Methods for Computing Physically Realistic Estimates of Electric Water Heater Demand Response Resource Suitable for Bulk Power System Planning Models

Elaine Hale, Matt Leach, Brady Cowiestoll, Yashen Lin, and Daniel Levie

National Renewable Energy Laboratory

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-6A40-82315
November 2022

Methods for Computing Physically Realistic Estimates of Electric Water Heater Demand Response Resource Suitable for Bulk Power System Planning Models

Elaine Hale, Matt Leach, Brady Cowiestoll, Yashen Lin, and Daniel Levie

National Renewable Energy Laboratory

Suggested Citation
Hale, Elaine, Matt Leach, Brady Cowiestoll, Yashen Lin, and Daniel Levie. 2022. *Methods for Computing Physically Realistic Estimates of Electric Water Heater Demand Response Resource Suitable for Bulk Power System Planning Models*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A40-82315.
https://www.nrel.gov/docs/fy23osti/82315.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-6A40-82315
November 2022

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Acknowledgments

This work was supported by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) Building Technologies Office (BTO). The authors especially thank Monica Neukomm (BTO) for her support. We also thank Ardelia Clarke, Peter DeWitt, and Trevor Stanley for computational assistance; Jiazi Zhang and Luke Lavin for improving our baseline production cost models; Marjorie Schott for graphical design; and Mike Meshek for technical editing. This report was improved by review comments from Jeff Macguire, Trieu Mai, Jaquelin Cochran (NREL), Richard O'Neill, Ookie Ma, and Amir Roth (DOE) who we thank wholeheartedly for taking the time to provide feedback.

Acronyms

DER distributed energy resource

DOE U.S. Department of Energy

DR demand response

DRA demand response asset

EMS energy management system

ERWH electric resistance water heater

HPWH heat pump water heater

ISO independent system operator

ISO-NE ISO New England

LMP locational marginal price

MISO Midcontinent Independent System Operator

PJM a regional transmission organization in the Eastern United States, formerly the Pennsylvania-New Jersey-Maryland Interconnection

RE renewable energy

VG variable generation

WH water heater

Notation

Table A. Units

Abbreviation	Name, Physical Quantity
°C	degrees Celsius, temperature
°F	degrees Fahrenheit, temperature
GW	gigawatt, power
GWh	gigawatt-hours, energy
K	kelvin, temperature
kg	kilograms, mass
kW	kilowatts, power
kWh	kilowatt-hours, energy
m	meters, length
MW	megawatts, power
MWh	megawatt-hours, energy
MW-h	megawatt-hours, reserves
yr	year, time

Table B. Set, Element, and Subscript Symbols

Symbol	Description
$\mathbb{K}$	set of water heaters in ResStock sampled portfolio of single family homes
k	single water heater or other controllable device
start	subscript indicating the notification period of a contingency reserve event or audit
response	subscript indicating the response period of a contingency reserve event or audit
rebound	subscript indicating the period of time immediately following a contingency reserve event or audit

Table C. Variables

Symbol	Description
L	aggregate flexibility energy storage variable, analogous to ΔS (MWh)
P	water heater electricity consumption (kW or MW)
ΔP	difference between actual (+) and baseline (-) water heater electricity consumption (kW or MW)

Continued

Symbol	Description
DP	alias for ΔP in time-discretized computational routines
S	thermal energy stored in one or more water heaters (kWh or MWh)
ΔS	difference between actual (+) and baseline (-) thermal energy stored in one or more water heaters (kWh or MWh)
DS	alias for ΔS in time-discretized computational routines
t	time
T	temperature of a thermostatically controlled load (°C or °F)
T_{tank}	water heater tank temperature (°C or °F)
ΔT_{tank}	difference between actual (+) and baseline (-) water heater tank temperatures
U	aggregate flexibility electricity consumption variable, analogous to ΔP (MW)

Table D. Parameters

Symbol	Description
α	thermal energy dissipation (h^{-1})
β_k	fraction of an aggregate dispatch signal to be allocated to resource $k \in \mathbb{K}$
η	efficiency of electricity to thermal energy conversion (kW-thermal/kW-electricity)
ρ	density (kg/m^3)
a	ambient temperature (°C or °F)
c	thermal capacitance (kWh/°C)
c_v	specific heat of water (kWh/kg·K)
$\underline{L}$	minimum allowable state of charge for an aggregate flexibility resource (MWh)
N	number of time-steps
$\tilde{P}$	baseline (without demand response interventions) electricity consumption (kW or MW)
$\overline{P}$	maximum power consumption
$\underline{\Delta P}$	minimum allowable difference between actual and baseline power consumption (kWh or MWh)
$\overline{\Delta P}$	maximum allowable difference between actual and baseline power consumption (kWh or MWh)
Q	exogenous heat gains within water heater tanks and other thermostatically controlled loads (kW)
r	thermal resistance (°C/kW)
δt	discretization time-step (minutes)

Continued

Symbol	Description
$\widetilde{S}$	baseline (without demand response interventions) thermal energy stored in one or more water heaters (kWh or MWh)
$\underline{\Delta S}$	minimum allowable difference between actual and baseline thermal energy stored (kWh or MWh)
$\overline{\Delta S}$	maximum allowable difference between actual and baseline thermal energy stored (kWh or MWh)
$\widetilde{T}$	baseline (without demand response interventions) temperature of a thermostatically controlled load
$\overline{T}$	maximum allowable temperature for a thermostatically controlled load (°C or °F)
$\underline{T}$	minimum allowable temperature for a thermostatically controlled load (°C or °F)
T_{sp}	temperature set point
$\widetilde{T}_{\text{tank}}$	baseline (without demand response interventions) water heater tank temperature (°C or °F)
V	water heater tank volume (m^3 or gal)
w	weight to place on energy storage, as opposed to power, capacity
w_k	sample weight corresponding to simulated water heater k

Executive Summary

Demand response is commonly called on to reduce load during system peak times or to respond to contingency events. In future power systems with higher shares of wind and solar generation (which we describe together as variable generation [VG]), demand response could have more opportunities to provide energy shifting or operating reserve services. Demand response and other forms of demand flexibility are often evaluated using historical or projected information about how grid systems operate, usually in the form of grid service price time series or marginal emissions time series. This price-taking assumption is convenient for demand-side modeling but only holds if the demand-side intervention does not change bulk power system build-out and dispatch enough to change grid service prices or marginal emissions. On the other hand, price-making approaches that directly represent demand response as a resource alongside supply-side options and are thus able to impact bulk power system build-out or dispatch usually make simplifying assumptions about round-trip efficiency, equipment headroom, and the aggregation of many kilowatt (kW) scale devices into megawatt (MW) scale flexibility resources that can misrepresent the size of the demand flexibility resource. This report develops and explores new computational methods for bridging these gaps by directly representing device-level flexibility and analyzing aggregation processes using a model form that can be incorporated directly into grid modeling workflows as price-making resources.

The methods are evaluated and refined by applying them to estimate the ability of residential electric water heaters, both electric resistance water heaters (ERWHs) and heat pump water heaters (HPWHs), to provide energy shifting and operating reserve services in envisioned future New England power systems. We start from detailed whole-building energy models that realistically represent New England single family home stock and use a battery-like surrogate model to represent operational flexibility in a form suitable for linear and mixed integer programming. This enables fast computation of optimal energy shifting for individual water heaters based on a fixed price profile, construction of aggregate, MW-scale models for quickly estimating contingency reserve resource, and in some cases the construction of aggregate, MW-scale models for energy shifting that can be directly included in bulk power system planning models. This report also explores the cost of aggregation—a battery-like model of a MW-scale resource cannot convey the exact capabilities of the individual resources it comprises. We therefore compute and compare inner and outer approximations, which are aggregate models that bound the actual flexibility of the comprising devices. Several limitations (described below) advise caution when interpreting the quantitative results of this study. However, that should not detract from the promise of new computational methods for more realistically representing the MW scale impacts of aggregated kW scale demand flexibility.

Findings

We estimate contingency reserve resource by fitting surrogate models to the ERWHs and HPWHs represented in two scenarios of the New England housing stock, aggregating the surrogate models, and using the aggregates to simulate event responses for each hour of the year. We select surrogate model parameters by comparing a much smaller, but still representative, number of surrogate model-simulated contingency event responses to EnergyPlus-simulated contingency event responses. The validation and resource estimation processes demonstrate that our results are sensitive to how available water heater measurements, especially temperatures, are mapped to

surrogate model parameters, and to the specific characteristics of the water heaters. For example, our per water heater-year estimates of contingency reserve resource are 0.8 to 1.2 MW-h for ERWHs and 1.0 MW-h for HPWHs, but those results are influenced by the details of the surrogate modeling and by large HPWH tanks (80 gallons compared to ERWH volumes that range between 20 and 60 gallons).

We explore energy shifting from ERWHs and HPWHs by dispatching surrogate models for individual water heaters against day-ahead prices from two operational models of future ISO-NE power systems: Near-Term VG and Mid-Term VG. The individual surrogate models are able to access and potentially shift all 1,547 GWh of ERWH load and 640 GWh of HPWH load modeled in two different scenarios of New England single family homes. Subject to the same ERWH and HPWH modeling caveats as above, we find ERWH energy cost savings of $38/WH-yr based on the Near-Term VG day-ahead prices and $39/WH-yr based on the Mid-Term VG day-ahead prices. The Near-Term VG and Mid-Term VG energy cost savings for HPWHs are $19/WH-yr and $22/WH-yr, respectively.

Aggregating surrogate models to the MW-scale for energy shifting service is more challenging than for contingency service and we only present such results for ERWHs, because we were unable to determine satisfactory ways to deal with HPWHs' time-varying and path dependent operational characteristics. For ERWHs, we compute inner approximation aggregate models that produce provably feasible dispatch instructions, and outer approximations that are upper-bound overestimates of actual flexibility. The best inner approximation aggregates present about 30% of the total ERWH load for energy shifting, but when those aggregate models are dispatched against Near-Term VG and Mid-Term VG day-ahead prices they only shift about 15% of the energy shifted by individual ERWH surrogate models under the same conditions due to very restrictive energy bounds. On the other hand, dispatching the outer approximation against day-ahead prices results in 7% to 17% more energy shifting compared to the individual ERWH surrogate models because the outer approximation is able to unrealistically pair power capacity and energy capacity from different water heaters. The inner and outer approximation models can also be directly dispatched in large-scale grid models. The inner approximation models produce unsatisfactory results because their restrictive energy bounds are computationally challenging and yield very conservative results. In comparison, the outer approximation models are more suitable, even though they overestimate flexibility on the order of 10% to 20%. By dispatching the ERWH outer approximations directly in the two ISO-NE power system day-ahead models we demonstrate that ERWH energy shifting can be directly coordinated with supply-side resources, e.g., it can impact large generator unit commitment decisions, displace less efficient forms of storage, and set day-ahead energy prices.

Limitations

The battery-like surrogate models used to predict the ability of individual water heaters to operate flexibly in response to grid signals are simplifications with key parameters that can be difficult to populate in a way that best represents flexibility around baseline operations. Specifically, we found that tank temperature is a key parameter that should be set to the controlled temperature reading minus the controller dead band, but it can be difficult to ascertain exactly what those values are, especially in tanks with multiple heating elements and especially in the real-world where measurements and other information on individual water heaters could be limited. This

study uses average tank temperature as a proxy for the controlled tank temperature minus the dead band. We found this to be a reasonable but not ideal proxy for our ERWH models, and an overly conservative one for the HPWH models.

Methodologically, we found that our provably dispatchable aggregation methods (inner approximations) are overly conservative, especially when used for energy shifting. In some cases, simply adding up the flexibility bounds of individual devices to compute outer approximations might produce acceptable representations of aggregate flexibility. However, in others such an approach will overestimate available flexibility by pairing, e.g., the ability of one water heater that is off but able to reduce tank temperature with the ability of another water heater that is below temperature set point to turn off. This study also does not suggest how to aggregate the flexibility of devices with time-varying and state-dependent parameters like heat pump water heaters and air conditioners. Certainly utilities and aggregators today use data-driven and heuristic methods to control such devices to provide demand response services. Nonetheless we remain curious about whether it is possible to estimate the aggregate flexibility of such resources using physics-based surrogate models.

The comparisons between ERWHs and HPWHs are impacted by two modeling artifacts. The version of ResStock used for this study had an ERWH modeling bug that set the element capacity too low, leading to longer heating cycles and less load to shed. This version of ResStock also only had an 80 gallon HPWH available; although 80 gallon HPWHs are available in the market, 50 gallon tanks tend to be more common. In combination, these two limitations will tend to overstate the flexibility of HPWHs as compared to ERWHs. Both of these issues have since been corrected in newer versions of ResStock.

Table of Contents

List of Figures

List of Tables

1 Introduction

Demand response, that is, electrical load contracted to provide a grid service by modifying demand-side operations, has a long history. Most commonly, demand response is called on to reduce load during system peak times or to respond to contingency events. Looking to future power systems with much higher shares of wind and solar generation, demand response could have the opportunity to provide more types of grid services more often, for example, energy shifting or various forms of operating reserve (Alstone et al. 2017; Hale, Stoll, and Novacheck 2018; Zhou and Mai 2021; Hale et al. 2021).

How can we evaluate the role of demand response in future power systems? Ideally demand response and supply-side resources (i.e., new and existing generators, transmission lines, and storages) would be evaluated simultaneously and in a technology-agnostic manner. This means capturing all key demand response characteristics and capabilities in grid planning tools and processes, which to first order has been done. Several studies have represented demand response as a zero-cost resource in grid capacity expansion and production cost models (Hummon et al. 2013; O'Connell et al. 2015; Stoll, Buechler, and Hale 2017; McPherson and Stoll 2020; Murphy et al. 2021), and other studies have developed demand response supply curves and other cost data (Alstone et al. 2017; Potter and Cappers 2017; Nubbe et al. 2021). Incorporating demand response dispatch capabilities, enablement costs, and participation supply curves into grid planning models allows demand response to be a *price-maker*; that is, demand response can change investment and operational decisions and thus directly inform grid service marginal prices.

However, the price-maker methodologies developed thus far, while capturing different time-varying ability per grid service, storage-like behavior for energy shifting, and sometimes additional constraints (e.g., number and duration of calls per day) still rely on assumptions such as 100% round-trip efficiency and non-physical estimates of equipment headroom (Hummon et al. 2013; Hale, Stoll, and Novacheck 2018; McPherson and Stoll 2020; Sun et al. 2020; Zhou and Mai 2021) that have not been calibrated to ensure sufficient realism or dispatchability. These drawbacks are largely the result of the need for grid models of demand response to fit into the large, megawatt (MW) to gigawatt (GW) scale linear and mixed-integer programs used for grid investment and operations planning (Bloom et al. 2016; Cohen et al. 2019) and the difficulties presented by trying to reflect demand flexibility characteristics that are well-known at the individual device level (kW-scale) (Cutler et al. 2013; Cole et al. 2014; O'Connell et al. 2015; MacDonald, Vrettos, and Callaway 2020; Luo, Langevin, and Chandra Putra 2021) at the aggregate MW-scale.

A parallel line of research has thus evaluated demand response from the demand-side perspective. Given historical or projected information about how grid systems operate, usually in the form of grid service price time series (e.g., locational marginal prices [LMPs], reserve prices, capacity prices, or time-of-use tariffs) or emissions time series (e.g., average, short-run marginal, or long-run marginal carbon, greenhouse gas, or other emissions), very detailed models of, for example, air conditioning, water heating, or industrial processes can be used to determine the value of responding to such signals (Mitra et al. 2012; Cole et al. 2014; Jin et al. 2017; Shah et al. 2020; Wang 2021; Langevin et al. 2021; Garfield et al. 2021). These methods are sometimes described as *price-taking* approaches, and they generally reverse the pro-con list of price-making models.

That is, price-taking methods provide ample opportunity for demand-side realism because they are not computationally constrained by the mathematical form or size of grid planning models, but they are separate from grid planning models and may therefore have less influence on grid planning processes. Furthermore, implementing results from a price-taking analysis might not have the intended effect of reducing system costs because of the resulting interactions with the supply side.

In this report, we describe methods for bridging the gap between price-taking and price-making demand response analysis for residential electric water heaters. Electric water heaters have provided demand response, including daily load shifting, for decades, but only in a few places (e.g., Minnesota and France) (Lescoeur and Galland 1987; Opalka 2013; Hopkins and Whited 2017). Because higher wind and solar generation shares combined with, for example, space and water heating electrification, might increase the need for regular energy shifting and diversify times of high grid stress to include winter peaks, times of low net-load[1] or high ramping, and times of low renewable availability (EPRI 2018; Murphy et al. 2021; Cochran et al. 2021), electric water heaters could become a more important source of power system capacity, contingency reserve, and energy shifting in many regions. Electric water heaters also provide an interesting case study in energy efficiency and demand response interactions, as heat pump water heaters (HPWHs) are much more efficient than ERWHs and might therefore have less grid services potential because each water heater has a smaller maximum capacity and represents less potentially shiftable load.

Recent related work examining the demand response potential of electric water heaters includes a DR pilot program that documented the ability of water heaters to contribute to peak load reduction during infrequent demand response events (Oschsner et al. 2011; BPA 2018), and to provide more regular energy-shifting (BPA 2018). Electric resistance and heat pump water heater flexibility has been tested and compared in a laboratory setting (Mayhorn et al. 2015). The grid value of electric resistance and heat pump water heaters has been evaluated analytically in terms of peak shaving, thermal storage, energy efficiency, and fast frequency response (i.e., regulation or balancing services) based on PJM and MISO market prices (Hledik, Chang, and Lueken 2016). HPWH energy shifting has been studied using simulation models calibrated to laboratory data for water heaters from multiple vendors and with both R134a and CO_2 refrigerants. Energy shifting value was estimated for multiple control strategies against both marginal energy costs and a time-of-use tariff (Carew et al. 2018).

In this study, we bridge price-taking and price-making approaches for analyzing demand response from water heaters using detailed building energy simulations (Figure 1, bottom left corner), dynamic surrogate models of individual device-level flexibility (upper left), and aggregation methods (center top), coupled with price-making (in a large-scale production cost model, lower right corner) and price-taking (simulation of response to contingency events or dispatch against modeled day-ahead prices, center bottom) methods. Because our goal is to improve the physical realism of demand response in grid models, we focus on the potential resource that could be offered by the residential water heating end use and do not evaluate enablement costs

[1]Net-load is commonly defined as electricity load minus wind, solar and other variable generation (VG). In high VG systems, net-load can be better correlated than load with times that are key for ensuring resource adequacy (Stephen, Hale, and Cowiestoll 2020). Periods of low net-load are often associated with times of VG curtailment.

nor likely participation rates. We also focus on contingency reserve and energy shifting grid services to the exclusion of other grid services. We examine contingency reserves because this is an important reliability service in its own right, and because the methods for estimating the amount of contingency reserve resource can be adjusted to estimate firm capacity contributions. Firm capacity and energy (which can be impacted by energy shifting) are the largest two value streams in bulk power systems (Neukomm, Nubbe, and Fares 2019; Denholm, Sun, and Mai 2019).

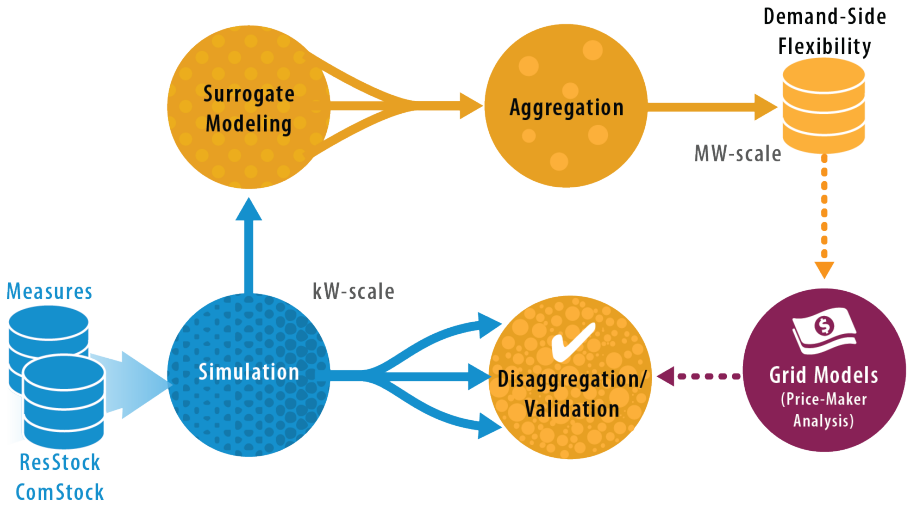


Figure 1. Overview of bottom-up engineering approach for estimating demand response resource.

Following this introduction (Chapter 1), the report introduces electric water heater flexibility by presenting basic simulation results for electric resistance water heaters (ERWHs) and heat pump water heaters (HPWHs) operating in New England, describing our dynamic surrogate model formulation and how we use building energy model data to define the surrogate model parameters for sample single family homes (Chapter 2). The building stock model used in this study is further described in Chapter 3, as are the grid models we use to analyze the potential value of electric water heater demand response resource in New England. Chapter 4 describes how we model contingency response from individual and aggregated ERWH and HPWH water heaters, validation results comparing surrogate model and EnergyPlus simulated responses, contingency resource estimates, and the impacts of making contingency reserves from ERWHs or HPWHs available to large-scale models of grid operations with different shares of variable generation (VG). Chapter 5 describes methods for estimating and evaluating the impact of electric water heaters shifting demand from higher to lower price times, as well as their potential bulk power grid impacts. Finally, the report concludes with Chapter 6.

2 Electric Water Heater Flexibility

Residential water heaters in the United States, which typically have tanks, are fueled by either natural gas or electricity and are thermostatically controlled to a fixed temperature set point. Because hot water is blended with cold water to produce the desired temperature for most applications, electric water heaters with tanks are a ubiquitous source of thermal energy storage that can be operated flexibly to provide grid services by modulating their power consumption while maintaining tank temperature within an acceptable range, rather than aiming for a specific set point.

In this report, we analyze the potential of the electric water heater fleet in New England to provide grid services. We rely on the ResStock model of U.S. single family homes as our "ground-truthed" description of the stock (Wilson et al. 2016). Using default settings, ResStock estimates that of 2,256,000 single-family homes in New England, about 27% of them have ERWHs (Figure 2) and less than 0.5% have HPWHs.

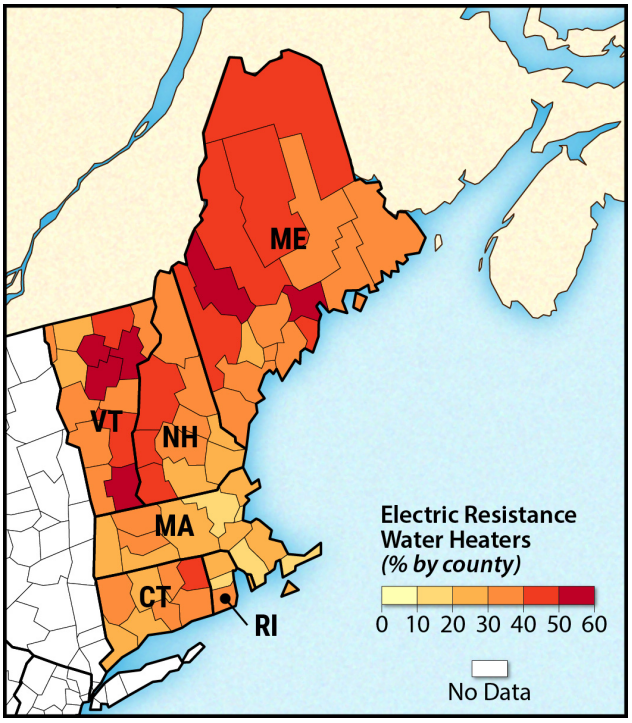
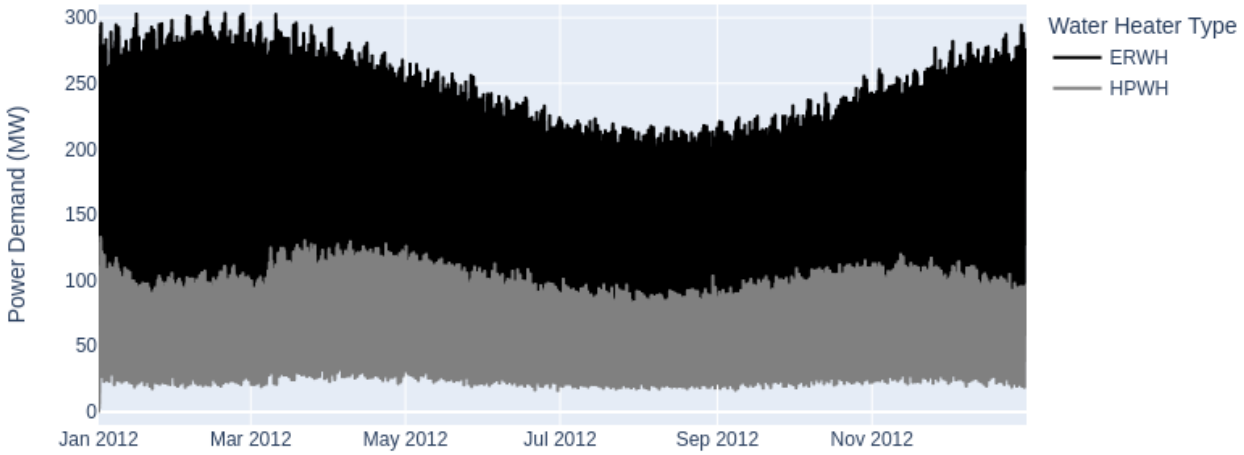
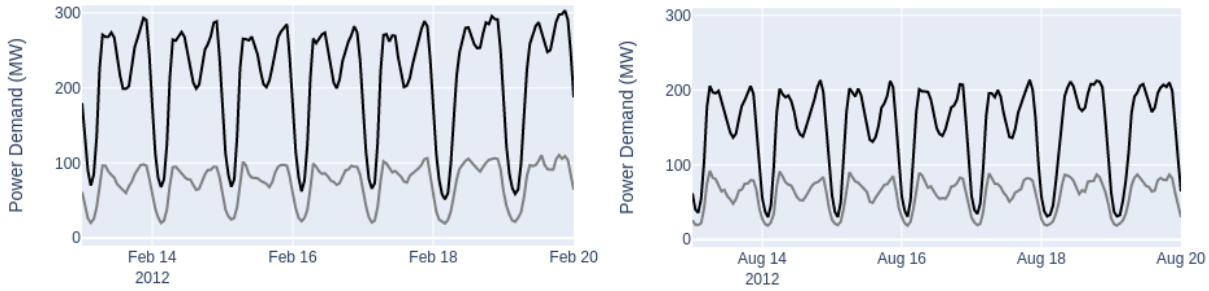


Figure 2. Fraction of New England single family homes with ERWHs, as estimated by ResStock

We examine the difference in potential flexibility between the current electric water heater fleet (as represented in ResStock with default settings) and a hypothetical scenario in which all electric water heaters are converted to 80 gallon HPWHs. Figure 3 shows the substantial difference in electricity use for 620,000 electric water heaters (approximately the current number of electric water heaters in New England) depending on type. Overall, ResStock estimates that 620,000 ERWHs in New England would use 1,589 GWh/yr of electricity whereas the same number of 80-gallon HPWHs would only use 641 GWh/yr. On a per-water-heater basis, that corresponds to 2,573 kWh/ERWH-yr and 1,034 kWh/HPWH-yr, on average.



(a) Full year hourly time series



(b) Example winter week **(c) Example summer week**

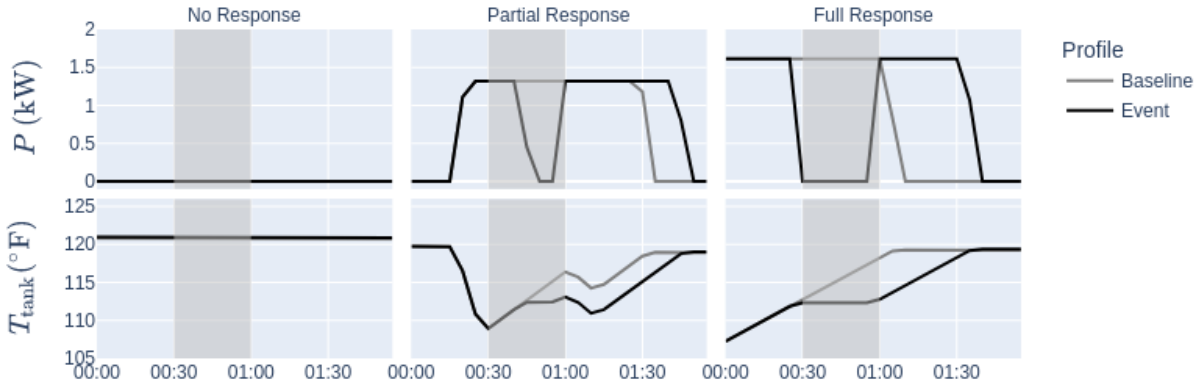
Figure 3. Power demand of 620,000 electric water heaters, either electric resistance (ERWH) or heat pumps (HPWH), as simulated by ResStock. Total annual energy is 2,573 kWh/ERWH-yr and 1,034 kWh/HPWH-yr. In (a) the profiles obscure each other and the time compression obscures daily and seasonal energy use trends especially for HPWHs. However, the relative magnitude of ERWH versus HPWH energy use is visible, and the ERWH profile shows the typical pattern of higher water heater energy use in the winter as compared to the summer.

Using ResStock, we can get a quick initial look at how the operations of these two types of water heaters can be perturbed to produce a change in power draw in response to a grid signal. Figure 4 shows data examples created by comparing otherwise identical ResStock runs with and without 20°F temperature set point reductions included during 30-minute event periods. Each of the three columns in the two subplots shows a different set of results, and together these results demonstrate that the actual response varies tremendously by water heater. The left column shows examples of no-responses that occur because the water heaters did not need to run during the event period under normal conditions (as evidenced by *P* Baseline values of zero). The right column shows "full" responses in which the water heater would have normally run during the entire event and can instead be shut off without violating the new, lower than normal, set point of 105°F (which is subject to an additional 10°F dead band). In the middle column, we see "partial" responses, which can occur because the water heater normally would not have run through the entire event period (as in the HPWH example, Figure 4b) or because the tank temperature reaches the lower set point during the event and thus turns on despite the ongoing event (as in the ERWH example,[2] Figure 4a). In general, Figure 4 shows that although the non-event temperature set point is 125°F in all cases, average tank temperatures can deviate significantly from that ideal for long periods of time based on the details of how water heaters are controlled (e.g., measurement locations, heating element locations, control dead bands). Figure 4a also highlights that this study's ERWHs were affected by a bug in an earlier version of ResStock that led to reduced capacity heating elements and was only triggered by certain water heater options not used in other ResStock projects at the time. Specifically, the ERWHs in this study have power capacities of 1.3 kW and 1.6 kW rather than the expected 4.5 kW and 5.5 kW (https://github.com/NREL/resstock/pull/804). This bug has since been fixed and to our knowledge only affected this study. Because the energy demands remain unchanged, the ERWH on-cycles are about three times too long compared to real-world water heaters, potentially impacting average outlet temperatures as well as load flexibility at timescales shorter than one hour.

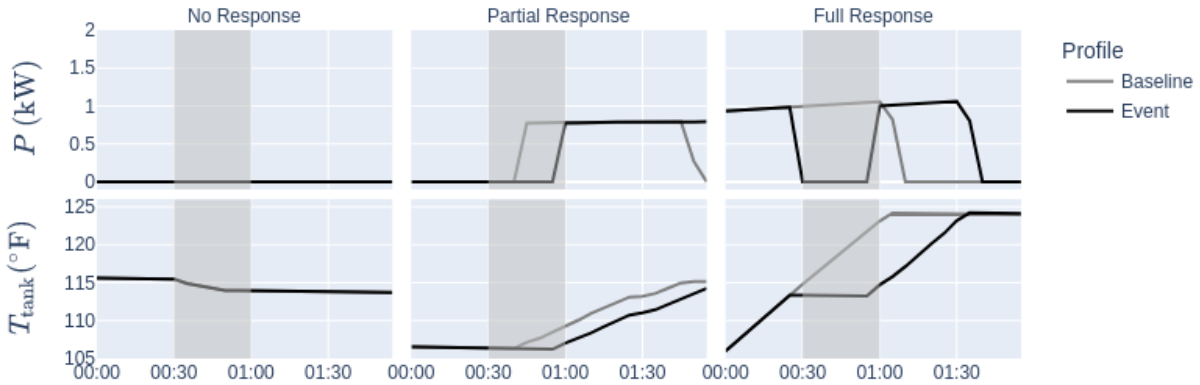
Although collecting and analyzing data like these is a fairly straightforward and intuitive process, because water heaters can be flexibly operated in many ways (e.g., for grid stress events of different duration and at different times, or to shift electricity use from high-price to low-price times, which can vary day-to-day, season-to-season, and year-to-year), we develop systematic representations of flexibility through a surrogate modeling framework.

---

[2]The reported temperature is the average tank temperature, which best captures the overall amount of thermal energy stored in the tank, but is usually higher than the temperature measurement used for control in our ERWH stratified tank models (Maguire 2018). Thus, although the baseline (average) tank temperature does not reach the event set point of 105°F, the EnergyPlus control logic is likely responding to a simulated measurement low in the tank and it is that temperature that is dropping below the lowered set point minus an additional 10°F dead band. Generally, electric resistance water heaters with tanks contain two heating elements, an upper element and a lower element, both of which respond to nearby temperature measurements. Different hot water draw patterns result in different cycling patterns of the two heating elements, different stratification states, and varying differences between average and controlled tank temperatures. In this study, the ERWH control temperatures tended to be significantly lower than the average tank temperatures, which resulted in water heaters often turning on mid-event as shown in Figure 4a.



(a) Example ERWH responses



(b) Example HPWH responses

Figure 4. Example responses for 30-minute demand response-type events, simulated in EnergyPlus by applying temperature set point reductions during the shaded event periods. Each row depicts a different metric: water heater power draw P (kW) or average water heater tank temperature $T_{\text{tank}}(^{\circ}\text{F})$. The columns depict different types of possible responses: No Response, Partial Response, or Full Response. Each plot shows data for both the Baseline and simulated Event runs. Responses relative to baseline can be computed by subtracting Baseline data from Event data.

2.1 Surrogate Model Formulation

Water heaters are thermostatically controlled loads similar to air conditioning, space heating, and refrigeration. The simplest type of physically realistic model for such loads is the resistance-capacitance model, a first-order ordinary differential equation with the controlled temperature $T(t)$ (°C) as the state variable. Here we write the model generally for heating loads, in which the ambient temperature $a(t)$ (°C) is generally lower than the controlled temperature, and injecting power into the system ($P(t) > 0$) increases the controlled temperature:

$$\frac{dT(t)}{dt} = \frac{a(t) - T(t)}{r(t)c(t)} + \frac{\eta(t)P(t)}{c(t)} + \frac{Q(t)}{c(t)} \tag{2.1}$$

$$0 \leq P(t) \leq \overline{P}(t) \tag{2.2}$$

The key parameters in the model are thermal resistance ($r(t)$ in $^{\circ}C$/kW); thermal capacitance ($c(t)$ in kWh/$^{\circ}C$); electricity to thermal energy conversion efficiency ($\eta(t)$, which is dimensionless and for HPWHs is also referred to as the coefficient of performance); and the maximum power of the equipment ($\overline{P}(t)$ in kW). Q (kW) represents the exogenous heat gains (or losses) for

the system, which for water heaters corresponds to heat loss associated with hot water use that is then backfilled by water at temperature of the main (i.e., tap).

To make this model more closely resemble a storage device, we can change the state variable to the amount of stored thermal energy S (kWh) defined as:

$$S(t) = c(t)\left(T(t) - a(t)\right). \tag{2.3}$$

For water heaters, the thermal capacitance is that of the water in the tank, which we can compute using the data in Table 1 assuming we know the volume and temperature of the tank.[3] We can then interpret S as the thermal energy stored in the water that is due to its temperature being higher than the surrounding environmental temperature $a(t)$, which could be that of a basement, garage, or storage closet where the water heater resides. Note that this is but one definition of stored thermal energy that we use to track how the tank interacts with its environment; however, from an end user perspective, the difference between tank temperature and mains temperature is likely more salient.

Table 1. Water Properties for Calculating Thermal Capacitance

T (°C)	c_v(kWh/kg·K)	ρ(kg/m^3)
0.0	0.001172	999.85
10.0	0.001164	999.70
20.0	0.001155	998.21
25.0	0.001149	997.05
30.0	0.001144	995.65
40.0	0.001132	992.22
50.0	0.001118	988.04
60.0	0.001105	983.20
70.0	0.001090	977.76
80.0	0.001076	971.79
90.0	0.001061	965.31
100.0	0.001047	958.35

Making the change of variables in Equation 2.1 results in:

$$\frac{dS(t)}{dt} = \eta(t)P(t) - \frac{S(t)}{r(t)c(t)} + \frac{S(t)}{c(t)}\frac{dc(t)}{dt} + Q(t) - c(t)\frac{da(t)}{dt}. \tag{2.4}$$

If we also have upper ($\overline{T}$ [°C]) and lower ($\underline{T}$ [°C]) bounds on acceptable temperatures, we can add bounds on the stored thermal energy:

$$c(t)\left(\underline{T}(t) - a(t)\right) \leq S(t) \leq c(t)\left(\overline{T}(t) - a(t)\right). \tag{2.5}$$

Equation 2.4, Equation 2.5, and Equation 2.2 can then be used together to represent the capability of water heaters (and other thermostatically controlled devices) to store thermal energy. However,

[3] We use constant volume thermal capacitance because the tank has a fixed volume and is full of water at the mains pressure.

what we really want to represent is the potential to operate water heaters flexibly *relative to how they would operate under baseline conditions*. To this end, we define a baseline temperature set point $\widetilde{T}$ (°C), which yields a baseline thermal energy storage quantity

$$\widetilde{S}(t) = c(t)\left(\widetilde{T}(t) - a(t)\right) \tag{2.6}$$

and if we assume that under baseline conditions the tank holds at a constant value $\widetilde{T}$ we can compute baseline power as:

$$\widetilde{P}(t) = \frac{\widetilde{S}(t)}{\eta(t)r(t)c(t)} - \frac{Q(t)}{\eta(t)}. \tag{2.7}$$

The ability to operate flexibly around these points can then be defined using the variables:

$$\Delta S(t) = S(t) - \widetilde{S}(t) \tag{2.8}$$

$$\Delta P(t) = P(t) - \widetilde{P}(t). \tag{2.9}$$

By substituting Equation 2.3 and Equation 2.6 into Equation 2.8 we see that:

$$\Delta S(t) = c(t)\left(T(t) - \widetilde{T}(t)\right) \tag{2.10}$$

and we can also derive:

$$\frac{d\Delta S(t)}{dt} = \eta(t)\Delta P(t) - \frac{\Delta S(t)}{r(t)c(t)} + \frac{\Delta S(t)}{c(t)}\frac{dc(t)}{dt} - c(t)\frac{d\widetilde{T}(t)}{dt} \tag{2.11}$$

which we simplify to:

$$\frac{d\Delta S(t)}{dt} = \eta(t)\Delta P(t) - \alpha(t)\Delta S(t) \tag{2.12}$$

by assuming $dc(t)/dt$ and $d\widetilde{T}(t)/dt$ are zero or small enough to ignore, and by defining the rate of thermal energy dissipation α (h^{-1}) to be equal to $1/r(t)c(t)$.

2.2 Estimating Individual Flexibility with ResStock

Given the surrogate flexibility model:

$$\frac{d\Delta S(t)}{dt} = \eta(t)\Delta P(t) - \alpha(t)\Delta S(t) \tag{2.13}$$

$$\underline{\Delta S}(t) = c(t)\left(\underline{T}(t) - \widetilde{T}(t)\right) \leq \Delta S(t) \leq c(t)\left(\overline{T}(t) - \widetilde{T}(t)\right) = \overline{\Delta S}(t) \tag{2.14}$$

$$\underline{\Delta P}(t) = -\widetilde{P}(t) \leq \Delta P(t) \leq \overline{P}(t) - \widetilde{P}(t) = \overline{\Delta P}(t). \tag{2.15}$$

and a ResStock simulation of residential housing stock, the parameters η, α, $\widetilde{S}$, $\underline{\Delta S}$, $\overline{\Delta S}$, $\widetilde{P}$, $\underline{\Delta P}$ and $\overline{\Delta P}$ can be specified several ways. In this report, we directly use the parameters specified in the ResStock EnergyPlus models when available.

For individual ERWHs, η, r, the volume of the water tank V and $\overline{P}$ are directly available and constant for each water heater. For η, we simply read the field Heater Thermal Efficiency. For

r, we read the fields Off-Cycle Loss Coefficient to Ambient Temperature and On-Cycle Loss Coefficient to Ambient Temperature. These two parameters describe the loss coefficients when the water heater is off and on and are equal in value for the case of electric (as opposed to natural gas) water heaters. We thus calculate r as:

$$r = \frac{1}{UA}, \tag{2.16}$$

where UA is the loss coefficient from the .idf file. V is the "Tank Volume" field, and this lets us calculate c using the data in Table 1 once we specify a temperature T and interpolate the table values:

$$c(T) = V\rho(T)c_v(T), \tag{2.17}$$

where V is the volume of the tank, ρ is the density of water, and c_v is the specific heat of water. For calculating α we assume $T = T_{\text{sp}}$, which is simply $125°F$ for all modeled homes in the version of ResStock used for this work. $\overline{P}$ corresponds to the EnergyPlus field Heater Maximum Capacity, equal in this study to 1.3 kW or 1.6 kW.[4] Baseline power draw $\widetilde{P}$ is a single time series result readily available from the ResStock simulation.

For individual HPWHs, r, V and α are defined the same as they are for ERWHs; however, η and $\overline{P}$ are no longer constants. EnergyPlus uses performance curves to capture how η and $\overline{P}$ vary with tank and ambient temperatures. The value of $\widetilde{P}$ is also more complex than in the electric resistance case as it is the sum of a compressor and a fan component. To report the timeseries values of these quantities, EnergyPlus Energy Management System (EMS) code is inserted into the models via OpenStudio measure. We also assume the back-up electric resistance element can be ignored (i.e., we do not attempt to model the additional flexibility available from accessing both the heat pump and electric resistance elements generally available in HPWHs). In our ResStock simulations, this is a reasonable assumption because the HPWHs only use resistance back-up heat if tank temperature falls to 76°F, and their large volumes (80 gallons in all cases) make this a rare occurrence.

For ERWHs and HPWHs we always pull the baseline power draw $\widetilde{P}$ directly from the baseline (constant temperature set point of $T_{\text{sp}} = 125°\text{F}$) EnergyPlus simulations, which were conducted at 5-minute resolution. Thus $\underline{\Delta P}$ and $\overline{\Delta P}$ are directly calculated as $-\widetilde{P}$ and $\overline{P} - \widetilde{P}$ respectively.

From this point, we explored two different ways to specify the remaining parameters. For the *Simple* model, we assume $\widetilde{S}$ corresponds to a constant tank temperature of T_{sp}, whereas for the *TankT* model, we calculate $\widetilde{S}$ using the actual time-varying baseline tank temperature $\widetilde{T}(t)$ as reported by EnergyPlus. In both cases, we use the actual ambient temperature, which is also time-varying and corresponds to the Living, Garage, Finished Basement, or Unfinished Basement thermal zones in our ResStock portfolios.

Thus, for the *Simple* model:

$$\widetilde{S}(t) = c(T_{\text{sp}})\left(T_{\text{sp}} - a(t)\right) \tag{2.18}$$

$$\underline{\Delta S} = c(\underline{T})\left(\underline{T} - T_{\text{sp}}\right) \tag{2.19}$$

$$\overline{\Delta S} = c(\overline{T})\left(\overline{T} - T_{\text{sp}}\right) \tag{2.20}$$

[4] This study was completed prior to the ResStock bug fix https://github.com/NREL/resstock/pull/804.

and for the *TankT* model:

$$\widetilde{S}(t) = c(\widetilde{T}(t))\left(\widetilde{T}(t) - a(t)\right) \tag{2.21}$$

$$\underline{S}(t) = c(\underline{T})\left(\underline{T} - a(t)\right) \tag{2.22}$$

$$\overline{S}(t) = c(\overline{T})\left(\overline{T} - a(t)\right) \tag{2.23}$$

$$\underline{\Delta S}(t) = \min\left(\underline{S}(t) - \widetilde{S}(t), 0\right) \tag{2.24}$$

$$\underline{\Delta S}(t) = \max\left(\overline{S}(t) - \widetilde{S}(t), 0\right). \tag{2.25}$$

In this report, we analyze the flexibility of three types of electric water heater represented in ResStock. Two are ERWHs with different amounts of insulation (Electric Standard and Electric Premium) and each comes in five different sizes—ResStock assigns tank size based on home size, number of occupants, and amount of hot water use. The third is an 80-gallon heat pump water heater: Electric Heat Pump. The choice of an 80-gallon tank for the HPWH is an artifact of what was available in the ResStock model library at the time as opposed to an assumption about what is likely to be installed in real buildings. As discussed throughout the report, these differences in modeled ERWH and HPWH storage volumes impact our demand flexibility estimates, generally biasing them to show increased HPWH flexibility as compared to what we would expect for a real-world mix of tank volumes. From Figure 5 we see that the thermal dissipation for all of these models is less than 1% per hour, such that we should not expect the α term in Equation 2.13 to have a large impact on our results, although dissipation is about four times larger in the HPWH models than in the ERWHs.

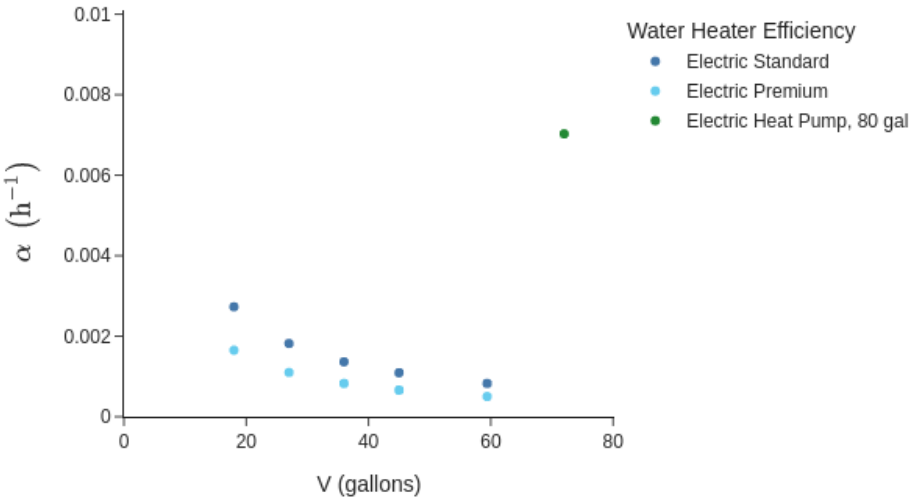


Figure 5. Water heater thermal energy dissipation versus tank volume for the ResStock water heater models used in this study

Water heater class, ERWH or HPWH, also significantly impacts η, baseline power draw, and bounds on how much the power draw can deviate from baseline. All these differences significantly impact the results we see later in the report (Chapter 4 and Chapter 5) and were previewed in Figure 3. The ERWH models take η directly from the EnergyPlus input files, which set the Heater Thermal Efficiency to 1.0. The observed distribution of η values for the HPWHs, whose interquartile range falls between 2.35 and 2.82, is shown in Figure 6.

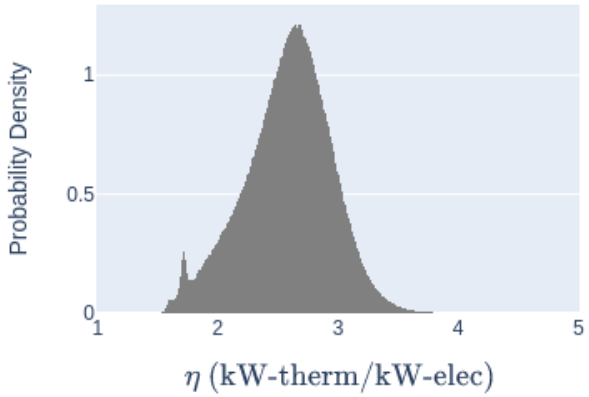


Figure 6. Distribution of HPWH electricity to thermal energy conversion efficiency, η

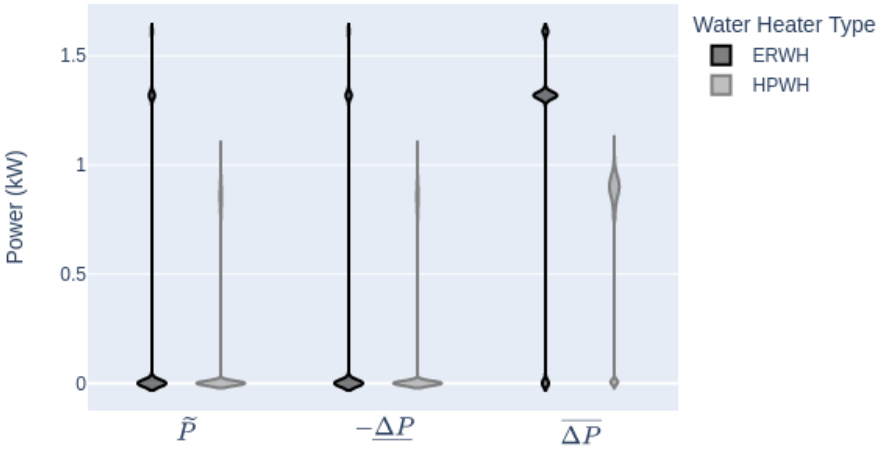


Figure 7. Distributions of time-varying baseline power draw and bounds on allowable change in power draw for ERWH and HPWH surrogate flexibility models

Figure 7 shows the distribution of baseline power draw by water heater type in the two left-most violin plots. Both ERWHs and HPWHs spend most of their time in the off-state, with $\widetilde{P}$ at or near zero. When on, ERWHs simply turn their resistance elements on, which leads to a concentration of points at 1.3 kW, the maximum power draw of the 83% of ERWHs that are of type "Electric Standard." The remaining 17% are "Electric Premium" ERWHs and pull a maximum of 1.6 kW. HPWH maximum power draw, on the other hand, varies with tank and ambient temperature, leading to a wide dispersion of on-states that tends to be less than 1 kW (the interquartile range of on-state data points is 0.81 to 0.89 kW). The distributions for $-\underline{\Delta P}$ are a simple repeat of the $\widetilde{P}$ distributions because the surrogate flexibility models can only reduce power draw when the water heaters are on under baseline conditions. $\overline{\Delta P}$ is how much the water heater power draw can be increased, which is the maximum possible power draw minus baseline power draw. For the ERWHs, the resulting distribution has its mode at 1.3 kW (corresponding to the power capacity of the majority of ERWHs) and significant density at 1.6 kW (power capacity of "Electric Premium" ERWHs) and 0 kW (value for all ERWHs when they are on). The HPWH $\overline{\Delta P}$ values factor in time-varying maximum power capacities, which yields less concentration in the distribution at the high end compared to the ERWH distribution.

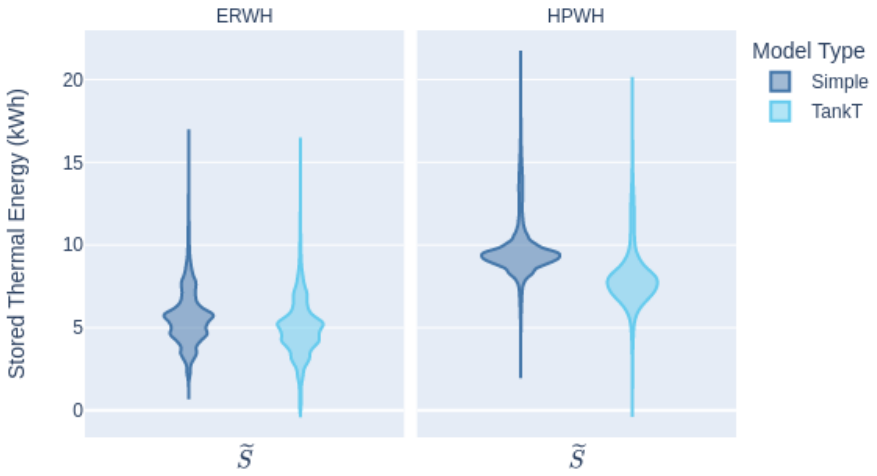


Figure 8. Distributions of the time-varying baseline thermal energy storage for ERWH and HPWH surrogate flexibility models

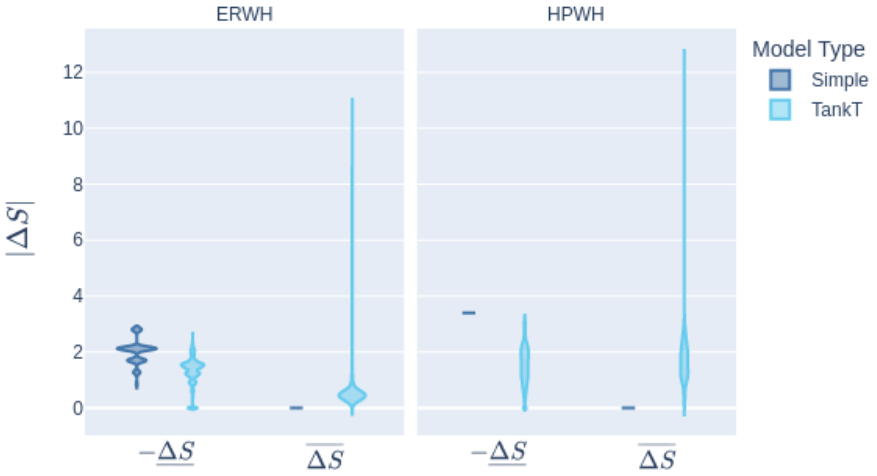


Figure 9. Distributions of the time-varying bounds on allowable change in thermal energy storage for ERWH and HPWH surrogate flexibility models

Because the two surrogate model types differ in how $\widetilde{S}$, $\underline{\Delta S}$, and $\overline{\Delta S}$ are calculated, we see differences in those parameters' distributions both when we compare the Simple and TankT models and when when we compare ERWHs and HPWHs (Figure 8 and Figure 9). Figure 8 shows distributions of the amount of thermal energy stored in the modeled ResStock water heaters under baseline conditions ($\widetilde{S}$). The three determinants of how much thermal energy is stored are (1) water heater volume, (2) ambient temperature, and (3) tank temperature. Because the HPWHs come in only one size and the tank temperature is assumed to be constant in the Simple model, the distribution shown for the Simple HPWH models only reflects the impact of different ambient temperatures. Ambient temperature varies over time for each water heater, each of which is modeled as being in a (conditioned) living zone, an (unconditioned) garage, or a basement (conditioned or unconditioned). Water heaters in conditioned spaces see less variation in ambient temperature than those in unconditioned spaces; across the portfolio, placement of water heaters in different types of thermal zones, outside weather, and indoor air temperature all contribute to the width of the HPWH-Simple $\widetilde{S}$ distribution.

The ERWH-Simple $\widetilde{S}$ distribution is essentially multiple single-volume distributions superimposed on each other. This and the smaller volume of the ERWHs relative to the HPWHs generally explains the differences between the ERWH-Simple and HPWH-Simple $\widetilde{S}$ distributions. The difference between the Simple models and the TankT models is that the former assumes the tank temperature is always at set point (here, 125°F) while the latter uses the actual tank temperature as reported by EnergyPlus. Because the EnergyPlus models use a constant set point with a dead band that only turns the heater on if the temperature falls 10°F below the set point (i.e., to 115°F) and turns the heater off when the set point is reached, the Simple model assumption is effectively an upper bound, and the distinct reduction in stored thermal energy in the TankT as compared to the Simple models reflects how often and how much the tank temperature is below set point in these simulations. This holds for both the ERWH and HPWH models. For all the distributions shown in Figure 8, the largest stored thermal energy values correspond to large-volume tanks at times and places in which the ambient temperatures are lowest (e.g., water heaters in unconditioned spaces experiencing very cold winter weather) while the tank temperatures are high (e.g., at set point). The smallest values correspond to times when the tank temperatures are very low relative to ambient temperature—for limited periods of time, the tank temperature can even be lower than the ambient temperature, which results in stored thermal energy values below zero, as the tank is recharged from the water mains, which can be colder than the air surrounding the water tank.

Because the surrogate models depicted here were developed for estimating response to load-shedding contingency events, the maximum temperature used to create thermal energy bounds was the same as the nominal set point (125°F). The lower tank temperature bound was 105°F. Thus, the distributions for $-\underline{\Delta S}$ shown in Figure 9 reflect the difference in thermal energy storage between 125°F and 105°F in the Simple surrogate models, and between the actual tank temperature (or 105°F, whichever is higher) and 105°F in the TankT surrogate models. The distributions of upper bound values, $\overline{\Delta S}$ (also shown in Figure 9), instead reflect the difference in thermal energy storage between 125°F and the tank temperature set point (that is, zero) and between 125°F and the actual tank temperature in the Simple and TankT models respectively. Thus, for the Simple models, the bounds on ΔS are constant per water heater and only vary between water heaters if water heaters have different tank volumes. On the other hand, the ΔS bounds in the TankT models are time-varying; the magnitude of the $\underline{\Delta S}$ bounds are always smaller in the TankT than in the Simple surrogate models, and the magnitude of the $\overline{\Delta S}$ bounds are correspondingly larger.

3 Case Study Setting

To explore aggregated water heater flexibility, we focus on technical potential resource size and the associated grid service value in New England, whose power system is operated by the independent system operator (ISO) ISO-New England (ISO-NE) (Figure 10).

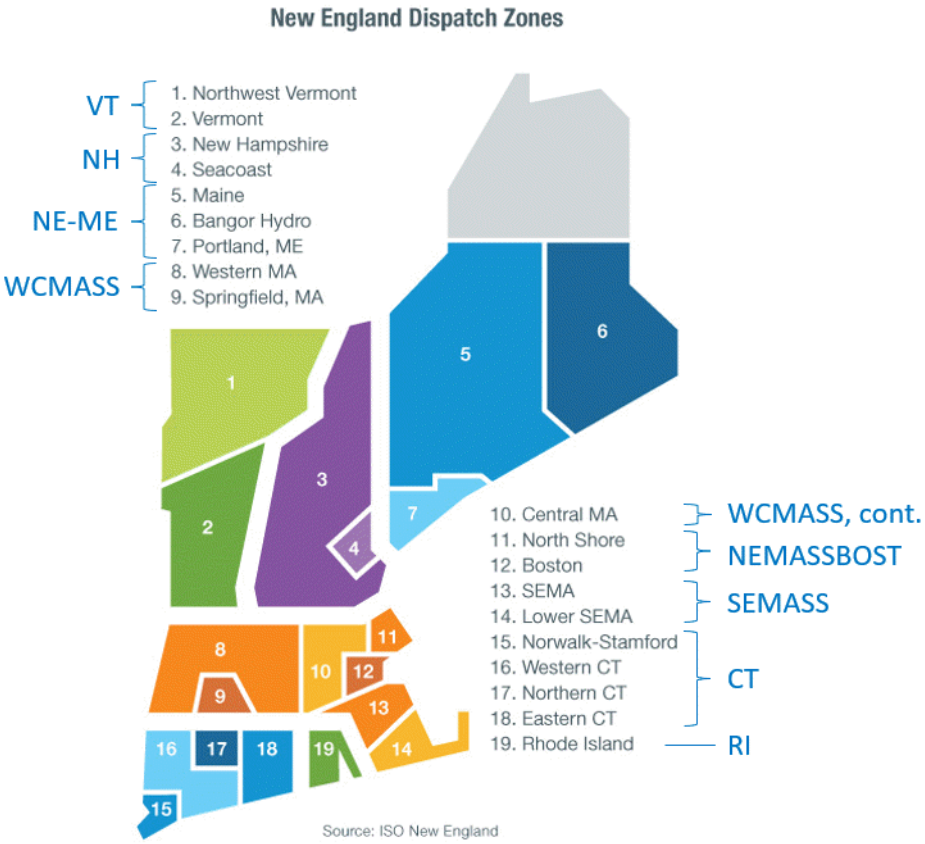


Figure 10. ISO-NE dispatch zones and load regions. The eight load regions are the indicated aggregations of 1 to 4 dispatch zones.

On the demand-side, our starting point is ResStock run with an overall sampling rate of one single-family home simulated per 228.6 actual single family homes. The run is filtered down to just the counties in New England and homes with electric water heaters. This results in about 2,700 samples each with sample weight 228.6 and assigned to a specific New England county.

On the supply-side, our representation of ISO-NE is a PLEXOS production cost model extracted from the SEAMS Study (Bloom et al. 2021; Energy Exemplar). Production cost models simulate grid operations by computing the minimum cost dispatch of grid resources that satisfies demand and provides sufficient reserve capacity to operate reliably at all times. Emulating real-world grid operations, production cost models are often run at multiple time scales and resolutions. For example, hourly day-ahead runs compute optimal dispatch for the following day and account for the unit commitment constraints of large generators, e.g., minimum generation levels, minimum up times, and minimum down times; while sub-hourly real-time runs compute optimal dispatch for the next 5 to 15 minutes and fix most unit commitment variables because most generators cannot start up or shut down that quickly. Production cost models typically run through a complex series

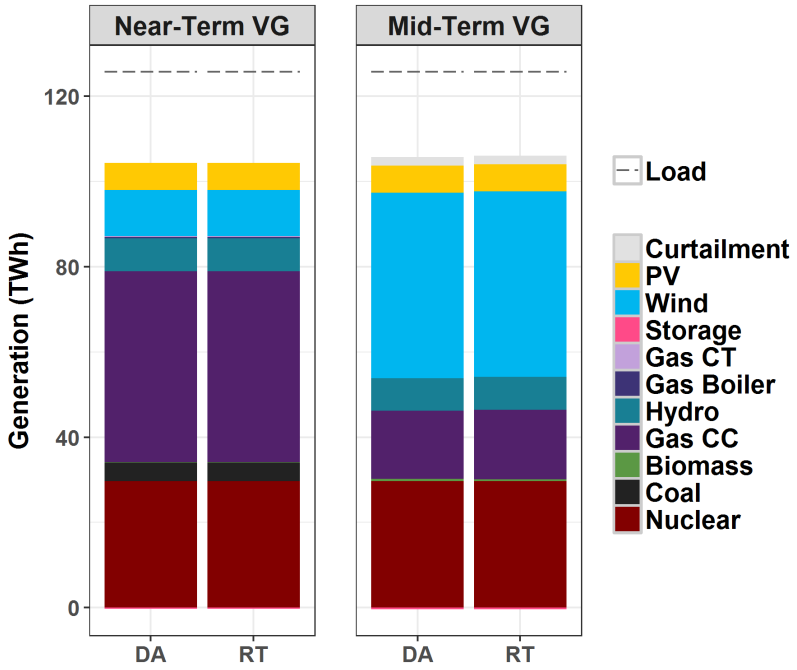


Figure 11. Annual generation in the Near-Term VG and Mid-Term VG PLEXOS models of ISO-NE before adding demand response resource from electric water heaters. Net imports serve the gap between generation and load. Storage is used to shift some of the generation shown here to other times, subject to round-trip efficiency losses.

of these types of optimization problems to simulate an entire year of grid operations, thereby quantifying a power system's annual production costs and its behavior across all seasons.

In this study, there are two production cost model scenarios, one representing a projected "2024" system and the other representing a projected "2038" system. In the remainder of this paper we refer to these scenarios as the "Near-Term VG" and "Mid-Term VG" scenarios to emphasize that they include different amounts of wind and solar generation, the two major forms of variable generation (VG) (Figure 11), but utilize the same load profiles. Some basic statistics on the scenarios are provided in Table 2. ISO-NE is well-connected with its neighbors. We balance computational tractability and realism by only modeling ISO-NE while representing connections to neighboring regions with fixed import and export schedules that were determined by running a version of the entire SEAMS model for each scenario. We additionally incorporate a high-priced generator at each of the interconnection points, to help the model deal with changes that occur in the power flow from the various resources we add throughout this report.

The water heaters modeled in ResStock are mapped into the ISO-NE power system by:

1. Computing the fraction of each county's residential utility customers that are in each census tract (U.S. DOE EERE 2019). There are 67 counties and 3,353 census tracts in New England.

[5]Zero and low carbon, comprising PV, wind, hydro, and nuclear generation
[6]PV, wind, and hydro generation
[7]PV and wind generation

Table 2. ISO-NE Production Cost Model Summary

	Near-Term VG	Mid-Term VG
Clean[5] Generation Share (%)	52	82
RE[6] Generation Share (%)	24	56
VG[7] Generation Share (%)	16	49
Curtailment (%)	0.02	3.9
Production Costs (Million $)	2400	1100
Unserved Reserve (GWh)	3.2	0.04

2. Determining the closest PLEXOS model transmission load node to each census tract. There are 783 load nodes in the 67 New England counties.
3. Combining those two maps to determine what fraction of a ResStock building's sample weight (all of which has already been allocated to a single county) should be assigned to each transmission node.

The water heater resources can then be represented at the nodal level, or at the more-aggregated dispatch zone or load region levels. The 19 ISO-NE dispatch zones are directly shown in the Figure 10 map. The 8 load regions are direct aggregations of dispatch zones as shown by the light blue text and brackets in that same figure.

We evaluate water heater flexibility in the context of two grid services, one that sheds load in times of grid stress and another that shifts load from higher-price to lower-price times.

Although demand response is most typically used to shed load during system peak times, that is, during the hottest afternoons (for summer peaking systems) or the coldest mornings (for winter peaking system) of the year, shed-demand response (DR) can also serve year-round as contingency reserves. In this case, a certain amount of demand response along with other resource capacity (i.e., generation or storage) is reserved by the system operator at all times but is only called upon to reduce load (or increase generation) in the case that there is an unexpected generation or transmission outage. Typical contingency events only require response for 10's of minutes to an hour (Denholm, Sun, and Mai 2019). ISO-NE has codified this for demand response assets (DRAs) by enabling them to sign up and then be audited for performance-based participation in their Claim10 and Claim30 programs (ISO New England Inc. 2022; Nichols and Haag 2018).

In this study, we model contingency response based on descriptions of ISO-NE's Claim10 and Claim30 auditing process (Nichols and Haag 2018). Claim10 is shorthand for a 10-minute notification service, that is, DRAs performing this service have 10 minutes from the time of event notification to ramp their response up to their target output. They are then expected to maintain their target output for 50-minutes, so that the whole auditing process (from notification to the end of response) is one hour. Claim30 is audited similarly, but with 30-minute notification. The total audit time is still one hour, that is, Claim30 resources are expected to ramp up to their target output by the 30-minute mark and then to maintain that level of output over a 30-minute response time.

We model demand response shifting service on an economic basis. That is, we either endogenously dispatch electric water heater surrogate flexibility models in the PLEXOS production cost model as a no-variable cost storage-like resource, or we dispatch surrogate flexibility models in a price-taking sense against the day-ahead locational marginal prices output by PLEXOS. In both cases we can directly observe the dispatch profiles. We can also observe and compare changes in generator dispatch and production costs by applying the endogenous and price-taking dispatch day-ahead profiles to the PLEXOS real-time model.[8] The individual water heater surrogate models are the same as what is used for the contingency analysis and described in Section 2.2 except that the upper tank temperature bound is set to 145°F such that the overall allowable tank temperature range is 125 ± 20°F.

[8]Compared to the day-ahead problems, the real-time problems fix the commitment status of longer start time thermal generators, and fix the dispatch of pumped hydro and water heater loads.

4 Contingency Service

To determine the technical potential of electric water heaters providing contingency reserve in New England, we simulate contingency events using the individual flexibility surrogate models described in Section 2.2 and aggregate flexibility models that are in the same mathematical form but whose parameter values are determined using the aggregation mathematics described below and in Appendix A. We apply the validation step of comparing surrogate model-simulated Claim10 and Claim30 responses to EnergyPlus-simulated Claim10 and Claim30 responses to (1) choose either the Simple or the TankT individual surrogate model form and (2) inform parameter values used in the aggregation steps. Having completed these steps, we compute final surrogate models, estimate hourly contingency reserve resource at different levels of aggregation, and add these resources to our ISO-NE production cost models.

4.1 Methods

4.1.1 Aggregated Flexibility

We can make two naïve assumptions to estimate the aggregate contingency reserve resource available from electric water heaters. First, we can assume contingency responses are short enough that all water heater load is sheddable for up to 1 hour. In this case, the average hourly water heater load (e.g., as depicted in Figure 3) becomes our estimate of contingency resource. Second, we can create individual flexibility models of the form Equation 2.13 - Equation 2.15 or discretized with time-step δt and represented per water heater $k \in \mathbb{K}$ as:

$$\Delta S_k(t+\delta t) = (1-\alpha_k(t)\cdot\delta t)\,\Delta S_k(t) + \eta_k(t)\Delta P_k(t)\cdot\delta t \tag{4.1}$$

$$\underline{\Delta S_k}(t) = c_k(t)\left(\underline{T_k}(t) - \widetilde{T}_k(t)\right) \le \Delta S_k(t) \le c_k(t)\left(\overline{T_k}(t) - \widetilde{T}_k(t)\right) = \overline{\Delta S_k}(t) \tag{4.2}$$

$$\underline{\Delta P_k}(t) = -\widetilde{P}_k(t) \le \Delta P_k(t) \le \overline{P_k}(t) - \widetilde{P}_k(t) = \overline{\Delta P_k}(t). \tag{4.3}$$

and then simply sum the individual bounds while also applying sample weights w_k to compute an aggregate model where $L(t)$ performs the role of ΔS and $U(t)$ performs the role of ΔP:

$$L(t+\delta t) = \left(1 - \frac{\sum_k w_k\alpha_k(t)}{\sum_k w_k}\cdot\delta t\right)L(t) + \frac{\sum_k w_k\eta_k(t)}{\sum_k w_k}U(t)\cdot\delta t \tag{4.4}$$

$$\underline{L}(t) = \sum_k w_k\underline{\Delta S_k}(t) \le L(t) \le \sum_k w_k\overline{\Delta S_k}(t) = \overline{L}(t) \tag{4.5}$$

$$\underline{U}(t) = \sum_k w_k\underline{\Delta P_k}(t) \le U(t) \le \sum_k w_k\overline{\Delta P_k}(t) = \overline{U}(t). \tag{4.6}$$

As shown below, it turns out that these estimates are often true *overestimates* of the actual aggregate resource. That is, the individual load reductions implied by shedding all water heater load are not all sustainable when lower bounds on tank temperature are accounted for, and the model created by simply summing all the individual bounds (shown above and referred to as the *outer approximation* in Appendix A and what follows) can incorrectly pair the ability of some water heaters to shed load with the ability of other water heaters to tolerate reductions in their tank temperatures to essentially ignore individual water heater constraints.

As described in Appendix A, if the model parameters η_k and α_k are constant, it is possible to construct an *inner approximation* of the aggregate resource that is provably dispatchable. The construction relies on assigning a fraction β_k to each resource in the aggregation and then distributing any aggregate dispatch request to those resources according to that fraction. To ensure dispatchability, the aggregate flexibility model parameters are assigned in a worst case sense, that is, the aggregate bounds $\underline{L}$, $\overline{L}$, $\underline{U}$ and $\overline{U}$ are set so that:

$$\underline{\Delta S_k}(t) \leq \beta_k \underline{L}(t) \leq \Delta S_k(t) \leq \beta_k \overline{L}(t) \leq \overline{\Delta S_k}(t) \; \forall k \in \mathbb{K} \tag{4.7}$$

and

$$\underline{\Delta P_k}(t) \leq \beta_k \underline{U}(t) \leq \Delta P_k(t) \leq \beta_k \overline{U}(t) \leq \overline{\Delta P_k}(t) \; \forall k \in \mathbb{K}. \tag{4.8}$$

As described in Appendix A, η_k and α_k are also tracked properly and used to define η and α parameters for the aggregate model to ensure overall dispatchability. The inner approximations are tunable in the sense that the β_k can be chosen to maximize:

$$w\,|\underline{L}| + (1-w)\eta\,|\underline{U}|\,, \tag{4.9}$$

where here we focus on the ability to reduce stored thermal energy and power draw because those are the capabilities of interest when examining contingency response.

In what follows, aggregations can be either an outer approximation or an inner approximation with energy weight w. An inner approximation with $w = 1$ is called an "inner_maxS" approximation whereas an inner approximation with $w = 0$ is called an "inner_maxP" approximation. For w values between 0 and 1 we use notation like "inner_w0p5" where the p stands in for the decimal point. Aggregation can also be done at different geographic scales. For contingency resource we aggregate all water heaters that map either to the same node, the same dispatch zone, or the same load region, which results in 783, 19, or 8 aggregate flexibility models respectively.

4.1.2 Dispatch of Surrogate Flexibility Models

Whether we are working with individual water heater or aggregate surrogate flexibility models, we evaluate contingency reserve resource by simulating Claim10 and Claim30 audit events using a simple dispatch algorithm. We run the dispatch algorithm in two modes: "Greedy", which always tries to reduce load as much as possible now without anticipating how much energy capacity will be left to sustain the response and "Flat", which tries to produce a sustained, flat response while respecting both power and energy bounds.

The dispatch algorithm expects to simulate an event consisting of three periods: starting notification, response, and rebound. In all cases, we take the duration of the notification plus the response period to be one hour, and the rebound period to be one hour. Then, Claim10 and Claim30 only differ in that Claim10 has a 10-minute notification period and a 50-minute response period whereas Claim30 has a 30-minute notification period and a 30-minute response period. The dispatch algorithm does nothing during the notification period, maximizes reductions in load (greedily or with a flat profile) during the response period, and gets back to $\Delta S = 0$ as quickly as possible (and if possible) during the rebound period.

The dispatch algorithm is shown in Algorithm 1. The dispatch simulation uses 5-minute timesteps. If the algorithm is supplied with 1-hour data, the data are interpolated to 5 minutes before

running the event simulation. After the dispatch algorithm is run, we have the resulting ΔP and ΔS profiles.

4.1.3 Computation of Contingency Reserve Resource

Given a surrogate flexibility model at hourly resolution and for a whole year, we estimate Claim10 or Claim30 resource by simulating a response in each hour of the year (per Algorithm 1 with N_{start} and N_{response} set properly for the Claim type) and estimate each hour's contingency resource to be the average response during the response period for that hour. We then multiply the total resource by 1.053 to account for a 5.3% distribution loss factor (Cohen et al. 2019). That is, we estimate that a 1 kWh reduction in load behind-the-meter reduces bulk power system load by 1.053 kWh. If we estimate contingency resource for a collection of aggregate resources that together cover all of ISO-NE, the total amount of resource is just the sum over all the aggregate models.

Algorithm 1 Contingency Event Simulation

function COMPUTE_STEP($DS, DP, \eta, \alpha, \delta t$)
    return $DS + \eta \cdot DP \cdot \delta t - \alpha \cdot DS \cdot \delta t$
end function

function COMPUTE_POWER($DS, DS\_target, \eta, \alpha, \delta t$)
    return $(DS\_target - DS + \alpha \cdot DS \cdot \delta t)/(\eta \cdot \delta t)$
end function

procedure DISPATCH(dtype, $\delta t, N_{\text{start}}, N_{\text{response}}, N_{\text{rebound}}, \underline{\Delta P}(t), \overline{\Delta P}(t), \underline{\Delta S}(t), \overline{\Delta S}(t), \eta(t), \alpha(t)$)
    ▷ dtype can be "Flat" or "Greedy". Greedy sets $DP(t)$ to $\underline{\Delta P}(t)$ if feasible.

    ▷ Compute response target if Flat dispatch
    if dtype == Flat **then**
        $DP\_target \leftarrow \max\left(\underline{\Delta P}(i \cdot \delta t) | N_{\text{start}} \leq i < N_{\text{start}} + N_{\text{response}}\right)$ ▷ Vector indices start at 0
        for $i \in \left[N_{\text{start}} + 1, N_{\text{start}} + 2, \ldots, N_{\text{start}} + N_{\text{response}}\right]$ **do**
            $DP\_target\_trial \leftarrow \underline{\Delta S}(i \cdot \delta t)/\left(\delta t \cdot (i - N_{\text{start}})\right)$
            if $DP\_target\_trial > DP\_target$ **then**
                $DP\_target \leftarrow DP\_target\_trial$
            end if
        end for
    end if

    ▷ Initialize outputs
    $DS(t) \leftarrow 0$
    $DP(t) \leftarrow 0$

    ▷ Simulate response
    for $i \in \left[N_{\text{start}}, N_{\text{start}} + 1, \ldots, N_{\text{start}} + N_{\text{response}} - 1\right]$ **do**
        if dtype == Flat **then**
            $DP(i \cdot \delta t) \leftarrow DP\_target$
        else
            $DP(i \cdot \delta t) \leftarrow \underline{\Delta P}(i \cdot \delta t)$
        end if
        $DS((i+1) \cdot \delta t) \leftarrow$ COMPUTE_STEP$(DS(i \cdot \delta t), DP(i \cdot \delta t), \eta(i \cdot \delta t), \alpha(i \cdot \delta t), \delta t)$
        if $DS((i+1) \cdot \delta t) < \underline{\Delta S}((i+1) \cdot \delta t)$ **then**
            $DP(i \cdot \delta t) \leftarrow$ COMPUTE_POWER$(DS(i \cdot \delta t), \underline{\Delta S}((i+1) \cdot \delta t), \eta(i \cdot \delta t), \alpha(i \cdot \delta t), \delta t)$
            if $DP(i \cdot \delta t) > \overline{\Delta P}(i \cdot \delta t)$ **then**
                ▷ Issue infeasibility warning–$\underline{\Delta S}$ bound will be violated
                $DP(i \cdot \delta t) \leftarrow \overline{\Delta P}(i \cdot \delta t)$
                $DS((i+1) \cdot \delta t) \leftarrow$ COMPUTE_STEP$(DS(i \cdot \delta t), DP(i \cdot \delta t), \eta(i \cdot \delta t), \alpha(i \cdot \delta t), \delta t)$
            else
                $DS((i+1) \cdot \delta t) \leftarrow \underline{\Delta S}((i+1) \cdot \delta t)$
            end if
        end if
    end for

▷ Simulate rebound
for $i \in \left[N_{\text{start}} + N_{\text{response}}, N_{\text{start}} + N_{\text{response}} + 1, \ldots, N_{\text{start}} + N_{\text{response}} + N_{\text{rebound}} - 1\right]$ **do**
  if $DS(i \cdot \delta t) == 0$ **then**
    ▷ Achieved goal of getting to no deviation from baseline energy storage
    BREAK
  end if
  $DS((i+1) \cdot \delta t) \leftarrow$ COMPUTE_STEP$(DS(i \cdot \delta t), \overline{\Delta P}(i \cdot \delta t), \eta(i \cdot \delta t), \alpha(i \cdot \delta t), \delta t)$
  if $DS((i+1) \cdot \delta t) > 0$ **then**
    $DP(i \cdot \delta t) \leftarrow$ COMPUTE_POWER$(DS(i \cdot \delta t), 0, \eta(i \cdot \delta t), \alpha(i \cdot \delta t), \delta t)$
    if $DP(i \cdot \delta t) < \underline{\Delta P}(i \cdot \delta t)$ **then**
      ▷ Issue warning about overshooting desired level of energy storage
      $DP(i \cdot \delta t) \leftarrow \underline{\Delta P}(i \cdot \delta t)$
      $DS((i+1) \cdot \delta t) \leftarrow$ COMPUTE_STEP$(DS(i \cdot \delta t), DP(i \cdot \delta t), \eta(i \cdot \delta t), \alpha(i \cdot \delta t), \delta t)$
    else
      $DS((i+1) \cdot \delta t) \leftarrow 0$
    end if
  else
    $DP(i \cdot \delta t) \leftarrow \overline{\Delta P}(i \cdot \delta t)$
  end if
end for
end procedure

4.2 Validation

We test the individual surrogate flexibility models and their aggregates by simulating their Claim10 and Claim30 responses and comparing them to EnergyPlus responses created by lowering water heater tank temperature set points during the response periods.

4.2.1 EnergyPlus "Ground Truth"

Claim10 and Claim30 responses are simulated with ResStock by injecting a DR Event Schedule and using it to trigger water heater tank temperature set point reductions. The DR Event Schedule specifies a DR Event to happen every day of the year at the same, top-of-the-hour local time each day. Because each event is separated by 24 hours, the water heaters settle back to baseline conditions between events. Each event is evaluated by taking time series data from ResStock portfolio buildings that include the events and subtracting the corresponding time series data from the baseline portfolio buildings. The main time series of interest are power and tank temperature profiles. Successful execution of this procedure requires having baseline and with-event portfolios that are otherwise identical.

We simulated a total of six "claim" runs, in addition to a baseline run, for two different ResStock portfolios. The first ResStock portfolio was the default ResStock portfolio filtered down to New England homes with electric water heaters. At that point a post-processing step further selected just the ERWHs with tanks (the default portfolio contains a few tankless models as well as a few heat pump water heaters). The second ResStock portfolio was also filtered to just include New England homes with electric water heaters, but then all of the existing electric water heaters were replaced by an "Electric Heat Pump, 80 gal," using ResStock upgrade logic.

Three different claim "times" were simulated: random, 2 p.m., and 6 p.m. The random claim runs choose a random hour for the events to occur for each ResStock sample building; that is, each sample building's "random" events start at the top of the same local-time-hour each day, but the hour chosen is different from building to building. In this way, we can check the degree to which our surrogate models can or cannot match EnergyPlus dispatch of a contingency-like event at all times of day. The 2 p.m. and 6 p.m. claim simulations are synchronized across all the buildings. This enables validation of aggregate responses.

The resulting ResStock runs and brief descriptions of how they factor into our analysis are summarized in Table 3.

4.2.2 Individual Surrogate Model Validation

Some examples of Claim10 responses for individual ERWHs are shown in Figure 12. The middle and rightmost columns demonstrate that the Simple and TankT surrogate models are both able to replicate EnergyPlus responses in many cases, especially when dropping all baseline load ($\widetilde{P}$) during the response period (shaded in light grey) does not result in tank temperature set point violations. However, on the left we see a more challenging event during which the water heater would normally be running during the whole response period in part to make up for water draws during that time. In this case, the baseline average tank temperature hovers around 110°F for the whole period despite the temperature set point of 125°F and constant water heater operation. Because the Simple model does not see the low tank temperature, it estimates that all the load can be shed; but the TankT model anticipates the water heater will need to run some during the

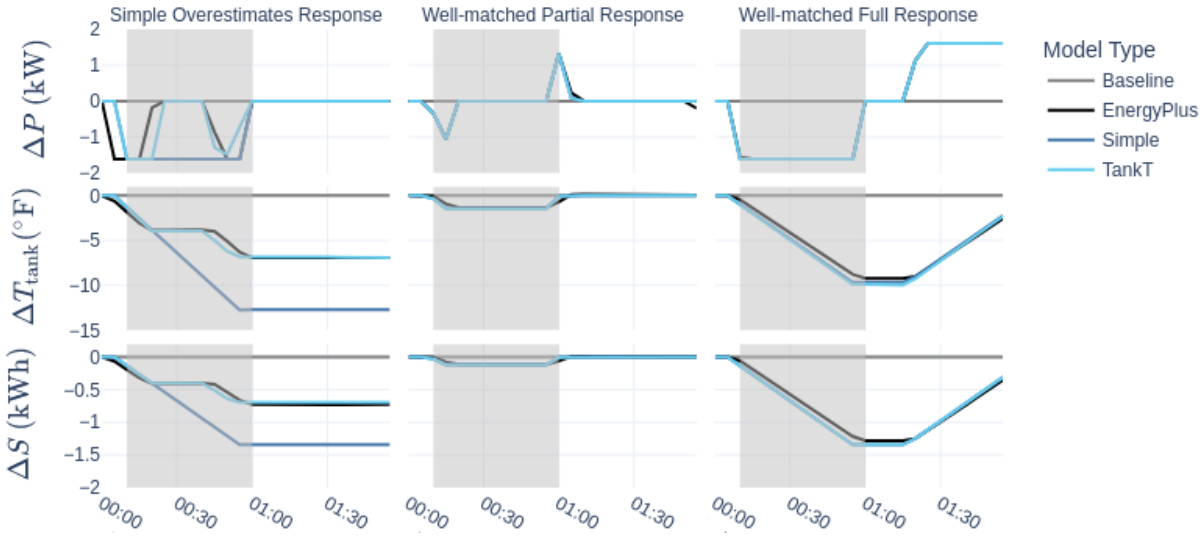
Table 3. ResStock Portfolios Used to Create Baseline and Simulated Contingency Event Data in EnergyPlus. All portfolios were filtered to New England homes with electric water heaters. A total of 14 portfolios are used in this analysis–all portfolio types were produced for both water heater types.

Water Heater Type	Portfolio Types	Portfolio Description
ERWH	Baseline	Data for estimating flexibility models and comparison point for Claim cases
HPWH	Claim10–Random	ISO-NE Claim10 audit-like events simulated at all times of day; validation of individual models
	Claim30–Random	ISO-NE Claim30 audit-like events simulated at all times of day; validation of individual models
	Claim10–2 p.m.	ISO-NE Claim10 audit-like events simulated at 2 p.m. for all buildings; validation of aggregate response
	Claim30–2 p.m.	ISO-NE Claim30 audit-like events simulated at 2 p.m. for all buildings; validation of aggregate response
	Claim10–6 p.m.	ISO-NE Claim10 audit-like events simulated at 6 p.m. for all buildings; validation of aggregate response
	Claim30–6 p.m.	ISO-NE Claim30 audit-like events simulated at 6 p.m. for all buildings; validation of aggregate response

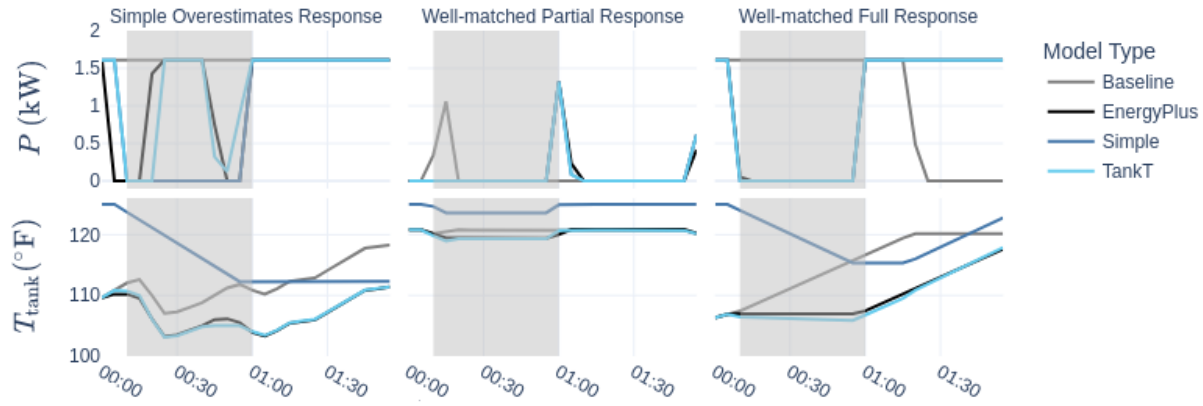
response period to respect the lowered tank temperature set point. The resulting TankT response mimics the actual EnergyPlus response fairly well, although not exactly.

The story for Claim30 responses (Figure 13) is similar, except that in the "Simple Overestimates Response" column we see that while the TankT model again performs better than the Simple model, in this case the EnergyPlus control logic reacts more strongly to low tank temperatures than the TankT model predicts. In responses like this and in what follows for HPWHs, we are reminded that to accurately predict individual responses the details matter. Because we are modeling the ERWHs using average tank temperatures, but the simulated tanks are stratified with one heating element and an associated temperature measurement near the bottom of the tank and another measurement-heating element pair near the top, the TankT model is still missing important details about the state of the tank and how it is controlled. In this case, the "real-world" EnergyPlus response is more muted than our models expect whereas in other cases our relatively simple models are sufficient to describe behavior under DR event conditions.

Across all responses, we see that the TankT surrogate model better replicates the EnergyPlus average reduction in load during the response period (reported as average kilowatts for each water heater and each response) than does the Simple surrogate model (Figure 14). However, while the number of full responses (around 1.3 kW and 1.6 kW) is less in the TankT models than the Simple models, they are still more than EnergyPlus shows. Overall, the EnergyPlus distribution is wider and more spread throughout the 0.1 kW to 0.8 kW region than the distributions for the

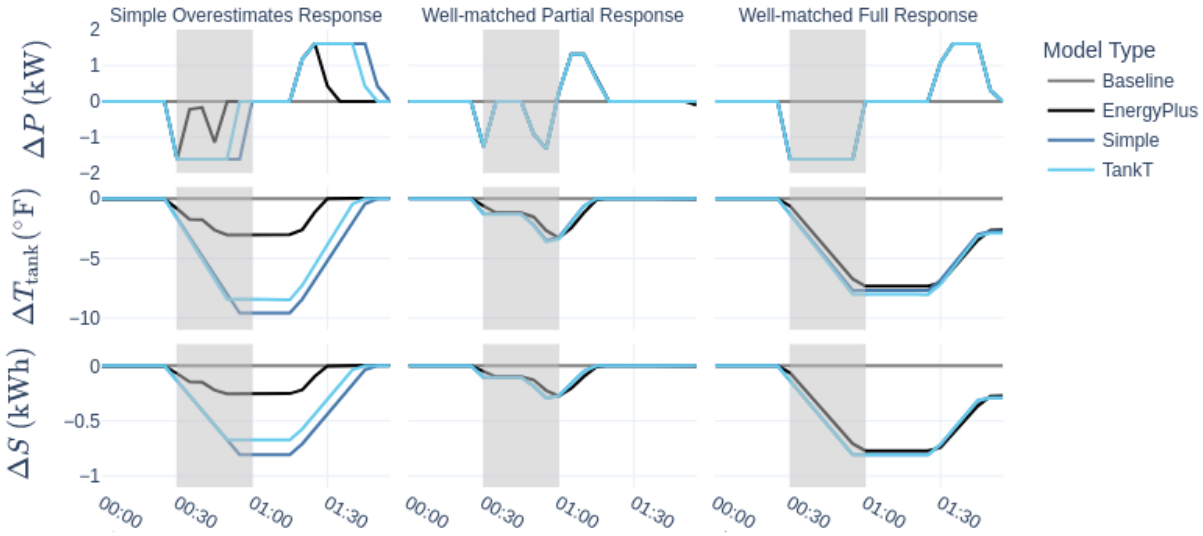


(a) Simulated event responses (differences from baseline)



(b) Simulated power draws and tank temperatures

Figure 12. Examples of individual ERWH Claim10-Random responses simulated with EnergyPlus and surrogate flexibility models.



(a) Simulated event responses (differences from baseline)

(b) Simulated power draws and tank temperatures

Figure 13. Example of individual ERWH Claim30-Random responses simulated with EnergyPlus and surrogate flexibility models.

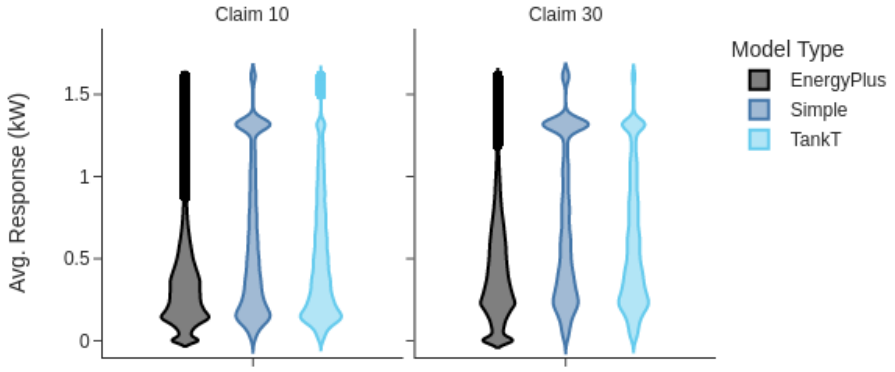


Figure 14. Distributions of individual ERWH average responses during Claim10-Random and Claim30-Random events: EnergyPlus, Simple surrogate models, and TankT surrogate models. Only data from non-null responses are included.

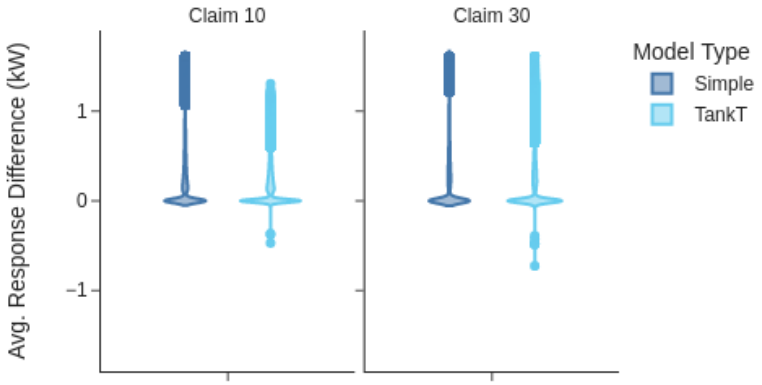


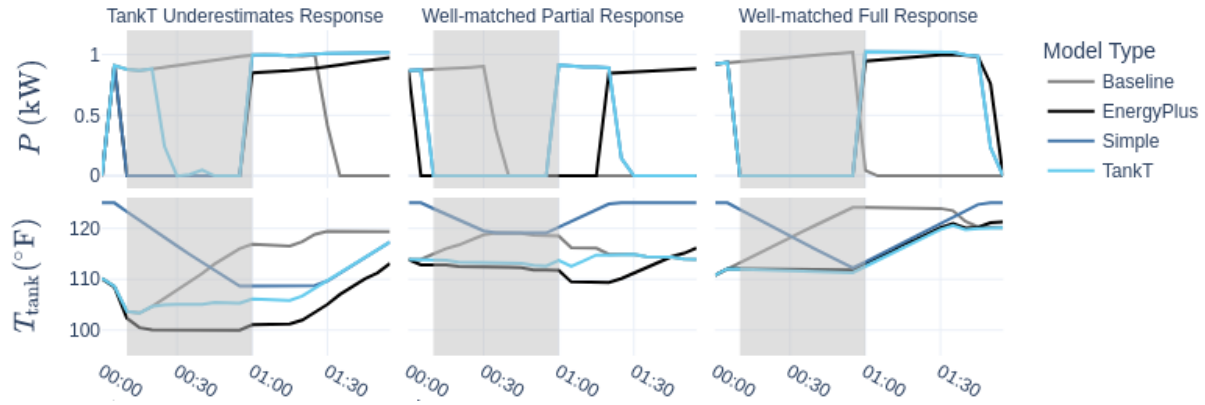
Figure 15. Distributions of differences between surrogate model and EnergyPlus individual ERWH average responses during Claim10-Random and Claim30-Random events. Only data from non-null responses are included.

surrogate models, and this holds for both Claim10 and Claim30 responses (though it is more pronounced for the longer Claim10 events).

Looking directly at surrogate model average responses minus those from EnergyPlus (Figure 15), we see confirmation of the surrogate models' biases toward overestimating response. However, we also see that the modes of the difference distributions are around zero and that the TankT model has some instances of underpredicting response relative to EnergyPlus. Overall, these results further confirm that although the TankT model is not perfect, it does a significantly better job than the Simple model in estimating EnergyPlus ERWH contingency responses.

The example HPWH individual responses shown in Figure 16 and Figure 17 demonstrate that EnergyPlus simulated HPWHs behave fairly differently than EnergyPlus ERWHs during the prescribed claim events. Focusing first on the T_{tank} row of the Claim10 plots in Figure 16b, we see the HPWH average tank temperatures do sometimes get to 125°F (Baseline profile, "Well-matched Full Response" column) but they may also turn off before the average tank temperature reaches that point ("TankT Underestimates Response" column, T_{tank} and P Baseline profiles). Then, we see in the "TankT Underestimates Response", ΔP facet (Figure 16a) that the HPWHs are more likely than the ERWHs to provide a full response even if their baseline average tank

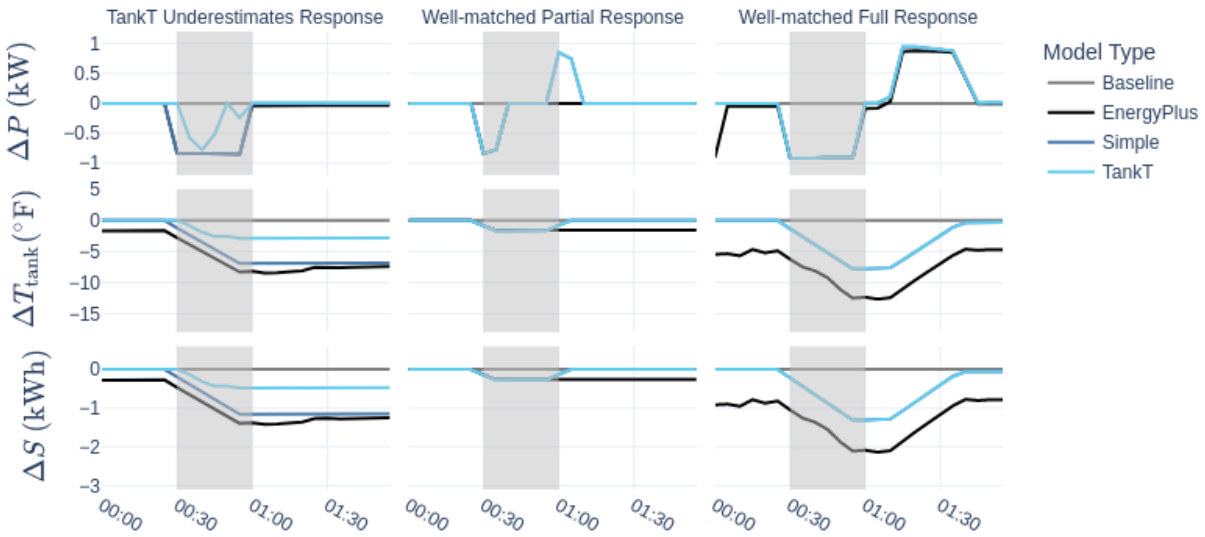
(a) Simulated event responses (differences from baseline)



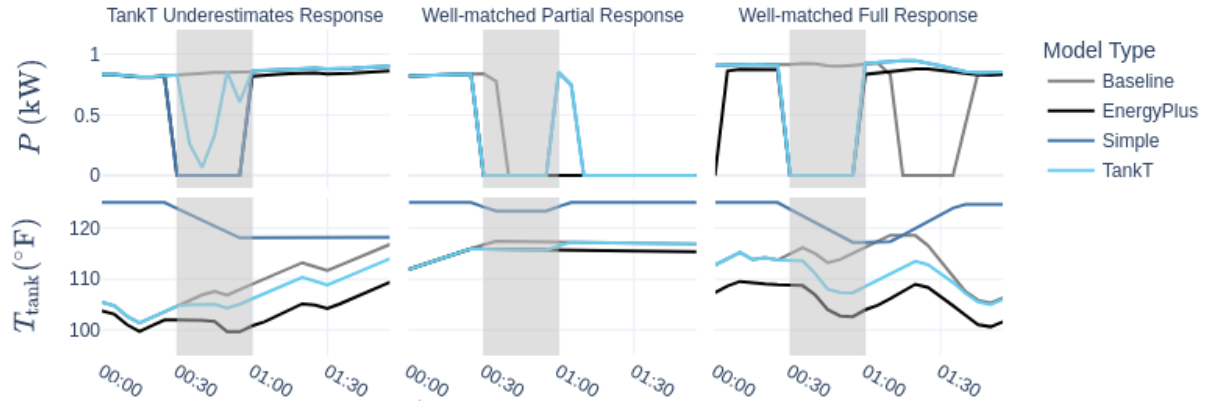
(b) Simulated power draws and tank temperatures

Figure 16. Example of individual HPWH Claim10-Random responses simulated with EnergyPlus and surrogate flexibility models.

temperatures (Figure 16b) are fairly low. Indeed, we find that for this study's HPWHs the Simple surrogate models often better represent what EnergyPlus does during a contingency event than the TankT models do. This fact is initially surprising, because on its face the Simple model seems guaranteed to be an overstatement of actual flexibility (since the tank temperature is not actually always at set point). However, average tank temperature is not the exact measurement the thermostat is responding to, and is thus an imprecise indicator of expected control behavior. In this particular case, it turns out that the TankT models are overly conservative in predicting ability to respond during the prescribed claim events–the HPWHs are typically able to delay turning on even if the average tank temperature is below the lowered set point of 105°F. As for why, in addition to the 10°F controller dead band that is included in all of the EnergyPlus ERWH and HPWH models, it is likely that the HPWH control temperature is often considerably higher than the average tank temperature. This is because EnergyPlus weighs the HPWHs' upper element temperature more heavily (at a ratio of 3:1) than the lower element temperature when determining if the heat pump needs to cycle. The large storage volumes (80 gallons) assumed in this study likely amplify this effect by buffering any hot water draws and inducing more temperature stratification.



(a) Simulated event responses (differences from baseline)



(b) Simulated power draws and tank temperatures

Figure 17. Example of individual HPWH Claim30-Random responses simulated with EnergyPlus and surrogate flexibility models.

The Claim30 "TankT Underestimates Response" shown in Figure 17 is an even more dramatic example of the Simple model aligning well with the EnergyPlus response while the TankT model is overly conservative. Another key difference between the surrogate models and EnergyPlus is visible in the ΔP rebounds (change in power draw after the response period). Generally in this plot and in Figure 16 the surrogate and EnergyPlus rebound sizes are similar, but EnergyPlus sometimes delays or foregoes recovery relative to the surrogate models ("Well-matched Partial Response" in Figure 16 and Figure 17).

The fairly good alignment between the Simple HPWH surrogate models and the EnergyPlus HPWH responses visible in the individual response plots holds up across the entire set of responses. In Figure 18, this shows up as EnergyPlus and the Simple surrogate models having similar distributions of average responses, in contrast to the TankT model predicting smaller and more-often-zero responses on average. Turning to the distributions of response differences (Figure 19), we see that the Simple model is generally aligned with the EnergyPlus contingency response estimates while the TankT model systematically underestimates response.

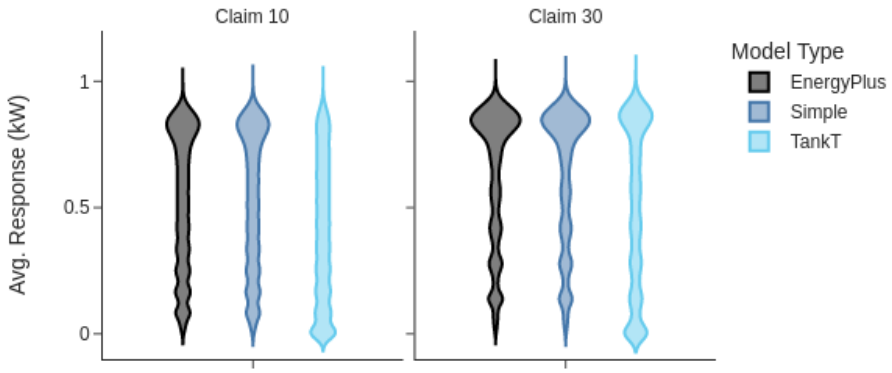


Figure 18. Distributions of individual HPWH average responses during Claim10–Random and Claim30–Random events: EnergyPlus, Simple surrogate models, and TankT surrogate models. Only data from non-null responses are included.

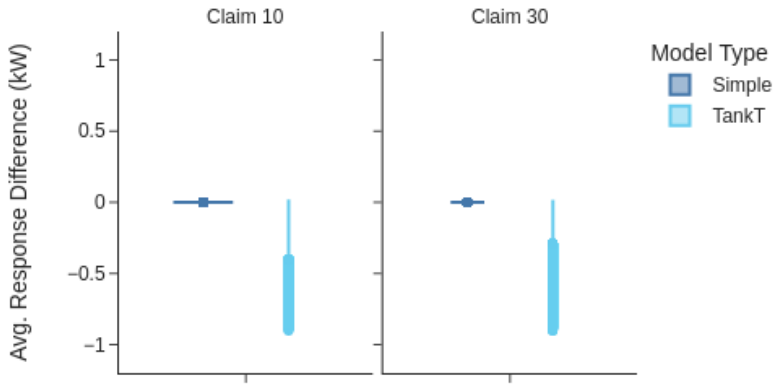


Figure 19. Distributions of differences between surrogate model and EnergyPlus individual HPWH average responses during Claim10–Random and Claim30–Random events. Only data from non-null responses are included.

Thus, from these results we see that TankT models best represent individual contingency responses for ERWHs and the Simple models are the better match for HPWH individual contingency responses.

4.2.3 Aggregate Response Validation: EnergyPlus, Outer Approximation, and Inner Approximations

To validate aggregate contingency responses, we compare the sum of responses over all of ISO-NE for individual EnergyPlus buildings and for surrogate models aggregated to the node (783 models), dispatch zone (19 models), or load region (8 models) level. The more aggregation there is (which corresponds to fewer models), the more conservative the inner approximations are since the worst case power and energy bounds are computed over a larger population of water heaters. In contrast, the outer approximation responses stay relatively constant when analyzed at the whole-system level, since the outer approximation surrogate models are computed by simply summing up the power and energy bounds of each water heater (with minor adjustments for variations in charging efficiencies and dissipation rates).

Comparing the ERWH surrogate model responses to the ERWH EnergyPlus responses in Figure 20, we see that the outer and inner approximations form an envelope containing the actual response, as expected based on the aggregation models' constructive proofs. Also, the Inner - Node aggregations most closely approximate the EnergyPlus responses on the conservative side, while the dispatch zone and load region aggregations are nearly identical and more conservative. Thus, from an aggregator's perspective and at this level of analysis the Inner - Node models are the best option shown here for bidding contingency resource into a wholesale market—they are provably dispatchable and significantly less conservative compared to dispatch zone and load region aggregations. Note that, as computed for this study, the inner approximations are not able to anticipate the expected post-response rebound, because worst-case bounds for ability to increase power draw are set on an hourly basis assuming that the tank temperature cannot exceed set point. This modeling deficiency can be corrected, but was not because ISOs currently do not anticipate rebound from demand response resources providing contingency reserve. It is also the case that if rebound became a system-level concern, aggregators would likely shape rebound responses to be less spiky than those shown for EnergyPlus and outer approximation.

The specific inner approximations shown in Figure 20 are the TankT, $w = 0.1$ models for both the dispatch zone and load region aggregate responses, and the TankT, $w = 0.05$ and TankT, $w = 0$ (maxP) models for the nodal Claim10 and Claim30 models respectively. Figure 21 summarizes all of the aggregate validation results using a Bland-Altman plot of response differences (surrogate model minus EnergyPlus) versus average response. The top two plots show that both the outer and the inner approximation surrogate models tend to overestimate response when the Simple model is used, unless w is 0.5 or greater. Mathematically, the inner approximations should never overestimate response if the individual flexibility models are a good representation of actual flexibility–thus this provides further evidence that the TankT models better reflect actual resource as measured by EnergyPlus simulations. The bottom two plots show that inner TankT approximations can match the aggregate EnergyPlus response well (differences near zero) in the Claim 30 case, but more often tend to under-estimate response, that is, they are true inner (conservative) approximations of actual resource.

Figure 20. Example of aggregate ERWH responses for contingency events at 6 p.m. on July 20. Simulated responses from EnergyPlus and several surrogate models are compared.

We summarize relative error of response in Figure 22, this time with results broken out by aggregation level (Node, Dispatch Zone or Load Region) and claim time (2pm or 6pm) as well as by surrogate model type, claim type, and aggregation type. The top sub-figure again shows how the Simple model overestimates response. In the bottom sub-figure of TankT results we see that for all claim types and times inner nodal aggregations give the most accurate results. Furthermore, for these contingency responses, w values < 0.5 are always more accurate than w = 0.5 or w = 1 (maxS). For nodal Claim30 aggregations, w < 0.05 is preferred. For Claim 10 responses, which are longer and thus have more need for energy capacity, and for higher levels of aggregation (dispatch zone or load region), w = 0.05 or w = 0.1 can be as good or better than the smallest w values.

All the HPWH aggregate surrogate responses in Figure 23 show good agreement with the corresponding EnergyPlus responses (black lines) during the actual response period (grey shading). The outer approximation then overestimates the speed with which the water heaters can recover during the rebound period that immediately follows, while the inner approximations do not capture the rebound period at all (because their parameters are defined in a worst-case sense and chosen to maximize contingency response). The inner approximations all started from Simple individual surrogate models and energy-weighting parameter $w = 0.0$ (maxP). As stated above, the rebound period does not factor into our contingency resource analysis. As such, it seems quite straightforward in this case to estimate HPWH contingency response—an inner maxP approximation at any level of aggregation could be used, as could the outer approximation, as they all give similar estimates in line with the EnergyPlus simulated responses.

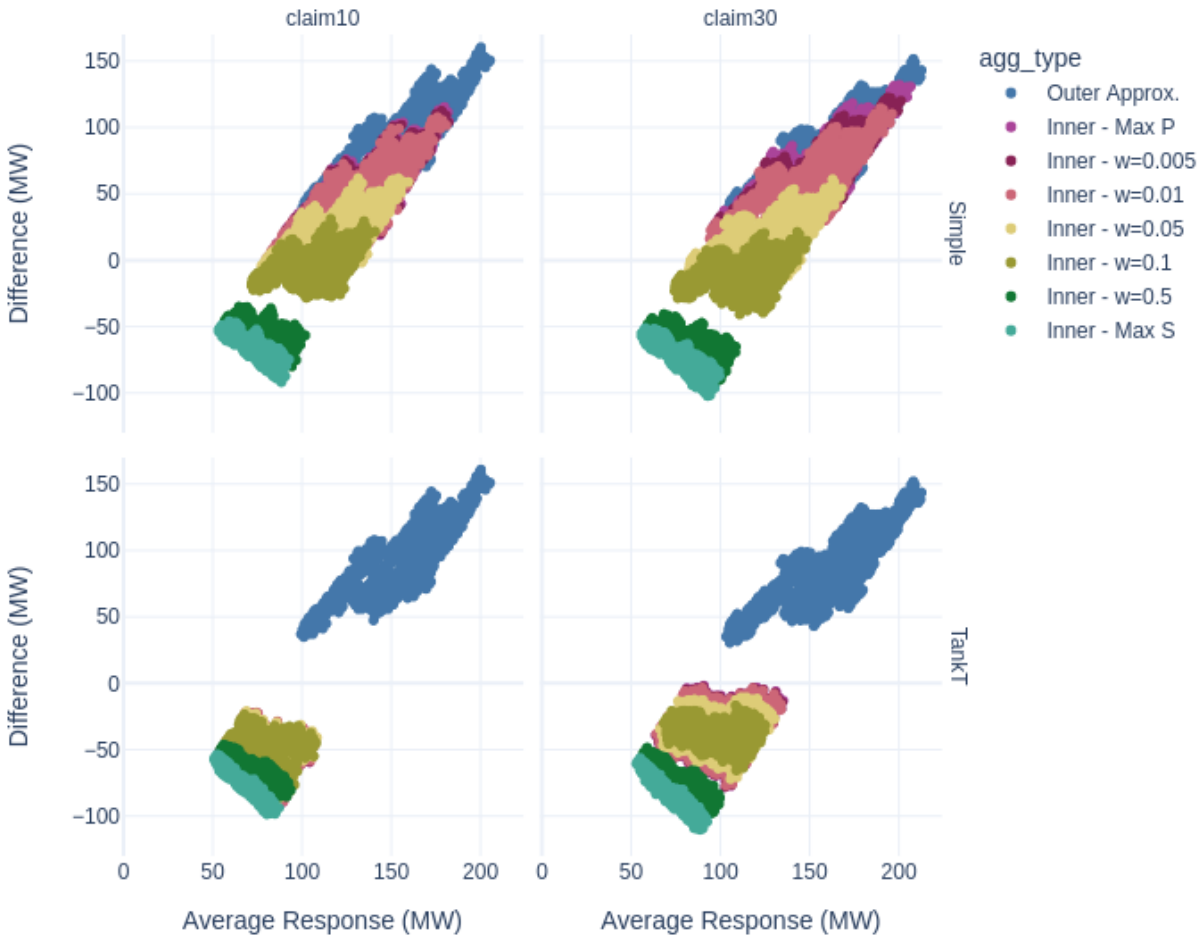
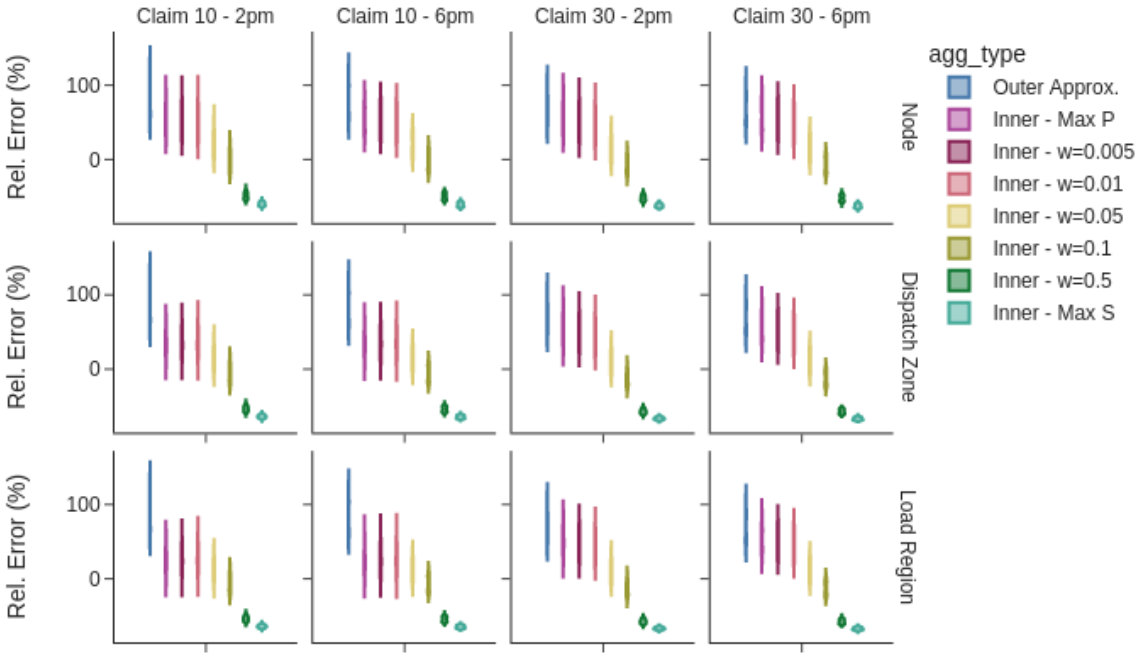
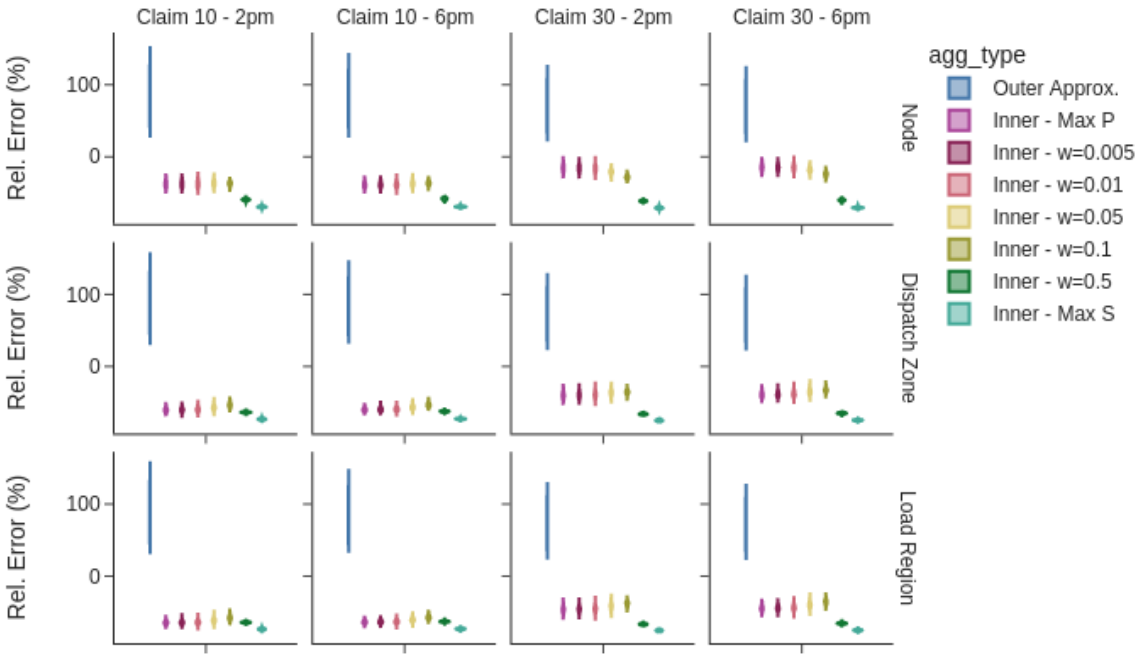


Figure 21. ERWH difference in responses (modeled minus EnergyPlus) versus average response plotted for all levels of aggregation and broken out by claim type and surrogate model type.



(a) Simple surrogate models



(b) TankT surrogate models

Figure 22. Distributions of ERWH aggregate response relative error

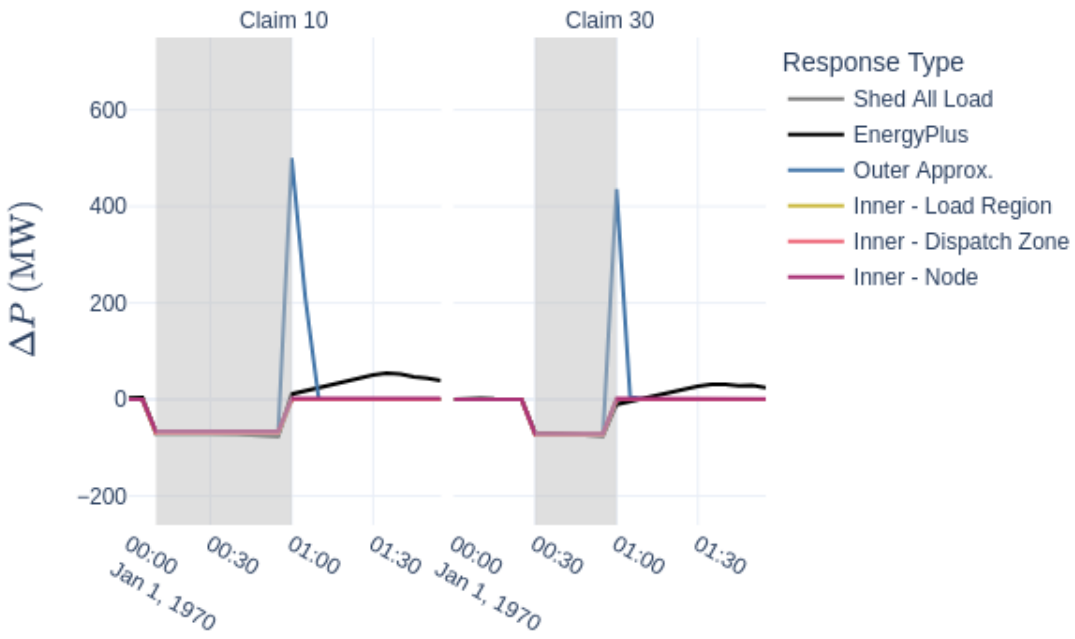


Figure 23. Examples of aggregate HPWH responses for contingency events at 6 p.m. on July 20. Simulated responses from EnergyPlus and several surrogate models are compared.

This conclusion is supported by analyzing all of the response data. Figure 24 shows the Bland-Altman plot for all simulated HPWH responses. Overall, the responses at 2 p.m. and 6 p.m. on each day of the year range between about 50 MW and 125 MW. The outer approximation does a good job of approximating response in this case, with differences generally less than 10 MW. The Simple Inner - MaxP approximation performs similarly. All other approximations are overly conservative in comparison. Increasing values of w, which put higher weight on energy, as compared to power, capacity of the aggregates are associated with increasing conservatism. And the TankT surrogate models are fundamentally more (and overly) conservative compared to the Simple surrogate models.

These data are summarized with more breakout of geographic aggregation and simulated claim time in Figure 25. These findings are particular to this version of ResStock's HPWH models in the sense that real-world HPWHs often have smaller tank sizes than those simulated in this study (80 gallons) and smaller tank sizes generally lead to less ability to sustain response over long periods of time. However, it is the case that real-world HPWHs are more likely than ERWHs to be running at any given time, such that the magnitude of response per water heater that we observe is likely realistic, especially for Claim30 responses.

Based on validating individual and aggregate responses from ERWHs and HPWHs, we conclude that contingency resource from ERWHs should be estimated using TankT model inner approximations with small values of w, and HPWH contingency resource should be estimated using Simple model inner approximations with w = 0 (Max P). In our context, ERWHs often are not able to provide a full response because their tank temperatures drop so low that the heater must

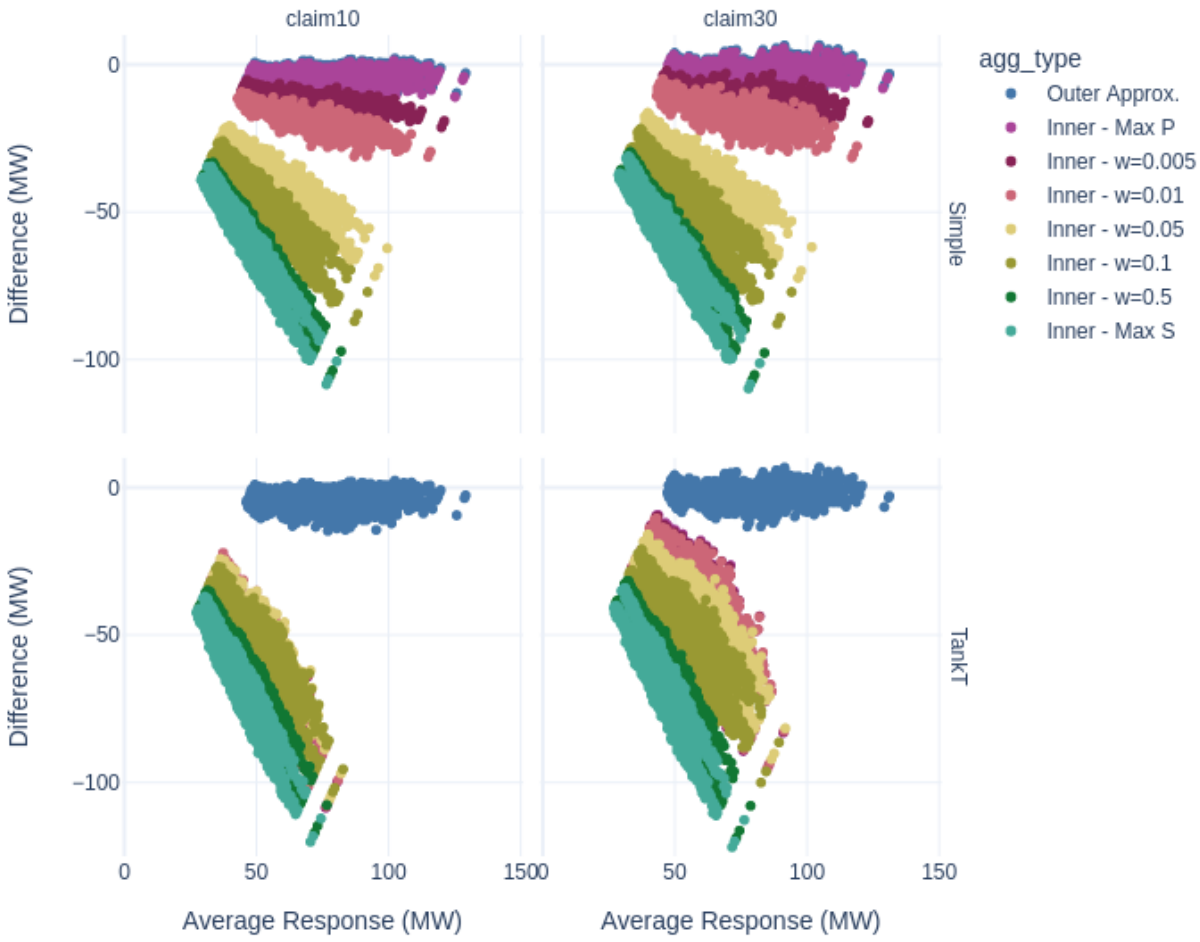
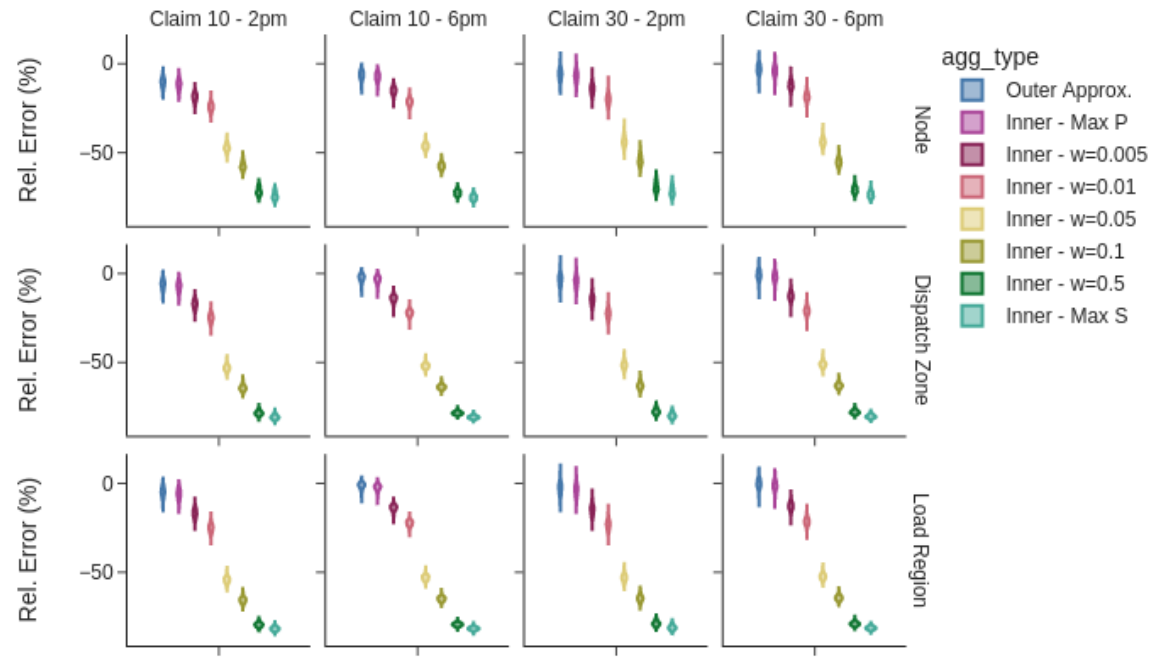
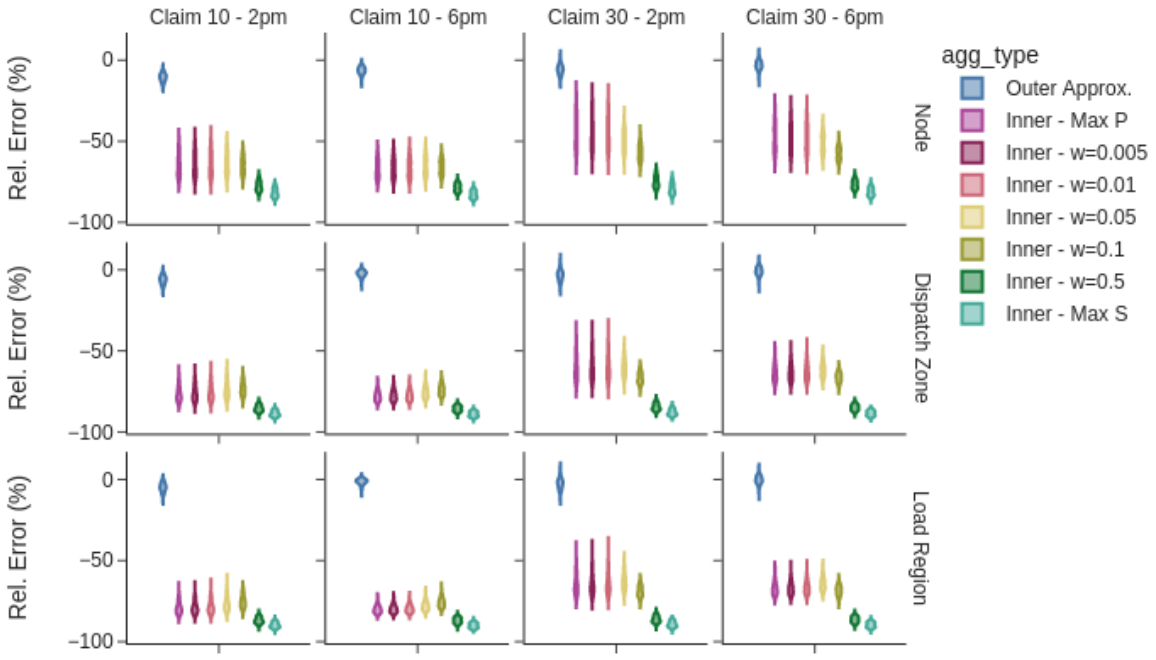


Figure 24. HPWH difference in responses (modeled minus EnergyPlus) versus average response plotted for all levels of aggregation and broken out by claim type and surrogate model type.



(a) Simple surrogate models



(b) TankT surrogate models

Figure 25. Distributions of HPWH aggregate response relative error

be turned on mid-event, but HPWHs do not have this problem and are generally able to shed all load that otherwise would have occurred during the response period. Again, this difference in model fit is most likely due to a combination of higher modeled storage volume for HPWHs and various modeling details, including unrealistically low power capacities for ERWHs.

4.3 Resource

To characterize the contingency resource available from electric water heaters for the entire year, we use inner aggregate approximations to simulate Claim10 and Claim30 responses for each hour of the year. Then, the average response (in MW) during the response period becomes the reported contingency resource for that aggregate-hour. Summing the resource over all aggregates in a set gives us the entire estimated resource in New England for that hour. This mimics the process of each aggregate being bid into the market and dispatched separately, thereby presenting MW-scale resources at multiple places (8 regions, 19 dispatch zones, or 783 nodes) in the transmission system. Calculating responses with EnergyPlus would be computationally prohibitive as calculating Claim10 and Claim30 resource for a single instance of ResStock would require running ResStock 49 times, once for the baseline case and once for each of the 24 hours in a day for both Claim10 and Claim30 responses. In contrast, the methods described in this report requires only a single baseline ResStock run and relatively fast event simulations using the surrogate models.

We summarize annual resource in Table 4. For each combination of claim type and water heater type, we highlight in blue the maximum amount of annual resource (in GWh) captured by the inner approximations and in the "Fraction Captured" row note the fraction of annual load that maximum estimate represents. Because ERWHs are energy-limited, we use the TankT model and find that less aggregation is able to capture more resource—the nodal resources are 30% to 50% larger than the regional resources. Also, the maximum resource for ERWHs usually corresponds to inner aggregations with $w > 0$, that is, aggregations that represent a balance of power and energy capacity. In contrast, because we find the EnergyPlus HPWHs can generally shed all load during contingency events with response periods of up to 50 minutes, for HPWHs we use the Simple model and find that the Max P inner aggregations always maximize resource quantity. The correlation between resource size and aggregation level also runs in the opposite direction—more aggregation enables more power capacity, and because these resources are not energy-limited, more resource is captured overall.

Selecting the maximal realistic resource estimates for each technology: nodal aggregations for ERWH and regional aggregations for HPWH, Inner - Max P aggregations except in the ERWH, nodal, Claim 10 case where we use the Inner - $w = 0.05$ aggregations; we can summarize the overall resource in terms of load fraction, total GWh per year, and MWh per water heater year. The ERWH scenario with about 603,400 water heaters is able to present 47% of annual load as Claim30 resource and 32% of annual load as Claim10 resource, which corresponds to 722 GWh and 488 GWh, and 1.2 MWh/water heater and 0.81 MWh/water heater, respectively. The HPWH scenario, which represents about 619,000 water heaters, is able to present 97% of the annual load for Claim30 service and 93% for Claim10 service, which correspond to 612 GWh and 595 GWh, and 0.99 MWh/water heater and 0.96 MWh/water heater respectively. We therefore find that contingency service resource from similar numbers of ERWHs and HPWHs can be similar if ERWH responses tend to be energy-limited and HPWH responses do not.

Table 4. Summary of Annual Contingency Reserve Resource. The inner approximation that captures the most resource for each combination of claim type and water heater type is highlighted in blue. The fraction of total load this represents is listed in the Fraction Captured rows.

Claim Type	Annual Resource (GWh)	ERWH (TankT)			HPWH (Simple)		
		Node	Dispatch Zone	Load Region	Node	Dispatch Zone	Load Region
N/A	**Total Demand**	1547	1547	1547	639.8	639.8	639.8
Claim10	Outer Approx.	1444	1466	1470	573.4	596.4	601.5
	Inner - Max P	479.1	316.4	285.5	566.6	589.8	595.0
	Inner - w = 0.005	480.3	318.9	287.8	520.5	525.8	526.2
	Inner - w = 0.01	481.4	321.3	290.2	483.6	476.6	474.1
	Inner - w = 0.05	487.6	340.8	309.4	335.3	297.3	289.0
	Inner - w = 0.1	487.4	364.1	333.9	268.2	225.1	216.6
	Inner - w = 0.5	321.3	283.0	277.4	176.1	135.1	128.3
	Inner - Max S	236.6	203.9	199.5	160.1	120.6	114.3
	Fraction Captured	0.32	0.24	0.22	0.88	0.92	0.93
Claim30	Outer Approx.	1489	1502	1504	602.9	615.7	618.6
	Inner - Max P	722.3	523.3	475.0	596.3	609.2	612.0
	Inner - w = 0.005	719.9	526.7	478.7	548.2	543.0	541.3
	Inner - w = 0.01	717.3	530.0	482.5	509.9	492.4	487.8
	Inner - w = 0.05	688.1	552.2	511.4	356.6	307.9	297.8
	Inner - w = 0.1	633.6	560.0	536.8	286.9	233.5	223.4
	Inner - w = 0.5	337.1	291.1	284.6	190.4	140.4	132.4
	Inner - Max S	248.8	209.8	204.6	173.4	125.3	118.0
	Fraction Captured	0.47	0.36	0.35	0.93	0.95	0.97

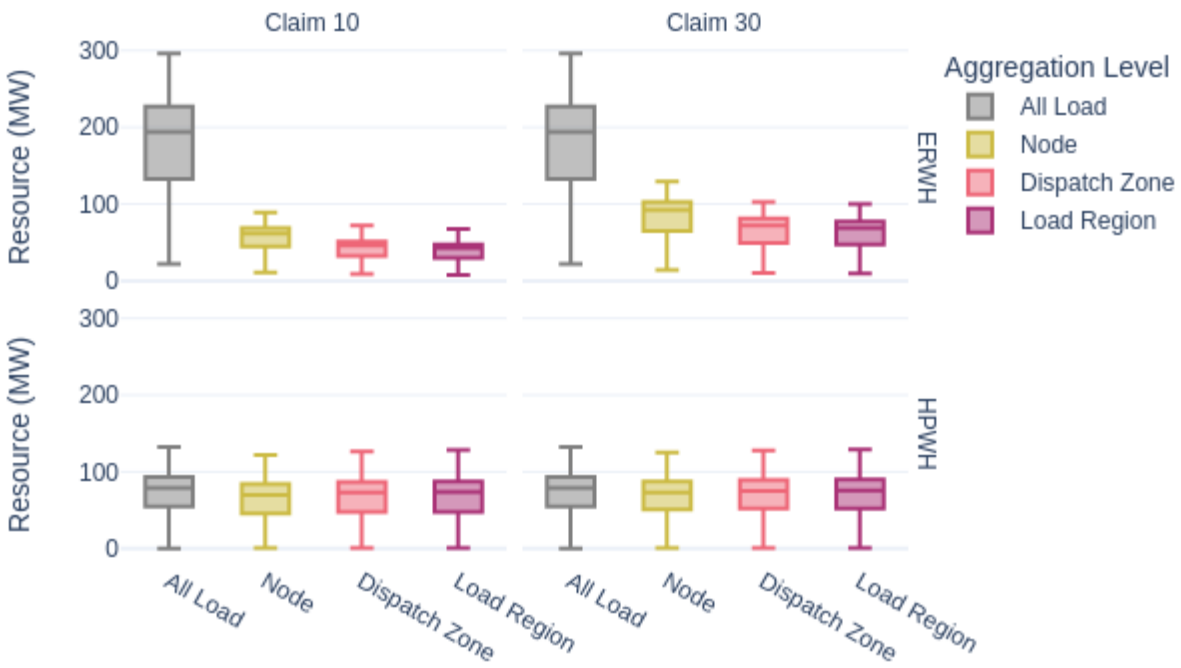
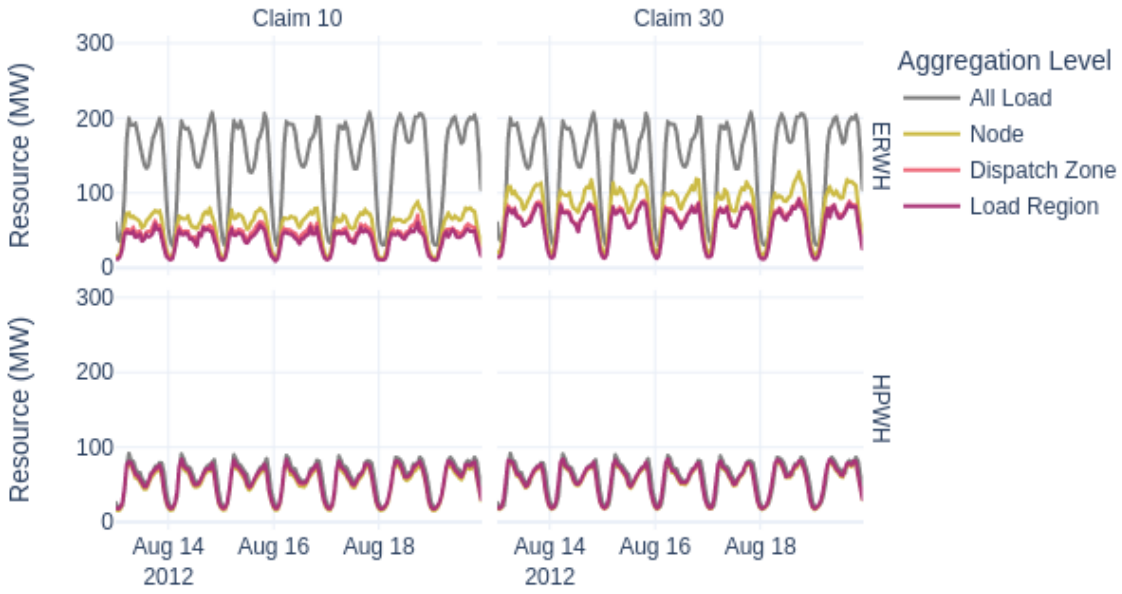


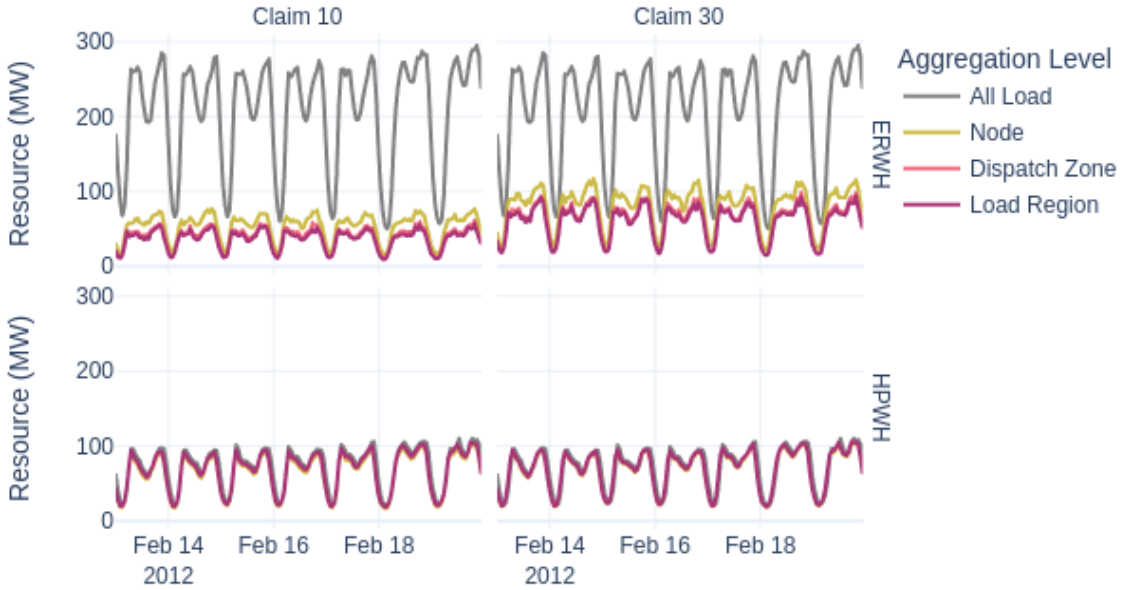
Figure 26. Box plots of contingency resource estimates as compared to all ERWH or HPWH load

Figure 26 tells the same story graphically and provides distributional detail. Estimated at the nodal level, the ERWH resource interquartile ranges are 45 MW–69 MW for Claim10 and 65 MW–102 MW for Claim30. At the Load Region level, the HPWH interquartile ranges are 48 MW–88 MW for Claim10 and 52 MW–90 MW for Claim30. Again, the ERWH resource is much more energy-limited, which limits how much of the total load is usable for contingency response and yields significant differences between Claim10 and Claim30 resource. Median HPWH Claim30 resource is larger than median HPWH Claim10 resource, but by only 2%, compared to 48% for for ERWHs.

The examples of weekly profiles in Figure 27 show that contingency resource tracks water heater load diurnal profiles, which are fairly high from about 6 a.m. to midnight, and only very low during the early morning hours, according to ResStock. The profiles also show a typical residential double-peak profile, with highest use in the mornings and evenings as people start and end their days. Also, water heating requires more energy in winter than summer, because both air and water temperatures are lower while water heater temperature set points remain unchanged. Finally, this gives yet another view of our persistent finding that for this version of ResStock, contingency resources for ERWHs and HPWHs are of similar magnitude even though ERWHs use 2–3 times more energy on average. Note that this conclusion would likely shift if we were to model more realistic ERWH heating capacities; while overall ERWH energy consumption would likely not change significantly, three times larger heating capacities leading to a reduced number and duration of heating cycles and a greater ability to increase and maintain tank temperatures would create more opportunity to shed during claim events.



(a) Summer week



(b) Winter week

Figure 27. Contingency resource example profiles–best inner approximations compared to hourly profile of all ERWH or HPWH load

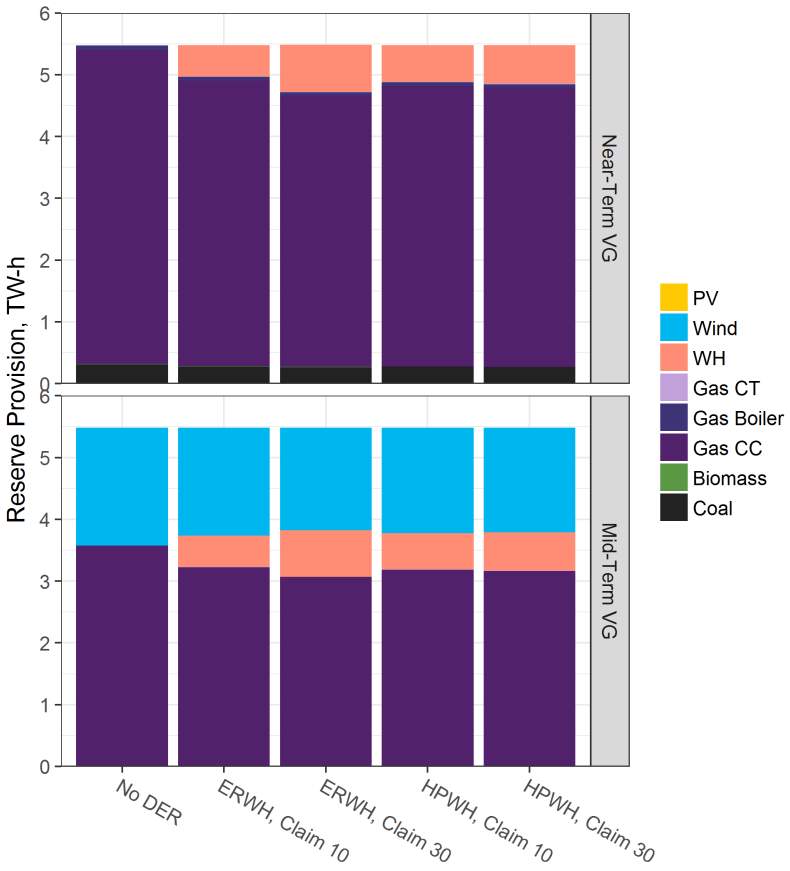


Figure 28. Provision of annual reserve in the Near-Term VG and Mid-Term VG ISO-NE models with ERWH or HPWH, Claim10 or Claim30 contingency reserve available from water heaters

4.4 Grid Impacts

When contingency reserve from water heaters is included as a resource in our PLEXOS models of ISO-NE as a no cost resource, water heaters (WH) provide as little as 9% and as much as 14% of the annual reserve depending on the water heater type and the length of the notification period (Figure 28). The low bound is set by ERWHs providing Claim10 service because this combination is most energy-limited (unable to maintain full response for the entire event duration due to tank temperatures dropping too low). Although ERWH Claim30 service is also more energy-limited than the HPWH responses, it sets the upper bound of contingency reserve provided by water heaters because the countervailing effect of starting with more load (compared to HPWHs) overcomes the energy effects in this case. Both types of water heaters are impacted by the duration of the expected response. Because we estimate that only 30 minutes of response has to be provided by Claim30 resources, compared to 50 minutes of response for Claim10 resources, the hourly resource in MW for Claim30 is larger than it is for Claim10. In either case, water heaters displace reserve provision from Gas combined-cycle (CC) units in the Near-Term VG model. In the Mid-Term VG model, reserve from water heaters displaces reserve that would otherwise come from both Gas CC and Wind units.

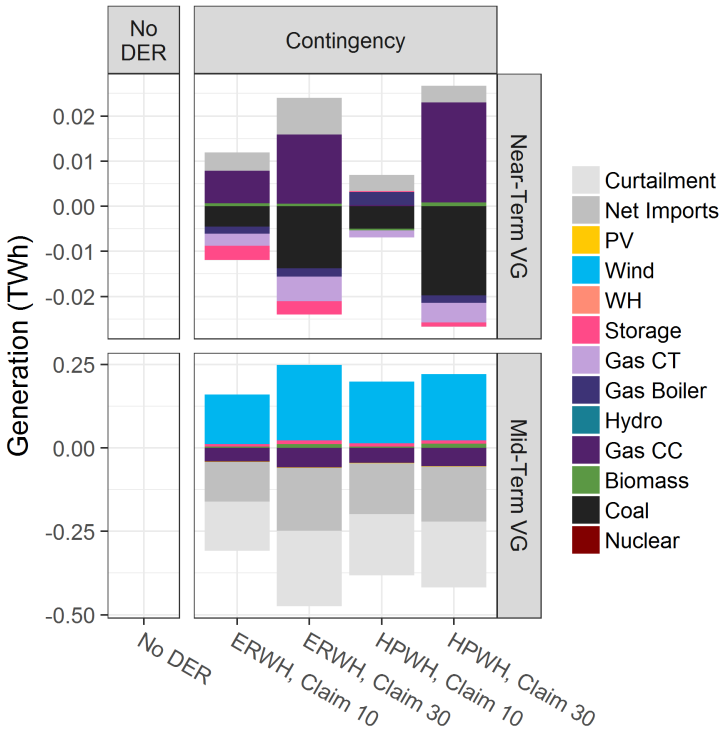


Figure 29. Annual generation differences (from the No DER case, Figure 11) in the Near-Term VG and Mid-Term VG ISO-NE models with ERWH or HPWH, Claim10 or Claim30 contingency reserve available from water heaters

Procuring reserve from electric water heaters has the follow-on effect of modifying the overall dispatch stack, as the capacity previously held for reserve is now available for generation. Figure 29 shows the resulting net change in annual generation. The first thing to notice is that the changes are small relative to the total annual generation in New England, which is about 105 TWh in our models. However, while the changes are on the order of 0.01% of annual generation in the Near-Term VG model, they are on the order of 0.1% in the Mid-Term VG model. The types of changes seen are fairly consistent within the two models. Except for the HPWH, Claim 10 case, in the Near-Term VG model coal, gas boiler, and gas combustion turbine (CT) generation decrease and are replaced by more gas CC generation and fewer exports (represented as an increase in net imports). Storage losses also increase (represented as a decrease in storage "generation"), which means that there is more energy shifting from pumped hydro plants than in the No DER case. Across all water heater contingency types, the Mid-Term VG model shows a clear pattern of using reserve from water heaters to reduce curtailment of wind generation, which also results in less gas CC generation and more exports (negative net imports).

Similar to generation, the total system cost changes from using water heaters for reserve is fairly small, and the relative differences are much smaller for the Near-Term VG model (0.01% - 0.03%) than they are for the Mid-Term VG model (0.2% - 0.3%) (Figure 30). The Mid-Term VG model percent savings are larger not only because the absolute magnitudes of the savings are several times larger than those in the Near-Term VG model, but also because the total cost denominator is about half as large due to more generation from zero or low marginal cost resources. Where cost savings come from is also different between the two models–WH reserve mostly

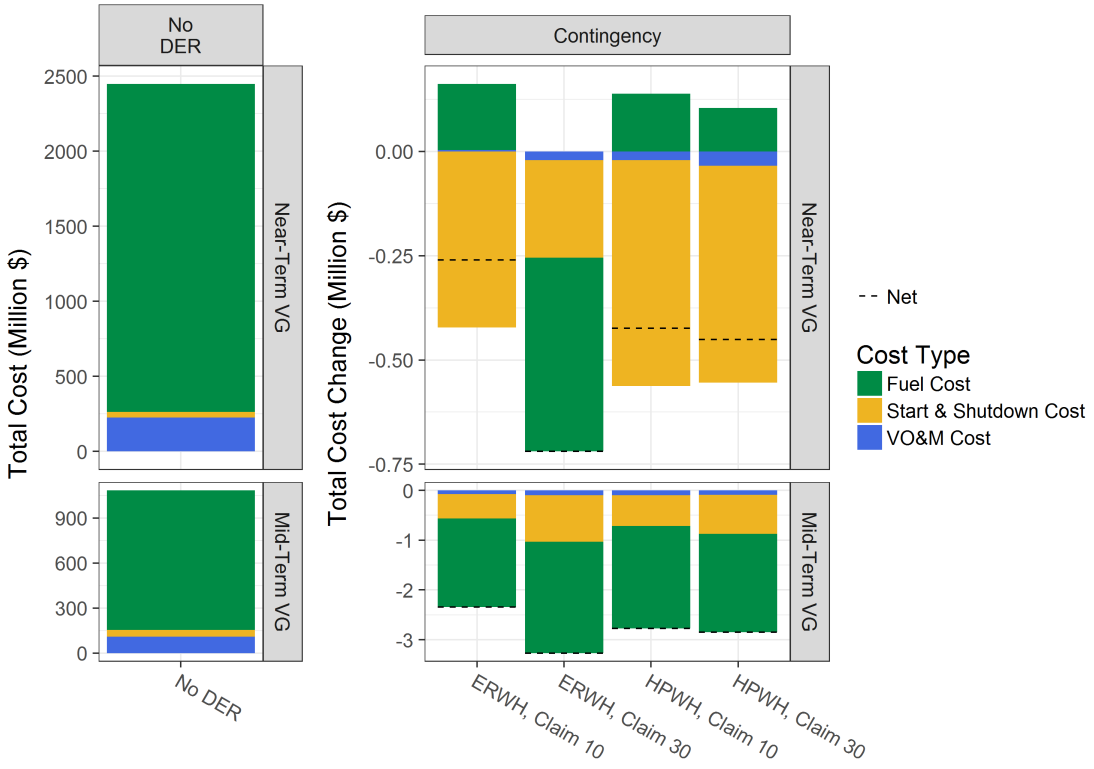


Figure 30. Total operational system costs and difference for claim cases versus the No DER case in the Near-Term VG and Mid-Term VG ISO-NE models with ERWH or HPWH, Claim10 or Claim30 contingency reserve available from water heaters

avoids Start & Shutdown Costs in the Near-Term VG model whereas savings are weighted more toward Fuel Costs in the Mid-Term VG model.

Thus we see that while the potential contingency reserve resource from electric water heaters in New England is modest, harnessing this resource could make a measurable difference in the New England power system, especially in a future with significantly more wind generation. While the use of water heater reserve in our Near-Term VG and Mid-Term VG models is similar, the impacts on generation and system costs are on the order of 0.01% in the Near-Term VG model but on the order of 0.1% in the Mid-Term VG model in which water heater reserve enables the use of otherwise-curtailed wind generation. While these percentages are small, they represent $0.3 to 0.7 million total savings ($0.40/WH-yr to $1.20/WH-yr) in the Near-Term VG model and $2.3 to 3.2 million savings ($3.80/WH-yr to $5.30/WH-yr) in the Mid-Term VG model. As expected, Claim30 (if that is sufficient from an operational perspective) is more impactful than Claim10 because it is able to present more megawatts of capacity to the system. Similarly, the value of ERWHs is higher than that of HPWHs in the Claim30 case because ERWHs are simply a larger resource than HPWHs on a heater-for-heater basis. Our HPWH results also highlight the extent to which water heaters with larger tank volumes can provide more resource and more value for services that require a longer response (Claim10 in this case).

5 Shifting Service

The individual surrogate models developed in the previous section are the starting point for evaluating the ability of water heaters to shift energy use from higher- to lower-price times. Because grid-supportive shifting profiles can be different from day to day and could be implemented via either set point changes or direct load control, here we rely on the validation done for contingency response to limit our focus to TankT models for ERWHs and Simple models for HPWHs and do not perform shifting-specific validation against EnergyPlus. We evaluate shifting at the individual water heater level and at an aggregate level. For the former we perform price-taking dispatch of hourly individual surrogate models against hourly energy prices from our PLEXOS day-ahead production cost models. For the latter, we are only able to analyze ERWHs because our aggregation mathematics do not currently apply to HPWHs with time-varying (and tank temperature-dependent) efficiency (η) and maximum power ($\overline{P}$) parameters. However, when we are able to create aggregate surrogate models, we represent them directly in our PLEXOS day-ahead production costs models so as to observe the grid impacts in a price-making sense; that is, in that case the water heater flexibility is dispatched alongside the supply-side resources and is thus able to impact unit starts and shutdowns, as well as help to set the energy price.

5.1 Methods

5.1.1 Price-taking Dispatch

Many individual surrogate models developed as described in Chapter 2 can be dispatched in parallel to maximize total profit given a price profile $p(t)$. PLEXOS provides such price profiles in units of \$/MWh. For this project we created an optimization formulation in GAMS that maximizes:

$$\sum_{k,t} -p(t) \cdot \Delta P_k(t) \tag{5.1}$$

subject to each discretized flexibility model's Equation 4.1 to Equation 4.3 with individual water heaters indexed by k and all of the parameters, whether constant across all devices, constant for each device, or completely time varying, handled and imported automatically such that the same formulation and supporting Python code is used for the two different types of water heaters (and can also be used for other flexible devices such as behind-the-meter battery systems and electric vehicles).

Each water heater is dispatched against prices for the node it is connected to, load-weighted dispatch zone prices, or load-weighted regional prices. This provides direct outputs of (a) shifted load profiles and (b) estimated price-taking profits. We also export the shifted load profiles for import into PLEXOS and are thus able to obtain the grid impacts as measured by the PLEXOS real-time model, which leaves the unit commitment and storage dispatch profiles from the day-ahead model in place for all non-quick-start units[9] and then adjusts the dispatch of available units to balance supply and demand intra-day.

For both price-taking dispatch and in communicating shiftable resource size to PLEXOS, we account for a 5.3% distribution loss factor (Cohen et al. 2019). This means that any $\pm$ 1 kWh change at the electrical meter is expected to have a $\pm$ 1.053 kWh impact at the substation, where

[9] Quick start units include Gas CT units

nodal prices are estimated. Thus we increase price-taking profits by a factor of 5.3% and also increase the magnitude of all aggregate resource bounds by 5.3% to estimate the impact of behind-the-meter changes in load on the bulk power system.

5.1.2 Aggregation of Electric Resistance Water Heaters

We aggregate ERWHs as described in Subsection subsection 4.1.1 and Appendix A, but special care is needed when selecting which water heaters to aggregate to form inner approximations. First, to ensure dispatchability, the fractions β_k need to be held constant over the whole day so that the energy deferred in one hour is actually repaid in a different hour. If the fractions varied from hour-to-hour, there would be no guarantee that the same water heaters who reduced load would be the same ones that increase load at another time. Then, because the aggregation sets energy and power bounds (in both directions) based on the worst-case values, to obtain any resource at all the water heaters being aggregated need to have similar-enough change-in-power and change-in-energy bounds such that there are some times available for both increasing and decreasing load.

When calculating inner approximations, we group water heaters with similar profiles together by applying k-means clustering to each water heater's $\underline{\Delta P}_k$ and $\underline{\Delta S}_k$ profiles normalized by $\overline{P}_k$. This is the last step in processing individual flexibility models prior to aggregation. The whole process is:

1. *Optionally* group the water heaters by a geography, that is, by node, dispatch zone or load region.
2. Split each flexibility model's timeseries data by day to produce separate flexibility model sets. At the end of this step, each day has its own set of surrogate flexibility models.
3. Apply k-means clustering to the resulting groups of surrogate flexibility models. Each group represents a single day and may also come from a single geography (if step 1 was applied).

At this point, aggregation proceeds largely as before, but is applied separately to each cluster, day, and (optionally) geography combination. Furthermore, the inner aggregations specify constant β_k values for the whole period even though the power and energy bounds for each water heater vary over the day. This makes it so that the objective function we used to select optimal β_k for contingency service, Equation 4.9, is no longer well-defined. Instead we apply proxies for how each water heater's bounds translate to $|\underline{L}|$ and $|\underline{U}|$. We estimate the overall energy capacity of each water heater in a worst-case sense as

$$\min\left(\min_t |\underline{\Delta S}_k(t)|, \min_t |\overline{\Delta S}_k(t)|\right) \tag{5.2}$$

and use $\overline{P}_k$ as a constant power capacity estimate. These values are then used to replace $c_k \min\left(\widetilde{T}_k - \underline{T}_k, \overline{T}_k - \widetilde{T}_k\right)$ and $\widetilde{P}_k(t)$, respectively, in Equation A.72. However, that step is applied after screening out resources with insufficient energy capacity, which are defined as resources whose values of Equation 5.2 are smaller than

$$0.1 \cdot \max\left(\max_{k,t} |\underline{\Delta S}_k(t)|, \min_{k,t} |\overline{\Delta S}_k(t)|\right). \tag{5.3}$$

Table 5. ERWH Groupings Used for Shifting Service Aggregation

Name	No. Geographies	No. Clusters per Geography	No. Resources
k5, Region	8	5	40
k15, Region	8	15	120
k30, Region	8	30	240
k40, ISO-NE	1	40	40
k120, ISO-NE	1	120	120
k240, ISO-NE	1	240	240

In testing, we found that more shifting resource is obtained when k-means clustering, as opposed to geography, is the primary way water heaters are grouped together. This is a natural consequence of the worst-case selection of inner approximation aggregate model parameters—the more similar profiles of water heaters in the same cluster compared to a similar number of water heaters selected in another way yield larger power and energy bounds. From prior experience, we also know that the PLEXOS model will perform better with fewer and larger resources. We therefore explore six different ways to select water heaters for aggregation–three that first group by load region and then apply k-means, and three that do not group at all by geography. The two sets match in terms of total numbers of resources as shown in Table 5. Also in all cases we examined inner approximations with $w \in \{0, 0.5, 1.0\}$ in price-taking analyses and $w = 1.0$ in PLEXOS, as well as the outer approximation computed using the k5, Region grouping.

5.1.3 Price-making Dispatch in PLEXOS

Representing Equation A.54 to Equation A.55 in PLEXOS requires translation from that mathematical form to an hourly discretized pumped hydro model with a head storage, a tail storage, and a turbine (generation) block. Generation is bounded between 0 and $-\underline{U}(t)$. Pumping (from the tail storage to the head storage) is bounded between 0 and $\overline{U}(t)$. Pumping efficiency is η and generation efficiency is $1/\eta$, although in our case (ERWHs originally modeled in EnergyPlus) these values are both simply equal to 1. The energy bounds can then be modeled by defining

$$L_{\text{base}} = \max\left(\max_t |\underline{L}(t)|, \max_t |\overline{L}(t)|\right), \tag{5.4}$$

setting head storage bounds as

$$\underline{L}(t) + L_{\text{base}} \leq L_{\text{head}}(t) \leq \overline{L}(t) + L_{\text{base}}, \tag{5.5}$$

and setting tail storage bounds as

$$L_{\text{base}} - \overline{L}(t) \leq L_{\text{tail}}(t) \leq L_{\text{base}} - \underline{L}(t). \tag{5.6}$$

The head and tail storages are also guaranteed to stay between 0 and $2L_{\text{base}}$. Although it should be possible to model the dissipation term ($-\alpha L(t)$) in PLEXOS using waterways and/or custom constraints, we leave that for future work given the small α values (less than 0.004 h^{-1}) in our models.

5.2 Resource

We estimate shifting resource starting with ERWH TankT surrogate models and HPWH Simple surrogate models computed with $\overline{T}_{\text{tank}} = 145°\text{F}$ and $\underline{T}_{\text{tank}} = 105°\text{F}$, which represent $\pm 20°\text{F}$ bounds on the nominal set point of 125°F. The resulting resources analyzed at the individual water heater level are summarized in the first two rows of Table 6. The individual models capture all of the baseline load and are able to shift it to other times subject to energy and power bounds. However, each resource is only a single water heater and is thus small, on the order of 1 kW and 3 kWh.

The third row of the table describes an aggregate ERWH outer approximation, which, like the set of individual ERWH models, is able to shift all ERWH load. By representing the shiftability of 603,500 water heaters using only 240 battery-like models, the size of each resource is greatly increased. The median maximum energy capacity is 35.5 MWh, the median maximum ability to decrease load is 4.7 MW, and the median maximum ability to increase load is 18.4 MW. Based on how these outer approximation models are constructed, we expect these aggregate models to strictly overestimate the ability of ERWHs to shift load from one time to another. Below, we see that this is the case for this particular outer approximation, which was made using the same groupings of water heaters as the k30, Region inner approximation also shown in Table 6. The effect is even more pronounced in more aggregated outer approximations such as the ones computed starting from the k5, Region groupings of water heaters, only use 40 resources to summarize ERWH load flexibility in ISO-NE.

Of the 18 sets of ERWH inner approximation aggregations (rows with aggregation levels "k5, Region" through "k240, ISO-NE"), the most important factors that determine how much of the original baseline load is available for shifting are total number of resources and whether resources are first grouped geographically or are only grouped according to baseline profiles. The range of outcomes across these two aggregation choices is large. Creating 40 resources starting with water heaters grouped by load region only captures 1.1% of the 1547 GWh of total ERWH load. Increasing the number of resources to 240 but still grouping first by load region captures 20% of total load. Staying with 240 resources but ignoring water heater location and only grouping by profile similarity increases the amount of load captured to 35%, which is similar to our hourly aggregation results for contingency resource at the dispatch zone and load region levels (Table 4).

The w weighting between power ($w = 0$) and energy ($w = 1$) has little impact on how much baseline energy is captured in the inner approximation aggregate power bounds, but does impact the amount of energy capacity available to each resource. Median energy capacity is about 1.5 to 1.8 times larger with $w = 1$ compared to $w = 0$. Median maximum power capacities are well above the 100 kW (0.1 MW) minimum suggested in FERC Order 2222 (FERC 2020), although that bound is violated for several of the resources computed for each day (since water heaters with little baseline power consumption are grouped together) and even relatively large resources will have hours in which they are not able to reduce load at all. For example, the data for June 3 for the k240, ISO-NE dataset includes 9 resources with $\underline{U}$ equal to zero in all hours and 4 resources with all $|\underline{U}| < 100$ kW. Of the 227 resources that at times exceed the 100 kW threshold, average size above the threshold ranges from 119 kW to 3.1 MW and the fraction of time above the threshold ranges from 4% to 96%, with a median of 40%.

Table 6. Annual Shifting Resource Summary

Water Heater Type	Aggregation Level	w	No. Resources	w_k	Shiftable Load (GWh)	Shiftable Load (% of Annual)	Median $\min_t \underline{L}(t)$ (MWh)	Median $\max_t \overline{L}(t)$ (MWh)	Median $\min_t \underline{U}(t)$ (MW)	Median $\max_t \overline{U}(t)$ (MW)
HPWH	None	N/A	2708	228.6	640	100	−0.0034	0.0033	−0.00093	0.0010
ERWH	None	N/A	2640	228.6	1547	100	−0.0020	0.0073	−0.0013	0.0013
	Outer Approx.	N/A	240	1	1547	100	−35.5	35.5	−4.72	18.4
	k5, Region	0	40	1	17.4	1.13	−26.9	26.9	−6.69	26.3
	k5, Region	0.5	40	1	17.4	1.13	−33.6	33.6	−6.69	24.9
	k5, Region	1	40	1	17.4	1.12	−42.0	42.0	−6.69	24.1
	k15, Region	0	120	1	169	10.9	−15.5	15.5	−4.92	17.8
	k15, Region	0.5	120	1	169	10.9	−20.1	20.1	−4.92	17.8
	k15, Region	1	120	1	168	10.9	−26.0	26.0	−4.92	16.8
	k30, Region	0	240	1	313	20.2	−12.0	12.0	−3.29	13.7
	k30, Region	0.5	240	1	312	20.2	−14.4	14.4	−3.29	13.6
	k30, Region	1	240	1	311	20.1	−18.6	18.6	−3.29	12.8
	k40, ISO-NE	0	40	1	180	11.6	−61.3	61.3	−14.5	69.2
	k40, ISO-NE	0.5	40	1	180	11.6	−81.6	81.6	−14.5	69.0
	k40, ISO-NE	1	40	1	179	11.6	−111	111	−14.5	66.0
	k120, ISO-NE	0	120	1	384	24.8	−35.1	35.1	−7.23	42.5
	k120, ISO-NE	0.5	120	1	382	24.7	−45.6	45.6	−7.23	42.4
	k120, ISO-NE	1	120	1	380	24.5	−62.4	62.4	−7.23	40.6
	k240, ISO-NE	0	240	1	544	35.2	−24.9	24.9	−4.52	31.7
	k240, ISO-NE	0.5	240	1	542	35.0	−28.6	28.6	−4.52	31.2
	k240, ISO-NE	1	240	1	537	34.7	−37.9	37.9	−4.47	29.5

5.3 Grid Impacts

All of the energy shifting resources, at the individual water heater and aggregate levels, are evaluated by dispatching them against day-ahead energy prices extracted from our PLEXOS Near-Term VG and Mid-Term VG ISO-NE models (Subsection 5.3.1). An outer approximation of ERWH flexibility (30 aggregates computed for each load region, that is, 240 flexibility resources in total) is also dispatched directly (i.e., endogenously) in the PLEXOS day-ahead models, which lets ERWH flexibility impact unit commitment and help set prices. We directly compare the endogenous dispatch and price-taking dispatch results by running all the resulting changes in load through the PLEXOS real-time models to ascertain impacts on supply-side dispatch and production costs (Subsection 5.3.2).

5.3.1 Price-taking Dispatch

The price-taking dispatch results are computed using nodal day-ahead energy prices from PLEXOS, as well as load-weighted average energy prices for each dispatch zone, load region, and for ISO-NE as a whole. Thus, flexibility resources can be dispatched against prices at any of these aggregation levels: node, dispatch zone, load region, or ISO-NE. Note that the more aggregated price data (e.g., load region or ISO-NE averages) will tend to be less volatile than price data at finer levels of resolution (e.g., node, dispatch zone). This can be seen for dispatch zone prices versus average ISO-NE prices by examining the price duration curves shown in Figure 31 and Figure 32. Figure 31 shows all hours but cuts off the y-axis at $100/MWh ($0.1/kWh). Figure 32 shows the full range of prices for the top 100 hours. Note that the very high prices in Figure 32 are set by various penalties in the PLEXOS model, e.g., for violating constraints such as reserve requirements and line limits. Although somewhat artificial in an economic sense, the high prices do represent real reliability concerns, and so we keep them as-is to reflect the high value of reducing load during those times. The highest prices in the Mid-Term VG model are lower than the highest prices in the Near-Term VG model because of transmission capacity adjustments made to reduce transmission congestion to more reasonable levels. Based on correspondence with ISO-NE our adjusted transmission assumptions are also largely in line with planned transmission system refurbishment.[10]

The Near-term VG model has almost no zero-price hours, fairly smooth price duration curves in the non-top hours, and little price variation between dispatch zones in the non-top hours, except that Springfield MA's prices are often higher than the others'. In contrast, the Mid-Term VG model has over 1000 zero-price hours, and significant price variations across dispatch zones for hours 3500 to 7000 (Figure 31). In the 100 top hours, in both models the highest-priced hours are limited to approximately the top 20 hours, and the Near-Term VG model appears to have more severe constraint violations with an average highest price of $5048/MWh as compared to $1579/MWh in the Mid-Term VG model (Figure 32).

The outcomes of dispatching the resources summarized in Table 6 against the day-ahead prices partially shown in Figure 31 and Figure 32 are summarized in Table 7 for the Near-Term VG model and Table 8 for the Mid-Term VG model. The first four rows of each table summarize outcomes for directly dispatching individual water heater flexibility against prices, and as such

[10]The key exception to this statement is our assumption of increased transmission capacity into the Boston area, which would require adding another circuit to an existing underground cable.

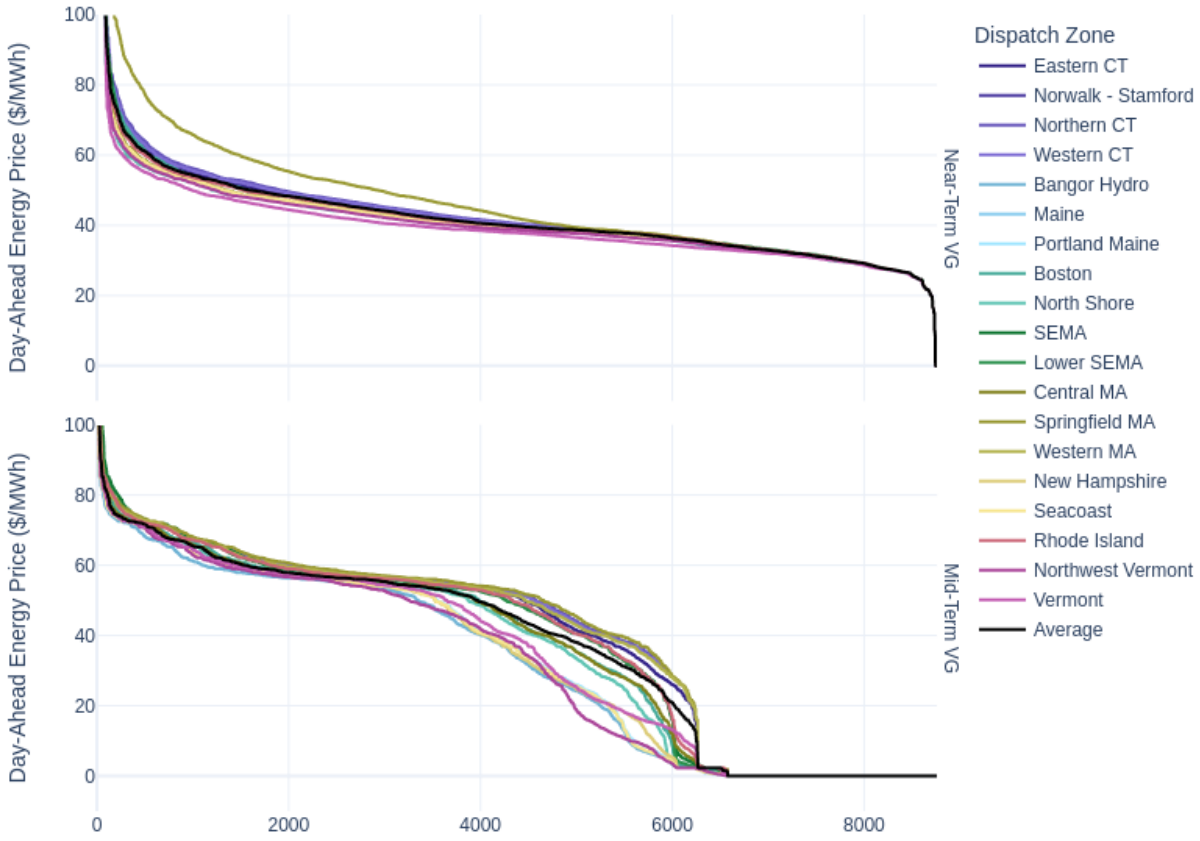


Figure 31. Load-weighted energy price duration curves for the PLEXOS ISO-NE Near-Term VG and Mid-Term VG models for the whole system (Average) and per dispatch zone. All hours to $100/MWh. Dispatch zones in the same load region are given similar colors.

provide benchmarks for the actual amount and value of shifting ERWH and HPWH load. The individual dispatch results are largely as expected. Dispatching against nodal prices results in slightly more profit (about 0.5%) than dispatching against the less volatile dispatch zone prices. ERWH shiftability is more valuable than HPWH shiftability. Energy shifting is more valuable under Mid-Term VG than Near-Term VG conditions. Somewhat surprisingly, ERWH shifting is only 75% more to about twice as valuable as HPWH shifting even though it represents more than a doubling of load (ERWH load is 240% of HPWH load). However, this is explainable based on the ERWH models having too-low power capacities and smaller tank volumes compared to the HPWH models (see Section 2.2). We observe a round-trip efficiency boost for HPWHs based on shifting energy from lower-efficiency to higher efficiency times, which means that the shifted HPWH profiles actually use less energy overall than the baseline profiles. However, we are not capturing the impact of shifting on efficiency but are rather directly using the hourly average efficiencies from the baseline run, so the advantages of this effect (round trip efficiency of 108% to 109%) might be overstated.

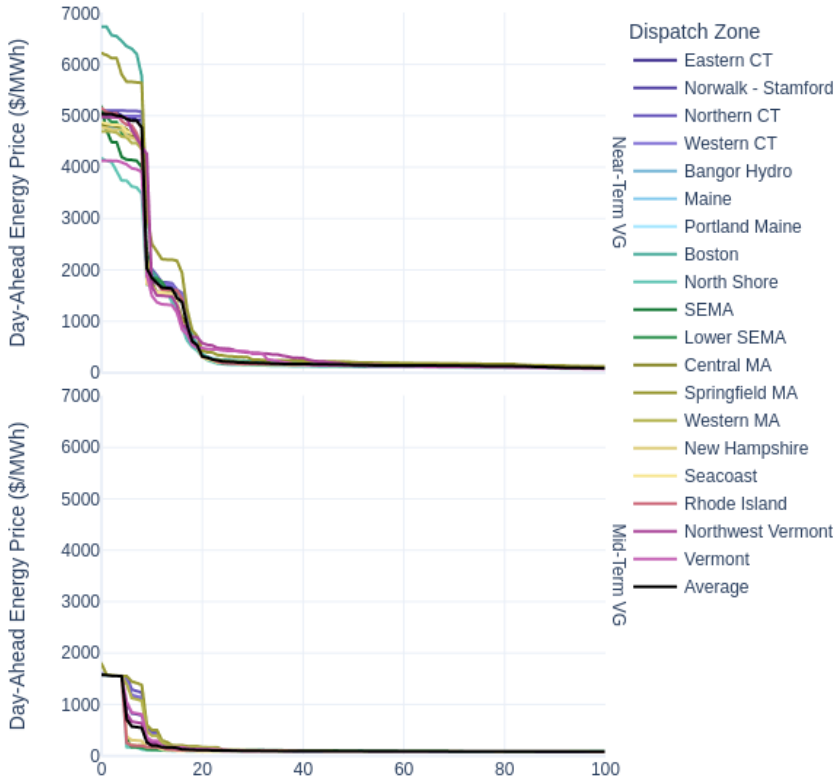


Figure 32. Load-weighted energy price duration curves for the PLEXOS ISO-NE Near-Term VG and Mid-Term VG models for the whole system (Average) and per dispatch zone. Top 100 hours.

The first thing to notice in comparing the aggregated ERWH results to the individual ERWH results is that the outer and inner approximations of aggregate ERWH flexibility provide upper and lower bounds on the amount of ERWH water heater load that can be shifted, and the expected profits mostly follow this rule as well. The inner approximations (with non-null w values) strictly underestimate the shiftability of ERWH load and likewise underestimate profitability. For example, although the ERWH k240, ISO-NE aggregated resources have about 30% of the water heater load available to them as a shiftable resource, dispatching those aggregate resources against day-ahead prices only shifts about 15% as much energy as the individual flexibility models and similarly only captures about 10% of the profits. On the other hand, the outer approximation (which was estimated using the same groupings of water heaters as the k30, Region inner approximation), shifts 7% (Near-Term VG) to 17% (Mid-Term VG) more load than the individual water heater dispatch. However, the impact of this extra shifting on profits is much smaller or even negative—the outer approximation produces 1.4% more profit and 0.5% less profit than is actually available based on dispatching the individual water heater flexibility against Near-Term VG and Mid-Term VG day-ahead prices, respectively. This result highlights the competing effects of the outer approximation artificially increasing resource size (because its constraints are more relaxed than those of the individual water heaters) and the more-aggregated prices (i.e., at the

regional instead of the nodal level) being less volatile and thus less amenable to arbitrage. The latter effect mitigates the former effect under both grid conditions, but in the Mid-Term VG case the latter effect overwhelms the former effect to produce less overall profit in the outer approximation scenario as compared to the no-aggregation scenarios. Note that this finding is sensitive to the amount of aggregation—more-aggregated groupings like k5, Region do show more profit for the outer approximation with regional Mid-Term VG prices as compared to individual water heaters dispatched against nodal Mid-Term VG prices.

Overall we find that the aggregation processes described in this report either overestimate flexibility (outer approximation) or do not maintain overall shifting resource quality in proportion to the amount of load nominally made available for shifting (inner approximations). By comparing inner approximation results within the same aggregation level but with different w values, we see that different w values do not qualitatively change the results but it is the case that $w = 1$ (max S) inner aggregations consistently produce higher profits than $w = 0$ (max P) or $w = 0.5$. For this reason, for the remainder of this report we focus on $w = 1.0$ (max S) results when discussing inner approximations.

Demand response studies often find significantly more value for energy shifting under higher-VG conditions, but that effect is muted here because of the expanded transmission capacity and subsequent reduction in penalty prices in the Mid-Term VG model as compared to the Near-Term VG model. Overall the differences in profit between the Near-Term VG and Mid-Term VG models are relatively small, about 16% more for HPWHs and 3% more for ERWHs under Mid-Term as compared to Near-Term VG conditions. HPWHs probably experience more of a boost under Mid-Term conditions because their larger energy capacities and smaller power capacities are better suited to taking advantage of numerous zero price hours in the Mid-Term VG model as opposed to mitigating the few high price hours in the Near-Term VG model. However, the advantage lies with ERWHs on a per water heater basis, with price-taking profits on the order of $40/year compared to about $20/year for HPWHs.

The diurnal profiles shown in Figure 33 illustrate how resource magnitude differs between water heater and aggregation types, and how shifted load shapes differ by season and grid conditions. First, comparing the "ERWH, k240, Max S" profiles to the others, we observe how conservative the shifted profiles are—only a small amount of load is moved into the early morning hours and away from peak morning and evening hours. All of the baseline load shown in that column is available for shifting, but the worst-case selection of bounds results in severe energy constraints that limit how much that load can be shifted in time. The individual HPWHs and the ERWHs, on the other hand, tend to move significant quantities of load into the early morning hours and away from the morning and evening residential load peaks. In the winter, the morning peak is avoided nearly as much as the evening peak under all grid conditions. Shifted ERWH load keeps its two peak shape, but now with a large peak in the early morning and a smaller peak in the early afternoon. The HPWH shifted profiles look similar, but with a smaller early afternoon peak relative to the early morning peak. The story is more complex in the shoulder and summer seasons. In the Near-term VG model, morning hours are obviously preferred to afternoon and evening hours in both the shoulder and (especially) summer months. However, in the Mid-Term VG model, the preference for morning hours is shifted later (toward daytime and away from early morning) and attenuated (especially in shoulder months). The desire to avoid evening hours is also attenuated,

especially in the early evening when there may still be some solar generation on the grid. As expected, the ERWH outer approximation results are similar to, but more exaggerated than, the ERWH individual water heater results. Because the outer approximation is able to shift the same amount of load but with fewer constraints (40 sets of aggregate constraints rather than 2,640 sets of individual constraints), the additional shifting represents the grid preferences encoded in day-ahead PLEXOS prices but is unrealistic from an individual water heater point of view (that is, if these shifts were realized we would expect water heater tank temperatures to stray outside of the allowed ranges, either too cold or too hot).

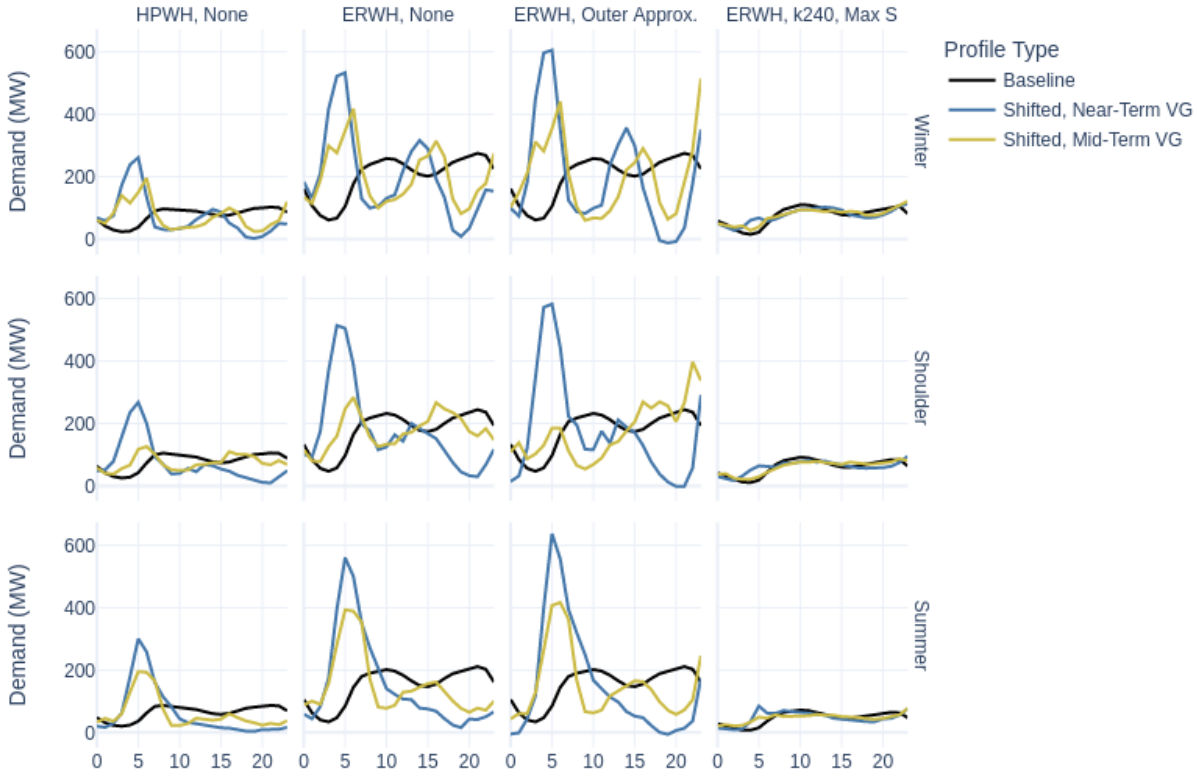


Figure 33. Baseline and shifted diurnal profiles, by season (rows), resource type (columns), and grid model (lines)

Table 7. Price-taking Dispatch Results with Near-Term VG Day-Ahead Prices

Water Heater Type	Aggregation Level	w	Pricing Zones	Shiftable Load (GWh)	Shifted Load (GWh)	Shifted Load (%)	Roundtrip Efficiency	Profit (million $/yr)	Profit per WH ($/WH-yr)
HPWH	None	N/A	Disp. Zone	639.8	358.8	56.1	1.09	11.576	18.70
	None	N/A	Node	639.8	363.0	56.7	1.09	11.615	18.76
ERWH	None	N/A	Disp. Zone	1547	709.3	45.8	0.997	22.821	37.81
	None	N/A	Node	1547	713.5	46.1	0.997	22.910	37.96
	Outer Approx.	N/A	Region	1547	762.8	49.3	0.992	23.233	38.50
	k5, Region	0	Region	17.46	1.24	7.1	0.997	0.043	0.07
	k5, Region	0.5	Region	17.45	1.28	7.3	0.996	0.045	0.07
	k5, Region	1	Region	17.40	1.32	7.6	0.996	0.048	0.08
	k15, Region	0	Region	168.8	7.80	4.6	0.996	0.237	0.29
	k15, Region	0.5	Region	168.7	8.03	4.8	0.995	0.245	0.41
	k15, Region	1	Region	168.3	8.23	4.9	0.995	0.250	0.42
	k30, Region	0	Region	312.8	23.98	7.7	0.996	0.729	1.21
	k30, Region	0.5	Region	312.4	24.39	7.8	0.996	0.737	1.22
	k30, Region	1	Region	311.1	24.60	7.9	0.996	0.744	1.23
	k40, ISO-NE	0	ISO-NE	180.3	8.18	4.5	0.994	0.220	0.36
	k40, ISO-NE	0.5	ISO-NE	180.1	8.57	4.8	0.994	0.236	0.39
	k40, ISO-NE	1	ISO-NE	179.4	8.95	5.0	0.993	0.249	0.41
	k120, ISO-NE	0	ISO-NE	383.7	42.43	11.0	0.996	1.032	1.71
	k120, ISO-NE	0.5	ISO-NE	382.5	43.29	11.3	0.996	1.056	1.75
	k120, ISO-NE	1	ISO-NE	379.7	43.61	11.5	0.996	1.076	1.78
	k240, ISO-NE	0	ISO-NE	544.1	96.49	17.7	0.998	2.477	4.10
	k240, ISO-NE	0.5	ISO-NE	542.0	97.32	18.0	0.997	2.498	4.14
	k240, ISO-NE	1	ISO-NE	537.1	96.83	18.0	0.997	2.497	4.14

Table 8. Price-taking Dispatch Results with Mid-Term VG Day-Ahead Prices

Water Heater Type	Aggregation Level	w	Pricing Zones	Shiftable Load (GWh)	Shifted Load (GWh)	Shifted Load (%)	Roundtrip Efficiency	Profit (million $/yr)	Profit per WH ($/WH-yr)
HPWH	None	N/A	Disp. Zone	639.8	314.1	49.1	1.086	13.482	21.78
	None	N/A	Node	639.8	317.6	49.6	1.083	13.534	21.86
ERWH	None	N/A	Disp. Zone	1547	612.1	39.6	1.000	23.323	38.65
	None	N/A	Node	1547	618.3	40.0	1.000	23.435	38.83
	Outer Approx.	N/A	Region	1547	724.8	53.2	0.998	23.317	38.64
	k5, Region	0	Region	17.46	1.30	7.4	0.999	0.031	0.05
	k5, Region	0.5	Region	17.45	1.33	7.6	0.999	0.032	0.05
	k5, Region	1	Region	17.40	1.36	7.8	0.999	0.033	0.06
	k15, Region	0	Region	168.8	8.00	4.7	0.998	0.226	0.37
	k15, Region	0.5	Region	168.7	8.22	4.9	0.998	0.234	0.39
	k15, Region	1	Region	168.3	8.25	4.9	0.998	0.241	0.40
	k30, Region	0	Region	312.8	23.77	7.6	0.999	0.695	1.15
	k30, Region	0.5	Region	312.4	24.09	7.7	0.998	0.710	1.18
	k30, Region	1	Region	311.1	23.93	7.7	0.998	0.722	1.20
	k40, ISO-NE	0	ISO-NE	180.3	8.54	4.7	0.997	0.222	0.40
	k40, ISO-NE	0.5	ISO-NE	180.1	8.88	4.9	0.997	0.233	0.39
	k40, ISO-NE	1	ISO-NE	179.4	9.02	5.0	0.997	0.245	0.37
	k120, ISO-NE	0	ISO-NE	383.7	44.87	11.7	0.999	1.046	1.73
	k120, ISO-NE	0.5	ISO-NE	382.5	45.53	11.9	0.999	1.076	1.78
	k120, ISO-NE	1	ISO-NE	379.7	45.13	11.9	0.998	1.099	1.82
	k240, ISO-NE	0	ISO-NE	544.1	100.7	18.5	0.999	2.265	3.75
	k240, ISO-NE	0.5	ISO-NE	542.0	100.3	18.5	0.999	2.304	3.82
	k240, ISO-NE	1	ISO-NE	537.1	99.93	18.6	0.999	2.321	3.85

5.3.2 Price-making Dispatch and Impacts

We created aggregate ERWH flexibility resources with the intention of including them directly in large-scale grid models. In this section we fulfill that intention by examining the performance of an outer approximation of ERWH flexibility endogenously dispatched by PLEXOS in the Near-Term VG and Mid-Term VG day-ahead models. As in Section 5.3.1, the outer approximation is computed using the k30, Region groupings of water heaters, which in PLEXOS yields 240 storage-like resources with time-varying power and energy bounds that are able to impact day-ahead unit commitment decisions and energy prices. We also worked with inner approximations in the PLEXOS day-ahead model but found their energy constraints to be overly conservative and difficult for the model to dispatch. We therefore exclude those results from this report.

The impact of both endogenously dispatched (price-making) and price-responsive (price-taking) water heater flexibility on supply-side dispatch can be observed using PLEXOS real-time models with fixed storage dispatch (i.e., fixed profiles for water heaters, pumped hydro, etc.) and fixed unit commitment decisions for coal and other slow-start resources. The only differences between the real-time models for the two dispatch mechanisms, endogenous dispatch and price-responsive dispatch, is that storage dispatch and unit commitment decisions are taken from day-ahead models with or without the outer approximation of ERWH flexibility included, respectively; and that the real-time water heater load profiles are set to match the results summarized in Section 5.3.1 in the case of price-responsive dispatch. The real-time model runs so constructed allow us to describe and quantify the changes in grid system costs associated with how the remainder of the system responds to water heater energy shifting no matter which type of dispatch mechanism is used, thereby producing comparable supply-side results. Note, however, that this modeling approach does not replicate other common differences between real-world day-ahead and real-time models such as load, wind, and solar forecast errors; finer time resolution; and unexpected outages or other deviations from day-ahead plan.

The annual generation differences that result from including shifted water heater load profiles in our PLEXOS ISO-NE real-time models are shown in Figure 34. First, we see that endogenous dispatch, as expected, has a larger impact than all instances of price-responsive dispatch because those scenarios rely on an outer approximation of aggregate ERWH flexibility and are able to impact day-ahead unit commitment and pumped hydro storage dispatch decisions. The magnitude of these two effects can be estimated by comparing the "ERWH, None" and "ERWH, Outer Approx." Price-Responsive Dispatch results to see the impact of using an aggregate outer approximation rather than individual water heater constraints; and by comparing the "ERWH, Outer Approx." results for Endogenous Dispatch and Price-Responsive Dispatch to see the impact of the same resource being able to impact day-ahead decisions in the former, but not the latter, case.

Focusing on the Near-Term VG results, across all dispatch mechanisms we see that the overall amount of change in annual dispatch by generation type is small relative to the amount of load that was shifted. For example, per the price-taking results, individual ERWHs and HPWHs shift about 710 GWh and 360 GWh, but the annual generation differences in the real-time PLEXOS model are only about 50 GWh and 30 GWh, respectively. This means that, when viewed through an annual net change lens, for the most part these resources are shifting energy between generators of the same type, e.g., from higher-price gas generators to lower-price gas generators. The

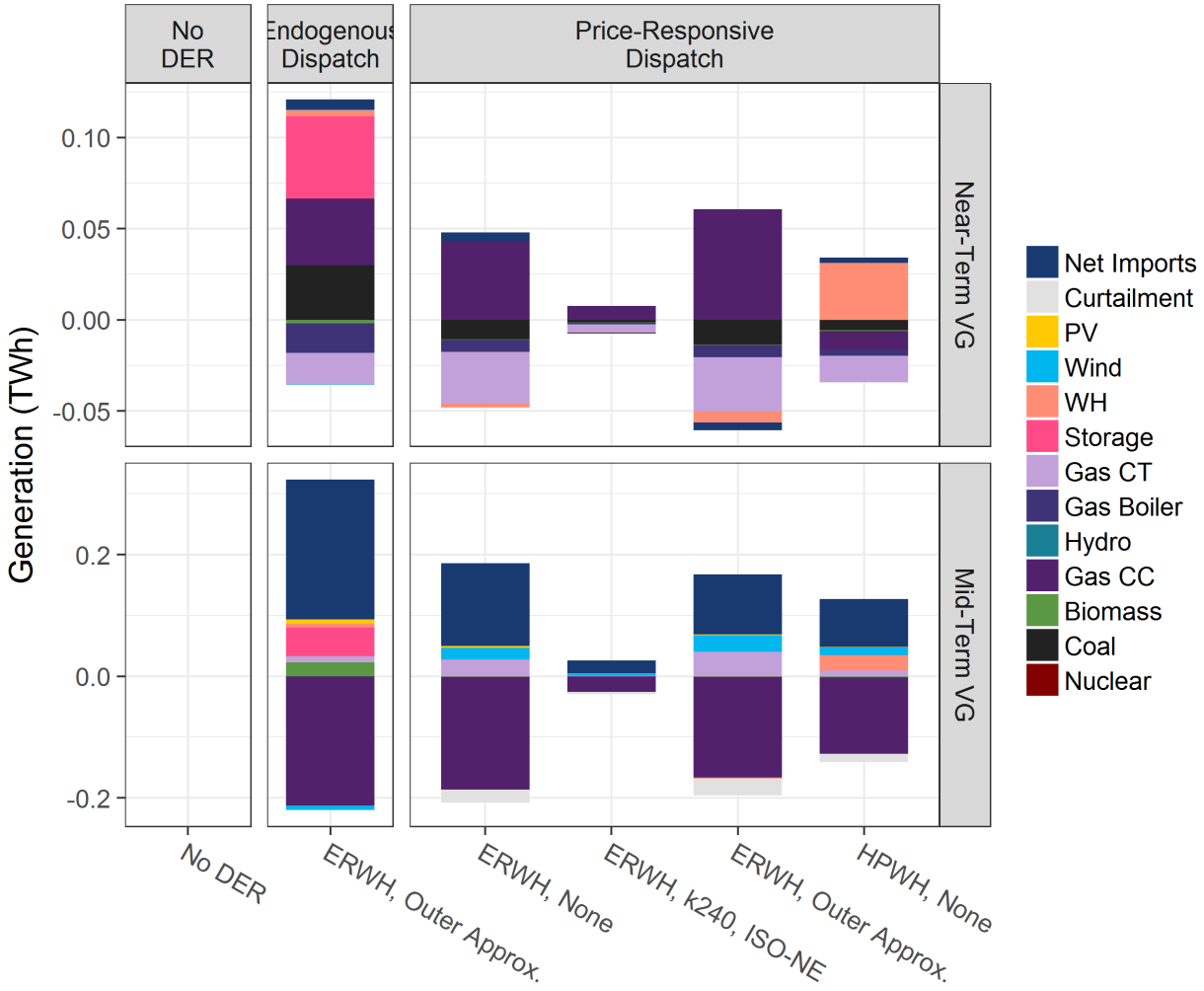


Figure 34. Annual generation differences (from the No DER case, Figure 11) in the Near-Term VG and Mid-Term VG ISO-NE models with energy shifting available from water heaters

annual generation by type changes that do result are qualitatively different depending on dispatch type and water heater type. The endogenously dispatched ERWH, Outer Approx. resources reduce (pumped hydro) storage losses, increase gas CC generation, increase coal generation, and decrease gas boiler and gas CT generation. Thus, these resources are able to displace a less efficient form of storage and make better use of gas CC and coal capacity that is already online (thereby preventing coal shut downs and gas generator start ups). Because price-responsive ERWH shifting is not able to change unit commitment decisions for coal nor dispatch decisions for pumped hydro storage, the individual and aggregated ERWHs dispatched against day-ahead prices increase gas CC generation, decrease coal and other gas generation, and do not impact the amount of storage losses in the system. Across the three ERWH aggregation types (None; k240, ISO-NE; and Outer Approx.), the types of changes seen in the price-responsive results are largely the same while magnitude of changes tracks the amount of shifting ability reflected by the different aggregations. The main effect of the individual HPWH shifting is to reduce the amount of electricity generation based on its 109% round-trip efficiency, which might be overstated because we did not model how shifting would change hourly average efficiency from baseline conditions.

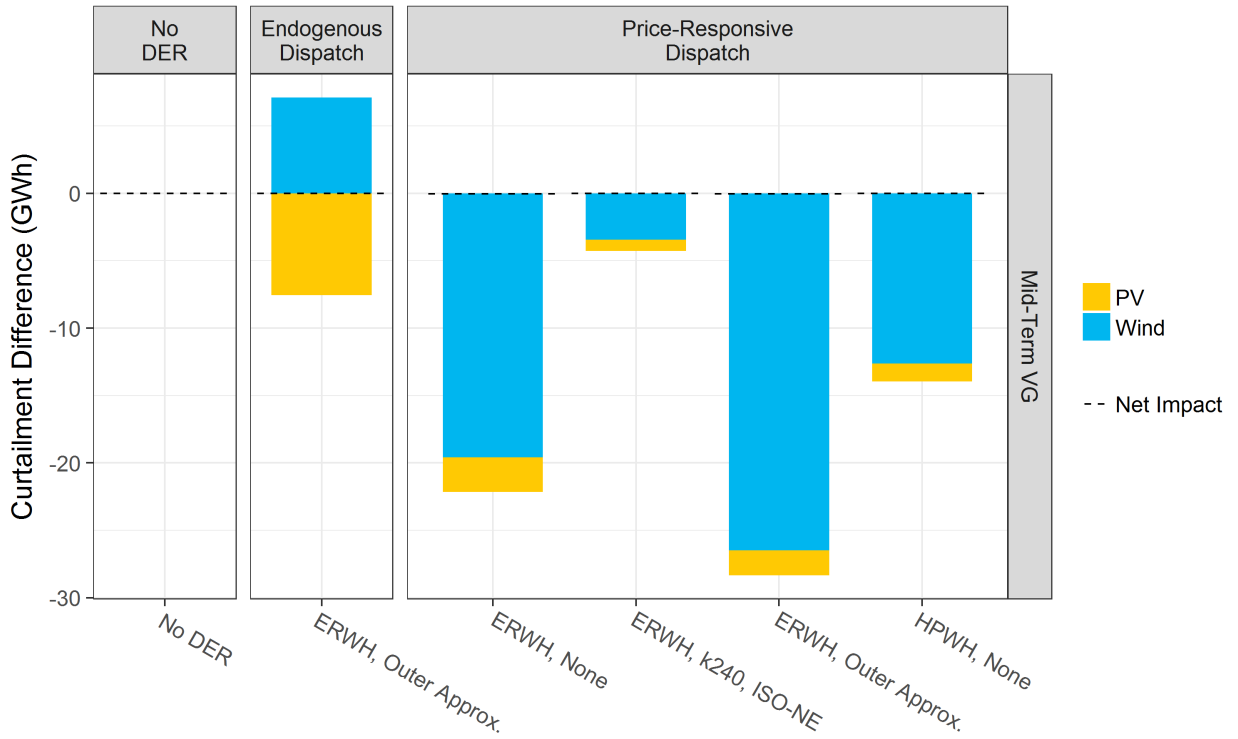


Figure 35. Annual curtailment differences (from the No DER case, Figure 11) in the Mid-Term VG ISO-NE model with energy shifting available from water heaters

Water heater energy shifting impacts supply-side dispatch more substantially under Mid-Term VG, as compared to Near-Term VG, conditions. In this case, price-taking dispatch of individual water heaters results in about 615 GWh and 315 GWh of ERWH and HPWH shifting, respectively, and annual generation by type differences have a similar order of magnitude—about 190 GWh and 130 GWh, respectively. Across all dispatch mechanisms and water heater types the predominant change is an increase in net imports and a decrease in gas CC generation. The increase in net imports actually corresponds to a decrease in exports—the water heater load shifting is relieving various physical and reliability constraints in the system that enables, e.g., within-system gas CC generation to be used locally rather than being exported to neighboring regions. Aside from this first-order effect, we do see different second-order impacts when we compare across dispatch types and water heater types. Endogenous dispatch of ERWHs leaves overall VG curtailment relatively static (Figure 35 shows a trade off between increased wind curtailment and decreased solar curtailment) while reducing the use of pumped hydro storage (shown as positive net generation from reduced storage losses in Figure 34) and increasing biomass and gas CT generation. Price-responsive dispatch of ERWHs generally reduces curtailment (increased wind and solar generation in Figure 34, details in Figure 35) and also increases gas CT generation. Finally, the two findings that (1) the ERWH inner approximation of aggregated resource (ERWH, k240, ISO-NE) has very little impact on annual generation by type and (2) reductions in electricity generation induced by 108% HPWH energy shifting round-trip efficiencies are both apparent in the Mid-Term VG results just as they were in the Near-Term VG results. By comparing Figure 35 to Figure 11 and Table 2 we see that the curtailment changes across all dispatch mechanisms and

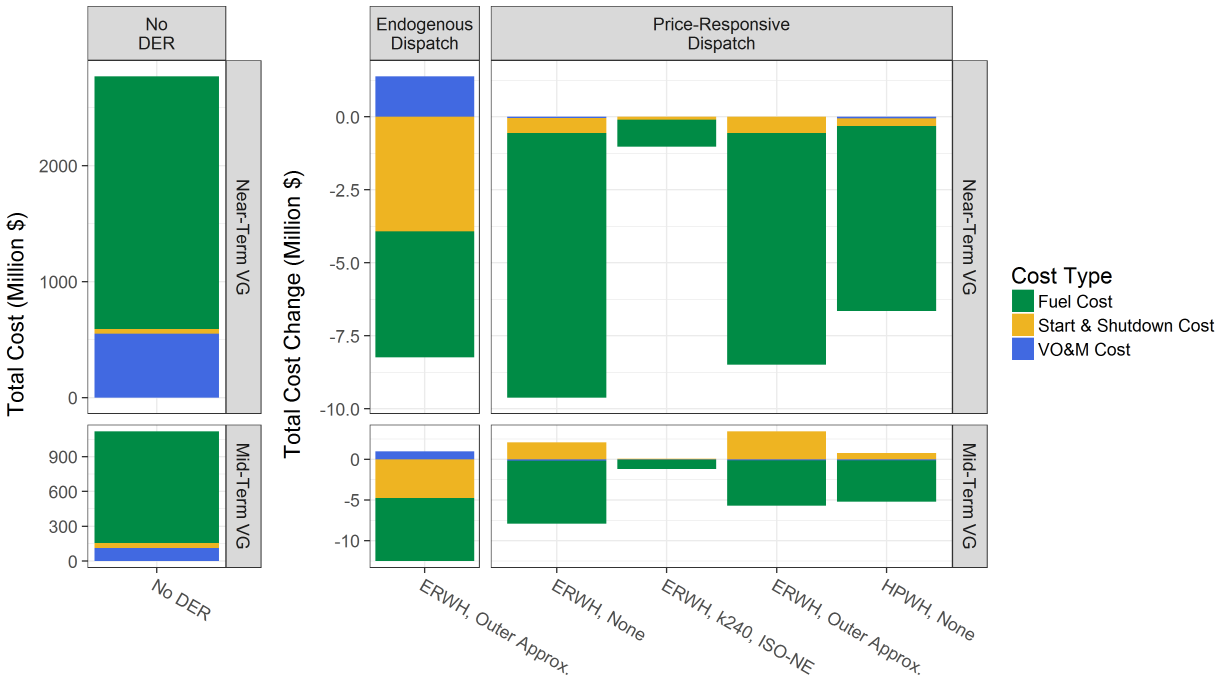


Figure 36. Total operational system costs for the No DER scenarios and differences in costs between real-time models with and without water heater energy shifting under Near-Term VG and Mid-Term VG conditions.

water heater types are small relative to the baseline amount of VG curtailment in the Mid-Term VG system, on the order of 1% or less.

The production cost savings from endogenous and price-responsive dispatch of electric water heater energy shifting are shown in Figure 36. One of the first things to notice is that endogenous dispatch of ERWH, Outer Approx. energy shifting achieves a lot of its savings in both models by reducing start and shutdown costs. This makes sense based on this resource being included in the day-ahead model where it can impact slow start unit commitment decisions, but it is still interesting to observe the direct comparison with price-responsive dispatch. Price-responsive dispatch reduces start and shutdown costs much less than does endogenous dispatch in the Near-Term VG model, and in the Mid-Term VG model price-responsive dispatch actually increases start and shutdown costs. In the Mid-Term VG model in particular, there are qualitative differences between how endogenous and price-responsive dispatch achieve some of their net production cost savings: endogenous dispatch of ERWH, Outer Approx. reduces start and shutdown costs as well as storage losses while keeping VG curtailment unchanged, and price-responsive dispatch of both water heater types reduces curtailment while actually increasing start and shutdown costs.

The overall system cost savings for all the dispatch mechanisms and water heater types are summarized in Table 9. We list the results for the ERWH k240, ISO-NE inner approximation for informational purposes, but note that its overly conservative shifting bounds result in production cost savings an order of magnitude smaller than all of the other flexibility model types, and so do not discuss it further. For the Near-Term VG model, the potential savings range from $6.4 million/yr to $9.3 million/yr which are 0.2% to 0.3% of the total production costs. The Mid-Term VG savings range from $2.0 million/yr to $11.4 million/yr, or 0.2% to 1.0% of total production

Table 9. Production Cost Savings Summary for Most Effective Energy Shifting Resources

Metric	Shifting Resource	Near-Term VG	Mid-Term VG
Total Cost (million $)	N/A	2800	1200
Total Savings (million $)	Endogenous ERWH, Outer Approx.	6.7	11.4
	Price-responsive ERWH, None	9.3	5.6
	Price-responsive ERWH, k240, ISO-NE	1.0	1.2
	Price-responsive ERWH, Outer Approx.	8.2	2.0
	Price-responsive HPWH, None	6.4	4.4
Savings ($/WH-year)	Endogenous ERWH, Outer Approx.	11.0	18.8
	Price-responsive ERWH, None	15.5	9.3
	Price-responsive ERWH, k240, ISO-NE	1.6	2.0
	Price-responsive ERWH, Outer Approx.	13.6	3.3
	Price-responsive HPWH, None	10.4	7.1

costs. (The cost of operating ISO-NE under Mid-Term VG conditions is less than half that of operating under Near-Term VG conditions because of increased zero marginal cost generation; therefore the same amount of savings translates into higher percent savings in the Mid-Term VG, as compared to to the Near-Term VG, models.)

As expected, individual HPWHs performing price-responsive dispatch produce less system cost savings than ERWHs dispatched in the same manner. However, as in the price-taking results they produce more savings than the size of their load would indicate. That is, HPWHs present about 40% of the load for shifting as compared to ERWHs (Table 7), but their cost savings are about 70% of the ERWHs' in both the Near-Term VG and Mid-Term VG models (Table 9). The relative rankings of the various ERWH results (other than the price-responsive inner aggregation) are not entirely as expected. Overall one would expect the endogenously dispatched outer approximation to produce the most system cost savings, which did happen in the Mid-Term VG model but did not happen in the Near-Term VG model. One would also expect the price-responsive outer approximation to out-perform price-responsive individual water heaters, but this does not happen for either the Near-Term VG model nor the Mid-Term VG model. It is beyond the scope of this report to fully explain these observations, but (a) we saw even in the price-taking results that at what level in the system (e.g., node, dispatch zone, or load region) flexibility is modeled can significantly impact outcomes, and (b) ISO-NE is a complex system with connections to neighboring regions that were difficult to capture in our PLEXOS models. Nonetheless, despite these idiosyncrasies the results do demonstrate that aggregated ERWH flexibility can be dispatched by PLEXOS and that doing so qualitatively changes the types of impacts such flexibility can have on the system as compared to purely price-responsive mechanisms that are not anticipated in load forecasts.

6 Conclusions

The research documented in this report developed and tested new methods for representing how electric water heaters can be operated flexibly in response to grid signals. The methods start with detailed building energy models as created by ResStock and then summarize the flexibility of water heaters in small, dynamic, battery-like surrogate models. The surrogate model forms were designed to be simple enough to include in linear and mixed-integer programs, and to be potentially aggregatable to the MW-scale, both of which are required for the models to be compatible with typical bulk grid investment and operational planning optimization formulations. We demonstrated the methodology in the context of (a) New England housing stock as modeled by ResStock and (b) prospective New England power grids as modeled in PLEXOS by extracting the ISO-NE system from the Interconnections Seam Study. We examined two scenarios for both the housing stock and the grid. The housing stock was modeled both with default ResStock parameters and with all electric water heaters replaced with 80 gallon HPWH models. The grid was modeled under both Near-Term VG and Mid-Term VG conditions.

In comparing the ability of ERWHs and HPWHs to provide contingency reserve, we found that the modeled representations of ERWHs and HPWHs responded differently to prescribed claim events and thus called for slightly different surrogate modeling methods. The main differences between how ERWHs and HPWHs were modeled for this study is that the HPWHs have larger absolute storage volumes (80 gallons compared to 50 gallons or less) and relative heating capacities (due to unintentionally modeling ERWHs with heating capacities more than three times less than typical) compared to ERWHs than is typical. Higher storage volume and higher heating capacity results in less frequent heating cycles and greater stratification (meaning that the temperature at the upper element can be significantly higher than both the average tank temperature and the temperature at the lower element). And because the modeled HPWH control logic weights the upper element temperature more heavily than the lower element temperature (at a ratio of 3:1), the modeled HPWHs are more likely than the modeled ERWHs to (1) be able to provide a full contingency response for the full duration of claim events; and (2) to outperform the predicted response of the TankT models (because the control temperature for HPWHs is often higher than the average tank temperature). On the other hand, ERWHs are much more likely to cycle due to low temperature at the lower element (which is weighted heavily by the ERWH controls and is often less than the average tank temperature). Practically, these differences resulted in different surrogate modeling assumptions for the two different water heater types–"TankT" assumptions for ERWHs to capture their limited ability to shed load due to lower storage volumes and lower control temperatures, and "Simple" assumptions for HPWHs to represent that they were generally able to take advantage of higher storage volumes and higher control temperatures to ride through almost all claim events. These different assumptions resulted in better matches between the contingency responses simulated with EnergyPlus and our surrogate models, but this result is likely dependent on specific equipment configurations and control logic. That is, "TankT" for ERWHs and "Simple" for HPWHs is likely more a reflection of the specific ERWH and HPWH systems modeled in this study rather than a general description of HPWH technology versus ERWH technology. The fact that storage volume and implemented control logic varies significantly across the stock of real-world water heaters indicates that demand flexibility potential (and corresponding model representation) will be device-dependent and cannot be generalized by technology type.

We estimated contingency reserve resource by aggregating individual water heater surrogate flexibility models and then simulating a contingency response for each hour of the year. This method is much less computationally intensive and more transferable to different types of contingency response than is directly simulating all possible contingency events in EnergyPlus with ResStock. As part of validating our surrogate models and aggregation methods, we demonstrated that we can compute aggregate models that bound the actual responses observed from directly dispatching individual water heaters with EnergyPlus for a few example events. The naïve aggregation method of simply summing up individual bounds to construct aggregated bounds forms an "outer approximation" that strictly overestimates actual flexibility. The provably dispatchable "inner approximations," on the other hand, are conservative in that individually dispatching water heaters with full information can provide a much larger response than simply dispatching the "inner approximation" aggregate. Given the aforementioned differences in simulated behavior between the ERWHs and the HPWHs, using an inner, as opposed to an outer, approximation for the ERWHs was very important (so as not to overstate the actual amount of resource), but was less so for the HPWHs because they tended not to be energy-limited (due to higher storage volume and higher control temperatures) when providing 30 minute responses for ISO-NE Claim30 events or 50 minute responses for ISO-NE Claim10 events. Relatedly, ERWH resource was significantly attenuated by the aggregation process which made it so that HPWHs were sometimes estimated to be able to provide more contingency resource than ERWHs even though their baseline load is 60% smaller. For example, we estimated ERWH Claim10 resource to be 334 GW-h/yr to 488 GW-h/yr, and HPWH Claim10 resource to be 567 GW-h/yr to 595 GW-h/yr, depending on the aggregation level. The amount of contingency reserve resource varied diurnally and seasonally for both types of water heaters. Estimated at the nodal level, the ERWH resource interquartile ranges were 45 MW–69 MW for Claim10 and 65 MW–102 MW for Claim30. At the Load Region level, the HPWH interquartile ranges were 48 MW–88 MW for Claim10 and 52 MW–90 MW for Claim30.

The process to produce inner approximation aggregations relies on assigning each individual resource a fraction of the aggregate dispatch signal and then setting the aggregate's bounds so that after the signal is partitioned out to individual devices none of the individual bounds are violated. This worst-case bounds construction places several practical constraints on its application that come to the fore when considering how to create aggregate resources for energy shifting. First, the disaggregation fractions need to be held constant over the response period (e.g., a whole day) to ensure that each individual water heater is actually returned to a known (baseline) state before being asked to shift energy again. Second, worst-case bound setting means that if a single individual resource is unable to move in a particular direction (e.g., cannot increase power or cannot reduce the amount of thermal energy stored in its tank), the entire aggregation will be prohibited from moving in that direction. Finally, time-varying and path-dependent parameters such as HPWH electricity to thermal energy conversion efficiency and maximum power consumption do not carry neatly through the aggregation mathematics. These limitations caused us to only look at direct dispatch of aggregated energy shifting resources in grid models for the case of ERWHs. When constructing inner (as opposed to outer) approximations we also had to group ERWHs primarily by power and energy profiles, rather than by geography, prior to aggregation to ensure that resources would end up with some ability to actually move energy from higher- to lower-price times.

Although our ability to investigate direct dispatch of aggregated electric water heater flexibility in large-scale grid models was more limited than we would have liked, we did dispatch individual surrogate models against day-ahead energy prices, inject the resulting load changes into real-time PLEXOS models of ISO-NE, and then observe the resulting changes in supply-side dispatch and costs. The aggregate resources we were able to include directly in the day-ahead models were then also run through the same real-time models to produce comparable results. In a price-taking sense, dispatching individual and aggregate flexibility models against PLEXOS day-ahead prices produced profits of about $20/WH-yr for individual HPWHs, about $38/WH-yr for individual ERWHs and ERWHs aggregated using an outer approximation, and up to $4/WH-yr for ERWHs aggregated using an inner approximation. When the supply-side response to these changes and the endogenous (price-making) dispatch of ERWH flexibility in the PLEXOS day-ahead models is accounted for, the grid system cost savings ranged from $10.4/WH-yr to $15.5/WH-yr under Near-Term VG conditions and from $3.3/WH-yr to $18.8/WH-yr under Mid-Term VG conditions. Overall these energy shifting experiments showed that while outer approximations of aggregate ERWH shiftability do overstate flexibility, the magnitude of overestimation can be modest (i.e., less than 20%) and the gain of being able to include that flexibility in bulk grid models could be worth it because ERWHs can then influence day-ahead unit commitment and other longer-horizon planning decisions. In contrast, the severe conservatism of the inner approximations of aggregate ERWH flexibility make them unattractive in the energy shifting context. Finally, while we were not able to aggregate individual HPWH flexibility models, dispatching them directly against prices showed that in the energy shifting, as in the contingency, context, the HPWHs in this study were able to provide more value compared to ERWHs than one would expect simply based on the relative amounts of electricity both types of water heaters consume.

This study provided an unprecedentedly detailed look at how electric resistance and heat pump water heaters could provide contingency reserve and energy shifting services in future grid systems. Nonetheless, several areas of investigation remain and our results are subject to several caveats. While this study did use realistic models of future ISO-NE power systems, such models are known to produce less volatile prices than real-world operations, such that our estimates of value might be understated. We also did not capture fully detailed 5-minute grid operations with forecast errors, nor attempt to adjust energy-shifting and other storage dispatch at that timescale. Finally, more heuristic and less provably dispatchable aggregation methods, as well as aggregation methods suitable for end-uses with significantly time-varying performance characteristics like HPWHs, might be able to unlock more price-making value for energy shifting and other grid services.

References

Alstone, P., et al. 2017. *2025 California Demand Response Potential Study-Charting California's Demand Response Future: Final Report on Phase 2 Results*. Technical Report LBNL-2001113. Berkeley, California: Lawrence Berkeley National Laboratory. http://eta-publications.lbl.gov/sites/default/files/lbnl-2001113.pdf.

Bloom, A., et al. 2016. *Eastern Renewable Generation Integration Study*. Tech. rep. NREL/TP-6A20-64472. Golden, CO: NREL. http://www.nrel.gov/docs/fy16osti/64472.pdf.

Bloom, A., et al. 2021. "The Value of Increased HVDC Capacity Between Eastern and Western U.S. Grids: The Interconnections Seam Study". Conference Name: IEEE Transactions on Power Systems, *IEEE Transactions on Power Systems*: 1–1. ISSN: 1558-0679. doi:10.1109/TPWRS.2021.3115092.

BPA. 2018. *CTA-2045 Water Heater Demonstration Report including A Business Case for CTA-2045 Market Transformation*. Technical Report BPA Technology Innovation Project 336. Bonneville Power Administration (BPA). https://www.bpa.gov/EE/Technology/demand-response/Documents/Demand%20Response%20-%20FINAL%20REPORT%20110918.pdf.

Carew, N., B. Larson, L. Piepmeier, and M. Logsdon. 2018. *Heat pump water heater electric load shifting: A modeling study*. Tech. rep. Ecotope, Inc. https://ecotope-publications-database.ecotope.com/2018_001_HPWHLoadShiftingModelingStudy.pdf.

Cochran, J., et al. 2021. *The Los Angeles 100% Renewable Energy Study (LA100)*. Tech. rep. NREL/TP-6A20-79444. National Renewable Energy Lab. (NREL), Golden, CO (United States). Visited on 01/05/2022. doi:10.2172/1774871. https://www.osti.gov/biblio/1774871.

Cohen, S. M., et al. 2019. *Regional Energy Deployment System (ReEDS) Model Documentation: Version 2018*. Tech. rep. NREL/TP-6A20-72023. National Renewable Energy Lab. (NREL), Golden, CO (United States). Visited on 01/05/2022. doi:10.2172/1505935. https://www.osti.gov/biblio/1505935/.

Cole, W. J., J. D. Rhodes, W. Gorman, K. X. Perez, M. E. Webber, and T. F. Edgar. 2014. "Community-scale residential air conditioning control for effective grid management". *Applied Energy* 130:428–436. Visited on 04/11/2017. http://www.sciencedirect.com/science/article/pii/S0306261914005728.

Cutler, D., J. Winkler, N. Kruis, C. Christensen, and M. Brandemuehl. 2013. *Improved Modeling of Residential Air Conditioners and Heat Pumps for Energy Calculations*. Technical Report NREL/TP-5500-56354. Golden, CO (United States): National Renewable Energy Laboratory (NREL). https://www.nrel.gov/docs/fy13osti/56354.pdf.

Denholm, P., Y. Sun, and T. Mai. 2019. *Introduction to Grid Services: Concepts, Technical Requirements, and Provisions from Wind*. Technical report NREL/TP-6A20-72578. National Renewable Energy Laboratory (NREL). https://www.nrel.gov/docs/fy19osti/72578.pdf.

Energy Exemplar. *PLEXOS*. http://energyexemplar.com/software.

EPRI. 2018. *U.S. National Electrification Assessment*. Technical Report. Electric Power Research Institute (EPRI). https://www.epri.com/research/products/000000003002013582.

FERC. 2020. *Participation of Distributed Energy Resource Aggregations in Markets Operated by Regional Transmission Organizations and Independent System Operators [Docket No. RM18-9-000; Order No. 2222]*. https://www.ferc.gov/sites/default/files/2020-09/E-1_0.pdf.

Garfield, D. J., P. N. Jadun, S. Hwang, M. O'Malley, and M. F. Ruth. 2021. *Opportunities for Industry to Provide Flexibility While Increasing Profitability*. Tech. rep. NREL/TP-6A50-75784. National Renewable Energy Lab. (NREL), Golden, CO (United States). Visited on 11/01/2021. https://www.osti.gov/biblio/1822824/.

Hale, E., et al. 2021. *Chapter 3: Electricity Demand Projections*. Technical Report NREL/TP-6A20-79444-3. Golden, CO (United States): National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy21osti/79444-3.pdf.

Hale, E. T., B. L. Stoll, and J. E. Novacheck. 2018. "Integrating solar into Florida's power system: Potential roles for flexibility". *Solar Energy* 170 (): 741–751. ISSN: 0038-092X, visited on 06/22/2018. doi:10.1016/j.solener.2018.05.045. https://www.sciencedirect.com/science/article/pii/S0038092X1830478X.

Hao, H., B. M. Sanandaji, K. Poolla, and T. L. Vincent. 2013. "A generalized battery model of a collection of thermostatically controlled loads for providing ancillary service". In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, 551–558. IEEE.

Hledik, R., J. Chang, and R. Lueken. 2016. *The Hidden Battery: Opportunities in Electric Water Heating*. Tech. rep. Prepared by the Brattle Group for NRECA, NRDC, and PLMA. https://www.electric.coop/wp-content/uploads/2016/07/The-Hidden-Battery-01-25-2016.pdf.

Hopkins, A. S., and M. Whited. 2017. *Best Practices in Utility Demand Response Programs: With Application to Hydro-Québec's 2017-2026 Supply Plan*. Technical Report. Cambridge, Massachusetts: Synapse Energy Economics Inc. https://www.synapse-energy.com/sites/default/files/Utility-DR-17-010.pdf.

Hummon, M., D. Palchak, P. Denholm, and J. Jorgenson. 2013. *Grid Integration of Aggregated Demand Response, Part 2: Modeling Demand Response in a Production Cost Model*. Technical Report NREL/TP-6A20-58492. Golden, CO: NREL (National Renewable Energy Laboratory (NREL). http://www.nrel.gov/docs/fy14osti/58492.pdf.

ISO New England Inc. 2022. *Demand Resources*. Visited on 01/05/2022. https://www.iso-ne.com/markets-operations/markets/demand-resources/.

Jin, X., K. Baker, D. Christensen, and S. Isley. 2017. "Foresee: A user-centric home energy management system for energy efficiency and demand response". *Applied Energy* 205 (): 1583–1595. ISSN: 0306-2619, visited on 01/05/2022. doi:10.1016/j.apenergy.2017.08.166. https://www.sciencedirect.com/science/article/pii/S0306261917311856.

Langevin, J., C. B. Harris, A. Satre-Meloy, H. Chandra-Putra, A. Speake, E. Present, R. Adhikari, E. J. H. Wilson, and A. J. Satchwell. 2021. "US building energy efficiency and flexibility as an

electric grid resource". *Joule* 5, no. 8 (): 2102–2128. ISSN: 2542-4351, visited on 09/15/2021. doi:10.1016/j.joule.2021.06.002. https://www.sciencedirect.com/science/article/pii/S2542435121002907.

Lescoeur, B., and J. B. Galland. 1987. "Tariffs and Load Management : The French Experience". Conference Name: IEEE Transactions on Power Systems, *IEEE Transactions on Power Systems* 2, no. 2 (): 458–464. ISSN: 1558-0679. doi:10.1109/TPWRS.1987.4335150.

Luo, N., J. Langevin, and H. Chandra Putra. 2021. "Quantifying the effect of multiple demand response actions on electricity demand and building services via surrogate modeling" (). Visited on 01/05/2022. https://escholarship.org/uc/item/0t72t7sk.

MacDonald, J. S., E. Vrettos, and D. S. Callaway. 2020. "A Critical Exploration of the Efficiency Impacts of Demand Response From HVAC in Commercial Buildings". Publisher: IEEE, *Proceedings of the IEEE*.

Maguire, J. 2018. *Verifying Simulation Models of Water Heaters*. https://www.aceee.org/sites/default/files/pdf/conferences/hwf/2018/1b-jmaguire.pdf.

Mayhorn, E., S. Widder, S. Parker, R. Pratt, and F. Chassin. 2015. *Evaluation of the demand response performance of large capacity electric water heaters*. Technical Report PNNL-23527. Richland, Washington: Pacific Northwest National Laboratory. https://labhomes.pnnl.gov/documents/PNNL_23527_Eval_Demand_Response_Performance_Electric_Water_Heaters.pdf.

McPherson, M., and B. Stoll. 2020. "Demand response for variable renewable energy integration: A proposed approach and its impacts". *Energy* 197 (): 117205. ISSN: 0360-5442, visited on 08/04/2020. doi:10.1016/j.energy.2020.117205. http://www.sciencedirect.com/science/article/pii/S0360544220303121.

Mitra, S., I. E. Grossmann, J. M. Pinto, and N. Arora. 2012. "Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes". *Computers & Chemical Engineering* 38 (): 171–184. ISSN: 0098-1354, visited on 11/01/2021. doi:10.1016/j.compchemeng.2011.09.019. https://www.sciencedirect.com/science/article/pii/S0098135411003012.

Murphy, C., T. Mai, Y. Sun, P. Jadun, M. Muratori, B. Nelson, and R. Jones. 2021. *Electrification Futures Study: Scenarios of Power System Evolution and Infrastructure Development for the United States*. Tech. rep. NREL/TP-6A20-72330. National Renewable Energy Laboratory (NREL); Northern Arizona University; Evolved Energy Research. Visited on 01/05/2022. doi:10.2172/1762438. https://www.osti.gov/biblio/1762438.

Neukomm, M., V. Nubbe, and R. Fares. 2019. *Grid-interactive Efficient Buildings Technical Report Series: Overview*. Technical Report. Washington, DC, USA: U.S. DOE Office of Energy Efficiency and Renewable Energy. https://www1.eere.energy.gov/buildings/pdfs/75470.pdf.

Nichols, J., and K. Haag. 2018. *Demand Response Auditing*. Webinar Training. https://www.iso-ne.com/static-assets/documents/2018/03/20180327-dr-auditing.pdf.

Nubbe, V., K. Lee, A. Valdez, E. Barbour, and J. Langevin. 2021. *Grid-Interactive Efficient Building Technology Cost, Performance, and Lifetime Characteristics*. Tech. rep. LBNL-2001375. Berkeley, CA (United States): Lawrence Berkeley National Laboratory (LBNL). Visited on 01/05/2022. https://escholarship.org/uc/item/44t4c2v6.

O'Connell, N., E. Hale, I. Doebber, and J. Jorgenson. 2015. *On the Inclusion of Energy-Shifting Demand Response in Production Cost Models: Methodology and a Case Study*. Technical Report NREL/TP-6A20-64465. Golden, Colorado: National Renewable Energy Laboratory.

Opalka, B. 2013. "How the unheralded water heater is on the cutting-edge of demand response". *Utility Dive* (). Visited on 09/23/2019. https://www.utilitydive.com/news/how-the-unheralded-water-heater-is-on-the-cutting-edge-of-demand-response/193871/.

Oschsner, H., J. Stewart, K. Bushman, J. Keeling, and H. Haeri. 2011. *Kootenai DR Pilot Evaluation: Full Pilot Report*. Technical Report. Prepared by The Cadmus Group, Inc. for Bonneville Power Admininstration. https://www.bpa.gov/EE/Technology/demand-response/Documents/Final%20Evaluation%20Report%20for%20KEC%20Peak%20Project_28Dec11.pdf.

Potter, J., and P. Cappers. 2017. *Demand Response Advanced Controls Framework and Assessment of Enabling Technology Costs*. Technical Report LBNL-2001044. Berkeley, California: Lawrence Berkeley National Laboratory. Visited on 07/10/2018. https://escholarship.org/uc/item/0229m0q2.

Shah, M., D. Cutler, J. Maguire, Z. Peterson, X. Li, J. Reyna, and J. Pohl. 2020. *Metrics and Analytical Frameworks for Valuing Energy Efficiency and Distributed Energy Resources in the Built Environment: Preprint*. Tech. rep. NREL/CP-6A20-77888. National Renewable Energy Lab. (NREL), Golden, CO (United States). Visited on 01/05/2022. https://www.osti.gov/biblio/1710148.

Stephen, G., E. Hale, and B. Cowiestoll. 2020. *Managing Solar Photovoltaic Integration in the Western United States: Resource Adequacy Considerations*. Technical Report NREL/TP-6A20-72472. Golden, CO: National Renewable Energy Laboratory.

Stoll, B., E. Buechler, and E. Hale. 2017. "The value of demand response in Florida". *The Electricity Journal*, Energy Policy Institute's Seventh Annual Energy Policy Research Conference, 30, no. 9 (): 57–64. ISSN: 1040-6190, visited on 12/11/2017. doi:10.1016/j.tej.2017.10.004. http://www.sciencedirect.com/science/article/pii/S1040619017302609.

Sun, Y., P. Jadun, B. Nelson, M. Muratori, C. Murphy, J. Logan, and T. Mai. 2020. *Electrification Futures Study: Methodological Approaches for Assessing Long-Term Power System Impacts of End-Use Electrification*. Technical Report NREL/TP-6A20-73336. Golden, CO (United States): National Renewable Energy Laboratory (NREL). https://www.nrel.gov/docs/fy20osti/73336.pdf.

U.S. DOE EERE. 2019. *Low-Income Energy Affordability Data (LEAD) Tool*. Dataset. Visited on 03/02/2020. https://data.openei.org/submissions/573.

Wang, J. 2021. "Occupant-centric Modeling and Control for Low-Carbon and Resilient Communities". PHD Thesis, Architectural Engineering.

Wilson, E., C. Christensen, S. Horowitz, and H. Horsey. 2016. “A High-Granularity Approach to Modeling Energy Consumption and Savings Potential in the U.S. Residential Building Stock”. *IBPSA-USA Journal* 6 (1).

Zhou, E., and T. Mai. 2021. *Electrification Futures Study: Operational Analysis of U.S. Power Systems with Increased Electrification and Demand-Side Flexibility*. Tech. rep. NREL/TP-6A20-79094. National Renewable Energy Lab. (NREL), Golden, CO (United States). Visited on 01/05/2022. doi:10.2172/1785329. https://www.osti.gov/biblio/1785329.

Appendix A. Aggregation Mathematics

Hao et al. (2013) has demonstrated methods for creating generalized battery models that represent aggregates of thermostatically controlled loads. Here we generalize their methods by accounting more explicitly for the storage characteristics of water heaters (and other DER resources) and the possibility of using these aggregation methods for longer-duration grid services than their context, which was frequency regulation. To do this, we start with a collection of water heaters (or similar DERs) with individual devices identified with indices $k \in \mathbb{K}$ and express their individual shiftability as:

$$\frac{d\Delta S_k(t)}{dt} = \eta_k(t)\Delta P_k(t) - \alpha_k(t)\Delta S_k(t) + \gamma_k(t) \tag{A.1}$$

$$\underline{\Delta S_k}(t) \leq \Delta S_k(t) \leq \overline{\Delta S_k}(t) \tag{A.2}$$

$$\underline{\Delta P_k}(t) \leq \Delta P_k(t) \leq \overline{\Delta P_k}(t) \tag{A.3}$$

Then we seek a similar model formulation to represent their aggregate flexibility:

$$\frac{dL(t)}{dt} = \eta(t)U(t) - \alpha(t)L(t) + \gamma(t) \tag{A.4}$$

$$\underline{L}(t) \leq L(t) \leq \overline{L}(t) \tag{A.5}$$

$$\underline{U}(t) \leq U(t) \leq \overline{U}(t) \tag{A.6}$$

Following Hao et al. (2013), the general procedure for estimating the parameters $\eta(t)$, $\alpha(t)$, $\gamma(t)$, $\underline{S}(t)$, $\overline{S}(t)$, $\underline{U}(t)$, and $\overline{U}(t)$ required to define the aggregate model is relatively straightforward in the Laplace domain, not in an exact sense, but in the sense of being able to create bounding models. Because we are most concerned with the flexibility realized in terms of the aggregate shifted energy

$$\sum_k w_k \Delta P_k(t), \tag{A.7}$$

where w_k is a sample weight for individual resource k, the focus of model construction is on allowable shifting profiles at both the individual and aggregate levels. That is, if we define the sets of possible realizations at the individual level to be

$$\mathbb{P}_k = \left\{\Delta P_k(t) \,|\, \exists \Delta S_k(t) \text{ with individual model (A.1) - (A.3) for device } k \in \mathbb{K} \text{ satisfied}\right\}, \tag{A.8}$$

and the possible realizations for the aggregate to be

$$\mathbb{U}(t) = \left\{ \sum_k w_k \Delta P_k(t) \,\middle|\, \Delta P_k(t) \in \mathbb{P}_k, \forall k \in \mathbb{K} \right\}, \tag{A.9}$$

we can work to define pairs of generalized storage model parameters for which the resulting models define sets of allowable power profiles $U(t)$ that are either strictly smaller or strictly larger than $\mathbb{U}(t)$.

To this end, let

$$\Phi = \left(\eta(t), \alpha(t), \gamma(t), \underline{L}(t), \overline{L}(t), \underline{U}(t), \overline{U}(t)\right) \tag{A.10}$$

be any realization of the aggregate storage model and

$$\mathbb{U}_{\Phi}(t) = \{U(t) \,|\, \exists L(t) \text{ with the aggregate model (A.4) - (A.6) defined by } \Phi \text{ satisfied}\}. \quad \text{(A.11)}$$

Then we seek $\underline{\Phi}$ and $\overline{\Phi}$ such that

$$\mathbb{U}_{\underline{\Phi}} \subset \mathbb{U} \subset \mathbb{U}_{\overline{\Phi}}. \quad \text{(A.12)}$$

Thus $\underline{\Phi}$ represents a lower-bound, inner-approximation, sufficient estimate of $\mathbb{U}$; whereas $\overline{\Phi}$ represents an upper-bound, outer-approximation, necessary estimate of $\mathbb{U}$, and the size of the difference between them is a metric of estimate tightness.

A.1 Aggregation of Individual Device Models in the Constant Parameters Case

In the case of the parameters η, α, and γ in (A.1) being time-invariant, the techniques of Hao et al. (2013) are easily extended to this more-general context.

Theorem 1. ***Outer approximation of general model with constant parameters.***

Assume η_k, α_k and γ_k in (A.1) are time-invariant for all individual devices $k \in \mathbb{K}$. Then with constants η and α arbitrarily chosen and

$$\gamma = \sum_k w_k \frac{\eta}{\eta_k} \gamma_k \quad \text{(A.13)}$$

the aggregate model represented by (A.4) - (A.6) with

$$\overline{L}(t) = -\underline{L}(t) = \sum_k w_k \frac{\eta}{\eta_k} \left(1 + \left|1 - \frac{\alpha_k}{\alpha}\right|\right) \max\left(|\underline{\Delta S_k}(t)|, |\overline{\Delta S_k}(t)|\right), \quad \text{(A.14)}$$

$$\underline{U}(t) = \sum_k w_k \underline{\Delta P_k}(t), \quad \overline{U}(t) = \sum_k w_k \overline{\Delta P_k}(t), \quad \text{(A.15)}$$

defines an outer-approximation, necessary model $\mathbb{U}_{\overline{\Phi}}(t)$ for $\mathbb{U}(t)$. That is,

$$\overline{\Phi} = \left(\eta, \alpha, \sum_k w_k \frac{\eta}{\eta_k} \gamma_k, \underline{L}(t), \overline{L}(t), \underline{U}(t), \overline{U}(t)\right), \quad \text{(A.16)}$$

with $\underline{L}(t)$ and $\overline{L}(t)$ defined in (A.14), and $\underline{U}(t)$ and $\overline{U}(t)$ defined in (A.15) constitutes an outer bound such that $\mathbb{U}(t) \subset \mathbb{U}_{\overline{\Phi}}(t)$.

Proof of Theorem 1. First, we take the Laplace transform of (A.1) after dropping the time-dependence of η_k, α_k, and γ_k:

$$s\Delta S_k(s) = \eta_k \Delta P_k(s) - \alpha_k \Delta S_k(s) + \frac{\gamma_k}{s} \quad \text{(A.17)}$$

and solve for $\Delta S_k(s)$:

$$\Delta S_k(s) = \frac{\eta_k}{s + \alpha_k} \Delta P_{c,k}(s) + \frac{\gamma_k}{s(s + \alpha_k)} \quad \text{(A.18)}$$

We then similarly transform (A.4):

$$sL(s) = \eta U(s) - \alpha L(s) + \frac{\gamma}{s}, \quad \text{(A.19)}$$

and solve for $L(s)$:

$$L(s) = \frac{\eta}{s+\alpha} U(s) + \frac{\gamma}{s(s+\alpha)}. \tag{A.20}$$

To connect these individual and aggregate models, we define

$$U(t) = \sum_k w_k \Delta P_k(t), \tag{A.21}$$

such that

$$U(s) = \sum_k w_k \Delta P_k(s). \tag{A.22}$$

This allows us to bring the individual models into (A.20) by first using (A.22):

$$L(s) = \frac{\eta}{s+\alpha} \sum_k w_k \Delta P_k(s) + \frac{\gamma}{s(s+\alpha)} \tag{A.23}$$

and then substituting in (A.18) solved for $\Delta P_{c,k}(s)$:

$$L(s) = \frac{\eta}{s+\alpha} \sum_k w_k \left(\frac{s+\alpha_k}{\eta_k} \Delta S_k(s) - \frac{\gamma_k}{s\eta_k} \right) + \frac{\gamma}{s(s+\alpha)}. \tag{A.24}$$

Rearranging we arrive at:

$$L(s) = \sum_k w_k \frac{\eta}{\eta_k} \left(1 + \frac{\alpha_k - \alpha}{s+\alpha} \right) \Delta S_k(s) + \frac{1}{s(s+\alpha)} \left(\frac{\gamma}{|\mathbb{K}|} - w_k \frac{\eta}{\eta_k} \gamma_k \right). \tag{A.25}$$

The upper-bound, necessary estimate $\overline{\Phi}$ is constructed by specifying energy, charging, and discharging capacities that envelop all possible realizations of the individual devices' shiftability. Thus we construct upper and lower bounds on charging capacity:

$$\underline{U}(t) = \sum_k w_k \underline{\Delta P_k}(t) \leq U(t) \leq \sum_k w_k \overline{\Delta P_k}(t) = \overline{U}(t). \tag{A.26}$$

To bound energy capacity, we first specify a value for η and then define γ as

$$\gamma = \sum_k w_k \frac{\eta}{\eta_k} \gamma_k \tag{A.27}$$

so that the final term of (A.25) drops out. We then let

$$H_k(s) = w_k \frac{\eta}{\eta_k} \left(1 + \frac{\alpha_k - \alpha}{s+\alpha} \right). \tag{A.28}$$

Then because

$$L(s) = \sum_k H_k(s) \Delta S_k(s), \tag{A.29}$$

we know that

$$L(t) = \sum_k H_k(t) \circledast \Delta S_k(t), \tag{A.30}$$

where $\circledast$ indicates the convolution operator. By the triangle inequality (first $\leq$) and Young's convolution inequality (second $\leq$) we can bound the aggregate energy capacity as in:

$$\|L(t)\|_r \leq \sum_k \|H_k(t) \circledast \Delta S_k(t)\|_r \leq \sum_k \|H_k(t)\|_p \|\Delta S_k(t)\|_q, \tag{A.31}$$

where

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1. \tag{A.32}$$

Inverting $H_k(s)$ we have

$$H_k(t) = w_k \frac{\eta}{\eta_k} \left(\delta(t) + (\alpha_k - \alpha) e^{-\alpha t}\right) \tag{A.33}$$

where $\delta(t)$ is the Dirac delta function. Then letting $p = 1$, $q = r = \infty$ we can compute

$$\|H_k(t)\|_1 = \int_0^\infty \left| w_k \frac{\eta}{\eta_k} \left(\delta(t) + (\alpha_k - \alpha) e^{-\alpha t}\right) \right| dt \tag{A.34}$$

$$= w_k \frac{\eta}{\eta_k} \left(1 + |\alpha_k - \alpha| \int_0^\infty e^{-\alpha t} dt\right) \tag{A.35}$$

$$= w_k \frac{\eta}{\eta_k} \left(1 + \frac{|\alpha_k - \alpha|}{\alpha}\right) \tag{A.36}$$

and specify

$$\|L(t)\|_\infty \leq \overline{L}(t) = -\underline{L}(t) = \sum_k w_k \frac{\eta}{\eta_k} \left(1 + \left|1 - \frac{\alpha_k}{\alpha}\right|\right) \max\left(|\underline{\Delta S_k}(t)|, |\overline{\Delta S_k}(t)|\right) \tag{A.37}$$

as an outer-bound on energy capacity. □

Theorem 2. ***Inner approximations of general model with constant parameters.***

Choose a β_k for each $k \in \mathbb{K}$ such that $0 \leq w_k \beta_k \leq 1$ and $\sum_k w_k \beta_k = 1$. Then with

$$\eta = \sum_k w_k \beta_k \eta_k \text{ and } \Delta P_k(t) = \beta_k U(t), \tag{A.38}$$

we see that

$$U(t) = \sum_k w_k \Delta P_k(t) \tag{A.39}$$

and we can show that $\mathbb{U}_{\underline{\Phi}}(t) \subset \mathbb{U}(t)$, where $\mathbb{U}(t)$ is defined by (A.9) and

$$\underline{\Phi} = \left(\sum_k w_k \beta_k \eta_k, \alpha, \gamma, -\overline{L}(t), \overline{L}(t), \max_k \frac{\underline{\Delta P_k}(t)}{\beta_k}, \min_k \frac{\overline{\Delta P_k}(t)}{\beta_k}\right), \tag{A.40}$$

where α and γ are arbitrarily chosen and

$$\overline{L}(t) \leq \frac{\alpha_k}{\alpha_k + |\alpha - \alpha_k|} \frac{\eta}{\beta_k \eta_k} \min\left(\frac{\Delta\gamma_k}{\alpha_k}\left(1 - e^{-\alpha_k t}\right) - \underline{\Delta S_k}(t), \overline{\Delta S_k}(t) - \frac{\Delta\gamma_k}{\alpha_k}\left(1 - e^{-\alpha_k t}\right)\right) \quad \forall k \in \mathbb{K}, \tag{A.41}$$

with

$$\Delta\gamma_k = \gamma_k - \frac{\beta_k \eta_k}{\eta} \gamma. \tag{A.42}$$

Proof of Theorem 2. Taking the Laplace transform of the expression for ΔP_k in (A.38) and substituting into (A.18) yields

$$\Delta S_k(s) = \eta_k \frac{\beta_k}{s+\alpha_k} U(s) + \frac{\gamma_k}{s(s+\alpha_k)}. \tag{A.43}$$

Rearranging (A.20) to solve for $U(s)$ and substituting in yields

$$\Delta S_k(s) = \beta_k \frac{\eta_k}{\eta} \left(1 + \frac{\alpha - \alpha_k}{s+\alpha_k}\right) L(s) + \frac{1}{s(s+\alpha_k)} \left(\gamma_k - \beta_k \frac{\eta_k}{\eta} \gamma\right). \tag{A.44}$$

Now let

$$G_k(s) = 1 + \frac{\alpha - \alpha_k}{s+\alpha_k} \text{ and } \Delta\gamma_k = \gamma_k - \frac{\beta_k \eta_k}{\eta} \gamma \tag{A.45}$$

and apply the inverse Laplace transform:

$$\Delta S_k(t) = \beta_k \frac{\eta_k}{\eta} G_k(t) \circledast L(t) + \frac{\Delta\gamma_k}{\alpha_k} \left(1 - e^{-\alpha_k t}\right). \tag{A.46}$$

Using Young's convolution inequality and the derivation of $H_k(t)$ in the proof of Theorem 1 we have

$$\|G_k(t) \circledast L(t)\|_\infty \leq \|G_k(t)\|_1 \|L(t)\|_\infty \tag{A.47}$$

$$= \left(1 + \frac{|\alpha - \alpha_k|}{\alpha_k}\right) \|L(t)\|_\infty. \tag{A.48}$$

Then to ensure that (A.2) is always satisfied we set

$$\|L(t)\|_\infty \leq \frac{\alpha_k}{\alpha_k + |\alpha - \alpha_k|} \frac{\eta}{\beta_k \eta_k} \min\left(\frac{\Delta\gamma_k}{\alpha_k}\left(1 - e^{-\alpha_k t}\right) - \underline{\Delta S_k}(t), \overline{\Delta S_k}(t) - \frac{\Delta\gamma_k}{\alpha_k}\left(1 - e^{-\alpha_k t}\right)\right) \quad \forall k \in \mathbb{K} \tag{A.49}$$

and similarly require

$$\underline{U_c}(t) \geq \frac{\underline{\Delta P_k}(t)}{\beta_k} \text{ and } \overline{U_c}(t) \leq \frac{\overline{\Delta P_k}(t)}{\beta_k} \quad \forall k \in \mathbb{K} \tag{A.50}$$

to satisfy (A.3). □

A.2 Electric Water Heaters

Electric resistance water heaters can be assumed to have constant η and α parameters, and a zero γ term as long as the baseline temperature set-point $\widetilde{T}$ is constant.[11] HPWHs are problematic on two counts. First, η and $\overline{P}$ vary with ambient (air surrounding the water heater) temperatures. This issue is mitigated if the water heater is located in a conditioned space with constant (or at least not widely-varying) temperature set-point. Second, η and $\overline{P}$ also vary with tank temperature. This is perhaps more problematic as flexible operation necessitates allowing the tank temperature to float over a wider range; but is mitigated if the range is limited. In what follows, we nevertheless treat η as a constant. What this means in practice in the body of the report is that we only apply aggregation methods to HPWHs for contingency service, and in that case we use

[11] Assuming a constant α also involves ignoring the dependence of water thermal capacitance on temperature.

hourly data and assume η and $\overline{P}$ are at the levels indicated by the baseline portfolio in each hour. In what follows, we also assume the bounds on acceptable tank temperatures, $\underline{T} \leq T(t) \leq \overline{T}$ are constant.

Under these simplifying assumptions, the shiftability of individual water heaters can be written as:

$$\frac{d\Delta S_k(t)}{dt} = \eta_k \Delta P_k(t) - \alpha_k \Delta S_k(t) \tag{A.51}$$

$$c_k\left(\underline{T}_k - \widetilde{T}_k\right) \leq \Delta S_k(t) \leq c_k\left(\overline{T}_k - \widetilde{T}_k\right) \tag{A.52}$$

$$-\widetilde{P}_k(t) \leq \Delta P_k(t) \leq \overline{P_k} - \widetilde{P}_k(t), \tag{A.53}$$

where $\alpha_k = 1/rc$. Thus, the parameters η_k and α_k in (A.51) are constant as are the bounds on $\Delta S_k(t)$, but the bounds on $\Delta P_k(t)$ are time-varying, as they depend on time-varying baseline operations.[12]

This model structure leads us to construct aggregate models of shiftability of the form:

$$\frac{dL(t)}{dt} = \eta U(t) - \alpha L(t) \tag{A.54}$$

$$\underline{L} \leq L(t) \leq \overline{L} \tag{A.55}$$

$$\underline{U}(t) \leq U(t) \leq \overline{U}(t) \tag{A.56}$$

Lemma 1. ***Outer approximation of aggregated electric water heaters.***

Assume η_k and α_k in (A.51) are time-invariant for all individual devices $k \in \mathbb{K}$. Then with constants η and α arbitrarily chosen the aggregate model represented by (A.54) - (A.56) with

$$\overline{L}(t) = -\underline{L}(t) = \sum_k w_k \frac{\eta}{\eta_k}\left(1 + \left|1 - \frac{\alpha_k}{\alpha}\right|\right) \max\left(c_k\left|\underline{T}_k - \widetilde{T}_k\right|, c_k\left|\overline{T}_k - \widetilde{T}_k\right|\right), \tag{A.57}$$

$$\underline{U}(t) = -\sum_k w_k \widetilde{P}_k(t), \quad \overline{U}(t) = \sum_k w_k\left(\overline{P_k} - \widetilde{P}_k(t)\right) \tag{A.58}$$

defines an outer-approximation, necessary model $\mathbb{U}_{\overline{\Phi}}(t)$ for $\mathbb{U}(t)$ as it is expressed in (A.9). That is, $\mathbb{U}(t) \subset \mathbb{U}_{\overline{\Phi}_{WH}}(t)$, where

$$\overline{\Phi}_{WH} = \left(\eta, \alpha, 0, \underline{L}(t), \overline{L}(t), \underline{U}(t), \overline{U}(t)\right), \tag{A.59}$$

with $\underline{L}(t)$ and $\overline{L}(t)$ defined in (A.57), and $\underline{U}(t)$ and $\overline{U}(t)$ defined in (A.58).

Proof of Lemma 1. Lemma 1 is a simplification of Theorem 1. □

[12] This representation therefore represents the Simple form of the surrogate model. In the TankT form of the surrogate model $\widetilde{T}_k(t)$ is the time-varying baseline tank temperature from EnergyPlus. Because making the bounds on $\Delta S_k(t)$ time-varying does not fundamentally change the mathematics that follow, we continue to use the Simple model in this section.

Lemma 2. ***Inner approximations of aggregated electric water heaters.***

Choose a β_k for each $k \in \mathbb{K}$ such that $0 \leq w_k \beta_k \leq 1$ and $\sum_k w_k \beta_k = 1$. Then with

$$\Delta P_k(t) = \beta_k U(t) \tag{A.60}$$

we can show that $\mathbb{U}_{\underline{\Phi}_{HW}}(t) \subset \mathbb{U}(t)$, where $\mathbb{U}(t)$ is defined by (A.9) and

$$\underline{\Phi}_{HW} = \left(\eta, \alpha, 0, -\overline{L}(t), \overline{L}(t), \max_k \frac{-\widetilde{P}_k(t)}{\beta_k}, \min_k \frac{\overline{P_k} - \widetilde{P}_k(t)}{\beta_k} \right), \tag{A.61}$$

where η and α are arbitrarily chosen and

$$\overline{L}(t) \leq \frac{\alpha_k}{\alpha_k + |\alpha - \alpha_k|} \frac{\eta}{\eta_k} \min \left(\frac{c_k \left| \underline{T}_k - \widetilde{T}_k \right|}{\beta_k}, \frac{c_k \left| \overline{T}_k - \widetilde{T}_k \right|}{\beta_k} \right) \quad \forall k \in \mathbb{K}. \tag{A.62}$$

Proof of Lemma 2. Lemma 2 is a simplification of Theorem 2. □

In the following lemma, we show how the β_k for each water heater can be be chosen optimally to satisfy a desired balance between energy and power capacity. The version shown here only optimizes the ability to reduce load and amount of stored thermal energy. Although the lemma and proof are easily extended to also incorporate an objective function term for ability to increase load, in practice we have found this to be unimportant. Ability to increase load is not in play at all for providing a load shedding contingency service. For shifting load, we found the ability to increase load to be relatively abundant as compared to reducing load. For example, our aggregate ERWH shifting resources performed slightly better if we grouped resources based only on $\underline{\Delta S}(t)$ and $\underline{\Delta P}(t)$ profiles; including $\overline{\Delta S}(t)$ and $\overline{\Delta P}(t)$ profiles as well yielded results that were mostly unchanged but slightly less effective in terms of amount of load shifted and price-taking profits.

Lemma 3. ***Optimal inner approximations of aggregated electric water heaters.*** *Building on the setting in Lemma 2, the β_k that maximize*

$$w\overline{L} + (1-w)\,|\underline{U}| \tag{A.63}$$

with $\overline{L}$ defined in (A.62), and $|\underline{U}| = -\underline{U}$ defined as in (A.61), are equal to

$$\beta_k = \frac{\phi_k}{\sum_k w_k \phi_k}, \quad \phi_k = w \frac{\eta}{\eta_k} \frac{\alpha_k c_k}{\alpha_k + |\alpha - \alpha_k|} \min \left(\widetilde{T}_k - \underline{T}_k, \overline{T}_k - \widetilde{T}_k \right) + (1-w)\widetilde{P}_k(t). \tag{A.64}$$

Proof of Lemma 3. We seek to solve

$$\max_{\beta_k, \overline{L}, \underline{U}} w\overline{L} + (1-w)\eta\,|\underline{U}| \tag{A.65}$$

$$\text{s.t. } \beta_k \overline{L} \leq \frac{\eta}{\eta_k} \frac{\alpha_k c_k}{\alpha_k + |\alpha - \alpha_k|} \min \left(\widetilde{T}_k - \underline{T}_k, \overline{T}_k - \widetilde{T}_k \right) \forall k \in \mathbb{K} \tag{A.66}$$

$$\beta_k |\underline{U}| \leq \widetilde{P}_k(t) \qquad \forall k \in \mathbb{K} \tag{A.67}$$

$$0 \leq w_k \beta_k \leq 1 \qquad \forall k \in \mathbb{K} \tag{A.68}$$

$$\sum_k w_k \beta_k = 1 \tag{A.69}$$

Using (A.66) - (A.67) we can write for each $k \in \mathbb{K}$:

$$\beta_k \left(w\overline{L} + (1-w)\eta|\underline{U_c}| \right) \leq \eta \left(w \frac{c_k}{\eta_k} \frac{\alpha_k}{\alpha_k + |\alpha - \alpha_k|} \min\left(\widetilde{T}_k - \underline{T}_k, \overline{T}_k - \widetilde{T}_k \right) + (1-w)\widetilde{P}_k(t) \right). \tag{A.70}$$

Then defining

$$\psi = \frac{1}{w\overline{L} + (1-w)\eta|\underline{U}|} \tag{A.71}$$

and

$$\phi_k = w \frac{c_k}{\eta_k} \frac{\alpha_k}{\alpha_k + |\alpha - \alpha_k|} \min\left(\widetilde{T}_k - \underline{T}_k, \overline{T}_k - \widetilde{T}_k \right) + (1-w)\widetilde{P}_k(t) \tag{A.72}$$

we have

$$\beta_k \leq \eta \phi_k \psi \tag{A.73}$$

$$1 = \sum_k w_k \beta_k \leq \psi \eta \sum_k w_k \phi_k. \tag{A.74}$$

Noting that minimizing ψ is equivalent to maximizing our objective (A.65), we see that (A.74) implies

$$\psi \geq \frac{1}{\eta \sum_k w_k \phi_k}, \tag{A.75}$$

such that the optimal ψ^* is

$$\psi^* = \frac{1}{\eta \sum_k w_k \phi_k}. \tag{A.76}$$

Changing (A.73) to an equality and thereby defining

$$\beta_k^* = \eta \phi_k \psi^* = \frac{\phi_k}{\sum_k w_k \phi_k} \quad \forall k \in \mathbb{K} \tag{A.77}$$

is then seen to yield the optimal values for β_k since one can verify that $\sum_k w_k \beta_k^* = 1$ and substituting the β_k^*'s into (A.61) yields $\overline{L}$, $|\underline{U_c}|$, and $\overline{U_c}$ values that satisfy (A.66) - (A.67). □

Between Lemmas 2 and 3 we see some difficulties arising because of the time-varying nature of water heater use. If we construct inner approximation aggregate models using constant β_k values, then there may be times when we cannot increase or decrease load based on $\widetilde{P}_k(t) = 0$ or $\widetilde{P}_k(t) = \overline{P_k}$ for some $k \in \mathbb{K}$. One way to avoid such issues is to estimate baseline power draw $\widetilde{P}_k$ over a coarse time frame $t \in [\underline{t}, \overline{t}]$, treat the aggregation problem within that window as non-time-varying, and then to ensure that $L(\overline{t}) = 0$. It might also be reasonable to estimate time-varying aggregate model parameters based on the time-varying $\beta_k(t)$ suggested by Lemma 3. However, additional conditions would almost certainly be needed to ensure that such an aggregate model was truly an inner approximation of the aggregate flexibility in the sense of an allowable aggregate dispatch not resulting in out-of-bounds operation of any of the individual devices.

In this study we take the former approach to ensure feasible dispatch. For contingency service, this means that we use different β_k for each hour and do not track what happens in the rebound period after the event. For shifting service we estimate constant β_k for an entire day and ensure that $L(t)$ is brought back to zero at the switchover time (midnight of each day). In both cases, we

apply tolerances to screen out any resources who cannot provide a sufficiently large individual response (and would therefore set overly restrictive bounds for the aggregate as a whole). For the shifting service we also cluster water heaters with similar profiles prior to aggregating since over an entire day it is not possible to escape, e.g., times when at least some water heaters are simply not on and thus are unable to reduce load.