FORECASTING

STREAMGAGE ELEVATIONS AT USGS NODES

WWW.THOMASOTT.IO

October 21, 2024

CONTENT

ABOUT ME

30+ years in Civil Engineering, NJ PE, Machine Learning/Data Science Expert, Dogs Love me.

Engineering consultancy in Northern New Jersey providing Site & Drainage Engineering, Variance Applications, and Green Infrastructure Design

POMPTON RIVER CROSSING AT ROUTE 23, NEW JERSEY

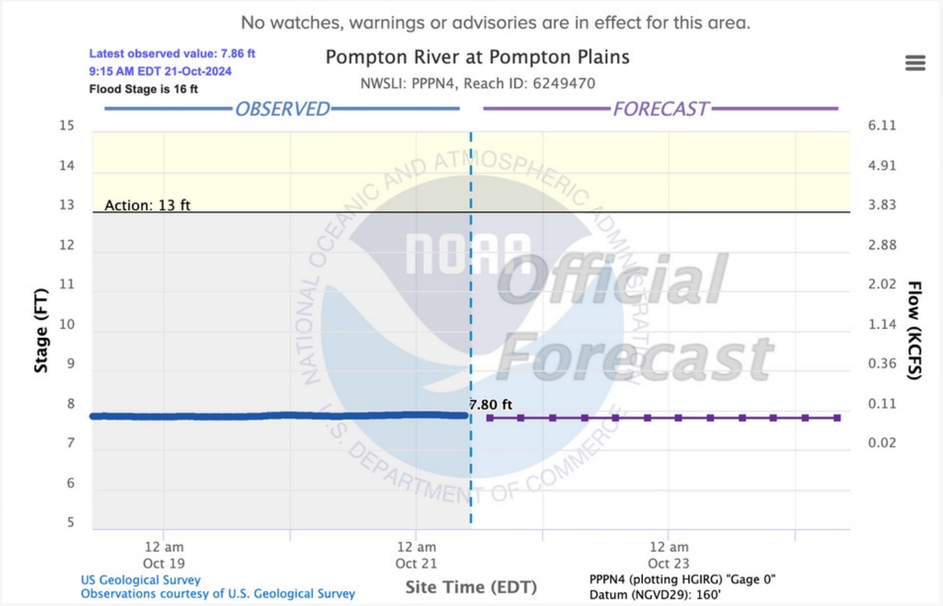
THE USE CASE

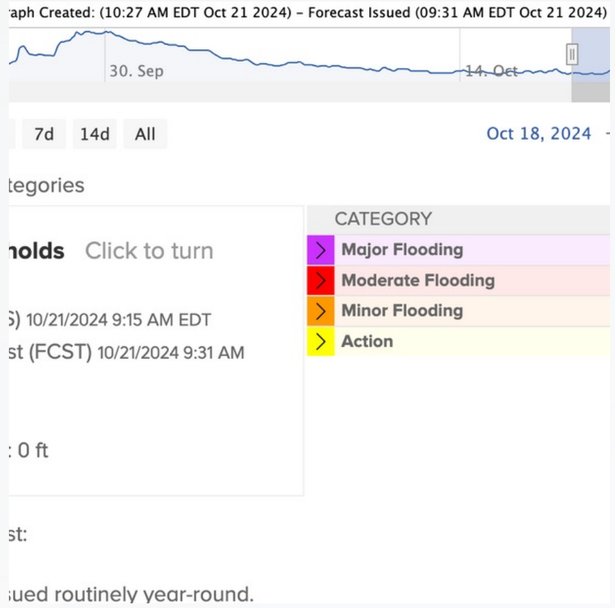
Wayne NJ August 30, 2011 - Hurricane Irene

WHY ANOTHER FLOOD PREDICTION PROCESS

NOAA Streamgage Forecast

- Existing Flood Stage Simulations exist already (NWM)*
- Why Another Flood Prediction Process?
- Physics based vs data driven approach



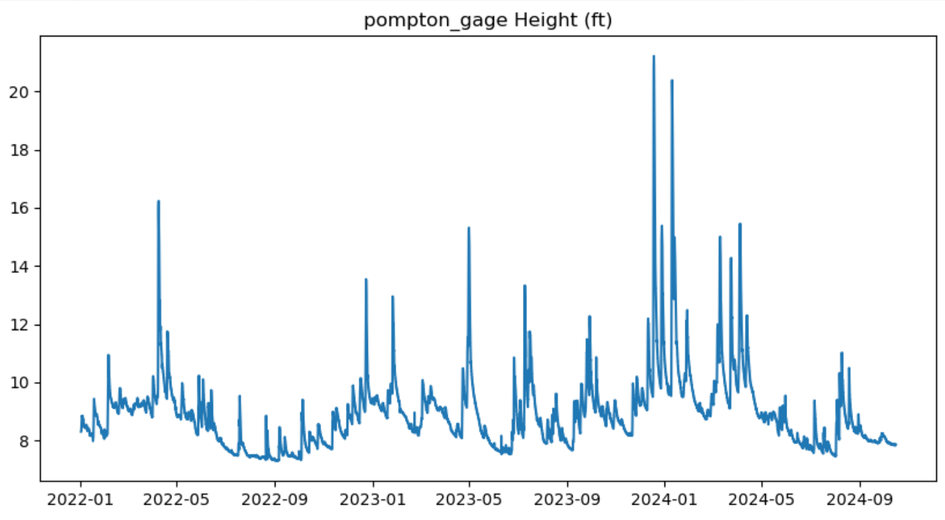


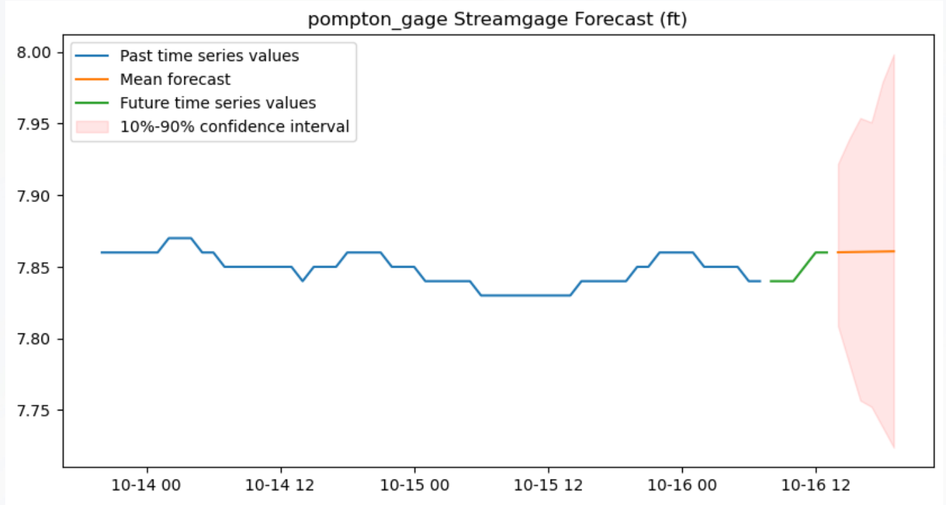
*NWM - National Water Model

WHY ANOTHER FLOOD PREDICTION PROCESS

USGS Streamgage Nodes

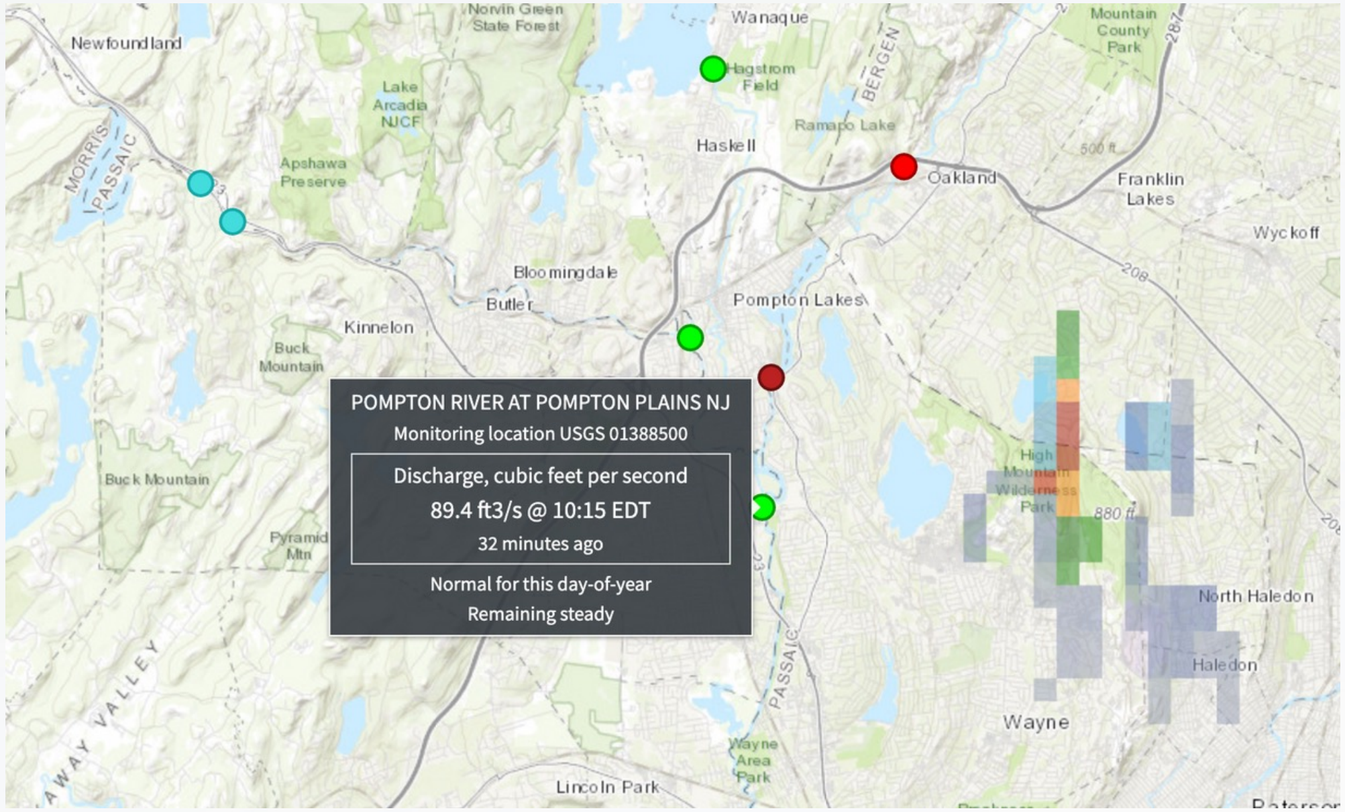
- Why? Apply a data driven approach for flood predictions
- Explore the possibility of inferencing at the streamgage node.
- Forecast 1 hour, 3 hour, and 6 hour streamgage elevation using machine learning
- Use inferencing results to inform USGS researchers and turn on/off different instrumentation (remotely)





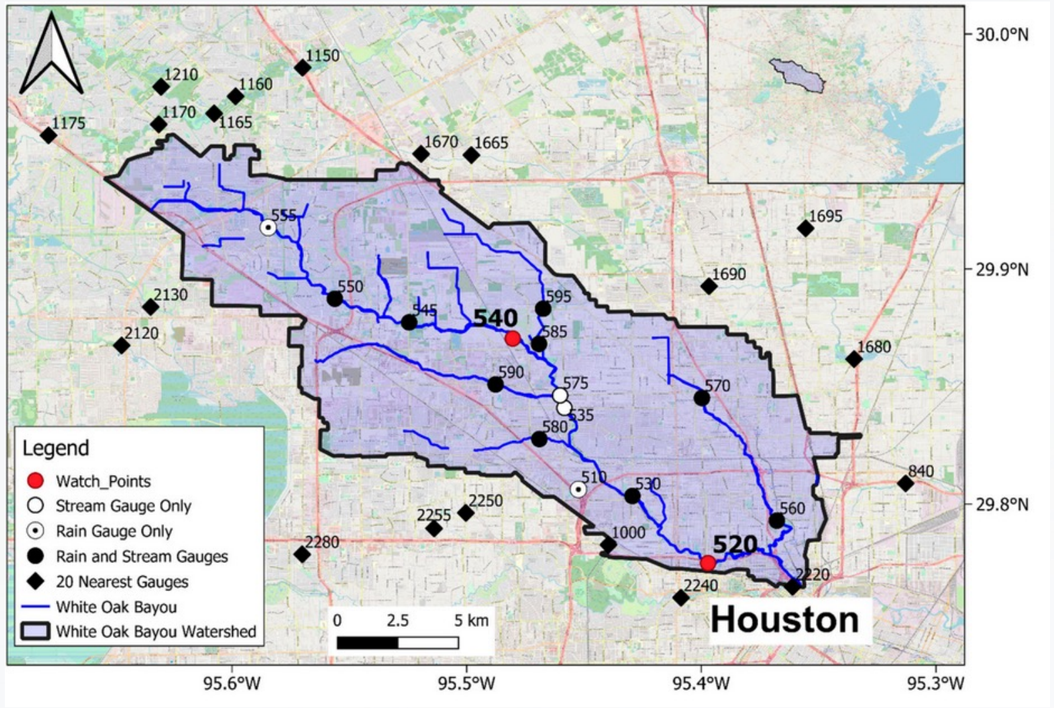
POMPTON RIVER CROSSING - WAYNE NJ

- Pompton River cross at Route 23 in Wayne, NJ
- Confluence of Ramapo & Pequannock River just north of bridge
- USGS Streamgage #01388500
- Route 23 typically closes at moderate flood stage
- Moderate flood stage elevation = 18 feet



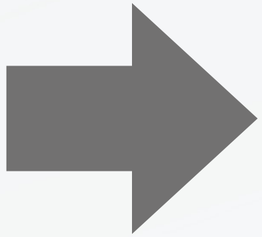
APPLYING DATA DRIVEN MODELS

- "Data-Driven Flood Alert System (FAS) Using Extreme Gradient Boosting (XGBoost) to Forecast Flood Stages"
- Research focused on streamgages in Houston watershed
- Data driven approach
- Apply open source Autogluon library



APPLYING DATA DRIVEN MODELS

- Data Inputs
  - Pequannock River Streamgage #01382800 (height elevation)
  - Ramapo River Streamgage (height elevation)
  - Ramapo Precipitation (inches)
  - Pompton Plains Streamgage (height elevation)
  - Pompton Plains Precipitation (inches)



- No stream flows (CFS) used / directly correlated
- No streamgage rate of change (ROC) used
- Direct input of streamgage heights and precipitation data only

- Utilize Autogluon library
- Iterates over many different algorithms
- Easy PKL file deployment
- Optimized for RMSE (less sensitive to spikes/penalizes large errors)

APPLYING DATA DRIVEN MODELS

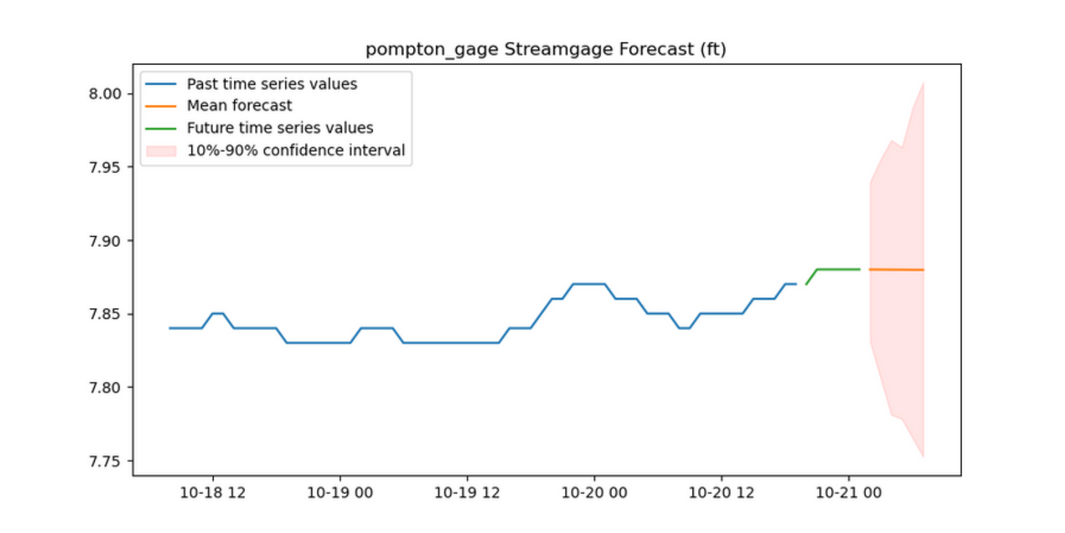
- Experiment Setup
  - Train & Test data split
  - Forecast for 1, 3, 6, 12, and 24 hours
  - Generate forecast plots
- Optimize for RMSE
- Train using “fast_training” / smaller selection of models
- Evaluate Naive, Seasonal Naive, Theta, and ETS models

- AutoGluon Results
  - 1 hr - RMSE: 0.002 / Weighted Ensemble
  - 3 hr - RMSE: 0.005 / Weighted Ensemble
  - 6 hr - RMSE: 0.007 / Theta
  - 12 hr - RMSE: 0.013 / Theta
  - 24 hr - RMSE: 0.012 / Weighted Esemble

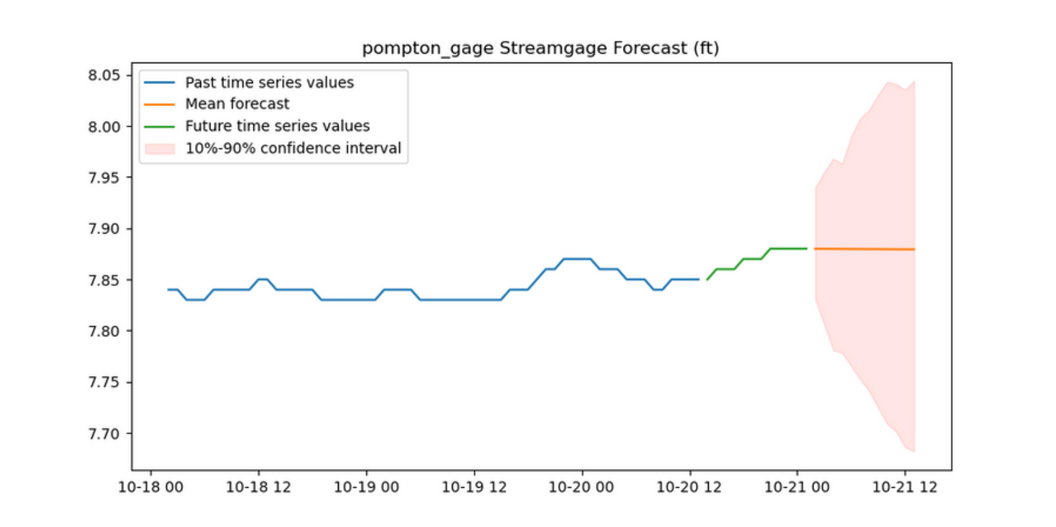
- Weighted Ensemble is a bagged meta-model
- Theta is a decomposition approach to forecasting

AutoGluon iterates through multiple models and hyper-parameters to select the best perform model

SCORING DATA DRIVEN MODELS



3 Hour Forecast w/Confidence Levels

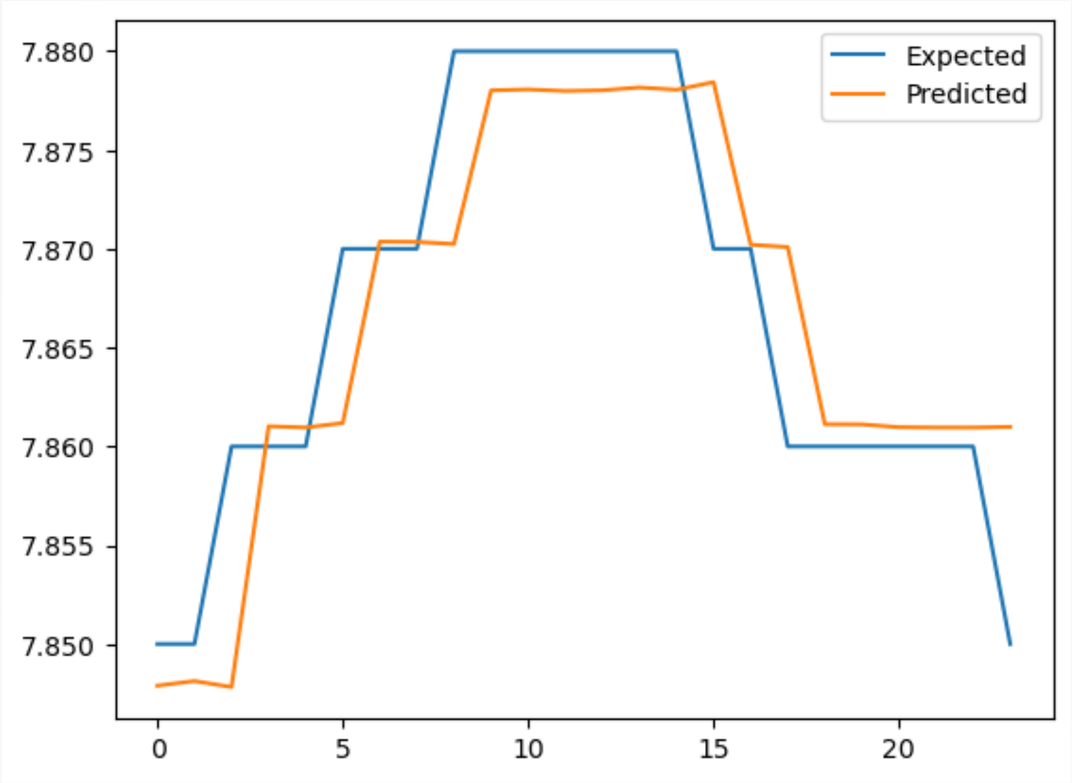


12 Hour Forecast w/Confidence Levels

WHAT ABOUT XGBOOST?

Is XGBoost all you need?

- XGBoost Foward Walk Results
  - 1 hr – RMSE: 0.011
  - 3 hr – RMSE: 0.006
  - 6 hr – RMSE: 0.005
  - 12 hr – RMSE: 0.006
  - 24 hr – RMSE 0.006



NEXT STEPS

1. Train Autogluon and XGBoost Models on a Raspberry Pi or Dell 3200
2. Evaluate Memory Consumption/usage
3. Create Training and Inference Templates for Auto-training and inferencing at streamgage
4. In field deployment

THANK YOU

Q&A

Contact:
Thomas N. Ott, PE
phone: 973-750-8877
email: info@thomasott.io