Probabilistic temperature forecasts that are actually calibrated.

26 US airport stations • 0–48h lead • GEFS ensemble post-processed with EMOS/SAR-SEMOS • verified on 11,174 resolved bracket samples.

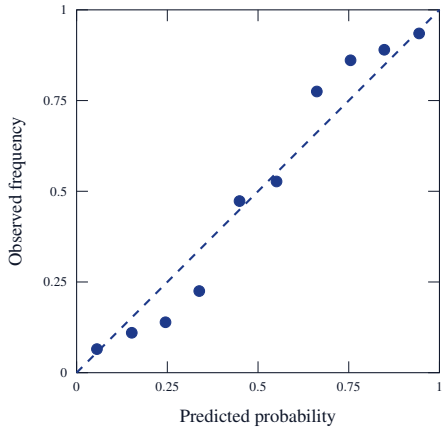
Per-station reliability curves published. 99.97% pipeline uptime over 80 continuous days of operation. CRPS competitive with the NWS NBM public baseline.

Per-station forecast skill

Station	CRPS	vs raw	Tier
KBUR	0.72	2.3×	**Best**
KLAX	0.73	2.8×	**Best**
KORD	0.79	2.1×	**Best**
KLAS	0.82	1.1×	**Best**
KMIA	0.88	1.3×	**Best**
KPHL	0.92	1.0×	**Best**
KDAL	0.94	0.9×	**Best**
KPHX	0.94	1.3×	**Best**
KPDX	0.95	1.1×	**Best**
KDEN	0.96	1.4×	**Best**
KDFW	0.98	0.9×	**Best**
KIAH	1.01	1.1×	Strong
KBOS	1.18	0.8×	Strong
KDCA	1.18	1.1×	Strong
KMDW	1.19	1.2×	Strong
KOKC	1.22	1.0×	Strong
KCVG	1.32	1.6×	Strong
KMSY	1.40	1.0×	Strong
KMSP	1.46	1.4×	Strong
KLGA	1.50	1.1×	Acceptable
KSEA	1.54	0.7×	Acceptable
KNYC	1.56	0.7×	Acceptable
KATL	1.65	1.0×	Acceptable
KSFO	1.67	0.7×	Acceptable
KAUS	1.82	1.0×	Acceptable
KSAC	1.93	0.7×	Acceptable

CRPS in °F, lower is better. NWS NBM baseline at 24h lead: 1.0–1.5. "vs raw" = MAE improvement over raw GEFS ensemble mean; > 1 means Aeolus beats raw point accuracy.

Reliability curve



Post-calibration reliability, n=1,984 resolved bracket samples. Marker size ∝ bin count. Perfect calibration falls on the diagonal.

Parametric trigger design — set deductibles you trust because the probability mass is honest, not adjusted upward by opaque vendor confidence.

Day-ahead load forecasting & degree-day pricing — distributional forecasts drop into stochastic load models and HDD/CDD swap calculations directly.

Drop-in JSON API — per-station, per-lead-hour distributional forecasts (mu, sigma); 4× daily refresh; sub-200ms p95 read; bearer-token auth; webhook delivery.

Xavier Briggs • xbriggs03@gmail.com