https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Discussions

Machine learning based streamflow prediction in a hilly catchment for future scenarios using CMIP6 data

Dharmaveer Singh[1]*, Manu Vardhan[2], Rakesh Sahu[3], Debrupa Chatterjee[1], Pankaj Chauhan[4], and Shiyin Liu[5*]

1. Symbiosis Institute of Geo-informatics, Symbiosis International (Deemed University), Pune-411004 (India).
2. Computer Science and Engineering Department, National Institute of Technology Raipur, Raipur-492010 (India).
3. Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai-400076 (India)
4. Wadia Institute of Himalayan Geology, Dehradun-248001 (India)
5. Institute of International Rivers and Eco-security, Yunnan University, Kunming - 650091 (China)

*Corresponding author's email: veermnnit@gmail.com and shiyin.liu@ynu.edu.cn

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences Discussions
Open Access

Abstract

The alteration in river flow patterns, particularly those that originate in the Himalayas, has been caused by the increased temperature and rainfall variability brought on by climate change. Due to the impending intensification of extreme climate events, as predicted by the Intergovernmental Panel on Climate Change (IPCC) in its sixth assessment report, it is more essential than ever to predict changes in streamflow for future periods. Despite the fact that some research has utilised machine learning-based models to predict streamflow patterns in response to climate change, very few studies have been undertaken for a mountainous catchment, with the number of studies for the western Himalaya being so small as to be considered insignificant. This study investigates the capability of different machine learning (ML) models, including the Gaussian Linear Regression Model (GLM), Gaussian Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), Artificial Neural Network (ANN), and Random Forest (RF), in streamflow prediction over the Sutlej River Basin in western Himalaya during the periods 2041-2070 (2050s) and 2071-2100 (2080s) for two greenhouse gas trajectories (SSP245 and SSP585). Coupled Model Intercomparison Project Phase 6 (CMIP6) bias corrected data downscaled at grid resolution of 0.25°× 0.25° for 6 General Circulation Models (GCMs) were used for this purpose. Four different rainfall scenarios (R_0, R_1, R_2, and R_3) were applied to the models trained with daily data (1979-2009) at Kasol (the outlet of the basin) in order to better understand how catchment size and the geo-hydro-morphological aspects of the basin affect runoff. RF model with rainfall scenario R_3 which outperformed other models during the training and testing period therefore was chosen to simulate streamflow in the Sutlej River in the 2050s and 2080s under the SSP245 and SSP585 scenarios. The mean ensemble of model results show that the mean annual streamflow of the Sutlej River is expected to rise between 2050s and 2080s by 5.51 to 6.04% for SSP585 and by 6.65 to 6.75% for SSP245. The seasonal

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

streamflow also is expected to increase in the 2050s and 2080s under both emission scenarios, with the exception of the pre-monsoon, where a decline in streamflow is anticipated for SSP585 in the 2080s. However, under both the emission scenarios, there seemed to be significant variation in the streamflow simulations among the individual models for various time periods. It has been found that the pattern of this variability is highly correlated with the pattern of precipitation and temperature predicted by various GCMs for future emission scenarios. The present study will therefore assist in strategy planning for ensuring the sustainable use of water resources downstream by acquiring a knowledge of the nature and causes of unpredictable streamflow patterns.

Keywords: Machine learning models; streamflow; climate change; CMIP6; western Himalaya

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

1. Introduction

Human-induced global warming has altered patterns of the rainfall worldwide (Goswami et al., 2006; Trenberth, 2011), and also increased risks of extreme events such as the droughts and floods (Easterling et al., 2000; Trenberth et al., 2015; Otto et al., 2017). It has impacted hydrology of many river basins globally, including variation in streamflow (Gerten et al., 2008; Lutz et al., 2014; Nepal and Shrestha, 2015; Singh et al. 2015a; Ali et al., 2018; Singh et al., 2022). A study of long-term (1948-2004) streamflow (discharge) data of 200 largest rivers of the globe showed considerable change in their annual discharge, however, results were statistically significant only for 64 rivers. Out of which 45 were marked with decreasing trends and the remaining 19 showed increasing trends in their annual discharge (Dai et al., 2009). Similar trends in discharge of the rivers were reported also at regional scale: Asia (Kundzewicz et al., 2009; Krysanova et al., 2015), Europe (Stahl et al., 2010; Stahl and Tallaksen, 2012) and America (Pasquini and Depetris, 2007). Moreover, it has been established that the effects of rainfall variation and extreme events on annual discharge are likely strong compared with other drivers (Miller et al., 2012; Van der Wiel et al., 2019). A projected rise of ~2°C to 5°C in mean annual global temperature by 2100 under higher greenhouse gas emission scenarios as predicted from the General Circulation Models (GCMs) (Gao et al., 2017) will considerably affect the rainfall pattern (intensity and amount) and may have adverse effects on hydrological cycles (Okai and Kanae 2006; Haddeland et al., 2014). This would subsequently impact availability of water resources and present challenges for their management since a rise in the demand of water is also predicted (Lutz et al., 2019). Therefore, it is indispensable to know the underlying hydrological dynamics occurring within a basin in context of climate change for effective management and sustainable use of the water resources.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

The underlying hydrological processes controlling rainfall-runoff generation in a basin can be understood with the use of a hydrological model which is based on complex mathematical equations and theoretical laws governing physical processes in the basin (Kirchner, 2006; Singh et al., 2019). It simulates/or predicts response of the basin to climatological forcings such as the rainfall (Sood and Smakhtin, 2009) and generate synthetic time series of hydrological data that could be used by water managers and scientists for varied applications ranging from water budgeting and partitioning (Conan et al., 2003; Schreiner-McGraw and Ajami, 2020; Masse-Dufresne et al., 2021) to inundation mapping and modelling (Mahto et al., 2022). A hydrological model is supposed not only to have a good predictive power but also the ability of capturing relationships among the forcing factors and catchment response so that an accurate estimate of rainfall-runoff could be made (Shortridge et al., 2016). However, until now, there is no hydrological model that can simulate basin-behaviour universally well against all the hydrological challenges inflicted from climate change and human-interventions (Yang et al., 2019). As a result, many hydrological models have been devised considering functioning and robustness of models in explaining underlying complexity in quantifying basin-scale response to small-scale spatial complexity of physical processes (Shortridge et al., 2016; Herath et al., 2021). Broadly, these can be grouped into two categories: physical or process-based models and empirical or data-driven models (Yang et al., 2019; Kabir et al., 2020). The latter category of models uses a mathematical relationship established between runoff and affecting factors in the basin for deriving the runoff (Muhammad Adnan, 2019).

It is purported that the data-driven model despite of inherited limitations over physical interpretability of processes has outperformed the physical models in terms of prediction accuracy in many hydrological applications (Shortridge et al., 2016; Muhammad Adnan, 2019; Kabir et al., 2020; Herath et al., 2021). Also, they are preferred over the physical models for rainfall-runoff modelling/or streamflow prediction modelling due to limited requirements of data as inputs, where data limitation is the major challenge (Beven, 2011). These models in past were heavily criticised on the ground of being incompetent to model the non-linear behaviour of streamflow (Yang et al., 2019). But recent developments in computational intelligence, in the area of Machine Learning (ML) in particular, have greatly expanded the capabilities of empirical modelling. This resulted in the development of many non-linear models such as the Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machine (SVM), which can capture and model non-stationarity of the rainfall-runoff relationships (Shortridge et al., 2016; Muhammad Adnan, 2019; Yang et al., 2019). Despite of the significance potentials of the ML models in streamflow prediction, relevant studies assessing the application of these models for streamflow prediction under future scenarios over the mountainous basins are limited due to non-availability of long-term data. Hence, it is important to test whether machine learning approaches can be effectively used over a mountainous river basin to predict streamflow using hydro-meteorological variables and CMIP6 scenarios as the input data.

With a catchment area of 56874km^2 (up to Bhakara Dam), the Sutlej also pronounced as 'Satluj' is an important river in the western Himalayas and runs through diverse climatic zones. The flow in the upper and middle catchment is primarily impacted by glacier/snow melt induced by seasonal temperature shift and preceding winter precipitation, while the lower section of the catchment area is mostly regulated by rainfall both in the winter and during the monsoon season (Singh and Jain, 2002; Archer, 2003; Miller et al., 2012). Based on data from the period 1986–1996, Singh and Jain (2002) estimated the mean yearly contribution of snow/glacier melt and rainfall to the Sutlej River as being 59% and 41%,

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

respectively. However, the discharge in the river peaks is directly related to the peak in rainfall during the monsoon (Lutz et al., 2014). Recent studies on this basin has raised concerns about the implications of climatic changes on streamflow since a warming climate has brought changes in the amount and spatial-temporal distribution of precipitation (Singh et al., 2014; Singh et al., 2015b). Previous research has only used process-based hydrological models to date when examining the effects of climate change (past and present) on streamflow patterns in the region (Singh and Jain, 2002; Singh et al., 2015a; Ali et al., 2018; Shukla et al., 2022), which leaves a gap in the use of machine learning models. This study very first time examines the potential of various ML models namely, Gaussian Linear Regression Model (GLM), Gaussian Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), ANN and RF in streamflow prediction over the Sutlej River Basin (rainfall dominated zone) in western Himalaya during the period 2041-2070 (2050s) and 2071-2100 (2080s) and its relationship to climate variability.

2. Study Area

Sutlej is a Trans Himalayan river with its origin in the Rakshastal Lake (elevation:4570m) near to the Mansarovar Lake in the western Himalaya. It is the largest tributary of the Indus River. The total catchment area of the Sutlej River upto Bhakara Dam is about 56874km^2, of which 22305 km^2 (extending between Spiti Valley and Bhakara) falls within India. It is a high relief catchment (elevation: 350-6558 m) dominated by rainfall, but also received significant contribution from glacier/snow melt. The selected study area is a sub-catchment within the basin (Figure 1), with an area of 2457 km^2. It is dominated mostly by forests (56.20%), grassland (26.4%), agricultural lands (17.1%), and glaciers and snow covers (0.3%) (Singh et al., 2015a). The details of the sub-catchment are summarised in Table 1.

Figure 1: The location of the sub-catchment within Sutlej River Basin

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences Discussions

The basin bestowed with the hydropower potential of about 9226.75MW is climatologically sensitive and, at present, facing the challenges created due to climate change and human's interventions (Singh et al., 2015b and 2015c). Change in future climate will alter patterns of flow in river and further could affect water resources and hydroelectric power production (Singh et al., 2014). Therefore, the present study will provide useful insight to devise better strategy for the management of water resources in the Sutlej basin.

Table 1: Characteristics of the study catchment over the evaluation period of 1979–2009

3. Description of the Data and Methods

The methodology involved in predicting streamflow for the period 2021-2100 in the Sutlej River include: 3.1) collection of hydro-meteorological data, 3.2) selection of machine learning models, and 3.3) performance evaluation of the developed models. These are described in details under following sub-headings:

3.1 Hydro-meteorological data

The daily rainfall, temperature (T_{max} and T_{min}), relative humidity, solar radiation, wind speed and discharge data used to study performance of the different machine learning models on streamflow modelling were collected for 31 years i.e. 1979-2009. Rainfall, temperature and discharge data were obtained from the Bhakara Beas Management Board (BBMB), while relative humidity, solar radiation and wind data were extracted from the Global Weather Data (http://globalweather.tamu.edu/). The outputs from the CMIP6, the latest generation of climate models, were used for streamflow prediction. Even by using downscaled GCMs, however, regional climate change projections inherit biases from the GCM boundary conditions, which were corrected in the dataset detailed in Mishra et al. (2020) for South

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences Discussions
Open Access
EGU

Asia. This dataset provides bias-corrected downscaled climate change projections for 13 CMIP6 models and four GHG emission scenarios (SSP126, SSP245, SSP370, and SSP585), the latter are briefly summarised in Riahi et al. (2017). The data are available at a daily time-scale and horizontal spatial resolution of 0.25°×0.25°. This study used the bias-corrected downscaled regional climate change scenarios from six CMIP6 GCMs: 1) Beijing Climate Centre Climate System Model version 2 (BCC-CSM2-MR), 2) Russian Institute for Numerical Mathematics Climate Model Version 4.8 (INM-CM4-8), 3) Russian Institute for Numerical Mathematics Climate Model Version 5.0 (INM-CM5-0), 4)Australian Community Climate and Earth System Simulator-Earth System Model version 1 (ACCESS- ESM1.5), 5) Australian Community Climate and Earth System Simulator-coupled model version 2 (ACCESS-CM2), and 6) Earth Consortium-Earth 3 Veg Model (EC-Earth-veg) which were selected based on their performances in simulating precipitation and temperature over the western Himalaya to examine future patterns in streamflow for the period 2021-2100 in the Sutlej River according to two GHG emissions scenarios: SSP245 and SSP585.

3.2 Selection of machine learning models for streamflow modelling

In this study, five machine learning models namely GLM, GAM, MARS, ANN and RF were selected and their performances in predicting streamflow in Sutlej River were compared. These are regression based models which capture relationship between the predictors (dependent variables) and predictand (independent variables) and provide value of the output variables (Muhammad Adnan, 2019; Kabir et al., 2020). The models were trained with daily data recorded during 1979-2009 at Kasol (the gauging site), the outlet of the basin. However, prior to building the models, all of the data were normalized using standard normalization techniques to get features on a common scale. Further, the entire data set was split into training and testing datasets since a cross-validation method was adopted in this study. The

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

training dataset (80%) was used for fitting the models whereas testing dataset was used for checking model accuracy (20%). Under the cross-validation method, the process was repeated until every part of the allocated data was used in testing (Kabir et al., 2020). This technique enables the buildout of generalising models, that if tested correct could be applied in predicting flow for ungauged watersheds of similar geographical and climatic characteristics where past rainfall/runoff are not available.

Five different program codes were written in python language for GLM, GAM, MARS, ANN and RF simulations. Out of these five selected models, GLM, GAM and MARS are linear models whereas other two i.e. ANN and RF are non-linear in nature (Shortridge et al., 2016; Yang et al., 2019; Herath et al., 2021). Additionally, excluding GLM all of the remaining ML models are based on non-parametric regression approach where functional relationship between predictor and predictand are not predetermined but can be adjusted to capture unusual or unexpected features of the data (Shortridge et al., 2016). A detailed description of these models can be found elsewhere (Shortridge et al., 2016; Muhammad Adnan, 2019; Yang et al., 2019; Kabir et al., 2020; Herath et al., 2021). Six variables namely rainfall, T_{max}, T_{min}, relative humidity, solar radiation and wind speed were used as the inputs for developing the models. Additionally, these models were simulated under four rainfall scenarios: rainfall on the same day (R_0), rainfall lagged by one day (R_1) and rainfall lagged by two days (R_2) and rainfall lagged by three days (R_3) to understand control of catchment size and geo-hydro-morphological characteristics of the basin in generating runoff. While, remaining meteorological parameters were held constant during the processes.

3.3 Model performance evaluation

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

It has been found that overfitting in a model may lead to large errors in out-of-sample predictions (Hastie et al., 2009). Therefore, it has been evaded by establishing model parameters for GLM, GAM, MARS, ANN and RF through automated hyperparameter tuning methods. 500 bootstrap resamples of the training data set were generated for each parameter value to be assessed. Table 2 presents the information on the specific parameters evaluated for each model.

Table 2: The information on hyper parameters used for estimating model parameters

Three evaluation criteria were used to assess the daily streamflow predictions of different models. These were coefficient of determination (R^2), ratio of the root mean square error to the standard deviation of measured data (RSR) and mean absolute error (MAE). The RSR and MAE are calculated as given in equation 4 and 5, respectively.

$$RSR = \frac{\sqrt{\sum_{i=1}^{n} \frac{(P_i - O_i)^2}{n}}}{STDEV_{Obs}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$$MAE = \frac{\sum_{i=1}^{n} |P_i - O_i|}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

where P_i are the predicted values and O_i are the observed values, n accounts for the number of samples, $\bar{O}$ represents the mean of observed data, and P_i is the mean of predicted data and $STDEV_{Obs}$ refers to the standard deviation of observed values. R^2 explains the correlation between observed and predicted values, lies between 0 and 1. R^2 values between 0.6 to 0.7 are considered satisfactory, 0.85 to 1 are very good and below 0.4 are unsatisfactory. The lower is the RSR value, the better is the model. Values greater than 0.7 are unsatisfactory whereas those lying between 0 and 0.5 come in the very good range. Mean absolute error accounts for the mean of the absolute differences in the observed and predicted values. Lower MAE is favourable.

4. Results and Discussion

4.1 Streamflow simulation and evaluation of model performance

The simulation (1979-2009) results generated under different rainfall scenarios (R_0, R_1, R_2 and R_3) on daily time scale for all five models (GLM, GAM, MARS, ANN and RF) is shown in Table 3. R^2 values across models ranged from 0.71 to 0.90 and from 0.73 to 0.81 during training and testing, respectively. Likewise, it was found that RSR and MAE varied from 0.55 to 0.31 and from 123.25 to 71.95 during training, as well as from 0.56 to 0.46 and from123.06 to 106.64 during testing, respectively. However, within models, RF performed better at runoff prediction under all rainfall scenarios (R_0, R_1, R_2, and R_3) compared to the other models, while GLM showed the poorest results. R^2, RSR and MAE values for the RF model during the training ranged from 0.88 to 0.90, 0.32 to 0.34, 71.95 to 77.49, respectively. Similar results were revealed during testing, with these falling between 0.76 and 0.78, 0.47 and 0.49, and 106.64 and 111.85, respectively. On the other hand, throughout training, the R^2, RSR, and MAE values for the GLM model varied from 0.69 to 0.71, 0.54 to 0.55, and 134.80 to 140.56, respectively. During testing, they varied between 0.69 and 0.71, 0.54 and 0.56, and 134.35 and 141.26, respectively.

Table 3: Summary of model performance in simulating streamflow at Kasol

Figure 2, 3, 4 and 5 shows comparison of observed and simulated streamflow under rainfall scenarios of R_0, R_1, R_2 and R_3 for all the models at Kasol, the outlet of the basin. As observed from the Figures (2-5), RF was able to follow the curve better compared to the other models.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

It is also deduced from the comparison of scatter plots wherein a relatively smaller deviation in the observed and estimated discharge of streamflow was found for the RF model. GLM performed the worst out of the five models with respect to the time variation graphs. A limitation faced by all the five models was the simulation of peak values. The models slightly underperformed at the prediction of higher values of streamflow. The model performed slightly better during training than testing periods. Furthermore, it was observed that the models with rainfall scenario R_3 had revealed reasonably better results in comparison to R_0, R_1 and R_2 scenarios, indicating delayed contribution of rainfall-runoff to the river.

Figure 2: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_0

Figure 3: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_1

Figure 4: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_2

Figure 5: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_3

R^2, RSR and MAE values across models under the rainfall scenario R_3 were found to range from 0.71 to 0.90, 0.34 to 0.54, and 72 to 134.80 during training and from 0.71 to 0.90, 0.47 to 0.54, and 106.6 to 134.80 during testing. These findings led to the ultimate decision to use the RF model with rainfall scenario R_3 to predict streamflow in the Sutlej River in the future (2050s and 2080s) under the SSP245 and SSP585 scenarios.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

4.2 Annual streamflow changes in 2050s and 2080s under SSP245 and SSP585

Figure 6 lists the projected change in mean annual streamflow for the Sutlej River in 2050s and 2080s with respect to the base period (1979-2009) under different emission scenarios. Out of six models, four (BCC-CSM2-MR, EC-Earth-3, INM-CM4.8, and INM-CM5.0) projected an increase in annual streamflow in the 2050s and 2080s for both emission scenarios. Depending on the model, it will be in the range of 0.46 to 2.84% in the 2050s and from 1.32 to 4.37% in the 2080s under SSP245 scenario. The models with the lowest increases in mean annual streamflow are BCC-CSM2-MR (0.46%) in the 2050s and INM-CM4.8 (1.32%) in the 2080s, while INM-CM5.0 and EC-Earth-3 have shown the highest increases (2.84% and 4.37%) in 2050s and 2080s, respectively. Similar to this, GCM models predict that rise under SSP585 will vary between 0.18% (ACCESS- ESM1.5) and 3.16% (EC-Earth-3) and 0.60% (ACCESS- ESM1.5) to 3.50% (EC-Earth-3) between 2050s and 2080s. However, a decrease in mean streamflow is predicted by ACCESS- ESM1.5 and ACCESS- ESM2 in 2050s and 2080s under SSP245 and by ACCESS- ESM2 under SSP585. It will range from -0.21 to -2.69% in 2050s and -0.35 to -2.21% in 2080s under SSP245 and from -2.75 to -4.08% under SPP585. These variations in streamflow prediction from GCMs may be attributed to their spatial resolution and parametrization levels (Singh et al., 2015c). Therefore, in order to reduce uncertainty in projection of streamflow related to individual GCMs, the yearly streamflow pattern of the Sutlej River was analysed also using the mean ensemble of all six GCMs. According to the mean ensemble of the models, between 2050 and 2080, the Sutlej River's annual stream flow will increase by 5.51 to 6.04 percent for SSP585 and by 6.65 to 6.75 percent for SSP245. In general, the rise is expected to be higher in 2080s as compared to 2050s under both the scenarios. This rise in mean annual streamflow is impacted largely by the substantial increase in precipitation predicted over the basin during

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

2050s and 2080s. It is also established by previous research on the Himalayan catchments which revealed that changes in the amount and direction of precipitation are significantly more powerful predictors of water availability in the catchment than the presence of glaciers (Miller et al., 2022).

Figure 6: Predicted change in mean annual streamflow of the Sutlej River with respect to the reference period (1979-2009) in 2050s and 2080s under SSP245 and SSP585 for different GCMs

Further, the projected streamflow patterns for the Sutlej River under SSP245 and SSP585 for the 2050s and 2080s show similar tendencies, but with differing magnitudes, that have been found by past researchers using process-based hydrological models. For instance, Singh et al. (2015 a) used the SWAT (Soil Water Assessment Tool) model, a semi-distributed hydrological model, to simulate streamflow for future periods using two CMIP3 models (CGCM3 and HadCM3), and they discovered that the Sutlej River's mean annual streamflow would increase in the range of 0.6 to 7.8% in the 2080s under higher emission scenario of A2, which is equivalent to the SSP585 of the CMIP6. Similar to this, using the Variable Infiltration Capacity (VIC) and SWAT models, respectively, Ali et al. (2018) and Shukla et al. (2022) estimated increases in the Sutlej River's mean annual streamflow for the 2050s and 2080s under RCP4.5 and RCP8.5. The study of Shukla et al. (2022) estimates that under RCP4.5 and RCP8.5, the mean streamflow of the river will increase by 14 and 21%, respectively, in the 2080s. The relatively higher increase in the projected streamflow may be attributed to the model's input variables (rainfall and temperature), which were derived from CORDEX CCSM4 experiments, a regional climate model.

4.3 Seasonal streamflow changes in 2050s and 2080s under SSP245 and SSP585

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

The projected change in seasonal streamflow of the Sutlej River in 2050s and 2080s is shown in the Figure 7. It is observed that all models for all scenarios indicated an increase in streamflow during the post-monsoon and winter season in the 2050s and 2080s, while a decrease in streamflow is seen in 2080s under higher GHG scenarios during the pre-monsoon. Four out of six models predict rise in streamflow under both the scenarios in 2050s (excluding under SSP245) and 2080s during monsoon season. Further, across the models and within scenarios, there is a considerable difference in the seasonal streamflow predictions. It will vary from +9.77% (BCC-CSM2-MR) to +15.56% (ACCESS- ESM2) in the 2050s and from +10.77% (ACCESS- ESM1.5) to 18.28% (ACCESS- ESM2) in the 2080s during winter season, from +5.17% (INM-CM5.0) to +39.86% (BCC-CSM2-MR) and from +1.85% (INM-CM4.8) to +46.98% (BCC-CSM2-MR) during post-monsoon season, from -4.31% (EC-Earth-3) to +7.49% (INM-CM5.0) and from -5.65% (EC-Earth-3) to +5.61% (INM-CM5.0) during pre-monsoon seasons, and from -5.20% (ACCESS- ESM2) to +1.91% (INM-CM5.0) and from -7.90% (ACCESS-ESM2) to +5.00% (INM-CM54.8) during monsoon season under SSP245 scenario. Similarly, under scenario SSP585, it will vary from +9.80% (BCC-CSM2-MR) to +15.15% (INM-CM5.0) in the 2050s and from +9.34% (BCC-CSM2-MR) to +18.39% (INM-CM5.0) in the 2080s during winter season, from +1.30% (INM-CM5.0) to +41.94% (BCC-CSM2-MR) and from +0.42% (ACCESS- ESM1.5) to +48.54% (BCC-CSM2-MR) during post-monsoon season, from -7.28% (INM-CM4.8) to -0.11% (ACCESS-ESM1.5) and from -10.03% (INM-CM4.8) to -2.22% (ACCESS- ESM1.5) during pre-monsoon seasons, and from -7.95% (ACCESS- ESM2) to +6.78% (EC-Earth-3) and from -6.29% (ACCESS- ESM2) to +8.15% (INM-CM4.8) during monsoon season. It is observed that within seasons, maximum (+48.54%) and minimum (-10.03%) changes are predicted in 2080s under SSP585 during post-monsoon and pre-monsoon seasons, respectively.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Discussions

Figure 7: Predicted change in seasonal streamflow pattern of the Sutlej River with respect to the reference period (1979-2009) in 2050s and 2080s under SSP245 and SSP585 for different GCMs

However, mean ensembles of models predicted a rise in streamflow of the Sutlej River in all seasons in the 2050s and 2080s under both emission scenarios, with the exception of the pre-monsoon, where streamflow is estimated to decrease in the 2080s under SSP585. In the 2050s, this will range from +5.60 to +5.99% the winter, +10.28 to +13.16% during the post-monsoon, +2.02 to 6.66% during the pre-monsoon, and +8.75 to +10.63% during the monsoon. In the 2080s, this will range from +6.30 to +6.47% percent during the winter, +9.93 to +10.14% during the post-monsoon, -1.09 to +3.48% during the pre-monsoon, and +11.07 to +11.43% during the monsoon.

4.4. Monthly Streamflow Changes in 2050s and 2080s under SSP245 and SSP585

Figure 8 illustrates how mean monthly streamflow patterns will alter in 2050s and 2080s under both GHGs scenarios (SSP245 and SSP585) for different GCMs. A considerable variation in the streamflow pattern is observed within months and for the models. All six models generally in both scenarios expected an increase in monthly streamflow from January to April and a reduction in June. It will range from +0.78 to +11.89% in January, from +5.07 to +18.17% in February, from +3.30 to +30.45% in March, from +13.20 to +50.52% in April, and from -21.35 to -6.97% in June. However, no clear pattern is seen for the other months of the year. For instance, three (BCC-CSM2-MR, EC-Earth-3 and INM-CM4.8) out of six GCMs projected a reduction in streamflow for the months of May (-15.97 to -0.10%), while four (ACCESS- ESM2.0, BCC-CSM2-MR, EC-Earth-3 and INM-CM4.8) predicted reduction ranging between -16.35 to -0.75% for July under all scenarios. However, all GCMs

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

showed an increase in streamflow during the months of November (+0.94 to +29.75%) and December (+0.10 to +7.81%), with the exception of INM-CM5.0 (in November) and ACCESS-ESM1.5 (in December) that predicted decline in the streamflow ranging from -2.42 to -0.80% and to -1.08 to -0.14% for the same period.

Figure 8: Predicted change in monthly streamflow pattern of the Sutlej River with respect to the reference period (1979-2009) in 2050s and 2080s under SSP245 and SSP585 for different GCMs

The pattern of projected streamflow derived from the mean ensembles of the models, on the other hand, is consistent for both the future periods under SSP245 and SSP585, and it exhibits relatively less variability across the months (Figure 8). With the exception of June, November and December, the increase in streamflow is anticipated for every month of the year under both scenarios in 2050s and 2080s. The maximum increase (+27.01%) in monthly streamflow is predicted in April in 2050s under SSP245 whereas minimum (-4.42%) in June under SSP585 for the same period. The findings of the current research work, which showed that the majority of models, including mean ensembles of the mean, reveal a decline in the pattern of the streamflow under both scenarios for all of the periods, were validated by the findings of Ali et al (2018) and Shukla et al (2022) studies' predictions of a decline in streamflow during May and June (Figure 9). This implies that the RF model can be successfully applied to streamflow simulation and modelling in the Himalayan environment where data availability is a constraint.

Figure 9: Comparison of monthly observed (1979-2009) and projected discharge of the multi-model ensembles for period 2050s and 2080s under SSP245 and SSP585 scenarios

5. Conclusion

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Discussions

Five machine learning models, GLM, GAM, MARS, ANN, and RF, were tested in this study to simulate rainfall-runoff responses over the hilly Sutlej River Basin in order to determine the best model for simulating streamflow response to future climate change in the 2050s and 2080s under SSP245 and SSP585 using CMIP6 data. In terms of runoff prediction, RF outperformed the other models, as per the statistical evaluation criteria (R^2, RSR, and MAE), whereas GLM produced the worst results. The RF model's R^2, RSR, and MAE values during training varied from 0.88 to 0.90, 0.32 to 0.34, and 71.95 to 77.49, respectively. However, during testing, it ranged between 0.76 and 0.78, 0.47 and 0.49, and 106.64 and 111.85, respectively. The developed RF model was then used to simulate streamflow responses for six GCMs (ACCESS- ESM1.5, ACCESS- ESM2, BCC-CSM2-MR, EC-Earth-3, INM-CM4.8, and INM-CM5.0) and the mean ensembles of the models to investigate the implications of future climate change on the Sutlej River pattern in the 2050s and 2080s under SSP245 and SSP585 emission scenarios.

Within months, seasons, years, and for the models, a considerable fluctuation in the streamflow pattern is seen. These variations in streamflow prediction may be illustrated by the variations in spatial resolution and parametrization levels of GCMs, which led to a noticeable fluctuation in the anticipated amounts of temperature and precipitation during the study periods. Therefore, the monthly, seasonal, and annual streamflow patterns of the Sutlej River were also studied using the mean ensemble of all six GCMs in order to reduce uncertainty in streamflow projection due to individual GCMs. Both emission scenarios predict an increase in the Sutlej River's mean annual streamflow (6.65-6.75% under SSP245 and 5.51-6.04% under SSP585) as well as its seasonal streamflow in the 2050s and 2080s, with the exception of the pre-monsoon season. The post-monsoon season (10.28–13.16%) in the 2050s and the monsoon season (11.07–11.43%) in the 2080s are anticipated to experience

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

the highest seasonal increases. Similar to this, the rise in streamflow is predicted for every month of the year under both the 2050s and 2080s scenarios, with the exception of June, November, and December. The largest increase in streamflow is observed for April in the 2050s under SSP245 while the least increase is predicted for June with SSP585 over the same time period. Additionally, the projected changes in the mean annual and seasonal streamflow of the river are consistent with earlier research done using process-based physical hydrological models. Thus, the outcomes of the overall study indicate that the random forest model is efficient for simulating streamflow in the Himalayan catchment, and that water availability will rise as a result of an increase in catchment precipitation, which would eventually lead to an increase in hydropower generation. The administrators of local water resources and the government organizations in charge of maintaining reservoirs downriver may find these details on streamflow patterns to be of great use.

Code Availability: The codes developed for this study is made available to the readers on reasonable request.

Data Availability: The observed station data are confidential and authors do not have permission for sharing the data.

Author's Contribution

DS and SL conceptualized the problems, supervised the entire research activity from its inception to the completion, contributed in data collection, processing, interpretation and wrote the research paper. MV and RS contributed in the development of model, generation of figures and analysis of data. PC contributed in the data analysis and interpretation.

Statements and Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Authors acknowledge National Natural Science Foundation of China (NFSC Grant no. 42171129) for funding this research work, and Bhakara Beas Management Board (BBM), India for the hydro-meteorological data used in this study. We thank Ms. Pratibha Shrivastava, a fourth-year B.Tech computer science student from NIT Raipur, for her help in building the model.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

References

Ali, S. A., Aadhar, S., Shah, H. L., & Mishra, V. (2018). Projected increase in hydropower production in India under climate change. Scientific Reports, 8(1), 1-12.

Archer, D. (2003). Contrasting hydrological regimes in the upper Indus Basin. *Journal of Hydrology*, *274*(1-4), 198-210.

Beven, K. J. (2011). Rainfall-runoff modelling: the primer. John Wiley & Sons.

Conan, C., de Marsily, G., Bouraoui, F., & Bidoglio, G. (2003). A long-term hydrological modelling of the Upper Guadiana river basin (Spain). Physics and Chemistry of the Earth, Parts A/B/C, 28(4-5), 193-200.

Dai, A., Qian, T., Trenberth, K. E., & Milliman, J. D. (2009). Changes in continental freshwater discharge from 1948 to 2004. Journal of climate, 22(10), 2773-2792.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., & Mearns, L. O. (2000). Climate extremes: observations, modeling, and impacts. Science, 289(5487), 2068-2074.

Gao, Y., Gao, X., & Zhang, X. (2017). The 2 C global temperature target and the evolution of the long-term goal of addressing climate change—from the United Nations framework convention on climate change to the Paris agreement. Engineering, 3(2), 272-278.

Gerten, D., Rost, S., von Bloh, W., & Lucht, W. (2008). Causes of change in 20th century global river discharge. Geophysical Research Letters, 35(20).

Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., & Xavier, P. K. (2006). Increasing trend of extreme rain events over India in a warming environment. Science, 314(5804), 1442-1445.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Herath, H. M. V. V., Chadalawada, J., & Babovic, V. (2021). Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. Hydrology and Earth System Sciences, 25(8), 4373-4401.

Kabir, S., Patidar, S., & Pender, G. (2020, April). Investigating capabilities of machine learning techniques in forecasting stream flow. In Proceedings of the Institution of Civil Engineers-Water Management (Vol. 173, No. 2, pp. 69-86). Thomas Telford Ltd.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).

Krysanova, V., Wortmann, M., Bolch, T., Merz, B., Duethmann, D., Walter, J., ... & Kundzewicz, Z. W. (2015). Analysis of current trends in climate parameters, river discharge and glaciers in the Aksu River basin (Central Asia). Hydrological Sciences Journal, 60(4), 566-590.

Kundzewicz, Z. W., Nohara, D., Tong, J., Oki, T., Buda, S., & Takeuchi, K. (2009). Discharge of large Asian rivers–Observations and projections. Quaternary International, 208(1-2), 4-10.

Lutz, A. F., ter Maat, H. W., Wijngaard, R. R., Biemans, H., Syed, A., Shrestha, A. B., ... & Immerzeel, W. W. (2019). South Asian river basins in a 1.5 C warmer world. Regional Environmental Change, 19(3), 833-847.

Mahato, P. K., Singh, D., Bharati, B., Gagnon, A. S., Singh, B. B., & Brema, J. (2022). Assessing the impacts of human interventions and climate change on fluvial flooding

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

using CMIP6 data and GIS-based hydrologic and hydraulic models. Geocarto International, 1-26.

Masse-Dufresne, J., Barbecot, F., Baudron, P., & Gibson, J. (2021). Quantifying floodwater impacts on a lake water budget via volume-dependent transient stable isotope mass balance. Hydrology and Earth System Sciences, 25(6), 3731-3757.

Miller, J. D., Immerzeel, W. W., & Rees, G. (2012). Climate change impacts on glacier hydrology and river discharge in the Hindu Kush–Himalayas. *Mountain Research and Development, 32*(4), 461-467.

Mishra V, Bhatia U, Tiwari A.D. 2020. Bias-corrected climate projections for South Asia from Coupled Model Intercomparison Project-6. Sci. Data. 7, 338. https://doi.org/10.1038/s41597-020-00681-1.

Muhammad Adnan, R., Yuan, X., Kisi, O., Yuan, Y., Tayyab, M., & Lei, X. (2019, June). Application of soft computing models in streamflow forecasting. In Proceedings of the institution of civil engineers-water management (Vol. 172, No. 3, pp. 123-134). Thomas Telford Ltd.

Nepal, S., & Shrestha, A. B. (2015). Impact of climate change on the hydrological regime of the Indus, Ganges and Brahmaputra river basins: a review of the literature. International Journal of Water Resources Development, 31(2), 201-218.

Otto, F. E., Skeie, R. B., Fuglestvedt, J. S., Berntsen, T., & Allen, M. R. (2017). Assigning historic responsibility for extreme weather events. Nature Climate Change, 7(11), 757-759.

Pasquini, A. I., & Depetris, P. J. (2007). Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview. Journal of Hydrology, 333(2-4), 385-399.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., ... & Tavoni, M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environmental Change*, *42*, 153-168.

Schreiner-McGraw, A. P., & Ajami, H. (2020). Impact of uncertainty in precipitation forcing data sets on the hydrologic budget of an integrated hydrologic model in mountainous terrain. Water Resources Research, 56(12), e2020WR027639.

Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences, 20(7), 2611-2628.

Shukla, S., Jain, S. K., & Kansal, M. L. (2021). Hydrological modelling of a snow/glacier-fed western Himalayan basin to simulate the current and future streamflows under changing climate scenarios. Science of The Total Environment, 795, 148871.

Singh, D., Gupta, R. D., & Jain, S. K. (2014). Study of long-term trend in river discharge of Sutlej River (NW Himalayan region). Geography, Environment, Sustainability, 7(3), 87-96.

Singh, D., Gupta, R. D., & Jain, S. K. (2015a). Assessment of impact of climate change on water resources in a hilly river basin. Arabian Journal of Geosciences, 8(12), 10625-10646.

Singh, D., Gupta, R. D., & Jain, S. K. (2015b). Statistical analysis of long term spatial and temporal trends of temperature parameters over Sutlej river basin, India. Journal of Earth System Science, 124(1), 17-35.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Singh, D., Jain, S. K., & Gupta, R. D. (2015c). Statistical downscaling and projection of future temperature and precipitation change in middle catchment of Sutlej River Basin, India. Journal of Earth System Science, 124(4), 843-860.

Singh, D., Rai, S. P., & Rai, D. (2019). Application of geospatial techniques in hydrological modelling. In Sustainable Green Technologies for Environmental Management (pp. 167-173). Springer, Singapore.

Singh, P., & Jain, S. K. (2002). Snow and glacier melt in the Satluj River at Bhakra Dam in the western Himalayan region. *Hydrological Sciences Journal*, *47*(1), 93-106.

Singh, D., Zhu, Y., Liu, S., Srivastava, P. K., Dharpure, J. K., Chatterjee, D., ... & Gagnon, A. S. (2022). Exploring the links between variations in snow cover area and climatic variables in a Himalayan catchment using earth observations and CMIP6 climate change scenarios. *Journal of Hydrology*, *608*, 127648.

Sood, A., & Smakhtin, V. (2015). Global hydrological models: a review. Hydrological Sciences Journal, 60(4), 549-565.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., Van Lanen, H. A. J., Sauquet, E., ... & Jódar, J. (2010). Streamflow trends in Europe: evidence from a dataset of near-natural catchments. Hydrology and Earth System Sciences, 14(12), 2367-2382.

Stahl, K., Tallaksen, L. M., Hannaford, J., & Van Lanen, H. A. J. (2012). Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble. Hydrology and Earth System Sciences, 16(7), 2035-2047.

Trenberth, K. E. (2011). Changes in precipitation with climate change. Climate Research, 47(1-2), 123-138.

Trenberth, K. E., Fasullo, J. T., & Shepherd, T. G. (2015). Attribution of climate extreme events. Nature Climate Change, 5(8), 725-730.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Van der Wiel, K., Wanders, N., Selten, F. M., & Bierkens, M. F. P. (2019). Added value of large ensemble simulations for assessing extreme river discharge in a 2 C warmer world. Geophysical Research Letters, 46(4), 2093-2102.

Yang, Q., Zhang, H., Wang, G., Luo, S., Chen, D., Peng, W., & Shao, J. (2019). Dynamic runoff simulation in a changing environment: A data stream approach. Environmental Modelling & Software, 112, 157-165.

Lutz, A. F., Immerzeel, W. W., Shrestha, A. B., & Bierkens, M. F. P. (2014). Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. Nature Climate Change, 4(7), 587-592.

Zhu, Y., Liu, S., Yi, Y., Xie, F., Grünwald, R., Miao, W., ... & Singh, D. (2021). Overview of terrestrial water storage changes over the Indus River Basin based on GRACE/GRACE-FO solutions. *Science of The Total Environment*, *799*, 149366.

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Open Access
Discussions

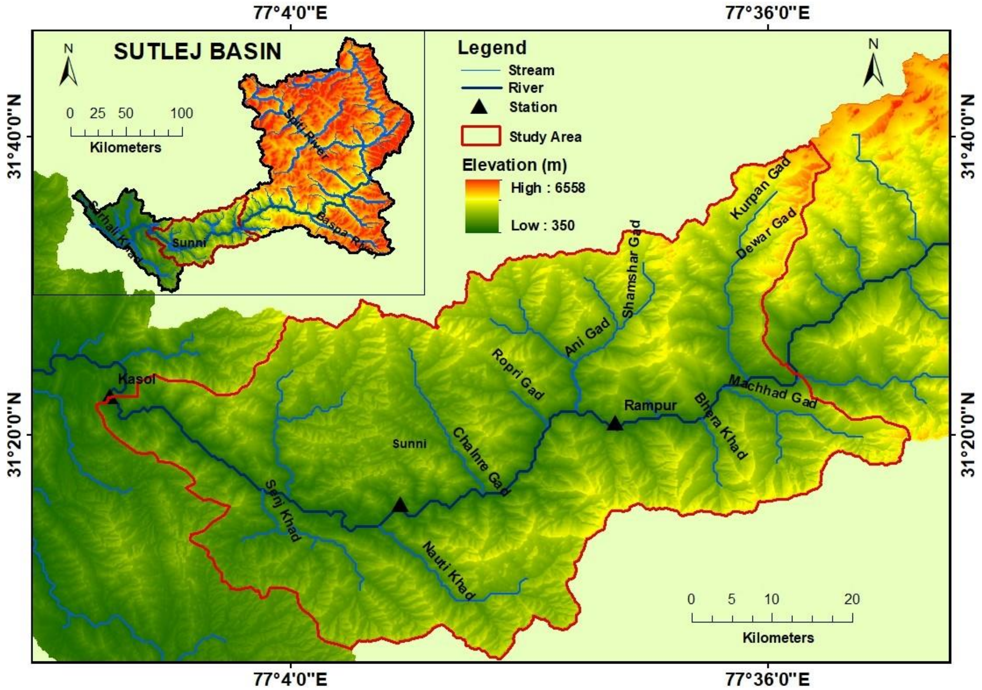


Figure 1: The location of the sub-catchment within Sutlej River Basin

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

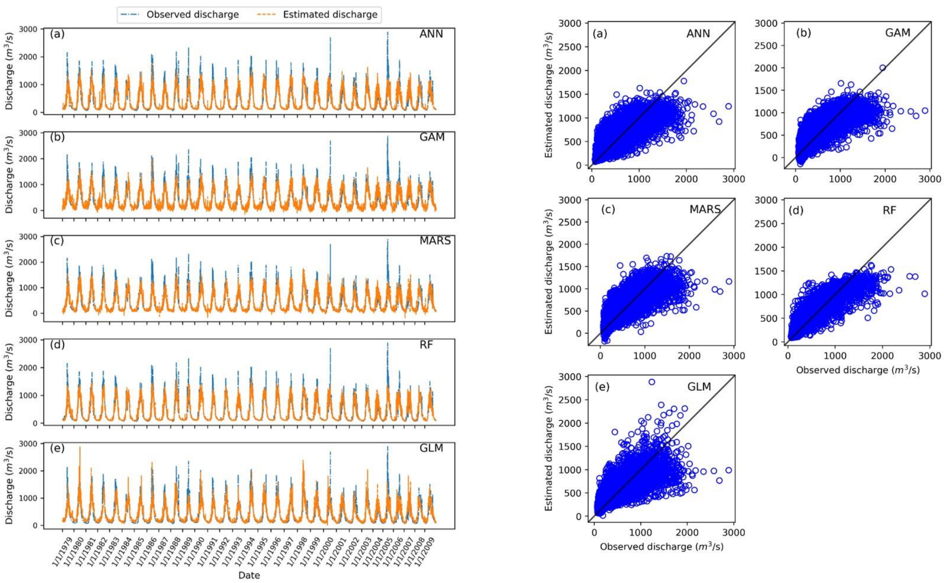


Figure 2: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_0

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Discussions

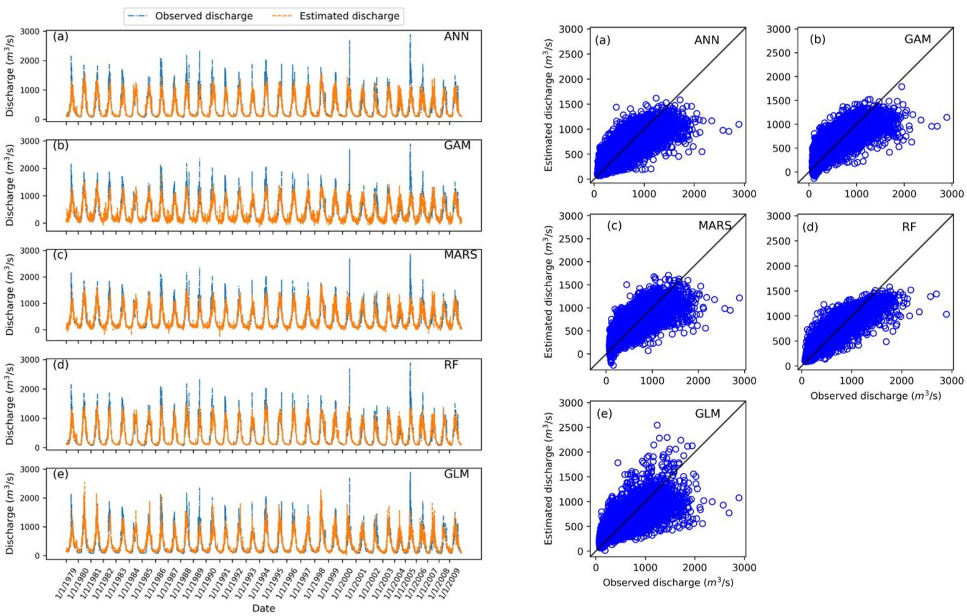


Figure 3: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_1

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Open Access
Discussions

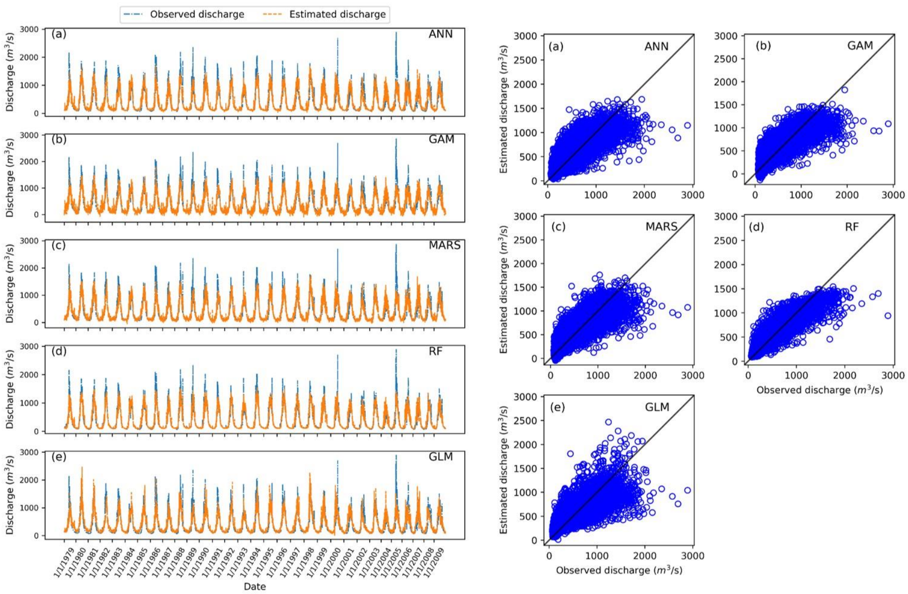


Figure 4: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_2

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

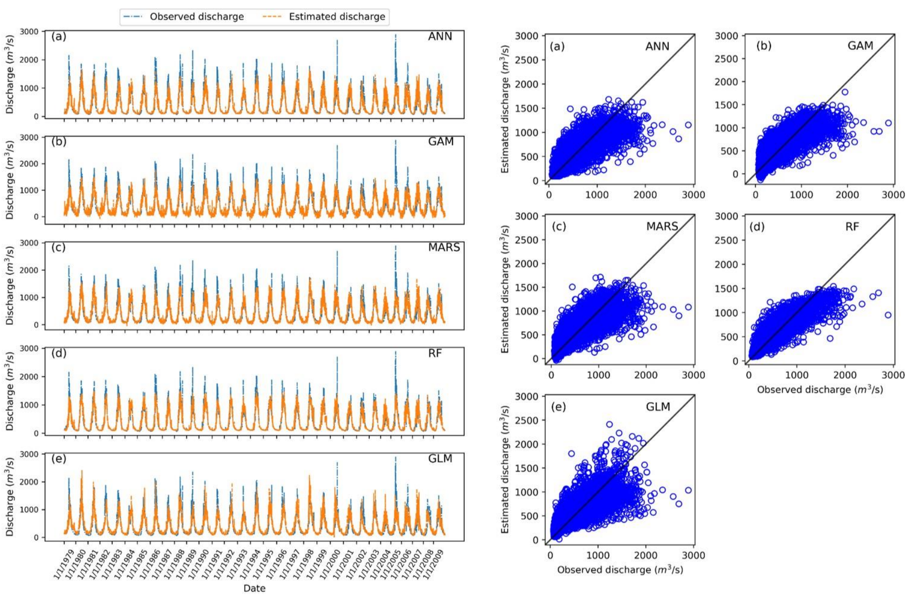
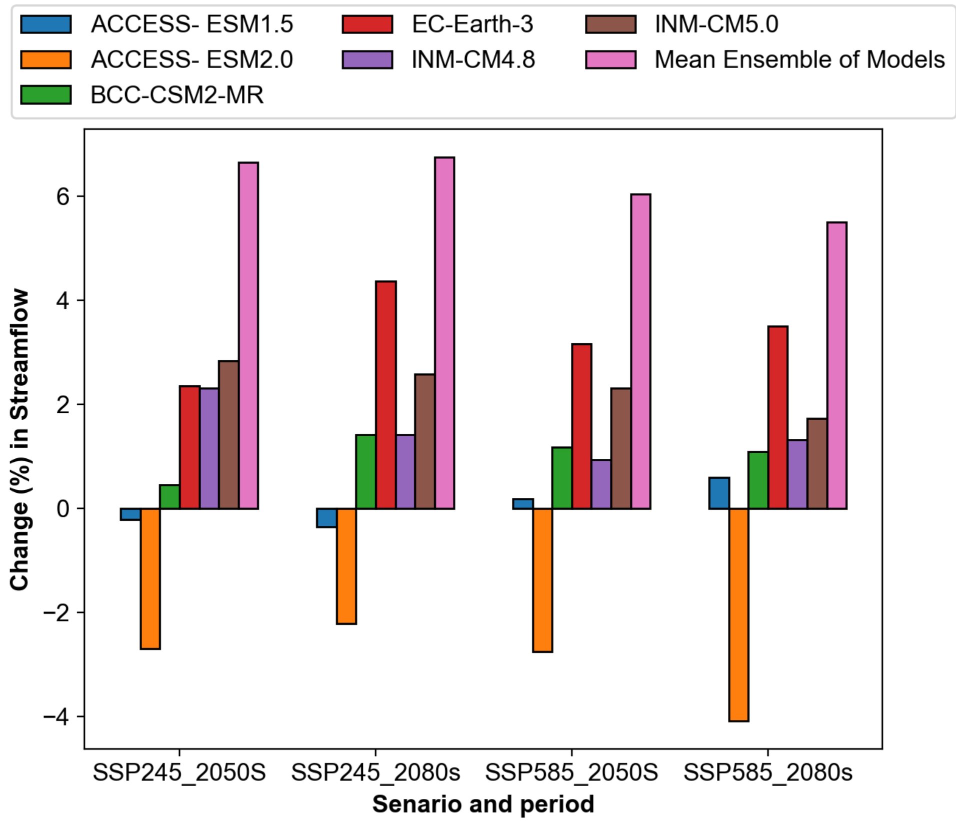


Figure 5: Comparison of observed and simulated streamflow for all five models (GLM, GAM, MARS, ANN and RF) under rainfall scenarios R_3

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Discussions



642

643 Figure 6: Predicted change in mean annual streamflow of the Sutlej River with respect to the

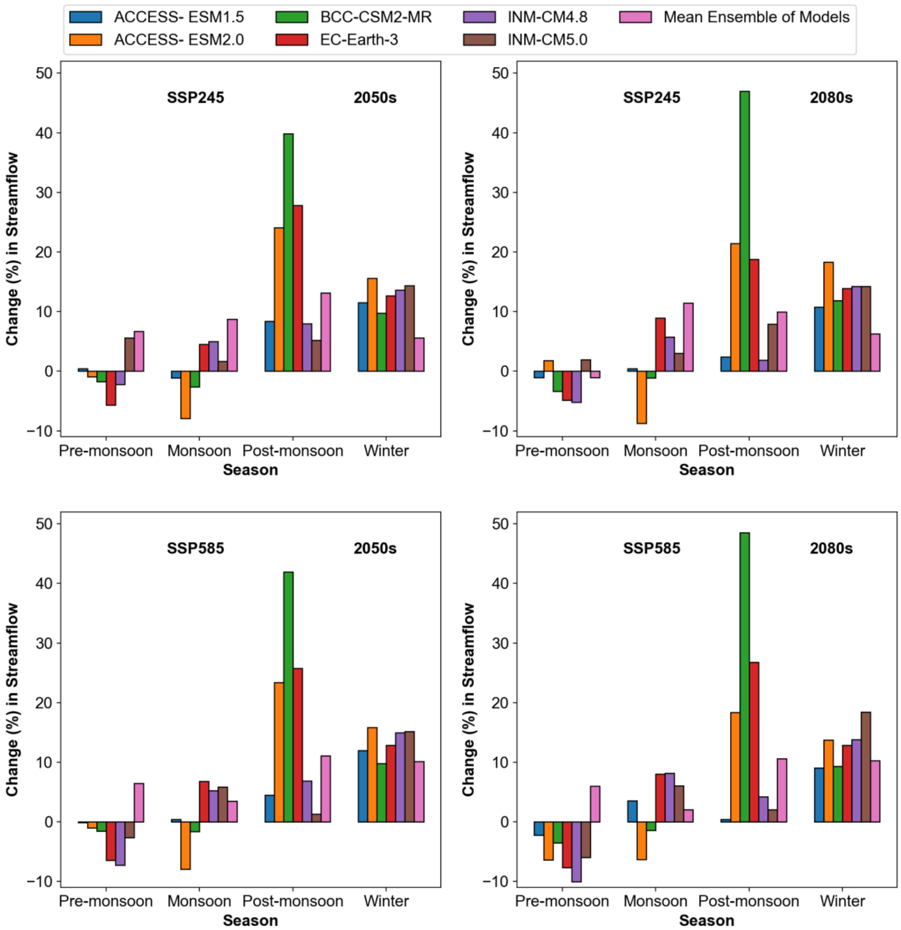
644 reference period (1979-2009) in 2050s and 2080s under SSP245 and SSP585 for different

645 GCMs

646

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Discussions



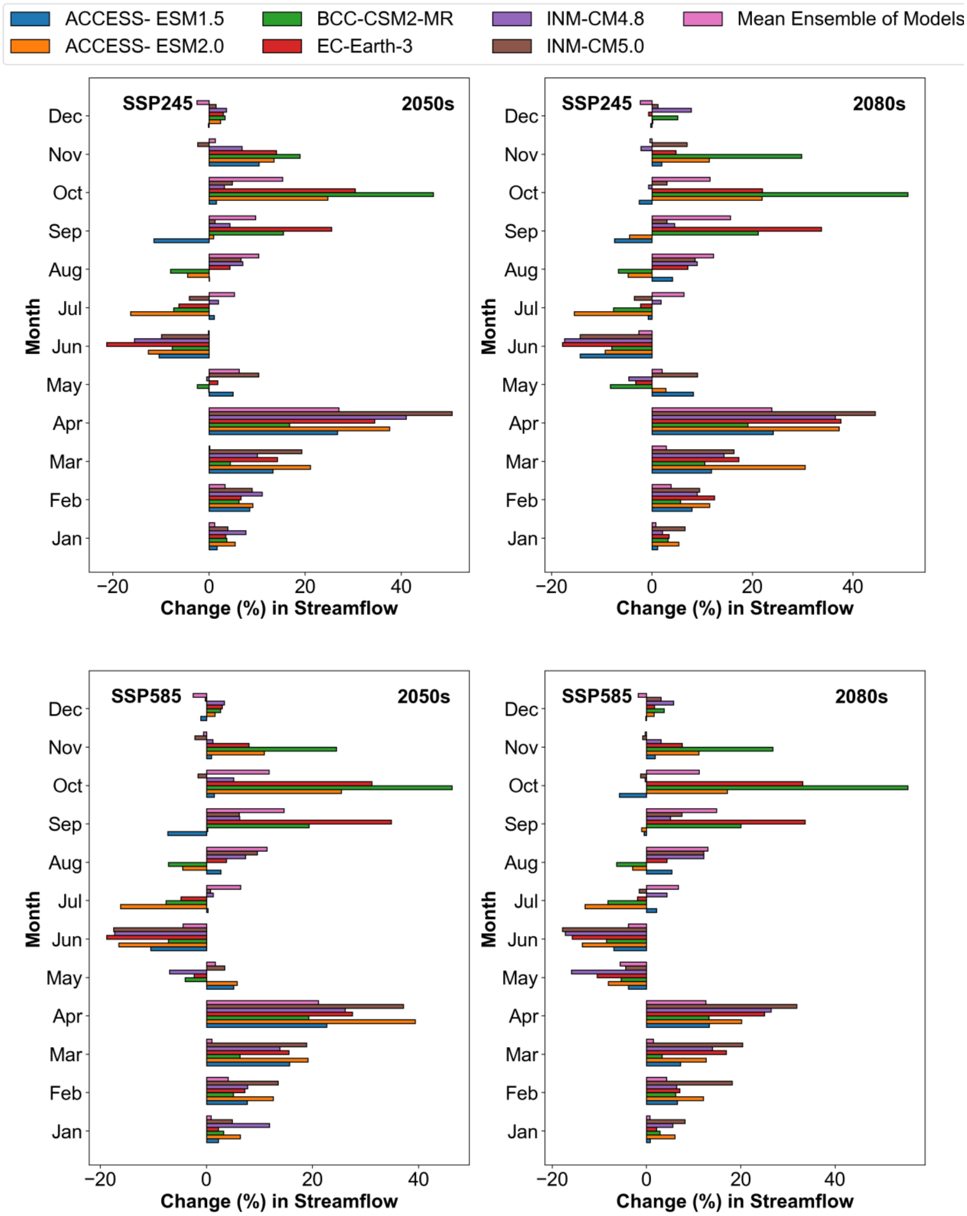
647

648 Figure 7: Predicted change in seasonal streamflow pattern of the Sutlej River with respect to
649 the reference period (1979-2009) in 2050s and 2080s under SSP245 and SSP585 for different
650 GCMs

651

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Discussions



652

653 Figure 8: Predicted change in monthly streamflow pattern of the Sutlej River with respect to

654 the reference period (1979-2009) in 2050s and 2080s under SSP245 and SSP585 for different

655 GCMs

656

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

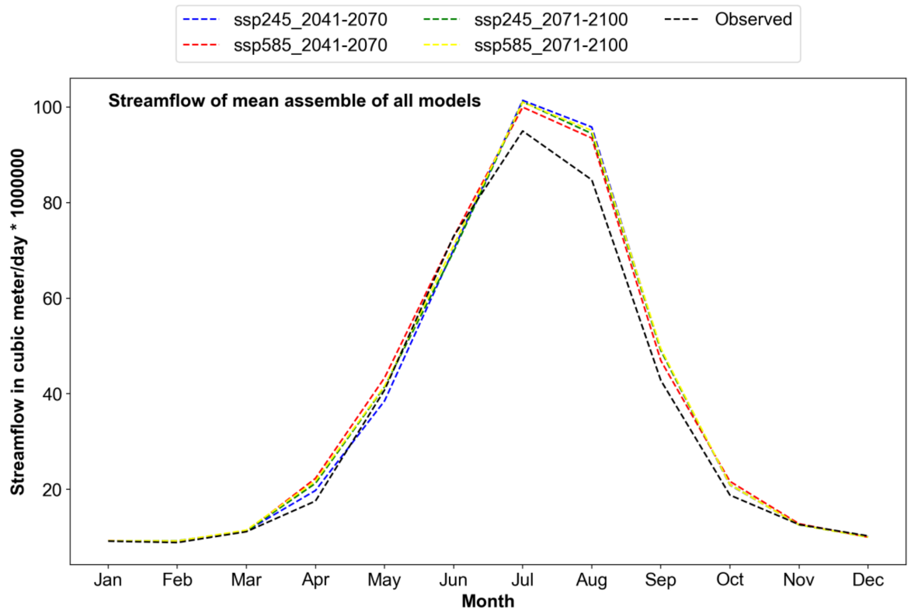


Figure 9: Comparison of monthly observed (1979-2009) and projected discharge of the multi-model ensembles for period 2050s and 2080s under SSP245 and SSP585 scenarios

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Table 1: Characteristics of the study catchment over the evaluation period of 1979–2009

Parameters	Details
Details of the sub-catchment	
Drainage area of the sub-catchment (km^2)	2457 km^2
Altitude	~500-5000 m
Slope	0-80°
Geology	Granite, Jutogh formation and Chail/Salkhala/Hemanta formation
Soil	Dystric cambisols, dystric regosols, and eutric fluviosols.
Streamflow measured at the outlet (Kasol) of the sub-catchment	
Average of annual streamflow	12469.43 m^3/s
Minimum streamflow	64.30 m^3/s
Maximum streamflow	2891.00 m^3/s
Standard deviation (SD) of annual streamflow	1750.70 m^3/s
Coefficient of variation (CV) of annual streamflow	0.14 m^3/s
Rainfall integrated over the sub-catchment	
Average of annual rainfall	1001.32mm
Average of monsoon rainfall (July-September)	403.08mm
Average of winter rainfall (December-March)	277.35mm
Temperature integrated over the sub-catchment	
Average annual maximum temperature (T_{max})	28.35°C
Average annual minimum temperature (T_{min})	13.98°C

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Table 2: The information on hyper parameters used for estimating model parameters

Model Name	Hyperparameter	Values
Random Forest (RF)	n_estimators,	value=500
	criterion,	value="squared_error"
	max_depth,	value=None
	min_samples_split,	value = 2
	min_samples_leaf,	value = 5
	min_weight_fraction_leaf,	value = 0.0
	max_features,	value = auto
	max_leaf_nodes,	value = None
	min_impurity_decrease,	value = 0.0
Generalized Linear Model (GLM)	endog,	value = 1D
	exog,	value = 1D
	family,	value = sm.families.Gaussian(sm.families.links.log())
	offset,	value = None
	exposure,	value = None
	freq_weights,	value = None
	var_weights,	value = None
	missing,	value = str
Artificial Neural Network (ANN)	build_fn,	value = build_regressor
	warm_start,	value = False
	random_state,	value = None
	optimizer,	value = rmsprop
	loss,	value = None
	metrics,	value = None
	batch_size,	value = 64
	validation_batch_size,	value = None
	verbose,	value = 1
	callbacks,	value = None
	validation_split,	value = 0.0
	shuffle,	value = True
	run_eagerly,	value = False
	epochs,	value = 500
Multivariate Adaptive Regression Splines (MARS)	max_terms,	value = 20
	max_degree ,	value = 3
	allow_missing,	value = False
	penalty,	value = 3.0
	endspan_alpha,	value = 0.005
	endspan,	value = -1
	minspan_alpha,	value = 0.005
	minspan,	value = -1
	thresh ,	value = 0.001
	zero_tol,	value = 1e-12
	min_search_points,	value = 100
	check_every,	value = -1
	allow_linear,	value = True

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

	use_fast, fast_K, fast_h, smooth, enable_pruning, feature_importance_typ e, feature_importance_typ e,	value = False value = 5 value = 1 value = False value = True value = None value = 0
Generalized Additive Model (GAM)	formula, family, data, weights, subset , na.action,offset, method, optimizer, control, scale, select, knots, sp, min.sp, H, gamma, fit, paraPen, G, drop.unused.levels, drop.intercept, discrete,	value = None value = gaussian() value = list() value = Null value = Null value = Null value = "GCV.Cp" value = c("outer","newton") value = list(), value = 0 value = False value = Null value = Null value = Null value – Null, value = 1 value = True value = Null value = Null value = True value = Null value = False

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Table 3: Summary of model performance in simulating streamflow at Kasol

Model	Model inputs	Training			Testing		
		R^2	RSR	MAE	R^2	RSR	MAE
RF	R_0 (rainfall on the same day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.88	0.34	77.5	0.76	0.49	111.9
	R_1 (rainfall lagged by one day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.89	0.33	74.4	0.77	0.48	109.2
	R_2 (rainfall lagged by two day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.90	0.32	73.1	0.78	0.47	107.5
	R_3 (rainfall lagged by three day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.90	0.32	72.0	0.78	0.47	106.6
GLM	R_0 (rainfall on the same day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.69	0.55	140.6	0.69	0.56	141.3
	R_1 (rainfall lagged by one day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.70	0.54	137.3	0.70	0.55	137.5
	R_2 (rainfall lagged by two day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.71	0.54	136.0	0.71	0.54	135.4
	R_3 (rainfall lagged by three day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.71	0.54	134.8	0.71	0.54	134.5
ANN	R_0 (rainfall on the same day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.73	0.52	123.3	0.73	0.52	123.0
	R_1 (rainfall lagged by one day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.74	0.51	119.3	0.74	0.51	118.9
	R_2 (rainfall lagged by two day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.75	0.50	119.3	0.75	0.50	118.2
	R_3 (rainfall lagged by three day), T_{max}, T_{min}, relative humidity, solar radiation and	0.75	0.50	117.7	0.75	0.50	117.4

https://doi.org/10.5194/hess-2022-339
Preprint. Discussion started: 4 October 2022

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

	wind speed						
MARS	R_0 (rainfall on the same day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.77	0.48	118.9	0.79	0.45	112.8
	R_1 (rainfall lagged by one day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.75	0.50	126.1	0.74	0.51	126.7
	R_2 (rainfall lagged by two day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.75	0.50	124.8	0.75	0.50	125.0
	R_3 (rainfall lagged by three day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.75	0.50	125.1	0.75	0.50	125.5
GAM	R_0 (rainfall on the same day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.72	0.53	139.0	0.72	0.53	139.4
	R_1 (rainfall lagged by one day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.74	0.51	136.2	0.73	0.52	136.9
	R_2 (rainfall lagged by two day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.74	0.51	134.7	0.74	0.51	135.2
	R_3 (rainfall lagged by three day), T_{max}, T_{min}, relative humidity, solar radiation and wind speed	0.74	0.50	133.8	0.74	0.51	134.8