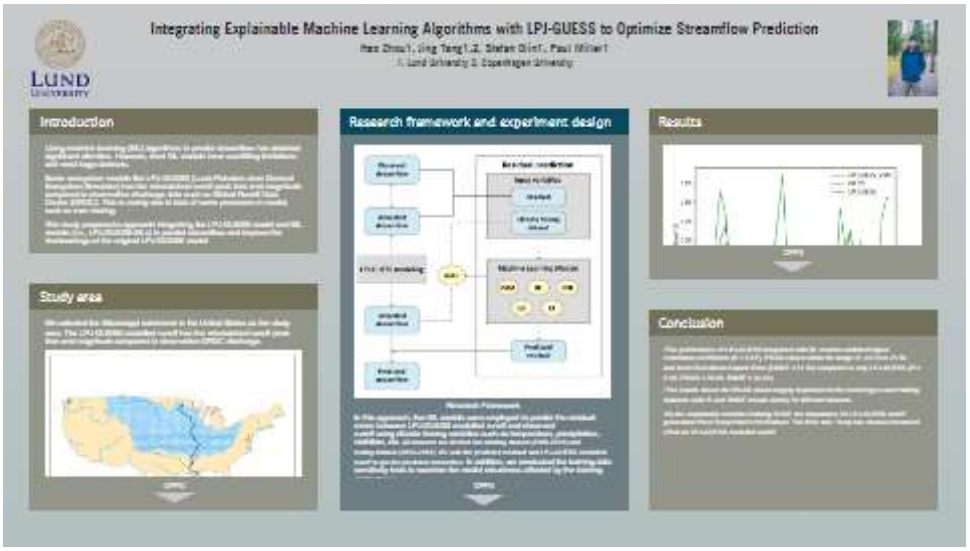
Integrating Explainable Machine Learning Algorithms with LPJ-GUESS to Optimize Streamflow Prediction

Hao Zhou[1]

[1]Affiliation not available

April 15, 2024

Integrating Explainable Machine Learning Algorithms with LPJ-GUESS to Optimize Streamflow Prediction



Hao Zhou1, Jing Tang1,2, Stefan Olin1, Paul Miller1

1. Lund University 2. Copenhagen University

PRESENTED AT:

AGU23 WIDE. OPEN. SCIENCE.

INTRODUCTION

Using machine learning (ML) algorithms to predict streamflow has obtained significant attention. However, most ML models have overfitting limitations and need large datasets.

Some ecosystem models like LPJ-GUESS (Lund-Potsdam-Jena General Ecosystem Simulator) has the mismatched runoff peak time and magnitude compared to observation discharge data such as Global Runoff Data Centre (GRDC). This is mainly due to lack of some processes in model, such as river routing.

This study proposed a approach integrating the LPJ-GUESS model and ML models (i.e., LPJ-GUESS-MLs) to predict streamflow and improve the shortcomings of the original LPJ-GUESS model.

STUDY AREA

We selected the Mississippi catchment in the United States as the study area. The LPJ-GUESS modelled runoff has the mismatched runoff peak time and magnitude compared to observation GRDC discharge.

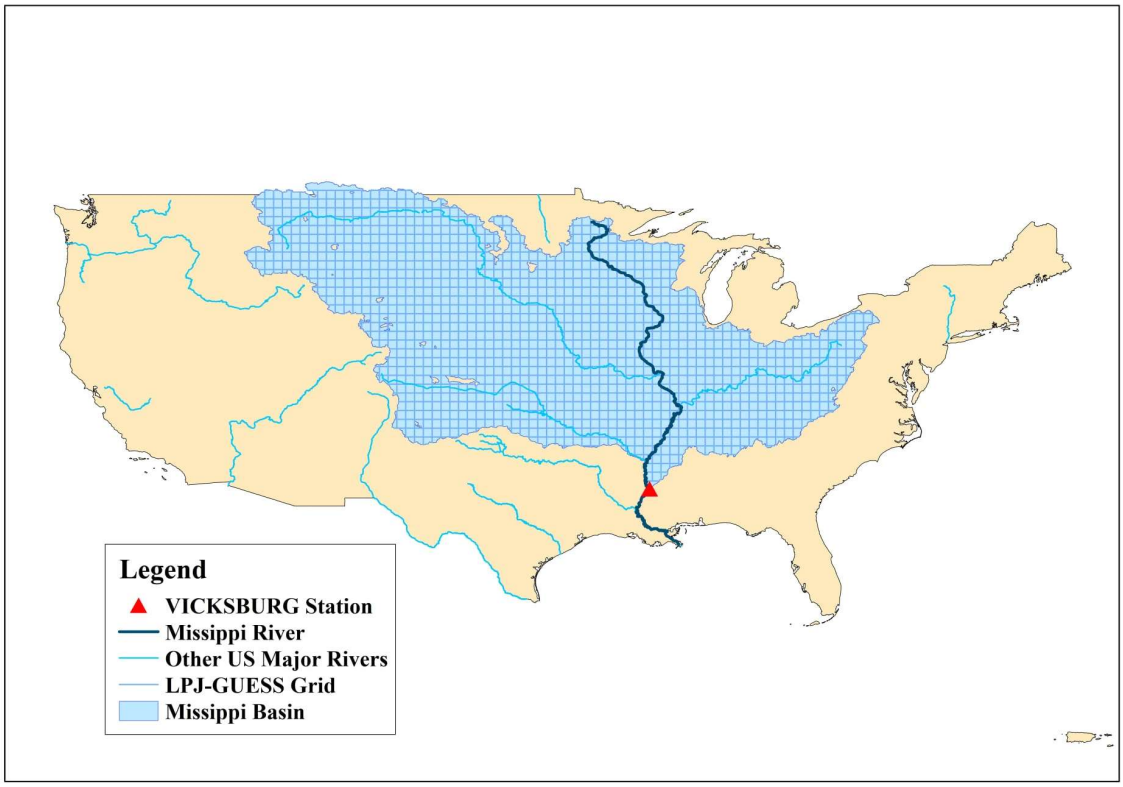


Fig1. Study area map

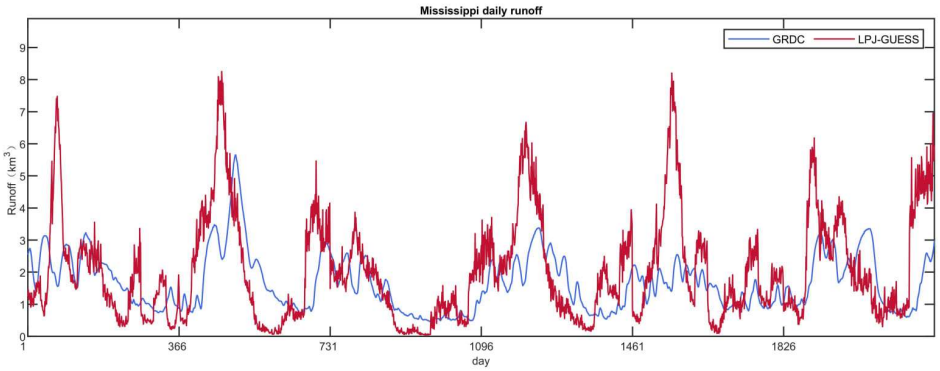


Fig2. Daily runoff of LPJ-GUESS and GRDC

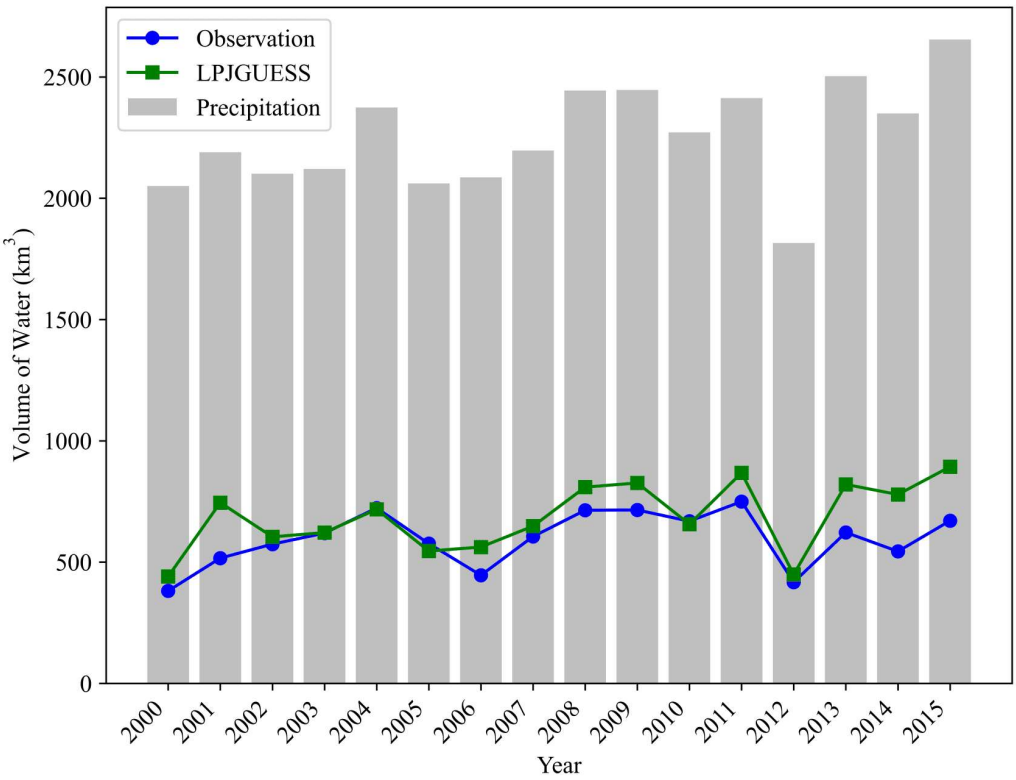


Fig3. LPJ-GUESS modelled runoff , GRDC runoff and input precipiation

RESEARCH FRAMEWORK AND EXPERIMENT DESIGN

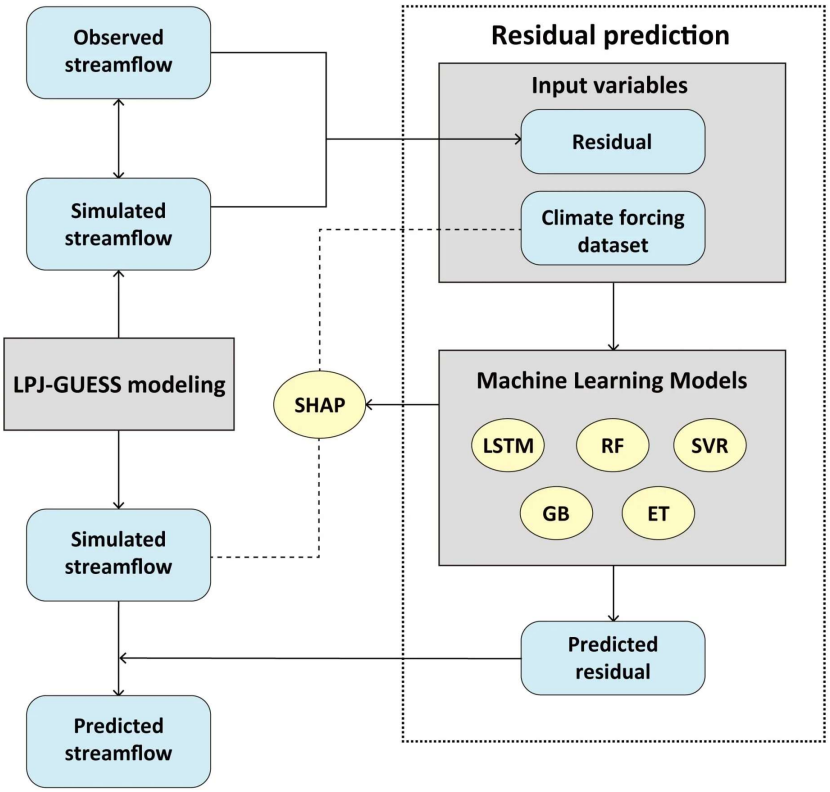


Fig 4. Research Framework

In this approach, five ML models were employed to predict the residual errors between LPJ-GUESS modelled runoff and observed runoff using climate forcing variables such as temperature, precipitation, radiation, etc. All datasets are divided into training dataset (2000-2012) and testing dataset (2013-2015). We add the predicted residual and LPJ-GUESS modelled runoff to get the predicted streamflow. In addition, we conducted the training data sensitivity tests to examine the model robustness affected by the training data's timing.

We set up and compared three different model configurations: LPJ-GUESS integrated with ML models, only LPJ-GUESS, and only ML models.

Furthermore, the explainable ML method SHAP was used to identify the driving factors and their interacted effect of LPJ-GUESS modelled runoff.

Table 1. Daily variables for three kinds of models:

Relative Humidity	*%*	LPJ-GUESS-ML & ML-only
Incoming radiation	*J/day*	
Mean wind speed at 10m height	*m/s*	
Mean air temperature	*deg C*	
Precipitation	*Km^3/day*	
Residual error (Observed – simulated (LPJ-GUESS) inflow)	*Km^3/day*	LPJ-GUESS-ML
Observed inflow	*Km^3/day*	ML-only

PS: Precipitation(Prec), Temperature(Temp), Radiation(Rad), Wind speed(U10), Relative humidity(Relhum)

Evaluation metrics：

•Pearson correlation coefficient (R)

•Root Mean Squared Error (RMSE)

•Percent Bias (PBIAS)

•Mean Absolute Error (MAE)

•Percentage Absolute Error (PAE)

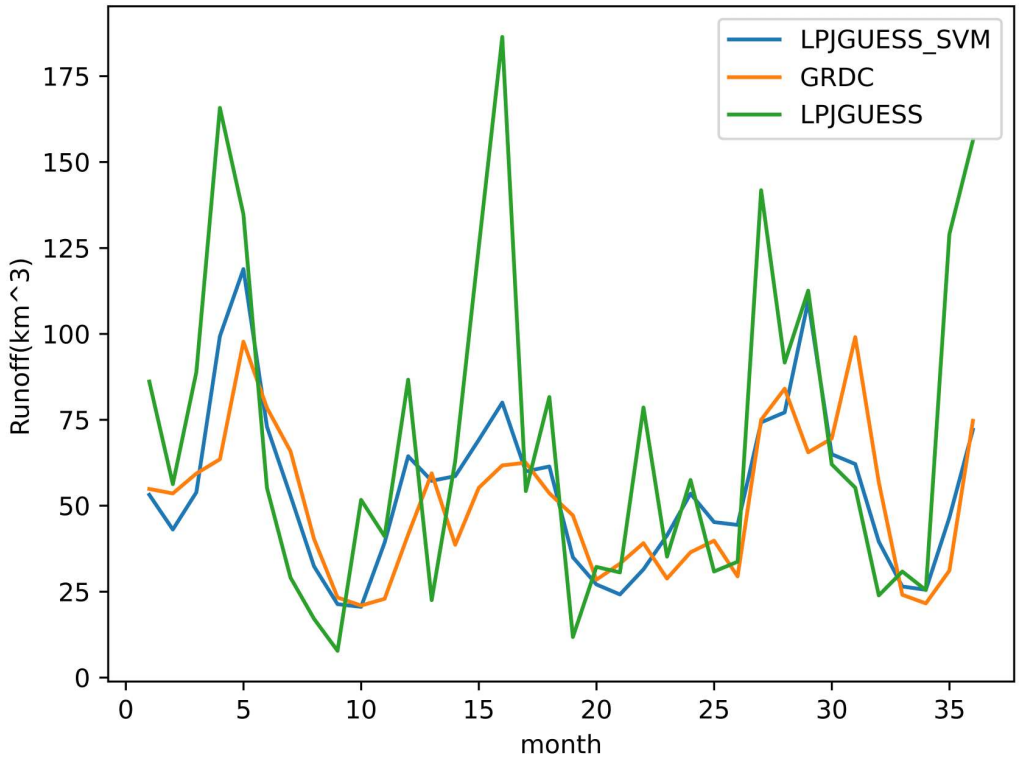


Fig 5. Monthly runoff of LPJ-GUESS-SVM (R: 0.78, RMSE: 14.53 km^3), LPJ-GUESS (R: 0.48, RMSE: 44.84 km^3) and GRDC

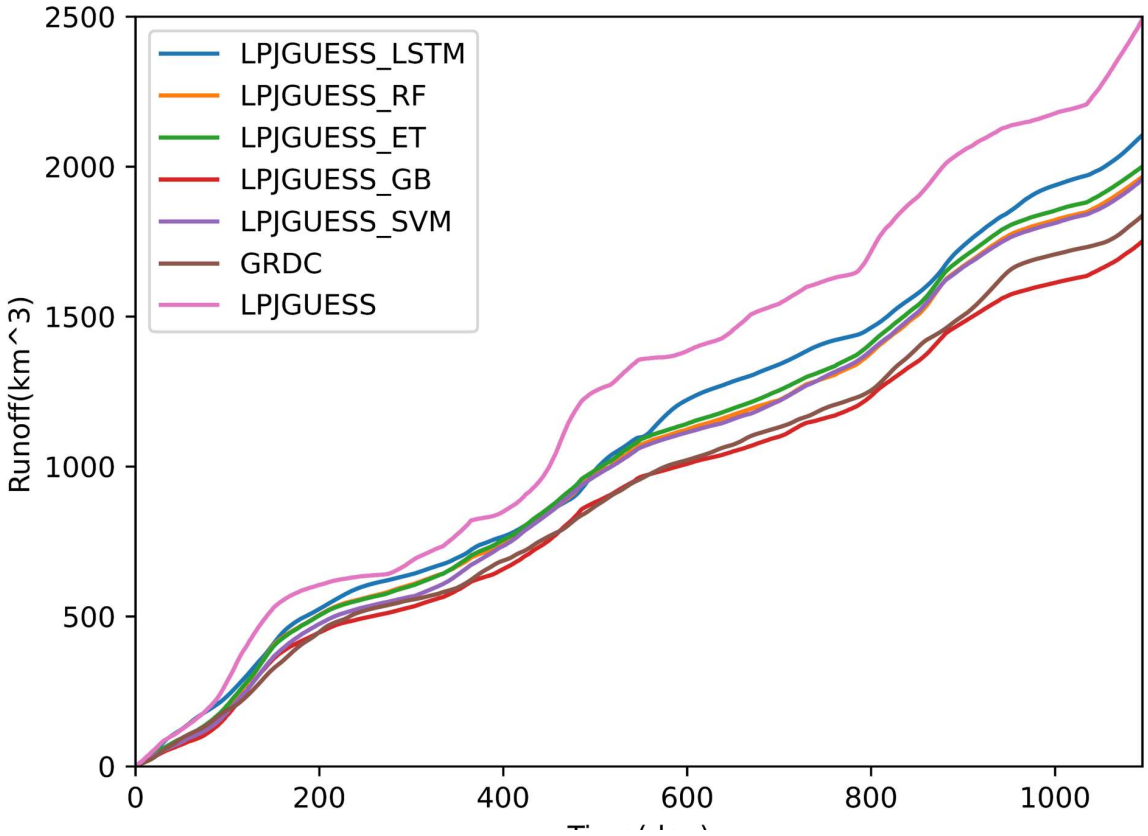


Fig 6. Cumulative daily runoff

•The performance of LPJ-GUESS integrated with ML models exhibited higher correlation coefficients (R > 0.67), PBIAS values within the range of -16.23 to 21.54, and lower Root Mean Square Error (RMSE < 21.54) compared to only LPJ-GUESS (R = 0.48, PBIAS = 35.66, RMSE = 44.84).

Training period	Predicting period	R	PBIAS	RMSE	MAE	PAE
2000-2012	2013-2015	0.77	6.62	15.68	11.72	0.75
2000-2009& 2013-2015	2010-2012	0.78	-15.10	19.97	13.92	0.72
2000-2005& 2010-2015	2007-2009	0.74	-14.97	18.12	13.15	0.77
2000-2002& 2006-2015	2003-2005	0.74	-14.63	17.38	12.54	0.78
2003-2015	2000-2002	0.71	19.75	16.11	13.48	0.59

Fig 7. Metrics range for sensitivity test

•The sensitivity test results shows the PBIAS values largely depended on the model types and training datasets while R and RMSE remain steady for different datasets.

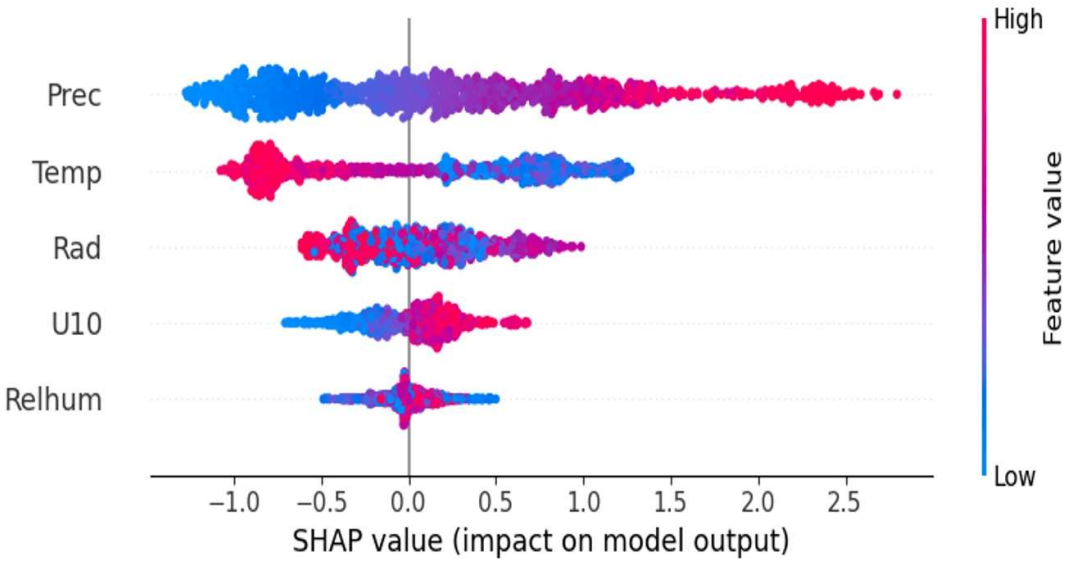
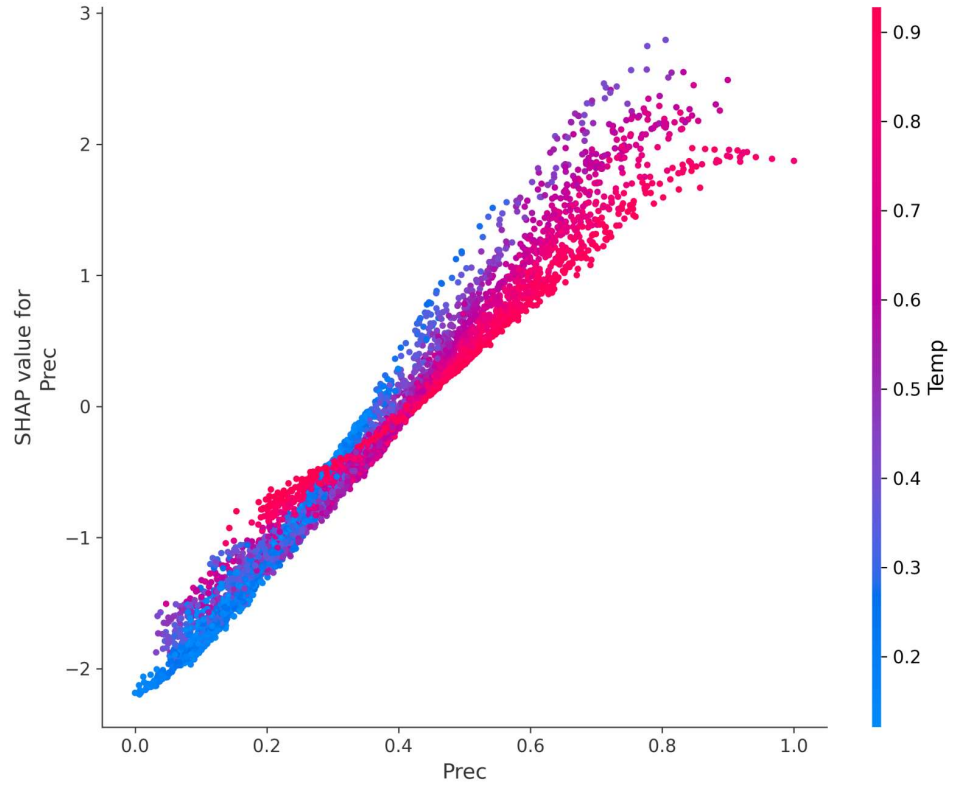
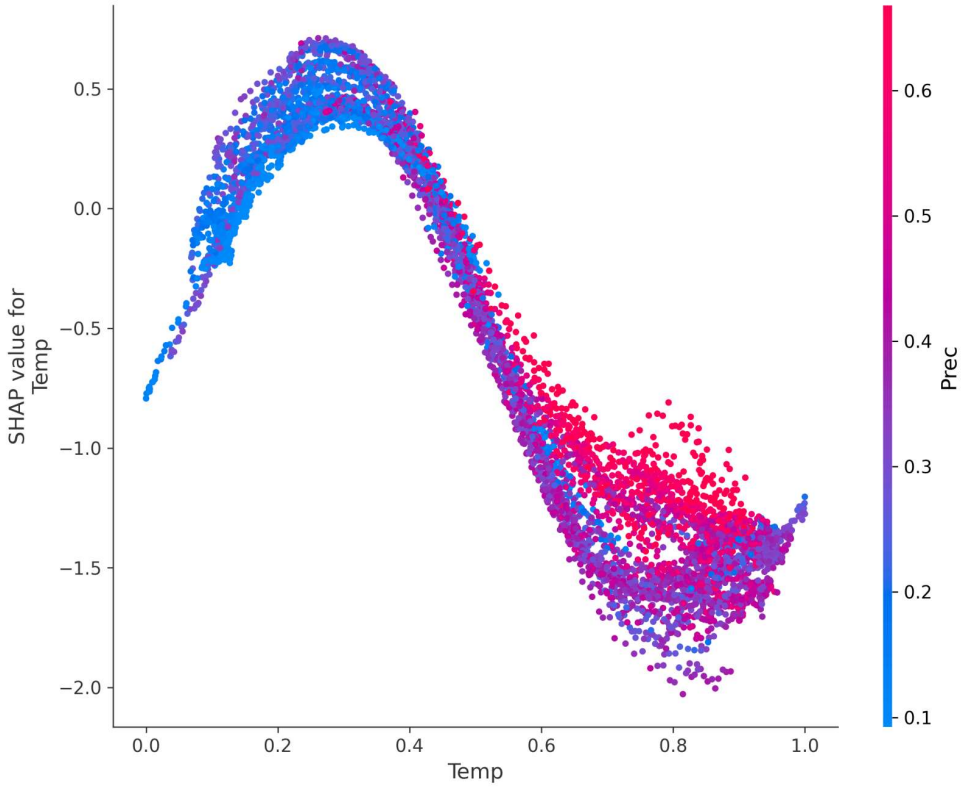


Fig 8. SHAP value summary plot


SHAP value for
Prec
Prec
Temp
3
2
1
0
−1
−2
0.0
0.2
0.4
0.6
0.8
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2

(a)


SHAP value for
Temp
Temp
Prec
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
0.6
0.5
0.4
0.3
0.2
0.1

(b)

Fig 9. Interacted effect between prec and temp

•By the explainable machine learning SHAP, the importance for LPJ-GUESS runoff generation:Prec>Temp>Rad>U10>Relhum. The Prec and Temp has obvious interacted effect on LPJ-GUESS modelled runoff.

CONCLUSION

In this study, we proposed an approach that integrates LPJ-GUESS with ML models to produce LPJ-GUESS-ML models that aim to improve catchment streamflow simulations. Tests to examine the sensitivity of the approach to training datasets show the robustness of model.

Furthermore, identifying the driving factors of the LPJ-GUESS modelled runoff using SHAP and MLs could find the interacted effect of climate variables on LPJ-GUESS modelled runoff.

AUTHOR INFO

Hao Zhou, PhD-student

Email: hao.zhou@nateko.lu.se

Department of Physical Geography and Ecosystem Science

Lund University

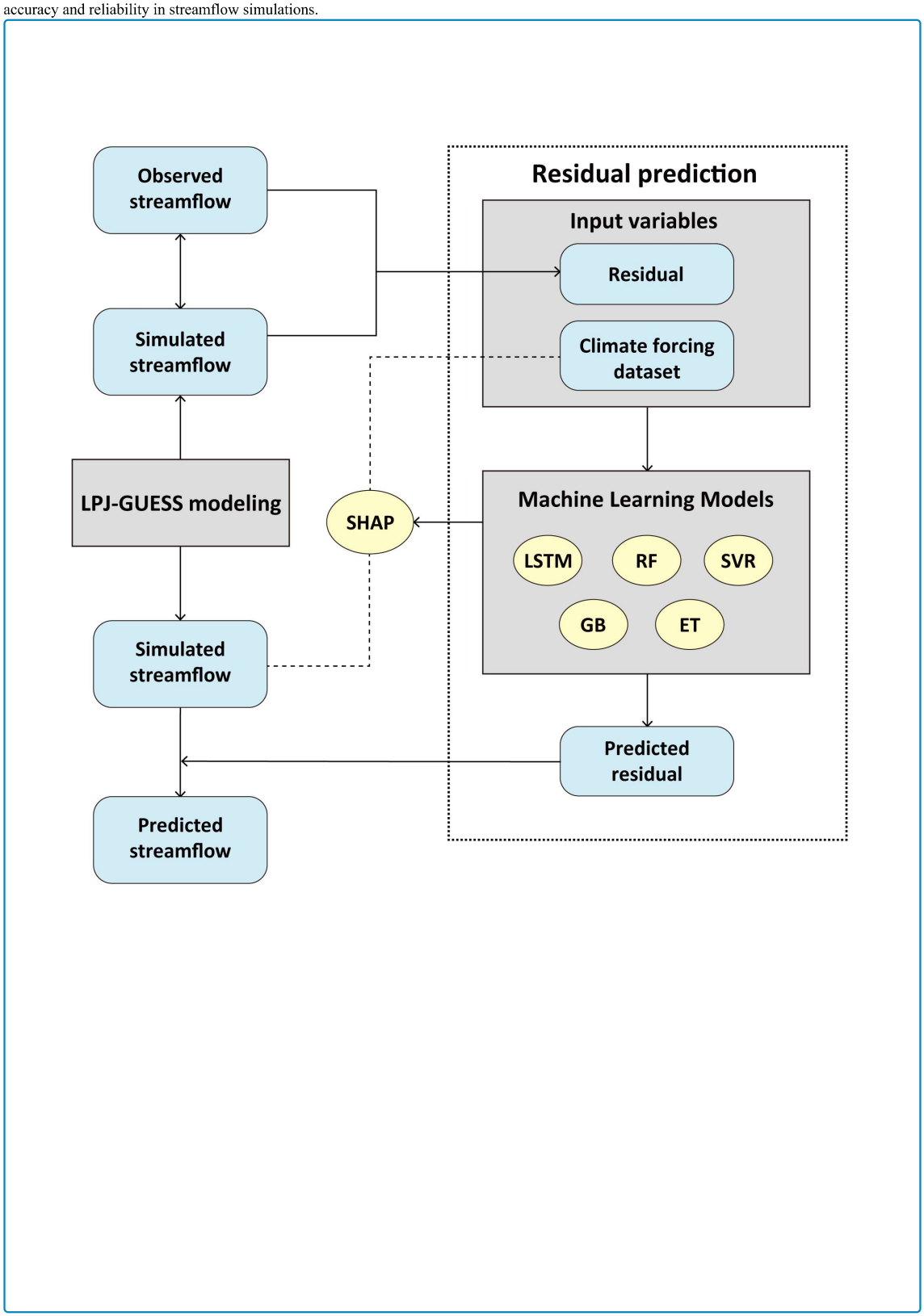
Sweden

TRANSCRIPT

ABSTRACT

Utilizing machine learning (ML) algorithms to help physically based models to predict streamflow has garnered significant attention as a potential improvement solution. However, most ML models have overfitting limitations and need large datasets. To address these shortcomings, we proposed an approach that integrates the dynamic global vegetation model LPJ-GUESS (Lund-Potsdam-Jena General Ecosystem Simulator) with multiple ML models to improve streamflow simulations. In this approach, ML models were employed to predict the residual errors of LPJ-GUESS using physically related model climate forcing variables such as temperature, precipitation, radiation, etc. We selected the Mississippi catchment in the United States as the study area for streamflow prediction, employing three different configurations: LPJ-GUESS integrated with ML models, only LPJ-GUESS, and only ML models. Evaluating with in-situ streamflow data of Vicksburg station from Global Runoff Data Centre (GRDC), the results demonstrate the superior performance of LPJ-GUESS integrated with ML models, exhibiting higher correlation coefficients ($R > 0.67$), PBIAS values within the range of -16.23 to 21.54, and lower Root Mean Square Error ($RMSE < 21.54$) compared to only LPJ-GUESS ($R = 0.48$, $PBIAS = 35.66$, $RMSE = 44.84$). In terms of the R and RMSE, LPJ-GUESS integrated with the ML models have an overall better value than the corresponding only ML models. Furthermore, a training data sensitivity experiment showes that the PBIAS values largely depends on the model types and training datasets while R and RMSE remain steady for different datasets. The analysis of LPJ-GUESS runoff generation driving factors by explainable machine learning SHAP shows the dominant importance of precipitation and temperature. Notably, LPJ-GUESS integrated with ML models successfully captures residual errors and effectively reduces inherent uncertainties, thus surpassing the performance of solely ML-based methods. Our study highlights the promising potential of integrating ML algorithms with LPJ-GUESS for streamflow prediction. This approach not only overcomes existing limitations but also offers a more robust representation of physical constraints, thereby fostering improved

accuracy and reliability in streamflow simulations.



(https://agu.confex.com/data/abstract/agu/fm23/2/2/Paper_1349022_abstract_1178201_0.jpg)