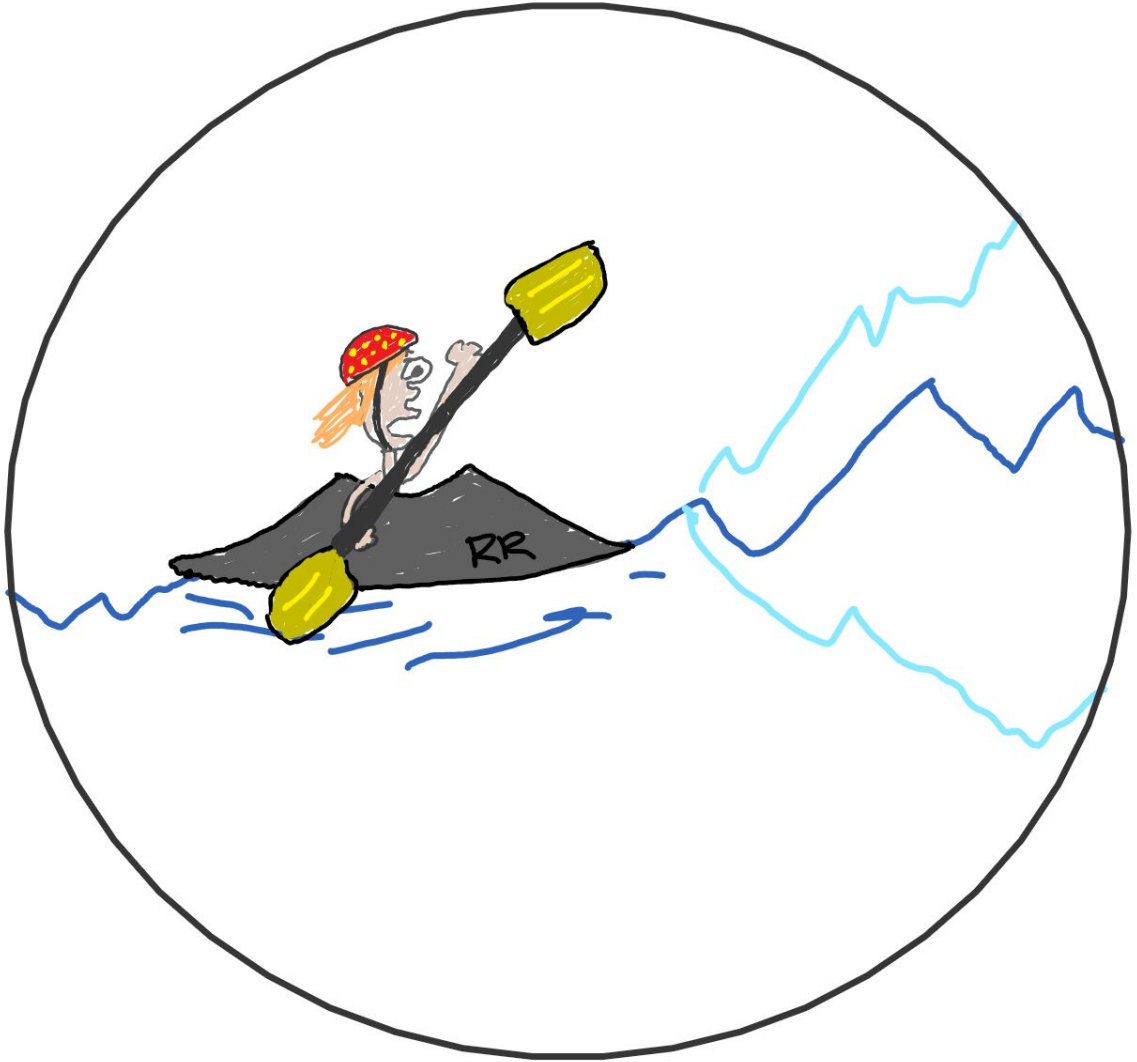
River Runner

Whitewater Kayaking Predictions

Ryan Bald, Kenten Danas, Luke Waninger

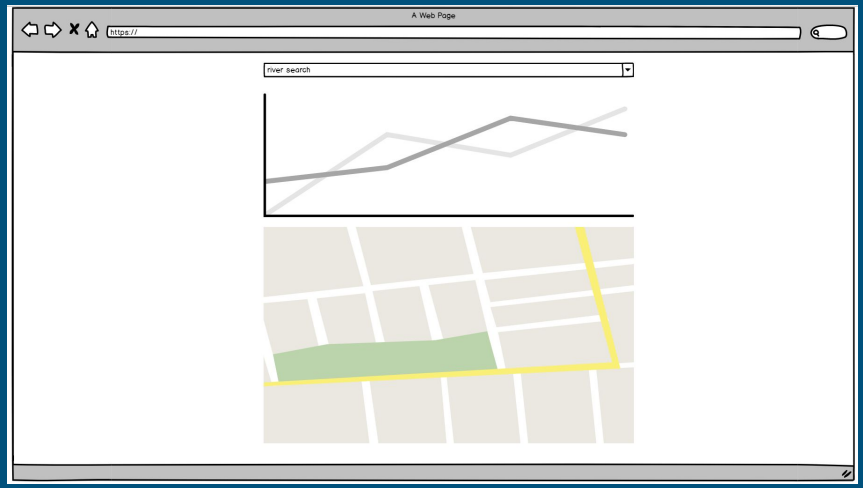


Background

- Whitewater kayaking is a popular sport in western Washington
- But, rivers are highly dynamic. Flow rates can change drastically in short time periods.
- There are currently no reliable tools to predict a river's flow rate, meaning paddlers are often left to figure out whether they can run a given river less than a day in advance

Background

- River Runner aims to provide predictions for whitewater kayaking run flow rates up to a week in advance
- Predictions are made based on ARIMA models of past flow rates along with exogenous weather predictors
- Paddlers will be able to access predictions for a chosen run(s) via a publically available graphical user interface



Data Sources: Professor Paddle

- Used for:
  - River run data including river run names and max/min runnable flow rates
- Obtained by:
  - Single use scraping functions
- Limitations:
  - This site does not have max/min runnable flow rates for all runs

River Name	Run Name	Mapplet / Media / Beta / Rapid #	Class	Flow & Units	Updated
American	1. Lodgepole Campground to Hell's Crossing		II-III+	A 913 cfs	05-25-2018 05:30:00
American	2. Hell's Crossing to American Forks		III-IV	A 913 cfs	05-25-2018 05:30:00
American	3. American Forks to Sawmill Flat (on Naches)		II+	A 913 cfs	05-25-2018 05:30:00
Austin Creek	Triple drop Park n Huck		IV-V		
Bacon Creek	Falls Creek to Skagit River confluence		II+(V)	A 974 cfs	05-25-2018 05:00:00
Bailie's Creek	Top to Slide		IV		
Baker Creek	Old Bridge to Middle Satsop	(1)	III-IV	B 443 cfs	05-25-2018 05:15:00
Baker River	North Cascades National Park to Baker Lake		II		
Beckler	Rapid River to S Fork Skykomish		II-III	A 10300 cfs	05-25-2018 05:30:00
Big Creek	Bridge above Narrows to South Boundary Road Bridge		III		
Big Quilcene	1 - bl. Tunnel Creek to Rainbow Campground (Upper)		V+	B 163	05-25-2018 05:30:00
Big Quilcene	2 - Rainbow Campground to fish hatchery (Lower)	(2)	IV-V		
Big Sheep Creek	Sheep Creek Rd. mile 7.5 to Sheep Creek C.G.		II(III)		
Bogachiel	Park Boundary to Hwy 101		II	**	05-25-2018 05:00:00
Boulder	Boulder Falls to mouth		III		
Box Canyon Creek	Box Canyon Creek to Kachess Reservoir		III-IV+(V)		

Data Sources: USGS

- Used for:
  - Historical and ongoing streamflow data
- Obtained by:
  - Repeated calls to the USGS Instantaneous Values REST web service
- Limitations:
  - Restricts the amount of data returned
  - Not every station measured streamflow

Data Sources: NOAA

- Used for:
  - Historical exogenous predictor data
    - snowpack, precipitation, temperature, humidity
- Obtained by:
  - Manual requests submitted through NOAA’s climate data online portal
- Limitations
  - Restricted how much data could be accessed with a single manual request
  - Data was provided via an email link and could not be automated for ongoing retrieval

Data Sources: DarkSky

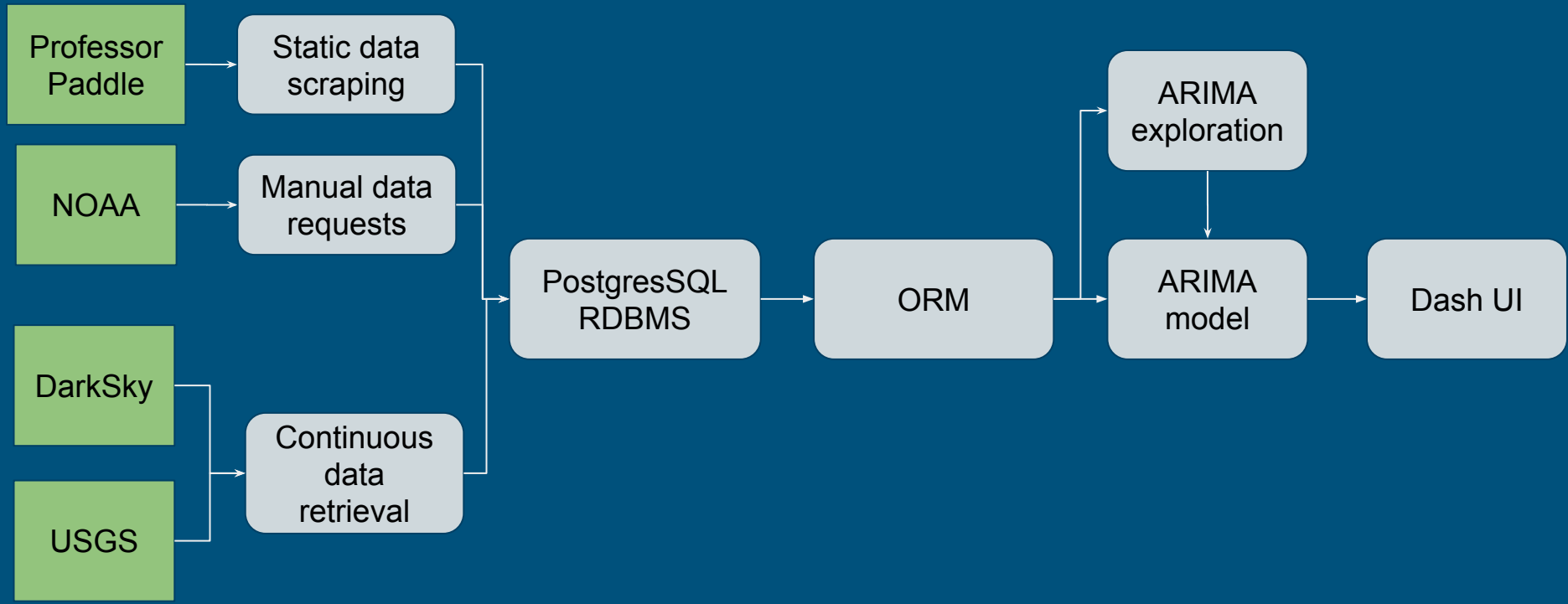
- Used for:
  - Ongoing exogenous predictor data
    - precipitation, temperature, humidity
- Obtained by:
  - Continuous API calls to DarkSky (https://api.darksky.net/forecast)
- Limitations
  - Data is retrieved for the previous day, every day, and works as expected. We found no limitations in regards to our project.

Use Cases

- Paddlers use RiverRunner to help plan kayaking trips by getting a prediction of a selected river's flow rate, which provides an indication of whether the river will be runnable
- User accesses the tool through a webpage, and selects a river run by name from the drop-down filter or by data point on the map
- The system responds with a plot for that run which includes:
  - 3 weeks of historic flow rate
  - 7 days of predicted flow rate
  - Highlighted band between the maximum and minimum runnable flow rates (where available)

Demo

Design


Professor Paddle
Static data scraping
NOAA
Manual data requests
DarkSky
USGS
Continuous data retrieval
PostgresSQL RDBMS
ORM
ARIMA exploration
ARIMA model
Dash UI

Design - Ongoing Data Collection

- Daily script that retrieves all weather/river-related data from the previous day and computes predictions
- Script that can retrieve all weather/river-related data over a specified date range if need be

```
def daily_run():
    context = Context(settings.DATABASE)
    session = context.Session()

    get_weather_observations(session)
    get_usgs_observations()
    compute_predictions(session)

    session.close()
```

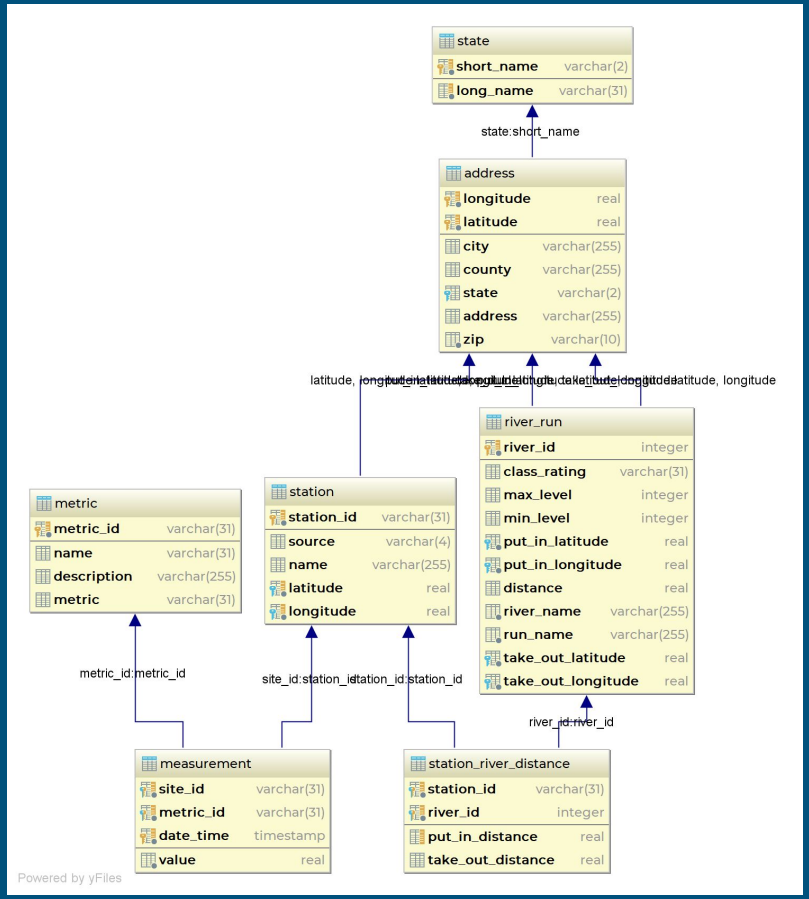
```
predictions = arima.arima_model(run.run_id)

to_add = [
    Prediction(
        run_id=run.run_id,
        timestamp=pd.to_datetime(d),
        fr_lb=round(float(p), 1),
        fr=round(float(p), 1),
        fr_ub=round(float(p), 1)
    )
    for p, d in zip(predictions.values, predictions.index.values)
]

repo.clear_predictions(run.run_id)
repo.put_predictions(to_add)
log(f'predictions for {run.run_id}-{run.run_name} added to db')
```

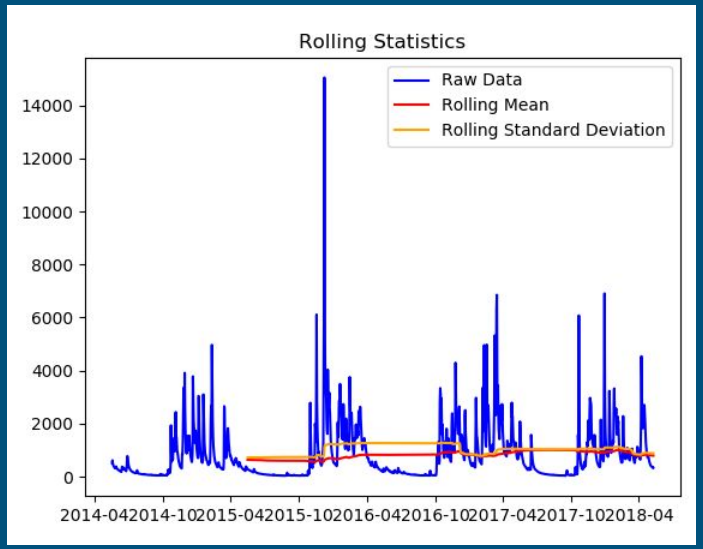
Design - RDBMS

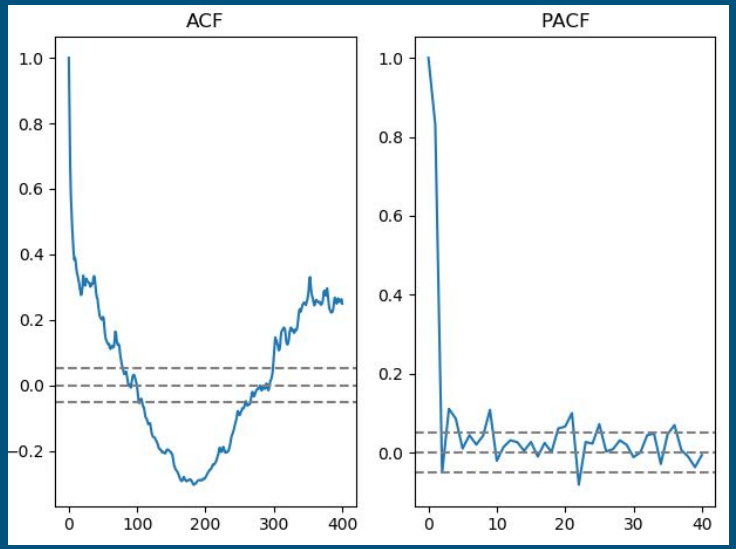
- Persistence is managed through an RDBMS - PostgresSQL 10.3 - Ubuntu Server 16.04 LTE



Design - ARIMA Exploration

- Static module used once for data exploration to determine best model parameters
- Analysis included visual and statistical (Dickey-Fuller) tests for stationarity, autocorrelation analysis for AR order and partial autocorrelation analysis for MA order

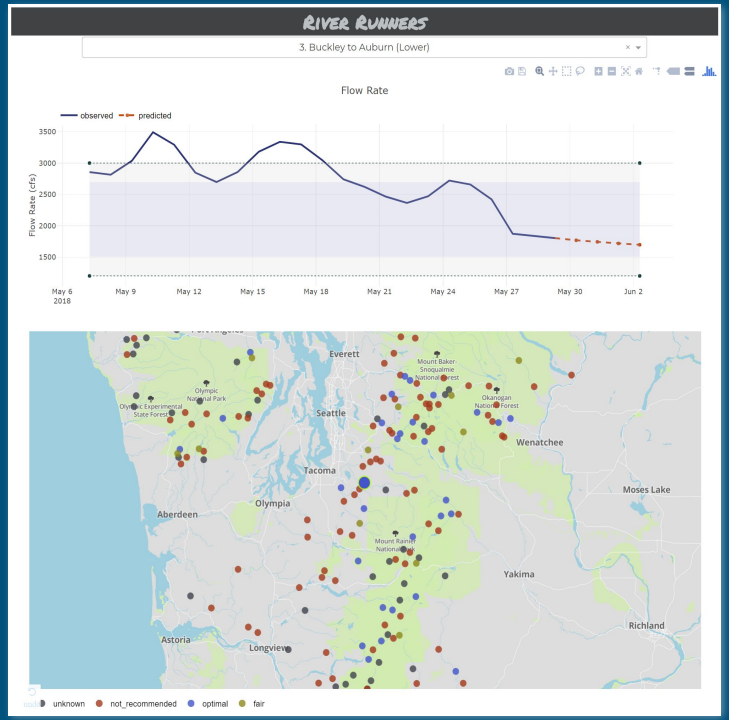




Design - ARIMA Model

- Future river flow rates are predicted using an autoregressive integrated moving average (ARIMA) model generated from historic USGS river flow rate time series data
- Temperature and precipitation are included in the models as exogenous predictor variables
- Models are generated using the past four years of historical data up to the current day, and predictions are made for the future seven days.
- Modelling is completed using built in Python functions

Design - Dash Front End



CI Testing - TeamCity

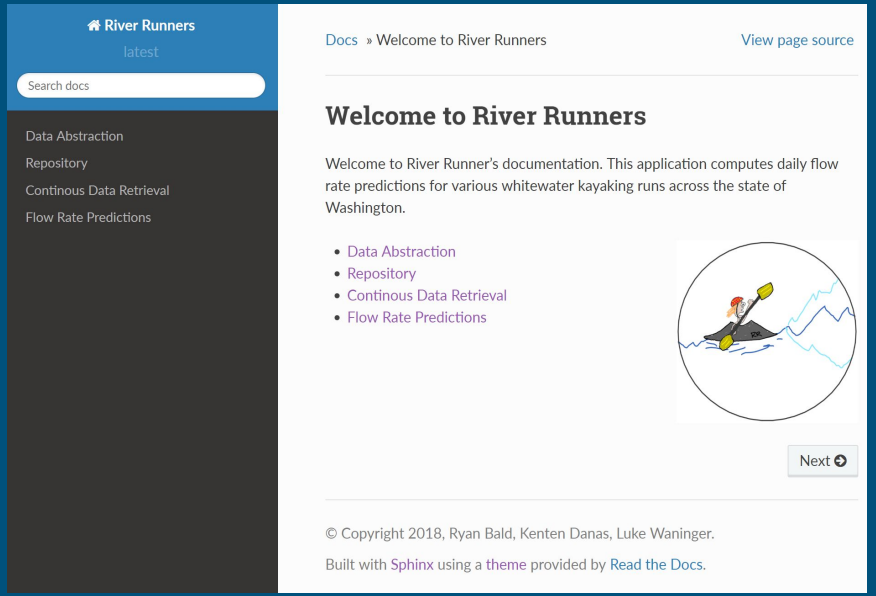
TC build | success

Project Structure

High Level

- RiverRunner/
  - doc/
  - examples/
  - riverrunner/
    - tests/
  - .gitignore
  - LICENSE
  - README.md

https://github.com/kentdanas/RiverRunner



Lessons Learned & Future Work

- Lessons Learned
  - Getting one ARIMA model to converge for 300 different datasets is difficult
  - Dash is a quick and effective tool for web visualization
  - Database must be prepped for daily modeling
- Future Work
  - Further refine the models used to improve accuracy and extend the forecasted flow rates beyond 7 days
  - Integrate weather forecast data for exogenous predictors
  - Expand scope of project to include more river runs

Questions?