Case-Studies in the Economics of Ancillary Services of Power Systems in Support of High Wind Penetrations

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Engineering & Public Policy

Todd Ryan

B.S.E., Chemical Engineering, Tufts University

Carnegie Mellon University

Pittsburgh, PA

May 2016

Acknowledgements

I would like to acknowledge the countless hours of effort and infinite patience of Paulina Jaramillo and Gabriela Hug. I would like to thank Paul Hines, Mario Berges, M. Granger Morgan, and the entire Engineering & Public Policy Team for their contributions to this thesis.

I wish to express my deepest thanks to David Luke Oates, Amy Wesolowski, Justin Glier, Steve Rose, Roger Lueken, and the many other friends who contributed to this thesis through their countless caffeinated conversations.

.Thank you to Jamie Gracer, Judy Shanks, and Sue Ryan for their love and support and Joshua Bernard Ryan, my son. Most of all, I would like to thank my wife, Talia Gracer, for her love, patience, support, and partnership.

To the memory of Bernard Francis Ryan, to whom I owe everything.

This work was supported by the Doris Duke Charitable Foundation, the RK Mellon Foundation, the Heinz Endowments, and the Electric Power Research Institute through their support to the RenewElec project at Carnegie Mellon University. The results and conclusion of this paper are the sole responsibility of the authors and do not represent the views of the funding source.

ABSTRACT

This thesis analyzes two potential means of mitigating the cost increase of ancillary services that is expected with the decarbonization of the U.S. electricity network. The first method, balancing area consolidation, addresses this cost rise by reducing the demand for ancillary services. This research quantifies the economic benefit of consolidation in the frequency regulation market by estimating the resulting reductions in frequency regulation requirements and cost. The results show that this policy leads to a reduction in frequency regulation cost of approximately $0.1 per MWh of total load. These results do not significantly change with the inclusion of 20% wind, suggesting that in the near term, wind's interaction in the frequency regulation market is not a prime motivation for consolidation. This analysis does not consider all the benefits or costs of BA consolidation, and is not meant as an assessment of net-benefits. Though the results show consolidation could lead to an increase in emissions of some air pollutants, which suggest that there may be significant trade-offs associated with the decision to consolidate balancing areas.

The second means of addressing the expected increase in ancillary services costs is to increase the supply of ancillary services by leveraging residential demand response. We developed methods that optimally schedule ancillary service capacity on demand response resources while accounting for the risk of customer response fatigue. The model is used to test the efficacy of hourly caps on demand response penetration in ancillary service markets. The results show that residential demand response could provide a significant portion of the total ancillary service requirements attributable to residential loads: between 50% and 75%. Hourly caps on demand response participation are shown to be economically inefficient. With a 25% market cap, residential demand response is scheduled to

provide 25% of the hourly total market value, while the risked-based optimization schedules residential demand response to provide 82% of the total value. Methods like the ones presented in this paper, that can appropriately weight the benefits and risks of committing residential demand response will be critical to efficiently and effectively use this resource for ancillary services.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1: INTRODUCTION

America is at a major crossroads: does it believe in anthropogenic climate change and does it want to reduce green-house gas emissions to climate stabilizing levels. If the answer to both of these questions is *yes*, then a critical step will be decarbonizing the electricity network. Today there are many state and federal policies that encourage the procurement of renewable energy such as renewable portfolio standards [1], regional greenhouse gas markets [2], and the Environmental Protection Agency's proposed clean power plan [3]. These policies, in conjunction with other environmental regulations, *e.g.,* the Mercury and Air Toxin Standards [4], will likely result in large amounts of coal-fueled generator to be retired as well as a significant increase in energy efficiency, demand-side management and large amounts of renewable generation. Most of this additional renewable energy is likely to come from wind [5].

This dramatic shift in both fossil-fueled and renewable generation will undoubtedly affect the cost of energy as well as the cost of ancillary services, a set of additional services that allow the commodity of energy to be reliably delivered. With this additional wind in the electricity network comes additional variability that is inherent to wind power [6]. The added variability will increase the need for balancing using flexible resources providing ancillary services. At the same time, thermal generation resources, traditionally supply these ancillary services are being retired due to environmental regulations. This decreases the supply of ancillary services. The net economic impact of these is still highly uncertain but will most likely result in higher prices and costs for the balancing of wind (Figure 1.1).

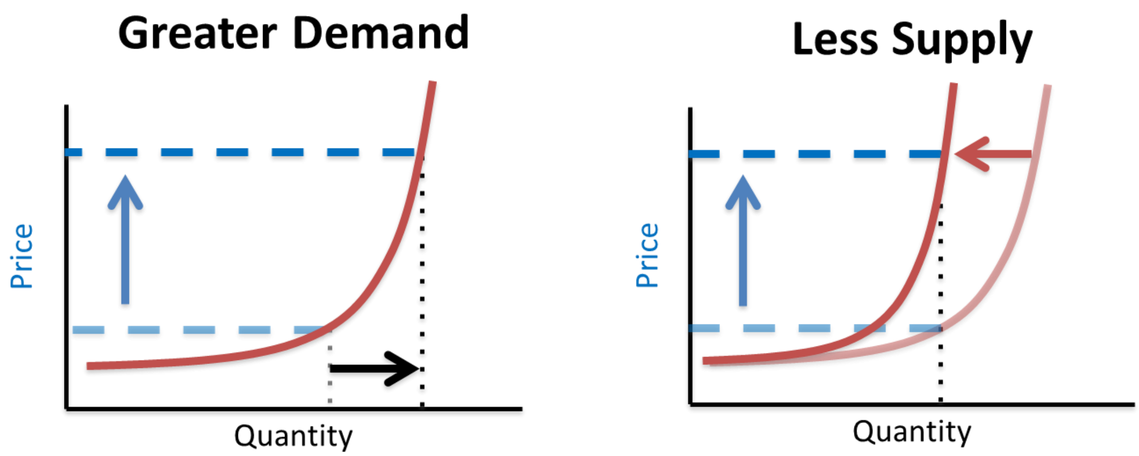

Figure 1.1. Illustration of the economic effects of increase demand and decreased supply. Both changes to the market result in higher prices, and therefore higher cost.

This thesis research will contribute to this field by evaluating two potential opportunities for mitigating the cost increase of ancillary services that is expected given high wind penetrations.

The first technology aims to reduce the demand for ancillary services by expanding the geographic footprint over which utilities coordinate ancillary services. This is known as balancing area consolidation and chapter two investigates its economic effects. Specifically, this chapter quantifies the effect of geographic diversity on aggregate wind power and on the requirements for frequency regulation, a key input to the co-optimized economic dispatch model. This model simulates prices for energy and frequency regulation, and calculates the cost of frequency regulation before and after consolidation. This chapter uses data from the Midcontinent Independent System Operator (MISO) but is not meant to be an assessment of the total or net benefits of the MISO consolidation. Rather we use MISO as the basis to create a hypothetical consolidation scenario using a coherent and realistic set of data that includes load profiles, generation fleets, and wind data. The results show that this policy leads to a reduction in frequency regulation cost of approximately \$0.1 per MWh of total load. These results do not significantly change with the inclusion of 20% wind, suggesting that in the near term, wind's interaction in the frequency regulation market is not a prime motivation for consolidation. Furthermore, the results show consolidation could lead to an increase in emissions of some air pollutants, which suggest that there may be significant trade-offs associated with the decision to consolidate balancing areas. This analysis does not consider all the benefits or costs of BA consolidation, and is not meant as an assessment of net-benefits.

Chapter three analyzes the potential for residential demand response as a new source of ancillary services. The analysis assumes that all residential houses have the ability to receive ancillary service dispatch instructions (*e.g.*, broadband or cable) and the necessary intra-household communication (*e.g.*, Wi-Fi) to transmit these instructions to individual flexible loads. We developed methods that optimally schedule ancillary service capacity on demand response resources while accounting for the risk of customer response fatigue. The model is used to test the efficacy of hourly caps on demand response penetration in ancillary service markets. The results show that residential demand response could provide a significant portion of the total ancillary service requirements attributable to residential loads: between 50% and 75%. Hourly caps on demand response participation are shown to be economically inefficient. With a 25% market cap, residential demand response is scheduled to provide 25% of the hourly total market value, while the risked-based optimization schedules residential demand response to provide 82% of the total value. Methods like the ones presented in this paper, that can appropriately weight the benefits and risks of committing residential demand response will be critical to efficiently and effectively use this resource for ancillary services.

This work aims to inform the policy discussion regarding the expected rise in ancillary service cost with high renewable generations. It is not meant to be an exhaustive text on the subject, but rather case-studies that demonstrate the subtlety and detail required to analyze the policy effects on the electric transmission system.

1.1 References

1. Commission CPU (2015) Renewables Portfolio Standard Quarterly Report - 1st Quarter 2015. CPUC, http://www.cpuc.ca.gov/NR/rdonlyres/8EB096A9-8000-4A1D-9623-DB3EB9F1BFB5/0/2015Q1RPSReport.pdf

2. RGGI (2015) RGGI Fact Sheet. http://www.rggi.org/docs/Documents/RGGI_Fact_Sheet.pdf. Accessed 6/9/2015 2015

3. EPA (2015) Clean Power Plan. Carbon Pollution Emission Guidelines for Existing Stationary Sources: Electric Utility Generating Units. Federal Registry Vol 79, No. 117

4. EPA (2014) Mercury and Air Toxics Standards. National Emission Standards for Hazardous Air Pollutants: Coal- and Oil-Fired Electric Steam Generating Units, vol 40 CFR Part 63. Federal Registry Vol 79, No. 223, page 68795

5. EIA (2015) Analysis of the Impacts of the Clean Power Plan. http://www.eia.gov/analysis/requests/powerplants/cleanplan/pdf/powerplant.pdf

6. Apt J (2007) The spectrum of power from wind turbines. Journal of Power Sources 169 (2):369-374

Chapter 2: Near-term Economic Effects of Wide-Scale Balancing Area Consolidation on Frequency Regulation Markets

This paper is based on a working paper. "Near-term Economic Effects of Wide-Scale Balancing Area Consolidation on Frequency Regulation Market." Department of Engineering and Public Policy, Carnegie Mellon University.

2.1 Introduction

Wind generation in the U.S. has grown rapidly in recent years [1] and this increase in wind power will need to continue in order for the U.S. to meet its renewable energy goals. Wind power's inherent variability occurs on all time scales [2] and can significantly affect the electricity grid's stability and reliability, as discussed in many integration studies, *e.g.*, the Eastern Wind Integration and Transmission Study [3]. When analyzing the effect of variability on the grid, it is important to consider that the system is geographically divided into balancing areas (BAs) where a system operator has to maintain the balance between demand and supply of electricity. There are over 100 BAs in the U.S. varying in size between 70 MW and 153 GW [4,5]. Balancing area consolidation has many potential benefits; for example BA consolidation can lead to improved coordination in the unit commitment process and transmission utilization [6]. Additionally, a set of BAs can lower their cost of energy by optimizing over their newly shared set of resources [7,8]. BA consolidation can also make it easier to follow load by smoothing the relative variability of net-load, lowering the needed amount of ramping capacity, ramping rate, and cycle frequency [9-12]. Previous work that has focused on this effect has either demonstrated that the integration of geographically diverse wind resources leads to reductions in net-load's relative variability [2,13,14], or has addressed the economic effects of sharing resources in the energy market [e.g., 8]. There is a gap, however, in understanding the combined effects of reducing variability and sharing resources. This work addresses this gap by simultaneously modeling and quantifying the reduction in net-load's relative variability; the reduction in frequency regulation requirements due to reduced relative variability; and the economic benefits of the sharing of frequency regulation resources and reduced frequency regulation requirements.

Although there are multiple ways in which BA consolidation leads to social benefits, this research focuses on the benefits in the frequency regulation[1] market and compares them to the benefits obtained in the energy market. The model also estimates how consolidation affects the emissions of air pollutants from power plants. The analysis focuses on a case-study using the balancing areas eventually consolidated into MISO Energy as a means of creating a consistent and realistic consolidation scenario. This research is not an attempt to replicate or estimate all the benefits and costs of the MISO consolidation; rather this research uses the MISO consolidation as a source for consistent data that can be used to examine the *ceteris parabis* economic effects of consolidation on frequency regulation markets.

2.2 Data and Methods

BA consolidation can reduce the cost of frequency regulation in two ways: i) it creates a more geographically diverse set of renewable resources, reducing the variability of the aggregate renewable output, and therefore reduces the quantity of frequency regulation needed; ii) it results in a new larger portfolio of generating assets, ensuring that the cheapest set of resources is used to provide frequency regulation.

This chapter quantifies the effect of geographic diversity on aggregate wind power and on the requirements for frequency regulation, a key input to the co-optimized economic dispatch model. This model simulates prices for energy and frequency regulation, and calculates the cost of frequency regulation before and after consolidation.

---

[1] Frequency regulation is an ancillary service that balances the electricity grid over short-times scales by modifying the output of flexible resources.

2.2.1 Midcontinent ISO Case-Study.

This chapter uses data from the Midcontinent Independent System Operator (MISO) but is not meant to be an assessment of the total or net benefits of the MISO consolidation. Rather we use MISO as the basis to create a hypothetical consolidation scenario using a coherent and realistic set of data that includes load profiles, generation fleets, and wind data.

MISO was founded in 1998 [15]; became a regional transmission operator in 2001; launched its energy market and started centralized dispatch of resources in 2005 [16]; and opened its ancillary services market and became a balancing authority in 2009 [17], when it consolidated 26 balancing areas at one time [15]. MISO has grown and shrunk and it currently encompasses what used to be 31 different balancing areas before 2006. Based on data availability, this chapter uses a subset of these historic balancing areas, grouped into sixteen pre-consolidation BAs. A list of these balancing areas is available in the Supporting Information.

In order to maintain the high correlation between wind and load data, it is critical that these data match temporally and geographically. In addition, the characteristics of the conventional generation fleet are needed. For most BAs in what is now MISO, load data are only available from 2006-2008 while simulated wind data are only available for 2004-2006. For this reason, the model uses load data from 2006 [18]. Power plant data are from the eGRID database, which provides information on every generating unit in the U.S. including unit size, heat rate, fuel type, emissions data, and to which power control area each unit belongs. We matched power plants in eGrid to the modeled balancing areas using the NERC CPS Bounds reports [4].

Time-series data of wind power are necessary in order to model a power system at any wind penetration. In order to model a system with higher wind penetration than exists today, time-series data from nonexistent, hypothetical wind farms are needed. The Eastern Wind Dataset (EWD) provides three years' worth of simulated ten-minute average wind power output for 1,300 hypothetical wind sites in the Eastern Interconnect [3]. We assigned EWD sites to historic balancing areas in MISO based on capacity factor and geographic location. Using Ventyx Velocity Suites [19], we mapped the footprints of the balancing areas and the location of the EWD sites. We further selected sites that fall within the footprint of a BA in order of descending capacity factor until the BA passes 20% wind by energy, based on the expected production of each wind farm. Some BAs do not have enough wind farms within their geographic boundary to reach 20% penetration in this first step. For these BAs, we added unassigned wind farms (not assigned in the first step) until each BA reaches above 20% wind penetration. This method results in each BA in our scenario having approximately 20% wind by energy; Figure 2.1 displays all the BAs and wind farms used in this chapter. While we considered additional means of allocating wind farms to BAs, they all considered produce similar results as detailed in the Supporting Information.

The EWD provides ten-minute average wind production data, which we averaged to hourly values for use in the economic dispatch model, such that the wind and load data have the same sample rate. This paper models two different wind penetration levels (0%, and 20% by energy). The model can choose to curtail wind when it is economic to do so, or if it is required in order to meet physical constraints. The 20% wind case is representative of near-term renewable portfolio standards that exist in the United States.

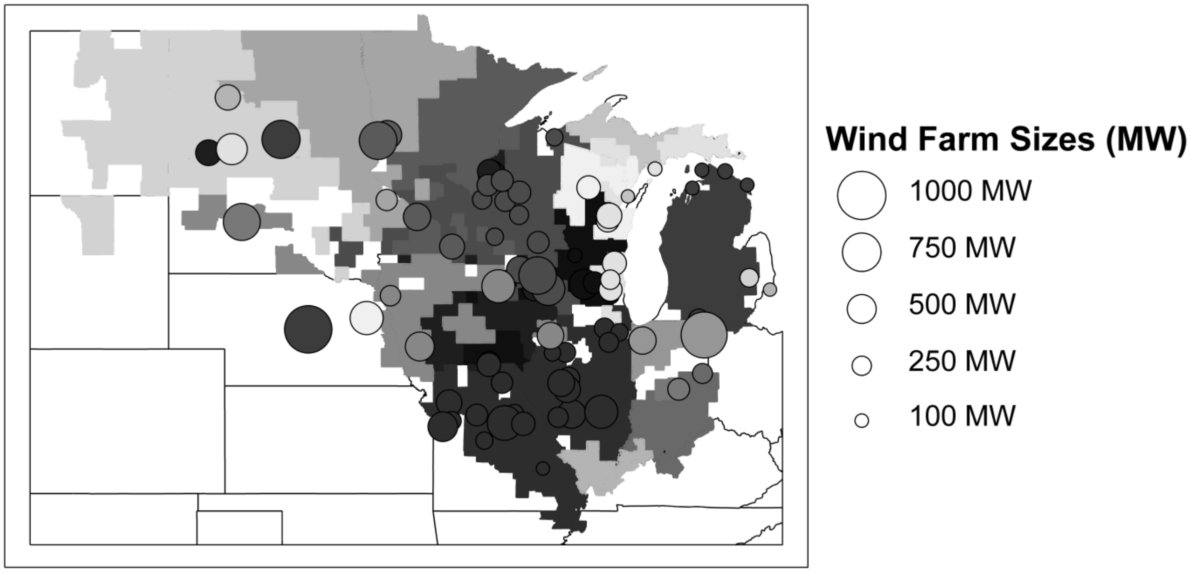

Figure 2.1. The sixteen pre-consolidation regions and their assigned wind farms (circles sized by MW capacity) in matching color.

2.2.2 Quantifying the Effects of Consolidation on Variability.

The grid needs to balance the variability of load and renewables and this variability plays a major role in setting reliability standards and the procurement requirements of ancillary services, such as frequency regulation. Adding more renewables, and therefore variability, to the grid may increase ancillary service requirements and the cost of providing reliable electricity to consumers [3,20]. Balancing area consolidation can reduce the need for frequency regulation, when comparing the sum of all the pre-consolidated requirements against the post-consolidated requirement. Previous research suggests that consolidation reduces frequency regulation requirement by reducing peak load and by reducing the variability of renewable resources through the smoothing effect of geographic diversity. The reduction in peak load is a result of non-coincidental peak loads across the different BA, *i.e.,* the peak load in one BA may not occur at the same time as the peak of a different BA. Once consolidated, the resulting peak load of the CBA will be lower than the sum of all the individual BAs' peak load. This reduction of peak load is easily quantifiable and is presented in the results section.

Geographic diversity has been shown to smooth the aggregate power of wind [21,22,13,3,20,11,14]. The variance of the distribution of aggregated wind power (W^T) and the variance of the distribution of step-changes in aggregate wind power (ΔW^T) provide the basis for evaluating the effects of geographic diversity on aggregate wind power variability. Katzenstein et al. [13] found that consolidating a few close wind farms reduces the absolute variability of aggregate wind power output. However, consolidating wind resources over larger and more distant regions does not reduce the absolute variability [14]. Such large scale consolidations only reduce the relative variability of the wind power as represented by the coefficient of variation (c_v), defined as the sample standard deviation (σ) divided by the sample mean (μ) [23]. Therefore, the metrics of variability presented in this paper are in relative terms (coefficient of variation).

2.2.3 Ancillary Service Requirements.

Regardless of the size of the effect geographic diversity has on the aggregate wind power, the change in frequency regulation requirements is the key driver of cost reductions in the market for frequency regulation (assuming fixed supply). In this paper we use two different methods to estimate the frequency regulation requirement of a balancing area ($FR\ Req$), both of which are based on the Western Wind and Solar Integration Study (WWSIS) [3].

The first heuristic suggests that the hourly operating reserves requirement is 3% of a BA's forecasted daily peak load ($\hat{L}$) plus 5% of the hourly forecasted aggregate BA wind output (W^T) [24]. WWSIS also suggests that approximately one third of this operating reserve should be in the form of frequency regulation, resulting in a forecast-based heuristic described in Equation 2.1. This method is consistent with a long-standing industry estimate of the frequency regulation requirement for BA's with little to no wind being approximately equal to 1% of daily peak load, *e.g.,* [25]. This forecast-based heuristic captures the effect of peak load reduction but does not capture the reduction in variability due to geographic diversity of wind resources.

$$FR\ Req_{\,t,j} = \left(0.01 * \hat{L}_{j,d\ni t}\right) + \left(\frac{0.05}{3} * W_{j,t}^{T}\right) \qquad 2.1$$

The second heuristic (also derived from the WWSIS study) can capture the change in the variance in wind that results from consolidating geographically diverse resources. This second heuristic sets the operating reserve requirement equal to three standard deviations (σ) of the 10-minute step-changes (Δ) in net-load ($3\sigma\Delta$). A step change (Δ) of any time series is the difference between two successive data points in a time-series. Net-load is the BA's load minus the BA's aggregate wind power. Given that approximately one third of this operating reserve should be in the form of frequency regulation, the heuristic becomes equal to one-sigma-delta ($1\sigma\Delta$).

This heuristic requires ten-minute net-load data, which requires both ten-minute wind and load data. However, only hourly load data is available for MISO during the modeled time-period. Based on a NREL report [26], we approximate the variance of ten-minute load based on hourly load; we then directly calculate the variance ten-minute wind data and sum these two variances using the mathematical properties of variance. A detailed description of this mathematical process is available in the Supporting Information. Equation 2.2 gives the final formulation used in this paper for this frequency regulation heuristic, which is a function of the daily peak load ($\hat{L}$) and the distribution of step-changes in aggregate BA wind power (ΔW^{T}). For the remaining of this paper, we refer to this method as the variance-based heuristic.

$$FR\ Req_{\,t,j} = \left(\left(0.01 * \hat{L}_{j,d\ni t}\right)^{2} + Var\left[\Delta W_{j,t}^{T}\right]\right)^{\frac{1}{2}} \qquad 2.2$$

This paper explores the effects of using these two separate frequency regulation demand heuristics on the benefits (or costs) of balancing area consolidation by including the demand for these services in the economic dispatch model. Comparing the results from simulations using these different heuristics, we can isolate the effect of geographic diversity on FR requirements and costs.

2.2.4 Economic Dispatch Model.

This analysis uses an economic dispatch model that minimizes the total cost of providing energy and frequency regulation for a specific BA, subject to the system constraints. These constraints include meeting system demand for electricity and frequency regulation, and adhering to generator limits for energy, regulation, and hourly ramping. We first perform this optimization for each of the consolidating BAs individually and then again for all the regions as one consolidated balancing area (CBA). For each hour, the model produces dispatch instructions for energy and frequency regulation for each generator, and curtailment instructions for each wind farm. The complete mathematical formulation of this optimization problem is available in the Supporting Information.

To solve for the least-cost dispatch schedule, the economic dispatch model includes the short run marginal costs of generation for individual power plants, as well as the cost for each plant to provide ancillary services. For fossil-based power plants, the short run marginal cost depends on the fuel cost and the power plant efficiency. As previously described, power plant efficiency is available from the eGRID database. Table 2.1 shows the fuel costs used in this model, as well as the short run marginal costs of non-fossil fueled power plants. Some power plants listed in the eGrid database use more than one fuel. For these plants, the model uses a weighted average of the short-run marginal cost for the primary and secondary fuels. The weighting is based on the unit's historic fuel consumption.

Table 2.1. Assumed fuel prices and short-run marginal cost of energy by fuel type (USD 2009). Fuel prices are average values. Short-run marginal costs are taken from Newcomer et al (2008).

Fuel Type	Fuel Price ($/mmBTU)	Short-Run Marginal Cost ($/MWh)
Biomass		$50.00
Bituminous Coal	$2.72	
Distillate Fuel Oils	$13.14	
Hydro		$10.00
Lignite Coal	$1.59	
Natural Gas	$4.90	
Nuclear	Considered must-run, i.e., $0 / MWh	
Petroleum Coke	$1.50	
Subbituminous Coal	$1.64	
Wind	Considered must-run, i.e., $0 / MWh	

The eGrid database does not provide information that can be used to directly estimate the marginal cost of providing frequency regulation, and MISO data were unavailable for 2006. Given this lack of information, we used historic bids for frequency regulation from the New York Independent System Operator (NYISO) to develop a simple heuristic to predict a unit's bid for frequency regulation. The NYISO provide the basis for this heuristic because this system is the most alike to MISO: both have bi-directional frequency regulation markets, with similar compensation mechanisms that have single clearing prices and include the marginal unit's opportunity cost. Using a year's worth of bids from the NYISO, we developed a deterministic model based on generator size to estimate the average bid quantity (MW_{FR}) and the bid price ($\$/MW_{FR}$-Hr). More details about the regression model are available in the Supporting Information. As a result of the bid analysis, generators with a capacity between 200 and 300 MW bid 14% of their capacity in the frequency regulation market; all other units bid 6% of their capacity for frequency regulation (Equation 2.3). Table 2.2 shows the regulation bid prices used in our optimization.

$$Q_{FR-Bid,i} = \begin{cases} 0.06 * \overline{\overline{\mathrm{MW}}}_{\mathrm{i}} & \text{for} \quad \overline{\overline{\mathrm{MW}}}_{\mathrm{i}} < 200, \overline{\overline{\mathrm{MW}}}_{\mathrm{i}} > 300 \\ 0.14 * \overline{\overline{\mathrm{MW}}}_{\mathrm{i}} & \text{for} \quad 200 < \overline{\overline{\mathrm{MW}}}_{\mathrm{i}} < 300 \end{cases} \qquad 2.3$$

Table 2.2. Frequency Regulation bid prices (USD 2009) by generator size based on regressions of historic frequency regulation bids.

Generator Size (MW)	Bid Price ($/MW-Hr)
< 100	$21
100-200	$7.4
200-300	$23
300-400	$35
400-500	$102
500-600	$60
600-700	$8.0
700-800	$200
800-900	-
900-1,000	$200
> 1,000	$200

2.2.5 Emissions Modeling

The economic dispatch produces data on how much electricity each generating unit produces at each time interval. These results determine the generator's emissions of carbon dioxide (CO_2), nitrogen oxides (NO_X), sulfur dioxide (SO_2), and particulate matter ($\leq$ 2.5 μm, $PM_{2.5}$). We used unit specific emission factors for each pollutant from AP 42 (for $PM_{2.5}$) [27] and the eGrid database (for all other pollutants) [28].

2.2.6 Sensitivity and Scenarios

A model of this nature has a large number of details and parameters that will affect the results of this analysis: generation fleet and expected energy bids; wind penetration; frequency regulation bid quantity and price; frequency regulation requirements; and fuel prices (natural gas, coal) are just a few of the key variables that drive the results of this model. This paper tests for the sensitivity of the results to these parameters and present the scenarios that show the highest sensitivity. Specifically, this paper describes scenarios that include two different levels of wind penetration, two methods of estimating the frequency regulation requirement, three natural gas prices, and two different levels of frequency regulation bid pricing. Table 2.3 summarized these scenarios. Note that the 'High Gas & Frequency Regulation' cases are meant to push the boundaries of the economic effect of the frequency regulation market by simultaneously raising gas prices to $10/mmBTU and multiplying all frequency regulation bid prices shown in Table 2.2 by a factor of 1.5.

Table 2.3. Summary of modeled scenarios that vary wind penetration, frequency regulation requirement heuristic, and natural gas price.

Scenario Name	Wind	FR Heuristic	Gas Price & Frequency Regulation Pricing
LG0	**0%**	**Forecast-Based** [A]	**Low Gas Price** [C]
B0	**0%**	**Forecast-Based** [A]	**Base Gas Price** [D]
HG0	**0%**	**Forecast-Based** [A]	**High Gas Price** [E]
HGR0	**0%**	**Forecast-Based** [A]	**High Gas Price & High FR Prices** [F]
LG1	**20%**	**Forecast-Based** [A]	**Low** [C]
B1	**20%**	**Forecast-Based** [A]	**Base** [D]
HG1	**20%**	**Forecast-Based** [A]	**High Gas** [E]
HGR1	**20%**	**Forecast-Based** [A]	**High Gas Price & High FR Prices** [F]
LG2	**20%**	**Variance-Based** [B]	**Low** [C]
B2	**20%**	**Variance-Based** [B]	**Base** [D]
HG2	**20%**	**Variance-Based** [B]	**High Gas** [E]
HGR2	**20%**	**Variance-Based** [B]	**High Gas Price & High FR Prices** [F]

A – **Forecast-Based** refers to the forecast-based frequency regulation heuristic and frequency regulation bids as described earlier
B – **Variance-Based** refers to the variance-based frequency regulation heuristic ($1\sigma\Delta$) and frequency regulation bids as described earlier
C – **Low Gas** refers to a gas price of $4/MMBTU (USD2009) and frequency regulation bids as described earlier
D – **Base Gas** refers to a gas price of $4.90/MMBTU (USD2009) and frequency regulation bids as described earlier
E – **High Gas Price** refers to a gas price of $10/MMBTU (USD2009) and frequency regulation bids as described earlier
F – **High Gas Price & High FR Prices** refers to a gas price of $10/MMBTU and an assumed 1.5 multiplier on all frequency regulation bids.

2.2.7 Model Limitations & Biases

There are a few limitations to this model that could induce different biases. A first bias derives from the strict assumption built into the disaggregated counterfactual that prior to consolidation, balancing areas meet their energy needs only through resources that are located physically in their area. This is not true for today's system as BAs are easily capable of importing and exporting power to neighbors. Given the available data, it is impossible to accurately estimate the counter-factual import and export for any modeled BA. Therefore, we had to assume that there is no import or export in the counterfactual case. This assumption likely induces a positive bias on the benefits to the energy market. That is to say, any actual import/export of energy between a pair of BAs today is, by definition, economically advantageous by both parties and reduces the cost of operating the system and move the pre-consolidated cost closer to that of the fully consolidated. However, this bias would be smaller for the frequency regulation where it would likely over-estimate the reduction in frequency regulation cost through the opportunity cost component of price. This potential bias likely does not change the order-of-magnitude of the benefits to the frequency regulation market.

In the chapter, transmission constraints or unit commitment have been excluded, which certainly affects the final numerical results. The lack of transmission constraints, however, does not affect the pre-consolidation costs because of the aforementioned strict counterfactual assumption. No inter-regional transmission is required if all the BAs are meeting their own needs with internal resources. The inclusion of transmission would therefore only limit the extent to which BAs could cooperate when consolidated. This would likely only reduce the potential benefits when consolidated. Therefore, the omission of transmission is a positive bias for benefits and likely small for the benefits of frequency regulation.

The exclusion of unit commitment constraints does create a negative bias. As previous research has shown, there are benefits to the unit commitment process from consolidation [6]. The bias in the benefits in the frequency regulation market, the focus of this work, is likely to be small given the amount of wind generation. Hence, the inclusion of unit commitment is beyond the scope of this paper.

2.3 Results & Discussion

This research quantifies the economic benefit of consolidation in the frequency regulation market by estimating the resulting reductions in frequency regulation requirements and cost. This analysis does not consider all the benefits or costs of BA consolidation, and is not meant as an assessment of net-benefits. Additionally, the many modeling assumptions, limitations, and trade-offs (see Appendix A.1.3 Economic Dispatch - Additional Details) mean that these results and conclusions may not be applicable to all power systems and scenarios.

2.3.1 Quantifying the effects of Consolidation on Variability

Figure 2.2 shows the coefficients of variation of aggregated wind power for each balancing area and the consolidated balancing area (CBA), using the methods outlined in Fertig et al. [23]. The results suggest that there is a reduction in the relative variance in wind power after consolidation and the results are consistent with those of Fertig et al. [14]. This figure shows diminishing marginal returns in the reduction of variance: the effect of geographic diversity is less significant in large-scale consolidation (over a gigawatt) than it is for the first gigawatt.

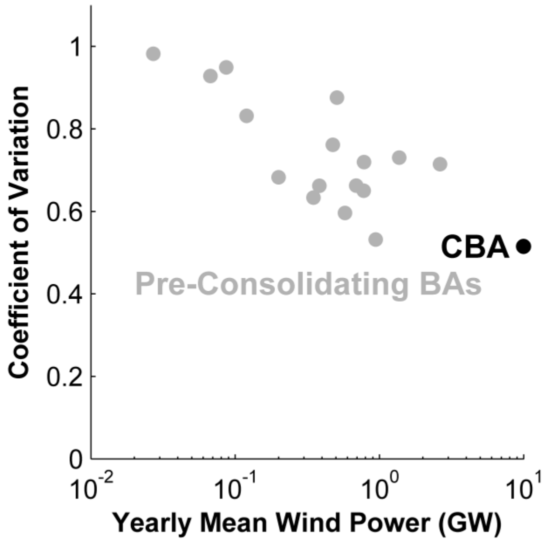

Figure 2.2. The coefficient of variation of wind power versus mean wind power. The relative variability, as represented by the coefficient of variation, decreases with increased wind power.

2.3.2 Quantifying the Effects of Consolidation on Ancillary Service Requirements

The frequency regulation requirements for each BA, for each hour, are based on the two heuristics described in section 3.3. The hourly average frequency regulation requirement for the CBA is compared to the sum of all the pre-consolidated BAs' requirements in Table 2.4.

Table 2.4. The average hourly frequency regulation requirement Summary of modeled scenarios that vary wind penetration, frequency regulation requirement heuristic, and natural gas price. for CBA (post-consolidation) compared to the sum of the requirements for the pre-consolidated BAs. The presented data are for scenarios based on the two wind penetrations (no wind and ~20% wind by energy) and the two frequency regulation heuristics (forecast-based and variance- based) assuming the base-case natural gas fuel price ($4.10/mmBTU).

	No Wind Forecast-Based (MW_{FR})	Wind 1 Forecast-Based (MW_{FR})	Wind 2 Variance-Based (MW_{FR})
Sum of BAs (Pre-Consolidation)	477 MW_{FR}	643 MW_{FR}	686 MW_{FR}
CBA (Post-Consolidation)	472 MW_{FR}	638 MW_{FR}	517 MW_{FR}
Change (MW / %)	-5 MW_{FR} (-1%)	-5 MW_{FR} (-0.7%)	-169 MW_{FR} (-25%)

The forecast-based heuristic predicts a small reduction in the frequency regulation requirements after consolidation: 1% without wind and 0.7% with wind (see the first two columns of Table 2.4).

This small change in frequency regulation requirements is due to the reduction of daily peak load. The variance-based heuristic predicts a 24% drop in frequency regulation requirement after consolidation. This suggests that the reduction of the relative variance of aggregate wind power, due to consolidation, does significantly reduce the amount of frequency regulation after consolidation.

Figure 2.3 further highlights this reduction in frequency regulation requirements. The forecast-based heuristic (1% peak + 1.67% wind) predicts a frequency regulation requirement that is fairly linear across the different balancing area sizes. This is because this heuristic is a linear function of forecasted wind and peak daily load, both of which scale with balancing area size. The variance-based heuristic does account for the change in relative variance associated with geographic diversity, visually demonstrated by the reduction in mean and variance in the relative hourly frequency regulation requirement (hourly frequency regulation requirement normalized by the daily peak). It can be seen, however, that there are diminishing returns in the reduction of frequency regulation as the size of the consolidated area increases. On the far right of the figure, the sum of the pre-consolidated BAs is compared to that of the CBA. The CBA on average has a frequency regulation requirement equal to 1.1% of its daily peak load; the sum of the pre-consolidated BAs FR requirements is equal to 1.46% of the CBA's daily peak load.

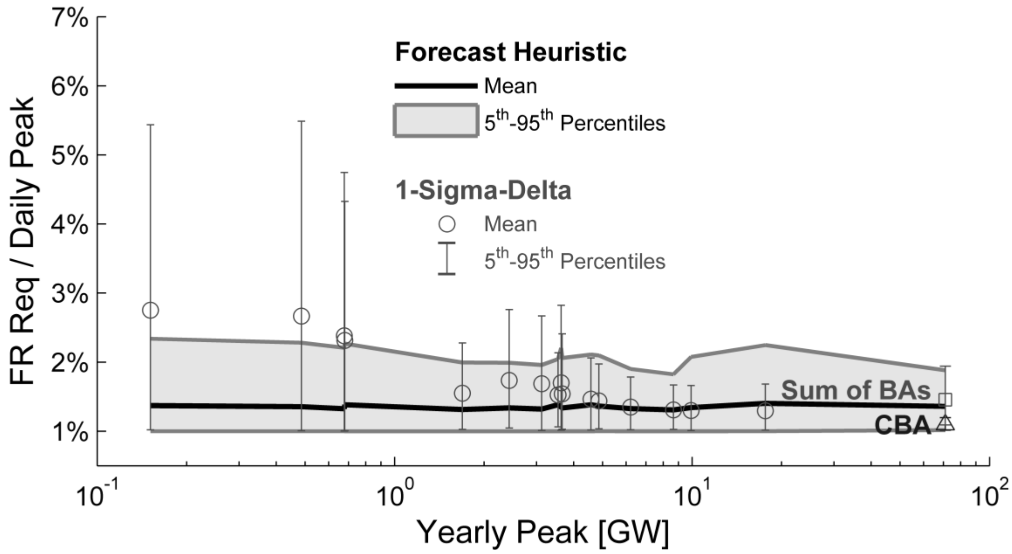

Figure 2.3. Balancing area frequency regulation requirements relative to the BA's size, for the previously described modeled BAs with 20% wind, based on two different heuristics: forecast-based and variance-based. The distribution shown for each BA represents the variation in hourly requirements throughout one year.

2.3.3 Quantifying the Economic and Environmental Effects of Consolidation

Figure 2.4 shows the distributions of economic gains in the frequency regulation market. The figure shows that the economic benefit of consolidation in the frequency regulation market, as approximated by the change in social surplus, is between \$0 and \$0.4 per MWh of total load. These results even hold in the extreme case of high natural gas prices and high frequency regulation bids (cases labeled HGR0, HGR1, and HGR2).

The economic benefit in the frequency regulation market is due to two effects: sharing of resources and the smoothing effect of geographic diversity. We estimate the contribution of each of these factors by comparing the results of the two different frequency regulation heuristics. The forecast-based heuristic (top row middle panel) does not account for the smoothing effect of geographic diversity. The economic benefit of geographic diversity in the frequency regulation market can thus be calculated by subtracting the benefits in the forecast-based heuristic (between \$0.04 and \$0.08 per MWh of total load) from the benefits in the variance-based heuristic ($1\sigma\Delta$, \$0.12

and \$0.18 per MWh of total load - top row right panel). Accordingly, the benefits of geographic diversity in the frequency regulation market are \$0.04-\$0.1 per MWh of load given an approximate wind penetration of 20%, by energy.

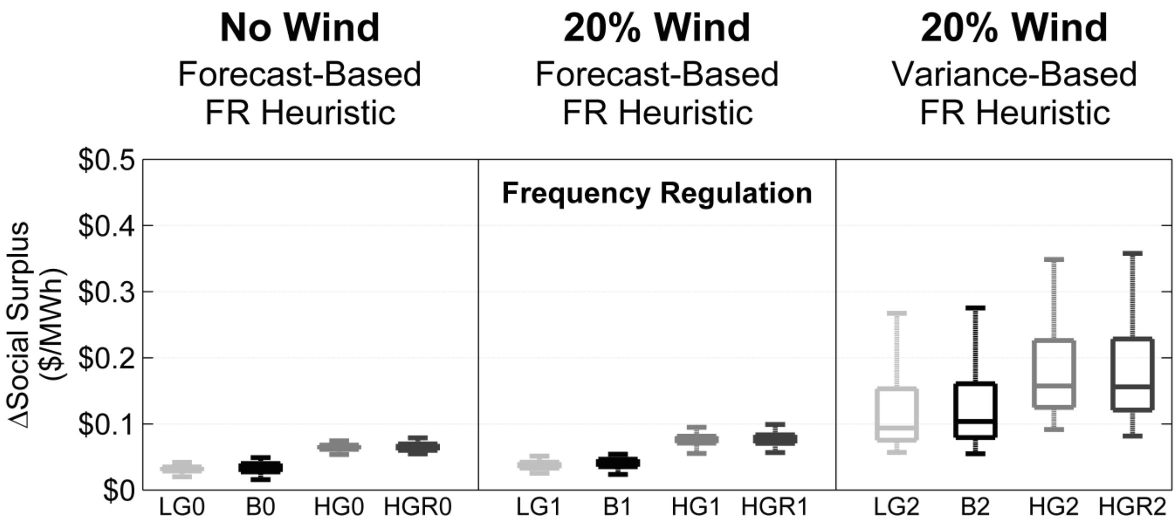

Figure 2.4. The change in social surplus (USD 2009) for the frequency regulation market is represented by the box plots above for the modeled scenarios. The variation shown in each boxplot represents how the economic benefits change in time; each datum being a weekly average value of the change in social surplus.

The economic benefits in the energy market (Figure 2.5) give further context to the size of the economic benefit of consolidation in the frequency regulation market. Without wind, the economic benefit of BA consolidation in the energy market is approximately \$1.5 per MWh of total load; with 20% wind, this average economic benefit is less than \$1 per MWh of total load. This lower benefit is due to the additional supply of zero marginal cost energy from wind energy, which suppresses energy prices. Note that the total economic benefit in the frequency regulation market is an order of magnitude lower than the benefit in the energy market.

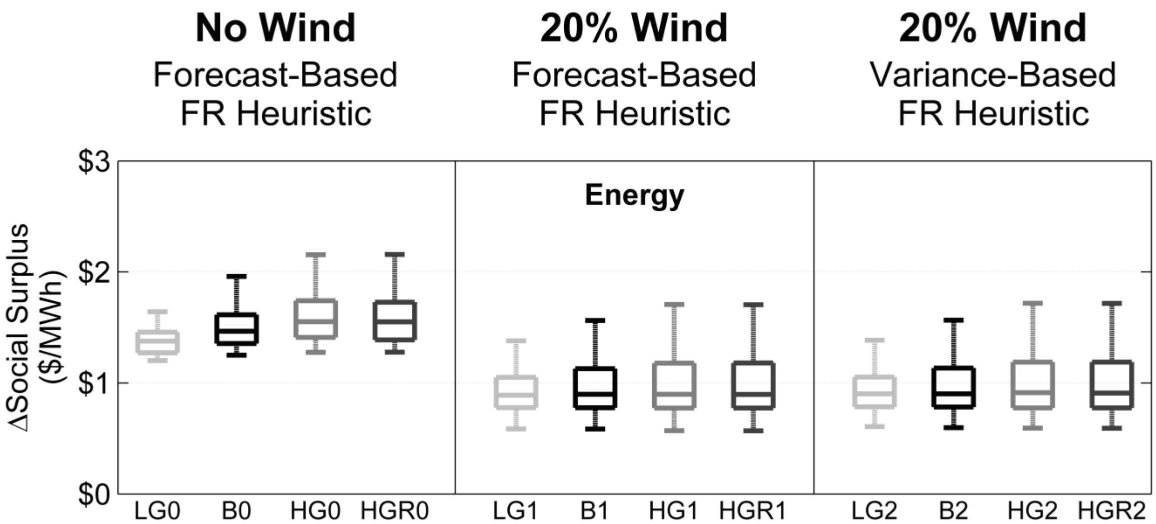

Figure 2.5. The change in social surplus (USD 2009) for the energy market is represented by the box plots above for the modeled scenarios. The variation shown in each boxplot represents how the economic benefits change in time; each datum being a weekly average value of the change in social surplus.

As part of this work, we estimate the emissions from each generation unit before, and after, consolidation. Figure 2.6 presents the percentage change in emissions (post-consolidation minus pre-consolidation) for four pollutants: carbon dioxide (CO_2), nitrogen oxides (NO_X), sulfur dioxide (SO_2), and particulate matter ($\leq$ 2.5 μm, $PM_{2.5}$). Carbon dioxide and particulate matter emissions increase; nitrogen oxides dramatically decrease; and sulfur dioxide moderately decreases. This change in emission rates is consistent with fuel switching from natural gas to coal. In our model, the low-cost coal generation is limited in the pre-consolidation state by the constraint on energy imports and exports; higher-priced natural gas units provide the missing energy. After consolidation, the low-cost coal units can produce more through exporting energy to other consolidated balancing areas, at the same time reducing natural gas generation. While this result is specific to a coal-heavy power plant fleet, it none-the-less demonstrates that there may be unintended consequences to balancing area consolidation.

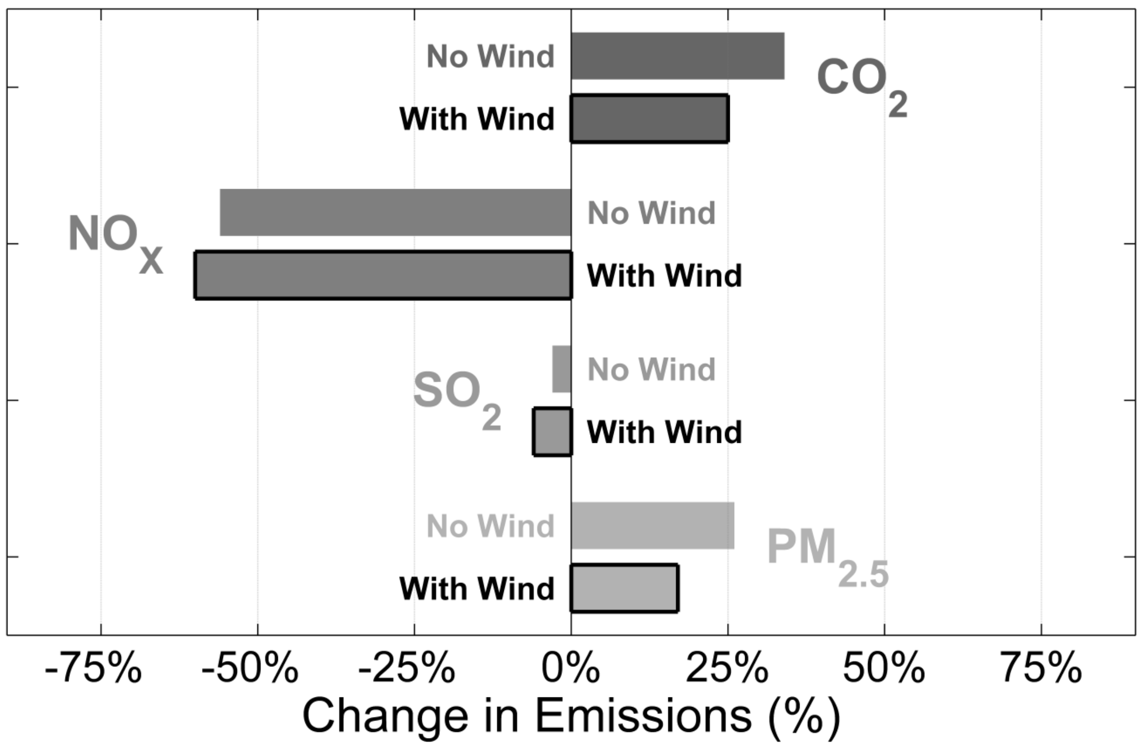

Figure 2.6. The percentage change in emissions for four pollutants: carbon dioxide (CO_2), nitrogen oxides (NO_X), sulfur dioxide (SO_2), and particulate matter (≤ 2.5 µm, $PM_{2.5}$) (post-consolidation minus pre-consolidation).

2.4 Conclusions and Policy Implications

This research focuses on quantifying the potential near-term economic benefit of wide-scale consolidation in the frequency regulation market; it is not intended to make a case for, or against, the policy of adding more wind energy to our electricity system. The results suggest that large wind penetration increases social surplus from the frequency regulation and energy markets (~$4 per MWh) and reduces air emissions, when compared to the non-wind cases. The rest of the discussion focuses on the comparisons between the consolidated and pre-consolidated cases – not between scenarios with and without wind. Additionally, this research does not include an exhaustive search for all of the benefits and costs of BA consolidation.

BA Consolidation reduces the relative variability of net load, which reduces the aggregate requirement for frequency regulation by approximately 25%. The data suggest that this effect exhibits diminishing marginal returns, meaning consolidating the first few BAs produces a larger reduction than the final few BAs. The reduced frequency regulation requirement, combined with the effect of shared resources, leads to a reduction in frequency regulation cost of approximately $0.1 per MWh of total load, or $30 million per year. These results do not significantly change with the inclusion of 20% wind, suggesting that in the near term, wind's interaction in the frequency regulation market is not a prime motivation for consolidation.

The benefit of this wide-scale consolidation on the energy market is approximately $1/MWh of total load. It is not surprising that the benefit to the energy market is larger than in the regulation market as the energy market is far bigger and more valuable than all of the ancillary service markets combined. It should be noted again that this analysis does not consider all the benefits or costs of BA consolidation, and is not meant as an assessment of net-benefits.

The data show that CO_2 and $PM_{2.5}$ emissions increase, while NO_X emissions dramatically decrease, and SO_2 slightly decreases. Given the relative health and climate risks each of these pollutants, it is likely that this change in emissions results in a net cost to society. This suggests that while there may be economic benefits to BA consolidation, BA consolidation may not be as beneficial today as previous literature suggest depending on the generation portfolio of the considered system and the level of penetration of renewable generation. Hence, a blanket policy to support balancing area consolidation in fossil-based systems may result in unexpected consequences such as increased emission. Those who may be considering consolidation should be careful that all benefit and all costs are considered and that the net-benefits are in line with their specific goals before proceeding.

2.5 References

1. EIA (2011) Electric Power Annual 2011: Summary Statistics for the U.S.

2. Apt J (2007) The spectrum of power from wind turbines. Journal of Power Sources 169 (2):369-374

3. Corbus D, King J, Mousseau T, Zavadil R, Heath B, Hecker L, Lawhorn J, Osborn D, Smit J, Hunt R (2010) Eastern wind integration and transmission study. NREL (http://wwwnrelgov/docs/fy11osti/47078pdf) CP-550-46505

4. NERC (2009) 2009 CPS2 Bounds Report. http://wwwnerccom/docs/oc/rs/CPS2%20Bounds%202009(Revised%2010-1-2009)pdf

5. NERC (2012) 2012 CPS2 Bounds Report. http://wwwnerccom/docs/oc/rs/2012%20CPS2%20Bounds%20Report%20Final(Update20120419)pdf

6. Van der Weijde AH, Hobbs BF (2011) Locational-based coupling of electricity markets: benefits from coordinating unit commitment and balancing markets. Journal of Regulatory Economics 39 (3):223-251

7. Rose J, McCarthy C, Collison K, Parmar H (2007) Independent Assessment of Midwest ISO Operational Benefits. ICF International (Ed), Fairfax, VA

8. Corcoran BA, Jenkins N, Jacobson MZ (2012) Effects of aggregating electric load in the United States. Energy Policy 46 (0):399-416. doi:http://dx.doi.org/10.1016/j.enpol.2012.03.079

9. Oneal A (1995) A simple method for improving control area performance: area control error (ACE) diversity interchange ADI. Power Systems, IEEE Transactions on 10 (2):1071-1076

10. Milligan M, Kirby B (2007) Impact of Balancing areas Size, Obligation Sharing, and Ramping Capability on Wind Integration: Preprint. 43 pp. NREL Report No. CP-500-41809. Presented at WindPower 2007, Los Angeles, CA,

11. Makarov YV, Etingov PV, Zhou N, Ma J, Samaan NA, Diao R, Malhara SV, Guttromson RT, Du P, Sastry C (2010) Analysis Methodology for Balancing Authority Cooperation in High Penetration of Variable Generation.

12. Milligan MR, Kirby B, Beuning S (2010) Combining Balancing Areas' Variability: Impacts on Wind Integration in the Western Interconnection. National Renewable Energy Laboratory,

13. Katzenstein W, Fertig E, Apt J (2010) The variability of interconnected wind plants. Energy Policy 38 (8):4400-4410

14. Fertig E, Apt J, Jaramillo P, Katzenstein W (2012) The effect of long-distance interconnection on wind power variability. Environmental Research Letters 7 (3):034017

15. FERC (1998) Order conditionally authorizing establishment of Midwest Independent Transmission System Operator and establishing hearing procedures. Federal Energy Regulatory Commission

16. FERC (2005) Order on certification and conditional approving revised tariff sheets. Federal Energy Regulatory Commission

17. NERC (2008) NERC Balancing Authority (BA) Certification of the Midwest Independent System Operator (Midwest ISO). North American Electric Reliability Council

18. FERC (2014) Form No. 714 - Annual Electric Balancing Authority Area and Planning Area Report. https://www.ferc.gov/docs-filing/forms/form-714/data.asp.

19. Ventyx (2011) Ventyx Velocity Suit. http://www.ventyx.com/en/solutions/business-operations/business-products/velocity-suite.

20. GE-Energy (2010) Western wind and solar integration study. National Renewable Energy Laboratory (NREL), Golden, CO.,

21. Zavadil R, King J, Samaan N, Lamoree J (2006) Final Report-2006 Minnesota Wind Integration Study: Volume I. The Minesota Public Utilities Comission, Saint Paul, MN, Final Rep

22. Efficiency E (2008) 20% Wind Energy by 2030: Increasing Wind Energy's Contribution to US Electricity Supply. DOE/GO-102008-2567, US Department of Energy, May. http://www. nrel. gov/docs/fy08osti/41869. pdf,

23. Fertig E, Apt J, Jaramillo P, Katzenstein W (2012) Supplemental Information: The effect of long-distance interconnection on wind power variability. Environmental Research Letters 7 (3):034017

24. GE-Energy (2010) Section 7.4 of WWSIS: Concepts for Simple Reserve Rules. National Renewable Energy Laboratory (NREL), Golden, CO.,

25. PJM (2011) PJM Manual 11: Energy and Ancillary Services Market Operations: Section 3.2.4.: Determination of Regulation Requirement.
http://pjmcom/~/media/documents/manuals/m11ashx

26. GE-Energy (2010) Section 7.3.5 and Figure 7.68 of WWSIS: 1% of peak daily load is a good proxy for the variance of ten-minute load. National Renewable Energy Laboratory (NREL), Golden, CO.,

27. EPA (2014) Emissions Factors & AP 42, Compilation of Air Pollutant Emissions Factors. 2014

28. EPA (2012) eGRID2012 Version 1.0. http://www.epa.gov/cleanenergy/energy-resources/egrid/index.html

Chapter 3: Residential Load Providing Ancillary Services: Estimates of potential given the risk of response fatigue

This paper is based on a working paper. "Residential load providing ancillary services: estimates of potential given the risk of response fatigue." Department of Engineering and Public Policy, Carnegie Mellon University.

3.1 Introduction

Most electricity consumers have very little conscious knowledge of the operations of the power system and their interactions with this complex system. Few people, for example, think about the quantity or price of the electricity they use before they turn on any household appliance. Nor do they think about the many reliability services that exist to make sure that the electricity they demand is always available. Demand response (DR) corresponds to a set of technologies and policies that allow customers to more actively participate in the electricity markets by 1) allowing them to be price sensitive, *i.e.,* increasing or decreasing their demand based on the price of electricity; and 2) allowing consumers to receive payments to modify their consumption automatically in a way that provides different types of ancillary services to the grid. This increased customer participation directly affects many aspects of the power system including generation, emissions, economics, and reliability.

Electric loads (or end-uses) come in many shapes and forms, and all of them have some inherent amount of flexibility. At a minimum, an end-use is flexible by switching between the 'on' and 'off' states. Other loads have greater flexibility and are capable of varying their power consumption with time. The idea of using load's inherent flexibility through direct load control traces back as far as 1969 [1] but only since 1999 [2] has the idea of using electric loads for ancillary services been considered in theory or practice.

Many demand response studies focus on demonstrating that load can provide ancillary services. Ancillary services are a set of balancing services provided by flexible resources in order to help maintain the grid's reliability. These services act in a hierarchical fashion to correct for grid imbalances over different time-scales and their technical requirements vary accordingly. There are up to twelve ancillary service types [3] but their specific names and requirements vary regionally [e.g., 4,5]. Table 3.1 shows a summary of six general ancillary services, and their respective technical requirements [6,7]. Other ancillary services such as black-start are excluded from Table 3.1 because DR cannot provide these services.

Table 3.1. Ancillary service characteristics: Response speed, duration, cycle time, quantity and price. The average quantity and price are for ERCOT data from 5/1/2013 to 4/1/2014

Generic Name (ERCOT Name)	Response[A] Speed	Duration[A]	Cycle[A] Time	Average[B] Quantity (MW/Hr)	Average[B] Price ($/MW)
Frequency Response[C] (Governor Response)	Seconds	Minutes	Minutes	2,000[C]	Not Applicable[D]
Frequency Regulation (Up Regulation & Down Regulation)	~1 min	Minutes	Minutes	464 (Up) 415 (Down)	$12 (Up) $8 (Down)
Spinning Reserves (Responsive Reserves)	Seconds to ~10 minutes	10 to 120 minutes	Hours to days	2,800	$13
Non-Spinning Reserves (Non-Spin)	< 30 minutes	2 hours	Hours to days	1,440	$4
Peak Reserves[E] (None)	~1 hour	6-8 hours	Daily	Not Applicable[E]	

A – Ancillary service characteristics are based off of Kirby et al [8]

B – Quantity and price data are based on ERCOT data from 5/1/2013 to 4/1/2014 - http://www.ercot.com/mktinfo

C – Frequency Response quantity is based on ERCOT's current frequency response obligation 286 MW/0.1 Hz and an under-frequency load shed threshold of 59.3 Hz [9]. (60Hz-59.3Hz)*2860MW/Hz = 2,002 MW.

D – Frequency Response is an ancillary service that is not currently priced markets. This is likely to change given FERC's approval of NERC's BAL-003 standard [10], the corresponding changes ERCOT has proposed to its ancillary services market [9], and FERC's recent notice of proposed rulemaking [11].

E – Peak reserves are not typically considered an ancillary service. In some markets however (e.g., PJM), peak reserves has its own unique forward auction market that is very different from the day ahead and real-time markets for energy and ancillary services. ERCOT does not have a peak reserves market, but ERCOT has and is developing programs that use direct load control to reduce load in peak hours [12,13].

Early demand response studies that focus on ancillary services show how large commercial or industrial loads can provide ancillary services [14,2,15,16]. These resources are large enough to provide such grid support without needing to aggregate multiple loads. More recent studies tend to focus on the theoretical control of thermostatically controlled loads (TCLs) providing a specific type of ancillary service [17-20]. TCLs (*e.g.,* air conditioning, heating, water heaters, *etc.*) operate with the goal of keeping a specific temperature within an acceptable range. These loads are particularly flexible and capable of providing ancillary services due to their inherent thermal mass that can be used, similar to a battery, to shift consumption in time. Much of the literature on TCLs covers the theoretical provision of all the main ancillary services [21-25], but few studies include demonstration projects that use actual hardware [17,24,26,27]. Additionally, the economic effect of DR providing ancillary services is often outside of the scope of these studies.

Papers that do show the net economic benefits of DR [28-32] focus on price sensitivity, estimating how consumers weigh the trade-off between the benefits gained from consuming electricity and its cost. For the most part, these studies ignore the ability of consumers to provide ancillary services and support grid operations. An exception is a recent set of studies [33-38] that established precise methods for assessing the economic implications of providing ancillary services through DR. Therefore, some of the methods presented in this paper are based on Olsen *et al* [35], which give a means to estimate ancillary service availability from demand response resources. We build on the methods presented in Olsen *et al* [35] which give a means to estimate ancillary service availability from demand but ignore the fact that response fatigue reduces the effective availability. Here, response fatigue refers to the fact that calling on DR resources to modify their consumption leads to the customer being dissatisfied, uncomfortable, and unwilling to respond to future calls for response [39,40]. The most common method for addressing consumer comfort and fatigue is to place constraints on state variable in the controls of the load aggregation that is providing demand response [21-25]. Most, if not all, of the studies used to define the flexibility parameters (section 3.5) addressed consumer comfort in this way [see supporting information section 8.1]. These additional constraints could lead to situations where the load aggregation won't be able to meet the actual dispatch of ancillary services associated with the amount of ancillary service capacity scheduled on that load, a major concern for grid operators. Some markets address this risk by placing caps on the amount of ancillary service capacity that can be assigned to DR resources as a method of mitigating the risk of response fatigue. For example, PJM limits demand response resources' contribution in the frequency regulation market to 25% of the overall requirement [41], while MISO limits interruptible loads to 10% of the spinning reserve requirement and 25% of the supplemental reserve requirement [42]. This type of policy is conservative in that it errs on the side of reliability; but it is also coarse in that it treats all hours of the year as equally risky and deserving of the same cap level. It is likely that there are hours in which this type of cap is justified, but it is unlikely that a fixed cap is appropriate in all hours.

This paper presents a novel new way of addressing comfort and consumer fatigue by trying to eliminate situations where scheduling ancillary service capacity is not worth the risk of being short on ancillary services due to consumer comfort and fatigue. The methods statistically model the *ex post* dispatch of ancillary services and assess the risk of consumer fatigue as estimated by each load aggregation's deferred energy: the difference between its counterfactual consumption and its modified consumption while providing ancillary services. Deferred energy is difficult to incorporate directly into a dispatch model because it is a function of the unpredictable reliability needs of the grid (*ex post*), while the reservation of ancillary service capacity is economically allocated ahead of time (*ex ante*). For this purpose, we develop two statistical models based on ERCOT data [43]to create potential realizations of deferred energy, due to ancillary service deployment, which we use to approximate the risk of DR's response fatigue. Given this estimate of risk, we can optimally schedule ancillary service capacity on aggregations of loads. This paper also compares the optimally scheduled DR resources against traditional means of scheduling with a fixed hourly cap on DR penetration in ancillary services markets.

Unlike some studies that attempt to estimate how much ancillary services potential from demand response could exist today [35,44,45,32,46], this paper focuses on a near future where most houses have efficient and controllable appliances that can easily communicate with grid operators, possibly through broadband and Wi-Fi systems. In order to model this future, this research uses data from the Pecan Street project, a Smart Grid demonstration project consisting of over 150 homes near the city of Austin, Texas. These homes have efficient modern appliances and their residents have agreed to high frequency monitoring of energy consumption at the appliance level. The Pecan Street sample is not meant to be representative of the average American household, but provides a vision of a plausible future for DR.

3.2 Methods

This paper quantifies the future potential and availability of ancillary services from residential loads. The analysis assumes that all houses have the ability to receive ancillary service dispatch instructions (*e.g.,* broadband or cable) and the necessary intra-household communication (*e.g.,* Wi-Fi) to transmit these instructions to individual flexible loads. Under these assumptions there are two key criteria that determine the amount of ancillary services residential load can supply. First, the load must be plugged in and consuming (or able to consume) electricity so that it can modify its consumption and provide DR. Second, the load has to be able to modify its hourly consumption from its counterfactual hourly consumption: what it would have consumed if it were not providing ancillary services. To identify loads that meet these requirements, it is critical to have accurate load data for each individual residential load type, which is why this chapter uses empirical data from the

Pecan Street project instead of relying on modeled load data, such as that from the National Energy Modeling System (NEMS) [47]. Section 3.1 describes the Pecan Street data set further.

Figure 3.1 is a schematic overview of this paper's model and the method used to optimally schedule ancillary services on residential demand response resources. The first step is to determine how much of each ancillary service could be made available by the residential load over time. Each of the resulting time-series of ancillary services availability is mutually exclusive; *i.e.,* if the entire air conditioning load is scheduled to provide frequency regulation, it cannot also be scheduled for spinning reserve. Therefore, the available ancillary service needs to be optimally allocated to ensure that the DR is being put to its best possible use. Finally and most importantly, this chapter presents a novel method of assessing the risk of scheduling DR resources beyond their ability to be flexible. This risk is based on the statistical likelihood that a load defers energy due to ancillary services deployment. The result is the scheduling of ancillary services capacity from residential DR resources that considers the inter-temporal risk of ancillary services deployment.

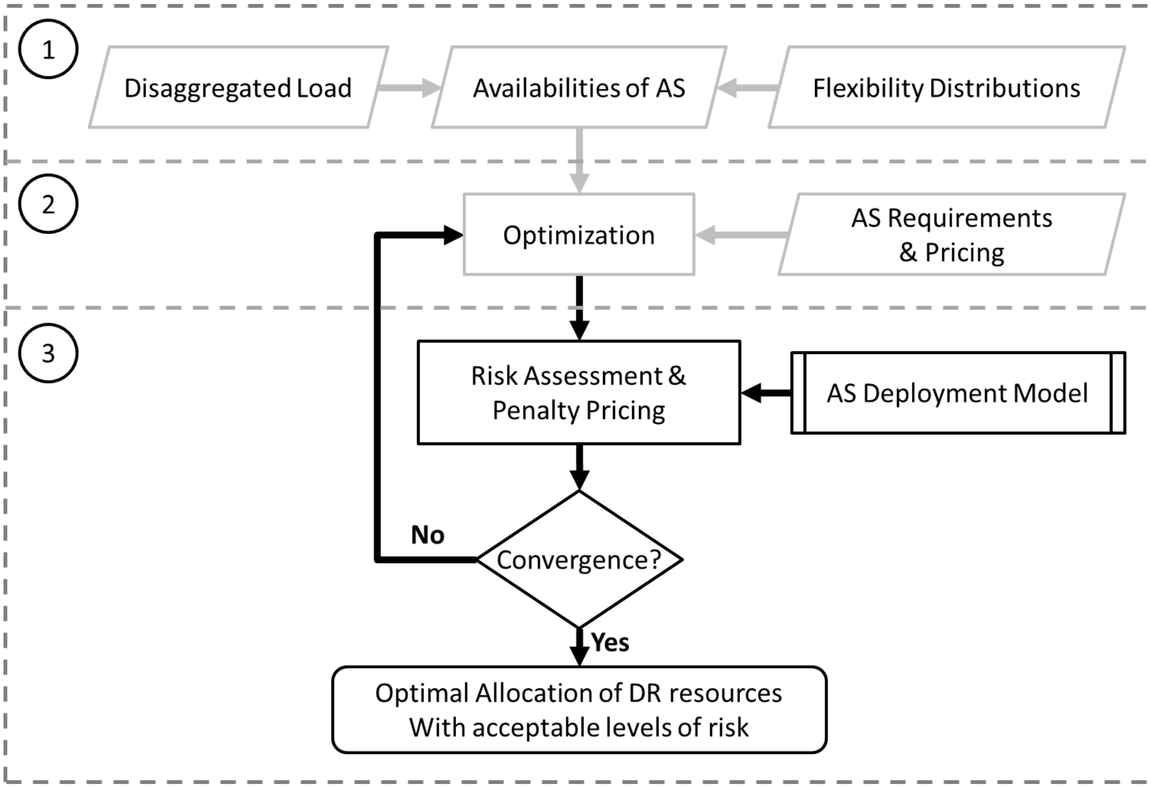

Figure 3.1. Skematic for the optimal allocation of DR providing AS. The boxes shown in gray are based on the methods in Olsen *et al.* The portions in black represent the paper's contribution to the methods.

3.2.1 Residential Demand Data

The model in this paper requires the use of data that represent the behavior of loads over the course of the day, week, month, and year. Unlike other studies that rely on top-down models for load profiles, this paper uses empirical data from the Pecan Street project, which includes over 150 homes near the city of Austin, Texas [48]. This chapter limits the sample to a subset of Pecan Street homes that contain valid refrigerator and air-conditioning data for one continuous year. For example, some excluded homes had air conditioning data with large gaps, null data, or data inconsistent with the dynamics of refrigeration. Thirty seven of the homes do not have reliable refrigerator data; these homes were included in our sample but we replaced the invalid refrigerator data using an exponential model [21,22,49,50]. Table 3.2 summarizes the final data set from the 92 homes (55 with empirical data; 37 with modeled data).

Table 3.2. Summary of the penetration of different loads for the considered 92 home sample from the Pecan Street Project

End-Use	Household Penetration	
	Count (#)	Percentage (%)[A]
Air Conditioning	90	99%
Refrigerators	92 (55)[B]	100% (60%)[B]
Electric Space Heating	80	88%
Electric Water Heating	8	9%
Lighting	92	100%
2nd Fridge or Freezer	1	1%
Clothes Washer	57	63%
Electric Clothes Dryers	81	89%
Dishwasher	62	68%
Pumps or Irrigation	6	7%
Electric Vehicle	37	41%
Solar PV	53	58%
Jacuzzi	5	5%

A – Count divided by 92, expressed as a percentage.

B – Fifty five (55) houses have usable refrigerator data for the entire sample period. For the balance of the sample (37 homes) the refrigerator data is simulated.

The ten end-use categories listed in Table 3.3 group all of the metered loads in each house. Aggregate values result from the sum of all loads in each category and over all houses. The result is a set of ten time-series of the sample's total residential consumption, by category, with a one-minute resolution. Figure 3.2 shows the daily average of these data. Grid operators set the commitment of ancillary services on an hourly basis through their *ex ante* security-constrained unit commitment and economic dispatch models. This frequency difference (between the one minute data and the hourly schedules) means that we have a few options when choosing the maximum basis for setting the hourly schedule. For example, given a load aggregation that is expected to vary its aggregate load between 100 kW and 110 kW during an hour, one could schedule this load to provide 100, 105, or

110 kW using minimum, average, and maximum as the basis, respectively. We took a conservative approach and used the minimum value in each hour, of the one-minute load data for an aggregation of a single end-use, as the maximum basis for scheduling ancillary services capacity.

Table 3.3. End-Use Categories and their respective loads

Index	End-Use Categories	Loads in Category
1	Space Heating	Space heating
2	Cooling	Air conditioning
3	Water Heating	Water heaters, Jacuzzis
4	Refrigeration	Refrigerators, refrigerator-freezer combinations, beer and wine specific refrigerators
5	Freezers	Freezers, ice makers
6	Delayable Appliances	Clothes washers, clothes dryers, dish washers, pumps, pools (assumed to be pumps), aquariums (assumed to be pumps), sprinklers and irrigations
7	Lighting	Lighting
8	Cooking	Microwaves, electric stoves, anything labeled "kitchen"
9	N/A	All others – including electric vehicles.

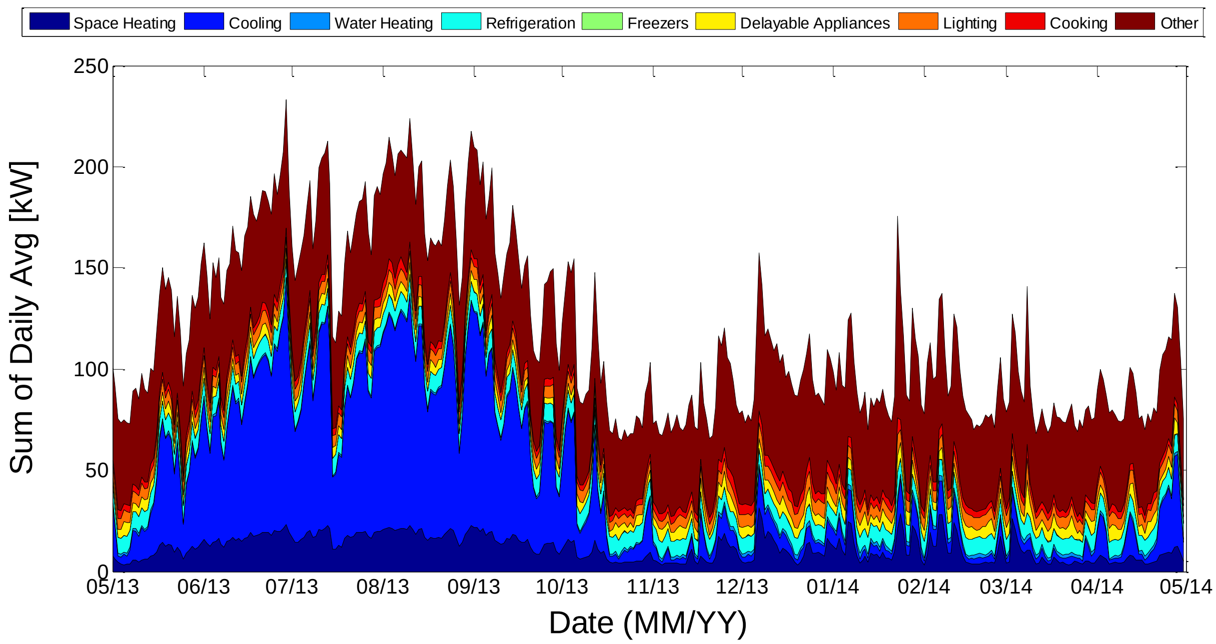

Figure 3.2. Contribution to total load of the various end-uses as represented by the daily average consumption. Note that the total value is dominated by air conditioning and the unknown other end-uses. The other category includes specific room loads (*i.e.*, bathroom, living room, *etc.*).

It should be noted that this chapter does not consider electric vehicles, as there is an entirely separate body of literature around the optimal charging of electric vehicles, *e.g.*, [51] and is outside of the scope of this project. Demand from electric vehicles in the Pecan Street Data is included in the "other" category, which we assumed is unable to provide ancillary services.

3.2.2 Ancillary Service Requirements

Ancillary service requirements set the market demand for each service in each hour. To obtain such data for the simulations, this paper uses time-series data of wholesale ancillary service requirements from ERCOT that are temporally coherent with the Pecan Street data. The Pecan Street sample is only a small fraction of the size of ERCOT, hence, it is necessary to estimate a residential ancillary services requirement that is proportional to the Pecan Street sample. The hourly requirement for frequency response, spinning reserves, and non-spinning reserves are all linked to the yearly peak of system load $\left(\hat{\hat{L}}\right)$, while hourly frequency regulation (up and down) requirements are more tied to daily peak load $(\hat{L})$ [6,9,52]. Therefore, ancillary services requirement attributable to the Pecan Street load is based on the appropriate ratio of sample peak load to the ERCOT peak load. We therefore determine the effective hourly reserve requirement for the sample (Q_t^r) by multiplying the ERCOT hourly requirement $\left(Q_t^{r,ERCOT}\right)$ by the appropriate ratio of peak loads (in Equation 3.1).

$$Q_t^r = \begin{cases} \dfrac{\hat{\hat{L}}}{\hat{\hat{L}}^{ERCOT}} * Q_t^{r,ERCOT} & r = \{frequency\ response, Spin\ \&\ NonSpin\} \\ \\ \dfrac{\hat{L}}{\hat{L}^{ERCOT}} * Q_t^{r,ERCOT} & r = frequency\ regulation \end{cases} \tag{3.1}$$

Frequency response and peak reserves need specific consideration. Frequency response is an ancillary service that is not currently priced in the market; instead ERCOT requires that all generators over the size of 20 MW be able to provide this service. This is likely to change given FERC's recent notice of proposed rulemaking [11], NERC's BAL-003 standard [10], and the corresponding changes ERCOT has proposed to its ancillary services market [9]. Therefore, this paper includes frequency response in the set of ancillary services with an assumed price of $1/MW. The required quantity is based on ERCOT's current frequency response obligation set at 2,860 MW/Hz and an under-frequency load shed threshold of 59.3 Hz[2] [9].

Peak reserves are typically not considered an ancillary service. In some markets however (*e.g.*, PJM), peak reserves have their own unique forward auction market that is very different from the day ahead and real-time markets for energy and ancillary services. ERCOT does not have a peak reserves market, but ERCOT is developing programs that use direct load control to reduce load in

[2] $(60\ Hz - 59.3\ Hz) * 2{,}860\ MW/Hz = 2{,}002\ MW$.

peak hours [12,13]. This paper approximates DR's role in providing peak reserves based on the 20 highest load hours of the year in ERCOT and the sample's load contribution to ERCOT's peak load; this same method is used in DOE studies [35,36]. In those 20 hours, the peak load reduction requirement is proportional to the difference between that hour's load $\left(\hat{\hat{L}}^{ERCOT}{}_{j=1-20}\right)$ and the 21st highest load hour's load $\left(\hat{\hat{L}}^{ERCOT}{}_{j=21}\right)$ (Equation 3.2). The price of this peak reduction is equivalent to the electricity price in that hour.

$$Q_{r,j=1\dots20} = \frac{\hat{\hat{L}}}{\hat{\hat{L}}^{ERCOT}} * \left(\hat{\hat{L}}_j^{ERCOT} - \hat{\hat{L}}_{21}^{ERCOT}\right)$$
$$\forall j \in \{1, \dots, 20\}, where\ \hat{\hat{L}}_1^{ERCOT} < \hat{\hat{L}}_2^{ERCOT} < \cdots < \hat{\hat{L}}_{21}^{ERCOT}, \qquad r = peak\ reserves \qquad 3.2$$

3.2.3 Ancillary Service Deployment & Deferred Energy

The reservation of ancillary service capacity is economically allocated (*ex ante*) but the deployment of these resources is not. Instead the deployment is based on unpredictable reliability needs in real-time (*ex post*). Ahead of the hour, the operator sends each resource a base-point from which the resource will deviate in order to provide ancillary service. For ancillary services that require load to decrease its consumption when deployed, the base-point may be the same as its counterfactual consumption. For ancillary services that require load to increase its consumption when deployed the base-point will be a lower level of consumption than the counterfactual. In real-time, the system operator will send a control signal to ancillary services resources to move up or down away from their base-point. Most papers model the scheduling of ancillary service resources but do not attempt to model the effects of the *ex post* dispatch of these services. This paper closes this gap by defining statistical models that approximate the effect of this *ex post* dispatch and incorporating this information into the *ex ante* decision making.

Over the course of the hour, the net effect of the dispatch will be that the resources' average consumption may be different than their scheduled base-points. We define two key metrics of this movement:

Deferred energy $\left(\varphi_t^{i,r}\right)$ – the number of kWh of deviation from the counterfactual consumption per kW of ancillary service during hour *t*.

Deployment energy $\left(\delta_t^{i,r}\right)$ – the number of kWh deviation from the base-point of one kW of ancillary services capacity during hour *t*.

The key difference between these two metrics is that the base-point for providing an ancillary service may not be the same as the counterfactual consumption. For all of the reserves that deploy "up" (*i.e.,* increase generation, or decrease load: *frequency response, frequency regulation up, spin, non-spin, and peak reserves*), the base-point for providing the ancillary service can be the counterfactual load (Equation 3.3). But for reserves that deploy "down" (*i.e.,* decreasing generation or increasing load, frequency regulation down), the load has to be curtailed to a new base-point (less load than the counterfactual) in order to leave the ability to be deployed "down", *i.e.*, increasing load when called. A more detailed example of this is given in the appendix.

$$\varphi_t^{i,r} = \begin{cases} 1\,kW * 1hour - \delta_t^{i,r} & for\ r = Down\ Regulation \\ \delta_t^{i,r} & \forall r \neq Down\ Reg \end{cases} \qquad 3.3$$

The deployment energy for each ancillary service is estimated using statistical models based on ERCOT data [7,43,53]. The appendix provides the full details and the derivation of the parameter values of these models. The statistical deployment models for frequency regulation up and down are exponential distributions $\left(\sim Exp(\lambda)\right)$ defined by Equation 3.4, with fit parameter (λ) equal to 0.16

and 0.12, respectively. The statistical models for contingency reserves are a set of cascading Bernoulli trials $(\sim B(\rho))$, where each subsequent reserve is dependent on the deployment of the reserve previous in the hierarchy; the response order is frequency response, spinning reserve, and then finally non-spinning reserves (defined by Equations 3.5 – 3.7). The conditional probabilities for these three are: 0.9, 0.25, 0.5, for ρ_1, ρ_2, and ρ_3, respectively. When frequency response is deployed, the duration is defined by a Weibull distribution $(\sim W(\lambda, k))$ with fitted parameter values of $\lambda = 31.665$, and $k = 1.563$. Spinning and non-spin are assumed to be either fully deployed for the entire hour $(\delta_t^{i,r} = 1)$ or not deployed at all $\left(\delta_t^{i,r} = 0\right)$. Peak reserves do not require a statistical model because the deployment of peak reserves is economically dispatched (*ex ante*) and therefore peak reserve resources are always fully dispatched.

$$\delta_t^{i,r} \sim Exp(\lambda) \qquad \forall i, \forall t, r = frequency\ reguation\ up\ and\ down \tag{3.4}$$

$$\delta_t^{i,r} \sim B(\rho_1)\, W(\lambda, k) \qquad \forall i, \forall t, r = frequency\ response \tag{3.5}$$

$$\delta_t^{i,r} \sim B\left(\rho_2 \middle| \delta_t^{i,frequency\ response} > 0\right) \qquad \forall i, \forall t, r = spin \tag{3.6}$$

$$\delta_t^{i,r} \sim B\left(\rho_3 \middle| \delta_t^{i,spin} > 0\right) \qquad \forall i, \forall t, r = non - spin \tag{3.7}$$

The result of these models are realizations of potential ancillary services deployments that can be used to calculate the amount of energy a specific load aggregation will defer while providing ancillary services.

3.2.4 Ancillary Service Availability

A recent set of studies outlines methods and impacts of DR providing ancillary services [33-38]. Their approach, which we also employ here, is to “filter” time-series data of load down to only the portion that is flexible and able to provide ancillary services. This is represented in Equation 3.8, where load data $\left(L_t^i\right)$ is multiplied by a filter $\left(\phi_t^{i,r}\right)$ to create a time-series of ancillary service capacity $\left(A_t^{i,r}\right)$. The filter is the product of two parameters: *participation rate* (β), which represents the percentage of loads willing and able to respond; and *flexibility parameter* (f), which is representative of how flexible the load can be when providing a specific ancillary service; this is represented in Equation 3.9. In the context of this paper, the term *flexibility* is inherent to a load type, not dependent on a consumer's willingness to participate in ancillary service markets. The participation rate is the minimum of two parameters: *controllability* (η) – the portion of the load that has the necessary communications requirements to be controlled, and *acceptability* (α) – the "willingness of customers to shed their load in a particular hour", represented in Equation 3.10.

$$A_t^{i,r} = L_t^i \ \phi_t^{i,r} \quad 0 \leq \phi_t^{i,r} \leq 1 \ \forall i, \forall r, \forall t \tag{3.8}$$

$$\phi_t^{i,r} = \ \beta_t^i \ f_t^i \qquad 0 \leq \beta_t^i \leq 1 \ \forall i, \forall t \quad 0 \leq f_t^i \leq 1 \ \forall i, \forall r \tag{3.9}$$

$$\beta_t^i = \min\{\eta_t^i, \alpha_t^i\} \, 0 \leq \eta_t^i \leq 1 \ \forall i \quad 0 \leq \alpha_t^i \leq 1 \ \forall i, \forall t \tag{3.10}$$

In this chapter, we assume that all houses have the necessary broadband connection to receive ancillary service dispatch instructions and the necessary intra-household communication (*e.g.,* Wi-Fi) to transmit these instructions to flexible loads. This is equivalent to assuming that the controllability parameter $\left(\eta_t^i\right)$ is equal to one for all loads. This paper also assumes that the state-of-the-art technology will improve to a point where customers are willing to provide ancillary services, as long as the modified consumption is not noticeably different from their counterfactual consumption. This is equivalent to assuming that the acceptability parameter $\left(\alpha_t^i\right)$ is equal to one for all loads and time. These assumptions may seem generous given the level of today's demand response participation, the current state-of-the-art technologies used to control load aggregations, and our current understanding of the behavioral economics of the customers. However, we utilize these assumptions to assess the maximum potential for residential loads to contribute to the ancillary services market. Thus, the result of all these assumptions the absolute upper-bound for the potential for DR to provide ancillary services can be estimated using Equation 3.11.

$$A_t^{i,r} = L_t^i f_t^{i,r} \tag{3.11}$$

3.2.5 Loads' Flexibility & Flexibility Parameters

Residential load can provide ancillary services based on the aggregated load's ability to adjust consumption and each load type has a different innate ability to provide such flexibility based on its physical attributes, such as thermal mass. Further, each aggregate load's ability to provide an ancillary service depends on the type of ancillary service. For example, a load may be able to provide a different amount of frequency regulation than the amount it can provide for spinning reserve. Therefore, a unique estimate of a load's ability to provide ancillary services, or *flexibility*, is needed for each possible combination of end-use and ancillary service. This paper defines the flexibility parameter as the quantity of ancillary service capacity (kW) that can be provided by one kilowatt of load.

Table 3.4 shows the expected value and range for each flexibility parameter based on a literature review. The appendix provides a comprehensive list of sources for these data. In most cases, there are studies that demonstrate the ability of a specific end-use to provide a specific ancillary service. Where no such study exists, there is often a closely related study, from which a datum can be estimated. For example, no literature was found for freezers (meaning stand-alone freezer without a refrigerator) providing frequency response. However, there are many studies showing refrigerators (referring to the standard combination refrigerator-freezer) providing this service. Given how similar the dynamics of these two loads are, it is likely that they can provide similar amounts of flexibility. Similarly, many studies focus on spinning reserves but almost none include contingency reserves. Given that these two services are very similar, data can be estimated for contingency reserves based off of studies relating to spinning reserves.

Table 3.4. Summary of the literature review regarding flexibility parameters: DR's ability to provide ancillary services. The expected value of the distribution is in bold [minimum / maximum].

	Frequency Response	Frequency Regulation (Up & Down)	Spin & Non-Spinning	Energy (Peak Reserves)
	Expected [Minimum / Maximum]			
Space Heating	**0.20** [0.05 / 0.35]	**0.20** [0.05 / 0.35]	**0.60** [0.10 / 0.90]	**0.10** [0.01 / 0.20]
Cooling	**0.20** [0.05 / 0.35]	**0.20** [0.05 / 0.35]	**0.68** [0.10 / 1.00]	**0.10** [0.01 / 0.20]
Water Heating	**0.10** [0.05 / 0.20]	**0.10** [0.05 / 0.20]	**0.50** [0.10 / 0.90]	**0.25** [0.10 / 0.50]
Refrigeration	**0.40** [0.01 / 0.64]	**0.40** [0.01 / 0.64]	**0.50** [0.25 / 0.75]	**0.25** [0.10 / 0.50]
Freezers[A]	**0.40** [0.01 / 0.64]	**0.40** [0.01 / 0.64]	**0.50** [0.25 / 0.75]	**0.25** [0.10 / 0.50]
Delayable Loads[B]	-	-	**0.50** [0.10 / 1.00]	**0.50** [0.10 / 1.00]
Lighting	-	-	**0.10** [0.01 / 0.25]	**0.10** [0.01 / 0.25]
Cooking	-	-	-	-
Other	-	-	-	-

A – **The category "freezers" includes ice-makers and stand-alone freezers. The standard household freezer that is attached to the refrigerator is included in the refrigerator category.**

B – **The category of "delayable loads" includes pumping and irrigation loads as well as delayable residential appliances such as dishwashers, clothes washing machines, and electric clothes dryers.**

Note that there is significant variation for each unique flexibility parameter estimate. For example, the flexibility parameter for space heating to provide frequency response ranges between 0.05 and 0.35. Much of this uncertainty comes from variations in how the original study where the data came from was conducted (*e.g.,* empirical or theoretical; aggregated loads or single large loads, consideration of consumer comfort, *etc.*). The model uses triangular distributions, specified by the values in Table 3.4, to help define scenarios and run Monte Carlo simulations to understand the uncertainty in our answer. Triangular distributions are used over uniform distributions to give more weight to the median estimate. Flexibility parameters are drawn with a positive correlation between those of the same load-type. The appendix provides more details about the draw method and alternatives that were considered.

In addition to Monte Carlo analysis using the triangular distribution, three scenarios are considered: **Olsen *et al***, where the flexibility parameter are defined by Olsen et al [35]; **Expected**, where only the expected value of the distributions in Table 3.4 are used; and **Optimistic**, where only the upper end of the distribution in Table 3.4 are used.

3.2.6 Load Shifting and Charging

Many DR studies analyze the benefits of shifting consumption from the daily peak to off-peak hours. This chapter, however, focuses on ancillary services and only allows this type of "peak shaving" through the peak reserves ancillary service for the twenty highest peak hours of the year. The only other energy that is shifted in time is that which is deferred due to ancillary service deployment. Like previous studies [35], this chapter limits the ability to increase counterfactual load, or "charge", to specific hours of the day and to specific load types (Table 3.5). Additionally, an aggregation of a specific load that is allowed to charge $(\hat{C}_t^i)$ is only allowed to increase its load between the counterfactual load and that load's daily peak power (Equation 3.12).

$$\begin{cases} \hat{C}_t^i = \hat{L}_t^i - L_t^i & \forall i \in \{chargable\} \cap \ t \in Allowable\ hours \\ \hat{C}_t^i = 0 & \forall i \notin \{chargable\} \cup \ t \notin Allowable\ hours \end{cases} \quad 3.12$$

Table 3.5. Charging constraints by End-Use Categories. Based on Olsen *et al* [35].

End-Use Categories	Charging Allowed	Acceptable Charge Hours
Space Heating	No	
Cooling	Yes	6am-6pm
Water Heating	Yes	All hours
Refrigeration	Yes	All hours
Freezers	Yes	All hours
Delayable Appliances	Yes	All hours
Lighting	No	
Cooking	No	
Other	No	

3.2.7 Risk-Based Optimal Allocation of DR resources

The data resulting from Equation 3.11 are a set of time-series representing how much of each ancillary service each single end-use can provide. As previously mentioned, much of these data are mutually exclusive; *i.e.,* using one megawatt of air conditioning load to provide frequency regulation excludes that same megawatt from providing spinning reserves. This is similar to the dispatch dilemma of generators, where providing one megawatt of one product (*e.g.,* energy, spinning reserve, frequency regulation, *etc.*) excludes it from using that same megawatt of capacity for another product. This problem is easily solved when all of these mutually exclusive products are co-optimized.

This paper uses an analogous co-optimization to find the best allocation of the available ancillary services from residential DR. This is similar to previous studies, *e.g.,* [36], except that it only co-optimizes the ancillary services provided by the residential resources and does not consider generation resources. This is because, compared to ERCOT, the 92 homes in this sample is small - the total yearly load in ERCOT is 3 million times greater and the coincidental yearly peak load is 2.6 million times greater. The relatively small sample of residential homes would add a negligible amount of additional supply of ancillary services to ERCOT and would not affect generation dispatch or market prices. Therefore, it is unnecessary to model the full ERCOT market. Hence, the model treats all pricing as exogenous. As the number of customers participating in demand response grows, the effects on the market outcomes cannot be neglected any more. However, the methods outlined in this paper can be incorporated into a full production-cost-based, unit-commitment and economic dispatch model, but it is outside of the scope of this paper and none of the presented results are biased based on the simpler formulation presented herein.

The model maximizes the value of the ancillary services that residential DR resources can provide given the risk of response fatigue. This is equivalent to the benefit of reserving ancillary services capacity on DR less the cost of pre- or re-charging thermal loads due to the ancillary services deployment and less the cost of potential response fatigue (Equation 3.13). The next two subsections provide details on how to value the risk of response fatigue. For now, it should be noted that a penalty price $\left(\tilde{p}_t^{i,r}\right)$ of assigning one kilowatt of AS capacity to a load is estimated. When that penalty price is greater than the benefit (p_t^r), the optimization will assign less of that ancillary service to that load. On the other hand, the potential cost of customer response fatigue is a function of how the resources are committed and dispatched. Therefore, we formulate it as an iterative optimization procedure. The tilde $\left(\tilde{p}_t^{i,r}\right)$ on the penalty price is there to signify the fact that this penalty price is based on the allocation of DR in the previous iteration. The algorithm converges easily and produces allocations of DR that account for the risk of loads deferring too much energy and not being flexible in future hours.

$$\max_{q,c} \sum_{I} \sum_{T} \left(\sum_{R} q_t^{i,r} \left(p_t^r - \tilde{p}_t^{i,r} + E\left[\varphi_t^{i,r}\right] p_t^E \right) - c_t^i p_t^E \right) \qquad 3.13$$

Subject to:

$$\sum_{r} \frac{q_t^{i,r}}{f_t^i} \leq L_t^i \quad \forall i \; \forall t \qquad 3.14$$

$$\sum_{i} q_t^{i,r} \leq Q_t^r \quad \forall r \; \forall t \qquad 3.15$$

$$0 \leq q_t^{i,r} \qquad \forall i \; \forall r \; \forall t \qquad 3.16$$

$$0 \leq c_t^i \leq \hat{C}_t^i \quad t \in Allowable\ hours \qquad 3.17$$

$$\sum_{t \in d} c_t^i - \sum_{t \in d} \sum_{R} q_t^{i,r} \mathbb{E}\left[\varphi_t^{i,r}\right] = 0 \;\; \forall d \in D, \forall i \epsilon \{chargable\} \qquad 3.18$$

The optimization is subject to a number of constraints (Equations 3.14 – 3.18). The first constraint ensures that the amount of load needed to support the scheduled ancillary services capacity is less than the counterfactual load value (Equation 3.14). The second constraint requires that scheduled quantity of ancillary services is never more than the requirement (Equation 3.15). In standard production cost models, this would be an equality constraint; but given that we may not be able to satisfy the reserve requirement for all ancillary services, in all hours, and that generators are available to cover the remaining requirement, this criterion is relaxed to an inequality constraint. Equation 3.16 requires that scheduled ancillary service quantities $\left(q_t^{i,r}\right)$ are greater than zero.

The final two constraints (Equations 3.17 – 3.18) deal with how specific loads can pre-charge or re-charge in order to compensate for energy they may forgo while providing ancillary services. Not all loads can pre-charge or re-charge in order to make up for deployments. For example, it is unrealistic to try to make up for lost lighting load by increasing the lumens later on. As previously shown, Table 3.5 and Equation 3.12 (same as Equation 3.17) show which loads can charge, how much they can charge, and the hours in which they are allowed to charge. The final charging constraint (Equation 3.18) makes sure that the daily total energy consumption of a specific load is as close to the counterfactual consumption as possible. It requires that the total quantity of charging, in kilowatt-hours $\left(\sum_{t\in d} c_t^i\right)$, is greater or equal to the expected amount of deferred energy $\left(\sum_{t\in d}\sum_R q_t^{i,r}\mathbb{E}\left[\varphi_t^{i,r}\right]\right)$. The reason this constraint is based on the expectation, and not the actual amount, of deferred energy is that we do not know at the time of assigning ancillary services capacity (*a priori*), what the amount of deferred energy will be for a given ancillary service. The commitment of ancillary services capacity takes place ahead of real-time (*ex ante*) but its dispatch takes place based on real-time reliability needs (*ex post*). Therefore, the charging constraint in Equation 3.18 has to be based on an expectation of the effect of ancillary services deployment $\left(\mathbb{E}\left[\varphi_t^{i,r}\right]\right)$.

3.2.8 Assessment of Risk & Penalty Pricing

This section describes one possible way to value the risk of consumer fatigue for residential demand resources providing ancillary services. This method demonstrates the ability to incorporate an expectation of *ex post* risk for consumer fatigue into *ex ante* demand response scheduling.

The model schedules DR resources for ancillary services based on the expectation of being deployed. It runs one hundred realizations of the statistical ancillary services deployment model and calculates the risk, and its value, for each realization. This model assumes that the risk of consumer fatigue is directly tied to the amount of deferred energy; more specifically, how much deferred energy is unaccounted for after scheduling charging based on the expectation of deferred energy (Equation 3.18). Net-Deferred Energy $\left(\Delta_{t\in d}^i\right)$ is defined as the difference between the daily deferred energy and the daily charging energy of a load (Equation 3.19). It is reasonable to assume that one kilowatt-hour of net-deferred energy on a one kilowatt-hour aggregate load leads to much more consumer fatigue than one kilowatt-hour of deferred energy for a one megawatt-hour aggregate load. Therefore, percent charging error $\left(\Psi_{i,d}\right)$, is defined in Equation 3.20 as the ratio of the net-deferred energy to the total daily load. Note how similar this is to the charging constraint in the optimization; the key difference is that this equation uses statistical realizations of deferred energy rather than the expectation of deferred energy.

$$\Delta_{t\in d}^{i} = \sum_{t}\sum_{R}\varphi_{t}^{i,r} - \sum_{t} c_{i,t} \quad \forall i, \forall t \in d \tag{3.19}$$

$$\Psi_{d}^{i} = \frac{\Delta_{t\in d}^{i}}{\sum_{t\in d} L_{t}^{i}} \quad \forall i, \forall t \in d \tag{3.20}$$

It seems reasonable to assume that a little net-deferred energy would be acceptable; therefore, any realization with less than 10% percent charging error is not assessed a penalty. After 10% charging error, any additional net-deferred energy is charged with the highest energy price of the day (Equation 3.21). The result is a total dollar value representing the market risk (Π_{d}^{i}) for each realization. From these data, the model produces a cumulative distribution function of the penalties costs that are likely to occur in that day, representing the risk profile of the given allocation of DR resources. From this distribution, a decision maker (*i.e.,* ISO or RTO) can select a value according to their risk aversion (*e.g.,* median or mean value for risk neutral, high values for the risk adverse, *etc.*). This chapter uses the maximum penalty cost for each load as the representative risk to model a risk-adverse decision-maker.

$$\Pi_{d}^{i} = \begin{cases} \left(\Psi_{d}^{i} - 0.10\right)\sum_{t\epsilon d} L_{i,t} \max[p_{t\in d}^{E}] & if\ \Psi_{d}^{i} > 0.1 \\ 0 & if\ \Psi_{d}^{i} \leq 0 \end{cases} \tag{3.21}$$

It is important to remember that there are multiple ancillary services contributing to each load's net-deferred energy and each ancillary service may not be contributing equally. The model allocates the total market risk ($) to each of the ancillary services based on their contribution to the net-deferred energy. First, it calculates the contribution ($w_t^{i,r}$, Equation 3.22) of each ancillary service to the daily quantity of net-deferred energy, for a given load. This is done by dividing the total daily deferred energy from each ancillary service on a single load by the total daily deferred energy for that load. For example, if air conditioning is providing frequency regulation and spinning reserve, Equation 3.22 calculates how much of the air conditioning's daily deferred energy comes from each of the two ancillary services.

$$w_{t\in d}^{i,r} = \frac{\sum_{t\in d}\left(q_t^{i,r} * \varphi_t^{i,r}\right)}{\sum_{t\in d}\sum_r q_t^{i,r} * \varphi_t^{i,r}} \qquad \forall i, \forall r \qquad (3.22)$$

A penalty price is calculated (Equation 3.23) by multiplying the respective penalty weight by the total penalty cost and dividing it by the respective scheduled quantity of each ancillary service type. The result creates a daily penalty price ($/kW-Hr of ancillary services capacity) specific to each load type and ancillary service, and represents the approximate daily risk of scheduling a unit of ancillary services capacity on a specific load aggregation.

$$\tilde{p}_{t\in d}^{i,r} = \frac{\Pi_d^i * w_t^i}{\sum_{t\in d} q_t^{i,r}} \quad \forall t \in d \qquad (3.23)$$

3.3 Results & Discussion

The results presented herein use two different metrics to measure the impact of residential demand response: ancillary service quantity $(q_t^{i,r})$ and percentage of hourly market value (τ_t). The quantification of ancillary services provided by residential DR, as compared to the reserve requirements, gives a good indication for the size of the impact that residential DR can provide. We could also report the market value of all DR resources $\left(\$ = \sum_r \sum_i p_t^r q_t^{i,r}\right)$. However, this is a misleading as this paper uses exogenous pricing data, the DR resources add a non-marginal additional supply, and the paper specifically attempts to measure the impact of DR resources that are ubiquitous. The market value of DR resources would overestimate the benefits of residential DR providing ancillary services because it does not incorporate the elasticity of supply and the resulting depression in market prices and costs that would accompany a non-marginal change in supply. Instead, this paper uses the percentage of hourly market value of the DR resources (Equation 3.24) as a means of creating one quantitative datum, by which the scenarios can be compared to each other. The hourly market value addresses the aforementioned issues by normalizing the market value of the DR resources by the total market value. This is essentially a price-weighted average of the total ancillary service quantity provided in each hour and provides a quantitative measure of how large the effect of DR resources in the market for that hour is.

$$\tau_t = \frac{\sum_r \sum_i p_t^r q_t^{i,r}}{\sum_r p_t^r Q_t^r} \qquad 3.24$$

3.3.1 Without Risk-based Feedback

The model in this paper optimally allocates ancillary services on residential demand resources with, or without, an estimate of the *ex ante* risk of consumer response fatigue. This section provides results from the simulations without any incorporation of risk. Figure 3.3 shows the quantity of ancillary services capacity (kW-Hr) scheduled on each load type over a year, using the data from the 92 houses extracted from the Pecan Street Project data. Approximately 90% (9.6×10^4 of 10.8×10^4 kW-Hr) of all the scheduled ancillary services capacity is scheduled on cooling (5×10^4 kW-Hr, 47%), refrigeration (2.6×10^4 kW-Hr, 24%), and electric space heating (2×10^4 kW-Hr, 19%). Delayable appliances (5%), Lighting (4.5%), and water heating (0.5%) together contribute the remaining amount.

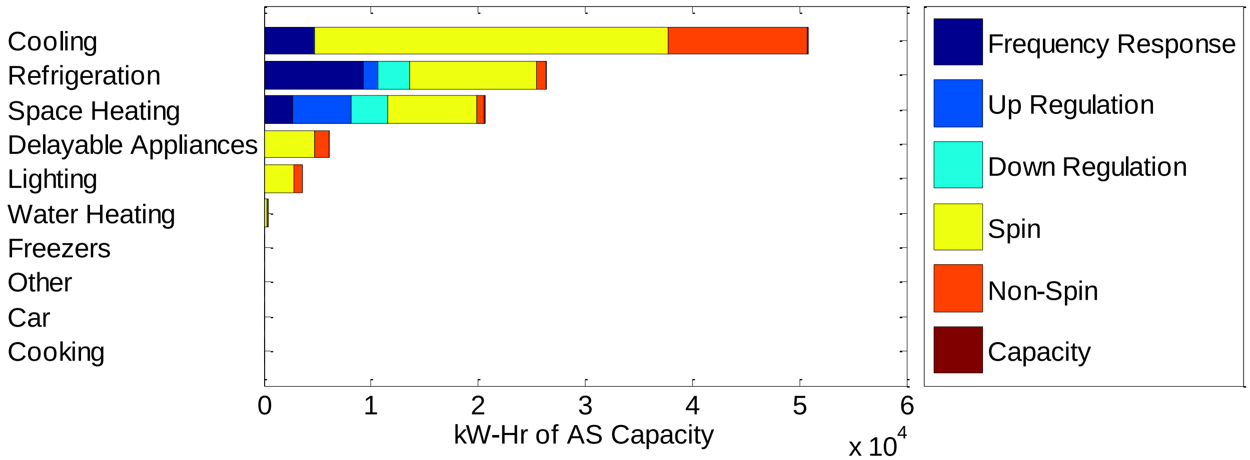

Figure 3.3 Amount of ancillary services capacity scheduled on each load type.

In order to provide context to these numbers, these data are compared against the sample's contribution to ERCOT ancillary service requirements. Figure 3.4 shows the residential DR resource's contribution to the hourly ancillary service markets in terms of quantity (top) and value (bottom). Each boxplot is a different realization from the Monte Carlo analysis and each datum in a boxplot represents an hour in a year. The median contribution of residential DR is between 50% to 75% of the hourly total ancillary service requirements and 75% to 95% of hourly market value (τ). The uncertainty in these ranges represents the uncertainty in the flexibility parameters. Even in pessimistic draws (left side), residential DR resources are likely to provide over 50% of the total requirements and 75% of the hourly market value, on average.

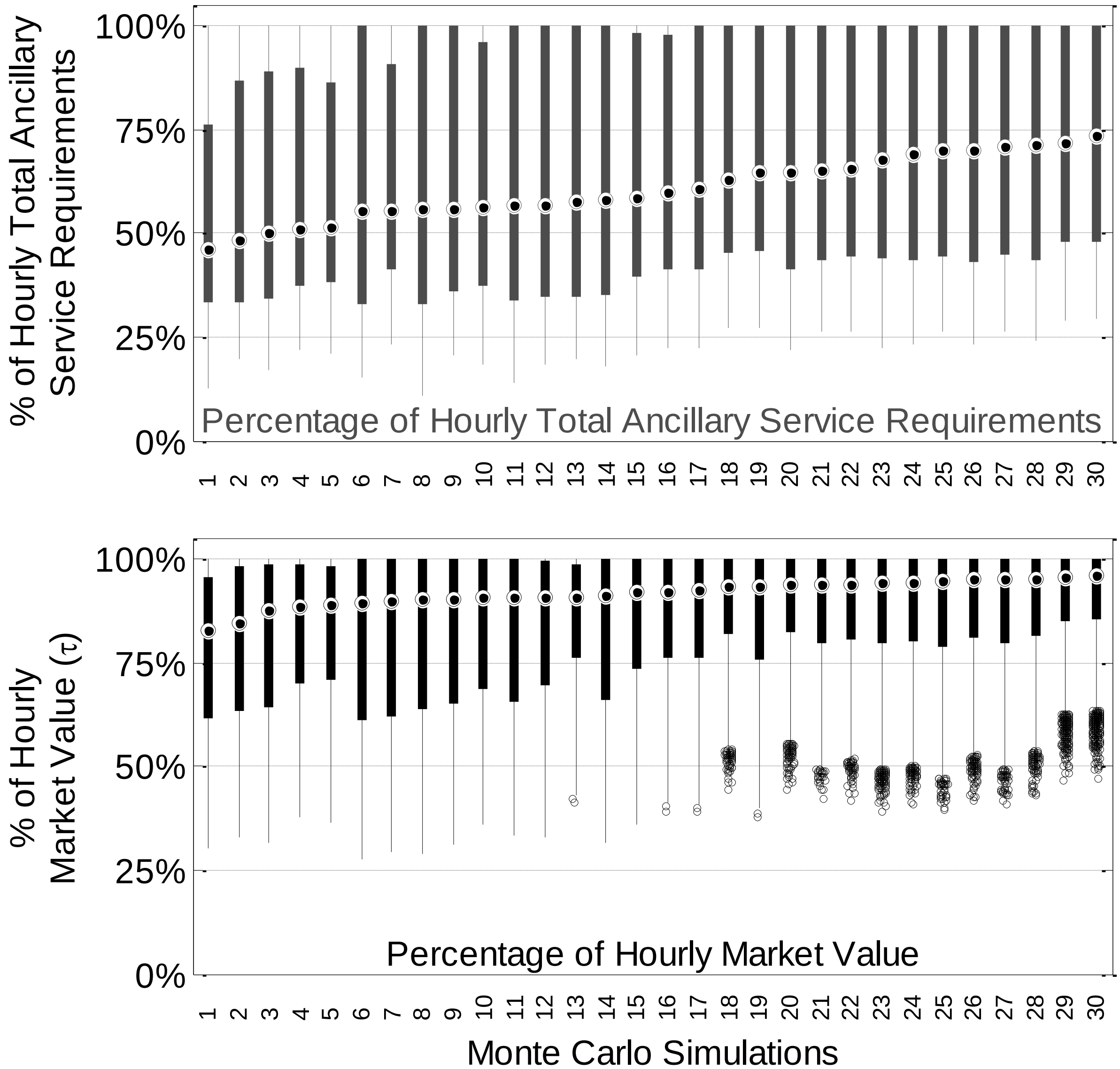

Figure 3.4. Percentage of hourly total ancillary service requirements (top) and percentage of hourly market value (τ, bottom)for each of the thirty modeled Monte Carlo simulations, ordered in ascending median value. Even in pessimistic draws (left side), residential DR resources are likely to provide over 50% of the total ancillary service requirements and 75% of the hourly market value, on average.

The key difference between the two metrics in Figure 3.4 is that the quantity metric (top) values all ancillary services equally; where the value metric (bottom) values each ancillary service based on their respective price. The optimization is choosing the higher value ancillary services first, resulting in more scheduled capacity of higher valued ancillary services and a larger effect on hourly market value than the total ancillary service requirements. This trade-off can be seen in Figure 3.5, which depicts what percentage of the ancillary services requirement (attributable to our residential load) is scheduled on DR in each hour of the year, for each type of ancillary service.

The shaded region represents the range of results from the Monte Carlo analysis; the individual lines represent different scenarios: **Olsen *et al***, where the flexibility parameter are defined by Olsen *et al* [35]; **Expected**, where only the expected value of the distributions in Table 3.4 are used; and **Optimistic**, where only the upper end of the distribution in Table 3.4 are used. For each ancillary service type, a non-trivial portion of the requirement is scheduled on DR resources. Note that residential DR resources are more likely to meet the full requirements for peak reserves, the highest valued ancillary service, than frequency response, the lowest valued product.

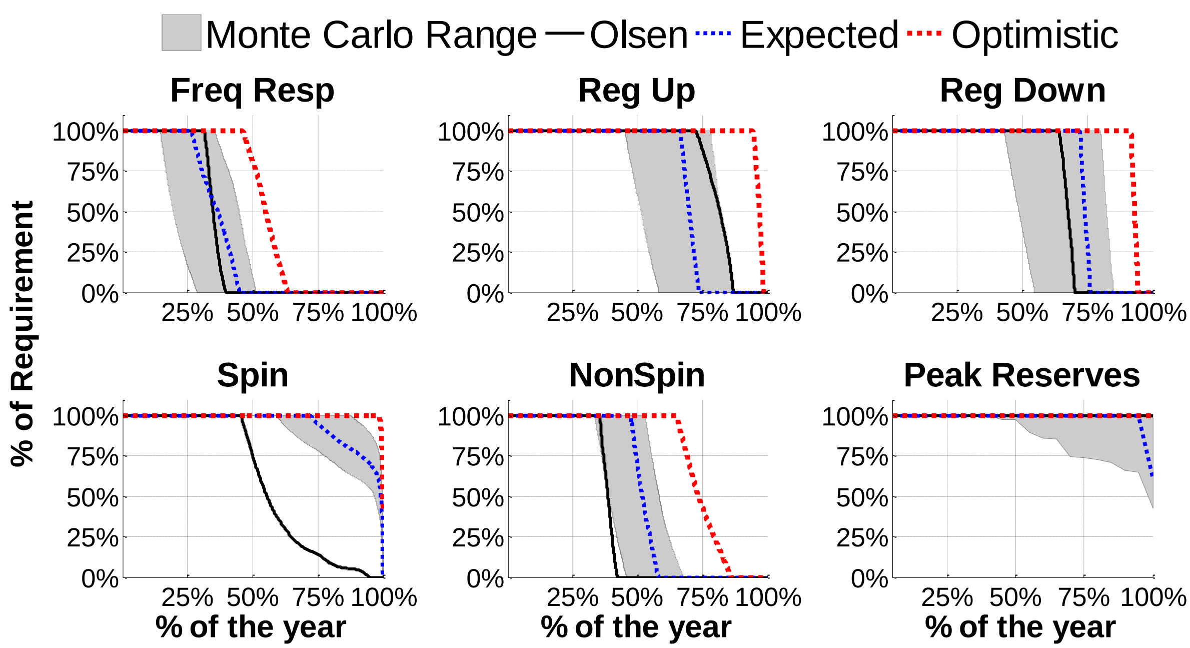

Figure 3.5 Duration curves showing the percentage of ancillary services requirement attributable to the Pecan Street load that are met by residential DR. Three cases are presented: Olsen represents the parameters used in Olsen *et al* (2013) [35]. The expected and optimistic cases are based on the expected value and maximum value (respectively) of each flexibility parameter distribution (see Table 3.4).

3.3.2 With Feedback

In this section, we now employ the method of assessing the *ex post* risk and incorporating this information into the *ex ante* scheduling of residential DR resources for ancillary services as described earlier. The model is used to test the effects of limiting DR's participation in ancillary service markets as compared to the risk-based feedback. DR's participation is capped at an hourly maximum percentage of supply of each ancillary service. Table 3.6 shows how the different hourly market caps affect the scheduling of demand response resources, as compared with the risk-based feedback without an hourly cap. Not surprisingly, hourly caps significantly reduce the quantity of ancillary services capacity that is scheduled and the value of the ancillary service capacity that is provided by DR resources. For example, a cap of 25% of the hourly ancillary services requirement schedules 4.6 kW-Hr of total ancillary service capacity from DR resources, only 37% of the 12.3 kW-Hr of total ancillary service capacity that is scheduled using risk-based feedback. Additionally, a 25% hourly cap limits the contribution from residential DR resources to ancillary services to approximately 25% of the hourly market value; the risk-based feedback method schedules 86% of the hourly market value on average. Using 10x, 100x, and 1,000x the maximum daily energy price as the basis for calculating the penalty price shows no significant change in the results. More details are provided in the appendix.

Table 3.6. Comparison of hourly market caps to a risk-based optimization. For each hourly market cap, the average amount of ancillary services scheduled on residential DR (q) is compared to the ancillary services requirement (Q). Also shown is the percentage of hourly market value of the scheduled ancillary services (τ) scheduled on the residential DR resources.

Market Cap	Average AS capacity (kW-Hr)			Average Percentage of Hourly Market Value
	q	Q	$\frac{q}{Q}$	τ
No Cap No Feedback	**12.3**		**67%**	**86%**
75%	**10.4**		**56%**	**70%**
50%	**8.0**	**18.4**	**43%**	**49%**
25%	**4.6**		**25%**	**25%**
Risk-Based Feedback	**12.1**		**65%**	**85%**

3.3.3 Thermal Battery Models and Modeling Tradeoffs

A majority of the loads modeled in this chapter are aggregations of homogeneous thermostatically controlled loads (TCLs). There are many different types of mathematical models that essentially treat TCLs loads as a thermal battery, shifting consumption of electricity in time. Future research could repeat this chapter's methods using a thermal battery model for each end-use aggregation instead of using the flexibility parameters. The section highlights the advantages of each of these methods.

There are many different ways to model an aggregation of TCLs but they can most easily be understood as a thermal battery model. The thermostatically controlled loads (*e.g.,* air conditioner) have constraints on the temperature range (*e.g.,* 70-75°F) for a thermal mass (*e.g.,* the house) and properties that explain how the system transitions from one state to another (*e.g.,* heat transfer coefficients, thermal loadings, time constants, *etc.*). The result is that the consumption of each individual load can be modified based on its particular parameters such that energy can be stored (*e.g.,* pre-cooling) or withdrawn (*e.g.,* delaying cooling) as if it were a battery. Each individual load may only be able to "*store*" a small amount of energy for a short amount of time, but as an aggregation the quantity and duration of the "*battery*" grows.

The ability to modify the total consumption of an aggregation of TCL's is its *flexibility*. The flexibility model eliminates the much of the complexity of modeling aggregations of TCL's by assuming a time-independent parameter that represents how well, on average, an aggregation of TCL's can perform a particular ancillary service.

Given that the flexibility model uses average values to represent an aggregation's *flexibility*, the battery model would show more variation in the load's ability to provide ancillary services. There will be some hours where the thermal battery model shows a load is more flexible and other ours when it is less flexible than the average flexibility parameter. Whether this greater variation creates a bias in our results, it is unclear. However, it is not likely that adding this additional complexity would change the conclusions drawn in this chapter.

The advantage of using a thermal battery model for each individual load aggregation TCLs is that the model implicitly captures the aggregation's flexibility based on its current state and allows for this flexibility to change with time as the aggregation's state changes. This would be equivalent to creating a time-dependent flexibility parameter in this chapter's methods. The downside to creating a battery model for each individual aggregation is that it adds significant complexity to an already large model. The benefit of the flexibility model is that it could be more easily adopted by utilities looking to estimate their demand response potential, without needing to hire an expert in modeling heterogeneous thermostatically controlled loads.

The battery method might be categorized as a bottom up method, as most thermal battery models use a known, and fixed, quantity of homogeneous loads. The flexibility model on the other hand can be used in a top-down analysis where nothing is known about the individual loads. The empirical data in this chapter poses a problem when using a thermal battery model. The empirical data has a known quantity of loads in each aggregation but the number that are '*on*' varies with time and is not known for all load types. Additionally, the parameters of each load (*e.g.,* thermal mass, heat transfer coefficients, time-constants, *etc.*) are not explicitly known. This makes it very difficult to know the precise characteristics of the aggregation. This hurdle could potentially be overcome using machine learning techniques [54,55], but we consider this to be beyond the scope of this thesis.

In short, building a "correct" battery model would take a substantial research effort that is likely not needed in order to address the research goals of this chapter.

3.3.4 Uncertainty

There are a number of sources of uncertainty in the results that are worth noting. First, the Pecan Street sample is not geographically representative of the average U.S. home. There is significant variation in the types of loads, their contribution to a home's total energy consumption [56], as well as variation in market designs and ancillary service requirements. Therefore, these results are only representative of homes in similar climates but we can still infer from these results how DR resources in other regions may impact their respective markets.

There is a large amount of uncertainty in these results and around the assumptions made in this chapter. First, the analysis presumes an extreme future where communication technologies and controllable appliances are extremely prevalent and acceptable to customers, making direct load control possible in every home. Changing this assumption to a less optimistic future simply scales down our results. A second major contributor to uncertainty is the estimate of how flexible a certain load can be. Loads could be more or less flexible based on our ability to control aggregated loads while keeping consumers happy and willing to participate. The answer of how much of a market impact residential DR could have greatly depends on this assumption. We are encouraged by our results with regards to this uncertainty for two reasons: i) the highest value ancillary services are selected first, meaning that a 1% reduction in the participating demand response will not create less than a 1% reduction in economic value; ii) large ancillary service deployments are rare and can be hedged through smart scheduling. Despite these uncertainties, the results show that residential DR can provide a significant portion of residential load's contribution to wholesale ancillary services requirements.

Finally, customer response fatigue is a phenomenon that is not well understood. This paper presents one potential mechanism: that significant deviation from daily counterfactual consumption leads to response fatigue. This mechanism is far from proven and more research is needed before this phenomenon can be accurately modeled and predicted. The methods presented here are a framework for optimally scheduling DR resources for ancillary services and can be adapted to incorporate any new understanding of consumer response fatigue.

3.4 Conclusions and Policy Implications

Residential demand response could be a significant contributor to ancillary services markets in the future. In our model, cooling, refrigeration, and electric space heating provide approximately 90% of all the ancillary services capacity scheduled on the Pecan Street residential demand resources. Delayable appliances (5%), Lighting (4.5%), and water heating (0.5%) together contribute the balance. The results suggest that residential DR could provide between 35% and 60% of residential load's contribution to wholesale ancillary services requirements and up to 80% of its respective market value. This could be a game-changer in ancillary service markets that could be facing higher demand with rising renewable penetration and lower supply given retirements of traditional thermal generation. The novel risk-based optimization framework presented here could have a number of real world applications including use in load aggregators' and grid operators' scheduling and resource optimization.

Additionally, the results suggest that hourly caps are not an economically efficient means of addressing consumer response fatigue as compared to incorporating that *ex ante* risk into the commitment of ancillary services capacity on DR resources. Future research to better quantify how much flexibility load can provide; the adoption of enabling technologies; and the behavioral economics that affect how much load will customers be willing to turn over to direct load control could further refine the model in this paper. The results presented here suggest that the impact of residential DR is great enough to justify further research into these policies.

The results of our simulation of hourly market caps suggest that this type of policy is not an economically efficient means of addressing consumer response fatigue as compared to incorporating that ex ante risk into the commitment of ancillary services capacity on DR resources. We hope that grid operators' who employ this or similar approaches seriously consider advancing their method and more precisely target the risk to their system. This could result in a more economically efficient market and significant savings to the rate-payers.

Future research to better quantify how much flexibility load can provide, the adoption of enabling technologies, and the behavioral economics that affect how much load will customers be willing to turn over to direct load control could further refine the model in this paper. The results presented here suggest that the impact of residential DR is great enough to justify further research into these policies.

3.5 References

1. Oliver J Radio control of water heaters and distribution station voltage regulators. In: IEEE Summer Power Meeting, Dallas, Texas, 1969.

2. Kirby B, Hirst E (1999) Load as a resource in providing ancillary services. Lockheed Martin Energy Research, Oak Ridge National Laboratory Oak Ridge, TN

3. Kirby B, Hirst E, Vancoevering J (1995) Unbundling electric generation and transmission services. Energy 20 (12):1191-1203

4. MISO (2004) Midwest Independent Transmission System Operator, Inc's Open Access Transmission and Energy Markets Tariff dated 3/31/04 under ER04-691. Volume 2, Modules A & B. http://elibrary.ferc.gov/idmws/common/opennat.asp?fileID=10105516

5. PJM (2011) PJM Manual 11: Energy and Ancillary Services Market Operations: Section 3.2.4.: Determination of Regulation Requirement.
http://pjmcom/~/media/documents/manuals/m11ashx

6. NERC (2013) Reliability Standards for the Bulk Electric System of North America.
http://wwwnerccom/docs/standards/rs/Reliability_Standards_Complete_Setpdf

7. Kirby B (2007) Ancillary services: Technical and commercial insights.

8. Kirby B (2006) Demand Response For Power System Reliability: FAQ.

9. ERCOT (2013) Future Ancillary Services in ERCOT. ERCOT CONCEPT PAPER. ERCOT,
http://www.ercot.com/content/committees/other/fast/keydocs/2014/ERCOT_AS_Concept_Paper_Version_1.1_as_of_11-01-13_1445_black.doc

10. FERC (2014) Frequency Response and Frequency Bias Setting Reliability Standard. 146 FERC¶ 61,024. NERC,
http://www.nerc.com/FilingsOrders/us/FERCOrdersRules/Frequency%20Response%20and%20Frequency%20Bias%20Setting%20Reliability%20Standard.pdf

11. FERC (2015) Third-Party Provision of Primary Frequency Response Service - Notice of proposed rulemaking. Docket No. RM15-2-000, vol 150 FERC ¶ 61,092;.
http://www.ferc.gov/whats-new/comm-meet/2015/021915/E-2.pdf

12. ERCOT (2013) Load Participation in the ERCOT Nodal Market: Financial Opportunities for Reducing Electricity Load. version 2.0 edn.,
http://www.ercot.com/content/services/programs/load/Load%20Participation%20in%20the%20ERCOT%20Nodal%20Market%20v2.doc

13. ERCOT (2014) Load Resource Participation in the ERCOT Markets. ERCOT.

14. Kueck J, Kirby B, Eto J, Staunton R, Marnay C, Martinez C, Goldman C (2001) Load as a reliability resource in restructured electricity markets. ORNL/TM2001/97, LBNL-47983 June

15. Kirby BJ, Kueck JD (2003) Spinning Reserve from Pump Load: A Technical Findings Report to the California Department of Water Resources. United States. Department of Energy,

16. Todd D, Caufield M, Helms B, Generating AP, Starke IM, Kirby B, Kueck J (2008) Providing reliability services through demand response: A preliminary evaluation of the demand response capabilities of alcoa inc. ORNL/TM 233

17. Eto JH, Nelson-Hoffman J, Torres C, Hirth S, Yinger B, Kueck J, Kirby B, Bernier C, Wright R, Barat A (2007) Demand response spinning reserve demonstration.

18. Kiliccote S, Piette MA, Koch E, Hennage D Utilizing Automated Demand Response in commercial buildings as non-spinning reserve product for ancillary services markets. In: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, 2011. IEEE, pp 4354-4360

19. Hovgaard TG, Larsen LFS, Jorgensen J Flexible and cost efficient power consumption using economic mpc a supermarket refrigeration benchmark. In: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, 2011. IEEE, pp 848-854

20. Zhang Y, Lu N Demand-side management of air conditioning cooling loads for intra-hour load balancing. In: Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES, 2013. IEEE, pp 1-6

21. Callaway DS (2009) Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. Energy Conversion and Management 50 (5):1389-1400

22. Koch S, Mathieu JL, Callaway DS Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services. In: Proc. PSCC, 2011. pp 1-7

23. Hovgaard TG, Larsen LF, Edlund K, Jørgensen JB (2012) Model predictive control technologies for efficient and flexible power consumption in refrigeration systems. Energy 44 (1):105-116

24. Sullivan M, Bode J, Kellow B, Woehleke S, Eto J (2013) Using Residential AC Load Control in Grid Operations: PG&E's Ancillary Service Pilot.

25. Zhao P, Henze GP, Plamp S, Cushing VJ (2013) Evaluation of Commercial Building HVAC Systems as Frequency Regulation Providers. Energy and Buildings

26. Gifford W (2010) Customer Impact Evaluation for the 2009 Southern California Edison Participating Load Pilot.

27. Hammerstrom DJ, Brous J, Chassin DP, Horst GR, Kajfasz R, Michie P, Oliver TV, Carlon TA, Eustis C, Jarvegren OM (2007) Pacific Northwest GridWise™ Testbed Demonstration Projects; Part II. Grid Friendly™ Appliance Project. Pacific Northwest National Laboratory (PNNL), Richland, WA (US),

28. Borenstein S (2005) The long-run efficiency of real-time electricity pricing. The Energy Journal:93-116

29. Borenstein S (2005) Time-Varying Retail Electricity Prices: Theory and Practice. Electricity deregulation: choices and challenges:317

30. Borenstein S, Holland SP (2005) On the Efficiency of Competitive Electricity Markets with Time-Invariant Retail Prices. RAND Journal of Economics 36 (3):469-493

31. Su C-L, Kirschen D (2009) Quantifying the effect of demand response on electricity markets. Power Systems, IEEE Transactions on 24 (3):1199-1207

32. FERC (2009) A national assessment of demand response potential. Government Printing Office, Washington, DC. Federal Energy Regulatory Commission,

33. Ma O, Alkadi N, Cappers P, Denholm P, Dudley J, Goli S, Hummon M, Kiliccote S, MacDonald J, Matson N (2013) Demand response for ancillary services. Smart Grid, IEEE Transactions on 4 (4):1988-1995

34. Starke M, Alkadi N, Ma O (2013) Assessment of Industrial Load for Demand Response across US Regions of the Western Interconnect. Technical report ORNL/TM-2013/407. Oak Ridge, TN: Oak Ridge National Laboratory,

35. Olsen DJ (2013) Grid Integration of Aggregated Demand Response, Part 1: Load Availability Profiles and Constraints for the Western Interconnection.

36. Hummon M, Palchak D, Denholm P, Jorgenson J, Olsen DJ, Kiliccote S, Matson N, Sohn M, Rose C, Dudley J (2013) Grid Integration of Aggregated Demand Response, Part 2: Modeling Demand Response in a Production Cost Model.

37. Hummon M, Denholm P, Jorgenson J, Palchak D, Kirby B, Ma O (2013) Fundamental Drivers of the Cost and Price of Operating Reserves. National Renewable Energy Laboratory (NREL), Golden, CO.,

38. Ma O, Alkadi N, Bhatnagar D, Cappers P, Currier AB, Denholm P, J. Dudley J, Goli S, Hummon H, Jorgenson J, Kiliccote S, MacDonald J, Matson N, Olsen D, Palchak D, Rose C, Sohn MD, Starke M, Kirby B, O'Malley M (*Forthcoming*) Demand Response and Energy Storage Integration Study. http://www1.eere.energy.gov/analysis/response_storage_study.html

39. Goldman C, Hopper N, Bharvirkar R, Neenan B, Cappers P (2007) Estimating demand response market potential among large commercial and industrial customers: a scoping study. Lawrence Berkeley National Laboratory

40. Callaway DS, Hiskens IA (2011) Achieving controllability of electric loads. Proceedings of the IEEE 99 (1):184-199

41. PJM (2014) Manual 11: Energy & Ancillary Service Market Operations; 3.2.4 Regulation Requirement Determination. http://www.pjm.com/documents/manuals.aspx

42. MISO (2014) BPM 02 Attachment B: Day-Ahead Energy and Operating Reserve Market Software Formulations and Business Logic; 3.34 Generation-Based Operating Reserve Requirement Factor.

https://www.misoenergy.org/Library/BusinessPracticesManuals/Pages/BusinessPracticesManuals.aspx

43. ERCOT (2014) Fast Responding Regulation Service Pilot. ERCOT.

44. FERC (2006) Assessment of demand response and advanced metering. Docket AD-06-2-000. Federal Energy Regulatory Commission,

45. FERC (2008) Assessment of demand response and advanced metering. Department of Energy. Federal Energy Regulatory Commission,

46. FERC (2011) Assessment of Demand Response and Advanced Metering. Federal Energy Regulatory Commission,

47. EIA (2009) The National Energy Modeling System: An Overview 2009. U.S. Energy Information Administration, http://www.eia.gov/oiaf/aeo/overview/pdf/0581(2009).pdf

48. Smith CA (2010) The Pecan Street Project: developing the electric utility system of the future.

49. Kara E, Bergés M, Hug G Modeling Thermostatically Controlled Loads to Engage Households in the Smart Grid: Lessons Learned from Residential Refrigeration Units. In: Computing in Civil and Building Engineering (2014). ASCE, pp 2032-2039

50. Mathieu JL (2013) Modeling, analysis, and control of demand response resources.

51. Weis A, Jaramillo P, Michalek J (2014) Estimating the potential of controlled plug-in hybrid electric vehicle charging to reduce operational and capacity expansion costs for electric power systems with high wind penetration. Applied Energy 115:190-204

52. ERCOT (2014) ERCOT Nodal Protocols. Section 3.16: Standards for Determining Ancillary Service Quantities. Electric Reliability Council Of Texas http://www.ercot.com/mktrules/nprotocols/current

53. Dumas J (2011) Historical Non-Spin Deployment Data. ERCOT, http://www.ercot.com/content/meetings/rdtf/keydocs/2011/0906/NS_history_and_floor_proposals.pdf

54. Berges M, Goldman E, Matthews HS, Soibelman L Learning systems for electric consumption of buildings. In: ASCI international workshop on computing in civil engineering, 2009.

55. Berges ME, Goldman E, Matthews HS, Soibelman L (2010) Enhancing electricity audits in residential buildings with nonintrusive load monitoring. Journal of industrial ecology 14 (5):844-858

56. EIA (2014) RESIDENTIAL ENERGY CONSUMPTION SURVEY (RECS). http://www.eia.gov/consumption/residential/. 2014

Chapter 4: CONCLUSIONS

This dramatic shift in both fossil-fueled and renewable generation will undoubtedly affect the cost of energy as well as the cost of ancillary services. The additional variability inherent to wind and solar power will increase the need for balancing using flexible resources providing ancillary services. At the same time, thermal generation resources are retiring due to environmental regulations, decreasing supply these ancillary services. The net economic impact of these is still highly uncertain but could result in higher prices and costs for ancillary services if nothing is done to mitigate this rise.

Balancing area consolidation provides economic benefit in the energy and frequency regulation market. Wide-scale consolidation reduces the relative variability of net load, which reduces the aggregate requirement for frequency regulation. The data suggest that this effect exhibits diminishing marginal returns, meaning consolidating the first few BAs produces a larger reduction than the final few BAs. This this analysis does not consider all the benefits or costs of BA consolidation, and is not meant as an assessment of net-benefits. That said, the reduced frequency regulation requirement, combined with the effect of shared resources, leads to a reduction in frequency regulation cost of approximately $0.1 per MWh of total load. These results do not significantly change with the inclusion of 20% wind, suggesting that in the near term, wind's interaction in the frequency regulation market is not a prime motivation for consolidation. The data show that CO_2 and $PM_{2.5}$ emissions increase, while NO_X emissions dramatically decrease, and SO_2 slightly decreases. Given the relative health and climate risks each of these pollutants, it is likely that this change in emissions results in a net cost to society. This suggests that while there may be economic benefits to BA consolidation, BA consolidation may not be as beneficial today as previous literature suggest depending on the generation portfolio of the considered system and the level of penetration of renewable generation. Hence, a blanket policy to support balancing area consolidation in fossil-based systems may result in unexpected consequences such as increased emission.

Residential demand response could be a significant contributor to ancillary services markets in the future. In our model, cooling, refrigeration, and electric space heating provide approximately 90% of all the ancillary services capacity scheduled on the Pecan Street residential demand resources. Delayable appliances (5%), Lighting (4.5%), and water heating (0.5%) together contribute the balance. The results suggest that residential DR could provide between 35% and 60% of residential load's contribution to wholesale ancillary services requirements and up to 80% of its respective market value. This could be a much needed, currently untapped, supply of ancillary services to offset the decrease in supply due to thermal generation retirements and expected increases in demand due to renewable energy. Additionally, the results suggest that hourly caps are not an economically efficient means of addressing consumer response fatigue as compared to incorporating that *ex ante* risk into the commitment of ancillary services capacity on DR resources. Future research to better quantify how much flexibility load can provide; the adoption of enabling technologies; and the behavioral economics that affect how much load will customers be willing to turn over to direct load control could further refine the model in this paper. The results presented here suggest that the impact of residential DR is great enough to justify further research into these policies.

Those who are concerned with mitigating the costs of renewables integration and ancillary services costs should consider policies that decrease the demand for and/or increase the supply of ancillary services. Balancing area consolidation and residential demand response both create the desired reduction in ancillary services cost. This thesis aims to inform the policy discussion as to the details of each of these technologies. In the case of balancing area consolidation, one should be careful to measure all feasibly measured benefits and costs, including emissions related health effects prior to judging the merits of a specific consolidation. Residential demand response appears to be a great supply of ancillary services if one can advance the acceptance of direct load-control and correctly manage the risk of consumer fatigue.

The challenge of renewable integration and the expected scarcity of ancillary services is a solvable challenge. We hope that this thesis has helped to inform the discussion forward on this topic and contributes to the U.S. meeting its energy policy goals.

Chapter 5: Appendix

A.1 Near-term Effects of Wide-Scale Balancing Area Consolidation on Frequency Regulation Markets

A.1.1 MISO Balancing Areas Modeled

Table 5.1. MISO Balancing Areas

Power control area name	Missing Data	Modeled	Pre-consolidated with another BA
Alliant - East		X	
Alliant - West		X	
Ameren -Illinois		X	
Ameren - Missouri		X	Ameren - Illinois
Big Rivers Electric Corp.	EWD		
Cleco Corp.	EWD & Load		
Columbia MO City of		X	Ameren – Illinois
Consumers Energy Co.		X	
Dairyland Power Coop.		X	
Detroit Edison Co.	EWD		
Entergy	EWD & Load		
Great River Energy		X	
Hoosier Energy REC		X	
Indianapolis Power & Light		X	
Lafayette, city of	EWD & Load		
Madison Gas and Electric Co.		X	Alliant - East
Michigan Electric Coordinated Systems	Load		
MidAmerican Energy Co.		X	
Minnesota Power		X	Great River Energy
Muscatine Power and Water	Load		
Northern Indiana Public Service		X	
Northern States Power		X	Dairyland Power Coop.
Otter Tail Power Company		X	
South Mississippi Elect. Pow.	EWD & Load		
Southern Illinois Power Cooperative		X	
Southern Indiana Gas & Electric Co.		X	S. Illinois Power Coop.
Southern Minnesota Mun. Power Agcy.		X	Great River Energy
Springfield IL - CWLP		X	Ameren – Ill.
Upper Peninsula Power Co.		X	
Wisconsin Energy Corp.		X	
Wisconsin Public Service Corp.		X	

A.1.2 Variance Based Frequency Regulation Requirement Heuristic (1σΔ)

The model uses two heuristics to estimate the frequency regulation requirements of a balancing area (BA), both of which are derived from the WWSIS study [1]. Both heuristics are described in the main paper but this section includes a short derivation of the second heuristic. This second heuristic is a function of the variance of net-load and therefore will capture the smoothing effects of geographic diversity.

First, the step-changes of ten-minute net-load (Δnl_{10}) are divided into its components (Equations 5.1, 5.2, 5.3) and rearranged and regrouped (5.4 and 5.5). The result is that the variance of the step-changes in net-load, $Var(\Delta nl_{10})$, are equal to the variance of the sum of the step-changes of load and wind, $Var(\Delta L_{10} - \Delta W_{10}^{T})$ (Equation 5.5). Equation 5.6 follows based on the mathematical properties of variance.

$$Var(\Delta nl_{10}) = Var(\Delta nl_{10}) \qquad \textbf{5.1}$$

$$Var(\Delta nl_{10}) = Var\left(nl_{10,t+1} - nl_{10,t}\right) \qquad \textbf{5.2}$$

$$Var(\Delta nl_{10}) = Var\left(\left(L_{10,t+1} - W_{10,t+1}^{T}\right) - \left(L_{10,t} - W_{10,t}^{T}\right)\right) \qquad \textbf{5.3}$$

$$Var(\Delta nl_{10}) = Var\left(\left(L_{10,t+1} - L_{10,t}\right) - \left(W_{10,t+1}^{T} - W_{10,t}^{T}\right)\right) \qquad \textbf{5.4}$$

$$Var(\Delta nl_{10}) = Var(\Delta L_{10} - \Delta W_{10}^{T}) \qquad \textbf{5.5}$$

$$Var(\Delta nl_{10}) = Var[\Delta L_{10}] + Var[\Delta W_{10}^{T}] - 2 * Cov[\Delta L_{10}, \Delta W_{10}^{T}] \qquad \textbf{5.6}$$

Assume that the covariance of step-changes in load and wind is equal to zero ($Cov[\Delta l_{10}, \Delta w_{10}] = 0$). This assumption represents an upper bound estimate for the variance of ten-minute step-changes in net-load, as any covariance would reduce the variance of the ten-minute step-changes (5.7). The equality in 5.7 follows from this assumed upper bound. The model uses this upper bound as our estimate of the variance of step-changes in net-load.

$$Var(\Delta nl_{10}) \leq Var[\Delta L_{10}] + Var[\Delta W_{10}^{T}] \qquad \textbf{5.7}$$

$$Var(\Delta nl_{10}) = Var[\Delta L_{10}] + Var[\Delta W_{10}^{T}] \qquad \textbf{5.8}$$

The model estimates that the frequency regulation requirement heuristic should be equal to 1σΔ of 10-minute net-load (Equations 5.9) which can be transformed into Equation 5.10 and substituted with Equation 5.8 to create Equation 5.11.

$$Reg_{Req} = 1\sigma(\Delta nl_{10}) \qquad \textbf{5.9}$$

$$Reg_{Req} = Var(\Delta nl_{10})^{\frac{1}{2}} \qquad \textbf{5.10}$$

$$Reg_{Req} = (Var[\Delta L_{10}] + Var[\Delta W_{10}^{T}])^{\frac{1}{2}} \qquad \textbf{5.11}$$

WWSIS says that one standard deviation of ten-minute changes in load is approximately 1% of daily peak load (Equations 5.12 and 5.13). This approximation is substituted into 5.11, to get the final result, Equation 2.2.

$$\sigma[\Delta l_{10}] \cong 0.01 * \hat{L} \qquad \textbf{5.12}$$

$$Var[\Delta l_{10}] \cong \left(0.01 * \hat{L}\right)^{2} \qquad \textbf{5.13}$$

A.1.3 Economic Dispatch - Additional Details

The objective of the economic dispatch model is to minimize the costs of producing energy and providing frequency regulation to instantaneously match load. The body of the paper shows the most important details of the model. Below are additional details that may help the reader further understand the optimization problem.

The economic dispatch model used in this paper is an inter-temporal, co-optimized economic dispatch model with the following formulation:

Co-Optimized Economic Dispatch Formulation

$$\min_{e,fr,c} Energy\ Costs + Frequency\ Regulation\ Costs$$

Energy Constraints

$\sum_{i=1}^{n} e_{i,t} + \sum w_{i,t} - c_t = L_{j,t}$	Generation + wind - curtailment = load
$\sum_{i=1}^{n} r_{i,t} = FR\_Req$	Sum of FR capacity = FR requirement
$e_{i,t} \leq \overline{\overline{e_i}}$	Can't over schedule a resource for energy
$-e_{i,t} \leq 0$	Only positive energy production
$e_{i,t} \leq e_{i,t-1} + ramp_i$	Ramp-up limitation
$-e_{i,t} \leq -e_{i,t-1} + ramp_i$	Ramp-down limitation
$-e_{i,t} \leq -\underline{e_i}\ \forall i \in \{nuclear\}$	Must produce a minimum amount of energy*

*This constraint is only for nuclear generation

Frequency Regulation Constraints

$r_{i,t} \leq \overline{\overline{r_i}}$ Can't over schedule a resource for regulation

$-r_{i,t} \leq 0$ Only positive frequency regulation capacity

Coupling Constraints

$e_{i,t} + r_{i,t} \leq \overline{\overline{MW_i}}$ Can't over schedule a resource

$-e_{i,t} + r_{i,t} \leq 0$ Must produce energy to provide frequency regulation

A.1.3.1 Ramp Constraints.

The model imposes hourly ramp constraints only on coal and nuclear units. No hourly ramp constraint is placed on single cycle or combined cycle units as these generators can ramp over their entire capacity within one hour (Katzenstein and Apt 2009). For nuclear units, the hourly ramp limit is 1% of the unit's capacity per hour. This limit allows for some minimal amount of movement but encourages constant production. For coal units, the hourly ramp limit is 20% of its capacity per hour, which is consistent with the ramp limits used in previous literature (WECC 2009).

A.1.3.2 Minimum Generation Limits for Nuclear Plants.

Nuclear units are considered "must-run" in MISO and have a minimum generation limit. However, there is little reliable data on what this limit should be. Figure 5.1 shows the cumulative distribution of the annual capacity factors for nuclear plants (EPA 2012). The data show that the majority of the nuclear plants (~80%) have annual capacity factors over 80%, which we choose as the minimum generation limit for the nuclear power plants in this model.

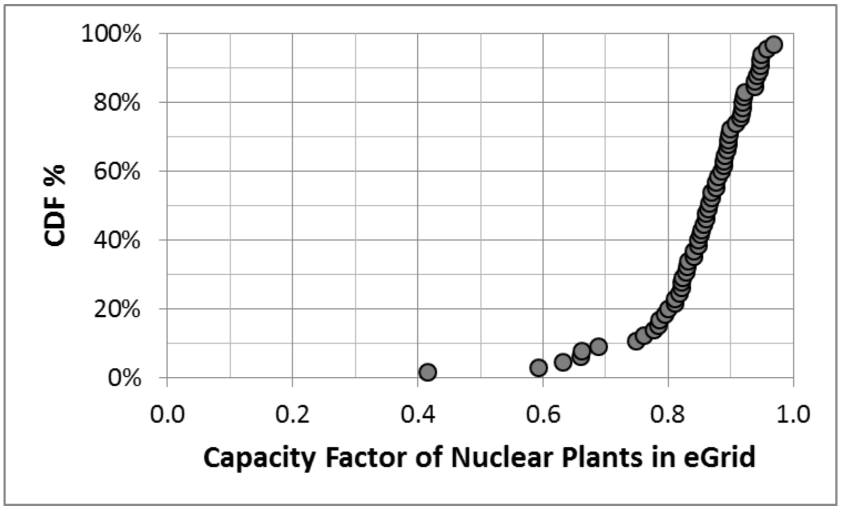

Figure 5.1. Approximate cumulative distribution function of nuclear capacity factors in the eGrid data set [2].

A.1.4 Frequency Regulation Details

Frequency regulation is an ancillary service that is at the center of this analysis. So the following sections provide clarifications regarding the methods of generating frequency regulation bids and the scheduling of frequency regulation capacity.

A.1.4.1 Frequency Regulation Bids.

We regress a year's worth of bids from the NYISO against the generator's size to estimate the average bid quantity (MW_{FR}) and the bid price ($\$/MW_{FR}$-Hr) of resources bidding into the frequency regulation market.

A linear regression does a good job of predicting the quantity (MW), or size of the frequency regulation bid, as shown in Figure 5.2 and Figure 5.3, with the exception of units between 200 and 300 MW. These units bid a higher percentage of their capacity into the frequency regulation market (14%). As a result of this analysis we use a heuristic in which units that are between 200-300 MW bid 14% of their capacity in the frequency regulation market. We further assume that all other units bid 6% of their capacity for frequency regulation. Figure 5.3 shows the results of this deterministic model, plotted with the actual bids.

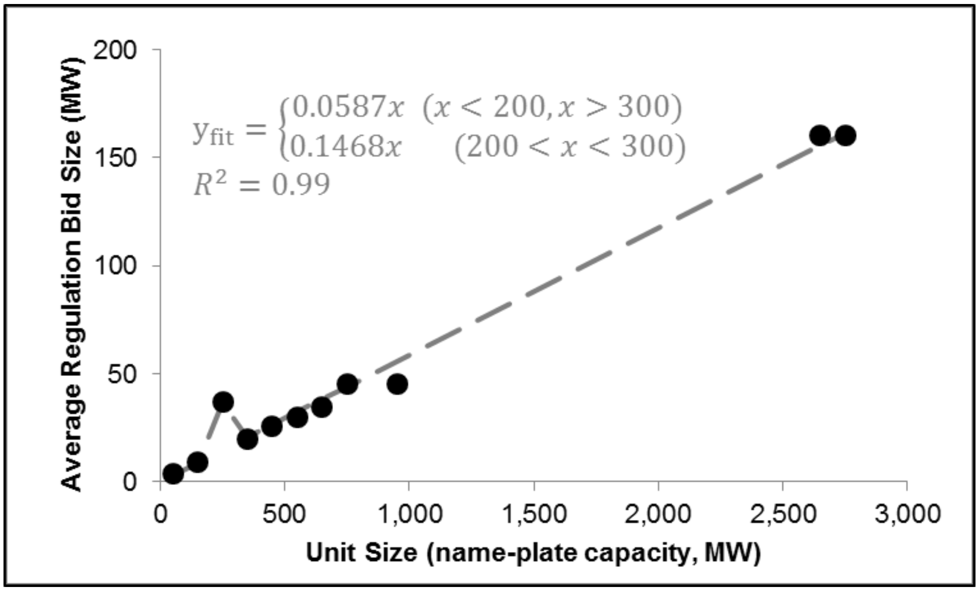

Figure 5.2. The average bid quantity (MW) into the frequency regulation market, by generator size. Based on 2009 NYISO bid data

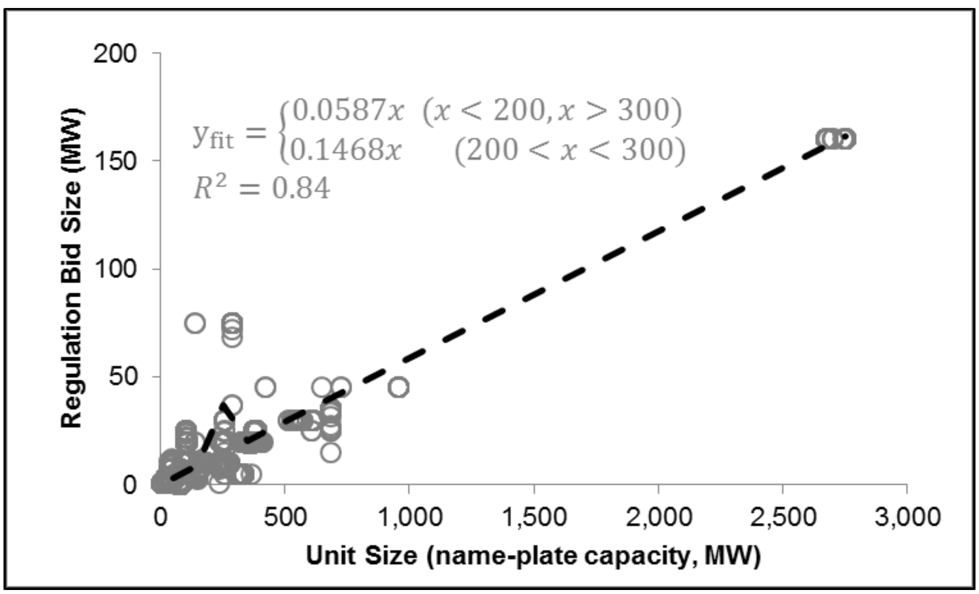

Figure 5.3. The bid quantity (MW) into the frequency regulation market plotted with our fitted model using the unit's capacity as the explanatory variable

Historic bid data from the NYISO are again used to estimate a bid price ($\$/MW_{FR}$-Hr) to units that are bidding into the frequency regulation market [3]. The average price for each bin is used to assign a bid price to each selected generator (Figure 5.4).

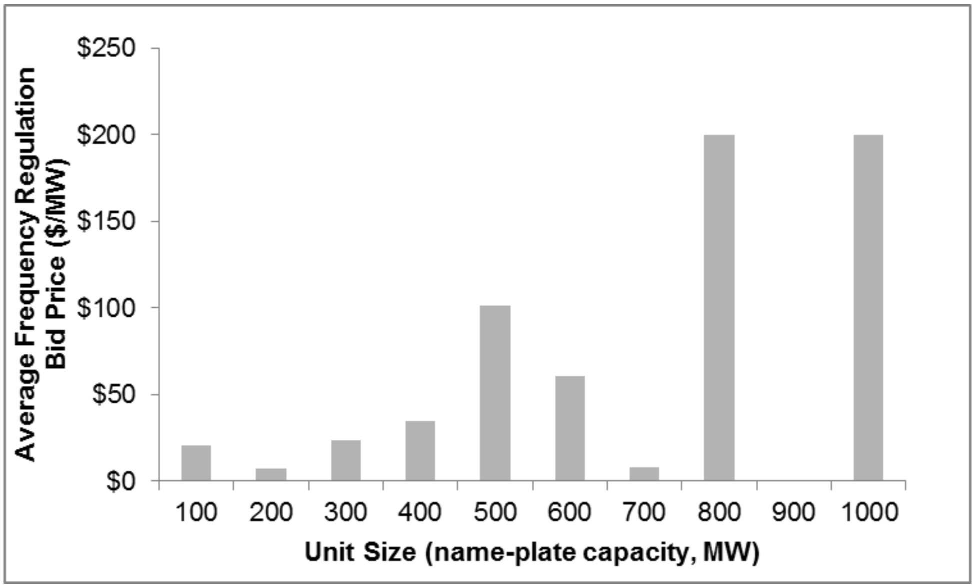

Figure 5.4. The average Bid Price ($/MW) into the frequency regulation Market, by generator size. Based on 2009 NYISO bid data [3].

It is often said that single cycle natural gas turbines ‘should’, or ‘are better equipped’, or ‘are likely to provide’ frequency regulation. However, no reliable source could be found to limit the set of generators providing frequency regulation to a specific type. Some limited evidence exists that coal units do provide frequency regulation, although Kirby [4] criticizes one plant for doing so poorly. Additionally Kirby et al [5] provide evidence that nuclear plants tend not to provide frequency regulation. As they put it, “Nuclear power plants choose not to participate in AGC because of the philosophy that reactor power is to be controlled only by the nuclear plant operator, and not by outside variables.” The model in this paper assumes that nuclear plants and wind farms do not bid into the frequency regulation market. All generators, other than nuclear and wind generators, bid into the frequency regulation market and are assigned a frequency regulation bid quantity and price based on its size.

A.1.4.2 Frequency Regulation Capacity.

Frequency regulation capacity is an amount of capacity, measured in megawatts (MW), that is held back from providing energy in order to counteract imbalances. These imbalances are the result of unexpected changes in imports, exports, generation, and load. Therefore frequency regulation capacity requirements are set ahead of time (*ex ante*) in order to reserve capacity for this balancing service. The amount of capacity is based on the historic distribution of imbalances, historic performance on NERC balancing standards (CPS1 & CPS2), and common heuristics (*i.e.*, 1% of peak load). The *ex ante* nature of reserves means that we do not need to model the sub-hourly dispatch of reserves in order to measure the economic effects of BA consolidation as long as the reserve requirements are sufficient. All economic effects are determined *ex ante*.

A.1.5 Additional Results

These sections highlight additional model results that may be of interest, including the alternative methods of wind farm allocation (A.1.5.1) and the model's resulting fuel mix (A.1.5.2).

A.1.5.1 Alternative Methods of Wind Farm Allocation

In section 2.2, we present a method for allocating hypothetical wind farms to BA's based on the capacity factor and location of each wind farm. The method presented in the main text ensures that each individual BA has approximately 20% wind, by energy. We show in section 2.4 of the manuscript, that consolidating all these BAs into one consolidated BA results in a reduction in the coefficient of variation of net-load – one method of quantifying the smoothing effect of consolidating geographically diverse wind farms. In this Supporting Information, we also consider another method of allocating wind farms to BA's. Here, we only use the wind farms with the highest capacity factors. This is to represent the fact that these wind farms are likely to be the most economically attractive and therefore are the most likely to be built. Figure 5.5 shows a plot of the coefficient of variation of net-load for both methods: the evenly dispersed 20% (in circles) and the highest capacity factor method (diamonds). Both methods result in the same downward trend that is characteristic of geographic diversity. However, the two methods are not substantially different. Both have similar effects on net-load, frequency regulation requirements, and therefore similar economic effects.

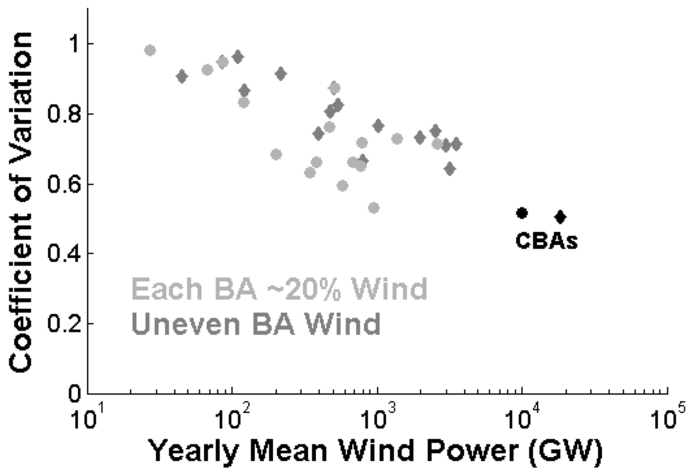

Figure 5.5. **The coefficient of variation of wind power versus mean wind power. The relative variability, as represented by the coefficient of variation, decreases with increased wind power. The circles represent the evenly distributed 20% wind allocation method presented in section 2.3 of the paper. The diamonds represent a method of wind allocation where only the wind farms with the highest capacity factors are included in the model. Therefore, there is significant variation in the wind penetration between regions.**

A.1.5.2 Fuel Mix

The BAs that we modeled have a lot of coal-fired generation, so it is not a surprise that coal dominates the fuel mix with, or without wind (see Tables 5.2 and 5.3).

Table 5.2. Fuel mix before and after consolidation for all scenarios without wind

Scenario	Without Wind Fuel Mix Pre- (Post-) Consolidation				
	Coal	**Gas**	**Nuclear**	**Oil**	**Biomass**
Low Gas	**80%** **(82%)**	**3%** **(1%)**	**17%** **(17%)**	**0%** **(0%)**	**0%** **(0%)**
Base Case	**81%** **(82%)**	**2%** **(0%)**	**17%** **(17%)**	**0%** **(0%)**	**0%** **(0%)**
High Gas	**82%** **(82%)**	**1%** **(0%)**	**17%** **(17%)**	**0%** **(0%)**	**0%** **(0%)**
High Gas & Frequency Regulation	**82%** **(82%)**	**1%** **(0%)**	**17%** **(17%)**	**0%** **(0%)**	**0%** **(0%)**

Table 5.3. Fuel mix before and after consolidation for all scenarios with ~20% wind

Scenario	With Wind (~20%) Fuel Mix Pre- (Post-) Consolidation					
	Coal	Gas	Nuclear	Oil	Biomass	Wind
Low Gas	59% (58%)	2% (1%)	16% (17%)	0% (0%)	0% (0%)	24% (24%)
Base Case	59% (59%)	1% (0%)	16% (17%)	0% (0%)	0% (0%)	24% (24%)
High Gas	60% (59%)	0% (0%)	16% (17%)	0% (0%)	0% (0%)	24% (24%)
High Gas & Frequency Regulation	60% (59%)	0% (0%)	16% (17%)	0% (0%)	0% (0%)	24% (24%)

A.1.6 References

1. GE-Energy (2010) Western wind and solar integration study. National Renewable Energy Laboratory (NREL), Golden, CO.,

2. EPA (2012) eGRID2012 Version 1.0. http://www.epa.gov/cleanenergy/energy-resources/egrid/index.html

3. NYISO (2009) NYISO Bid Data: 2009. http://mis.nyiso.com/public/P-27list.htm

4. Kirby B (2007) Ancillary services: Technical and commercial insights.

5. Kirby B, Kueck J, Leake H, Muhlheim M Nuclear generating stations and transmission grid reliability. In: Power Symposium, 2007. NAPS'07. 39th North American, 2007. IEEE, pp 279-287

A.2 RESIDENTIAL LOAD PROVIDING ANCILLARY SERVICES

A.2.1 SOURCES FOR FLEXIBILITY PARAMETER ESTIMATES

Table 5.4 Sources for the flexibility parameter for DR providing AS

Year/ Author	End-Use(s)	Ancillary Service	Min	$\mathbb{E}[]$	Max	Notes
2003 Kirby [1]	Pumps	Spinning Reserve	0.00	0.50	1.00	This suggest that pumps might be similar to a delayable appliance
2007 Eto [2]	Cooling	Spinning Reserve	0.00	0.15	0.38	Actual Demonstration. Customers didn't notice a 20 minute curtailment. That means 3 households could be sequentially curtailed sequentially to make one full hour of curtailment.
2007 Hammerstrom [3]	Water Heating	Spin	0.00	-	1.00	Demonstrates the ability to respond to under-frequency load sheds
2007 Hammerstrom [3]	Clothes Dryers	Spin	0.00	-	1.00	Demonstrates the ability to respond to under-frequency loadsheds
2007 Stadler [4]	Refrigeration	Load Following	0.10	0.25	0.50	Table 2shows how long a load reduction can be sustained as a function of two variables: 1) the spread parameter and 2) the percentage load reduction achieved. Looking at these data backwards (from the values up to the columns), one can infer what services could be provided at a given reduction. E.g., 10% reduction can be held for nearly 2 hours; 25% for approximately 1.5 hrs; 50% for about an hour; ...For load following you need about an hour...
2007 Stadler [4]	Refrigeration	Spin	0.25	0.50	0.75	Based Table 2 from the 2007 Stadler Load Following
2007 Stadler [4]	Refrigeration	Frequency Regulation	0.38	0.41	0.45	Based on Table 2 from the 2007 Stadler Load Following.

Year/ Author	End-Use(s)	Ancillary Service	Min	$\mathbb{E}[]$	Max	Notes
2009 Callaway [5]	Cooling	Frequency Regulation / Load Following	0.00	0.05	-	The paper is highly theoretical but uses 14kW average consumption TCLs and shows a response of "roughly 0.5 kWh and 0.75 kW of responsiveness per load". Though this wasn't maximum capable, it was maximum needed to respond to a wind turbine error signal.
2011 Kiliccote [6]	HVAC	Non-Spin	0.00	0.22	-	
2011 Hovgaard [7]	Refrigeration	Frequency Regulation	0.00	0.20	-	Really "Primary Response" – Does not include a marginal willingness to pay for electricity. 20% based on a statement in the conclusion.
2011 Keep [8]	Heat	Frequency Regulation & Response	0.00	0.05	0.35	
2011 Keep [8]	Heat	Frequency Regulation & Response	0.00	0.05	0.35	
2012 Angeli [9]	Refrigeration	Frequency Response	0.00	0.64	0.98	Droop characteristic for these are likely between 0.4-0.6%. The flexibility parameter is miss-leading because each 'fridge is providing very little response – not optimized.
2012 Huesan [10]	Lighting	Spin	0.00	0.23	-	This study does consider "noticable" changes and avoid them
2012 Huesan [10]	Lighting	load shifting	0.00	0.23	-	This does consider "noticable" changes and avoid them
2012 Perfumo [11]	Cooling	Spin & Non-Spin	0.00	0.34	-	Incorporates a "comfort" metric. Spin chosen based on signal and duration
2013 Aunedi [12]	Refrigeration	Frequency Response	0.00	0.01	-	Not optimized, the loads are clearly not being taxed fully. Very mininimal amount of response from each (<0.01% curtailment).

Year/ Author	End-Use(s)	Ancillary Service	Min	$\mathbb{E}[]$	Max	Notes
2013 Halamay [13]	**Water Heating**	**Frequency Regulation**	**0.00**	**0.05**	**0.08**	**Does the balancing of a single wind farm.**
2013 Halamay [13]	**Water Heating**	**Load Shifting**	**0.00**	**0.00**	**0.06**	**Does the balancing of a single wind farm.**
2013 Hao [14]	**Cooling**	**Frequency Regulation**	**0.00**	**0.36**	**-**	**Used the PJM RegD (aka Fast) signal. This is not representative of all frequency regulation signals.**
2013 Harsha [15]	**Cooling**	**Peak Shaving**	**0.02**	**0.02**	**0.03**	
2013 Lu [16]	**HVAC**	**Frequency Regulation**	**0.00**	**0.17**	**-**	
2013 Sullivan [17]	**Cooling**	**Spinning Reserve**	**0.00**	**0.68**	**0.80**	**Customers who experiences > 60 curtailments reported noticing ~2.79 events. Customers who only experienced one curtailment reported noticing 3.23 events.**
2013 Zhang [18]	**Cooling**	**Frequency Regulation & Response**	**0.00**	**0.13**	**-**	**Used the PJM RegD (aka Fast) signal. This is not representative of all frequency regulation signals.**
2013 Zhang [18]	**Cooling**	**Peak Shaving**	**0.00**	**0.13**	**-**	**Assumes the same "flexibility" as when used for frequency regualtion - 2.5MW reduction from the same number of ACs**
2013 Zhang [18]	**Cooling**	**Load Following**	**0.00**	**0.17**	**-**	
2013 Zhao [19]	**HVAC**	**Frequency Regulation**	**0.00**	**0.20**	**0.35**	**Used the PJM RegD (aka Fast) signal. This is not representative of all frequency regulation signals.**

A.2.2 Statistical Ancillary Service Deployment Model

These statistical models were based on ERCOT data [20].

A.2.3 Deployment of Frequency Regulation Reserves

Frequency regulation resources are constantly on call to move up and down in response to real-time imbalances of the grid, but the amount dispatched varies throughout time. We develop a statistical model that approximates the correlated distributions of dispatch energy $(\delta_{r,t})$ associated with frequency regulation up and down. Up regulation is dispatched when the frequency is low and down regulation is dispatched when frequency is high; given this, it makes sense that the dispatch of these two reserves are negatively correlated. We approximated the correlated regulation deployment energy values as bivariate exponentially distributed random variables with a correlation coefficient of -0.5. The up regulation deployment is exponentially distributed with expectation of 0.16 kWh/kW $\left(\delta_{r,t} \sim Exp(0.16)\right)$; and the down regulation deployment is exponentially distributed with an expectation of 0.12 kWh/kW $\left(\delta_{r,t} \sim Exp(0.12)\right)$.

A.2.4 Deployment of Contingency Reserves

This paper defines contingency reserves as frequency response, spinning, and non-spinning reserves. Although this is not the formal NERC definition [21], all of these reserves respond in the event of a contingency. For this paper, we model the three reserves (frequency response, spin, and non-spin) as a cascading Bernoulli trial. A probability (p_1) suggests that an under-frequency event occurs, which triggers frequency response providers to respond. There is also a conditional probability, p_2, that the low frequency event cascades to the next reserve – spin. Alternatively, there is a probability $(1 - p_2)$ that the initial event is resolved without needing to call spinning reserves. If spinning reserves are called, however, there is a probability (p_3) that non-spin will be called and a

$(1 - p_3)$ chance that the event resolves without calling non-spin. This process can be repeated hundreds of time to develop a set of possible contingency calls that are used to calculate the risk of DR violating its charging requirement (section 3.2.7)

Based on the ERCOT data [20], frequency response is deployed in approximately 90% of all hours. Therefore p_1 in our model is 0.9. The amount of frequency response deployment, given a call, can be modeled as a Weibull distribution (with fit parameters A = 31.665 and B = 1.563 using ERCOT data). With regards to spinning reserves, it is likely that they are called between 20 and 200 times a year [22,23]. We assume a worst case and use 200 spin calls a year. Lastly, data from an ERCOT presentation [23] suggest that non-spin is called approximately 80 times a year. However this value is for a single year and there is no indication that this is a typical year for ERCOT. Therefore, we assume an order of magnitude estimate of 100 non-spin calls a year. The resulting conditional probabilities for spin and non-spin are: $p_2 = 0.025$; $p_3 = 0.50$. The amount of deployment (δ) for both spin and non-spin are assumed to be equal to one. In other words, a DR resource providing one kW of spinning, or non-spinning, reserve will have one kWh of deployment energy when called upon.

A.2.5 Deferred Energy Example

The amount of energy a specific load defers in order to provide ancillary services is a function of three factors: 1) the type of ancillary service; 2) the amount of that ancillary service capacity scheduled on load; and 3) the amount of deployment for that ancillary service. Figure 5.6 provides an example of how the amount of deferred energy differs depending on the type of ancillary service.

For example, let's assume that a load with maximum power of $\mathbf{L}_{\mathbf{max}}$ will naturally consume an hourly average of $\mathbf{L}_1$ kWh when it isn't providing any form of demand response. *When providing DR, any reduction in consumption from* $\mathbf{L}_1$ *will count as deferred energy*. If the load is scheduled to provide regulation up (see left panel in Figure 5.6), it first needs to reserve the ability to reduce its consumption from $\mathbf{L_1}$ to $\mathbf{L_3}$, but it will only need to lower its consumption from $\mathbf{L_1}$ when it is deployed. Let's now assume that it is deployed for regulation up to an hourly average of $\mathbf{L_2}$ kWh. Therefore, the amount of deferred energy is equal to the deployment energy multiplied by one hour: $\mathbf{L_1 – L_2}$.

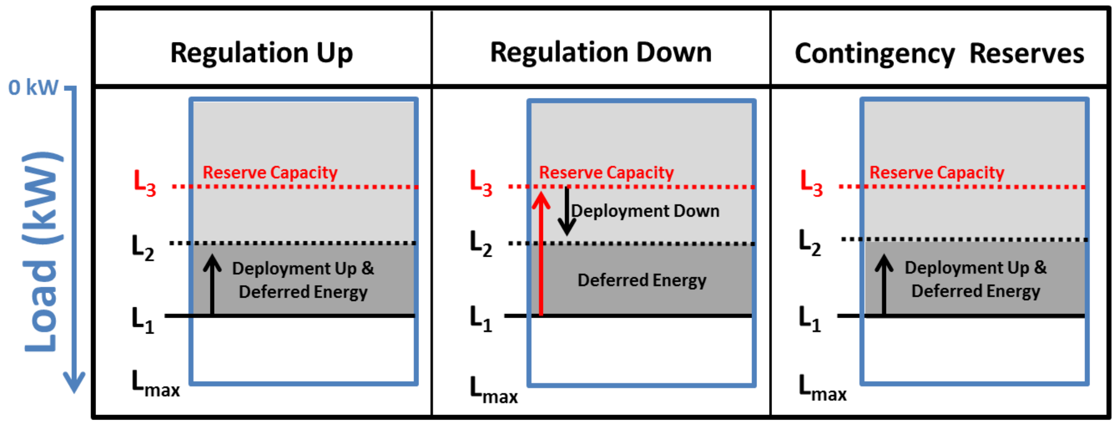

Figure 5.6. Depiction of how AS deployment and the amount of deferred energy varies with different AS types. (from grid perspective)

The important part here is that in order to reserve the regulation up capacity, the load does not need to change consumption (from its counterfactual consumption – $\mathbf{L_1}$) unless it is deployed up. The same is true for providing contingency reserves. However, in order to reserve regulation down capacity (center panel in Figure 5.6), a load needs to defer from consuming energy, from $\mathbf{L_1}$ to $\mathbf{L_3}$, in order to reserve the ability to provide regulation down. This is a key difference for only regulation down. Given no deployments, the deferred energy is equal to the regulation down capacity times one hour, ($\mathbf{L_1 – L_3}$)*1 kWh. The deferred energy for a deployed load providing regulation down is the amount of capacity of regulation down (times one hour) minus the hourly average amount of deployment down: ($\mathbf{L_1 – L_3}$)*1 kWh minus ($\mathbf{L_2 – L_3}$) kWh. This is equivalent to $\mathbf{L_1 – L_2}$.

A.2.6 Independent and Correlated Draws for Monte Carlo Simulations

One question that arose during this analysis regards how different flexibility parameters $\left(f_t^{i,r}\right)$ are correlated. The possibilities are bound between totally independent parameters to perfectly correlated parameters. This analysis tests three cases to see if the level of correlation has any significant effect on the results. The three cases are i) totally independent, ii) correlated by load, and iii) correlated by load and ancillary service.

Completely independent draws mean that a realization of one flexibility parameter does not affect the realization of a draw from any of the other distributions of flexibility parameters. Another possibility is that the parameters could be correlated. For example, if someone finds an approach to make the energy consumption of refrigerators very flexible and able to provide a large amount of frequency regulation, then the same technology will likely make refrigerators more able to provide spinning reserve. The correlated-by-load case represents this scenario and does correlated draws for each set of flexibility parameters in each load type.

Similar logic applies to the correlated by load and ancillary service case. If a researcher finds the means to extract lots of flexibility from refrigerators to provide frequency regulation, then the same method may apply to other thermostatically controlled loads as well. Therefore, there should be a correlation between the flexibility parameters across load types and ancillary services. Another way of thinking about this case is that there is a single draw that represents the 'state of technology' for aggregated loads providing ancillary services. If this draw is low, then the 'state of technology' is less advanced and all the flexibility parameters will reflect this.

Figure 5.7 shows how much of each ancillary service requirement is optimally scheduled on residential DR resources, in the form of duration curves, for each of the methods of drawing flexibility parameters. The correlated by load and ancillary service (in red) shows the greatest variation in the scheduling of ancillary services. The independent and correlated-by-load cases have less variation.

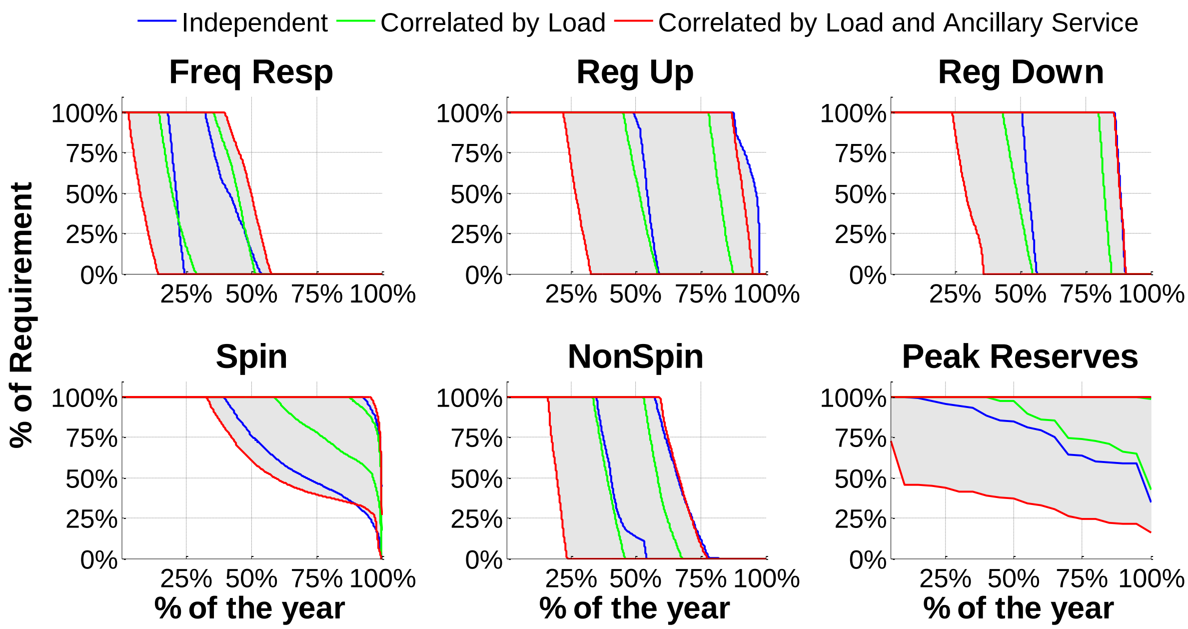

Figure 5.7. Duration curves of the percent of ancillary service requirements optimally scheduled on residential DR resources, each representing a method of drawing flexibility parameters. The correlated by load and ancillary service case (red) has the most variation. The shaded area represents the total span of all cases of drawing flexibility parameters. The results shown in Figure 3.5 use the correlated by load assumption.

The variation in scheduling, shown in Figure 5.7, also exists in the results for the percentage of hourly market value. Figure 5.8 shows the percentage of hourly market value of the Monte Carlo simulations for each draw method. Again, the correlated by load and ancillary service case has the most variation and a few pessimistic draws (bottom left of Figure 5.8) for which the results fall significantly far from those presented in the main paper. Despite this variation, the overall methods and conclusions presented in the paper are not changed by the different draw methods.

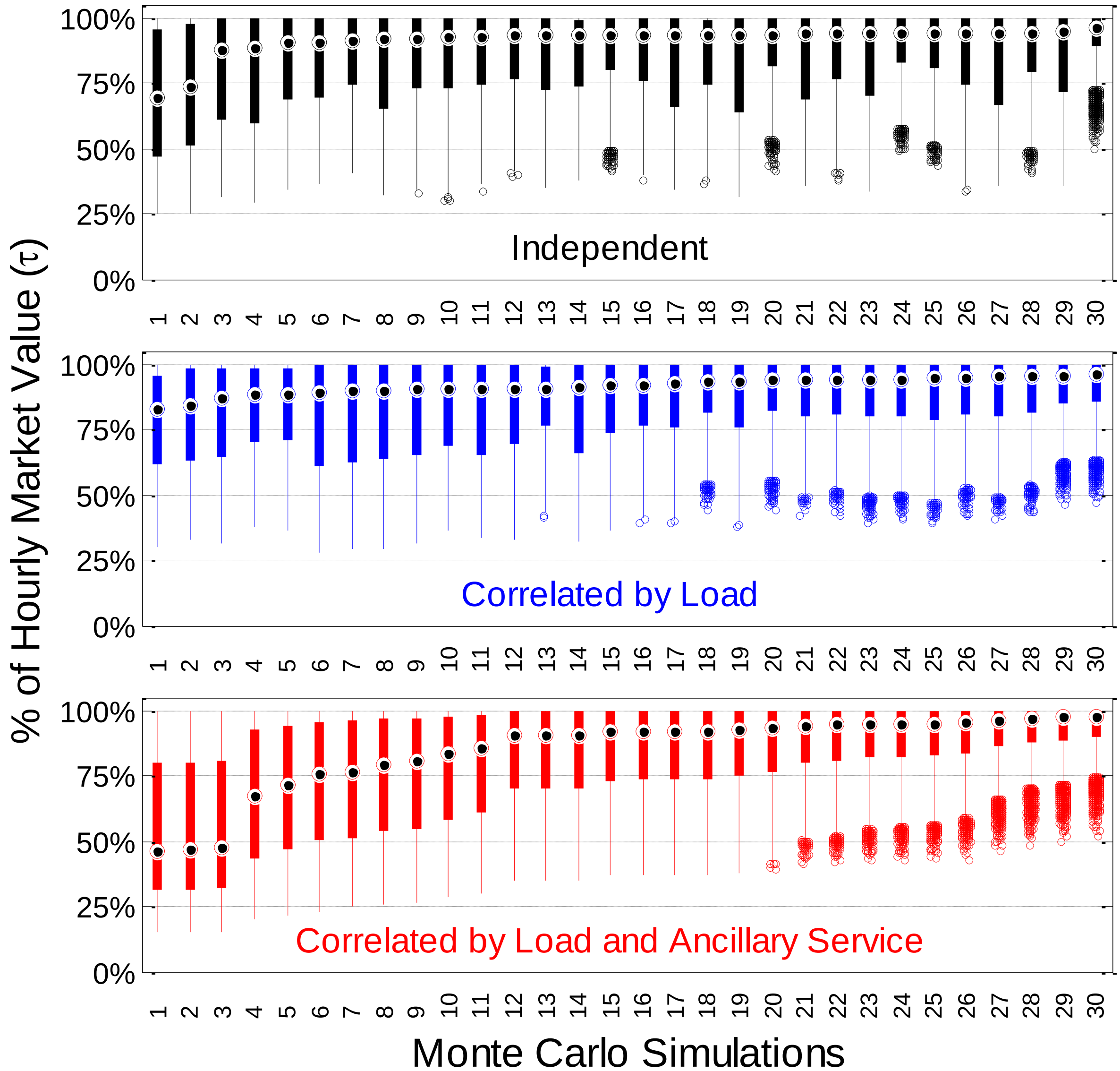

Figure 5.8. Percentage of hourly market value for three different methods of drawing flexibility parameters: independently drawn (black, top), correlated by load (middle, blue), and correlated by load and ancillary service (bottom, red).

A.2.7 Penalty Function

The penalty function in this analysis assumes consumers care more when energy they want does not get delivered. Therefore the penalty function places a cost when loads under-charge, meaning that the energy deferred for ancillary service deployment is greater than the energy it pre-charged or could make up post-deployment. There is a ten percent tolerance to allow for some under-charging to occur (Figure 5.9). After this tolerance, a cost equal to the maximum energy price for that day is incurred for every deficient kWh.

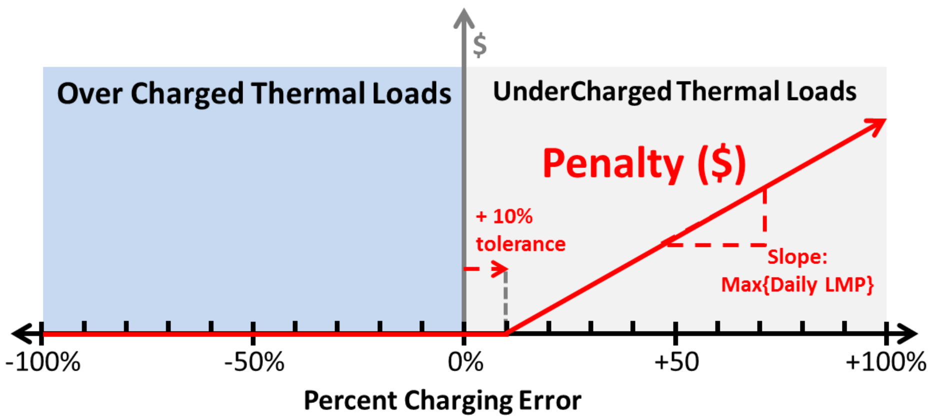

Figure 5.9. Depiction of the penalty function with respect to the percent charging error

If the percent charging error is less than zero we do not assign a penalty. This would occur if a thermal load pre-charged (*ex anti*) more than necessary given its actual deferred energy (*ex post*). As an example, imagine an aggregation of air conditioning units is scheduled to charge for an expectation of one kilowatt-hour of deferred energy but in real-time it is not deployed for ancillary services and therefore has no deferred energy. In this case, it is likely that the risk for customer response fatigue is low or non-existent. Primarily this is because the dynamic controls of the aggregation may sense the over-charging and direct the aggregation to reduce consumption, thereby eliminating the negative charging error. If over-charging still exists, it is possible that the customer will be happy at the slightly cooler temperature. Extremely negative values of charging error are unlikely because the optimization already discourages this in the objective function. The outcome of introducing such penalty function is a cost (in dollars) for each load type, for each of the one hundred realizations of ancillary services deployment, for that particular day of the year.

However, the methods are not predicated on this particular penalty function. Given better understanding of consumer preferences a more accurate penalty function could be incorporated. Given that disclaimer, below is a depiction of the penalty function used in

A.2.8 Penalty Price Sensitivity

Error! Reference source not found. 5.10 shows how different hourly caps affect DR's ability to provide market value and compares this to the value DR would provide when the risk of response fatigue is included in the ancillary services capacity assignment. Note that adding the penalty price that represents the risk of response fatigue only slightly reduces the quantity of ancillary services capacity assigned to DR resources. However, a fixed hourly cap significantly reduces the value that DR can provide through ancillary services.

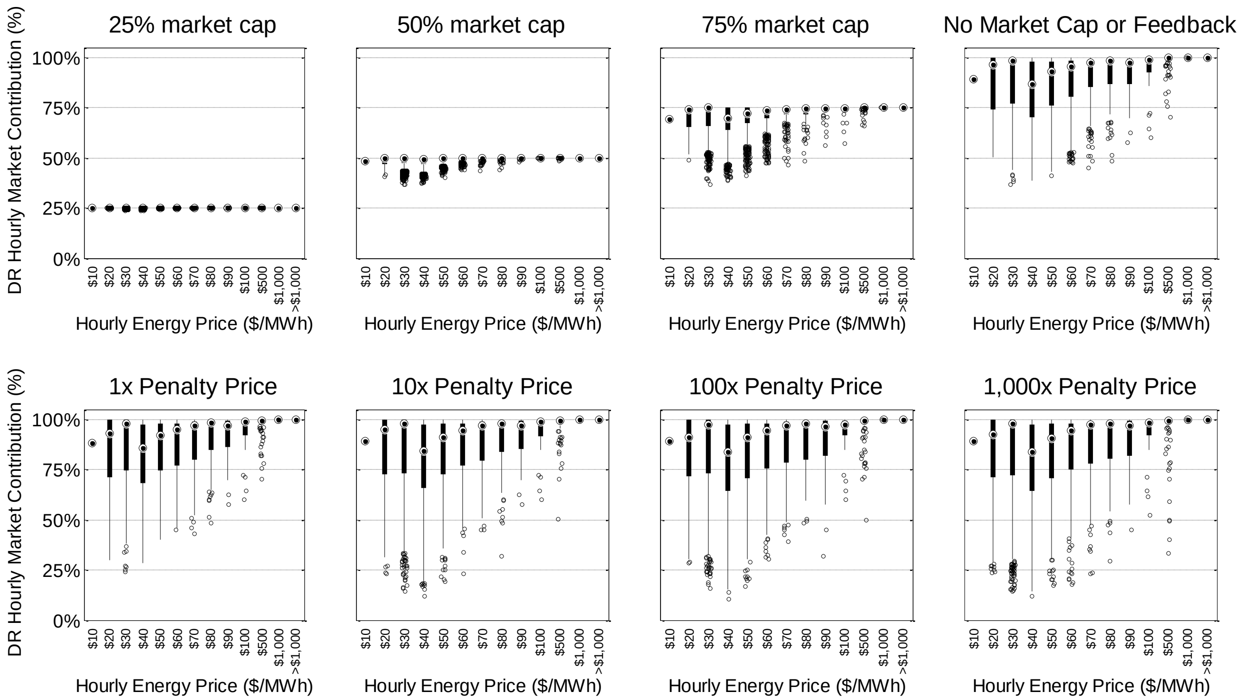

Figure 5.10. This plot shows how much (%) of the market value of all the ancillary services capacity being scheduled on residential DR using different coarse market caps on DR penetration. The solid lines do not include the risk-based feedback in optimization; the circles are results using the feedback.

A.2.9 Refrigerator Model

Missing or invalid refrigerator data from our sample is simulated using an exponential refrigerator model [5,24-26]. The model calculates the next average refrigerator temperature (θ_{t+1}) as the sum of the increase in temperature due to heat transfer, the decrease in temperature due to the refrigerator's heat rejection, and a random noise parameter (w_t), the three terms in in Equation 5.14, respectively. The increase temperature due to heat transfer from the surrounding room is modeled as an exponential function of the time step (h), the thermal capacity (C), and thermal resistance (R) (Equation 5.15). The heat rejection (second term in Equation 5.14) is dominated by the state of the compressor motor (m), either on or off, which is determined by a dead-band control described by Equation 5.16.

$$\theta_{t+1} = \alpha\theta_t + (1-\alpha)\left(\theta_t^{ambient} - m * R * P\right) + w_t \qquad \textbf{5.14}$$

$$\alpha = \exp\left(-\frac{h}{CR}\right) \qquad \textbf{5.15}$$

$$m_{t+1} = \begin{cases} 0, & \theta_t < \delta_{low} \\ 1, & \theta_t > \delta_{high} \\ m_t, & otherwise \end{cases} \qquad \textbf{5.16}$$

Previous models draw random values from distributions of refrigerator properties – including compressor motor size (kW), thermal capacitance (kWh/°C), and thermal resistance (°C/kW). One issue that can arise from this method is that a random refrigerator could draw a very low motor size while also drawing a very low thermal resistance. This leads to a situation where the refrigerator is absorbing more heat from the ambient surroundings than is being rejected from the refrigerator's cooling coil, even with the motor always on. We address this issue by correlating the refrigerators' power and thermal resistance values. We also tuned all of the parameters to match the average refrigerator in our sample. Table 5.5 provides parameters used in this paper and other papers.

Table 5.5. Refrigerator model parameters

Modeling Parameter	Mathieu [26]	Kara & Berges [25]	Tuned Range
Temperature Setpoint (*C)	**1.7-3.3**	**2.5**	**2**
Deadband Width (*C)	**1-2**	**1.5**	**1**
Power (kW)	$\sim U(x)$ $0.1 \leq x \leq 0.5$	$\sim U(x)$ $0.4 \leq x \leq 0.8$	$\sim LN(\mu, \sigma)^*$ $\mu = 0.163, \sigma = 1.08$
Thermal Resistance (*C/kW)	$\sim U(x)$ $80 \leq x \leq 100$	$\sim U(x)$ $80 \leq x \leq 100$	$\sim U(x)$ $200 \leq x \leq 220$
Thermal Capacitance (kWh/*C)	$\sim U(x)$ $0.4 \leq x \leq 0.8$	$\sim U(x)$ $0.2 \leq x \leq 1.0$	$\sim U(x)$ $0.4 \leq x \leq 0.8$

* Power and thermal resistance are modeled as correlated variables with a correlation of -0.8

A.2.10 References

1. Kirby BJ, Kueck JD (2003) Spinning Reserve from Pump Load: A Technical Findings Report to the California Department of Water Resources. United States. Department of Energy,

2. Eto JH, Nelson-Hoffman J, Torres C, Hirth S, Yinger B, Kueck J, Kirby B, Bernier C, Wright R, Barat A (2007) Demand response spinning reserve demonstration.

3. Hammerstrom DJ, Brous J, Chassin DP, Horst GR, Kajfasz R, Michie P, Oliver TV, Carlon TA, Eustis C, Jarvegren OM (2007) Pacific Northwest GridWise™ Testbed Demonstration Projects; Part II. Grid Friendly™ Appliance Project. Pacific Northwest National Laboratory (PNNL), Richland, WA (US),

4. Stadler M, Krause W, Sonnenschein M, Vogel U The Adaptive Fridge–Comparing different control schemes for enhancing load shifting of electricity demand. In: 21st Conference Informatics for Environmental Protection-Enviroinfo Warsaw, 2007. pp 199-206

5. Callaway DS (2009) Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. Energy Conversion and Management 50 (5):1389-1400

6. Kiliccote S, Piette MA, Koch E, Hennage D Utilizing Automated Demand Response in commercial buildings as non-spinning reserve product for ancillary services markets. In: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, 2011. IEEE, pp 4354-4360

7. Hovgaard TG, Larsen LFS, Jorgensen J Flexible and cost efficient power consumption using economic mpc a supermarket refrigeration benchmark. In: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, 2011. IEEE, pp 848-854

8. Keep TM, Sifuentes FE, Auslander DM, Callaway DS Using load switches to control aggregated electricity demand for load following and regulation. In: Power and Energy Society General Meeting, 2011 IEEE, 2011. IEEE, pp 1-7

9. Angeli D, Kountouriotis P-A (2012) A stochastic approach to "dynamic-demand" refrigerator control. Control Systems Technology, IEEE Transactions on 20 (3):581-592

10. Husen SA, Pandharipande A, Tolhuizen L, Wang Y, Zhao M Lighting systems control for demand response. In: Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES, 2012. IEEE, pp 1-6

11. Perfumo C, Kofman E, Braslavsky JH, Ward JK (2012) Load management: Model-based control of aggregate power for populations of thermostatically controlled loads. Energy Conversion and Management 55:36-48

12. Aunedi M, Kountouriotis P-A, Calderon JO, Angeli D, Strbac G (2013) Economic and Environmental Benefits of Dynamic Demand in Providing Frequency Regulation.

13. Halamay DA (2013) Development of a grid-scale energy storage solution via demand response using model predictive control of multiple residential water heaters.

14. Hao H, Sanandaji BM, Poolla K, Vincent TL A Generalized Battery Model of a Collection of Thermostatically Controlled Loads for Providing Ancillary Service. In: proceedings of the 51-th Annual Allerton Conference on Communication, Control and Computing, 2013.

15. Harsha P, Sharma M, Natarajan R, Ghosh S A Framework for the Analysis of Probabilistic Demand Response Schemes.

16. Lu N, Zhang Y (2013) Design Considerations of a Centralized Load Controller Using Thermostatically Controlled Appliances for Continuous Regulation Reserves.

17. Sullivan M, Bode J, Kellow B, Woehleke S, Eto J (2013) Using Residential AC Load Control in Grid Operations: PG&E's Ancillary Service Pilot.

18. Zhang W, Lian J, Chang C-Y, Kalsi K (2013) Aggregated Modeling and Control of Air Conditioning Loads for Demand Response.

19. Zhao P, Henze GP, Plamp S, Cushing VJ (2013) Evaluation of Commercial Building HVAC Systems as Frequency Regulation Providers. Energy and Buildings

20. ERCOT (2014) Fast Responding Regulation Service Pilot. ERCOT.

21. NERC (2013) Reliability Standards for the Bulk Electric System of North America. http://wwwnerccom/docs/standards/rs/Reliability_Standards_Complete_Setpdf

22. Kirby B (2007) Ancillary services: Technical and commercial insights.

23. Dumas J (2011) Historical Non-Spin Deployment Data. ERCOT, http://www.ercot.com/content/meetings/rdtf/keydocs/2011/0906/NS_history_and_floor_proposals.pdf

24. Koch S, Mathieu JL, Callaway DS Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services. In: Proc. PSCC, 2011. pp 1-7

25. Kara E, Bergés M, Hug G Modeling Thermostatically Controlled Loads to Engage Households in the Smart Grid: Lessons Learned from Residential Refrigeration Units. In: Computing in Civil and Building Engineering (2014). ASCE, pp 2032-2039

26. Mathieu JL (2013) Modeling, analysis, and control of demand response resources.