A Probabilistic Estimation of PV Capacity in Distribution Networks from Aggregated Net-load Data

Lewis Waswa¹, Munyaradzi Justice Chihota¹, and Bernard Bekker¹
¹Department of Electrical Engineering, Stellenbosch University, Stellenbosch, 7600, South Africa

Corresponding author: Lewis Waswa (e-mail: waswa@sun.ac.za)

This work was supported by the Centre for Renewable and Sustainable Energy Studies, Stellenbosch, South Africa.

ABSTRACT Globally, solar photovoltaic (PV) installations on distribution LV feeders have increased significantly. The increased penetration leads to several technical problems on existing networks and impacts utilities’ business models as energy sales drop. For proactive management of these challenges, utilities need to continually monitor the capacity of installed PV. To this end, some utilities typically require PV installations to be registered and sometimes use GIS mapping to approximate the installed PV capacity. However, these GIS PV capacity estimations methods are unreliable. Therefore, to obtain reliable PV capacity estimates at a distribution level, comprehensive modeling is required to accurately represent the generation output and the distribution load. This paper proposes a novel probabilistic PV estimation method that uses time-series historical sets of load and irradiance data to estimate the embedded PV capacity while considering the uncertainty in solar irradiance and the measured net load. Uncertainty characterization is implemented using empirical probability density functions, and simulation is performed stochastically using Monte-Carlo methods. A novel quantile analysis approach is developed and used in the computation of the final PV estimates. The proposed methodology is tested using measured load data from Ausgrid customers, Australia, and achieves reasonable accuracy (between 86% and 90% for the tested cases) and a 10% mean absolute percentage error. This approach is robust to the effects of input uncertainty and can be used by distribution utilities to estimate and monitor PV capacity installed on the distribution networks without incurring extra advanced metering investment costs.

INDEX TERMS Disaggregation, Monte-Carlo Simulation, photovoltaic systems, probability density function, quantile analysis

I. INTRODUCTION

A. CONTEXT

The global uptake of solar photovoltaics (PV) is reported to be on the rise due to various factors including the need to mitigate climate change [1]. Research indicates that increasing PV penetration may result in several technical challenges such as increased voltage rise as well as thermal loading issues [2]. These challenges are more difficult to control on low voltage (LV) networks in comparison to medium and high voltage networks. This is because most LV networks are not remotely monitored. As such, it is difficult to determine the network’s condition, particularly the contribution from PV, at any given time. Therefore, maintaining optimal network operational performance within the required constraints often comes down to regulating the amount of PV, i.e., PV penetration, on a specific LV network.

Various strategies are employed to regulate PV penetration [3], [4]. In South Africa, for example, PV penetration limits are regulated in two ways. Firstly, the PV system size is restricted based on each connected customer’s nominal maximum load, determined by their supply-side circuit breaker size. Secondly, the PV system size is restricted based on the ratio of the aggregated installed solar PV to the maximum MV feeder load and the capacity of the supplying transformer [3]. The underlying approach to these regulations is to limit the aggregate PV capacity below the respective feeder’s hosting capacity. Adequate knowledge of the
installed aggregated PV capacity, therefore, provides a vital indicator of a LV feeder’s operating conditions based on regulated penetration limits. This information is critical in the early mitigation of potential faults, low quality of supply, or thermal overloads [4].

Most utilities require customers to register their installations to gather information on the total installed PV capacity. However, the records of installations may not be reliable as some customers might not follow stipulated registration processes or may install additional capacity after registration [5]. A more reliable approach to gathering the relevant information is to deploy smart devices, but this requires widespread installation of smart devices and the supporting technology for live monitoring, which is costly for both the utility and customers [6]. Given the challenges with registrations and live monitoring, it is important to develop reliable methods that can inform the distribution network operators (DNOs) on the capacity of PV on their networks.

The following section looks at the existing literature on the PV estimation processes.

B. LITERATURE REVIEW

This section provides a broad literature review of PV estimation methods. These methods can be broadly classified into geographical information systems (GIS) and disaggregation methods. The following sub-sections examine these two classes and their application.

1) GEOGRAPHICAL INFORMATION SYSTEMS METHODS

GIS methods have been applied in various studies, including mapping useful lands, urban planning, and mapping energy resources [7]. The most widely used GIS method in PV capacity estimation and related studies is remote sensing [8]. This method has been applied in the assessment and automatic detection of potential available roof space for PV installation [8], [9], and the estimation of the number of PV panels already installed within an area [7]. These reported GIS methods cannot directly estimate the PV capacity embedded in feeder load for the following reasons. Firstly, the collected imagery requires complex, data-intensive post-processing to detect PV panels [10], [11]. Secondly, this method is prone to detection errors, which coupled with the fact that the panel rating is unknown, leads to potentially significant estimation errors [7]. Finally, GIS methods do not provide accurate information as to whether the customer assessed is connected to the network or not. From this discussion, it can be deduced that GIS methods are useful in the determination of the spatial distribution of rooftop PV in an area. However, the methods cannot determine the PV capacity on a feeder.

2) DISAGGREGATION METHODS

Load disaggregation is a method by which composite load data could be broken down to obtain the contribution of different sources of energy to the aggregate signal [12]. In contrast to the GIS methods, disaggregation models have been used to estimate the PV capacity installed at the customer level using customer load data. These methods have also been used to inform other areas of energy informatics.

Disaggregation methods, as formulated in [13] and discussed in [14], have been used in a range of applications with different objectives. These objectives include household demand forecasting [6], determination of PV site power output [15], real-time estimation of PV generation from active power flow measurements [16] as well as detection, verification, and estimation of customer PV capacity from customer load data [17]. Similarly, in [13], the method is used in the separation of different sources of home energy consumption based on smart meter data. In [14], disaggregation is used in the decomposition of household net meter data into solar generation and energy consumption.

Other studies have focused solely on the estimation of PV power output from different areas using clustering techniques. Studies reported in [15], [18] proposed a data-driven approach that applies clustering techniques to predict power output of unknown solar sites with characteristics similar to those of a known proxy site using AMI data. Geographical correlations are used in the clustering process. Additionally, the literature includes a wide range of studies that are focused on estimating instantaneous PV generation from aggregated load data where the installed PV capacity is known. These are discussed as follows. In [16], [19], the authors applied multiple linear regression (MLR) method to disaggregate real-time PV output from aggregated net load data at a substation level. The study in [19] extended the concept defined in [16] by adding new regressors and weights. In [20] the authors applied an unsupervised disaggregation of PV production from power flow measurement by modeling PV generation as a function of irradiance. In contrast, the study in [21] proposed the application of a hybrid game theory combined with clustering techniques to implement the same problem. Hidden Markov Model (HMM) regression was applied in [22] to estimate solar PV generation. MLR model was proposed in [23]. This work uses AMI data to predict which homes have PV installed and to extract the solar PV generation from the aggregated load measurements.

From the above discussion, it can be deduced that disaggregation can be applied to various problems. A number of these studies attempt to estimate the installed PV capacity at a customer level and are reported in [17], [24]–[26]. None of the studies is focused on estimating PV at a distribution level. The study reported in [17] presented a change-point detection algorithm that analyses abnormal changes in time-series load data to detect PV’s presence. In [26], the authors combine ANN and dimension reduction techniques to identify customers with PV based on both net load and AMI data. Studies reported in [24], [25] proposed a two-stage support vector machine (SVM) approach that classifies different customer net loads based on the discrepancies observed in customers with PV and those without. A support vector classification model was then used to determine whether the customer had PV and a second stage was applied to determine...
the customer PV capacity. These four studies are the closest works related to our proposed method. However, there exist fundamental differences in terms of their proposed approach and the underlying objectives, which limit the application of these methods in the distribution level PV estimation as discussed in the following paragraphs.

To accurately estimate the PV capacity installed on a distribution feeder, several considerations need to be addressed. Firstly, model inputs must be correctly characterized. In this regard, probabilistic modeling of both solar irradiance and the load is needed. Feeder load is an aggregation of individual customer loads that have different use patterns. The representation of underlying uncertainties in these inputs is therefore critical and has not been undertaken by any of these studies. Secondly, these methods need large amounts of data from AMI which requires the installation of measurement infrastructure. This is a costly undertaking for the utility. In addition to these limitations, these methods have different objectives from what is presented in this paper.

The accurate estimation of PV capacity at the distribution level, as opposed to the customer level, is critical for various reasons. Firstly, the knowledge of the installed PV capacity at a distribution level is critical in network planning and reinforcement. Given that the planner is unaware of the location and the size of the distributed PV on the system, accurate estimation of the installed capacity on each of the network feeders is crucial in the planning exercise. Secondly, information relating to the installed PV capacity on a particular feeder is key in operations and maintenance decisions. Such decisions are likely to be made based on the overall condition of the feeder as opposed to that of an individual customer. As such, it is more important to know the capacity of PV installed at the distribution level as opposed to that installed at a customer level. Lastly, the most easily available data to system operators is at the feeder level. In this regard, the adoption of a method estimating PV at a distribution feeder level provides a simpler and less costly approach as compared to the one focused on customer-level PV. As such, an appropriate method of estimation that considers these practical factors are required, and considering this, the proposed method is developed to meet these practical requirements.

3) KEY GAPS IN LITERATURE
This paper addresses critical gaps in the existing literature concerning the estimation of PV capacity. Firstly, we could identify no method in the literature applied to the problem of estimating the capacity of PV installed on a distribution feeder: those that focus on PV capacity estimation do so at a customer level. Secondly, the methods addressing customer level PV estimation cannot be applied at a distribution level for various reasons including the lack of comprehensive modeling of the irradiance and load uncertainties. This paper develops an uncertainty-based approach, which adequately represents these input uncertainties at each aggregation level. Lastly, this paper addresses the need for a simplified approach that does not require extensive use of AMI data.

4) CONTRIBUTIONS, LIMITATIONS, AND OVERVIEW
The main contributions of this paper are as follows. Firstly, it proposes a practical method that can adequately inform DNOs about the capacity of PV embedded in their networks. Secondly, this method introduces novel processes including stochastic expansion and aggregation. Additionally, a probabilistic quantile analysis estimation method is introduced. These methods can be applied in other areas of distributed energy resource estimation. Thirdly, this method accounts for the uncertainty characterizing the load and solar irradiance by using Monte Carlo simulation (MCS) processes to adequately represent these main inputs. Finally, the method requires common sets of data to estimate the PV capacity on a distribution feeder as opposed to extensive AMI data seen in related studies. This makes the estimation process less costly to the DNOs.

This paper focuses on residential LV PV systems where PV systems are allowed to export power back to the grid. In most cases, the regulation and monitoring of such installations is more challenging for DNOs. The efficacy of the proposed method is based on a data-based distinction between two load datasets which include a defined historical dataset characterizing an initial state before the introduction of PV, and the net load measurements, defining the present state with PV effects. In the likely case of system expansion or reconfiguration, the method’s accuracy depends on the validity of the assumptions regarding the two datasets. More specifically, cases of system expansion leading to increased system load require the incorporation of these changes in the historical load before the application of the proposed method. This is to avoid underestimation of the embedded PV. The proposed method hinges on probabilistic and risk-based formulation and is robust to variations in power generation including incidences of curtailment. However, as currently formulated, the method does not directly address issues such as curtailment which may be applied by the DNOs.

The rest of this paper is organized as follows. In section II, the overview of the proposed methodology is presented. Then, in section III, the proposed methodology is described in detail with the characterization of the model inputs and the formulation of the concepts of stochastic expansion, aggregation, and summation. The developed methodology is validated in section IV, and, in section V, its efficacy is demonstrated with various test cases. The paper concludes with section VI.

II. OVERVIEW OF THE PROPOSED METHODOLOGY
In this section, a high-level overview of the proposed methodology is described, as shown in the process flow of Figure 1. First, the general method to determining the PV capacity estimate is presented in an objective function. Then, the main processes contributing to solving the objective function are described in brief. These processes include the
description of data inputs, modeling of input variables, the stochastic simulation of customer loads and PV generation, and finally, the probabilistic estimation of the installed PV capacity. Each of these processes is described below and further explored in detail in section III of this paper.

A. THE OBJECTIVE FUNCTION

The proposed methodology is based on the hypothesis that the installed PV capacity can be estimated by comparing the measured net-load to simulated stochastic net-load scenarios based on aggregated historical load data and PV penetration scenarios. The simulated stochastic net-load models the range of possible net-load values considering the uncertainty in the customer loads and solar irradiance. As such, the objective function, \( f(x) \), is formulated to determine the best estimate of the measured net-load based on confidence intervals from a solution space of stochastic net-load profiles, \( S(x) \), as follows:

\[
S(x) = (S_1)_t = H_t - x \cdot k_t \; \forall x \in X = [y_0, y_f] 
\]

\[
\text{maximize: } f(x) = \frac{\sum \text{g}(P'(a_x S_M x'b_t))_{\tau=t}}{\tau} \times 100\% \quad (2)
\]

Probabilistic variables are distinguished from deterministic variables using boldface formatting. In the system of equations, \( x \) is a candidate solution of the aggregated PV capacity estimate in a defined selected solution space \([y_0, y_f]\). \( S_t, H_t, \) and \( k_t \) denote the simulated net-load, customer loads, and the PV capacity factors in time interval \( \tau \) and aggregated for the whole network. The function \( \text{g}(\cdot) \) is a Boolean function that tests – for all \( x \) and at confidence level \( \tau \) – whether the measured net-load \( M_{\tau} \) lies within the confidence interval \([a_x, b_t]\) of the simulated net-load, in a given time interval \( t \). Lastly, \( f(\cdot) \) is an objective function that searches for the optimal estimate with the maximum proportion of the measured net-load data points falling within the simulated net-load’s two-sided confidence interval applied over a defined period \([1, T]\).

The following sections describe the four processes that contribute to solving the objective function (2).

B. PROCESS FLOW TO SOLVING THE OBJECTIVE FUNCTION

The appropriate preparation of data inputs is necessary to ensure quality results. The Data Input segment of the proposed methodology specifies the data requirements in terms of sampling resolution, time-series range, and statistical significance. The data inputs are pre-processed in the input modeling segment to generate probabilistic time-series profiles of the customer loads and PV generation. The load profiles characterize load patterns at the customer level, with characteristics of load stochasticity and diversity. The PV generation models represent the expected PV generation considering a set of PV capacity scenarios, with attributes of solar irradiance uncertainty.

The load and PV generation profiles form the inputs to the input modeling segment. Two critical processes based on the MCS method are carried out: stochastic aggregation and summation. The stochastic aggregation process generates probabilistic aggregated feeder load profiles from the representative customer load profiles. The stochastic summation process maps corresponding scenarios of the net-load – termed simulated net-load hereafter – from the aggregated feeder load and PV generation profiles. The simulated net-load has probabilistic characterization as it replicates the uncertainty in the load and PV generation inputs.

The simulated net-load and the measured net-load profile form the Estimation Engine’s inputs, where the optimal estimate of the embedded PV capacity is determined. The optimal embedded PV capacity estimate is achieved by a statistical comparison of the probabilistic range of simulated net-load profiles to the singular measured net-load using a quantile density analysis.

III. METHODOLOGY

This section details the proposed methodology components: from the input requirements and the respective modeling processes to the conclusion of the PV capacity estimated.

A. DATA INPUT REQUIREMENTS

The proposed methodology requires three forms of input data: (1) time-series load data characterizing the electrical behavior of individual customers without PV or with a known PV capacity, which we term historical to denote pre-PV characteristics, (2) time-series historical solar irradiance of the location of the network, and (3) the measured net-load in the period of estimation. Data cleaning and checking are applied to ensure high-fidelity data inputs.

1) DESCRIPTION OF DATA VARIABLES

Historical load data: this data is required at the customer level to allow accurate modeling of the load diversity between...
customers. Ideally, the historical load data’s timeframe should match the measured net-load, such as new measurements from a sample of customers without PV. Where not available, data from previous years may be used but with caution. Where the available historical load data is too old for the period of estimation, adjustments may be required to account for trends significantly impacting the energy consumption, i.e., quantity and time-of-use patterns, on a long-term basis. Such trends may include energy efficiency, tariffs, and technological changes, e.g., the adoption of solar water heating systems and energy storage systems.

Solar irradiance data: in this work, we use instantaneous time-series global horizontal irradiance (GHI) to model the expected power yield from rooftop PV systems. GHI data is easily accessible for many locations worldwide and is easily integrable with most PV design software such as PVsyst. The required irradiance data is obtained from online irradiance resources [27] or a national weather station, such as SAURAN, in the case of South Africa [28], and other resources such as Solcast [29]. The availability of extensive historical data makes possible uncertainty characterization per time interval, favoring accurate PV generation modeling.

Another aspect affecting the accuracy of the developed PV models is spatial resolution. This paper assumes the network or feeder under test is small enough that there is an insignificant variation of GHI values and PV outputs between customer locations. Accordingly, the GHI data can be area-generalized and is applicable to the aggregated PV generation computation.

Measured net-load data: this data represents the load measurements with the impacts of the unknown capacity of embedded PV generation. The measured net-load is obtained as aggregated data at the feeder or transformer level. This form is the most common form in which the data exists and is convenient for aggregated PV capacity estimation. The section that follows discusses data quality characteristics.

2) STATISTICAL SIGNIFICANCE, CADENCE, AND TIME-SERIES RANGE
When modeling a technical variable’s characteristics, two aspects of critical importance are variability (the changes with time) and uncertainty (unknown state of the variable in each time instant). The time-series range and frequency of samples, also termed cadence determine the extent of variability captured. On the other hand, the number of samples per instant affects the degree of representativity of the associated uncertainty. Figure 2 illustrates these aspects.

For the focus application, hourly samples over a year’s period are adequate to capture the diurnal and seasonal variability, which are of most importance in PV studies. A matched cadence is required for the historical load data and measured net-load data. For detailed uncertainty representation of the historical load and irradiance data (measured net-load is fixed, hence deterministic), samples with adequate statistical significance are required [30]. In general, the bigger the sample size, all other factors being equal, the more reliable the test statistics.

Depending on the initial form of the presented data inputs, data manipulation may be required to drive the inputs to a preferred form that supports the generation of well-diversified customer load profiles and PV output profiles. The input modeling component deals with these aspects.

B. INPUT MODELLING
The input modeling component receives, as inputs, the set of ‘raw’ customer load profiles and irradiance profiles for the study area, in hourly cadence, over a year. From these data inputs, probabilistic profiles are generated through two modeling processes, as follows.

1) STOCHASTIC EXPANSION
The stochastic expansion component prepares the inputs for stochastic simulation, where the customer loads are aggregated and summed with the PV generation using the MCS method. These MCS processes typically require numerous scenarios in the order of thousands for convergence or adequate accuracy [31].

Suppose N samples of profiles are required for sufficient accuracy, and the data inputs are of a smaller sample size, n. Then, the data inputs need to be expanded accordingly. For this, we define a stochastic expansion process that generates a larger pool of profiles from a smaller pool using a random sampling technique. Given the stochasticity of customer loads, the method considers the original profile samples as the observed or sampled behavior of the concerned random variable in a broad population of profiles. As such, the profile samples represent the characteristic profiles from which we develop inferential models per interval (as illustrated in Figure 2) using the non-parametric probability distribution function (PDF).

For grouped or clustered customers with similar usage patterns, the PDFs in each interval represent the likely load characteristic of customers belonging to that cluster. As such, MCS samples of ‘characteristic’ loads, with preservation of the time-series variation and correlation, can be generated from each interval’s characteristic non-parametric PDF function to model a typical customer’s load profile. The
process is repeated for a selected number of iterations, \( N \). The expanded set of data with \( N \) profiles represents the range of possible load profiles for the grouped customers, considering load diversity within time intervals and the influence of stochasticity. To ensure the expanded historical data, \( H \), is an accurate statistic of the original load data, \( \tilde{h} \), the density functions of corresponding intervals must be computable:

\[
f_p(H_t) \approx f_p(\tilde{h}_t); \quad \forall t \in [1,T]
\]

where \( f_p() \) is a probability function that fits a non-parametric PDF to the argument data.

The expansion process is also applied to the historical solar irradiance data profiles, \( I_N \), to achieve an expanded set of characteristic profiles, \( I_{N'} \). These irradiance profiles need further processing to achieve a corresponding set of PV generation profiles.

2) PV GENERATION MODELING

PVsys is used to model, from irradiance data, the output power per installed kWP. Performing this for the probabilistic range of \( N \) irradiance profiles, \( I_{N'} \), would be computationally intensive. To minimize computational burden, we map the output power, \( P_{out} \), for a single representative profile, \( I_0 \), using PVsys. Then, to get the probabilistic PV output profiles, \( k_t \), we scale the remaining profiles in the expanded irradiance profiles, \( I_{N'} \), according to the following transform:

\[
k_t = (P_{out}/I_0) \times I_N
\]

The probabilistic PV output profiles, \( k_t \), represent the uncertainty in a 1 kWp system’s expected PV output. Other PV capacities’ power output profiles can be achieved by simple, linear scaling.

3) PV OUTPUT ADJUSTMENT

Software estimates of the expected PV output are generally higher than the actual performance [32]. The differences arise from disparities between the simulated and practical models of the analyzed PV systems. For instance, there are site-dependent factors such as dust accumulation that potentially reduce PV systems’ performance by margins up to 15% and may not be adequately captured by the PV software in calculating the factor \( k_t \) [33]. The assumption of optimal panel orientation in simulation models, which may not be implementable practically, also contributes to further disparities [32], [34]. Accordingly, software-derived PV output models require adjustment to drive them towards realistic expectations.

Studies investigating PV systems’ performance ratios of the ideal to actual PV output can guide the adjustment process: PV system performance ratios can be used as adjustment or correction factors that downscale the simulated PV output, such as the one in (4). The selection of the relevant adjustment factor should consider geospatial characteristics, which have been proved to significantly affect performance ratios [35], [36]. In this study, we apply an adjustment factor of 80%, supported by the performance ratios for Sydney reported in [32].

C. STOCHASTIC SIMULATION

This component develops aggregated net load scenarios based on candidate PV capacity solutions. Two processes are required to achieve this: the aggregation of individual customer loads, what we term stochastic aggregation, and the summation of the aggregated loads and PV generation profiles, what we term stochastic summation.

1) STOCHASTIC AGGREGATION OF CUSTOMER LOADS

Stochastic aggregation is a process for simulating scenarios of aggregated load profiles at the feeder level. The simulation process involves the random selection and summation of profiles from the pool of characteristic probabilistic profiles obtained from the stochastic expansion process. We use the MCS method with a limited number of trials to generate representative scenarios from the large spectrum of combinatorial possibilities. To ensure accuracy, we conduct convergence tests that indicate the sufficiency of the applied number of trials.

For a feeder with \( n \) customers, an MCS aggregation procedure with \( m \) trials is as follows:

i. Using the MCS, randomly select \( n \) customer load profiles from the probabilistic customer load profiles.

ii. Sum the selected profiles, interval by interval, to achieve a single aggregated load profile scenario.

iii. Repeat steps (i) and (ii), with replacement, \( m \) times to achieve \( m \) scenarios of aggregated load profiles.

The generated aggregated load profiles, \( A_t \), form part of the inputs to the stochastic summation process.

2) STOCHASTIC SUMMATION

This component simulates scenarios of aggregated net-load profiles using the aggregated load and PV output profiles as inputs. Since the said inputs are probabilistic variables, a random simulation process must also model representative net-load scenarios. We use the MCS for this summation process, with PV output profiles modeled as negative loads.

The simulation procedure is similar to that of stochastic aggregation, except that only two profiles are summed at a time: a randomly selected aggregated load profile and an aggregated PV output profile. Also, the summation is repeated with different PV output profiles scaled according to the embedded PV capacity scenarios, \( x \), in a defined candidate solution set \( X \), according to (1). The selected range of \( X \) depends on the anticipated levels of PV penetration of the analyzed feeder, and its adequacy is assessed based on the results from the estimation engine. The selection of the extent of MCS trials, \( m \), should be guided by convergence tests. In this study, we apply 1,000.

The stochastic summation process leads to a set of probabilistic, simulated net-load profiles for various PV capacity solution scenarios. We use the term ‘simulated’ to distinguish these net-load profiles from the measured ones. The estimation engine tests the fitness of the solution scenarios.
D. ESTIMATION ENGINE

The estimation engine determines the optimal estimate of the installed PV penetration by finding the PV capacity solution scenario with the simulated aggregated net-load that best compares to the measured net-load profile. The estimation engine is modeled probabilistically to cater to the uncertainty reflected in the simulated net-load variables. This approach allows a risk-based comparison of the probabilistic range of simulated net-load profile scenarios and a single measured net-load profile.

1) QUANTILE DENSITY ANALYSIS

The probabilistic range of the simulated net-load is constrained by the maximum and the minimum net-load scenarios in each time interval. However, to avert the risk of evaluating outliers, we restrict the range of net-load scenarios to a band defined by a selected risk level or, conversely, statistical confidence intervals. For instance, a 10% risk level is defined by a quantile band between the 5th and 95th quantiles and captures 90% of the population of simulated net-load profiles.

The simulated net-load’s quantile band – defined in each time interval – is then compared to the measured net load. Figure 3 demonstrates, for a given time interval $t$, the comparison between a measured net-load, $M_t$, and the probabilistic range of simulated net-load with $x$ kWp PV capacity, $\mathcal{S}_x$, restricted to the quantile band $[x_a,t, x_b,t]$.

To determine the fitness of the PV capacity candidate solutions, we conduct a 3-step test protocol:

i. For each time interval in the time-series range $(1, T)$, test the measured net loads' incidence within the simulated net loads' quantile band:

\[
g(x, t) = \begin{cases} 1, & x_a \leq M_t \leq x_b, \forall t \in [1,T] \\ 0, & \text{otherwise} \end{cases}
\]  

ii. Evaluate the density of the measured net-load points (or time intervals) that test positive in step (i). We term this the quantile density, and is expressed as a percentage of the total number of time intervals:

\[
Q_x = \frac{\sum_{t=1}^{T} g(x,t)}{T} \times 100 \%
\]  

iii. Repeat (i) and (ii) for all candidate solutions $x$. The optimal estimate is obtained by evaluating the tested capacity for which the quantile density is maximum.

This analysis is carried out on sun-hours and evaluated for an ordinary day assumed to have nine sun-hours that range from 8 am to 4 pm.

E. SUMMARY OF THE PROPOSED METHOD

A summary of the implementation algorithm is presented below. It highlights the steps taken in the formulation of the probabilistic quantile estimation process. The various methodology components are underlined, and any relevant equations are referenced in parenthesis.

**Algorithm 1: Probabilistic quantile estimation process**

1. **Input:** Historical customer load profiles, $h_t$, Irradiance profiles, $I_h$, Measured net load data, $M_t$, number of customers on the feeder, $n$
2. **Stochastic Expansion:** Generate 5,000 MCS random customer load profile, $H_t$, from the pool, $h_t$, and 1,000 random area-specific irradiance profiles, $I_N$, from pool $I_h$
3. **for** $m = 1:1000$ – Stochastic Aggregation for generating aggregate load profiles, $A_t$
   4. Select $n$ random customer load profiles from $H_t$
5. **end for**
6. for $x = y$ to $y$
7. **PV output modeling:** Convert irradiance Profiles, $I_N$, to an expected PV output profiles, $k_t$ (equation 4)
8. Scale $k_t$ according to the aggregated PV capacity scenario $x$
9. **for** $m = 1:1000$
10. **Stochastic summation:** Generate simulated netload profiles, $S_x$, from the aggregated load, $A_t$, and aggregated PV capacity scenario, $x \cdot k_t$
11. **end for**
12. **end for**
13. **Quantile Density Analysis:** Test the incidence of $M_t$ within the quantile band of $S_x$ at a selected risk (equation 5)
14. For each scenario $x$, evaluate the density of points $Q_x$ satisfying the test of line 15 (equation 6)
15. **Output:** Return the capacity estimate $x$ at which $Q_x$ is maximum

![Figure 3: Comparison of the measured and simulated net-load variables](Image)
IV. VALIDATION OF THE METHODOLOGY

This section presents the validation of the proposed methodology. First, it deals with the input modeling processes (stochastic expansion and aggregation), vital for probabilistic customer load profiles that are accurate models of the original data. Secondly, it deals with the stochastic summation and estimation engine, linked to the estimation process’s outcomes. Lastly, the section explores the methodology’s performance under various test conditions.

A. TEST DATA

All validation tests are based on South Africa’s domestic load research (DLR) data for residential customers [37]. The DLR data is a one-year data at a 5-minute resolution. The data describes the passive (without PV) consumption characteristics of a group of 73 residential customers in a relatively affluent area, classified by a living standard measure (LSM) 10. Though without PV customers, the DLR data is convenient for generating ‘synthetic’ (not measured from physically installed systems) net-load data with varied aggregated installed capacity, allowing for testing the proposed methodology’s performance under ideal cases.

To generate the synthetic net-load data, we take the historical load data and a selected scenario of aggregated PV capacity, including its uncertainty, and perform a stochastic summation. A synthetic measured net-load is then achieved by selecting a single synthetic net-load scenario using a random selection process or percentiles. As a base case, a synthetic net-load with 75 kWp embedded PV capacity is used.

B. VALIDITY OF INPUT MODELING PROCESSES

Two conditions influence the representativeness of sample-based probabilistic models: first, the adequacy of selected samples in representing the behavior of the original system, and second, consistency of statistical properties, i.e., uncertainty in each time interval and variability in the time series. Several statistical tests are conducted to ensure the conditions stipulated above are met for the stochastic expansion and aggregation processes.

1) SAMPLE SIZE ADEQUACY

A convergence test was carried out to determine the optimal sample size in creating the probabilistic customer load profiles through stochastic expansion. In this test, the 5th percentile, the mean, and 95th percentile profiles for different sample sizes are compared to the original data’s corresponding properties. Mean errors per interval are recorded, and histograms of all errors in the time series are plotted. The sample size is increased in steps of 500 until the tolerance level, ε, is less than 0.2 A, representing less than 1% error in the average customer load.

Figure 4 compares the mean error distributions for the various sample sizes. It is observed that a sample size of 5,000 achieves the selected tolerance of 0.2 A. Tests with other datasets, including solar irradiance, confirm that a sample size of 5,000 is generally valid.

The convergence test protocol was also used to determine the optimal sample size for the simulated aggregated load profiles, which was determined to be 1,000.

2) STATISTICAL COHERENCE

The consistency of statistical properties between the expanded probabilistic load profiles from the stochastic expansion process and the original customer load profiles is tested by analyzing the intra-interval and inter-interval variations.

Intra-interval coherence measures how well the expanded customer load profiles represent the load diversity of the original profiles. In contrast, inter-interval coherence deals with time-series variability. Conformance of the simulated samples to these characteristics can be referred to as statistical coherence, and the term is consistently used in the paper.

Figure 5 compares non-parametric PDFs of customer load currents between the expanded and original customer load profiles in randomly selected time intervals. The consistency between the PDFs indicates the stochastic expansion process preserves the original load data’s diversity characteristic in each time interval.

Figure 6 compares the 5th percentile, mean, and 95th percentile load profiles between the expanded and original load datasets. The plot shows the original and expanded load profiles follow the same trend with minor differences (lower than 2 A mean errors per interval) observed in the lower and upper percentile profiles. The percentiles’ error margin is expected as the percentiles towards the distribution tails are sensitively affected by outlier samples.

The two validation tests for the adequacy of simulation samples and statistical coherence establish confidence in the critical concepts around stochastic simulation, anchoring the proposed methodology. Though demonstrated using load data, the same approach is equally applicable to the PV data expansion process. The validity of the stochastic summation and the estimation engine is demonstrated next through various case studies.

C. VALIDITY OF THE ESTIMATION PROCESS

The estimation engine’s efficacy pivots on the quantile density analysis, through which a candidate solution’s fitness in
modeling the measured net-load is assessed. The selection of the risk level determining the quantile band’s range is critical.

As aforementioned, taking the full simulated net loads probabilistic range (the min-max range) would weigh the effects of outliers in the estimation. Figure 7 is a quantile density ($Q_x$) plot for a range of PV capacity solutions, x. Generally, the quantile density plot’s interpretation is as follows: a low quantile density signifies an estimate whose simulated net-load band poorly captures the measured net-load profile, and the global maximum of the plot indicates the optimal estimate with the highest capture of measured net-load data points. Applied to Figure 7, the PV capacity estimation is inconclusive as several estimates in the range 0 – 35 kWp result in a quantile density of 100%. The plot illustrates the problem of a minimum-maximum approach and highlights the importance of a risk-based approach to averting the assessment of outliers.

To select the best risk level, we study the effect on PV capacity estimation of the selected risk level and quantile band’s width. Table 1 compares the PV capacity estimates from various risk levels applied in the estimation engine.

The results of Table 1 indicate that the estimate accuracy deteriorates with the increase in the risk level and, conversely, the reduction of the probability band’s width. The results indicate that the method performs well at lower risk levels, which is expected as the confidence interval is large. Accordingly, a 1 % risk level is concluded as most suitable for robust estimation considering the uncertainty in the net-load’s practical scenario. This risk level is applied hereafter as the default in the estimation engine.

Figure 8 illustrates how PV capacity candidate solutions lead to distinct quantile densities. The plot depicts a portion of the simulated net-load profiles and the capture of the synthetic net-load by the respective quantile bands at 1 % risk. Differences in the simulated net-load profiles are observed during the sun-hour period (8 am – 6 pm) with different duck-curve depths depending on the PV capacity candidate solution. The 75 kWp solution better captures the measured net-load profile within its quantile band than other candidate solutions in the selected time-series period. Taken over the whole year, Figure 9 confirms 75 kWp as the optimal estimate with a distinct quantile density maximum over the entire solution space. The results also show consistency in accuracy for all cases of random synthetic measured net-load profiles. The conclusion is the validity of the 1 % risk level in achieving clear conclusions on the optimal PV capacity estimate.
D. ROBUSTNESS

To ensure the proposed methodology is widely applicable to various contexts, we test its response to various installed PV capacities and uncertainty conditions in the measured net-load scenario.

1) VARIED PV CAPACITIES

In this test case, the proposed methodology’s performance is tested under various PV capacity or penetration scenarios. Four penetration levels are considered, and five scenarios of net-load profiles are taken from the stochastic range of the synthetic measured net-load profiles. The results shown in Table II indicate that the proposed probabilistic estimation method achieves results with high accuracy. There are tolerable discrepancies observed where an unlikely or outlier synthetic net-load profile is applied. The relative errors observed in these cases range from 2.4% for the highest tested capacity to 10% for the lowest tested capacity. Given that only the extreme cases are likely to have the worst errors, the results validate the proposed methodology’s underlying theory with high confidence.

2) UNCERTAINTY CONDITIONS

In this case study, we test the influence of uncertainty characterization (for customer loads and solar irradiance) on the accuracy of PV capacity estimates. The proposed method’s performance (with uncertainty characterization) is compared to that of our baselined deterministic method.

Baseline deterministic study: To illustrate the importance of uncertainty representation, a deterministic method reported in [5] is adopted. The deterministic formulation characterizes the loads and solar irradiances using single profiles without considering the associated uncertainty and applies the Pearson-Tukey estimation approach in generating both feeder historical load and PV capacity factor profiles. The estimation of PV capacity is carried out by least error search for the normalized relative error between the measured and the simulated net load. Such a direct method has also been reported in [24].

The performance of this method is evaluated at a capacity of 75 kWp and the results are compared to the proposed method. The results are tabulated in Table III. Figure 10 illustrates the results obtained using a deterministic method and the resulting PV estimation curves based on normalized least error analysis. For the same system with 75kWp, the deterministic method’s PV capacity estimation becomes inconclusive as multiple estimates are achieved. As expected with deterministic solutions, accurate results are achieved only for selected measured net-load scenarios close to the mean of the simulated range of stochastic net-load profiles. Moving away from the mean scenario, e.g., the 5th and 97.5th percentiles, the deterministic methodology’s performance deteriorates. In this case, significant errors of more than 30% are observed. This method is thus inconsistent. On the other hand, the probabilistic method is robust, as demonstrated in Figure 9, showing estimation plots for the same measured net-load scenarios as those of Figure 10. The PV capacity estimate is conclusive at 75 kWp, with

![Figure 8. Illustrating the quantile density concept](image)

![Figure 9. PV estimation curves for random profiles at 1% risk using the probabilistic methodology.](image)

| TABLE II | ROBUSTNESS OF ESTIMATION ACCURACY TO PV CAPACITY |
|----------------|-----------------------------------------------|
| Actual installed capacity (kWp) | PV capacity estimates for various synthetic net-load scenarios (kWp) | 5th | mean | 97.5th | Rnd1 | Rnd2 |
| 75 (base case) | 75 | 74 | 72 | 71 | 74 |
| 100 | 100 | 98 | 98 | 96 | 99 |
| 200 | 201 | 200 | 199 | 194 | 196 |
| 300 | 302 | 299 | 298 | 296 | 298 |

| Net load profile | Quantile analysis | 3-point estimation |
|------------------|------------------|-------------------|
| Probabilistic estimates (kWp) | Deterministic estimates (kWp) |
| 5th | 75 | 90 |
| mean | 74 | 75 |
| 97.5th | 72 | 50 |
| Rnd1 | 71 | 75 |
| Rnd2 | 74 | 50 |
| MAPE | 2.4 | 17.33 |
| SE | 1.64 | 12.55 |
insignificant deviations, for all stochastic scenarios of the measured net-load. The findings establish the validity of a probabilistic estimation method to cater to input uncertainties.

V. A PRACTICAL CASE STUDY
To demonstrate the proposed methodology’s efficacy and practical application, we present a case study involving field gross-metering measurements from residential PV installations on the Ausgrid in the New South Wales region, Australia [38]. In this study, we compared the performance of our proposed method and that of our baseline study.

A. TEST DATA
Prosumer data from Australian utility, AusGrid, is available for 300 homes with rooftop PV systems, each of which is monitored by a gross meter, which records the PV power gross generation (GG) every 30 minutes. The data also includes half-hourly load measurements in two forms: general consumption (GC), which records the total amount of electricity supplied to the customer for standard consumption, and controlled load (CL), which records the additional electricity supplied under specific tariff regulations [38]. Each customer’s installed PV capacity in kWp is also provided.

B. TEST CASES
The Ausgrid dataset is used as follows. The sum of the GC and CL measurements is taken as the historical load (H) since it represents the customer’s gross electricity consumption without PV. Irradiance data obtained from Solcast [30] is used in modeling the simulated net load for a range of candidate PV capacity solutions in the set [0, 100] kWp. The measured net load is generated by discounting the total load (GG and CL) from the GG measurements and aggregated for all customers in a selected service area.

The service area containing the 300 customers is quite large and may have significantly varying irradiance and customer load characteristics. To enhance the quality of input models, the spatial resolution was increased by taking a subset of 100 customers separated into two smaller areas – A and B, shown in Figure 11 – based on customer density. Area A has 63 customers, and Area B has 37 customers. Customer segmentation for each area is also applied according to load type: customers with and without controlled loads. The characteristics of the tested datasets are presented in Table IV.

The segmented data lead to four test cases. In each case, the proposed methodology’s accuracy is quantified by calculating the percentage ratio of the PV capacity estimate to the actual aggregated installed capacity.

C. RESULTS
Figure 12 shows the estimation curves for the customers segmented by area and load type. Each curve’s optimal estimate is highlighted with a broken blue line and the actual installed PV capacity with a broken red line.

The shapes of the estimation curves show a clear conclusion on the installed PV capacity taken to be the global maxima. The quantile densities’ maxima corresponding to the optimal estimates (> 70%) indicate good estimation accuracy. Tables V and VI quantify the estimation accuracy for the test cases segmented by area and load type, based on the method used. From Table VI, the proposed method shows an average

\[ \text{PV generation estimation} \]
Undoubtedly, this performance is dismal and disqualifies the levels between 3.3% to 25% very low estimation accuracy. The method records accuracy hand, applying the deterministic method to this data yields eliminates PV output over the same one used in the estimation methodology, which dataset used to generate synthetic measured net achieved with synthetic data. In the latter, the solar irradiance trend reflects the over-estimation in PV output power because of the limited accuracy in the available solar irradiance data. This is the main difference between these results and those achieved with synthetic data. In the latter, the solar irradiance dataset used to generate synthetic measured net-load is the same one used in the estimation methodology, which eliminates PV output over-estimation errors. On the other hand, applying the deterministic method to this data yields very low estimation accuracy. The method records accuracy levels between 3.3% to 25% as illustrated in Table V. Undoubtedly, this performance is dismal and disqualifies the application of this estimation method. The standard error (SE) and MAPE are analyzed and presented in Table VII.

### VI. CONCLUSION

This paper presented a novel probabilistic method for estimating the capacity of PV embedded in a network’s measured aggregated net-load using available datasets of historical load and historical irradiance data. Various uncertainty handling techniques are implemented including new concepts of stochastic expansion and aggregation essential in expanding and aggregating stochastic customer loads. A quantile density analysis formulation enables risk-based comparative analysis of probabilistic simulated net-load scenarios with the measured net-load leading to optimal estimates of the installed capacity. The method is validated using synthetic and practical data.

A wide range of tests based on synthetic data demonstrate the robustness of the proposed method in response to various feeder conditions such as feeder size, customer population, and the total installed capacity. Sensitivity tests using the same data demonstrate the impact on estimation accuracy of simulation characteristics such as the extent of stochastic simulation, selection of risk factors in the probabilistic framework, and the representation of load and PV uncertainties. Tests using real smart meter measurements and a practical test system from Ausgrid demonstrate the proposed method’s practical application. Consistent performance with an average accuracy of approximately 90% is achieved over four test cases with varying customer population, load conditions, and spatial coverage. In addition, the proposed
method achieves MAPE of 10.6% which is a good performance indicator.

Compared to other formulations, the proposed method is relatively simpler yet more comprehensive since it considers uncertainty. The proposed probabilistic method outperforms a deterministic three-point estimate method, and the significance of the differences in the methods’ results highlights the relevance of extensive uncertainty characterization and simulation in the estimation process.

In practice, this method can provide a reliable monitoring tool for utilities to systematically monitor the changes in the PV capacity on their feeders without extra metering investment costs. However, the quality of input data for load and PV generation modeling – particularly customer or area segmentation for high-resolution input modeling – affects output accuracy. Other areas that would benefit from further research include the consideration of different DG types, the impact of regulated DG operation dynamics, such as curtailment, and spatiotemporal modeling aspects on estimation accuracy.

REFERENCES
[1] L. S. Waswa and B. Bekker, “Impact of PV small scale embedded generation on South Africa’s system demand profile,” in South African Solar Energy Conference, 2018.
[2] N. Mararakanje and B. Bekker, “Renewable energy integration impacts within the context of generator type, penetration level and grid characteristics,” Renew. Sustain. Energy Rev., vol. 108, pp. 441–451, Jul. 2019.
[3] South Africa Bureau of Standards (SABS), “NRS 097-2-3 : 2014 -A simplified utility connection criteria for low-voltage connected generators," NRS, South Africa, 2014.
[4] C. T. Gaunt, “Implications of renewable energy intermittency for power system supply planning,” in 2017 IEEE AFRICON: Science, Technology and Innovation for Africa, AFRICON 2017, 2017, no. March, pp. 1095–1100.
[5] L. S. Waswa, “A probabilistic estimation of the capacity of solar PV SSEGs installed on a LV feeder network,” Stellenbosch University, 2020.
[6] C. M. Cheung, W. Zhong, C. Xiong, A. Srivastava, R. Kannan, and V. K. Prasanna, “Behind-the-meter solar generation disaggregation using consumer mixture models,” in 2018 IEEE International Conference on Communications, Control and Computing Technologies for Smart Grids, 2018.
[7] Department of Agriculture Western Cape, “The mapping of agricultural commodities and infrastructure in the Western Cape,” Department of Agriculture, Government of Western Cape, Elsenburg, Stellenbosch, 2018.
[8] L. López-fernández, S. Lagüela, I. Picón, and D. González-aguilera, “Large scale automatic analysis and classification of roof surfaces for the installation of solar panels using a multi-sensor aerial platform,” Remote Sens., vol. 7, no. February, 2015.
[9] O. Reinecke, C. Leonard, K. Kritzinger, B. Bekker, J. L. van Niekerk, and J. Thilo, “Unlocking the rooftop PV market in South Africa,” Stellenbosch, 2013.
[10] K. Bradbury, L. C. Pratt, T. J. Nicholas, J. M. Pratt, and R. N. Nicholas, “Solar Power Estimation Through Remote Sensing,” Duke Energy Initi., p. 2007, 2015.
[11] J. M. Malof, Rui Hou, L. M. Collins, K. Bradbury, and R. Newell, “Automatic solar photovoltaic panel detection in satellite imagery,” in 2015 International Conference on Renewable Energy Research and Applications (ICRERA), 2015, pp. 1428–1431.
[12] A. Ebrihaim and O. Mohammed, “Pre-processing of energy demand disaggregation based data mining techniques for household load demand forecasting,” Inventions, vol. 3, no. 45, pp. 1–16, 2018.
[13] M. Wytock and J. Z. Kolter, “Contextually supervised source separation with application to energy disaggregation,” Proc. Natl. Conf. Artif. Intell., vol. 1, pp. 486–492, 2014.
[14] D. Chen and D. Irwin, “SunDance: Black-box Behind-the-Meter Solar Disaggregation,” e-Energy, 2017.
[15] H. Shaker, H. Zareipour, and D. Wood, “A data-driven approach for estimating the power generation of invisible solar sites,” IEEE Trans. Smart Grid, vol. 7, no. 99, pp. 1–11, Sep. 2016.
[16] E. C. Kara, C. M. Roberts, M. Tabone, L. Alvarez, D. Callaway, and E. M. Stewart, “Disaggregating solar generation from feeder-level measurements,” Sustain. Energy, Grids Networks, vol. 13, pp. 112–121, Mar. 2018.
[17] X. Zhang and S. Grijalva, “A data-driven approach for detection and estimation of residential PV installations,” IEEE Trans. Smart Grid, vol. 7, no. 5, pp. 2477–2485, 2016.
[18] H. Shaker, H. Zareipour, and D. Wood, “Estimating power generation of invisible solar sites using publicly available data,” IEEE Trans. Smart Grid, vol. 7, no. 5, pp. 2456–2465, Sep. 2016.
[19] E. Vrettos, E. C. Kara, E. M. Stewart, and C. Roberts, “Estimating PV power from aggregate power measurements within the distribution grid,” J. Renew. Sustain. Energy, vol. 11, no. 2, p. 23707, Mar. 2019.
[20] F. Sossan, L. Nespoli, V. Medici, and M. Paolone, “Unsupervised disaggregation of photovoltaic production from composite power flow measurements of heterogeneous prosumers,” IEEE Trans. Ind. Informatics, vol. 14, no. 9, pp. 3904–3913, Sep. 2018.
[21] F. Bu, K. Dehhchanpour, Y. Yuan, Z. Wang, and Y. Zhang, “A data-driven game-theoretic approach for behind-the-meter PV generation disaggregation,” IEEE Trans. Power Syst., vol. 35, no. 4, pp. 3133–3144, 2020.
[22] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, “Estimation of behind-the-meter solar generation by integrating physical with statistical models,” in 2019 IEEE International Conference on Communications, Control and Computing Technologies for Smart Grids, 2019.
[23] M. Tabone, S. Kiliccote, and E. C. Kara, “Disaggregating solar generation behind individual meters in real time,” in Proceedings of the 5th Conference on Systems for Built Environments - BuildSys '18, 2018, pp. 43–52.
[24] F. Wang et al., “A distributed PV system capacity estimation approach based on support vector machine with customer net load curve features,” Energies, vol. 11, no. 7, 2018.
[25] K. Li, F. Wang, Z. Li, M. Fotuhi-Firuzabad, N. Duı́, and T.
Wang, “Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation,” Appl. Energy, vol. 253, no. July, 2019.

[26] D. L. Donaldson and D. Jayaweera, “Effective solar prosumer identification using net smart meter data,” Int. J. Electr. Power Energy Syst., vol. 118, no. July 2019, p. 105823, 2020.

[27] SodaPro, “SoDa products - www.soda-pro.com,” Helio_Clim Database, 2019. [Online]. Available: http://www.soda-pro.com/soda-products. [Accessed: 21-Apr-2019].

[28] M. . Brooks, “‘SAURAN: A new resource for solar radiometric data in southern Africa,”” J. Energy South. Africa, vol. 26, pp. 2–10, 2015.

[29] Solcast, “Soleast,” Solcast Australia, 2020. [Online]. Available: https://toolkit.solcast.com.au/. [Accessed: 15-Jul-2020].

[30] C. T. Gaunt, R. Herman, and B. Bekker, “Probabilistic methods for renewable energy sources and associated electrical loads for Southern African distribution systems,” 2009 CIGRE / IEEE PES Jr. Symp. Integr. Wide-Scale Renew. Resour. into Power Deliv. Syst., no. March 2019, pp. 1–7, 2009.

[31] M. J. Chihota and B. Bekker, “Modelling and simulation of uncertainty in the placement of distributed energy resources for planning applications,” 2020 Int. Conf. Probabilistic Methods Appl. to Power Syst. PMAPS 2020 - Proc., 2020.

[32] N. Haghdadi, J. Copper, A. Bruce, and I. Macgill, “Operational performance analysis of distributed PV systems in Australia,” in Asia-pacific Solar Research Conference, 2016.

[33] A. Bonkaney, S. Madougou, and R. Adamou, “Impacts of cloud cover and dust on the performance of photovoltaic module in Niamey,” J. Renew. Energy, vol. 2017, no. 2013, pp. 1–8, 2017.

[34] N. Umar, B. Bora, C. Banerjee, and B. S. Panwar, “Comparison of different PV power simulation softwares: case study on performance analysis of 1 MW grid-connected PV solar power plant,” Int. J. Eng. Sci. Invent., vol. 7, no. 7, pp. 11–24, 2018.

[35] B. Shiva Kumar and K. Sudhakar, “Performance evaluation of 10 MW grid connected solar photovoltaic power plant in India,” Energy Reports, vol. 1, pp. 184–192, 2015.

[36] D. Okello, E. E. Van Dyk, and F. J. Vorster, “Analysis of measured and simulated performance data of a 3.2 kWp grid-connected PV system in Port Elizabeth, South Africa,” Energy Convers. Manag., vol. 100, pp. 10–15, 2015.

[37] University of Cape Town, “Domestic Load Research Project,” Energy Research Data Portal for South Africa, 2018. [Online]. Available: http://energydata.uct.ac.za/th/organization/about/dlr. [Accessed: 17-Oct-2018].

[38] AusGrid, “Solar Home Electricity Data,” AusGrid, 2020. [Online]. Available: https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data. [Accessed: 15-Jul-2020].

LEWIS WASWA obtained the B.Sc., and M.Eng. degrees in Electrical Engineering from the University of Nairobi, Kenya, and Stellenbosch University, South Africa, in 2016 and 2020, respectively. He is currently enrolled for PhD. in Electrical Engineering at Stellenbosch University, where his research interests include distributed energy resources integration studies, rural electrification, low voltage distribution network planning, and energy policy. Lewis works as an engineer at the Centre for Renewable and Sustainable Energy Studies (CRSES).

MUNYARADZI J. CHIHOTA received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the University of Cape Town, Cape Town, South Africa, in 2011, 2015, and 2019, respectively.

From 2019 to 2020, he was a Research Engineer with the Centre for Renewable and Sustainable Energy Studies (CRSES), Stellenbosch, South Africa. Since 2020, he has been a Post-doctoral Fellow with the Electrical and Electronic Engineering Department, Stellenbosch University, South Africa. His research interests include distribution network planning using probabilistic techniques, load modeling and customer characterization, distributed energy resource integration, and autonomous control of inverter-based energy resources.

BERNARD BEKKER (M’19) received the B.Eng. and M.Eng. degrees in electrical engineering from Stellenbosch University, South Africa, in 1996 and 2004, and the Ph.D. degree in electrical engineering from the University of Cape Town, South Africa, in 2010.

He currently holds the Eskom Research Chair in Power System Simulation at Stellenbosch University and is the Associate Director of the Centre for Renewable and Sustainable Energy Studies at the same university. His research interest includes the grid integration of distributed energy resources, probabilistic modeling, and variable renewable energy forecasting.