Supplementary Material for

Multi-Objective Hyperparameter Optimization Improves the Interpretability of LSTM Rainfall-Runoff Models

Qiuyang Tan[1,2,4], Jianming Shen[2,4], Youqing Wang[3], Moyuan Yang[2,4,*], Lin Zhu[1], Juan Zhang[2,4], Yang Liu[2], Zeyuan Chen[2], Chao Zhai[2], and Yun Zhu[3]

[1] College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China

[2] Beijing Water Science and Technology Institute, Beijing 100048, China

[3] Beijing Water Resources Dispatching and Management Affairs Center, Beijing 100048, China

[4] Beijing Key Laboratory of Urban Groundwater Monitoring, Early Warning and Safety Control, Beijing 100048, China

[*] Correspondence: ymoyuan@163.com

The following are included as supporting information for this paper:

Number of texts: 3

Number of tables: 1

Number of figures:7

Text S1. Calculation method of the baseflow index (BFI)

In this study, a digital filtering method was applied to separate baseflow from total runoff[S1]. The baseflow index (BFI) represents the contribution ratio of baseflow to total runoff. The principle of digital filtering is based on the different hydrological response times of various runoff components. Specifically, surface runoff generally has a higher frequency and shorter response time, whereas baseflow signals exhibit lower frequency and longer response times. This method is simple and effective, and its performance has been extensively evaluated in previous studies[S2-4]. The calculation formula for baseflow is as follows:

$$b_k = ab_{k-1} + \frac{(1+a)}{2}(y_k - y_{k-1}) \quad \text{(S1)}$$

where b_k represents the baseflow at time k, y_k denotes the total streamflow at time k, and a is the filter parameter. Previous studies based on catchments in Australia[S2], the United States[S3], and Türkiye[S4] have demonstrated that a=0.925 can achieve reliable baseflow separation. Therefore, a was set to 0.925 in this study.

Text S2. Calculation method of Autocorrelation Function (ACF)

The Autocorrelation Function (ACF) reflects the relationship between observations in a time series and their past values, and it is an effective tool for time series analysis[S5, S6]. The equation for calculating the autocorrelation function is as follows:

$$R_k = \frac{\frac{1}{N-k}\sum_{i=1}^{N-k}(X_i-\bar{X})(X_{i-k}-\bar{X})}{\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2} \quad \text{(S2)}$$

where R_k represents the autocorrelation coefficient at lag time k, X_i denotes the observed value on day i, N is the total number of observations, k represents the lag time, and $\bar{X}$ is the mean value of the entire time series.

Text S3. Calculation of HV and IGD

(1) Hypervolume (HV):

HV measures the volume of the region enclosed by the solution set generated by the algorithm and a reference point in the objective space[7]. HV evaluates both the convergence and diversity of the obtained solution set. It is calculated as follows:

$$HV(P) = \lambda(\cup_{p_i \in P}\{h | C(p_i) < h < C(p_{ref})\}) \quad \text{(S3)}$$

$$f_i^{norm} = \frac{f_i - f_i^{min}}{f_i^{max} - f_i^{min}} \quad \text{(S4)}$$

where P denotes the obtained Pareto solution set, p_{ref} represents the reference point. In this study, the reference point was defined as 1.1 times the maximum normalized objective values derived from equation S4[8]. λ is the Lebesgue measure. A higher HV value indicates a Pareto front with better convergence and diversity.

(2) Inverted Generational Distance (IGD):

IGD measures the average Euclidean distance from uniformly distributed points on the entire Pareto front (PF) to their nearest solutions in the obtained solution set. IGD evaluates the convergence performance and coverage of the obtained PF. It is calculated as follows:

$$IGD(P, P^{true}) = \frac{\sum_{i=1}^{|P^{true}|} d(P_i^{true}, P)}{|P^{true}|} \quad \text{(S5)}$$

where P^{true} represents a set of points uniformly distributed along the true Pareto front in the objective space, and P denotes the approximate PF obtained by the algorithm. $d(P_i^{true}, P)$ represents the minimum Euclidean distance between P_i^{true} and the obtained solution set P. A lower IGD value indicates that the algorithm has achieved better performance.

Table S1. Correlations among LSTM performance metrics ($RMSE_{all}$, $RMSE_{high}$, $RMSE_{low}$, Dv, NSE, and KGE) at different lead times (1d, 3d, and 7d).

Lead time (day)	lead time	$RMSE_{all}$	$RMSE_{high}$	$RMSE_{low}$	Dv	NSE
1	$RMSE_{high}$	0.44*				
	$RMSE_{low}$	0.48*	-0.57*			
	Dv	-0.30*	-0.51*	0.17		
	NSE	-1.00*	-0.44*	-0.48*	0.31*	
	KGE	-0.02	-0.48*	0.46*	0.04	0.03
3	$RMSE_{high}$	-0.12				
	$RMSE_{low}$	0.61*	-0.85*			
	Dv	-0.20	-0.29*	0.10		
	NSE	-1.00*	0.12	-0.61*	0.20	
	KGE	-0.05	-0.69*	0.50*	0.24	0.04
7	$RMSE_{high}$	0.34*				
	$RMSE_{low}$	0.08	-0.91*			
	Dv	-0.04	-0.03	-0.06		
	NSE	-1.00*	-0.34*	-0.08	0.03	
	KGE	-0.36*	-0.80*	0.68*	-0.02	0.36*

* and gray-shaded cells indicate statistically significant correlations ($p < 0.05$).

Figure S1. HV convergence curves of the MOPSO-LSTM model with lead times of (a) 1 day, (b) 3 days, and (c) 7 days.

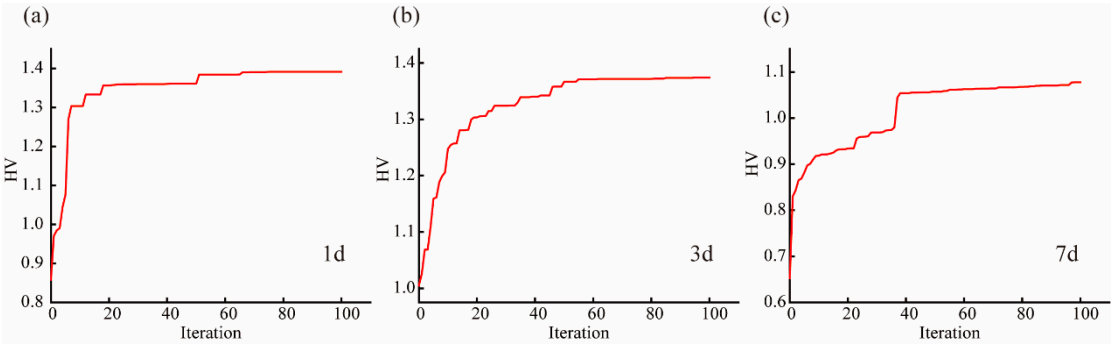


Figure S2. IGD convergence curves of the MOPSO-LSTM model with lead times of (a) 1 day, (b) 3 days, and (c) 7 days.

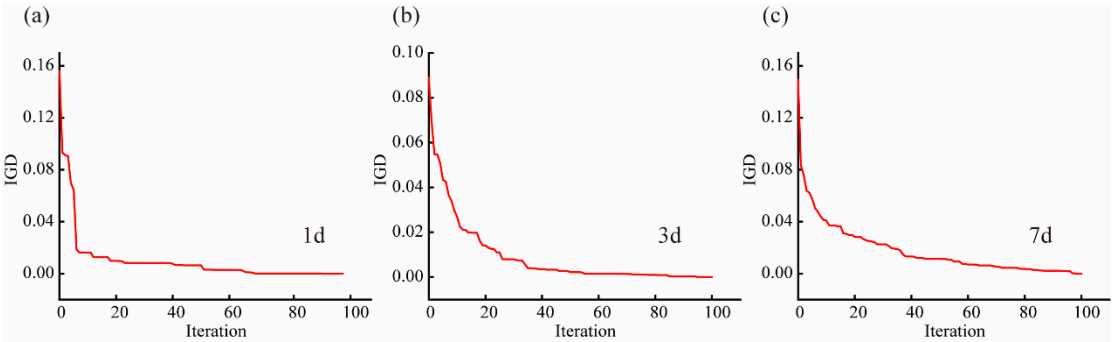


Figure S3. Computational time of each generation during the 100-generation MOPSO-LSTM training process with forecast lead times of (a) 1 day, (b) 3 days, and (c) 7 days.

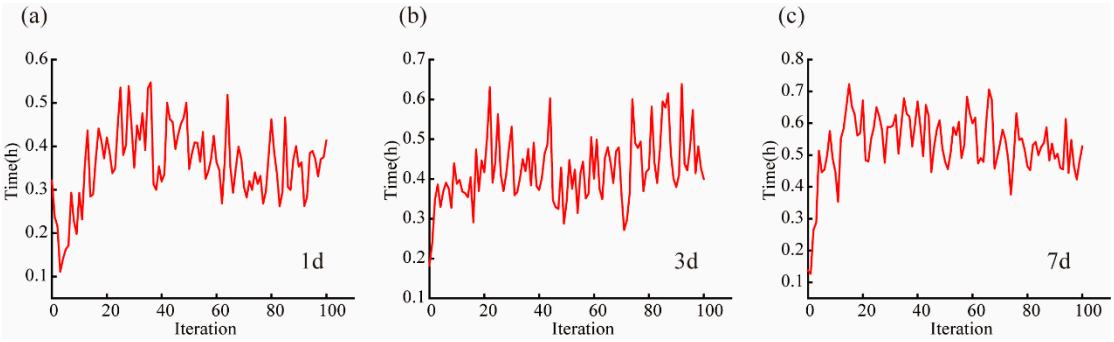


Figure S4. Partial dependence plots of LSTM hyperparameters for $RMSE_{all}$ under different lead times. (a) 1d, (b) 3d, (c) 7d. The X-axis represents hyperparameter values, and the Y-axis represents the simulated performance metric under averaged effects of other hyperparameters.

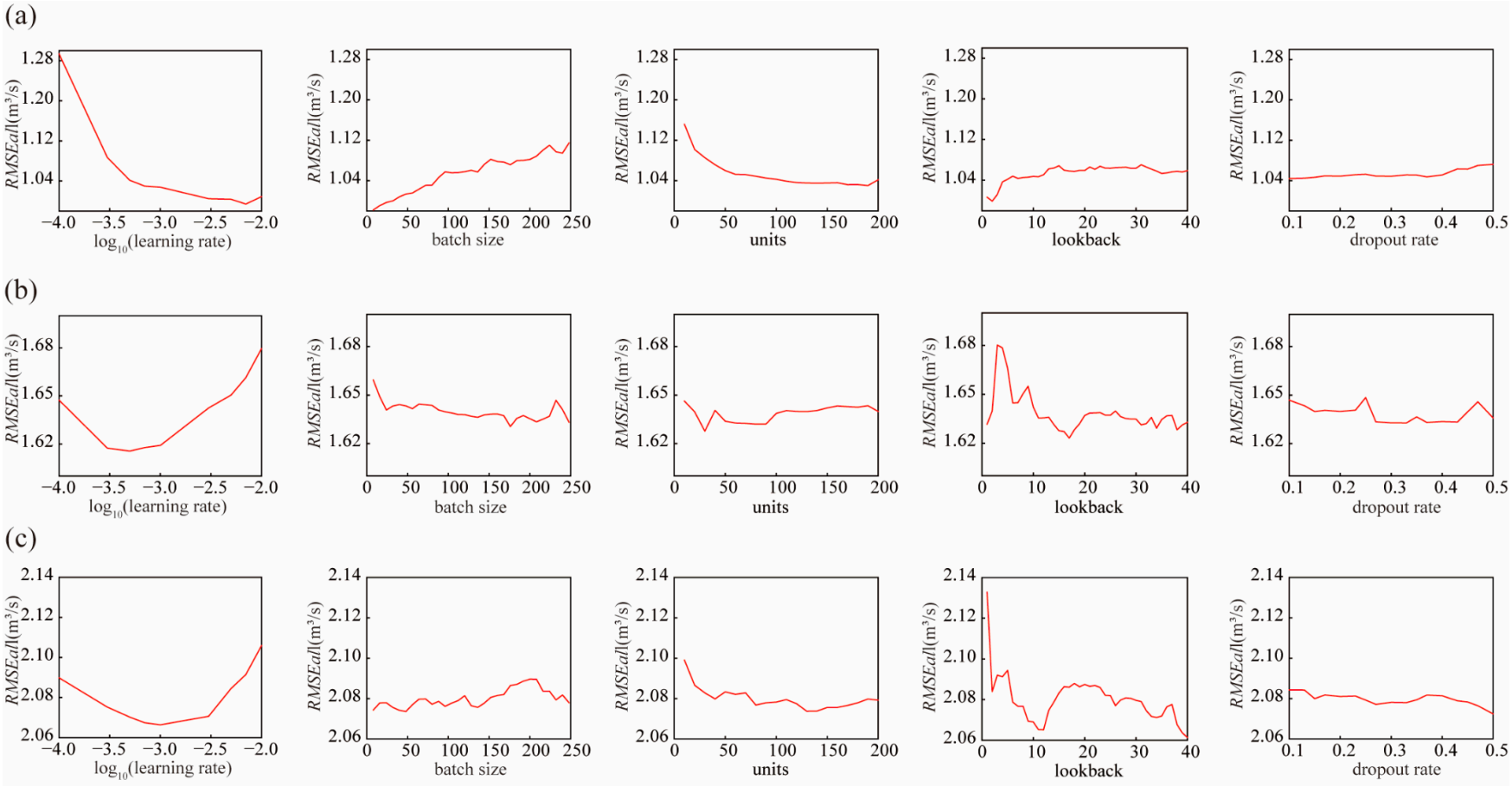


Figure S5. Partial dependence plots of LSTM hyperparameters for $RMSE_{high}$ under different lead times. (a) 1d, (b) 3d, (c) 7d. The X-axis represents hyperparameter values, and the Y-axis represents the simulated performance metric under averaged effects of other hyperparameters.

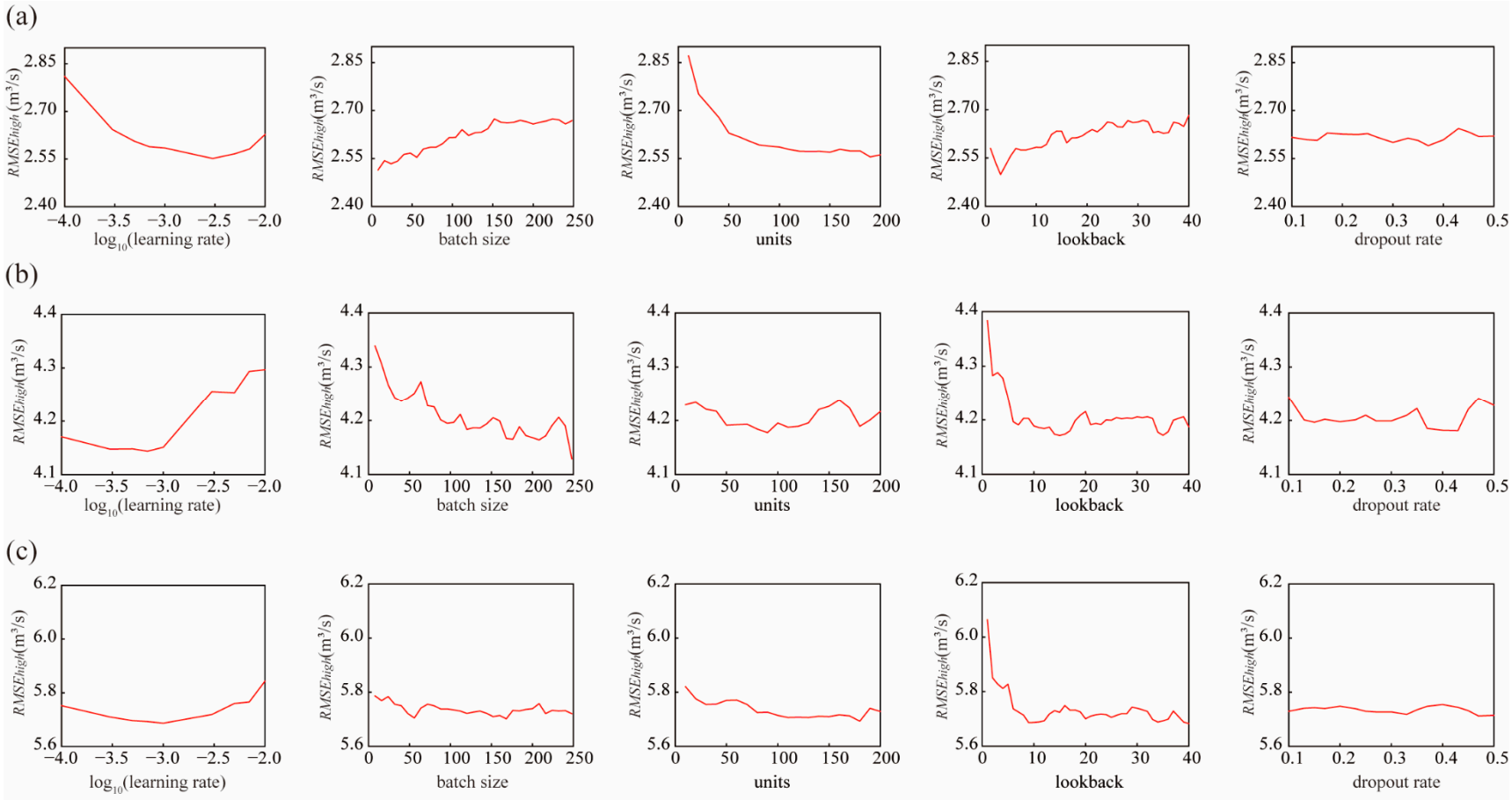


Figure S6. Partial dependence plots of LSTM hyperparameters for $RMSE_{low}$ under different lead times. (a) 1d, (b) 3d, (c) 7d. The X-axis represents hyperparameter values, and the Y-axis represents the simulated performance metric under averaged effects of other hyperparameters.

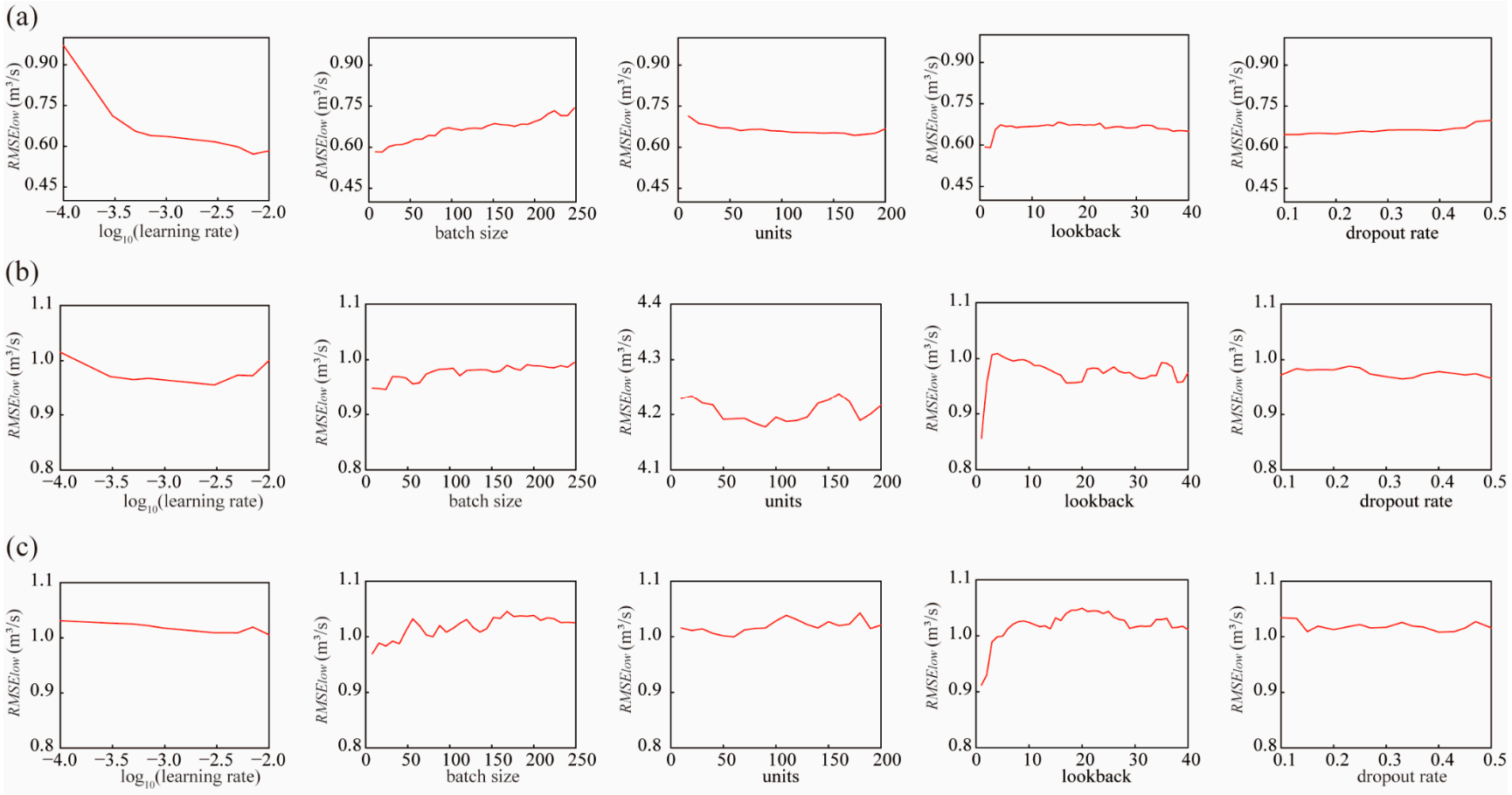
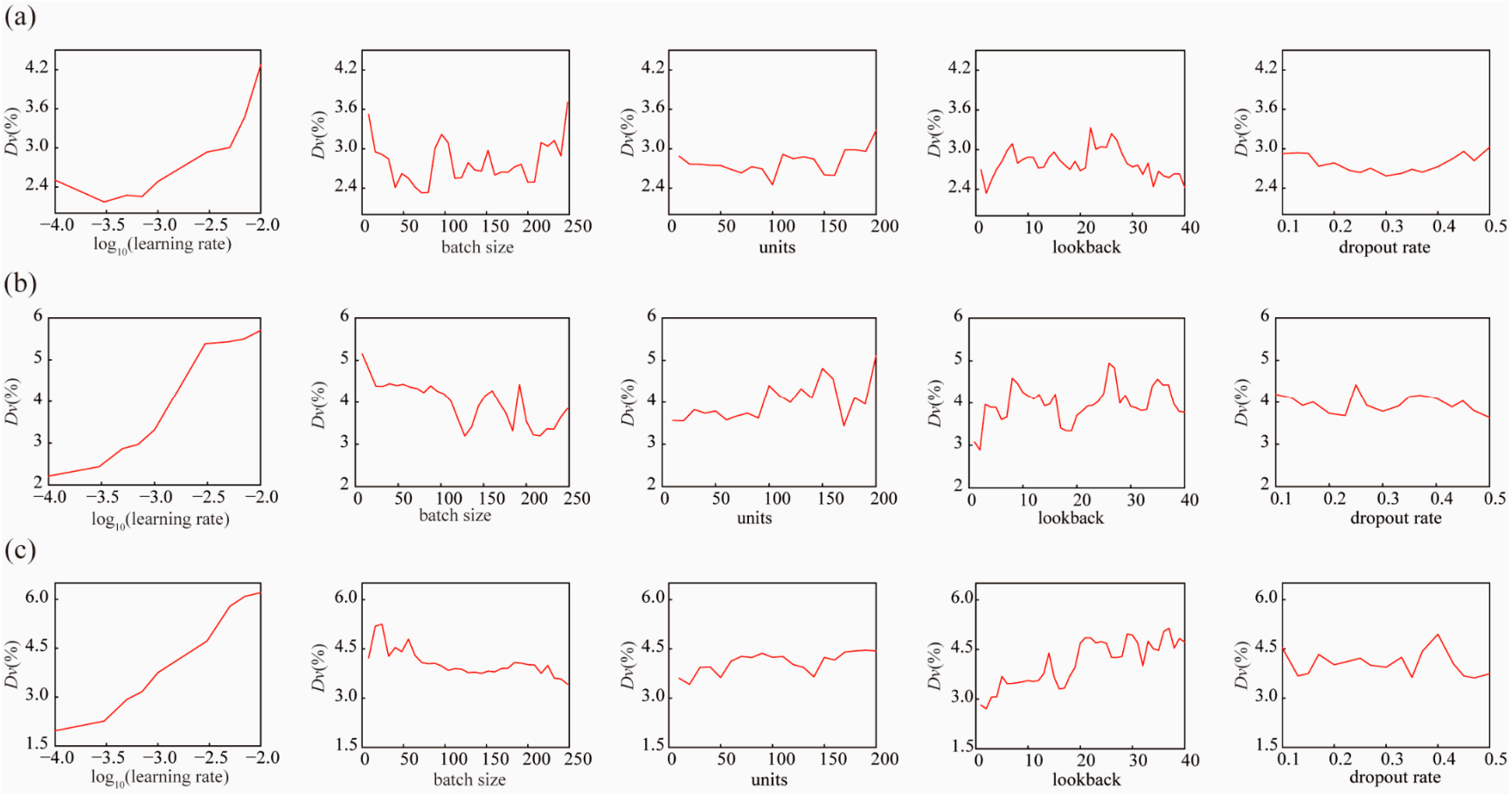


Figure S7. Partial dependence plots of LSTM hyperparameters for *Dv* under different lead times. (a) 1d, (b) 3d, (c) 7d. The X-axis represents hyperparameter values, and the Y-axis represents the simulated performance metric under averaged effects of other hyperparameters.



Reference

S1. Lyne, V.; Hollick, M. Stochastic time-variable rainfall-runoff modelling. In Hydrology and Water Resources Symposium, Perth, Australia, 10–12 September **1979**, pp. 89–92.

S2. Nathan, R. J.; McMahon, T. A. Evaluation of automated techniques for base flow and recession analyses. *Water Resour. Res.* **1990**,26,1465-1473.

S3. Arnold, J.; Allen, P.; Muttiah, R.; Bernhardt, G. Automated base flow separation and recession analysis techniques. *Groundwater* **1995**,33,1010-1018.

S4. Güngör, M. N.; Ozkaya, A. Baseflow index estimation using digital and minima methods in the Büyük Menderes Basin. Sustain. *Water Resour. Manag.* **2025**,11,129.

S5. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

S6. Christian, K.;Roy, A. F.;Yudianto, D.; Zhang, D. Application of optimized support vector machine in monthly streamflow forecasting: using autocorrelation function for input variables estimation. *Sustain. Water Resour. Manag.* **2021**,7,29.

S7. Fu, J.;Du, L.; Li, H. Hybrid intelligent multi-objective dynamic optimization based on MOEPSO-GA. *IEEE Trans. Autom. Sci. Eng.* **2026**,

S8. Yin, S.;Luo, Q.; Zhou, Y. IBMSMA: An indicator-based multi-swarm slime mould algorithm for multi-objective truss optimization problems. *J. Bionic Eng.* **2023**,20,1333-1360.