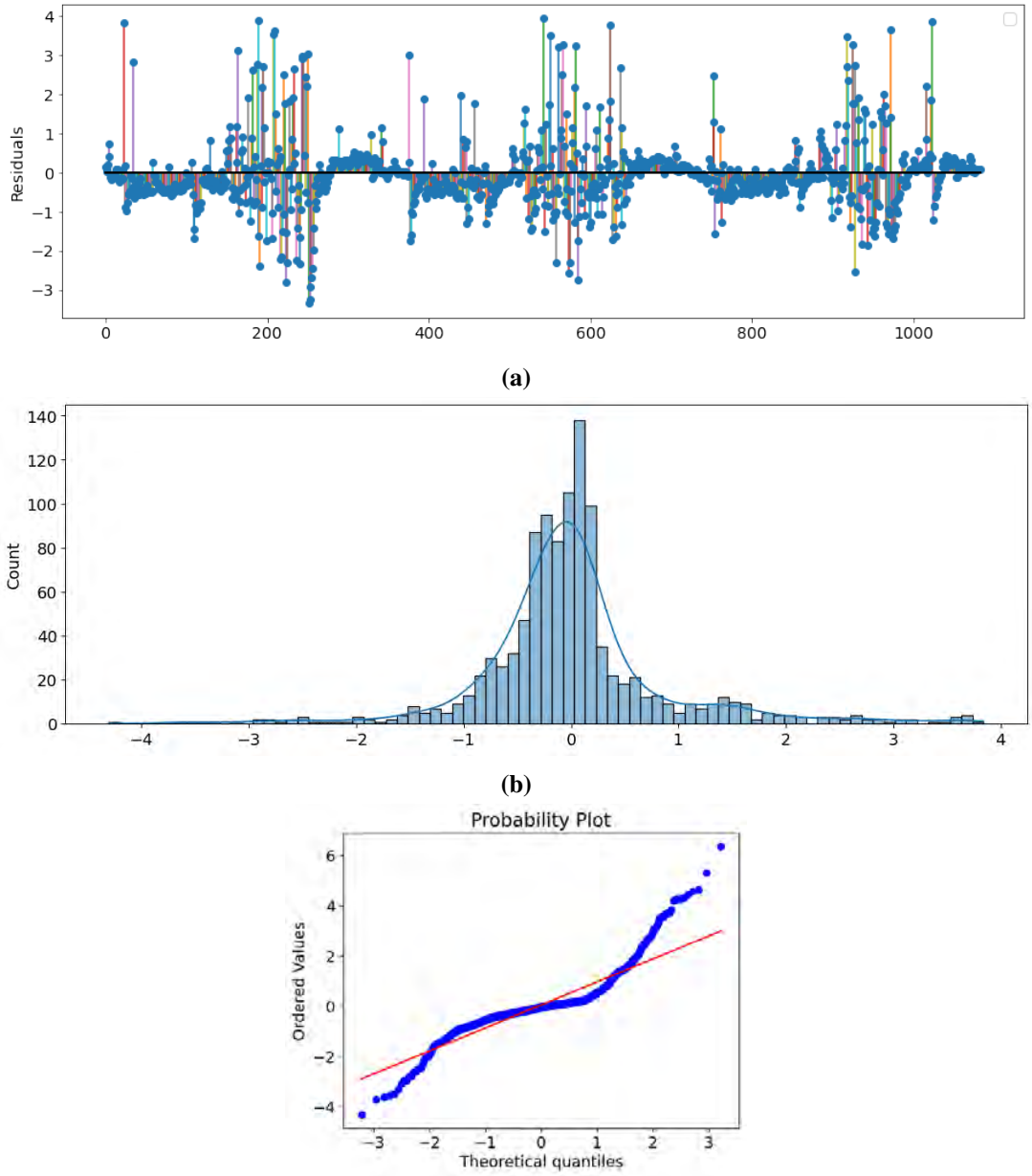


(b)



(c)

Fig. 4.15: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the SVR model in daily wind speed forecasting at Pokhran, Rajasthan.



(a)

(b)

(c)

Fig. 4.16: (a) Residual plot, (b) histogram plot, and (c) the P-P plot of the standardized residuals corresponding to the SVR model in daily GHI forecasting at Pokhran, Rajasthan.

4.4 Summary

Machine learning models, by leveraging advanced algorithms and historical data, revolutionize forecasting by delivering reliable predictions and valuable insights across diverse industries. In this chapter, we have implemented seven machine learning methods, namely SVR, ANN, LSTM, BiLSTM, encoder-decoder LSTM, attention layer LSTM, and CNN for forecasting of wind speed and solar irradiance at four different timescales and four selected locations. Each of these machine learning models is controlled by a variety of variables, including the type and number of hidden layers, the activation function, the optimization technique, the loss function, the epochs, and the learning rate. We identify the best parameter values from a range of potential parameter values so that the error values turn out to be the minimum. We compare these seven models based on the RMSE values and note that no single model is globally best across time and space. For example, we observe that the ANN model has the least error values in monthly wind speed forecasting for only two out of four study sites, whereas, the ANN model has the least error values in monthly GHI forecasting across all four study sites. This behavior may be attributed to varying data size, inherent stochasticity, seasonal and non-linear variability in the dataset. We have provided tables and figures reflecting the comparative performance of the implemented models, highlighting the best fit models. In addition, the results are supported through the residual analysis. Therefore, in summary, the present chapter provides a comprehensive analysis of several state of the art machine learning models in renewable energy forecasting.

Chapter 5

Hybrid Models for Renewable Energy Forecasting

"*The formulation of the problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill.*"

– ALBERT EINSTEIN

This chapter provides a generic introduction of the hybrid models in renewable energy prediction and explains relevant methodology in two steps. In the first step, an ARIMA model is used to analyze the linear part of the problem. In the second step, a neural network model (ANN, CNN, and LSTM) is developed to model the residuals obtained through the ARIMA model. Since the ARIMA model is unable to capture the non-linear structure of the data, the residuals of linear model will contain information about the non-linearity. In this regard, the neural networks can be effectively utilized to model error terms of the ARIMA model. Therefore, a hybrid model essentially combines the strengths of the ARIMA model as well as machine learning models in analyzing different linear and non-linear patterns. As a result, it is advantageous to model linear and non-linear patterns separately by using different models and then combine the forecasts to improve the overall forecasting performance. In view of the above, we develop three hybrid models, namely ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM for forecasting of daily, weekly, and monthly wind speed and GHI data. The emanated results based on RMSE and MAPE values reveal that the ARIMA-ANN model has the best performance in wind speed forecasting across time and space. In case of GHI forecasting, the ARIMA-ANN has the best performance except one location for which the ARIMA-LSTM model has the best representation. At the end, we provide a comprehensive comparison of the results obtained from Chapter 3, Chapter 4, and Chapter 5.

Contents

5.1 Introduction

Making accurate forecasts of solar irradiance and wind speed is challenging since it is difficult to assess the exact characteristics of the time series data of the underlying process. The challenges of forecasting become more complex when the data are non-linear, missing, or multidimensional [153]. There are several approaches to time series analysis and forecasting, and each approach has its own advantages and drawbacks. Based on the adopted methodology, forecasting techniques are categorized in three classes: (i) statistical time series methods, (ii) artificial intelligence or machine learning methods, and (iii) hybrid methods. The advantage of time series models (discussed in Chapter 3) is that they establish a linear relationship among observed data points and future data points, whereas the artificial intelligence methods (discussed in Chapter 4) are utilized for modeling the big data, studying non-linear components, and learning from the data. Neither linear nor non-linear approaches are adequate general models that can be applied in all circumstances. The approximation of time series models to complex non-linear problems may not be adequate. On the other hand, using neural nets to model linear problems yielded mixed results. Due to this, hybrid approaches [158] are introduced in the literature to capture both the linear and non-linear patterns resulting in improvement in the forecasting performance [168]. These methods integrate the benefits of many separate models. For this, heterogeneous models, such as linear and non-linear models [11, 72, 138] or homogeneous models, such as neural networks with various configurations [36, 55, 76] may be combined to form hybrid models. Researchers have regularly employed the linear statistical time series ARIMA model [10] from linear techniques and artificial neural networks [2] from non-linear approaches in forecasting of renewable energy resources such as solar irradiance and wind speed [70]. Below, we provide a few relevant studies to highlight the efficacy and limitations of ARIMA based hybrid models.

5.2 Literature Survey

In recent years, it has been asserted that hybrid approaches, especially those developed by fusing the advantages of several models, produce outcomes that are noticeably superior to those of the individual methods. The following is a summary of few research works on hybrid models, which integrate the benefits of two or more distinct models for solar irradiance and wind speed forecasting.

In 2011, Wu and Chain [57] developed a unique hybrid model for hourly solar irradiance forecasting that combines ARMA and time delay neural network (TDNN) because of each having their own advantages in time series analysis. The proposed set-up produces more accurate

results as compared to the involved standalone models in the study sites.

In 2013, Huang et al. [51] highlighted an effective hybrid method to predict one-hour ahead hourly solar irradiation on cloudy days in Mildura, Australia. The hybrid model combines a dynamical system model and an autoregressive (AR) model and introduces the coupled autoregressive dynamical system (CARDS) model. The forecasting accuracy increases by 30% in comparison to models without this adjustment.

In 2017, Mukaram and Yusof [86] developed a hybrid model based on SARIMA and ANN models for daily and monthly solar irradiance forecasting in Peninsular, Malaysia. The experimental results reflect that the hybrid model provides comparatively accurate forecasts of solar radiation data when compared to standalone ANN and standalone SARIMA model. This indicates that the daily and monthly average solar radiation arise from a combination of linear and non-linear processes. The study also suggested to consider few alternative hybrid methods such as SVM-ANN and SARIMA-GARCH [59].

In 2017, Nair et al. [91] compared the performance of ANN, ARIMA, and hybrid ARIMA-ANN models for the forecasting of wind speed at three different locations from Tamil Nadu, India. The study has noted the effectiveness of hybrid model at three different timescales based on various error metrics. They concluded that the hybrid model performs better than both of the ANN model and the ARIMA model since the behavior of the wind speed is both linear as well as non-linear.

In 2021, Huang et al. [52] designed a hybrid model for wind speed forecasting using exponential smoothing ARMA model that extracts the linear patterns hidden in the time series, and the back propagation neural network optimized by the cuckoo search algorithm [66] to extract the non-linear patterns in the data. Experimental results based on nine datasets from the Penglai wind farm in China ensure that the prediction accuracy of the novel hybrid system is higher than that of the single methods.

In 2021, Huang and Hui [50] proposed a hybrid model for solar irradiance forecasting using four components, namely the signal decomposition (EWT), neural network (NARX), Adaboost, and ARIMA. The experiment examined nine models and performed one, three, and five steps ahead predictions to demonstrate the resilience of the multi-step prediction model for four different datasets from Changde weather station in Hunan, China.

In view of the above studies, this chapter considers linear ARIMA model as it has been extensively used in solar irradiance and wind speed forecasting. In addition, due to the effective performance of ANN [136], CNN, and LSTM models observed in Chapter 4, we consider them for modeling the non-linear component to finally build the hybrid models. Thus, three hybrid models, namely ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM are implemented for daily, weekly, and monthly forecasting of solar irradiance and wind speed at four selected locations in

India. Finally, we compare forecasting efficacy of standalone models with these hybrid models based on RMSE and MAPE error values.

5.3 Methodology

In their respective linear or non-linear domains, both ARIMA and machine learning models have been successful. However, none of them serves as a generalized model applicable to all situations [168]. Sometimes, the ARIMA models' approximations for complex non-linear problems may be inadequate. Similarly, the outcomes from modeling linear problems with machine learning models have been inconsistent [147]. Since it is difficult to completely describe the characteristics of the data in a real problem, hybrid methodology that has both linear and non-linear modeling capabilities can be a good strategy for practical use [20]. Thus, various aspects of the underlying patterns can be suitably studied by combining different standalone models in a hybrid setup. Mathematically, a time series y_t can be represented as a combination of a linear structure L_t and a non-linear component N_t as follows.

$$y_t = L_t + N_t \tag{5.1}$$

The proposed methodology of the hybrid models (ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM) consists of the following steps:

- Step 1: Preprocess the time series data to remove any outliers and to resample at desired timescales.
- Step 2: Apply the Box-Jenkins methodology to identify the optimal ARIMA model for the linear component of the time series. This involves testing for stationarity, differencing (if necessary), estimating the model parameters, and checking the model adequacy.
- Step 3: Obtain the residuals of the fitted ARIMA model and carry out residual analysis. The residuals are given by

  $$e_t = y_t - \hat{L}_t, \tag{5.2}$$

  where e_t is the residual at time t, y_t is the actual value of the time series at time t, and $\hat{L}_t$ is the predicted value by the ARIMA model at time t. This expression represents the difference between the actual and predicted values of the time series by the ARIMA model. A linear model is considered insufficient if its residuals exhibit any non-linear patterns. However, it may be noted that the residual analysis alone can not determine the presence of any non-linear patterns in the data. In fact, there is no stringent diagnostic statistics to decide non-linear autocorrelation connections.

- Step 4: Use the residuals from the ARIMA model as the input for the machine learning model (ANN, CNN, and LSTM) so as to capture the non-linear component of the time series by learning from the error patterns. Then, train the neural network model using a suitable learning algorithm, such as backpropagation, and select the optimal network architecture, such as the number of hidden layers and neurons, based on some performance criteria, such as RMSE or MAPE.

- Step 5: Combine the outputs of the ARIMA and machine learning models to obtain the final forecast for the time series. The forecast is given by:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{5.3}$$

where, $\hat{y}_t$ is the final forecast for time t, $\hat{N}_t$ is the output of the machine learning model at time t, and $\hat{L}_t$ is the output of the ARIMA model at time t. This expression represents a sum of forecasted linear and non-linear components of the time series.

This methodology is based on the work of Zhang et al. [168], who proposed a hybrid ARIMA-ANN model for time series forecasting and demonstrated its superior performance over individual models in various applications. In addition, Wahedi et al. [147] and Sahin et al. [20] have also utilized the same methodology of the ARIMA based hybrid models in time series forecasting.

5.4 Results

We discuss the results in terms of RMSE and MAPE values obtained through the three implemented hybrid models for daily, weekly, and monthly forecasting of wind speed and GHI data in Section 5.4.1. It may be noted that the model parameters for the ARIMA are same as used in Chapter 3 and model parameters for the ANN, CNN, and the LSTM are same as used in Chapter 4. As stated in the above methodology section, we first implemented the ARIMA model and carried out the corresponding residual analysis. We applied Jarque-Bera test to check for normality of the residuals. The test statistic and p-value at the significance level of $\alpha = 5\%$ for the standalone models and the hybrid models are listed in Section 5.4.2. For comparison, the results of the standalone models in terms of error values are also included in this section.

5.4.1 Results of Hybrid Models

The RMSE and MAPE values of ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM models for wind speed and GHI forecasting are tabulated in Table 5.1 and Table 5.2, respectively. We

observe that the ARIMA-ANN model has the best performance in wind speed forecasting for monthly, weekly, and daily data across all the four locations. In case of GHI forecasting, the ARIMA-ANN has the best performance except one location (Pavagada) where the ARIMA-LSTM model has the best representation.

Table 5.1: Results of hybrid models in wind speed forecasting

Location	**Model**	**Monthly**		**Weekly**		**Daily**	
		RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
Pokhran, Rajasthan	ARIMA-ANN	**0.382**	0.093	**0.767**	0.251	**0.821**	0.232
	ARIMA-CNN	0.690	0.154	0.807	0.279	0.895	0.273
	ARIMA-LSTM	0.559	0.140	0.910	0.327	0.931	0.312
Bitta, Gujarat	ARIMA-ANN	**0.356**	0.085	**0.654**	0.176	**0.815**	0.194
	ARIMA-CNN	0.619	0.150	0.700	0.204	0.944	0.204
	ARIMA-LSTM	0.701	0.173	0.673	0.183	1.002	0.207
Pavagada, Karnataka	ARIMA-ANN	**0.495**	0.134	**1.024**	0.467	**0.724**	0.019
	ARIMA-CNN	0.930	0.332	1.079	0.544	0.875	0.196
	ARIMA-LSTM	1.683	0.499	1.449	0.599	0.928	0.214
Ramagundam, Telangana	ARIMA-ANN	**0.327**	0.125	**0.700**	0.398	**0.770**	0.240
	ARIMA-CNN	0.649	0.369	0.717	0.518	0.957	0.310
	ARIMA-LSTM	0.914	0.700	0.826	0.553	0.915	0.281

Table 5.2: Results of hybrid models in GHI forecasting

Location	**Model**	**Monthly**		**Weekly**		**Daily**	
		RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
Pokhran, Rajasthan	ARIMA-ANN	38.239	0.011	**39.512**	0.047	**78.560**	0.191
	ARIMA-CNN	48.159	0.073	48.159	0.055	98.495	0.371
	ARIMA-LSTM	**8.831**	0.012	47.048	0.055	92.125	0.312
Bitta, Gujarat	ARIMA-ANN	**16.705**	0.024	**46.945**	0.066	**76.298**	0.114
	ARIMA-CNN	50.494	0.074	52.293	0.062	80.494	0.342
	ARIMA-LSTM	18.832	0.027	49.515	0.066	78.832	0.187
Pavagada, Karnataka	ARIMA-ANN	10.254	0.015	60.587	0.085	90.254	0.315
	ARIMA-CNN	37.871	0.056	63.816	0.096	99.871	0.356
	ARIMA-LSTM	**9.528**	0.014	**55.686**	0.070	**87.528**	0.314
Ramagundam, Telangana	ARIMA-ANN	**24.670**	0.040	**65.343**	0.096	**124.670**	0.340
	ARIMA-CNN	52.357	0.110	74.354	0.109	152.357	0.410
	ARIMA-LSTM	26.015	0.041	65.851	0.083	126.015	0.341

5.4.2 Hybrid versus Standalone Models

After discussing the relative performance of hybrid models in the previous section, this section provides a comparison of standalone ARIMA, ANN, CNN, and LSTM models versus hybrid ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM models in terms of error values. In addition, for each standalone and hybrid models, we present the observed value of the Jarque-Bera test statistic of the residuals and associated p-value at $\alpha = 5\%$ in Table 5.3 – Table 5.8. The instances where the hybrid models outperform the standalone models either in terms of lower RMSE or smaller test statistic values are highlighted in bold. From these tables, we note the following:

1. In monthly GHI, there are only three instances where the ARIMA-LSTM model outperforms the standalone ARIMA and standalone LSTM in terms of RMSE values.

2. In weekly and daily GHI, none of the hybrid models provides lesser RMSE value than that of the standalone models.

3. In wind speed forecasting, across all timescales, none of the hybrid models provides lesser RMSE value than that of the standalone models.

4. Nevertheless, the test statistic from the Jarque-Bera test reveals significant improvement towards the Gaussian behavior of residuals in all the hybrid models as compared to the standalone models across both time and space.

5. In wind speed forecasting, as we move from longer to shorter timescale, i.e, from monthly to weekly, the p-value reduces; in daily timescale, it is always closer to zero. In GHI forecasting, except monthly timescale, the p-value is always close to zero. This suggests deviation of residuals from the normal distribution in a statistical sense.

Table 5.3: Monthly wind speed forecasting

Location	Model	RMSE	MAPE	Test Statistic	$p-$ Value
Pokhran, Rajasthan	ARIMA	0.363	0.103	26.066	~ 0
	ANN	0.114	0.032	0.687	0.708
	CNN	0.122	0.032	1.040	0.594
	LSTM	0.549	0.122	8.432	0.392
	ARIMA-ANN	0.382	0.093	11.123	0.003
	ARIMA-CNN	0.690	0.154	1.905	0.385
	ARIMA-LSTM	0.559	0.140	**6.331**	0.042
Bitta, Gujarat	ARIMA	0.442	0.067	11.861	0.002
	ANN	0.104	0.028	2.673	0.262
	CNN	0.114	0.029	0.081	0.960
	LSTM	0.488	0.119	3.261	0.105
	ARIMA-ANN	0.356	0.085	1.826	0.394
	ARIMA-CNN	0.619	0.150	1.199	0.549
	ARIMA-LSTM	0.701	0.173	**0.913**	0.633
Pavagada, Karnataka	ARIMA	0.468	0.140	1.811	0.404
	ANN	0.102	0.026	1.790	0.408
	CNN	0.100	0.026	8.863	0.011
	LSTM	0.697	0.167	3.614	0.253
	ARIMA-ANN	0.495	0.134	**0.115**	0.943
	ARIMA-CNN	0.930	0.332	**0.156**	0.924
	ARIMA-LSTM	1.683	0.499	**0.316**	0.853
Ramagundam, Telangana	ARIMA	0.442	0.170	0.146	0.929
	ANN	0.118	0.039	0.029	0.985
	CNN	0.107	0.039	0.381	0.826
	LSTM	0.429	0.146	0.620	0.714
	ARIMA-ANN	0.327	0.125	0.521	0.770
	ARIMA-CNN	0.649	0.369	0.808	0.667
	ARIMA-LSTM	0.914	0.700	0.248	0.882

Table 5.4: Monthly GHI forecasting

Location	Model	RMSE	MAPE	Test Statistic	$p-$Value
Pokhran, Rajasthan	ARIMA	20.012	0.084	1.280	0.527
	ANN	9.444	0.010	4.453	0.107
	CNN	18.488	0.017	2.303	0.316
	LSTM	16.838	0.022	15.862	0.002
	ARIMA-ANN	38.239	0.011	108.895	~ 0
	ARIMA-CNN	48.159	0.073	1.448	0.484
	ARIMA-LSTM	**8.831**	0.012	**0.273**	0.872
Bitta, Gujarat	ARIMA	39.506	0.060	0.394	0.821
	ANN	9.876	0.012	18.293	0.001
	CNN	15.720	0.191	3.509	0.172
	LSTM	27.366	0.038	15.449	0.004
	ARIMA-ANN	16.705	0.024	**0.147**	0.927
	ARIMA-CNN	50.494	0.174	0.631	0.729
	ARIMA-LSTM	**18.832**	**0.027**	0.092	0.630
Pavagada, Karnataka	ARIMA	20.107	0.034	1.980	0.371
	ANN	7.608	0.012	0.337	0.845
	CNN	10.849	0.015	1.807	0.404
	LSTM	21.073	0.029	16.707	0.002
	ARIMA-ANN	10.254	0.015	**1.792**	0.408
	ARIMA-CNN	37.871	0.056	**1.300**	0.521
	ARIMA-LSTM	**9.528**	0.014	**1.731**	0.420
Ramagundam, Telangana	ARIMA	22.704	0.041	2.965	0.227
	ANN	14.648	0.022	0.171	0.917
	CNN	19.113	0.022	3.229	0.199
	LSTM	35.058	0.055	13.624	0.001
	ARIMA-ANN	24.670	0.040	0.969	0.615
	ARIMA-CNN	52.357	0.110	**0.248**	0.883
	ARIMA-LSTM	26.015	0.041	**0.962**	0.618

Table 5.5: Weekly wind speed forecasting

Location	Model	RMSE	MAPE	Test Statistic	$p-$Value
Pokhran, Rajasthan	ARIMA	0.801	0.237	47.084	~ 0
	ANN	0.449	0.121	4.098	0.128
	CNN	0.442	0.117	5.195	0.074
	LSTM	0.485	0.125	6.983	0.063
	ARIMA-ANN	0.767	0.251	**2.249**	0.324
	ARIMA-CNN	0.807	0.279	**1.430**	0.489
	ARIMA-LSTM	0.910	0.327	**5.649**	0.059
Bitta, Gujarat	ARIMA	0.657	0.194	25.868	~ 0
	ANN	0.338	0.080	22.685	~ 0
	CNN	0.348	0.083	18.961	~ 0
	LSTM	0.390	0.091	15.960	~ 0
	ARIMA-ANN	0.654	0.176	**6.526**	0.038
	ARIMA-CNN	0.700	0.204	**3.118**	0.210
	ARIMA-LSTM	0.673	0.183	**1.617**	0.445
Pavagada, Karnataka	ARIMA	0.868	0.238	34.947	~ 0
	ANN	0.405	0.109	3.394	0.183
	CNN	0.413	0.107	3.557	0.168
	LSTM	0.424	0.114	4.079	0.150
	ARIMA-ANN	1.024	0.467	21.242	~ 0
	ARIMA-CNN	1.079	0.544	28.131	~ 0
	ARIMA-LSTM	1.449	0.599	31.919	~ 0
Ramagundam, Telangana	ARIMA	0.637	0.271	101.172	~ 0
	ANN	0.342	0.119	4.641	0.098
	CNN	0.345	0.126	0.095	0.953
	LSTM	0.370	0.124	4.760	0.057
	ARIMA-ANN	0.700	0.398	**4.619**	0.099
	ARIMA-CNN	0.717	0.518	1.130	0.568
	ARIMA-LSTM	0.826	0.553	**1.620**	0.444

Table 5.6: Weekly GHI forecasting

Location	Model	RMSE	MAPE	Test Statistic	$p-$Value
Pokhran, Rajasthan	ARIMA	49.931	0.082	318.256	~ 0
	ANN	27.449	0.030	147.687	~ 0
	CNN	27.560	0.030	392.568	~ 0
	LSTM	31.728	0.063	137.999	~ 0
	ARIMA-ANN	39.512	0.047	**90.532**	~ 0
	ARIMA-CNN	48.159	0.055	**56.466**	~ 0
	ARIMA-LSTM	47.048	0.055	**102.122**	~ 0
Bitta, Gujarat	ARIMA	48.657	0.194	248.858	~ 0
	ANN	25.374	0.031	333.962	~ 0
	CNN	19.232	0.040	410.395	~ 0
	LSTM	21.992	0.068	368.963	~ 0
	ARIMA-ANN	46.945	0.066	**30.301**	~ 0
	ARIMA-CNN	52.293	0.062	**83.509**	~ 0
	ARIMA-LSTM	49.515	0.066	**36.551**	~ 0
Pavagada, Karnataka	ARIMA	45.912	0.124	169.256	~ 0
	ANN	34.038	0.047	22.998	~ 0
	CNN	36.762	0.047	86.941	~ 0
	LSTM	30.438	0.098	161.808	~ 0
	ARIMA-ANN	60.587	0.085	**54.991**	~ 0
	ARIMA-CNN	63.816	0.096	**67.308**	~ 0
	ARIMA-LSTM	55.686	0.070	**44.170**	~ 0
Ramagundam, Telangana	ARIMA	40.974	0.175	116.326	~ 0
	ANN	44.849	0.067	174.135	~ 0
	CNN	45.374	0.066	163.233	~ 0
	LSTM	34.594	0.125	161.397	~ 0
	ARIMA-ANN	65.343	0.096	**31.227**	~ 0
	ARIMA-CNN	74.354	0.109	**16.832**	~ 0
	ARIMA-LSTM	65.851	0.083	**50.910**	~ 0

Table 5.7: Daily wind speed forecasting

Location	Model	RMSE	MAPE	Test Statistic	$p-$Value
Pokhran, Rajasthan	ARIMA	1.317	0.349	147.687	~ 0
	ANN	0.743	0.247	48.700	0.020
	CNN	0.775	0.225	92.683	0.010
	LSTM	0.720	0.241	37.241	0.244
	ARIMA-ANN	0.821	0.232	**8.964**	0.487
	ARIMA-CNN	0.895	0.273	**42.448**	0.240
	ARIMA-LSTM	0.931	0.312	**16.273**	0.347
Bitta, Gujarat	ARIMA	1.126	0.227	201.394	~ 0
	ANN	0.642	0.176	74.962	~ 0
	CNN	0.629	0.151	103.395	~ 0
	LSTM	0.662	0.175	68.963	~ 0
	ARIMA-ANN	0.815	0.194	**10.147**	0.072
	ARIMA-CNN	0.944	0.204	**56.631**	0.029
	ARIMA-LSTM	1.002	0.207	80.092	0.030
Pavagada, Karnataka	ARIMA	1.119	0.309	252.908	~ 0
	ANN	0.523	0.155	90.007	0.006
	CNN	0.533	0.144	186.415	~ 0
	LSTM	0.534	0.163	85.708	0.112
	ARIMA-ANN	0.724	0.019	**18.792**	0.408
	ARIMA-CNN	0.875	0.196	**71.002**	0.052
	ARIMA-LSTM	0.928	0.214	**31.371**	0.206
Ramagundam, Telangana	ARIMA	0.926	0.375	109.820	~ 0
	ANN	0.583	0.219	149.006	~ 0
	CNN	0.548	0.192	134.330	0.009
	LSTM	0.570	0.217	207.341	0.070
	ARIMA-ANN	0.770	0.240	**84.969**	0.115
	ARIMA-CNN	0.957	0.310	**74.723**	0.083
	ARIMA-LSTM	0.915	0.281	**62.780**	~ 0

Table 5.8: Daily GHI forecasting

Location	Model	RMSE	MAPE	Test Statistic	$p-$Value
Pokhran, Rajasthan	ARIMA	73.330	0.069	291.280	~ 0
	ANN	83.030	0.319	147.687	~ 0
	CNN	73.054	0.281	392.568	~ 0
	LSTM	68.804	0.113	137.999	~ 0
	ARIMA-ANN	78.560	0.191	**76.708**	~ 0
	ARIMA-CNN	98.495	0.371	146.097	~ 0
	ARIMA-LSTM	92.125	0.312	**87.650**	~ 0
Bitta, Gujarat	ARIMA	77.952	0.092	248.005	~ 0
	ANN	97.730	0.399	333.962	~ 0
	CNN	77.406	0.300	410.395	~ 0
	LSTM	74.006	0.159	368.963	~ 0
	ARIMA-ANN	76.298	0.114	**40.147**	~ 0
	ARIMA-CNN	80.494	0.342	**60.631**	~ 0
	ARIMA-LSTM	78.832	0.187	**55.092**	~ 0
Pavagada, Karnataka	ARIMA	87.392	0.116	41.980	~ 0
	ANN	92.132	0.327	42.998	~ 0
	CNN	75.906	0.315	86.941	~ 0
	LSTM	74.043	0.103	16.808	~ 0
	ARIMA-ANN	90.254	0.315	51.792	~ 0
	ARIMA-CNN	99.871	0.356	67.300	~ 0
	ARIMA-LSTM	87.528	0.314	41.731	~ 0
Ramagundam, Telangana	ARIMA	75.660	0.102	121.609	~ 0
	ANN	101.844	0.416	449.001	~ 0
	CNN	91.404	0.386	328.749	~ 0
	LSTM	88.996	0.123	289.144	~ 0
	ARIMA-ANN	124.670	0.340	**154.969**	~ 0
	ARIMA-CNN	152.357	0.410	437.248	~ 0
	ARIMA-LSTM	126.015	0.341	**160.962**	~ 0

5.5 A Comprehensive Summary of All Implemented Models

To recall, we implemented the statistical time series methods in Chapter 3 where we highlighted the best fit models (specially the WS-ARIMA for daily and weekly forecasting) for wind speed and solar irradiance at different timescales. Similarly, in Chapter 4, we compared the efficacy of different machine learning models in renewable energy forecasting and listed the best fit models

at specific timescales. Finally, a comprehensive comparison of the efficacy of the implemented models is required to provide the generic guidelines to choose the best fit model in wind speed and GHI forecasting at a desired timescale. In this section, we compare the results of all the models implemented in Chapter 3, Chapter 4, and Chapter 5. A comparative performance in terms of RMSE and MAPE values of the time series models, machine learning models, and hybrid models for wind speed and GHI data is presented in Table 5.9 and 5.10, respectively. Results of the best fit models are highlighted in bold. We also provide a list of three best suitable models in wind speed and GHI forecasting at different timescales in Table 5.11 and 5.12, respectively.

In monthly wind speed forecasting, the ANN model has the least RMSE values in Rajasthan and Gujarat, whereas the CNN model has the least error values at the other two sites. In weekly wind speed forecasting, the CNN has the least error in Rajasthan, ANN has the least error in Gujarat and Telangana, and the WS-ARIMA has the least error value in Karnataka. In daily wind speed forecasting, the WS-ARIMA has the best representation across all four study sites. In hourly wind speed prediction, the bilSTM has the best results in Rajasthan and Karnataka, whereas, the encoder-decoder LSTM has the best performance in the other two study sites. It may be noted that for hourly dataset, we consider machine learning models for comparison as we implemented the time series models based on three years of data (please refer to Chapter 3).

In monthly GHI data, the ARIMA-LSTM model has the least RMSE in Rajasthan, whereas the ANN model has the best representation at three other study sites. In weekly forecasting, the ANN has the least RMSE in Rajasthan, CNN has the least RMSE in Gujarat, the LSTM has the least RMSE in Karnataka, the WS-ARIMA model has the least error in Teangana. In daily forecasting, the WS-ARIMA has the least error across all four study sites, whereas in hourly forecasting, the SVR model has the best representation across all four study sites. As similar to hourly wind speed forecasting, we have considered only the machine learning models for performance comparison in hourly GHI forecasting, .

Table 5.9: Results obtained from different models for wind speed data at four study sites (L1: Pokhran, Rajasthan, L2: Bitta, Gujarat, L3: Pavagada, Karnataka, L4: Ramagundam, Telangana)

Monthly Forecasting								
	L1		**L2**		**L3**		**L4**	
Model	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
AR	0.488	0.116	-	-	-	-	-	-
MA	0.518	0.112	-	-	-	-	-	-
ARIMA	0.363	0.103	0.442	0.067	0.468	0.140	0.442	0.170
SARIMA	0.428	0.105	0.319	0.075	0.383	0.110	0.354	0.102
WS-ARIMA	0.457	0.111	0.557	0.111	0.746	0.168	0.384	0.159
SVR	0.417	0.087	0.328	0.087	0.422	0.092	0.359	0.104
ANN	**0.114**	0.032	**0.104**	0.028	0.102	0.026	0.118	0.039
CNN	0.122	0.032	0.114	0.029	**0.100**	0.026	**0.107**	0.039
LSTM	0.549	0.122	0.488	0.119	0.697	0.167	0.429	0.146
BiLSTM	0.552	0.134	0.750	0.172	0.603	0.134	0.363	0.126
Encoder-Decoder	0.383	0.086	0.362	0.085	0.428	0.110	0.369	0.111
Attention Layer	1.516	0.380	0.441	0.112	1.973	0.436	1.549	0.488
ARIMA-ANN	0.382	0.093	0.356	0.085	0.495	0.134	0.327	0.125
ARIMA-CNN	0.690	0.154	0.619	0.150	0.930	0.332	0.649	0.369
ARIMA-LSTM	0.559	0.140	0.701	0.173	1.683	0.499	0.914	0.700
Weekly Forecasting								
AR	1.288	0.323	-	-	-	-	-	-
MA	1.617	0.412	-	-	-	-	-	-
ARIMA	0.801	0.237	0.657	0.194	0.868	0.238	0.637	0.271
SARIMA	1.281	0.424	1.302	0.376	1.072	0.331	0.940	0.319
WS-ARIMA	0.483	0.129	0.453	0.109	**0.393**	0.101	0.377	0.128
SVR	0.663	0.174	0.571	0.132	0.709	0.170	0.542	0.174
ANN	0.449	0.121	**0.338**	0.080	0.405	0.109	**0.342**	0.119
CNN	**0.442**	0.117	0.348	0.083	0.413	0.107	0345	0.126
LSTM	0.485	0.125	0.390	0.091	0.424	0.114	0.370	0.124
BiLSTM	0.669	0.171	0.592	0.131	0.697	0.166	0.528	0.192
Encoder-Decoder	0.957	0.242	0.802	0.187	0.970	0.226	0.727	0.257
Attention Layer	0.831	0.236	0.732	0.169	0.893	0.213	1.809	0.574
ARIMA-ANN	0.767	0.251	0.654	0.176	1.024	0.467	0.700	0.398
ARIMA-CNN	0.807	0.279	0.700	0.204	1.079	0.544	0.717	0.518

ARIMA-LSTM	0.910	0.327	0.673	0.183	1.449	0.599	0.826	0.553
Daily Forecasting								
AR	1.383	0.329	-	-	-	-	-	-
MA	1.521	0.357	-	-	-	-	-	-
ARIMA	1.317	0.349	1.126	0.227	1.119	0.309	0.926	0.375
SARIMA	2.067	0.762	2.212	0.701	2.162	0.664	1.352	0.504
WS-ARIMA	**0.372**	0.113	**0.282**	0.073	**0.127**	0.036	**0.232**	0.093
SVR	0.709	0.202	0.642	0.177	0.526	0.151	0.578	0.209
ANN	0.743	0.247	0.642	0.176	0.523	0.155	0.583	0.219
CNN	0.775	0.225	0.629	0.151	0.533	0.144	0.548	0.192
LSTM	0.720	0.241	0.662	0.175	0.534	0.163	0.570	0.217
BiLSTM	0.783	0.219	0.699	0.157	0.524	0.156	0.543	0.190
Encoder-Decoder	0.832	0.257	0.671	0.184	0.557	0.165	0.592	0.233
Attention Layer	1.141	0.360	1.028	0.295	1.082	0.295	0.940	0.362
ARIMA-ANN	0.821	0.232	0.815	0.194	0.724	0.019	0.770	0.240
ARIMA-CNN	0.895	0.273	0.944	0.204	0.875	0.196	0.957	0.310
ARIMA-LSTM	0.931	0.312	1.002	0.207	0.928	0.214	0.915	0.281
Hourly Forecasting								
SVR	0.174	0.063	0.186	0.059	0.186	0.057	0.209	0.086
ANN	1.002	0.391	0.930	0.328	0.917	0.287	1.031	0.369
CNN	0.701	0.240	0.653	0.201	0.700	0.281	0.494	0.192
LSTM	0.164	0.075	0.174	0.087	0.158	0.082	0.174	0.086
BiLSTM	**0.154**	0.057	0.153	0.049	**0.150**	0.082	0.154	0.072
Encoder-Decoder	0.155	0.055	**0.149**	0.052	0.152	0.048	**0.153**	0.064
Attention Layer	0.177	0.063	0.161	0.052	0.153	0.041	0.160	0.054

Table 5.10: Results obtained from different models for GHI data at four study sites (L1: Pokhran, Rajasthan, L2: Bitta, Gujarat, L3: Pavagada, Karnataka, L4: Ramagundam, Telangana)

Monthly Forecasting								
	L1		**L2**		**L3**		**L4**	
Model	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
AR	26.996	0.094	-	-	-	-	-	-
MA	32.514	0.101	-	-	-	-	-	-
ARIMA	20.012	0.084	39.506	0.060	20.107	0.034	22.704	0.041
SARIMA	15.437	0.020	22.765	0.030	15.821	0.027	19.384	0.029
WS-ARIMA	49.156	0.069	54.000	0.079	39.689	0.058	34.079	0.060
SVR	13.193	0.020	24.945	0.035	18.756	0.028	30.638	0.048
ANN	9.444	0.010	**9.876**	0.012	**7.608**	0.012	**14.648**	0.022
CNN	18.488	0.017	15.720	0.191	10.849	0.015	19.113	0.022
LSTM	73.054	0.281	77.406	0.300	75.904	0.315	91.404	0.386
BiLSTM	18.353	0.026	29.498	0.039	21.492	0.028	35.492	0.059
Encoder-Decoder	19.319	0.024	25.884	0.032	25.594	0.026	35.449	0.051
Attention Layer	33.109	0.128	25.612	0.051	38.532	0.098	36.795	0.076
ARIMA-ANN	38.239	0.011	16.705	0.024	10.254	0.015	24.670	0.040
ARIMA-CNN	48.159	0.073	50.494	0.074	37.871	0.056	52.357	0.110
ARIMA-LSTM	**8.831**	0.012	18.832	0.027	9.528	0.014	26.015	0.041
Weekly Forecasting								
AR	55.509	0.160	-	-	-	-	-	-
MA	68.961	0.195	-	-	-	-	-	-
ARIMA	49.931	0.082	48.657	0.194	45.912	0.124	40.974	0.175
SARIMA	45.934	0.063	48.321	0.062	43.178	0.112	38.877	0.172
WS-ARIMA	28.127	0.031	30.848	0.041	32.037	0.056	**19.785**	0.061
SVR	38.353	0.050	42.852	0.053	55.551	0.076	61.179	0.092
ANN	**27.449**	0.030	25.374	0.031	34.038	0.047	44.849	0.067
CNN	27.560	0.030	**19.232**	0.040	36.762	0.047	45.374	0.066
LSTM	31.728	0.063	21.992	0.068	**30.438**	0.098	34.594	0.125
BiLSTM	40.913	0.047	44.761	0.058	52.521	0.074	54.602	0.083
Encoder-Decoder	57.162	0.066	63.107	0.071	66.412	0.088	76.699	0.106
Attention Layer	48.590	0.059	52.177	0.153	47.752	0.139	83.107	0.149
ARIMA-ANN	39.512	0.047	46.945	0.066	60.587	0.085	65.343	0.096
ARIMA-CNN	48.159	0.055	52.293	0.062	63.816	0.096	74.354	0.109

ARIMA-LSTM	47.048	0.055	49.515	0.066	55.686	0.070	65.851	0.083
Daily Forecasting								
AR	109.118	0.086	-	-	-	-	-	-
MA	89.580	0.118	-	-	-	-	-	-
ARIMA	73.330	0.069	77.952	0.092	87.392	0.116	75.660	0.102
SARIMA	81.677	0.093	84.718	0.106	92.064	0.240	100.941	0.215
WS-ARIMA	**34.804**	0.041	**33.671**	0.045	**33.776**	0.061	**52.633**	0.085
SVR	57.515	0.063	59.343	0.075	71.482	0.095	93.687	0.151
ANN	83.030	0.319	97.730	0.399	92.132	0.327	101.844	0.416
CNN	73.054	0.281	77.406	0.300	75.904	0.315	91.404	0.386
LSTM	68.804	0.113	74.006	0.159	74.043	0.103	88.996	0.123
BiLSTM	59.451	0.055	64.720	0.068	69.806	0.087	85.456	0.118
Encoder-Decoder	61.237	0.070	64.263	0.077	74.143	0.099	96.029	0.155
Attention Layer	79.095	0.101	79.052	0.113	82.962	0.186	83.987	0.179
ARIMA-ANN	78.560	0.191	76.298	0.114	90.254	0.315	124.670	0.340
ARIMA-CNN	98.495	0.371	80.494	0.342	99.871	0.356	152.357	0.410
ARIMA-LSTM	92.125	0.312	78.832	0.187	87.528	0.314	126.015	0.341
Hourly Forecasting								
SVR	**63.591**	0.178	**61.382**	0.169	**83.040**	0.289	**81.592**	0.323
ANN	89.904	0.340	84.017	0.316	94.893	0.397	97.800	0.403
CNN	76.034	0.157	81.911	0.185	84.621	0.1849	92.015	0.200
LSTM	65.731	0.187	64.407	0.144	84.161	0.185	82.772	0.147
BiLSTM	67.974	0.162	67.901	0.168	89.582	0.187	88.167	0.190
Encoder-Decoder	68.773	0.184	67.782	0.144	91.693	0.183	89.052	0.174
Attention Layer	91.141	0.214	104.065	0.286	86.993	0.290	98.005	0.276

Table 5.11: Ranking of the studied models for wind speed forecasting at different timescales

Timescale		L1	L2	L3	L4
Monthly	Rank 1	ANN	ANN	CNN	CNN
	Rank 2	CNN	CNN	ANN	ANN
	Rank 3	ARIMA	SARIMA	SARIMA	SVR
Weekly	Rank 1	CNN	ANN	WS-ARIMA	ANN
	Rank 2	ANN	CNN	ANN	CNN
	Rank 3	WS-ARIMA	LSTM	CNN	LSTM
Daily	Rank 1	WS-ARIMA	WS-ARIMA	WS-ARIMA	WS-ARIMA
	Rank 2	SVR	CNN	ANN	BiLSTM
	Rank 3	LSTM	SVR	BiLSTM	CNN
Hourly	Rank 1	BiLSTM	Encoder-Decoder	BiLSTM	Encoder-Decoder
	Rank 2	Encoder-Decoder	BiLSTM	Encoder-Decoder	BiLSTM
	Rank 3	LSTM	Attention Layer	Attention Layer	Attention Layer

Table 5.12: Ranking of the studied models for GHI forecasting at different timescales

Timescale		L1	L2	L3	L4
Monthly	Rank 1	ARIMA-LSTM	ANN	ANN	ANN
	Rank 2	ANN	CNN	ARIMA-LSTM	CNN
	Rank 3	SVR	ARIMA-ANN	ARIMA-ANN	SARIMA
Weekly	Rank 1	ANN	CNN	WS-ARIMA	LSTM
	Rank 2	CNN	LSTM	ANN	WS-ARIMA
	Rank 3	WS-ARIMA	ANN	CNN	ANN
Daily	Rank 1	WS-ARIMA	WS-ARIMA	WS-ARIMA	WS-ARIMA
	Rank 2	SVR	SVR	BiLSTM	Attention Layer
	Rank 3	BiLSTM	Encoder-Decoder	SVR	BiLSTM
Hourly	Rank 1	SVR	SVR	SVR	SVR
	Rank 2	LSTM	LSTM	LSTM	LSTM
	Rank 3	BiLSTM	Encoder-Decoder	CNN	BiLSTM

5.6 Summary

This chapter discusses the implementation of ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM hybrid models developed through time series ARIMA model and three machine learning models, namely ANN, CNN, and LSTM. The methodology works in a two-fold manner: first, implementing an ARIMA model to analyze the linear part of the data and then, implementing a neural network (ANN, CNN, and LSTM) to model the residuals of the ARIMA. Based on the RMSE values corresponding to daily, weekly, and monthly forecasting, we note that the ARIMA-ANN model has the best performance in wind speed forecasting across time and space, whereas in case of GHI forecasting, the ARIMA-ANN has the best performance except one location in which the ARIMA-LSTM has the best representation.

In order to assess the efficacy of the hybrid models in comparison to the corresponding standalone models, we compare their RMSE values and residuals (through the observed values of Jarque-Bera test statistic). Based on RMSE values, we notice that except a few instances, none of the hybrid models provides better results than that of the standalone models. Nevertheless, the Jarque-Bera test suggests that hybrid models enable a significant improvement towards the normal behavior of residuals. However, the ineffective performance of hybrid models does not necessarily discourage the users to implement and explore the hybrid models in renewable energy forecasting since the present study has focused on a specific setup among several available setups, such as decomposition based hybrid models [56, 103, 106, 159, 163], homogeneous hybrid models [36, 55, 63, 76], and other heterogeneous models as mentioned at the beginning of this chapter.

Finally, in order to provide a generic guideline to the users on choice of the best fit models in wind speed and GHI forecasting at a desired time horizon, we provide a comprehensive summary of all forecasting models implemented in Chapter 3, Chapter 4, and Chapter 5. We also list out the best three ranked models for each time horizon. Therefore, this chapter not only presents the implemention of hybrid models but also provides an extensive comparison among all fifteen implemented models.

Chapter 6

Conclusions and Future Works

"*To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science.*"

– ALBERT EINSTEIN

The present thesis has addressed the concept of renewable energy forecasting using statistics and machine learning. For this, Chapter 1 has provided an overview and rationale of the thesis along with the principal objective and scope of the thesis. Chapter 2 has carried out statistical assessment and preprocessing of the hourly, daily, weekly, and monthly wind speed and solar irradiance data collected at four different locations in India. Chapter 3 has presented a comprehensive analysis of various time series models in renewable energy forecasting. Chapter 4 has dealt with the implementation of several machine learning methods. Chapter 5 has discussed the implementation of a few hybrid models along with a comprehensive summary of all implemented models in the thesis. Finally, the present chapter (Chapter 6) summarizes research outcomes of the thesis accompanying with future research directions. The overall content of this chapter is provided below.

Contents

6.1 Research Objectives and Their Conclusions

In Chapter 1, the main objective of the thesis was framed as the forecasting of wind speed and GHI using statistical time series, machine learning, and hybrid methods for the Indian subcontinent. To accomplish this objective, four sub-objectives (for the sake of simplicity, called "objectives" hereafter) were framed. A brief summary of each objective and associated concluding remark is provided in the following sub-sections.

6.1.1 Research Objective 1: To carry out preliminary data analysis and to explore the best fit probability distribution(s) for wind speed and GHI data

To accomplish the first research objective, in Chapter 2, we performed statistical analysis based on 15 years (2000–2014) of wind speed and solar irradiance data at four different timescales as well as four selected locations in India. The analysis was carried out in four major steps. (i) First, we examined data characteristics using descriptive measures and time series plots. We noted that the highest mean wind speed comes from Gujarat $(3.53\ m/s)$, whereas the lowest mean wind speed has been recorded in Telangana $(2.34\ m/s)$. Similarly, the highest mean GHI comes from Karnataka $(579.35\ W/m^2)$, whereas the lowest mean GHI has been recorded in Telangana $(537.40\ W/m^2)$. The time series plots exhibit significant variations among the datasets as well as yearly seasonal pattern (ii) Then, we performed a time series decomposition to decompose the data into seasonal, trend, and irregular components. Although there are long term seasonal patterns in the data, there is no significant upward or downward trend. (iii) After this, we implemented the ADF test to check for stationarity in the data. The results at $\alpha = 5\%$ suggest overall stationarity of data. (iv) Finally, we studied five probability distributions, namely exponential, gamma, lognormal, Weibull, and exponential Weibull to determine the most suitable probability model for GHI and wind speed data. We carried out statistical inference using maximum-likelihood estimation and K-S test for parameter estimation and goodness of fit, respectively. We observe that the exponentiated Weibull has the best representation across wind speed and GHI data, suggesting it as a highly useful model for data on renewable energy sources. As a conclusion, the present study on preliminary data analysis and exploration of the best fit probability distribution(s) for wind speed and GHI data has successfully accomplished the research objective 1.

6.1.2 Research Objective 2: To implement various statistical time series methods for renewable energy forecasting

To accomplish the second research objective, we implemented several time series methods in Chapter 3. Particularly, we focused on the forecasting of wind speed and GHI data at hourly, daily, weekly, and monthly timescales using five statistical methods, namely AR, MA, ARIMA, SARIMA, and WS-ARIMA. We chose these time series models since the datasets exhibit seasonality, stationarity, and randomness. We adopted a grid search method to find optimum values of model parameters and used root mean square error (RMSE) to assess the performance of the studied models. Finally, we performed a residual analysis as a post-processing step to examine any systematic bias in the implemented models. The experimental results revealed that (i) for monthly forecasting, the SARIMA model has the best performance, (ii) for daily and weekly wind speed and GHI data, the WS-ARIMA method consistently outperforms the conventional time series methods with significant improvement in the forecasts across time and space, and (iii) for hourly forecasting, the WS-ARIMA and ARIMA model have comparable performance based on three years (2012–2014) of data. In addition, the results emphasize that the inclusion of sliding windows in conventional ARIMA model significantly improves the forecasting performance. Therefore, from the above discussions, it is concluded that the research objective 2 is successfully accomplished through the study of statistical time series methods in renewable energy forecasting.

6.1.3 Research Objective 3: To implement several machine learning techniques for short term, intermediate term, and long term renewable energy forecasting

To accomplish the third objective, in Chapter 4, we implemented various machine learning methods, namely SVR, ANN, LSTM, BiLSTM, encoder-decoder LSTM, attention layer LSTM, and CNN for wind speed and GHI forecasting. Each of these machine learning models is controlled by a variety of variables, including the type and number of hidden layers, the activation function, the optimization technique, the loss function, the epochs, and the learning rate. We identified the best parameter values from a range of potential parameter values so that the error values turn out to be the minimum. We compared these seven models based on the RMSE values and noted that no single model turned out to be globally best across time and space. For example, we have observed that the ANN model has the least RMSE values in monthly wind speed forecasting for only two study sites, whereas, the ANN model has the least error values in monthly GHI forecasting across all four study sites. We anticipated that this behavior may be attributed to the model sensitivity in learning from varying data size, inherent stochasticity,

seasonal and non-linear variability in the dataset. We have provided relevant tables and figures to summarize the comparative performance of the studied machine learning models. In addition, we performed residual analysis for model validation. Therefore, the research objective 3 is successfully accomplished through the implementation of seven state of the art machine learning models for hourly, daily, weekly, and monthly forecasting of renewable energy resources at four study sites in India.

6.1.4 Research Objective 4: To explore hybrid setups for renewable energy forecasting and to compare their efficacy with time series and machine learning models

To achieve the fourth objective, in Chapter 5, we implemented few hybrid models in renewable energy data. Specifically, we considered ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM hybrid models which are developed through time series ARIMA model and three machine learning models, namely ANN, CNN, and LSTM. The methodology had two connected steps: first, implementing an ARIMA model to analyze the linear part of the data and then, implementing a neural network (ANN, CNN, and LSTM) to model the residuals of the ARIMA. Based on the RMSE values corresponding to daily, weekly, and monthly forecasting, we noted that the ARIMA-ANN model has the best performance in wind speed forecasting across time and space, whereas in case of GHI forecasting, the ARIMA-ANN has the best performance except one location. Furthermore, in order to assess the efficacy of these hybrid models in comparison to the corresponding standalone models, we compared their RMSE values and residual characteristics. We noticed that except a few instances, none of the hybrid models provided lesser RMSE than that of the standalone models. Nevertheless, the Jarque-Bera test suggests that hybrid models enable a significant improvement towards the Gaussian behavior of residuals. However, the ineffective performance of hybrid models does not necessarily persuade against the importance of hybrid models in renewable energy forecasting since the present study has focused on a specific setup among several available setups, such as decomposition based hybrid models, homogeneous hybrid models, and other heterogeneous models. Towards the end of Chapter 5, we provided a comprehensive summary of all forecasting models studied in the thesis. Therefore, the research objective 4 has been successfully accomplished through the implementation of hybrid models along with an extensive comparison among all fifteen implemented models.

6.2 Major Findings of the Thesis

The major findings of the thesis are highlighted below.

1. The time series of wind speed and GHI data exhibit the presence of long term yearly seasonal pattern.

2. Among the five implemented probability distributions, the exponentiated Weibull is deemed to be the best fit model for both wind speed and GHI data.

3. Generally, both wind speed and GHI data reveal stationary behavior. As a consequence, in each case, the implemented time series ARIMA model turns out to be a simple ARMA model.

4. Among time series methods, the SARIMA model consistently produces the best performance for monthly forecasting, whereas for the weekly and daily forecasting, the WS-ARIMA model yields significant RMSE reduction. For hourly forecasting, based on three years (2012–2014) of data, both WS-ARIMA and ARIMA have comparable performance. Based on this, the present study brings out the importance of WS-ARIMA in renewable energy forecasting.

5. Among seven implemented machine learning models (i.e., SVR, ANN, CNN, and four variants of LSTM), based on the RMSE values, no single model appears to be globally best across time and space. This behavior may be attributed to varying data size, inherent stochasticity, seasonal and non-linear variability in the data.

6. In order to examine the linear and non-linear components of data simultaneously, we studied three hybrid models in Chapter 5. Based on RMSE values, except a few instances, none of the hybrid models provides better results than that of the standalone models. However, the residuals corresponding to hybrid models have desirable characteristics compared to standalone models.

6.3 Contributions through This Research

The present research has the following contributions towards a better understanding of renewable energy forecasting using statistics and machine learning in the Indian region.

1. The stationary behavior of wind speed and GHI data in Indian region, as observed here, is noteworthy for the renewable energy community to better analyze relevant data characteristics.

2. The best fit distributions presented in this study, highlighting the efficacy of the exponentiated Weibull distribution, are helpful for a number of practical applications, such as energy economics, energy trading, and pattern recognition and classification.

3. The effectiveness of the WS-ARIMA model in wind speed and GHI forecasting leads to reliable resource management, demand and supply analysis, and electric grid and market operations.

4. The optimal model parameters presented in this thesis may be used as benchmark to initiate other forecasting techniques, such as dynamic linear models, decomposition based hybrid models, and satellite image based space-time energy prediction.

5. Finally, the list of the best three ranked models among 15 implemented forecasting models provides a generic guideline to the users on choice of the best fit forecasting technique at a desired time horizon. Moreover, the uncertainty and bias estimates for the relevant forecasts are useful for risk assessment and decision making in the renewable energy sector.

6.4 Future Scope of the Present Research Work

A number of future researches that can be developed and/or integrated from the present investigation are mentioned below.

1. **Study of mixed probability distributions for modeling wind speed and GHI data**: We observed that both wind speed and GHI exhibit a complex and non-Gaussian distribution, depending on the position of Sun, atmospheric conditions, cloud cover, shading, and other factors. The GHI data usually have a high frequency of zero values, a peak near the maximum possible value, and a skewed distribution. Therefore, conventional probability distributions are inadequate in capturing data patterns. In this regard, the mixture of two or more probability distributions, such as Weibull-extreme value distribution, Weibull-lognormal, and gamma-Weibull may be employed for a reliable estimate of mean, variance, quartiles, and other statistical measures of data variables. As a consequence, use of mixed distributions leads to better assessment of energy availability, variability, uncertainty, and renewable energy potential at different target locations and timescales [62, 92].

2. **Study of dynamic linear models for wind speed and GHI forecasting**: Among various state space models, dynamic linear models (DLMs) are one of the most popular models due to their explainability and ability to incorporate regressors with dynamic coefficients.

They offer a very generic framework to analyze time series data with dynamic and non-linear features. Therefore, the DLMs can incorporate various types of information in wind speed and GHI data, such as trend, seasonality, cyclic behavior, and noise into the model structure. In addition, the DLMs can effectively address the structural changes in the renewable energy data [101, 122].

3. **Statistical methods for hourly forecasting**: We noted that the short term hourly data exhibits highly complex and stochastic behavior due to various climatic factors. Therefore, hourly forecasting of wind speed and GHI through classical time series methods is a challenging task. In light of this, data preprocessing based on statistical differences and similarities in an hourly manner or finding the optimal window length in the WS-ARIMA model may provide improved accuracy at shorter timescales [108].

4. **Decomposition based forecasting techniques**: Due to data characteristics, such as high seasonal fluctuations, trend, and noise, the performance of standalone forecasting models is limited. To overcome this issue, the decomposition based forecasting methods are generally recommended. These methods consist of two subsequent steps: first decomposition and then prediction. In the decomposition step, the original data are decomposed through wavelet transform, empirical mode decomposition, ensemble empirical mode decomposition, or variational mode decomposition into several components that represent different characteristics of the data, such as trend, seasonality, cycle, and noise. In the prediction step, each component is predicted separately using suitable models, such as neural networks, regression, or time series models. Thereafter, the predicted components are combined to obtain the final forecast [56, 103, 106, 159, 163].

5. **Homogeneous and heterogeneous hybrid models based on time series and artificial intelligence techniques**: The hybrid methods integrate the benefits of many independent models. For this, heterogeneous models (e.g., ARIMA-ANN, ARIMA-CNN, and ARIMA-LSTM), such as linear and non-linear models [11, 72, 138] or homogeneous models (e.g., SVR-ANN, LSTM-CNN, and CNN-LSTM), such as neural networks with various configurations [36, 55, 63, 76] may be combined to form hybrid models. Further, there are several ways of combining the results from standalone base models. For example, the study by Temur et al. [138] combines the error values from both ARIMA and LSTM models (calculated and through normalization process) through the weights assigned to the standalone models, whereas the present thesis directly combines the predictions of ARIMA and residuals processed through machine learning models. Thus, the exploration of various hybrid setups is required in order to appraise the best fit model in renewable energy prediction.

6. **Multivariate studies**: Although we noted that wind speed and GHI have high correlation with temperature, pressure, and other environmental factors, the present thesis considers the implementation of univariate forecasting models. Therefore, a forecasting method that uses multiple variables, such as temperature, humidity, pressure or wind direction may account for the correlation and interaction among different meteorological factors, resulting reliable forecasts of wind speed and GHI data [44, 73, 132].

7. **Post-processing techniques for validation**: The present thesis uses RMSE and MAPE error metrics which provide information on the overall collective error rather than the instantaneous error of the forecast. The information of instantaneous error may be more relevant from an operator point of view. Regarding this, we performed a preliminary residual analysis through residual plots, histograms, and P-P plots of the standardized residuals. However, there are explicit post-processing techniques which consider bias, accuracy, uncertainty, reliability, and resolution to validate the results of the forecasting models (Chapter 9 in [112]). Thus, a dedicated post-processing analysis will be helpful in producing accurate probabilistic forecasts by correcting the systematic errors [143].

8. **Spatio temporal analysis**: In spite of the fact that the information of spatio temporal characterization is important for several practical applications, such as potential location identification, cost effective analysis, and energy optimization, the present thesis has provided insights to the temporal variation of the underlying renewable energy process. Therefore, the future work may consider the implementation of a few spatio temporal techniques, such as Kriging, empirical orthogonal function, and variogram analysis [71].

9. **Satellite data based renewable energy prediction**: The lack of sufficient and reliable data sources, especially for offshore wind farms or remote areas, limit the availability and quality of input data for forecasting models. It could be advantageous to explore and use satellite data or remote sensing techniques to obtain wind speed and GHI data for forecasting purposes across space and time [37, 93, 130].

10. **Integration of forecasting, optimization, and intelligent decision making models**: As the developers and users in the renewable energy community often come from interdisciplinary domains, the integration of various contributors related to forecasting, optimization, and management offers a secure, affordable, and green energy to each and every relevant application.

Bibliography

[1] H. Abbasimehr and R. Paki, "Improving time series forecasting using LSTM and attention models," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 1–19, 2022.

[2] M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," in *2015 North American Power Symposium (NAPS)*, IEEE, 2015, pp. 1–5.

[3] R. Adhikari and R. K. Agrawal, "An introductory study on time series modeling and forecasting," *arXiv preprint arXiv:1302.6613*, 2013.

[4] K. Aimeur, L. S. Saoud, and R. Ghorbani, "Short-term solar irradiance forecasting and photovoltaic system management using octonion neural networks," *Applied Solar Energy*, vol. 56, pp. 219–226, 2020.

[5] M. M. Alayat, Y. Kassem, and H. Camur, "Assessment of wind energy potential as a power generation source: A case study of eight selected locations in northern Cyprus," *Energies*, vol. 11, no. 10, pp. 2697–2719, 2018.

[6] D. Alberg and M. Last, "Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms," *Vietnam Journal of Computer Science*, vol. 5, pp. 241–249, 2018.

[7] M. H. Alsharif, M. K. Younes, and J. Kim, "Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea," *Symmetry*, vol. 11, p. 240, 2019.

[8] D. R. Anderson, D. J. Sweeney, T. A. Williams, J. D. Camm, and J. J. Cochran, *Statistics for Business and Economics*. Cengage Learning, 2016.

[9] P. Arora and A. Balyan, "Comparative analysis of LSTM, encoder-decoder and GRU models for stock price prediction," in *Computational Intelligence: Select Proceedings of InCITe 2022*, Springer, 2023, pp. 399–410.

[10] S. Atique, S. Noureen, V. Roy, V. Subburaj, S. Bayne, and J. Macfie, "Forecasting of total daily solar energy generation using ARIMA: A case study," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2019, pp. 114–119.

[11] C. N. Babu and B. E. Reddy, "A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data," *Applied Soft Computing*, vol. 23, pp. 27–38, 2014.

[12] V. Bali, A. Kumar, and S. Gangwar, "Deep learning based wind speed forecasting- a review," in *2019 9th International Conference on Cloud Computing, Data Science and Engineering (Confluence), IEEE*, 2019, pp. 426–431.

[13] M. Bebbington, C. D. Lai, and R. Zitikis, "Modeling human mortality using mixtures of bathtub shaped failure distributions," *Journal of Theoretical Biology*, vol. 245, no. 3, pp. 528–538, 2007.

[14] B. Belmahdi, M. Louzazni, and A. El Bouardi, "One month-ahead forecasting of mean daily global solar radiation using time series models," *Optik*, vol. 219, p. 165 207, 2020.

[15] M. Bhaskar, A. Jain, and N. V. Srinath, "Wind speed forecasting: Present status," in *2010 International Conference on Power System Technology*, 2010, pp. 1–6.

[16] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, 1992, pp. 144–152.

[17] M. Bouzerdoum, A. Mellit, and A. M. Pavan, "A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant," *Solar Energy*, vol. 98, pp. 226–235, 2013.

[18] B. Brahma and R. Wadhvani, "Solar irradiance forecasting based on deep learning methodologies and multi-site data," *Symmetry*, vol. 12, no. 11, p. 1830, 2020.

[19] J. Brownlee, "Encoder-decoder long short-term memory networks," *Machine Learning Mastery*, 2017.

[20] U. C. Buyuksahin and S. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing*, vol. 361, pp. 151–163, 2019.

[21] E. Cadenas and W. Rivera, "Wind speed forecasting in the south coast of Oaxaca, Mexico," *Renewable Energy*, vol. 32, pp. 2116–2128, 2007.

[22] E. Cadenas, W. Rivera, R. Campos-Amezcua, and C. Heard, "Wind speed prediction using a univariate ARIMA Model and a multivariate NARX model," *Energies*, vol. 9, pp. 1–15, 2016.

[23] Y. P. Chaubey, "Modeling of distribution for excess-of-loss insurance data," 2007.

[24] T. Chen, G. Chen, W. Chen, S. Hou, Y. Zheng, and H. He, "Application of decoupled ARMA model to modal identification of linear time-varying system based on the ICA and assumption of 'short-time linearly varying'," *Journal of Sound and Vibration*, vol. 499, p. 115 997, 2021.

[25] D. E. Choe, G. Talor, and C. Kim, "Prediction of wind speed, potential wind power, and the associated uncertainties for offshore wind farm using deep learning," in *ASME Power Conference*, American Society of Mechanical Engineers, vol. 83747, 2020.

[26] M. Claesen, F. De Smet, J. A. Suykens, and B. De Moor, "Fast prediction with SVM models containing RBF kernels," *arXiv preprint arXiv:1403.0736*, 2014.

[27] *Contribution of fossil fuels in energy production*, https://www.iea.org/data-and-statistics/charts/share-of-cumulative-power-capacity-by-technology-2010-2027, [accessed last in February, 2023].

[28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[29] A. R. Daniel and A. A. Chen, "Stochastic simulation and forecasting of hourly average wind speed sequences in Jamaica," *Solar Energy*, vol. 46, pp. 1–11, 1991.

[30] K. Dineva and T. Atanasova, "Systematic look at machine learning algorithms–advantages, disadvantages and practical applications," *International Multidisciplinary Scientific GeoConference: SGEM*, vol. 20, no. 2.1, pp. 317–324, 2020.

[31] P. Ditthakit, S. Pinthong, N. Salaeh, J. Weekaew, T. T. Tran, and Q. B. Pham, "Comparative study of machine learning methods and GR2M model for monthly runoff prediction," *Ain Shams Engineering Journal*, vol. 14, no. 4, p. 101 941, 2023.

[32] H. Dong, X. Guo, H. Reichgelt, and R. Hu, "Predictive power of ARIMA models in forecasting equity returns: A sliding window method," *Journal of Asset Management*, vol. 21, pp. 549–566, 2020.

[33] Y. Duan, H. Wang, M. Wei, L. Tan, and T. Yue, "Application of ARIMA-RTS optimal smoothing algorithm in gas well production prediction," *Petroleum*, vol. 8, no. 2, pp. 270–277, 2022.

[34] A. K. Dubey, A. Kumar, V. Garcia-Diaz, A. K. Sharma, and K. Kanhaiya, "Study and analysis of SARIMA and LSTM in forecasting time series data," *Sustainable Energy Technologies and Assessments*, vol. 47, p. 101 474, 2021.

[35] S. Franco, V. R. Mandla, and K. R. Mohan Rao, "Estimation of bright roof areas for large scale solar PV applications to meet the power demand of megacity hyderabad," *Applied Solar Energy*, vol. 52, pp. 284–289, 2016.

[36] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, "Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks," *Renewable Energy*, vol. 162, pp. 1665–1683, 2020.

[37] P. M. Garniwa, R. A. Rajagukguk, R. Kamil, and H. Lee, "Intraday forecast of global horizontal irradiance using optical flow method and long short-term memory model," *Solar Energy*, vol. 252, pp. 234–251, 2023.

[38] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep learning for solar power forecasting—an approach using AutoEncoder and LSTM Neural Networks," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2016, pp. 002 858–002 865.

[39] M. Ghofrani and M. Alolayan, *Time Series and Renewable Energy Forecasting*. IntechOpen, 2018, vol. 10.

[40] *Global environmental pollution level*, https://www.iea.org/news/global-co2-emissions-rebounded-to-their-highest-level-in-history-in-2021, [accessed last in September, 2023].

[41] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term non-parametric probabilistic forecasting of renewable energy generation with application to solar energy," *IEEE Transactions Power System*, pp. 3850–3863, 2016.

[42] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[43] G. Guariso, G. Nunnari, and M. Sangiorgio, "Multi-step solar irradiance forecasting and domain adaptation of deep neural networks," *Energies*, vol. 13, no. 15, p. 3987, 2020.

[44] P. Gupta and R. Singh, "Combining a deep learning model with multivariate empirical mode decomposition for hourly global horizontal irradiance forecasting," *Renewable Energy*, vol. 206, pp. 908–927, 2023.

[45] R. C. Gupta, P. L. Gupta, and R. D. Gupta, "Modeling failure time data by Lehman alternatives," *Communications in Statistics-Theory and Methods*, vol. 27, no. 4, pp. 887–904, 1998.

[46] R. D. Gupta and D. Kundu, "Theory & methods: Generalized exponential distributions," *Australian & New Zealand Journal of Statistics*, vol. 41, no. 2, pp. 173–188, 1999.

[47] K. W. Hipel and A. I. McLeod, *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, 1994.

[48] R. V. Hogg and A. T. Craig, "Introduction to Mathematical Statistics. (5th edition)," *Englewood Hills, New Jersey*, 1995.

[49] H. S. Hota, R. Handa, and A. K. Shrivas, "Time series data prediction using sliding window based RBF neural network," *International Journal of Computational Intelligence Research*, vol. 13, pp. 1145–1156, 2017.

[50] J. Huang and H. Liu, "A hybrid decomposition-boosting model for short-term multi-step solar radiation forecasting with NARX neural network," *Journal of Central South University*, vol. 28, no. 2, pp. 507–526, 2021.

[51] J. Huang, M. Korolkiewicz, M. Agrawal, and J. Boland, "Forecasting solar radiation on an hourly time scale using a coupled autoregressive and dynamical system (CARDS) model," *Solar Energy*, vol. 87, pp. 136–149, 2013.

[52] X. Huang, J. Wang, and B. Huang, "Two novel hybrid linear and nonlinear models for wind speed forecasting," *Energy Conversion and Management*, vol. 238, p. 114 162, 2021.

[53] S. Impram, S. V. Nese, and B. Oral, "Challenges of renewable energy penetration on power system flexibility: A survey," *Energy Strategy Reviews*, vol. 31, p. 100 539, 2020.

[54] E. Isaksson and M. Karpe Conde, *Solar Power Forecasting with Machine Learning Techniques*, 2018.

[55] S. M. J. Jalali, S. Ahmadian, A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "Automated deep CNN-LSTM architecture design for solar irradiance forecasting," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 54–65, 2021.

[56] K. U. Jaseena and B. C. Kovoor, "Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks," *Energy Conversion and Management*, vol. 234, p. 113 944, 2021.

[57] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, vol. 85, no. 5, pp. 808–817, 2011.

[58] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 2*. John Wiley & Sons, 1995, vol. 289.

[59] I. L. Kane and F. Yusof, "Assessment of risk of rainfall events with a hybrid of ARFIMA-GARCH," *Modern Applied Science*, vol. 7, no. 12, p. 78, 2013.

[60] O. Karakus, E. E. Kuruoglu, and M. A. Altinkaya, "One-day ahead wind speed/power prediction based on polynomial autoregressive model," *IET Renewable Power Generation*, vol. 11, pp. 1430–1439, 2017.

[61] A. K. Khamees, A. Y. Abdelaziz, M. R. Eskaros, M. A. Attia, and A. O. Badr, "The mixture of probability distribution functions for wind and photovoltaic power systems using a metaheuristic method," *Processes*, vol. 10, no. 11, p. 2446, 2022.

[62] R. Kollu, S. R. Rayapudi, S. Narasimham, and K. M. Pakkurthi, "Mixture probability distribution functions to model wind speed distributions," *International Journal of Energy and Environmental Engineering*, vol. 3, pp. 1–10, 2012.

[63] P. Kumari and D. Toshniwal, "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance," *Journal of Cleaner Production*, vol. 279, p. 123 285, 2021.

[64] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[65] J. Li, J. K. Ward, J. Tong, L. Collins, and G. Platt, "Machine learning for solar irradiance forecasting of photovoltaic system," *Renewable Energy*, vol. 90, pp. 542–553, 2016.

[66] X. Li and M. Yin, "Modified cuckoo search algorithm with self adaptive parameter method," *Information Sciences*, vol. 298, pp. 80–97, 2015.

[67] W. H. Lin, P. Wang, K. M. Chao, H. C. Lin, Z. Y. Yang, and Y. H. Lai, "Wind power forecasting with deep learning networks: Time-series forecasting," *Applied Sciences*, vol. 11, no. 21, p. 10 335, 2021.

[68] H. Liu, X. W. Mi, and Y. F. Li, "Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network," *Energy Conversion and Management*, vol. 156, pp. 498–514, 2018.

[69] H. Liu, H. Q. Tian, X. F. Liang, and Y. F. Li, "Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks," *Applied Energy*, vol. 157, pp. 183–194, 2015.

[70] H. Liu, H. Tian, and Y. Li, "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction," *Applied Energy*, vol. 98, pp. 415–424, 2012.

[71] J. Liu, G. Ren, J. Wan, Y. Guo, and D. Yu, "Variogram time-series analysis of wind speed," *Renewable Energy*, vol. 99, pp. 483–491, 2016.

[72] M. D. Liu, L. Ding, and Y. L. Bai, "Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction," *Energy Conversion and Management*, vol. 233, p. 113 917, 2021.

[73] S. X. Lv and L. Wang, "Multivariate wind speed forecasting based on multi-objective feature selection approach and hybrid deep learning model," *Energy*, vol. 263, p. 126 100, 2023.

[74] S. Malakar, S. Goswami, B. Ganguli, A. Chakrabarti, S. S. Roy, K Boopathi, and A. Rangaraj, "Designing a long short-term network for short-term forecasting of global horizontal irradiance," *SN Applied Sciences*, vol. 3, no. 4, pp. 1–15, 2021.

[75] I. Maqsood, M. R. Khan, and A. Abraham, "An ensemble of neural networks for weather forecasting," *Neural Computing & Applications*, vol. 13, pp. 112–122, 2004.

[76] F. P. Marinho, P. A. Rocha, A. R. Neto, and F. D. Bezerra, "Short-term solar irradiance forecasting using CNN-1D, LSTM, and CNN-LSTM deep neural networks: A case study with the folsom (USA) dataset," *Journal of Solar Energy Engineering*, vol. 145, no. 4, p. 041 002, 2023.

[77] L. Martin, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, "Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning," *Solar Energy*, vol. 84, no. 10, pp. 1772–1781, 2010.

[78] N. Masseran, "Evaluating wind power density models and their statistical properties," *Energy*, vol. 84, pp. 533–541, 2015.

[79] H. Mehdi, Z. Pooranian, and P. G. V. Naranjo, "Cloud traffic prediction based on fuzzy ARIMA model with low dependence on historical data," *Transactions on Emerging Telecommunications Technologies*, vol. 33, pp. 1–17, 2022.

[80] E. Melikhova and A. Rogachev, "Computer optimization of ANN hyperparameters for retrospective information processing," in *International School on Neural Networks, Initiated by IIASS and EMFCSC*, Springer, 2022, pp. 723–730.

[81] M. Mishra and M. Srivastava, "A view of artificial neural network," in *2014 International Conference on Advances in Engineering & Technology Research (ICAETR-2014)*, IEEE, 2014, pp. 1–3.

[82] Z. E. Mohamed, "Using the artificial neural networks for prediction and validating solar radiation," *Journal of the Egyptian Mathematical Society*, vol. 27, pp. 1–13, 2019.

[83] J. Moon, M. B. Hossain, and K. H. Chon, "AR and ARMA model order selection for time-series modeling with ImageNet classification," *Signal Processing*, vol. 183, p. 108 026, 2021.

[84] G. S. Mudholkar and D. K. Srivastava, "Exponentiated Weibull family for analyzing bathtub failure rate data," *IEEE Transactions on Reliability*, vol. 42, no. 2, pp. 299–302, 1993.

[85] G. S. Mudholkar, D. K. Srivastava, and M. Freimer, "The exponentiated Weibull family: A reanalysis of the bus-motor failure data," *Technometrics*, vol. 37, no. 4, pp. 436–445, 1995.

[86] M. Z. Mukaram and F. Yusof, "Solar radiation forecast using hybrid SARIMA and ANN model: A case study at several locations in Peninsular Malaysia," *Malaysian Journal of Fundamental and Applied Sciences Special Issue on Some Advances in Industrial and Applied Mathematics*, vol. 13, pp. 346–350, 2017.

[87] B. P. Mukhoty, V. Maurya, and S. K. Shukla, "Sequence to sequence deep learning models for solar irradiation forecasting," in *2019 IEEE Milan PowerTech*, IEEE, 2019, pp. 1–6.

[88] D. P. Murthy, M. Xie, and R. Jiang, *Weibull Models*. John Wiley & Sons, 2004.

[89] H. Muthiah, U. Saadah, and A. Efendi, "Support vector regression (SVR) model for seasonal time series data," in *The Second Asia Pacific International Conference on Industrial Engineering and Operations Management*, 2021, pp. 3191–3200.

[90] Y Nagaraja, T Devaraju, M. V. Kumar, and S. Madichetty, "A survey on wind energy, load and price forecasting: (Forecasting methods)," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 783–788.

[91] K. R. Nair, V Vanitha, and M Jisma, "Forecasting of wind speed using ANN, ARIMA and hybrid models," in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, IEEE, 2017, pp. 170–175.

[92] N. Natarajan, M. Vasudevan, and S. Rehman, "Evaluation of suitability of wind speed probability distribution models: A case study from Tamil Nadu, India," *Environmental Science and Pollution Research*, vol. 29, no. 57, pp. 85 855–85 868, 2022.

[93] M. M. Nezhad, A Heydari, E Pirshayan, D Groppi, and D. A. Garcia, "A novel forecasting model for wind speed assessment using sentinel family satellites images and machine learning method," *Renewable Energy*, vol. 179, pp. 2198–2211, 2021.

[94] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Systems with Applications*, vol. 105, pp. 233–261, 2018.

[95] T. B. Ouarda, C. Charron, J.-Y. Shin, P. R. Marpu, A. H. Al-Mandoos, M. H. Al-Tamimi, H. Ghedira, and T. Al Hosary, "Probability distributions of wind speed in the UAE," *Energy conversion and management*, vol. 93, pp. 414–434, 2015.

[96] M. Pal, M. M. Ali, and J. Woo, "Exponentiated Weibull distribution," *Statistica*, vol. 66, no. 2, pp. 139–147, 2006.

[97] S. Pasari and O. Dikshit, "Impact of three-parameter Weibull models in probabilistic assessment of earthquake hazards," *Pure and Applied Geophysics*, vol. 171, pp. 1251–1281, 2014.

[98] S. Pasari and V. S. S. K. Nandigama, "Statistical modeling of solar energy," in *Enhancing Future Skills and Entrepreneurship: 3rd Indo-German Conference on Sustainability in Engineering*, Springer International Publishing Cham, 2020, pp. 157–165.

[99] S. Pasari and A. Shah, "Time series auto-regressive integrated moving average model for renewable energy forecasting," *Sustainable Production, Life Cycle Engineering and Management*, 2020.

[100] S. Pasari, A. Shah, and U. Sirpurkar, "Wind energy prediction using artificial neural networks," in *Enhancing Future Skills and Entrepreneurship, Springer*, 2020, pp. 101–107.

[101] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic linear models with R*. Springer Science & Business Media, 2009.

[102] M. Pirani, P. Thakkar, P. Jivrani, M. H. Bohara, and D. Garg, "A comparative analysis of ARIMA, GRU, LSTM and BiLSTM on financial time series forecasting," in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, IEEE, 2022, pp. 1–6.

[103] V. Prema and K. U. Rao, "Time series decomposition model for accurate wind speed forecast," *Renewables: Wind, Water, and Solar*, vol. 2, no. 1, pp. 1–11, 2015.

[104] T. Pukkila, S. Koreisha, and A. Kallinen, "The identification of ARMA models," *Biometrika*, vol. 77, no. 3, pp. 537–548, 1990.

[105] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, 2018.

[106] Z. Qu, W. Mao, K. Zhang, W. Zhang, and Z. Li, "Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network," *Renewable Energy*, vol. 133, pp. 919–929, 2019.

[107] A. Rabehi, A. Rabehi, and M. Guermoui, "Evaluation of different models for global solar radiation components assessment," *Applied Solar Energy*, vol. 57, pp. 81–92, 2021.

[108] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, pp. 342–349, 2009.

[109] G. Reikard and C. Hansen, "Forecasting solar irradiance at short horizons: Frequency and time domain models," *Renewable Energy*, vol. 135, pp. 1270–1290, 2019.

[110] G. Reikard, S. E. Haupt, and T. Jensen, "Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models," *Renewable Energy*, vol. 112, pp. 474–485, 2017.

[111] S. I. Resnick, *Heavy-tail phenomena: Probabilistic and statistical modeling*. Springer Science & Business Media, 2007.

[112] M. Rivero, A. Reyes, M. Escalante, and O. Probst, "Forecasting of renewable energy generation for grid integration," *Transforming the Grid Towards Fully Renewable Energy*, pp. 1–39, 2018.

[113] T. J. J. Ryan, *LSTMs explained: A complete, technically accurate, conceptual guide with keras*, 2021.

[114] H. Saima, J. Jaafar, S. Belhaouari, and T. Jillani, "Intelligent methods for weather forecasting: A review," in *2011 National Postgraduate Conference*, 2011, pp. 1–6.

[115] E. S. Salami, M. Ehetshami, A. Karimi Jashni, M. Salari, S. Nikbakht Sheibani, and A. Ehteshami, "A mathematical method and artificial neural network modeling to simulate osmosis membrane's performance," *Modeling Earth Systems and Environment*, vol. 2, pp. 1–11, 2016.

[116] M. Santhosh, C. Venkaiah, and D. M. V. Kumar, "Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review," *Engineering Reports*, vol. 2, pp. 1–20, 2020.

[117] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press, 2002.

[118] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The national solar radiation data base (NSRDB)," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 51–60, 2018.

[119] H. Sharadga, S. Hajimirza, and R. S. Balog, "Time series forecasting of solar power generation for large-scale photovoltaic plants," *Renewable Energy*, vol. 150, pp. 797–807, 2020.

[120] *Share of renewables in energy generation*, https://cea.nic.in, [accessed last in April, 2023].

[121] S. Sheoran, R. Badekar, S. Pasari, and R. Kulshrestha, "Wind speed forecasting using time series methods: A case study," . *In Chamola BP, Kumari P, Kaur L (eds) Emerging Advancements in Mathematical Sciences, Nova publishing*, pp. 125–133, 2022.

[122] S. Sheoran and S. Pasari, "Efficacy and application of the window-sliding ARIMA for daily and weekly wind speed forecasting," *Journal of Renewable and Sustainable Energy*, vol. 14, no. 5, 2022.

[123] S. Sheoran, S. Shukla, S. Pasari, R. S. Singh, and R. Kulshrestha, "Wind speed forecasting at different time scales using time series and machine learning models," *Applied Solar Energy*, vol. 58, no. 5, pp. 708–721, 2022.

[124] S. Sheoran, R. S. Singh, S. Pasari, and R. Kulshrestha, "Forecasting of solar irradiances using time series and machine learning models: A case study from india," *Applied Solar Energy*, vol. 58, no. 1, pp. 137–151, 2022.

[125] J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, pp. 1064–1069, 2012.

[126] A. ShobanaDevi, G. Maragatham, M. R. Prabu, and K. Boopathi, "Short-term wind power forecasting using RLSTM," *International Journal of Renewable Energy Research*, vol. 11, no. 1, pp. 392–406, 2021.

[127] S. Shukla, R. Ramaprasad, S. Pasari, and S. Sheoran, "Statistical analysis and forecasting of wind speed," in *2022 4th International Conference on Energy, Power and Environment (ICEPE)*, IEEE, 2022, pp. 1–6.

[128] S. Shukla, S. Sheoran, and S. Pasari, "Exploration of solar irradiance in thar desert using time series model," *Applied Solar Energy*, vol. 58, no. 6, pp. 876–888, 2022.

[129] O. B. Shukur and M. H. Lee, "Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA," *Renewable Energy*, vol. 76, pp. 637–647, 2015.

[130] Z. Si, Y. Yu, M. Yang, and P. Li, "Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks," *IEEE Transactions on Industry Applications*, vol. 57, no. 1, pp. 5–16, 2020.

[131] S. Siami Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *2019 IEEE International Conference on Big Data*, IEEE, 2019, pp. 3285–3292.

[132] M. Sibtain, H. Bashir, M. Nawaz, S. Hameed, M. I. Azam, X. Li, T. Abbas, and S. Saleem, "A multivariate ultra-short-term wind speed forecasting model by employing multistage signal decomposition approaches and a deep learning network," *Energy Conversion and Management*, vol. 263, p. 115 703, 2022.

[133] S. Singh, A. Mohapatra, *et al.*, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renewable Energy*, vol. 136, pp. 758–768, 2019.

[134] *Solar power in india*, https://en.wikipedia.org/wiki/Solar_power_in_India, [accessed last in August, 2023].

[135] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," *North American Power Symposium 2010*, pp. 1–8, 2010.

[136] Z. Tang and P. A. Fishwick, "Feedforward neural nets as models for time series forecasting," *ORSA Journal on Computing*, vol. 5, no. 4, pp. 374–385, 1993.

[137] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[138] A. S. Temur and S. Yildiz, "Comparison of forecasting performance of ARIMA, LSTM and hybrid models for the sales volume budget of a manufacturing enterprise," *Istanbul Business Research*, vol. 50, no. 1, pp. 15–46, 2021.

[139] V. L. Tran, "Stochastic models of solar radiation processes," *HAL*, vol. 2013, 2013.

[140] A. Ucar and F. Balo, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, pp. 1912–1928, 2018.

[141] N. Vakitbilir, A. Hilal, and C. Direkoglu, "Hybrid deep learning models for multivariate forecasting of global horizontal irradiation," *Neural Computing and Applications*, vol. 34, no. 10, pp. 8005–8026, 2022.

[142] S. M. Valdivia-Bautista, J. A. Dominguez-Navarro, M. Perez-Cisneros, C. J. Vega-Gomez, and B. Castillo-Tellez, "Artificial intelligence in wind speed forecasting: A review," *Energies*, vol. 16, no. 5, p. 2457, 2023.

[143] S. Vannitsem, J. B. Bremnes, J. Demaeyer, G. R. Evans, J. Flowerdew, S. Hemri, S. Lerch, N. Roberts, S. Theis, A. Atencia, *et al.*, "Statistical postprocessing for weather forecasts–review, challenges and avenues in a big data world," *Bulletin of the American Meteorological Society*, pp. 1–44, 2020.

[144] V. N. Vapnik, "The nature of statistical learning theory," *New York: Spring-Verlag*, vol. 286, 1995.

[145] R. A. Verzijlbergh, L. J. De Vries, G. Dijkema, and P. Herder, "Institutional challenges caused by the integration of renewable energy sources in the European electricity sector," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 660–667, 2017.

[146] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[147] H. Wahedi, K. Wrona, M. Heltoft, S. Saleh, T. R. Knudsen, U. Bendixen, I. Nielsen, S. Saha, and G. S. Borup, "Improving accuracy of time series forecasting by applying an ARIMA-ANN hybrid model," in *IFIP International Conference on Advances in Production Management Systems*, Springer, 2022, pp. 3–10.

[148] J. Wang, Y. Song, F. Liu, and R. Hou, "Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models," *Renewable and Sustainable Energy Reviews*, pp. 960–981, 2016.

[149] M. Wang and K. Rennolls, "Tree diameter distribution modelling: Introducing the logit logistic distribution," *Canadian Journal of Forest Research*, vol. 35, no. 6, pp. 1305–1313, 2005.

[150] S. Wang, C. Li, and A. Lim, "Why are the ARIMA and SARIMA not sufficient," *arXiv preprint arXiv:1904.07632*, 2019.

[151] W Weibull, "A statistical theory of the strength of materials," *Royal Academy of Engeneering Science*, vol. 15, 1939.

[152] W. Weibull, "A statistical distribution function of wide applicability," *Journal of Applied Mechanics*, 1951.

[153] S. Wheelwright, S. Makridakis, and R. J. Hyndman, *Forecasting: Methods and Applications*. John Wiley & Sons, 1998.

[154] *Wind power in india*, https://en.wikipedia.org/wiki/Wind_power_in_India, [accessed last in August, 2023].

[155] *World energy data*, https://ourworldindata.org/energy-substitution-method, [accessed last in September, 2023].

[156] *World energy transactions outlook*, https://www.irena.org/publications/2021/Jun/World-Energy-Transitions-Outlook, [accessed last in March, 2023].

[157] L. Wu, J. Park, J. Choi, J. Cha, and K. Y. Lee, "A study on wind speed prediction using artificial neural network at Jeju island in Korea," in *2009 Transmission and Distribution Conference and Exposition: Asia and Pacific, IEEE*, 2009, pp. 1–4.

[158] Y. Xiao, J. Xiao, and S. Wang, "A hybrid forecasting model for non-stationary time series: An application to container throughput prediction," *International Journal of Knowledge and Systems Science (IJKSS)*, vol. 3, no. 2, pp. 67–82, 2012.

[159] J. Xie, H. Zhang, L. Liu, M. Li, and Y. Su, "Decomposition-based multistep sea wind speed forecasting using stacked gated recurrent unit improved by residual connections," *Complexity*, vol. 2021, pp. 1–14, 2021.

[160] A. K. Yadav and S. Chandel, "Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model," *Renewable Energy*, vol. 75, pp. 675–693, 2015.

[161] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining," *Solar Energy*, vol. 168, pp. 60–101, 2018.

[162] ——, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining," *Solar Energy*, vol. 168, pp. 60–101, 2018.

[163] H. Yao, Y. Tan, J. Hou, Y. Liu, X. Zhao, and X. Wang, "Short-term wind speed forecasting based on the EEMD-GS-GRU model," *Atmosphere*, vol. 14, no. 4, p. 697, 2023.

[164] H. Ye, B. Yang, Y. Han, Q. Li, J. Deng, and S. Tian, "Wind speed and power prediction approaches: Classifications, methodologies, and comments," *Frontiers in Energy Research*, vol. 10, p. 901 767, 2022.

[165] X. Ying, "An overview of overfitting and its solutions," in *Journal of Physics: Conference Series*, vol. 1168, 2019, p. 022 022.

[166] P. S. Yu, S. T. Chen, and I. F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3–4, pp. 704–716, 2006.

[167] Q. Yu, L. Jibin, and L. Jiang, "An improved ARIMA-based traffic anomaly detection algorithm for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 12, pp. 1–9, 2016.

[168] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[169] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2253–2262, 2015.

[170] D. X. Zhou, "Universality of deep convolutional neural networks," *Applied and Computational Harmonic Analysis*, vol. 48, no. 2, pp. 787–794, 2020.

List of Publications

Journal Publications

- S. Shukla, **S. Sheoran**, S. Pasari, "Exploration of solar irradiance in Thar desert using time series models", *Applied Solar Energy*, pp. 876–888, 2022 (Scopus).
- **S. Sheoran**, S. Shukla, S. Pasari, R. S. Singh, R. Kulshrestha, "Wind speed forecasting at different time scales using time series and machine learning models", *Applied Solar Energy*, pp. 708–721, 2022 (Scopus).
- **S. Sheoran**, S. Pasari, "Efficacy and application of the window-sliding ARIMA for daily and weekly wind speed forecasting", *Journal of Renewable and Sustainable Energy*, p. 053305, 2022 (SCIE) (AIP Publishing).
- **S. Sheoran**, R.S. Singh, S. Pasari, R. Kulshrestha, "Forecasting of solar irradiances using time series and machine learning models: A case study from India", *Applied Solar Energy*, pp. 137–151, 2022 (Scopus).
- **S. Sheoran** and S. Pasari, "Exploration of exponentiated family of probability distributions in pattern recognition of renewable energy data" (submitted).
- **S. Sheoran** and S. Pasari, S. Shukla, "Study of univariate and multivariate support vector regression in wind speed and GHI forecasting" (under preparation).
- **S. Sheoran** and S. Pasari, S. Shukla, "Multidisciplinary approaches to renewable energy forecasting in Indian region" (under preparation).

Conference Proceedings/Book Chapters

- S. Shukla, **S. Sheoran** and S. Pasari, "Prediction of Solar Energy using Time Series Methods," *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, pp. 72–76, 2022.
- **S. Sheoran**, S. Pasari and S. Shukla, "Application of Window Sliding ARIMA in Wind Speed and Solar Irradiance Forecasting," *2022 7th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, Mangalore, India, pp. 321–325, 2022.
- S. Shukla, Ramaprasad R, S. Pasari and **S. Sheoran**, "Statistical Analysis and Forecasting of Wind Speed," *2022 4th International Conference on Energy, Power and Environment (ICEPE)*, Shillong, India, pp. 1–6, 2022.

- K. Gupta, S. Shukla, S. Pasari and **S. Sheoran**, "Wind Speed Prediction Using Sentinel-1 OCN Products," *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, pp. 1145–1149, 2023.

- T. Agrawal, **S. Sheoran**, S. Pasari and S. Shukla, "Wind Speed Prediction with Spatio Temporal Correlation using Convolutional and Spiking Neural Networks," *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, pp. 197–200, 2022.

- A. Rathi, S. Pasari and **S. Sheoran**, "Live Sign Language Recognition: Using Convolution Neural Networks," *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, pp. 502–505, 2022.

- **S. Sheoran**, R. Bavdekar, S. Pasari, R. Kulshrestha, "Wind speed forecasting using time series methods: a case study," in Chamola BP, Kumari P, Kaur L (eds) *Emerging Advancements in Mathematical Sciences* (Nova publishing), pp. 125–133, 2022.

List of Attended Conferences/Workshops/Schools

Presented works in international conferences

1. Presented paper entitled "Forecasting of wind speed using time series methods" in International Conference on Recent Advances in Mathematical Sciences and its Applications (RAMSA-2021), organized by the Department of Mathematics, Jaypee Institute of Information Technology, Noida, December 02–04, 2021.

2. Presented seminar entitled "Residual analysis for time series modeling" in International Conference on Advances in Mechanics, Modelling, Computing and Statistics (ICAMMCS 2022), organized by the Department of Mathematics, BITS Pilani, Pilani Campus, Rajasthan, March 19–21, 2022.

3. Presented paper entitled "Application of window sliding ARIMA in wind speed and solar irradiance forecasting" in IEEE International Conference on Recent Advances and Innovations in Engineering (7th IEEE ICRAIE 2022), organized by National Institute of Technology Karnataka, Surathkal, December 1–3, 2022.

4. Presented paper entitled "Wind speed prediction with spatio temporal correlation using convolutional and spiking neural networks" in International Conference on Cloud Computing, Data Science and Engineering (Confluence 2022), organized by Amity University, Noida, January 17–28, 2022.

Attended conferences/schools

1. Attended Indian Women and Mathematics (IWM) Annual Conference 2019: Department of Mathematics, organized by Indian Institute of Technology Bombay, June 10–12, 2019.

2. Attended the Advanced Instructional School (AIS) - Stochastic Processes - level (1), organized by NISER, Bhubaneswar, June 17–July 12, 2019.

Brief Biography of the Supervisor

Prof. Sumanta Pasari is an Associate Professor in the Department of Mathematics at Birla Institute of Technology and Science, Pilani (Pilani Campus). He obtained his Ph.D. in Civil Engineering with a specialization in Geoinformatics from the Indian Institute of Technology Kanpur (IIT Kanpur) and also holds a Masters in Mathematics from the same institute. His research focuses on a variety of topics, including crustal deformation and active tectonics, statistical seismology, and renewable energy modeling. He has published numerous research articles in recognized journals and proceedings. Two of his Ph.D. students, Dr. Yogendra Sharma and Dr. Neha, recently graduated from his research group. Five Ph.D. scholars are conducting research on crustal deformation analysis, machine learning in geosciences, and renewable energy prediction under his supervision.

Brief Biography of the Co-Supervisor

Prof. Rakhee is a Professor in the Department of Mathematics at Birla Institute of Technology and Science, Pilani (Pilani Campus). She has completed her Ph.D. from the Centre for Information and Decision Sciences, Dr. B.R. Ambedkar University, Agra in 2003. She worked as a visiting faculty at the Institute of Engineering and Technology, Dr. B.R. Ambedkar University, Khandari Campus, during August 2003–June 2004. In July 2004, she joined Banasthali University, Rajasthan. She joined as an Assistant Professor in the Department of Mathematics at Birla Institute of Technology and Science, Pilani (Pilani Campus) in November 2008. Her research interests include the areas of applied probability, performance analysis of communication networks, inventory and supply chain management, circular economy, and renewable energy modeling. There are 39 research publications in refereed international/national journals/proceedings and a monograph to her credit. Prof. Rakhee received DST-DFG Bilateral Co-operation Fellowship, awarded by DFG, Germany to work with Prof. Raik Stolletz, University of Mannheim, Germany during December 1, 2015–January 15, 2016. She also has one patent (with Dr. Savita Kumari and Prof. Seema Verma, Banasthali University). She has participated in more than 40 international and national conferences in India and abroad and visited many reputed Universities/Institutes in Germany, Turkey, and Singapore. During her tenure at Banasthali University, she co-supervised Dr. Savita Kumari. At BITS Pilani, one scholar (Dr. Shruti) graduated from her research group. Currently, she is guiding five Ph.D. students (Ms. Sarita, Ms. Pooja, Mr. Vijaypal Poonia, Mr. Ajay Singh, and Ms. Pooja Yadav) at BITS Pilani. Prof. Rakhee is a member of many international and national bodies, like FIM, ORSI, ISPS, and ISMS. In addition, she is serving as an executive council member of Vijnana Parishad of India.

Brief Biography of the Candidate

Ms. Sarita graduated with B.Sc. (H) degree in Mathematics from the Department of Mathematics, Maharashi Dayanand University, Rohtak, Haryana in 2013, and post-graduated with M.Sc. degree in Mathematics from the Department of Mathematics, Indian Institute of Technology, Delhi in 2016. Currently, she is working towards a Ph.D. degree from Birla Institute of Technology and Science, Pilani Campus, Pilani. Her research interests lie in modeling of renewable energy resources such as wind speed and solar irradiance. She has 11 research publications in peer-reviewed journals and conference proceedings to her credit. She has qualified GATE in Mathematics and she is a recipient of the UGC NET-JRF. She has attended six international conferences and summer schools during her Ph.D.