Memorandum

|	
TO:	ERCOT AS Methodology Team
FROM:	Jeff McDonald, Andrew Reimers
DATE:	August 5, 2025
RE:	IMM Comments on 2026 AS Methodology

A. Executive Summary

As the Independent Market Monitor for the ERCOT wholesale market, we appreciate the opportunity to present our concerns with ERCOT's proposed AS Methodology for 2026. Our concerns focus on two key issues: (1) the proposed methodology is not aligned with reliability outcomes, resulting in excessive AS procurements, and (2) it will undermine the performance of ERCOT's energy-only market. We begin by outlining our comments and conclusions surrounding these key issues. Next, we recommend changes that would achieve comparable reliability outcomes as the proposed methodology at a lower cost. Finally, we provide a more detailed discussion of each model input's contribution to the excessive procurement volumes, as well as a chronology of how ERCOT arrived at this conservative operational paradigm.

1. The AS Methodology is Misaligned with Reliability Outcomes

- ERCOT's operational objective is to avoid Watch conditions instead of reliability outcomes such as preventing load shedding.

- The input parameters for individual AS products are not based on the risks they are designed to manage. For NSRS, the forecast error time horizon is set at 6 hours, and the duration requirement for ESRs is set at 4 hours, even though NSRS should only needed for no more than one hour before elevated prices incentivize additional commitments from thermal resources.

- The current model unnecessarily discounts 75% of available headroom during daytime hours which creates an artificial scarcity that inflates the AS plan further.

Together, these inputs result in the proposed AS Methodology that procures 133% more than what is required to satisfy the 1-in-10 reliability standard for load shed, *the last 3 GW of which provide no additional reliability*. This is illustrated in the following figure. The blue curve represents the probability of firm load shed at different procurement levels of ECRS and NSRS. This curve reflects the IMM base case and uses a one hour load forecast error and a one hour duration requirement for energy storage resources. The orange curve reflects the probability of entering a Watch condition at different procurement levels of ECRS and NSRS and uses the

ERCOT base case of a six hour load forecast error and a four hour duration requirement for energy storage resources. The procurement level required to achieve a one-in-ten (or 10%) probability of each outcome is identified by the intersection of horizontal black line with these two curves.
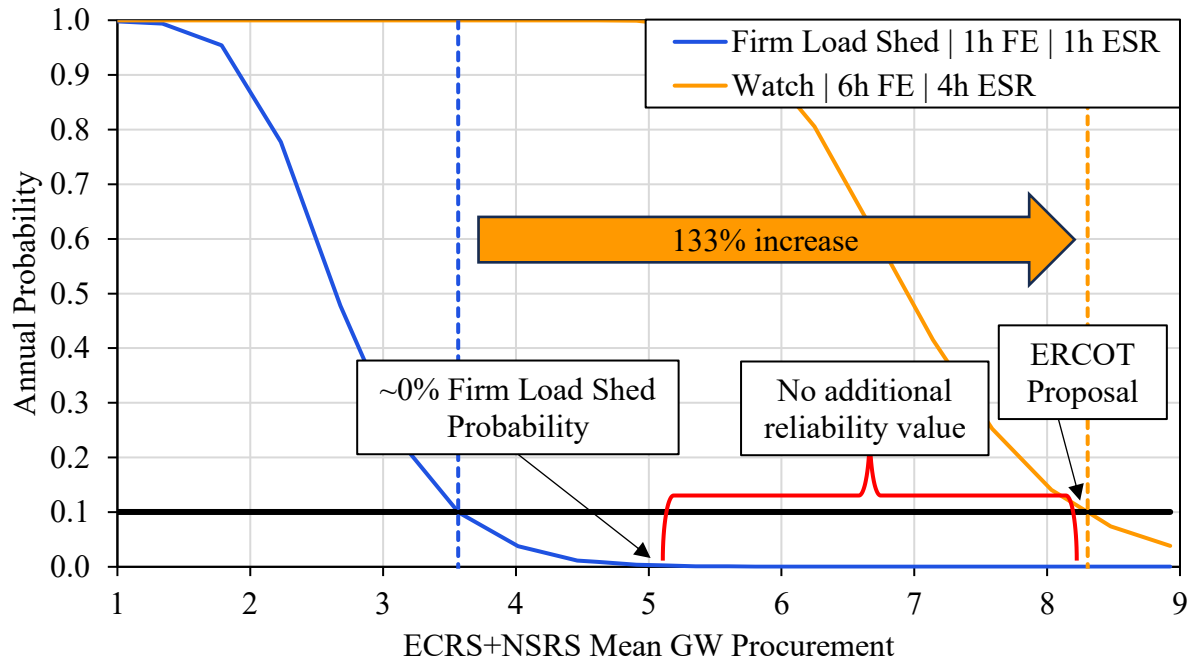
This figure shows that the level of total ECRS and NSRS needed to meet the one-in-ten (0.1 probability) reliability standard would average roughly 3.6 GW. In contrast, ERCOT's proposal would procure 8.3 GW of reserves, more than 3 GW of which would have no reliability value.

2.   The AS Methodology will Undermine the Performance of the Energy-Only Market

ERCOT's energy-only market design must allow for scarcity pricing to send market signals to incent investment in new generation. Procuring excessive volumes of ancillary services will inefficiently inject additional non-price-setting energy into the real time market and reduce prices. Ultimately, it can undermine the ability of shortage pricing to provide the efficient long-term signals for new investment.

The implementation of RTC could mitigate the impact of excessive operating reserve procurements if the ORDC and associated ASDCs scaled up with the AS procurement targets. However, the fact that the ASDCs will drop to the price floor at quantities well below the procurement targets, and that ERCOT will use out-of-market RUC commitments to meet the targets, will undermine the formulation of efficient scarcity pricing. In other words, the actions taken to procure and maintain the excessive levels of operating reserves will inefficiently prevent the supply margin falling into ranges where the market will provide substantial and efficient scarcity pricing.

B. Recommended Compromise Proposal

We understand that the heightened reliability concerns following Winter Storm Uri together with the looming threat of rampant load growth is motivating a conservative AS methodology. However, the ERCOT proposal goes too far and will result in significant excess cost with no added reliability benefit. We identify a number of issues with the methodology proposed by ERCOT but have limited our recommendations to three key areas. As a compromise, we recommend:

- A 3-hour forecast horizon for determining the target volumes for NSRS.

- Using a 1-in-10 criteria based on the LOLP curve in place of the Watch criteria.

- Using a 1-hour discharge horizon for ESRs rather than 4 hours.

Regarding the three-hour forecast horizon, ERCOT already has a track record of using such an input in their AS Methodology from 2016 to 2022 without any adverse reliability consequences caused by inadequate NSRS capacity. This compromise, detailed in Scenario 2.2 in Table 1.1:

- Represents a 34% increase in AS procurement relative to the IMM baseline scenario;

- Achieves a near-0% probability of load shed, and

- Avoids setting reserve targets beyond this point that do not add any reliability value.

Table 1.1: IMM Compromise - Near-0% Load Shed Probability

#	Scenario name	Convergence Criteria (MW)	Forecast Error Time Horizon (Hours)	ESR Duration	Annual Probability	ECRS + NSRS Plan (MW)	Absolute Increase from IMM Base Case (MW)	Relative Increase from IMM Base Case (%)
0	IMM Base Case	1,500	1	1	1/10	3,566		
2.2	3-HA Forecast Error	1,500	3	1	1/10	4,767	1,201	33.7%
4	ERCOT Base Case	3,000	6	4	1/10	8,304	4,738	132.9%

C. Comprehensive Analysis on AS Methodology

1. Introduction

We applaud ERCOT's efforts to incorporate a stochastic risk methodology into their formulation of the AS Plan. We have not yet developed a position on the statistical or probabilistic techniques used in their proposed methodology. However, we remain concerned about many of

the inputs to this methodology that are upstream of the new and improved modeling techniques. We have raised these concerns consistently for several years.

Much of our critique relates to "Conservative Operations," the operational paradigm ERCOT adopted in the aftermath of Winter Storm Uri. Conservative Operations refers to the set of policies and practices oriented around maintaining a larger operating reserve margin than what was standard practice in ERCOT before Winter Storm Uri to avoid going into a Watch and exceeding the required reserve levels of all other ISO in the US, as shown in Figure 1 in the Figures and Tables section at the end of these comments. It is implausible that this operating posture would have prevented the issues experienced during Winter Storm Uri. Further, several resilience mechanisms have been implemented since Winter Storm Uri, outside of the wholesale market, that will significantly reduce the impact of similar future winter storms. We classify the following input parameters to the proposed AS Methodology as part of the Conservative Operations paradigm:

- Setting operating reserve targets to achieve a one-in-ten standard for the probability of entering a "Watch," nominally defined as dropping below 3,000 MW, rather than the probability of firm load shed, which is defined as reserves dropping below 1,500 MW. This parameter increases the target level of the AS Plan by approximately 43%.

- Assessing the operational risk that NSRS is meant to manage based on the six hour-ahead forecast error for demand and generation from intermittent renewables. This parameter increases the target level of the AS Plan by approximately 57% compared to our recommendation of using a one hour-ahead forecast error.

- Accrediting the available headroom of ESRs according to the power output they can sustain for four hours, rather than our recommendation of one hour. This parameter increases the target level of the AS Plan by approximately 29%.

- Severely discounting the capacity of headroom in the system capable of providing operating reserves in the event of a contingency or forecast error that results in a reduction of reserves. *We are still estimating the relative impact of ERCOT's "risk credit" methodology in formulating the AS Plan.*

Together, this AS Plan is 133% larger than the IMM base case which achieves a 1-in-10 yearly probability of load shed. The last 3 GW of this plan provide no additional reliability value on average, as illustrated in Figure 2 (the same figure presented in the executive summary). We recommend that ERCOT revise their AS Methodology and consider operating reserve targets that reflect the reliability value they provide in mitigating real-time operational risks. The remainder of these comments detail our concerns with each of these parameters. A summary of the independent impact of each of these parameters on the size of the AS Plan is included in Table 1 at the end of these comments, and we include portions of Table 1 throughout.

2. Convergence Criteria

Prior to Winter Storm Uri, ERCOT had a policy of issuing a voluntary conservation notice referred to as a Watch whenever operating reserves dropped below 3,000 MW. Such events

were relatively commonplace, with an average of 82 hours per year with PRC below 3,000 MW for 2016-2020, as shown in Figure 3. The aftermath of Winter Storm Uri shone a spotlight on the reliability of the ERCOT grid and, given the recent legislative session and a gubernatorial election around the corner, it was decided that avoiding Watch conditions was of paramount importance. Thus, Conservative Operations was born, the primary feature of which was to run the system with a high level of operating reserves, whether procured through RUC or through the AS Methodology, as discussed in more detail in the following section.

In the summer of 2024, the IMM and ERCOT collaborated on the AS Study mandated by PURA 35.004(g), wherein we proposed a stochastic risk methodology that would more effectively capture the risks that real-time operating reserves, particularly ECRS and NSRS, were meant to address. This stochastic risk methodology is a Monte Carlo simulation of historical system conditions that accounts for the probability of forced outages and forecast error. The distribution of outcomes produced by these simulations determines the underlying probability of adverse system conditions such as firm load shed. That determination depends on "convergence criteria," i.e., the level of reserves below which an iteration is flagged as an outage. For our analysis, we used a convergence criterion of 1,500 MW of operating reserves, the same criteria ERCOT uses to signal an EEA3 and the point at which ERCOT nominally begins firm load shed.

ERCOT has repeatedly argued that we should use the probability of going into a Watch so that the results would reflect the de facto operating policy. That is, the analysis should work backwards from the decision that Watch conditions were to be avoided and that an exceptionally high level of operating reserves was the way to achieve that goal. We have consistently maintained that formulating the AS Methodology according to the probability of going into a Watch is inappropriate because it is an arbitrary, ill-defined threshold that is only significant because it functions as the first official warning sign that conditions are becoming tight. Alternatively, we proposed that if more conservative reliability targets were desirable, they should be defined according to objective reliability criteria, e.g., a one-in-twenty probability of firm load shed rather than a more typical one-in-ten standard as shown in Table 1.2.

Table 1.2: Base Case Comparison with 1-in-20 Criteria and Watch Criteria

#	Scenario name	Convergence Criteria (MW)	Forecast Error Time Horizon (Hours)	ESR Duration	Annual Probability	ECRS + NSRS Plan (MW)	Absolute Increase from IMM Base Case (MW)	Relative Increase from IMM Base Case (%)
0	IMM Base Case	1,500	1	1	1/10	3,566		
0.1	1/20 Standard for Firm Load Shed	1,500	1	1	1/20	3,927	361	10.1%
1.1	"Watch" Criteria	3,000	1	1	1/10	5,112	1,546	43.4%

Table 1.2 shows that a one-in-twenty standard for firm load shed would increase the target level of operating reserves by a little more than 10% rather than the 43% increase implied by the one-in-ten standard for going into a Watch.

ERCOT continued to assert that it was appropriately cautious to formulate the AS Methodology based on the probability of going into a Watch. The PUC adopted ERCOT's position in the form of guidance to that effect.[1] Thus, ERCOT's current proposed AS Methodology is formulated to achieve a one-in-ten standard of entering a Watch, a significantly more expansive criteria than is used by any other ISO in the US.

ERCOT's convergence criteria includes another even stricter component that escalates the bias for excess reserves in their proposal. Their convergence criteria selects the maximum of either 3,000 MW (i.e., a Watch) or the sum of the target volumes for RegUp and RRS. The latter criterion is the stricter of the two in 67% of hours in our study period, with RegUp and RRS summing on average to 3,149 MW. In Table 1.3, scenarios 1.2 and 1.3 show that the distinction between 1,500 MW and 3,000 MW in the convergence criteria is overwhelmed by the sum of RegUp and RRS, both resulting in target volumes approximately 50% higher than the IMM base case. The rationale for including this stricter criterion in their formulation is not clear, as the thresholds for Watch conditions and firm load shed are both defined according to PRC, rather than shortages of RegUp and RRS.

Table 1.3: RegUP and RRS Sum Comparison with Watch Criteria

#	Scenario name	Convergence Criteria (MW)	Forecast Error Time Horizon (Hours)	ESR Duration	Annual Probability	ECRS + NSRS Plan (MW)	Absolute Increase from IMM Base Case (MW)	Relative Increase from IMM Base Case (%)
0	IMM Base Case	1,500	1	1	1/10	3,566		
1.2	Max of Firm Load Shed or Reg + RRS	MAX(1,500, RRS+REG)	1	1	1/10	5,334	1,768	49.6%
1.3	Max of "Watch" or Reg + RRS	MAX(3,000, RRS+REG)	1	1	1/10	5,349	1,783	50.0%

---

[1] PUC guidance was to continue to operate to the Watch criteria until further data and assessment could be used to assess the appropriateness of operating to a Watch, Emergency Alert, or Load Shed event. The assessment is due to the PUC no later than the 2027 AS Methodology process.

3.   Time Horizon for Forecast Error

One of the main risks that ECRS and NSRS are intended to manage is the impact of forecast errors for demand and intermittent renewable generation. Under-forecasted demand and over-forecasted renewable generation have the effect of discouraging commitment from thermal resources, which can result in a supply shortfall in real-time. The AS Methodology should reasonably incorporate the probability of such shortfalls. A key parameter to that end is the time horizon over which these forecast errors are calculated.

All inputs to the risk assessment should be based on the operational risks that the product is meant to manage. NSRS, for example, is designed to mitigate the risks associated with forced outages or forecast errors over a time horizon of approximately one hour. In the event of such risks manifesting themselves, market pricing should be sufficient to incentivize commitment of offline resources to maintain the reliability of the system. Our analysis has consistently shown that ERCOT has significant capacity from offline reserves that can start in an hour or less under tight system conditions, as illustrated in Figure 4. These conditions produce higher prices which in turn elicit a market response of self-commitment from available off-line resources. Setting this parameter to six hours nearly doubles the perceived risk associated with forecast error, as shown in Figure 5. This parameter effectively assumes that none of the offline capacity will self-commit in response to evolving system conditions, which is contrary to historical experience.

ERCOT's operational history demonstrates that six hours is an unnecessarily long time-horizon to run the system reliably. From 2016 to 2022, the methodology for determining the target volumes for NSRS used a three hour ahead forecast error. Over that time, there was no firm load shed caused by a lack of sufficient NSRS.

Figure 6 shows the increase in the unit hours of resources committed through RUC and the increase in the average lead time between the commitment instruction and real-time. The following year, to achieve its desired level of operating reserves through the market rather than through overreliance on RUC, ERCOT increased the time horizon used for calculating forecast error in the NSRS methodology to six hours. Thus, *the methodology followed from the operational paradigm,* which itself was not supported by sound reliability criteria.

In summary, the operational response to Winter Storm Uri was to run the system with a higher level of operating reserves with the goal of avoiding Watch conditions. The AS Methodology was adjusted to reflect these decisions and not based on objective reliability criteria.

4.   Duration Accreditation for ESRs

We have consistently argued that the duration requirement for ESRs to carry ECRS or NSRS should both be set at one hour to reflect the time that may be needed to provide energy while thermal resources start-up in response to market conditions after a contingency. ERCOT has maintained the position that the duration requirement for ESRs to carry NSRS should be set at four hours. We covered our objection to this parameter in detail in our comments on NPRR

1282.[2] Incorporating this four-hour duration requirement into the AS Methodology discounts the reliability value provided by ESRs and increases the target volume for NSRS by more than 29%.

5.   Risk Credit Accounting

ERCOT's risk credit parameters unreasonably discount the headroom in the system. Risk Credit Accounting refers to accounting of the amount of capacity that can respond within 30 minutes. During daytime hours, ERCOT accounts for only 25% of available headroom, excluding capacity already carrying ancillary services. This decision is inconsistent with their convergence criteria, which is based on the probability of a Watch. A Watch, like the probability of firm load shed, is a function of PRC, and ERCOT's risk credit methodology effectively discounts 75% of the available PRC in their estimation of the probability of a Watch. We are still estimating how much ERCOT's risk credit methodology inflates the target level of reserves set by the AS Plan.

It is unclear to us why any discounting of the available capacity is needed. The stochastic risk methodology is meant to accurately reflect the probability of adverse system conditions and the corresponding response to such conditions by the grid operators and the real-time market dispatch. In our portion of the AS Study, we account for all available capacity that can respond to a contingency such as a forced outage or under-commitment of thermal resources caused by forecast error. We accounted for capacity that is already carrying AS, unloaded headroom from generation and energy storage resources, and demand response from price-responsive loads. The rationale for this accounting is twofold. First, it reflects the likelihood that operators will take advantage of any capacity available in the system before initiating firm load shed. Second, it reflects the functioning of the real-time market with co-optimization of energy and operating reserves. Following a contingency, some volume of operating reserves must be converted to energy, after which operating reserves are then reallocated to the remaining unloaded headroom.

We acknowledge that there are legitimate shortcomings to this type of accounting, particularly when based on historical system conditions. The level of available operating reserves depends on the composition of the resource mix, which is changing rapidly with the deployment of solar and energy storage resources and resulting commitment decisions of thermal resources, which are implicitly influenced by the target level of operating reserves procured in the market. That is, reducing the target level of operating reserves procured in the market would reduce the level of commitment, thus also reducing the level of free headroom available in the system. We would advise that forecasts of the changing generation mix and corresponding impacts on thermal commitments be incorporated into the AS Methodology. This is a more reasonable approach than discounting the capacity available in the system as ERCOT has proposed.

6.   Operating Reserves and the Energy-Only Market

We have sought to comprehensively document the effect of these parameters on increasing the target operating reserve volumes set by the AS Methodology and to demonstrate that they were

---

2 https://www.ercot.com/mktrules/issues/NPRR1282

set based on upstream policy decisions not connected to objective reliability criteria. Zooming out from the specific impacts on the AS Methodology, policies associated with Conservative Operations have been implemented all along the interface of operations and market design, often with inconsistent and conflicting implications. Consider the decision to increase the target volumes set by the NSRS Methodology in 2022. The motivation for this decision was to achieve the desired level of operating reserves through the Day-Ahead Market without as much reliance on RUC. Empirical analysis demonstrating the need to maintain an increased level of operating reserves was not provided to support the increased procurement level. Increasing the NSRS volumes procured in the DAM was a plausible strategy for achieving this goal because in the current market design, DAM always procures the entirety of the AS Plan.

Once RTC is implemented, however, sloped ASDCs will be incorporated into the day-ahead and real-time markets. As currently formulated, these demand curves do not scale with the target volumes defined by the AS Plan. The ASDCs are fixed despite significant variation in the hourly volumes of target reserves, and the AORDC is only defined up to 10,000 MW. At the extremity of the AORDC, the shortage price for reserves is so low as to allow the market to go short on AS in periods of modest scarcity but low probability of reliability issues. This dynamic is, in fact, how RTC is meant to function, where shortages of operating reserves are a critical aspect of price formation. ERCOT operations, however, has signaled skepticism of this feature of RTC, indicating their intention to use RUC to increase operating reserves when the market does not procure the entirety of the AS Plan *that was inflated specifically to avoid overreliance on RUC*. The logic leading to this set of decisions is circular, but ERCOT continued to work in this construct by imposing a $15/MWh floor on any shortages of AS in the day-ahead and real-time markets.

The ultimate effect of this kind of excessive operating reserve policy will be to simultaneously increase costs for consumers while suppressing the price signals needed for maintaining resource adequacy. Consumers will be subject to increased costs due to uplift from RUC and from DAM make-whole payments and the $15/MWh floor on the AORDC. At the same time, the excess supply of reserves and reluctance to endure any shortage of reserves will suppress the price signal that the ASDCs are meant to produce. Ultimately, this suppression of genuine shortage pricing will reduce the incentive for new entry, and the modest increase in revenues under minimal shortage conditions will serve only as a transfer payment to incumbent generators above the reliability value their resources provide.

Beyond our recommendations on the AS Methodology, we call for a more general reconsideration of Conservative Operations with the goal of cost-effectively achieving objective reliability targets. This reconsideration should be applied both to the AS Methodology and to the corresponding formulation for the ASDCs.[3] The ASDCs and the AS Methodology should be

---

[3] The PUC did provide guidance at the end of the AS Study requiring ERCOT to provide analysis for operating to a Watch, Emergency Alert, and Load Shed event no later than the 2027 AS Methodology process. We note that this guidance did not preclude ERCOT from providing that analysis for the 2026 AS Methodology process and given the excessive procurement of reserves under the current set of assumptions, we believe that such analysis should be performed for the 2026 AS Methodology.

explicitly linked such that the shortage pricing represented by the ASDCs reflects with the marginal reliability value of the corresponding ancillary service. We elaborate on the proper formulation of the ASDCs and on the AS Methodology in our most recent edition of the State of the Market report in recommendation 2024-1.

D. Figures and Tables

**Table 1. Summary of Impacts of Input Parameters for ERCOT's 2026 AS Methodology**

#	Scenario name	Convergence Criteria (MW)	Forecast Error Time Horizon (Hours)	ESR Duration	Annual Probability	ECRS + NSRS Plan (MW)	Absolute Increase from IMM Base Case (MW)	Relative Increase from IMM Base Case (%)
0	IMM Base Case	1,500	1	1	1/10	3,566		
0.1	1/20 Standard for Firm Load Shed	1,500	1	1	1/20	3,927	361	10.1%
1.1	"Watch" Criteria	3,000	1	1	1/10	5,112	1,546	43.4%
1.2	Max of Firm Load Shed or Reg + RRS	MAX(1,500, RRS+REG)	1	1	1/10	5,334	1,768	49.6%
1.3	Max of "Watch" or Reg + RRS	MAX(3,000, RRS+REG)	1	1	1/10	5,349	1,783	50.0%
2.1	6 HA Forecast Error	1,500	6	1	1/10	5,599	2,033	57.0%
2.2	3 HA Forecast Error	1,500	3	1	1/10	4,767	1,201	33.7%
3	4 hr. ESR Duration	1,500	1	4	1/10	4,613	1,047	29.4%
4	ERCOT Base Case	3,000	6	4	1/10	8,304	4,738	132.9%

**Figure 1: Reserve Requirements Across Various US Electricity Markets**
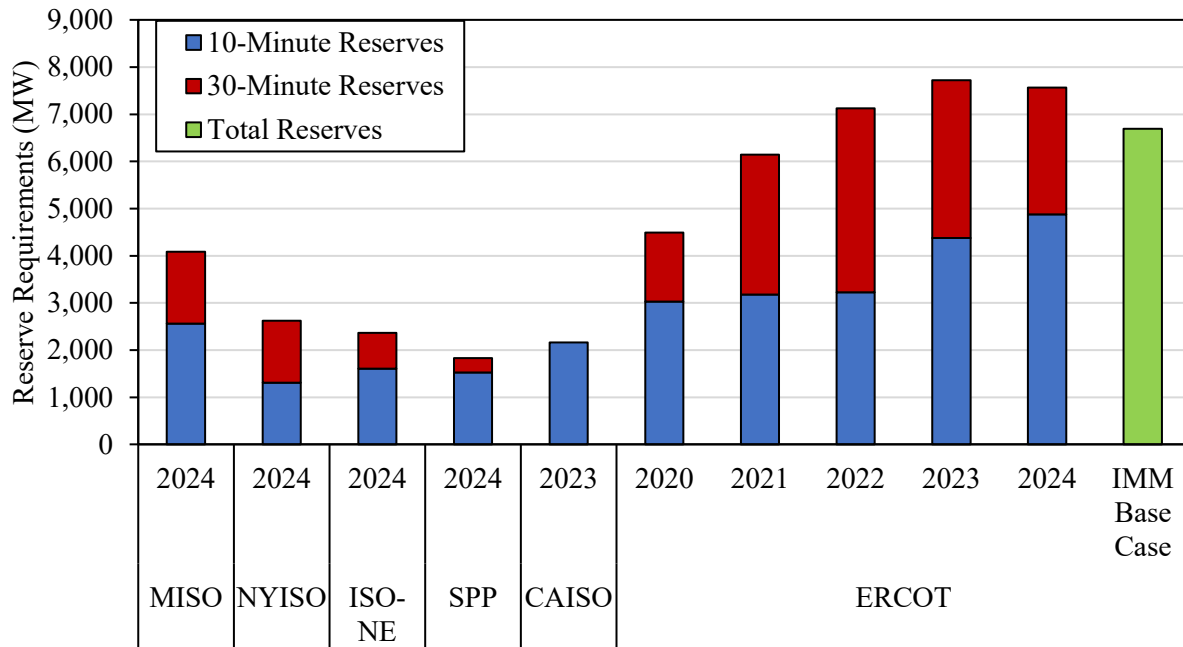
On average, ERCOT procures 153% more 10- and 30-minute reserves than other ISOs.

Figure 2: Annual Probability of Firm Load Shed and entering Watch Conditions*
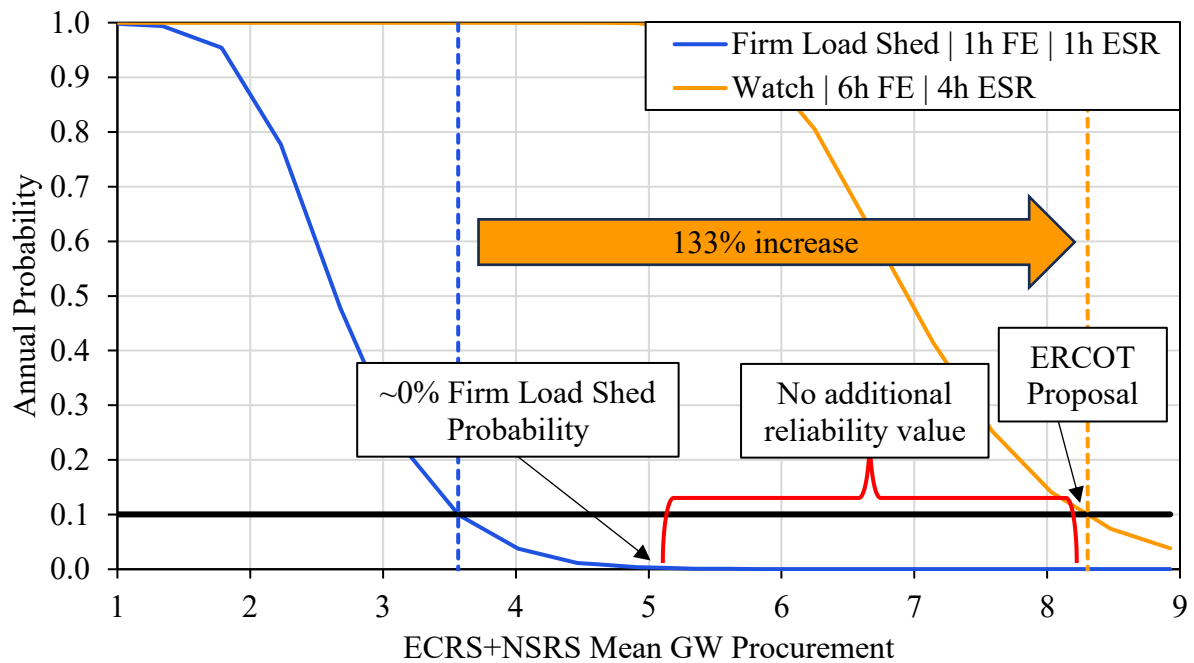
Note that our estimate of the target volumes for ERCOT's base case will be even higher after we account for their risk credit methodology.

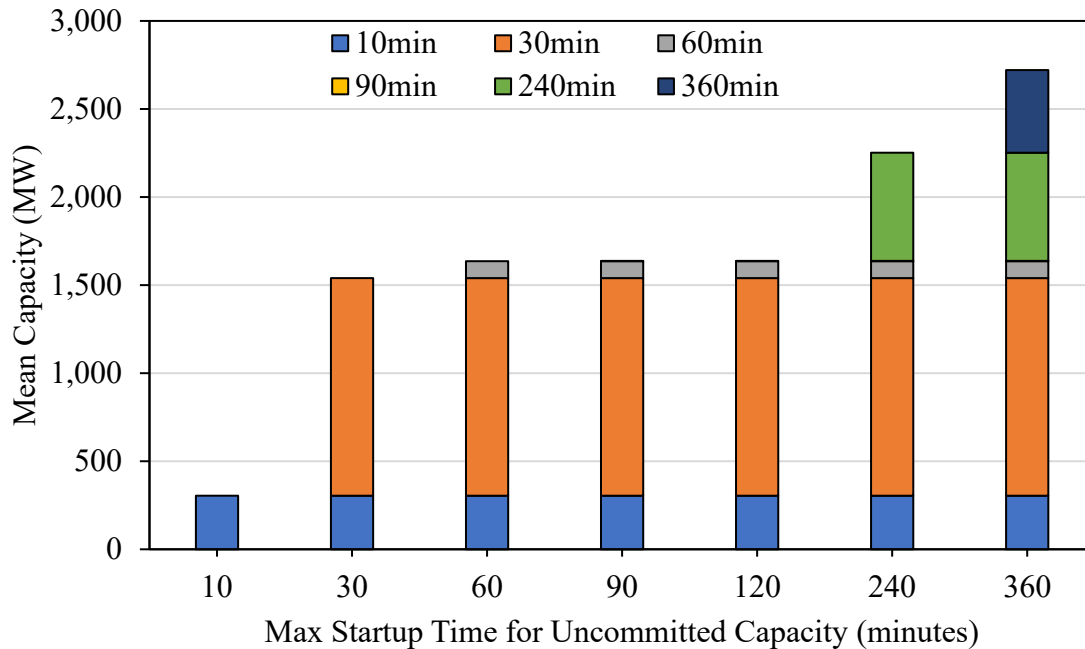**Figure 3: Annual duration of PRC less than 3,000 MW between 2016 and 2025**
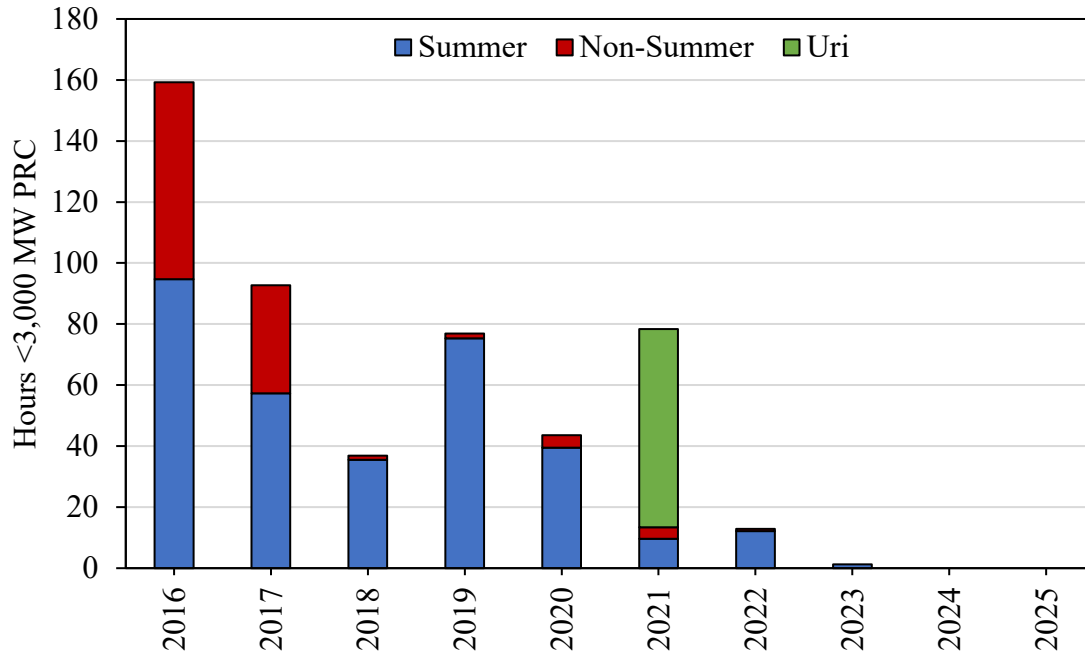
For the five years before Winter Storm Uri, PRC dropped below 3,000 MW for 82 hours on average every year.

**Table 2: RegUp + RRS across June 2023 - December 2024**

Parameter	Value
Min MW	2,403 MW
Max MW	3,889 MW
Mean MW	3,149 MW
Hours >3,000 MW	67%

Figure 4: Resource Availability Across Response Time Intervals


The cumulative capacity of reserves available to start under tight system conditions (PRC < 5,000 MW) includes more than 1,500 MW of capacity capable of starting in an hour or less.

Figure 5: Net Load Forecast Error Across Different Time Horizon Durations
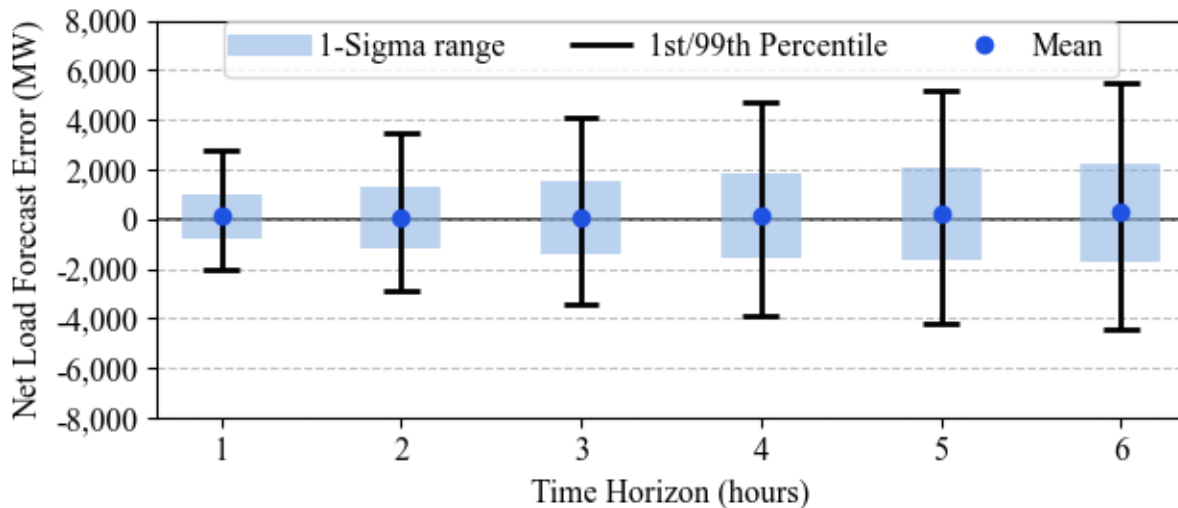
Forecast error risk is nearly twice as high six hours ahead as one hour ahead, ignoring the likelihood of self-commitment from quick-start resources.

Figure 6: RUC Commitments Between 2016 and 2024
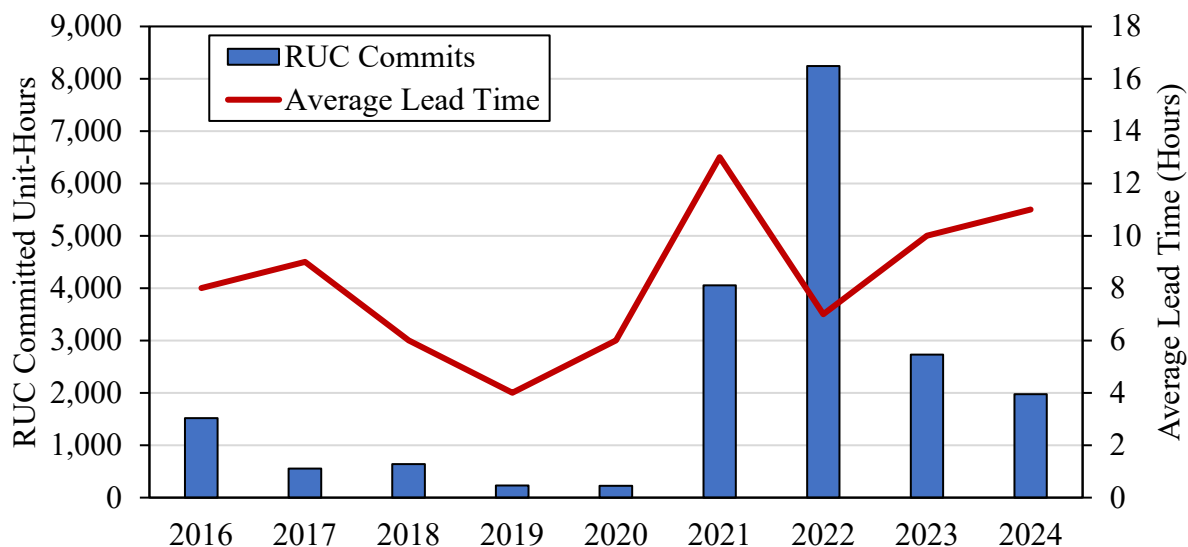
After Winter Storm Uri in 2021, the frequency of RUC commits increased substantially, and the average RUC instruction was given further in advance of real-time.