

臺灣區域再分析資料之最新進展

連國淵、鄧雯心、黃子茂、林怡萱⁺、沈彥志⁺、趙子瑩

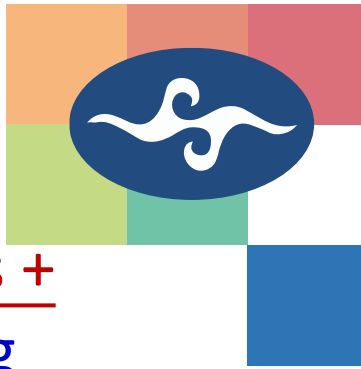
中央氣象署科技發展組

⁺現職單位：交通部民用航空局

2026/09/03

115 年第四十屆天氣分析與預報研討會

(Atmospheric) Reanalysis



- Using a system similar to the operational NWP (model + observations + data assimilation) to perform **retrospective runs**, aiming at generating **analysis fields for a past period**.
- **Purposes:**
 1. To obtain analysis data for historical periods that are likely of higher quality than the operational analyses at the time, by employing a more advanced NWP system nowadays.
 2. To obtain long-term analysis data that are of relatively consistent quality using a fixed-version NWP system.

Note: The quality of the observing systems inevitably varies over time.
- The reanalysis has been even more important in the era of AI weather prediction.

Global and regional reanalysis datasets



- **Global reanalysis:**

- **NCEP:** R1, R2, CFSR
- **ECMWF:** ERA-40, ERA-interim, ERA5
- **JMA:** JRA-55, JRA-3Q

- **Regional reanalysis:**

- **EU:** Copernicus Regional Reanalysis for Europe (CERRA) / Europe area; 5.5-km grid
- **DWD:** COSMO Regional Reanalysis (COSMO-REA) / German area; 2-km grid
- **BOM:** Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA) / Australian area; 12-km grid
- **Japan:** RRJ-ClimCORE / Japan area; 5-km grid

- **Taiwan's regional "reanalysis-like data":**

- **TCCIP/NCDR Taiwan ReAnalysis Downscaling data (TReAD)**
 - WRF dynamical downscaling based on ERA5
 - Taiwan area; 2-km grid; 40+ years (80 years for v2)

CWA's TAiwan ReAnalysis (**TARRA**)

- **Goals:** Deliver a high-resolution regional reanalysis dataset for the Taiwan area to support a variety of applications, e.g.,
 - Weather and climate studies
 - Machine-learning weather prediction (MLWP) model training
- **Concepts and approaches:**
 - A high-resolution convection-allowing model
 - Rapid-update data assimilation of unique convective-scale observations
→ To provide high-temporal-resolution data
 - A regional domain covering the vicinity of Taiwan
 - **[For the 1st version]** A period spanning ~10 years (2017 – 2026)
- **Developed based on CWA RWRF system**

Reanalyses in the Age of Machine Learning: Why Dataset Curation Matters Now More than Ever

Mimi Rose Abel^{},^a Alexander J. Thompson,^{a,b} Ethan D. Gutmann,^c Kelly Mahoney,^a
Rachel Rose McCrary,^c Russ S. Schumacher,^d and Laura C. Slivinski^a

AFFILIATIONS: ^a NOAA/Physical Science Laboratory, Boulder, Colorado; ^b Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado; ^c NSF National Center for Atmospheric Research, Boulder, Colorado; ^d Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

<https://doi.org/10.1175/BAMS-D-25-0149.1>

Some key points from Abel et al. (2026)

- Reanalysis is a highly **labor- and computationally-intensive** task.
 - Producers must balance desired features with available resources, so advanced methods or optimal configurations are not always feasible.
 - Observations that could possibly degrade the analysis are discarded.
- The term “reanalysis” originally had a strict definition, but nowadays an increasing number of “reanalysis-adjacent” products are ambiguously labeled as reanalyses.
 - The authors advocate reserving “reanalysis” for datasets that explicitly combine **models** and **observations** through **data assimilation**.
- Reanalysis is becoming increasingly important in the AI/ML era, but **the potential for its misuse is also growing**.
- The characteristics of reanalysis variables vary substantially:
 - Some variables are strongly constrained by observations, so they are accurate enough and close to observations.
 - Many others are essentially **model-derived quantities**, with limited observational constraint. **Treating these variables as direct observations or “ground truth” can be problematic.**

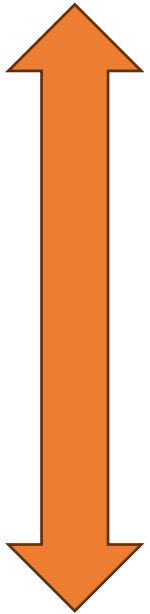
Traditional reanalysis vs. “Reanalysis adjacent” (Abel et al. 2026)

	Category	Examples
Traditional reanalysis (model + observation + data assimilation)	Global atmospheric reanalysis	ERA5, MERRA-2, JRA-3Q
	Reanalysis for adjacent disciplines	Oceanic reanalysis, atmospheric composition reanalysis
	Regional reanalysis	CERRA, RRJ-ClimCORE, TARRA
“Reanalysis adjacent” (No or limited DA)	Downscaling with nudging (or blending) toward global reanalysis	CONUS404, TReAD
	“Replay” method	UFS-Replay
	Reanalysis-forced simulations (often forced by surface fluxes)	ERA5-land

“Variable classification system” (Kalnay et al. 1996)

“The reanalysis gridded fields have been classified into four classes, depending on the relative influence of the observational data and the model on the gridded variable.”

Most
reliable



Less
reliable

Class	Descriptions
A	The analysis variable is strongly influenced by observed data, and hence it is in the most reliable class.
B	Although there are observational data that directly affect the value of the variable, the model also has a very strong influence on the analysis value.
C	There are no observations directly affecting the variable, so that it is derived solely from the model fields forced by the data assimilation to remain close to the atmosphere.
D	A field that is obtained from climatological values and does not depend on the model.

“Variable classification system” (Kalnay et al. 1996)

Pressure-level variables

Class	Field type
A	Zonal wind
A	Meridional wind
A	Geopotential height
A	Virtual temperature
A	Absolute vorticity
B	Vertical velocity (1000–100 hPa only)
B	Specific humidity (1000–300 hPa only)

Single-level variables

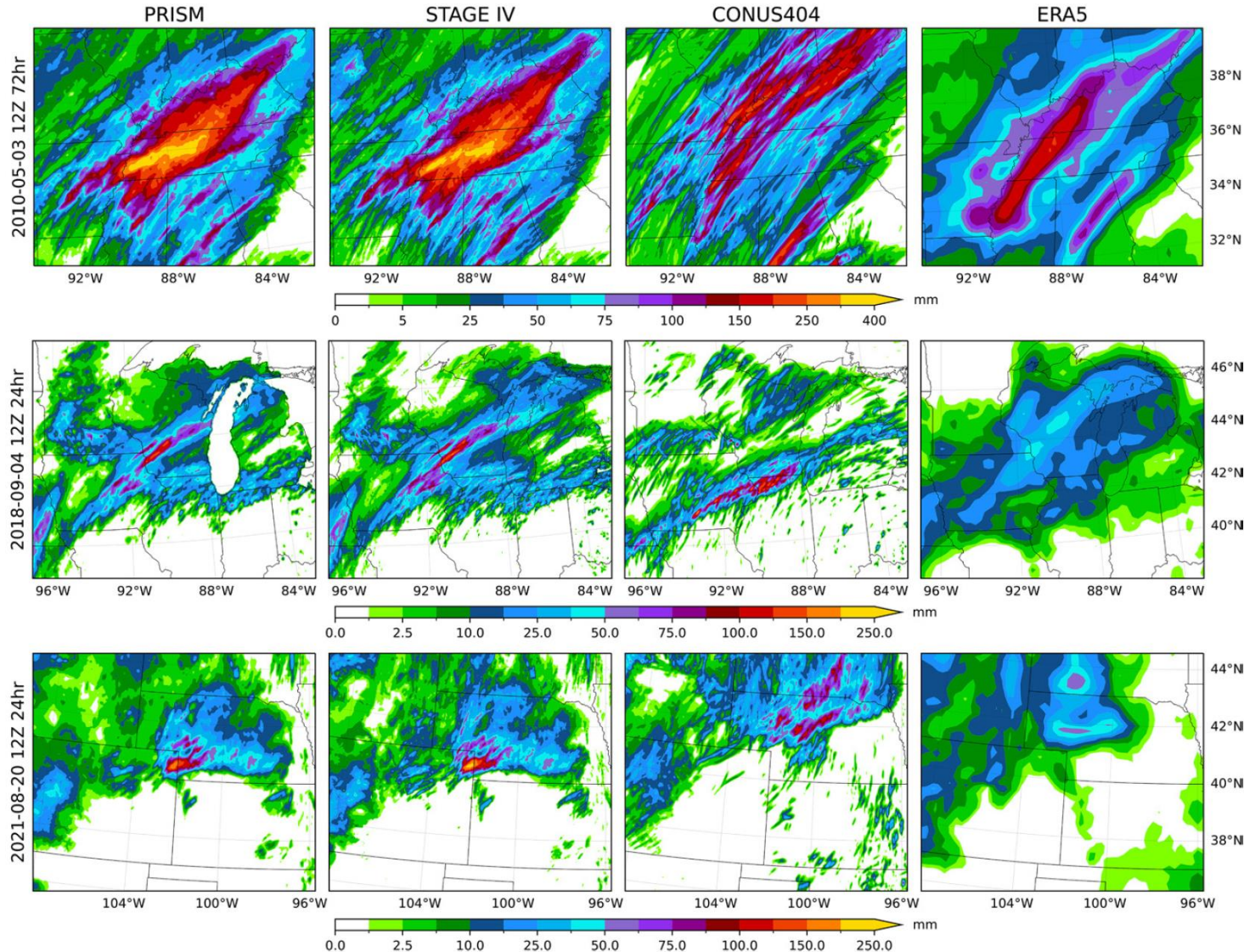
Class	Field type
B	Surface temperature
C	Skin temperature
B	2-m temperature
B	Surface pressure
D	Albedo
C	Surface sensitivity and latent fluxes
C	Top-of-the-atmosphere fluxes
B	Zonal wind at 10 m
B	Meridional wind at 10 m
C	Surface wind stress
A	Mean sea level pressure
B	Precipitable water
C	Snow depth
B	Snow cover
C	Precipitation (total and convective)
B	Mean relative humidity (multiple layers)
C	Soil wetness and temperature
C	Surface runoff
C	Cloud fraction (high, middle, low)
C	Cloud forcing, clear-sky fluxes
C	Gravity wave drag
B	Max and min temperature

Tables 3 and 4 in Kalnay et al. (1996)

Example: Issue with precipitation variables in reanalysis

Observations-based

Reanalysis/Reanalysis-adjacent



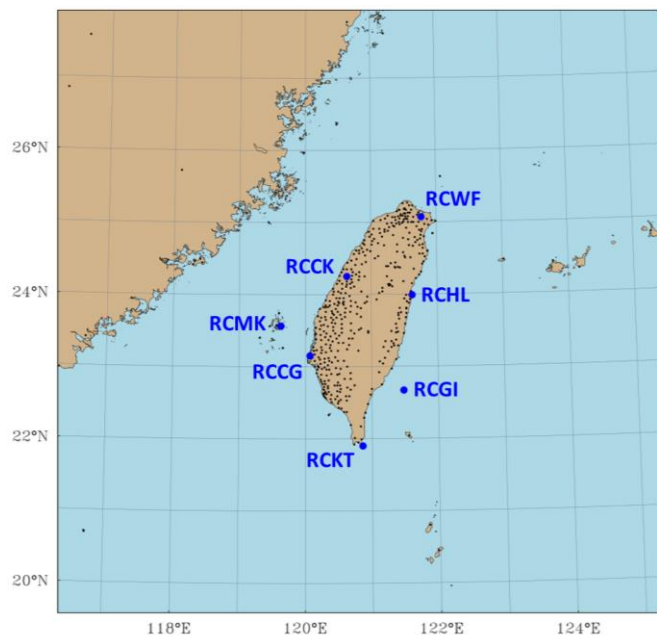
Generally speaking,
**we should avoid using precipitation
variables from reanalysis data
(when better alternatives are available).**

- In observation-rich areas, gauge-based or radar-gauge-blended products such as STAGE IV (U.S.) or QPESUMS (Taiwan) should be much better “precipitation observations.”

Figure 1 in Abel et al. (2026)

FIG. 1. Precipitation totals from PRISM, Stage IV, CONUS404, and ERA5 on (top) 72-h accumulation ending 1200 UTC 3 May 2010, (middle) 24-h accumulation ending 1200 UTC 4 Sep 2018, and (bottom) 24-h accumulation ending 1200 UTC 20 Aug 2021.

CWA RWRF system



- Based on WRF and WRFDA
- **2-km resolution** for the Taiwan area
- **Hybrid 3DEnVar** coupled with RWRF-LETKF members
 - 30-minute cycle for radar DA
 - Hourly cycle for surface DA
- **6-hourly blending with the GFS downscaling**
- 13-hour forecasts initialized every hour
- **Radar data assimilation:**
 - Reflectivity (through hydrometeor retrievals) and radial velocity
- **Surface data assimilation:**
 - Temperature, humidity, winds

Observations

Surface stations

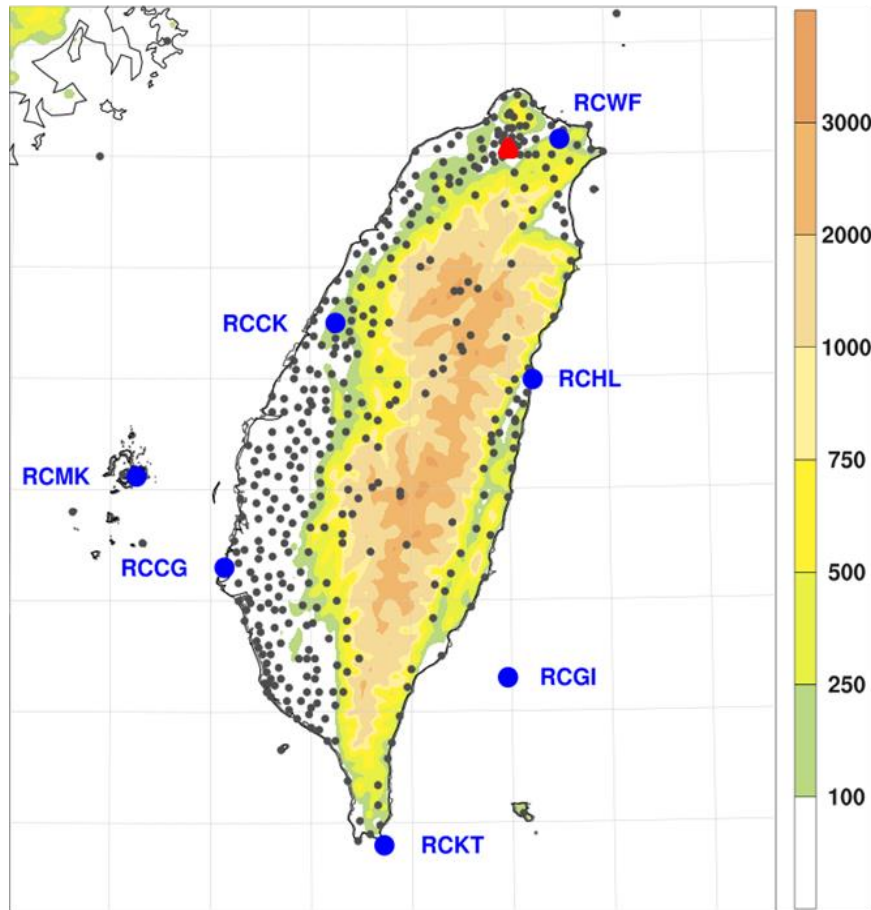
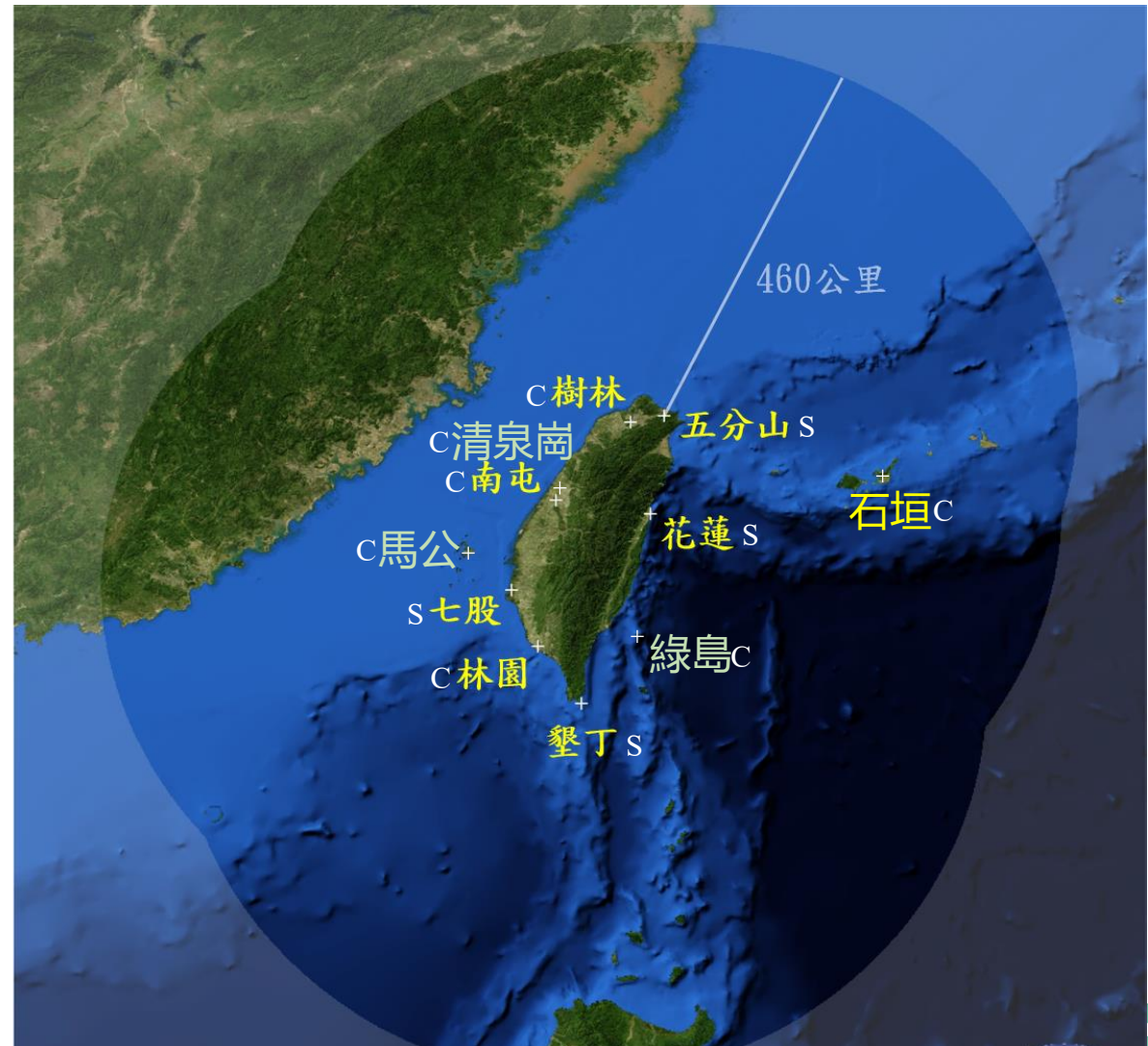


Figure 4 of Chen et al. (2020)

Thank the CWA radar team for helping restore some data missing in the operational archive.

Radar sites



Development of RWRF-based regional reanalysis



- **Most things are based on a decade of development of CWA RWRF:**
 - Radar data assimilation (蔡等人 2019, 大氣科學)
 - Surface data assimilation (Chen et al. 2020, WAF)
 - Large-scale blending (Jiang et al. 2021, TAO)
 - Improved surface observation operators (沈等人 2023, 天氣分析與預報研討會)
- **For reanalysis purpose, several additional developments have been further undertaken:**
 - **Develop an “incremental blending update” method** to remove the discontinuity introduced by the 6-hourly blending scheme in RWRF, and thus **improve the temporal smoothness**. (沈等人 2025, 天氣分析與預報研討會)
 - **Redesign the RWRF scripts** to make them suitable for long-term production (e.g., optimizing disk usage, automation, and efficiency), and also **improve the integration with verification tools**.

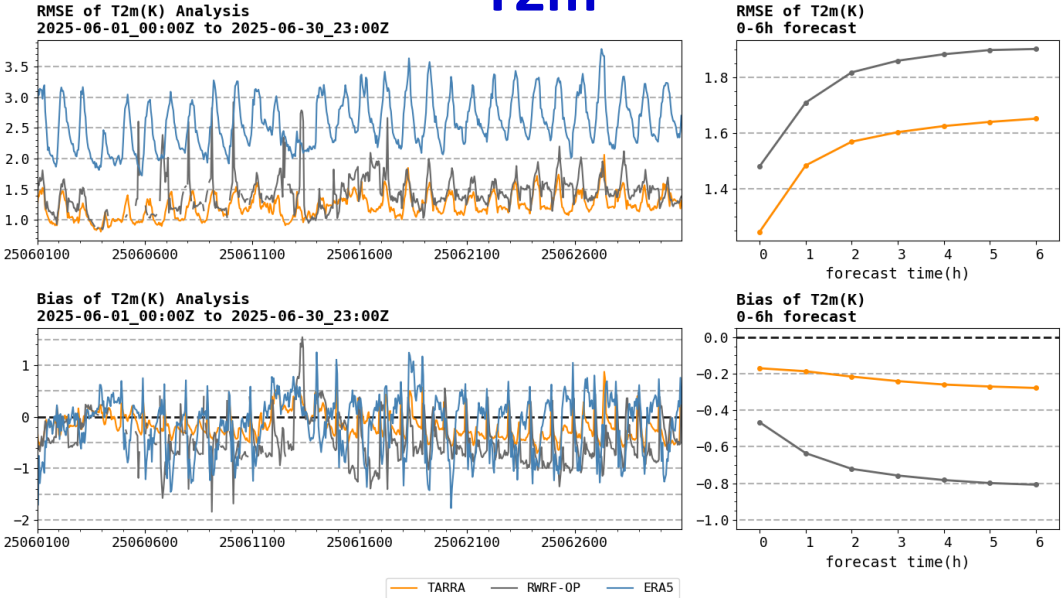
Taiwan Regional ReAnalysis (TARRA): Comparison of operational RWRF and the reanalysis system

		Operational RWRF * Only list recent two major versions		Reanalysis
		2021 – May 2024	June 2024 –	TARRA version 1
Model		CWA OP441 (based on WRF v3.8.1)	CWA OP50 (based on WRF v4.4.2)	CWA OP51 (based on WRF v4.4.2)
Radar DA (reflectivity & radial velocity)	Method	3DVar	Hybrid 3DEnVar	3DVar
	Cycle length	30 minutes		1 hour
Surface DA	Method	Hybrid 3DEnVar		3DVar
	Cycle length	1 hour		1 hour
	Observation operator	Default (via modified observed values)		Improved method (via model diagnosed values)
	Digital filter initialization	V		—
Incremental analysis update (IAU)		—		V (10-min window)
Large-scale blending		V Traditional (instantaneous)		V Incremental
Boundary conditions (lateral & surface) & data for blending		NCEP GFS → WRF 10-km domain forecast		ERA5

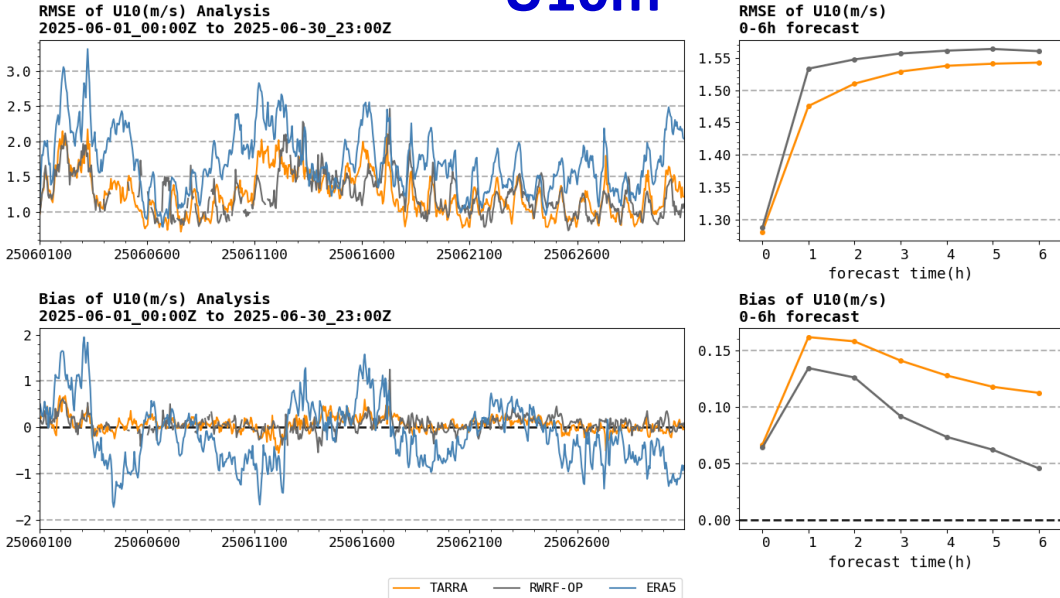
Verification – June 2025



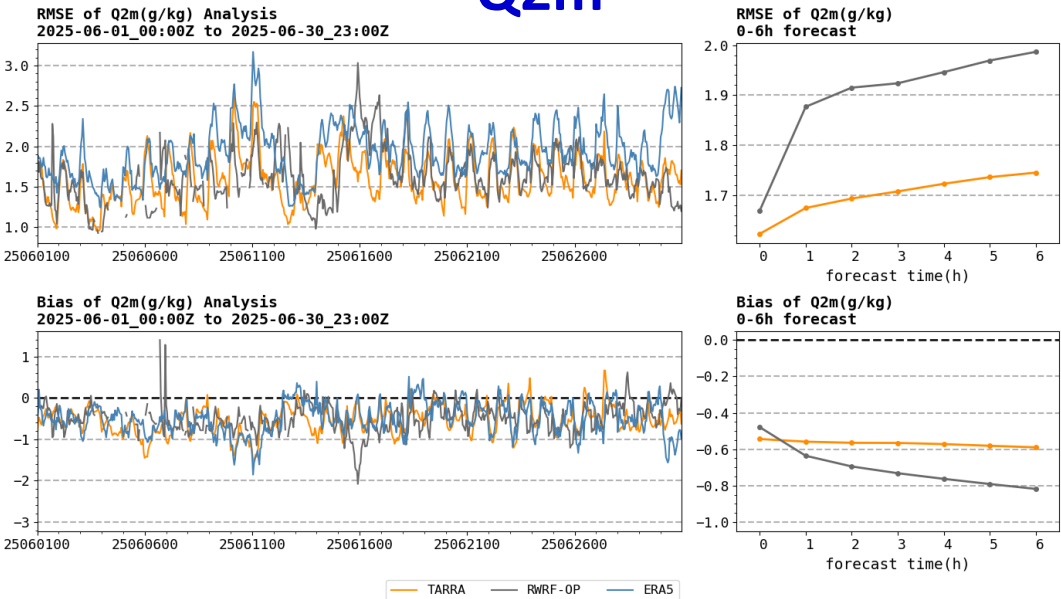
T2m



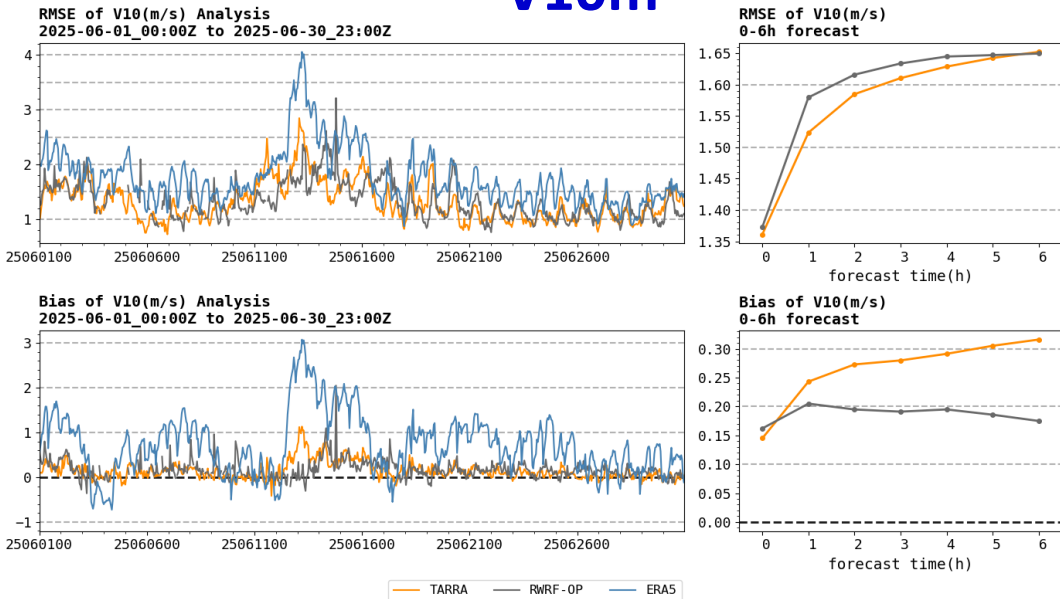
U10m



Q2m



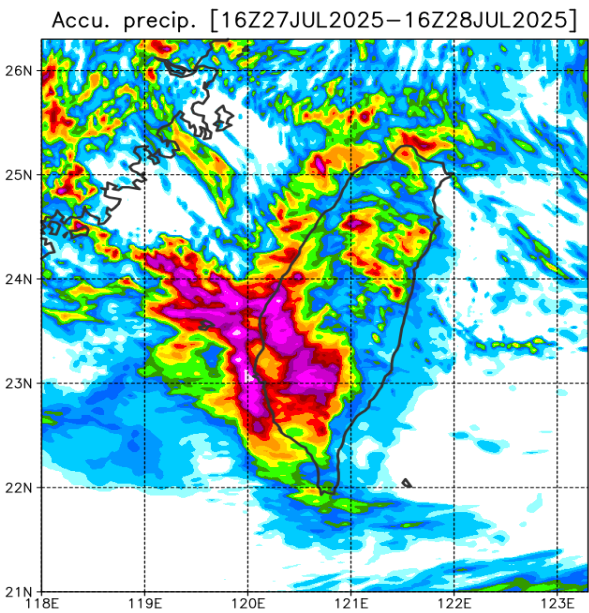
V10m



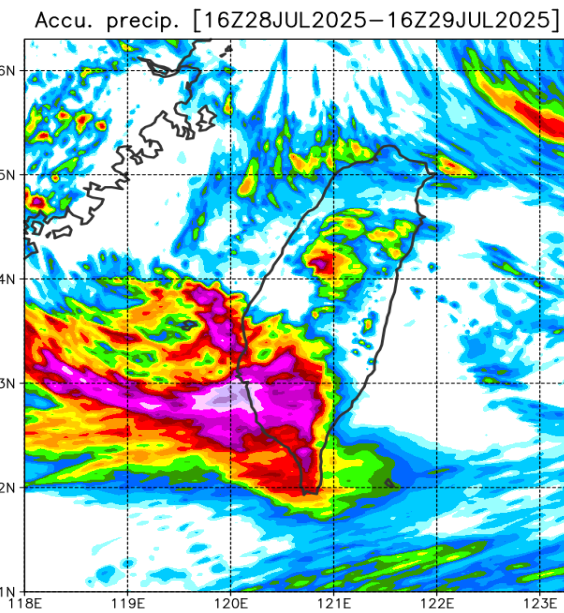
Verification – Southwesterly flow heavy rainfall event (28 July – 3 August 2025)

Top: Sum of 0–1 h precipitation forecast from each cycle (i.e., Sum over 24 fcst. to obtain daily accumulation)
Bottom: Observation

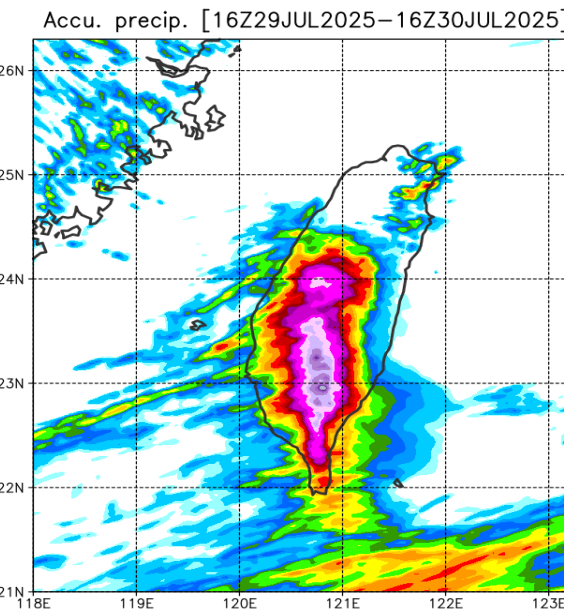
28 July 2025



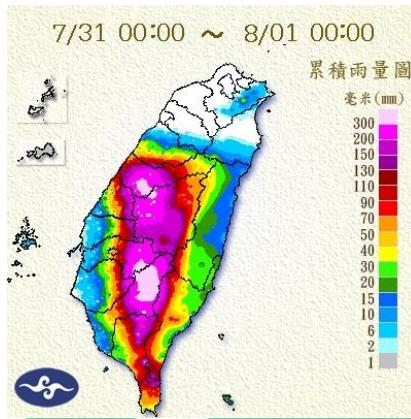
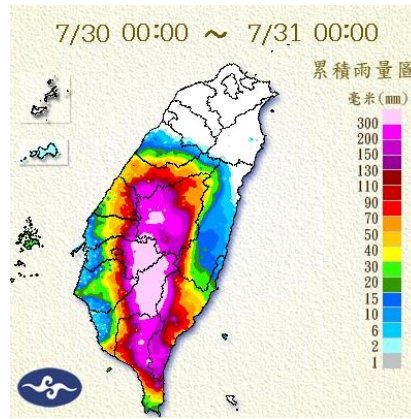
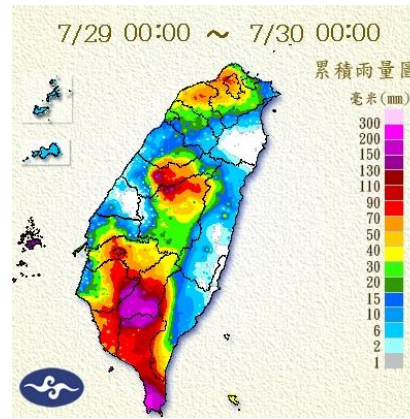
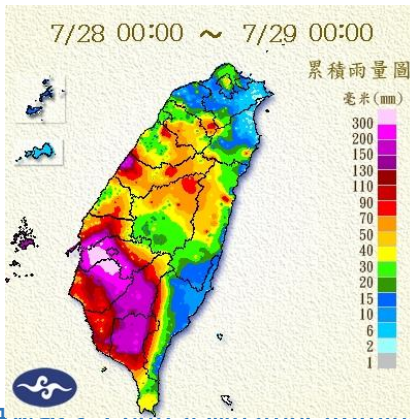
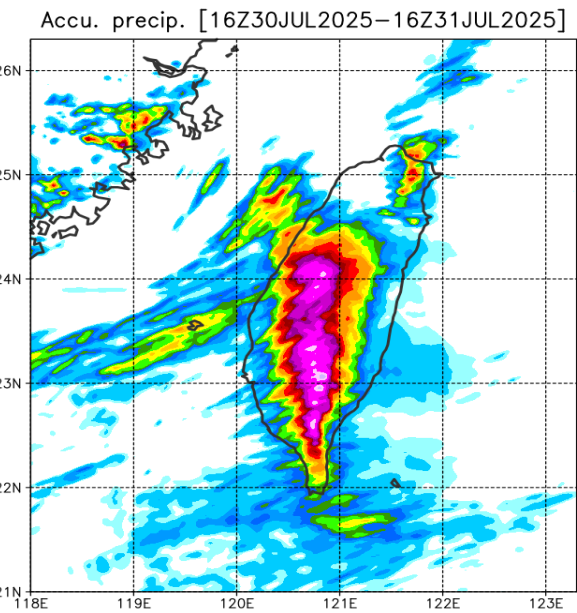
29 July 2025



30 July 2025



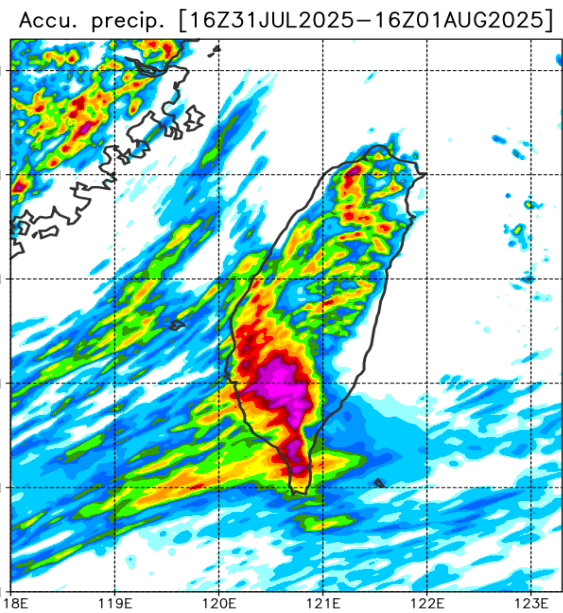
31 July 2025



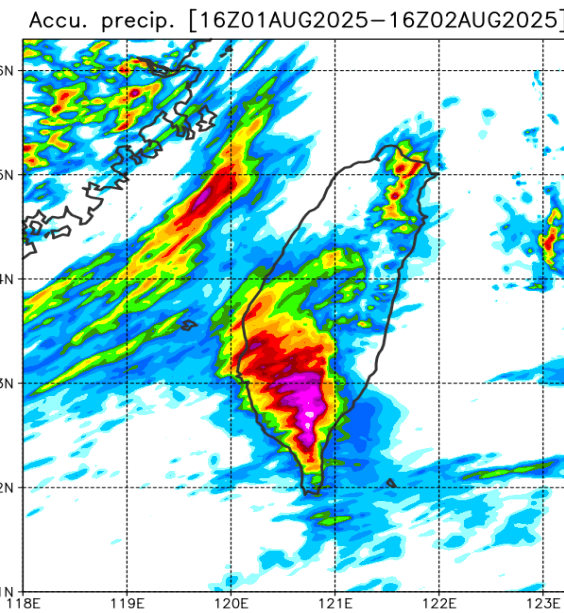
Verification – Southwesterly flow heavy rainfall event (28 July – 3 August 2025)

Top: Sum of 0–1 h precipitation forecast from each cycle (i.e., Sum over 24 fcst. to obtain daily accumulation)
Bottom: Observation

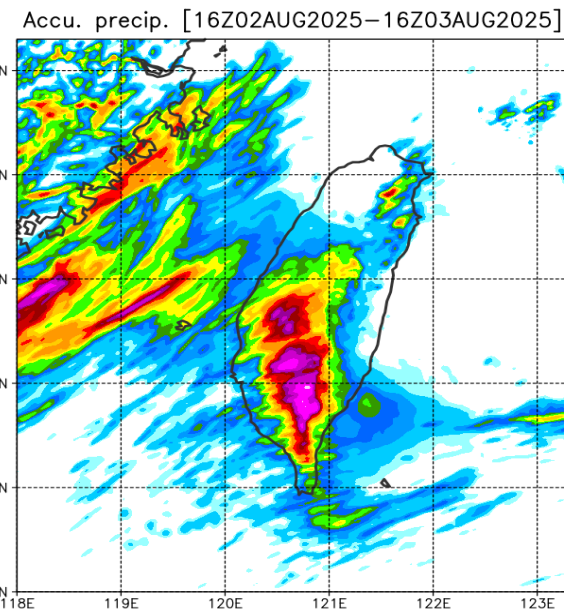
1 August 2025



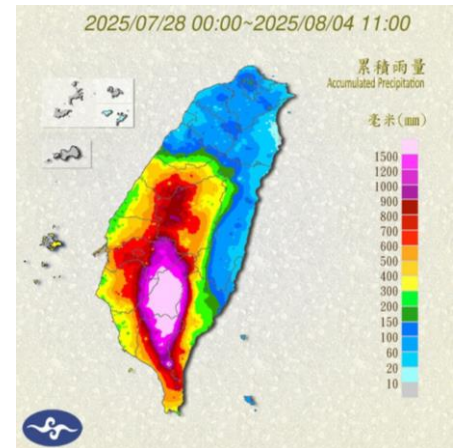
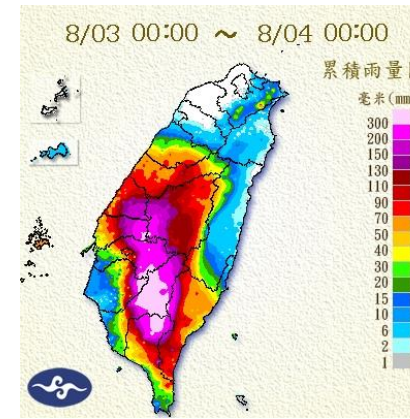
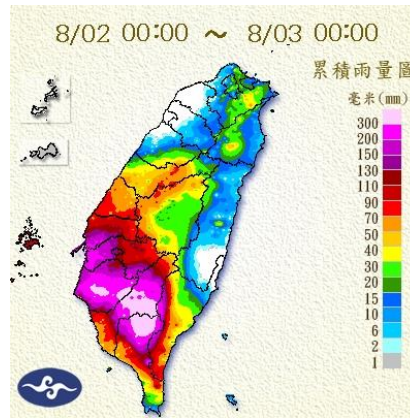
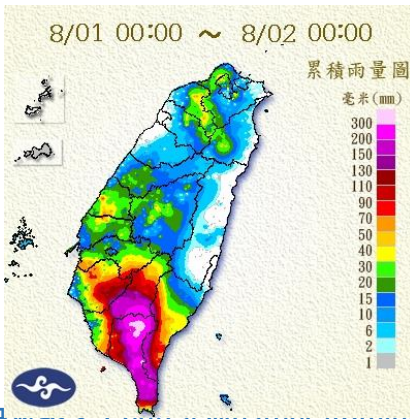
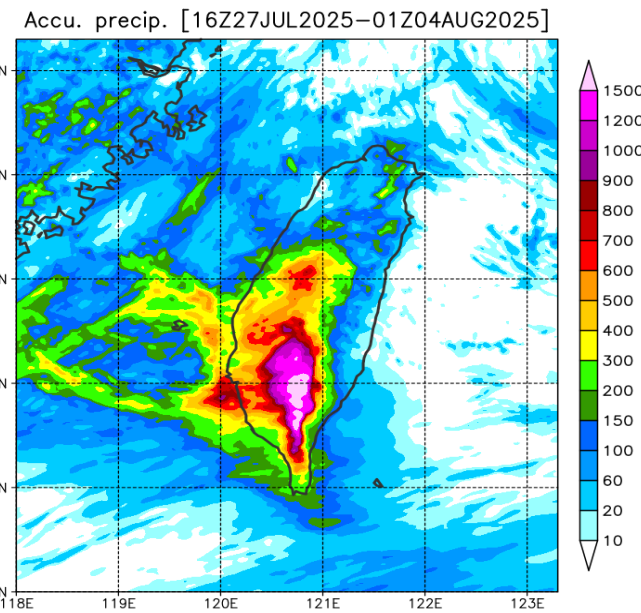
2 August 2025



3 August 2025



Total:
27 July – 4 August 2025



TARRA production status

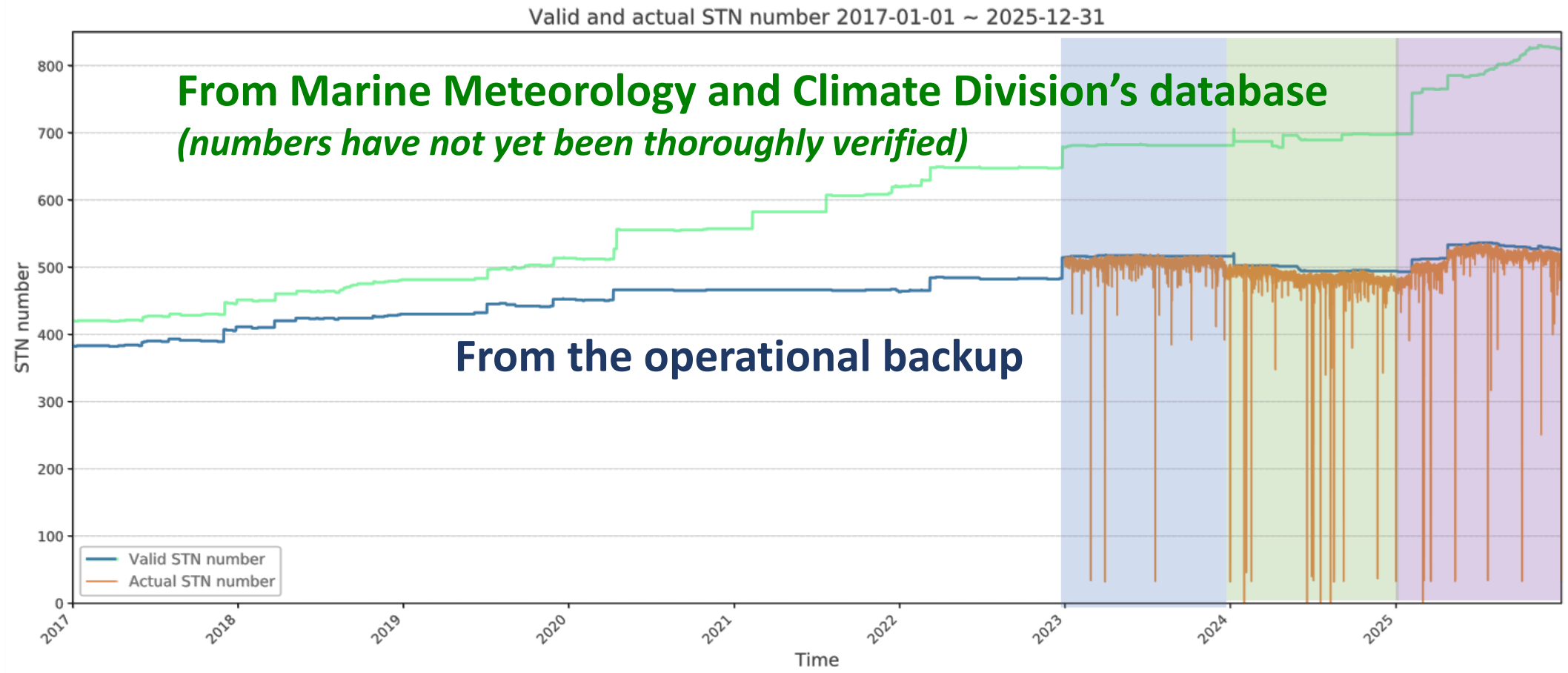


- **Finished production of the 2025 data !!**
 - Verification results indicate that **the data quality meets expectations.**
- **However, some issues were found (totally unsurprisingly) during the 2025 production:**
 - Some “artifacts” in the near-coastal SST field
 - Short-wave noises in some fields
 - Different sources of surface observations
- **Plan:**
 - 1) Withhold the production of the next period
 - 2) Adjust the reanalysis system to fix the issues (as far as possible)
 - 3) Continue the production of other years (2021 – 2024)
 - 4) Redo the 2025 data (possibly early next year)

Different sources of surface observations



- A more complete surface observation dataset may be used in the assimilation.



Summary



- CWA has been working on producing a high-resolution regional reanalysis for Taiwan.
 - Based on CWA RWRF: **2-km resolution**; hourly radar and surface data assimilation
 - Special emphasis on ensuring *temporal smoothness*.
- Comprehensive verification has shown satisfactory data quality. [A2-17 鄧雯心]
- Timeline:
 - Production of **the 2025 data** has been finished.
 - However, production is currently withheld to resolve some issues.
 - 3+ years of data are expected by the end of 2026.
- This regional reanalysis, once completed, will constitute **the first dataset of its kind for the Taiwan region** and is expected to support a wide range of applications.