UNIVERSITY OF CALIFORNIA SAN DIEGO

Selection of Representative Days in Microgrid Planning

A thesis submitted in partial satisfaction of the
requirements for the degree
Master of Science

in

Engineering Sciences (Mechanical Engineering)

by

Kelsey Taylor Fahy

Committee in charge:

Professor Jan Kleissl, Chair
Professor Keiko Nomura
Professor George Tynan
Professor David G. Victor

2019

The thesis of Kelsey Taylor Fahy is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

To my parents and my husband, for your unwavering faith, support, and encouragement.

EPIGRAPH

The Road goes ever on and on,

Down from the door where it began.

Now far ahead the Road has gone,

And I must follow, if I can,

Pursuing it with eager feat,

Until it joins some larger way

Where many paths and errands meet.

And whither then? I cannot say.

—John Ronald Reuel Tolkien

TABLE OF CONTENTS

LIST OF SYMBOLS

C	Total annual energy cost
L	Annual hourly demand data
$c_{utility}$	Utility purchases and fees
c_{invest}	Annualized investment costs of technologies
$c_{O\&M}$	Technology operation and maintenance costs
u	Electricity purchased (kWh)
g	Electricity produced and/or consumed by technologies
j	Index for technology unit
$L_{m,d,h}$	
$RD_{m,pk}$	Representative Peak Demand Profile for month m
$RD_{m,wd}$	Representative Weekday Demand Profile for month m
$RD_{m,we}$	Representative Weekend Demand Profile for month m
σ_m	Upper bound on number of peak days for month m
σ_m^p	Number of peak days tested for month m, iteration p
$N_{m,wd}^p$	Number of weekdays for month m, iteration p
$N_{m,we}^p$	Number of weekends for month m, iteration p

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Jan Kleissl for his support as the chair of my committee. Working and writing under his guidance has been a pleasure.

I would also like to acknowledge my employers Adib Nasle and Michael Stadler for providing me the opportunity to do this research, and my coworker and friend Zachary Pecenak for guiding me through the adventure of research and paper writing.

And finally, a special acknowledgement to the best labmates. Your invaluable advice and support improved my work at every step, while your friendship kept me sane.

This thesis, in full, is currently being prepared for submission for publication of the material. Fahy, Kelsey; Stadler, Michael; Pecenak, Zachary; Haghi, Hamed Valizadeh; Kleissl, Jan. The thesis author was the primary investigator and author of this material.

ABSTRACT OF THE THESIS

Selection of Representative Days in Microgrid Planning

by

Kelsey Taylor Fahy

Master of Science in Engineering Sciences (Mechanical Engineering)

University of California San Diego, 2019

Professor Jan Kleissl, Chair

Optimization tools require prohibitive computational time to model energy systems with annual hourly input data, and input is typically reduced into representative periods to increase solver speed. Data reduction in microgrid optimizations impacts the objective function accuracy. Methods preserving demand data fluctuations through reduction into representative days show improved accuracy with increases in number of representative periods. This work presents a method of data reduction that aggregates annual hourly demand data into typical weekdays and weekends, while explicitly preserving demand peaks in distinct representative profiles. The proposed method is tested in an energy system optimization using historical 15-minute resolution annual demand data from a gymnasium in La Jolla, California, and the system is optimized in

terms of total annual costs. Results are in good agreement with a full-resolution optimization of the energy system, establishing the validity of the proposed technique. Additionally, a comparison of method performance demonstrates a significant improvement in accuracy with the inclusion of peak demand profiles.

Chapter 1

Introduction

1.1 General Background

The tradeoff between accuracy and computational speed is an important research question as optimization tools are increasingly used for the planning and design of least cost, maximum CO2 reduction microgrids. Effective planning can reduce investment costs, and operational costs can be minimized with smart dispatch of distributed energy resources (DER) to reduce demand charges, engage in price arbitrage opportunities, and generate revenue through providing services to the utility. Optimization tools are advantageous in their ability to determine a system configuration that efficiently and economically utilizes the available technologies.

Optimization tools using annual hourly granularity for time series data can take multiple days to complete a single run[SSF+18], [GGMM18], yet the influence of seasonally varying location-dependent variables (such as demand and solar insolation) on system design and operation requires optimization tools to represent these variables with a high level of granularity. These variables play a particularly important role for systems subject to time-of-use (TOU) rates and demand charges, as the interaction between fluctuating electricity prices and technology dispatch significantly impacts sizing decisions. Reducing the computational complexity required to solve

optimization problems while capturing key features of input data variability presents a challenge to the widespread use of optimization tools for microgrid planning.

A commonly used technique to increase the computation speed of energy systems optimization tools is the reduction of high resolution time series input data to sets of representative periods. Researchers have used 3-6 representative days with 1, 2, or 4-hour resolution to capture seasonal trends, such as in [YHI02], [ZLZC17], [MGRM17], [CPR09], and [MDFG08]. Aggregating time series data into 12 monthly representative days with 1-hour resolution is another approach seen in [MDFG08], [LRS10], [SSH$^+$17], and [MSMP13]. One commonality between all of these methods is that they do not apply any methodology to preserve the peak demand.

The following papers account for peak demand in their selection of representative days. [WY14], [WY15], and [WKY16] also use three typical seasonal days, as well as two additional days to capture seasonal winter and summer peak behavior. [LRCS09] uses two representative 1-hour resolution days for each month: one to represent typical working days, and one to represent festive-weekend days. Both [FBM$^+$14] and [DMCLCAGS11] account for extreme behavior by adding peak days as insulated clusters, noting this step as necessary for properly sizing the system. [BKS$^+$17] adds peak days to aggregated input data to account for thermal extremes. [MDFG08] reduces load data to seasonal profiles, preserving the maximum load value of all months within a season.

The focus of the previously mentioned works has been on the development of the optimization model. Several other works have actively tested the impact of a preferred data reduction technique on computation time and error. The performance quality of a k-means method integrated with a parametric ε-constraints optimization technique is evaluated in [FBM$^+$14]. Both [GGMM18] and [GFMM19] evaluate the performance of a k-means method that selects and integrates peak days. In [DMCLCAGS11] and [STSM], the use of k-medoids clustering methods to reduce demand data is tested.

Recently, researchers have begun to directly compare the performance of multiple methods

of data reduction, seeking to identify whether there is one technique or group of techniques that stands out in minimizing error across a range of models for power systems and microgrids. Teichgraeber and Brandt [TB19] compare k-means, k-medoids, and hierarchical clustering methods, as well as two shape-based clustering methods. Clustering methods are applied to hourly electricity prices and tested using two MILP formulations. Green et al.[GSV14] applied a k-means algorithm to cluster electricity demand and wind output data for Great Britain, and from within each cluster, tested 15 methods of selecting representative days. Pfenninger [Pfe17] presents a systematic analysis of downsampling, clustering, and heuristic techniques, with a focus on large power systems with high shares of renewable generation.

Schutz [SSF$^+$18] compares the performance of six different representative day selection techniques, testing four clustering methods (k-means, k-medoids, k-centers, k-medians) and two aggregation methods (monthly and seasonal). No methods for representing extreme behavior days are tested. Electricity and heat demand and solar irradiation are clustered, and an apartment building and a single-family house with generation and storage technologies are modeled. Schutz evaluates the deviation of the objective function value, comparing representative day solutions to a recalculation of the same energy system using annual hourly resolution input data. Schutz finds that as few as 4 days selected using k-medoids most closely and reliably approximate demand related costs, but all methods determine energy systems close to the optimal system chosen by the reference model.

Kotzur et al.[KMRS18] tested the performance of four different techniques: averaging, k-means, and two medoid based methods, a k-medoid clustering algorithm and a hierarchical algorithm. They applied the methods to two residential systems and a larger islanded system and use the objective function as a performance metric. They found that medoid-based methods tended to perform better than centroid-based or aggregation methods without inclusion of peak periods. However, a significant increase in objective function error is observed when moving from the residential cases to the islanded case. This is attributed to the reliance on storage in the islanded

system, as the optimizations using representative periods do not model inter-day storage. Although the objective function error and discrepancy of system design both improve with increased period length and increased number of representative periods, Kotzur et al.[KMRS18] conclude that representative periods are inadequate for modeling systems reliant on long-term storage, and that data reduction method suitability primarily depends on the system being modeled.

Common findings and methodologies are as follows: Pfenninger [Pfe17] found that heuristic methods of selection showed promise in handling inter-annual variations. Both Kotzur et al.[KMRS18] and Schutz [SSF+18] found that k-medoids performed better than other clustering methods when applied to demand profiles with greater fluctuations. Schutz [SSF+18] represents electricity prices as a single flat rate, and Kotzur et al.[KMRS18] does not specify whether flat or varying rates are used. This misses a key interaction between variations in building demand, the utility, and technology for demand charge reduction and price arbitrage. Pfenninger, Schutz, Gabrielli, and Kotzur each compare results of optimizing energy systems with storage using uncoupled representative periods against reference metrics calculated using annual hourly resolution optimizations. As a result, they introduce discrepancies caused by both data reduction and by storage operation within uncoupled representative periods, and do not directly assess how much of the objective function error is caused by one or the other. None of the researchers that directly compare data reduction techniques found a significant enough difference in performance to recommend an overall technique.

1.2 Contribution of Paper

The primary contribution of this paper is testing the integration of peak demand profiles through a comparison of demand data reduction techniques. A method for constructing representative weekday and weekend profiles while explicitly preserving demand peaks is introduced, and hereafter referred to as Monthly Peak Preservation (MPP). Nine different demand data reduction

approaches are compared, varying the method and the number of constructed peak days used in MPP. The testing methodology directly isolates the impact of the demand data reduction methods, avoiding additional discrepancies caused by storage dispatch between uncoupled days, and by data reduction of separate input time series.

The major contributions of this paper are:

- Introduction of demand data reduction method which explicitly captures demand peaks (Monthly Peak Preservation)
- Comparison of two classes of demand data reduction methods
- Demonstration of the importance of capturing demand peaks
- Isolate the impact that different input sources have on objective function error

1.3 Layout of Paper

The layout of the paper is as follows: Section 2.1 introduces the data reduction methods that will be tested. The MILP formulations used are introduced in Section 3.1, followed by an explanation of the testing procedure in Section 3.2. I identify the performance metrics that will be used to assess data reduction method performance in Section 3.2.2. Case studies and energy systems modeled are described in Section 3.3. Results are then discussed in Section 4.

Chapter 2

Data Aggregation Methodology

2.1 Demand Data Reduction Methods

2.1.1 Overview

In this paper, demand data is reduced into representative days using two classes of demand data reduction methods: a k-means clustering method, and the Monthly Peak Preservation method (MPP) introduced here. For both classes, the reduced demand data represents weekdays and weekends. MPP also includes representative profiles for peak days. These representative days $RD_{m,DT}$ are 24-hour, hourly resolution profiles for each month, and are constructed from annual hourly demand data that has been separated into weekday and weekend data sets. This ensures that the level of granularity of the reduced demand data matches or exceeds that of typical monthly and hourly variations in TOU rates and demand charges. (Holidays are disregarded, as the infrequency of holiday billing periods is not expected to significantly influence results.)

Similarly to the approach in [SGCM14], sizing and operation are optimized for 2-6 days per month, and variables such as annual consumption for these representative periods are projected out to obtain monthly and annual characteristics of the energy system. The scaling factors used in this projection are equal to the modeled number of days each optimized demand

profile represents within a month, and are referred to as $n_{m,DT}$. Multiple approaches within each class are considered, and all methods are introduced and formulated in the following sections.

2.1.2 Clustering: K-means

K-means is a commonly used approach in the manipulation of large data sets to create representative, but smaller subsets of the data [Jai10]. It is formulated as a greedy optimization algorithm which assigns points among clusters by calculating cluster centroids which minimize the sum of the within-cluster sums of point-to-centroid distance (the distance measure $d(x_i,\mu_k)$). The distance measure used is the squared Euclidean distances between the points x_i within a cluster and the cluster centroid μ. The distance minimization is shown in Eq. (2.1). K-means iterates to reassign points and recalculate cluster centroids, decreasing the total sum of distances and the number of reassignments until the algorithm reaches a minimum. The number of clusters are defined a priori. Final centroids are the empirical means of their clusters, rather than selected members of the clusters.

$$\min[\sum_{k=1}^{N_k}\sum_{i=1}^{N_i} d(x_i,\mu_k)\times z_{i,k}] \tag{2.1}$$

$$d(x_i,\mu_k) = \sum_{g=1}^{N_g}(x_{g,i}-\mu_{g,k})^2$$

In the formulation above, i is the index for the candidates considered for typical period k, and g is the time step index within a period. $z_{i,k}$ is a binary variable equal to 1 if candidate i is assigned to cluster k.

In this paper, the k-means clustering is applied to the annual hourly energy demand data. Total annual demand is conserved by summing all cluster centroids multiplied by the corresponding number of cluster members. However, variations are smoothed out and demand

peaks are not preserved. The candidates $x_{g,i}$ are the demand data, as shown in (2.2). The weekday demand and the weekend demand of every month are each grouped into c clusters. The cluster centroids are used as the representative 24-hour, hourly resolution profiles $RD_{m,DT,c}$. Three sets of representative profiles are constructed, using one cluster ($c = 1$), two clusters ($c = 2$), or three clusters ($c = 3$), and are referred to as K_1, K_2, K_3, respectively. An example of the clustering approach is shown in Fig 2.1. The built-in MATLAB k-means function is used for testing.

$$x_{g,i} = L_{g,w,m},\ g \in (1:24),\ w \in (weekday, weekend),\ m \in (1:12) \tag{2.2}$$

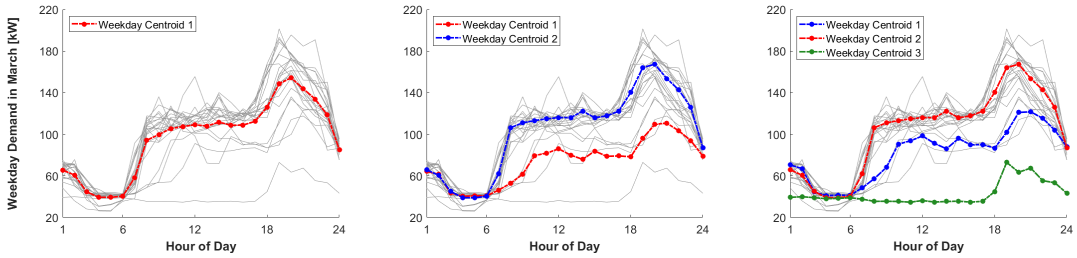

Figure 2.1: The cluster centroids selected by k-means are used to construct the representative demand profiles for a typical March weekday demand profile. From left to right are shown the centroids chosen for sets K1, K2, and K3.

2.1.3 Monthly Peak Preservation

An averaging method is specifically formulated for this paper to preserve total annual energy demand and diurnal demand peak behavior. Similarly to the application of k-means, representative weekday and weekend demand profiles are constructed by reducing the annual hourly weekday and weekend demand. However, unlike k-means, MPP explicitly preserves monthly demand peaks in a third type of representative profile, called the peak day. The annual hourly demand is separated into weekday and weekend demand for each month, and monthly demand peaks are subtracted. The remaining data is averaged to construct representative weekday and weekend demand profiles for month m and hour h, and the demand peaks populate the peak

day. The impact of demand data reduction on energy systems under tariffs that include demand charges is seen by including these peak days.

Peak Demand Profile Twelve representative hourly resolution days are created to preserve peak demand values for each month. I construct profiles representing maximum monthly demand at every hour, $RD_{m,pk,h}$. To create these peak day profiles, the maximum demand for month m and hour h over all days in the month is selected as the value for the peak demand profile. An example is shown in Fig 2.2.

$$RD_{m,pk,h} = \max_d(L_{m,h,d}) \tag{2.3}$$

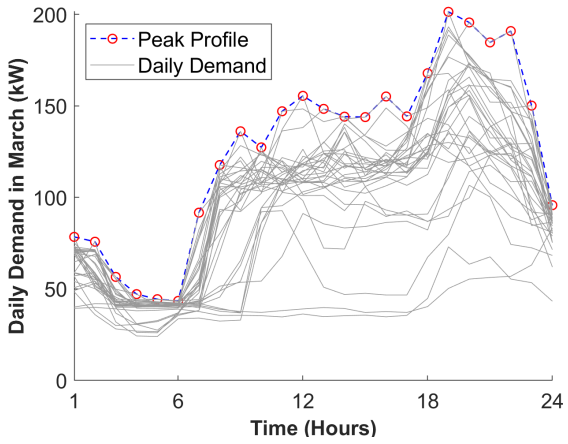

Figure 2.2: The representative energy demand profile for demand peaks in March is constructed by selecting the maximum demand of each hour across all days.

Weekday and Weekend Demand Profiles To construct the weekday and weekend demand profiles, I first separate the total annual hourly demand data into weekday demand and weekend demand for each month. I then use $RD_{m,pk,h}$ to modify the weekday and weekend data sets prior to averaging them into the representative demand profiles $RD_{m,wd,h}$ and $RD_{m,we,h}$. Demand peak values at hour h are removed from the demand data sets, based on the occurrence of the peak on a weekday or a weekend, as well as on the number of peak days represented in the month. For example, if 3 peak days are represented for month m, and the monthly peak at hour h fell on a

weekday, then $3 \times RD_{m,wd,h}$ is subtracted from the total demand summed over all weekdays for month m at hour h. This ensures that the addition of the third daytype to represent peak behavior avoids overestimating the annual energy demand, as calculated in Eq. (2.4).

$$\sum_{m=1}^{12} \sum_{DT} \sum_{h=1}^{24} n_{m,DT} RD_{m,DT,h} \; \forall DT \in (peak, weekday, weekend) \tag{2.4}$$

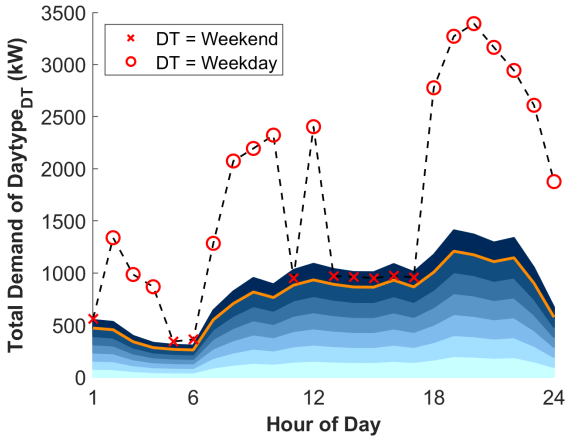

Figure 2.3: The dashed line represents the sum of either the weekday or weekend demand, depending on the occurrence of the peak. Each shaded area represents the demand peak at each hour, $RD_{m,pk,h}$. The demand peaks are stacked until a multiple of the demand first intersects the sum of the demand. The upper bound on the number of demand peaks that fit underneath the sum of the demand is marked by the solid line.

In this paper, I test the impact of varying the number of peak days by applying the load data reduction method as I iterate through a range of values for $n_{m,pk}$. I set an upper bound for how many peak days can be represented in each month by calculating the total load to peak ratio, σ_m. Fig 2.3 illustrates how I determine σ_m, ie, the number of peak days that can be subtracted when calculating the weekday and weekend load profiles. The shaded segments each represent the peak demand at each hour, and are stacked in multiples. For instance, at hour 5, the peak occurred on a weekend, so the value of the marker is the sum of the load across all weekends in March at hour 5. The solid line indicates the boundary before the stacked areas first intersect the marker line. The number of stacked areas under the boundary is the upper bound on the number

of peaks that can be subtracted across all hours for this month, and I therefore set σ_m equal to the upper bound.

Algorithm 1 Construct sets of representative weekday and weekend profiles

for iterator $m \in (1,12)$ **do**
  Construct peak profile $RD_{m,pk,h}$
  Track peak occurrence
  Calculate total load to peak ratio σ_m
  for iterator $p \in (0,\max_m(\sigma_m))$ **do**
    if $\sigma_m < p$ **then**
      $\sigma_m^p = \sigma_m$
    else
      $\sigma_m^p = p$
    end if
    Subtract $\sigma_m^p RD_{m,pk}$ from demand data
    Average remaining demand data to construct $RD^p_{m,wd}$, $RD^p_{m,we}$
    Calculate $n^p_{m,wd}$ and $n^p_{m,we}$
  end for
end for

We modify $n_{m,wd}$ and $n_{m,we}$ as I test values for $n_{m,pk}$, as shown in Eq. (2.5). The actual number of weekdays and weekends in a month (NWD_m and NWE_m) is reduced by the number of peak days tested, weighted by η_m, the ratio of maximum demand values that occurred on a weekday. This maintains the equivalence of Eq. (2.4) with the total annual energy demand, and the sum of $n_{m,wd}$, $n_{m,we}$, and $n_{m,pk}$ is equal to the actual number of days in month m.

$$n_{m,DT} = \begin{cases} \sigma_m, & DT = Peak. \\ NWD_m - \eta_m \sigma_m, & DT = Weekday. \\ NWE_m - (1-\eta_m)\sigma_m, & DT = Weekend. \end{cases} \tag{2.5}$$

$$\eta_m = \frac{\sum_{h=1}^{24} PI_{m,h}}{24}, \ \forall PI_{m,h} = \begin{cases} 1, & \text{Demand peak occurred on a weekday.} \\ 0, & \text{Else.} \end{cases} \tag{2.6}$$

Varying Number of Peaks An iteration described by Algorithm 1 imposes an additional limit on the number of peak days, ranging the upper limit from 0 peaks to $max_m(\sigma_m)$. I obtain $max_m(\sigma_m)+1$ sets of 3 representative demand profiles for each month, 36 in total, and all corresponding values of $n_{m,DT}$. The calculation of $RD^p_{m,wd}$ and $RD^p_{m,we}$, as well as $n^p_{m,DT}$, is subject to the restriction on σ_m as I vary p. I refer to the sets of demand profiles as $M0$ for the set constructed using $p=0$, $M1$ for the set constructed using $p=1$, etc.

2.2 PV Data Reduction Method

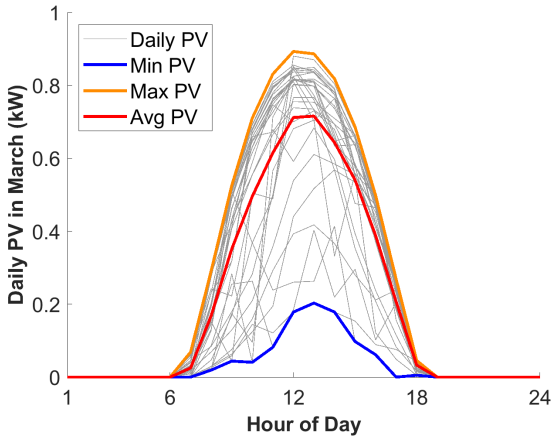

Figure 2.4: Representative PV system profiles constructed using Eqs (2.7) and (2.8). Also shown are the daily PV system performance data profiles used in constructing the representative profiles.

We create twelve representative hourly days to characterize the average monthly PV system performance. The representative profiles RPV^{avg}_m are created by averaging the daily solar irradiance $PV_{m,d}$, as shown in Eq. (2.7).

$$RPV^{avg}_{m,h} = \frac{1}{ND_m}\sum_{d=1}^{ND_m} PV_{m,d,h},\ ND_m = \text{Number of Calendar Days in } m \tag{2.7}$$

Additionally, I create a set of representative profiles to characterize the minimimum monthly PV system performance, RPV^{min}_m, and a set of representative profiles to characterize

the maximum monthly PV system performance, RPV_m^{max}. They are constructed by taking the minimum and the maximum values of the daily solar irradiance, as shown in Eq. (2.8). An example of the average, minimum, and maximum representative PV profiles is shown in Fig 2.4.

$$RPV_{m,h}^{min} = \min_d(PV_{m,d,h}) \qquad RPV_{m,h}^{max} = \max_d(PV_{m,d,h}) \tag{2.8}$$

Chapter 3

Testing Methodology

3.1 Optimization Schemes

3.1.1 8760-Timestep Optimization

A 8760-timestep optimization is developed for this paper to provide reference results for data reduction method performance. The 8760-timestep optimization is a mixed-integer linear program (MILP) based on DER-CAM [SGCM14], and minimizes the total annual costs of providing energy services to a system by optimizing the technology portfolio and operation of an energy system. The input for the MILP characterizes the system with annual hourly time series data, including building demand and solar irradiance profiles.

Decision variables are optimized for each timestep $t \in (1, 8760)$. The optimization considers all costs associated with meeting system energy demand, including (but not limited to) monthly fixed utility purchases, volumetric electricity purchases, demand charges, annualized technology investment costs, and operational costs. Cost-related decision variables are summed and minimized in the objective function, shown in simplified form in Eq. (3.1). The optimization is subject to operational constraints and an energy balance constraint.

$$\min C = \sum_{t} c_{utility} + \sum c_{invest} + \sum_{t} c_{O\&M} \tag{3.1}$$

$$s.t.$$

$$L_t = u_t + \sum_{j} g_{j,t}$$

3.1.2 Representative Days Optimization

For testing the demand data reduction methods, I alter the MILP formulation (3.1) by replacing timestep t with $m \in (1,12)$, $d \in (1, ND_m)$, and $h \in (1,24)$. The resulting optimization takes demand data input in the form of representative load profiles $RD_{m,DT,h}$ and the corresponding number of days $n_{m,DT}$, as well as the monthly average PV system performance profile described in Eq. (2.7). The MILP formulation optimizes selection and intra-day operation without connecting the representative days, which is a focus of future work. The variables characterizing monthly and annual quantities are determined by projecting out the daily variables using $n_{m,DT}$.

$$\min C = \sum_{m,d,h} c_{utility} + \sum c_{invest} + \sum_{m,d,h} c_{O\&M} \tag{3.2}$$

$$s.t.$$

$$L_{m,d,h} = u_{m,d,h} + \sum_{j} g_{j,m,d,h}$$

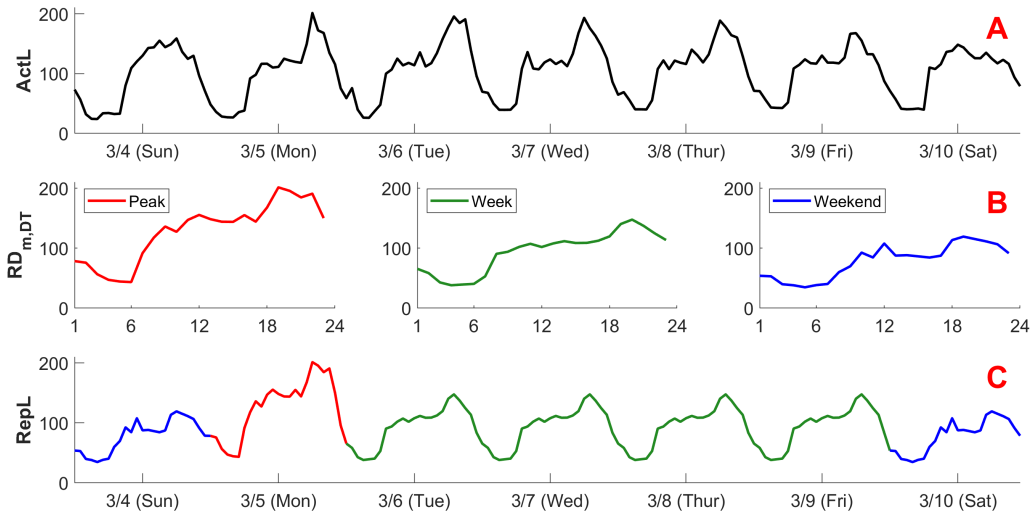

Figure 3.1: A is a segment of ActL, the actual demand data. B shows the representative demand profiles for March, constructed using a chosen method from those described in 2.1. C is the repeating annual hourly demand profile, reconstructed from the representative profile as described in 3.2.1. Both A and C are over the same time interval, March 4th through March 10th, as an example.

3.2 Testing Framework

3.2.1 Testing

We test the performance of the data reduction methods by comparing the results of the representative daytype formulation against reference results obtained from the 8760-timestep formulation. Prior to running optimizations, I prepare the input data needed for all testing. Using the actual data for system energy demand (ActL) and PV system performance (ActP), which are unmodified annual hourly time series, I construct sets of representative demand and PV profiles by applying the data reduction methods described in sections 2.1 and 2.2. Further, I create annual hourly demand and PV data time series with intra-daily variations identical to the representative demand profiles created from them, as well as total monthly consumption/solar insolation identical to that calculated by projecting out the representative profiles using $n_{m,DT}$. I do this by populating each day of the year with their representative profiles $RD_{m,DT}$ and RPV_m^{avg},

respectively. These reconstructed annual hourly time series are referred to as RepL and RepP. A visual representation is shown in Fig 3.1 and 3.2, where the top plot shows a week-long segment of the data used to construct the reduced representative profiles, which are shown in the middle plot. The bottom plot shows the representative profiles populated across the days through the same time interval as the top plot.

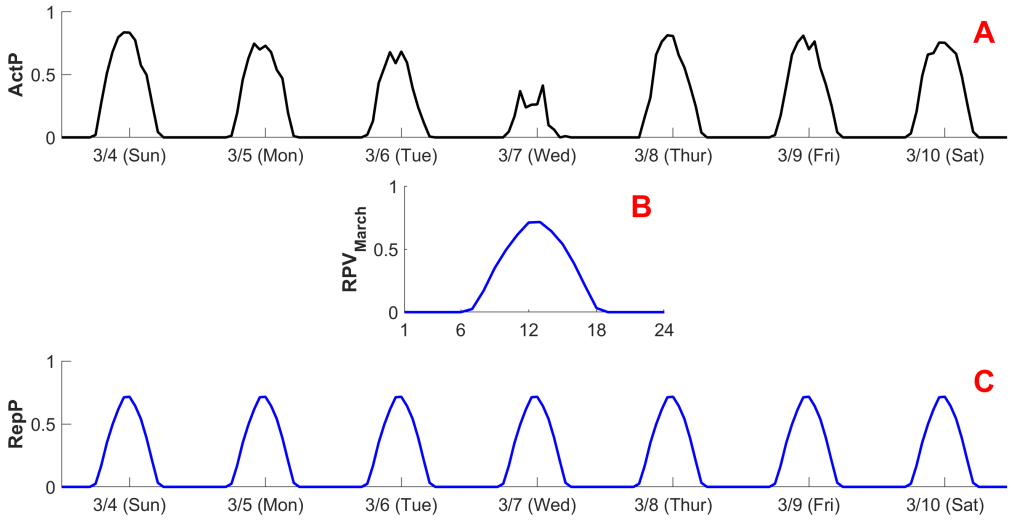

Figure 3.2: A is a segment of ActP, the actual PV system performance data. B is the representative average 24-hour PV profile for March, constructed according to Eq. (2.7). C is the repeating annual hourly PV profile, reconstructed from the representative profile as described in 3.2.1. Both A and C are over the same time interval, March 4th through March 10th.

Following input data preparation, testing proceeds as follows: An 8760-timestep optimization and a representative days optimization are run for the same energy system, with identical technology parameters. Existing system characteristics are defined by a demand profile and a PV profile chosen from the pre-prepared set of input data. The sizing and operation are each determined by the 8760-timestep and the representative days optimizations. The results of each energy system designed by the 8760-timestep and the representative days optimizations are compared using the performance metrics described in 3.2.2. Fig 3.3 provides a visualization of the overall testing procedure.

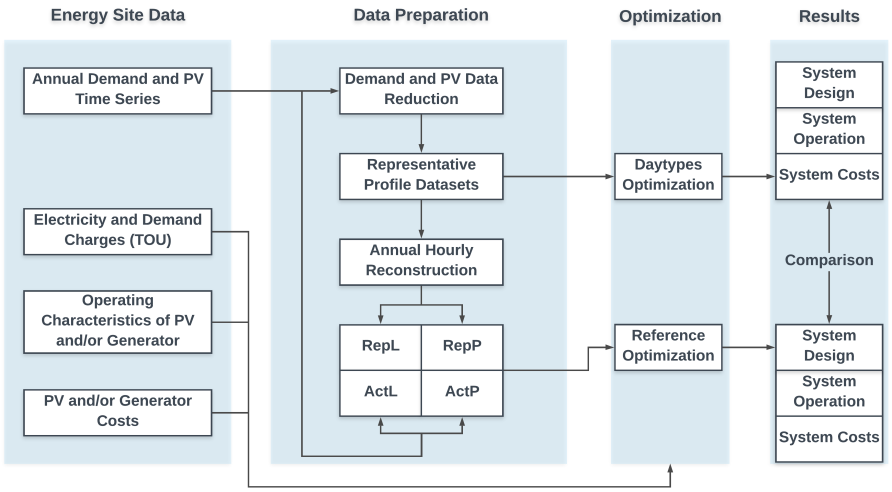

Figure 3.3: An overview of each step of the testing procedure for assessing data reduction method quality. The particular input data chosen for each optimization depends on the reduction method being tested.

To test the effects of demand data reduction, I use an 8760-timestep optimization with RepP and ActL as a reference. The purpose of using RepP is to isolate the impact of the demand data reduction on results, as seen in the comparison between the two energy systems designed by the 8760-timestep optimization and the representative days optimization. As noted earlier, the magnitude and hourly variation of RepP is identical to that of the annual PV profile that is projected out from RPV_m^{avg}, and all input data representing the energy system are completely identical between the two models, apart from the demand data input. Therefore, I can identify all discrepancies in performance metrics as being caused by the demand data reduction.

We additionally examine the impact of PV data reduction by applying a similar approach. I run the representative day optimization once, using $RD_{m,DT}$ and $n_{m,DT}$ created by a single demand data reduction method. The results of the representative day optimization are compared against those of the 8760-timestep optimization using RepL constructed from the chosen $RD_{m,DT}$ and ActP. The impact of the data reduction method being tested is again isolated: Total consumption and intra-daily variations are perfectly identical between the projection of the representative days $RD_{m,DT}$ and the annual hourly demand profile RepL. All discrepancies in performance metrics

Table 3.1: Performance metrics used in assessing data reduction method quality, using results from reference cases solved with the 8760-timestep MILP formulation.

Performance Metric	Information Gained
Objective Function	Evaluate quality of data reduction method
Technology Sizing	Impact on use as sizing tool
Demand Charges	Evaluate method of handling peak behavior
Energy Charges	Evaluate solver differences in minimizing cost

between the reference results and the representative days optimization results can be therefore attributed to the PV data reduction.

We include additional combinations of reconstructed annual hourly demand and PV system performance profiles to gain insight as to how combinations of data reduction influence the performance metrics described in 3.2.2. Each of these demand data and PV data annual hourly input data is used to run the 8760-timestep optimization, and the results are compared against both the representative days optimization and the reference results.

3.2.2 Performance Metrics

Data reduction methods performance will be evaluated based on objective function, capacity, demand charges, and energy charges, shown in Table 3.1. The primary metric for data reduction method quality is the objective function. The discrepancies in demand charges, energy charges, and sizing for the representative day optimization results and the 8760-timestep optimization results will be useful in understanding solver differences. The discrepancy between the representative day optimization results and the 8760-timestep optimization results will be reported.

3.3 Energy System and Case Study Description

A building on the UCSD campus in La Jolla, California is chosen for testing. The building is a gym that is open through all days of the week, and is subject to closures during holidays and school breaks. The demand data is historical 15-minute resolution real power data from the gym. Only electricity demand is modeled, and a snapshot of actual demand data recorded at the gym is shown in Fig 3.4, as well as the load duration curve for the annual demand. The annual hourly PV system performance data is the AC power output per kW of system capacity, calculated based on location, the choice of module and array type, and the tilt, among other parameters.

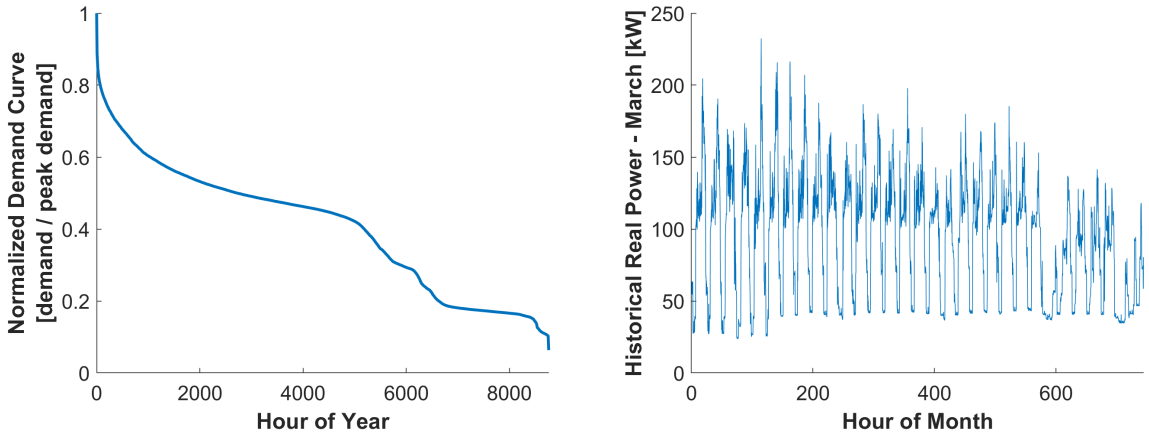

Figure 3.4: The load duration curve for Main Gym in 2018 (left) and raw historical real power data recorded for Main Gym in March 2018 (right). The building experiences frequent variations in demand, with extreme fluctuations during vacations and periods of high use.

The building is in the San Diego Gas and Electric service area. The tariff structure selected is Schedule AL-TOU Secondary, which applies to non-residential customers with a monthly demand exceeding 20 kW. The tariff includes summer and winter energy charges and demand charges for on-peak, semi-peak, and off-peak periods. The demand charge is applied to the maximum demand hourly demand for a given month. The on-peak period is from 4 pm to 9 pm for all days of the year.

In Case 1, the choice for generation technology is a photovoltaic (PV) system. In Case 2,

the choice includes both PV and a natural gas generator. The optimizer selects and sizes the PV capacity on a continuous basis, and selects and sizes the number of generator units on a discrete basis. The PV system modeled is a fixed ground-mounted array with an approximate nominal efficiency of 19%. PV system installation cost is set to $1700 per kW of capacity, and operation and maintenance costs are set to $1.4167 per kW of capacity per month. The natural gas generator modeled is based on the Generac SG100, and has a power rating of 100 kW. Installation costs are set at $200,000 per generator, with variable operation and maintenance costs of $0.02 per kWh of output energy. Lifetime of both technologies is set to 30 years.

Chapter 4

Results

4.1 Total Annual Demand

The first set of results I look at is the error in calculating the total annual consumption. Shown in Fig 4.1 is the percent error of the calculated annual demand against the actual annual demand, $\sum_{t=1}^{8760} L_t$. All methods estimate the total annual demand to within 0.7% accuracy. K1, K2, K3, and M0 simply calculate mean representative demand profiles without including any peak days, and therefore show zero error in total annual consumption.

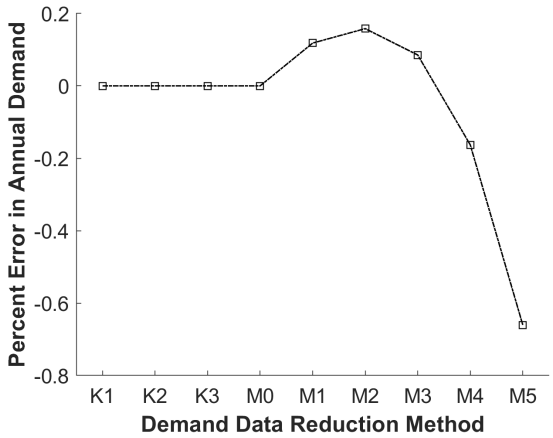

Figure 4.1: The percent error for total annual demand as calculated by projecting out $RD_{m,DT}$ with $n^p_{m,DT}$, as described in Eq. (2.4).

The error in total consumption in M1 through M5 is a consequence of the demand data reduction method. Demand peaks are subtracted from weekend and weekday data sets and used to construct peak profiles on an hour-by-hour basis, but daily profiles are projected out on a daily basis by a single value per profile in the optimization. As I begin to increase the number of peak days represented in each month, the calculated annual demand is first overestimated, then changes to underestimated when I increase p from 3 to 4. The error increases with the number of peak days represented, as it increases the demand that is subtracted from the weekend and weekday demand data prior to constructing $RD_{m,wd}$ and $RD_{m,we}$ and more significantly reduces the representative weekday and weekend consumption.

4.2 Influence of Demand Data Reduction

4.2.1 Case 1

The objective function error of the optimizations using reduced demand data is shown in Fig 4.2. The data reduction methods that perform the best in terms of objective function error are M1 through M5, which are the methods that include at least one peak day profile for each month. These methods all reach an objective function value that deviates from the actual objective function by no more than 0.3%. The demand data reduction methods that do not consider representative profiles for demand peaks show much more significant objective function errors.

The inclusion of an artificially constructed peak demand profile plays a much greater role in accurately representing the system demand than the number of separate profiles used to represent typical weekdays and weekends. The error is greatest for K1 and M0, which only use two representative demand profiles for each month, one weekday profile and one weekend profile. Adding more representative profiles for each month, four in K2 and six in K3, captures more variation in demand, and therefore, reaches a more accurate objective function. However, without

demand peaks captured in a peak day profile, three typical weekday and three typical weekend profiles per month can only reduce the objective function error to 4.485%.

In contrast, the addition of at least one peak demand profile per month is sufficient to drop the objective function error to 0.2213%. The degree to which the objective function error varies as the number of peak days changes is minimal.

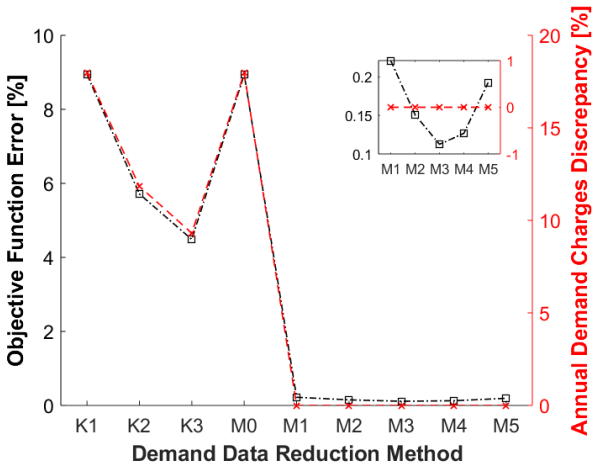

Figure 4.2: Case 1: The objective function error and the total annual demand charge discrepancy are shown for all demand data reduction methods. A clear correlation exists between objective function error and demand charge discrepancy.

The significant impact that the inclusion of a peak profile has on the results is seen by examining the annual demand charge discrepancy for each of the demand data reduction methods. The variation of the demand charge discrepancy from method to method is nearly identical to the variation in objective function error. This correlation between objective function error and demand charge discrepancy over various demand data reduction methods leads us to conclude that the demand charge discrepancy significantly contributes to the objective function error. The methods that have the greatest discrepancy in demand charge are the ones that do not have peak day profiles. The degree to which monthly demand peaks are captured accurately drives the degree to which the objective function deviates from the actual objective function. The significant contribution of demand charges to the total annual cost is caused by the timing of the monthly demand peaks, the timing of the on-peak TOU period, and the absence of dispatchable generation.

The times at which peak demand occurs does not coincide with significant PV production hours, so there is no opportunity for demand charge reduction, and therefore total annual costs are underestimated to the same degree by which maximum monthly demand, and therefore total annual utility demand charges, are underestimated by a given demand data reduction method. This highlights the value of using a peak demand profile in accurately capturing total annual costs. The impact of the way and the extent to which extremes in demand are captured is most immediately seen in the demand charge discrepancy, which highlights the need to test data reduction methods in the context of TOU tariffs that include demand charges.

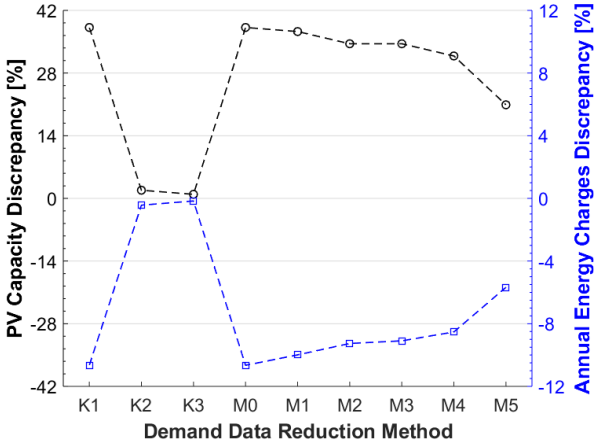

Figure 4.3: Tradeoff between reducing annualized investment costs of installing additional PV vs reducing utility purchases.

The discrepancy in PV sizing and energy charges, shown in Fig 4.3, provides insight into how the solver minimizes the total annual cost for this energy system. The sizing deviates significantly from the capacity chosen by the 8760-timestep optimization using ActL and RepP. Each of the reduced representative profile optimizations oversize the PV system. The significant variations in sizing do not correspond to the variations seen in the objective function error. The M1-M5 methods recommend between 20.91% to 37.27% additional PV compared to the reference case. These methods had an objective function error of less than 0.3%, and 0% discrepancy in annual demand charges, and by examining the degree to which the annual energy charges are

underestimated for each case, and how the underestimation varies from method to method, I can see that the solver is choosing to minimize cost by investing in more PV and reducing energy charges for the daytypes optimizations, whereas for the 8760-timestep optimization, it is choosing to minimize cost by selecting less PV capacity and reducing annualized investment costs rather than reducing energy charges.

Additionally, I can surmise that the averaging process increases the overlap between the demand profile and the averaged PV system performance profile, and that there is therefore additional benefit to investing in more PV for a system in the representative days optimizations, which have smoothed, more consistent demand profiles. The overlap increases opportunities for energy charge reduction via PV power output. The sizing variations are unsurprising in the context of this system, which has no dispatchable generation. The second case tested includes a dispatchable technology, and therefore addresses the sizing fluctuations.

4.2.2 Case 2

As in Case 1, the demand data reduction methods that consider peak days easily outperform those that do not, and the inclusion of an artificially constructed peak demand profile more significantly reduces objective function error than increasing the number of representative profiles or increasing p. The objective function error and the demand charge discrepancy of the demand data reduction methods tested are both plotted in Fig 4.4. The same trends for both the objective function error and the demand charge discrepancy observed in the results for Case 1 are seen in the results for Case 2, and methods M1-M5 again have the smallest objective function error.

We again see that fewer representative profiles are necessary to reduce the objective function error simply by ensuring that the reduction preserves demand peaks. Adding more representative profiles in K2 and K3 reduces the objective function error as more demand variation is captured as c increases. However, K3 uses six representative profiles per month for typical weekdays and weekends, yet has a greater error than M1, which only uses three

representative profiles per month. K3 has an objective function error of 1.83%. Increasing the value of p from 1 to 4 does improve the objective function error, but these improvements are minimal compared to the improvement seen when adding a peak profile, as shown by going from M0 to M1. Similar patterns are seen in the demand charge discrepancy results, although the correlation between the objective function error and the demand charge discrepancy is not as strong here as in Case 1. This is due to the presence of a dispatchable technology, which allows the solver to find more solutions to minimize cost than in Case 1.

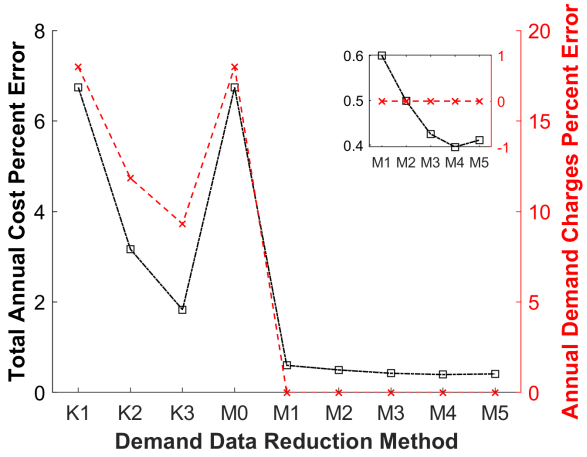

Figure 4.4: Case 2: The objective function error and the total annual demand charge discrepancy are shown for all demand data reduction methods. Demand charge discrepancy is correlated with objective function error.

We continue to observe a significant impact of including a peak demand profile $RD_{m,pk}$ by examining the sizing results for Case 2, shown in Fig 4.5. Again, there is a clear distinction between the demand data reduction methods that include a peak demand profile and the methods that do not. M1 through M5 select the same number of generators as the reference case, whereas K1, K2, K3, and M0 all select one less generator. The PV sizing discrepancy is smoothed out by the additional of dispatchable generation, particularly for M1 through M5, which have a PV sizing discrepancy of no more than 2.273%.

The results for generator sizing are a direct consequence of the demand peaks represented in the energy system when modeling a choice of dispatchable energy as a means to reduce

demand charges during evening peaks. By failing to capture the maximum demand values, the representative days optimizations using either k-means or M0 underestimate the value of demand charge reduction that would be provided by a second generator. This further highlights the significance of evaluating demand data reduction methods in the context of energy systems subject to varying TOU rates and demand charges.

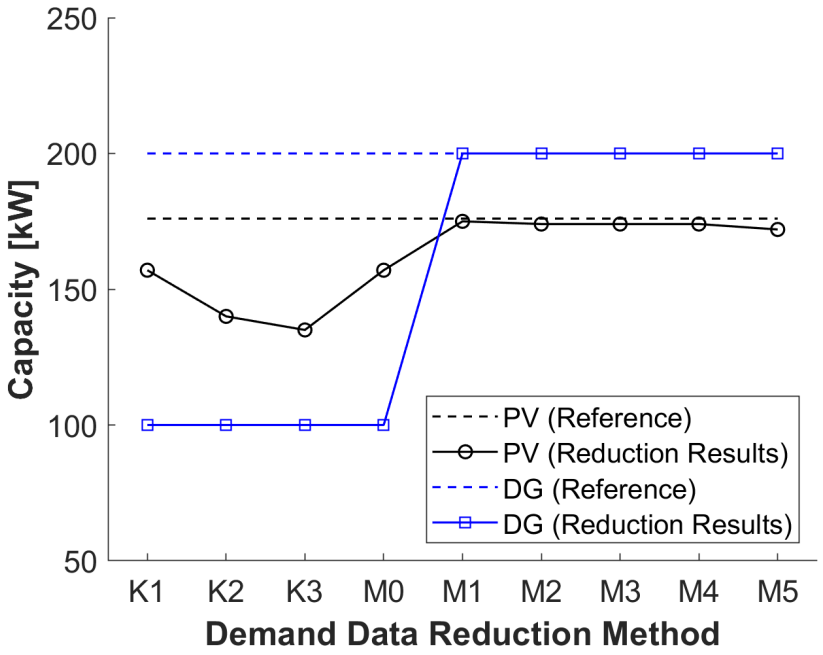

Figure 4.5: Case 2: PV and generator capacity selected for all demand data reduction methods. Also shown are the PV and generator capacity selected for the reference results.

4.3 Influence of PV Data Reduction

Fig. 4.6 shows the results of testing three different methods of PV system performance data reduction testing, applied to Case 2. The impact of PV system performance data reduction is isolated by using (RepL,ActP) as a reference case to compare the results of the representative profile optimization. M3 was chosen as the representative profile set for constructing RepL.

We first assess the isolated impact of PV system performance data reduction methods

by examining the objective function results for M3 and (RepL,ActP). Of the three methods tested, averaging the PV system performance data (RPV_m^{avg}) has the best performance, with an objective function error of 0.85042%, compared to a more significant underestimation of total annual cost by 6.0821% using maximized PV system performance data (RPV_m^{max}). These cost underestimations are accompanied by overestimations in PV sizing. The minimization PV data reduction method causes costs to be overestimated by 3.9196%, and no PV is selected in the solution.

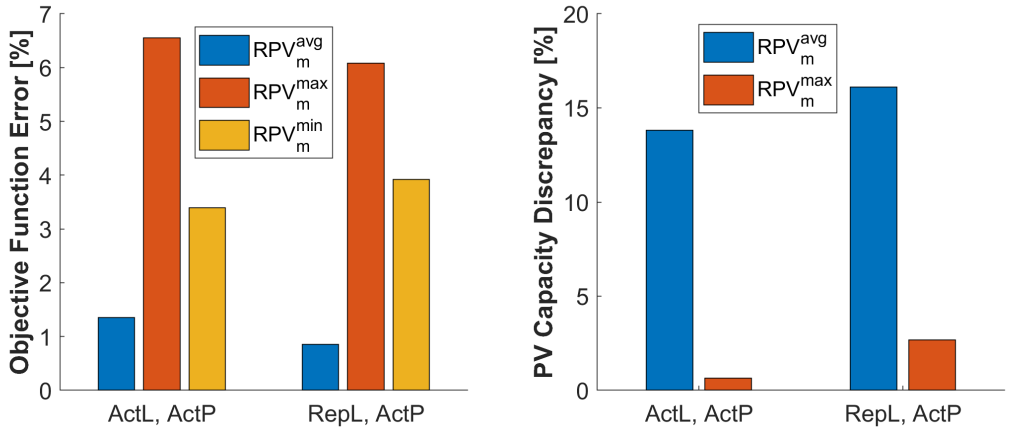

Figure 4.6: Objective function error and PV capacity discrepancy of the tested PV system performance data reduction methods, with respect to each of the 8760-timestep optimizations run with different data combinations. No PV was selected using RPV_m^{min}.

For Case 2, reducing PV system performance data has a greater impact on both the objective function and the sizing results than reducing demand data. As seen by examining the impact of reducing both demand and PV system performance data, the objective function error increases over the comparisons that only test either demand data reduction or PV system performance data reduction, but is still under 2% for the averaged PV system performance data reduction.

4.4 Run times

Using representative demand profiles reduces the run time by over 90%. There is no meaningful, consistent difference that would motivate a choice of one of the tested methods solely on the basis of computational speed. Therefore, I again identify M1 through M5 as the methods with best performance, losing less than 1% accuracy in exchange for a significant gain in run time savings.

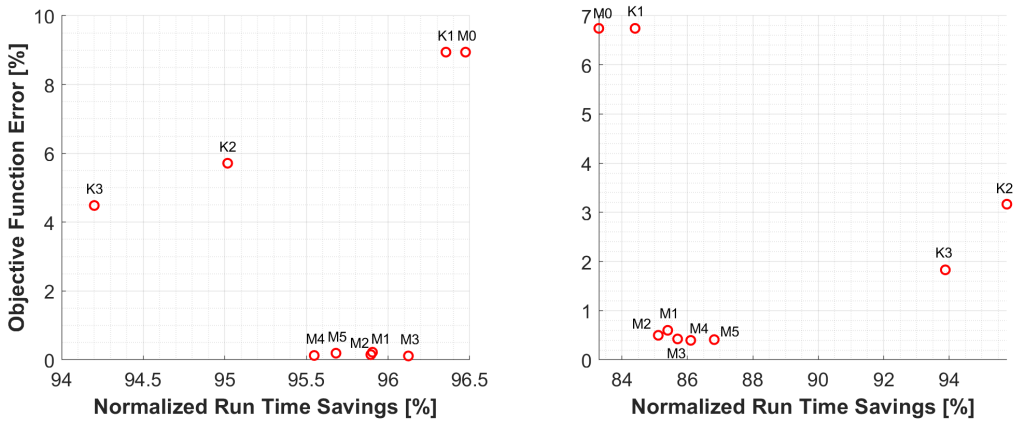

Figure 4.7: The run time savings of each demand data reduction method, normalized by the run time of the 8760-timestep optimization, plotted against the objective function error for Case 1 (left figure) and Case 2 (right figure).

Chapter 5

Conclusion

This paper assesses the quality of various demand and PV system performance data reduction methods for use in an energy system design optimization problem. An in-house approach to averaging demand data while preserving demand peaks is presented and compared against existing clustering techniques. The testing methodology is designed to isolate the effects of reducing time series data, providing new insight into the impact of using representative demand profiles and the comparative effects of implementing various data reduction methods. Data reduction method performance is evaluated on the basis of objective function error from reference results, as well as discrepancies in demand charges, technology sizing, and energy charges.

Results indicate that methods which include a peak demand profile outperform those that do not, thereby demonstrating the importance of accounting for extreme demand behavior and validating the presented method of preserving peak demand values. The method introduced in this paper reduces annual hourly demand data to 36 representative 24-hour demand profiles, using one profile per month to preserve peak demand, and two profiles per month to capture average weekday and weekend demand. For the case of a grid-connected building subject to TOU rates and demand charges, this method has less than 1% objective function error for the energy system modeled. Comparatively, methods which use 48 or 72 representative profiles to capture weekday

and weekend demand without preserving demand peaks show an error between 1.83% and 5.7%. The number of peak days per month tested also had relatively little effect on the objective function error, compared to the effect of including vs removing peak demand profiles entirely.

Objective function error shows strong correlation with demand charge discrepancy in the cases studied, further demonstrating the effect of including peak demand profiles in demand data reduction. DER sizing discrepancy is also clearly correlated with objective function error for the case that includes options for dispatchable generation. Methods that did not preserve peak demand values consistently undersized both generator and PV capacity, which is unsurprising, given that systems subject to TOU rates and demand charges especially benefit from demand charge reduction through DER deployment, and therefore, accurately modeling energy systems for such systems especially benefit from the inclusion of peak demand profiles.

Further work is necessary to continue to develop understanding on how data reduction influences energy system design. Recent works have modeled systems with both electrical and thermal loads as well as a wide selection of generation and storage options. However, research on data reduction methods in the application of energy system optimization has thus far focused on the impact of data reduction as a whole, without identifying the error contributions of each time series reduction. I clearly show that isolating different inputs has specific effects, and I recommend that future research apply a similar approach in analyzing individual aspects of using representative profile MILP formulations. Additionally, the MILP formulation does not model storage dispatch across representative days, which may have significant consequences for scenarios in which that is important. A follow-up paper will address this challenge and present an evaluation of the impact of modeling uncoupled representative days on storage dispatch and sizing.

Appendix A

Final notes

A.1 Additional Equations for Monthly Peak Preservation Method

Total Load to Peak Ratio:

$$MPSF_m = \min(\text{floor}(\frac{PI_{m,h}\sum_{d=1}^{NWD_m} WDL_{m,h}(d) + (1-PI_{m,h})\sum_{d=1}^{NWE_m} WEL_{m,h}(d)}{PL_{m,h}})) \quad \text{(A.1)}$$

Explicit Formula for Representative Weekday Profile Calculation

$$WDP_{m,h} = \frac{\sum_{d=1}^{NWD_m} WDL_{m,h}(d) - PSF_m PL_{m,h} PI_{m,h}}{NWD_m - PSF_m PI_{m,h}} \quad \text{(A.2)}$$

Explicit Formula for Representative Weekend Profile Calculation

$$WEP_{m,h} = \frac{\sum_{d=1}^{NWE_m} WEL_{m,h}(d) - PSF_m PL_{m,h}(1-PI_{m,h})}{NWE_m - PSF_m(1-PI_{m,h})} \quad \text{(A.3)}$$

This thesis, in full, is currently being prepared for submission for publication of the material. Fahy, Kelsey; Stadler, Michael; Pecenak, Zachary; Haghi, Hamed Valizadeh; Kleissl, Jan. The thesis author was the primary investigator and author of this material.

Bibliography

[BKS+17] B. Bahl, A. Kümpel, H. Seele, M. Lampe, and A. Bardow. Time-series aggregation for synthesis problems by bounding error in the objective function. *Energy*, 2017:900–912, 2017.

[CPR09] M. Casisi, P. Pinamonti, and M. Reini. Optimal lay-out and operation of combined heat power (chp) distributed generation systems. *Energy*, 34(12):2175–2183, 2009.

[DMCLCAGS11] F. Domínguez-Muñoz, J.M. Cejudo-López, A. Carrillo-Andrés, and M. Gallardo-Salazar. Selection of typical demand days for chp optimization. *Energy and Buildings*, 43:3036–3043, 2011.

[FBM+14] S. Fazlollahi, S.L. Bungener, P. Mandel, G. Becker, and F. Maréchal. Multi-objectives, multi-period optimization of district energy systems: I. selection of typical operating periods. *Computers Chemical Engineering*, 65:54–66, 2014.

[GFMM19] P. Gabrielli, F. Fürer, G. Mavromatidis, and M. Mazzotti. Robust and optimal design of multi-energy systems with seasonal storage through uncertainty analysis. *Applied Energy*, 238:1192–1210, 2019.

[GGMM18] P. Gabrielli, M. Gazzani, E. Martelli, and M. Mazzotti. Optimal design of multi-energy systems with seasonal storage. *Applied Energy*, 219:408–424, 2018.

[GSV14] R. Green, I. Staffell, and N. Vasilakos. Divide and conquer? k-means clustering of demand data allows rapid and accurate simulations of the british electricity system. *IEEE Transactions on Engineering Management*, 61(2):251–260, 2014.

[Jai10] A.K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[KMRS18] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten. Impact of different time series aggregation methods on optimal energy system design. *Renewable Energy*, 117:474–487, 2018.

[LRCS09] M.A. Lozano, J.C. Ramos, M. Carvalho, and L.M. Serra. Structure optimization of energy supply systems in tertiary sector buildings. *Energy and Buildings*, 41(10):1063–1075, 2009.

[LRS10] M.A. Lozano, J.C. Ramos, and L.M. Serra. Cost optimization of the design of chcp (combined heat, cooling and power) systems under legal constraints. *Energy*, 35(2):794–805, 2010.

[MDFG08] G. Mavrotas, D. Diakoulaki, K. Florios, and P. Georgiou. A mathematical programming framework for energy planning in services' sector buildings under uncertainty in load demand: The case of a hospital in athens. *Energy Policy*, 36(7):2415–2429, 2008.

[MGRM17] S. Moradi, R. Ghaffarpourb, A.M. Ranjbar, and B. Mozaffari. Optimal integrated sizing and planning of hubs with midsize/large chp units considering reliability of supply. *Energy Conversion and Management*, 148:974–992, 2017.

[MSMP13] E. Mehleri, H. Sarimveis, N. Markatos, and L. Papageorgiou. Optimal design and operation of distributed energy systems: Application to greek residential sector. *Renewable Energy*, 51:331–342, 2013.

[Pfe17] S. Pfenninger. Dealing with multiple decades of hourly wind and pv time series in energy models: a comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Applied Energy*, 197:1–13, 2017.

[SGCM14] M. Stadler, M. Groissböck, G. Cardoso, and C. Marnay. Optimizing distributed energy resources and building retrofits with the strategic der-camodel. *Applied Energy*, 132:557–567, 2014.

[SSF+18] T Schütz, M.H Schraven, M. Fuchs, P. Remmen, and D. Müller. Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis. *Renewable Energy*, 129:570–582, 2018.

[SSH+17] T. Schütz, L. Schiffer, H. Harb, M. Fuchs, and D. Müller. Optimal design of energy conversion units and envelopes for residential building retrofits using a comprehensive milp model. *Applied Energy*, 185:1–15, 2017.

[STSM] J. Schiefelbein, J. Tesfaegzi, R. Streblow, and D. Müller. Design of an optimization algorithm for the distribution of thermal energy systems and local heating networks within a city district.

[TB19] H. Teichgraeber and A.R. Brandt. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. *Applied Energy*, 239:1283–1293, 2019.

[WKY16] T. Wakui, H. Kawayoshi, and R. Yokoyama. Optimal structural design of residential power and heat supply devices in consideration of operational and capital recovery constraints. *Applied Energy*, 163:118–133, 2016.

[WY14] T. Wakui and R. Yokoyama. Optimal structural design of residential cogeneration systems in consideration of their operating restrictions. *Energy*, 64:719–733, 2014.

[WY15] T. Wakui and R. Yokoyama. Optimal structural design of residential cogeneration systems with battery based on improved solution method for mixed-integer linear programming,. *Energy*, 84:106–120, 2015.

[YHI02] R. Yokoyama, Y. Hasegawa, and K. Ito. A milp decomposition approach to large scale optimization in structural design of energy supply systems. *Energy Conversion and Management*, 43(6):771–790, 2002.

[ZLZC17] Q. Zhu, X. Luo, B. Zhang, and Y. Chen. Mathematical modelling and optimization of a large-scale combined cooling, heat, and power system that incorporates unit changeover and time-of-use electricity price. *Energy Conversion and Management*, 133:385–398, 2017.