Prediction of Tropical Monsoon Hydrology Using Gridded Meteorological Products Over the Cau River Basin in Vietnam

DuyMinh Dao [1,2], Jianzhong Lu [1,*], Xiaoling Chen [1], PhamChiMai Phan [1,3] and DinhKha Dang [4]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; xiaoling_chen@whu.edu.cn

[2] Faculty of Geography, Thai Nguyen University of Education, 20 Luong Ngoc Quyen, Thai Nguyen 250000, Vietnam ; daoduyminh87@gmail.com

[3] Faculty of Tourism, Thai Nguyen University of Sciences, Tan Thinh ward, Thai Nguyen 250000, Vietnam; phanphamchimai87@gmail.com

[4] Department of Hydrology and Water Resources, University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi 100000, Vietnam; dangdinhkha@hus.edu.vn

* Correspondence: lujzhong@whu.edu.cn; Tel.: +86-27-68778755

Abstract: Gridded meteorological products are generated with different spatial/data and methods, and it will be sensitive to different regions for hydrological models. Therefore variables including temperature and precipitation should be evaluated before applying them in studies. To improve knowledge of this matter, the potential of two reanalysis products (RPs) including the China Meteorological Assimilation Driving Datasets for the SWAT model (CMADS) and Climate Forecast System Reanalysis (CFSR) is for the first time compared with the ground-based meteorological data in 5 years from 2008 to 2013 over the Cau River Basin (CRB), Northern of Vietnam. The statistical indicators, and the Soil and Water Assessment Tool (SWAT) model are employed to investigate the hydrological performance of the CRPs against the 17 rain gauses placed across the CRB. The result showed that there is a strong correlation of the temperature reanalysis in both CMADS, CFSR with ground-observed (correlation coefficient-CC is from 0.92 to 0.97). The division indicated clearly when CFSR data overestimated precipitation (about 88%) at both daily and monthly scales, whereas a slight variation of CMADS product was found in the high terrain. The flow simulation results also show that the performance of CMADS-SWAT (with value R2 > 0.75 and NSE > 0.78) is more accurate than CFSR-SWAT on the monthly scale. The assessment of the potential of CRPs especially CMADS will further provide an additional quick alternative for water resource research and management in basins with similar hydro-meteorological conditions.

Keywords: Cau river basin; CFSR; CMADS; SWAT model; tropical monsoon; reanalysis data; extreme hydrologic event

PACS: J0101

1. Introduction

Accurate and complete weather information has always been an important input, which is widely used in hydrological models, flood forecasting, and climate change as well as providing scientific guidance for water resources management [1], [2]. Practically, satellite estimates with wide coverage and high resolution have become potential data sources to supplement the rapid substitution in flow patterns especially in areas where surface observations are limited [3]. Together with The Soil and Water Assessment Tool – SWAT [4], the commonly used climate products from satellite data include Climate Forecast System Reanalysis NECP-CFSR [4] and China Meteorological Assimilation Driving Datasets for the SWAT model – CMADS. The study's results have shown encouraging performances of these products at different catchment scales [2], [8]. According to our latest understanding, the applicability of the two data above has yet been confirmed outside of China.

In this study, for the first time, the performance of CFSR and CMADS reanalysis data products (hereinafter referred collectively to as Climate Reanalysis Products - CRPs) as well as its reliability, were used in hydrological evaluation on a specific river basin in Vietnam. Therefore, the Cau River Basin (CRB) has been selected to carry out the studies with the following steps: (1) Compare and evaluate the efficiency of temperature and precipitation estimates of CFSR, CMADs data using Ground-based meteorological station (GMS) data with statistical indicates at temporal, spatial scales; (2) Analyze the capture of extreme weather events, especially (temperature, precipitation) occurring over the CRB; (3) Use the SWAT model to evaluate the capabilities of these products, which is the input data in hydrological research from 2008 to 2013.

2. Materials and Methods

2.1. Study area and data

The Cau river basin (21.070N-22.180N; 105.280E-106.080E) originates in the high mountains in the Northwest of Bac Kan province with an area of nearly 6300 km2. The river flows in a north-south direction with a total length of about 1602 km, of which the maintream length is 290 km. The average annual temperature in the CRB ~ 220C but there are 3 months (from December to February) when the average monthly temperature is below 200C. The mean annual precipitation ~ 1680mm with the rainy season (May to October) accounts for more than 80%, the rest is in the dry season. The mountains which are over 1000m high in the Northern, Northeast and Southwest have created climatic changes according to elevation and ridge effect (Figure 1).

Table 1. The information datasets used in this study

Data type	Products	Spatial Resolution	Temporal Resolution	Sources
Weather Data	CFSR	0.31x0.310	Daily (1979-7/2014)	NECP
	CMADS v.1.1	0.25x0.250	Daily (2008-2016)	CAU
	GMS	Point (17 stations)	Daily (2008-2013)	MORNE
Geography	DEM	30x30m	2005	USGS
	Soil	1x1km	2005	FAO
	Land use	1x1km	2005	MORNE
Hydrology	Discharge	Point (1 station)	Daily (2008-2103)	MORNE

Note: NECP: National Centers for Environmental Prediction; CAU: China Agricultural University; USGS: United States Geological Survey, FAO: The Food and Agriculture Organization, MORNE: Ministry of Natural Resources and Environment, Vietnam.

The input spatial data for the SWAT model in CRB include the Digital elevation model - DEM, soil data, land use and CFSR, CMADS, GMS weather datasets. The parameters used to evaluate the efficiency and simulate flow at CRB include: maximum / minimum temperature, daily precipitation from 01-01-2008 to 31-12-2013 to ensure consistency in evaluation and comparison between CRPs and gauge observed. We only use Gia Bay hydrological station in the flow simulation over the CRB basin (operating from 1997 to present) because other stations have stopped working. Information on these products is described in detail in Figures 1 and Table 1.

2.2. Modeling method and evaluation indicators

2.2.1. SWAT model

Arc SWAT version 2012, the interface in ArcGIS 10.2, was used to build a hydrological model for the CRB. SWAT basin delineation tool is based on baseline data (30 m DEM and river network)

dividing the CRB into 45 sub-basins and 350 hydrologic response units (HRUs) – consisting of homogenous topography, soil characteristics, land use and management. In all flow simulations by CFSR, CMADs and GMS data the warmup period was 2008, the calibration period was 2009-2010 and the validation period was 2011-2013. The calibration and validation of the flow simulation were conducted independently at daily and monthly time steps, results were compared with the values measured at Gia Bay hydrological station at the same time in this study.
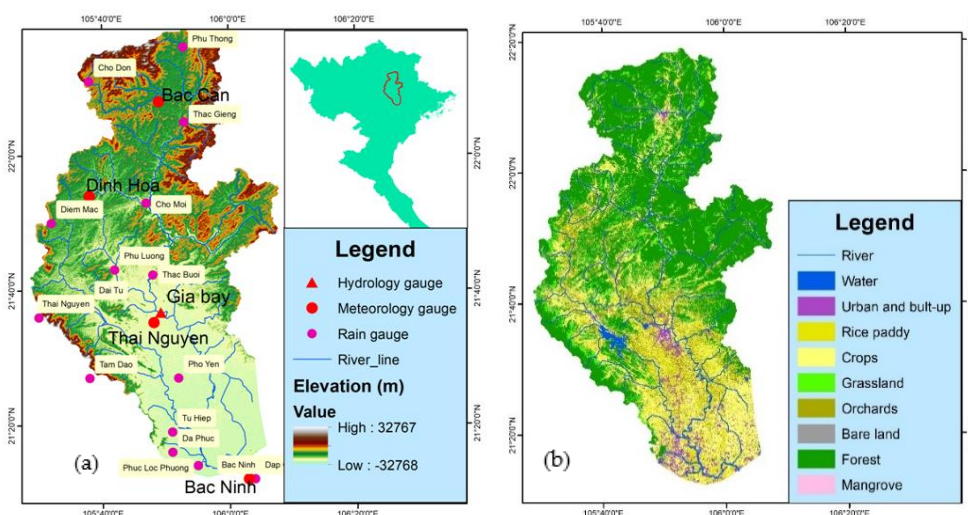
Figure 1. Study region map of the Cau River Basin: (a) including the location, digital elevation model, drainage network, and hydrological and meteorological stations; (b) land use map

2.2.2. Evaluation indicators

Within the CRB, there are only 4 meteorological stations that have temperature data information (maximum, minimum), while the numbers of grid points of CMADS / CFSR are relatively high, therefore the temperature data authentication was conducted only for climate stations. Meanwhile, the precipitation data are pulled together from 13 stations from 2008-2013 at the catchment scale. For GMS stations, the point to pixel assessment was applied by selected the closest CRPs grid points as references for data validation.

Aim for assessing the quantity of CRPs in collecting temperature and precipitation, the following indicators have been used: (i) Four basic statistical indicators such as Correlation coefficient (CC), Mean absolute error (MAE), Root mean square error (RMSE), and Percentage bias (PBIAS); (ii) Three statistical-categorical indicators to evaluate precipitation events, including Probability of detection (POD), False alarm ratio (FAR) and Critical success index (CSI). Calculation formula, unit, range of values, and their significance synthesized from other studies [2], [7], [8].

To assess the performance of CRPs in capturing extreme weather events, this study selected indices recommended by the Expert Team on Climate change Detection and Indices [9] and Circular regulates techniques and processes for dangerous hydro-meteorological forecasting of the Ministry of Natural Resources and Environment, Vietnam (2016). For precipitation, the values were chosen are the number of days with precipitation ≥ 10mm/50mm/100mm/day per year (R10, R50, R100 mm), and the extreme phenomena of temperature were: the number of cold days (with daily average temperature - Tav≤15^{0}C) per year. While the number of days when the strong sun occurs (with the daily maximum temperature range from 37^{0}C to 39^{0}C – Tmax37-39^{0}C) and the number of days scorching sun occurs (with daily maximum temperature ≥39^{0}C – Tmax39^{0}C) for the 2008-2013 period. Meanwhile, the performance of flow simulations in the CRB was assessed by NSE, R2, and PBIAS statistics with ranking criteria applied for daily and monthly scales [10].

3. Results and Discussion

3.1. CFSR and CMADS temperature validation using GMS data

Mean, CC, MAE, RMSE, MBE values are chosen to evaluate the accuracy of daily and monthly temperature data. The results from Table 2 show that CMADS as well as CFSR temperature data at all stations have a strong correlation with GMS data. With temperature data from CFSR data, the CC mean value is 0.92, and with CMADS data, it's 0.97. Simultaneously, ME value (Tmin ranging from 1.0 to 2.81 and Tmax ranging from 0.95 to 2.59), RMSE value (Tmax ranging from 1.39 to 2.47 and Tmin from 1.27 to 1.48) indicated that there's only a negligible difference in temperature between CRPs and GMS.

Table 2. analysis of CFSR and CMADS maximum temperature (Tmax) and minimum temperature (Tmin) in the CRB.

Tempe-rature	Values	Bac Kan		Dinh Hoa		Thai Nguyen		Bac Ninh	
		CFSR	CMADS	CFSR	CMADS	CFSR	CMADS	CFSR	CMADS
Tmax	Mean (^{0}C)	25.43	26.48	25.43	26.44	26.07	27.39	27.68	27.60
	CC	0.92	0.97	0.92	0.97	0.92	0.97	0.91	0.97
	MAE (mm)	2.81	1.5	2.67	1.51	2.37	1.17	2.12	1.0
	RMSE (mm)	2.47	1.53	2.46	1.57	2.51	1.57	2.85	1.39
	RB	-8.05	-4.22	-7.34	-3.63	-5.19	-0.39	0.33	0.05
Tmin	Mean (^{0}C)	17.8	19.66	17.8	19.94	18.47	20.68	19.9	20.81
	CC	0.97	0.97	0.96	0.97	0.96	0.97	0.97	0.97
	MAE (mm)	2.01	0.94	2.41	1.05	2.59	0.95	1.57	1.01
	RMSE (mm)	1.4	1.29	1.48	1.38	1.47	1.27	1.34	1.28
	RB	-9.64	-0.21	-11.62	-1.02	-12.11	-157	-6.37	-2.11

Almost all RB values are negative (except Tmax in BacNinh station), which implies that both CMADS and GMS data tend to underestimate the actual temperature. For Tmax from CRPs data, RB value gradually decreases from the North (high latitude and mountainous terrain) to the South (downstream), while for Tmin this trend is opposite, meanwith the RB's largest values in the downstream (except Tmin-CFSR at Bac Ninh station). This might be because CRPs consistent with estimates of small day-night temperature difference in mountains (Tmin) while in flat terrain the differences are larger (Tmax). Despite being slightly different, both CFSR and CMADS temperature data can be used as GMSs in CRB meteorological-hydrological studies.

3.2. CFSR and CMADS Precipitation validation using GMS data

3.2.1. At the temporal scales

Table 3 describes the results of continuous statistical indicators of precipitation on daily, monthly and seasonal time scales at the CRB. At the daily scale, CMADS data tends to underestimate precipitation, with PBIAS is -16.64% while CFSR data overestimate, with PBIAS is 99.2%. Therefore, MAE values are also quite different with 8.01mm / day for CFSR and 5.7mm / day for CMADS. As expected, this trend is also expressed on the monthly scale when the MAE and RMSE indices of CFSR data are larger than CMADS. The errors on the daily scale are canceled out due to aggregation so higher accuracy can be seen more at monthly scale (the CC value of CFSR, CMADS is 0.82, 0.84, respectively), but it still does not affect the major differences in its rating trends. Overall, CMADS data is more accurate than CFSR data and has a relatively good alignment with observed precipitation at monthly steps.

The PBIAS value distribution indicates that CFSR data has the worst performance during the rainy season (from May to October). CMADS rainfall distribution between dry and rainy seasons was

16% and 84%, while these values for the CMADS data were 11% and 89%, respectively (2008-2013 period). CMADS precipitation in the dry season is underestimated compared to the gauge observation, with the PBIAS value is -40.9% (Table 3). This is possibly related to the reduced efficiency of CMADS due to the CMORPH satellite data having rain detection errors below 4mm [6].

Table 3. statistical analysis of the CMADS and CFSR precipitation data on the CRB period 2008-2013.

| Time Scale	Products	Mean (mm)	CC	MAE (mm)	RMSE (mm)	PBIAS (%)
Daily	CFSR	9.04	0.45	8.01	15.5	99.2
	CMADS	3.81	0.31	5.7	15.63	-16.64
Monthly	CFSR	275.18	0.82	145.72	151.78	99.2
	CMADS	115.95	0.84	58.44	80.31	-16.64
Rainy season (V-X)	CFSR	2783.2	-0.2	1463.9	469.3	109.4
	CMADS	1227.2	0.4	256.5	199.7	-8.5
Dry season (XI-IV)	CFSR	523.1	0.5	277	111	114.8
	CMADS	145.1	0.6	121.5	73.3	-40.9

3.2.2. Accuracy of rainfall events detection

Values of 0.1mm / day were selected as the rainfall detection threshold [7], and POD, FAR, CSI indices were used to evaluate the ability of CRPs to detect precipitation. The POD average value with CFSR data is 0.98, indicates that it tends to capture all daily rain events. Concurrently, the FAR average value is 0.72 (vary from 0.56 to 0.74) indicates that only nearly 30% of the rain events forecasted from CFSR data are accurate. On the contrary, CMADS data shows the harmony in the forecasting, with POD and FAR of 0.6 and 0.2, respectively; consistent with rain forecasting success CSI of 43%. Overall, CMADS precipitation data is more accurate in estimating rainfall events while the CFSR data excels at its ability to detect rain but should still be validated with rain gauge data.

3.2.3. Ability evaluation of capturing extreme weather events

Statistics show that from 2008 to 2013, CFSR data had a total number of 252 hot days (with Tmax≥370C), larger than CMADS (118 days), and GMS (117 days). This difference relates to excessive misunderstandings about the buffer surface between datasets. For example, at Dinh Hoa station, the number of hot days is very high (32 days Tmax37-390C and 35 days Tmax390C), possibly due to the widely captured data of CFSR for low mountainous areas lying on the sheltered slopes in summer, meanwhile Bac Ninh station is located in the plain area with many industrial-construction activities thus the estimate of maximum temperature was incorrect due to the buffer surface.

With the cold days (Tav≤150C), CMADS is compared parallel with GMS because CFSR data is not available for this information. Statistics show that both of these datasets have similar changes with the number of cold days decreasing in high latitudes, mountainous terrain (Bac Kan, Dinh Hoa) to low latitudes, flat terrain (Thai Nguyen, Bac Ninh). The number of cold days is relatively high in both CMADS and GMS with 17.2 and 19.3 days/year, respectively, showing a strong influence of winter monsoons on CRB. The appearance of cold airwaves not only lower the area's heat base (from December to February, the temperature drops below 200C) but also very little rainfall during this period. It is clear that the assessment of extreme weather events at time scales is closely related to the precipitation, which directly affects the flow distribution in the CRB.
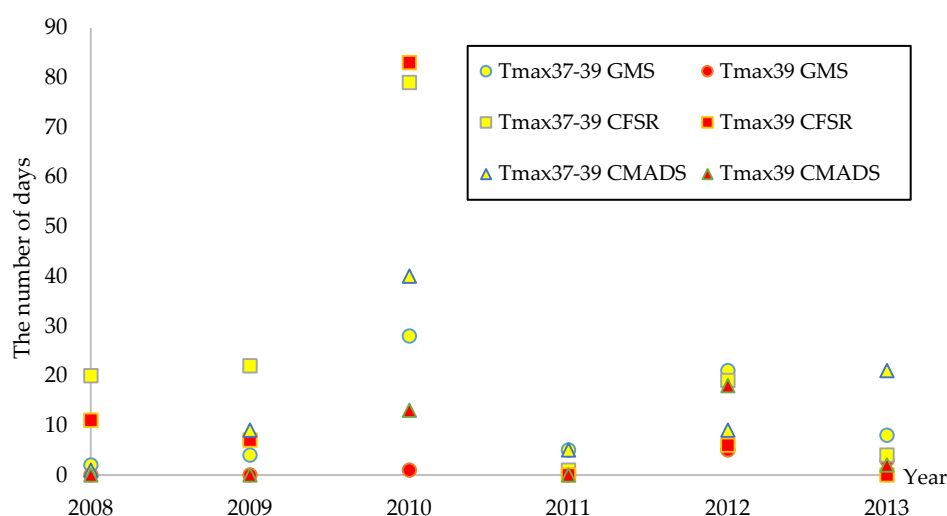
Figure 2. number of days when hot weather phenomena appeared in the period of 2008-2013.

Visual inspection at temporal scale distributions reveals that CFSR data had superior hot days (Tmax≥37^{0}C) compared to CMADS and CFSR data (Figure 2). Notably, an unusual increase in the number of hot days was found in all three datasets in 2010 (except GMS data at Tmax39^{0}C). According to information collected, in May and June 2010, the North of Vietnam experienced the longest heatwave in 27 years. These findings suggest that while there is a similar error at maximum temperature events, but when calibrated with ground observed temperature, the CRPs can provide an additional viable alternative to predicting and capturing extreme events on the temporal and spatial scale.

The assessment results of the ability to collect extreme heavy rainfall are calculated by the average value of the corresponding grid-points/station during the 2008-2013 period. From the CFSR precipitation, there are 609 R10mm days, captured more R10mm days than GMS (265 days), and CMADS data (224 days) at the coincident station. Whereas at the R50mm event, GMS and CFSR precipitation both captured the value of 49 days compared to 34 days for CMADS data. The R100mm value for GMS precipitation is 9 days, while for CFSR and CMADS, the values are 5 days, which indicates that CRPs tend to underestimate the results at heavy rain threshold (50mm, 100mm). Although the frequency of occurrence is low (2.24%), the heavy rain class (> 50mm) contribute up to 37% of the total annual rainfall as it is related to the typical summer rain as well as the effects of tropical storms on this area.

3.3. Evaluate the Performance of Hydrological processes in the Cau river basin

The statistical indices (R2, NSE and PBIAS) are summarized in Table 4 for the SWAT simulations driven by GMS, CFSR and CMADS data in the period of 2009-2013. Overall, the SWAT model based on GMS data was the most suitable in both daily and monthly scales during the calibration and validation period. The simulated streamflow reproduced by GMS data at Gia Bay station is "good", with NSE > 0.67 and R2 > 0.77. The simulation using the CMADS- driven model tends to appreciate the observed flow with the PBIAS value varying from 17.35 to 18.95% however with R2> 0.84 and NSE> 0.73 also identified as "satisfactory" at monthly scale. Finally, the CFSR data resulted in a relatively high overestimation of observed streamflow throughout the simulation period (as indicated by the high PBIAS value of -38%) and tends to capture the peaks streamflow (Figure 3). Generally, the CFSR- driven model is not suitable for flow simulation on the CRB basin with the R2 and NSE values as "unsatisfactory" based on the given criteria [10].
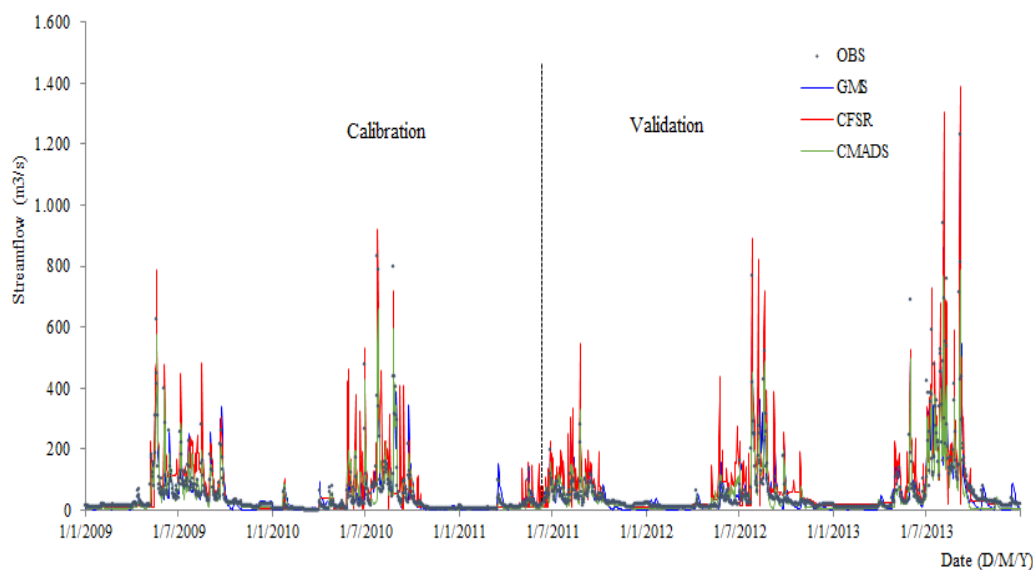
Figure 3. between the daily observed streamflow and results simulated using the GMS-, CMADS-, and CFSR- driven models in the period of 2009–2013.

These results are consistent with some studies in the Asian monsoon region where CFSR data are not really suitable for flow simulation [10], [11]. It is really difficult for estimated products like CFSR to accurately capture climatic conditions in areas with very complex climates like northern Vietnam (tropical monsoon climate with cold winters). Furthermore, differences in catchment area and topography (including the elevation and direction of the ridge) will also lead to changes in the algorithm, interpolation and model parameters. The CMADS data set is integrated with the CMROPH data and is collected from the automatic measuring stations in the region for reverse interpolation, so it can be widely used and increase accuracy in Chinese territory. Compared to published studies, we find that the performance of this data needs to be validated in areas within the coverage. Generally, the analytical results show that the CMADS- driven model will have a good performance if the input data was validated with the gauge observation.

Table 4. Statistical indices of the streamflow simulation for the GMS-, CFSR-, and CMADS- driven models.

Period	Index	Daily			Montly		
		GMS	CFSR	CMADS	GMS	CFSR	CMAD
Calibration (2009-2010)	R2	0.71	0.38	0.67	0.93	0.62	0.81
	NSE	0.91	0.32	0.68	0.87	0.58	0.83
	PBIAS (%)	0.06	-44.78	15.9	3.14	-35.23	17.35
Validation (2011-2013)	R2	0.67	0.22	0.62	0.89	0.51	0.75
	NSE	0.9	0.51	0.72	0.92	0.49	0.78
	PBIAS (%)	-0.68	-45.29	19.92	6.15	-38.81	18.95

4. Conclusions

The usefulness and suitability of the climate reanalysis products have been evaluated in this study. The CMADS and CFSR temperature datasets both performed well in comparison to GMS and promise rapid replacement in areas with a low number of observation stations. Verification of rainfall of CRPs as well as flow simulation results of the SWAT model on CRB shows that CMADS data has more suitable results; meanwhile, it is recommended that the overall CFSR data be evaluated before application in hydrological research where the conditions are similar. The advantages and disadvantages of CFSR, CMADS, and GMS data suggest that local knowledge/information is also

very useful in hydro-meteorological research to avoid excessive misunderstandings of gridded climate products.

Author Contributions: This manuscript is a result of the research of D.M.D. under the supervision of J.Z.L, X.L.C, and the advice of P.C.M.P and D.K.D. This paper was finished under the efforts of all authors. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. A. Aghakouchak, A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, "Evaluation of satellite-retrieved extreme precipitation rates across the central United States," J. Geophys. Res. Atmos., 2011, doi: 10.1029/2010JD014741.

2. J. Lu, Z. Liu, W. Liu, X. Chen, and L. Zhang, "Assessment of CFSR and CMADS Weather Data for Capturing Extreme Hydrologic Events in the Fuhe River Basin of the Poyang Lake," J. Am. Water Resour. Assoc., 2020, doi: 10.1111/1752-1688.12866.

3. M. Faramarzi et al., "Setting up a hydrological model of Alberta: Data discrimination analyses prior to calibration," Environ. Model. Softw., vol. 74, pp. 48–65, 2015, doi: 10.1016/j.envsoft.2015.09.006.

4. A. Saleh et al., "Application of SWAT for the Upper North Bosque River Watershed," Trans. Am. Soc. Agric. Eng., vol. 43, no. 5, pp. 1077–1087, 2000, doi: 10.13031/2013.3000.

5. Y. Cao et al., "Application of SWAT model with CMADS data to estimate hydrological elements and parameter uncertainty based on SUFI-2 algorithm in the Lijiang River basin, China," Water (Switzerland), vol. 10, no. 6, 2018, doi: 10.3390/w10060742.

6. L. Zhang, X. Meng, H. Wang, M. Yang, and S. Cai, "Investigate the applicability of CMADS and CFSR reanalysis in Northeast China," Water (Switzerland), vol. 12, no. 4, 2020, doi: 10.3390/W12040996.

7. K. D. Dinh, T. N. Anh, N. Y. Nguyen, D. D. Bui, and R. Srinivasan, "Evaluation of grid-based rainfall products and water balances over the Mekong river Basin," Remote Sens., vol. 12, no. 11, 2020, doi: 10.3390/rs12111858.

8. M. L. Tan, P. W. Gassman, and A. P. Cracknell, "Assessment of three long-term gridded climate products for hydro-climatic simulations in tropical river basins," Water (Switzerland), vol. 9, no. 3, 2017, doi: 10.3390/w9030229.

9. T. R. Karl, N. Nicholls, and A. Ghazi, "CLIVAR/GCOS/WMO Workshop on Indices and Indicators for Climate Extremes - Workshop summary," Clim. Change, vol. 42, no. 1, pp. 3–7, 1999, doi: 10.1023/A:1005491526870.

10. D. N. Moriasi, M. W. Gitau, N. Pai, and P. Daggupati, "Hydrologic and water quality models: Performance measures and evaluation criteria," Trans. ASABE, vol. 58, no. 6, pp. 1763–1785, 2015, doi: 10.13031/trans.58.10715.

11. M. T. Vu, S. V. Raghavan, and S. Y. Liong, "SWAT use of gridded observations for simulating runoff - A Vietnam river basin study," Hydrol. Earth Syst. Sci., vol. 16, no. 8, pp. 2801–2811, 2012, doi: 10.5194/hess-16-2801-2012.