# Supplementary Material

# Machine Learning Models for Prediction of Shade-affected Stream Temperatures

Efrain Noa-Yarasca[1], Meghna Babbar-Sebens[1], Chris E. Jordan [2]

[1] School of Civil & Construction Engineering, Oregon State University, Corvallis, OR 97331, USA

[2] Conservation Biology Division, NWFSC, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2032 SE OSU Dr., Newport, OR, 97365 USA.

## Introduction

This supplementary material complements the manuscript titled "Machine Learning Models for Prediction of Shade-affected Stream Temperatures." The study assesses the selection of predictor variables for machine learning (ML) models predicting stream temperatures influenced by riparian shading. It explores two model development scenarios: *Forecasting ML Models* for predicting shade-affected stream temperatures at monitored locations and *Interpolation ML Models* for predicting at unmonitored locations using data from nearby monitored gauges. This document is organized into three appendices. Appendix 1 illustrates the categorization of predictor variables utilized in the study. Appendix 2 outlines the process of computing the "shade factor," incorporating key variables, diagrams, equations, and results. Appendix 3 provides in-depth tables and figures, offering detailed insights into specific aspects of the research results.

**Supplementary material Appendix 1**

## 1. Categorization of predictor variables

**Table S1**: *Variables categorized into four classes—weather, stream & watershed characteristics, seasonality, and site-location characteristics—corresponding to each component of the mass and energy transfer process.*

| Class of variable | Sub-class | Mass transfer | | | | Energy transfer | | |
|---|---|---|---|---|---|---|---|---|
| | | Surface flow | Lateral flow | Goundwater | Snow flow | Shortwave radiation | Longwave radiation | Evaporation & conduction |
| Weather | | Precipitation | Precipitation | Precipitation | Max air temp | Solar radiation | Air temperature | Air temperature |
| | | Air temperature | Air temperature | Groundwater temp. | Humidity | | Humidity | Wind speed |
| | | Streamflow* | | | Wind speed | | | |
| | | | | | Solar radiation | | | |
| Stream & watershed characteristics | Watershed | Drainage area | Drainage area | | % Snow area | | | |
| | | | slope | | | | | |
| | Stream | Length (stream) | Length slope | | | Stream aspect | | |
| | | Slope | | | | Channel width | | |
| | | n Manning channel | | | | water depth | | |
| | | Sinuosity | | | | | | |
| | Land cover | Land cover | | | | Riparian vegetation | | |
| | | n Manning overland | | | | (Tree height) | | |
| | Soil | Hydrologic group | Hydrologic group | Baseflow recession | | | | |
| | | | Available water storage | Drainage class | | | | |
| | | | Drainage class | Available water storage | | | | |
| Seasonality | Year | | | | Day of the year | Day of the year | | |
| Site location | | | | | Maritime location | | | |
| | | | | | Latitude | Latitude | | |
| | | | | | Longitude | Longitude | | |
| | | | | | Elevation | | | |

**Supplementary material Appendix 2**

## 2. Shade factor modeling

### 2.1. Shade factor approach

The shade factor was calculated as the ratio of blocked solar radiation (due to topography and riparian vegetation) to potential solar radiation that would reach the stream surface. Key variables in the equations are outlined below.

**Solar angle and solar azimuth**

The solar angle is measured between the observer's horizon and the sun. It is a function of the stream latitude, declination of the sun, and the time of the day.

$$\alpha = \sin^{-1}(\sin\phi . \sin\delta + \cos\phi . \cos\delta . \cos\tau) \quad \text{(S1)}$$

$$\delta = 23.45 \left(\frac{2\,\pi}{360}\right) cos\left(\frac{2\,\pi\,(172 - JD)}{365}\right) \quad \text{(S2)}$$

$$\tau = \left(180 - long - t_m - (360\, hr/24)\right)(2\pi/360) \quad \text{(S3)}$$

Where: $\alpha$ is the solar altitude (solar angle), $\phi$ is the stream latitude, $\delta$ is the declination of the sun, $\tau$ is the local hour angle of the sun, $JD$ is the Julian day (1-365), $long$ is the stream longitude, $t_m$ is the local time zone meridian (degrees), and $hr$ is the hour of the day. These equations are explained in depth by Boyd (Boyd 2003).

The solar azimuth is the angle formed by the north and the horizontal projection of the sun (on the observer's horizon) measured clockwise.

$$Sun_{az} = \cos^{-1}\left(\frac{\sin\delta - \sin\alpha .\ \sin\phi}{\cos\alpha .\ \cos\phi}\right) \quad \text{(S4)}$$

Stream azimuths were measured from the north to the stream center line in the flow direction.

**Sub daily solar radiation**

Solar radiation for sub-daily time scales was obtained using the Kaplanis approach (Kaplanis, 2006; Khatib & Elmenreich, 2015). This approach proposes solar radiation at any time as a cosine function limited by the sunrise and sunset and conditioned to the day.

$$h_{ij} = a.n_j + b.n_j\, cos\,(\frac{2\,\pi\, t_{SS}}{24}) \quad \text{(S5)}$$

Where $h_{ij}$ is the solar radiation at any time within the day, $t_i$ is the time in hours, $n_j$ is the Julian day, and $a$ and $b$ are coefficients determined for any site and any day. The sub-daily solar radiation was determined for time intervals of 0.01 hours and accumulated during the day (from sunrise to sunset).

**Shadow over the stream**

The length of the shadow (Laz) (either by riparian or the topography) parallel to the solar azimuth, and length of the shadow normal to the streamflow are obtained by geometry (**Figure S1**).

$$Laz = \frac{h\_tree}{tan\,(\alpha)} \quad \text{(S6)}$$

$$L_n = L_{az} \,.sin\,(sun_{az} - strm_{az}) \quad \text{(S7)}$$

Where $h\_tree$ is the tree height in riparian vegetation, $\alpha$ is the solar angle, $sun_{az}$ is the solar azimuth, and $strm_{az}$ is the stream azimuth.

The normal shadow was then multiplied by the stream length. Thus, three shading scenarios on the stream can be observed: no shadow over the stream, partial shadow over the stream, and full shadow over the stream. In this calculation, the shade factor corresponding to the topography, left bank and right bank (defined in the direction of flow) has been identified and then calculated separately. The sum of these three components was the total SF at the control station.

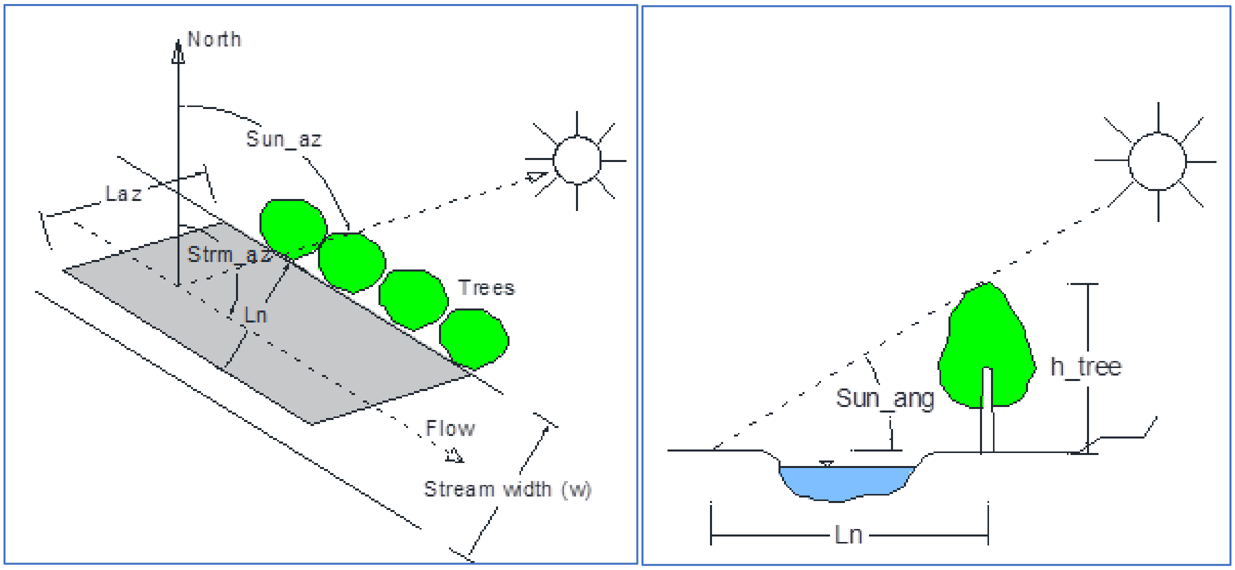
$$Shade = Strm_{length} \,.L_n \quad \text{(S8)}$$

Finally, the shade factor for each day and each station was obtained by dividing the accumulated amount of blocked solar radiation by the potential solar radiation representing the solar heat flux (both diffuse and direct beam) that would reach the stream surface without barriers.

$$SF_{ijk} = \frac{\sum_{k=t_{sr}}^{t_{ss}} Shade_{ijk} \,.h_{ijk}}{L_j.W_j\, H_{ijk}} \quad \text{(S9)}$$

Where $i$ indicates the station or sub-basin, $j$ is the day in the year (from 1 to 365), $k$ is the time in the day, $Shade_{ijk}$ is the shade of the barrier on stream, $h_{ijk}$ is the solar radiation at the time $k$, $L_j$ is the stream length, $W_j$ is the surface water width determined by the no linear equation suggested by Allen et at (Allen et al., 1994), $H_{ijk}$ is the registered daily solar radiation.

**Figure S1**. *Diagram showing the variables used to calculate the length of the shadow parallel to the azimuth (Laz) and perpendicular to the streamline (Ln).*

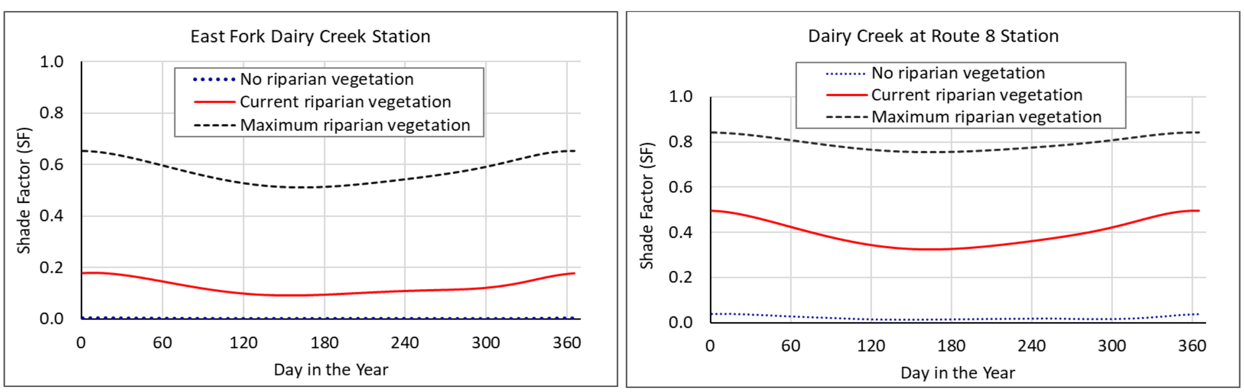



This calculation process was developed in the Python environment, which is available at https://github.com/noayarae/SF_model.git.

**2.2. Shade factor results**

In all three scenarios (No-riparian, current-riparian, and full-riparian vegetation), shade factors show similar patterns over the year with slightly higher SF values in winter than in summer. Shade factors in the non-riparian vegetation scenario are practically zero. Shade factors in the full riparian vegetation scenario are greater than the SFs in the current scenario (**Figure S2**).

**Figure S2**: *Annual shade factor variations for (a) East Fork Dairy Creek station (left) and (b) Dairy Creek at Route 8 station (right), considering three riparian vegetation scenarios: No riparian, current riparian, and maximum riparian vegetation.*



**Supplementary material Appendix 3**

**Table S2.** *Top five models in best subset selection for EFDC and DCR8 stations across three ML models. Models with four predictors share common predictors, though not always in the same order.*

| # variables | XGB - ML model | MLP - ML model | CNN - ML model |
|---|---|---|---|
| **East Fork Dairy Creek Station** | | | |
| 1 | Ta | Ta | Ta |
| 2 | Ta, SF | Ta, flow | Ta, DY |
| 3 | Ta, SF, flow | Ta, flow, SF | Ta, DY, flow |
| 4 | Ta, SF, flow, DY | Ta, flow, SF, DY | Ta, DY, flow, SF |
| 5 | Ta, SF, flow, DY, pp | Ta, flow, SF, DY, SR | Ta, DY, flow, SF, pp |
| **Dairy Creek at Route 8 Station** | | | |
| 1 | Ta | Ta | Ta |
| 2 | Ta, SF | Ta, DY | Ta, DY |
| 3 | Ta, SF, DY | Ta, DY, SF | Ta, DY, SF |
| 4 | Ta, SF, DY, flow | Ta, DY, SF, flow | Ta, DY, SF, flow |
| 5 | Ta, SF, DY, flow, SR | Ta, DY, SF, flow, SR | Ta, DY, SF, flow, SR |

**Table S3.** *Variables included in the 6-predictor Interpolation ML Model.*

| # variables | Set of predictors |
|---|---|
| 1 | Ta |
| 2 | Ta, W_elev |
| | Ta, S_elev |
| | Ta, hg_cd |
| | Ta, SR |
| | Ta, DY |
| 3 | Ta, SR, W_elev |
| | Ta, SR, hg_cd |
| | Ta, DY, W_elev |
| | Ta, W_elev, SR |
| | Ta, hg_cd, SR |
| 4 | Ta, SR, hg_cd W_elev |
| | Ta, SR, hg_cd str_wd |
| | Ta, S_elev, RH W_elev |
| | Ta, S_elev, RH str_wd |
| | Ta, SR, W_elev str_wd |
| 5 | Ta, SR, hg_cd W_elev, DY |
| | Ta, S_elev, RH W_elev, str_wd |
| | Ta, SR, hg_cd W_elev, hg_b |
| | Ta, SR, hg_cd W_elev, dc_3 |
| | Ta, SR, hg_cd W_elev, aws_150 |
| 6 | Ta, SR, hg_cd W_elev, DY, RH |
| | Ta, SR, hg_cd W_elev, DY, W_area |
| | Ta, SR, hg_cd W_elev, DY, W_slp |
| | Ta, SR, hg_cd W_elev, DY, str_len |
| | Ta, SR, hg_cd W_elev, DY, str_wd |

**Table S4:** *Variables featured in models with six predictors and their frequency of appearance*.

| Variable | Variable code | # of times in the model | Type of data |
|---|---|---|---|
| Air temperature | Ta | 26 | Weather data |
| Watershed elevation | W_elev | 17 | Stream/watershed feature |
| Solar Radiation | SR | 17 | Weather data |
| hydrological soil group C/D | hg_cd | 14 | Soil data |
| Day in the Year | DY | 8 | Seasonality |
| Stream width | str_wd | 5 | Stream/watershed feature |
| Site elevation | S_elev | 4 | Site-location feature |
| Relative Humidity | RH | 4 | Weather data |
| hydrological soil group B | hg_b | 1 | Soil data |
| Somewhat poorly drained soil | dc_3 | 1 | Soil data |
| Available water storage | aws_150 | 1 | Soil data |
| Watershed area | W_area | 1 | Stream/watershed feature |
| Watershed slope | W_slp | 1 | Stream/watershed feature |
| Stream length | str_len | 1 | Stream/watershed feature |

**Table S5***: Categories of variables in six-predictor models, along with their respective frequencies.*

| Type of variable | # of times in the model |
|---|---|
| Weather data | 21 |
| Stream/watershed features | 25 |
| Soil data | 17 |
| Seasonality | 8 |
| Site-location features | 4 |

**Figure S3***: DMW map displaying RMSE values achieved by the optimal Interpolation ML model with six predictors.*

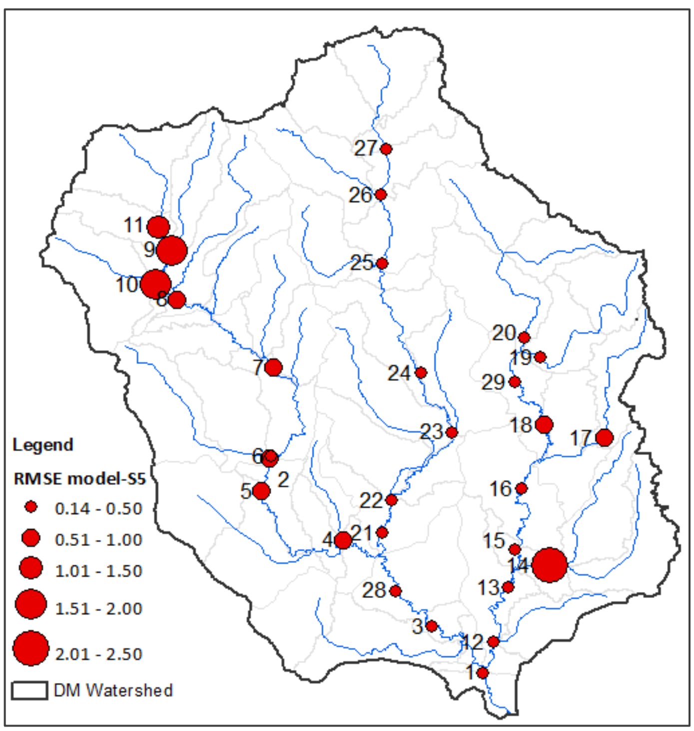


**Figure S4:** *Validation of stream temperature forecasting at East Fork Dairy Creek (top) and Dairy Creek at Route 8 (bottom) stations.*

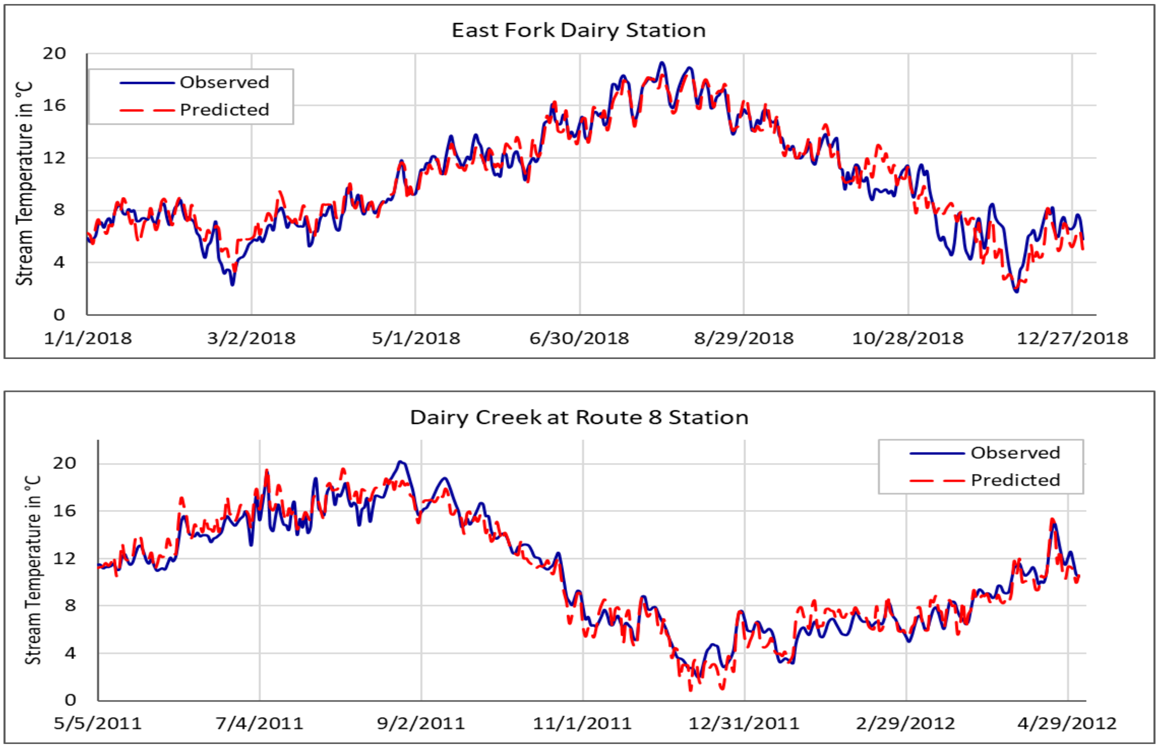
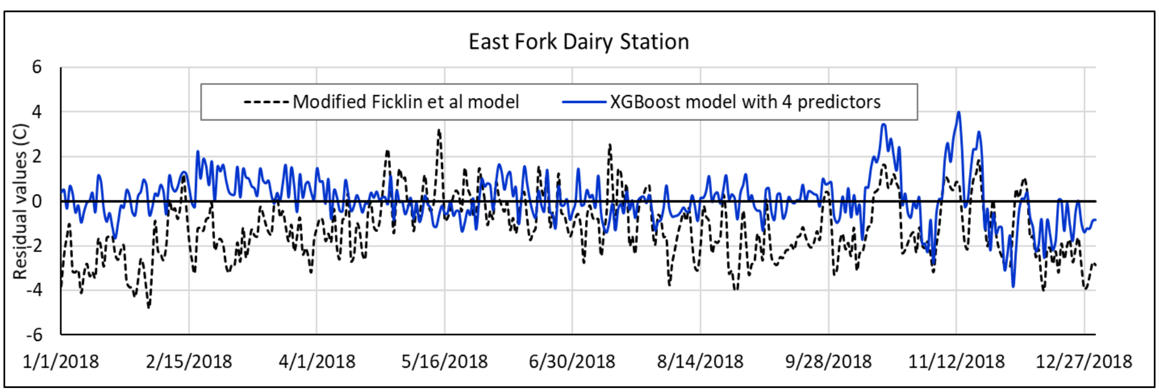



**Figure S5:** *Residual stream temperature comparison between the XGB model with four predictors and the modified Ficklin et al. model at East Fork Dairy station (top) and Dairy Creek Route 8 station (bottom).*



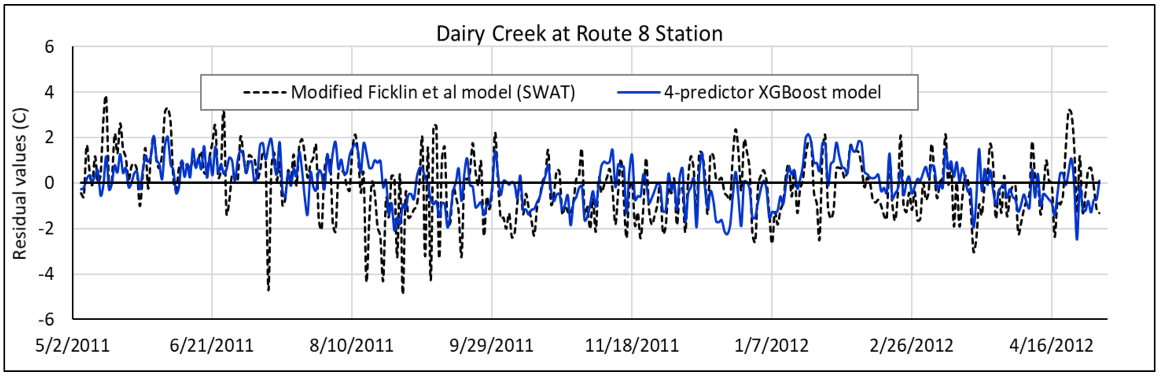



**Figure S6.** *Predicted stream temperature for EFDC (top) and DCR8 (bottom) stations in summer validation, comparing different riparian conditions. Observed temperature shown as dashed line.*