The Weather Commons: Public and Private Innovation in Forecast Accuracy

Soomi Kim* Jane Wu†

June 16, 2026

Extended Abstract

Although weather forecasting has become so ubiquitous that it can often be taken for granted, modern numerical forecasts are the culmination of a century-long, "quiet revolution" in science (Bauer, Thorpe, and Brunet 2015). These forecasts play critical roles both in public safety and across a wide range of economic sectors, from agriculture to transportation to retail (Dell, Jones, and Olken 2014; Cruz and Rossi-Hansberg 2024; Lemoine and Kapnick 2024; Song 2024; Shrader et al. 2026; Molina and Rudik 2026). As climate change intensifies extreme weather events, demand for more frequent and precise forecasts is expected to grow.

While much of the prior work has examined how weather and weather forecasts impact key economic outcomes, in this paper, we instead study the production of weather forecasts themselves. Specifically, we focus on weather forecast accuracy and explore the growing role of private firms in shaping forecasts relative to public agencies.

Weather forecasting provides a particularly attractive setting to study innovation in predictive technologies and the role of public and private providers. First, weather forecasting was one of the earliest fields to combine data, models, and computing for large-scale prediction—well before recent advances in AI (Lynch 2008). In many settings, prediction quality is hard to observe, and performance is often assessed against curated benchmark datasets. Weather forecasting, however, offers an objective, high-frequency measure of prediction quality: forecasts are issued multiple times a day by various providers and validated directly against the ground truth of observed weather. These features make weather forecast accuracy an unusually clean measure of innovative output.

Second, weather forecasting has evolved from a largely public enterprise into a heterogeneous public-private ecosystem. While governments continue to provide the underlying infrastructure,

*Columbia University
†University of California, Los Angeles

private firms increasingly build on these shared inputs and add proprietary models and data to develop their own forecasts.

Understanding where private firms can add value requires first outlining how forecasts are produced. Forecasting begins with the data collection of current weather conditions from instruments such as satellites, radar, and weather balloons. These observations serve as inputs into physics-based models that run on supercomputers, which generate raw numerical predictions. The raw predictions are then further processed by additional statistical refinement and judgment from human meteorologists before being issued to consumers as final forecasts. Much of this forecasting infrastructure is publicly operated and made broadly available to downstream users.

Given that all providers have access to the same public input, how do private firms differentiate, both against each other and the public baseline? What is the impact of increasing privatization of a public good? We explore three possibilities. First, private firms may add little value, repackaging public inputs without meaningfully improving accuracy. Second, private firms may push the frontier, but selectively, reflecting priorities that differ from those of public providers and potentially reshaping access to innovation. Third, private firms may perform uniformly better at forecast accuracy.

To explore these possibilities, we build a unique dataset that captures multiple stages of weather forecasting: (i) the raw numerical predictions generated by weather models, (ii) the final forecasts issued to consumers after additional processing and refinement, and (iii) for a subset of final forecasts, whether human meteorologists were involved.

From the THORPEX Interactive Grand Global Ensemble (TIGGE), we obtain numerical weather prediction output from two global models: the European Centre for Medium-Range Weather Forecasts' Integrated Forecasting System (the "European" model) and the American Global Forecast System (the "American" model). These publicly operated models do not provide forecasts directly to consumers but generate raw, model-based predictions that serve as a starting point for the forecasts ultimately issued by downstream public and private providers.

We obtain the final consumer-facing forecasts from ForecastAdvisor (Intellovations, LLC), which tracks daily temperature and precipitation forecasts of 21 public, commercial, and open-source providers at 1-, 3-, and 5-day horizons.[1] These providers transform the underlying numerical predictions into final forecasts through a combination of additional statistical processing, proprietary data, and human judgment.

By linking data from TIGGE and ForecastAdvisor, we can measure the value added by downstream forecast providers beyond the underlying raw models. In addition, for the National Weather Service (the US public provider), we observe two types of forecasts: an automated digital forecast (NWX) that statistically post-processes the raw models, and a human-edited forecast in which staff meteorologists revise the digital product before publication (NWS).

1. The providers in the sample represent a concerted effort from our data provider to capture all consumer-facing weather forecasts available in the US, with the exception of those with policies strictly against collecting their forecasts.

Our final data consists of randomly selected 200 weather stations in the US between 2007 and 2024 (Figure 1). The panel yields roughly 5.2 million provider-station-day-horizon observations, of which 1.8 million are at the 3-day horizon that anchors our analysis (Table 1). As shown in Figure 2, at the beginning of the sample there were a limited number of private providers, but entry by private providers has grown in recent years.

A simple linear regression suggests that on average private providers fare approximately 0.10°F worse than public in terms of mean absolute error (MAE) of their forecasts. This aggregate result, however, hides significant heterogeneity in performance. As shown in Figure 3, some firms greatly outperform the public baseline (the red dotted line), while others are clearly worse. In contrast to our priors, this points to private firms building on public foundations in heterogeneous ways. To better understand the value that the public and private providers add, we conduct the following analyses.

First, we examine whether the performance of private firms depends on geographic location. Some locations are intrinsically more difficult to predict (e.g., higher elevation), while some locations may be economically more important (e.g., higher household income), which may incentivize greater private investment. Table 2 shows suggestive evidence that private providers' relative advantage depends on geographic location.

Second, we benchmark each forecast against a naive blend of the raw numerical predictions, defined as the average of the European and American models. This benchmark represents the accuracy attainable from publicly available numerical weather predictions without any subsequent refinement. As shown in Table 3, the naive blend yields a pool mean absolute error of 3.81°F at the 3-day horizon. The NWS human-edited forecast recovers 0.42°F off this baseline, and the top decile of private providers recovers an additional 0.38°F beyond NWS. This implies that human-edited public refinement and private refinement contribute roughly comparable shares of the value above the raw numerical weather predictions that are publicly available. Six commercial and open-source providers fail to clear the naive blend at all, suggesting that simply layering proprietary branding on top of free numerical weather predictions can leave consumers worse off than the public floor itself.

Finally, we examine the identity of the frontier provider (defined as ranking first on the station-day-horizon "leaderboard"). We find that the frontier rotates substantially across providers (Figure 4). During the sample period, no single provider holds state-of-the-art at more than a modest share of station-days at a time, and the Herfindahl-Hirschman index of frontier-holding has declined substantially since 2007. The effective number of providers reaching state-of-the-art, defined as the inverse of the HHI, has roughly tripled over the sample period. This suggests that competition for the accuracy frontier has broadened over time.

Weather forecasting has a history of belonging to the commons: a shared public resource that helps individuals and organizations make decisions, across urban and rural, wealthy and poor

communities. The setting offers a particularly clear window into who gains from public and private innovation, under what competitive conditions, and the evolving comparative advantage of human judgment versus purely algorithmic prediction.

References

Bauer, Peter, Alan Thorpe, and Gilbert Brunet. 2015. "The Quiet Revolution of Numerical Weather Prediction." *Nature* 525:47–55.

Cruz, José-Luis, and Esteban Rossi-Hansberg. 2024. "The Economic Geography of Global Warming." *Review of Economic Studies* 91 (2): 899–939.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2014. "What Do We Learn from the Weather? The New Climate-Economy Literature." *Journal of Economic Literature* 52 (3): 740–798.

Lemoine, Derek, and Sarah Kapnick. 2024. "Financial Markets Value Skillful Forecasts of Seasonal Climate." *Nature Communications* 15 (4059).

Lynch, Peter. 2008. "The Origins of Computer Weather Prediction and Climate Modeling." *Journal of Computational Physics* 227 (7): 3431–3444.

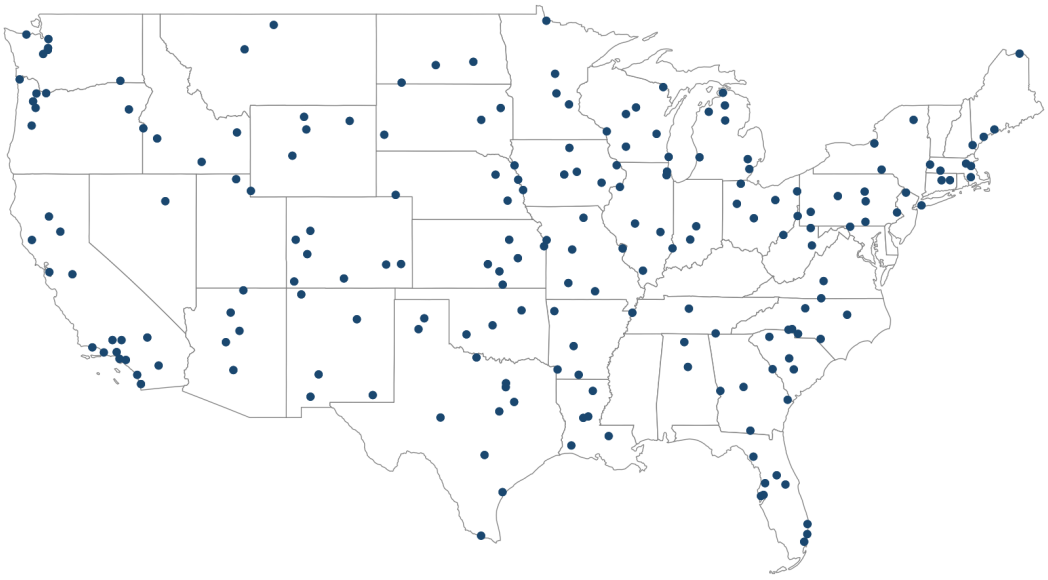
Molina, Renato, and Ivan Rudik. 2026. "The Social Value of Hurricane Forecasts." Forthcoming, *American Economic Journal: Economic Policy.*

Shrader, Jeffrey G., Stephan Thies, Laura A. Bakkensen, Manuel Linsenmeier, and Derek Lemoine. 2026. "Weather Forecasts Become More Important for Reducing Mortality as the Climate Warms." *Proceedings of the National Academy of Sciences* 123 (16).

Song, Yuqi. 2024. "The Value of Weather Forecasts: Evidence from Labor Responses to Accurate versus Inaccurate Temperature Forecasts in China." *Journal of Environmental Economics and Management* 125 (16): 102970.

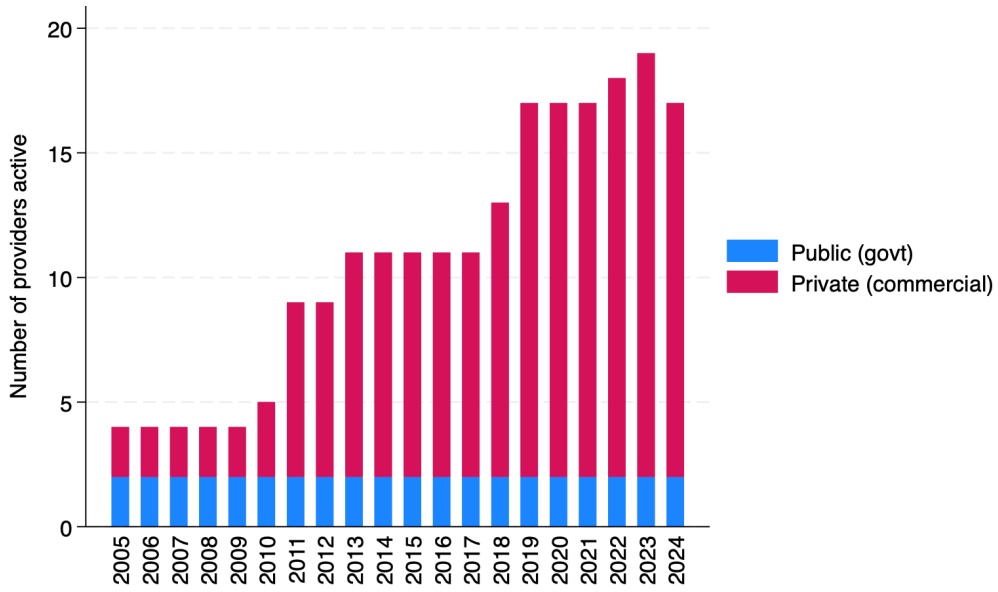
Figures and Tables

Figure 1: Map of Weather Stations



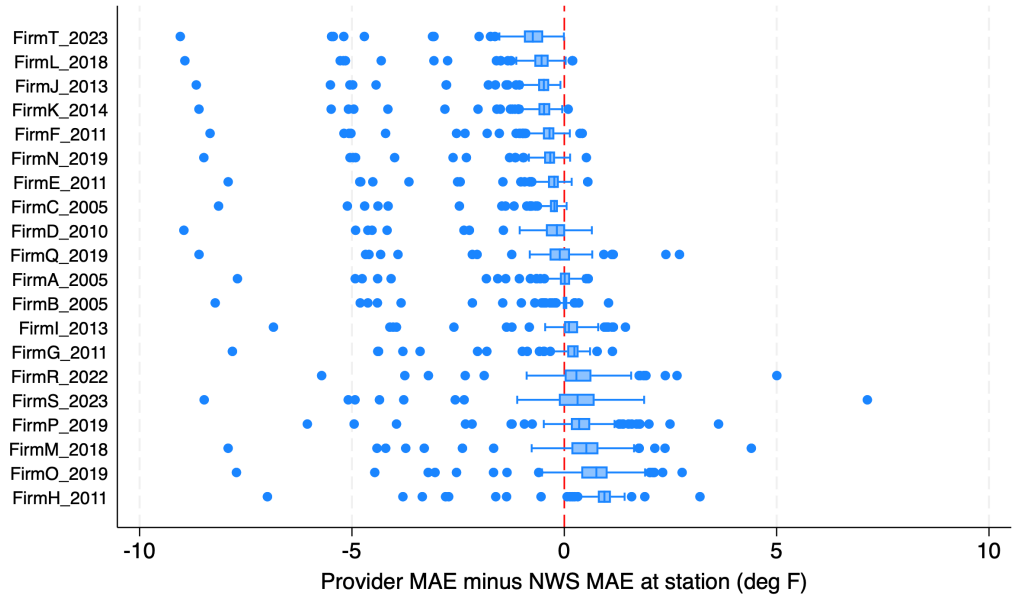
Notes: This figure plots the 200 US weather stations in our sample. Due to data sharing agreement with ForecastWatch, these stations were randomly selected from the full set of 711 available stations.

Figure 2: Private Provider Entry over Time



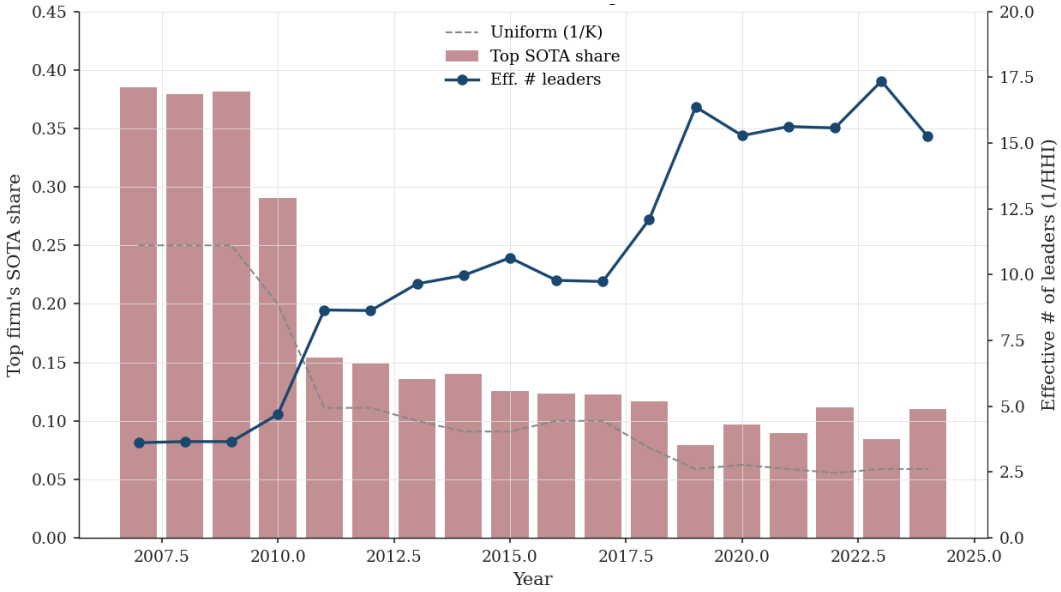
Notes: This figure plots the entry and exit of consumer-facing providers in the US from 2005-2024.

Figure 3: Distribution of Private Provider Forecast Accuracy vs. NWS



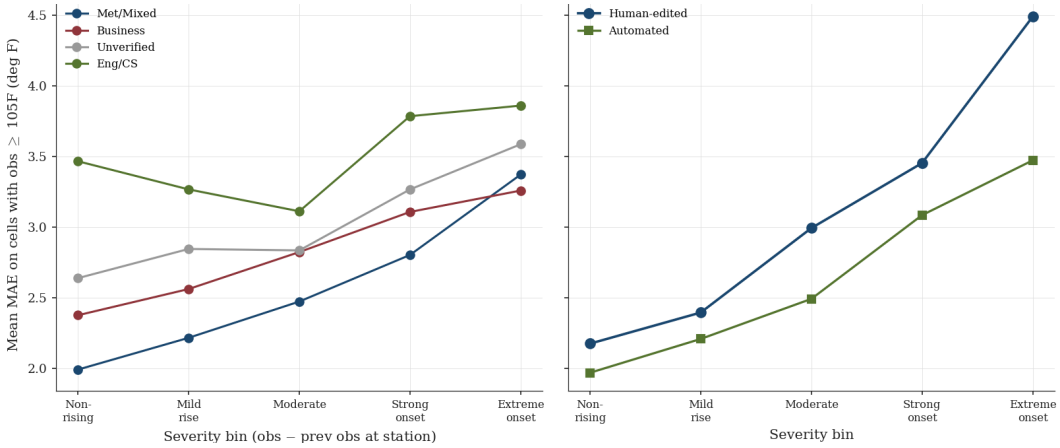
Notes: This figure plots the distribution of private provider accuracy relative to the National Weather Service (NWS). Each row corresponds to a private provider, and each blue dot represents a provider–station pair. The x-axis shows the difference between the provider's forecast error and the NWS's forecast error at the same station, where forecast error is measured using mean absolute error (MAE). A value of zero (red line) indicates that the private provider and the NWS performed identically. Blue dots to the left of zero indicate that the private provider outperformed NWS (lower forecast error), while blue dots to the right indicate that the NWS outperformed the private provider.

Figure 4: Leaderboard Dispersion Increasing over Time



Notes: Leadership is assigned to the provider with the lowest absolute forecast error at each station-day at the 3-day horizon. The left panel reports the annual share of station-day cells held by each of the top providers; the right panel reports the corresponding Herfindahl-Hirschman index of leadership shares and the effective number of leaders (1/HHI). Sample is restricted to the balanced set of providers reporting in each year.

Figure 5: Human Verification Value-Add Tapers at Tail Events



Notes: Severity bins are constructed on the difference between the current and previous observed daily high at the same station: non-rising ($< 5°$F), mild rise (5–10°F), moderate (10–15°F), strong onset (15–∞ within range), and extreme onset ($\geq 15°$F representing an unprecedented jump). The left panel compares mean absolute error by CEO/founder background of the provider's leadership team (meteorology-trained, business-trained, computer-science-trained, or unverified). The right panel compares the NWS human-edited forecast ("Human-edited") to the NWX automated digital forecast ("Automated"), which share inputs and infrastructure but differ only in whether meteorologists edit the output before release.

Table 1: Sample Descriptives

	Mean	SD	Min	Median	Max	N
Panel A: Station-level descriptive statistics (N = 200)						
Forecasts in panel	9251.2	207.6	8585	9226	9600	200
Providers ever covering station	21	0	21	21	21	200
Geographic Variables						
Elevation (m)	422.3	516.3	-36	237	2297	200
Distance to coast (km)	535.8	445.8	1	418	1482	200
Economic Variables						
Population (block group)	207327.8	716497.7	755	41866	8478072	200
Median age (years)	38.5	6.3	20	38	62	200
Median household income (USD)	66414.9	21078.1	32128	64112	238444	198
Mean commute time (min)	20.4	5.6	9	20	41	200
Panel B: Provider-level descriptive statistics (all providers, N = 21)						
For-profit (0/1)	0.90	0.30	0.00	1.00	1.00	21
Years active in panel	10.06	6.13	1.06	9.54	19.97	21
Stations covered	200.00	0.00	200.00	200.00	200.00	21
Panel C: Private-provider only (N = 19)						
Year founded	1999.78	16.35	1962.00	2004.00	2021.00	18
US-based (0/1)	0.47	0.51	0.00	0.00	1.00	19
Has a consumer app (0/1)	0.74	0.45	0.00	1.00	1.00	19
Founder is a meteorologist (0/1)	0.50	0.52	0.00	0.50	1.00	12
Forecasts in panel	228697.16	162976.50	29800.00	160640.00	569423.00	19

Notes: Sample covers 21 commercial, public, and open-source forecast providers across 200 randomly selected US ICAO weather stations between 2007 and 2024. Observations are at the provider-station-day-horizon level. Forecast accuracy is reported as the mean absolute error (MAE), in °F, between the provider's predicted daily high temperature and the subsequently observed daily high at the same station. Provider entry years are the first observed forecast date in the ForecastAdvisor archive.

Table 2: Forecast Accuracy (MAE) by Difficulty and Importance

	(1)	(2)	(3)	(4)	(5)	(6)
Private	0.102***	0.101***	0.101***	0.101***	0.102***	0.103***
	(0.031)	(0.031)	(0.030)	(0.030)	(0.031)	(0.030)
Geographic Variables						
Private × Elevation (km)		-0.234**	-0.314***	-0.307***	-0.284***	-0.271***
		(0.093)	(0.099)	(0.104)	(0.103)	(0.103)
Private × Distance to coast (100 km)			0.021***	0.021***	0.018***	0.023***
			(0.005)	(0.006)	(0.005)	(0.007)
Economic Variables						
Private × ln(Population)				0.008	0.010	-0.001
				(0.015)	(0.016)	(0.017)
Private × Median HH income (10k)					-0.017	-0.026*
					(0.012)	(0.015)
Private × Commute time (min)						0.011
						(0.009)
Station FE	Yes	Yes	Yes	Yes	Yes	Yes
Forecast-day FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,771,635	1,771,635	1,771,635	1,771,635	1,754,443	1,754,443
R-squared	0.071	0.071	0.071	0.071	0.072	0.072

Notes: OLS at the provider-station-day level; at 3 days out, pooled across providers. Outcome is the within-cell mean of abs(mos_high- obs_mos_high). Standard errors clustered at the station level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Value-Add: Decomposing the Pipeline

	Mean MAE (deg F)	Δ vs Naive blend	Δ vs NWS	N
Raw models				
Raw European model	3.754	-0.054	+0.367	1,590,626
Raw American model	4.499	+0.691	+1.112	1,590,626
Naive blend	3.808	—	+0.421	1,590,626
Public				
NWS	3.388	-0.421	—	144,073
NWX	3.488	-0.321	+0.100	146,331
Private				
Top private	3.012	-0.796	-0.375	798,734
Bot private	3.859	+0.051	+0.471	472,611
Open-source	3.715	-0.093	+0.327	28,877

Notes: The table reports the mean absolute error (MAE), in °F, of each entity at the 3-day forecast horizon. "Naive blend" is the unweighted average of the raw European and American numerical weather prediction outputs, (European model + American model)/2, evaluated against the observed daily high at the same station. "NWS" refers to the National Weather Service human-edited forecast and "NWX" to the corresponding automated digital forecast; "Top private" is the set of commercial providers whose aggregate MAE clears the public NWS baseline; "Bot private" is the remaining commercial providers; "Open-source" includes both public and community open-source projects. The column Δ vs Naive blend reports the difference between each entity's MAE and the naive-blend MAE computed on the same provider-station-day cells, so that the threshold is invariant to differences in provider coverage across stations and years. The column Δ vs NWS is computed analogously on the cells the entity forecasts. N denotes the number of provider-cell observations used to compute the row.