Automated Water Resources Analysis Using United States Geological Survey (USGS) Data in Chesapeake Bay Watershed

5 August, 2020

Ruoyu Zhang, Yibo Wang, Hyunglok Kim, Keyu Chen & Yusheng Jiang

Introduction

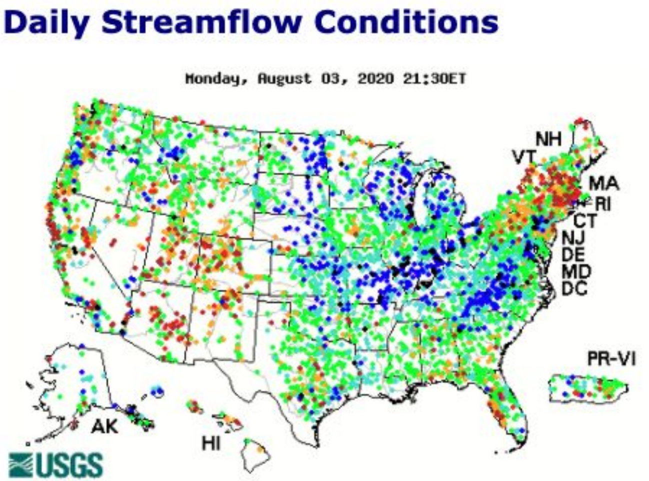
1. Stream water is the major freshwater sources for human activities
2. Due to climate change:
   - The spatial distribution of stream water may change
   - We will experience more frequent hydrologic hazards (i.e., flood and drought)

Before we predict the future, we need to understand how water behaved in the past

Questions to Address

1. What is the best way to quickly download the discharge data?
2. Besides discharge data, what other data can we get for each gage?
3. How does the monthly and yearly trends match at gages in the Chesapeake Bay Watershed?
4. Are there clusters of the spatial patterns of the trends in our study area?

Program Design



USGS science for a changing world

Parent Class: USGS_Gage(USGS_id)

Attributes

id: *required input parameter* USGS Gage ID, string

vars_info: variables information, dictionary

geo_info: gage location info, dictionary

Methods

getVarsMetaData():
save variable names, start/end dates to vars_info, return a pandas dataframe

getGeoMetaData():
save gage's county, state, and coordinates to geo_info, return a dictionary

Child Class: USGS_Gage_DataRetriever()

Attributes

ismetric: *optional input parameter* if save data in metric unit, Boolean, default True

data: time-series of daily discharge data, pandas DataFrame

autoDates: *optional input parameter* if use the metadata period, Boolean, default True

startdate: *optional input parameter* start date to download data, str, 'YYYY-MM-DD'

enddate: *optional input parameter* end date to download data, str, 'YYYY-MM-DD'

Methods

getDailyDischarge(drop=True):
download data, drop no data value (<0) by default, return a pandas dataframe

getStatistics():
get min, median, max, mean, and std of the data, return a dictionary

getUnit():
return the unit of discharge cubic feet/second or cubic meter/second

plotTimeSeries(savefig=path):
plot discharge time-series, can save to a *path (optional)*

findLargestEvents(top_x):
return a dataframe with largest X events in history (*top_x = X*)

trendTest(time_scale, least_records, target_alpha, plot):
perform Mann-Kendall test on data to detect trend within the define time
time_scale: "M" (monthly) or "Y" (yearly) to detect trend at define scale
least_records: least months or years to perform the test, number
target_alpha: significant level of the trend, 0.05 by default
plot: if plot the resampled data with the trend line, Boolean, False by default

https://github.com/ruoyu93/USGS_DataRetriever

Unit test

```python
###############################################################################
# unit test 11 - test trendTest()
###############################################################################

class USGS_DataRetriever_Test(unittest.TestCase):
    def test_trendTest(self):
        # set up
        site = USGS_Gage_DataRetriever("02024752")
        stat = test11.getStatistics()
        # test if all the data meet our expectation
        trend_result_m, slope_result_m, R_TS_m, R_MK_m, reason = site.trendTest('M', 120, target_alpha = 0.05)
        trend_result_y, slope_result_y, R_TS_y, R_MK_y, reason = site.trendTest('Y', 10,target_alpha = 0.05)
        list_test_11_m = [trend_result_m, slope_result_m, R_TS_m, R_MK_m, reason]
        list_test_11_y = [trend_result_y, slope_result_y, R_TS_y, R_MK_y, reason]
        # test if trend_result_m, slope_result_m, R_TS_m, R_MK_m, reason meet our expectation
        # test if trend_result_y, slope_result_y, R_TS_y, R_MK_y, reason meet our expectation

        self.assertEqual(list_test_11_m,[1,0.14427400917512528,(0.14427400917512528, 62.65598104129241,0.008914136718965527, 0.32995723856417797),(1, 0.2
        self.assertEqual(list_test_11_y,[1, 4.649560190190382, (4.649560190190382, 69.51532936073576, 1.5004534761252657, 7.679532771685558), (1, 4.81872


if __name__ == '__main__':
```

Output (

All 11 tests passed!

```
----------------------------------------------------------------------
USGS Gage 02024752 has following variables:
    Discharge(Mean) from 2005-10-01 to 2020-08-01
----------------------------------------------------------------------

    Setting new dates from 2005-10-01 to 2020-08-01
Summary of flow from 2005-10-01 to 2020-08-01
    Min: 13.677037095414589
    Median: 56.633693976871996
    Max: 1458.317619904454
    Mean: 107.0119775066119
    Standard Deviation: 136.498088589308

.
----------------------------------------------------------------------
Ran 1 test in 2.301s

OK
```

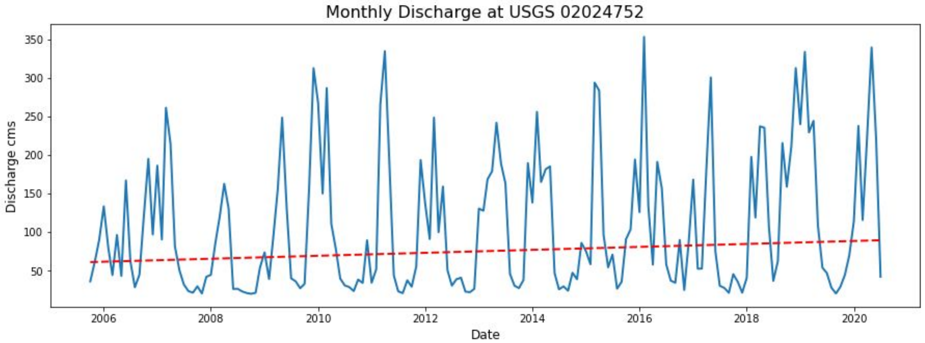
Obtain Trend Data

We looped through all USGS gages within the Chesapeake Bay watershed

All data obtained are saved into a Pandas DataFrame

	GageID	Start Data	End Data	Min	Median	Max	Mean	slope (M)	p-value (M)	slope (Y)	p-value (Y)	County	CountyID	State	Coordinate
0	1656120	1996-10-01	2000-01-18	0.46	49.0	5720.0	170.910846	NaN	NaN	NaN	NaN	Prince	51153	VA	(38.64150829, -77.5124873)
1	1654000	1947-10-01	2020-07-31	0.00	12.0	3600.0	29.781916	0.007347	0.000253	0.111807	0.005671	Fairfax	51059	VA	(38.81289066, -77.2283158)
2	2032000	1943-08-05	1946-09-29	0.00	4.6	1340.0	18.271615	NaN	NaN	NaN	NaN	Albemarle	51003	VA	(38.13485668, -78.7358548)
3	1669000	1951-10-01	2020-07-31	0.00	22.0	1540.0	31.552949	-0.001507	0.142754	0.020466	0.183291	Essex	51057	VA	(37.8770819, -76.900521)
4	2039500	1926-04-01	2020-07-31	0.07	162.0	28000.0	284.408989	-0.020130	0.036976	-0.090548	0.197380	Cumberland	51049	VA	(37.3070965, -78.388607)

Trend analysis: Mann-Kendall Test



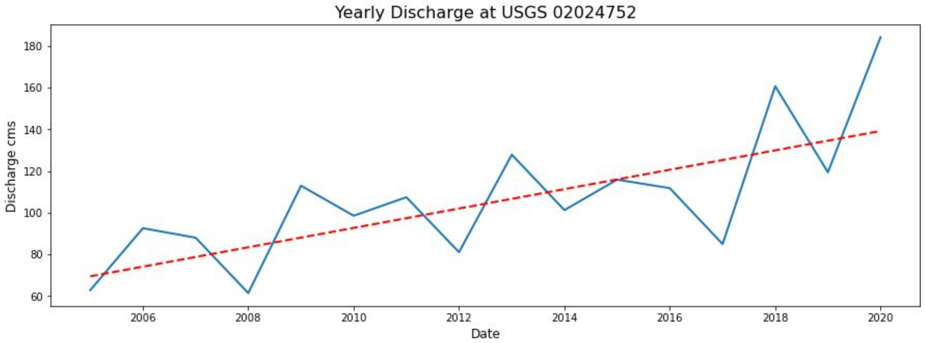
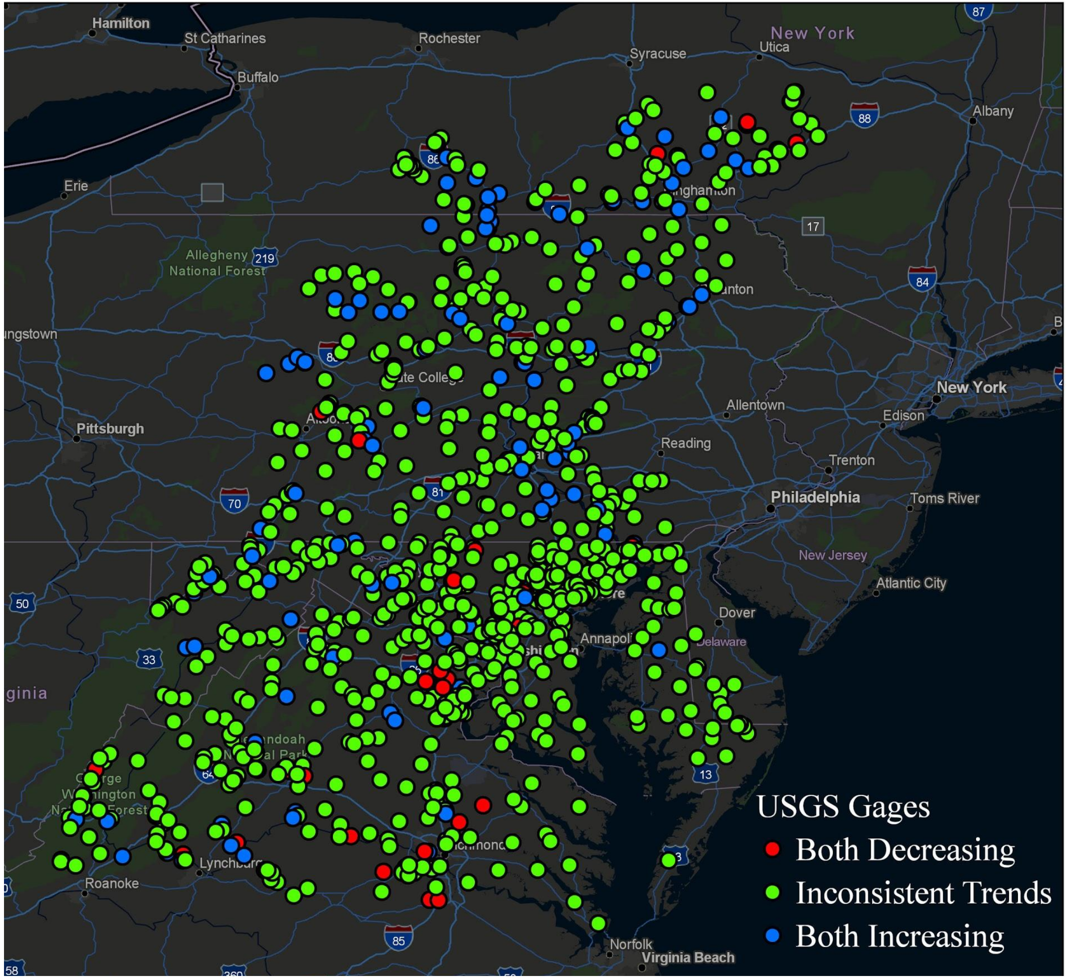


Figure. Time series of the discharge at 02024752 USGS site, James River at Blue Ridge Parkway near Big Island, VA, with trend line

- If a gage shows significant monthly/yearly trends, the program draws the trend line (red) automatically
- Each gage’s statistics can be stored and presented (next slide)

Results

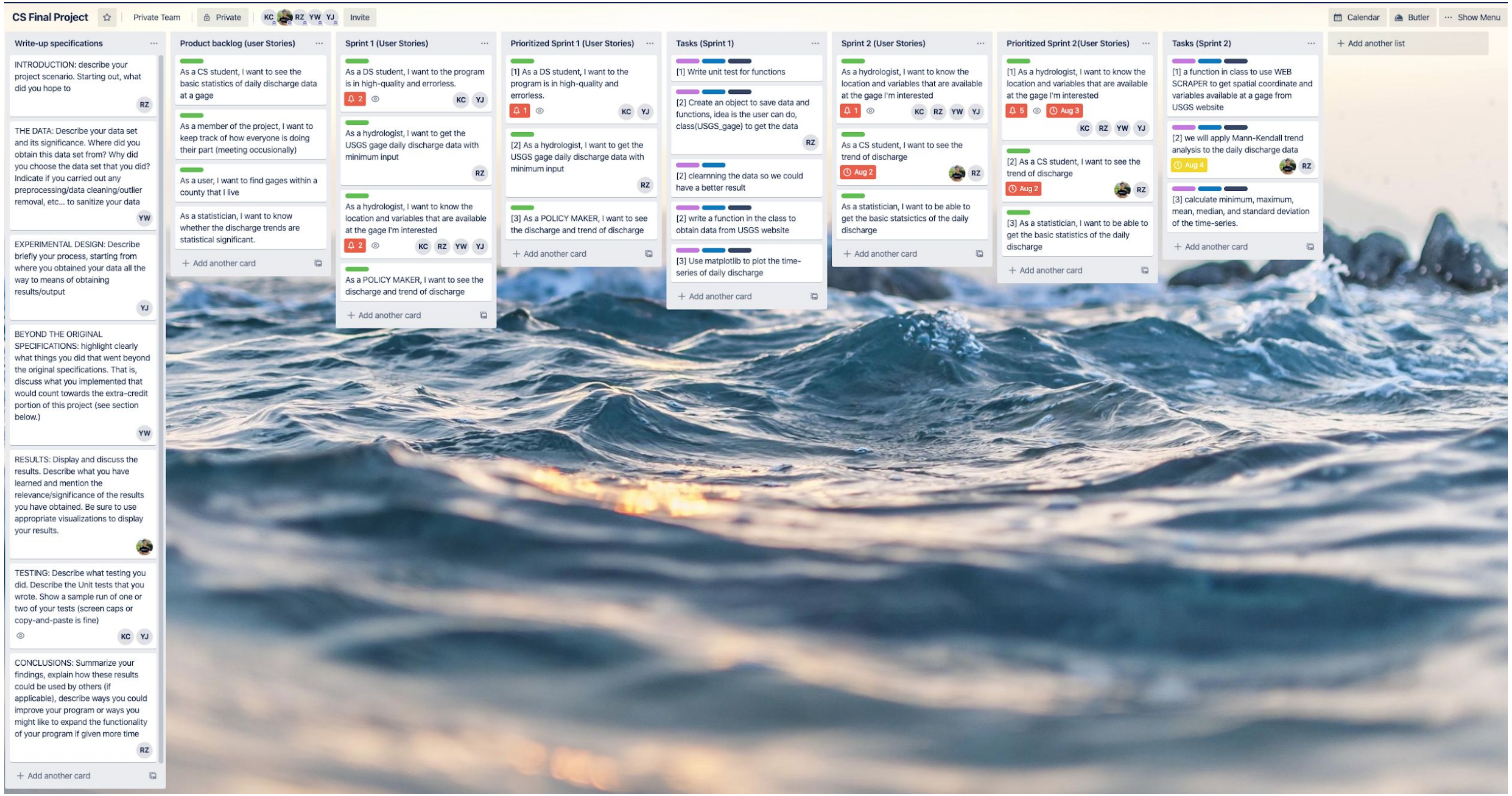
- Overall, 25.5 % of stations show a significant increasing trend and 6.7 % of stations show a significant decreasing trend
- NOVA and Richmond decreasing trend clustering
- Central and Northern PA increasing trend clustering
- Most gages show no trend



https://nbviewer.jupyter.org/github/ruoyu93/USGS_DataRetriever/blob/ruoyu93-patch-1/test.ipynb

SCRUM

How we accelerate the work process



Conclusion

1) We developed a Python program to download the USGS daily data automatically through web-scrape. The program also acquire metadata the gage.

2) Overall, 67.8% of stations do not show significant trend, however, 25.5% and 6.7% of stations show a significant increasing and decreasing trend, respectively.

3) The map shows that North Virginia and Richmond have clustering of significant decreasing trend

4) Future applications

- Downloading 15-min discharge data
- Supplying data for flash flood forecasting [Machine learning]
- Extending analysis to the entire United States