Portland State University

PDXScholar

Master of Environmental Management Project
Reports

Environmental Science and Management

2018

Development of an Effective Shade Model for Water Quality Management in Oregon

Erin Costello
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/mem_gradprojects

Part of the [Water Resource Management Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Costello, Erin, "Development of an Effective Shade Model for Water Quality Management in Oregon" (2018). *Master of Environmental Management Project Reports*. 39.
https://pdxscholar.library.pdx.edu/mem_gradprojects/39
<https://doi.org/10.15760/mem.42>

This Project is brought to you for free and open access. It has been accepted for inclusion in Master of Environmental Management Project Reports by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Development of an effective shade model for water quality management in Oregon

by

Erin Costello

A report submitted in partial fulfillment of the requirements for the degree of
Masters of Environmental Management

Committee:

Dr. Eugene Foster

Dr. Yangdong Pan

Dr. Daniel Sobota

Ryan Michie

The Department of Environmental Science and Management

Portland State University

2018

Abstract

Warming stream temperatures degrade water quality by stressing cold-water species, promoting harmful algal blooms, lowering dissolved oxygen, and increasing the toxicity of some compounds. The influence of stream temperature on the survival of native salmon is of particular concern in the Pacific Northwest. Effective shade, the fraction of solar radiation blocked from reaching a stream, is directly related to in-stream temperature and can be manipulated by riparian management actions. Mechanistic models of effective shade are prohibitively data and cost intensive. The objective of this research was to develop an empirical model (Simple Shade) for evaluating effective shade across the state of Oregon over time on a consistent and cost effective basis. Model inputs included Normalized Difference Vegetation Index (NDVI) values derived from high-resolution aerial imagery, watershed area, topographic shading, and channel characteristics derived from digital elevation models (DEMs) and a training dataset of effective shade output from the process-based shade model Heat Source. The study area was the Middle Fork Coquille River, Oregon, and imagery data was collected in summer 2009. Principal components analysis and random forest models were used to analyze variable importance and predict effective shade. The Simple Shade model explained 45% of variation in effective shade and had a mean of squared residuals of 2.7% of effective shade units. The most important predictor variables were watershed area, mean NDVI and topographic shade angle to the South. This research demonstrated that an empirical model of effective shade has the potential to match the performance of a mechanistic model of effective shade. The correlation between shade and mean NDVI also suggested that NDVI could be a surrogate measure of effective shade. However, variability in imagery and longitudinal watershed characteristics likely obscured the relationship between NDVI and effective shade. An improved technique for normalizing image and longitudinal variability within a watershed is needed to implement Simple Shade as a tool for assessing riparian conditions across Oregon and elsewhere.

Acknowledgements

Thank you to Dr. Gene Foster for the resources, insight and guidance that made this project possible and even fun. Thank you to Ryan Michie for sharing your technical expertise and for pushing me to produce meaningful work. Thank you to Dr. Dan Sobota for being a sounding board for this project and for providing essential feedback throughout the report writing process. Thank you to Colin Donald for technical assistance, scripting wizardry and cubicle commiseration. Thank you to the rest of the Oregon DEQ Watershed Management Section for making space for me in their busy schedules and for their genuine interest in the project. Thank you to my academic advisor Dr. Yangdong Pan for his incisive critiques of my statistical analysis. And lastly, thank you to my sweetie Tommy Yacoe for supporting my pursuit of knowledge and purpose with love and enthusiasm.

Table of Contents

Abstract	i
Acknowledgements	ii
List of Figures	v
List of Tables	vi
1. Introduction	1
1.1 Stream Temperature	1
1.1.1 Drivers of warming	2
1.1.2 Impacts of warming in the Pacific Northwest	3
1.2 Water quality regulation	4
1.3 Effective shade	4
1.3.1 Measuring effective shade in the field	6
1.3.2 Modeling effective shade	9
1.4 NDVI and effective shade	10
1.5 Objective and hypotheses	12
2. Methods	12
2.1 Study Area	12
2.2 Heat Source Modeling	15
2.3 Simple Shade model	18
2.3.1 Response variable	18
2.3.2 Topographic shade	20
2.3.3 Vegetative shade	22
2.3.4 Upstream watershed area	23
2.4 Summary statistics	24
2.5 Principal components analysis	24
2.5.1 Data transformation	25
2.5.2 Principal components	26
2.5.3 Variable importance	26
2.6 Random forest modeling	27
2.6.1 Model setup	28
2.6.2 Variable reduction	28
2.6.3 Variable interaction	29
2.6.4 Model alternatives	29
3. Results	30
3.1 Summary statistics	30
3.2 Principal components analysis	34
3.3 Random forest modeling	38
3.3.1 Model trials	38
3.3.2 Variable interaction	40
3.3.3 Model alternatives	42
4. Discussion	45
4.1 Model performance	45
4.2 Model limitations	48
4.3 Recommendations for future research	49
References	52

APPENDIX A: Effective shade field data for Heat Source calibration.....	58
APPENDIX B: Simple Shade model data	59
APPENDIX C: Script for computing zonal statistics on an NDVI raster.....	60
APPENDIX D: Script for principal components analysis	61
APPENDIX E: Script for Simple Shade random forest modeling	62
APPENDIX F: Script for partial dependence plots	63

List of Figures

Figure 1. Diagram of how (A) vegetation, (B) topography and (C) channel width influence effective shade.	6
Figure 2. Lidar data products (in blue) available from the Oregon Lidar Consortium (DOGAMI 2009).	10
Figure 3. The reflectance of blue, green, red and near-infrared light from different land cover types.	11
Figure 4. Study area map of the Middle Fork Coquille River.	13
Figure 5. Map of 2011 National Land Cover Database (NLCD) land cover classification and percent composition in the Middle Fork Coquille Watershed. Most "Shrub/Scrub" land cover is harvested timberland.	14
Figure 6. Distribution of Heat Source sample nodes along the Middle Fork Coquille River.....	16
Figure 7. Diagram of effective shade definition and calculation used by Heat Source.	17
Figure 8. Predicted (line) versus observed (points) effective shade for the Middle Fork Coquille River (Bloom 2017).	18
Figure 9. Heat Source estimates of effective shade for the Middle Fork Coquille River.	19
Figure 10. Example of land cover sampling performed by TTools.	21
Figure 11. Example of topographic shade angles (θ , α).	21
Figure 12. Polygon shapes used to extract zonal statistics from (A) true color imagery and (B) NDVI imagery at Simple Shade sample nodes.	23
Figure 13. Example diagram of the random forest machine algorithm.	27
Figure 14. Histogram of effective shade binned into five classes.	30
Figure 15. Correlation matrix of effective shade and all topographic shade angle variables. Scatter plots with LOESS line fits between variables are below the histograms and the correlation coefficient between variables is above the histograms. The larger the magnitude of a correlation coefficient is, the larger the font.	32
Figure 16. Correlation matrix of effective shade, watershed area and NDVI variables. Scatter plots demonstrating LOESS line fits between variables are below the histograms and the correlation coefficient between variables is above the histograms. The larger the magnitude of a correlation coefficient is, the larger the font.	33
Figure 17. Biplot of PCA results displaying (A) scatter plot of samples and (B) the eigenvectors of each variable in ordination space compared to effective shade correlation. Scatter plot labels in (A) indicate the sample node identification number, with 0 at the mouth and 628 at the headwaters.	36
Figure 18. Effective shade versus PC1 and PC2. River kilometer increases with color shift from red to yellow (Red= lower watershed and yellow= upper watershed).	37

Figure 19. Predictor variable importance for effective shade in the Middle Fork Coquille River, OR. The horizontal axis represents the percent increase in mean squared error (MSE) if the associated predictor was removed from the model.	39
Figure 20. Partial dependence plots demonstrating the relationship between effective shade and (A) watershed area, (B) maximum NDVI, (C) topographic shade angle to the South and (D) mean NDVI. All other explanatory variables are held at an average value while the variable of interest is varied from minimum to maximum.	41
Figure 21. Variable importance plot for an upper watershed random forest model of effective shade in the Middle Fork Coquille River.	42
Figure 22. Variable importance plot for a lower watershed random forest model of effective shade in the Middle Fork Coquille River.	43
Figure 23. Variable importance plot for a classification random forest model of effective shade for the Middle Fork Coquille River.	44

List of Tables

Table 1. Comparison of effective shade measurement methods (OWEB 1999). Approximate costs updated to reflect 2018 prices.	8
Table 2. List of variables in principal components analysis.	25
Table 3. Summary statistics for variables in principal components analysis.	31
Table 4. Summary of principal component (PC) importance. Bold values indicate where more variance was explained by the PC than the broken stick model.	34
Table 5. Individual predictor variable loadings on PC1 and PC2, listed from smallest to largest. Loading importance is based on absolute value with sign (positive or negative) indicating the association between variable and the principal component.	35
Table 6. Summary of random forest model trials.	38
Table 7. Confusion table comparing observed versus predicted effective shade class.	44

1. Introduction

1.1 Stream Temperature

Increasing temperatures in streams and rivers pose significant threats to aquatic life and human health worldwide. Elevated stream temperatures degrade water quality by stressing cold-water species, promoting harmful algal blooms, lowering dissolved oxygen levels and increasing the toxicity of some compounds (Carlisle et al. 2009). Aquatic species depend on temperature for cueing specific physiological functions and life cycle components (Richter & Kolmes 2005). Extreme changes of water temperature can have adverse impacts on fertility, growth, life cycle, or distribution, and could lead to a loss of biodiversity (Poole & Berman 2001).

The temperature of a river reflects a combination of heat inputs, storage and outputs. Heat inputs include solar radiation, condensation and precipitation, advection from ground water and upstream, and from tributary inflows (Walling & Webb 1992). Heat energy is lost from the water through reflection of solar, atmospheric and forest radiation; by back radiation from the water; by evaporation; and by outflow from the stream reach. Energy can be gained or lost by convection and by conduction to or from the atmosphere, streambed, and banks (Brown 1969). These parameters, and the volume of water, will affect the rate of temperature change in streams and rivers.

The largest source of heat for aquatic ecosystems is solar radiation (Brown 1969). Absorption of solar energy affects the thermal structure, stratification and hydrodynamics of lakes and rivers. In addition, the productivity of fresh waters is driven and controlled by energy derived directly from the solar energy utilized in photosynthesis. The amount of direct solar radiation a stream receives varies with channel shading, depth and shape. The temperature of large rivers is less influenced by shading, as their size ensures that they are largely exposed to the

sun. In small streams, however, shading can substantially influence temperature regimes. Gray and Edington (1969) closely monitored temperature in a small stream in Northern England before and after much of the catchment was deforested. The year after trees were harvested, the summer maximum temperature increased from 15 to 21.5 °C. Loss of streamside vegetation due to agriculture and other human activities has likely increased the summer temperatures of many streams around the world (Karr & Schlosser 1978).

1.1.1 Drivers of warming

Temperatures are rising globally due to anthropogenic and natural processes that increase heat transfer to streams (Diabat et al. 2013). Natural processes that contribute to stream temperature increase include geothermal activity and channel morphology. Wide, shallow channels have more surface area for heat exchange to take place than narrow, deep channels (Trimmel et al. 2016). For this reason, lakes and wetlands tend to be sources of heat to connecting streams and rivers. Additionally, geothermal activity can create hot springs that transfer heat from below the earth's surface to waterways.

Climate change causes the atmosphere to warm, which increases heat transfer at the water's surface. Impacts of climate change on freshwater biota have already been observed around the world (Schindler 1997). Durance and Ormerod (2007) published a decline in macroinvertebrate abundance at a small catchment in Wales with increasing stream temperatures over a 25-year period. Decline in freshwater biodiversity has also been linked to climate-induced changes in water temperature (Lake et al. 2000; Xenopoulos et al. 2005).

Another anthropogenic impact on water temperature is thermal pollution. Thermal pollution is any discharge into a water body that is significantly warmer than the natural water

source (Clark 1969). This includes point source discharges, commonly from municipal and industrial effluent. Water is used as a cooling agent by manufacturing plants, and if it is not cooled prior to being discharged, temperature will rise sharply in the receiving water body, which alters natural environments and destroys habitat (Clark 1969). Thermal pollution is also caused by runoff from parking lots or other impervious surfaces. Water that flows over these surfaces absorbs much of their heat and transfers it to nearby streams or rivers (Perlman 2013).

Degradation of riparian vegetation also contributes to the anthropogenic warming of streams and rivers. The removal of riparian vegetation increases the exposure of water bodies to direct sunlight, which increases water temperature. Streams that are shaded by vegetation or other objects will not absorb as much heat as sunlit water. Riparian deforestation is linked to multiple land use types such as timber harvest, agricultural activities, livestock grazing and urban development (Perlman 2013).

1.1.2 Impacts of warming in the Pacific Northwest

In the Pacific Northwest, many culturally and economically valuable aquatic species are impacted by warming stream temperatures. Native salmonids and trout are highly adapted to narrow thermal regimes, and experience weight loss, disease, competitive displacement, reduced fecundity and death when exposed to overwarm stream temperatures (Richter & Kolmes 2005). More than \$80 million dollars was invested in watershed enhancement projects aimed at increasing native salmon populations from 2015-2017 in Oregon alone (OWEB 2017). Restoring thermal regimes is a major component of salmonid conservation and water quality management, yet many streams in Oregon are still too warm to support healthy native fish populations (ODEQ 2012).

1.2 Water quality regulation

Excessively warm temperature is the most common water quality impairment of Oregon's waterways (ODEQ 2012). In Oregon, the Department of Environmental Quality (ODEQ) regulates water quality according to Clean Water Act framework. The ODEQ sets water quality standards according to the beneficial uses a water body supports. Beneficial uses include swimming, fishing, drinking water supply, and fish and aquatic life. Fish are typically the most sensitive beneficial use to water temperature increases, with native salmon and trout species being particularly vulnerable (ODEQ 2011).

The ODEQ monitors water quality in streams, rivers, lakes, and reservoirs across the state to identify places and times with violations of standards. When a standard is violated, the impaired stream is placed on the Environmental Protection Agencies' (EPA) 303(d) list and a pollution reduction plan called a Total Maximum Daily Load (TMDL) must be developed as required under the Federal Clean Water Act (CFR 130.7(c)). A TMDL provides the amount of pollutant a waterbody can receive over time and still meet water quality standards. For temperature violations, heat is the pollutant of concern. Heat can enter a stream by many different natural and man-made processes. A temperature TMDL defines the amount of thermal load that can be discharged into a water body without exceeding temperature standards, allocates heat loading from different sources and identifies actionable targets that designated management agencies must implement in order to restore instream temperature (ODEQ 2011).

1.3 Effective shade

Effective shade is the fraction of solar radiation flux blocked from reaching the stream surface over a specified time period. TMDLs in Oregon commonly identify effective shade as a primary target for temperature restoration. This includes temperature TMDLs for the Tualatin,

Willamette, and Nehalem Rivers (ODEQ 2001, ODEQ 2006, ODEQ 2003). Effective shade is a useful target due to the strong relationship between effective shade and stream temperature flux which is often the primary cause of stream temperature warming (Gebhardt & Fischer 1999), as well as the ease it is manipulated via management actions. Replanting streamside vegetation is a cost effective and measurable way to reduce in stream temperature by reducing exposure to direct solar radiation (OWEB 2017). While direct solar radiation is only part of the heat budget for aquatic systems, it is the most important for temperature regulation, particularly in mid-latitude regions during the summer months (Beschta 1997; Li et al. 2012).

The amount of effective shade on a stream is especially important during the hottest time of the year. In Oregon, this is usually July, August and September. Spawning salmon, eggs, juveniles and fry are especially susceptible to heat induced stress and mortality during these months (Richter & Kolmes 2005). To account for different salmonid life cycle phases and use patterns, temperature standards in Oregon vary over the course of a year. For example, during periods of rearing and migration, when juveniles are developing and swimming to the ocean, the seven-day-average maximum temperature may not exceed 18.0°C (ODEQ 2011). During designated spawning periods, when adult salmon lay and fertilize eggs, the seven-day-average maximum temperature may not exceed 13.0°C (ODEQ 2011).

Three primary mechanisms control effective shade: topographic shade, vegetative shade and channel width (Figure 1). Topographic shade is the shading of a stream by landscape features such as hills, mountains and stream banks. The amount of effective shade provided by a topographic feature depends on its height, distance to stream and orientation relative to seasonal solar paths. Vegetative shade is the shading of a stream by riparian vegetation. The amount of effective shade provided by vegetation varies with canopy height, canopy density and season.

Channel width influences the proportion of stream surface available to be shaded by topography or vegetation.

Figure 1. Diagram of how (A) vegetation, (B) topography and (C) channel width influence effective shade.

The removal of riparian vegetation, and resulting decrease in effective shade, is the primary non-point source of heat pollution to streams in Oregon (ODEQ 2012). To increase effective shading, over 500 miles of riparian planting projects were completed between 2015 and 2017 (OWEB 2017). However, 283 streams are still listed for temperature violations in Oregon (ODEQ 2012). Thus, information on the overall status of effective shading to waterbodies across the state, as well as methods to assess the change in effective shade over time in response to restoration activities, are critical for informing water quality management and riparian restoration activities.

1.3.1 Measuring effective shade in the field

Effective shade can be measured using a variety of devices and techniques ranging from hand-held spherical mirrors to specialized cameras costing thousands of dollars (Table 1). All of these methods generally measure the proportion of sky that is shaded by vegetation or

topography. Determining the best tool for use depends on the ease of use, cost, level of data precision desired, and the questions asked of the monitoring data. While these techniques can be applied effectively on individual systems of interest, widespread application across Oregon would be extremely labor intensive and cost prohibitive at the current time. Until the time and cost for widespread field measurements become more tractable, modeling and remote sensing approaches to quantifying will continue to be the most efficient way to estimate effective shade over large regions and over time.

Table 1. Comparison of effective shade measurement methods (OWEB 1999). Approximate costs updated to reflect 2018 prices.

Method	Description	Advantages	Disadvantages	Cost
Densiometer	Small spherical mirror with grid reflects sky. Grid intersections are counted to determine % canopy cover	Inexpensive, quick, easy, and indestructible. Small, light weight device. Procedure has been widely used.	Difficult to keep hand-held device level. Taking into consideration different vegetation qualities is difficult. Measures canopy cover, not shade directly.	~~\$100-~~ \$150
Clinometer	Measures angle from horizon to open sky. Gives percentage of 180-degree arc that is covered by vegetation or topography.	Inexpensive, quick, easy and fairly rugged. Small, light weight device. Procedure has been widely used.	Internal moisture can obscure reading and foul moving parts if dropped in stream. Taking into consideration different vegetation qualities is difficult. Requires good vision in two eyes. Measures angles to open sky, not shade directly. Tends to lump a site into high or low with no gradation.	~~\$100-~~ \$150
Hemispherical photography	180 degree photograph of the sky is computer analyzed	Produces high-quality, permanent canopy cover records. Less prone to user error than other methods. Computer analysis enables more complex data manipulation, analysis, and storage. Directly measures shade.	Expensive, heavy, and delicate. Not simple and easy to use. Different lighting conditions can cause problems. Requires more data reduction than other methods. This is a fairly new technology that has had limited use.	\$4000 to \$8000
Solar Pathfinder	180 degree diagram of sky is hand drawn. Open area on diagram gives amount of solar energy reaching the stream.	Fairly easy, and quick to use (not as quick and easy as densiometer). Inexpensive and light weight. Produces permanent canopy record. Measurements are not effected by lighting conditions. Directly measures shade.	Light weight plastic parts are not particularly rugged. Adjusting for different vegetation shade qualities is possible but not automatic. Prone to user error. Operating equipment in center of rapidly flowing stream can be challenging. Requires more data reduction than other methods.	~~\$200-~~ \$300

1.3.2 Modeling effective shade

In Oregon, the Heat Source model (Boyd and Kasper 2006) is typically used to predict effective shade in stream temperature modeling. This process-based model produces estimates of effective shade based on simulations of solar inputs and shading from topographic features and riparian vegetation. Topographic shading is characterized with digital elevation models and solar paths are created according to a user determined time step. Vegetation height surfaces can be derived from Lidar data or digitized by hand from aerial imagery. Lidar is an active remote sensing method that emits light and measures the time it takes to reflect off an object and return to the sensor. This time is converted to distance which can then be used to generate three-dimensional elevation models. This process is similar to how bats and dolphins use sound pulses to map their environments. Although Lidar provides the highest quality data, these data are expensive, infrequently collected and unavailable for large swaths of Oregon (Figure 2). While hand digitization for specific projects can be accomplished, using this approach over large areas and over time is not feasible from a labor or cost point of view. For these reasons, current models and approaches that require vegetation height to predict effective shade cannot assess riparian condition across the entire state on a recurring basis.

Figure 2. Lidar data products (in blue) available from the Oregon Lidar Consortium (DOGAMI 2009).

Shade models have been developed, such as Penumbra (Halama 2017) and the Soil and Water Assessment Tool (SWAT) (Gassman et al. 2007), that do not require high-resolution vegetation height inputs. However, the outputs of these models are too coarse to inform reach scale water quality restoration projects and meet the ODEQ’s water quality management goals. These models predict the solar radiation received by an entire watershed over daily, or longer, time steps rather than predict effective shade at discrete points relevant for restoration activities.

1.4 NDVI and effective shade

The Normalized Difference Vegetation Index (NDVI) is a measure of vegetation health that uses the proportion of near-infrared (NIR) and red light reflected from a surface. It is a unitless index of plant “greenness” or photosynthetic activity (Glenn et al., 2008). NDVI can be used

for land cover classification since different surfaces reflect visible and invisible light differently (Figure 3) (Gandhi et al., 2015). For example, the National Land Cover Database (NLCD) uses NDVI to derive vegetation phenology (Homer et al. 2015). NDVI has also been directly correlated to biomass, leaf chlorophyll content, vegetation water stress and plant height (Jensen 2000, Liang 2004, Payero et al. 2004, Sheffield et al. 2010). Based on the robust relationships between NDVI and vegetation characteristics that are directly related to effective shade (i.e. canopy cover, height) NDVI should correlate with effective shade. The formula for calculating NDVI is:

$$(1) \text{NDVI} = ((\text{NIR} - \text{Red})) / ((\text{NIR} + \text{Red})).$$

Where NIR is near-infrared light and Red is light in the red spectrum.

Figure 3. The reflectance of blue, green, red and near-infrared light from different land cover types.

1.5 Objective and hypotheses

The objective of this thesis is to develop an empirical model of effective shade that allows ODEQ to assess status and trends in effective shade, as a surrogate for water temperature, on two-year intervals for reports summarizing water quality statewide.

The approach and model I describe uses statistical analyses to relate remotely sensed landscape predictors to effective shade. Landscape predictors are derived from high resolution aerial imagery and digital elevation models. I hypothesize that NDVI can be used to predict effective shade and that predictions from an empirical model of effective shade will be comparable to those from a data-intensive mechanistic model.

2. Methods

2.1 Study Area

The Middle Fork Coquille watershed encompasses one USGS 10-digit hydrologic unit (1710030501) in the South Coast Basin of Oregon (Figure 4). The river is 64 km long and transects Douglas and Coos counties. The watershed is more than 60% forested (Figure 5) and land use is dominated by commercial timber production (Homer et al. 2015). The Camas Valley is an agricultural center and rural community located in the upper watershed. Oregon State Route 42 runs along the river from the Camas Valley to the mouth.

Figure 4. Study area map of the Middle Fork Coquille River.

Figure 5. Map of 2011 National Land Cover Database (NLCD) land cover classification and percent composition in the Middle Fork Coquille Watershed. Most "Shrub/Scrub" land cover is harvested timberland.

The ownership of the Middle Fork Coquille Watershed is divided between public lands administered by the Bureau of Land Management (BLM), forest products companies, tribal lands, Oregon Department of Transportation highway right of way, private agricultural activities, small timber owners and rural residences (Follansbee 2003). Major tributaries to the Middle Fork Coquille River that are also listed for temperature impairment include Belieu Creek, Rock Creek, Twelvemile Creek, Bingham Creek, and Reed Creek (ODEQ 2012).

2.2 Heat Source Modeling

The EPA and ODEQ have listed the Middle Fork Coquille River for violation of temperature water quality standards. The beneficial uses threatened by overwarm stream temperatures include salmon and steelhead spawning and rearing and cold water aquatic life (ODEQ 2012). This listing status requires a TMDL that identifies and quantifies sources of temperature pollution (heat) to the river. As part of TMDL development, ODEQ staff used the Heat Source model to make predictions of effective shade at 100-meter intervals for the entire Middle Fork Coquille River (Figure 6) (Boyd and Kasper 2003; Bloom 2017).

Figure 6. Distribution of Heat Source sample nodes along the Middle Fork Coquille River.

The Heat Source stream thermodynamics model was developed at Oregon State University in 1996 (Boyd and Kasper 2003). It has since been maintained and updated ODEQ. Heat Source is a data intensive mechanistic model that uses watershed, riparian, and water quality inputs to produce high-resolution estimates of stream temperature, effective shade and associated parameters. It has been used for water quality regulation across the United States (Watanabe et al. 2005) and in temperature studies around the world (Bond et al. 2015, Diabat et al. 2013, Holzapfel et al. 2013, Justice et al. 2017, Lawrence et al. 2014, Loheide and Gorelick 2006, Trimmel et al. 2016, and Woltemade and Hawkins 2016).

Heat Source contains a module that simulates solar radiation flux at the stream surface, which is used to calculate effective shade. Effective shade is simulated as a function of geographic positioning, near stream vegetation, topography and channel morphology (Figure 7) (Boyd and Kasper 2003).

Source: Heat Source Methodology (Boyd & Kasper 2003)

* Does not account for near stream vegetation; represents potential solar insolation without any interference from vegetation.

Figure 7. Diagram of effective shade definition and calculation used by Heat Source.

The Middle Fork Coquille River Heat Source model was calibrated to field measurements of effective shade (Figure 8; Appendix A). The mean of squared residuals was 2.4% of effective shade units. Elevation and vegetation height surfaces used in model calculations were derived from Lidar data collected in 2009 (Bloom 2017). Because of the widespread acceptance of Heat Source as a standard by which to accurately measure effective shade and the performance of the

calibrated model, I used the effective shade outputs of the Middle Fork Coquille Heat Source model to train the Simple Shade model.

Figure 8. Predicted (line) versus observed (points) effective shade for the Middle Fork Coquille River (Bloom 2017).

2.3 Simple Shade model

2.3.1 Response variable

In order to develop Simple Shade outputs at the same spatial resolution of Heat Source estimates of effective shade, the same sample nodes generated by that model were used in Simple Shade (Figure 9). The Heat Source effective shade outputs were used as the response variable. This enabled Simple Shade to test a much larger set of data than would have been possible with sparse field observations.

Figure 9. Heat Source estimates of effective shade for the Middle Fork Coquille River.

2.3.2 Topographic shade

Topographic shade angle is the angle between a sample node and a point on the horizon that contributes the most shade to the sample node. It is calculated using TTools (Michie 2018), which is a suite of Python (Version 2.7) scripts used in conjunction with ArcGIS (Version 10.5) (ESRI 2016) to sample and assemble geospatial data inputs for the Heat Source model. The maximum topographic shade angle was calculated to the East, Southeast, South, Southwest and West of each sample node (Figure 10). In each direction TTools incrementally samples a digital elevation model (DEM) for the maximum elevation, then computes the angle to that point (the “topographic shade angle”; Figure 11). The Lidar derived DEM for the Middle Fork Coquille was provided by the Oregon Spatial Data Library and has a resolution of 1 meter (DOGAMI 2016). The formula for calculating the topographic shade angle is:

$$(2) \theta_T = \tan^{-1}\left(\frac{z_T - z_i}{L_T}\right)$$

where z_i is elevation of the stream data node,

z_T is elevation of the topographic features,

L_T is distance from stream data node to topographic feature, and

θ_T is topographic shade angle.

Figure 10. Example of land cover sampling performed by TTools.

Figure 11. Example of topographic shade angles (θ , α).

2.3.3 *Vegetative shade*

Digital aerial imagery collected by the National Agricultural Imagery Program (NAIP) in 2009 was used to assess near stream vegetation. This imagery has a 1-meter spatial resolution and a 4-band spectral resolution (red, green, blue, and near-infrared). It is collected during the agricultural growing season, or “leaf-on” period (June, July, August, September), and is used as a base layer by the USDA Farm Service Agency and other public agencies and private organizations (USDA 2017). The multispectral nature of this imagery enables the calculation of a Normalized Difference Vegetation Index (NDVI) value for every pixel in the study area.

NDVI values were summarized for each sample node using polygons that extended 50 meters clockwise from the Northeast to the Northwest of each node (Figure 12). This distance is consistent with the largest riparian buffer widths used in regional land management practices, including forestry (DeWalle 2008). The zonal statistics calculated include minimum, maximum, mean and standard deviation of NDVI. This data analysis was completed using the raster package in R (Version 2.6-7) (Hijmans 2017).

Figure 12. Polygon shapes used to extract zonal statistics from (A) true color imagery and (B) NDVI imagery at Simple Shade sample nodes.

2.3.4 Upstream watershed area

Channel width influences the capacity for a stream to be shaded. A direct measurement of channel width requires left and right bank digitization - a resource intensive process that requires high-resolution, hydro-conditioned DEMs and hand digitization. Because a primary goal of the Simple Shade model is to provide a cost effective alternative to resource intensive models, I used upstream watershed area as a surrogate for channel width because it is highly correlated with channel width and readily derived from widely available data products (Lyon 2003). Watershed area was calculated using the ESRI Watershed Tool in ArcGIS. Inputs were point locations (sample nodes) and a DEM and outputs were watershed polygons

2.4 *Summary statistics*

Summary statistics were used to describe parameter values within the Simple Shade dataset and a correlation matrix was used for exploratory analysis. Histograms were used to assess the frequency distribution of select variables. Histograms were inspected for skewness, outliers and artifacts that may influence modeling results.

2.5 *Principal components analysis*

I used a principal components analysis (PCA) to examine correlations between predictor variables and to consolidate predictor variables into similar groups. Reducing the dimensionality of data with high covariance allows me to efficiently summarize the data, identify inter-site relationships, and detect underlying patterns. Data were log-transformed and scaled in order to minimize outlier and skewness effects that could dominate the PCA. All PCA and related analysis completed with R statistical computing software (Version 3.4.3) (R Core Team 2017).

Table 2. List of variables in principal components analysis.

Parameter	Description
NDVI Maximum	Maximum NDVI value within a sample zone.
NDVI Mean	Mean NDVI value within a sample zone.
NDVI Minimum	Minimum NDVI value within a sample zone.
NDVI Standard Deviation	Standard deviation of NDVI values within a sample zone.
Topographic Shade Angle to the East	Angle in degrees to a point on the horizon to the East contributing the most shading to a sample node.
Topographic Shade Angle to the Northeast	Angle in degrees to a point on the horizon to the Northeast contributing the most shading to a sample node.
Topographic Shade Angle to the Northwest	Angle in degrees to a point on the horizon to the Northwest contributing the most shading to a sample node.
Topographic Shade Angle to the South	Angle in degrees to a point on the horizon to the South contributing the most shading to a sample node.
Topographic Shade Angle to the Southeast	Angle in degrees to a point on the horizon to the Southeast contributing the most shading to a sample node.
Topographic Shade Angle to the Southwest	Angle in degrees to a point on the horizon to the Southwest contributing the most shading to a sample node.
Topographic Shade Angle to the West	Angle in degrees to a point on the horizon to the West contributing the most shading to a sample node.
Watershed Area	Upstream drainage area of a sample node in square kilometers.

2.5.1 Data transformation

The “powerTransform” function in the car package (Version 3.0-0) (Fox and Weisberg 2011) indicated the appropriate transformation for improving the normality of predictor variable data. This function uses the maximum likelihood-like approach of Box and Cox (1964) to select a transformation for a multivariate dataset. Due to negative values in the dataset, the Yeo-Johnson power family of transformations was used. The function returns the most likely transformation for improving normality as a lambda value. Lambda ranges from -5 to 5 and indicates an exponent to which all data for a specific parameter should be raised. The exception

is when lambda equals zero. In this situation, the recommended transformation is a logarithm with base e .

2.5.2 *Principal components*

The PCA determined the interaction of predictor variables and produced synthetic principal components that combined multiple predictor variables. The number of principal components is equal to the number of original variables. The proportion of variance explained by principal components summarizes the importance of that principal component. Because the importance of each PC decreases sequentially from PC1 to PC k (1 to k variables) and all PCs are independent of each other, dimensional reduction of the multivariate data can be achieved by retaining only the top few PCs without losing too much information.

A broken stick model determined which principal components were meaningful. This technique retains components that explain more variance than would be expected by randomly dividing the variance into as many parts as there are variables (Jackson 1993). Only principal components that met these criteria were analyzed. This analysis was completed using the MASS package in R (Version 7.3-47) (Venables and Ripley 2002).

2.5.3 *Variable importance*

The eigenvector of each variable defines the linear relationship between that variable and the corresponding principal component (Venables and Ripley 2002). This vector is equivalent to a coefficient in a regression. Eigenvectors highlight patterns between variables and sample sites within a dataset. They are also helpful for understanding predictor variable importance. Effective shade was plotted as an environmental variable in ordination space alongside predictor variable eigenvectors using the envfit function within the vegan package (Version 2.5-2) (Oksanen et al.

2018). This demonstrated the relationship between effective shade, PC1 and PC2, and identified variables that were correlated with effective shade.

2.6 Random forest modeling

Random forest is a machine learning ensemble modeling technique that can be used for both regression and classification problems. “Forest” refers to the creation and averaging of multiple decision trees (usually >500 trees) to get a more accurate and stable prediction than a single decision tree (Breiman 2001). Decision trees are statistical analysis tools that model complex variable interactions relatively simply. However, individual trees may not predict well and trees with a large number of splits can be difficult to interpret (De’ath 2007). “Random” refers to (1) the repeated bootstrapping of the original sample with replacement and (2) the random subset of variables used at each decision tree branch (Figure 13). Learning algorithms, such as random forest, greatly improve upon the capacity to predict complex systems and render simple graphical and numerical interpretations of complex relationships (De’ath 2007).

Figure 13. Example diagram of the random forest machine algorithm.

Each tree in the random forest model is constructed using a bootstrap sample of the original data. Approximately one-third of observations are left out of bootstrap sampling and are used for out-of-bag (OOB) cross-validated error estimation (Breiman 2001). The proportion of times that the prediction is not equal to the true value class is the OOB error estimate. The OOB error is used to calculate importance scores for each variable (Breiman 2001). Estimates of what variables are important in the classification or regression model are then generated. If the number of variables is very large, random forest can be run once with all variables, then run again using only the most important variables identified in the first run. Useful features of random forest include the capacity to process large datasets with many input variables, the lack of assumptions required by other statistical inference methods and the ability to detect variable interactions (Breiman 2001).

2.6.1 Model setup

The Simple Shade random forest model inputs consisted of one response variable (effective shade) and 12 predictor variables ($n = 628$). Each observation represented one sample node from the Middle Fork Coquille Heat Source Model. I used the default of 500 decision trees to create the model as determined by the `randomForest` function. The number of random variables tested at each split was determined using the “`tuneRF`” function. The `randomForest` package (Version 4.6-12) (Liaw and Wiener 2002) was used for this analysis.

2.6.2 Variable reduction

In order to simplify the model without losing a significant amount of information, I created multiple random forest models with a reduced number of predictors. Variables were

dropped based on the results of the full model that tested all 12 predictors. The reduced model included the top five most important variables from the full model.

2.6.3 Variable interaction

I used partial dependence plots to better understand how effective shade changed over gradients of important variables identified in the random forest models. This technique varies the predictor of interest from minimum to maximum while all other variables are held at an average value. The “partialPlot” function in the randomForest package (Version 4.6-12) (Liaw and Wiener 2002) was used for this analysis.

2.6.4 Model alternatives

Additional random forest models were generated based on model results and partial dependence plots in order to improve explanatory power. First, the dataset was subset according to watershed location and two different models were used to predict effective shade in the upper and lower watershed separately. Next, a random forest classification algorithm was used to predict classes of effective shade in order to determine whether this approach was more appropriate than a regression for the dataset. Effective shade was binned into five equal categories (Figure 14). The variable importance plot for classification models compared mean decrease accuracy to identify significant variables. The larger the decrease in accuracy, the more important a variable is considered.

Figure 14. Histogram of effective shade binned into five classes.

3. Results

3.1 Summary statistics

The model variables displayed a variety of ranges and distribution types (Table 3). Effective shade ranged from 0.09 to 0.96 and has a median of 0.43. The distribution of effective shade is skewed slightly to the right and most observations were less than 0.60 (Figure 15).

Table 3. Summary statistics for variables in principal components analysis.

Parameter	Mean	Median	Minimum	Maximum
Effective Shade	0.45	0.43	0.09	0.96
NDVI Maximum	0.32	0.31	-0.03	0.96
NDVI Mean	0.16	0.33	-0.86	0.90
NDVI Minimum	-0.27	-0.16	-1.0	0.28
NDVI Standard Deviation	0.29	0.25	0.06	0.61
Topographic Shade Angle to the East	12.28	10.93	2.14	35.22
Topographic Shade Angle to the Northeast	13.72	12.85	1.36	40.03
Topographic Shade Angle to the Northwest	13.51	11.42	1.04	40.03
Topographic Shade Angle to the South	15.67	15.89	1.86	46.67
Topographic Shade Angle to the Southeast	13.49	12.15	2.19	40.74
Topographic Shade Angle to the Southwest	14.19	13.22	1.04	43.25
Topographic Shade Angle to the West	12.51	10.30	2.47	46.67
Watershed Area	368.04	303.82	10.54	798.60

Figure 15. Correlation matrix of effective shade and all topographic shade angle variables. Scatter plots with LOESS line fits between variables are below the histograms and the correlation coefficient between variables is above the histograms.

Effective shade showed very little correlation with the topographic shade angle variables and the topographic shade angles were all positively correlated with each other (Figure 15). All topographic shade angles have a right skewed distribution with a high frequency of low angle observations. The median of topographic shade angles ranged from 10 to 16 degrees (Table 3).

Effective shade was most correlated with Mean NDVI ($r = 0.33$) of all NDVI variables (Figure 16). Mean NDVI displays a roughly normal distribution with very few observations near the minimum and maximum range of the data. The scatter plots of NDVI minimum and NDVI mean show distinct groupings of observations.

Figure 16. Correlation matrix of effective shade, watershed area and NDVI variables. Scatter plots demonstrating LOESS line fits between variables are below the histograms and the correlation coefficient between variables is above the histograms.

Effective shade was most correlated with watershed area out of all of the predictor variables. Watershed area has a non-normal distribution with a high frequency of observations near the minimum and maximum values (Figure 16). The median of watershed area for the reach

was 303.82 km² and there was a gap in values between 500 and 600 km². This was attributed to a rapid increase in watershed area due to tributary inputs.

3.2 Principal components analysis

Two principal components were able to explain over half of the variability within the dataset (Table 4). Only principal component 1 (PC1) and principal component 2 (PC2) were retained for analysis. PC1 and PC2 explain 36.6% and 18.9% of total variance in the predictor variables, respectively. The PCA was performed on log-transformed variables.

Table 4. Summary of principal component (PC) importance. Bold values indicate where more variance was explained by the PC than the broken stick model.

	Proportion of Variance	Broken Stick Proportion of Variance
PC1	0.366	0.232
PC2	0.189	0.160
PC3	0.099	0.125
PC4	0.094	0.101
PC5	0.068	0.083
PC6	0.047	0.069
PC7	0.044	0.057
PC8	0.032	0.047
PC9	0.021	0.038
PC10	0.010	0.030
PC11	0.009	0.023
PC12	0.007	0.016
PC13	0.004	0.010
PC14	0.003	0.005

NDVI and watershed characteristic variables tended to have inverse loadings along each principal component, but more strikingly along the PC1 axis (Table 5). The PC2 axis was more associated with topographic shade variables. Two distinct clusters of observations are visible in ordination space (Figure 18A). Clustering aligns with sample node label; smaller numbers are grouped in the upper left corner of the plot and larger numbers are grouped in the lower right corner of the plot. This indicates that grouping is based on watershed location. Effective shade was more strongly correlated with PC2 than PC1, most similar in loading and direction to the NDVI mean eigenvector, and most dissimilar to the watershed area eigenvector (Figure 18B).

Table 5. Individual predictor variable loadings on PC1 and PC2, listed from smallest to largest. Loading importance is based on absolute value with sign (positive or negative) indicating the association between variable and the principal component.

Variable	PC1 Loading
Watershed Area	-0.36
Topographic Shade angle to the Southwest	-0.32
NDVI Maximum	-0.31
Topographic Shade angle to the South	-0.30
NDVI Standard Deviation	-0.28
Topographic Shade angle to the Northeast	-0.25
Topographic Shade angle to the East	-0.23
Topographic Shade angle to the West	-0.22
Topographic Shade angle to the Southeast	-0.17
Topographic Shade angle to the Northwest	-0.10
NDVI Mean	0.31
NDVI Minimum	0.38
Variable	PC2 Loading
Topographic Shade angle to the Southeast	-0.40
Topographic Shade angle to the Northwest	-0.39
NDVI Mean	-0.35
Topographic Shade angle to the East	-0.25
NDVI Minimum	-0.24
Topographic Shade angle to the West	-0.22
Topographic Shade angle to the South	-0.22
Topographic Shade angle to the Northeast	-0.21
Topographic Shade angle to the Southwest	-0.14
NDVI Standard Deviation	-0.11
NDVI Maximum	0.04
Watershed Area	0.25

Figure 17. Biplot of PCA results displaying (A) scatter plot of samples and (B) the eigenvectors of each variable in ordination space compared to effective shade correlation. Scatter plot labels in (A) indicate the sample node identification number, with 0 at the mouth and 628 at the headwaters.

Very little pattern exists between effective shade and the principal components, other than a slightly negative correlation between effective shade and PC 2 (Figure 19). There is a relationship between sample site location and principal component, demonstrated by point color. Samples from the upper watershed have stronger PC1 loadings and samples from the lower watershed have stronger PC2 loadings.

Figure 18. Effective shade versus PC1 and PC2. River kilometer increases with color shift from red to yellow (Red= lower watershed and yellow= upper watershed).

3.3 Random forest modeling

3.3.1 Model trials

The full random forest model testing all 12 predictor variables explained 45.0% of variation in effective shade (Table 6). The mean of squared residuals, which compares the difference between predicted and observed effective shade values, was 0.027 fractional effective shade units. Watershed area was the most influential predictor of effective shade, followed by topographic angle to the South, mean NDVI, maximum NDVI and topographic angle to the Southeast respectively (Figure 20).

Table 6. Summary of random forest model trials.

	<i>Full Model</i>	<i>Reduced Model</i>
<i>Random Forest Type</i>	Regression	Regression
<i>Number of Trees</i>	500	500
<i>No. of variables tried at each split</i>	6	2
<i>Mean of squared residuals</i>	0.027	0.029
<i>% Variance explained</i>	45.0	41.4

Figure 19. Predictor variable importance for effective shade in the Middle Fork Coquille River, OR. The horizontal axis represents the percent increase in mean squared error (MSE) if the associated predictor was removed from the model.

The top five predictors would each result in more than a 20% increase in mean squared error if dropped from the model. To examine these variables more in-depth, I created a random forest model with only the top five most influential predictor variables. This model explained 41.4% of variance in effective shade and identified watershed area, topographic shade angle to the South, mean NDVI, maximum NDVI and topographic shade angle to the Southeast as the most important predictors, respectively.

3.3.2 *Variable interaction*

To explore the relationships between individual predictor variables and effective shade along the Middle Fork Coquille River, partial dependence plots were created for watershed area, mean NDVI, maximum NDVI and topographic shade angle to the South (Figure 21). These variables were consistently identified as the top four most important variables in every random forest model trial. A partial dependence plot displaying effective shade response to topographic shade angle to the Southwest is not included because it matched the relationship between effective shade and topographic shade angle to the South.

Effective shade had generally positive correlations with mean NDVI and topographic shade angle to the South. Both plots show thresholds between effective shade and the gradient of mean NDVI and topographic shade angle before increasing to a plateau. The partial dependence plot of effective shade and mean NDVI shows three plateaus, with effective shade forming a stair step relationship over the gradient of mean NDVI. Effective shade and maximum NDVI had a negative relationship, with effective shade decreasing as maximum NDVI increased. Effective shade exhibited a strong threshold over the gradient of watershed, with distinct grouping above and below 500 km².

Figure 20. Partial dependence plots demonstrating the relationship between effective shade and (A) watershed area, (B) maximum NDVI, (C) topographic shade angle to the South and (D) mean NDVI. All other explanatory variables are held at an average value while the variable of interest is varied from minimum to maximum.

3.3.3 Model alternatives

To refine the predictions of effective shade, I developed separate models informed by the results of the partial dependence plots. First, I separated the dataset based on upper and lower watersheds because effective shade fell into two groups based on watershed area. Using only data from the upper watershed, the model explained 24.1% of variance in effective shade with a mean of squared residuals of 0.034 effective shade units; three variables were tried at each split. The most influential predictor variables were topographic shade angle to the South, topographic shade angle to the Southwest, topographic shade angle to the Southeast and watershed area respectively (Figure 22).

Figure 21. Variable importance plot for an upper watershed random forest model of effective shade in the Middle Fork Coquille River.

The lower watershed model explained 36.7% of variance in effective shade with a mean of squared residuals equal to 0.013; six variables were tried at each split. The most influential predictor variables were mean NDVI, watershed area, and minimum NDVI respectively (Figure 23).

Figure 22. Variable importance plot for a lower watershed random forest model of effective shade in the Middle Fork Coquille River.

I also used a classification random forest model to predict effective shade classes. Effective shade was binned into five equal classes. Most effective shade values fell between 0.20 and 0.80. The out of bag error for this model, which measures prediction error, was relatively high at 61% and three variables were tried at each split. The class representing the highest effective shade values had the highest misclassification rate as well as the lowest number of

observations (Table 7). The most importance variables in this model were watershed area, minimum NDVI and mean NDVI respectively (Figure 24).

Table 7. Confusion table comparing observed versus predicted effective shade class.

<i>n</i> =628		<i>Predicted Class</i>					Classification error
<i>Actual Class</i>		1	2	3	4	5	
	1	56	34	12	3	0	0.47
	2	30	66	54	26	3	0.63
	3	9	50	49	45	7	0.69
	4	6	14	39	64	9	0.51
	5	2	5	9	27	9	0.83

Figure 23. Variable importance plot for a classification random forest model of effective shade for the Middle Fork Coquille River.

4. Discussion

The objective of this thesis was to develop an empirical model that makes use of readily available remotely sensed imagery to assess effective shade with an accuracy comparable to outputs from a data-intensive mechanistic model. I hypothesized that NDVI would serve as a strong predictor of effective shade, thereby allowing assessments of effective shade over time that are not currently possible with Lidar data or hand delineation. Results of my analysis and model construction suggest that, both qualitatively and quantitatively, NDVI can predict different levels of effective shade after accounting for other watershed characteristics that influence effective shade. In this discussion, I focus on several components of the Simple Shade model that can be improved to prepare this tool for use in assessing riparian conditions in Oregon and elsewhere.

4.1 *Model performance*

The mean squared error of the Simple Shade model was 2.7%, and the mean squared error of the Middle Fork Coquille River Heat Source model of effective shade was 2.4% (Bloom 2017). These values suggest that the precision of Simple Shade is similar to the mechanistic Heat Source model. When compared to other models used to assess water quality, Simple Shade performed as well, or better with respect to precision of model estimates. Duda et al. (2012) suggested that a less than 7% difference between simulated and observed values is very good performance for the Hydrological Simulation Program Fortran (HSPF) watershed model. The ODEQ uses this model to simulate nutrients, sediment, temperature and bacteria in streams. This is a costly model to develop, and the EPA estimates that it would require five full time employees a minimum of one year to develop an HSPF model for a watershed larger than 250 square miles (EPA 1996). The Willamette Basin is approximately 11,000 square miles. The

performance of the Automated Geospatial Watershed Assessment (AGWA) hydrologic model is considered acceptable if simulated values are within 30% of observed values (Al-Quarashi et al. 2008). Simple Shade predicted effective shade as well as the Heat Source model in the Middle Fork Coquille Watershed, and satisfied criteria used to determine acceptable precision for other water quality models.

Mechanistic models require a high degree of environmental data that are often unavailable or prohibitively expensive to collect. Empirical models, like Simple Shade, simplify complex systems by using observed relationships between important and easily measured parameters. This makes them an attractive solution to modeling complex systems with limited data and resources. Ample literature has compared mechanistic and empirical models. In general, it is uncommon that an empirical model performs equally well to a mechanistic model. For example, Caffi et al. (2007) concluded that empirical models of grapevine downy mildew infection were too over-simplified to simulate correctly complex biological phenomenon influencing infection. The mechanistic approach was preferred as it provided accurate dynamic simulations and required neither calibration nor correction. On the other hand, Estes et al. (2013) concluded that empirical models could achieve the same or better accuracy as mechanistic models for predicting suitability and productivity of crops in South Africa. A successful empirical model must use surrogates for mechanisms that accurately mimic the effect of the mechanism on the system and align with how the system has behaved in the past (Gustafson 2013).

The Simple Shade model captured mechanistic processes influencing shading by using relevant predictor variables. The mean NDVI of the riparian area to the South (as determined by the sampling polygons) of each sample node differentiated between land cover types that

contribute different amounts of shade to the stream. For example, where effective shade was highest, mean NDVI was highest and land cover was generally mature trees. Where effective shade was lowest, mean NDVI was lowest and land cover was generally bare earth or grass. Thus, NDVI qualitatively captured the critical relationship between riparian vegetation height and density on stream shading (Beschta 1997).

Maximum NDVI was also included in the model as a possible surrogate for riparian vegetation; but exhibited an unexpected negative correlation with effective shade. This is likely because the maximum value was less representative of riparian vegetation characteristics for a given area than an average value. Maximum NDVI captures the characteristics of a single tree while mean NDVI captures overall tree height and canopy density for a given area.

Topographic shade is generally less influential on stream heat flux than vegetative shade. However, this physical landscape feature may confound the predictive relationship between riparian vegetation and effective shade. For example, if effective shade is high and mean NDVI is low (indicating little vegetation cover), there may be a large hill adjacent of the stream. Predictions of effective shade inconsistent with riparian vegetation condition can therefore be explained by including directional topographic shade angles in the Simple Shade model. The topographic shade angle incorporates both distance to stream and feature height, which are the metrics used by process based models to calculate the amount of incoming solar radiation blocked by topography.

Including watershed area in Simple Shade was intended to act as a surrogate for channel width, which influences the amount of stream surface available to be shaded. However, basin-specific drainage characteristics resulted in highly nonlinear longitudinal variation in channel width as opposed to the expected continuous channel widening with basin area observed across a

large number of river basins (Leopold et al. 1964). This was due to watershed specific characteristics (the confluence of multiple tributaries) that caused a rapid increase in upstream watershed area over a relatively short distance. This suggests that longitudinal variation in stream characteristics complicated the relationship between effective shade and channel width and affected model performance.

4.2 Model limitations

A major limitation of this research was the presence of shadows in the imagery prior to NDVI calculation. This is because shadows change how a surface reflects and absorbs light, which is what is used to calculate NDVI. When using NDVI to determine land cover type, the surface must be shadow free to avoid misclassification. For example, if a forested area falls under shadow, the NDVI for that area will be lower than expected, indicating a vegetation free land cover type (like road or water). Conversely, the NDVI of open water is artificially increased when shadows are present, indicating a vegetated land cover type. These errors affect the Simple Shade model by confounding the assumed relationship between mean riparian NDVI and effective shade. A sample node may have high levels of effective shade, but due to the shading of the area in an image, riparian NDVI is low when it should be high. It is contradictory that the presence of shadows in an image decreases mean NDVI, which indicates less effective shade to the model. This limitation likely weakened the relationship between mean NDVI and effective shade and influenced overall model performance.

An additional limitation is the comparison of aerial imagery collected on different days at different times. Daily variation in atmospheric condition and sun angle influences how light is reflected, which affects the calculation of NDVI (Sheffield et al. 2010). For example, clouds reflect visible light and attenuate NIR bands, which results in low NDVI. Therefore, imagery

collected on cloudier days is likely to produce lower NDVI values. This issue could be exacerbated in larger watersheds that require more time to image. The larger the variation in dates and times of imagery, the larger the variation in NDVI throughout the watershed. NDVI also varies with physiological plant function, so if a plant is stressed NDVI is reduced. This is a well-known observation used in precision agriculture to identify failing or diseased crops (USDA 2017). In the context of this research, it means that the health of plants could influence the relationship between NDVI and effective shade though canopy cover and height may not be visibly affected.

Lastly, since Simple Shade was trained using modeled values of effective shade, there is potential for compounding error. This would result in the model being less accurate than the reported mean squared error. For instance, if Heat Source had an error of 3% and Simple Shade had an error of 3%, the total absolute error between observed and predicted effective shade would be 6%. In the future, field measurements of effective shade should be used to calibrate and validate the Simple Shade model.

4.3 Recommendations for future research

This study found that longitudinal variation in the Middle Fork Coquille River influenced model performance disproportionately more than the physical mechanisms that provide shade. To move towards a more general model, watershed characteristics that vary dramatically with river kilometer need to be normalized. This could be achieved by using separate models to predict shade according to stream order, for example a model for the headwaters (stream order 1 to 3), one for the mid reaches (stream order 4 to 6), and one for the lower reaches (stream order greater than 6). This would prevent steep gradients of longitudinal stream characteristics from dominating model parameter relationships.

To address the major limitations of this research, shadows should be masked or removed from the imagery prior to analysis. This would ensure that NDVI values are representative of land cover type rather than temporary shading. This would also enable the accurate classification of land cover type based on NDVI values. Trends could then be examined between effective shade and general vegetation type (i.e. bare earth, shrubs, trees), which would be informative to water quality managers. They could use NDVI to identify areas of land cover types associated with low effective shade that could then be prioritized for replanting projects. Shadow removal could be completed manually, though this could be time consuming depending on the watershed and imagery. An automated technique could be used to identify shadows in an image, similar to the spectral-unmixing algorithm developed by Adler-Golden et al. (2002). They used shadow detection, atmospheric correcting, and reflectance filtering to identify the shadow fraction for each pixel, and then applied a pixel-by-pixel filter that produced visually realistic de-shadowed images.

Levins (1966) argued that there is a necessary trade-off between generality, precision and realism in mathematical models of complex biological systems. In order to meet the research objective of developing a model to monitor changes in shade across Oregon, the generality and precision of Simple Shade must be robust. While Simple Shade was relatively precise, its generality is currently limited to the Middle Fork Coquille River and its realism depends on the known parameters included in the model that affect effective shade. Oregon has a diverse landscape with nine distinct ecoregions, and Simple Shade needs to be tested and calibrated in river systems from different Oregon ecoregions to improve its generality. The precision of the Simple Shade model should be evaluated by comparing predictions to field measurements of effective shade and riparian condition over time. In particular, it is critical to continue to assess

the major hypothesis of this research: mean NDVI increases in response to vegetation growth, both in height and canopy cover.

This research is the first step to developing an empirical model of effective shade. NDVI was identified as an important predictor and a potential surrogate measure of effective shade. However, variability in imagery and longitudinal watershed characteristics complicated the relationship between NDVI and effective shade. A technique for normalizing image and longitudinal variability within a watershed is needed to improve the generality and precision of Simple Shade so that it can be used for assessing riparian conditions as part of water quality management in Oregon and elsewhere.

References

- 40 CFR § 130.7(c), Total maximum daily loads (TMDL) and individual water quality-based effluent limitations, Code of Federal Regulations.
- Adler-Golden, Steven & W. Matthew, Michael & P. Anderson, Gail & W. Felde, Gerald & A. Gardner, James, 2002, An Algorithm for De-Shadowing Spectral Imagery, Proceedings of SPIE - The International Society for Optical Engineering, 4816. 10.1117/12.451691.
- Allen, D., Dietrich, W. et al., 2007, Development of a mechanistically based basin-scale stream temperature model: Applications to cumulative effects modeling, USDA Forest Service Gen.Tech. Rep. PSW-GTR-194.
- Allen, J. D. 1995. Stream ecology: Structure and function of running waters. Chapman & Hall, New York, 388 pp
- Beschta, R., 1997, Riparian shade and stream temperature: An alternative perspective, Society for Range Management, 19(2):25-28.
- Bloom, James R., 2017, "DRAFT : Middle Fork Coquille River, Oregon Heat Source Temperature Model Simulations", State of Oregon Department of Environmental Quality, Watershed Management Section.
- Bond, R.M, Stubblefields, A.P., Van Kirk, R.W. 2015, Sensitivity of summer stream temperatures to climate variability and riparian reforestation strategies, *Journal of Hydrology: Regional Studies*, 4(B): 267-279.
- Boyd, M. and Kasper, B., 2003, "Analytical methods for dynamic open channel heat and mass transfer: Methodology for heat source model Version 7.0".
- Breiman, L., 2001, Machine Learning (2001), Accessed March 2018 at: <https://doi.org/10.1023/A:1010933404324>.
- Brown, G.W., 1969, Predicting temperatures of small streams, *Water Resources Research*, 5:68-75.
- Caffi, T. , Rossi, V. , Cossu, A. and Fronteddu, F. (2007), Empirical vs. mechanistic models for primary infections of *Plasmopara viticola*. EPPO Bulletin, 37: 261-271. doi:10.1111/j.1365-2338.2007.01120.x.
- Clark, J.R., 1969, Thermal Pollution and Aquatic Life, *Scientific American*, v. 220, p. 18–27, doi: 10.2307/24926305.
- De'ath, G., 2007, BOOSTED TREES FOR ECOLOGICAL MODELING AND PREDICTION, *Ecology*, 88: 243-251, doi:10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.
- DeWalle, D.R., 2008, Guidelines for Riparian Vegetative Shade Restoration Based Upon a Theoretical Shaded-Stream Model, *JAWRA Journal of the American Water Resources Association*, v. 44, p. 1373–1387, doi: 10.1111/j.1752-1688.2008.00230.x.

- Diabat, M., Haggerty, R., Wondzell, S.M., 2013, Diurnal timing of warmer air under climate-change affects magnitude, timing and duration of stream temperature change, *Hydrological. Processes*, 27(16): 2367–2378.
- Dobrowski, S., Safford, H., Cheng, Y., & Ustin, S., 2008, Mapping Mountain Vegetation Using Species Distribution Modeling, Image-Based Texture Analysis, and Object-Based Classification, *Applied Vegetation Science*, 11(4), 499-508. Retrieved from <http://www.jstor.org/stable/25488487>
- Durance, I., and S. J. Ormerod, 2007, Climate change effects on upland stream invertebrates over a 25 year period, *Global Change Biol.*, 13, 942–957.
- Environmental Systems Research Institute (ESRI), 2016, ArcGIS Release 10.5, Redlands, CA.
- EPA, 1996, TMDL development Cost Estimates: Case Studies of 14 TMDLs, Washington, DC, EPA-R-96-01.
- Follansbee, B., 2003, Middle Fork Coquille River Riparian Shade Assessment Technical Data for Water Quality Management Plan Prepared for Coquille Watershed Association, Accessed June 2017 at: <https://www.oregon.gov/deq/FilterDocs/scCRMiddleforkassessment.pdf>.
- Gassman, P.W., M. R. Reyes, C. H. Green, J. G. Arnold. 2007. The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. *Transactions of the ASABE*. Vol. 50(4): 1211-1250.
- Gebhardt, K. and Fischer, H., 1999, Using surrogates for stream temperature nonpoint source water quality loading allocation. *Proceedings of the Seventh Biennial Watershed Management Conference: Water Resources Center Report No. 98*, University of California Davis.
- Gray, J.R.A. and J.M. Edington, 1969, Effect of woodland clearance on stream temperature, *Journal of the Fisheries Research Board of Canada*, 26:399-403, <https://doi.org/10.1139/f69-038>.
- Gustafson, Eric J., 2013, When relationships estimated in the past cannot be used to predict the future: using mechanistic models to predict landscape ecological dynamics in a changing world, *Landscape Ecology*, 28:1429-1437, doi:10.1007/s10980-013-9927-4.
- Halama, Jonathan J., 2017, Penumbra: A Spatiotemporate Shade-Irradiance Analysis Tool with External Model Integration for Landscape Assessment, Habitat Enhancement, and Water Quality Improvement, Doctor of Philosophy dissertation presented on January 6, 2017.
- Hijmans, Robert J., 2017, “raster: Geographic Data Analysis and Modeling”, R package version 2.6-7, <https://CRAN.R-project.org/package=raster>.
- Holzappel, G., Weihs, P., and Rauch, H.P., 2013, Use of the Shade-a-lator 6.2 model to assess the shading potential of riparian purple willow (*Salix purpurea*) coppices on small to medium sized rivers: *Ecological Engineering*, v. 61, p. 697–705, doi: 10.1016/J.ECOLENG.2013.07.036.

- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information, *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354.
- Jensen, J. R., 2000, Remote Sensing of the Environment: An Earth Resource Perspective, 544, Prentice Hall, Upper Saddle River, New Jersey, USA.
- Justice, C., White, S.M., McCullough, D.A. Graves, D.S., Blanchard, M.R, 2017, Can Stream and riparian restoration offset climate change impacts to salmon populations?, *Journal of Environmental Management*, 188: 212-227.
- Karr, J., and Schlosser, I., 1978, Water Resources and the Land-Water Interface, *Science*, 201:229-34.
- Lake, P. S., and Coauthors, 2000, Global change and the biodiversity of freshwater ecosystems: Impacts on linkages between above-sediment and sediment biota, *Bioscience*, 50, 1099–1107.
- Lawrence, D.J., Stewart-Koster, B., Olden, J.D., Ruesch, A.S., Torgersen, C.E., Lawler, J.J., Butcher, D.P., Crown, J.K., 2014, The interactive effects of climate change, riparian management, and a nonnative predator on stream-rearing salmon, *Ecological Applications*, 24(4): 895-912.
- Leopold, L. B., Gordon Wolman, M., and Miller, J. P., 1964, *Fluvial processes in geomorphology*, W. H. Freeman, San Francisco.
- Li, G., Jackson, C. et al., 2012, Modeled riparian stream shading: Agreement with field measurements and sensitivity to riparian conditions, *Journal of Hydrology*, 428-429, 142-151.
- Liang, S., 2004, Quantitative Remote Sensing of Land Surfaces, John Wiley & Sons, Inc., New Jersey, USA.
- Liaw, A. and M. Wiener, 2002, “Classification and Regression by randomForest”, *R News* 2(3), 18-22.
- Loheide S.P., Gorelick, S.M., 2006, Quantifying stream–aquifer interactions through the analysis of remotely sensed thermographic profiles and in situ temperature histories, *Environmental Science and Technology*, 40(10): 3336-3341.
- Lyon, J., 2003, Deriving Stream Channel Morphology Using GIS-Based Watershed Analysis, *GIS for Water Resource and Watershed Management*, London: CRC Press.
- Michie, R., 2018, TTools, GitHub repository, <https://github.com/rmichie/TTools>.
- Millennium Ecosystem Assessment, 2005, Ecosystems and human well-being: Biodiversity synthesis. World Resources Institute Millennium Ecosystem Assessment Rep., 86 pp.
- Oksanen, J.F., Guillaume, B., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G., Solymos, P., Henry, M., Stevens, H., Szoecs, E., and Wagner,

- H., 2018, vegan: Community Ecology Package. R package version 2.5-2.
<https://CRAN.R-project.org/package=vegan>.
- Oregon Department of Environmental Quality (ODEQ), 2001, “Tualatin Subbasin Total Maximum Daily Load”, ODEQ, TMDL Program,
<https://www.oregon.gov/deq/FilterDocs/tmdlwqmp.pdf>.
- Oregon Department of Environmental Quality (ODEQ), 2003, “North Coast Subbasins Total Maximum Daily Load”, ODEQ, TMDL Program,
<https://www.oregon.gov/deq/FilterDocs/NCStmdl.pdf>.
- Oregon Department of Environmental Quality (ODEQ), 2006, “Willamette Basin Temperature Total Maximum Daily Load”, ODEQ, TMDL Program,
<https://www.oregon.gov/deq/FilterDocs/chpt4temp.pdf>.
- Oregon Department of Environmental Quality (ODEQ), 2011, Temperature Standard Rule, OAR-340-041- 0028, Accessed May 2018 at:
http://arcweb.sos.state.or.us/pages/rules/oars_300/oar_340/340_041.html.
- Oregon Department of Environmental Quality (ODEQ), 2012, Oregon’s 2012 Integrated Report Assessment Database and 303(d) List, Accessed March 2018 at:
<http://www.oregon.gov/deq/wq/Pages/2012-Integrated-Report.aspx>.
- Oregon Department of Environmental Quality (ODEQ), 2018, Total Maximum Daily Loads, Accessed March 2018 at <https://www.oregon.gov/deq/wq/tmdls/Pages/default.aspx>.
- Oregon Department of Geology and Mineral Industries (DOGAMI), 2009, Lidar Data Viewer, Accessed March 2018 at: <https://gis.dogami.oregon.gov/maps/lidarviewer/>.
- Oregon Department of Geology and Mineral Industries (DOGAMI), 2016, Oregon-Lidar-Derived Digital Elevation Model Mosaic, Accessed March 2017 at:
<http://spatialdata.oregonexplorer.info/geoportals/details?id=0516198835684a759a422bc0d3cef9d8>.
- Oregon Watershed Enhancement Board (OWEB), 1999, Water Quality Monitoring Technical Guide Book Ch. 14 Stream Shade and Canopy Cover Monitoring Methods. 775 Summer ST. NE., Suite 360, Salem, OR, 97301-1290.
- Oregon Watershed Enhancement Board (OWEB), 2017, 2015-2017 Oregon Plan biennial report, Oregon Government Publications Collection, Accessed July 2018 at
<https://digital.osl.state.or.us/islandora/object/osl:62075>.
- Payero, J.O., Neale, C.M.U., and Wright, J.L., 2004, Comparison of eleven vegetation indices for estimating plant height of alfalfa and grass: Applied Engineering in Agriculture, v. 20, p. 385–393, doi: 10.1017/CBO9781107415324.004.
- Perlman, H., 2013, Water Properties: Temperature, In The USGS Water Science School, Accessed July 2018 at: <http://ga.water.usgs.gov/edu/temperature.html>.

- Poole, G.C., and Berman, C.H., 2001, PROFILE An Ecological Perspective on In-Stream Temperature: Natural Heat Dynamics and Mechanisms of Human-Caused Thermal Degradation: *Environmental Management*, v. 27, p. 787–802, doi: 10.1007/s002670010188.
- Richter, A. and Kolmes, S., 2005, Maximum temperature limits for Chinook, Coho, and Chum salmon, and Steelhead Trout in the Pacific Northwest, *Reviews in Fisheries Science*, 13, 23-49.
- Santhi, C., R. Srinivasan, J.G. Arnold, J.R. Williams, 2005, A modeling approach to evaluate the impacts of water quality management plans implemented in a watershed in Texas, *Environmental Modelling & Software*, 21 (2006): 1141-1157.
- Schindler, D. W., 1997, Widespread effects of climate warming on freshwater ecosystems in North America, *Hydrol. Processes*, 11, 1043–1067.
- Sheffield, Kathryn & Abuzar, Mohammad & Whitfield, Des & Mcallister, Andy & O'Connell, Mark & Mcclymont, Lexie, 2010, SATELLITE REMOTE SENSING BASED MEASURES OF RIPARIAN VEGETATION STATUS. Conference Paper, 15th Australian Remote Sensing and Photogrammetric Conference.
- SolarPathfinder™, 2016, Instruction Manual For The SolarPathfinder Unit™, Accessed July 2018 at: <https://www.solarpathfinder.com/pdf/pathfinder-manual.pdf>.
- Trimmel, H. Gangneux, C. Kalny, G. Weihs, P. 2016. Application of the model ‘Heat Source’ to assess the influence of meteorological components on stream temperature and simulation accuracy under heat wave conditions. *Meteorologische Zeitschrift*. 25(4): 389-406.
- United States Department of Agriculture (USDA), Farm Service Agency, 2017, NAIP Imagery, Accessed December 2017 at: <https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/index>.
- United States Department of Agriculture (USDA), Farm Service Agency, 2017, NAIP Imagery, Accessed December 2017 at: <https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/index>.
- United States Geological Survey (USGS), 2018, NDVI, the foundation of remote sensing phenology, Accessed June 2018 at: https://phenology.cr.usgs.gov/ndvi_foundation.php.
- Walling, D.E. and B.W. Webb, 1992, Water Quality, I. Physical Characteristics, in P.Calow and G.E. Petts, eds. *The Rivers Handbook*, I. Hydological and ecological principles, Blackwell Scientific Publishing, Oxford, pp.48-72.
- Watanabe, M., R.A. Adams, J. Wu, J.P. Bolte, M.M. Cox, S.L. Johnson, W.J. Liss, W.G. Boggess and J.L. Ebersole, 2005, Toward efficient riparian restoration: integrating economic, physical and biological models, *Journal of Environmental Management*, 75:93- 104.
- Wetzel, G.R., 2001, *Limnology: Lake and River Ecosystems*, 3rd. 10.1016/B978-0-08-057439-4.50017-4.

- Woltemeade, C.J., and Hawkins, T.W., 2016, Stream temperature impacts because of changes in air temperature, land cover, and stream discharge: Navarro River watershed, California, USA, *River Research Applications*, 32(10): 2020-2031.
- Xenopoulos, M. A., D. M. Loddge, J. Alcamo, M. Marker, K. Schulze, and D. P. van Vuuren, 2005, Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biol.*, 11, 1557–1564.

APPENDIX A: Effective shade field data for Heat Source calibration.

Transect	Stream KM (from TTools)	Solar Pathfinder Effective Shade	Heat Source Effective Shade	Bankfull Width (m)
MC 27	51	83.0%	51.0%	13.1
U/S Sandy Creek Bridge #1	25.35	17.0%	23.0%	13.4
U/S Sandy Creek Bridge #2	25.35	47.0%	43.0%	18.3
MM 35	25.1	66.2%	78.0%	NA
ML 1	14.05	61.9%	48.0%	3.0
RM 3.2 #1	5.15	24.0%	35.0%	37.8
RM 3.2 #2	5.15	30.0%	43.0%	29.6

Mean of squared residuals: 0.024

APPENDIX B: Simple Shade model data

APPENDIX C: Script for computing zonal statistics on an NDVI raster

APPENDIX D: Script for principal components analysis

APPENDIX E: Script for Simple Shade random forest modeling

APPENDIX F: Script for partial dependence plots