Development of global climate datasets in CMA supporting data requirement of TPRCC

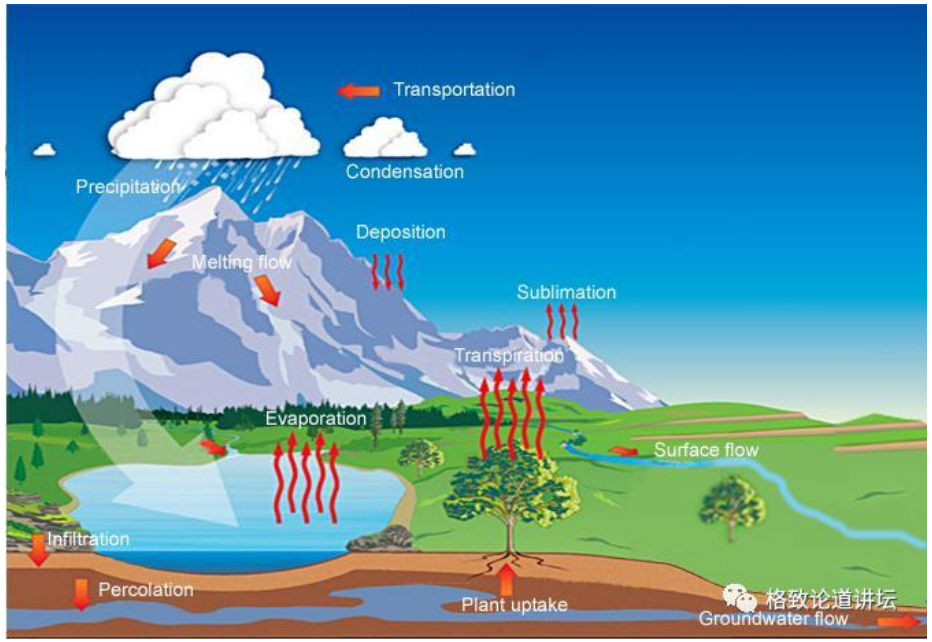
Su YANG
National Meteorological Information Centre, CMA
June 2024, Lijiang

国家气象信息中心
National Meteorological Information Centre

Contents

Back Ground

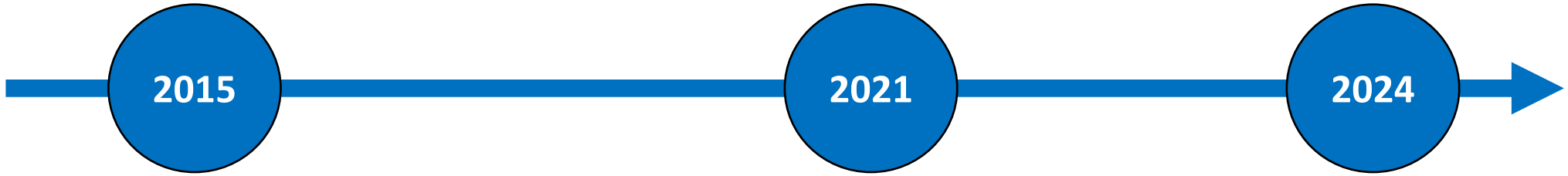
- Data have been more and more crucial for our lives, and also for global and regional climate works, such as the study on TPR (Third Pole Region).
- The demands for data over TPR are much greater than other regions because of it's important role in global climates change and it's shortage of data for long time.



Main processes in global climate data by NMIC/CMA

China Meteorological Administration (CMA) set up a research team led by NMIC focusing on the development of global climate products including global base data and reanalysis products (CRA).

The global base data product develops a suit of datasets and the CRA improves it's assimilation scheme and update the version of product to 1.5.

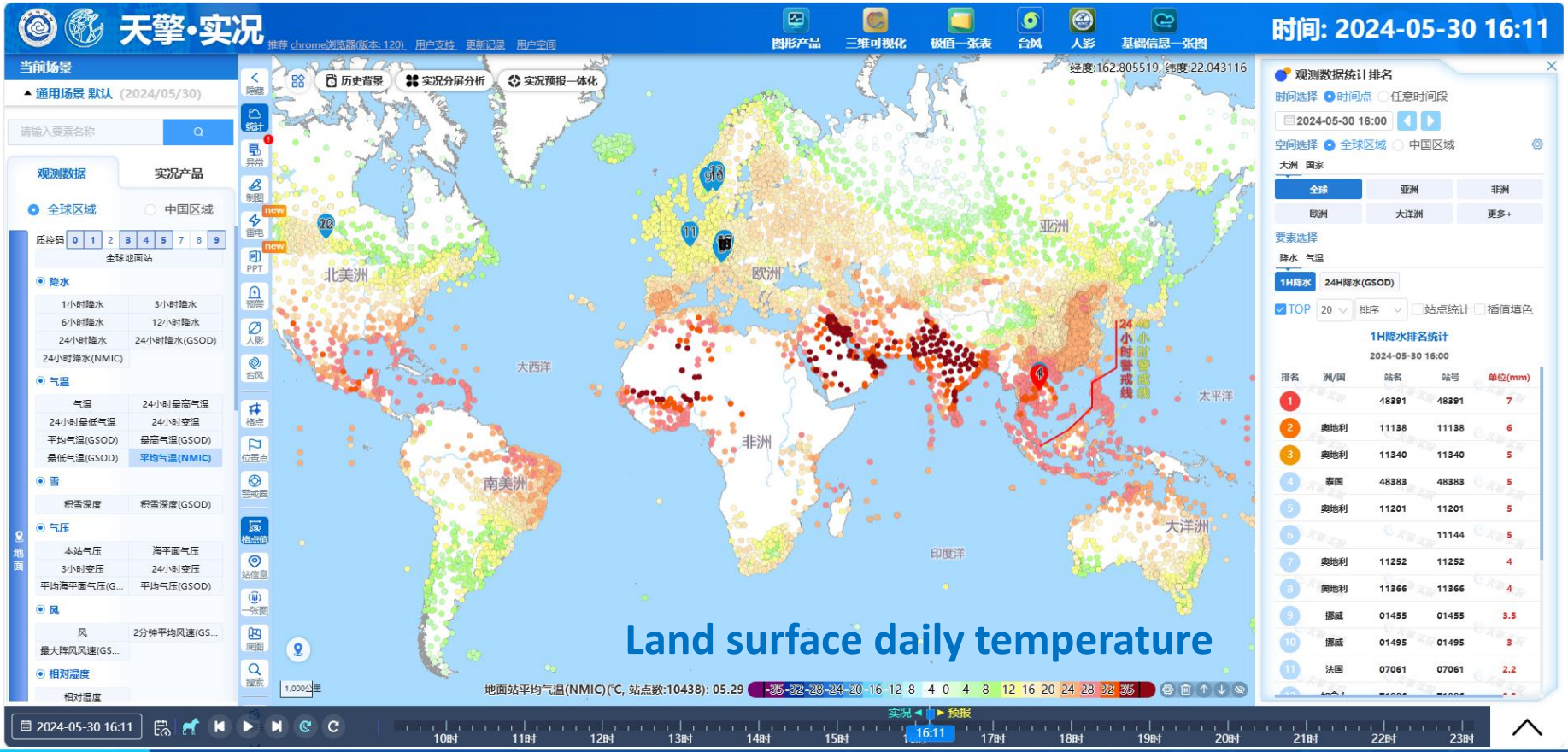


The primary global base data product and first generation CRA were accomplished.

Contents

The development of global base data product in CMA

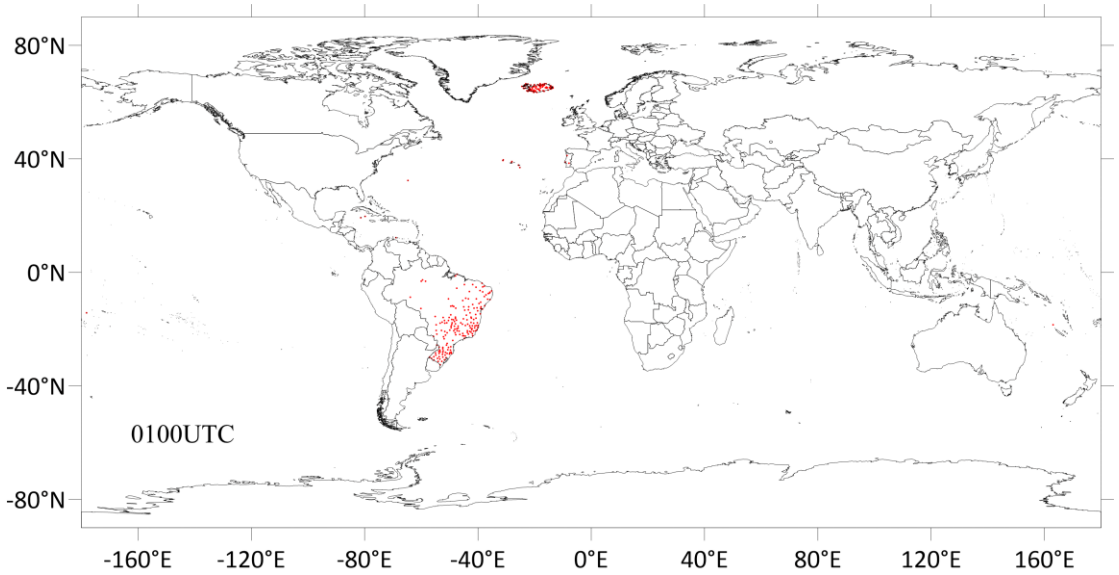
- The base data (in-situ) have always been the foundation of climate change monitoring, the referees of models and even gradually being the driving factors of the model, e.g. AI.
- We aim to establish a global base data producing system that contains full meteorological measurements, updates data near-real time, and provides substantial variabilities and items.



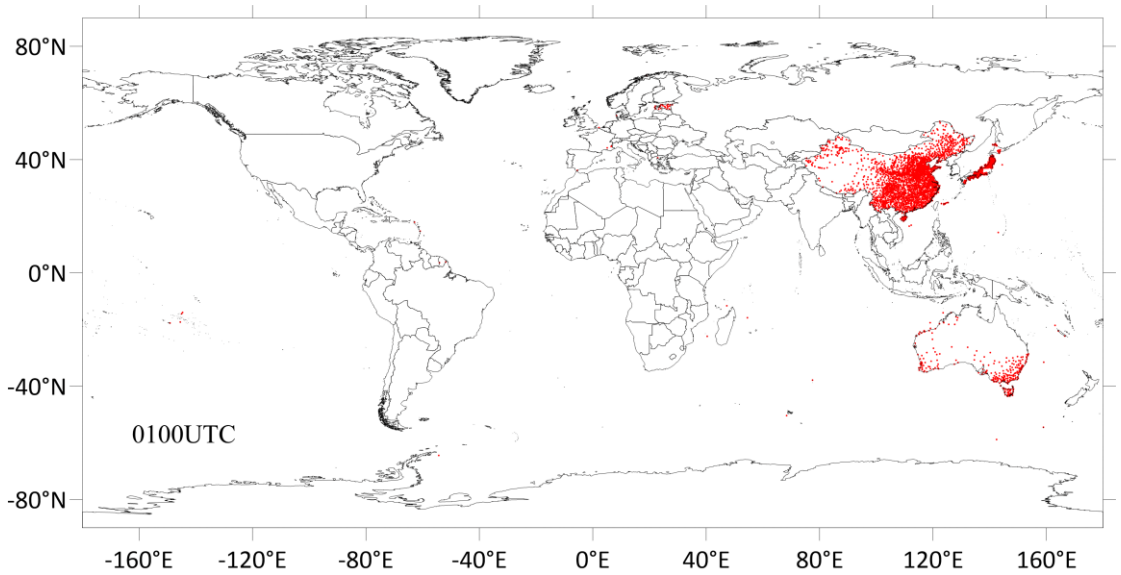
Keys determining the global data amounts and quality

- **1. Converting the sub-daily records to daily data fully**

- Billions of sub-daily synoptic reports are transmitted around the world everyday via GTS (and the forthcoming WIS 2.0). This is one of the most important way of global data exchange.
- It is not an easy task to transform them into daily data because the sub-daily reports are always recorded at diverse observational time by different meteorological data centers .



The spatial distribution of **Tmin-12h** data at each clock in 2022

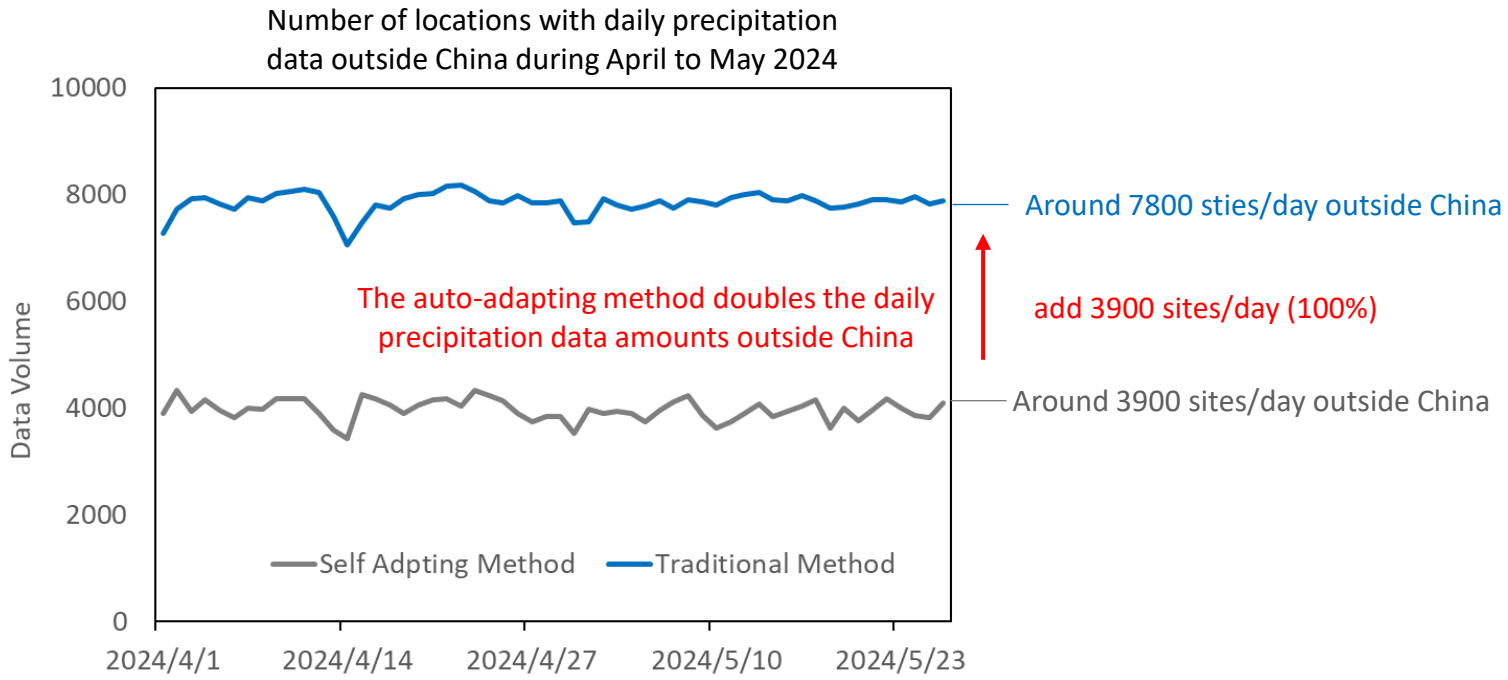


The spatial distribution of **Tmin-24h** data at each clock in 2022

Keys determining the global data amounts and quality

More daily data derived from the synoptic reports

- In order to obtain as much daily data from the synoptic reports as possible, we have developed an **'auto-adapting'** method.
- It decomposes the sub-daily records into shorter ones and rearranges them into daily values.
- Remarkably more daily data are obtained, especially for the precipitation amounts (double).



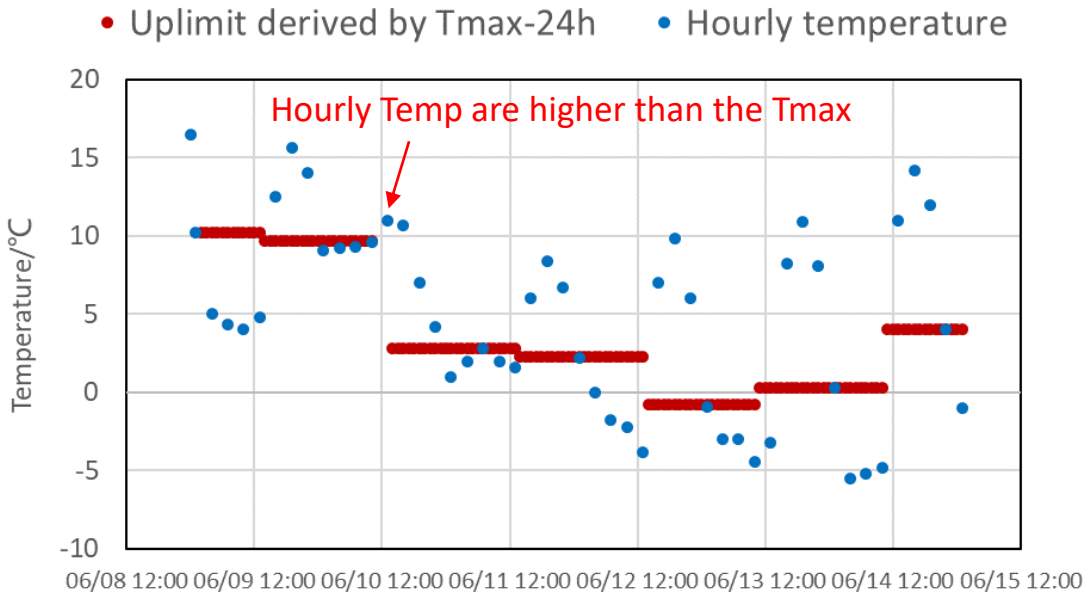
				1h	3h	6h	12h	24h
2016	1	1	18	0	999999	999999	0	999999
2016	1	1	19	0	999999	999999	999999	999999
2016	1	1	20	0	999999	999999	999999	999999
2016	1	1	21	999999	999999	999999	999999	999999
2016	1	1	22	0	999999	999999	999999	999999
2016	1	1	23	0	999999	999999	999999	999999
2016	1	2	0	999999	999999	0	999999	999999
2016	1	2	1	0	999999	999999	999999	999999
2016	1	2	2	0	999999	999999	999999	999999
2016	1	2	3	999999	999999	999999	999999	999999
2016	1	2	4	0	999999	999999	999999	999999
2016	1	2	5	0	999999	999999	999999	999999
2016	1	2	6	0	999999	999999	0	0
2016	1	2	7	0	999999	999999	999999	999999
2016	1	2	8	0	999999	999999	999999	999999
2016	1	2	9	999999	999999	999999	999999	999999
2016	1	2	10	−10	999999	999999	999999	999999
2016	1	2	11	−10	999999	999999	999999	999999
2016	1	2	12	0	999999	−10	999999	999999
2016	1	2	13	−10	999999	999999	999999	999999
2016	1	2	14	0	999999	999999	999999	999999
2016	1	2	15	999999	999999	999999	999999	999999
2016	1	2	16	0	999999	999999	999999	999999
2016	1	2	17	0	999999	999999	999999	999999
2016	1	2	18	0	999999	999999	−10	999999
2016	1	2	19	0	999999	999999	999999	999999
2016	1	2	20	0	999999	999999	999999	999999
2016	1	2	21	999999	999999	999999	999999	999999
2016	1	2	22	0	999999	999999	999999	999999
2016	1	2	23	0	999999	999999	999999	999999
2016	1	3	0	0	999999	0	999999	999999
2016	1	3	1	−10	999999	999999	999999	999999
2016	1	3	2	−10	999999	999999	999999	999999
2016	1	3	3	999999	999999	999999	999999	999999
2016	1	3	4	1	999999	999999	999999	999999
2016	1	3	5	1	999999	999999	999999	999999
2016	1	3	6	6	999999	999999	8	8

Su Yang, et al., 2019. IJOC. DOI: 10.1002/joc.6367

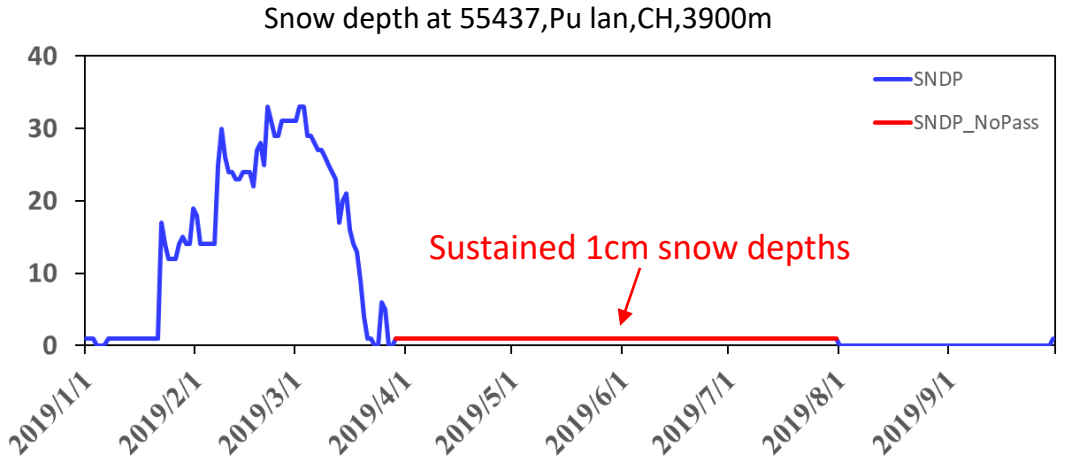
Keys determining the global data amounts and quality

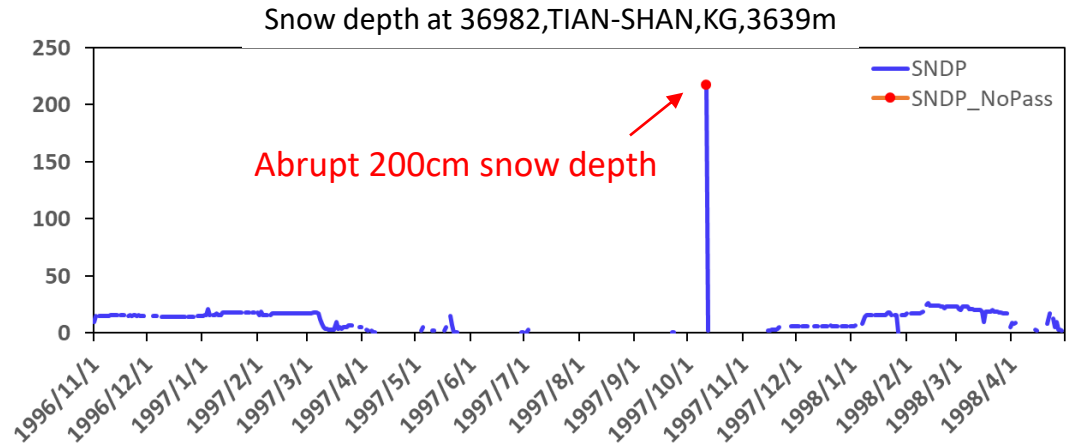
2. Reduce the hurt to data quality and value caused by incorrect records

- The odd records always come as surprises, e.g. typing errors, sustained or abrupt values, etc.
- In order to identify these errors effectively, quality control methods have been developed for them (Temp, Prcp, Snow Depth and so on), separately.



Type Tmax-12h as Tmax-24h at 87448, Villa Reynolds, AR in 2018





Key tasks in global base data producing

Eliminate the influences of inhomogeneity issues on historic data

- Inhomogeneity is another trouble for long-term measurements(series). It can even fake climate changes.
- A homogenizing method including mathematical tests (Rhtests[1]) and metadata checks has been applied to eliminate these non-climatic changes caused by relocations, instrument and observation schedule changes.

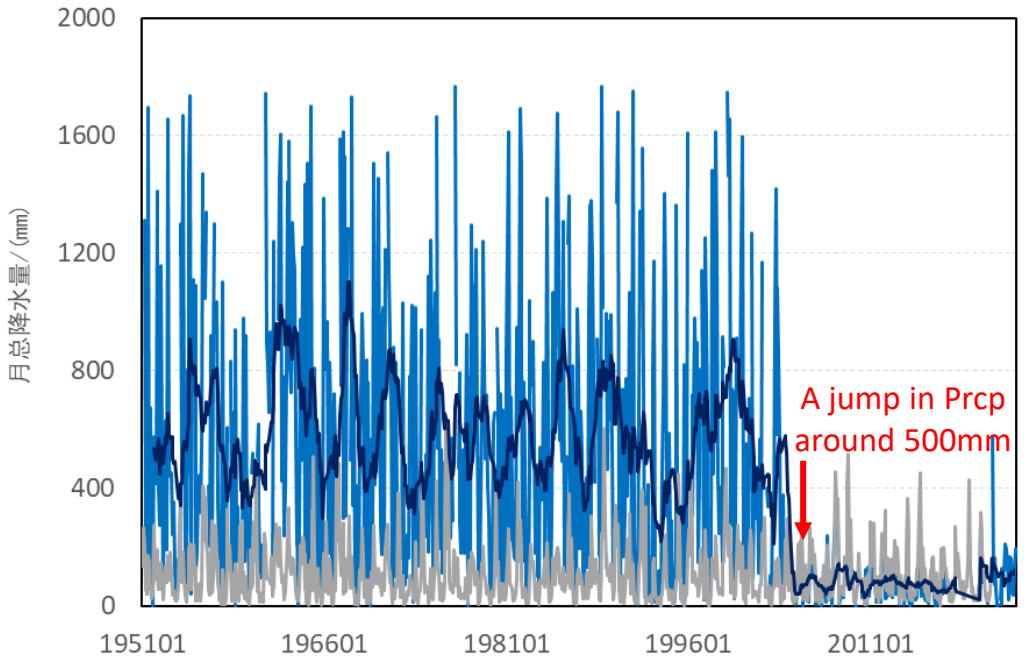


Fig. Koumac, New Caledonia monthly accumulated precipitation suddenly dropped around 500mm (from 580mm to 79mm) in 2003 that did not occur at nearby stations

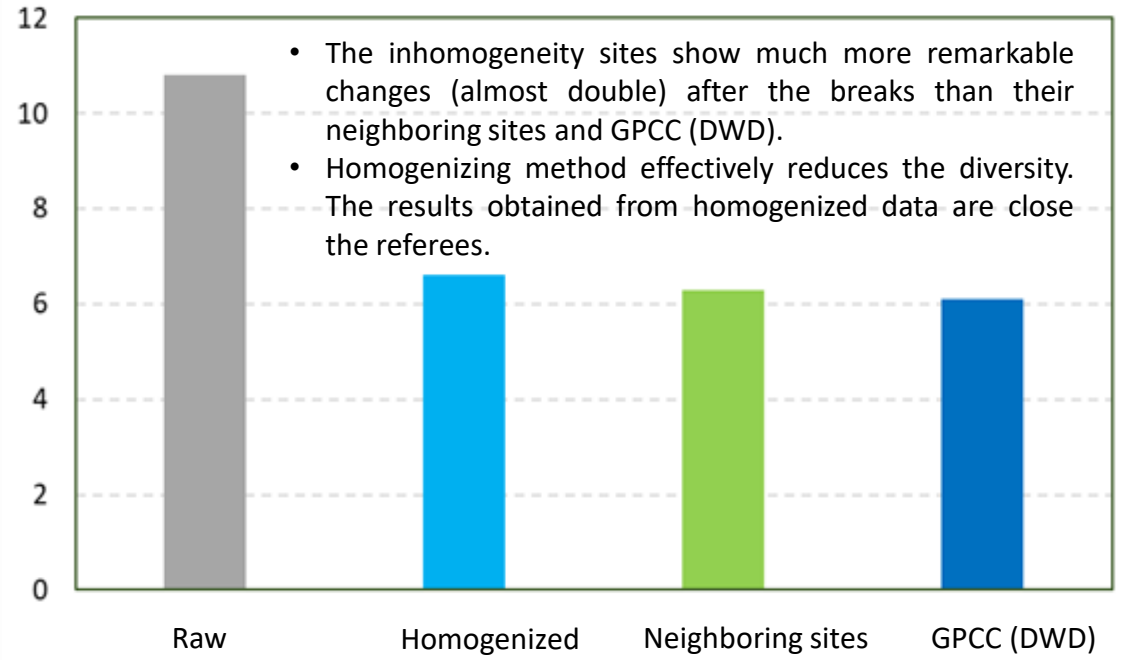


Fig. The average difference in monthly precipitation from 842 inhomogeneous sites between the segments before and after the breaks

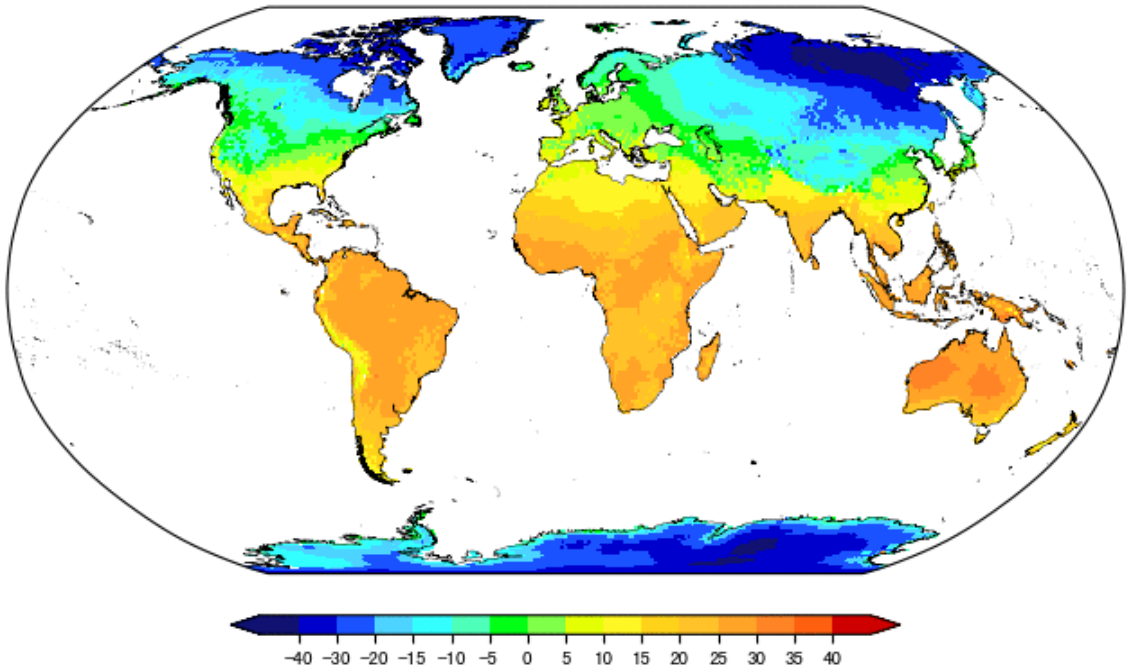
1. Wang XL. 2007, 2008; Wang XL and Feng, 2013

The datasets composing of the global base data product in NMIC/CMA

- Now, global base datasets have in-situ and gridding spatial forms and 8 meteorological items.
- Most of them update every day/month.

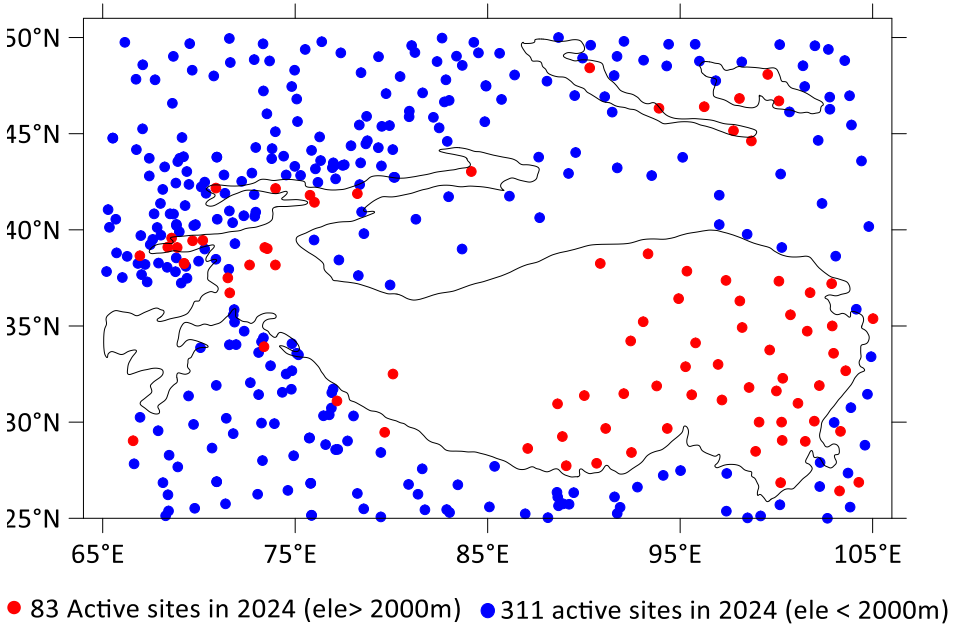
Data type	Spatial form	Item
1. Near-real time dataset	In-situ, Around 10000 sites (day)$^{-1}$	Pres, Tave, Tmax, Tmin, Prcp, Windspeed
	0.5 × 0.5° Gridding (coming)	Tave, Prcp
2. Full dataset	In-situ, Around 19000 sites (day/mon)$^{-1}$	Tave, Prcp
	1.0 × 1.0° ; 2.0 × 2.0° Gridding	Tave
3. Analysis dataset	In-situ, Around 10000 sites	Climatic Normal Extreme records

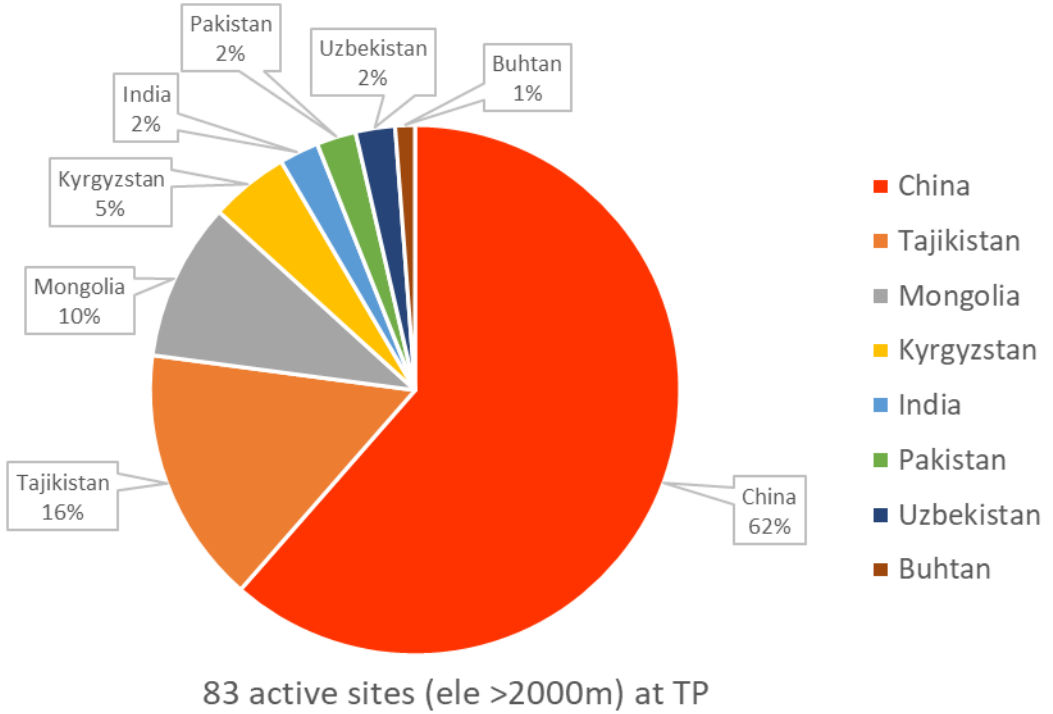
Global land 1 × 1° monthly Tave



The base data in TPR

- There are about 390 active sites in 2024, 83 of which are located at above 2000m.
- Most of the active sites above 2000m are situated in China, Tajikistan and Mongolia.

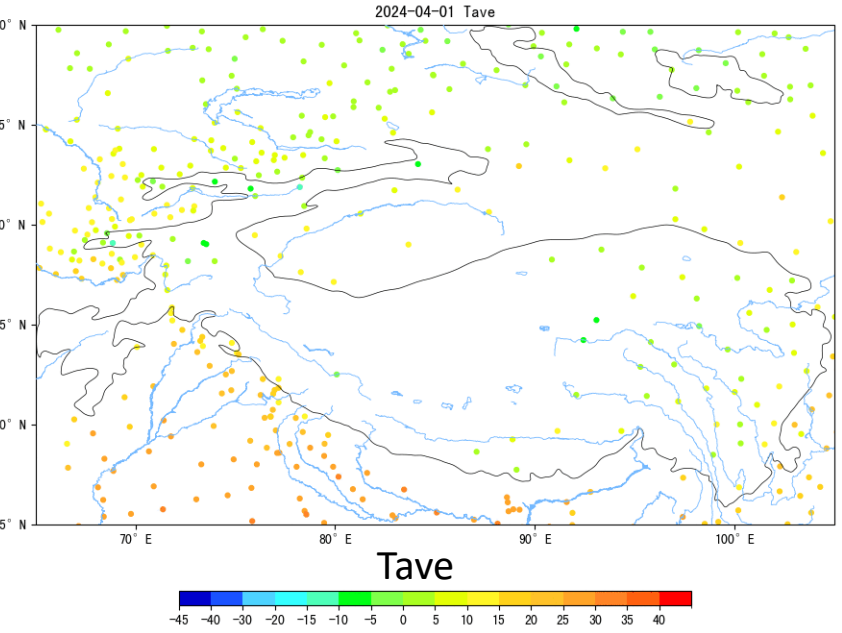


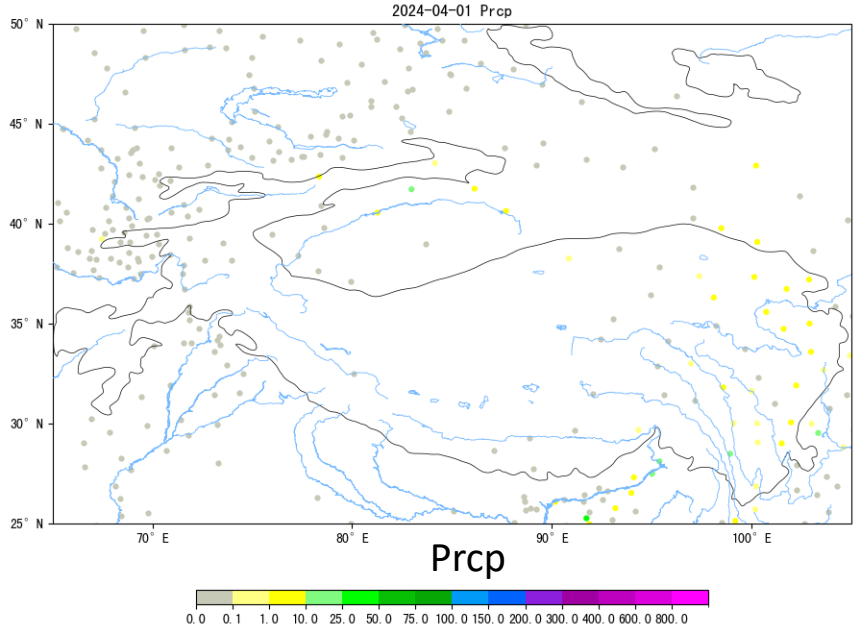


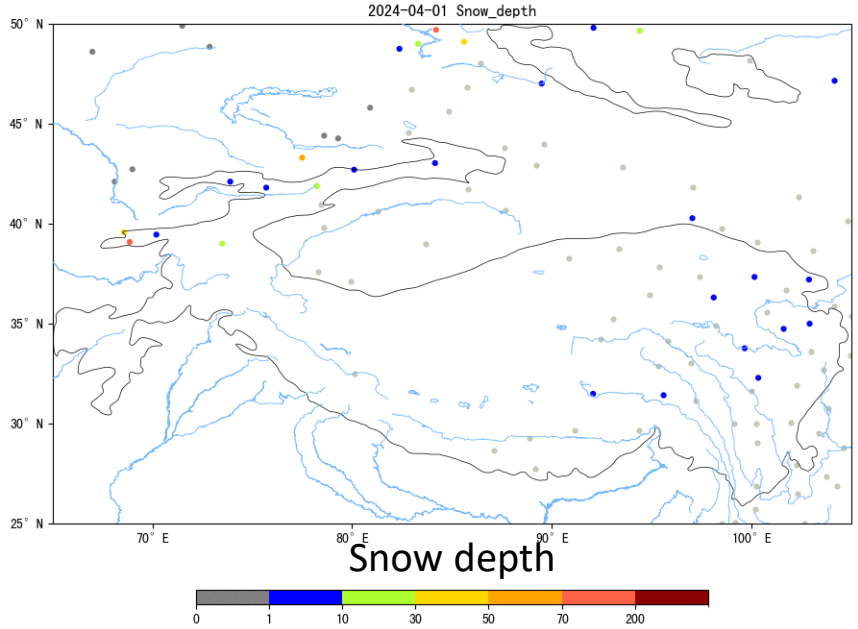
83 active sites (ele >2000m) at TP

The daily data in TPR

- The in-situ daily data in TPR update everyday.
- Tave shows higher spatial density and better temporal continuity in TPR than other meteorological variabilities, e.g. Prcp, Snow depth.



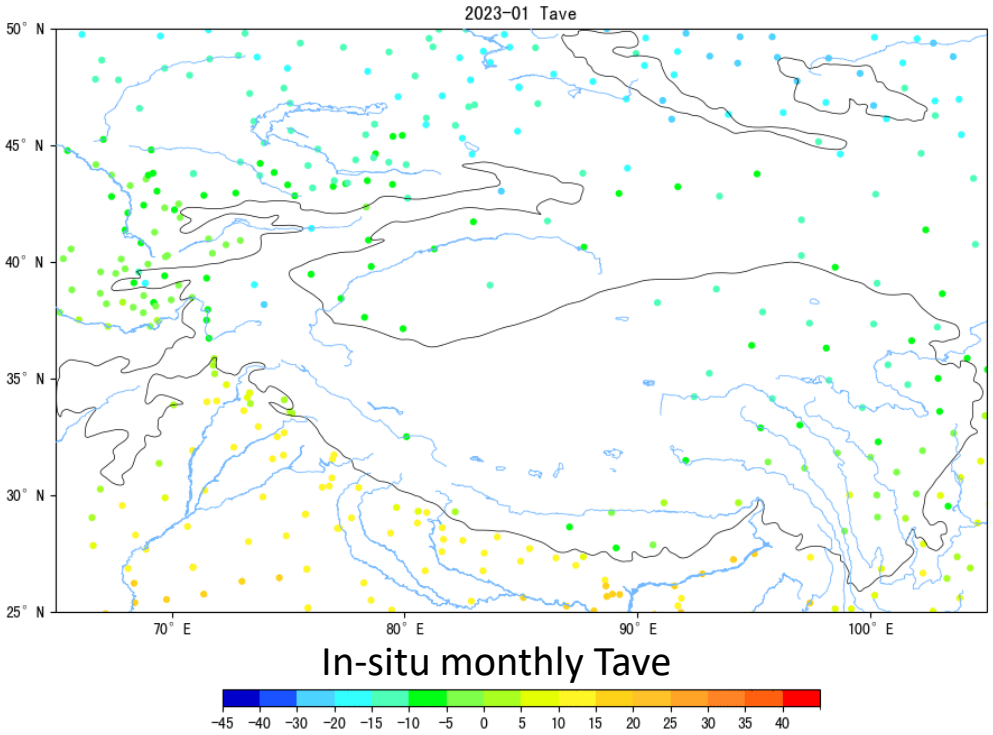




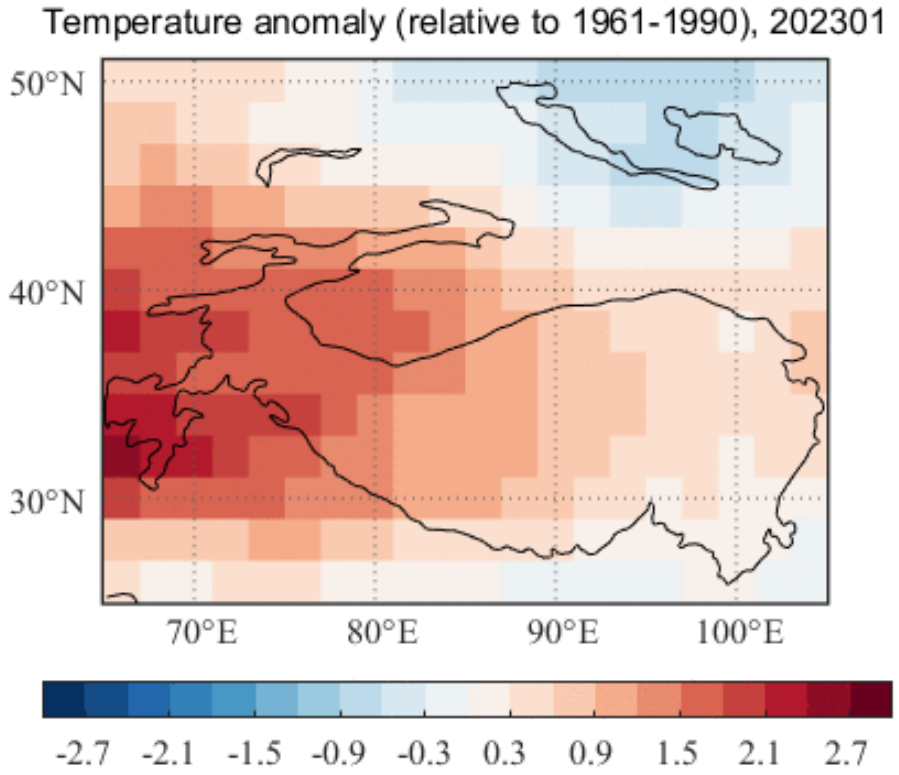
Spatial distribution of in-situ daily data in last 2 months (202404-202405)

The monthly data in TPR

- The monthly temperature and precipitation data in TPR update every month.
- The monthly gridding temperature anomalies (2 × 2 °) are produced synchronously .



In-situ monthly Tave

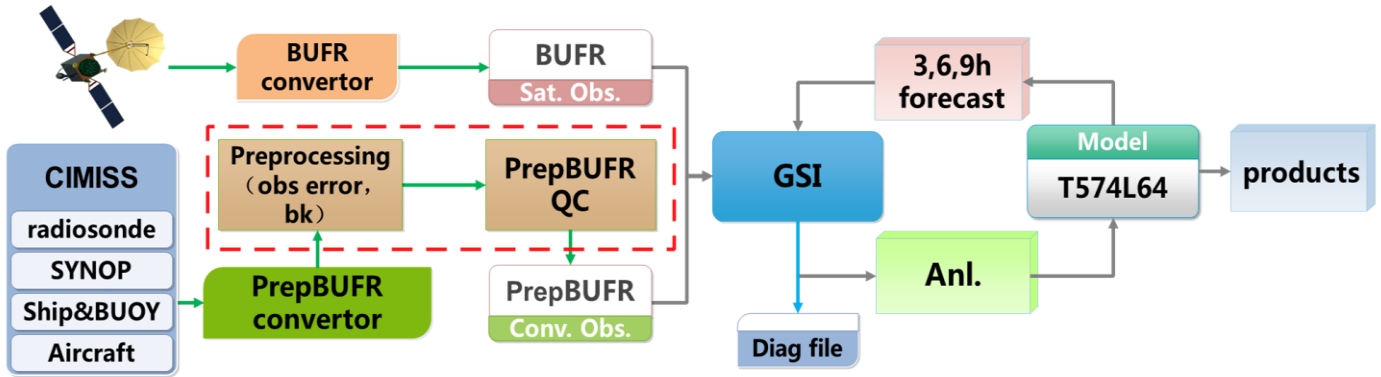


Spatial distribution of monthly data in last 16 months (202301-202404)

Contents

The development of CMA reanalysis product (CRA-40)

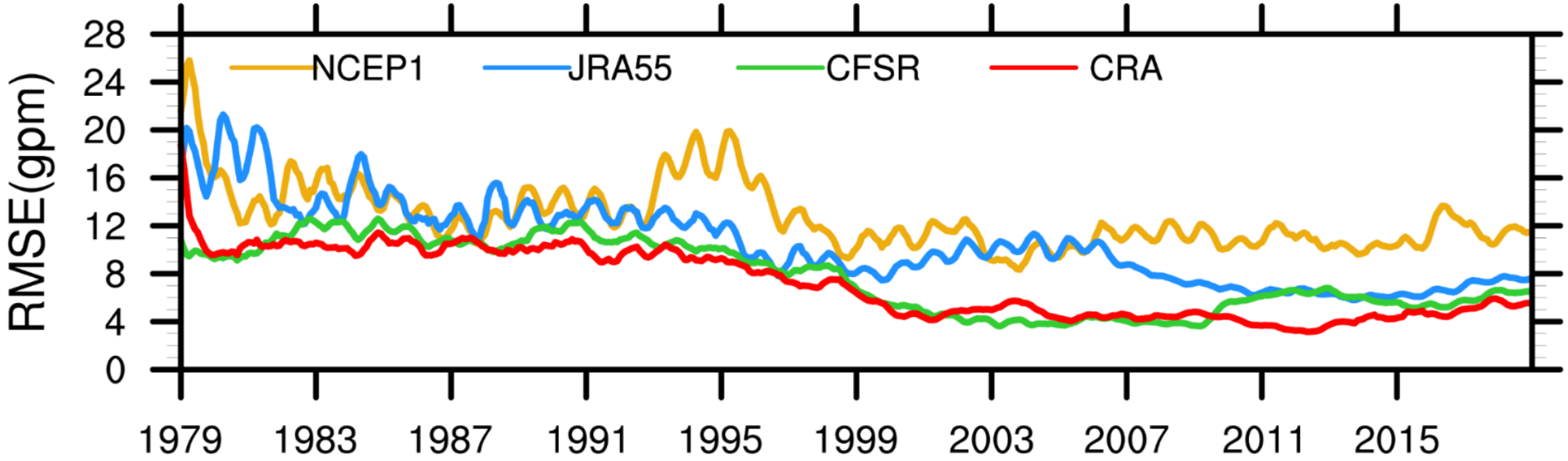
- The reanalysis product combined numerous types of observational data under atmosphere dynamic and physical mechanisms. It is good at temporal and spatial coverage and is a powerful tool for meteorological works.
- The CRA-40 (1979-present) is produced by GFS model and GSI assimilation system, assimilating a large amount of measurements(including Fengyun satellite data) based on 3DVar assimilation methods.



Product	CRA-40
Varidables	geopotential height, temperature, humidity, U/V wind, etc, 204 variables
Format	GRIB2
Resolution	0.25°/6h
Region	global
Timeliness	6.5h
Range	1979-present
Levels	Isobaric surface layer (47 layers), model layer (64 layers), surface layer, equipotential vortex layer, isentropic layer, specific altitude layer, specific ground altitude layer, other characteristic layers, etc

Overall performance of CRA-40 over world

- In order to evaluate the performance of CRA-40, an overall comparison with other reanalysis products was conducted including ERA5 and ERA-interim, CFSR, NCEP1, and JRA55.
- Taking ERA5 as the reference, the RMSE of global upper air potential height (500hPa) of CRA-40 is smaller than that of the NCEP1, and is equivalent to that of JRA55 and CFSR.

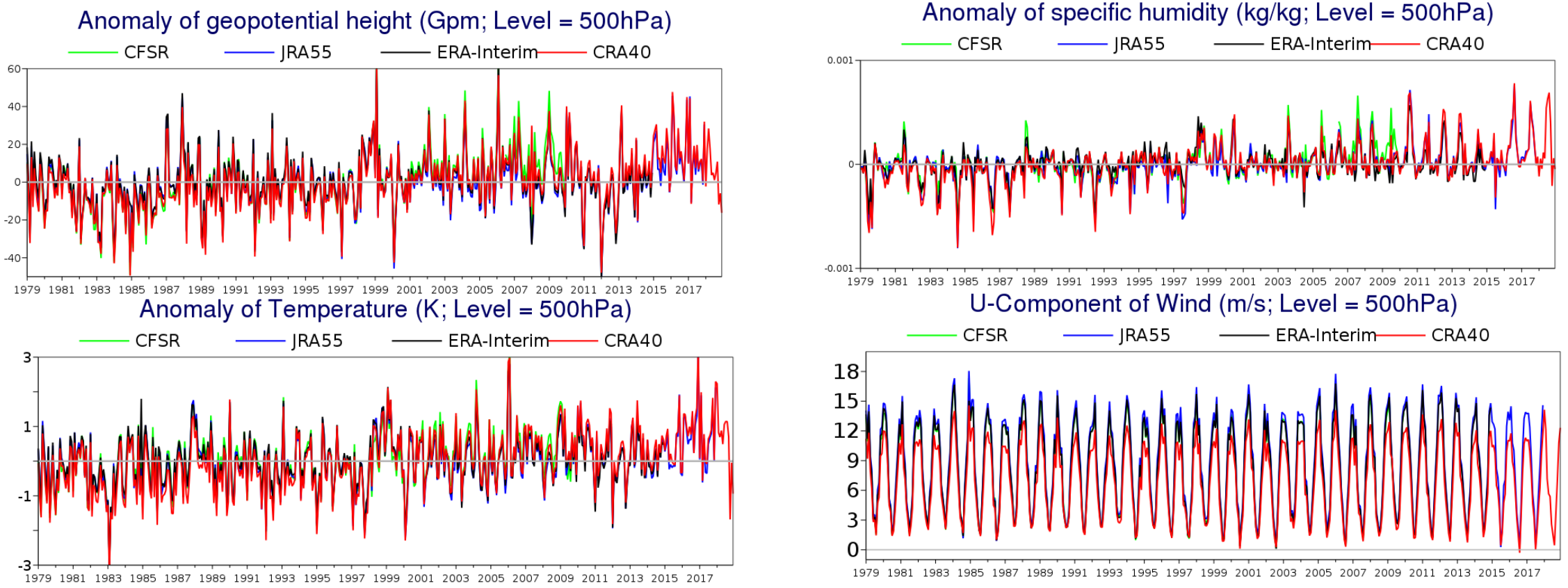


Multiple sets of reanalysis 500hPa potential height root mean square error time series (with era5 as reference)

Liu, Z. Q et al., 2023. J. Meteor. Res. doi: 10.1007/s13351-023-2086-x.

Overall performance of CRA-40 in Tibet Plateau(TP)

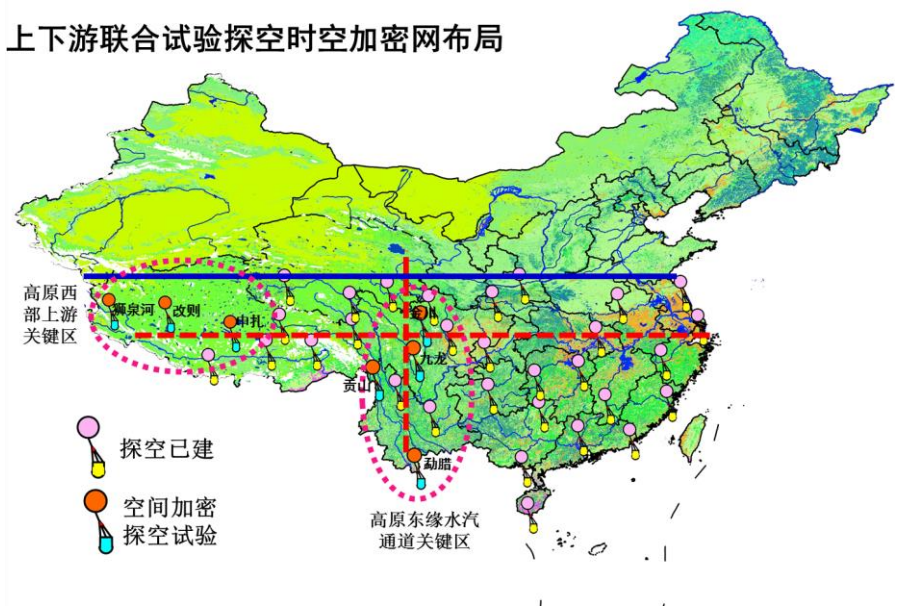
- Close and comprehensive attentions were paid to the performance of CRA-40 in Tibet Plateau(TP)
- CRA-40 shows similar long-term changes in TP in upper air geopotential height, temperature and humidity (500hPa) to that derived from JRA-55, ERA-interim, and CFSR.



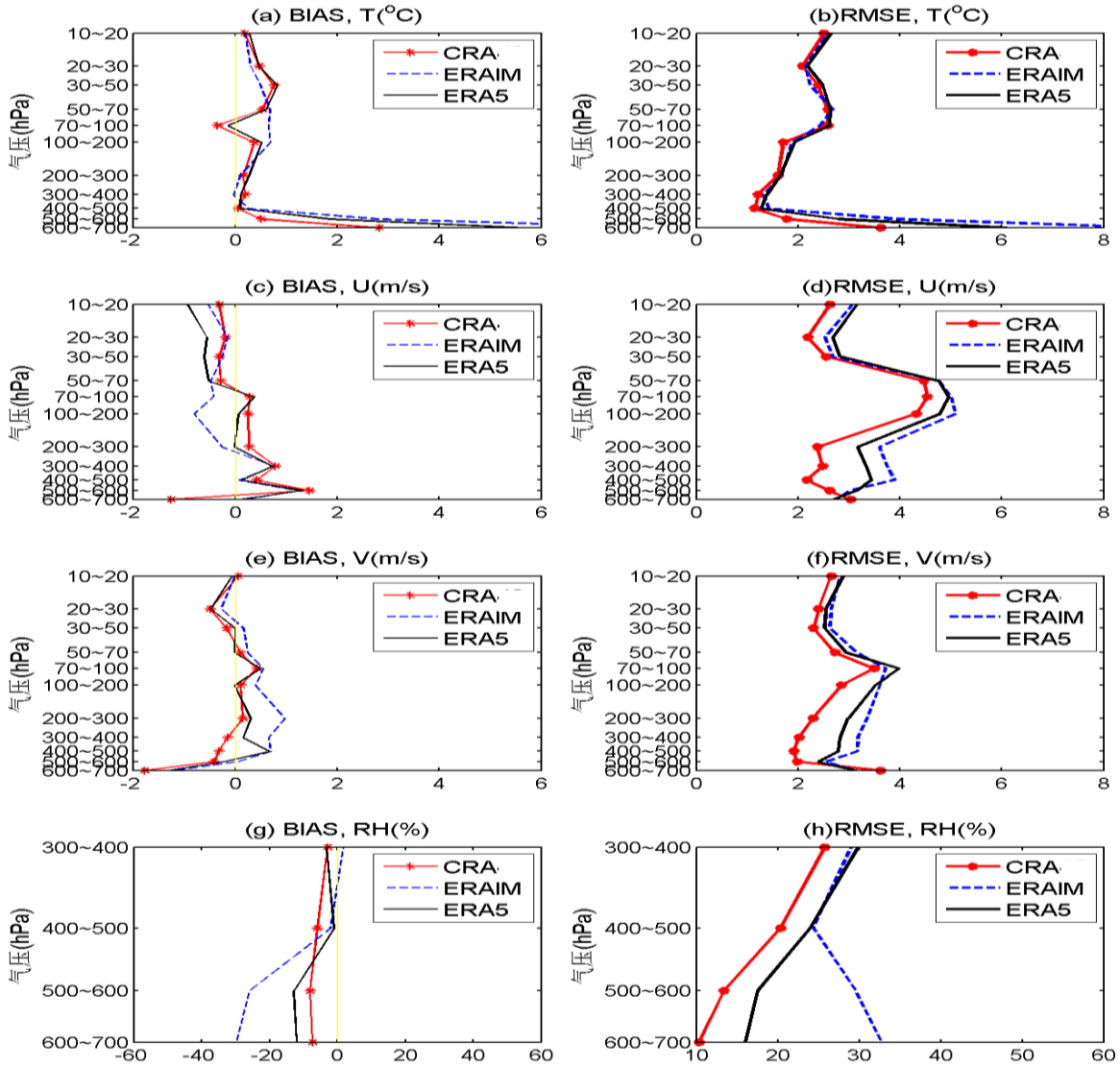
Time series of 500hPa geopotential height, humidity, temperature and wind over the Tibetan Plateau

Comparison with in-situ data at Tibet Plateau(TP) up air

- The upper air temperature, wind, humidity in TP from CRA-40, ERA-interim, and ERA5 were compared with the radiosonde measurements.
- CRA-40 showed a little higher correspondences and lower diversities with the observational data in TP than others



2015.6- 2016.10， 12UTC



Comparison with in-situ data over Tibet Plateau(TP) land surface

- The surface temperature change trends in TP derived from reanalysis products were also compared with that from in-situ observations.
- CRA-40 shows higher agreement of long-term surface temperature changes with the in-situ measurements than that from ERA-interim, CMFD(China Meteorological Forcing Dataset).

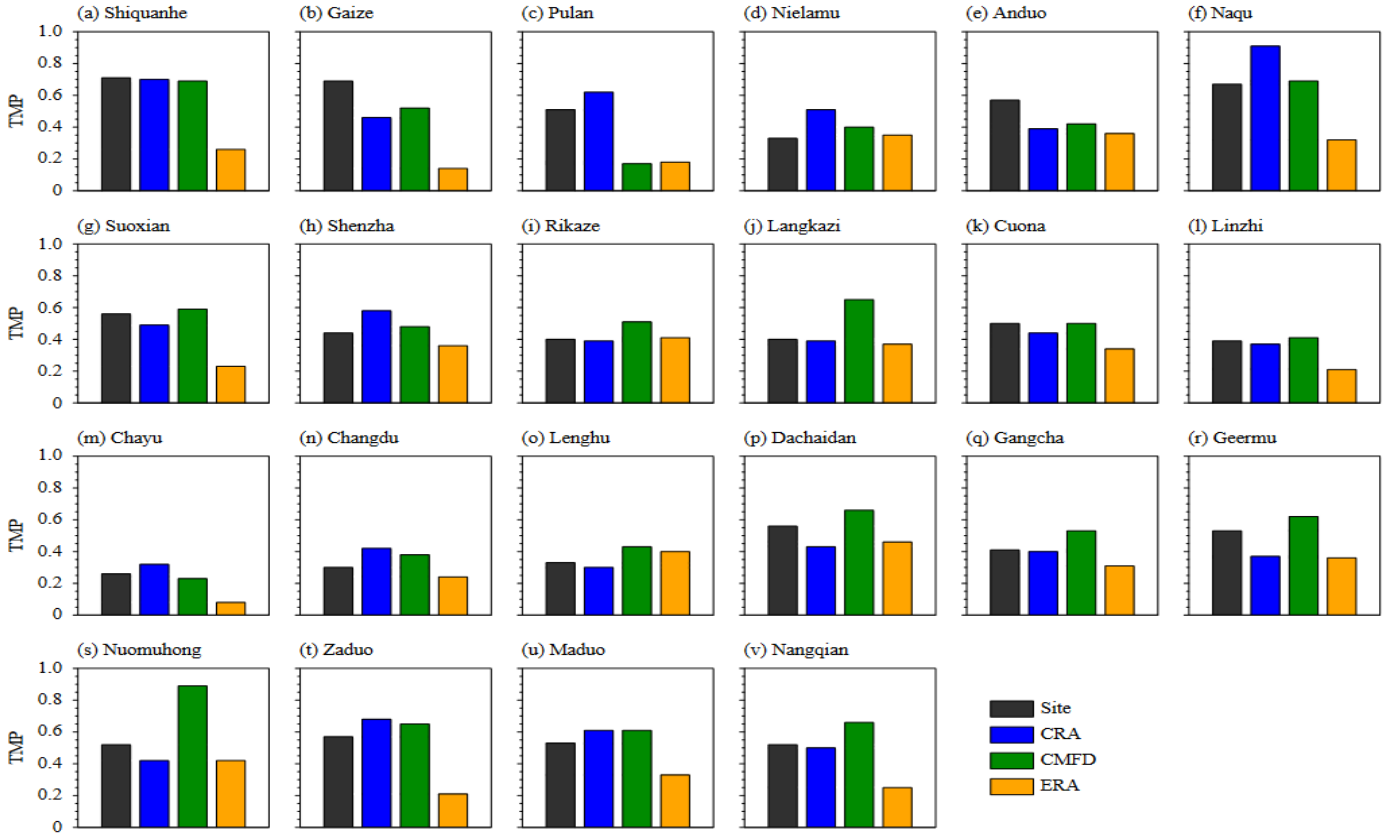
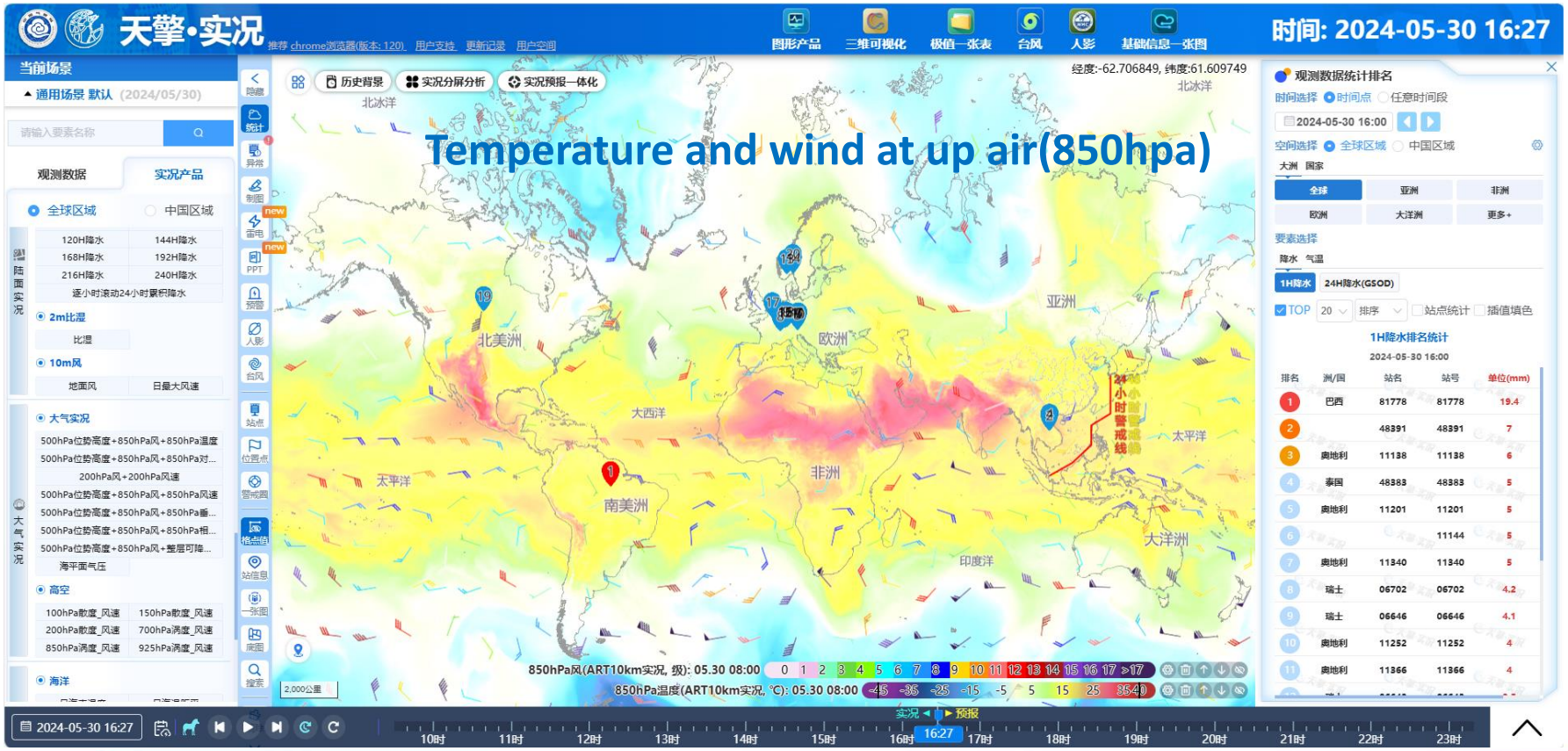


Fig. 5. Changes in the daily mean surface air temperature [TMP; °C (10 yr)$^{-1}$] from 1979 to 2018 derived from different datasets for each in-situ station.

Jiaxi YANG, et al. 2021. J. Meteor. Res.

The coming CRA V1.5

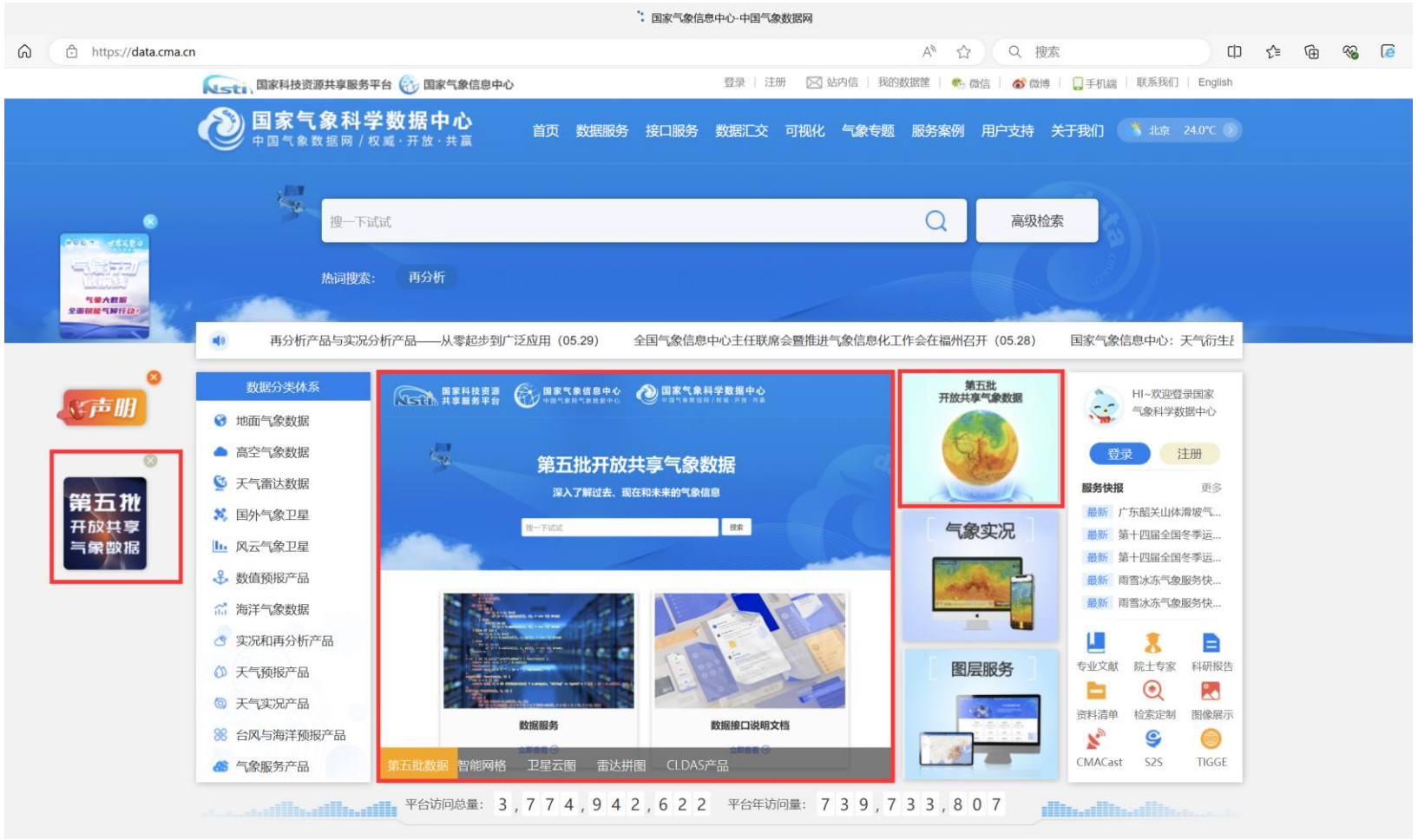
- Compared with the CRA V1.0, the horizontal resolution of the system is upgraded from T574 (about 34km) to T1534 (about 13km), the assimilation method is upgraded from 3DVar assimilation to hybrid-4DEnVar, and the update frequency move from six hours to one hour on a case-by-case basis.
- The primary assessment shows a significant improvement compared with CRA V1.0, e.g. the RMSE of the upper air (500-hPa) potential height between CRA and ERA5 during 1979 to 2023 decreases 20% relative to CRA V1.0.



Contents

Data open to the public

- We have been working on the data supports.
- A group of datasets have been open by CMA since May 2024, including global base data and CRA product.
- Welcome all members of TPRCC to apply and test and our data products, and we are long to your suggestions.
- Meanwhile, we also aim to the improve the ways of data access, e.g. the application, the download speed by internet etc.



https://data.cma.cn

Data development and support plan for TPRCC

Global base data product: enrich the meteorological variabilities contained in the datasets, e.g. the daily(monthly) humidity, weather, max wind speed etc., and achieve the gridding daily temperature and precipitation data. This would provide TPRCC with more first-hand observational variabilities.

CRA product: fully evaluate the new CRA V1.5 product and then update it's spatial resolution to 5×5km by improving the models and establish a global atmospheric real-time analysis system. This would provide TPRCC with more detailed and finer reanalysis data.

Thanks for your attention