Supplement of Earth Syst. Sci. Data, 17, 3987–4004, 2025
https://doi.org/10.5194/essd-17-3987-2025-supplement

Open Access Earth System Science Data

Supplement of

An upgraded high-precision gridded precipitation dataset for the Chinese mainland considering spatial autocorrelation and covariates

Jinlong Hu et al.

Correspondence to: Chiyuan Miao (miaocy@bnu.edu.cn)

Contents of this file

Tables S1 to S8.

Table S1. List of abbreviations used throughout this paper.

Abbreviation	Full Term
LGBM	Light gradient boosting machine
PRISM	Parameter-elevation regression on independent slopes model
IDW	Inverse distance weighting
CMA	China Meteorological Administration
GHCND	Global Historical Climatology Network-Daily
NCDC	National Climatic Data Center
NOAA	National Oceanic and Atmospheric Administration
SRTM	Shuttle Radar Topography Mission
DEM	Digital Elevation Model
CGIAR-CSI	Consortium for Spatial Information, Consultative Group for International Agricultural Research
NASA	National Aeronautics and Space Administration
GSMaP	Global Satellite Mapping of Precipitation
PERSIANN-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
JAXA	Japan Aerospace Exploration Agency
GLDAS NOAH	Global Land Data Assimilation System Noah Land Surface Model
NDVI	Normalized Difference Vegetation Index
IMERG	Integrated Multi-satellitE Retrievals for GPM
CMA-HD	High-density automatic rain gauge stations across Chinese mainland
NEC	North East China
NC	North China
SCC	South and Central China
IM	Inner Mongolia
NWC	North West China
SWC	South West China
QT	Qinghai-Tibet Plateau
CDD	Correlation decay distance
ADW	Adaptive distance weighting
GBDT	Gradient-boosted decision tree
AE	Absolute error
KGE	Kling-Gupta efficiency
RSD	Relative standard deviation
HSS	Heidke skill score
FAR	False alarm ratio
POD	Probability of detection

Table S2. Gridded precipitation dataset used for comparison.

Dataset Name	Data Generation Methods	Spatial Resolution	Temporal Resolution	Data Availability Time Range
CHM_PRE V1	Gauge-based Interpolation	0.10°	1d	1961–2022
GSMaP V8	Remote Sensing	0.10°	1h	1998–Present
PERSIANN-CDR	Remote Sensing	0.25°	1d	1983–Present
GLDAS_Noah025 V2.1	Reanalysis	0.25°	3h	2000–Present
IMERG Final L3 V7	Remote Sensing	0.10°	1d	2000.06–Present

Table S3. The variables used in the precipitation retrieval.

Variable Type	Variable Name	Description
Spatial autocorrelation variables	Lat	Latitude of the grid center
	Lon	Longitude of the grid center
	Interp. Prec.	Gridded precipitation based on gauge interpolation
Precipitation-related covariates	DEM	Average elevation of the grid
	Slope	Average slope of the grid
	GLDAS Prec.	Precipitation of the grid from GLDAS
	Prec. RS	Satellite-derived precipitation of the grid
	GLDAS SM	Soil moisture of the grid from GLDAS
	NDVI	NDVI of the grid
	Annual Prec.	Annual total precipitation of the grid
	Monthly Prec.	Monthly total precipitation of the grid
	1st-day prior Prec.	Daily precipitation one day before the current date
	2nd-day prior Prec.	Daily precipitation two day before the current date
	3rd-day prior Prec.	Daily precipitation three day before the current date
	4th-day prior Prec.	Daily precipitation four day before the current date
	5th-day prior Prec.	Daily precipitation five day before the current date

Table S4. Contingency table for comparing the precipitation and no-precipitation events detected by gauge and products.

Threshold	**Product≥threshold**	**Product＜threshold**
Gauge≥threshold	Hits (TP)	Misses (FN)
Gauge＜threshold	False alarms (FP)	Correct negatives (TN)

Table S5. Precipitation accuracy of different datasets validated by high-density gauge data. The bolded numbers in the column represent the optimal accuracy values for that metric.

Dataset Name	MAE (mm/day)	KGE	Bias	RSD
CHM_PRE V2	**1.48**	**0.79**	1.05	**0.88**
CHM_PRE V1	1.67	0.70	1.12	0.78
GSMaP	2.94	0.48	**1.04**	0.80
IMERG	3.27	0.44	1.12	0.84
PERSIANN-CDR	3.70	0.29	1.12	0.70
GLDAS	3.69	0.31	1.04	0.79

Table S6. Precipitation accuracy validated by high-density gauge data in different regions. The bolded numbers in the column represent the optimal accuracy values for that metric.

Region Abbreviation	Dataset Name	MAE (mm/day)	KGE	Bias	RSD
NEC	CHM_PRE V2	**1.00**	**0.63**	**1.19**	**0.79**
	CHM_PRE V1	1.06	0.58	1.23	0.73
	GSMaP	1.74	0.42	1.12	0.77
	GLDAS	2.21	0.15	1.25	0.55
	PERSIANN-CDR	2.35	0.10	1.35	0.54
	IMERG	2.08	0.29	1.32	0.76
NC	CHM_PRE V2	**0.85**	**0.76**	**1.11**	**0.85**
	CHM_PRE V1	0.97	0.66	1.16	0.75
	GSMaP	1.87	0.42	1.12	0.75
	PERSIANN-CDR	2.41	0.15	1.24	0.53
	IMERG	2.11	0.33	1.22	0.73
	GLDAS	2.37	0.22	1.14	0.66
SCC	CHM_PRE V2	**1.94**	**0.80**	1.04	**0.90**
	CHM_PRE V1	2.21	0.71	1.11	0.79
	GSMaP	3.92	0.48	1.03	0.80
	IMERG	4.35	0.45	1.10	0.86
	PERSIANN-CDR	4.88	0.32	1.09	0.73
	GLDAS	4.89	0.33	**1.01**	0.82
SWC	CHM_PRE V2	**1.92**	**0.69**	0.96	0.85
	CHM_PRE V1	2.06	0.64	1.07	0.78
	GSMaP	3.03	0.41	0.92	0.82
	IMERG	3.39	0.38	**1.02**	**0.89**
	PERSIANN-CDR	3.87	0.23	1.03	0.78
	GLDAS	4.03	0.23	1.05	0.87
IM	CHM_PRE V2	**0.57**	**0.71**	1.15	**0.85**
	CHM_PRE V1	0.61	0.65	1.18	0.78
	GSMaP	1.12	0.41	**1.11**	0.83
	PERSIANN-CDR	1.43	0.14	1.27	0.56
	IMERG	1.26	0.32	1.26	0.78
	GLDAS	1.45	0.16	1.27	0.65
NWC	CHM_PRE V2	**0.47**	**0.56**	1.12	0.73
	CHM_PRE V1	0.52	0.49	1.18	0.66
	GSMaP	0.71	0.33	**0.96**	**0.81**
	PERSIANN-CDR	0.96	-0.02	1.26	0.41
	IMERG	0.82	0.21	1.15	0.67
	GLDAS	0.91	0.09	1.07	0.64
QT	CHM_PRE V2	**1.18**	**0.71**	**1.02**	0.84
	CHM_PRE V1	1.27	0.66	1.11	0.79
	GSMaP	2.17	0.37	1.01	0.81

	IMERG	2.35	0.34	1.03	0.93
	PERSIANN-CDR	2.91	0.15	1.35	0.73
	GLDAS	2.82	0.23	1.14	**0.95**

Table S7. Precipitation event accuracy of different datasets validated by high-density gauge data. The bolded numbers in the column represent the optimal accuracy values for that metric.

Dataset Name	HSS	F1 Score	Accuracy	POD	FAR
CHM_PRE V2	**0.68**	**0.80**	**0.85**	0.84	**0.24**
CHM_PRE V1	0.58	0.75	0.79	**0.93**	0.37
GSMaP	0.50	0.67	0.78	0.65	0.31
IMERG	0.39	0.62	0.71	0.69	0.43
PERSIANN-CDR	0.21	0.54	0.59	0.70	0.55
GLDAS	0.29	0.54	0.68	0.55	0.47

Table S8. Precipitation event accuracy validated by high-density gauge data in different regions. The bolded numbers in the column represent the optimal accuracy values for that metric.

Region Abbreviation	Dataset Name	HSS	F1 Score	Accuracy	POD	FAR
NEC	CHM_PRE V2	**0.61**	**0.72**	**0.84**	0.84	**0.37**
	CHM_PRE V1	0.54	0.68	0.79	**0.90**	0.45
	GSMaP	0.49	0.62	0.81	0.63	0.39
	IMERG	0.38	0.56	0.74	0.68	0.53
	GLDAS	0.37	0.55	0.73	0.69	0.54
	PERSIANN-CDR	0.16	0.45	0.54	0.78	0.68
NC	CHM_PRE V2	**0.67**	**0.74**	**0.88**	0.80	**0.30**
	CHM_PRE V1	0.55	0.67	0.81	**0.90**	0.46
	GSMaP	0.47	0.59	0.82	0.61	0.44
	IMERG	0.34	0.51	0.73	0.67	0.58
	GLDAS	0.27	0.45	0.72	0.55	0.61
	PERSIANN-CDR	0.16	0.42	0.54	0.78	0.71
SCC	CHM_PRE V2	**0.68**	**0.83**	**0.84**	0.84	**0.19**
	CHM_PRE V1	0.56	0.79	0.77	**0.94**	0.32
	GSMaP	0.47	0.69	0.74	0.66	0.27
	IMERG	0.37	0.66	0.68	0.68	0.36
	GLDAS	0.25	0.56	0.63	0.53	0.40
	PERSIANN-CDR	0.24	0.61	0.62	0.67	0.44
SWC	CHM_PRE V2	**0.67**	**0.81**	**0.84**	0.85	**0.23**
	CHM_PRE V1	0.58	0.77	0.78	**0.94**	0.35
	GSMaP	0.54	0.71	0.78	0.69	0.26
	IMERG	0.47	0.69	0.74	0.74	0.35
	PERSIANN-CDR	0.39	0.65	0.70	0.70	0.40
	GLDAS	0.37	0.62	0.70	0.63	0.39
IM	CHM_PRE V2	**0.66**	**0.73**	**0.89**	0.82	**0.35**
	CHM_PRE V1	0.59	0.68	0.85	**0.90**	0.45
	GSMaP	0.48	0.57	0.85	0.58	0.43
	IMERG	0.38	0.51	0.79	0.61	0.56
	GLDAS	0.32	0.47	0.76	0.58	0.61
	PERSIANN-CDR	0.20	0.41	0.61	0.77	0.72
NWC	CHM_PRE V2	**0.52**	**0.61**	0.85	0.80	0.51
	CHM_PRE V1	0.45	0.56	0.79	**0.89**	0.59
	GSMaP	0.41	0.50	**0.85**	0.51	**0.51**
	IMERG	0.31	0.44	0.76	0.65	0.67
	GLDAS	0.21	0.36	0.75	0.48	0.71
	PERSIANN-CDR	0.12	0.32	0.53	0.78	0.80
QT	CHM_PRE V2	**0.61**	**0.77**	**0.81**	0.88	**0.32**
	CHM_PRE V1	0.55	0.75	0.77	**0.94**	0.38

GSMaP	0.51	0.69	0.77	0.69	0.32
IMERG	0.41	0.65	0.71	0.74	0.42
GLDAS	0.33	0.57	0.69	0.56	0.42
PERSIANN-CDR	0.19	0.57	0.56	0.82	0.56