Supplement of Hydrol. Earth Syst. Sci., 30, 3425–3438, 2026
https://doi.org/10.5194/hess-30-3425-2026-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Symbolic regression-based regionalization of baseflow separation parameter using catchment-scale characteristics

Yongen Lin et al.

Correspondence to: Dagang Wang (wangdag@mail.sysu.edu.cn)

S1. Derivation of SEC based on mass balance relationships

For a two-component streamflow composition scenario, in which total streamflow is composed of a rapid surface/shallow subsurface flow component and a slower baseflow component, the mass balance relationships can be expressed as:

$$SEC_t Q_t = SEC_{s,t} S_t + SEC_{b,t} B_t \quad , \qquad (\text{S1})$$

$$Q_t = S_t + B_t \quad , \qquad (\text{S2})$$

where Q_t and SEC_t denote the total streamflow and the corresponding environmental tracer concentration at time step t, respectively. S_t and B_t are the surface flow and baseflow contributions at time step t. $SEC_{s,t}$ and $SEC_{b,t}$ are the tracer concentration of surface flow and baseflow. Combining equations S1 and S2 to solve for SEC_t, we obtain:

$$\widehat{SEC}_t = \frac{SEC_{s,t}(Q_t - B_t) + SEC_{b,t} B_t}{Q_t} \quad . \qquad (\text{S3})$$

S2. Diebold–Mariano (DM) test

To evaluate whether the predictive performance of different formulas is statistically distinguishable, we applied the Diebold–Mariano (DM) test (Diebold and Mariano, 1995). The DM test compares predictive accuracy based on the loss differential between two competing models. First, the loss differential at time t is defined as

$$d_t = L(e_{A,t}) - L(e_{B,t}) \ , \qquad (\text{S4})$$

where $e_{A,t}$ and $e_{B,t}$ denote the prediction errors of formulas A and B at time t, respectively, and $L(\cdot)$ is the loss function. In this study, the squared error was used to test for the significant differences between different R^2 values:

$$L(e_t) = e_t^2 \quad . \qquad (\text{S5})$$

The DM test statistic is then computed as

$$DM = \frac{\bar{d}}{\sqrt{\widehat{Var}(\bar{d})}} \quad , \qquad (\text{S6})$$

where $\bar{d}$ is the mean loss differential and $\widehat{Var}(\bar{d})$ is the long-run variance of $\bar{d}$. Under the null hypothesis of equal predictive accuracy, the DM statistic asymptotically follows a standard normal distribution. A two-sided test was used to determine whether the predictive performance of two SR formulas differed significantly at the level of 0.01.

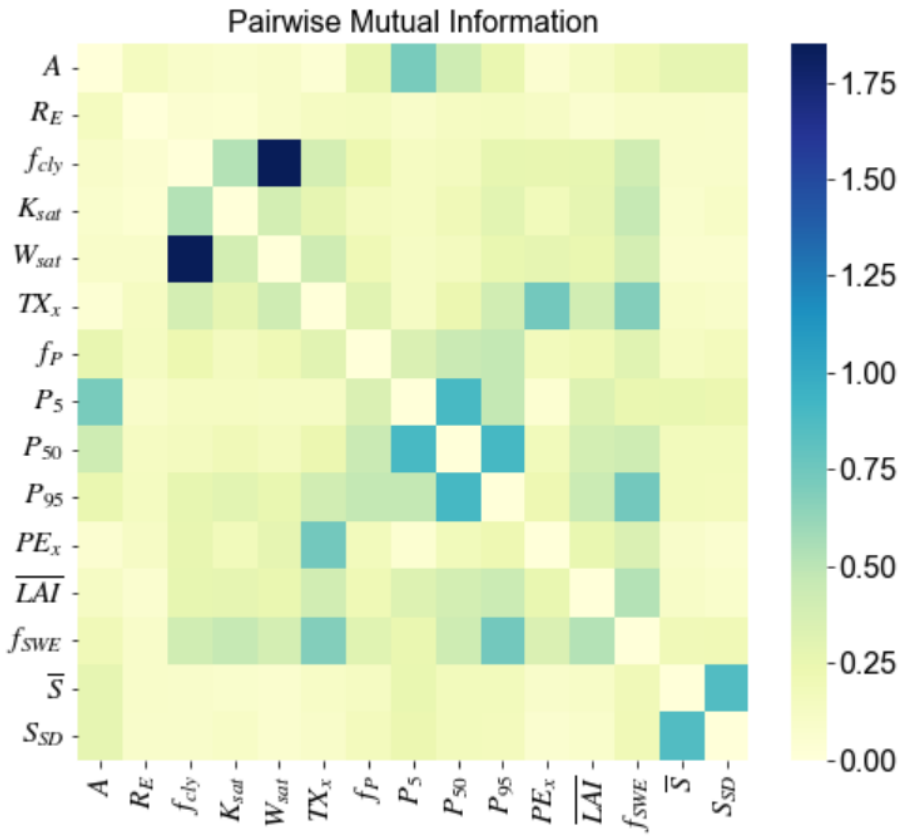

Figure S1. Mutual information matrix of candidate predictor variables.

References

Diebold, F.X., Mariano, R.S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13(3): 253-263. DOI:10.1080/07350015.1995.10524599