Applicability of Land Surface Assimilation and Reanalysis Precipitation Data in Inner Mongolia: A Postprint

Authors: Song Haiqing

Date: 2021-12-12T00:00:00+00:00

Abstract

The reliability of four land surface assimilation and reanalysis precipitation datasets (FLDAS, ERA5, CRA40/Land, and GLDAS) in Inner Mongolia was evaluated using monthly precipitation observations from 115 meteorological stations in the region during 1982–2018. The results indicate that: (1) All four precipitation datasets can reasonably characterize the spatiotemporal variation features of precipitation in Inner Mongolia, specifically the decreasing trend from northeast to west and the pattern of low precipitation in winter and high precipitation in summer. (2) Calculations of correlation coefficient, mean absolute error, root mean square error, and Nash-Sutcliffe efficiency coefficient reveal that the four precipitation datasets perform best in summer and worst in winter, and exhibit better performance in the eastern semi-humid and semi-arid regions than in the western arid and extremely arid regions. (3) The ERA5 dataset shows overestimation of precipitation in Inner Mongolia during most periods, while GLDAS exhibits underestimation, particularly in winter when GLDAS has almost no capability to observe solid precipitation; the newly released FLDAS dataset demonstrates better performance. Overall, compared with ERA5 and GLDAS, the FLDAS and CRA40/Land precipitation datasets show the lowest differences from observations and possess the most favorable statistical characteristics.

Full Text

Arid Zone Research, Vol. 38 No. 6, Nov. 2021 (ChinaXiv Partner Journal)

Applicability of Land Surface Assimilation and Reanalysis Precipitation Datasets in Inner Mongolia

SONG Haiqing[1,2], **ZHU Zhongyuan**[1], **LI Yunpeng**[2]

[1]Water Conservancy and Civil Engineering College, Inner Mongolia Agricultural University, Hohhot 010018, Inner Mongolia, China
[2]Ecology and Agrometeorology Center of Inner Mongolia, Hohhot 010051, Inner Mongolia, China

Abstract

Using monthly precipitation observations from 115 meteorological stations across Inner Mongolia during 1982–2018, this study systematically evaluated the reliability of four land surface assimilation and reanalysis precipitation datasets: the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS), the fifth-generation ECMWF reanalysis (ERA5), the China Meteorological Administration Global Land Surface Reanalysis (CRA40/Land), and the Global Land Data Assimilation System (GLDAS). The results demonstrate that all four datasets effectively capture the characteristic spatiotemporal precipitation patterns of Inner Mongolia, specifically the gradient decreasing from northeast to west and the seasonal cycle of low winter precipitation and high summer precipitation. Statistical analyses including correlation coefficients, mean absolute errors, root mean square errors, and Nash–Sutcliffe efficiency coefficients reveal that ERA5 consistently overestimates precipitation across most of the region, while GLDAS exhibits underestimation, particularly in winter, where it shows virtually no capacity to capture solid precipitation. The newly released FLDAS dataset demonstrates superior performance in this regard. Overall, FLDAS and CRA40/Land exhibit the smallest deviations from observational data and display the most favorable statistical characteristics. Performance is optimal during summer and poorest during winter, with better accuracy in the eastern semi-humid and semi-arid zones compared to the western arid and extremely arid zones.

Keywords: precipitation; land assimilation data; reanalysis data; applicability

1. Introduction

Precipitation is a critical controlling factor in the global water cycle and one of the most dynamic components of weather and climate, influencing and regulating climate systems. Accurate and reliable precipitation data are essential for weather and climate research, hydrological modeling, and water resource management. Currently, precipitation data are obtained through three primary approaches: rain gauge observations, weather radar or satellite remote sensing retrievals, and numerical model simulations. While gauge observations provide the most accurate measurements, their spatial representativeness is limited, and they are affected by topography and other conditions with uneven spatial distribution; complex terrain regions and remote areas often have sparse observation networks. In recent years, satellite-based retrieval of high-resolution gridded precipitation data has developed rapidly, with commonly used products including TRMM, GPM, CMORPH, and CHIRPS. Additionally, land data assimilation and reanalysis datasets provide spatially and temporally continuous global and regional precipitation data, compensating for deficiencies in observational coverage and becoming indispensable for precipitation research and applications in ungauged regions.

With rapid advances in computing power, continuous improvements in numerical models, and innovative data assimilation methods, land data assimilation and reanalysis datasets have achieved higher spatiotemporal resolution and substantially improved accuracy. Numerous studies have evaluated the quality and performance of various precipitation datasets across different regions, particularly newly developed products, yielding valuable results for precipitation and hydrological research in mountainous areas, continents, and specific regions of China. For example, Chen et al. assessed the performance of multiple reanalysis precipitation datasets in Nepal on the southern slope of the Himalayas using station observations, finding similar performance at low elevations but that only high-resolution datasets could reflect topographic effects on precipitation at high altitudes. Dinku et al. evaluated satellite precipitation products in Africa and South America, revealing systematic underestimation. Studies across China have shown that reanalysis datasets generally tend to overestimate precipitation while satellite products tend to underestimate, with satellite data showing slightly better quality overall. The accuracy of these products is influenced by topography, elevation, and seasonality, exhibiting distinct spatiotemporal variations.

Inner Mongolia, located in northern China at the boundary between monsoon and non-monsoon regions, serves as a crucial ecological barrier in northern China. The region transitions from semi-humid in the east to semi-arid, arid, and extremely arid zones toward the west, with complex and diverse topography. Covering approximately one-eighth of China' s total area, Inner Mongolia is vast and sparsely populated with relatively sparse meteorological stations, particularly lacking observations in forested and arid regions. Consequently, evaluating multiple satellite and reanalysis gridded precipitation datasets is of significant importance. This study employs monthly precipitation observations from 115 national meteorological stations across Inner Mongolia to systematically assess the spatiotemporal distribution and statistical characteristics of four high-resolution land assimilation and reanalysis precipitation datasets, providing a foundation for better understanding precipitation variability and distribution patterns in the region.

2. Data and Methods

2.1 Data Sources

2.1.1 Observational Precipitation Data Monthly precipitation data from 115 national surface observation stations were obtained from the Inner Mongolia Meteorological Bureau. To evaluate dataset performance across different climate regimes, Inner Mongolia was divided into four climate zones based on mean annual precipitation: semi-humid (>400 mm), semi-arid (200–400 mm), arid (100–200 mm), and extremely arid (<100 mm) zones. Climate zone divisions and station statistics are presented in .

2.1.2 Land Surface Assimilation and Reanalysis Precipitation Datasets Four high-resolution monthly precipitation datasets were selected for evaluation:

1. **FLDAS**: The Famine Early Warning Systems Network Land Data Assimilation System precipitation dataset, developed by NASA for food and water security monitoring. The dataset uses a 0.1°$×$0.1° spatial resolution with monthly temporal resolution, covering 1982–2018. Outside the 50°S-50°N coverage area, CHIRPS precipitation data are used for gap-filling.
2. **ERA5**: The fifth-generation global reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF), featuring 31 km spatial resolution and 4DVar assimilation. The precipitation data span 1982–2018 at monthly resolution.
3. **CRA40/Land**: China' s first-generation land surface reanalysis dataset produced by the China Meteorological Administration, with 34 km spatial resolution and monthly temporal resolution covering 1982–2018. This product merges global gauge analysis data and satellite precipitation using data fusion methods.
4. **GLDAS**: The second-generation Global Land Data Assimilation System product from NASA, with 0.25°$×$0.25° spatial resolution and monthly temporal resolution from 1982–2018. GLDAS-2.1 uses the ITPCAS dataset as background and merges CMA routine precipitation observations, demonstrating good quality across China.

2.2 Methods

To quantitatively evaluate the applicability of the four precipitation datasets in Inner Mongolia, we calculated correlation coefficients (CC), mean absolute errors (MAE), mean bias errors (MBE), root mean square errors (RMSE), and Nash–Sutcliffe efficiency coefficients (NSE) against observational data:

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(M_i - O_i)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - O_i)^2}$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}(M_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$$

where n is the sample size, i is the time variable, M represents modeled values, O represents observed values, and $\bar{O}$ is the mean observed precipitation. Taylor diagrams were also employed to visually assess overall performance.

Given the varying spatial resolutions of the datasets, bilinear interpolation was used to regrid all products to the GLDAS resolution (0.25°$×$0.25°). To minimize errors from interpolating gridded data to station locations, the nearest grid point to each station was used for comparison.

3. Results and Analysis

3.1 Spatial Distribution of Multi-Year Mean Precipitation

The spatial distribution of multi-year mean precipitation from the four datasets shows consistent patterns with observations, all capturing the characteristic decrease from northeast to west across Inner Mongolia [Figure 1: see original paper]. While the spatial patterns are similar among datasets, ERA5 generally shows overestimation across most regions. CRA40/Land and GLDAS show smaller differences from observations, with FLDAS performing best in capturing local details.

Bias distributions relative to observations reveal larger discrepancies in Hulunbuir (northeast) and central-southern Inner Mongolia [Figure 2: see original paper]. FLDAS shows the best performance with biases mostly within $±$20 mm, slightly underestimating precipitation in most areas except for minor overestimation in northern Hulunbuir and southeastern Inner Mongolia. ERA5 overestimates precipitation across most of the study area, particularly in Hulunbuir and central-southern regions with biases exceeding 80 mm. CRA40/Land overestimates precipitation in most of Hulunbuir (exceeding 80 mm in the north) while underestimating in central-southern areas and southeastern Alxa League. GLDAS underestimates precipitation across most regions except for slight overestimation in southern Tongliao, with particularly large underestimation (>80 mm) in central-western areas.

3.2 Temporal Variation

Comparisons of monthly precipitation time series show strong consistency between all four datasets and observations across Inner Mongolia and its climate zones [Figure 3: see original paper]. All datasets capture the seasonal pattern of maximum precipitation in summer and minimum in winter. The temporal characteristics of the semi-humid and semi-arid zones most closely resemble the regional average, while arid and extremely arid zones show poorer agreement, likely due to lower precipitation amounts and sparser station networks.

Monthly bias time series reveal distinct seasonal patterns [Figure 4: see original paper]. ERA5 consistently overestimates precipitation across all climate zones. GLDAS shows overestimation in semi-humid, semi-arid, and arid zones but less underestimation in extremely arid zones. Notably, GLDAS demonstrates virtually no capacity to capture solid precipitation in winter, indicating poor snow observation capability. The upgraded GLDAS-2.1 shows improved winter performance, suggesting enhanced solid precipitation detection. FLDAS exhibits overestimation across Inner Mongolia, semi-humid, and semi-arid zones, with bias trends increasing over time, possibly due to inadequate solid precipitation detection. CRA40/Land shows underestimation during 1991–1995 and 2003–2005, with otherwise small biases.

3.3 Statistical Analysis of Monthly Precipitation

Statistical characteristics were calculated for each month [Figure 6: see original paper]. Correlation coefficients exceed 0.6 for most datasets across all months except winter, when GLDAS shows significantly lower values, indicating poor representation of winter precipitation variability. All datasets exhibit clear seasonal patterns in MAE and RMSE, with smallest errors in winter and largest in summer, reflecting higher precipitation variability during warm seasons. In terms of bias, FLDAS shows the smallest mean bias, while ERA5 shows the largest. CRA40/Land demonstrates the best NSE values, followed by FLDAS and GLDAS, with ERA5 performing worst.

Spatial distributions of correlation coefficients show a decreasing gradient from northeast to west [Figure 7: see original paper], with all datasets performing better in the semi-humid and semi-arid regions of eastern and central Inner Mongolia than in western arid and extremely arid zones. CRA40/Land shows the highest overall correlation coefficients. RMSE distributions [Figure 8: see original paper] similarly show larger errors in semi-humid and semi-arid zones where precipitation amounts are greater.

Taylor diagrams for monthly precipitation [Figure 9: see original paper] demonstrate that all four datasets perform well in the semi-humid and semi-arid zones, with correlation coefficients above 0.7 and normalized standard deviations close to 1. CRA40/Land shows the best overall performance, followed by FLDAS, while ERA5 and GLDAS exhibit larger biases. Performance is poorest in the extremely arid zone, where all datasets struggle to capture precipitation patterns.

3.4 Overall Performance Assessment

Comprehensive evaluation indicates that FLDAS and CRA40/Land exhibit the highest quality among the four datasets. While ERA5 and GLDAS show relatively high correlation coefficients, their biases are substantial. All datasets demonstrate superior performance in semi-humid and semi-arid zones compared to arid and extremely arid zones, likely due to limitations in simulating low-precipitation regions. The overall performance varies by season, with best performance in summer and poorest in winter.

4. Discussion

The large east-west gradient in precipitation across Inner Mongolia reflects strong regional heterogeneity, primarily influenced by latitude, elevation, slope, and topography. Major mountain ranges including the Da Hinggan, Yinshan, and Helan Mountains lie along the East Asian monsoon margin. The eastern and southern slopes receive more moisture from the monsoon, while northwestern slopes receive less, creating higher precipitation in the southeast. These mountainous areas pose challenges for numerical simulation. Reanalysis precipitation calculations are affected by model resolution and physical parameterizations; the 34 km grid spacing cannot resolve fine-scale orographic precipitation. The 0.1°$×$0.1° resolution of FLDAS provides better representation of local precipitation details.

In terms of data fusion, FLDAS incorporates more observational data including satellite infrared observations, CHIRPS monthly precipitation, CHPclim climatology, and TRMM3B42 products, supplemented by gauge observations. This extensive observational input contributes to its superior statistical performance compared to GLDAS, which relies primarily on the Princeton meteorological forcing dataset. CRA40/Land and GLDAS are both land data assimilation products, but CRA40/Land’ s higher resolution (0.1°$×$0.1°*versus*0.25°×$0.25°) and incorporation of more Chinese gauge observations result in better performance across Inner Mongolia.

The notable negative bias in CRA40/Land during 1991–1995 and 2003–2005 requires further investigation. While FLDAS demonstrates the best overall quality, its consistent overestimation warrants attention. The higher spatial resolution of FLDAS and CRA40/Land enables better representation of local details, making them valuable supplements to sparse observational networks in ungauged regions and providing essential data sources for land surface and hydrological modeling.

5. Conclusions

This study systematically evaluated four monthly precipitation datasets against gauge observations across Inner Mongolia and its climate zones from 1982–2018. Key conclusions are:

1. All four datasets capture the observed precipitation gradient decreasing from northeast to west. FLDAS and CRA40/Land, with their higher spatial resolutions, best reproduce the spatial distribution, though FLDAS

tends to overestimate while GLDAS underestimates, particularly in winter where GLDAS fails to capture solid precipitation.

2. Temporally, all datasets show highest biases in summer and lowest in winter. FLDAS and CRA40/Land exhibit the smallest deviations from observations, while ERA5 consistently overestimates and GLDAS underestimates. The bias trends in FLDAS increase after 2000.
3. Statistically, correlation coefficients are highest in summer and lowest in winter for all datasets. Spatially, correlations decrease from northeast to west, with better performance in semi-humid and semi-arid zones. CRA40/Land shows the best overall statistical characteristics, followed by FLDAS.
4. Overall, FLDAS and CRA40/Land demonstrate the highest quality among the evaluated datasets, with superior performance in semi-humid and semi-arid regions compared to arid and extremely arid regions. Seasonal performance is best in summer and poorest in winter.

High-resolution, high-quality gridded precipitation datasets provide valuable supplements to sparse observational networks, offering pathways for monitoring precipitation-related disasters and providing data sources for land surface and hydrological modeling in data-scarce regions.

References

[1] Kidd C, Huffman G. Global precipitation measurement[J]. Meteorological Applications, 2011, 18(3): 334-353.

[2] Chen Y, Sharma S, Zhou X, et al. Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya[J]. Atmospheric Research, 2020, 250: 105365.

[3] Yan Yan, Liu Gang, He Jun, et al. Assessment of satellite and reanalysis precipitation data in Chongqing[J]. Plateau Meteorology, 2020, 39(3): 594-608.

[4] Tapiador F J, Turk F J, Petersen W, et al. Global precipitation measurement: Methods, datasets and applications[J]. Atmospheric Research, 2012, 104-105: 70-97.

[5] Wang Jun, Wang Huijun, Hong Yang. A high resolution flood forecasting and monitoring system for China using satellite remote sensing data[J]. Chinese Science Bulletin, 2016, 61(Suppl.): 518-528.

[6] Liu Tian, Yang Kun, Qin Jun, et al. Construction and applications of time series of monthly precipitation at weather stations in the central and eastern Qinghai Tibetan plateau[J]. Plateau Meteorology, 2018, 37(6): 1449-1457.

[7] Yu Jingjing, Shen Yan, Pan Yang, et al. Improvement of satellite based precipitation estimates over China based on probability density function matching method[J]. Journal of Applied Meteorological Science, 2013, 24(5): 544-553.

[8] Pan Yang, Gu Junxia, Yu Jingjing, et al. Test of merging methods for multi source observed precipitation products at high resolution over China[J]. Acta Meteorologica Sinica, 2018, 76(5): 755-766.

[9] Lu D, Yong B. Evaluation and hydrological utility of the latest GPM IMERG V5 and GSMaP V7 precipitation products over the Tibetan Plateau[J]. Remote Sensing, 2018, 10(12): 2022.

[10] Sharma S, Chen Y, Zhou X, et al. Evaluation of GPM Era satellite precipitation products on the southern slopes of the central Himalayas against rain gauge data[J]. Remote Sensing, 2020, 12(11): 1836.

[11] Dinku T, Connor S J, Ceccato P. Comparison of CMORPH and TRMM-3B42 over Mountainous Regions of Africa and South America: Satellite Rainfall Applications for Surface Hydrology[M]. Dordrecht: Springer Press, 2010.

[12] Funk C, Peterson P, Landsfeld M, et al. The climate hazards infrared precipitation with stations: A new environmental record for monitoring extremes[J]. Scientific Data, 2015, 2(1): 1-21.

[13] Sahlu D, Moges S A, Nikolopoulos E I, et al. Evaluation of high resolution multi satellite and reanalysis rainfall products over East Africa[J]. Advances in Meteorology, 2017, 2017: 1-14.

[14] Jiang S, Ren L, Hong Y, et al. Comprehensive evaluation of multi satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method[J]. Journal of Hydrology, 2012, 452-453: 213-225.

[15] Wang Wen, Wang Xingchen, Liu Lichao. Characteristics of precipitation in the surrounding area of Tengger Desert in 1957-2017[J]. Journal of Desert Research, 2020, 40(4): 163-170.

[16] Wang Yudan, Chen Hao, Liu Canran, et al. Applicability of ITPCAS and CMORPH precipitation datasets over Shaanxi province[J]. Arid Zone Research, 2018, 35(3): 579-588.

[17] Zhang Meng, Huang Anning, Ji Xiaolong, et al. Validation of satellite precipitation products over Qinghai Xizang Plateau Region[J]. Plateau Meteorology, 2016, 35(1): 34-42.

[18] Dee D P, Uppala S M, Simmons A J, et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system[J]. Quarterly Journal of the Royal Meteorological Society, 2011, 137(656): 553-597.

[19] Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis[J]. Quarterly Journal of the Royal Meteorological Society, 2020, 146(730): 1999-2049.

[20] Saha S, Moorthi S, Pan H L, et al. The NCEP climate forecast system reanalysis[J]. Bulletin of the American Meteorological Society, 2010, 91(8): 1015-1058.

[21] He J, Yang K, Tang W, et al. The first high resolution meteorological forcing dataset for land process studies over China[J]. Scientific Data, 2020, 7(25): 1-12.

[22] Li Chunxiang, Zhao Tianbao, Shi Chunxiang, et al. Evaluation of daily precipitation product in China from the CMA global atmospheric interim reanalysis[J]. Journal of Meteorological Research, 2020, 34(1): 117-136.

[23] Molod A, Takacs L, Suarez M, et al. Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2[J]. Geoscientific Model Development, 2015, 8(5): 1339-1356.

[24] McNally A, Arsenault K, Kumar S, et al. A land data assimilation system for sub-Saharan Africa food and water security applications[J]. Scientific Data, 2017, 4(1): 1-19.

[25] McNally A, Verdin K, Harrison L, et al. Acute water scarcity monitoring for Africa[J]. Water, 2019, 11(10): 1968.

[26] Shen Luting, Zhang Fangmin, Huang Jin, et al. Climate characteristics of day and night precipitation during the growing season in Inner Mongolia from 1961 to 2018[J]. Arid Zone Research, 2020, 37(6): 1519-1527.

[27] Liu Chuan, Yu Ye, Xie Jin, et al. Applicability of soil temperature and moisture in several datasets over Qinghai Xizang Plateau[J]. Plateau Meteorology, 2015, 34(3): 653-665.

[28] Smith R B, Barstad I. A linear theory of orographic precipitation[J]. Journal of Atmospheric Sciences, 2004, 61(12): 1377-1391.

[29] Blocken B, Poesen J, Carmeliet J. Impact of wind on the spatial distribution of rain over micro-scale topography: Numerical modeling and experimental verification[J]. Hydrological Processes, 2006, 20: 345-368.

[30] Fallah A, Rakhshandehroo G R, Berg P, et al. Evaluation of precipitation datasets against local observations in southwestern Iran[J]. International Journal of Climatology, 2020, 40(9): 4102-4116.

[31] Wang Y, Yang K, Zhou X, et al. Synergy of orographic drag parameterization and high resolution greatly reduces biases of WRF-simulated precipitation in central Himalaya[J]. Climate Dynamics, 2020, 54: 1729-1740.

Note: Figure translations are in progress. See original paper for figures.