A Comparison of Data-Driven Models for Predicting Stream Water Temperature

Helen Weierbach[1], Aranildo R. Lima[2], Danielle Christianson[1], Boris Faybishenko[1], Val Hendrix[1], and Charuleka Varadharajan[1]

[1]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720
[2]Aquatic Informatics, Vancouver, BC, Canada

Abstract

Changes to the Earth's climate are expected to negatively impact water resources in the future. It is important to have accurate modelling of river flow and water quality to make optimal decisions for water management. Machine learning and deep learning models have become promising methods for making such hydrological predictions. Using these models, however, requires careful consideration both of data constraints and of model complexity for a given problem. Here, we use machine learning (ML) models to predict monthly stream water temperature records at three monitoring locations in the Northwestern United States with long-term datasets, using meteorological data as predictors. We fit three ML models: a Multiple Linear Regression, a Random Forest Regression, and a Support Vector Regression, and compare them against two baseline models: a persistence model and historical model. We show that all three ML models are reasonably able to predict mean monthly stream temperatures with root mean-squared errors (RMSE) ranging from 0.63-0.91 °C. Of the three ML models, Support Vector Regression performs the best with an error of 0.63-0.75 °C. However, all models perform poorly on extreme values of water temperature. We identify the need for machine learning approaches for predicting extreme values for variables such as water temperature, since it has significant implications for stream ecosystems and biota.

1 Introduction

The Earth's water resources are increasingly under stress due to changes in climate and land use. Current projections estimate decreases in water availability and worsening of water quality over the coming decades. There is an urgent need for accurate predictions of water quantity and quality at seasonal to decadal time-scales, and at local to regional spatial scales, to enable water managers to make better decisions with an uncertain future. A particular variable of importance is stream water temperature (WT), which affects many stream processes and also impacts fish and other aquatic biota [Heck et al., 2018].

Data-driven modeling of watershed variables such as discharge and groundwater levels that make use of large-scale meteorological and watershed data has become increasingly possible with growing amounts of publicly available water datasets [Kratzert et al., 2019, Müller et al., 2019]. Machine learning (ML) and deep learning models are also being increasingly considered for predictions of water quality, and specifically water temperatures due to the limitations of physics-based models to conduct these predictions and the ability of ML to reduce complexity of models particularly at large spatial scales [Carlisle et al., 2010, Zhu and Piotrowski, 2020, Jia et al., 2020].

Tackling Climate Change with Machine Learning workshop at NeurIPS 2020.

Making accurate water quality and other hydrological predictions presents several challenges since watersheds are complex systems. In particular, several factors can influence stream temperature including meteorological conditions, river discharge, snow melt, groundwater temperature, reservoir and thermal power plant operations. It is typically difficult to find co-located datasets for potential predictor variables. The choice of spatial and temporal scales for prediction is also dependent on constraints of data availability and stakeholder needs. These decisions are problem dependent, changing with both spatial and temporal scales.

Our goal is to determine if machine learning models can accurately predict water quality at the multiple spatial (point to basin-scale) and temporal (sub-daily to decadal-scale) scales that are relevant for disturbance events such as floods and droughts. Here, we start by using baseline ML models for predicting monthly water temperature at individual station locations using publicly available data. Understanding data dependencies with low-order models will help to better constrain the problem and give a low-complexity baseline comparison for predicting water quality at larger spatial scales and higher temporal resolutions. We pay special attention to how our models perform at predicting extreme values of water temperature. These extreme temperatures typically determine the suitability of the streams for fish habitat, and will be increasingly important since disturbances that adversely impact stream temperature in the short-term are projected to be more frequent with climate change.

2 Methods

We develop models that predict water temperature using data from the CAMELS (Catchment Attributes and Meteorology for Large Sample studies) dataset [Addor et al., 2017] for three stations in the Sandy River Basin. Stream temperature data for each station were retrieved from the United States Geological Survey (USGS) National Water Information System (NWIS). We select model input features (minimum and maximum air temperature, solar radiation, and month of year) through exploratory data analysis. Following data preprocessing, we fit five models: a Multiple Linear Regression (MLR), a Random Forest Regression (RF), a Support Vector Regression (SVR), a historical model, and a persistence model. The historical, persistence, and MLR models do not have any hyperparameters, however RF and SVR models both have several (such as the number of trees and the regularization parameter.) Hyperparameter tuning for these models is performed using a random search k-fold cross validation. More information on data, model construction and methods are located in the appendix.

3 Results

Table 1 shows the root mean squared error (RMSE), mean absolute error (MAE) and the R^2 value comparing observations to modeled results for each model at the three different stations for the test period. All error metrics are reported with hyperparameters tuned using cross-validation with RMSE and MAE in units °C. For RF, error is reported as the mean of 30 different initializations. The standard deviations of ensemble are around 0.002 for all stations and thus are not shown. The three statistical models (MLR, SVR, RF) outperform the baseline historical and persistence models with considerably lower RMSE and MAE for all three stations. SVR performs the best at each station with RMSE ranging from 0.63-0.75 °C. These results suggest that at monthly time scales, stream water temperature can be predicted well using primarily air temperature and solar radiation data. The ability to predict stream temperature at the monthly resolution with meteorological data can greatly increase locations where predictions can be made by reducing the need for other co-located observations.

Other studies predicting daily water temperature using data driven approaches such as Artificial Neural Networks, Bootstrap Aggregated Decision Trees, and Multilayer Perceptron neural networks [Zhu et al., 2018] have shown optimal RMSE around 0.5-3 °C at individual stations. While our results are not directly comparable with prior modeling results due to differences in temporal resolution, our results suggest that regression ML models are appropriate for monthly station-based predictions.

Despite relatively good model performance (R^2 0.9 or greater), we find model error is consistently high for extreme values of water temperature (Appendix, Figure 1). For example, for the Fir Creek station, all models tend to over-estimate temperature for low temperature values (below 4 °C), and underestimate the high temperature extremes (above 12 °C). Peaks and troughs of water temperature, as observed in the time-series, are mostly missed by all models for all stations (Appendix, Figure 2).

Table 1: Model Errors (°C). The predictions with the least error are highlighted in bold.

Station	**MLR**	**SVR**	**Random Forest**	**Persistence**	**Historical**
Fir Creek RMSE	0.70	**0.63**	0.72	1.81	0.80
Fir Creek MAE	0.55	**0.49**	0.54	1.51	0.65
Fir Creek R^2	0.95	**0.96**	**0.96**	-	-
South Fork RMSE	0.91	**0.75**	0.82	2.10	0.95
South Fork MAE	0.71	**0.59**	0.64	1.78	0.76
South Fork R^2	0.93	**0.95**	**0.95**	-	-
North Fork RMSE	0.91	**0.70**	0.76	1.58	0.78
North Fork MAE	0.72	**0.53**	0.58	1.30	0.61
North Fork R^2	0.89	**0.94**	0.92	-	-

4 Conclusions and Future Work

We have used regression-based ML models to predict monthly water temperatures with long-term time-series data at individual monitoring stations. At monthly frequencies, SVR and RF models predict water temperature relatively well with RMSE ranging from 0.63-0.86 °C, depending on the station and model used. However, we find that the models perform less accurately at extreme values of stream temperature, for many possible reasons including the lack of adequate training observations and classical model error minimization approaches. We also find that discharge is not an important input feature at the monthly resolution for these locations, but may be important for other locations or daily predictions.

Our next steps are to expand the use of the models to greater spatial coverage and temporal scales. This will require training models at more locations, increasing the temporal resolution to predict daily temperatures, and using more complex architectures (if needed) with additional input features. We will also examine model generalizability by fitting one model for multiple locations. At the daily frequency, our exploratory data analysis has shown that seasonal meteorological and discharge data have both auto-correlations and cross-correlations with water temperature that we will have to consider. To make daily predictions, we will also integrate more complex gap filling methods to enable use of stream temperature datasets with longer gaps (> 3 months). Our preliminary results using the regression models for daily temperature predictions suggest that more input features and likely lagged variables are required. We will also investigate use of more complex models such as Multilayer Perceptrons, Long Short-Term Memory Networks, and Convolutional Neural Networks using the approach outlined in Müller et al. [2019]. We will compare the outcomes of these predictions with those from process-based and ML models of stream temperature. Finally, we will test the sensitivity of our models to different meteorological data sources and input features. Our preliminary results indicate that monthly model predictions are not significantly sensitive to changes in different meteorological forcing data or combinations of input features.

Broader Impact

The overarching goal of this research is to determine whether data-driven models can predict water quality responses of streams to disturbance events such as floods and droughts. These extreme events occur across a variety of spatial and temporal scales- from local to regional, and from days to months to years. Our preliminary results examine how ML models perform at the monthly time scale, illustrating that although baseline regression-based ML models can reasonably predict mean monthly temperatures, they perform poorly for extremes values. By improving methodology, scaling our problem up to multiple locations and using more complex models we hope to accurately predict stream temperatures across a range of spatiotemporal scales, although new mathematical approaches may still be needed to capture the behavior following extreme events. Comparing models at different scales and with different complexities will be helpful toward predicting water quality during disturbances at decision-relevant spatial and temporal scales. These predictions are increasingly important to the health of watersheds in the face of climate change, where disturbances such as floods and droughts will become more frequent.

Acknowledgments and Disclosure of Funding

This material is based upon work supported by the Early Career Research Program funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under the Berkeley Lab Contract Number DE-AC02-05CH11231.

References

Nans Addor, Andrew J. Newman, Naoki Mizukami, and Martyn P. Clark. The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10):5293–5313, October 2017. ISSN 1027-5606. doi: https://doi.org/10.5194/hess-21-5293-2017. URL https://hess.copernicus.org/articles/21/5293/2017/. Publisher: Copernicus GmbH.

Daren M. Carlisle, James Falcone, David M. Wolock, Michael R. Meador, and Richard H. Norris. Predicting the natural flow regime: models for assessing hydrological alteration in streams. *River Research and Applications*, 26(2):118–136, 2010. ISSN 1535-1467. doi: 10.1002/rra.1247. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rra.1247. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rra.1247.

Michael P Heck, Luke D Schultz, David Hockman-Wert, Eric C Dinger, and Jason B Dunham. Monitoring stream temperatures—a guide for non-specialists. Technical report, US Geological Survey, 2018.

Xiaowei Jia, Jacob Zwart, Jeffery Sadler, Alison Appling, Samantha Oliver, Steven Markstrom, Jared Willard, Shaoming Xu, Michael Steinbach, Jordan Read, and Vipin Kumar. Physics-guided recurrent graph networks for predicting flow and temperature in river networks, 2020.

Frederik Kratzert, Daniel Klotz, Guy Shalev, Günter Klambauer, Sepp Hochreiter, and Grey Nearing. Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine-Learning Applied to Large-Sample Datasets. *arXiv:1907.08456 [cs, stat]*, November 2019. URL http://arxiv.org/abs/1907.08456. arXiv: 1907.08456.

Juliane Müller, Jangho Park, Reetik Sahu, Charuleka Varadharajan, Bhavna Arora, Boris Faybishenko, and Deborah Agarwal. Surrogate optimization of deep neural networks for groundwater predictions. *Journal of Global Optimization*, pages 1–29, 2019.

AJ Newman, MP Clark, Kevin Sampson, Andrew Wood, LE Hay, A Bock, RJ Viger, D Blodgett, L Brekke, JR Arnold, et al. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous usa: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1):209, 2015.

C. Varadharajan, D. A. Agarwal, W. Brown, M. Burrus, R. W. H. Carroll, D. S. Christianson, B. Dafflon, D. Dwivedi, B. J. Enquist, B. Faybishenko, A. Henderson, M. Henderson, V. C. Hendrix, S. S. Hubbard, Z. Kakalia, A. Newman, B. Potter, H. Steltzer, R. Versteeg, K. H. Williams, C. Wilmer, and Y. Wu. Challenges in building an end-to-end system for acquisition, management, and integration of diverse data from sensor networks in watersheds: Lessons from a mountainous community observatory in east river, colorado. *IEEE Access*, 7:182796–182813, 2019.

Senlin Zhu and Adam P. Piotrowski. River/stream water temperature forecasting using artificial intelligence models: a systematic review. *Acta Geophysica*, September 2020. ISSN 1895-7455. doi: 10.1007/s11600-020-00480-7. URL https://doi.org/10.1007/s11600-020-00480-7.

Senlin Zhu, Emmanuel Karlo Nyarko, and Marijana Hadzima-Nyarko. Modelling daily water temperature from air temperature for the missouri river. *PeerJ*, 6:e4894, 2018.

Appendices

A Additional Methods and Results

We develop models that predict water temperature using data from the CAMELS (Catchment Attributes and Meteorology for Large Sample studies) dataset [Addor et al., 2017]. Stream temperature data for each station were retrieved from the United States Geological Survey (USGS) National Water Information System (NWIS) using our custom data integration tool BASIN-3D (Broker for Assimilation, Synthesis and Integration of eNvironmental Diverse, Distributed Datasets), which provides a generic framework to synthesize diverse, multiscale data across a variety of additional data sources and environmental data types [Varadharajan et al., 2019].

The CAMELS product consists of daily meteorological, discharge, and catchment attribute data for monitoring stations in 672 pristine catchments across the continental U.S from 1980-2014. For this analysis, we use Daymet meterological data at 1 km resolution that is mapped to the point stations in the CAMELS dataset [Newman et al., 2015]. Although we chose the CAMELS product since it has compiled and quality checked data for the predictor variables, less than 50 of the 671 basins in the CAMELS dataset have associated water temperature records, of which only three stations in Oregon had near-complete stream temperature records from 1980-2014, with minimal gaps (<3 months). We selected these three stations for our models, which are all located in the Sandy River basin of the Northwestern United states – Fir creek near Brighton, OR (USGS HUC: 14138870), South Fork of Bull Run River near Bull Run, OR (USGS HUC: 14139800) and North Fork of Bull Run River near Multnomah Falls, OR (USGS HUC: 14138900).

We perform exploratory data analysis to understand the seasonality of the meteorological and discharge variables and calculate time-lagged cross-correlations of water temperature with deseasonalized versions of the variables to understand which variables are important at the monthly resolution. Minimum and maximum air temperature and solar radiation are highly correlated with water temperature, with a strong seasonal influence, and hence were chosen as input features for the ML models and along with month of the year (indicative of seasonality). These input features are also identified by Zhu and Piotrowski [2020] as common and meaningful features in predicting river and stream temperatures for other modeling efforts. Notably, for these stations discharge and precipitation did not have significant cross-correlations with water temperature.

Data are preprocessed by re-sampling to monthly frequency (using monthly means) and performing minimal gap filling (< 3 months) by linear interpolation for one station. Data are next split into training and test data at a 70/30 split. A standard scalar is fit separately to the training data for both the input features ($\mathbf{X}$) and predictors ($\mathbf{y}$). This scalar is used to transform the data to zero mean and standard variance for training (and again after predicting to transform back to the original units).

Following the data preprocessing, we fit five models: a Multiple Linear Regression (MLR), a Random Forest Regression (RF), a Support Vector Regression (SVR), and baseline historical and persistence models. The historical, persistence, and MLR models do not have any hyperparameters, however RF and SVR methods both have several (such as the number of trees and the regularization parameter). Hyperparameter selection is performed using a random search k-fold cross validation provided by

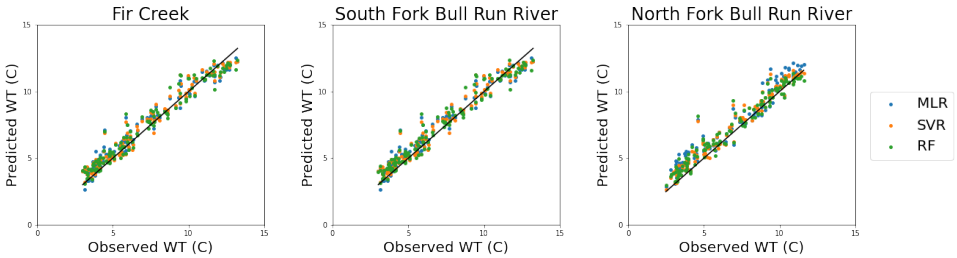

Figure 1: Observed (x-axis) vs. predicted (y-axis) water temperature for MLR, RF and SVR models at each of the three stations. Mean values are evenly distributed above/below the 1:1 line while extreme values are more skewed.

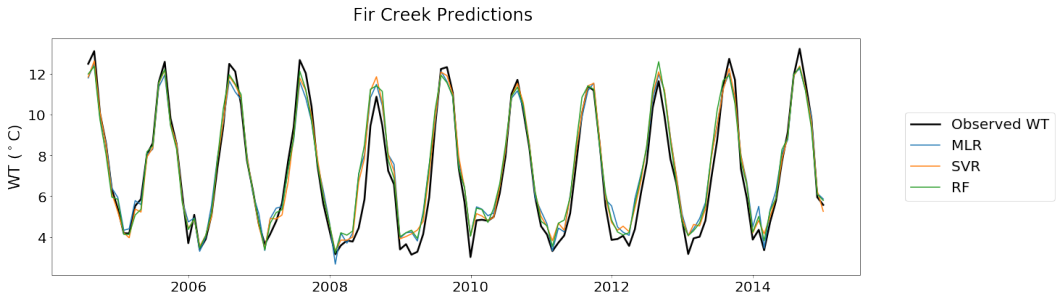

Figure 2: Example time series for Fir Creek station. Observed (blue), and predicted water temperature for each model is shown for the test period (07/2004-12/2014). MLR predictions (blue), SVR (orange) and RF (green) all tend to perform well at mean conditions but have difficulty capturing extremes.

sci-kit learn. For each of 5 folds, a random search is conducted across a pre-specified grid of model hyperparameters. The parameters selected by cross-validation are next used to predict data in the test region of the time series and calculate model error.

Observed and predicted stream temperature for each ML model and station are shown in Figure 1. A time series of model fits for one station (Fir Creek) is shown in Figure 2. Peaks and troughs of water temperature are mostly missed by all three models for all three stations.