Modeling Wave Celerity in Desert Rivers Using a Data Driven Approach

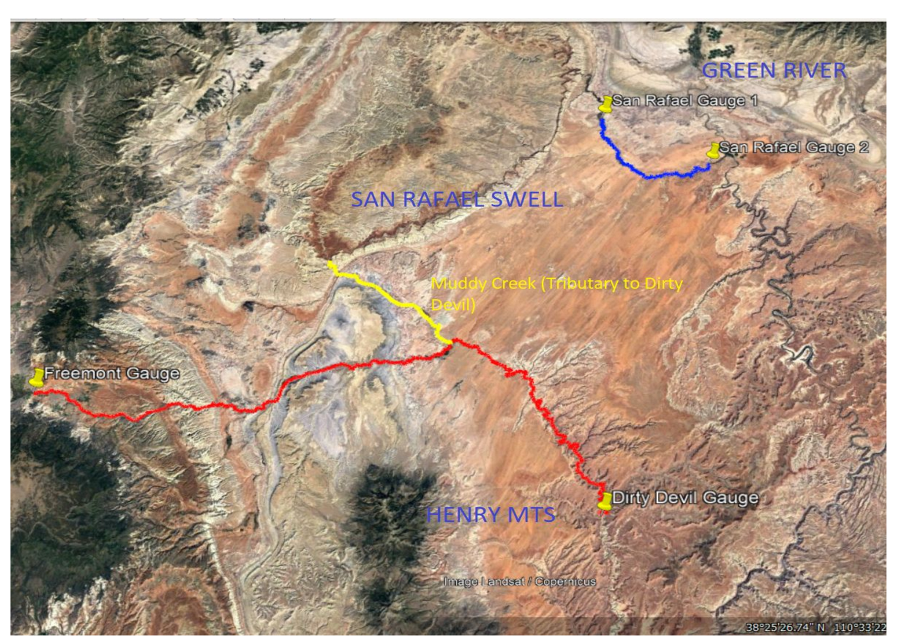
Torin Mcdonald, Jake Osterloh, Stone Mele aka Flow Riders

Problem

- Celerity refers to the speed at which a spike in water discharge travels down a river bed.
- If the celerity for a river is known, then we can estimate the time for a water surge to travel a certain distance.
- Calculating the celerity for river is traditionally done by methods that are either inaccurate or expensive.
- Our motivation for this project is to explore a data driven approach to calculating celerity by using river discharge data and various regression techniques we learned in class.

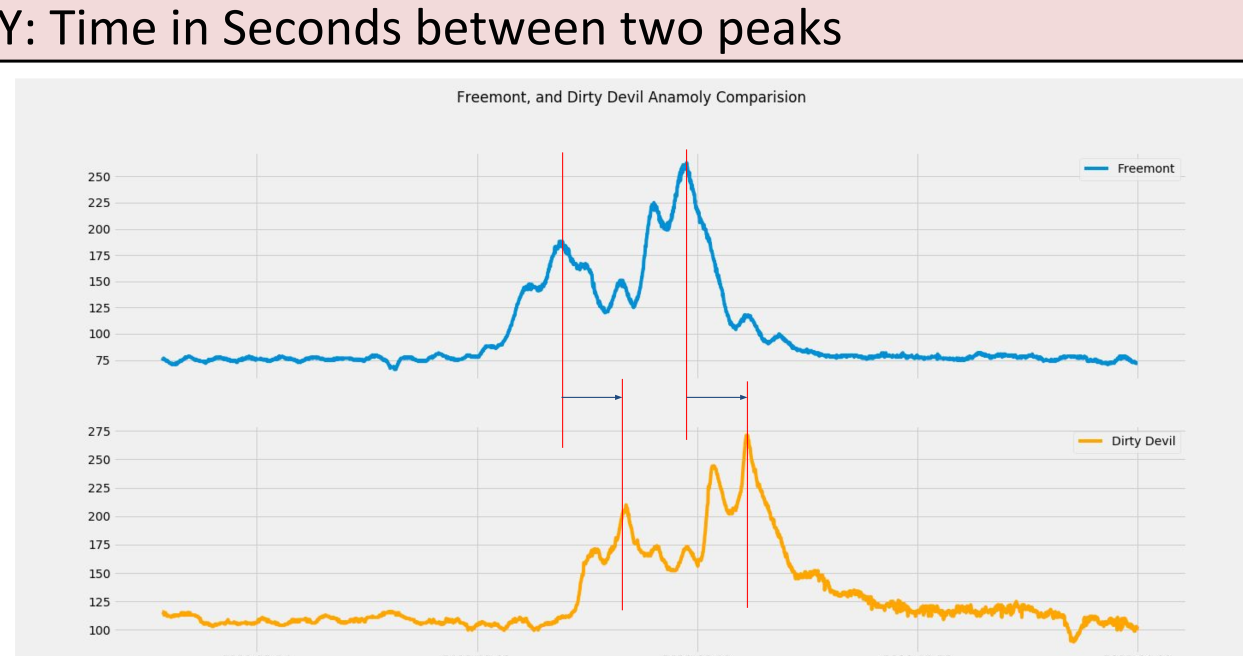
Data Collection

- Data was retrieved from the United States Geological Survey
- Discharge(cfs) measurements every 15 minutes
- San Rafael Upper-> San Rafael Lower (50.2km)
- Fremont -> Dirty Devil (182km)



Data Processing

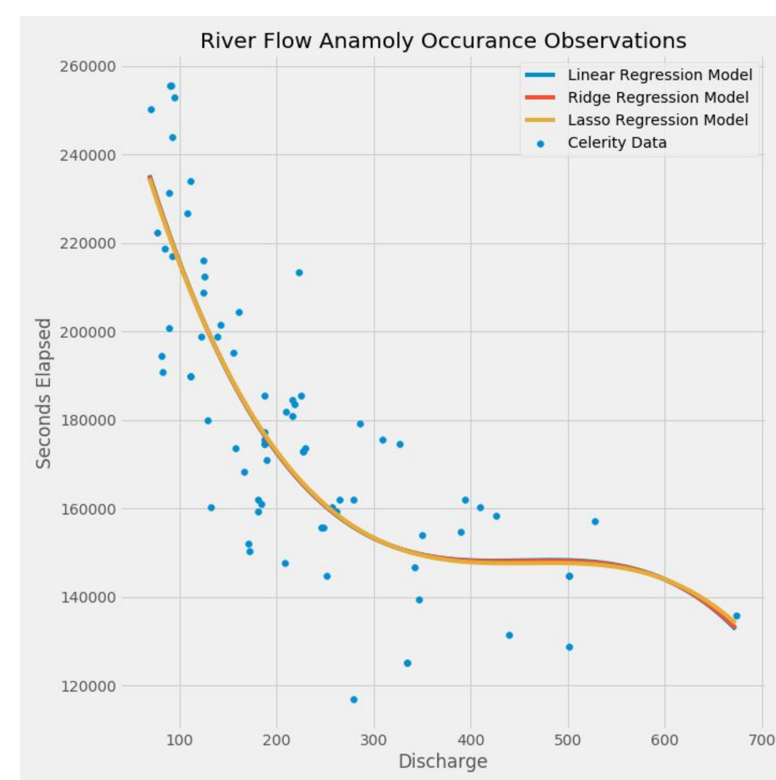
- Used peak detection tool for detecting peaks in both datasets
- Matched peaks on corresponding data sets
- Used this to calculate time it took for water to travel downstream
- Leaves us with 2 Dimensional Dataset
  - X: Discharge Value
  - Y: Time in Seconds between two peaks



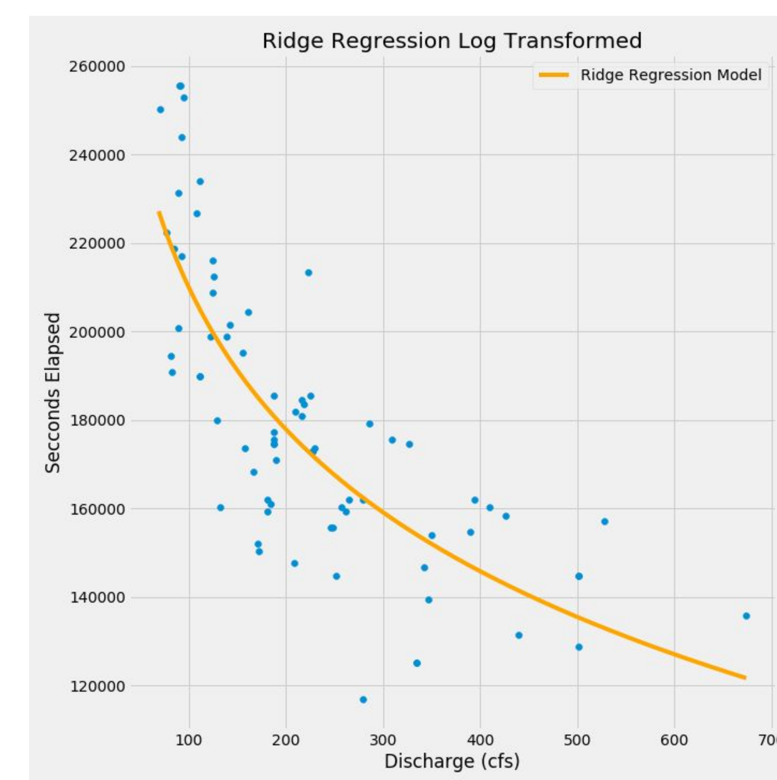
Methods

- Least Squares Linear Regression did not perform well on untransformed data
- Extended linear regression techniques to polynomial features
  - $(x, y) \Rightarrow (1, x, x^2, x^3, y)$
- Extended linear regression techniques to log transformed x-value
  - $(x, y) \Rightarrow (\log(x), y)$
- Regression Methods Used:
  - Least Squares
  - Ridge
  - Lasso

Polynomial



Logarithmic



Results

- 80% Training, 20% Testing 5 Fold Cross Validation
- Used R-Squared Metric to compare models
  - $1 - ||M(x) - y|| / ||\hat{y} - y||$

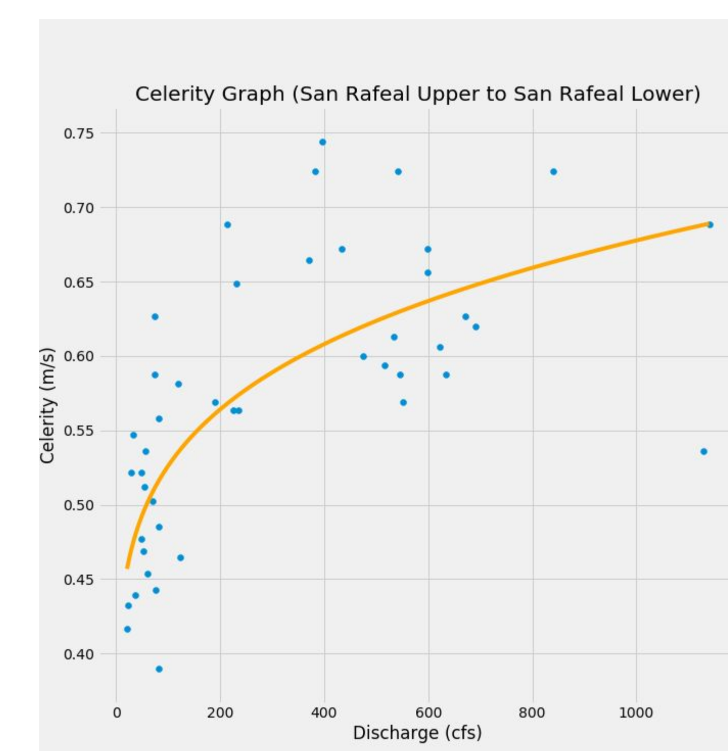
Table 1: K-Fold Results for San Rafael

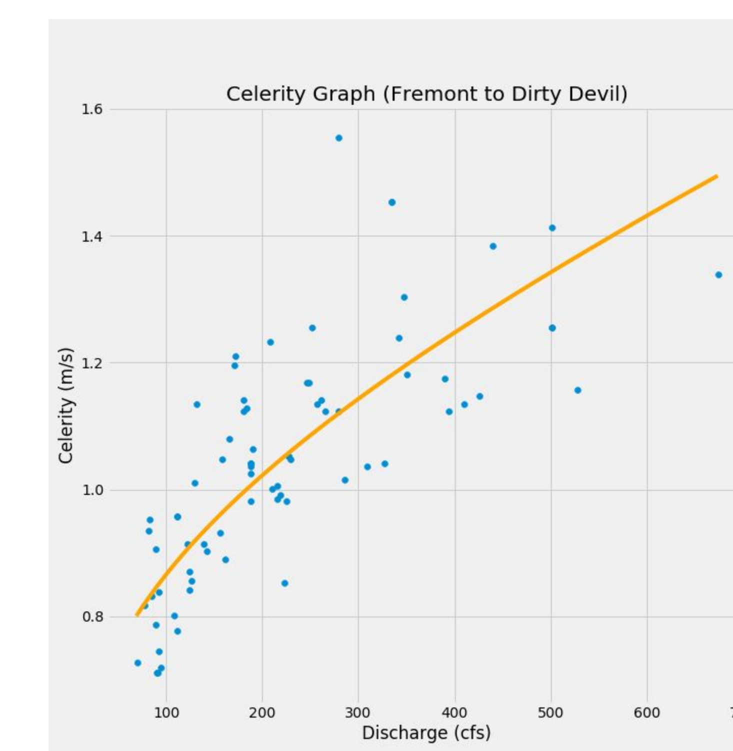
Regression Type	5-Fold Average Score	Hyperparameters:	
		Degree	Alpha
Normal Regression Polynomial Features	0.44535	3	Not Applicable
Ridge Polynomial Features	0.44535	3	.001
Lasso Polynomial Features	0.44573	3	.01
Normal Regression Log Scaled	0.48498	Not Applicable	Not Applicable
Ridge Log Scaled	0.48736	Not Applicable	1
Lasso Log Scaled	0.48573	Not Applicable	100

Table 2: K-Fold Results for Fremont/Dirty Devil

Regression Type	5-Fold Average Score	Hyperparameters:	
		Degree	Alpha
Normal Regression Polynomial Features	0.55823	3	Not Applicable
Ridge Polynomial Features	0.55859	3	10
Lasso Polynomial Features	0.56898	3	10
Normal Regression Log Scaled	0.56762	Not Applicable	Not Applicable
Ridge Log Scaled	0.57161	Not Applicable	10
Lasso Log Scaled	0.56828	Not Applicable	100

Graphs of Celerity





Conclusion

- Celerity best modeled with ridge regression applied to a logarithmically transformed dataset
- San Rafael best K-Fold Result: 0.48736
- Fremont/Dirty Devil best K-Fold Result: 0.57161

Future Work

- Automating the data collection process
- Extending the model to predict discharge values along with time
- Investigate data transformation that would allow us to generalize one rivers model to another