Supplement of Geosci. Model Dev., 15, 3021–3039, 2022
https://doi.org/10.5194/gmd-15-3021-2022-supplement


Geoscientific Model Development
Open Access

*Supplement of*

# AI4Water v1.0: an open-source python package for modeling hydrological time series using data-driven methods

**Ather Abbas et al.**

*Correspondence to:* Jong Ahn Chun (jachun@apcc21.org) and Kyung Hwa Cho (khcho@unist.ac.kr)

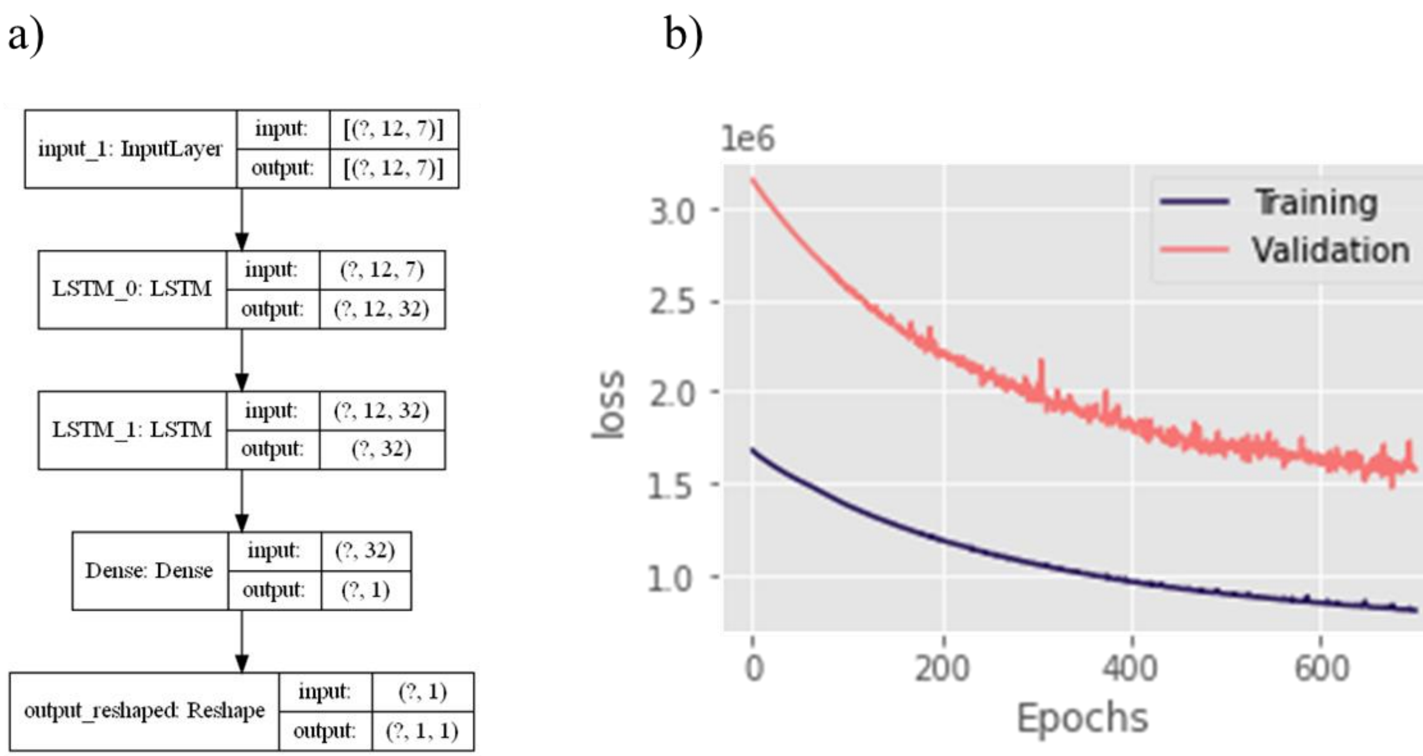


**Figure S1:** Architecture of a four-layer neural network used for prediction of streamflow at catchment number 224206 of CAMELS-AUS. Seven climate variables were used and 12 days of historical data was used for training the model. A) The model consisted of 2 LSTM layers followed by a Dense layer as output layer. The output was finally reshaped into 3d array. B) Training and validation loss curves during model training. The model was trained for 700 epochs.

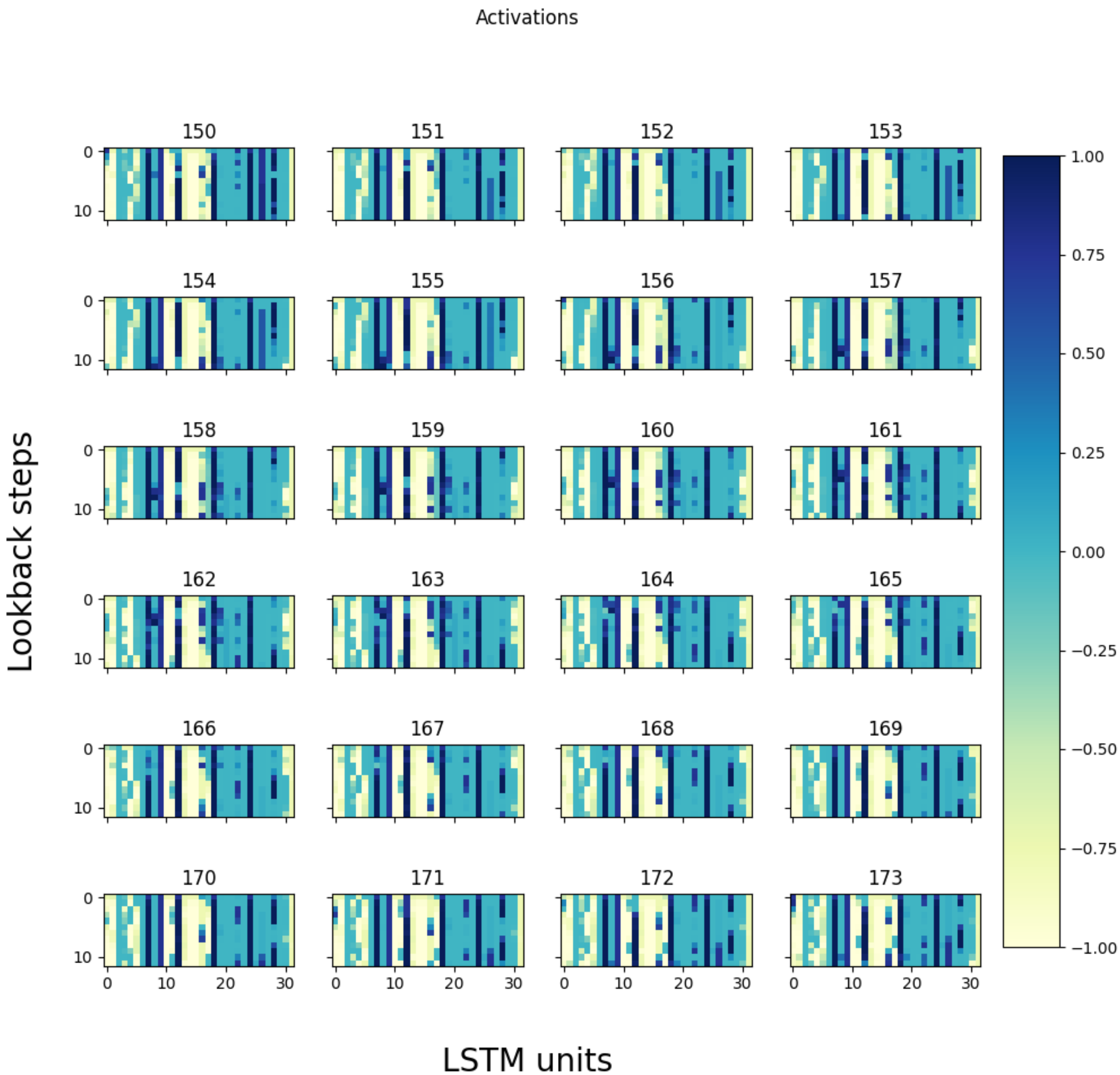


**Figure S2:** Output of first LSTM for 24 days. The model consisted of two LSTM layers with 32 units for each LSTM. The lookback steps indicate the number of historical days used by the model to predict value for next day. The titles for each subplot indicate Julian day for the year 2000.

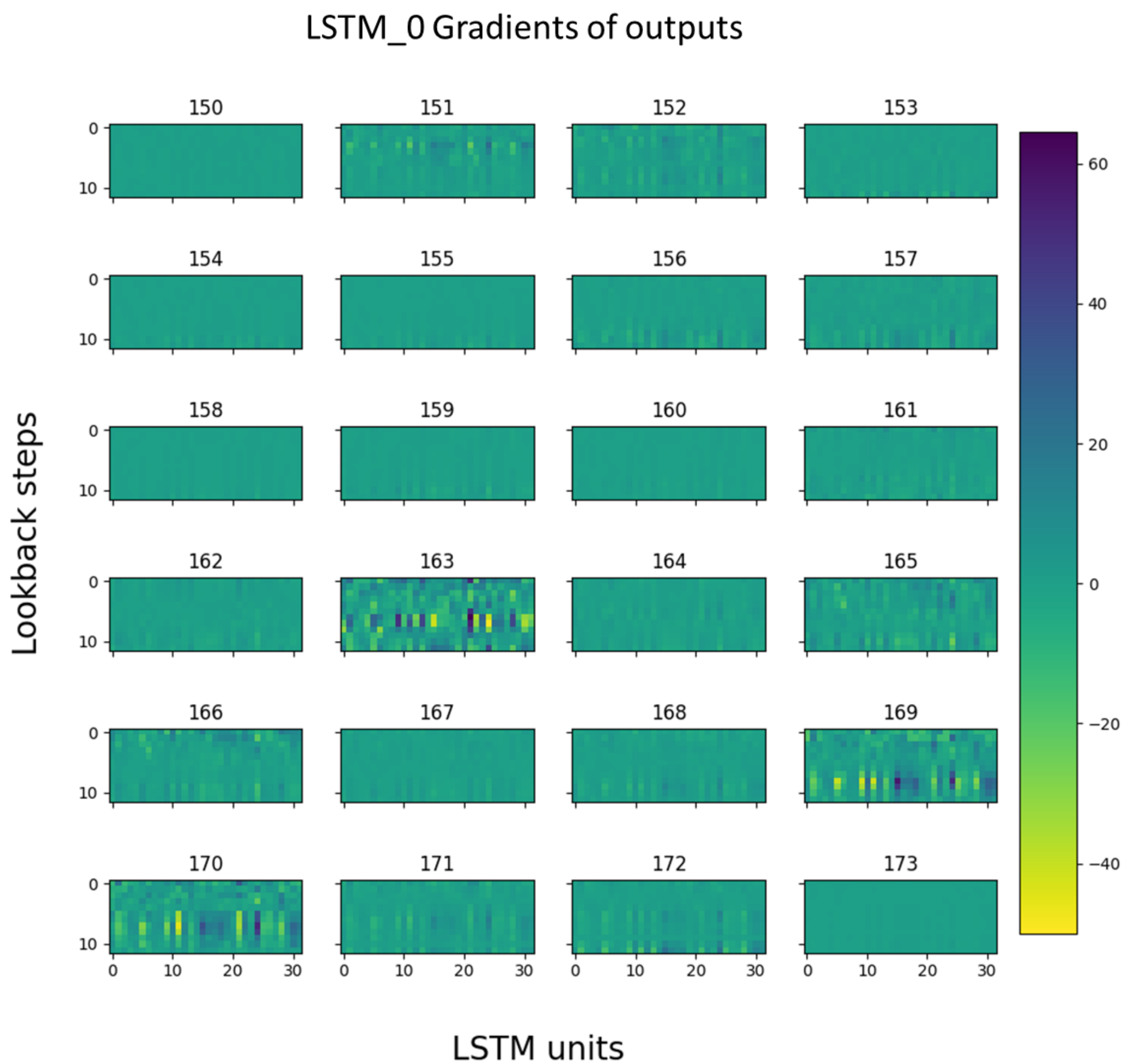


**Figure S3:** Gradients of outputs of first LSTM for 24 days. The model consisted of two LSTM layers with 32 units for each LSTM. The lookback steps indicate the number of historical days used by the model to predict value for next day. The titles for each subplot indicate Julian day for the year 2000.

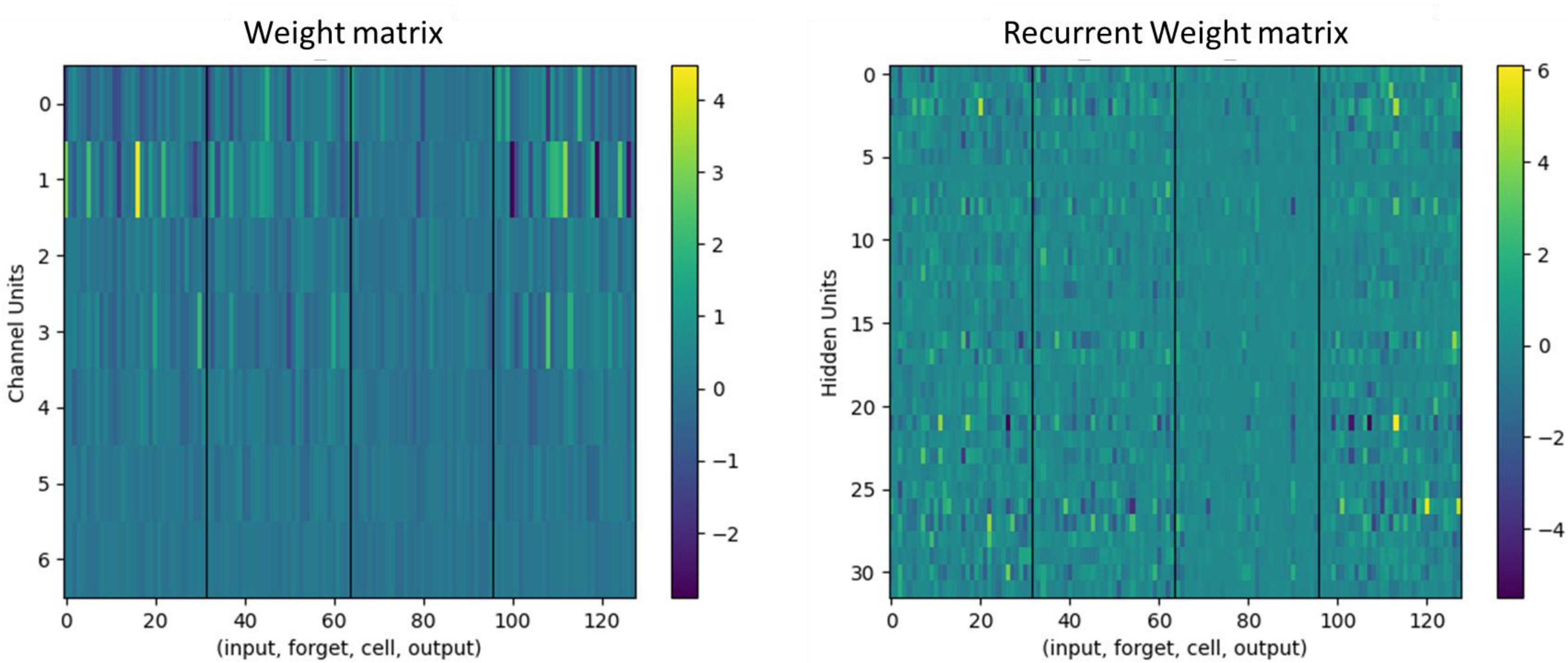


**Figure S4:** Weight matrices of LSTM layer. The LSTM layer consists of two weight matrices. The portion of weight matrices responsible for input gate, forget gate, output gate and cell state are highlighted by lack lines.

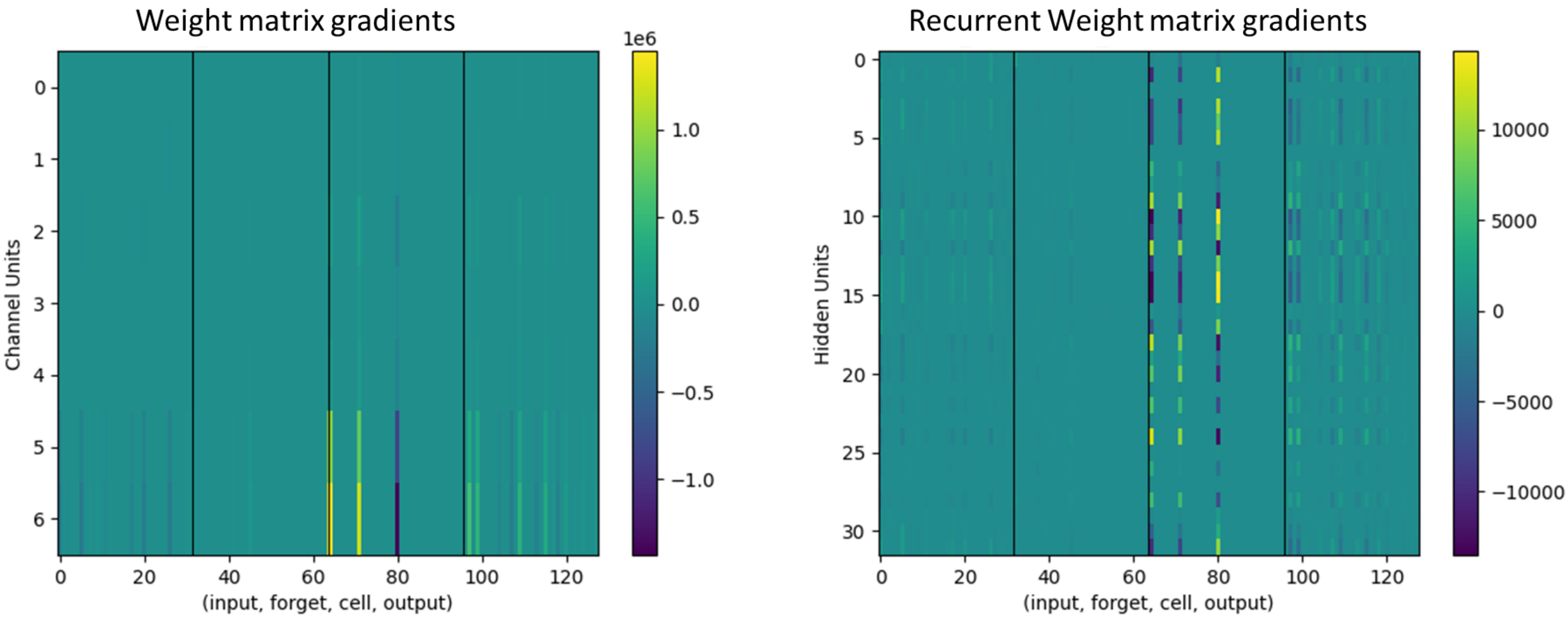


**Figure S5:** Gradients of weight matrices of LSTM layer. The LSTM layer consists of two weight matrices. The portion of weight matrices responsible for input gate, forget gate, output gate and cell state are highlighted by lack lines.

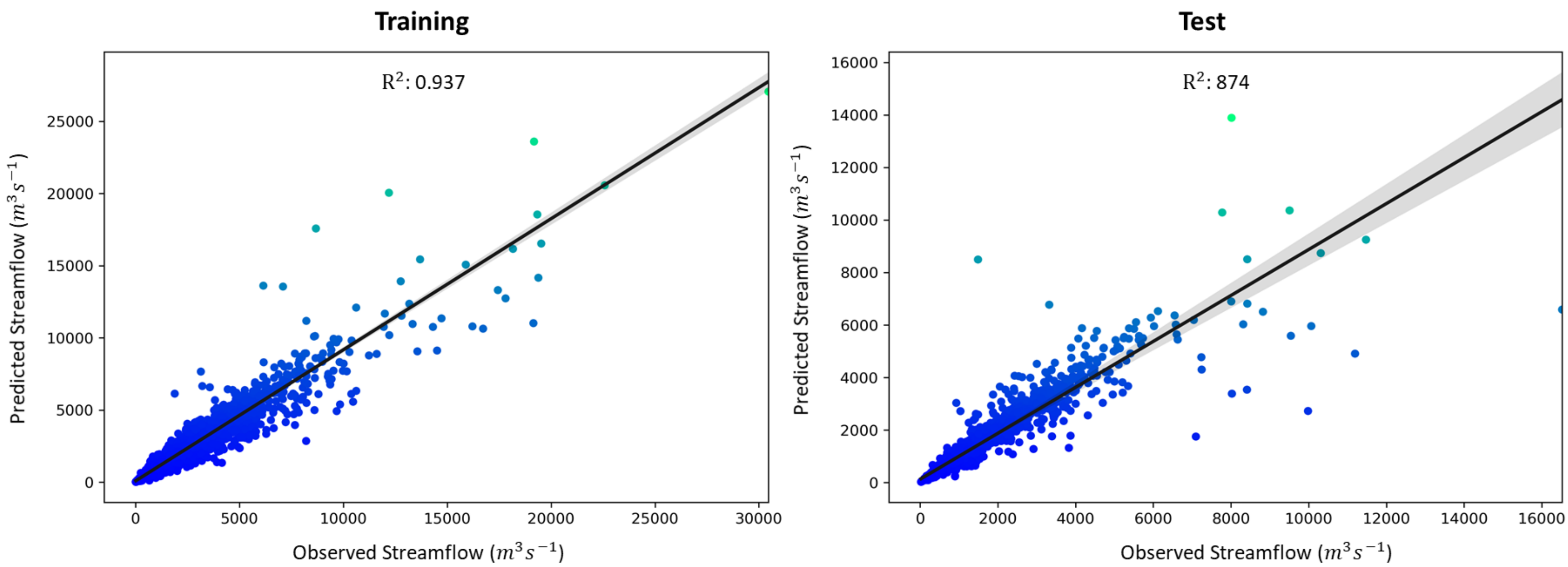


**Figure S6:** Scatter plots between observed and predicted streamflow for training and test dataset for rainfall-runoff modeling in catchment number '401203' of CAMELS Australia dataset. The units of streamflow are cubic meter per second ($m^3 s^{-1}$).

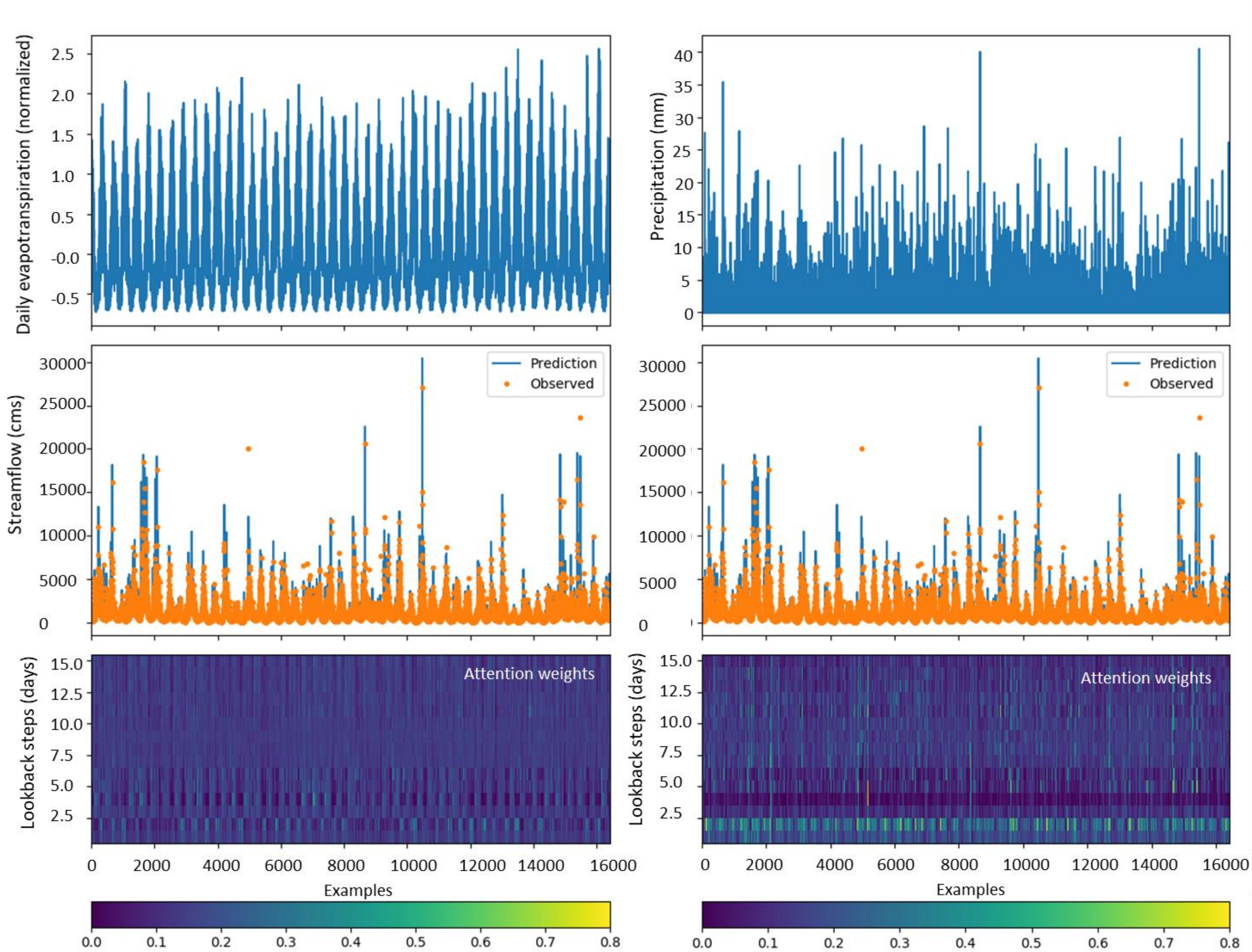


**Figure S7:** Attention weights for Morton Evapotranspiration and Precipitation along lookback steps in dual-stage attention model for rainfall-runoff modeling in catchment number '401203' of CAMELS Australia dataset.

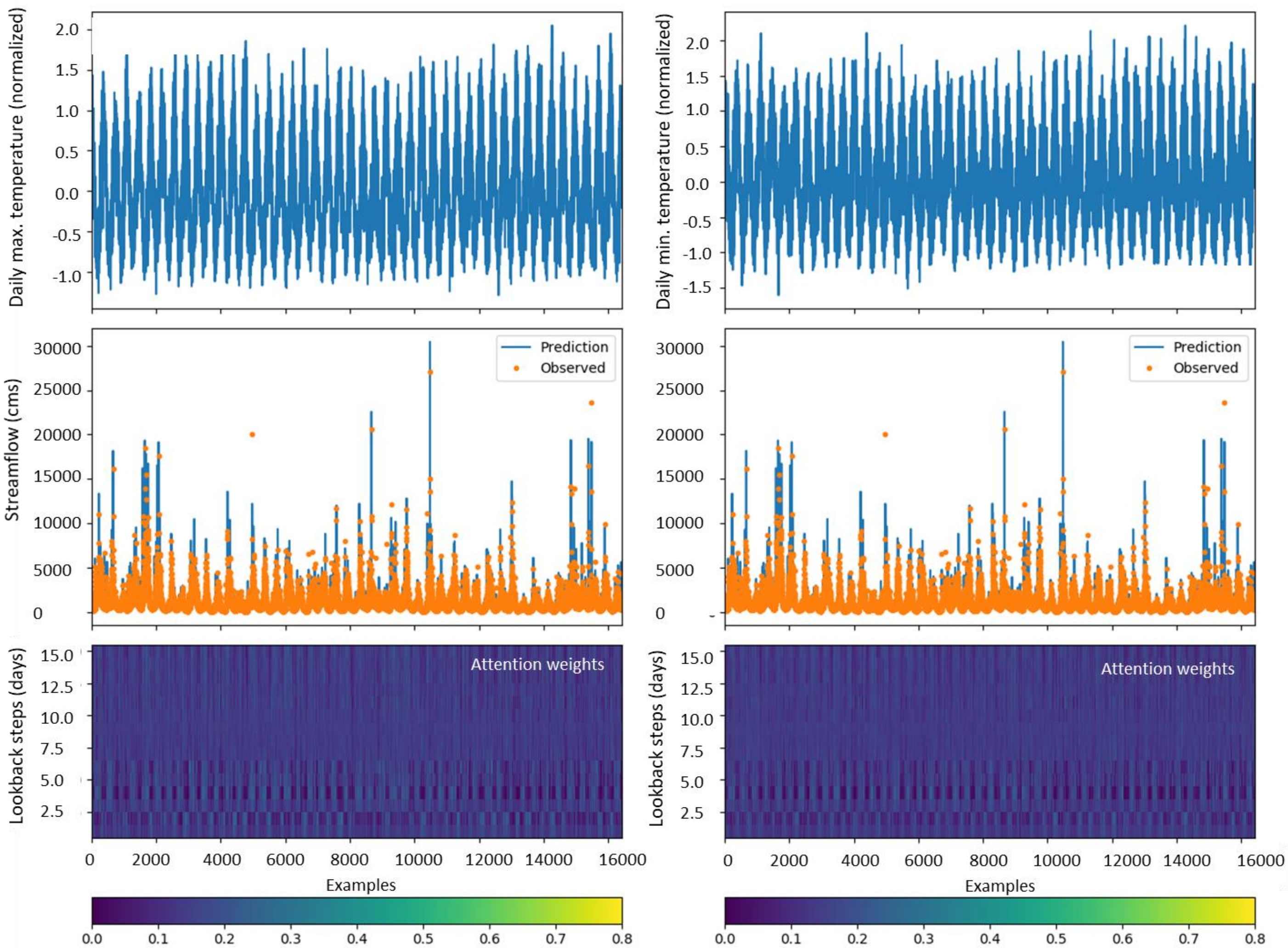


**Figure S8:** Attention weights for maximum and minimum temperature along lookback steps in dual-stage attention model for rainfall-runoff modeling in catchment number '401203' of CAMELS Australia dataset. The units of streamflow are cubic meter per second ($m^3 s^{-1}$).

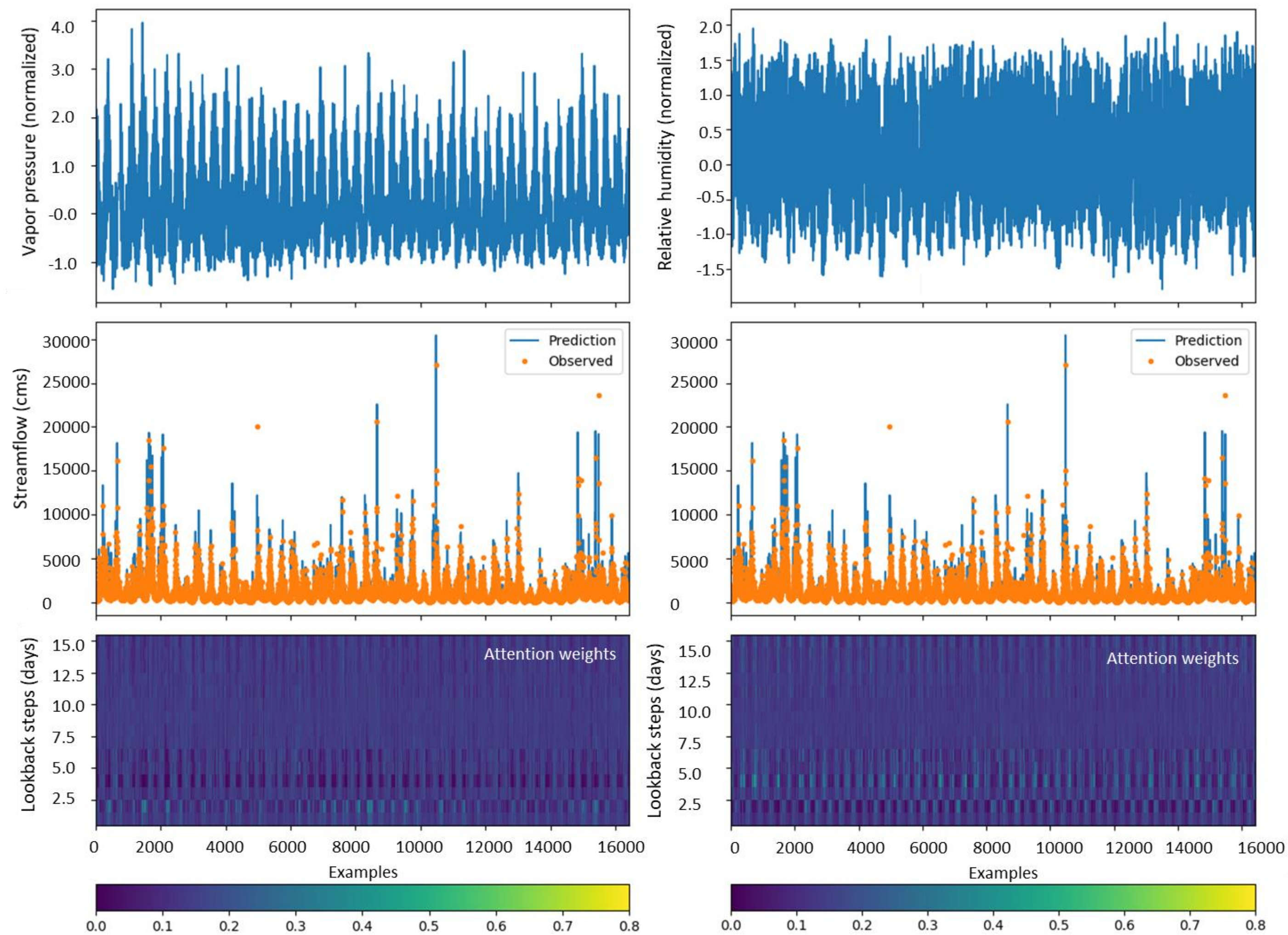


**Figure S9:** Attention weights for vapor pressure and relative humidity along lookback steps in dual-stage attention model for rainfall-runoff modeling in catchment number '401203' of CAMELS Australia dataset.

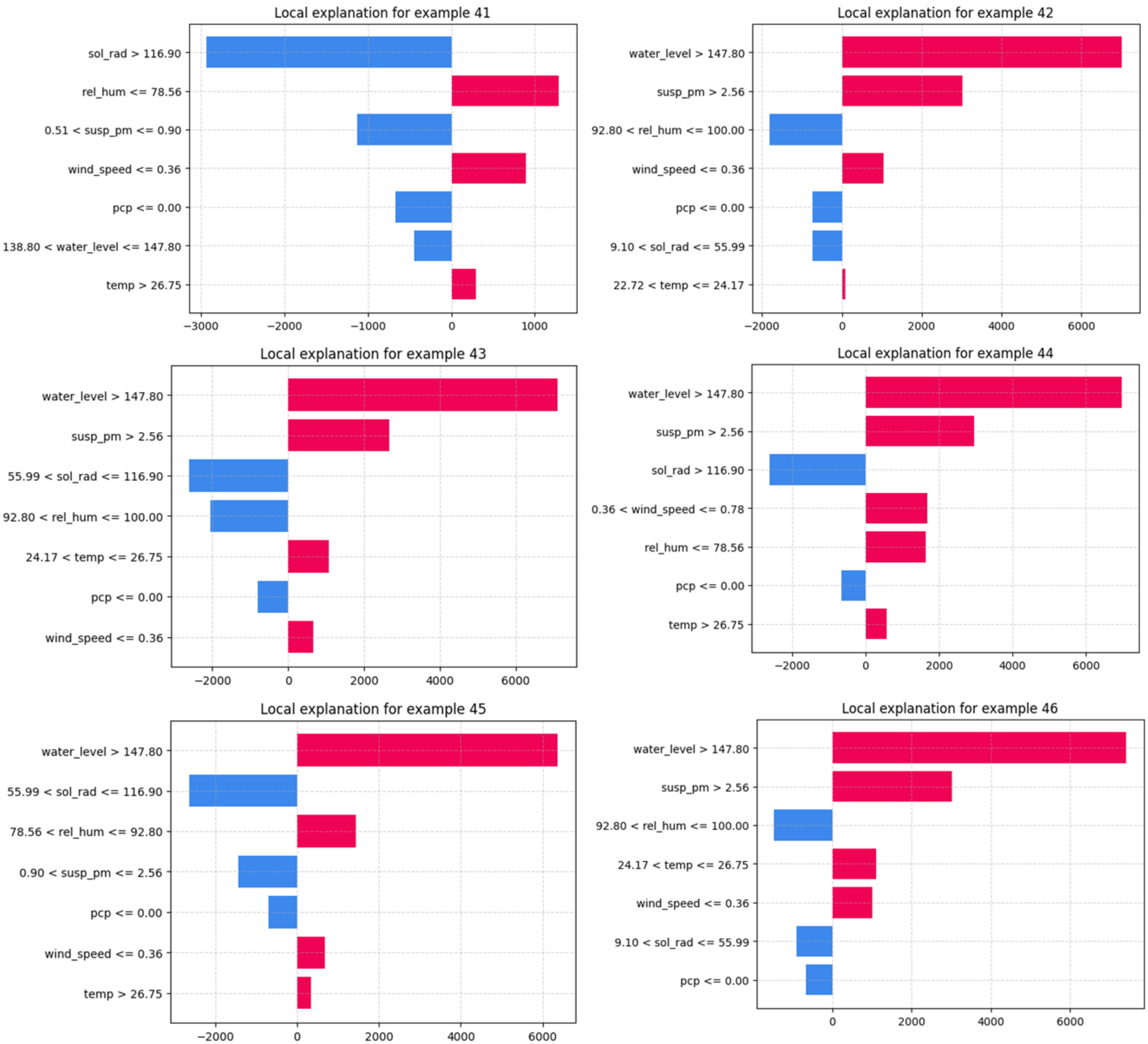


**Figure S10.** Explanation of XGBoost model for *E. coli* prediction using LIME method for six selected examples from test data. The explanations show the importance for each input feature by the model.

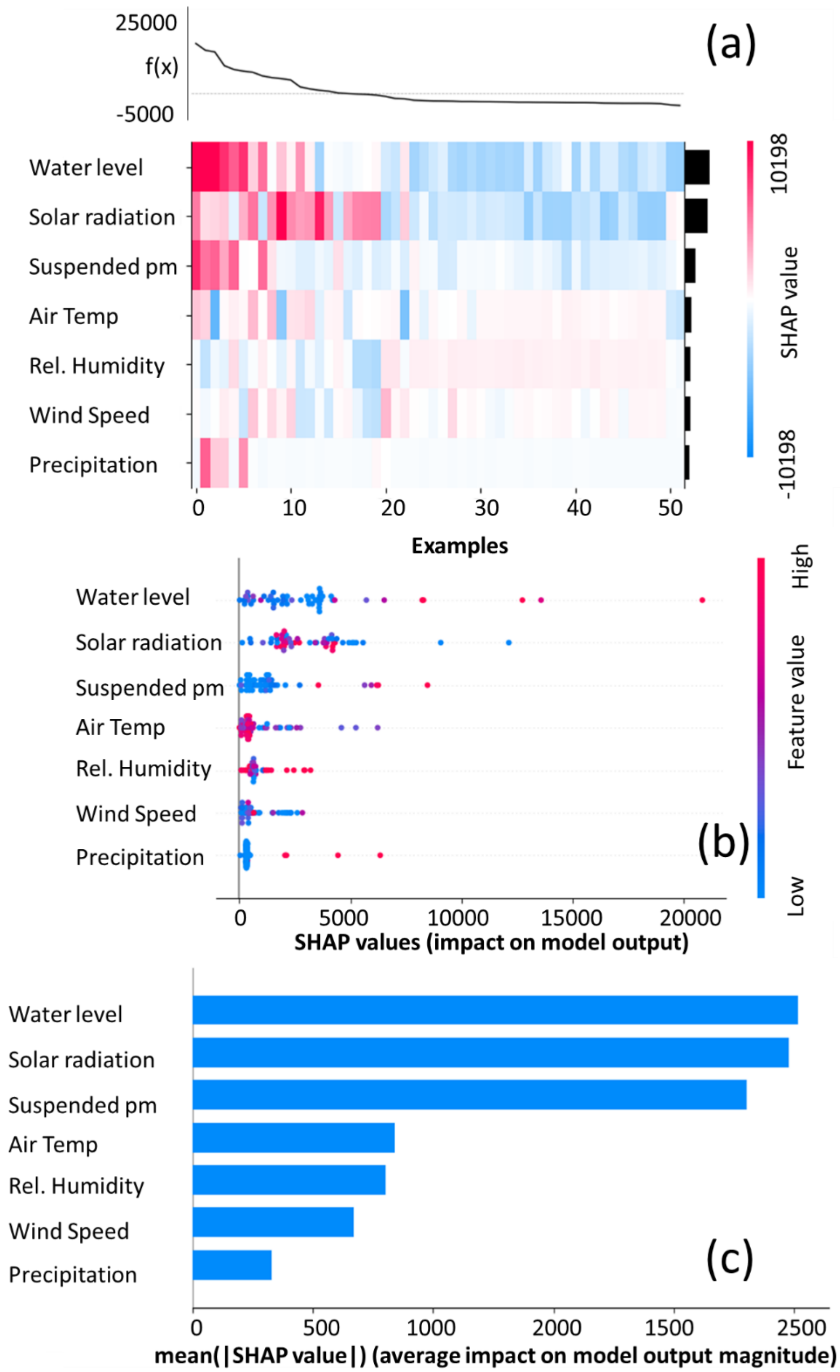


**Figure S11.** Explanation of XGBoost model for *E. coli* prediction using SHAP method. The explanations show the importance for each input feature by the model. (a) SHAP values as heat map. $f(x)$ indicates sum of SHAP values of all input features. The prediction of machine learning model is equal to sum of $f(x)$ and base value. The base value is the sum of models' prediction on training data which was 4661.082. (b) SHAP values for individual examples in test data, (c) global feature importance based upon SHAP values.

**Table S1:** Hydrological response unit (HRU) definitions supported in *AI4Water*. For using a specific HRU definition, corresponding required shape files must be provided.

| HRU definitions | Input shape file requirement |
|---|---|
| Unique land use | Land use |
| Unique sub-basin | Sub-basin delineation |
| Unique slope class | Slope classes |
| Unique soil type | Soil types |
| Unique land use with unique sub-basin | Land use, sub-basin delineation |
| Unique land use with unique soil type | Land use, soil types |
| Unique land use in unique slope class | Land use, slope classes |
| Unique soil type in unique slope class | Soil types, slope classes |
| Unique soil type in unique sub-basin | Soil types, sub-basin delineation |
| Unique slope class in unique sub-basin | Slope classes, sub-basin delineation |
| Unique land use with unique sub-basin and with unique soil type | Land use, sub-basin delineation, soil types |
| Unique land use with unique sub-basin and with unique slope | Land use, sub-basin delineation, slope classes |
| Unique soil type with unique slope class and unique sub-basin | Soil types, slope classes, sub-basin delineation |

**Table S2:** Names of evapotranspiration methods implemented in *AI4Water*. The 'name' column consists of the name used in *AI4Water*. The last columns give the input data requirement for a given method.

| Evapotranspiration Method | Name | Input parameters |
|---|---|---|
| Abtew (Abtew, 1996) | Abtew | temperature |
| Albrecht (Albrecht, 1950) | Albrecht | vapor pressure, temperature |
| Blaney Criddle (McMahon et al., 2013) | BlaneyCriddle | relative humidity, sunshine hours, wind |
| Brutsaert Strickler (Brutsaert and Stricker, 1979) | BrutsaertStrickler | temperature, vapor pressure, wind, radiation |
| Chapman (Chapman, 2003) | ChapmanAustralia | temperature, relative humidity, radiation, wind |
| Granger Gray (Granger and Gray, 1989) | GrangerGray | temperature, vapor pressure, wind, radiation |
| Hamon (Hamon, 1964) | Hamon | temperature |
| Haude (Haude, 1954) | Haude | temperature |
| Hargreaves Samani (Hargreaves and Samani, 1985) | HargreavesSamani | temperature |
| Jensen and Haise (Jensen and Haise, 1965) | JensenHaise | temperature |
| Kharrufa (Kharrufa, 1985) | Kharrufa | temperature |
| Linacre (Linacre, 1977) | Lincacre | temperature |
| MattShuttleworth (Shuttleworth and Wallace, 2009) | MattShuttleworth | temperature |
| McGuinnesBorndne | McGuinnesBordne | temperature |
| Makkink (de Bruin, 1981) | Makkink | temperature, relative humidity |
| Penman Monteith (Allen et al., 1998) | PenmanMonteith | temperature, relative humidity, radiation, wind, vapor pressure |
| Penman (Penman, 1948) | Penman | temperature, relative humidity, radiation, wind |
| Penman pan (Penman, 1956) | PanPan | temperature, relative humidity, radiation, wind |
| Priestley Taylor (Priestley and Taylor, 1972) | PriestelyTaylor | temperature, relative humidity, radiation |
| Romanenko (Romanenko V A, 1961) | Romanenko | temperature |

| | | |
|---|---|---|
| SzilagyiJozsa (Szilagyi et al., 2011) | SzilagyiJozsa | temperature, relative humidity, radiation, wind |
| Turc (Turc, 1961) | Turc | temperature, relative humidity |

**Table S3:** Exhaustive list of performance metrics which are calculated using *SeqMetrics* module in *AI4Water*. The second column gives the name of function in *SeqMetrics* module.

| **Metric name** | **Function name** |
|---|---|
| Absolute Percent Bias | abs_pbias |
| Agreement Index (Willmott, 1981) | agreement_index |
| Aitchison Distance (Zhang et al., 2020) | aitchison |
| Akaik's Information Criterion (Akaike, 1998) | aic |
| Alpha decomposition of NSE (Gupta et al., 1998) | nse_alpha |
| Anomaly correction coefficient (Langland and Maue, 2012) | acc |
| Bias (Gupta et al., 1998) | bias |
| Beta decomposition of NSE (Gupta et al., 2009) | nse_beta |
| Bounded NSE (Mathevet et al., 2006) | nse_bound |
| Bounded KGE (Mathevet et al., 2006) | kge_bound |
| Brier Score (BRIER, 1950) | brier_score |
| Correlation Coefficient | corr_coeff |
| Coefficient of determination | r2 |
| Centered root mean square | centered_rms_dev |
| Covariance | covariance |
| Decomposed mean square error | dcecomposed_mse |
| Explained Variance Score | exp_var_score |
| Euclid Distance | euclid_distance |
| Geometric Mean Difference | gmean_diff |
| Geometric Mean Relative Absolute Error | game |
| Geometric Mean Absolute Error | gmrae |
| Inertial Roost Squared Error | irse |
| Integral Normalized Root Squared Error | inrse |
| Inter-percentile normalized RMSE | nrmse_ipercentile |
| Jensen-Shannon divergence | JS |
| Kling Gupta Efficiency (Gupta et al., 2009) | kge |
| Legate-McCabe Efficiency Index | lm_index |
| Logarithmic NSE | log_nse |
| Logarithmic Probability distribution | log_prob |
| Maximum Error | max_error |
| Mean absolute percentage error (de Myttenaere et al., 2016) | mape |
| Mean absolute percentage deviation | mapd |
| Mean absolute error | mae |

| | |
|---|---|
| Mean absolute relative error (Despotovic et al., 2015) | mare |
| Mean absolute scaled error (Hyndman, 2006) | mase |
| Mean arctangle absolute percentage error | maape |
| Mean bias error (Willmott and Matsuura, 2006) | mean_bias_error |
| Mean Bounded relative error | mbrae |
| Mean error | me |
| Mean gamma deviance | mean_gamma-deviance |
| Mean log error | mle |
| Mean normalized RMSE | nrmse_mean |
| Mean percentage error | mpe |
| Mean Poisson deviance | mean_poisson_deviance |
| Mean relative absolute error | mrae |
| Mean squared error | mse |
| Mean square logarithmic error | msle |
| Mean variance | mean_var |
| Mean absolute error | mae |
| Median absolute percentage error | med_bs_error |
| Median bias error (Despotovic et al., 2015) | mbe |
| Median error | mde |
| Median relative absolute error | mdrae |
| Median squared error | med_seq_error |
| Mielke-Berry R (Mielke, 2007) | mb_r |
| Modified Agreement of Index | mod_agreement_index |
| Modified KGE (Kling et al., 2012) | kge_mod |
| Modified NSE (Krause et al., 2005) | nse_mod |
| Nash-Sutcliff Efficiency | nse |
| Non-parametric KGE (Pool et al., 2018) | kge_np |
| Normalized absolute error | norm_ae |
| Normalized absolute percentage error | norm_ape |
| Normalized Euclid distance | norm_euclid_distance |
| Normalized RMSE | nrmse |
| Peak flow bias of flow duration curve (Yilmaz et al., 2008) | fdc_fhv |
| Pearson correlation coefficient (Pearson, 1895) | pearson_r |
| Percent Bias (Moriasi et al., 2015) | pbias |
| Range normalized RMSE | nrmse_range |
| Refined agreement of index (Willmott et al., 2012) | ref_agreement_index |

| | |
|---|---|
| Relative Agreement of index | rel_agreement_index |
| Relative absolute error | rae |
| Relative RMSE | relative_rmse |
| Relative NSE | relative_nse |
| Root Mean Square Error | rmse |
| Root mean square log error (Abimbola et al., 2020) | rmsle |
| Root mean square percentage error | rmspe |
| Root mean square scaled error | rmsse |
| Root median squared scaled error | rmsse |
| Root relative squared error | rrse |
| RSR (Moriasi et al., 2015) | rsr |
| Spearman correlation coefficient (Zhang et al., 2020) | spearman_corr |
| Skill score of Murphy (Murphy, 1988) | skill_score_murphy |
| Spectral Angle | sa |
| Spectral correlation | sc |
| Spectral gradient angle | sga |
| Spectral information divergence | sid |
| Symmetric Kullback-Leibler divergence | KLsym |
| Symmetric mean absolute percentage error | smape |
| Symmetric median absolute percentage error | smdape |
| Sum of squared errors | sse |
| Volumetric efficiency | ve |
| Volume Error (Reynolds et al., 2017) | vol_error |
| Unscaled mean bounded relative absolute error | umbrae |
| Watterson’s M | watt_m |
| Weighted mean absolute percent errors | wmape |
| Weighted Absolute percentage error | wape |

**Table S4:** Names of classical machine learning (ML) algorithms supported by *AI4Water*

| Abbreviation of ML algorithm | Complete name of algorithm |
|---|---|
| XGBoostRegressor | XGBoost regression (Chen and Guestrin, 2016) |
| NuSVR | Nu Support Vector Regression |
| RandomForestRegressor | Random Forest Regression ((Liaw and Wiener, 2002)) |
| XGBoostRFRegressor | XGBoost Random Forest Regression (Chen and Guestrin, 2016) |
| ExTraTreesRegressor | Extra-tree Regression (Geurts et al., 2006) |
| HistGradientBoostingRegressor | Histogram based gradient boosting regression Tree |
| SVR | Epsilon-Support Vector Regression (Platt and others, 1999) |
| CATBoostRegressor | CatBoost Regression (Prokhorenkova et al., 2018) |
| LGBMRegressor | Light Gradient Boosting Machine (Ke et al., 2017) |
| GradientBoostRegressor | Gradient Boosting for Regression (Friedman, 2001) |
| ADABoostRegressor | AdaBoost Regressor (Freund and Schapire, 1997) |
| BaggingRegressor | Bagging Regressor (Ho, 1998) |
| ExtraTreeRegressor | |
| KNeighborsRegressor | Nearest Neighbors Regressor |
| LassoLarsCV | Ridge Regression with Cross Validation |
| LassoLarsIC | Lasso model fit with Lars using BIC or AIC for model selection (Zou et al., 2007) |
| Ridge | Linear least square with L2 Regularization |
| RidgeCV | Linear least square with L2 Regularization and Cross Validation |
| BayesianRidge | Bayesian Ridge Regression (Tipping, 2001) |
| KernelRidge | Kernel Ridge Regression (Murphy, 2012) |
| LassoCV | Linear model with L1 regularization |
| ElasticNetCV | |
| LinearRegression | Ordinary Least Square linear regression |
| ARDRegression | Bayesian regression model with ARD (MacKay, 1994) prior |
| DecisionTreeRegressor | Decision Tree regression |
| TweedieRegressor | Generalized linear model with Tweedie distribution |
| PoissonRegressor | Generalized linear model with Poisson distribution |
| Lars | Least Angle Regression |
| LarsCV | Least Angle Regression with Cross validation |
| OrthogonalMatchingPursuitCV | |
| HuberRegressor | Linear Regression robust to outliers |
| LinearSVR | Linear Support Vector Regression |
| GaussianProcessRegressor | |

| | |
|---|---|
| RansacRegressor | Random Sample Consensus (Choi et al., 2009) |
| OrthogonalMatchingPursuit | Orthogonal Matching Pursuit (Rubinstein et al., 2008) |
| OneClassSVM | One Class Support Vector Machine(Chang and Lin, 2011) |
| MLPRegressor | Multi-layer perceptron regression |
| TheilsenRegressor | Theil-Sen Estimator (Dang et al., 2009) |
| SGDRegressor | Linear Model with Stochastic Gradient Descent |
| LassoLars | |

**Table S5:** File types and their extensions accepted by *AI4Water*.

| File extension | File type |
| --- | --- |
| **.csv** | Comma separated file |
| **.xlsx** | Microsoft Excel |
| **.npz** | Numpy zipped file |
| **.parquet** | Parquet |
| **.feather** | Feather |
| **.nc** | netCDF5 |
| **.mat** | MATLAB |

## **References**:


Abimbola, O. P., Mittelstet, A. R., Messer, T. L., Berry, E. D., Bartelt-Hunt, S. L., and Hansen, S. P.: Predicting Escherichia coli loads in cascading dams with machine learning: An integration of hydrometeorology, animal density and grazing pattern, 722, https://doi.org/10.1016/j.scitotenv.2020.137894, 2020.

Abtew, W.: Evapotranspiration measurements and modeling for three wetland systems in south Florida, 32, https://doi.org/10.1111/j.1752-1688.1996.tb04044.x, 1996.

Akaike, H.: Information Theory and an Extension of the Maximum Likelihood Principle, https://doi.org/10.1007/978-1-4612-1694-0_15, 1998.

Albrecht, F.: Die Methoden zur Bestimmung der Verdunstung der natürlichen Erdoberfläche, 2, https://doi.org/10.1007/BF02242718, 1950.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration – Guidelines for computing crop water requirements. – FAO Irrigation and drainage paper 56 / Food and Agriculture Organization of the United Nations., 1998.

BRIER, G. W.: VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY, 78, https://doi.org/10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2, 1950.

de Bruin, H. A. R.: The determination of (reference crop) evapotranspiration from routine weather data., 28, 1981.

Brutsaert, W. and Stricker, H.: An advection-aridity approach to estimate actual regional evapotranspiration, 15, https://doi.org/10.1029/WR015i002p00443, 1979.

Chang, C. C. and Lin, C. J.: LIBSVM: A Library for support vector machines, 2, https://doi.org/10.1145/1961189.1961199, 2011.

Chapman, T. G.: Estimation of evaporation in rainfall-runoff models, 2003.

Chen, T. and Guestrin, C.: XGBoost: A scalable tree boosting system, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, https://doi.org/10.1145/2939672.2939785, 2016.

Choi, S., Kim, T., and Yu, W.: Performance evaluation of RANSAC family, in: British Machine Vision Conference, BMVC 2009 - Proceedings, https://doi.org/10.5244/C.23.81, 2009.

Dang, X., Peng, H., Wang, X., and Zhang, H.: The Theil-Sen Estimators in a Multiple Linear Regression Model, 2009.

Despotovic, M., Nedic, V., Despotovic, D., and Cvetanovic, S.: Review and statistical analysis of different global solar radiation sunshine models, https://doi.org/10.1016/j.rser.2015.08.035, 2015.

Freund, Y. and Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, 55, https://doi.org/10.1006/jcss.1997.1504, 1997.

Friedman, J. H.: Greedy function approximation: A gradient boosting machine, 29, https://doi.org/10.1214/aos/1013203451, 2001.

Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, 63, https://doi.org/10.1007/s10994-006-6226-1, 2006.

Granger, R. J. and Gray, D. M.: Evaporation from natural nonsaturated surfaces, 111, https://doi.org/10.1016/0022-1694(89)90249-7, 1989.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, 34, https://doi.org/10.1029/97WR03495, 1998.

Gupta, H. v., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, 377, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hamon, W. R.: Computation of direct runoff amounts from storm rainfall, 1964.

Hargreaves, G. H. and Samani, Z. A.: Reference Crop Evapotranspiration from Temperature, 1, https://doi.org/10.13031/2013.26773, 1985.

Haude, W.: Zur praktischen Bestimmung der aktuellen und potentiellen Evaporation und Evapotranspiration, Schweinfurter Dr. und Verlages, 1954.

Ho, T. K.: The random subspace method for constructing decision forests, 20, https://doi.org/10.1109/34.709601, 1998.

Hyndman, R. J.: Another look at forecast-accuracy metrics for intermittent demand, 4, 2006.

Jensen, M. E. and Haise, H. R.: Closure to "Estimating Evapotranspiration from Solar Radiation," 91, https://doi.org/10.1061/jrcea4.0000342, 1965.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T. Y.: LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems, 2017.

Kharrufa, N. S.: Simplified equation for evapotranspiration in arid regions, 5, 1985.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, 424–425, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, 5, https://doi.org/10.5194/adgeo-5-89-2005, 2005.

Langland, R. H. and Maue, R. N.: Recent Northern Hemisphere mid-latitude medium-range deterministic forecast skill, 64, https://doi.org/10.3402/tellusa.v64i0.17531, 2012.

Liaw, A. and Wiener, M.: Classification and Regression with Random Forest, 2, 2002.

Linacre, E. T.: A simple formula for estimating evaporation rates in various climates, using temperature data alone, 18, https://doi.org/10.1016/0002-1571(77)90007-3, 1977.

MacKay, D. J. C.: Bayesian nonlinear modeling for the prediction competition, in: ASHRAE Transactions, https://doi.org/10.1007/978-94-015-8729-7_18, 1994.

Mathevet, T., Michel, C., Andréassian, V., and Perrin, C.: A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins, 2006.

McMahon, T. A., Peel, M. C., Lowe, L., Srikanthan, R., and McVicar, T. R.: Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: A pragmatic synthesis, 17, https://doi.org/10.5194/hess-17-1331-2013, 2013.

Mielke, P. W. Jr. , B. K. J.: Permutation methods: a distance function approach, 2nd ed., Springer Science & Business Media., 2007.

Moriasi, D. N., Gitau, M. W., Pai, N., and Daggupati, P.: Hydrologic and water quality models: Performance measures and evaluation criteria, 58, https://doi.org/10.13031/trans.58.10715, 2015.

Murphy, A. H.: Skill scores based on the mean square error and their relationships to the correlation coefficient, 116, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, 1988.

Murphy, K. P.: Machine Learning - A Probabilistic Perspective - Table-of-Contents, 2012.

de Myttenaere, A., Golden, B., le Grand, B., and Rossi, F.: Mean Absolute Percentage Error for regression models, 192, https://doi.org/10.1016/j.neucom.2015.12.114, 2016.

Pearson, K.: VII. Note on regression and inheritance in the case of two parents, 58, https://doi.org/10.1098/rspl.1895.0041, 1895.

Penman, H. L.: Natural evaporation from open water, bare and grass, 193, 1948.

Penman, H. L.: Evaporation: an introductory survey., 4, https://doi.org/10.18174/njas.v4i1.17768, 1956.

Platt, J. and others: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, 10, 1999.

Pool, S., Vis, M., and Seibert, J.: Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency, 63, https://doi.org/10.1080/02626667.2018.1552002, 2018.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, 100, https://doi.org/10.1175/1520-0493(1972)100<0081:otaosh>2.3.co;2, 1972.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A.: Catboost: Unbiased boosting with categorical features, in: Advances in Neural Information Processing Systems, 2018.

Reynolds, J. E., Halldin, S., Xu, C. Y., Seibert, J., and Kauffeldt, A.: Sub-daily runoff predictions using parameters calibrated on the basis of data with a daily temporal resolution, 550, https://doi.org/10.1016/j.jhydrol.2017.05.012, 2017.

Romanenko V A: Computation of the autumn soil moisture using a universal relationship for a large area., in: Proc. of Ukrainian Hydrometeorological Research Institute, 12–25, 1961.

Rubinstein, R., Zibulevsky, M., and Elad, M.: Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit, 2008.

Shuttleworth, W. J. and Wallace, J. S.: Calculating the water requirements of irrigated crops in Australia using the matt-shuttleworth approach, 52, 2009.

Szilagyi, J., Jozsa, J., and Kovacs, A.: A Calibration-Free Evapotranspiration Mapping (CREMAP) Technique, in: Evapotranspiration, https://doi.org/10.5772/14277, 2011.

Tipping, M. E.: Sparse Bayesian Learning and the Relevance Vector Machine, 1, https://doi.org/10.1162/15324430152748236, 2001.

Turc, L.: Estimation of Irrigation water requirements, potential evapotranspiration: a simple climatic formula evolved up to date., 12, 1961.

Willmott, C. J.: On the validation of models, 2, 1981.

Willmott, C. J. and Matsuura, K.: On the use of dimensioned measures of error to evaluate the performance of spatial interpolators, 20, https://doi.org/10.1080/13658810500286976, 2006.

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, 32, https://doi.org/10.1002/joc.2419, 2012.

Yilmaz, K. K., Gupta, H. v., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, 44, https://doi.org/10.1029/2007WR006716, 2008.

Zhang, M., Shi, W., and Xu, Z.: Systematic comparison of five machine-learning models in classification and interpolation of soil particle size fractions using different transformed data, 24, https://doi.org/10.5194/hess-24-2505-2020, 2020.

Zou, H., Hastie, T., and Tibshirani, R.: On the “degrees of freedom” of the lasso, 35, https://doi.org/10.1214/009053607000000127, 2007.