RESEARCH

Developing ensemble mean models of satellite remote sensing, climate reanalysis, and land surface models

Mohammad Valipour[1,2] · Jörg Dietrich[2]

Abstract

This study aims to access the selected satellite remote sensing, climate reanalysis, and land surface models to estimate monthly land surface air temperature (LSAT), solar radiation (SR), and precipitation (P) at the global scale. To this end, we apply six datasets including Modern-Era Retrospective Analysis for Research and Applications-version 2 (MERRA-2), European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis-version 5 (ERA-5), ERA-5-Land version (ERA5-Land), Global Land Data Assimilation System (GLDAS), Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FL and Global Precipitation Climatology Project (GPCP). In terms of SR, we compare the selected products against the National Oceanic and Atmospheric Administration (NOAA)-Cooperative Institute for Research in Environmental Sciences (CIRES)-Department of Energy (DOE) Twentieth Century Reanalysis (20CR) (NOAA-CIRES-DOE 20CR) dataset from 1982 to 2015. For LSAT and P, we consider NOAA Climate Prediction Center (CPC) (NOAA-CPC) as the reference dataset in the periods of 1982–2020 and 1983–2019, respectively, based on available data. ERA5-Land, MERRA-2, and GLDAS show the best results with root mean square difference (RMSD) equal to 19.03 W/m^2, 1.93 °C, and 37.61 mm/month for SR, LSAT, and P estimates compared to NOAA datasets. Since there are uncertainties in all of the products, here we introduce new datasets based on merging the best products concerning their accuracy. The evaluation results can be used also as feedback to developers to improve the products and to facilitate the users to understand the status of the products and better use them for practical applications on a global scale.

1 Introduction

Land surface air temperature (LSAT), solar radiation (SR), and precipitation (P) are the main descriptors of terrestrial environmental conditions with relevance to the hydrological cycle across the earth. These three variables play crucial roles on our planet and having accurate knowledge about them can help experts, researchers, managers, and stakeholders to adapt their objectives, missions, and policies to improve water resources management and decision support

to achieve sustainable agriculture in the future. However, measuring these three variables is time-consuming, and there are not enough weather stations to cover all regions of the globe.

The state-of-the-art reanalysis, land surface models, and remote sensing retrievals can help us to estimate these three variables to save time and cost. There are a lot of products to capture LSAT, SR, and P dynamics. However, previous investigations indicate better performance of some of them compared to others. Those are the National Oceanic and Atmospheric Administration (NOAA)-Cooperative Institute for Research in Environmental Sciences (CIRES)-Department of Energy (DOE) Twentieth Century Reanalysis (20CR) (NOAA-CIRES-DOE 20CR) project (Slivinski et al. 2019), NOAA Climate Prediction Center (CPC) (NOAA-CPC) (Xie et al. 2007; Chen et al. 2008), Modern-Era Retrospective analysis for Research and Applications-version 2 (MERRA-2) (Gelaro et al. 2017), European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis-version 5 (ERA-5) (Albergel et al., 2018; Hersbach et al. 2020), ERA5-Land version (ERA5-Land) (Muñoz-Sabater

✉ Mohammad Valipour
mvalipou@msudenver.edu

Jörg Dietrich
dietrich@iww.uni-hannover.de

[1] Department of Engineering and Engineering Technology, Metropolitan State University of Denver, Denver, CO 80217, USA

[2] Institute of Hydrology and Water Resources Management, Leibniz University Hannover, Appelstraße 9A, 30167 Hannover, Germany

et al. 2021), Global Land Data Assimilation System (GLDAS) (Rodell et al. 2004), Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) (McNally et al. 2017), and Global Precipitation Climatology Project (GPCP) (Huffman et al. 1997, 2009; Adler et al. 2003). Here, we review the literature based on regional and global scales.

1.1 Regional studies

Han et al. (2020) compared the GLDAS dataset against in-situ LSAT measurements in China. The results indicated that GLDAS is significantly correlated with observations. However, more caution is necessary when using the data in mountain regions as the accuracy of GLDAS gradually decreases with increasing altitude due to the lack of enough observational stations. Liu et al. (2021) compared the ECMWF Reanalysis-Interim version (ERA-Interim), GLDAS, National Centers for Environmental Prediction (NCEP), and ERA-5 in terms of estimation of LSAT in the Tibetan Plateau, in China. They suggest that GLDAS and ERA-5 are superior products compared to the other datasets for measuring weather sites. He et al. (2021) indicated that SR overestimation decreased from 15.88 W/m^2 in ERA-Interim to 10.07 W/m^2 in ERA-5 over China from 1979 to 2014. Similar to this, Jiang et al. (2019) claimed ERA-5 overestimates SR across China. In the other study, Zhang et al. (2020) observed an overestimation of MERRA-2 and ERA-Interim in terms of SR in China. Based on their results, cloud coverage, aerosol optical depth, and water vapor content are the main factors of errors to estimate SR.

Jiang et al. (2021) evaluated the accuracy of ERA-5 P in China. They indicated ERA-5 has higher root-mean-square difference (RMSD) values in the tropical and subtropical regions with a relatively wet climate. ERA-5 tends to overestimate (underestimate) light (moderate and heavy) P events compared to satellite-based P. ERA-5 can recognize the P distribution and center, but underestimates extreme P. Song and Wei (2021) compared ERA-5 and MERRA-2 in terms of P estimation over the North China Plain. They reported that the land surface is more strongly coupled with the atmosphere in MERRA-2 dataset compared to ERA-5. Indeed, MERRA-2 P is more sensitive to precipitable water. The low-level wind field is more divergent in MERRA-2 than in ERA-5, which causes weaker ascending motions, less precipitable water, and lower P efficiency. The less favorable atmospheric conditions for P, exacerbated by the strong land–atmosphere coupling, lead to occurring errors MERRA-2 in terms of P estimates.

Chen et al. (2020) successfully applied GLDAS and MERRA-2 P data to estimate drought in Northern China. Wang et al. (2016) successfully simulated P and LSAT using the GLDAS dataset in China from 1979 to 2010. Chen et al.

(2021) estimated P in the Yangtze River Basin by using GLDAS. Based on their results, it is feasible to apply satellite-based grid P products to replace the measured data for the regional studies of extreme P. Wang et al. (2020) successfully merged gauge, climate reanalysis, and satellite P datasets (including MERRA-2 and GLDAS) for the largest river basin of the Tibetan Plateau in China. Yao et al. (2020) evaluated NOAA, GPCP, ERA-Interim, and MERRA-2 to estimate P in the arid region of northwestern China. According to the results, all the products reasonably reproduce patterns of P in the study area, with a bias of less than 1.5 mm/day. However, there are differences among estimated patterns of P by NOAA, GPCP, ERA-Interim, and MERRA-2. NOAA, GPCP, and MERRA-2 show more accurate results, respectively. However, ERA-Interim overestimates P in mountainous zones due to the lack of enough observational stations. They mentioned that a systematic assessment of the differences between multiple products is critical to reduce discrepancies.

In Europe, Babar et al. (2019) reported overestimation (underestimation) of SR by using ERA-5 (Cloud, Albedo, Radiation dataset Edition 2 (CLARA)) at high latitudes in Norway. Khatibi and Krauter (2021) showed that the correlation coefficient (*R*) between MERRA-2 SR (LSAT) and measured data is 0.97 (0.99) in Germany.

In Africa, Gleixner et al. (2020) assessed the performance of ERA-Interim and ERA-5 in terms of LSAT and P estimation. According to the results, in ERA-5, the climatological biases in LSAT and P are significantly decreased, and the representation of interannual variability is improved over most of Africa. However, ERA-Interim and ERA-5 performed less well to capture the observed long-term trends, despite a slightly better accuracy of ERA-5 compared to ERA-interim. The representation of the annual cycle of P is substantially improved in ERA-5 by decreasing the wet bias during the rainy season over East Africa. In addition, ERA-5 performs better in terms of the spatial distribution of P during extreme years. Khalil et al. (2021) showed that the R between ERA-5 SR and measured data, in Egypt, is 0.96.

In Asia, Mokhtari et al. (2018) successfully applied the GLDAS dataset to estimate SR. They suggest GLDAS outperforms a satellite observation model (Satellite Application Facility on Climate Monitoring, CM-SAF) in Iran, in case of a lack of meteorological data. Hamal et al. (2020) revealed that MERRA-2 can estimate P in Nepal with $R = 0.63$ in comparison with gauge-based data.

In America, Tarek et al. (2020) compared the performance of ERA-5 and ERA-Interim against in situ measurements in terms of P and LSAT. They showed that ERA-5 outperforms ERA-Interim in North America. In ERA-Interim, there are large biases in mountainous regions, where observation networks are generally considered less robust. However, ERA-5 well corrects the biases in ERA-Interim.

1.2 Global studies

Ji et al. (2015) compared GLDAS LSAT against NOAA over the globe from 2000 to 2011. They showed that the values of RMSD range between 3.76 and 3.93 °C. Hinkelman (2019) evaluated the performance of MERRA and MERRA-2 to estimate SR at the global scale. She mentioned clouds are overrepresented over the tropical oceans in MERRA and MERRA-2, and somewhat underrepresented in marine stratocumulus regions. MERRA-2 also shows signs of excess cloudiness in the Southern Ocean. Delgado-Bonal et al. (2020) analyzed changes in the complexity of climate in the last four decades using MERRA-2 SR. They suggest the observed behavior of climatic complexity could be due to the changes in cloud amount, and they assess that possibility by evaluating its evolution from a complexity perspective by information from the International Satellite Cloud Climatology Project (ISCCP).

Sun and Fu (2021) merged the Tropical Rainfall Measuring Mission (TRMM) and ERA-5 P to build a robust dataset. The results appear that the accuracy of the combined dataset is reasonable. Reichle et al. (2017) compared MERRA, MERRA-2, and MERRA-Land against GPCP dataset. MERRA-2 outperforms MERRA and MERRA-Land because in MERRA-2, a merged satellite-gauge P is applied instead of the gauge-only information applied for MERRA and MERRA-Land. Correcting the P within the coupled atmosphere-land modeling system allows the MERRA-2 LSAT and humidity to respond to the improved P forcing. At high latitudes, however, the lack of sufficient and reliable P observations results in undesirable land spin-up effects that impact MERRA-2 P estimates.

Hobeichi et al. (2020) evaluated MERRA-2 and GPCP to estimate P over the globe. The results show better accuracy of GPCP, especially over the tropics. However, both products suffer to capture P dynamics in the Middle East and North Africa (MENA). Li et al. (2021) compared Japanese 55-year reanalysis (JRA-55) and MERRA-2 P against GPCP. There is a good agreement between the reanalysis datasets and GPCP. However, JRA-55 produces more intense P with a larger bias, particularly over the Atlantic and Pacific intertropical convergence areas. Adler et al. (2017, 2018) and Smith et al. (2006) showed that GPCP is well able to capture extreme P and El Niño-Southern Oscillation events at the global scale. Nogueira (2020) compared ERA-Interim and ERA-5 against GPCP dataset. ERA-5 indicates lower bias and RMSD, as well as higher R, compared to ERA-Interim. ERA-Interim reveals significant P underestimation over the mid-latitude oceans because of underestimation of deep convection and moisture flux convergence. In addition, the results show an improved representation of the moisture sink/source patterns over the tropical oceans in ERA-5. However, there are significant differences in the P patterns of the three products, particularly in tropical Africa.

To the best of our knowledge, there is no study to compare new versions of climate reanalysis, remote sensing retrievals, and land surface models to estimate LSAT, SR, and P neither on a regional nor global scale. Therefore, the objective of this study is a comparison of all the mentioned successful models to assess the estimation of LSAT, SR, and P against NOAA products to identify more skillful products. The evaluation results can be used as feedback to developers to help them further improve the products, and to facilitate the users to understand the status of the products and better use them for practical applications. Finally, a new dataset is represented based on the combination of the best products in terms of LSAT, SR, and P. The new combined dataset can be applied to assess land-atmospheric systems and prepares an excellent opportunity for multi-source data analysis as well as for model simulations.

2 Datasets

2.1 NOAA

Using a state-of-the-art data assimilation system and surface pressure observations, the NOAA-CIRES-DOE 20CR project has generated a four-dimensional global atmospheric dataset of weather spanning 1836 to 2015 to place current atmospheric circulation patterns into a historical perspective (Slivinski et al. 2019). The reanalysis of NOAA-CIRES-DOE 20CR and gauge-based NOAA-CPC applies upgraded data assimilation approaches involving an adaptive inflation algorithm; has a newer, higher-resolution forecast model which specifies dry air mass; and assimilates a larger set of pressure observations. These changes have improved the ensemble-based estimates of confidence, removed spin-up effects in the P fields, and removed the sea-level pressure biases. Other developments include more accurate representations of storm intensity, smaller errors, and large-scale reductions in model bias (Slivinski et al. 2019).

One of the advantages of NOAA products is their ability to capture extreme events such as droughts, floods, and hurricanes (e.g., the Great Blizzard of 1888) (Slivinski et al. 2019). Previous studies show the successful application of NOAA-CIRES-DOE 20CR and NOAA-CPC in terms of LSAT (Smith and Reynolds 2005; Smith et al. 2008; Foster and Rahmstorf 2011; Ma et al. 2020), SR (Fallahi et al. 2018; Sengupta et al. 2018), and P (Chen et al. 2008; Xie et al. 2007). Therefore, in this study, we consider NOAA-CIRES-DOE 20CR as the reference for SR, and NOAA-CPC as the reference for LSAT and P, and compare the other products against these two NOAA datasets.

2.2 ERA-5

ERA-5 is the fifth generation ECMWF atmospheric reanalysis of the global climate. ERA-5 merges vast amounts of historical information into global estimates using advanced modeling and data assimilation systems (Muñoz-Sabater et al. 2021). Indeed, ERA-5 replaces the ERA-Interim reanalysis and has a finer spatiotemporal resolution than ERA-Interim. In ERA-5, the data cover the earth on a 30-km grid and 137 levels from the surface up to a height of 80 km are available. ERA-5 represents uncertainties for all variables. In addition, monthly and daily real-time data are available with a 3-month and 5-day delay, respectively (Muñoz-Sabater et al. 2021). In this study, the ERA5.1 has been used. This dataset contains ERA5.1 surface-level analysis parameter data for the period 2000–2006. ERA5.1 is the ECWMF ERA-5 reanalysis project re-run for 2000–2006 to improve upon the cold bias in the lower stratosphere seen in ERA-5 (European Centre for Medium-Range Weather Forecasts 2021).

2.3 ERA5-Land

ERA5-Land is a reanalysis dataset representing a consistent view of the evolution of land variables over several decades (Muñoz-Sabater et al. 2021). ERA5-Land replays the land component of the ECMWF ERA-5 climate reanalysis. ERA5-Land has a finer spatial resolution compared to ERA-5. ERA5-Land data are available from 1979 onward with 2–3 months behind the real-time data (Hersbach et al. 2020). ERA5-Land applies a revised land surface hydrology (HTESSEL) to address shortcomings of the land surface scheme, particularly the lack of surface runoff and the choice of a global uniform soil texture (Balsamo et al. 2009).

2.4 GLDAS

The goal of GLDAS is to ingest satellite- and ground-based observational datasets, by using advanced land surface modeling and data assimilation approaches, to generate optimal fields of land surface states and fluxes (Rodell et al. 2004). GLDAS integrates a huge quantity of observation-based information and executes globally at high resolutions (2.5° to 1 km). GLDAS can produce results in near-real-time (Rodell et al. 2004).

2.5 FLDAS

The goal of the FLDAS project is to achieve more effective use of limited available hydroclimatic observations (McNally et al. 2017). FLDAS is associated with food security assessment in data-sparse, developing country settings. Adopting a land information system allows FLDAS to

leverage existing land surface models and produce ensembles of climate variables based on multiple meteorological inputs or land surface models. FLDAS has a finer spatial resolution compared to the GLDAS dataset (McNally et al. 2017).

2.6 MERRA-2

MERRA-2 has been developed by the National Aeronautics and Space Administration (NASA) (Gelaro et al. 2017). Additional advances in both the Goddard Earth Observing System (GEOS) model and the Gridpoint Statistical Interpolation (GSI) assimilation system are included in MERRA-2. MERRA-2 replaces the original MERRA dataset and has a finer spatial resolution compared to MERRA and MERRA-Land datasets (Gelaro et al. 2017).

2.7 GPCP

GPCP has been developed by Meteorological Organization/World Climate Research Program/Global Energy and Water Experiment (WMO/WCRP/GEWEX). GPCP is one of the several GEWEX global analyses of components of the water and energy cycle organized under the GEWEX Radiation Panel (Huffman et al. 1997, 2009; Adler et al. 2003). GPCP information is essential to quantify the global water cycle, verify numerical models, and develop the background climate statistics for practical water resource projects. The GPCP dataset is developed and maintained as an international project among various universities and researchers (Huffman et al. 1997, 2009; Adler et al. 2003).

Table 1 shows general information of all datasets used in this study. We consider monthly data in the periods of 1982–2015, 1983–2019, and 1982–2020 to compare NOAA datasets with the other products in terms of SR, LSAT, and P, respectively. In addition, all the datasets are re-gridded to a spatial resolution of $1° \times 1°$ ($0.5° \times 0.5°$) using the Inverse Distance Weight Interpolation (IDWI) method (Burrough 1986) to be compatible with those of NOAA-CIRES-DOE 20CR (NOAA-CPC) (see also Chen et al. 2013; Wanders et al. 2014; Zhang et al. 2018). Since P cannot be retrieved directly from the FLDAS dataset, we have used GPCP version 3.1 as a source of P data to compare with NOAA.

3 Statistical metrics

In this study, we consider three statistical indices to evaluate the difference between NOAA products and the other datasets. These statistical indices are Pearson's correlation coefficient (R), root mean square difference (RMSD), and mean absolute difference (MAD):

Table 1 General information of all datasets used in this study

Product	Spatial resolution	Temporal resolution	Start year	End year	Reference
NOAA-CIRES-DOE 20CR	$1° \times 1°$	Monthly	1836	2015	https://psl.noaa.gov/
NOAA-CPC	$0.5° \times 0.5°$	Monthly	1948	Present	https://psl.noaa.gov/
ERA-5	$0.25° \times 0.25°$	Hourly	1979	Present	https://www.ecmwf.int/
ERA5-Land	$0.1° \times 0.1°$	Hourly	1979	Present	https://www.ecmwf.int/
GLDAS	$0.25° \times 0.25°$	Hourly	1948	Present	https://disc.gsfc.nasa.gov/
FLDAS	$0.1° \times 0.1°$	Hourly	1982	Present	https://disc.gsfc.nasa.gov/
MERRA-2	$0.5° \times 0.625°$	Hourly	1980	Present	https://disc.gsfc.nasa.gov/
GPCP	$0.5° \times 0.5°$	Daily	1983	2019	https://disc.gsfc.nasa.gov/

$$R = \frac{\sum_{i=1}^{i=N}\left(NOAA_i - \overline{NOAA}\right)\left(Product_i - \overline{Product}\right)}{\sqrt{\sum_{i=1}^{i=N}\left(NOAA_i - \overline{NOAA}\right)^2}\sqrt{\sum_{i=1}^{i=N}\left(Product_i - \overline{Product}\right)^2}} \quad (1)$$

where i is a counter, N is the number of data, $NOAA_i$ is ith NOAA observation, $\overline{NOAA}$ is the average of NOAA observations, and $Product_i$ is the ith value of products employed in this study (i.e., ERA-5, ERA5-Land, GLDAS, FLDAS, MERRA-2, and GPCP), $\overline{Product}$ is the average of values of products employed in this study. R is ranged between -1 (the strongest possible negative correlation) and 1 (the strongest possible positive correlation), R = 0 indicates no correlation.

$$RMSD = \sqrt{\sum_{i=1}^{i=N}\frac{\left(NOAA_i - Product_i\right)^2}{N}} \quad (2)$$

$$MAD = \frac{\sum_{i=1}^{i=N}\left|NOAA_i - Product_i\right|}{N} \quad (3)$$

4 Results and discussions

4.1 Comparison of NOAA-CIRES-DOE 20CR and the other products in terms of SR

Figure 1 exhibits the mean monthly SR of the products from 1982 to 2015. As can be seen, ERA-5 and ERA5-Land have closer results to NOAA. However, FLDAS (GLDAS and MERRA-2) overestimate (underestimate) SR compared to NOAA. Mean monthly SR based on NOAA, ERA-5, and ERA5-Land are 196.70, 194.12, and 193.40 W/m², respectively.

Figure 2 illustrates R between NOAA and the other products in terms of SR. It should be noted that only significant
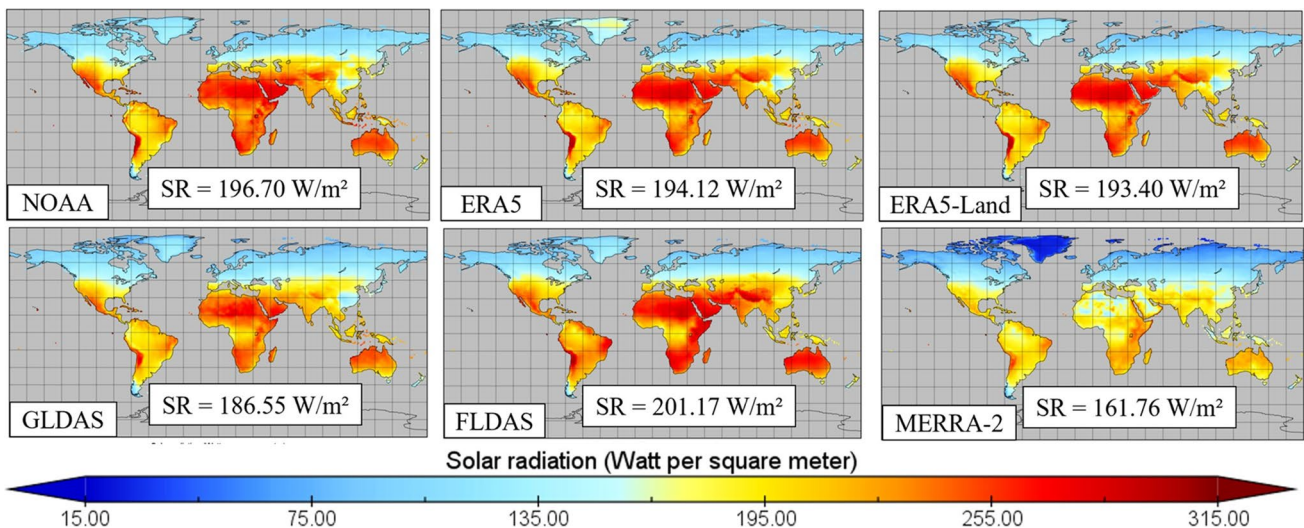

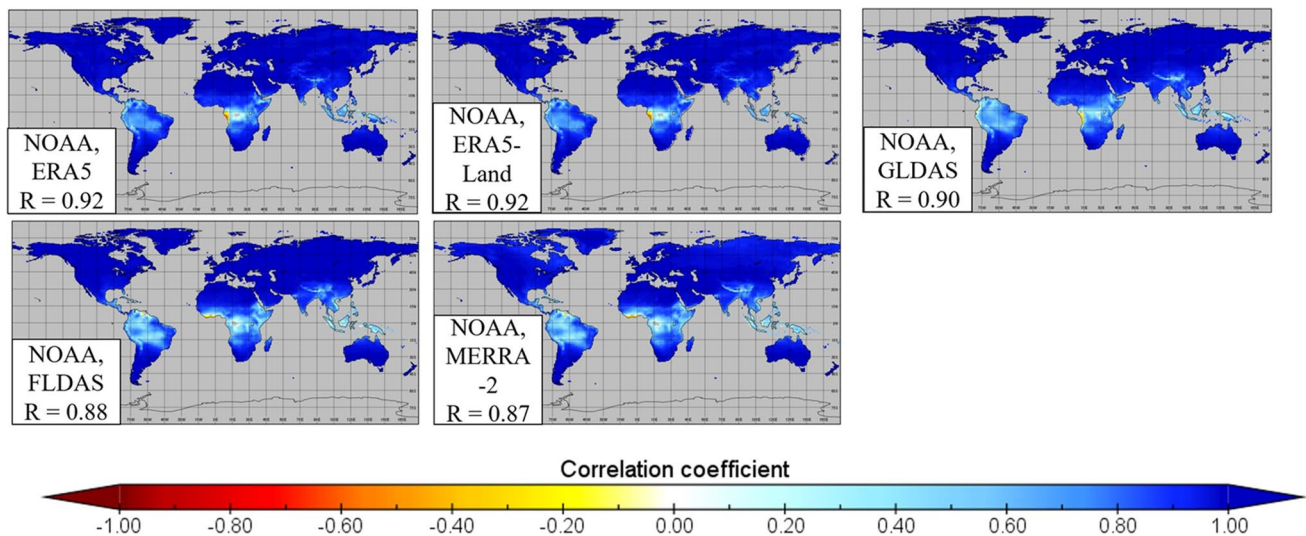**Fig. 1** Mean monthly SR of the products from 1982 to 2015

Fig. 2 *R* between NOAA and the other products in terms of SR. Only significant correlations at the confidence level of 95% have been shown

correlations at the confidence level of 95% have been shown. Based on the obtained results in Fig. 1, as expected, ERA-5 and ERA5-Land (MERRA-2) have the highest (has the lowest) *R* with NOAA data (Fig. 2). The values of *R* are 0.92, 0.92, 0.90, 0.88, and 0.87 between NOAA and ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. Past studies support our findings; Khatibi and Krauter (2021) showed that the *R* between MERRA-2 SR and measured data is 0.97 in Germany which is similar to our results for Germany (see dark blue in Fig. 2). Khalil et al. (2021) appeared that the *R* between ERA-5 SR and measured data in Egypt is 0.96 which is in line with Fig. 2. In all of the maps, the areas between 15 S and 15 N reveal the least *R* compared to the rest of the world.

Figure 3 is plotted to have a better view of the variations of *R* and latitudes in terms of SR. Similar to Fig. 2,

the lowest *R* values are seen nearby the equator in all of the graphs. The minimum *R* (*R*min) values for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2 are 0.57, 0.57, 0.50, 0.47, and 0.48, respectively. ERA-5 and ERA5-Land have the highest *R*min and can be introduced as the best datasets to estimate SR concerning NOAA products.

Previous investigations show that Southeastern Asia, Central Africa, and South America have the highest cloudiness in the globe (ISCCP, 2021; Geerts and Linacre, 2021). These regions are the area where the lowest *R* can be seen based on Fig. 2. Delgado-Bonal et al. (2020) suggest the observed behavior of climatic complexity in terms of MERRA-2 SR could be due to the changes in cloud amount concerning ISCCP information. The impact of cloud coverage on the estimation of SR was also reported
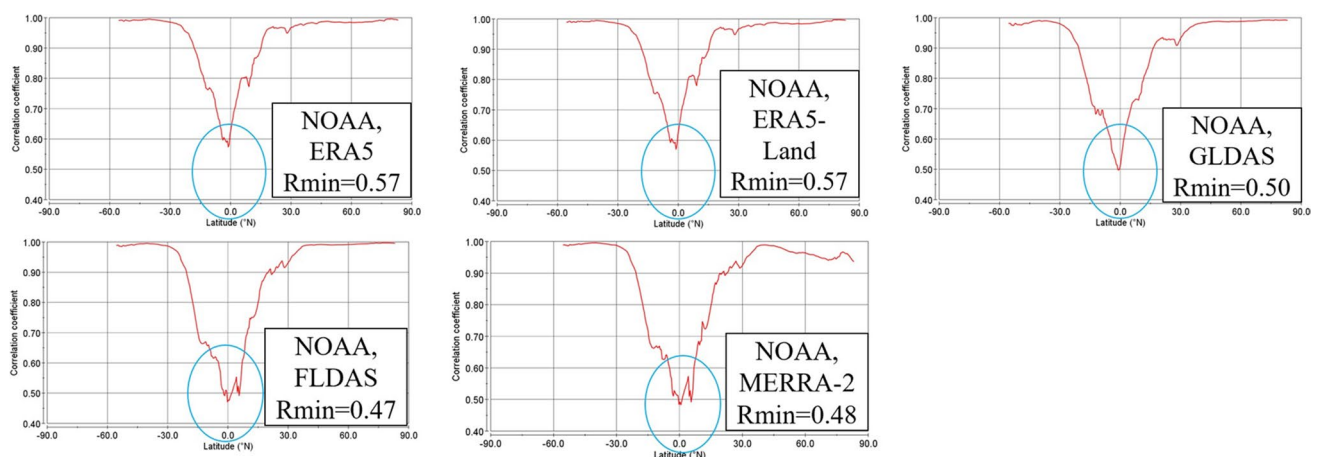

Fig. 3 Variations of the *R* between NOAA SR and the other products with latitudes

by Zhang et al. (2020) and Hinkelman (2019) over China and the globe, respectively.

Estimation of low cloud (stratocumulus) is particularly difficult because it requires many different parametrizations to interact correctly with each other to produce an accurate estimate (Met Office 2017). When the low cloud is accompanied by a high "cirrus" cloud, the satellite only sees the high cloud, making the full extent of the low cloud difficult to determine (Met Office 2017). Klein et al. (2013) showed that as in nature, clouds in climate models strongly affect the radiation balance as a function of space and time. Indeed, the impact of clouds on the top-of-atmosphere radiation budget is too small for passive sensors to detect. Therefore, using

satellites with active sensors instead of passive sensors can be an option to improve the quality of the retrieved images by satellites. Hannak et al. (2017) indicated that many models simulate too low near-surface relative humidity, leading to insufficient low cloud cover and abundant SR.

Figures 4 and 5, respectively, indicate RMSD and MAD between NOAA and the other products in terms of SR. The values of RMSD are 19.14, 19.03, 22.94, 22.18, and 44.88 W/m² for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. According to Fig. 4, ERA5-Land and ERA-5 show the best agreement (the lowest RMSD) with NOAA data while MERRA-2 represents the worst performance, especially in MENA and Greenland. These results
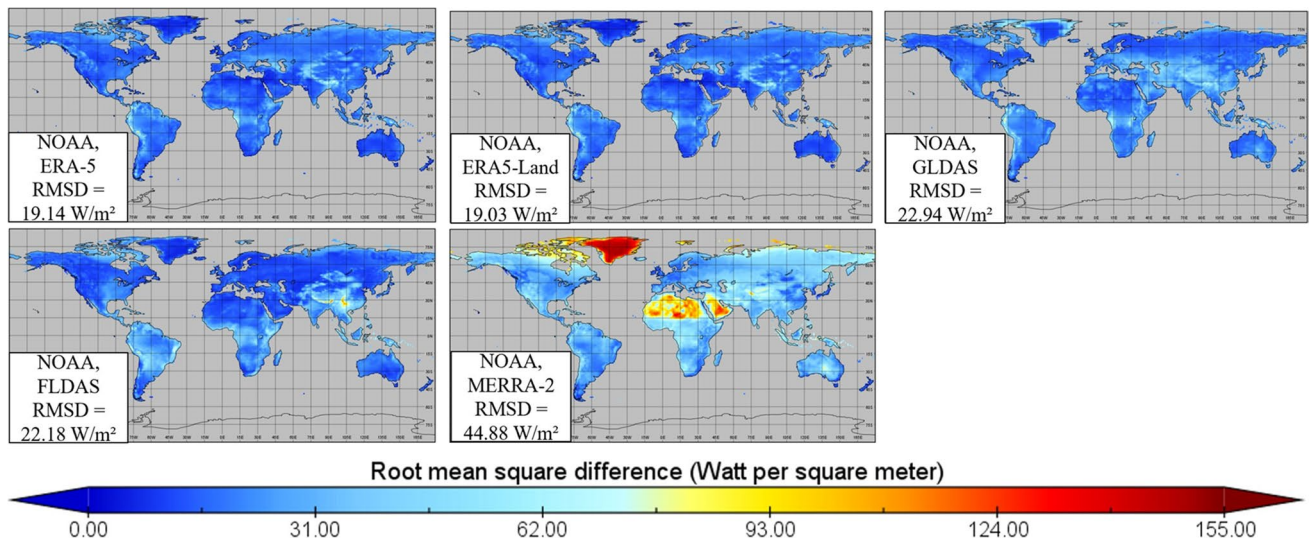

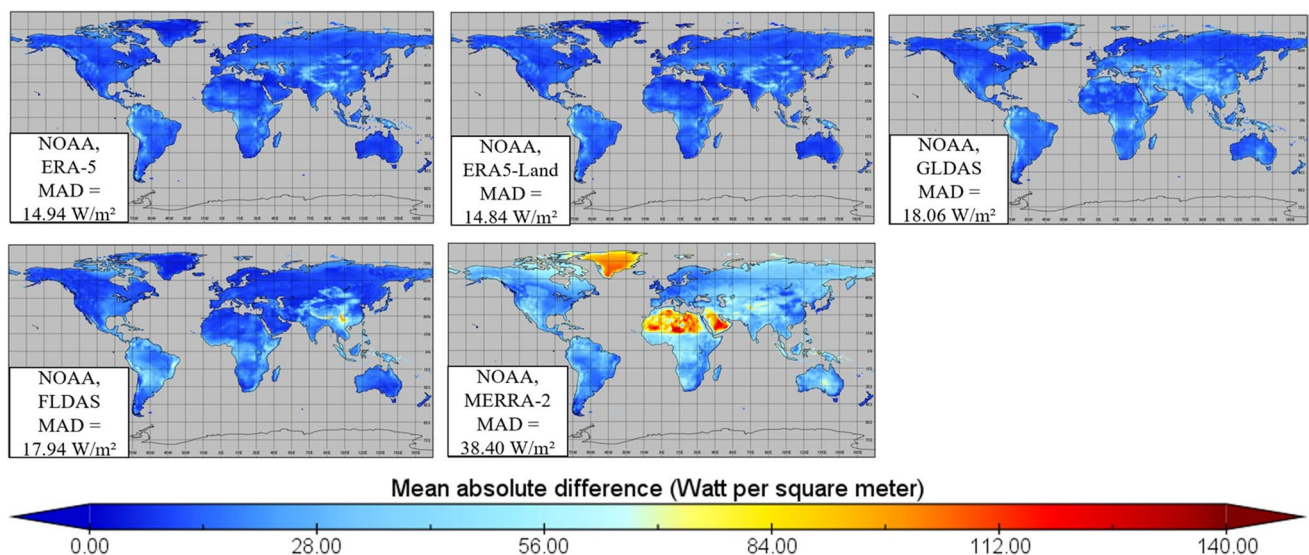**Fig. 4** RMSD between NOAA and the other products in terms of SR

Fig. 5 MAD between NOAA and the other products in terms of SR

are similar to Fig. 5 in terms of MAD. The values of MAD are 14.94, 14.84, 18.06, 17.94, and 38.40 W/m² for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. Similar to Fig. 4, in Fig. 5, ERA5-Land and ERA-5 show the best agreement with NOAA data while, MERRA-2 has the worst estimates, particularly in MENA and Greenland. It is worth mentioning that these regions (i.e., MENA and Greenland) have the highest and lowest values of SR concerning Fig. 1. Indeed, the performance of MERRA-2 can be drastically affected by extreme events (i.e., snow and ice coverage) in terms of SR estimation, which can be seen in both Figs. 4 and 5. Hobeichi et al. (2020) reported the poor performance of remote sensing and reanalysis datasets in MENA. Cloud coverage, aerosol optical depth, and water vapor content are the main factors of errors to estimate SR (see also Zhang et al. 2020). Aerosol optical depth is a quantitative measure of the extinction of SR by aerosol scattering and absorption between the point of observation and the top of the atmosphere. Alghoul et al. (2009) indicated that SR is very influenced by an increase or decrease of aerosol optical depth. Obregón et al. (2020) showed that the spatial distributions of cloud coverage, aerosol optical depth, and water vapor are closely linked to the spatial distributions of their effects on solar radiation at the surface. The highest aerosol optical depth values are located in North Africa, due to the influence of the Sahara Desert. In the case of water vapor, the highest values are obtained over water-covered surfaces since these constitute sources of moisture. Cloud coverage, aerosol optical depth, and water vapor effects are negative, indicating a reduction of SR reaching the surface due to cloud coverage, aerosol optical depth, and water vapor effects. The analysis of the spatial distribution of cloud coverage, aerosol optical depth, and water vapor effects shows that the highest effects

occur over MENA, coinciding with the areas with the greatest influence of aerosols and water vapor when considered individually.

4.2 Comparison of NOAA-CPC and the other products in terms of LSAT

Figure 6 illustrates the mean monthly LSAT of the products from 1982 to 2020. As can be seen, ERA-5, GLDAS, FLDAS, and MERRA-2 show the results close to each other with only a 0–0.03 °C difference. Mean monthly LSAT based on NOAA, ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2 are 13.71, 13.44, 13.15, 13.46, 13.46, and 13.47 °C, respectively.

Figure 7 exhibits R between NOAA and the other products in terms of LSAT. It should be noted that only significant correlations at the confidence level of 95% have been shown. ERA-5 (GLDAS) has the highest (lowest) R with NOAA data. The values of R are 0.94, 0.91, 0.89, 0.92, and 0.92 between NOAA and ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. Khatibi and Krauter (2021) reported that the R between MERRA-2 LSAT and measured data is 0.99 in Germany which is in agreement with our results for Germany (see dark blue in Fig. 7). Similar to Fig. 2, in all of the maps, the areas between 15 S and 15 N reveal the least R compared to the rest of the world.

Figure 8 is plotted to have a better view of the variations of R and latitudes in terms of LSAT. Similar to Fig. 7, the lowest R values are seen nearby the equator in all of the graphs. The Rmin values for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2 are 0.65, 0.53, 0.34, 0.62, and 0.63, respectively. ERA-5 and MERRA-2 have the highest Rmin
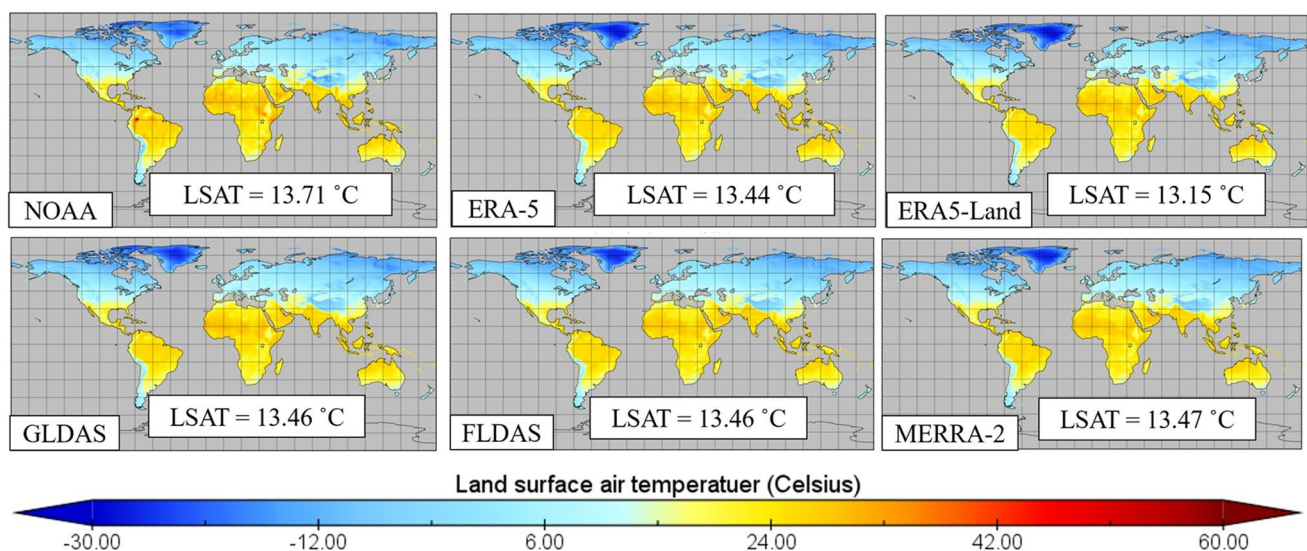

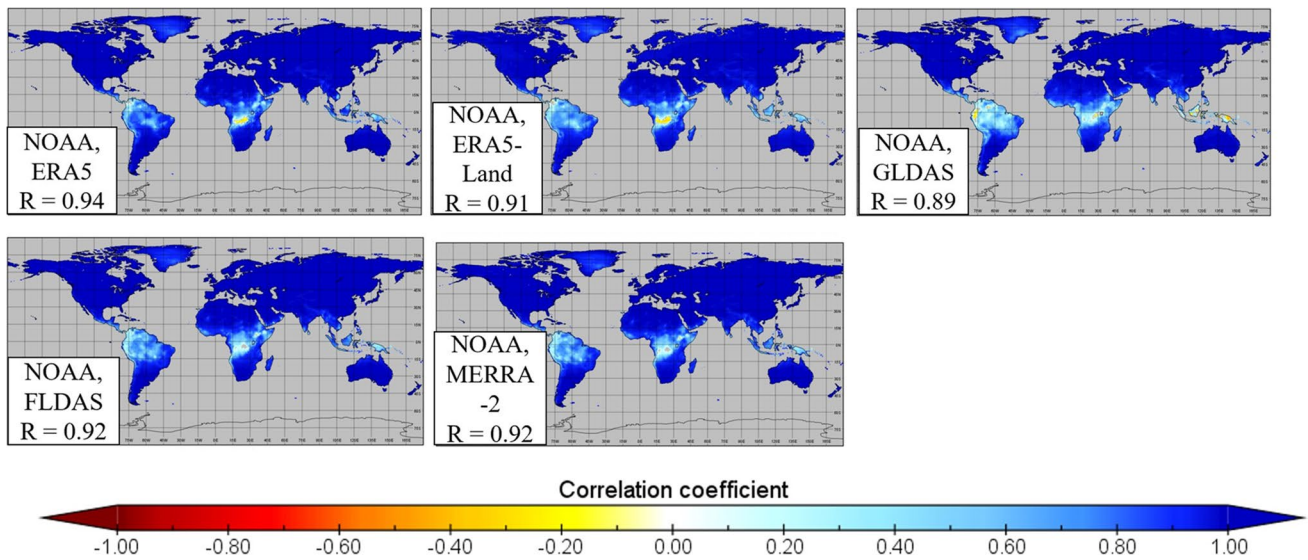**Fig. 6** Mean monthly LSAT of the products from 1982 to 2020

Fig. 7 The correlation coefficient between NOAA and the other products in terms of LSAT. Only significant correlations at the confidence level of 95% have been shown
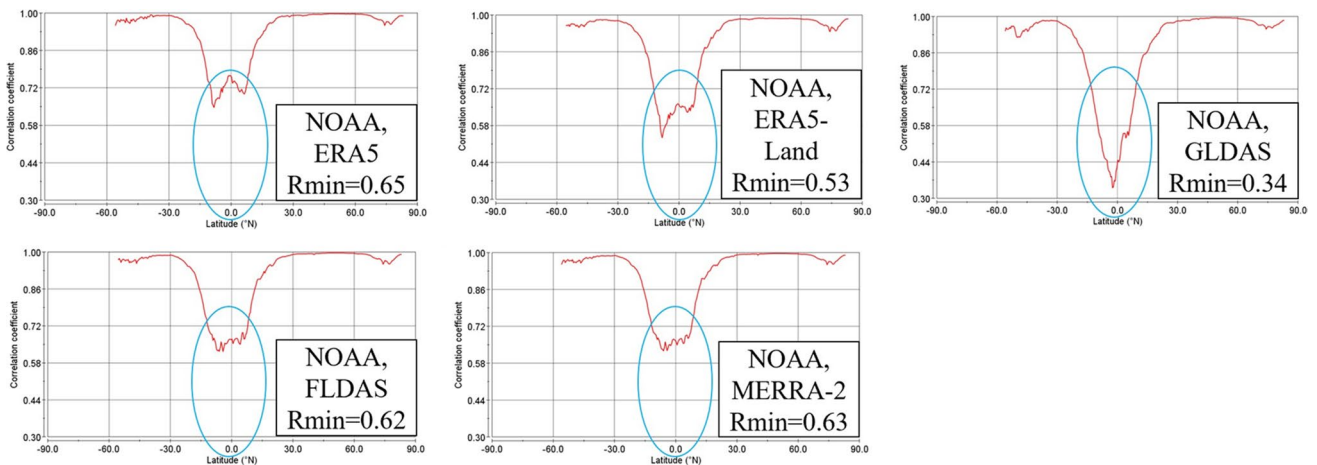

Fig. 8 Variations of the *R* between NOAA LSAT and the other products with latitudes

and can be introduced as the best dataset to estimate SR concerning NOAA products.

Similar to Figs. 2, 3, 4, and 5, Southeastern Asia, Central Africa, and South America show the lowest agreement between NOAA and the other datasets because of the higher cloudiness in these areas compared to the rest of the world (see also Delgado-Bonal et al. 2020; ISCCP, 2021; Geerts and Linacre, 2021; Zhang et al. 2020; and Hinkelman 2019; Met Office 2017; Klein et al. 2013; Hannak et al. 2017).

Figures 9 and 10, respectively, indicate RMSD and MAD between NOAA and the other products in terms of LSAT. The values of RMSD are 1.94, 2.40, 1.98, 2.13, and 1.93 °C for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. MERRA-2 has the best agreement with the

NOAA dataset based on RMSD. The values of MAD for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, are 1.71, 1.98, 1.66, 1.87, and 1.67 °C, respectively. GLDAS shows the best agreement with the NOAA dataset based on MAD. This is consistent with Han et al. (2020) who compared GLDAS LSAT against in situ measurements in China. They showed that the performance of GLDAS can be affected in mountain regions due to having fewer weather sites (see also Figs. 9 and 10). Ji et al. (2015) compared GLDAS LSAT against NOAA over the globe from 2000 to 2011. They showed that the values of RMSD range between 3.76 and 3.93 °C. Compared to their research, our study shows better agreement between GLDAS and NOAA (RMSD = 1.98 °C). The reason could be due to improving
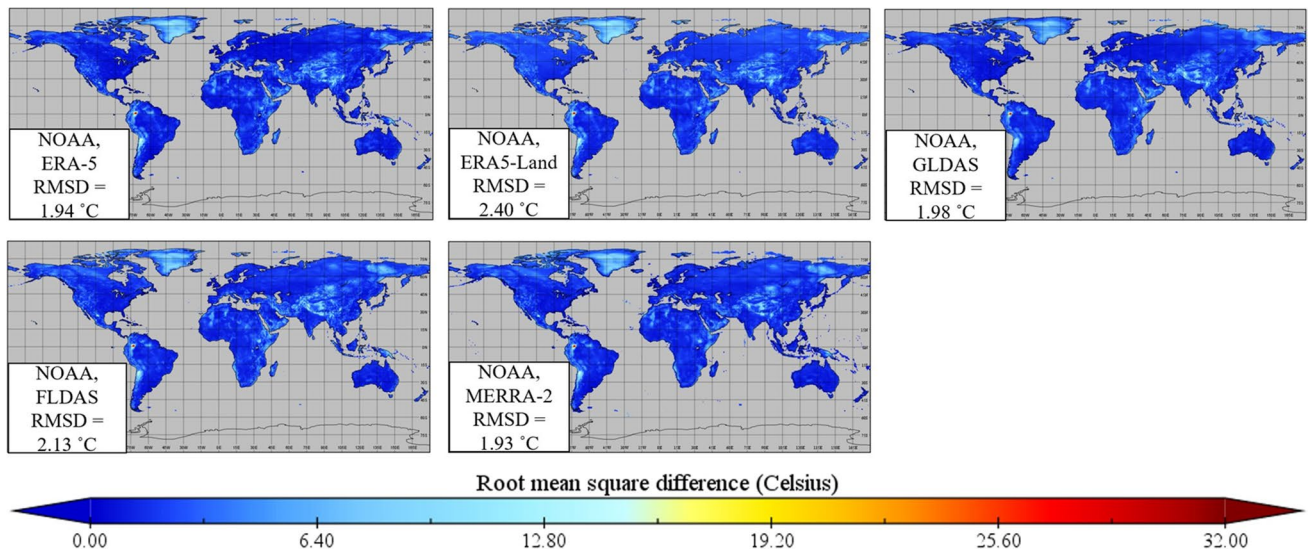
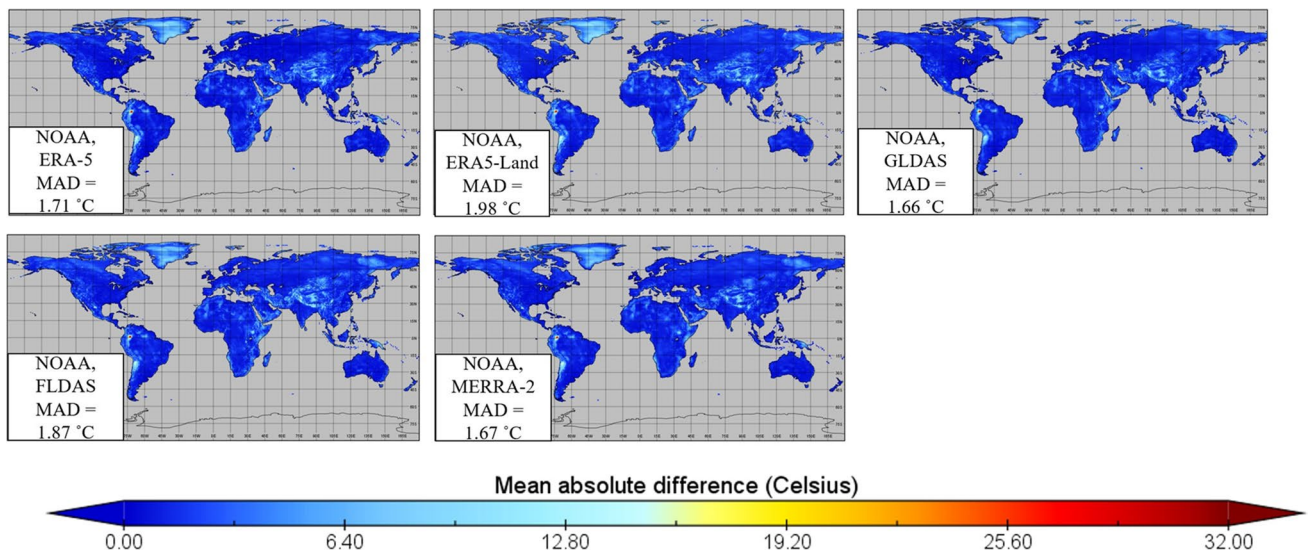**Fig. 9** RMSD between NOAA and the other products in terms of LSAT

Fig. 10 MAD between NOAA and the other products in terms of LSAT

forcing data in the new version of GLDAS which is used in this study (see also Liu et al. 2020; Wu et al. 2021; Qi et al. 2020). Similar to SR, all of the products suffer to capture accurately LSAT dynamics in Greenland since Greenland has the lowest mean monthly LSAT in the globe (see Fig. 6).

4.3 Comparison of NOAA-CPC and the other products in terms of P

Figure 11 shows the mean monthly *P* of the products from 1983 to 2019. As can be seen, GLDAS (MERRA-2) has the closest (farthest) result to NOAA. Mean monthly *P* based on NOAA, ERA-5, ERA5-Land, GLDAS, GPCP,

and MERRA-2 are 58.05, 77.38, 75.97, 70.08, 73.94, and 92.07 mm/month, respectively.

Figure 12 exhibits *R* between NOAA and the other products in terms of P. It should be noted that only significant correlations at the confidence level of 95% have been shown. GPCP (MERRA-2) has the highest (lowest) *R* with NOAA data. The values of *R* are 0.73, 0.74, 0.75, 0.78, and 0.68 between NOAA and ERA-5, ERA5-Land, GLDAS, GPCP, and MERRA-2, respectively. Our results are in line with Hamal et al. (2020) who showed that MERRA-2 can estimate P in Nepal with *R* = 0.63 in comparison with gauge-based data. In all of the maps, the areas between
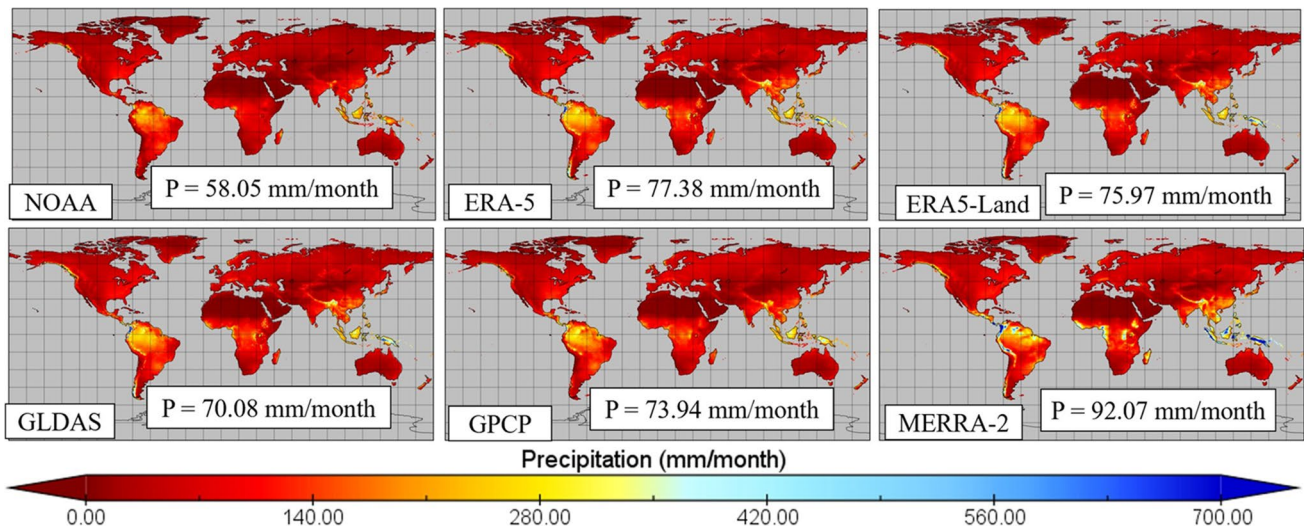
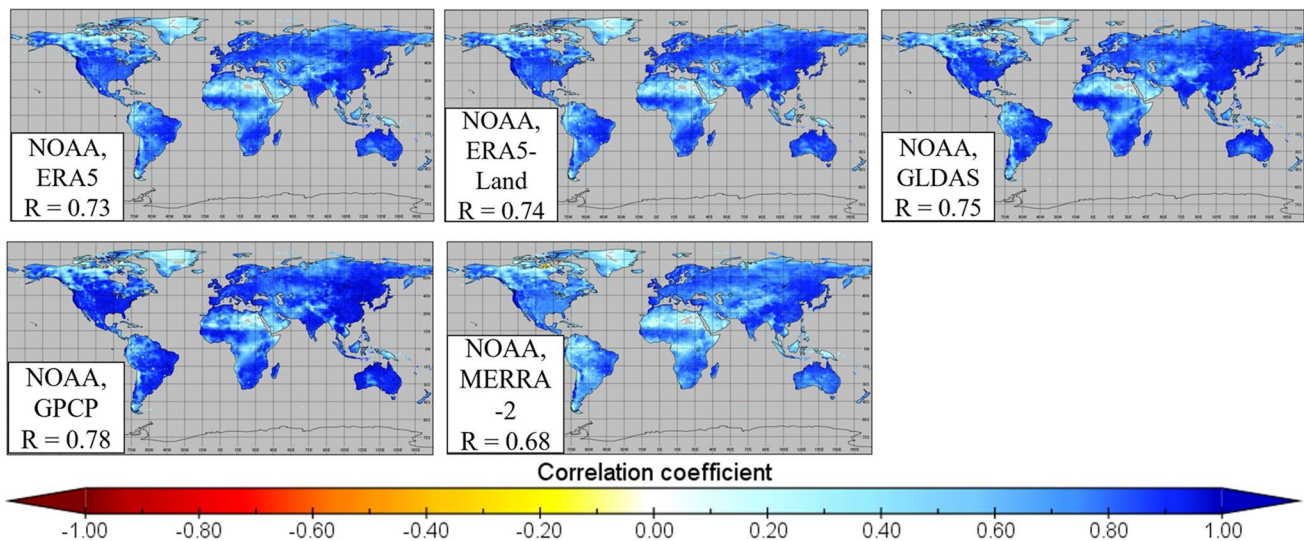**Fig. 11** Mean monthly P of the products from 1983 to 2019

Fig. 12 R between NOAA and the other products in terms of P. Only significant correlations at the confidence level of 95% have been shown

15 S and 15 N, MENA, and Greenland reveal the least *R* compared to the rest of the world.

Figure 13 is plotted to have a better view of the variations of *R* and latitudes in terms of P. The *R*min values for ERA-5, ERA5-Land, GLDAS, GPCP, and MERRA-2 are 0.21, 0.14, 0.34, 0.26, and 0.19, respectively. GLDAS and ERA5-Land have the highest and lowest *R*min, respectively. GLDAS can be introduced as the best dataset to estimate SR concerning NOAA products.

As we discussed, Southeastern Asia, Central Africa, and South America have the highest cloud coverage across the world (ISCCP 2021; Geerts and Linacre, 2021). These areas also have the highest rate of P according to Fig. 11. Jiang

et al. (2021) indicated that ERA-5 has higher RMSD values in the tropical and subtropical regions with a relatively wet climate. Moreover, similar to our results, Tarek et al. (2020) showed that ERA-5 overestimates P in North America, and that might be associated with the quality of the observation datasets in the remote northern catchments. In addition, MENA and Greenland show the lowest rate of P based on Fig. 11. These regions are the areas where the lowest R can be seen based on Fig. 12. Our results are in agreement with Hobeichi et al. (2020) who evaluated MERRA-2 and GPCP to estimate P across the world. They reported better accuracy of GPCP, especially over the tropics. However, both products suffer to capture P dynamics in MENA. Therefore,
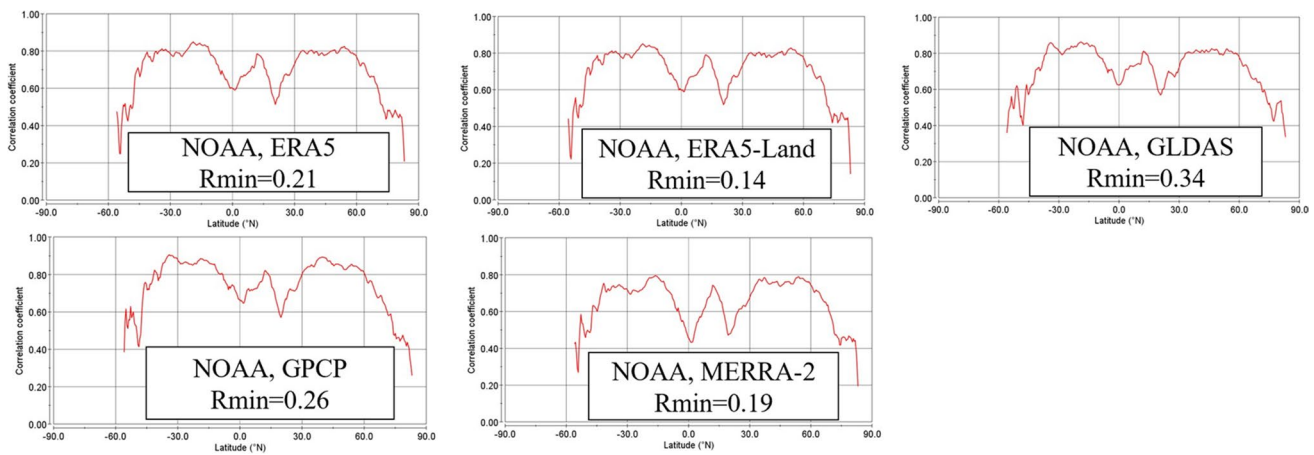
Fig. 13 Variations of the *R* between NOAA P and the other products with latitudes

cloudiness and extreme conditions impact the accuracy of datasets in terms of estimating P.

Figures 14 and 15, respectively, illustrate RMSD and MAD between NOAA and the other products in terms of P. The values of RMSD are 47.05, 46.20, 37.61, 37.92, and 67.55 mm/month for ERA-5, ERA5-Land, GLDAS, GPCP, and MERRA-2, respectively. Similar results can be seen for MAD. The values of MAD are 33.23, 32.62, 25.66, 25.97, and 48.16 mm/month for ERA-5, ERA5-Land, GLDAS, GPCP, and MERRA-2, respectively. GLDAS represents the best agreement with NOAA data while, MERRA-2 indicates the worst performance especially between 15 S and 15 N. Indeed, the performance of MERRA-2 can be drastically affected by cloud coverage in terms of P estimation. The high error indices of MERRA-2 are consistent with Song and Wei (2021). They compared ERA-5 and MERRA-2 in

terms of P estimation in China and reported that the land surface is more strongly coupled with the atmosphere in MERRA-2 dataset compared to ERA-5. Indeed, MERRA-2 P is more sensitive to precipitable water. The low-level wind field is more divergent in MERRA-2 than in ERA-5, which causes weaker ascending motions, less precipitable water, and lower P efficiency. The less favorable atmospheric conditions for P, exacerbated by the strong land–atmosphere coupling, lead to occurring errors MERRA-2 P in terms of P estimates. Yao et al. (2020) evaluated NOAA, GPCP, ERA-Interim, and MERRA-2 to estimate P over China. Although all the products reasonably reproduce patterns of P in the study area, there are differences among estimated patterns of P by NOAA, GPCP, ERA-Interim, and MERRA-2. NOAA, GPCP, and MERRA-2 that show more accurate results, respectively, which is consistent with our findings.
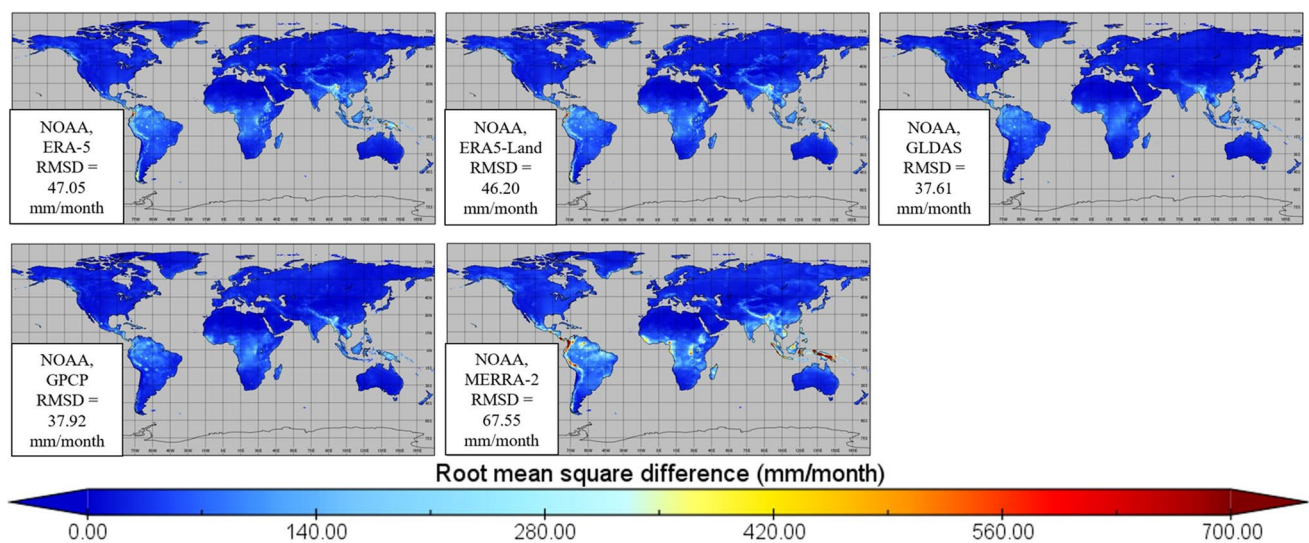

Fig. 14 RMSD between NOAA and the other products in terms of P
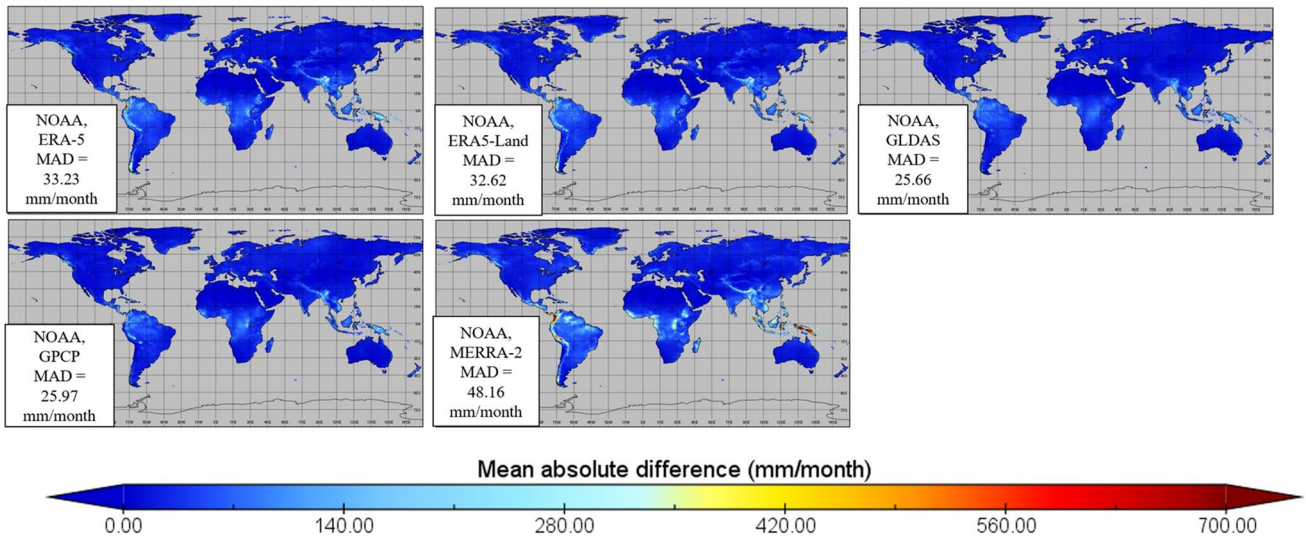
Fig. 15 MAD between NOAA and the other products in terms of P

It should be taken into account that we suppose NOAA datasets and especially their in situ measurements are perfect. However, in practice, at the catchment scale, one would expect that the measurements would be far from perfect and involve errors because of location representativeness, P under the catch, and missing data because of site malfunction and/or instrument replacement (see also Tarek et al. 2020). Therefore, we should consider these factors as sources of uncertainties for this study.

It is worth mentioning that we could consider GPCP as the reference dataset like some past studies. However, since GPCP is a remote sensing-based product and NOAA-CPC is a gauge-based dataset, we decided to consider NOAA-CPC as the reference. Furthermore, since we use NOAA as the reference for SR and LSAT, it would be better to have the same product (i.e., NOAA) as the base dataset to be consistent with our other comparisons.

4.4 Ensemble mean dataset

The results reveal that the employed datasets have different performances to estimate SR, LSAT, and P. All the products have some advantages and disadvantages. Since there are uncertainties in all of the products, developing new datasets based on merging the best products concerning their accuracy, may be useful. In addition, different datasets employed in this study, use various sets of forcing data, and none of them has superiority over the others in all terms. Therefore, the goal is to build ensemble mean models in terms of SR, LSAT, and P, in which the most accurate datasets are combined. This can improve the reliability aspect of the obtained results. Table 2 shows datasets used in the ensemble mean models based on their performance in the estimation of

Table 2 Datasets used in the ensemble mean models based on their performance in estimation of LSAT, SR, and P in comparison with NOAA

Dataset	Ensemble mean LSAT	Ensemble mean SR	Ensemble mean P
ERA-5	✓	✓	✓
ERA5-Land	✓	✓	✓
GLDAS	✓	✓	✓
FLDAS	✓	✓	✗
MERRA-2	✓	✗	✗
GPCP	✗	✗	✓

LSAT, SR, and P in comparison with NOAA which was discussed in the previous sections (i.e., Sects. 4.1, 4.2, and 4.3). The ensemble mean models are the mean of all used datasets in Table 2. According to Table 2, ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2 are merged to build an ensemble mean LSAT dataset. Also, ERA-5, ERA5-Land, GLDAS, and FLDAS have selected datasets to build an ensemble mean SR product. Finally, ERA-5, ERA5-Land, GLDAS, and GPCP are considered for an ensemble mean model in terms of P.

Figure 16 exhibits the monthly average of the ensemble mean models and their comparisons with NOAA in terms of SR (above panel), LSAT (middle panel), and P (bottom panel). As can be seen, in terms of SR, the ensemble mean model has a mean monthly equal to 194.04 W/m². This value is close to NOAA and ERA-5 based on Fig. 1. The difference between NOAA and the ensemble mean dataset is 2.40 W/m² at the global scale. In terms of LSAT, the mean monthly ensemble mean model is 13.38 °C. This value is more close
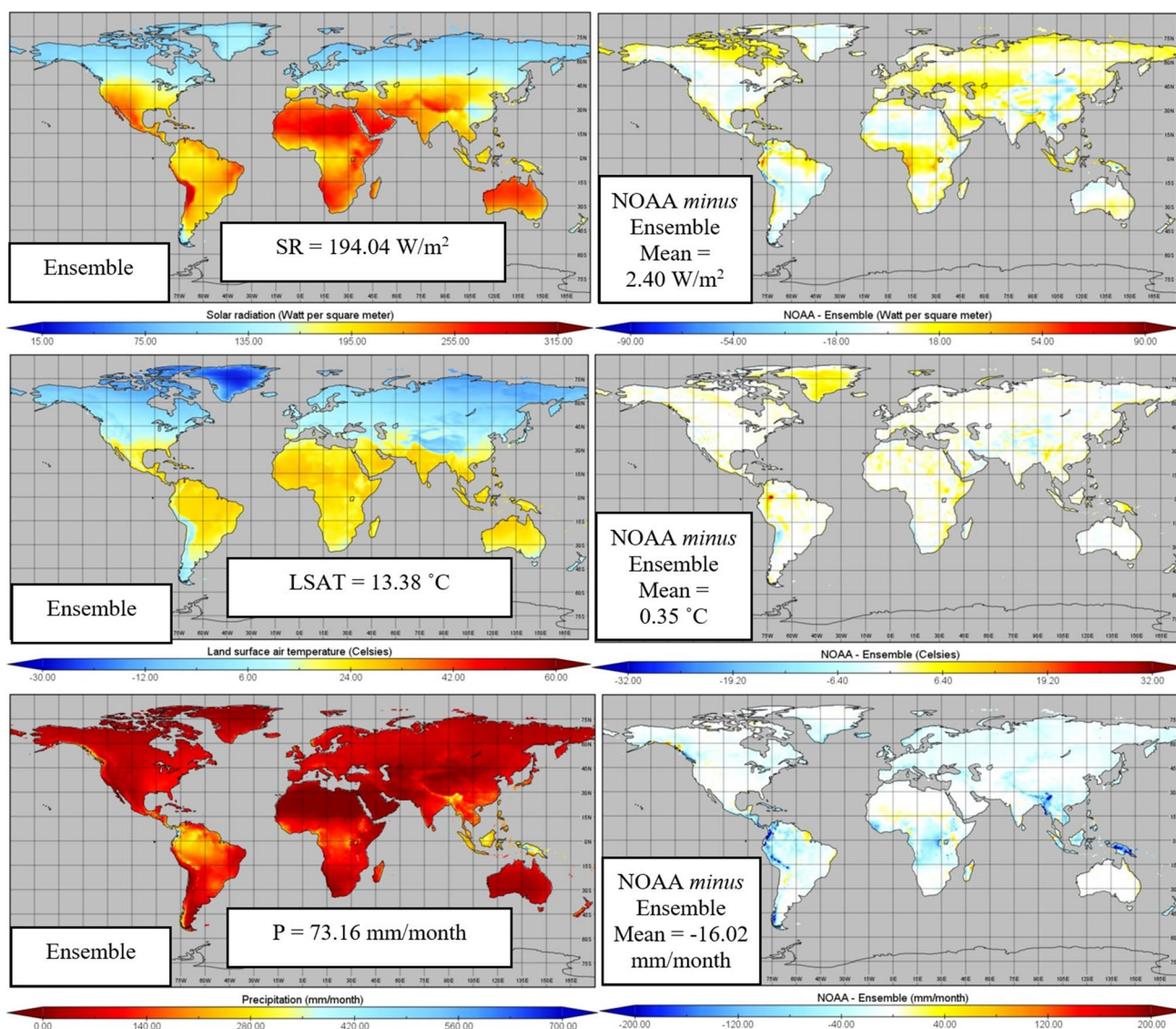
Fig. 16 Monthly average of the ensemble mean models and their comparisons with NOAA in terms of SR (above panel), LSAT (middle panel), and P (bottom panel)

to GLDAS and FLDAS concerning Fig. 6. The difference between NOAA and the ensemble mean dataset is only 0.35 °C at the global scale which indicates the good performance of the ensemble mean model. Regarding P, the mean monthly of the ensemble mean model is 73.16 mm/month. This value is more close to GPCP according to Fig. 11. The difference between NOAA and the ensemble mean dataset is − 16.02 mm/month at the global scale.

Other investigations also indicate the advantages of the ensemble mean models. Wang et al. (2020) successfully merged gauge, climate reanalysis, and satellite P datasets (including MERRA-2 and GLDAS) in China. In the other study in China, Yao et al. (2020) evaluated NOAA, GPCP, ERA-Interim, and MERRA-2 to estimate P. According to their results, all the products reasonably reproduce patterns

of P; however, there are differences among estimated patterns of P by NOAA, GPCP, ERA-Interim, and MERRA-2. They mentioned that a systematic assessment of the differences between multiple products is critical to reduce discrepancies. Sun and Fu (2021) merged the Tropical Rainfall Measuring Mission (TRMM) and ERA-5 P to build a robust dataset at the global scale. The results appear that the accuracy of the combined dataset is reasonable.

In general, the ensemble mean P overestimates NOAA which is in line with our findings in Fig. 11, while ensemble mean SR and LSAT slightly underestimate NOAA. Although there are many regions with high accuracy in the ensemble mean models (see the regions with white color), there are still some areas in which the ensemble mean models suffer in estimating SR, LSAT, and P accurately. These areas are

Southeastern Asia, South America, Central Africa, MENA, and Greenland due to their specific hydrological conditions as we discussed. Hence, the next investigations should focus on those regions to insure how we can improve the quality of remote sensing, reanalysis, and land surface models to capture SR, LSAT, and P dynamics on regional and global scales. According to Fig. 16, the results of the ensemble models are consistent. For example, the highest amount of SR is observed in MENA where the lowest amount of P can be seen.

Although the ensemble mean construction may include bias correction terms, further postprocessing to fulfill the requirements of bias corrections to represent the water balance or to have more precise plant growth conditions is needed in future studies. In addition, other databases such as the Land Use Model Intercomparison Project (LUMIP) and the Global Energy Balance Archive (GEBA) may be considered as a baseline instead of the NOAA dataset (Lawrence et al. 2016; Alexander et al. 2020; Chakraborty and Lee 2021; Wild et al. 2017). LUMIP has a resolution of $0.25° \times 0.25°$ (Lawrence et al. 2016). GEBA has continuously been expanded and updated and contains in its 2017 version around 500,000 monthly mean entries of various surface energy balance components measured at 2500 locations (Wild et al. 2017).

5 Summary and conclusions

This study compares six datasets including ERA-5, ERA5-Land, GLDAS, FLDAS, MERRA-2, and GPCP against NOAA products in terms of SR, LSAT, and P. These datasets are selected based on their successful performances in previous studies. Three statistical metrics (i.e., R, RMSD, and MAD) are used to check the difference between NOAA, as a reference product, and the other datasets. Based on the obtained results:

- Mean monthly SR based on NOAA, ERA-5, and ERA5-Land are 196.70, 194.12, and 193.40 W/m^2, respectively. In addition, the values of R between NOAA SR and the other products are 0.92, 0.92, 0.90, 0.88, and 0.87 for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. Also, the values of RMSD (MAD) are 19.14, 19.03, 22.94, 22.18, and 44.88 (14.94, 14.84, 18.06, 17.94, and 38.40) W/m^2 for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. ERA-5 and ERA5-Land show the best agreement with NOAA data while MERRA-2 represents the worst performance.
- ERA-5, GLDAS, FLDAS, and MERRA-2 LSAT show the results close to each other with only a 0–0.03 °C difference. Mean monthly LSAT based on NOAA, ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2

are 13.71, 13.44, 13.15, 13.46, 13.46, and 13.47 °C, respectively. The values of R are 0.94, 0.91, 0.89, 0.92, and 0.92 between NOAA and ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively, in terms of LSAT. In addition, the values of RMSD (MAD) are 1.94, 2.40, 1.98, 2.13, and 1.93 (1.71, 1.98, 1.66, 1.87, and 1.67) °C for ERA-5, ERA5-Land, GLDAS, FLDAS, and MERRA-2, respectively. MERRA-2 and GLDAS have the best agreement with the NOAA dataset based on RMSD and MAD, respectively.

- Mean monthly P based on NOAA, ERA-5, ERA5-Land, GLDAS, GPCP, and MERRA-2 are 58.05, 77.38, 75.97, 70.08, 73.94, and 92.07 mm/month, respectively. Meanwhile, GPCP (MERRA-2) has the highest (lowest) R with NOAA data. The values of R are 0.73, 0.74, 0.75, 0.78, and 0.68 between NOAA and ERA-5, ERA5-Land, GLDAS, GPCP, and MERRA-2, respectively, in terms of P. The values of RMSD (MAD) are 47.05, 46.20, 37.61, 37.92, and 67.55 (33.23, 32.62, 25.66, 25.97, and 48.16) mm/month for ERA-5, ERA5-Land, GLDAS, GPCP, and MERRA-2, respectively. GLDAS represents the best agreement with NOAA data while MERRA-2 indicates the worst performance. The performance of MERRA-2 can be drastically affected by cloud coverage in terms of SR and P estimation.

It is worth mentioning that, at the catchment scale, one would expect that the measurements would be far from perfect and involve errors because of location representativeness, P under the catch, and missing data because of site malfunction and/or instrument replacement. Therefore, we should consider these factors as sources of uncertainties. Quantifying these uncertainties is beyond the subject of this study. However, one should be aware that all datasets derived from ground data are affected by various uncertainties associated with the local measurement and areal estimation of climatic variables.

Since there are uncertainties in all of the products, developing new datasets based on merging the best products concerning their accuracy, may be useful. In addition, different datasets employed in this study, use various sets of forcing data, and none of them has superiority over the others in all terms. Therefore, the goal is to build ensemble mean models in terms of SR, LSAT, and P, in which the most accurate datasets are combined. This can improve the reliability aspect of the obtained results. To this end, three ensemble mean datasets are suggested as the average of the best products in terms of SR, LSAT, and P. According to the obtained results of the ensemble mean models:

- The difference between NOAA and the ensemble mean datasets are 2.40 W/m^2, 0.35 °C, and -16.02 mm/month

at the global scale for the ensemble mean SR, LSAT, and P, respectively.

- There are some regions in which the datasets cannot estimate SR, LSAT, and P accurately. Those areas are Southeastern Asia, South America, Central Africa, MENA, and Greenland because of their specific hydrological conditions such as the occurrence of extreme events and high rate of cloudiness.

- Therefore, there are some open avenues for the next investigations on those areas to insure how we can improve the quality of remote sensing, reanalysis, and land surface datasets to capture SR, LSAT, and P dynamics on regional and global scales, particularly concerning climate change and variability (Valipour 2017; Almazroui et al. 2021).

- It is notable that, different products use different spatial and temporal resolutions as well as satellite types which can impact on the results.

- Moreover, if we want to expand the results for regional and local cases, we will need finer spatial resolution which could be the next generation of satellites.

- In this study, the goal was to show that ensemble models are more reliable than estimating SR, LAST, and R by each model individually. However, it depends on our decision to pick the best models. For example, we can only pick one climate reanalysis alongside only one of the land surface models and develop the ensemble model accordingly. Even in that case, the performance of the ensemble model would be more reliable instead of recommending one certain model.

- Since some of the products do not support water surfaces, in this study, we have focused on grid points on the land surfaces. However, in the future and developing coverage of the products, it would be useful to add water grid points and compare the ability of products in the estimation of LSAT, SR, and P in water grid points.

- The ensemble models are useful for any future development of climate reanalysis, land surface models, and remote sensing retrievals especially in terms of estimating LAST, SR, and P.

Declarations

References

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie PP, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A (2003) The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). J Hydrometeorol 4(6):1147–1167

Adler RF, Gu G, Sapiano M, Wang JJ, Huffman GJ (2017) Global precipitation: means, variations and trends during the satellite era (1979–2014). Surv Geophys 38(4):679–699

Adler RF, Sapiano MR, Huffman GJ, Wang JJ, Gu G, Bolvin D, Chiu L, Schneider U, Becker A, Nelkin E, Xie P (2018) The Global Precipitation Climatology Project (GPCP) monthly analysis and a review of 2017 global precipitation. Atmosphere 9(4):138

Albergel C, Dutra E, Munier S, Calvet JC, Munoz-Sabater J, de Rosnay P and Balsamo G, (2018). ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? Hydrol Earth Sys Sci 22(6).

Alexander LV, Bador M, Roca R, Contractor S, Donat MG, Nguyen PL (2020) Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products. Environ Res Lett 15(5):055002

Alghoul MA, Khamies H, Sulaiman MY, Assadeq J, Yahya M, Alfegi M, Zaharim A, Sopian K (2009) Impact of aerosol optical depth on solar radiation budget. In Proc. 3rd WSEAS Int. Conf Energy Plan Energy Saving, Environ Educ EPESE 9:386–393

Almazroui M, Saeed F, Saeed S, Ismail M, Ehsan MA, Islam MN, Abid MA, O'Brien E, Kamil S, Rashid IU, Nadeem I (2021) Projected changes in climate extremes using CMIP6 simulations over SREX regions. Earth Sys Environ 5(3):481–497

Babar B, Graversen R, Boström T (2019) Solar radiation estimation at high latitudes: assessment of the CMSAF databases, ASR and ERA5. Sol Energy 182:397–411

Balsamo G, Viterbo P, Beljaars A, van den Hurk B, Hirschi M, Betts AK, Scipal K (2009) A revised hydrology for the ECMWF model. J Hydro Met. https://doi.org/10.1175/2008JHM1068.1

Chakraborty T, Lee X (2021) Large differences in diffuse solar radiation among current-generation reanalysis and satellite-derived products. J Clim 34(16):6635–6650

Chen M, Shi W, Xie P, Silva VB, Kousky VE, Wayne Higgins R, Janowiak JE (2008) Assessing objective techniques for gauge-based analyses of global daily precipitation. J Geophys Res: Atmos 113:D4

Chen Y, Yang K, Qin J, Zhao L, Tang W, Han M (2013) Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau. J Geophys Res 118(10):4466–4475

Chen N, Li R, Zhang X, Yang C, Wang X, Zeng L, Tang S, Wang W, Li D, Niyogi D (2020) Drought propagation in Northern China Plain: a comparative analysis of GLDAS and MERRA-2 datasets. J Hydrol 588:125026

Chen Z, Zeng Y, Shen G, Xiao C, Xu L, Chen N (2021) Spatiotemporal characteristics and estimates of extreme precipitation in the Yangtze River Basin using GLDAS data. Int J Climatol 41:E1812–E1830

Delgado-Bonal A, Marshak A, Yang Y, Holdaway D (2020) Analyzing changes in the complexity of climate in the last four decades using MERRA-2 radiation data. Sci Rep 10(1):1–8

European Centre for Medium-Range Weather Forecasts (2021) ECMWF ERA5.1: surface level analysis parameter data for 2000-2006. Centre for Environmental Data Analysis, date of citation. https://catalogue.ceda.ac.uk/uuid/8aa70a91378d455ea63a2a195 3858a7f

Fallahi S, Amanollahi J, Tzanis CG, Ramli MF (2018) Estimating solar radiation using NOAA/AVHRR and ground measurement data. Atmos Res 199:93–102

Foster G, Rahmstorf S (2011) Global temperature evolution 1979–2010. Environ Res Lett 6(4):044022

Geerts B, Linacre E (2021) Variation of cloudiness with latitude. http://www-das.uwyo.edu/~geerts/cwx/notes/chap08/cloud_lat.html#:~:text=On%20average%2C%20cloudiness%20is%20highest,at%20around%2060%EF%BF%BDS. Accessed 5/5/2021

Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, Randles CA, Darmenov A, Bosilovich MG, Reichle R, Wargan K (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim 30(14):5419–5454

Gleixner S, Demissie T, Diro GT (2020) Did ERA5 improve temperature and precipitation reanalysis over East Africa? Atmosphere 11(9):996

Hamal K, Sharma S, Khadka N, Baniya B, Ali M, Shrestha MS, Xu T, Shrestha D, Dawadi B (2020) Evaluation of MERRA-2 precipitation products using gauge observation in Nepal. Hydrology 7(3):40

Han S, Liu B, Shi C, Liu Y, Qiu M, Sun S (2020) Evaluation of CLDAS and GLDAS datasets for Near-surface Air Temperature over major land areas of China. Sustainability 12(10):4311

Hannak L, Knippertz P, Fink AH, Kniffka A, Pante G (2017) Why do global climate models struggle to represent low-level clouds in the West African summer monsoon? J Clim 30(5):1665–1687

He Y, Wang K, Feng F (2021) Improvement of ERA5 over ERA-Interim in simulating surface incident solar radiation throughout China. J Clim 34(10):3853–3867

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049

Hinkelman LM (2019) The global radiative energy budget in MERRA and MERRA-2: evaluation with respect to CERES EBAF data. J Clim 32(6):1973–1994

Hobeichi S, Abramowitz G, Contractor S, Evans J (2020) Evaluating precipitation datasets using surface water and energy budget closure. J Hydrometeorol 21(5):989–1009

Huffman GJ, Adler RF, Arkin P, Chang A, Ferraro R, Gruber A, Janowiak J, McNab A, Rudolf B, Schneider U (1997) The global precipitation climatology project (GPCP) combined precipitation dataset. Bull Am Meteor Soc 78(1):5–20

Huffman GJ, Adler RF, Bolvin DT, Gu G (2009) Improving the global precipitation record: version 2.1. Geophys Res Lett 36:17

ISCCP: International Satellite Cloud Climatology Project (2021). https://isccp.giss.nasa.gov/products/onlineData.html. Accessed 5/5/2021

Ji L, Senay GB, Verdin JP (2015) Evaluation of the Global Land Data Assimilation System (GLDAS) air temperature data products. J Hydrometeorol 16(6):2463–2480

Jiang H, Yang Y, Bai Y, Wang H (2019) Evaluation of the total, direct, and diffuse solar radiations from the ERA5 reanalysis data in China. IEEE Geosci Remote Sens Lett 17(1):47–51

Jiang Q, Li W, Fan Z, He X, Sun W, Chen S, Wen J, Gao J, Wang J (2021) Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. J Hydrol 595:125660

Khalil SA, Khamees AS, Morsy M, Hassan AH, Rahoma UA, Sayad T (2021) Evaluation of global solar radiation estimated from ECMWF-ERA5 and validation with measured data over Egypt. Turkish J Comput Math Educ (TURCOMAT) 12(6):3996–4012

Khatibi A, Krauter S (2021) Validation and performance of satellite meteorological dataset MERRA-2 for solar and wind applications. Energies 14(4):882

Klein SA, Zhang Y, Zelinka MD, Pincus R, Boyle J, Gleckler PJ (2013) Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator. J Geophys Res: Atmos 118(3):1329–1342

Lawrence DM, Hurtt GC, Arneth A, Brovkin V, Calvin KV, Jones AD, Jones CD, Lawrence PJ, de Noblet-Ducoudré N, Pongratz J, Seneviratne SI (2016) The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design. Geosci Model Dev 9(9):2973–2998

Li C, Zhao T, Shi C, Liu Z (2021) Assessment of precipitation from the CRA40 dataset and new generation reanalysis datasets in the global domain. Int J Climatol

Liu L, Gu H, Xie J, Xu YP (2021) How well do the ERA-Interim, ERA-5 GLDAS-2.1 and NCEP-R2 reanalysis datasets represent daily air temperature over the Tibetan Plateau? Int J of Climatol 41(2):1484–1505

Ma J, Zhou J, Göttsche FM, Liang S, Wang S, Li M (2020) A global long-term (1981–2000) land surface temperature product for NOAA AVHRR. Earth Sys Sci Data 12(4):3247–3268

McNally A, Arsenault K, Kumar S, Shukla S, Peterson P, Wang S, Funk C, Peters-Lidard CD, Verdin JP (2017) A land data assimilation system for sub-Saharan Africa food and water security applications. Sci Data 4(1):1–19

Met Office (2017) Why is forecasting low cloud so difficult? https://blog.metoffice.gov.uk/2017/04/07/why-is-forecasting-low-cloud-so-difficult/#:~:text=As%20with%20most%20weather%20forecasting,world%20as%20a%20pixellated%20image. Accessed 5/7/2021

Mokhtari A, Noory H, Vazifedoust M (2018) Performance of different surface incoming solar radiation models and their impacts on reference evapotranspiration. Water Resour Manage 32(9):3053–3070

Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B (2021) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Sys Sci Data 13:4349–4383

Nogueira M (2020) Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last 40 years: process-based analysis of systematic and random differences. J Hydrol 583:124632

Obregón MA, Costa MJ, Silva AM, Serrano A (2020) Spatial and temporal variation of aerosol and water vapour effects on solar radiation in the Mediterranean basin during the last two decades. Remote Sensing 12(8):1316

Qi W, Liu J, Yang H, Zhu X, Tian Y, Jiang X, Huang X, Feng L (2020) Large uncertainties in runoff estimations of GLDAS Versions and in China. Earth and Space Science 7(1), p.e2019EA000829

Reichle RH, Liu Q, Koster RD, Draper CS, Mahanama SP, Partyka GS (2017) Land surface precipitation in MERRA-2. J Clim 30(5):1643–1664

Rodell M, Houser PR, Jambor UEA, Gottschalck J, Mitchell K, Meng CJ, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK (2004) The global land data assimilation system. Bull Am Meteor Soc 85(3):381–394

Sengupta M, Xie Y, Lopez A, Habte A, Maclaurin G, Shelby J (2018) The national solar radiation data base (NSRDB). Renew Sustain Energy Rev 89:51–60

Slivinski LC, Compo GP, Whitaker JS, Sardeshmukh PD, Giese BS, McColl C, Allan R, Yin X, Vose R, Titchner H, Kennedy J (2019) Towards a more reliable historical reanalysis: improvements for version 3 of the Twentieth Century Reanalysis system. Q J R Meteorol Soc 145(724):2876–2908

Smith TM, Reynolds RW (2005) A global merged land air and sea surface temperature reconstruction based on historical observations (1880–1997). J Clim 18:2021–2036

Smith T, Yin X, Gruber A (2006) Variations in annual global precipitation (1979–2004) based on the Global Precipitation Climatology Project analysis. Geophysi Res Lett 33(6)

Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). J Clim 21:2283–2293

Song Y, Wei J (2021) Diurnal cycle of summer precipitation over the North China Plain and associated land-atmosphere interactions: evaluation of ERA5 and MERRA-2. Int J Climatol

Sun L, Yunfei Fu (2021) A new merged dataset for analyzing clouds, precipitation and atmospheric parameters based on ERA5 reanalysis data and the measurements of the Tropical Rainfall Measuring Mission (TRMM) precipitation radar and visible and infrared scanner. Earth Sys Sci Data 13(5):2293–2306

Tarek M, Brissette FP, Arsenault R (2020) Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. Hydrol Earth Syst Sci 24(5):2527–2544

Valipour M (2017) Calibration of mass transfer-based models to predict reference crop evapotranspiration. Applied Water Sci 7(2):625–635

Wanders N, Bierkens MFP, Jong SM, Roo A, Karssenberg D (2014) The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. Water Resour Res 50:6874–6891

Wang W, Cui W, Wang X, Chen X (2016) Evaluation of GLDAS-1 and GLDAS-2 forcing data and Noah model simulations over China at the monthly scale. J Hydrometeorol 17(11):2815–2833

Wang Y, Wang L, Li X, Zhou J, Hu Z (2020) An integration of gauge, satellite, and reanalysis precipitation datasets for the largest river basin of the Tibetan Plateau. Earth Sys Sci Data 12(3):1789–1803

Wild M, Ohmura A, Schär C, Müller G, Folini D, Schwarz M, Hakuba MZ, Sanchez-Lorenzo A (2017) The Global Energy Balance Archive (GEBA) version 2017: a database for worldwide measured surface energy fluxes. Earth Sys Sci Data 9(2):601–613

Wu Z, Feng H, He H, Zhou J, Zhang Y (2021) Evaluation of soil moisture climatology and anomaly components derived from ERA5-Land and GLDAS-in China. Water Resour Manag 35(2):629–643

Xie P, Chen M, Yang S, Yatagai A, Hayasaka T, Fukushima Y, Liu C (2007) A gauge-based analysis of daily precipitation over East Asia. J Hydrometeorol 8(3):607–626

Yao J, Chen Y, Yu X, Zhao Y, Guan X, Yang L (2020) Evaluation of multiple gridded precipitation datasets for the arid region of northwestern China. Atmos Res 236:104818

Zhang Q, Fan K, Singh VP, Sun P, Shi P (2018) Evaluation of remotely sensed and reanalysis soil moisture against in situ observations on the Himalayan-Tibetan Plateau. J Geophys Res: Atmos 123(14):7132–7148

Zhang X, Lu N, Jiang H, Yao L (2020) Evaluation of reanalysis surface incident solar radiation data in China. Sci Rep 10(1):1–20