International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1622-1628www.ijaem.net ISSN: 2395-5252

Integrating climate reanalysis data and ground station observations in the SWAT model on the Cau river basin, Vietnam

Phan Pham Chi Mai [1], Hoang Hai Thai [2], Ha Thi To Uyen [2]Dao Duy Minh [2*]

[1]*Thai Nguyen University of Science, Vietnam*
[2*] *Thai Nguyen University of Education, Vietnam*
Corresponding Author: Dao Duy Minh

Date of Submission: 20-04-2023 Date of Acceptance: 30-04-2023

ABSTRACT: Climate reanalysis data products offer the potential to represent weather in remote areas, poor quality, and short-term observations. Moreover, this dataset is also widely used in hydrological models, especially SWAT models. This paper addresses the potential of NECP-CFSR and CMADS reanalysis datasets in simulated flow performance investigations. The results extracted from the model output show encouraging signals (only R2 and NSE are "good") and can be reliably used in river basins with similar conditions.
KEYWORDS: NECP-CFSR, CMADS, The Cau river basin, SWAT model, hydrological.

I. INTRODUCTION

Nowadays, it is recognized that the critical role of hydrological models in water resource planning is to meet diverse needs for their sustainable use. The Soil and Water Assessment Tool (SWAT), a hydrological model and water quality assessment developed by the United States Department of Agriculture, is used to analyze the impact of land use changes for discharge, erosion, sedimentation, and water quality. This versatile model has been widely used with over 4.000 publications and applied by many government and private organizations, universities, and research centers [1].

Spatial data used as input for this model include weather information, land data, land use, and land cover... In which weather data source is the most important and decisive variable. Determine the performance of the model. Based on advances in remote sensing and GIS techniques, several satellite-based precipitation products with global high resolution (up to 0.25°) are now available, such as CFSR and CMADS... has become a potential source of additional data for hydrological models performed in watersheds with a shortage of observation stations. In general, CFSR and CMADS are analytical datasets (developed based on satellite data sources and then calibrated based on terrestrial observations to obtain better estimates) with much potential use because they have higher spatial resolution than other products and are publicly available to the study areas including temperature and precipitation.

NECP-CFSR is a reanalysis product developed by the US National Center for Environmental Forecasting [2]. This system was collected by NCEP from 1979 to July 2014 with a resolution of 0.31° (~38km) based on the assimilation of sea-ocean-land ice system data globally on the same scale. Among the earliest NCEP-CFSR evaluation studies were performed in China's Three Gorges Reservoir, the Upper Blue Nile Basin, and five US watersheds. These studies show that NCEP-CFSR can simulate flow satisfactorily using SWAT. Since then, many similar studies have been carried out in different parts of the world, with some reporting unsatisfactory performance by NCEP-CFSR, especially in tropical and subtropical regions [3].

Recently, CMADS has been a new, reanalyzed climatological data for the SWAT model that has been used more frequently in hydrological studies. The CMADS dataset is developed by the Agricultural University of China based on the integration of the Local Forecasting and Analysis System/Space-Time Multilayer Analysis System (LAPS / STMAS), the CPC MORPHing technique (CMROPH), and China Meteorological Information

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1622-1628www.ijaem.net ISSN: 2395-5252

Center national precipitation data. The data sources used for the CMADS series are available from 2008 to 2016 and cover all of East Asia (0°-65° N; 60°-160°E). The studies on this data source focus mainly on river basins in China, such as the Lijiang River, Yangtze River, Poyang Lake, and Northeast region... with a quite encouraging performance. Based on its full Asia coverage and improved accuracy, CMADS promises to be one of the most useful satellite-derived weather datasets for meteorological and hydrological research [4].

The Cau river basin (CRB) is the most important tributary of the Thai Binh river system in the North of Vietnam. With a total basin of more than 6,300km^2, this river flows through 6 provinces and cities, from high mountains to midland terrain and plains. The river basin is associated with many production activities of ethnic minorities, and the natural environment is undergoing drastic changes [5]. To our knowledge, the applicability of CMADS and CSFR data for hydrometeorological studies has not yet been fully explored outside China, including river basins in Vietnam.

Stemming from the above issues, building a hydrological model in the basin to understand the change in the water cycle, thereby managing freshwater in this basin in a more appropriate way, is a very necessary task. Set. Furthermore, little is known about both CMADS and SWAT applications in CRB. Therefore, this study aims to: (1) validate weather data at the daily scale of NCEP-CFSR and CMADS, such as maximum and minimum temperatures; (2) to evaluate the ability of the Ground-based meteorological station (GMS), NCEP-CFSR, CMADS in flow simulation using SWAT in CRB.

II. MATERIALS AND METHODS

2.1 Study area

CRB is located in an area with a tropical monsoon climate with cold winters. The average temperature in the area ranges from 22-24^{o}C. However, due to the influence of the polar air mass in winter from the Siberian high pressure, the winter lasts three months (December to February), and the temperature drops below 20^{o}C. Wetness is reflected in the average rainfall of over 1.500mm, with 85% in the rainy season (May to October) and the rest in the dry season.

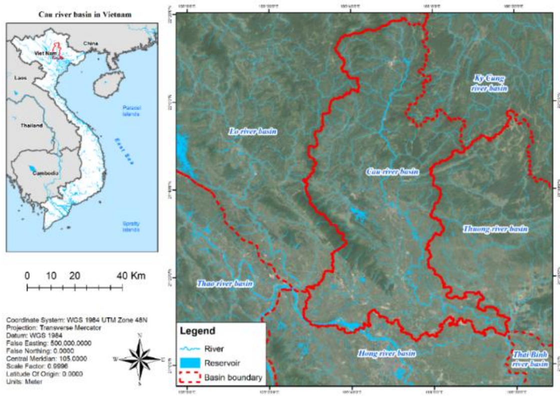

Figure 1. Geographic location of the CRB

Located in the Northern Midlands and Mountains, the main area of the basin is mainly natural forest land, grasslands, and shrubs distributed in watershed areas, high mountainous terrain in Bac Kan province, Thai Nguyen (accounting for about 48%). The type of agricultural land accounts for about 36%, distributed along the river valley where the terrain is flat. Construction land (8%) is located in the lower river basin, where the population is densely populated, and the economic and development centers are located. Develop. Hydrologic variables (water level and discharge) on the mainstream were observed at seven stations, but some stations stopped measuring discharges in the 1980s and 1990s. Recent river discharge is only at Gia Bay station (Thai Nguyen province) in the center of the river basin.

2.2 Input data of the model

a. Digital Elevation Model (DEM)

For this study, the spatial resolution of 30m SRTM (Shuttle Radar Topography Mission) was extracted from the USGS (https://lpdaac.usgs.gov/data_access)/data_pool) and used as an input dataset.

b. Soil type and characteristics

In this study, different soil types were identified based on the soil type database compiled by the Food and Agriculture Organization of the United Nations (FAO, 2005). Ten soil forms were identified in the study basin, which was then grouped into seven soil groups and mapped for the area based on the criteria of Fey (2001). Meanwhile, 2005 land use data were collected from the Ministry of Natural Resources and Environment (MONRE) and classified based on the SWAT format.

c. Reanalysis datasets

- NECP-CFSR data: In CRB, daily meteorological data of 15 grid points were downloaded from the official website (https://globalweather.tamu.edu/).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1622-1628www.ijaem.net ISSN: 2395-5252

- CMADS data: In this paper, we use CMADS v1.1, with a spatial resolution of 0.25° and 24 grid points available (this product is available from the following website: http://www.cmads.org/)

2.3 Run the SWAT model

The SWAT model was used to estimate all components of the water balance in the study basin. ArcSWAT 2012 version, an interface in ArcGIS 10.2, is used to build hydrological studies in CRB. Installation instructions and research papers related to the SWAT model are available at https://swat.tamu.edu/, as well as in an online open-source document.

In the simulation process, watershed delineation is the first procedure. Once the basin delineation is complete, the identification of hydrological response units (HRUs) is continued. Using 30m DEM data and river networks, the watershed delineation in ArcSWAT creates 42 sub-basins in the CRB. These sub-basins are further subdivided into 405 HRUs with different soil, topographic, and soil management characteristics. All necessary climate variables are then included in the model, including precipitation, minimum and maximum temperature, relative humidity, average wind speed, and solar radiation data. The GMS, CFSR, and CMADS weather data are then provided as input to the flow simulation (hereinafter referred to as the CFSR_, CMADS_ and GMS-driven SWAT model), and the results are verified with the Gia Bay hydrology station.

2.4 Evaluation indicators

The assessment of NCEP-CFSR and CMADS is divided into two main parts: (1) about the climate aspect: confirm the rainfall, Tmax/Tmin with the climatometer at the ground station; and (2) hydrological aspect comparing monthly flows simulated by SWAT using measuring station, NCEP-CFSR, and CMADS with observed flow. In general, products with grids provide regional climate information for each respective grid, while the measuring station measures climate variability at a specific point. Grid-to-grid and point-to-grid are two approaches commonly used to compare products and gridded meters. The grid-to-grid approach requires a high density of gauges in a particular grid for robust comparisons. In fact, the lack of measuring stations is a major problem in various parts of the world, including the CRB. In addition, some uncertainties may arise during the interpolation of multiple measuring stations into grid-based observable data. The climate aspect comparison is therefore made using the point-to-grid approach, where the measuring stations are compared directly with their respective grid values. The evaluation is limited to grids containing at least one measure.

The ability of NCEP-CFSR and CMADS to estimate rainfall and maximum and minimum temperatures from 2008 to 2014 at daily and monthly scales was assessed using three widely applied statistical measures, such as Correlation Coefficient (CC), Root Mean Squared Error (RMSE) and Percent bias (PBiAs). CC, RMSE, and PBiAs are commonly used to evaluate the fit, mean error magnitude, and systematic deviation, respectively, between the measuring station and grid station data. A reliable grid product can be declared if the CC value is greater than 0.7 and the PBiAs value is between -10% and 10%.

The hydrological aspect involved the comparison of flows simulated using observed climate data, NCEP-CFSR and CMADS. The factor of Determination (R2), Nash-Sutcliffe Efficiency (NSE), and RB was used to evaluate the performance of SWAT in the discharge simulation, showing the performance ratings to measure the efficiency of SWAT was proposed by Moriasi et al. and is widely applied in SWAT studies. Flow comparison only focuses on monthly scale assessment due to the difficulty of collecting reservoir management data. Indeed, a similar situation occurs in other river basins of Vietnam, where reservoir management data is considered very sensitive data to the public.

III. RESULTS

3.1 Parameter sensitivity analysis and calibration

Parameter sensitivity analysis shows that CN2 and ESCO are among the most sensitive parameters in CRB SWAT calibration, while GW_Delay and SOL_AWC are less sensitive than others.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1622-1628www.ijaem.net ISSN: 2395-5252

Table 1. Sensitivity values of the parameters used for flow simulations by GMS_, using the SWAT model in the Cau river basin.

Parameter	Name	Range	Calibrated value
v__ALPHA_BF.gw	Baseflow alpha factor	0-1	0.7
v__GW_DELAY.gw	Groundwater delay time	0-500	250
v__GWQMN.gw	Threshold water depth in the shallow aquifer for flow	0-5000	2500
v__GW_REVAP.gw	Groundwater re-vaporisation coefficient	0.02-0.2	0.11
v__REVAPMN.gw	Threshold depth of water in the shallow aquifer for “revap” to occur	0-500	250
v__CANMX.hru	Maximum canopy storage	0-100	10
v__SURLAG.bsn	Surface runoff lag coefficient	0.05-24	16.815
v__CH_N2.rte	Manning value for main channel	-0.01-0.3	0.207
v__CH_K2.rte	Effective hydraulic conductivity in main channel alluvium	-0.01-500	449.99
r__CN2.mgt	Initial SCS runoff curve number for moisture condition II	-0.42-0.065	-0.275
r__SOL_AWC.sol	Available water capacity of soil layer	0-3.347	2.342
r__SOL_K.sol	Saturated hydraulic conductivity	0-306.6	30.66
r__ESCO.hru	Soil evaporation compensation factor	0-1	0.1
r__EPCO.hru	Plant uptake compensation factor	0-1	0.7

3.2 Compare CFSR and CMADS temperatures using GMS data

In general, Tmax and Tmin tend to increase gradually from north to south, i.e., from high latitudes and hilly areas (e.g., Bac Kan station, Dinh Hoa station) to lower margins (e.g., North station) Ninh). Both the CFSR and CMADS temperature data showed strong correlations with the GMS data collected at the observation stations. The mean CC values obtained from CFSR were 0.92 for Tmax and 0.97 for Tmin, while that of CMADS was 0.96 and 0.97, respectively. The average RMSE was 1.8, varying from 1.27 to 2.85; these values show that the CFSR and CMADS data agree with the observed temperatures at the stations. Figure 2 shows a box plot of the maximum and minimum temperatures of the CFSR, CMADS, and ground station data in the CRB. The boxes show the CFSR temperature with the most extensive range of change. The mean values and temperature ranges between the analytes were consistent in CMADS compared with those in the CFSR. Although the data sets differ slightly, CFSR and CMADS temperature data can be used as GMS station data in CRB hydro-meteorological studies.

The PBIAS value is negative at most stations (except for Tmax at Bac Ninh station), reflecting that both the CFSR and CMADS data tend to underestimate the maximum and minimum temperatures compared to the observed temperatures. Specifically, the PBIAS value is much smaller at the Bac Ninh station than at other stations, showing that the CFSR and CMADS data have higher accuracy in low-terrain areas. In contrast, the increasing PBIAS values (especially with CFSR data) observed in regions with more elevated topography may be related to temperature estimation errors that occur due to the influence of mountain clouds, and minimum temperature values usually cause high PBIAS in winter.

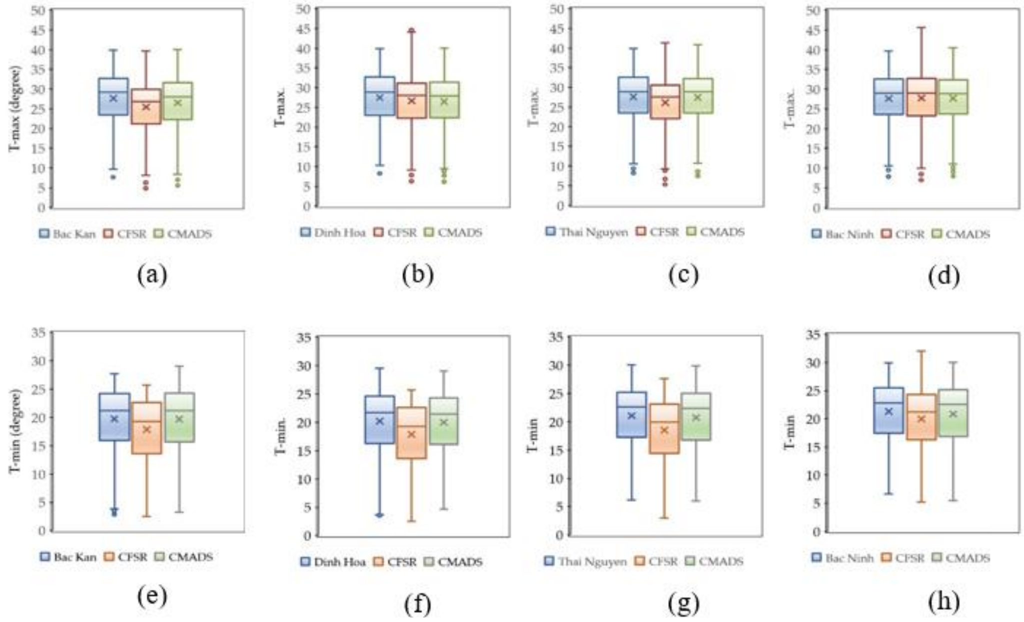

Figure 2. Box plots of daily maximum (a-d) and minimum (e-h) temperatures from CFSR and CMADS at the Bac Kan, Dinh Hoa, Thai Nguyen and Bac Ninh meteorological stations

Table 2. Statistical indicators used to evaluate temperature (maximum, Tmax/minimum, Tmin) in the CFSR and CMADS data in the CRB

Tempe-rature	Values	Bac Kan		Dinh Hoa		Thai Nguyen		Bac Ninh	
		CFSR	CMADS	CFSR	CMADS	CFSR	CMADS	CFSR	CMADS
Tmax	Mean (°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
	RMSE (°C)	2.47	1.53	2.46	1.57	2.51	1.57	2.85	1.39
	PBIAS	-8.05	-4.22	-7.34	-3.63	-5.19	-0.39	0.33	0.05
Tmin	Mean (°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
	RMSE (°C)	1.4	1.29	1.48	1.38	1.47	1.27	1.34	1.28
	PBIAS	-9.64	-0.21	-11.62	-1.02	-12.11	-1.57	-6.37	-2.11

3.3 Calibration and validation of the SWAT model against parameters calibrated from GMS and CMDAS weather data

As the studies done by [6] carried out on the CRB concluded that the CFSR rain data is overestimated compared to the measured station rainfall. This is due to differences in the data and algorithms used in these satellite rain products. CFSR rainfall data taken from the global forecasting system tends to overestimate precipitation. Estimation products such as CFSR are quite difficult to accurately capture very complex climatic conditions, such as in Northern Vietnam, combined with differences in basin area and topography (including elevation and mountain orientation) will also change the results of modeling algorithms and interpolation... Compared with CFSR rain data, CMADS rainfall shows a better linear correlation when compared to GMS and generally has a lower bias. Therefore, the CMADS precipitation data may be more representative than the CFSR data in the study area.

Because the study aims to evaluate the ability of satellite data in hydrological studies, we propose an option to combine input data for the SWAT model in this section. Accordingly, the optimal sensitivity parameters of GMS, CFSR temperature data, and CMADS precipitation (from now on referred to as the **SWAT** model driven by **G**round station sensitivity parameter and **M**ulti-source **C**limate reanalysis data, **SWAT_G-MC**) are used as inputs in the model. Simulated flows

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1622-1628www.ijaem.net ISSN: 2395-5252

will be compared with the Gia Bay hydrological station flow for the period of calibration (2009-2011) and validation (2012-2013). The results from Figure 3 shows that the flow curve taken by SWAT_G-MC has a good agreement with the observed flow at a monthly scale. R2>0.8 value; NSE>0.75, so the simulated flow is said to be "good" according to the criteria of Moriasi (2007).

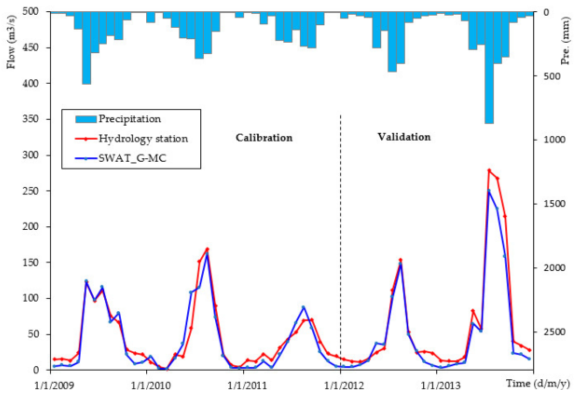

Figure 3. Calibration and validation with SWAT model driven by Ground station sensitivity parameter and Multi-source Climate reanalysis data, SWAT_G-MC uses monthly observed flow at Gia Bay hydrological station

The PBIAS value of -3.5% in the calibration phase and -5.2% in the validation phase also shows encouraging results. For extreme flows, differences are found in the curve for the dry months (November-April), as shown in the figure showing that the angle of SWAT_G-MC lies below the implied that it is underrated. This may be related to the decline in rainfall during this period of the estimates from the CMADS than the actual rainfall reported by [7]. Even so, the flow made by combining these sources of information can give satisfactory results and reproduce the basic flow in some cases where the observed flow is deficient.

3.4 Discussion

In addition to the quality of meteorological data, the number and spatial distribution of meteorological stations are two important factors affecting hydrological simulations. Upstream and downstream of the CRB are wide and narrow, respectively. Four meteorological stations are respectively distributed in the upstream and downstream, midstream and downstream. According by [8], suitable rain density is in the range of 1.0~1.4/1000km^2; however, the density is 0.33/1000 km^2 in CRB and SWAT_GMS achieves good performance in flow simulation. This is possible because meteorological stations are relatively evenly distributed in the CRB, which can reflect the rainfall situation well and offset the scarcity of stations. Unfortunately, meteorological stations are often unevenly distributed over most areas, in which case CMADS can be a supplement to GMS.

Normally, parameterization in a SWAT model is done by calibration and validation against different meteorological data sets because the non-uniqueness of that parameter is an inherent property of the inverse model, so there is no best parameter. This strategy was also suggested in the studies of Lu, Meng, and Tan to overcome the uncertainty in the data [9]. This study proposes a method to check the model results with other data sources. Specifically, we combined using the parameters corrected from the terrestrial meteorological station and weather data (CFSR temperature combined with CMADS precipitation) in the SWAT model (hereinafter referred to as: SWAT model driven by Ground station sensitivity parameter and Multi-source Climate reanalysis data, SWAT_G-MC) for comparison in simulation. The comparison has shown that this method has significantly improved the performance of the model compared with the conventional strategy based on the parameters as stated. The model results at the hydrological station with R2 > 0.8 and NSE > 0.7 show that SWAT_G-MC can be applied in case the ground weather data is difficult to access. This approach provides yet another novel solution to the potential of reanalyzed data for hydrological studies if the parameters are corrected to improve model performance. Obviously, if the model input, especially the precipitation variable, is verified prior to application in hydrological studies (e.g., CMADS), it will give the modeler confidence about the model output.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1622-1628www.ijaem.net ISSN: 2395-5252

IV. CONCLUSION

(1) Verification of temperature in CRB shows that CFSR and CMADS can be represented as terrestrial temperature measurement stations in meteorological and hydrological studies. Both CMADS and CFSR temperature data sets perform well against GMS data so that CMADS and CFSR temperature data can be used reliably in areas with a low number of observation stations.

(2) In general, the SWAT controlled by G-MC achieves good performance in flow simulation. Better results are found at peak flow, and G-MC degradation curves closely match the actual flow. This suggests an additional approach for watersheds for which available data sources are lacking and incomplete, such as the CRB basin, using satellite weather data combined with parameters obtained from on-site measurements.

REFERENCES

[1]. Tan, M.L.; Gassman, P.W.; Srinivasan, R.; Arnold, J.G.; Yang, X. A Review of SWAT Studies in Southeast Asia: Applications, Challenges and Future Directions. Water **2019**, 11, 914

[2]. Fuka, D.R.; Walter, M.T.; MacAlister, C.; Degaetano, A.T.; Steenhuis, T.S.; Easton, Z.M. Using the Climate Forecast System Reanalysis as weather input data for watershed models. Hydrol. Process. **2014**, 28, 5613–5623

[3]. Bressiani, D.A.; Gassman, P.W.; Fernandes, J.G.; Garbossa, L.H.P.; Srinivasan, R.; Bonumá, N.B.; Mendiondo, E.M. Review of Soil and Water Assessment Tool (SWAT) applications in Brazil: Challenges and prospects. Int. J. Agric. Biol. Eng. **2015**, 8, 9–35

[4]. Meng, X.; Wang, H.; Shi, C.; Wu, Y.; Ji, X. Establishment and Evaluation of the China Meteorological Assimilation Driving Datasets for the SWAT Model (CMADS). Water **2018**, 10, 1555

[5]. Bui, H.H.; Ha, N.H.; Nguyen, T.N.D.; Nguyen, A.T.; Pham, T.T.H.; Kandasamy, J.; Nguyen, T.V. Integration of SWAT and QUAL2K for water quality modeling in a data scarce basin of Cau River basin in Vietnam. Ecohydrol. Hydrobiol. **2019**, 19, 210–223

[6]. Dao, D.M.; Lu, J.; Chen, X.; Kantoush, S.A.; Binh, D.V.; Phan, P.; Tung, N.X. Predicting Tropical Monsoon Hydrology Using CFSR and CMADS Data over the Cau River Basin in Vietnam. Water **2021**, 13, 1314

[7]. Zhang, D.; Tan, M.L.; Dawood, S.R.S.; Samat, N.; Chang, C.K.; Roy, R.; Tew, Y.L.; Mahamud, M.A. Comparison of NCEP-CFSR and CMADS for Hydrological Modelling Using SWAT in the Muda River Basin, Malaysia. Water **2020**, 12, 3288

[8]. Arnold, J.G.; Fohrer, N. SWAT2000: Current capabilities and research opportunities in applied watershed modelling

[9]. Jianzhong Lu; Zixuan Liu; Weizhe Liu; Xiaoling Chen; Ling Zhang. Assessment of CFSR and CMADS Weather Data for Capturing Extreme Hydrologic Events in the Fuhe River Basin of the Poyang Lake. Journal of the American Water Resources Association **2020**, 56, 917 -934