Photovoltaic Systems: Forecasting for Demand Response Management and Environmental Modelling to Design Accelerated Aging Tests

by

Nikhil Kadloor

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2017 by the
Graduate Supervisory Committee:

Joseph Kuitche, Co-Chair
Rong Pan, Co-Chair
Teresa Wu

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Distributed Renewable energy generators are now contributing a significant amount of energy into the energy grid. Consequently, reliability adequacy of such energy generators will depend on making accurate forecasts of energy produced by them. Power outputs of Solar PV systems depend on the stochastic variation of environmental factors (solar irradiance, ambient temperature & wind speed) and random mechanical failures/repairs. Monte Carlo Simulation which is typically used to model such problems becomes too computationally intensive leading to simplifying state-space assumptions. Multi-state models for power system reliability offer a higher flexibility in providing a description of system state evolution and an accurate representation of probability. In this study, Universal Generating Functions (UGF) were used to solve such combinatorial problems. 8 grid connected Solar PV systems were analyzed with a combined capacity of about 5MW located in a hot-dry climate (Arizona) and accuracy of 98% was achieved when validated with real-time data. An analytics framework is provided to grid operators and utilities to effectively forecast energy produced by distributed energy assets and in turn, develop strategies for effective Demand Response in times of increased share of renewable distributed energy assets in the grid. Second part of this thesis extends the environmental modelling approach to develop an aging test to be run in conjunction with an accelerated test of Solar PV modules. Accelerated Lifetime Testing procedures in the industry are used to determine the dominant failure modes which the product undergoes in the field, as well as predict the lifetime of the product. UV stressor is one of the ten stressors which a PV module undergoes in the field. UV exposure causes browning of modules leading to drop in Short Circuit Current. This thesis presents an environmental modelling approach for the

hot-dry climate and extends it to develop an aging test methodology. This along with the accelerated tests would help achieve the goal of correlating field failures with accelerated tests and obtain acceleration factor. This knowledge would help predict PV module degradation in the field within 30% of the actual value and help in knowing the PV module lifetime accurately.

To,

My parents, Nanda Kumar Kadloor and Chandrika Kadloor, and other friends and family members for their constant support and love

ACKNOWLEDGMENTS

I would like to thank Dr. Govindsamy Tamizhmani and Dr. Joseph Kuitche for giving me an opportunity to work at Photovoltaic Reliability Lab. As a result, I have learned a great deal and gained a lot of hands on experience in working in the PV industry. I have also gained experience putting analytics into use for the Energy Industry and easing Grid Integration of renewables. Their continued guidance and support has helped me gain confidence to continue my work. I would also like to Dr. Pan for agreeing to be my thesis Co-Chair and for always supporting me. I would like to thank Dr. Wu for agreeing to be my thesis committee member. Finally, I would also like to thank Sai Tatapudi, Aravind Srinivasan and all my other colleagues at PRL for their support and love.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

PART 1: PHOTOVOLTAIC SYSTEMS: FORECASTING FOR DEMAND RESPONSE MANAGEMENT

1.0 INTRODUCTION

1.1 Background

As the threat of climate change looms, countries across the world are setting up ambitious targets to achieve the goal of cutting carbon emissions. A big part of those plans is deploying renewable energy sources like Solar, Wind, etc. to comprise 20-30% of the country's electric grid. The biggest challenges in doing so are two-fold: 1. Tackling the intermittent nature of these renewable sources of energy which calls for effective forecasts to maintain predictability of energy supply. 2. Successful integration of independent distributed sources of energy into the grid. These challenges are key in bringing cost of renewable sources of energy equivalent to other conventional forms of energy. Markets have been very instrumental in bringing cost of manufacturing and deployment of Solar PV down to 12 cents/W. Aided with renewable tax credit (RTC) here in the US, Solar PV deployment is increasing at a rapid pace where many states have carved out ambitious targets for Solar PV.

As we move towards increased deployment of Solar PV, the focus of the industry has been slowly shifting from investing in efficiency gains to investing in quality and reliability. With most of the deployment slated to happen in the coming decade, Crystalline Silicon technologies have demonstrated their leadership in the market. This has called for increased focus on reliability of Crystalline Silicon PV modules. Although significant studies have been performed on PV module reliability, both from the perspective of module manufacturers and the PV system owners, not many significant

studies have been conducted evaluating PV system reliability from the perspective of a grid operator.

As more Solar PV rooftop systems come online, grid operators face an unique challenge. The electricity grid has been an aging system, built almost a century ago supporting only one-way transfer of energy. With the rise of renewable sources of energy, energy production is no longer concentrated but more distributed in nature, making people owning these systems not just mere consumers of energy from the grid but traders of energy. The trade of energy is entirely based on supply-demand fluctuations and has created the need for utilities and grid operators to come up with a customer friendly net-metering policies. However, with a grid built on one-way transfer of power, grid operators have seen escalating costs and reliability issues caused due to Solar PV energy production peaking at times of low-demand. Not only the nature of production of energy is changing, also the nature of energy consumption is changing with smart appliances, connected home devices and electric vehicles changing energy consumption patterns.

Smart meters and smart grids have been the way forward for many utilities, who have made significant investments to make the grid support such two-way energy transfer. Combining this with sophisticated IT systems powered with analytics, utilities and grid operators have been exploring new ways to remain profitable in this changing landscape. Concept of Virtual Power Plant and using analytics to manage Demand-Response rather than changing the supply side has proven viable for many utilities.

These changes have to brought in by focusing on the underlying energy forecasts and system reliability calculations, on which business models and operating models of these grid operators depend.

1.2 Scope of Work

This thesis deals with PV system reliability in the context of energy supplied. PV module degradation affects the reliability with which the Module supplies the rated energy. PV modules have two major sources of stochasticity- fluctuating weather and internal mechanical degradation. In this thesis, quantitative and statistical models were explored to model the weather and internal mechanical degradation. Also, a technique was developed in evaluating both the sources of fluctuations probabilistically and calculate its impact against satisfying the load requirements of the system. To demonstrate this, 8 rooftop ASU PV systems were considered and this analysis was carried out against each building's load. The long-term system reliability was measured using the reliability indices Loss of Load Expectation (LOLE) and Expected Energy Not Supplied (EENS).

There was also a need to develop short term energy forecasts which can be accurate enough so that the grid operators or the utilities can forecast energy production by each of their customers having Solar PV rooftop systems. This thesis provides with various machine learning models which can learn from previous environment data to provide accurate forecasts of energy.

The methodology detailed in this thesis provides an analytics framework for a grid operator to assess long term reliability of distributed energy assets which will help them with capacity planning (no. of backup transformers, energy storage, etc.) and financial planning. Short term forecasts help utilities design targeted Demand-Response plans using predictive analytics which can be specific to every customer's energy production/usage patterns.

2. LITERATURE REVIEW

Distributed Grid resource modelling has been a field which has gained a lot of importance as the share of renewable energy in the energy grid has increased. This has called for accurate mapping of resources which affect the generation of renewable energy. Traditionally simulation techniques were widely used to understand system behavior. These techniques are well developed and are widely used in many different industries to predict the behavior of systems. To develop a system reliability approach for a distributed grid, modelling behavior of every component in the grid is of utmost importance. This section is further divided into subsections detailing the prevailing literature on modelling individual system components and it's demonstrated use in the energy industry to effectively lower costs for grid operators and make it attractive for customers to depend on renewables to partly power their energy demands.

2.1 Weather Models

Weather fluctuation is the major cause for PV energy generation intermittency. The effect of solar irradiance fluctuation on PV Energy production and the power quality is well described by a study on 14 PV power plants by Patsalides M et al [1]. It was concluded that during low irradiance levels system generation was severely impacted. Effect of wind speed, Cell temperature and Irradiance levels on the performance of PV modules and PV arrays were modelled by King et al [2]. Electrical, thermal and optical characteristics of the PV modules are addressed in the above referred publication.

Weather modelling is of interest to many different industries and the science of numerical prediction of weather is well captured by Bauer et al [3]. Markov models are used majorly to model the transition between different weather states, as used by Billinton and Gao for assessing wind energy states [4]. For this thesis, emphasis was placed on statistical methods of weather fluctuation. Irradiance, Temperature and Wind speed were the variables of choice, which needed to be evaluated. R E Thomas et al developed an approach of "Environmental Cell" which is a discrete approach to model the changes in weather variables [5]. It involves measuring statistics pertaining to frequency of occurrence of a particular weather condition, and analyzing transition between different states of weather conditions using descriptive statistics. These help us achieve probabilistic states of weather which are used in Universal Generating Function approach developed by Ushakov [6].

Continuous models, namely time-series models are very popular in modelling weather variables like temperature. Pan et al demonstrate a ARIMA time-series approach to model the temperature of a hot-dry climate [7]. Though continuous models are more useful in describing the variable's fluctuations, their ability to create long term forecasts breaks down due to error accumulation. The environmental cell approach and the ARIMA approach were majorly explored in this thesis to achieve a simple probabilistic forecast for future time-periods. Other continuous models were further explored from the intention of generating short-term forecasts.

2.2 Degradation models

In Reliability studies, to effectively model system reliability one must carefully consider component degradation. For PV modules, Jordan et al have reviewed the various

degradation modes and assessed thousands of modules to find the median degradation rates for different climates [8]. It's concluded that for most of the PV modules, linear degradation has been observed, although it should be noted that for the study only 4% of the total population of PV modules have been evaluated, so such a conclusion might change subject to more degradation studies. Kuitche et al performed FMEA and data-mining techniques to find out the dominant failure modes and its impact on PV module reliability and performance [9]. Statistical reliability theory suggests various ways of modelling degradation [10]. Meeker & Escobar present various methods to assess time-to-failures of right censored and left censored data which can be used to assess time to failures of PV modules [11]. Pan et al used an ARIMA model of temperature and found a correlation between temperature and its impact on PV module degradation [7]. Mancenido demonstrated an online monitoring methodology to capture the non-stationary degradation as a stochastic process [12]. This method involves fitting a ARIMA model to power output of PV modules to model the degradation path. Though such methods are suitable, it's difficult to find raw KWh data of the lifetime of PV modules for such models to fit and PV module lifetime analysis is plagued by a lot of censored data.

Degradation can be modelled by various techniques. Markovian models are used widely in literature to model stochastic non-stationary process. Zuo et al demonstrated an approach to fit a markov model to degradation data to obtain failure and repair rates for the system [13]. Meeker presented a Monte-Carlo simulation based method to obtain failure rate and repair rate of a system using degradation data [14]. Park et al demonstrate the use of 10-state markov model, to model irradiance variation and the degradation function [15].

For Multi-state Reliability modelling Theristis et al demonstrated the use of Markov chains to model the system degradation and PV system Uptime and Downtime [16]. Using Markov chains, they calculate system Reliability Indices as Loss of Load Probability and Expected Energy Not Supplied, using which system sizing methodology and its impact on LCOE is shown. Markov models require less computation and steady state probabilities obtained by markovian models have proved their usage in reliability models. Kumar et al demonstrated the use of markovian models with 4 states of module operation- Working, partial degradation, major failures and complete failures. [17] Albeit the approach used in this thesis would be to characterize states as number of modules working or not working, where a module is considered not working when it reaches 20% degradation.

2.3 System Reliability

Calculating the system reliability is a combinatorial problem, involving combinations of multiple operating levels of different components and developing an analytical model to find out the net effect of all possible combinations. Billinton & Gao applied multi-state reliability models to wind generating units to assess system reliability [4]. Massim et al do the same to assess repairs/degradation of components [18].

Khattam et al demonstrate this using Monte-Carlo Simulation to find out the stability of the grid [19]. Monte-Carlo simulation being computationally expensive, often leads to simple binary state representations of complex systems. This leads to drop in accuracy. To evaluate the entire distributed grid, a less computationally expensive method is required. Ushakov demonstrated the use of UGF (Universal Generating Functions), which converts discrete probabilistic states into a polynomial, on which simple

mathematical operations can be done [6]. Li et al demonstrated the use of UGF to evaluate a distributed grid, although the solar PV states were simplified due to lack of degradation data [20]. This approach can be extended to this thesis, where significant degradation studies have been published by PRL, using which more accurate reliability estimates can be found out. Coupled with short term forecasts, a complete analytical tool can be developed giving grid operators long-term reliability estimates to plan and design the grid and the short term forecasts helping in better targeting of consumers for Demand-Response programs.

2.4 Short Term Forecasts

Forecasting is a very important science, which has gained relevance in the advent of very accurate forecasting techniques available today, due to the rise of machine learning and cheap computation. The recent advances in Deep Neural Networks has sparked interest in using the other learning algorithms traditionally used in the domains of computer vision and commerce.

Forecasting energy demand and supply has become ever important in a time where renewables are increasing their share in the grid. Since these sources are intermittent, there is a growing need to accurately estimate energy production of Solar PV assets. Perera et al have demonstrated the use of machine learning in generating energy forecasts for distributed grid components like Wind Generators and Solar PV generators [21]. Such a forecasting scheme is explained from the context of its application in a smart grid, complete with sensors at every point of the grid helping both energy producers and consumers have

more information. Zurborg explains how a concept of "Virtual Power Plant" powered by smart grids can help utilities and grid operators reduce operational cost as well as incentivize users to use distributed energy resources by offering better net-metering deals [22]. The analytical backbone for such a system requires accurate short term energy forecasts as well as long term reliability indices. Long term reliability indices have been explained in the previous sections. Short term forecasts of Solar Energy requires accurate forecasts of Irradiance and Ambient temperature, the two main variables effecting Solar Generator output in the short term. Pan et al have developed a ARIMA time series approach to generate forecasts of Temperature [7]. Various techniques like Neural networks have been used in previous literature in fault detection of PV modules [23]. Ren at al have described all the techniques prevalent today in Solar Irradiance and Wind forecasting, doing a comparative analysis between various ensemble techniques [24]. Out of all the techniques, Support Vector Machines have proved to be the best performing method to forecast Irradiance, especially its ability to model non-linear parameters well. SVMs have been extremely effective at non-linear learning tasks [25]. In PV energy forecasting, Sharma et al demonstrated use of Support Vector Regression in being accurate to predict Solar irradiance in the short term (1-2 days) [26]. Yerrapragada et al further developed this approach to produce a modified SVR called least squares SVR which gave more accurate results [27]. A Support Vector Regression technique was expanded in this thesis using a combination of all tuning parameters mentioned in the above literature, also with a different set of features from that in the literature. Comparable accuracy was obtained for relatively higher forecast period (7-10 days).

With short term energy forecasts, grid operators can move towards a data-driven Demand Response schemes to manage supply-demand of energy as well as manage distributed energy resources like Solar, Wind, Electric Vehicles, etc.

3.0 METHODOLOGY

When modelling the energy output of a solar generator, the meteorological data of the site is of primary importance. Shaping, formatting and transformation of that data to fit the requirements is one of the vital components of this thesis. Obtaining degradation data for the modules to obtain probabilistic states of existence was done by the periodic performance data collected by PRL at various PV power plants. The flowchart given below gives an overview of the steps required to be taken to forecast Loss of Load (LOL)

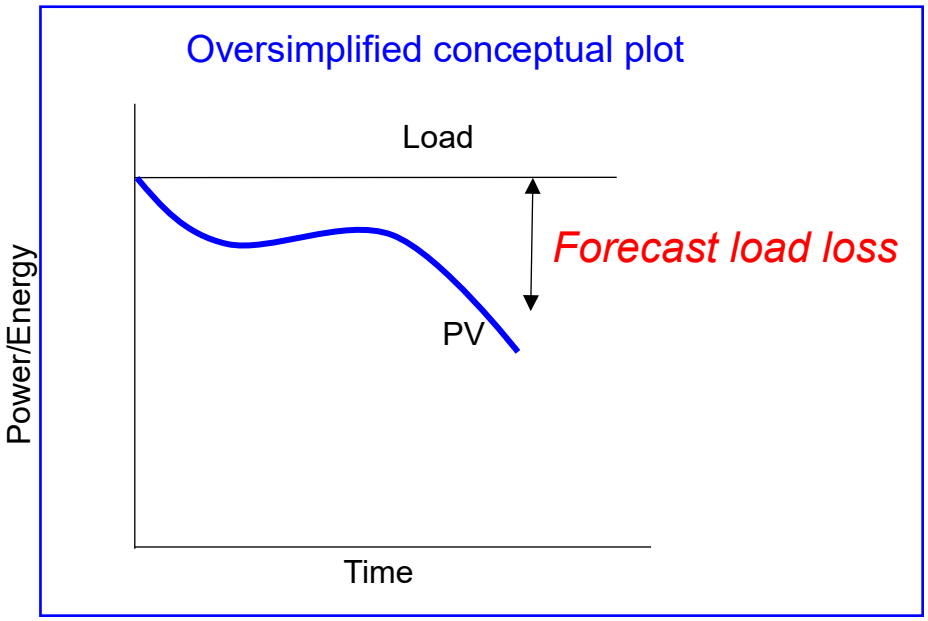

Figure 1-Loss of load simplified plot

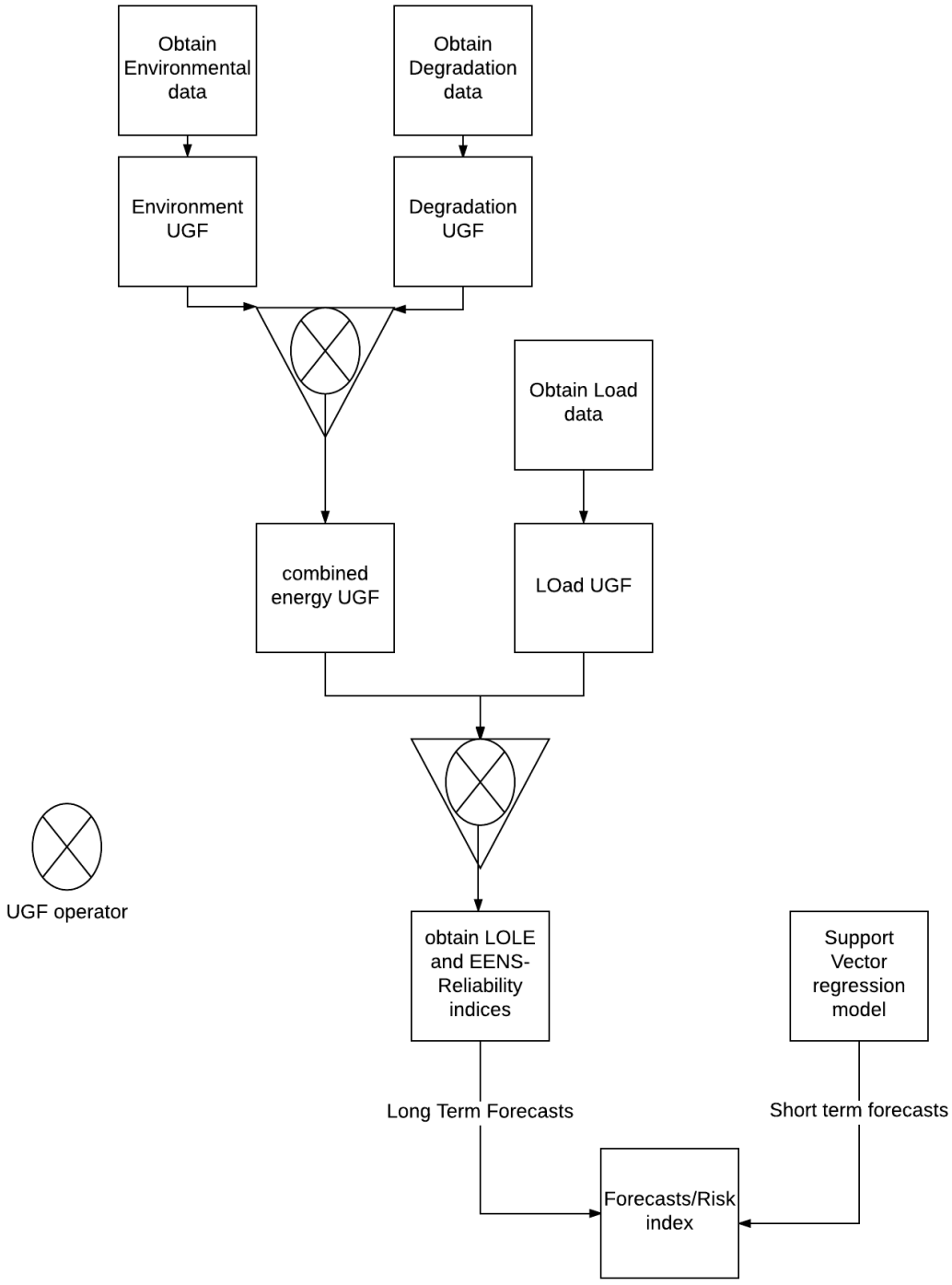

Figure 2- Flow chart representing the logical flow of the program developed to calculate the reliability indices of a given Solar PV system

3.1 Environmental Model

Input for the developed forecasting tool is the environmental model. It can be defined as an appropriate way to represent the field conditions of the site in which the PV system is located, over the long term. The subject of interest was to see how fluctuations in weather affect the energy production of the modules, using which probabilistic states of energy were developed. Five years of data from Solar Anywhere database for the Tempe area was used to build the environmental model. The input variables to build the model included Direct Normal Irradiance (DNI) and Ambient temperature. The frequency of the observations were 10 minutes i.e. the data was captured every 10 minutes. This gave a very high resolution of fluctuations in weather. Once the data was shaped and formatted, which included removing missing values it was compiled and imported into the R programming environment for further analysis. Two types of weather models were considered for analysis- Discrete model and continuous model. Discrete model of weather involved dividing different input variables into different ranges called bins and observing the occurrence of those discrete bins. Continuous model involved observing the weather variables as a continuous function and then observing the fluctuations of weather over time.

3.1.1 Discrete Environmental Model

Discretization process involves converting a continuous variable like Temperature and Irradiance into a non-continuous format by dividing the variables into different range bins and observing which bin does each observation fall into. To create a discrete

model, the weather data was imported into the R programming environment, since it's convenient for statistical analysis. Primary interest in analyzing weather data was to observe the fluctuations in energy production by each individual solar module. The energy produced by each module is a function of both irradiance and ambient temperature. Ambient temperature has an indirect effect on energy produced by the module, it has a direct impact on different variables like Isc, Voc, FF which in turn affect the energy produced by each solar module.

$$P(S_i, n_{sm}) = n_{sm} \cdot FF \cdot V_y \cdot I_y$$

$$I_y = S_i \cdot [I_{SC} + \alpha(T_c - 25)]$$

$$V_y = V_{OC} - \beta \cdot T_c$$

$$T_c = S_i e^{(-3.56-0.075xW.S)} + T$$

$$FF = \frac{V_{MP} \cdot I_{MP}}{V_{OC} \cdot I_{SC}}$$

Where:

$P(S_i, n_{sm})$ is the output power of the solar generator at Solar irradiance Bin S_i with n_{sm} functioning PV modules

α and β are the current temperature coefficient (A/ºC) and voltage temperature coefficients (V/ºC) respectively

T_c is the module temperature

T is the ambient temperature

FF is the fill factor.

$P(S_i, n_{sm})$ shown above is the total power produced by the solar generator. Total power produced is a function of fluctuating environment variables (Ambient

temperature & Solar irradiance) and fluctuating functioning solar modules (varying Degradation rates of modules). Environmental variables are modelled using the discrete environmental model described below. The fluctuating functioning solar modules are modelled using degradation modeling described in the subsequent section. The ambient temperatures were converted to module temperatures using the Sandia King model after considering other temperature models [2]. Once the data was in the desirable format with Irradiance and Module Temperature being consistent without any missing values, the data was used for further analysis. The discretizing scheme used here was used by JPL et. al to study the impact of environmental stresses on Solar Module degradation [5]. The approach involved discretizing Irradiance and Module Temperature into different bins and considering the combined effect of both the variables in the form of a bin. For example, a bin code of 12 represents irradiance state as bin 1 (range= 0-300 W/sq. m) and a module temperature bin of 2 (range= 35-50^0 C). The figure given below shows the environmental bin model visually. All three variables considered by JPL et al are represented by their discrete bins in the form of a 3-Dimensional cube, where each smaller cubes referring to simultaneous values of Temperature, Humidity and Irradiance. Such a discretizing process considers the simultaneous occurrence of environmental variables which affect the energy generation of solar modules

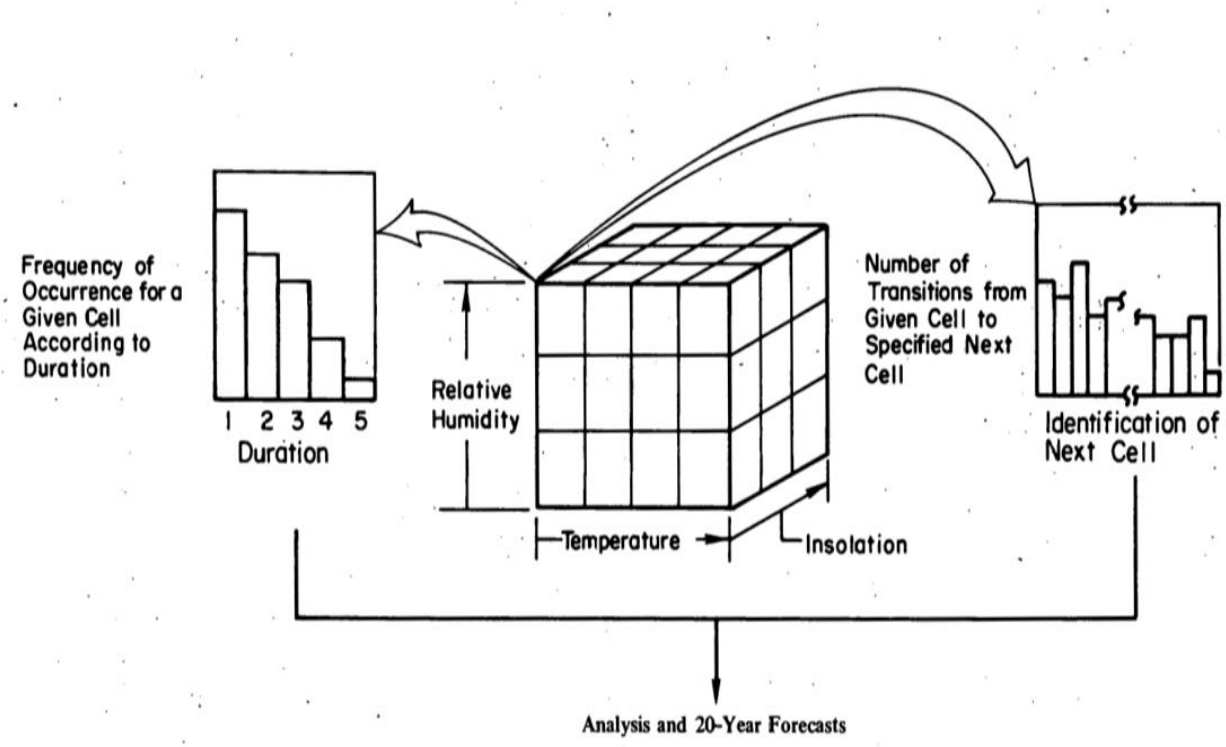

Figure 3- Environmental bin representation. The different axes show different variables and each smaller cube represents one environmental bin [5]

Once the discretizing process was completed and for each observation (row) a bin was assigned. Then a simple counting algorithm was used to generate the frequencies of each bin and probabilities were calculated by simply dividing the frequencies by the total count. These probabilities were stored for its subsequent use in the Universal Generating Function which will be explained in the coming sections.

Bins	Frequencies
11	138838
12	9114
13	1
21	7774
22	6004
23	337
31	15288
32	32308
33	15970
34	11

Table 1- Table showing the number of times each bin has occurred in the past

Bin statistics can be used to achieve more insight into which bins are occurring more than the others. To achieve higher resolution into the bins of interest, further discretization of those specific ranges can be done. For example, if the bin 11, bin 32 and bin 12 have high counts of frequency, those bins can be analyzed more closely. On doing that its analyzed that bin 11 is the low irradiance low module temperature bin, which likely doesn't effect energy production to a high extent. But bin 32 is the high irradiance bin which needs some more resolution. Adding another discrete range can help get more detail. The irradiance bin 3 (600-1000 W/sq. m) can be divided into 2 bins of ranges 600-800 W/sq.m and 800-1000 W/sq. m. Doing so increases the number of computations, but gives more accurate results. This discretization process to achieve a greater level of detailed bins can be done continually so, but it comes at a cost which in this case is computational power. The higher the number of bins used, the computational complexity increases exponentially in the UGF structure. More on UGF will be explained in the subsequent sections. The above-mentioned procedure can help us convert continuous random variables (Irradiance, Temperature) into discrete

random variables. But for specific calculations, each bin needs to represent a value. Multiple methods were used to select the bin value. A simple average of the upper limit and lower limit of the bin can give us a value for each bin. But this is least likely to be accurate, because the univariate distributions of irradiance and temperature don't follow the uniform distributions. They tend to be more normal in nature, with Means of those distributions being the most accurate measure of its central tendency. Medians also tend to be a good measure of central tendency in the case of data containing lot of outliers, but since the data was cleaned and discretized, Mean seems to be a better a measure to use. Also, the outliers tend to be more concentrated in the lower range values of Irradiance (Due to Pyranometers malfunctions) [26]**.**

After using filters multiple times, the means of each bin are recorded. The advantage of this method is, the bin means are taken from many years of detailed historical data. This will help us make accurate forecasts of bins in the future which are most likely to follow a similar pattern in the future as they did in the past.

3.1.2 Continuous Environmental Model

Although the discrete modelling is very convenient and easily interpretable there is likely to be some loss of details associated with discretization process. Continuous models use the data to fit a continuous, smooth model. Such a model might be beneficial for studying the variations in the environmental conditions. Pan et al had developed a time-series model for temperature using the historical dataset for Mesa, AZ. Time series models are usually preferred for modelling variables which vary over time.

3.1.2.1 Time Series Models

ARIMA models and Holt Winter's smoothing methods are dominantly used in research and industry to explain variations in variables with respect to time [28][12][7].

Our interest here was to model the variables of our choice- Irradiance and Module Temperature. The approach was to see the comparison between individual Univariate Time series models for each variable i.e. study the variation of each of our variables with respect to time taken one at a time. When two different ARIMA models for each of the variables were obtained, the models can be combined to see how environment has behaved in the past. A seasonal ARIMA model approach is used to model both irradiance and temperature. Such a model works well on data which has seasonality, daily fluctuations and has an yearly pattern.

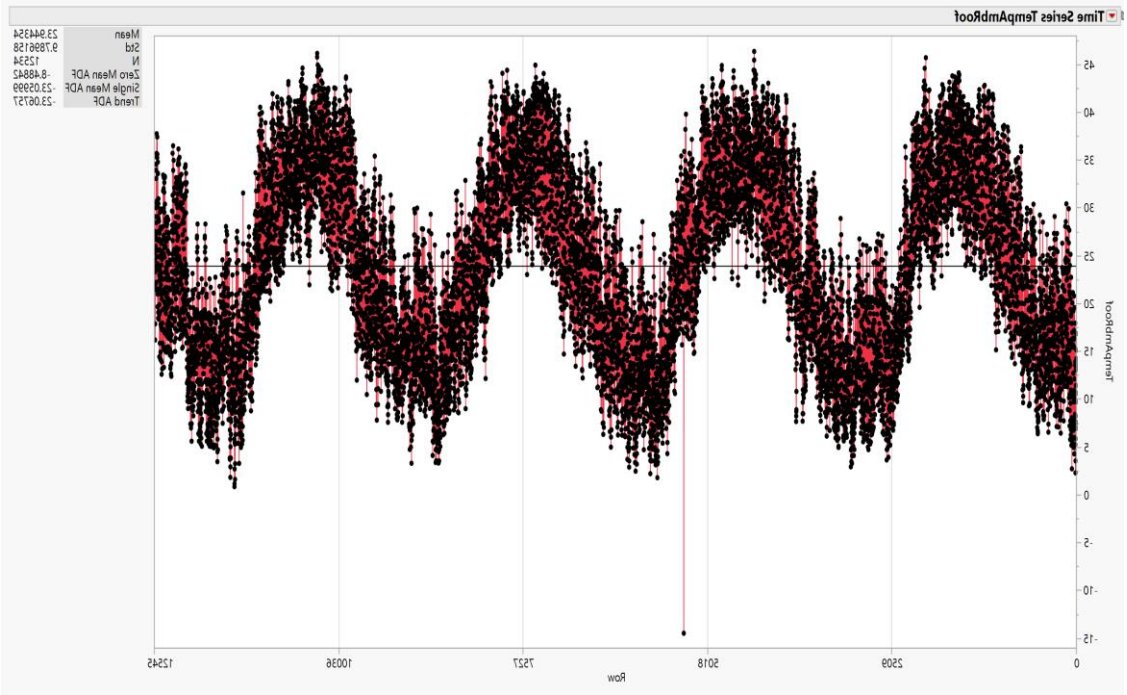

Figure 4-Time series plot of temperature

Model: Seasonal ARIMA(1, 1, 1)(1, 1, 1)8

Model Summary

DF	12498	Stable	Yes
Sum of Squared Errors	38228.5191	Invertible	Yes
Variance Estimate	3.05877093		
Standard Deviation	1.74893423		
Akaike's 'A' Information Criterion	49475.0292		
Schwarz's Bayesian Criterion	49512.1978		
RSquare	0.96811148		
RSquare Adj	0.9681013		
MAPE	6.62249858		
MAE	1.24285733		
-2LogLikelihood	49465.0292		

Failed: Cannot Decrease Objective Function Hessian is not positive definite.

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	-0.9623184	0.0115855	-83.06	<.0001*
AR2,8	2	8	0.1990141	0.0119026	16.72	<.0001*
MA1,1	1	1	-0.9779850	0.0091261	-107.2	<.0001*
MA2,8	2	8	0.8761629	0.0068786	127.38	<.0001*
Intercept	1	0	-0.0000974	.	.	.

Constant Estimate	Mu
-0.000153	-9.7365e-5

Forecast

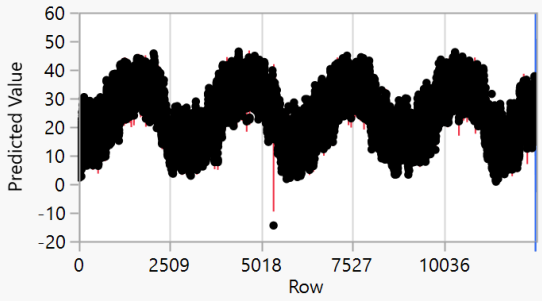

Figure 5- ARIMA model of temperature

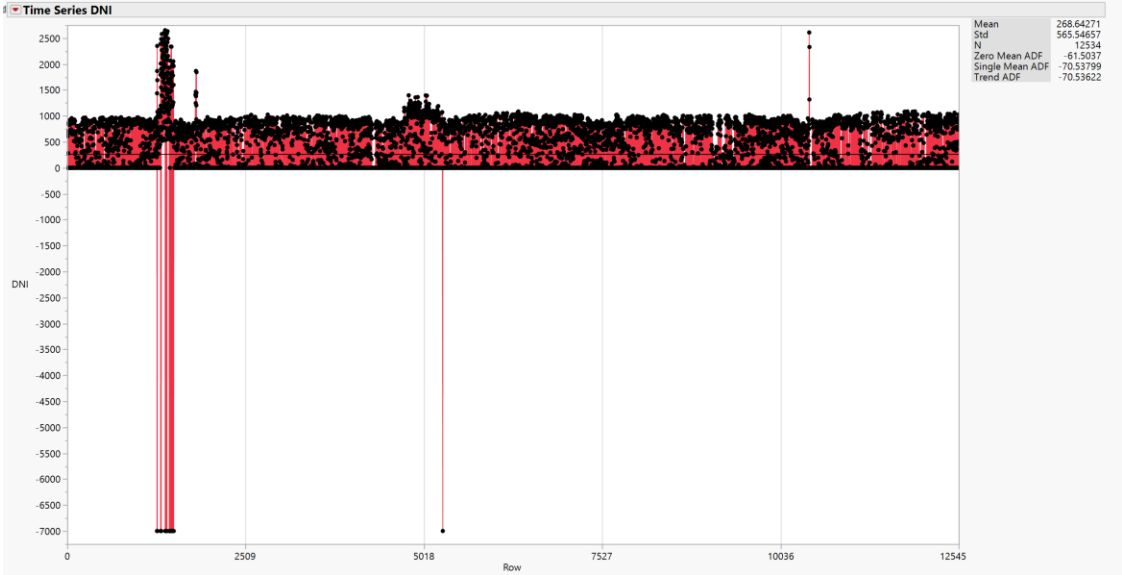

Figure 6-Time series of Irradiance

Model: Seasonal ARIMA(1, 1, 1)(1, 1, 1)8

Model Summary

DF	12498	Stable	Yes
Sum of Squared Errors	2444856465	Invertible	Yes
Variance Estimate	195619.816		
Standard Deviation	442.28929		
Akaike's 'A' Information Criterion	187839.219		
Schwarz's Bayesian Criterion	187876.387		
RSquare	0.38981901		
RSquare Adj	0.38962406		
MAPE	.		
MAE	172.750723		
-2LogLikelihood	187829.219		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant	Mu
AR1,1	1	1	0.2839486	0.0094805	29.95	<.0001*	Estimate	-0.0426117
AR2,8	2	8	0.0875642	0.0103338	8.47	<.0001*	-0.0278404	
MA1,1	1	1	0.9195696	0.0034740	264.70	<.0001*		
MA2,8	2	8	0.9236053	0.0047389	194.90	<.0001*		
Intercept	1	0	-0.0426117	0.0856043	-0.50	0.6187		

Forecast

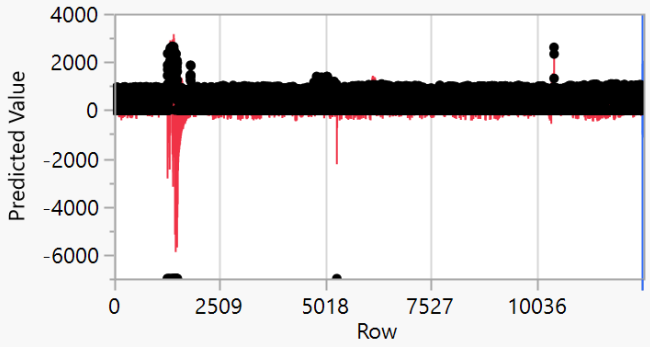

Figure 7-Statistics of ARIMA model for irradiance

As noted in the summary statistics for the temperature model, a seasonal ARIMA (1,1,1)(1,1,1) 8 has been used. It means the temperature model can be split into two parts- one is the variation of temperature with time and another is the variation of temperature with cyclic variation (diurnal variations) and seasonal variations. Appropriate differencing order, Autoregressive order and Moving average order is chosen by the JMP program by conducting successive iterations of model fitting and choosing the model based on best values of fit (R-Square), MAPE (Mean Absolute

error), RMSE (Root mean squared error), information criterion (AIC and BIC). The model achieves a good fit of R-square 97%

Similarly, a seasonal ARIMA model for irradiance is built using data for 5 years. The picture above shows that the model fits poorly. It has an R-square of 39% and almost every other measure is in poor shape. On closer observation of residuals, the model has high residuals during the night hours when irradiance drops to 0. This seemed to distort the model towards a lot of noise. Removing all the night time observations and running a similar model on data from 8am to 7pm seems to improve the accuracy a little. It improves the fit to 77% and all the other accuracy measures improve, but it's still not a good model.

A multi-variate time series of Irradiance and Temperature was considered, but preliminary analysis revealed very weak cross-correlation between Irradiance and Temperature. Such a weak correlation between the two variables might not improve the fit of the model.

On further investigation, it seems logical to conclude that irradiance models are usually non-linear in nature [29]**.** Irradiance seems to vary in relation to many other parameters in the weather data, like temperature, cloud cover, chance of precipitation, wind speed etc. There is a need to find a model which can model irradiance based on all the above presented variables and effectively capture all the non-linear relationships between the above given variables.

3.1.2.2 Support Vector Machines

Machine learning methods were explored to develop a model to predict Solar Irradiance. Sharma et al used a support vector regression method to model the Solar irradiance [26]. Yerrapragada et al used a variation of support vector machines called Least Squares Support vector regression to achieve accurate models to predict solar irradiance [27]

A support vector regression model was built to predict Solar Irradiance. Data was transformed to a format where the output variable was the target hour of solar irradiance, and the independent variables were the solar irradiance of the same target hour but with a lag of 7,8,9,10,11 days. These variables made sure the time-series dependence of Irradiance was maintained. Along with these, temperature, wind speed of the target hour was added too. This technique is significantly different from the techniques presented in the literature. It doesn't use any other parameter like cloud cover, precipitation data or any other meteorological inputs from satellite data and yet produces a reliable model.

A regular Support Vector Regression model using a RBF kernel is used with a complexity parameter of 1000 and a gamma value of 0.00001. When tested on a test set of a week ahead values, the predictions produced were highly accurate with a MAE (Mean absolute Error) of 93 W/sq. m. So, this proved that the model fits well on both seen and unseen data.

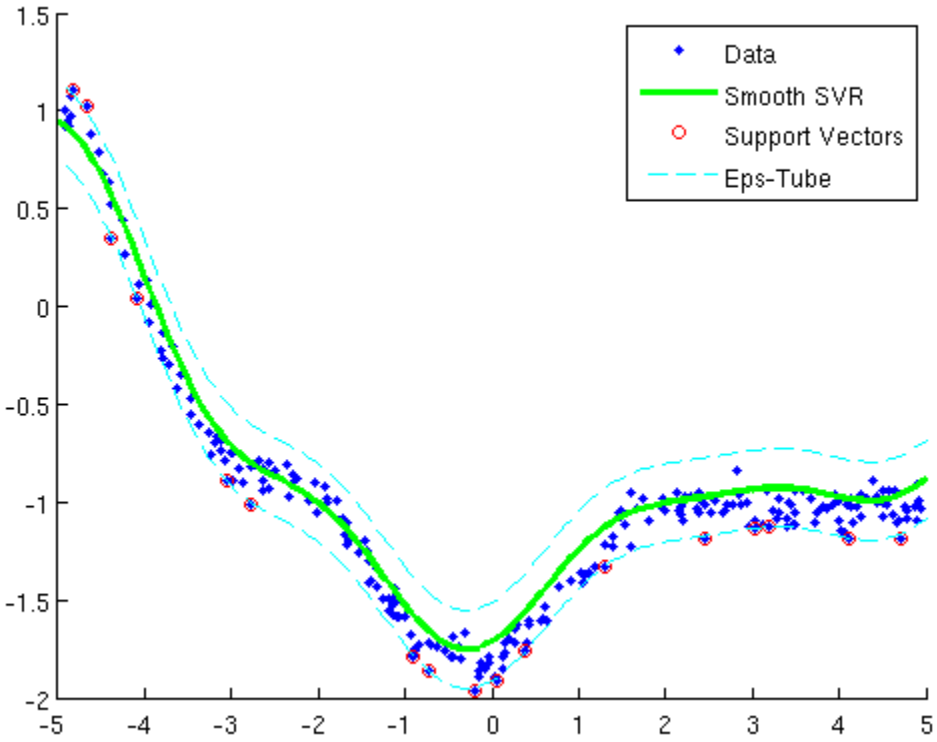

Figure 8- Representation of Support Vector Regression. The legend shows different decision boundaries for different types of SVM models and tuning parameters.

Continuous models provide a greater fit to the data than the discrete model, however generating long term forecasts of PV Loss of Load and Energy Supply was of greater interest. It was learnt that continuous models are useful in providing highly accurate short term forecasts whereas discrete models perform well for long term forecasts. For the UGF approach discrete model was chosen, whereas for the short-term forecast generation continuous models were used.

Keeping in mind this distinction it was decided to produce two types of forecasts-

1. Long term forecasts of Loss of Load which would help give an insight into Grid Stability, Contribution of Distributed sources to the grid.

2. Short term forecasts to help Utilities gain unique insights into customers owning PV systems on residential and commercial buildings. These insights include- better targeting customers who have PV systems in Load Shedding programs during peak demand times, potentially reducing costs for the utilities.

3.2 Degradation Models

Any long-term PV Energy forecasts must consider the effect of degradation of modules over time. Several degradation studies have shown the linear nature of degradation of rated Power of Crystalline silicon modules [7][10]. Multiple failure modes have been identified and NREL has identified 87 types of defects which affect the safety and performance of a PV module [8] System modeling software packages like SAM and PVSyst do consider the effect of degradation over the long-term energy forecast of the system. But due to lack of degradation data they assume a mean degradation rate of 0.8% with a user input of a RMS value. Jordan et al have reviewed the mean degradation rates for PV modules by different regions and climates around the world [8]. They conclude the median degradation rates have been consistently around 0.5% with the mean being around 0.8%.

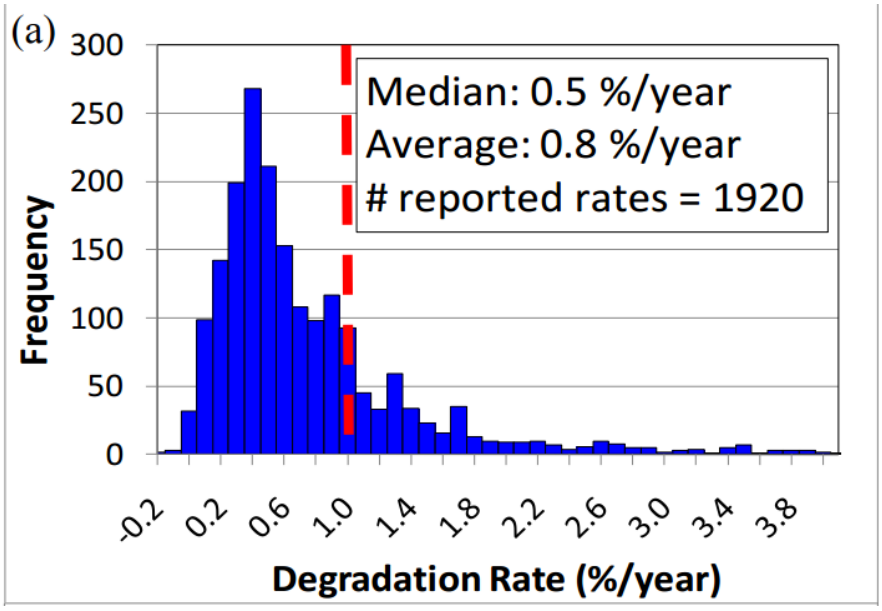

Figure 9- Distribution of Degradation [8]

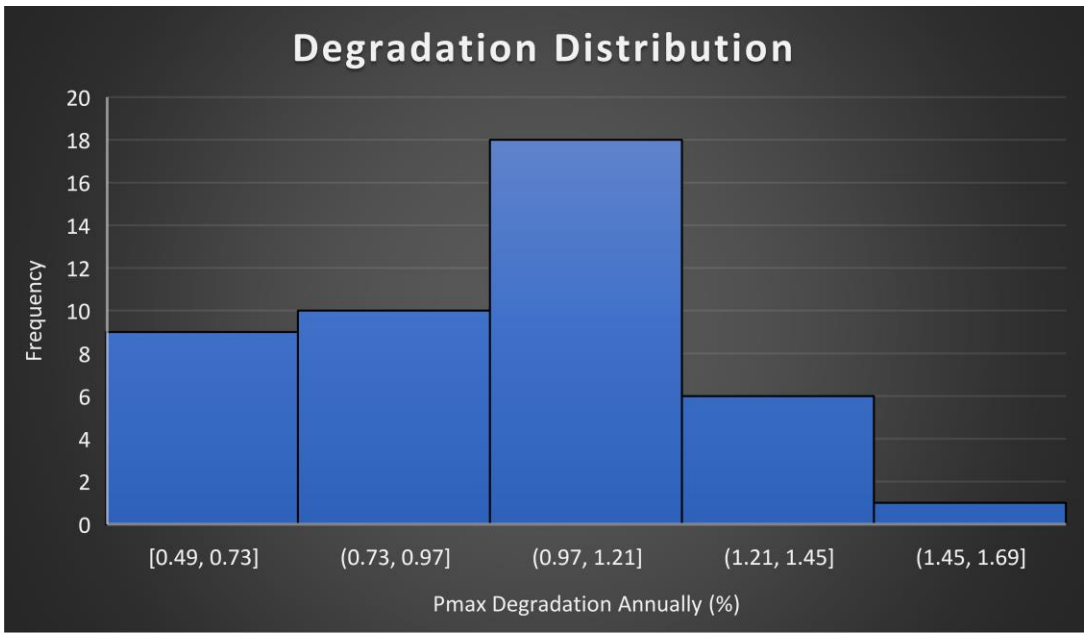

Figure 10-Tempe warehouse degradation distribution

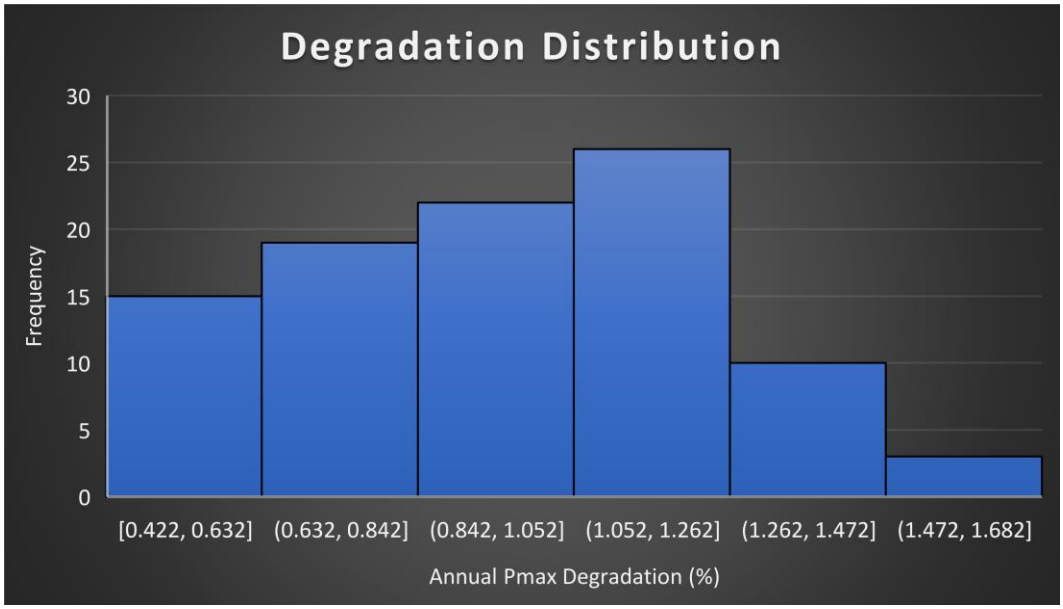

Figure 11- Rogers1 degradation distribution

For this thesis, PV power plant degradation was our main concern. When analyzing individual module, I-V data done over the years for the Tempe Warehouse and Rogers-1 power plants, the distribution of degradation rates for different modules confirm Jordan et al findings about the nature of the distribution, though higher degradation rates were observed for the power plant under study. The distribution was normal, but skewed to the left. For such distributions median is a better representation of central tendency than the mean. The median degradation was found to be 0.99 % /year. So, the failure rate for each module was defined as the rate at which the modules reach a degradation of 20%. This failure rate was found to be 0.050 / year. This is important, because the failure rates are used in the Markov Process calculations to achieve steady state probabilities of different working-not working modules. More on markov processes will be explained in the forthcoming section.

3.3 Markov Processes

A markov process is a stochastic process which is used to represent the chance that a random variable is going to take a particular value. Markov processes are based on a central assumption that, the next value of a process (random variable) depends on the current value only and is independent of the previous states' values. The random variable is represented by multiple values/states, the transition between states being represented by failure rates and repair rates.

Markov processes are a very popular method of modelling stochastic processes. A process is called stochastic when its variations are random in nature. For example, random variables like number of machines breaking down can be represented as markov chains. At any given point, to find out how many machines are broken down needs us to evaluate multiple failure rates and repair rates of many machines together. Failure rates and repair rates of a machine are competing processes which decide if at a given time a machine has failed or is still working. The problem becomes more complex when there are multiple machines as described above which have competing failure and repair rates. It cannot be stated certainly that at a given time how many machines would be working or not working. But using markov processes one can find out the probability of being in every state. This is called the steady state probability of markov process.

From the context of this thesis, markov processes was used to model the random event of multiple modules degrading in a power plant, leading to multiple competing processes. At a given time it's difficult to predict the proportion of modules performing

at different degradation levels, since each module has its own degradation rate as a result of difference in manufacturing quality and various other factors. So one might wonder as to how to find the proportion of modules working at different levels of degradation. Markov processes offers an effective way to model this phenomenon. The mean failure rate, mean repair rate (assuming if the modules are replaced after they degrade 20%) is found out and the expected probabilities of being in each state can be calculated.

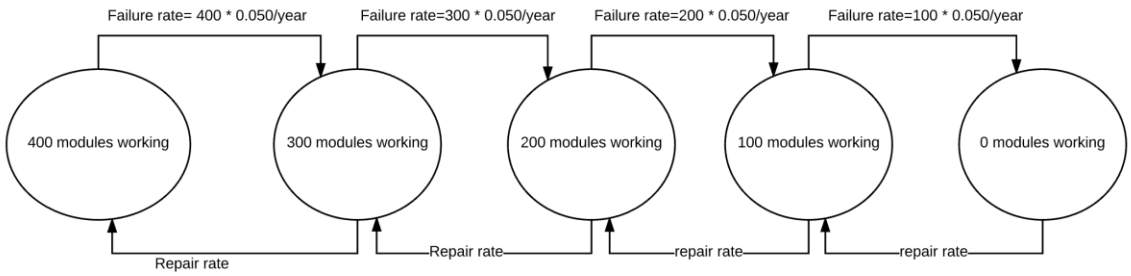

Figure 12- Markov Model showing transition between different functioning states. The failure rates and the repair rates represent the transition between different states.

The figure above shows us the different markov chain. If our system has 400 modules, it can be discretized into different states. It should be noted that, a module is represented as working (100% Pmax) and not working (Pmax<80% of original). For a system of 400 modules, it can be discretized into 5 states as represented in the figure. For a state representing 300 modules working, means that 300 modules are working at 100% Pmax, and 100 are working at 80% Pmax degradation. This can be represented mathematically by a transition diagram as shown below. Assuming the states are 1, 2, 3….., r; the transition matrix is given below.

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{bmatrix}.$$

Figure 13- Transition matrix representing transition probabilities/failure rates of a markov process. P_{ij} represents probability/rate of transition from state i to j

Through recursive matrix multiplication, the steady state probabilities of each state can be found out. This tells us, at a given time when looked at the power plant, the chance of finding the plant in the respective states is given by the steady state probability calculated. It's evident by looking at the explanation that the more states the process is divided into, more detailed representation of the working of the power plant can be obtained.

3.4 Universal Generating Functions

Ushakov et al demonstrated the use of universal generating functions to evaluate joint probability distribution of multiple discrete random variables. Universal Generating Functions (UGF) convert a probability distribution into a discrete polynomial function. Such a representation is very convenient to carry out mathematical operations like multiplication, addition and subtraction. The general format of a UGF is given below.

$$u(IR) = \sum(p_i^{IR} * Z^{y(i)})$$

Here 'p_i^{IR}' represents the probability of the discrete variable i, value of which is given by $y(i)$. More discrete states result in a long UGF.

3.4.1 UGF of Solar Energy States

As per the discrete environmental model explained before, the output obtained there is the different environmental bins and the respective frequencies and probabilities of the bins. Using the following information, energy produced by each bin is given by a energy function.

Energy produced by a solar module is a function of environmental conditions. It directly depends on Solar Irradiance and the operating module temperature [2].

$$P(one\ solar\ module) = FF \cdot V_y \cdot I_y$$

$$I_y = S_i \cdot [I_{SC} + \alpha(T_c - 25)]$$

$$V_y = V_{OC} - \beta \cdot T_c$$

$$T_c = S_i e^{(-3.56-0.075xW.S)} + T$$

$$FF = \frac{V_{MP} \cdot I_{MP}}{V_{OC} \cdot I_{SC}}$$

V_y and I_y are calculated for each bin. And then the energy produced by each environmental bin is calculated. So for each bin, the probability associated with it and the energy produced by it is obtained. This can be expressed as a UGF.

$$u_Y(Z) = \sum_{i=0}^{n_E-1} p(Y = i) Z^{h_i}$$

Where:

h_i denotes the power output of a single solar module at state i

$p(\cdot)$ denotes the state probability distribution

This represents the UGF of solar energy states.

3.4.2 UGF of Solar Module Degradation

Energy produced by a solar power plant depends not only on the weather states, but also on the number of functional PV modules. By functional, if a PV module is working at 100% of its rated power, then its fully functional. If it is working at 80% or lower of its rated power then its considered as not working. From the markov process model of degradation, the steady state probability of each degradation state is calculated. That's given below.

States	**Modules effectively working - $g(i)$**	**Probability of the state**
400 modules working	400	p_1^{deg}
300 modules working	380	p_2^{deg}
200 modules working	360	p_3^{deg}
100 modules working	340	p_4^{deg}
0 modules working	320	p_5^{deg}

Table 2- Steady state probabilities of number of functioning modules. A module is considered functioning if degradation is 0%. A module is considered failed if degradation is greater than 20%.

the UGF of Degradation states can be written as given below

$$u_X(Z) = \sum_{i=0}^{n_M-1} p(X = i)Z^{g_i}$$

Where:

g_i denotes the number of working modules at state i

$p(\cdot)$ denotes the state probability distribution

3.4.3 Combined Energy UGF

When the combined effect of modules degrading over time and fluctuations in weather is analyzed, the actual Energy output expected from the PV system can be calculated. Using the UGF approach, every fluctuation is reduced to probabilities or chance. When the UGFs are combined, its equivalent to taking the joint probability distribution of multiple random variables. Essentially, it's like calculating the chance that all the fluctuations (fluctuation in weather & degradation) occur simultaneously and estimating its impact on energy generation.

System modelling software like PVSyst perform this operation, by reducing weather fluctuation to a single TMY file which is generated using iterative sub sampling techniques. Even the degradation states are randomly subsampled and then combined in a Monte-Carlo simulation to see the combined effect. UGF provides us a method which is computationally less expensive and doesn't lose a lot of detail as in the case of TMY compression.

Combined representation of Solar Energy States and Module Degradation States is given below.

Assuming independence between X and Y, the u-function for the solar generator power output can be written as:

$$u_S(Z) = u_X(Z) \otimes_{\times} u_Y(Z)$$

The $\otimes$ symbol is like the multiplication operator. It is used here in context of multiplying two UGFs. Multiplying the two UGFs must be done in a fashion as shown below.

$$u_S(Z) = \sum_{i=0}^{n_M-1} \sum_{j=0}^{n_E-1} p(X=i)p(Y=j)Z^{g_i \times h_j}$$

3.4.4 Load Data

The load data of the site was obtained and a simple discretization rule was followed. The load data was divided into bins, and the mean of each bin was found from the actual load file. Once frequencies were obtained for each bin, it was converted to probabilities. This was represented in the form of a UGF for the load file.

$$u_{load}(Z) = \sum_{i=0}^{n_{load}-1} p(load=i)Z^{l_i}$$

3.4.5 Loss of Load Estimate (LOLE)

LOLE is a reliability index which is used frequently by utilities to characterize the stability of a power system with respect to meeting energy needs. Its represented in percentages and it quantifies the amount of time the load requirement won't be satisfied by the solar generator. The general procedure to calculate LOLE involves comparing hourly Solar Generator output to that of the hourly load demand. The number of hours

load exceeds the amount of energy produced is divided by the total number of hours to result in a fraction. This fraction gives us the reliability of a Solar generator, signifying how effective is the solar generator in satisfying the load requirement.

$$LOLE = \frac{number\ of\ hours\ (energy\ produced < load\ demand)}{total\ number\ of\ hours\ simulated}$$

For the UGF approach, the combined UGF obtained above is compared to the load UGF using a special operator as defined below.

$$LOLE = \sum_{i=1}^{N_t} Pr\ (G_i < L_i)$$

$$\Psi(pZ^{g-W}) = \begin{cases} p, & if\ g \geq W \\ 0, & if\ g < W \end{cases}$$

Here Gi represents the solar generator output and Li represents the load demand. The distributive operator Ψ is used to carry out this desired operation. In simple terms, what the distributive operator does is, it carries out a simple polynomial multiplication of combined energy UGF and the load UGF and keeps the probabilities of all the terms where G_i (generated energy at the i^{th} hour) is less than the L_i (load requirement at the i^{th} hour). Once all the probabilities are obtained, it can be simply summed up to get the total probability. This total probability represents the LOLE in probability terms. So instead of getting percentages of LOLE, fractional values are obtained. If this probability is multiplied with time (number of hours in a year), our LOLE estimate would then indicate how many hours in a year the system wouldn't satisfy the load demand. This can be represented mathematically as given below.

$$LOLE = N_t \cdot \Psi\left(\sum_{i=0}^{n_G-1}\sum_{j=0}^{n_L-1} p_j^L p_i^G Z^{g_j^L - g_i^G}\right)$$
$$= N_t \cdot \sum_{i=0}^{n_G-1}\sum_{j=0}^{n_L-1} \Psi(p_j^L p_i^G Z^{g_j^L - g_i^G})$$

Here the distributive operator Ψ is being applied on the multiplication of the combined energy UGF and Load UGF, the result being multiplied again by the total time steps N_t.

3.4.6 Expected Energy Not Supplied (EENS)

In the previous subsection, it was demonstrated how to obtain the LOLE which represented the time the load demand won't be met by the solar generator. It would also be of importance to know the combined difference of the load demand and actual energy produced. It would help us estimate by how much does the solar generator fall short of the load requirement. For utilities, such information helps them know their customer's energy needs better, since they would know the quantity by which these solar generators miss load demand. It also helps the system designers to design systems more effectively.

For this study, EENS is calculated from UGFs. The EENS would be calculated by the equation given below.

$EENS = \Sigma(load - energy\ produced)$….[when load > energy produced]

$$EENS = N_t \cdot \sum_{i=0}^{n_G-1}\sum_{j=0}^{n_L-1}(g_j^L - g_i^G)\Psi(p_j^L p_i^G Z^{g_j^L - g_i^G})$$

Here g_j^L-g_i^G represents the difference of the load demand and the energy produced by the solar generator at that time step. Similar to LOLE, a simple polynomial

multiplication is performed between the combined energy UGF and the Load UGF. The distributive operator is used to keep the combined probabilities only whose load demand exceed the energy supplied. In addition to that, the difference by which the load demand is missed is also multiplied with the above probabilities, as shown in the formula above. The result of the summation would give us the amount of energy not supplied per unit time. When this term is multiplied by the total time, it results in the total energy not supplied.

4.0 RESULTS AND DISCUSSIONS

The discrete and continuous environmental models are discussed first. Following which, their outputs are plugged into the Universal Generating Functions (UGF). In the subsequent sections calculations showing combining energy and degradation UGFs to obtain combined energy UGFs and the load UGF is shown and discussed. Calculations of LOLE and EENS as a function of number of chosen states are shown and discussed.

Following which, comparisons with PVSyst outputs for similar systems is shown to get an estimate the performance and accuracy of the model presented.

The system in consideration was Tempe Warehouse. The details of the system is given below in the table.

Date	**Isc (A)**	**Voc (V)**	**Imax (A)**	**Vmax (V)**	**FF (%)**	**Pmax (W)**
23/4/2013	6.50	60.00	5.80	51.00	75.85	300.00

Table 3- Tempe Warehouse module specifications

Some of the key assumptions to be made before analyzing the results are-

1. The degradation data is compiled using the I-V curves taken at the individual module level. Not all the modules are tested, so its assumed that the sample of modules tested represents the entire population of the modules of the power plant.
2. This analysis can be easily translated to string level analysis by taking the raw KWh degradation data.

4.1 Discrete Environmental Model UGF

Initially for the first iteration, the irradiance was divided into 3 ranges which were- 0-300, 300-600, 600-1000 W/m^2. This was done by taking into account the most dense regions of the Irradiance distribution [5].

The module temperatures were divided into 5 ranges- 0-35, 35-50, 50-65, 65-80, 80-100. So overall there are 3x5, 15 bins. But not all the bins have occurred, so the table given below summarizes only the bins that have occurred.

Bin	Probability
11	0.614488
12	0.040338
13	4.43E-06
21	0.034407
22	0.026573
23	0.001492
31	0.067664
32	0.142993
33	0.070682
34	4.87E-05

Table 4-Bins and their probability of occurrence

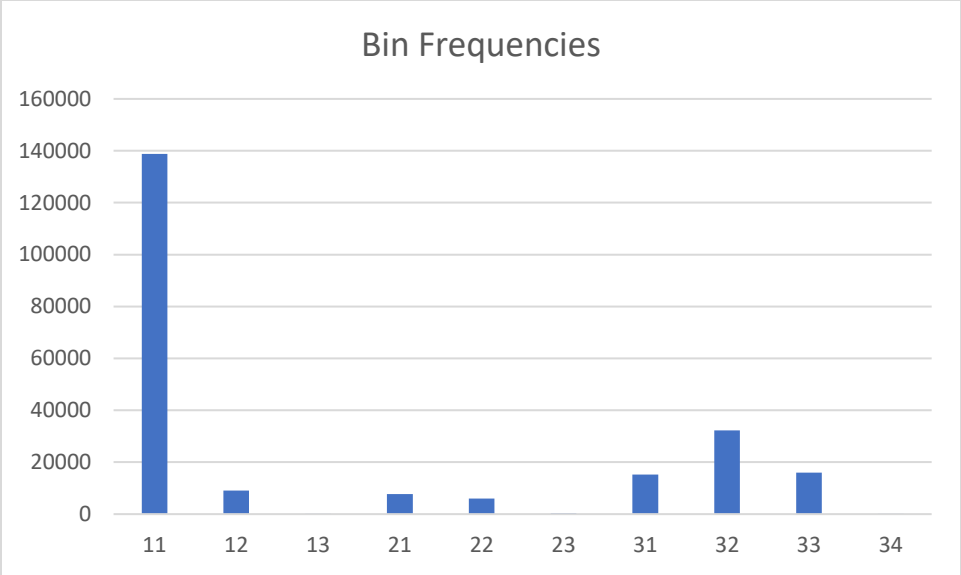

Figure 14-Bar chart of bin frequencies

The range of the bins are given below and the average values of Irradiance and Temperature for each bin is given below in the figure.

Bins	Irradiance Range (W/m^2)	T-cell Range (^{0}C)
11	0-300	0-35
12	0-300	35-50
13	0-300	50-65
21	300-600	0-35
22	300-600	35-50
23	300-600	50-65
31	600-1000	0-35
32	600-1000	35-50
33	600-1000	50-65
34	600-1000	65-80

Table 5- Bin ranges

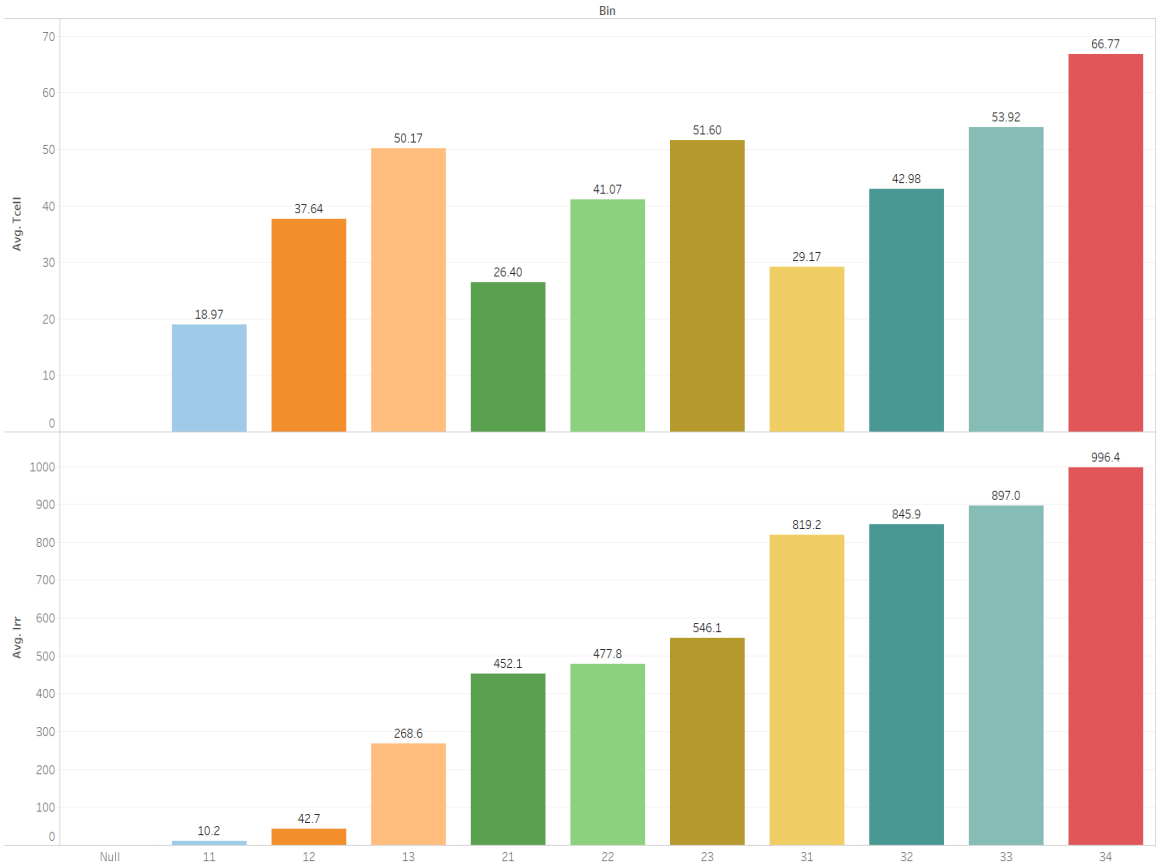

Figure 15-Average Irradiance and Average module temperature for each bin

Since a continuous data format is converted to a discrete one by dividing into bins, one can easily convert this to a UGF format. As mentioned in the previous section, a UGF converts a discrete probability distribution like the ones for the environmental bins, into a polynomial format. For doing so, one would need the energy output of each bin. This can be summarized as in the table below.

Bin	Irradiance mean (W/m^2)	T-cell Mean (^0C)	V_y	I_y	Power (W)	Probability
11	10.16476	18.21282	61.40495	66.10542	3.122337	0.614488
12	42.71431	37.63772	57.38399	277.3731	12.24318	0.040338
13	268.6	50.17182	54.78943	1742.519	73.4368	4.43E-06
21	452.09789	21.00461	60.82704	2939.539	137.5357	0.034407
22	477.76128	41.07223	56.67305	3101.609	135.2082	0.026573
23	546.10208	51.60014	54.49377	3542.4	148.4854	0.001492
31	819.19571	10.44618	63.01264	5330.733	258.377	0.067664
32	845.92158	42.978	56.27855	5490.886	237.6975	0.142993
33	896.97351	53.91979	54.0136	5817.358	241.6953	0.070682
34	996.36364	66.76615	51.35441	6455.557	255.0062	4.87E-05

Table 6- State Probabilities and Energy values

$u_X(Z) = \sum_{i=0}^{n_M-1} p(X = i)Z^{g_i}$

$u_X(Z) = [\ 0.6144Z^{3.12W} + 0.040Z^{12.24W} + \ 0.00000443Z^{73.43W} + 0.034Z^{137.53W} + 0.0265Z^{135.2082W} + 0.00149Z^{148.48W} + 0.067Z^{258.37W} + 0.14Z^{237.69W} + \ 0.0707Z^{241.69W} + \ 0.000048Z^{255W}]$

Above given UGF, $u_X(Z)$ is described for a single Solar Module. It is represented as a 10 state UGF for a single solar module in a solar generator. But the energy produced by the entire solar power plant is the product of the energy produced by the module and the number of modules present in the power plant.

4.2 UGF of Degradation States

The UGF of degradation states gives us the variation in the number of modules at a power plant. For example, if out of 400 modules 10 modules are working at 80% Pmax rating and rest are at 100% , then the effective number of modules of the power plant are 390 + 0.8*10 which is 398 effectively working modules. Markov chains are used to find out the different probability states involving modules working or not working as explained in previous section.

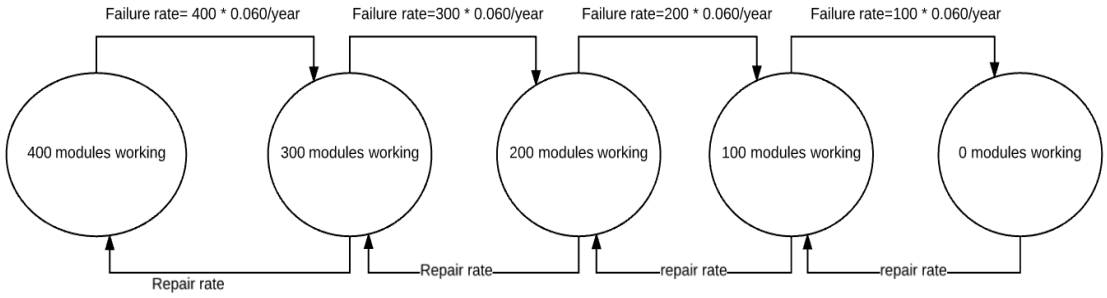

Figure 16- Markov model for Tempe Warehouse

An excel tool was built to calculate the steady state probability. The transition matrix, steady state matrix and the steady state probabilities are shown in the image below.

Five-state Markov Chain (ergodic case)

Λ						check sum
	-23.6	23.6	0	0	0	0
	0.0003699	-17.70037	17.7	0	0	0
	0	0.0003699	-11.8004	11.8	0	0
	0	0	0.00037	-5.90037	5.9	0
	0.5	0	0	0.00037	-0.50037	0

Steady state Λ^T

-23.6	0.0004	0	0	0.5
23.6	-17.7	0.0004	0	0
0	17.7	-11.8	0.0004	0
0	0	11.8	-5.9	0.0004
0	0	0	5.9	-0.5

Failure rate/yr	repair rate/yr
=0.050	0.00037

hours in a yr			
8760	0.0288	523.85	

last row replaced

					right hand side
-23.6	0.0004	0	0	0.5	0
23.6	-17.7	0.0004	0	0	0
0	17.7	-11.8	0.0004	0	0
0	0	11.8	-5.9	0.0004	0
1	1	1	1	1	1

Solution:

0.018
0.024
0.036
0.0721
0.8499

Figure 17- Markov States calculator tool. On entering the failure rate and repair rate, we obtain the steady state probabilities associated with each state.

Steady state probability	Modules working	Effective number of Modules
0.018	400	400
0.024	300	380
0.036	200	360
0.0721	100	340
0.8499	0	320

Table 7- Markovian steady states

The UGF of the following degradation model can be written below

$$u_Y(Z) = \sum_{i=0}^{n_E-1} p(Y=i)Z^{h_i}$$

h_i denotes the number of working modules at state i

$p(\cdot)$ denotes the state probability distribution

$$u_Y(Z) = [\,0.018Z^{400} + 0.024Z^{380} + 0.036Z^{360} + 0.0721Z^{340} + 0.8499Z^{320}]$$

This UGF describes the state-space model of the degrading modules. The degradation UGF and the UGF for a single Solar Module is combined to see the combined effect of both the varying parameters.

4.3 UGF of the Combined Energy Model

Both the UGFs obtained above are combined using the $\otimes$ operator. This operator uses polynomial multiplication to achieve the combinatorial representation.

Assuming independence between X and Y, the u-function for the solar generator power output can be written as:

$$u_S(Z) = u_X(Z) \otimes_{\times} u_Y(Z)$$

$$u_S(Z) = \sum_{i=0}^{n_M-1} \sum_{j=0}^{n_E-1} p(X=i)p(Y=j)Z^{g_i \times h_j}$$

$$u_S(Z) = [0.6144Z^{3.12W} + 0.040Z^{12.24W} + 0.00000443Z^{73.43W} + 0.034Z^{137.53W} + 0.0265Z^{135.2082W} + 0.00149Z^{148.48W} + 0.067Z^{258.37W} + 0.14Z^{237.69W} + 0.0707Z^{241.69W} + 0.000048Z^{255W}] \otimes_{\times} [0.018Z^{400} + 0.024Z^{380} + 0.036Z^{360} + 0.0721Z^{340} + 0.8499Z^{320}]$$

$$u_S(Z) = [0.0110Z^{1.2480KW} + 0.01477Z^{1.186KW} + \cdots \ldots \ldots \ldots \ldots \ldots \ldots \ldots + 0.0000415Z^{81.65KW}] \{50 \text{ terms}\}$$

4.4 Load UGF

The load file is simulated to approximately mimic the actual load on the system. The load variation is shown in the image below. Its average load is 85KW. The file simulated is for a year.

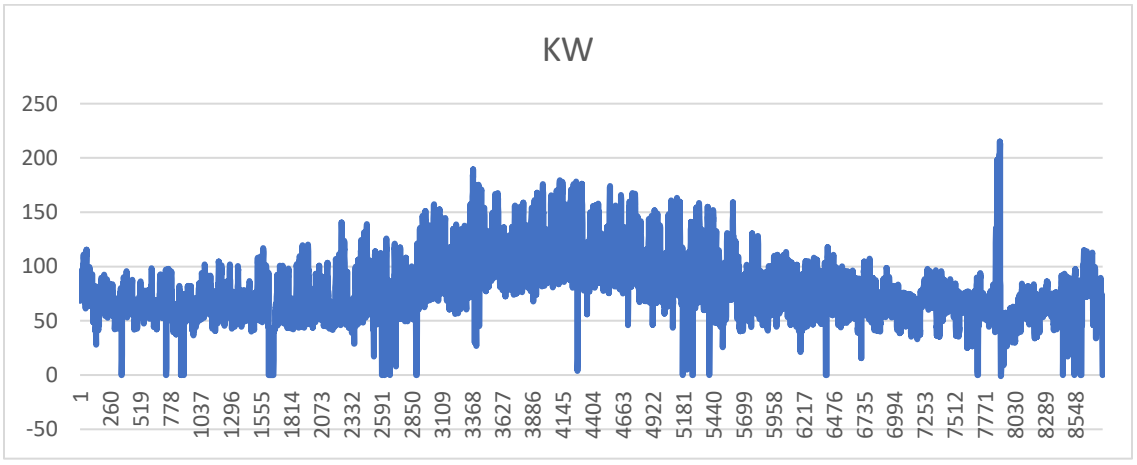

Figure 18- Time series of the simulated load file

The load file is partitioned in different bins which is shown below in the table.

Bin Average (KW)	Frequency	Probability
30	1631	0.185679
80	5159	0.587318
150	1989	0.226434
225	4	0.000455

Table 8- Load file UGF

The UGF for the load file can be written as

$$u_{load}(Z) = 0.1856Z^{30KW} + 0.5873Z^{80KW} + 0.2264Z^{150KW} + 0.000455Z^{225KW}$$

4.5 LOLE and EENS

The LOLE and EENS are calculated by comparing the combined energy UGF and the load UGF. The equations used to obtain the indices was described in the previous section.

The LOLE is obtained by using the Ψ operator which returns the product of the probabilities of the Combined energy UGF and the load UGF only when the load term exceeds the energy UGF term.

$$\Psi\{u_S(Z) \otimes u_{load}(Z)\} = \Psi\{[\ 0.0110Z^{1.2480KW} + 0.01477Z^{1.186KW} +$$

$$\cdots \ldots \ldots \ldots \ldots \ldots \ldots \ldots + 0.0000415Z^{81.65KW}] \otimes [\ 0.1856Z^{30KW} + 0.5873Z^{80KW} +$$

$$0.2264Z^{150KW} + 0.000455Z^{225KW}]\}$$

$$\Psi\{u_S(Z) \otimes u_{load}(Z)\} = 0.8381$$

LOLE= 8784 x 0.8381= 7361.8704 hours/year

As shown above, the Loss of Load Estimate for this system for the given load is 7361.8704 hours per year for its whole lifetime. This means that 83.81% of the time in a year the system doesn't satisfy the load demand.

EENS is obtained by multiplying the difference in load demand states and the energy produced states. It can be represented as the formula given below.

EENS= $8736 * [(g_{i,load-}g_{j,energy}) * \Psi\{u_S(Z) \otimes u_{load}(Z)\}]$ =557.370 MWh/ year

This represents the sum of energy deficit of the system, i.e. the average sum of all the missed load demand for an year for the given system. The value is 557.370 MWh/ year. Both LOLE and EENS quantifies the reliability of the system. These values help us predict for a given system, what is the expected energy output over its lifetime. It be must be noted here that these two reliability indices might be giving us a conservative estimate for the first few years of the operation of the system as there is no significant impact of degradation during the early years of the power plant. But when looking at the average impact of power degradation over its lifetime, it can be seen that LOLE and EENS indices are close to what was predicted by the developed program.

To get an insight into the functioning of the program, the number of states were varied. Particularly the number of bins of irradiance was varied to achieve more states and hence increase the resolution of the data. Increasing the number of states increases the complexity of calculations, but it achieves better accuracy. The tables given below summarizes all the possible combinations which were tried and it's impact on LOLE and EENS was recorded.

No. Irradiance bins	No. of Temperature bins	Total environment bins	No. of Degradation states	Total States	LOLE	EENS
3	5	15	5	75	0.8381	557.370
4	5	20	5	100	0.82766	553.541
5	5	25	5	125	0.81955	548.690
6	5	30	5	150	0.81065	547.866

Table 9- LOLE and EENS as a function of states

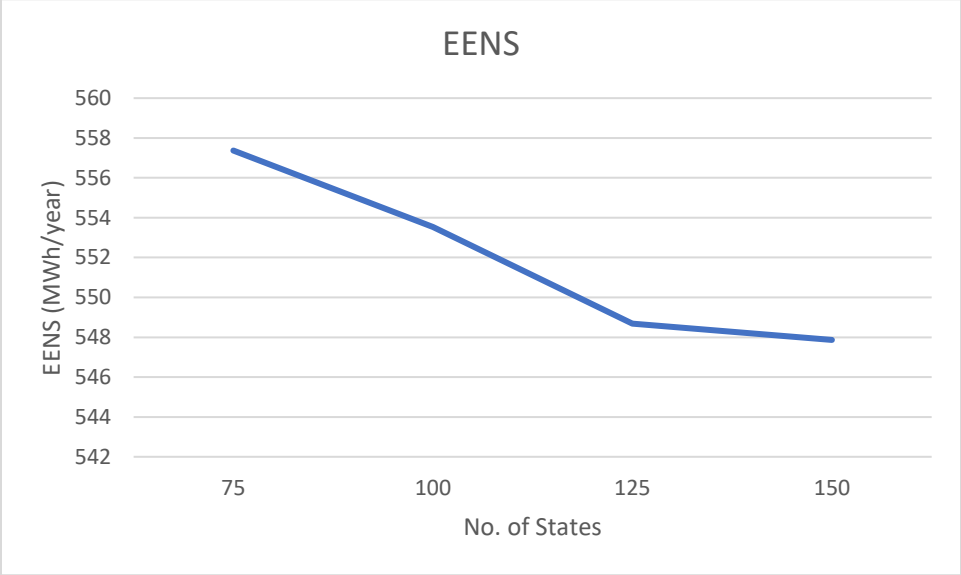

Figure 19- EENS V/s No. of states

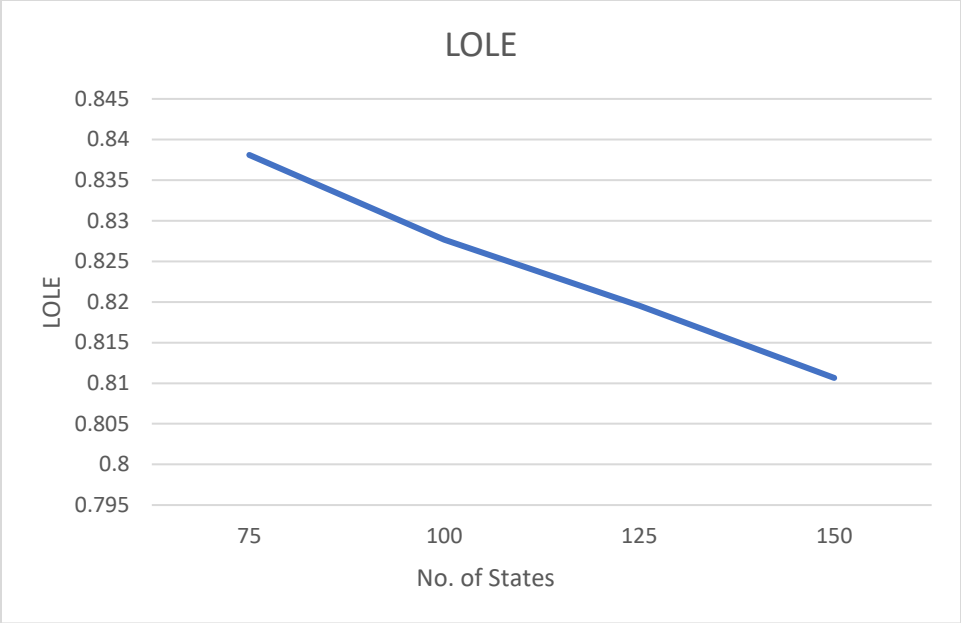

Figure 20- LOLE V/s no. of states

The graphs shown above show, as the number of states are increased the results get more precise. As the solar irradiance bins are increased, more states of solar irradiance are captured. This improves the energy estimates and both LOLE and EENS go down.

The same procedure was expanded to calculate the system reliability for the ASU renewable energy systems. ASU has established one of the largest installed Solar PV systems, providing energy to its buildings and facilities, with a total energy production of 24MW/hour. The energy produced by the renewable energy systems and the load requirements of the building is calculated and documented by ASU and made available as a part of ASU metabolism. This web interface provides us with the load files for an year at an hourly frequency. The specification details of each ASU Solar PV system are maintained at ASU-PRL and can be found in the appendix. These specifications and degradation studies done at those sites were utilized for our study here.

Structure	LOLE	EENS (MWh/year)	Actual EENS (MWh/Year)	Accuracy (%)
Barrett	0.998	296.1	301.4	98.24
Computing Commons	0.998	180.1	182.4	98.73
Cowden	0.827	236.1	236.8	99.67
Hayden	0.847	112.5	123.1	91.42
ISTB1	0.998	838.4	840.5	99.75
Weatherup	0.877	621.2	626	99.23
Wrigley	0.997	580.2	603.4	96.15

Table 10(a)-LOLE and EENS for different systems on a 24-hourly basis. The table compares the actual EENS values obtained from the real-time data of ASU systems to the EENS values calculated by the program developed. Accuracy scores show that the Program developed is highly accurate.

8 ASU systems were analyzed, Solar Energy UGF is developed using the historical environmental data used for Tempe Warehouse system and the degradation data which is specific to each system.

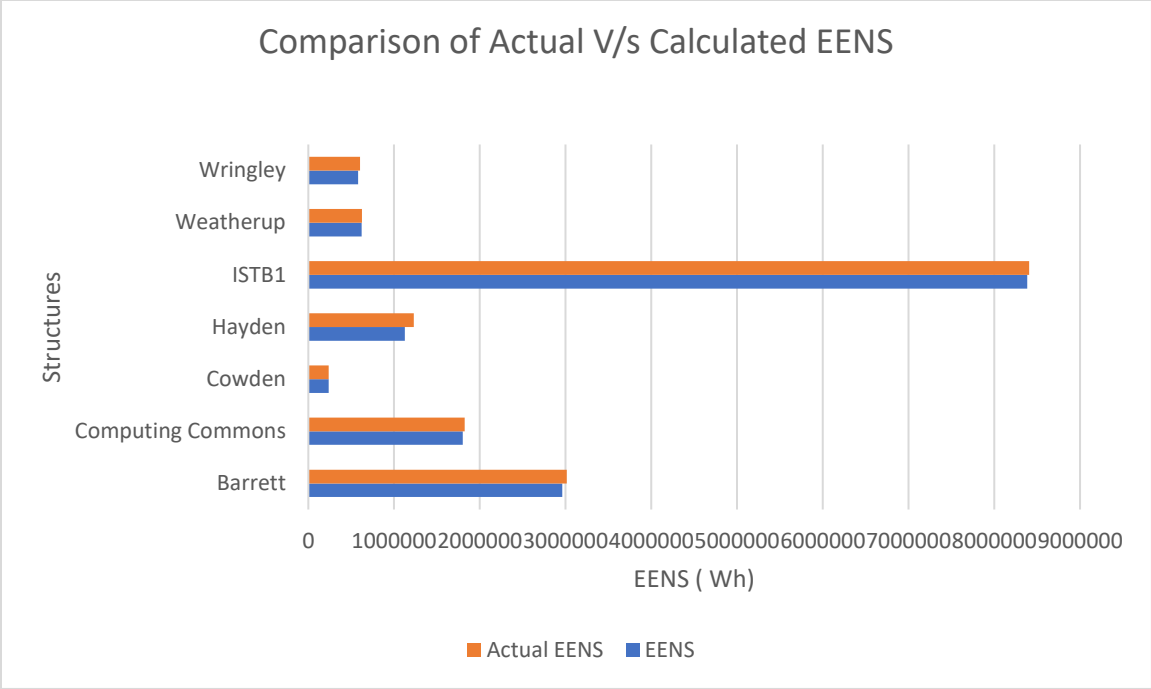

Figure 21-Comparison of Actual V/s Calculated EENS

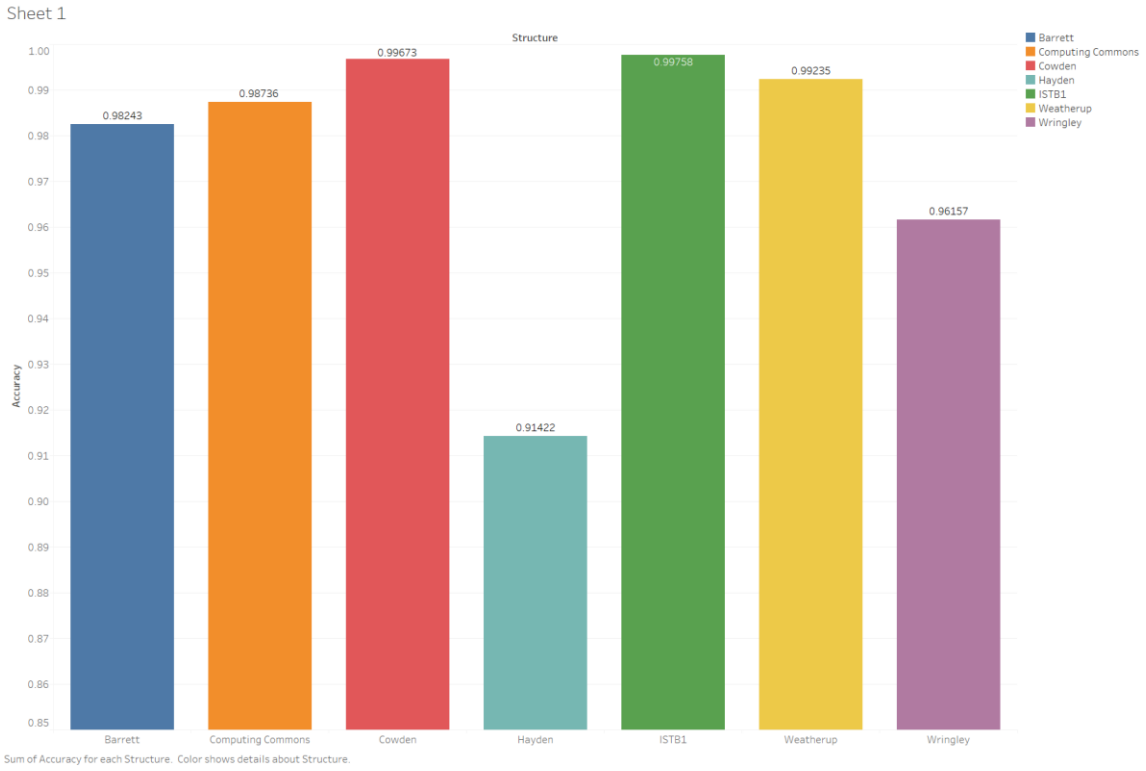

Figure 22-Comparison of Accuracy scores for different systems

To benchmark the UGF program developed, there is a need to compare it to a real-world measure which will validate the approach and prove the underlying assumptions to be a good approximation of the natural energy production pattern of a PV system. For this purpose, ASU Systems data was the benchmark as the data is continuously monitored and using smart systems its continually being relayed to a web interface. The actual energy produced by the Solar PV systems is captured and yearly summaries are available on ASU metabolism. LOLE reliability index is a measure which is good to approximate the system's reliability measure but it's an analytical measure. It is not captured by the energy data acquisition system. But EENS is a reliability index which can be calculated from live real time data. ASU metabolism provides us the yearly load on a system and yearly energy

generation of each system. The difference of these two parameters give us the actual Energy Not Supplied (EENS).

$$Actual\ EENS = \ Load\ requirement - Yearly\ energy\ production(KWhr/yr)$$

When this was compared with the calculated EENS, accuracy measure is obtained.

The figure above shows that the calculated EENS is very accurate and the model is performing well.

4.6 Short Term Energy Forecasts

Support vector regression proved to be the most accurate forecasting technique to forecast irradiance values. For utilities and grid operators, short term solar energy forecasts are very important to assess the energy production of solar assets connected to the grid. Since most grids are decades old, they were built for one-way energy transfer, i.e. from the grid to the consumer. But the rise of distributed renewable sources like Solar PV has seen the emergence of two-way energy flow. Since the grids are not clearly built for such two-way transfer of energy, an increase in share of renewables will cause much damage to the grid. Accurate forecasts in the short term would help utilities measure energy outputs into the grid by their customers who have Solar modules installed on their assets. Such forecasts also give an insight into which customers are likely to participate in load reduction programs and also helps the utilities create better pricing structure for such customers.

For the weather dataset for the Tempe area, a training set was built using the data from the year 2000-2003. The test set was the first few weeks of year 2004. The data was filtered

only to include the peak sun hours and night time values were discarded. The prediction tool would predict solar irradiance with a Mean absolute error of 93 W/m^2. for up to a week in the future by observing the past. Stretching the prediction interval to more than one week can lead to accumulation of errors. The graph given below shows how the predicted values compare to the actual values observed during the first week of 2004.

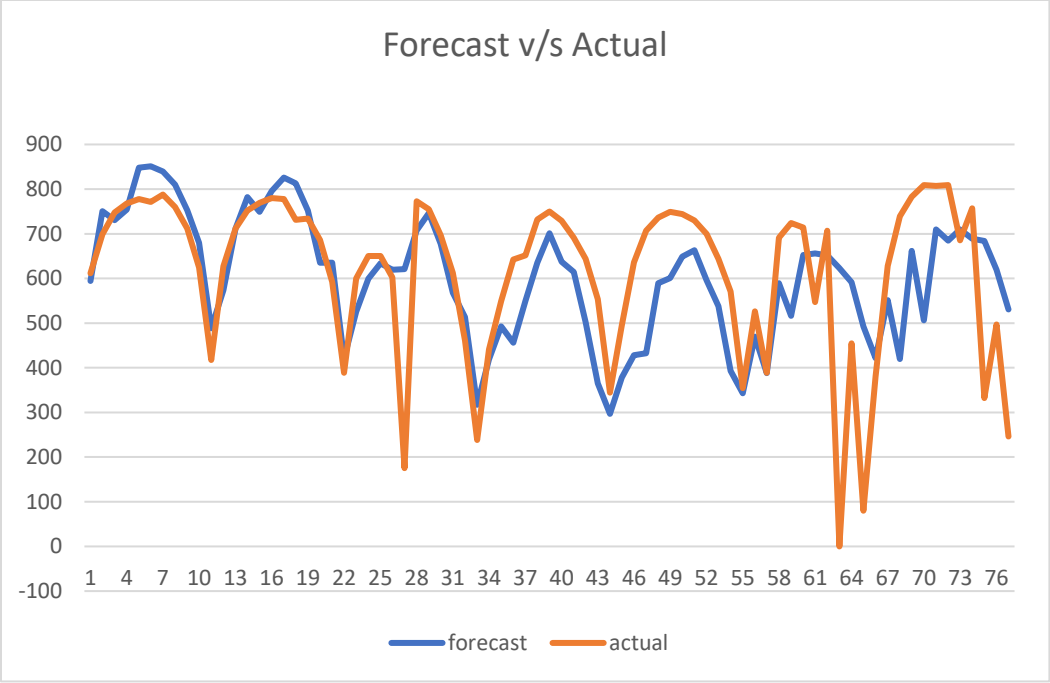

Figure 23- Irradiance comparison. Actual irradiance values observed is compared the forecasts developed for the same period. The plot shows the hourly irradiance plots and shows that forecasts seem to be very closely moving with actual irradiance values. Except for hour 61-64 where sudden drop in actual irradiance might be due to sensor malfunction where forecasts don't match the actual values.

CV=10, Radial											
Y~ t.7+t.8+t.9+temp+wind, cost = 1300, epsilon= 0.1											
Gamma	RMSE	MAE									
0.00001	145.5662	102.209									
0.0001	143.4722	103.0585									
0.001	132.9151	95.00883									
0.01	164.6386	133.5802									
Y~ t.7+t.8+t.9+temp+wind, epsilon= 0.1, gamma = 0.001											
Cost	RMSE	MAE									
1300	132.9151	95.00883									
1400	132.4574	94.60482									
1500	132.7152	94.82422									
Y~ t.7+t.8+t.9+temp+wind, cost = 1300, gamma = 0.001											
Episilon	RMSE	MAE									
0.1	132.9151	95.00883									
0.01	133.8429	94.30791									
0.001	133.3623	93.90498									
0.0001	133.1546	93.71868									
Y~ t.7+t.8+t.9+temp+wind, cost = 1300, epsilon= 0.0001											
Gamma	RMSE	MAE	MAPE								
0.01	161.2736	128.6586	0.31557								
0.001	133.7187	93.71868									
0.0001	144.5915	103.2527									
0.00001	145.5922	99.80352									
Y~ t.7+t.8+t.9+temp+wind, epsilon= 0.1, gamma = 0.00001											
Cost	RMSE	MAE									
1300	145.5662	102.209									
1400	146.0646	102.7776									
1500	146.0047	102.7252									
Y~ t.7+t.8+t.9+temp, train, cross = 10, kernel = "radial", cost = 1300, epsilon= 0.1, gamma = 0.001											
	RMSE	MAE									
W/O wind	145.1355	95.83027									

Table 11- SVR parameter configuration. For cost=1300, epsilon=0.0001 and gamma=0.001, we achieve the least generalization error rate on the test set.

The figure above shows all the configurations tried to achieve the best accuracy score. The Mean Absolute error is a good estimate of the generalization error. For the highlighted configuration, it achieved the best MAE estimate of 93 W/m^2. A high cost parameter "c", makes sure that any test set misclassification is penalized over training set error. This gives us a better estimate of generalization error but there are significantly more training errors.

The results obtained from SVR were encouraging, though further work could be done LSTM (Long-Short Term Memory) Neural networks, which are found to work the best on time series datasets. Including more features like cloud cover index, precipitation, etc can fetch better accuracy scores.

5.0 CONCLUSION

As described in the previous chapter, the UGF approach was applied to 8 systems and their LOLE and EENS values were calculated, when compared their load data. It was seen that as the number of states are increased, the LOLE and EENS values decrease marginally and those results are more accurate as more granularity is achieved. The values of the above-mentioned reliability indices will help utilities and grid operators assess the impact of renewables on the grid as the amount of energy not supplied by these assets is obtained. When this program works at scale, it can evaluate the impact of multiple Solar PV systems simultaneously, reducing complex combinatorial calculations to simple probabilistic calculations using Universal Generating Functions. The Support Vector Regression model developed forecasts Solar Irradiance for a week with an accuracy of 93 W/m^2. This short-term forecasting tool can forecast energy accurately, helping grid operators gain more information on customers' energy consumption/production. Such information can provide utilities the base to design custom demand-response programs for its customers as managing demand-response is easier than regulating the supply-side of energy. This has the potential to lower operational costs for utilities, which might result in better net-metering incentives for customers. Increase in adoption of smart grids and smart meters, along with industrial internet will pave the need for efficient forecasts.

The UGF approach can be extended to other components of the distributed grid. For a PV power system, inverters can be modelled into the UGF equation. Wind energy assets and Electric Cars can be modelled probabilistically into the UGF. The distributed grid reliability can be modelled by UGF approach and a fault tree diagram of the entire distributed grid can be performed. Instead of modelling the degradation states using steady

state markov assumptions, different iterations of state probabilities can be performed and compared. When extended to other components of the grid, a realistic estimate of system reliability of a distributed grid can be obtained as well as a direction to reduce costs and increase profitability for the grid operators and utilities.

PART 2: ENVIRONMENTAL MODELLING TO DESIGN AN AGING TEST

1.0 INTRODUCTION

1.1 Background

Accelerated tests are critical in defining quality and performance standards for PV modules. As Solar PV assets are increasingly being financed by high proportion of debt, a 1 percent change in performance output might mean a 10% difference in return on investment. Reliability of the PV modules are important in defining the Levelized cost of energy being produced by them. PV module degradation has been proven as the primary reason of long term drop in performance of a PV system when compared to other components in the system. Accelerated Lifetime Testing (ALT) procedures have been developed to effectively predict lifetimes of PV modules, study the dominant failure modes and their effect on reduction in performance output. The qualification tests do not identify all the possible field failures and are often minimum requirements to start a reliability testing methodology.

The traditional approach to Accelerated Testing has involved identifying the dominant failure modes using FMEA by looking at field failure data. Different environmental stresses were identified and ALT specification had been developed. Acceleration factors for different climate types and for different environmental stress variables have been demonstrated. There is a high confidence in the quality of the modules if the module has undergone Accelerated Lifetime Testing. But such a procedure is expensive and testing protocol has not been developed. There is need to develop climate-specific and technology specific ALT methodology which can be used to develop acceleration factors and predict lifetime of modules specific to region.

To develop a climate specific testing methodology, an effective way to model the environment as well as correlating the field environmental conditions to the actual testing environment is required. A methodology to validate the physical models of degradation with climate specific field degradation is also needed.

1.2 Scope

This thesis is mainly focused on developing an environmental model for the accelerated test to study the effect of UV stressor and develop climate specific ALT with Phoenix (hot-dry) and New York (Cold-dry). UV radiation mainly affects the Encapsulant by causing browning and delamination. The aim of this accelerated test is to predict the degradation rates in the field for Arizona within 30% accuracy of the actual field degradation rates.

This thesis presents the discrete environmental modelling approach which can be used to design an aging test, which will serve as a validation for the Arrhenius model which serves as a base for the Accelerated Test.

2. LITERATURE REVIEW

To design an appropriate environmental model to mimic the field conditions, JPL had developed an environmental cell/bin approach [5]. This approach was used to assess combined impact of different environmental variables like ambient temperature, Irradiance (UV exposure) and Relative Humidity on the degradation of PV modules. This approach forms the basis of developing an aging test in conjunction with an accelerated test. Accelerated testing is done mainly to identify all the potential failure modes and compress field degradation into a shorter time, so potential impacts of degradation can be measured in testing conditions. Tamizhmani et al describe and review various qualification standards and accelerated testing parameters, summarizing the differences between different types of accelerated tests [30]. For every environmental stressor, its respective acceleration testing standards are described. There is a need to develop a climate-specific and technology agnostic rating system and a lifetime rating system. JPL demonstrated an approach using reliability ratio to generate universal ratings for modules [31]. This environmental qualification testing methodology was one of the first steps in moving towards a climate specific approach. To determine the different modelling approaches which have conventionally been used, Tamizhmani et al have described various white box and black box modelling techniques used to calculate AF is described [30]. Also JPL has done significant work on understanding the effects of UV stressor on PV module encapsulants [32]. Based on reviewing work in the field, it's found that there is a need to develop a database of climate-specific degradation mechanisms and correlate it to the defects and performance degradation occurred in the field. Although a lot of testing protocols have been existing, with increased data sharing a need to develop universal standards has risen.

The focus on climate-specific results has led to the need to accurately develop environmental models which can correlate the physical models with actual field degradations. C-Si technology provides us with ample field data to design the kind of aging test presented in the thesis, which can be used in conjunction with an accelerated test.

3.0 METHODOLOGY

To fulfill the objectives of predicting the acceleration factor of UV degradation a combination of accelerated tests and sequential tests were designed. The objective of this thesis is to scientifically design the sequential aging test, to replicate the field conditions in the weather chambers and report the degradation which is seen. This aging test is used as a validation tool for the accelerated test. The accelerated tests are being designed to obtain the Activation Energy in the Arrhenius model, which is believed to be the primary mechanism behind UV caused degradation. Having an aging test would help verify the degradation mechanism and quantify the degradation due to each set of conditions.

3.1 Environmental Model

Environmental variables are continuous in nature, i.e. they vary continuously with time and don't have any breaks in their measurements. Examples of continuous functions include the sine function, exponential function etc. Examples of discrete functions are step functions, binary outputs, etc. There are several ways to model the environment. But two of the most relevant classifications can be made namely- discrete method and continuous method. For this thesis, the environmental model was restricted to the discrete model, since the requirement for designing the aging test was discrete values of the test conditions.

The JPL approach to environmental modelling was the primary focus of this thesis. This approach involved discretizing different environmental variables into different ranges called "bins". The purpose of this modelling approach is to seek the combined effect of all the environmental variables taken together. So henceforth, each bin represents a collective

state of all the considered environmental bins. For designing an aging test for PV modules, there was need to identify the environmental variables which are the major stressors for a PV module. As the literature suggests, Module temperature (function of ambient temperature and wind speed), Irradiance and Relative Humidity are the main PV stressors [32].

Irradiance contains Ultraviolet radiation as a part of the spectrum of light, which causes encapsulant browning and bleaching [32].

The approach used involved discretizing the 3 main environmental variables- Irradiance, Module temperature and Relative Humidity. To design an effective accelerated test, the weather conditions of the region of interest, their interplay and the effect of the conditions on the field modules must be understood [1]. Most tests focus mainly on one variable and base their results on the impact that variable had on the test specimen. But the combined effect of all these variables on the modules in field is of interest here. The environmental data was compiled from different sources and was merged. The frequency of the observation was set to every 10 minutes since the data was available at such a detailed level. For long term climatic exposures, 10 minute interval data is adequate to estimate the environmental conditions prevailing at the particular region. Also, to observe how the environment condition shifts from one bin to the other, it's important to keep the resolution of the data at as high resolution as possible.

Once the data was compiled, a simple algorithm was written in the R programming language to discretize the variables into different bins. Discretization process was done keeping in mind the range of Temperature, Irradiance and Relative Humidity values observed in the field. The values of temperature obtained were the ambient temperature

values. To design the aging test for PV modules, Module temperatures were of higher significance as they directly correlate with some of the failure modes. The ambient temperatures were converted to module temperatures using the Sandia-King model. This model represents module temperature as a function of Ambient temperature and Wind speed. Both these variables are available in the Raw data. Using this information, the three variables of importance are decided to be- Irradiance, Module Temperature and Relative Humidity.

The range values set were setup appropriately to capture all the variation in these variables.
Temperature was divided into 4 ranges; 0-35 °C, 35-50 °C, 50-65 °C, 65- 80 °C
Irradiance was divided into 3 ranges: 0-400 W/m2, 400-700 W/m2, 700-1050 W/m2
Relative Humidity was divided into 3 ranges: 0-30%, 30-60%, 60-100%

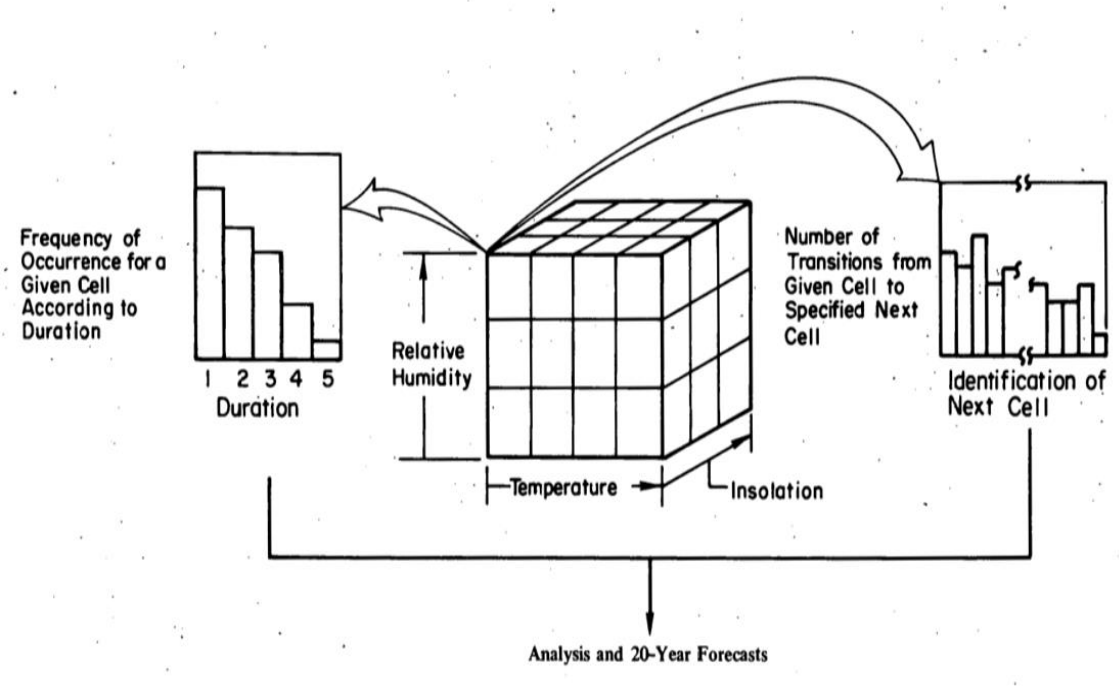

Figure 24- Environmental Bin representation [5]

Overall there are 4× 3 × 3 = 36 environmental bins which are created because of discretizing the environmental variables. The above figure represents the concept of environmental bin. The environmental bin can be visualized as a 3-dimensional cube, with each axis representing a particular environmental variable. The number of divisions on each axis represents the ranges the variable is divided into. Looking at the figure it can be seen as each bin being a smaller cube within the bigger cube.

Analytically, each bin represents an environment condition. Each bin is a given a code. For example, if at a given time the temperature is 26 °C, Irradiance is 1000W/m^2 and Relative Humidity is 90 % then the bin code is 131. With the environmental bins created and mapped, a frequency distribution for every bin is obtained along with how long the conditions stay in that bin.

3.2 Environmental Bin Statistics

A lot of insight can be gained about the nature of the changing environmental conditions by observing different Bin Statistics. Bin statistics refers to statistically analyzing how the bins are changing over time. For the aging test, two things are of utmost importance which need to be analyzed from the past data.

1. How many times is a particular environmental condition (Bin) occurring
2. For how long does the Environment stay in the same Bin

For the first condition, a simple counting algorithm was written in the R language, which generated a frequency distribution of all the bins and its respective counts for the previous 5 years of data. This distribution helps us see which Bins have occurred the most amount of time. When designing an aging test, the bins with the highest frequencies have to be picked, and approximate the ones which occur less number of times by not including them in the test. It also helps identify environmental conditions which are more important from a degradation point of view. By knowing the exact distribution of environmental conditions (Bins), the nature of the environment for the future can be predicted, assuming that over the long-term weather fluctuations remain more or less similar to that of the past. If a large dataset is analyzed, then it can be approximated that, the patterns observed in the past remain similar as what would be observed in the future. If short term forecasts are analyzed, they might be different from that of the past values at that time. But if averaged it to a long period, then the environment fluctuations have a typical behavior for every region.

Another interesting Environment Bin statistic is the frequency of duration of a environmental condition. This is shown in the figure below.

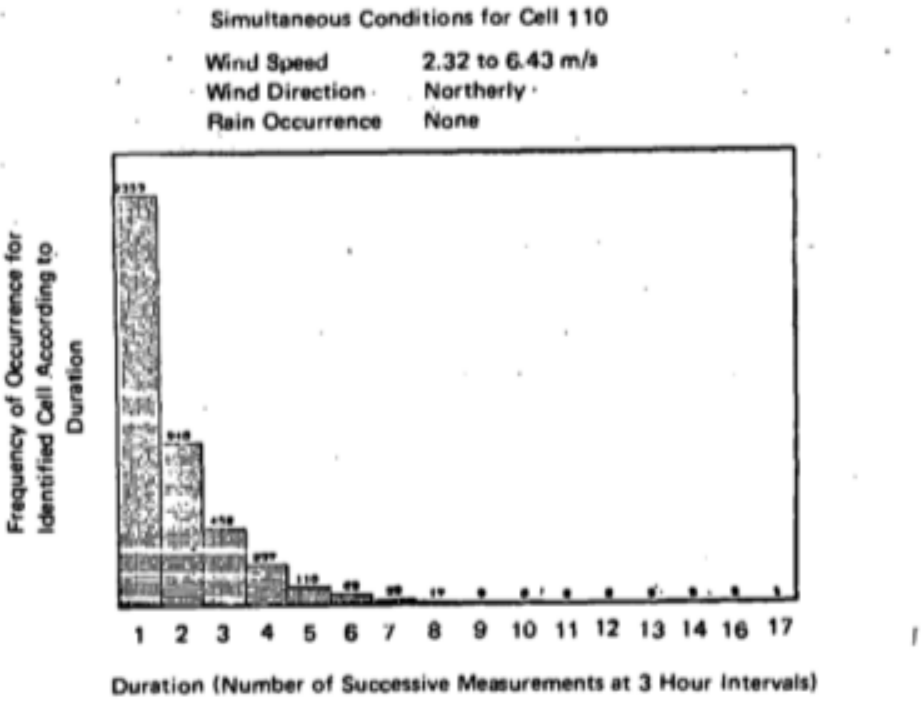

Figure 25- Environmental Bin duration frequencies [5]

This represents the frequency distribution of successive repetitions of the same bin consecutively. Such graphs are produced for each of the 36 bins and analyzed. When choosing the specific bins conditions which will be used for the aging test, this frequency distribution is also analyzed. If a bin has two or more successive occurrences, it is given more importance in getting chosen. For example, 3 successive measurements in the past data shows that for 30 minutes (3x10) the environment remained in the same condition. Frequency of this phenomenon helps analyze the duration of the condition. If the nature of the duration of a condition is understood, then correlation of a specific degradation

mechanism with such exposures can be understood too. In some cases, degradation mechanisms are not simply additive, i.e. they only take place after sufficient exposure to an environmental factor is seen [5]. Along with frequency of the bin, the nature of its duration also needs to be studied.

3.3 Aging Test

To validate the acceleration factor obtained from the accelerated test, an aging test needs to be designed which exactly mimics the field conditions. The acceleration factor is obtained using the Arrhenius model, by running the accelerated test at 3 different temperatures. This gives the required slope to calculate the activation energy, which in turn helps calculate the acceleration factor. To verify that the acceleration factor obtained corresponds to the actual field degradation, simultaneously testing of the field degradation on similar modules needs to be done. The degradation rates obtained by the aging test run at field exposure conditions are verified by the Arrhenius model used by Shantanu et al at ASU-PRL.

The primary method used to run the aging test is by exactly replicating the field conditions in the aging test. For this test, the artificial weather chamber is used. This chamber can vary the Module temperature, UV dosage and Relative humidity is kept constant at around 20%. The UV lamps are used to simulate the UV intensity present in sunlight, which is around 5% of actual irradiance values. The UV intensity also influences the operating module temperature. To achieve variations in module temperature, a differential rack is being developed which can change the distances between the UV lamps and the modules. This difference in distances between the modules and the UV lamps act as a Module temperature

differential. With the physical construction completed, the test conditions needed to be decided. Since the one run of the entire aging test might take several months, the number of environmental bins to be simulated is limited. Some of the bins can be eliminated.

The UV induced browning takes place mainly as a function of UV insolation, i.e. product total UV irradiance and the time. The different Irradiance bins and their relative frequencies are converted to Insolation and then 5% of those values are UV insolation values. The total testing time is calculated based on the rating of the UV lamp present at PRL's environment chambers and the temperature at which the modules are exposed.

Literature shows that for module temperatures below 35^0C, Isc degradation doesn't take place [30]. Also when the Air mass value is above 1.5 the UV intensity is too low to be of any significance to the Encapsulant browning as the UV photons are dispersed in the atmosphere, hence the data is filtered to consider data only from 9AM- 5PM. Taking into account the following changes, the aging test conditions were designed which is elaborated in the subsequent section.

The modules used for the test were M55 modules. These modules are similar in construction type and technology to the MSX60 modules undergoing the accelerated test. This makes it scientifically accurate to compare the two types of modules.

3.4 Nesting of Test Conditions

While designing the aging test, it needs to be noted that the test is sequential. It's important to design the test conditions as well as the order in which the sequential test should progress. UV stressor is the primary reason which causes Encapsulant Browning. This problem is accelerated due to higher module operating temperatures. But at high temperatures oxygen causes a bleaching reaction which causes the browning to reduce. So there is a need to capture only the effect of browning for which oxygen bleaching's effects should be negated.

This makes designing an aging test very challenging. Nesting of test conditions ensures that the aging test proceeds in the decreasing order of environmental stress. This is done to preserve the combinatorial nature of stresses, i.e. it makes a difference when trying to achieve the combined effect of UV stressors (UV dosage and High Module Temperatures) and relative humidity. Considering the effect of high values for only one stressor won't achieve the actual degradation mode, so there is a need to see the effect of all the stressors which causes a net high environmental stress. In this test, high values of UV dosage, Module temperature and relative humidity causes the highest stress. So the nested conditions for the environmental variables are arranged in the decreasing order of their values. Relative humidity doesn't have a significant impact on the browning reaction, hence an average humidity of 20% is maintained.

The test times are calculated based on the time taken to reach the UV insolation in field, which is calculated from the environmental bins produced.

4.0 RESULTS AND DISCUSSION

Raw data used here was obtained from the Solar Anywhere database, for the years 2000-2004. The data consisted of 265566 data points taken at 10 minute intervals. The variables of focus were Direct Normal Irradiance, Relative Humidity and Ambient Temperature (Module Temperature).

```
> head(initialdata)
  Unit Year Date TOD WindAV WindPk WindDir Rain    GHI   DNI Tracker8Temp
1  113 2000    1   0  3.854  6.199   189.4    0 -2.269 0.484        12.22
2  113 2000    1  10  3.884  5.480   197.6    0 -2.655 0.232        12.16
3  113 2000    1  20  4.204  6.055   194.5    0 -2.531 0.366        12.28
4  113 2000    1  30  4.800  8.500   218.7    0 -2.307 0.751        12.64
5  113 2000    1  40  3.924  6.918   226.6    0 -2.710 0.426        12.41
6  113 2000    1  50  5.145  7.780   244.2    0 -2.408 0.822        12.94
  WindAvMetal WindPkMet Humid TempAmbRoof
1       1.231     4.045 49.17       13.08
2       0.953     3.724 49.22       13.01
3       1.808     4.299 47.89       13.05
4       2.466     6.483 45.83       13.25
5       1.096     5.552 44.44       13.37
6       3.411     6.686 43.03       13.55
```

Figure 26- Weather Dataset showing all the parameters.

This figure shows the format of the raw data and the columns. This raw data was filtered in the following ways using R programming language

1. Missing values were replaced with the mean values of their neighborhood points.
2. Data was filtered to include only values from 9AM-5PM at 10 minute frequencies.
3. The data obtained was converted to bins and it's frequencies were noted.
4. Initial analysis with all module temperatures were considered
5. Finally, module temperatures above 35^0C were chosen to reduce the test time.

Initially the bins generated for all the module temperature bins. This is done to verify that the UV insolation obtained by modelling matches the actual studies done on UV insolation. The frequency distribution of the Environmental Bins is shown below. This distribution represents for all module temperatures.

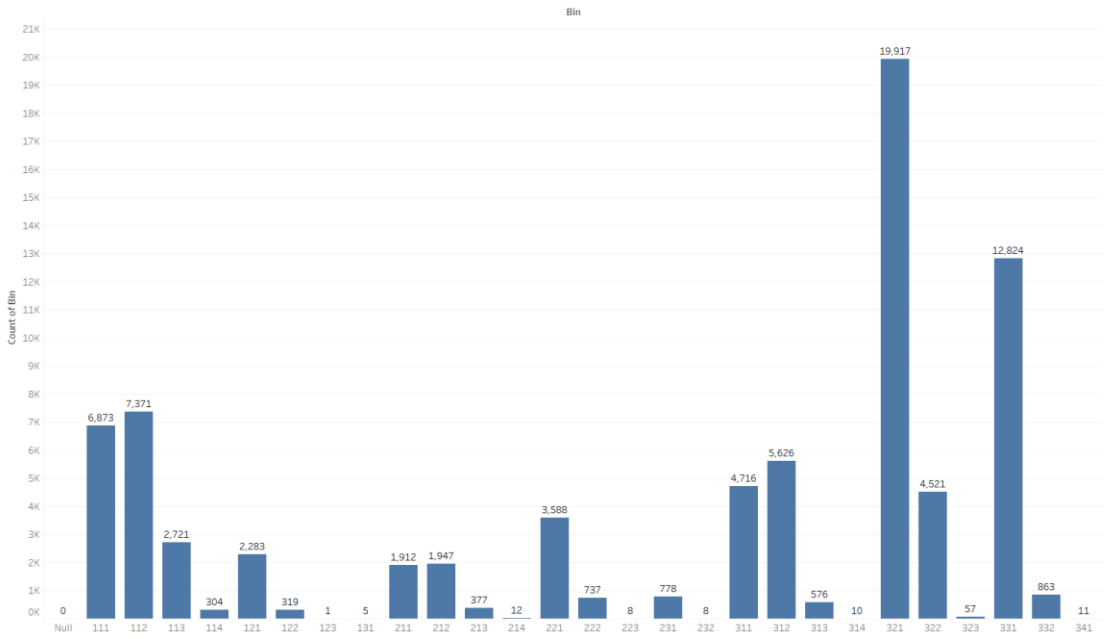

Figure 27- Environmental Bin Frequencies for years 2000-2004

Only the high frequency bins are considered and the respective bin details are represented in the table given below. The UV insolation is obtained from the mean Irradiance of each bin multiplied by its frequency of occurrence. The Histogram given below also shows us the distribution of Irradiance values between 9AM-5PM. It shows majority of the values are lying towards the right at higher irradiance levels. It's evident which irradiance bins are repeating in high frequency. Irradiance bins 2 & 3 are the ones which are repeated the most.

Bins	Irradiance range	Module temperature range	Irradiance Mean (W/m^2)	Module Temp. mean (C)	Humidity range	Frequency (10 min interval)	Insolation of UV(Wh/m^2)
341	700-1050	65-85	996	67	0-30	11	100.43
332	700-1050	50-65	913	52.57	30-60	863	7222.59
331	700-1050	50-65	900.5	54.15	0-30	12824	105856.77
321	700-1050	35-50	863	43.8	0-30	19917	157560.06
322	700-1050	35-50	822.4	41.79	30-60	4521	34082.31
221	400-700	35-50	521	42.3	0-30	3588	17135.69
121	0-300	35-50	150	39.9	0-30	2283	3139.12
311	700-1050	0-35	893	29	0-30	4716	38604.39
312	700-1050	0-35	860	30	30-60	5626	44351.63
211	400-700	0-35	557	25	0-30	3859	19703.41
112	0-400	0-35	73.6	20	30-60	7371	4972.96
111	0-400	0-35	84	11	0-30	6873	5292.21
113	0-400	0-35	38.4	20	60-100	2721	957.79

Table 12- Insolation of UV for 5 years of data

Bins	Irradiance range	Module temperature range	Humidity range	Irradiance Mean (W/m^2)	Insolation of UV (yearly)(Whr/m^2)
341	700-1050	65-85	0-30	996	23.63058824
332	700-1050	50-65	30-60	913	1699.433137
331	700-1050	50-65	0-30	900.5	24907.47686
321	700-1050	35-50	0-30	863	37072.95706
322	700-1050	35-50	30-60	822.4	8019.367529
221	400-700	35-50	0-30	521	4031.927059
121	0-300	35-50	0-30	150	738.6176471
311	700-1050	0-35	0-30	893	9083.385882
312	700-1050	0-35	30-60	860	10435.67843
211	400-700	0-35	0-30	557	4636.096667
112	0-400	0-35	30-60	73.6	1170.110118
111	0-400	0-35	0-30	84	1245.225882
113	0-400	0-35	0-30	38.4	225.3628235

Table 13- Insolation of UV Annually

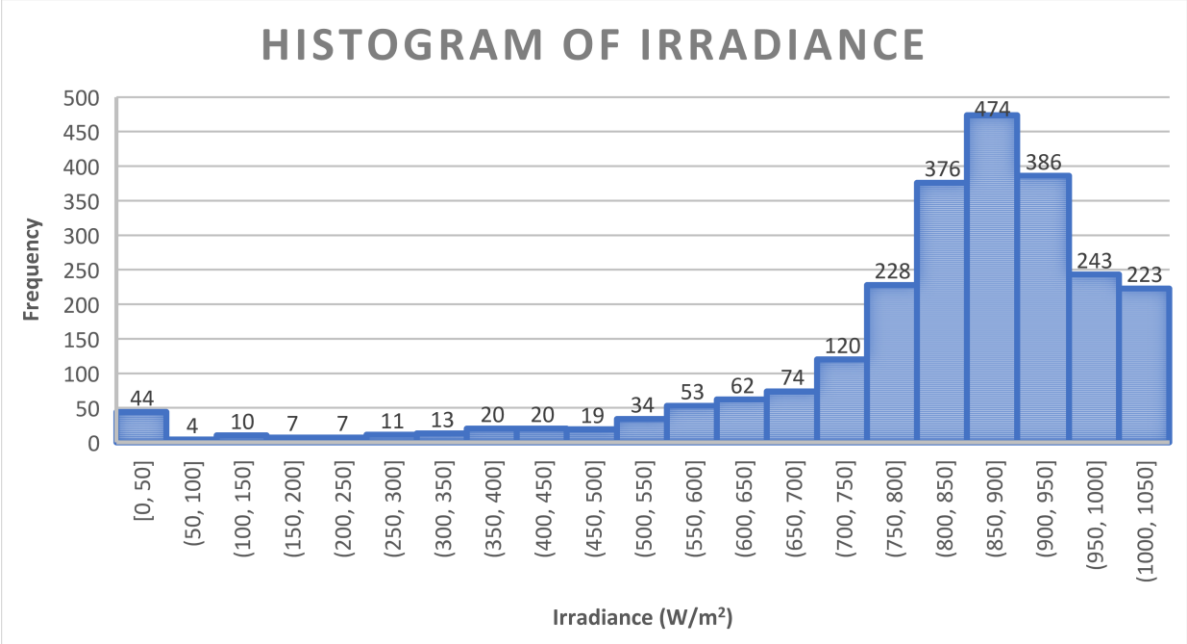

Figure 28-Irradiance Histogram. X-axis represents different bins of histogram and y-axis represents its frequency. It can be seen that for the filtered 9am-5pm data, irradiance is mostly in the range 800-1000 W/m^2 for Phoenix, AZ.

This data from the field is compared to UV studies done in the past. Koehl et al conducted UV insolation studies at PRL [33]. According to these studies, annual UV insolation in Arizona (hot-dry) climate is around 120 KW/m^2. This data was verified and proven to be the most accurate UV estimate obtained for Phoenix area. According to the data analyzed in this thesis, a UV insolation of 440 KW/m^2 is obtained. The data considered for 2000-2004 contains a lot of missing values in the year 2004, so the number of years considered comes up to 4.25 years and not 5 years. Taking this into account, it turns out that annual UV insolation according to the environmental bin approach is around 104 KWh/m^2 as shown in Table 13. This is close to the values obtained previously, validating the environmental bin approach used. The graphs below show a comparison between UV insolation obtained in literature to that obtained by environmental modelling.

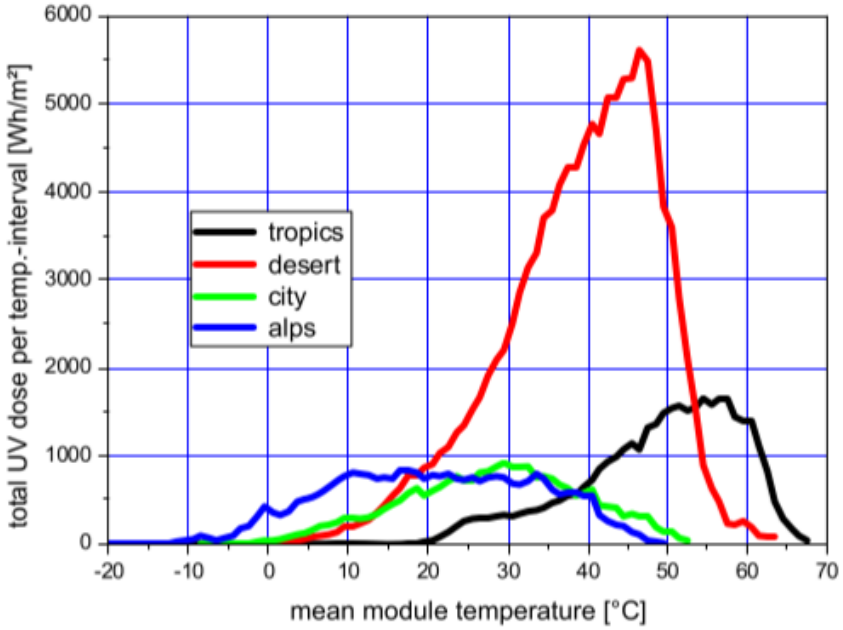

Figure 29- Annual UV insolation for different Module Temperatures. The desert climate plot is of interest here. The area under the Red Curve comes up to 120KW/m^2 [33]

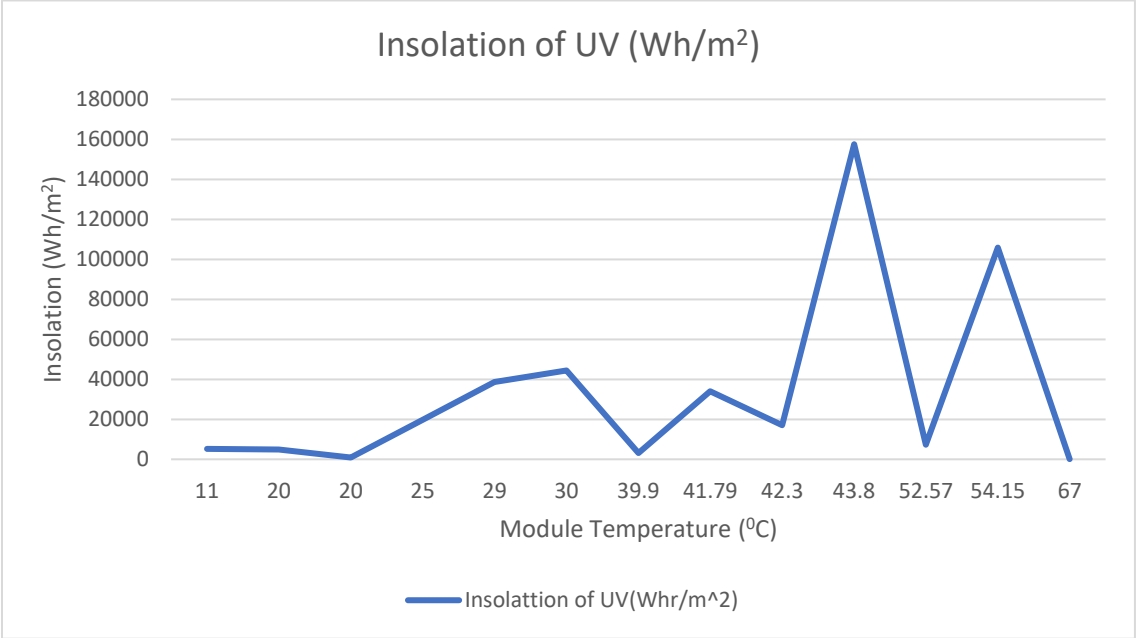

Figure 30- UV Insolation as a function of Module Temperature obtained from environmental bin approach. This is for years 2000-2004.

The plot shown here is like the one obtained by Koehl et al. The plot obtained here is not as smooth since it's based on a discrete environment model. The sum of the area under the curve gives 440KW/ m^2.

Looking at Table 12, necessary test conditions can be designed. Since there are a lot of bins, bins are nested together to reduce the number of conditions. To not lose any detail, bins are nested based on common conditions. For the UV browning reaction, the module temperature decides the rate of the reaction. So, the bins with the same module temperature bins were nested together and their respective UV insolation doses were added. Bin 341 occurs only 11 times in 5 years of data, so it can be ignored. Three testing conditions to be run at three different module temperatures and UV doses are obtained. Time to simulate the aging test is calculated based on how long it will take to reach the UV dosage for each bin. The UV lamps used at PRL have a rating of 250 W/m^2. This reduces the time of the aging test as its being accelerated.

Module Temperature	Insolattion of UV(Whr/m^2)	Time to simulate 5 years of exposure (Hours)	Temperature factor	Aging test time for 5 years (Hours)	Aging test time for 7 years (Hours)	Aging test time (Days)
22	113882.4042	455.529617	0.25	1822.118467	3001.136	125.0473458
42	211917.1945	847.668778	1	847.668778	1396.16	58.17334751
54	113079.3675	452.31747	2	226.158735	372.4967	15.5206975

Table 14- Final conditions for the aging test for 7 years in hot-dry climate, Phoenix AZ (2000-2004). Note: These conditions which were used are for the data recorded between 9am and 5pm at 10 minute intervals.

When its assumed that the future environmental conditions remain the same as the past, the bin frequencies and the observed times can be extrapolated to the future as well. The MS55 modules are field exposed for 18 years. After being exposed for 25 years, the

degradation mechanisms change as widely reported in literature. For the aging test, we have limited time and limited runs. The aim here is to simulate aging for additional 7 years. Table 13 shows, to achieve the given UV insolation for each bin, the insolation values are divided by UV lamps' rating. This would help achieve the time to run the test for simulating 5 years in the field. Temperature factor is taken into account, because its reported in previous studies that with every 10^0C rise in temperature the reaction rate doubles [33]. Considering both the factors, aging test times to simulate 7 years in the field is generated. As seen in Table 13, 3 different conditions are generated. The variables to be controlled in the environmental chambers are- module temperature and the time of exposure. The time to run this test comes up to 200 days. If the effect of module temperatures below 35^0C is not considered, then the first condition can be eliminated. This brings down the test time to 70 days.

5.0 CONCLUSION

An experimental methodology in developing an aging test in conjunction with a traditional accelerated test was presented in this thesis. The different environmental conditions and their respective run-times in the environmental chambers is quantified and presented. These results provide a framework in simulating actual environmental conditions of the field responsible for encapsulant browning failure mode for a hot-dry climate. This aging test proves as a validation for the physical Arrhenius model driving the Browning reaction mechanism. The activation energy obtained from running the accelerated test at 3 different conditions can be verified by measuring the performance drop experienced by the aging test. This validation will prove useful when trying to predict the field degradation for other modules in the future.

When conducted successfully and the validation is successful, a design of experiments approach could be used to statistically validate the degradation being caused by the failure modes attributed to UV degradation. Such an experiment setup was out of scope for the current project, but could be developed to quantify the degradation caused by UV Stressor. Designed experiment should be done in such a way that it falls within the time constraint, as well as doesn't induce different failure modes. By avoiding these pitfalls further quantifiable studies could be made, which will validate physical models hence improving our definition of module reliability and at the same time developing new standards which educates the industry about the nature of such failures.

REFERENCES

[1] M. Patsalides *et al.*, "The Effect of Solar Irradiance on the Power Quality Behaviour of Grid Connected Photovoltaic Systems."

[2] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Photovoltaic array performance model," *Sandia Rep. No. 2004-3535*, vol. 8, no. December, pp. 1–19, 2004.

[3] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.

[4] R. Billinton and Y. Gao, "Multistate wind energy conversion system models for adequacy assessment of generating systems incorporating wind energy," *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 163–170, 2008.

[5] R. E. Thomas and D. C. Carmichael, "Terrestrial Service Environments for Selected Geographic Locations," 1976.

[6] I. Ushakov, *Probabilistic Reliability Models*. 2012.

[7] R. Pan, J. Kuitche, and G. T. (Mani), "Degradation analysis of solar photovoltaic modules: Influence of environmental factor," *2011 Proc. - Annu. Reliab. Maintainab. Symp.*, pp. 1–5, 2011.

[8] D. C. Jordan and S. R. Kurtz, "Photovoltaic degradation rates - An Analytical Review," *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 1. pp. 12–29, 2013.

[9] J. M. Kuitche, "A Statistical Approach to Solar Photovoltaic Module Lifetime Prediction," no. December, 2014.

[10] N. Gorjian, L. Ma, M. Mittinty, P. Yarlagadda, and Y. Sun, "A REVIEW ON DEGRADATION MODELS IN RELIABILITY ANALYSIS."

[11] W. Meeker and L. Escobar, "Statistical methods for reliability data," p. 680, 1998.

[12] M. V Mancenido, "Online Degradation Monitoring of Solar Photovoltaic Modules Using One-step-ahead ARIMA Forecasts."

[13] M. J. Zuo, J. Renyan, and R. C. M. Yam, "Approaches for reliability modeling of continuous-state devices," *IEEE Trans. Reliab.*, vol. 48, no. 1, pp. 9–18, 1999.

[14] C. J. Lu, W. Q. Meeker, and Q. Meeker, "Measures to Estimate Using Degradation a Distribution," *Technometrics*, vol. 35, no. 2, pp. 161–174, 1993.

[15] J. Park, W. Liang, J. Choi, A. A. El-Keib, M. Shahidehpour, and R. Billinton, "A probabilistic reliability evaluation of a power system including solar/photovoltaic cell generator," *2009 IEEE Power Energy Soc. Gen. Meet. PES '09*, no. 1, pp. 1–6, 2009.

[16] M. Theristis and I. A. Papazoglou, "Markovian reliability analysis of standalone

photovoltaic systems incorporating repairs," *IEEE J. Photovoltaics*, vol. 4, no. 1, pp. 414–422, 2014.

[17] M. Studies, E. S. Kumar, and A. Sarkar, "Markov chain modeling of performance d egradation of photovoltaic syst," vol. 12, no. 1, pp. 39–44, 2012.

[18] Y. Massim, A. A. Zeblah, A. M. Benguediab, A. A. Ghouraf, and R. Meziane, "Reliability evaluation of electrical power systems including multi-state considerations."

[19] W. El-Khattam, Y. G. Hegazy, and M. M. A. Salama, "Investigating Distributed Generation Systems Performance Using Monte Carlo Simulation," *IEEE Trans. POWER Syst.*, vol. 21, no. 2, 2006.

[20] Y. F. Li and E. Zio, "A multi-state power model for adequacy assessment of distributed generation via universal generating function," *Reliab. Eng. Syst. Saf.*, vol. 106, pp. 28–36, 2012.

[21] K. S. Perera, Z. Aung, and W. L. Woon, "Machine Learning Techniques for Supporting Renewable Energy Generation and Integration: A Survey," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8817, pp. 81–96, 2014.

[22] A. Zurborg, "Unlocking Customer Value : The Virtual Power Plant," *World Power*, pp. 1–5, 2010.

[23] C. B. Jones, J. S. Stein, S. Gonzalez, and B. H. King, "Photovoltaic system fault detection and diagnostics using Laterally Primed Adaptive Resonance Theory neural network," *2015 IEEE 42nd Photovolt. Spec. Conf.*, no. 1, pp. 1–6, 2015.

[24] Y. Ren, P. N. Suganthan, and N. Srikanth, "Ensemble methods for wind and solar power forecasting—A state-of-the-art review," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 82–91, 2015.

[25] V. Vapnik and a Lerner, "Pattern recognition using generalized portrait method," *Autom. Remote Control*, vol. 24, pp. 774–780, 1963.

[26] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting Solar Generation from Weather Forecasts Using Machine Learning," *Proceedings, 2011 IEEE Int. Conf. Smart Grid Commun.*, pp. 528–533, 2011.

[27] J. B. Yerrapragada, "Short-Term Power Forecasting of Solar PV Systems Using Machine Learning Techniques," no. July, pp. 1–5, 2013.

[28] Y. Guo, X. Li, G. Bai, and J. Ma, "Time series prediction method based on LS-SVR with modified Gaussian RBF," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7664 LNCS, no. PART 2, pp. 9–17, 2012.

[29] P. E. Bett and H. E. Thornton, "The climatological relationships between wind and solar energy supply in Britain," *Renew. Energy*, vol. 87, pp. 96–110, 2016.

[30] G. TamizhMani and J. Kuitche, “Accelerated Lifetime Testing of Photovoltaic Modules Solar America Board for Codes and Standards,” *Sol. ABC*, no. July, p. 106, 2013.

[31] A. R. Hoffman and R. G. Ross Jr., “Environmental Qualification Testing of Terrestrial Solar Cell Modules.,” *Conf. Rec. IEEE Photovolt. Spec. Conf.*, pp. 835–842, 1978.

[32] M. D. Kempe, “Accelerated UV Test Methods for Encapsulants of Photovoltaic Modules,” *33rd IEEE*, no. May, p. 5, 2008.

[33] M. Koehl, D. Philipp, N. Lenz, and M. Zundel, “Development and application of a UV light source for PV-module testing,” *SPIE Reliab. Photovolt. Cells, Modul. Components, Syst. II*, vol. 7412, pp. 741202–741207, 2009.

APPENDIX-A

R-CODE TO OBTAIN THE MULTI-STATE RELIABILITY MODELS AND SVM MODELS

```
initialdata<- read.csv("finaldataset-compiled.csv")
datanew<- initialdata[, 1:20]
# A is Temperature, B is Humidity, C is Irradiance
myvar<- c("Year", "Date", "TOD", "Tracker8Temp" , "Humid", "DNI", "WindAV")
datanew<- datanew[myvar]
names(datanew)[7]<- "ws"
names(datanew)[6]<- "irr"
names(datanew)[4]<- "temp"
datanew$irr<- ifelse(datanew$irr<0, 0, datanew$irr)
datanew$irr<- ifelse(datanew$irr>1050, 1050, datanew$irr)
datanew$tcell<- datanew$irr*exp(-3.56-0.075*datanew$ws)+datanew$temp
datanew$tcellbin<- ifelse(datanew$tcell<35, 1, (ifelse(datanew$tcell<50,2,(ifelse(datanew$tcell<65
datanew$irrbin<- ifelse(datanew$irr<200, 1, (ifelse(datanew$irr<400,2,(ifelse(datanew$irr<750,3,if
#concatanated 3 columns to give a new column containing the cell number for each observation

datanew$bin<- str_c(datanew$irrbin,datanew$tcellbin)
#performing the pivot table operation aggregating irr and tcell by mean
datanew<- data.table(datanew)
setkey(datanew, bin)
msmodel= as.data.frame(datanew[ , j=list(irrmean= mean(irr, na.rm= TRUE), tcellmean=mean(tcell, na
msmodel<- msmodel[-1, ]
msmodel$Vy<- 60- 0.207*(msmodel$tcellmean-25)
msmodel$Iy<- msmodel$irrmean*(6.5- 0.0005*(msmodel$tcellmean-25))
msmodel$power<- msmodel$Vy*msmodel$Iy*0.7692/1000


w= table(datanew$bin)
msprob= as.data.frame(w)
msprob$prob<-  msprob$Freq/225941

degstates<- c(400, 380, 360, 340, 320)
degprob<- c(0.018, 0.024, 0.036, 0.0721, 0.8499)

#polynomial multiplication of energy and degradation using "mouter" function
#install matlab package
energyugfprob<- outer(msprob$prob, degprob)
#converting them to vectors
as.vector(t(energyugfprob))
energyugfpower<- outer(msmodel$power, degstates)
as.vector(t(energyugfpower))
```

```
#the load UGF
loadugfpower<- c(30000, 80000, 150000, 225000)
as.vector(loadugfpower)
loadugfprob<- c(0.185679, 0.5873, 0.2264, 0.000455)
as.vector(loadugfprob)

#LOLE
LOLE<-0
i<-0
j<-0
while(i< length(energyugfpower)) {
  while(j< length(loadugfpower)) {
    if(energyugfpower[i+1]<loadugfpower[j+1])
      LOLE<- LOLE + energyugfprob[i+1]*loadugfprob[j+1]
    j<- j+1
  }
  i<- i+1
  j<- 0
}
LOLE

#EENS
EENS<-0
i<-0
j<-0
while(i< length(energyugfpower)) {
  while(j< length(loadugfpower)) {
    if(energyugfpower[i+1]<loadugfpower[j+1])
      EENS<- EENS + energyugfprob[i+1]*loadugfprob[j+1]*(loadugfpower[j+1]-energyugfpower[i+1])
    j<- j+1
  }
  i<- i+1
  j<- 0
}
EENS<- 8736*EENS/1000
EENS
```

```
initialdata<- read.csv("finaldataset-compiled.csv")
datadoe<- initialdata[, 1:20]
#to convert 10 min readings to 3 hourly readings
datadoe<- subset(datadoe, TOD<=1510 & TOD>=900)
datadoe<- subset(datadoe, TOD==900 | TOD==1200 | TOD==1500 | TOD==1700)
# A is Temperature, B is Humidity, C is Irradiance
myvar<- c("Year", "Date", "TOD", "Tracker8Temp" , "Humid", "DNI", "WindAV")
datadoe<- datadoe[myvar]
names(datadoe)[7]<- "ws"
names(datadoe)[6]<- "irr"
names(datadoe)[4]<- "temp"
datadoe$irr<- ifelse(datadoe$irr<0, 0, datadoe$irr)
datadoe$irr<- ifelse(datadoe$irr>1050, 1050, datadoe$irr)
datadoe$irrbin<- ifelse(datadoe$irr<400, 1, (ifelse(datadoe$irr<600,2,(ifelse(datadoe$irr<1060,3,4
datadoe$tcell<- datadoe$irr*exp(-3.56-0.075*datadoe$ws)+datadoe$temp
datadoe$tcellbin<- ifelse(datadoe$tcell<35, 1, (ifelse(datadoe$tcell<50,2,(ifelse(datadoe$tcell<65
datadoe$Humidbin<- ifelse(datadoe$Humid<30, 1, (ifelse(datadoe$Humid<60,2,(ifelse(datadoe$Humid<10

#concatanated 3 columns to give a new column containing the cell number for each observation
datadoe$bin<- str_c(datadoe$irrbin, datadoe$tcellbin, datadoe$Humidbin)
write.csv(datadoe, file="9-3data.csv")
w= table(datadoe$bin)
doe= as.data.frame(w)
doe
# converted the new columns to a vector
vectornewcol<- as.vector(datanew$bin)
# sample for finding the count statistic of cell 121 ( for example)
count<- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j<- 1
i<- 1
while(i<12546)
{ if(vectornewcol[i]=="21")
{count[j]<- count[j]+1
j<- j+1
}

  else
  {
    while((j-1) >1)
    { count[j-2]<- count[j-2]-1


    j<- j-1
    }
    if(j!=1)
    {
      j<- 1

    }
  }
  i<- i+1
}
```

```
svmmodel<- read.csv("train2000.csv")
Testset <- read.csv("Supplied.csv")
svmmodel$Y<- ifelse(svmmodel$Y<0, 0, svmmodel$Y)
Testset$Y<- ifelse(Testset$Y<0, 0, Testset$Y)
Testset$fault<- ifelse(Testset$Y<0, 1, 0)
svmmodel$fault<- ifelse(svmmodel$Y<10, 1, 0)
tc <- tune.control(cross = 5)
model<- svm(Y~t.7+t.8+t.9+wind+temp+fault, cross=10, data= svmmodel, kernel= "radial", cost=1300, eps
predictedy<- predict(model, svmmodel)
rmse(svmmodel$Y, predictedy)
mae(svmmodel$Y, predictedy)
predict_test <- predict(model, Testset)
rmse(Testset$Y, predict_test)
mae(Testset$Y, predict_test)
tuned<- tune.svm(Y~t.7+t.8+t.9+temp+wind, svmmodel, cross=10, kernel= "radial", cost=1300, epsilon=0.
write.csv(predict_test, file="predictions2.csv")
```