Quantifying the Benefits of Imperfect Demand Response

Patricia Levi Doctoral Candidate, Stanford University

Department of Management Science & Engineering
475 Via Ortega, Stanford, CA 94305
pjlevi@stanford.edu, (510) 604-0009

Abstract

Decarbonizing the electricity sector requires a massive shift away from fossil-fired generation, towards a diverse suite of newer technologies. Demand response (DR) is one of these promising technologies. It encompasses a range of techniques that adjust demand levels in response to system conditions. It could make the grid more flexible and reduce peaks in demand, potentially helping integrate renewables and operate the grid more efficiently. DR can have operating characteristics that differ from those of traditional generators. To realize the potential value of DR to a decarbonizing grid, we must understand how these unique properties affect its system-wide value. This study contributes to our understanding by characterizing the relative value of different possible properties of DR, so that market participants can focus their efforts on the most useful types of DR resources. We use a two-stage stochastic unit-commitment model, with ERCOT as our test system. Features examined include advance notification requirements, restrictions on when DR is available, the number of startups, the number of hours of operation, and the amount of energy shed. Results suggest that inexpensive DR that requires advance notification may still be quite valuable to the grid, and these limitations affect the value of DR less than other usage restrictions. Availability for early afternoon ramps and peaks is key for realizing reductions in system costs and ramp rates among thermal generation, and may be more important than the ability to respond to the real time market. The understanding gained from this study can guide the development of new DR products that provide higher system-wide value and better consumer satisfaction.

1 Introduction

There are great expectations for demand response (DR), which encompasses the idea of electricity demand reducing or shifting its load in response to incentives that are linked to wholesale market conditions. The greater flexibility in demand that it could provide would be very valuable on a low- or no-carbon grid. DR could help integrate intermittent renewables by making the grid more flexible, and it could help utilities meet their resource adequacy requirements, which will be increasingly difficult as dispatchable generation is replaced with variable renewables [15]. It could also help manage and limit new investment needed in distribution grids with a large number of small energy resources, like home solar panels, small storage units, and car chargers. However, there are many different ways of implementing DR, each with their own benefits and drawbacks that are not all well understood. We examine some of the options for DR in an attempt to quantify their relative benefits.

There are many types of DR. This study focuses on the types of DR that, at an aggregate level, are able to receive instructions from wholesale electricity markets regarding the amount of energy to shed. This kind of DR is defined by two key features. First, the incentives or instructions for load shedding are dynamic and able to respond to unexpected grid conditions like heat waves, reliability events, or high-priced hours. This feature is in contrast to 'static' programs like time-of-use pricing, in which a different rates are offered for on-peak and off-peak hours that do not change regardless of system conditions. Second, the DR resource must be able to provide a specified level of load reduction when called upon by some type of system operator. This arrangement is in contrast to dynamic pricing programs which expose consumers to a varying price signal and allow them react as they see fit. The combination of these two features creates a resource that can be counted on by utilities and system operators to help balance the grid.

We focus on this type of DR because it represents the form of DR that is able to bid in wholesale markets in the US. In 2011, FERCs Order 745 paved the way for the expansion of this type of DR by requiring that it be compensated in wholesale markets at equivalent rates to a generator. As a result, the volume of DR included in wholesale markets has been on the rise. Major market operators including CAISO, ISO-NE, PJM, and ERCOT have programs for generic DR resources to bid in their wholesale, ancillary services, and capacity markets where applicable.

Individual electricity customers that might provide demand response do not bid in these markets on their own. Typically, electricity customers providing demand response participate in wholesale markets via intermediaries, either third party aggregators (termed demand response providers), or utility programs that moderate the relationship between wholesale market and consumer. This intermediary aggregates a number of electricity customers together and bids their combined load-reducing capabilities into the wholesale energy markets. If that bid is won, the intermediary translates that award into an instruction sent to the customers to reduce their load. That instruction could take many forms, including automatic, remote control of a device or appliance, or an automated message requesting that a consumer take a certain action, or offering a certain incentive for reducing their electricity demand.

Many technologies and end-uses can provide demand response. Even after focusing within the category of dynamic, energy-based DR, we still find large variation in how companies have chosen to implement DR schemes, to say nothing of the potential variation in unexplored arrangements. For example, demand response providers like OhmConnect aggregate residential consumers and bid their load into the wholesale market. Businesses, shopping centers, industrial facilities, and even electric car chargers could provide demand response by reducing and deferring demand. Although the ideal DR resource is available all the time, as much as needed, in reality each of these cases comes with their own abilities regarding when, how often, and for how long they can be called upon to reduce their load.

This variation results in many different types of restrictions in how DR can be dispatched. For example, many residential customers, especially those without connected smart home appliances, need significant advance warning to reduce their demand. Deferrable car charging may come with limits on the amount of energy that can be reduced in a given time frame or the duration of the reduction to ensure that cars are charged up in time for their drivers to use them. Industrial applications may only be able to reduce their load a certain number of times per month or year to hit their production targets.

With so many different ways of implementing DR available, aggregators face a decision regarding which types to target. The knowledge of which limits are preferable would help the aggregator target more valuable types of consumers for inclusion in demand response aggregations. Is the deferral of car charging or control of building HVAC systems more valuable? Should an aggregator try to use contracts that limit what times of day customers are called on, or rather limits how many hours in a week they'll be used? With this knowledge, aggregators could choose to market to customers who are more able to provide DR that is more valuable to the system they are selling to. System operators, as they are updating bid structures in their markets to accommodate DR, can make sure to include parameters that accurately represent the forms of DR that are more desirable.

These issues point to the importance of deepening our understanding of the value of DR that is imperfect, i.e. DR that is subject to some limitations like when and how much it can be used, or what type of notification timing it needs to work properly. Many of these possible types of limitations look different for DR than for traditional generators, so there is not an existing body of research on how these types of limitations might affect system operation. Our study seeks to help answer these questions so that market participants can focus on the types of DR that create the most value and flexibility for the grid.

To shed some light on these issues, we use a two-stage stochastic unit commitment model of the ERCOT system to simulate the impact of DR subject to five different types of limitations within the wholesale market context. These limits affect: (1) the number of startups in a given time frame, (2) the number of hours of operation in a given time frame, (3) the amount of energy use reduced in a given time frame, (4) which hours DR is available to be dispatched, and finally (5) how far in advance DR must be given commitment and production decisions. By comparing these qualities in the same modeling framework, we can identify their relative importance.

In existing markets, bid structures do not fully recognize the nuances of DR functions and capabilities, especially the load shifting aspects of DR. Only the load shedding capabilities of DR are recognized and compensated. This structure allows DR to be represented in a similar way to traditional generators in market models. (At least one pilot program in California is targeting DR that increases load during times of surplus supply, but this remains the exception rather than the rule.) As long as levels of DR are low enough that load shifting does not create significant new peaks or ramps in demand, only representing the demand reduction may be a workable approximation.

In this study, we represent DR in the same way as done in these wholesale markets. This approach means results will necessarily be limited by the models inability to represent the deferment of demand, which can create new peaks in extreme cases [2]. However, this approach provides computational simplicity and allows our results to be mapped more easily to existing markets.

1.1 Prior literature

Existing literature examines the features and potential value of DR, though only a minority of studies examine DR as it is represented in wholesale electricity markets or otherwise take a system-wide perspective of its value and include a somewhat detailed representation of the grid's operation. The studies that do attempt to determine the system-level value of DR have examined the impact of several types of restrictions in the operation of DR. [2], [1] and [7] look at the impact of the reliability of DR on its capacity and reliability value. The degree of centralized control of DR can also affect its value in a similar way, and is investigated by [10]. [8] and [2] examine the impact of availability hours; [8] additionally examines the maximum duration of response. Payback or rebound, i.e. the amount of shed load that must be shifted to another period and how long it can be shifted, could also have a significant impact on the value of DR, especially at high penetrations, as shown in [21]. This paper builds on the above literature by comparing different DR qualities side-by-side in the same system, instead of one-by-one.

2 Methods

2.1 Stochastic Unit Commitment Model

To represent the impact of DR on the operation of the grid, we use a two-stage stochastic unit commitment and economic dispatch model. Our model is complex enough to capture the factors that might have a substantial impact on the dispatch of DR, including ramping constraints, startup costs, and generator minimum load thresholds. We make several simplifying assumptions to limit the computational complexity and data requirement. Although the precise cost estimates from the model may not be accurate, the trends shown between different scenarios should be representative of what we might expect in an actual system.

Our model takes its inspiration from [9] and [10]. Its formulation is given in Appendix A. DR is represented as it is in wholesale markets: as a shed only resource that does not account for load shifting. As a result, DR is represented similarly to traditional generators, but with special properties: no startup costs, no ramping constraints, and a near-zero minimum generation level. The use of a unit commitment and economic dispatch model allows us to simulate more detail in ramping and startup constraints. The model's stochastic nature allows us to mimic a day-ahead and real-time market by representing the day-ahead uncertainty in the demand forecast. This feature enables us to tease apart the value of DR that can be scheduled in real-time instead of a day in advance.

Uncertainty is driven by day-ahead demand forecast error. First stage decision variables must hedge against possible realizations of demand, and represent choices made in the day-ahead market regarding how much generation must be committed for the following day. Second stage decisions are analogous to real-time market decisions, and are made for each possible demand scenario. Generators are divided into fast and slow types; slow types must have their startup and commitment decisions made in the first stage, and their power output levels are decided in the second stage, subject to their commitment status. Fast types have their startup, commitment, and production decisions made in the second stage. Commitment decisions are binary, but for the purposes of our model the binary constraint is relaxed to the range between 0 and 1 for computational efficiency.

The operational time-frame used in this model is 5 days. A single day is the more traditional unit of simulation in similar studies [10, 20], but this formulation does not allow for several important features of our study. First of all, we are interested in looking at restrictions in the dispatch of DR over the course of several days; for example, some customers may only want to reduce their load once a week. A single day time horizon does not allow for these types of longer time-horizon restrictions. Second, unit commitment models are subject to end effects for a few hours near the start and end of the simulation, due to startup costs and minimum production costs. A longer time horizon allows us to eliminate the suspect hours on either end and focus on the more robust middle.

Finally, the ERCOT system features two large nuclear plants which must be on for almost a full day to justify their startup costs. Especially in times of high renewables production, they may not be dispatched immediately within our model. This operation would be unrealistic because it would overly weight the startup costs against the variable costs due to the short time horizon. A longer horizon results in more realistic weighting of startup against variable costs.

Our selection of time-frame makes a few assumptions about the system operation that are important to note. Since we are using day-ahead uncertainty but making decisions for five days at a time, we effectively assume the system operator has much better foresight than they actually have because when first stage decisions are made, the demand on days 1-5 has the same level of uncertainty. As a result, the operation of the system would be a best-case scenario, providing a lower bound on costs. For instances where DR is restricted to a certain amount of usage over the simulation, we assume much better foresight regarding when DR will be most valuable than a system operator would typically have, again leading to a best-case scenario for DR dispatch.

Instead of modeling the whole year for every scenario, we choose a sub-selection of days based on those days that had the highest prices, ramp prates, and first stage committed capacity. We grouped these days into 22 five-day simulations. The simulations overlap by 24 hours, allowing us to eliminate the 12 hours on either end of each simulation. In total, the results for each scenario are based on eighty-eight days.

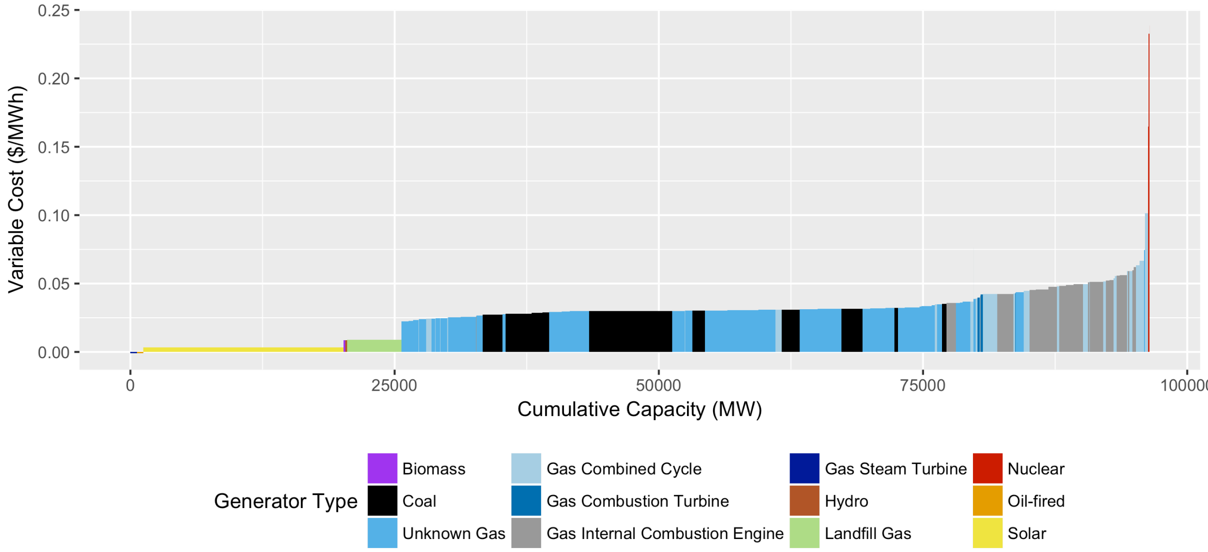


Figure 1: *Supply curve for ERCOT in 2016 as modeled.*

2.2 Data

The system modeled is ERCOT, which serves most of Texas, and uses data from 2016, the most recent year with complete data at the time of the study. Transmission costs and constraints are not considered, and the system is modeled as a single node.

Generator data is taken from EIA Form 860, and information regarding heat rates is taken from the EPAs eGrid database [16], [19]. We remove plants that are retired or not yet built. Where values are missing in eGrid, we substitute the average by generator type. Generator capacities are confirmed via the ERCOT Capacity Demand and Reserve Report [3], which is also used to weed out some generators which may not have been operating on the ERCOT system in 2016. Variable costs of production for each plant are composed of fuel costs (using costs from [17] and the heat rate from [19]), and variable operations and maintenance costs taken from [14]. Startup costs are assigned by technology type from [6]. No differentiation is made in this study between cold and hot startups. Values for these parameters and for ramp rate limits are given in Appendix B. The supply curve of the resulting generator set is shown in Figure 1.

Historical wind and solar production, as well as demand, are all taken from the year 2016 to preserve any correlation between renewable generation and load. Wind data by generator is available from ERCOT ([4]). For wind generators in the EIA database without historical production data, we assign data from the nearest available generator. Solar availability is estimated from historical insolation data according to [11, 13], assuming single-axis tracking. We gathered production estimates on a 2x2 lat-lon grid across the ERCOT region and assigned the closest point to each solar installation. Historical demand and day-ahead demand forecasts are taken from EIA Form 930 [18]. Where NAs are present, data from adjacent times is used to interpolate.

2.3 Statistical Model of Uncertainty

In our two-stage stochastic model, uncertainty is represented by a suite of possible scenarios for demand. We create these scenarios using historical demand and day-ahead demand forecasts, from which we determine the forecast error. We represent the forecast error as a percent of the total forecasted demand and use an autoregressive process to model it. The model parameters are found by maximum likelihood estimation using a Kalman filter [5, 12]. To generate demand scenarios, we apply the autoregressive model to a randomly generated timeseries drawn from the same distribution as the original forecast error, which is used for the 'burn in' period. Twenty-five different scenarios for forecast error are used in total. A sample of these resulting demand from a sample of these scenarios over a two day period is shown in Figure 2. All scenarios are given equal weight in the model.

2.4 Implementation of DR and Restrictions

The baseline DR resource is an optimistic representation to illustrate what might be possible with an ideal DR implementation. This resource has no startup cost, a minimum generation level near zero, constant availability, and a marginal cost of $35/MWh in the base case, putting it on par with a coal plant. In reality, most consumers would not be able to or want to modify their load as frequently as this resource does. This idealistic case serves as a reference point for more realistic restricted DR implementations. We assume 1000 MW of this resource is available to the wholesale market in the form of many homogeneous smaller resources, which we model together as a single 'pseudo-generator' with a relaxed binary commitment variable.

Five types of restrictions on DR usage are compared in this study, which can be divided into pure usage restrictions and notification restrictions. The usage restrictions are on the number of startups for DR, the number of hours it is committed, the amount of energy shed, and the hours of the day during which DR is available. These restrictions are for the five-day period of each simulation. To ensure that DR cannot be committed in all hours costlessly to avoid startups, the minimum generation of DR is set to 0.1% of its capacity.

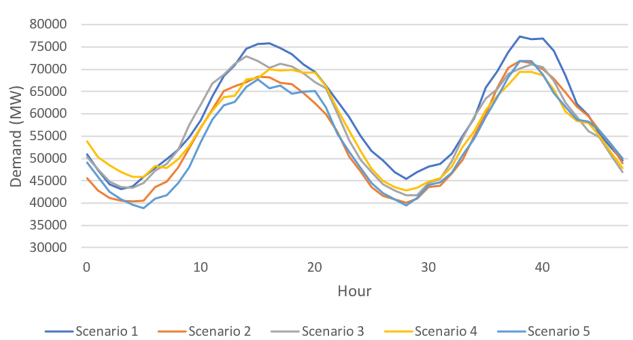


Figure 2: *Five sample realizations of demand over 48 hours generated using an autoregressive model.*

Notification restrictions are represented by modifying the formulation of the commitment and production variables for DR. Commitment and production can each be first stage variables, meaning they must be decided before uncertainty is revealed, or second stage variables, meaning that they can be decided uniquely for each scenario after uncertainty is revealed. In the most optimistic case DR is a fast resource, meaning that both types of decisions can be made in the second stage, which corresponds to DR that can react to real-time market awards. In the middle case, DR is the same as a slow generator, in which commitment is a first stage decision but production can be decided in the second stage. In the most restrictive case, DR requires full day-ahead notification of both its commitment and production decisions, so they are both first-stage variables. Each of these representations could be imagined in DR resources with varying levels of automation and user control.

3 Results

We can compare these different types of DR restrictions in several ways. A change in total system costs indicates their overall value. Shifts in the distribution of ramp rates in dispatchable generation gives us a sense of the flexibility value of DR (albeit on an hourly timescale). The total MWh shed by DR resources, and the number of hours in which DR was committed are also useful metrics for examining how the DR resource was used. We compare these restrictions to a base case of DR with no limits, as well as a model of the system without an additional DR resource.

3.1 Usage Limits

We examined four types of usage restrictions: limits on the amount of energy used, the hours of commit-

Type of DR	Reduction in 99th percentile ramp (MW/hr)
Fast	40.67
Advance Commitment	40.67
Advance Production	33.45
Afternoon & Evening	0.00
Daytime	40.67
Energy Limit (10GWh)	17.67
Energy Limit (5GWh)	0.00
Hour Limit (10)	17.67
Hour Limit (5)	0.00
5 startups	40.67
3 startups + 30 hour limit	37.09
1 startup	40.67

Table 1: *Shift in the distribution of upward ramps, as illustrated by the change in the 99th percentile upward ramp relative to a case with no DR.*

ment, the number of startups, and the times of availability. They all resulted in somewhat different levels of use of the DR resources, making it less straightforward to compare their value. We would expect that infrequently-dispatched resources are likely to be dispatched only in high-value hours, while more frequently dispatched resources will be used in both high and lower-value hours. As a result, we expect that types of DR that are used to shed less energy would have a higher value per MWh shed. As a result when we see types of DR resources that are dispatched more than others but create less value per MWh, it tells us that that restriction has a significant impact on the value of DR to the system.

Energy and hour limits interestingly have similar value to the system at similar levels. (e.g., a 5000 MWh energy limit is the same as a 5 hour limit for a 1000 MW resource.) These limits reduce the amount that the DR resource is dispatched more than in other scenarios, resulting in a much lower system benefit for the same amount of DR capacity. However, since this DR is saved for high-value hours, the system cost savings per MWh shed are much higher than other types of restrictions (Figure 4).

Availability restrictions examined two different timeframes for DR usage. The first, daytime-only DR (7am-10pm), might represent the HVAC functions of large shopping centers or office buildings; it also corresponds with ERCOT's on-peak hours[1]. Afternoon/evening DR (3-9pm) might represent appliance and AC use by households returning from school and work. Unsurprisingly, daytime-only DR provides almost the same level of system benefits as always-available DR (the base case), as cost-driving ramp rates and high marginal prices are almost never seen at night. Evening-only DR, on the other hand, is dramatically less valuable. Summertime peaks in Texas often begin before 3pm, so this type of DR is not available to reduce high afternoon ramps (see Table 1) nor is it available to reduce peak demand in the early part of the peak. The cost savings per MWh shed are equivalent to the base case (Figure 4c), but since this type of DR is used to shed demand in dramatically fewer hours than the base case, we would expect it to have a higher value per hour, as seen in the energy and hour limits.

Startup restrictions are, in theory, a practical way of implementing a restriction on the number of unique 'events' that a DR resource experiences. However, in practice they do not work very well in the absence of other restrictions. With nothing stopping the system operator from maintaining DR operation at a low level for a long period of time, startup limits can be met by simply never 'shutting down' DR. This is what we see when we restrict the model to only one DR startup per day, as illustrated in Figure 3. In this case, DR sheds very slightly (0.42%) more MWh than the base DR case over the study period, but it is committed in five times the number of hours.

Even in the case of a relatively relaxed 5 startup limit over 5 days, we see an elevated level of commitment. Additional restrictions like a no-load cost, a response-duration constraint, or a limit to the number of hours of operation are needed for a startup restriction to create a desired number of unique 'events'. A secondary scenario with a limit of 3 startups coupled with a 30-

[1] ERCOT Glossary http://www.ercot.com/glossary/o

hour commitment limit to control the over-commitment problem still saw higher amounts of commitment than in the base case. These results are partly driven by the low marginal cost of DR in this model, the low minimum energy shed from DR, and the resulting low cost of commitment. The option value of being able to dispatch such a cheap resource is worth paying a small commitment cost. For DR with a high marginal cost and/or a high commitment cost, we would expect different levels of operation and commitment.

Besides pure system-cost savings, we are also interested in the ability of DR to reduce major ramps in generation from dispatchable plants. These types of ramps are expensive for thermal generators, especially those that were not built with ramping capabilities in mind, and higher levels of renewables may increase the frequency and size of these ramps. Evening-only DR does not reduce the frequency of high-ramp periods relative to a case with no DR; the most dramatic ramps in Texas occur in the early afternoon before this type of DR is available (see Table 1). DR that is subject to hour and energy limits is also less able to reduce ramps, likely because it is being deployed in times of high marginal cost of production, instead of during shoulder times that

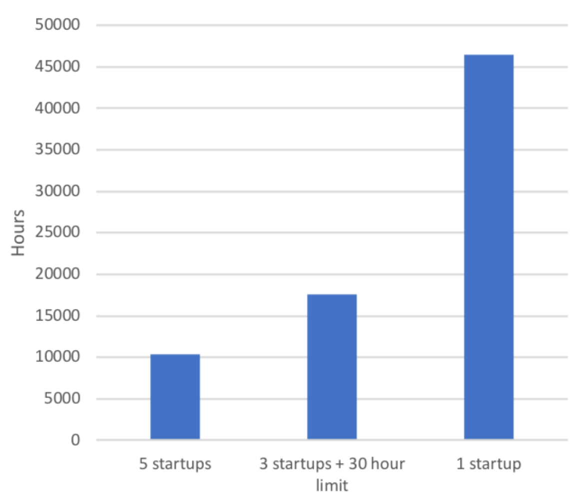


Figure 3: *Expected number of hours that DR is committed during the 88 'key' days modeled. Three types of startup restrictions are: 5 startups, 3 startups and 30 hours of commitment, and 1 startup, all applied over a 5-day time-frame.*

feature strong ramps.

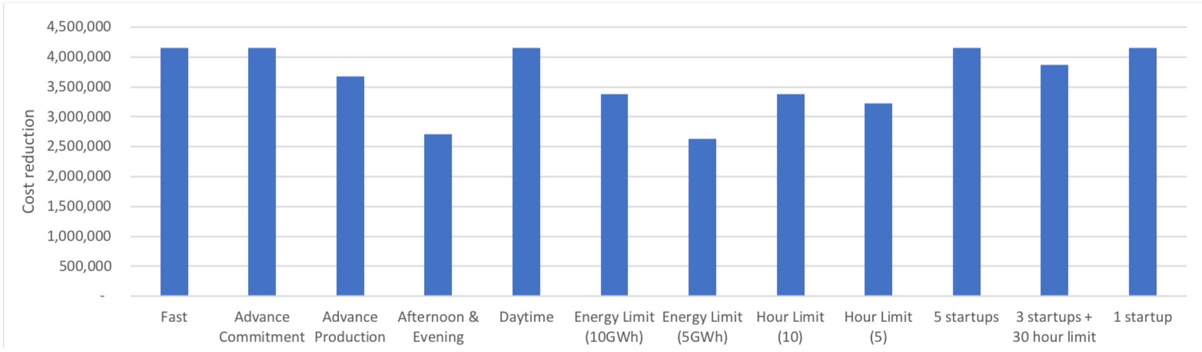
Interestingly, DR that requires an advance production decision is also somewhat less effective at reducing maximum ramps. This result points to the influence of uncertain changes in demand on the maximum ramp rates of the system (a result which could also be cause by uncertain RE production in real systems) . It seems the keys for reducing ramps among dispatchable generation on the ERCOT system are, in order of importance, availability in the early afternoon (or whenever times of maximum ramp and load are), availability for multi-hour events that stretch from the ramp to the peak, and fast operation to respond to unexpected ramps.

3.2 Notification Limits

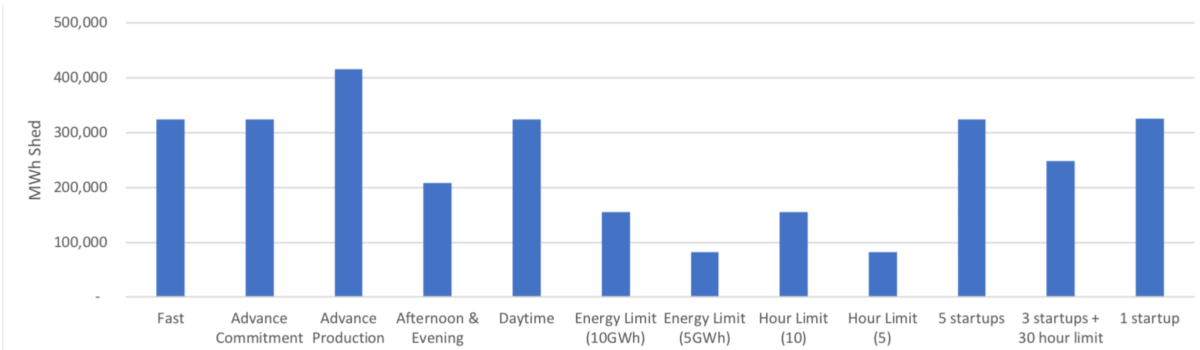
Demand response with notification limits can still be valuable, but much higher amounts of commitment or production are needed to achieve similar value to the base DR case. We examine two types of notification requirements: advance commitment demand response (AC DR, which functions the same as a slow generator), and advance production DR (AP DR) which requires production schedules in the first stage (i.e. day-ahead market) that cannot be modified. For both, we find that the notification limit results in more, not less, dispatch of DR (either through commitment or production), indicating that even this imperfect version of DR is preferable to a more expensive or ramp-limited generator.

AC DR is committed in more than twice as many hours as in the base case, yet it is used to shed virtually the same amount of energy for the same level of cost savings, as shown in Figures 4 and 5. In this model we assign a small cost for commitment (via a very small minimum generation level) for demand response, so that it cannot be committed all the time costlessly. In our model, this small cost incurred by committing AC DR in the day-ahead market is worth it for the option value of dispatching AC DR in the real time market.

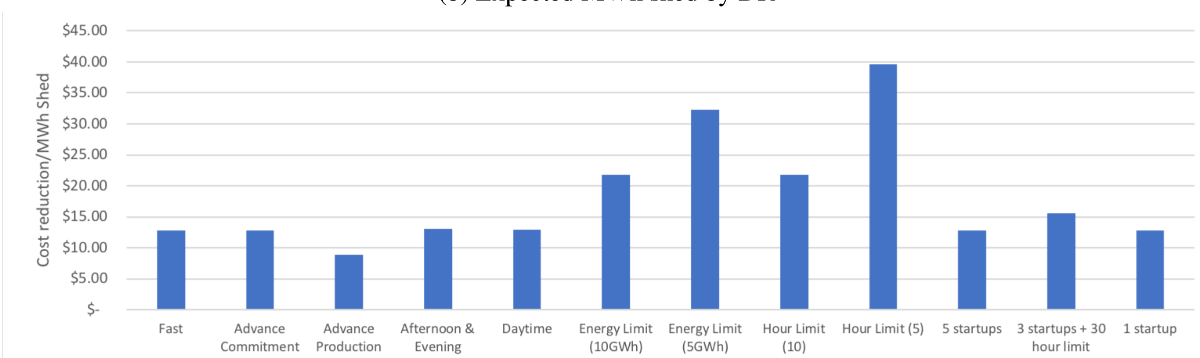
AP DR is not as valuable as AC DR or base DR.



(a) Expected cost savings relative to a scenario with no DR



(b) Expected MWh shed by DR



(c) Expected cost reduction per MWh shed by DR

Figure 4: *Effects of introducing DR with various limitations. All values are expected values across 25 demand scenarios, for the 88 most important days simulated. Fast DR, on the left, represents the base case.*

Since the model must decide on production levels without knowing the true levels of demand, sometimes it will be used to produce in an hour when it is uneconomic. Although it produces more than the base DR, it results in less cost reduction and also less ramp reduction than the base case or the AC DR case, as shown in Figure 4 and Table 1. However, it is still more valuable from a cost and ramping perspective than several of the usage restrictions.

AP DR also shifts costs towards fast generators and away from slow generators relative to base DR. Both production and startup costs are higher for fast generators and lower for slow generators when AP DR is used . This result points to the reduced flexibility value of AP DR and the ability of base DR to displace fast peaker plants. As the time resolution of the model is only hourly, the advantage of DR that can be dispatched in real time may be even higher if sub-hourly effects are taken into account. However the flexibility of generators in the model is likely somewhat overestimated, so the relative advantage of DR may be somewhat less than shown in this model.

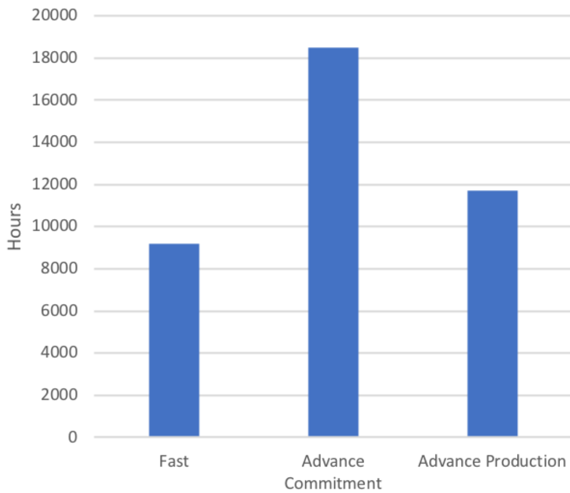


Figure 5: *Expected number of hours that DR is committed during the 88 'key' days modeled. The three types of DR restrictions are: Fast, also referred to as base DR, which can be dispatched in the real time market, advance commitment which gets a unit commitment decision in the day-ahead market, and advance production which gets a production schedule in the day ahead market.*

4 Discussion and Conclusion

We examined the value of different types of demand response (DR) to the electricity system within a unit commitment and economic dispatch model of the ERCOT system in Texas. A two stage stochastic model mimicked the impact of forecast uncertainty on the day ahead and real time markets. We evaluated four types of usage restrictions and two types of advance notification requirements, and compared them to 'perfect' DR resources that are available all the time with no advance notification, as well as to a system without additional DR.

These results begin to inform a discussion about what types of 'imperfect' DR are more preferable, a question that developers of demand response must address, given that consumers most likely want some type of limit placed on how much they can be used to shed or defer demand.

Given the choice between notification restrictions or usage restrictions, the former appears more valuable to the system. It is possible that consumers may be willing to defer or reduce their demand more often if they are given more advance notice. DR that requires an advance commitment provides similar value to unrestricted DR in terms of system cost reduction and offsetting large system ramps; this value comes at the cost of being committed in many hours where it may ultimately not be used, and is made possible by low commitment costs. DR that requires advance production schedules is slightly less valuable, but still reduces costs and ramps significantly. This type of DR is dispatched more frequently, which could lead to difficulties measuring baseline usage.

Startup restrictions, although an appealing way to control the number of 'events' a consumer experiences in a given time-frame, are relatively ineffective at accomplishing this goal without another type of usage restriction, or some type of 'no load' commitment cost. In the absence of these other restrictions, the system operator can maintain DR at a low level of demand shedding without turning it 'off,' thus avoiding another startup. A representation of the deferral of demand by the system operator, not just the shedding of demand, would also avoid this phenomenon, but there is significant uncertainty surrounding how long demand can be deferred and how much of the shed load must be recovered.

Energy and hour-based limits turn out to have quite similar effects; DR developers should consider which one makes the most sense for the type of customer resource they are contracting with. If it makes use of a consumer's battery backup, for example, an energy-based limit may make the most sense. Otherwise, an hour-based limit may be more intuitive.

These results call for further study. There are other types of DR characteristics that should be studied, like how reliably DR does what it is dispatched to do, how long it can shed load for, and different hours of availability. Combinations of new and already-studied characteristics might represent the true operational characteristics of known DR resources. Different marginal costs for DR, or fuel costs for existing thermal generators, could shift the conclusions. Refinements to the modeling structure would introduce additional complexity but perhaps answer questions with a higher fidelity. Sub-hourly timescales could shed light on shorter ramping events; combined with more detailed modeling of ramping and startup costs, we could gain more insight into the flexibility value of these resources. Finally, different systems with different types of variable renewables and load shapes may respond differently.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1656518. The author would like to thank Prof. John Weyant and Prof. Jim Sweeney for their continued advising and assistance. Additional thanks to John Taggart for his helpful comments.

References

[1] Kenneth Bruninx et al. "Valuing Demand Response Controllability via Chance Constrained Programming". In: *IEEE Transactions on Sustainable Energy* 9 (1) (2018), pp. 178–187. ISSN: 19493029.

[2] Robert Earle, Edward Kahn, and Edo Macan. "Measuring the Capacity Impacts of Demand Response". In: *The Electricity Journal* 22 (6) (2009).

[3] ERCOT. *Capacity, Demand and Reserves Report - December 2016*. 2016. URL: http://www.ercot.com/content/wcm/lists/96607/CapacityDemandandReserveReport-Dec2016.xlsx.

[4] ERCOT. *Hourly Aggregated Wind Output*. 2016. URL: http://www.ercot.com/gridinfo/generation.

[5] G. Gardner, A. C. Harvey, and G. D. A. Phillips. "Algorithm AS 154: An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering". In: *Applied Statistics* 29 (3) (1980), p. 311. ISSN: 00359254. URL: https://www.jstor.org/stable/10.2307/2346910?origin=crossref.

[6] N. Kumar et al. *Power Plant Cycling Costs*. Tech. rep. Sunnyvale, California: National Renewable Energy Laboratory, Intertech APTECH, 2012.

[7] Hyung Geun Kwag and Jin O. Kim. "Reliability modeling of demand response considering uncertainty of customer behavior". In: *Applied Energy* 122 (2014), pp. 24–33. ISSN: 03062619. URL: http://dx.doi.org/10.1016/j.apenergy.2014.01.068.

[8] Sheila Nolan et al. "A Methodology for Estimating the Capacity Value of Demand Response". In: (2014).

[9] Anthony Papavasiliou and Shmuel S Oren. *Multi Area Stochastic Unit Commitment for High Wind Penetration in a Transmission Constrained Network*. Tech. rep. FERC 2011 Software Conference, 2011. URL: https://www.ferc.gov/CalendarFiles/20110628073950-Jun28-SesD3-Papavasiliou-UCBerkeley.pdf.

[10] Anthony Papavasiliou and Shmuel S Oren. "Multiarea Stochastic Unit Commitment for High Wind Penetration in a Transmission Constrained Network". In: *OPERATIONS RESEARCH* 61 (3) (2013), pp. 578–592. ISSN: 1526-5463. URL: http://dx.doi.org/10.1287/opre.2013.1174.

[11] Stefan Pfenninger and Iain Staffell. "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data". In: *Energy* 114 (2016), pp. 1251–1265. ISSN: 03605442. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360544216311744andhttps://www.renewables.ninja/.

[12] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018.

[13] Michele M. Rienecker et al. "MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications". In: *Journal of Climate* 24 (14) (2011), pp. 3624–3648. ISSN: 0894-8755. URL: http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00015.1.

[14] Rick Tidball et al. "Cost and Performance Assumptions for Modeling Electricity Generation Technologies". In: *National Renewable Energy Lab - Subcontract Report* (November) (2010), pp. 37–41. URL: http://www.osti.gov/servlets/purl/993653-tJWyM6/.

[15] U.S. Department of Energy. *Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them*. Tech. rep. U.S. Department of Energy, 2006. URL: https://eetd.lbl.gov/sites/all/files/publications/report-lbnl-1252d.pdf.

[16] U.S. Energy Information Administration. *EIA Form 860*. 2016. URL: https://www.eia.gov/electricity/data/eia860/.

[17] US Energy Information Administration. "EIA Form 923". In: (2016).

[18] U.S. Energy Information Administration. *EIA Form 930*. 2016. URL: https : / / www . eia . gov / realtime{\ _ }grid / {\ # } / status ? end = 20190323T01.

[19] U.S. Environmental Protection Agency. *Emissions & Generation Resource Integrated Database (eGrid)*. 2016. URL: https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid.

[20] Chaoyue Zhao et al. "Multi-Stage Robust Unit Commitment Considering Wind and Demand Response Uncertainties". In: 28 (3) (2013), pp. 2708–2717.

[21] Yutian Zhou, Pierluigi Mancarella, and Joseph Mutale. "Modelling and assessment of the contribution of demand response and electrical energy storage to adequacy of supply". In: *Sustainable Energy, Grids and Networks* 3 (2015), pp. 12–23. ISSN: 2352-4677. URL: http://dx.doi.org/10.1016/j.segan.2015.06.001.

Appendix A: Model Formulation

Sets:

f: fast generators

s: slow generators

d: demand response pseudo-generators

g: all generators: the union of f, s, and d

t: timesteps

ω: scenarios

Parameters:

$PMIN$: min generation levels per generator

$PMAX$: max generation levels per generator

F: production cost ($/MWh)

C: startup cost

A: generator availability

R: max ramp rate, as a fraction of total capacity

S: startup limit (if used)

E: energy limit (if used)

H: hour limit (if used)

$prob$: probability of scenario ω

Variables:

p: production (second stage)

q: production by advance production DR (first stage)

z: slow generator startup (first stage)

v: all generator startup (second stage)

w: slow generator commitment (first stage)

u: all generator commitment (second stage)

c: number of total startups (second stage)

Equations

Objective: Minimize expected cost

$$min\ Cost = \sum_{\omega} [c_g * C_g + p_{gt\omega} F_g] prob_{\omega} \tag{1}$$

Subject To

Supply demand balance

$$\sum_{g} p_{gt\omega} \geq D_{t\omega} \; \forall \, t, \omega \tag{2}$$

Respect generator min and max

$$p_{gt\omega} \geq PMIN_g \, u_{gt\omega} \, A_{gt} \quad \forall \, g, t, \omega \tag{3}$$

$$p_{gt\omega} \leq PMAX_g \, u_{gt\omega} \, A_{gt} \quad \forall \, g, t, \omega \tag{4}$$

Startup, commitment, and initialization

$$z_{gt} = w_{gt\omega} - w_{g,t-1,\omega} \quad \forall \, g \in s, t > 1, \omega \tag{5}$$

$$v_{gt} = u_{gt\omega} - u_{g,t-1,\omega} \quad \forall \, g, t > 1, \omega \tag{6}$$

$$z_{gt} = w_{gt\omega} \quad \forall g \in s, t = 1, \omega \tag{7}$$

$$v_{gt} = u_{gt\omega} \quad \forall g, t = 1, \omega \tag{8}$$

Non-anticipitivity constraints for slow generators

$$v_{s,t,\omega} = z_{s,t} \quad \forall \, s, t, \omega \tag{9}$$

$$u_{s,t,\omega} = w_{s,t} \quad \forall \, s, t, \omega \tag{10}$$

Ramping constraints

$$p_{g,t,\omega} - p_{g,t-1,\omega} \leq R_g \quad \forall g, t, \omega \tag{11}$$

Startup detection

$$c_{g,t,\omega} \geq v_{g,t,\omega} \quad \forall g, t, \omega \tag{12}$$

$$c_{g,t,\omega} \geq 0 \tag{13}$$

For advance production DR

$$q_{g,t} \leq PMAX_d \, w_{g,t} \, A_{g,t} \qquad \forall g \in d, t, \omega \tag{14}$$

$$p_{g,t,\omega} = q_{g,t} \qquad \forall g \in d, t, \omega \tag{15}$$

Startup constraints

$$\sum_{t} c_{gt\omega} \leq S \quad \forall g \in d, \omega \tag{16}$$

Energy constraints

$$\sum_{t} p_{g,t,\omega} \leq E \quad \forall g \in d, \omega \tag{17}$$

Hour limits

$$\sum_{t} u_{g,t,\omega} \leq H \quad \forall g \in d, \omega \tag{18}$$

Relaxed binary and non-negative constraints

$$0 \leq w, u \leq 1 \tag{19}$$

$$p, q, c \geq 0 \tag{20}$$

Appendix B: Generator Data

Plant Type	Max ramp rate (fraction of capacity)	Variable O&M ($/MWh)	Startup costs ($/MW)	Slow or Fast?
Biomass	1	8.742	0	Slow
Coal <300 MW	0.6	4.831	157	Slow
Coal	0.6	4.831	65	Slow
Other Gas	1	3.385	55	Slow
Gas Combined Cycle	1	2.828	55	Slow
Gas Combustion Turbine	1	3.942	126	Fast
Gas Internal Combustion Engine	1	5.850	55	Fast
Gas Steam Turbine	1	4.831	58	Fast
Hydroelectric Plant	1	0.000	0	Slow
Landfill gas	1	5.850	0	Slow
Nuclear	0.17	1.050	300	Slow
Petroleum	0.9	0.000	55	Fast
Solar PV	1	0.000	0	Fast
Solar Thermal	1	0.044	0	Fast
Wind	1	3.364	0	Fast

Table 2: Default data used in this model. Sources are given in the document text.

Fuel	Cost	Units
Coal	$2.311	$/MMBTU
Natural Gas	$3.789	$/MMBTU
Petroleum	$15.83	$/MMBTU
Nuclear	$7.7	$/MWh

Table 3: Fuel costs used to calculate variable production costs. Sources are given in the document text.