FRAMEWORK FOR SCENARIO GENERATION AND REDUCTION IN PHOTOVOLTAIC-INTEGRATED GENERATION COMMITMENT

by

Lee Mckenzie Luis

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Electrical Engineering

Charlotte

2018

Approved by:

Dr. Sukumar Kamalasadan

Dr. Valentina Cecchi

Dr. Umit Cali

ABSTRACT

LEE MCKENZIE LUIS. Framework for Scenario Generation and Reduction in Photovoltaic-Integrated Generation Commitment. (Under the direction of DR. SUKUMAR KAMALASADAN)

Photo-voltaic (PV) power generation is promising from an environmental and economic standpoint. However, increased penetration of PV power into the electric grid can present significant challenges to decision-making in energy markets. To maintain grid robustness, especially with renewable sources like PV power that are highly uncertain and variable, system operators seek the most accurate information available on the generation characteristics. In this regard, a scenario generation and reduction framework that captures the uncertainty and variability of PV power is proposed. This work characterizes the forecast error via a set of uncertainty and variability indices. A large set of scenarios is generated using a pseudo-random number generation process. Next, a scenario reduction framework to improve computational tractability is proposed. Finally, the efficacy of these methodologies is proven by observing their impact on the unit commitment solution via a cost-benefit analysis. The proposed work is tested using measured and forecasted PV power data for a specific geographical location in North Carolina. The entire framework is built using MATLAB/Simulink and GAMS Optimization software. The results indicate that incorporating scenarios in the deterministic unit commitment model improves overall operational costs while capturing the uncertainty and variability of PV generation.

ACKNOWLEDGEMENTS

I take this opportunity to express my sincere gratitude to everyone who contributed toward the successful completion of my thesis work, academic coursework, research work, and overall life on campus, here at the University of North Carolina at Charlotte.

First and foremost, I sincerely thank my advisor Dr. Sukumar Kamalasadan. Dr. K has been a constant source of rich technical-expertise and moral support throughout my graduate studies and research. His patience and positive approach to problem-solving has helped me overcome several obstacles faced in my research work. I feel truly fortunate to have studied under his thoughtful mentorship. Without his guidance and never-ending support, I would not have been able to complete this research work.

I would like to thank my committee members for their efforts and support in my research and academic work during the past two years: Dr. Valentina Cecchi and Dr. Umit Cali. Dr. Cecchi supported me through a research position in the very first semester of my graduate studies; an experience that gave me first-hand experience in academic research. With her help, we found UNCC's first IEEE Power and Energy Society (PES) Student Chapter and later represented the school at the US-Canada IEEE PES Student Congress in Boston, Massachusetts; a memorable experience that I will always be grateful for. I thank Dr. Cali for providing me the opportunity to assist him in a research project during my first summer. His constant support and motivation will always be appreciated by me.

I would also like to thank my research colleagues and dear friends, especially, Abilash, Jakir, Tumininu, Kaylee, and Rojan. I owe it to them for making my journey through graduate school one that was academically-fulfilling and a joyful experience. Additionally, I would like to thank Schweitzer Engineering Laboratories (SEL) for providing me an opportunity to intern in the final semester. The industrial experience has most definitely been immensely challenging and productive. I greatly acknowledge the Energy Production and Infrastructure Center (EPIC) department and my research lab, Power Energy and Intelligent Systems Lab (PEISL) for the academic and research support.

Ultimately, I would like to thank my ever-supportive family back in India. Living a thousand miles away from home can at times be difficult, but they have always been extremely supportive, financially and emotionally. I owe my success to them and God.

DEDICATION

To my beloved family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CAISO California Independent System Operator.

DA Day-ahead.

DER Distributed energy resource.

DUC Day-Ahead Unit Commitment.

EENS Expected energy not served.

FFS Fast-forward selection.

ISO Independent System Operator.

ITL Information theoretic learning.

LR Lagrangian Relaxation.

MILP Mixed-Integer Linear Programming.

MISO Mid-continent Independent System Operator.

NLP non-linear optimization problems.

NREL National Renewable Energy Laboratory.

PV Photovoltaic.

RES Renewable energy source.

RNG Random number generator.

RUC Risk-based Unit Commitment.

SP Stochastic Programming.

UC Unit Commitment.

CHAPTER 1: INTRODUCTION

"The scientific man does not aim at an immediate result. He does not expect that his advanced ideas will be readily taken up. His work is like that of the planter - for the future. His duty is to lay the foundation for those who are to come, and point the way."- Nikola Tesla

In modern power systems, the uncertain and variable nature of renewable energy sources (RES) pose significant challenges in continuing electricity-market operations in a reliable and cost-effective manner. It is predicted that the penetration of solar power in the electric grid is increasing steadily, and that it could provide as much as 14% of U.S. electricity demand by 2030 and 27% by 2050[1]. Unlike with conventional generation units, the production of renewable energy such as wind and solar power cannot be predicted with perfect accuracy (i.e.uncertainty).

Even with state-of-the-art forecasting methods [2] [3] [4], the predictions seem to become more uncertain as the planning time horizon exceeds a couple of hours. Moreover, if renewable energy production were to be accurately forecasted, generation varies with time (variability) and depending on the weather conditions, these variations can be different from what is predicted by forecasters. As a result, numerous researchers and power system operators alike, have come to appreciate the need for improved power system flexibility, specifically, higher ramping-capability and better resource management, all in efforts to keep up with the unpredictable and rapid variations of the net load (total demand minus renewable production).

Additionally,serving net-loads has become very critical to system operators, since failure to meet these requirements can have undesirable ramifications; power-balance violations, large out-of-market corrections, high volatility of electricity prices, and underutilization of renewable generation. To tackle these challenges in an economically-efficient way, advanced short-term scheduling strategies have been developed that are largely influencing the unit commitment and economic dispatch frameworks. For example, in the United States, several system operators like in Minnesota (MISO) [5] and California (CAISO) [6] utilize the deterministic day-ahead unit commitment (DUC) to schedule generation plants, in which the net load is modeled as a single forecast and the associated uncertainty is handled using ad-hoc reserve requirements which can be fixed during the course of a day or on an hourly basis. Additionally, probabilistic methodologies exist to quantify utility-based requirements for operating reserves under increased penetration of renewable energy.

1.1 Research Objectives

The research objectives of this thesis are the following:

1. To develop a framework for characterizing the uncertainty and variability of photovoltaic generation using historical and prediction data.

2. To develop and propose scenario generation methods based on the uncertainty and variability of photovoltaic generation. Additionally, to apply clustering techniques to sample scenarios based on statistical information of photovoltaic generation.

3. To develop and propose scenario reduction techniques based on probabilistic distances between scenarios.

4. To assess the efficacy of these proposed methodologies using real-world data and under realistic weather conditions utilizing a unit commitment formulation.

1.2 Data Description

In this thesis, we built data-sets using NREL's Solar Power Data for Integration Studies. A solar power plant of 100MW capacity is chosen arbitrarily to represent a PV generation plant. This data contains the actual solar power generation as well as day-ahead forecast data for the year 2006. The forecast data is based on the Weather Research and Forecasting model developed by NREL. We consider hourly data in our analysis, and as such, re-sample data if required using MATLAB scripts.

1.3 Organization of the Following Chapters

This thesis is divided into six individual chapters. 1.1.

Chapter 2 introduces the notion of optimization in power systems, with a special emphasis on the unit commitment problem. The differences in stochastic and deterministic optimization formulations are evaluated. Furthermore, the importance of scenario-based analysis in decision-making processes is discussed. Additionally, a brief introduction and supporting literature review on scenario generation and reduction methods is also presented.

Chapter 3 proposes two methods for scenario generation, one based on uncertainty and variability indices and the other on clustering techniques, both applied to photovoltaic generation in particular, and using historical and forecasted data. These methods employ statistical and mathematical techniques to capture the uncertainty of forecasts and the variability of PV generation in a finite set of scenarios. Illustrative examples demonstrating the usefulness of these two methodologies is presented at the end of the chapter.

Chapter 4 proposes two methods for scenario reduction, one based on a probability distance metric called Kantorovich distance, and the second based on K-means clustering technique. Reduction of scenario-set to a pre-defined cardinality is achieved using both the methods. Ultimately, following the scenarios generated in the example from chapter 3, an illustrative example for scenario reduction using the proposed methodologies is presented.

Chapter 5 presents several case studies demonstrating the application of the proposed methodologies from Chapters 3 and 4 on real-world data described earlier in chapter 1. Furthermore, the final scenario set is then tested on a unit commitment model and a cost-benefit analysis is conducted.

Chapter 6 summarizes the contributions of this thesis, discusses its conclusions, and recommends future work.

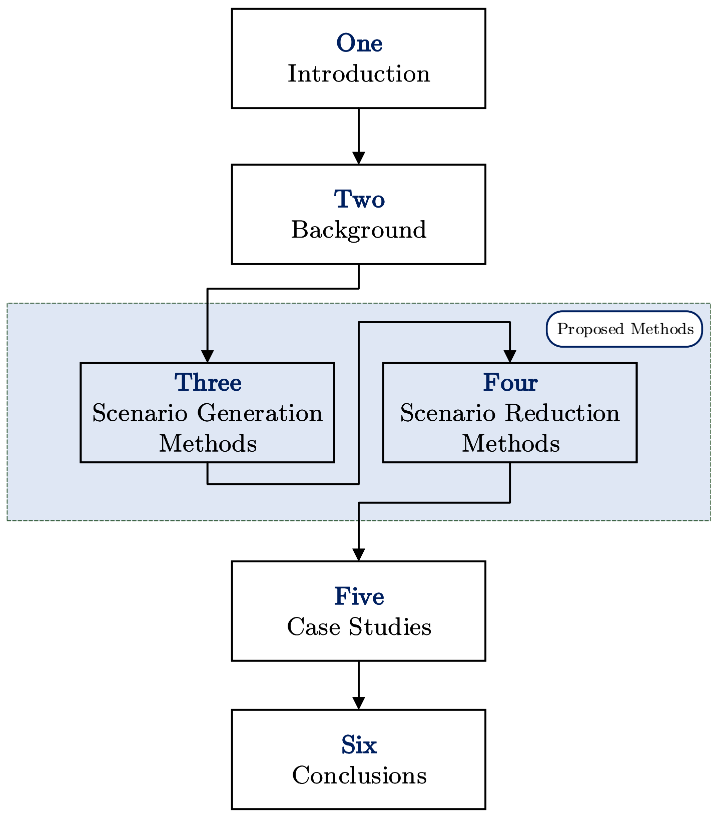

Fig. 1.1: Thesis Outline

CHAPTER 2: BACKGROUND

In this chapter, we provide a background to several concepts and philosophies adopted in this thesis. First, a background to optimization in power systems is presented. Here, we discuss the differences in deterministic and stochastic approaches to solving non-linear global optimization problems. Additionally, a brief introduction to a special optimization problem in power systems called *unit commitment* is provided. Second, a descriptive analysis of scenario generation and reduction is presented. Here, we first discuss how to define the quality of a scenario created. Next, the role of scenario generation and reduction in optimization problems is presented. Additionally, the role of scenario generation in energy markets is examined. Third, an extensive literature review on relevant approaches to unit commitment, scenario generation, and scenario reduction is presented.

2.1 Optimization in Modern Power Systems

In its simplest form, an optimization problem consists of finding the most *optimal* result by maximizing or minimizing a real-valued function. This function is called the *objective function*, and it can be bounded by a set of rules or constraints. Additionally, most real-world optimization problems are non-linear in nature and as such, we will refer to only non-linear functions. Mathematically, non-linear optimization problems (NLPs) are defined as:

$$
\begin{aligned}
\min_x \quad & f(x) \\
& l \leq g(x) \leq u \\
& x^L \leq x \leq x^U
\end{aligned}
\tag{2.1}
$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R} \to \mathbb{R}$,$f : \mathbb{R} \to \mathbb{R}$,$l, u \in \mathbb{R}^m$ are the lower and upper bounds of the constraints, and $x^L, x^U \in \mathbb{R}^n$ are the lower and upper bounds to the variables. The functions f and g are, in general, non-convex.

The nature of the optimization problem can be classified as either *deterministic* or *stochastic*. Let us take a look at the two approaches.

2.1.1 Deterministic Model

In the deterministic model, the optimization problem is to deliver a unique solution for the given set of inputs. Refer to the following figure. Here, the *situation* represents the the real-world problem under consideration. The various assumptions and deterministic information are applied to obtain a mathematical model. Next, an *algorithm* is utilized to find the optimal decision represented by the *decision x*.

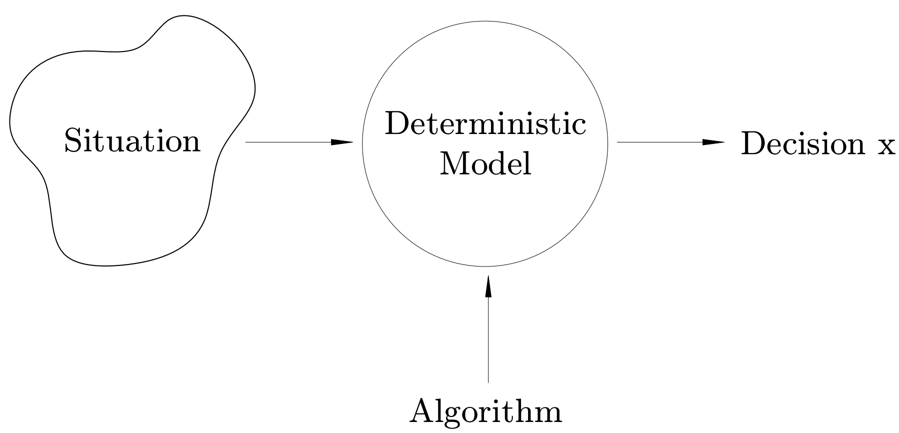

Fig. 2.1: Deterministic approach to decision-making

In this approach, the decision x is optimum for the model, and not the situation. Hence, for real-world problems which have an inherent level of uncertainty, the deterministic approach might not be suitable. To add a little uncertainty or randomness to the decision, a stochastic approach can be used.

2.1.2 Stochastic Model

In the stochastic approach, we have a model describing the process, however, a current representation of the real-world problem is presented and the decision pertaining to this representation are considered in the optimization results. Moreover, the model also provides for several alternative futures, or scenarios. The parameters of each future is deterministic, but they differ from each other.

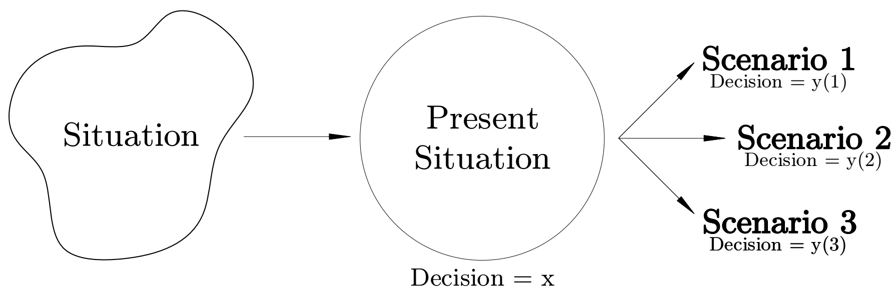

Fig. 2.2: Stochastic approach to decision-making

Hence it is only natural that stochastic methods are the choice of approach for modeling optimization problems that involve uncertainty. As opposed to the deterministic approach, wherein the decision variable has a singular realization as part of the objective function, the stochastic approach considers different scenarios along with their probability of occurrence. It should be noted that the efficiency of stochastic methods highly depends on the quality of sampled points. In theory, uniformly distributed deterministic sequences provide more accurate results than purely random sequences. This plays a huge role in the formulation of a scenario generation method, as you will witness in the next chapter.

2.1.3 Unit Commitment

The *Unit Commitment* problem is an optimization problem of special interest to power system operators. According to Y. Huang et al. [7], it is defined as a combinatorial optimization problem, that given a set of physical conditions such as generation capacity, minimum ON/OFF time, reserve requirements, as well as generation costs such as, startup/shutdown costs, and fuel costs, finds the optimal generation schedule in terms of either cost (*minimize*) or profit (*maximize*).

2.2 Importance of Scenario-based Analysis

Humans strive for a controlled and predictable working environment. Predictability helps us anticipate different possible outcomes and plan accordingly for them. However, while dealing with most natural processes, the outcomes are often unpredictable. Therefore, in such cases, the process of decision-making should be flexible enough to accommodate a certain level of uncertainty. In engineering, problem solvers will often convert a given problem into a mathematical model. Mathematical models capture the nature of the problem, that is, the characteristics, processes, dependencies, and any other useful feature that can be exploited to yield optimal solutions. For example, in the field of optimization, the following framework is used to construct a mathematical model:

- The processes are written as constraints,
- The results are obtained using an objective function,
- And an appropriate solver is used to solve the optimization problem.

The results obtained from such a model are only as good as the inputs given in the first place. That is, it follows a garbage-in-garbage-out scheme. Now, for a problem with absolute certainty - inputs are known with absolute certainty - we more often that not, arrive at an optimal solution. With uncertainty present, however, a deterministic input will often yield a sub-optimal result. Hence, rather than providing only one input, we provide a range of possible inputs within a certain level of confidence, or in other words, we provide different "scenarios" as inputs. By doing so, we take care of a limited amount of unexpected variation in inputs, consequently yielding results that are robust and optimal for the given conditions.

Furthermore, when we deal with optimization problems in power systems with renewable energy sources,the problems are observed to be inherently large in nature and size. The problem becomes even larger when we utilize hundreds of scenarios to represent the uncertainty and variability of renewable sources. While getting an optimal result is important, the usability of the solution is dependent on the timeliness of arriving at the solution, and on ability of the solver to handle the computational burden. For these reasons, and to alleviate computational intractability, it is imperative that we employ scenario reduction techniques.

The most important objective of scenario reduction is to first and foremost, reduce the number of scenarios in a manner that allows us to retain the probability distribution and intrinsic characteristics of the original scenario set. That is, while the number of scenarios is reduced, the reduced set should give a good - if not better - approximation of the uncertainty and variability of the process.

2.2.1 Qualitative Metrics for Scenarios

Zenios in [8] defined three metrics for identifying the quality of scenario generation - Correctness, Accuracy and Consistency. These metrics are explained below:

2.2.1.1 Correctness

- Scenarios should contain properties that are prevalent from the academic research point of view. For example, the term structure should exhibit mean reversion and changes. The term structure consists of changes in level, slope and curvature as examined in academic research.

- Scenarios should also cover all relevant past history. Furthermore, scenarios should account for events that were not observed, but are plausible under current conditions.

2.2.1.2 Accuracy

- As in many cases, scenarios represent a discretization of a continuous process. Accumulating a number of errors in the discretization is unavoidable. Different approaches can be used to ensure the sampled scenarios still represent the underlying continuous distribution function.

- Accuracy is ensured when, for instance, the first and higher moments of the scenarios match those of the underlying theoretical distribution. (Moments and property matching are often used in order to ensure that the scenarios keep the theoretical moments of the distribution they represent).

- The accuracy demand can lead to a large number of scenarios generated. That is, in order to create a fine discretization of the continuous distribution and to achieve the accuracy considered appropriate and acceptable for the application at hand.

2.2.1.3 Consistency

- When scenarios are generated for different test cases, they need to be consistent in capturing the underlying uncertainty.

  For example, in photovoltaic generation, scenarios generated for an overcast day may numerically differ from those generated for, say, a sunny day. However, the robustness and level of uncertainty of the scenarios should be consistent among both days.

2.2.2 Scenario Generation and Reduction in Optimization

An overview of the process of optimization is shown in the following figure.

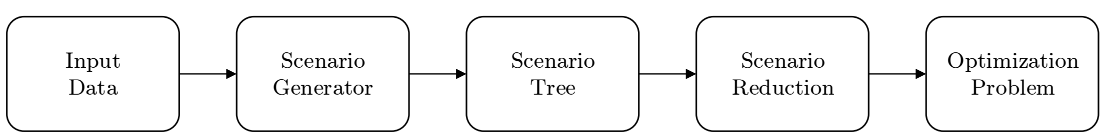

Fig. 2.3: Outline for Scenario Generation and Reduction in Optimization

Scenario generation algorithms aim to capture the underlying uncertainties in an optimization problem. In this section we broadly introduce the different attributes of an optimization process with special emphasis on the role of scenario generation.

2.2.2.1 Input Data

Data that is believed to provide insights into the intrinsic properties of the process are passed to the scenario generator as inputs. This data can be historical data, expert opinions, theoretical relations between input variables, and any other features that are could be important in capturing the uncertainty of the process.

2.2.2.2 Scenario Generator

Next, the scenario generator utilizes specific algorithms to exploit the relationships between the input parameters and capture the underlying uncertainty and variability of the process. Next, the algorithm generates a pre-defined number of scenarios in the form of a scenario tree. Let us now define what is a scenario tree and how it serves as the input to a stochastic optimization process.

2.2.2.3 Scenario Tree

A scenario tree is the collection of all generated scenarios. Each point on the scenario tree is called a node. The tree begins with one node, called the original node. This node is then connected to each node in the next stage. The interconnection of nodes from subsequent time stages ultimately creates a branch, or what we call, a scenario. As an example, consider the following illustration. Here, the scenario tree is comprised of four stages and three scenarios. This particular *fan-type* scenario tree has visibly distinct scenarios across all stages. That is, every node, other than the original node, is connected to only one other node from the previous and subsequent stages. For this thesis, we shall consider fan-type scenarios. Each scenario here is a distinct realization and has a probability of occurrence. Next, we discuss how the number of scenarios can be optimally reduced.

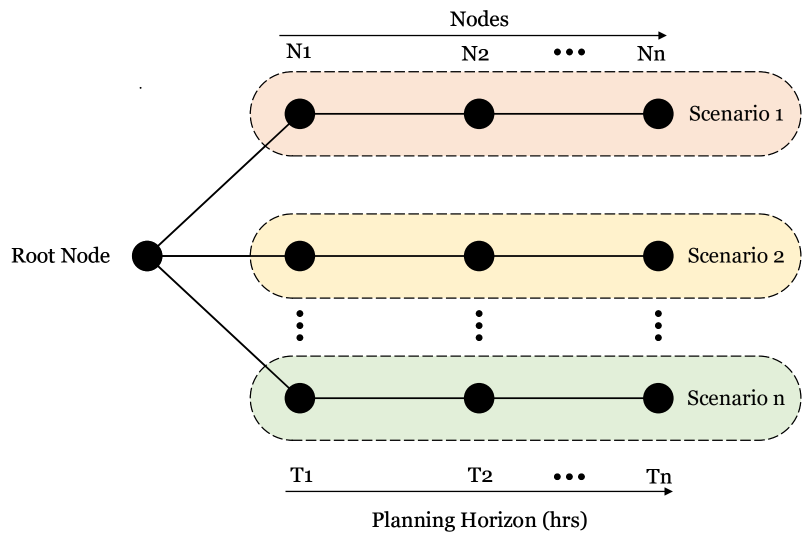

Fig. 2.4: Illustration of a Scenario Tree

2.2.2.4 Scenario Reduction

Here, for the purpose of reducing the computational burden and improving the tractability of the optimization problem, we reduce the number of scenarios to a predetermined number. In order to retain the statistical properties of the original set, we make use of specific probability distances and clustering techniques to preserve the best scenarios. Once we have this set, we solve the optimization problem as described in the next step.

2.2.2.5 Optimization Problem

An optimization problem has an objective function, a set of constraints, and a solver to find the value of the objective function for a given input. In nature, it is quite normal for the environment to change over time. Hence, the solution to a problem now may not be the optimal - or even a good - solution to the corresponding problem in the future. This is the rationale behind using stochastic optimization. The scenario tree, as described in the previously, represents many possible future outcomes along with their probabilities of occurrence, and the stochastic optimization problem takes this scenario tree as an input and computes the cost function for each scenario in a deterministic manner. finally, the scenario with the least cost is chosen as the optimal result.

2.2.3 Role of Scenario Generation and Reduction in Energy Markets

High levels of renewable energy penetration pose significant challenges to the operation, scheduling, and planning of modern power systems. Conventional generation plants - for example, nuclear plants - are relatively easier to schedule, as their generation is controlled and hence predictable. By nature, renewable energy sources (RES) are highly intermittent and uncertain. This makes it increasingly difficult for system operators to accurately define their dispatchability on the electric grid. Failure to accurately predict necessitates the use of other generation units with higher ramping capabilities or increasing reserves allocations in the conventional generation units. This increases the overall operation costs. An important step toward overcoming these challenges is to model the uncertainty and variability of RES, and thereby improve their predictability in the immediate future (for example, day-ahead and four-hour ahead predictions). By doing so, system operators can mitigate as much corrective generation as possible; this is reflected in the energy imbalance services in the energy markets which take care of hourly deviations between advance generation schedules and real-time dispatch. To tackle this, one suitable approach is to create multiple generation schedules based on day-ahead (DA) forecasts and historical measurement data, that is, by the process of scenario generation.It should be noted that scenario generation methods must jointly address the following two criteria:

- Size of the scenarios set should be large enough to retain statistical properties of the stochastic process. (scenario generation)
- Cardinality of the scenarios set should be small enough to preserve the computational tractability of the stochastic problem. (scenario reduction)

2.3 Literature Review

This section presents a literature review of the traditional deterministic as well as stochastic unit commitment formulations, scenario generation techniques, and scenario reduction techniques.

2.3.1 Unit Commitment

Over the years, the unit commitment formulation has evolved in its general effectiveness and complexity. Initially, the UC problem was solved by *Exhaustive Enumeration*. This method comprised of enumerating all the possible combinations of generating units, followed by a search for the most cost-effective combination. In [9] and [10], Kerr *et al.*, and Hara *et al.* respectively solved the UC problem successfully for the Florida Power Corporation by using this technique. However, this method is limited in terms of scalability, in that, it is not suitable for larger-sized utilities. Another approach was to use a technique called *priority listing*, in which the generating units were arranged based on their operational-cost characteristics. Then, this predetermined order was sequentially used to increase or decrease the generation capacity until the system load requirement was reached. Leading pioneers of this method were Burns *et al.* [11] and Lee [12].Furthermore, Lee *et al.* in [13] solved the single and multi-area UC problem using priority listing with a classical index. Soon engineers realized that these methods were becoming extremely tedious for larger systems and hence started to adopt optimization techniques to solve the UC problem. *Dynamic Programming* became mainstream among electric utilities around the world due to its simplicity and effectiveness in solving the UC problems for a variety of system sizes [14] [15]. In 1971, Happ [16] discussed the use of personal-computers (PCs) to solve the hourly UC-problem using dynamic programming. Hobbs *et al.* [17] developed and implemented a realistic UC model for an energy management system. Later, Li *et al.* [18] developed a novel UC formulation that schedules the generation units in a *de-commitment* manner. That is, initially all the active units are committed in the schedule, and as the demand changes, a single unit de-commitment is implemented via dynamic programming.

Integer and Linear Programming (LP) was the next class of techniques that became a popular approach to the UC problem. Dillon *et al.* [19] extended the branch-and-bound method to developed an integer-programming formulation of the UC problem. Here, the objective function was partitioned into a pure integer non-linear component and a non-linear dispatch problem. Unlike in mixed-integer linear programming (MILP) methods [20, 21, 22, 23, 24], the LP method is solved by either decomposing the whole problem into sub-problems, or by simplex techniques [25]. The *Branch-and-Bound* method was presented by Lauer *et al.* and Cohen *et al.* in [26] [27] respectively. Here, a repetitive elimination of subsets of the solution space, followed by creation of upper and lower bounds is used to arrive at the solution.

Another popular methodology that is adopted by some utilities [28] [29] [30] is based on the *Lagrangian Relaxation* (LR) method. Using the LR method, we include the constraints within the objective function and thus make the problem "relaxed", that is, an un-constrained problem. This is achieved by multiplying the constraints with multiplies - called as Lagrangian multipliers - and including them in the objective function. Merlin *et al.* in [31] presented a new method based on the LR approach.

A more intuitive class of methodologies include the *Expert Systems*. Here, expert domain-knowledge is used to improve the inferences in the problem solution process. In the case of UC, Mokhtari *et al.* in [32] adopted this technique to develop an expert-system based consultant for assisting generation scheduling. With the introduction of cognitive-based approaches, two classes of techniques, *Fuzzy Systems* [33] [34] [35] [36] and *Artificial Neural Networks* [37] [38] were used to address the UC problem when the load demand and generation variables were inaccurately known. Additionally, more complex methodologies that combined various models were later adopted.

Some examples are given in [39] [40] [41].

Stochastic Programming-based Unit Commitment has become increasingly popular due to the introduction of uncertainty and variability caused by renewable energy sources in the utility-scale generation. Generally, there are four classes of UC models that have been studied and developed to tackle the uncertainty of such sources. The *scenario-based stochastic programming* technique found in [42] [43] incorporates the dispatching of multiple wind power realizations scenarios with day-ahead scheduling. Next, the *robust UC model* found in [44] [45] minimizes the schedule cost of the worst-case scenario. The *chance-constrained UC model* in [46] sets the constraints with stochastic variables using probability limits at which the constraints should hold. This ensures that real-operation constraint violations are less probable to occur. Finally, *risk-based UC (RUC) model* are often used to consider the operational risks such as, expected energy not served (EENS) and wind power curtailment, in the objective function and the constraints of the UC model.

2.3.2 Scenario Generation

Scenario-based analysis has been popular across various engineering disciplines in testing and analyzing the response of a system to multiple scenarios. An important characteristic of credible scenario generation techniques is the ability to model the stochastic processes, and capture the uncertainty and variability associated with it. Several works addressing this problem can be found in the literature, however, we will focus of the applications pertaining specifically to power systems. We will now briefly discuss few of these techniques.

Time-series models are often used for representing natural processes. Once a model is built, a set of parameters are estimated to fit the historical data available. These models are then discretized and simulated to generate scenarios. Continuous time-series models are often used to represent processes like diurnal production of photovoltaic power and wind energy [47, 48, 49, 50]. Additionally, we can find application of historical data analysis for scenario generation in the works on Konno *et al.* [51] and Young [52]. *Bootstrapping* - a re-sampling technique in statistics - is widely used for generating scenarios [53, 54]. It is performed by re-sampling and replacement of the same data that was previously available. This technique is simple as it only used available data and does not require any mathematical models. Furthermore, due to extensive use of historical data, the correlation between the parameters are preserved. However, this technique heavily depends on the quality of the initial sample, and as such may not accurately capture the uncertainty while generating scenarios.

Conditional Sampling is one of the most common methods for generating scenarios. At every node of a scenario tree, we sample several values from a stochastic process. This is done either by sampling directly from the distribution of the stochastic process, or by evolving the process according to an explicit formula.[55] Traditional sampling methods can sample only from a uni-variate random variable. When we want to sample a random vector, we need to sample every marginal (the uni-variate component) separately, and combine them afterwards [56] [57]. *Moment-matching* is another method typically used when the distribution functions of the marginals are not known [57, 56, 58]. Using this method, we define the marginals in terms of their moments - mean, variance, skewness, kurtosis, etc. - and additionally define the correlation matrix. A well-known set of approaches pioneered by Dupačovã are classified as *path-based methods* [59]. Here, each scenario is created completely, in parallel to each other forming a fan-type scenario set. Then, the fan is converted to a scenario tree by clustering techniques. An example of this can be found in [60].

2.3.3 Scenario Reduction

From a power system operations standpoint, optimal reserve allocations plays an important role in tackling uncertainty introduced by renewable sources such as PV and wind energy. From the previous sections, we have seen how scenario-based unit commitment formulations have contributed to this problem. Moreover, it has been seen that generating more scenarios leads to better approximation of the uncertainty in the original stochastic process. This, however, may lead to computational intractability and hence scenario reduction techniques has been widely investigated in the past decades [61] [62] [63] [64] [65].

In general, scenario reduction is a probabilistic way of retaining scenarios which are representative of the entire scenario set. That is, in reducing the scenario set, we aim to preserve most of the characteristics from the original set in the reduced set [66]. One class of methods include forward selection and backward reduction-based greedy algorithms, first introduced by Dupačovã *et al.* in [67]. Later, Nicole *et al.* proposed a scenario tree construction and reduction framework for improved computational performance in [61]. With the introduction of an optimization platform called *GAMS*, Nicole adapted their method in a routine called *scenred* which made use of state-of-art optimization solvers for even more improved computational performance [68]. A majority of the methods are based on the *Fast Forward Selection* (FFS) method [68] which is based on quantification of pair-wise distances between scenarios. For example, Feng *et al.* [69] proposed a solution-sensitivity based scenario reduction technique that begins with clustering of scenarios and them sampling each cluster centroid with FFS. Additionally, Gomez-Martinez [70] presented a case-study in the UK adopting this technique. Building upon this work, Sumali *et al.* [71] proposed a clustering-based scenario reduction technique using an information theoretic learning (ITL) mean shift algorithm. However, one of the most elegant clustering technique is based on K-means clustering [72] and its advance variations [73, 74].

2.4 Summary

This chapter has presented an elaborate background to the concepts of scenario generation and reduction, along with an insightful discussion on optimization in power systems, specifically on the unit commitment problem. First, a brief introduction to optimization in power systems was presented, differentiating between stochastic and deterministic approaches for problem-solving. Furthermore, a brief introduction to unit commitment was presented along with differences in deterministic and stochastic formulations. It is evident from our discussions and the literature review that stochastic UC formulations are better suited for capturing the uncertainty and variability in decision making involving renewable energy sources. We then discussed about scenario-based analysis and its role in improving stochastic solutions. This was followed by a detailed review of past literature on scenario generation and reduction techniques. In the next chapter, we take a look at the proposed methodologies for generating credible scenarios for photovoltaic generation.

CHAPTER 3: PROPOSED METHODOLOGIES FOR SCENARIO GENERATION

In this chapter, two methods for scenario generation are proposed, one based on the concept of probability distances, and the other, based on clustering techniques. Several approaches to generate scenarios were discussed in chapter 2. However it can be noted that majority of the work concerns wind-based systems. In this work,we utilize indices to quantify uncertainty and variability of renewable energy sources and apply them to photovoltaic-integrated systems.

We begin with a brief discussion on the concept of forecast error, followed by the construction of uncertainty and variability indices. Next, we develop algorithms and sub-routines for the proposed methods. Once explained in detail, an illustrative example is then provided for the reader to gain a practical understanding of the methodologies.

3.1 Definitions

We shall now define some important terms before presenting the proposed methodologies.

3.1.1 Forecast Error

In its most generic form, the forecast error is defined as the difference between the forecasted value and the measured value in the future. Mathematically this is written as,

$$(f.e.) = x_{forecasted} - x_{measured} \tag{3.1}$$

The forecast error can either be positive or negative. That is,

$$f.e. = \begin{cases} positive & \text{for} \quad x_{forecasted} >= x_{measured} \\ negative & \text{for} \quad x_{forecasted} < x_{measured} \end{cases} \tag{3.2}$$

However, in this work, we consider only the absolute values as we only model the magnitude of the error terms.

3.1.2 Uncertainty and Variability Indices

Renewable energy sources inherently are know for their unpredictability and rapid variations.we will know define two terms that are key to modeling variable generation processes like in PV and wind generation.

Uncertainty is defined as the forecast error, that is, the different between the forecast and actual value for time.

$$uncertainty = x_{forecasted,t} - x_{measured,t} \tag{3.3}$$

If the forecast error is large, it implies that the uncertainty in predicting those values is greater. The converse is also true.

Variability is defined as the difference between the rate of change of an entity from the current hour to the next hour. It is defined as,

$$variability = x_{measured,t} - x_{measured,t+1} \tag{3.4}$$

In other words, variability could be interpreted as the ramp rate of the generation processes. Variable generation like solar and wind generally have high ramp rates as they are heavily dependent on weather fluctuations, which historically are known to be very variable.

Now that we have defined uncertainty and variability, we present indices on a scale

Table 3.1: Uncertainty Indices

Index	Value	Description
u_1	0.2	predictable
u_2	0.4	mostly predictable
u_3	0.6	uncertain
u_4	0.8	moderately uncertain
u_5	1.0	highly uncertain

Table 3.2: Variability Indices

Index	Value	Description
v_1	0.2	constant
v_2	0.4	mostly constant
v_3	0.6	variable
v_4	0.8	moderately variable
v_5	1.0	highly variable

of 0 to 1 for each parameter as shown in tables 3.1 and 3.2, respectively.

3.1.3 Error Terms

Based on a general analysis of the historical dataset, we create a set of error terms to define the magnitude of forecast error.

Table 3.3: Error Terms

Term	Magnitude (p.u.)
e_1	0.05
e_2	0.1
e_3	0.2
e_4	0.25
e_5	0.35

3.1.4 Clustering

Clustering or cluster analysis is the process of grouping a set of items in a manner that similar items are in the same group. The 'similarity' of items is decided based on different metrics, depending on the given problem.Alternatively, it can be described as the method of grouping items that are relatively more similar to each other than to items in other groups. Clustering is a form of unsupervised learning and it primarily focuses on finding some structure in a collection of unlabeled data.

3.1.5 Centroid

Each of the clusters generated from cluster analysis has a central point known as the centroid of the cluster. That is, the centroid is the point having the lowest distances from all other points within the cluster. Moreover, the centroid can be considered as the representative point of a cluster.

3.1.6 Squared Euclidean Distance Measure

The squared Euclidean distance (d) measure is the square value of the euclidean distance between point a and point b. The formula is given as:

$$d = (a_1 - b_1)^2 + (a_2 - b_2)^2 + ... + (a_n - b_n)^2 \tag{3.5}$$

This distance reveals the proximity or *nearness* of scenarios either within a cluster or with respect to a central scenario.

3.2 Methodology I: Uncertainty and Variability Indices-Based Scenario Generation Method

In this section we introduce a methodology based on the modeling of uncertainty and variability of photovoltaic generation. This method based on modeling of inherent stochastic dependencies of solar energy by creating a set of indices to capture the uncertainty and variability of diurnal (during the day) PV generation. This method can be broadly divided into four distinct steps, data collection, computation of forecast error, assignment of uncertainty and variability indices to data, and finally scenario generation. The flowchart in figure 3.1 summaries the methodology.

3.2.1 Data Collection

In order to create PV generation scenarios, we first collect PV irradiance and PV power data. As described in 1.2,real-data is collected from NREL's Solar Power Data for Integration Studies database. In this dataset, PV irradiance and PV power in 5-min resolution are provided for the year 2006 for different generation capacity plants all over the United States. In this thesis, we consider a 100MW plant located in the state of North Carolina as our test plant. Additionally, this data consists of historical forecasted data as well as measured data, both of which is used as input data for the process of scenario generation.

3.2.2 Computation of Forecast Errors

Next, we compute the difference between the historical forecasted values and measured power values.This gives us the forecast error.

3.2.3 Assignment of Uncertainty and Variability Indices to Data

Before we assign these indices to the historical forecast error data, estimation of uncertainty and variability for each data point is required. Following the definitions provided in 3.1 we assign an index value for each data point in the historical data set.

Now that every data point has an uncertainty and variability level, we calculate

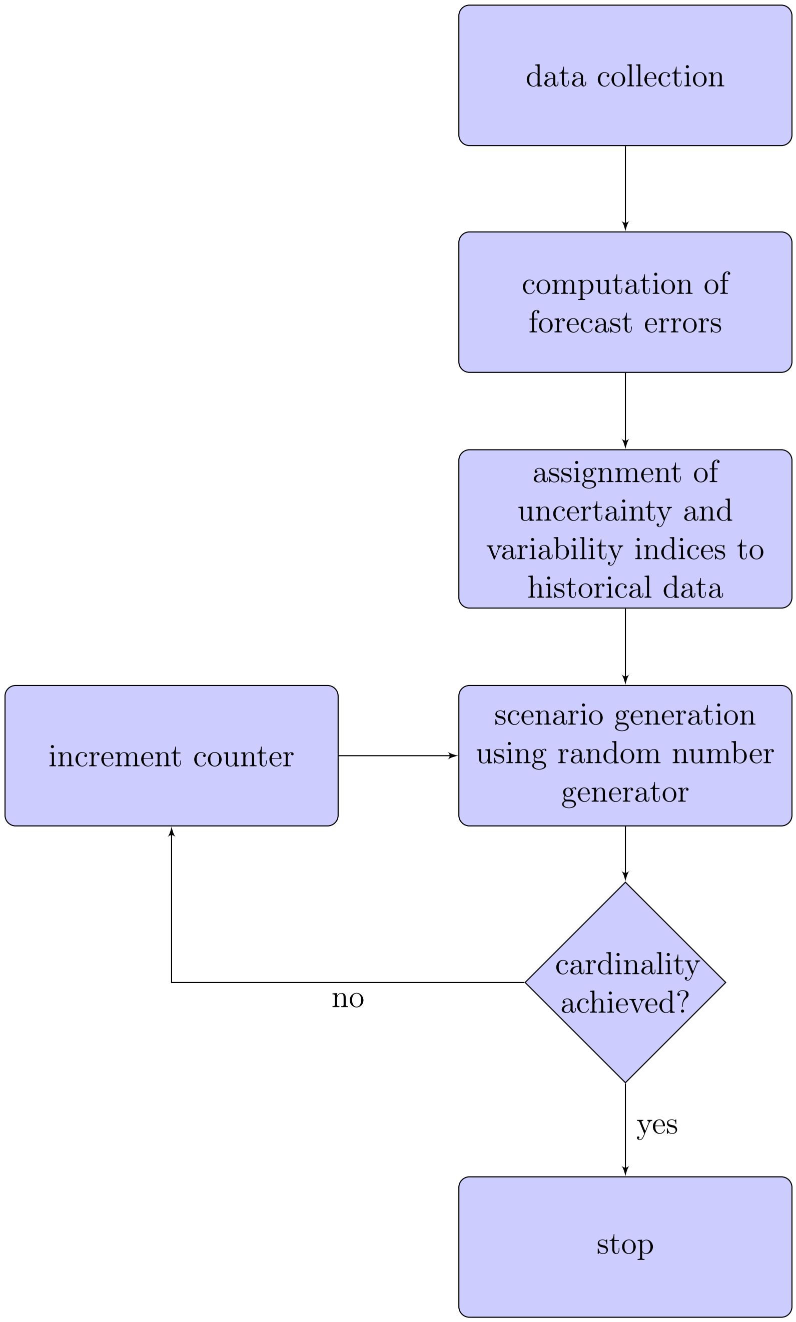

Fig. 3.1: Flowchart for uncertainty and variability indices-based scenario generation method

the year-long average of hourly uncertainty and variability levels prior to the day of operation. This helps in determining on an average how the uncertainty and variability levels varied during every hour of the day. Since seasonality largely influences PV generation, we restrict this analysis to only three-months prior to the day of operation. That is, we consider a three-month average of hourly uncertainty and variability levels.

The algorithm for assigning uncertainty and variability levels to historical data are given below. Here, we consider a year-long (8760) dataset for the historical data and the indices described in 1.2.

Algorithm 1: Assignment of Variability Levels

```
Result: Variability Index Matrix V(i)
Let  P  → measured PV power matrix
     v_i → variability index
Initialize V = zeros(size(P))
for i = 2 : length(P) do
    if 0 ≤ |P_i − P_{i−1}| ≤ e_1 then
        v_1 → V(i);
    else if e_1 ≤ |P_i − P_{i−1}| ≤ e_2 then
        v_2 → V(i);
    else if e_2 ≤ |P_i − P_{i−1}| ≤ e_3 then
        v_3 → V(i);
    else if e_3 ≤ |P_i − P_{i−1}| ≤ e_4 then
        v_4 → V(i);
    else if e_4 ≤ |P_i − P_{i−1}| ≤ e_5 then
        v_5 → V(i);
    else
        v_6 → V(i);
    end
end
```

3.2.4 Scenario Generation using Random Number Generator

Finally, the process of scenario generation is carried out using a random number generator (RNG) function in MATLAB. First, we define an upper limit and lower limit for the RNG between which a fixed number of random scenarios are generated.

Algorithm 2: Assignment of Uncertainty Levels

Result: Uncertainty Index Matrix $U(i)$

P $\rightarrow$ measured PV power matrix

Let DA $\rightarrow$ historical day-ahead forecast PV power matrix

u_i $\rightarrow$ uncertainty index

Initialize $U = zeros(size(P))$

for $i = 1 : length(P)$ **do**

  if $0 \leq |DA_i - P_i| \leq e_1$ **then**

    $u_1 \rightarrow U(i)$;

  else if $e_1 \leq |DA_i - P_i| \leq e_2$ **then**

    $u_2 \rightarrow U(i)$;

  else if $e_2 \leq |DA_i - P_i| \leq e_3$ **then**

    $u_3 \rightarrow U(i)$;

  else if $e_3 \leq |DA_i - P_i| \leq e_4$ **then**

    $u_4 \rightarrow U(i)$;

  else if $e_4 \leq |DA_i - P_i| \leq e_5$ **then**

    $u_5 \rightarrow U(i)$;

  else

    $u_6 \rightarrow U(i)$;

  end

end

These limits are decided by looking at the uncertainty and variability levels of the hour and the appropriate error terms as described in 3.1. The generic algorithm is as follows.

Once the desired cardinality of the reduced scenario set is achieved, the scenario generation process is halted,or else the generator continues to randomly sample data. A rule-of-thumb for number of scenarios is that approximately 10^s pseudo-random scenarios are typically necessary for statistical credibility, where s is the dimensionality of the model. If the generation of scenarios is quasi-random, this requirement can be far lesser.

Algorithm 3: Scenario Generation using Random Number Generator

Result: Scenario-Set (S)

Let
$\delta \rightarrow$ number of required scenarios
$r \rightarrow$ row
$c \rightarrow$ column
$\Psi_{3mo} \rightarrow$ three-month hourly average of uncertainty levels
$\Phi_{test} \rightarrow$ day-ahead forecast for PV generation

Initialize $S = zeros(size(\delta, 24))$
for $i = 1 : \delta.24$ **do**
  if $r \mod \delta = 0$ **then**
    $r = 1, c = c + 1;$
  else if $\Psi_{3mo}(c, 1) \geq 0$ ***and*** $\Phi_{test}(c) = 0$ **then**
    $S(r, c) = 0;$
  else if $0 \leq \Psi_{3mo}(c, 1) \leq u_1$ **then**
    $S(r, c) = random(\Phi_{test}(c), \Phi_{test}(c) + e1);$
  else if $u_1 \leq \Psi_{3mo}(c, 1) \leq u_2$ **then**
    $S(r, c) = random(\Phi_{test}(c), \Phi_{test}(c) + e2);$
  else if $u_2 \leq \Psi_{3mo}(c, 1) \leq u_3$ **then**
    $S(r, c) = random(\Phi_{test}(c), \Phi_{test}(c) + e3);$
  else if $u_3 \leq \Psi_{3mo}(c, 1) \leq u_4$ **then**
    $S(r, c) = random(\Phi_{test}(c), \Phi_{test}(c) + e4);$
  else if $u_4 \leq \Psi_{3mo}(c, 1) \leq u_5$ **then**
    $S(r, c) = random(\Phi_{test}(c), \Phi_{test}(c) + e5);$
  $r = r + 1;$
end

3.3 Methodology II: K-means Clustering-based Scenario Generation Method

In this section we introduce a methodology based on grouping scenarios that are similar to each other. The similarity of scenarios is determined by observing the standard euclidean distance between each scenario to a base-case scenario (In this case, the day-ahead forecast PV generation profile.) This method can be simplified into the four distinct steps, namely, data collection, computation of forecast-errors, application of K-means clustering to forecast-errors, and finally, centroid-based scenario generation.

3.3.1 Data Collection

In order to create PV generation scenarios, we first collect PV irradiance and PV power data. As described in 1.2,real-data is collected from NREL's Solar Power Data for Integration Studies database. In this dataset, PV irradiance and PV power in 5-min resolution are provided for the year 2006 for different generation capacity plants all over the United States. In this thesis, we consider a 100MW plant located in the state of North Carolina as our test plant. Additionally, this data consists of historical forecasted data as well as measured data, both of which is used as input data for the process of scenario generation.

3.3.2 Classification of data into clusters

Next, apply k-means clustering algorithm to the historical measured data to classify them into clusters. The number of clusters is said to be equal to the cardinality of the scenario-set. Once the data is segregated into clusters, we find the mean, standard deviation and variance of each cluster.

Algorithm 4: K-Means Clustering

Result: Cluster centroids (c) and label for each data-point ($c^{(i)}$)

Let
$$\begin{aligned} \mu_k \in \mathbb{R}^n &\rightarrow \text{initial cluster centroids} \\ x^{(i)} \in \mathbb{R}^n &\rightarrow \text{data points} \\ \text{k} &\rightarrow \text{number of required clusters} \end{aligned}$$

Initialize $\mu_1, \mu_2, ..., \mu_k \in R^n$ randomly

repeat

For every i, set $\quad c^{(i)} = \arg\min_x \left\| x^{(i)-\mu_j} \right\|^2$

For every j, set $\quad \mu_j = \dfrac{\sum_{i=1}^m \{c^i = j\} x^{(i)}}{\sum_{i=1}^m \{c^i = j\}}$

until centroids do not change

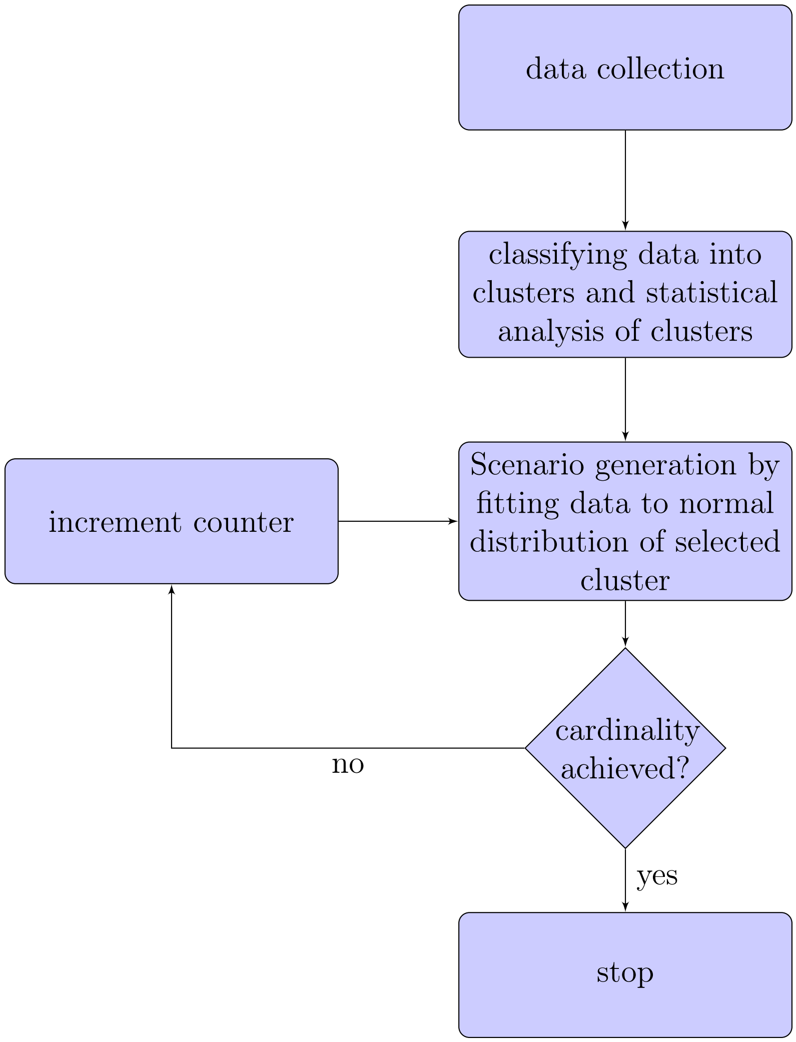

Fig. 3.2: Flowchart for K-means clustering-based scenario generation method

3.3.3 Centroid-based scenario generation

Once the centroids (or error terms) are determined, we first input the day-ahead PV forecast data which is considered as the central scenario. Next, depending on which cluster the day-ahead forecast data-point is closest, we generate the required number of scenarios using a normal distribution. The normal distribution is based on the statistical information derived in the previous step.

3.4 Illustrative Example for Scenario Generation

In this illustrative example, we will consider scenario generation for a relatively small system. The historical data provided is a set of 5 PV generation profiles recorded over 4 time stamps.Also, the historical forecast data for the same profiles are provided. Refer to table 3.4. Additionally, a day-ahead forecast PV generation profile is provided in table 3.5. The goal of this example is to show how scenarios are generated with the aforementioned methodologies.

Consider the following data:

Table 3.4: Historical data for PV Generation

Data-set #	1	2	3	4	5
Measured PV Generation for $t = 1$	0.40	0.45	0.35	0.20	0.50
Forecasted PV Generation for $t = 1$	0.40	0.40	0.40	0.35	0.55
Measured PV Generation for $t = 2$	0.55	0.50	0.60	0.65	0.48
Forecasted PV Generation for $t = 2$	0.50	0.55	0.60	0.60	0.55
Measured PV Generation for $t = 3$	0.60	0.80	0.75	0.55	0.90
Forecasted PV Generation for $t = 3$	0.45	0.70	0.70	0.65	0.80
Measured PV Generation for $t = 4$	0.40	0.55	0.45	0.43	0.35
Forecasted PV Generation for $t = 4$	0.40	0.40	0.45	0.40	0.40

Table 3.5: Day-ahead (DA) PV Generation Forecast

Time stamp #	1	2	3	4
PV Generation	0.4	0.45	0.35	0.5

Table 3.6: Measured PV Generation on Test Day

Time stamp #	1	2	3	4
PV Generation	0.40	0.47	0.40	0.45

Generate 10 scenarios for the day-ahead forecast using:

1. Methodology I (*use error-terms given in table 3.3*)

2. Methodology II (*create 5 clusters*)

Solution:

1. Methodology I: Based on Uncertainty and Variability Indices

   The scenarios are produced using the steps described in flowchart 3.1.

   Step 1: Data collection

   First, we evaluate the given data. In this example, we are provided with 5 sets of PV generation data each for 4 time periods. Second, a day-ahead forecast profile is provided for the same duration. Third, the measured PV generation data on the day of the test is provided for comparison. Additionally, an irradiance forecast is provided for the test day.

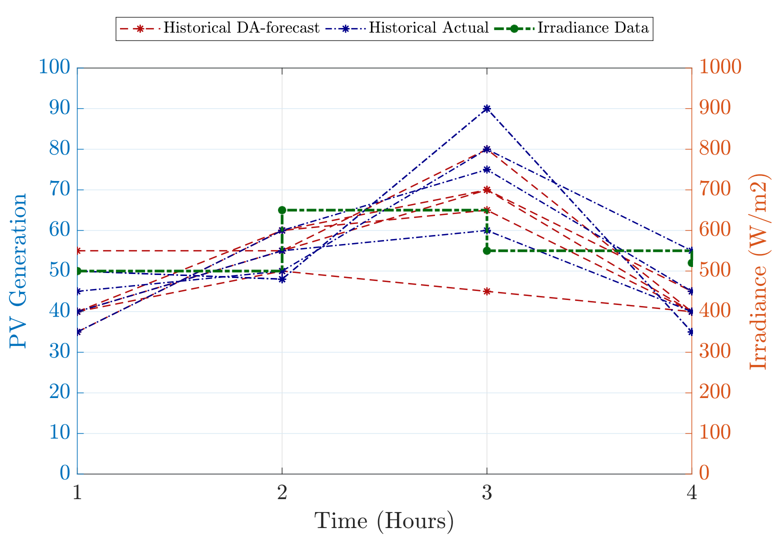

Fig. 3.3: Historical Forecast and Measured Data

Step 2: Computation of forecast error

In this step, we compute the absolute difference between the historical forecasted data and historical measured data. This helps us assess the accuracy of the forecast model used in the given forecasted data, and therefore determine how much weightage should be given to this forecasted data in our model.

The following figure shows the forecast error.

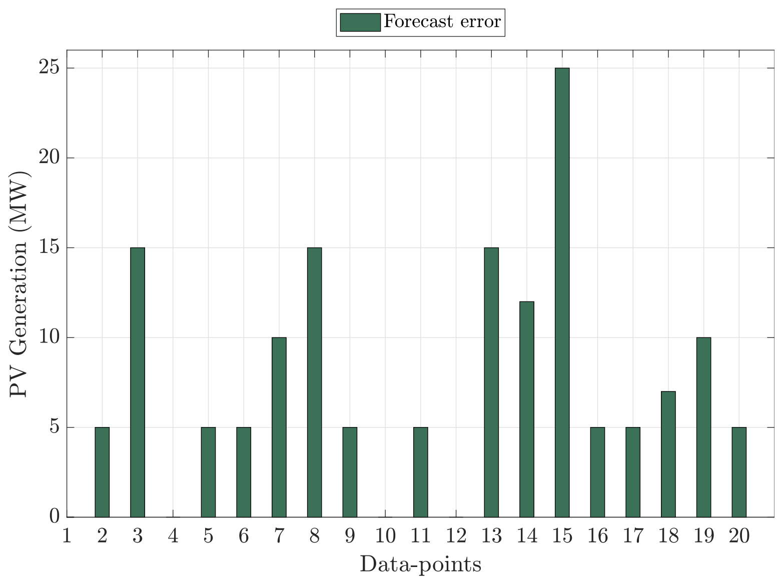

Fig. 3.4: Computation of Forecast Error

Step 3: Assignment of uncertainty and variability levels to historical data

Based on equations 3.3, 3.4 and historical irradiance data, we deduce the uncertainty and variability levels for all data points. In order to do this, we first define the error terms, uncertainty levels, and the variability levels as described in tables 3.3, 3.1 and 3.2.

The following figures show the uncertainty and variability levels for all data points in the scenario set.

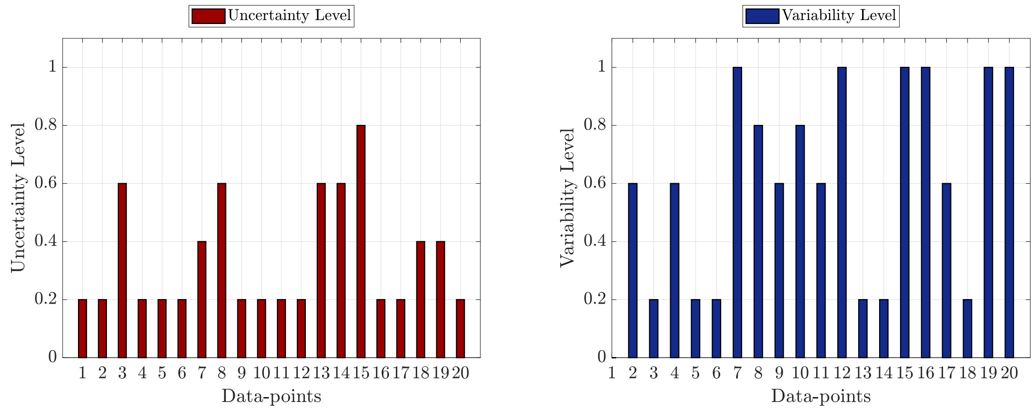

Fig. 3.5: Uncertainty and Variability Levels for Historical Data

Next, we determine the hourly uncertainty and variability levels across the entire time horizon.

Step 4: Determination of Upper Limit and Lower Limit for Scenario Generation

Based on a central PV profile, which in this case is the day-ahead PV generation provided, we determine the upper limit and lower limit for scenarios generated for each time stamp. This is done by following the steps outlined in algorithm 3.The following figure depicts the limits. All the scenarios will be generated within these limits.

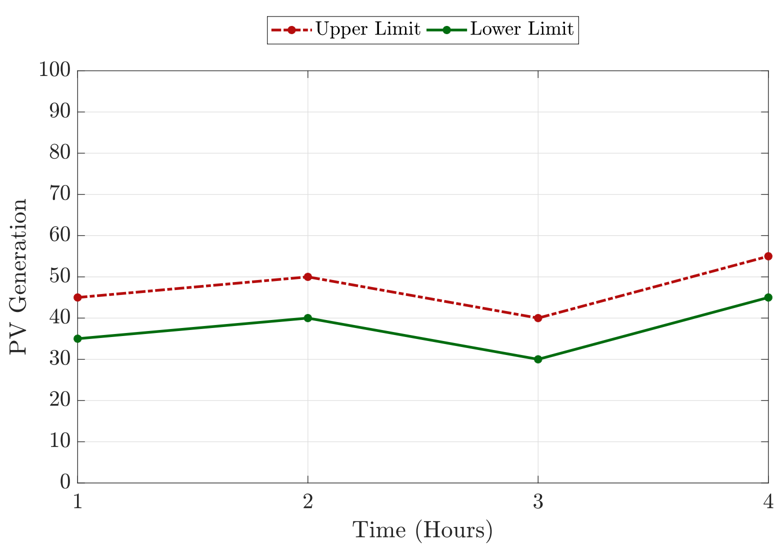

Fig. 3.6: Upper and Lower Limits for Scenario Generation

Step 5: Scenario generation using random number generation

Finally, using the limits generated from step 4, we generate 10 scenarios using the random number generator (RNG) function in MATLAB. Figure 3.7 shows the 10 scenarios generated around the central day-ahead forecast.

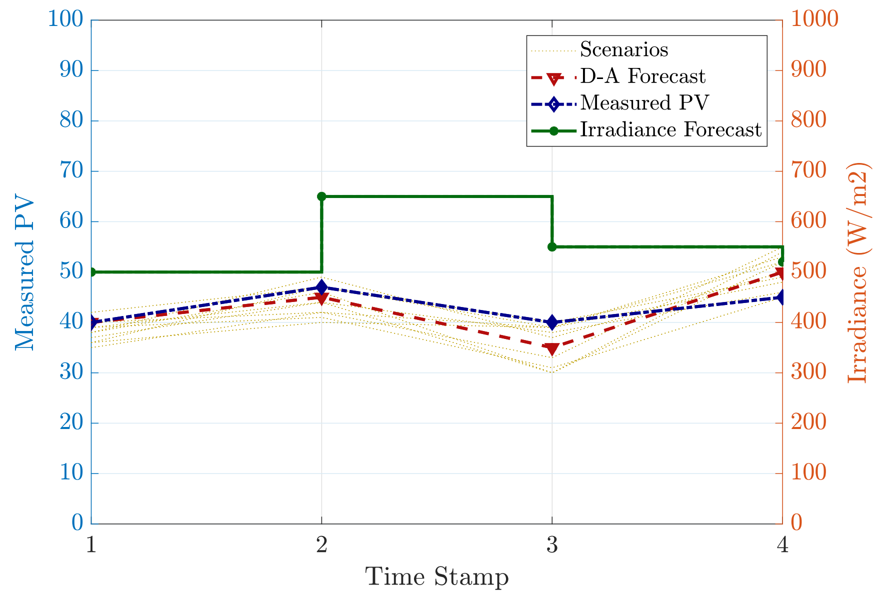

Fig. 3.7: Scenario Generation

Step 6: Cardinality check

Although this step is not explicitly shown, the algorithm for scenario generation mentioned in algorithm 3 ensures that the required cardinality of the scenarios set is reached. Once this is achieved, the process of scenario generation is stopped.

2. Methodology II: Based on K-Means Clustering

   The scenarios are produced using the steps described in flowchart 3.2.

   Step 1: Data collection

   First, we evaluate the given data. In this example, we are provided with 5 sets of

PV generation data each for 4 time periods. Second, a day-ahead forecast profile is provided for the same duration. Third, the measured PV generation data on the day of the test is provided for comparison. Additionally, an irradiance forecast is provided for the test day.

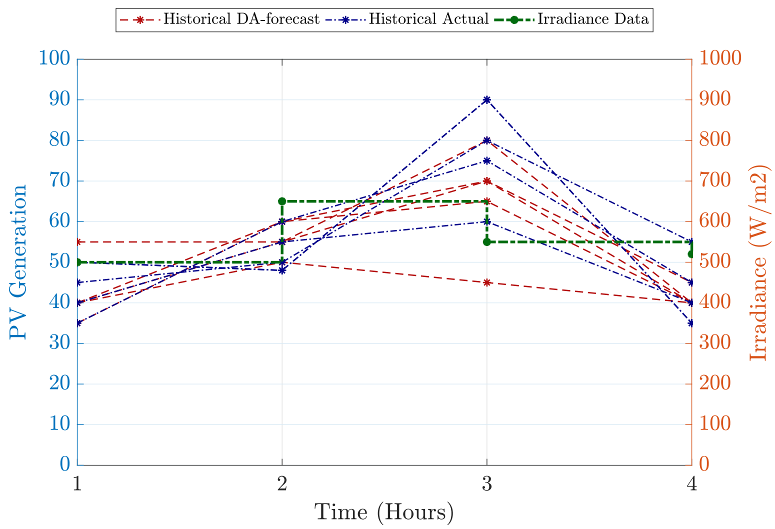

Fig. 3.8: Historical Forecast and Measured Data

Step 2: Apply K-means clustering algorithm and determine the clusters

In this step, we apply group the historical measured data into clusters based on the K-means clustering algorithm. Once the clusters are generated, we determine the centroid of each of these clusters.

Table 3.7: Representative Centroids and Cluster data-points

Cluster	Centroid	Data-points						
Cluster 1	37	45	50	45	50	48	50	48
Cluster 2	48	90	90					
Cluster 3	57.5	55	60	55	60			
Cluster 4	77.5	80	75					
Cluster 5	90	40	40	35	35	35		

Step 3: Statistical analysis of cluster data-points

In this step we determine the mean, standard deviation, and variance of each cluster. This knowledge of the distribution of data-points in each cluster will later help us generate scenarios which are statistically similar to those in the clusters.

Table 3.8: Statistical Analysis of Clusters

Cluster	Mean(μ)	Standard Deviation(σ)	Variance(σ^2)
Cluster 1	48	2.24	5
Cluster 2	90	0	0
Cluster 3	57.5	2.88	8.33
Cluster 4	77.5	3.53	12.5
Cluster 5	37	2.73	7.5

Step 4: Scenario generation using probability distribution

Using equation 3.5, we calculate the pair-wise squared euclidean distance between centroids and each data-point in the given day-ahead forecast profile. For every data-point of the day-ahead forecast, we select the the pair with the lowest distance. Next, we consider the statistical properties of that cluster to which the centroid belongs, and generate the required number of scenarios. The scenarios generated follow a normal distribution based on the standard deviation and variance of the selected cluster. The following figure depicts the generated scenarios.

Step 5: Cardinality check

Although this step is not explicitly shown, the algorithm for scenario generation mentioned in algorithm 4 ensures that the required cardinality of the scenarios set is reached. Once this is achieved, the process of scenario generation is stopped.

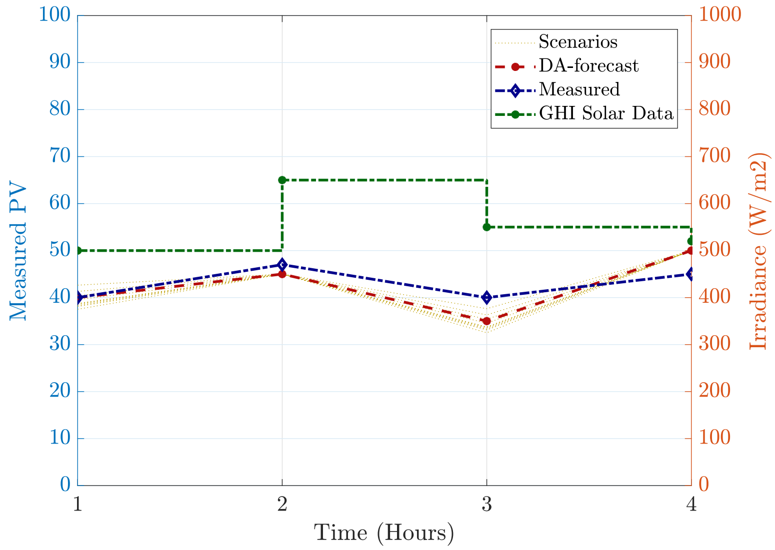

Fig. 3.9: Scenario Generation using K-means Clustering

3.5 Summary

This chapter elaborated on methodologies proposed for generating scenarios using two distinctive approaches: one based on uncertainty and variability indices, and the second, adopting a widely known clustering technique called k-means clustering. These approaches generate scenarios that both, statistical represent the historical data, as well as takes into account the forecasted day-ahead data while generating scenarios. The first method merits in capturing the uncertain and variable nature of PV generation and creates scenarios which are heavily dependent on the quality of PV generation forecast results. The second method has a sound mathematical foundation, in that, a formal approach is utilized in clustering historical data. Furthermore, the scenario data-points are generated based on statistical properties of the clusters they belong to, thereby rendering the scenarios similar characteristics. In the next chapter, we introduce the concept of scenario reduction and in a similar way, propose methodologies to reduce the cardinality of the scenario set with the intent to improve computational tractability.

CHAPTER 4: PROPOSED METHODOLOGIES FOR SCENARIO REDUCTION

In this chapter, two methods for implementing scenario reduction are proposed. The first method is based on the closeness of scenarios within the initially generated scenario set.

We begin by defining the important terms used in the chapter. This is followed by an elaboration on each of the two proposed methodologies. In each section, the steps are detailed with relevant algorithms and diagrams. Finally, an illustrative example is provided for a relatively small problem to gain insights into the methodologies.

4.1 Definitions

We shall now define some important terms before presenting the proposed methodologies.

4.1.1 Kantorovich Distance

The Kantorovich or *Wasserstein* distance is a non-parametric distance between probability measures, and is defined between two probability distributions Q and Q'. It is obtained by assigning the probabilities of non-selected scenarios $\omega \in \Omega, \Omega_S$ to the closest scenario ω' in the selected scenario set Ω_S and can be expressed as[75]:

$$D_K(Q, Q') = \sum_{\omega \in \Omega, \Omega_S} \pi(\omega)(\|y(\omega) - y(\omega')\|) \tag{4.1}$$

4.1.2 Euclidean Distance Matrix

Euclidean distance matrix is an nxn matrix representing the spacing of a set of n points in Euclidean space. If A is a Euclidean distance matrix and the points are defined on m-dimensional space, then the elements of A are given by:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{4.2}$$

4.2 Methodology I: Probability Distance-based Scenario Reduction

This method is based on the *closeness* of the scenarios within the scenario set S generated earlier using the proposed methodologies in Chapter 3. In this method, we observe the probability distances between scenario pairs utilizing a well-known non-parametric measure called Kantorovich distance metric. To explain the methodology in detail, a flowchart (see 4.1) has been created. In the following subsections, we provide a detailed elaboration on each step of the flowchart.

4.2.1 Data Collection

First, we collect the scenarios generated using any given scenario generation method. In this method, all scenarios will be considered equally probable, and hence the probability of all the scenarios will be the same. If there are n scenarios in the scenario set S, and π is the probability of a scenario, then the probability of each scenario is given by,

$$\pi_n = \frac{1}{n} \tag{4.3}$$

Setting all the scenarios with an equal probability of occurrence ensures that non of the scenarios are biased initially. This is important so as to avoid ignoring outliers scenarios.

Additionally, we collect the day-ahead forecast for PV generation at the specific location of the PV plant and day of operation. Also, the cardinality of the reduced set is pre-determined by the user. If m is the number of scenarios generated, n is the number of time stamps, r is the number of scenarios in the reduced set, and the scenario reduction percentage is pre-determined as $s_{\%}$, then the cardinality of the reduced set S_R is given by,

$$|S_R| = s_{\%} * |S| \qquad \forall \quad s_{\%} \in \{0, 1\} \tag{4.4}$$

4.2.2 Computation of Distance Matrix (D)

In order to observe the *closeness* of the generated scenarios, we compute the distance matrix (D) for the scenario set S. The mathematical definition of the distance matrix is given in equation 4.2. The resulting matrix is then used to find the scenario with the lowest pair-wise distances.

4.2.3 Calculation of Kantorovich Distances

In this step, we calculate the Kantorovich distances using the probabilities of the scenarios generated and the distance matrix obtained in step 2. This gives us a measure of the probability distances between all scenario pairs.

4.2.4 Selection of Reduced Set Scenarios

Once we calculate the Kantorovich distances, we then select the scenario with the minimum probability distance as the fist scenario in the reduced set. To do this, we extract the scenario with the lowest Kantorovich distance first. Next, we update the distance matrix to reflect the change in scenario set. We then recalculate the Kantorovich distances among the remaining scenarios. Once we do this, we repeat the step again to choose the next scenario for the reduced set. This step is repeated till the required cardinality of the reduced set is achieved.

4.2.5 Ranking of Scenarios

Once the reduced scenario set is filled with the required number of scenarios, the next step is to rank the scenarios according to their probability of occurrence. It must be noted that we started assuming all scenarios are equally-probable. We know update the probability of the reduced scenarios by transferring the probability of the non-selected scenarios to the selected scenarios closest to them. The algorithm to achieve transfer of probabilities of scenarios and rank them accordingly, is given below.

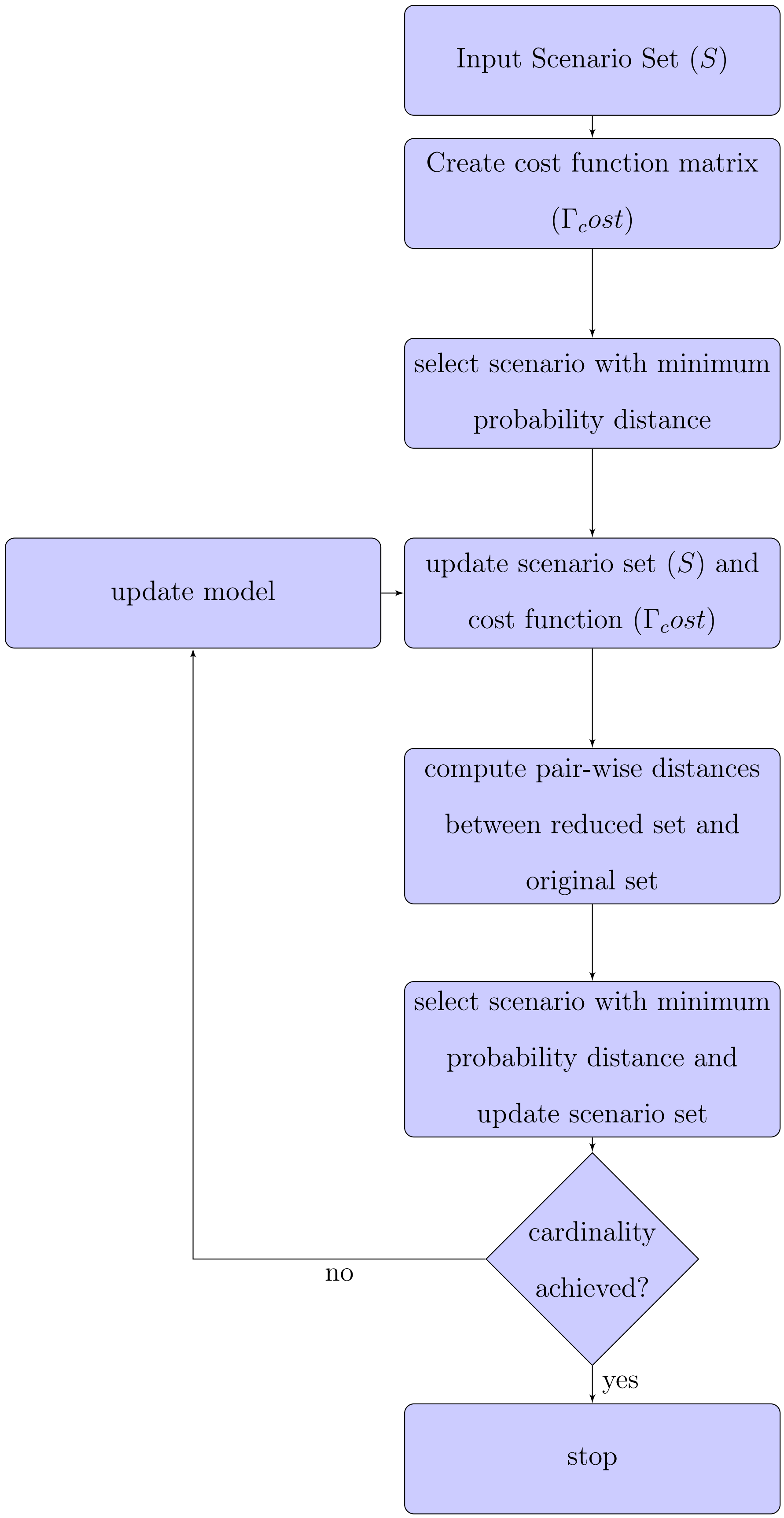

Fig. 4.1: Flowchart for scenario reduction via probability distance metrics

4.3 Methodology II: K-Means Clustering-based Scenario Reduction

In this methodology, we apply the concept of clustering to the generated set of scenarios. The key steps in the methodology are elaborated in the following subsections. We now present a flowchart for the methodology in figure 4.2.

4.3.1 Data Collection

In this step, we collect the scenario set S generated using any of the scenario generation techniques mentioned in chapter 3. Additionally, we input the day-ahead forecast for the day of operation.

4.3.2 Grouping of Scenarios into Clusters

Next, we apply K-means clustering to the entire scenario set and determine the clusters. The number of clusters is equal to the required number of scenarios. Once the clusters are formed, we analyze and extract the statistical information from each cluster, namely, the mean, standard deviation, and variance.

Algorithm 5: K-Means Clustering

Result: Cluster centroids (c) and label for each data-point ($c^{(i)}$)

$$\textbf{Let} \quad \begin{array}{rcl} \mu_k \in \mathbb{R}^n & \rightarrow & \text{initial cluster centroids} \\ x^{(i)} \in \mathbb{R}^n & \rightarrow & \text{data points} \\ \text{k} & \rightarrow & \text{number of required clusters} \end{array}$$

Initialize $\mu_1, \mu_2, ..., \mu_k \in R^n$ randomly

repeat

For every i, set $\quad c^{(i)} = \arg\min_{x} \left\| x^{(i) - \mu_j} \right\|^2$

For every j, set $\quad \mu_j = \dfrac{\sum_{i=1}^{m} \{c^i = j\} x^{(i)}}{\sum_{i=1}^{m} \{c^i = j\}}$

until centroids do not change

4.3.3 Local Ranking of Scenarios within Clusters

Once the centroids (or error terms) are determined, we first input the day-ahead PV forecast data. Second, we calculate pair-wise distance between the day-ahead forecast set and the error terms set.Accordingly, we choose the cluster that is closest to a particular data point. Next, we determine the statistical distribution of the cluster, and generate scenarios or data points from this distribution. Once the desired cardinality is achieved, we stop the process.

4.3.4 Probability Re-distribution within Clusters

In this step, we consider the representative scenario - defined as the centroid of the cluster - to be preserved in the final reduced set. Therefore, the the probabilities of the non-selected scenarios within each cluster are transferred to the representative scenario.

4.3.5 Ranking of Scenarios

Once all the representative scenarios are extracted along with their individual cluster probabilities, the scenarios are ranked according to their cumulative probabilities.

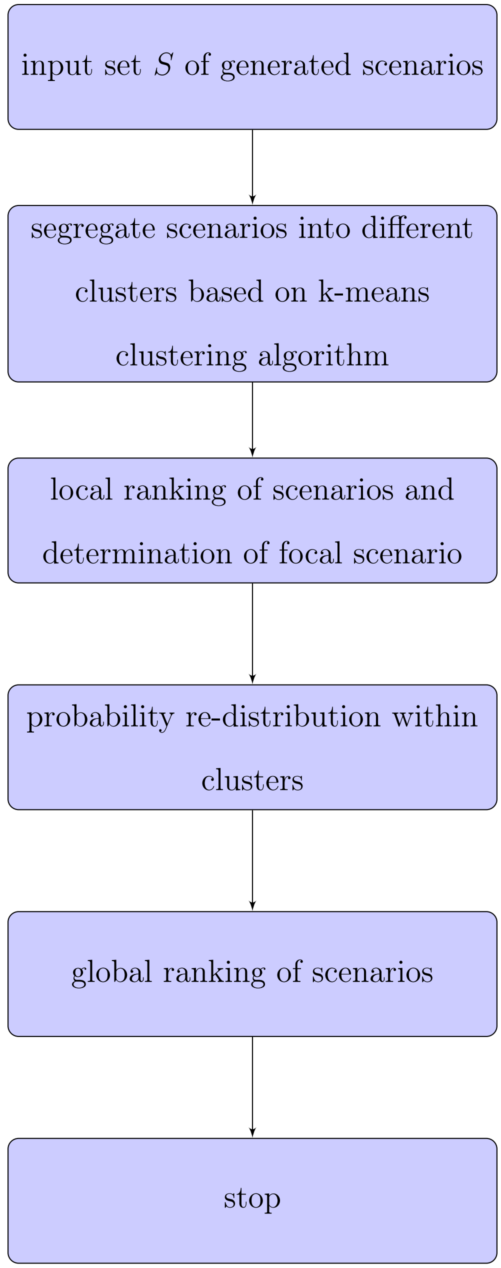

Fig. 4.2: Flowchart for K-means clustering-based scenario reduction method

4.4 Illustrative Example for Scenario Reduction

Consider the following data:

Table 4.1: Scenarios Generated in Chapter 3

Scenario #	1	2	3	4	5	6	7	8	9	10
t_1	39	42	35	38	36	36	37	38	39	40
t_2	44	47	42	49	40	47	44	46	41	42
t_3	38	40	33	37	39	39	30	30	31	39
t_4	46	49	55	50	52	48	52	54	45	53

Table 4.2: Day-ahead (DA) PV Generation Forecast

Time stamp #	1	2	3	4	5
PV Generation	0.4	0.45	0.35	0.20	0.5

Reduce the cardinality of the scenario set to 5 using the following:

1. Methodology I

2. Methodology II

Solution:

1. Methodology I: Based on Kantorovich distance

   The scenarios are produced using the steps described in flowchart 4.1.

   Step 1: Data collection

   The scenarios generated earlier in chapter 3 are considered as in the initial scenario set. Additionally, we consider the the day-ahead forecast profile given in 4.2. Before we begin the scenario reduction process, we define the cardinality of the reduced scenario set. That is, the user pre-determines the reduction percentage. Set $s_{reduction} = 50\%$.

Step 2: Compute the distance matrix (D)

In this step, we compute a distance matrix (D) which is a square matrix (two-dimensional array) containing the distances, taken pairwise, between the elements of the scenario set. This matrix helps us determine the relative closeness of scenarios with respect to each other.

The following figure shows a graphical representation of the distance matrix for this problem.

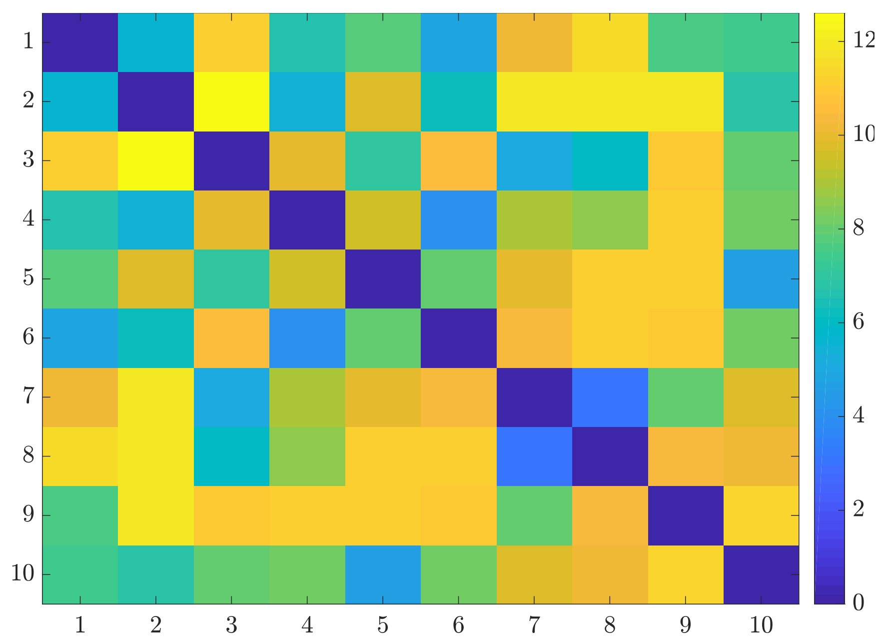

Fig. 4.3: Distance Matrix for Scenario Set (S)

Step 3: Calculate Kantorovich distance and choose scenarios with minimum distance

Based on equation 4.1, we calculate the Kantorovich distances between every scenario pair. Since we begin with hall scenarios that are equally-probable, the probability of occurrence of each scenario is $s_{prob} = 1/10 = 0.1$.

The following figure shows the Kantorovich distances for the first scenario pair.

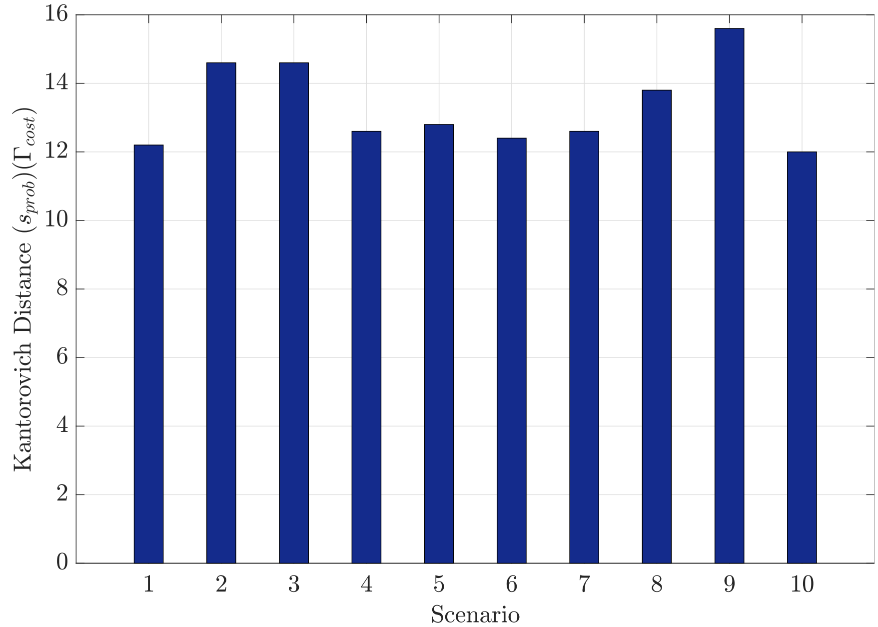

Fig. 4.4: Kantorovich Distances for First Scenario Pair

Step 4: Updating distance matrix and selecting scenarios for reduced set

Once we compute the Kantorovich distances, the next step is to select the scenario with the minimum distance. This is selected as the first scenario in the reduced set. Next, we update the distance matrix. The process is continued till the required cardinality of the reduced set is achieved.

The following graph shows the reduced scenario set. Here the number of scenarios are reduced from 10 to 5. When the cardinality is achieved, the algorithm ends, and the probabilities of the non-selected scenarios get transferred to the those scenarios in the reduced set that are the closest.

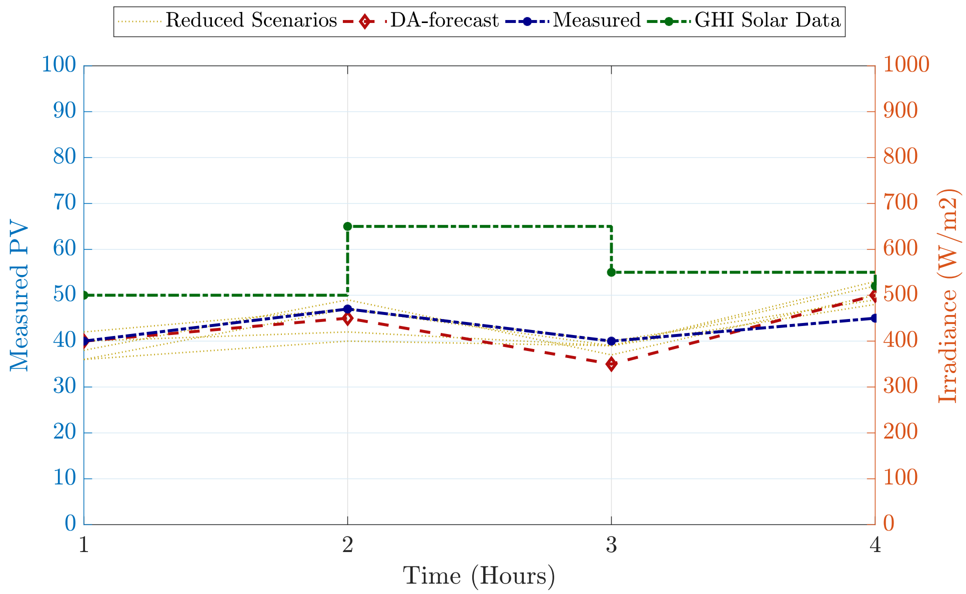

Fig. 4.5: Scenario Reduction based on Kantorovich Distance

Step 5: Rank of scenarios

The scenarios in the reduced set are ranked according to their probability of occurrences.

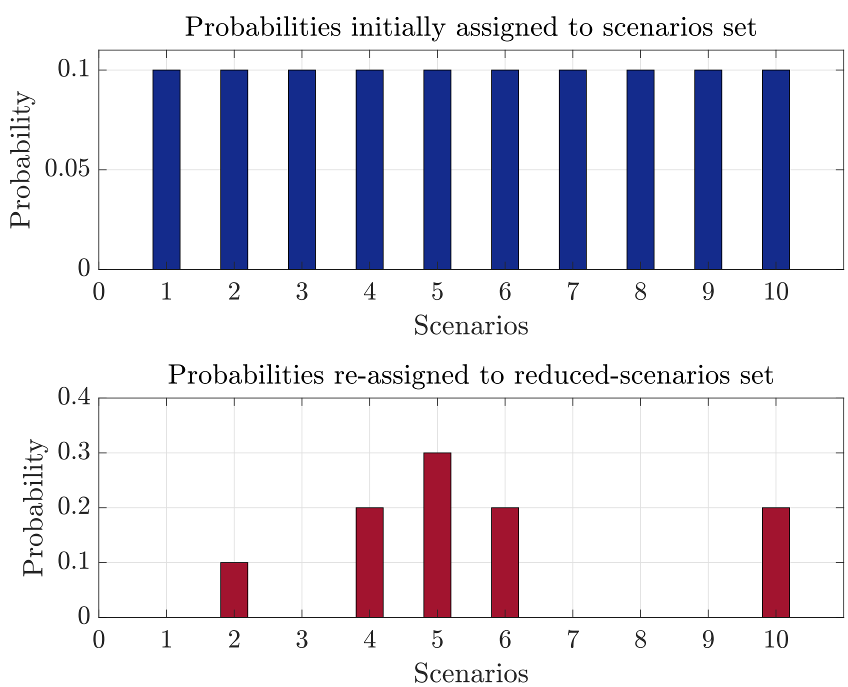

Fig. 4.6: Rank of Scenarios in Reduced Set

2. Methodology II: Based on K-Means Clustering

The scenarios are produced using the steps described in flowchart 4.2.

Step 1: Data collection

The scenarios generated earlier in chapter 3 are considered as the initial scenario set. Additionally, we consider the day-ahead forecast profile given in 4.2. Before we begin the scenario reduction process, we define the cardinality of the reduced scenario set. That is, the user pre-determines the reduction percentage. Set $s_{reduction} = 50\%$.

Step 2: Grouping scenarios into clusters

In this step, k-means clustering technique is applied to the scenario set [74]. This groups similar scenarios into the clusters. Note that K-means clustering technique utilizes a pre-determined number of clusters. Here, the number of clusters is set equal to the cardinality of the reduced scenario set. If α_{cls} is the number of clusters, then $\alpha_{cls} = s_{reduction}$.

The following table shows the distribution of scenarios among 5 clusters.

Table 4.3: Cluster-Grouping of Scenarios

Cluster	Scenarios #
Cluster 1	5 10
Cluster 2	7 8
Cluster 3	3
Cluster 4	9
Cluster 5	1 2 4 6

Step 3: Pair-wise distance between day-ahead forecast and all scenarios within clusters

Next, we compute the pair-wise distances between the given day-ahead forecast and all the scenarios within each cluster. For this method, we use the squared euclidean distance as a measure of distance between the elements of day-ahead forecast set and the scenario set.

Step 4: Local ranking scenarios within each cluster

Based on results obtained in step 3, the scenarios are locally ranked within each cluster based on the pair-wise distance metric. The scenario with the minimum distance receives the highest rank and conversely, scenarios with larger distances receive a lower rank. Once the ranking is completed, the scenarios with the highest rank in their respective cluster are selected and added to the reduced scenario set.

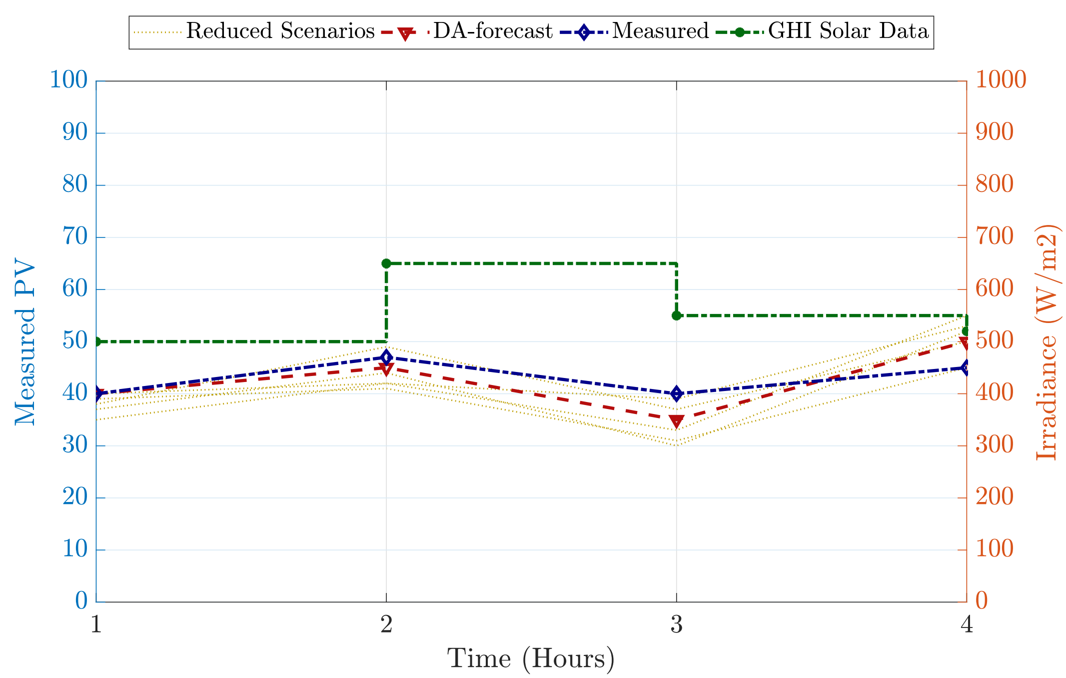

Fig. 4.7: Scenario Reduction based on K-means Clustering

Step 5: Probability re-distribution within each cluster

From step 4, we know which scenarios are preserved in the reduced scenario set. Since we began with every scenario having an equal probability of occurrence, we now transfer the probabilities of the non-selected scenarios to the scenario selected within their respective cluster.

The following figure shows the updated probabilities of scenarios. Note that the probability of non-selected scenarios is reduced to zero, while the scenarios preserved in the reduced set have a probability associated to them.

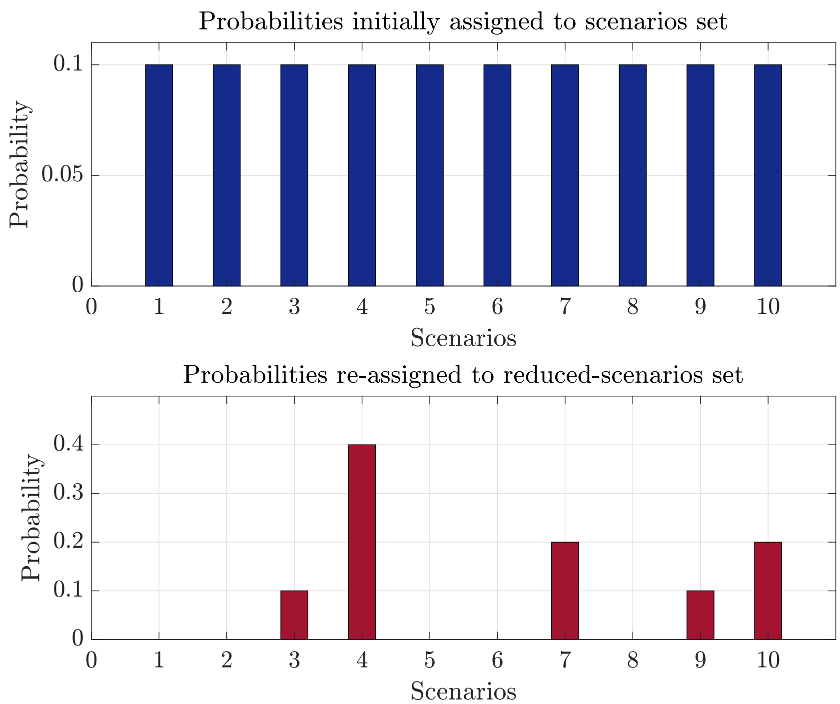

Fig. 4.8: Probability Re-distribution Among Scenarios

Step 6: Rank of scenarios

The scenarios in the reduced set are ranked according to their probability of occurrence. It should be noted that unlike the local-ranking shown in step 4, this ranking is global. That is, the scenarios in the reduced set are ranked according to the cumulative probability of their respective clusters.

Table 4.4: Rank of Reduced Scenarios

Rank	1	2	3	4	5
Scenario #	4	10	7	3	9

4.5 Summary

In this chapter two methodologies for scenario reduction were presented. The first method utilizes a non-parametric probability distance metric called Kantorovich distance to assess the resemblance of scenarios and derive a reduced set that have the minimum probability distance to the original set. The second methodology is based on a clustering technique called K-means clustering. Here, we pre-determine a fixed number of clusters equal to the cardinality of the reduced scenario set, and group similar scenarios based on the squared euclidean pair-wise distance between the day-ahead forecast and the scenarios. In the following chapter we consider numerous case studies to assess the aforementioned proposed methodologies.

CHAPTER 5: CASE STUDIES

In this chapter, several case studies involving adoption of the proposed scenario generation and reduction techniques will be presented. We will first present a study on scenario generation for various pre-defined cardinalities. Furthermore, we will demonstrate four scenario reduction levels and how this affects the ability of scenarios to capture the uncertainty and variability of PV generation. Next, we present case studies demonstrating the application of the proposed scenario generation and reduction methodologies described in chapters 3 and 4 on real-world data described in 1. In order to show the temporal capability of these proposed methods, we tests the methodologies on four days representing the seasonal PV generation in a given year.

5.1 Demonstrations

5.1.1 Scenario Generation using real-world data

In this section, we consider a random day in the yearly data available to us, say, the 52nd day. As depicted in chapter 3, we generate:

- 1000 scenarios,
- 500 scenarios,
- 100 scenarios, and
- 10 scenarios

using the first proposed methodology. The results are plotted in figure 5.1.

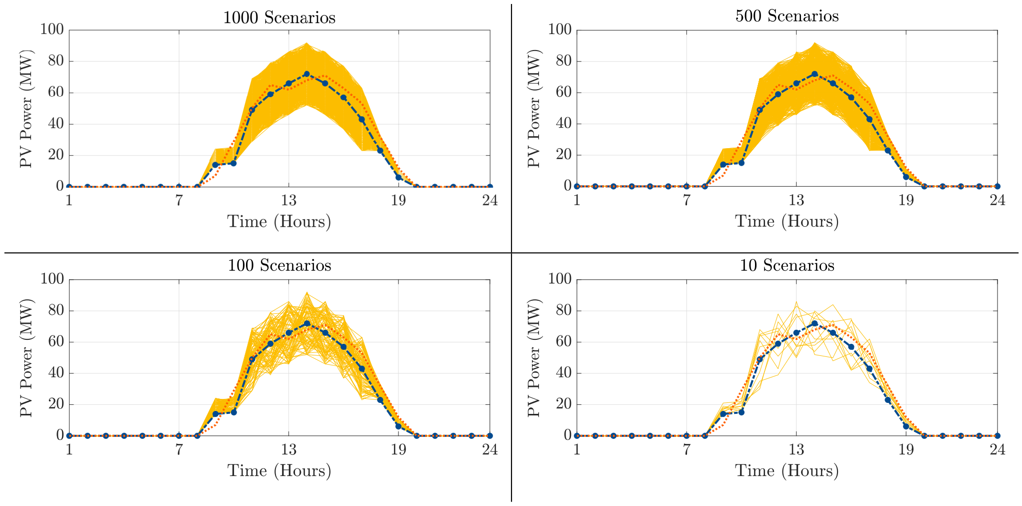

Fig. 5.1: Demonstration of Scenario Generation for different cardinalities

Observations: With increasing number of scenarios, a higher degree of uncertainty and variability is visually seen to be captured. It can be observed that generating 10 scenarios, for instance, generates scenarios that are sparsely distributed and hence does not cover all possible scenarios. However, generating 1000 scenarios can be said to have *over-generated* scenarios, that is, handling too large of a scenario set will adversely affect the computational tractability of the optimization problem needed to solve for these scenarios. Through empirical observations, we have decided to generate 100 scenarios since it not too less to not capture the uncertainty and variability, and not too much to lead to computational intractability.

5.1.2 Scenario Reduction for different reduction percentages

In this section, scenario reduction is demonstrated for four different reduction percentages:

- 75%
- 50%
- 25%

- 5%

Figure 5.2 show the results obtained using methodology II described in 4.2.

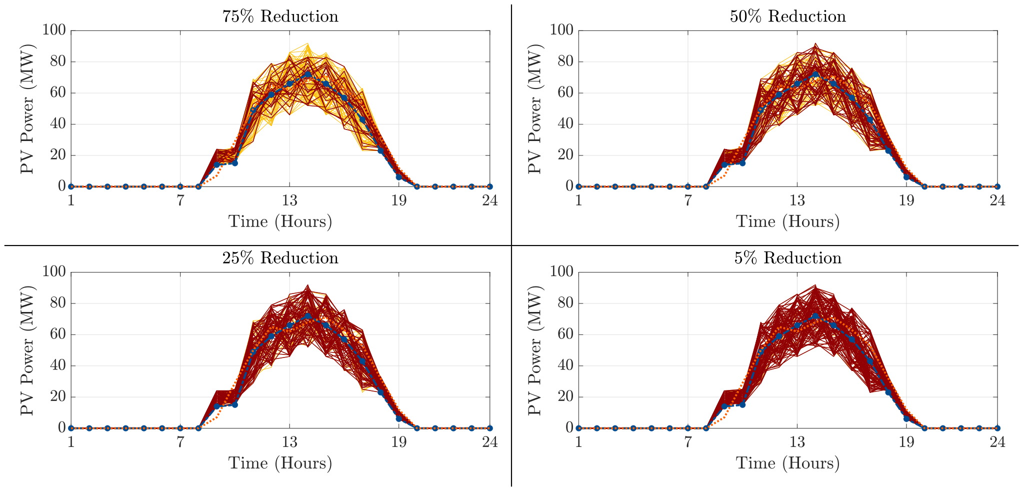

Fig. 5.2: Demonstration of Scenario Generation for different cardinalities

Observations: It can be observed that with increasing reduction in the scenario set, we arrive at a more conservative set of scenarios. While the proposed methodologies that care in preserving outliers in the scenario set - extreme scenarios that may affect the reliability of the system - we need to pre-determined the reduction percentage. It can be observed that for an initial set of 100 scenarios, a 75% reduction gives a set of 25 scenarios with a probability of occurrence attached to each scenario.

For the remainder of the chapter, we will generate 100 scenarios and apply 75% reduction in the scenario reduction process. Additionally, we refer to the proposed methodologies as the following:

Table 5.1: Short-hand Terms for Proposed Methodologies

Term	Method
Scengen Methodology I	Uncertainty and Variability Indices-based Scenario Generation
Scengen Methodology II	Clustering-based Scenario Generation
Scenred Methodology I	Kantorovich Distance-based Scenario Reduction
Scenred Methodology II	Clustering-based Scenario Reduction

The four different days will be mentioned as the following:

Table 5.2: Day-types for Case Studies

Term	Description
Day-type I	Sunny Day
Day-type II	Rainy Day
Day-type III	Winter Day
Day-type IV	Overcast Day

The following figure shows the different hourly PV generation profiles for each type of day.

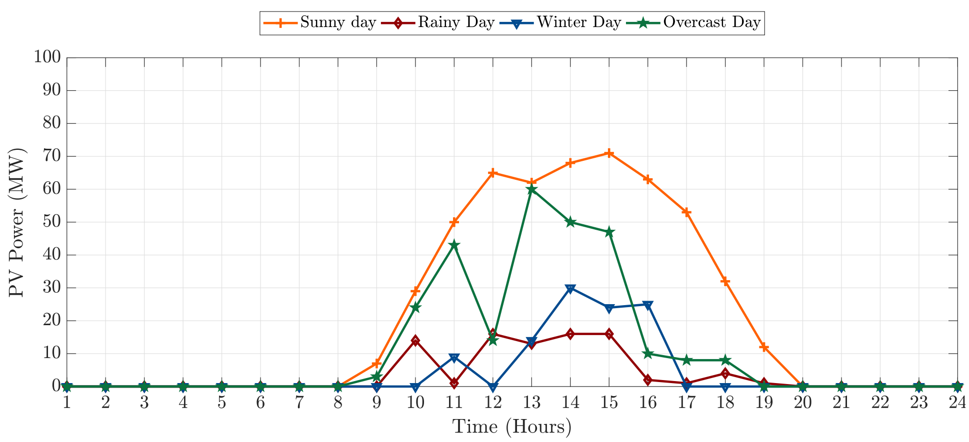

Fig. 5.3: Hourly PV Generation for each Day-Type

5.2 Scenario Generation and Reduction using Real-World Data

In this section, we present four case studies. Cases I and II refer to studies based on proposed scenario generation methodologies. Cases III and IV present case studies for proposed scenario reduction methodologies. We select PV generation profiles that represents the diurnal PV generation pattern as seen on the different day-types mentioned in table 5.2. A day-ahead forecast and actual PV power curve is shown for the purpose of comparison. Additionally, we pre-determine the number of scenarios generated. In all the cases we have decided to generate 100 scenarios around the day-ahead forecast. The PV plant considered for these studies has a generation capacity of $100MW$.

5.2.1 Case I: Scenario Generation Methodology I

In this subsection, we present the results of scenario generation using methodology I: Uncertainty and Variability Indices-Based Scenario Generation.

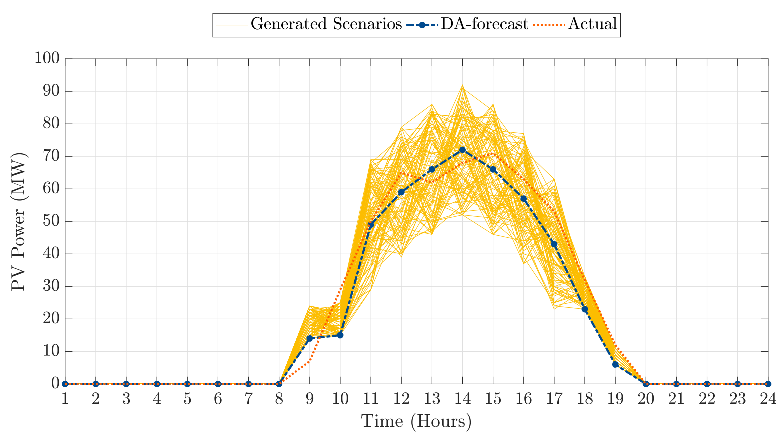

Fig. 5.4: Scenario generation methodology I as applied to day-type I (Sunny Day)

Comment: A sunny-day PV generation profile is typically a bell-shaped curve, with a peak at around noon. In this case, the PV generation initially has an almost uniform ramp, steadily increasing from around $8:00$, hits a peak at $14:00$, and finally gradually decreasing til it reaches zero PV generation at $20:00$. Based on the average uncertainty and variability index values for each hour, this methodology assumes higher level of uncertainty during peak hours for this particular case. It should be noted that these indices are based on historical data and hence do not indicate exactly what is contributing to the uncertainty and variability. It can be seen that the scenarios generated cover a wide range of possible PV generation for the following day. This is evident when the actual PV generation is compared with the scenarios. As seen, the scenarios cover the differences in day-ahead forecasts and the actual PV generation.

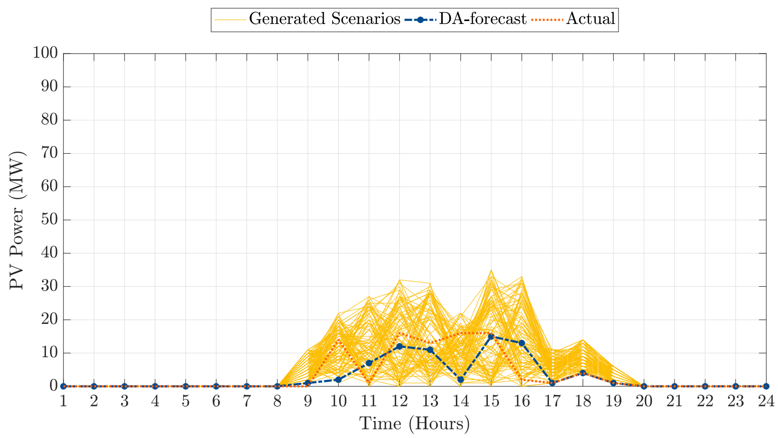

Fig. 5.5: Scenario generation methodology I as applied to day-type II (Rainy Day)

Comment: The case deals with scenario generation for a rainy-day. A rainy-day is particularly characterized by high levels on uncertainty and high variability due to uneven cloud cover and a generally low irradiance reception. The PV generation starts at $8:00$ and is seen to be highly variable due to possible cloud cover. This trend continues into noon time and finally the generation returns to zero at $20:00$. Based on the average uncertainty and variability index values for each hour, this methodology generates scenarios that are lower in generation - below 40% capacity - and vary between larger limits. As seen in the figure, the scenarios were later seen to be successful in covering the actual variations on the day of operation.

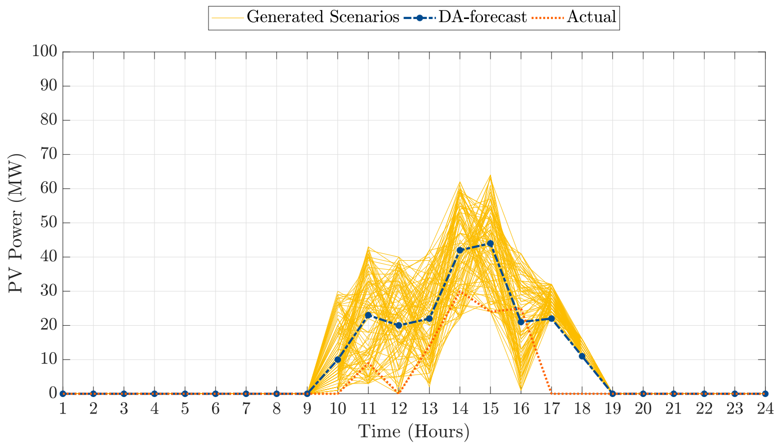

Fig. 5.6: Scenario generation methodology I as applied to day-type III (Winter Day)

Comment: A winter-day is marked by lower temperatures and a higher level of uncertainty. As observed in Figure 5.6, the actual generation begins at 10 : 00 and while it generally performed as forecasted, the generation is highly uncertain. The uncertainty and variability indices based on historical data observed this trend during similar days and generated wider limits for the scenario generation process. This led to large *bands* of scenarios earlier in the morning - between 9 : 00 to 13 : 00 as well. However, the scenario generation process was unable to capture the the generation possibilities in the evening.

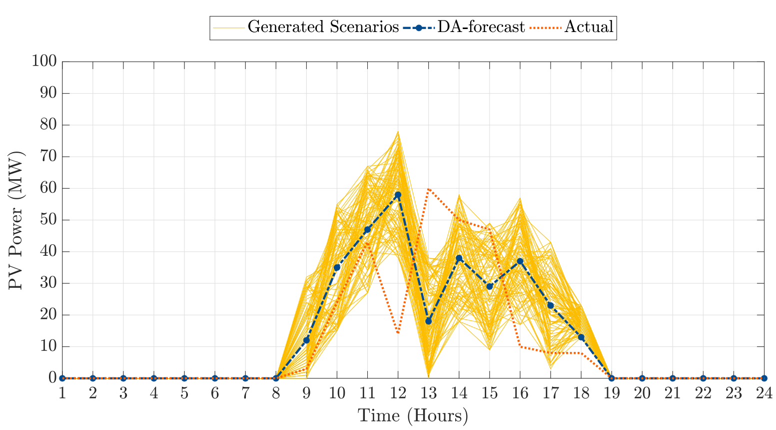

Fig. 5.7: Scenario generation methodology I as applied to day-type IV (Overcast Day)

Comment: In this case, we observe the scenario generation process for an overcast day. An overcast day is marked with generally moderate generation accompanied by high levels of variability and uncertainty, and as expected, presents the most difficult case for accurate scenarios. As observed from figure 5.7, the scenario generator was less efficient in capturing the uncertainty and variability. Between the time 12 : 00 and 14 : 00, due to the fact that the scenario generator generates scenarios around the day-ahead forecast, and that the forecast was inaccurate, the scenarios generated where unable to predict the reverse ramp witnessed on the day of operation. Apart from this, the scenario generator did moderately well in capturing the uncertainty and variability.

5.2.2 Case II: Scenario Generation Methodology II

In this subsection, we present the results of scenario generation using methodology II: Clustering-Based Scenario Generation.

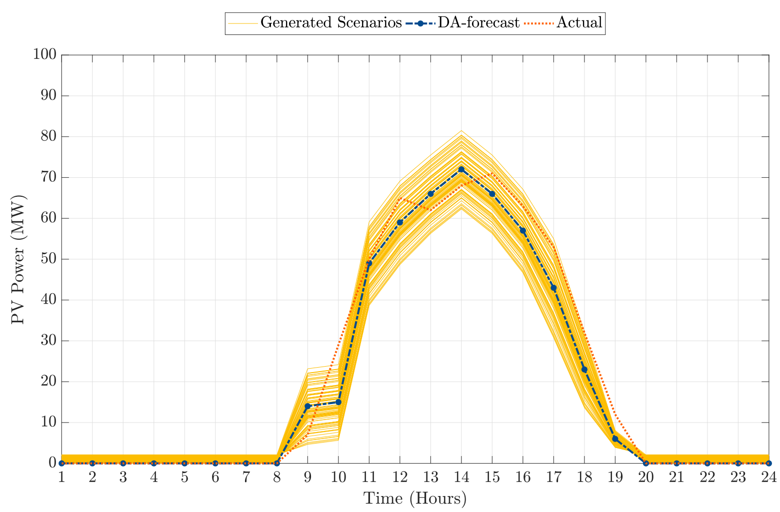

Fig. 5.8: Scenario generation methodology II as applied to day-type I (Sunny Day)

Comment: The scenario data-points are generated using a normal distribution based on the statistical properties of the cluster they belong. Hence the generated scenarios are uniformly distributed around the central day-ahead forecast. It can be observed that the scenarios have well covered the actual PV generation on the day of operation.

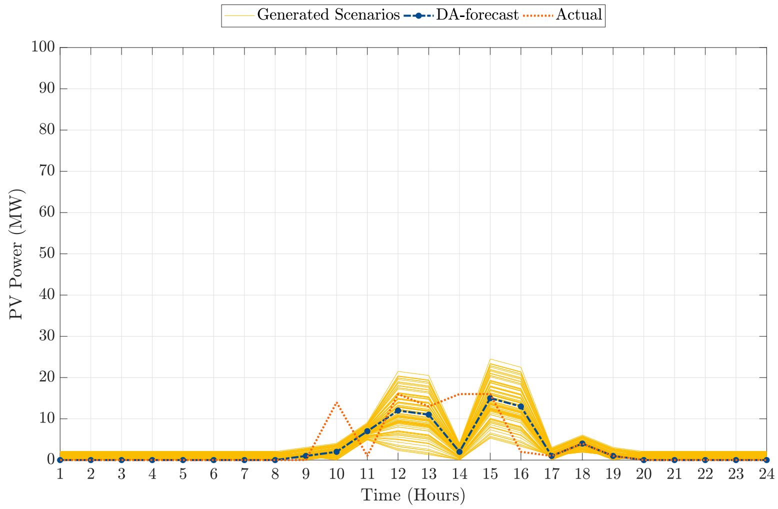

Fig. 5.9: Scenario generation methodology II as applied to day-type II (Rainy Day)

Comment: Due to the highly uncertain and variable natures of rainy-days, the scenario generation process finds it difficult to capture all possibilities as seen in this case. Here, while the uncertainty was captured for most of the hours, the variability, specifically between $9:00$ and $11:00$ was not captured.

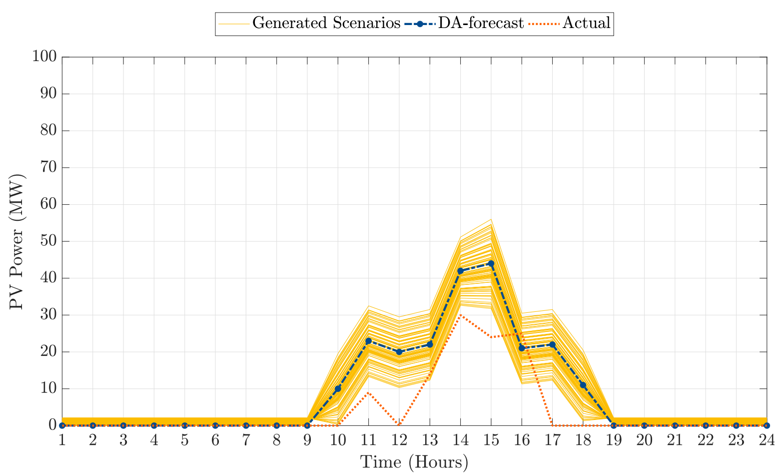

Fig. 5.10: Scenario generation methodology II as applied to day-type III (Winter Day)

Comment: As depicted is the previous case study, the winter day relatively predictable however, highly variable. Here, we observe that the generated scenarios were able to reduce the difference between the forecast and actual PV generation, with the exception of few hours of operation, specifically, between 16 : 00 and 18 : 00.

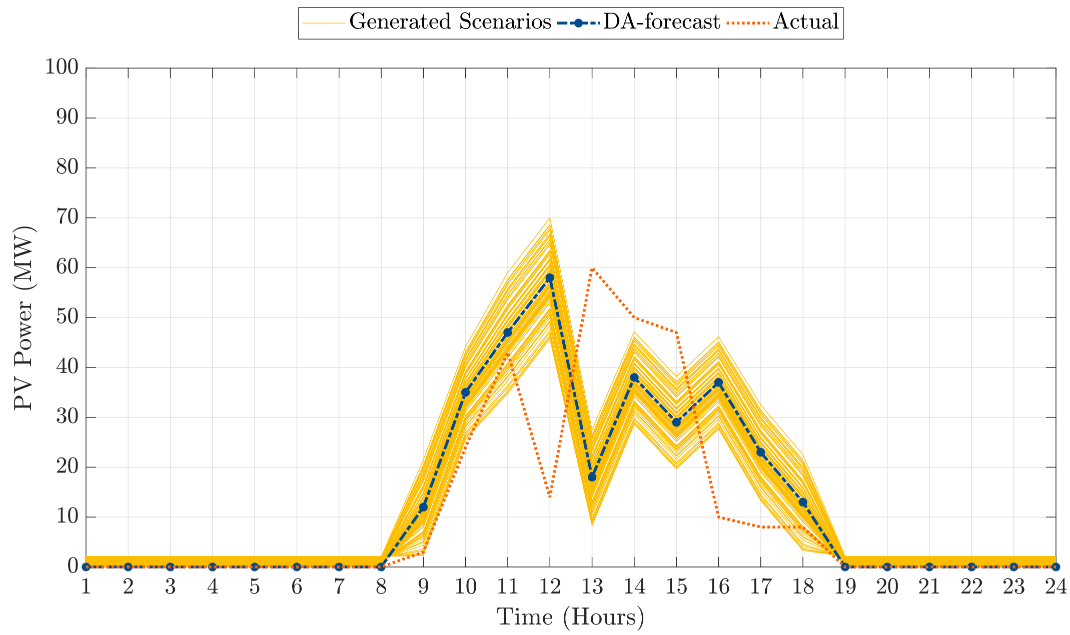

Fig. 5.11: Scenario generation methodology II as applied to day-type IV (Overcast Day)

Comment: This methodology was able to generate scenarios that reduced the gap between the day-ahead forecast and actual PV generation, however as observed with overcast days in general, the generation was very unpredictable. It is observed that between 13 : 00 and 14 : 00 there is a considerable mismatch in the generation. This may be attributed to a delay in an anticipated cloud cover, causing a mismatch between the forecast and actual generation by almost one hour.

5.2.3 Case III: Scenario Reduction Methodology I

In this subsection, we present the results of scenario reduction using methodology I: Probability Distance-Based Scenario Reduction.

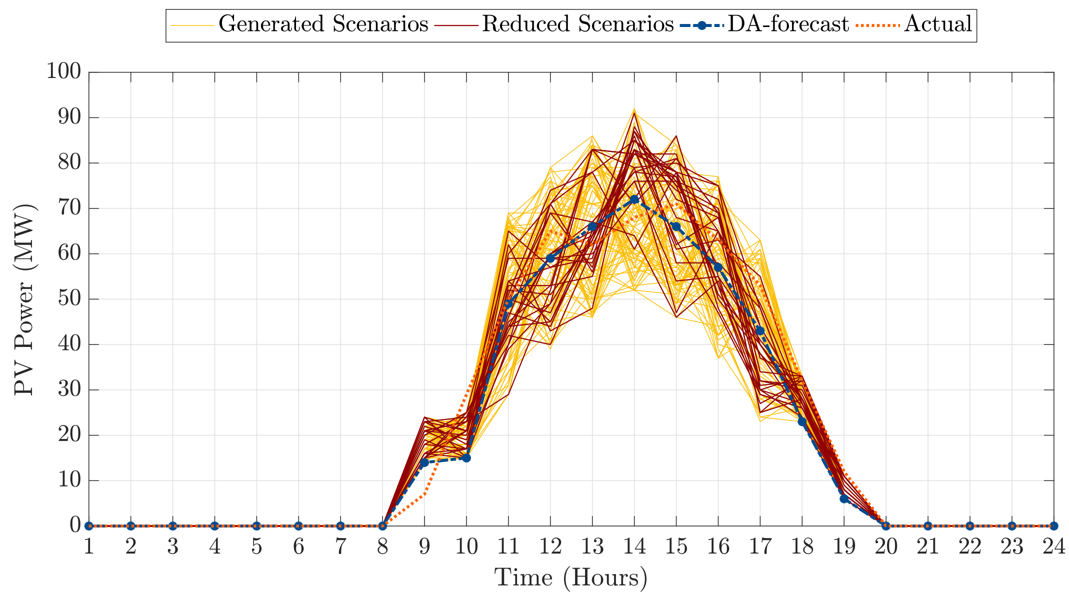

Fig. 5.12: Scenario reduction methodology I as applied to day-type I (Sunny Day)

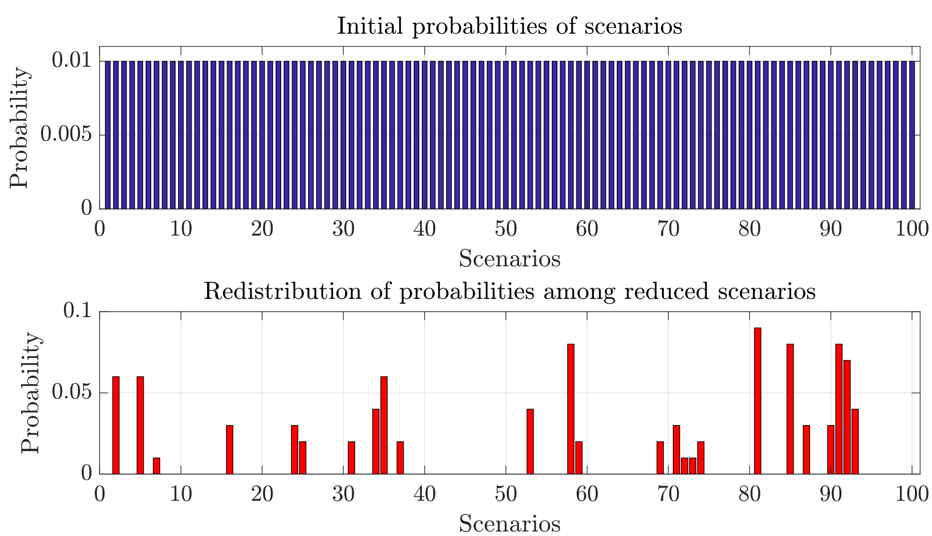

Fig. 5.13: Probability Re-distribution

Comment: In this case, scenario reduction at 75% reduction is applied to the initial scenario set generated in Case Study I. This results in the preservation of 25 scenarios, as displayed in figure 5.13. Initially, all the scenarios were equally probable. However, on reduction, redistribution of probabilities among the reduced set was performed. In this case, for example, scenario 81 was assigned the highest probability of 0.09.

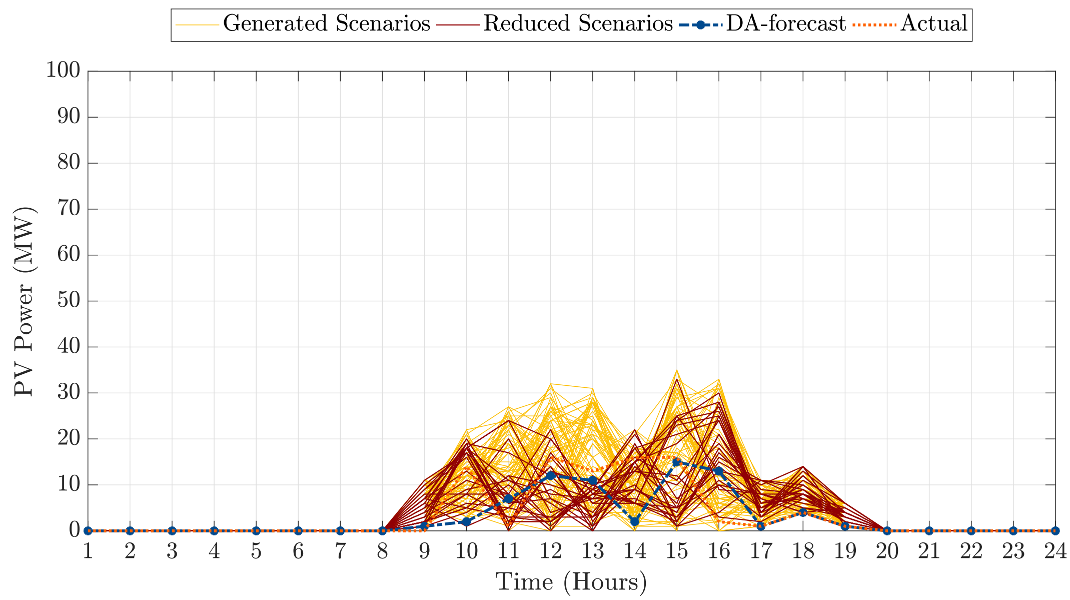

Fig. 5.14: Scenario reduction methodology I as applied to day-type II (Rainy Day)

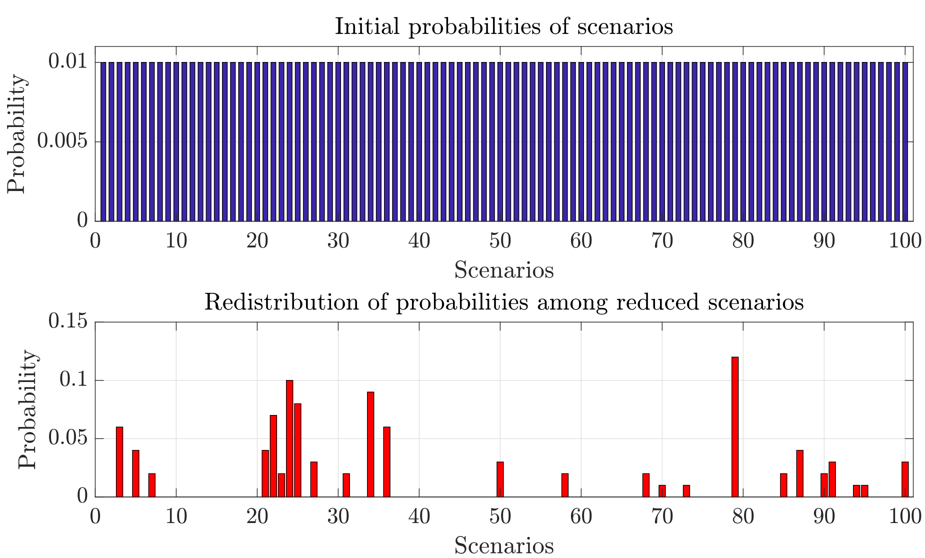

Fig. 5.15: Probability Re-distribution

Comment: As mentioned before, the PV generation is highly volatile in nature on rainy days in general. Therefore, it is important to preserve scenarios with similar uncertainty and variability indices. Figure 5.14 presents the scenario reduction for this case. It can be observed that, highly variable scenarios were preserved to reflect the non-predictability on such type of days.

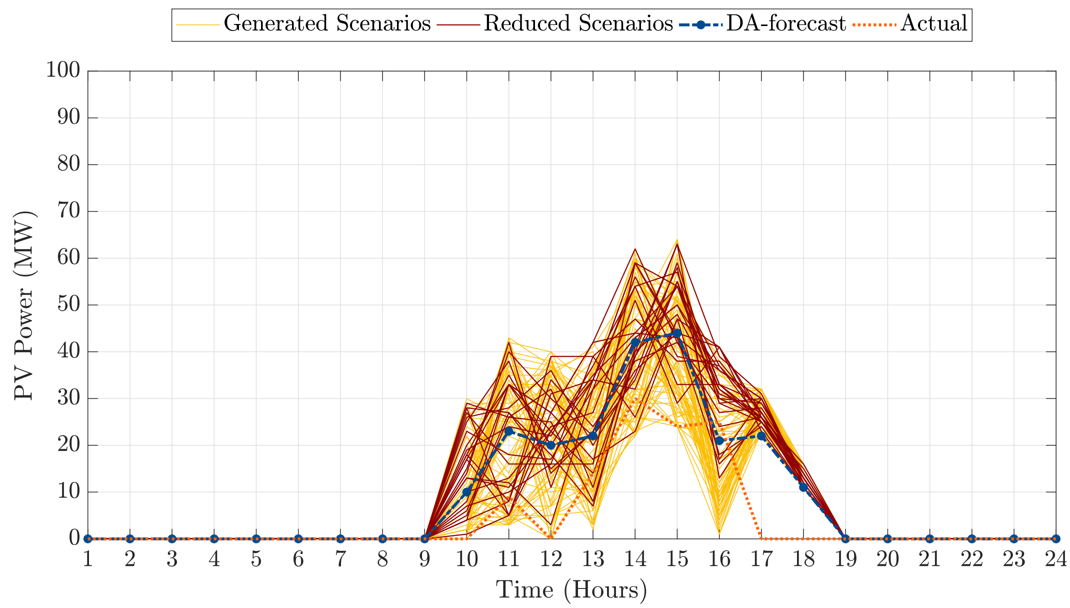

Fig. 5.16: Scenario reduction methodology I as applied to day-type III (Winter Day)

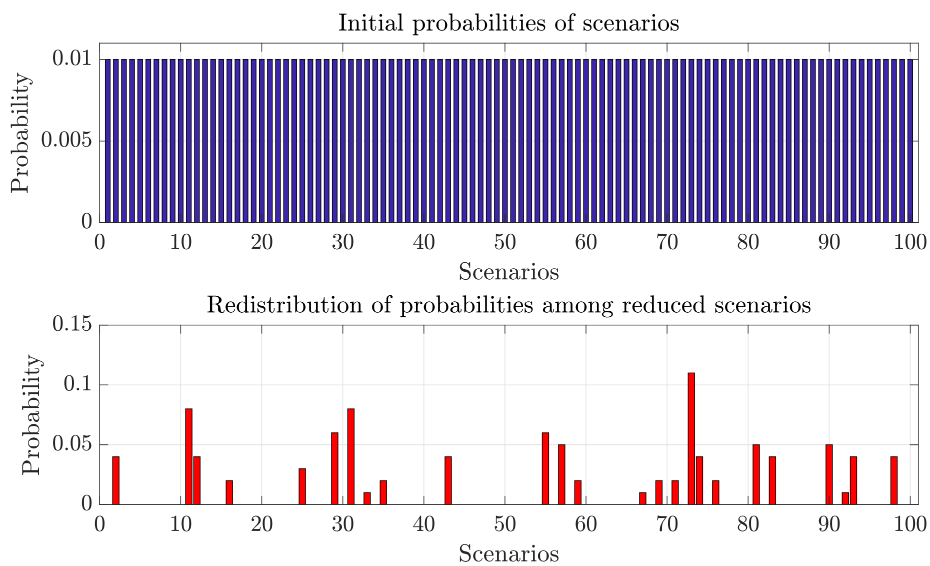

Fig. 5.17: Probability Re-distribution

Comment: It can be observed that outlier scenarios that are extreme but important were preserved in this case. This helps to capture the uncertainty on winter-type of days. For example, between $9:00$ and $1:00$, it was observed that such days have a higher uncertainty level. Hence, the band of scenarios was greater to accommodate higher temporal variations.

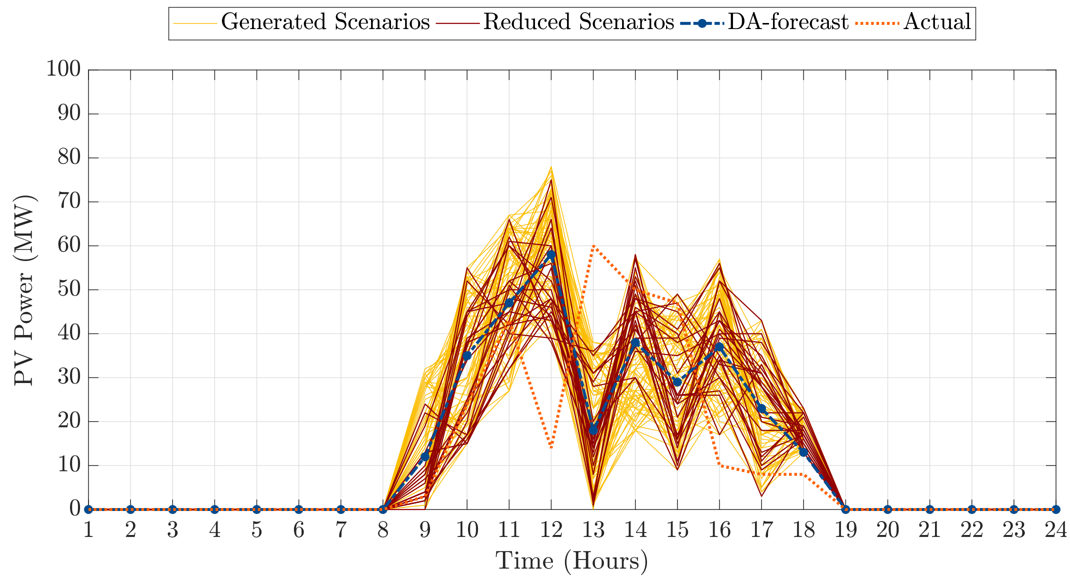

Fig. 5.18: Scenario reduction methodology I as applied to day-type IV (Overcast Day)

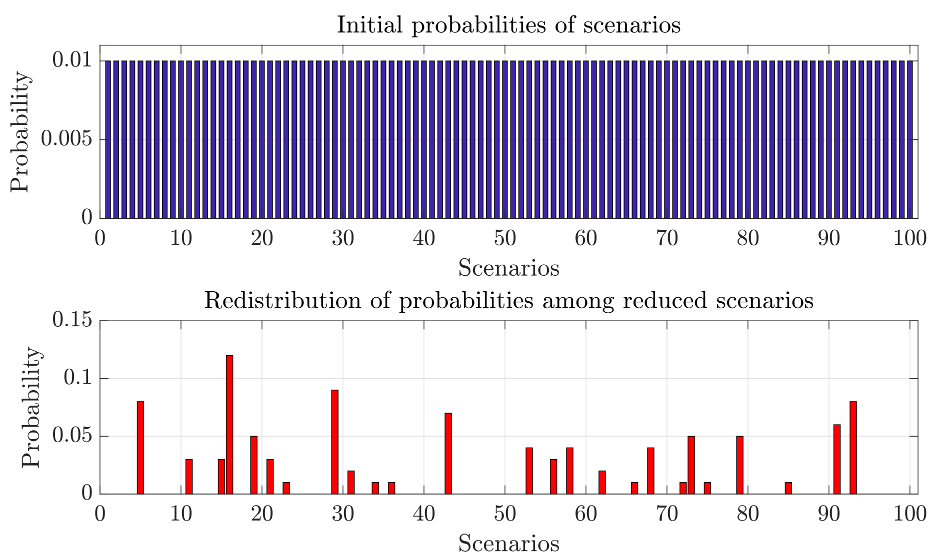

Fig. 5.19: Probability Re-distribution

Comment: As with the rest of the cases, the overcast type of day presents uncertainty and variability in scenarios generated. By measuring the probability distances between scenario pairs, the resulting reduced set contains only scenarios that are probabilistically speaking closer to each other and the day-ahead forecast.

5.2.4 Case IV: Scenario Reduction Methodology II

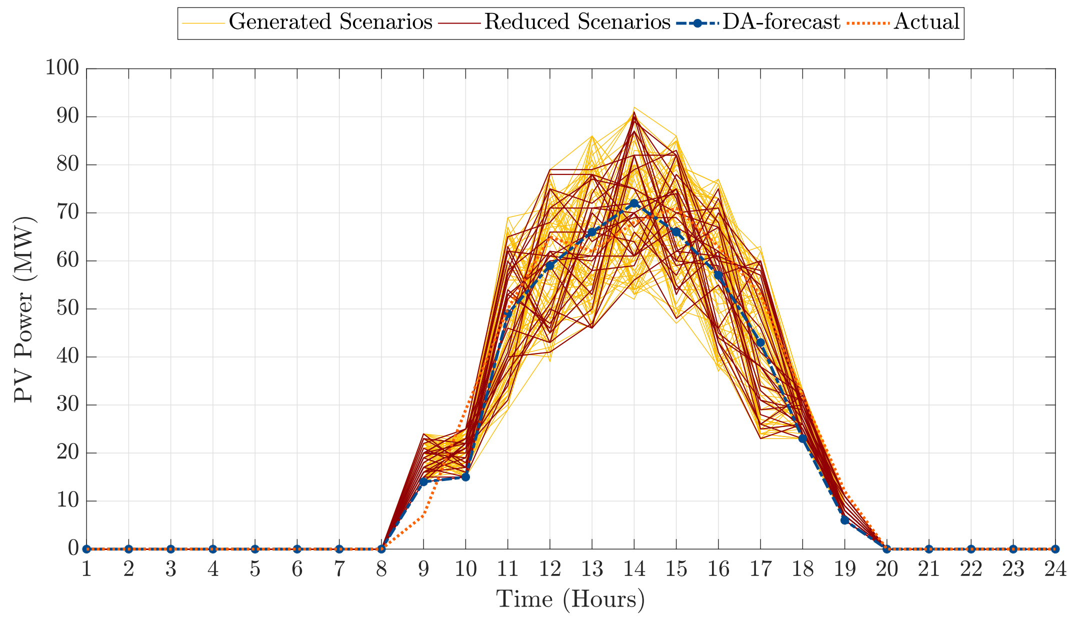

Fig. 5.20: Scenario reduction methodology II as applied to day-type I (Sunny Day)

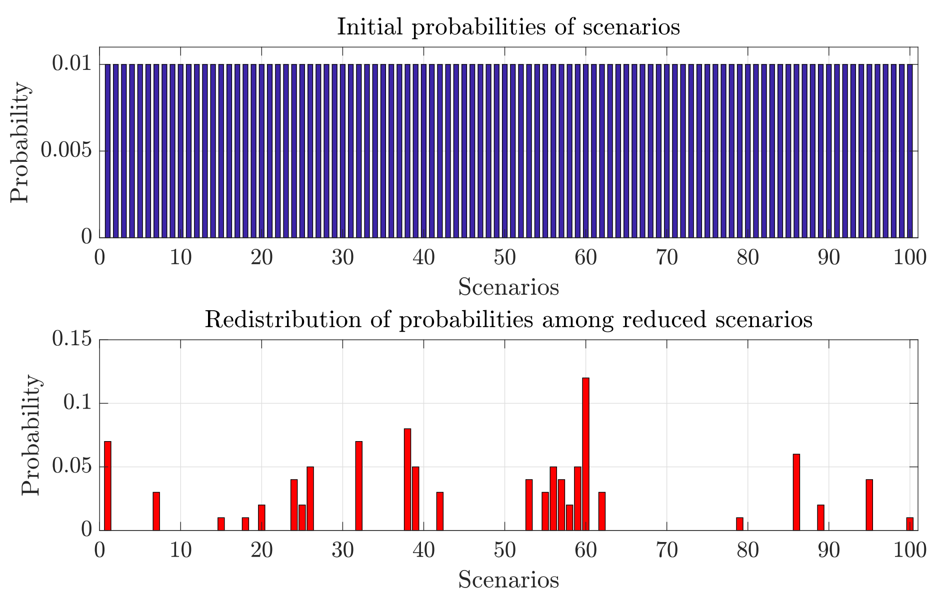

Fig. 5.21: Probability Re-distribution

Comment: In this case, the scenarios generated initially are grouped into clusters based on the square euclidean distance between individual scenarios and the day-ahead forecast. Since the variability is captured in the scenario generation process, results of the scenario reduction are highly dependent on the former. Probability redistribution is implemented at the end of the scenario reduction process.

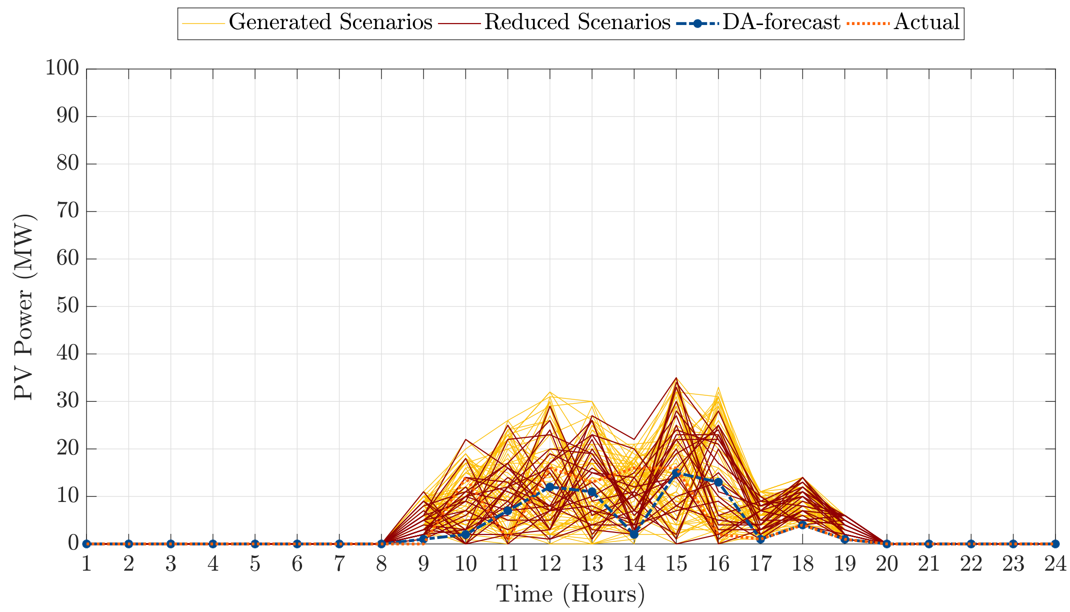

Fig. 5.22: Scenario reduction methodology II as applied to day-type II (Rainy Day)

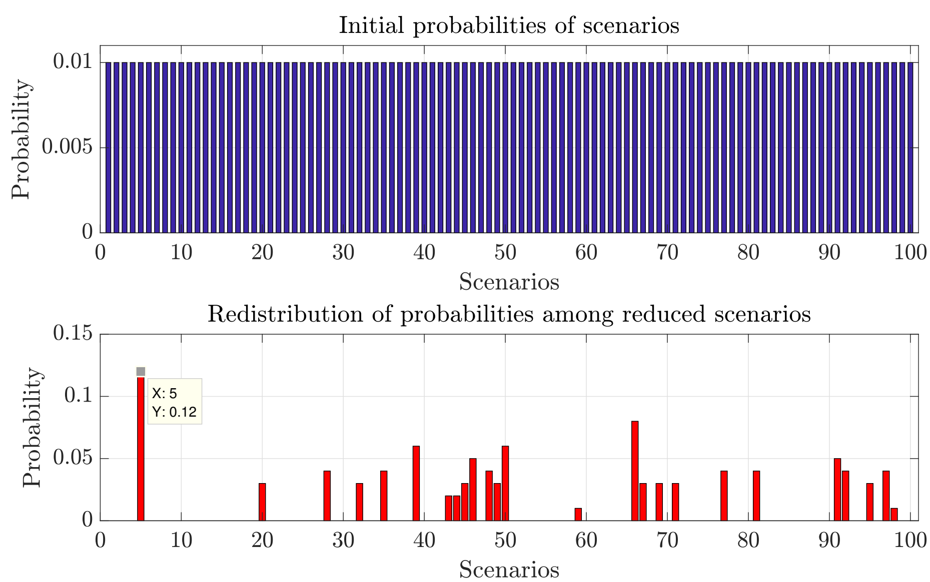

Fig. 5.23: Probability Re-distribution

Comment: Similar to the other methodology applied, the scenarios are reduced for better computational tractability. For this case, the uncertainty and variability is observed to be very high, hence the scenarios possessing similar characteristics were preserved.

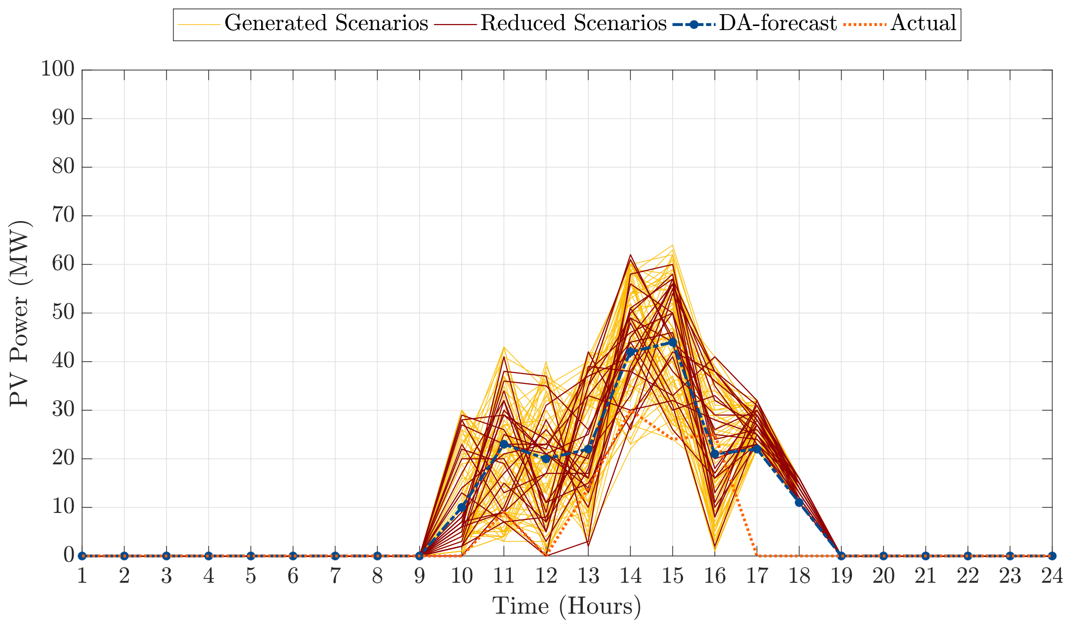

Fig. 5.24: Scenario reduction methodology II as applied to day-type III (Winter Day)

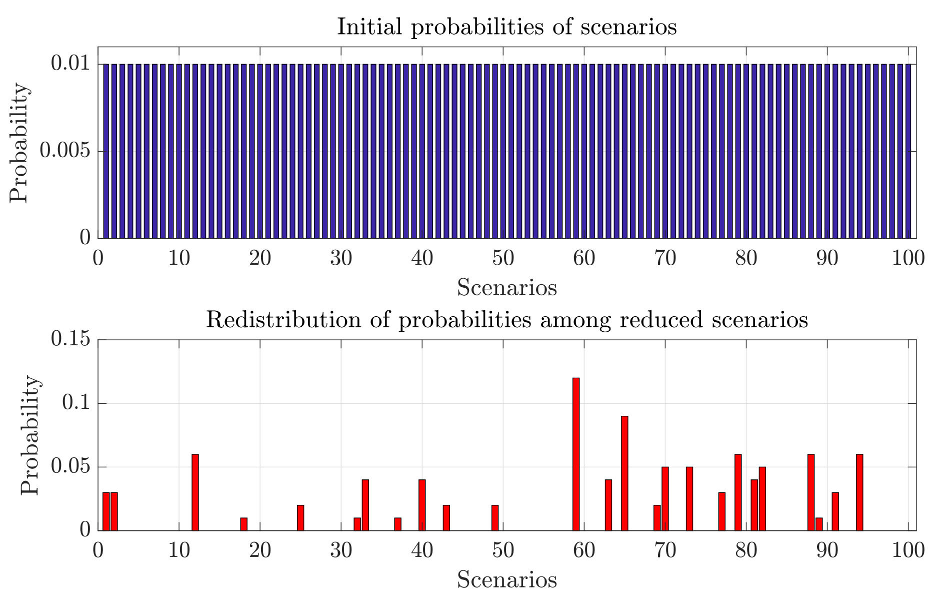

Fig. 5.25: Probability Re-distribution

Comment: In this case, the variability in the earlier hours - from 9 : 00 to 13 : 00

was high. Hence such scenarios were preserved in the reduced set.

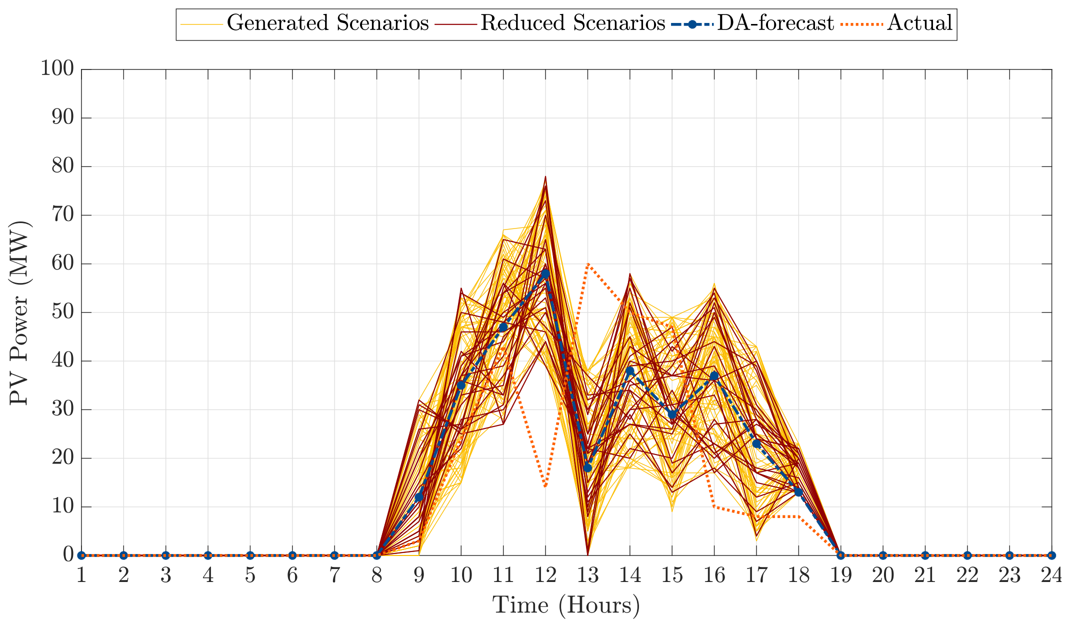

Fig. 5.26: Scenario reduction methodology II as applied to day-type IV (Overcast Day)

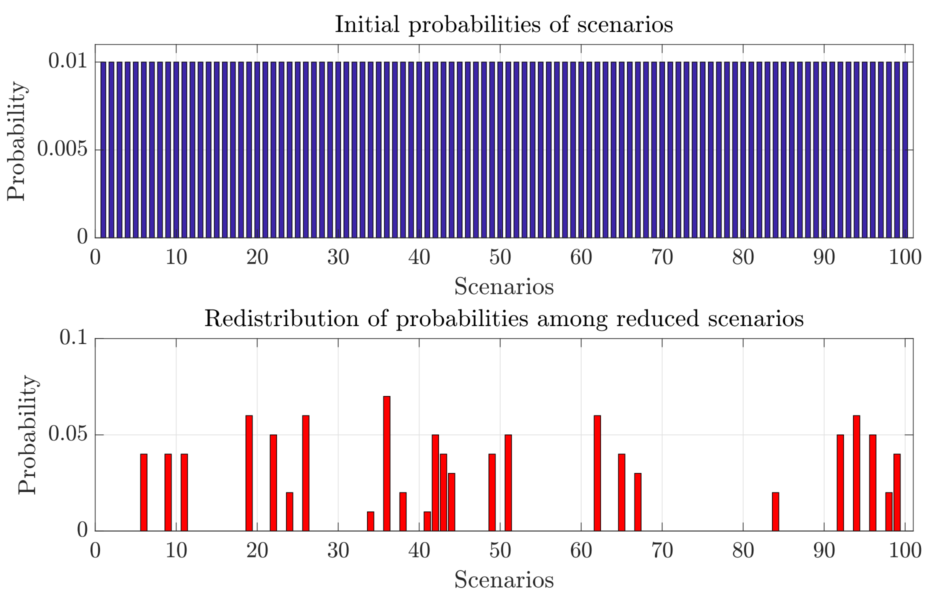

Fig. 5.27: Probability Re-distribution

Comment: In this case study, we observe that the reduced scenario set consists of a differently ramped scenarios, and not just those similar to the day-ahead forecast.

The clustering-based methodology provides an advantage in that we set the number of clusters required equal to the cardinality of the reduced matrix. This ensures that every scenario in the reduced set is representative of a different cluster of scenarios, thus capturing the statistical characteristics of the entire original set of scenarios.

5.3 Impact on Unit Commitment Solutions

In this section we will evaluate the impact on costs when scenarios are considered in the unit commitment problem formulation. We begin with a brief introduction to the mathematical formulation of deterministic UC formulation, followed by a description of the test system under study. Next, we evaluate a base case - without PV generation - followed by the use of day-ahead forecast to form the benchmark case for evaluation of unit commitment solutions. Ultimately we demonstrate the use of scenarios in this formulation and thereby provide a method for selection of the scenario that is most likely to occur according to our proposed methodologies.

5.3.1 Deterministic UC Formulation

The objective function of unit commitment is to achieve the minimum total operational cost over a planned time horizon. A deterministic UC objective function is composed of two component costs, related to two-stage decisions. The first component cost is a influenced by day-ahead schedules, typically dictated by the status of the generating units. The following mathematical formulation is adapted from [7]. The first-stage decisions include start-up decision v_gt and the shutdown decision w_gt. These indicate when the generators will be tuned on or shut down and the resulting schedule is not changed during the next-day operation hours. The second component cost includes the total operational costs in the second stage, while is made up of fuel cost and possible energy penalty 2.3.1.

Objective function

The objective function is typically given as the following:

$$\min \sum_{g \in G} \sum_{t \in T} (SU_g v_{gt} + SD_g wgt) + \sum_{g \in G} \sum_{t \in T} F_g(p_{gt}) + VOLL \sum_{i \in N} \sum_{t \in T} \delta_{it} \tag{5.1}$$

where

SU_g start-up cost of unit g

SD_g shut-down cost of unit g

$F_g(.)$ fuel cost function for unit g

$p_g t$ thermal power generation/dispatch amount of unit g at time t

$VOLL$ value of loss load [$\$/MWh$]

$\delta_i t$ load loss at bus i at time t

In general, fuel cost function are represented a s quadratic function of the dispatch/production level p and take the form, $F_g(p) = a + bp + cp^2$ where a,b and c are positive cost coefficients.

Constraints

Any generation unit g in the unit commitment schedule will have the basic constraint on their minimum ON time and minimum OFF time, thus specifying the startup action and shutdown action on each unit at a time period t respectively.

minimum ON time constraint

$$u_{gt} - u_{g(t-1)} \leq u_{g\tau} \qquad \forall g \in G, t \in T, \tau = t, ..., \min\{t + L_g - 1, |T| \tag{5.2}$$

minimum OFF time constraint

$$u_{g(t-1)} - u_{gt} \leq 1 - u_{g\tau} \qquad \forall g \in G, t \in T, \tau = t, ..., \min\{t + l_g - 1, |T| \tag{5.3}$$

where

$u_g t$ commitment decision a generator commits online if $u_{gt} = 1$, otherwise $u_{gt} = 0$.

L_g minimum ON duration

l_g minimum OFF duration

τ time alias (possible operating time period starting from time t)

$|T|$ duration of planning horizon

The startup and shutdown actions are determined by the generator commitment statuses in the previous time period $t-1$ and the current time period t.

Start-up action constraint

$$v_{gt} \geq u_{gt} - u_{g(t-1)} \qquad \forall g \in G, t \in T \tag{5.4}$$

Shutdown action constraint

$$W_{gt} \geq -u_{gt} + u_{g(t-1)} \qquad \forall g \in G, t \in T \tag{5.5}$$

$$u_{gt}, v_{gt}, w_{gt} \in [0,1] \qquad \forall g \in G, t \in T \tag{5.6}$$

5.4 Case Study: Application of Scenarios in Deterministic Unit Commitment

In this case study, we will evaluate the cost-benefit of using scenarios in a deterministic UC formulation. As for the scenarios used in this case study, we consider the results generated by application of proposed method 2 - based on k-means clustering - in Chapter 5. To match the generating capacity of the overall system, the scenarios are scaled up by a factor of 100.

A Procedure to determine the best scenario for UC schedule

Below we begin with the case study, we present a procedural technique to determine which of the generated scenarios would most likely be the best in terms of objective cost minimization.

1. First, we begin by running the UC problem for the day-ahead forecast as well as for all the scenarios generated.

2. Next, we determine the objective costs of each of the cases. The objective cost for net-load using day-ahead forecast is selected as a benchmark value.

3. We then select a subset of the scenarios which result in an objective cost less than that resulting from step 3.

4. Once we obtain the subset, we look at the rank of these scenarios. This ranking is observed from the results of scenario reduction process and is based on their probability of occurrence.

5. The scenario with the highest probability of occurrence - and therefore the highest rank - is selected as the most optimal scenarios for minimizing objective costs.

5.4.1 Test System Description

For this study, we consider a 200-generator test system. The generator units parameters are described in the Appendix **??**. Also, the unit commitment formulation was implemented in the optimization software, GAMS, and solved using a state-of-the-art solver called *BARON*[76].

Base load profile for 200-generator bus system:

The 24-hour load profile for the overall system is shown below in figure 5.28.

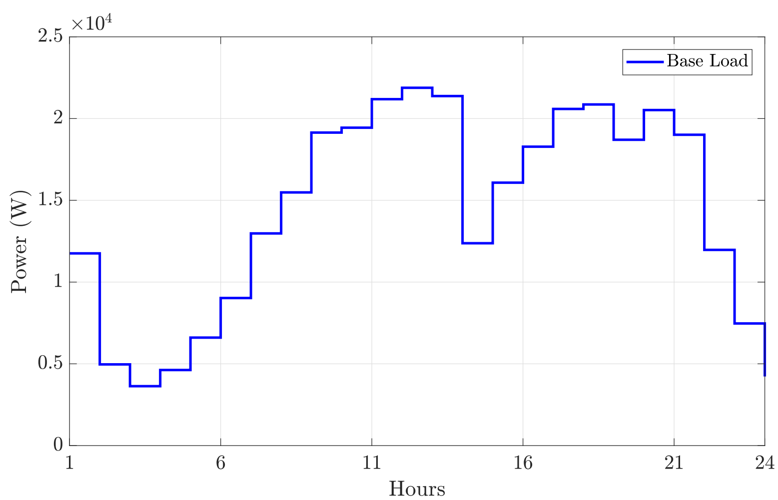

Fig. 5.28: 24-hour base-load profile for 200-generator system

This load profile represents the overall load variations for the entire test day.

Scenarios for UC formulation

In our test case, we apply method 1 of scenario generation and method 2 of scenario reduction to produce a set of 25 scenarios for day type I (sunny day). The results are shown in figure 5.29.

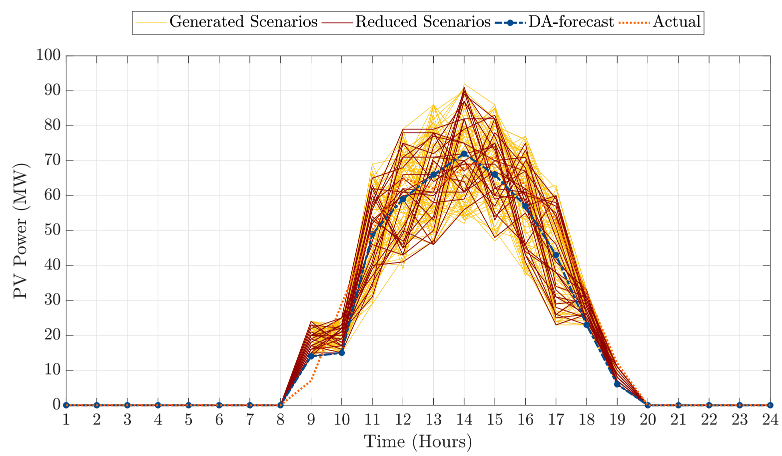

Fig. 5.29: Scenario Reduction for Test Day

Based on their probability of occurrence, the scenarios are ranked and shown in figure 5.30. Also, shown in figure is the percent error of the scenarios with respect to the actual PV generation on the test day.

5.4.2 Net-Load Profile for Representation of PV Generation in UC

If P_{load} is the base load profile, and P_{PV} is the PV generation contributing to the load demand, then the net load is given by:

$$P_{NET} = P_{load} - P_{PV} \tag{5.7}$$

Net load profile considering day-ahead forecast:

By applying equation 5.7 to the day-ahead forecast - shown in figure 5.31 for PV generation, we get a net-load profile. The resulting net-load is shown in figure 5.31.

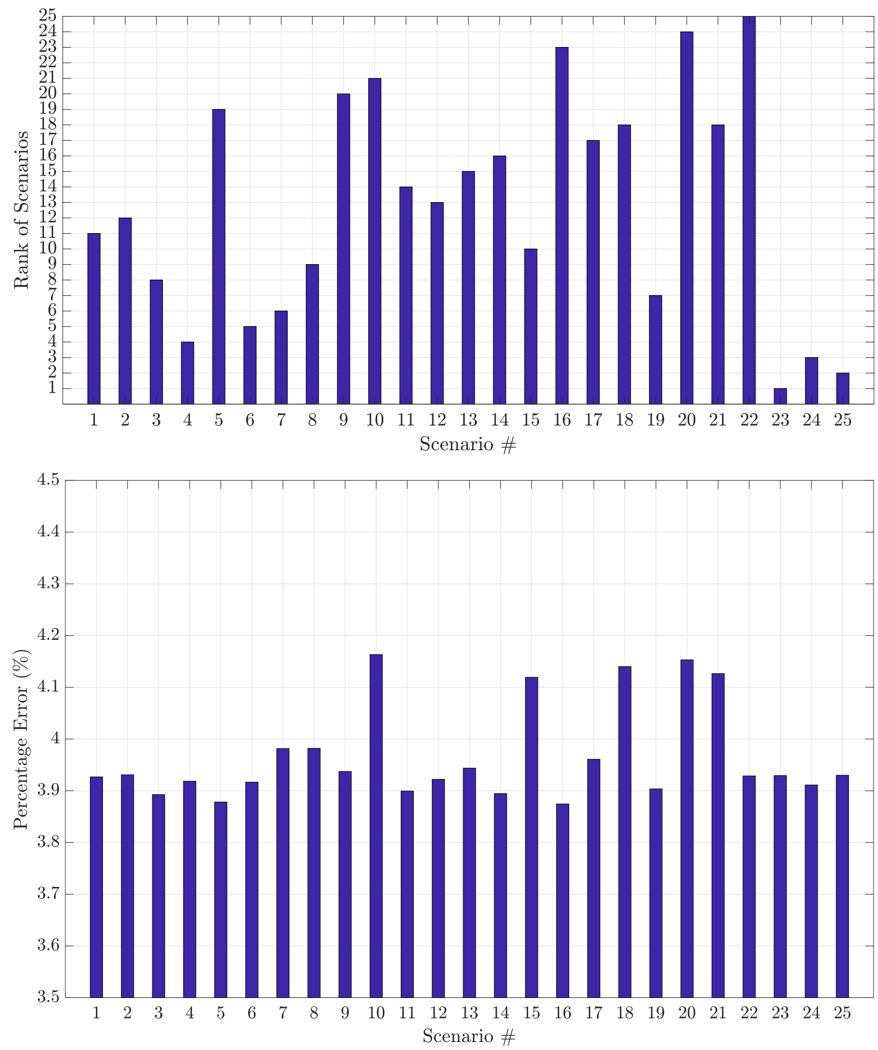

Fig. 5.30: Scenario Ranking and Percentage Error of Scenarios

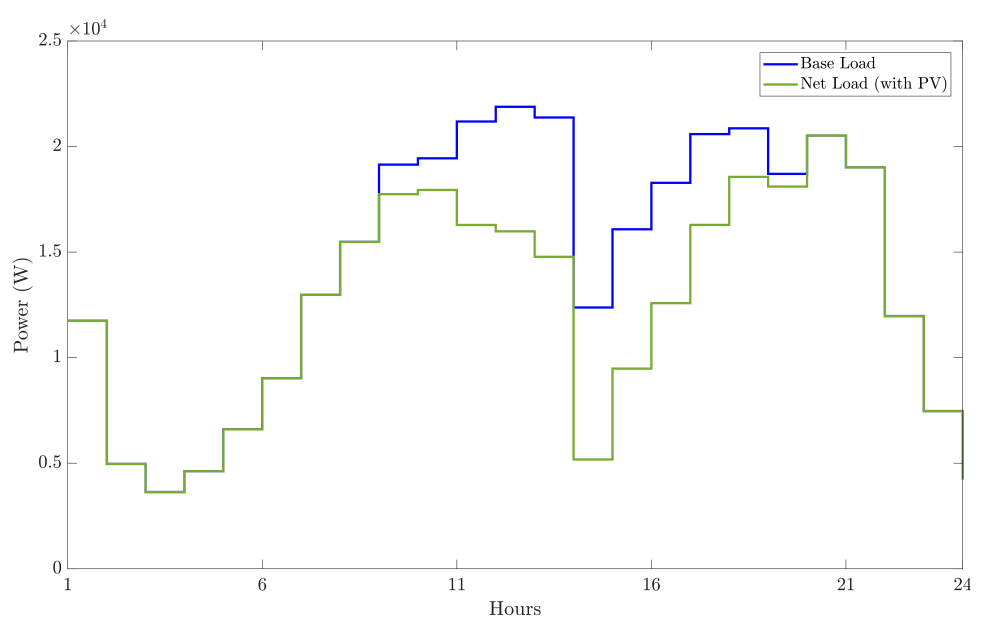

Fig. 5.31: 24-hour base-load profile with day-ahead predicted PV generation

It is observed that the overall load is reduced from time 7 : 00 to 20 : 00 due to PV generation present during this time.

Additionally, we apply the same equation to the 25 scenarios generated and calculate the net-load for each scenario. The resulting net-loads are shown in figure 5.32 along with a visual representation of PV generation as a percentage of the total load demand.

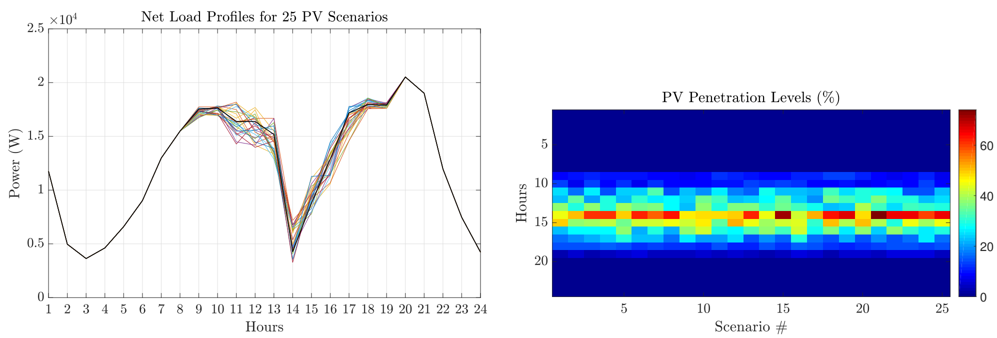

Fig. 5.32: Net-load Profiles for Scenarios and PV Penetration Levels

5.4.3 Deterministic Unit Commitment Solutions

Next, we run the deterministic UC problem for the following cases and find the total objective costs:

- Base-load case (without PV)
- Net-load with Actual PV generation
- Net-load with Day-ahead PV generation forecast
- Net-loads with 25 Scenarios for PV Generation

The hourly generation for the base-load case is given in figure 5.33.

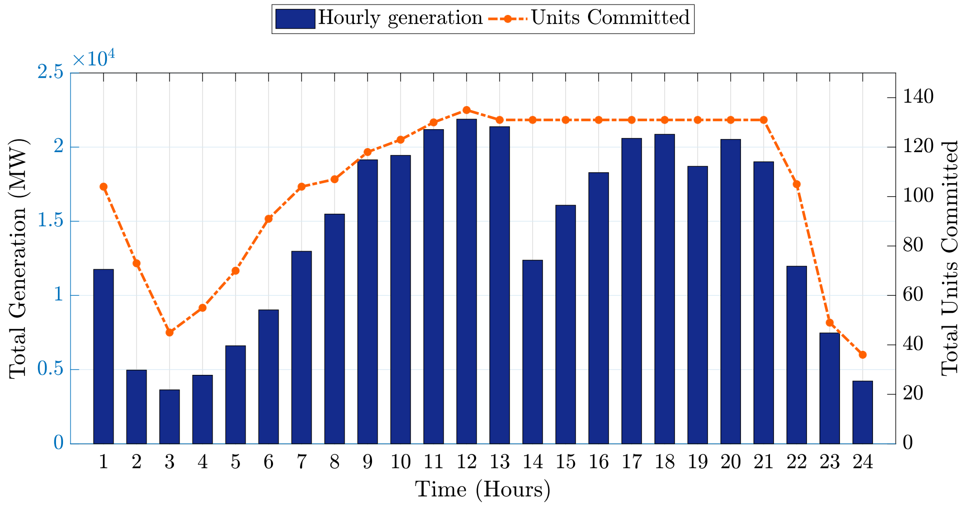

Fig. 5.33: Hourly Generation and Units Committed for Base-load Case

Table 5.3 summarizes the total objective costs generated for all the aforementioned cases:

Analysis:

There are a couple of observations that can be made from this table.

- The total objective costs generated using 25 scenarios ranges from 5.0627 to 7.3985 million dollars.

- The objective cost when day-ahead forecast was used is found to be 5.9573 million dollars. Let us call this value OBJ_{DAF}

- From a power system operator stand-point, scenarios with a total objective cost lesser than this value. This is because, in terms of minimizing costs, the scenarios are utilized to find a PV profile that results in a lesser cost compared to what is predicted by the day-ahead forecast.

- There are only *12* scenarios resulting in a objective less than OBJ_{DAF}. These 12 scenarios are highlighted in figure 5.35.

Table 5.3: Total Objective Costs for UC Solutions

Case	Objective Cost (million $)
Base Case	5.846368672
Actual PV	5.682048108
DA-Forecast	5.957341024
Scenario 1	5.124536757
Scenario 2	5.266452792
Scenario 3	6.359869023
Scenario 4	6.194680509
Scenario 5	5.11576512
Scenario 6	6.376362058
Scenario 7	5.588765234
Scenario 8	6.393572432
Scenario 9	5.211168849
Scenario 10	5.150977541
Scenario 11	5.22492091
Scenario 12	5.248129265
Scenario 13	6.15694809
Scenario 14	5.111185572
Scenario 15	6.966832853
Scenario 16	5.062652865
Scenario 17	5.44720502
Scenario 18	6.92948379
Scenario 19	7.133574159
Scenario 20	5.155918588
Scenario 21	7.398528485
Scenario 22	6.971609672
Scenario 23	6.594616808
Scenario 24	6.00691421
Scenario 25	6.588567798

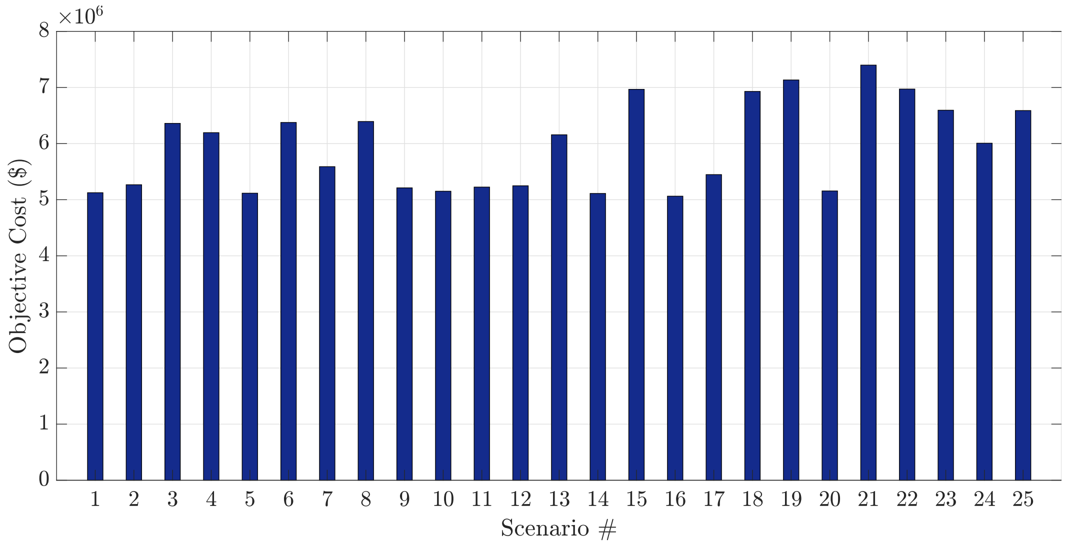

Fig. 5.34: Objective costs for 25 scenarios

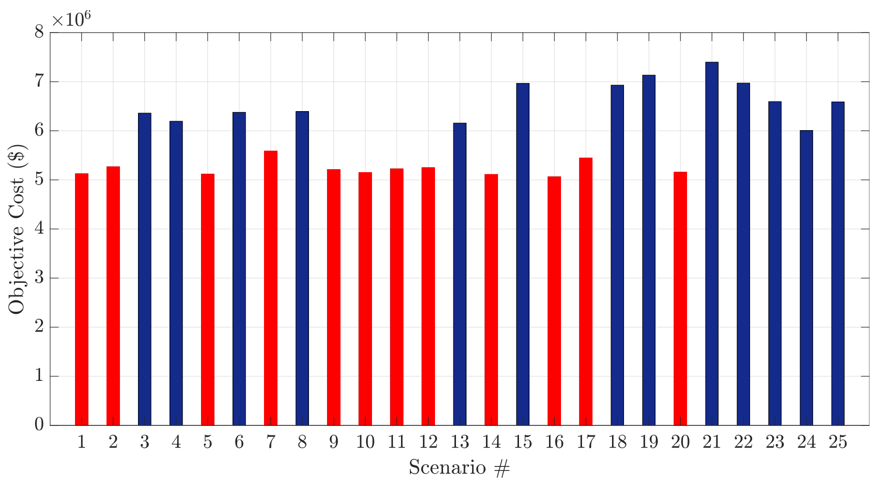

Fig. 5.35: Scenarios with objective cost less than OBJ_{DAF}

- Now, based on the ranking of scenarios shown in figure 5.30, we observe that among these 12 scenarios, the 7^{th} scenario has the highest ranking.

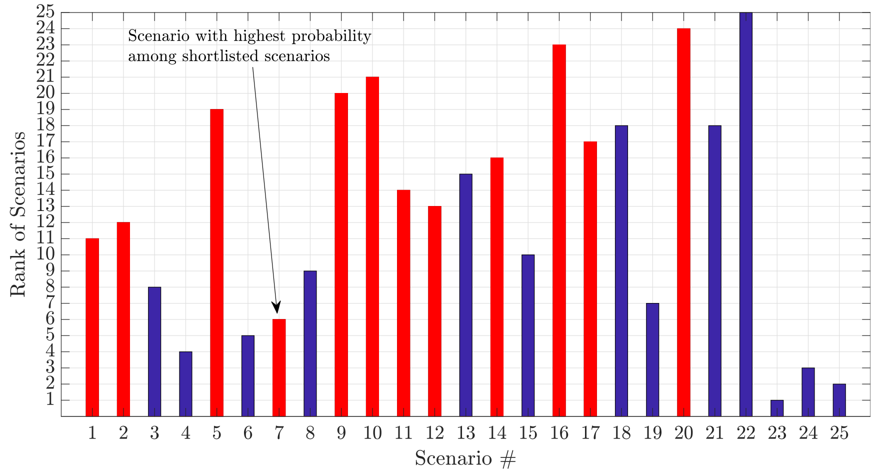

Fig. 5.36: Selection of most-likely scenario

- The objective cost of the resulting scenario is 5.5888 million dollars which is less than OBJ_{DAF} and is most-probable, hence representing the most optimal scenario from the given set of scenarios.

The net-load for the selected scenario is given in figure 5.37.

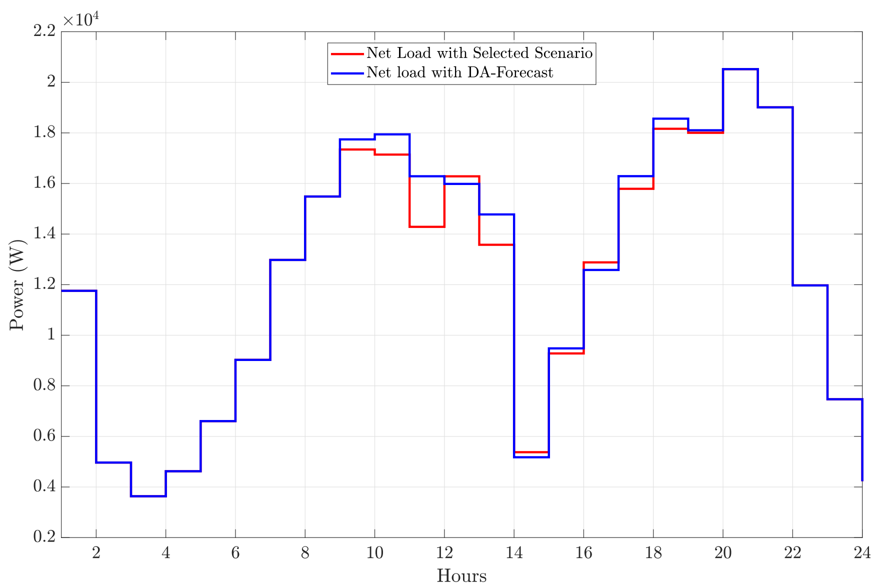

Fig. 5.37: Net-load profile for 7^{th} Scenario

The hourly generation and total units committed are shown in the figure below.

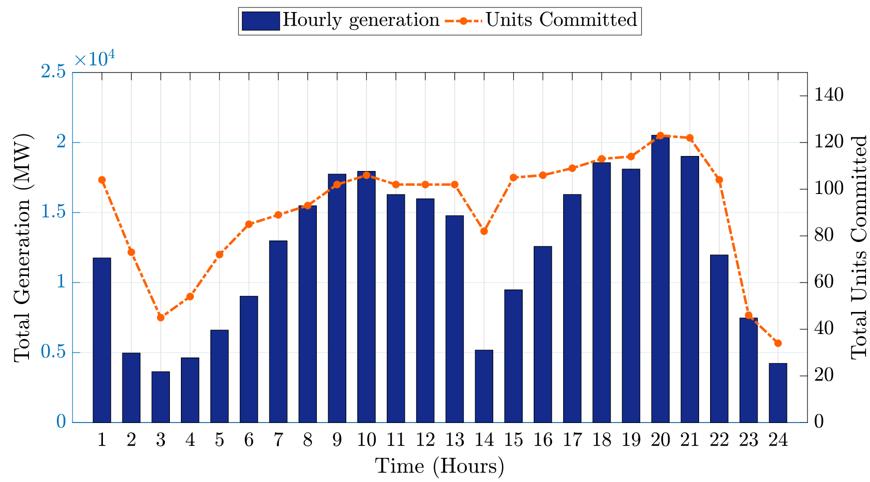

Fig. 5.38: Hourly Generation and Units Committed for 7^{th} Scenario

5.5 Summary

In this chapter, real-world data was used to demonstrate the proposed scenario generation and reduction methodologies. First, a demonstration of scenario generation for different sizes of scenario sets were presented. Second, using the scenarios generated in the first section, we demonstrate scenario reduction at difference reduction percentages. In the next section, several case studies were formulated to investigate the performance of scenario generation and reduction processes for different types of days in a year.It was observed that each of the proposed methodologies have a distinct advantage in terms of either capturing the uncertainty and variability of PV generation, or strategically reducing the scenario set for computational gains. Finally, the resulting reduced scenario set were applied to a deterministic unit commitment model to assess the cost benefits in generation scheduling with scenarios.

CHAPTER 6: CONCLUSIONS

6.1 Summary and Research Contribution

This thesis proposes a framework for scenario generation and reduction involving renewable energy sources, specifically, photovoltaic generation for the purpose of integrating with the generation commitment formulations in the energy markets.
This thesis proposes two sets of methodologies:

- Two scenario generation methodologies based on uncertainty and variability indices, and clustering technique respectively.

- Two scenario reduction methodologies based on probability distances, and k-means clustering technique respectively.

These methodologies together form a framework for transforming statistical knowledge of PV generation to realistic scenarios that can aid in, one way, improving the PV dispatchability on the electric grid. Through a set of illustrative examples and real-world data from NREL, the scalability of these methodologies were showcased. Additionally, we observed the performance of each method for different diurnal PV generation patterns across four different type of days. These studies were performed in MATLAB environment and algorithms are implemented in the MATLAB script language.

The general contributions of this thesis are as follows:

- Framework to characterize uncertainty and variability of PV generation in terms of set of indices by analyzing historical measured and forecasted data. Additionally, a method to model the forecast errors.

- Scenario generation methodology based on the day-ahead forecast and the uncertainty and variability indices.

- Scenario generation methodology based on clustering historical data into predefined clusters and sampling technique based on comparison between day-ahead forecast points and clusters information.

- Adapted framework to reduce scenario-set according to probabilistic distances between generated scenarios.

- Scenario reduction technique based on the application of K-means clustering technique to initial scenario set.

While there is adequate literature on scenario-based analysis in problem-solving, to the best of my knowledge, there is little work done towards applying these techniques to PV-integrated generation commitment problems. There are several ways these works can be extended in future.

6.2 Future Work

This thesis proposed a novel way to apply statistical techniques and expert knowledge of power systems to PV-integrated generation commitment.In addition to this application, these ideas can be applied to other power systems operations, such as in wind power systems, power system planning, and microgrid applications. It is suggested that the future works are focused on the following:

- Extend the idea of scenario generation and reduction to intra-day market-decisions, e.g., scenarios can be generated for day-ahead unit commitment decisions, and with a receding time horizon approach, the scenarios can be improved as we approach the operating time. This approach has potential to deliver better results as the forecast tends to improve as the lead time decreases.

- Cloud cover plays a crucial role in determining the uncertainty and variability of the actual PV generation. It is suggested to develop a framework toward modeling cloud cover into the uncertainty and variability indices proposed in chapter 3.

- The proposed scenario reduction methodologies in chapter 4 assume a predefined cardinality and thus applies the same reduction percentage, irrespective of the statistical information of the scenario set. An extension of this work is to optimally determine the reduction percentage for reducing a scenario set. An obvious caveat, however, is the increased computational burden with respect to the scenario reduction framework as a whole.

- Application of scenarios to stochastic unit commitment (SUC) formulations can be conducted to assess the impact of scenarios in a probabilistic manner. It is suggested to adopt SUC and compare the cost-benefits with the deterministic results.

- Extended the study to measure the impact of scenarios on a deterministic UC formulation.

REFERENCES

[1] R. Margolis, C. Coggeshall, and J. Zuboy, "Sunshot vision study," *US Department of Energy*, 2012.

[2] B. Espinar, J.-L. Aznarte, R. Girard, A. M. Moussa, and G. Kariniotakis, "Photovoltaic forecasting: A state of the art," *Proceedings 5th European PV-Hybrid and Mini-Grid Conference*, pp. 250–255, 2010.

[3] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, pp. 65–67, 2013.

[4] R. Perez, "Comparison of numerical weather prediction solar irradiance forecasts in the us," *Solar Energy*, pp. 305–326, 2013.

[5] N. Navid and G. Rosenwald, "Ramp capability product design for miso markets: Available online at https://www.misoenergy.org/Library/Repository," *US Department of Energy.*

[6] L. Xu and D. Tretheway, "Flexible ramping products," *California ISO*, 2014.

[7] Y. Huang, "Electrical power unit commitment," *Briefs in Energy, Springer*, 2017.

[8] S. Zenios, "Practical financial optimization: Decision making for financial engineers," *Blackwell Publications*, pp. 421–428, 2007.

[9] R. Kerr, A. J. Fontana, and J. Wiley, "Unit Commitment," *IEEE Transactions on Power Apparatus and Systems*, no. 5, pp. 417–421, 1966.

[10] K. Hara, M. Kimura, and N. Honda, "A Method for Planning Economic Unit Commitment and Maintenance of Thermal Power Systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-85, no. 5, pp. 427–436, 1966.

[11] R. Burns and C. A. Gibson, "Optimization of priority lists for a unit commitment program," *IEEE Power Eng. Soc. Summer Meeting*, 1975.

[12] F. Lee, "Short-term thermal unit commitment-a new method," *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 421–428, 1988.

[13] F. N. Lee and Q. Feng, "Multi-area unit commitment," *IEEE Trans. Power Syst.*, vol. 7, pp. 591–599, 1992.

[14] R. Nieva, A. Inda, and J. Frausto, "CHT: a digital computer package for solving short term hydro-thermal coordination and unit commitment problems," *IEEE Trans. Power Syst.*, vol. PAS-1, pp. 168–174, 1986.

[15] W. Snyder, H. D. Jr. Powell, and J. C. Jr. Rayburn, "Dynamic programming approach to unit commitment," *IEEE Trans. Power Syst.*, vol. 2, pp. 339–347, 1987.

[16] H. Happ, R. C. Johnson, and W. J. Wright, "Large scale hydro-thermal unit commitment-method and results," *IEEE Trans. Power Syst.*, vol. PAS-90, pp. 1373–1384, 1971.

[17] W. Hobbs, G. Hermon, and G. B. Warner, S.and Sheble, "An enhanced dynamic programming approach for unit commitment," *IEEE Trans. Power Rev.*, vol. PAS-90, p. 70, 1988.

[18] C. Li, R. Johnson, and A. Svoboda, "A new unit commitment method," *IEEE Trans. Power Syst.*, vol. 12, pp. 113–119, 1997.

[19] T. Dillon, K. Edwin, H. Kochs, and R. Taud, "Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination," *IEEE Trans. Power Syst.*, vol. PAS-97, pp. 2154–2166, 1978.

[20] C. Chang, G.W.and Su, "A practical mixed integer linear programming-based short-term hydro scheduling," *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*, vol. 3, pp. 1606–1610, 2002.

[21] W. Chang, J. M. J. Aganagic, M. Waight, T. Burton, S. Reeves, and M. Christofordis, "Experiences with mixed integer linear programming based approaches on short-term hydro scheduling," *IEEE Transactions on Power Syst*, vol. 16, pp. 743–749, 2001.

[22] G. S. L. T. K. Chang, W.and Chuang, "A simplified combined-cycle unit model for mixed integer linear programming-based unit commitment," *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pp. 1–6, 2008.

[23] L. S. Guedes, P. De Mendonca Maia, A. C. Lisboa, D. A. G. Vieira, and R. R. Saldanha, "A Unit Commitment Algorithm and a Compact MILP Model for Short-Term Hydro-Power Generation Scheduling," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3381–3390, 2017.

[24] X. Ke, N. Lu, D. Wu, and M. Kintner-Meyer, "A modified priority list-based MILP method for solving large-scale unit commitment problems," *IEEE Power and Energy Society General Meeting*, vol. 2015-September, 2015.

[25] G. B. Sheble and G. N. Fahd, "Unit commitment literature synopsis," *IEEE Transactions on Power Systems*, vol. 9, no. 1, pp. 128–135, 1994.

[26] S. Lauere, N. R. Sandell, D. P. Bertsekas, and T. A. Posbergh, "Solution of large-scale optimal unit commitment problems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 1, pp. 79–86, 1982.

[27] A. Cohen and M. Yoshimura, "A Branch-and-Bound Algorithm for Unit Commitment," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 2, pp. 444–451, 1983.

[28] A. I. Cohen and S. H. Wan, "A Method for Solving the Fuel Constrained Unit Commitment Problem," *IEEE Transactions on Power Systems*, vol. 2, no. 3, pp. 608–614, 1987.

[29] R. Nieva, A. Inda, and I. GuilléN, "Lagrangian Reduction of Search-Range for LargeâScale Unit Commitment," *IEEE Transactions on Power Systems*, vol. 2, no. 2, pp. 465–473, 1987.

[30] S. Tong and S. Shahidehpour, "Combination of Lagrangian-relaxation and linear-programming approaches for fuel-constrained unit-commitment problems," *IEE Proceedings C Generation, Transmission and Distribution*, vol. 136, p. 162, 1989.

[31] A. I. Cohen and S. H. Wan, "A Method for Solving the Fuel Constrained Unit Commitment Problem," *IEEE Transactions on Power Systems*, vol. 2, no. 3, pp. 608–614, 1987.

[32] S. Mokhtari, J. Sing, and B. Wollenberg, "A unit commitment expert system [power system control]," *Power Systems, IEEE Transactions on*, vol. 3, no. 1, pp. 272–277, 1988.

[33] M. Mazumdar and A. Kapoor, "Stochastic models for power generation system production costs," *Electric Power Systems Research*, vol. 35, no. 2, pp. 93–100, 1995.

[34] N. P. Padhy, V. R. Ramachandran, and S. R. Paranjothi, "Fuzzy decision system for unit commitment risk analysis," *Int. J. Power Energy Syst.*, vol. 19, no. 2, pp. 180–185, 1999.

[35] S. K. Tong, "Hydro-thermal unit commitment with probabilistic constraints using segmentation method," *IEEE Trans. Power Syst.*, vol. 5, no. 1, pp. 276–282, 1990.

[36] N. P. Padhy, "A new fuzzy expert decision making approach for unit commitment with reliable risk reserve and emission constraints," *J. Energy Environment*, vol. 1, pp. 25–36, 1999.

[37] H. Sasaki, M. Watanabe, J. Kubokawa, N. Yorino, and R. Yokoyama, "A Solution Method of Unit Commitment by Artificial Neural Networks," *IEEE Transactions on Power Systems*, vol. 7, no. 3, pp. 974–981, 1992.

[38] C. Wang and S. M. Shahidehpour, "Effects of ramp-rate limits on unit commitment and economic dispatch," *IEEE Transactions on Power Systems*, vol. 8, no. 3, pp. 1341–1350, 1993.

[39] Y.-Y. Su, C.-C., Hsu, "Fuzzy Dynamic Programming: An Application to Unit Commitment," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1231–1237, 1991.

[40] M. Hindsberger and H. Ravn, "Multiresolution modeling of hydro-thermal systems," *PICA 2001. Innovative Computing for Power - Electric Energy Meets the Market. 22nd IEEE Power Engineering Society. International Conference on Power Industry Computer Applications (Cat. No.01CH37195)*, 2001.

[41] N. P. Padhy, "Unit commitment using hybrid models: a comparative study for dynamic programming expert system fuzzy system and genetic algorithms," *J. Energy Environment*, vol. 23, pp. 827–836, 2000.

[42] W. Lei, M. Shahidehpour, and L. Tao, "Stochastic security-constrained unit commitment," *J. Energy Environment*, vol. 22, pp. 800–811, 2007.

[43] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 306–316, 2008.

[44] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 800–810, 2012.

[45] C. Zhao, J. Wang, J. P. Watson, and Y. Guan, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2708–2717, 2013.

[46] Q. Wang, Y. Guan, and J. Wang, "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 206–215, 2012.

[47] J. Morales, R. MÃÂnguez, and A. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843,855, 2010.

[48] R. Karki, P. Hu, and R. Billinton, "A simplified wind power generation model for reliability evaluation," *Energy Conversion IEEE Transactions on Power Energy and Industry Applications*, vol. 21, no. 2, pp. 533,540, 2006.

[49] A. Sturt and G. Strbac, "Time series modelling of power output for large-scale wind fleets," *Wind Energy*, vol. 14, no. 8, pp. 953–966, 2011.

[50] T. Peiyuan Chen, B. Pedersen, B. Bak-Jensen, and B. Zhe Chen, "Arima-based time series model of stochastic wind power generation," *Power Systems IEEE Transactions on Power Energy and Industry Applications*, vol. 25, no. 2, pp. 667,676, 201005.

[51] H. Konno and H. Yamazaki, "Mean-absolute deviation portfolio optimization model and its application to tokyo stock market," *Management Science*, vol. 37, pp. 519–531, 1991.

[52] M. Young, “A minimax portfolio selection rule with linear programming solution,” *Management Science*, vol. 44, pp. 673–683, 1998.

[53] G. Albeanu, M. Ghica, F. Popeniu-VlË, and A. , “On using bootstrap scenario-generation for multi-period stochastic programming applications,” vol. 3, pp. 156–161, 01 2008.

[54] O. F. Demirel and T. R. Willemain, “Generation of simulation input scenarios using bootstrap methods,” *The Journal of the Operational Research Society*, vol. 53, no. 1, pp. 69–78, 2002.

[55] R. R. P. Kouwenberg, “Scenario generation and stochastic programming models for asset liability management,” *European Journal of Operational Research*, vol. 134(2), p. 51â64, 2001.

[56] D. R. Carino, T. Kent, D. H. Myers, C. Stacy, M. Sylvanus, A. L. Turner, K. Watanabe, and W. T., “The russell-yasuda kasai model: an asset liability model for a japanese insurance company using multistage stochastic programming,” *INTERFACES*, vol. 24(1), p. 29â49, 1994.

[57] H. Chen, R. Li, Y. Wang, and C. Xu, “Wind power forecasting based on refined lstar-garch model,” *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2590–2593, 2017.

[58] P. M. Lurie and M. S. Goldberg, “An approximate method for sampling correlated random variables from partially-specified distributions,” *Management Science*, vol. 44(2), p. 203â218, 1998.

[59] J. Dupacova, G. Consigli, and S. Wallace, “Scenarios for multistage stochastic programs,” *Annals of Operations Research*, vol. 100, no. 1, pp. 25,53, 2000.

[60] I. Corp., “Ibm optimization library stochastic extensions users guide,” 1998.

[61] N. Growe-Kuska, H. Heitsch, and R. W., “Scenario reduction and scenario tree construction for power management problems,” *Power Tech Conference Proceedings, 2003 IEEE Bologna*, vol. 3, p. 7 pp. Vol.3, 2003.

[62] Y. Dvorkin, Y. Wang, H. Pandzic, and D. Kirschen, “Comparison of scenario reduction techniques for the stochastic unit commitment,” pp. 1–5, 2014.

[63] J. M. Morales, S. Pineda, A. J. Conejo, and M. Carrion, “Scenario reduction for futures market trading in electricity markets,” *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 878–888, 2009.

[64] Y. Feng and S. M. Ryan, “Scenario reduction for stochastic unit commitment with wind penetration,” *2014 IEEE PES General Meeting | Conference Exposition*, pp. 1–5, 2014.

[65] K. Bruninx and E. Delarue, “Scenario reduction techniques and solution stability for stochastic unit commitment problems,” *2016 IEEE International Energy Conference (ENERGYCON)*, pp. 1–7, 2016.

[66] Y. Wang, “Energy storage operation under uncertainty,” 2017.

[67] J. Dupacova, N. Grajwe-Kuska, and W. Rajmisch, “Scenario reduction in stochastic programming,” *Mathematical Programming*, vol. 95, no. 3, pp. 493–511, 2003.

[68] G. Nicole, “Gams/scenred tutorial: Technical report,”

[69] F. Yonghan and R. Sarah, “Solution sensitivity-based scenario reduction for stochastic unit commitment,” *Computational Management Science*, vol. 13, pp. 29–62, 2016.

[70] J. Gomez-Martinez, M. Ortega-Vazquez, and L. F. Ochoa, “Scenario selection of wind forecast errors for stochastic unit commitment: A uk case study,” *IEEE PES Innovative Smart Grid Technologies, Europe*, pp. 1–6, 2014.

[71] J. Sumaili, H. Keko, V. Miranda, Z. Zhou, A. Botterud, and J. Wang, “Finding representative wind power scenarios and their probabilities for stochastic models,” *Intelligent System Application to Power Systems (ISAP), 2011 16th International Conference*, pp. 1–6, 2011.

[72] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.

[73] T. Sapatinas, “The elements of statistical learning,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 167, 2004.

[74] D. Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding,” *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, 2007.

[75] M. Sarfati, “Modeling the Diversification Benefit of Transmission Investments Modeling the Diversification Benefit of Transmission Investments in the Presence of Uncorrelated Generation Sources,” 2012.

[76] “Baron solver,” *http://archimedes.cheme.cmu.edu/?q=baron.*

APPENDIX A: Ranking of Reduced Scenarios

A.1 Scenario Reduction Methodology I: Case I to IV

Table A.1: Methodology I: Ranking and Probability distribution of Reduced Scenarios

Rank	Case I		Case II		Case III		Case IV	
	Scenario #	Probability	Scenario #	Probability	Scenario #	Probability	Scenario #	Probability
1	81	0.09	79	0.12	73	0.11	16	0.12
2	58	0.08	24	0.1	11	0.08	29	0.09
3	85	0.08	34	0.09	31	0.08	5	0.08
4	91	0.08	25	0.08	29	0.06	93	0.08
5	92	0.07	22	0.07	55	0.06	43	0.07
6	2	0.06	3	0.06	57	0.05	91	0.06
7	5	0.06	36	0.06	81	0.05	19	0.05
8	35	0.06	5	0.04	90	0.05	73	0.05
9	34	0.04	21	0.04	2	0.04	79	0.05
10	53	0.04	87	0.04	12	0.04	53	0.04
11	93	0.04	27	0.03	43	0.04	58	0.04
12	16	0.03	50	0.03	74	0.04	68	0.04
13	24	0.03	91	0.03	83	0.04	11	0.03
14	71	0.03	100	0.03	93	0.04	15	0.03
15	87	0.03	7	0.02	98	0.04	21	0.03
16	90	0.03	23	0.02	25	0.03	56	0.03
17	25	0.02	31	0.02	16	0.02	31	0.02
18	31	0.02	58	0.02	35	0.02	62	0.02
19	37	0.02	68	0.02	59	0.02	23	0.01
20	59	0.02	85	0.02	69	0.02	34	0.01
21	69	0.02	90	0.02	71	0.02	36	0.01
22	74	0.02	70	0.01	76	0.02	66	0.01
23	7	0.01	73	0.01	33	0.01	72	0.01
24	72	0.01	94	0.01	67	0.01	75	0.01
25	73	0.01	95	0.01	92	0.01	85	0.01

A.2 Scenario Reduction Methodology II: Case I to IV

Table A.2: Methodology II: Ranking and Probability distribution of Reduced Scenarios

Rank	Case I		Case II		Case III		Case IV	
	Scenario #	Probability	Scenario #	Probability	Scenario #	Probability	Scenario #	Probability
1	60	0.12	5	0.12	59	0.12	36	0.07
2	38	0.08	66	0.08	65	0.09	62	0.06
3	32	0.07	50	0.06	12	0.06	19	0.06
4	1	0.07	39	0.06	79	0.06	94	0.06
5	86	0.06	46	0.05	88	0.06	26	0.06
6	56	0.05	91	0.05	94	0.06	22	0.05
7	59	0.05	97	0.04	70	0.05	96	0.05
8	39	0.05	35	0.04	82	0.05	51	0.05
9	26	0.05	28	0.04	73	0.05	42	0.05
10	95	0.04	92	0.04	40	0.04	92	0.05
11	57	0.04	48	0.04	33	0.04	65	0.04
12	24	0.04	77	0.04	81	0.04	99	0.04
13	53	0.04	81	0.04	63	0.04	11	0.04
14	7	0.03	69	0.03	1	0.03	49	0.04
15	55	0.03	95	0.03	2	0.03	6	0.04
16	42	0.03	71	0.03	77	0.03	9	0.04
17	62	0.03	49	0.03	91	0.03	43	0.04
18	89	0.02	45	0.03	69	0.02	44	0.03
19	25	0.02	32	0.03	43	0.02	67	0.03
20	58	0.02	67	0.03	25	0.02	98	0.02
21	20	0.02	20	0.03	49	0.02	24	0.02
22	100	0.01	44	0.02	32	0.01	38	0.02
23	18	0.01	43	0.02	18	0.01	84	0.02
24	79	0.01	59	0.01	89	0.01	34	0.01
25	15	0.01	98	0.01	37	0.01	41	0.01

APPENDIX B: Code Snippets

B.1 Analysis: Forecast Error Calculations and Data Initialization

```
capacity_MW=100; %capacity of PV power plant
Power_pu=Power_MW/capacity_MW;
DA_Power_pu=DA_Power_MW/capacity_MW;
forecast_error=abs(DA_Power_MW-Power_MW); %normalised
    forecast error ranging from 0 to 1
forecast_error_pu=forecast_error/capacity_MW;
avgdata=(Power_pu+DA_Power_pu)/2;
data=horzcat(Power_pu,DA_Power_pu,GHISolar,avgdata,
    forecast_error);
```

B.2 Uncertainty and Variability Indices

B.2.1 Assignment of uncertainty levels to hourly data

```
%  assign uncertainty levels to hourly data:
u=zeros(size(Power_pu));
for(i=1:length(Power_pu))
    if  0<=abs(DA_Power_pu(i,1)-Power_pu(i,1))&& abs(
        DA_Power_pu(i,1)-Power_pu(i,1))<=e1
        u(i,1)=u1;
    elseif e1<abs(DA_Power_pu(i,1)-Power_pu(i,1))&& abs(
        DA_Power_pu(i,1)-Power_pu(i,1))<=e2
        u(i,1)=u2;
    elseif e2<abs(DA_Power_pu(i,1)-Power_pu(i,1))&& abs(
        DA_Power_pu(i,1)-Power_pu(i,1))<=e3
        u(i,1)=u3;
    elseif e3<abs(DA_Power_pu(i,1)-Power_pu(i,1))&& abs(
```

```
        DA_Power_pu(i,1)-Power_pu(i,1))<=e4
        u(i,1)=u4;
    elseif e4<abs(DA_Power_pu(i,1)-Power_pu(i,1))&& abs(DA_Power_pu(i,1)-Power_pu(i,1))<=e5
        u(i,1)=u5;
    else e5<abs(DA_Power_pu(i,1)-Power_pu(i,1))
        u(i,1)=u5;
    end
    i=i+1;
end
uncertainty=u;
```

B.2.2 Assignment of variability levels to hourly data

```
%assign variablity levels to hourly data:
v=zeros(size(Power_pu));
for(i=2:length(Power_pu))
    if 0<=abs(Power_pu(i,1)-Power_pu(i-1,1))&& abs(Power_pu(i,1)-Power_pu(i-1,1))<=e1
        v(i,1)=u1;
    elseif e1<abs(Power_pu(i,1)-Power_pu(i-1,1))&& abs(Power_pu(i,1)-Power_pu(i-1,1))<=e2
        v(i,1)=u2;
    elseif e2<abs(Power_pu(i,1)-Power_pu(i-1,1))&& abs(Power_pu(i,1)-Power_pu(i-1,1))<=e3
        v(i,1)=u3;
    elseif e3<abs(Power_pu(i,1)-Power_pu(i-1,1))&& abs(Power_pu(i,1)-Power_pu(i-1,1))<=e4
```

```
            v(i,1)=u4;
        elseif e4<abs(Power_pu(i,1)-Power_pu(i-1,1))&& abs(
          Power_pu(i,1)-Power_pu(i-1,1))<=e5
            v(i,1)=u5;
        else e5<abs(Power_pu(i,1)-Power_pu(i-1,1))
            v(i,1)=u5;
        end
        i=i+1;
end
variability=v;
```

B.3 Code for Scenario Generation Methodologies

B.3.1 Scenario Generation Methodology I: Based on Uncertainty and Variability Indices

```
scengen=zeros(scenum,24); %creating scenario matrix of size
    scenum x 24
r=1;
c=1;
for p=1:scenum*24
        if mod(r,scenum)==0
            r=1;
            c=c+1;
        else if havguc(c,1)>=0 && DA_test(c)==0
                scengen(r,c)=0;
            else if havguc(c,1)>0 && havguc(c,1)<=u1
                    scengen(r,c)=randi([(DA_test(c)), DA_test(c)+
                        e1*capacity_MW]);
```

```
            else if havguc(c,1)>u1 && havguc(c,1)<=u2
                scengen(r,c)=randi([(DA_test(c)), DA_test(c)+e2*capacity_MW]);
                else if havguc(c,1)>u2 && havguc(c,1)<=u3
                    scengen(r,c)=randi([(DA_test(c)-e3*capacity_MW), DA_test(c)+e3*capacity_MW]);
                    else if havguc(c,1)>u3 && havguc(c,1)<=u4
                        scengen(r,c)=randi([(DA_test(c)-e4*capacity_MW), DA_test(c)+e4*capacity_MW]);
                        else if havguc(c,1)>u4 && havguc(c,1)<=u5
                            scengen(r,c)=randi([(DA_test(c)-e5*capacity_MW), DA_test(c)+e5*capacity_MW]);
                        end
                    end
                end
            end
        end
    end
    r=r+1;
  end
end
```

B.3.2 Scenario Generation Methodology II:Based on Clustering

```
%% Scengen using K-Means Clustering
clustnum=5;
[idx,error] = kmeans(Power_MW,clustnum);
error=sort(error); %sorts in ascending-order. (predictable, most predictable, uncertain, moderately uncertain, highly uncertain)
%grouped real data into respective clusters
cluster_index=horzcat(idx,Power_MW);
ind1 = cluster_index(:,1) == 1;
ind2 = cluster_index(:,1) == 2;
ind3 = cluster_index(:,1) == 3;
ind4 = cluster_index(:,1) == 4;
ind5 = cluster_index(:,1) == 5;
cluster1 = cluster_index(ind1,:);
cluster2 = cluster_index(ind2,:);
cluster3 = cluster_index(ind3,:);
cluster4 = cluster_index(ind4,:);
cluster5 = cluster_index(ind5,:);
%clusters
cluster1(:,1) =[];
cluster2(:,1) =[];
cluster3(:,1) =[];
cluster4(:,1) =[];
cluster5(:,1) =[];
cluster={cluster1,cluster2,cluster3,cluster4,cluster5};
```

```
clear ind1 ind2 ind3 ind4 ind5 cluster_index
%% Statistical Analysis of Clusters
stats1 = [mean(cluster1) std(cluster1) var(cluster1)];
stats2 = [mean(cluster2) std(cluster2) var(cluster2)];
stats3 = [mean(cluster3) std(cluster3) var(cluster3)];
stats4 = [mean(cluster4) std(cluster4) var(cluster4)];
stats5 = [mean(cluster5) std(cluster5) var(cluster5)];
%%
D = pdist2(error,DA_test','squaredeuclidean');
[M,I]=min(D);
scenum=scenum+1;
scengen=zeros(scenum,24); %creating scenario matrix of size
    scenum x 24
for p=1:24

    rng('default');
    rng(1); % for reproducilbility
    cls=I(p);
    a=std(cell2mat(cluster(cls)));
    b=DA_test(p);
    c=round(var(cell2mat(cluster(cls))));
    if p==1
        scengen = abs(a.*randn(scenum,1) + b); %generate
            scenarios from normal distribution of cluster
    else
        scengen1 = abs(a.*randn(scenum,1) + b);
        scengen=horzcat(scengen,scengen1);
```

```
    end
end
```

B.4 Code for Scenario Reduction Methodologies

B.4.1 Scenario Reduction Methodology I: Based on Probability Distances

```
%% Reducing scenario set to consider only diurnal data
%(this is done to avoid sparsity in cost function matrix)
scengen =abs(scengen(1:end-lastn,:));
scengen_SR=(scengen); %% only positive elements from scenarios
zm=[];
for m=1:24
    if mean(scengen_SR(:,m))==0
        zm=horzcat(zm,m);
    end
    m=m+1;
end
scengen_SR(:,zm)=[];
scenum_SR=size(scengen_SR,1);

clearvars r c p lastn m
clear zm
%% Scenario Reduction
scenrednum=round((1-lambda)*scenum); %desired cardinality of reduced scenario set
tp=size(scengen_SR,2); %number of time stamps
%------scenred code starts here------%
```

```
% calculating cost function using distance matrix
p=1; nn=1;
CF=zeros(scenum_SR); %cost function

while p<=tp
    c=1; r=1;
    for nn=1:(scenum_SR+numel(CF))
        if mod(c,scenum_SR+1)==0
            c=1;
            r=r+1;
        else
            CF(r,c)=CF(r,c)+abs(scengen_SR(r,p)-scengen_SR(c,p)); %calculating distance between scenario pairs at each time stamp
            c=c+1;
        end
        nn=nn+1;
    end
    p=p+1;
end
clearvars nn c r p
CF_original(:,:,1)=CF;
% calculating Kantorovich Distances
CF_nonselect=zeros(scenum_SR,scenum_SR);%initialize CF for the first time
k_dist=zeros(); %Initializing kantorovich distances matrix
scen_prob=ones(1,scenum_SR)/scenum_SR; %Initializing scenario
```

```
probability matrix
k_dist=(sum(scen_prob.*CF,2)); %Initial kantorovich distances
% finding the scenario with the minimum distance (M: value, I: index)
[Mvalue,Mindex]=min(k_dist);
selected_scen=Mindex; %matrix of selected scenarios
nonselected_scen=(find(k_dist~=Mvalue)); % matrix of non-selected scenarios
p=1;
nn=1;
s=[];
s(p)=selected_scen;
while p<=scenrednum
    %Initializing reduced scenario index set
    while nn<=scenrednum
        if nn==1
            break
        else
            %computing intermediate CF matrix for k_dist calculations
            CF_nonselect=CF;
            scen_prob1=scen_prob;
            s( :, ~any(s,1) ) = [];
            CF_nonselect(:,s)=0;
            CF_nonselect(s,:)=0;
            scen_prob1(:,s)=0;
            k_dist1=(sum(scen_prob1.*CF_nonselect,2));
```

```
        k_dist1(s,:)=max(k_dist1)+1;
        [Mvalue,Mindex]=min(k_dist1);
        selected_scen=Mindex; %matrix of selected scenarios
        s(1,p)=selected_scen;
        nonselected_scen=[1:scenum_SR]';
        nonselected_scen(s,:)=[]; % matrix of non-selected scenarios
        if nn~=scenrednum
            break
        else
            nn=scenrednum+1;
        end
    end
end
nn=nn+1;
%% Updating cost matrix
c=1;
r=1;
for q=1:(scenum_SR+numel(CF))
    if mod(c,scenum_SR+1)==0
        c=1;
        r=r+1;
    else if r==c
            CF(r,c)=0;
        else if c==selected_scen
                CF(r,c)=CF(r,c);
```

```
                    else if r==selected_scen
                            CF(r,c)=CF(r,c);
                        else
                            CF(r,c)=min(CF(r,selected_scen),CF(r,
                                c));
                        end
                    end
                end
                c=c+1;
            end
            q=q+1;
            CF1(:,:,p)=CF;
        end
        p=p+1;
end
nonselected_scen=(nonselected_scen)';
s=sort(s,2);
nonselectednum=length(nonselected_scen);
```

B.4.2 Scenario Reduction Methodology II:Based on K-means Clustering

```
%% SCENARIO REDUCTION
scenrednum=round((1-lambda)*scenum); %desired cardinality of
    reduced scenario set
tp=size(scengen_SR,2); %number of time stamps

%% K-Means Clustering for Scenario Reduction
%---------scenred code starts here---------%
```

```
clustnum=scenrednum;
[idx,error] = kmeans(scengen,clustnum);
%idx:   classification of all scenarios by clusters. (scenum
    x 1)
%error: entities of all clusters (scenrednum x 24)

%grouping of scenarios into respective clusters
cluster_index=horzcat(idx,scengen);
%finding scenarios with same cluster index
for x=1:clustnum
    ind{x}=cluster_index(:,1)==x;
end
%grouping scenarios according to cluster indices
for y=1:clustnum
    cluster_grp{y}=cluster_index(cell2mat(ind(y)),:);
end
%creating cell comprising of differently-sized clusters
cluster={};
for z=1:clustnum
    intcluster=cell2mat(cluster_grp(z));
    intcluster(:,1)=[];
    cluster(z)=mat2cell(intcluster,size(intcluster,1),size(
        intcluster,2));
    clear intcluster
end

%Calculating pair-wise distances between DAF and cluster
```

```
entities
for i=1:clustnum
    D{i} = pdist2(DA_test,cell2mat(cluster(i)),'euclidean');
    Dsz(i,:)=size(D{i});                                  % Size Of Each Vector
end
Colmax=max(Dsz(:,2));                                     % Maximum # Columns
Dmtx=NaN(i,Colmax);                                       % Preallocate
for ii=1:i
    Dmtx(ii,1:Dsz(ii,2))=D{ii};                           % Fill Matrix
end
clear x y z i ii Dmtx Colmax Dsz

%ranking scenarios within each cluster
rankscen={}; d={};
for i=1:clustnum
    d(i)=D(i);
    d1=cell2mat(d(i));
    [scen,rnk]=sort(d1,'descend');
    rankscen{i}=vertcat(scen,rnk);
    clear scen rnk d1  i
end

%% extracting highest-ranked scenario from each cluster
```

```
selectedscenrank=zeros(1,clustnum);
selectedscen=zeros(clustnum,24);
for i=1:clustnum
    a=cell2mat(cluster(i));
    b=cell2mat(rankscen(i));
    selectedscenrank(i)=b(end);                 %selecting
       the scenario with the lowest probability distance
    selectedscen(i,:)=a(selectedscenrank(i),:); %creating
       matrix of reduced scenarios
end
```

B.4.3 Probability Re-distribution Among Reduced Scenarios

```
%% Probability distribution within clusters:
scen_prob=ones(1,scenum)/scenum;                 %determining
    initial probabilities of all scenarios
clustsize=zeros(1,clustnum);
scenredprob=zeros(1,clustnum);
for i=1:clustnum
    aa=cell2mat(cluster(i));
    clustsize(i)=size(aa,1);                     %determining
        size of each cluster
    scenredprob(i)=scen_prob(1,1)*clustsize(i);  %
       distributing probabilites among clusters
end
```

B.4.4 Ranking of Scenarios

```
%% Finding the original scenario number of reduced scenarios
[C,ia,ib]=intersect(scengen,selectedscen,'rows');
```

```
D=horzcat(ia,ib);
D(ib)=ia;
D(:,2)=[];
s_ranked=D';
%ia: scenario number in original scenario set
%ib: cluster number to which scenario belongs

%% Ranking the final set of reduced scenarios
scenred=vertcat(s_ranked,scenredprob);
scenred=sortrows(scenred',2,'descend');
```

APPENDIX C: NREL Data Graphical Description

C.1 Comparison of 8760 Data

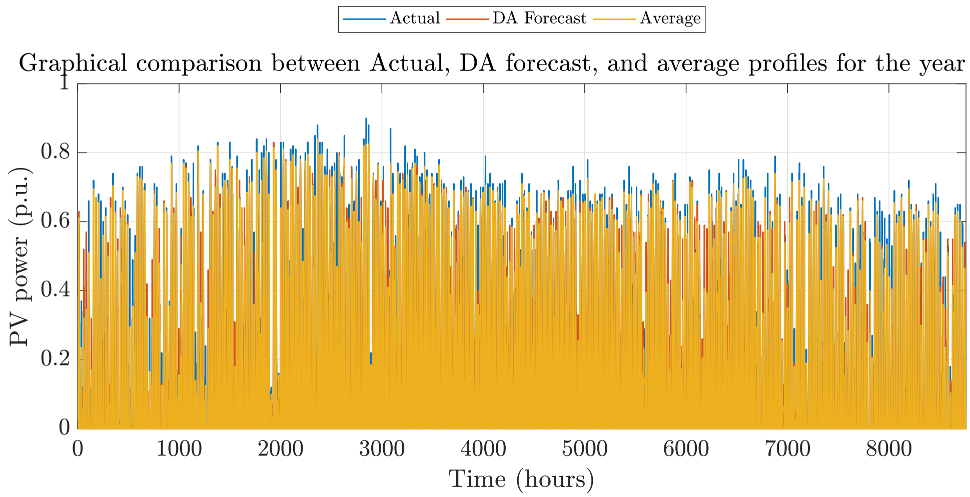

Fig. C.1: Comparison of Actual, Day-Ahead Forecast, and 8760 PV Generation Data

C.2 Forecast Errors

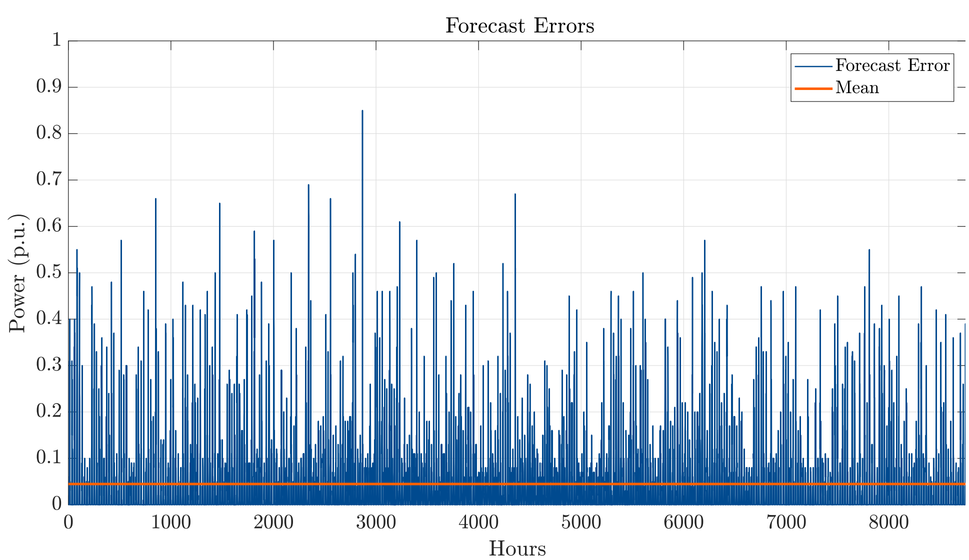

Fig. C.2: Forecast Errors for 8760 Data

APPENDIX D: 200-Generator Test-System: Generator Parameters

The following figures provide the generator parameters for the units in the 200-Generator test system using in unit commitment case study of chapter 5.

GENERATION UNITS DATA									
Sr. No.	**Pmin**	**Pmax**	**a**	**b**	**c**	**UH**	**DH**	**UR**	**DR**
1	78.1906	215.878	0.005359	7.1279	529.473	6	6	54.1205	35.4206
2	76.7302	222.955	0.000617	7.79228	433.175	6	6	43.9067	59.6891
3	63.1443	179.084	0.063177	7.22072	294.823	2	3	44.2867	55.7218
4	64.0049	197.525	0.047074	7.29861	195.44	2	3	42.5249	34.9529
5	43.2432	124.509	0.088557	4.12656	136.242	1	1	161.954	65.342
6	67.3907	191.995	0.052698	7.30746	323.715	2	3	55.8034	38.3005
7	66.4489	163.24	0.037892	6.96832	269.177	2	3	28.6486	26.1962
8	40.5936	119.494	0.099686	4.35908	136.145	1	1	118.781	132.171
9	49.3115	121.374	0.071523	4.75838	133.424	1	1	98.2978	41.8115
10	89.3823	329.552	0.024416	8.31428	409.696	6	6	76.8804	60.671
11	36.2111	110.945	0.085868	4.87963	132.379	1	1	127.48	45.1175
12	72.1113	292.199	0.015899	7.5389	402.651	6	6	65.5171	89.5873
13	87.3351	201.928	0.000864	7.97401	543.069	6	6	43.5863	55.6145
14	61.0929	174.705	0.049772	7.30887	214.952	2	3	47.9905	41.5853
15	68.6248	189.604	0.035124	7.34268	271.213	2	3	58.9548	45.3146
16	34.8805	103.16	0.074892	4.76901	147.246	1	1	78.9251	35.7119
17	45.8666	104.114	0.07405	4.63344	148.848	1	1	39.8788	60.9546
18	65.4674	161.216	0.049561	7.18776	204.707	2	3	37.6862	39.6775
19	51.3764	186.285	0.040397	6.85097	335.058	2	3	34.3958	46.0225
20	88.2891	312.032	0.029186	7.48167	502.304	6	6	64.2044	67.6675
21	60.5014	155.159	0.034328	7.04958	244.767	2	3	44.0188	30.302
22	41.5818	120.098	0.095331	4.2306	125.083	1	1	56.5887	44.1197
23	39.1647	110.429	0.06953	4.4105	117.335	1	1	76.5312	37.9127
24	54.4942	198.37	0.060153	6.74463	276.292	2	3	46.4188	51.8247
25	71.8915	318.741	0.029351	7.01035	531.14	6	6	63.7774	78.3845
26	61.8925	184.158	0.046314	7.22613	195.718	2	3	33.586	57.7822
27	63.7431	180.457	0.050864	7.10342	288.162	2	3	31.9625	32.9796
28	83.4284	265.26	0.00005	7.54682	456.874	6	6	81.1196	52.5663
29	81.7878	295.936	0.010227	7.09833	501.384	6	6	87.6094	71.7396
30	67.0971	165.044	0.040403	6.77676	207.461	2	3	25.9076	30.8019
31	52.1198	187.257	0.034203	7.06522	209.703	2	3	43.0802	38.5724
32	80.5774	293.682	0.020431	7.4716	474.032	6	6	57.0957	92.5366
33	42.4906	100.328	0.082121	4.34284	102.835	1	1	37.4791	56.0049
34	74.1182	266.109	0.013644	7.46258	488.108	6	6	91.8549	79.4514
35	30.304	110.381	0.076134	4.39427	149.907	1	1	54.8107	91.1382
36	35.9389	125.608	0.094907	4.7571	146.078	1	1	118.547	71.3343
37	90.7547	318.923	0.003715	7.23461	402.083	6	6	92.2897	79.4938
38	89.0646	240.118	0.01564	7.18133	433.262	6	6	46.695	43.8668
39	66.375	182.4	0.036803	7.05428	212.96	2	3	34.9044	57.0428
40	50.8625	163.977	0.054947	7.08893	283.968	2	3	55.2112	56.4009
41	86.8087	262.36	0.030382	8.16138	512.993	6	6	80.5845	51.8812
42	95.2025	215.275	0.029634	7.26514	528.233	6	6	46.3274	55.1376

Fig. D.1: Generator Parameters Part I

43	99.4498	231.033	0.012448	7.60253	510.224	6	6	41.7385	33.2495
44	87.9229	210.486	0.006821	7.37927	432.337	6	6	59.2456	54.4929
45	47.3583	106.494	0.085908	4.61596	148.895	1	1	63.6532	66.0486
46	47.9638	120.428	0.076732	4.89569	105.15	1	1	41.9778	59.5984
47	91.4606	318.61	0.007947	7.91093	532.998	6	6	72.397	74.7694
48	83.3671	320.605	0.014307	8.3845	462.143	6	6	79.9175	104.22
49	73.2511	256.901	0.005512	8.38977	432.539	6	6	48.2639	67.3762
50	50.6049	161.043	0.051303	7.1649	276.814	2	3	28.5195	41.1341
51	38.7985	113.513	0.085142	4.59093	147.101	1	1	85.1149	59.8755
52	72.2944	289.786	0.006839	7.03323	458.17	6	6	105.88	61.7971
53	47.8112	123.929	0.096994	4.4279	143.439	1	1	62.4013	69.8426
54	60.7865	193.951	0.051615	7.37635	319.496	2	3	46.7185	56.8064
55	68.3331	161.19	0.049021	7.23916	202.232	2	3	26.7329	24.4962
56	88.2754	271.616	0.013483	7.60051	456.192	6	6	84.4107	50.616
57	84.593	208.693	0.028716	7.03914	493.931	6	6	33.5886	39.8437
58	41.9797	123.942	0.099689	4.20719	110.802	1	1	51.8781	141.164
59	78.294	273.921	0.026565	7.74096	480.519	6	6	70.8612	75.9149
60	75.9502	252.386	0.026461	8.33767	405.8	6	6	70.5979	65.3501
61	33.0641	113.32	0.089307	4.75381	119.538	1	1	46.6998	64.9797
62	44.116	124.516	0.099533	4.02438	122.987	1	1	55.947	45.1431
63	53.5261	152.393	0.048417	7.00782	316.244	2	3	44.4946	46.7658
64	96.0924	297.642	0.01304	8.29391	441.2	6	6	85.615	76.8445
65	65.8812	184.153	0.061093	6.89259	260.705	2	3	52.0442	47.5722
66	68.4466	171.946	0.052362	6.85319	282.293	2	3	35.5779	28.8887
67	82.4992	279.312	0.014761	7.37533	523.925	6	6	77.7023	58.9821
68	50.3595	197.913	0.056543	6.83651	219.874	2	3	39.8345	61.0756
69	73.9753	303.779	0.014825	8.22808	444.926	6	6	61.591	102.712
70	80.6754	231.795	0.023751	8.14797	430.099	6	6	39.2241	55.7213
71	41.8851	116.199	0.070949	4.3538	149.605	1	1	49.5466	115.424
72	60.7669	181.251	0.035695	6.93371	251.476	2	3	50.1142	53.7056
73	88.8888	311.889	0.031087	8.36285	532.765	6	6	99.9841	109.364
74	31.0779	100.18	0.077614	4.26377	115.397	1	1	60.38	67.1156
75	70.9027	309.088	0.028148	7.59534	513.822	6	6	115.888	107.992
76	53.5603	198.909	0.037749	7.23046	312.094	2	3	52.0746	65.4407
77	69.8181	193.39	0.053421	6.7313	224.176	2	3	50.5628	39.8193
78	96.9063	205.888	0.015173	7.07466	461.96	6	6	28.7788	45.3816
79	95.6246	232.517	0.01596	8.21036	433.308	6	6	58.1497	43.8068
80	44.2875	116.368	0.088154	4.541	120.562	1	1	103.344	114.147
81	78.9111	307.033	0.012415	7.1307	514.792	6	6	113.053	90.1454
82	58.9956	166.814	0.056857	6.87951	312.914	2	3	36.6422	29.6383
83	44.5116	126.995	0.088358	4.13376	135.51	1	1	73.4949	69.7023
84	37.7346	112.503	0.097589	4.60033	135.141	1	1	98.7278	38.6528
85	66.7486	173.925	0.04502	6.81809	310.214	2	3	31.3066	33.1357
86	36.686	114.147	0.073092	4.3976	123.142	1	1	40.0412	43.3394
87	35.9261	124.493	0.092563	4.49516	139.987	1	1	91.164	126.805
88	52.3572	175.602	0.037446	6.89453	316.557	2	3	44.3786	32.9856

Fig. D.2: Generator Parameters Part II

89	43.2737	125.737	0.070867	4.96146	128.224	1	1	53.8461	51.1898
90	35.1155	122.038	0.080611	4.0585	126.675	1	1	80.2898	55.3726
91	54.1475	165.461	0.061599	7.17195	315.448	2	3	32.4941	37.4161
92	60.2585	184.03	0.040815	6.85685	199.434	2	3	31.7863	43.0559
93	54.9477	161.788	0.0556	6.73826	232.072	2	3	44.3334	42.0581
94	90.0754	249.474	0.032945	8.44044	522.272	6	6	46.7307	50.5439
95	52.2932	159.316	0.048798	6.81991	291.058	2	3	41.7259	48.5815
96	87.7993	233.271	0.022216	7.04692	488.351	6	6	52.3889	39.7036
97	59.7439	191.296	0.042506	6.93452	242.965	2	3	33.6217	48.3899
98	72.3118	282.121	0.018676	8.15081	438.856	6	6	75.6357	70.097
99	88.1179	230.382	0.00279	8.17008	424.528	6	6	56.5246	47.1973
100	61.0971	169.889	0.034377	6.92664	296.449	2	3	32.0378	51.2402
101	82.1439	240.257	0.026024	7.26904	459.255	6	6	66.9177	47.9491
102	30.7794	120.849	0.070568	4.17472	145.57	1	1	87.3052	128.318
103	45.9813	115.258	0.090806	4.47472	136.064	1	1	39.7788	56.3319
104	79.4623	268.418	0.001057	7.14383	431.706	6	6	55.3505	68.5372
105	99.0872	243.57	0.016	8.20098	455.061	6	6	37.7636	53.7094
106	61.2662	171.619	0.042012	7.29156	239.674	2	3	33.0688	31.8456
107	33.9796	104.726	0.072319	4.59639	140.329	1	1	60.6021	48.0854
108	56.9826	171.891	0.063171	7.0126	195.254	2	3	37.4163	38.859
109	39.7214	108.872	0.085774	4.8037	122.926	1	1	119.409	43.0786
110	37.5405	129.058	0.092524	4.1583	127.163	1	1	78.4623	70.0814
111	95.5824	264.462	0.005521	7.19112	511.144	6	6	54.7446	78.047
112	47.6879	113.945	0.09112	4.89502	140.809	1	1	39.6501	63.1447
113	65.1085	167.518	0.061258	6.89897	251.843	2	3	31.7528	46.4693
114	49.9335	111.508	0.097534	4.22263	101.48	1	1	38.24	37.1009
115	98.2888	218.786	0.027208	7.64451	410.813	6	6	35.5557	53.288
116	37.8488	108.091	0.090244	4.59261	129.463	1	1	40.2476	51.2046
117	40.0333	112.313	0.087892	4.35136	125.546	1	1	94.7493	42.9346
118	69.8419	176.337	0.038218	7.21961	207.779	2	3	53.1715	30.2754
119	50.539	157.321	0.04159	7.19363	313.378	2	3	41.3118	31.9072
120	69.2022	195.346	0.041841	6.98238	311.63	2	3	50.8395	32.4301
121	64.2308	187.101	0.062155	7.14847	316.132	2	3	39.7938	46.8522
122	53.202	192.474	0.048533	6.78631	339.096	2	3	45.6636	43.3697
123	88.7762	255.365	0.030605	8.04909	491.166	6	6	53.2631	80.4726
124	44.5152	106.001	0.081074	4.33	141.699	1	1	34.5878	89.9536
125	88.3404	241.249	0.02651	8.37077	475.309	6	6	63.2637	38.7241
126	96.4934	215.263	0.015314	8.37246	407.732	6	6	36.1247	39.4441
127	96.9576	277.166	0.011777	7.21809	505.596	6	6	48.7999	62.1084
128	40.549	102.565	0.093	4.83306	137.616	1	1	40.5021	37.0221
129	39.2074	123.308	0.069897	4.51155	114.476	1	1	89.8569	42.5549
130	59.8917	151.695	0.052701	6.72452	224.366	2	3	41.061	31.2066
131	86.3152	224.693	0.02384	7.5264	524.396	6	6	54.6227	55.865
132	51.568	181.988	0.045795	7.07191	290.413	2	3	58.7322	33.5216
133	45.2715	105.482	0.089698	4.89895	107.933	1	1	103.165	61.5819
134	63.3671	194.429	0.05495	7.12529	317.299	2	3	35.2975	47.4583

Fig. D.3: Generator Parameters Part III

135	82.2282	203.011	0.016714	7.18462	484.176	6	6	43.876	31.7383
136	56.4882	158.272	0.055292	7.0862	287.23	2	3	42.1605	31.3472
137	34.1401	128.326	0.072829	4.86877	144.955	1	1	52.4124	61.4016
138	32.8614	118.047	0.071105	4.59667	116.268	1	1	65.4595	48.9098
139	40.3824	114.422	0.0908	4.88549	141.749	1	1	56.4756	49.5026
140	95.1173	327.147	0.015698	7.16549	444.24	6	6	66.4847	68.9633
141	62.045	185.585	0.058737	7.32452	237.111	2	3	50.3823	33.5919
142	51.9245	186.789	0.044264	6.81474	327.66	2	3	37.8944	39.2913
143	50.539	165.29	0.045429	7.36932	223.575	2	3	43.9647	33.5191
144	83.965	266.402	0.007804	7.71221	425.091	6	6	48.3159	46.4457
145	54.4356	167.371	0.048224	7.11207	211.28	2	3	50.7051	32.1575
146	68.4143	171.319	0.040055	7.27515	275.637	2	3	26.9521	38.8823
147	63.1474	160.295	0.043522	7.06475	238.346	2	3	28.0923	28.5673
148	79.5648	326.719	0.002344	7.95126	424.972	6	6	88.9769	97.9286
149	36.0555	122.077	0.07549	4.41783	149.8	1	1	92.406	106.451
150	94.7447	206.507	0.023391	8.26356	489.688	6	6	28.0346	39.6057
151	90.3967	324.426	0.006043	8.07614	496.66	6	6	89.7859	77.0139
152	78.3197	309.235	0.009754	7.14379	514.431	6	6	99.9233	80.0016
153	43.5539	129.485	0.092399	4.59352	131.301	1	1	124.765	67.0394
154	30.8881	119.554	0.069842	4.21146	104.976	1	1	114.666	64.6391
155	88.367	204.884	0.012744	8.09775	500.391	6	6	39.1348	37.9567
156	47.5903	111.176	0.068938	4.64507	134.458	1	1	42.1241	36.5767
157	86.3701	254.369	0.006217	8.09066	509.688	6	6	59.7831	49.0373
158	58.3139	197.339	0.034471	6.87909	223.795	2	3	55.825	60.4794
159	34.5912	109.391	0.079292	4.14655	116.564	1	1	53.4129	51.6633
160	87.7123	221.603	0.023013	7.80619	480.326	6	6	49.7322	34.0118
161	83.1465	214.267	0.028829	7.84904	536.647	6	6	43.4157	35.6638
162	72.9618	324.648	0.01745	7.28854	402.229	6	6	112.732	79.4524
163	40.0107	100.631	0.087702	4.36601	146.179	1	1	42.7565	43.8859
164	53.379	179.217	0.037486	7.36776	250.202	2	3	40.2968	32.3079
165	62.1464	150.478	0.053726	6.77844	211.749	2	3	22.7093	23.1767
166	49.9762	100.244	0.074626	4.78512	102.324	1	1	48.1296	51.1733
167	74.5283	325.096	0.000727	7.59058	430.003	6	6	77.8159	107.886
168	47.7905	103.415	0.074117	4.83786	133.48	1	1	95.3668	28.2395
169	47.7526	122.948	0.087785	4.94717	107.83	1	1	45.1019	57.6394
170	79.7827	307.211	0.025637	8.39727	526.859	6	6	97.787	109.272
171	55.9651	193.049	0.050797	7.08417	279.895	2	3	57.2311	35.0887
172	38.8607	107.223	0.068648	4.79797	107.787	1	1	55.6796	87.0638
173	96.1116	315.503	0.018438	8.38414	476.037	6	6	68.8305	77.6815
174	74.3544	215.358	0.019607	7.82208	438.705	6	6	65.8626	44.8761
175	35.2742	113.373	0.077248	4.4676	146.789	1	1	68.7769	63.5315
176	45.269	123.593	0.09856	4.7839	144.923	1	1	67.482	57.077
177	44.3919	128.367	0.09997	4.67745	112.209	1	1	78.5465	91.8149
178	49.4031	103.853	0.079005	4.03861	139.557	1	1	32.3116	93.6316
179	34.9159	125.217	0.07372	4.20347	106.87	1	1	49.057	57.7792
180	99.165	224.15	0.025655	7.95085	449.619	6	6	53.8439	44.6751

Fig. D.4: Generator Parameters Part IV

181	45.6359	117.787	0.090659	4.66088	136.785	1	1	76.2728	68.1441
182	57.5759	174.552	0.040663	6.85381	323.91	2	3	35.4595	50.8121
183	54.2726	159.557	0.0421	7.3047	285.374	2	3	51.6043	32.438
184	68.6883	156.626	0.043449	7.01224	346.587	2	3	41.5097	25.306
185	53.4712	168.334	0.036568	6.97227	223.946	2	3	37.4558	34.4492
186	43.8682	129.348	0.085368	4.73873	100.803	1	1	103.351	64.7836
187	98.6276	268.017	0.008628	7.19062	439.817	6	6	74.4499	79.91
188	40.839	125.564	0.088657	4.62444	127.01	1	1	104.333	160.668
189	89.407	230.811	0.021246	7.75515	460.573	6	6	35.8571	53.6258
190	68.3599	174.427	0.050629	6.77943	238.439	2	3	30.5829	48.8077
191	84.1288	271.564	0.027114	8.45752	471.445	6	6	57.2417	52.703
192	79.1949	250.838	0.014908	7.38586	422.546	6	6	80.5603	54.4309
193	84.9547	314.769	0.025655	7.93808	531.455	6	6	89.2685	66.0196
194	50.1337	170.653	0.03661	7.06877	226.08	2	3	33.2878	49.4153
195	86.1998	272.885	0.027996	7.66446	489.312	6	6	58.1122	53.3278
196	53.1495	194.192	0.037633	7.27447	317.797	2	3	38.733	40.5969
197	83.13	296.666	0.012238	7.64414	481.196	6	6	74.6291	64.2828
198	65.902	170.289	0.047282	7.35379	265.383	2	3	31.9083	29.7647
199	60.2133	173.109	0.045695	6.99413	230.392	2	3	35.0009	35.346
200	39.5938	108.946	0.082988	4.00925	135.691	1	1	68.0724	55.8756

Fig. D.5: Generator Parameters Part V