Data Science for Storm Events

Andrew Porter, Benjamin Gildersleeve, Mike Ni, Laura Dietz, Adam Wymore

University of New Hampshire Department of Computer Science

Introduction

The objective of our project is to discern a relationship between flow rate and water solute concentrates. Doing so, will allow current researchers at the University of New Hampshire to aid in the development of plans to clean watersheds of solute, after storms have passed

Questions

- Is there a relationship between discharge and solute concentration?
- Do storms behave differently depending on the season (fall, spring, winter, etc.)?

Classification

How can we predict when a storm has occurred?

- Manually label historical storm data
- Create training and test set
- Use flow rate (Q) to predict on storm
- Normalize features for cross site predictions

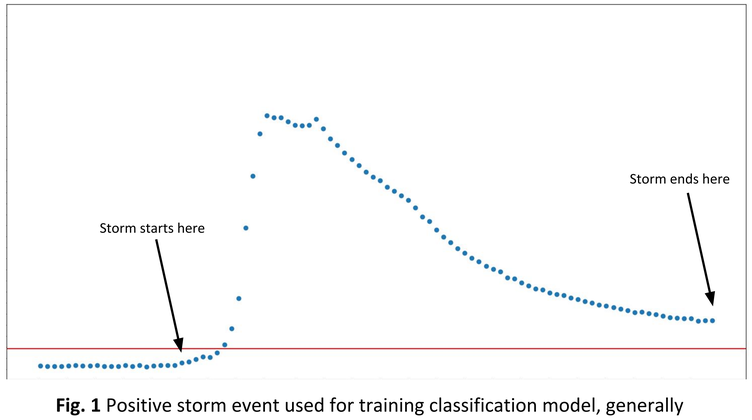

Fig. 1 Positive storm event used for training classification model, generally labeling points above the read mean flow rate line.

Normalization

Min/Max Normalization:

- Taking labeled dataset based on marked "isStorm"
- Flow rate on a 0-1 scale, 1 being the max flow rate for a specific site

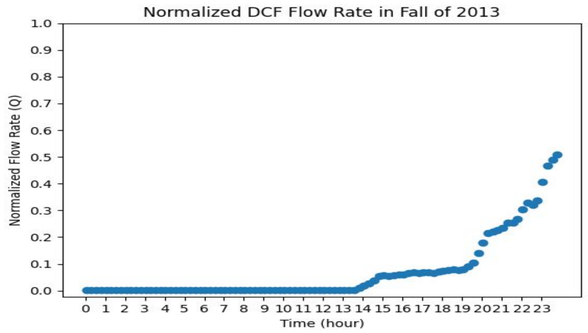

Fig. 2 Normalized storm events showing flow rate versus time

Data Cleaning

- Recode
- Distribution (shown below)
- Using classification model to locate oddly/questionably shaped trends

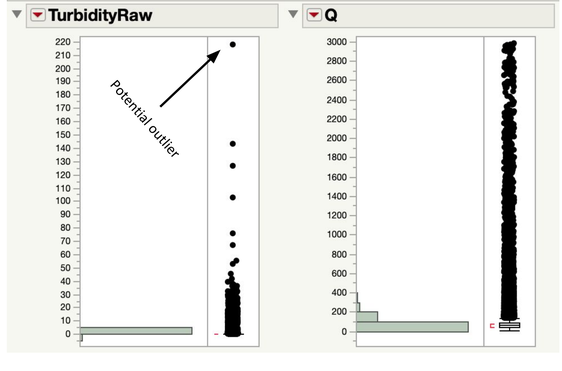

Fig. 3 Two plots from the Distribution platform that plots the variables TurbidityRaw and Q variables from the MCQ site.

Results

- 97.2% prediction accuracy
  - 58.7% accuracy on positive storm events (true storm events that were predicted as such between sites)

Precision (proportion of positive identifications)	65.56%
Recall (proportion of actual positives identifies)	58.72%

- 95.69% cross-site prediction accuracy
  - 11.55% on positive storm events

Conclusions & Future Work

- Flow rate alone is not enough
- Multi-site model bad at predicting positive storm events
- Test trained models for full year instead of cyclical basis
- Include more features into classification model
- Train with Z-score normalized data
- Gap-filled data produced abnormal trends
- Self data cleaning program possible but complicated

References

Wymore, Adam S., et al. "Hysteretic Response of Solutes and Turbidity at the Event Scale Across Forested Tropical Montane Watersheds." *Frontiers in Earth Science*, vol. 7, 2019, doi:10.3389/feart.2019.00126.