STAT 536 HW 1 Report

Ty Hawkes and Brigg Trendler

2024-09-30

Abstract

1. Introduction: Problem Statement and Understanding

2. Methodology

2.1 Proposed Methods

2.2 Model Evaluation

3. Results

3.1 What are the biggest climate / river network / human factors that impact overall river flow?

The table below shows the beta coefficients of the factors that improved the BIC of our model. Each variable in our model had a significant impact on overall river flow.

Term	Category	Estimate	Std. Error	Pr(>
(Intercept)		103.18286	22.61833	1.61e-05 ***
bio15	Climate	-0.59865	0.13978	4.65e-05 ***
CumPrec05	Climate	2.20575	0.63710	0.000826 ***
mPDC_SomewhatExcessive	Land Cover	0.42028	0.16888	0.014682 *
bio18	Climate	0.73360	0.37259	0.052074 .
cls2	Land Cover	317.43807	65.55078	5.37e-06 ***
Lon		-0.08278	0.02067	0.000128 ***
cls9	Land Cover	-1.32608	0.34508	0.000228 ***
bio14	Climate	-1.25629	0.39629	0.002090 **
gord	Network	0.44521	0.11739	0.000271 ***
mPDC_ModeratelyWell	Land Cover	-0.19627	0.09066	0.033074 *
cls8	Land Cover	2.17050	1.09753	0.051066 .
cls1	Land Cover	0.17930	0.09525	0.063035 .

Each category had at least one significant factor affecting river flow. Climate factors included bio15 (Precipitation Seasonality (Coefficient of Variation)), CumPrec05 (Cumulative May precipitation for the watershed upstream of grdc station), bio18 (Precipitation of Warmest Quarter), and bio14 (Precipitation of Driest Month); Land cover factors included meanPercentDC_SomewhatExcessive (mean somewhat excessive drainage class), cls2 (Evergreen_Broadleaf), cls9 (Urban), meanPercentDC_ModeratelyWell (mean moderately well drained soil), cls8 (Regularly Flooded Vegetation), and cls1 (Evergreen_Dec_Needle_Trees); and network factors included gord (global stream order from stream dem (Predicted relationship with area)).

3.2 How well do the factors you listed in #1 explain overall flow?

Our model does a decent job explaining the overall flow of rivers. With these selected co-variates, our model explains nearly 82% of the in-sample variance as measured by the r-squared value.

3.3 How predictive of overall flow are these identified factors?

We estimate that our model will do a reasonable job predicting overall flow on our-of-sample data. Using a leave-one-out cross validation, our RMSE was .4077, a significant improvement over the base intercept model with an RMSE of .8845.

4. Conclusions

Teamwork