Evaluation of ERA5 Reanalysis over the Deserts in Northern China[1]

Chengzhi Hou[1], Danqing Huang[2]*, Hao Xu[3,4], Zhiwei Xu[1,5]*

1 School of Geography and Ocean Science, Nanjing University, China

2 School of Atmospheric Sciences, Nanjing University, China

3 Institute of Forestry and Grassland Ecology, Ningxia Academy of Agricultural and Forestry Sciences, China

4 Ningxia Key Laboratory of Desertification Control and Soil and Water Conservation, China

5 Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, China

Email: huangdq@nju.edu.cn; zhiweixu@nju.edu.cn

Abstract: Deserts cover a vast area of the world's land surface, but the study of desert climate has been impeded by a lack of ground meteorological observations. In recent years, the climate reanalysis products provide an important data source to investigate climate change in observation-limited areas, however, their accuracy in desert regions has been poorly investigated. Here, we first comprehensively evaluated the performance of the latest ERA5 reanalysis datasets on the climatic conditions over the deserts of northern China (DNC), including temperature and precipitation climatology, climate extremes and detection skills of daily precipitation, in comparison with ground observations (OBS) as well as other datasets. The results show that ERA5 well captures the observed pattern of annual and seasonal temperature, as well as the warming trend during the past 40 years in DNC, compared to the OBS. However, both annual and seasonal precipitation have been greatly overestimated over DNC and large bias of ERA5 is found in estimating precipitation trend. Compared to the OBS, ERA5 has lower variability in interannual precipitation, smaller precipitation intensity and maximum 1-day precipitation, shorter continuous dry days and more wet days. The bias in extreme precipitation may be traced to the overestimation of rainfall

[1] This study was supported by the National Natural Science Foundation of China (Nos. 41871012, 42075020), the Fundamental Research Funds for the Central Universities of China (0209-14380091), and the Chinese Academy of Sciences Interdisciplinary Innovation Team.

occurrence in ERA5, because the false alarm ratio (FAR) is 66% over the whole DNC and can be as high as 90% in extremely arid western deserts. Overall, ERA5 outperforms other reanalysis such as MERRA2, NCEP1, NCEP2 and NOAA-20C for precipitation and extreme precipitation over DNC, although the performance is not as good as gauge-based precipitation datasets. This study provides reference for use of ERA5 in climate change studies over DNC and other observation-limited desert regions.

Key words: ERA5, Deserts in Northern China, desert climate, climate extremes, rainfall occurrence

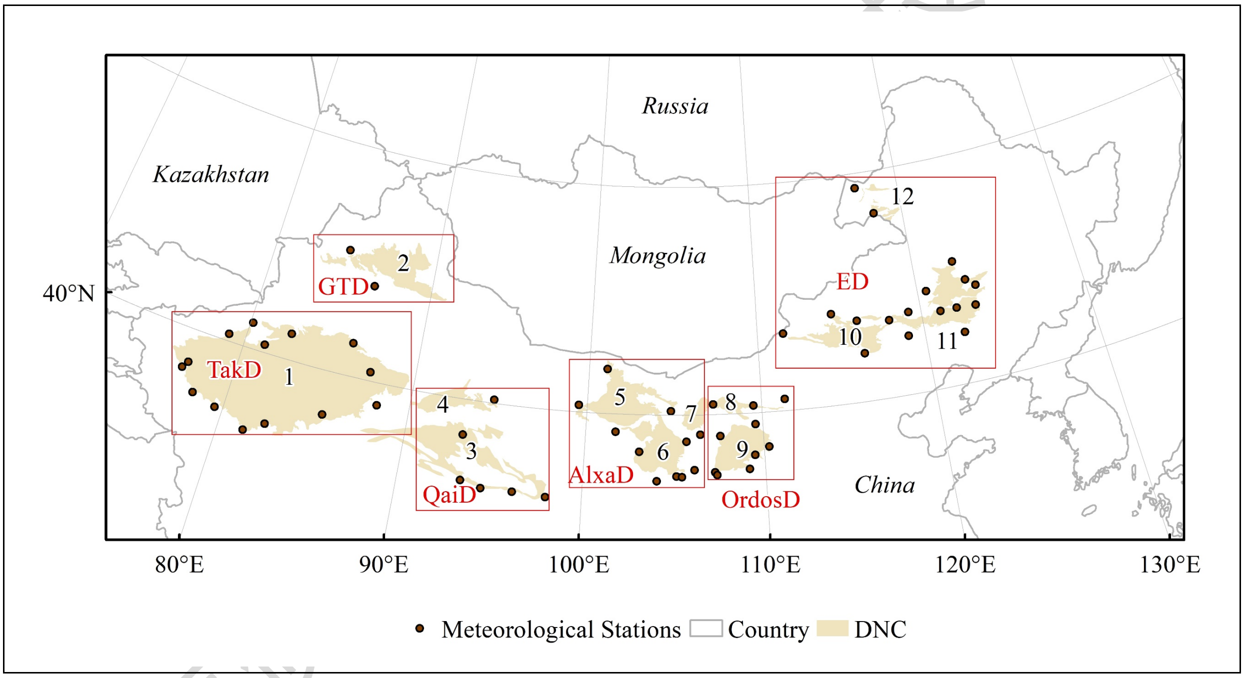

Fig.1. Spatial distribution of DNC and meteorological stations. The shade in the map refers to the deserts and dune fields while the dots indicate the meteorological stations. The box labels each sub-region. 1: Taklamakan Desert, 2: Gurban Tunggut Desert, 3: Qaidam Desert, 4: Kumtag Desert, 5: Badain Jaran Desert, 6: Tengger Desert, 7: Ulan Buhe Desert, 8: Hobq Desert, 9: Mu Us Dune Field, 10: Otindag Dune Field, 11: Horqin Dune Field, 12: Hulunbuir Dune Field.

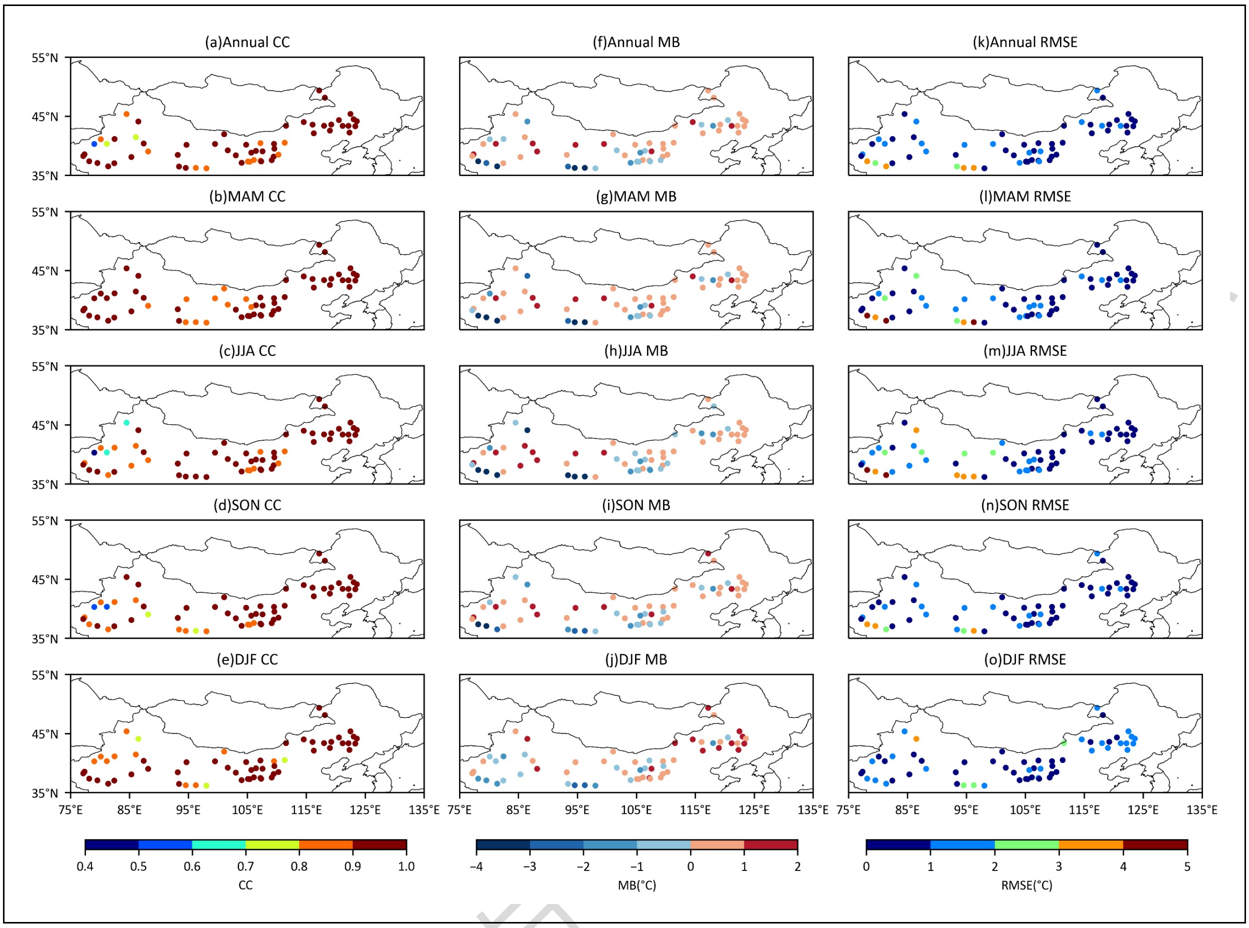

Fig. 2. Spatial distribution of statistical indices of annual and seasonal temperature at each station. (a)-(e): Correlation coefficient (CC) between OBS and ERA5. (f)-(j): Mean bias (MB) between OBS and ERA5. (k)-(o): Root mean square error (RMSE) between OBS and ERA5. The dots represent the geographical location of each station.

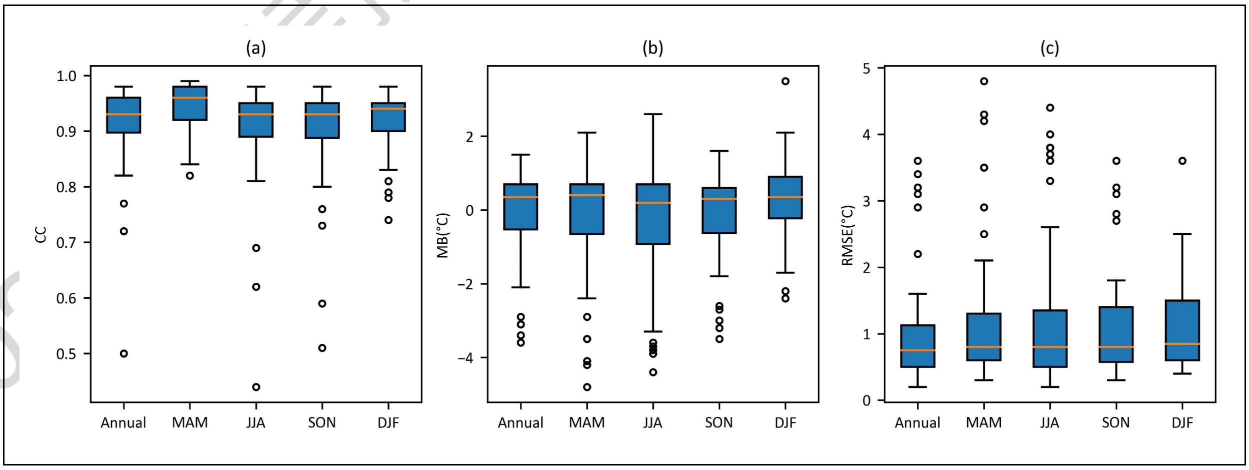

Fig. 3. Boxplot of statistical indices of annual and seasonal temperature. (a): Correlation coefficient (CC) between OBS and ERA5. (b): Mean bias (MB) between OBS and ERA5. (c): Root mean square error (RMSE) between OBS and ERA5.

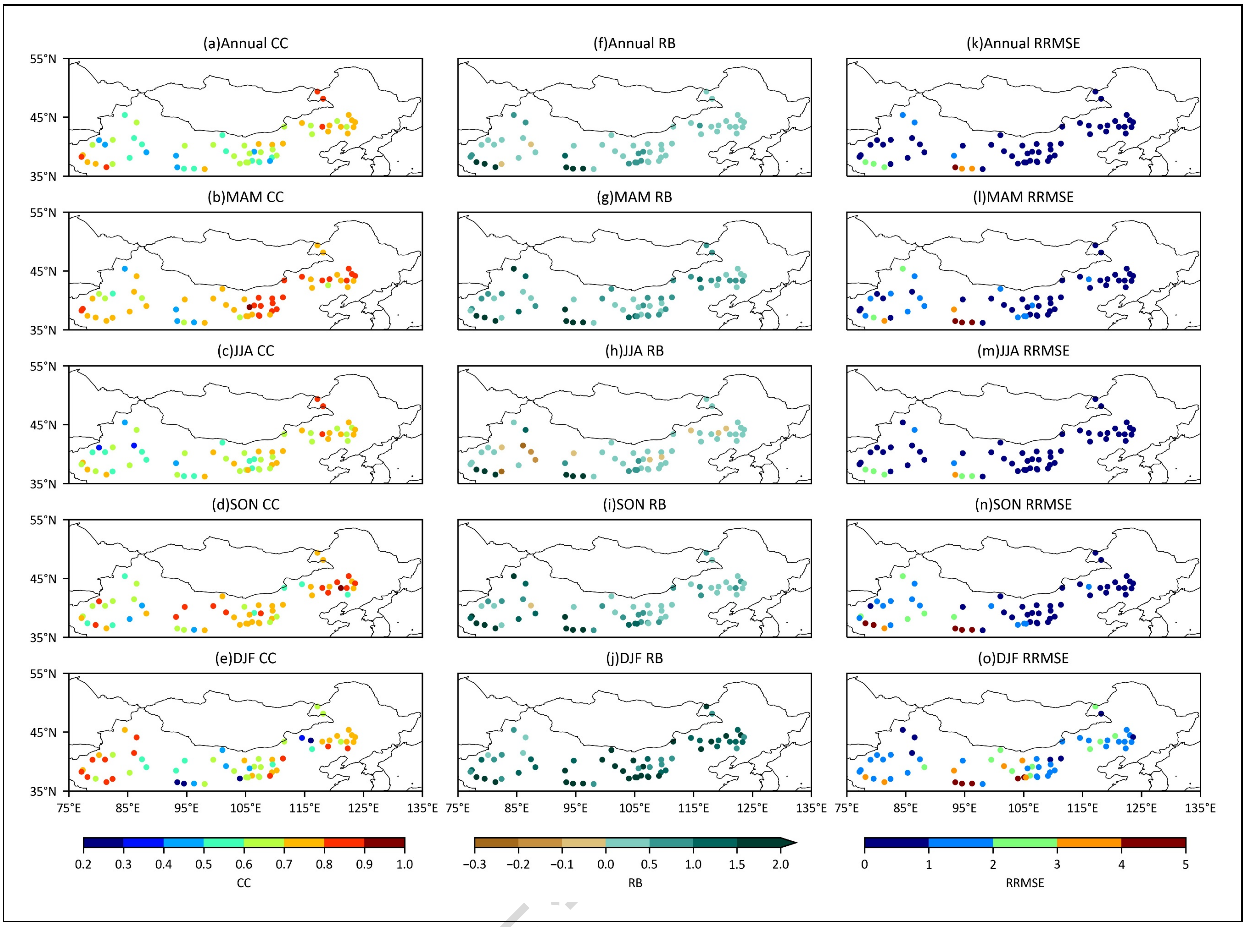

Fig. 4. Spatial distribution of statistical indices of annual and seasonal precipitation at each station. (a)-(e): Correlation coefficient (CC) between OBS and ERA5. (f)-(j): Mean bias (MB) between OBS and ERA5. (k)-(o): Root mean square error (RMSE) between OBS and ERA5. The dots represent the geographical location of each station.

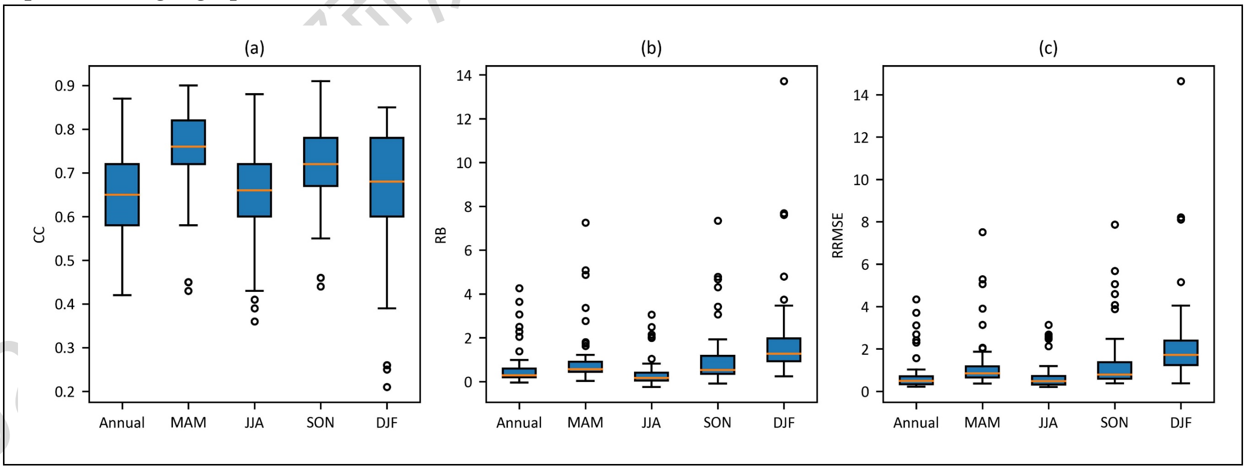

Fig. 5. Boxplot of statistical indices of annual and seasonal precipitation. (a): Correlation coefficient (CC) between OBS and ERA5. (b): Mean bias (MB) between OBS and ERA5. (c): Root mean square error (RMSE) between OBS and ERA5.

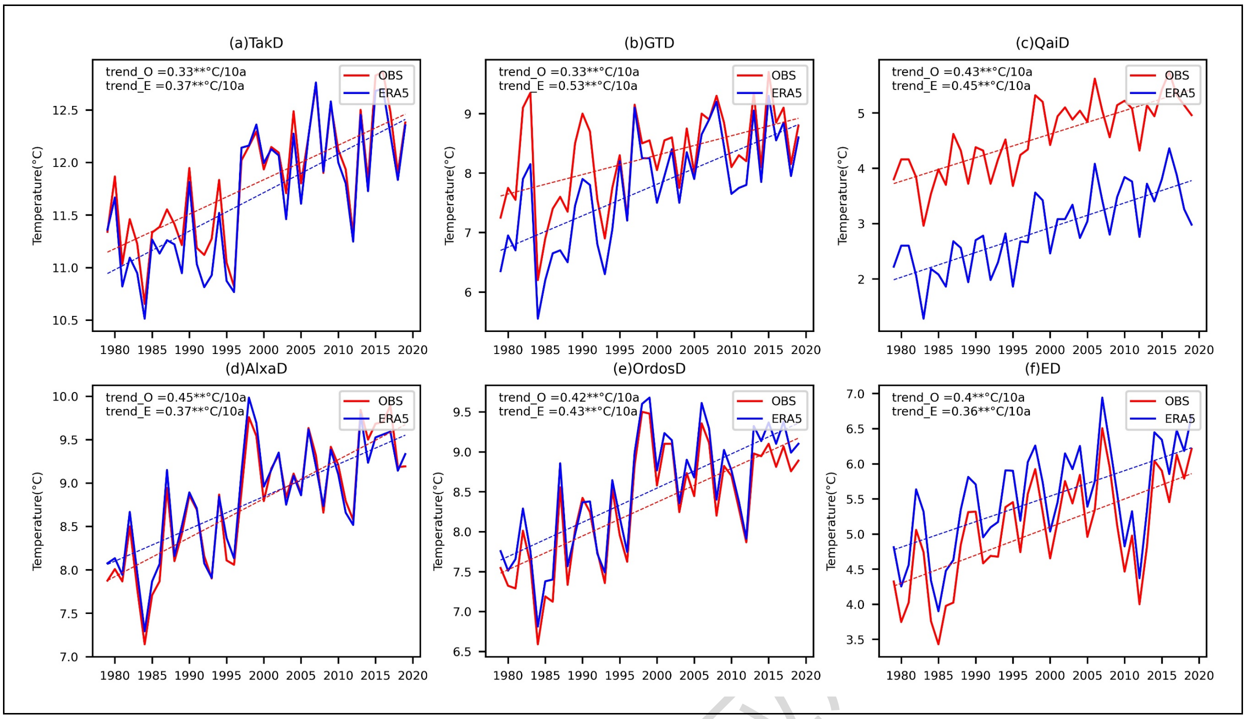

Fig. 6. Time series and linear trend of annual mean temperature based on the OBS and ERA5 during 1979-2019 over six sub regions of DNC. 'trend_E' and 'trend_O' are the linear trend in ERA5 and OBS, respectively. The significant level of the trend was determined by Mann-Kendall test. **, * indicate 99%, 95% significant levels respectively.

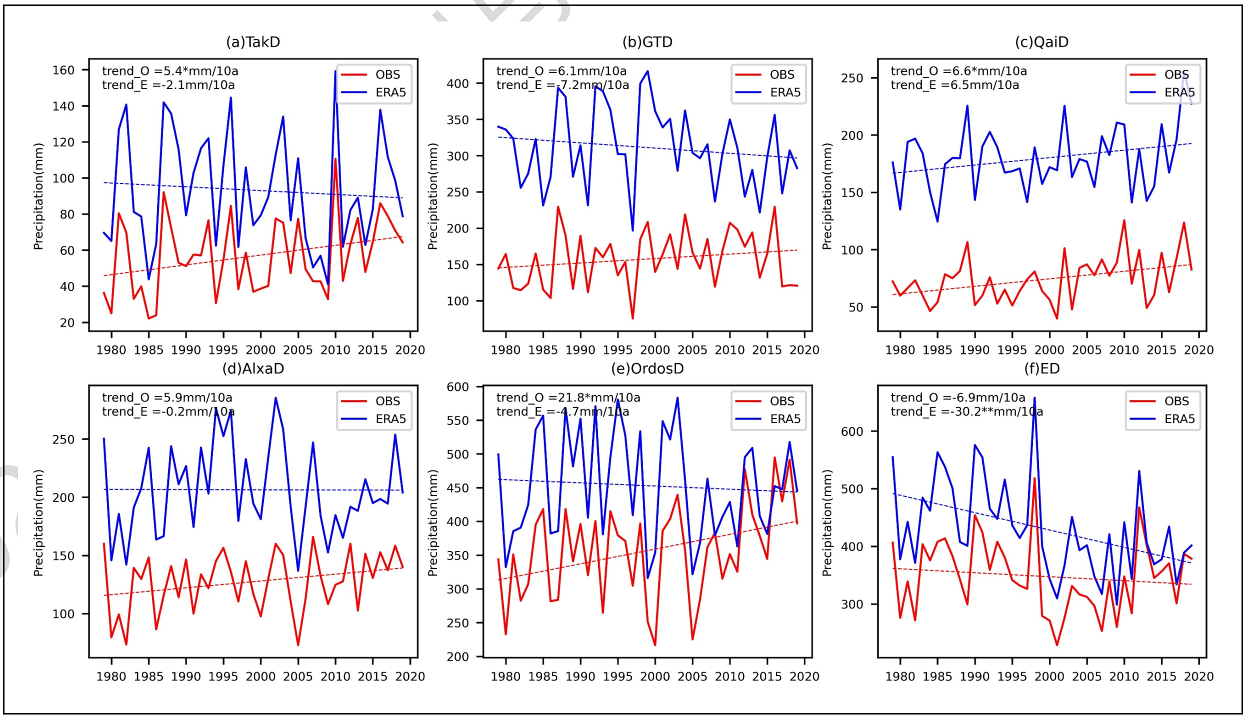

Fig. 7. Time series and linear trend of annual precipitation based on the OBS and ERA5 during 1979-2019 over six sub regions of DNC. 'trend_E' and 'trend_O' are the linear trend in ERA5 and OBS, respectively. The significant level of the trend was determined by Mann-Kendall test. **, *

indicate 99%, 95% significant levels respectively.

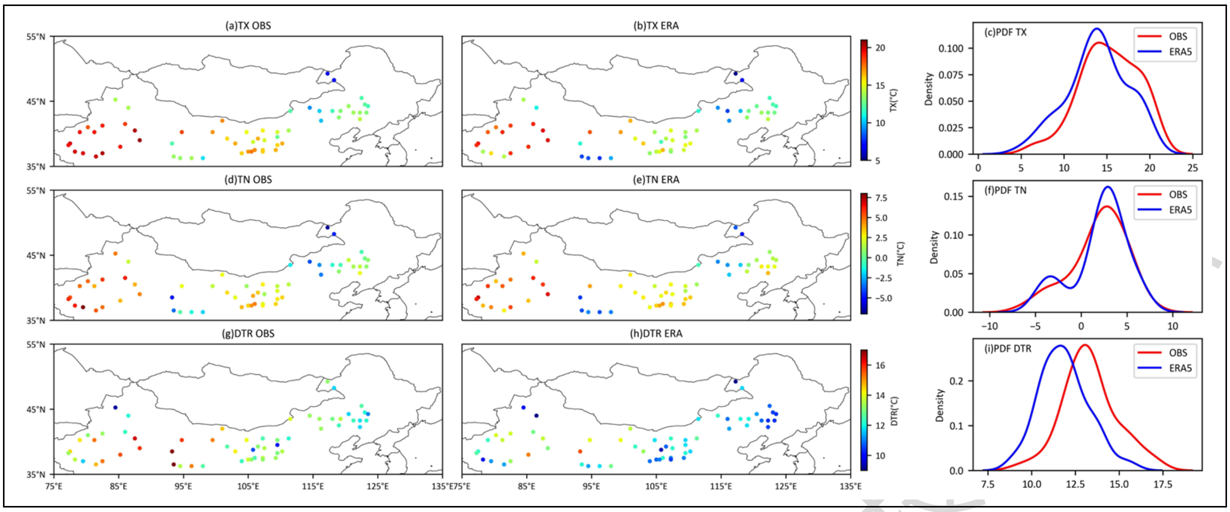

Fig. 8. Spatial distributions of temperature extreme indices based on OBS and ERA5, respectively, and their probability density distribution (the rightmost panel). (a): TX calculated from OBS. (b): TX calculated from ERA5. (c): Probability density distribution of TX based on OBS and ERA5. (d): TN calculated from OBS. (e): TN calculated from ERA5. (f): Probability density distribution of TX based on OBS and ERA5. (g): DTR calculated from OBS. (h): DTR calculated from ERA5. (i): Probability density distribution of DTR based on OBS and ERA5.

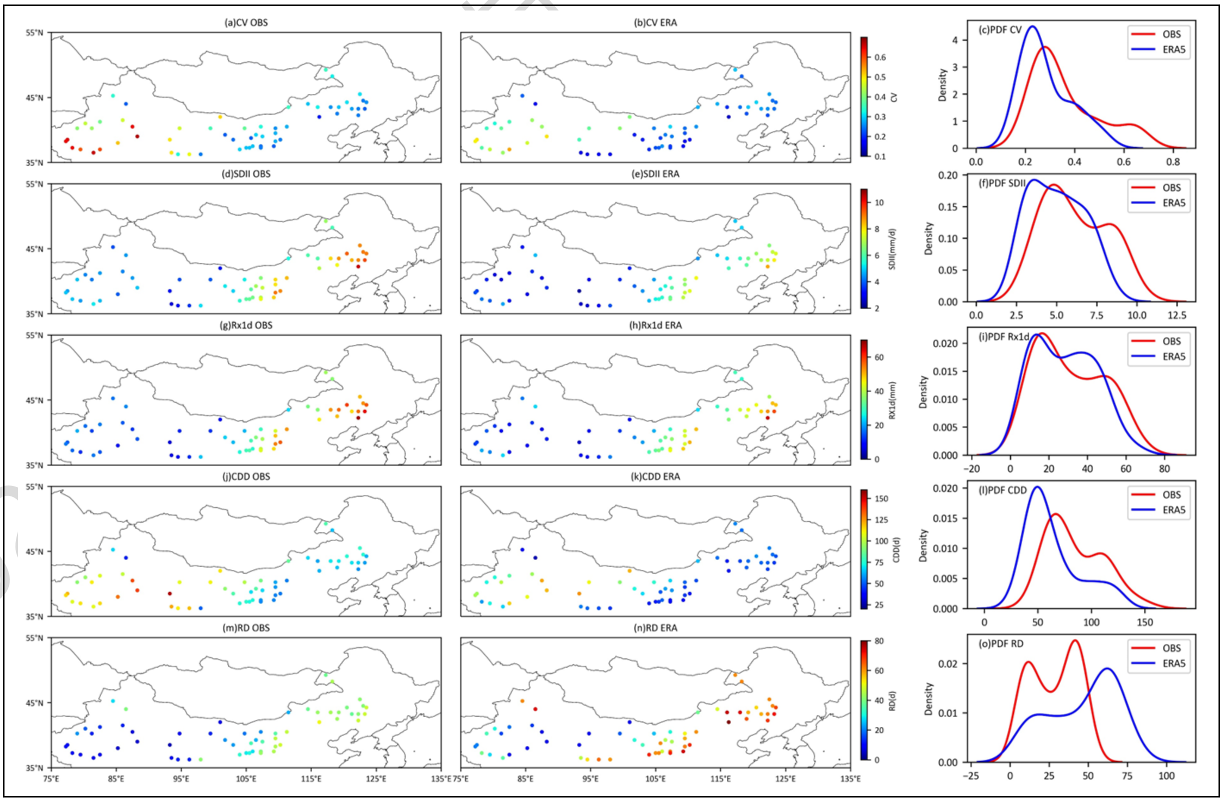

Fig. 9. Spatial distributions of five different precipitation extreme indices based on OBS and ERA5, respectively, and their probability density distribution (the rightmost panel). (a): CV

calculated from OBS. (b): CV calculated from ERA5. (c): Probability density distribution of CV based on OBS and ERA5. (d): SDII calculated from OBS. (e): SDII calculated from ERA5. (f): Probability density distribution of SDII based on OBS and ERA5. (g): RX1d calculated from OBS. (h): RX1d calculated from ERA5. (i): Probability density distribution of RX1d based on OBS and ERA5. (j): CDD calculated from OBS. (k): CDD calculated from ERA5. (l): Probability density distribution of CDD based on OBS and ERA5. (m): RD calculated from OBS. (n): RD calculated from ERA5. (o): Probability density distribution of RD based on OBS and ERA5.

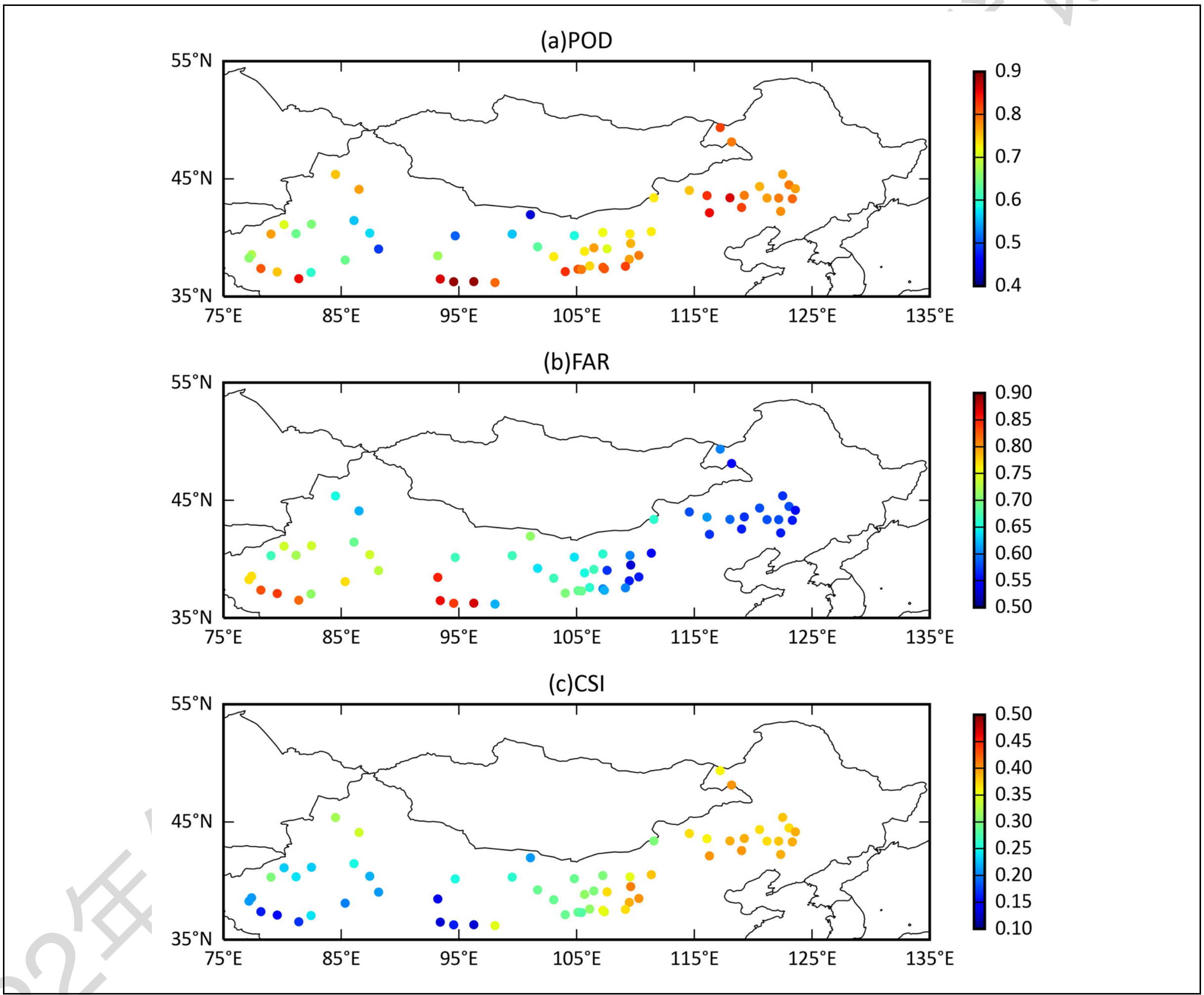

Fig. 10. Spatial distribution of daily precipitation indices that used to evaluate the capacity of ERA5 in detecting daily precipitation events. (a): Probability of detection (POD). (b): False alarm ratio (FAR). (c): Critical success index (CSI).

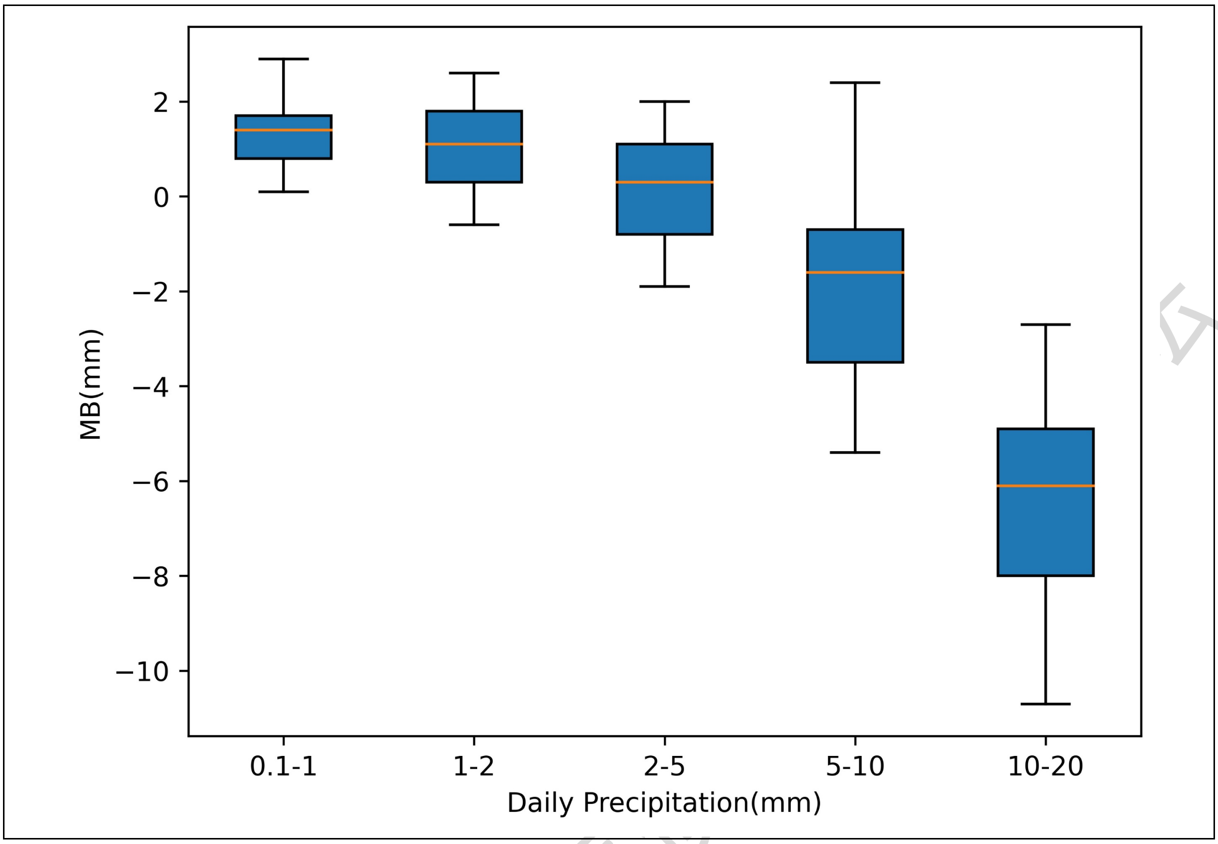

Fig. 11. Boxplot of mean bias calculated for daily precipitation of different grades.

Table 1 Information of sub-regions of DNC.

Short Name	Full Name	Deserts/ Dune fields	Geographical Location
TakD	Taklamakan Desert Region	Taklamakan Desert (1)	Tarim Basin south of the Tianshan Mountains
GTD	Gurban Tunggut Desert Region	Gurbantunggut Desert (2)	Junggar Basin north of the Tianshan Mountains and west of the Altai Mountain
QaiD	Qaidam Desert Region	Qaidam Desert and Kumtagh Desert (3, 4)	Qaidam Basin south of Qilian Mountain
AlxaD	Alxa Desert Region	Badain Jaran, Tengger and Ulan Buh Deserts (5, 6, 7)	Alxa Plateau west of Helan Mountain and north of Qilian Mountain
OrdosD	Ordos Desert Region	Hobq Desert and Mu Us dune field (8, 9)	Ordos Plateau east of Helan Mountain and west of Taihang Mountain
ED	Eastern Desert Region	Otindag, Horqin and Hulunbuir dune fields (10, 11, 12)	Mongolia Plateau in northeastern China

Table 2 Information of the statistical indices used in this study. In the formula, n is the length of time series, E_i is the climate variables in ERA5 during i period, and O_i is the climate variables based on ground meteorological observations, $\bar{E}$ and $\bar{O}$ are the average value in corresponding

periods.

Short name	Full name	Formula
CC	Correlation coefficient	$CC = \frac{\sum_{i=1}^{n}(E_i - \bar{E})(O_i - \bar{O})}{\sqrt{\ }}$
MB	Mean bias	$MB = \frac{1}{n}\sum_{i}^{n}(E_i - O_i)$
RB	Relative bias	$RB = \frac{MB}{\bar{O}}$
RMSE	Root mean square error	$RMSE = \sqrt{\frac{1}{n}\sum_{i}^{n}(E_i - O_i)^2}$
RRMSE	Relative root mean square error	$RRMSE = \frac{RMSE}{\bar{O}}$

Table 3 POD, FAR, and CSI are used to evaluate the capacity of ERA5 in detecting daily precipitation events. H is the number of hits, which means the frequency of precipitation events recorded in both ERA5 and meteorological stations. M is the number of misses, which means that the frequency of precipitation events recorded in meteorological stations but missed in ERA5. F is the number of false alarms which means the frequency of precipitation event occurring in ERA5 but not appearing in the meteorological stations.

Short name	Full name	Formula
POD	Probability of detection	$POD = \frac{H}{H + M}$
FAR	False alarm ratio	$FAR = \frac{F}{H + F}$
CSI	Critical success index	$CSI = \frac{H}{H + F + M}$

Table 4 Specific climate indices for desert climate in DNC.

Indices	Unit	Definition
TX	℃	Averaged daily maximum temperature
TN	℃	Averaged daily minimum temperature
DTR	℃	Averaged daily temperature range
CV	\	Coefficient of variation of annual precipitation
SDII	mm/d	Averaged precipitation intensity of all wet days (>1mm)
Rx1d	mm	Maximum 1-day precipitation

CDD	d	Maximum number of consecutive days with precipitation < 1mm
RD	d	Number of days with precipitation >= 1mm

Table 5 Evaluation results of several reanalysis and gauge-based gridded precipitation datasets for annual precipitation and daily precipitation detection skills.

Datasets		Error analysis			Daily precipitation detection capacity		
		CC	RB(%)	RRMSE	POD	FAR	CSI
Reanalysis datasets	ERA5	0.65	61.4	0.76	0.74	0.66	0.30
	MERRA2	0.53	66.3	0.95		\	
	NCEP1	0.43	186.2	2.27	0.32	0.84	0.11
	NCEP2	0.43	85.6	1.37	0.24	0.81	0.11
	NOAA-20C	0.46	159.7	1.73	0.54	0.84	0.14
Gridded precipitation dataset	PREC/L	0.70	17.6	0.42		\	
	GPCP	0.55	60.4	0.80		\	
	CPC	0.71	13.7	0.38	0.36	0.80	0.15

Table 6 Same as Table 5, but for extreme precipitation results.

		CV			SDII			RX1d			CDD			RD		
		CC	RB(%)	RRMSE	CC	RB(%)	RRMSE	CC	RB(%)	RRMSE	CC	RB(%)	RRMSE	CC	RB(%)	RRMSE
Reanalysis datasets	ERA5	0.82	-21.4	0.31	0.43	-18.6	0.34	0.34	-7.1	0.59	0.36	-25.2	0.42	0.56	80.8	0.91
	NCEP1	0.22	-9.6	0.42	0.21	-7.90	0.37	0.14	-13.8	0.70	0.20	-14.7	0.50	0.42	169.5	1.98
	NCEP2	0.45	4.9	0.36	0.18	-3.60	0.38	0.14	-13.8	0.69	0.28	-3.70	0.47	0.45	71.7	1.12
	NOAA-20C	0.61	-46.7	0.56	0.27	-43.7	0.52	0.20	-46.2	0.71	0.18	-25.0	0.56	0.44	153.8	1.73
Gridded precipitation dataset	CPC	0.92	2.1	0.18	0.62	-18.2	0.32	0.54	-16.1	0.54	0.57	-4.0	0.31	0.58	29.8	0.47