Bridger Trail Predictions - Follow-Up

Dr. Meridith L. Bartley

Extrapolation of out-of-sample trail use

This follow up project applies a model fit previously to trail use counter data in the Bridger Mountain trail system to additional out-of-sample trails in the same mountains. Deliverable product is a table of predicted trail use estimates (i.e. extrapolations), standard errors, and (95%) confidence intervals for the following six trail segments listed below in an Excel spreadsheet.

a. Bridger Foothills: Sypes to Middle Cottonwood

b. Bridger Foothills: Middle Cottonwood to Truman

c. Shafthouse Hill: Upper Shafthouse

d. Sypes Canyon: Four way to Baldy ("Upper Sypes")

e. Bridger Foothills: Truman Gulch to Ross Pass

f. Bridger Foothills: Ross Pass to Sacagawea Pass

The hierarchical generalized additive mixture model (HGAMM) applied to these out-of-sample trail (or trail subsections) was fit using 23 trail subsections for days with camera counter data (primarily summer 2021). Several model classes were compared in previous work, but model GI (single common smoother plus group-level smoothers with differing wiggliness) was found to be better at predicting in-sample fits for all trail subsections and thus is the model class applied to obtain out-of-sample predictions. This analysis was conducted in '`HE_Trailuse-R\output\report-bookdown\05-Analyze_AllTrails_HGAM.Rmd`' and full model details may be found in the statistical analysis report previously provided.

Some minor data cleaning was required for this work and included the following:

- The Sypes Canyon trail (included in the original model fit) was split into "Lower Sypes" and "Upper Sypes" sections. This split was done using Strava edge IDs assigned to each subsection provided by Headwaters Economics. Due to the method of combining Strava-provided trail use counts (i.e. combining segments using the maximum count among all segments) then the "Lower Sypes" segment should be close if not identical to that used

during model fit as this subsection is where the trail head and trail use camera were located.

- Previously provided trail subsections of the Bridger Foothills trail, Middle Cottonwood to Bostwick and Bostwick to Truman, were combined to create a unified "Middle Cottonwood to Truman" subsection.
- Previously provided trail subsections of the Bridger Foothills trail, Truman to Jones and Jones to Ross Pass, were combined to create a unified "Truman to Ross Pass" subsection.

An updated `strava_day` data frame with the new Sypes Canyon subsections was saved to `data/processed/strava_day_updated.Rdata.` With this file loaded (instead of the original raw strava_day.xlsx file) I re-ran `01-LoadData.R` and saved a new version of `allData` at `data/processed/allData_updated.Rdata`. This version of allData was filtered to only include the six trail subsections outlined above and the relevant (model) covariates and then used to obtain (extrapolate) new trail use estimates.

All associated files from this project will be directly provided, but for future work they have also been included in the same `HE-TrailUse-R` R Project folder in a new top-level folder named `follow-up`. New data files provided were added to `data/raw/follow-up`. Extrapolated estimates, standard errors, and confidence intervals for the six trail subsections along with the associated covariates used are located at `output/predictions` in a file called `predict_ModelGI_outofsample.csv`.

Caveat Reminder for Extrapolated Values

A common issue with any model, the generalized additive modelling framework included, is how to extrapolate beyond the range of data used to train the model. Temporal extrapolation is particularly tricky with temporal (time series) data. For GAMs, the issue arises as this framework uses splines to learn from the data via the basis functions. The splines are often set up directly related to the data included in the training set and it's not always clear how these should extend past that range of data. Improved forecasting could be accomplished with several years of year-round observations.