Optimization of Environmentally Powered Wireless Sensor Networks for Efficient Energy Harvesting

A Thesis

submitted in partial fulfillment of the requirements for the award of the degree of

Doctor of Philosophy

in

Electronics and Communication Engineering Department

by

Amandeep Sharma
(Reg no: 901306003)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Thapar Institute of Engineering and Technology
Patiala-147001
Punjab, India
March 2019

Candidate Declaration

I hereby certify that the work which is being presented in the thesis entitled **Optimization of Environmentally Powered Wireless Sensor Networks for Efficient Energy Harvesting** in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy** and submitted to the institution is an authentic record of my own work carried out during the period **July 2013** to **March 2019** under the supervision of **Dr. Ajay Kakkar**. I have also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

The matter presented in this thesis has not been submitted elsewhere for the award of any other degree or diploma from any institution.

Date: _____

Amandeep Sharma
Candidate

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date: _____

Dr. Ajay Kakkar
Supervisor

.....dedicated to all young generation researchers

Abstract

Focusing on environmentally powered Wireless Sensor Networks (WSNs), this thesis studies optimized operation of individual sensor node in terms of average duty cycle. In particular, the focus lies in gaining high average duty cycle with high stability. To achieve this objective, an energy neutral approach based efficient power management system is introduced and investigated in different working conditions.

WSNs deployed in ad hoc manner comprise of numerous sensing nodes organised in a network for the sake of checking and balancing the environmental factors. Each node has sensing, computation, communication and locomotion capabilities but operates with limited battery life. Energy harvesting is a way of powering these WSNs by harvesting energy from the environment. Using harvesting energy as source, certain considerations are different from that battery operated networks. Nondeterministic energy availability with respect to time is the reason behind these differences which put a limit on the maximum rate at which energy can be used. Thus, power management is of prime importance in self-powered networks.

The thesis begins with development of efficient solar forecasting algorithm for accurate estimation of energy availability. Reliable knowledge of solar radiation is essential for informed design, deployment planning and optimal management of energy in rechargeable sensor networks. In the proposed work, an optimized Pro-Energy algorithm is developed using level and trend factors in time series analysis for future solar irradiance estimation. The performance of proposed algorithm has been compared with EWMA, WCMA, and Pro-Energy on the basis of prediction error. The problem of solar irradiance forecasting has been further addressed by machine learning methodologies over historical data set. In proposed work, forecasts have been done using FoBa, leapForward, spikeslab, Cubist and bagEarthGCV models. To achieve more precise and consistent forecast, four Statistical Ensemble (SE) approaches have been presented. To validate the effectiveness of these methodologies, a series of experimental evaluations have been presented in terms of forecast accuracy, correlation coefficient and Root Mean Square Error (RMSE). The *R* interface has been used as simulation platform for these evaluations.

Based on forecasted solar energy profiles, an integrated approach of energy assignment principles with adaptive duty cycling has been introduced to efficiently utilize the available energy. For this purpose, four factors of the system including energy generation rate, energy consumption rate, storage and controlled energy allotment to the sensor node have

been formulated in to a theoretical model. The dynamic programming has also been used for theoretical analysis of the system. The analytical models of solar powered wireless sensor node have been used to validate the effectiveness of proposed work. The extensive simulations have been conducted on real time solar energy profiles in terms of magnitude and stability of sensors average duty cycle. The experimental results shows that the proposed approach offers perpetual node operation with high energy efficiency.

Keywords: Wireless sensor networks, Energy harvesting, Solar forecasting, Photovolatic (PV) systems, Average node duty cycle, EWMA, WCMA, Pro-Energy, FoBa, leapForward, spikeslab, Cubist, bagEarthGCV.

Acknowledgements

First, I would like to express my deep gratitude to my supervisor **Dr. Ajay Kakkar** for his invaluable advice and encouragement at every step of my Ph.D. program. Without his unfailing support and belief in me, this thesis would not have been possible. His contribution to this thesis goes well beyond his role as an academic supervisor and includes constant support on a personal level without which this journey may never have been completed. And for this, I am truly grateful. He is a great mentor for my life as well.

I would like to express my gratitude to my director **Prof. Prakash Gopalan** for his constant motivation and encouragement. I would like to thank **Dr. Alpana Agarwal**, Head, Electronics and Communication Engineering Department who guides for the quality research work and self motivation. I would like to thank my research committee members **Dr. S. S. Bhatia**, **Dr. Rajesh Kumar Khanna** and **Dr. Amit Kumar Kohli** for their guidance and support.

My sincere thanks goes to **Dr. Kuldeep Singh** and **Dr. Prashant Singh Rana** for their invaluable insights, guidance and support, especially during the difficult conceptual development stage. I remain indebted for their understanding and support at various stages of research.

I also wish to thank non-teaching staff of the institute for their help and support.

I would like to express my sincere and deep gratitude to my parents Mr. Jagdish Kumar and Mrs. Kamla Sharma, my in-laws Mr. Davinder Sharma and Mrs. Surindera Sharma and family members for their love, encouragement, care and support. Finally thanks to my husband Mr. Vivek Batish for having faith on me and supporting me at every step. Without his support, I would not have completed my Ph.D. program and extend lot of love to my son Vaasu Batish.

Finally I would like to give special acknowledgement to my fellow Ph.D. scholars from whom I learn the art of happiness and never give up approach.

Amandeep Sharma

List of Publications

- Amandeep Sharma, Ajay Kakkar, **"Machine learning based optimal renewable energy allocation in sustained wireless sensor networks"**, Wireless Networks (Springer Nature), pp. 1-29, January 2019
  https://doi.org/10.1007/s11276-018-01929-w
  Impact Factor: 1.981

- Amandeep Sharma, Ajay Kakkar, **"Dynamic programming based optimal renewable energy allocation in sustained wireless networks"**, Journal of Renewable and Sustainable Energy (AIP), Vol. 10, pp. 063705:1-34, December 2018
  https://doi.org/10.1063/1.5043126
  Impact Factor: 1.337

- Amandeep Sharma, Ajay Kakkar, **"Forecasting daily global solar irradiance generation using machine learning"**, Renewable and Sustainable energy reviews (Elsevier), Vol. 82, Issue 3, pp. 2254-69, February 2018.
  https://doi.org/10.1016/j.rser.2017.08.066
  Impact Factor: 9.14

- Amandeep Sharma, Ajay Kakkar, **"Development of modified Pro-Energy algorithm for future solar irradiance estimation using level and trend factors in time series analysis"**, Journal of Renewable and Sustainable Energy (AIP), Vol. 9, Issue 3, pp. 033701-16, May 2017.
  https://doi.org/10.1063/1.4982749
  Impact Factor: 1.337

- Amandeep Sharma, Ajay Kakkar, **"Comparative study of energy harvesting in WSN"**, International Conference on Emerging Technologies in Electronics and Communication, pp. 305-307, 20-22 December, 2013, GNDU, Amritsar.

Table of Contents

List of Figures

List of Tables

List of Notations

I_{max}	maximum solar radiation
ξ	solar declination
ϖ	hour angle
θ_{ϕ}	geographical latitude
θ_{φ}	longitude of the PV sites
T_{LC}	local mean time
TZ	time zone
T_E	time equation
E_M	mean value
n	number of time slots
ρ	smoothing constant for level
L_t	smoothed level component
T_t	smoothed trend component
β	smoothing constant for trend
S_m	seasonal component
κ	smoothing constant for seasonal component
i_m^+	seasonal indices
x_n	energy harvested by the end of n_{th} time slot
$\mu_n^{(d)}$	average of energy predicted by the end of n_{th} time slot
D	number of past days considered
N	total number of time slots
K	number of past intervals considered
E_n	average of observed energy
$M_{(d,n+1)}$	predicted energy for next time slot
GAP_K	weighting factor to account changing weather conditions
$C_{(t+1)}^{predict}$	predicted energy value of future time slot
F	future time slots for which prediction will be carried out
G	future time slots having correlation with current slot t
WP	combined weighted profile
P	selected days from the pool
y_t	data series
Cy_t	periodic cycles
I_t	irregularities
α	weighting factor for short term forecasting

υ	weighting factor for medium and long term forecasting
E^d	most similar profile among the days to the current day
s	pool of most similar days
E_{fn}	error function
E_{actual}	real irradiance value
$E_{predicted}$	predicted irradiance value
$R(w)$	cost function
$w \epsilon R^d$	linear prediction function
$f_j \epsilon R^n$ (j=1,...,d)	feature vectors
δ	Forward/backward step square error reduction
GCV	generalized cross validation
$E_{available}$	available harvested energy
$E_{consumed}$	energy consumed by a load
B_0	initial battery level
η	charging efficiency
E_{leak}	leakage energy
ϱ	rate of energy availability
σ	signal burstiness
B	battery size
L_T	life time
E_c	energy consumption by the node
$Duty(i)$	node duty cycle
S_{sun}	sun slot
S_{Dark}	dark slot
$E_{remaining}$	remaining energy
E_{backup}	battery backup
E_{budget}	energy budget
$E_{maximum}$	maximum allocated energy
$E_{minimum}$	minimum allocated energy
$E_{deviation}$	difference between predicted and real energy profiles
ΔE	energy loss
D_{avg}	average duty cycle

List of Abbreviations

ALADIN	Aire Limitee et Adaptation Dynamique
AM	Air Mass
ANN	Artificial Neural Networks
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANOVA	Analysis of Variance
AODV	Ad-hoc on Demand Distance Vector
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARX	Autoregressive with Exogenous Inputs
ASEA	Advanced Solar Energy Allocation Scheme
AWGN	Additive White Gaussian Noise
BAPs	Budget Assigning Principles
B-C	Bristow-Campbell
BMA	Bayesian Model Averaged
BMS	Baseline Measurement System
BRPS	Bit Reversal Permutation Sequence
CCM	Cross Correlation Method
CCAM	Conformal Cubic Atmospheric Model
CDM	Classical Decomposition Method
CLG	Combined Local and Global
DARPA	Defence Advanced Research Project Agency
DHI	Diffuse Horizontal Irradiance
DIMA	DARPA Interference Multiple Access
DNI	Direct Normal Irradiance
DirRec	Direct Recursive
DPM	Dynamic Power Management
ECMWF	European Centre for Medium-Range Weather Forecasts
EMD	Empirical Mode Decomposition
ENN	Elman Neural Network
ESC	Energy Synchronised Communication
ESRA	European Solar Radiation Atlas
ETD	Expected Transmission Delay
EWMA	Exponentially Weighted Moving Average
FoBa	Forward Greedy Algorithm with Adaptive Backward Step

FSMC	Finite State Markov Channel
GA	Genetic Algorithm
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GAF	Geographical Adaptive Fidelity
GBR	Gradient Boosting Regression
GBRT	Gradient Boosted Regression Trees
GFM	Generalized Fuzzy Model
GFS	Global Forecast System
GHI	Global Horizontal Irradiance
GME	Generalized Mean Ensemble
GNET	Generalized Elastic Net
GP	Genetic Programming
GPR	Gaussian Process Regression
GRNN	Generalized Regression Neural Network
HMM	Hidden Markov Model
LES	Linear Exponential Smoothing
LLR	Linear Regression Model
LS-SVM	Least Square-SVM
MA	Moving Average
MAE	Minimum Absolute Error
MARS	Multivariate Adaptive Regression Splines
MDP	Markov Decision Process
MISMO	Multiple Input Several Multiple Output
ML	Maximum Likelihood
MLP	Multilayer Perceptron
MLR	Multivariate Linear Regression
MOS	Model Output Statistics
MQAM	M-ary Quadrature Amplitude Modulation
MTE	Minimum Total Energy
NNARX	Neural Network Autoregressive Model with Exogenous Inputs
NOAA	National Oceanic and Atmospheric Administration
NREL	National Renewable Energy Laboratory
NWP	Numerical Weather Prediction
NWS	National Weather Services
OP-ELM	Optimally Pruned Extreme Learning Machine
PEP	Parameter Exponential Polynomial
PP	Parameter Polynomial

Pro-Energy	PROfile Energy Prediction Algorithm
PV	Photovoltaic
RBF	Radial Basis Function
RBNN	Radial Basis Neural Network
RFR	Random Forest Regression
RMSE	Root Mean Square Error
RSSI	Received Signal Strength Indication
RW	Random Walk
SA	Simulated Annealing
SAT	Satellite Data
SE	Statistical Ensemble
SES	Simple Exponential Smoothing
SM	Stationary Markovian
SMAC	Sensor Medium Access Control
SOM-OPELM	Self Organizing Map-OPELM
SRRL	Solar Radiation Research Laboratory
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
TGPR	Temporal GPR
TL	Link Turbidity
TOA	Top of Atmosphere
TSK	Takagi-Sugeno-Kang
UD-WCMA	Universal Dynamic-WCMA
UVI	Ultraviolet Index
VOF	Variational Optical Flow
VP-RAD	Vapour Pressure Radiation
VSTR	Very Short Term Reactive
WCMA	Weather Conditioned Moving Average
WMO	World Meteorological Organization
WP	Weighted Profile
WRF	Weather Research and Forecasting
WSN	Wireless Sensor Networks

Chapter 1

Introduction

1.1 Wireless Sensor Networks

Wireless Sensor Networks (WSNs) gives the opportunity to replace wired embedded systems and to frame autonomous systems without the requirement of wiring up an entire network. This advanced technology allows the complex sensor systems to reduce installation/maintenance cost and overall power consumption[1]. Spatially distributed sensor nodes are scattered over a targeted sensed area and interconnected in an ad hoc manner via radio communication links as shown in Figure1.1. These sensor nodes are microcontroller based embedded computing devices which sense and measure the information from the area they are deployed into. For storage and processing of data, they collectively pass the gathered information to central sink node or base station through wireless communication link[2][3][4].

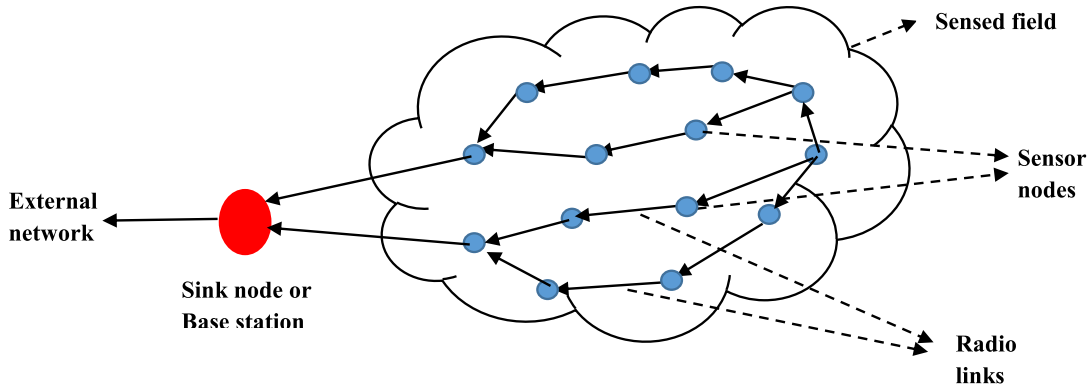


Figure 1.1: Structure of WSN communication

At the boundary of sensor field, data has been transferred to the sink node which is considered as a gateway. The sink node has high processing and memory capacity to perform data processing before transferring the information to external network[5]. Further, High volume of data and random propagation delays affects the overall system performance and quality of service. The network takes more time to process large data and this processing time will increase if the propagation delays are uncertain. In some applications

such as audio and video transmission, real time delivery is of prime concern. In order to meet the QoS guarantees, multiple techniques have been introduced by the researchers including data link layer QoS scheme, integrated multiple layer approaches, QoS routing, an efficient admission and reservation control.

1.2 Prime Concerns about Resources and Designing of Sensor Nodes

WSNs are employed in a wide range of applications but their internal architecture poses significant challenges during network design[6]. A sensor node is the key element of WSNs and designed with a power source, a sensor unit, a microcontroller and a wireless radio unit as the basic components. Figure 1.2 shows architecture of a typical wireless sensor node. Separate modules of a sensor node are constrained by limited resources, summarized as processing, storage, and energy constraints.

The power unit with its regulating circuitry enable the wireless systems to work. The power source is generally a battery of some form which has certain exceptions to address and it is going to deplete over time. Thus, efficient utilization of available energy is the prime issue for WSNs. This problem has been addressed in the literature by dynamically adjusting operating voltages and frequencies at the node level, increasing the efficiency of the cell within the battery, increasing the power density of the cell, and designing of energy efficient network protocols[7].

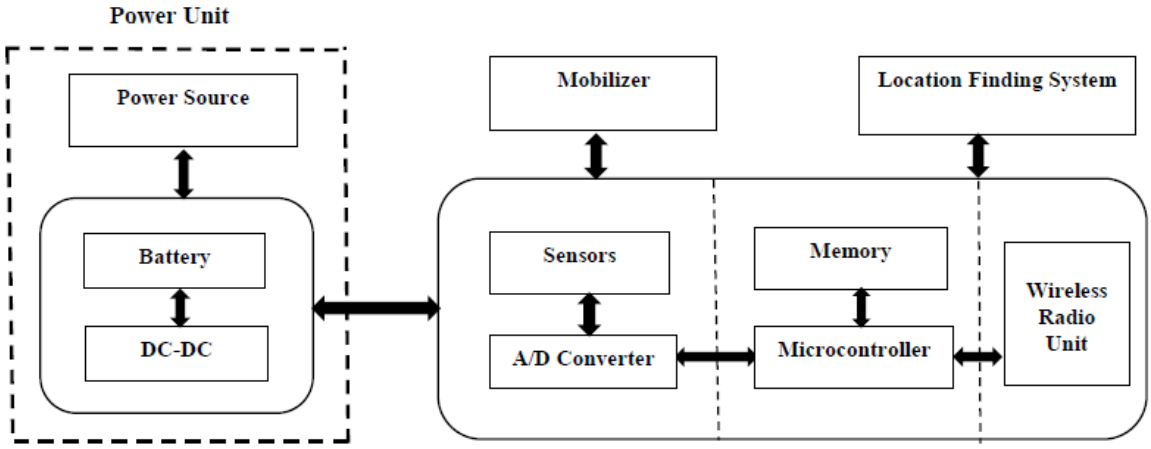


Figure 1.2: Architecture of a wireless sensor node

Another constraint for WSN is characterized by computational power. Nodes have optimized processing unit which is designed to operate on low power embedded devices.

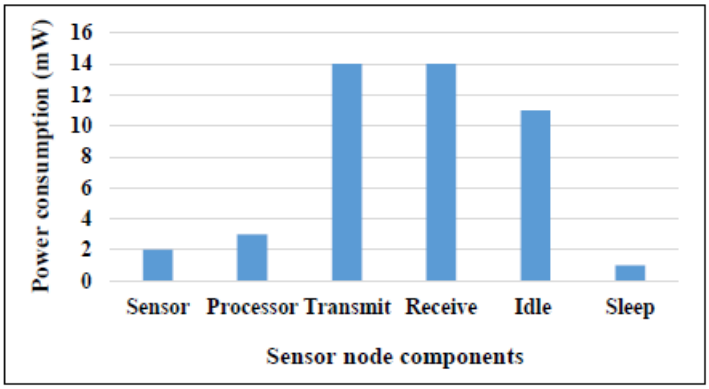


Figure 1.3: Power consumption in WSNs

To perform node level data processing, a microcontroller run at a frequency from 8 to 16 MHz with a memory capacity of kilobytes to some megabytes. Microcontrollers are characterized as low memory processing unit that makes the sensor nodes programmable and performs required processing on data rather than transmitting information myopically. Here, the trade-off is to maintain adequate energy consumption by limiting the complexity of tasks and computation capabilities of the node.

As a third source of constraints, wireless radio unit has a transceiver with many associated components that have an impact on node performance and overall cost of the system. This is because of the fact that the radio communication is an energy demanding process and need consideration to path losses and interference. The power consumption in this unit is high for transmission, reception, and idle state but drops to at least one order of magnitude in a sleep state. Figure 1.3 shows energy expenditure in different operating modes of sensor.

Further, the design of sensing unit is application dependent and may have significant energy consumption. Finally, cost of designing, implementing, and installing WSN is one of the driving issues. With respect to this constraint, sensor nodes should be inexpensive, small, and replaceable. Table 1.1 enlists different commercial sensor nodes and compare their hardware specifications.

1.3 Wireless Network Structure

The network structure can be classified in two basic topologies with respect to the way in which nodes are structured[7]. The first is single hop star networks in which all the sensor nodes can directly communicate with a base station or sink node as shown in Figure 1.4. This structure simplifies the protocols required for network operation but size is limited by the communication range of the sensor nodes. Further, the sensor nodes do not have

Table 1.1: Hardware specifications of different wireless sensor nodes

Sensor Node	**Tiny node** [8]	**MicaZ** [9]	**Imote2** [10]	**Tmote sky** [11]
Supply Voltage	2.4V - 3.6V	2.7V - 3.3V	3.2V - 4.5V	2.1V - 3.6V
Microcontroller	TI MSP430	AT mega 128L	Intel PXA271	TI MSP430
RAM	10 KB	4 KB	256 KB	10 KB
External/ SDRAM	512 KB	512 KB	32 MB	1024 KB
Flash Memory	48 KB	128 KB	32 MB	48 KB
MCU Sleep	6.5 μA	$< 15\mu$A	390μA	5.1μA
Data Rate	152.3 kbps	250 kbps	250 kbps	250 kbps
Radio Chip	XE1205	CC2420	CC2420	CC2420
Radio Frequency	868 MHz	2.4 GHz	2.4 GHz	2.4 GHz
MCU On, Radio off	2.1 mA	8 mA	31 mA	1.8 mA
MCU On, Radio Tx	27 mA	25.4 mA	66 mA	19.5 mA
MCU On, Radio Rx	18 mA	27.7 mA	66 mA	21.8 mA

forwarding property i.e. they do not transmit or receive on the behalf of other nodes. The nodes are independent to each other and can communicate through the sink node only. Further, the hop distance in this configuration is same for all nodes.

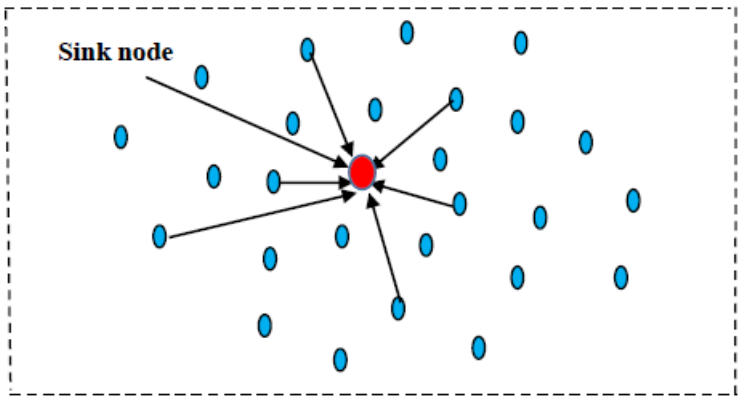


Figure 1.4: Single hop star network

Second is the multi-hop networks in which the distance between the sensor and sink node is greater than one hop. The nodes can communicate with each other and hop distance is not same for all nodes. Multi-hop networks cover a wider area having sensor nodes with forwarding properties but require communication and routing protocols[12] that make the system complicated and energy demanding. According to the internal configuration of nodes, multi-hop networks have been further divided into flat and hierarchical (cluster based) networks. In flat networks, all the nodes in the network are equal and have same duties. Figure 1.5 shows the internal structure of a flat network.

In hierarchical configuration (Figure 1.6), the network is subdivided into single hop groups with special nodes called cluster heads[13][14]. All regular nodes transmit their informa-

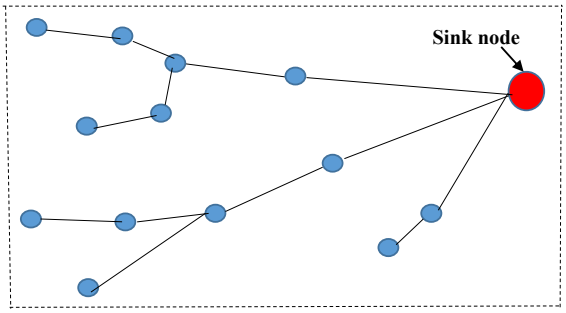


Figure 1.5: Flat network infrastructure

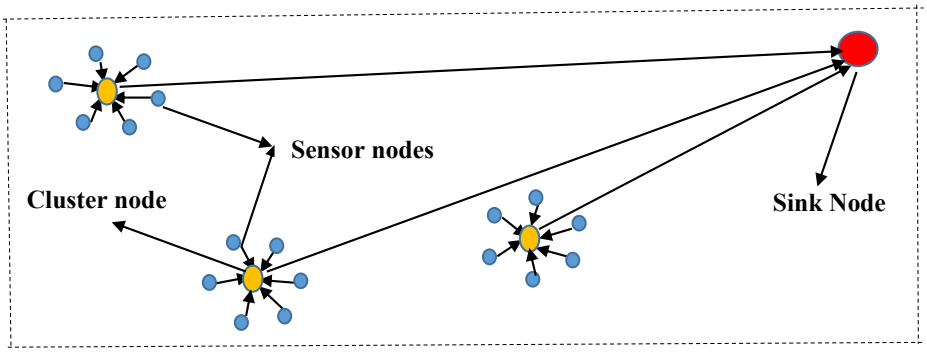


Figure 1.6: Hierarchical network infrastructure

tion to respective cluster heads and not forward any information to other nodes. Cluster nodes are different from regular nodes logically/physically and are responsible for the entire group. A communication set up with sink node has been set up using dynamic protocols.

1.4 Practical Challenges in Wireless Sensing and Networking

Energy consumption is one of the main challenges that is faced by a WSN node. Energy becomes a gridlock especially for those applications where the network is expected to operate for the long time period. Thus, efficient management of battery energy is essential because of the following factors:

- A sensor node will be dead if its battery runs out and overall network performance degraded accordingly. Frequent battery replacement is infeasible and increases the cost of network maintenance because of large and inaccessible installations[15].
- Small size of the sensor node is important for the ease of installation. Batteries represent the majority of weight and volume of the sensor node. The requirement

of a small volume of the device limits the amount of energy that batteries can store. On the other hand, small batteries with very low power density cannot be utilized to power the node because it sharply reduces the entire lifetime of the node. Thus, there is always a trade-off between the node size and power density of the battery. Research has received substantial focus in the direction of enhancing the energy density of energy storage elements[16][17][18].

- Several energy efficient mechanisms were proposed in the literature to prolong node life cycle and improve network performance level. These include adaptive duty cycle strategies[19][20], energy efficient Medium Access Control (MAC) protocols[21][22], power-aware routing and scheduling[23][24][25], adaptive sensing rate[26], tiered system architecture[27], and redundant node placement[28][29]. Although these energy efficient mechanisms can extend the life of sensor node but lifetime is still restricted because of limited energy resources. A sensor node consumes all energy after a specific period of time and stops functioning.

Table 1.2 includes the power consumption of some commercially available sensor nodes in different operating modes[30]. It also highlights the impact of node duty cycle and storage capacity on the lifespan of nodes.

Table 1.2: Power consumption in WSN

Node	Transmit mode (mA)	Receive mode (mA)	Sleep mode (mA)	Duty cycle	Operating voltage (V)	Batteries (Alkaline AA cells)	Battery life cycle (days)
$MicaZ^{TM}$	-	-	-	0.70	2.7	2	170
$XBee^{TM}$	45	50	10	0.51	2.8	2	230
Fleck3	36.8	18.4	80	0.27	3.3	3	440

1.5 Energy Harvesting Wireless Sensor Networks

Strict energy constraints and requirement of the uninterrupted life cycle of sensor node makes the ground for incorporating energy harvesting mechanism in wireless sensor networks. The term strict energy constraints points to the fixed charging capacity of conventional batteries. Because once the battery is fully discharged, replacement is the only option. Especially in remote areas (e.g. military or under water scenarios) where the replacement is not possible, the sensor node is dead after energy depletion state. Energy harvesting technologies conquer the major constraints in WSNs that arises because of high power requirements of the sensor node and limited lifecycle of energy storage elements. The environmental sources power the WSN by small but infinite energy providing that

the lifetime of the sensor node is not limited by stored energy but dependent upon the reliability of WSN hardware/software and efficiency of energy harvesting module. Figure 1.7 shows the comparison of energy usage in battery operated and energy harvested sensor nodes. It is depicted that battery operated node discharged monotonically without any recharge option. On the other hand, energy harvested node recover its charge at different time instants with respect to availability of solar energy. The different harvesting intervals are the different time instants at which solar energy is available to recharge the node and that can be defined in minutes, hours or days[31].

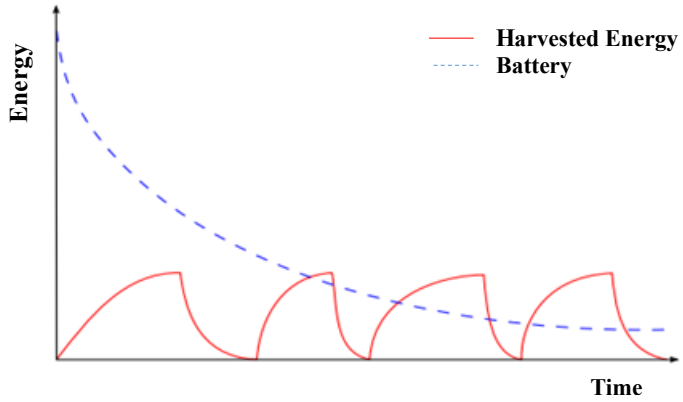


Figure 1.7: Comparison of energy usage in battery operated and energy harvested sensor nodes

Energy can be harvested from several environmental sources including solar and wind in outdoor installations, artificial light from indoor deployments, and heat from radiators. The energy harvesting wireless sensor networks have advanced sensor designs that are capable of harvesting energy from the environment and have the potential for perpetual network operations. The advantages of environmentally powered WSN over battery operated WSN are the decreased cost of maintenance, minimum need for battery replacement, environmentally suitable energy source, reduced battery waste, and reduced human involvement[32].

1.5.1 Energy Harvesting WSN Node

Energy harvesting WSN node has the same hardware configuration as in conventional node except the power supply unit[33]. The fundamental components of a wireless sensor node with energy harvesting capability is shown in Figure 1.8. Energy harvesting unit, such as solar panel has the ability to harness the ambient energy and convert it into electrical form. This form of energy support all node operations. The harvester output is collected by power management unit. This unit delivers energy to the other system

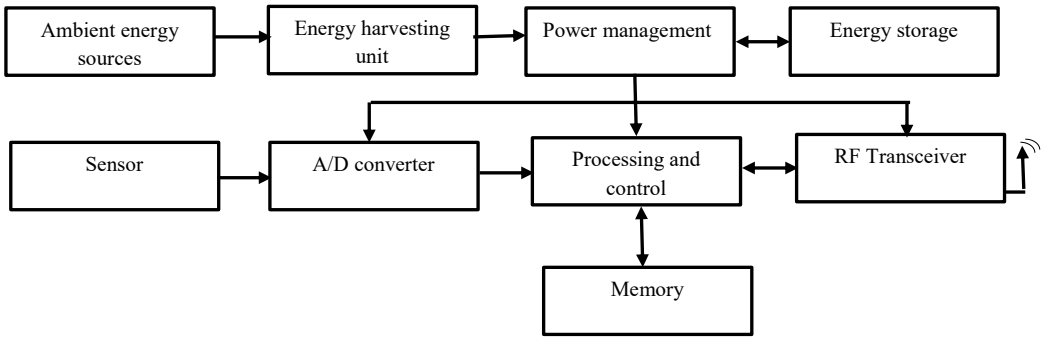


Figure 1.8: Architecture of wireless sensor node with energy harvester

components in an adequate proportion to prevent energy depletion state and store the excess energy in energy storage unit for future usage. The rest of circuitry including sensory unit, A/D converter, processing unit, transceiver unit, and memory performs the same operation as in conventional WSN node but their operation has been controlled by power management unit.

1.5.2 Energy Harvesting Technologies and Energy Sources

Ambient energy is available in many forms that have been considered for energy harvesting [34][35][36][37]. These technologies have different potential power densities that ranges from kilo watt to micro watt. Figure 1.9 depicts different energy harvesting sources. A brief explanation of these sources with appropriate references is given below:

Biochemical Energy Harvesting uses electrochemical reactions to convert oxygen and endogenous substances into electric power. The biofuel cells will act as enzymes/ catalysts and used for harvesting the biochemical energy that is further converted into electricity. Even a human body contains many types of substances that have harvesting ability such as glucose. Biochemical energy harvesting technique is biocompatible and generates continuous power but its performance depends upon the availability of fuel cells[38].

Wireless Energy Harvesting is of two types namely RF energy harvesting and resonant energy harvesting.

- In RF energy harvesting, electromagnetic waves are converted into electricity. The rectifying antenna is used for this purpose. The energy source can be RF power from Wi-Fi communications, mobile phones, microwave, television, and radio broadcasting or electromagnetic waves of a specific wavelength. RF energy harvester can generate power in micro-watts to milli-watts range. This factor depends upon the

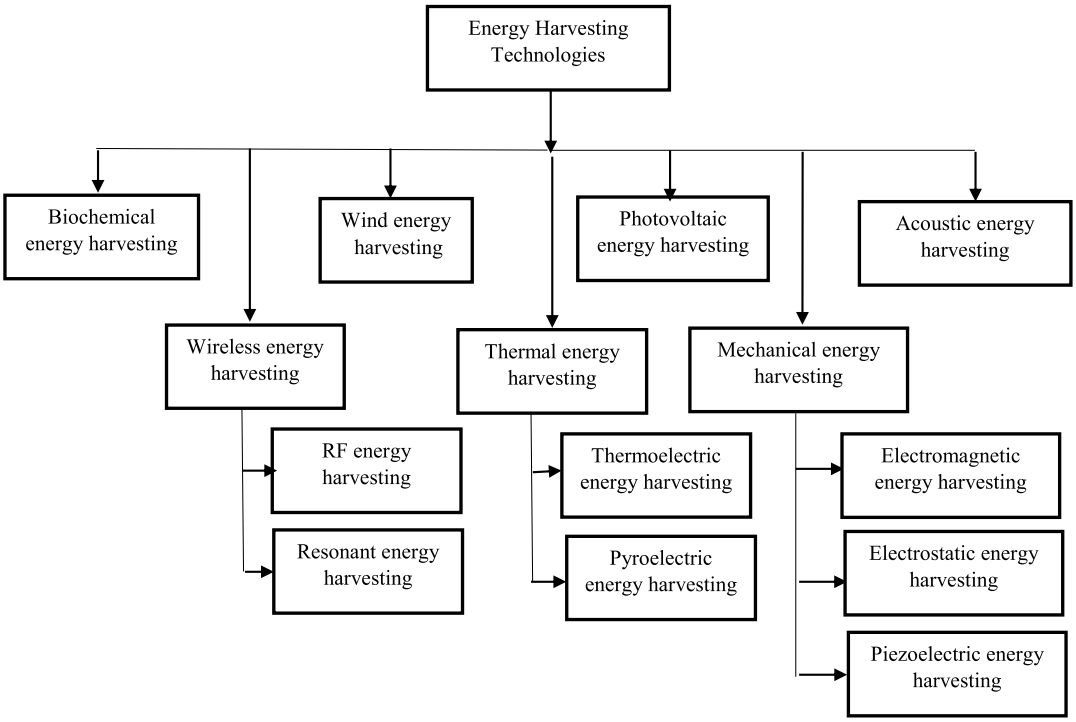


Figure 1.9: Different energy harvesting technologies

distance between RF source and energy harvester unit[39][40].

- Resonant energy harvesting generates electrical energy using resonant inductive coupling between two coils which are tuned at the same frequency[41]. An inductive transformer coupled to primary coil transmit power to the secondary coil through the air. The time-varying magnetic flux generated by primary coil crosses the secondary coil and induce a proportional voltage. A strong electromagnetic coupling between resonant devices can generate power over large distances.

Wind Energy Harvesting converts the air flow energy into electricity. A linear motion of a large-sized wind turbine has been used for electricity generation. Miniature wind turbines generate sufficient energy level to power the sensor nodes. Unpredictable flow and fluctuations in wind strength are the main research issues in this approach[42][43][44][45].

Thermal Energy Harvesting has been categorized into thermoelectric and pyroelectric energy harvesting.

- Thermoelectric energy harvesters generate electrical energy by thermal gradients using thermoelectric power generators. The process is based on the Seebeck effect. According to this principle, an electric voltage is generated when two different metals joined at two junctions are kept at different temperatures. This process

will produce a voltage which is directly proportional to the temperature difference between hot and cold plates. The energy is harvested till the temperature difference has been maintained[46].

- Pyroelectric energy harvesting uses time-varying temperature changes of pyroelectric materials. Time-dependent heating or cooling of these materials change the crystal structure of materials by altering atom positions. The crystal should continuously be subject to temperature changes for proportional voltage generation otherwise generated voltage will diminish due to leakage[47].

Thermoelectric energy harvesters generate high energy levels as compared to pyroelectric energy harvesting but energy efficiency of later technique is high.

Photovoltaic Energy Harvesting converts the light from indoor and outdoor sources such as solar or artificial light into electric form by PV cells. Photovoltaic cell releases electrons when exposed to the light[48]. Photovoltaic energy harvesting provides high power levels as compared to other harvesting technologies and suitable for large-scale energy harvesting units[49][50]. The efficiency of photovoltaic energy based systems is highly dependent upon weather conditions[51][52].

Mechanical Energy Harvesting considers mechanical stress, strain, vibrations, pressure, rotational movements, mechanical force, and fluid as mechanical energy that is converted in electrical form. The conversion of oscillations and displacements of a spring mounted device within the harvester into electricity is the basic principle of mechanical energy harvesting[53]. This type of harvesting is divided into three categories including electromagnetic, electrostatic, and piezoelectric energy harvesting.

- Electromagnetic energy harvesting follows the Faraday's law of electromagnetic induction. The electric voltage is induced by moving a magnetic material in a magnetic field that is created by a stationary magnet. The movement of magnetic material changes the flux and proportional voltage is generated[54]. Reliability is the main advantage of this type of harvesting because of the absence of the mechanical contacts between parts. The large size of electromagnetic components is the main disadvantage for which it is difficult to design with WSN node.

- Electrostatic energy harvesting utilizes the phenomenon of changing the capacitance of vibration dependent variable capacitor[55]. Vibrations change the capacitance of the fully charged capacitor by displacing the plates and proportional voltage is generated. Due to the small size, these type of harvesters can be integrated into sensor nodes but require extra voltage source to charge the capacitor initially.

Table 1.3: Comparison of different energy harvesting techniques

Energy harvesting technology		Potential power density
Mechanical Energy	Electromagnetic	4.0 μW/ cm^3
	Electrostatic	3.8 μW/ cm^3
	Piezoelectric	500 μW/ cm^3
Photovoltaic Energy	Solar Indoor	10 μW/ cm^2
	Solar outdoor (Direct Sun)	15 mW/ cm^2
	Solar outdoor (Cloudy Day)	0.15 mW/ cm^2
Thermal Energy	Thermoelectric	Human: 30 μW/ cm^2
		Industry: 1 to 10 mW/ cm^2
	Pyroelectric	8.64 μW/ cm^2
Wireless Energy Harvesting	RF	0.1 μW/ cm^2 for GSM 900/ 1800 MHz
		0.01 μW/ cm^2 for Wi-Fi 2.4 GHz
	Resonant	3 to 14 mW/ cm^2
Wind Energy Harvesting		380 μW/ cm^3
Acoustic Energy Harvesting		0.96μW/ cm^3 at 100 dB
		0.003 μW/ cm^3 at 75 dB

- In Piezoelectric energy harvesting, vibrations or pressure is converted into electrical energy by using the piezoelectric effect. The principal is to put a strain on piezoelectric material that causes charge separation in the harvester. It generates a proportional electric field to produce voltage with respect to the stress generated[56]. The main advantage of piezoelectric material is a direct generation of proportional voltage without an additional source of supply. Charge leakage and breakage are the limitations of piezoelectric materials.

Acoustic Energy Harvesting uses continuous acoustic waves from the environment to convert it into electrical energy by acoustic transducer[57]. The acoustic emissions are present in the form of transverse, longitudinal, bending or hydrostatic waves with low to high frequencies. This type of energy can only be harvested in the very noisy environment and can generate a power level up to $0.96\mu W/cm^3$ which is lower than other harvesting technologies.

Table 1.3 summarizes different energy harvesting sources with their potential power densities[33]. From the above survey, it has been observed that solar energy is the most readily available source of energy with sufficient energy conversion efficiency. Hence, outdoor employment of WSNs commonly employs solar energy as harvesting source. Some existing solar energy harvesting platforms are Heliomote, Prometheus, and Ambimax with practical deployments. Table 1.4 shows the specifications of different solar energy harvesting sensor nodes including solar panel size, panel output, energy available (mWh/day), sensor node used, battery type, and battery capacity.

Table 1.4: Specifications of different solar energy harvesting sensor nodes

Node Name	**Solar Panel Size** (in * in)	**Solar Panel Power (mW)**	**Energy Availability** (mWh/day)	**Sensor Node**	**Storage Type**	**Battery Type**	**Battery Capacity (mAh)**
Heliomote	3.75*2.5	190	1140	Mica2	Battery	Ni-MH	1800
AmbiMax	3.75*2.5	400	1200	Telos	Supercapcitor (two 22pF) and battery	Li-Poly	200
HydroWatch	2.3*2.3	276	139	TelosB	Battery	Ni-MH	2500
Prometheus	3.23*1.45	130	780	Telos	Supercapcitor (two 22pF) and battery	Li-Poly	200
Everlast	2.25*3.75	450	2700	-	Supercapcitor (100F)	-	-
Sunflower	-	4 Pin photo diodes 20mW	100	-	Supercapcitor (0.2F)	-	-
SolarBiscuits	2*2	150	900	-	Supercapcitor (1F)	-	-

1.5.3 Energy Usage

The way of energy usage by the sensor nodes depends upon the designed architecture of the system and is categorised as: (i) Harvest-Use and, (ii) Harvest-Store-Use configuration[33]. Figure 1.10 and 1.11 depict these two architectures and show the additional components required in the later design.

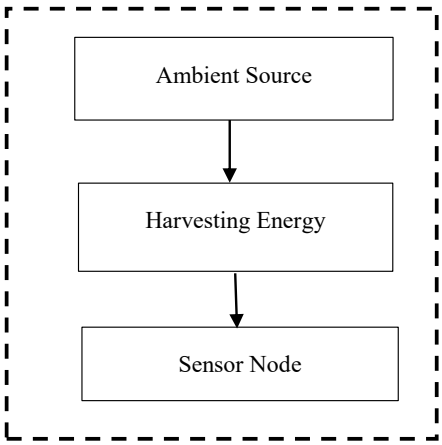


Figure 1.10: Harvest-Use architecture

The Harvest-Use architecture employs all harvested energy directly to power the sensor nodes without any energy buffer. If the energy is below threshold level, the node will stop functioning. The required supply voltages and drawn currents of different sensor nodes have been enlisted in table 1.1. For an uninterrupted operation of the node, the energy harvester should provide the power output above the minimum threshold level

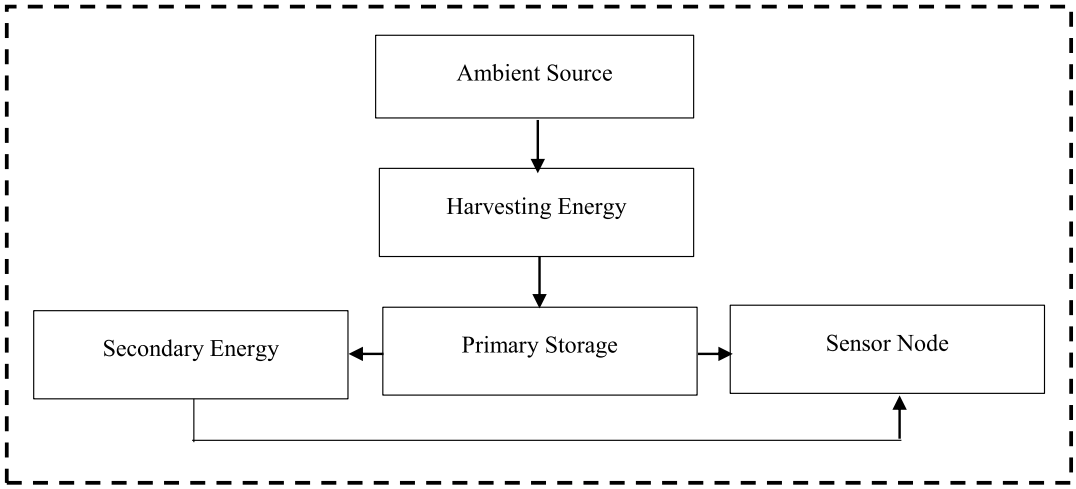


Figure 1.11: Harvest-Store-Use architecture

continuously otherwise node will stop working. The node resumes its operation only when sufficient energy is harvested again. Another shortcoming of this configuration is the wastage of unused energy[33]. The efficiency of operation degrades in this type of configuration as the node is dead in case of energy deficiency.

These limitations have been addressed in Harvest-Use-Store architecture by integrating energy storage components in system architecture. In this configuration, energy buffers have been added in system design which are charged by energy harvester (producer) and discharged by sensor node (consumer). The system will behave like Harvest-Use configuration when production rate is equal to or less than consumption rate. In those conditions when the harvested energy is more than the required energy, the excess energy will be saved by storage components for later use. The advantage of above-mentioned configuration is the uninterrupted operation of the node in case of unsuitable energy harvesting conditions or high consumption rate than energy generation. These two architectures have their specific characteristics and suitable for different applications.

1.5.4 Energy Neutrality for Sustainable Operation

In energy harvesting WSNs, certain considerations are different from battery operated networks. Instead of limiting the maximum energy usage, the network has a limit on the maximum rate at which the energy can be utilized. Secondly, harvested energy is non-deterministic in nature, so efficient energy utilization is of prime importance. As long as the energy generation rate is higher or equal to the consumption rate, energy does not put a limitation on the life cycle of the node. This operating state is known as the Energy Neutral Operation (ENO) or sustainable state[58]. The operating conditions

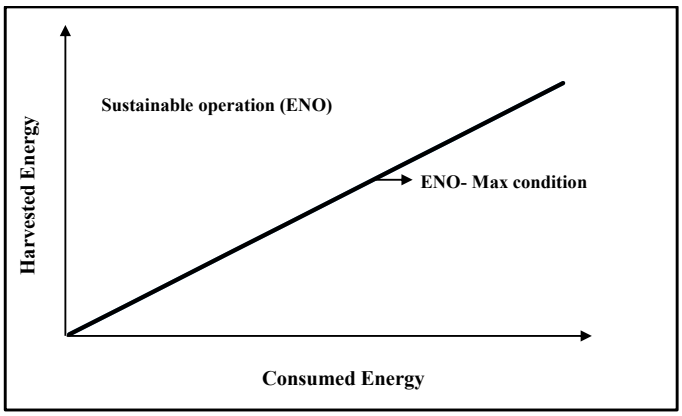


Figure 1.12: Operation of sensor node in different energy expenditures

where the generation rate of harvested energy is much higher than required, the excess energy should be utilized to optimize the network performance level. Thus, the energy harvested WSNs have a twofold goal: the first is to achieve sustainable operation and second is to maximize the system performance. This state is referred to as ENO-Max state in which optimal sustainable performance level with a desired operating point is achieved while balancing the energy generation and consumption rates. Figure 1.12 [58] shows the operation of a sensor node in different energy expenditures. The node operates in sustainable mode if it harvests more energy than it consumes and should put all energy that it harvests to maximize the performance (ENO-Max). On the other hand, if the expenditure is more than energy generation rate, energy put a limit on sensor life cycle. To achieve the goal of sustainability and system performance, all the network protocols and services should utilize the available energy efficiently. Secondly, energy harvesting arises the need for adaptability. Due to the random and unpredictable nature of ambient sources, the network protocols and services should adapt their energy consumption profiles with respect to energy availability[58].

A practical solution of sustainable operation and maximized performance is the use of energy buffers or Harvest-Store-Use approach that satisfies the sudden additional energy requirements while maintaining the residual energy level of buffer between the maximum and minimum thresholds[33][58].

1.6 Harvesting Aware Power Management

For efficient use of energy resources and effective management of node hardware, harvesting aware power management techniques are required with intelligent hardware solutions. These power management techniques schedule the energy in such a way that the network

performance is efficiently maximized without the chances of the battery under or overflow and leads to an uninterrupted node operation [58][59].

1.6.1 Adaptive Power Management Approach

Adaptive power management approach adjusts the node duty cycle with respect to energy availability. The node operates on decreased duty cycle when energy is not sufficient to maintain the high-performance level and vice-versa. Based on the historical data set, a prediction model estimates the future availability of energy and power control unit adjust the operational parameters accordingly. The adaptive duty cycle approach will significantly enhance the overall network performance by ensuring[58], (a) energy consumption by the load should not be more than that of energy availability (direct and stored energy), (b) adjusts the dynamics of the system according to the energy source in the field and, (c) enhance the network performance with respect to application utility model.

1.7 Motivation for the Research

Several energy harvesting techniques were implemented in the last few years by recognizing their long-term autonomous operation. Ambient energy powered wireless sensor nodes can run over long time period without any human involvement. When a specific type of energy harvesting system is implemented for a typical application area, a number of system design parameters need to be considered to enhance the efficiency of the harvesting platform and to meet the application requirements. High system efficiency means that weight, cost, size, and life cycle of the system should be as efficient as expected. Thus, designing of a simulation level system model is important to integrate the physical, electrical, and other system modules practically.

The objectives of the work is to optimize the performance of environmentally powered, self-sustaining wireless sensor node. By considering all constraints, the wireless sensor node should follow energy efficient and parameter optimization principles. The proposed work aims to optimize environmentally powered wireless sensor network by considering theoretical and practical implementations. To achieve this goal, an analytical approach has been adopted that consider the trade off between energy generation, energy consumption, and power management of the system. A systematic solution to this problem has been proposed that is validated by extensive simulations.

1.8 Thesis Statement

This thesis focuses on statistical analysis and analytical modeling of the energy harvesting wireless sensor nodes. The research objectives of this study have been divided into four parts and summarized as follows:

1. Development of modified prediction algorithm for energy harvesting.
2. Optimization of forecasting algorithm to predict future energy more accurately.
3. To optimize average duty cycle for a node in an adaptive duty cycle algorithm to meet performance requirements.
4. Comparison of the proposed technique with the existing ones.

1.9 Research Methodology

To fulfil the objectives of the thesis, the work has been segregated in three stages. Figure 1.13 shows the methodology of the proposed work. In the first stage, work is carried out for accurate pre-estimation of solar energy. For this, the Exponential Weighted Moving Average (EWMA), Weather Conditioned Moving Average (WCMA), Pro-Energy algorithms are analyzed analytically and time series approach is adopted to optimize the Pro-Energy algorithm. The precise solar estimations for few hours ahead time horizons have been achieved. In the second stage, work is done for long-term forecasting and adopt machine learning approach for solar prediction. Five machine learning models viz; FoBa, leapForward, Spikeslab, Cubist and bagEarthGCV are applied and their ensemble approach is introduced for highly stable and precise results. In the third stage, work is done for the optimization of node duty cycle. The forecasted information about solar irradiance is used to calculate energy budgets for different operations of the sensor node. Dynamic programming approach is used to formulate different energy models of a node aimed for high and stable average node duty cycle.

1.10 Thesis Organization

A brief summary about individual chapters is presented in the following section:

Chapter 1 presents the introduction about energy harvesting wireless sensor networks. It includes wireless network structures and practical concerns about resources and de-

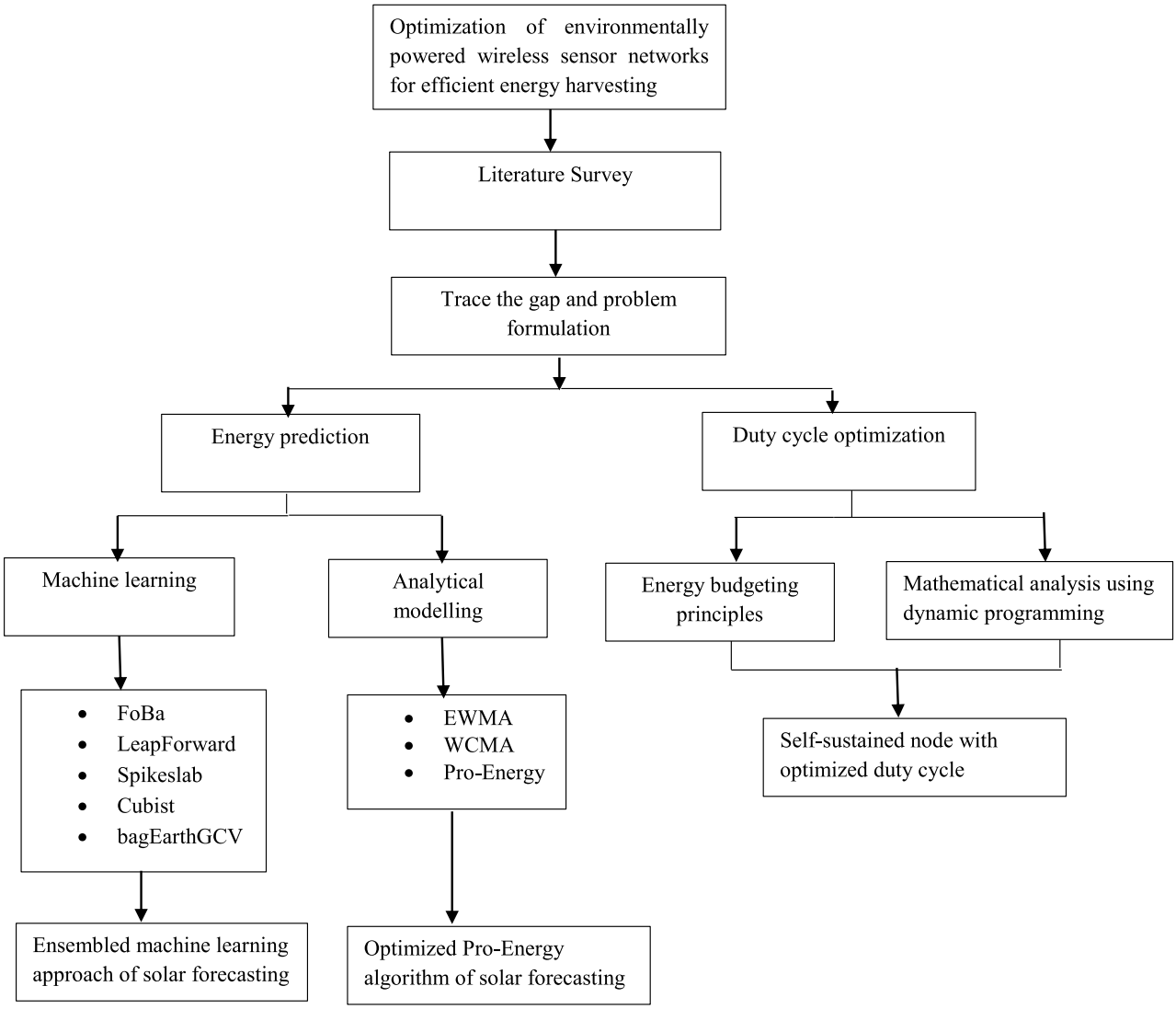


Figure 1.13: Research methodology

signing of sensor nodes. Different energy harvesting technologies are discussed and analysed.

Chapter 2 provides in-depth survey of energy harvested wireless sensor networks. It includes review on different solar irradiance forecasting approaches and analysis of their advantages and disadvantages. A detailed review has also been presented on adaptive power management techniques and summarizes the solutions in terms of detailed study of different protocols, transmission and reception techniques, scheduling, and routing strategies.

Chapter 3 presents the basics of solar irradiance and solar forecasting. Based on the energy constraints in this area, a detailed theoretical analysis of energy harvesting systems is carried out. Solar irradiance prediction empowers the system to make dynamic tradeoffs

between energy consumption, system operation, and operational fidelity to attain an energy neutral operation. From the detailed literature survey, it is observed that a further improvement in prediction accuracy is required to predict the trace close to real irradiance profile. In the proposed work, a modified Pro-Energy algorithm for energy forecasting is developed using the time series analysis. By using the trend and level components of a time series in solar irradiance forecasting, the average growth of irradiance (increasing or decreasing) and a constant value for irradiance has been estimated at the end of each time slot. The estimated trend represents a long-term change in irradiance level of the solar energy. Thus, it provides the approximation about irradiance levels for the future time slots. The methodology for optimization of Pro-Energy algorithm is presented and each step is explained in detail. For the validation of effectiveness of proposed approach, comparison with existing approaches is presented in section 3.9. This chapter covers the first and fourth objective of proposed thesis.

Chapter 4 explains the basics of machine learning methodology together with the ensembled techniques. Machine learning is an approach of data analysis for the analytical model building. Brief theory of five machine learning models with procedure is presented and advantages of ensembled approach over independent models are given. Machine learning is applied to historical solar intensity observations as a training dataset to calculate future solar irradiance for different forecasting horizons irrespective of seasonal variation and input parameters availability. To achieve more precise forecast, Statistical Ensemble (SE) approach is used on machine learning models to circumvent the effect of forecasting horizon and seasonal variability. Optimal characteristics of accuracy and stability are obtained by aggregating base machine learning models. The statistical ensemble performs generalized mean ensemble including Pythagorean means (arithmetic, harmonic and quadratic means) and median rule. The performance of presented models is analysed for different weather conditions, different forecasting horizons, and different historical parameters. These results are given in section 4.5.2. This chapter covers the second objective of proposed thesis.

Chapter 5 presents the proposed energy efficient power management approach for energy harvested wireless sensor node. Analytical models of different energy profiles of a node are described. In proposed work, an integrated approach of energy assignment principles with adaptive duty cycling is introduced to efficiently utilize the available energy. Initially, the node duty cycle is dynamically adjusted with respect to the estimation of available energy and then subject to real-time changes according to measured solar irradiance profile. Machine learning based solar irradiance prediction is used to pre-estimate the node duty cycle. Real-time changes in computed duty cycle are implemented by

proposed adaptive duty cycle algorithm. The linear programming based intense calculations are replaced by simple mathematical programming which offers low computational complexity. The proposed work offers minimum variation in node duty cycle in case of unstable environmental conditions and undesired hardware degradations. The effectiveness of proposed work is validated by extensive simulations on real-time solar energy profiles in terms of magnitude and stability of sensors average duty cycle. The proposed work is compared with existing power management approach and shown in section 5.3. This chapter covers the third and fourth objective of proposed thesis.

Chapter 6 concludes this thesis by summarizing the findings of the proposed work, highlighting the future work, and presenting final remarks.

All the work proposed in this thesis aim to optimize an environmentally powered wireless sensor network that has a perpetual life cycle and satisfies energy efficiency and network performance.

Chapter 2

Literature Survey

This section describes the work carried out by the various researchers so far in the field of energy harvesting in wireless sensor networks. A literature survey has been carried independently for solar forecasting and node level power management.

2.1 Solar Forecasting

Dependency on meteorological conditions causes renewable energy resources to be inconsistent. Under this constraint, reliable solar irradiance forecast on different time horizons is essential for developing and utilizing solar energy based systems. As a sequel, research on solar irradiance forecasting has been germinated along with the areas of forecasting theory[60][61][62], solar physics[63], stochastic processes[64][65], and machine learning[66][67].

Meza *et al.*[68] presented a comparison between temperature dependent and sunshine hour dependent models. It was observed that temperature-dependent empirical models have large correlation coefficient than cloud cover dependent models in weather conditions of Chile. The main issue with empirical models was the availability of various meteorological parameters and their calibration for different locations. Reikard *et al.*[69] run forecasting tests over regression in logs, Autoregressive Integrated Moving Average (ARIMA), transfer functions, Neural Networks (NN), and hybrid models. It was observed that at high resolutions, a transfer function using cloud cover achieved high accuracy than ARIMA. At very high resolutions, NN and hybrid models improved the solar forecast. ARIMA obtained good results in logs with time-varying coefficients and effective in capturing the diurnal cycle of solar radiation. Khatib *et al.*[70] presented a review of solar radiation modeling and prediction techniques. Global and diffuse solar energy models for radiation modeling and linear, nonlinear, and artificial intelligence modeling techniques for prediction were studied. It was observed that ambient temperature, sunshine ratio, and relative humidity were the most relevant parameters for solar energy prediction.

Inman *et al.*[60] reviewed different forecasting methodologies and solar forecasting based

applications at the utility scale level. The stochastic learning methods including Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and ARIMA take the advantage of correlated patterns of solar irradiance and applied to different forecasting horizons. Artificial Neural Networks (ANN) were non-linear approximators and successfully forecast the solar irradiance on intra hour to yearly ahead time horizons. The satellite data in conjunction with regression techniques provided an alternative to ground-based pyranometers and forecasting on intra-hour to intra-day ahead time horizon. Limited spatial resolution of satellite images and satellite sampling frequency put spatial and temporal limitations on these methods. Numerical Weather Prediction (NWP) methods gain high attention for long term forecasting horizons and lead to modeling of temporal progression of cloud cover in poor ground network areas. Integration of stochastic learning methods with satellite/NWP methods and assimilating meteorological data from sky imagers was suggested to achieve high forecasting accuracy in different time horizons.

Diagne *et al.*[71] reviewed statistical models and NWP models for solar forecasting. In statistical models, time series models, sky image-based models, satellite data-based models, Wavelet models, and ANN models were included. NWP models utilized the reproduction of physical systems. It was observed that selection of a particular forecasting approach was highly influenced by the availability of data and forecasting horizon. Simulation results showed that for short-term forecasting horizon, ARIMA achieved satisfactory results whereas for long-term forecasting, European Centre for Medium-Range Weather Forecasts (ECMWF) with Model Output Statistics (MOS) post-process gained high prediction accuracy. In the condition of high atmospheric uncertainty, hybrid models and sky imagery approach improved the forecasting results. Yadav *et al.*[72] and Kadiyala *et al.*[73] reviewed different ANN based solar radiation prediction models on horizontal and inclined surfaces. The Multilayer Perceptron (MLP), Radial Basis Function (RBF), Generalized Regression(GM), and Bessian Neural Network(BNN) were analysed with back propagation Levenberg-Marquardt (LM) training algorithms. Simulation results showed that MLP with LM learning algorithm and hyperbolic tangent sigmoid in hidden layer gained high prediction accuracy than Angstrom, conventional, linear, nonlinear and fuzzy logic based models.

Besharat *et al.*[74] presented a review of global solar radiation models and categorized them in four sections including sunshine based models, cloud information-based models, temperature-based models, and other meteorological parameters based models on the basis of model input. All the models carried empirical constant that depended upon geographical location and season of the year. The results showed that sunshine based

models were more accurate among the four categories. In the absence of sunshine information, air temperature based models were good alternatives. The temperature range was the main influencing factor for model accuracy. Cloud information based models also provided good results but were sensitive to human biasing. The other meteorological parameters based models reported good prediction of solar irradiance but a large number of required input parameters were not readily available.

Wang *et al.*[75] investigated three ANN models for solar irradiance prediction in different weather conditions. The Multilayer Perceptron (MLP), Generalized Regression Neural Network (GRNN), and Radial Basis Neural Network (RBNN) used sunshine duration, air temperature, air pressure, relative humidity, and water vapour pressure as input parameters to predict daily global solar radiation. The Bristow-Campbell (B-C) model was also investigated. The results showed that MLP and RBNN models provided high prediction accuracy than GRNN and B-C models. Antonanzas *et al.*[76] presented a review on different photovoltaic power forecasting techniques. Statistical approaches were found more accurate than parametric ones and resource variability would be reduced by spatial averaging than independent sites. Results showed that for long-term forecasting, NWP models achieved high-performance matrix. Recent research in this area validated the effectiveness of machine learning approaches in terms of prediction accuracy and ease of modeling without the need of various meteorological parameters. Chaturvedi *et al.*[77] studied different solar forecasting techniques. In physical models, cloud imagery, satellite-based, and NWP models were considered. In the results, first two models were found effective in those cases where no other data was available and suffered from spatial and temporal limitations. The NWP models were suitable for long-term forecasting.

Mohanty *et al.*[78] presented a survey on the present status of India in Photovoltaic (PV) energy harnessing and use of renewable energy sources. The requirement of solar forecasting was discussed and various forecasting approaches were analyzed. Different balancing methods for effective control infrastructure were analyzed for power grid management in India. Raza *et al.*[79] analyzed different factors related to PV power profile and PV power forecast that were considered as inputs for forecasting model to improve prediction accuracy. Solar radiation, speed and direction of wind, atmospheric temperature, aerosol index, and humidity were considered as relevant factors that affect PV plant output. Different stochastic and artificial intelligence approaches were also investigated. It was concluded that AR, MA, ARMA, and ARIMA based statistical approaches were suitable when less number of meteorological parameters were available while ANN and fuzzy/genetic algorithm based approaches were effective in dynamic environment for providing improved nonlinear approximations and effective in handling uncertainty in solar

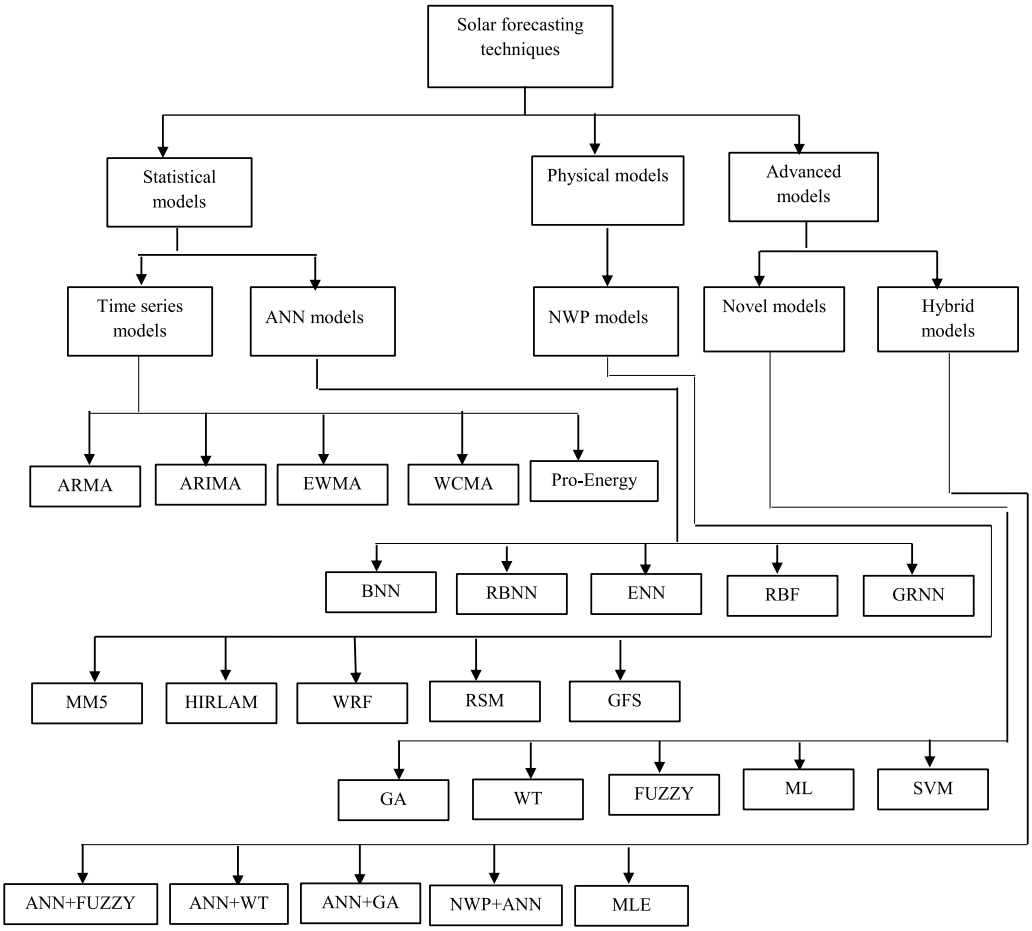


Figure 2.1: Different solar forecasting techniques

radiation profiles. Voyant *et al.*[80] investigated different solar forecasting methods using machine learning approaches. Simulation results showed that Support Vector Machine (SVM), regression tree, and random forest gave satisfactory results and exhibited similar error statistics. It was also observed that predictor ensemble approach gave high prediction accuracy than individual predictors.

Gagne *et al.*[81] analyzed different combinations of statistical learning methods and weather data from different sites to identify the best in terms of low prediction error. The results demonstrated that the configurations of statistical model have a great impact on sharpness versus smoothness of the model fit. The use of a large and diverse training dataset was also suggested to improve the performance of forecasting models. Results also showed the effect of spatial and temporal profiles of input variable on model performance. Different solar forecasting techniques are depicted in Figure 2.1 and a brief description of these techniques is given in the subsequent sections.

2.1.1 Statistical Models

2.1.1.1 Time Series Models

Thornton *et al.*[82], Trabea *et al.*[83], Meza *et al.*[68], Winslow *et al.*[84], El-Metwally *et al.*[85], and Liu *et al.*[86] investigated the correlation between measured solar irradiance and different meteorological parameters. The work was done to trace a set of empirical models for accurate prediction of solar radiation using relevant meteorological parameters. It included mean values of maximum temperature, relative humidity, vapour pressure, sea level pressure, and sunshine hours. In the obtained results, correlation coefficient varied from 0.86 to 0.99 for different geographical parts. Winslow *et al.* presented a Vapour Pressure Radiation (VP-RAD) model for daily solar radiation prediction. Liu *et al.* concluded that B-C model had high correlation of its parameters with other factors that make it suitable for solar radiation measurement without calibration. El-Metwally *et al.* proposed three solar radiation prediction methods which were based on cloud cover information, maximum and minimum temperature values and extraterrestrial global radiation. The models were non-linear and included Four Parameter Polynomial (4PP), Five Parameter Polynomial (5PP), and Five Parameter Exponential Polynomial (5PEP). The standard least square error procedure was used to fit these equations. A comparison was carried out with Supit-Van Kppel model and Angstrom-Prescott method on daily, monthly and seasonal basis, and under different sky conditions. The results showed that the method outperform the Supit-Van Kppel method and had similar error matrix with Angstrom-Prescott method. Since later method was a sunshine duration method, in the absence of sunshine duration, the given method could be used for solar irradiance estimation. Hocaoglu *et al.*[87] introduced a visualization approach to analyse and model one year hourly solar radiation data. Using this approach, nine analytical two dimensional surface fitting models were designed and compared. The gaussian surface model was found most accurate for data characterization purposes. Results revealed that in comparison to neural networks, the analytical surface model provides more generalized form and suitable for various geographical locations.

Janjai *et al.*[88], Pandey *et al.*[89], Lee *et al.*[90], and Wang *et al.*[91] developed semi-empirical models for the prediction of clear sky global and direct normal irradiances in tropical environment. Janjai *et al.* used aerosol parameters, total column ozone, precipitation, zenith angle, wavelength exponent, and air mass to model global and direct normal irradiances as empirical functions. Pandey *et al.* employed day of the year, hour of the day, solar constant, latitude and longitude of the selected location, and optimized parameter values to calculate solar isolation parameter for implementation of empirical

transmission function to estimate solar irradiance. Wang *et al.* proposed Empirical Mode Decomposition (EMD) based solar forecasting algorithm while considering environmental factors into forecasting models of individual decomposed sub series. The results showed that the accuracy of EMD model was comparable with existing physical models. Lee *et al.* observed that maximum value could be taken at inclined angles between 20° and 40° with azimuth angles from 150° to 210°. Khorasanizadeh *et al.*[92] analysed thirteen pre-existing clearness index based diffuse radiation models. The models were analysed on the basis of diffuse fraction and diffuse coefficient with clearness index as the sole input variable. Based on analysis data, twenty-six new models were established for daily and monthly average solar irradiance predictions for Kerman, Iran. It was observed that diffuse fraction model in linear form was most suitable for daily solar predictions while diffuse modeling coefficient in linear form gains high prediction accuracy for monthly average solar predictions. The suggested models were computationally simple as clearness index was the only parameter that is required for global solar irradiance forecasting. Zhu *et al.*[93] presented a semi-empirical model to estimate clear sky Direct Normal Irradiance (DNI). The wavelet technology based forecasting model decomposed the non-stationary series of clear sky index to obtain four sub-series with different frequencies. The forecasting models were designed according to the characteristics of corresponding sub-series and estimate intra hour DNI under all sky conditions.

Renner *et al.*[94] presented an adaptive slotting scheme for real time adaption of energy harvesting patterns by sensor node. An integration of given slotting approach with established forecasting algorithms lead to significant improvement in short and long term forecast accuracy while preserving low memory footprints. The approach achieved depletion safe operation of node and efficient task scheduling with few adjustments in duty cycle. Hyndman *et al.*[95], Piorno *et al.*[96], Taylor *et al.*[97], Dong *et al.*[98], and Dehwah *et al.*[99] provided a state space framework that include an extended range of exponential smoothing methods. Hyndman *et al.* carried out the computation of prediction intervals, likelihood, and model selection criteria to the data from M-competition and M3-competition. Taylor *et al.* proposed Singular Value Decomposition (SVD) based exponential smoothing formulation for one day ahead solar forecasting. Dong *et al.* used exponential smoothing approach with fourier trend model and validated that the given approach showed better performance in comparison with Autoregressive Integrated Moving Average (ARIMA), Linear Exponential Smoothing (LES), Simple Exponential Smoothing (SES), and Random Walk (RW) models, tested on same data set. The application could be extended to two dimensional irradiance forecast where a forecast irradiance map could be generated using 2D spline interpolated trend surfaces. Dehwah *et al.* presented adaptive weighting parameter based Weather Conditioned Moving Average (WCMA) so-

lar forecasting algorithm that considered current and past day weather conditions and successfully reduced the relative mean error to 10% during unstable weather conditions. In conjunction with energy management algorithm, it effectively utilized the available energy under real working conditions. Consideration of real time energy measurements, a set of historical profiles, and lack of tuning parameters made the method robust and suitable for distributed energy harvesting based applications such as wireless sensor networks.

Bacher *et al.*[100], Dazhi *et al.*[101], Salcedo-sanz *et al.*[102], Trapero *et al.*[103], Verzijlbergh *et al.*[104], and David *et al.*[105] used adaptive linear time series models including Autoregressive (AR) and Autoregressive with Exogenous Inputs (ARX) from Numerical Weather Predictions (NWP) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) to predict normalized solar power. Dazhi *et al.* used three types of meteorological input parameters including Global Horizontal Irradiance (GHI), Diffuse Horizontal Irradiance (DHI) and Direct Normal Irradiance (DNI). It was observed that ARX model with NWP inputs and consideration of cloud cover conditions make solar power observations suitable for a horizon up to 36 hours ahead. The collaboration of GARCH with ARMA allowed consideration of the prediction interval associated with point forecast that lead to a probabilistic forecast. Salcedo-sanz *et al.* proposed Gaussian Process Regression (GPR) for solar radiation estimation. The associated parameters were deduced by type-II Maximum Likelihood (MLII) to make analytical expressions of observed variations that lead to computationally efficient model. The results showed the effectiveness of Temporal GPR (TGPR) over other statistical regression techniques in terms of prediction accuracy and bias. Trapero *et al.* used univariate dynamic harmonic regression model for short term forecasting using Global Horizontal Irradiation (GHI) and Direct Normal Irradiation (DNI). Instead of two steps procedure consisting of deseasonalizing the time series and modeling the residuals, the procedure involved in single step recursive method with adaptive predictions. To enhance the forecast accuracy of NWP models, Verzijlbergh *et al.* presented a Model Output Statistics (MOS) routine which was based on stepwise linear multivariate regression method. The results showed that large set of meteorological variables improved the prediction accuracy and lead to probabilistic forecast.

Li *et al.*[106] and Kaplanis *et al.*[107] proposed a trigonometric model with sine and cosine waves for global solar radiation prediction on horizontal surface and compared with three existing trigonometric correlation models proposed by Bulut *et al.*[108] and Al-salaymeh *et al.*[109]. Kaplanis *et al.* proposed cosine function based average daily solar irradiance predictor on a horizontal plane. For a latitude of 0° to 23°, two cosine functions were

obtained by regression analysis using a sum of 3 to 8 gaussian functions and for a latitude of 23° to 71°, two cosine parameters were obtained by a sum of two exponential functions or a product of an exponential and a cosine function. Number of days was the only required parameters for solar prediction. The model was successful in those areas where meteorological data is not readily available.

Cammarano *et al.*[110], Hocaoglu *et al.*[111], and Akarslan *et al.*[112] presented a novel multi-source energy prediction model for short and medium term forecasting horizon which was based on the selection of similar day from the past records to predict present day. Prediction was performed by calculating the Euclidean distance between search vector and data belonging to similar day. A search vector was built that carried data from early morning to the desired prediction hour. The algorithm allowed combining multiple energy profiles to improve energy prediction accuracy. Comparison was performed with EWMA and WCMA and results showed an improvement in prediction accuracy up to 60%. Hocaoglu *et al.* introduced Mycielski model that traces the most similar sub-array pattern from the recorded hourly solar radiation data. The model started with the last recorded value and searches for longest matching solar radiation data array which was considered as prediction. In case of multiple sub-arrays, a Markov chain was used to set probabilistic relation of sub-patterns.

Yang *et al.*[113] and Amod *et al.*[114] developed a finite fourier series model to analyse monthly average solar insolation and presented a forecast method which was based on extended fourier series model in the least squares to predict monthly movement of average solar insolation. The forecast model was designed using optimum fourier coefficients with least square method using previous monthly movements. The obtained fourier coefficients with harmonics were used to predict its movements at subsequent time slots. The fourier series evaluation model analysed monthly average solar radiation using 6-term fourier series. The extended fourier series forecast model in the least square sense predicted monthly movements by less than 6-terms fourier series.

Akarslan *et al.*[115] and Diagne *et al.*[116] applied linear filtering approach for solar forecasting. Akarslan *et al.* converted extraterrestrial irradiance and ambient temperature in the form of 2-D image and correlation between different images was computed using multi-dimensional linear prediction filters. Diagne *et al.* used Kalman filtering method to improve an hour ahead GHI forecasting from measured GHI at the ground and forecasted GHI from Weather Research and Forecasting (WRF). The relevant inputs for Kalman filter were determined by bias error analysis. The solar zenith angle and clear sky index were used for this analysis. The results illustrated that the approach achieved better prediction accuracies in the range of 1% to 30% for multi-dimensional models as compared to

two dimensional ones. Kaplanis *et al.*[117] and Prema *et al.*[118] used time-based composite covariance that encoded time dependent features including seasonal trend, magnitude, smoothness of periodic component, and variation in regular periodicity to account for the nonstationary temporal behaviour of the signal. Kaplanis *et al.* implemented a stochastic model for hourly profile prediction of global solar radiation based on one, two, or three morning measurements of corresponding day. The model was validated with solar measurements taken for spring, autumn, and winter season in Patra, Greece.

2.1.1.2 Artificial Neural Networks

Maier *et al.*[119], Atiya *et al.*[120] and Parlos *et al.*[121] reviewed neural network based prediction models and discussed their designing steps including pre-processing and division of data, appropriate input parameters and network architecture, training weight optimization, choice of performance criteria, and model validation. The impact of algorithm optimization on prediction accuracy, training speed, and effect of stopping criteria in determining optimum number of hidden layer nodes were also considered.

Kemmoku *et al.*[122] proposed multistage neural network for solar forecasting. The first stage of Neural Network (NN), forecasted the average atmospheric pressure of next day from previous day atmospheric pressure data. The second stage NN predicted insolation level of next day from average atmospheric pressure and weather data of past day. The third stage of NN forecasted the solar insolation of next day from insolation level and weather data of previous day. Results showed a reduction in the mean error from 30% (single-stage) to 20% (multi-stage). Jiang *et al.*[123] proposed an ANN based model for monthly average diffuse solar radiation prediction. The solar radiation data from nine different places of China with different climatic conditions was used for training and testing the ANN models. The results were compared with empirical regression models that validated the ability of ANN models to produce accurate solar estimations. Mellit *et al.*[124] proposed a Multilayer Perceptron (MLP) model for 24 hours ahead solar radiation forecasting. Daily average solar irradiance and air temperature were used as input parameters. The configuration with one input layer within three neurons, two hidden layers within 11 and 17 neurons and 24 neurons in the output layer were found suitable for forecasting. Quaiyum *et al.*[125] and Marquez *et al.*[64] introduced ANN based medium term solar forecasting models, feeded with both endogenous and exogenous variables. Quaiyum *et al.* presented feed forward neural network with thirty neurons in the input layer and one neuron in the output layer. Marquez *et al.* used Gamma test combined with genetic algorithm for the selection of input variables and collected meteorological parameters from National Weather Services (NWS) forecasting database. Results showed that

the sky cover, probability of precipitation, maximum and minimum temperature were the most important meteorological variables.

Fidan *et al.*[126] presented variational spectrum model using neural network over harmonic data for the solar prediction. In the presented two step prediction model, hourly solar radiation data of single day was considered and its periodic fourier series coefficients were calculated in the first step. First two coefficients were considered for mathematical modeling using Kaplanis approach[107]. In the second step, feed forward neural networks with solar radiation data and model generated fourier coefficients as input parameters were used to estimate solar behaviour. Alsina *et al.*[127], Ding *et al.*[128], and Aguiar *et al.*[129] presented monthly average estimation of daily global solar radiation using ANNs. Alsina *et al.* proposed feed-forward ANN with one hidden layer and used back-propagation algorithm with seven climatologic parameters including Top of Atmosphere (TOA) radiation, day length, number of rainy days, average rainfall, latitude, and altitude. Aguiar *et al.* proposed ANN models with integration of ground measurements with satellite generated exogenous inputs and NWP data for intraday solar forecast. Bayesian frame work[130] was applied to optimize the ANN architecture by selecting relevant inputs and to improve ANN learning process. It was observed from the obtained results that best results were obtained by combining NN, European Centre for Medium Range Weather Forecast(ECMWF) and Satellite Data(SAT).

Voyant *et al.*[131] applied multilayer perceptron to four types of uncertainties including the error due to measurement, machine learning uncertainty, the variability of time series, and error related to the horizon. By considering these error components, a global uncertainty related to prediction efficiency could be estimated. A reliability index was defined to estimate the validity of prediction.

2.1.2 Physical Models

Huang *et al.*[132] presented a Conformal Cubic Atmospheric Model (CCAM) and Global Forecast System (GFS) to show the effects of simulated regional weather variability on the solar irradiance forecast accuracy. It was observed that high resolution CCAM produced more accurate results than low resolution GFS. The results showed that optimum scale of spatial averaging reduced the extent of solar variability and enhanced the performance of both models. Larson *et al.*[133] presented least square optimization of Numerical Weather Prediction (NWP) for hourly averaged day ahead solar irradiance forecasting in the American southwest[134][135][136][137]. The proposed methodology could be applied without any prior knowledge of NWP model. It was observed that the method was

suitable for modeling the complexity of power plant operations and compensated the systematic errors for day ahead solar forecast.

2.1.3 Advanced models

2.1.3.1 Novel models

- **Support Vector Machine**

  Zeng *et al.*[138] presented a Least Square Support Vector Machine (LS-SVM) model for short term prediction. Two dimensional atmospheric transmissivity, relative humidity, sky cover, and wind speed were used as model inputs and predict atmospheric transmissivity which was converted in solar power with respect to latitude of the site and time of the day. Results showed that the model outperformed AR model and RBNN models in terms of prediction accuracy. Jiang *et al.*[139], Li *et al.*[66], and Belaid *et al.*[140] proposed nonlinear support vector machine based approach for daily and monthly average global solar radiation prediction. Jiang *et al.* used hard penalty function based optimization model to forecast solar radiation. Support vector machine approach was converted into a regularization problem with ridge penalty and a hard penalty function was added to select the radial basis functions. Glowworm swarm optimization algorithm was used to determine optimal parameters of the model. Li *et al.* analysed Hidden Markov Model (HMM) and Support Vector Machine (SVM) regression for short term solar irradiance forecasting under different weather conditions. Use of irradiance cloud models generated by sky cameras and data fusion techniques were suggested to improve the forecasting results. Belaid *et al.* considered ambient temperature, sunshine duration, and extraterrestrial solar radiation as model inputs and presented a comparative analysis with existing ANN models. The results showed that SVM models achieved relatively same performance level compared to ANN based models.

- **Extreme Learning Machine**

  Grigorievskiy *et al.*[141] applied Optimally Pruned Extreme Learning Machine (OP-ELM) for long term time series prediction. OP-ELM was used with recursive, direct, and DirRec strategies with linear least square model and Least Square Support Vector Machines (LS-SVM). It was observed from the results that DirRec strategy with LS-SVM lead to heavy computations in comparison to OP-ELM that allowed reasonable computational time for DirRec strategy. It was also shown that an ensemble of three strategies for OP-ELM could highly improve the prediction ac-

curacy. Wu *et al.*[142] introduced Self Organizing Map-Optimally Pruned Extreme Learning machine(SOM-OPELM) model for short term solar radiation prediction. Data segregation was done by SOM and prediction model was built by OPELM. Multi-step prediction was also performed using three time series strategies including recursive, DirRec, and Multiple Input Several Multiple Output approach(MISMO). Results showed that the SOM-OPELM approach with DirRec strategy lead to accurate solar predictions. Deo *et al.*[143] proposed an Extreme Learning Machine (ELM) based computationally efficient data-driven forecasting approach for diffuse and global, Very Short-Term Reactive (VSTR) Ultraviolet Index (UVI) forecast. The model utilized the solar zenith angle data to identify real-time sun protection behavior recommendations.

- **Machine Learning**

  Friedman *et al.*[144] introduced Multivariate Adaptive Regression Splines (MARS), a method for regression modeling of high dimensional data. The approach combined recursive partitioning and spline fitting while retaining the positive aspects of both. MARS procedure retained the adaptive property of recursive partitioning as it permits the recursive splitting of all basis functions in the model and generates continuous models by replacing step functions. MARS approach showed satisfactory interpretability through Analysis of Variance (ANOVA) decomposition. Quinlan *et al.*[145] presented a method for combining model based and instance based learning for prediction. The composite approach was implemented with three model learning methods including linear regression, model trees and neural networks. Results validated that the composite method achieved high prediction accuracy than either model based or instance based approaches.

  Wang *et al.*[146] proposed reconstruction of a machine learning method M5 for inducing model trees from empirical data. A public domain scheme was provided to induce models from data that involved continuous classes. The advantages of the method were compact model design and efficient dealing with enumerated and missing attribute values[147]. Zou *et al.*[148] introduced elastic net, a regularization and variable selection method. The elastic net is a generalization of Lasso approach which is a suitable tool for feature selection and model fitting. The given approach generated sparse model with high prediction accuracy and effective grouping effect. The elastic net offered low misclassification errors on microarray data and performed automatic gene selection.

  Moghaddamnia *et al.*[149] used Gamma test in combination with different nonlinear modeling techniques to predict daily solar irradiance from meteorological

parameters. Extraterrestrial radiation, daily average temperature, daily maximum temperature, wind velocity, and precipitation were used for prediction of solar radiation. The required data size was calculated using M-Test. Local Linear Regression Model (LLR), Elman Neural Network (ENN), Multilayer Perceptron (MLP), Neural Network Autoregressive Model with Exogenous Inputs (NNARX), and Adaptive Neuro-Fuzzy Inference System (ANFIS) with LM algorithm were analyzed. It was concluded that the performance of ANFIS model was not satisfactory for solar prediction but LLR and NNARX were very much suitable for this research area. Ishwaran *et al.*[150] described *R* software based Spikeslab package for implementing spikes and slab prediction and variable selection approach. Zhang *et al.*[151] investigated the problem of learning sparse representations with greedy algorithms. The work was focused on feature selection and reconstruction of the target function from noisy observations. A combination of Forward Greedy Algorithm with Adaptive Backward Step Feature (FoBa) was proposed. Results showed that FoBa gave smaller training error than forward greedy or L1 regularization problem for any level of sparsity that lead to better estimation results. For correlated high dimensional problems, weighted ridge regression offered satisfactory performance level. The Bayesian spikes and slab models could efficiently compute such estimators and effective in forecasting.

Gala *et al.*[152] applied Support Vector Regression (SVR), Gradient Boosting Regression (GBR) and Random Forest Regression (RFR) to generate a hybrid model that improve solar radiation forecast. It was observed that the forecast could be improved by aggregating daily solar radiation and generating hourly radiation values by disaggregating daily radiation using ECMWF. Then again generating hourly values using empirical curve yields the best prediction results. Persson *et al.*[153] proposed a non-parametric machine learning approach for multisite solar power generation prediction on 1 to 6 hours ahead of time horizon. The Gradient Boosted Regression Trees (GBRT) was analyzed and optimized using 5-fold cross-validation. Results showed that fitting the GBRT model on raw and normalized data result in the consistent response of the model to the scaling of variables. Further, GBRT could efficiently trace non-linear patterns between input variables and target response.

- **Genetic Algorithm**

  Orfila *et al.*[154] presented a dynamic genetic algorithm based model of solar cycle for long-term solar forecasting. Singular spectrum analysis was used to reconstruct the deterministic part of time series. Darwinian theory *et al.*[155] based evolution-

ary algorithm for natural selection was suited for non-linear time series. The obtained results showed that the technique was useful for long-term solar forecasting. Jafarzadeh *et al.*[156] introduced type-1 and type-2 Takagi-Sugeno-Kang (TSK) fuzzy systems for solar power plant prediction. The type-1 and type-2 antecedents and crisp consequents were considered. Methodology for tuning of antecedents and consequents of each model was also given. Solar irradiance and forecasted temperature was considered to implement fuzzy logic based models. The results showed that type-2 TSK models with type-2 antecedents and crisp consequents achieved high prediction accuracy and offered uncertainty range with the forecasted value of power generation. For the solar plants involved in the electricity market, information of uncertainty is important to develop good bidding strategies.

- **Sky Imagers**

  Chow *et al.*[157] presented a ground-based sky imager for intra-hour, sub-kilometre cloud shadow nowcasting and forecasting at the University of California (UC) San Diego. High cloud cover variability increased the cloud forecast error with increasing forecasting horizon. It was observed that sky imagery would be combined with satellite and numerical forecast for more accurate forecast in different forecasting horizon. Marquez *et al.*[158] presented a framework to forecast 1-minute Direct Normal Irradiance (DNI) for 3 to 15 minutes time horizon. The image processing procedures were applied including velocity field calculations, a spatial transformation of images, and cloud classification. It was observed that incorporation of stochastic learning techniques with improved cloud classification approach increased the forecast accuracy at short time horizons.

  Chow *et al.* [159] proposed a Variational Optical Flow (VOF) technique to estimate the cloud motion using a ground-based sky imaging systems. VOF determined the sub-pixel accuracy of cloud motion for every pixel. The cloud locations up to 15 minutes ahead could be estimated. The method was compared with Cross-Correlation Method (CCM) and image persistence method. Results validated the effectiveness of VOF over the rest two methods. Alonso-Montesinos *et al.*[160] presented short and medium term solar irradiance predictor that predicted three solar components including beam, diffuse, and global irradiance under all sky conditions. The Heliosat-2 method was utilized to estimate atmospheric attenuation and European Solar Radiation Atlas (ESRA) model to forecast three solar components. Jiang *et al.*[161] and Kaur *et al.*[162] analysed different moving target and feature matching algorithms including block motion estimation and Combined Local and Global (CLG) optical flow algorithm for measuring cloud movement images.

Results showed that CLG algorithm outperformed block motion estimation algorithm in terms of prediction accuracy and computation speed. Grantham *et al.*[163] presented a data-driven approach for probabilistic forecasting which accounts heteroscedasticity of solar radiation and non-parametric bootstrap. The bootstrap ensemble approach developed a full predictive density of global horizontal irradiation that lead to calibrated ensembles of one-hour forecasts. The nonparametric bootstrap idea was also presented that could be extended to autoregressive conditionally heteroscedastic models to track nonlinear dynamics of solar radiation.

- **Fuzzy Logic**

  Boata *et al.*[164] and Kaur *et al.*[165] proposed fuzzy logic based daily global solar radiation predictor. Daily clearness index was used to measure the stochastic component of solar radiation and followed the rules of autoregressive fuzzy algorithm. For input linguistic variables, membership functions were designed using fuzzy c-means clustering. Results validated that the approach convert past two days measurements in to an actual prediction of clearness index with high accuracy. Yang *et al.*[166] proposed a fuzzy information granule having time-dependent membership function to transform numerical time series into granular time series. The fuzzy inference based long-term prediction model was designed using fuzzy rule-based interpolation. Results validated the effectiveness of given approach over AR models, nonlinear autoregressive neural networks, and Support Vector Regression (SVR) approaches.

2.1.3.2 Hybrid Models

Mellit *et al.*[167] described an adaptive wavelet-network with Infinite Impulse Response (IIR) filter for predicting solar irradiance. The model obtained predicted values using historical data set and results showed that the given approach gave satisfactory performance matrix as compared to AR, ARMA, recurrent, and radial basis function networks. A high speed of convergence was the main advantage of wavelet-network model. Hocaoglu *et al.*[168] proposed a two-dimensional representation model of hourly solar radiation data. The approach provided a compact and unique visualization of data that lead to accurate forecasting using image processing methods[169]. Using the 2-D representation of data, an image model was formed in raster scan where rows and columns represented days and hours respectively. The between-day correlation along the same hour segment provided vertical correlation of the image. To evaluate the forecasting efficiency of the model, nine different linear filters with different filter tap configurations were tested and generated.

The performance of 2D forecasting method was tested using feed-forward neural network. The obtained results showed that 2D model was effective for both linear filters and neural network prediction methods.

Voyant *et al.*[170] proposed a hybrid ARMA/ANN model for global solar radiation prediction with data availability from NWP model. A hydrostatic model, Aire Limitee et Adaptation Dynamique (ALADIN) developed by Meteo-France in collaboration with ECMWF was employed to generate meteorological forecast with NWP model. After the optimization of MLP by endogenous and ALADIN forecast data, the model was combined with ARMA approach for the analysis of hourly data series. Finally, the prediction was generated using confidence intervals in order to achieve more reliable information for the end user. Mandal *et al.*[171] proposed a hybrid approach with a combination of Wavelet Transform (WT) and Artificial Intelligence (AI) techniques for one hour ahead solar power prediction. The sudden changes in time series data of solar photovoltaic (PV) power were filtered by WT. AI model captured the non-linear fluctuations in the solar PV power plant in a significant manner. The results demonstrated that the given model performed well for all seasons of a year with high prediction accuracy.

Mostafavi *et al.*[172] introduced an integrated approach of global solar radiation prediction using Genetic Programming (GP) and Simulated Annealing (SA) called GP/SA. The monthly average global solar radiation data for last 6 years was used. Different meteorological and climatological parameters were employed to formulate solar radiation. The results obtained from GP/SA approach were compared with conventional Angstrom model and revealed that the given approach was a promising solution for energy conversion and management problems. Voyant *et al.*[173] combined ARMA that accurately described clear sky data and ANN that utilized endogenous and exogenous inputs for the analysis of hourly data series. Because of its non-linear nature, ANN was suitable to predict cloudy days. Based on this consideration and utilizing best configurations of ARMA and ANN approaches, three hybrid models were suggested in the work. In the first approach, ARMA model was used for six months of summer and spring and for rest of the year ANN model was employed. The second model worked on the same concept as first but incorporated seasonal learning parameter in forecast. The last suggested approach followed the same division of models as in previous two approaches but also consider the forecast error occurred in the previous hour. The results showed that coupling of ARMA and ANN models achieved an improvement of more than 1% with the maximum in autumn season (3.4%) and minimum in the winter season (0.9%).

Chen *et al.*[174] proposed neural and fuzzy logic based solar radiation forecast model. The fuzzy logic was used to optimize the clustering technique and to reduce the number of sky

classes. The neural network based model was used for hourly and daily solar prediction using temperature and sky information from National Environment Agency (NEA). Long *et al.*[175] applied Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbour (kNN), and Multivariate Linear Regression (MLR) to develop the solar prediction model. These models were evaluated with and without considering meteorological parameters in multi-step ahead forecasting horizon. The results revealed that prediction accuracy decreased with the increase of forecasting horizon. Amrouche *et al.*[176] proposed a combination of spatial and ANN modeling for daily global solar irradiance on the horizontal plane. The National Oceanic and Atmospheric Administration (NOAA) provided the information about meteorological parameters. The results showed that ANN based daily GHI modeling achieved satisfactory performance level.

Lin *et al.*[177] developed Evolutionary Seasonal Decomposition Least Square Support Vector Regression (ESDLS-SVR) model for monthly solar prediction. To effectively handle seasonal influence, the model used seasonal decomposition approach and least square support vector regression method. The genetic algorithm was used to select parameters for the proposed model. Results showed that ESDLS-SVR performed better than ARIMA, Seasonal ARIMA (SARIMA), LS-SVR and Generalized Regression Neural Network (GRNN) models. Jimenez-Perez *et al.*[178] presented a data-mining approach based solar radiation forecast method. A clustering algorithm was used to identify different types of days in the dataset. Different classification algorithms were combined with regression algorithms to forecast hourly global solar radiation. The decision trees, support vector machine classification, support vector machine regression, and artificial neural networks were used to estimate different parameters that characterized each type of day. Regression algorithms were used to estimate the clearness index. Li *et al.*[179] proposed a simplified method with reduced exogenous inputs and without the need of solar irradiance data to predict 15 minutes, 1 hour and, 24 hours ahead solar irradiance. The models were developed using ANN, SVR, and online meteorological services with photovoltaic's historical data. A hierarchical forecasting approach was proposed in which machine learning tool at the micro level was used for each inverter prediction and combined the output at the macro level for the whole plant. The results showed that the hierarchical approach outperformed the traditional methods in terms of accuracy. Table 2.1 summarizes the observations drawn from above mentioned solar forecasting techniques.

Table 2.1: Observations from existing solar forecasting techniques (Cont.)

Author	**Year**	**Parameters**	**Technique**	**Observations**
[82], [83], [68], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93]	1999, 2000, 2001, 2009, 2011, 2012, 2013, 2014, 2016, 2017	Meteorological Parameters	Analytical modeling, empirical modeling	•Trace correlation between measured solar irradiance and different meteorological parameters to improve the forecast accuracy. •In comparison to neural networks, the analytical/ empirical models provide more generalized form suitable for various geographical locations.
[95], [96], [97], [98], [99]	2002, 2012, 2013	Historical solar irradiance profiles	Exponential smoothing	•Computation of prediction intervals, likelihood, and model selection criteria to the data from M-competition and M3-competition. •Offers state space framework that include an extended range of exponential smoothing methods with less tuning parameters.
[100], [101], [102], [103], [105]	2009, 2012, 2014, 2015, 2016	GHI, DNI and DHI	AR, ARX, GARCH, HR, GPR	•ARX model with NWP inputs and consideration of cloud cover conditions make solar power observations suitable for a horizon up to 36 hours ahead. •Collaboration of GARCH with ARMA allow consideration of the prediction interval associated with point forecast that leads to a probabilistic forecast.
[109], [108], [106], [107]	2006, 2007, 2010, 2016	GHI, Latitude	Trigonometric modelling	•Large meteorological data is not required for solar forecasting.
[110], [111], [112]	2012, 2017	Historical solar irradiance profiles	Similar day approach	•Similarity with past sub-patterns is traced and used to identify most similar solar pattern to the current day for estimating current weather conditions.
[115], [116]	2014, 2014	Zenith angle, clear sky index, GHI	Linear filtering approach	•Multidimensional models offer high prediction accuracy than two dimensional approach.

to be cont'd on next page

Table 2.1: Observations from existing solar forecasting techniques

[122], [64], [124], [64], [125], [129], [131]	1999, 2010, 2011, 2016	Atmospheric pressure, insolation level, temperature, precipitation	Multi-stage neural networks	•Multi-stage neural networks achieve low mean error as compared to single stage configuration.
[132], [133]	2013, 2016	NWP	Physical models	•Selection of optimum scale of spatial averaging reduce the variability in solar prediction.
[138], [139], [66], [140]	2013, 2016	Meteorological parameters	Support Vector Machine	•SVM models achieve relatively same performance level compared to ANN based models.
[141], [142], [143]	2014, 2016, 2017	Sea water temperature, sun spots, recorded dataset	Extreme Learning Machine	•SOM-OPELM approach with DirRec strategy achieve high solar prediction accuracy.
[144], [145], [146], [148], [150], [151], [152], [153]	1991, 1993, 1997, 2005, 2009, 2010, 2011, 2016, 2017	Meteorological parameters	Machine Learning (Regression modelling)	•Machine learning approaches efficiently trace linear and non-linear patterns between input variables and target response.
[154], [156], [172]	2002, 2013	Recorded solar irradiance, temperature	Genetic Algorithm	•Singular spectrum analysis and evolutionary algorithms estimate the uncertainty range which is useful to develop bidding strategies.
[157], [158], [159], [160]	2011, 2013, 2015	Cloud images	Sky Imagers	•Integration of sky images with stochastic techniques and satellite images will improve the prediction accuracy.
[164], [166]	2012, 2017	Clearness index	Fuzzy Logic	•Perform better than regression models but results are probabilistic.
[167], [168], [170], [171], [173], [174], [175], [176], [177], [178], [179]	2006, 2008, 2012, 2013, 2014, 2016	Meteorological parameters	Hybrid Models	•Collaboration of different forecasting models lead to high and stable prediction accuracy than independent approaches.

2.2 Node Level Power Management

Raghunathan *et al.*[180] reviewed various strategies to overcome the limited energy problem in wireless sensor networks including low power Medium Access Control (MAC) protocols, hierarchical architectures, energy aware sensing, and energy harvesting. Raghunathan *et al.* discussed their operation strategies, network requirements, and summarize their advantages and disadvantages. Yick *et al.*[1] presented a review on WSN including internal platform, underlying operating system, communication protocol stack, network services, provisioning, and deployment. Issues on communication architectures, security, and management were considered.

Gilbert *et al.*[181] provided a survey on different energy harvesting sources and different energy conversion approaches for making them suitable for wireless sensor nodes. The estimate of energy requirements by a sensor nodes on practical platform was focused. Anastasi *et al.*[16] surveyed different solutions of energy conservation in wireless sensor networks. The characterization and interactions between different protocols were presented and investigated that integration of different energy conservation approaches improved the results. The sparse sensor network architecture was discussed and concluded that efficient use of collector nodes by communication protocols made the network more robust. Alippi *et al.*[59] presented a classification and review of different techniques of energy management for sensor node. A framework for adaptive sensing strategies and adaptive duty cycle approach based data acquisition strategies were introduced. It was observed that activity based adaptive sampling techniques became more energy efficient by using spatio-temporal correlation but require complex computations.

Sudevalaym *et al.*[33] surveyed energy harvesting based sensor systems including architecture, energy sources, and storage technologies. The implications of node recharging on sensor network design and node level operations were presented. Banos *et al.*[61] provided a review on optimization algorithms for design, planning, and control problems in renewable and sustainable energy scenario. One class of optimization methods was based on traditional approaches including linear programming, Lagrangian relaxation, quadratic programming, and Nelder-Mead simplex method. The recent research involves heuristic optimization methods especially, particle swarm optimization, genetic algorithms[182], and pareto optimization technique. Kausar *et al.*[183] provided a review on scope, challenges, and approaches of energy harvesting for wireless sensor nodes. Various projects conducted by different research communities were reviewed and discussed their advantages and weaknesses. The need of low power sensors, ultra-low power RFID circuitry, and advanced micro machined embodiments was observed. Khan *et al.*[184] surveyed

different battery driven and energy harvesting based approaches for wireless sensor networks. It was observed that consideration of energy provisioning and efficient energy consumption profile is necessary for effective energy management schemes and efficient network wide operations. Energy provisioning approach summarized the energy source characteristics and developing algorithms. It included battery driven, energy harvesting, and energy transference based schemes. Energy consumption schemes included data driven, adaptive duty cycling, and mobility based energy management approaches.

Bhatnagar *et al.*[185] focused on low power system applications including various energy harvesting techniques, power conversion, and characteristics of mini and microscale self-sustaining power generation systems. Ahmed *et al.*[186] reviewed recent research in the field of energy harvesting communications and presented a survey on resource allocation in the area of energy harvesting. The issue of limited network life span with random energy arrival was outlined and scope of research was summarized. Babayo *et al.*[187] presented a review on energy management schemes and proposed a classification according to different application requirements. It was investigated that 41% of energy management schemes were throughput efficient, 33% of them enhance the data rate and 16% of them take duty cycle as performance metric. Shaikh *et al.*[188] presented a review on energy harvesting techniques that grasp energy from ambient sources to power wireless sensor networks. A survey of energy harvesting mechanism, hardware configurations of harvester, efficiency of harvester to capture, and ambient energy conversion was presented. Various energy prediction models were also surveyed that could lead to adaption of energy harvested WSNs while replacing the conventional ones.

2.2.1 Energy Allocation Approach

2.2.1.1 Dynamic Programming

Schal *et al.*[189] presented a dynamic programming based Markovian decision model for average optimality condition with compact action space and general state space. The average cost was considered as main criteria. The optimal inequality was shown in terms of minimum average cost. A relative value function was defined in terms of relative compactness in a real-valued state space. The simulation results showed that the pointwise relative compactness could be done in a set of real numbers if the generalized lower limit of a function was considered. Fu *et al.*[190] presented a dynamic programming approach for optimal energy allocation to a single satellite in earth orbit. For each unit of energy, the satellite encounters different demand levels, financial rewards, and different

atmospheric conditions. The authors were focused to maximize the reward by optimizing the energy consumption. For general case, suboptimal heuristics were provided that depended upon certainty equivalent control. For the special case where demand was unlimited, closed-form optimal solution was proposed. A value function was induced and proved by concavity property. This parameter lead to reduced computation time and provide scalability.

Ho *et al.*[191] proposed Generalized Markovian (GM) model to analyze harvested environmental energy which was based on empirical measurements and utilized Bayesian Information Criteria (BIC) to evaluate the suitability of the proposed model for harvested energy. Simulation results showed that for solar energy, a first-order stationary model was suitable. To model burst and erratic nature of piezoelectric energy, the generalized model was adequate. Ho *et al.*[192] investigated the problem of throughput maximization in point to point wireless communication powered by ambient sources over a finite horizon. Energy allocation problem was solved by considering the channel conditions and time-varying nature of energy sources. Two types of information regarding the harvested energy and channel conditions were assumed to be available including causal information of past/present time slots and full information about past, present and future time slots. Dynamic programming and convex optimization approaches were used for optimal energy allocation. A first-order Markov process was used to model harvested energy and channel conditions. Simulation results showed increased throughput per slot with given energy allocation schemes.

2.2.1.2 Linear Programming

Moser *et al.*[193] introduced multiparametric linear programming based formal model to precompute application parameters offline including constraints concerning buffer size, rates and timings of energy arrivals with respect to environmental conditions, and present system state. The control theory laws were integrated with software design to increase the robustness of adaptive power management. A hierarchical control design was proposed in which lower layer showed the role of power saving techniques in conjunction with upper control layer which prevented the node from running out of energy. The simulation results showed that by dividing the control theory into two subproblems significantly reduce the computational complexity. Noh *et al.*[194] proposed linear programming based two solar energy allocation algorithms to meet the requirement of fixed demand with variable supply. The basic and advanced expectation models were formulated to estimate the energy available for each time slot in advance. The first Simple Solar Energy Allocation (SSEA) scheme used a basic expectation model with low resource allocation cost and designed for

a resource-constrained node. The Advanced Solar Energy Allocation Scheme (ASEA) was based on advanced expectation model for more precise energy allocation and designed for a node with sufficient resources to support additional computation. Simulation results showed that both resource allocation schemes lead to maximum utilization of harvested energy with minimum variability and upgrade the network performance.

2.2.1.3 Channel Capacity

Goldsmith *et al.*[195] calculated the capacity of fading AWGN channel under average power constraint and different channel side information conditions at the transmitter and receiver both, and at the receiver alone. When channel information was known at the transmitter and receiving end, adaptive transmission scheme used the water-pouring technique for power adaption and a variable rate multiplexed coding scheme. Gaussian signalling formulations for transmission over a Gaussian fading channel was employed. The results showed that these techniques lead to high channel capacity with low complexity than nonadaptive transmission using receiver side information only. Sadeghi *et al.*[196] discussed the principles of Finite State Markov Channel (FSMC) modeling of fading channels for wireless communication systems. The relationship of FSMC parameters with flat fading and frequency selective channel statistics was presented. The proposed FSMC model was used to estimate channel state, equalization, and decoding techniques at the receiving end. The authors reported that the number of FSMC states, the amplitude in FSMC model, memory order, and state transition probabilities affect the information rate loss through actual and physical fading channel.

Ozel *et al.*[197] addressed the issue of stochastic energy arrivals in Additive White Gaussian Noise (AWGN) channel. It was observed that the capacity of AWGN channel with mean power constraint was equal to the average recharge rate. Two capacity achieving schemes were introduced including save and transmit and best effort transmit scheme. The first achieved upper bound of recharge rate and relied on zero code symbols appended to a portion of total block length which were neglected when block length becomes larger. The second approach was the best effort transmit scheme in which code symbols were transmitted on the channel when sufficient energy was available and a zero was put on channel instead in less residual energy availability case. The results showed that second scheme gained transmission rate very close to the channel capacity. Further, the formulation was extended to achieve offline optimal power allocation for maximum average throughput when recharge rate was not ergodic.

2.2.1.4 Traffic Control Task Scheduling

Dovrolis *et al.*[198] discussed Guaranteed (G) service class with an assumption that user can select their traffic parameters called token bucket parameters in order to reduce reserve rate in the network. Two schemes were suggested for this purpose. One was for an optimum solution called OptRate but with complex mathematical formulations. The second was the MaxRate scheme with a simple solution and was suitable for the multicast session. In this approach, every user requested different bounds of delay and different path related parameters. Farkas *et al.*[199] addressed dynamic range coordination problem, distributed power control techniques, trade space, and power-aware scheduling approaches for multiuser detection based receivers. The Defence Advanced Research Project Agency(DARPA)[200] Interference Multiple Access (DIMA) platform was used to formulate Power-Aware Scheduling (PAS) and Power Control (PC) algorithms. Escolar *et al.*[201] proposed an idea of programmed sensor node with alternative scheduling plans and introduced an algorithm for the selection of relevant scheduling strategy to maintain energy neutral condition and improve quality of service. A re-optimization approach was also introduced to adapt present energy conditions. For simulation, the energy generation was modeled from MSX-005F solar cell and energy consumption with MicaZ sensor node. An algorithm for optimal scheduling plans was proposed to optimize the quality of service of the sensor and sink while maintaining energy neutral condition. The sensor activities were considered as tasks and their cost was defined in terms of energy required by the task per unit time. The level of quality of service and energy production by a solar cell were also modeled using dynamic programming. Based on these mathematical models, scheduling strategies for different tasks were derived in order to maximize the quality of service while keeping the node energy neutral.

2.2.1.5 Data Aggregation

Kalpakis *et al.*[202] and [203] focused to trace an efficient way of data gathering and transmission by energy constrained sensors in order to maximize their life cycle. A polynomial time algorithm was given to solve Maximum Lifetime Data Aggregation (MLDA) problem. The approach worked efficiently for small networks but computationally complex for large sensor networks. The clustering based heuristics strategies were also proposed to prolong the lifetime of large data gathering and aggregation networks. The results showed that MLDA performed 1.15 to 2.32 times better than existing data aggregation procedures for small sensor networks. For large sensor networks, an increase by a factor of 2.61 was observed by clustering based approaches. Baek *et al.*[204] formulated

compression and aggregation approaches for large-scale sensor networks, sampled on a spatial-temporal field. The reduction in energy expenditure was targeted by reducing the network traffic load using distributed source coding with header compression and data aggregation approaches and secondly, by the use of energy efficient routing and distributed medium access control. A hierarchical model was also proposed with multiple sinks, aggregation, and compression nodes. Optimal organisation of devices was well implemented by Johnson-Mehl tessellation with their locations. The average cost associated with such organisations was evaluated and a comparison was made with proximity-based organisations using stochastic geometry. The results revealed an energy saving of 8% to 28% by using optimized aggregation and compression structures for sensor networks.

Nuggehalli *et al.*[205] designed a transmission schedule to maximize battery life while considering constraints including deadline and average delay constraint. A channel coding approach was utilized for reduced power level transmission that resulted in extended transmission durations. Further, an electrochemical mechanism in batteries was adapted that allowed energy recovery during idle periods but require idle transmitter slots. To balance these two conditions the bursty traffic load should be distributed during idle periods that lead to extended battery life. Seyedi *et al.*[206] proposed adaptive transmission policies to maximize the likelihood of detecting and reporting the events by sensor nodes. To select a relevant transmission mode, such policies require the knowledge of current energy level of node, data generation rate, and battery recharge rate. Markov Decision Process (MDP) was adapted for transmission scheduling and comparative analysis was done with energy balancing and aggressive policies. The simulation results showed that MDP formulation achieved better performance than rest of two policies. In some conditions including the number of dead slots and an average number of consecutive messages, energy balancing policy outperformed the other two.

Antepli *et al.*[207] formulated the problem of transmission completion time minimization offline for a given number of bits per user in the energy harvesting broadcast link. The Additive White Gaussian Noise (AWGN) broadcast channel was considered where the end user meet the random amount of energy at random time instants and no one to one correspondence between transmission power and rate point. It was observed that for optimum scheduling, average rates should be proportional to the number of bits and transmission to all users started and finished at the same time. To handle this condition, flow right algorithm[208] which was an iterative offline algorithm and was adopted by the authors with different parameters. Ozel *et al.*[209] considered an energy harvester transmitter for the optimization of a point to point data transmission under the constraints of wireless fading channel and limited battery capacity. The dynamic programming ap-

proach was used to maximize channel throughput and minimize transmission completion time in a fading channel with random energy arrivals. A water filling algorithm was also introduced to formulate the optimality conditions and control the time sequence of power transmission with respect to residual energy level and causality constraints.

Castiglione *et al.*[210] addressed the issue of energy allotment over data acquisition and transmission for energy harvesting sensors. Energy management policies were presented for a single sensor with the aim of minimum distortion and a stable queue connecting a source and channel encoders. Energy allocation between source acquisition and data transmission was performed by considering the current status of channel SNR, data queue, and statistics of energy harvesting process. A framework for multiple sensors and time division scheduling techniques was also introduced to minimize average distortion and stability of all data queues at all sensors. Results validated the effective resource allocation for energy neutral operation. Reddy *et al.*[211] considered transmission and reception power management issues in energy harvested WSNs including power loss of the circuitry, battery inefficiency, and its storage capacity under energy neutral constraint. A dual stage power management algorithm was designed in which outer stage scheduled the power that was utilized by the inner stage algorithm to maintain energy neutral state and long-term energy utility. The inner stage optimized the communication parameters for the case of Rayleigh scatting channel.

Yang *et al.*[212] considered packet scheduling problem in energy harvesting wireless sensor networks for a single user. The transmission rate was adaptively selected with respect to traffic load and available energy in order to reduce the transmission delay. The offline scheduling strategies were formulated with the assumption that energy harvesting time and rate were known in advance. Causality constraints were also considered on data/ energy arrivals and channel fading in formulations. Tutuncuoglu *et al.*[213] considered the transmission policies to maximize the amount of data transmitted in a finite time horizon and minimize the transmission completion time.

2.2.1.6 Modulation Approach

Cui *et al.*[214] analysed different modulation techniques in order to achieve minimized energy consumption for a particular number of bit transmission with error correcting codes and to gain required bit error rate. Different parameters including constellation size, optimized transmission time, coding gain, and bandwidth expansion were considered. The simulation results revealed that in Multiple Frequency Shift Keying (MFSK) systems, coding could reduce energy consumption only for large transmission distances

and desirable in power limited applications. The uncoded form of MFSK performed better than coded MFSK for short range applications in terms of bandwidth and energy efficiency. The trellis coded M-ary Quadrature Amplitude Modulation (MQAM) was energy efficient in coding and efficiency increased with transmission distance.

2.2.1.7 Routing Strategies

Cruz *et al.*[215] formulated the bounds on delay, throughput, and buffering requirements with a particular routing strategy for a communication network operating in packet switched mode. To characterize the traffic, burstiness constraint was considered that was said to be satisfied by data if the quantity of the data stream at any time instant was less than a value that depend upon the length of that interval. The authors also considered communication network models with interconnections of network elements and proposed a method to calculate burstiness constraints for traffic between network elements. An idea was also given to employ regulator elements within the network that resulted in smaller delay and large throughput. Xu *et al.*[216] proposed Geographical Adaptive Fidelity (GAF) algorithm with an aim to reduce energy expenditure in wireless sensor networks. The approach was to identify nodes with same routing probabilities and switch off the unnecessary nodes while maintaining routing fidelity. GAF approach was independent to network routing protocols and required system level and application level information. The approach was applicable to intermediate nodes of the network while the source and sink nodes were always kept on. The proposed approach was analysed over ad hoc on Demand Distance Vector (AODV) [217][218] and Dynamic Source Routing (DSR) protocol[219]. The results showed 40% to 60% less energy consumption than unmodified routing protocols. Younis *et al.*[220] and Sharma *et al.*[221] proposed clustering based energy aware routing of gathered data. The approach was to design a centralized gateway node that acts as network manager and set routes to deliver data based on the energy expenditure and residual energy level of every node. The gateway node also monitored the delay and throughput among the node cluster and judge medium access among the sensor nodes. An algorithm was also proposed to adopt network topology dynamically in the cluster to reduce energy consumption in data transmission and consumption.

Chang *et al.*[222] addressed network lifetime problem by optimizing routing algorithms with linear programming formulations. It was observed that Minimum Total Energy (MTE) approach was not effective for efficient utilization of transmission energy. The shortest cost path routing approach was introduced where link cost comprised of energy consumption during transmission/reception and residual energy at two end nodes. The results revealed that approach gain near-optimal performance for constant information

gathering rates as well as random ones. Bhattacharjee *et al.*[223] introduced an energy efficient load balancing routing scheme that considered node cost in packet transmission from source to sink. The node cost included energy efficiency and overhearing cost to locate energy efficient next hop node and residual energy node for even distribution of load in order to balance energy consumption. The approach distributed the traffic load among different energy efficient paths rather than relying on a single path for packet transmission and an improved end to end delay. A comparison was presented with shortest path routing protocol, shortest path aggregation tree-based routing protocol[224], power efficient data gathering/aggregation protocol[225], and distributed energy balanced routing[226]. The results validated the effectiveness of proposed protocol in terms of network lifetime, an end to end delay, number of alive nodes, and energy efficiency.

Voigt *et al.*[227] proposed two routing protocols to perform solar radiation aware routing. The first solar aware protocol was based on local information which was a simplified version of directed diffusion. The solar aware directed diffusion was the second suggested protocol which was an extension of standard directed diffusion approach. Different network topologies were suggested where traffic was routed via solar-powered nodes and some nodes could transmit/receive without consuming limited battery resources. The results showed that the first protocol was suitable for small sensor networks because it was not capable to trace a path for one hop away. The second protocol was effective for large sensor networks but its efficiency was dependent upon various network parameters. It was observed that by designing transmission rates adaptive and making routing protocols energy aware, a significant energy saving in energy harvested wireless sensor networks is achieved. Sharma *et al.*[228] obtained mean delay optimal and throughput optimal energy management policies and identified a greedy policy in low SNR regime that make the system to work under energy neutral condition. While simulation, the effects of fading channel, battery leakage/inefficiency, and energy loss due to data processing and sensing were included. Simulation results showed that energy storage feature in the node lead to high stability and low average delay.

2.2.2 Network Life Bounds

Based on the analytical evaluation and network simulation on energy efficiency, Bhardwaj *et al.*[229] derived upper bounds on sensor network lifetime under different scenarios. The response of energy source at a particular location, the number of nodes, energy status, energy efficiency of node, and path loss characteristics were considered for evaluation. A technique was proposed for bounding network lifetime by partitioning the

problem into subproblems where bounds were easy to derive and showed the impact of the source region, source behaviour, the number of nodes, base station location, radio energy parameters, and path loss. The calculated network lifespan bounds were allowed to calibrate network strategies and protocols regularly aimed to increase overall network lifetime. Bhardwaj *et al.*[230] derived the upper bounds on sensor lifetime by combining the networks with relevant topologies and source movement. The bounds were formulated using Feasible Role Assignment (FRA) approach where an FRA must be assigned to every bit received by the base station as it allow sensing in a non-redundant manner. An integration of continuous density Hidden Markov Model (HMM) with a Generalized Fuzzy Model (GFM) was introduced for solar irradiance prediction. Interdependency of different meteorological parameters including sunshine hours, temperature, humidity, atmospheric pressure, and wind speed were utilized for solar irradiance estimation. The Viterbi algorithm generated hidden states to partition the measured meteorological data into clusters by calculating correlation coefficients which were further modelled by GFM. A test sequence was allocated to a proper cluster by implementing log-likelihood ratio test and output was predicted using these trained clusters. The results showed an optimized lifetime bounds for data gathering sensor networks.

Rahimi *et al.*[231] proposed an approach of energy management and equalization using mobile nodes to increase network life cycle. The basic idea was to employ some autonomous nodes as mobile robots which move in search of energy, recharge themselves, and come back to the main service area to charge immobile energy depleted nodes. The Robomotes were used as the mobile nodes that send communication signals to the network and get the reply to locate themselves. Results showed that by the inclusion of about 40% mobile nodes in the conventional sensor networks make the networks self-sustaining. Giridhar *et al.*[232] proposed analytical solutions of a linear program for sensor network topologies. A linear array and a planer circularly symmetric network were considered. An upper bound on the functional lifetime of linear networks was derived as a function of data quantity and initial energies. The results showed that strategy of transmitting the data to its nearest neighbor in the direction of collector lead to optimal network lifetime in scaling sense. Kansal *et al.*[233] proposed a framework for formulating a relationship between network lifetime and achieved distortion performance. It was observed that different rate allocations had different energy expenditure that depends upon network topology. The authors derived this relation in terms of rate-distortion bounds and formulated heuristics to evaluate these bounds. Finally, a low complexity approximation of network lifetime was proposed with respect to computed distortion.

Jeong *et al.*[234] presented practical design issues of micro-solar power system over a

long time scale and considered different environmental conditions including geographical location, time, and present weather. The proposed design was divided in six functional parts including solar harvester, external environment, energy storage, input regulator, output regulator, and load. These components were correlated in terms of energy usage, operating range, and energy efficiency.

2.2.3 Storage Element Analysis

Chiasserini *et al.*[235] and Singh *et al.*[236] proposed battery management techniques with the aim to reduce energy consumption. A framework was presented to model battery characteristics/behaviour and trace the dynamics of recovery. The gain for stochastic pulse discharge was also formulated and generated by discharge demand processes. The discharge process was adjusted by a shaping technique to improve the energy efficiency of the battery. If the battery level dropped below a certain threshold, the discharge process was interrupted at the terminal side to recover the charge and enhance the performance. Results showed that when the parameters of discharge shaping algorithm were matched with storage cell parameters, battery performance improved significantly. Zhao *et al.*[237] presented an eScan approach to access updated status of sensor networks in terms of resources and running application activities, deployed in an uncertain environmental condition. Global Positioning System(GPS) was used to track the location of the node and made use of Advanced Power Management(APM) or Advanced Power Configuration and Power Interface (ACPI) to measure residual energy level. Energy efficient in-network aggregation approach was designed for scanning residual energy status to estimate the remaining energy distribution in the network. At each aggregation step, all local scans were autoscaled in terms of resolution and combined to make a composite scan. An idea of incremental updates was also proposed where a node send a partial update to a scan when its local state changes. The results showed that aggregation of individual scans did not carry detailed information about the residual energy level of the individual node but a compact representation of information reduced the processing and communication cost. Jiang *et al.*[238] introduced Prometheus, a multi-stage energy transfer system that combined positive attributes of different energy storage elements and utilized the intelligence of microprocessor to gain uninterrupted network operation. The proposed work was implemented on solar energy operated Berkeleys Telos mote and considered two-stage storage system using supercapacitors as a primary buffer and lithium rechargeable batteries as a secondary buffer. The relationship between system components and application specific optimal hardware specifications was also discussed. The results showed that system runs for 43 years with 1% load, 4 years under 10% load, and

1 year under 100% load. Mora-Merchan *et al.*[239] introduced a simulation tool, Mtossim (TinyOS Simulator), to estimate the battery lifetime and power consumption over long-term analysis. Received Signal Strength Indication (RSSI) attenuation was estimated by considering radio power consumption and battery discharging in the emitter. The model was also applicable to study of coverage range, localization, and time estimation between network maintenance.

2.2.4 Adaptive Duty Cycle Approach

Sinha *et al.*[26] proposed an operating system based Dynamic Power Management (DPM) technique for energy harvested sensor nodes. The DPM reduced power consumption by controlling the active and idle period of the node. It was observed that while effective sleep state transitioning lead to substantial energy saving, additional energy could be saved by optimizing the performance of node in the active state. In the given approach, processors voltage and frequency were dynamically adjusted to match the workload and processing requirements. A workload prediction algorithm was also introduced which was based on adaptive filtering of the previous workload and made a comparative analysis of previous filtering techniques. The workload information from three different processors over different day times was taken. The results showed that maximum power saving was achieved when processing rate was kept equal to the average load. Kansal *et al.*[240] presented harvesting theory for energy harvested wireless sensor networks. A model was designed to elaborate the characteristics of environmental sources. Harvesting theorems were formulated and proved to summarize the conditions for energy neutral mode of operation. The derived concepts were directly implemented in practical scheduling techniques. The harvesting theory was implemented on an embedded sensor system that was powered by solar energy. The results showed that the network automatically adjusted its performance level with respect to the status of energy resources. A task scheduling approach was also proposed that considered spatiotemporal characteristics of ambient sources and achieved improve energy efficiency in the distributed systems.

Raghunathan *et al.*[241] discussed various issues in the designing of solar powered wireless harvesting system and their effect on harvesting efficiency. The difference between solar powered and battery operated networks was also summarized. The essential design features of a harvesting module was enlisted to allow harvesting aware power management in other components of the system and to compute the effectiveness of harvesting based operation compared to battery operated operation in terms of the life cycle of a node. Finally, a practical demonstration was done on the design strategy, implementation, and

performance of Heliomote, a solar energy harvesting module for Berkeley/Crossbow motes and validated the applicability of proposed methodology.

Kansal *et al.*[58] analyzed complex time-varying nature of environmental sources with analytically tractable models for prior knowledge of energy availability. Different issues in enabling the energy neutral state of operation were discussed while considering the characteristics of energy sources and loads. A theoretical framework for performance optimization under energy neutral mode of operation including storage inefficiencies was proposed. A methodology was given to dynamically adapt the real-time energy availability using adaptive duty cycling approach. It was observed that the integration of link layer and routing methods with energy harvesting technologies lead to a practical adaption of energy neutral mode of operation in various applications. Niyato *et al.*[242] proposed a queuing analytical model to analyze different sleep and awake strategies in a solar power operated wireless sensor network. A linear battery model with relaxation effect was used to model battery recharge process and a stochastic process was used to model solar radiation process. The game theoretic formulation was used to generate optimal parameters for a particular sleep and wakeup strategy. The Nash solution was obtained for equilibrium point between sleep and wakeup probabilities. The packet blocking and dropping probabilities were considered as performances matrix for the proposed work.

Vigorito *et al.*[243] presented a control theory based adaptive duty cycle mechanism to achieve energy-neutral operation, performance maximization, and duty cycle stability. The closed-form optimal control laws were applied to formulate linear quadratic tracking problem in which dynamics of the system were linear whereas noise, controls of the system, and cost function to be minimized were quadratic in the outputs. Results showed that because of battery centric nature of the algorithm, sensor node recovered from energy depletion states and lead to constant computational time and constant storage requirements. A two-third variance reduction in duty cycle profile was observed in the results. To keep energy demands and supplies in balance, Gu *et al.*[244] introduced Energy Synchronised Communication (ESC) as a transparent middleware between data link layer and network layer that control the amount and timing of RF activity at receiving nodes. ESC was implemented on MicaZ node using TinyOS platform in nesC. A cross-traffic delay model was designed for the individual node to observe the effect of the low duty cycle. The working schedule of the node was adjusted to reduce the cross traffic delays with changing duty cycles. It was observed from the results that ESC can reduce delay and increase delivery ratio while synchronizing sensor node activity with available energy. Merlin *et al.*[19] proposed control theory based adaptive duty cycle approach

for low listening MAC protocols in wireless sensor networks. The work was focused to respond bursts of packets caused by the sudden occurrence of an event. The Dynamic Duty Cycle Control (DDCC) approach was introduced to control the duty cycle in order to achieve the target rate of packet transmission while consuming minimum energy. The simulation results showed that DDCC approach achieved 20% more energy saving and a significant reduction in the number of dropped packets as compared to fixed duty cycle approach.

Byun *et al.*[245] proposed duty cycle control through queue management under variable traffic rates. The approach was based on control theory where a feedback controller adopted the node sleep time according to the traffic rates while maintaining the queue length at a predetermined value. To make synchronization between active time slots of sensor nodes, an active pattern based synchronization scheme was also presented. Results revealed that the given scheme achieved high power saving with minimum delay by fast adaption of duty cycle with respect to changing traffic patterns. Jaggi *et al.*[246] designed an adaptive sensor activation algorithm that adjusted the subsequent sleep intervals with an additive increase and multiplicative decrease concept while considering sensors current energy level. Results showed that the algorithm increased the event detection probability and offered high stability with respect to current energy level and sleep interval variation.

Tadayon *et al.*[247] designed Sensor Medium Access Control (SMAC) mechanism for environmentally powered WSNs. The performance of the proposed model was analyzed for three test cases including throughput of SMAC in energy harvested WSNs, calculation of energy packets with respect to throughput, and optimization of the duty cycle to achieve long network lifespan and high quality of service. As the energy packets in the node was a decreasing function of duty cycle, the work met the sensor lifetime requirements by setting maximum and minimum duty cycle in order to achieve the quality of service at the same time. Valera *et al.*[248] introduced dynamic wakeup scheduling scheme for sensor networks that was based on Bit-Reversal Permutation Sequence (BRPS). This approach evenly distributed the wake-up slots and reduced the sleep latency. An Expected Transmission Delay (ETD) metric was formulated that considered wireless link quality and sleeps latency simultaneously. The simulation results showed that BRPS used with ETD achieved high packet delivery ratio with low end to end delay. Peng *et al.*[249] proposed a real-time adaptive Prediction Free Energy Neutral (P-FREEN) power management approach to implement Budget Assigning Principles (BAPs) with respect to energy harvesting rate and battery residual energy level. The BAPs were focused to maximize the sensors average duty cycle by minimizing the energy loss due

to storage inefficiencies. Results showed that the given approach reduced the variations in residual energy level that improve the battery life cycle. Valera *et al.*[250] addressed the issue of bulk data transfer in energy harvested WSNs while considering the duty cycle constraint. A TinyOS implemented Pump-and-Nap packet forwarding technique was presented which was based on adaptive feedback control to approximate optimal packet length for transmission and reception. A controller was proposed to compute prior input/output observations to estimate capacity allocations and accordingly duty cycle usage and to compensate wireless link quality variations. Simulation results showed that the approach achieved high throughput while maintaining required duty cycle. Table 2.2 enlists different power management techniques for wireless sensor nodes and summarizes the drawn observations.

Table 2.2: Observations from existing node level power management approaches(Cont.)

Author	Year	Parameter	Technique	Observations
[189], [190], [191] [192]	1993, 2003, 2010, 2012	Optimal cost function,	Dynamic programming	•Optimal energy management policies with the aim to maximize the throughput and minimize the mean delay.
[193], [194]	2010, 2011	Buffer size, energy rates, timings	Linear programming	•Precompute different network parameters for different working conditions. •Optimized utilization of harvested energy and minimized variability in energy allocation.
[198] , [199], [201]	1997, 2014	Quality of service	Traffic control, Task scheduling	•Reduce reserve rate in the network and controlled end-to-end delay. •Selection of relevant scheduling strategy to maintain energy neutral condition and improve quality of service.
[195], [196], [197]	1997, 2008, 2010	Channel capacity	Gaussian signalling, FSMC modeling	•Capacity of a AWGN channel under stochastic energy arrivals and different channel conditions.
[202], [204], [205], [206], [207], [209] , [210], [211], [212]	2003, 2004, 2006, 2010, 2011, 2012	Average distortion, stability, time constraint, battery capacity	Data aggregation	•An efficient way of data gathering and transmission by energy constrained sensors in order to maximize their life cycle.
[214]	2005	Bandwidth and energy efficiency	Modulation approach	•Minimized energy consumption for a particular number of bit transmission with error correcting codes and high bit error rate.
[215], [216], [220], [222], [228], [223]	1991, 2001, 2002, 2004, 2010, 2013	Delay, throughput	Routing strategies	•Load balanced routing scheme that considered node cost in packet transmission from source to sink.
[229], [230], [231], [232], [233], [234]	2001, 2002, 2003, 2005, 2005, 2012	Network life bounds	Analytical modeling	•Calibrated network strategies and protocols to improve the network life span.
[235]	2001	Storage element analysis	Analytical modeling	•Estimated battery lifetime over long-term analysis.

to be cont'd on next page

Table 2.2: Observations from existing node level power management approaches

[237], [238] [239]	2002, 2005 2013		e-scan approach TinyOS simulator	
[26], [58], [242], [248], [249],[250]	2001, 2007, 2009, 2013, 2014, 2017	Adaptive duty cycle	Sleep and wake strategies	•Adjusted working schedule of individual node for changing duty cycle with respect to energy availability.
[243], [19]	2007, 2010		Control theory	•Adaptive duty cycle mechanism to achieve energy-neutral operation, performance maximization and duty cycle stability.
[244], [247]	2009, 2013		MAC protocols	•Low listening MAC protocols for adaptive duty cycle control.

2.3 Research Gaps

From the literature survey and observations, the following research gaps are drawn:

- In the first part of literature survey, different approaches for solar irradiance forecasting are analysed. It is observed that hybrid techniques leads to high prediction accuracy but still there is a need for more precise estimation of solar energy as it is becoming an alternative to conventional energy sources. The main focus of the proposed work is to achieve accurate approximations of future energy, mainly in different forecasting horizons as it is essential for effective power management in various activities of a network.
- In the second part of the literature survey, different power management approaches for energy harvested wireless sensor networks are studied and analysed. It is observed that integration of energy harvesting technology with the conventional network designs is the upcoming research area. Specially, the issue of efficient power management is gaining more importance because of stochastic nature of ambient sources. There is a need to standardize existing systems with harvesting technologies. The proposed work consider this issue and introduced energy efficient power management approach by mathematically modeling different node parameters.

By considering the importance of optimized environmentally powered wireless sensor networks, the objectives of the work have been defined and enlisted in section 1.8.

Chapter 3

Development of Modified Pro-Energy Prediction Algorithm for Solar Energy

The growth of solar power based systems is anticipated to continue as it is an emission free, clean energy source. It avoids global warming and pollution that is caused by fossil fuel based power generation methods. The utilization of the Photovoltaic (PV) systems make it possible to deploy solar systems in small scale residential to large scale industries.

3.1 Concept of Solar Photovoltaic

The solar PV system is used to convert solar energy in to electricity[251]. It converts the photon (light) energy in to electric voltage using the photovoltaic material[252][253]. The basic concept has been shown in Figure 3.1. The photon has high energy in sunlight and possesses short wavelength[254]. It causes the electron of photovoltaic material to be free from its atom. When an electric field is added, the electron comes to a metallic contact and proportional electric current is created[255][256].

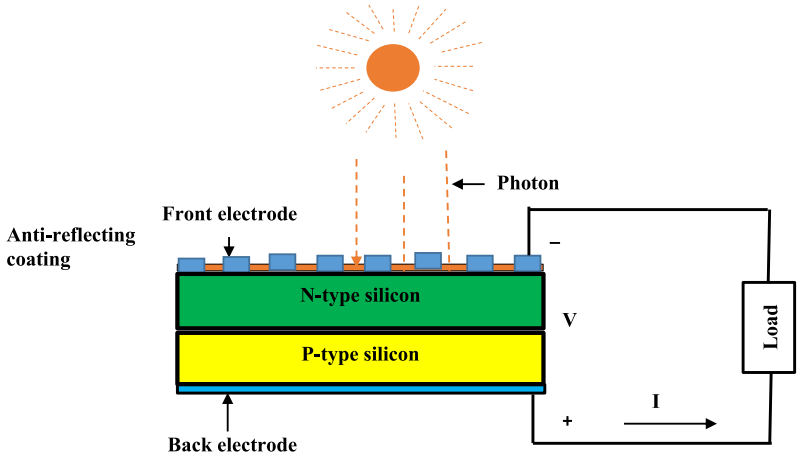


Figure 3.1: Electricity generation by photovoltaic effect

The proposed work has been published in **Journal of Renewable and Sustainable Energy, 9(3), pp. 033701, 2017**.

Advantages of PV systems

- The PV plants have low operational and maintenance cost as no conventional fuels (coal, gas etc.) are required for operation[257][258].
- No harmful or toxic gases are produced by solar PV plants. Thus, electricity generation is eco-friendly and reduces global warming[259].
- Silicon and glass that have been used for the designing of PV modules can be recycled. This feature helps in decreasing the demand of energy for producing such materials and clean the environment.

3.2 Solar Fundamentals

The solar energy reaches on the earth in the form of electromagnetic radiation. The sun and other stars act as nuclear reactors to synthesize the hydrogen atoms together for the generation of helium atoms. A large amount of energy is released in this process. The present section summarizes the basics of solar radiation that are necessary to model the solar energy profiles.

3.2.1 Solar Irradiance and Solar Power

The irradiance produced by the sun to the earth in the form of electromagnetic waves is called solar irradiance and measured as power density (W/m^2). For the measurement of solar power, the effective collector area of the device is also considered. It is measured as received solar irradiance by solar conversion module multiplied by the effective collector area of the module ($W/m^2 * m^2 = W$)[260].

3.2.2 Components of Solar Irradiance

Solar irradiance is visible, ultraviolet, and near to infrared energy in wavelength ranging from 300 to 3000 nm. According to World Meteorological Organization (WMO), solar irradiance has three basic components[261] and illustrated in Figure 3.2. When solar radiation passes through upper atmosphere, its energy decreases to 1367 W/m^2, called solar constant. When it reaches to atmosphere, some energy is absorbed and scattered by dust, air molecules, and clouds called Diffuse Horizontal Irradiance (DHI). The solar radiation that reaches directly on the surface of PV modules without scattering and

absorption is called Direct Normal Irradiance (DNI). The solar irradiance that is reflected by ground is called ground reflected solar radiation.

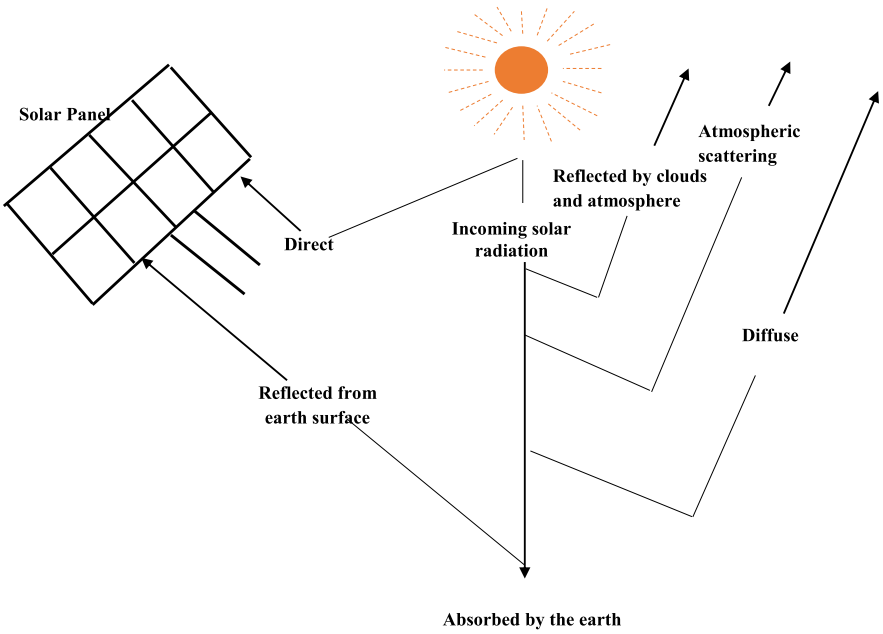


Figure 3.2: Components of solar irradiance

3.2.2.1 Direct Normal Irradiance (DNI)

The amount of solar radiation received by per square meter horizontal earth surface which is perpendicular to the sun rays is called direct normal irradiance. These rays reaches to the surface in a straight line from the sun without any atmospheric losses. DNI is measured by Pyrheliometers with acceptance half angles ranging from 2.5° to 5°. For high concentration factor, the aperture of solar concentrating collectors will be less than Pyrheliometers acceptance angle. This parameter is important for solar installations including Concentrated Photovoltaic Systems (CPV) and concentrated solar power systems[261].

3.2.2.2 Diffuse Horizontal Irradiance (DHI)

The amount of solar radiation received by per square meter horizontal earth surface in an indirect path from sun and is scattered or diffused by atmospheric components including cloud particles, aerosol particles, air molecules, or other components is called diffuse horizontal irradiance. To measure DHI, pyranometer is used that consider location and time dependent correction factor in measurement and correction[261].

3.2.2.3 Global Horizontal Irradiance (GHI)

GHI includes Direct Normal Irradiance (DNI) from the sun, Diffuse Horizontal Irradiance (DHI) from the sky and ground reflected radiation. It is considered as the total amount of shortwave irradiance received by a surface horizontal to the ground. Practically, ground reflected radiation is insignificant compared to the DNI and DHI component. Thus, for all photovoltaic solar installations, GHI is considered as the sum of direct and diffuse radiation only and represented as:

$$GHI = DNI + DHI * \cos\theta \tag{3.1}$$

where θ is the solar zenith angle and defined as the angle between direction of sun and the overhead direction (Zenith). It is 0° vertically above the location and 90° horizontally. GHI is measured by pyranometer, mounted horizontally with a hemispherical view angle of 180°. GHI is the key parameter for comparing the radiation in different climatic zones and for the calculation of irradiance on tilted planes[261].

3.3 Patterns of Solar Radiation

Sun emits the radiation in all the directions. The energy of global solar radiation is affected by the length of path through the atmosphere. Longer distances weaken the strength of global solar radiation and vice versa. This is the reason that polar regions receives less solar radiation since sun light has to travel longer distances to polar regions than tropic areas[262].

3.3.1 Linke Turbidity(TL)

This parameter is used to model the atmospheric absorption and scattering of solar irradiance under clear skies. It represents the number of dry and clean atmospheres that are necessary to generate observed irradiance at earth surface and approximates the optical thickness of the atmosphere due to water vapour and aerosol particles relative to a clear and dry atmosphere. It denotes the transparency of cloudless atmosphere. For dry and clean sky, Linke Turbidity(TL) is equal to 1; for deep blue sky, TL is just above 1 and in turbid or polluted atmosphere, TL is close to 6 or 7. The larger the TL, large is the attenuation of solar radiation by the atmosphere[262].

3.3.2 Air Mass

To measure the path length of solar radiation, Air mass is used. The value of air mass depends upon the position of the sun. It is 0 if the radiation is in upper atmosphere and equal to 1 when sun is directly overhead. For the installation of PV systems, 1.5 is the standard value for air mass. The maximum value of solar irradiance that is received by earth surface is estimated by considering the air mass as[262]:

$$I_{max} = 1367 * 0.7^{AM^{0.678}} \tag{3.2}$$

where I_{max} is the maximum solar radiation and AM is the air mass.

3.3.3 Seasonal Variation in Solar Radiation

Summer days have more solar radiation hours than winter days. Also, high solar intensity is received during day time and becomes weak in afternoon. These seasonal variations are due to the change in relative positions of the sun and the earth as shown in Figure 3.3 [261].

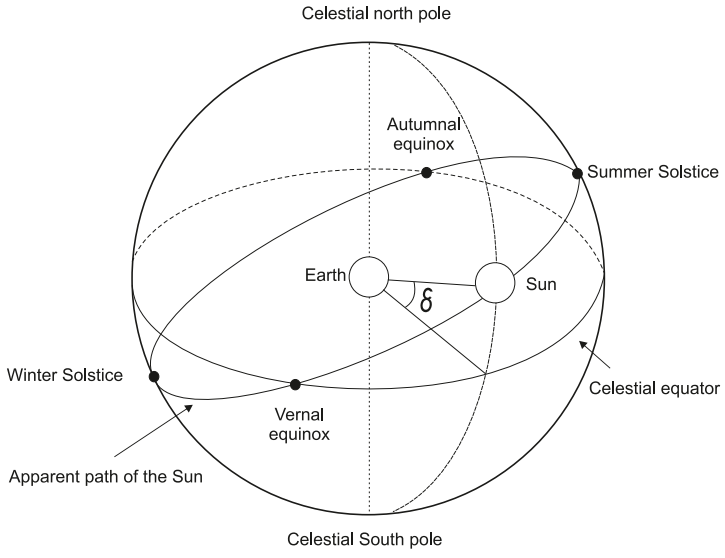


Figure 3.3: Movement of sun

The apparent path of sun and the celestial equator have an angle of 23.45° between them. The angle between the deviation of the line linking the centre of the earth and the sun with equatorial plane is called solar declination ξ and calculated as (in radians):

$$\xi = \pi \frac{23.45}{180} \sin 2\pi \frac{284 + D_{current}}{365} \tag{3.3}$$

where $D_{current}$ is the current day. The angle ξ is assumed constant within one day. For March 21 or 22 and September 22 or 23, the solar declination angle is 0° and for June 21 or 22 and December 21 or 23, the angle is 23.45°. This angular difference is the reason behind a long day in summer and a short day in winter.

3.3.3.1 Daily Variations in Solar Radiation

The apparent daily movements of the sun causes variation in solar intensity on daily basis. Figure 3.4(a) shows apparent position between the sun and earth in the celestial plane. Figure 3.4(b) depicts the occurrence of day and night because of this daily movement. The geographical latitude of the observation location on the earth is represented by θ_ϕ. The instantaneous point of the sun is represented by hour angle ϖ [261] and given as:

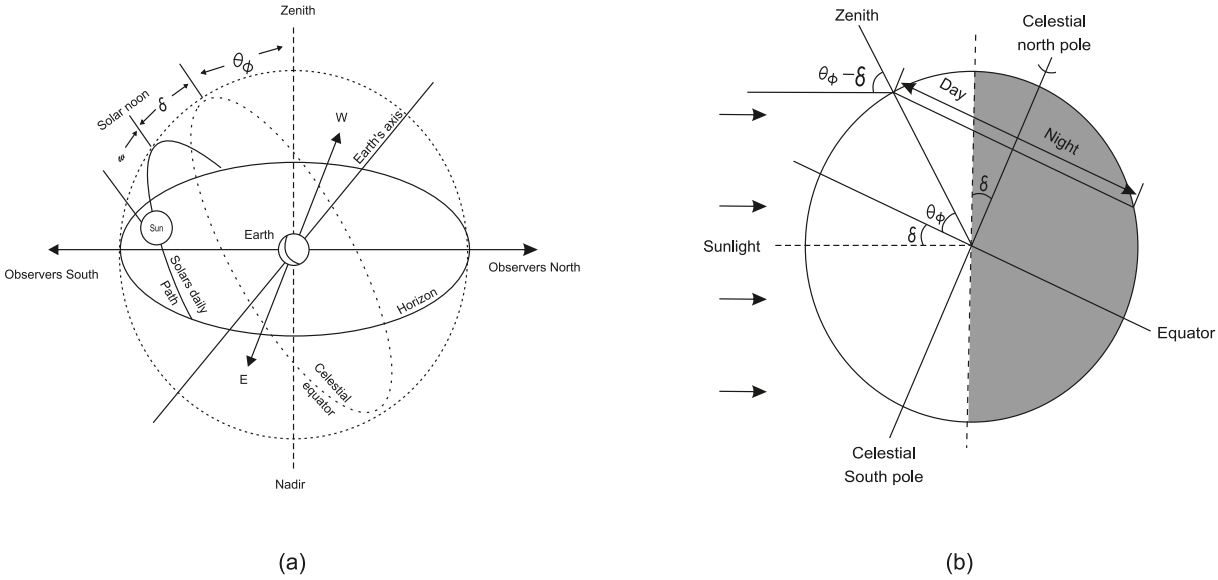


Figure 3.4: Daily variation in solar radiation

$$\varpi = 15(T_{LC} - TZ - 12) + \theta_\varphi + T_E/4 \tag{3.4}$$

where θ_φ is the longitude of the PV sites, T_{LC} is the local mean time and TZ represents the time zone. T_E is the time equation which is calculated as:

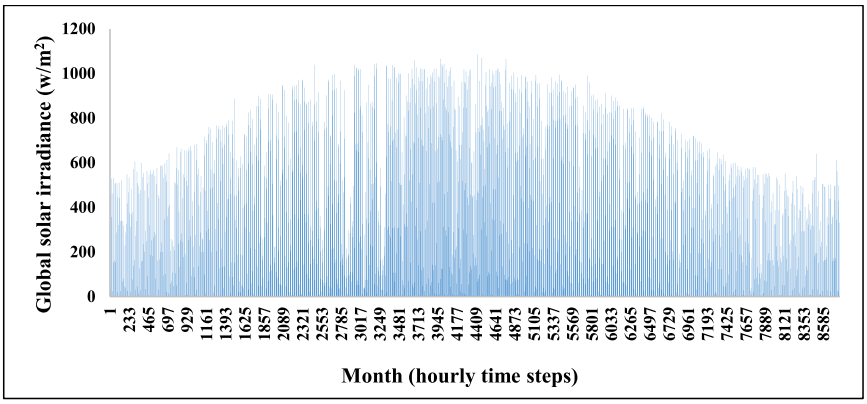
$$T_E = 229.1831(0.000075 + 0.001868\cos\theta_\varsigma - 0.032077\sin\theta_\varsigma - 0.014615\cos 2\theta_\varsigma - 0.040849\sin 2\theta_\varsigma) \tag{3.5}$$

where

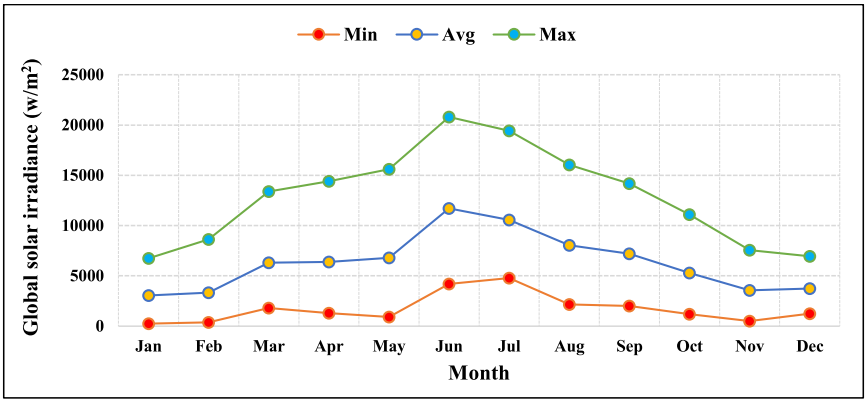
$$\theta_\varsigma = \frac{360}{354} * (D_{current} - 1) \tag{3.6}$$

3.4 Requirement of Solar Irradiance Forecasting

Solar irradiance is restrained by different meteorological conditions (atmospheric temperature, pressure, humidity, aerosol concentration, water vapour etc.), seasonal variability (time of the day, season of the year), geographical constraints (location of the sensor node, solar elevation angle, site altitude, etc.) and intra-hour solar intensity. Uncertain generation of solar energy causes variable power generation in different time scales. Figure



(a)



(b)

Figure 3.5: (a)Solar irradiance (hourly),(b) Minimum, maximum and average solar irradiance (monthly)

3.5 exhibits a typical monthly statistics based global solar radiation on horizontal surface from January-December, 2016. Dataset is adopted from Solar Radiation Research Laboratory (SRRL) under National Renewable Energy Laboratory (NREL)[263] with CMP-22 pyranometer as solar radiation sensor[264]. NREL is primary national laboratory in US for renewable energy that uses Baseline Measurement System (BMS) with latitude 39.742 north, longitude 105.18 west, and elevation 1828.8 meters with time zone GMT-7. Per minute solar dataset is obtained from NREL on horizontal plane with 1440 samples per day.

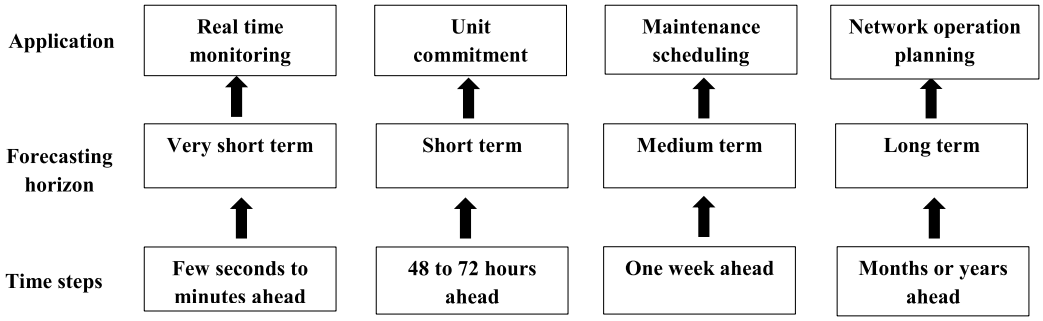


Figure 3.6: Forecasting horizon and concerned applications

In Figure 3.5(a), twelve months period is depicted with 8760 hours on x-axis and y-axis shows the variation in solar irradiance with respect to different months of the year. It is observed that intensity of solar radiation of a particular day is highly influenced by seasonal effects as well as the present day weather conditions including various meteorological parameters. To corroborate the observations from Figure 3.5(a); Figure 3.5(b) exhibits the maximum, minimum, and average solar irradiation with respect to different months of the year. It is clear from the Figure 3.5(b) that in year 2016, peak solar intensity is available during summer season (June, July, and August), moderate in spring season (March, April, and May) and autumn season (September, October, and November) and lowest in winter season (December, January, and February). It is also observed from Figure 3.5(a) that summer and winter season offer less alterations for short time periods whereas large alterations are more probable in spring and autumn season. Solar forecasting diminishes the effect of resource variability and uncertainty by targeting different forecast time horizons. With the concern of practical use, Figure 3.6 shows different forecasting horizons and related activities in solar powered WSNs. Very short term forecasting is essential for real time monitoring of battery status. Short term forecasting is critical for decision making activities including unit commitment etc. Medium term forecasting is effective for maintenance scheduling and spinning of power unit. Long term forecasting is useful in planning the network operations. Precise solar forecasting ensure reliable and stable rechargeable sensor operation with improved control algorithms for battery backup and make dynamic tradeoffs between energy consumption, system operation, and operational fidelity to attain an energy neutral condition.

3.5 Selection of Forecasting Model

Selection of a forecasting model depends on many factors including forecasting conditions, availability and applicability of data, forecasting horizon, expected forecast accuracy and time for making the analysis. These factors should be weighted constantly. With the

study of basic features and limitations of different forecasting techniques, forecasting problem can be formulated effectively with independent or blending (hybrid) of techniques. In proposed work, time series analysis based statistical approach is used for short term forecasting as it provides a solution to deal with temporal effects. Generalization is carried out by statistical tests on training dataset but relies on firm initial assumptions about data series to achieve an optimal solution. By taking this constraint in to consideration, machine learning based models are used for medium and long term forecasting. Machine learning has the features of reasoning, perception, and redundancy that make it capable to handle large dataset with various features. Also, no firm observations about data series are necessary; different heuristics are exercised for a good solution[265].

3.6 Solar Forecasting Models

For non-controllable and predictable energy sources, forecasting models are used to predict the energy availability and allow the system to take decision about the utilization of available energy. The present section aims to provide an overview of basic statistical energy prediction models that makes the ground of the proposed algorithm. The statistical models are time series analysis based mathematical models that make use of historical data to predict future values. The main focus is on past patterns, change in patterns, and use statistical equations to identify these patterns and trends.

3.6.1 Exponentially Weighted Moving Average(EWMA)

EWMA[58] is based on an exponentially weighted moving average approach and follow the assumption that amount of energy available at a given time of the day is similar to the energy observed at the same time on the previous days. Time is divided in N time slots of equal duration. EWMA predict the energy as a weighted average of the energy received at the same time over a set of previous days. In weighted moving average approach, each mean is multiplied by a pre-specified weight to follow the data series more closely. The weighted moving average assign more weight to recent data points and decreases in arithmetic progression. The weights are determined by the number of periods selected and must sum to 1. Mathematically, EWMA predicts the amount of energy in n_{th} time slot by the following equation:

$$\mu_n^{(d)} = \alpha \cdot x_n + (1 - \alpha) \cdot \mu_n^{(d-1)} \tag{3.7}$$

where x_n is the energy harvested by the end of n_{th} time slot, $\mu_n^{(d-1)}$ is average harvested

energy over the past $(d-1)$ days in their n_{th} time slot and α is the weighting factor that ranges from 0 to 1. EWMA considers the seasonal variations and adapts diurnal solar energy cycles. The results of algorithm are satisfactory for unchanging weather conditions but introduce significant prediction errors when weather is highly variable.

3.6.2 Weather Conditioned Moving Average (WCMA)

The proposed algorithm[96] is based on EWMA principle but with reduced prediction error. WCMA reduces the prediction error by taking in to account the hourly changes as well as seasonal changes in weather conditions. To this purpose, past days energy values are maintained in a matrix M of size $D \times N$ where rows represent the past days under consideration and column represents the different time slots of each day. For current day energy observation, a vector C is maintained of size N. WCMA also maintains a vector E of size N whose n_{th} entry E_n keeps the average of observed energy during the time slot n in past D days.

$$E_n = \frac{1}{D}.\sum_{i=(d-1)}^{(d-D)} M_{(i,n)} \tag{3.8}$$

E_n is updated at the end of each day with recent energy observations. Energy prediction equation for WCMA algorithm is given as:

$$M_{(d,n+1)} = \rho.M_{(d,n)} + (1-\rho).E_{(n+1)}.GAP_k \tag{3.9}$$

where $M_{(d,n+1)}$ is the predicted energy for next time slot, ρ is the weighting factor similar to EWMA, $M_{d,n}$ is the observed solar energy during time slot n, and $E_{(n+1)}$ is the average harvested energy during time slot n+1 for last D days. The GAP_k is a new weighting factor that gives information about changing weather conditions during time slot n of current day with respect to previous D days. k is number of past intervals considered. The value of k should be chosen in such a way that it is related to present weather conditions of the current day and reject those time slots that have no impact on present day energy profiles. A vector V_K is defined to calculate GAP_k and given as:

$$V_K = [v_1, v_2, v_3, ..., v_k] \tag{3.10}$$

The individual value of vector is calculated as:

$$v_k = M_{(d,n-K+k-1)} / \frac{\sum_{i=(d-1)}^{d-D} M_{(i,n-K+k-1)}}{D} \tag{3.11}$$

The calculation in equation 3.11 returns a value whose magnitude represents the difference in weather conditions of present day and previous days in a particular time slot. A value greater than 1 indicates better solar conditions than previous day and value smaller than one represents a cloudy day. Further, for better predictions, more importance is given to the elements of vector V those are close to the current time instant. To this purpose, a weighting vector P of length k is defined and individual element of P is calculated as:

$$P_k = \frac{k}{K} \tag{3.12}$$

The factor GAP_K for day d using K past values at an interval n is calculated using the equation 3.13:

$$GAP_k = \frac{V.P}{\sum P} \tag{3.13}$$

WCMA performs well when used for short-term predictions of solar energy but prediction accuracy degrades for medium and long term forecasting horizons.

3.6.3 Pro-Energy (PROfile Energy Prediction Algorithm)

Pro-Energy prediction model[110] is based on similar past energy observations to the current day for predicting energy availability for half an hour to few hours ahead prediction horizons. A day has been divided in N time slots and classified as sunny, cloudy or rainy day. The current day energy profiles are stored in a vector C of length N. A pool of past energy profiles is maintained in matrix M of size $D \times N$ where D is the number of past days considered. The stored profiles depicts the energy obtained during past D days. Pro-Energy predicts the energy availability for the next time slot by tracing the most similar profile to the current day. The similarity between the energy profiles of the current day and days stored in the pool is computed in terms of Euclidean distance between two vectors and given as:

$$E^d = \min \sum_{i=t-k}^{t} \frac{1}{k} * abs(C_i - M_i^d) \tag{3.14}$$

where E_d is the most similar profile among the days stored in pool M and k is past slots considered. The prediction of energy for future time slot is carried out using the equation 3.15:

$$C_{(t+1)}^{predict} = \alpha * C_t + (1 - \alpha) * M_{(t+1)}^d \tag{3.15}$$

where $C_{(t+1)}^{predict}$ is the predicted energy value of future time slot, C_t is the energy harvested during the slot t of current day, $M_{(t+1)}^d$ is the energy available during slot $t+1$ of the

selected day from the pool and α is the is weighting factor ranging from 0 to 1. Equation 3.15 is suitable for short term energy prediction. For few hours ahead predictions, consideration of correlation between successive time slots is important to determine the number of past time slots which are having impact on prediction of future time slots. With all these considerations, a new weighting parameter υ is used to predict energy values for F future time slots using equation 3.16:

$$\upsilon_i = \begin{cases} \alpha * (1 - \frac{i-1}{G}) & \text{if} \quad i \leq G \\ 0 & \text{if} \quad i > G \end{cases} \qquad \forall \mathrm{i}, 1 \leq \mathrm{i} \leq \mathrm{F} \tag{3.16}$$

where weighting factor υ_i is used to assign weights to past values from the pool and treated same as α is used for short term prediction, i is the position of the future time slot with respect to current time slot t, G is the number of time slots having correlation with current slot t and F is future time slots for which prediction will be carried out. The forecasting equation used for medium and long term equation is given as:

$$C_{(t+i)}^{predict} = \upsilon_i * C_t + (1 - \upsilon_i) * M_{(t+i)}^{d} \tag{3.17}$$

3.7 Development of Modified Prediction Algorithm for Energy Harvesting

Time series analysis is important for predicting a quantity which is changing over time and estimating its future values using the behaviour of past historical data. Time series analysis works on ordered sequence of observations of a quantity over equally distributed time intervals and accounts the fact that these data points have internal structure between them (autocorrelation, trend patterns, seasonality)[266][267]. These structures should be accounted to explore hidden insights of data.

Results shown in section 3.9 depicts that EWMA and WCMA have high prediction error in the different forecasting horizons; therefore, the estimated solar irradiance traces have large deviations from the real irradiance profiles. The Pro-Energy algorithm increases the prediction accuracy to some extent but still has prediction error of the order of 20.80%, 22.78%, 28.82%, and 29.69% for 30, 60, 90, and 120 intervals respectively. From the results obtained, it is observed that a further improvement in accuracy is required to predict the trace close to real irradiance profile. Hence, it is intended to develop a modified Pro-Energy prediction algorithm for energy harvesting using the time series analysis. By availing the trend and level components of a time series in solar irradiance

forecasting, the average growth of irradiance (increasing or decreasing) and a constant value for irradiance is estimated at the end of each time slot. The system model shown in Figure 3.7 is used to implement the modified prediction algorithm.

3.7.1 Description of Proposed Algorithm

The periodic nature of solar irradiance is suitable for estimating future updates using the neighbouring values. In the proposed work, 24 hours time period is divided into fixed length time slots and the solar irradiance prediction is carried out after every half an hour interval. Figure 3.8 shows real time solar irradiance profile of January 1, 2016 for the experimental development. The time duration of eight hours (7:00am to 3:00pm) is considered because most of the solar characteristics are accounted during this time interval. The initial measurements of current day irradiance profile are decomposed into the trend component using Classical Decomposition Method (CDM)[268].

The estimated trend represents a long-term change in irradiance level of the solar energy. Thus, it provides the approximation about irradiance levels for the future time slots. The calculated level and the trend components goes through the smoothing process in order to separate a fundamental composition of a data series from randomness. Based on the estimated irradiance level of future time slots for the current day, a pool of most similar days is extracted from the historical data set. The solar dataset of January 2013, 2014, and 2015 is considered as past historical data to predict January 2016. For the extraction of similar days from last three years, available time slot of considered day (January 1, 2016) is compared with 31 days of January 2013, 2014 and 2015. The past days with the Minimum Absolute Error (MAE) of the current day comprises a pool of most similar days. The size of the pool is selected on the basis of iterate simulation.

For the short-term forecasting, the irradiance profile of most similar day from the pool is used to predict the irradiance availability in future together with the recent information of the current day. But for few hours ahead forecasting, instead of relying on a single past day, a group of most similar days is used for prediction. A combined weighted profile of selected days is calculated to take consideration of large variations in the environmental conditions. The information of previous slots of the current day in conjunction with combined weighted profile of past days significantly reduces the prediction error. Finally, the performance evaluation of proposed work together with EWMA, WCMA, and existing Pro-Energy algorithm is carried out on the basis of prediction error in 30, 60, 90, and 120 minutes ahead of the forecasting horizons.

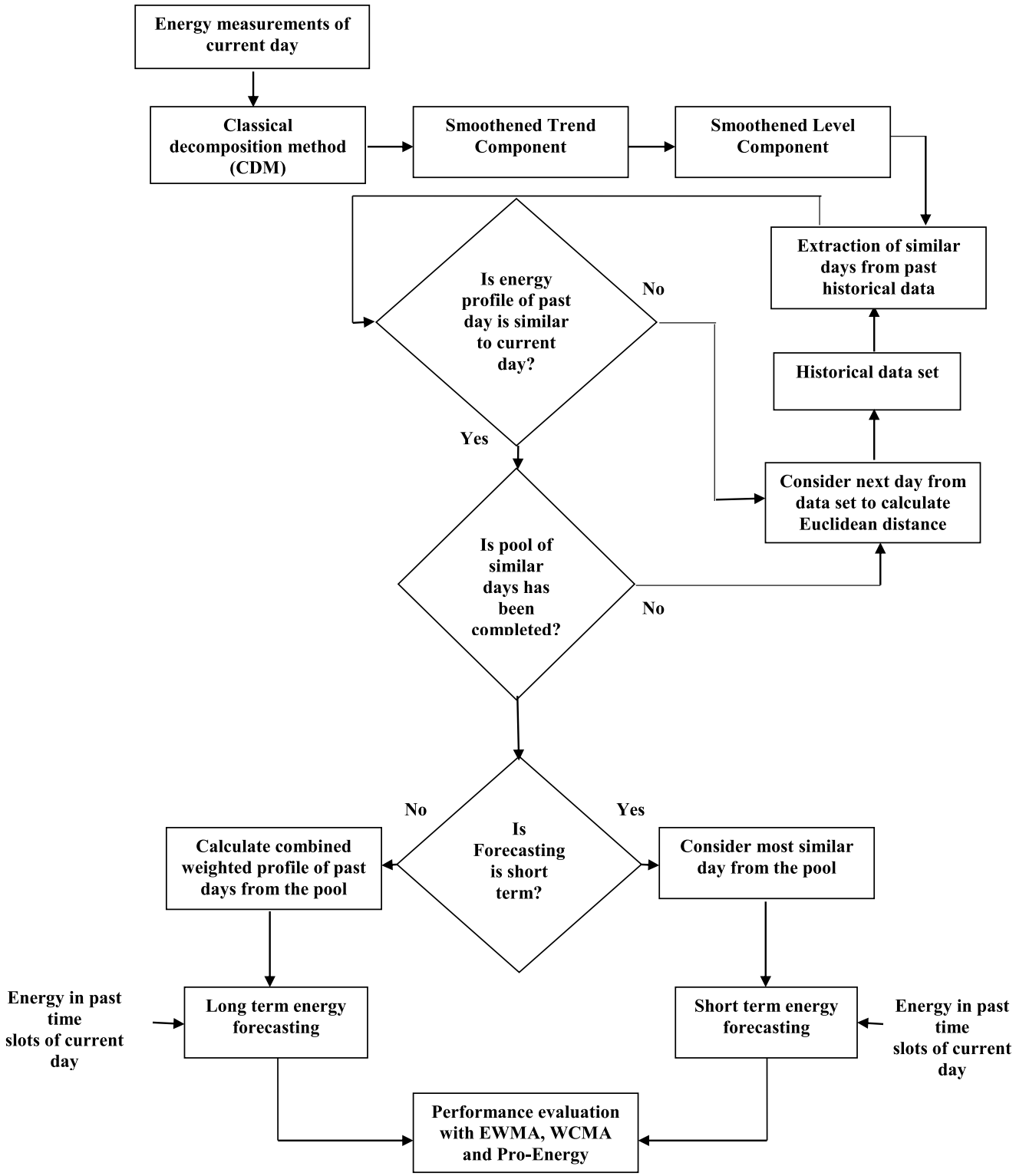


Figure 3.7: The modified Pro-Energy algorithm for solar irradiance prediction using level and trend factors

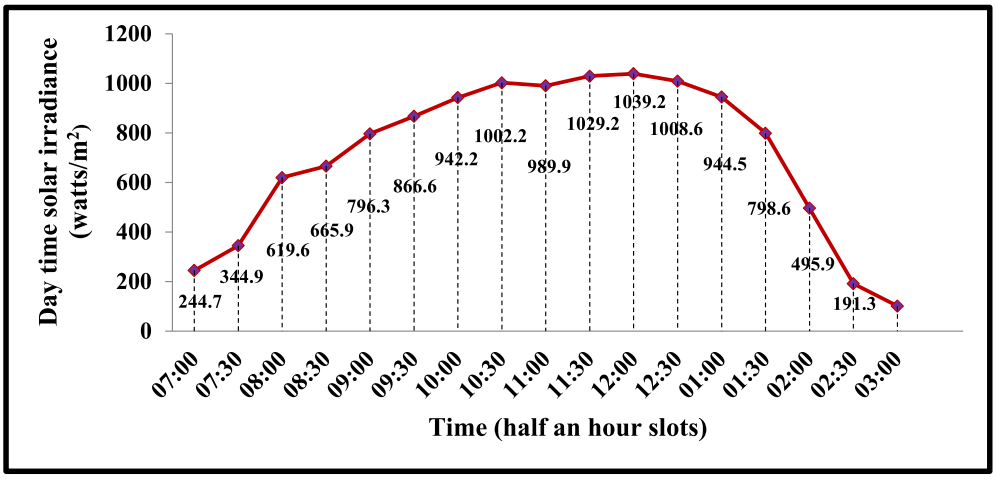


Figure 3.8: Day time solar irradiance taken at every half an hour instant

3.7.1.1 Classical Decomposition Method (CDM)

The time series analysis is based on the assumption that data points taken at different time instants have some correlated patterns. CDM[269] statistically analyses these patterns and split them into several sub patterns so that the elementary components that affect the individual values of data series can be recognized. The analysis of data points is done on the basis of the rate of change of series and emerges in important components of a time series. This technique provides an useful abstract model of a time series by splitting it in level, trend, seasonality, and random components. These individual components represents different pattern categories of a time series that lead to effective analysis of the series and improve the forecast. Mathematically, a data series composed of basic components is represented as:

$$y_t = \{T_t S_t C y_t I_t\} \tag{3.18}$$

where

y_t represents data series,

T_t represents trend of the data series and give estimated average growth(upward or downward) of solar irradiance at the end of each time period,

S_t represents the seasonality component,

Cy_t represents the periodic cycles and shows upward and downward fluctuations in solar irradiance measured over long time period. The time period of a periodic cycle is longer than a seasonal component and carries more variation than a seasonal pattern,

I_t represents the irregularities present in the solar profiles and give short term

variations that arises because of unexpected events of nature.

Basic Structure

There are basic two structures that are considered for CDM approach:

- Additive CDM
$$y_t = T_t + S_t + Cy_t + I_t \tag{3.19}$$
The additive model is suitable for a data series in which seasonal fluctuations will not vary with the level of time series and relatively constant over time.

- Multiplicative CDM
$$y_t = T_t * S_t * Cy_t * I_t \tag{3.20}$$
The multiplicative model is suitable for the data series in which seasonal fluctuations are almost proportional to the level of time series and increases over time.

Additive CDM is used for the implementation of the proposed work. The effect of individual component is estimated by the CDM to explore the behaviour of data series. The CDM is preferred among other time series methods because of its less complexity.

Trend Model

The trend component of a decomposed time series is acclimated to predict the irradiance availability in future. The selection of a trend model depends upon the nature of solar radiation. The basic radiation patterns are linear, quadratic, exponential, and s-shaped curve. If solar irradiance is increasing or decreasing at almost constant rate, it is then framed in linear trend model.

1. Initialization of Trend Component: The best fit linear equation has been calculated for the initial data set. A linear trend line has been calculated by ***Least square method***[270] for time t which is given as:
$$x' = L + Tt \tag{3.21}$$
$$T = \frac{\sum tx - (\sum x)(\sum t)/n}{\sum t^2 - ((\sum t)^2)/n} \tag{3.22}$$
$$L = \frac{\sum x}{n} - T(\frac{\sum t}{n}) \tag{3.23}$$
where x is the actual value, L is the intercept, T is the slope of the line, n is the number of time slots and x' is the forecast at time t.

To validate the above procedure, ***Backcasting***[271] approach is applied to generate

initial values for trend model. In this process, the data series is reversed and past values are forecasted instead of future values. The starting values for slope and intercept are generated by this approach. The algorithm will run in regular manner after this procedure.

2. Trend Component for Subsequent Time Slots

   The following equation generates trend component for the subsequent time slots using linear trend model:

$$T_t = (L_t - L_{t-1}) + T_{t-1} \tag{3.24}$$

   The two estimates from the trend equation are, i) how much the irradiance level is changed from last period to this period, and b) the trend estimation of the last period shows average growth rate of time series in $(t-1)$ time slot. By virtue of trend component given by Chatfield[272] and Box et al.[273], irradiance level, L_t for present time slot is calculated as:

$$L_t = x_t + (L_{t-1} + T_{t-1}) \tag{3.25}$$

   The first part $(x(t))$ of equation denotes the solar irradiance at time t. The second part is the summation of irradiance level of previous time slot and how much level should have changed from last period to this period that represents trend. Thus, equation 3.25 is not a forecast of next time slot but estimate of level for the same time period t.

3.7.1.2 Smoothing Process

In some locations, the seasonal variations are so strong that they conceal any trend and level component. It causes a reduction in the prediction accuracy. The smoothing process[274] removes these random fluctuations from the data series and provides perceptible view of series to predict its behaviour more accurately. It provides weighted estimate of level and trend of future irradiance with more importance to recent observations and exponential decrease of weights because measurements become more distant. The smoothing method can be static or dynamic and selected on the basis of the nature of data series. Figure 3.9 summarize different smoothing techniques and their applicability.

The static methods follow patterns that are stable over time. On the other hand, dynamic methods process patterns that changes over time and predictions are updated using

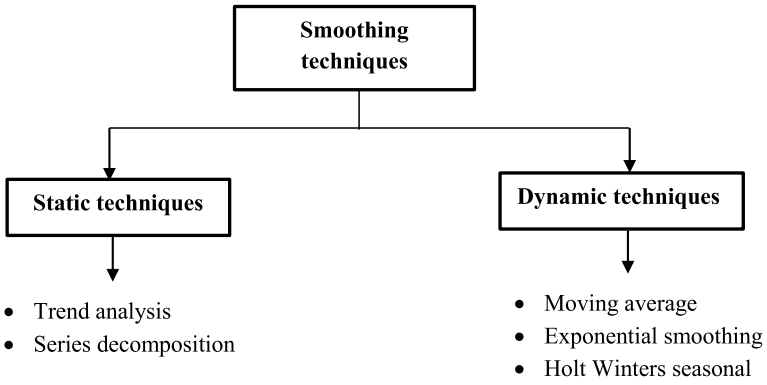


Figure 3.9: Different smoothing techniques

neighbouring values. These two methods are integrated to improve the forecast accuracy using double exponential smoothing or Holt winter's method[275]. Further, trend analysis and time series decomposition are combined to utilize different trend models offered by trend analysis.

The modified Pro-Energy algorithm is implemented using Double Exponential Smoothing (DES)or Holt′s Linear Trend Model[276] [95]. Single exponential smoothing does not trace trend factor effectively in data series because of the single smoothing parameter ρ. Double exponential smoothing make use of two smoothing parameters to update the trend and level components in each time slot. The smoothing constant ρ is used to estimate the level in the forecast and second smoothing constant β accounts the trend component in the time series. The advantage of using the double exponential smoothing method is the rates at which level and trend are tracked and can be revised. The estimates for trend and level are exponentially smoothened by parameters ρ and β and formulated as follows:

$$L_t = \rho * x_t + (1 - \rho) * (L_{t-1} + T_{t-1}) \tag{3.26}$$

$$T_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * T_{t-1} \tag{3.27}$$

where ρ is smoothing constants of level with range $(0 \leq \rho \leq 1)$ and β is the smoothing constant of trend with range $(0 \leq \beta \leq 1)$.

Optimized Smoothing Constants

The choice of smoothing constants ρ and β is done using iterative procedure called grid search in parameter space to reduce prediction error. A set of forecasts is generated using each value of smoothing constant and results are compared with real irradiance values. A value with least prediction error is chosen as smoothing constant. The preferred value

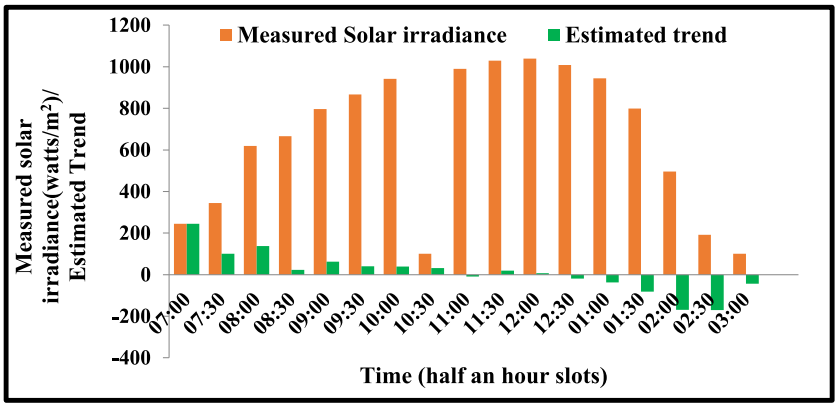


Figure 3.10: Estimated trend component of predicted solar irradiance

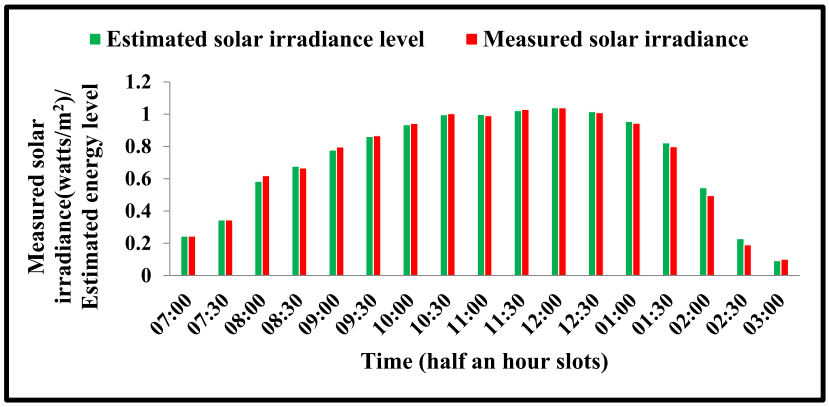


Figure 3.11: Estimated energy level of measured solar irradiance

of ρ and β is in between 0.5 to 0.9. Figure 3.10 and Figure 3.11 shows that smoothing constant ρ as 0.8 and β as 0.7 achieve optimized trend and level components respectively. The obtained trend component has positive and negative slopes to represent the variation in solar irradiance. The accurate estimation of trend component leads to close approximation of solar irradiance level in future time slots. It helps to increase the possibility of choosing most similar energy profiles to current day using level component.

3.7.1.3 Extraction of Similar Days from Past Historical Data

The historical data of 2013, 2014, and 2015 has been considered to extract similar irradiance profiles to current day. By using the equation 3.28, a pool S of most similar past solar energy profiles is maintained in matrix M of size $D \times N$ where D is the number of past days considered and N is the number of time slots.

$$Diff(d) = abs(L_{j-1} - M(d, j)) \tag{3.28}$$

where d denotes past days ranging from 1 to D (past days in the matrix M) and j denotes time slots ranging from 1 to N. L_j is the calculated solar irradiance level of current day for particular time slot and compared with one-time slot ahead of past days from the pool. The size of the pool varies with respect to forecasting horizon and decided by iteration procedure. To adapt the changing weather conditions, the pool is periodically updated. For the short-term prediction, solar irradiance profile of most similar day is considered for reference. For 60, 90 and 120 minutes ahead prediction D is taken as 14, 9 and 14 days respectively.

3.7.1.4 Solar Irradiance Forecasting

1. 30 Minutes Ahead Solar Irradiance Prediction

   Solar irradiance prediction for the next time slot is carried out by considering last time slot of current day in combination with irradiance availability in next time slot of similar past day. The estimated trend component is added to the energy profile of the past day to reduce the deviation from the current day energy profile. The equation for the short term solar irradiance prediction is given as:

$$C_{t+1}^{predict} = \alpha \times C_t + (1-\alpha) \times (d_{t+1} + T_t) \tag{3.29}$$

   where

   $C_{t+1}^{predict}$ is predicted irradiance value of future time slot,

   C_t is irradiance harvested during the slot t of current day,

   d_{t+1} is most similar stored profile up to time slot t,

   T_t is estimated trend component at time slot t,

   α is weighting factor; $0 \leq \alpha \leq 1$.

2. 60, 90, and 120 Minutes Ahead Solar Irradiance Prediction

   For 60, 90, and 120 minutes ahead predictions, consideration of correlation between successive time slots is important to determine the number of past time slots having impact on prediction of future time slots. For this reason, a new weighting factor, υ_i is used to assign weights to past values from the pool and calculated as:

$$\upsilon_i = \begin{cases} \alpha * (1 - \frac{i-1}{G}), & if \;\; i \leq G \\ 0, & if \;\; i > G \end{cases} \qquad \forall 1 \leq i \leq F \tag{3.30}$$

where

i is the position of future time slot with respect to current time slot t

G is the number of future time slots having correlation with current slot t

F is the future time slots for which the prediction will be carried out

The parameter υ_i combine the current irradiance observation with the stored irradiance profiles and works same as weighting factor α in equation 3.29 but in equation 3.30, υ_i decreases progressively when predicting irradiance for time slots that have low correlation with current time slots. Equation 3.31 depicts the basic formulation of Pearson correlation coefficient which is used to calculate such correlation.

$$r = \frac{N\sum E_{past}E_{future} - (\sum E_{past})(\sum E_{future})}{\sqrt{[N\sum(E_{past})^2 - (\sum E_{past})^2][N\sum(E_{future})^2 - (\sum E_{future})^2]}} \quad -1.0 \leq r \leq +1.0 \tag{3.31}$$

where r represents the correlation coefficient and ranges from -1 to +1. A value of 1 represents the perfect positive correlation; 0 shows no correlation, and -1 shows perfect negative correlation between data points. N represents the time slots considered. E_{past} and E_{future} is the solar irradiance at past and future time slots respectively. Table 3.1 specifies values of parameter G and υ_i for different forecasting horizons. For the time slots which exceeds the limit of G slots, weighting factor is set to be zero because of negligible correlation.

Table 3.1: Values of parameter G and υ_i for different forecasting horizons

Forecasting horizon	**60 minutes**	**90 minutes**	**120 minutes**
G	5	5	6
υ_i	0.32	0.27	0.22

3. Calculation of Combined Weighted Profile of Past Days

In case of highly unstable weather conditions, consideration of single day solar irradiance profile may results in low prediction accuracy. The consideration of multiple profiles leads to accurate predictions by taking into account potential changes in weather. The basic approach to improve medium and long term prediction accuracy of proposed algorithm is to use a set of similar days from the pool S instead of using a single similar day and combine them to form a Weighted Profile (WP) for prediction. The combined weighted profile is computed from the sorted list of past days. The day with minimum solar irradiance difference with current day is considered on priority. This feature helps to evaluate the current day for different

possible weather conditions as for every half an hour the sorted list is revised with respect to current weather condition. The similarity between the irradiance profiles of the current day and days stored in the pool is computed in terms of Mean Absolute Error (MAE) and given as:

$$E^d = \min_{E^d \epsilon S} \sum_{i=t-k}^{t} \frac{1}{k} * abs(C_i - E_i^d) \tag{3.32}$$

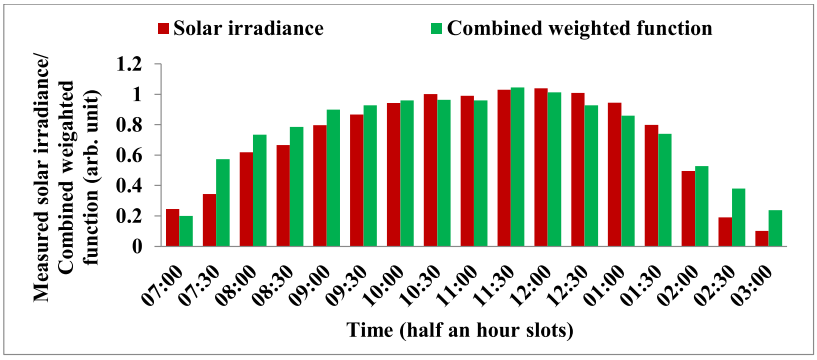
where E^d is the most similar profile among the days to the current day C from the pool S at time slot t and k signifies last time slots considered. If environmental conditions vary during current day, considering only last k observations reduce the probability of selecting a wrong irradiance profile as well as the computational overhead. MAE measures the absolute value of difference between two irradiance profiles without considering their direction. MAE is the natural measure of average errors where individual differences are weighted equally. Let E^{d1},E^{d2},..,E^{dp} is the ordered list of most similar days from the pool S i.e. days having minimum absolute error are arranged in ascending order. The combined Weighted Profile (WP) of p days for the next time slot $(t + i)$ is calculated as:

$$WP_{t+i} = \frac{1}{p-1} \sum_{j=1}^{p} w_j * E_{t+i}^{d_j} \tag{3.33}$$

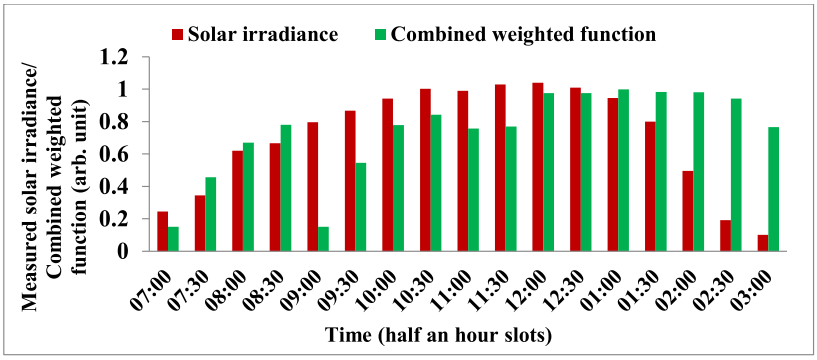
where w_j is the computed weight of individual day using the MAE of k past slots between the current day and the day selected from the pool which is given as:

$$w_j = 1 - \frac{MAE_k(E^{d_j}, C)}{\sum_{j=1}^{p} MAE_k(E^{d_j}, C)} \tag{3.34}$$

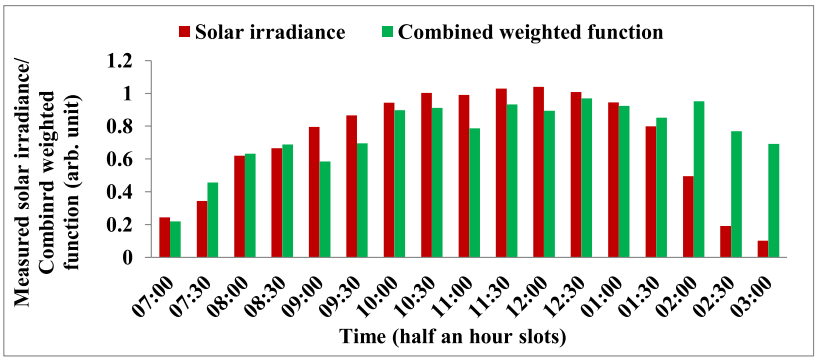
$E_{t+i}^{d_j}$ is the group of similar days from the pool where j varies from 1 to P. Figure 3.12 depicts the calculated combined weighting function in four different forecasting horizons. Solar irradiance level at each time slot of current day is shown together with the combined weighting function for that particular slot. It is depicted in the Figure 3.12 that various weather evolutions of the current day are considered using multiple energy profiles that lead to improve prediction accuracy and reduce the impact of forecasting horizon. In the proposed work, arbitrary units (arb. unit) are used to depict two different quantities on y-axis as shown in Figure 3.12.



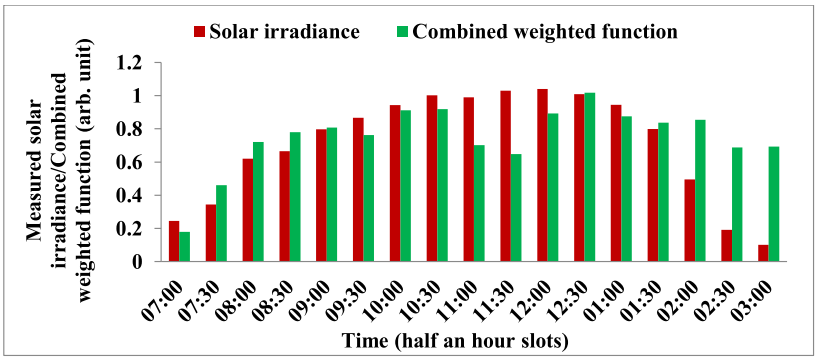
(a) Combined weighted function in 30 minutes ahead prediction



(b) Combined weighted function in 60 minutes ahead prediction



(c) Combined weighted function in 90 minutes ahead prediction



(d) Combined weighted function in 120 minutes ahead prediction

Figure 3.12: Combined weighted profile in different forecasting horizons

The equation for i period ahead solar irradiance forecasting using combined weighting function has been given as:

$$C_{t+i}^{predict} = \upsilon_i \times C_t + (1 - \upsilon_i) \times (WP_{t+i} + (i \times T_t)) \tag{3.35}$$

where $C_{t+i}^{predict}$ is the predicted irradiance and WP_{t+i} is the combined weighted function for $(t + i)$ time slots.

The simulation results in section 3.9 shows that by calculating the level and trend components of a time series together with the reference of past days, the future irradiance values of few hours ahead in time are calculated accurately. The algorithm 3.1 summarizes the procedure of proposed algorithm.

3.8 Performance Evaluation Function

The divergence of predicted irradiance value from the real irradiance intake at a particular time instant is the basis of performance evaluation of the proposed algorithm. The Mean Absolute Percentage Error (MAPE) is used as error function for computing the prediction error and given as:

$$E_{fn} = \frac{1}{N} \sum_{i=1}^{N} abs(1 - \frac{E_{actual}}{E_{predicted}}) \times 100 \tag{3.36}$$

where E_{actual} is the real irradiance value and $E_{predict}$ is the predicted irradiance value for N time slots (half an hour duration each). In order to have higher accuracy of results, the time slots where the harvested irradiance level is lesser than 10% of maximum irradiance of the day, are discarded in performance evaluation.

3.9 Results and Discussions

The performance of the proposed algorithm is evaluated against EWMA, WCMA and Pro-Energy algorithm and the results are discussed. The values for different parameters including the smoothing constants (α, ρ, β), the number of past days (D) considered to make similar day pool, and the number of past time slots considered (k) are selected where algorithms[58][96] and [110] provide their optimized results. The solar irradiance predictions are carried out for four forecasting horizons and are given as:

Algorithm 3.1 Modified Pro-Energy algorithm for solar irradiance prediction

Input parameters:

Real solar irradiance measurement in initial time slot, C_t

Historical solar irradiance measurements of past D days

output parameters: Forecasted solar irradiance, $c_{t+i}^{predict}$

- Calculation of initial values of level and trend component of the current day vector, C

$$x' = L + Tt$$

$$T = \frac{\sum tx - (\sum x)(\sum t)/n}{\sum t^2 - ((\sum t)^2)/n}$$

$$L = \frac{\sum x}{n} - b(\frac{\sum t}{n})$$

- Trend Component for Subsequent Time Slots:

$$T_t = (L_t - L_{t-1}) + T_{t-1}$$

$$L_t = x_t + (L_{t-1} + T_{t-1})$$

- Smoothing Process

$$T_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * T_{t-1}$$

$$L_t = \rho * x_t + (1 - \rho) * (L_{t-1} + T_{t-1})$$

- Extraction of Similar Days from Past Historical Data

$$Diff(d) = abs(L_{j-1} - m(d, j))$$

- Short Term Solar Irradiance Prediction

$$C_{t+1}^{predict} = \alpha \times C_t + (1 - \alpha) \times (d_{t+1} + T_t)$$

- Medium and Long Term Solar Irradiance Prediction

$$\upsilon_i = \begin{cases} \alpha * (1 - \frac{i-1}{G}), & if \;\; i \leq G \\ 0, & if \;\; i > G \end{cases} \qquad \forall 1 \leq i \leq F$$

- Calculation of Combined Weighted Profile of Past Days

$$E^d = \min_{E^d \epsilon S} \sum_{i=t-k}^{t} \frac{1}{k} * abs(C_i - E_i^d)$$

$$WP_{t+i} = \frac{1}{p-1} \sum_{j=1}^{p} w_j * E_{t+i}^{d_j}$$

$$w_j = 1 - \frac{MAE_k(E_j^d, C)}{\sum_{j=1}^{p} MAE_k(E_j^d, C)}$$

- i period ahead forecast

$$C_{t+i}^{predict} = \upsilon_i \times C_t + (1 - \upsilon_i) \times (WP_{t+i} + (i \times T_t))$$

3.9.1 30 minutes Ahead Solar Irradiance Prediction

One period ahead prediction estimates the irradiance intake for next 30 minutes duration. The table 3.2 illustrate different parameter values and manifest the performance of different algorithms for short term forecasting horizon.

Table 3.2: Different parameters for EWMA, WCMA, Pro-Energy and modified Pro-Energy for 30 minutes ahead prediction evaluation

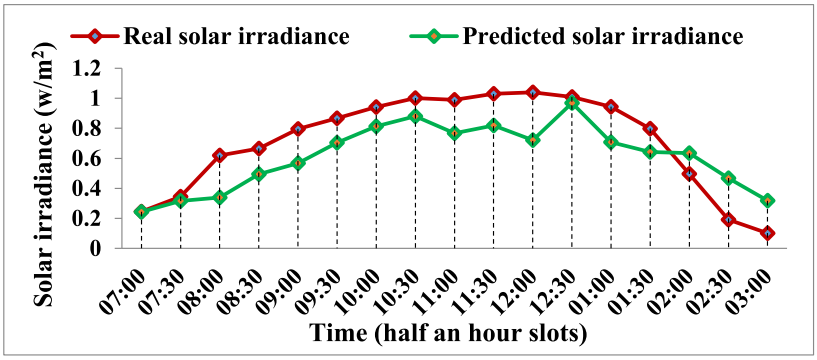
EWMA		WCMA				Pro-Energy				Modified Pro-Energy					
α	$E_{fn}\%$	α	D	k	$E_{fn}\%$	α	D	k	$E_{fn}\%$	ρ	β	α	D	K	$E_{fn}\%$
0.5	30.98	0.7	14	2	15.92	0.5	14	2	20.8	0.8	0.7	0.5	14	2	13.12

The Figure 3.13 exhibits the predicted irradiance traces with the actual irradiance harvested for half an hour time period ahead prediction. Figure 3.13(a) shows the performance of EWMA algorithm. The weighting factor, ρ is considered 0.5, and the prediction error ($E_{fn}\%$) of 30.98% is observed. Figure 3.13(b) reveals the WCMA prediction error of 15.92% that is almost half in comparison to EWMA. A pool of past 14 days (D) is used to fetch present day environmental conditions. WCMA gain optimum results by considering past two-time slot (K) of each day in pool with the weighing constant as 0.7. To evaluate the performance of Pro-Energy, a pool of past 14 day is maintained. By allocating ρ as 0.5 to previous time slots an error of 20.80% is perceived as apparent in Figure 3.13(c). It is evident from Figure 3.13(d) that the predicted solar irradiance trace of the proposed algorithm is best fit to actual irradiance values. By taking previous two time slots (K) of selected similar days and taking level factor, trend factor and weighting factor as 0.8, 0.7 and 0.5 respectively, an error function is reduced up to 13.12% which is less as compared to all considered algorithms and an improvement in prediction accuracy of 7.68% is achieved from existing Pro-Energy algorithm.

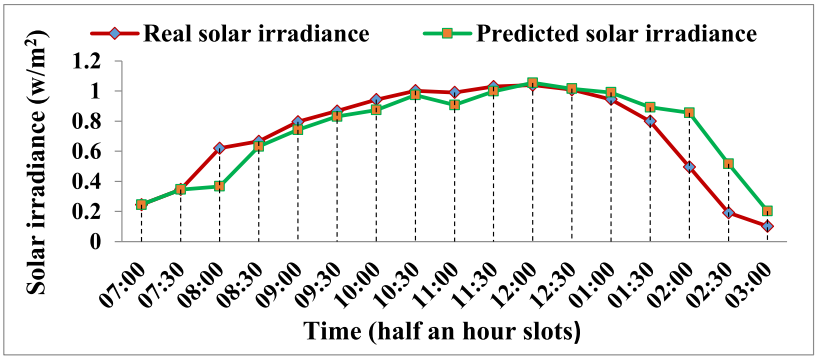
3.9.2 60 Minutes Ahead Solar Irradiance Prediction

The second prediction horizon forecasts the solar irradiance for one hour ahead time period. The concept of combined weighted profile is used for few hours ahead prediction horizons in Pro-Energy and modified Pro-Energy algorithm in order to increase the prediction accuracy. Table 3.3 illustrate different parameter values used to evaluate the performance of considered algorithms.

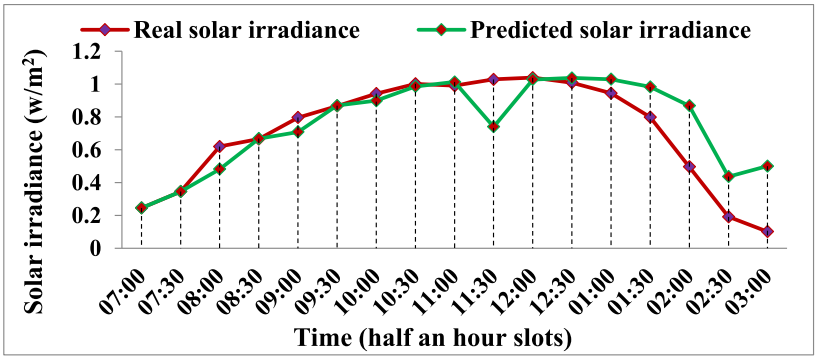
The Figure 3.14 indicates that error in estimating the future irradiance increases as we move ahead with respect to the current time slot. Figure 3.14(a) shows a relative



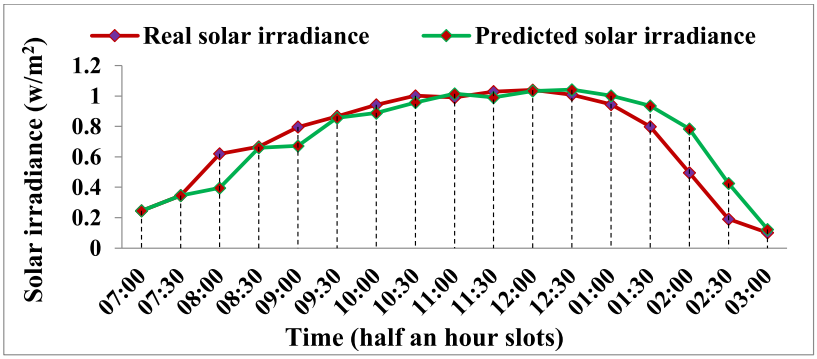
(a) 30 minutes ahead solar irradiance prediction by EWMA



(b) 30 minutes ahead solar irradiance prediction by WCMA

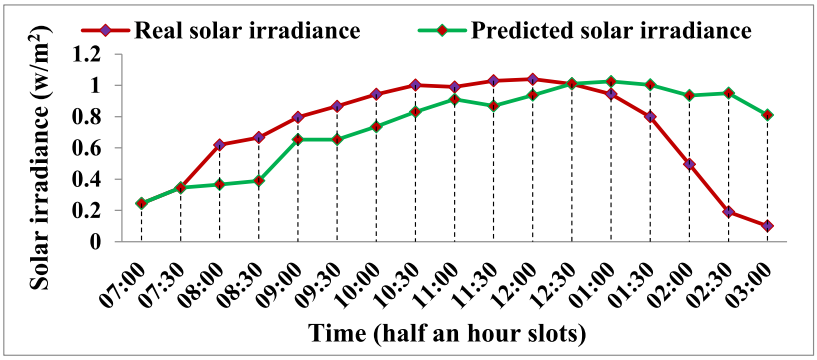


(c) 30 minutes ahead solar irradiance prediction by Pro-Energy

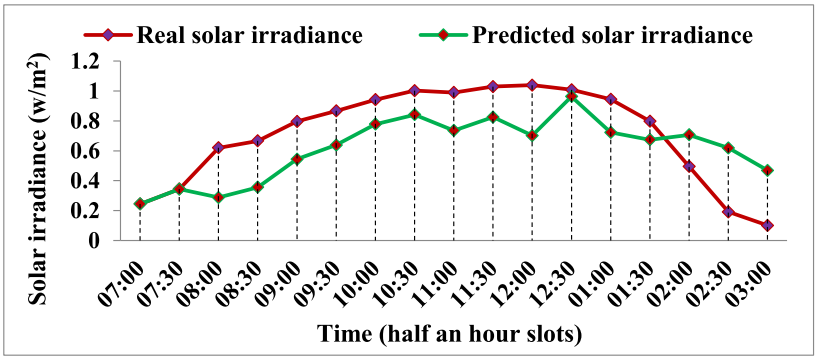


(d) 30 minutes ahead solar irradiance prediction by modified Pro-Energy

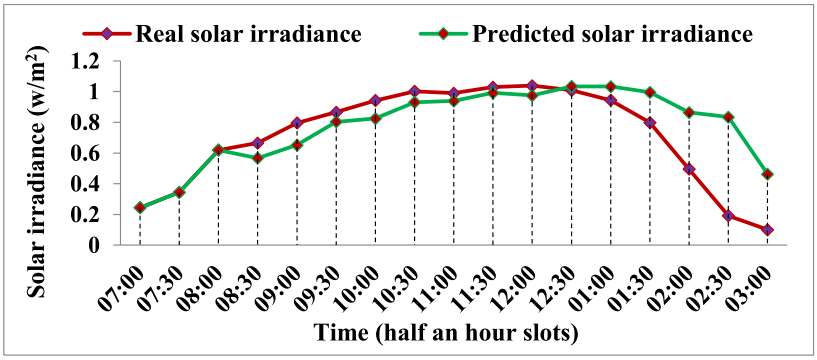
Figure 3.13: 30 minutes ahead solar irradiance prediction by different algorithms



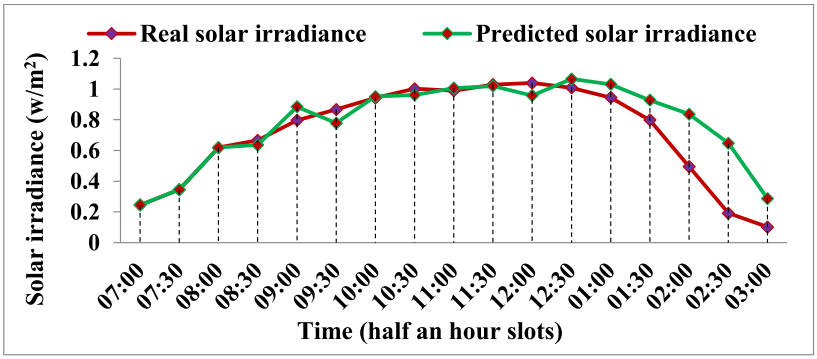
(a) 60 minutes ahead solar irradiance prediction by EWMA



(b) 60 minutes ahead solar irradiance prediction by WCMA



(c) 60 minutes ahead solar irradiance prediction by Pro-Energy



(d) 60 minutes ahead solar irradiance prediction by modified Pro-Energy

Figure 3.14: 60 minutes ahead solar irradiance prediction by different algorithms

Table 3.3: Different parameters for EWMA, WCMA, Pro-Energy and modified Pro-Energy for 60 minutes ahead prediction evaluation

EWMA		WCMA)				Pro-Energy					Modified Pro-Energy						
α	$E_{fn}\%$	α	D	k	$E_{fn}\%$	υ	D	P	k	$E_{fn}\%$	υ	ρ	β	D	P	k	$E_{fn}\%$
0.5	35.40	0.7	14	2	24.26	0.32	14	9	2	22.78	0.32	0.8	0.7	14	9	2	18.54

increase of 5% in the prediction error in case of EWMA with weighting factor as 0.5. For 60 minutes ahead forecasting, the prediction error in WCMA is 24.26% which is 8.34% more than the short term forecasting as depicted in Figure 3.14(b). In this scenario, the estimation of future irradiance intake is more accurate in Pro-Energy algorithm as compared to EWMA and WCMA. It is clear from the results in Figure 3.14(c), an error of 22.78% and a relative error increase of 1.98% have been observed in Pro-Energy which is less in comparison to EWMA and WCMA algorithms. In modified Pro-Energy algorithm, a combined weighting factor (υ) is taken as 0.32 and the prediction error is reduced to 18.54%, clearly shown in predicted irradiance traces in Figure 3.14(d). An improvement of 16.86%, 5.72%, and 4.24% is gained in modified Pro-Energy algorithm as compared to EWMA, WCMA and Pro-Energy respectively.

3.9.3 90 Minutes Ahead Solar Irradiance Prediction

The different parameter values including smoothing constants (υ, ρ, β), number of past days (D), number of past time slots considered (k), and value of error function of considered algorithms for one and half an hour ahead time is provided in table 3.4. Similar to previous test cases, in this forecasting horizon error in estimating future solar irradiance is minimum in proposed algorithm as compared to EWMA, WCMA, and Pro-Energy.

Table 3.4: Different parameters for EWMA, WCMA, Pro-Energy and modified Pro-Energy for 90 minutes ahead prediction evaluation

EWMA		WCMA)				Pro-Energy					Modified Pro-Energy						
α	$E_{fn}\%$	α	D	k	$E_{fn}\%$	υ	D	P	k	$E_{fn}\%$	υ	ρ	β	D	P	k	$E_{fn}\%$
0.5	42.58	0.7	14	2	34.78	0.27	9	7	6	28.82	0.27	0.8	0.7	9	7	6	24.52

In Figure 3.15, it is observed that the predicted irradiance traces deviate from actual irradiance values heavily when the distance increases in prediction horizons with respect to the current time. An error of 42.58% is detected in EWMA which is 7.18% more in comparison to 60 minutes ahead forecasting as shown in Figure 3.15(a). A similar response is observed in WCMA, given in Figure 3.15(b), where an error of 34.78% is noticed and a relative increase of 10.52% is obtained from the previous forecasting horizon.

Figure 3.15(c) depicts that Pro-Energy algorithm offers 28.82% prediction error with υ as 0.27 and a relative increase of 6.04% from 60 minutes ahead forecasting error. The use of the time series components (level and trend) in the proposed algorithm attain the actual irradiance traces and reduce the prediction error. The results are drawn in Figure 3.15(d) that indicate that the prediction error of 24.52% is observed and 18.06%, 10.26%, and 4.3% error reduction are achieved in the modified Pro-Energy algorithm as compared to EWMA, WCMA and Pro-Energy algorithm.

3.9.4 120 Minutes Ahead Solar Irradiance Prediction

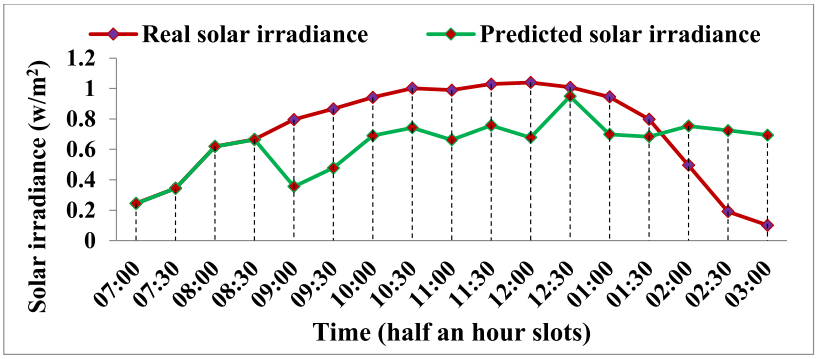
The results are drawn for two hours ahead prediction revealed that the minimum error is provided by the modified Pro-Energy algorithm in this horizon. Table 3.5 shows the different parameters and their associated values taken for simulation. The error reduction of 23.85% is observed in the modified Pro-Energy algorithm with values of υ, ρ, β, D, k as 0.22, 0.8, 0.7, 14, and 1 respectively. An error reduction of 27.33%, 23.2%, and 5.84% is achieved in the proposed algorithm as compared to rest of three algorithms. It is justified by the Figure 3.16 that the estimated irradiance values by the proposed algorithm are best fitted with the actual solar irradiance values.

Table 3.5: Different parameters for EWMA, WCMA, Pro-Energy and modified Pro-Energy for 120 minutes ahead prediction evaluation

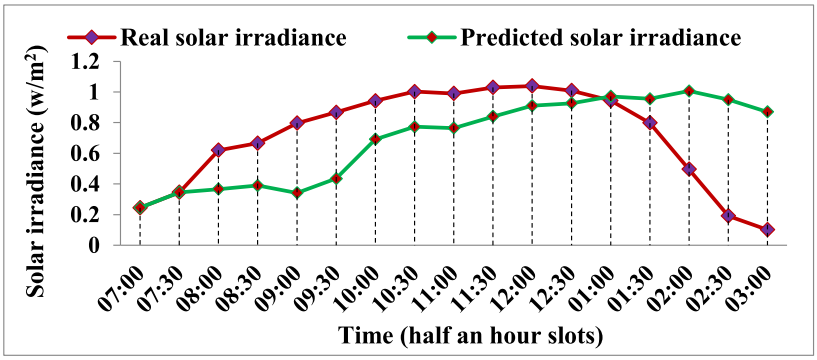
EWMA		WCMA)				Pro-Energy					Modified Pro-Energy						
α	E_{fn}%	α	D	k	E_{fn}%	υ	D	P	k	E_{fn}%	υ	ρ	β	D	P	k	E_{fn}%
0.5	51.18	0.7	14	2	47.05	0.22	14	9	1	29.69	0.22	0.8	0.7	14	9	1	23.85

The Figure 3.17 shows the average prediction errors of considered algorithms in four forecasting horizons. It clearly illustrates that the proposed algorithm consistently outperforms EWMA, WCMA and the Pro-Energy algorithm for all forecasting horizons in terms of prediction accuracy.

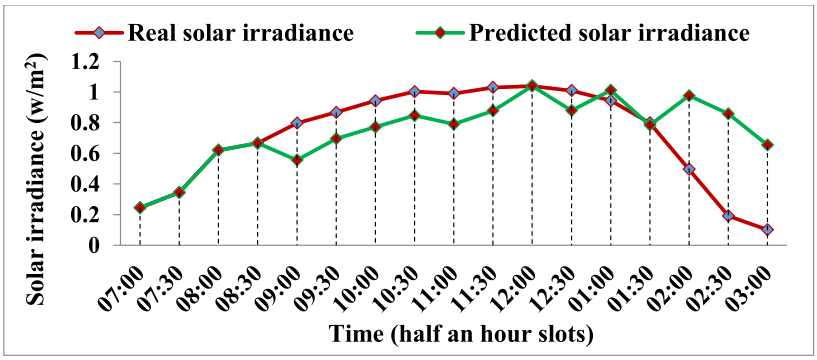
The results revealed that by calculating the level and trend components of a time series together with the reference of past days, the future irradiance values up to few hours ahead in time are calculated accurately. The weather conditions are uncertain and are best estimated by evaluating the energy profiles of similar past days. Specially, by estimating the trend of a time series, the sudden weather changes are traced more precisely.



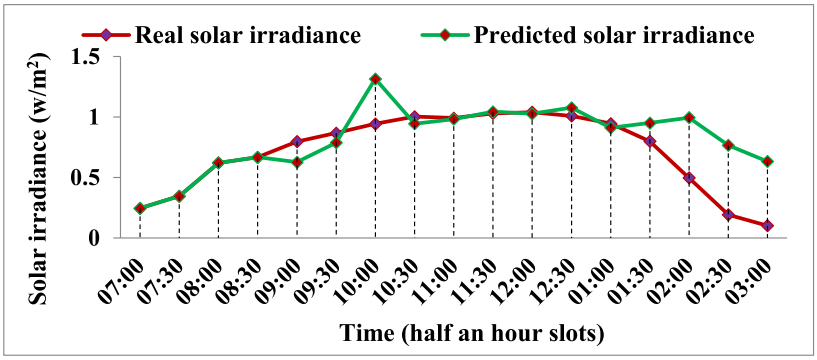
(a) 90 minutes ahead solar irradiance prediction by EWMA



(b) 90 minutes ahead solar irradiance prediction by WCMA

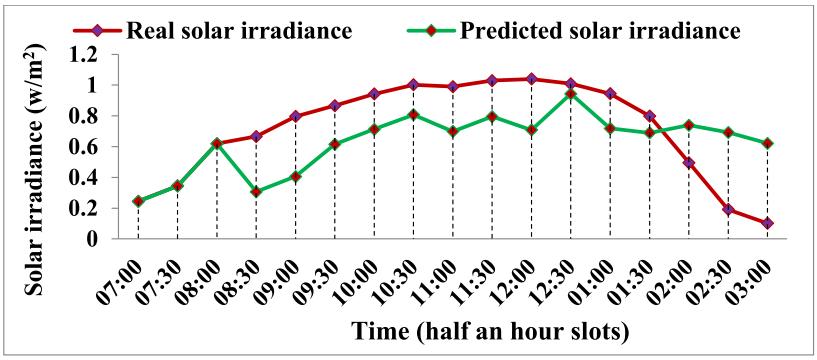


(c) 90 minutes ahead solar irradiance prediction by Pro-Energy

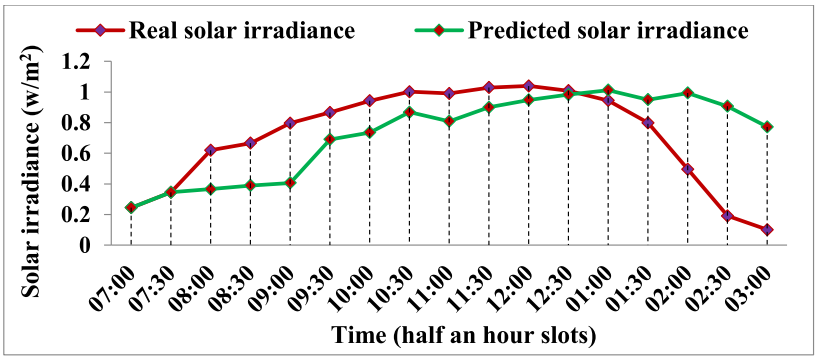


(d) 90 minutes ahead solar irradiance prediction by modified Pro-Energy

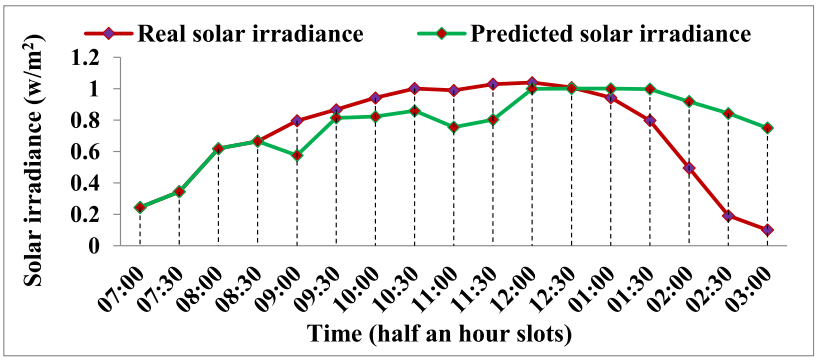
Figure 3.15: 90 minutes ahead solar irradiance prediction by different algorithms



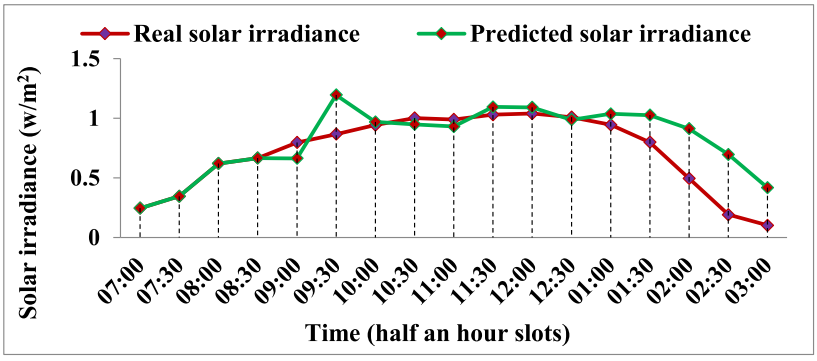
(a) 120 minutes ahead solar irradiance prediction by EWMA



(b) 120 minutes ahead solar irradiance prediction by WCMA



(c) 120 minutes ahead solar irradiance prediction by Pro-Energy



(d) 120 minutes ahead solar irradiance prediction by modified Pro-Energy

Figure 3.16: 120 minutes ahead solar irradiance prediction by different algorithms

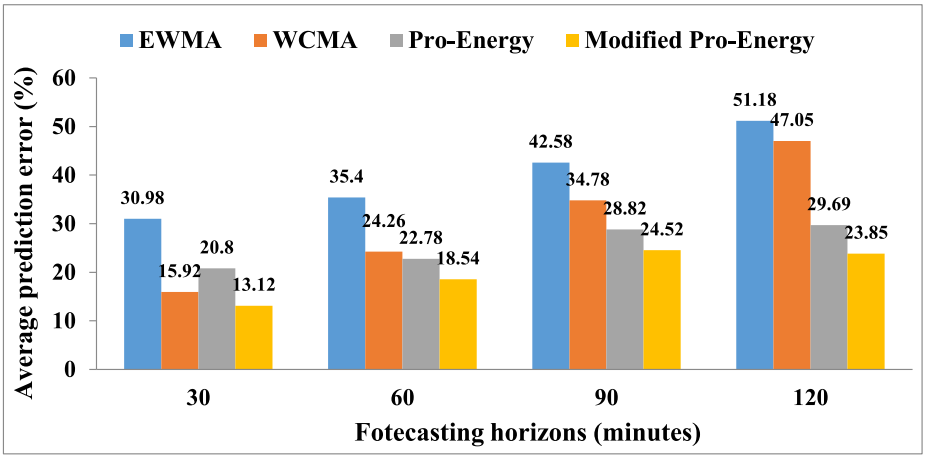


Figure 3.17: Average prediction error in different forecasting horizons

3.10 Concluding Remarks

The critical analysis on statistical models is done for solar irradiance prediction. It is found that EWMA and WCMA are suitable for the very short-term forecasting horizons (30 minutes) whereas the Pro-Energy prediction algorithm is well suited for few hours ahead forecasting horizon. The time series analysis is taken as the main methodology to modify the Pro-Energy prediction algorithm. The smoothened level and the trend components of the time series are computed to estimate the future irradiance level. The extraction of similar days from past historical data is done to maintain a pool having similar characteristics as the current day. The short term solar irradiance forecasting is carried out using the most similar day from the pool having a minimum absolute error with the current day. For few hours ahead forecasting, a set of most similar days is taken instead of a single day to improve prediction accuracy. The combined weighted profile of selected days is computed to consider different weather conditions. The simulation results validate the effectiveness of the proposed work in term of reduced prediction error in 30, 60, 90, and 120 minutes ahead prediction in comparison with EWMA, WCMA, and the pre-existing Pro-Energy algorithm.

For 30 minutes ahead solar irradiance prediction, the modified Pro-Energy algorithm offers 13.12% prediction error which is acceptable in short term forecasting. The proposed approach shows better response in comparison to existing EWMA, WCMA, and Pro-Energy algorithms which offers prediction error of the order of 30.98%, 15.92%, and 20.8% respectively.

The response of the modified Pro-Energy algorithm is not very much affected for 60 minutes ahead solar irradiance prediction. The proposed method offers an error of the order

of 18.54% which is less in comparison to EWMA, WCMA, and Pro-Energy algorithms (35.40%, 24.26%, and 22.78% respectively) and results in close estimation of future energy profiles.

For 90 minutes ahead solar irradiance prediction, the modified Pro-Energy algorithm offers 24.52% prediction error which is less in comparison to 42.58%, 34.78%, and 28.82% offered by EWMA, WCMA, and Pro-Energy algorithms respectively.

For 120 minutes ahead solar irradiance prediction, the proposed approach results in accurate estimations of solar profiles with prediction error of the order of 23.85%. Similar to previous test cases, in this forecasting horizon the resultant prediction error is less as compare to EWMA, WCMA, and Pro-Energy algorithms (51.18%, 47.05%, and 29.69% respectively).

Chapter 4

Optimized Solar Forecasting Approach Using Machine Learning

Machine learning based solar forecasting is carried out in the present chapter. The work is divided in two parts. In Part A, five independent machine learning models including FoBa, leapForward, Spikeslab, Cubist, and bagEarthGCV are analysed and simulated. In Part B, ensembled approach of these models is proposed and analysed to improve the prediction accuracy.

Part A

Rechargeable wireless sensor networks mitigate the life span and cost constraints propound in conventional battery operated networks. Reliable knowledge of solar radiation is essential for informed design, deployment planning, and optimal management of self-powered nodes. The problem of solar irradiance forecasting is well addressed by machine learning methodologies over historical data set. The characteristic of machine learning models to trace relation between input and output parameters allow this methodology to be successful in various domains including classification, data mining, and solar forecasting. Classification and data mining have been considered as the initial step for machine learning based models as pre-processing of data has been required with big datasets. In proposed work, forecasts are done using FoBa, leapForward, Spikeslab, Cubist, and bagEarthGCV models to track effective solar forecasting models and analyse prediction accuracy of each model. Machine learning is applied on historical solar intensity observations as training dataset to calculate future solar irradiance for different forecasting horizons irrespective of seasonal variation and input parameters availability. To validate the effectiveness of these methodologies, a series of experimental evaluations are presented in terms of forecast accuracy, correlation coefficient, and Root Mean Square Error (RMSE). The *R* interface is used as simulation platform for these evaluations. The dataset from National Renewable Energy Laboratory (NREL) is used for experiments. The experimental results exhibits that from few hours to two days ahead solar irradi-

The proposed work has been published in **Renewable and Sustainable Energy Reviews journal, 82(3), pp. 2254-2269, 2017**.

ance prediction is precisely estimated by machine learning based models irrespective of seasonal variation in weather conditions.

4.1 Solar Irradiance Forecasting Platform

4.1.1 Machine Learning Methodology

All machine learning based algorithms works to trace a predictive model that estimates a particular type of data with high accuracy. Large dataset is essential for the learning algorithm to understand the behaviour of the system. Figure 4.1 exhibits the basic machine learning methodology. First step for machine learning based system is data procurement. Collected data is divided from different prospectives and summarizes in useful information. The steps included in this process is data cleansing and data segregation. Data is segregated in three disjoint sets, training, testing, and blind set. Training dataset is applied for model training and testing dataset is used for model optimization and evaluation. Blind dataset is used for result analysis and cross validation.

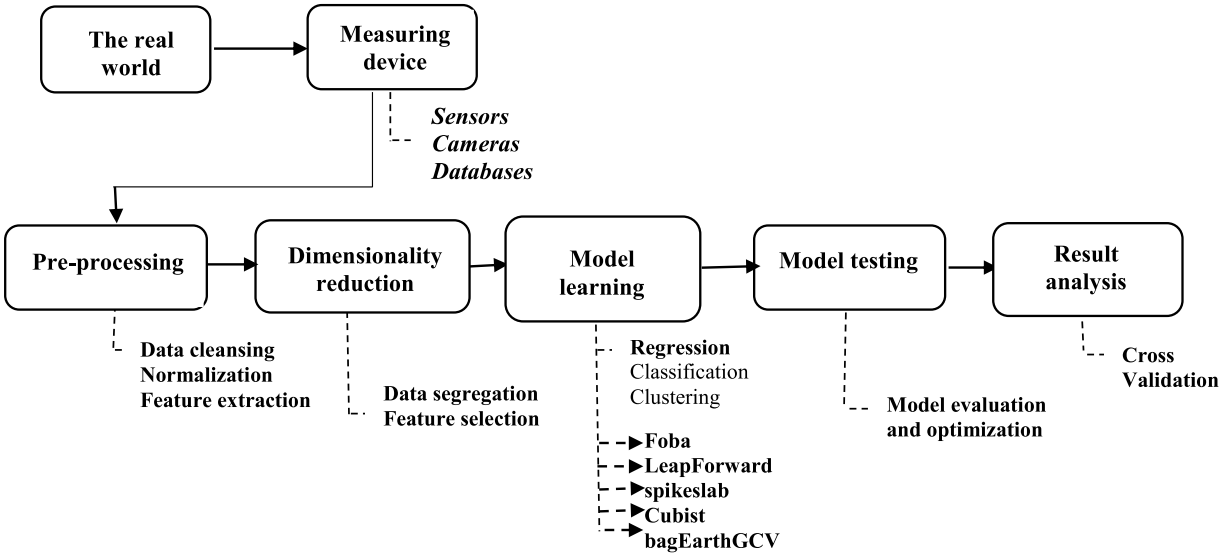


Figure 4.1: Flow diagram of machine learning methodology

4.1.2 Machine Learning Models

In proposed work, machine learning time series models which are based on historically observed solar irradiance as input parameter are used for solar forecasting called endogenous forecasting. Figure 4.2 summarizes five forecasting models used in proposed work with their methodologies and explained in following section:

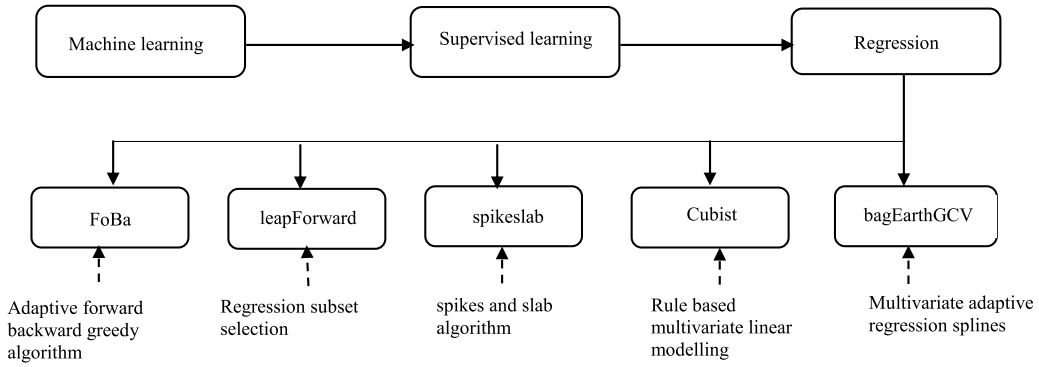


Figure 4.2: Machine learning based models

4.1.2.1 FoBa (Adaptive Forward Backward Greedy Algorithm)

FoBa is based on forward greedy algorithm with adaptive backward steps[151][277]. The objective is to remove any error caused by earlier forward steps and avoid large number of basis functions. Adaptive backward steps ensure that any backward greedy step will not erase gain made in forward steps. Consider n input vectors $x_i \epsilon R^d$ (i=1, 2, ..., n), d feature vectors $f_j \epsilon R^n$ (j=1, 2, ..., d) with output variables $y \epsilon R^n$. Each f_j is equivalent to j_{th} feature component of x_i that corresponds to $f_{j,i} = x_{i,j}$. By considering a linear prediction model f(x)= w^Tx and imposing emphasis on weight w to generate a regularized problem with main constraint as sparsity, corresponds to non convex L_0 regularization. By defining $\|w\|_0 = |j : w_j \neq 0| = sp$, L_0 regularization is computed as:

$$w^{\cdot} = \arg\min_{w \epsilon R^d} \frac{1}{n} \sum_{i=1}^{n} \phi(w^T x_i, y_i) \qquad \|w\|_0 \leq sp \tag{4.1}$$

where sp is sparsity parameter. If sp is not given, it is considered as tuning parameter and selected by cross validation method. This process is known as subset selection and important in machine learning. The prediction accuracy is evaluated by loss function $\phi(w^T x_i, y_i)$. For simplification, equation 4.1 is re-written as non convex L_0 regularization problem with sparsity parameter sp and cost function $R(w)$ and given as:

$$w^{\cdot} = \arg\min_{w \epsilon R^d} R(w) \tag{4.2}$$

subject to $\|w\|0 \leq sp$ where w=$[w_1, w_2, ..., w_d] \epsilon R^d$. For least square regression, R(w) is a real valued cost function and calculated as:

$$R(w) = n^{-1} \| \sum_{j=1}^{d} w_j f_j - y \|_2^2 \tag{4.3}$$

A backward step is taken when the increase of cost function is no more than half of the decrease of cost function in earlier forward steps i.e. if l forward steps are taken, the cost function should be decreased by at least by an amount of $l_\varepsilon/2$. This means that if $R(W) \geq 0$ for $allw \epsilon R^d$, the algorithm terminates after no more than $2\mathrm{R}(0)/\varepsilon$ steps. The procedure for FoBa is summarized in the algorithm 4.1.

Algorithm 4.1 FoBa: Forward- Backward greedy algorithm

Input parameters: f_1,.....,f_d, $y \epsilon R^n$ and $\varepsilon > 0$
output parameters: F^{sp} and w^{sp}

Let $F^{(0)}=\phi$ and $W^{(0)}=0$
Let sp=0
while
Let sp=sp+1 // Forward step
Let i^{sp}=arg $\min_i \min_\alpha \mathrm{R}(W^{SP-1}+\alpha e_i)$
Let $F^{sp}=(i^{sp}) \cup (F^{sp-1})$
Let $w^{sp}=w_\cdot(F^{sp})$
δ=Forward/backward step square error reduction
Let δ^k=R($w^{(sp-1)}$)-R(w^{sp})
if ($\delta^{sp} \leq \varepsilon$)
sp=sp-1
break
endif

// Backward step(after few forward steps)
while
Let j^{sp}=arg $\min_{j \epsilon F^{(sp)}}$ R($w^{(sp)}$-$w_j^{(sp)} e_j$) where e_j = zero vector and $e_j \epsilon R^d$
Let δ^x=R($w^{(sp)} - w_{j^{(sp)}}^{(sp)} e_{j^{(sp)}} - R(w^{(sp)})$)
if δ^x =$(0.5)\delta^{sp}$
break
sp=sp-1
$F^{sp} = F^{sp+1} - j^{sp+1}$
Let $w^{sp}=w_\cdot(F^{sp})$
end
end

4.1.2.2 leapForward

leapForward model is based on linear regression with forward subset selection. It performs exhaustive search using match-select-action cycle for tracing best subset of predicting variables[278]. Selecting subset refers to finding a small set of independent variables that offers less prediction error in predicting the dependent variables with measures of goodness-of-fit. With a given set of variables X_1, X_2, ..., X_k, the focus is to trace a subset

of $p < k$ variables $X_{(1)}$, $X_{(2)}$, ..., $X_{(p)}$ which minimizes equation 4.4 in least square sense,

$$S = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} b_{(j)}x_{i,(j)})^2 \tag{4.4}$$

$x_{i,(j)}$, y_i are the i_{th} observations of variables $X_{(j)}$ and Y signifies the predicted data points, b_j denotes the least square regression coefficient.

In the forward selection procedure, the first variable selected is that variable X_j for which

$$S = \sum_{i=1}^{n}(y_i - b_{(j)}x_{i,j})^2 \tag{4.5}$$

is to be minimized. b_j minimizes S for variable X_j. The value of b_j is given as

$$b_j = \sum_{i=1}^{n} x_{(i,j)}y_i / \sum_{i=1}^{n} x_{ij}^2 \tag{4.6}$$

By using the value of b_j in equation 4.5, the new equation follows that

$$S = \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} x_{(i,j)}y_i)^2 / \sum_{i=1}^{n} x_{ij}^2 \tag{4.7}$$

Thus, the variable should be selected that maximizes

$$(\sum_{i=1}^{n} x_{(i,j)}y_i)^2 / \sum_{i=1}^{n} x_{ij}^2 \tag{4.8}$$

If the equation 4.8 is divided by $\sum_{i=1}^{n} y_i^2$, the square of the cosine of the angle between vectors X_j and Y is generated. The cosine is the correlation between the variables X_j and Y, if the mean is subtracted from each variable. If $X_{(1)}$ is denoted as the first selected variable, the residuals $Y - X_{(1)}b_{(1)}$ are orthogonal to $X_{(1)}$. For the reduction of the sum of squares by adding further variables, the space orthogonal to $X_{(1)}$ should be selected. Thus, the next variable selection is formulates as:

$$X_{j.(1)} = X_j - b_{j,(1)}X_1 \tag{4.9}$$

where $b_{j,(1)}$ is the least square regression coefficient of X_j based on $X_{(1)}$. The equation 4.8 will maximize when Y is replaced by $Y - X_{(1)}b_{(1)}$ and X_j is replaced by $X_{j.(1)}$.

The consecutive terms of sum of squares and products is calculated directly from pre-

vious sum of squares and product terms without calculating the individual orthogonal components for n observations. If the mean is subtracted from each variable, the new selected variable has the largest correlation with variable Y.

Thus, the selected variables $X_{(1)}$, $X_{(2)}$, ..., $X_{(p)}$ are sequentially added to the prediction equation in order to minimize residual sum of squares and to improve prediction accuracy.

4.1.2.3 Spikeslab

The Spikeslab model is a generalization of elastic net with Bayesian estimates[148][150]. These computed estimates follows weighted generalized ridge regression approach to handle correlated large datasets. The model is implemented in three stages shown in Figure 4.3 and listed below:

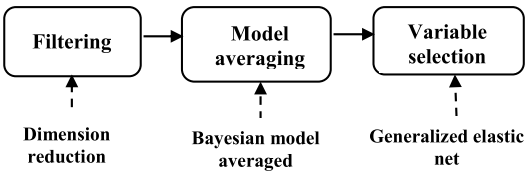


Figure 4.3: Flow diagram of Spikeslab

In step 1, filtering process carries top $(n \times F)$ variables where n is the sample size and $F > 0$ is the user defined fraction. Rest of the variables are filtered out to reduce the dimension. The posterior mean coefficient is calculated using Gibbs sampling[279] for appropriate ordering of variables.

Step 2 refit the model using only those variables that are not filtered in step 1. Gibbs sampler is used for model fitting. The algorithm returns fast execution process because of the reduced variables from the large dataset. A blocking technique is used to further reduce the computation time. The posterior mean of regression coefficients is calculated referred to as Bayesian Model Averaged (BMA) and used to estimate the regression coefficients.

Generalized Elastic Net (gnet) in step 3 is used for variable selection. Variables obtained from restricted BMA from step 2 are classified in groups. Grouping forces the variables to share a common regularization parameter. BMA obtain l_2 regularization parameters and solution path is determined with respect to l_1 regularization parameters using $lars$ R *package*[280]. There is no limit on number of groups. A variable that does not appear in the list will be assigned to a default group that has its own group specific regularization parameter.

4.1.2.4 Cubist

Cubist was developed by Quinlan[145] for inducing trees for regression models. Cubist is a rule based predictive model where each rule carries a multivariate linear model. These models works on the predictions of previous splits[281][146]. When a case satisfies all rule based conditions, the associated model is used for prediction. Figure 4.4 shows the flow diagram of Cubist model.

In first stage, recursive partitioning (divide and conquer) of training cases is exercised to generate piecewise linear model in the form of regression based model tree. Each training case has a set of attributes and associated target value.

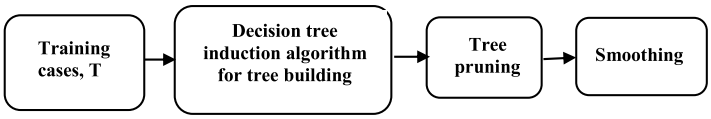


Figure 4.4: Flow diagram of Cubist model

The basic approach is to generate a model that relates target values of the training cases to their values of other attributes. A splitting criteria is used to minimize the intra subset variation in the class values instead of maximizing the information gain at each interior node. The splitting criteria is based on computing standard deviation of target values in T. The attribute that minimizes standard deviation is chosen. If $sd(T_i)$ is considered as standard deviation of the target in T_i cases where i=$1, 2, .., n$, reduction in standard deviation is calculated as:

$$Error = sd(T) - \sum_{i=1}^{n} \frac{T_i}{T} \times sd(T_i) \tag{4.10}$$

where T is the set of cases that reaches to the node and T_1, T_2 are selected cases after splitting the node according to chosen attributes.

At the second stage, pruning is carried out by estimating the expected error that will be experienced at each node for test data. Each linear model is simplified by eliminating parameters in order to reduce its estimated error. Parameters are eliminated one by one so long as the error estimate decreases. Each internal node of the tree has both a simplified model and a model subtree. The one with lowest estimated error is chosen. Firstly, the absolute difference between actual and predicted data value is averaged for each training case that reach the node. The calculated parameter is multiplied by a factor of $(n + p)/(n - p)$ where n is the training cases and p is the number of parameters that represents the class value at each node. Further, for each interior node a linear model

is estimated by using the standard regression approach. The induced tree is simplified by dropping the certain cases in order to minimize the estimated error. The terms are dropped sequentially as long as the error estimates decreases. Finally, when tree is induced for each interior node, tree is pruned back from leaves.

Finally, smoothing process is carried out to compensate abrupt discontinuities between adjacent linear models of a pruned tree. A predicted value from a leaf model is adjusted to account models at the nodes along the path from root to that leaf. Smoothing is carried out by multiplying output from a leaf model with the value predicted by linear model of that node. The calculation is as follows:

$$Predicted" = \frac{g \times Predicted' + g \times Predicted}{n + g} \tag{4.11}$$

where $Predicted"$ is the prediction passed on the next higher nodes, $Predicted'$ is prediction passed to this node from below, $Predicted$ is the predicted value at this node, n is the number of training cases that reach the node below and g is a constant. Smoothing process increases the prediction accuracy significantly.

4.1.2.5 bagEarthGCV

The bagEarthGCV is a non-parametric regression technique and is based on Multivariate Adaptive Regression Splines (MARS)[144][282]. Open source implementations of MARS are termed as Earth which are licenced to Salford systems[283]. The relationship between independent and dependent variables is constructed using a set of coefficients and two sided truncated basis functions that are calculated using regression data. The MAR-Splines model is implemented by the following equation:

$$y = f(x) = W_0 + \sum_{u=1}^{U} W_u h_u(x) \tag{4.12}$$

where y is dependent output variable, $f(x)$ is predicted as a function of predictor variable x, intercept parameter W_0 and weighted sum of one or more basis functions $h_u(x)$. Each W_u is a constant coefficient. To increase the measure of goodness-of-fit, more basis functions are added until a user defined complexity level is reached.

- ***Pruning***

  After forward stepwise selection of basis functions, a backward procedure called pruning is applied to remove those basis functions which are least concerned with increase in goodness-of-fit. The generalized cross validation error is a measure of

goodness-of-fit that considers residual error together with model complexity and formulated as:

$$GCV = \frac{\sum_{i=1}^{N}(y_i - f(x_i))^2}{(1 - \frac{B}{N})^2} \qquad B = 1 + b.df \tag{4.13}$$

In equation 4.13, N signifies number of cases in dataset, df is degree of freedom which is equal to independent basis functions and b corresponds to the penalty of adding a basis function.

4.2 Description of Database

To determine the effectiveness of above mentioned models, historical solar data set including daily solar irradiance over a period of six years (January 1, 2010 to December 31, 2015) is collected from National Renewable Energy Laboratory (NREL). The data set obtained from NREL is sampled on per minute basis on horizontal plane with 1440 samples per day. Historical data for the same month from all the six years is taken for training. For instance, if March 11, 2016 is to be predicted, the March dataset is taken from the years from 2010 to 2015 for training. This is done to circumvent variable maximum and minimum solar irradiance and different duration of sunshine for different months. Best input selection is the primary goal to train models for different forecasting horizons. Different configurations of historical data set are tested to downright the training process. Season (spring, summer, monsoon, and winter) base training of different models is performed. The models are trained for ± 5 to ± 25 past days $(D_{d-1}, D_{d-2}, ..., D_{d-n})$ and for 6:00 am onwards past time slots $(T^{t-1}, T^{t-2}, ..., T^{t-n})$. These features are selected by exhaustive simulation process with respect to correlation coefficient (r), coefficient of determination (r^2), RMSE, and accuracy. Machine learning based models assure similar meteorological behaviour in training and testing datasets while keeping a ratio of 70% training and 30% testing. The proposed methodology is generalized and applicable to any year or country. As the presented approach utilize the historical dataset, the only condition is to select past historical records from the same geographical locations, similar weather conditions and from the same seasons of the year for which the prediction is to be made. Further, in cloudy versus sunny weather conditions, definitely there is a difference in expected energy levels. The proposed methodology is not specific to any particular year or to a particular country. As the presented approach utilize the historical dataset, the only condition is to select past historical records from the same geographical locations, similar weather conditions and from the same seasons of the year for which the prediction is to be made. Further, in cloudy versus sunny weather conditions, definitely there is a difference in expected energy levels.

4.3 Performance Matrixes

To precisely evaluate the prediction accuracy of previously described models, four statistical quality measures are adopted. These measures are executed by the equations 4.14 and 4.15, given as:

- Correlation is a statistic that interpret the degree of correspondence between two variables and is computed as:

$$r = \frac{N\sum E_{actual}E_{predicted} - (\sum E_{actual})(\sum E_{predicted})}{\sqrt{[N\sum(E_{actual})^2 - (\sum E_{actual})^2][N\sum(E_{predicted})^2 - (\sum E_{predicted})^2]} \quad -1.0 \leq r \leq +1.0 \tag{4.14}$$

  where N is the number of samples, $E_{Predicted}$ is predicted solar irradiance, and E_{Real} is the real time solar irradiance.

- Coefficient of determination, r^2 $(0 \leq r^2 \leq 1)$ exhibit the proportional variation of one variable which is predicted from other variable by measuring the predicted values that are best fit to regression line. Square of correlation coefficient, r^2 is taken as coefficient of determination in linear least square regression.

- To calculate the difference between real time measurements and specific model predicted values, Root Mean Square Error (RMSE) or Root Mean Square Deviation (RMSD) is used:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(E_{actual} - E_{predicted})^2} \tag{4.15}$$

  where T is the time slots under consideration.

- Accuracy measures the proximity of the analytical results to the actual value. In proposed work an acceptance error of $\pm 20\%$ is considered.

The model performance has not been evaluated for night test samples as solar irradiance is not available during night hours.

4.4 Simulation and Result Discussion

A series of simulation experiments are performed to study the accuracy of five solar irradiance forecasting models. A common platform is used to run all experiments. The present section narrate different experimental results. Historical data of last six years from NREL is used for training. To evaluate the forecasting models in different weather

conditions, four days, 11^{th} March (spring), 25^{th} June (summer), 30^{th} August (monsoon) and 31^{st} December (winter) from different seasons of year 2016 are used for testing. It is seen from Figure 4.5 that the data on 11^{th} March and 30^{th} August is smooth as these are sunny days, data on 25^{th} June varies throughout the day because of cloudy weather conditions, data on 31^{st} December possesses smooth behaviour with lower maximum and minimum solar intensity thresholds because of winter season.

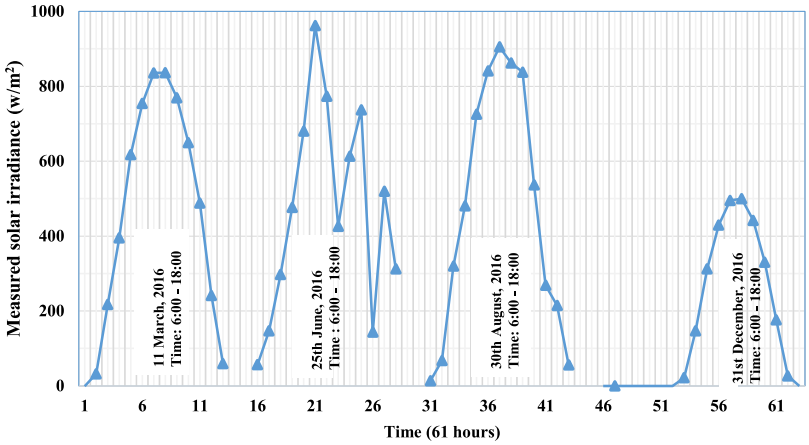


Figure 4.5: Seasonal variation in solar irradiance with days of a year

4.4.1 Experiments

Three test cases are carried out to access prediction effectiveness based on input parameter selection. In the first one, impact of number of past days on prediction accuracy is evaluated. Second experiment evaluate the performance by varying number of past time slots. In the third experiment, different forecasting horizons are considered to evaluate the forecasting models.

4.4.1.1 Prediction Accuracy With Respect to Number of Past Days

In test 1, simulation is performed with 10 to 50 past days for training and 5 to 25 past days for testing for all five forecasting models. As shown in table 4.1, for 11^{th} March, leapForward shows maximum accuracy of 96.08% with ±15 training days and 15 past days for testing. Cubist achieves maximum accuracy of 78.95% for 25^{th} June with ±5 training days and 5 historical days for testing. For 30^{th} August, an accuracy of 89.47% is achieved by bagEarthGCV with ±5 training days and 5 historical days for testing. An accuracy of 98.21% is gained by FoBa with ±20 training days and 20 days for testing for 31^{st} December. The performance ranking of above mentioned models is complicated

because relationship between past days weather metrics and present day solar intensity is random in nature. The choice of most accurate model with adequate number of past days depends upon on present day weather conditions and is iteration specific. However, it is observed that prediction accuracy is satisfactory if model is trained for the same month i.e. ± 15 days for which the prediction to be done.

Table 4.1: Comparison results of five forecasting models for selection of historical days for forecasting (Test 1)

Season(2016)		**11 March (Spring)**					**25 June (Summer)**					**30 August (Monsoon)**					**31 December (Winter)**				
Past days (Training)		±5	±10	±15	±20	±25	±5	±10	±15	±20	±25	±5	±10	±15	±20	±25	± 5	±10	±15	±20	±25
Past days (Testing)		5	10	15	20	25	5	10	15	20	25	5	10	15	20	25	5	10	15	20	25
FoBa	**r**	0.94	0.98	0.98	0.96	0.96	0.96	0.98	0.95	0.96	0.89	0.97	0.95	0.94	0.93	0.97	0.99	0.97	0.97	1	0.98
	r^2	0.88	0.96	0.96	0.92	0.92	0.92	0.96	0.9	0.92	0.79	0.94	0.9	0.88	0.86	0.94	0.98	0.94	0.94	1	0.96
	RMSE	55.41	32.25	35.77	36.51	38.36	64.18	47.22	56.59	57.88	98.98	33.71	33.23	28.54	31.68	21.69	14.77	17.05	15.99	2.92	11.07
	accuracy	47.37	70.27	52.94	66.07	67.86	57.89	64.86	50.91	54.55	29.09	42.11	56.76	65.45	57.14	78.57	63.16	89.19	78.18	98.21	85.71
leapForward	**r**	0.89	0.98	0.99	0.96	0.96	0.99	0.98	0.95	0.96	0.87	1	0.87	0.94	0.91	0.96	0.99	0.96	0.97	0.98	0.98
	r^2	0.79	0.96	0.99	0.92	0.92	0.98	0.96	0.9	0.92	0.76	1	0.76	0.88	0.83	0.92	0.98	0.92	0.94	0.96	0.96
	RMSE	63.67	32.03	3.07	42.11	38.3	44.42	47.14	56.44	57.88	100.3	14.69	41.31	29.61	33.18	21.67	12.9	17.32	14.14	13.49	12.17
	accuracy	47.37	48.65	96.08	57.14	67.86	42.11	64.86	50.91	54.55	23.64	73.68	64.86	61.82	73.21	83.93	84.21	78.38	80	78.57	82.14
Spikeslab	**r**	0.94	0.98	0.99	0.96	0.97	0.98	0.99	0.95	0.97	0.88	1	0.94	0.95	0.93	0.97	1	0.97	0.98	0.95	0.93
	r^2	0.88	0.96	0.98	0.92	0.94	0.96	0.98	0.9	0.94	0.77	1	0.88	0.9	0.86	0.94	1	0.94	0.96	0.97	0.92
	RMSE	47.96	27.98	25.62	36.17	34.34	35.53	35.89	55.57	53.46	93.77	12.52	30.64	28.8	30.33	19.86	9.74	17.06	13.75	16.34	20.56
	accuracy	57.89	67.57	66.67	66.07	62.5	63.16	56.76	43.64	54.55	36.36	89.47	72.97	63.64	67.86	80.36	78.95	72.97	83.64	75.88	70.66
Cubist	**r**	0.93	0.99	0.99	0.98	0.97	0.99	0.98	0.95	0.98	0.86	0.99	0.92	0.95	0.93	0.97	0.99	0.97	0.97	0.99	0.98
	r^2	0.86	0.98	0.98	0.96	0.94	0.98	0.96	0.9	0.96	0.74	0.98	0.85	0.9	0.86	0.94	0.98	0.92	0.94	0.98	0.96
	RMSE	53.62	21.7	14.31	25.5	34.28	35.18	36.79	49.61	35.47	100.9	10.98	37.1	26.71	30.62	18.51	12.39	17.55	14.66	9.24	10.38
	accuracy	63.16	75.68	84.31	76.79	64.29	78.95	59.46	60	61.82	47.27	84.21	64.86	74.55	71.43	76.89	78.95	78.38	78.18	87.5	89.29
bagEarthGCV	**r**	0.95	0.98	1	0.97	0.96	0.98	0.98	0.95	0.96	0.86	0.99	0.89	0.95	0.93	0.97	1	0.96	0.97	0.98	0.98
	r^2	0.9	0.96	1	0.94	0.92	0.96	0.96	0.9	0.92	0.74	0.98	0.79	0.9	0.86	0.94	1	0.92	0.94	0.96	0.96
	RMSE	35.92	25.39	9.49	27.04	39.05	50.68	44.81	60.85	54.53	109	11.27	38.15	29.87	31.68	19.65	8.49	19.82	15.14	12.14	11.27
	accuracy	63.16	75.68	88.24	71.43	50	52.63	43.24	43.64	45.45	36.36	89.47	67.57	65.45	66.07	78.57	84.21	75.68	81.82	82.14	89.29

4.4.1.2 Prediction Accuracy With Respect to Initial Past Time Slots

In test 2, initial time slots from 6:00 am to 10:00 am are considered for performance evaluation. Simulation results in table 4.2 exhibits the effect of initial past time slots on prediction accuracy. For different four days (11^{th} March, 25^{th} June, 30^{th} August, and 31^{st} December) of the year 2016, FoBa offers maximum accuracy (70.27%, 64.86%, 66.07%, and 98.21% respectively) with past samples taken from 7:00am onwards. The methodology of leapForward model has least effect of initial past samples on prediction accuracy. For 11^{th} March, 25^{th} June, 30^{th} August, and 31^{st} December maximum accuracies are 96.08%, 64.86%, 83.93%, and 84.21% respectively for almost all time slots. Experimental results shows that similar to leapForward model, Spikeslab offers less variation with respect to initial time slot consideration. For 11^{th} March, 25^{th} June, and 31^{st} December, maximum accuracy (78.38%, 68.42%, 85.45% respectively) is gained with past time slots from 10:00 am onwards. For 30^{th} August maximum accuracy (78.95%) is gained with past samples from 6:00 am onwards. Cubist model achieves high prediction accuracy with large past sample consideration (6:00 am onwards). For 11^{th} March, 25^{th} June, 30^{th} August, and 31^{st} December maximum accuracies are 90.2%, 78.95%, 84.21%, and 85.71% respectively. Similar results are obtained for bagEarthGCV where prediction accuracy of the order of 94.12%, 63.16%, 94.74%, and 89.29% is achieved for 11^{th} March, 25^{th} June, 30^{th} August and 31^{st} December respectively for different past samples. It is observed from the above test that prediction accuracy is affected by those time slots which are having considerable solar intensity. It is seen in table 4.2 that by considering solar irradiance observation from 7:00am onwards, satisfactory level of measures of goodness-of-fit is achieved.

4.4.1.3 Prediction Accuracy with Respect to Forecasting Horizon and Seasonal Validation

To investigate the prediction accuracy of five forecasting models with respect to different forecasting horizons, simulation tests have been conducted for 1 hour ahead, 24 hours ahead, and 48 hours ahead solar forecasting in test 3.

1. 11^{th} ***March 2016***

   For 11^{th} March, 2016, historical days from 25^{th} February to 27^{th} March ($\pm$15 days)(2010-2015) are used for training. The testing data is taken from 25^{th} February 2016 to last considered horizon. Initial past time slots are taken from 7:00am.

Table 4.2: Comparison results of five forecasting models for selection initial past time slots for forecasting (Test 2)

Season(2016)		**11 March (Spring)**					**25 June (Summer)**					**30 August (Monsoon)**					**31 December (Winter)**				
Initial past time slots		6:00 AM	7:00 AM	8:00 AM	9:00 AM	10:00 AM	6:00 AM	7:00 AM	8:00 AM	9:00 AM	10:00 AM	6:00 AM	7:00 AM	8:00 AM	9:00 AM	10:00 AM	6:00 AM	7:00 AM	8:00 AM	9:00 AM	10:00 AM
FoBa	**r**	0.92	0.98	0.97	0.98	0.98	0.96	0.98	0.95	0.95	0.96	0.97	0.97	0.95	0.96	0.96	0.99	1	0.99	0.99	0.99
	r^2	0.85	0.96	0.94	0.96	0.96	0.92	0.96	0.9	0.9	0.92	0.94	0.94	0.9	0.92	0.92	0.98	1	0.98	0.98	0.98
	RMSE	63.03	32.25	34.49	29.89	34.37	76.3	47.22	75.02	75.02	64.81	23.33	23.33	33.23	28.8	29.08	14.77	2.92	14.49	11.18	12.46
	accuracy	49.02	70.27	60.78	62.75	62.75	32.43	64.86	43.24	43.24	40.54	66.07	66.07	56.76	57.14	57.14	63.16	98.21	63.16	78.95	78.95
leapForward	**r**	-0.59	1	1	1	1	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.96	0.96	0.96	0.99	0.99	0.99	0.99	0.99
	r^2	0.35	1	1	1	1	0.96	0.96	0.96	0.96	0.96	0.92	0.92	0.92	0.92	0.92	0.98	0.98	0.98	0.98	0.98
	RMSE	654.5	3.07	3.07	13.55	3.07	47.14	47.14	47.14	47.14	47.14	21.67	21.67	21.67	21.67	21.67	12.9	12.9	12.9	12.9	12.9
	accuracy	0	96.08	96.08	78.43	96.08	64.86	64.86	64.86	64.86	64.86	83.93	83.93	83.93	83.93	83.93	84.21	84.21	84.21	84.21	84.21
Spikeslab	**r**	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.99	1	0.94	0.95	0.99	1	0.98	0.97	0.97	0.97	0.97
	r^2	0.96	0.96	0.96	0.96	0.96	0.96	0.98	0.96	0.96	0.98	1	0.88	0.9	0.98	1	0.96	0.94	0.94	0.94	0.94
	RMSE	27.98	27.98	28.19	28.23	29.18	35.13	35.89	40.42	33.24	32.36	13.9	30.64	28.8	20.84	15.63	14.18	17.06	13.48	13.62	13.18
	accuracy	70.27	67.57	70.27	67.57	78.38	63.16	56.76	63.16	63.16	68.42	78.95	72.97	63.64	57.89	57.89	83.64	72.97	85.45	81.82	85.45
Cubist	**r**	0.99	0.99	0.97	0.98	0.97	0.98	0.99	0.98	0.98	0.98	0.99	0.92	0.99	0.99	0.99	0.98	0.97	0.98	0.98	0.98
	r^2	0.98	0.98	0.94	0.96	0.94	0.96	0.98	0.96	0.96	0.96	0.98	0.85	0.98	0.98	0.98	0.96	0.92	0.96	0.96	0.96
	RMSE	11.72	21.7	16.62	17.17	34.28	47.16	35.18	50.23	45.21	39.39	10.98	37.1	12.78	12.78	14.87	11.22	17.55	10.84	12.47	12.41
	accuracy	90.2	75.68	90.2	86.27	64.29	57.89	78.95	47.37	52.63	57.89	84.21	64.86	84.21	84.21	84.21	85.71	78.38	83.93	83.93	82.14
bagEarthGCV	**r**	0.99	0.98	1	0.99	1	0.99	0.98	0.99	0.96	0.98	0.99	0.99	1	0.99	1	0.98	0.96	0.97	0.98	0.98
	r^2	0.98	0.96	1	0.98	1	0.98	0.96	0.98	0.92	0.96	0.98	0.98	1	0.98	1	0.96	0.92	0.94	0.96	0.96
	RMSE	10.29	25.39	8.8	10.41	7.45	34.08	44.81	46.4	71.59	40.84	13.54	11.27	9.84	16.02	11.64	11.22	19.82	13.32	11.16	10.59
	accuracy	90.2	75.68	90.2	94.12	86.27	63.16	43.24	42.11	42.11	52.63	84.21	89.47	94.74	78.95	78.95	85.71	75.68	83.93	87.5	89.29

Table 4.3: Observations form 11^{th} March 2016 predictions

11^{th} **March 2016**			
Parameters	**1 hour ahead forecasting**	**24 hours ahead forecasting**	**48 hours ahead forecasting**
Optimized model	bagEarthGCV	Cubist	Spikeslab
Accuracy(%)	90.2	79.17	69.09
r^2	1	0.94	0.88
RMSE	8.52	38.22	45.15

In one hour ahead prediction, highest prediction accuracy (90.2%) is achieved by bagEarthGCV with r^2 value of 1 and 8.52 RMSE (Figure 4.6). In 24 hours ahead solar forecasting for 11^{th} March, 2016, prediction accuracy is reduced from 90.2% to 79.17% and offered by Cubist model. The values of correlation coefficient and RMSE are observed as 0.94 and 38.22 respectively (Figure 4.7). In 48 hours ahead solar forecasting for 11^{th} March 2016, maximum prediction accuracy (69.09%) is offered by Spikeslab model with 0.88 correlation coefficient and 45.15 RMSE (Figure 4.8). Table 4.3 reports the performance indicators in terms of correlation coefficient (r^2), RMSE and accuracy(%)

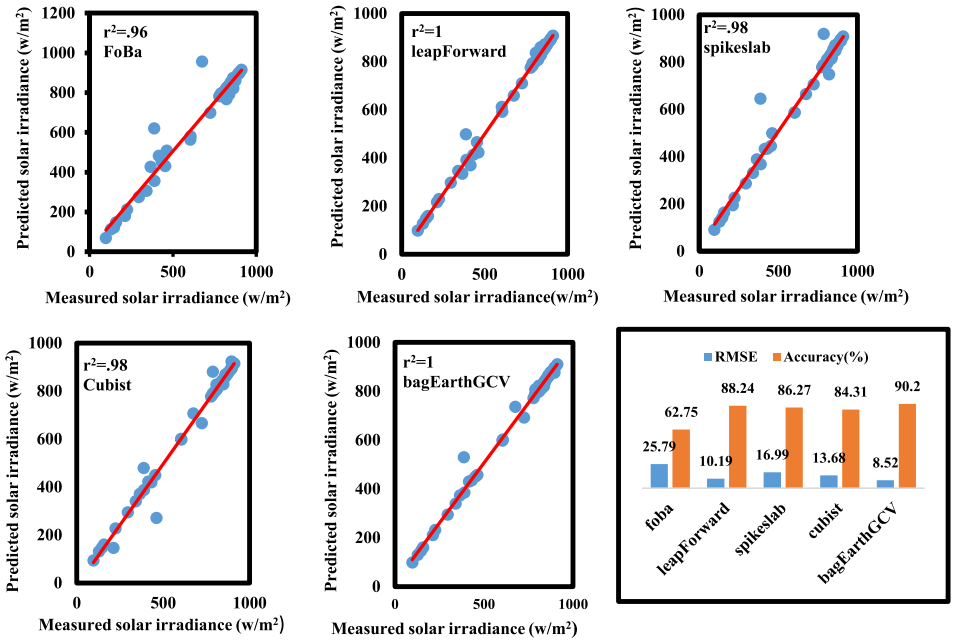


Figure 4.6: Correlation between predicted and measured solar irradiance for 11^{th} March, 1 hour ahead prediction

It has been observed that for 11^{th} March 2016 in all three forecasting horizons, performance matrix is satisfactory and effectiveness of a particular model is weather specific.

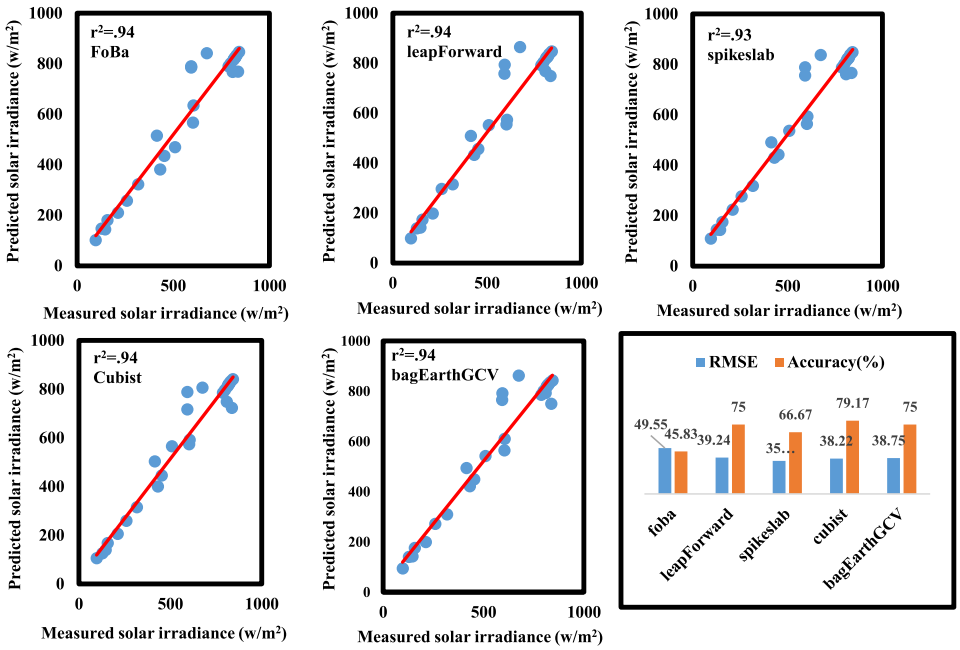


Figure 4.7: Correlation between predicted and measured solar irradiance for 11^{th} March, 24 hours ahead prediction

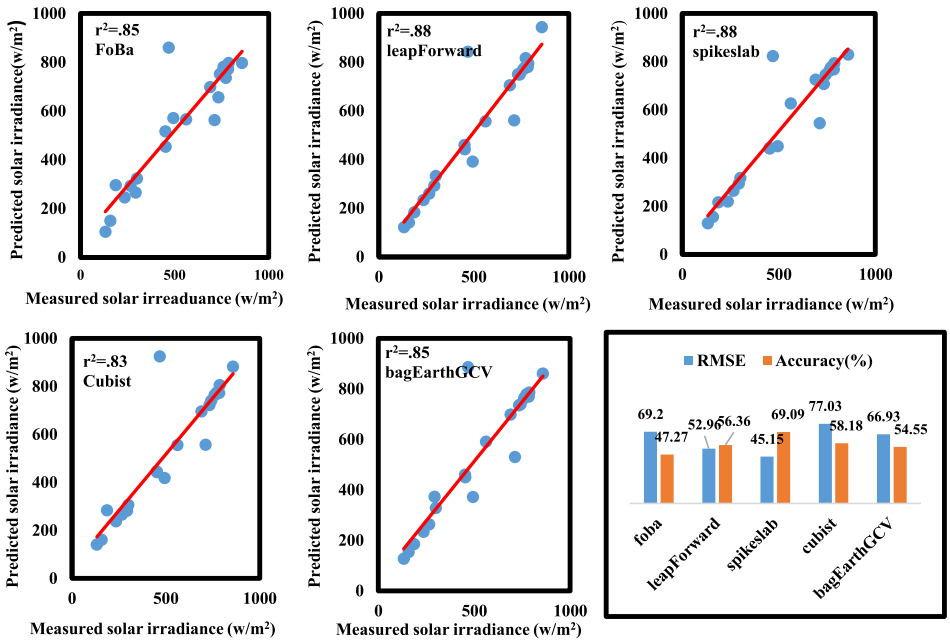


Figure 4.8: Correlation between predicted and measured solar irradiance for 11^{th} March, 48 hours ahead prediction

2. 25^{th} ***June 2016***

For 25^{th} June, 2016 in 1 hour ahead prediction, bagEarthGCV gains the highest prediction accuracy (73.08%) with 0.85 correlation coefficient and 35.58 RMSE (Figure 4.9). In 24 hours ahead prediction, Cubist model offers the highest accuracy (65.38%) with 0.94 correlation coefficient and 46.34 RMSE (Figure 4.10). In 48 hours ahead prediction of 25^{th} March 2016, maximum prediction accuracy (62.5%) is achieved by Spikeslab model with 0.94 correlation coefficient and 53.34 RMSE (Figure 4.11). The historical days from 10^{th} June to 10^{th} July (±15 days) (2010 to 2015) are used for training. Days from 10^{th} June 2016 to last available slot are used for testing. Initial past time slots are taken from 7:00am. Table 4.4 enlists the results of analysis.

Table 4.4: Observations form 25^{th} June 2016 predictions

25^{th} **June 2016**			
Parameters	**1 hour ahead forecasting**	**24 hours ahead forecasting**	**48 hours ahead forecasting**
Optimized model	bagEarthGCV	Cubist	Spikeslab
Accuracy(%)	73.08	65.38	62.5
r^2	0.85	0.94	0.94
RMSE	35.58	46.34	53.34

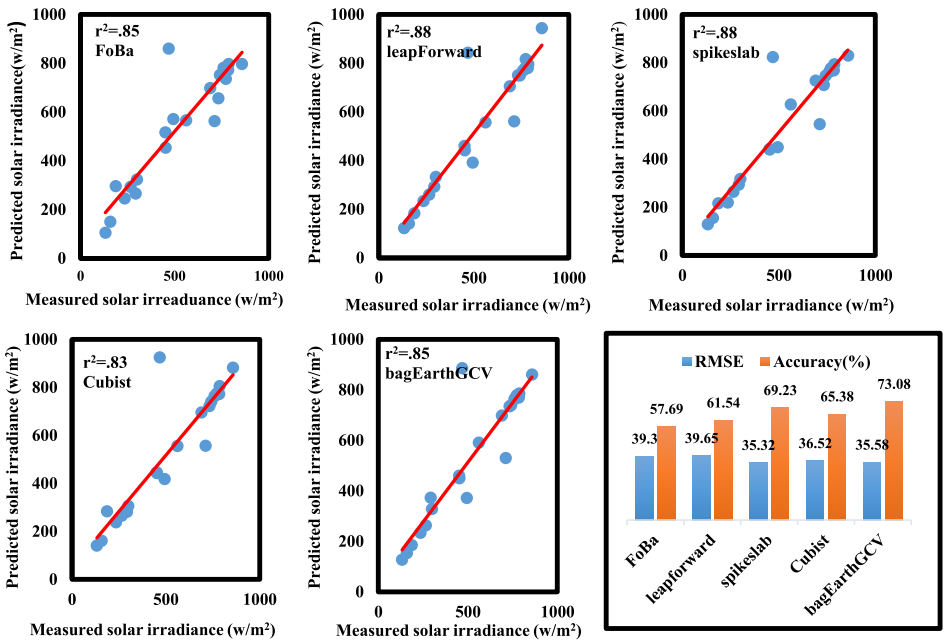


Figure 4.9: Correlation between predicted and measured solar irradiance for 25^{th} June, 1 hour ahead prediction

It is observed that unstable weather conditions as shown in Figure 4.5 and low

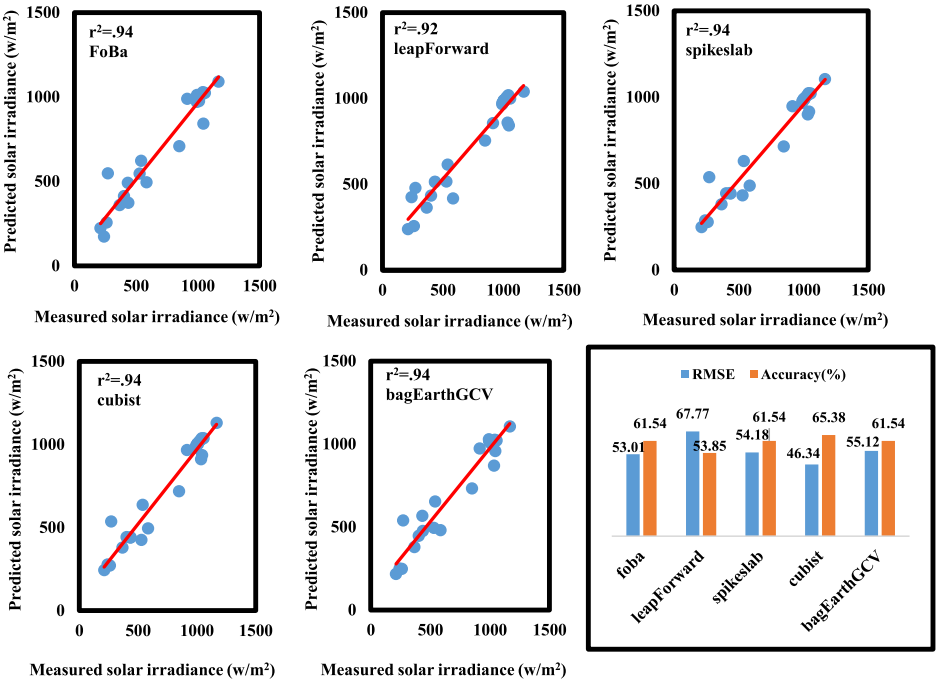


Figure 4.10: Correlation between predicted and measured solar irradiance for 25^{th} June, 24 hours ahead prediction

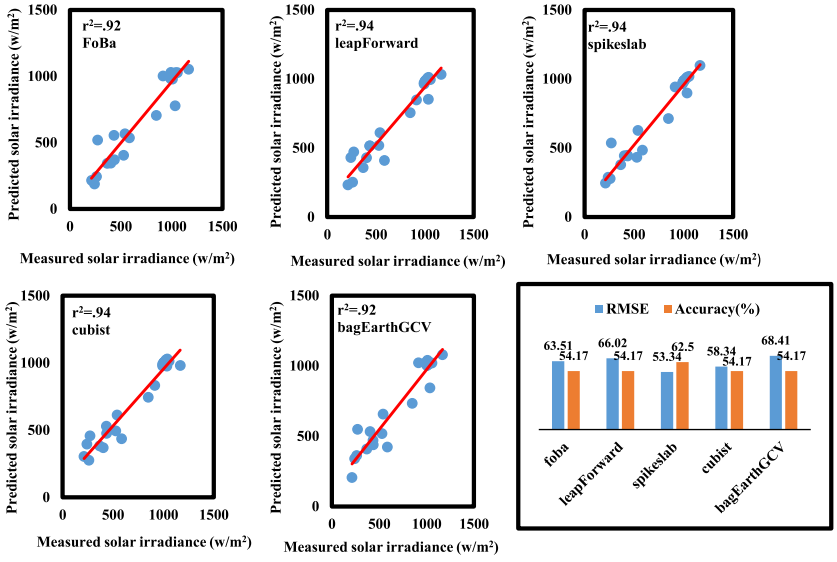


Figure 4.11: Correlation between predicted and measured solar irradiance for 25^{th} June, 48 hours ahead prediction

correlation with past days is the reason of low prediction accuracy and high RMSE in all forecasting horizons for 25^{th} June 2016.

3. 30^{th} ***August, 2016***

In one hour ahead prediction for 30^{th} August 2016, maximum prediction accuracy (83.78%) is achieved by Spikeslab with 0.98 correlation coefficient and 22.18 RMSE (Figure 4.12). In 24 hours ahead solar forecasting for 30^{th} August 2016, Spikeslab offers the maximum prediction accuracy (88.46%) with 0.92 correlation coefficient and 23.37 RMSE (Figure 4.13). In 48 hours ahead prediction for 30^{th} August 2016, maximum accuracy (91.67%) is gained by Spikeslab model with 0.92 correlation coefficient and 22.73 RMSE (Figure 4.14). The historical days from 15^{th} August to 14^{th} September ($\pm$15 days) (2010 to 2015)are used for training and days from 15^{th} August 2016 to last slot available are used for testing. Table 4.5 enlists the results of analysis.

Table 4.5: Observations form 30^{th} August, 2016 predictions

30^{th} **August, 2016**			
Parameters	**1 hour ahead forecasting**	**24 hours ahead forecasting**	**48 hours ahead forecasting**
Optimized model	Spikeslab	Spikeslab	Spikeslab
Accuracy(%)	83.78	88.46	91.67
r^2	0.98	0.92	0.92
RMSE	22.18	23.37	22.73

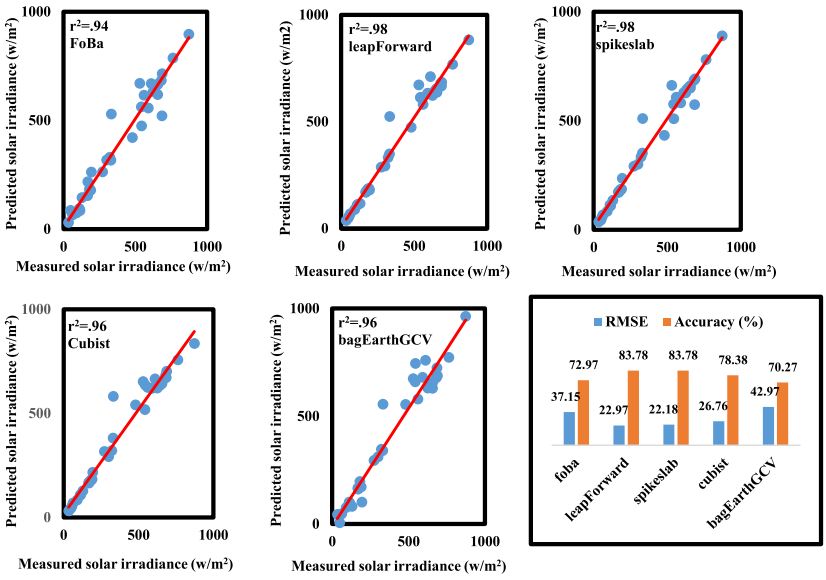


Figure 4.12: Correlation between predicted and measured solar irradiance for 30^{th} August, 1 hour ahead prediction

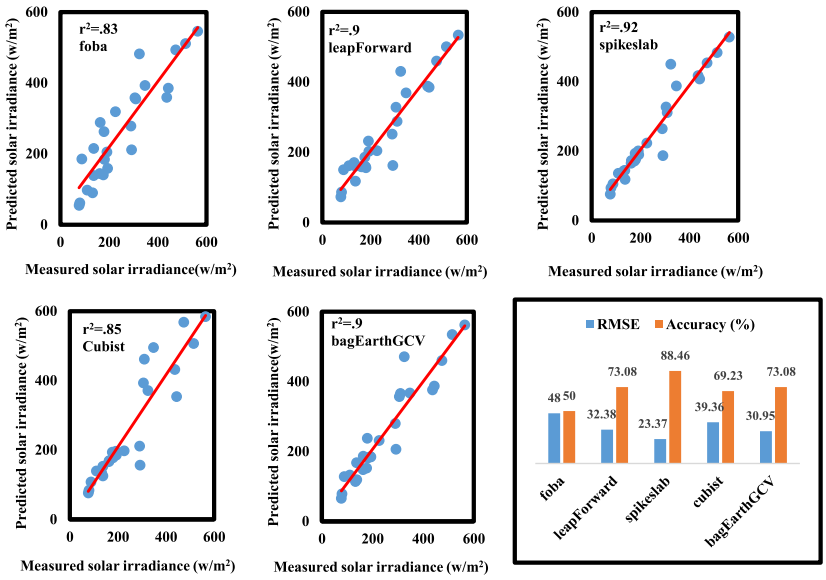


Figure 4.13: Correlation between predicted and measured solar irradiance for 30^{th} August, 24 hours ahead prediction

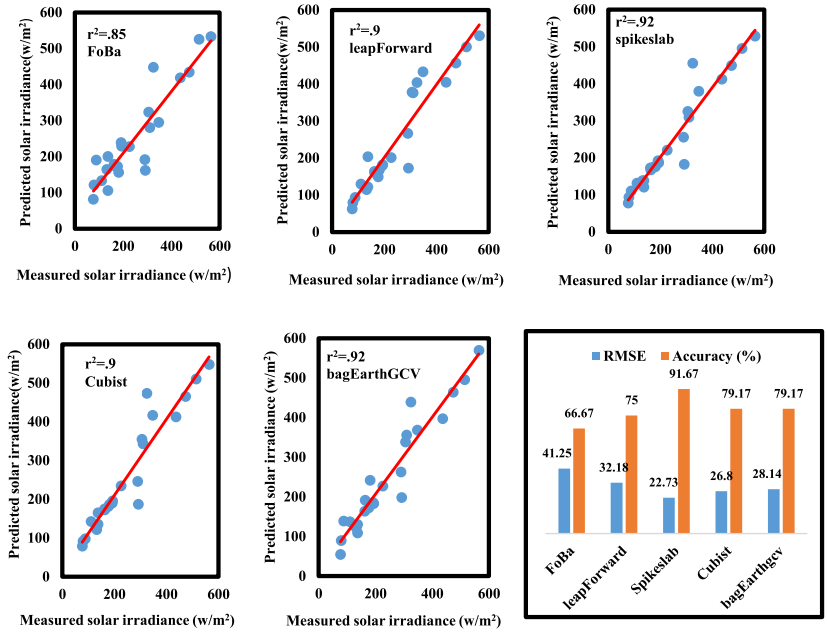


Figure 4.14: Correlation between predicted and measured solar irradiance for 30^{th} August, 48 hours ahead prediction

4. 31^{st} ***December 2016***

In 1 hour ahead solar forecasting for 31^{st} December 2016, maximum accuracy (93.86%) is given by Cubist model with 0.92 correlation coefficient and 10.3 RMSE Figure 4.15. The Spikeslab model offers highest prediction accuracy (92.7%) for 30^{th} August, 2016 in 24 hours ahead prediction horizon with 0.92 correlation coefficient and 10.73 RMSE (Figure 4.16). For 48 hours ahead forecasting, Spikeslab offers highest prediction accuracy (91.67%) with 0.94 correlation coefficient and 22.73 RMSE (Figure 4.17). The historical days from 16^{th} December to 15^{th} January (±15 days) (2010 to 2015) are used for training and days from 16^{th} December 2016 to last slot available are used for testing. Initial past time slots are taken from 7:00am. Table 4.6 enlists the results of analysis.

Table 4.6: Observations form 31^{st} December, 2016 predictions

31^{st} **December 2016**			
Parameters	**1 hour ahead forecasting**	**24 hours ahead forecasting**	**48 hours ahead forecasting**
Optimized model	Cubist	Spikeslab	Spikeslab
Accuracy(%)	93.86	92.7	91.67
r^2	0.92	0.92	0.94
RMSE	10.3	10.73	22.73

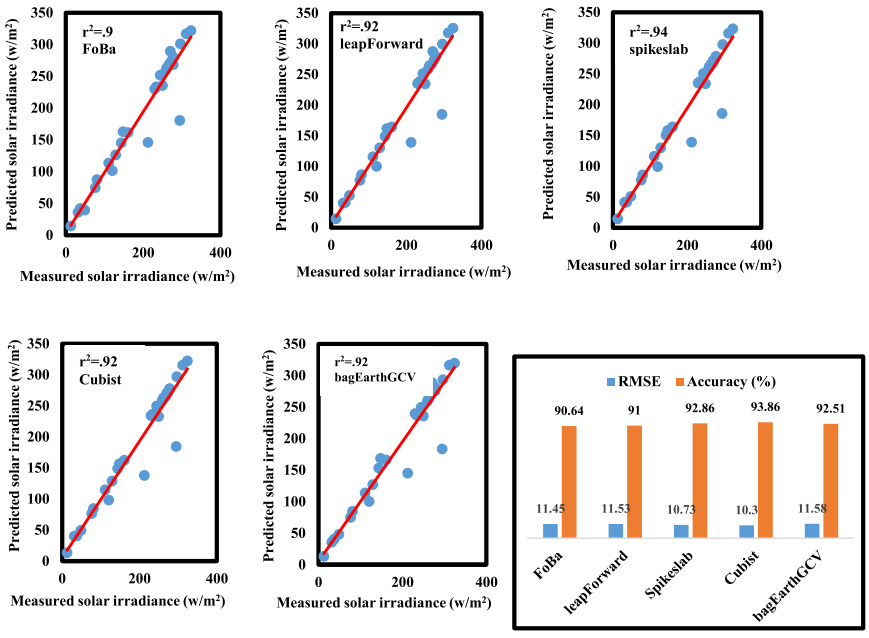


Figure 4.15: Correlation between predicted and measured solar irradiance for 31^{st} December, 1 hour ahead prediction

It is observed from the results obtained in the above section that Spikeslab and Cubist model achieves high prediction accuracy with respect to different forecasting

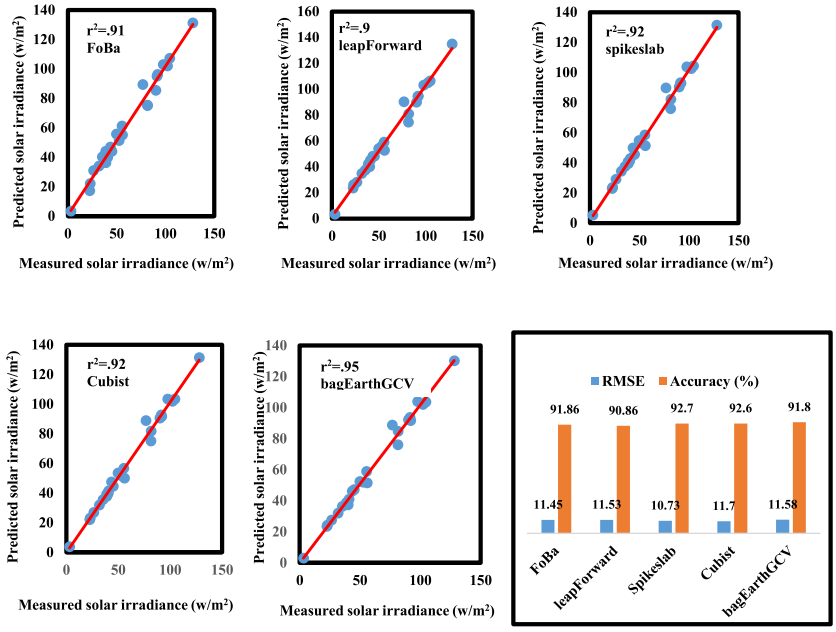


Figure 4.16: Correlation between predicted and measured solar irradiance for 31^{st} December, 24 hours ahead prediction

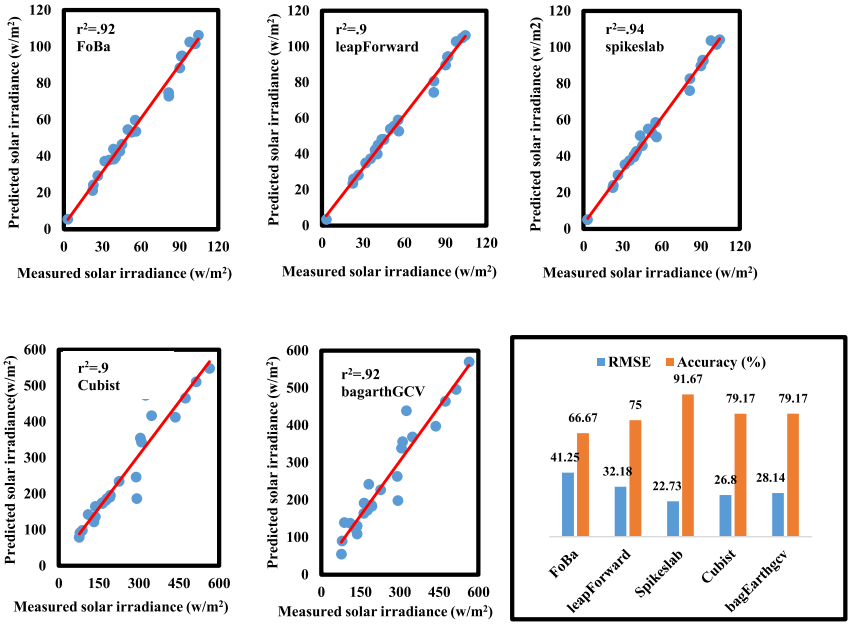


Figure 4.17: Correlation between predicted and measured solar irradiance for 31^{st} December, 48 hours ahead prediction

horizons for all seasons of a year. But not a particular model is equally good in all the test cases. this issue is addressed by ensemble modelling approach and presented in section 4.5.

4.4.2 Concluding Remarks

The applicability of five machine learning models, FoBa, leapForward, Spikeslab, Cubist, and bagEarthGCV in modeling solar irradiance prediction has been investigated and evaluated under seasonal effects using the same test platform and datasets. The main contribution is the best parameter selection for these models to improve the prediction accuracy in different forecasting horizons ranging from 1 hour ahead to 48 hours ahead. The performance is evaluated by statistical indices including correlation coefficient, RMSE, and prediction accuracy(%) for each model. Regarding the results obtained in test 1, accuracy of a model depends upon quality of selected data for model training. For different days of a year (11^{th} March, 25^{th} June, 30^{th} August, and 31^{st} December), the performance matrix (r^2, RMSE, and accuracy) for leapForward (0.99, 3.07, 96.08%), Cubist (0.98, 35.18, 78.95%), bagEarthGCV (0.98, 11.27, 89.47%) and FoBa (1, 2.92, 98.21%) respectively is achieved.

In test 2, it is observed that FoBa, leapForward, and Cubist performs well with large set of past time slots according to solar irradiance availability (7:00am onwards) whereas Spikeslab works with less past samples. It is observed that the past samples with considerable solar intensity affect the prediction accuracy and this parameter is season dependent. The results obtained in test 3 shows that results obtained by Spikeslab and Cubist model are very promising and stable with respect to different forecasting horizons. The prediction accuracy with different forecasting horizons (1 hour ahead, 24 hours ahead, and 48 hours ahead) gained by Spikeslab for 11^{th} March (86.27%, 66.67% and, 69.09% respectively), for 25^{th} June (69.09%, 61.54%, and 62.5% respectively), 30^{th} August (83.78%, 88.46%, and 91.67% respectively) and 31^{st} December (92.86%, 92.8%, and 91.67% respectively) are satisfactory and stable. Similarly, Cubist achieves (84.31%, 79.17%, and 58.18% respectively) for 11^{th} March, (58.18%, 65.38%, and 54.17% respectively) for 25^{th} June, (78.38%, 69.23%, and 79.17% respectively) for 30^{th} August and (93.86%, 92.6%, and 79.17% respectively) for 31^{st} December.

From the results, it is concluded that solar irradiance forecasting with such machine learning models is recent. The productive study in this field leads to optimized solar forecasting approach than conventional methods.

Part B

4.5 Statistical Ensemble Machine Learning Approach for Long Term Solar Irradiance Forecasting

The proposed work presents a novel machine learning framework of nine models for two days ahead solar forecasting with hourly resolution. The objective is to implement an approach that forecast solar irradiance traces close to real solar profiles. FoBa, leapForward, Spikeslab, Cubist, and bagEarthGCV models are evaluated for future irradiance estimation. To achieve more precise and consistent forecast, four Statistical Ensemble (SE) approaches are presented. The trained model performances are verified in terms of correlation coefficient, Root Mean Square Error (RMSE), and forecast accuracy(%). The simulations are conducted using *R* interface and on real data set collected from National Renewable Energy Laboratory (NREL). The obtained results validate the strong potential of these models for long term forecasting. Specially, Generalized Mean Ensemble (GME) approach gives highly accurate and consistent results. Above 70% prediction accuracy is achieved in different seasonal days and for different forecasting horizons ranging from 1 hour to 48 hours ahead. Further, in stable weather conditions, prediction accuracy of the order of 92% is achieved for two days ahead predictions.

4.5.1 Ensembled Forecasting

Machine learning based prediction models works to trace a suitable hypothesis space that perform accurate predictions for a particular problem. In ensemble forecasting individual prediction model is trained as forecasting member and multiple hypothesis are combined to generate a better hypothesis which is expected to be more accurate than any other ensemble member. In proposed work, Statistical Ensemble (SE) approach is applied on machine learning models to circumvent the effect of forecasting horizon and seasonal variability. Statistical ensemble performs generalized mean ensemble including pythagorean means (arithmetic, harmonic, and quadratic means) and median rule. The methodology is shown in Figure 4.18 and is described below:

1. If a is a non-zero real number and z_1,...,z_n are positive real numbers, the generalized

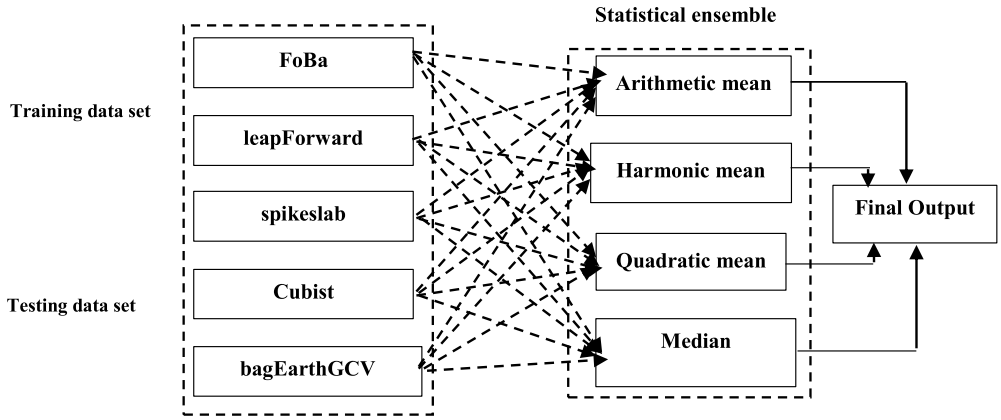


Figure 4.18: Statistical ensemble approach

mean with exponent a of these real numbers can be defined as:

$$GM_a(z_1, ..., z_n) = (\frac{1}{n}\sum_{i=1}^{n} z_i^a)^{\frac{1}{a}} \qquad -\infty \leq a \leq +\infty \tag{4.16}$$

where GM is the generalized mean and n is the number of terms. Different mean functions are evaluated by considering different value of a.

2. Arithmetic mean is the special case with $a = 1$ and defined as:

$$GM_1(z_1, ..., z_n) = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{4.17}$$

3. Harmonic mean is defined as the quotient of number of given values and sum of reciprocals of the given values. It is the special case with $a = -1$ and used to calculate the average of variables expressed as the ratio of two measuring units. Equation 4.18 gives the mathematical formula for harmonic mean.

$$GM_{-1}(z_1, ..., z_n) = (\frac{1}{n}\sum_{i=1}^{n} z_i^{-1})^{-1} = \frac{n}{\sum_{i=1}^{n}\frac{1}{z_i}} \tag{4.18}$$

4. Quadratic mean assign greater weight to larger values in the set and always equal to or greater than arithmetic mean. It is the special case with $a = 2$ and mathematically given in equation 4.19:

$$GM_2(z_1, .., z_n) = (\frac{1}{n}\sum_{i=1}^{n} z_i^2)^2 \tag{4.19}$$

4.5.2 Simulation Results and Discussions

A series of simulation experiments are performed to check the accuracy of five solar irradiance forecasting models and their ensemble approach. A common platform is used to run all experiments. To evaluate the forecasting models in different weather conditions, four days, 10^{th} April (spring), 18^{th} June (summer), 26^{th} November (autumn), and 20^{th} January (winter) from different seasons of year 2016 are used for testing. It is seen in Figure 4.19 maximum sunshine hours and peak irradiance level is highly dependent upon seasons. On 18^{th} June (summer) maximum solar irradiance with large sunshine hours is observed. In spring season (10^{th} April), high irradiance is available during ideal hours of the day. Autumn (26^{th} November) and winter (20^{th} January) possesses smooth irradiance with low maximum and minimum solar intensity thresholds and less sunshine hours.

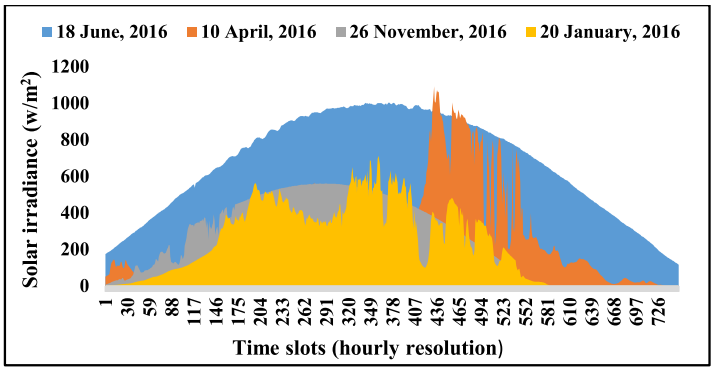


Figure 4.19: Variation in solar irradiance with different seasons

To validate the prediction effectiveness of proposed models, experiments are conducted in two phases. In the first phase, all five models are evaluated for two days ahead solar forecasting with appropriate number of past time slots and past days selection. In the second experiment, a statistical ensemble approach is applied to mitigate the uncertain performance measures of models with respect to forecasting horizons and seasonal effects.

4.5.3 Model Evaluation

The performance comparison between different models is difficult because of various time scales of predicted data, variability in meteorological conditions, and different forecasting horizons. To estimate the accuracy of models, graphical tool (scatter plot) is used with experimental measurements. The scatter plot shows the systematic bias and range of deviation between measured and predicted solar irradiance. The accuracy and RMS

plots are used for statistical analysis. Test data from year 2016 is used for evaluation and a pool of last 30 days is selected from each year (2010-2015) for training.

4.5.3.1 For 20^{th} January, 2016 (winter)

Highest prediction accuracy (87.27%) (Figure 4.22(a)) is achieved by Spikeslab with r^2 value of 0.98 (Figure 4.21a) and Figure 4.20(a)) and 9.89 RMSE (Figure 4.22(b)) for 1 hour ahead prediction. In this horizon minimum accuracy is offered by FoBa (70.91%) with 16.21 RMSE.

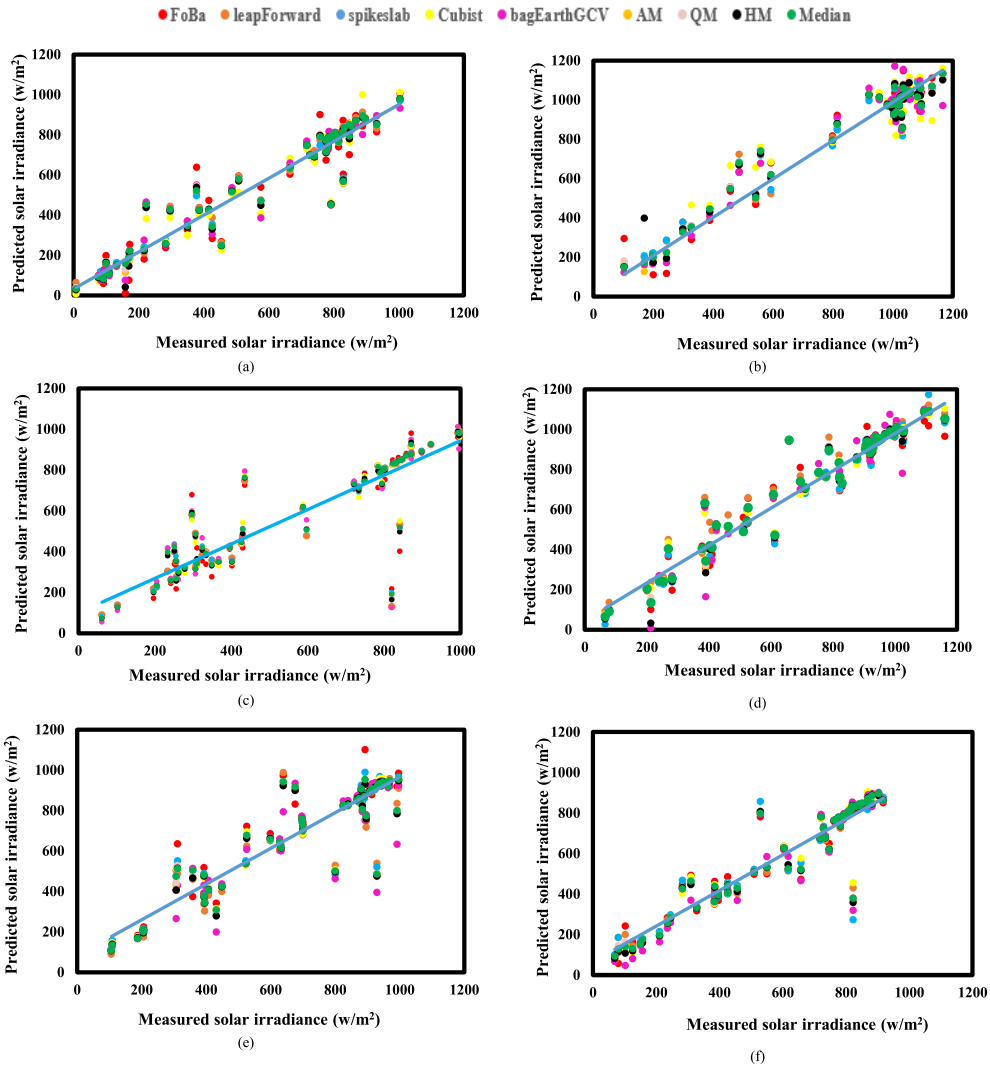


Figure 4.20: Correlation between measured and predicted solar irradiance for (a)1 hour (b)6 hours (c)12 hours (d)18 hours (e)24 hours, and (f)48 hours ahead forecasting horizon for 20^{th} January, 2016

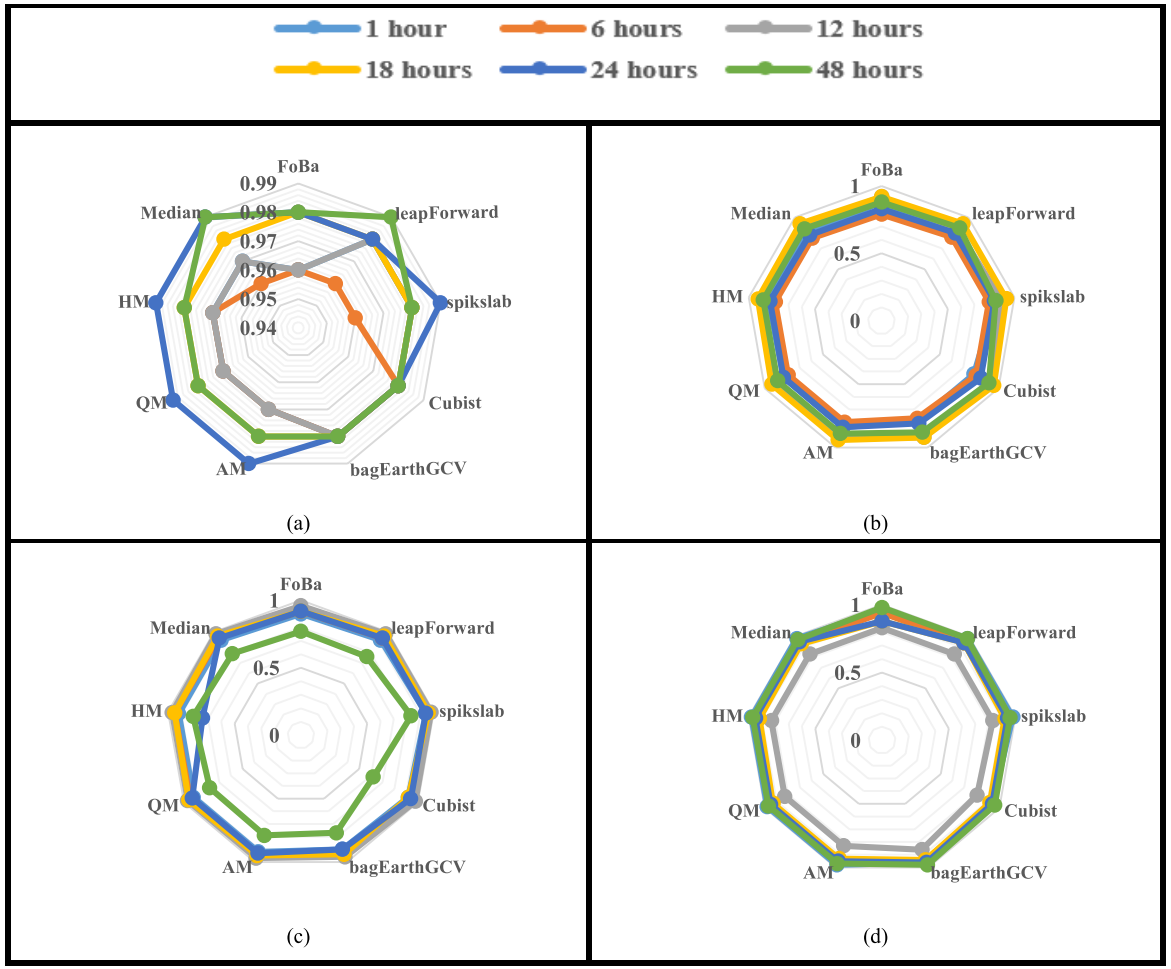


Figure 4.21: Correlation coefficient

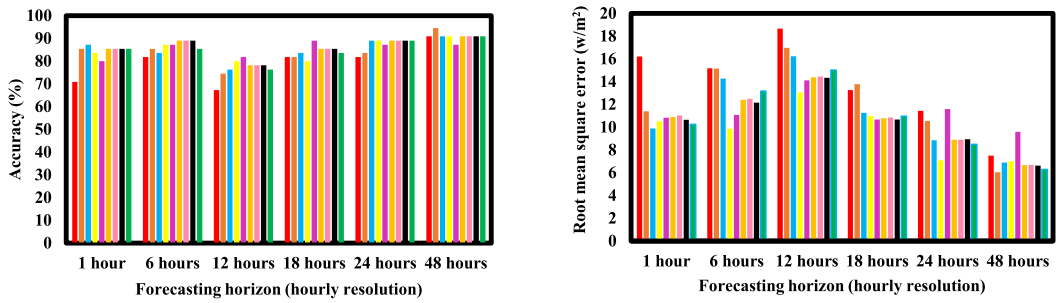


Figure 4.22: (a)Prediction accuracy and (b)RMS error in six forecasting horizons for 20^{th} January, 2016

For 6 hours ahead prediction, Cubist gain maximum accuracy (87.27%) (Figure 4.22(a)) with 0.98 r^2 (Figure 4.21a and Figure 4.20(b)) and 9.87 RMSE (Figure 4.22(b)). FoBa results in minimum accuracy (81.82%) with 15.17 RMSE. For next forecasting horizon (12 hours ahead), bagEarthGCV attain maximum accuracy (81.82%) (Figure 4.22(a)) with 0.98 r^2 (Figure 4.21a and Figure 4.20(c)) and 14.13 RMSE (Figure 4.22(b)). Minimum accuracy (67.27%) is given by FoBa with 18.66 RMSE. For 18 hours ahead prediction bagEarthGCV offer high accuracy (89.09%) (Figure 4.22(a)) with r^2 value of 0.98 (Figure 4.21a and Figure 4.20(d)) and 10.68 RMSE (Figure 4.22(b)). Cubist Gains minimum accuracy (80%) with 11.99 RMSE. Cubist achieves highest accuracy (89.09%) (Figure 4.22(a)) with 0.98 r^2 (Figure 4.21a) and Figure 4.20(e)) and 7.09 RMSE (Figure 4.22(b)) for 24 hours ahead prediction. Lowest performance level is shown by FoBa with (81.82%) prediction accuracy and 11.44 RMSE. For 48 hours ahead forecasting leapForward gain maximum accuracy (94.55%) (Figure 4.22 (a)) with r^2 value of 0.99 (Figure 4.21a) and Figure 4.22(b)) and 6.04 RMSE (Figure 4.20(h)). Minimum prediction accuracy (87.27) is achieved by using bagEarthGCV with 9.6 RMSE. The performance of different models for 20^{th} January is summarized in table 4.7.

4.5.3.2 For 10^{th} April, 2016 (spring)

The performance of different models for 10^{th} April is summarized in table 4.8. Highest prediction accuracy (69.09%) (Figure 4.24(a)) is achieved by Spikeslab with r^2 value of 0.86 (Figure 4.21(a) and Figure 4.23(a)) and 46.47 RMSE (Figure 4.24(b)) for 1 hour ahead prediction. FoBa gains minimum accuracy (60%) with 46.93 RMSE. For 6 hours ahead prediction, leapForward gain maximum accuracy (70.91%) (Figure 4.24(a)) with 0.81 r^2 (Figure 4.21(b) and Figure 4.23(b)) and 56.54 RMSE (Figure 4.24(b)). Minimum accuracy (63.64%) is offered by Cubist with 58.6 RMSE.

Table 4.7: Performance of nine machine learning models for 20^{th} January, 2016

Machine learning methodologies							Ensemble approaches			
20-01-2016(Winter)		**FoBa**	**leapForward**	**Spikeslab**	**Cubist**	**bagEarthGCV**	**Arithmetic mean**	**Harmonic mean**	**Quadratic mean**	**Median**
1 hour	r	0.98	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98
	r^2	0.96	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97
	RMSE	16.21	11.4	9.89	10.51	10.84	10.9	11.03	10.65	10.25
	accuracy	70.91	85.45	87.27	83.64	80	85.45	85.45	85.45	85.45
6 hour	r	0.98	0.98	0.98	0.99	0.99	0.98	0.98	0.98	0.98
	r^2	0.96	0.96	0.96	0.98	0.98	0.97	0.97	0.97	0.96
	RMSE	15.17	15.16	14.28	9.87	11.08	12.4	12.5	12.17	13.17
	accuracy	81.82	85.45	83.64	87.27	87.27	89.09	89.09	89.09	85.45
12 hour	r	0.98	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98
	r^2	0.96	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97
	RMSE	18.66	16.97	16.23	13.06	14.13	14.39	14.46	14.35	15.02
	accuracy	67.27	74.55	76.36	80	81.82	78.18	78.18	78.18	76.36
18 hour	r	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	r^2	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	RMSE	13.26	13.78	11.27	10.99	10.68	10.78	10.85	10.67	10.98
	accuracy	81.82	81.82	83.64	80	89.09	85.45	85.45	85.45	83.63
24 hour	r	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	r^2	0.98	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.99
	RMSE	11.44	10.54	8.85	7.09	11.6	8.89	8.89	8.95	8.5
	accuracy	81.82	83.64	89.09	89.09	87.27	89.09	89.09	89.09	89.09
48 hour	r	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	r^2	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.99
	RMSE	7.51	6.04	6.9	7.02	9.6	6.67	6.7	6.63	6.3
	accuracy	90.91	94.55	90.91	90.91	87.27	90.9	90.9	90.9	90.9

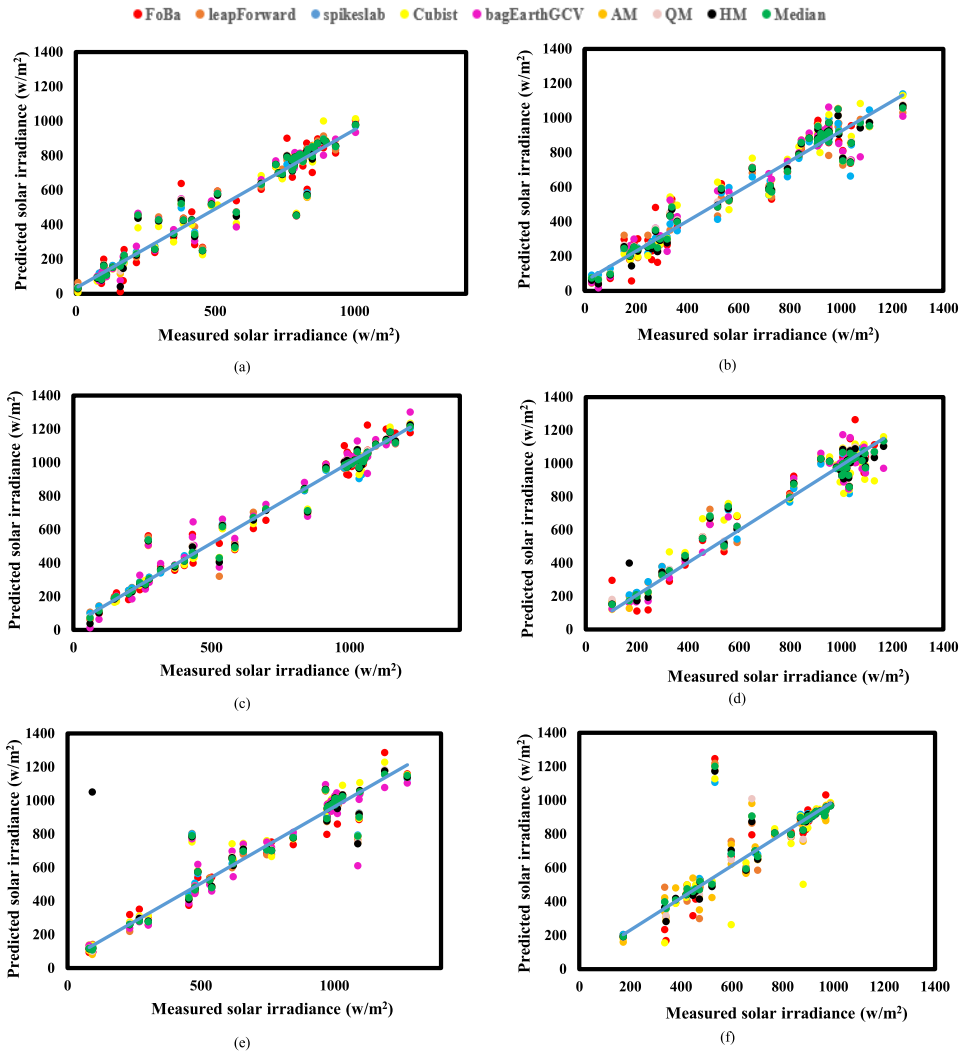


Figure 4.23: Correlation between measured and predicted solar irradiance for (a)1 hour (b)6 hours (c)12 hours (d)18 hours (e)24 hours, and (f)48 hours ahead forecasting horizon for 10^{th} April, 2016

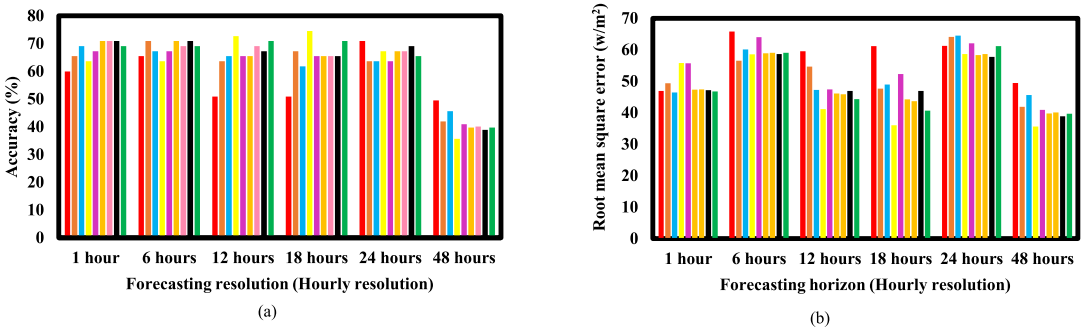


Figure 4.24: (a)Prediction accuracy and (b)RMS error in six forecasting horizons for 10^{th} April, 2016

Table 4.8: Performance of nine machine learning models for 10^{th} April, 2016

Machine learning methodologies							Ensemble approaches			
10-04-2016(Spring)		**FoBa**	**leapForward**	**spikeslab**	**Cubist**	**bagEarthGCV**	**Arithmetic mean**	**Harmonic mean**	**Quadratic mean**	**Median**
1 hour	**r**	0.94	0.92	0.93	0.89	0.89	0.92	0.92	0.92	0.92
	r^2	0.88	0.85	0.86	0.79	0.79	0.84	0.84	0.85	0.85
	RMSE	46.93	49.36	46.47	55.78	55.71	47.3	47.4	47.14	46.75
	accuracy	60	65.45	69.09	63.64	67.27	70.9	70.9	70.9	69.09
6 hours	**r**	0.89	0.9	0.9	0.9	0.88	0.89	0.89	0.89	0.89
	r^2	0.79	0.81	0.81	0.81	0.77	0.8	80	0.8	0.8
	RMSE	65.83	56.54	60.12	58.6	64.03	58.89	59.02	58.62	59.07
	accuracy	65.45	70.91	67.27	63.64	67.27	70.9	69.09	70.9	69.09
12 hour	**r**	0.96	0.94	0.95	0.96	0.96	0.95	0.95	0.95	0.95
	r^2	0.92	0.88	0.9	0.92	0.92	0.91	0.91	0.91	0.91
	RMSE	59.54	54.64	47.22	41.13	47.45	46.13	45.83	46.9	44.3
	accuracy	50.91	63.64	65.45	72.73	65.45	65.45	69.09	67.27	70.9
18 hour	**r**	0.96	0.97	0.97	0.98	0.96	0.97	0.97	0.96	0.97
	r^2	0.92	0.94	0.94	0.96	0.92	0.94	0.94	0.93	0.94
	RMSE	61.14	47.65	48.96	36	52.3	44.21	43.69	46.96	40.67
	accuracy	50.91	67.27	61.82	74.55	65.45	65.45	65.45	65.45	70.9
24 hour	**r**	0.91	0.92	0.92	0.92	0.9	0.92	0.91	0.92	0.91
	r^2	0.83	0.85	0.85	0.85	0.81	0.84	0.84	0.84	0.83
	RMSE	61.22	64.13	64.51	58.61	62.06	58.36	58.65	57.73	61.19
	accuracy	70.91	63.64	63.64	67.27	63.64	67.27	67.27	69.09	65.45
48 hour	**r**	0.94	0.95	0.93	0.96	0.94	0.94	0.94	0.94	0.94
	r^2	0.88	0.9	0.86	0.92	0.88	0.89	0.89	0.89	0.89
	RMSE	49.48	41.91	45.67	35.66	40.9	39.77	40.13	38.91	39.72
	accuracy	65.45	67.27	76.36	76.36	70.91	70.9	70.9	74.54	72.72

For next forecasting horizon (12 hours ahead), Cubist attain maximum accuracy (72.73%) (Figure 4.24(a)) with 0.92 r^2 (Figure 4.21(b) and Figure 4.23(c)) and 41.13 RMSE (Figure 4.24(b)). Minimum performance level is gained by FoBa with 50.91% prediction accuracy and 59.54 RMSE. For 18 hours ahead prediction, Cubist offer high accuracy (74.55%) (Figure 4.24(a)) with r^2 value of 0.96 (Figure 4.21(b) and Figure 4.23(d)) and 36 RMSE (Figure 4.24(b)). FoBa Offers minimum prediction accuracy (50.91%) with 61.14 RMSE. FoBa achieves highest accuracy (70.91%) (Figure 4.24(a)) with 0.83 r^2 (Figure 4.21(b) and (Figure 4.23(e)) and 61.22 RMSE (Figure 4.24(b)) for 24 hours ahead prediction. The Spikeslab model gains minimum accuracy (63.64%) with 64.51 RMSE. For 48 hours ahead forecasting, Cubist gain maximum accuracy (76.36%) (Figure 4.24(a)) with r^2 value of 0.92 (Figure 4.21(b) and Figure 4.23(f)) and 35.66 RMSE (Figure 4.24(b)). Minimum prediction accuracy (65.45%) is given by FoBa with 49.48 RMSE.

4.5.3.3 For 18^{th} June, 2016 (summer)

The performance of different models for 18^{th} June is summarized in table 4.9. Highest prediction accuracy (74.55%) (Figure 4.26(a)) is achieved by Spikeslab with r^2 value of 0.94 (Figure 4.21(c) and (Figure 4.25(a)) and 43.68 RMSE (Figure 4.26(b)) for 1 hour ahead prediction. The bagEarthGCV model gains minimum prediction accuracy (60%) with 56.35 RMSE. For 6 hours ahead prediction, Cubist gain maximum accuracy (56.36%) (Figure 4.26(a)) with 0.94 r^2 (Figure 4.21(c) and Figure 4.25(b)) and 55.77 RMSE (Figure 4.26(b)). Minimum prediction accuracy (23.64%) is offered by FoBa with 78.03 RMSE. For next forecasting horizon (12 hours ahead), Cubist attain maximum accuracy (80%) (Figure 4.26(a)) with 0.98 r^2 (Figure 4.21(c) and Figure 4.25(c)) and 27.05 RMSE (Figure 4.26(b)). FoBa gives minimum prediction accuracy (60%) with 41.98 RMSE. For 18 hours ahead prediction Spikeslab offer high accuracy (67.27%) (Figure 4.26(a)) with r^2 value of 0.96 (Figure 4.21(c) and Figure 4.25(d)) and 44.6 RMSE (Figure 4.26(b)). Minimum prediction accuracy (58.18%) is given by FoBa with 58.48 RMSE. The Spikeslab model achieves highest accuracy (76.36%) (Figure 4.26(a)) with 0.94 r^2 (Figure 4.21(c) and Figure 4.25(e)) and 42.55 RMSE (Figure 4.26(b)) for 24 hours ahead prediction. Lowest performance level has been offered by bagEarthGCV with 65.45% prediction accuracy and 47.53 RMSE. For 48 hours ahead forecasting leapForward gain maximum accuracy (81.82%) (Figure 4.26(a)) with r^2 value of 0.76 (Figure 4.21(c) and Figure 4.25(f)) and 42.84 RMSE (Figure 4.26(b)). In this horizon FoBa offers minimum prediction accuracy (72.73%) with 41.72 RMSE.

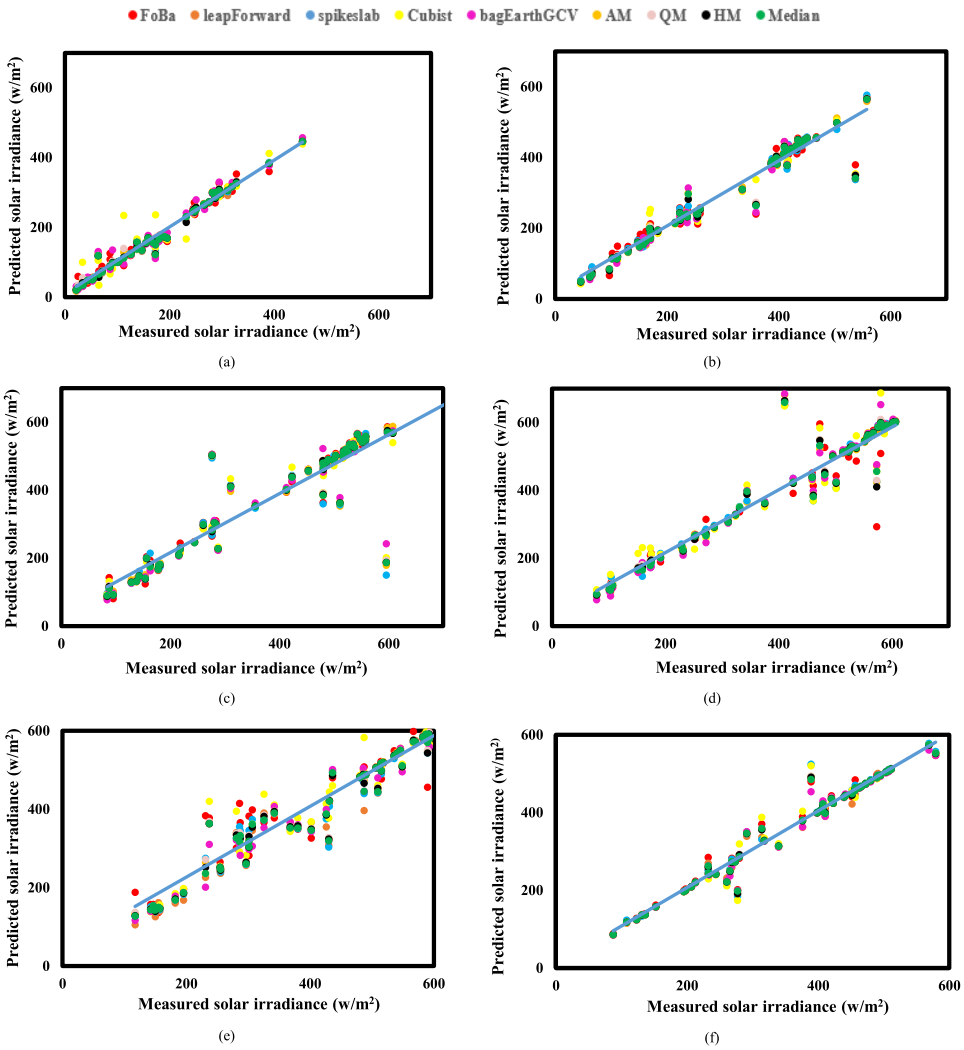


Figure 4.25: Correlation between measured and predicted solar irradiance for (a)1 hour (b)6 hours (c)12 hours (d)18 hours (e)24 hours, and (f)48 hours ahead forecasting horizon for 18^{th} June, 2016

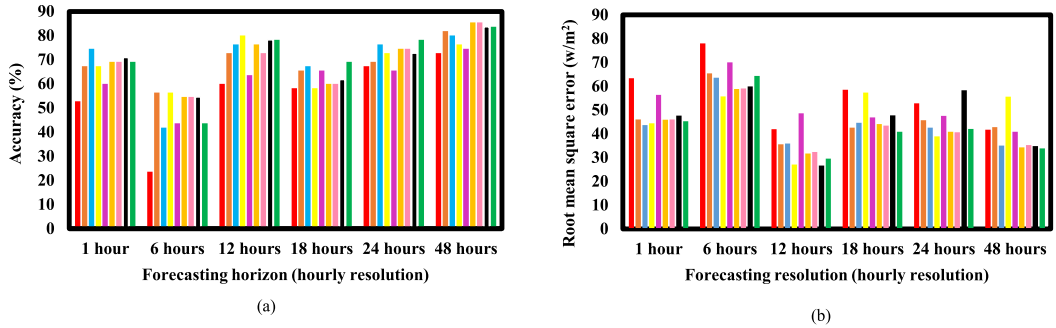


Figure 4.26: (a)Prediction accuracy and (b)RMS error in six forecasting horizons for 18^{th} June, 2016

Table 4.9: Performance of nine machine learning models for 18^{th} June, 2016

Machine learning methodologies							Ensemble approaches			
18-06-2016(Summer)		**FoBa**	**leapForward**	**spikeslab**	**Cubist**	**bagEarthGCV**	**Arithmetic mean**	**Harmonic mean**	**Quadratic mean**	**Median**
1 hour	**r**	0.98	0.96	0.97	0.96	0.95	0.96	0.96	0.96	0.96
	r^2	0.9	0.92	0.94	0.92	0.9	0.92	0.92	0.92	0.92
	RMSE	63.42	46.03	43.68	44.45	56.35	45.88	46.05	47.66	45.27
	accuracy	52.73	67.27	74.55	67.27	60	69.09	69.09	70.9	69.09
6 hours	**r**	0.96	0.97	0.98	0.97	0.96	0.97	0.97	0.97	0.97
	r^2	0.92	0.94	0.96	0.94	0.92	0.95	0.95	0.95	0.95
	RMSE	78.03	65.4	63.6	55.77	70.02	58.89	59.11	59.89	64.31
	accuracy	23.64	56.36	41.82	56.36	43.64	54.54	54.54	54.54	43.63
12 hour	**r**	0.98	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.99
	r^2	0.96	0.98	0.98	0.98	0.96	0.97	0.97	0.97	0.98
	RMSE	41.98	35.53	35.85	27.05	48.6	31.74	32.3	26.65	29.56
	accuracy	60	72.73	76.36	80	63.64	76.36	72.72	78.18	78.18
18 hour	**r**	0.96	0.98	0.98	0.96	0.97	0.97	0.98	0.97	0.98
	r^2	0.92	0.96	0.96	0.92	0.94	0.95	0.96	0.95	0.96
	RMSE	58.48	42.63	44.6	57.35	46.92	44.12	43.39	47.7	40.87
	accuracy	58.18	65.45	67.27	58.18	65.45	60	60	61.81	69.09
24 hour	**r**	0.96	0.97	0.97	0.97	0.95	0.96	0.96	0.86	0.97
	r^2	0.92	0.94	0.94	0.94	0.9	0.93	0.93	0.74	0.94
	RMSE	52.81	45.66	42.55	38.87	47.53	40.84	40.62	58.36	42.01
	accuracy	67.27	69.09	76.36	72.73	65.45	74.54	74.54	72.72	78.18
48 hour	**r**	0.88	0.87	0.91	0.79	0.88	0.88	0.88	0.9	0.89
	r^2	0.77	0.76	0.83	0.62	0.77	0.79	0.78	0.81	0.79
	RMSE	41.72	42.84	35.02	55.65	40.82	34.29	35.22	34.82	33.84
	accuracy	72.73	81.82	80	76.36	74.55	85.45	85.45	83.63	83.63

4.5.3.4 For 26^{th} November, 2016 (autumn)

The performance of different models for 26^th November is summarized in table 4.10. Highest prediction accuracy (90.91%) (Figure 4.28(a)) is achieved by Spikeslab with r^2 value of 0.98 (Figure 4.21(d) and Figure 4.27(a)) and 6.41 RMSE (Figure 4.28(b)) for 1 hour ahead prediction. Minimum accuracy (78.18%) is given by FoBa with 13.14 RMSE. For 6 hours ahead prediction, spikeslab gain maximum accuracy (85.45%) (Figure 4.28(a)) with 0.94 r^2 (Figure 4.21(d) and Figure 4.27(b)) and 14.37 RMSE (Figure 4.28(b)). Minimum accuracy of the order of 63.64% has been offered by FoBa with 19.77 RMSE. For next forecasting horizon (12 hours ahead), leapForward attain maximum accuracy (78.18%) (Figure 4.28(a)) with 0.83 r^2 (Figure 4.21(d) and Figure 4.27(c)) and 28.38 RMSE (Figure 4.28(b)). FoBa gains minimum accuracy (65.45%) with 33.03 RMSE. For 18 hours ahead prediction leapForward offer high accuracy (83.64%) (Figure 4.28(a)) with r^2 value of 0.94 (Figure 4.21(d) and Figure 4.27(d)) and 17.81 RMSE (Figure 4.28(b)). Cubist offers minimum prediction accuracy (65.45%) with 27.45 RMSE. Cubist achieves highest accuracy (74.55%) (Figure 4.28(a)) with 0.94 r^2 (Figure 4.21(d) and Figure 4.27(e)) and 19.95 RMSE (Figure 4.28(b)) for 24 hours ahead prediction. Minimum accuracy (58.18%) is given by FoBa with 32.22 RMSE. For 48 hours ahead forecasting spikeslab gain maximum accuracy (89.09%) (Figure 4.28(a) with r^2 value of 0.96 (Figure 4.21(d) and Figure 4.27(f)) and 11.58 RMSE (Figure 4.28(b)). The bagEarthGCV model achieves minimum prediction accuracy (80%) with 10.84 RMSE. It is observed from the results that forecasting accuracy achieved by different machine learning models is satisfactory with respect to all forecasting horizons and in different seasons but not a particular model always perform well as compare to rest of others. Thus, there is a need to combine the output of independent models to generate a single optimized output. To address this issue, an ensembled machine learning approach is proposed in section 4.5.4.

4.5.4 Ensemble Evaluation

In proposed work, statistical ensemble approach is used to generate a single optimized output by combining output of different models. In statistical ensembling, four approaches (arithmetic mean, quadratic mean, harmonic mean, and median) are evaluated.

In comparison with independent model prediction as in previous section, for 20^{th} January (table 4.7) in 1 hour ahead prediction, median achieves 85.45% prediction accuracy (Figure 4.22(a)) with 0.97 r^2 (Figure 4.21(a) and Figure 4.20(a)) and 10.25 RMSE (Figure 4.22(b)).

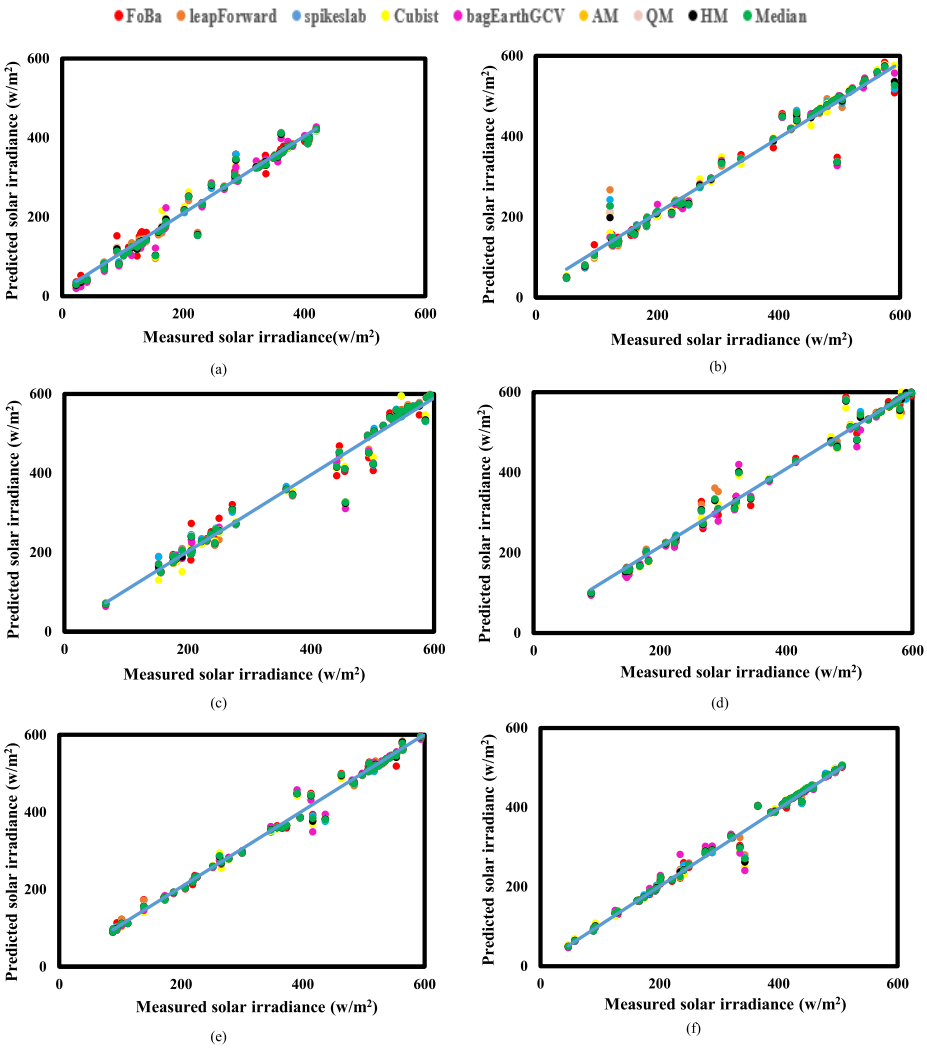


Figure 4.27: Correlation between measured and predicted solar irradiance for (a)1 hour (b)6 hours (c)12 hours (d)18 hours (e)24 hours, and (f)48 hours ahead forecasting horizon for 26^{th} November, 2016

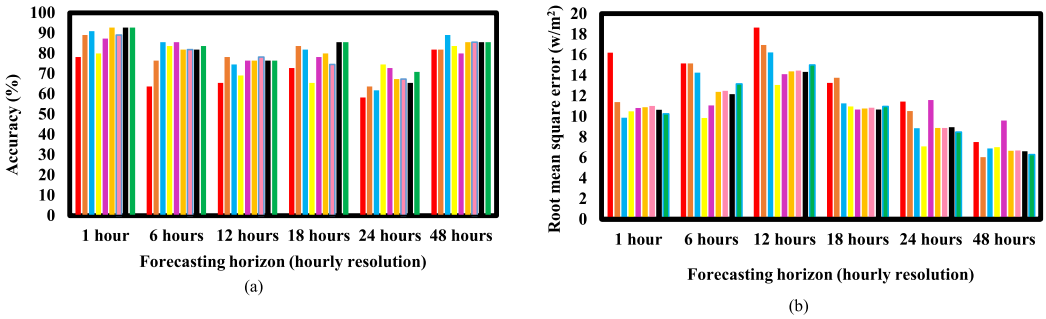


Figure 4.28: (a)Prediction accuracy and (b)RMS error in six forecasting horizons for 26^{th} November, 2016

Table 4.10: Performance of nine machine learning models for 26^{th} November, 2016

Machine learning methodologies							Ensemble approaches			
26-11-2016(Autumn))		**FoBa**	**leapForward**	**spikeslab**	**Cubist**	**bagEarthGCV**	**Arithmetic mean**	**Harmonic mean**	**Quadratic mean**	**Median**
1 hour	**r**	0.98	0.99	0.99	0.97	0.98	0.99	0.99	0.99	0.99
	r^2	0.96	0.98	0.98	0.94	0.96	0.98	0.98	0.98	0.98
	RMSE	13.14	6.91	6.41	13.09	11.91	8.42	8.19	7.44	8.49
	accuracy	78.18	89.09	90.91	80	87.27	92.72	89.09	92.72	92.72
6 hours	**r**	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	r^2	0.94	0.94	0.94	0.94	0.94	0.95	0.95	0.95	0.94
	RMSE	19.77	15.92	14.37	13.16	14.53	13.52	13.56	13.47	13.44
	accuracy	63.64	76.36	85.45	83.64	85.45	81.81	81.81	81.81	83.63
12 hour	**r**	0.91	0.91	0.91	0.9	0.93	0.91	0.91	0.91	0.91
	r^2	0.83	0.83	0.83	0.81	0.86	0.83	0.83	0.83	0.83
	RMSE	33.03	28.38	30.98	32.5	26.52	6.59	29.33	29.58	28.88
	accuracy	65.45	78.18	74.55	69.09	76.36	76.36	78.18	76.36	76.36
18 hour	**r**	0.94	0.97	0.96	0.96	0.97	0.96	0.96	0.96	0.96
	r^2	0.88	0.94	0.92	0.92	0.94	0.93	0.93	0.92	0.93
	RMSE	25.26	17.81	21.04	27.45	19.12	19.62	19.74	19.45	18.28
	accuracy	72.73	83.64	81.82	65.45	78.18	80	74.54	85.45	85.45
24 hour	**r**	0.94	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97
	r^2	0.88	0.94	0.94	0.94	0.96	0.95	0.95	0.95	0.95
	RMSE	32.22	23.19	24.62	19.95	16.99	20.51	20.75	20.15	19.79
	accuracy	58.18	63.64	61.82	74.55	72.73	67.27	67.27	65.45	70.9
48 hour	**r**	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.98	0.98
	r^2	0.98	0.98	0.96	0.96	0.98	0.97	0.97	0.97	0.97
	RMSE	12.38	10.85	11.58	12.51	10.84	10.69	10.74	10.6	10.55
	accuracy	81.82	81.82	89.09	83.64	80	85.45	85.45	85.45	85.45

Table 4.11: Comparison between four ensemble approaches for different seasonal days

	20^{th} **January, 2016**				10^{th} **April, 2016**				18^{th} **June, 2016**				26^{th} **November, 2016**			
Hours	**AM**	**HM**	**QM**	**Median**	**AM**	**HM**	**QM**	**Median**	**AM**	**HM**	**QM**	**Median**	**AM**	**HM**	**QM**	**Median**
1 hour	0.98	0.98	0.98	0.98	0.92	0.92	0.92	0.92	0.96	0.96	0.96	0.96	0.99	0.99	0.99	0.99
	0.97	0.97	0.97	0.97	0.84	0.84	0.85	0.85	0.92	0.92	0.92	0.92	0.98	0.98	0.98	0.98
	10.9	11.03	10.65	10.25	47.3	47.4	47.14	46.75	45.88	46.05	47.66	45.27	8.42	8.19	7.44	8.49
	85.45	85.45	85.45	85.45	70.9	70.9	70.9	69.09	69.09	69.09	70.9	69.09	92.72	89.09	92.72	92.72
6 hours	0.98	0.98	0.98	0.98	0.89	0.89	0.89	0.89	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	0.97	0.97	0.97	0.96	0.8	80	0.8	0.8	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.94
	12.4	12.5	12.17	13.17	58.89	59.02	58.62	59.07	58.89	59.11	59.89	64.31	13.52	13.56	13.47	13.44
	89.09	89.09	89.09	85.45	70.9	69.09	70.9	69.09	54.54	54.54	54.54	43.63	81.81	81.81	81.81	83.63
12 hours	0.98	0.98	0.98	0.98	0.95	0.95	0.95	0.95	0.98	0.98	0.98	0.99	0.91	0.91	0.91	0.91
	0.97	0.97	0.97	0.97	0.91	0.91	0.91	0.91	0.97	0.97	0.97	0.98	0.83	0.83	0.83	0.83
	14.39	14.46	14.35	15.02	46.13	45.83	46.9	44.3	31.74	32.3	26.65	29.56	6.59	29.33	29.58	28.88
	78.18	78.18	78.18	76.36	65.45	69.09	67.27	70.9	76.36	72.72	78.18	78.18	76.36	78.18	76.36	76.36
18 hours	0.99	0.99	0.99	0.99	0.97	0.97	0.96	0.97	0.97	0.98	0.97	0.98	0.96	0.96	0.96	0.96
	0.98	0.98	0.98	0.98	0.94	0.94	0.93	0.94	0.95	0.96	0.95	0.96	0.93	0.93	0.92	0.93
	10.78	10.85	10.67	10.98	44.21	43.69	46.96	40.67	44.12	43.39	47.7	40.87	19.62	19.74	19.45	18.28
	85.45	85.45	85.45	83.63	65.45	65.45	65.45	70.9	60	60	61.81	69.09	80	74.54	85.45	85.45
24 hours	0.99	0.99	0.99	0.99	0.92	0.91	0.92	0.91	0.96	0.96	0.86	0.97	0.97	0.97	0.97	0.97
	0.99	0.99	0.99	0.99	0.84	0.84	0.84	0.83	0.93	0.93	0.74	0.94	0.95	0.95	0.95	0.95
	8.89	8.89	8.95	8.5	58.36	58.65	57.73	61.19	40.84	40.62	58.36	42.01	20.51	20.75	20.15	19.79
	89.09	89.09	89.09	89.09	67.27	67.27	69.09	65.45	74.54	74.54	72.72	78.18	67.27	67.27	65.45	70.9
48 hours	0.99	0.99	0.99	0.99	0.94	0.94	0.94	0.94	0.88	0.88	0.9	0.89	0.98	0.98	0.98	0.98
	0.98	0.98	0.98	0.99	0.89	0.89	0.89	0.89	0.79	0.78	0.81	0.79	0.97	0.97	0.97	0.97
	6.67	6.7	6.63	6.3	39.77	40.13	38.91	39.72	34.29	35.22	34.82	33.84	10.69	10.74	10.6	10.55
	90.9	90.9	90.9	90.9	70.9	70.9	74.54	72.72	85.45	85.45	83.63	83.63	85.45	85.45	85.45	85.45

For 6 hours ahead prediction, QM gains 89.09% prediction accuracy (Figure 4.22(a)) with 0.97 r^2 (Figure 4.21(a) and Figure 4.20(b)) and 12.17 RMSE (Figure 4.22(b)). For 12 hours ahead prediction QM gives an accuracy of 78.18% (Figure 4.22(a)) with 0.97 r^2 (Figure 4.21(a) and Figure 4.20(c)) and 14.35 RMSE (Figure 4.22(b)). For 18 hours ahead prediction, QM offers 85.45% of prediction accuracy (Figure 4.22(a)) with 0.98 r^2 (Figure 4.21(a) and Figure 4.20(d)) and 10.67 RMSE (Figure 4.22(b)). For 24 hours ahead prediction, median achieves 89.09% prediction accuracy (Figure 4.22(a)) with 0.99 r^2 (Figure 4.21(a) and Figure 4.20(e)) and 8.50 RMSE (Figure 4.22(b)). For 48 hours ahead prediction, median (90.90%) achieves high prediction accuracy (Figure 4.22(a)) with 0.99 r^2 (Figure 4.21(a) and Figure 4.20(e)) and 6.3 RMSE (Figure 4.22(b)). From the results obtained for 20^{th} January, 2016, it is observed that QM and median approaches offer high prediction accuracy (78% to 90%) in all forecasting horizons.

For 10^{th} April (table 4.8) in 1 hour ahead prediction, QM achieves 70.90% prediction accuracy (Figure 4.24(a)) with 0.85 r^2 (Figure 4.21(b) and Figure 4.23(a)) and 47.14 RMSE (Figure 4.24(b)). For 6 hours ahead prediction, QM achieves highest prediction accuracy (70.90%) (Figure 4.24(a)) and 0.80 r^2 (Figure 4.21(b) and Figure 4.23(b)) and 58.62 RMSE (Figure 4.24(b)). For 12 hours ahead prediction, median achieves 70.90% prediction accuracy (Figure 4.24(a)) with 0.91 r^2 (Figure 4.21(b) and Figure 4.23(c)) and 44.30 RMSE (Figure 4.24(b)). For 18 hours ahead prediction, median achieves 70.90% prediction accuracy (Figure 4.24(a)) and 0.94 r^2 (Figure 4.21(b) and Figure 4.23(d)) and 44.67 RMSE (Figure 4.24(b)). For 24 hours ahead prediction, QM achieves highest prediction accuracy (69.09%) (Figure 4.24(a)) with 0.84 r^2 (Figure 4.21(b) and Figure 4.23(e)) and 54.73 RMSE (Figure 4.24(b)). For 48 hours prediction, QM achieves highest prediction accuracy (74.54%) (Figure 4.24(a)) with 0.89 r^2 (Figure 4.21(b) and Figure 4.23(f)) and 38.91 RMSE (Figure 4.24(b)).

Results validate the effectiveness of QM approach for 10^{th} April, 2016 also. In all test cases, high and stable prediction accuracy (70% to 74%) is gained by QM ensemble approach.

For 18^{th} June (table 4.9) in 1 hour ahead prediction, highest prediction accuracy (70.90%) (Figure 4.26(a)) is achieved by QM method with 0.92 r^2 (Figure 4.21(c) and Figure 4.25(a)) and 47.66 RMSE (Figure 4.26(b)). For 6 hours ahead prediction, QM achieves highest prediction accuracy (54.54%) (Figure 4.26(a)) and gains 0.95 r^2 (Figure 4.21(c) and Figure 4.25(b)) and 59.89 RMSE (Figure 4.26(b)). For 12 hours ahead prediction, 78.18% prediction accuracy (Figure 4.26(a)) is achieved by QM with 0.97 r^2 (Figure 4.21(c) and Figure 4.25(c)) and 26.65 RMSE (Figure 4.26(b)). For 18 hours ahead prediction, highest prediction accuracy (69.09%) (Figure 4.26(a)) has been gained by median

with 0.96 r^2 (Figure 4.21(c) and Figure 4.25(d)) and 40.87 RMSE (Fig. 13(b)). For 24 hours ahead prediction, median achieves prediction accuracy of the order of 78.18% (Figure 4.26(a)) with 0.94 r^2 (Figure 4.21(c) and Figure 4.25(e)) and 42.01 RMSE (Figure 4.26(b)). For 48 hours ahead prediction accuracy, QM gains highest prediction accuracy (83.63%) (Figure 4.26(a)) with 0.81 r^2 (Figure 4.21(c) and Figure 4.25(f)) and 34.82 RMSE (Figure 4.26(b)).

For 26^{th} November (table 4.10) in 1 hour ahead prediction, QM achieves highest prediction accuracy (92.72%) (Figure 4.28(a)) with 0.98 r^2 (Figure 4.21(d) and Figure 4.27(a)) and 7.44 RMSE (Figure 4.28(b)). For 6 hours ahead prediction, 83.63% prediction accuracy (Figure 4.28(a)) is gained by median with 0.94 r^2 (Figure 4.21(d) and Figure 4.27(b)) and 13.44 RMSE (Figure 4.28(b)). For 12 hours ahead prediction, maximum accuracy (78.18%) is offered by HM (Figure 4.28(a)) with 0.83 r^2 (Figure 4.21(d) and Figure 4.27(c)) and 29.33 RMSE (Figure 4.28(b)). For 18 hours ahead prediction, median achieves 85.45% prediction accuracy (Figure 4.28(a)) with 0.93 r^2 (Figure 4.21(d) and Figure 4.27(d)) and 18.28 RMSE (Figure 4.28(b)). For 24 hours ahead prediction maximum accuracy (70.90%) (Figure 4.28(a)) is offered by median with 0.95 r^2 (Figure 4.21(d) and Figure 4.27(e)) and 19.89 RMSE (Figure 4.28(b)). For 48 hours ahead prediction, median achieves 85.45% prediction accuracy (Figure 4.28(a)) with 0.97 r^2 (Figure 4.21(d) and Figure 4.27(f)) and 10.55 RMSE (Figure 4.28(b)).

It is validated from the simulation results that rather than being uncertain about the performance of a particular machine learning model in different test conditions, ensemble approach overwhelm the uncertainties associated with individual models by offering a single optimized output by combining the output of independent models. As observed from the results, QM, and median ensemble approach perform equally good in different seasons and for different forecasting horizons. A comparison summary for different ensemble approaches has been tabulated in table 4.11.

4.5.5 Concluding Remarks

Machine learning based statistical ensemble approach has been proposed for two days ahead solar irradiance forecasting with hourly resolution. Simulations is performed using *R* interface and conducted on real time solar data from NREL. Optimal characteristics of accuracy and stability have been obtained by aggregating base machine learning models using generalized mean approach (arithmetic mean, harmonic mean, quadratic mean and median). Statistical comparison between different models and proposed approach is carried out with respect to seasonal variation and for forecasting horizons ranging from

1 hour ahead to 48 hours ahead. From the simulation results, it is seen that statistical ensembling achieve relatively high and stable prediction accuracy and hence overcome the shortcomings of independent machine learning models. Quadratic mean and median approaches achieves high performance matrix in terms of prediction accuracy, correlation coefficient RMSE and stability.

Regarding the results obtained for 20^{th} January, 2016 (winter), highest performance indices is achieved by QM and median approach ranging from 78.18% to 90.90% for 1 hour to 48 hours ahead forecasting. The results obtained for 10^{th} April (spring), 2016, shows an accuracy of 69.09% to 74.54% gained by QM and median approach for all forecasting horizons. An accuracy of the order of 69.09% to 78.18% is accounted by QM and median method for 18^{th} June, 2016 (summer). Similar results have been observed for 26^{th} November (autumn), 2016 where 70.90% to 92.72% prediction accuracy is achieved by QM and median approach up to 48 hours ahead prediction.

The results are evident that solar irradiance forecasting with such machine learning models is recent and productive study in this field leads to accurate solar forecasting than conventional methods.

Chapter 5

Adaptive Duty Cycle Algorithm for Energy Harvesting Wireless Sensor Networks

Designing of low power electronic devices and advancement in miniaturization leads to deployment of WSNs for long term observations of different processes at large spatial and temporal distribution. Geographical and biological process monitoring, industrial infrastructure monitoring and failure warning, health monitoring, and wild life surveillance monitoring are some exemplary deployments of WSNs. These application scenarios are designed to meet different objectives. The systems deployed to target their respective goals exhibit several requirements including reliable operation of the node over long time periods in different working conditions; cheap manufacturing of nodes because of the requirement of large number of nodes and their nonrecoverable nature; light and small design of nodes to reduce the installation cost and interference in the environment.

In the inaccessible and remote location deployments where continuous power supply is not possible, WSN nodes are powered by finite capacity batteries. This feature put limits on system life time and performance level of the node specially in high energy demanding scenarios such as Global Positioning System (GPS) receivers. Since node life time is proportional to the ratio of energy generation and consumption profiles, node has been equipped with high capacity batteries to increase the life span. The increased capacity batteries tends to the increase size and cost of the node. Wireless rechargeable sensor node addresses this issue and draw all or a part of required energy from the environment. Though ambient source of energy is infinite and allow sustainable node operation over long time periods, but there is a limit on the rate at which the ambient energy can be utilized. Thus, designing and operation of energy harvested wireless sensor node faces challenges in terms of both hardware and software. To meet the expected life time requirement, real time power management between harvesting source and load is primary requirement with the following two primary considerations:

- Energy neutral mode of operation: To ensure that energy consumption is always less than or equal to available energy over a given time period, the node should

The proposed work has been published in **Wireless Networks, pp. 1-29, January 2019**.

operate in energy neutral mode. This condition is necessary for a system to achieve uninterrupted operation over long time periods.

- Maximize Network Performance: The power management unit should make efficient utilization of available energy to maximize network performance. To achieve this, scheduling the energy should be such that energy will not waste due to overflow and no node depletion due to underflow.

5.1 Introduction

High and stable duty cycle is an essential feature for sustainable sensor nodes though the power consumption varies vastly in microwatts during standby and milli watts during active sessions. Further, uncertain solar irradiance together with the substantial decline of the energy storage devices are the inclusive factors of duty cycle variation. Through consideration of all these constraints, an adaptive duty cycle algorithm has been developed aiming to achieve improved average duty cycle with high stability. The proposed work focuses on improving average duty cycle with high stability throughout the day regardless of prediction horizon and irrespective of solar irradiance variation.

Figure 5.1 depicts the methodology adapted for optimized operation of energy harvested wireless sensor networks. For efficient power management, availability of solar irradiance is pre-estimated. Initially, the node duty cycle is dynamically adjusted with respect to estimation of available energy and then subject to real time changes according to measured solar irradiance profile. The proposed work is a collaboration of budget assigning principles with adaptive duty cycle algorithm. The work is segregated in two parts, a) Pre-estimation of node duty cycle using predicted solar irradiance (W/m^2), and b) Real time adaptivity for pre-computed node duty cycle. To verify the effectiveness of proposed algorithm, the results are compared with kansal *et al.*[58] for short, medium and long term prediction horizons. The main contribution of proposed work in the research area of energy harvested sensor node is as follows:

1. **Uninterrupted node operation**: The future estimation of ambient energy leads to precompute the node duty cycle and related performance parameters in advance to prevent the condition of energy depletion.

2. **Real time adaptivity**: Precomputed value of duty cycle gets adjusted relevantly in case of increase or decrease of actual energy level from its predicted value.

3. **Low variance**: The proposed work offers minimum variation in node duty cycle in

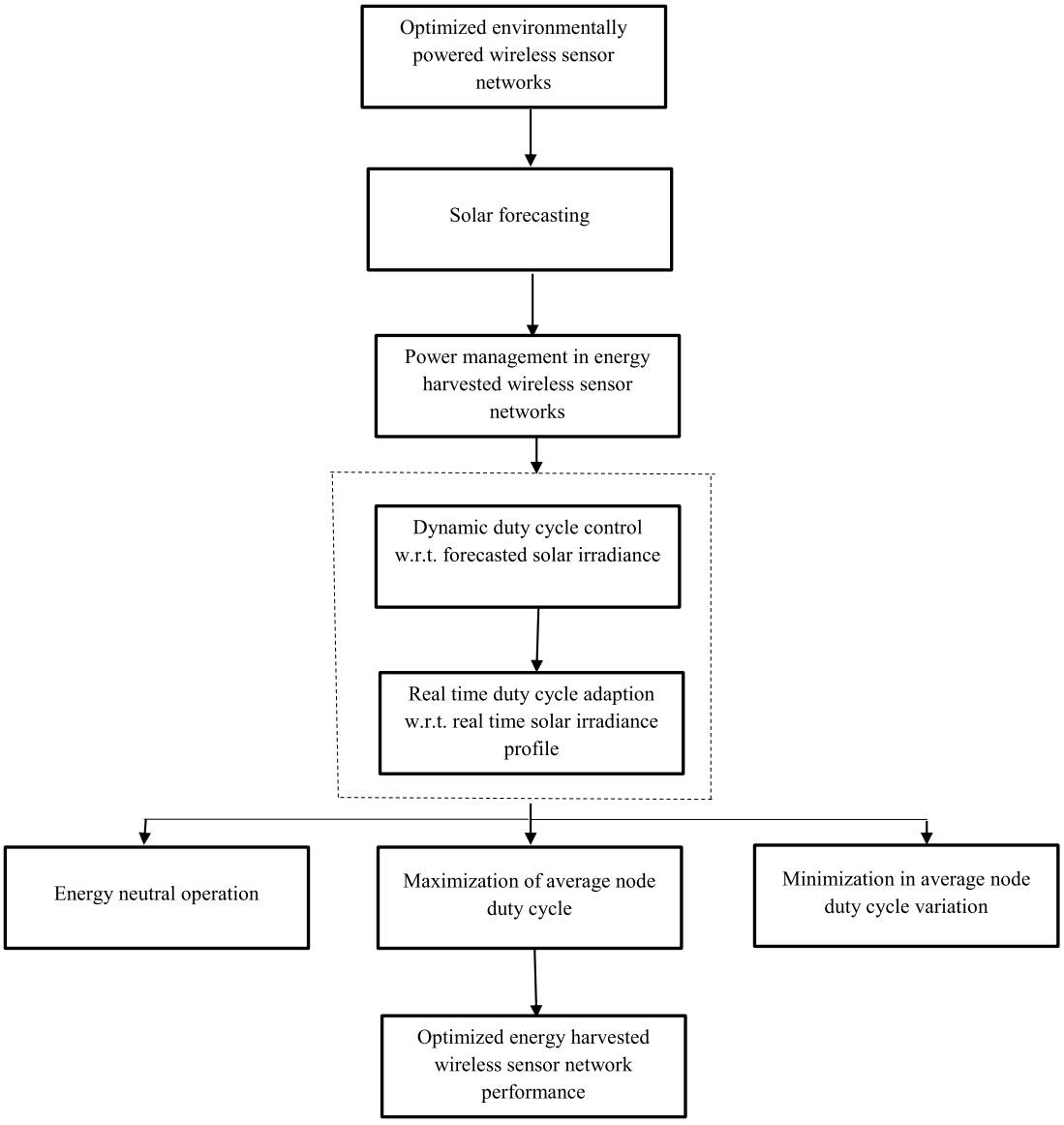


Figure 5.1: Methodology for optimized energy harvested wireless sensor networks

case of unstable environmental conditions and undesired hardware degradations.

4. **Low computational overhead**: The linear programming based intense calculations are replaced by simple mathematical programming which offers low computational complexity.

5.1.1 System Model

Figure 5.2 depicts the system model for proposed work. The machine learning based Cubist model[145] is used to predict the solar irradiance availability in future. The historical data sets are maintained to facilitate the Cubist algorithm and taken from

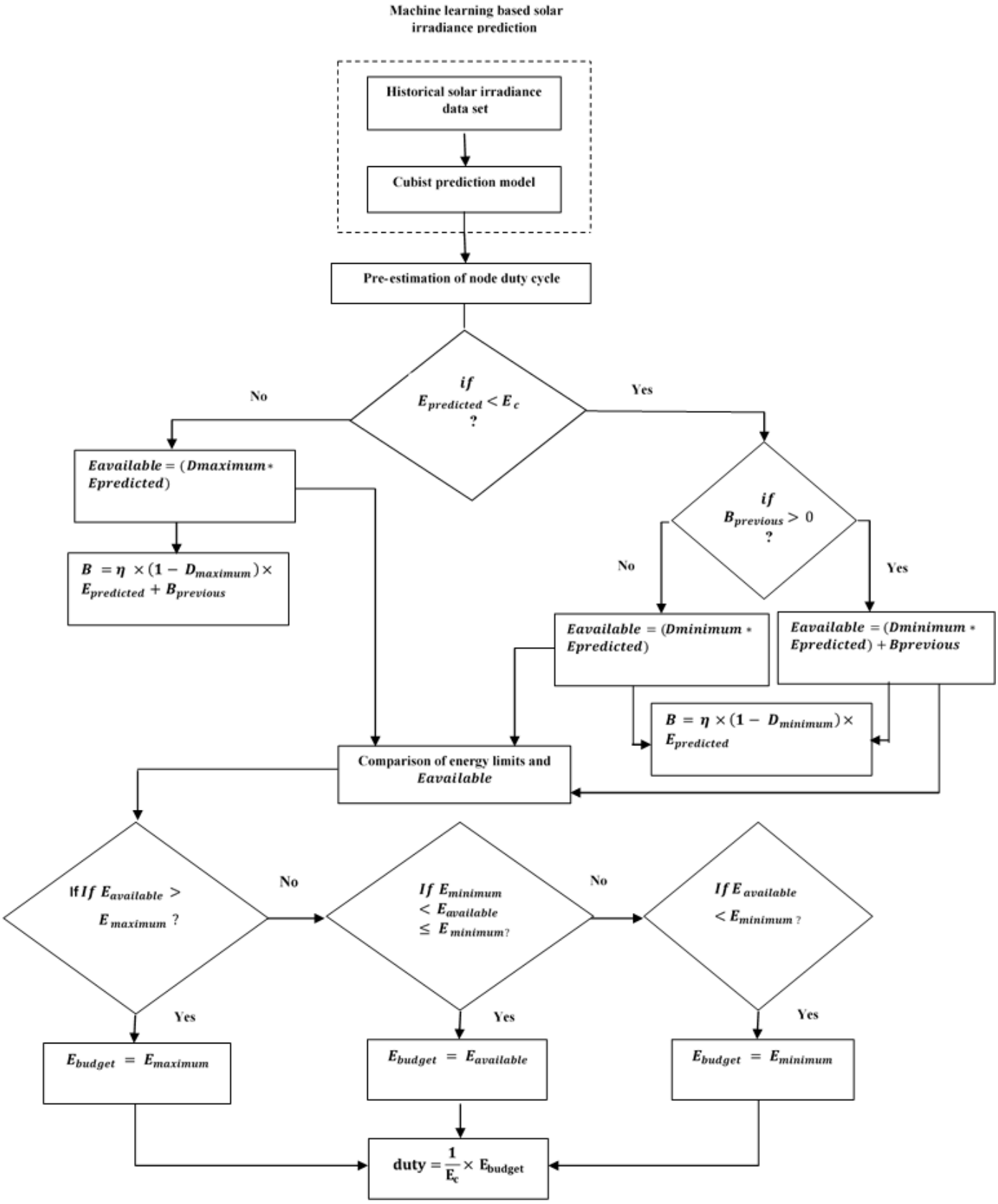


Figure 5.2: System model

the Solar Radiation Research Laboratory (SRRL) of the National Renewable Energy Laboratory (NREL). The data set obtained from NREL is per 30 minutes global solar irradiance measured on horizontal plane. The predicted profiles are used to calculate energy budgets[284] for energy allocation to nodes with respect to the estimated generation rates. The aim is to achieve energy neutral state of operation in situations of bulk ambient energy availability or in energy depletion states while considering the energy storage element. To reduce the effect of battery inefficiencies, the harvest-use-store approach is adopted for the proposed algorithm. In this approach, node is directly powered by harvested energy through power management unit. Only excess energy will be kept in energy storage element[33]. Due to the different meteorological and geographical factors, real-time energy profile may be different from estimated ones. In response to the real-time solar irradiance profiles, the estimated duty cycle is adaptively updated to maintain a balance between energy generation and consumption rates with respect to energy available in actual rather than predicted ones.

5.1.2 Characterization of Harvesting System with Different Storage Profiles

If $E_{available}(t)$ is the available harvested energy and $E_{consumed}(t)$ is the energy consumed by a load such as a sensor node at time t, the harvesting system can be categorized in three different models:

- **Harvesting system without storage**: The first model is a harvesting system without energy storage. The harvested energy from the source is directly utilized by load. In such harvesting systems, the operating condition is:

$$E_{available}(t) \geq E_{consumed}(t) \qquad \forall t \tag{5.1}$$

  Any surplus energy ($E_{available}(t)$) -$E_{consumed}(t)$) and energy received during $E_{available}(t) < E_{consumed}(t)$ is wasted.

- **Harvesting system with ideal energy storage**: An ideal energy storage refers to a device with no leakage of charge, no inefficiency in charging, and large energy storage capacity. This stored energy support the node operation in less energy availability. In such harvesting systems, the operating condition is:

$$\int_0^T E_{available}(t)dt + B_0 \geq \int_0^T E_{consumed}(t)dt \qquad \forall T(0,\infty) \tag{5.2}$$

  where B_0 is initial battery level of ideal storage unit.

- **Harvesting system with practical energy storage conditions**: A practical energy storage refers to a device with energy losses through leakage, limited charging efficiency, $\eta < 1$, and limited energy storage capacity. To define the operating condition in such harvesting systems, a rectifier function is introduced and given as:

$$[x]^+ = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{5.3}$$

Following the equation 5.3, energy conservation in sensor node leads to:

$$\begin{aligned} B_0 = \eta \int_0^T [E_{available}(t) - E_{consumed}(t)]^+ dt - \int_0^T [E_{consumed}(t) \\ - E_{available}(t)]^+ dt - \int_0^T E_{leak}(t) dt \geq 0 \qquad \forall T(0, \infty) \end{aligned} \tag{5.4}$$

where E_{leak}(t) is the leakage energy from the storage unit. Capacity of the storage unit is calculated by using the equation 5.4 and given as:

$$\begin{aligned} B_0 = \eta \int_0^T [E_{available}(t) - E_{consumed}(t)]^+ dt - \int_0^T [E_{consumed}(t) - \\ E_{available}(t)]^+ dt - \int_0^T E_{leak}(t) dt \leq B \qquad \forall T(0, \infty) \end{aligned} \tag{5.5}$$

where B is the capacity of battery. For a harvesting system with practical energy storage conditions, effective power management techniques is required. For this, an analytical model is developed to analyse the variability and energy generation rate of the harvesting source and consumption profile of the load.

5.1.3 Modelling of Practical Energy Harvesting System

If ϱ is defined as average rate at which energy is available and σ_1 and σ_2 defines the burstiness in the signal, the function $(\varrho, \sigma_1, \sigma_2)$ represents a continuous, positive, and bounded function $E_{available}(t)$. The $E_{available}(t)$ is considered as energy source if the following conditions are satisfied between time interval τ and T:

$$\int_\tau^{\tau+T} E_{available}(t) dt \leq \varrho T + \sigma_1 \tag{5.6}$$

$$\int_\tau^{\tau+T} E_{available}(t) dt \geq \varrho T - \sigma_2 \tag{5.7}$$

For load $E_{consumed}(t)$, the function is defined in terms of $(\varrho_2, \sigma_3, \sigma_4)$. The equation 5.4 is

re-written in term of $E_{available}(t)$ and $E_{consumed}(t)$ as follows:

$$B_0 + \eta.\min\{\int_T E_{available}(t)dt\} - \max\{\int_T E_{consumed}(t)dt\} - \int_T E_{leak}(t)dt \geq 0 \quad (5.8)$$

$$\Rightarrow B_0 + \eta(\varrho_1 T - \sigma_2) - (\varrho_2 T + \sigma_3) - \varrho_{leak} T \geq 0 \quad (5.9)$$

5.1.4 Physical Conditions for Energy Neutrality

- The initial energy level required in the battery is obtained by considering T=0 in equation 5.9 and given as:

$$B_0 \geq \eta\sigma_2 + \sigma_3 \quad (5.10)$$

- By considering T as ∞ in equation 5.9 yields :

$$\eta\varrho_1 - \varrho_{leak} \geq \varrho_2 \quad (5.11)$$

  By substituting equation 5.10 and 5.11 in 5.5, the resultant equation is:

$$\Rightarrow B_0 + \eta(\varrho_1 T + \sigma_1) - (\varrho_2 T - \sigma_4) - \varrho_{leak} T \leq B \quad (5.12)$$

- By substituting $T = 0$ in equation 5.12, we obtain required battery size, given as:

$$B_0 + (\eta\sigma_1 - \sigma_4) \leq B \quad (5.13)$$

  and using the equation 5.10

$$B \geq \eta(\sigma_1 + \sigma_2) + \sigma_3 - \sigma_4 \quad (5.14)$$

  Also, by taking $T = \infty$, the equation 5.12 yields:

$$\eta\varrho_1 - \varrho_{leak} \leq \varrho_2 \quad (5.15)$$

  The equation 5.14 defines the maximum battery size required to cover up the burstiness of energy generation and consumption rates. The limiting case $T = \infty$ signifies the long term sustainable behaviour without bursts.

- Energy Neutral Operation:

  By considering a practical harvesting system where energy source is defined by the function $(\varrho, \sigma_1, \sigma_2)$, load is defined by (ϱ_2, σ_3) and energy storage element is defined

by storage efficiency η and leakage ϱ_{leak}, the following conditions are necessary for energy neutral operation:

$$\varrho_2 \leq \eta\varrho_1 - \varrho leak \tag{5.16}$$

$$B_0 \geq \eta\sigma_2 + \sigma_3 \tag{5.17}$$

$$B \geq B_0 \tag{5.18}$$

where B denotes the capacity of battery and B_0 denotes the initial charge in the battery. This parameter helps in selecting a particular storage technology to satisfy the capacity requirements.

- Relation Between Battery Size and Node Life Time:

  If B is the size of the battery, ϱ_1 is the harvested energy and ϱ_2 is the load consumption, the achieved life time, L_T is calculated as:

$$L_T = \frac{B}{\varrho_2 - \eta\varrho_1} \tag{5.19}$$

It is clear from the equation 5.19 that with a finite source of energy, a node cannot continue energy neutral mode of operation and a large battery size is required to survive for longer time periods. Energy harvesting based systems are appropriate alternatives if these are feasible and cost effective for a particular application.

5.2 Proposed Power Management Approach

An algorithm is proposed to achieve **Energy neutral operation** for a predictable and non-controllable energy source (solar). The approach is to characterize the nature of harvesting source and then adapt the performance parameters accordingly. **Adaptive duty cycling approach** is adopted for performance scaling of the node by controlling the energy consumption rate, ϱ_2. The following parameter are considered to estimate node duty cycle for N (i=1, ..., N) time slots with duration ΔT:

- $E_{predicted}(i)$ is the estimated energy availability in time slot i from the harvesting source. This parameter is assumed constant over a slot duration.
- E_c is the energy consumption by the node in active state. The sleep mode energy consumption is negligible.
- $Duty(i)$ is the node duty cycle.

- $B(i)$ is the battery back up at the start of slot i and depend upon duty cycle $Duty(i)$. The residual battery at the end of last time slot is considered as $B(N+1)$.

The relation between energy consumption rate and node duty cycle is given as:

$$\varrho_2 = Duty * E_c \tag{5.20}$$

Thus, by controlling the node duty cycle, energy consumption rate of the node is controlled. The focus of the proposed work is to maintain satisfactory performance level of the system by maximizing the average node duty cycle with minimum variations.

In a time slot t, there are two possible conditions regarding the availability of harvesting energy $E_{predicted}$, it may be either higher or lower than the energy consumption E_c. When $E_{predicted}$ is higher than E_c, the node is directly powered by the source and rest of the energy is supplied to the battery. otherwise, the required energy is borrowed from the battery. By using the rectifier function defined in equation 5.3, energy utilized from the battery at the end of each time slot is computed as :

$$\begin{aligned} B(i) - B(i+1) &= \Delta T.Duty(i)[E_c - E_{predicted}(i)]^+ - \eta \Delta T E_{predicted}(i)\{1 - Duty(i)\} \\ &- \eta \Delta T.Duty(i)[E_{predicted}(i) - E_c]^+ \quad \forall i(1,..,N) \end{aligned} \tag{5.21}$$

The first part of equation on right hand side shows the energy taken from battery when $E_{predicted}<E_c$, second term denotes the stored energy in battery during sleep state and last term shows the stored energy in active state when $E_{predicted}>E_c$. For the attainment of energy neutral state, the battery back up at the end of each time slot should be equal to or greater than initial energy level of the battery. Hence, the selection of time slots N is important so that even in cloudy days, battery will not be energy depleted. These considerations leads to optimal duty cycle conditions and are stated below:

$$\sum_{i=1}^{N} Duty(i) \tag{5.22}$$

$$B(N+1) \geq B_0 \tag{5.23}$$

$$Duty(i) \geq D_{min} \qquad \forall i \epsilon (1,...,N) \tag{5.24}$$

$$Duty(i) \leq D_{max} \qquad \forall i \epsilon (1,...,N) \tag{5.25}$$

For the solution of above optimization problems, harvesting aware power management approach is carried out in three steps. In the first step, past energy profiles are tracked and use to estimate future energy availability. In the second step, node duty cycle is pre-estimated based on the predicted solar energy profiles. Dynamic programming approach is used to achieve the same. In the third step, dynamic adaption of node duty cycle is carried out with respect to real time solar energy profiles as energy neutral condition should be sustained with respect to real energy profiles rather than predicted values.

5.2.1 Solar Irradiance Prediction

In third and fourth chapter of the thesis, solar irradiance forecasting methods for short and long term forecasting were proposed. To follow the first step of proposed power management approach (Section 5.2), *Cubist*, machine learning model is used for the estimation of future energy availability to power up the sensor node. Figure 5.3 depicts energy profiles of January 1, January 4, and January 7, 2016. It shows the variations in solar irradiance at different time slots of a day. It is also seen that January 1 and January 7, 2016 have similar energy profiles but January 7, 2016 have low solar irradiance availability through out the day. Solar irradiance prediction is carried out during the sun hours (7:30am to 16:30pm). The total time period is divided into fixed length time slots of duration 30 minutes. Figure 5.4 represents predicted energy profiles of January 1, 2016 for 60, 90, 120, and 150 minutes ahead prediction horizons. Results shows that a close approximation of real time energy values is obtained by the prediction model with high correlation coefficients as shown in Figure 5.5.

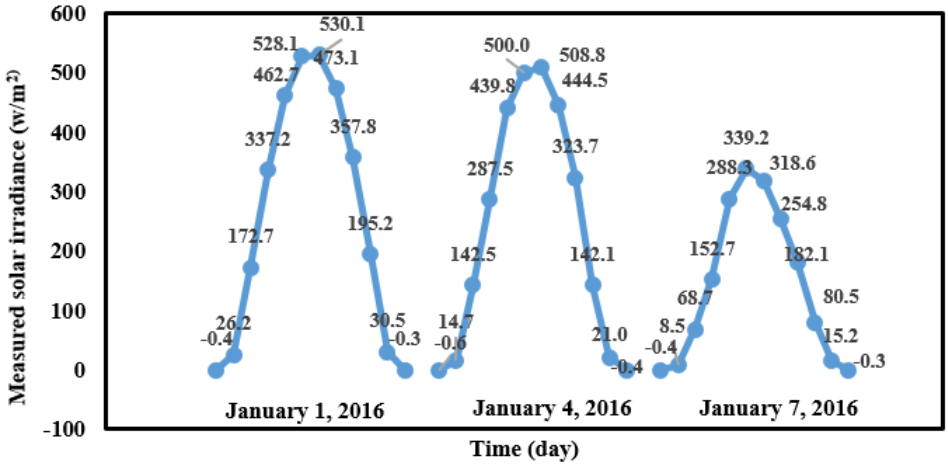
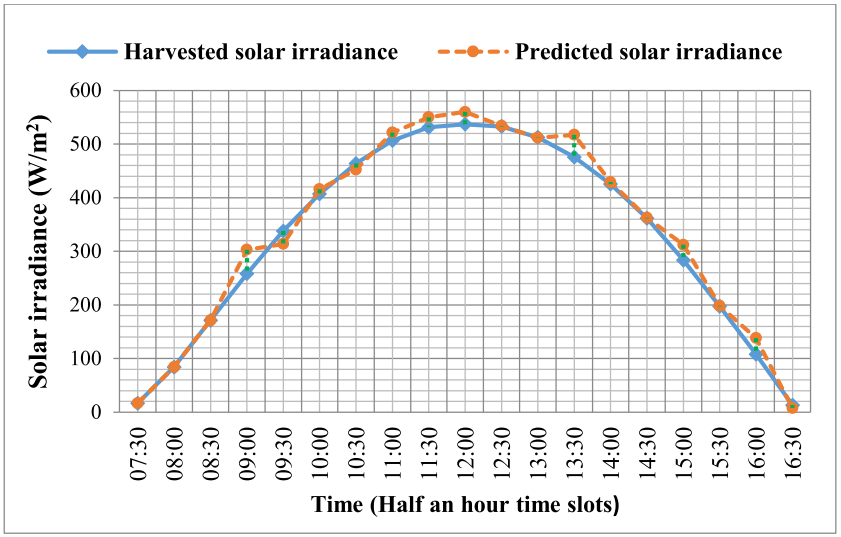
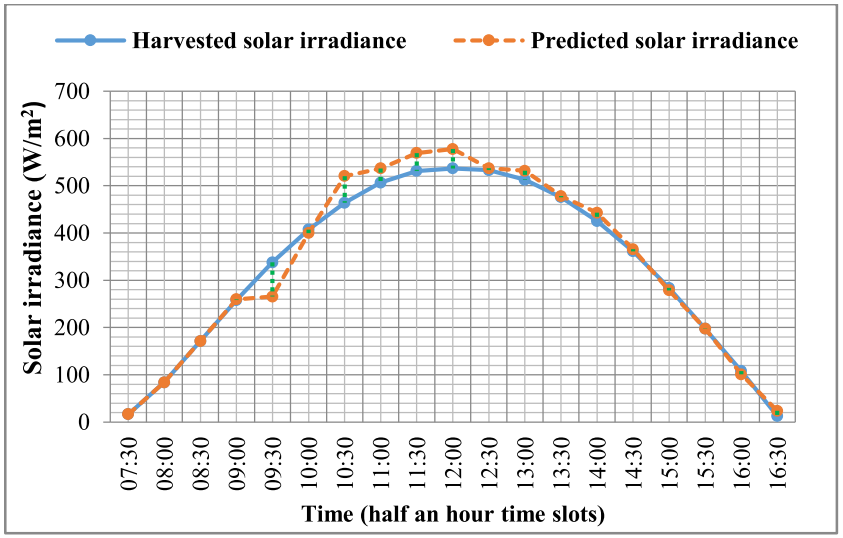


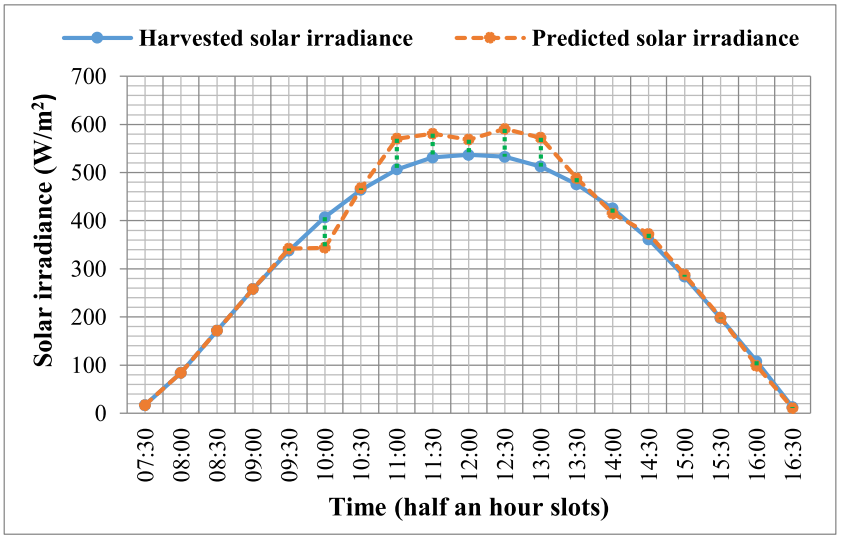
Figure 5.3: Solar irradiance profile of January 1, January 4 and January 7, 2016



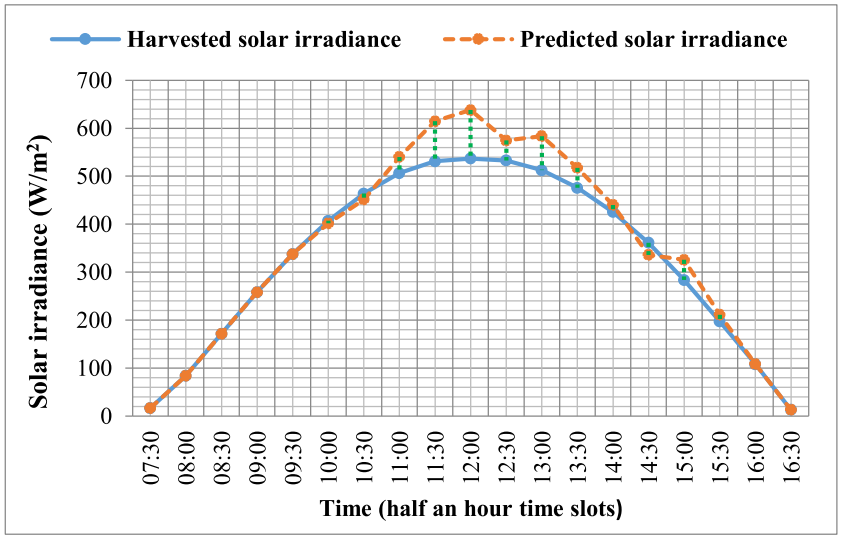
(a) 60 minutes ahead prediction for January 1, 2016.



(b) 90 minutes ahead prediction for January 1, 2016.

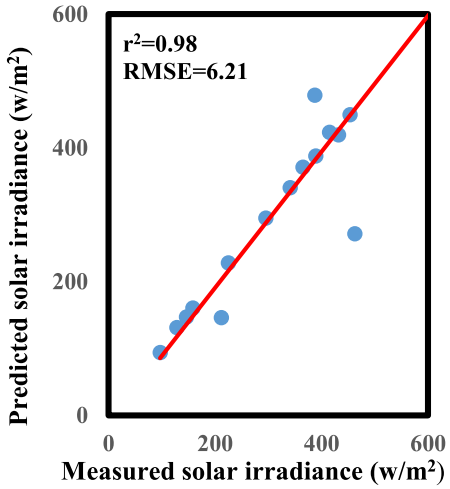


(c) 120 minutes ahead prediction for January 1, 2016.

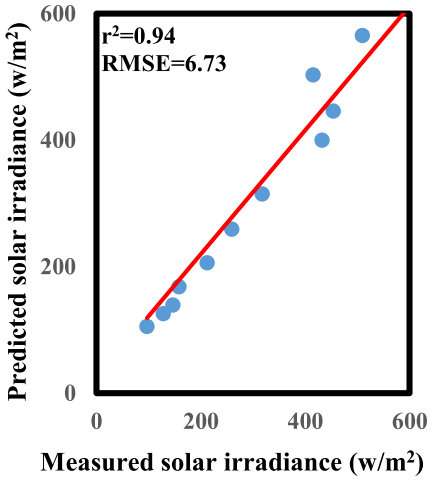


(d) 150 minutes ahead prediction for January 1, 2016.

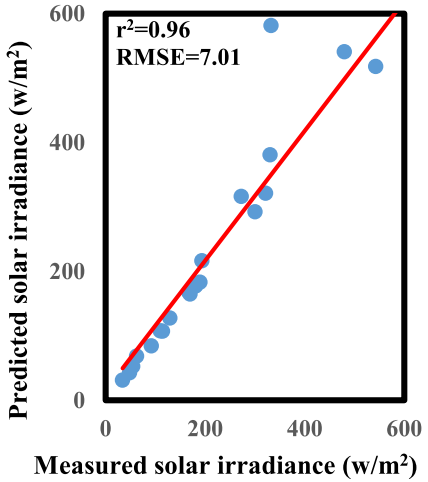
Figure 5.4: Solar irradiance prediction for January 1, 2016.



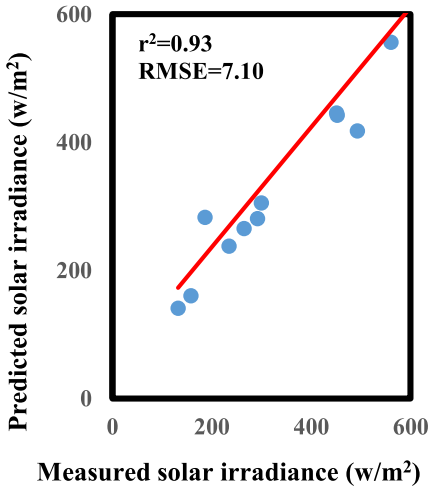
(a) 60 minutes ahead solar irradiance prediction for January 1, 2016.



(b) 90 minutes ahead solar irradiance prediction for January 1, 2016.

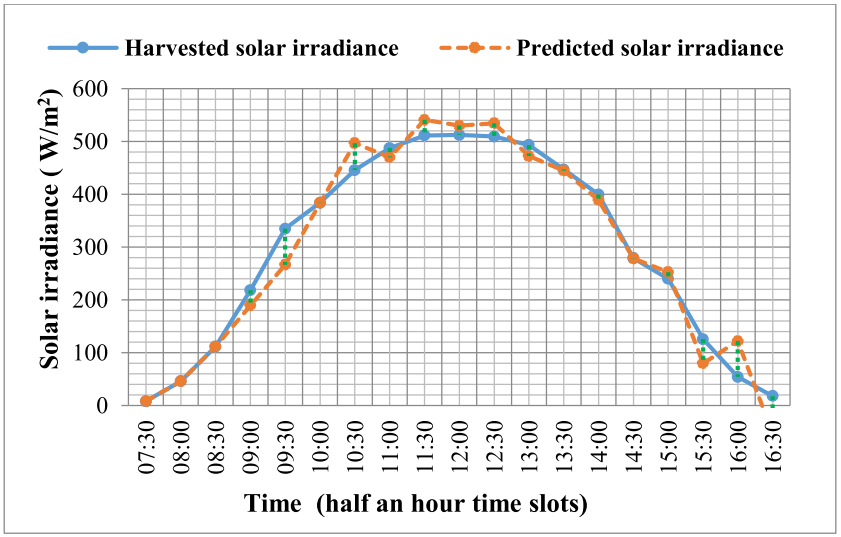


(c) 120 minutes ahead solar irradiance prediction for January 1, 2016.

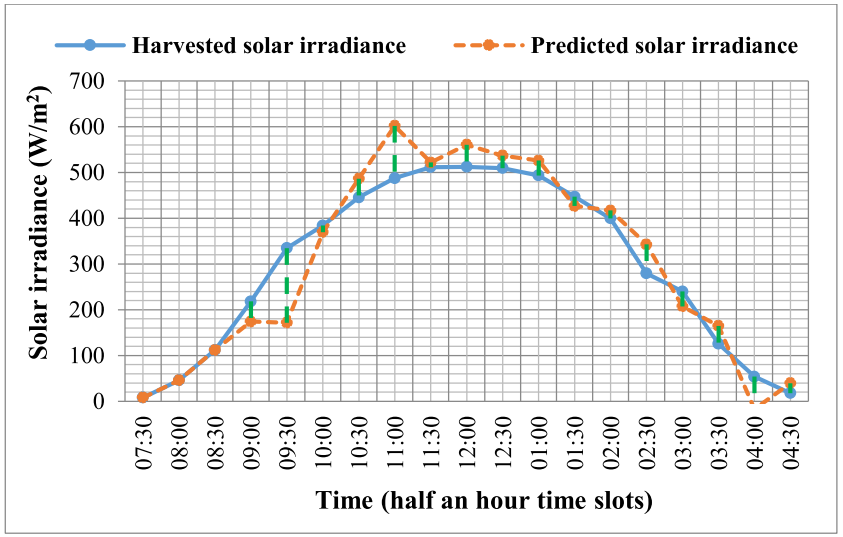


(d) 150 minutes ahead solar irradiance prediction for January 1, 2016.

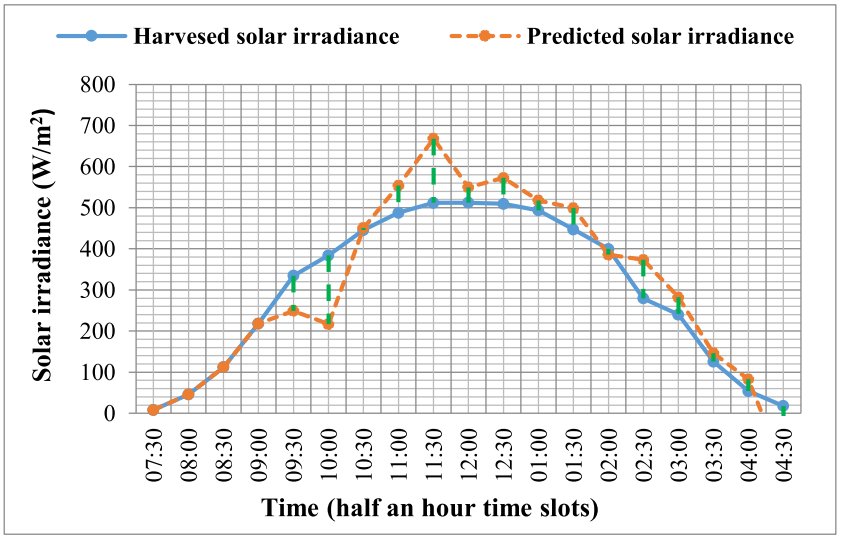
Figure 5.5: Correlation coefficient for January 1, 2016.



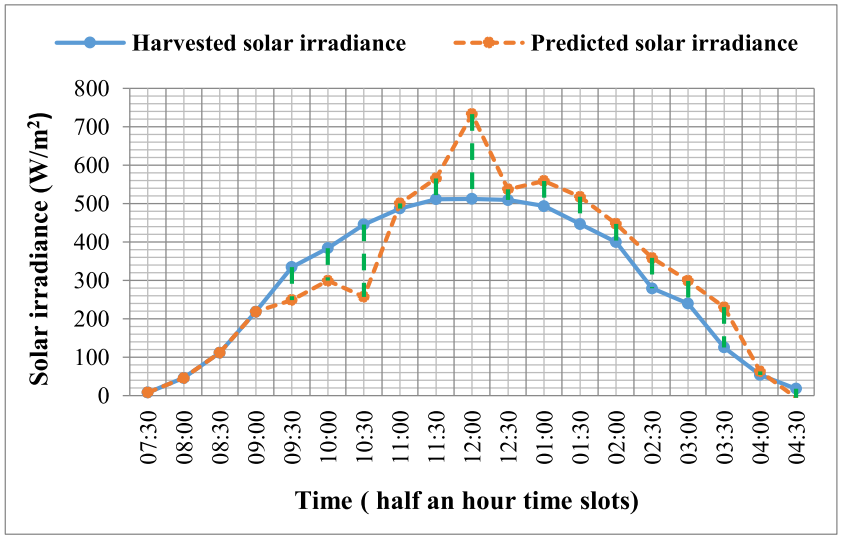
(a) 60 minutes ahead prediction for January 4, 2016.



(b) 90 minutes ahead prediction for January 4, 2016.

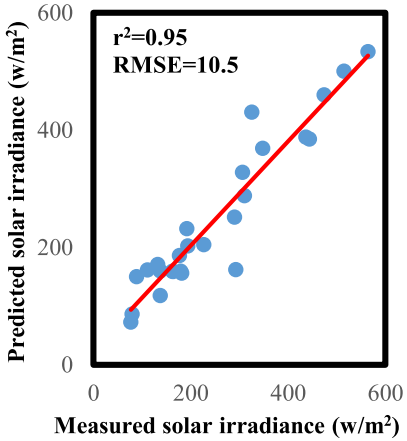


(c) 120 minutes ahead prediction for January 4, 2016.

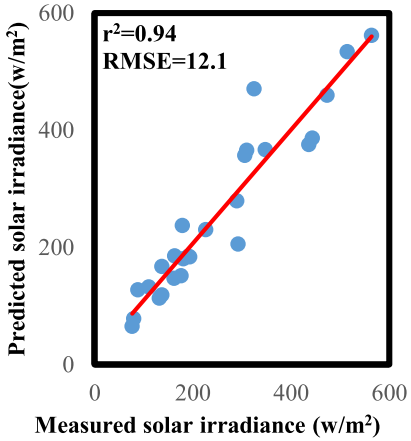


(d) 150 minutes ahead prediction for January 4, 2016.

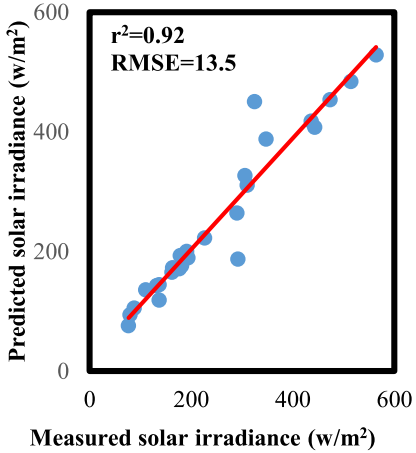
Figure 5.6: Solar irradiance prediction for January 4, 2016.



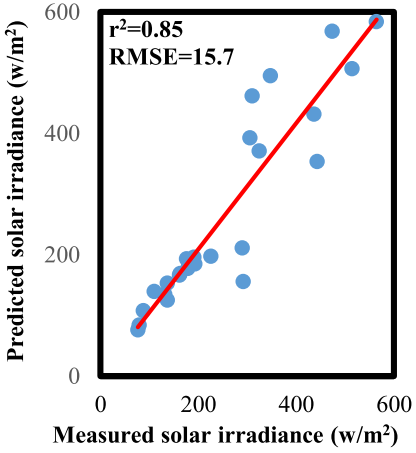
(a) 60 minutes ahead solar irradiance prediction for January 4, 2016.



(b) 90 minutes ahead solar irradiance prediction for January 4, 2016.

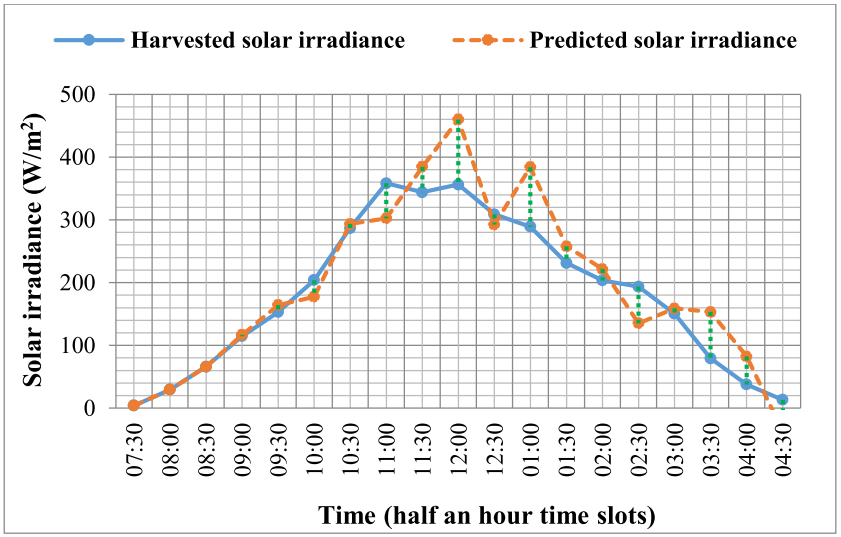


(c) 120 minutes ahead solar irradiance prediction for January 4, 2016.

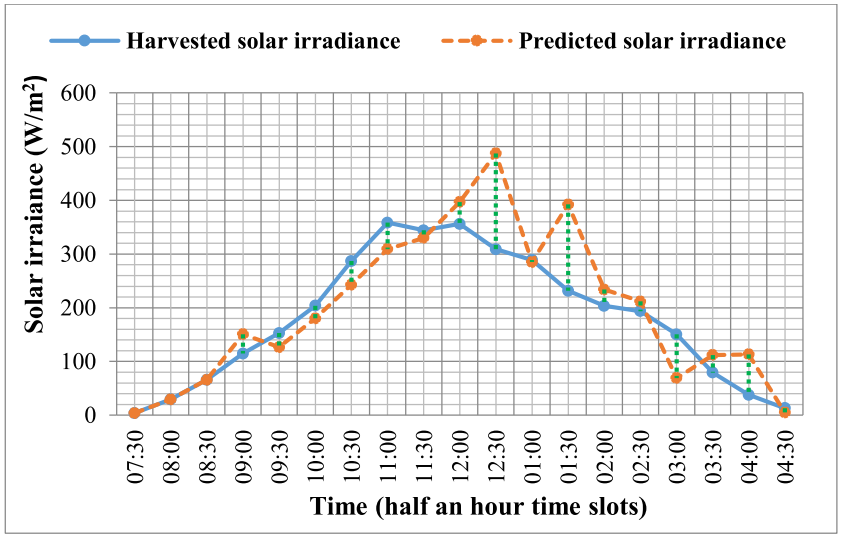


(d) 150 minutes ahead solar irradiance prediction for January 4, 2016.

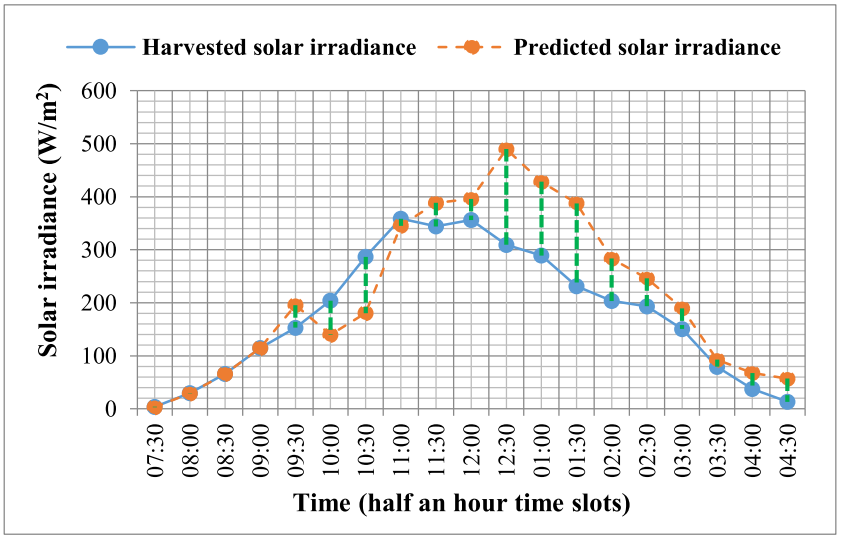
Figure 5.7: Correlation coefficient for January 4, 2016.



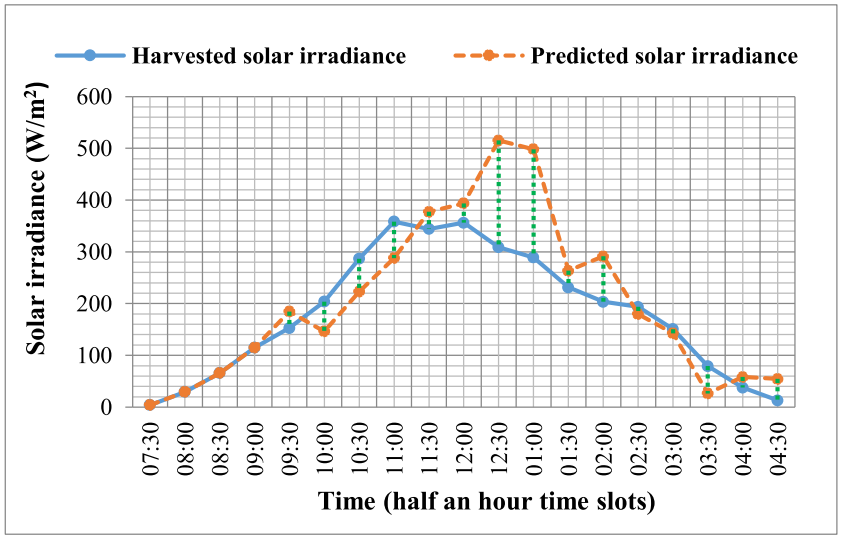
(a) 60 minutes ahead prediction for January 7, 2016.



(b) 90 minutes ahead prediction for January 7, 2016.

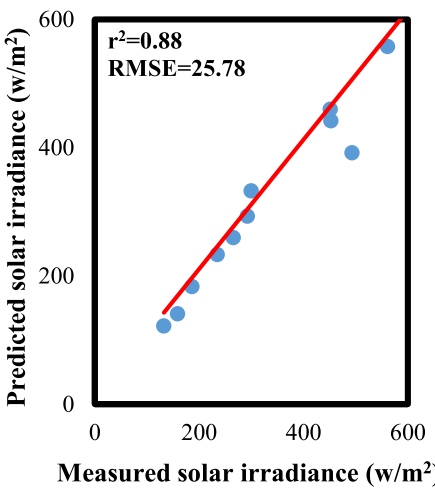


(c) 120 minutes ahead prediction for January 7, 2016.

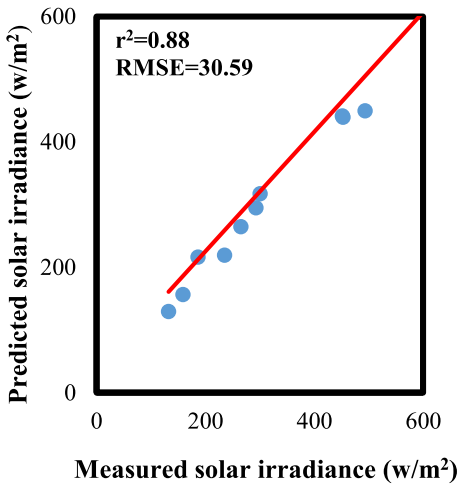


(d) 150 minutes ahead prediction for January 7, 2016.

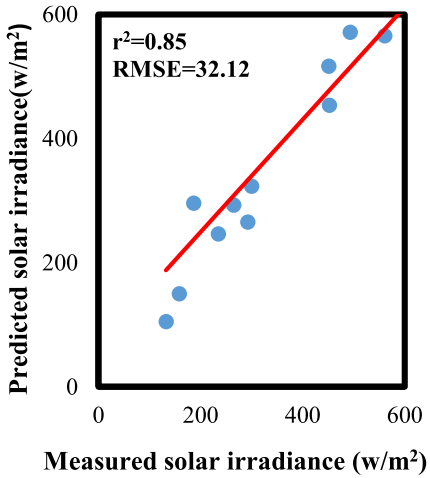
Figure 5.8: Solar irradiance prediction for January 7, 2016.



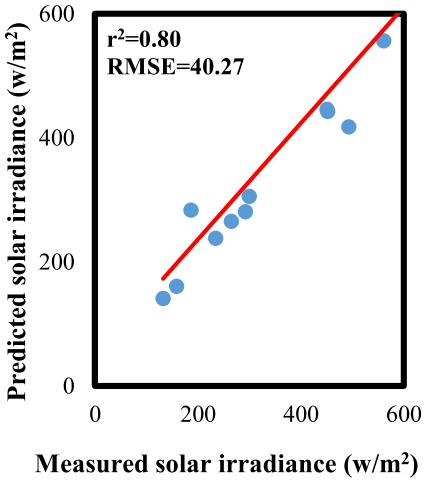
(a) 60 minutes ahead solar irradiance prediction for January 7, 2016.



(b) 90 minutes ahead solar irradiance prediction for January 7, 2016.



(c) 120 minutes ahead solar irradiance prediction for January 7, 2016.



(d) 150 minutes ahead solar irradiance prediction for January 7, 2016.

Figure 5.9: Correlation coefficient for January 7, 2016.

For 60 minutes ahead prediction (Figure 5.4(a)), estimated traces of solar irradiance is very close to real solar profile. The Root Mean Square Error (RMSE) is 6.73 (Figure 5.5(a)). A small deviation is observed during peak hours (10:30am to 1:00pm) in 90 (Figure 5.4(b)), 120 (Figure 5.4(c)), and 150 (Figure 5.4(d)) minutes ahead predictions and results in 6.73 (Figure 5.5(b)), 7.01 (Figure 5.5(c)), and 7.10 (Figure 5.5(d)) RMSE respectively. Figure 5.6 represents predicted energy profiles of January 4, 2016 for 60 (Figure 5.6(a)), 90 (Figure 5.6(b)), 120 (Figure 5.6(c)), and 150 (Figure 5.6(d)) minutes ahead prediction horizons. Results shows that similar accuracy level is achieved as for January 1. The values of RMSE and correlation coefficient are depicted in Figure 5.7

Figure 5.8 represents predicted energy profiles of January 7, 2016 for 60 (Figure 5.8(a)), 90 (Figure 5.8(b)), 120 (Figure 5.8(c)), and 150 (Figure 5.6(d)) minutes ahead prediction horizons. Results shows that predicted energy profiles is deviated from actual energy traces in all the prediction horizons. This situation arises when environmental conditions of a particular day are highly different from historical data. The values of RMSE and correlation coefficients are shown in Figure 5.9. The Root Mean Square Error (RMSE) of the order of 25.78, 30.59, 32.12, and 40.27 is observed during 60 (Figure 5.9(a)), 90 (Figure 5.9(b)), 120 (Figure 5.9(c)), and 150 minutes (Figure 5.9(d)) ahead prediction.

5.2.2 Pre-estimation of Node Duty Cycle

The forecasted values of solar irradiance for next N slots are used to compute the duty cycle for these slots. For this, the slots are divided in two classes, given as:

$$S_{sun} = E_{predicted}(i) - E_c \geq 0 \qquad \forall i \tag{5.26}$$

$$S_{Dark} = E_{predicted}(i) - E_c < 0 \qquad \forall i \tag{5.27}$$

In any particular slot, if estimated harvested energy is greater than node consumption, slot is referred as sun slot and vice-versa. By using slot classification in equation 5.21, the resultant equation is:

$$\begin{aligned} &\sum_{i=1}^{N} B(i) - B(i+1) = \sum_{i \epsilon S_{Dark}} \Delta T.Duty(i)(E_c - E_{predicted}(i)) - \sum_{i=1}^{N} \eta\Delta \\ &TE_{predicted}(i) + \sum_{i=1}^{N} \eta\Delta T E_{predicted}(i) Duty(i) - \sum_{i \epsilon S_{sun}} \eta\Delta T \\ &Duty(i)(E_{predicted}(i) - E_c) \end{aligned} \tag{5.28}$$

For an energy neutral condition, the left side of equation 5.28 is equated to zero. By

equating right side of the equation 5.28 equal to zero, the equation is reformulated as:

$$\sum_{i=1}^{N} E_{predicted}(i) = \sum_{i \epsilon S_{Dark}} Duty(i) \\ [\frac{E_c}{\eta} + E_{predicted}(i)(1 - \frac{1}{\eta})] + \sum_{i \epsilon S_{Sun}} E_c Duty(i) \tag{5.29}$$

Left hand side of equation 5.29 shows the estimation of total energy harvested in N slots. The first term on right side shows total energy consumed in dark slots and second term give energy consumption in sun slots. Initially, energy is allocated to maximize duty cycle in sun slots and assign minimum allowed values for dark slots as given in equation 5.30:

Initial assignment

$$Duty(i) = D_{max} \qquad \forall i \epsilon S_{sun} \tag{5.30}$$

$$Duty(i) = D_{min} \qquad \forall i \epsilon S_{dark} \tag{5.31}$$

Subsequent cases

Initial energy assignment given in equations 5.30 and 5.31 is neither optimal nor feasible for different energy states. Thus, duty cycle adjustments are required with respect to estimation of energy availability. The subsequent cases are described below:

- *Estimated harvested energy is greater than total energy consumption*:

$$\sum_{i=1}^{N} E_{predicted}(i) > \sum_{i \epsilon S_{Dark}} Duty(i)[\frac{E_c}{\eta} + E_{predicted}(i)(1 - \frac{1}{\eta})] + \sum_{i \epsilon S_{Sun}} E_c Duty(i) \tag{5.32}$$

  In this case, the excess energy is utilized to increase the duty cycle in dark slots as sun slots are already set to D_{max}. For this, the dark slot coefficients ($\frac{E_c}{\eta} + E_{predicted}(i)(1 - \frac{1}{\eta})$) are arranged in ascending order and assign D_{max} to smallest dark slot coefficient. This process continues until the excess energy is not sufficient to assign D_{max} further. The remaining energy increase the duty cycle of next slot j to a level which is calculated as:

$$Duty(j) = \frac{E_{remaining}}{(E_{predicted}(j) - E_c)/\eta - E_{predicted}(j)} + D_{min} \tag{5.33}$$

  If there is remaining energy after allocating D_{max} to all slot, the node will operate on maximum performance level and this energy is stored in battery for future use.

- *Estimated Harvested Energy is Lower than Total Energy Consumption*:

  Initial duty cycle assignment require more energy and energy deficiency is calculated as:

$$L = \sum_{i \epsilon S_{Dark}} Duty(i)[\frac{E_c}{\eta} + E_{predicted}(i)(1 - \frac{1}{\eta})] + \sum_{i \epsilon S_{Sun}} E_c Duty(i) - \sum_{i=1}^{N} E_{predicted}(i) \tag{5.34}$$

  In this case, the residual energy level in the battery is utilized to maintain the assigned duty cycles. If this energy is not sufficient, the duty cycles of sun slots D_{max} are reduced uniformly by an equal amount as duty cycle in dark slots cannot be reduced below D_{min}.

5.2.2.1 Maximum Available Energy, $E_{available}$

The maximum available energy, $E_{available}$ for node operation is bounded by maximum and minimum duty cycles (D_{max}, D_{min}) as well as the battery backup, E_{backup}.

- If predicted energy is greater than average energy consumption in a particular time slot, $E_{available}$ is allocated with respect to D_{max} only and given as:

$$E_{available}(i) = D_{max} \times E_{predicted}(i) \tag{5.35}$$

  It is observed from equation 5.35, the battery backup is not considered in time slots where bulk energy is available for use. The excess energy is used to increase the duty cycle in dark slots and rest energy is stored in the storage device for future time slots.

- In the time slots with insufficient energy availability for node operation, the energy is allocated with respect to D_{min}, and E_{backup}. If E_{backup} is sufficient to maintain the duty cycle, the node utilize sum of harvested energy and battery charge for its operation otherwise node operates with low duty cycle as insufficient harvested energy is available. Both the cases are formulated as:

$$E_{available}(i) = \begin{cases} Duty(i)(E_{predicted}(i) + \eta(E_c - E_{predicted}(i))) & E_{backup} > 0 \\ Duty(i)(E_{predicted}(i)) & E_{backup} \leq 0 \end{cases} \tag{5.36}$$

5.2.2.2 Battery Backup for Future Time Slots

- When sufficient amount of environment energy is available, battery stores the excess energy while considering the constraints of battery efficiency, η and sleep period of duty cycle (1-D_{max}). The battery backup for these slots is computed as:

$$E_{backup}(i) = \eta \times (1 - D_{max}) \times E_{predicted}(i) + \eta \times (E_{predicted(i)} - E_c) + E_{backup}(i-1) \quad (5.37)$$

  The first part on the right hand side of the equation 5.37 shows the saved energy in sleep mode, second part denotes the surplus energy than required by the load and last term shows the battery backup in last time slot.

- The time slots with less energy availability will make use of battery backup to continue node operation. The backup in the battery is dependent upon battery efficiency, η, inactive period of duty cycle and $E_{predicted}$. At the end of time slot t remaining battery backup is given as:

$$E_{backup}(i) = \begin{cases} \eta \times (1 - Duty_(i)) \times E_{predicted}(i) & E_{backup}(i) \leq 0 \\ \eta \times (1 - Duty_(i)) \times E_{predicted}(i) + & \\ E_{backup}(i-1) - (Duty(i) \times \eta \times (E_c - E_{predicted}(i))) & E_{backup}(i) > 0 \end{cases} \quad (5.38)$$

  The first case in equation 5.38 denotes the saved energy in sleep mode. No residual energy is available in this condition. In the second case from equation 5.38, battery charge is used by the node to compensate the energy deficiency. The battery back up at the end of the slot is the sum of saved energy in sleep mode and remaining residual energy in the battery.

5.2.2.3 Energy Allocation Rules

The energy allocation for a slot is computed at the start of the slot t and taken as E_{budget}. The battery storage inefficiency ($\eta \leq 1$) is taken into account as harvest-usestore approach is used as basic architecture of sensor module. The amount of energy allotment for each slot is dependent upon present weather conditions and battery back-up. The parameter $E_{available}$ is used as key parameter to calculate E_{budget} for a particular slot. The upper and lower limits of energy budget are decided by D_{max} and D_{min} and given as:

$$E_{maximum} = D_{max} \times E_c \quad (5.39)$$

$$E_{minimum} = D_{min} \times E_c \quad (5.40)$$

Based on the equations 5.39 and 5.40, the energy assignment to a node at the start of each time slot is given as:

$$E_{budget}(i) = \begin{cases} E_{maximum}(i) & if E_{available}(i) > E_{maximum}(i) \\ E_{available}(i) & if E_{minimum}(i) < E_{available}(i) \leq E_{maximum}(i) \\ E_{minimum}(i) & if E_{available}(i) \leq E_{minimum}(i) \end{cases} \tag{5.41}$$

For a time slot i, the duty cycle $Duty(i)$ is computed using the parameter, $E_{budget}(i)$ and defined as:

$$Duty(i) = \frac{1}{E_c} \times E_{budget}(i) \tag{5.42}$$

E_c is assumed constant for N time slots and shows the energy consumption when node is running on D_{max}. The algorithm 5.1 summarizes the procedure.

5.2.3 Real Time Adaptivity to Pre-estimated Node Duty Cycle with Low Variance

To maintain the condition of energy neutrality, it is necessary to dynamically adopt the pre-estimated duty cycle with respect to real time energy measurements. If $Duty(i)$ is the pre-estimated duty cycle, $E_{predicted}(i)$ is the predicted solar irradiance, and $E_{actual}(i)$ is the real time energy profile, the difference between two energy profiles, $E_{deviation}(i)$ is given as:

$$E_{deviation}(i) = \begin{cases} E_{actual}(i) - E_{predicted}(i) & if E_{actual}(i) > E_c \\ E_{actual}(i) - E_{predicted}(i) - & \\ Duty(i)[E_{predicted}(i) - E_{actual}(i)](1 - \frac{1}{\eta}) & if E_{actual}(i) \leq E_c \end{cases} \tag{5.43}$$

- If measured solar irradiance, $E_{actual}(i)$ is greater than the energy consumption, E_c, the $E_{deviation}(i)$ is simply the difference between predicted and measured energy values.

- If measured solar irradiance, $E_{actual}(i)$ is lower than the energy consumption, E_c, the $E_{deviation}(i)$ also consider the extra energy drawn from the battery to maintain energy neutrality.

- If $E_{deviation}(i)$ is positive, the excess energy will be used to increase the node duty cycle. The slots with high energy availability are arranged in descending order and iteratively increase the duty cycle to D_{max} for performance maximization.

Algorithm 5.1 Pre-estimation of node duty cycle

- Condition for Energy neutral state and optimal duty cycle

$$\sum_{i=1}^{N} Duty(i)$$

$$B(N+1) \geq B_0$$

$$Duty(i) \geq D_{min} \qquad \forall i \epsilon (1, ..., N)$$

$$Duty(i) \leq D_{max} \qquad \forall i \epsilon (1, ..., N)$$

- Classification of sun and dark slots

$$S_{sun} = E_{predicted} - E_c \geq 0 \qquad \forall i$$

$$S_{Dark} = E_{predicted} - E_c < 0 \qquad \forall i$$

- Estimation of battery back-up

$$\sum_{i=1}^{N} B(i) - B(i+1) = \sum_{i \epsilon S_{Dark}} \Delta T Duty(i)(E_c - E_{predicted}(i)) - \sum_{i=1}^{N} \eta \Delta$$

$$TE_{predicted}(i) + \sum_{i=1}^{N} \eta \Delta T E_{predicted}(i) Duty(i) - \sum_{i \epsilon S_{sun}} \eta \Delta T$$

$$Duty(i)(E_{predicted}(i) - E_c)$$

- For energy neutrality the following condition should be satisfied:

$$\sum_{i=1}^{N} E_{predicted}(i) = \sum_{i \epsilon S_{Dark}} Duty(i)[\frac{E_c}{\eta}$$

$$+ E_{predicted}(i)(1 - \frac{1}{\eta})] + \sum_{i \epsilon S_{Sun}} E_c Duty(i)$$

- Initial assignment

$$Duty(i) = D_{max} \qquad \forall i \epsilon S_{sun}$$

$$Duty(i) = D_{min} \qquad \forall i \epsilon S_{dark}$$

- Subsequent cases
  1. If estimated Harvested Energy is greater than Total Energy Consumption, then $E_{available}$ is:

$$E_{available}(i) = D_{max} \times E_{predicted}(i)$$

  2. If Estimated Harvested Energy is Lower than Total Energy Consumption, then $E_{available}$ is:

$$E_{available}(i) = \begin{cases} D(i)(E_{predicted}(i) + \eta(E_c - E_{predicted}(i))) & E_{backup} > 0 \\ D(i)(E_{predicted}(i) & E_{backup} \leq 0 \end{cases}$$

- Battery Backup for Future Time Slots

$$E_{backup}(i) = \eta \times (1 - D_{max}) \times E_{predicted}(i) + \eta \times (E_{predicted(i)-E_c}) + E_{backup}(i)$$

$$E_{backup}(i) = \begin{cases} \eta \times (1 - duty(i)) \times E_{predicted}(i) & if E_{backup}(i) \leq 0 \\ \eta \times (1 - duty(i)) \times E_{predicted}(i) + E_(backup(i-1)) & \\ -(duty(i) \times \eta \times (E_c - E_{predicted}(i))) & if E_{backup}(i) > 0 \end{cases}$$

- Energy Allocation Rules

$$E_{maximum} = D_{max} \times E_c$$

$$E_{minimum} = D_{min} \times E_c$$

$$E_{budget}(i) = \begin{cases} E_{maximum}(i) & if E_{available}(i) > E_{maximum}(i) \\ E_{available}(i) & if E_{minimum}(i) < E_{available}(i) \leq E_{maximum}(i) \\ E_{minimum}(i) & if E_{available}(i) \leq E_{minimum}(i) \end{cases} \tag{5.44}$$

- Estimated node duty cycle

$$Duty(i) = \frac{1}{E_c} \times E_{budget}(i)$$

Updated Duty Cycle:

To account the change in duty cycle, the residual energy, $E_{residual}(i)$ is computed. This quantity signifies the amount of energy usage by node because of changed duty cycle and

given as:

$$E_{residual}(i) = \begin{cases} E_c \times (D_{max} - Duty(i)) & if E_{actual} \geq E_c \\ (D_{max} - Duty(i))[\frac{E_c}{\eta} + E_{actual}(i)(1 - \frac{1}{\eta})] & if E_{actual} < E_c \end{cases} \tag{5.45}$$

$$Duty(i) = \begin{cases} D_{maximum} & if E_{residual}(i) \leq E_{deviation}(i) \\ Duty(i) + \frac{E_{deviation}(i)}{E_c} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) > E_c \\ Duty(i) + \frac{E_{deviation}(i)}{\frac{E_c}{\eta} + E_{actual}(i) \times (1 - \frac{1}{\eta})} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) \leq E_c \end{cases} \tag{5.46}$$

- If $E_{deviation}(i)$ is negative, the duty cycle of subsequent time slots is reduced to compensate the shortfall of energy. The slots with less energy availability are arranged in ascending order and iteratively decreases the duty cycle to D_{min} in order to reduce energy consumption by an amount $E_{deviation}$.

Updated Duty Cycle:

$$E_{residual}(i) = \begin{cases} E_c \times (D_{min} - Duty(i)) & if E_{actual} \geq E_c \\ (D_{min} - Duty(i))[\frac{E_c}{\eta} + E_{actual}(i)(1 - \frac{1}{\eta})] & if E_{actual} < E_c \end{cases} \tag{5.47}$$

$$Duty(i) = \begin{cases} D_{min} & if E_{residual}(i) \leq E_{deviation}(i) \\ Duty(i) + \frac{E_{deviation}(i)}{E_c} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) > E_c \\ Duty(i) + \frac{E_{deviation}(i)}{\frac{E_c}{\eta} + E_{actual}(i) \times (1 - \frac{1}{\eta})} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) \leq E_c \end{cases} \tag{5.48}$$

The proposed adaptive duty cycle algorithm compensate for the prediction errors but cannot be considered as a substitute for a good prediction method. The accurate estimation solar irradiance is necessary for optimal duty cycle allocation whereas the dynamic duty cycle adaption ensures the energy availability. The summarized procedure has been given in algorithm 5.2.

5.3 Performance Analysis of Proposed Adaptive Duty Cycle Algorithm in Different Forecasting Horizons

The performance of proposed algorithm is analysed in terms of average duty cycle obtained and effect of different prediction horizons on the stability of computed duty cycle.

The performance analysis of proposed algorithm is carried out in two steps. In the first step, pre-estimation of node duty cycle is done and analysed with respect to predicted solar irradiance. For this analysis, upper and lower limit for average duty cycle is taken as 70% and 1% respectively. In the second step, real time changes in node duty cycle are observed with respect to measured solar irradiance.

5.3.1 Analysis of Pre-estimated Duty Cycle

January 1, January 4, and January 7 of the year 2016 are considered for the analysis of proposed algorithm. Calculation of duty cycle is carried out during sun hours i.e. from 7:30am to 4:30pm and total time period is divided in half an hour time slots. Duty cycle is pre-estimated from one hour to two and half an hour ahead time period.

- ***Node Duty Cycle for January 1, 2016***

Algorithm 5.2 Real Time Adaptivity to Pre-estimated Node Duty Cycle with Low Variance

- Difference between predicted and real time energy profile

$$E_{deviation}(i) = \begin{cases} E_{actual}(i) - E_{predicted}(i) & if E_{actual}(i) > E_c \\ E_{actual}(i) - E_{predicted}(i) - Duty(i)[E_{predicted}(i) & \\ -E_{actual}(i)](1 - \frac{1}{\eta}) & if E_{actual}(i) \leq E_c \end{cases}$$

- If $E_{residual}(i)$ is positive, then updated duty cycle Updated Duty Cycle:

$$E_{residual}(i) = \begin{cases} E_c \times (D_{max} - Duty(i)) & if E_{actual} \geq E_c \\ (D_{max} - Duty(i))[\frac{E_c}{\eta} + E_{actual}(i)(1 - \frac{1}{\eta})] & if E_{actual} < E_c \end{cases}$$

$$Duty(i) = \begin{cases} D_{maximum} & if E_{residual}(i) \leq E_{deviation}(i) \\ Duty(i) + \frac{E_{deviation}(i)}{E_c} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) > E_c \\ Duty(i) + \frac{E_{deviation}(i)}{\frac{E_c}{\eta} + E_{actual}(i) \times (1 - \frac{1}{\eta})} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) \leq E_c \end{cases}$$

- If $E_{residual}(i)$ is negative, then updated duty cycle Updated Duty Cycle:

$$E_{residual}(i) = \begin{cases} E_c \times (D_{min} - Duty(i)) & if E_{actual} \geq E_c \\ (D_{min} - Duty(i))[\frac{E_c}{\eta} + E_{actual}(i)(1 - \frac{1}{\eta})] & if E_{actual} < E_c \end{cases}$$

$$Duty(i) = \begin{cases} D_{min} & if E_{residual}(i) \leq E_{deviation}(i) \\ Duty(i) + \frac{E_{deviation}(i)}{E_c} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) > E_c \\ Duty(i) + \frac{E_{deviation}(i)}{\frac{E_c}{\eta} + E_{actual}(i) \times (1 - \frac{1}{\eta})} & if E_{residual}(i) > E_{deviation(i)} \wedge E_{actual}(i) \leq E_c \end{cases}$$

Duty cycle of an energy harvested node is dependent upon the available solar irradiance and residual battery level. It is shown in Figure 5.10, these three quantities are depicted simultaneously. It is assumed that initially the battery is fully charged. For 1 hour ahead prediction with 6.21 RMSE, an average duty cycle of 66.31% is computed as depicted in Figure 5.10(a). It is also clear from the Figure 5.10(a) that in the absence of the solar energy, residual battery charge maintains the duty cycle to an acceptable level. During the slots where considerable solar intensity is available, solar irradiance drives the node and remaining energy recharge the battery.

For 90 minutes ahead prediction, Cubist results in RMSE of the order of 6.73 that offers similar solar profile as in previous case. For this prediction horizon, 65.78% average duty cycle is pre-estimated (Figure 5.10(b)). For 120 and 150 minutes ahead prediction horizon, Cubist offers 7.01 and 7.10 RMSE that results in 66.84% and 66.31% average duty cycle respectively (Figure 5.10(c) and (d)).

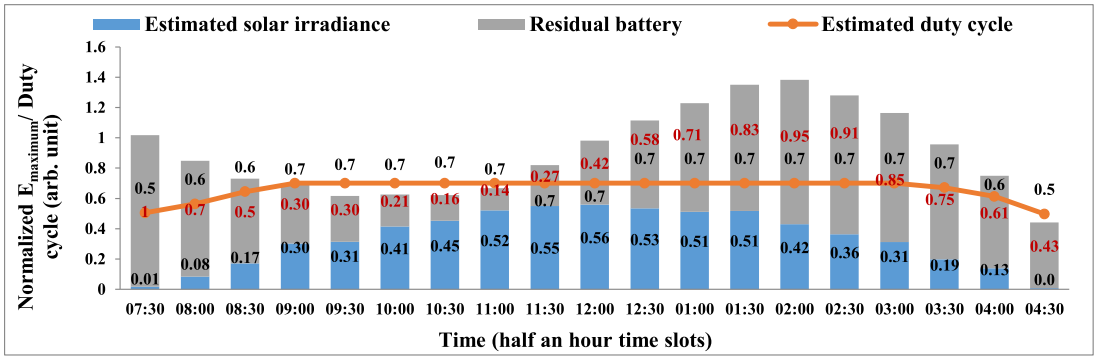
It is observed that duty cycle of 60% to 70% is maintained for slots from 8:00am to 4:00pm in all the prediction horizons and reduced to 50% only in those slots where solar energy is not available and battery backup is used to operate the node.

- ***Node Duty Cycle for January 4, 2016***

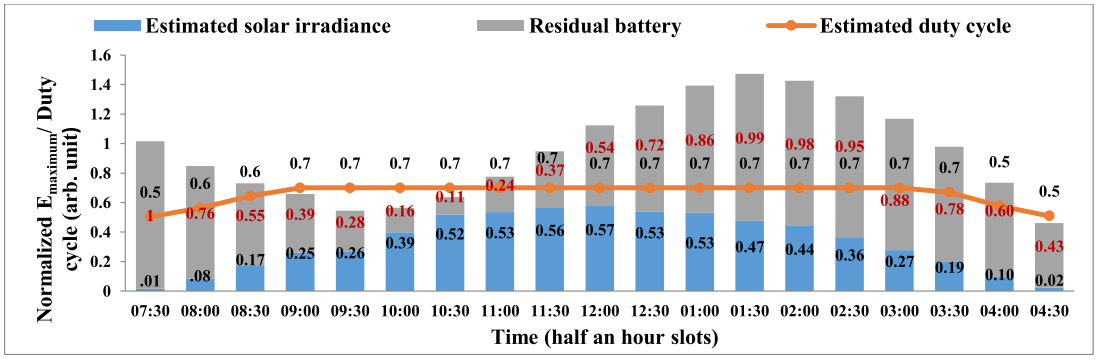
  By considering the energy profile of January 4, 2016 as depicted in Figure 5.3, it is observed that solar intensity level is similar to January 1, 2016. For 60 minutes ahead prediction, Cubist offers 10.5 RMSE. It is seen in the Figure 5.11(a) that minimum duty cycle of 1% is estimated during 4:00pm when solar intensity is very low and battery get energy depleted. In the next time slot when battery recharge itself, the duty cycle raised to 40%. With respect to this prediction horizon, the average node duty cycle is 62.5%.

  For 90 minutes ahead prediction, low irradiance level is observed during initial time slots. Thus, the node will not attain maximum duty cycle during these slots. Further, negligible solar intensity is observed at 4:00pm but residual battery level is sufficient to operate the node on 50% duty cycle. For this forecasting horizon with 12.1 RMSE, 64.0% average duty cycle is obtained (Figure 5.11(b)).

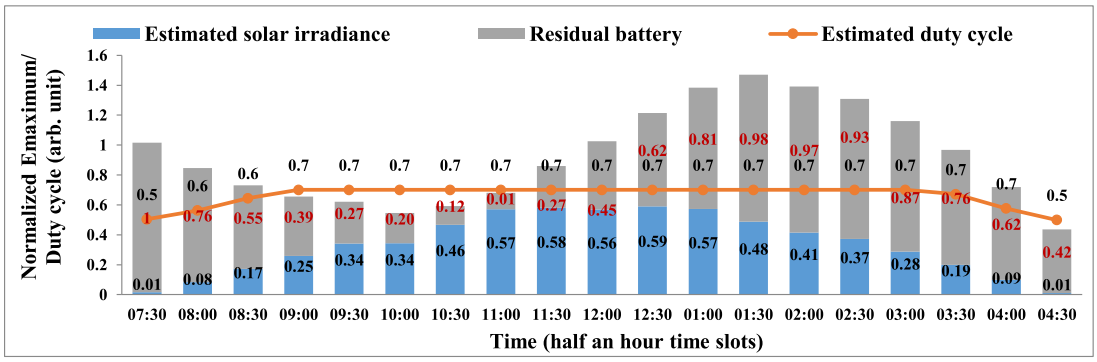
  With respect to 120 minutes ahead prediction, 65% average node duty cycle is estimated. Minimum duty cycle is obtained at 4:30pm where solar intensity is negligible and node is powered by battery only (Figure 5.11(c)). Similarly, 65.5% average duty cycle is achieved with respect to 150 minutes ahead solar prediction (Figure 5.11(d)).



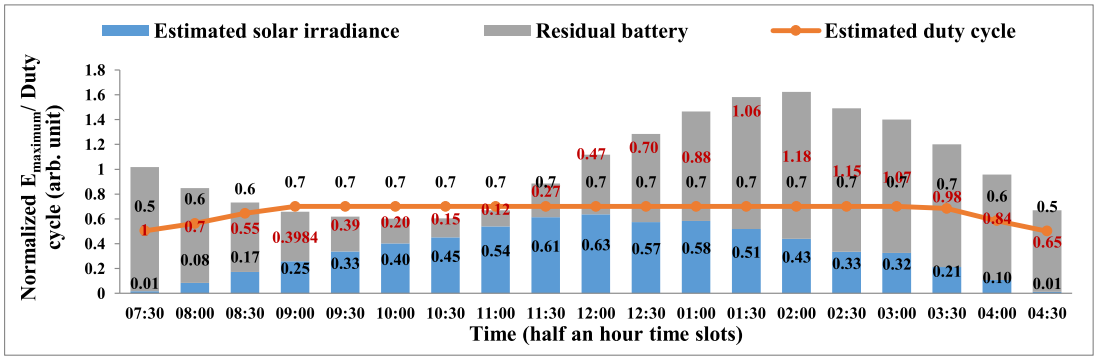
(a) Pre-estimation of node duty cycle in 60 minutes ahead prediction horizon for January 1, 2016



(b) Pre-estimation of node duty cycle in 90 minutes ahead prediction horizon for January 1, 2016

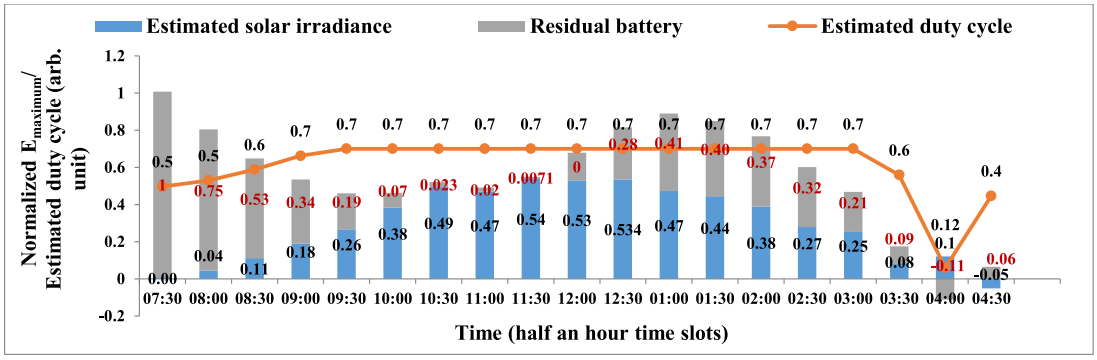


(c)Pre-estimation of node duty cycle in 120 minutes ahead prediction horizon for January 1, 2016

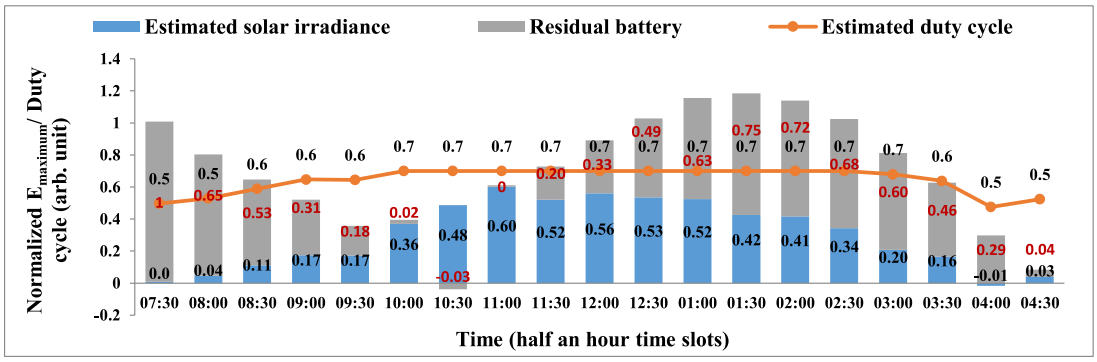


(d) Pre-estimation of node duty cycle in 150 minutes ahead prediction horizon for January 1, 2016

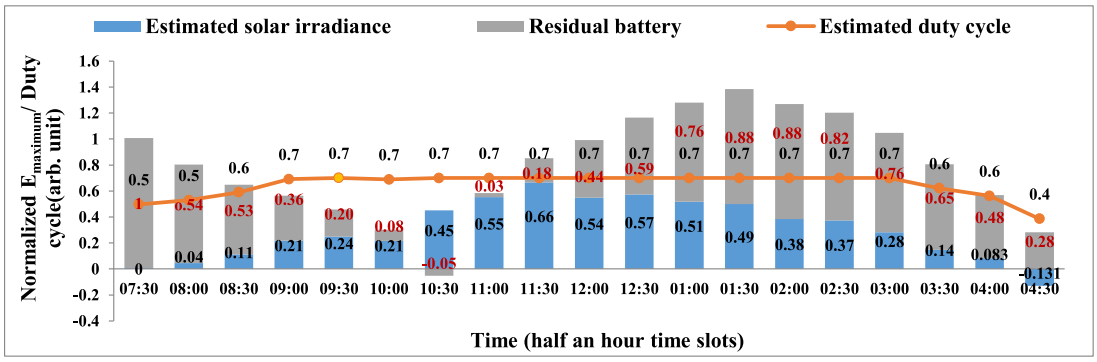
Figure 5.10: Pre-estimation of node duty cycle for January 1, 2016



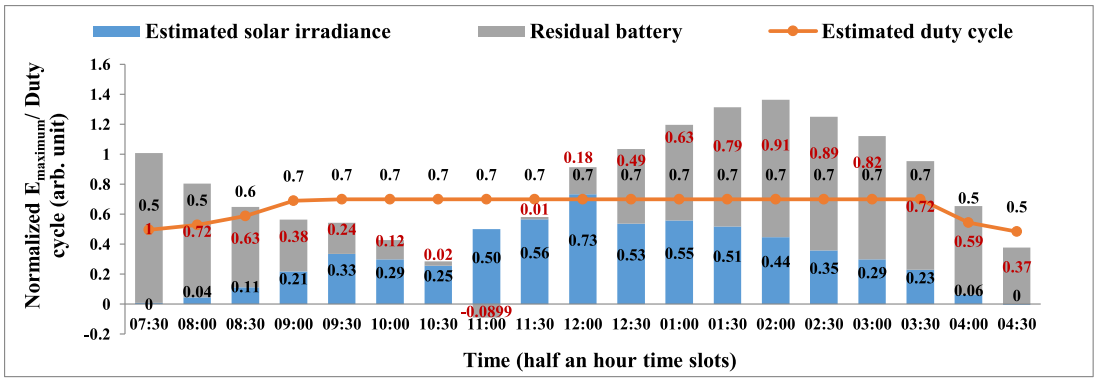
(a) Pre-estimation of node duty cycle in 60 minutes ahead prediction horizon for January 4, 2016



(b) Pre-estimation of node duty cycle in 90 minutes ahead prediction horizon for January 4, 2016



(c) Pre-estimation of node duty cycle in 120 minutes ahead prediction horizon for January 4, 2016



(d) Pre-estimation of node duty cycle in 150 minutes ahead prediction horizon for January 4, 2016

Figure 5.11: Pre-estimation of node duty cycle for January 4, 2016

- ***Node Duty Cycle for January 7, 2016***

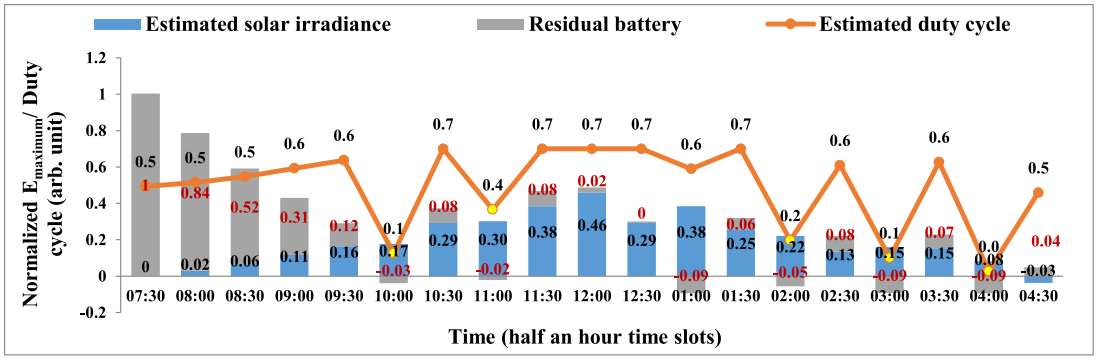
It is seen from the Figure 5.3 that predicted solar intensity is low on January 7, 2016 as compared to January 1 and January 7, 2016. Low harvested energy causes poor charging of the storage device and low average duty cycle as it is depicted in Figure 5.12. Further, different weather conditions with respect to past days also affect the prediction accuracy.

For one hour ahead prediction horizon, Cubist offers 25.78% prediction error. It is seen from the Figure 5.12(a) that during initial sun hours (7:00am to 10:00am), very low solar irradiance is estimated. In this condition, the node is driven by residual energy of battery. During 10:00am, the duty cycle drops to 1% as battery get depleted. The proposed algorithm maintains the continuous operation of node until the available energy is insufficient to operate the node. It is observed from the Figure 5.12(a) that node is shut down at 4:00am when harvested energy is negligible and battery get energy depleted. The average node duty cycle is estimated as 50%. The same response is observed for 90 minutes (5.12(b)), 120 minutes (5.12(c)), and 150 (5.12(d)) minutes ahead prediction horizons where average duty cycle is observed of the order of 48.94%, 50%, and 48.94% respectively.

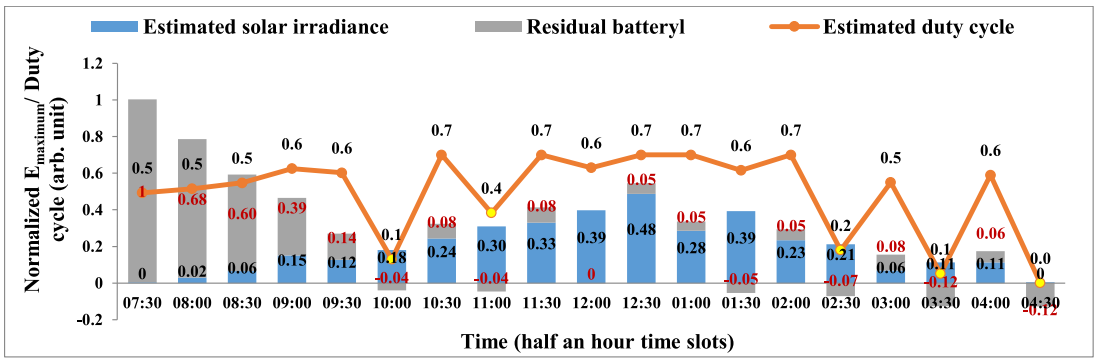
It is observed from the above analysis that the proposed algorithm attain high average duty cycle in sunny weather conditions and avoid node shut down in the condition of energy depletion (low battery backup and low solar irradiance). Table 5.1 summarizes different parameters obtained during pre-estimation of node duty cycle.

Table 5.1: Parameters of pre-computed duty cycle.

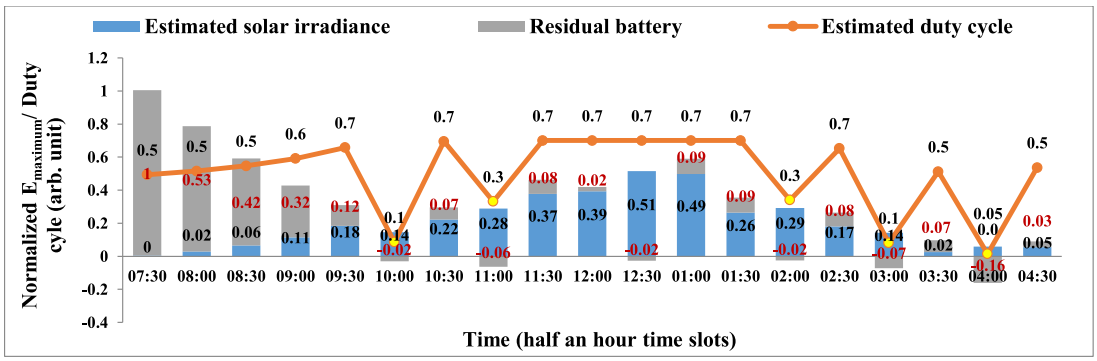
Days under consideration	**January 1, 2016**				**January 4, 2016**				**January 7, 2016**			
Prediction horizon (minutes)	60	90	120	150	60	90	120	150	60	90	120	150
Maximum duty cycle (%), D_{max}	70	70	70	70	70	70	70	70	70	70	70	70
Minimum duty cycle (%),D_{min}	49.65	50.41	49.92	50.41	5.95	47.56	38.61	48.52	2.72	0.01	1.36	0.01
Average duty cycle (%), D_{avg}	65.31	66.78	66.84	66.31	62.5	64.01	65.0	65.5	50.0	48.94	50.0	48.94



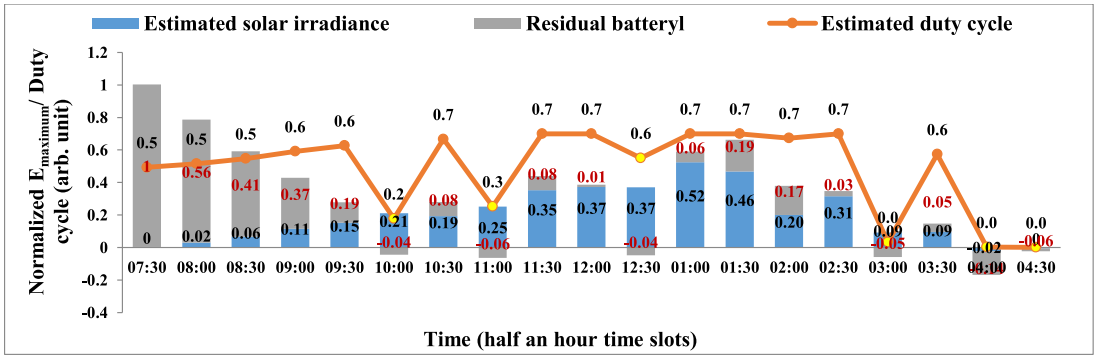
(a) Pre-estimation of node duty cycle in 60 minutes ahead prediction horizon for January 7, 2016



(b) Pre-estimation of node duty cycle in 90 minutes ahead prediction horizon for January 7, 2016



(c)Pre-estimation of node duty cycle in 120 minutes ahead prediction horizon for January 7, 2016

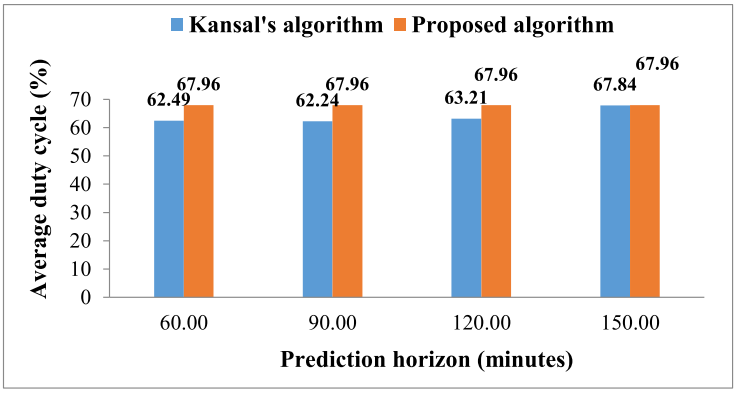


(d) Pre-estimation of node duty cycle in 150 minutes ahead prediction horizon for January 7, 2016

Figure 5.12: Pre-estimation of node duty cycle for January 7, 2016.

5.3.2 Analysis of Real-time Duty Cycle

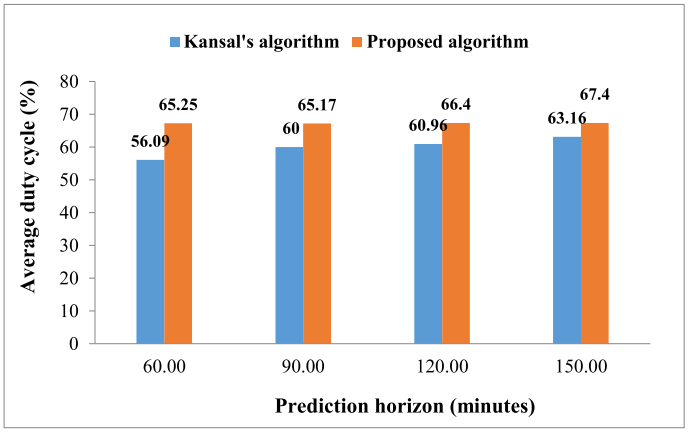
In this analysis, real time duty cycle is measured with respect to estimated solar irradiance in different forecasting horizons. The effectiveness of the proposed work is calculated in terms of magnitude and stability of computed duty cycle. This measure of goodness-of-fit allow pre-estimation of node duty cycle for network scheduling and planning.



Performance comparison for January 1, 2016.

Figure 5.13: Performance comparison for January 1, 2016.

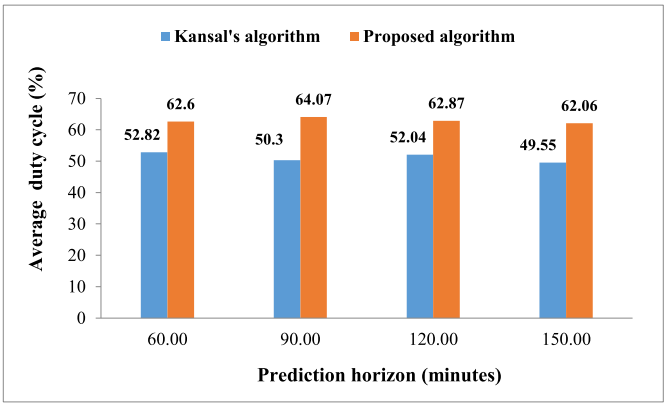
Figure 5.13 shows a performance comparison between proposed algorithm and Kansal *et al.*[58] for real time adaption of duty cycle. A stable duty cycle of 67.96% is observed over all prediction horizons by the proposed work whereas a variation of 6% (62.49% to 67.84%) is shown by [58] approach. Further, a maximum deviation of 2.18 % is observed between estimated and real time duty cycle. For January 4, 2016, Figure 5.14 shows



Performance comparison for January 4, 2016.

Figure 5.14: Performance comparison for January 4, 2016.

average duty cycle measured for the node with respect to different forecasting horizons and maximum duty cycle of the order of 67.4% is calculated. Further, the proposed



Performance comparison for January 7, 2016.

Figure 5.15: Performance comparison for January 7, 2016.

adaptive duty cycle algorithm shows a deviation of 2.23% whereas a variation of 7.07% (56.09% to 63.16%) is observed in [58] approach.

Further, for January 7, 2016 where low duty cycle is estimated because of low solar intensity, the proposed approach is effective in increasing the duty cycle by efficiently utilizing the available energy and also circumvent the prediction errors. It is seen in the Figure 5.15 that for estimated average duty cycles (50%, 48.94%, 50%, and 48.94%) for 60, 90, 120, and 150 minutes ahead time period respectively, the proposed adaptive duty cycle algorithm results in 62.6%, 64.07%, 62.87% and 62.06% average duty cycle for 60, 90, 120, and 150 minutes ahead time period respectively. It is evident from the the Figure 5.15 that the computed duty cycle offers high magnitude and stability in comparison to Kansal *et al.* [58] approach.

It is observed from the results that the proposed algorithm outperforms Kansal *et al.*[58] for all the three days. Results validate that the proposed approach offers high and stable duty cycle by compensating the variations in solar energy profiles and efficiently utilizing the available energy.

5.4 Concluding Remarks

An improved adaptive duty cycle algorithm for energy harvested wireless sensor networks is proposed. The key feature of the proposed algorithm is the energy assignment principles along with the ability to adapt real time changes in pre-computed duty cycle. Pre-estimation of minimum, maximum, and average node duty cycle prevent node from shut down condition in energy depletion state.

The accuracy of solar irradiance prediction algorithm plays a significant role in the proposed approach. Based on the predicted solar irradiance values and by considering storage element efficiency, the node duty cycle is pre-estimated. Simulation results shows in section 5.3.1 that algorithm results in stable and significant magnitude of duty cycle of the order of 66.84%, 65.5%, and 50% for January 1, January 4, and January 7, 2016 respectively. Energy budget principles efficiently schedule the predicted energy and avoid the node running out of energy.

Further, real time adaption to duty cycle compensate the deviation between predicted and measured solar profiles. Under stable weather conditions, less deviation is expected between these profiles and proposed approach leads to stable duty cycle with respect to all horizons. Results also validate that the real time duty cycle is very close to predicted ones. The proposed algorithm also handles the low residuals levels of storage elements and low irradiance levels of solar energy by dynamically adopting the duty cycle with respect to current energy states in the network. In case of low solar irradiance availability as for January 7, 2016 as shown in section 5.3.2 where pre-estimated average duty cycle has been calculated in range of 49% to 50% in different forecasting horizons, effective energy management policies in real time duty cycle adaption algorithm increase the node duty cycle to 64%.

The abrupt changes in the duty cycle are prevented by controlling the energy assignment for node operation at different time instants. This phenomenon reduces the variations in duty cycle between the different time slots as well as when the duty cycle is estimated in different prediction horizons. The adaptive duty cycle lead to efficient power management for the wireless sensor node.

Chapter 6

Conclusion and Future Directions

,

The research objective of this thesis is to optimize the operation of environmentally powered wireless sensor nodes in terms of average duty cycle. An energy efficient power management approach is proposed to achieve high average duty cycle with high stability for optimizing WSN performance. The main findings from the research are listed below:

1. ***A modified solar energy prediction algorithm is developed to gain close approximations of real time energy profiles.***

   In this thesis, a modified form of time series analysis based ***Pro-Energy*** solar forecasting algorithm is proposed. To account the intrahour and intraday variability, a similar day based data processing is done. For this purpose, the historical data is statistically modelled to investigate the nature of past energy profiles. The future energy patterns is interpreted from the pre-established patterns. The ***Trend*** component is employed to depict time series patterns. Trend is a systematic, linear/nonlinear component that changes over time, give long run direction of energy profiles (increase/decrease), and not repeat itself over specified range of considered data. To eliminate the non-systematic components from the data series, double exponential smoothing process is used. The statically manipulated data series is used to extract past similar days to the current day energy profile with minimum Euclidean distance. To give selective weightage to selected days, concept of combined weighted profile is used. The performance of proposed algorithm is compared with EWMA, WCMA, and Pro-Energy on the basis of prediction error. The simulation results shows that proposed algorithm achieves large reduction in prediction error over 30 minutes to few hours ahead prediction horizon in comparison with Pro-Energy algorithm. Simulations are performed using MATLAB interface and conducted on real time solar data from NREL.

2. ***The solar forecasting algorithm is optimized using machine learning approach to predict future energy more accurately.***

To meet with the second objective, ***Machine learning*** approach is adopted which is a sub-field of artificial intelligence. Similar day based methodology is used to enable the machine learning models into self learning mode. The trained models adapts to new data independently by utilizing the concept of pattern recognition and by learning previous computations. The approach has the ability to analyze large amount of data, interpret, and extract the useful information and filter out the rest data. By utilizing this feature, the data from six years (2010-2015) is used to execute second objective. Simulations are performed using R interface and conducted on real time solar data from NREL.

Further, the selection of a particular model depends upon the nature of data series. A number of regression based machine learning models are evaluated in terms of prediction accuracy. FoBa, leapForward, Spikeslab, Cubist, and bagEarthGCV models are used for simulation. Three independent test cases are conducted to evaluate the models in terms of prediction accuracy. It includes number of past days considered, number of past time slots considered, and simulation over different forecasting horizons. Seasonal effects are also considered while evaluating the performance. The simulation results shows that machine learning models provide high prediction accuracy over 1 hour to 48 hours ahead prediction horizons.

To further enhance the prediction accuracy, generalized mean based ensemble method is proposed that assures high consistency and accuracy in combined behavior of individual predictors. Optimal characteristics of accuracy and stability are obtained by aggregating base machine learning models using generalized mean approach (arithmetic mean, harmonic mean, quadratic mean, and median). Statistical comparison between different models and proposed approach is carried out with respect to seasonal variation and for forecasting horizons ranging from 1 hour ahead to 48 hours ahead. From the simulation results, it has seen that statistical ensembling achieve relatively high and stable prediction accuracy and hence overcome the shortcomings of independent machine learning models. Quadratic mean and median approaches achieves high performance matrix in terms of prediction accuracy, correlation coefficient, RMSE, and stability.

3. ***An adaptive duty cycle algorithm is proposed to optimize average duty cycle for a node to meet the performance requirements.***

   For an uninterrupted operation, sensor node should efficiently utilize the energy for different node operations including sensing, actuation, processing, and communication. For this purpose, a power management unit in the power section assist the node to tradeoff energy availability from energy harvesting source and battery

backup of node. In this thesis, an adaptive duty cycling based power management approach is proposed that employ energy budgeting principles for allocating energy to the node with respect to the present status of energy availability while considering harvesting source, battery backup, and load requirements.

Since solar energy is nondeterministic in nature, the pre-estimated solar energy profiles generated by Cubist model is used to model the energy generation rate, battery backup in future slots, and energy consumption rate. The prior information of all these parameters overcome the condition of energy depletion state in case of low or no harvesting energy. These estimated profiles are used to implement energy budgets for a sensor node that directly control the power consumption by the node and hence the duty cycle. This predicted duty cycle provides the estimate of node performance in future time slots. Further, based on the real time energy measurements, the predicted node duty cycle is updated accordingly. The results shows that the proposed approach achieves high average duty cycle with high stability and hence the improved network performance. Simulations are performed using MATLAB interface

4. ***The proposed algorithms are compared with the existing ones.***

   - The proposed modified Pro-Energy algorithm is compared with EWMA, WCMA, and existing Pro-Energy algorithm over 30 minutes to 120 minutes ahead prediction horizon in terms of Mean Absolute Error (MAPE). The results validate the effectiveness of proposed work over existing techniques for all prediction horizons.

   - The proposed adaptive duty cycle algorithm is compared with existing kansal *et al.*[58] approach. The first evaluation criteria is forecasting horizon over which the node duty cycle is pre-estimated. This parameter is important as it specify the time gap over which the pre-estimation of node status is formulated. The second criteria is the magnitude and stability of real time node duty cycle as it directly affect the network performance. The simulation results shows that the proposed power management approach has effectively respond to long time horizons and results in high and stable average duty cycle in comparison to existing one.

The proposed thesis has achieved all the research objectives presented in the objective one. To enhance the lift time of energy harvested wireless sensor networks, the future work should be focused on the following factors:

- The lifetime of a single energy harvested wireless sensor node has been considered.

As the life time of whole network is important, the future work can focus to increase the life cycle of energy harvested wireless sensor network.

- Solar energy based prediction model for wireless networks has been considered. By considering the cost and feasibility, the future can include other environmental energy resources such as thermal, RF and vibration energy to design sensor network and evaluate the performance.
- Finally, the most important research issue is overall system efficiency. Work can be extended to improve the circuit efficiency specially of the storage medium and the power management unit, cost, size and power overhead as these are the deciding factors of overall system efficiency.

References

[1] Jennifer Yick, Biswanath Mukherjee, and Dipak Ghosal. Wireless sensor network survey. *Computer Networks*, 52(12):2292–2330, 2008.

[2] Parag Aggarwal and Aditya Trivedi. Secure wireless communication using friendly noise through an untrusted relay. *Wireless Personal Communications*, 84(4):2927–2941, 2015.

[3] Rajiv K Tripathi, Yatindra Nath Singh, and Nishchal K Verma. Two-tiered wireless sensor networks–base station optimal positioning case study. *IET Wireless Sensor Systems*, 2(4):351–360, 2012.

[4] Rajeev K Shakya, Yatindra Nath Singh, and Nishchal K Verma. A novel spatial correlation model for wireless sensor network applications. 9^{th} *International Conference on Wireless and Optical Communications Networks, WOCN 2012*. 1-6. IEEE, 2012.

[5] Waltenegus Dargie and Christian Poellabauer. *Fundamentals of Wireless Sensor Networks: Theory and Practice*. John Wiley & Sons, 2010.

[6] Abdelmalik Bachir, Mischa Dohler, Thomas Watteyne, and Kin K Leung. Mac essentials for wireless sensor networks. *IEEE Communications Surveys & Tutorials*, 12(2):222–248, 2010.

[7] H Karl and A Willig. Protocols and Architectures for Wireless Sensor Networks: John Wiley. 2005.

[8] Tinynode 584 users manual. Shockfish SA. www.tinynode.com.

[9] MICAz: Wireless measurement system. Crossbow Technology Inc. www.xbow.com.

[10] Crossbow Technology Inc. and Imote2: High-performance wireless sensor network node. Intel Corporation. www.xbow.com.

[11] Tmote Sky: Low power wireless sensor module. Sentilla Corporation. www.sentilla.com.

[12] Sukhchandan Lally and Ljiljana Trajkovic. Performance analysis of routing protocols for wireless Ad-Hoc networks. *OPNETWORK*. 1–8, 2011.

[13] Vinay Kumar, Sanjay B Dhok, Rajeev Tripathi, and Sudarshan Tiwari. A review study of hierarchical clustering algorithms for wireless sensor networks. *International Journal of Computer Science Issues (IJCSI)*, 11(3):92, 2014.

[14] Ajay Singh Raghuvanshi, Sudarshan Tiwari, Rajeev Tripathi, and Nand Kishor. Optimal number of clusters in wireless sensor networks: a FCM approach. *International Journal of Sensor Networks*, 12(1):16–24, 2012.

[15] Chulsung Park, Kanishka Lahiri, and Anand Raghunathan. Battery discharge characteristics of wireless sensor nodes: An experimental analysis. 2^{nd} *Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, IEEE SECON 2005.* 430–440, IEEE, 2005.

[16] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, and Andrea Passarella. Energy conservation in wireless sensor networks: A survey. *Ad hoc Networks*, 7(3):537–568, 2009.

[17] Huige Wei, Daowei Ding, Xingru Yan, Jiang Guo, Lu Shao, Haoran Chen, Luyi Sun, Henry A Colorado, Suying Wei, and Zhanhu Guo. Tungsten trioxide/zinc tungstate bilayers: electrochromic behaviors, energy storage and electron transfer. *Electrochimica Acta*, 132:58–66, 2014.

[18] Jiahua Zhu, Minjiao Chen, Honglin Qu, Xi Zhang, Huige Wei, Zhiping Luo, Henry A Colorado, Suying Wei, and Zhanhu Guo. Interfacial polymerized polyaniline/graphite oxide nanocomposites toward electrochemical energy storage. *Polymer*, 53(25):5953–5964, 2012.

[19] Christophe J Merlin and Wendi B Heinzelman. Duty cycle control for low-power-listening MAC protocols. *IEEE Transactions on Mobile Computing*, 9(11):1508–1521, 2010.

[20] Robin Singh Bhadoria, Geetam Singh Tomar, and Sungmin Kang. Proficient energy consumption aware model in wireless sensor network. *International Journal of Multimedia and Ubiquitous Engineering*, 9(5):27–36, 2014.

[21] SR Khara and R Tripathi. An efficient polling and information-request packet transfer for wireless networks in fading and interference environments. *Journal of High Speed Networks*, 14(3):235–242, 2005.

[22] Rajeev K Shakya, Yatindra N Singh, and Nishchal K Verma. A correlation model for MAC protocols in event-driven wireless sensor networks. *IEEE Region 10 Conference TENCON 2012*. 1–6, IEEE, 2012.

[23] Kemal Akkaya and Mohamed Younis. A survey on routing protocols for wireless sensor networks. *Ad hoc Networks*, 3(3):325–349, 2005.

[24] Mano Yadav, Vinay Rishiwal, and KV Arya. Routing in wireless Ad hoc networks: A new horizon. *Journal of Computing*, 1(1):204–208, 2009.

[25] Stephen Makonin, William Sung, Ryan Dela Cruz, Brett Yarrow, Bob Gill, Fred Popowich, and Ivan V Bajic. Inspiring energy conservation through open source metering hardware and embedded real-time load disaggregation. *IEEE PES Asia-Pacific Power and Energy Engineering Conference, APPEEC 2013*. 1–6, IEEE, 2013.

[26] Amit Sinha and Anantha Chandrakasan. Dynamic power management in wireless

sensor networks. *IEEE Design & Test of Computers*, 18(2):62–74, 2001.

[27] Jeongyeup Paek, Ben Greenstein, Omprakash Gnawali, Ki-Young Jang, August Joki, Marcos Vieira, John Hicks, Deborah Estrin, Ramesh Govindan, and Eddie Kohler. The Tenet architecture for tiered sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 6(4):34, 2010.

[28] Mohamed Younis and Kemal Akkaya. Strategies and techniques for node placement in wireless sensor networks: A survey. *Ad hoc Networks*, 6(4):621–655, 2008.

[29] Arpit Tripathi, Pulkit Gupta, Aditya Trivedi, and Rahul Kala. Wireless sensor node placement using hybrid genetic programming and genetic algorithms. *Organizational Efficiency through Intelligent Information Technologies*. 125–144, IGI Global, 2013.

[30] Chris Knight, Joshua Davidson, and Sam Behrens. Energy options for wireless sensor nodes. *Sensors*, 8(12):8037–8066, 2008.

[31] Winston KG Seah, Zhi Ang Eu, and Hwee-Pink Tan. Wireless sensor networks powered by ambient energy harvesting (WSN-HEAP)-Survey and challenges. 1^{st} *International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009.* 1–5, IEEE, 2009.

[32] Mummadi Veerachary, Tomonobu Senjyu, and Katsumi Uezato. Neural network based maximum power point tracking of coupled inductor interleaved-boost-converter-supplied PV system using fuzzy controller. *IEEE Transactions on Industrial Electronics*, 50(4):749–758, 2003.

[33] Sujesha Sudevalayam and Purushottam Kulkarni. Energy harvesting sensor nodes: Survey and implications. *IEEE Communications Surveys & Tutorials*, 13(3):443–461, 2011.

[34] Abha Rajoriya and Eugene Fernandez. Hybrid energy system size optimization and sensitivity evaluation for sustainable supply in a remote region in India. *International Journal of Sustainable Energy*, 32(1):27–41, 2013.

[35] Olga Moraes Toledo, Delly Oliveira Filho, Antonia Sonia AC Diniz, Jose Helvecio Martins, and Maria Helena Murta Vale. Methodology for evaluation of grid-tie connection of distributed energy resources-case study with photovoltaic and energy storage. *IEEE Transactions on Power Systems*, 28(2):1132–1139, 2013.

[36] Aurobinda Panda, Mukesh kumar Pathak, and Satya Prakash Srivastava. Solar direct torque controlled induction motor drive for industrial applications. *International Journal of Renewable Energy Research (IJRER)*, 3(4):794–802, 2013.

[37] Stephen Makonin, Laura Guzman Flores, Robyn Gill, Roger Alex Clapp, Lyn Bartram, and Bob Gill. A consumer bill of rights for energy conservation. *IEEE Canada*

International Conference on Humanitarian Technology Conference, IHTC 2014. 1–6, IEEE, 2014.

[38] Chung-Yang Sue and Nan-Chyuan Tsai. Human powered MEMS-based energy harvest devices. *Applied Energy*, 93:390–403, 2012.

[39] Rudolf Heer, Jurgen Wissenwasser, M Milnera, L Farmer, C Hopfner, and M Vellekoop. Wireless powered electronic sensors for biological applications. *Annual International Conference on Engineering in Medicine and Biology Society (EMBC), 2010*. 700–703, IEEE, 2010.

[40] Pooja Lakhlan and Aditya Trivedi. Energy harvesting-based two-hop D2D communication in cellular networks. *International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*. 328–332, IEEE, 2016.

[41] Soumyajit Mandal, Lorenzo Turicchia, and Rahul Sarpeshkar. A low-power, battery-free tag for body sensor networks. *IEEE Pervasive Computing*, 9(1):71–77, 2010.

[42] Fei Fei, John D Mai, and Wen Jung Li. A wind-flutter energy converter for powering wireless sensors. *Sensors and Actuators A: Physical*, 173(1):163–171, 2012.

[43] Aoife M Foley, Paul G Leahy, Antonino Marvuglia, and Eamon J McKeogh. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1):1–8, 2012.

[44] Saurabh S Soman, Hamidreza Zareipour, Om Malik, and Paras Mandal. A review of wind power and wind speed forecasting methods with different time horizons. *North American Power Symposium, NAPS 2010*. 1–8, IEEE, 2010.

[45] Y Mishra, S Mishra, Fangxing Li, and ZY Dong. Eigenvalue analysis of a DFIG based wind power system under different modes of operations. *Wind Power Systems*. 191–213, Springer, 2010.

[46] Nicholas S Hudak and Glenn G Amatucci. Small-scale energy harvesting through thermoelectric, vibration, and radiofrequency power conversion. *Journal of Applied Physics*, 103(10):5, 2008.

[47] Xin Lu and Shuang-Hua Yang. Thermal energy harvesting for WSNs. *IEEE International Conference on Systems Man and Cybernetics, SMC 2010*. 3045–3052, IEEE, 2010.

[48] Sahar Ayazian, Vahid A Akhavan, Eric Soenen, and Arjang Hassibi. A photovoltaic-driven and energy-autonomous CMOS implantable sensor. *IEEE Transactions on Biomedical Circuits and Systems*, 6(4):336–343, 2012.

[49] DS Karanjkar, S Chatterji, and Amod Kumar. Design and implementation of a linear quadratic regulator based maximum power point tracker for solar photovoltaic system. *International Journal of Hybrid Information Technology*, 7(1):167–

182, 2014.

[50] S Karak, SK Ray, and A Dhar. Improvement of efficiency in solar cells based on vertically grown copper phthalocyanine nanorods. *Journal of Physics D: Applied Physics*, 43(24):245101–245105, 2010.

[51] Ajit K Katiyar, S Mukherjee, M Zeeshan, Samit K Ray, and AK Raychaudhuri. Enhancement of efficiency of a solar cell fabricated on black Si made by inductively coupled plasma–reactive ion etching process: A case study of a n-CdS/p-Si heterojunction cell. *ACS Applied Materials & Interfaces*, 7(42):23445–23453, 2015.

[52] Suellen CS Costa, Antonia Sonia AC Diniz, and Lawrence L Kazmerski. Solar energy dust and soiling R&D progress: Literature review update for 2016. *Renewable and Sustainable Energy Reviews*, 82(3):2504–2536, 2017.

[53] Paul D Mitcheson, Eric M Yeatman, G Kondala Rao, Andrew S Holmes, and Tim C Green. Energy harvesting from human and machine motion for wireless electronic devices. *IEEE Proceedings*. 96(9):1457–1486, 2008.

[54] Rohit Moghe, Yi Yang, Frank Lambert, and Deepak Divan. A scoping study of electric and magnetic field energy harvesting for wireless sensor networks in power system applications. *IEEE Conference on Energy Conversion Congress and Exposition, ECCE 2009.* 3550–3557, IEEE, 2009.

[55] Shadrach Joseph Roundy. *Energy Scavenging for Wireless Sensor Nodes with a Focus on Vibration to Electricity Conversion*. PhD thesis, University of California, Berkeley, CA, 2003.

[56] Jose Gerardo Rocha, Luis Miguel Goncalves, PF Rocha, Marco P Silva, and Senentxu Lanceros-Mendez. Energy harvesting from piezoelectric materials fully integrated in footwear. *IEEE Transactions on Industrial Electronics*, 57(3):813–819, 2010.

[57] Fei Liu, Alex Phipps, Stephen Horowitz, Khai Ngo, Louis Cattafesta, Toshikazu Nishida, and Mark Sheplak. Acoustic energy harvesting using an electromechanical Helmholtz resonator. *The Journal of the Acoustical Society of America*, 123(4):1983–1990, 2008.

[58] Aman Kansal, Jason Hsu, Sadaf Zahedi, and Mani B Srivastava. Power management in energy harvesting sensor networks. *ACM Transactions on Embedded Computing Systems (TECS)*, 6(4):1–32, 2007.

[59] Cesare Alippi, Giuseppe Anastasi, Mario Di Francesco, and Manuel Roveri. Energy management in wireless sensor networks with energy-hungry sensors. *IEEE Instrumentation & Measurement Magazine*, 12(2):16–23, 2009.

[60] Rich H Inman, Hugo TC Pedro, and Carlos FM Coimbra. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*,

39(6):535–576, 2013.

[61] Raul Banos, Francisco Manzano-Agugliaro, FG Montoya, Consolacion Gil, Alfredo Alcayde, and Julio Gomez. Optimization methods applied to renewable and sustainable energy: A review. *Renewable and Sustainable Energy Reviews*, 15(4):1753–1766, 2011.

[62] Anil Kumar Sahu, Arun M Shandilya, and S Bhardwaj. Forecasting and management of load for rural areas. *International Journal of Scientific and Research Publications*, 3(1):216, 2013.

[63] Jorgen Christensen-Dalsgaard. Physics of solar-like oscillations. *Solar Physics*, 220(2):137–168, 2004.

[64] Ricardo Marquez and Carlos FM Coimbra. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy*, 85(5):746–756, 2011.

[65] MR Lohokare, Swapna Devi, Shyam S Pattnaik, Bijaya K Panigrahi, and Jayant G Joshi. Modified biogeography-based optimisation (MBBO). *International Journal of Bio-Inspired Computation*, 3(4):252–266, 2011.

[66] Jiaming Li, John K Ward, Jingnan Tong, Lyle Collins, and Glenn Platt. Machine learning for solar irradiance forecasting of photovoltaic system. *Renewable Energy*, 90:542–553, 2016.

[67] Nesreen K Ahmed, Amir F Atiya, Neamat El Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5):594–621, 2010.

[68] Francisco Meza and Eduardo Varas. Estimation of mean monthly solar global radiation as a function of temperature. *Agricultural and Forest Meteorology*, 100(2):231–241, 2000.

[69] Gordon Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342–349, 2009.

[70] Tamer Khatib, Azah Mohamed, and Kamaruzzaman Sopian. A review of solar energy modeling techniques. *Renewable and Sustainable Energy Reviews*, 16(5):2864–2869, 2012.

[71] Maimouna Diagne, Mathieu David, Philippe Lauret, John Boland, and Nicolas Schmutz. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27:65–76, 2013.

[72] Amit Kumar Yadav and SS Chandel. Solar radiation prediction using artificial neural network techniques: A review. *Renewable and Sustainable Energy Reviews*, 33:772–781, 2014.

[73] A Kadiyala, A Kumar, and D Kaur. Comparative performance of back propagation and radial basis function networks in predicting indoor air quality in a transit microenvironment. *International Conference on Emerging Technologies for Sustainable Environment*. 262-271, 2010.

[74] Fariba Besharat, Ali A Dehghan, and Ahmad R Faghih. Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, 21:798–821, 2013.

[75] Lunche Wang, Ozgur Kisi, Mohammad Zounemat-Kermani, German Ariel Salazar, Zhongmin Zhu, and Wei Gong. Solar radiation prediction using different techniques: Model evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 61:384–397, 2016.

[76] J Antonanzas, N Osorio, R Escobar, R Urraca, FJ Martinez-de Pison, and F Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar Energy*, 136:78–111, 2016.

[77] DK Chaturvedi. Solar power forecasting: A review. *International Journal of Computer Applications*, 145(6):28–50, 2016.

[78] Sthitapragyan Mohanty, Prashanta K Patra, Sudhansu S Sahoo, and Asit Mohanty. Forecasting of solar energy with application for a growing economy like India: Survey and implication. *Renewable and Sustainable Energy Reviews*, 78:539–553, 2017.

[79] Muhammad Qamar Raza, Mithulananthan Nadarajah, and Chandima Ekanayake. On recent advances in PV output power forecast. *Solar Energy*, 136:125–144, 2016.

[80] Cyril Voyant, Gilles Notton, Soteris Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte, and Alexis Fouilloy. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105:569–582, 2017.

[81] David John Gagne, Amy McGovern, Sue Ellen Haupt, and John K Williams. Evaluation of statistical learning configurations for gridded solar irradiance forecasting. *Solar Energy*, 150:383–393, 2017.

[82] Peter E Thornton and Steven W Running. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology*, 93(4):211–228, 1999.

[83] AA Trabea and MA Mosalam Shaltout. Correlation of global solar radiation with meteorological parameters over Egypt. *Renewable Energy*, 21(2):297–308, 2000.

[84] Jerome C Winslow, E Raymond Hunt, and Stephen C Piper. A globally applicable model of daily solar irradiance estimated from air temperature and precipitation data. *Ecological Modelling*, 143(3):227–243, 2001.

[85] Mossad El-Metwally. Simple new methods to estimate global solar radiation based on meteorological data in Egypt. *Atmospheric Research*, 69(3):217–239, 2004.

[86] Xiaoying Liu, Xurong Mei, Yuzhong Li, Qingsuo Wang, Jens Raunso Jensen, Yanqing Zhang, and John Roy Porter. Evaluation of temperature-based global solar radiation models in China. *Agricultural and Forest Meteorology*, 149(9):1433–1446, 2009.

[87] Fatih Onur Hocaoglu, Omer NezIh Gerek, and Mehmet Kurban. Novel models for hourly solar radiation using a $2D$ approach. *International Journal of Innovative Computing, Information and Control*, 5(4):855–864, 2009.

[88] S Janjai, K Sricharoen, and S Pattarapanitchai. Semi-empirical models for the estimation of clear sky solar global and direct normal irradiances in the tropics. *Applied Energy*, 88(12):4749–4755, 2011.

[89] Pramod K Pandey and Michelle L Soupir. A new method to estimate average hourly global solar radiation on the horizontal surface. *Atmospheric Research*, 114:83–90, 2012.

[90] Kwanho Lee, Hochun Yoo, and Geoff J Levermore. Quality control and estimation hourly solar irradiation on inclined surfaces in South Korea. *Renewable Energy*, 57:190–199, 2013.

[91] Yamin Wang and Lei Wu. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. *Energy*, 112:208–220, 2016.

[92] Hossein Khorasanizadeh, Kasra Mohammadi, and Navid Goudarzi. Prediction of horizontal diffuse solar radiation using clearness index based empirical models; a case study. *International Journal of Hydrogen Energy*, 41(47):21888–21898, 2016.

[93] Tingting Zhu, Haikun Wei, Xin Zhao, Chi Zhang, and Kanjian Zhang. Clear-sky model for wavelet forecast of direct normal irradiance. *Renewable Energy*, 104:1–8, 2017.

[94] Christian Renner and Volker Turau. Adaptive energy-harvest profiling to enhance depletion-safe operation and efficient task scheduling. *Sustainable Computing: Informatics and Systems*, 2(1):43–56, 2012.

[95] Rob J Hyndman, Anne B Koehler, Ralph D Snyder, and Simone Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002.

[96] Joaquin Recas Piorno, Carlo Bergonzini, David Atienza, and Tajana Simunic Rosing. Prediction and management in energy harvested wireless sensor nodes. 1^{st} *International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, Wireless VITAE 2009.* 6–10, IEEE, 2009.

[97] James W Taylor. Short-term load forecasting with exponentially weighted methods.

IEEE Transactions on Power Systems, 27(1):458–464, 2012.

[98] Zibo Dong, Dazhi Yang, Thomas Reindl, and Wilfred M Walsh. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy*, 55:1104–1113, 2013.

[99] Ahmad H Dehwah, Shahrazed Elmetennani, and Christian Claudel. Ud-wcma: An energy estimation and forecast scheme for solar powered wireless sensor networks. *Journal of Network and Computer Applications*, 90:17–25, 2017.

[100] Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Solar Energy*, 83(10):1772–1783, 2009.

[101] Dazhi Yang, Panida Jirutitijaroen, and Wilfred M Walsh. Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy*, 86(12):3531–3543, 2012.

[102] Sancho Salcedo-Sanz, Carlos Casanova-Mateo, Jordi Munoz-Mari, and Gustau Camps-Valls. Prediction of daily global solar irradiation using temporal gaussian processes. *IEEE Geoscience and Remote Sensing Letters*, 11(11):1936–1940, 2014.

[103] Juan R Trapero, Nikolaos Kourentzes, and Alberto Martin. Short-term solar irradiation forecasting based on dynamic harmonic regression. *Energy*, 84:289–295, 2015.

[104] Remco A Verzijlbergh, Petra W Heijnen, Stephan R de Roode, Alexander Los, and Harm JJ Jonker. Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications. *Solar Energy*, 118:634–645, 2015.

[105] M David, F Ramahatana, Pierre-Julien Trombe, and Philippe Lauret. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy*, 133:55–72, 2016.

[106] Huashan Li, Weibin Ma, Yongwang Lian, and Xianlong Wang. Estimating daily global solar radiation by day of year in China. *Applied Energy*, 87(10):3011–3017, 2010.

[107] Socrates Kaplanis, Jatin Kumar, and Eleni Kaplani. On a universal model for the prediction of the daily global solar radiation. *Renewable Energy*, 91:178–188, 2016.

[108] Husamettin Bulut and Orhan Buyukalaca. Simple model for the generation of daily global solar-radiation data in Turkey. *Applied Energy*, 84(5):477–491, 2007.

[109] A Al-Salaymeh. Modelling of global daily solar radiation on horizontal surfaces for Amman city. *Emirates Journal for Engineering Research*, 11(1):49–56, 2006.

[110] Alessandro Cammarano, Chiara Petrioli, and Dora Spenza. Pro-energy: A novel energy prediction model for solar and wind energy-harvesting wireless sensor networks. 9^{th} *IEEE International Conference on Mobile Adhoc and Sensor Systems*

(MASS), 2012 . 75–83, IEEE, 2012.

[111] Fatih Onur Hocaoglu and Fatih Serttas. A novel hybrid (Mycielski-Markov) model for hourly solar radiation forecasting. *Renewable Energy*, 108:635–643, 2017.

[112] Emre Akarslan and Fatih Onur Hocaoglu. A novel method based on similarity for hourly solar irradiance forecasting. *Renewable Energy*, 112:337–346, 2017.

[113] Zong-Chang Yang. Modeling and forecasting monthly movement of annual average solar insolation based on the least-squares Fourier-model. *Energy Conversion and Management*, 81:201–210, 2014.

[114] Amod Kumar, Sneh Anand, and LN Yaddanapudi. Fuzzy model for estimating induction dose for general anesthesia. *Journal of Scientific and Industrial Research*, 65:325–328, 2006.

[115] Emre Akarslan, Fatih Onur Hocaouglu, and Rifat Edizkan. A novel multi-dimensional linear prediction filter approach for hourly solar radiation forecasting. *Energy*, 73:978–986, 2014.

[116] Maimouna Diagne, Mathieu David, John Boland, Nicolas Schmutz, and Philippe Lauret. Post-processing of solar irradiance forecasts from WRF model at Reunion Island. *Solar Energy*, 105:99–108, 2014.

[117] S Kaplanis and E Kaplani. Stochastic prediction of hourly global solar radiation for Patra, Greece. *Applied Energy*, 87(12):3748–3758, 2010.

[118] V Prema and K Uma Rao. Development of statistical time series models for solar power prediction. *Renewable Energy*, 83:100–109, 2015.

[119] Holger R Maier and Graeme C Dandy. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 15(1):101–124, 2000.

[120] Amir F Atiya, Suzan M El-Shoura, Samir I Shaheen, and Mohamed S El-Sherif. A comparison between neural-network forecasting techniques-case study: river flow forecasting. *IEEE Transactions on Neural Networks*, 10(2):402–409, 1999.

[121] Alexander G Parlos, Omar T Rais, and Amir F Atiya. Multi-step-ahead prediction using dynamic recurrent neural networks. *Neural Networks*, 13(7):765–786, 2000.

[122] Y Kemmoku, S Orita, S Nakagawa, and T Sakakibara. Daily insolation forecasting using a multi-stage neural network. *Solar Energy*, 66(3):193–199, 1999.

[123] Yingni Jiang. Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. *Energy Policy*, 36(10):3833–3837, 2008.

[124] Adel Mellit and Alessandro Massi Pavan. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy*, 84(5):807–821, 2010.

[125] Salman Quaiyum, Shahriar Rahman, and Saidur Rahman. Application of artificial neural network in forecasting solar irradiance and sizing of photovoltaic cell for standalone systems in Bangladesh. *International Journal of Computer Applications*, 32:51–56, 2011.

[126] Mehmet Fidan, Fatih Onur Hocaouglu, and Omer Nezih Gerek. Harmonic analysis based hourly solar radiation forecasting model. *IET Renewable Power Generation*, 9(3):218–227, 2014.

[127] Emanuel Federico Alsina, Marco Bortolini, Mauro Gamberi, and Alberto Regattieri. Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy Conversion and Management*, 120:320–329, 2016.

[128] Qingye Ding, Zhida Li, Prerna Batta, and Ljiljana Trajkovic. Detecting BGP anomalies using machine learning techniques. *IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016* . 003352–003355, IEEE, 2016.

[129] L Mazorra Aguiar, B Pereira, Philippe Lauret, F Diaz, and Mathieu David. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*, 97:599–610, 2016.

[130] David JC MacKay. A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3):448–472, 1992.

[131] Cyril Voyant, Gilles Notton, Christophe Darras, Alexis Fouilloy, and Fabrice Motte. Uncertainties in global radiation time series forecasting using machine learning: The multilayer perceptron case. *Energy*, 125:248–257, 2017.

[132] Jing Huang, Małgorzata Korolkiewicz, Manju Agrawal, and John Boland. Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model. *Solar Energy*, 87:136–149, 2013.

[133] David P Larson, Lukas Nonnenmacher, and Carlos FM Coimbra. Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. *Renewable Energy*, 91:11–20, 2016.

[134] Alexander Baklanov, Ulrik Korsholm, Alexander Mahura, Claus Petersen, and Allan Gross. Enviro-hirlam: On-line coupled modelling of urban meteorology and air pollution. *Advances in Science and Research*, 2(1):41–46, 2008.

[135] William C Skamarock. Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132(12):3019–3032, 2004.

[136] Sultan Al-Yahyai, Yassine Charabi, and Adel Gastli. Review of the use of numerical weather prediction (NWP) models for wind energy assessment. *Renewable and Sustainable Energy Reviews*, 14(9):3192–3198, 2010.

[137] William C Skamarock and Joseph B Klemp. A time-split nonhydrostatic atmo-

spheric model for weather research and forecasting applications. *Journal of Computational Physics*, 227(7):3465–3485, 2008.

[138] Jianwu Zeng and Wei Qiao. Short-term solar power prediction using a support vector machine. *Renewable Energy*, 52:118–127, 2013.

[139] Baotan Jiang, Zhibin Pan, Yuehong Qiu, Zhi Chen, Dalei Yao, and Fan Bu. Intra-hour cloud movement detection for solar forecasts based on ground imaging system. *Optik-International Journal for Light and Electron Optics*, 127(19):7803–7808, 2016.

[140] S Belaid and A Mellit. Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. *Energy Conversion and Management*, 118:105–118, 2016.

[141] Alexander Grigorievskiy, Yoan Miche, Anne-Mari Ventela, Eric Severin, and Amaury Lendasse. Long-term time series prediction using OP-ELM. *Neural Networks*, 51:50–56, 2014.

[142] Yujie Wu and Jianzhou Wang. A novel hybrid model based on artificial neural networks for solar radiation prediction. *Renewable Energy*, 89:268–284, 2016.

[143] Ravinesh C Deo, Nathan Downs, Alfio V Parisi, Jan F Adamowski, and John M Quilty. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. *Environmental Research*, 155:141–166, 2017.

[144] Jerome H Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19:1–41, 1991.

[145] J Ross Quinlan. Combining instance-based and model-based learning 10^{th} *International Conference on Machine Learning, 236–243*. 1993.

[146] Yong Wang and Ian H Witten. Inducing model trees for continuous classes 9^{th} *European Conference on Machine Learning, 128–137*. 1997.

[147] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and Regression Trees*. CRC press, 1984.

[148] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[149] A Moghaddamnia, R Remesan, M Hassanpour Kashani, M Mohammadi, D Han, and J Piri. Comparison of LLR, MLP, Elman, NNARX and ANFIS models with a case study in solar radiation estimation. *Journal of Atmospheric and Solar-Terrestrial Physics*, 71(8):975–982, 2009.

[150] Hemant Ishwaran, Udaya B Kogalur, and J Sunil Rao. spikeslab: Prediction and variable selection using spike and slab regression. *R Journal*, 2(2), 2010.

[151] Tong Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE Transactions on Information Theory*, 57(7):4689–4708, 2011.

[152] Yvonne Gala, Angela Fernandez, Julia Diaz, and Jose R Dorronsoro. Hybrid machine learning forecasting of solar radiation values. *Neurocomputing*, 176:48–59, 2016.

[153] Caroline Persson, Peder Bacher, Takahiro Shiga, and Henrik Madsen. Multi-site solar power forecasting using gradient boosted regression trees. *Solar Energy*, 150:423–436, 2017.

[154] A Orfila, JL Ballester, R Oliver, A Alvarez, and J Tintore. Forecasting the solar cycle with genetic algorithms. *Astronomy & Astrophysics*, 386(1):313–318, 2002.

[155] Alex Mesoudi. *Cultural Evolution: How Darwinian theory can explain human culture and synthesize the social sciences*. University of Chicago Press, 2011.

[156] Saeed Jafarzadeh, M Sami Fadali, and Cansın Yaman Evrenosoglu. Solar power prediction using interval type-2 TSK modeling. *IEEE Transactions on Sustainable Energy*, 4(2):333–339, 2013.

[157] Chi Wai Chow, Bryan Urquhart, Matthew Lave, Anthony Dominguez, Jan Kleissl, Janet Shields, and Byron Washom. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, 85(11):2881–2893, 2011.

[158] Ricardo Marquez and Carlos FM Coimbra. Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, 91:327–336, 2013.

[159] Chi Wai Chow, Serge Belongie, and Jan Kleissl. Cloud motion and stability estimation for intra-hour solar forecasting. *Solar Energy*, 115:645–655, 2015.

[160] Joaquın Alonso-Montesinos and Francisco Javier Batlles. Solar radiation forecasting in the short-and medium-term under all sky conditions. *Energy*, 83:387–393, 2015.

[161] He Jiang and Yao Dong. A nonlinear support vector machine model with hard penalty function based on glowworm swarm optimization for forecasting daily global solar radiation. *Energy Conversion and Management*, 126:991–1002, 2016.

[162] Sukhmeet Kaur, M Singh, and Rajeev Agarwal. Vhdl implementation of non restoring division algorithm using high speed adder/subtractor. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 2(7):3317–3324, 2013.

[163] Adrian Grantham, Yulia R Gel, and John Boland. Nonparametric short-term probabilistic forecasting for solar radiation. *Solar Energy*, 133:465–475, 2016.

[164] Remus St Boata and Paul Gravila. Functional fuzzy approach for forecasting daily global solar irradiation. *Atmospheric Research*, 112:79–88, 2012.

[165] Devinder Kaur and Haricharan Pulugurta. Comparative analysis of fuzzy decision tree and logistic regression methods for pavement treatment prediction. *WSEAS*

Transactions on Information Science and Applications, 5(6):979–990, 2008.

[166] Xiyang Yang, Fusheng Yu, and Witold Pedrycz. Long-term forecasting of time series based on linear fuzzy information granules and fuzzy inference system. *International Journal of Approximate Reasoning*, 81:1–27, 2017.

[167] Adel Mellit, M Benghanem, and Soteris A Kalogirou. An adaptive wavelet-network model for forecasting daily total solar-radiation. *Applied Energy*, 83(7):705–722, 2006.

[168] Fatih O Hocaouglu, Omer N Gerek, and Mehmet Kurban. Hourly solar radiation forecasting using optimal coefficient 2D linear filters and feed-forward neural networks. *Solar Energy*, 82(8):714–726, 2008.

[169] Amandeep Singh and Maninder Lal Singh. Automated color prediction of paddy crop leaf using image processing. *IEEE Conference on Technological Innovation in ICT for Agriculture and Rural Development, TIAR 2015*. 24–32, IEEE, 2015.

[170] Cyril Voyant, Marc Muselli, Christophe Paoli, and Marie-Laure Nivet. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy*, 39(1):341–355, 2012.

[171] Paras Mandal, Surya Teja Swarroop Madhira, Julian Meng, and Ricardo L Pineda. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. *Procedia Computer Science*, 12:332–337, 2012.

[172] Elham Sadat Mostafavi, Sara Saeidi Ramiyani, Rahim Sarvar, Hashem Izadi Moud, and Seyyed Mohammad Mousavi. A hybrid computational approach to estimate solar global radiation: an empirical evidence from Iran. *Energy*, 49:204–210, 2013.

[173] Cyril Voyant, Marc Muselli, Christophe Paoli, and Marie-Laure Nivet. Hybrid methodology for hourly global radiation forecasting in Mediterranean area. *Renewable Energy*, 53:1–11, 2013.

[174] SX Chen, HB Gooi, and MQ Wang. Solar radiation forecast based on fuzzy logic and neural networks. *Renewable Energy*, 60:195–201, 2013.

[175] Huan Long, Zijun Zhang, and Yan Su. Analysis of daily solar power prediction with data-driven approaches. *Applied Energy*, 126:29–37, 2014.

[176] Badia Amrouche and Xavier Le Pivert. Artificial neural network based daily local forecasting for global solar radiation. *Applied Energy*, 130:333–341, 2014.

[177] Kuo-Ping Lin and Ping-Feng Pai. Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression. *Journal of Cleaner Production*, 134:456–462, 2016.

[178] Pedro F Jimenez-Perez and Llanos Mora-Lopez. Modeling and forecasting hourly global solar radiation using clustering and classification techniques. *Solar Energy*, 135:682–691, 2016.

[179] Zhaoxuan Li, SM Rahman, Rolando Vega, and Bing Dong. A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting. *Energies*, 9(1):55, 2016.

[180] Vijay Raghunathan, Saurabh Ganeriwal, and Mani Srivastava. Emerging techniques for long lived wireless sensor networks. *IEEE Communications Magazine*, 44(4):108–114, 2006.

[181] James M Gilbert and Farooq Balouchi. Comparison of energy harvesting systems for wireless sensor networks. *International Journal of Automation and Computing*, 5(4):334–347, 2008.

[182] Akhil Kadiyala, Devinder Kaur, and Ashok Kumar. Application of MATLAB to select an optimum performing genetic algorithm for predicting in-vehicle pollutant concentrations. *Environmental Progress & Sustainable Energy*, 29(4):398–405, 2010.

[183] ASM Zahid Kausar, Ahmed Wasif Reza, Mashad Uddin Saleh, and Harikrishnan Ramiah. Energizing wireless sensor networks by energy harvesting systems: Scopes, challenges and approaches. *Renewable and Sustainable Energy Reviews*, 38:973–989, 2014.

[184] Junaid Ahmed Khan, Hassaan Khaliq Qureshi, and Adnan Iqbal. Energy management in wireless sensor networks: A survey. *Computers & Electrical Engineering*, 41:159–176, 2015.

[185] Vikrant Bhatnagar and Philip Owende. Energy harvesting for assistive and mobile applications. *Energy Science & Engineering*, 3(3):153–173, 2015.

[186] Imran Ahmed, M Majid Butt, Constantinos Psomas, Amr Mohamed, Ioannis Krikidis, and Mohsen Guizani. Survey on energy harvesting wireless communications: Challenges and opportunities for radio resource allocation. *Computer Networks*, 88:234–248, 2015.

[187] Aliyu Aliyu Babayo, Mohammad Hossein Anisi, and Ihsan Ali. A review on energy management schemes in energy harvesting wireless sensor networks. *Renewable and Sustainable Energy Reviews*, 76:1176–1184, 2017.

[188] Faisal Karim Shaikh and Sherali Zeadally. Energy harvesting in wireless sensor networks: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 55:1041–1054, 2016.

[189] Manfred Schal. Average optimality in dynamic programming with general state space. *Mathematics of Operations Research*, 18(1):163–172, 1993.

[190] Alvin C Fu, Eytan Modiano, and John N Tsitsiklis. Optimal energy allocation and admission control for communications satellites. *IEEE/ACM Transactions on Networking*, 11(3):488–500, 2003.

[191] Chin Keong Ho, Pham Dang Khoa, and Pang Chin Ming. Markovian models for

harvested energy in wireless communications. *IEEE International Conference on Communication Systems, ICCS 2010*. 311–315, IEEE, 2010.

[192] Chin Keong Ho and Rui Zhang. Optimal energy allocation for wireless communications with energy harvesting constraints. *IEEE Transactions on Signal Processing*, 60(9):4808–4818, 2012.

[193] Clemens Moser, Lothar Thiele, Davide Brunelli, and Luca Benini. Adaptive power management for environmentally powered systems. *IEEE Transactions on Computers*, 59(4):478–491, 2010.

[194] Dong Kun Noh and Kyungtae Kang. Balanced energy allocation scheme for a solar-powered sensor system and its effects on network-wide performance. *Journal of Computer and System Sciences*, 77(5):917–932, 2011.

[195] Andrea J Goldsmith and Pravin P Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, 1997.

[196] Parastoo Sadeghi, Rodney A Kennedy, Predrag B Rapajic, and Ramtin Shams. Finite-state markov modeling of fading channels - a survey of principles and applications. *IEEE Signal Processing Magazine*, 25(5), 2008.

[197] Omur Ozel and Sennur Ulukus. Information-theoretic analysis of an energy harvesting communication system. 21^{st} *International Symposium on Personal, Indoor and Mobile Radio Communications Workshops, PIMRC Workshops 2010*. 330–335, IEEE, 2010.

[198] Konstantinos Dovrolis, Maruthy Prasad, and Vedam Parameswaran Ramanathan. The selection of the token bucket parameters in the IETF guaranteed service class. pages 1–28, 1997.

[199] Joseph Farkas, Brandon Hombs, John Tranquilli, Shaomin Mo, Matthew Sherman, John Gu, and Bruce Fette. Power aware scheduling and power control techniques for multiuser detection enabled wireless mobile ad-hoc networks. *Military Communications Conference, MILCOM 2010*. 110–115, IEEE, 2010.

[200] DARPA. Darpa energy-harvesting projects. http://www.darpa.mil/projects.html.

[201] Soledad Escolar, Stefano Chessa, and Jesus Carretero. Energy management in solar cells powered wireless sensor networks for quality of service optimization. *Personal and Ubiquitous Computing*, 18(2):449–464, 2014.

[202] Konstantinos Kalpakis, Koustuv Dasgupta, and Parag Namjoshi. Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. *Computer Networks*, 42(6):697–716, 2003.

[203] Sushruta Mishra and Hiren Thakkar. Features of WSN and data aggregation techniques in WSN: A survey. *International Journal of Engineering and Innovative*

Technology(IJEIT), 1(4):264–273, 2012.

[204] Seung Jun Baek, Gustavo De Veciana, and Xun Su. Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation. *IEEE Journal on Selected Areas in Communications*, 22(6):1130–1140, 2004.

[205] Pavan Nuggehalli, Vikram Srinivasan, and Ramesh R Rao. Energy efficient transmission scheduling for delay constrained wireless networks. *IEEE Transactions on Wireless Communications*, 5(3):531–539, 2006.

[206] Alireza Seyedi and Biplab Sikdar. Energy efficient transmission strategies for body sensor networks with energy harvesting. *IEEE Transactions on Communications*, 58(7):2116–2126, 2010.

[207] Mehmet Akif Antepli, Elif Uysal-Biyikoglu, and Hakan Erkal. Optimal packet scheduling on an energy harvesting broadcast link. *IEEE Journal on Selected Areas in Communications*, 29(8):1721–1731, 2011.

[208] Elif Uysal-Biyikoglu and Abbas El Gamal. On adaptive transmission for energy efficiency in wireless data networks. *IEEE Transactions on Information Theory*, 50(12):3081–3094, 2004.

[209] Omur Ozel, Kaya Tutuncuoglu, Jing Yang, Sennur Ulukus, and Aylin Yener. Transmission with energy harvesting nodes in fading wireless channels: Optimal policies. *IEEE Journal on Selected Areas in Communications*, 29(8):1732–1743, 2011.

[210] Paolo Castiglione, Osvaldo Simeone, Elza Erkip, and Thomas Zemen. Energy management policies for energy-neutral source-channel coding. *IEEE Transactions on Communications*, 60(9):2668–2678, 2012.

[211] Srinivas Reddy and Chandra R Murthy. Dual-stage power management algorithms for energy harvesting sensors. *IEEE Transactions on Wireless Communications*, 11(4):1434–1445, 2012.

[212] Jing Yang and Sennur Ulukus. Optimal packet scheduling in an energy harvesting communication system. *IEEE Transactions on Communications*, 60(1):220–230, 2012.

[213] Kaya Tutuncuoglu and Aylin Yener. Optimum transmission policies for battery limited energy harvesting nodes. *IEEE Transactions on Wireless Communications*, 11(3):1180–1189, 2012.

[214] Shuguang Cui, Andrea J Goldsmith, and Ahmad Bahai. Energy-constrained modulation optimization. *IEEE Transactions on Wireless Communications*, 4(5):2349–2360, 2005.

[215] Rene L Cruz. A calculus for network delay. I. network elements in isolation. *IEEE Transactions on Information Theory*, 37(1):114–131, 1991.

[216] Ya Xu, John Heidemann, and Deborah Estrin. Geography-informed energy conservation for Ad hoc routing. 7^{th} *Annual International Conference on Mobile Computing and Networking, 2001.* 70–84, ACM, 2001.

[217] Charles Perkins, Elizabeth Belding-Royer, and Samir Das. Ad hoc on-demand distance vector (AODV) routing. 2^{nd} *Workshop on Mobile Computing Systems and Applications, WMCSA'99.* Technical report, 1–37, 2003.

[218] Rakesh Kumar and KV Arya. A modified approach for route maintenance using alternate path in AODV. *International Conference on Computational Intelligence and Communication Networks, CICN 2011.* 37–41, IEEE, 2011.

[219] David B Johnson, David A Maltz, Josh Broch, et al. Dsr: The dynamic source routing protocol for multi-hop wireless Ad hoc networks. *Ad hoc Networking*, 5:139–172, 2001.

[220] Mohamed Younis, Moustafa Youssef, and Khaled Arisha. Energy-aware routing in cluster-based sensor networks. 10^{th} *IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems, MASCOTS 2002.* 129–136, IEEE, 2002.

[221] Tripti Sharma, GS Tomar, Radhika Gandhi, Srishti Taneja, and Kiran Agrawal. Optimized genetic algorithm (OGA) for homogeneous WSNs. *International Journal of Future Generation Communication and Networking*, 8(4):131–140, 2015.

[222] Jae-Hwan Chang and Leandros Tassiulas. Maximum lifetime routing in wireless sensor networks. *IEEE/ACM Transactions on Networking*, 12(4):609–619, 2004.

[223] Sanghita Bhattacharjee and Subhansu Bandyopadhyay. Lifetime maximizing dynamic energy efficient routing protocol for multi hop wireless networks. *Simulation Modelling Practice and Theory*, 32:15–29, 2013.

[224] Dijun Luo, Xiaojun Zhu, Xiaobing Wu, and Guihai Chen. Maximizing lifetime for the shortest path aggregation tree in wireless sensor networks. *IEEE International Conference on Computer Communications (INFOCOM), 2011* . 1566–1574, IEEE, 2011.

[225] Huseyin Ozgur Tan and Ibrahim Korpeolu. Power efficient data gathering and aggregation in wireless sensor networks. *ACM Sigmod Record*, 32(4):66–71, 2003.

[226] Chang-Soo Ok, Seokcheon Lee, Prasenjit Mitra, and Soundar Kumara. Distributed energy balanced routing for wireless sensor networks. *Computers & Industrial Engineering*, 57(1):125–135, 2009.

[227] Thiemo Voigt, Hartmut Ritter, and Jochen Schiller. Utilizing solar power in wireless sensor networks. 28^{th} *Annual IEEE International Conference on Local Computer Networks, 2003. LCN'03, 2003.* 416–422, IEEE, 2003.

[228] Vinod Sharma, Utpal Mukherji, Vinay Joseph, and Shrey Gupta. Optimal energy

management policies for energy harvesting sensor nodes. *IEEE Transactions on Wireless Communications*, 9(4):1326–1336, 2010.

[229] Manish Bhardwaj, Timothy Garnett, and Anantha P Chandrakasan. Upper bounds on the lifetime of sensor networks. *IEEE International Conference on Communications, ICC 2001.* 785–790, IEEE, 2001.

[230] Manish Bhardwaj and Anantha P Chandrakasan. Bounding the lifetime of sensor networks via optimal role assignments. 21^{st} *Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2002.* 1587–1596, IEEE, 2002.

[231] Mohammed Rahimi, Hardik Shah, Gaurav S Sukhatme, John Heideman, and Deborah Estrin. Studying the feasibility of energy harvesting in a mobile sensor network. *IEEE International Conference on Robotics and Automation, ICRA'03.* volume 1. 19–24, IEEE, 2003.

[232] Arvind Giridhar and PR Kumar. Maximizing the functional lifetime of sensor networks. 4^{th} *International Symposium on Information Processing in Sensor Networks, IPSN 2005* . 5–12, IEEE, 2005.

[233] Aman Kansal, Aditya Ramamoorthy, Mani B Srivastava, and Gregory J Pottie. On sensor network lifetime and data distortion. *International Symposium on Information Theory, ISIT 2005.* 6–10, IEEE, 2005.

[234] Jaein Jeong and David Culler. A practical theory of micro-solar power sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 9(1):9, 2012.

[235] C-F Chiasserini and Ramesh R Rao. Improving battery performance by using traffic shaping techniques. *IEEE Journal on Selected Areas in Communications*, 19(7):1385–1394, 2001.

[236] S Sanajaoba Singh and Eugene Fernandez. Method for evaluating battery size based on loss of load probability concept for a remote PV system. 6^{th} *IEEE Conference on Power India International Conference, PIICON 2014.* 2014.

[237] Yonggang Zhao, Ramesh Govindan, and Deborah Estrin. Residual energy scans for monitoring wireless sensor networks. *IEEE Conference on Wireless Communication and Networking, WCNC 2002.* 2002.

[238] Xiaofan Jiang, Joseph Polastre, and David Culler. Perpetual environmentally powered sensor networks. 4^{th} *International Symposium on Information Processing in Sensor Networks, 2005.* 65–70, IEEE, 2005.

[239] JM Mora-Merchan, DF Larios, Julio Barbancho, Francisco Javier Molina, Jose Luis Sevillano, and Carlos Leon. mtossim: A simulator that estimates battery lifetime in wireless sensor networks. *Simulation Modelling Practice and Theory*, 31:39–51, 2013.

[240] Aman Kansal, Dunny Potter, and Mani B Srivastava. Performance aware tasking for environmentally powered sensor networks. *ACM SIGMETRICS Performance Evaluation Review*, 32(1):223–234, 2004.

[241] Vijay Raghunathan, Aman Kansal, Jason Hsu, Jonathan Friedman, and Mani Srivastava. Design considerations for solar energy harvesting wireless embedded systems. 4^{th} *International Symposium on Information Processing in Sensor Networks, 2005* . 457–462, IEEE, 2005.

[242] Dusit Niyato, Ekram Hossain, and Afshin Fallahi. Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: Performance analysis and optimization. *IEEE Transactions on Mobile Computing*, 6(2):221–236, 2007.

[243] Christopher M Vigorito, Deepak Ganesan, and Andrew G Barto. Adaptive control of duty cycling in energy-harvesting wireless sensor networks. 4^{th} *Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON'07.* 21–30, IEEE, 2007.

[244] Yu Gu, Ting Zhu, and Tian He. Esc: Energy synchronized communication in sustainable sensor networks. 17^{th} *IEEE International Conference on Network Protocols, ICNP 2009.* 52–62, IEEE, 2009.

[245] Heejung Byun and Junglok Yu. Adaptive duty cycle control with queue management in wireless sensor networks. *IEEE Transactions on Mobile Computing*, 12(6):1214–1224, 2013.

[246] Neeraj Jaggi, Sreenivas Madakasira, Sandeep Reddy Mereddy, and Ravi Pendse. Adaptive algorithms for sensor activation in renewable energy based sensor systems. *Ad hoc Networks*, 11(4):1405–1420, 2013.

[247] Navid Tadayon, Sasan Khoshroo, Elaheh Askari, Honggang Wang, and Howard Michel. Power management in SMAC-based energy-harvesting wireless sensor networks using queuing analysis. *Journal of Network and Computer Applications*, 36(3):1008–1017, 2013.

[248] Alvin C Valera, Wee-Seng Soh, and Hwee-Pink Tan. Energy-neutral scheduling and forwarding in environmentally-powered wireless sensor networks. *Ad hoc Networks*, 11(3):1202–1220, 2013.

[249] S Peng and CP Low. Prediction free energy neutral power management for energy harvesting wireless sensor nodes. *Ad hoc Networks*, 13:351–367, 2014.

[250] Alvin C Valera, Wee-Seng Soh, and Hwee-Pink Tan. Enabling sustainable bulk transfer in environmentally-powered wireless sensor networks. *Ad hoc Networks*, 54:85–98, 2017.

[251] Sukumar Mishra and Dushyant Sharma. Control of Photovoltaic Technology. *Electric Renewable Energy Systems*, 2015.

[252] Olga Moraes Toledo, Delly Oliveira Filho, and Antonia Sonia Alves Cardoso Diniz. Distributed photovoltaic generation and energy storage systems: A review. *Renewable and Sustainable Energy Reviews*, 14(1):506–511, 2010.

[253] Maria Roo Ons, S Shynu, Max Ammann, Sarah McCormack, and Brian Norton. Novel techniques for the integration of antennas and photovoltaic cells. *Royal Irish Academy Research Colloquium on Wireless as an Enabling Technology: Innovation for a Critical Infrastructur*. 2010.

[254] Manus Kennedy, SJ McCormack, John Doran, and Brian Norton. Ray-trace modelling of reflectors for quantum dot solar concentrators. *SPIE Optics and Photonics and Solar Energy Conference*. International Society for Optics and Photonics, 6649, 664905–664908, 2007.

[255] Gilbert M Masters. *Renewable and Efficient Electric Power Systems*. John Wiley & Sons, 2013.

[256] Navid Haghdadi, Jessie Copper, Anna Bruce, and Iain MacGill. A method to estimate the location and orientation of distributed photovoltaic systems from their generation output data. *Renewable Energy*, 108:390–400, 2017.

[257] James Hazelton, Anna Bruce, and Iain MacGill. A review of the potential benefits and risks of photovoltaic hybrid mini-grid systems. *Renewable energy*, 67:222–229, 2014.

[258] Sebastian Oliva, Iain MacGill, and Rob Passey. Estimating the financial costs and benefits of distributed grid-connected photovoltaics for different electricity industry participants. 2013.

[259] Sebastin Oliva, Iain MacGill, and Rob Passey. Estimating the net societal value of distributed household PV systems. *Solar Energy*, 100:9–22, 2014.

[260] Jan Kleissl. *Solar Energy Forecasting and Resource Assessment*. Academic Press, 2013.

[261] Tomas Markvart. *Solar Electricity*. John Wiley & Sons, 2000.

[262] Aden B Meinel and Marjorie P Meinel. Applied solar energy: An introduction. *NASA STI/Recon Technical Report A*, 77, 1977.

[263] NREL [online]. Available. www.nrel.gov.

[264] Pyranoometer [online]. Available. www.kippzonen.com.

[265] Lise Getoor and Ben Taskar. *Introduction to Statistical Relational Learning*. MIT press, 2007.

[266] Bidisha Ghosh, Biswajit Basu, and Margaret Omahony. Multivariate short-term traffic flow forecasting using time-series analysis. *IEEE Transactions on Intelligent Transportation Systems*, 10(2):246–254, 2009.

[267] Bidisha Ghosh, Biswajit Basu, and Margaret OMahony. Bayesian time-series model

for short-term traffic flow forecasting. *Journal of Transportation Engineering*, 133(3):180–189, 2007.

[268] George Adomian. *Solving Frontier Problems of Physics: The Decomposition Method*. 60, Springer Science & Business Media, 2013.

[269] Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman. *Forecasting Methods and Applications*. John Wiley & Sons, 2008.

[270] Douglas C Montgomery, Lynwood A Johnson, and John S Gardiner. *Forecasting and Time Series Analysis*. McGraw-Hill Companies, 1990.

[271] Aumnad Phdungsilp. Futures studies backcasting method used for strategic sustainable city planning. *Futures*, 43(7):707–714, 2011.

[272] Chris Chatfield. *The Analysis of Time Series: An Introduction*. CRC press, 2016.

[273] George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. Stochastic models and their forecasting. *Time Series Analysis, Fourth Edition*, 2008.

[274] Baki Billah, Maxwell L King, Ralph D Snyder, and Anne B Koehler. Exponential smoothing model selection for forecasting. *International Journal of Forecasting*, 22(2):239–247, 2006.

[275] James W Taylor. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805, 2003.

[276] Sarah Gelper, Roland Fried, and Christophe Croux. Robust forecasting with exponential and holt–winters smoothing. *Journal of Forecasting*, 29(3):285–300, 2010.

[277] Tong Zhang. Adaptive forward-backward greedy algorithm for sparse learning with linear models. *Advances in Neural Information Processing Systems, NIPS 2008*. 2009.

[278] Alan Miller. *Subset Selection in Regression*. CRC Press, 2002.

[279] Kurt Hornik, Friedrich Leisch, and Achim Zeileis. Jags: A program for analysis of Bayesian graphical models using Gibbs sampling. 3^{rd} *International Workshop on Distributed Statistical Computing, DSC 2003*. 1–10, 2003.

[280] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least Angle Regression. *The Annals of Statistics*, 32(2):407–499, 2004.

[281] J Ross Quinlan. *C4. 5: Programs for Machine Learning*. Elsevier, 2014.

[282] Michael S Balshi, A DAVID Mcguire, Paul Duffy, Mike Flannigan, John Walsh, and Jerry Melillo. Assessing the response of area burned to changing climate in western boreal North America using a multivariate adaptive regression splines (MARS) approach. *Global Change Biology*, 15(3):578–600, 2009.

[283] D Steinberg, PL Colla, and Kerry Martin. Mars User Guide. *San Diego, CA: Salford Systems*, 1999.

[284] Christian Renner, Volker Turau, and Kay Romer. Online energy assessment with supercapacitors and energy harvesters. *Sustainable Computing: Informatics and Systems*, 4(1):10–23, 2014.