Jul 29th, 11:00 AM - 1:30 PM

Development of a Water Quality Index Calculation Tool using Excel

Rowan McCarthy
University of Northern Iowa

Mohammad Iqbal Ph.D.
University of Northern Iowa

Development of a Water Quality Index Calculation Tool using Excel

University of Northern Iowa
Earth & Environmental Sciences

Rowan McCarthy
Mentor: Dr. Mohammad Iqbal

Introduction

With many parameters by which water can be measured, it can be difficult to understand whether water is high quality or low quality. To solve this problem, many water quality indices have been developed over the last 60 years. The UNI Hydrology Lab uses the National Sanitation Foundation Water Quality Index (NSFWQI). This index provides a methodology by which Dissolved Oxygen, Fecal Coliform, pH, Biochemical Oxygen Demand, Temperature Change, Total Phosphate, Nitrate, Turbidity, and Total Solids can be combined to provide a single number indicative of the overall water quality. In previous years, when WQI calculations were needed, the bulk of the work was done either by hand, aligning rulers on printed NSF graphs or one at a time with online calculators. This process is extremely time and labor intensive. This project sought to make a high-volume calculator with which thousands of data could be calculated in a matter of minutes.

The development of this tool was staggered. Initially q-values for the 9 parameters were scraped from an online calculator and used as points on a curve. These curves proved to be a poor fit for this project, as the original NSF graphs are joined line-segments. Eventually, using the FORECAST function in excel, and treating the scraped NSF q-value data as arrays, linear interpolation was achieved, automating the manual method in a computer program. Applying this to all the arrays, and building in the rules unique to each graph, the calculator was ready to be used.

Since May 2011, the UNI Hydrology Lab has collected samples regularly from the Dry Run Creek and Cedar River. These data were taken from UNI Hydrology Lab website and WQI values were calculated for all 1341 samples published on the website. A process which would have taken nearly 100 hours (2 parameters per minute) to perform by hand was completed in less than 1 hour.

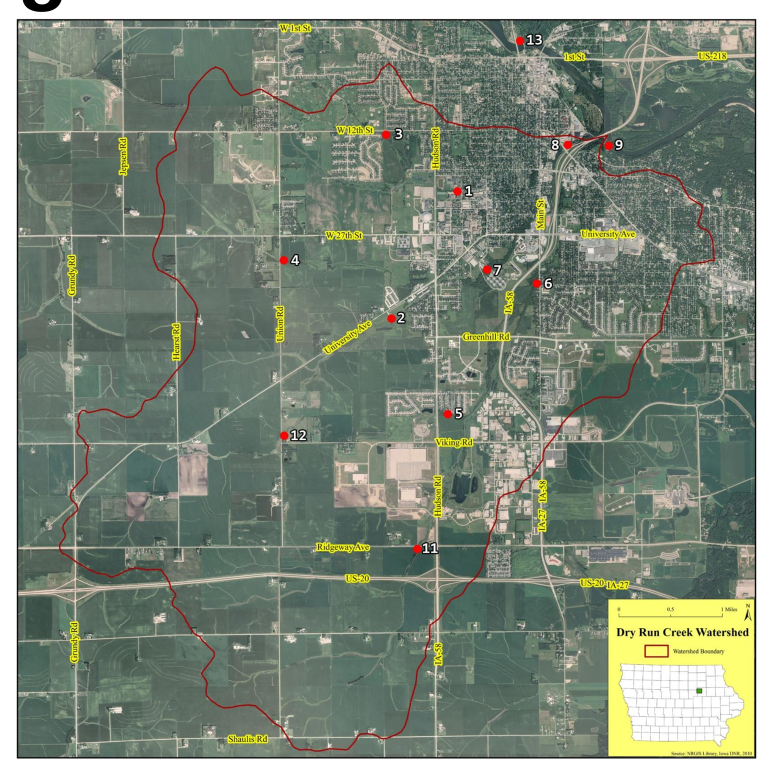
Date	Temp	DO	BOD	pH	TDS	TSS	Turb	Cond	NO3	Ecoli	PO4	DO%	Temp Change per Mile	Final WQI
	(°C)	(mg/L)	(mg/L)		(mg/L)	(mg/L)	(ntu)	(µs/cm)	(mg/L)	(col/ 100 mL)	(mg/L)	%	(°C/mile)	
												0.00	na	#N/A
												0.00	na	#N/A
												0.00	na	#N/A
												0.00	na	#N/A
												0.00	na	#N/A
												0.00	na	#N/A

This is the user interface for the project. Each parameter's raw data is entered. If it is unavailable, the spreadsheet will recognize many variations of N/A. DO% is automatically calculated as many tools lack the capability of direct DO% measurement. The final WQI is then calculated for each sample. As a spreadsheet, this interface can be copied downward until the spreadsheet limits, as long as it does so in tandem with the calculating cells on other sheets.

Background

Water Quality Index values are single number representations of overall water quality. This single WQI value is determined from by calculating Q-Values for each of the 9 parameters in the NSF guidelines. Each Q-Value can be found manually by examining a graph of Q-Values charted over the unit of measurement for a particular parameter (right). Each of these Q-Values is then multiplied by a Weight Factor, adjusting the effect of the parameter on the overall WQI based off how important the parameter is to water quality as shown in the chart below (e.g., Fecal Coliform is weighted twice as highly as turbidity). The products of these Q-Values and Weight Factors are then summed and reported as a WQI value. If one or more parameters are missing the final WQI is then divided by the sum of the Weight Factors of the parameters considered in calculation, approximating what the WQI would be if the missing parameter had been present and in line with the other parameters.

Parameters	Weight Factor
DO%	0.17
Fecal Coliform (E.	0.16
BOD	0.11
pH	0.11
Temp Change per	0.1
Nitrates (NO3)	0.1
Phosphate	0.1
Turb	0.08
TDS	0.07



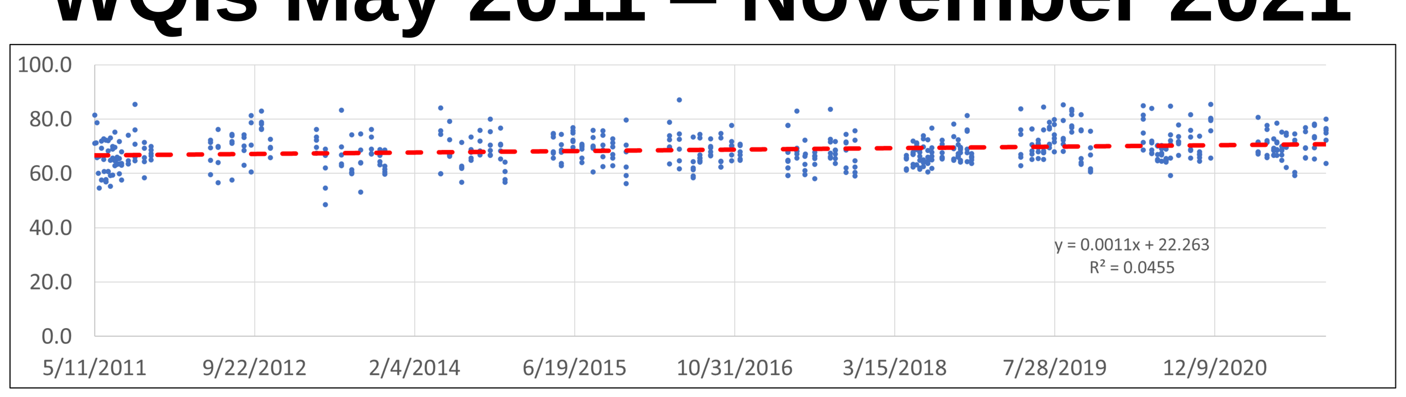
Since May 2011, the University of Northern Iowa has conducted routine water monitoring on the above sites, testing for many of the NSFWQI parameters. These data have not benefited comprehensive WQI computation in the last 11 years. This demonstrates a need for a bulk processing system.

Platform Choice

Many possible platforms exist upon which this system could be built. These are the reasons this project was built on Microsoft Excel:

- Microsoft Excel is ubiquitous, most professionals have it.
- Microsoft adheres to backwards compatibility; an update won't break this tool.
- Excel is readable, easily understood by familiar users
- Training resources are abundant
  - Almost all skills used in this project were learned in June

UNI Hydrology Collected Samples WQIs May 2011 – November 2021



Shown above is an example of the application of this tool. Last year, when calculating WQI values for samples takin in Nepal, my peak efficiency was approximately 2 parameters per minute. Assuming this efficiency, this project would have taken about 88 hours. Using this tool, the WQI's were generated and graphed in 38 minutes

Trials and Errors

My first attempt at making this spreadsheet was disastrous. Not only was every q-value off by several points, but the formulas themselves were a nightmare to read. Below is an example of these difficult to parse formulas.

```
=IF(G3="Not Available","Not Available",IF('DRC Site 1'!F3="nd",100,IF('DRC Site 1'!F3>30,2,
IF((-0.00001*'DRC Site 1'!F3^5)+(0.0014*'DRC Site 1'!F3^4)-(0.053*'DRC Site 1'!F3^3)+(1.0834*'DRC Site 1'!F3^2)-(13.825*'DRC Site 1'!F3)+103.82>100,100,
(-0.00001*'DRC Site 1'!F3^5)+(0.0014*'DRC Site 1'!F3^4)-(0.053*'DRC Site 1'!F3^3)+(1.0834*'DRC Site 1'!F3^2)-(13.825*'DRC Site 1'!F3)+103.82))))
```

Note that the above formula begins with a series of IF statements. This is essentially checking whether a cell has appropriate data in it, as well as apply rules specific to this parameter outlined in the NSF WQI methods. Below that is the problem. This first attempt used Excel to automatically produce trendlines that hit on several points along the NSF graphs in order to emulate the graphs. This was a mistake, as the NSF graphs are connected line segments, with sharp turns and non-exponential slopes. Also note, this "trendline" was exponential to the 5th power. This was inaccurate and extremely labor intensive to write out correctly.

```
=IF(G3="Not Available","Not Available",IF('Calcator Interface'!D3="nd",100,IF('Calcator Interface'!D3>30,2,
(FORECAST('Calcator Interface'!D3,
INDEX('WQI Datasets'!$B$2:$B$18,MATCH('Calcator Interface'!D3, 'WQI Datasets'!$A$2:$A$18,1)):INDEX('WQI Datasets'!$B$2:$B$18,MATCH('Calcator Interface'!D3, 'WQI Datasets'!$A$2:$A$18,1)+1),
INDEX('WQI Datasets'!$A$2:$A$18,MATCH('Calcator Interface'!D3, 'WQI Datasets'!$A$2:$A$18,1)):INDEX('WQI Datasets'!$A$2:$A$18,MATCH('Calcator Interface'!D3, 'WQI Datasets'!$A$2:$A$18,1)+1))))))
```

Above is the same cell from my second attempt. Note the identical first line. The second line sees the divergence of the two. Using the FORECAST statement in Excel, I was able to simulate linear interpolation between two values. To select the values, the next two lines MATCH the value entered for the parameter in the Interface tab to the array appropriate for the parameter. Using INDEX statements, this cell determines what the values in the array above and below the Interface value are. It then calculates a straight line between the two values and FORECASTS where the Interface value would land on that line. This comes out as a correct application of the manual method, without the same risk of human error.

TempChange	TempChangeQI	TDS	TSQI	Phosphate	PhosphateQI	Turbidity	TurbidityQI
-10	55	0	79	0	100	0	99
0	93	10	82	0.2	92	3	90
5	73	20	84	0.5	60	8	80
9.5	47	30	84.5	0.7	50	13	70
12	36	40	86	1	40	15	67
14.5	32	60	87	1.3	34	20	61
21	20	70	86	1.6	30	30	53
30.00001	10	150	79	2	27	40	45
		450	40	3.2	20	50	39
		500.00001	31	4	17	60	33
				5	13	70	29
				6	10	80	25
				7	8	90	22
				8	7	100.00001	17
				10.00001	7		

Above are four of the aforementioned arrays. Each column of parameter values is aligned with corresponding Q-Values.
This allows the function in the above calculations to fill in the below "working cells", which are available to view by the user to assist in education and troubleshooting. Note that the last number in each parameter column has seemingly extraneous zeros. This is because if an entered value is identical to the last value in the array the linear interpolation formula cannot function, as there is no array value higher than that. As such, I raised the values by 0.00001 as it is just above the limit enforced by the first line in the interpolation formulas and beyond the specificity of our measuring tools.

Date	Temp Change per Mile (Be sure Units are Correct)	DO%	BOD	pH	TDS	Turb	NO3	Ecoli	Phosphate
	Q-Value	Q-Value	Q-Value	Q-Value	Q-Value	Q-Value	Q-Value	Q-Value	Q-Value
11/22/2021	Not	2	100	0	79	99	97	#N/A	100

Temp Change per Mile Weighting	DO% Weighting	BOD Weighting	pH Weighting	TDS Weighting	Turb Weighting	NO3 Weighting	Ecoli Weighting	Phosphate Weighting	Total Weight	Final WQI
0.1	0.17	0.11	0.11	0.07	0.08	0.1	0.16	0.1	1	
0	0.17	0.11	0.11	0.07	0.08	0.1	#N/A	0.1	#N/A	#N/A

The left section are the cells which contain the interpolation formula from above. The right section determines the final WQI for each sample and it's relative weight out of 1. This line of cells can be copied to the limits of the spreadsheet, as long as the Interface is also copied to the same row.

Challenges and Future Research

Though Excel is an accessible platform, it is not the *most* accessible platform, as it is inherently costly to non-university affiliated individuals. The future of this project may see this work translated into a freely available platform such as Google Drive, or a standalone application.

Additionally, in preparation for the inevitable day on which something in this tool breaks, a user guide may be developed to troubleshoot and fix future problems.

Dr. Iqbal has noted the educational utility of a tool like this. As such, video lessons to describe how it works and what the results mean may be generated to accompany the excel file.

As part of the write-up for this research, a summarized history of water-quality initiatives taken in the Cedar Falls area will be used to demonstrate possible reasons *why* the WQI values may be improving.

Acknowledgements

AJ Fleming, RWTH
Ben Wonderlin, Yale
Black Hawk County Grant
Brian Oram, KnowYourH2O
Daniel Laske, UW Madison
Department of Earth & Environmental Sciences, UNI
Hydrology Lab, UNI
Joshua Balk, Black Hawk County
Tyler Larson, UNI
Gautam Kambhampati, Imperial College London: Dept. of Physics
Steve Smith, UNI

Digital Poster, Related Links