Methods - Multiple Regression Stream Temperature Model

The intent of this project is to outline the basic steps needed to build multiple regression statistical models useful for assessing spatial or temporal patterns in stream temperature throughout river networks. The approach requires basic GIS skills and stream temperature data that are now routinely collected by National Forests and other land management agencies. An example application of the models for delineating suitable habitat for bull trout is provided in an accompanying bull trout monitoring report (Isaak et al.).

A. Stream temperature database. One of the first steps in building a stream temperature model is assembling and organizing a stream temperature database. The necessary data can either come from existing sources or be collected during subsequent field efforts.

Existing temperature data.—The advent of cheap, reliable digital thermographs has made collection of stream temperature data relatively routine. If these data exist within the area of interest, they can often be easily integrated to the development of a stream temperature model. The main requirement is that the data are referenced to a specific time and place on an individual section of stream. If the data were recorded with digital thermographs, the time-series of temperature measurements may also be available. A variety of temperature metrics can be summarized from these time-series using the SAS temperature macro. Most metrics are highly correlated and designed to describe temperature means, maximums, or variability (Dunham et al. 2005). In our temperature summaries, we focus on a standard summer period from July 15 - September 15, which usually encompasses the warmest portion of the year when spatial variation in temperatures is most pronounced and fish growth greatest. If the original time-series of temperature measurements are not available, the investigator must use whichever summary statistics were originally calculated or will have to discard the data.

New temperature data.—Although enough data may exist in some areas to develop preliminary stream temperature models, these data were probably collected for a variety of studies and may not provide a representative sample of the study area. To develop this sample, temperatures observations should be obtained across the range of major environmental gradients that are likely to affect stream temperatures. Within mountain landscapes, elevation and stream size (length or width) are often two of the dominant gradients that also have the advantage of being easily described in a GIS. Elevation creates a strong gradient because cooler air temperatures and greater snow accumulations at higher elevations cool stream temperatures through cold ground water inputs and convective heat exchange. Stream size creates a gradient because longer streams have been exposed to solar insolation over a greater length and wider streams are less shaded by adjacent riparian vegetation. Other factors may have important local effects on stream temperatures (e.g., riparian management activities, fires, large lakes), and if these are of interest to managers, will require additional sampling.

Example sampling design.—In 2006, RMRS scientists assisted biologists on the Payette National Forest to develop a network-scale stream temperature model for the Secesh River (Isaak et al.). To determine an efficient sampling strategy for this project, a digital representation of the stream network was developed and used to summarize gradients in elevation and stream size (contributing area is often used as a surrogate for stream size; Figure 1). The scatterplot summary was divided into nine quadrants and several stream segments within each quadrant were selected for thermograph deployments.

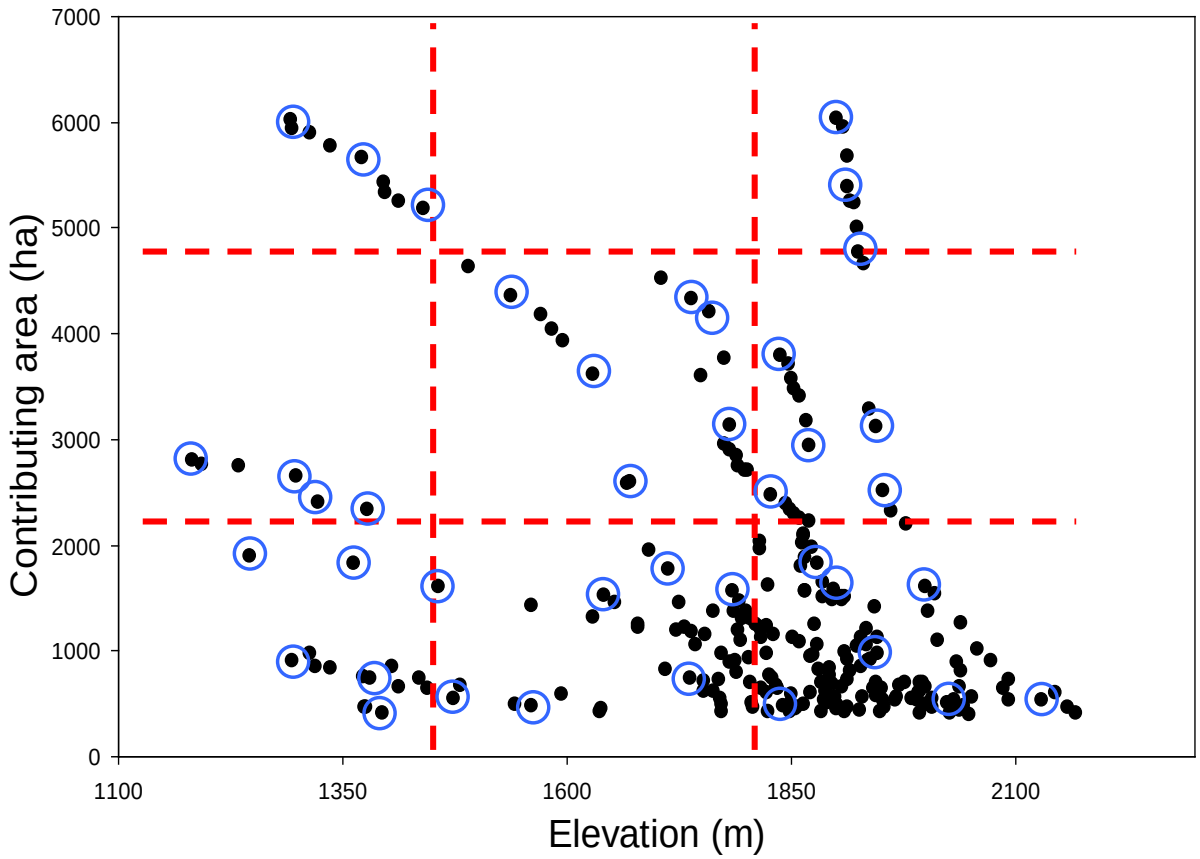

Figure 1. Scatterplot of stream segment contributing area values versus gradients from a digital representation of the river network in the Secesh watershed. Dashed lines indicate quadrants used to apportion the temperature observations; blue circles indicate stream segments where thermographs were deployed.

If stream temperature data had already existed in several of these quadrants, fewer thermographs could have been deployed in these areas. In the final sampling design, notice that thermographs were often deployed near stream mouths to facilitate logistical efficiency (Figure 2). Thermographs were also deployed immediately upstream/downstream from strong local influences like the large lake on Loon Creek and near recent fire perimeters to isolate the effects of these factors on stream temperatures. On several longer streams that were of particular interest (e.g., Lake and Lick Creeks), multiple thermographs were deployed, although an exact interval was not set. At each site, thermographs were placed in deep areas of flowing water after being mounted inside opaque cylinders that provided shade as described in Dunham et al. (2005)

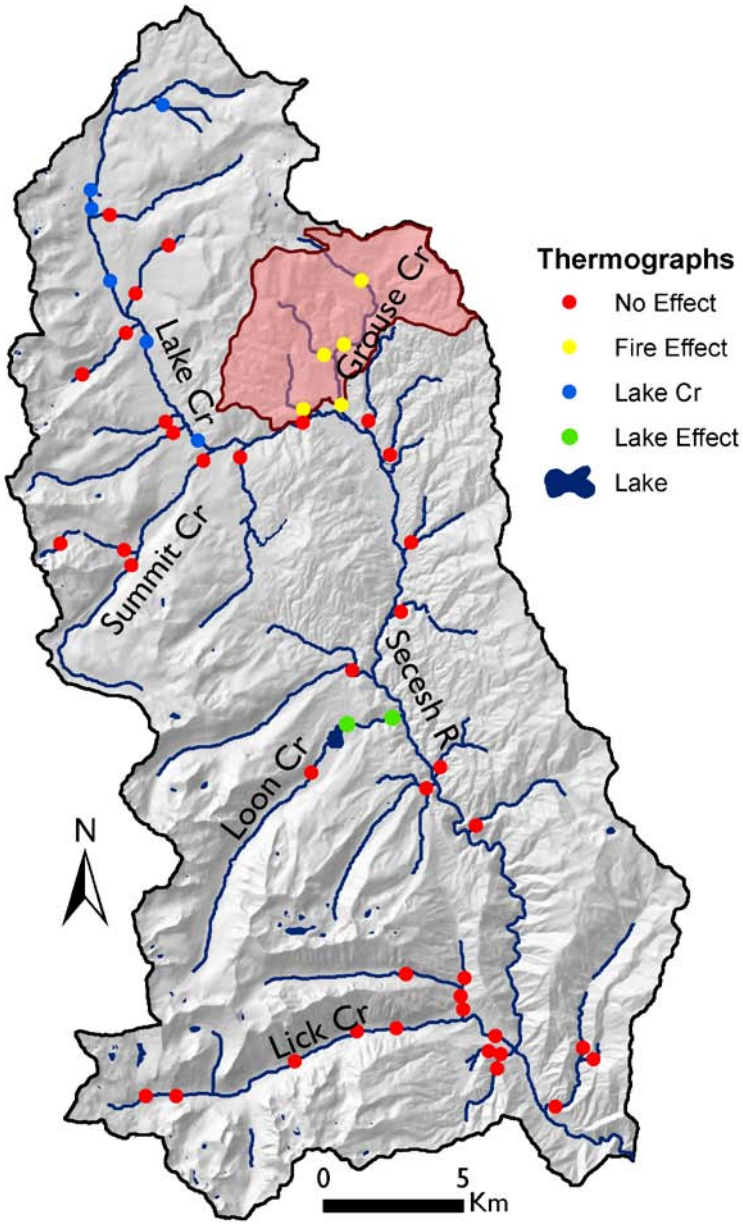

Figure 2. Thermograph locations in the Secesh River basin in 2006. Points are colored based on local thermal influences. The shaded area shows the extent of a fire that burned Flat, Sand, and Grouse creeks in 2000.

After retrieving thermographs, the investigator should create a point file in ArcMap that represents the locations of each thermograph site (How to Create a Shapefile of Thermograph Sites).

When developing a statistical model, approximately 10 observations are needed to estimate the parameter associated with each variable to ensure the accuracy and stability of these estimates. In our example, therefore, a minimum of approximately 30 stream temperature observations would be needed for a rudimentary temperature model that included parameters for the elevation, contributing area, and intercept parameters. If additional predictor variables are desired, additional observations should be obtained. If resources are limited, obtaining the necessary observations in a given year may be difficult, but sampling could also be spread across multiple years. This approach reduces the number of thermographs that are needed and can also capture inter-annual differences in air temperatures and stream flows that cause variation in stream temperatures.

B. Predictor variables. Predictor variables are used to describe spatial or temporal attributes associated with the stream network that may affect stream temperature. Many of these variables can be downloaded from the web or calculated from a digital elevation model (DEM) using a geographic information system (GIS). Once these variables are calculated, they can be used to develop temperature sampling designs like the one described above or linked to stream temperature observations to develop predictive statistical models.

Stream network delineation.—We have used two different sources for obtaining stream networks. One program is TauDEM (Terrain Analysis Using Digital Elevation Models), which is available at 1:24,000-scale. The TauDEM program was developed by Dave Tarboton at Utah State University (http://hydrology.neng.usu.edu/taudem/) and is a free toolbar plugin that uses DEMs. For details on downloading DEMs for specific areas and creating a digital stream network see How to Create Stream Networks Using DEM and TauDEM.

A second source is NHDPlus (http://www.horizon-systems.com/nhdplus/index.php), which uses the medium resolution, 1:100,000-scale, National Hydrography Dataset (NHD). For instructions on downloading a stream network from NHDPlus see How to Download NHDPlus Streams.

Geomorphic predictors.—Once developed, the digital stream network and DEM can be used to quantify a variety of different geomorphic predictors that may affect stream temperatures. For purposes of this example, we will use contributing area, channel slope, and elevation. Area and slope are calculated by TauDEM (TauDEM Attributes) while elevation can be derived from the DEM and attributed to stream segments using ArcToolbox (How to Attribute Elevation to a Stream Segment).

The NHDPlus program provides contributing area, slope, and elevation in its extensive list of attributes (NHDPlus Attributes) and these values are easily joined to stream segments (How to Join Elevation, Contributing Area, and Slope to NHDPlus Stream Segments).

Climate predictors.—Climatically influenced factors like air temperature and stream flow may create considerable inter-annual variation and have important consequences for stream temperatures. Air temperature affects stream temperature through convective heat exchange near the surface of the stream and by influencing temperatures of near-surface groundwater, which is an important component of summer flows. Stream flow determines the volume of water available for heating, with larger flows having greater flow velocities and thermal capacities, which makes them less responsive to heating. It will often be impractical to measure climatic variables at

individual thermograph sites, but these data can be derived from weather or flow gaging stations in or near the study area. Time-series of daily or hourly values can be downloaded from USGS and regional climate center websites that summarize this information (http://water.usgs.gov/; http://www.wrcc.dri.edu/index.html).. Summary metrics for use as predictors can be obtained using the same SAS macro that was applied earlier to stream temperature data.

Categorical predictors.—In some instances, discrete landscape features (e.g., a large lake, riparian alterations associated with a wildfire boundary, hot-springs) can have pronounced effects on local stream temperatures. The potential effects of such features are difficult to measure with a GIS, but should not be ignored if possible. One solution is to treat these features as categorical predictors by coding the effect as present (1) or absent (0) within different portions of the stream network.

C. Developing the statistical models. Once the set of predictor variables has been derived for all segments of the stream network, the ArcMap point file that represents thermograph locations must be overlaid on the network and these points attributed with the values of the predictor variables in the respective stream segment (How to Attribute Points with TauDEM or NHDPlus). Once this is done, these predictor values are linked to the stream temperatures summarized from each thermograph to complete the database needed for the regression model. The Excel file "MultipleRegressionExampleData.xls" shows this database from the Secesh River project for the thermograph locations in Figure 2. Open the file and see that the predictor variables for our model include three continuous, geomorphic variables (*elevation*, *contributing area*, and *slope*) and three categorical variables. The categorical variables include a *Lake Effect* for thermographs downstream from a large lake on Loon Creek (n=2), a *Fire Effect* for thermographs inside a recent fire boundary (n=5), and a *Lake Creek Effect* for thermographs on Lake Creek (n=6), which exhibited consistently higher stream temperatures than nearby streams, perhaps due to differences in underlying geology.

The response variable for the model is *Summer Maximum Weekly Temperature* (Sum_MWMT). The Regression tool in Excel is used to estimate the predictor variable parameters and output the regression statistics (see: How to Model Stream Temperature Using ArcMap, step 1; for TauDEM, or NHDPlus). In our TauDEM example*, the model accounts for 60% of the variation in Sum_MWMT and all model parameters, except elevation, are statistically significant ($p < 0.05$). Before selecting a final temperature model, the investigator should do routine checks to ensure that assumptions for normality, homogeneity, and independence are met and may also want to employ model selection procedures to determine the most parsimonious set of predictors. Details regarding these procedures are beyond the scope of this project, but are summarized in many introductory statistical texts.

D. Application of stream temperature model. After a final stream temperature model has been developed, individual parameters can be examined to assess the relative importance of predictor variables, stream temperatures can be predicted throughout the network (Figure 3), or temporal trends examined if the model contains climatic predictors. Using these predictions, suitable habitat distributions may be mapped for individual species or individual life stages by selecting appropriate thermal criteria (see: How to Model Stream Temperature Using ArcMap, steps 2-3; for TauDEM, or NHDPlus). An application for bull trout is provided in Isaak et al.

*Because the NHDPlus data are at a lower resolution than TauDEM, NHDPlus results differ from the TauDEM results.

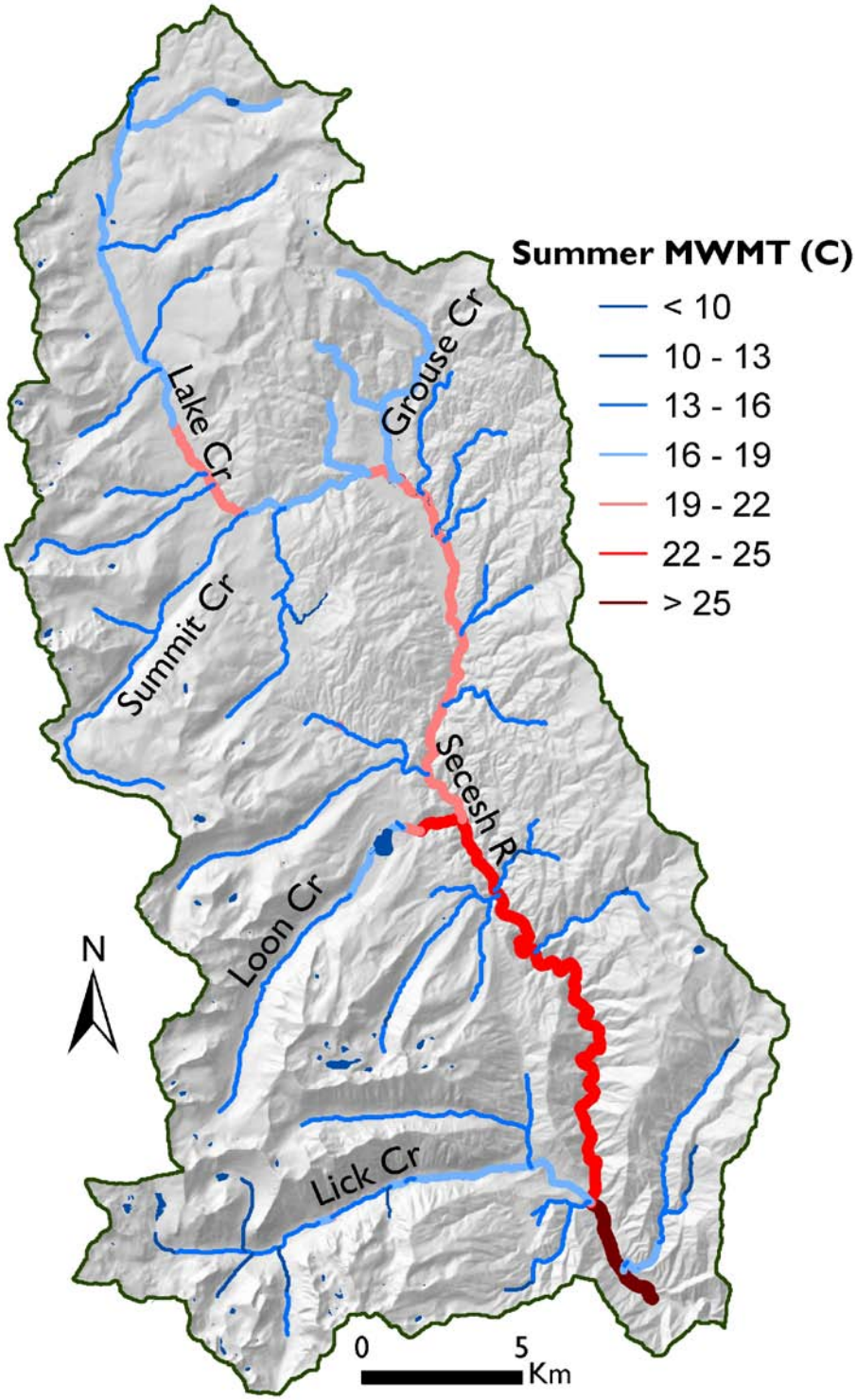

Figure 3. Predictions from the multiple regression model for maximum summer stream temperature mapped back to stream segments throughout the Secesh River network.