

Berkeley Earth Data Pipeline 2.0

*Building more robust and **resilient** climate data infrastructure*

After a series of upstream data disruptions last year, Berkeley Earth committed to building a more resilient data pipeline, one that no longer depends on any single source of observations. Pipeline 2.0 is that system: reinforced, expanded, and delivering the most comprehensive and most resilient observational temperature record available.

Why we rebuilt it

Berkeley Earth's land record is assembled from many overlapping archives. In the archive used for BEST-HR, more than 200,000 candidate source records reduce, after de-duplication and cross-referencing, to 58,311 unique station series. On average, each station appears in about four of the input datasets.

That redundancy in the inputs did not extend to delivery. A small number of aggregated products carried most of the timely data, and when delivery of GSOD, GHCN-Daily and GHCN-Monthly was interrupted from August 2025, monthly analyses were delayed. **Pipeline 2.0 was designed so that the loss of any one upstream product no longer interrupts the record.**

What changed in the ingest layer

The number of independent ingest sources has grown nearly threefold, to 32. For a growing set of countries, observations now come directly from national networks, including Météo-France, Deutscher Wetterdienst, the UK Met Office, Environment and Climate Change Canada, the Japan Meteorological Agency, Brazil's INMET and Australia's SILO.

New international collections add depth in under-reported regions and periods: WMO WIS2, SASSCAL in southern Africa, DWD CLIMAT and the HARD historic Arctic archive. NOAA's SSOD now takes the place of the static GSOD product within the source hierarchy.

FIG. 1

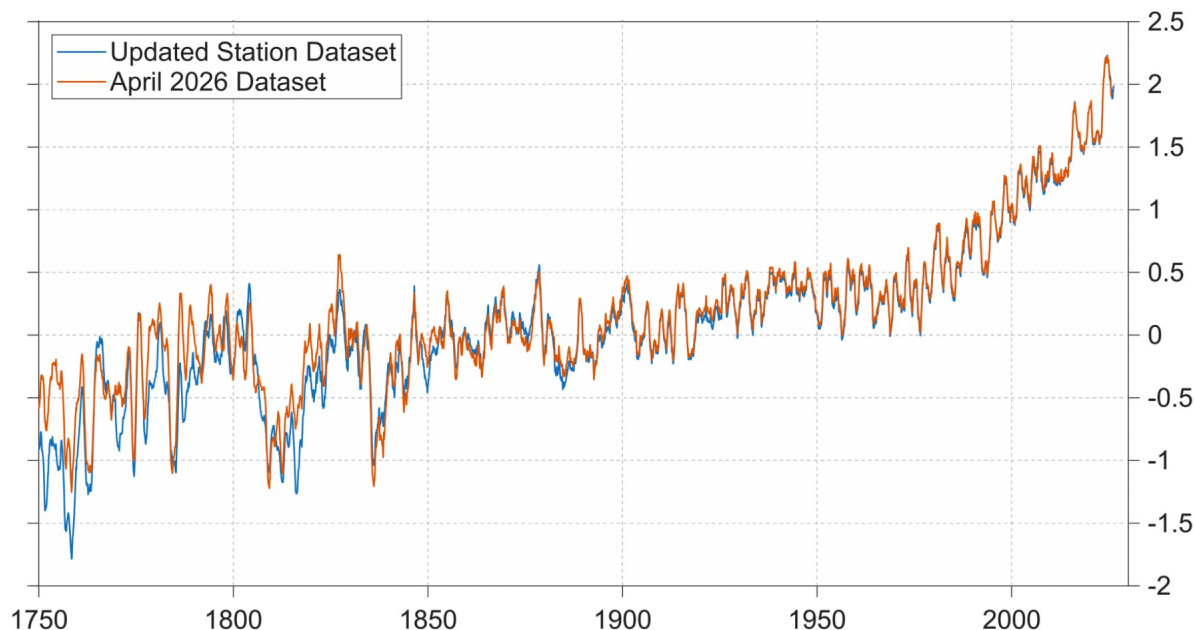
Stations in the Berkeley Earth archive, before and after the rebuild



Each dot represents about 250 stations. Counts follow automated de-duplication. Data: Berkeley Earth station database, April 2026 and updated Pipeline 2.0 builds.

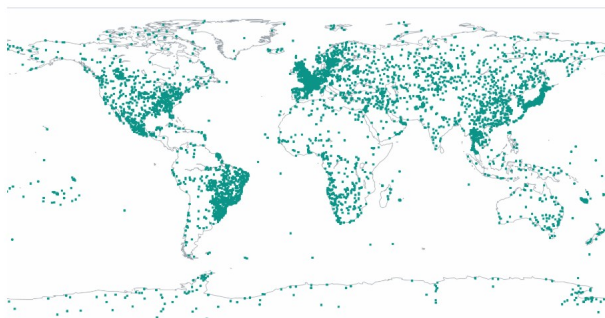
Coverage and consistency

FIG. 2 Global land average, reconstructed from the April 2026 and updated station archives



Temperature anomaly in °C. Blue: updated Pipeline 2.0 archive; orange: April 2026 archive. The two agree closely after 1850; earlier differences fall where few stations report and uncertainty is largest.

FIG. 3 New station locations



Most additions sit where national agencies now report directly: France, the UK, Germany, Japan, Brazil, Mexico and Thailand. SASSCAL adds stations across southern Africa, and international collections add smaller numbers worldwide. Few additions fall where coverage was thinnest, which is why the long-term record holds steady.

TABLE 1 Selected new collections in Pipeline 2.0

Source	Coverage	Stations
National networks		
Météo-France	France	5,017
Deutscher Wetterdienst	Germany	613
Met Office	United Kingdom	1,245
Environment and Climate Change Canada	Canada	4,300
SILO (Queensland Government)	Australia	756
Japan Meteorological Agency	Japan	946
INMET	Brazil	627
Servicio Meteorológico Nacional	Mexico	5,096
Thai Meteorological Department	Thailand	227
International collections		
WMO WIS2	Global, real time	168
SASSCAL WeatherNet	Southern Africa	157
DWD CLIMAT	Global, monthly	1,247
HARD	Historic Arctic	64

A selection of the sources added in Pipeline 2.0. Counts are stations in the merged archive that draw on each source; many draw on several. Where daily and monthly feeds exist, the larger count is shown.

Where the upgrade sits

Pipeline 2.0 replaces the two foundation layers of the Berkeley Earth processing chain. Every layer built on them runs as before.

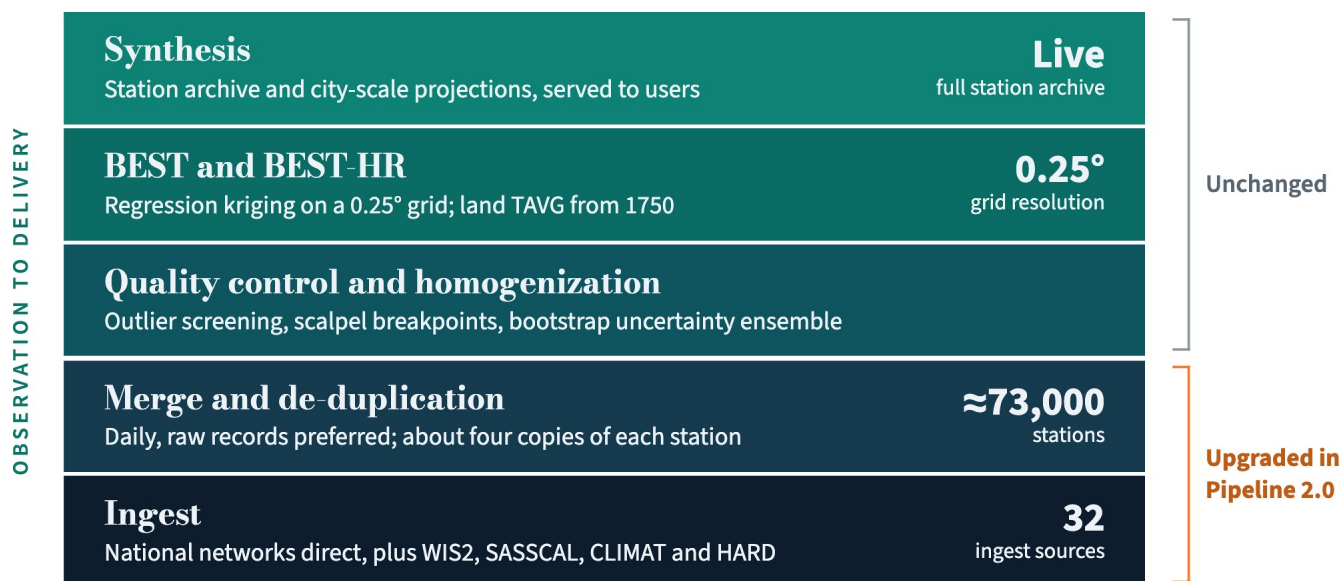


FIG. 4 The Berkeley Earth processing chain, from ingest (bottom) to delivery (top).

What hasn't changed

Only the way data comes in has changed. Everything after that works exactly as before: the same quality checks, the same corrections for changes at individual stations, and the same methods for building the record and estimating its uncertainty.

Most new stations fall in regions that were already well sampled, such as the UK, France and Germany, and many records are short. They add density and redundancy rather than filling historical gaps, and they do not materially change the global land average relative to its stated uncertainty.

Faster delivery

With fewer points of failure upstream, each month's analysis can begin sooner, and updated results reach users earlier than before.

Now on Synthesis

The complete Berkeley Earth station archive, including every station added in this rebuild, is now available on Synthesis, with near-real-time updates to follow.

Data	berkeleyearth.org/data and Synthesis (synthesis.berkeleyearth.org)
Licence	CC BY-NC 4.0. For commercial licensing, contact admin@berkeleyearth.org .
Cite	Rohde, R. A., Calvert, B., and Hausfather, Z.: Berkeley Earth Surface Temperature – High-Resolution (BEST-HR), Earth Syst. Sci. Data Discuss. [preprint], doi:10.5194/essd-2026-412 , 2026.