Determining Probabilistic Outcomes for Data Center Load Growth in the Southeast Using Monte Carlo Modeling

Kavin Manickaraj[1], Youngsun Baek[1], Amy Sharma[2], Etan Gumerman[1]
[1]Greenlink Analytics, [2]Science For Georgia

ABSTRACT

Across the Southeastern U.S., utilities are proposing more steel-in-the-ground generation projects to meet increased demand forecasts that specifically highlight data center growth as one of its primary drivers (IEEFA 2025). While the utilities are guaranteed cost recovery on capacity projects, the financial risk of overbuilding often passes to the ratepayers. Therefore, it is essential to critically examine the assumptions that go into building data center growth forecasts.

Using a Monte Carlo (MC) method, we performed an uncertainty analysis to offer a reasonable range of data center-specific load growths in the Southeast. Rather than single-point forecasts, as utilities tend to report in their Integrated Resource Plans (IRPs), the MC analysis generates a probabilistic range of outcomes that contextualizes how technology scenarios drive future demand. This allows for an understanding of the likelihoods of specific needs over time. These simulations incorporate a range of variables that capture inherent uncertainties including 1) shifts in domestic and international data center markets, 2) advancements in hardware technology, and 3) improvements in computing algorithms.

Compared with this study's Monte Carlo approach, utility predictions for the expected load growth are consistently higher than the vast majority of simulated forecasts. This discrepancy and the historical evidence of utility forecasting patterns both suggest the need for additional independent perspectives to determine the appropriate capacity needed to ensure grid stability while minimizing ratepayer burdens.

INTRODUCTION

Utilities and pipeline companies across the Southeastern United States are currently proposing a significant, long-term expansion of natural gas infrastructure, citing projected data center growth as a primary driver[1]. Within the vertically integrated utility structures of 'Southeast,' defined for purposes of this analysis as Alabama, Georgia, North Carolina, and South Carolina, these capital investments are financially incentivized, often encouraging the construction of surplus generation capacity as a hedge against service disruptions. However, since at least 2007, several Southeastern utilities have consistently overestimated peak demand growth in their ten-year forecasts.

Figures 1 and 2 show the longitudinal data from Federal Energy Regulatory Commission (FERC) Form 714 filings for peak demand forecasts compared to actual peak demand within the Georgia Power territory between 2007-2016 and between 2017-2023, respectively. The lines represent the actual peak demand as they change year after year while the colored lines show utility growth forecasts filed through FERC Form 714 at different starting years (FERC 2025). Figure 1 shows consistently overestimated peak demand forecasts compared to actual peak demand. However, Figure 2 shows a different pattern for GPC forecasts in 2017 and through the COVID-19 pandemic, aligning more closely with historical load demand until the 2023 load forecast (in pink) which once again projects a much steeper peak demand forecast. Given that such overestimations are both common and financially advantageous for utilities, it is essential to critically examine the assumptions behind current forecasts, particularly those linked to the rapidly evolving data center sector. This historical critique is essential for identifying the potential for "surplus generation capacity" incentives within vertically integrated utility structures that may drive current aggressive infrastructure expansion requests.

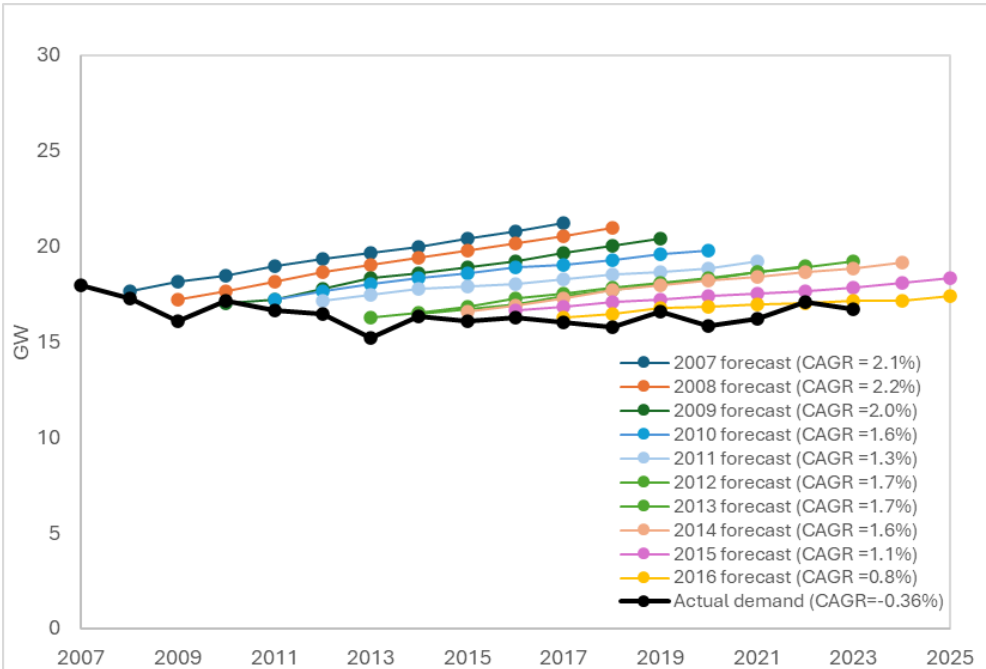

Figure 1. Georgia Power's peak load forecasts (various colors: 2007-2016) compared to actual demand (black line). *Source*: FERC 2025.

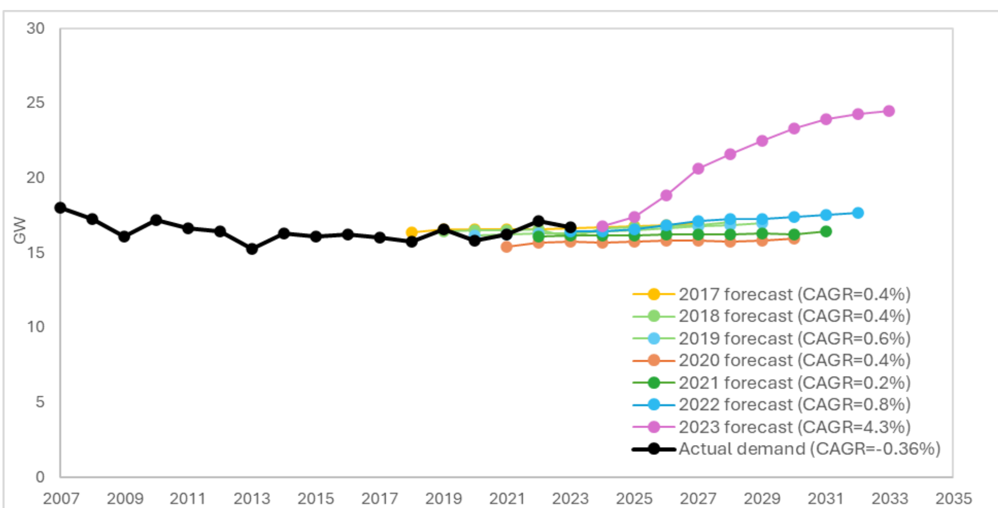

Figure 2. Georgia Power's peak load forecasts (various colors: 2017-2023) compared to actual demand (black line). *Source*: FERC 2025.

Reliably anticipating regional data center load growth is challenging due to inconsistent reporting and gaps in targeted regional research. The lack of transparency and the complexity in a volatile global market situation makes forecasting data center load growth highly variable (LEI 2025). Because international, national, and regional markets are highly interdependent, this study adopts the central assumption that Southeastern markets will track national and global market demand trends. According to the International Energy Agency (IEA), the U.S. market remains a constant proportion of the global data center market from 2024-2030 (IEA 2025); therefore, this analysis assumes that global and domestic growth rates can be viewed interchangeably.

To quantify the inherent uncertainty of these projections, this study utilizes Monte Carlo (MC) simulations to generate a probability distribution of potential outcomes rather than a single-point forecast. This approach provides a statistically robust means to evaluate future demand under uncertainty, mirroring the stochastic modeling techniques often employed by state Public Utility Commissions (PUCs) during Integrated Resource Planning (IRP) processes. This analysis specifically examines the service territories of Duke Energy (DEC & DEP), Georgia Power (GPC), Alabama Power (APC), Santee Cooper (SCPSA), and Dominion Energy South Carolina (DESC) which represent the majority of the annual load across the four-state region[1].

This approach is not intended to serve as the sole predictive engine for the volatile and uncertain future of data center expansion throughout the domestic energy market but rather a tool for challenging the often opaque technical evidence produced by utilities who may be incentivized to meet large load customer demand – primarily from data centers – through preemptive generator capacity expansion.

METHODOLOGY

Two distinct forecasts were developed from the MC simulations to account for varying technological contingencies. The first forecast incorporates a conservative efficiency outlook based on expert-estimated shifts in domestic and global markets, while the second considers more optimistic efficiency gains achieved through advancements in hardware technology and computing algorithms. These independent forecasts are then compared against an aggregation of the utilities' own data center load growth estimates. By prioritizing evidence-based inputs and stochastic modeling, the analytical framework seeks to quantify the range of uncertainty inherent to the rapidly evolving data center sector. The approach is divided into three primary phases: the establishment of a historical baseline through comprehensive data gathering, the development of both conservative and optimistic Monte Carlo simulations to produce load growth trajectories, and the comparison against utility-filed forecasts.

DATA ACQUISITION AND BASELINE ESTABLISHMENT

The Southeast region's 2024–2025 baseline data center capacity – the power and cooling operational requirements – was anchored at 4.3 GW, derived by cross-referencing market intelligence with utility disclosures (S&P 2025; Aterio 2025). This baseline was selected as an intermediate-conservative starting point, positioning it between the more aggressive estimates favored by industry lobbyists and the potentially opaque figures provided in utility filings that

[1] Using the EIA-861 2024 ER data and the most recent Integrated Resource Plan (IRP) documents, the total peak loads across four states and that of the six utilities were compared. It was found that the combined peak load of the six utilities represents approximately 86–90% of the total peak load of the four states. Due to differences in boundary definitions across datasets and literature, as well as variations in reporting timelines and temporal resolution, these figures are presented as best estimates.

may put the figure somewhere between 5.1 and 6.7 GW (Duke Energy 2023; NCUC 2024; Santee Cooper 2024; GPC 2025). It also accounts for the peak power draw of facilities already in operation across the service territories previously listed.

To better assess the regional baseline, global and national datasets were also investigated. Recognizing market interdependence, International Energy Agency (IEA) research showing the U.S. maintains a 43% to 44% share of the global data center market allowed for the interchangeability of growth trends where regional research was scarce. Insights from consultants like CBRE, McKinsey, and the Boston Consulting Group provided additional context for estimating current capacities. Because of the overall limited availability of information, referencing external, third-party sources was valuable in evaluating the likely outcomes in this study's MC approach.

Geographic load growth attribution. In recent years, the potential cumulative growth of data center load has been a recurring topic in state energy forums, committee meetings, and IRP proceedings. Some utilities, such as GPC, have not publicly disclosed the attributions of data centers to load growth in terms of megawatts (MW) in their official forecasts (GPSC 2025).

To estimate expected data center load growth for the Southeast utilities over the next decade, each demand forecast collected for a utility was interpreted against their public statements to estimate data center growth's contributions. The resulting approximate potential cumulative growth attributed to data centers is shown in Table 1.

Table 1. Southeast utilities' cumulative data center capacity forecast (MW)

Year	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035
MW	4.3	5.8	7.4	9.7	11.6	13.0	14.2	14.6	15.0	15.2	15.4

* S&P-anchored 2025 base with approximate utilities' data center growth forecasts.

In GPC's 2025 projections, roughly 80% of new generation is anticipated for data centers (Dunlap 2025); this factor was applied to projected winter peak load growth. Duke Energy (DEC and DEP) estimated 45% of large-load growth is attributable to data centers based on its 2024 supplemental analysis. Santee Cooper projects data centers account for 70–80% of growth, so a 75% factor was applied. For DESC, 2024 IRP updates suggest 256 MW by 2032, while APC signed agreements totaling 1 GW of capacity as of May 2025 (Kelly 2025).

These attributions suggest utilities expect at least 10 GW of data center-related growth (a cumulative total of 14.3 GW) by 2031. A recent stipulated agreement with the Georgia Public Service Commission authorized procurement of approximately 10 GW of generation capacity for new large load customers, mainly data centers, just within Georgia Power's territory (GPSC 2025; Kann 2025). Although this renders the original 10 GW of expected growth to be a conservative estimate, this paper's analysis moves forward with the utilities' capacity forecast profile described in Table 1.

Load growth assumptions. According to the International Energy Agency (IEA), the US share of the global datacenter market remains consistent, 43–44% through 2030, allowing global and domestic growth trends to be used interchangeably (IEA 2025). Forecasts were reviewed from the IEA, which provided four global projections split by term, and Lawrence Berkeley National Labs (LBNL), which calculated US growth through 2028 (IEA 2025; Shehabi 2024). Additional forecasts were gathered from Boston Consulting, Enverus Intelligence Research, Goldman

Sachs, McKinsey & Company, and S&P Global (Lee 2023; Enverus 2024; Goldman Sachs 2025; McKinsey 2025a; S&P 2025). Notably, McKinsey's estimates shifted downward from 22% to 13–24% over just seven months, illustrating high market volatility (McKinsey 2024a, 2024b, 2025a). Table 2 summarizes these compound annual growth rates (CAGR). Because demand typically plateaus after 2030, 10-year projections were estimated by halving 5-year growth rates for 2030–2035. This remains likely to overestimate growth, as IEA models typically show growth slowing by at least two-thirds after 2030.

Table 2. Data center load growth estimates in domestic and global markets

Forecast Name	2024-2030 CAGR	2030-2035 CAGR	Modeling Location
IEA Base	15%	5%	Global
IEA Lift Off	20%	6%	Global
IEA High Efficiency	11%	4%	Global
IEA Headwinds	8%	1%	Global
LBNL (low)	13%	6.7%*	US
LBNL (high)	27%	13.5%*	US
Boston Consulting (2023)	15%	7.5%*	US
Enverus (flat)	7%	7%	US
Enverus (curved)	15%**	6%**	US
Goldman Sachs	13%	6.5%*	Global
Goldman Sachs (Conservative)	11%	5.5%*	Global
McKinsey (constrained)	13%	6.5%*	Global
McKinsey (sustained)	18%	9%*	Global
McKinsey (accelerated)	24%	12%*	Global
S&P Global	18%	9%	US

* Estimated by halving the previous range's estimates.
** The Enverus report stated an estimated demand for 2050, which was used to generate a flat CAGR from 2024 until 2050. A growth rate of approximately double the flat CAGR was assumed from now until 2030 and then halved after 2030.
† Note: When analyzing the difference in growth rates between the earlier and later years, significant variation across sources was observed. Based on a comparative analysis, the CAGR in the later years ranged from roughly one-tenth to one-half of the earlier years' rate. To incorporate a conservative assumption in the MC simulations, the CAGRs for later years was assumed to be half of the earlier years'.

PROBABILISTIC MODELING AND MONTE CARLO SIMULATION

The core framework is a Monte Carlo (MC) simulation, a stochastic technique executing 100,000 iterations to generate a probability distribution. For each iteration, the model randomly sampled two CAGRs from Table 2—applied to 2024–2030 and 2030–2035—anchored to the 4.3 GW starting capacity. These simulations ensure statistically robust results capturing "tail risks," mirroring stochastic modeling required by PUCs for fuel and weather variability but applied here to data center demand.

A unique feature is the explicit modeling of efficiency gains. The "Conservative Efficiency" case assumes moderate technology advancements already baked into the annual growth rate samples, while the "Optimistic Efficiency" case incorporates aggressive hardware breakthroughs—like co-packaged optics and liquid cooling—and algorithmic optimizations.

Isolating these variables tests how innovation might offset physical facility expansion. This modeling assumes that energy efficiency measures are uniformly applied across data center operations, and that energy savings translate linearly to data center capacities for a given unit of computational work. In reality, this is likely a more complex function of variables on its own that may be explored in future work. The forecasted outputs for this case serve as lower-bound estimates subject to a few caveats: the effect of commercial viability barriers, the rebound effect (Jevons' Paradox), and limited but potential double-counting – fewer than seven of the fifteen models explicitly account for any technology advancements.

Hardware and software power savings. To incorporate insights from market experts, each MC simulation randomly selects a CAGR from Table 2 and applies it to the 4.3 GW starting point to generate a demand curve. To reflect insights from technological experts, an additional set of simulations, the Optimistic Efficiency Case, are produced to evaluate the effects of three hardware and three software improvements on these forecasts. Some of the market experts' growth projections already incorporate some technological improvements, such as EIA's and LBNL's, while others are less transparent or ignore these effects entirely. The Optimistic Efficiency Case did not explicitly discount the expert growth rates references that already account for this, however, the resulting outcomes have limited double-counting (as stated above) and serve as a lower bound growth forecast. Figure 3 illustrates this decision-tree selection process for the Optimistic Efficiency Case.

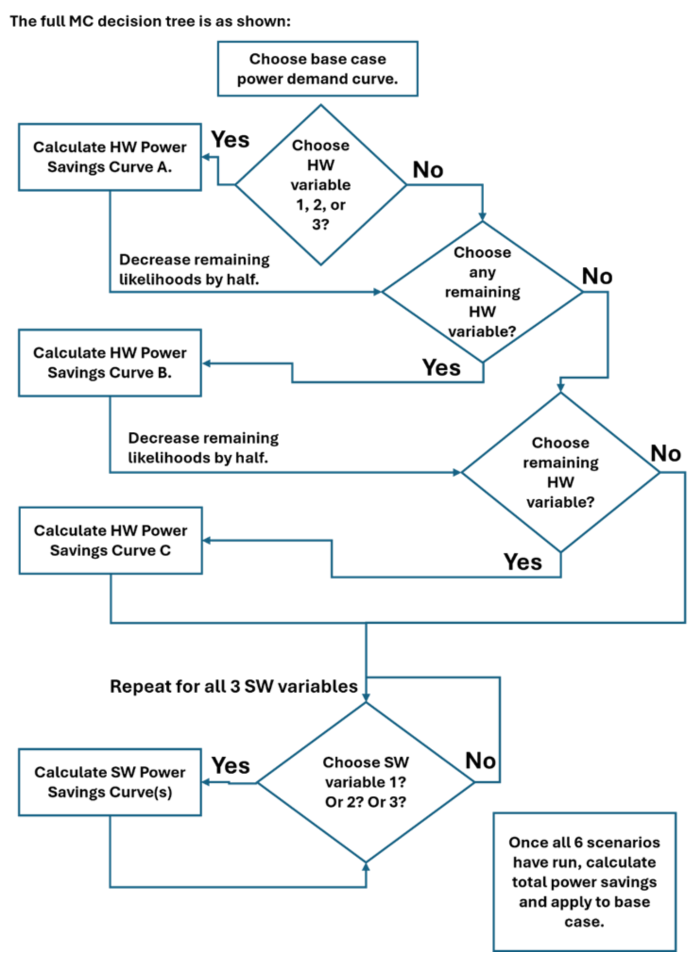

Figure 3. MC simulator full power savings decision tree.

Hardware (HW in Figure 3) and software (SW) groups are independent, but within each, successful implementation reduces the next improvement's probability by 50%. This reflects market dynamics where dominant technologies limit alternative solutions. To accurately model the impact of emerging technologies on overall power consumption, each emerging technology is defined using four key variables:

- Market Adoption Likelihood – Probability that the technology will move beyond the "innovators" stage.
- Start Time – When the technology first enters the market.
- End Time (Full Adoption) – Time required to reach market saturation.
- Impact on Power Consumption – Estimated change in energy use (e.g., 1%, 5%, 100%).

Hardware implementation requires capital and installation, resulting in standard 4-to-6-year refresh cycles (DCK 2022; Horizon 2025). Consequently, hardware not currently in experimental stages is unlikely to impact the market within 10 years; therefore, the start time for modeled technologies—limited to those in experimental phases—is set to 2025. Three innovations—thermal innovations, compute proximity, and optics in networking—are identified in Table 2 for potential power savings, while others were discounted due to negligible impact, double counting, or rebound effects.

Table 2. Hardware technology advancement variables

Hardware Technology	Likelihood	Savings	Start Year	Years to Full Adoption
Thermal Innovations	75%	6% - 36%	2025	8-14
Compute Proximity	40%	2% - 18%	2025	20
Optics in Networking	40%	1.5% - 3.5%	2025	20

According to McKinsey, AI accounted for about 50% of data center workload and is expected to account for about 70% of workload in 2030 (McKinsey 2025a). Exploring more efficient AI algorithms can have a significant impact on data center power draw, because it will have an impact on a large majority of data center workload and consequently data center power draw.

A software change can happen through the introduction of a new product or as a version upgrade to an existing project and can represent either a sea change in how people work or an optimization of an existing system. As such, market adoption speed can vary widely. Herein, three types of software upgrades that will each have their own time-to-market and power savings potentials are outlined in Table 3 and are defined as Established Algorithms, Emerging Algorithms, and Highly Experimental.

Table 3. Software technology advancement variables

Software Algorithm	Likelihood	Savings	Start Year	Years to Full Adoption
Established Algorithms	80%	10% - 22%	2025-2031	2
Emerging Algorithms	30%	11% - 24%	2025-2031	4

Software Algorithm	Likelihood	Savings	Start Year	Years to Full Adoption
Highly Experimental	5%	15% - 30%	2025	12-18

Thermal innovations. As chip density and workloads increase, traditional air cooling is no longer enough to keep computer hardware cool. Companies are utilizing liquid-cooling techniques such as in-rack, direct-to-chip, rear-door heat exchanges, and immersion cooling to keep computing resources cool. These new technologies are already being implemented, and surveys have shown they have reached the "early majority" technology implementation stage and are predicted to reach the "late majority" in the next few years (Alissa et al. 2025; Moore 2025; Vertiv 2023; Weston 2021). This study assumes an energy savings range of 6% to 36% for improvements in thermal innovations based on technology expert assessments (Haghshenas et al. 2022; Kleyman 2025).

Compute proximity. Involved in moving the data storage closer to the computer processors, compute proximity cuts down on moving data back and forth when utilizing and training AI and models. This technology is being tested by companies that can build their own custom hardware and is at the beginning of the market adoption curve (proof-of-concept/innovators stage) (Ali et al. 2022; Derbyshire 2023; Khan et al. 2024; Fay 2024). This study assumes an energy savings range of 2% to 18% for improvements in compute proximity (Falevoz, Yann, and Legriel 2024).

Optics in networking. In addition to data being moved between storage and compute inside data centers, large quantities of data must also be uploaded to and downloaded from data centers. Smoother movement of data at the networking, routing, and switching level translates into energy savings. In the past few years, optics and co-packaged optics have been shown to be more energy efficient and have started to be utilized in data centers. Current adoption is in a small number of hyperscale centers (the proof-of-concept stage) (Chung 2024; McKinsey 2025b). This study assumes an energy savings range of 1.5% to 3.5% for improvement in optics in networking (Tate 2025).

Established algorithms. Software already on the market will integrate power-savings algorithms into their established products as version upgrades. As companies are constantly innovating to stay competitive, these improvements were given a high likelihood of occurring (80%), representing the likelihood that an energy-saving algorithm could appear anytime in the next six years (2025-2031), and that once it is released, it would take 2 years for it to saturate the market (Tang et al. 2020). For example, BERT was released in a paper in Oct 2018 (Devlin et al. 2019), and by Oct 2019 (Nayak 2019), Google began using BERT in its production search algorithms. As AI algorithms grow in popularity, overall energy savings from algorithm improvements would mainly come from improvements to modeling algorithms. Using McKinsey's estimates that 50% of the data center workload is from AI (McKinsey 2025a) and reports that servers are 60% of a data center's power usage (IEA 2025), the overall energy savings would be 30% of the energy savings realized by an algorithmic improvement. A conservative energy savings range is from 1.6x to 3.7x, or 37.5% to 73% (Yang et al. 2017), which is approximately the same as early reporting on DeepSeek's energy savings, which range from 40% to 75% (Rinnovabili 2025). With the previously calculated 30% discount, this translates to 11% to 32% overall energy savings for established algorithms.

Emerging algorithms. AI software that will emerge onto the market as a new product, such as the next AI software system (e.g., a new ChatGPT rival), must first break through into the mainstream market, giving it a lower likelihood of influencing data center energy use compared to existing software. Again, this is an area of active research, and a breakthrough could occur anytime in the next six years, and energy savings could range from 11-24%. Given that it is not already established, a longer market adoption curve of 4 years is assumed as organizations must justify moving to new software. As an example, ChatGPT was commercially launched in Nov 2022, rapidly expanded during 2023, and has reached late majority today, 3 years later. The energy saving for emerging algorithms has been cited to range from 36% to 78% in one source (Mao et al. 2024) to having an upper end of 80% to 95% (Różycki 2025). Taken together, 36% to 80% is a conservative estimate for algorithms that are not yet mainstream but still mentioned in survey papers. Combined with the 30% discount, this translates to 11% to 24% overall savings for emerging algorithms.

Highly experimental. A radical departure from the norm, this is software that is unlike things people have seen before. It disrupts current workflow or creates entirely new workflows and business models. Approximately 95% of all highly experimental products fail (MITPE 2025), and those that succeed will take much longer to saturate the market, as they are based on a perceived new market segment or market demand. For example, Salesforce debuted in about 2000 and is reaching maturity now; Kubernetes debuted in 2015 and is already reaching maturity now. As such, the market adoption time is set between 12 and 18 years, and the start time to 2025 (or it would take too long to make any appreciable impact). Studies mention 100x (99% savings) (Europawire 2025) and 160x (99.3%) (Xu et al. 2024) and even speak of using less than 1 photon per calculation (Wang et al. 2022). Setting the consumption reduction range as 50% to 99% and applying the 30% discount, the power savings range from 15% to 30% for experimental algorithms.

COMPARATIVE BENCHMARKING AND STATISTICAL VALIDATION

The final stage of the methodology involved benchmarking the aggregated utility forecasts against the results of the MC simulations. This comparison was designed to determine where the utilities' projected data center growth of approximately 10 GW (Kann 2025) to a total of 14.3 GW falls within the simulated probability space. By calculating the percentage of simulations that met or exceeded the threshold, the study was able to assign a statistical likelihood to the utilities' planning assumptions. This validation process is essential for identifying whether proposed infrastructure investments are based on probable demand or on extreme, low-probability scenarios that risk creating stranded assets and increasing costs for the region’s ratepayers. Through this rigorous comparative analysis, the methodology provides a framework for regulators to evaluate the necessity of massive methane gas infrastructure expansions in the Southeast.

RESULTS

The results of this study's probabilistic modeling reveal a significant divergence between independent market-based projections and the load growth forecasts currently being utilized by Southeastern utilities to justify infrastructure expansion. Starting from a 2025 baseline of 4.3 GW for regional data center demand, the MC simulations provided a distribution of 100,000 potential outcomes for the subsequent five-to-six-year period. In the "Conservative Efficiency" case seen in Figure 4, the simulations produced a 75% probability range (p12.5 to p87.5) for a load growth between 6.7-11 GW (adding 2.4-6.7 GW, respectively) by 2031 as described in Table 4. This range reflects the most statistically probable trajectory for the region based on a synthesis of fifteen expert growth forecasts in Table 1.

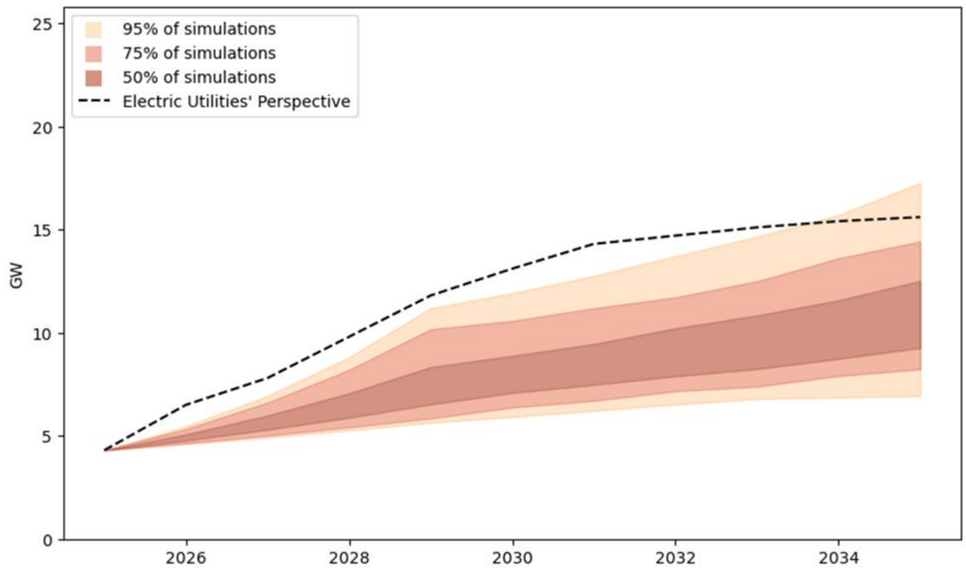

Figure 4. Conservative Efficiency simulation data center load forecast in the Southeast compared to utilities' forecast.

Table 4. Data center capacity (MW) forecasts in Conservative Efficiency by percentile

Percentile	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035
p97.5	4.3	5.5	6.9	8.8	11.2	11.9	12.7	13.6	14.7	15.7	17.3
p87.5	4.3	5.3	6.6	8.2	10.2	10.6	11.0	11.7	12.5	13.6	14.4
p75	4.3	5.1	6.0	7.1	8.3	8.9	9.5	10.1	10.8	11.6	12.5
p25	4.3	4.8	5.3	5.9	6.5	7.1	7.5	7.9	8.2	8.7	9.3
p12.5	4.3	4.6	5.0	5.4	5.9	6.4	6.7	7.2	7.4	7.9	8.2
p2.5	4.3	4.6	4.9	5.3	5.6	5.9	6.2	6.5	6.8	6.9	6.9

When benchmarked against these results, the aggregated growth forecast of approximately 10 GW to a total of 14.3 GW by 2031 provided by regional utilities appears as an extreme statistical outlier. The utility-projected growth exceeds 99.7% of all simulations in the conservative efficiency case – in the 100,000 iterations of the model, the total load exceeded the utility threshold in only 218 instances. Statistically, this indicates that utility planning is predicated on a future with a 1-in-500 likelihood of occurring, a probability of less than 0.22%. This implies the market would need to achieve and sustain growth rates that fall well outside the consensus of technology and market analysts.

The “Optimistic Efficiency” case seen in Figure 5 suggests that the capacity growth will be even more moderate. In this scenario, significant expansion in data center capacity and computing power could continue while the associated electricity load growth is largely offset by technological evolution. Seen in Table 5, the simulations produced a 75% probability range (p12.5 to p87.5) for a load growth between 4.7-8.7 GW (adding 0.4-4.4 GW, respectively) by 2031. Under these parameters, the gap between independent projections and utility forecasts widens even further – by 2031 the simulated forecasts exceed the 14.3 GW total only 13 out of 100,000 times, or 0.013%. This suggests that current utility models may be failing to account for the rapid pace of demand-side innovation within the technology sector.

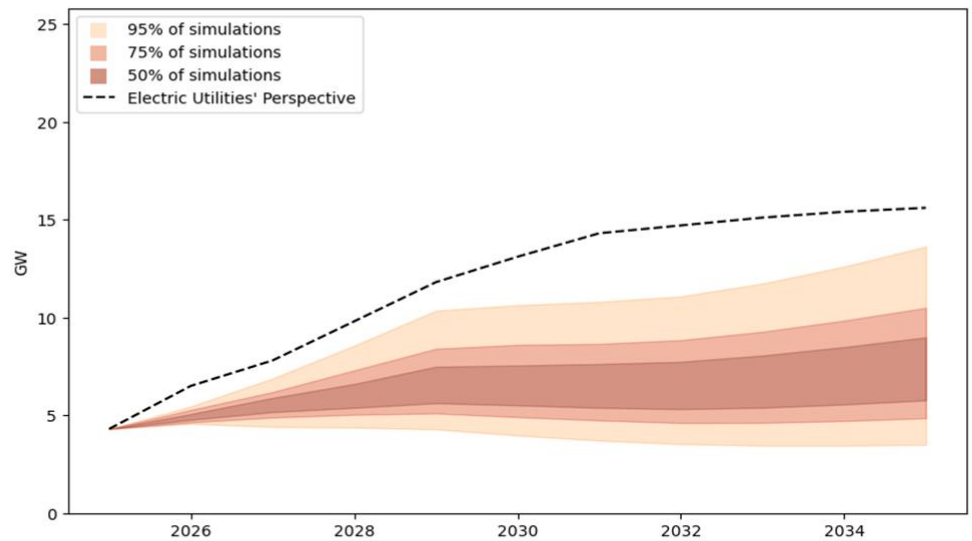

Figure 5. Optimistic Efficiency simulation data center load forecast in the Southeast compared to utilities’ forecast.

Table 5. Data center capacity (MW) forecasts in Optimistic Efficiency by percentile

Percentile	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035
p97.5	4.3	5.5	6.9	8.6	10.4	10.6	10.8	11.1	11.8	12.6	13.6
p87.5	4.3	5.3	6.2	7.3	8.4	8.6	8.7	8.8	9.3	9.8	10.5
p75	4.3	5.1	5.9	6.6	7.5	7.6	7.6	7.7	8.1	8.5	9.0
p25	4.3	4.8	5.2	5.4	5.6	5.5	5.4	5.3	5.4	5.6	5.8
p12.5	4.3	4.6	4.9	5.0	5.1	4.9	4.7	4.6	4.6	4.7	4.9
p2.5	4.3	4.6	4.4	4.4	4.3	4.0	3.7	3.5	3.5	3.5	3.5

DISCUSSION

The discrepancies between the utilities’ forecast and those produced by this probabilistic model highlight an important question: is new capacity expansion justified without data transparency or consensus with multiple third-party forecasts? It is true that utilities’ obligations to meet national reliability standards likely result in forecast inflators that were not considered in this study. However, a comprehensive probabilistic approach accounting for this may address how much financial risk may fall onto ratepayers should there be significant overestimation.

Current reporting opacity also creates a significant information asymmetry between utilities and regulators. Without independent verification, regulators rely on utility-provided data that historically trends toward overestimation. This study's analysis of FERC Form 714 filings confirms that Southeastern utilities have consistently overestimated peak demand since 2007; data center growth may simply be the latest iteration of this bias, amplified by the profile of artificial intelligence.

Finally, while the "rebound effect" is a valid concern, data center hardware advancements could dampen load growth more effectively than utilities acknowledge. Innovations in co-packaged optics, liquid cooling, and model compression have the potential to decouple computing growth from electricity demand. By favoring linear growth models over these variables, utilities risk building infrastructure for demand that technological progress may render obsolete.

CONCLUSION

This study concludes that the Southeastern utilities' data center load forecasts to be mostly improbable when compared to the results of this Monte Carlo forecasting approach using available market data – less than a 0.22% chance under conservative efficiency assumptions and a 0.013% chance under more aggressive efficiency assumptions. Without more transparency into the utilities' methods of forecasting their data center-specific load growth, these results question whether existing steel-in-the-ground expansion plans are commensurate with the uncertainty of data center energy needs even with reliability obligations. In a rapidly evolving landscape, failing to investigate large-load customers projections through various forecasting methods risks making multi-billion-dollar decisions that may result in increased ratepayer burdens.

To mitigate these risks, State Public Utility Commissions could mandate the use of independent modeling—such as, but not limited to, the Monte Carlo techniques employed in this paper—within the Integrated Resource Planning (IRP) process. By requiring utilities to plan for the most probable range of outcomes rather than extreme outliers, regulators can ensure a more resilient, cost-effective, and sustainable energy future for the Southeast. Future research should continue to refine these models as more transparent state-level data becomes available, ensuring that regional infrastructure planning remains grounded in empirical evidence.

REFERENCES

Ali, M., et al. 2022. "Compute-in-Memory Technologies and Architectures for Deep Learning Workloads." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 30 (11): 1615–1630. ieeexplore.ieee.org/document/9899381.

Alissa, H., et al. 2025. "Using Life Cycle Assessment to Drive Innovation for Sustainable Cool Clouds." *Nature* 641: 331–338. doi.org/10.1038/s41586-025-08832-3.

Aterio. 2025. "Data Centers in the United States." Accessed July 16. www.aterio.io/insights/us-data-centers.

Chang, Y. 2024. "Co-Packaged Optics (CPO): Evaluating Different Packaging Technologies." IDTechEx, August 22. www.idtechex.com/en/research-article/co-packaged-optics-cpo-evaluating-different-packaging-technologies/31608.

DCK (Data Center Knowledge). 2022. "Data Center Hardware Refresh Cutback by Microsoft – What's Next?" August 25. www.datacenterknowledge.com/hyperscalers/data-center-hardware-refresh-cutback-by-microsoft-what-s-next-?.

Derbyshire, K. 2023. "Increasing AI Energy Efficiency With Compute In Memory." *Semiconductor Engineering*, November 16. semiengineering.com/increasing-ai-energy-efficiency-with-compute-in-memory.

Devlin, J., et al. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1. au1206.github.io/assets/pdfs/BERT.pdf.

Duke Energy. 2023. *Supplemental Planning Analysis: Carolinas Resource Plan.* Charlotte, NC: Duke Energy. www.duke-energy.com/-/media/pdfs/our-company/carolinas-resource-plan/supplements/supplemental-planning-analysis.pdf.

Dunlap, S. 2025. "Georgia Power's Plan to Support Potential Data Center Surge with Fossil Fuel Energy Faces Scrutiny." *Georgia Recorder*, June 1. georgiarecorder.com/2025/06/01/georgia-powers-plan-to-support-potential-data-center-surge-with-fossil-fuel-energy-faces-scrutiny.

Enverus. 2024. "Returning to Growth: US Power Demand Forecast Highlights Impact of Data Centers, EVs, and Solar." July 16. www.enverus.com/newsroom/returning-to-growth-us-power-demand-forecast-highlights-impact-of-data-centers-evs-and-solar/.

Europawire. 2025. "TUM Researchers Develop 100x Faster Method to Cut Energy Consumption in AI Neural Network Training." Press Release, March 7. news.europawire.eu/tum-researchers-develop-100x-faster-method-to-cut-energy-consumption-in-ai-neural-network-training/eu-press-release/2025/03/07/11/31/01/149862/.

Falevoz, Y., and J. Legriel. 2024. "Energy Efficiency Impact of Processing in Memory: A Comprehensive Review of Workloads on the UPMEM Architecture." *Lecture Notes in Computer Science* 14352. doi.org/10.1007/978-3-031-48803-0_13.

Fay, M., et al. 2016. "Disentangling the Relationship Between the Adoption of In-Memory Computing and Firm Performance." In *Proceedings of the 24th European Conference on Information Systems*. Istanbul, Turkey.

FERC (Federal Energy Regulatory Commission). 2025. "FERC eForms Submission History." Accessed March. ecollection.ferc.gov/submissionHistory.

GPC (Georgia Power Company). 2025. *2025 Integrated Resource Plan.* Atlanta: Georgia Power. www.georgiapower.com/content/dam/georgia-power/pdfs/company-pdfs/2025-Integrated-Resource-Plan.pdf.

GPSC (Georgia Public Service Commission). 2025. *Stipulated Agreement in re: Georgia Power Company's Application for the Certification of Capacity from the 2029–2031 All-Source RFP and Supplemental Resources.* Docket Nos. 56298 & 56310. December 10. psc.ga.gov/search/facts-document/?documentId=224772.

Goldman Sachs. 2025. "AI to Drive 165% Increase in Data Center Power Demand by 2030." February 4. www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030.

Haghshenas, K., et al. 2022. "Enough Hot Air: The Role of Immersion Cooling." *arXiv* preprint, May 9. arxiv.org/abs/2205.04257.

Hering, G., and S. Dlin. 2025. "US Datacenter Power Draw to Double by 2028; States Tackle Supply Cost, Supply Concerns." *S&P Global Online*, July 10. www.spglobal.com/market-intelligence/en/news-insights/articles/2025/7/us-datacenter-power-draw-to-double-by-2028-states-tackle-cost-supply-concerns-91382267.

Horizon (Horizon Technology). 2025. "Navigating Hardware Refresh Cycles in the Data Center." September 9. horizontechnology.com/news/data-center-hardware-refresh-cycles/.

IEA (International Energy Association). 2025. *Energy and AI*. Paris: IEA. www.iea.org/reports/energy-and-ai.

IEEFA (Institute for Energy Economics and Financial Analysis). 2025. *Data Centers Drive Buildout of Gas Power Plants and Pipelines in the Southeast.* January. ieefa.org/resources/data-centers-drive-buildout-gas-power-plants-and-pipelines-southeast.

Kann, D. 2025. "Georgia Power, PSC staff strike deal for $16B expansion to power data centers." *The Atlanta Journal Constitution*, December 10. www.ajc.com/news/2025/12/georgia-power-psc-staff-strike-deal-to-allow-historic-data-center-expansion/.

Kelly, B. 2025. "Analysis: Where to Colocate a Data Center with an Underutilized Power Plant." *HData Blog*, May 7. blog.hdata.com/data-center-power-plant-colocation.

Khan, A. A., et al. 2024. "The Landscape of Compute-Near-Memory and Compute-in-Memory: A Research and Commercial Overview." *arXiv* preprint, January 24. arxiv.org/pdf/2401.14428v1.

Kleyman, B. 2025. "Liquid Cool: A Year in Review." *Data Center Frontier*, February 18. www.datacenterfrontier.com/sponsored/article/55266938/liquid-cooling-a-year-in-review.

Lee, V. 2023. "The Impact of GenAI on Electricity: How GenAI is Fueling the Data Center Boom in the U.S." Boston Consulting Group, September 13. www.linkedin.com/pulse/impact-genai-electricity-how-fueling-data-center-boom-vivian-lee/.

LEI (London Economics International LLC). 2025. *Uncertainty and Upward Bias Are Inherent in Data Center Electricity Demand Projections.* Southern Environmental Law Center. www.selc.org/wp-content/uploads/2025/07/LEI-Data-Center-Final-Report-07072025-2.pdf.

Mao, Y., et al. 2024. "Green Edge AI: A Contemporary Survey." *Proceedings of the IEEE* 112 (7): 880–911. ieeexplore.ieee.org/document/10637271.

McKinsey & Company. 2024a. "How Data Centers and the Energy Sector Can Sate AI's Hunger for Power." September 17. www.mckinsey.com/industries/private-capital/our-insights/how-data-centers-and-the-energy-sector-can-sate-ais-hunger-for-power#/.

———. 2024b. "AI Power: Expanding Data Capacity to Meet Growing Demand." October 29. www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-power-expanding-data-center-capacity-to-meet-growing-demand.

———. 2025a. "The Cost of Compute: A $7 Trillion Race to Scale Data Centers." April 28. www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-cost-of-compute-a-7-trillion-dollar-race-to-scale-data-centers.

———. 2025b. "Opportunities in Networking Optics: Boosting Supply for Data Centers." June. www.mckinsey.com/~/media/mckinsey/industries/technology%20media%20and%20telecommunications/high%20tech/our%20insights/opportunities%20in%20networking%20optics%20boosting%20supply%20for%20data%20centers/opportunities-in-networking-optics-boosting-supply-for-data-centers.pdf.

MIT Professional Education. n.d. "Why 95% of New Products Miss the Mark (and How Yours Can Avoid the Same Fate)." Accessed October 1, 2025. professionalprograms.mit.edu/blog/design/why-95-of-new-products-miss-the-mark-and-how-yours-can-avoid-the-same-fate/.

Moore, M. 2025. "Overheating Is a Big Problem for AI Hardware as Demand Rises - and Dell Thinks It Might Have the Answer." *Tech Radar*, May 26. www.techradar.com/pro/overheating-is-a-big-problem-for-ai-hardware-as-demand-rises-and-dell-thinks-it-might-have-the-answer.

Nayak, P. 2019. "Understanding Searches Better Than Ever Before." *Google Blog*, October 25. blog.google/products/search/search-language-understanding-bert/.

(NCUC) North Carolina Utilities Commission. 2024. *In the Matter of: Biennial Consolidated Carbon Plan and Integrated Resource Plans of Duke Energy Carolinas, LLC, and Duke Energy Progress, LLC.* Hearing Transcript, vol. 24. August 5. starw1.ncuc.gov/NCUC/ViewFile.aspx?Id=b3f65f27-eaba-4a2e-aa69-00c2d190bf7a.

Rinnovabili. 2025. "How Energy-Efficient Is DeepSeek, China's AI Disruptor?" January 29. www.rinnovabili.net/business/markets/deepseeks-energy-consumption-ais-75-power-cut/.

Różycki, R., D. A. Solarska, and G. Waligóra. 2025. "Energy-Aware Machine Learning Models—A Review of Recent Techniques and Perspectives." *Energies* 18 (11): 2810. doi.org/10.3390/en18112810.

S&P (S&P Global Market Intelligence). 2025. "451 Research Datacenter KnowledgeBase." Accessed June. www.spglobal.com/market-intelligence/en/solutions/datacenter-knowledgebase.

Santee Cooper. 2024. *2024 Integrated Resource Plan Update.* Public Service Commission of South Carolina. September 16. www.santeecooper.com/About/Integrated-Resource-Plan/Reports-and-Materials/Santee-Cooper-2024-IRP-Update.pdf.

Shehabi, A., et al. 2024. *2024 United States Data Center Energy Usage Report.* Berkeley, CA: Lawrence Berkeley National Lab (LBNL). eta-publications.lbl.gov/sites/default/files/2024-12/lbnl-2024-united-states-data-center-energy-usage-report_1.pdf.

Tang, X., et al. 2020. "The Pace of Artificial Intelligence Innovations: Speed, Talent, and Trial-and-Error." *Journal of Informetrics* 14 (4): 101094. doi.org/10.1016/j.joi.2020.101094.

Tate, G. 2025. "Photonics Speeds Up Data Center AI." *Semiconductor Engineering*, May 1. semiengineering.com/photonics-speeds-up-data-center-ai/.

Vertiv. 2023. "Quantifying the Impact on PUE and Energy Consumption When Introducing Liquid Cooling Into an Air-Cooled Data Center." February 15. www.vertiv.com/en-emea/about/news-and-insights/articles/blog-posts/quantifying-data-center-pue-when-introducing-liquid-cooling/.

Wang, T., et al. 2022. "An Optical Neural Network Using Less Than 1 Photon Per Multiplication." *Nature Communications* 13 (1): 123. www.nature.com/articles/s41467-021-27774-8.

Weston, S. 2021. "Microsoft Is Submerging Servers in Boiling Liquid to Prevent Teams Outages." *IT Pro*, April 7. www.itpro.com/server-storage/data-centres/359129/microsoft-submerges-servers-in-boiling-liquid-to-prevent-teams.

Xu, T., et al. 2024. "Control-Free and Efficient Integrated Photonic Neural Networks via Hardware-Aware Training and Pruning." *Optica* 11 (8): 1039–1049. opg.optica.org/optica/fulltext.cfm?uri=optica-11-8-1039.

Yang, T., et al. 2017. "Designing Energy-Efficient Convolutional Neural Networks Using Energy-Aware Pruning." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. arxiv.org/abs/1611.05128.

AI-assisted tools used: Google Gemini
Used for: Shortening report length; Organizing references section.
Extent: Light editing and applying references format to all sources already assembled by human authors.
Human review: The authors reviewed and verified all content, calculations, and citations.