Efficient modelling of water temperature patterns in river systems – benchmarking a set of machine learning approaches

Mathew Herrnegger[1] (mathew.herrnegger@boku.ac.at), Moritz Feigl[1], Katharina Lebiedzinski[1] and Karsten Schulz[1]

[1]Institute for Hydrology and Water Management (HyWa), University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

EGU General Assembly 2020

1 Motivation

Statistical approaches (e.g. regression models) are widely used for modelling river water temperature, but their flexibility of mapping processes and dynamics within the catchment (e.g. energy fluxes) is rather limited. Here, **machine learning (ML) methods** are providing approaches, which may **consider processes and dynamics within a catchment** by learning from data. Recent studies just showed that Artificial Neural Network approaches outperform the commonly used multiple regression approaches (Graf et al., 2019). This contribution goes one step further by analysing a set of different ML methods regarding their applicability by using different input data sets and catchments.

2 Data & Methods

Study area and input data sets

The ML approaches are tested on 10 catchments listed in Table 1. They have different characteristics, human impacts (e.g. hydropower, river regulation) and time series lengths (10 to 39 years). The catchment outlets are situated in the Austrian Alps or flatlands with areas ranging from 200 to 96.000 km² (see Figure 1).

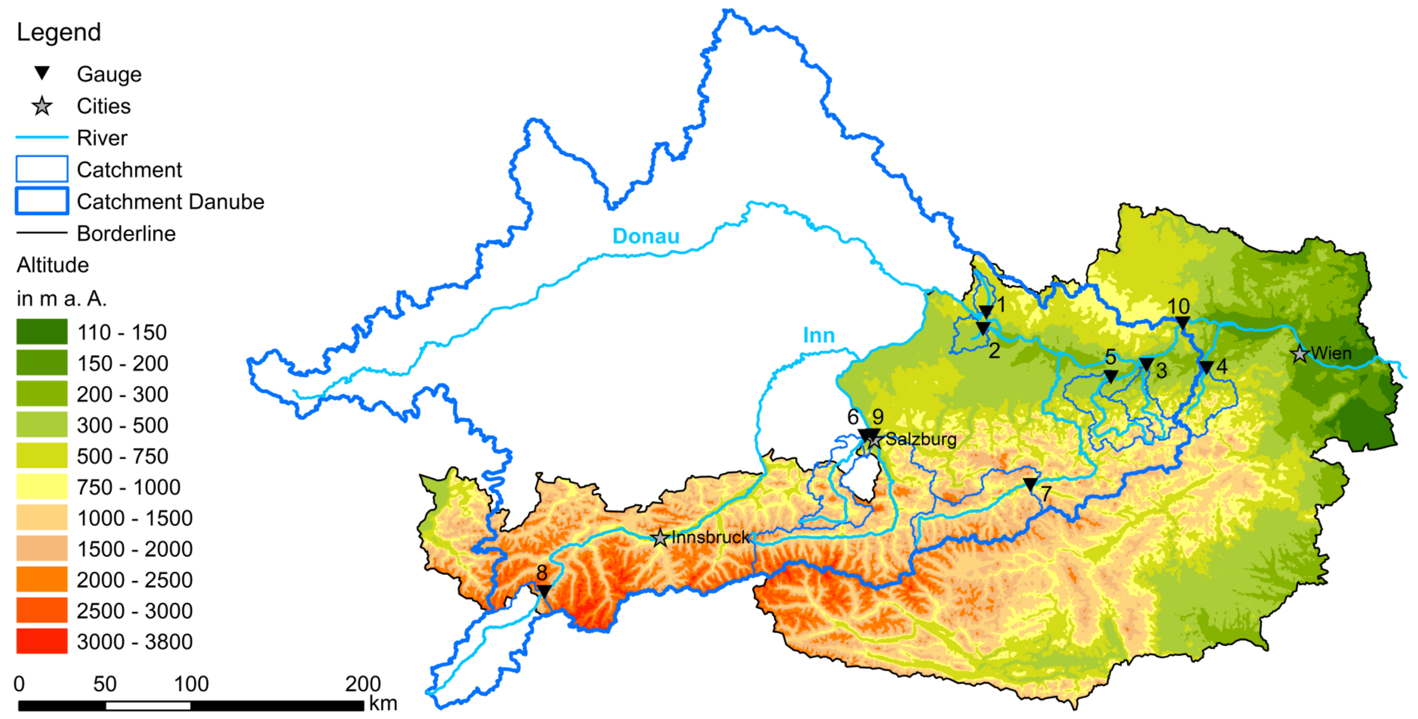

Figure 1: Overview map showing the study area in Austria, Switzerland and Germany.

Table 1: Catchment properties including mean daily river water temperature (WT mean), mean daily discharge (Q mean), mean daily precipitation (Prec. mean) and mean daily global radiation (Glob. Rad. mean).

ID	River	Gauge	HZB-Nr.	Timeseries	Years without gaps	Area km²	WT mean °C	Q mean m³/s	T mean °C	Prec. mean mm	Glob. Rad. mean W/m²
1	Kleine Muehl	Obermühl	204883	2002-2015	14,0	200,2	8,87	3,12	9	2,73	135
2	Aschach	Kropfmühle	205054	2004-2015	11,9	312,2	10,78	3,80	10	2,50	136
3	Erlauf	Niederndorf	207803	1980-2015	35,3	604,9	9,42	15,27	8	3,59	127
4	Traisen	Windpassing	207910	1998-2015	17,7	733,3	9,83	14,88	8	3,33	131
5	Ybbs	Greimpersdorf	207688	1981-2015	34,7	1.116,6	9,87	31,50	8	3,77	127
6	Saalach	Siezenheim	203570	2000-2015	16,0	1.139,1	8,50	39,04	7	4,60	135
7	Enns	Liezen	210799	2006-2015	10,0	2.116,2	1,19	67,56	6	3,60	137
8	Inn	Kajetansbrücke	201178	1997-2015	18,8	2.162,0	6,00	59,26	0	2,56	153
9	Salzach	Salzburg	203398	1977-2015	39,0	4.425,7	7,63	178,11	5	4,16	136
10	Donau	Kienstock	207357	2005-2015	11,0	95.970,0	10,77	1.798,31	10	2,13	131
Source	HZB	HZB	HZB	HZB	HZB	HZB	HZB	HZB	SPARTACUS	SPARTACUS	INCA (2007+)

Data preprocessing

Observed data variables are grouped to simple and advanced sets of input data in Table 2 to analyse possible data dependencies. Since river water temperature is largely controlled by processeses within the catchment, it can be described as a function of catchment properties. Hence, only water temperature is used from a point measurement (gauge) at the catchment outlet. Input variables with an integral effect over the catchment are aggregated to catchment means. Additionally, lags of air temperature and discharge were considered (1-4 days). The time was included as input by using fuzzy months. They are equal 1 on the 15th of each month and linearly decreaseing until they are zero on the 15th day of the previous and following month.

Table 2: Input data sets.

Input variable		Input data sets: simple	precipitation	radiation	all
T mean	mean daily air temperature (values, lags, differences)	x	x	x	x
Q mean	mean daily discharge (values, lags, differences)	x	x	x	x
Prec. mean	mean daily precipitation		x		x
Glob. Rad. Mean	mean daily global radiation			x	x
Time	fuzzy months	x	x	x	x

Models

The models compared in this study (see Figure 2) can be grouped into three categories:

(i) **Linear Models** are included in the investigation as a benchmark, since they are widely used in river water temperature modelling. We applied a multiple regression model with only air temperature and discharge as input (LM) and a regression model with step wise variable selection (LMstep). LMstep uses all data input sets and also interaction terms.

(ii) The applied **regression tree based models** are Random Forest (RF) (Breiman, 2001) and XGBoost (Chen & Guestrin, 2016). Random forest predictions rely on averageing an ensemble of regression tree predictions trained on a random subset of training data. XGBoost consists of an ensemble of regression trees where each tree aims to predict the residuals resulting from the prediction of the previous trees.

(iii) Three types of **Neural Networks** are applied: Feedforward Neural Networks (FNN) (Rosenblatt, 1961), Long short-term Models (LSTM) (Hochreiter & Schmidhuber, 1997) and Gated Recurrent Units (GRU) (Cho et al., 2014). FNNs are consisting of multiple fully connected hidden layers. LSTMs and GRUs are both Recurrent Neural Networks (RNN) that consists of nodes with internal states that can be used to model temporal sequences. GRUs have a simpler structure than LSTMs, as they include only one internal state whereas LSTM contain two internal states per node.

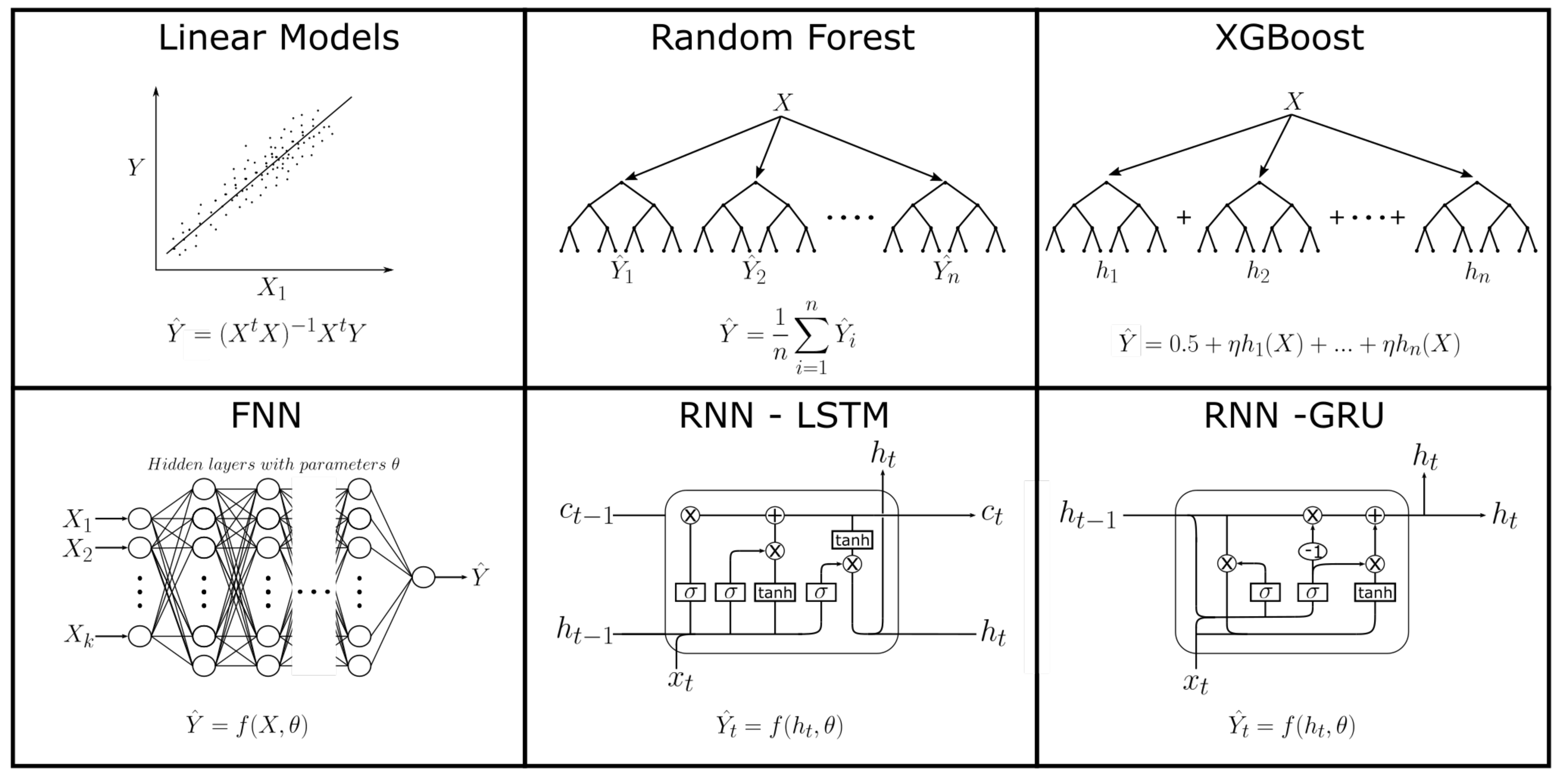

Figure 2: Overview of the applied models with $\hat{Y}$ denoting estimated water temperatures and X observed variables. $h_{1,...,n}$ defines the predicted residuals from individual XGBoost trees. $f(X,\theta)$ denotes a mapping from a FNN with the parameters θ. For a given time step, h_t denotes the hidden internal state of a RNN cell and c_t the internal cell state of a LSTM cell. RNNs consist of a cascade of cells, each feeding their internal states into the next cell, finally resulting in a single feedforward layer estimating $\hat{Y}$ from h_t.

Hyperparameter optimization and data splits

Hyperparameters are model parameters which values are set before the learning process is started. Therefore, to make the results of the models comparable and to ensure the best possible predictive ability, all hyperparameters in this study are optimized using the Bayesian Hyperparameter optimization. The activation functions for FNN was chosen to be SELU (Klambauer et al., 2017) activation function and was not optimized. The models are compared using their performance in the testing time period, which consists of the last 20% of the time series for each catchment. For hyperparameter optimization, the other 80% are further split into 60% training data and 20% validation data for the Neural Network models. The Linear Models and regression tree based models are validated using both 5 times repeated 10-fold Cross-validation and a time series Cross-validation.

3 Results & Discussion

Table 3 summarises the best performing models for each catchment. Our results demonstrate that the analysed **ML approaches outperforms LM methods** regardless of the catchment characteristics and input data set. The model RMSE indicate that modelling river water temperature with the applied set of ML approaches achieves **excellent prediction results** for all tested catchment sizes, but works best for catchments larger than approx. 300 km². The results does not determine a specific input data length, elevation or input data set to be most eligible. The best prediction results are achieved for the Alpine catchment Inn (RMSE 0,509 °C, NSE 0,975 °C) using XGBoost and all input data (see Figures 3 to 6). For this catchment, LM was applied with a simple data input, while LMstep, RF, FNN, GRU and LSTM showed best results using the precipitation input data set. Figure 6 is showing the ranked importance of input variables of the XGBoost model for catchment Inn. Although the best XGBoost model results were achieved using all input variables, the most important were mean daily air temperature on the day of prediction and the days before (lags). The introduced set of ML methods provide an attractive approach for **large-scale river temperature modelling**, where the requirements for using process-based models are not able to be met. The applied models will be introduced to the community as open source R package with a companion publication soon.

Table 3: Best performing model for each catchment compared to the standard LM model performance.

Catchment	Model	Input	Model RMSE °C	Model NSE °C	LM RMSE °C	LM NSE °C	Area km²
Kleine Muehl	RF	simple	0,795	0,979	1,734	0,899	200,2
Aschach	FNN	simple	0,709	0,987	1,764	0,922	312,2
Erlauf	XGBoost	simple	0,525	0,986	1,379	0,905	604,9
Traisen	FNN	all	0,578	0,979	1,261	0,912	733,3
Ybbs	FNN	simple	0,563	0,989	1,795	0,889	1116,6
Saalach	FNN	radiation	0,524	0,977	1,311	0,866	1139,1
Enns	FNN	radiation	0,520	0,979	1,421	0,842	2116,2
Inn	XGBoost	all	0,509	0,975	1,376	0,829	2162,0
Salzach	XGBoost	precip	0,514	0,979	1,332	0,862	4425,7
Donau	GRU	simple	0,582	0,989	2,146	0,842	95970,0

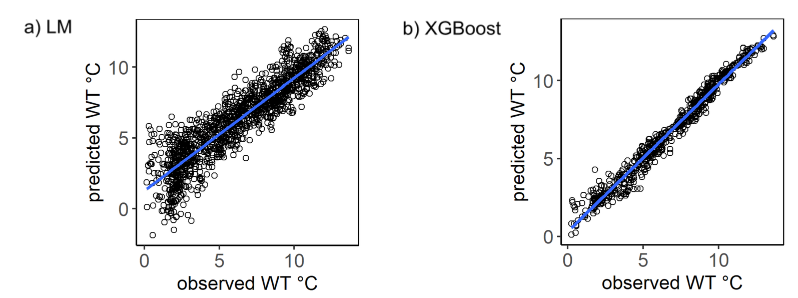

Figure 3: Correlation of observed and predicted mean daily river water temperatures for catchment Inn (dots) compared to a 1:1 line (blue line).

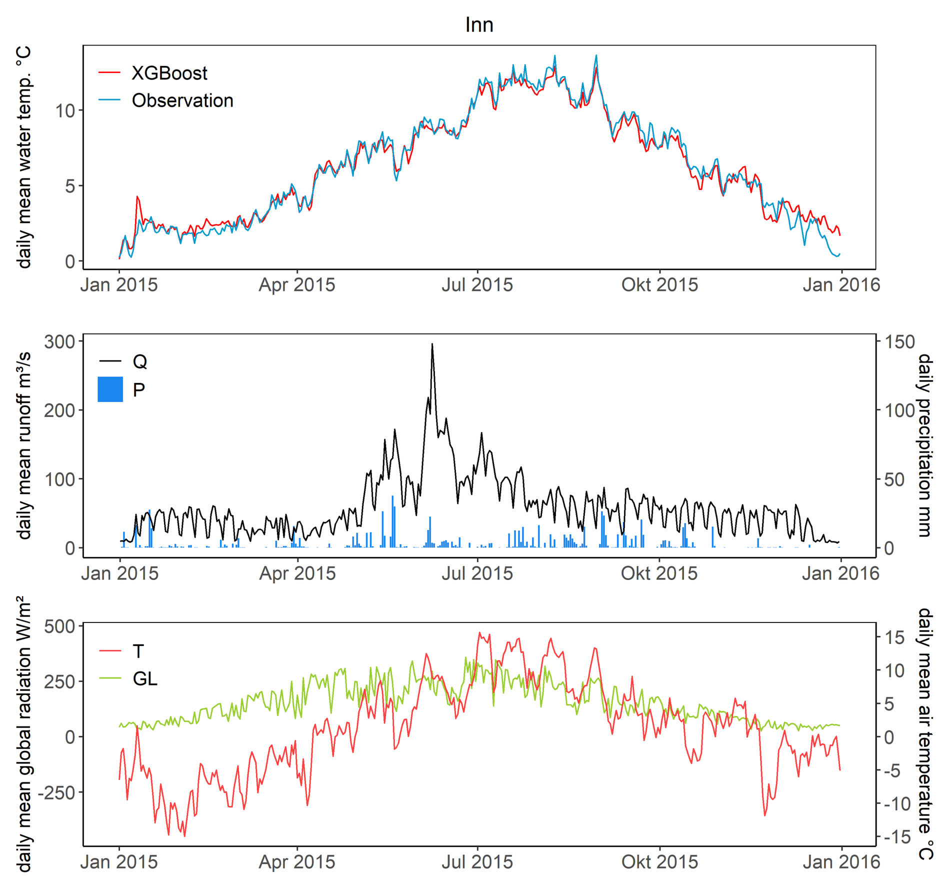

Figure 4: Time series of observed river water temperature, best model (XGBoost) and input variables for 2015 at catchment Inn.

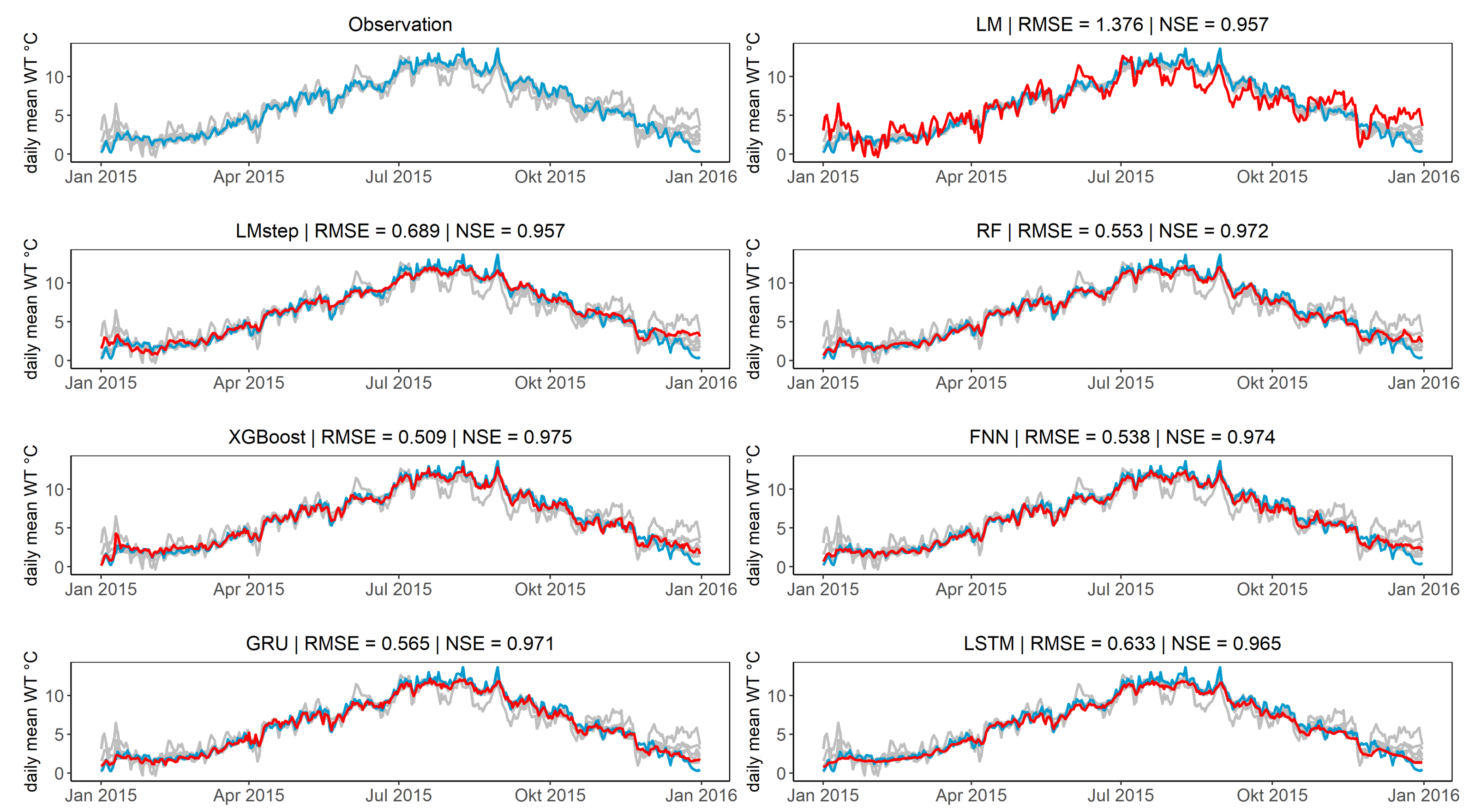

Figure 5: Comparison of observed river water temperature (blue line) and predicted river water temperature of the best perfoming models within each group (grey and red lines) for 2015 at catchment Inn.

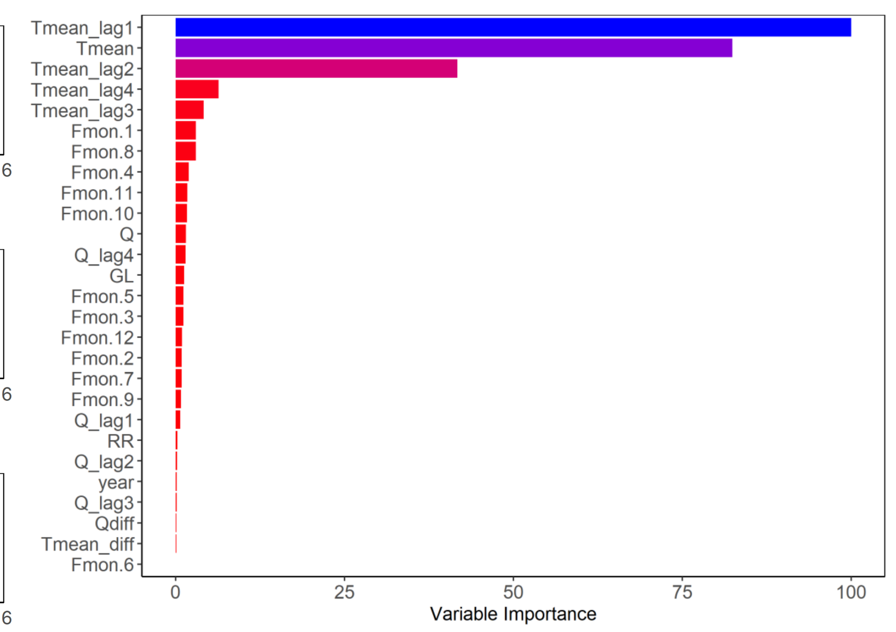

Figure 6: Importance plot showing the relevance of XGBoost variables for catchment Inn.

References

Breiman, L. (2001). Random Forests. Machine Learning, 45 (1), 5-32.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y.(2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Graf, R., Zhu, S. & Sivakumar, B. (2019). Rorecasting river water temperature time series using a wavelet-neural network hybrid modelling approach. Journal of Hydrology, 578, 124115.

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S.(2017). Self-normalizing neural networks. In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 972-981). Curran Associates Inc.

Rosenblatt, F. (1961). Principles of neurodynamics. perceptrons and the theory of brain mechanisms (No. VG-1196-G-8). Cornell Aeronautical Lab Inc Buffalo NY.