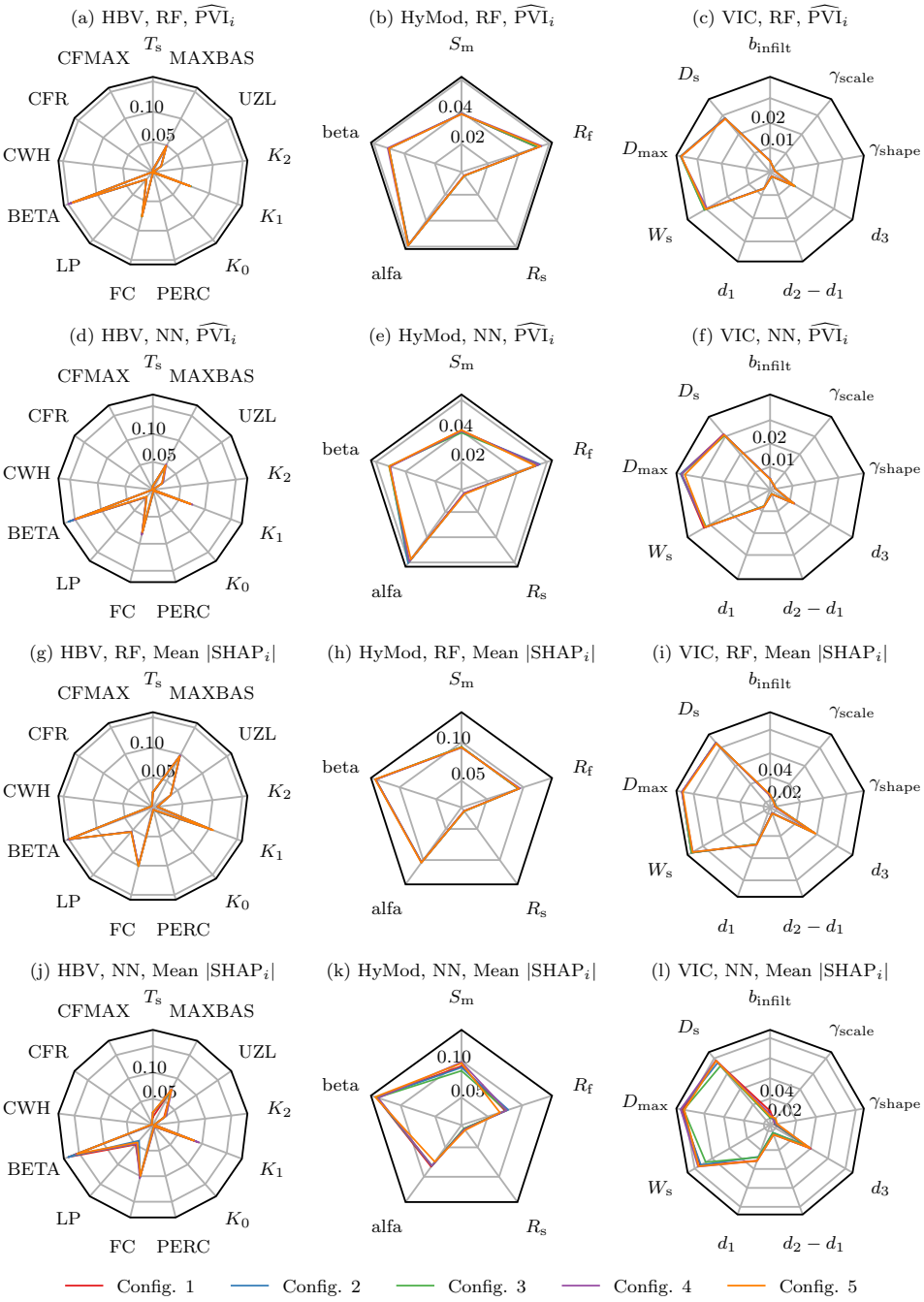

Figure S1. Sanity check of feature importance estimates for RF and NN metamodels across five overfitting configurations for the Schackendorf catchment. Panels (a–c) show $\widehat{\mathrm{PVI}}_i$ and panels (g–i) show mean $|\mathrm{SHAP}_i|$ for the RF metamodel applied to HBV, HyMod, and VIC, respectively, while panels (d–f) and (j–l) show the corresponding results for the NN metamodel. Each coloured line represents a different model configuration. The strong overlap of importance curves across configurations indicates that the feature importance estimates are robust to overfitting. Based on these results, configuration 1 was retained for RF and NN metamodels for all subsequent analyses.

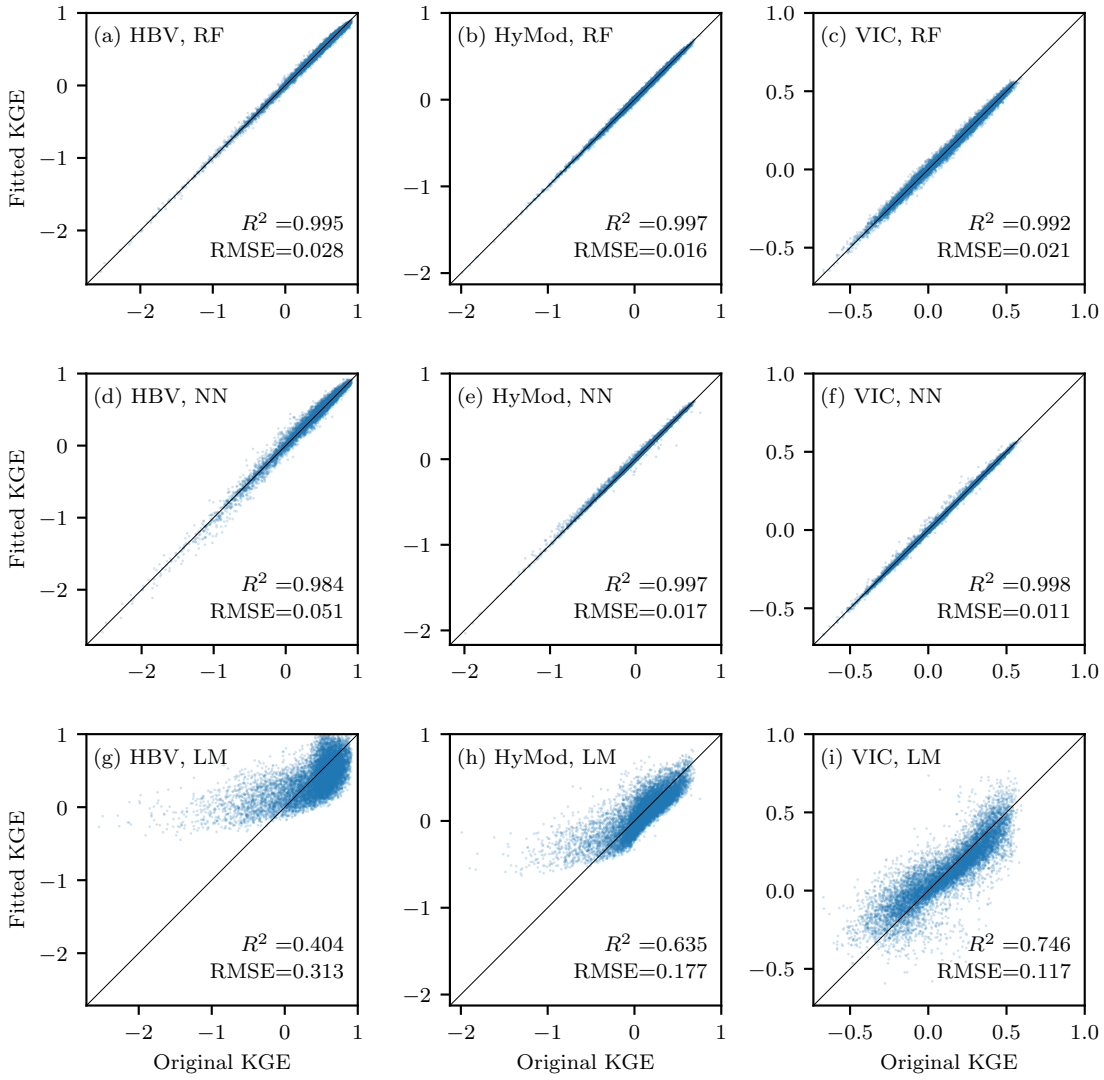

Figure S2. Predictive performance of the RF, NN, and LM metamodels for the three hydrologic models applied in the Schackendorf catchment. Scatter plots compare fitted and original KGE values for (a–c) RF, (d–f) NN, and (g–i) LM applied to HBV, HyMod, and VIC, respectively. The solid line indicates the 1:1 relationship. Coefficients of determination (R^2) and root mean square errors (RMSE) are reported in each panel.

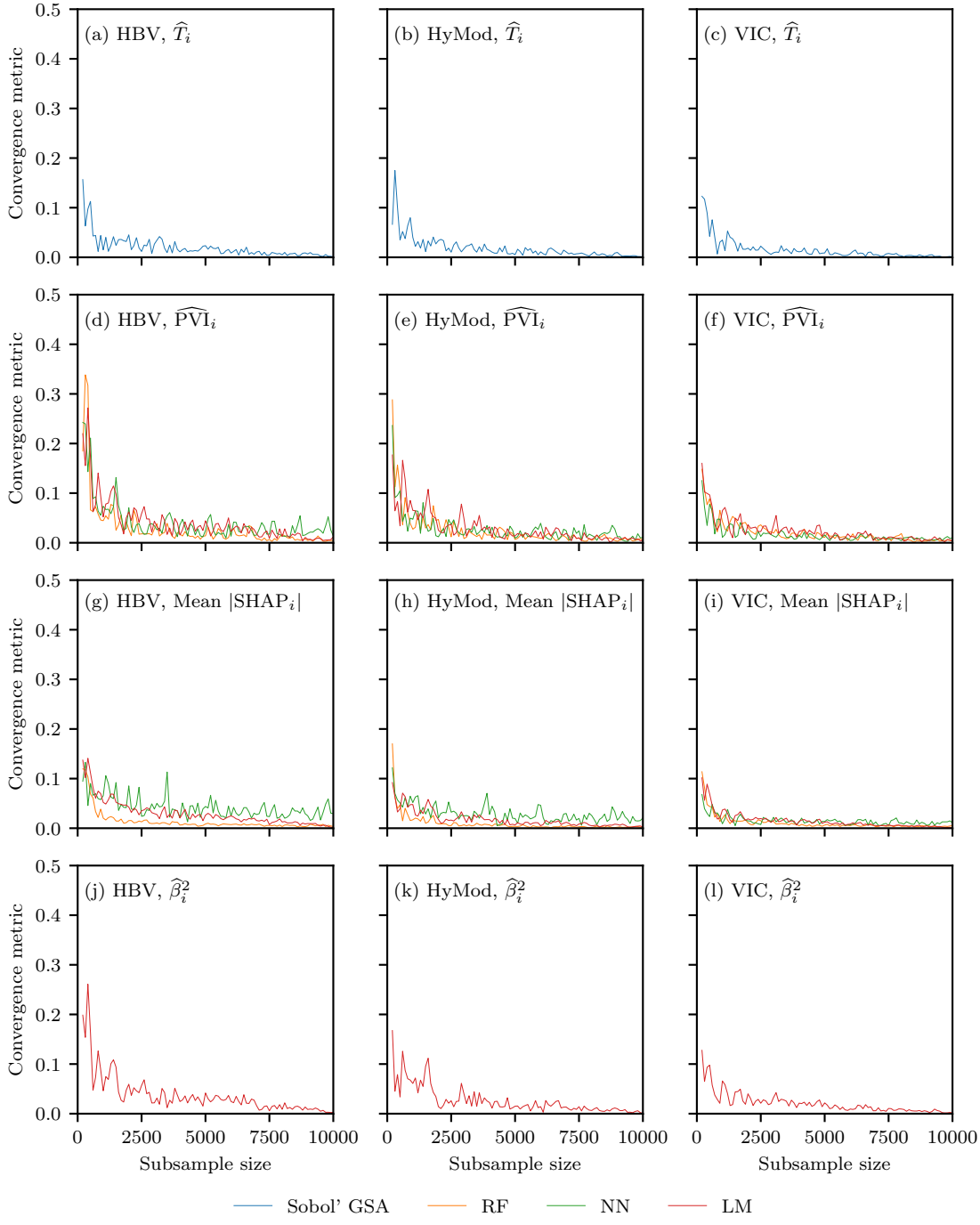

Figure S3. Convergence of sensitivity measures as a function of subsample size for the Schackendorf catchment. Panels (a–c) show the convergence of $\widehat{T}_i$, panels (d–f) of $\widehat{\mathrm{PVI}}_i$, panels (g–i) of mean $|\mathrm{SHAP}_i|$, and panels (j–l) of $\widehat{\beta}_i^2$ for HBV, HyMod, and VIC, respectively. The convergence metric is computed as the Euclidean distance between normalized sensitivity vectors obtained from consecutive subsample sizes. Normalization was performed by dividing by the sum of all sensitivities.

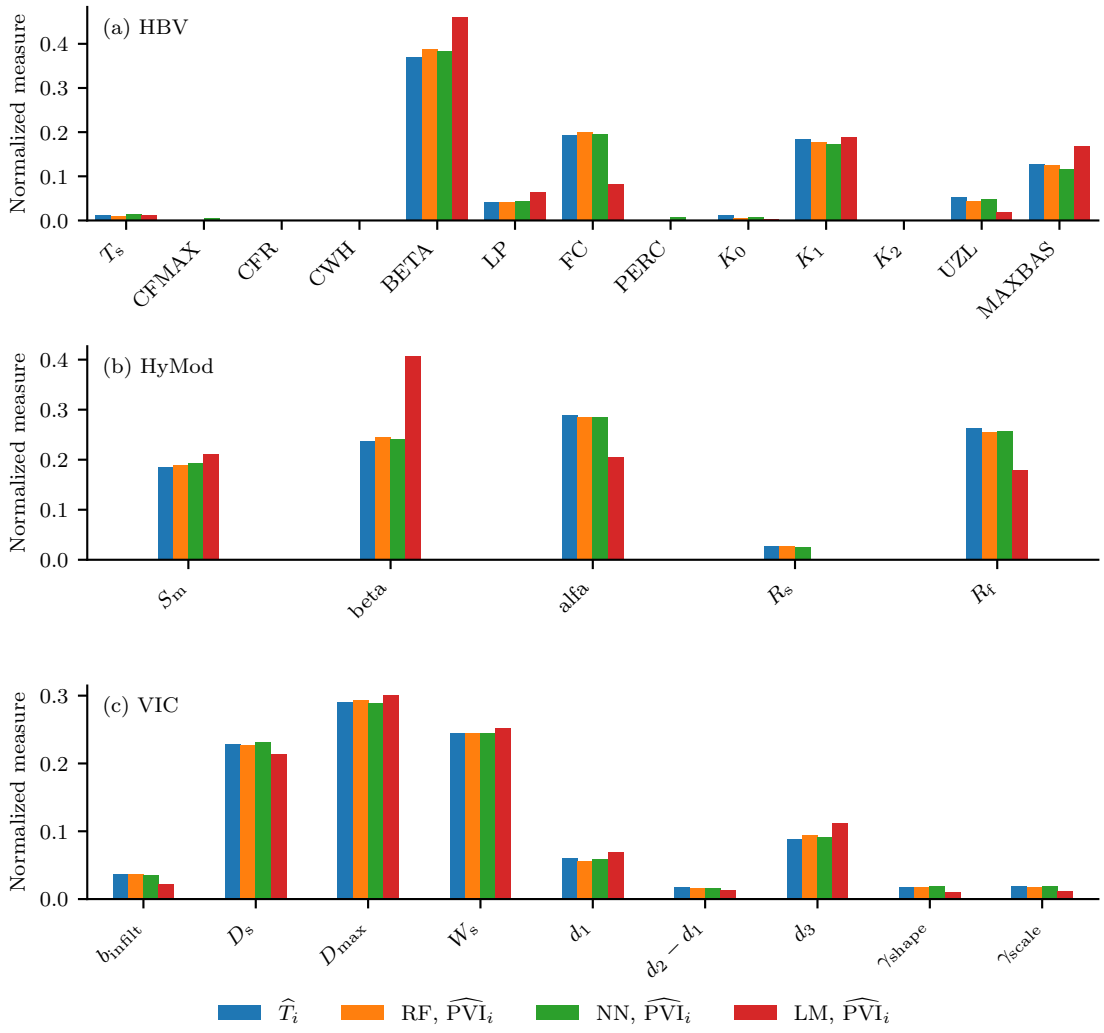

Figure S4. Parameter rankings for (a) HBV, (b) HyMod, and (c) VIC applied in the Schackendorf catchment. Bars show normalized $\widehat{T}_i$ from variance-based GSA and normalized $\widehat{\mathrm{PVI}}_i$ estimates obtained from RF, NN, and LM metamodels. Sensitivity measures were normalized by dividing each by the sum of all sensitivities, as in the convergence test.

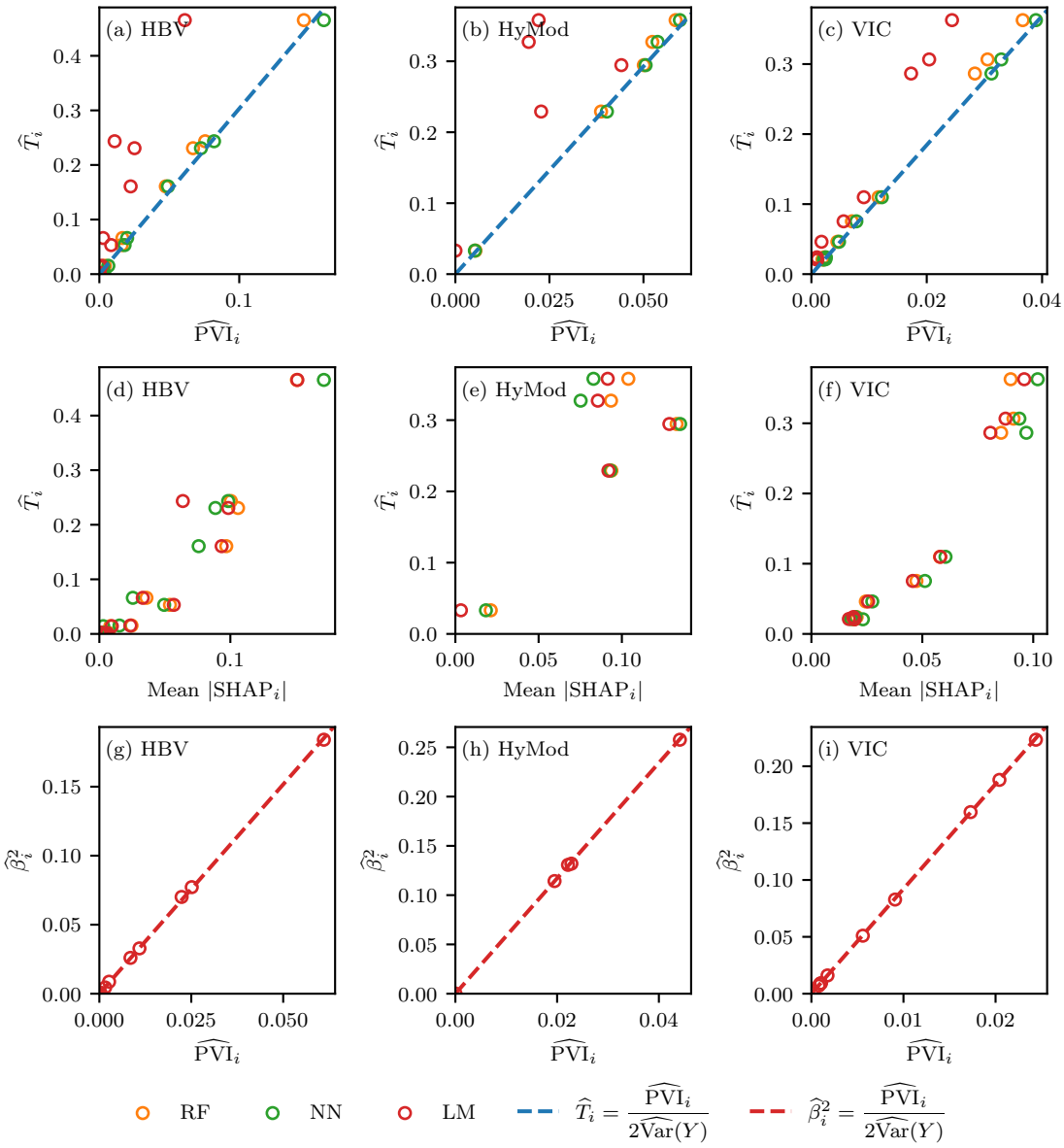

Figure S5. Comparison of different sensitivity estimates across hydrologic models for the Schackendorf catchment. Panels (a–c) compare $\widehat{\mathrm{PVI}}_i$ with $\widehat{T}_i$ for HBV, HyMod, and VIC, respectively. Panels (d–f) compare mean $|\mathrm{SHAP}_i|$ with $\widehat{T}_i$, and panels (g–i) compare $\widehat{\mathrm{PVI}}_i$ with $\widehat{\beta}_i^2$ for the LM metamodel. Dashed lines indicate the corresponding theoretical relationships. Colours denote the ML metamodel used.

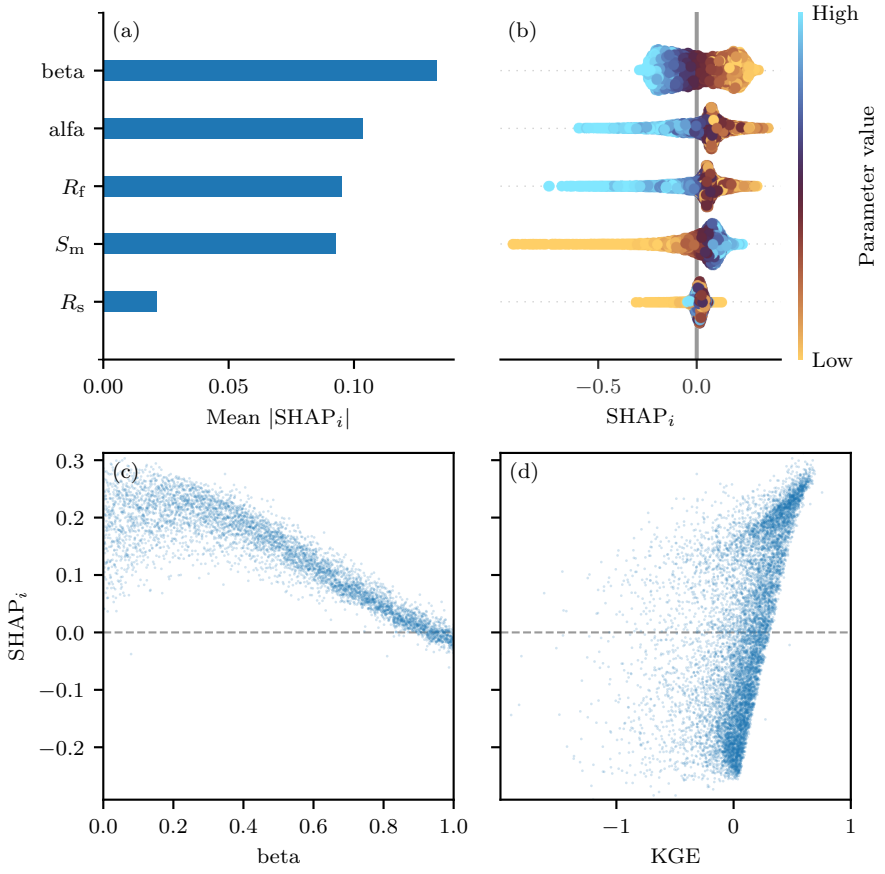

Figure S6. Detailed SHAP_i-based sensitivity analysis for HyMod applied in the Schackendorf catchment. Panel (a) shows the global parameter ranking based on mean $|\mathrm{SHAP}_i|$. Panel (b) displays the distribution of SHAP_i values across the parameter space, coloured by parameter value. Panel (c) illustrates the variation of SHAP_i for the most influential parameter, beta, across its range. Panel (d) shows SHAP_i values for beta as a function of KGE, highlighting how parameter contributions vary across model performance levels.

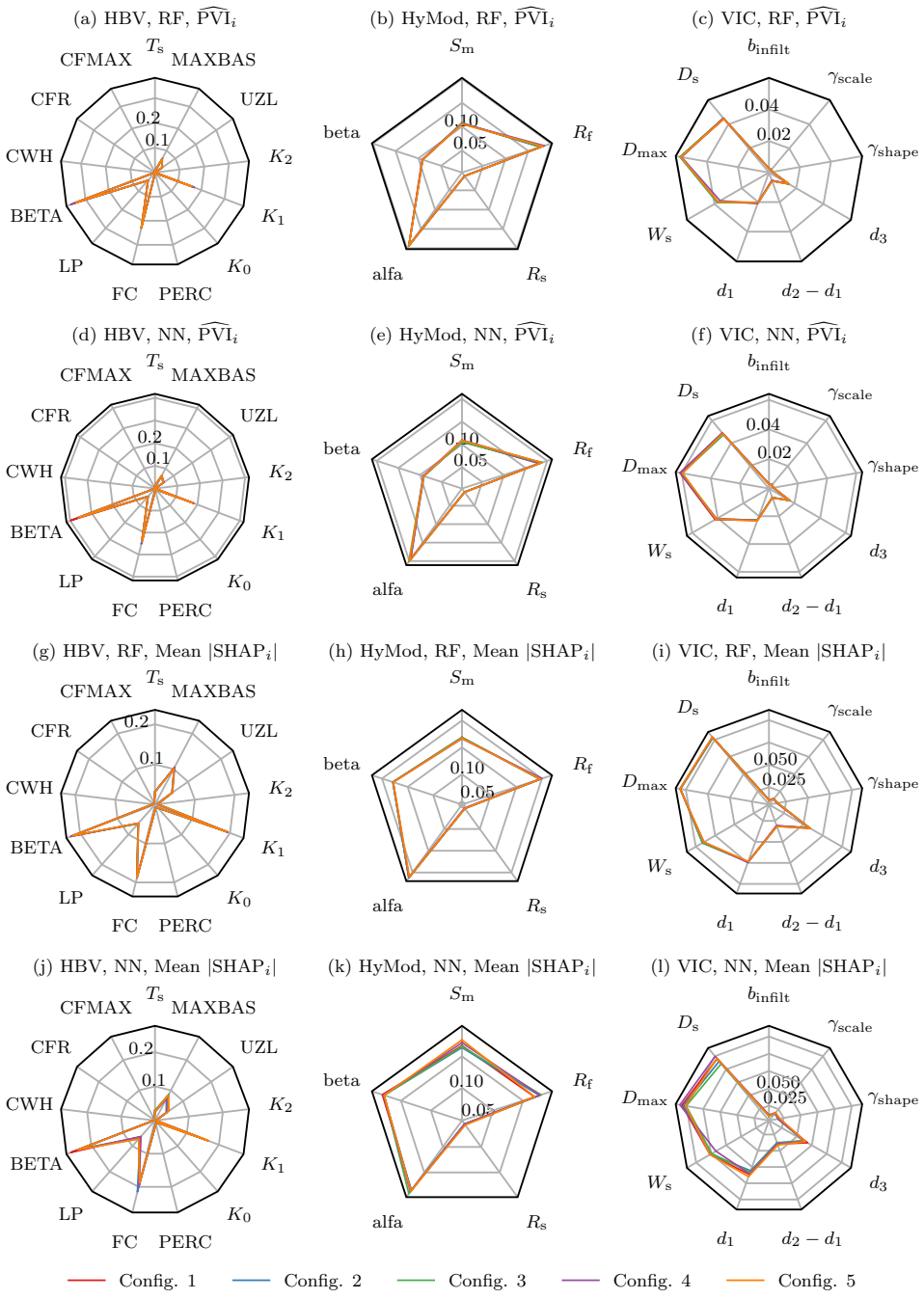

Figure S7. Same as Figure S1 but for the Augaard catchment.

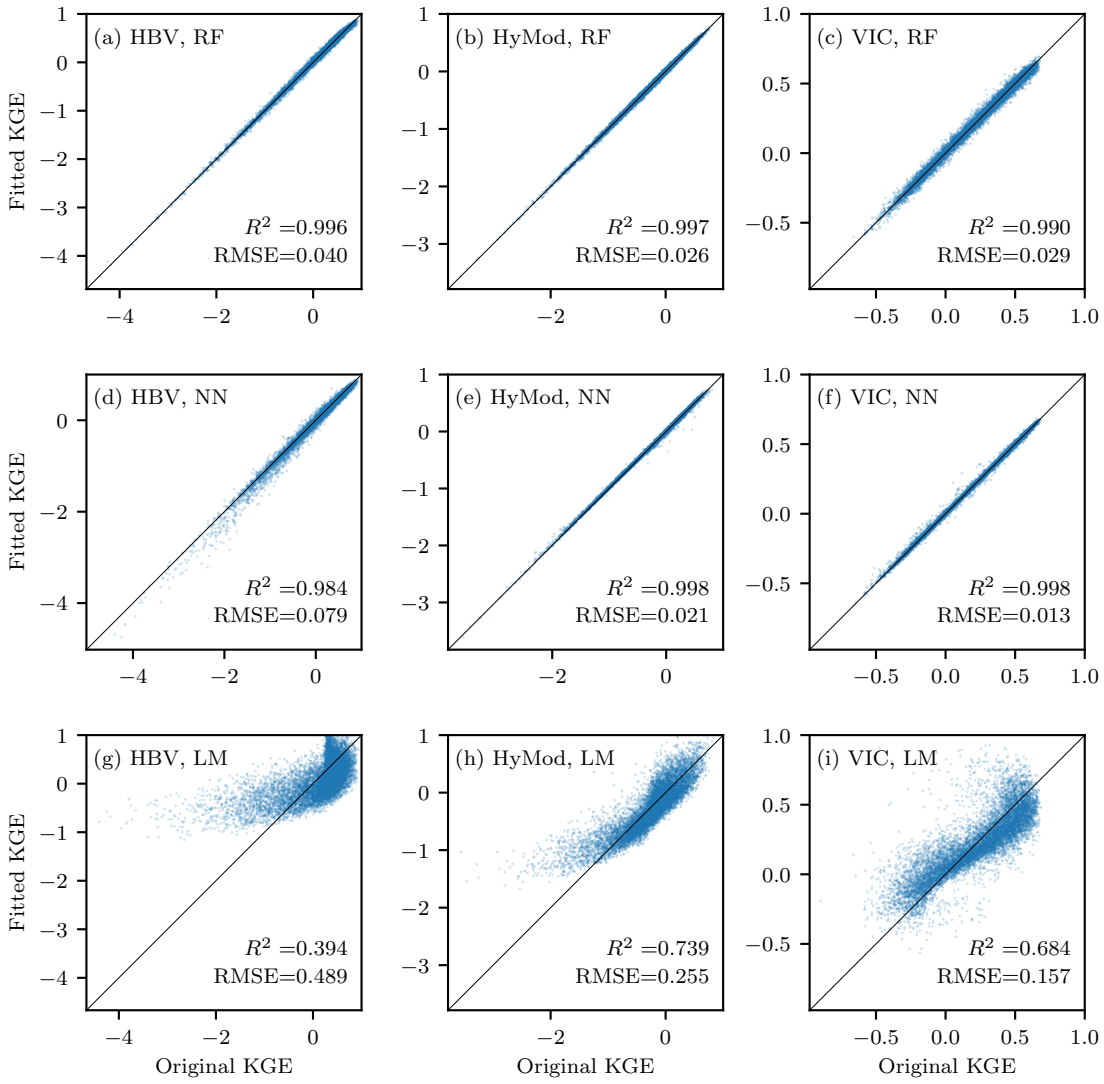

Figure S8. Same as Figure S2 but for the Augaard catchment.

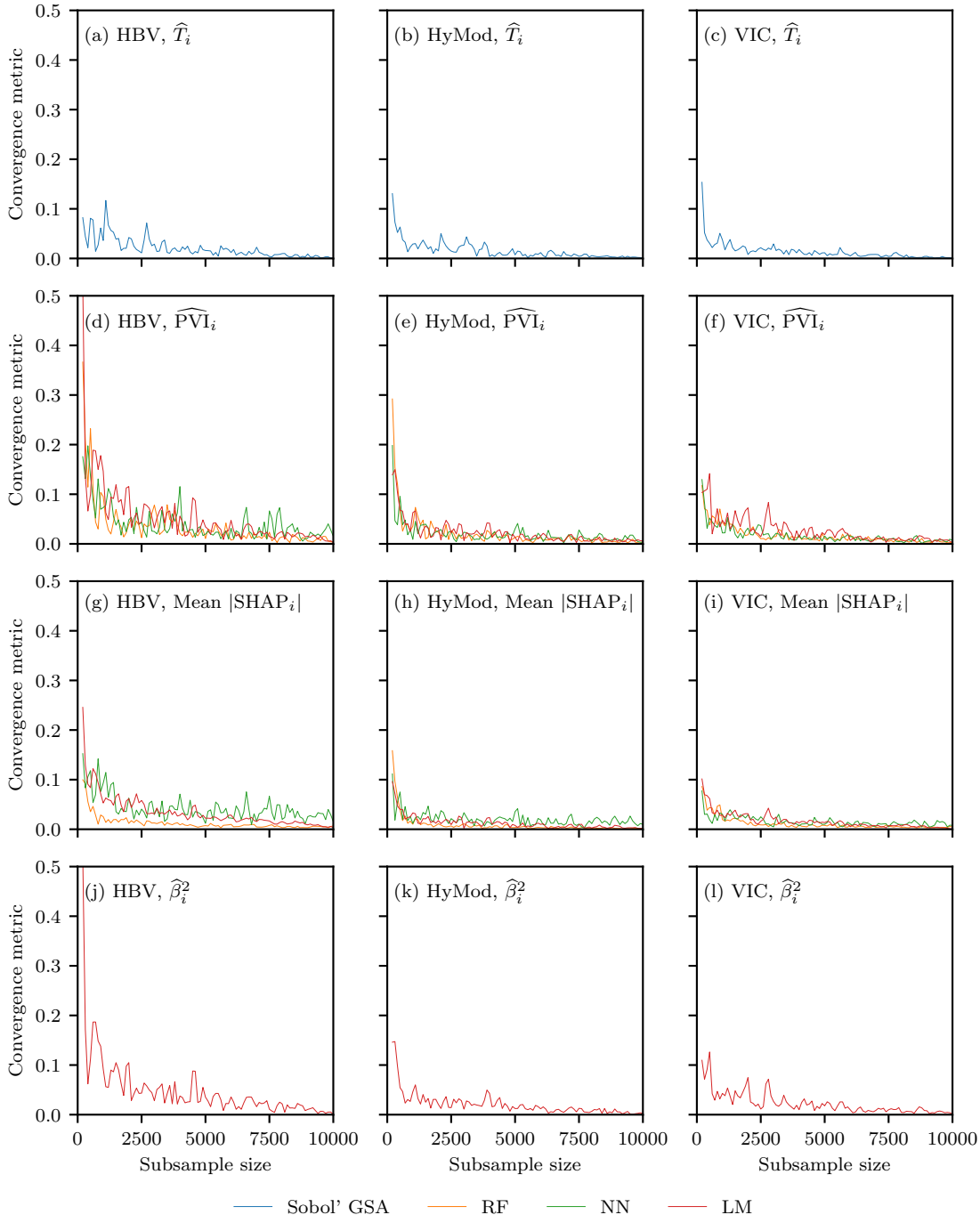

Figure S9. Same as Figure S3 but for the Augaard catchment.

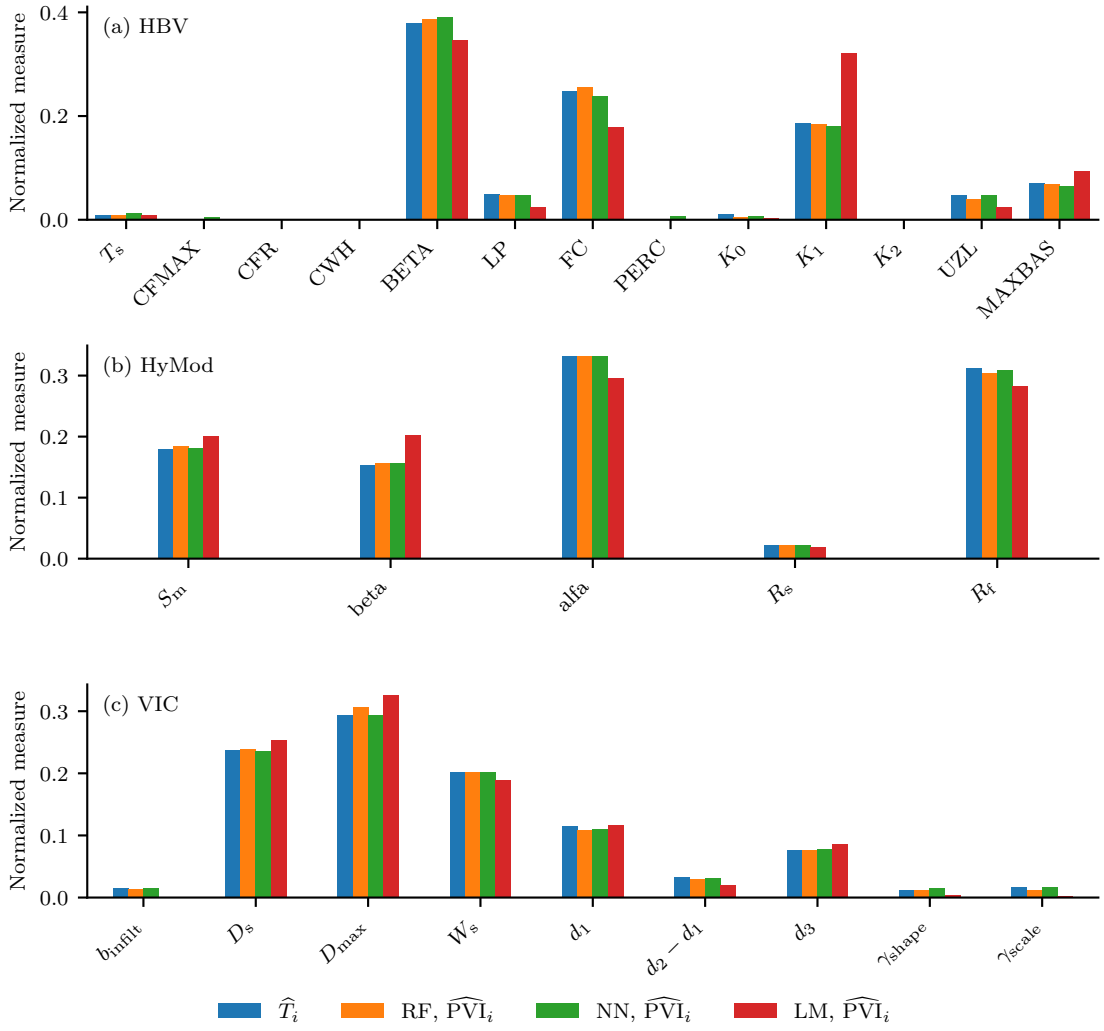

Figure S10. Same as Figure S4 but for the Augaard catchment.

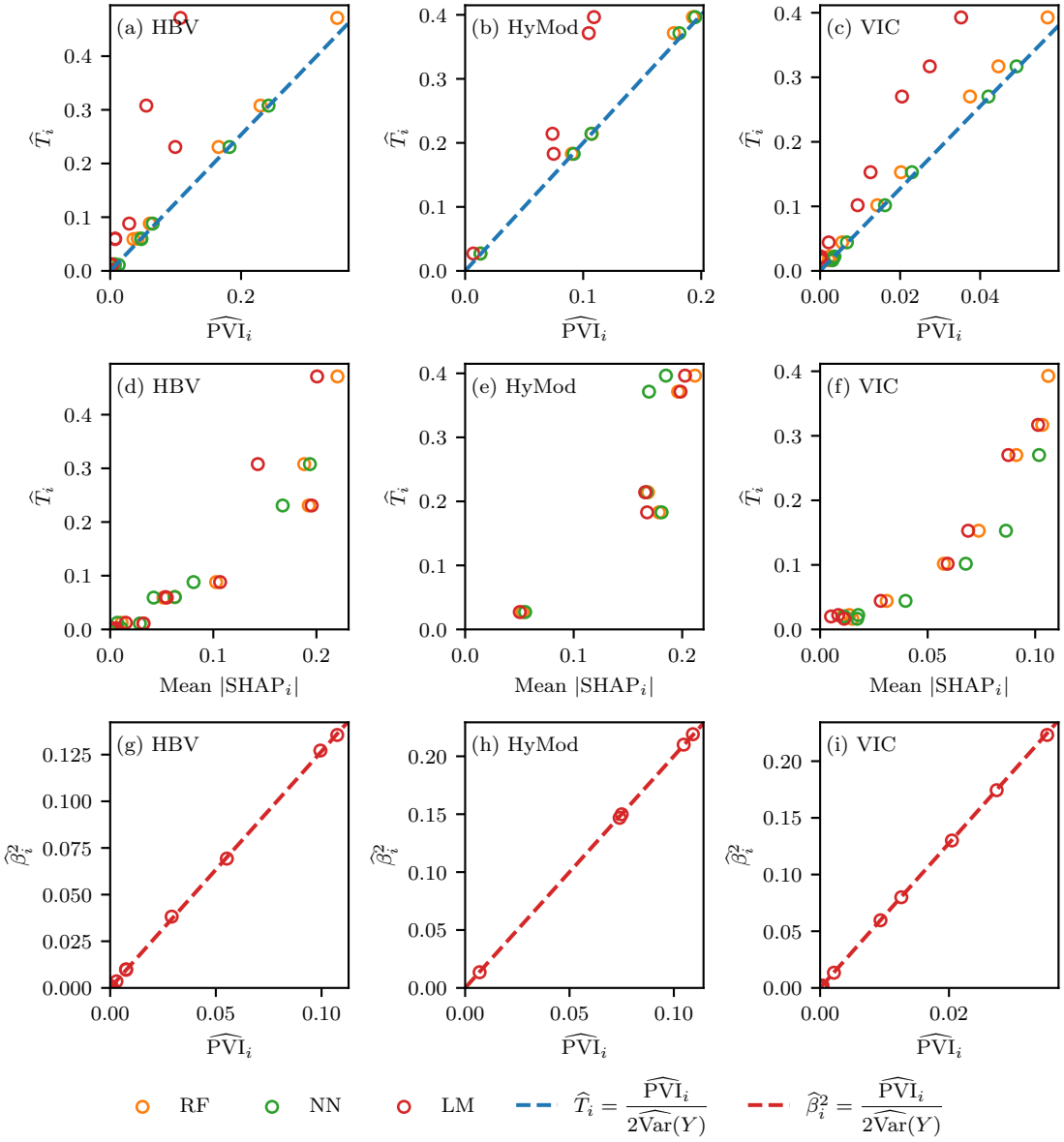

Figure S11. Same as Figure S5 but for the Augaard catchment.

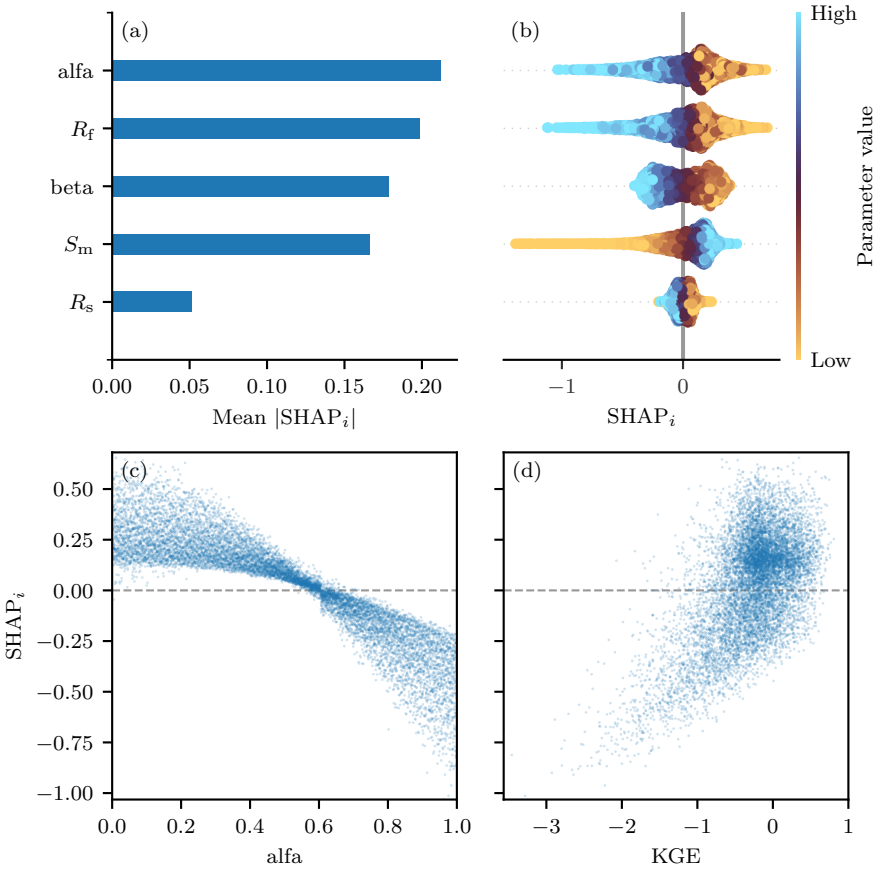

Figure S12. Same as Figure S6 but for the Augaard catchment. In this case, the most influential parameter is alfa.