Forecasting Natural Gas Demand

What weather, home heating, and electricity use tell us about future gas use.

Data through: August 2026 (gas demand)
Model results: November 1, 2026
Write-up: November 4, 2026

1. Introduction

Gas demand means the amount of natural gas people and businesses use. Cold weather increases the need for heating. Hot weather increases the need for electricity to run air conditioning. How much extra gas is needed depends on the heating fuels and power plants in each place.

I tested whether three changes could make monthly gas-use forecasts more accurate: giving states different importance in the weather average, making forecasts for individual states, and predicting electricity production before predicting gas use. I also compared my forecasts with saved forecasts from the U.S. Energy Information Administration (EIA) and explored whether electricity sales could help identify data-center demand.

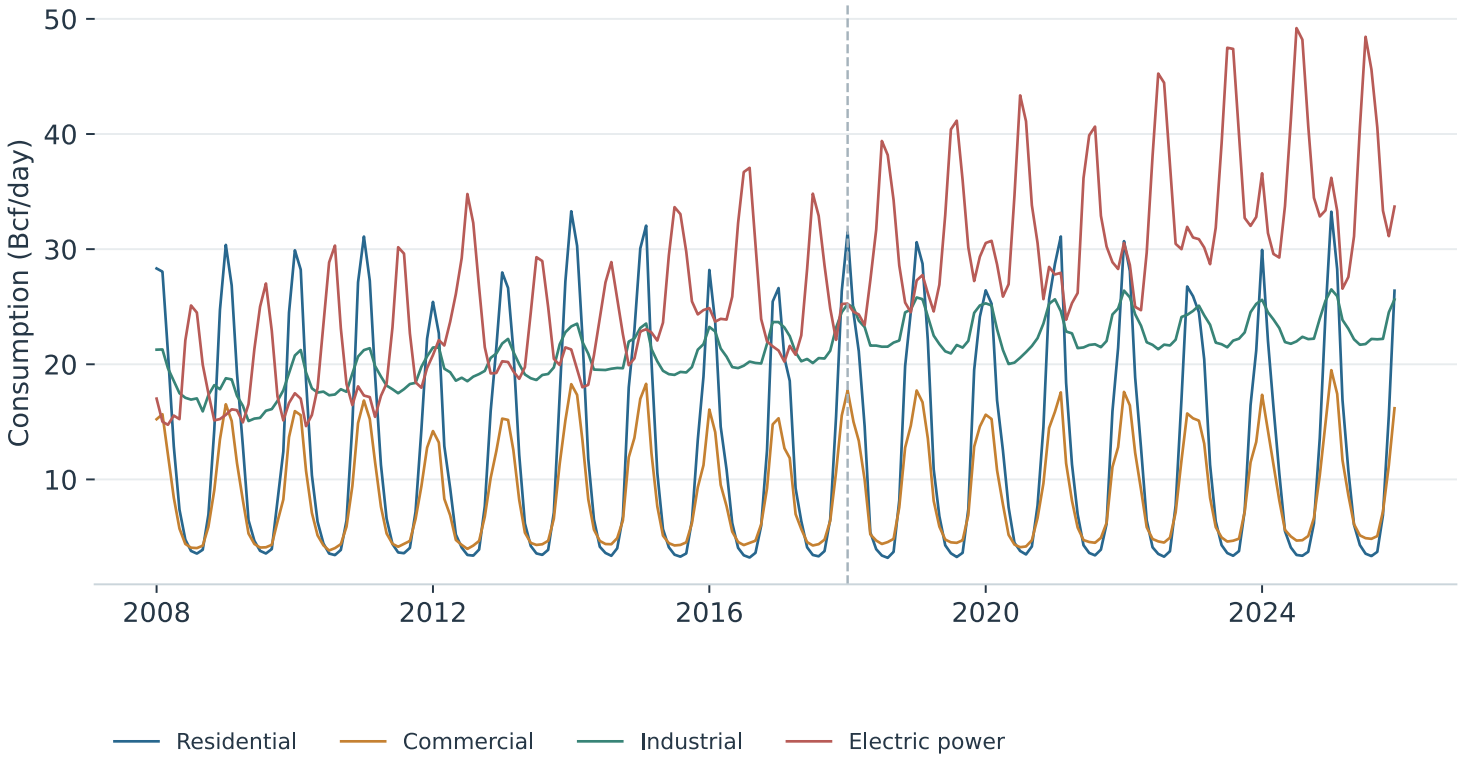

Figure 1. Four sectors, different demand patterns. Monthly gas use in the 48 connected states, measured in billion cubic feet per day (Bcf/day). The dashed line starts the forecast test in January 2018. Power-plant use grows over time; home use rises each winter. Chart data.

I study four groups of gas users (sectors):

- **Residential:** homes that use gas for heating, hot water, cooking, and other household needs.
- **Commercial:** buildings such as offices, shops, schools, and hospitals.
- **Industrial:** factories and other industrial users.
- **Electric power:** power plants that burn gas to make electricity.

The four-sector total is the sum of these uses. It leaves out gas used in pipelines, at gas wells and processing plants, and in vehicles. It therefore does not represent every use of natural gas in the country.

2. Data

How gas use is measured

The main study covers the 48 connected U.S. states and excludes Washington, DC. I divide each month's gas volume by its number of days. The result is billion cubic feet per day (Bcf/day), a daily rate that lets a short month and a long month be compared fairly.

Where every state reports a value, I add the states together. Some industrial values are withheld. In those cases, I use EIA's published U.S. total and subtract Alaska, Hawaii, and DC. Where both methods can be checked, their totals differ by at most 7 million cubic feet.

A missing value stays missing. I enter zero only when the official table marks a value as "Not Applicable."

The electricity data cover utility power plants and independent power producers. They include plants that make both electricity and useful heat, called combined heat and power plants. They exclude generators counted in the commercial and industrial sectors and small-scale solar. I check the electricity totals against regional records. Electricity sales measure what customers buy, but that electricity may have been produced in another state.

Measuring weather and deciding which states matter most

The National Oceanic and Atmospheric Administration (NOAA) provides degree-day data. Degree days summarize heating and cooling needs using 65°F as a reference. A heating degree day counts how far a day's average temperature is below that reference. A cooling degree day counts how far it is above. Larger totals mean more need for heating or cooling.

A national weather average needs rules for how much each state contributes to the average, which we can represent using weights. Population weights give more importance to states with more people. Gas-heated-home weights give more importance to states with more homes heated by utility gas. The second rule may better reflect where cold weather will raise gas use.

NOAA already uses 2010 population to combine weather observations *within* each state. I change only how states are combined with one another. There is one small mismatch: Maryland's weather area includes DC, but my gas-use total excludes DC.

I take annual population and gas-heated-home counts from the Census Bureau's American Community Survey (ACS). Gas-heated homes here means occupied homes heated with utility gas delivered through pipelines. It excludes bottled and tank gas, including liquefied petroleum (LP) gas such as propane. The series begins in 2006. Standard 2020 estimates were not released, so I carry the latest available estimate forward.

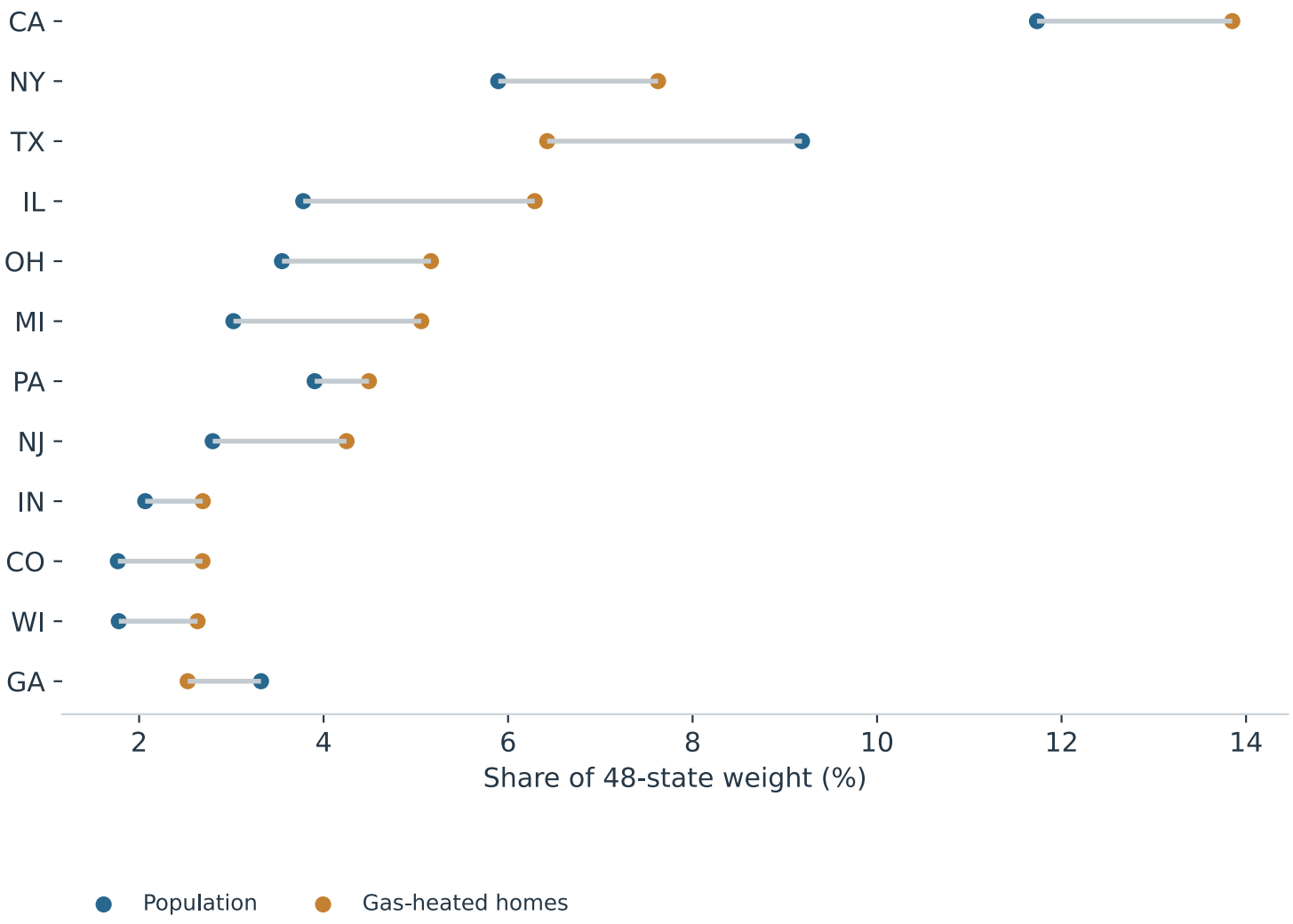

Figure 2. States contribute differently to the heating-weather average. The twelve largest shares of gas-heated homes in January 2025. A larger weight gives a state more influence on the national weather average. These ACS counts follow the two-year reporting rule. NOAA already combines weather within each state. This study changes how states are combined. Chart data.

I also try EIA’s counts of gas customers. A customer account is not necessarily a building, and a change of supplier can cause double counting. Home-heating counts are an imperfect measure for business heating too. They are a stand-in, rather than a direct count of gas-heated commercial buildings.

What information a forecast is allowed to use

Each forecast is dated the first day of the month it predicts. To allow for reporting delays, I use gas and electricity observations at least three months old and Henry Hub prices at least two months old. Annual state weights become available in the model no earlier than January two years after the year they describe.

For weather, the forecast uses the average of the previous 20 occurrences of that calendar month. Model fitting begins with January 2008. I score the forecasts on all 96 months from 2018 through 2025.

The historical data include revisions made after their first release. My delay rules reduce the risk of using information too early, but they cannot recreate exactly what a forecaster knew at every past date.

3. Models

The simple forecast and the fitted models

The benchmark is the previous year’s gas use per day for the same calendar month. It already captures much of the winter heating and summer cooling pattern. The charts call it seasonal naive.

The main national model uses earlier gas use, recent growth, the month of the year, longer-term trends, Henry Hub prices, and heating and cooling degree days. Its power-sector version also uses earlier electricity growth and the shares of generation supplied by gas, coal, and renewables.

This model is ridge regression. For home and business demand, I use gas-heated-home weather weights. For industrial demand, I use population weights. For power

demand, I use each state's electricity production in the same month of the previous year.

The state model, labeled "State panel" in the charts, keeps some common patterns but lets states have different demand levels and weather responses. I add the 48 state forecasts within each sector to get that sector's total. I also test gradient boosting. I recorded the settings before running the full historical test and did not tune them on the evaluation years.

Predicting power-plant gas use in two steps

The direct model predicts power plants' gas use in one step. The two-stage model first predicts how much electricity will be generated and what share will come from gas. It then uses the previous 12 available months to estimate how much gas is needed for each unit of gas-generated electricity.

> Power-plant gas use ≈ electricity produced per day × share produced with gas × gas needed per unit of gas-generated electricity.

The model limits the share and the conversion rate to avoid unrealistic estimates. This approach follows a physical link between electricity and fuel use, but a mistake at either stage can affect the final answer. Fuel costs, outages, hydroelectricity, renewables, and electricity transfers between states also influence which plants run.

How I check whether a model helps

I refit the models each month using earlier observations, including the weather already known for those earlier months. The main forecasts use normal weather for the month being forecast. A separate check gives the model the weather that actually happened. The saved results contain 3,744 predictions for the 48-state sector totals and 36,864 state predictions.

Mean absolute error (MAE) is the average size of a forecast miss, ignoring whether the forecast was too high or too low. Here it is measured in Bcf/day. The download also includes root mean square error (RMSE), which gives more weight to large misses, along with results for individual years, seasons, and states.

I planned nine main comparisons of weather weights, state models, and the two-stage power model. A smaller average error alone is not enough to call a change reliable. I also check how much the estimated improvement could vary if a different set of months had been observed.

To estimate that uncertainty, I resample matched forecast errors in 12-month blocks. This keeps groups of neighboring months together, because their errors can be related. The resulting 95% interval shows uncertainty about the average improvement. If it includes zero, both an improvement and no improvement remain consistent with that calculation.

The tests also account for related errors across time and for trying nine comparisons. The adjustment is called the Holm method. Its adjusted p-value is a test score, where a smaller value gives stronger evidence against equal forecast accuracy, under the test's assumptions. I count a gain as supported only when the average gain is positive, the whole interval is above zero, and the adjusted p-value is below 0.05. These rules reduce the chance of treating a lucky result as a useful improvement.

I also change the training window and ridge penalty to see how sensitive the answers are to those choices.

A prediction band is a range around a forecast, rather than a single guess. I build it from at least 24 earlier forecast errors whose true outcomes would already have been reported. The total-demand band uses errors for the combined total, because errors in different sectors can move together. Coverage is the percentage of later observations that fall inside the band. I report it instead of assuming the band works as intended.

4. Results

The national model improves on repeating last year

Average forecast miss (MAE), in Bcf/day, over 96 months in 2018–2025.

Sector	Last-year MAE	National-model MAE	State-model MAE	Boosting MAE	National improvement

Sector	Last-year MAE	National-model MAE	State-model MAE	Boosting MAE	National improvement
Residential	1.58	1.29	1.41	1.16	+18.4%
Commercial	0.81	0.70	0.77	0.63	+13.9%
Industrial	0.60	0.49	0.49	0.67	+18.8%
Electric power	2.11	1.58	2.00	2.61	+24.9%

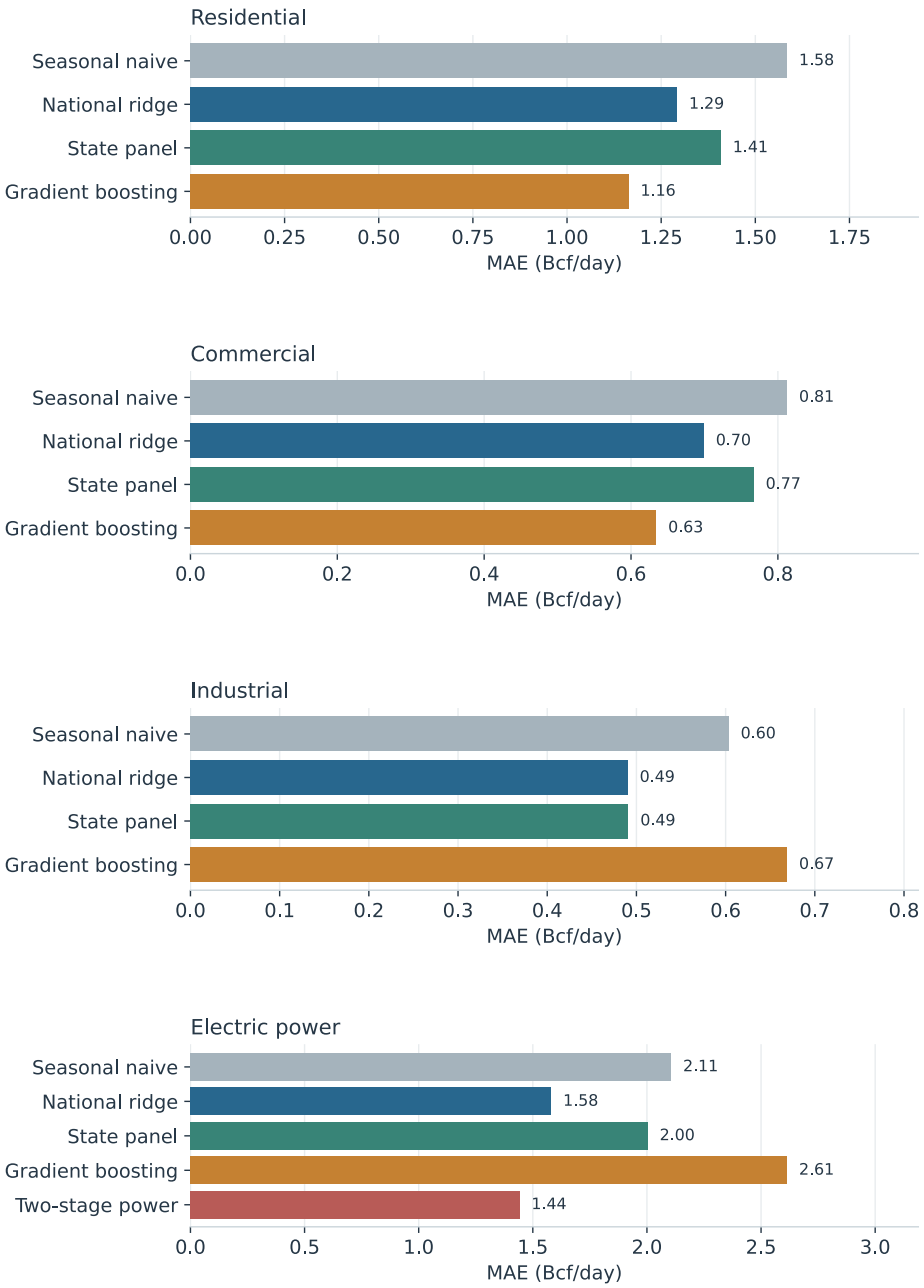

Figure 3. How far each model misses the actual demand. Each model is tested on the same 96 months, 2018–2025. The fitted models use normal weather. “Seasonal naive” repeats last year’s same-month demand. MAE is the average size of the miss, so shorter bars are better. Chart data.

Compared with the same-month-last-year benchmark, the main national model lowers MAE by 18.4% for homes, 13.9% for businesses, 18.8% for industry, and 24.9% for power plants. A positive gain means a smaller average miss. These comparisons describe how much the national model improves on a simple starting point. The nine planned tests below ask which specific model changes help.

Giving more importance to gas-heated homes helps

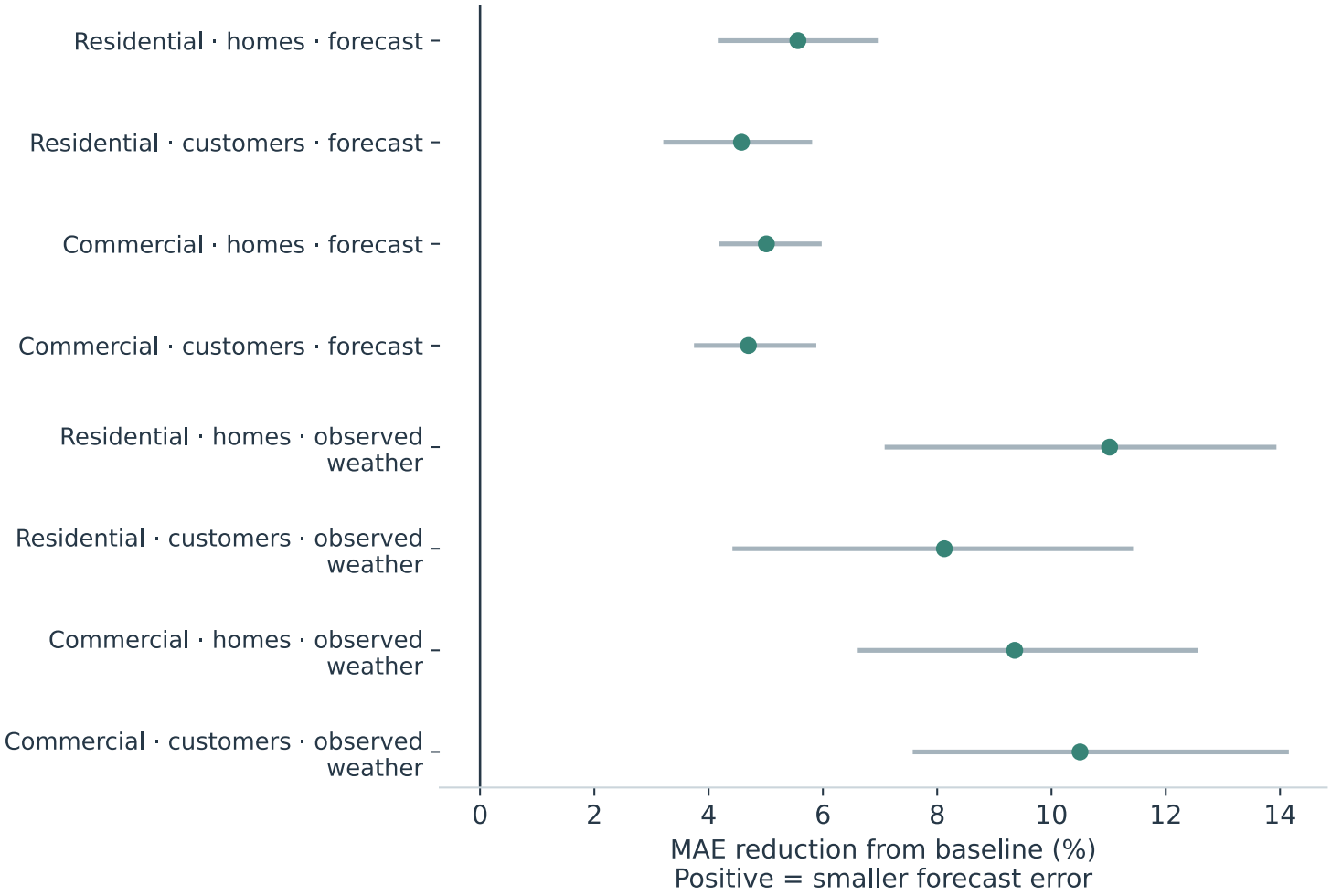

Figure 4. Heating-based weather weights lower forecast errors. The dots show how much average error falls compared with population weights. Positive values mean smaller errors. Gray lines show 95% uncertainty intervals, calculated from matched errors resampled in 12-month blocks. Results labeled “observed weather” use weather known afterward. The table below also accounts for testing several comparisons. Chart data.

Switching from population weights to gas-heated-home weights lowers home-use MAE by 5.6% and business-use MAE by 5.0%. Using gas-customer counts also helps both sectors. All four weighting comparisons pass the three checks described above. This clearest result is that weather averages should pay attention to where gas is actually used for heating.

The result does not mean the weather measures are completely different. They share much of the same seasonal pattern. The test is whether the smaller differences between them improve the forecasts.

In the table, H1 labels the weather-weight comparisons, H2 labels the state-model comparisons, and H3 labels the two-stage power comparison.

Forecasts using normal weather. A supported gain must be positive, have its whole uncertainty interval above zero, and have an adjusted p-value below 0.05.

Comparison	Average-error reduction	95% interval	Adjusted p-value	Passes all checks?
H1 · Residential · Gas-heated homes	+5.6%	+4.2% to +6.9%	<0.001	Yes
H1 · Residential · Gas customers	+4.6%	+3.2% to +5.8%	<0.001	Yes
H1 · Commercial · Gas-heated homes	+5.0%	+4.2% to +5.9%	<0.001	Yes
H1 · Commercial · Gas customers	+4.7%	+3.8% to +5.8%	<0.001	Yes
H2 · Residential · State panel	-9.0%	-14.3% to -5.7%	<0.001	No
H2 · Commercial · State panel	-9.7%	-15.5% to -6.3%	<0.001	No
H2 · Industrial · State panel	-0.2%	-15.2% to +14.8%	0.986	No
H2 · Electric power · State panel	-26.8%	-56.2% to +1.9%	0.130	No
H3 · Electric power · Two-stage power	+8.8%	+0.7% to +16.7%	0.130	No

Of the nine planned comparisons, 4 pass all three checks for the forecasts. The separate exercise that already knows the weather also has 4 supported gains. The table includes every planned comparison, including changes that raise errors and changes whose gains remain uncertain.

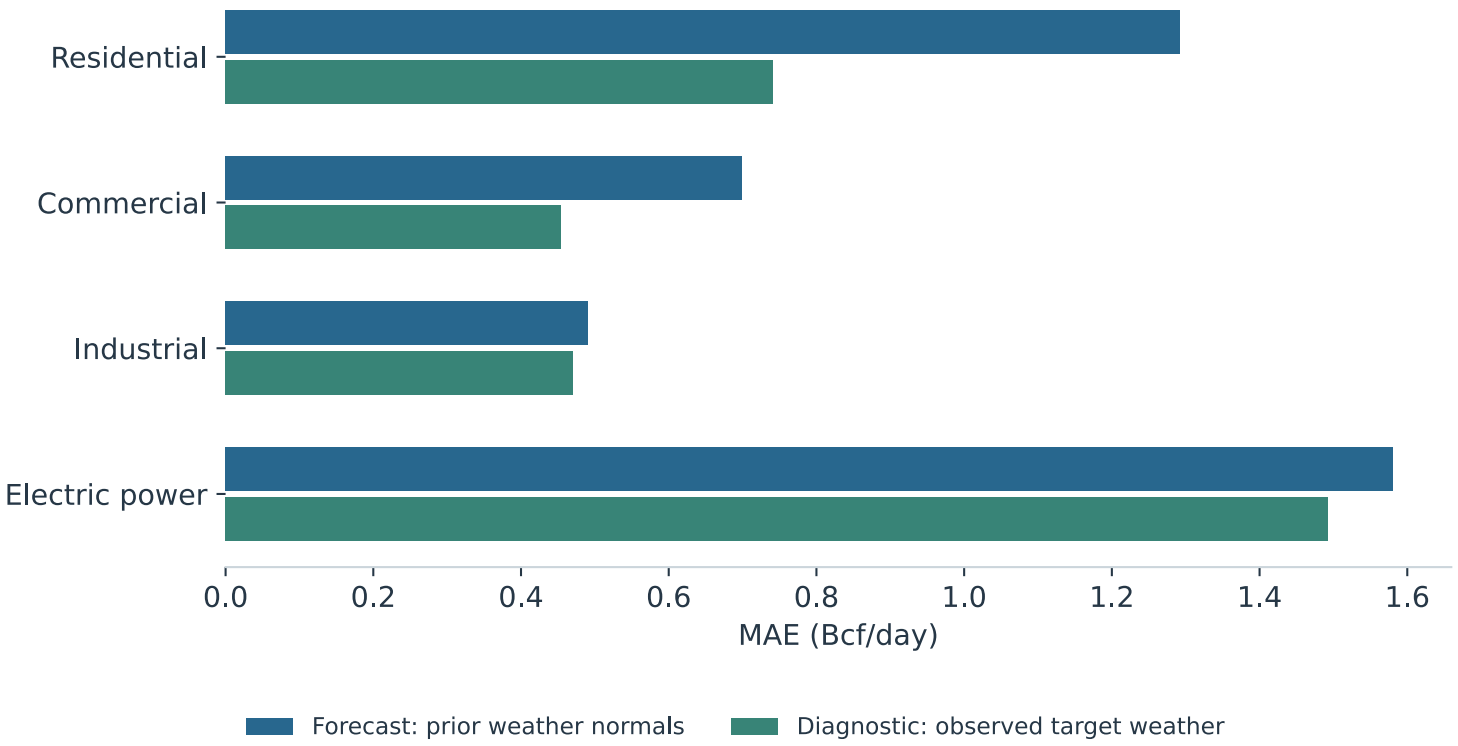

Figure 5. Knowing the actual weather changes the errors. The only input changed is weather in the month being predicted. Blue uses the historical average available at the forecast date; green uses the weather that actually happened. Chart data.

Giving the model the weather that actually occurred lowers MAE by 42.6% for homes and 5.5% for power plants.

Separate state forecasts do not improve the total

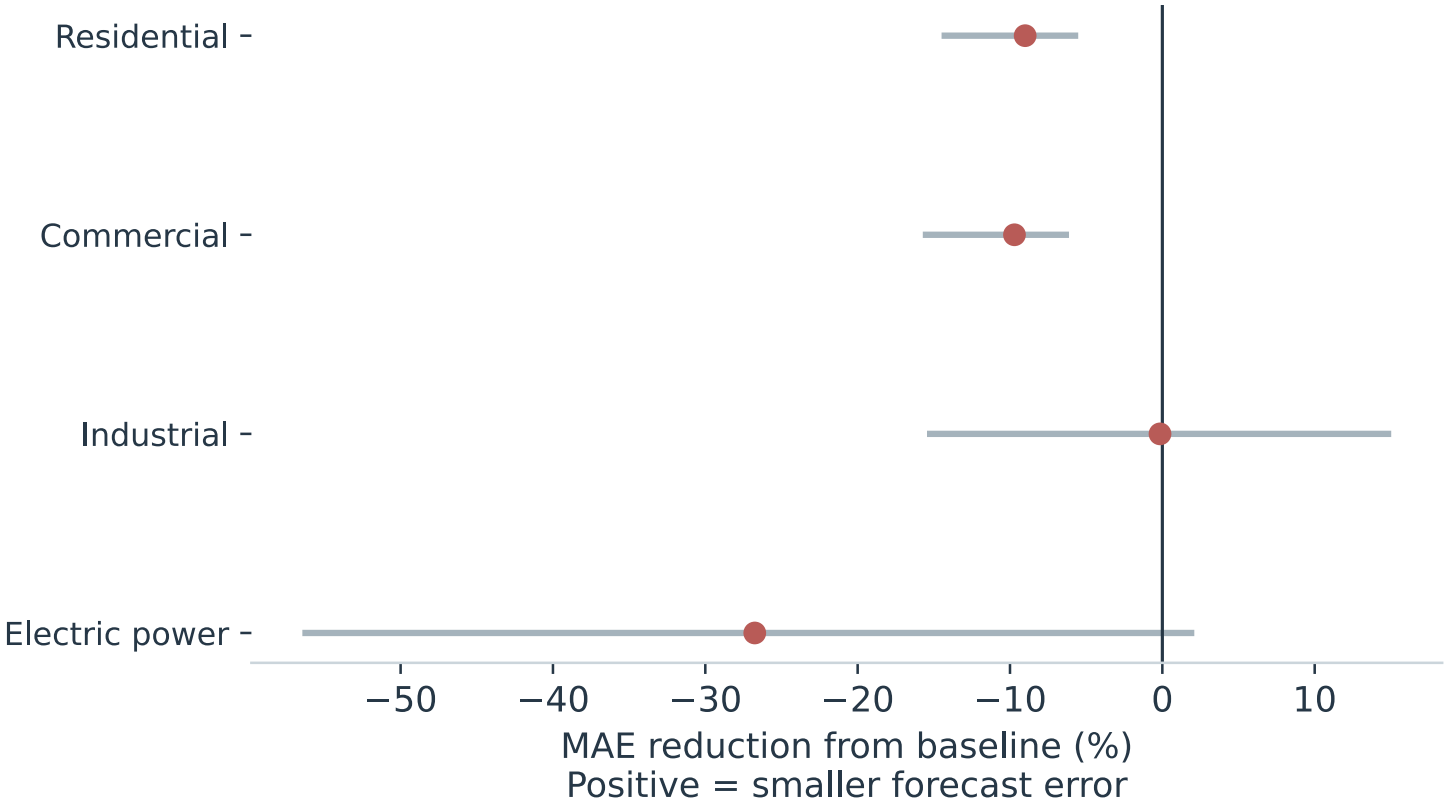

Figure 6. State forecasts do not improve the combined total. State forecasts are added within each sector and compared with the national model for the same 48 states. Positive values mean smaller average errors; gray lines show 95% uncertainty intervals. Missing state demand observations are left out of model fitting. Chart data.

The tested state model does not lower combined-demand MAE in any of the four sectors. State detail adds more effects to estimate, while some states have smaller or withheld samples. More detail therefore does not automatically mean a better forecast. This finding applies to this particular state model, not every possible use of state data.

The electricity-first approach needs more testing

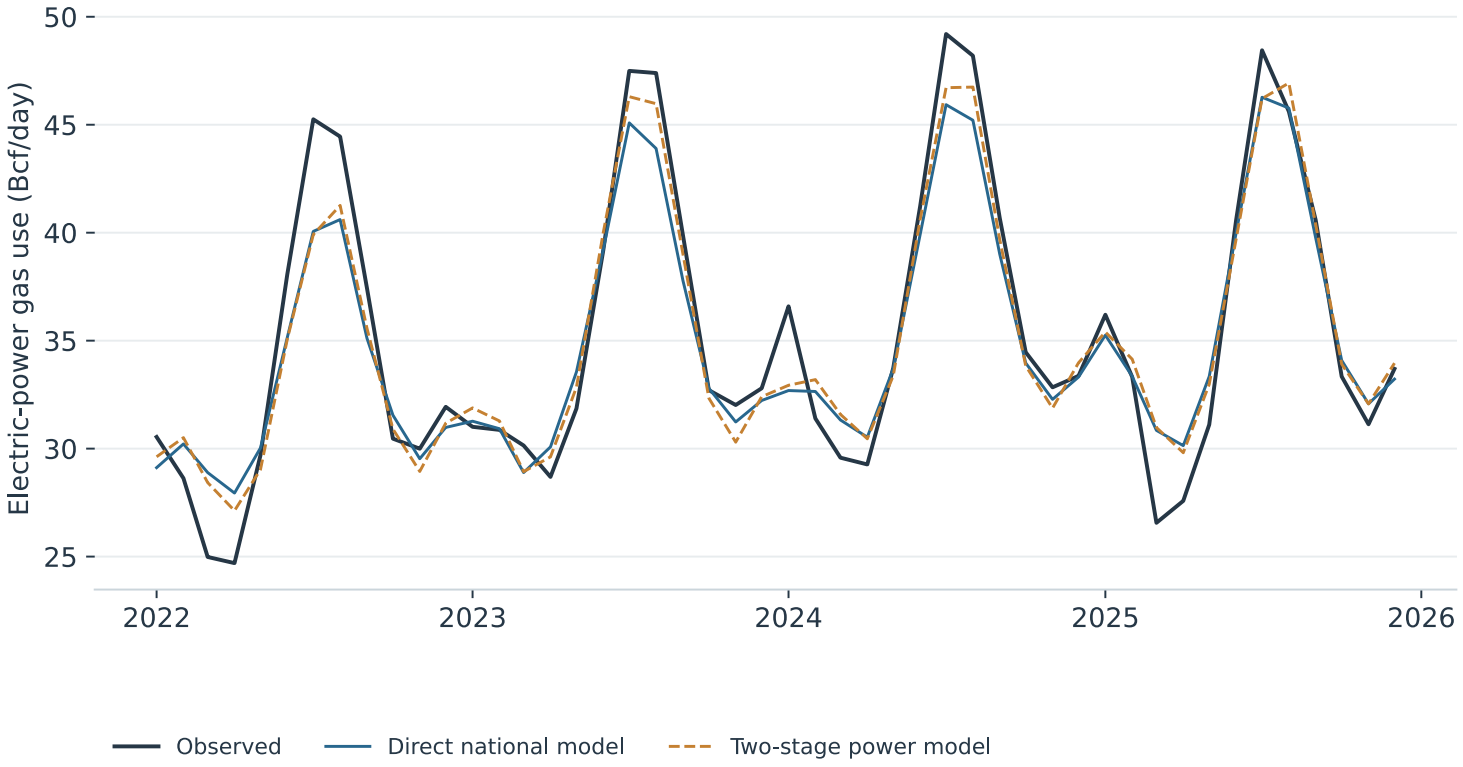

Figure 7. Two ways to predict gas use at power plants. This shows a recent part of the test period. The direct model predicts gas use in one step. The two-stage model first predicts electricity production and the portion made with gas, then calculates gas use. Both forecasts use normal weather. Chart data.

The two-stage model has an MAE of 1.44 Bcf/day. Its average error is 8.8% lower than the direct power model's. However, it does not pass all three checks. Its Holm-adjusted p-value is 0.130, above the 0.05 cutoff, even though the resampling interval is entirely positive. I therefore treat the apparent improvement as promising but unproven. The seasonal table shows where that average gain comes from. "Shoulder" means April, May, and October, between the defined summer and winter groups.

Season	Months	Direct-model MAE	Two-stage MAE
Shoulder	24	1.41	1.28
Summer (Jun-Sep)	32	2.02	1.60
Winter (Nov-Mar)	40	1.33	1.41

How the forecasts compare with EIA’s outlook

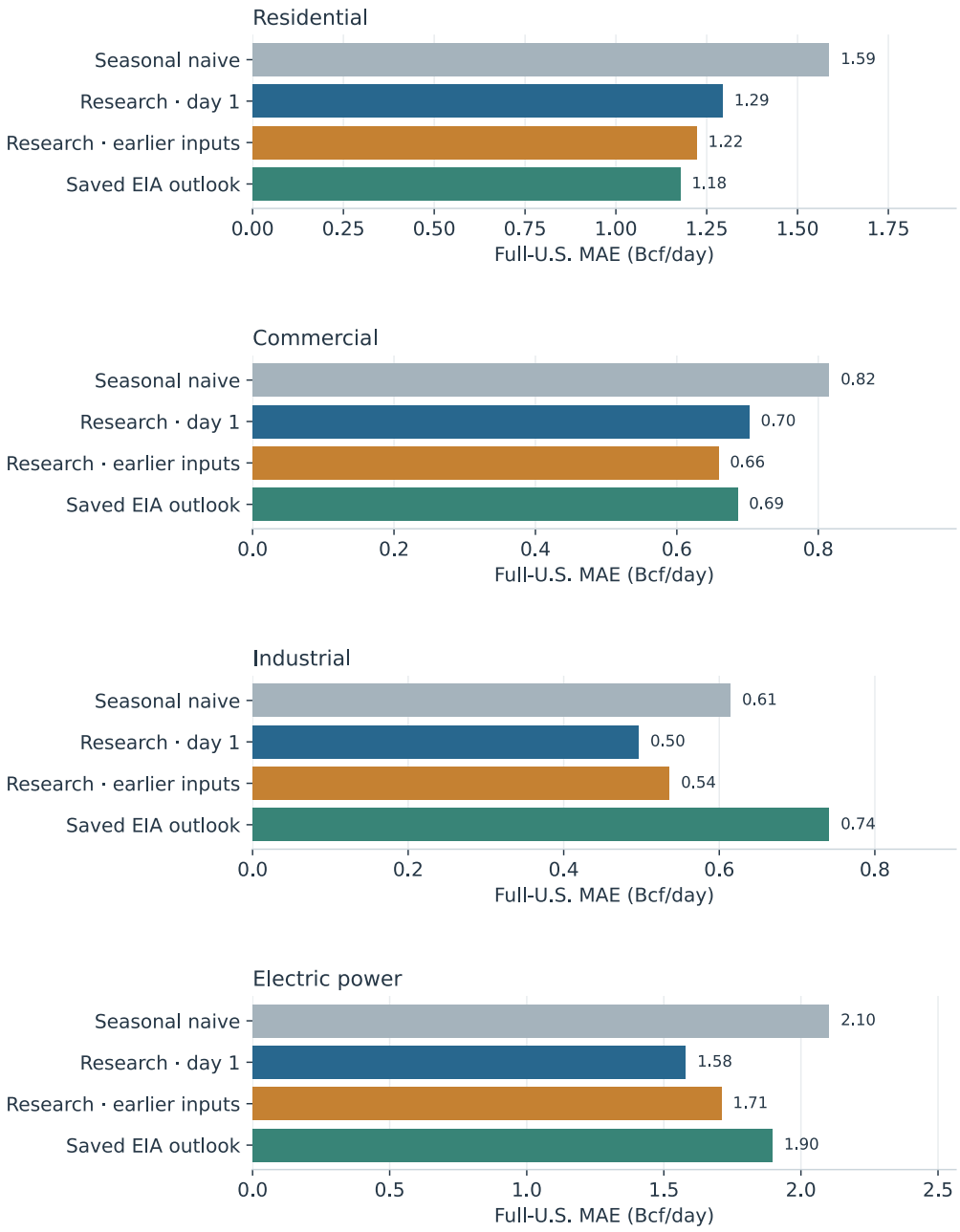

Figure 8. How the forecasts compare with EIA’s saved outlooks. All forecasts cover the same 96 months and the full U.S. total. The research forecasts add seasonal estimates for Alaska, Hawaii, and DC. STEO is EIA’s outlook from the preceding month. The earlier-input research model moves its cutoff back to make the comparison closer, but its historical data still contain later revisions. Chart data.

Average forecast miss for the full U.S., in Bcf/day, over all 96 months. Lower is better.

Sector	Repeat last year	Research: month start	Research: earlier inputs	Saved EIA outlook
Residential	1.59	1.29	1.22	1.18

Sector	Repeat last year	Research: month start	Research: earlier inputs	Saved EIA outlook
Commercial	0.82	0.70	0.66	0.69
Industrial	0.61	0.50	0.54	0.74
Electric power	2.10	1.58	1.71	1.90

EIA publishes the **Short-Term Energy Outlook (STEO)**, which includes gas-demand forecasts. Its archived forecasts have lower MAE than my earlier-input model in 1 of the four sectors. I use the preceding month's issue from the official EIA archive.

I also move my model's information cutoff earlier to make this a closer comparison. That version uses gas and electricity data at least four months old, prices at least three months old, and the annual weights available one month earlier. I add seasonal estimates for Alaska, Hawaii, and DC to match EIA's full-U.S. target.

This is still an imperfect comparison. My historical inputs contain later revisions, while EIA has a wider set of information and its own weather assumptions.

Do the results hold across years and model choices?

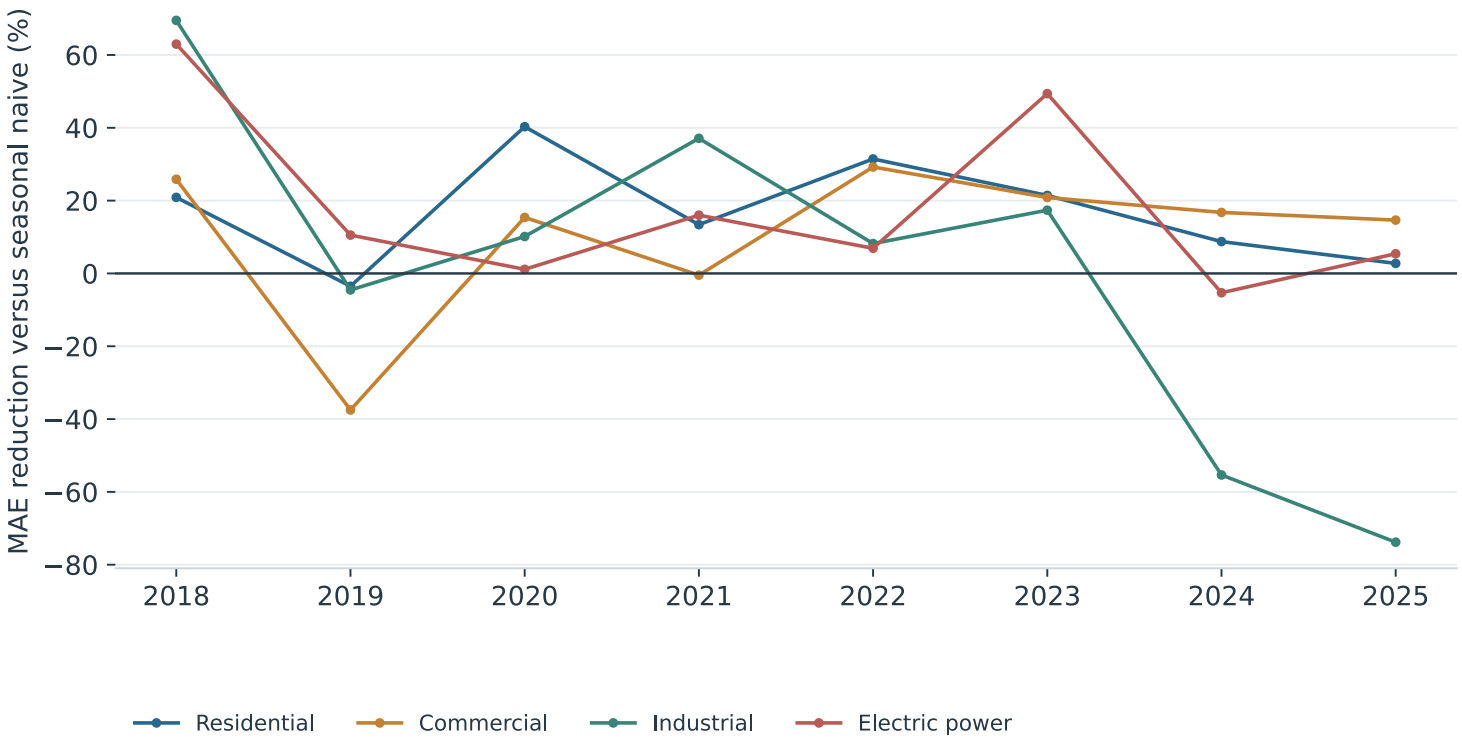

Figure 9. The size of the improvement changes from year to year. Each point compares the main national model with repeating last year's same-month demand. Positive values mean smaller average errors. There are only 12 months behind each annual point, so one year alone gives limited evidence. Chart data.

Average-error reduction compared with the main national model. Positive means lower error. These checks are not used to pick the winning settings.

Sector	Train from 2011	Rolling 8 years	Weaker penalty	Stronger penalty
Residential	+2.5%	+5.3%	+2.4%	-8.9%
Commercial	+3.8%	+4.4%	-0.9%	-3.0%
Industrial	-2.7%	+2.8%	-2.3%	-5.4%
Electric power	-6.9%	-2.1%	+1.1%	-8.9%

The alternative training windows and penalties show how much the scores change when the fitting choices change. Starting in 2011 also avoids much of the period affected by EIA's 2010 estimation change. "Rolling 8 years" means using only the

most recent eight years of training observations. These are checks on the result, not a search for the winning setting.

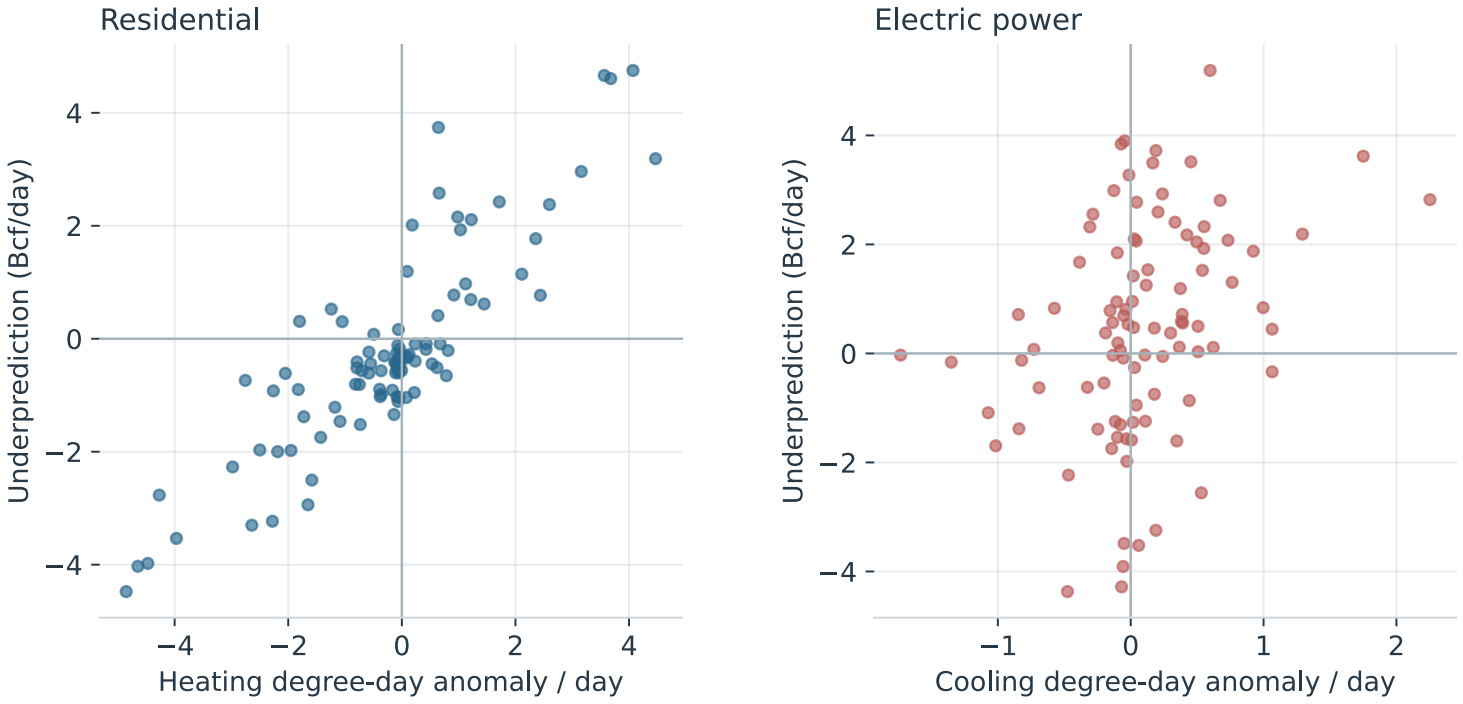

Figure 12. Unexpected weather helps explain forecast misses. Each dot is one test month. It compares weather minus its earlier 20-year average with actual demand minus the forecast. Chart data.

The dots compare unexpected weather with forecast misses. They use weather measured after the forecast date, so they help explain mistakes. A useful next test would use weather outlooks that were actually available when each gas forecast was made.

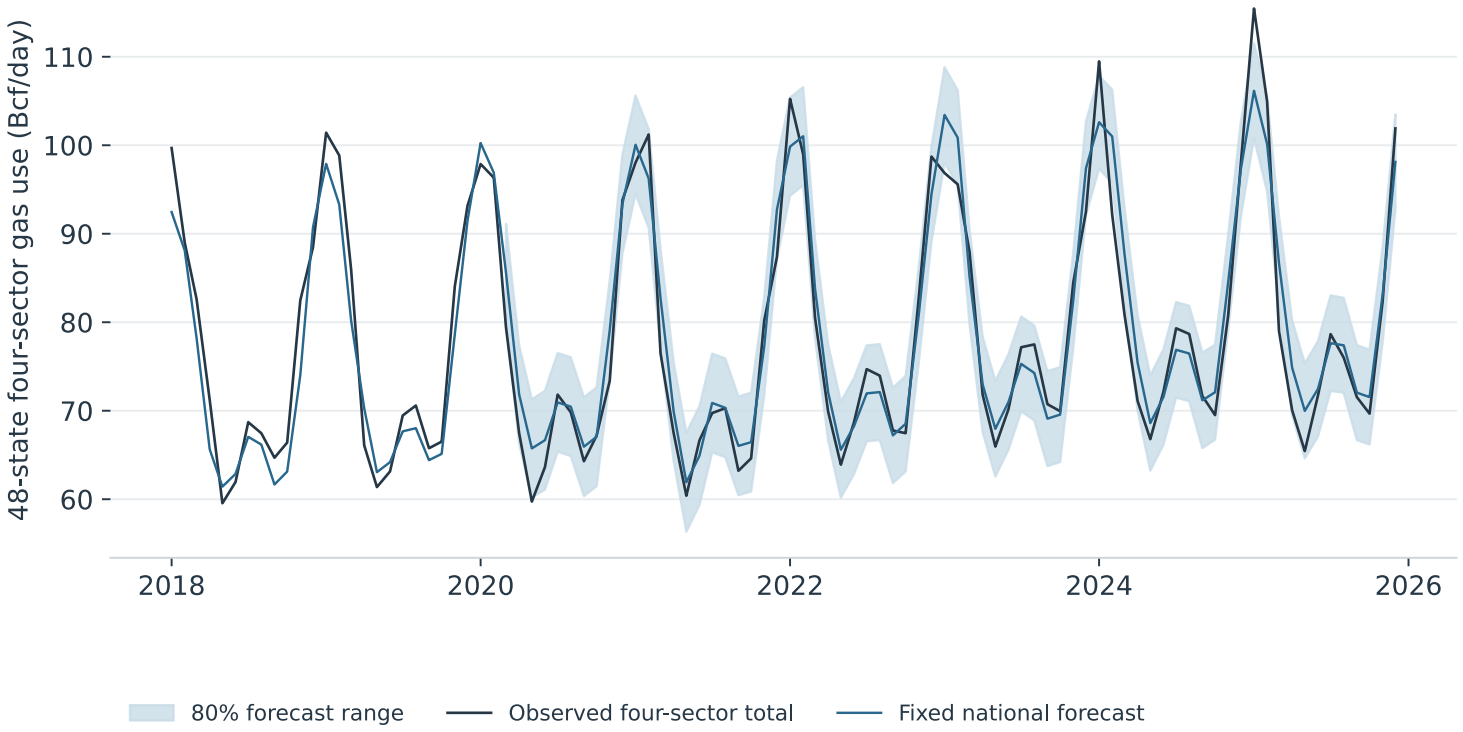

Figure 10. A range for demand, built from earlier forecast misses. The shaded range starts once 24 earlier errors would have been reported, allowing for the three-month delay. It uses errors for the combined demand total. This keeps the effect of sectors moving together, rather than adding separate sector ranges. The table below shows how often later observations actually fell inside. Chart data.

Forecast	Months	Outcomes inside range	Average range width
Commercial	70	82.9%	2.47 Bcf/day
Electric power	70	80.0%	5.59 Bcf/day
Industrial	70	77.1%	1.70 Bcf/day
Residential	70	82.9%	4.75 Bcf/day
Four-sector total	70	85.7%	10.88 Bcf/day

The band is intended to cover 80% of future outcomes. In this test it contains 85.7% of the 70 eligible four-sector totals. The bands start only when enough earlier errors are available. Demand patterns can change, so this measured coverage is more useful than assuming that an “80% band” will always cover 80% of later observations.

The data-center test does not isolate data-center gas use

The report from the U.S. Department of Energy and Berkeley Lab, published June 18, 2026, estimates that data centers used 192 terawatt-hours (TWh) of electricity in 2024, or 4.7% of U.S. use. TWh is a unit for an amount of electricity used over time. The report's 2030 reference scenario is 649 TWh, within a 521–843 TWh range.



Figure 11. Forecast misses do not show a clear data-center effect. The value is actual power-sector gas use minus the prediction, grouped by year, as a percentage of actual use. A positive value means demand was higher than predicted. The five states were grouped for this historical study, so the gaps cannot be treated as gas use caused by data centers. Chart data.

I look at five states—Virginia, Texas, Georgia, Ohio, and Arizona—as a possible place to see data-center effects. For their combined power-sector gas use, the model predicts too much in every year from 2020 through 2025. That does not show that data-center demand failed to grow. The model's trend and earlier demand already capture some growth.

This is a historical grouping of states, and I do not have direct measurements of their data-center electricity use. Power also moves across state lines. These limits mean the forecast misses cannot tell me how much gas demand data centers caused.

I then try growth in commercial electricity sales as a proxy, or stand-in measurement, for changing data-center activity. It does not improve forecasts. Its relative MAE

reduction is -1.3%, with a 95% interval of -5.6% to +1.0%. A negative reduction means the error got larger.

Weather cases for November 2026

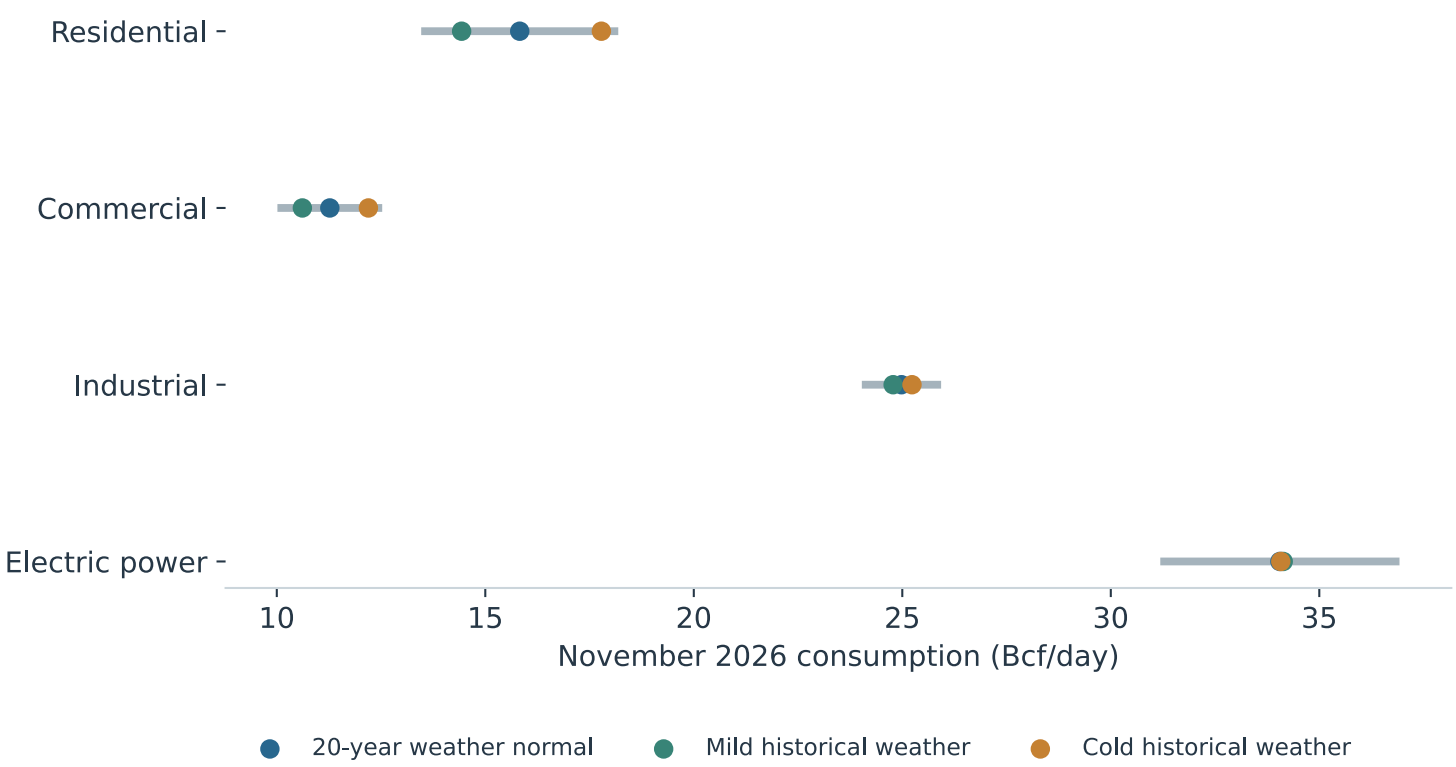

Figure 13. Possible weather cases for November 2026. Only weather changes between the dots. Each case reuses one whole historical pattern of state weather, with current weights and the same year in every sector. Gray bars show forecast ranges for normal weather. The other dots are possible cases, not probability estimates or the limits of those ranges. Chart data.

Gas use in Bcf/day under each weather case. The range for the total is estimated from total-demand errors, not by adding sector ranges.

Sector / total	Normal weather	80% forecast range	Mild-weather case	Cold / hot case
Residential	15.83	13.56–18.10	14.43	17.78
Commercial	11.27	10.11–12.43	10.61	12.20
Industrial	24.98	24.12–25.84	24.78	25.23
Electric power	34.06	31.28–36.83	34.13	34.08

Sector / total	Normal weather	80% forecast range	Mild-weather case	Cold / hot case
Four-sector total	86.13	80.86–91.40	83.96	89.29

This outlook uses gas and electricity observations through August 2026. The normal-weather case averages the previous 20 observations for the same calendar month. The mild and cold/hot cases reuse the full pattern of state weather from 2015 and 2019. I rank the earlier years by heating need in colder months or cooling need in warmer months. The chosen years are near the 10th and 90th percentiles.

5. Conclusion

The clearest lesson is to match the weather average to the way gas is used. Giving more importance to gas-heated homes lowers home and business forecast errors by 5.6% and 5.0% compared with population weights.

- **Useful change:** gas-heated-home and gas-customer weights improve heating-demand forecasts. These four comparisons account for the 4 supported gains among the nine main forecast tests.
- **Change that did not help:** adding up separate state forecasts does not beat the national model in any sector in this test.
- **Promising, but not established:** predicting electricity first gives an average 8.8% improvement for power-sector gas demand, but the evidence does not pass the study’s combined checks.
- **Question still unanswered:** commercial electricity sales do not improve this forecast or identify the amount of gas demand caused by data centers.

I would next test archived weather forecasts, recreate the original data releases more closely, find better measures of gas heating in commercial buildings, and connect measured data-center electricity use to the places that generate the power.

Data and code

Download data and code · Download results.