A Novel Stochastic Dynamic Modeling for PV Systems Considering Dust and Cleaning

Armaghan Cheema[a], M. F. Shaaban[a,b]*, Mahmoud H. Ismail[a]

[a]*Department of Electrical Engineering, American University of Sharjah, Sharjah, 26666, UAE*
[b]*Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON, N2M2C7, Canada*

ABSTRACT

Stochastic photovoltaic (PV) modeling that can be used for long-term planning of power systems is essential for future renewable power generation. One of the most prevalent problems that PV systems face is the accumulation of dust on the PV panel surface that negatively impacts the output power. Wind speed along with other weather variables including relative humidity, temperature, and precipitation are some of the major factors that contribute to dust accumulation. This paper presents a novel dynamic model of the PV output power profile including the dust accumulation using a Markov chain model. The proposed model incorporates the seasonal variations in ambient temperature, solar irradiance, dust accumulation, and rate of dust accumulation as well as the desired cleaning frequency, which affect the overall energy yield of the PV system. The outcome of the model is virtually generated scenarios that can be used by the investors to decide on the optimal size of the PV system and the optimal cleaning frequency for each season. The model outcome shows an error of less than 5% when compared to actual data collected from the field without cleaning. This error can be reduced by increasing the number of states, which affects the computational time. Various case studies are presented to show the effectiveness of the proposed model and its benefits.

* *Corresponding author.* Tel.: +971-06515-4912.
E-mail address: mshaaban@aus.edu, mostafa.shaaban@uwaterloo.ca

1. INTRODUCTION

The recent growth in solar energy in the Middle East and specifically the United Arab Emirates (UAE) has led to numerous solar plants being built in the region. This includes the Shams Solar Power Station and the Masdar 10 MW Solar Photovoltaic (PV) Farm in Abu Dhabi and the 1000 MW Mohammed Bin Rashid Al Maktoum Solar Park in Dubai [1]. Other solar farms in the region include the Benban Solar Park in Egypt with a 1650 MW capacity and the Sakaka Solar Plant in Saudi Arabia with a 300 MW capacity [2]. Similarly, in the USA, there are solar farms like the Solar Star with a 579 MW capacity and the Topaz Solar Farm with a 550 MW capacity as well [2]. However, a common concern in almost every solar plant is the accumulation of dust on the solar panel surface, which negatively impacts its performance [3]. While numerous factors including aging, radiation, shading, temperature, and pollution impact the performance of a PV, dust is one of the most problematic issues out of all of them [4]. As the rate of dust deposition on a PV surface increases, the efficiency and power output of the module would decrease. In desert climates, dust accumulation reduces the overall power output of PVs by close to 40% on average over a year [5].

The dust accumulation problem is affected by many factors such as particle size [6], panel tilt angle [7], wind velocity [8], temperature and relative humidity. The relative humidity plays a key role in long-term soiling as 40-80% relative humidity increases adhesion to 80% due to the capillary forces that stick dust particles onto a solar panel surface [9].

Another major factor dictating the rate of dust accumulation on a solar panel is how close it is to an area where dust is highly likely to travel airborne. In dry climates [10], airborne dust particles find themselves on the surface of solar panels primarily due to adhesive forces. In wet conditions, on the other hand, dust particles adhere to the surface of the solar panel due to the presence of fog, rain, and snow. In [11], the dust deposition rates of Egypt and Colorado indicated that desert climates were likely to have far greater dust deposition rates close to 300 mg/m2.

Currently, manual cleaning and automated cleaning are the most common cleaning methods for PV that have been implemented around the world [12], [13]. However, as water is a scarce resource and continuous cleaning will be costly from labour and resources points of view, it is important to optimally clean solar panels while using minimal resources. Consequently, there is a need to model the PV output power considering dust accumulation and a need for a mechanism to incorporate different cleaning frequencies. With such a model, the optimal cleaning frequency for PVs can be determined, which can later be used for long-term investments.

The contribution of this paper thus involves the development of a novel model for the stochastic nature of PV performance including the effect of dust accumulation using meteorological data. The model includes a unique mechanism that depicts dust accumulation levels at different cleaning frequencies in the PV system.

The paper is organized as follows: Section II details the related work. Section III includes an overview of the proposed model, its different stages, and how the model can be used to develop virtual scenarios. Section IV discusses an application of the proposed model to develop the power profile of a PV. Section V analyzes the virtual scenarios for dust accumulation with varying cleaning frequencies followed by a validation of the overall model. Finally, Section VI discusses the implementation of the proposed model in a PV power plant for a specific case study and Section VII concludes the work.

2. RELATED WORK

The modeling of PV and dust accumulation has been carried out through numerous techniques depending on the complexity of the model and the assumptions made. When it comes to modeling PV energy production alone, it has often been performed by clustering daily values of solar irradiance together or by using monthly-hourly data to guarantee greater precision [14]. In most research works, solar irradiance is generally modeled statistically as a beta distribution [15]. At the same time, others have modeled global solar irradiance through exponential, Weibull, gamma, normal, lognormal, beta, or geometric distributions [16]. Other models have used solar irradiance and air temperature data for short-term forecasting using a power probability density function that is based on the Bayesian autoregressive time-series model [17].

However, an accurate model of PV output should consider the numerous weather factors that impact it and, in particular, dust accumulation. A common way of modeling dust accumulation levels, as seen in [18], is by directly using the PV output and predicting the level of dust that accumulates on the surface of the panel based on the performance. Other models rely on Monte Carlo simulations (MCS) for stochastically generating possible soiling profiles on a daily basis [19]. Other research proposed the use of a fixed rate for dust accumulation [20]-[23]. However, the rate of dust accumulation is never fixed as the weather variables surrounding the PV at any time can either increase or decrease the rate of dust accumulation that occurs. Dust accumulation modeling has also been performed using particulate matter concentration to help predict future precipitation patterns [24].

PV yield can also be used to model dust accumulation and PV soiling loss as in [25] using the stochastic rate and recovery (SSR) method, which is based on MCS. Dust accumulation rates along with soiling rates have also been modeled in [26] using the Theil-Sen estimator assuming a dry period of at least 14 days.

Overall, while the modeling of dust as a constant factor may fit certain research criteria, a more realistic approach would be to develop a dynamic model for dust accumulation. Consequently, there is a lack of literature on the dynamic modeling of dust

accumulation and its corresponding impact on PV performance. Moreover, the cleaning rate should be varied across the year as the dust accumulation is not fixed. Thus, the model should be able to incorporate varying cleaning rates.

3. PROPOSED MODEL

The proposed model in this work takes into consideration ambient temperature, solar irradiance, dust accumulation, and rate of dust accumulation and their influence on dust accumulation through a Markov chain model. This section provides a description to the proposed model, which is divided into three stages, as shown in Fig. 1. The outcome of this model can be described as virtual scenarios of ambient temperature, solar irradiance, and rate of dust accumulation based upon the dust accumulation state. The advantage of using virtual scenarios is that they include as many variations as possible of the weather variables mentioned earlier. While historical data of a random variable only represents one way of how the variable changes, the use of virtual scenarios allows the model to capture the stochastic nature of the random variable.

3.1. Stage 1: Data Acquisition and Processing

The PV power measurements and climate data used in this paper were for N_y years of hourly historical data from Arizona City in the United States of America. This included data regarding ambient temperature (T_{ambient}), solar irradiance (SI), dust accumulation (D_{acc}), and rate of dust accumulation (RDA) as all of these factors impact power. The PV measurements were made from a solar panel with a fixed tilt angle of 30 degrees. While other factors including wind speed, relative humidity, and precipitation are also valid weather variables to include, the data for dust accumulation would account for any changes in the previously mentioned three weather variables. Furthermore, RDA was calculated using the dust accumulation data by noting the change in dust levels at any given hour in comparison with the one succeeding it seen in (1). Additionally, as the panels were initially clean, dust levels began at zero. At this point, we have four data sets, which include T_{ambient}, SI, D_{acc}, and RDA.

$$RDA = \frac{D_{acc(t)} - D_{acc(t-1)}}{\Delta t} \tag{1}$$

To analyze and use the hourly historical data, the discretization of all data points was necessary so that it could then be used in the Markov chain model. Consequently, the most efficient way to discretize the data and cluster them based on the centroids was using the k-means clustering algorithm [30], which is based on the squared error function in (2).

$$\underset{S}{\arg\min} = \sum_{c=1}^{k} \sum_{x \in S_c} \| x - \mu_c \|^2 , \tag{2}$$

where k represents the number of centroids, c represents the index of the centroids, x represents the data, S represents the data divided into k sets and μ_c represents the mean of data points in subset S_c.

In this model, N_{RDA} is the number of states for the rate of dust accumulation, N_{SI} is the number of states for solar irradiance, N_{T} is the number of states for temperature, and N_{D} is the number of dust accumulation states. By increasing the number of states that the data is clustered into, the accuracy of the model increases but also the level of complexity increases, which is reflected in the computational time. The four data sets of T_{ambient}, SI, D_{acc}, and RDA were discretized after which the primary and secondary categorization is performed as follows.
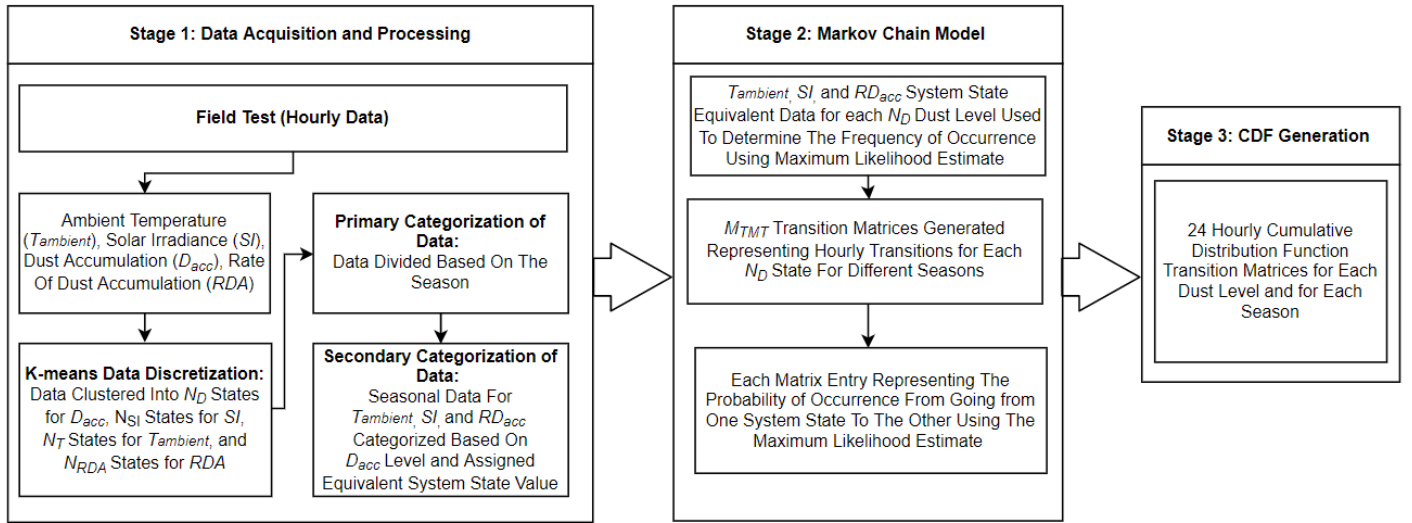
Fig. 1 - Overall system model flowchart.

3.1.1. Primary Categorization

The four data sets are then categorized according to the season such that the model could account for seasonal variations in ambient temperature, solar irradiance, dust accumulation, and rate of dust accumulation. Each data set is categorized as follows. The $T_{ambient}$ data set is organized in matrix $T = [T_{i,h}]$, the SI data set in matrix $SI = [SI_{i,h}]$, the D_{acc} data set in matrix $D = [D_{i,h}]$, and the RDA data set in matrix $RDA = [RDA_{i,h}]$. In each element of the above matrices, $i \in \mathcal{I}$, where i is the days in the original data and $\mathcal{I}$ is the original data set of days, and h is the hour of the day, which ranges from 1 to 24. When categorizing these data sets further according to the four seasons, the $T_{ambient}$ seasonal data is organized in a matrix $T^s = [T_{i,h}^s]$, the SI data set in a matrix $SI^s = [SI_{i,h}^s]$, the D_{acc} data set in a matrix $D^s = [D_{i,h}^s]$, and the RDA data set in a matrix $RDA^s = [RDA_{i,h}^s]$. In this case, sample $i \in \mathcal{I}_s \subset \mathcal{I}$, where $\mathcal{I}_s$ is the subset of days in each season (almost $90 \times N_y$ days) and $s \in \{1,2,3,4\}$ represents the seasons assuming that data is available for N_y years. At this point, there are a total of 16×24 data sets in total and 4×24 data sets for each season.

3.1.2. Secondary Categorization

The seasonal data sets $T_{ambient}$, SI, and RDA are then categorized further to N_D groups, as shown in Fig. 2, based on the dust level in the first hour of the day assuming that the dust level would not change for the rest of the day. By doing this, the $T_{ambient}$ seasonal data would then be represented in a matrix $T^{s,d} = [T_{i,h}^{s,d}]$, the SI data set in a matrix $SI^{s,d} = [SI_{i,h}^{s,d}]$, and the RDA data set in a matrix $RDA^{s,d} = [RDA_{i,h}^{s,d}]$. In these matrices, $d \in \{1,2,\dots,N_D\}$ would represent the dust level. At this point, there are $12 \times N_D$ data sets in total and $3 \times N_D$ data sets per season. In other words, for any day i in season s, the $T_{ambient}$, SI, and RDA data are categorized together if the dust level in the first hour of day i for that season matches the dust level in the D^s matrix. In this case, 25 data elements , which are made up of the 24 hours of the day in focus and the first hour of the following day from matrices T^s, SI^s, and RDA^s are categorized together to form matrices $T^{s,d}$, $SI^{s,d}$, and $RDA^{s,d}$, respectively. The first 24 points are necessary to understand the behavior of the different data sets within the specific day with respect to dust. In addition, the 25th point, which is the first hour of the next day, is necessary to transition to the next day and will be used in the Markov Chain Model explained in the next stage.

To analyze all the data together and not separately as three different matrices, it was important to put all the data under a single matrix. Hence, to generate an overall multi-state model, the three variables of temperature, solar irradiance, and rate of dust accumulation that affect the PV are combined into one matrix $M^{s,d} = [M_{i,h}^{s,d}]$. Each element of the matrix $M^{s,d}$ is composed of an element from each of the matrices $T^{s,d}$, $SI^{s,d}$, and $RDA^{s,d}$. The three matrix elements are then replaced with an equivalent system state value and stored in a single element in $M^{s,d}$. The total number of system states that describe all possible conditions should thus be

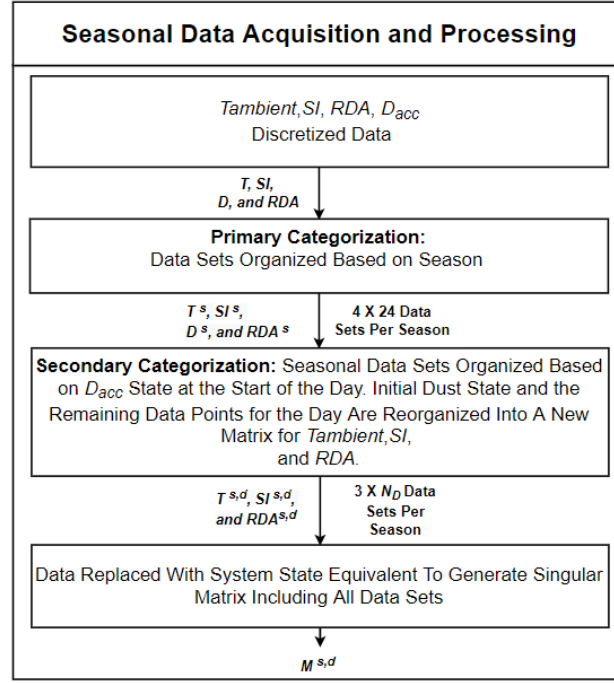$$N_{SYS} = N_{SI} \times N_T \times N_{RDA}. \tag{3}$$

Fig. 2 - Seasonal data organization.

Each data point is then assigned to one of the system states $1, \dots, N_{\text{SYS}}$ that would correspond to a specific state for $N_{\text{SI}}, N_{\text{T}}$, and N_{RDA}. By doing so, the N_{v} different weather variables, in this case four, could later be analyzed inside a single matrix. We assume there is no correlation between these random variables occurring at a certain hour of a certain season at the same dust level.

3.2. Stage 2: Markov Chain Model

While there are numerous stochastic modeling techniques used in research for PV modeling forecasting, one of the most common ones is the Markov chain model. While every forecasting model has its own advantages in solving complex real-world problems, the Markov process is commonly used when modeling dynamic stochastic systems and the state transitions that exist in complex stochastic systems. A Markov chain model is a discrete-time stochastic process that models how a random variable changes at discrete points of time. The Markov chain model has therefore been employed in this paper to model the behavior of numerous weather factors and their impact on dust accumulation to analyze the performance of a PV.

The discrete-time Markov chain $M(t)$ is a discrete-time stochastic process based on the idea that each time step t is occupied by one state E_μ in a series of states defined as $E_1, \dots, E_N$. Each of these states is defined stochastically on the basis of only the previous state and this satisfies the *Markov property*. In other words, the probability distribution of any state at any time step $t + 1$ is dependent on the state t and not dependent upon the previous states that lead to the state at time t. More importantly, the state transition that occurs between time step t and $t + 1$ is independent of time. The time steps involved in the entire process can be defined from $t = 1, \dots, T$ with $\mu = 1, \dots, N$ representing the index of the state the Markov chain is in. After that, the transition matrices must be generated. As the Markov process moves from time step t to the next time step at $t + 1$, the state of the process at time $t + 1$ can be determined from the state at time step t using the transition probabilities given as

$$P_{\mu v}(t) \equiv \text{Prob}\,(X_{t+1} = E_v \,|X_t = E_\mu). \tag{4}$$

It is important to note that (4) satisfies the Markov property. Using this, the transition matrices P, which are square matrices, can be generated with dimensions $N \times N$. For each N_{D} for dust accumulation, there are 24 different Markov transition matrices with each matrix representing the transition from a specific hour of the day to the following hour. As there are now 24 transition matrices for each dust level for each season, the total number of Markov transition matrices, M_{TMT}, can be calculated as follows:

$$M_{\text{TMT}} = 24 \times 4 \times N_{\text{D}.} \tag{5}$$

The transition matrices each have dimensions $N_{\text{SYS}} \times N_{\text{SYS}}$, with the columns and rows representing the different system map values. Each element within the transition matrix P represents the probability of state v occurring at time slot $t + 1$ given that the

previous state at time slot t is considered as μ. The way the probabilities for each matrix value were calculated is using the maximum likelihood (M_L) estimation, which can be found as

$$M_L = n_{\mu v} / \sum_p n_{\mu p},$$

(6)

where $n_{\mu v}$ is the number of transitions from state μ at the time instance t till time instance $t + 1$. The maximum likelihood estimate is used to calculate the probability, which is the frequency of occurrences divided by the total number of possible occurrences. To guarantee that each transition matrix was calculated accurately, the sum of each row for the transition matrices had be equal to 1 as it is an important characteristic of the Markov model.

3.3. Stage 3: Cumulative Distribution Function Generation

Using the transition probability matrices, the cumulative distribution function (CDF) could be created that would later be used for generating virtual scenarios. The CDF of a random variable X, when plotted, would form a staircase plot with the CDF of any random variable flat between x_k and x_{k+1}. The probability mass function (PMF), which is the data from the Markov transition matrices, is used to develop the CDF $F_X(x)$ as follows:

$$F_X(x) = \sum_{x_k \leq x} P_X(x_k).$$

(7)

3.4. Virtual Scenarios Generation

The output of the model can now be used to develop virtual scenarios showing dust accumulation levels across an entire year as seen in Fig. 3. The random virtual scenario instant is equal to the inverse CDF of a uniformly distributed random number between 0 and 1. Assuming there are $N_{scenarios}$ that are to be modeled, the dimensions of the virtual scenarios matrix will be $8760 \times N_{scenarios}$ with the columns representing each hour of the day for a complete year. This will be generated by using the inverse of the CDF generated from Stage 3. It is important to note that the first column of the virtual scenarios matrix was calculated differently than the rest of the columns as it is not based on any previous information before that first day. Therefore, for the first hour of the first day for each row of the virtual scenario matrix, the initial data was converted into a transition matrix without any separation based on correlating it with any level of dust accumulation. After that, it is converted into the system map, which is then systematically placed as the first column of the virtual scenario matrix. With the first column of the virtual scenario calculated, the remaining parts of the virtual scenarios were calculated with the information from the transition matrix CDF's and the first column of the virtual scenario matrix. Depending upon the value of the first number in the first column of the virtual scenario matrix, the corresponding row for the transition matrix CDF for hour one to hour two was calculated. To do this, a uniformly distributed random number between 0 and 1 was generated, and depending upon that value, the corresponding two numbers around that uniformly distributed random number were selected. For every uniformly distributed random number, there will be a number less than and greater than it in the transition matrix CDF. Between the two numbers, the number that is greater than it is chosen and selected as the second-row value in the second column of the virtual scenario matrix. This is because the values in the second column represent the transition from hour one to hour two of the first day that is virtually generated. A similar process is followed for the remaining hours of the day and for each day of the year.

An important aspect of the virtual scenario design is also the implementation of the desired cleaning pattern synthetically. Given the solar panels are cleaned every desired N_{clean} days throughout the year, the first column after the N_{clean} day utilizes the CDFs correposding to the clean dust level (the first dust state). Then, as the dust accumulation increases over the days, the corresponding transition matrix with a different dust level is chosen. After each hour of the virtual generation, the dust level is measured, using information from the rate of dust accumulation, to determine if the dust level has entered another category of N_d. If this occurs, a different hourly transition for a different dust state transition matrix is used to continue.

4. POWER PROFILE GENERATION MECHANISM

The proposed model can now be used to develop the power profile of a PV. In this section, a relationship between dust accumulation and the power output is developed and then applied to the virtual scenarios as shown in Fig. 4.
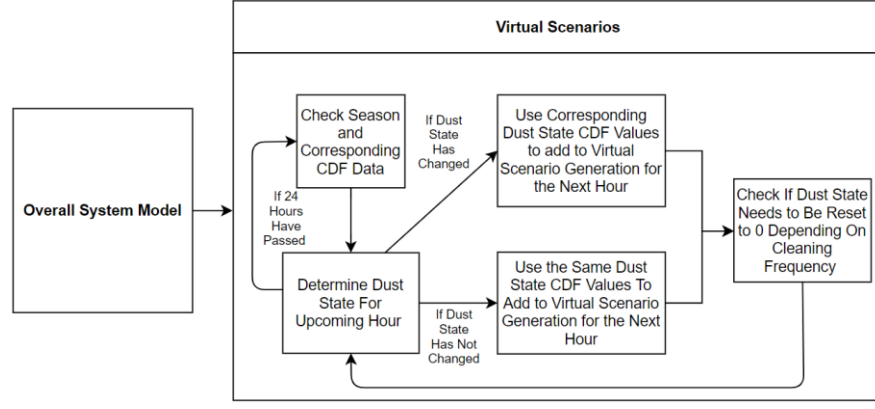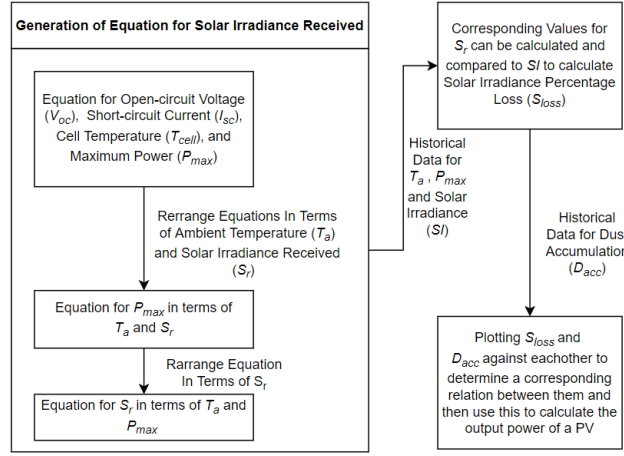
Fig. 3 - Application of overall model.

Fig. 4 - Dust accumulation and received solar irradiance relationship.

We start with the existing equations for the module short-circuit current (I_{sc}), module open-circuit voltage (V_{oc}), T_{cell}, and the PV maximum power output (P_{max}) in [28], which were modified as follows:

$$I_{sc} = I_{sc_{stc}} \times \left(1 + K_{isc} \times (T_{cell} - T_{stc})\right) \times \frac{S_R}{S_{stc}}, \tag{8}$$

$$V_{oc} = V_{oc_{stc}} \times \left(1 + K_{voc} \times (T_{cell} - T_{stc})\right), \tag{9}$$

$$T_{cell} = T_A + S_R \times \left(\frac{NOCT - T_{NOCT}}{S_{NOCT}}\right), \tag{10}$$

$$P_{max} = FF \times I_{sc} \times V_{oc} \times N_{PVModules}, \tag{11}$$

where $I_{sc_{stc}}$ represents the short-circuit current at standard test conditions, K_{isc} the temperature coefficient for I_{sc}, T_{cell} the temperature of the cell, T_{stc} the temperature at standard test conditions, S_R the solar irradiance received, S_{stc} the solar irradiance at standard test conditions, $V_{oc_{stc}}$ the open-circuit voltage at standard test conditions, K_v the temperature coefficient for V_{oc}, $NOCT$ the nominal operating cell temperature, T_{NOCT} the temperature at the nominal operating cell temperature, S_{NOCT} the solar irradiance at that specific nominal operating cell temperature, FF the fill factor, and finally, $N_{PVModules}$ the number of PV modules. Equations (8)-(10) were re-written so that they would only be two equations, one for I_{sc} and one for V_{oc}, and they would only be in terms of T_A and S_R. The reason why this is done is that every parameter in (8)-10) is a fixed variable based on any PV module characteristic and the only values that will change are T_A and S_R. In addition to that, dust accumulation results in a drop in solar irradiance received, which directly results in a drop in current and current directly impacts the overall power. Consequently, by re-writing (8)-(10) into two equations for I_{sc} and for V_{oc}, they can be substituted into (11) where the FF and $N_{PVModules}$ are also fixed variables, to determine P_{max}. The rearranged equation for P_{max} can be seen in (12) with C representing a constant that is generated during the re-arranging process.

$$P_{max} = T_A{}^2 \times S_R \times C. \tag{12}$$

At this point, there is a single equation to calculate P_{max} that is only in terms of T_A and S_R. Therefore, (12) can now be rewritten to solve for S_R with T_A and P_{max} being required for it as seen in (13).

$$S_R = \frac{P_{max}}{T_A{}^2 \times C}. \tag{13}$$

Using (13), historical data from N_y years regarding T_A and P_{max} can be acquired. In this case, P_{max} would represent the power that the solar panel is generating over the course of N_y years of historical data. The reason this information is relevant is that solar panels generally have built-in methods to determine the ideal solar irradiance available at a certain time of the day. However, when considering the existence of dust, the solar irradiance absorbed by the solar panel is not what the reference cell on the solar panel would suggest as it would be less.

Therefore, given P_{max} generated by a solar panel and the corresponding T_A for N_y years of historical data, the actual S_R that is unknown can be determined. After determining what S_R is, it can be compared to S_{ideal}, the optimal solar irradiance assuming no dust from a reference cell, to determine solar irradiance percentage loss S_{loss}, which can then be plotted versus dust accumulation for N_y years of historical data to determine a relationship between the two. This can be seen in Fig. 5.

Also, from Fig. 5, as dust accumulation increases, S_{loss} also increases. The best fit curve was determined after comparing the linear, quadratic, exponential, and quartic relation between the two data types. After determining the best fit curve to represent the relationship between S_{loss} and dust accumulation as the quartic relation, an overall relationship connecting dust accumulation and power output can be established. This relation is used to generate the power output in Fig. 6.

In Fig. 6, the power output of a PV for N_y years of historical data across a year can be seen and is represented by P_{actual}. It is important to note that in this specific case, there is no cleaning done and dust is expected to accumulate naturally on the panel overtime for the entire year. Also, P_{ideal} represents the solar panel power output assuming it receives maximum solar irradiance with no dust, which is recorded by the reference cell on the panel and calculated using (8)-(11). It is important to point out that the reference cell is not affected by the dust as the typical practice when taking measurements using the reference cell is to keep it covered and only uncover it when measurements of solar irradiance must be taken. Across the entire year, the average percentage difference is as high as 62.7% between P_{actual} and P_{ideal}. As expected, there is a huge difference between the actual power output of the panel and its ideal power output assuming no dust and no hindrance to overall solar irradiance. Using the relationship determined earlier between S_{loss} and dust accumulation, existing data for dust accumulation and maximum solar irradiance from the reference cell for the same panel can be used to determine the S_{loss} for any data set.

Now, using the calculated S_{loss}, S_R can be determined. Using S_R with T_A, P_{max} can then be calculated and is represented by P_{model} in Fig. 6. Across the entire year, the difference between P_{model} and P_{actual} is 2.08%, which highlights the accuracy of the relationships determined earlier between dust accumulation and S_{loss}. This also supports the fact that the overall relationship determined between dust accumulation and power output of a solar panel is accurate.

5. RESULTS AND ANALYSIS

This section explores how the dust behavior and dust accumulation models in the virtual scenarios can be used to determine optimal PV cleaning frequencies depending upon the season of the year.

5.1. Virtual Scenarios Analysis

We carried out investigations for two seasons; summer and winter, while varying the cleaning frequencies. The virtual scenarios that were generated focused on a 90-day period, which is close to a 3-month period that would account for a complete season. More specifically, the winter season was specified for 90 days after December 1st and the summer season would last for 90 days after June 1st. The five cleaning frequencies would range from cleaning every week to cleaning every five weeks.
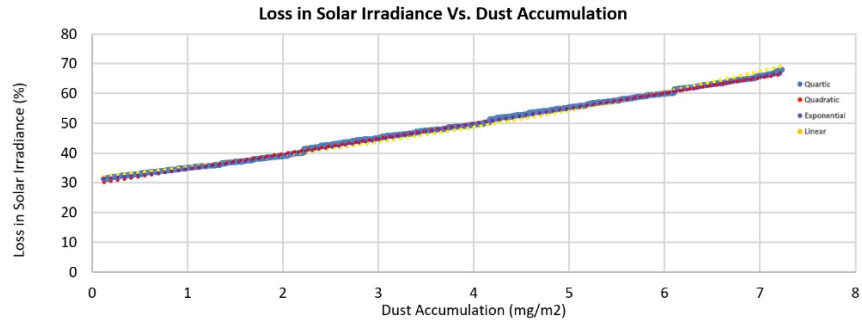
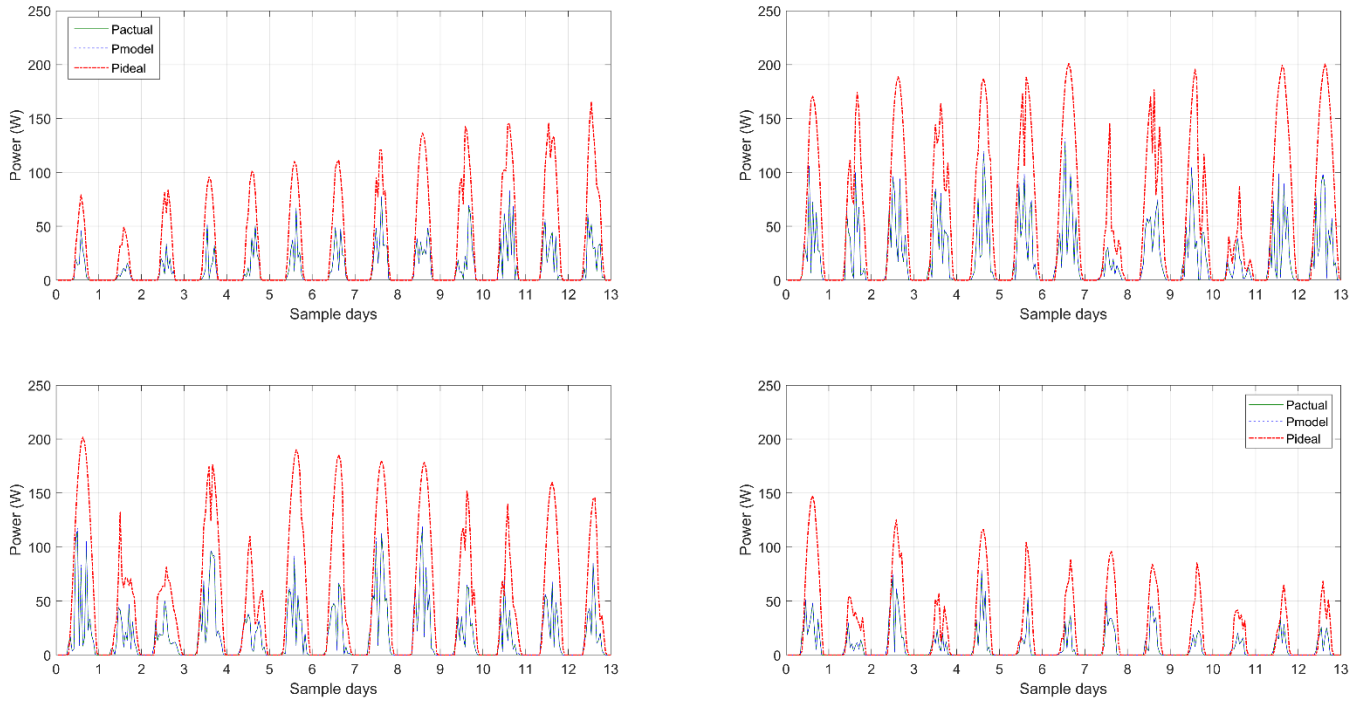Fig. 5 - Percentage loss in solar irradiance versus dust accumulation.

Fig. 6 - Power output of a solar panel under ideal, actual, and model conditions.
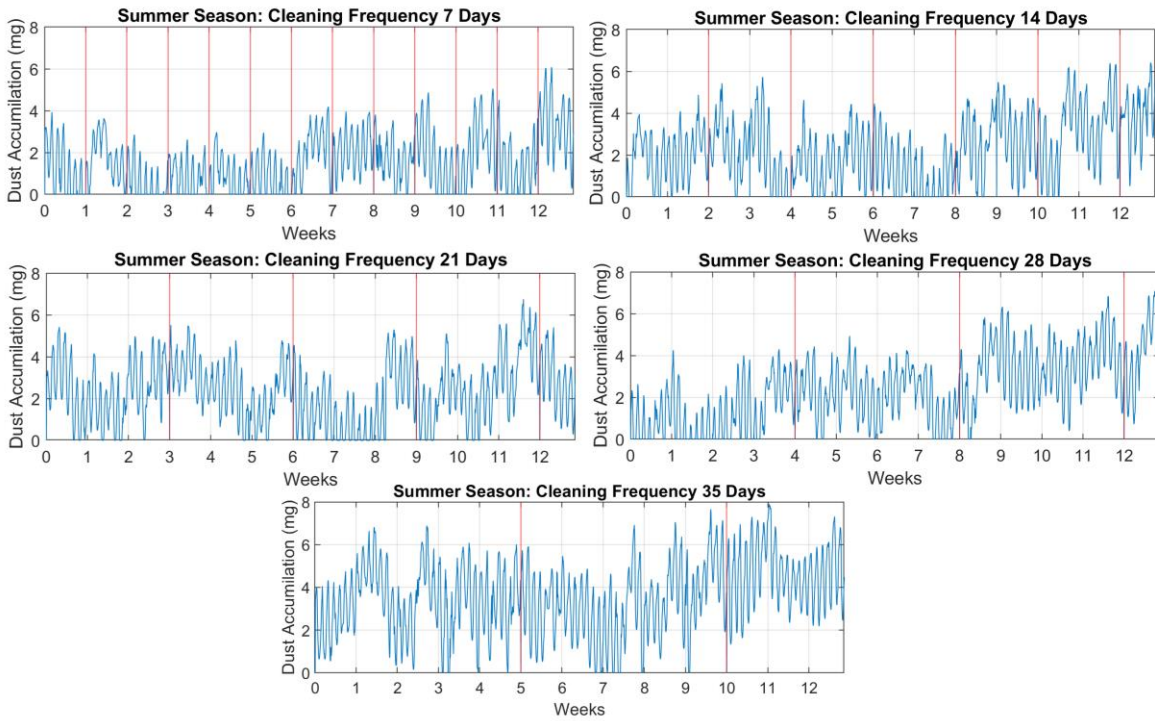
Fig. 7 - Virtually generated dust accumulation on solar panels during the summer season with five different cleaning frequencies.

A summary of the results for both the maximum and average dust levels across the summer and winter seasons for the five different cleaning frequencies can be seen in Tables 1 and 2 where the cleaning frequencies are compared. As seen from the Tables, the need to clean in the summer is far greater with dust levels reaching over 4 mg more than in winter. With cleaning every 5 weeks, the levels of dust during the summer reached a peak level of 7.9116 mg while, during the winter months, it reaches a maximum of 4.5734 mg. This indicates that the need for cleaning is far greater in the summer as the rate of dust accumulation is very close to double that in the winter. When comparing the average dust levels for both seasons, similar results can also be seen. The summer has an average dust level close to 3.6672 mg when cleaning occurs every five weeks while the winter has an average level of 1.4115 mg. Evidently, the average dust level during the summer is over double that in the winter, which indicates a far greater dust accumulation rate and a greater need for cleaning. When analyzing this further, it would take cleaning to occur every week for the average dust level to drop close to that when cleaning occurs every five weeks in the winter. If cleaning occurs every week in the summer, dust levels can drop significantly and the power losses that PVs face will be far less. In fact, the difference in the average dust level during the summer for every week and every five weeks is 1.7432 mg. When comparing this to the winter, the difference between the two extremes of cleaning frequencies is less than half at 0.6276 mg. Evidently, there is a far greater benefit of cleaning every week in the summer than there is in the winter as dust accumulation levels would decrease by close to 2 mg on average, which will make a huge difference to overall PV performance. Consequently, cleaning does not have to occur anywhere near as frequently during the winter season as the difference in average dust levels is far less.

5.2. Model Validation

The virtual scenarios generated depict results for dust accumulation throughout the year, which must be validated. To carry out the validation process, the mean value of dust accumulation for each hour of each season was gathered from the virtual scenarios and the N_y years of historical data. It is important to note that the generation of the virtual scenarios was done assuming no cleaning throughout the year.

TABLE 1 - Maximum dust levels across the summer and winter season

Cleaning Frequency	Summer	Winter
Every 7 Days	6.0785 mg	2.7246 mg
Every 14 Days	6.4206 mg	3.2258 mg
Every 21 Days	6.7535 mg	4.1370 mg
Every 28 Days	7.0889 mg	4.0536 mg
Every 35 Days	7.9116 mg	4.5734 mg

TABLE 2 - Average dust levels across the summer and winter season

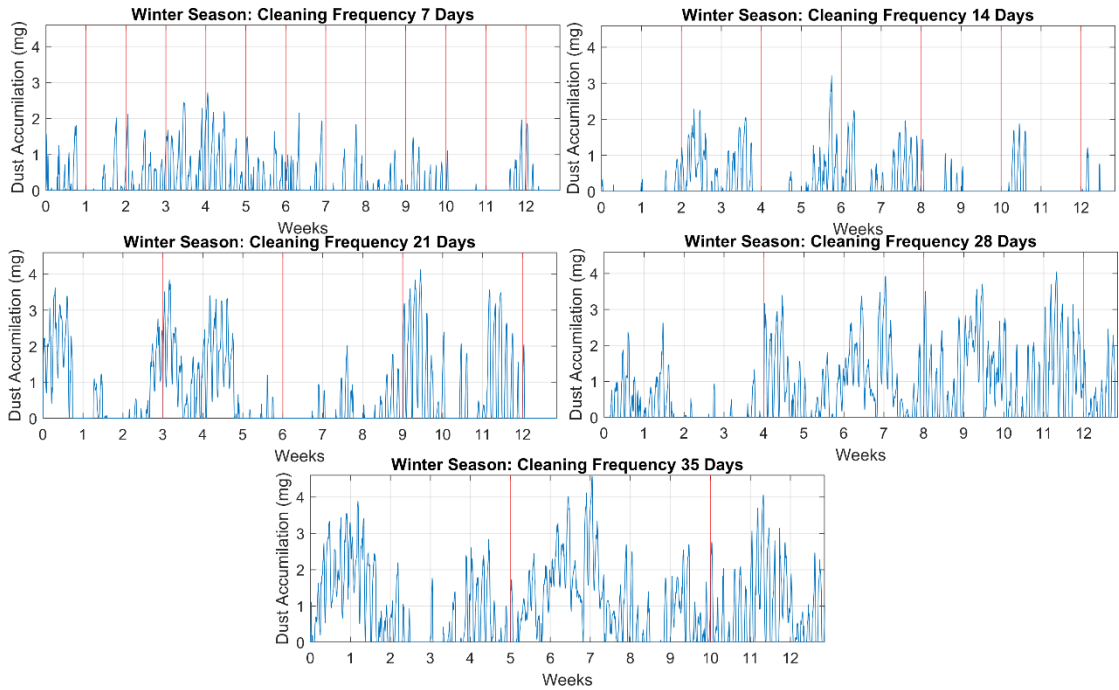| Cleaning Frequency	Summer	Winter
Every 7 Days	1.9240 mg	0.7839 mg
Every 14 Days	2.5347 mg	0.8267 mg
Every 21 Days	2.5704 mg	1.2431 mg
Every 28 Days	2.7221 mg	1.4115 mg
Every 35 Days	3.6672 mg	1.4792 mg

Fig. 8 - Virtually generated dust accumulation on solar panels during the winter season with five different cleaning frequencies.

This is because the existing N_y years of historical data is of dust accumulation with no cleaning. The mean values were then plotted together for an entire year and can be seen in Fig. 9.

As seen in Fig. 9, there is a seasonal variation between the mean values of the levels of dust accumulation. An important observation is the large increase in dust accumulation levels between the seasons as is visible from the spring to the summer seasons. The reason behind this large increase is because the first hour of summer represents the mean value of dust accumulation for every first hour of the day over the entire season. Consequently, there is no gradual increase in dust accumulation but a sudden one. In order to interpret the results of the mean values of dust accumulation for the virtual scenarios and the N_y years of historical data, the percentage difference between the mean values was calculated for the entire year and can be seen in Table 3. Over the course of the year, the average percentage difference was 4.57% and the maximum percentage difference was 9.78%, which highlights the accepTable accuracy of the model. In addition, the standard deviation was also calculated for the virtual scenarios and the N_y years of historical data. Using the standard deviation, the coefficient of variation (CV) was calculated as

$$CV = \frac{\sigma}{\varphi}, \tag{14}$$

where σ represents the standard deviation and φ represents the mean. For the virtual scenario generated, a standard deviation of 0.078 and a mean of 0.457 gave a CV of 17.06%, which is an accepTable 5alue. Hence, there is not a massive dispersion of the data around the mean and, instead, the estimates from the modeled virtual scenarios are precise.

6. PV SYSTEM SIZING USING THE PROPOSED MODEL

The proposed model can be implemented to size PV power plants, where the proposed model will impact economical aspects severely. To highlight this impact, a case study was developed for a 100-MW PV power plant. The parameters that were varied in this case study was the cleaning frequency and the method of cleaning. There were five different cleaning frequencies, which included daily, weekly, biweekly, monthly, and no cleaning. In terms of the method of cleaning, there was automated and manual cleaning. Table 4 summarizes the parameters of the case study that were used to get a cost estimate. All costs related to cleaning were acquired from an existing case study in [29].

6.1. Cleaning Cost Per Cycle for Different Cleaning Methods

The first step to calculate the cost estimate of cleaning was to determine the overall cost of a single cleaning cycle for the entire PV power plant assuming the different cleaning methods. Starting with manual cleaning, the cost of water, labor, and other materials was calculated to reach a total sum of $17,921 per cleaning cycle. It is important to note that in this calculation, there is no inclusion of any capital cost as this is not automated cleaning. For automated cleaning, costs were divided into either running or capital. With the running costs calculated first, the total automated running costs when including water, labor, and other materials is $2,344 while the automated total capital cost is $57,620.
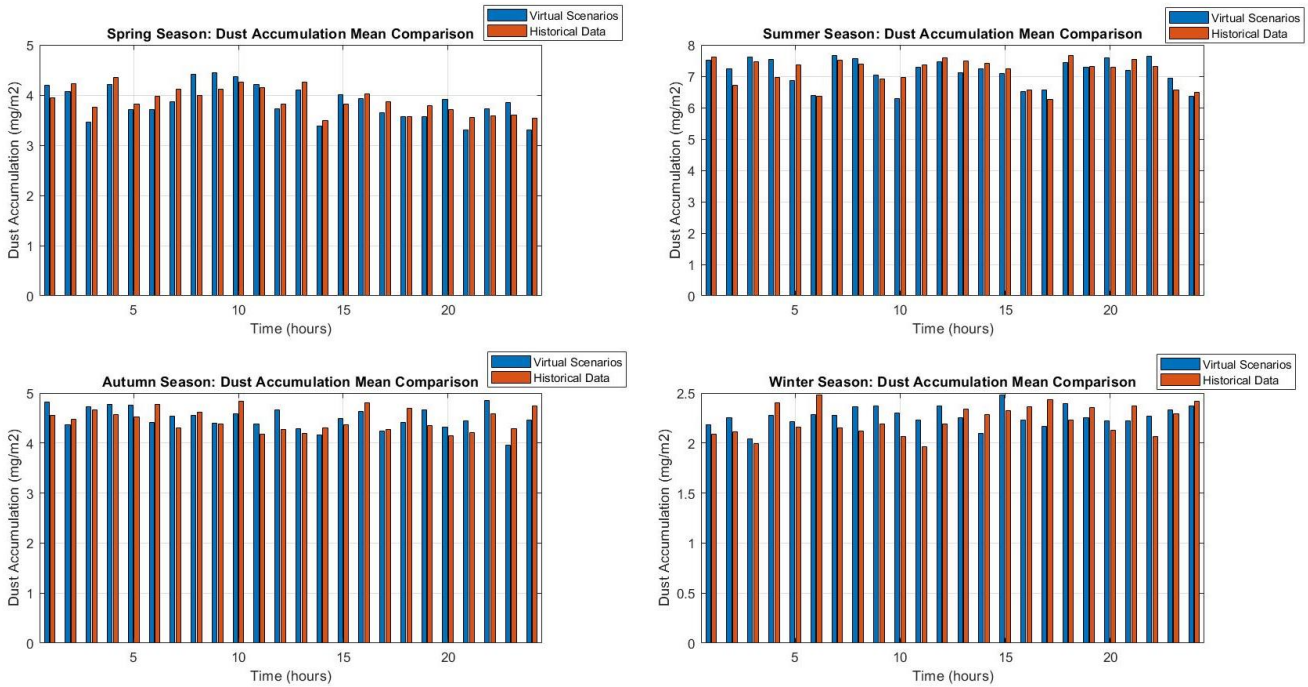

Fig. 9 - Dust accumulation mean comparison for different seasons.

TABLE 3 – Maximum and minimum percentage difference across all seasons

Seasons	Maximum Percentage Difference (%)	Minimum Percentage Difference (%)
Spring	9.78	0.27
Summer	8.34	0.41
Autumn	9.33	0.47
Winter	9.15	1.92

TABLE 4 - Parameters for cleaning a 100 mw PV power plant

Parameter	Value	Parameter	Value
Power Plant Size (MW)	100	Capital Equipment ($)	90,000
Power Plant Area (m^2)	717,949	Consumable & Maintenance Cost ($/hr)	7
Module efficiency	14%	Allocated Capital Cost ($/hr)	3.12
Cost of Water ($/liter)	0.0024	Automated Cleaning Rate (m^2/hr)	4460
Water Consumption (liter/m^2)	0.5	Discount rate	5%
Time to Clean 1 Panel (min)	0.5	Escalation rate	1%
Labor Rate ($/hr)	4.45	Loan period (yrs)	20
Cost of Other Materials ($/m^2)	0.0053		

6.2. Levelized Cleaning Costs

To calculate the yearly cleaning cost depending on the different frequencies, the levelized cleaning cost of the different cleaning frequencies must first be calculated. To do that, the levelizing factor (LF) was first calculated using

$$LF = \left(\frac{(1+d')^{n_{loan}}-1}{d'(1+d')^{n_{loan}}}\right) \times \left(\frac{d(1+d)^{n_{loan}}}{(1+d)^{n_{loan}}-1}\right), \tag{15}$$

$$d' = \frac{d-e}{1+e}, \tag{16}$$

where d represents the nominal discount rate, and d' is the effective discount rate with escalation, e the escalation rate, and n_{loan} the loan term. The LF can now be multiplied with the total cleaning cost per cycle for both cleaning types and with the cleaning frequency as well. The results of the levelized cleaning cost based on cleaning frequency for a year can be seen in Table 5. As seen in Table 5, the cost for using automated cleaning is far cheaper no matter what the cleaning frequency is. Consequently, only the automated cleaning option will be used in further calculations.

TABLE 5 - Levelized cleaning cost based on cleaning frequency

Cleaning Frequency	Manual Cleaning	Automated Cleaning
Daily	$7,158,522	$994,074
Weekly	$1,019,844	$191,033
Biweekly	$509,922	$124,327
Monthly	$235,349	$88,408

6.3. Cost Analysis

With the costs of automated cleaning for different cleaning frequencies acquired, the proposed model can now be simulated and sized to fit a 100-MW PV power plant. However, the model will also vary the cleaning frequency for an entire year and determine the average energy production for different seasons. A summary of the results can be seen in Table 6. With the cost of electricity assumed at $60 per MWh, the levelized annual cost per season can be calculated without any cleaning cost and can be seen in Table 7. Using the levelized cleaning costs based on the cleaning frequency calculated earlier, the total net profit including cleaning costs can be seen in Table 8.

As seen in Table 8, depending on the season, certain cleaning frequencies would be more profiTable. In this specific case with the cost of electricity set at $60 per MWh, weekly cleaning would be the ideal cleaning frequency in spring, daily cleaning in the summer and autumn season, and weekly cleaning again in the winter season.

TABLE 6 - Average energy in MWh for different seasons depending on cleaning frequency

Cleaning Frequency	Spring	Summer	Autumn	Winter
Daily	18,232	35,683	34,451	13,138
Weekly	16,562	32,193	30,943	11,860
Biweekly	15,419	30,642	29,276	11,203
Monthly	14,336	28,500	26,867	10,351
None	12,815	24,742	23,845	8,946

TABLE 7 - Levelized annual cost per season in US Dollars

Cleaning Frequency	Spring	Summer	Autumn	Winter	Total
Daily	1,197,187	2,343,047	2,262,175	862,665	6,665,076
Weekly	1,087,494	2,113,939	2,031,824	778,766	6,012,025
Biweekly	1,012,487	2,012,075	1,922,365	735,654	5,682,583
Monthly	941,383	1,871,421	1,764,197	679,668	5,256,670
None	841,509	1,624,618	1,565,752	587,424	4,619,305

TABLE 8 - Total net profit in US Dollars

Cleaning Frequency	Spring	Summer	Autumn	Winter	Total	Percentage Increase Compared to No Cleaning
Daily	948,669	2,094,528	2,013,657	614,147	5,671,002	22.8%
Weekly	1,039,736	2,066,181	1,984,066	731,007	5,829,992	26.2%
Biweekly	981,406	1,980,994	1,891,284	704,572	5,558,256	20.3%
Monthly	919,281	1,849,319	1,742,095	657,566	5,168,262	11.9%
None	841,509	1,624,618	1,565,752	587,424	4,619,305	0%
Maximum (Fixed Cleaning Schedule)	1,039,736	2,066,181	1,984,066	731,007	5,829,992	26.2%
Maximum (Varying Cleaning Schedule)	1,039,736	2,094,528	2,013,657	731,007	5,878,930	27.3%

If the same cleaning frequency was followed for every season, profits would not be maximized. The final row of Table 8 indicates the maximum values for each of the seasons to get the highest net total profit. The percentage difference between the highest net total profit by choosing the optimal cleaning frequency in different seasons and no cleaning is the most at 27.3%. By keeping the same cleaning frequency for the entire year, the percentage difference with no cleaning frequency is 22.8%, 26.2%, 20.3%, and 11.9%, for daily, weekly, biweekly, and monthly cleaning, respectively. In monetary terms, the loss in profit when not cleaning when compared to implementing the optimum cleaning frequency depending on the season is estimated at $1,259,625. Even by keeping monthly cleaning for the entire year, the loss amount is as large as $710,667. Hence, there is a need to optimally choose the ideal cleaning frequency depending upon the season and not use the same cleaning frequency for the entire year. Overall, the maximum total net profit is an accepTable and realistic amount for a 100-MW PV power plant. In addition, depending upon the cost of electricity, the recommended cleaning frequency would change depending upon the season to maximize profit. To understand this further, a summary of the result of varying the cost of electricity starting from $40 per MWh going up in steps of $20 per MWh until $160 per MWh can be seen in Table 9.

As seen in Table 9, the recommended cleaning frequency when the price of electricity was $40 was weekly cleaning for all seasons throughout the year for maximum profit. As the price of electricity increased, the recommended cleaning frequencies for the summer and autumn season became daily. This pattern remained the same until the cost of electricity was $120 where the spring recommended cleaning frequency became daily. Finally, at an electricity cost of $160, the cleaning frequency for all seasons became daily. More importantly, when comparing the percentage increase in profit when compared to no cleaning for the specified cost of electricity in Table 9, it was 23.9%, 27.3%, 30.5%, 32.4%, 33.9%, 35.1%, and 36.2%. The importance of percentage increase in profit when there is a varying cleaning frequency based on the proposed model can be seen when comparing it to the percentage increase in profit when there is a fixed cleaning frequency in Table 10. For every cost of electricity, the varying cleaning frequency allows for equal or a greater percentage increase in profit with there being

no cases where having a fixed cleaning frequency is more profiTable than a varying one.

In addition to the above, Table 8 indicates how the 100-MW PV plant would have varying degrees of financial benefit depending on the cleaning frequency they use over an entire year. As the objective of the case study was to vary the cleaning frequencies in a 100-MV PV plant, the benefits of implementing the proposed model can be seen from a financial point of view. If the 100-MW PV farm was to maintain the same cleaning frequency with electricity costing $60 per MWh for an entire year, it would not maximize its net profit. With a weekly cleaning frequency, the 100 MW PV farm would make $5,829,992 total net profit. Similarly, with a daily, bi-weekly, monthly, or no cleaning, the PV farm would make $5,671,002, $5,558,256, $5,168,262, and $4,619,305, respectively. However, after implementing the model, Table 8 indicates that by varying the cleaning frequency in different seasons, a higher total net profit can be made. In this case, if the spring season is cleaned on a weekly basis, the summer and autumn seasons cleaned daily, and the winter season cleaned on a weekly basis, a higher total net profit of $5,878,930 can be achieved.

TABLE 9 – Recommended cleaning frequency for different seasons for varying Feed-in tariff (FiT)

Cost of Electricity ($/MWh)	Recommended Cleaning Frequency to Maximize Profit (Variable Frequency)					
	Spring	Summer	Autumn	Winter	Profit ($)	Percentage Increase Compared to No Cleaning
40	Weekly	Weekly	Weekly	Weekly	3,816,984	23.9
60	Weekly	Daily	Daily	Weekly	5,878,930	27.3
80	Weekly	Daily	Daily	Weekly	8,036,091	30.5
100	Weekly	Daily	Daily	Weekly	10,193,253	32.4
120	Daily	Daily	Daily	Weekly	12,369,040	33.9
140	Daily	Daily	Daily	Weekly	14,562,765	35.1
160	Daily	Daily	Daily	Daily	16,779,463	36.2

TABLE 10 - Fixed cleaning frequency for different seasons for varying Feed-in tariff (FiT)

Cost of Electricity ($/MWh)	Recommended Cleaning Frequency to Maximize Profit (Fixed Frequency)					
	Spring	Summer	Autumn	Winter	Profit ($)	Percentage Increase Compared to No Cleaning
40	Weekly	Weekly	Weekly	Weekly	3,816,984	23.9
60	Weekly	Weekly	Weekly	Weekly	5,820,992	26.0
80	Weekly	Weekly	Weekly	Weekly	7,825,001	27.0
100	Daily	Daily	Daily	Daily	10,114,386	31.4
120	Daily	Daily	Daily	Daily	12,336,089	33.5
140	Daily	Daily	Daily	Daily	14,557,771	35.1
160	Daily	Daily	Daily	Daily	16,779,463	36.2

7. CONCLUSION

There has been a need for modeling dust accumulation as a function of time and not a fixed variable in order to achieve higher accuracy. The proposed work is able to achieve this by introducing the impact of ambient temperature, solar irradiance, dust accumulation, and rate of dust accumulation to better mimic the realistic behavior of dust accumulation. The virtual scenarios generated were able to show that dust accumulation would increase at certain times and decrease as well depending upon the aforementioned weather variables in different seasons of the year. As a result, the paper was able to address what was lacking in literature today by modeling dust as a function of time which is a major benefit of this work. Additionally, another important outcome of the proposed model is determining the optimum cleaning frequency for a PV farm to maximize net profit. After the proposed model was tested a case study for a 100-MW PV power plant, results showed that depending upon the cost of electricity, the need for cleaning at different frequencies could profit PV power plants more or less. By allocating all resources towards automated cleaning, as it was shown to be more economical than manual cleaning, a PV power plant could maximize profit by determining at which frequency they would want to clean in a specific season. Furthermore, the paper also developed an accurate relationship between the output power of a PV and dust accumulation.

In terms of recommendations for future work, it is important to note that the proposed model only included data for solar irradiance, ambient temperature, dust accumulation, and rate of dust accumulation. While other factors including relative humidity, wind speed, and precipitation are also important for dust accumulation, their information was assumed to be embedded within the data for dust accumulation as noted in Section III. Consequently, including separate data for relative humidity, wind speed, and precipitation along with dust accumulation would further improve the accuracy of the model. In addition, other factors including dust particle size, PV tilt angle, and pollutants in the environment can also be taken into consideration as they also dictate the behavior of dust accumulation. Therefore, by including the aforementioned factors affecting the behavior of dust, the proposed model would more accurately depict dust accumulation on a PV. The resulting output of the improved model would better decide when to optimally clean large PV systems that suffer greatly from dust like in the Middle East depending on the season.

REFERENCES

[1] D. Llamas, "Dubai unveils 1,000 MW solar energy park", HELIOSCSP, 2020. Accessed on: November 13, 2020. [Online]. Available: http://helioscsp.com/dubai-unveils-1000-mw-solar-energy-park/.

[2] S. Chakrabarti, "Top 19 Biggest Solar Farms in the World | SolarFeeds Marketplace", *SolarFeeds Marketplace*, 2020. [Online]. Available: https://solarfeeds.com/solar-farms-in-the-world/.

[3] K. Narimane, D. Ahmed, B. Khadidja, D. M. Bilal, B. Halima and A. Gougui, "In-situ Investigation the Effect of Dust Deposition on the polycrystalline silicon photovoltaic modules in Ouargla region," *2018 International Conference on Communications and Electrical Engineering (ICCEE)*, El Oued, Algeria, pp. 1-5, 2018.

[4] M. Piliougine, C. Cañete, R. Moreno, J. Carretero, J. Hirose, S. Ogawa, and M. Sidrach-de-Cardona, "Comparative analysis of energy produced by photovoltaic modules with anti-soiling coated surface in arid climates," Appl. Energy, vol. 112, pp. 626–634, 2013.

[5] A. Ahmed and T. Massier, "Impact of Climate Change and High PV Penetration on Power Factor Profile," 2019 IEEE Industry Applications Society Annual Meeting, Baltimore, MD, USA, pp. 1-6, 2019.

[6] D. L. Alvarez, A. S. Al-Sumaiti and S. R. Rivera, "Estimation of an Optimal PV Panel Cleaning Strategy Based on Both Annual Radiation Profile and Module Degradation," in IEEE Access, vol. 8, pp. 63832-63839, 2020.

[7] Å. Skomedal, H. Haug and E. S. Marstein, "Endogenous Soiling Rate Determination and Detection of Cleaning Events in Utility-Scale PV Plants," in IEEE Journal of Photovoltaics, vol. 9, no. 3, pp. 858-863, 2019.

[8] Karaki, S.H., Chedid, R.B., Ramadan, R.: 'Probabilistic performance assessment of autonomous solar-wind energy conversion systems', IEEE Trans. Energy Convers., vol. 14, no. 3, pp. 766–772, 1999.

[9] Arthur, Y.D., Gyamfi, K.B., Appiah, S.K.: 'Probability distributional analysis of hourly solar irradiation in Kumasi-Ghana', Int. J. Bus Soc. Res., vol. 3, no. 3, pp. 63–75, 2013.

[10] Dissou, A.Y., Tabi, O.F.: 'Monthly cluster of hourly solar irradiation in Kumasi-Ghana', Int. J. Bus. Soc. Sci., vol. 3, pp. 123–131, 2012.

[11] Jeantya, P., Delsaut, M., Trovaleta, L., et al.: 'Clustering daily solar radiation from reunion island using data anal. methods'. Proc. Int. Conf. Renewable Energies and Power Quality, Bilbao, Spain, pp. 1-6, 2013.

[12] C. Miqdam, "Impact of Some Environmental Variables with Dust on Solar Photovoltaic (PV) Performance: Review and Research Status," International J of Energy and Environment, vol. 7, pp. 152-159, 2013.

[13] M. Saidan, A. Albaali, E. Alasis and J. Kaldellis, "Experimental study on the effect of dust deposition on solar photovoltaic panels in desert environment", Renewable Energy, vol. 92, pp. 499-505, 2016.

[14] Bracale, A., Caramia, P., Carpinelli, G., et al.: 'A Bayesian method for shortterm probabilistic forecasting of photovoltaic generation in smart grid operation and control', Energies, vol. 6, no. 2, pp. 733–747, 2013.

[15] Lai, C.S., Jia, Y., McCulloch, M.D., et al.: 'Daily clearness index profiles cluster analysis for photovoltaic system', IEEE Trans. Ind. Inf., vol. 13, no. 5, pp. 2322–2332, 2017.

[16] M. G. Deceglie, L. Micheli and M. Muller, "Quantifying Soiling Loss Directly From PV Yield," in IEEE Journal of Photovoltaics, vol. 8, no. 2, pp. 547-551, 2018.

[17] L. Micheli, M. Muller, and S. Kurtz, "Determining the effects of environment and atmospheric parameters on PV field performance," in Proc. IEEE 43rd Photovolt. Specialist Conf., pp. 1724–1729, 2016.

[18] L. Micheli, M. G. Deceglie, and M. Muller, "Predicting photovoltaic soiling losses using environmental parameters: An update," Prog. Photovolt. Res. Appl., vol. 27, no. 3, pp. 210–219, 2019.

[19] M. G. Deceglie, L. Micheli, and M. Muller, "Quantifying year-to-year variations in solar panel soiling from PV energy-production data," in Proc. IEEE 44th Photovolt. Spec. Conf., pp. 2804-2807, 2017.

[20] P. K. Sen, "Estimates of the regression coefficient based on Kendall's tau," J. Amer. Statist. Assoc., vol. 63, no. 324, pp. 1379–1389, 1968.

[21] J. Tanesab, D. Parlevliet, J. Whale, and T. Urmee, "The effect of dust with different morphologies on the performance degradation of photovoltaic modules," Sustainable Energy Technologies and Assessments, vol. 31, pp. 347–354, 2019.

[22] A. Pan, H. Lu, and L.-Z. Zhang, "Experimental investigation of dust deposition reduction on solar cell covering glass by different self-cleaning coatings," Energy, vol. 181, pp. 645–653, 2019.

[23] H. Lu and W. Zhao, "Effects of particle sizes and tilt angles on dust deposition characteristics of a ground-mounted solar photovoltaic system," Applied Energy, vol. 220, pp. 514–526, 2018.

[24] H. Theil, "A rank-invariant method of linear and polynomial regression analysis," in Henri Theil's Contributions to Economics and Econometrics, vol. 1, no. 2, pp. 345–381, 1992.

[25] L. Zhu et al., "Spatiotemporal characteristics of particulate matter and dry deposition flux in the cuihu wetland of beijing," PLOS ONE, vol. 11, no. 7, pp. 1-16, 2016.

[26] G. Sehmel and S. L. Sutter, "Particle deposition rate on a water surface as a function of particle diameter and air velocity," J. De Recherches Atmospheriques, vol. 8, pp. 911–920, 1974.

[27] D. Rajashreee and D. Rasmita, "Comparative Analysis of Supervised and Unsupervised Discretization Techniques," in *International Journal of Advances in Science and Technology*, vol. 2, no. 3, pp. 29-37, 2011.

[28] A. S. Al-Sumaiti, M. H. Ahmed, S. Rivera, M. S. El Moursi, M. M. A. Salama and T. Alsumaiti, "Stochastic PV model for power system planning applications," in *IET Renewable Power Generation*, vol. 13, no. 16, pp. 3168-3179, 2019.

R. K. Jones et al., "Optimized Cleaning Cost and Schedule Based on Observed Soiling Conditions for Photovoltaic Plants in Central Saudi Arabia," in IEEE Journal of Photovoltaics, vol. 6, no. 3, pp. 730-738, 2016.