Title
US Representative Feeder Sets for Distribution Grid Economics and Policy Applications

Journal
IEEE Transactions on Energy Markets Policy and Regulation, 4(1)

ISSN
2771-9626

Authors
Seres, Albane
Heleno, Miguel

Publication Date
2026-03-01

DOI
10.1109/tempr.2025.3626480

US Representative Feeder Sets for Distribution Grid Economics and Policy Applications

Albane Seres, *Member, IEEE*, and Miguel Heleno, *Senior Member, IEEE*

***Abstract*—This paper proposes a methodology to generate sets of distribution feeders representative of any U.S. county for economic and policy studies. The methodology modifies prototypical feeders and organizes them into sets that (i) reproduce historical economic and infrastructure investment patterns and (ii) accommodate county-specific demand characteristics, including energy, peak, and building types. The approach relies on publicly available data, enabling systematic application across all U.S. counties and adaptation to other regions. The use of the resulting feeders is demonstrated through an electrification impact study in Alameda County, California.**

***Index Terms*—Distribution grid economics, distribution infrastructure, electrification.**

NOMENCLATURE

The mathematical symbols used throughout this paper are classified below as follows.

Sets

$\mathcal{F}_t$	Set of taxonomy feeders representative of the climate zone.
$\mathcal{F}_{cz}$	Set of climate-zone representative feeders with modified assets.
$\mathcal{F}_{county}$	Set of county-specific feeders with modified assets and county building stock.
WFS_{county}	Set of weighted feeders representative of the county.
$\mathcal{C}$	Utility catalog of capacitor, lines and transformers.
$\mathcal{B}$	Set of branches.
$\mathcal{B}_{viol}$	Set of branch violations.
$\mathcal{N}_{viol}$	Set of nodes with undervoltages.
$\mathcal{M}_f^T$	Multiset of building types for feeder f.
$\mathcal{B}_f$	Set of buildings associated with feeder f.
$\mathcal{N}_f$	Set of end-nodes for feeder f.
$\mathcal{L}_f$	Set of loads for feeder f.
T	Set of building types.
P	Set of growth rates.

Parameters

$Uc_f(\bar{\rho})$	Upgrade cost of feeder f for a homothetic load growth $\bar{\rho}$.
Uc_f^{baseline}	Baseline upgrade cost of feeder f.
α	Weighting factor for county energy, peak load, and historical investment residual.
β	Weighting factor for building-type energy residual.
γ	Reduction factor for baseline upgrade cost.
δ	Threshold for iterative optimization.
P_f	Total active power of the feeder f.
P_n	Active power of a feeder end-node n.
$L_b(t)$	Load profile for building b.
$L_f(t)$	Aggregated load profile of buildings in feeder f.
E_f	Energy of the buildings in feeder f.
E_f^τ	Energy of building type τ in feeder f.
E_{county}	Total energy of the county.
L_{county}	Peak load of the county.

Decision variables

$x_{b,n}$	Binary decision variable for building b assigned to feeder end-node n.
ϵ_f^n	Residual error for building to end-node load assignment at t^{peak}.
δ_f^n	Max residual error for building to end-node load assignment.
n_f	Optimal weighting factor for feeder f.
r_{county}^E	Residual error for county energy consumption.
r_{county}^L	Residual error for county peak load.
r_ρ^{Uc}	Residual error for historical investment for a load increase of ρ.
r_τ^E	Residual error for building type τ energy consumption.

I. INTRODUCTION

IN THE United States, the electrification of all end-use sectors could increase annual electricity consumption by up to 44% from 2016 to 2050 under high adoption scenarios, that assume rapid uptake of electric vehicles and electric heating technologies [1]. The new electric demand has the potential to change load shapes and increase the risk of grids congestions [2], [3], which will require significant upgrades of the distribution grid infrastructure and increase the overall costs of the electricity sector. Thus, to develop efficient electrification policies, it is necessary to systematically estimate these future infrastructure

Received 9 April 2025; revised 30 August 2025; accepted 20 October 2025. Date of publication 30 October 2025; date of current version 17 March 2026. This work was supported by the U.S. Department of Energy under Contract DE-AC02-05CH11231. Paper no. TEMPR-00090-2025. *(Corresponding author: Albane Seres.)*

The authors are with Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA (e-mail: seresalbane@gmail.com; miguelheleno@lbl.gov).

Digital Object Identifier 10.1109/TEMPR.2025.3626480

costs, particularly in what concerns to distributions grid upgrades, which nowadays already represent the highest share of electric utilities' capital expenditures [4].

Traditionally, distribution grid upgrade costs are estimated empirically, by analyzing historical utility investment data and extrapolating these trends to predict future costs [2], [5]. This approach involves estimating costs linearly from peak load using incremental reference values of investment rates, expressed in $/kW, derived from historical utility upgrades and peak demand gathered from regulatory data sources, such as the Federal Energy Regulatory Commission (FERC) Form 1 [6]. However, projecting future costs based on the past capital expenditures implicitly assumes that utilities will continue investing in traditional "poles and wires" infrastructure and that equipment cost trends will remain the same.

Alternatively, forward-looking methodologies to assess distribution upgrades involve projecting future load scenarios for individual feeders and calculating future capital costs. For example, a California study in [7] assessed the impact of behind-the-meter PV integration on distribution grid costs, using a set of 75 distribution feeders. Another recent California study evaluated how electric vehicle (EV) penetration and net energy metering tariffs affect distribution grid infrastructure [8]. This study analyzed all individual feeders of the three main utilities by projecting EV scenarios and calculating the capacity upgrade costs for each. A work in Switzerland [9] estimated distribution upgrade costs associated with the deployment of solar, EV and space heating, considering a sample of real networks serving 170 thousand consumers. A common characteristic among these studies is the use of real feeder data, which is often proprietary and difficult to obtain. In addition, using real feeders in large-scale national studies presents a trade-off: modeling every feeder is impractical, while relying on feeder samples may raise concerns about representativeness.

Real distribution network data can be represented by prototypical feeders, specifically developed to reproduce electric design characteristics that are found in real networks [10]. This is the case, for example, of the IEEE PES Test Feeders [11], which consist of 11 radial distribution networks, or the low voltage (LV) feeders developed by Cigre to benchmark the integration of distributed generation [12]. Comprehensive sets of these feeders can be developed to represent an entire utility territory or geographical region. For example, in Italy the ATLANTIDE Project created a digital archive with three MV reference networks representing rural, industrial, and urban areas [13], while in Sweden a similar project established two MV reference networks for reliability studies [14]. In the US, a variety of works over the last decades have used clustering techniques to analyze real feeders and derive representative networks of a particular utility territory [15] or a state [16]. At regional or country level, 25 taxonomy feeders representative of specific climate zones were developed, using hierarchical agglomerative clustering of real distribution networks across the nation [17]. Similar clustering techniques have also been applied in Australia to derive taxonomy feeders, including the consideration of LV networks [18] and the integration of distributed solar [19].

A different approach to approximating realistic distribution networks is the generation of synthetic feeders over street maps. This is the case with the Reference Network Model (RNM), a tool used to produce substation and feeder designs while considering geolocation, electricity consumption, and a realistic distribution equipment catalog [20]. RNM was used to derive the SMART-DS dataset, which contains thousands of feeders across four U.S. regions [21]. Besides the synthetic representation of electrical parameters — from LV through MV and up to sub-transmission — the dataset was enhanced with individual building-level demand and PV profiles [22]. Unlike the test and taxonomy feeders [11], [17], these synthetic networks do not reproduce the real characteristics and electric designs of a particular region. However, they have the advantage of being geolocated and offering detailed demand profiles.

Despite their use in various distribution operation studies, both taxonomy and synthetic geolocated networks are inadequate for distribution grid economics and policy studies. First, although they rely on realistic electrical installations, they do not intend to reproduce infrastructure or economic costs of a particular utility or location. Second, despite including realistic peak loads and sometimes load profiles, they do not mean to capture the entire energy demand of a service territory, nor they are available for every single location. Therefore, designing electrification policies and projecting future distribution grid costs requires a new set of feeders that can reproduce real distribution infrastructure costs and techno-economic characteristics of a particular location.

To bridge this gap, this paper proposes a methodology to generate sets of representative feeders in the US, with county-level granularity, for the purpose of distribution economic and policy studies. For a given county, we show that our methodology is able to generate a population of representative feeders that:

- i) contains the demand characteristics of the county (including energy, peak and building types);
- ii) reproduces the economics and the distribution infrastructure investment patterns historically observed in the area;
- iii) is based on publicly available data and, therefore, can be systematically applied to every single county of the US territory.

We illustrate the methodology and the use of the datasets by applying it to an electrification impact study in Alameda County, CA, US.

II. Methodology

To generate feeder sets that reproduce the economics of local distribution grids, we depart from prototypical feeders representative of the electric installations of a given location. In this work, we use the US taxonomy feeders [17], but the methodology can be applied to any representative feeder set. Without changing the topology, the proposed methodology modifies the electric equipment rating in those feeders so that they reproduce the rate of investments historically observed in the utilities of a particular climate zone. Additionally, it creates different instances of feeders with electric load profiles that are

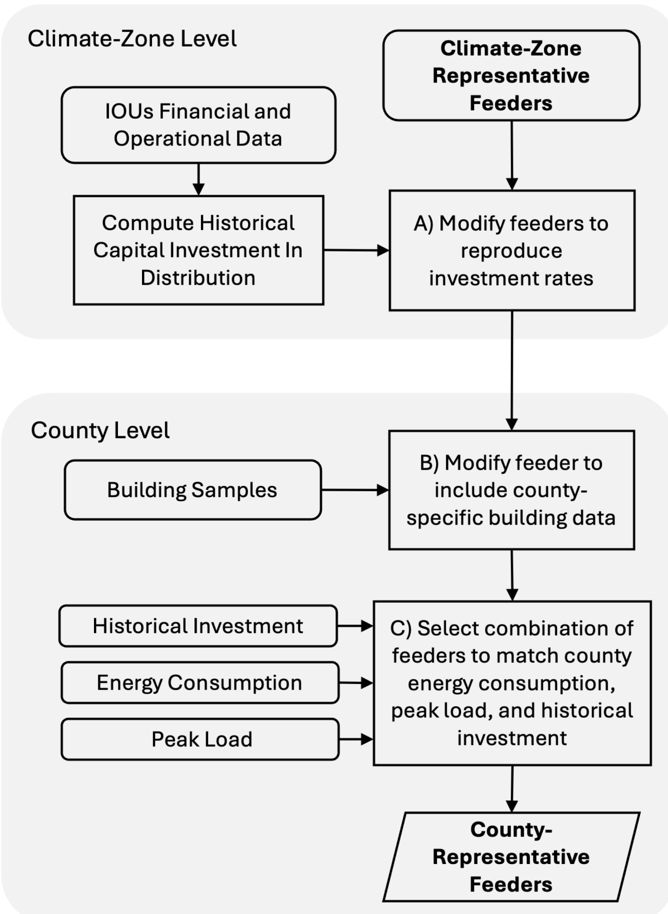


Fig. 1. Methodology workflow.

characteristic of the building stock of a given county. Finally, the methodology ensures that the resulting representative set of feeders can reproduce both the historical rate of investments as well as the total energy, peak, and building types of each county. The overall methodology workflow is presented in Fig. 1.

A. Modifying Feeders to Reproduce Investment Rates

This step of the methodology focuses on modifying the equipment of taxonomy feeders, which are representative of Building America climate zones, to ensure their investment rates align with the historical capital investment rate of utilities operating within those climate zones. The modifications include adjustments to line and transformer capacities as well as the addition of capacitors. The targeted rate of investment can be derived from the historical data on local utilities' capital expenditures and associated demand growth, using the methodology outlined in [23]. Financial and operational data from major US investor-owned utilities (IOUs) are collected and a multivariate regression model with fixed effects was used:

$$\mathrm{Capex}_{kW,i,t} = \beta_0 + \beta_{\mathrm{G}}\mathrm{G}_{i,t} + \mathrm{Y}_t + \mathrm{U}_i + \epsilon_{i,t} \quad (1)$$

where $\mathrm{Capex}_{kW,i,t}$ represents the capital expenditure per kW for utility i in year t; $\mathrm{G}_{i,t}$ is the growth rate of system capacity for

Algorithm 1: Investment-Adjusted Representative Feeders.

Input: Taxonomy Feeder Set $\mathcal{F}_t$, reference load increase $\bar{\rho}$, utility catalog of capacitors, lines and transformers $\mathcal{C}$.
Output: Investment-adjusted feeders per climate zone $\mathcal{F}_{cz}$.
for all $f \in \mathcal{F}_t$ **do**
  for a target load increase $\bar{\rho}$, get feeder overloading and costs to upgrade it using **Algorithm 2**: $Uc_f(\bar{\rho}), \mathcal{N}_c, \mathcal{B}_{viol}, \leftarrow$ Algorithm 2$(f, \bar{\rho}, \mathcal{C})$
  calculate the difference between historical and observed upgrade costs: $\Delta c_f \leftarrow Uc^{Hist}(\bar{\rho}) - Uc_f(\bar{\rho})$
  for $i = 1$ to N_{var} **do**
    create $\hat{f}_i$, a copy of feeder f
    randomly add capacitors to nodes $\mathcal{N}_c$ of $\hat{f}_i$
    if $\Delta c_f < 0$ **then**
      randomly upgrade branches $\mathcal{B}_{viol}$ in $\hat{f}_i$ using $\mathcal{C}$.
    else
      randomly downgrade branches $\mathcal{B} \setminus \mathcal{B}_{viol}$, in $\hat{f}_i$ using $\mathcal{C}$.
    end if
  end for
  $Uc_{\hat{f}_i}(\bar{\rho}) \leftarrow$ Algorithm 2$(\hat{f}_i, \bar{\rho}, \mathcal{C})$
end for
return $\mathcal{F}_{cz} = \{(\hat{f}_i, Uc_{\hat{f}_i}(\bar{\rho})) | i \in 1, \ldots, N_{\mathrm{var}}\}$

utility i in year t; β_{G} represents the change in per-kW costs in response to a one percentage point increase in load growth; Y_t is a fixed effect for each year, U_i is a fixed effect for each utility, and $\epsilon_{i,t}$ is the error term. The intercept $\beta_0 + \overline{\mathrm{Y}_t} + \overline{\mathrm{U}_i}$ represents the average per-kW distribution investment for the case of no growth, denoted by I_0.This model isolates the effect of growth rate on capital expenditure per kW while accounting for the fixed effects of capital expenditures per utility and year that are not attributed to load growth (e.g. reliability enhancements). The total cost per new kilowatt of capacity, $R(\rho)$, including both the base cost and the incremental cost due to load growth, is derived for a growth rate ρ of the initial capacity C_i as follows:

$$R(\rho) = \frac{\overbrace{((I_0 + 100 \times \beta_{\mathrm{G}} \times \rho) \times C_i \times (1+\rho))}^{\text{Investment after load growth}} - \overbrace{(I_0 \times C_i)}^{\text{Initial investment}}}{\underbrace{C_i \times \rho}_{\text{New kW of capacity}}} \quad (2)$$

$$R(\rho) = I_0 + 100 \times \beta_{\mathrm{G}} \times (1+\rho) \quad (3)$$

Given the lack of coherent publicly available data and the presence of multiple utilities within a single county, historical investment records are aggregated by climate zone and analyzed separately. Thus, for each climate zone, historical β_{G}, I_0 is obtained. For a reference demand increase $\bar{\rho}$, it is possible to calculate the rate of utility upgrade investments $Uc^{Hist}(\bar{\rho})$ observed. This value expresses the trend of distribution utility investments in \$per-kW of new peak load in a particular climate-zone.

Algorithm 1 modifies the equipment in the taxonomy feeders $\mathcal{F}_t$ so that they match the historical trend $Uc^{Hist}(\bar{\rho})$. The algorithm applies a load growth $\bar{\rho}$ to the feeder f and estimates the upgrade costs $Uc_f(\bar{\rho})$ required to keep the feeder within the voltage and loading limits, using Algorithm 2. The heuristic is designed to capture exclusively the capital investment costs associated with physical upgrades needed to accommodate load growth. Operational expenditures (O&M), administrative adders, and other utility overheads are intentionally excluded, consistent with the derived historical $per-kW of new peak load, which reflects only capital investments directly attributable to increased load. This feeder cost is then compared to the historical investment empirically observed in the area $Uc^{Hist}(\bar{\rho})$. If a feeder's upgrade cost significantly exceeds historical rate of investments, the capacity of its overloading lines and transformers is increased. Thus, with these new capacities, the feeder would need less upgrades when a load growth $\bar{\rho}$ is seen, and their investment needs approximate $Uc^{Hist}(\bar{\rho})$. Conversely, when $Uc_f(\bar{\rho})$ is lower than the historical, lines and transformer ratings are decreased, in order to create overloading and consequent investments in the presence of load growth $\bar{\rho}$. Finally, in this process of adjusting the investment needs to match the historical, it is possible to simulate N_{var} variations of the same feeder, by randomly adding capacitors to the nodes and randomizing the lines and transformers selected to increase or decrease their rating. This approach results in a new set of investment-adjusted feeders for each climate zone $\mathcal{F}_{cz}$, with the same topology and design of the original taxonomy feeders, $\mathcal{F}_t$, but with different equipment loading levels, which would make feeder investment needs approximate the historical $Uc_f(\bar{\rho})$.

Algorithm 2: Heuristic to Obtain Overloading and Upgrade Cost.

Input: Feeder f, reference load increase $\bar{\rho}$, utility catalog of capacitors, lines and transformers $\mathcal{C}$.
Output: Total upgrade cost $U^f_{cost}(\bar{\rho})$, Set of candidate nodes for capacitor additions $\mathcal{N}_c$, Set of branches with violation $\mathcal{B}_{viol}$.

```
Calculate load increase: l ← l × (1 + ρ̄)   for l ∈ L_f
Run power flow (PF) to obtain undervoltages and branch
  violations: N_viol, B_viol ← PF(f, L_f)
for b ∈ B_viol do
  b^up ← arg min_{bc∈C_bc} {S(bc) > S(b)}
  Uc_f(ρ̄) ← Uc_f(ρ̄) + Cost(b^up)
end for
Z_v ← voltage zone clusters from N_viol
while N_viol ≠ ∅ do
  for z ∈ Z_v do
    n ← arg min_{n∈z} {V_pu(n)}
    c_n^up ← arg min_{n∈N_c} {Cap(c_n)}
  end for
  N_viol ← PF(f, L_f, N_c)
end while
Uc_f(ρ̄) ← Uc_f(ρ̄) + Σ_{n∈N_c} Cost(c_n^up)
return Uc_f(ρ̄), N_c, B_viol
```

S represents the apparent power rating, V_{pu} represents the voltage per-unit, and Cap represents the capacity of the component.

B. Modifying Feeder to Include County-Specific Building Data

Here we present a method to further extend the set of feeders $\mathcal{F}_{cz}$, tailoring them to a particular county, $\mathcal{F}_{county}$. We achieve this by associating to each node of the feeder set a group of building load profiles representative of the county's building stock.

It is important to note that each taxonomy feeder is already associated with pre-established representation of building types [17], including specific configurations of residential and commercial consumers. To create a maximum variety of feeders with different building-type representations, while ensuring that they remain realistic, the representative share of each building type w^τ_f in the feeder f is slightly modified as $\hat{w}^\tau_f \leftarrow w^\tau_f \cdot (1 + u^\tau_f)$, where u^τ_f is a random perturbation drawn from a uniform distribution.

Considering the building stock representative of a particular county, from datasets such as the NREL developed ResStock and ComStock [24], a weighted multiset of building types $\mathcal{M}^\mathcal{T}_f$ is constructed for each feeder, where the proportion of each building type τ is determined by the weighting factor $\hat{w}^\tau_f$. The sampling process involves two steps: first, a building type is randomly selected from the multiset, and second, a specific building profile is randomly sampled from the chosen building type. This process is repeated until the cumulative peak load of the selected building profiles matches the original feeder's peak load P_f. The result is a set of building profiles for each feeder ($\mathcal{B}_f$), in which the representativeness of building types ($\hat{w}^\tau_f$) and the aggregated peak load are aligned with the taxonomy feeder characteristics.

Next, it is necessary to distribute the building profiles among the nodes of the feeder, while respecting the original peak load of each node. This step can be formulated as an optimization model that assigns each building to a node $x_{b,n}$ and minimizes the overall nodal assignment error of the resulting load. The first component of this error, ϵ^n_f, is the relative difference between the node's active load and the aggregated power of the assigned buildings at the time of the feeder peak t^{peak}, given by:

$$t^{\text{peak}} = \max_t \left\{ \sum_{b \in \mathcal{B}_f} L_b(t) \right\}$$

where $L_b(t)$ represents the load of building b at time t, and $\mathcal{B}_f$ is the set of buildings associated with the feeder. The second component of the assignment error, δ^n_f, is the relative difference between the total peak load of the buildings connected to a node n and the original active power of that node. Thus, these two error components ensure that the buildings profiles selected for each node approximate, as much as possible, the original load both at the time of the local and system peaks.

Thus, given each feeder f representative of a climate zone $\mathcal{F}_{cz}$, we run the optimization model (4)–(8) to assign building profiles characteristic of the county to obtain $\mathcal{F}_{county}$.

$$\underset{x_{b,n},\epsilon^n_f,\delta^n_f}{\text{Minimize}} \quad \sum_n (\epsilon^n_f + \delta^n_f) \tag{4}$$

s.t.:

$$\left| \frac{P_n - \sum_b x_{b,n} \cdot L_b(t = t^{\text{peak}})}{P_n} \right| \leq \epsilon_f^n, \quad \forall n \in \mathcal{N}_f \tag{5}$$

$$\frac{\max_t \sum_b x_{b,n} \cdot L_b(t) - P_n}{P_n} \leq \delta_f^n, \quad \forall n \in \mathcal{N}_f \tag{6}$$

$$\epsilon_f^n \geq 0, \quad \forall n \in \mathcal{N}_f \tag{7}$$

$$\sum_n x_{b,n} = 1, \quad \forall b \in \mathcal{B}_f \tag{8}$$

It is important to note that, although this model minimizes the discrepancies between original feeder and the new profile loads, it is still possible that the building assignment introduces new overloads and, consequently, the need for new upgrades, which could change the assumptions discussed in Section II-A. Thus, we quantify these additional baseline upgrade costs Uc_f^{baseline} using the Algorithm 2 and will consider them later.

C. Feeder Population Representative of a County

At this point, we have a set $\mathcal{F}_{county}$ of taxonomy feeders that approximate utility's historical investments rates (from Section II-A) and have their nodes associated with building profiles sampled from the county's building stock (from Section II-B). However, the entire feeder set $\mathcal{F}_{county}$ does not match the load of the county nor does it reproduces the total county investment costs.

This section presents an optimization model to assign a factor to each taxonomy feeder in $\mathcal{F}_{county}$, denoted as n_f, that describes the number of feeders of type f presented in a given county. The objective is to ensure that the resulting population of feeders matches the county's aggregated energy consumption and peak load. Furthermore, the aggregated historical investment of the feeders should also correspond to the historical investment of the climate zone in which the county is located. To achieve this, we formulate the problem as an iterative Mixed-Integer Linear Programming (MILP) model, presented in (9)–(15).

$$\underset{n_f, r^E_{\text{county}}, r^L_{\text{county}}, r^{Uc}_\rho, r^E_\tau}{\text{Minimize}} \quad \alpha \cdot \left(r^E_{\text{county}} + r^L_{\text{county}} + \frac{1}{|P|} \sum_{\rho \in P} r^{Uc}_\rho \right) + \beta \cdot \sum_{\tau \in T} \left(r^E_\tau \cdot \frac{E^\tau_{\text{county}}}{E_{\text{county}}} \right) \tag{9}$$

$$\left| \frac{\sum_{f \in \mathcal{F}_{\text{county}}} n_f \cdot E_f - E_{\text{county}}}{E_{\text{county}}} \right| \leq r^E_{\text{county}} \tag{10}$$

$$\left| \frac{\sum_{f \in \mathcal{F}_{\text{county}}} n_f \cdot L_f(t^{\text{peak}}) - L_{\text{county}}(t^{\text{peak}})}{L_{\text{county}}(t^{\text{peak}})} \right| \leq r^L_{\text{county}} \tag{11}$$

$$\left| \frac{\sum_{f \in \mathcal{F}_{\text{county}}} n_f \cdot Uc_f(\rho) - Uc^{Hist}(\rho)}{Uc^{Hist}(\rho)} \right| \leq r^{Uc}_\rho, \quad \forall \rho \in P \tag{12}$$

$$\left| \frac{\sum_{f \in \mathcal{F}_{\text{county}}} n_f \cdot E^\tau_f - E^\tau_{\text{county}}}{E^\tau_{\text{county}}} \right| \leq r^E_\tau, \quad \forall \tau \in T \tag{13}$$

$$\sum_{f \in \mathcal{F}_{\text{county}}} n_f \cdot Uc_f^{\text{baseline}} \leq Uc_{\text{previous}}^{\text{baseline}} \cdot (1 - \gamma), \quad \text{if } i \neq 0 \tag{14}$$

$$r^E_{\text{county}}, \quad r^E_\tau, \quad r^L_{\text{county}}, \quad r^{Uc}_\rho \quad \geq 0 \tag{15}$$

The objective function (9) minimizes the mismatch of energy (r^E_{county}), peak load (r^E_{county}) and upgrade costs (r^U_ρ) between the representative feeder population and the actual observed values in each county. These are the primary errors to be minimized and therefore given weight of α. A secondary objective, with a weight of β is to reduce, as much as possible, the errors associated with the energy per building type τ. The constraints (10)–(13) define these errors as a function of the weight (n_f) of each feeder in the population.

The optimization model employs an iterative approach to minimize the total baseline upgrade cost while ensuring that the deviation from the optimal objective value does not exceed a predefined threshold δ. The reduction factor γ in the constraint (14) dictates the percentage by which the baseline upgrade cost must decrease in each iteration i. The term $Uc_{\text{previous}}^{\text{baseline}}$ correspond to the total baseline cost at the previous iteration and is calculated as:

$$Uc_{\text{previous}}^{\text{baseline}} = \sum_{f \in \mathcal{F}_{\text{county}}} n_{f,\text{previous}} \cdot Uc_f^{\text{baseline}}$$

where $n_{f,\text{previous}}$ represents the results from the previous iteration.

As a result of the optimization model, a weight n_f is defined for each feeder in $\mathcal{F}_{county}$. The resulting Weighted Feeder Set WFS_{county} is defined as the subset of $\mathcal{F}_{county}$ containing the feeders with a non-zero weight n_f.

$$WFS_{county} = \{f \in \mathcal{F}_{county} | n_f \neq 0\}$$

III. Results

This section presents three key results from the study: First, it details the data sources considered. Second, it introduces the developed Weighted Feeder Set for Alameda County, CA ($WFS_{Alameda}$), which can be accessed in [25], and provides a comparison with actual county data to assess its representativeness. Finally, it outlines the total distribution costs for Alameda County in the context of building electrification scenario, calculated using $WFS_{Alameda}$.

A. Data Sources

1) Climate Zone Specific Taxonomy Feeders: A study by PNNL [17] developed a taxonomy of 24 prototypical radial distribution feeders for the U.S., designed to reflect key attributes of distribution systems across five major climate regions. These feeders, categorized into three voltage classes (12.47 kV, 25.00 kV, 35.00 kV), incorporate climate-driven heating and cooling demand patterns that influence feeder design. While radial feeders represent typical rural systems, they exclude urban topologies (e.g., networked grids in large cities). In the PNNL taxonomy, the Marine climate zone lacks large urban feeder models; however, we address this limitation by including the IEEE 342-Low Voltage Network Test System (LVNTS) [26], a

TABLE I
TAXONOMY FEEDERS FOR THE MARINE CLIMATE ZONE ('R1' STANDS FOR 'REGION 1' AND 'GC' FOR 'GENERAL CATEGORY')

Feeder	kV	kVA	Description
R1-12.47-1	12.5	7152	Moderate suburban/rural
R1-12.47-2	12.47	2836	Moderate suburban/light rural
R1-12.47-3	12.47	1362	Small urban center
R1-12.47-4	12.47	5334	Heavy suburban
R1-25.00-1	24.9	2105	Light rural
GC-12.47-1	12.47	5200	Single large commercial
LVNTS	13.2	42210	Heavy urban and meshed

TABLE II
REGRESSION COEFFICIENTS, R^2 VALUES, AND COST PER NEW KW FOR DIFFERENT GROWTH RATES IN EACH CLIMATE ZONE

Climate Zone	I_0	β_{Growth}	R^2	R [\$/new kW]	
				$\rho = 5\%$	$\rho = 10\%$
Cold	10.44	0.24	0.70	35.64	36.84
Hot-Dry	26.34	1.72	0.87	206.94	215.54
Hot-Humid	8.69	0.31	0.66	41.24	42.79
Marine	12.64	0.44	0.83	58.84	61.04
Mixed-Humid	7.28	0.32	0.62	40.88	42.48

highly meshed, unbalanced 120/208 V urban network with eight 13.2 kV primary feeders supplying a 50 MVA load.

This study focuses on capturing the economic costs associated with feeder infrastructure rather than replicating their physical characteristics or topologies. Table I gives a summary of the feeders associated with the Marine climate zone.

2) Financial and Operational Data From IOUs: Financial and operational data are sourced from FERC Form 1, covering 99 major U.S. IOUs. The dataset includes historical cost and capacity data from 2000 to 2007, a period of significant electricity sales growth [23]. The analysis assumes non-IOUs (e.g., municipal utilities, co-ops) follow regional IOU investment trends, though this introduces potential bias due to exclusion of public utility data. Within this framework, $CapEx_{kW,i,t}$ represents capital investment in conductors and transformers per kilowatt for utility i in year t. Utilities are classified by climate zone, and the rate R—the additional investment per new kilowatt of capacity—is computed for each zone, as detailed in Table II.

The investment rate results have been corroborated with external data, specifically for Alameda County, CA. The E3 report for the California Public Utilities Commission [27] lists a long-term distribution capacity marginal cost of 57.3 \$/kW for PG&E, which serves this specific county. This figure is consistent with the calculated rate of 58.84 \$/kW (corresponding to a 5% growth rate) for the Marine climate zone, confirming alignment with real-world utility data for this region.

3) Distribution Assets Costs: The Distribution Grid Integration Unit Cost Database [28], published by NREL, contains a set of unit distribution grid upgrades, including traditional forms of reconductoring and transformer capacity upgrades for different locations in the US. The database includes costs of transformers, cables, lines, voltage regulators, capacitor banks, storage assets, among others.

4) County-Specific Building Samples: The ResStock and ComStock datasets, developed by NREL, provide representative building samples and associated load profiles for each county. ResStock focuses on residential buildings, detailing characteristics like home type (e.g., single-family, multi-family) and equipment, while ComStock models commercial buildings, covering activity types (e.g., office, restaurant) and technologies. Industrial buildings are excluded from ComStock's scope and were not included in this study. Additionally, both datasets do not account non-building loads such as EVs and distributed energy resources (DERs), which may require consideration in future analyses. Both datasets use the Actual Meteorological Year (AMY) 2018 for energy modeling, with the latest releases being ResStock 2022 Release 1 and ComStock 2023 Release 2.

5) County-Specific Energy Consumption and Peak Load: The ResStock (residential) and ComStock (commercial) datasets provide aggregated end-use load profiles by sector and building type (e.g., single-family homes, offices). Total county-level electricity consumption is calculated by summing these profiles across all timesteps, while peak load is identified as the maximum load value. These profiles also enable granular calculations of consumption and peak load for individual building types within each sector.

B. Representative Feeder Set of Alameda County, California

Alameda County, California, was selected as a case study due to its location in the Marine climate zone, with representative feeders from this zone forming the initial set $\mathcal{F}_t$. Feeder R1-25.00-1 was excluded from the analysis because industrial loads accounted for over 96% of its total load, which falls outside the scope of this study focused on residential and commercial building electrification. Historical data on investment rates, building samples, energy consumption, and peak load for Alameda County were collected from the data sources mentioned earlier.

This section provides an overview of the characteristics of the generated sets: $\mathcal{F}_{cz}$, $\mathcal{F}_{county}$, and $WFS_{Alameda}$, as described in Section II. It also compares $WFS_{Alameda}$ with actual county data to assess the accuracy of the proposed methodology in representing the energy consumption and economic features of Alameda County.

1) Obtaining Feeders That Reproduce Investment Rates Observed in the Climate Zone $\mathcal{F}_{cz}$: As described in Section II, we produced around 100 variations of each feeder in the climate zone taxonomy set $\mathcal{F}_t$, by modifying their assets to achieve investment rates aligned with the ones historically observed. It is important to note that the objective is not to exactly match that rate for every single feeder, but to create a diversity of feeders with investment rates around the historical. Fig. 2 illustrates the distribution of investment rates for the modified feeders in $\mathcal{F}_{cz}$, grouped by their corresponding initial feeder taxonomy in $\mathcal{F}_t$.

The box plots in Fig. 8 demonstrate that modified feeder investment rates adjust toward the historical rate for the climate zone (blue dashed line). When the initial feeder's rate (green cross) exceeds the historical rate, modified feeders are generally generated with lower rates; and they are adjusted upward when the initial rate falls below. While deviations can occur due to

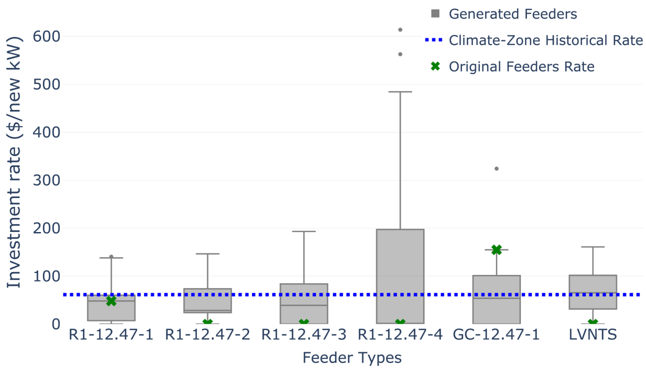


Fig. 2. Boxplot of modified feeders' investment rates ($/kW), categorized by the initial feeder $\mathcal{F}_t$.

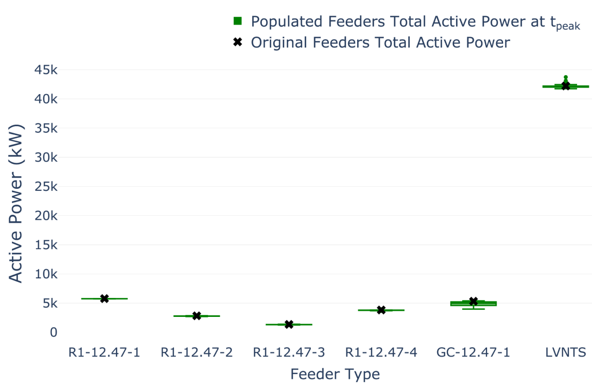


Fig. 3. Distribution of populated feeders' peak loads compared to the peak loads of their corresponding initial feeders in $\mathcal{F}_t$.

TABLE III
STATISTICS SUMMARY FOR POPULATED FEEDERS

Feeder	N_f	AFLA	ANLA (%)		
		(%)	Res.	Comm.	All
R1-12.47-1	81	99.6	95.4	82.0	95.0
R1-12.47-2	78	98.7	97.6	58.7	94.6
R1-12.47-3	79	97.5	98.6	68.9	69.7
R1-12.47-4	90	98.9	99.9	84.6	92.5
GC-12.47-1	90	91.9	—	91.7	91.7
LVNTS	82	99.5	96.44	71.6	88.5

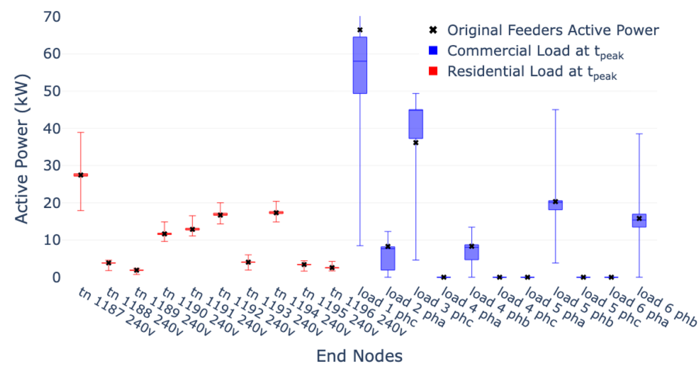


Fig. 4. Comparison of the distribution of assigned loads (across 98 populated feeders) and node active power capacities, for a subset of feeder R1-12.47-1 nodes.

randomness in the feeder generation process, these adjustments ensure closer alignment with historical investment patterns. Each box plot has different ranges, indicating that the diversity of investment rates that can be generated is specific to the topology of the initial feeder.

2) Formation of County-Specific Feeder Set $\mathcal{F}_{county}$: Each feeder in $\mathcal{F}_{cz}$ is populated with a set of buildings, denoted as $\mathcal{B}_f$, such that the aggregated peak load of these buildings matches the feeder's peak load. This process results in the formation of the set $\mathcal{F}_{county}$. Fig. 3 presents the distributions of the populated feeders' peak loads compared to the feeder's peak load for each initial feeder in $\mathcal{F}_t$.The Average Feeder Load Accuracy (AFLA) is calculated to quantify the accuracy of this matching. This metric represents the average, across all populated feeders of the same type, of the relative difference between the total active power of the feeder P_f and the aggregated load of all buildings at the time the feeder experiences its peak. It can be expressed as:

$$\text{AFLA } (\%) = \frac{1}{N_f} \sum_{i=1}^{N_f} \left| \frac{\sum_{b \in \mathcal{B}_{f,i}} L_b(t^{peak})}{P_f} \right| \tag{16}$$

The results presented in Table III indicate that the AFLA values are consistently high across all feeder types, with most values exceeding 97% . This demonstrates that for each feeder f in $\mathcal{F}_{county}$, the aggregated peak load of the selected set of buildings $\mathcal{B}_f$ effectively matches the feeder's total peak load.

The buildings in $\mathcal{B}_f$ are assigned to specific nodes of the feeder through the optimization process described in Section II-B). Fig. 4 shows the distribution of the buildings' loads at the time of the peak assigned to a node versus the given active power of the node P_n across a subset of the nodes of the initial feeder R1-12.47-1, as an example. This visualization helps to compare the actual loads assigned to each node with the maximum available active power. To assess the accuracy of the building-to-node assignment methodology across all populated feeders, the Average Node Load Accuracy (ANLA) can be calculated. This metric represents the average node matching accuracy across all nodes and all populated feeders of the same type. The ANLA can be calculated separately for residential nodes, commercial nodes, or all nodes combined. It can be expressed as:

$$\text{ANLA } (\%) = \frac{1}{N_n} \sum_{n=1}^{N_n} \overbrace{\left(\frac{1}{N_f} \sum_{f=1}^{N_f} \left(1 - \epsilon_f^n\right) \times 100 \right)}^{\text{Average accuracy for node n}}$$

where N_f is the number of populated feeders of the same type, N_n is the number of nodes considered in each feeder (residential only, commercial only, or both), and ϵ_f^n is the relative assignment error for the n-the node in the f-th feeder, calculated as in Section II-B.

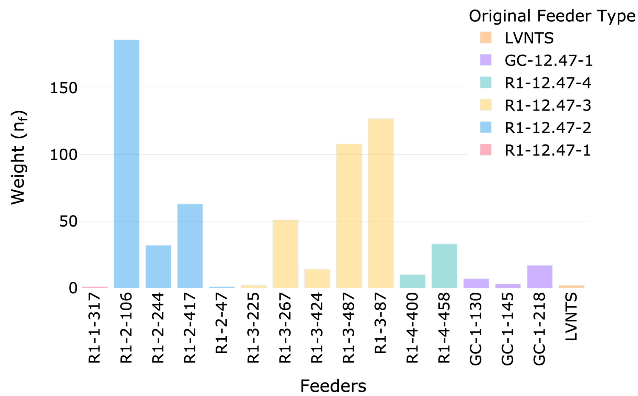


Fig. 5. Feeder weights in $WFS_{Alameda}$.

The ANLA values in Table III indicate high accuracy for residential nodes across all feeders, with values consistently exceeding 95% . This high accuracy reflects the effectiveness of the building-to-node assignment methodology for residential buildings, due to its relatively low power. In contrast, the ANLA values for commercial nodes are notably lower, particularly for feeders R1-12.47-2, R1-12.47-3 and LVNTS, due to the magnitude and heterogeneity of loads in commercial buildings, which makes them harder to match with distribution nodes. Overall, the ANLA calculated for all nodes combined remains high, with values exceeding 91% for most feeders. This high accuracy is primarily due to the predominance of residential nodes in most feeders.

3) Selection of Optimal Weighted Feeders Set (WFS_{county}): The optimization model outlined in Section II-C is used to select a subset of feeders from the generated set $\mathcal{F}_{county}$ by assigning corresponding weights to the feeders in the pool determined above. The primary objective of this selection process is to minimize the discrepancy between county-level characteristics and those represented by the feeders in WFS_{county}. The characteristics considered included total energy consumption, peak load, historical rate of investment, and, with lower priority, energy consumption per building type. The weights of the optimization function are therefore assigned as follows: $\alpha = 1$, and $\beta = 10^{-2}$. The model is run iteratively, reducing the baseline upgrade cost by half between each iteration by setting $\gamma = 0.5$, until the objective value differs by more than 10% ($\delta = 10\%$) from the previous iteration.

The final $WFS_{Alameda}$ comprises 16 selected feeders. Fig. 5 shows their weights and origin taxonomy feeder types. There is notable variability in the weights assigned to each feeder. For instance, the WFS includes one R1-12.47-1 feeder (count 1), four R1-12.47-2 feeders (counts 1–32-63-186), and a mix of R1-12.47-3 (counts 2–14-51-108-127) and R1-12.47-4 (counts 10-33) feeders. The inclusion of two LVNTS feeders (count 2) provides a networked urban topology, capturing characteristics absent in radial feeders. Overall, the optimization effectively combines feeders of different sizes and topologies to form a WFS that reflects Alameda County's energy and economic constraints.

TABLE IV
METRICS COMPARISON: $WFS_{Alameda}$ VS. ACTUAL ALAMEDA COUNTY DATA, WITH INVESTMENT RATE FOR A 10% LOAD INCREASE

Metrics	**WFS**	**County**	**Rel. Error [%]**
Total Energy [GWh]	7,582	7,582	<0.01
Total Peak Load [MW]	1,287	1,287	<0.01
Investment Rate [$/kW]	61.28	61.04	0.23

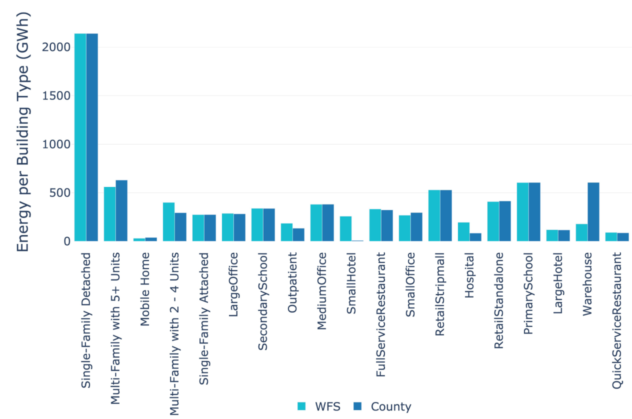


Fig. 6. Energy consumption per building type: $WFS_{Alameda}$ vs. Actual Alameda County Data.

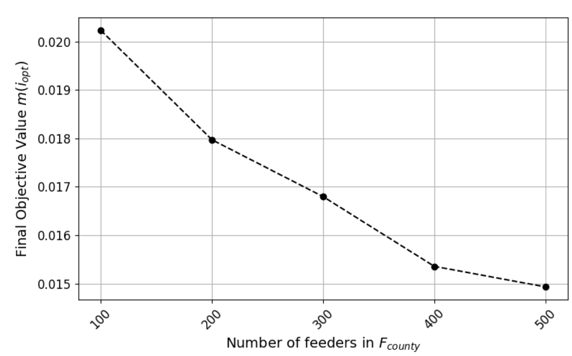


Fig. 7. Objective value vs. Number of feeders in $\mathcal{F}_{county}$.

The results presented in Table IV show that $WFS_{Alameda}$ aligns closely with Alameda County's total energy consumption, total peak load, and investment rate, with relative errors below 0.3% . Fig. 6 also shows that, beyond these aggregate results, the $WFS_{Alameda}$ can also replicate the energy of the individual buildings types of each county. However, discrepancies arise for "Warehouse," "Hospital," and "Small-Hotel" building types, which have high individual loads but are sparsely represented in the county. These characteristics makes the final match more dependent on the original pool of buildings of this type. We note that both the peak load and total energy used to evaluate the WFS exclude industrial energy, which is not represented in the ResStock or ComStock datasets. This limitation should be considered when interpreting the WFS performance, particularly in counties with significant industrial demand.

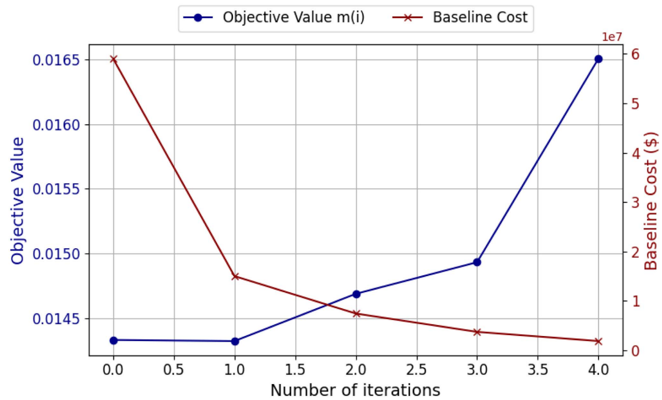


Fig. 8. Objective value and baseline upgrade cost vs. optimization model iterations.

Fig. 7 illustrates how the number of feeder variations generated in $\mathcal{F}_{county}$ influences the accuracy of the final selection in WFS. More feeders variations in $\mathcal{F}_{county}$ means more diversity in terms of loads, building types and capital costs, which allows to optimization model to find combinations of feeder weights that can more accurately represent the county. However, for very large sizes of $\mathcal{F}_{county}$ (more than 300 feeders), the added value of new feeders decreases and the improvements in accuracy may not justify the computational time to generate and select more feeders.

Fig. 8 shows the relationship between the objective value and baseline upgrade cost across iterations. The optimization process terminates at iteration 3, setting $i_{opt} = 3$, as the 10% improvement threshold is exceeded in iteration 4. This results in an optimal baseline upgrade cost of \$1.8 million.

Computational experiments were conducted on a workstation with an Apple M2 chip, 8-core CPU (4 performance + 4 efficiency cores), and 16 GB of unified memory. The most time-consuming step is building the MILP for assigning buildings to feeder end-nodes (Section II-B), as the number of variables increases with the feeder's number of nodes and peak power. For large feeders, constructing this model can take several minutes, whereas smaller feeders are processed in just a few seconds. In contrast, the MILP for selecting the optimal feeder subset and assigning weights (Section II-C) is relatively fast, typically converging within a few minutes, with a maximum of 10 iterations in this study.

C. County Distribution Grid Investment: A Case Study

This section explains how the WFS_{county} can be used to estimate the investment needed in distribution grid assets to meet growing electricity demand. In order to provide realistic forward-looking projections, we focus on Alameda County and construct scenarios that reflect state-level policy targets and technology adoption trends. These scenarios are then used to estimate the grid upgrades required to accommodate projected demand increases.

1) Forward-Looking Electricity Consumption: We begin by projecting future demand conditions for 2030 in Alameda County, guided by two key statewide adoption targets: heat pumps (HPs) and electric vehicles (EVs).

California has set a goal of installing 6 million heat pumps [29] and deploying 7.1 million light-duty plug-in EVs [30] across its 14.5 million residential units by 2030. Assuming adoption is distributed uniformly across the state, this corresponds to an estimated 41% of households in Alameda County adopting heat pumps and about 50% adopting EVs. For EVs, we further assume at most one vehicle per household.

To capture different potential adoption levels, we consider three scenarios combining heat pump adoption with varying EV adoption rates (0% , 25% and 50%) in residential units. Table V summarizes the three scenarios, showing the adoption rates of heat pumps and EVs in residential units to consider.

The data sources used to build these scenarios are as follows: heating and cooling load profiles for HP adoption were derived from the ResStock measure package "ENERGY STAR Air-to-Air Heat Pump With Electric Backup" [24]; EV charging load profiles were obtained from the Ochre database, sampling 1,000 household-level charging behaviors, with 85% battery electric vehicles (BEVs) and 15% plug-in hybrid electric vehicles (PHEVs).

Fig. 9 presents the peak day electricity consumption in Alameda County under four scenarios: the baseline (observed load) and the three demand scenarios outlined in Table V. The peak occurs during summer when cooling demand is greatest. In the heat pump (HP) adoption scenario, peak load increases by approximately 4% . Because heat pumps provide cooling, electricity consumption rises particularly in households that previously lacked air conditioning, thereby adding stress to the grid.

TABLE V
ADOPTION RATES OF HEAT PUMPS (HP) AND ELECTRIC VEHICLES (EV) IN RESIDENTIAL UNITS FOR ALAMEDA COUNTY IN 2030

Scenario	HP Adoption	EV Adoption
1	41%	0%
2	41%	25%
3	41%	50%

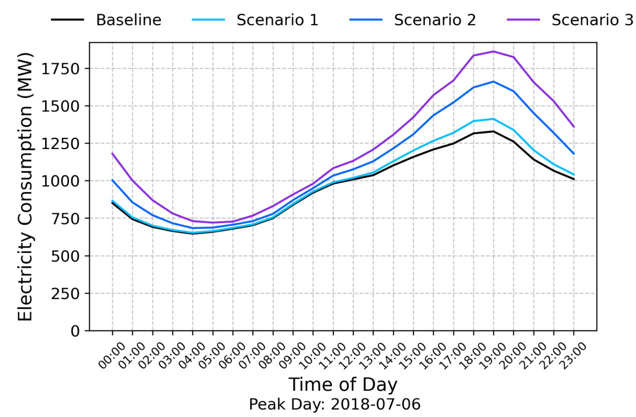
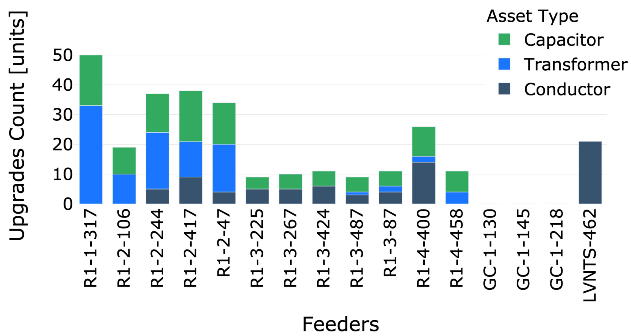
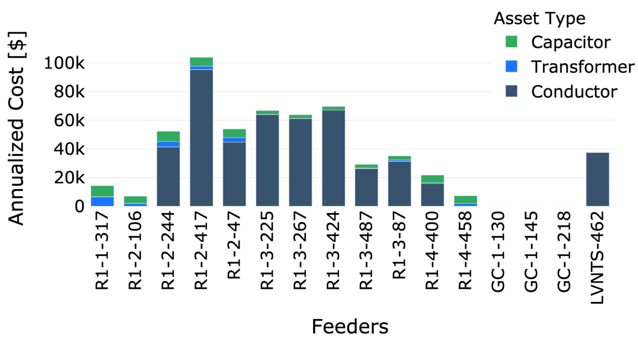


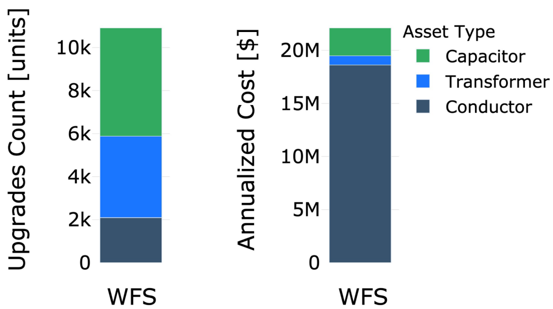
Fig. 9. Projected peak day electricity consumption in Alameda County under baseline (observed load) and the three demand scenarios.



(a) Count of Asset Upgrades/Additions per Feeder and per Asset



(b) Cost of Asset Upgrades/Additions per Asset and per Feeder



(c) Total Asset Upgrades/Additions and Associated Annualized Cost

Fig. 10. Analysis of asset upgrades and additions for scenario 3.

When electric vehicle adoption is combined with HP adoption, charging demand further amplifies peak loads, by 27% under 25% EV adoption and 42% under 50% adoption. The peak loads are concentrated in the evening when residential charging is most common. Since the peak occurs around 7 pm, solar PV would contribute little to reducing these peaks, while battery discharge and other DERs could help mitigate them. The integration of such technologies is left for future extensions of this work.

2) Distribution Grid Investment Upgrade Costs: To calculate the counts and associated costs of asset upgrades or additions, residential building load profiles were adjusted considering HP and EV adoption, and the Algorithm 2 was applied to determine feeder-level investment using an OpenDSS power flow analysis for the peak electricity consumption day in Alameda County. Although upgrade costs were computed for all scenarios, we highlight Scenario 3 here, as it represents the 2030 adoption projections.

Fig. 10(a) illustrate the total number of upgrades necessary for each feeder in $WFS_{Alameda}$ to operate without violations under the 2030 load projections (scenario 3), and Fig. 10(b) shows the corresponding annualized costs of these upgrades. Fig. 10(c) aggregate the results across all feeders, incorporating the weight of each feeder to calculate the total number of asset upgrades required for each type (line, transformer, capacitor) and the resulting total annualized cost at the county level.

Based on these results, the adoption of HPs and EVs planned for 2030 in Alameda County leads to violations across all residential feeders of $WFS_{Alameda}$. Commercial feeders originating from GC-12.47-1 do not experience load increase on peak day since the HP adoption is only considered for residential units. Feeders that originate from the same initial feeder tend to exhibit similar types of investment.

When considering the total number of grid upgrades across all feeders and their respective weights, capacitor additions account for 50% , transformer reinforcements for 30% , and line reinforcements for 20% . In terms of total costs, based on asset cost assumptions, line reinforcements represent the largest expense, comprising approximately 80% of the total investment required for this scenario, while capacitor upgrades account for about 18% .

For the other scenarios, the final distribution of asset investments in the WFS remains similar to those in scenario 3. In terms of annualized upgrade costs, scenario 1 amounts to 3 million USD, scenario 2 reaches 15.5 million USD, and scenario 3 reaches 22 million USD. This significant increases in scenario 2 and 3 highlight the direct impact of high EV adoption rates on distribution infrastructure costs.

This analysis illustrates how to leverage the representative feeder set to estimate distribution upgrade costs at the county level. The impact of HP and EV adoption can be evaluated on the feeders in $WFS_{Alameda}$ to project the effects for Alameda County. It is important to note that these estimates do not represent the total future costs, as they do not account for future load-decreasing factors such as load-decreasing DERs like PV systems and battery.

IV. Conclusion

In this work we demonstrated that by modifying taxonomy feeders, without compromising their original electric designs (in terms of topology and peak loads), it is possible to form sets of feeders that can accurately replicate the energy, power and distribution system costs of a county. These feeders sets can be used in forward-looking economic and policy studies to consistently capture distribution grid costs in future scenarios of load growth without historical precedent and for which empirical models, based on past utility investment records, are not applicable. Future work could extend this methodology to incorporate industrial loads and DERs, enabling the feeder set to better represent the diversity and flexibility of load types in each county and support the analysis of additional electrification scenarios.

ACKNOWLEDGMENT

The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

REFERENCES

[1] T. Mai et al., "Electrification futures study: Scenarios of electric technology adoption and power consumption for the United States," 2018. [Online]. Available: https://www.nrel.gov/docs/fy18osti/71500.pdf

[2] Y. Li and A. Jenn, "Impact of electric vehicle charging demand on power distribution grid congestion," *Proc. Nat. Acad. Sci.*, vol. 121, no. 18, 2024, Art. no. e2317599121. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.2317599121

[3] T. A. Deetjen, L. Walsh, and P. Vaishnav, "US residential heat pumps: The private economic potential and its emissions, health, and grid impacts," *Environ. Res. Lett.*, vol. 16, no. 8, 2021, Art. no. 084024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1748-9326/ac10dc/meta

[4] Edison Electric Institute, "Projected functional capital expenditures, 2015–2025, U.S. investor-owned electric utilities," Washington, D.C.: EEI Financial Analysis Department, Sep. 2025. [Online]. Available: https://www.eei.org/-/media/Project/EEI/Documents/Issues-and-Policy/Finance-And-Tax/Industry-Capital-Expenditures.pdf

[5] R. L. Fares and C. W. King, "Trends in transmission, distribution, and administration costs for U.S. investor-owned electric utilities," *Energy Policy*, vol. 105, pp. 354–362, 2017.

[6] "OEDI: FERC form 1 electric utility cost, energy sales, peak demand, and customer count data 1994-2019," 2017. [Online]. Available: https://data.openei.org/submissions/489

[7] DNV-GL, "Residential zero net energy building integration cost analysis," California Public Utilities Commission, Document No. 10007451-HOU-R-02-D, 2017.

[8] Kevala, Inc., "Electrification impacts study, Part 1: Bottom-up load forecasting and system-level electrification impacts cost estimates," in *Prepared California Public Utilities Commission, Energy Division, Proc. R.21-06-017*, San Francisco, CA, USA, May 9, 2023. [Online]. Available: https://www.kevala.com/resources/electrification-impacts-study-part-1

[9] R. Gupta et al., "Spatial analysis of distribution grid capacity and costs to enable massive deployment of PV, electric mobility and electric heating," *Appl. Energy*, vol. 287, 2021, Art. no. 116504.

[10] F. E. Marcos et al., "A review of power distribution test feeders in the United States and the need for synthetic representative networks," *Energies*, vol. 10, no. 11, 2017, Art. no. 1896. [Online]. Available: https://www.mdpi.com/1996-1073/10/11/1896/htm

[11] K. P. Schneider et al., "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181–3188, May 2018.

[12] K. Strunz, R. H. Fletcher, R. Campbell, and F. Gao, "Developing benchmark models for low-voltage distribution feeders," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2009, pp. 1–3.

[13] F. Pilo et al., "Atlantide — Digital archive of the Italian electric distribution reference networks," in *Proc. CIRED 2012 Workshop: Integration Renewables Distrib. Grid*, 2012, pp. 1–4.

[14] M. J. Ueda, O. Engblom, and K. Alvehag, "Representative test systems for swedish distribution networks," in *Proc. Int. Conf. Exhib. Electricity Distribution*, 2009, pp. 1–4.

[15] H. L. Willis, H. N. Tram, and R. W. Powell, "A computerized, cluster based method of building representative models of distribution systems," *IEEE Trans. Power App. Syst.*, vol. PAS-104, no. 12, pp. 3469–3474, Dec. 1985.

[16] R. J. Broderick, J. R. Williams, and K. Munoz-Ramos, "Clustering method and representative feeder selection for the california solar initiative," Sandia Nat. Lab. (SNL-NM), Albuquerque, NM, USA, Tech. Rep. SAND2014-1443, 2014. [Online]. Available: https://www.osti.gov/biblio/1200658

[17] K. P. Schneider, D. W. Engel, Y. Chen, S. E. Thompson, D. P. Chassin, and R. G. Pratt, "Modern grid initiative distribution taxonomy final report," 2008. [Online]. Available: http://www.ntis.gov/ordering.htm

[18] Y. Li and P. J. Wolfs, "Taxonomic description for Western Australian distribution medium-voltage and low-voltage feeders," *IET Gener., Transmiss. Distribution*, vol. 8, no. 1, pp. 104–113, 2014. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-gtd.2013.0005

[19] A. Berry, L. Collins, E. Oliver, and C. Perfumo, "Representative Australian electricity feeders with load and solar generation profiles. v1. CSIRO. Data Collection," 2025. [Online]. Available: https://doi.org/10.4225/08/5631B1DF6F1A0

[20] C. M. Domingo, T. G. S. Román, Á. Sánchez-Miralles, J. P. P. González, and A. C. Martínez, "A reference network model for large-scale distribution planning with automatic street map generation," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 190–197, Feb. 2011.

[21] C. Mateo et al., "Building large-scale U.S. synthetic electric distribution system models," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5301–5313, Nov. 2020. [Online]. Available: https://www.researchgate.net/publication/342321369_Building_Large-Scale_US_Synthetic_Electric_Distribution_System_Models

[22] "OEDI: SMART-DS synthetic electrical network data OpenDSS models for SFO, GSO, and AUS," 2024. [Online]. Available: https://data.openei.org/submissions/2981

[23] N. Rauschkolb, N. Limandibhratha, V. Modi, and I. Mercadal, "Estimating electricity distribution costs using historical data," *Utilities Policy*, vol. 73, 2021, Art. no. 101309.

[24] P. R. White et al., "ResStock 2024.2 release: Technical documentation," 2024, Accessed: Nov. 04, 2025. [Online]. Available: https://oedi-data-lake.s3.amazonaws.com/nrel-pds-building-stock/end-use-load-profiles-for-us-building-stock/2024/resstock_tmy3_release_2/resstock_documentation_2024_release_2.pdf

[25] A. Seres and M. Heleno, "US representative feeder sets for distribution grid economics and policy applications - Alameda county, CA case study, IEEE DataPort, *Ieee-dataport.org*, Apr. 09, 2025, Accessed: Nov. 04, 2025. [Online]. Available: https://ieee-dataport.org/documents/us-representative-feeder-sets-distribution-grid-economics-and-policy-applications-alameda

[26] K. Schneider, P. Phanivong, and J.-S. Lacroix, "IEEE 342-node low voltage networked test system," in *Proc. IEEE PES Gen. Meeting Conf. Expo.*, 2014, pp. 1–5.

[27] E3, "2024 distributed energy resources avoided cost calculator documentation," California Public Utilities Commission, San Francisco, CA, USA, Oct. 2024. [Online]. Available: https://www.ethree.com/public_proceedings/energy-efficiency-calculator/

[28] K. Horowitz, "2019 distribution system upgrade unit cost database current version"," *NREL Data Catalog*, Golden, CO: National Renewable Energy Laboratory, Jul. 22, 2025, doi: 10.7799/1491263.

[29] California Heat Pump Partnership, "Scaling California's heat pump market: The path to six million," 2023, Accessed: Nov. 4, 2025. [Online]. Available: https://heatpumppartnership.org/wp-content/uploads/2025/03/CAHPP-Blueprint-Final.pdf

[30] California Energy Commission, "Assembly bill 2127 second electric vehicle charging infrastructure assessment: Assessing charging needs to support zeroemission vehicles in 2030 and 2035," Sacramento, CA, USA: California Energy Commission, Tech. Rep. CEC-600-2024-003, 2024.