PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

Climate Change Service climate.copernicus.eu

IMPLEMENTED BY ECMWF

#EUSpace

Updates from the Copernicus Climate Change Service Global Land and Marine Observations Database

Robert Dunn (UKMO)

Simon Noone (NUIM)

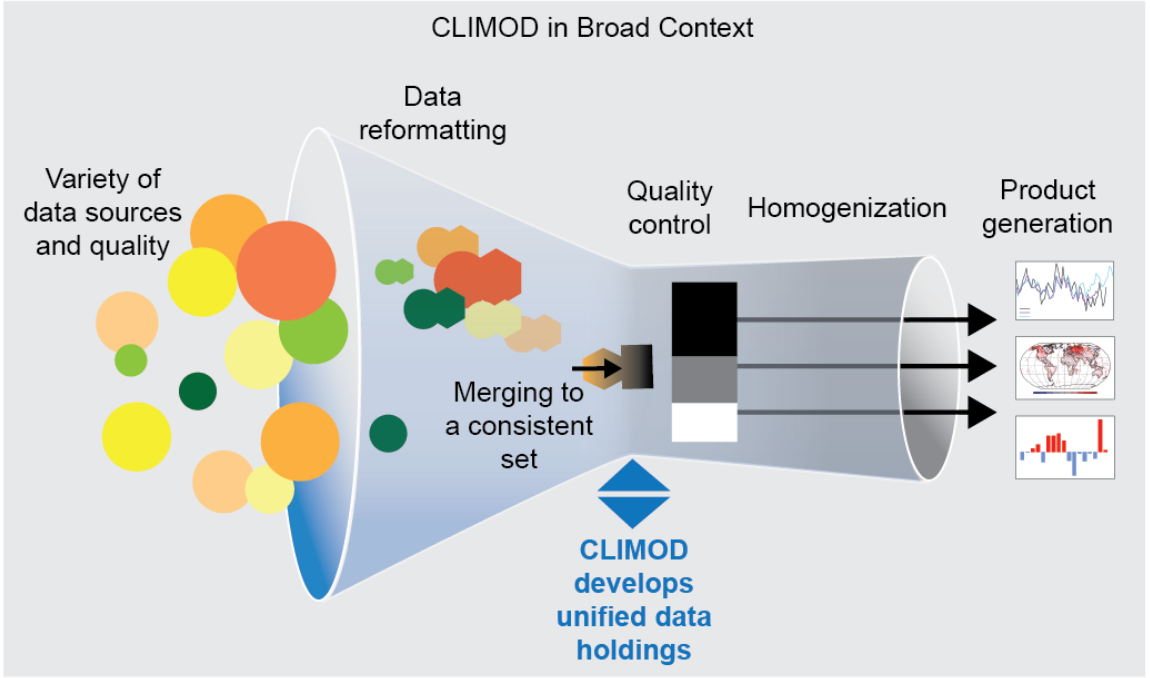
Matthew Menne (NOAA-NCEI)

Nancy Casey (CSS, Inc)

Peter Thorne (NUIM)

Concept – Global Land and Surface Marine Observations Database

- Thorne et al. (2017) outlined the requirements of a databank to improve the observations landscape
- Harmonise the fractured holdings of in situ land observations
  - Space
  - Variable
  - Timescale
- Unified format of integrated, discoverable holdings
  - Updated in near-real time
- Preserve known and unknown data and images



(CLIMOD – Comprehensive Land-based International Meteorological Observation Databank)

We cannot predict what is not observed, and we cannot analyse what is not archived.

PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

Climate Change Service climate.copernicus.eu

IMPLEMENTED BY ECMWF

#EUSpace

Overview of C3S 331 bis service

[Slide adapted from Peter Thorne]

C3S2 311 Bis contract

C3S2 311 Bis. Rescue, collection and processing of in-situ observations

- 34-month duration,
- 1st September 2025 to 30th June 2028
- 4.9 million Euros
- 14 subcontractors

WP0: Management
WP1: Climate data rescue
WP2: Station climate timeseries
WP3: Gridded climate products

Builds on previous contract(s) which addressed similar topics
Now all under one.

PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

Climate Change Service climate.copernicus.eu

IMPLEMENTED BY ECMWF

#EUSpace

In situ land data [WP2]

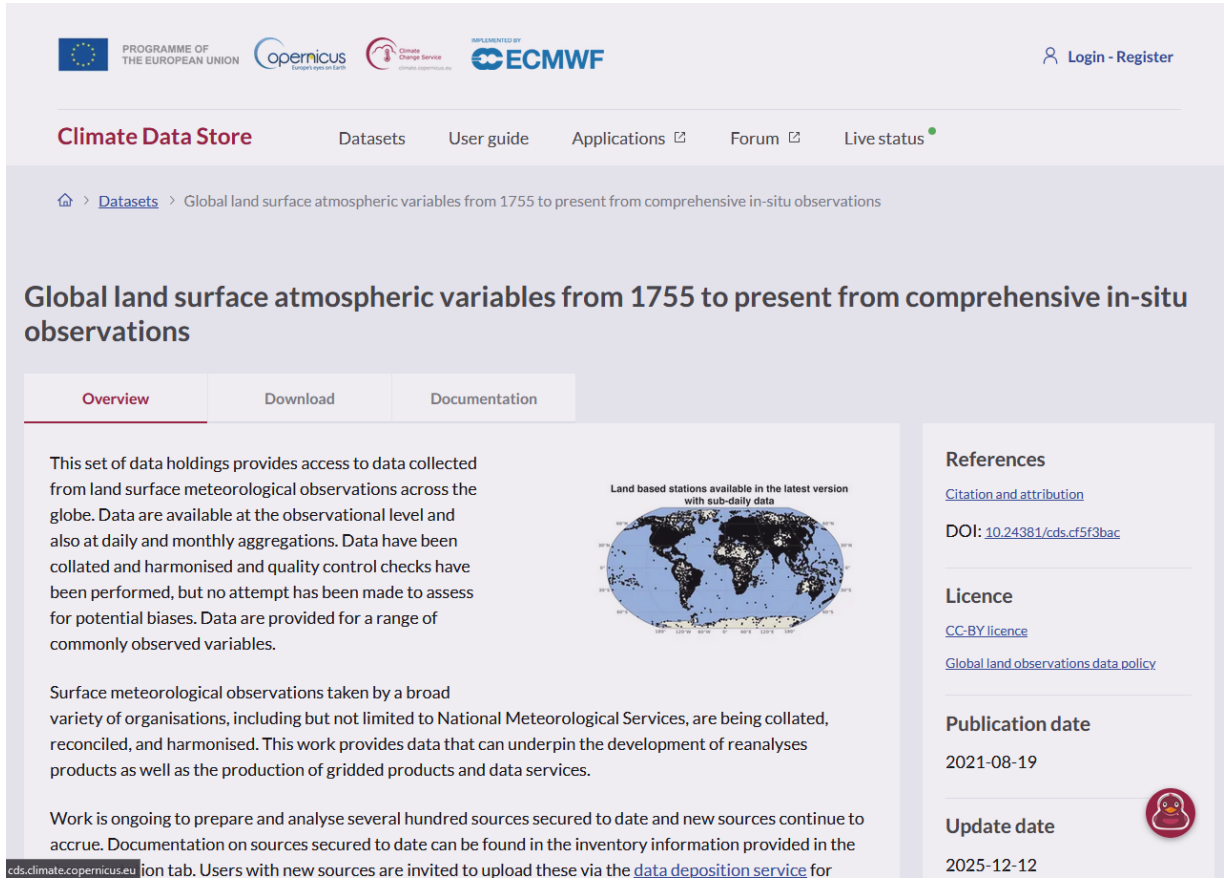
Task Lead: Simon Noone (NUIM)

Overview

- Pursuing additional sources of land data including those provided via the data deposition service.
- Preparation of sub-daily, daily and monthly harmonised holdings.
- Strong links and dependencies with Data Rescue and Digitization

- https://cds.climate.copernicus.eu/datasets/insitu-observations-surface-land?tab=overview
- Documentation and User Guides on Confluence

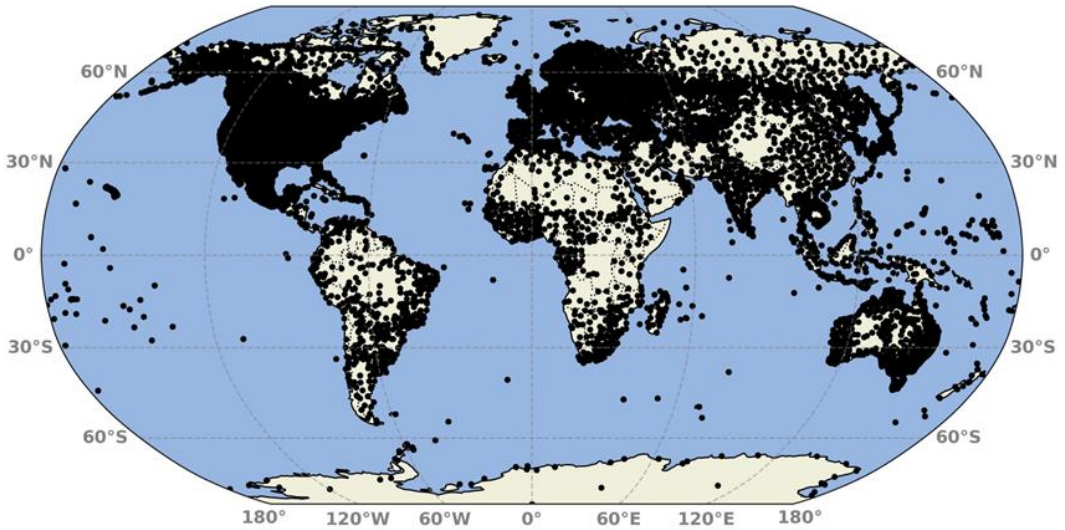
- Links to OSCAR-Surface and WIGOS Information System
  - E.g. To add identifiers for historical stations

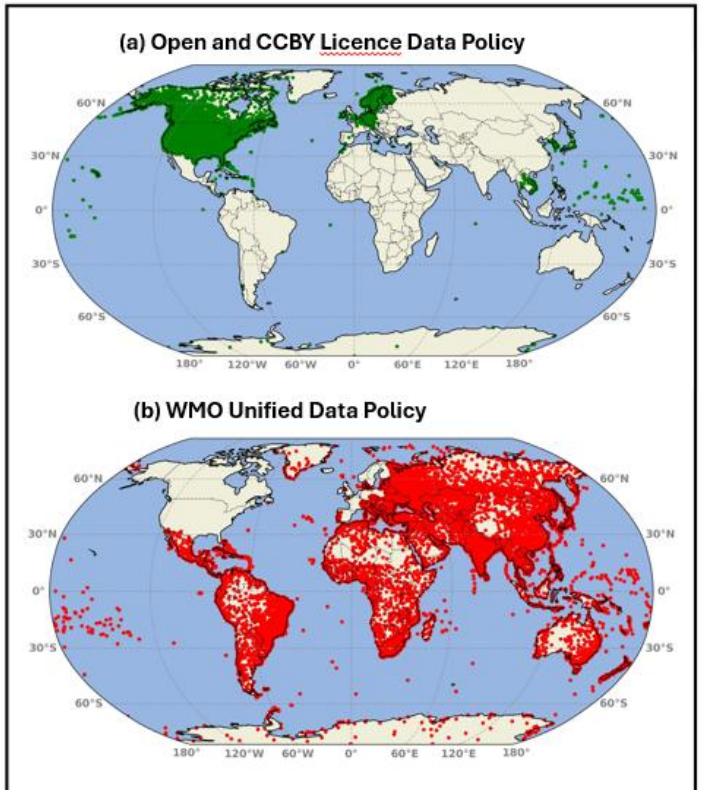


PROGRAMME OF THE EUROPEAN UNION

Daily and monthly holdings

- These closely mirror the GHCNd/GHCNm products
- Partnership with NOAA/NCEI
- Daily
  - ~177,000 inventoried stations
    - ~15,000 more than at start of service
  - 150 data sources
  - ~86,500 stations have at least 2 target ECVs
    - Precip, Temperature, Snow, Wind speed
  - 1763-2026
- Monthly
  - ~186,000 inventoried stations
  - 85 data sources
  - ~83,000 stations
  - Match daily station selection



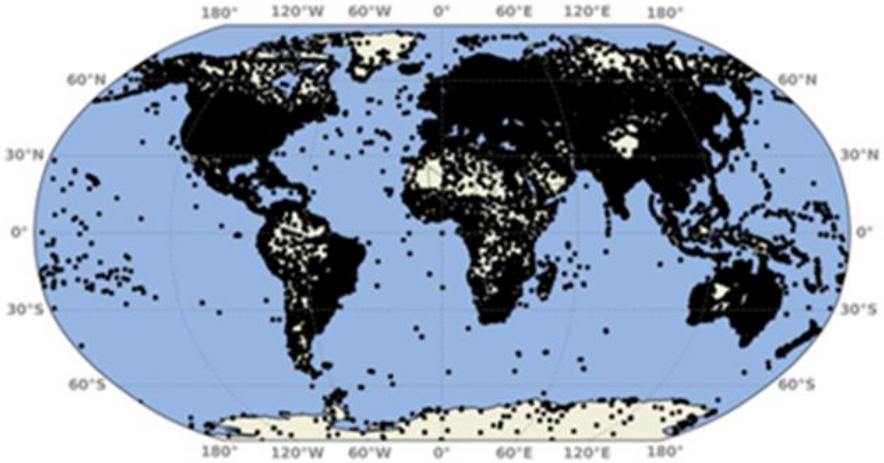


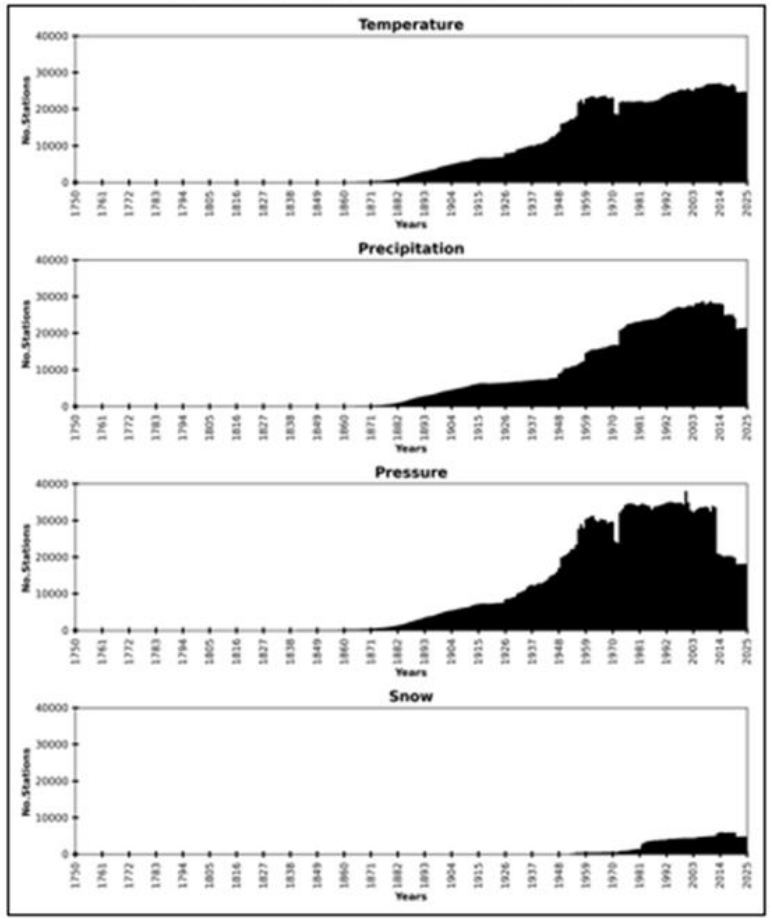

PROGRAMME OF
THE EUROPEAN UNION
Copernicus
Europe's eyes on Earth
Climate
Change Service
climate.copernicus.eu
IMPLEMENTED BY
ECMWF
#EUSpace
Subdaily
holdings

PROGRAMME OF THE EUROPEAN UNION

Data Sources

- Bulk of work has been on the subdaily holdings
- The ISD provided ~30k stations
  - Hard to maintain and use, and stopped in Aug 2025
  - Time for a new version, harmonising with GHCN products
- Data have been collected from a wide range of sources
  - Data rescue collections, some undertaken under C3S
  - Publicly available data feeds
  - Copernicus International Exchange Agreements
- Public release 3 / internal release 8.1 [March 2026]:
  - ~155,000 inventoried stations
  - 206 sources
  - ~35,000 stations (1716-2026)
  - 18 billion unique observations
- Focus initially on ECVs with large data volumes and achievable QC

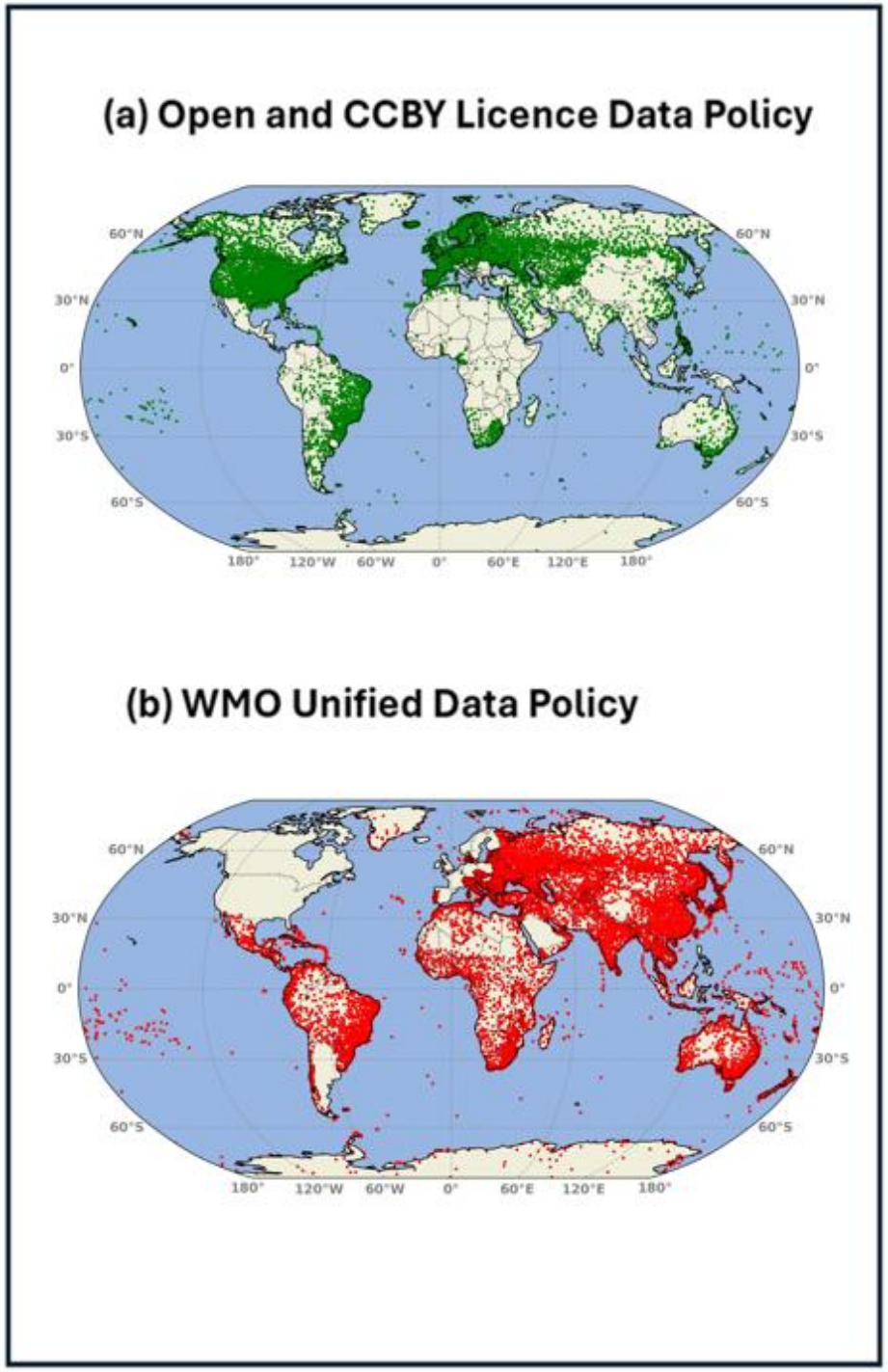




PROGRAMME OF THE EUROPEAN UNION

Data Sources

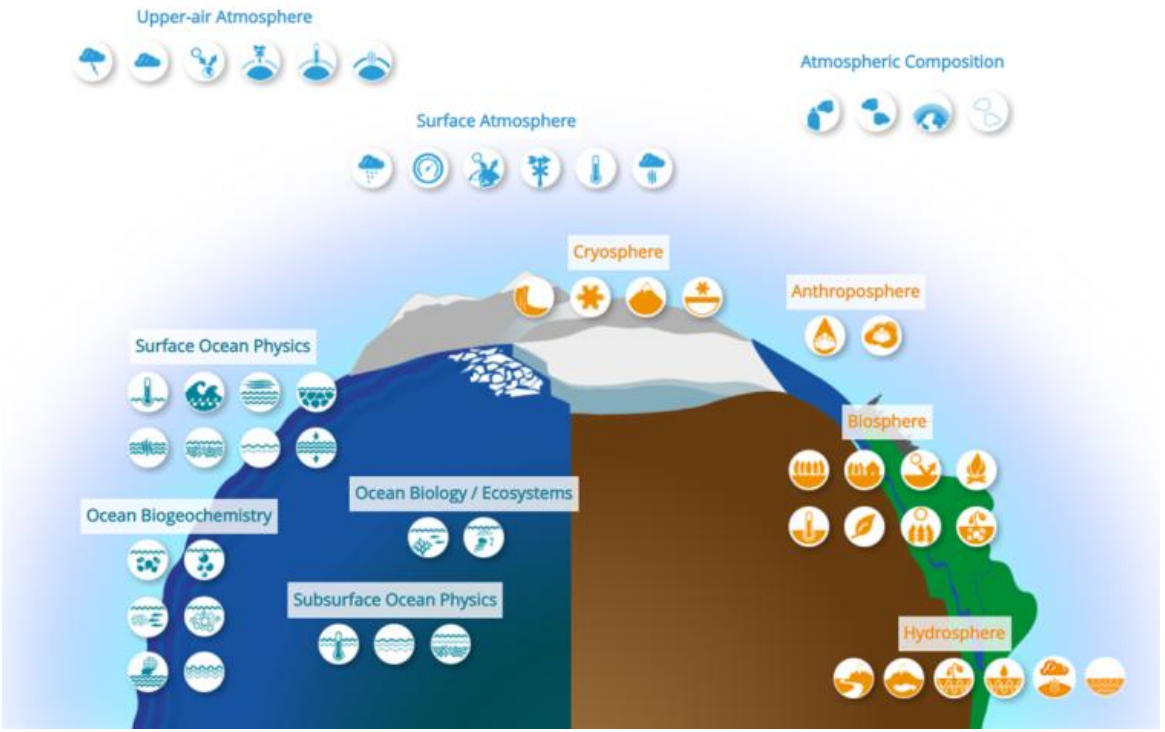
- Data policy noted and retained throughout processing chain
  - 145 sources have open data
  - 11 CCBY
  - 50 stated/implied WMO Unified Data Policy (e.g. Res 40)
  - 6 have mixed data policy
- Original sources are archived to protect against potential data loss
- Conversion to standardised internal format to enable further processing.
  - Currently “psv” (csv with pipe-’|’, also used for GHCNh)
  - Moving to parquet (binary) for i/o speed and storage volumes



Initial ECV selection

- Focussed on ECVs which were
  - Reported across all timescales
  - Reported frequently in the station record
  - Had identified user needs
  - Based on the daily holdings for overall selection
  - Using holdings in 2018

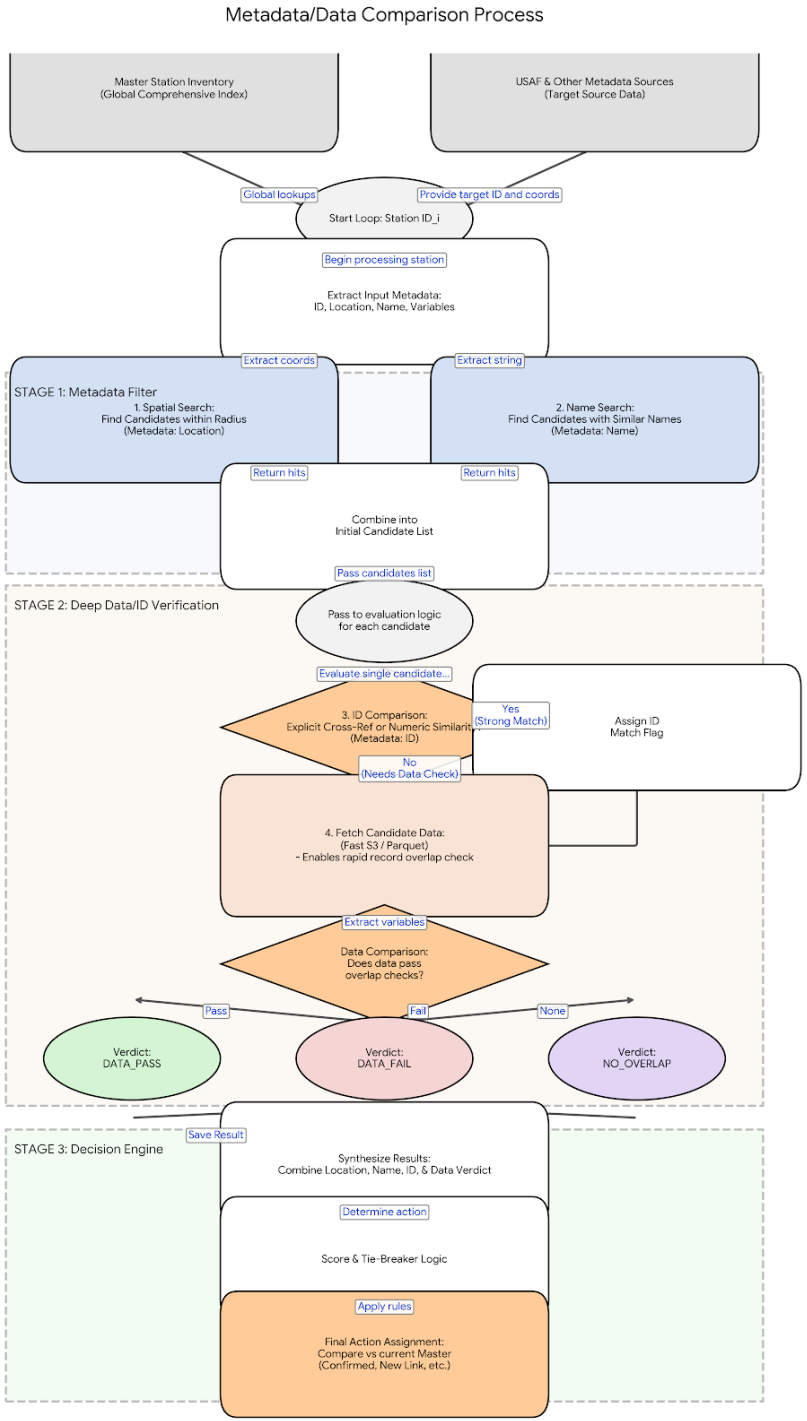
- For sub daily
  - Temperature, Dewpoint, Sea & Station Level Pressure, Wind speed & direction
- Future
  - Wet bulb temperature, relative humidity, cloud information
  - Snow cover and depth
  - Precipitation



GCOS ECVs

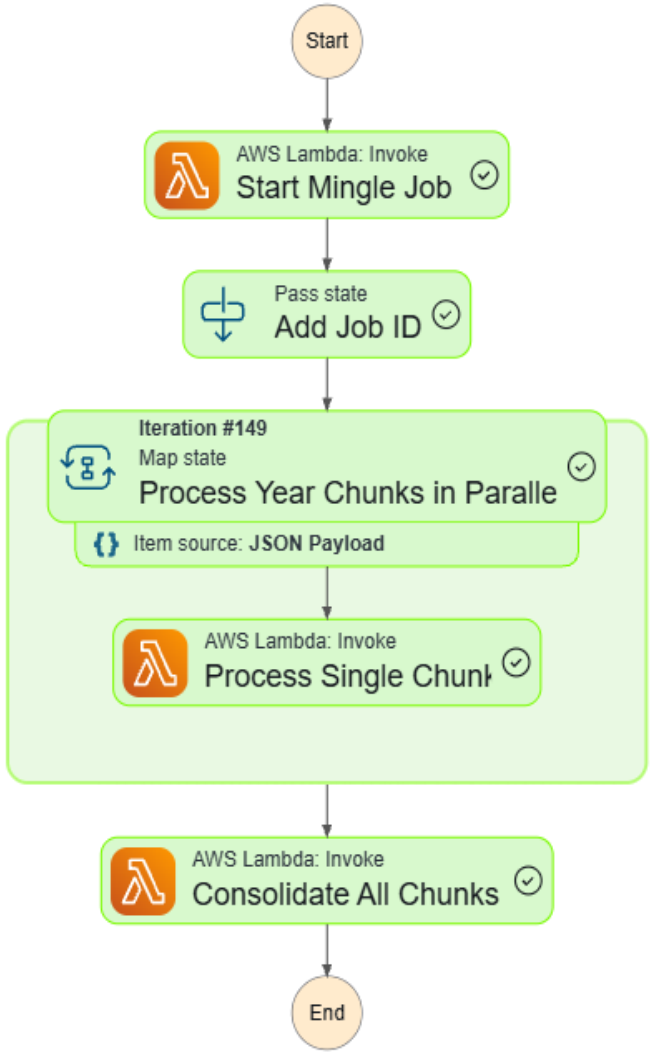
Merging

- ISD was heavily based on USAF holdings
  - Early NCEI/NCDC sources and Washington GTS
  - Merging approach not well documented
  - So far only using ICAO, WMO, WBAN identifiers
- Merge approach now follows GHCNd method
  - Menne et al, 2012
  - Each source assigned a priority, USAF as lowest
- Steps in comparison process
  1. Metadata comparison (coordinates, name etc)
  2. Determine overlap by variable between candidate source and other sources
  3. Decision engine based on metadata and data comparisons



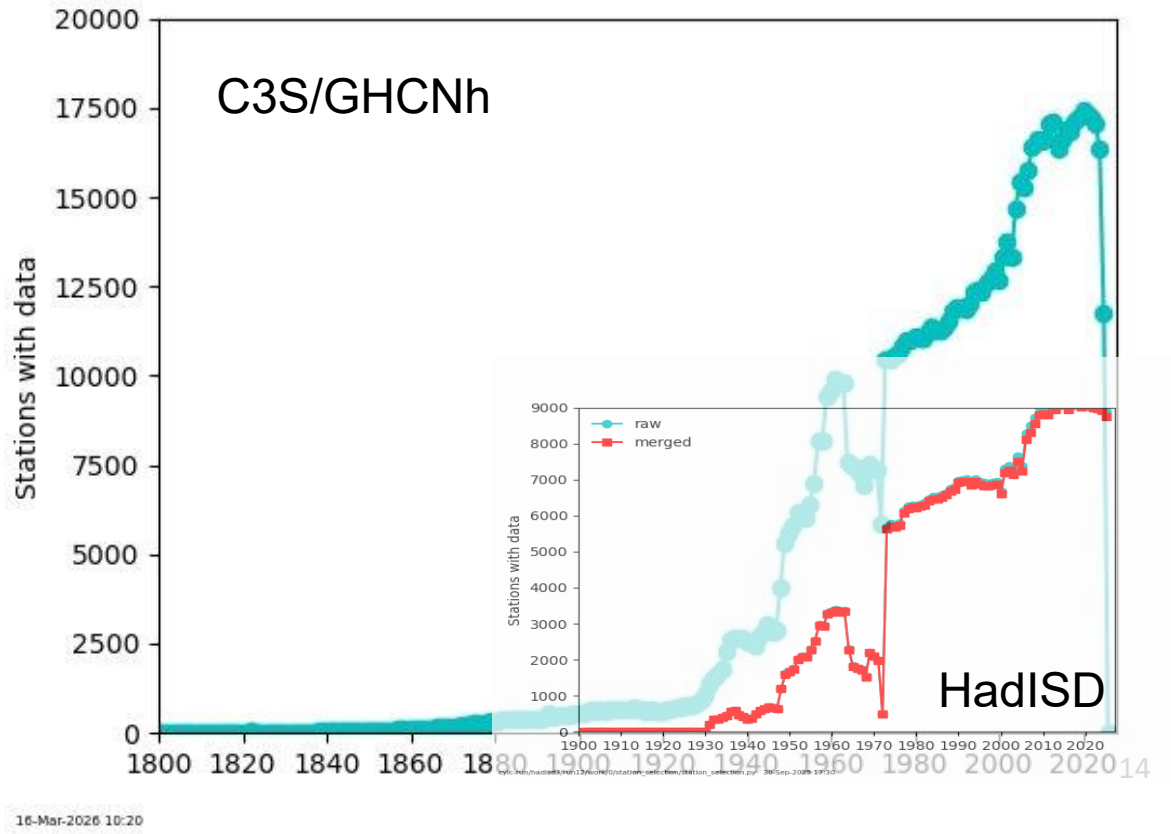
Refactoring of mingle to improve processing speed

- Recent work focussed on processing speed
- Centralization of metadata sources
- Use of AWS S3 and parquet files for rapid period of record/overlap determinations and data comparisons
- Modularization and “atomization”: Breaking complex logic into smaller, discrete functions.
- Can run 1000 jobs at once, which complete in 10-15 minutes (40 jobs for 40K stations)
- Reduces reprocessing time from weeks to hours
- Allows for fast additions of new sources and re-evaluation of existing sources
  - Not previously achievable.



Sub-daily Quality Control

- Upstream data have not undergone any consistent (or necessarily documented) QC
- Requirement to provide quality assured data to C3S Climate Data Store
  - Hence sub-daily precipitation not included so far
- Code and tests based on HadISD checks
  - Dunn et al, 2012, Dunn et al, 2016, Dunn 2019
  - These were inspired by GHCNd checks in Durre et al, 2010
- Initial data and logic checks
- “Internal” checks using observations themselves
- “Buddy” checks comparing with neighbouring stations

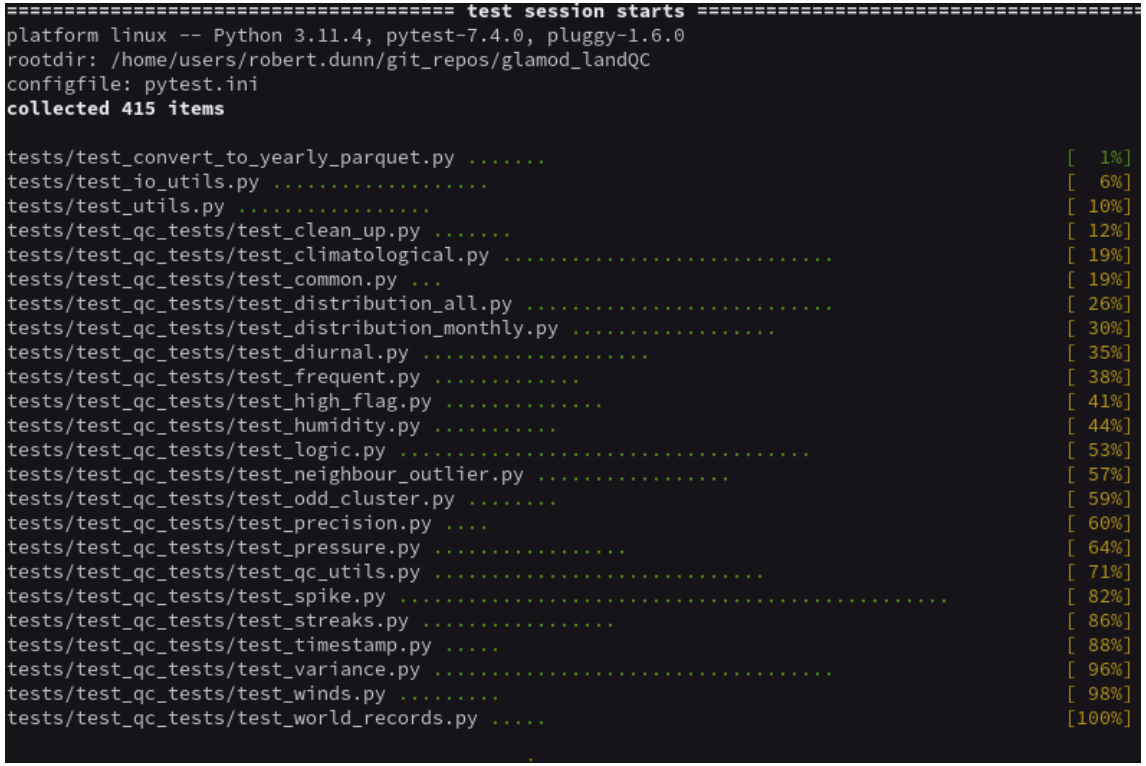


Sub-daily Quality Control

- Odd cluster (temporally isolated short runs of observations)
- Frequent values
- Diurnal cycle
- Distribution
- World records
- Repeated streaks (runs, excesses, whole days)
- Climatological
- Spike
- Humidity (supersaturation, dewpoint depression)
- Wind logical consistency
- Buddy checks
- Clean up (removes months with lots of flags)
- High flagging (>20% obs flagged in 2 ECVs)
- Logic checks
  - data and metadata consistency
  - No more than 0.5% of data falls outside of certain bounds
- Pressure
  - Sea level pressure and station level pressure consistency
  - With elevation if available
- Timestamp
  - No duplicated timestamps with different values
  - (ideally no duplicated timestamps at all!)
- Precision
  - If Td and T have different reporting precision
  - Can cause flags in humidity check
  - For power-users/humidity specialists [so far]

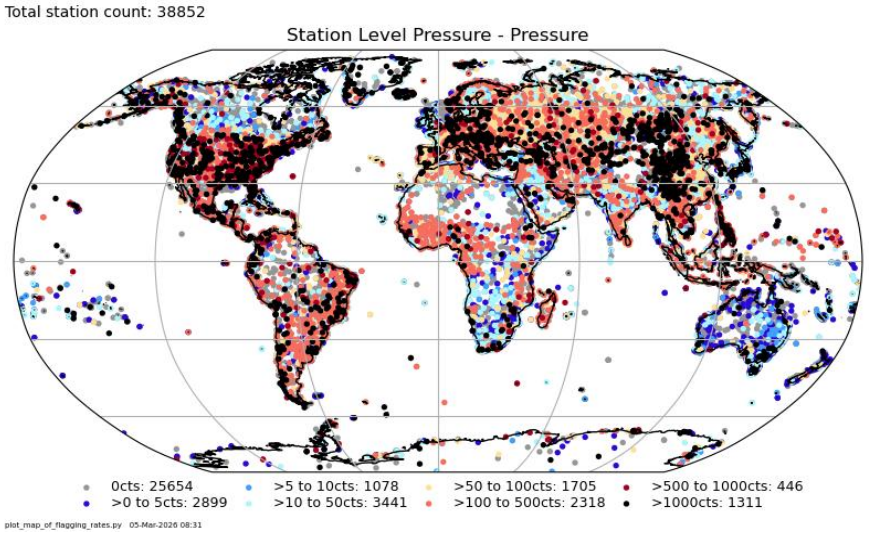
Recent progress in sub-daily QC

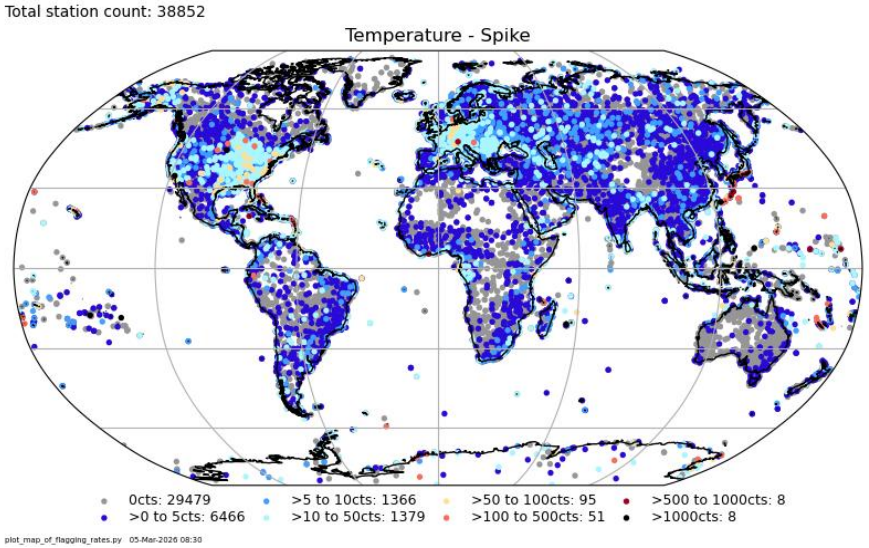
- Aims for improvement in 2025/6 have been on stability & robustness
  - Especially given compute resource challenges over recent releases
  - Unit testing of QC checks now complete
    - Though some utility routines still need additional tests
  - Improved processing on ICHEC compute resource
    - Maximising stable usage of the available CPUs and RAM
    - Better balance of processes (no long waits for final ones to finish)
  - Updated documentation
  - QC took about 6 days for all 39k stations with no issues



Recent progress in sub-daily QC

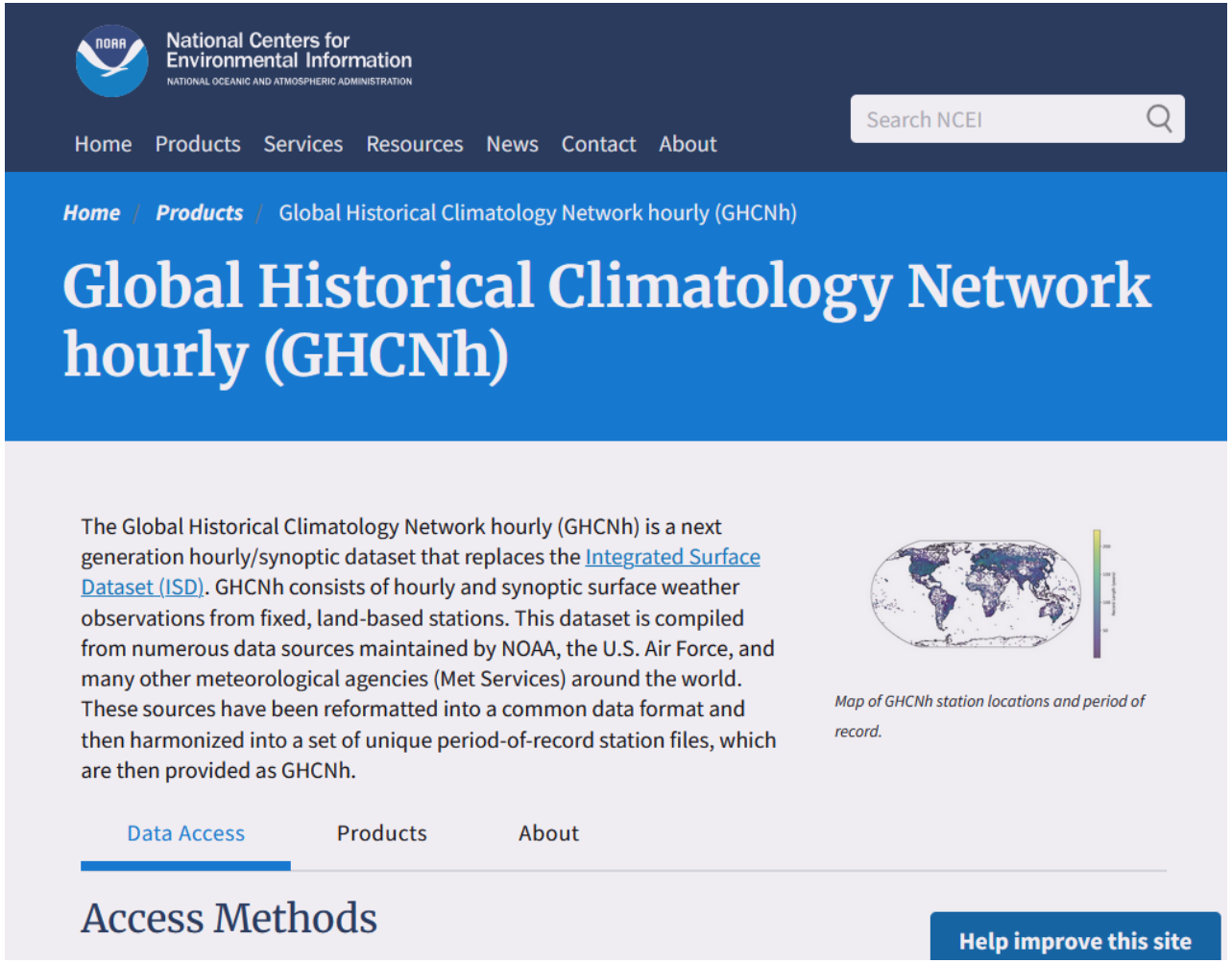
- Harmonisation with existing (legacy/upstream) flags
  - C3S QC ≤R8.0 had upper case and lowercase alpha-numeric chars
  - Upstream flags also had upper case alpha-numeric chars
  - Now C3S QC uses lower case alpha characters only.
- New test flagging where difference between sea-level and station-level pressure inconsistent with elevation
- New test comparing station pressure to that derived from standard normal SLP & station elevation
- Improved spike check (binning time-differences)
- Next steps
  - QC tests on additional variables (e.g. T_w, rh, clouds/base height..)
  - Pre-QC checks to help identify potential issues earlier (e.g. more wind measurement flags)





Partner products

- GHCNh
- Yearly and period of record stations available
- Near-real time updates
- Upstream QC flags
- Potentially better suited for “power-users”



https://www.ncei.noaa.gov/products/global-historical-climatology-network-hourly

Future Plans (next ~ 2yrs)

- All variables already served in GHCNh
  - But not all have been QC'd through this service
- Extension of QC to other variables
    - Wet-bulb temperature & relative humidity
    - Cloud & cloud base
    - Wind gust
    - ….precipitation…?
  - And then inclusion in C3S product.
- Development of near real-time updates
  - Using ECMWF GTS feed/MARS archive

- Ongoing increases in data holdings, including:
  - "MIDAS-Open" (Met Office Integrated Data Archive System) +500 new stations,
  - Data Deposit Service submitted data,
  - Ukraine hourly data.
- Fixes to data issues identified
  - Please do use the data, and report anything odd you find!
  - Lat == Lon != 0
  - Missing elevations
  - Wind measurement codes
  - Merge failures
- Investigate stations not served to C3S due to high flagging rates

PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

Climate Change Service climate.copernicus.eu

IMPLEMENTED BY ECMWF

Thank you