DATA ANALYTICS-BASED ADAPTIVE FORECASTING FRAMEWORK FOR
SMART ENERGY MANAGEMENT IN ELECTRICAL MICROGRIDS

APPROVED BY SUPERVISING COMMITTEE:

__________________________
Mo Jamshidi, Ph.D., Chair

__________________________
Rolando E. Vega, Ph.D.

__________________________
Mehdi Shadaram, Ph.D.

__________________________
Hariharan Krishnaswami Ph.D.

__________________________
Les E. Shephard, Ph.D.

Accepted: ____________________________
Dean, Graduate School
DEDICATION

This is a sample dedication text.
DATA ANALYTICS-BASED ADAPTIVE FORECASTING FRAMEWORK FOR
SMART ENERGY MANAGEMENT IN ELECTRICAL MICROGRIDS

by

YASHAR SAHRAEI MANJILI, M.Sc.

DISSERTATION
Presented to the Graduate Faculty of
The University of Texas at San Antonio
In Partial Fulfillment
Of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT SAN ANTONIO
College of Engineering
Department of Electrical and Computer Engineering
December 2014
ACKNOWLEDGEMENTS

This is a sample acknowledgement text.
DATA ANALYTICS-BASED ADAPTIVE FORECASTING FRAMEWORK FOR SMART ENERGY MANAGEMENT IN ELECTRICAL MICROGRIDS

Yashar Sahraei Manjili, Ph.D.
The University of Texas at San Antonio, 2014

Supervising Professors:
Mo Jamshidi, Ph.D. (Chair)
Rolando E. Vega, Ph.D. (Co-Chair)

This research establishes a framework to quantify and integrate the effects of renewable energy intermittency, energy market, policy decisions, and environmental effects into the grid. Novel methodologies are developed for autonomous energy management in electrical microgrids and for solar energy forecasting. Standard performance measures are considered based on the latest benchmarking references and performances of the proposed frameworks are discussed and compared to those of the current state-of-the-art. Potential benefit that the proposed microgrid energy flow control scheme brings to the local microgrid owner and the solar forecasting’s added value to the utility companies and load serving entities (LSE) are presented. Microgrid energy management is implemented using an evolutionary qualitative decision-making scheme based on Genetic-Fuzzy System (GFS) which combines the Fuzzy logic with Genetic Algorithms (GA) to find solutions to the multiple-objective optimization problem. Results are presented for various microgrid structures and control scenarios and it is demonstrated how control of storage energy flow in electrical microgrids can help improve efficiency of demand-supply management, reduce electricity pay rates, as well as reduce quantifiable amounts of air pollution. Microgrid owners looking into adopting a smart decision-making tool for battery management systems (BMS) can use the analytics developed in this work to see a return on investment (ROI) between 5 and 10.

In the future, distribution and transmission grids across the world will incorporate vast amounts of solar energy generation (or other renewable sources) to minimize the amount of net
load to be provided from conventional sources such as coal, petroleum and natural gas. As a result, a data analytics-based adaptive forecasting framework is developed combining data mining, statistical analysis and artificial intelligence. Application of the proposed system for day-ahead prediction of solar energy is considered. Models are developed, tested and verified utilizing a large dataset from the National Renewable Energy Laboratory (NREL) archive, the Automated Surface Observing System (ASOS), and the NREL solar position and intensity calculator (i.e. NREL-SOLPOS) sampled at 1-minute intervals during 8 years (2005 – 2012) at NREL site in Golden, Colorado, USA. A uniqueness of the developed framework is that an integrated serial time-domain analysis coupled with multivariate analysis was used for pre-processing to enhance the recorded data. The resulting enhanced dataset is used for adaptive training of the neural networks forecast engine. Standard performance measures are obtained. The forecast results are compared to those of the standard persistence approach and the state-of-the-art on solar energy forecasting methodologies. The proposed day-ahead solar energy forecasting framework brings an added value of 9.5% to 14.5% to the utility companies compared to a common persistence forecasting.

The methodology is now ready to be deployed in San Antonio, Texas with data collected in the 41 MW Alamo 1 solar PV plant, the largest in Texas, which is the first solar plant in Texas that is connected to the transmission grid allowing solar energy bidding into the market.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. iv
ABSTRACT ........................................................................................................................................ v
LIST OF TABLES ............................................................................................................................... ix
LIST OF FIGURES ............................................................................................................................ x
CHAPTER 1:  INTRODUCTION .......................................................................................................... 1
  1.1 Historical Perspective .................................................................................................................. 1
  1.2 Motivation ................................................................................................................................... 2
  1.3 Objectives and Hypothesis .......................................................................................................... 7
  1.4 Scope, Limitations, Methodology and Multidisciplinary Aspects of Research ......................... 11
  1.5 Intellectual Merit of Research and Contributions ....................................................................... 14
CHAPTER 2:  ELECTRICITY; SOURCES, MARKET, ENVIRONMENT AND GRID ............................ 17
  2.1 Electrical Energy Sources ......................................................................................................... 17
  2.2 Conventional (Non-renewable) Methods ..................................................................................... 21
  2.3 Sustainable (Renewable) Methods ............................................................................................... 22
  2.4 Environmental Protection Agency (EPA) and Air Pollution Standards ..................................... 28
  2.5 Smart Grid and Microgrids ........................................................................................................ 32
  2.6 Summary ..................................................................................................................................... 34
CHAPTER 3:  MICROGRID CONTROL AND ENERGY MANAGEMENT ............................................. 35
  3.1 About Microgrid Energy Flow Control ....................................................................................... 35
  3.2 Microgrid Model and Energy Management Scenarios ................................................................. 37
     3.2.1 Characteristics of Microgrid Buses in Scenario 1 .................................................................. 40
     3.2.2 Characteristics of Microgrid Buses in Scenario 2 .................................................................. 41
     3.2.3 Characteristics of Microgrid Buses in Scenario 3 .................................................................. 41
     3.2.4 Characteristics of Microgrid Buses in Scenario 4 .................................................................. 42
     3.2.5 Characteristics of Microgrid Buses in Scenario 5 .................................................................. 42
     3.2.6 Characteristics of Microgrid Buses in Scenario 4 .................................................................. 43
  3.3 Problem Statement ....................................................................................................................... 44
  3.4 Decision-Making Platform for Microgrid Energy Management ................................................ 48
  3.5 Fuzzy Logic-based Decision-Making for Microgrid Energy Management ................................. 55
LIST OF TABLES

Table 1 World Annual Electricity Generation from Solar [HDW 2013] .................................................. 23
Table 2 Input-output Configuration of Energy Management Scenarios ...................................................... 38
Table 3 Rule-base for Scenario 4 .............................................................................................................. 52
Table 4 Weekly Balance Values Obtained by the Six Scenarios for the Numerical Example .... 71
Table 5 Characteristics of current practices in solar energy forecast [Bing 2012] ...................... 76
Table 6: Performances of simulated approaches and persistence for next day forecast .......... 125
Table 7: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using German Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012) ... 129
Table 8: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using Swiss Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012) .... 130
Table 9: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using Austrian Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012) ... 132
Table 10: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using Spanish Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012) ... 133
Table 11 Performances of simulated approaches and persistence for day-ahead forecast ....... 136
Table 12: Actual and Relative Performance Measures for Day-Ahead Forecast using Austrian Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012). ................................................................. 139
Table 14 Summary of Market / Forecast Outcomes and Restrictions ......................................... 145
LIST OF FIGURES

Figure 1 World Energy Demand – Long term Energy Sources [Orr 2005] ........................................ 17
Figure 2 Global Energy Sources by Alternative World Energy Outlook [LBST AWEO 2006]............... 18
Figure 3 World Electricity Generation by Country [CIATWF Elec. 2012]........................................ 19
Figure 4 US Energy Consumption by Source; From 1800 to 2010 [EIA Energy 2010] ............... 20
Figure 5 World Solar Capacity; (a) Overall Capacity between 1995 and 2009 (b) Cumulative Capacity by Country in 2009 ................................................................. 24
Figure 6 US PV Market Forecast; (a) Overall US PV Capacity (b) Residential PV Capacity by States [Baca 2013]................................................................................................. 24
Figure 7 US PV Market between 2000 and 2012; a. Market Cost b. Market Growth [Baca 2013].............................................................................................................................. 25
Figure 8 Renewable Energy Market Forecast by US EIA [EIA Energy 2013].............................. 26
Figure 9 DOE-EERE Budget Allocation from 2000 to 2014 [Stevens 2013]............................... 27
Figure 10 Electricity Consumption in 51 States of the US in 2006 [EIA Elec. 2006] (a) Overall power in GigaWatts (b) Power per capita in kiloWatts......................................................... 29
Figure 11 US Carbon Dioxide Emission by State [EIA Emission 2013]........................................ 31
Figure 12 Schematic of a Smart Grid Consisting of a Number of Local Microgrids ............... 33
Figure 13 Schematic of a Generic Electrical Micro-grid .............................................................. 38
Figure 14 Three Bus Micro-grid Model ....................................................................................... 47
Figure 15 Structure of the real-time intelligent fuzzy decision-making approach; four input variables and one output variable.................................................................................. 50
Figure 16 Fuzzy Membership functions for input and output variables of the Fuzzy Controller; (a) price, load and generation (b) air pollution (c) output ......................................................... 51
Figure 17 Profiles of Price, Renewables, Load, and Pollution; First day ........................................ 56
Figure 18 Profiles of Electricity Price, Renewable Generation, Local Load, and Air Pollution; One Week .................................................................................................................... 57
Figure 19 Power Exchange with the Main Grid in p.u. value; First Day; Six Scenarios ............ 57
Figure 20 Normalized Micro-grid Contribution to Air Pollution; One Week; Six Scenarios...... 59
Figure 21 Normalized Battery Unit’s Stored Energy; First Day; Four Scenarios ....................... 59
Figure 22 Normalized Battery Power Flow; First Day; Four Scenarios ...................................... 60
Figure 41: Neural Networks Training and Test Process for 1 Day-Ahead Solar Energy Forecast

Figure 42: One Day-Ahead Forecast Performance versus Training Window Length; (a) Non-adaptive approach without ICA (b) Non-adaptive approach with ICA Correlation Coefficient (Solid Blue), RMSE (Dashed Green)

Figure 43: Dynamic User Interface Display; Forecast (dashed red), Target (Filled blue), Actual Error (filled red), Percent Error (Filled Pink), Average Forecast (Red Bar), Average Target (Blue Bar)

Figure 44: Main User Interface Display representing forecast results based on Monte Carlo uncertainty analysis approach for current day and the next day; forecast boundaries (bluish area), Forecast $p_{50}$ (magenta), Forecast $p_{90}$ (green), table represents the hourly average values of irradiance in Wm$^2$ for the next day for the three different percentiles of $p_{50}, p_{75}$ and $p_{90}$

Figure 45: Main User Interface Display representing pdf’s of forecast results based on Monte Carlo uncertainty from a different view angle

Figure 46: User Interface Display representing 1-Day ahead and 2-Days ahead forecast of the day before and current day percentiles based on Monte Carlo uncertainty analysis approach for performance tracking and evaluation; Target (blue), Forecast $p_{50}$ (magenta), Forecast $p_{90}$ (green), Current-Time (Black dot)

Figure 47: GHI profile for a single day (RED), and the accumulated Solar Irradiance (BLUE)

Figure 48: Hourly forecast and target irradiance for days 222 to 232 of year 2012

Figure 49: Next Day Solar Energy Forecasting Performance (a) Scatter Plot; Persistence, (b) Scatter Plot; Proposed Method, (c) Actual Error Histogram; Persistence, (d) Actual Error Histogram; Proposed Method, (e) Percent Error Histogram; Persistence, (f) Percent Error Histogram; Proposed Method

Figure 50: Monthly performance analysis for 2012; Actual Error (TOP Diagram) and Relative Error (BOTTOM Diagram) for Next Day Persistence (BLACK) and Proposed Method (BLUE)

Figure 51: Day-Ahead Solar Energy Forecasting Performance (a) Scatter Plot; Persistence, (b) Scatter Plot; Proposed Method, (c) Actual Error Histogram; Persistence, (d) Actual Error Histogram; Proposed Method, (e) Percent Error Histogram; Persistence, (f) Percent Error Histogram; Proposed Method.
Figure 52: Monthly performance analysis for 2012 Day-Ahead Solar Power Forecast of The Blue Wing Solar Plant; Actual Error (TOP Diagram), Relative Error (BOTTOM Diagram), Persistence (BLACK), Proposed Method (BLUE) ................................................................. 138
Figure 53 Instantaneous ERCOT Electricity Price for the Year 2012; Day-Ahead Market (DAM) and Real-Time Market (RTM) Locational Marginal Price (LMP) ................................................. 143
Figure 54 Annual Solar Energy Sales Revenue Obtained by the Proposed Forecast and the Persistence for the 14 MW Blue Wing Solar Plant in ERCOT Electricity Market versus Deviation Penalty Factor ................................................................. 149
Figure 55 Annual Revenue Relative to Perfect Forecast and the RIF Obtained by the Proposed Forecasting Method versus DPF for the Blue Wing Solar Plant during the Year 2012 .......... 151
Figure 56 Annual Revenue in terms of the Perfect Forecast and the Relative RIF for the Proposed Method at the ERCOT CPS Node During Year 2012 versus PV Plant Capacity and the DPF . 153
Figure 57: Angles describing the position of the sun: \( \theta \) - zenith angle; \( h \) - elevation angle; \( \mu_s \) – azimuth angle. Angles describing the position of the surface: \( \beta \) - slope or tilt angle; \( \mu \) - surface azimuth angle. The incidence angle \( \theta \) is the angle between the sun direction and the surface’s normal \( n \) [27]. ............................................................................................................. 159
CHAPTER 1: INTRODUCTION

1.1 Historical Perspective

Sun as the predominant source of energy in the nature, federal policies promoting and supporting clean energy industries, attraction of major companies toward sustainable energy and clean electricity generation, plus the decreasing cost of solar panels, whose average price dropped 50 percent in 2012 compared to 2011, according to the Solar Energy Industries Association (SEIA) [1], are some of the most important factors contributing to the rise in solar energy adoption into the grid. Besides, new financing methods that allow owners to lease systems long term will help them cut their current electricity costs with little or no upfront investment [2]. In 2011, about 1,855 megawatts of new photovoltaic capacity was installed in the US, more than double the 887 megawatts of the year before [1].

which are still in relatively early stages of development and yet need to become commercially available.

Extensive amount of work is done on energy management in microgrids and smart grid networks to address specific issues. Auction-based theory for pricing strategy in solar powered microgrid is studied in [Maity 2010]. In [Manjili WAC 2012], Authors considered Fuzzy decision-making to control battery storage unit in microgrid assuming different cases of infinite capacity and limited capacity storage, and calculated the overall costs and profits brought to the system in each case. In some most recent research, intelligent control approaches are proposed for battery storage unit control for energy management in microgrid systems using agent-based control algorithms [Yoo 2013].

This dissertation proposes evolutionary qualitative decision-making framework for energy management in electrical microgrids together with a novel data analytics-based adaptive forecasting scheme based on data mining, statistical analysis and artificial intelligence. The benefits and drawbacks of different energy management scenarios for adopted in this study will be addressed considering the electricity trade profit, transmission network loss, and carbon emission control. The proposed forecasting methodology will be deployed for day-ahead prediction of expected solar energy and its performance will be compared against the current state-of-the-art in the solar forecasting in Europe (e.g. NWP and Statistical methods) and to that of the persistence forecast, as the reference case.

1.2 Motivation

The traditional electrical grid is generally used to carry power from a few central generators to a large number of users or customers. In contrast, the new emerging smart grid uses two-way flows of electricity and information to create an automated and distributed advanced energy
delivery network which can be both more reliable and smarter, by means of artificial intelligence-based algorithms, than the older versions. Despite the novel institutional arrangements and network designs of the electrical grid, its power delivery infrastructures suffer aging power equipment, obsolete system layout, outdated engineering and lack and flexibility across the developed and dynamically changing world [Willis 2001].

Fast-paced increasing electricity demand in recent years, problems with already existing old electricity grid, information technology growth and huge data flow requirements, the need for a more flexible and more reliable electric grid are only some of the most vital factors in moving towards smart grids. Such a smart transmission and distribution network not only serves the consumers more reliably, but also helps save natural environment, more specifically the air and the ozone layer, by modifying and controlling its contribution to the air pollution through appropriate demand-response algorithms. Last but not least, smart grids are capable of increasing the financial benefits of either the Utility Companies or the microgrid owners and consumers by deploying appropriate management and decision-making approaches. In recent years many research projects have been conducted to explore the concept of smart grid. Most of these researches are mainly focused on three systems in smart grid- the infrastructure system, the management system, and the protection system [Fang 2011].

United States is the 4th leading country in the world with regard to solar energy capacity, after Germany, Spain, and Japan, with more than 3GW installed solar PV plants operating, and is expected to get close to 9 GW by 2016 [Baca 2013]. Obama administration has been positive towards solar and wind energy and has legislated policies supporting clean and sustainable energy industry including partial recoup of development costs as a cash grant, cost rebates and tax credit. Despite having lost the program that allowed developers to recoup 30 percent of their costs as a
cash grant, the solar industry is still eligible through 2016 for a tax credit to be taken over five
years, making its future seem in some ways more solid than that of the wind power industry for
which the tax credit that has been in place since 1992 is already expired at the end of 2012. The
mentioned factors lead to more than 3 GW solar energy capacity installed in the US during 2012.
The average solar installation cost has also dropped more than 27% from 2011 to 2012 and is now
less than $4 per Watt [Baca 2013].

Texas leads the nation in overall annual electricity generation. Texas is very rich in non-
hydroelectric renewable energy potential. Texas leads the nation on wind capacity and is among
the top 10 states with regard to solar energy installed capacity [Baca 2013]. High levels of direct
solar radiation are suitable to support large-scale solar power plants concentrated in West Texas.
Due to its large population and an energy-intensive economy, Texas also leads the Nation in energy
consumption, accounting for more than one-tenth of total U.S. energy use. Energy-intensive
industries in Texas include aluminum, chemicals, forest products, glass, and petroleum refining.
Texas has the 6th place in the US regarding per capita energy consumption and is the 5th most
expensive state in the US with respect to energy consumption expenditures [CPS 2013].

San Antonio is the 7th most populated city in the United States [USCB 2012]. It is the 7th
largest city in the US too [USCB 2010]. CPS Energy of San Antonio, Texas (formerly "City Public
Service") is the United States' largest municipally owned utility company, with combined electric
and natural gas service. Fourteen percent of all utility revenues are returned to the City of San
Antonio, and those revenues make up more than 20 percent of the City of San Antonio's annual
operating budget. Acquired by the City in 1942, CPS Energy serves over 728,000 electric
customers and 328,000 natural gas customers in its 1,566-square-mile (4,060 km²) service area,
which includes Bexar County and portions of its 7 surrounding counties. The CPS energy
transmission system is part of the interconnected transmission grid operated by the Electric Reliability Council of Texas (ERCOT). ERCOT serves as the regional security coordinator for the North American Reliability Council and the Independent System Operator (ISO) for the transmission grid serving the majority of Texas [SAEDF 2013]. The average annual electric rates for industrial users in Texas are amongst the lowest in the nation in 2013 [EIA 2013]. In addition, electricity bills in San Antonio are competitive when compared to other utilities across Texas and the US.

In a recent report by the U.S. Department of Energy, grid operators from 18 countries identified forecasting of variable energy resources as the most important prerequisite for successfully integrating variable generation capacity into power systems [L. E. Jones 2011]. Forecasting of solar power provides significant value when that forecast allows utilities to see this variable resource as a dispatchable power generation facility. Power can be sold to other utilities across Texas if the predictions are reliable in the day-ahead market, and demand and supply can be managed properly if the predictions are reliable for real-time operations. This is an important element of CPS Energy’s research portfolio because they are expanding their utility-scale solar power to 400 MW and because there is an increasing amount of distributed solar power generation across the city. This amount of solar energy can represent up to 20% of the electricity demand in San Antonio over the next 5-10 years, hence the risk management, revenue projection and the development of intelligence for an uncertain generation source like solar energy becomes critical for the city.

Solar forecasting is an area of research that is gaining more and more visibility due to the increasing penetration of solar energy into electricity grids. Once the amount of solar penetration into the grid reaches a certain threshold, the variability of the solar energy production may become
a problem for both grid stability and reliable bidding of electricity by electric utilities. Solar forecasting solutions provide electric utilities with predictions of power output from large-scale solar installations or from distributed solar generation with a time scale ranging from the next few minutes up to several days ahead forecast [Numnikoski 2013].

Databases offer a wealth of knowledge to users/operators but are often so restrictively large that information cannot be extracted, analyzed and synthesized quickly and efficiently enough for daily, hourly, and intra-hourly decision-making. This is where data analytic and data mining techniques can be leveraged. Data analytics seek to discover relevant data features and patterns through data mining processes and communicate that information in the most effective way to the end-user; either a human or a computerized process [Kohavi 2002, Fayyad 1996].

Various parties, such as system operators, utilities, project developers, and solar farm owners, can benefit from solar forecasting. For system operators, solar forecasts allow them to predict and manage the variability in solar energy to balance supply and demand on regional or national grid system. Moreover, knowing in advance when expected surges in cheap and clean solar energy production will occur could allow for grid operators to reduce costs through the power-down of more expensive natural gas-fired plants. In addition, there are costs associated with having excess units online, as well as from reduced unit efficiency and increased operations and maintenance. Improved solar energy forecasting can reduce these costs too. Energy providers and utilities can benefit from solar energy forecasts. Imbalance charges imposed on energy providers that result from deviations in scheduled output will increase energy providers’ operating costs. Solar energy forecasts can help minimize these penalties. Solar energy forecasts can also reduce the significant opportunity costs of being too conservative in bidding output into a forward market, due to uncertainty of availability.
1.3 Objectives and Hypothesis

An evolutionary qualitative decision-making framework for energy management in electrical microgrids together with a novel data analytics-based adaptive forecasting scheme based on data mining, statistical analysis and artificial intelligence. The benefits and drawbacks of different energy management scenarios for adopted in this study will be addressed considering the electricity trade profit, transmission network loss, and carbon emission control. The proposed forecasting methodology will be deployed for day-ahead prediction of expected solar energy and its performance will be compared the current state-of-the-art in the solar forecasting and to that of the persistence forecast, as the reference case, which is the industry’s current practice.

This research will establish an adaptive framework to quantify and integrate the effects of renewable energy intermittency, energy market, policy decisions, and environmental effects into the grid taking advantage of artificial Intelligence and pattern recognition, to bring the flexibility and robustness required for a control system. The ultimate goal is to demonstrate how control of energy flow in electrical microgrids will help improve efficiency of demand-response management, reduce electricity generation and consumption costs, as well as reduce quantifiable amounts of air pollution. The energy control framework consists of a evolutionary qualitative decision-making engine. The data analytics-based adaptive solar forecasting system combines data mining, statistical analysis and artificial intelligence in two different time-domain sequential analysis and multivariate parallel analysis dimensions. Outcomes of this research can help conserve electrical energy, save environment, bring financial benefits to either the utility companies or the microgrid owners, and improve the reliability of the power network itself. The energy management approach tackles the energy exchange problem between different nodes in a microgrid system. There is a predefined fuzzy rule base inside the decision-making engine which
determines the energy exchange rates between the storage unit in the microgrid and other nodes of the power network based on current measured values of the electricity price, solar PV electricity generation rate, electricity demand, and air pollution. Different scenarios are considered for microgrid structure and management policy, outcomes of every scenario will be considered and compared to each other and benefits and drawbacks of those approaches will be discussed so that the optimal energy management scenario can be determined for different conditions.

Solar forecasting is a rapidly expanding area of research where many different forecasting techniques have been tried over various time ranges, such as five minutes ahead to two days ahead [Nummikoski 2013]. The objective of solar energy forecast in this thesis is to develop an adaptive and intelligent framework based on big data portfolio, feature extraction, and artificial neural networks for prediction of solar energy available up to two days ahead so that the microgrid energy management engine can re-adjust its settings based on the forecast data and make optimal decisions based on current situation, physical constraints and predicted resource availability.

The patterns that weather and climate variables follow will affect the cloud formation in the sky which requires specific conditions for pressure, temperature and relative humidity with respect to the dew point and air particles present in the sky. The cloud formation consequently influences the amount of sun rays that reach the ground which are the sources of solar energy and can be transformed into electrical energy using a number of approaches including photovoltaic (PV) solar cells, concentrated solar power (CSP), etc.

There have been a number of approaches developed for tracking those patterns and estimating future behaviors of the climatic variables so that the required preparation and decision-making can be done for numerous purposes well in advance such as for traffic control, travel safety considerations, energy distribution, evacuation in case of natural disasters etc.
By developing solutions which are capable of predicting the cloud formation patterns in the sky the available solar energy can be reasonably estimated ahead of time which is a very important and critical. There are a number of approaches that focus on direct estimation of cloud formation, movement and spread patterns in the sky such as methods based on sky-imaging which are high fidelity approaches for intra-hour solar energy forecast. Other methods such as NWP or satellite-based approaches are used for recognition and prediction of cloud formation patterns in large temporal and spatial resolution.

NWP discretize space and time at the mesoscale/regional and national levels and solve the equations of motion, thermodynamics and mass transfer at pre-defined time and space intervals. The problem with NWP is that it cannot be considered site-specific but based on regional commingled datasets. Such commingled datasets propagate error from node to node and from time step to time step and its uncertainties are still too large, mean absolute error (MAE) of 20%-40% and root mean square error (RMSE) of 30%-50% for day-ahead predictions [Mathiesen 2011, Perez 2007, Bartholomy 2014 Marquez 2011, Lorenz 2009]. More recently, there has been an increased interest in smart modeling of weather and climate systems through a process of training and evolution so that the solar energy forecast can be done taking into account the cloud formation patterns [13, 14, 15, Nummikoski 2013,Fayyad 1996, Tannahill 2014, Manjili Patent 2014, Tannahill 2013, Coccccioni 2011, Amrouche 2014, Azadeh 2009, Mubiru 2008, Mellit 2010, Paoli 2010, Caputo 2010].

The proposed solar forecasting framework is not dependent on theories defined for physical processes that constrain natural processes to a set of equations. If we compare NWP to the solution of a textbook, the solution that this dissertation proposes using Adaptive Data Analytics-based Artificial Neural Networks can be compared to the experience of a veteran in weather who is able
to predict climate conditions accurately because he/she knows the local micro-meteorological features and has a robust memory to recall historical anatomies of weather and do inference based on interconnections between changing patterns of different variables with respect to one another.

The suggested approach has initially been demonstrated for the prediction of global horizontal irradiance of the next day using the irradiance information and weather sensory data of current day. Predicting irradiance on a point in space is more difficult than energy over larger areas since the local features cannot be averaged out. Therefore by developing the site-specific irradiance solar forecast tool the methodology addresses the forecast of energy over larger areas more accurately when aggregating the predictions from singular points in space.

The proposed framework takes advantage of adaptive training of an artificial neural network (ANN), as the core of the solar energy forecast engine, to optimally analyze and recognize patterns in meteorological and solar data inputs so that the ANN understand how different variables are related to each other and what would be the climate-related outcome of certain changing patterns in some variables with respect to others. The weather synoptic events detection is also an added value to the proposed forecasting framework which helps with the adaptive training process by making use of time-dependent orders within disorder to establish the time-varying coherence of these events and their effect on atmospheric cloud formation and its development. The adaptive training of the ANN ensures detection of the beginning and ends of the independent events in the weather system so that the best training sets with proven better performances can be used for generation of final forecast result. The features of the training data sets are first extracted using a powerful tool called independent component analysis (ICA) and the ANN is trained using the extracted features, i.e. the independent components or IC’s. Use of ICA drastically improves accuracy of forecasts. Forecast results are then fine-tuned according to a
hierarchical rule-base scheme obtained from study of climate patterns through three main weather components contributing to cloud formation including atmospheric pressure, temperature and relative humidity. The aforementioned procedures enable the proposed forecasting framework to adapt to both abrupt and long-term changes in climate and environmental conditions and modify its predictions accordingly.

1.4 Scope, Methodology and Limitations of Research

In many power system problems, the use of optimization techniques has been inevitable to reduce costs and losses of the system. One of the most important points in this process is the computational cost to find the sub-optimal solution. Number of constraints, number of variables, poor convergence speed, and getting stuck in local extrema are some examples of computational cost. Genetic algorithms (GA) have demonstrated to be a powerful tool to perform tasks such as generation of fuzzy rule base, optimization of fuzzy rule bases, generation of membership functions, and tuning of membership functions. All these tasks can be considered as optimization or search processes [Cordon 2001 a]. Fuzzy systems generated or adapted by genetic algorithms are called Genetic Fuzzy Systems (GFS) [Cordon 2001 b]. The combination of Fuzzy Systems with Genetic Algorithms have great acceptance in the scientific community, once these algorithms are robust and can search efficiently large solution spaces.

When designing a GFS it is often needed to first determine which part of the knowledge base is subject to optimization by the GA. The knowledge base of a fuzzy system does not constitute a homogeneous structure but is rather the union of qualitatively different components. For instance, the knowledge base of a Mamdani-type fuzzy system has two components: a data base, containing the definitions of the scaling factors and the membership functions of the fuzzy sets associated with the linguistic labels, and a rule base, constituted by the collection of fuzzy
rules. For the purposes of smart energy management in microgrid networks the GFS developed in this dissertation is designed so that the fuzzy membership functions are tuned using GA to reduce electricity consumption cost, increase energy trade profits for microgrid owner, and help reduce the carbon emission rate.

One of the very basic limitations with regard to power flow calculations and microgrid modeling was the need for a flexible tool which was able to do continual power flow calculations and also had the storage unit simulation capability, i.e. a specific node capable of switching from load state to generation state and vice versa with adjustable consumption and/or generation rates. After a broad survey conducted to find the suitable tool for doing continual power flow calculations with storage modeling capability the author came up with his own code-based solution which is implemented in MATLAB programming environment.

When the microgrid is connected to and is working synchronously with the main grid, it is assumed that the flow of electricity can be two-way, i.e. either from the main grid to the microgrid or vice-versa [Manjili WAC 2012]. Whenever the flow of electric power is from micro-grid towards the main grid, the micro-grid, or in general the customer, is making profit by selling energy to the main grid. Otherwise, the consumer must pay for the amount of electrical energy purchased from the main grid. Without loss of generality, for each time instant, it is assumed that the electricity cost for bidding and selling energy is the same under a dynamic electricity notion of price policy.

Energy management is done using a fuzzy decision-making engine based on human reasoning in combination with genetic algorithms to give the fine tuned settings of the fuzzy inference engine so that the optimal solutions are obtained with regard to increasing the microgrid owner’s benefits and reducing the air pollution.
With respect to solar energy forecast, the first challenge was the big data portfolio. National Renewable Energy Laboratory (NREL) archive, which is possibly the nation’s largest data set, is the reference big data set for solar energy forecast [NREL]. Data from more than 250 instruments, variables and readings were collected, including: global horizontal irradiance, diffuse horizontal irradiance, direct normal irradiance, solar zenith angle, solar azimuth angle, atmospheric pressure, atmospheric temperature, relative humidity, cloud cover %, wind speed, wind direction, and many more [Nummikoski 2013]. It was vital for the process to handle the huge NREL data set and to determine the most important information among more than 250 recorded variables. Measurement noise and sensor failures must have also been detected and statistically qualified. Variables added in new recordings should also be recognized and their previous null data must have been discarded to avoid erroneous solar irradiance estimation by the forecast engine. However, the cloud cover percentage was still an issue of concern since the image processing-based results were the only source determining the opaque cloud cover (OCC) and the total cloud cover (TCC) ratios and therefore, the recordings were -1 for the whole night time when there is not enough light available for imaging the sky. This added to complicacy of training the neural network and could increase the estimation error.

Adaptive artificial neural networks (AANN) in combination with pattern recognition approaches based on filtering, shifting, correlation analysis of variables, weather synoptic systems detection [Dr. Vega Thesis] and independent component analysis (ICA) constitute the core of the solar forecast engine. The forecast results will be further stabilized using the Monte Carlo statistical analysis to make the outcomes more reliable and to reduce the forecast error.

The GFS developed in this research is originally used for modifying the rules according to the optimization goals, and using weighting regimes based on the decision-making policies and
constraining factors of optimization problem. However, it can generally serve as a tool for solving multi-objective function problems where more than one target of interest exists. The proposed forecasting technique can also, due to its originality of implementation and generality of concept, be extended to other similar areas by minor changes, and the forecast results will be used to further enhance the optimization problem solutions in continuing and real-time applications. The proposed forecasting framework is a non-parametric approach which means the prediction can be made based solely based on the I/O samples of a process with no need of any representation or estimation of the internal mathematical relationships. Ultimately, the proposed forecasting framework, as a universal approach, can be applied to do predictions on any processes for which physical sensory data is available regardless of the nature of the process. Some major applications of the proposed forecasting framework include, but is not limited to, wind energy forecasting, financial/stock market forecasting, weather/meteorology forecasting, biological forecasting applications, medical/healthcare forecasting, traffic/transportation forecasting, and electrical load forecasting.

1.5 Intellectual Merit, Multidisciplinary Aspect of Research and Contributions

The evolutionary qualitative decision-making scheme for energy management in microgrids proposed in this dissertation uses a flexible GFS potentially capable of solving multi-objective function problems of various natures. Part of the intelligence added to the control system comes from the combination of GA with fuzzy decision-making engine.

Performance of the proposed intelligent energy management methodology is illustrated by numerical examples, and the net cost (which includes pricing measures, demand measures and supply measures) and air pollution measures are then compared in every scenario with the objective to identify best-practices for energy control and management of microgrids.
owners looking into adopting a smart decision-making tool for energy storage management may see a return on investment (ROI) between 5 and 10 [Manjili IASC 2013].

A number of novel contributions are associated with the proposed data analytics-based adaptive forecasting framework. The proposed method works as a non-parametric approach. This means that the suggested framework does the forecast based on the I/O samples of a process and does not need any representation or estimation of the internal mathematical relationships of the process model. This is a major intellectual merit of this research that makes the proposed framework a universal approach to forecasting which can be applied to various fields regardless of the nature of the process as long as sensory input-output data are available. Combination of ICA, as a powerful feature extraction approach, and the ANN, as one of the most popular modeling and estimation tools in engineering, with an algorithm for detection of weather synoptic events is an intellectual merit of this research. Algorithms of weather synoptic systems detection [Dr. Vega Thesis] and the filtering, shifting, cross correlation analyses of variables bring added intelligence to the forecast engine which leads to adaptive training of the ANN using appropriate training window and the most information-rich portions of the training subset. The proposed day-ahead solar energy forecasting framework brings an added value of 9.5% to 14.5% to the utility companies compared to the best-practice in persistence forecasting.

Further outcomes of the present research include one US patent, one journal article published, one journal article accepted for publication, one journal article ready for submission, and five conference papers/posters among them a best poster award for college of sciences research conference at the University of Texas at San Antonio (UTSA-COS 2011).
CHAPTER 2: ELECTRICITY; SOURCES, MARKET, ENVIRONMENT AND GRID

2.1 Electrical Energy Sources

Energy demand in the world is increasingly rising each day. Large and small scale industries, residential areas, commercial sector, and transportation are among the most notable consumers of energy. Figure 1 represents the world energy demand and the in the long term energy sources until 2055 based on statistical data and scientific predictions.

Figure 1 World Energy Demand – Long term Energy Sources [Orr 2005]

The energy sources are separated into two categories of sustainable and fossil fuels in Figure 1. In this categorization sustainable sources include biofuels, hydroelectric, solar, geothermal which, except for hydropower, are relatively new and growing, but their growth rate
is expected to be much higher around 2020 and after. Figure 2 represents the world’s energy sources projected from 1930 to 2100 as an alternative world energy outlook (AWEO 2006) by LBST organization in Germany. There’s a notable decline of the energy curve in between the years 2010 to 2030 which does not match with the world energy outlook report of 2006 (WEO 2006). This could have been happened because of world economic decline which seems to affect the world market until 2030.

![Figure 2: Global Energy Sources by Alternative World Energy Outlook [LBST AWEO 2006]](image)

As it can be inferred from diagram of Figure 2, the world energy supply is transition from fossil fuels to the renewable and sustainable resources in the coming decade. This means not only the rate of energy supply using fossil fuels is expected start to decline after reaching
its peak in the coming decade, but also the renewable resources will start to grow more rapidly than ever so that by 2100 the world primary energy supply is renewable resources, including hydro power, and the fossil fuels constitute a small percentage, less than 30%, of world’s energy production. 

Figure 3 shows the country-wise diagram of world electricity generation for the based on recordings of the United States Central Intelligence Agency (CIA) in 2012 and estimations of some of the countries production made in years 2010 and 2011. It can be seen from Figure 3 that the United States is the second country in the world, after China, with regard electricity generation. The two countries China and the US are far ahead of the rest of the world in this regard.

![Figure 3 World Electricity Generation by Country](CIATWF Elec. 2012)
Major portion of energy consumption in those countries listed in figure 3 is for the purpose of generating electricity which is the vital form of energy used in various industrial, commercial, residential, and somehow in transportation sectors.

Figure 4 represents United State’s annual energy consumption by source from based on the US Energy Information Administration (EIA) data [EIA Energy 2010] as a major energy leader and one of the most influential countries in the world. Majority of the energy sources are used in the process of generating electricity since electricity is the force behind most industrial, residential, and commercial applications.

Figure 4 US Energy Consumption by Source; From 1800 to 2010 [EIA Energy 2010]
As represented in the semi-logarithmic diagram of Figure 4, the prominent players in the energy industry are Petroleum, Natural Gas, and Coal, each of which contributes to the US energy industry more than 10 times that of Nuclear power, Wood, and Hydroelectricity. However, in late 1900s and early 21st century the advent of renewable forms of energy into the market is clearly visible. In 2010, Biofuels, Hydroelectricity, Wind, Geothermal and Solar energy were respectively the main contributors of renewable energy industry to electricity generation in the US.

Nowadays, there are a number of different approaches used to produce electricity around the world using the major energy sources mentioned in figures 1 and 2. However, regardless of the fact that the generating party is a large scale utility company owned by government, a private sector power plant, a small scale plant of a microgrid, or a micro-scale residential plant, the methods by which electricity is generated can be divided into two major categories: conventional and renewable.

2.2 Conventional (Non-renewable) Methods

Conventional methods for electricity generation involve methods which use coal, natural gas, oil and other carbon-emitting sources of energy to produce electricity by different means including coal-fired power plants or gas or steam turbines, etc. These fuels have much larger capacity than other kinds of energy sources in the nature and are mostly feasible for different scales of utilities. However, the most important issue with these sources of energy is that they contribute highly to carbon emission and add other toxic gasses to the atmosphere which are harmful to the environment and the living beings in the nature. As represented in figure 1, 2, and 3, coal, oil, and gas are the predominant source of energy currently used for electricity generation around the world.
world. Major reason for this is the economic feasibility of these resources compared to other sources such as renewable resources.

Some references also include the hydro-electricity generation in the conventional approaches for electricity production. Nuclear power plants are not included directly in the conventional sources of electricity generation since there are very little or zero carbon emission associated with the process of those plants. However, taking into account the indirect costs and emissions due to cement production required for building a nuclear power plant and its structures the result may show a huge negative environmental effect, apart from the nuclear plant’s potential threat for the natural environment [Brumfiel 2013].

2.3 Sustainable (Renewable) Methods

Renewable methods revolve around using nature’s energy sources such as sun, wind, biomass, geothermal, etc. to produce electricity. Solar power is the conversion of sunlight into electricity. This process can be done directly using photovoltaic (PV) cells, or indirectly using concentrated solar power (CSP) technology. Concentrated solar power systems use lenses or mirrors and tracking systems to focus a large area of sunlight into a small beam. Photovoltaics convert light into electric current using the photoelectric effect [DOE EnergySources 2011].

Photovoltaics (PV) were initially, and still are, used to power small and medium-sized applications, from the calculator powered by a single solar cell to off-grid homes powered by a photovoltaic array. They are an important and relatively inexpensive source of electrical energy where grid power is inconvenient, unreasonably expensive to connect, or simply unavailable. However, as the cost of solar electricity is falling, solar power is also increasingly being used even in grid-connected situations as a way to feed clean low-carbon energy into the grid.
Commercial concentrated solar power plants were first developed in the 1980s. The 354 MW SEGS CSP installation is the largest solar power plant in the world, located in the Mojave Desert of California. Other large CSP plants include the Solnova solar power station (150 MW) and the Andasol solar power station (150 MW), both in Spain. The 250+ MW Agua Caliente solar project in the United States, and the 221 MW Charanka solar park in India, are the world’s largest photovoltaic power stations. Table 1 represents the world’s total annual solar energy for the period between 2005 and 2012. An increasing growth rate can be seen by comparing the % of total in the third column for consecutive years.

Table 1 World Annual Electricity Generation from Solar [HDW 2013]

<table>
<thead>
<tr>
<th>Year</th>
<th>Energy (TWh)</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>3.7</td>
<td>0.02 %</td>
</tr>
<tr>
<td>2006</td>
<td>5</td>
<td>0.03 %</td>
</tr>
<tr>
<td>2007</td>
<td>6.7</td>
<td>0.03 %</td>
</tr>
<tr>
<td>2008</td>
<td>11.2</td>
<td>0.06 %</td>
</tr>
<tr>
<td>2009</td>
<td>19.1</td>
<td>0.09 %</td>
</tr>
<tr>
<td>2010</td>
<td>30.4</td>
<td>0.14 %</td>
</tr>
<tr>
<td>2011</td>
<td>58.7</td>
<td>0.27 %</td>
</tr>
<tr>
<td>2012</td>
<td>93.0</td>
<td>0.41 %</td>
</tr>
</tbody>
</table>

Figure 5 shows the linear diagram of world solar capacity and the pie diagram of major leading countries in solar industry [EIA Solar 2009]. It is seen that Germany was by far the leading country in the world with regard to using solar energy by having almost half of world installed capacity. The United States at the same time was and still is among the top five countries leading in solar energy industry.
The IEA (2010) forecasts an average annual market growth rate of 17% in the next decade, leading to a global cumulative installed PV power capacity of 200 GW by 2020 and 3000GW by 2040 (with repowering of older systems). This would represent roughly 11 percent of global energy demand should this scenario play out.
United States’ PV market forecast is represented in Figure 6 which shows the overall market of the US and the residential PV market for US states for the period of 2010 to 2016. The projected growth is expected to get to a point where almost 9 GW of electricity is generated from PV panels in 2016 more than 2 GW of which are used in residential sector.

Figure 7 US PV Market between 2000 and 2012;

Figure 7 represents the US photovoltaic market’s cost and growth from 2000 to 2012. The PV market cost had a descending trend while its growth trend was generally ascending with some deviations between 2004 and 2010.
Figure 8 represents the renewable energy market forecast done in 2011 which trends of renewable energy sources up to 2040. The Solar energy is expected to pass the Geothermal by 2040 as mentioned in the diagram. Wind will remain the dominant renewable resource of electricity generation and will still be growing, apart from hydropower which continues to keep a flat curve without much of a growth. It can be seen that the energy generated from municipal solid waste and landfill gas (MSW/LFG) will continue to be on the same level and won’t show an expected growth in rate or amount. EPA has a number of standards for MSW landfills to reduce environmental and public safety concerns.
Currently, in the US, more than 3 GW-dc of PV is installed and operating in the US to help utilities provide electricity. More than 10 GW-dc worth of projects are either under construction or their contract has been signed. Also, more than 24 GW-dc worth of projects are announced and undergoing the pre-contract procedures and have to secure a power off-taker.

In Figure 1 and Figure 2 it can be seen that the PV and wind energy will more important in the long term. Besides, in Figure 4 it is represented that the solar energy highest slope of growth compared to all other types of energy in the market in 2010 and is expected to have a growing share of electricity production in future as predicted in Figure 8. Hence, considering the budget allocation for development of renewable energy by the office of energy efficiency and renewable energy (EERE) at department of energy (DOE), as represented in Figure 9, it is fair to expect the solar energy to be a more and more important player in the energy sector.
in the years to come and will probably pass some other competitors, like Geothermal and even Biomass, behind.

2.4 Environmental Protection Agency (EPA) and Air Pollution Standards

Using some sources of energy for electricity production will result in emission of carbon and other gasses which are harmful to health and nature and can threaten the environment we live in. Mostly, the conventional methods of electricity generation contribute to those harmful emissions. The more the electricity consumption increases, the larger the generation rates should be in order to compensate for that and to provide the consumers with the required energy. In other words, the utilities have to supply the demand. Figure 10 shows the pure and per capita electricity consumption in the 51 states of the US in 2006.
Figure 10 Electricity Consumption in 51 States of the US in 2006 [EIA Elec. 2006]
(a) Overall power in GigaWatts   (b) Power per capita in kiloWatts
It is evident from Figure 10(a) that Texas has the highest net electricity among all other states of the US. California secures the second rank as stays afar from Florida which is in the third place. Figure 10(b) shows the per capita electricity Wyoming, and Louisiana are respectively the top three states in per capita consumption. Alaska and Wyoming have comparatively very low net electricity consumption, however, since those states have little population, their per capita consumption suddenly rises. Texas also has a relatively high per capita consumption and stays in the 5th place. The per capita electricity consumption in Texas is around 17 kW per person which is way higher than the nation’s average of 11.14 kW.

As represented in Figure 11, electricity generation has a positive correlation with emission. Texas and California lead the nation with regard to carbon emission with Texas being by far on top contributing to more than 650 million metric tons of carbon dioxide in 2010. Therefore, the need for emission-free or low-emission electricity generation is inevitable, specifically in states like Texas which have such an important leading role in technology and industry and have the resources available.
Environmental Protection Agency (EPA), the entity which is in charge of controlling the emission of dangerous gasses into the atmosphere, has introduced standard measures called air quality index for computing air pollution [EPA AQI 2012]. Carbon Dioxide (CO2) is the main contributor to air pollution and global warming when considering the emissions from power plants. A 1000 megawatt (MW) coal-fired power plant produces approximately the same amount of global warming as 1.2 million cars [Freese 2006]. According to EPA reports the coal-fired power plants emit 2249 pounds of CO2 per 1 megawatt-hour (MWh) of electricity generated while the power plants which use natural gas as the fuel to generate electricity produce around 1135 pounds of CO2 for the same amount of energy generated. EPA has released a new rule to regulate CO2 emissions from power plants. The new rule requires power plants to meet an output-based standard of 1,000 pounds or less of carbon emission per MegaWatt-hour (MWh) of electricity generated [EPA MSW 2012].
2.5 Smart Grid and Microgrids

In 1881 two electricians built the world's first power system at Godalming in England. It was powered by a power station consisting of two waterwheels that produced an alternating current that in turn supplied seven Siemens arc lamps at 250 volts and 34 incandescent lamps at 40 volts [Gravett 1981]. However, supply to the lamps was intermittent and in 1882 Thomas Edison and his company, The Edison Electric Light Company, developed the first steam powered electric power station on Pearl Street in New York City. The Pearl Street Station initially powered around 3,000 lamps for 59 customers [Williams 2007], [Grant 2013]. The power station used direct current and operated at a single voltage. Direct current power could not be easily transformed to the higher voltages necessary to minimize power loss during long-distance transmission, so the maximum economic distance between the generators and load was limited to around half-a-mile (800 m) [Klap 2013].

Smart grid is a modernized electrical grid that uses information and communications technology to gather knowledge and act on data, such as information about the behaviors of suppliers and consumers, in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity [DOE SmartGrid 2012]. The smart grid technology also implies a fundamental re-engineering of the electricity services industry, although typical usage of the term is focused on the technical infrastructure [Torriti 2012].

Smart grid networks allow two-way flows of electricity and information to create an automated and distributed advanced energy delivery network which can serve the customers in a more reliable and smarter way, taking advantage of artificial intelligence-based approaches,
compared to the traditional grids which are generally used to carry power from generating nodes to a number of consuming nodes.

A microgrid is a combination of loads and generators designed locally as a small-scale electrical grid for the purpose of providing energy and distributing it between local consumers. In other words, a micro-grid is an aggregation of multiple distributed generators (DGs) such as renewable energy sources, conventional generators, in association with energy storage units which
work together as a power supply network in order to provide both electric power and thermal energy for small communities which may vary from one common building to a smart house or even a set of complicated loads consisting of a mixture of different structures such as commercial buildings, industrial facilities, hospitals, schools, etc [1]. Typically, a micro-grid operates synchronously in parallel with the main grid. However, there are cases in which micro-grid operates in islanded mode, i.e. disconnected from the main grid [2]. Figure 12 represents general schematic that includes a smart grid structure including of a number of microgrids.

2.6 Summary

World electricity supply and demand is discussed in this chapter and sources of electricity generation are studied including renewable energy resources. Specifically, the solar PV market and percent share of electricity generation is considered. Air pollution measures and standards are introduced and main US states that contribute to harmful emissions are represented. Smart Grid and microgrid definitions are illustrated. In the next chapter the autonomous electricity flow control for energy management in microgrids will be implemented based on the information given here considering the EPA air pollution standards and the constraints imposed to the microgrids and power plants.
CHAPTER 3: MICROGRID CONTROL AND ENERGY MANAGEMENT

3.1 About Microgrid Energy Flow Control

Various algorithms and novel approaches are introduced for microgrid energy management. Auction-based theory for pricing strategy in solar powered micro-grid is studied in [3]. In [1], the authors proposed a fuzzy logic-based decision-making approach to control power exchange with battery storage unit in micro-grid considering an ideal storage unit with both the limited-capacity and unlimited-capacity cases, and investigated the overall costs and profits the fuzzy approach could bring to the system. In [4] the concept of micro-grids and applications of intelligent systems theory have been applied to various problems and issues in these power systems. An overview of micro-grid systems control is given in [5].

In this chapter, the GFS decision-making framework will be enhanced by weighting factors determined based on the difference between current time and the predicted peak pollution time for the day ahead in order to elevate performance of the system by reducing micro-grid owner’s losses due to the electricity purchased from the main grid and increasing his gains or gains due to the electricity sold to the main grid while helping reduce the air pollution and modify the peak air pollution due to micro-grid operation. Genetic algorithms (GA) is the distributed search method used to enhance the fuzzy system and to determine the settings of fuzzy sets so that sub-optimal results for energy and pollution management could be achieved. When the micro-grid is connected to the main grid and is working synchronously with it, it is assumed that the flow of electricity can be two-way, i.e. either from the main grid to the micro-grid or vice-versa [1]. Whenever the flow of electric power is from micro-grid towards the main grid, the micro-grid, or in general

35
customer, is making profit by selling energy to the main grid. Otherwise, the consumer must pay for the amount of electrical energy purchased from the main grid. Without loss of generality, for each time instant, it is assumed that the electricity cost for bidding and selling energy is the same. Demand-side load management is not considered since the micro-grid is supposed to fully provide the local load with the demanded energy and to meet the load profile.

A smart decision-making framework based on genetic fuzzy systems (GFS) is proposed for control and energy management of micro-grids. Objectives are to meet the demand profile, minimize electricity consumption cost, and to modify air pollution under a dynamic electricity pricing policy. The energy demand in the micro-grid network is provided by distributed renewable energy generation (coupling solar and wind), battery storage and balancing power from the electric utility. The fuzzy intelligent approach allows the calculation of the energy exchange rate of the micro-grid storage unit as a function of time. Such exchange rate (or decision-making capability) is based on (1) the electrical energy price per kilowatt-hour (kWh), (2) local demand (load), (3) electricity generation rate of renewable resources (supply), and (4) air pollution measure, all of which are sampled at predefined rates. Then, a cost function is defined as the net dollar amount corresponding to electricity flow between micro-grid and the utility grid. To define the cost function one must consider the cost incurred by the owner of the micro-grid associated to its distribution losses, in addition to its demand and supply costs, in such a way that a positive cost translates to owner losses and a negative cost is a gain. Six likely scenarios were defined to consider different micro-grid configurations accounting for the conditions seen in micro-grids today and also the conditions to be seen in the future. GA is implemented as a heuristic (DNA-based) search algorithm to determine the sub-optimal settings of the fuzzy controller. The aforementioned net cost (which includes pricing, demand and supply measures) and air pollution
measures are then compared in every scenario with the objective to identify best-practices for energy control and management of micro-grids. Performance of the proposed GA-fuzzy intelligent approach is illustrated by numerical examples, and the capabilities and flexibility of the proposed framework as a tool for solving intermittent multi-objective function problems are presented in detail. Micro-grid owners looking into adopting a smart decision-making tool for energy storage management may see an ROI between 5 and 10.

3.2 Microgrid Model and Energy Management Scenarios

The model used for simulation of the microgrid is a three-bus power network, integrating every bus on the power network as three different types. The methods presented in this article are applicable and easily transferable to micro-grids of any size, therefore the three-bus system was chosen for illustrative purposes only. One of the buses in the distributed generation model is assumed to serve the renewable generators including solar farm, wind farm, or any other renewable generation units either in association with battery storage unit or without any storage. Another bus is assumed to be there as the representative for connection between main grid, i.e. utility, and the local micro-grid through which two-way power flow is available. The excess energy demanded by local load, which renewable electricity generation system cannot afford to supply, is drawn from the main grid through this bus. The excess energy existing at the micro-grid side can also be delivered to the main grid from the same bus. The third bus will be the integration of all local loads. This load can be anything from a common building or a smart house, to a group of factories, or any combination of all. Figure 13 shows schematic of generic micro-grid including electricity generators and storage unit, utility, and local load.
Figure 13 Schematic of a Generic Electrical Micro-grid

Table 2 Input-output Configuration of Energy Management Scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Power Network Model Elements</th>
<th>Fuzzy Inputs</th>
<th>Fuzzy Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>Main grid (Utilities) Local Load</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>Main grid (Utilities) Local Load Renewables</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>Main grid (Utilities) Local Load Renewables Battery Storage</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>Main grid (Utilities) Local Load Renewables Battery Storage Fuzzy Control</td>
<td>( P_r(t) ) ( P_a(t) ) ( P_L(t) )</td>
<td>( P_B(t) )</td>
</tr>
<tr>
<td>Scenario 5</td>
<td>Main grid (Utilities) Local Load Renewables Battery Storage Enhanced Control</td>
<td>( P_r(t) ) ( P_a(t) ) ( C(t) )</td>
<td>( P_B(t) )</td>
</tr>
<tr>
<td>Scenario 6</td>
<td>Main grid (Utilities) Local Load Renewables Battery Storage Pollution Control</td>
<td>( P_r(t) ) ( P_a(t) ) ( C(t) )</td>
<td>( P_B(t) )</td>
</tr>
</tbody>
</table>
Six simulation scenarios are considered for the power network in this article. Specifications of these scenarios are given in Table 2 which include the elements of the power network model, the fuzzy inputs to the energy flow controller, and the fuzzy output from the decision-making engine for each and every scenario.

In Table 2, \( Pr(t) \) represents price per kWh of electrical energy in Cents at time \( t \), the electricity generation rate of renewable resources, which may include any combination of net production from either solar arrays or wind turbines, in kW at time instant \( t \), \( P_L(t) \) represents the local load, or the so called energy demand rate, in kW at time \( t \), \( C(t) \) represents the pollution measure which shows the micro-grid’s contribution in the amount of CO2 in the atmosphere, and \( P_B(t) \) is the energy exchange rate of storage unit at time instant \( t \).

It must be noted that in this study, the environmental-friendly aspect of the micro-grid control considers the effect that each and every scenario has on the air pollution, specifically speaking the amount of CO2 added to the air due to micro-grid energy demand. Hence, instead of using detailed mathematical representations of the fluid dynamics for propagation of CO2, the net amount of CO2 emitted from conventional sources of energy due to micro-grid’s excess energy demand and/or the amount of CO2 eliminated from the atmosphere due to well-supervised storage and use of utility grid’s power and renewable resources is studied. The pollution measure, \( C(t) \), represents the average CO2 concentration in the whole region of interest. The region of interest is referred to the city in which the micro-grid is located, and is not limited to the area around the micro-grid network or the area close to conventional power plants of the main grid. Assuming other sources of pollution keep a recurring profile of pollutant emission, which is a fair assumption for the simulation period of one week considered in this article, it will be possible to study how much micro-grid will incorporate to air pollution.
The micro-grid system model is simplified down to a three-node model, with three branches connecting every two nodes to each other, where the renewable generators are supposed to be located all together with the storage unit at a single node, the utility grid connections to the micro-grid network are summarized as the second node, and the consumers or the so called local load is considered as the third node. It is assumed that the renewable generators do not have any capability to provide the micro-grid with reactive power which means that the renewable resources bus should be considered as a PQ bus since at each instance of time both active and reactive power present at this bus are known for power flow calculations. The active power is known due to renewable generation available and the energy exchange with the storage unit, and the reactive power is also known to be zero all the time since it was the assumption that there’s no reactive power capacity available at this bus. Utility bus is the standard choice for a slack bus since it is assumed that the voltage magnitude and angle are constant at this bus regardless of the amount of active and/or reactive power exchange through it. Load bus, therefore, needs to be the PV bus in order for power flow calculations to be done properly. In order to make this possible, there’s this assumption that the load bus is associated with a synchronous compensator which helps keep the voltage magnitude constant at 1 p.u. regardless of the active or reactive power demand at the load side. Furthermore, the active power demand of the load is also known at each time instance, which together with the synchronous compensator make the necessary and sufficient conditions for a PV type bus to be there. In the following, characteristics of the three buses in the micro-grid model are mentioned for each simulation scenario.

3.2.1 Characteristics of Microgrid Buses in Scenario 1

The network only consists of two buses as follows:
• First bus is of type slack (reference) and is used as the utility (main grid) connection node.
• Second bus is of type PV; used as the local load bus with a synchronous compensator.

Renewable energy generators and storage units are not considered in this scenario. Hence, the local loads, i.e. the consumers, are only supplied by the main grid which was the typical case prior to introduction of the definition of micro-grid, renewable resources and storage unit to the industry. This scenario is predominant in majority of cases around the globe in order to provide electricity from utility companies to the consumers.

For scenarios 2 to 6 refer to Figure 13.

3.2.2 Characteristics of Microgrid Buses in Scenario 2

Renewable resources are deployed in the micro-grid without any storage units available. Characteristics of the three buses in the micro-grid network model are as follows in the second scenario:
• Bus 1 is a PQ bus and is used as the bus for renewable generation unit.
• Bus 2 is the slack (reference) bus and is used as the utility (main grid) connection node.
• Bus 3 is of type PV and is used as the local load bus with a synchronous compensator.

3.2.3 Characteristics of Microgrid Buses in Scenario 3

In this scenario, the storage unit is also added to the model. Characteristics of the three buses in the micro-grid network model simulated in this scenario are as follows:
3.2.4 Characteristics of Microgrid Buses in Scenario 4

Micro-grid structure in this scenario is the same as that of scenario 3. However, the intelligent decision-making approach, i.e. fuzzy control, is used as the energy control and management engine to provide the system with reduced costs and increased benefits. Buses have following characteristics:

- Bus 1 is a PQ bus and is used as the bus for renewable generation unit and storage unit.
- Bus 2 is the slack (reference) bus and is considered as the utility (main grid) connection node.
- Bus 3 is of type PV and serves as the local load bus with a synchronous compensator.

Three input variables to the intelligent control system include electricity price, local load demand, and renewable electricity generation rate. Output of the fuzzy control system will be the energy exchange rate between the storage unit and other elements of the micro-grid at each time instance.

3.2.5 Characteristics of Microgrid Buses in Scenario 5
Scenario 5 is essentially similar to scenario 4 with another input variable added to the fuzzy controller called the air pollution index. Rules of the fuzzy inference engine are also modified in such a way to take into account environmental constraints besides providing the micro-grid with reduced costs and increased benefits. However, the main constraint is to still provide local load with the required energy demand. Making gain for micro-grid owner and reducing micro-grid’s negative impact on the air pollution compared to the first scenario are both of second priority and a compromise has to be made for these two secondary goals when modifying the fuzzy rule-base. Characteristics of buses in this scenario are:

- Bus 1 is a PQ bus and is used as the bus for renewable generation unit and storage unit.
- Bus 2 will be the slack (reference) bus and is used as the utility (grid) connection node.
- Bus 3 is of type PV and is used as the local load bus with a synchronous compensator.

3.2.6 Characteristics of Microgrid Buses in Scenario 4

This scenario is almost a duplicate of the scenario five with only difference being the fuzzy decision-making engine is enhanced so that the control commands can be emphasized relative to the sensitivity of each condition. There will be a weighting factor which is determined based on the difference between current time and the estimated peak pollution time. Rules of the fuzzy inference engine are the same as those in scenario 5. However, the membership value of the fuzzy input variable air pollution index is weighted in such a way that whenever the current time-slot is close to the estimated peak pollution time, the emphasis on the membership value of high air pollution fuzzy set will increase and simultaneously the emphasis on the membership value of low
air pollution fuzzy set will decrease, and vice-versa. The main constraint is to still provide the local load with the demand profile. Making gain for micro-grid owner and reducing micro-grid’s negative impact on the air pollution compared to the first scenario are both of second priority and a compromise needs to be made for these two secondary goals when generating the fuzzy rule-base. Characteristics of buses in this scenario are:

- Bus 1 is a PQ bus and is used as the bus for renewable generation unit and storage unit.
- Bus 2 will be the slack (reference) bus and is used as the utility (grid) connection node.
- Bus 3 is of type PV and is used as the local load bus with a synchronous compensator.

The energy capacity of the storage unit is assumed to be limited and normalized. The state-of-charge (SOC) of battery unit will be computed at the end of each 15-minute period using the constant energy exchange rate for the same time interval determined by intelligent energy management approach. It is assumed that the stored energy curve is sufficiently linear within the 10% to 90% SOC range. The storage unit is assumed to be capable of supplying the micro-grid’s local load for up to 10 hours at full demand when fully charged. This means that, for instance, a micro-grid with full demand of 1 MW at the load side should be associated with a storage unit that is capable of storing 10 MWh energy.

3.3 Problem Statement

Dynamic pricing policy for electricity purchase means the bid price is not constant during the day. The update duration of electricity cost is assumed to be 15 minutes. This implies that the money consumers have to pay to the utility for the same amount of energy consumed during
different time-intervals might be different. Therefore, a function is introduced to take into account
the difference between amount of power given to the utility from the micro-grid, and the amount
of power taken from the utility by the micro-grid. This function gives us a cumulative sum of the
money that consumer must pay to the utility or, in some circumstances, the consumer gets from
the utility companies by selling the electricity to the them due to proper purchase, storage,
consumption, and sale policy. \( Eq. 1 \) represents this cost function:

\[
\text{Cost} = \int_0^T \text{Pr}(t) P_U(t) dt
\]

\( Eq. 1 \)

Where the electricity price \( \text{Pr}(t) \) is the sell and bid value per kilowatt-hour of electrical
energy. \( P_U(t) \) is the active power exchanged between micro-grid and the utility grid. If energy is
received from the utility companies \( P_U(t) \) is assumed to be positive, and if power is delivered to
the utilities \( P_U(t) \) will appear with a negative sign. The update duration of the electricity cost is
assumed to be 15 minutes. Therefore, during the 24-hour day period, there will be a total of \( \frac{24\text{h}}{0.25\text{h}} = 96 \) time intervals for each of which the electricity cost will be determined by
the utility companies and remains the same.

\( Eq. 2 \) represents the standard measure used for computing air pollution called air
quality index introduced by the environmental protection agency (EPA) for major air pollutants
[6]: 
\[ I = \frac{\text{Pollutant Concentration}}{\text{Pollutant Standard Level}} \times 100 \quad \text{Eq. 2} \]

In this study, Carbon Dioxide (CO2) is considered as the emission factor contributing to air pollution and global warming. A 1000 megawatt (MW) coal-fired power plant produces approximately the same amount of global warming as 1.2 million cars [7]. According to EPA reports the coal-fired power plants emit 2249 pounds of CO2 per 1 megawatt-hour (MWh) of electricity generated while the power plants which use natural gas as the fuel to generate electricity produce around 1135 pounds of CO2 for the same amount of energy generated. EPA has released a new rule to regulate CO2 emissions from power plants. The new rule requires power plants to meet an output-based standard of 1,000 pounds or less of carbon emission per MegaWatt-hour (MWh) of electricity generated [8]. Constraining relationship between the energy taken from the main grid and the CO2 added to air is shown in \textbf{Eq. 3}:

\[ p = \psi E \quad \text{Eq. 3} \]

Where \( p \) represents how much CO2 in units pound is added to the air, \( E = \int P(t)dt \) is the energy generated by the power plant during a specific time period where \( P(t) \) stands for the function representing output electrical power of the plant in megawatts (MW), and \( \psi \) is the restricting coefficient which is assumed to be \( 1000 \left( \frac{\text{lb}}{\text{MWh}} \right) \).

\textbf{Figure 14} in the following represents the three-bus model used for simulation of the micro-grid in different scenarios, as discussed earlier in Sections 3.2.1 to 3.2.6, along with the branch impedances and types of the buses.
Figure 14 Three Bus Micro-grid Model

Mathematical representation of the active power exchange between the micro-grid and main grid is given in Eq. 4:

$$\sum_{i=1}^{N} P_U^i(t) = \sum_{i=1}^{N} (P_L^i(t) - P_B^i(t)) + P_{\text{loss}}(t) \quad , \quad P_{\text{loss}}(t) = \sum_{i=1}^{N} \sum_{j=1}^{N} (|P_{ij}(t)|^2 \cdot Z_{ij})$$

Eq. 4

subject to:

- if $P_U^i(t) < 0$ energy is being stored in $i^{th}$ battery, i.e. storage can be regarded as a load
- if $P_U^i(t) > 0$ energy is being drawn from $i^{th}$ battery, i.e. storage acts like a generator

Where $P_U^i(t)$ stands for the power purchased from the main grid at $i^{th}$ bus, $P_U^i(t) > 0$, or sold to the main grid, $P_U^i(t) < 0$, through $i^{th}$ bus at time instant $t$. $P_L^i(t)$ is the local energy demand, i.e. the local load, at $i^{th}$ bus at time instant $t$. $P_{\text{loss}}(t)$ shows the distribution loss due to branch impedances at time instant $t$ and as represented in the equation can be computed by summing the amount of loss on each individual branch over the entire network. Considering $P_{\text{loss}}(t)$ in the cost function can help taking into account the distribution losses over the branches of the power network in very large-scale micro-grids. However, for small-scale micro-grids the distribution loss, i.e. $P_{\text{loss}}(t)$, is small enough to be neglected or is compensated for by the utility companies.
themselves. \( P_B^i(t) \) is the rate at which energy is given to storage unit, i.e. \( P_B^i(t) < 0 \), or is taken from it, i.e. \( P_B^i(t) > 0 \), at ith bus at time instant t. Accordingly, \( P_B^i(t) \) represents the electrical power produced by renewable generator at the ith bus at time instant t which, neglecting factors such as the reverse saturation current of solar farm when there’s no sunlight in the sky, is assumed to be either equal to or greater than zero.

The value \( P_B^i(t) \) will be determined by a management scheme for each 15 minute interval at the beginning of the interval based on samples of the input variables including electricity cost, renewable electricity generation rate, local load demand for scenarios four to six. However, only for scenarios five and six, air pollution or CO2 concentration will play a role as the fourth input.

Hence, using the value \( P_B^i(t) \) calculated by the energy management scheme for each 15-minute interval, plus the continuously sampled values of \( P_L(t) \) and \( P_R^i(t) \), the values of \( P_U^i(t) \) and also \( P_{\text{loss}}(t) \) can be determined by a power flow calculation algorithm in power networks since the impedances of the branches are known. For the sake of generality, as represented in Figure branches of the network are assumed to be alike with the same impedances.

There are a number of methods for calculation of power flow in the distributed generation network [9]. Four different types of buses are generally considered in a distributed generation network, the characteristics of which will be used for calculation in power flow algorithms. These four types include PQ, PV, slack, and isolated [10, 11]. For the simulation purposes of this paper, Gauss-Seidel iterative algorithm is implemented to do the power flow calculation [11].

### 3.4 Decision-Making Platform for Microgrid Energy Management

A framework based on fuzzy logic [12] is used for energy management in the micro-grid network by control of the power exchange with battery storage unit in order to improve the cost function introduced in Eq. 1. Slightly different versions of the fuzzy controller are applied in the
three scenarios four, five, and six. The three input variables to the fuzzy inference engine for scenario four include electricity cost per kWh or $Pr(t)$, renewable electricity generation rate or $Pr(t)$, and local load demand or $Pl(t)$. The fuzzy inference engine serves as the controller which determines the rate at which electrical energy must be exchanged with the battery unit during each 15 minute period, based on the samples of the three input variables at the beginning of that period.

In scenario five, a fourth input variable will be fed to the fuzzy inference engine called air pollution measure or $C(t)$. For the purposes of this study, an exemplary pollution profile is generated for a typical 24 hour period in order to examine capabilities of different scenarios. $C(t)$ is the average amount of CO2 on entire region of interest, neither only at specific points around the polluting power plants nor only around micro-grid local loads. Discrete-time mathematical representation for air pollution update is represented in Eq. 5 and Eq. 6:

$$C(k+1) = C(k) + \Delta C$$ \hspace{1cm} \text{Eq. 5}

$$\Delta C = \Delta p(k) - \Delta r(k)$$ \hspace{1cm} \text{Eq. 6}

Where $C(k)$ represents the measure of pollutant concentration, here CO2, at the end of $k^{th}$ 15-minute time interval. $\Delta C$ stands for the change in the CO2 measure during the $k^{th}$ time interval. The term $\Delta p(k) = p(k+1) - p(k) = \psi \int_{k\Delta t}^{(k+1)\Delta t} P_D(t) dt$ represents the amount of CO2 added to the air during $k^{th}$ time interval due to operation of the main grid’s conventional power plants when $\Delta t$ represents the duration of each time interval, i.e. 15 minutes. Finally, the term $\Delta r(k)$ represents the removal portion of pollution associated with chemical reactions and pollution’s dispersion in the atmosphere during $k^{th}$ time interval. It is assumed that the pollution removal
obeys the law of exponential decay. The coefficient for pollution removal is stochastically drawn out from a normal random distribution for each time interval.

![Diagram of the real-time intelligent fuzzy decision-making approach](Figure 15) 

Figure 15 Structure of the real-time intelligent fuzzy decision-making approach; four input variables and one output variable

... Figure 15 represents structure of the intelligent decision-making approach article. Quantitative sensory data are gathered from different locations of the micro-grid network and also from the utility. These data include electricity price, \( Pr(t) \), renewable electricity generation rate, \( PR(t) \), local load, \( PL(t) \), and measure of CO2 present in the air. These four variables undergo a fuzzification process in order to transform into linguistic terms which can be used for fuzzy decision-making in the rule-base of the inference engine. The output which results from the decision-making process, i.e. \( PB(t) \), is still in linguistic form however. Hence, a defuzzification process must be done in order to transform the output from linguistic or qualitative term into quantitative term so that it can be used for control and management actions in practical micro-grid network.

Original fuzzy sets for the four input variables and the only output variable of the fuzzy inference engine (see Figure 15) are shown in Figure 16.
Figure 16 Fuzzy Membership functions for input and output variables of the Fuzzy Controller; (a) price, load and generation (b) air pollution (c) output

According to Figure 15 and Figure 16, the numerical values for the three input variables price, load, and generation are normalized to the zero to one interval, and then are fuzzified using three fuzzy sets defined as Low (L), Medium (M), and High (H) as represented in figure 4a. Air pollution has two membership functions defined as Low (L) and High (H) shown in figure 4b. After fuzzification, the input variables will be fed to fuzzy inference engine where the rule-base is applied to them and the fuzzy output will be determined based on human reasoning. There is only one output variable for the fuzzy controller which determines the rate at which energy must be exchanged with the battery during the next 15-minute interval. As represented in figure 4c, output variable fuzzy set has five membership functions called Discharge Heavily (DH), Discharge Lightly (DL), Zero Exchange (ZE), Charge Lightly (CL), and Charge Heavily (CH). The power drawn from the batteries can be used to help the renewable electricity generation unit provide the local load with required demand, can be sold to the main grid, or can be partially used for both reasons [13]. Table 3 represents the fuzzy rule-base for scenario four where Pr represents the variable electricity price, PR is the electricity generation rate by renewable resources including solar farm and wind farm, and PL stands for the value of local load while PB is the symbol of power exchange between storage unit and the rest of the nodes in the microgrid.
<table>
<thead>
<tr>
<th>#</th>
<th>Pr</th>
<th>Pa</th>
<th>Pr</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>L</td>
</tr>
<tr>
<td>2</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>5</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>6</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>7</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>8</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>9</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>10</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>L</td>
</tr>
<tr>
<td>11</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>L</td>
</tr>
<tr>
<td>12</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>L</td>
</tr>
<tr>
<td>13</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>14</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>15</td>
<td>IF</td>
<td>L</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>16</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>17</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>18</td>
<td>IF</td>
<td>M</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>19</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>L</td>
</tr>
<tr>
<td>20</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>L</td>
</tr>
<tr>
<td>21</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>L</td>
</tr>
<tr>
<td>22</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>23</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>24</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>M</td>
</tr>
<tr>
<td>25</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>26</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>H</td>
</tr>
<tr>
<td>27</td>
<td>IF</td>
<td>H</td>
<td>AND</td>
<td>H</td>
</tr>
</tbody>
</table>

The rules mentioned in Table 3 are determined based on human reasoning via expert knowledge. These rules or, in other words, the value of PB, can also be regarded as the decisions made by an expert operator of such system based on combinations of different values for input variables. This means that the fuzzy inference engine can replace an experienced system expert and makes appropriate control actions based on current situation of the system. According
to Table 2 there are 27 rules for the fuzzy system as expected, i.e. (Input variables) Membership functions = $3^3 = 27$.

The role of fuzzy inference engine is critically important for obtaining satisfactory results. For instance, an example of Mamdani-based inference rule for scenario four is as follows:

- IF the Price is Medium, AND the Renewable Generation Rate is Low, AND the Load is Medium, THEN the Battery should be Lightly Discharged.

Same rule, modified to be strictly environmentally-friendly based on air pollution constraints in scenario six, reads as follows:

- IF the Price is Medium, AND the Renewable Generation Rate is Low, AND the Load is Medium, AND the Air Pollution is High, THEN the Battery should be Heavily Discharged.

The same modified rule, where membership value of the air pollution antecedent is modified by a weighting factor based on the difference between current time and estimated peak pollution time in scenario five for enhanced control is represented in the following:

- IF the Price is Medium, AND the Renewable Generation Rate is Low, AND the Load is Medium, AND the Modified Air Pollution is High, THEN the Battery should be Heavily Discharged.

Modified membership value of air pollution measure in scenario five, $\mu_M^{C(t)}$, is obtained as represented in equations $\text{Eq. 7}$ to $\text{Eq. 9}$ during a 24-hour period:

$$\mu_{L, M}^{C(t)} = \xi \mu_L^{C(t)} \quad \text{Eq. 7}$$
\[
\mu_{H,M}^{C(t)} = (1 - \xi) \mu_{H}^{C(t)}
\]

Eq. 8

\[
\xi = \max(b_L, \min\left(1 - \frac{|t - t_p|}{\max(t_p, 24 - t_p)}, b_H\right))
\]

Eq. 9

where \(\xi\) is the weighting factor. Parameter \(t\) is current time, and \(t_p\) represents the estimated peak pollution time, based on look-ahead prediction approach [14], for current day. \(b_L\) and \(b_H\) are two constant factors used as lower and upper boundaries for \(\xi\) to avoid some of the membership values and consequently some rules to be neglected at the transit time instants between consecutive days or when the estimated peak pollution time arrives. In our simulations, we set \(b_L = 0.0001\) and \(b_H = 0.9999\).

By using the weighting regime introduced in Eq. 7 to Eq. 9 the importance of air pollution will be emphasized and exaggerated as the expected peak pollution time approaches which in-turn results in stricter limitations on charging the storage unit around the expected peak pollution time and gives more freedom and priority to discharging the storage unit when needed to help mitigate the contribution to air pollution by reducing the amount of CO2 emission.

The primary objective in these simulations is for the micro-grid to meet the local load profile at all times. Under low-price electricity conditions, the action decided by the rules might even sometimes require the micro-grid network to purchase energy from grid and store it in the battery unit regarding the fact that the electricity price is low. This consequently results in more degree of freedom for the system to sell energy to the main grid during high-price periods, even under cases of high local load demand. Hence, having feasible rules predefined for the fuzzy
system helps improve the cost function drastically. Also, the proposed approach may sometimes bring some profit to the micro-grid rather than making it pay to the utility companies.

3.5 Fuzzy Logic-based Decision-Making for Microgrid Energy Management

Simulation is done on the three bus system shown in figure 2 for the duration of one week. The Gauss-Seidel algorithm is implemented using MatLab® for power flow calculation [15]. Sample profiles are generated for electricity price rate, load demand profile, renewable electricity generation rate, and air pollution. The profiles for local load and the renewable electricity generation were generated based on available data for daily consumption of the University of Texas at San Antonio (UTSA) and also based on the daily generation of the solar panels installed on campus at UTSA respectively. Air pollution is updated using Eq. 5 and Eq. 6 section 3.4. For scenarios 2, i.e. section 3.2.2, to 6, i.e. section 3.2.6, resulting air pollution is compared to that of scenario 1, i.e. section 3.2.1, and the difference is represented as a measure called air pollution change, i.e. Pol. In the same fashion, peak pollution change, i.e. Peak, refers to the difference between the peak values of pollution during the one week period of simulation for scenarios 2, i.e. section 3.2.2, to 6, i.e. section 3.2.6, with that of scenario 1, i.e. section 3.2.1. Final diagrams represent unit-less measures of the cost function (including distribution losses, demand and supply), pollution change, and peak pollution change.

Normalized profiles of the four input variables, all or some of which are fed to the fuzzy controller in different scenarios, are shown in Figure 17 for a typical 24-hour period. Variables include electricity price, renewable electricity generation rate, local load demand, and air pollution. The data is generated arbitrarily for simulation purposes only considering similarity to the real world issues and with regard to the fact that the peak electricity consumption for the entire
region of interest of the main grid occurs around 7:30 PM where the electricity price reaches its highest value.

Simulation results for six scenarios introduced in section 3.2 are represented in Figure 17 to 22. In the simulations, nominal power generation capacity of the renewable plant is assumed to be equal to the local demand’s peak value. Figure 17 represents values of the four variables electricity price, electricity generation rate by renewable generators, local energy demand, and the air pollution measure for one week. In Figure 18, active exchange between the micro-grid and the utility at bus number 2 - which is the connection point between the micro-grid and the utility or main grid - is shown for six scenarios for the duration of the first day. In Figure 19, when each curve is positive, it means the electrical energy is delivered to the micro-grid from the utility at the rate represented and the micro-grid owner must pay to the utility companies for purchase of energy. However, whenever each curve is negative it means that energy is being delivered to the main grid from micro-grid’s side for that scenario at that time instant which results in some gain for the micro-grid owner.
In Figure 19, it can be seen that under scenario 3 where the micro-grid has electricity generation unit and battery storage without any intelligent control system applied, there are cases when the power flow exceeds the value 1 p.u. which with high probability will be undesired for the system. This happens because of the fact that there is no smart decision-making or control approach is used for the storage unit, i.e. the battery unit is pre-adjusted to start from an initial condition and get charged to its full capacity before it starts discharging the stored energy.
into the rest of the system. Hence, this can be concluded that applying an efficient control method to micro-grid is of utmost importance when storage unit is present in the network.

Factors of randomness and intermittency are associated with electricity price, local load, and renewable electricity generation rate in order to provide realistic situations for the simulation. Air pollution is updated after each time interval using Eq. 5 and Eq. 6. Air pollution boundaries and its concentration needs to be estimated downstream of the sources based on factors such as wind blow and the pollution production rate, etc. [14]. The air pollution measure considered in this study is the average amount of CO2 present in the air in the entire area of interest which includes surrounding space of the micro-grid plus the regions being supplied by the main grid.

Figure 20 represents the normalized curves indicating how much CO2 is added environment for different scenarios during a one week period. It should be mentioned that curves depicted in this figure only represent the effect of the micro-grid on air pollution through considering the power exchange between the main grid and the micro-grid itself, without taking into account any effects of the pollution removal terms such as Δr(k) as mentioned in Eq. 6. Results are also normalized in order to provide a relative view of the environmental effects of those six scenarios. As expected, scenario 1 where no renewable generation system and no batteries are involved has the worst effects on environment, and scenario 2 which includes only the renewable generation unit without any storage units in the micro-grid, has the best results in this regard. In scenario 1 there is only the local load present in the micro-grid, and whatever the local load demands must be provided by the main grid. Therefore, there will never be any flow of energy from micro-grid towards main grid which results in the amount of CO2 in the air to be always increasing gradually. However, in scenario 2, there are renewable generators besides load in the
micro-grid network. In this scenario, whatever amount of energy generated by the renewable generators must either be consumed by the local load or must be delivered to the main grid since there is no storage unit available in the system. This will result in less amount of CO2 added to the environment compared to the other scenarios since the huge amount of renewable energy added to utility grid utterly reduces utilities’ dependence on fossil fuel for electricity generation.

Figure 20 Normalized Micro-grid Contribution to Air Pollution; One Week; Six Scenarios

Figure 21 Normalized Battery Unit’s Stored Energy; First Day; Four Scenarios
However, scenario 2 does not bring as much environmental profit to the microgrid owner as the last three scenarios, i.e. 4, 5, and 6, do. Scenario 6, which employs the air pollution control in fuzzy inference engine, on the other hand, stands right after scenario 2 on reducing CO2 emission.

In scenario 6 there are lots of profit for the micro-grid owner and the consumers, though, which is not the case for scenario 2.

Normalized energy stored in battery unit is depicted in Figure 21 for the period day under four scenarios 3, 4, 5 and 6 in which the storage unit is utilized. Scenario 3, which does not employ any smart control on storage unit, uses the battery capacity to the full extent according to the predefined simple charge/discharge policy. Scenario 4 in which fuzzy intelligent control with no environmental constraints is deployed utilizes the battery capacity more than the other last two scenarios, since the highest priority objective in this scenario is to provide as much profit for the micro-grid as possible. Scenarios 6 in which air pollution reduction is considered as the high priority target alongside cost reduction, represents relatively less usage of battery storage capacity.

The scenario 5 - which does dynamic weighted compromise between the air pollution reduction and the cost reduction - uses the battery capacity relatively more than the scenario 6 does. Figure
Figure 22 represents the power flow to the storage unit for one typical day under four scenarios storage unit was employed. Whenever the value of each curve is positive it means that the flow of energy was towards storage unit, i.e. electrical energy was being stored and vice versa. The red curve shows the power exchange with battery in scenario 3 when no control method was used. As it can be seen from the diagram, there are only two states in which the battery is being either charged or discharged and they follow each other in a periodic manner which is the most basic control approach implemented on storage unit and is not desirable. The green curve shows the power flow to/from battery unit in scenario 6 where pollution control is done. As expected, this curve represents the least variance compared to the other three curves since the modification of fuzzy rule-base in order to reduce the air pollution growth rate results in less storage of energy in the battery unit in order to reduce the amount of energy drawn from main grid.

Eq. 10 represents the relationship between balance, distribution loss and the overall the network.

\[ \text{Balance} = \text{Cost} - \text{Loss} \quad \text{Eq. 10} \]

Where “Loss” stands for the overall sum of multiplication of the electricity price and dissipated energy on distribution branches, i.e. \( S_1(t) \), for all 15-minute periods. Loss will always be greater than or equal to zero and is given by:

\[ \text{Loss} = \int_0^T \text{Pr}(t) \cdot P_{loss}(t) dt \quad \text{Eq. 11} \]

“Cost” is calculated using Eq. 14 and represents the monetary value that the microgrid owner has to pay to the main grid, if Cost > 0, or will get from the main grid, if Cost < 0. “Balance” will then be the measure of the money that microgrid owner had to pay to the main grid, i.e. Balance.
> 0, or the profit that micro-grid owner will get from the main grid, i.e. Balance < 0, in case the power network were lossless or the network losses could be neglected.

The center of gravity, i.e. centroid, defuzzification is used for computing crisp values of the output variable from union of the curves obtained by fuzzy rules as represented in Eq. 12

\[
y_{\text{crisp}}(k) = \frac{\sum_{i=1}^{n} (\max_j (\mu_{kij}) \times y_{ki})}{\sum_{i=1}^{n} \max_j (\mu_{kij})}
\]

Eq. 12

Where \( y_{\text{crisp}}(k) \) represents the energy exchange rate with the storage unit determined at the beginning of \( k \)th time interval, \( i \) refers to the number of discrete points in the universe of discourse of output variable, which is between 1 and \( n \). The factor \( j \) changes between 1 and the overall number of rules which in this case is \( 3 \times 3 \times 3 = 27 \) for scenario 4, and \( 2 \times 3^3 = 54 \) for scenarios 5 and 6, and represents the number of rule curves for each of which we consider the membership value of \( i \)th point in the universe of discourse of the output variable. In other words, \( \mu_{kij} \) represents the membership value of \( i \)th point of output variable’s universe of discourse in the \( j \)th fuzzy rule for \( k \)th time interval.
Figure 23 Measures of Balance, Pollution, Peak Pollution; Six Scenarios using Fuzzy Approach

Figure 23 represents the three final measures of balance, pollution change, and pollution change for all six scenarios obtained for the simulation of one week operation of the micro-grid. Scenario 1 is considered as the reference to which the pollution changes and peak pollution changes of all other scenarios will be compared. Therefore, the amount of pollution change and peak pollution change for the first scenario are assumed to be zero as the base of comparison, as represented in the diagram of Figure 23. The pollution measures represent much other scenarios contribute in increasing or decreasing the CO2 emissions compared to scenario 1.

The fitness values represented in Figure 23 for the scenarios four, five and six the measure of their fitness to the whole strategy based on fitness function given in Eq. 13 will later be introduced in details in Section 3.6.
As it can be seen in Figure 23, in the first Scenario the network cost is at its
which is expectable since there is only local load and the main grid. Therefore, the consumer must
only pay to the main grid in order to purchase the electrical energy required to provide the demand.
Scenario 4 where real-time intelligent fuzzy decision-making approach is used with no
environmental constraint brings the most profit to the micro-grid owner. This profit will be higher
as the value of balance measure drops more and more below zero, i.e. as the Balance value becomes
negative. Literally, negative balance can be interpreted as credit which is the definition the authors
have used in the diagrams of cost and balance measures in this article. On the other hand, Scenario 6 - where the fuzzy decision-making method with environmental considerations is incorporated -
results in better modification of air pollution compared to the Scenario 4, at the cost of losing some
small portion of profit. This means that in Scenario 6, better results for air pollution control will
be achieved while the cost reduction will be relatively degraded compared to Scenario 4. However,
Scenario 5 in which enhanced fuzzy approach, i.e. the weighted fuzzy based on pollution peak
instance estimation, is deployed with environmental constraints gives satisfactory results which
may be regarded as a compromise between the two Scenarios four and six. In Scenario 5, the air
pollution control is very close to Scenario 6 result and the profit value is also very close to the
profit measure of Scenario 4.

3.6 Intelligent GFS-based Decision-Making for Microgrid Energy Management

With the facts explained about figure 11, the need to have a merit function in order to get
a uniform measure of each scenario’s performance lead to the idea of using genetic algorithms as
a DNA-based evolution method [16, 17] in combination with fuzzy decision-making framework
in order to enhance performance of the proposed approach and get sub-optimal settings for fuzzy
sets of the input variables. Hence, cost reduction and pollution modification could be achieved
simultaneously without losing any of the objective functions due to over-emphasis on some other portions. Figure 24 represents the overall structure of the combined GA-fuzzy approach making based energy management and control in the micro-grid networks. The block diagram represented in Figure 24 can also be regarded as the flowchart of the genetic fuzzy system implemented to do intelligent decision-making for microgrid energy management.

---

**Figure 24 Block Diagram of the GFS-based Decision-Making Framework**
The first generation will be produced based on some random process. The next generations will be produced based on evaluation and selection process, according to which the best fits, i.e. the strongest creatures, will be present in consecutive generations, plus some genetic evolution which includes two major processes of cross-over and mutation. After a generation is formed, for each offspring, the fitness values for three last scenarios are computed and compared to each other for the highest fitness to be determined.

Then, among all population of a generation, the highest fitness values of different offsprings will be compared and the off-springs will be sorted in a descending order based on the value of their fitness. A selection process will take place based on the fitness values. The best ones will be transferred to the next generation plus some of the medium members and very few of the population with lower fitness values. After the selection process which determines a portion of the population for the next generation, some new off-springs will be generated based on cross-over and mutation processes applied to the members of current generation. These new off-springs will complement the preselected members in order to form the next generation. After the last generation is evaluated, the best off-spring will be chosen as the sub-optimal solution. In this study, the genetic algorithms were used for creating different combinations of fuzzy sets for the input variables of the fuzzy controller.

Eq. 13 represents the fitness function used for the genetic algorithms approach which be regarded as the merit function of the whole system in order to evaluate performance of each different scenario:
\[ F_{m_s}^g = \left( \max_g \left( \max_m \left( \max_s \left( \text{Bal} \right) \right) \right) - \text{Bal}_{m_s}^g \right) / \left( \max_g \left( \max_m \left( \max_s \left( \text{Bal} \right) \right) \right) \right) \]

\[ = -\alpha \left| \frac{\min_g \left( \min_m \left( \min_s \left( \text{Pol} \right) \right) \right) - \text{Pol}_{m_s}^g \right|}{\min_g \left( \min_m \left( \min_s \left( \text{Pol} \right) \right) \right)} \]

\[ = -\beta \left| \frac{\min_g \left( \min_m \left( \min_s \left( \text{Peak} \right) \right) \right) - \text{Peak}_{m_s}^g \right|}{\min_g \left( \min_m \left( \min_s \left( \text{Peak} \right) \right) \right)} \]

Eq. 13

Where \( F_{m_s}^g \) is the fitness value for scenario “s” of off-spring “m” in the generation “g”. The variable \( s \) varies between 4 and 6, where \( s=4 \) refers to scenario four, \( s=5 \) implies scenario five and \( s=6 \) stands for the last scenario, i.e. scenario six. The generation number, \( g \), changes between 1 and \( G \) where \( G \) is the total number of generations considered in the genetic algorithms. The offspring number, \( m \), changes between 1 and \( M \) where \( M \) is the number of population per generation.

The term \( \max_g \left( \max_m \left( \max_s \left( \text{Bal} \right) \right) \right) \) refers to the maximum value of Balance measure for every scenario of all populations in the entire generations. This maximum value occurs at scenario 1 where the only element in the network is the local load besides the main grid. The interesting point is that regardless of the off-spring or the generation, this maximum value is always the same since similar price and load profiles are used for homogeneous and uniform evaluation of various fuzzy system set-ups and fuzzy sets' characteristics. Hence, the maximum Balance value was already at hand and consistent and was equal to 232. Term \( \text{Bal}_{m_s}^g \) represents the Balance measure for scenario “s” of off-spring “m” in generation “g”. The terms \( \min_g \left( \min_m \left( \min_s \left( \text{Pol} \right) \right) \right) \) and \( \min_g \left( \min_m \left( \min_s \left( \text{Peak} \right) \right) \right) \) represent the minimum value of the pollution measure and the minimum...
value of the peak pollution measure respectively among all different scenarios for every off-spring of the entire generations. Fortunately, these values were also already at hand and consistently occurred at scenario 2, where the only two elements of the micro-grid are local load and renewable generators alongside the utility grid, and were equal to -3930 and -154 respectively. We chose the values of $\alpha$ and $\beta$ to be 2 and 4, respectively, so that more weight was given to the peak pollution reduction which is in most cases more important than the overall pollution reduction. Also, the air pollution is itself more critical and relatively more difficult than the cost to reduce, hence using the factors $\alpha$ and $\beta$ will compensate effort.

The genetic algorithms was implemented using 20 generations, i.e. G=20, with population of 10 per generation, i.e. M=10, with each off-spring having 4 chromosomes corresponding to the fuzzy system’s four input variables mentioned in 3.2. The chromosomes included appropriate number of cells in order to determine the edges and centers of the fuzzy sets for the relevant variable. Each chromosome for fuzzy sets of the variables price, renewable generation rate, and local load, included 9 cells to cover the edges and centers of its three membership functions. Air pollution index which has two membership functions - as represented in figure 4 - requires six cells per chromosome.

The best measures of Balance, Overall Pollution Change, and Peak Pollution Change obtained by the GFS-based approach are represented in Figure 25. Fitness value of the scenario, i.e. scenario 6 which focuses on pollution control and cost reduction simultaneously and equally, is the best of all based on Eq. 13.
Figure 25 Measures of Balance, Pollution, and Peak Pollution; Six Scenarios using GA-Fuzzy Approach

Results represented in Figure 25 also confirm the fact that maximum Balance occurs at scenario 1 and is of the same amount, i.e. 232. This figure also shows that the minimum, i.e. best, amount of air pollution change and also peak pollution change happens when scenario 2 is employed. The minimum values of unit-less measures for pollution change and peak pollution change are -3930 and -154 respectively.
Figure 26 represents the fuzzy sets of the four input variables for the best offspring obtained by combined GA-fuzzy approach. It can be seen that there are different fuzzy sets for each of the three variables price, renewable and local load. If we compare the three diagrams of Figure 26 with the diagram of Figure 16, where the three mentioned variables had similar fuzzy sets, the difference will be clearly revealed. The diagrams represented in Figure 26 show the edges and centers of the fuzzy sets for the best fit offspring of the GFS-based decision-making. This means that the control and energy management scheme will give the best results for these input variable fuzzy sets based on the fitness function of Eq. 13.

3.7 Numerical Example for Evaluation of Intelligent Energy Management

For any practical system of the same structure as mentioned for the micro-grid in this article, the actual numeric results can be obtained after replacing the variables in this study by the...
actual values of the practical system itself. As an example, if the peak load and the nominal power rating of the renewable electricity generation plant both are assumed to be 1 MW, i.e. $P_{L,max}(t) = 1 \text{ MW}$ and $P_{R,max}(t) = 1 \text{ MW}$, and the maximum electricity price is presumably 50 cents per kWh energy, i.e. $P_{r,max}(t) = 0.5 \left( \frac{\$}{\text{kWh}} \right)$, the resulting Balance values would be as represented in Table 4.

Table 4 Weekly Balance Values Obtained by the Six Scenarios for the Numerical Example

<table>
<thead>
<tr>
<th></th>
<th>Scn. 1</th>
<th>Scn. 2</th>
<th>Scn. 3</th>
<th>Scn. 4</th>
<th>Scn. 5</th>
<th>Scn. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bal.</td>
<td>$29043$</td>
<td>$3466$</td>
<td>$4917$</td>
<td>$-13761$</td>
<td>$-11264$</td>
<td>$-12204$</td>
</tr>
</tbody>
</table>

The balance values represented in Table 4 are the monetary values that microgrid owner has to pay to the utility companies or will receive from them. Positive values refer to a payment, and negative numbers represent the microgrid owner’s gain. The resulting balance values are based on a one-week simulation period. This means, assuming identical gains for a total of 52 weeks of the year, scenario 4 will bring a $716,000 gain to the microgrid using the GFS-based decision-making approach for energy control. On the other hand, in scenario 3 which has the same microgrid structure as that of scenario 4, without any intelligent control algorithms applied to the storage, the owner has to pay $256,000 to the electric utility for one year. Comparing results of scenarios 3 and 4 leads to the fact that using the GFS-based decision-making framework proposed in this research yields an annual benefit of around $970,000 for the microgrid owner. Approximating the cost of implementing such a decision making tool for energy management to be around $100,000 for the size of micro-grid described in this numerical example, this brings a Return-on-Investment (ROI) of approximately 10. It would be conservative to say that microgrid
owners looking into adopting intelligent decision-making tool for energy storage management may see an ROI between 5 and 10.

3.8 Summary

An intelligent GFS-based framework for decision-making and energy flow control in micro-grids through adjusting energy exchange rate of storage unit is introduced and simulated for cases with and without environmental considerations. Six microgrid structure and operation scenarios were considered. If an appropriate smart decision-making approach, such as the one presented in this article, is employed to manage energy storage, then adopting measures of cost and pollution will help reduce the net electricity consumption costs and increase the monetary gains for microgrid owners. Simulation results reveal the fact that considering air pollution control policy within the fuzzy inference engine and modifying the fuzzy rules accordingly will result in less pollution compared to all other intelligent control scenarios addressed in this study. While the pollution control scenario, i.e. scenario 6, helps reduce air pollution most significantly, the net cost of this scenario will be larger in comparison to the intelligent control scenario, i.e. scenario 4, which brings the highest financial benefit to the microgrid owner. The enhanced weighted fuzzy approach, i.e. scenario 5, turned out to be a compromise between the two cases of intelligent control, scenario 4, and pollution control, scenario 6. This means enhancing the fuzzy system by a weighting regime based on the predicted peak pollution time for the current day, the controller represents a balanced behavior of both intelligent control and pollution control. Hence, both objectives of cost reduction and air pollution modification could be achieved if the rules are modified accordingly and also enhanced by weighting factors.

The proposed GFS-based intelligent decision-making framework proved capable of cost-efficient energy management in microgrids under constraints of meeting the demand profile,
meeting battery SOC safety conditions, handling renewable electricity intermittency, and complying with environmentally-friendly considerations. Results suggested that the use of an intelligent decision-making approach for control of energy flow in the electrical microgrids where storage units are already deployed will bring an ROI of between 5 to 10 for the microgrid owner while giving the leverage to meet the environmental standards by reducing the amount of contribution to air pollution through enhanced rule-base if needed. In the next chapters the data analytics-based adaptive forecasting framework will be introduced and its application in solar energy forecasting will be considered. The performance and the added value that the forecasting of solar energy may bring to the utility companies will be analyzed.
CHAPTER 4: DATA ANALYTICS-BASED ADAPTIVE SOLAR FORECASTING

4.1 Current State-Of-The-Art in Solar Energy Forecasting

A report performed for the Regents of the University of California titled “Current State of the Art in Solar Forecasting” [4] best summarizes current status of the solar energy forecasting field of research. In the report Glassley et al. state that satellite and numerical weather prediction (NWP) are the preferred methods for longer duration solar forecasting (one hour to a few days) [4, 5]. As for intra-hour forecasts, sky imagery based forecasting methods are most common, but, unfortunately, several basic assumptions regarding cloud shape and linear cloud movement vectors reduce the potential accuracy of this type of forecasting [6, 7].

Making intra-hourly or site-specific forecasts using satellite data is also not very common due to the infrequent sampling interval (30 minutes) and the low image resolution [4]. This problem is only increased in the NWP method due to a larger sampling interval and lower cloud imaging precision. The limitations of NWP models do not allow for shorter time-scale and smaller spatial sampling to be accounted for [8]. Some of the error associated with the aforementioned problems can be corrected through a method known as Modeled Output Statistics (MOS), which determines statistical correlations in observed weather data and related imagery (satellite, sky imagery, etc.) [4, 8, 9, 10]

Practical methods for day-ahead up to week-ahead solar forecasting methods are mostly based on NWP and statistical approaches as already mentioned [4, 11, 5, 10]. A number of models are available for this purpose, among them the Global Forecasting Service and the European Center for Medium Range Weather Forecasting (ECMWF) [12] both of which are considered the state-
of-the-art of global forecast models. However, in order to increase spatial and temporal resolution of these models, other models have been developed which are generally called mesoscale models among which, the High Resolution Limited Area Model (HIRLAM) and the Weather Research and Forecasting (WRF) are widely used by different communities. A broad range of in-depth expertise is needed in order to obtain accurate results when running these models, due to the wide variety of parameters that can be configured. Also, sophisticated techniques such as data assimilation or statistical post-processing can be used in order to obtain a probabilistic point of view of the accuracy of the output. Usually techniques that mix outputs of different models are used for post processing which finally provide a better estimate of those variables along with a degree of uncertainty associated with the forecast results [7].

Table 5 Characteristics of current practices in solar energy forecast [Bing 2012]

<table>
<thead>
<tr>
<th>Technology</th>
<th>Time Horizon</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite</td>
<td>12 hrs to 7 days</td>
<td>Global</td>
</tr>
<tr>
<td>Meso-scale NWP models</td>
<td>12 hrs to months</td>
<td>Global / Regional</td>
</tr>
<tr>
<td>Aggregated Ground Sensors</td>
<td>1 hrs to 3 hrs</td>
<td>Regional</td>
</tr>
<tr>
<td>Sky-Imager</td>
<td>30 minutes to 3 hrs</td>
<td>2 to 10 km radius</td>
</tr>
<tr>
<td>Array Scale Sensors</td>
<td>1 to 30 minutes</td>
<td>Array Size</td>
</tr>
</tbody>
</table>
Table 5 summarizes the characteristics of current approaches to solar forecasting industry [Bing 2012]. It can be inferred from Table 5 that the NWP and satellite imagery-forecast techniques respectively are suitable for the longest and the second longest forecast time horizons, have the largest area coverage in solar energy forecast.

NWP discretize space and time at the mesoscale/regional and national levels and solve the equations of motion, thermodynamics and mass transfer at pre-defined time and space intervals [7-ISJ]. The problem with NWP is that it cannot be considered site-specific to any plant but based on regional commingled datasets and interpolation techniques. Such commingled datasets propagate error from node to node and from time step to time step and its uncertainties are still too large, in the order of mean absolute error (MAE) of 20%-40% and root mean square error (RMSE) of 30%-50% for day-ahead predictions [7-ISJ-11-ISJ]. Recently, smart forecasting of expected solar energy by modeling of weather and climate events through training of systems based on artificial intelligence, specifically techniques that involve neural networks, support vector machines, etc. has become more popular [13, 14, 15, 1-ISJ,3-ISJ,4-ISJ,5-ISJ,12-ISJ,13-ISJ,REF1-ISJ – REF6-ISJ]. The results of such modeling approaches can be extensively used in solar forecasting. However, the aforementioned machine learning-based approaches still need to improve the accuracy of their forecast engine, better stabilize and further fine-tune the forecast results in order to be comparable with already well-established NWP-based forecasting methods that have been in the market for the past decades.

NWP and satellite imagery-based forecast techniques are the two most versatile approaches for longer forecast horizons, i.e. have largest temporal capacity. The mentioned approaches also have the largest coverage, i.e. biggest spatial capacity, in solar energy forecast.
Regarding the Modeled Output Statistics (MOS) for a solar forecasting approach, regardless of the time scale, a database of instrumentation characteristics is necessary [9]. Unfortunately, the instrument characteristics database by itself does not translate to improve accuracies because the MOS solar forecasting model would need accurate correlations between weather elements or the instrumentation variables measured, and these would be delegated to a statistical model, which often includes outlier data that do not improve forecasting results [4, 10]. Furthermore, a statistical model is not learning from the decision-making and rule policies developed by its users/operators over time.

In this chapter a novel data analytics-based adaptive approach to solar energy prediction will be presented based on a combination of data mining techniques, statistical analysis and artificial intelligence.

4.2 Forecasting Framework Block Diagram and Flow Chart

The proposed solar energy forecasting framework is adaptively implemented based on correlation analysis and feature-extraction of weather and climate data. This is done in combination with artificial neural networks and synoptic systems detection algorithm which also takes advantage of a knowledge-base library that is ever growing in size and intelligence.
Figure 27: Solar Forecast System Input-Output Structure

Figure 27 represents the basic input-output setup of the proposed approach to forecasting in a black box diagram. The proposed forecast framework, as represented in Figure, is capable of giving estimations of solar irradiance of the next “k” sample times based on current time data and the extracted feature-matrix of the data recorded in the past.
Figure 28 Block Diagram of the Proposed Solar Energy Forecasting Framework

Figure 28 represents the comprehensive block diagram schematic of the proposed solar forecast framework. This block diagram can also be regarded as a flow chart of the forecast process. As it can be seen in the block diagram of Figure 28, the data coming from the sensors includes the weather and timing data together with the solar plant output, i.e., the amount of DC power generated by solar array. These sensory data is added to the existing data base. Monte Carlo statistical uncertainty analysis is performed. Fine-tuning and convergence check are performed. Units that store, transfer, present or display data are present. Control commands and switching signals are also present.
Carlo approach to uncertainty analysis requires a number of data sets similar to the recorded data base to be generated taking into account the uncertainty associated with sensory measurements, such as measurement noise and hardware failure or calibration issues, so that the output of the forecast engine can be stabilized in order to obtain a reliable forecast result. The input data is delayed, with respect to the plant output, based on the forecast horizon.

A number of training subsets with different lengths and coverage of training windows are formed, considering the results of synoptic events detection algorithm, and a consecutive training and test process will take place for the target forecast horizon using each and every one of the training subsets. The process goes on by passing each training subset through a correlation analysis where the correlation function between temperature, pressure, relative humidity and GHI input with the plant output, i.e. the generated electrical power or the on-site measured GHI, is computed after using low-pass filters of lengths from 1 minute to 3 hours on the mentioned input variables. After the best filter length and shift values, i.e. the values that maximize the correlation between the mentioned variables and the plant output, are determined for each and every one of the input variables mentioned above the rates of change (ROC) of pressure, temperature, relative humidity and the GHI are also computed as represented in Figure 3 in the ROC blocks. These ROC’s represent gradients in the atmospheric variables and constitute precursors of changes in other variables which indicate development or dissipation of clouds. The first and the second derivatives of the above-mentioned four atmospheric variables are considered as the ROCs.

The best fit variables obtained by correlation function analysis is fed together with the ROC’s, i.e. 17 variables in overall, to the feature extraction block which transforms the data into a new domain using Independent Components Analysis (ICA) in order to find the underlying independent features of the variables. These independent components or features will then be used
for training of the artificial neural network (ANN). After the predefined convergence conditions for the ANN are met control commands are sent to the connectors to switch from training phase to the test phase. In test phase, as seen in the lower part of the block diagram, results of dynamic performance evaluation will be considered for final aggregation of forecasts. As the first step in this section, the select variables of real-time data obtained from the plant, D1(t), are shifted and filtered using the best set of filter length and shift value, that are obtained from correlation analysis, for each and every training subset. Then, the rate of change for weather variables is computed. Then, the shifted and filtered variables are fed to the feature computation block together with the ROC’s of weather data. The data input to the feature computation unit is transformed into the independent components’ domain using the separation matrix W(t) obtained for each training subset, by applying the ICA approach to the same training subset. The computed features will then be fed to the ANN for the prediction of solar energy to take place. The output of the ANN will go through the Monte Carlo statistical uncertainty analysis for each and every training subset so that the percentiles can be determined. The fiftieth percentile, p_{50}, obtained for each training subset is regarded as the stabilized raw forecast result. For each single target day, the forecasts percentiles obtained by different training subsets will be aggregated based on a weighted averaging regime in order to form the final percentiles. The final percentiles will be then fine-tuned according to a set of rules extracted from a study performed on the weather and atmospheric variables influential on cloud formation patterns taking into account results of synoptic events detection algorithm. The fine-tuned signal is ready to be used by the grid operators and will be fed into the electrical grid energy management user interface for electric energy bid decisions into the day-ahead market (DAM) by the energy market operations (EMO) of utility companies.
In the following the major steps towards implementation of the proposed solar energy forecast framework, represented in Figure 28, is addressed as they form the organization of the rest of Chapter 4:

Section 4.3 introduces the big data base used for solar forecasting framework.

Section 4.4 discusses the algorithm of weather synoptic events detection.

Section 4.5 describes the procedure of correlation function analysis.

Section 4.6 explains the feature extraction process and the impact.

Section 4.7 explains the artificial neural network, adaptive training and test.

Section 4.8 outlines the uncertainty analysis based on Monte Carlo approach.

Section 4.9 demonstrates the dynamic performance evaluation and its outcomes.

Section 4.10 talks about the aggregation process and fine-tuning of forecast results.

Section 4.11 discusses the forecasting interface for the utility operator.

4.3 Data description
The National Renewable Energy Laboratory (NREL) archive, the nation’s largest combined climate and solar energy data set, is the reference big data set used for implementation of the solar energy forecast framework along with the Automated Surface Observing System (ASOS) information obtained from Iowa Environmental Mesonet (IEM) [16, 17, 18]. Data from more than 250 instruments, variables and readings were collected, including GHI, DHI, DNI, solar zenith and azimuth angles, weather and atmospheric variables such as pressure, temperature, relative humidity, wind speed, wind direction, cloud cover percentages including Total Cloud Cover (TCC) and Opaque Cloud Cover (OCC), and many more [1].

Figure 30: Removing the Undesired Data for Solar Energy Forecast; (a) Snapshot of NREL library in the form of a spreadsheet; 250+ variables sampled every 1 minute for 8 years (b) Scaled diagrams of three variables for 8 years; Temperature – RED, Pressure – GREEN, and GHI – BLACK
variables for the duration between 2004 and 2012 sampled at 1 minute intervals. (a)

represents a snapshot of the data set in the form of a spreadsheet including more than 250 variables for the duration between 2004 and 2012 sampled at 1 minute intervals. (a)

the GHI over the period of eight years. In (a)

shows diagrams of the three variables atmospheric temperature, atmospheric pressure and the GHI over the period of eight years. In (a)

single diagram only for representation purposes.

values of temperature, pressure and GHI are scaled to fit within the same range in a single diagram only for representation purposes.

Figure 31: Structure of the NREL Data Set; 8 years worth of data used for ANN training
One critical thing when working with a big data set such as that of NREL, as can be inferred from figures 4 and 5, is the know-how of handling large data sets and how to extract the meaningful information out of the pool of measurements and the recorded variables.

Some of the most important challenges of dealing with large data set are as follows:

- Removal of measurement noise
- Addressing sensor failures to eliminate invalid data and outliers
- Choosing the most information-rich part of the data set for efficient training
- Handling the growing number of new variables and measurement samples to keep consistency of the framework and to maintain the performance

In order to deal with the above-mentioned challenges, first, a solid set of variables was determined to be used for further simulations. Techniques were used to prepare the data for development of the forecasting framework. These techniques include elimination of invalid portions of data and replacing the outliers by valid intrapolated values, filtering of each and every variable with an appropriate low-pass or moving average filter, as represented in Figure 32. Irrelevant or temporary invalid spot measurements were also removed or replaced with reasonable estimates through either extrapolation of the previous samples or a random deviation introduced to the last valid sample both of which were quite appropriate since the sampling interval was 1 minute.
Choosing the best set of variables and recognizing the most information-rich part of the training subset are critically important for achieving reliable and accurate forecasts. The data set used for this purpose consists of weather and climate variables and sensory data of solar irradiance and plant output together with data related to positioning of the Earth with respect to Sun and the timing data.

4.4 Weather Synoptic Events Detection

Synoptic weather events occur in a periodic fashion but like within the nature of the turbulence behavior, these events are regarded as to be random and stochastic in essence. These events may seem random when analyzed in a historical statistical approach. However, when studied in a time-dependent manner, the time-varying coherence of those synoptic systems and the effects of individual variables and their features on atmospheric cloud formation and eventually their influence on the solar irradiance can be established.

The procedure for detection of weather synoptic events is based on locating the low pressure weather systems which happen in between two consecutive points of peak pressure that

Figure 32: Smoothening of Data Set; Original Variable (RED), Low-Pass Filtered Variable (BLUE)
meet certain conditions. Two basic criteria are introduced to distinguish a fully developed synoptic system from a weak one as established in [19].

Figure 33: Weather Synoptic Events Detection Process Block Diagram

As represented in Figure 33, the first and the second derivatives of the filtered through a set of rules in order for the minimum and maximum points to be figured out. Finally, based on the fact that what portion of the filtered curve of reference variable does each day fall within, types of consecutive days is determined which are directly used for the purpose of adaptive training of the ANN which is explained later in section 3.6. The outcomes of this algorithm including the types of days and the extrema points are also used in order to extract the rule-base for the process of fine-tuning which is discussed in section 3.9.

Mathematical representation of the filtering and derivation steps of the procedure for synoptic events detection is as follows:

\[
f(t_i) = \frac{\sum_{j=\min(1,\frac{i}{L})}^{\min(N,1+\frac{i}{L})} v(t_j)}{\min(N,i + \frac{1}{L}) - \max\left(1, i - \frac{\left\lceil \frac{1}{L} \right\rceil}{2}\right)}
\]

Eq. 14
\[
f'(t_i) = \left. \frac{df(t)}{dt} \right|_{t=t_i}, \quad f''(t_i)
\]

Eq. 15

\[
= \left. \frac{d^2f(t)}{dt^2} \right|_{t=t_i}
\]

Where \(i = 1,2,\ldots,N\) is the sample number. \(N\) and \(L\) respectively represent the number of samples for the input variable, \(v(t)\), and the low-pass filter length which in-fact works like a moving window used for averaging. The rules used to determine the points of maximum or minimum values for the input variable are as follows:

- IF \(f'(t_i) = 0\) AND \(f''(t_i) < 0\) \(THEN\) \(t_i\) is a MAX
- IF \(f'(t_i) = 0\) AND \(f''(t_i) > 0\) \(THEN\) \(t_i\) is a MIN

After the MAX and MIN points are found the next step is to determine high, i.e. Hi, and low, i.e. Lo, points of the input variable based on the criteria (a), (b) and (c) which are given in the following:

(a) If the difference in value of a MAX with its adjacent MIN is less than 1 unit the MAX point is neglected.

(b) If any two consecutive MAX points are less than 36 hours away the largest is defined as Hi, and the other local MAX is neglected.

(c) The smallest MIN point between each two consecutive Hi’s is considered as a Lo.

Two different signals are used for extracting the rules of climate change. The first signal is the atmospheric pressure which is used for synoptic events detection. The second signal is calculated based on the ratio of the three variables of atmospheric pressure, temperature and relative humidity, i.e. \(\frac{P \times T}{RH}\). When both of the signals are rising towards their Hi points the
atmospheric weather mostly behaves in a stable manner and little to no cloud formation can be expected.

Error! Reference source not found. represents the result of synoptic event detection algorithm based on the pressure signal and the atmospheric reference signal, i.e. \( \frac{P \times T}{RH} \), for a portion of the year 2012 between day 262 and day 312. It must be noted that in Error! Reference source not found., the pressure signal is vertically expanded by making its standard deviation 10 times the actual value only for better representation purposes.

![Figure 34 Weather Synoptic Event Detection Results; Atmospheric Pressure Signal (GREEN), Meteorological Reference Signal (YELLOW), Hi’s (RED) and Lo’s (BLUE)](image)

Four rules which are used to define types of days are as follows:

- IF day \( d_j \) has a \( t_j = \text{MAX} \) AND \( t_j \) meets the criteria (a), (b), (c) THEN \( d_j \) is Hi
- IF day $d_j$ has $t_i = \text{MIN}$ AND $t_j$ meets the criteria (a), (b), (c) THEN $d_j$ is Lo
- IF day $d_j$ is after a Lo AND $d_j$ is NOT a Hi THEN $d_j$ is a Rise
- IF day $d_j$ is after a Hi AND $d_j$ is NOT a Lo THEN $d_j$ is a Fall

Figure 35 below shows the types of days obtained based on the synoptic events algorithm implemented on the two signals of pressure and atmospheric reference $\frac{P \times T}{RH}$ for the same period of the year 2012 represented in Figure 34.

![Figure 35: Type of Day (TOD) based on Pressure Signal and Atmospheric Reference P*T/RH](image)

The type of day (TOD) represented in Figure 35 are used for extracting the rules which the fine-tuning of forecast results is implemented. Those rules are obtained for climate change by the authors after studying the three main components of cloud formation including pressure, temperature, and relative humidity.
4.5 Correlation Function Analysis

Correlation analysis and the algorithm for weather synoptic systems detection are the two first steps towards adding intelligence to the proposed solar forecasting solution and making the system perform in an adaptive way. The correlation function analysis is implemented in the time domain and, as represented in Figure 28, consists of a filtering and shifting block, in loop correlation analysis unit, followed by the smart training subset detection algorithm.

In the first stage of serial time-domain analysis, input data of temperature, atmospheric pressure, relative humidity and the GHI is passed through filters of different lengths and various shifted version of these filtered data is generated so that the correlation function for the mentioned four variables with the output of the solar plant over the training subsets are obtained. The characteristics of filter and shift for which peak cross-correlation between each of the four select input variables and the plant output occurs are used for modifying the input variables accordingly both in the training and test phases so that the training set is better aligned and becomes information-rich for prediction purposes. The correlation table of variables for original input data with no filtering of shifting applied to the variables is given in Figure 36 where the matrix represent the correlation coefficient, i.e. “r”, between each two pairs of variables.

\[
\text{abs}(M_{corr}) =
\begin{bmatrix}
\text{TA} & 1.0000 & 0.5605 & 0.5473 & 0.9910 & 0.0567 & 0.2922 & 0.1703 & 0.0566 & 0.7469 \\
\text{TCC} & 0.5605 & 1.0000 & 0.5741 & 0.5614 & 0.0160 & 0.0440 & 0.1458 & 0.0503 & 0.1967 \\
\text{OCC} & 0.5473 & 0.5741 & 1.0000 & 0.5440 & 0.0165 & 0.0390 & 0.1383 & 0.0343 & 0.1811 \\
\text{Zenith} & 0.9910 & 0.5614 & 0.5440 & 1.0000 & 0.0016 & 0.2812 & 0.1563 & 0.0390 & 0.7367 \\
\text{Azimuth} & 0.0567 & 0.0160 & 0.0165 & 0.0016 & 1.0000 & 0.1600 & 0.0768 & 0.0492 & 0.0035 \\
\text{Temp} & 0.2922 & 0.0440 & 0.0390 & 0.2812 & 0.1600 & 1.0000 & 0.8694 & 0.4923 & 0.4025 \\
\text{RH} & 0.1703 & 0.1458 & 0.1383 & 0.1563 & 0.0768 & 0.8694 & 1.0000 & 0.5283 & 0.3041 \\
\text{Pres} & 0.0566 & 0.0503 & 0.0343 & 0.0290 & 0.0492 & 0.4923 & 0.5283 & 1.0000 & 0.0984 \\
\text{GHI} & 0.7468 & 0.1967 & 0.1611 & 0.7367 & 0.0035 & 0.4025 & 0.3041 & 0.0984 & 1.0000
\end{bmatrix}
\]

Figure 36: Absolute Value of Correlation Matrix for Original Variables
It can be inferred from the absolute values of original correlation matrix that the following pairs of variables are strongly correlated during the year 2012:

- Time Angle and Zenith Angle ($r = 0.991$)
- OCC and TCC ($r = 0.9741$)
- Temperature and Relative Humidity ($r = 0.86941$)

Figure 37 represents the correlation functions for nine input variables with the GHI of the next day for the year 2012. The correlation coefficient values between each and every variable with the GHI are represented versus different amounts of shift and using different filter lengths applied to the input variables. Some of the weather and climate variables can play the role of drivers for cloud formation which in turn affects the amount of solar irradiance that reaches the ground. For instance, atmospheric pressure works as the driver for the weather synoptic system so that a high pressure system is regarded as the sign of a stable climate and a low pressure system can be followed by an unstable, stormy weather or overcast and highly cloudy days. Hence, by continuously computing the correlation functions between each of those variables with GHI over the training window, for each consecutive target day, we can not only obtain the refined training data set with respect to optimally correlated variables but also use the information obtained by some of those climatic variables in order to detect the weather synoptic systems beginnings and ends. Using the weather synoptic events detection algorithm [19] and finding its statistical information, the forecasting framework is able to adaptively determine the size of training window by eliminating portions of original training window which do not seem to be playing important role in forecast of the next day’s solar energy. This way, only the most information-rich part of the
training subset will be used for training of the forecast framework which is expected to result in more accurate predictions compared to training over a constant length of previously recorded data.

Figure 37: Cross Correlation Functions of Nine Input Variables with GHI Using Different Filters; (a) Time (b) Zenith (c) Azimuth (d) Temperature (e) Pressure (f) Relative Humidity (g) GHI (h) TCC (i) OCC
The process of correlation function analysis takes place continuously for each and every target day in order to find the low-pass filter length and the shift amount, both in number of samples, which result in the peak cross-correlation between each of the three variables pressure, temperature, and relative humidity with GHI over the training window to be determined. This must be noted that the minimum shift between the three variables already mentioned with respect to the GHI is considered to be equal to the forecast horizon, i.e. in order to do the forecast for 24 hours ahead the correlation function of each select variable, \( v(t) \), with the GHI, \( I(t) \), is computed neglecting the \( v(t) \) values recorded during the past 24 hours. This can be mathematically represented as follows:

\[
\begin{align*}
\text{Eq. 16} \\
r_{L}^{T_{sh}} = \text{corr}(f(t - T_{th} - T_{sh}), I(t)) = \frac{\sum_{i=N-N_{d}}^{N} \left( f(t_{i} - T_{sh}) - f(t_{i}) \right) \left( I(t_{i}) - I_{\bar{t}} \right)}{\sqrt{\sum_{i=N-N_{d}}^{N} \left( f(t_{i} - T_{sh}) - f(t_{i}) \right)^{2} \left( I(t_{i}) - I_{\bar{t}} \right)^{2}}} \\
\text{f}(t_{i}) = \frac{\sum_{j=\text{min}(N_{d}+\lfloor \frac{L}{2} \rfloor)}^{\text{max}(N_{d}+\lfloor \frac{L}{2} \rfloor)} \left( v(t_{j}) \right)}{\text{min}(N_{d}+\lfloor \frac{L}{2} \rfloor, \text{max}(N_{d}+\lfloor \frac{L}{2} \rfloor)} \right)}
\end{align*}
\]

Where \( T_{sh} \) and \( L \), respectively, represent the shift value and the length of low-pass filter which is applied to the signal for correlation analysis and \( T_{th} \) stands for the forecast horizon which is the default delay value between the variable \( v(t) \) and the irradiance \( I(t) \). Parameter \( N \) represents the total number of samples that exist in the vector \( v(t) \) while \( N_{d} \) is a positive integer showing the number of samples in the data subset \( \tau \) for target day \( d \). Each training subset \( \tau \) for target day \( d \), i.e. \( S_{d}^{\tau} \), is a predefined subset which includes the variables for a constant number of days in the past. In the proposed framework the number of days in the past which are considered as the predefined
lengths of training subsets are stored in a unit-less vector, $\varphi = [5,8,10,12,15,20]$. Term $T_{sh}$ ranges from 0 to $N_d^2$ while the filter length $L$ is between 1 and 1440 samples, i.e. minutes.

4.6 Feature Extraction

A fundamental problem in neural network research, as well as in many other disciplines, is finding a suitable representation of multivariate data, i.e. random vectors. For reasons of computational and conceptual simplicity, the representation is often sought as a linear transformation of the original data. In other words, each component of the representation is a linear combination of the original variables. Well-known linear transformation methods based on statistical concepts which use mathematical tools to obtain some information in different forms such as feature extraction, blind source separation, etc. include principal components analysis (PCA), factor analysis (FA), projection pursuit and independent components analysis (ICA).

In the proposed framework the underlying features of input data are be extracted before being fed to the artificial neural network for further pattern recognition and training purposes. There are a number of feature extraction approaches in the literature each of which have specific applications. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are used for feature extraction in this project.

4.6.1 Principal Components Analysis (PCA)

The PCA tries to find the internal modes of a data set by finding the eigenvectors of the covariance matrix of that data set. The degree of importance of each eigenvector, i.e. feature vector, is determined by its relevant eigenvalue. The larger the eigenvalue, the more important the eigenvector is which means the data is more widely spread along the line represented by that feature vector. PCA can also be used for dimensionality reduction, when the data set is high
dimensional, in order to effectively reduce computation load at the cost of losing accuracy to a feasible extent which is not the case for this research. Before mentioning the formulation of PCA let us introduce some preliminary statistical definitions.

Mean:

\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{Eq. 18} \]

Standard Deviation:

\[ \sigma_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \quad \text{Eq. 19} \]

Variance:

\[ \text{var}(x) = \sigma_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \quad \text{Eq. 20} \]

Covariance:

\[ \text{cov}(x, y) = \sigma_x \sigma_y = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad \text{Eq. 21} \]

PCA formulation can now be given as follows:

\[ U = M_{\text{pca}} \mathbf{X}_{\text{adj}}^T \quad \text{Eq. 22} \]

Where \( M_{\text{pca}} \) is the feature matrix obtained by PCA algorithm which consists of eigenvectors of the covariance matrix of adjusted input variables, \( \mathbf{X}_{\text{adj}} \), on its columns and \( \mathbf{X}_{\text{adj}} \) is the matrix of adjusted, i.e. zero mean, input variables where each column represents a variable, i.e. a dimension, and each row contains samples of all variables for a single data point containing each variable:

\[ \mathbf{X}_{\text{adj}} = \mathbf{X} - \bar{\mathbf{X}} \quad \text{Eq. 23} \]

Computing the Singular Value Decomposition (SVD) of the adjusted input matrix \( \mathbf{X}_{\text{adj}} \) yields:
\[ X_{\text{adj}} = P_{d \times d} D_{d \times d} (Q_{d \times d})^T \]  
Eq. 24

Where \( P \) and \( Q \) are orthogonal matrices and \( D \) is a diagonal matrix.

The definition for covariance matrix is as follows:

\[ M_{\text{cov}} = X_{\text{adj}} X_{\text{adj}}^T = (c_{i,j}, c_{j,i} = \text{cov}(\text{Dim}_i, \text{Dim}_j)) \]  
Eq. 25

Where \( M_{\text{cov}} \) is a square matrix with \( d \) rows and \( d \) columns. This means that the covariance matrix has dimensions equal to the dimension of the data set, and \( \text{Dim}_k \) refers to the \( k^{\text{th}} \) dimension.

The PCA transform matrix, i.e. the feature matrix, can then be computed as follows:

\[ [M_{\text{pca}} \lambda_{\text{pca}}] = \text{eig}(M_{\text{cov}}) \]  
Eq. 26

Which means the \( M_{\text{pca}} \) is equal to the matrix of eigenvectors of the covariance matrix, \( M_{\text{cov}} \), and \( \lambda_{\text{pca}} \) represents the diagonal matrix which includes the eigenvalues of the covariance matrix on its main diagonal.

4.6.2 Independent Components Analysis (ICA) and FastICA Algorithm

Independent components analysis (ICA) is a recently developed method in which the goal is to find a linear representation of non-Gaussian data so that the components are statistically independent, or as independent as possible. Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation.

In order to reveal the underlying features of an input set \( X \) the FastICA learning rule finds a direction, i.e. a unit vector \( w \) such that the projection \( w^T X \) maximizes non-Gaussianity. Non-Gaussianity is here measured by the approximation of negentropy \( J(w^T X) \) given as follows [20; 21]:

98
\[ J(y) \propto \left[ E[G(y)] - E[G(v)] \right]^2 \]

Eq. 27

For any non-quadratic practical function \(G\). However, by choosing \(G\) wisely, better results can be obtained. In particular, choosing \(G\) that does not grow too fast, one obtains more robust estimators. The following choices of \(G\) have proven very useful:

\[
G_1(u) = \frac{1}{a_1} \log \cosh(a_1 u) , \quad G_2(u) = -\exp(-u^2/2) 
\]

Eq. 28

Where \(1 \leq a_1 \leq 2\) is some suitable constant.

The variance of \(w^TX\) must here be constrained to unity; for whitened data this is equivalent to constraining the norm of \(w\) to be unity [22].

The FastICA is based on a fixed-point iteration scheme for finding a maximum of the non-Gaussianity of \(w^TX\). It can be also derived as an approximative Newton iteration [21]. Denote by \(g\) the derivative of the nonquadratic function \(G\) used mentioned before; for example the derivatives of the functions introduced as useful choices above:

\[
g_1(u) = \tanh(a_1 u) , \quad g_2(u) = u \exp(-u^2/2)
\]

Eq. 29

Where \(1 \leq a_2 \leq 2\) is some suitable constant, often taken as \(a_2 = 1\). The basic form of the FastICA algorithm is as follows:

1. Choose an initial (e.g. random) weight vector \(w\).
2. Let \(w^+ = E\{xg(w^TX)\} - E\{g'(w^TX)\} w\)
3. Let \(w = w^+/||w^+||\)
4. If not converged, go back to 2.
Convergence means that the old and new values of w point in the same direction, i.e. their dot-product is (almost) equal to 1. It is assumed that the data is pre-whitened. Figure 38 explains the effect of whitening on the joint distribution density of original variables:

The preprocessing options for ICA include:

- Centering: remove the mean of the data so that the data becomes zero-mean.
- Whitening: transform the observed vector x linearly so that we obtain a new vector \( \tilde{x} \) which is white, i.e. its components are uncorrelated and their variances equal unity. In other words, the covariance matrix of \( \tilde{x} \) equals the identity matrix:

\[
E \{ \tilde{x} \tilde{x}^T \} = I. \quad \text{Eq. 30}
\]

- PCA: one possible option is to use principal components analysis itself as a preprocessing for ICA.
4.7 Artificial Neural Network Structure, Adaptive Training and Test

Artificial Neural Networks (ANN), as represented in Figure 39, can be used for almost any types of systems due to the fact that their capability to mimic highly nonlinear, intermittent and time-varying functions is proven over the past two decades both in theory and in practice. The models obtained by ANN then can be used to serve for critical experiments which may not be feasible or possible at all to be done on the actual system in order to get an insight of what the system’s behavior is under specific situations.

Figure 39: Artificial Neural Network (ANN) General Structure

Figure 40 summarizes different scenarios in which ANN can be used for purposes. Those scenarios can be categorized into two main areas called estimation and forecast, each of which may be performed in an “open loop” or a “closed loop” manner. This makes a total...
of four different possibilities for using ANN as a predictor. Block diagrams “a” and “b” in figure 16 represent the training process for open loop and closed loop estimation respectively, while the two block diagrams “c” and “d” show the training procedure for open and closed loop forecast. The difference between estimation and forecast is in the fact that for estimation the input values to the system at each time instance, such as the value of temperature, pressure, cloud cover percentage, irradiance at previous time instance, etc., are known and the system is only supposed to give an estimate of the value of the expected irradiance at the same time instance assuming we don’t have the irradiance sensor available for that instance of time. This is not the case when doing forecast, i.e. the values of the inputs to the system for the day ahead are not known, since the next day has not happened yet. Therefore, this makes the problem even more sophisticated which means that the day-ahead forecast must be made using only the information available at current time and whatever is recorded in the past.

This must be noted that ANN have an internal feedback loop through which their weights are updated during training phase until convergence occurs, i.e. the error between the actual system’s output and the ANN output drops down under some predetermined level. However, the open-loop or closed-loop nature of the modeling process in general is determined based on the fact that if outputs of the actual system are being used for training the ANN or not. If the outputs of the actual system in the past are used for training the neural network, the process is called closed-loop. Otherwise, modeling process is referred to as open-loop.
A feed-forward (FF) back-propagation (BP) ANN is used with one input layer, one hidden layer and one output layer. The hidden layer consists of 20 neurons with tan-sig activation function. Convergence criteria in the training phase include either five validation check failures or 10 training epochs. The default mean squared error (mse) performance value and gradient of 0.00 and $10^{-5}$ are left untouched. The error back-propagation method is used for updating the weights in the training phase based on Levenberg-Marquardt algorithm which uses the gradient decent methodology. The ANN mentioned in the proposed framework uses the upcoming entries of the
dataset in order to retrain while it is used in real-time for forecasting. This way, the ANN will keep itself updated so that it can match with intentional or unintentional changes that may occur in a system’s general structure or in the environment.

The ANN mentioned in the proposed framework uses the upcoming entries of the data set in order to retrain while it is being used in real-time for modeling purposes either if it is for estimation or forecast. This means that instead of being trained once using a specific constant window of data set and keep running afterwards based on the initial settings and weight values, the ANN will continuously retrain itself using the data inside a moving window which sweeps the updated data-set and this way the ANN will keeps itself as updated as possible so that it can match with intentional or un-intentional changes that may occur in a system’s general structure or in the environment. This way, i.e. through adaptive training and test using training sets determined based on synoptic event detection algorithm, ANN results in more reliable models compared to training by a static training set or even by a moving training window with a constant length, and consequently will lead to more accurate and precise estimation/forecast outcomes.

Taking advantage of an ANN structure as the core of the forecast engine the proposed framework optimally analyze and recognize patterns in meteorological and solar data inputs and understands how different variables are related to each other and what would be the climate-related outcome of certain changing patterns in some variables with respect to others.
Figure 41: Neural Networks Training and Test Process for 1 Day-Ahead Solar Energy Forecast

The procedure for training the neural networks is visually mentioned in Figure 41. Figure 41 represents the consecutive training days in dark blue, the reference day in light blue and the target day in red. As it can be seen in the diagram during the training phase, the output of the next day is the target for inputs and output of the current day. This process continues until we get to the reference day. Then, the test phase begins. This means that the inputs and output of the reference day are fed into the already trained ANN and the output of the ANN is kept recorded in order to be compared to the actual target signal which will happen the next day. This target signal may be the output power of the solar plant or the amount of solar irradiance that reaches the ground.

The neural network training procedure can be illustrated by the fact that the output of the next day is the target for current day’s input-output (I/O) measured variables. This process continues until we get to the reference, or in some cases current, day. After training converges, the
test phase begins. The independent components associated with inputs and output of the reference day are fed into the already trained ANN and the output of the ANN is the forecast result which will be later on compared to the actual target signal. The target signal may be the output power of the solar plant or the measured solar irradiance that reaches the ground.

In the proposed framework, the original input data first is categorized into different training subsets for adaptive training and test purposes based on the outcomes of synoptic event detection algorithm. Data included in each training subset will pass through a correlation function analysis unit which helps determine the best filter and mutual shift required between the select variables with respect to GHI over the period of training subset in order to maximize the cross-correlation of those select variables and GHI. Then the features of the best fit variables, i.e. the shifted and filtered versions of training subset variables, is computed using ICA and a de-mixing matrix, W, is computed for each training subset. Those features will finally be fed into the ANN in training phase. In the test phase the recorded select variables of the reference day, which is usually the same as current day in solar energy forecast applications, is filtered and shifted using the best set of filter and shift obtained from correlation analysis in training process. After that, the matrix of modified variables is multiplied by the de-mixing matrix, W, obtained from applying ICA in training phase in order to extract the features of test data considering the same rules applies to the modified variables of reference day. The features of test data will then be given to the trained ANN as input for the estimation to be done for target day’s expected solar energy. For each target day the aggregation will take place based on the results of dynamic performance evaluation. When all aggregated results are out for the data sets generated by Monte Carlo approach and the percentiles are determined, the forecasts is fine-tuned using the climate pattern rules extracted from synoptic events detection outcomes.
Figure 42: One Day-Ahead Forecast Performance versus Training Window Length; (a) Non-adaptive approach without ICA (b) Non-adaptive approach with ICA
Correlation Coefficient (Solid Blue), RMSE (Dashed Green)

Figure 42 represents the forecast accuracy measures versus the training window’s for the original case without ICA and for the case where the ICA is used. The two performance curves represent the correlation coefficient (Solid Blue Curve) and the root mean square error (Dashed Green Curve). Simulation is done for all 366 days in the year 2012. It can be seen that
both correlation coefficient and the RMSE tend to improve as the training window length increases. The correlation coefficient and the RMSE have exponential saturation envelop and exponential decay envelop respectively.

It is evident that the ICA as a feature extraction approach significantly improves the performance when compared to the original forecast case. However, the forecast becomes more and more time-consuming as the training window length increases and there is no guarantee that the intra-daily weather patterns are accurately followed as the training window expands for us to get lower RMSE or higher correlation.

The predefined lengths for the adaptive training windows are stored in a vector called training specifications vector, \( \phi = [5, 8, 10, 12, 15, 20] \). Those predefined lengths, however, do not mean that every single days before current day will be included in the training until the number of days is reached. The days which will be put in the training subset will be determined based on some training rules defined the authors for achieving more accurate results. These rules which consider the day-types obtained by synoptic event detection algorithm on the pressure signal are as follows in order to form the training window are as follows:

- IF current day is “Lo” or “Rise” of “P” THEN include days of type “Lo”, “Rise” and “Hi” of “P”
- IF current day is “Hi” or “Fall” of “P” THEN include days of type “Hi”, “Fall” and “Lo” of “P”

Hence, by choosing the information-rich parts of each data subset for training purposes it is more possible that the forecasting will lead to results with higher accuracy and precision since the appropriate patterns of data relevant to the current climate conditions and changes are included in the training set.
In the training phase the input matrix to the ANN includes 17 variables where length of each variable, i.e. the batch size, is between $5 \times 1440$ and $20 \times 1440$ samples and the target is a vector of the same length as the batch size. In the test, i.e. forecasting, phase the input matrix consists of 17 variables of 1440 samples each. The target is a vector of length 1440 samples which is the expected GHI of target day.

### 4.8 Monte Carlo Uncertainty Analysis

In order to stabilize the forecast results and to obtain reliable predictions of the available solar energy in the days ahead it is required to consider the uncertainty associated with the recorded data. This uncertainty can be associated with the measurement noise, hardware or software sensory failure and technology limitations in providing highly accurate measurement devices. Therefore, 19 data sets similar to the originally recorded data base are generated using the sensor tolerances obtained from current industry state of the art producers. The variables considered for Monte Carlo approach to uncertainty analysis include atmospheric pressure, temperature, relative humidity, total and opaque cloud cover, and GHI. The sensory tolerance for each variable was the basis for generation of the other 19 seeds. We assume the recorded value to be the actual value for each variable and generate 19 other values around this actual value as follows:

$$X_i = \begin{cases} X & i = 1 \\ X(1 + \text{Rand}(\mu, \sigma)) & i = 2, 3, ..., 20 \end{cases}$$

Eq. 31

Where $X_i$ refers to the $i^{th}$ data set and $X$ is the original recorded data set. $\text{Rand}(\mu, \sigma)$ is a matrix of random variables with Gaussian distribution of mean $\mu$ and standard deviation $\sigma$. For the purposes of this study we needed $\mu = 0$ and $\sigma = X(v_{tol})$ where $v_{tol}$ is the tolerance vector which includes the tolerances, or coefficients of variation (COV), of sensors designed for each and every one of the variables considered in Monte Carlo analysis. For variables such as Time, Zenith angle and Azimuth angle the tolerance was considered to be 0 therefore the values of those variables are
the same in all of the 20 data sets. The COV’s are respectively chosen to be 0.03, 0.15, 0.01, 0.02, 0.04 and 0.03 for atmospheric pressure, temperature, relative humidity, GHI, total cloud cover percentage and opaque cloud cover percentage [23, 24, 25, 26].

After the forecast results are obtained for each and every one of the data sets generated by Monte Carlo approach, i.e. $Y_i, i = 1, 2, ..., 20$, the output percentiles are determined and the fiftieth percentile, i.e. $p_{50}$, is considered as the stabilized forecast result. The following relationships represent how to compute the percentiles:

\[ p_{90}(y) = \bar{Y} - z_{90} \times \sigma_Y \]  
\[ p_{75}(y) = \bar{Y} - z_{75} \times \sigma_Y \]  
\[ p_{50}(y) = \bar{Y} - z_{50} \times \sigma_Y \]

Where $\bar{Y}$ denotes the mean of all 20 outcomes of the 20 different data sets used for forecast and $\sigma_Y$ represents the standard deviation of those forecast results. Finally, $z$ stands for the z-score of each percentile. The z-scores for the aforementioned percentiles are as follows:

\[ z_{50} = 0 \quad , \quad z_{75} = 0.675 \quad , \quad z_{90} = 1.282 \]  

which will be used to determine the final forecast results of target day. The fiftieth percentile, i.e. $p_{50}$, is regarded to as the final forecast result. However, the seventy-fifth percentile, i.e. $p_{75}$, and the ninetieth percentile, i.e. $p_{90}$, are also used in some cases where more conservative predictions are preferred. The ninetieth percentile is the most conservative case which infers that 90 percent of the time the measured solar energy will be higher than the forecast. This may be the preferred choice for utilities that are interested in conservative bidding of energy into the day-
ahead electricity market to avoid penalties associated with over-prediction imposed by the independent system operators.

4.9 Dynamic Performance Evaluation

Dynamic evaluation of performance helps with the adaptive training by making the smart aggregation procedure possible which improves the accuracy of forecast results. The outcome of dynamic performance evaluation, i.e. the performance matrix, tends to be more reliable as time passes and more new data come into the database. In this process the performances of each and every training subset used for training of the ANN for the last target day is computed. Let’s denote those subsets by $S^\tau_{d-1}$, $\tau = 1, 2, ..., \tau$ where $\tau$ and $\tau$ respectively represent the training set number and the total number of data subsets considered for adaptive training and $d$ stands for current target day. This means the performances are calculated for $d-1$ target days so far. The aggregation coefficients will be determined based on the recorded forecast performances for each and every training subset. This will be discussed in more detail in section III.H.

Once the forecast for target day $d$ is obtained through the ANN that is trained by the training subset $\tau$, i.e. $S^\tau_{d}$, the accuracy of forecast result must be computed in order for the usefulness of the training subset to be evaluated. The performance measure, i.e. function, used in the process of dynamic performance evaluation is defined as the ratio of the correlation coefficient between forecast and target for each target day $d$ to the forecast RMSE of the same day:

$$\gamma^\tau_d = \frac{r^\tau_d}{\max(1, \text{RMSE}^\tau_d)} \quad \text{Eq. 36}$$

Where $\tau$ refers to the training set used in order to train the forecast smart engine for the specific target day $d$. The higher the cross-correlation between forecast and actual irradiance for a target day $d$, $r^\tau_d$, and the lower the forecast error for the same day, $\text{RMSE}^\tau_d$, the better and higher the performance value, $\gamma^\tau_d$, is. Considering the largest value of correlation coefficient which is 1 it
can be said that the highest possible value for performance is 1, which represents the best case in day-ahead forecast, since the denominator of performance function will not be smaller than 1.

For each target day \( d \) after the performance values of each and every training set are calculated those values will form a vector called the performance vector. This performance vector is added as a new line to the end of the performance matrix. The mathematical representation is as follows:

\[
\gamma_d = \left[ \gamma_d^1 \gamma_d^2 \ldots \gamma_d^\tau \right], \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{d-1} \end{bmatrix}
\]

Eq. 37

Where \( \gamma_d \) represents the performance vector for target day \( d \) and \( \gamma \) stands for the performance matrix. \( \gamma_{d-1} \) is the performance vector for the last target day where \( d \) is the number of target day that is currently being forecast. Each row of the performance matrix includes the performance measure for all of the adaptive training subsets for one single target day.

### 4.10 Aggregation Process and Fine-Tuning

After the performance matrix is obtained the aggregation coefficients need to be computed so that the weighted average regime can be applied to obtain the forecast result for target day \( d \).

The aggregation coefficients are determined as follows and are used to compute the forecast result of target day:

\[
c_d = \begin{cases} 
\frac{1}{v_d} & \text{if } d = 1 \\
\frac{1}{\sum_{\tau=1}^{d} v_\tau} & \text{if } d > 1
\end{cases}
\]

Eq. 38

Where \( c_d \) is the aggregation coefficient which is used as the weight factor for the forecast result for target day \( d \) obtained using the ANN trained by the training subset \( \tau \), i.e. \( S_d^\tau \). \( \alpha \) is a constant coefficient which can be used to give more dominance to the performance value of the
last target day, i.e. target day \(d - 1\), with respect to the performances of the previous target days. In our simulations the value of \(\alpha\) is set to 10.

Assuming that \(y^*_d\) is the forecast obtained for target day \(d\) by training set \(\tau\) the aggregated forecast can then be represented as follows:

\[
Y_d = \frac{\sum_{\tau} y^*_d \times c^*_d}{\sum_{\tau} c^*_d} \quad \text{Eq. 39}
\]

Where \(Y_d\) is the aggregated forecast for target day \(d\).

The rule-base used for fine-tuning of the forecast results for each and every target day \(d\) consists of the following rules which take into account the current day’s condition and the day-types for current and past days. In the rules mentioned below \(P\) and \(\frac{P \times T}{RH}\) respectively refer to the atmospheric pressure signal the second reference signal which is defined based on Bernoulli’s principle:

- IF Current day is “Fall” of \(\frac{P \times T}{RH}\) AND “Lo” of “P” THEN Next day is “Overcast”
- IF Current day is “Fall” of \(\frac{P \times T}{RH}\) AND “Hi” of “P” THEN Next day is “Scattered”
- IF Current day is “Hi” of \(\frac{P \times T}{RH}\) AND “Hi” of “P” THEN Next day is “Scattered to Clear”
- IF Current day is “Hi” of \(\frac{P \times T}{RH}\) AND “Lo” of “P” THEN Next day is “Scattered to Overcast”
- IF Current day is “NOT Clear” AND “Fall” of \(\frac{P \times T}{RH}\) AND “Fall” of “P” THEN Next day is “Overcast”
- IF Current day is “Overcast” AND “Fall” of \(\frac{P \times T}{RH}\) THEN Next day is “Scattered to Overcast”
The rules mentioned above work as a modifier of the raw forecast results obtained from the ANN-based forecast engine. The forecast results are adjusted by averaging them with overcast and clear sky reference curves according to the conditions mentioned in the relationships mentioned above.

4.11 Day-ahead Solar Energy Forecast User Interface

The user interface consists of three main displays. Figure 43 represents the which shows the forecasted irradiance curve in red and the actual GHI curve in filled blue for the target day along with the actual and percent error values and average values of target and forecast. This display will update as time passes and new samples of the actual target GHI are measured and recorded.

Figure 43: Dynamic User Interface Display; Forecast (dashed red), Target (Filled blue), Actual Error (filled red), Percent Error (Filled Pink), Average Forecast (Red Bar), Average Target (Blue Bar)
The second interface display given in Figure 44 represents the Monte Carlo analysis results. The two curves of 50th and 90th percentiles, i.e. $p_{50}$ and $p_{90}$, are shown on the diagram in pink and green respectively. This display also includes a table which summarizes the optimal market bid values for solar energy by representing hourly averaged solar irradiance forecasts based on three main percentiles obtained by Monte Carlo uncertainty analysis approach. For most conservative bids the ninetieth percentile, i.e. $p_{90}$ can be used.

![Figure 44: Main User Interface Display representing forecast results based on Monte Carlo uncertainty analysis approach for current day and the next day; forecast boundaries (bluish area), Forecast $p_{50}$ (magenta), Forecast $p_{90}$ (green), table represents the hourly average values of irradiance in W/m² for the next day for the three different percentiles of $p_{50}$, $p_{75}$ and $p_{90}$.](image)

The same results of Figure 44 are represented in
Figure 45

area depicted in figure 19 are actually envelops of a number of probability density functions (pdf) representing the probability of the solar irradiance intensity that reaches the ground at the specific site where the solar farm is located at different time periods based on the Monte Carlo approach to uncertainty analysis. All curves represented in Figure 44 are forecast results. The rest of the day dated DEC. 30th 2012 and the whole day tomorrow dated DEC. 31st 2012 are forecasted in this case.

Figure 45 from a different perspective. This shows the depth of data which means that the bluish area depicted in figure 19 are actually envelops of a number of probability density functions (pdf) representing the probability of the solar irradiance intensity that reaches the ground at the specific site where the solar farm is located at different time periods based on the Monte Carlo approach to uncertainty analysis. All curves represented in Figure 44 are forecast results. The rest of the day dated DEC. 30th 2012 and the whole day tomorrow dated DEC. 31st 2012 are forecasted in this case.
Figure 45: Main User Interface Display representing pdf’s of forecast results based on Monte Carlo uncertainty from a different view angle.

Figure 46: User Interface Display representing 1-Day ahead and 2-Days ahead forecast of the day before and current day percentiles based on Monte Carlo uncertainty analysis approach for performance tracking and evaluation; Target (blue), Forecast $p_{90}$ (magenta), Forecast $p_{50}$ (green), Current-Time (Black dot).
Top portion of Figure 46 represents the actual measured irradiance curve of the and current day, up to current time, together with the forecasts, i.e. the fiftieth and ninetieth percentiles, for the day before and the complete current day’s forecast. This must be noted that yesterday’s forecast results represented in this top section of Figure 46 are obtained using ahead forecast approach whereas the forecast for current day is obtained by 2 days-ahead forecast method.

Bottom portion of the Figure 46 represents current day’s measure irradiance, from minute in the morning up to current time represented by bolded black dot, together with the forecast results obtained by 1 day-ahead forecast approach. The reason of having results of two different forecast schemes represented on the same figure for current day is for operator to compare the performances of two approaches and to realize how using the most recent set of information can lead to higher accuracy. This way, the operator will also be able to get to know how climate patterns change by seeing two consecutive days on the top portion of the figure. The user will also observe how the forecast engine performs in detecting the climate patterns by comparing the accuracies of the next day forecast and the day-ahead forecast for each target day.

4.12 Summary

A solar forecasting tool is developed.
5.1 Overview

Metrics used to evaluate effectiveness of forecasting methods can be divided into three main categories: consistency, quality, and value [1_UCSD]. Consistency refers to the correspondence between the forecast and the judgments made by forecasters to determine the forecast (i.e. do the same inputs that are used to determine a forecast produce the same forecast). Quality refers to the difference between forecasts and observations. Finally, value refers to the incremental monetary benefits of forecasts to users [UCSD]. Solar power generators and system operators are two main entities that use solar forecasting and may benefit from using it. In an open market, solar power generators would primarily rely on the value criterion because forecast quality does not necessarily translate to forecast value [UCSD]. For instance, a forecast that is capable of providing the user with a higher quality, i.e. more accurate and precise forecasts, during peak net load times of the day, when energy prices are high and errors would be more costly, may be more valuable than a forecast that has an overall higher quality, but not necessarily at the critical times of day. System operators, on the other hand, are primarily concerned with the quality of a forecast, since their primary concern lies in grid reliability and accurate planning. The system operators’ secondary concern is to operate the energy market optimally reducing energy generation, transmission, and reserve costs on the grid for which reason the forecast value becomes the main factor since the amount and nature of forecast bias and the planning made based on forecasts can affect the marginal electricity price and cause congestion. For example, an under-forecast would result in over-procurement of energy at a higher marginal cost and possibly transmission
congestion near the solar power plant; an over-forecast would result in under-procurement of energy, purchase of energy from reserves or regulation, and potential operation of transmission lines below capacity [UCSD].

Consistency of the proposed solar forecasting framework is ensured through implementation of the Monte Carlo uncertainty analysis together with the adaptive nature of the methodology which is achieved by pre-processing of the data, extracting data features, and the synoptic events detection algorithm.

In this chapter the performance and the value of the proposed forecasting framework will be discussed. A recent topic of discussion between the forecasting experts and the electric grid entities is the mismatch between what is considered a good forecast for system operators and power generators. It is possible for a power generator to benefit economically from using biased (rather than neutral) forecasts, which is detrimental to the systems operator’s goal of reliability of the power grid [2_UCSD].

5.2 Performance Metrics

In order to evaluate performance in such a way to be comparable to the literature some standard and most popular measures were considered which already exist in the literature for similar areas of research. This section considers hour by hour performance evaluation through computing the hourly average values of actual GHI and the forecast GHI. Figure 47 represents the irradiance profile for a single day in red and the accumulated GHI in blue. By dividing the accumulated GHI to the duration of integration the average value can be easily found for the target day. In a similar fashion each single day is divided into 24 time slots, standing for the 24 hours of the day, and the average solar energy is calculated for each time slot which later
on is regarded as hourly averaged solar energy. The hourly performance of the proposed solar forecast framework is evaluated for the 366 days in the year 2012.

Figure 47: GHI profile for a single day (RED), and the accumulated Solar Irradiance (BLUE)

After obtaining the daily average values and the hourly average values for both target and forecast GHI the error indicators are computed as mentioned in Eq. 40-42:

- **Mean Error (ME) or Bias Error (BE)**
  \[
  \text{BE} = \frac{1}{N} \sum_{i=1}^{N} (T_i - F_i) \tag{40}
  \]

- **Mean Absolute Error (MAE)**
  \[
  \text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |T_i - F_i| \tag{41}
  \]

- **Root Mean Squared Error (RMSE)***

---

121
\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_i - F_i)^2}
\]

Eq. 42

where \(N\) represents the total number of points for which the error values are calculated.

The year 2012 is comprised of 366 day points and 366*24 = 8784 hour points. Therefore, in the case of daily averaged irradiance value there will be \(N = 366\) points for error computation in this year. However, for hourly averaged solar irradiance values the case is a little complicated since the night time hours must be excluded from the final set in order to obtain a clear understanding of system performance for day-time samples only, i.e. \(T_i = \frac{1}{n} \sum_{j=(i-1)\cdot n+1}^{i \cdot n+1} (\text{GHI})_j > 0\) where \(n\) is the number of samples in one hour. Hence, the number of day-time samples for hourly averaged solar energy will be around \(N = 4400\). Although the above-mentioned indicators are extremely popular and are regarded as standard measures of error in the literature they may fail when two different approaches are being compared to each other using two different data sets. Hence, relative error measures are to be used which take into account the ratio of error to the average value of the target value. These relative measures are obtained as follows:

- **Relative Mean Error (ME\(_\text{Rel}\)) or Bias Error (BE\(_\text{Rel}\))**

  \[
  \text{BE}_{\text{Rel}} = \frac{\frac{1}{N} \sum_{i=1}^{N} (T_i - F_i)}{\overline{T}}
  \]

  Eq. 43

- **Relative Mean Absolute Error (MAE\(_\text{Rel}\))**

  \[
  \text{MAE}_{\text{Rel}} = \frac{\frac{1}{N} \sum_{i=1}^{N} |T_i - F_i|}{\overline{T}}
  \]

  Eq. 44

- **Relative Root Mean Squared Error (RMSE\(_\text{Rel}\))**

  \[
  \text{RMSE}_{\text{Rel}} = \sqrt{\frac{\frac{1}{N} \sum_{i=1}^{N} (T_i - F_i)^2}{\overline{T}}}
  \]

  Eq. 45

where \(\overline{T}\) is the mean value of solar energy that reaches the ground during the day-time for the whole year 2012 which is computed as follows:
\[
T = \frac{1}{N} \sum_{i=1}^{N} T_i \quad T_i = \frac{1}{n} \sum_{j=i-n+1}^{i} (GHI)_j > 0
\]  

Eq. 46

The Percent Error is also considered for performance evaluation purposes and is defined as the ratio of the difference between the forecast and the target divided by the capacity at each time instance. Mathematical representation of the percent error is given in Eq. 47.

\[
e_i^\% = 100 \times \frac{(T_i - F_i)}{C_i} \quad T_i = \frac{1}{n} \sum_{j=i-n+1}^{i} (GHI)_j > 0
\]  

Eq. 47

where \(T_i\) and \(F_i\) are the \(i\)th sample of the target and forecast respectively and the \(C_i\) represents the maximum solar energy capacity, i.e. the maximum amount of solar energy that may reach the ground, at time instance or duration \(i\). The percent error is bounded within the range of -100\% and 100\%. Percent error’s mean value and standard deviation together with the mean absolute percent error (MAPE) will serve as performance measures which are computed as given in Eq. 48 to Eq. 50.

\[
\mu^e^\% = \frac{1}{N} \sum_{i=1}^{N} e_i^\%
\]

Eq. 48

\[
\sigma^e^\% = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (e_i^\% - \mu^e^\%)^2}
\]

Eq. 49

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} |e_i^\%| 
\]

Eq. 50

Lastly, one of the most important performance measures is the correlation coefficient between the target solar energy and its forecast. The correlation coefficient represents the strength of the linear association between target and forecast. In other words, it tells us how well is the forecast tracking the actual upcoming deviations in the solar irradiance curve. The correlation coefficient is computed as follows:
\[ r = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{T_i - \bar{T}}{\sigma_T} \right) \left( \frac{F_i - \bar{F}}{\sigma_F} \right) = \frac{\sum_{i=1}^{N}(T_i - \bar{T})(F_i - \bar{F})}{\sqrt{\sum_{i=1}^{N}(T_i - \bar{T})^2 (F_i - \bar{F})^2}} , \quad T_i = \frac{1}{n} \sum_{j=(i-1)n+1}^{in} (GHI_j) > 0 \]

Eq. 51

Where \( \sigma_T \) and \( \sigma_F \) respectively stand for the standard deviation of the target samples and their associated forecast. The more the correlation coefficient represented by "r" is close to 1 the better the changes in our forecast curve of solar energy is matched with those of the actual curve and the closer r is to zero the forecast patterns are less aligned with those of the actual solar irradiance curve.

5.3 Next-Day Solar Irradiance Forecast Performance Analysis for Golden, CO

Performance of the proposed solar forecasting framework is determined for prediction of the next day’s expected solar energy during the year 2012 using the NREL data set for their site in Golden, Colorado, United States. Table 6 summarizes the simulation results for next day solar energy forecast using different scenarios together with the next day persistence \(^1\) as the reference case for comparison. The methodologies that are compared in Table 6 include the standard next day persistence, the ICA-based forecast, the ICA-based forecast with Monte Carlo approach, the adaptive ICA-based forecast which takes advantage of the synoptic events detection algorithm and adaptive training of the forecast engine, and the adaptive ICA-based forecast with Monte Carlo analysis.

The measures represented in Table 6 include the correlation coefficient, bias error (BE), mean absolute error (MAE), root mean squared error (RMS), together with their relative

\(^1\) Persistence refers to the case where the future conditions are assumed to be exactly the same as current conditions. Hence, the persistence reference curve in the case of solar energy forecasting assumes that the next day’s solar irradiance will be exactly similar to current day’s solar irradiance curve. In the same fashion, n days-ahead persistence also assumes that today’s solar irradiance curve will be replicated on the next \( n \)th day.
values to the average day-time measured GHI for the whole year 2012 and the percent error mean, i.e. $\mu^\%$ and the percent error standard deviation, i.e. $\sigma^\%$.

Table 6: Performances of simulated approaches and persistence for next day forecast

<table>
<thead>
<tr>
<th>Approach</th>
<th>$r$</th>
<th>BE (W/m)</th>
<th>BE_rel ( % )</th>
<th>MAE (W/m)</th>
<th>MAE_rel ( % )</th>
<th>RMSE (W/m)</th>
<th>RMSE_rel ( % )</th>
<th>$\mu^%$</th>
<th>$\sigma^%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistence_ND</td>
<td>0.79</td>
<td>-0.23</td>
<td>0.01%</td>
<td>116.32</td>
<td>32.50%</td>
<td>194.55</td>
<td>54.36%</td>
<td>0.28%</td>
<td>29.66%</td>
</tr>
<tr>
<td>ICA_SF_ND</td>
<td>0.8</td>
<td>6.25</td>
<td>1.75%</td>
<td>121.84</td>
<td>34.03%</td>
<td>183.13</td>
<td>51.14%</td>
<td>-1.37%</td>
<td>38.34%</td>
</tr>
<tr>
<td>MC_ICA_SF_ND</td>
<td>0.823</td>
<td>3.03</td>
<td>0.85%</td>
<td>113.16</td>
<td>31.60%</td>
<td>171.71</td>
<td>47.95%</td>
<td>-3.50%</td>
<td>33.46%</td>
</tr>
<tr>
<td>Adapt_ICA_SF_ND</td>
<td>0.858</td>
<td>4.76</td>
<td>1.33%</td>
<td>104.21</td>
<td>29.12%</td>
<td>158.63</td>
<td>44.32%</td>
<td>2.58%</td>
<td>26.07%</td>
</tr>
<tr>
<td>MC_Adapt_ICA_SF_ND</td>
<td>0.869</td>
<td>8.91</td>
<td>2.49%</td>
<td>98.09</td>
<td>27.39%</td>
<td>149.29</td>
<td>41.69%</td>
<td>1.46%</td>
<td>22.88%</td>
</tr>
</tbody>
</table>

From Table 6 it is clear that the ICA-based approach alone gives almost the same performance of the persistence case however, when considering ICA-based method with Monte Carlo analysis the performance gets better compared to the persistence. The adaptive forecast also elevates the performance to the next level and when implemented according to the Monte Carlo analysis guarantees the highest performance of the proposed forecasting framework.

Figure 48 represents the target, i.e. the actual or measured solar irradiance, in blue together with the forecast result, obtained by the proposed framework, in red for the duration of 11 consecutive days in the year 2012.
In Figure 48, three major performance charts are represented as simulation results comparing the performance of proposed forecast engine with the standard reference case of persistence for next day forecast. These performance charts include: Target vs. forecast scatter plot, Actual error histogram, and the Percent error histogram. The measures are calculated only considering the hourly averaged solar energy of day-time samples.

From Figure 48, it is seen that the proposed approach strongly outperforms the next day forecast simulation based on the results obtained for year 2012 with respect to the standard performance measures.

Figure 48: Hourly forecast and target irradiance for days 222 to 232 of year 2012
Figure 49: Next Day Solar Energy Forecasting Performance
(a) Scatter Plot; Persistence, (b) Scatter Plot; Proposed Method, (c) Actual Error Histogram; Persistence, (d) Actual Error Histogram; Proposed Method, (e) Percent Error Histogram; Persistence, (f) Percent Error Histogram; Proposed Method.
Monthly analysis of the next day forecast performance is also considered for the proposed approach and is compared to that of the persistence reference case in order to get a better understanding of the improvements that the adaptive framework brings to the solar irradiance forecasting. Figure 50 in the following represents the results of monthly forecast analysis.
Comparison between performance of the proposed approach for next day solar energy forecast and some of the already benchmarked methods are given in the following. An overview on the statistical error measures for the complete German data set consisting of three stations is given in Table 7 for the next day solar energy forecast [5].

Table 7: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using German Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012).

<table>
<thead>
<tr>
<th>Approach</th>
<th>$\theta_{\text{RMSE}}$ (RMSE)</th>
<th>RMSE(W/m²) (RMSE Rel)</th>
<th>MAE(W/m²) (MAE Rel)</th>
<th>BE(W/m²) (BE Rel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ECMWF-OL</td>
<td>52 (36.1%)</td>
<td>92 (40.3%)</td>
<td>59 (26.2%)</td>
<td>-7 (-2.9%)</td>
</tr>
<tr>
<td>2 BLUE</td>
<td>50 (34.7%)</td>
<td>94 (41.5%)</td>
<td>62 (27.2%)</td>
<td>-3 (-1.4%)</td>
</tr>
<tr>
<td>3 MM-MOS</td>
<td>45 (31.3%)</td>
<td>99 (43.5%)</td>
<td>67 (29.6%)</td>
<td>-2 (-1.0%)</td>
</tr>
<tr>
<td>4 Proposed Method</td>
<td>45.26 (23.3%)</td>
<td>149.29 (41.69%)</td>
<td>98.09 (27.39%)</td>
<td>8.91 (2.49%)</td>
</tr>
<tr>
<td>(NREL Golden CO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 CENER</td>
<td>31 (21.6%)</td>
<td>113 (49.9%)</td>
<td>72 (31.5%)</td>
<td>13 (5.9%)</td>
</tr>
<tr>
<td>6 WRF-MT</td>
<td>26 (18.1%)</td>
<td>118 (51.8%)</td>
<td>74 (32.6%)</td>
<td>-1 (-0.3%)</td>
</tr>
<tr>
<td>German Persistence</td>
<td>0 (0.00%)</td>
<td>144 (63.5%)</td>
<td>92 (40.5%)</td>
<td>-4 (-1.8%)</td>
</tr>
<tr>
<td>NREL Persistence</td>
<td>0 (0.00%)</td>
<td>194.55 (54.36%)</td>
<td>116.32 (32.5%)</td>
<td>-0.23 (-0.06%)</td>
</tr>
</tbody>
</table>

Actual and relative values for RMSE, MAE and bias error are given in the Table 7 for the next day forecast in Germany for European NWP-based approaches and the persistence method as a reference for the years 2007 to 2008. Table 7 also includes results obtained using the method proposed in this article, i.e. UTSA Method, and the persistence approach for the
specific NREL site at Golden CO, USA for the complete year 2012. The actual improved RMSE values for each method with respect to the relevant local persistence are measures of positive change that happens in the RMSE by using the method over the persistence. The actual and relative RMSE improvement values are defined as follows:

\[
\theta_{\text{RMSE}} = \text{RMSE}_{\text{Persistence}} - \text{RMSE}_{\text{method}} \quad \text{Eq. 52}
\]

\[
\eta_{\text{RMSE}} = \frac{\theta_{\text{RMSE}}}{\text{RMSE}_{\text{Persistence}}} \times 100 \quad \text{Eq. 53}
\]

where \(\theta_{\text{RMSE}}\) is a factor representing the actual change in the forecast RMSE using the proposed method compared to the RMSE of the persistence and is simply equal to the difference between the RMSE obtained by the persistence and that of the method. The \(\theta_{\text{RMSE}}\), therefore, has the same unit with that of the RMSE itself. The relative improvement value in RMSE, i.e. \(\eta_{\text{RMSE}}\), is a unit-less factor that represents the percent change in the forecast RMSE using the proposed method compared to RMSE of the local persistence. A negative value for \(\theta_{\text{RMSE}}\) and \(\eta_{\text{RMSE}}\) means the proposed method is deficient and hence it is better to stick to the standard local persistence for that specific site and the relevant dataset.

Considering the relative improved RMSE values the proposed method outperforms the two approaches developed by CENER and the WRF-MT and stands in the fourth place after ECMWF-OL, BLUE and MM-MOS.

An overview on the statistical error measures for Switzerland is given in Table 8 for next day solar energy forecast [5] during the years 2007 and 2008 together with the results obtained by the proposed method for the NREL site at Golden Colorado during the year 2012.

Table 8: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using Swiss Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012).
Performance measures for the next day forecast in Switzerland for European NWP-based approaches and the persistence method as a reference for the years 2007 to 2008 are given in Table 8 together with the results of UTSA Method and the persistence approach for the specific NREL site at Golden CO, USA for the complete year 2012. Considering the relative improved RMSE values with respect to persistence our proposed method secures the fourth place after ECMWF-OL, BLUE and WRF-MT and performs better than the MM-MOS approach.

For Austria, in addition to the forecasts of the University of Oldenburg, Meteotest, CENER and the statistical method of Blue Sky, also the traditional synoptic method of the meteorologists of Blue Sky was also evaluated. This method is spatially restricted and only available for Austria. Results are represented in Table 9.

<table>
<thead>
<tr>
<th>Approach</th>
<th>$\delta_{\text{RMSE}}$ (RMSE)</th>
<th>RMSE(W/m²) (RMSE_rel)</th>
<th>MAE(W/m²) (MAE_rel)</th>
<th>BE(W/m²) (BE_rel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ECMWF-OL</td>
<td>51 (32.3%)</td>
<td>107 (39.6%)</td>
<td>70 (25.8%)</td>
<td>-1 (-0.3%)</td>
</tr>
<tr>
<td>2 BLUE</td>
<td>49 (31.0%)</td>
<td>109 (40.5%)</td>
<td>73 (27.0%)</td>
<td>-9 (-3.3%)</td>
</tr>
<tr>
<td>3 WRF-MT</td>
<td>39 (24.7%)</td>
<td>119 (44.2%)</td>
<td>76 (28.0%)</td>
<td>4 (1.3%)</td>
</tr>
<tr>
<td>4 Proposed Method (NREL Golden CO)</td>
<td>45.26 (23.3%)</td>
<td>149.29 (41.69%)</td>
<td>98.09 (27.39%)</td>
<td>8.91 (2.49%)</td>
</tr>
<tr>
<td>5 MM-MOS</td>
<td>36 (22.8%)</td>
<td>122 (45.0%)</td>
<td>85 (31.5%)</td>
<td>-18 (-6.6%)</td>
</tr>
<tr>
<td>Swiss Persistence</td>
<td>0 (0.00%)</td>
<td>158 (58.4%)</td>
<td>104 (38.7%)</td>
<td>-17 (-6.3%)</td>
</tr>
<tr>
<td>NREL Persistence</td>
<td>0 (0.00%)</td>
<td>194.55 (54.36%)</td>
<td>116.32 (32.5%)</td>
<td>-0.23 (-0.06%)</td>
</tr>
</tbody>
</table>
Table 9: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using Austrian Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012).

<table>
<thead>
<tr>
<th>Approach</th>
<th>$\delta_{\text{RMSE}}$ (%)</th>
<th>RMSE(W/m$^2$) (RMSE Rel)</th>
<th>MAE(W/m$^2$) (MAE Rel)</th>
<th>BE(W/m$^2$) (BE Rel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BLUE</td>
<td>43 (30.3%)</td>
<td>99 (44.6%)</td>
<td>61 (27.6%)</td>
<td>1 (0.5%)</td>
</tr>
<tr>
<td>2 ECMWF-OL</td>
<td>41 (28.9%)</td>
<td>101 (45.6%)</td>
<td>65 (29.2%)</td>
<td>16 (7%)</td>
</tr>
<tr>
<td>3 UTSA Method (NREL Golden CO)</td>
<td>45.26 (23.3%)</td>
<td>149.29 (41.69%)</td>
<td>98.09 (27.39%)</td>
<td>8.91 (2.49%)</td>
</tr>
<tr>
<td>4 SYNOP</td>
<td>30 (21.1%)</td>
<td>112 (50.4%)</td>
<td>70 (31.5%)</td>
<td>9 (0.4%)</td>
</tr>
<tr>
<td>5 WRF-MT</td>
<td>19 (13.4%)</td>
<td>123 (55.4%)</td>
<td>77 (34.7%)</td>
<td>47 (21.0%)</td>
</tr>
<tr>
<td>6 CENER</td>
<td>13 (9.2%)</td>
<td>129 (58.1%)</td>
<td>87 (39.0%)</td>
<td>30 (13.6%)</td>
</tr>
<tr>
<td>Austrian Persistence</td>
<td>0 (0.00%)</td>
<td>142 (64.3%)</td>
<td>91 (41.2%)</td>
<td>-14 (-6.4%)</td>
</tr>
<tr>
<td>NREL Persistence</td>
<td>0 (0.00%)</td>
<td>194.55 (54.36%)</td>
<td>116.32 (32.5%)</td>
<td>-0.23 (-0.06%)</td>
</tr>
</tbody>
</table>

Based on the relative improved RMSE values with respect to persistence given in Table 9 for Austria our proposed method stays in the third place after ECMWF-OL and BLUE. This means that UTSA Method outperforms WRF-MT and SYNOP. Notice that SYNOP is a real-time day-ahead approach done by human meteorology experts.

In Southern Spain with sunny climate, absolute error values are smaller than in Europe. Concerning relative error values, there is an additional effect by normalization to a higher average
irradiance in Spain. Results obtained by using Spanish data set are represented in Table 10 following:

Table 10: Actual and Relative RMSE, MAE, and BE for Next Day Forecast using Spanish Data Set (2007 and 2008) and for the Proposed Method using Golden CO, USA Data Set (2012).

<table>
<thead>
<tr>
<th>Approach</th>
<th>$\tilde{\eta}<em>{\text{RMSE}}$ (%$\eta</em>{\text{RMSE}}$)</th>
<th>$\text{RMSE(W/m}^2\text{)}$ (%$\eta_{\text{RMSE}}$)</th>
<th>$\text{MAE(W/m}^2\text{)}$ (%$\eta_{\text{MAE}}$)</th>
<th>$\text{BE(W/m}^2\text{)}$ (%$\eta_{\text{BE}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ECMWF-OL</td>
<td>44.88 (35.7%)</td>
<td>81 (20.8%)</td>
<td>48 (12.2%)</td>
<td>1 (0.1%)</td>
</tr>
<tr>
<td>2 WRF-UJAEN</td>
<td>36.24 (28.8%)</td>
<td>89.64 (22.9%)</td>
<td>57.64 (14.7%)</td>
<td>28.26 (7.2%)</td>
</tr>
<tr>
<td>3 Proposed Method (NREL Golden CO)</td>
<td>45.26 (23.3%)</td>
<td>149.29 (41.69%)</td>
<td>98.09 (27.39%)</td>
<td>8.91 (2.49%)</td>
</tr>
<tr>
<td>4 CENER</td>
<td>25.88 (20.6%)</td>
<td>100 (25.4%)</td>
<td>64 (16.3%)</td>
<td>-3.33 (-0.8%)</td>
</tr>
<tr>
<td>5 HIRLAM-CI</td>
<td>1.82 (0.01%)</td>
<td>124.06 (31.7%)</td>
<td>79.87 (20.4%)</td>
<td>58.87 (15.0%)</td>
</tr>
<tr>
<td>Spanish Persistence</td>
<td>0 (0.00%)</td>
<td>125.88 (32.1%)</td>
<td>64.91 (16.6%)</td>
<td>2.44 (0.6%)</td>
</tr>
<tr>
<td>NREL Persistence</td>
<td>0 (0.00%)</td>
<td>194.55 (54.36%)</td>
<td>116.32 (32.5%)</td>
<td>-0.23 (-0.06%)</td>
</tr>
</tbody>
</table>

With regards to the relative improved RMSE values the proposed method follows the ECMWF-OL and WRF-UJAEN in the third place. It means the proposed method outperforms CENER and HIRLAM-CI based on the performance improvement measure $\eta_{\text{RMSE}}$.

The relative performance improvement with respect to persistence, i.e. $\eta_{\text{RMSE}}$, is a universal reliable measure for comparing different approaches applied to different data sets. This measure accounts for the climate difference due to various geographical conditions and timings of
different studies by considering the percent improvement in RMSE or other performance criteria over the study period with respect to those of the local persistence. Since the persistence approach naturally accounts for site-specific climate patterns then the relative comparison of a specific method’s performance applied to a data set to that of the persistence on the same data set is much likely to account for special characteristics of different data sets and provide an almost unified measure to compare the performances of various methods applied to different datasets with each other. This is most useful when it is not either possible or feasible to apply different datasets to the same forecasting method for direct performance comparison.

5.4 Day-Ahead Solar Power Forecast Performance Analysis for San Antonio, TX

The day-ahead forecast of expected energy from the Blue Wing solar plant is implemented using the dataset obtained from automated surface observing system (ASOS) for San Antonio Airport (SAT) weather station and the DC power output of the Blue Wing\(^2\) solar plant for the duration of the year 2012. The ASOS-SAT dataset included the temperature, the dew point, the atmospheric pressure obtained by three different pressure sensors, the wind velocity and wind direction. The atmospheric pressure was computed by doing the averaging of the three sensory measurements for pressure in order to get a single value associated with the pressure variable. The temperature in combination with the dew point was used to compute the relative humidity (RH) according to the following simple approximation:

\[
RH \approx 100 - 5 \times (T - T_{DP})
\]  
Eq. 54

Where \(RH\) is the relative humidity, \(T\) and \(T_{DP}\) represent the temperature and the dew point temperature respectively. This can be inferred from Eq. 54 that:

---

\(^2\) The 14-MW Blue Wing Solar photovoltaic (PV) installation in southeast San Antonio began generation in 2011. [CPS Website] The Blue Wing solar plant is located in 29.289822˚ Latitude and -98.424967˚ Longitude. [Lat Long]
For every 1°C difference in the dew point temperature and the temperature, the relative humidity decreases by 5%, starting with RH = 100% when the dew point equals the ambient temperature. It must be noticed that the dew point temperature is always equal to or smaller than the dry bulb temperature.

The Zenith angle ($\theta_Z$) and the Azimuth angle ($\theta_A$) for the Blue Wing solar plant were computed using the NREL solar positioning (SOLPOS) calculator for the duration of the year 2012 considering horizontal solar panels. The latitude and the longitude of the Blue Wing solar plant were used for computing $\theta_Z$ and $\theta_A$. The Blue Wing solar plant is located on the latitude of N 29˚ 17' 23.3596" and longitude of W 98˚ 25' 29.883" which translate into latitude of N 29˚ 17.389326’ and the longitude of W 98˚ 25.49805’’. The final values of latitude and longitude are then obtained in degrees as +29.289822’’ and -98.424957’’ [Lat Long]. The values of latitude and longitude are fed into the NREL-SOLPOS calculator in order for the Zenith and the Azimuth angles to be computed during the year 2012.

The prediction of the day-ahead expected solar power during the year 2012 using the San Antonio data set for the SAT weather station and the Blue Wing solar plant is implemented considering current practice in the industry. This means that the forecast horizon is between 18 hours ahead to 42 hours ahead of current time [UCSD 2014]. Table 11 summarizes the results for day-ahead solar power forecast using different scenarios together with the day-ahead persistence as the reference case for comparison. The forecast
<table>
<thead>
<tr>
<th>Approach</th>
<th>r</th>
<th>BE (MW)</th>
<th>BE_rel ( % )</th>
<th>MAE (MW)</th>
<th>MAE_rel ( % )</th>
<th>RMSE (MW)</th>
<th>RMSE_rel ( % )</th>
<th>μ ( % )</th>
<th>σ ( % )</th>
<th>MAPE ( % )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day-Ahead Persistence</td>
<td>0.76</td>
<td>-0.02</td>
<td>-0.45%</td>
<td>1.8</td>
<td>34%</td>
<td>2.95</td>
<td>55.75%</td>
<td>0.73%</td>
<td>25.29%</td>
<td>17.21%</td>
</tr>
<tr>
<td>Day-Ahead Forecast</td>
<td>0.834</td>
<td>-0.49</td>
<td>-9.32%</td>
<td>1.64</td>
<td>30.96%</td>
<td>2.39</td>
<td>45.25%</td>
<td>-6.81%</td>
<td>20.79%</td>
<td>17.04%</td>
</tr>
</tbody>
</table>

Based on the performance measures represented in Table 11, it can be seen that the day-ahead proposed method outperforms the day-ahead persistence with respect to every aspect except for the bias error (BE) which can be regarded as a negative offset in the forecast results. This means that the forecast results are in average 9.32 percent less than the measured day-ahead values. However, in order to correct this offset some complicated statistical approaches are required so that appropriate corrections can be applied to the forecast results when required.

Target vs. forecast scatter plot, actual error histogram and percent error histogram are represented in Figure 51 for day-ahead proposed forecast method and the day-ahead persistence. The measures are calculated only considering the hourly averaged solar energy of daytime samples.
Figure 51: Day-Ahead Solar Energy Forecasting Performance
(a) Scatter Plot; Persistence, (b) Scatter Plot; Proposed Method, (c) Actual Error Histogram; Persistence, (d) Actual Error Histogram; Proposed Method, (e) Percent Error Histogram; Persistence, (f) Percent Error Histogram; Proposed Method.
From Figure 51 it is seen that the proposed approach strongly outperforms the day-ahead forecast based on the results obtained for the Blue Wing solar plant in San Antonio during the year 2012.

Figure 52: Monthly performance analysis for 2012 Day-Ahead Solar Power Forecast of The Blue Wing Solar Plant; Actual Error (TOP Diagram), Relative Error (BOTTOM Diagram), Persistence (BLACK), Proposed Method (BLUE)
Monthly analysis of the day-ahead forecast performance is also considered for the proposed approach in Figure 52 and is compared to that of the day-ahead persistence reference to get a better understanding of the improvements that the adaptive framework brings to the solar irradiance forecasting. The monthly performance charts are represented in Figure 52 in following.

Comparison between performance of the proposed day-ahead solar power forecasting approach and those of the day-ahead solar forecasting benchmarks are given in the following.

For Austria, the forecasts of the University of Oldenburg, Meteotest, CENER, the statistical method of Blue Sky, and the traditional synoptic method of the meteorologists of Blue Sky (SYNOP) were evaluated for day-ahead forecast. The SYNOP method is spatially restricted and only available for Austria. Results are represented in Table 12.

<table>
<thead>
<tr>
<th>Approach</th>
<th>$\eta_{\text{RMSE}}$ ($\eta_{\text{RMSE}}$)</th>
<th>RMSE (RMSE$_{\text{Rel}}$)</th>
<th>MAE (MAE$_{\text{Rel}}$)</th>
<th>BE (BE$_{\text{Rel}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> BLUE</td>
<td>57 (W/m²) (36.1%)</td>
<td>101 (W/m²) (45.2%)</td>
<td>64 (W/m²) (28.4%)</td>
<td>1 (W/m²) (0.5%)</td>
</tr>
<tr>
<td><strong>2</strong> ECMWF-OL</td>
<td>51 (W/m²) (32.3%)</td>
<td>107 (W/m²) (47.4%)</td>
<td>69 (W/m²) (30.6%)</td>
<td>14 (W/m²) (6.1%)</td>
</tr>
<tr>
<td><strong>3</strong> SYNOP</td>
<td>47 (W/m²) (29.7%)</td>
<td>111 (W/m²) (49.3%)</td>
<td>70 (W/m²) (30.9%)</td>
<td>7 (W/m²) (2.9%)</td>
</tr>
<tr>
<td><strong>4</strong> Proposed Method</td>
<td><strong>0.56 (MW)</strong> (19.0%)</td>
<td><strong>2.39 (MW)</strong> (45.25%)</td>
<td><strong>1.64 (MW)</strong> (30.96%)</td>
<td><strong>-0.49 (MW)</strong> (-9.32%)</td>
</tr>
<tr>
<td>(The Blue Wing Plant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5</strong> WRF-MT</td>
<td>26 (W/m²) (16.5%)</td>
<td>132 (W/m²) (58.5%)</td>
<td>84 (W/m²) (37.1%)</td>
<td>30 (W/m²) (13.1%)</td>
</tr>
<tr>
<td><strong>6</strong> CENER</td>
<td>24 (W/m²) (15.2%)</td>
<td>134 (W/m²) (59.5%)</td>
<td>90 (W/m²) (39.8%)</td>
<td>21 (W/m²) (9.4%)</td>
</tr>
</tbody>
</table>

Based on the relative improved RMSE values with respect to persistence given in Table 12 for Austria the proposed method stays in the fourth place after ECMWF-OL, BLUE and the SYNOP for day-ahead solar forecast with respect to the relative RMSE improvement measure, i.e. $\eta_{\text{RMSE}}$. This means that the proposed method outperforms WRF-MT and CENER.
The relative performance improvement with respect to persistence, i.e. $n_{RMSE}$ is a universal reliable measure for comparing different approaches applied to different data sets. This measure accounts for the climate difference due to various geographical conditions and timings of different studies by considering the percent improvement in RMSE or other performance criteria over the study period with respect to those of the local persistence. Since the persistence approach naturally accounts for site-specific climate patterns then the relative comparison of a specific method’s performance applied to a data set to that of the persistence on the same data set is much likely to account for special characteristics of different data sets and provide an almost unified measure to compare the performances of various methods applied to different datasets with each other. This is most useful when it is not either possible or feasible to apply different datasets to the same forecasting method for direct performance comparison.

5.5 Day-Ahead Solar Power Forecast Value Analysis for San Antonio, TX

5.5.1 Forecast Value Preliminaries

Forecasting of renewable energy sources such as solar and wind becomes crucial for reliable bidding of electricity into the market by load serving entities (LSE) including the utility companies. This becomes even more exaggerated when the amount of solar penetration reaches a certain threshold.

The value of a forecast can be evaluated both from a power generator’s perspective or that of a system operator, since both are users of forecasts. The information on the value of a solar energy forecast is helpful in determining the level of investment into the forecast. Modeling the value of a solar energy forecast to a system operator is very sophisticated and requires knowledge of resources available, costs of start-up and running costs of those resources, and the unit
commitment rules used by the system operator. The factors already mentioned may vary among system operators and consequently such a study would be specific to each ISO. Large-scale studies such as western wind and solar integration study (WWSIS) have already generalized such factors to investigate cost savings due to forecasts of wind and solar energy. In this section, however, the spread between the day-ahead market (DAM) and real-time market (RTM) prices and its correlation to day-ahead solar forecast error will be investigated to assign a more general value of solar forecasts to market participants such as solar power generators and load serving entities.

The CPS Energy of San Antonio is among the 270+ nodes in the ERCOT nodal market. CPS Energy has an ongoing project which targets to develop more than a total sum of 400 MW of grid-connected solar plants over the next four years [CPS Energy]. Currently, the 14 MW Blue Wing plant together with the 41 MW Alamo 1 project and three Sun Edison solar plants that total 36 MW are the solar providers for the San Antonio area [CPS Energy]. Considering the electricity locational marginal prices (LMP) in the day-ahead market (DAM) and the real-time market (RTM) of ERCOT ISO the general value of the proposed solar forecasting framework will be evaluated based on the methodology suggested in [UCSD].

5.5.2 Energy Market and Forecast Value Analysis Methodology

DAM LMP (the market price at which a day-ahead forecast is committed) and RTM LMP (the market price at which settlements are made) from January 1, 2011 to December 31, 2012 were used in this study. The hour-beginning DAM LMP (e.g. the 0900 DAM LMP applies to 0900–1000) are averaged to correspond to the instantaneous hourly forecast data (e.g. for 0900, the mean of the 0800 and 0900 DAM LMP is used). For the RTM, LMP are reported every five minutes and are valid through the next five minutes (e.g. the 0900 RTM LMP is used for 0900–0905). To determine hourly RTM LMP, prices were averaged between ±30 min from the hour (e.g. for 0900,
the mean of values from 0830–0925 RTM LMP were taken. Figure 53 represents the instantaneous electricity price in the ERCOT energy market for the year 2012.

![Figure 53 Instantaneous ERCOT Electricity Price for the Year 2012; Day-Ahead Market (DAM) and Real-Time Market (RTM) Locational Marginal Price (LMP)]

In order to calculate revenue from solar energy sales during the year 2012, it is assumed that solar photovoltaic (PV) plants participate in the wholesale energy market like other generating resources. Furthermore, it is assumed that the 2012 LMP does not change due to additional PV plants participating in the market. In reality, DAM prices would be reduced with increasing solar
generation penetrated into the market and the RTM price could fluctuate depending on systematic solar forecast biases and daily production values [UCSD_16].

Since forecasts are submitted in the DAM and the deviations are settled in the RTM, the value of a forecast must be dependent on the difference between the DAM and RTM LMPs in addition to the magnitude of forecast error, as represented in Table 13. For example, if LMP is greater than the DAM LMP, energy will be then more expensive to procure in real-time. Hence, if an over-forecast occurs there will be a large loss in revenue as additional units of energy will need to be purchased at the higher RTM price in order to compensate the under-delivery that is resulted due to the day-ahead forecast deviation in order to meet the real-time demand. Conversely, if the RTM LMP is greater than the DAM LMP, under-forecasts are potentially profitable as excess energy produced in the DAM can be sold at the higher RTM price. However, energy is not guaranteed to sell in the RTM. In fact, a significant excess of energy in the RTM at a particular node is likely to drive the local RTM price down, negating any potential profit. For this reason, excess energy sold in the RTM will be considered only as a potential gain in revenue and RTM LMP will be set to zero when under-forecasts occur. The case when the RTM price is less than the DAM price is also considered. Here, under-forecasts result in a loss of revenue as the full amount of produced energy was not sold in the higher-priced DAM. In this situation, over-forecasts produce higher revenues than accurate forecasts as energy can be procured in the RTM inexpensively [UCSD].

Table 13 summarizes the possible outcomes considering forecast error and the difference between the RTM and DAM and mentions the restrictions that are imposed on the variables in order to make the value analysis more realistic.
Table 13 Summary of Market / Forecast Outcomes and Restrictions

<table>
<thead>
<tr>
<th>LMP_{DAM} – LMP_{RTM}</th>
<th>Forecast Bias</th>
<th>Outcome</th>
<th>Restrictions Imposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0</td>
<td>Over-Forecast</td>
<td>Purchase additional energy at the higher RTM price: <em>loss of revenue</em></td>
<td>LMP_{RTM} = \max(0, LMP_{RTM})</td>
</tr>
<tr>
<td>&gt; 0</td>
<td>Over-Forecast</td>
<td>Purchase additional energy at the lower RTM price: <em>gain of revenue</em></td>
<td>LMP_{RTM} = \max(0, LMP_{RTM})</td>
</tr>
<tr>
<td>&lt; 0</td>
<td>Under-Forecast</td>
<td>Sell additional energy at higher RTM price: <em>gain of revenue</em></td>
<td>None'</td>
</tr>
<tr>
<td>&gt; 0</td>
<td>Under-Forecast</td>
<td>Sell additional energy at lower RTM price: <em>loss of revenue</em></td>
<td>None</td>
</tr>
</tbody>
</table>

If a deviation penalty factor (DPF) is not considered for value analysis then a restriction of LMP_{RTM} = \min(0, LMP_{RTM}) must be imposed in this case. This translates to only the slightest possibility of revenue gain by selling additional energy in RTM if penalty is not considered in the value analysis process.

5.5.3 Revenue and Forecast Value Analysis

To determine forecast value, the total annual revenue obtained from the sale of solar energy, \( R \), is calculated. In general, \( R \) is calculated by first selling the forecasted energy in the DAM and then settling (either selling excess energy or purchasing energy to meet shortfalls) in the RTM. In some cases, LMP (especially in the RTM) may be negative. This generally occurs when there is an oversupply of energy or transmission difficulties near a single node. In these instances, energy suppliers are paid to curtail production. To ensure that negative LMP are not unfairly beneficial, energy is never allowed to be purchased at a negative price in this study. However, energy is allowed to be procured at zero cost [UCSD]. Hence, when an over-forecast occurs, negative LMP_{RTM} values are set equal to zero, as described in Table 13. It is also known that when under-forecasts occur, there is no guarantee that energy can be sold in the RTM.
This can be simulated by setting positive values of LMP$_{RTM}$ to zero in those situations, as mentioned in Table 13 comment. To determine the value of forecasts, it is assumed that a non-revenue-biased bidding strategy$^3$ is used. In order to obtain more realistic value analysis results the deviation penalty, i.e. $P_{deviation}$, is also considered. A positive constant coefficient between 1 and 2 is used as the deviation penalty factor (DPF) in order to penalize the day-ahead forecast deviations from tomorrow’s actual measured solar power. The penalty is computed as represented in Eq. 55.

$$P_{deviation}[h] = DPF \times \max(0, LMP_{DAM}[h] - LMP_{RTM}[h])$$

$$\times |E_{Measurement}[h] - E_{Prediction}[h]|$$

Eq. 55

where DPF is the deviation penalty factor. Any DPF greater than one ensures that the maximum possible revenue results from using an unbiased forecast. In this study, the DPF ranges from 1 to 2. However, the results obtained by a factor of 1.5 are considered to not only ensure that the deviation penalty is always 50% larger than the possible revenue gain of a biased forecast but also make it possible to compare the results obtained from current research to other sources in the literature. A proper DPF ensures that the highest quality forecast is the most valuable forecast without excessively diminishing total revenue for the solar energy generator (i.e. if the DPF is too high solar energy generators will be harshly penalized for providing even near perfect forecasts) [UCSD]. Annual revenue obtained by using a forecast approach is computed as represented in Eq. 56.

---

$^3$ Non-revenue-biased bidding strategy means that the solar power providers bid exactly at the forecast level and do not consider possibilities of increasing their revenues through bidding at amounts which differ from what is represented by the forecast curve [UCSD].
\[ R_{Prediction} = \sum_{h=1}^{H} E_{Prediction}[h] \times (LMP_{DAM}[h] - LMP_{RTM}[h]) + E_{Measurement}[h] \times LMP_{RTM}[h] - P_{deviation}[h] \]  

Eq. 56

Where \( R_{Prediction} \) is the revenue from using the prediction curve obtained by any specific forecast method, \( E_{Prediction} \) stands for the predicted hourly averaged solar energy, \( LMP_{DAM} \) and \( LMP_{RTM} \) respectively represent the electricity price in the day-ahead market (DAM) and the real-time market (RTM), \( E_{Measurement} \) shows the measured hourly averaged solar energy in the target day and \( P_{deviation} \) is the deviation penalty due to the difference between the solar forecast curve and the actual measured solar energy. Finally, \( h \) represents the hour number and changes between 1 and \( H \) which is equal to 24*366 = 8784 for the year 2012.

The value analysis process is done for both the proposed forecast method and the persistence for the day-ahead prediction of solar energy at the Blue Wing solar plant in San Antonio, Texas. The electricity price in the ERCOT DAM and RTM are used for this analysis in the year 2012. The resulting revenue of solar energy sale in the ERCOT electricity market for the proposed method and the day-ahead persistence are given in the following. The revenue improvement that the proposed forecast framework brings to a load serving entity (LSE) such as the CPS Energy of San Antonio compared to the industry’s current practice, i.e. the day-ahead persistence, is determined and discussed for different values of DPF between 1 and 2. Two revenue improvement factors are introduced in Eq. 57 and Eq. 58.
\[ \theta_R = R_{\text{Forecast}} - R_{\text{Persistence}} \]  
\[ \eta_R = 100 \times \frac{\theta_R}{R_{\text{Persistence}}} \]

Eq. 57  
Eq. 58

Where \( \theta_R \) and \( \eta_R \) are called the actual and the relative revenue improvement factors with respect to the persistence approach. \( \theta_R \) has the same unit of revenue since it represents the monetary improvement in the revenue of the LSE using the proposed forecasting framework over the day-ahead persistence. \( \eta_R \), however, is represented in percent values since it is a measure of relative revenue improvement. Obviously, a negative value for \( \theta_R \), and consequently for \( \eta_R \), infers that the solar forecasting method the value of which is being analyzed does not outperform the day-ahead persistence and therefore, it would be better to rely on the day-ahead persistence for the relevant data set or region during the specific study period.
From Figure 54, it can be seen that using the proposed forecasting framework can increase the annual revenue of solar energy sales in the year 2012 by a factor of 100 \times \frac{410600 - 358700}{358700} \approx 14.5\% over the persistence approach if the DPF of 1.5 is applied to penalize the forecast deviations.
from the measured values. The relative improvement is about 9.5% for the DPF of 1 as represented in Figure 54 and Figure 55. As it can be seen from Figure 54 total proposed method and the day-ahead persistence drops as the DPF increases and this is because of harsher penalties applied to the forecast deviations from the measured solar energy. This must noted though, that the revenue obtained by the perfect forecast does not change for different values of DPF since there is no deviation in the forecasts from the measured solar energy to be penalized.

The red diagram in Figure 54 represents the amount of actual revenue improvement, i.e. DPF. It can be seen that the $\theta_{R}$ increases as the DPF grows larger. For the DPF equal to 1.5 the monetary value of actual revenue improvement using the proposed forecast framework over the day-ahead persistence is almost $52,000 for the 14 MW Blue Wing solar plant of San Antonio, TX in the year 2012.
Figure 55 Annual Revenue Relative to Perfect Forecast and the RIF Obtained by the Proposed Forecasting Method versus DPF for the Blue Wing Solar Plant during the Year 2012

As represented in Figure 55, the relative RIF for the proposed day-ahead solar forecasting framework over the day-ahead persistence changes from 9.5% to 14.8% for values of DPF between 1 and 1.5 used in the value analysis. This is the relative RIF obtained for the 14 MW Blue Wing solar plant during the year 2012. The annual revenue of solar energy sales in the
ERCOT market, on the other hand, drops from 72.6% to 56.6% of the percent revenue obtained relative to the perfect day-ahead forecast. This is because of the fact that the deviations in the forecast from the measured values are penalized harsher as the DPF increases and therefore the final revenue drops as the penalty factor grows. As it can be seen from the diagram of Figure the annual percent revenue obtained by the proposed forecast method is 56.6% in terms of the perfect forecast revenue. This is strongly in accordance with the results that is obtained by a similar study done by the University of California San Diego (UCSD) on 63 stations in the California ISO (CAISO) for the case of revenue performance of NAM_Pen [UCSD].

In Figure 5 below the result of extending the value analysis approach to other capacities is represented versus DPF. The results confirm that by increasing the DPF the percent improvement that using the proposed forecast brings to a load serving entity such as the CPS Energy of San Antonio increases. However, at the same time, the actual revenue is decreasing compared to that of a perfect forecast due to harsher penalizing effects of a larger DPF. The increase in the actual RIF and the relative RIF, then, can be explained by the fact that the revenue obtained by the persistence approach drops at a higher slope compared to that of the proposed forecast method.
As it can be seen from Figure 56, for an increase in the DPF, the relative RIF, i.e. while the annual percent revenue drops. This trend is similar for different capacities of solar plant considered in this study. PV plant capacities considered in this study include 14 MW, 50 MW, 100 MW, 200 MW and 500 MW and the DPF gets values between 1 and 2 with increments of 0.2. The ERCOT DAM and RTM electricity price for the year 2012 and the forecast results obtained using the proposed framework and the day-ahead persistence of for the Blue Wing solar plant are used to obtain approximate value analysis results for PV capacities beyond 14 MW.
5.6 Summary

Performance and value of the proposed solar forecasting framework were analyzed and results were discussed. Performance of the proposed framework for next day solar energy prediction at NREL site in Golden, Colorado and for the day-ahead solar power forecast of the Blue Wing solar plant at San Antonio were determined. The proposed solar forecasting framework improves the forecasting RMSE by more than 23% over the next day persistence. Value of the proposed forecasting framework for day-ahead solar power prediction of the 14 MW Blue Wing solar PV plant at San Antonio, Texas in year 2012 was evaluated and compared to similar studies in California. ERCOT DAM and RTM electricity price together with the proposed forecast and the day-ahead persistence of the Blue Wing solar plant were used for the value analysis process. The proposed forecasting framework improves solar energy sales annual revenue by 14.8% over the day-ahead persistence. The suggested forecasting framework strongly outperforms the persistence reference case in both performance and value. It also compares equal to the current benchmark methods in solar energy forecasting and performs better than some of the state-of-the-art methodologies in solar forecasting.
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

An intelligent GFS-based framework was introduced for evolutionary qualitative decision-making on storage energy flow control in electrical microgrids. Simulation was done for several cases including different microgrid structures and control strategies. The proposed GFS-based intelligent decision-making framework proved capable of cost-efficient energy management in microgrids under constraints of meeting the demand profile, meeting battery SOC constraints, handling renewable electricity intermittency, and complying with environmentally-friendly considerations. Results suggested that for the electrical microgrids where storage units are already deployed the use of an intelligent decision-making approach for storage energy flow control will bring an ROI of between 5 to 10 for the microgrid owner while gives the leverage to meet the environmental standards by reducing the amount of contribution to air pollution through rule-base modification if need be.

The proposed intelligent GFS-based framework can also be used as a tool for solving multi-objective function problems where more than one optimization target of interest exist. The use of genetic algorithms as a heuristic search and evolution tool in order to improve the capabilities of the fuzzy logic-based decision-making approach can be considered for further expanding the solutions to multi-objective function problems. Topics of further research related to this research include:

- Increasing the number of nodes in the microgrid network specifically regarding the number, size and location of storage units,
- Accounting for demand-side load management in the simulation and considering its effect on the overall process,
Enhancing the proposed decision-making framework by using the solar energy forecast and load profile forecast outcomes in order to prepare the system in advance for upcoming changes in the climate and in the network.

The adaptive data analytic-based solar energy forecasting framework introduced in this research combines data mining with statistical approaches and artificial intelligence in the form of artificial neural networks (ANN) structures in order to provide a novel methodology for non-parametric prediction. The proposed method results in more than 23% performance improvement and about 14.8% value improvement over the persistence method for solar forecasting. Due to the depth and breadth of possible approaches to solar energy forecasting and considering the amount of work, including the steps for data analytic, feature extraction, artificial intelligence, and possible forecasting scenarios, presented in this dissertation there are still whole new problems that can be tackled through implementing advanced procedures to enhance the already existing framework with more capabilities or to elevate its performance and value. A list of these potential works along with brief description is given in the following:

- Include the Luni-Solar variables, i.e. variables accounting for Earth’s position in the sky with respect to the Sun’s and the Moon’s position relative to that of Earth’s, inclination angle of earth with respect to its axis etc. in the database to enrich the data by some variables that contain information regarding the changes in months, seasons and years.

- Increase synoptic event detection algorithm resolution by considering the types of hours (TOH) in addition to the types of days (TOD) to achieve higher quality fine-tuning and better forecast accuracy and to prepare the framework for dynamic real-time applications.
Smart automatic rule update scheme in order to modify adaptive training rules and the fine-tuning rules due to climate characteristics or geographic diversities of different sites or because of long-term changes in the global weather conditions etc. so that the performance of the system can be maintained.

Modify the proposed forecasting framework accordingly in order to apply to other categories including but not limited to wind forecasting, electric load forecasting, weather variables forecasting such as temperature, pressure or relative humidity etc.
APPENDIX A: SOLAR IRRADIANCE PARAMETER DEFINITIONS

The following quantities associated to solar radiation are commonly measured:

- Direct Normal Irradiance (I_{DNI}): is the energy flux density, in (\frac{W}{m^2}), of the solar radiation incoming from the underlying solid angle by the sun’s disk on a unitary surface perpendicular to the rays.
- Diffuse Horizontal irradiance (I_{DHI}): represents the energy flux density of the solar radiation incoming from the entire sky dome on a horizontal surface, excluding the direct beam coming from the sun’s disk.
- Global Horizontal Irradiance (I_{GHI}): is the total energy flux density a horizontal surface on the earth is exposed to.

Lambert’s cosine law states that the energy flux density on a plane surface is directly proportional to the cosine of the incidence angle [27]. Since the incidence angle of the solar beam striking the horizontal ground is equal to the sun’s zenith angle, as represented in Figure 57, the relationship between GHI with DHI and DNI is represented by \textbf{Eq. 59}:

\[ I_{GHI} = I_{DHI} + I_{DNI} \times \cos(\theta_z) \]  
\textbf{Eq. 59}

The term “global” is associated to the fact that the solar radiation is received from the entire 2\pi solid angles of the sky vault [27].
Figure 57: Angles describing the position of the sun: $\theta_z$ - zenith angle; $h$ - elevation angle; $\mu$ - azimuth angle. Angles describing the position of the surface: $\beta$ - slope or tilt angle; $\mu$ - surface azimuth angle. The incidence angle $\theta$ is the angle between the sun direction and the surface’s normal $\vec{n}$ [27].

The global tilted irradiance ($I_{GTI}$) or the total irradiance received by a surface tilted with an angle $\beta$ with respect to the horizontal plane (Figure 57) is the sum of direct flux density, flux density, and the additional flux density of the solar radiation reflected from the ground ($I_{RTI}$), respectively. Usage of Eq. 59 yields:

$$I_{GTI} = I_{Direct} + I_{Diffused} + I_{Reflected}$$

s.t. \[ \begin{align*}
I_{Direct} &= I_{DNI} \times \cos(\theta) \\
I_{Diffused} &= R \times I_{DHI} \\
I_{Reflected} &= I_{RTI}
\end{align*} \] Eq. 60

where $\theta$ is the incidence angle (i.e., the angle between the sun direction and the normal to the surface (Figure 57)), $R$ is the conversion coefficient taking into account the sky view factor due to the tilt, and $I_{RTI}$ is the energy flux density of radiation reflected by the ground that is intercepted by the tilted surface. Models for estimating global solar irradiance on tilted surfaces differ generally in their treatment of $R$ which is considered the main potential source of errors [27].


[23] DIN IEC 751, “Permissible Deviations for Class A and Class B Platinum Sensors” *Form Number F038-00-00 Revision 0* — ©2000 Lake Shore Cryotronics, Inc. — 18 January 2000


JANE DOE

Jane Doe is from San Antonio, TX. She studied higher education administration and earned both a Bachelor’s and Master’s degree in Education from The University of Texas at San Antonio. Her future plans include attending a Ph.D. program.

Vita Instructions:

- The Vita is required.
- Indent the first line of each paragraph.
- The Vita is a short biographical summary. Students typically include information such as where they are from, what they have studied in the past, degrees they have earned, and future endeavors.
- Do not include contact information, such as addresses, emails, phone numbers, etc.
- Do not include a resume, curriculum vitae, or list of publications.
- The Vita should be written in THIRD PERSON, paragraph style.
- The text should be double-spaced.
- The Vita may not exceed one page.