7-12-2018

Huron-to-Erie Water Quality Data Platform

Lanyu Xu
Wayne State University, xu.lanyu@wayne.edu

Carol J. Miller
Wayne State University, cmiller@eng.wayne.edu

James Olson
Wayne State University, jamesols@mtu.edu

Michael Beaulac
Michigan Department of Natural Resources, beaulacm@michigan.gov

Huron-to-Erie Water Quality Data Platform

Lanyu Xu

Wayne State University

Email: xu.lanyu@wayne.edu

Carol J. Miller

Wayne State University

Email: cmiller@eng.wayne.edu

James Olson

Wayne State University

Email: jameso@mtu.edu

Michael Beaulac

Office of the Great Lakes, Michigan Department of Natural Resources

Email: beaulacm@michigan.gov

Abstract: To address the challenges of environmental degradation, creation of a sustainable urban environment, and increased public engagement and awareness, a mass-oriented, user-friendly and cloud-based data platform has been developed and deployed to provide integrative water quality data in one of the most critical urban corridors of the Laurentian Great Lakes system. In this paper, we describe the data platform developed for the watershed and connecting channels between Lake Huron and Lake Erie, including the St. Clair River, Lake St. Chair, and the Detroit River. This data platform greatly facilitates the access of data across data providers and agencies. Several example applications are provided of platform use for temporal and spatial characterization of intake water source quality and urban beach health through consideration of Escherichia coli, Dissolved Oxygen, pH, and blue-green algae detections along the Huron-to-Erie corridor. Although data collection for each of these parameters was designed for unique purposes and supported through varied agencies, this paper shows the collective advantages of applying the data beyond the original scope of collection.

Keywords: Data, Water Quality Monitoring, Lake Huron, Lake Erie, Detroit River, Blue-Green Algae, Laurentian Great Lakes

1 Introduction

In all regions of the USA, and in most international sites, multiple governmental and political jurisdictions are responsible for the evaluation of water quality for maintenance of public health and other purposes. The present project takes these data sets that are typically used in isolation for the specific regulatory application (eg., beach health or drinking water) and provides a sharing mechanism to allow the data obtained for single-purpose use to be combined into a more holistic data set that can, when integrated, address more complex issues than the original data base. Therefore, the novelty of the present approach lies in the multiple data types and sources used in the development of the data platform and the subsequent analysis.

Each type of data is organized into a particular project folder. Each project is associated with a particular environmental purpose. The two projects considered in this paper relate to water quality for drinking water and recreational swimming. Water quality sampling on a regular basis is mandated by the regional government for each of these functions. For drinking water, the state (Michigan Department of Environmental Quality) and federal (US Environmental Protection Agency) government have primary authority, while beach health is under local (County Health Department) jurisdiction, with federally mandated sampling protocols.

This cross-application use of data is important because the collection of water quality data is relatively expensive, and as budgets become more constrained for environmental efforts, use of existing, long-term, historical monitoring from multiple sources can provide an economical approach to provide information to assess questions of cause and effect and provide input for empirical models for the sampled geography. The approach has global/international relevance as most countries require water quality monitoring for multiple water use points. In many cases the water resource serves more than one function (swimming/bathing, consumption, fishing, etc.), and there are multiple data sets for the single water resource. Of course, this data is non-harmonious; that is, the timing of the samples are not synchronized, the parameters sampled are not consistent across projects, the location of the sample measurement is not consistent, and the sample resolution may be variable across platforms.

The environmental problems that can be addressed with these data sets are significant and varied. In the location of the present study, there is significant economic value associated with the water resource as a tourist destination. Beach closure due to high Escherichia Coli (e coli) levels can have a major impact on the local economy. Several of the beaches in the study area experience multiple shut-downs during the summer months and the closures have become, in addition to a health issue, an important economic issue. As such, there is great pressure to uncover greater insight into the source of the high e coli levels. One simple approach to gaining insight into the important variables can be found by augmenting the e coli data collected as part of the beach monitoring program with other water quality measures collected as part of the other projects. While the beach sampling requires e coli measurements, the drinking water monitoring program requires additional details and a greater frequency of sampling. In the present application, we augment e coli with dissolved oxygen (DO), pH, and blue-green algae (BGA) measurements. Weather data from the region provides the important rainfall measurements.

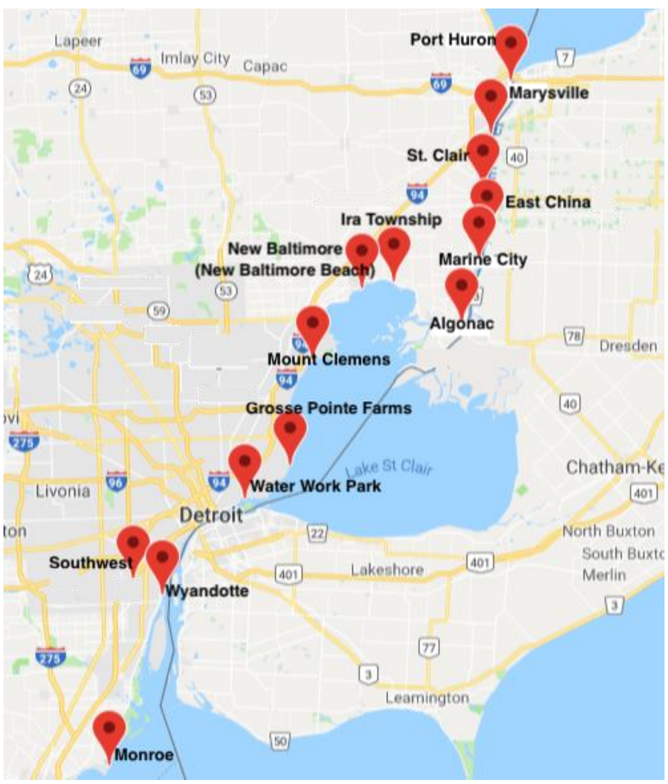


Fig. 1 Huron-to-Erie corridor.

It is important to provide background regarding the specific geographic region of the present study/application. The 80-mile Lake Huron to Lake Erie international corridor is a major global shipping route and its shores include both heavy manufacturing along downriver Detroit, Michigan USA and a concentrated network of petrochemical plants just south of Sarnia, Ontario CA. There are 14 U.S. Water Treatment Plants (WTP) (as shown in figure 1) along this international corridor that treat water and subsequently distribute clean drinking water through an extensive network that serves a population of about 4 million in southeast Michigan. These WTPs are owned and operated by 12 different local communities and the Great Lakes Water Authority.

The primary responsibility of a drinking water service provider is to protect public health. In order to protect public health, WTPs need to be aware of potential source water risks/threats to the drinking water plants such as accidental spills and emergency diversions into source water areas combined with nutrient-triggered algal blooms. These risks demonstrate historical occurrences and it is understood that they are impossible to completely eliminate. Despite this, the state and federal drinking water regulatory framework does not include specific requirements for source water monitoring.

The Huron-to-Erie Real-time Drinking Water Protection Network was created in the mid-2000s to counter those threats. The 14 communities and the Detroit Water and Sewer District (DWSD) agreed that in order to protect public health, they each needed to monitor (i.e. see in real-time) the quality of the source water entering the drinking water treatment plants in order to counteract any threats contained in the source water. The system network included many elements including a variety of monitors, data logging and website access. Unfortunately, a number of challenges, including the economic recession (dwindling budgets), complicated equipment and high maintenance costs, and limited staff resources (priorities) effectively reduced network participation and limited its effectiveness.

Federal and state funds were initially made available to acquire and maintain the monitoring equipment and create the network and databases used by the WTP operators and public. All WTPs were provided with YSI multi-parameter sondes (bundled probes) while about half received total organic carbon (TOC) analyzers, fluorometer and gas chromatograph/mass spectrometer (GC/MS) units.

The sondes were the least expensive equipment to purchase and the simplest to maintain and use. They included sensors for general water quality parameters such as DO, conductivity, pH, temperature, turbidity, oxidation reduction potential, BGA and chlorophyll. The TOC analyzers, used to detect total organic carbon, the fluorometers, used to identify hydrocarbons indicative of petroleum spills, and the GC/MS units for measuring volatile organic chemicals (VOCs), all required far greater expertise to use and were both more time-consuming and expensive to maintain than expected. Regular equipment maintenance was either provided by a competent engineering firm or by internal WTP staff with expertise, depending on available resources. The global financial meltdown significantly stressed those resources and added to maintenance and training inconsistencies, and eventually a drop-off in network participation. Accordingly, only about six of the 14 WTPs participating in the network still maintain functional monitoring equipment, with the majority being the multi-parameter sondes. The Healthy Urban Waters (HUW) program at Wayne State University (WSU) was asked to host and improve the Huron-to-Erie drinking water data platform which had been established through federal funding with lofty goals in 2007. All the data from the vintage Drinking Water Monitoring Network and other supporting monitoring projects have now been migrated by HUW to a public-facing and user-friendly water quality data platform and database for long-term research and public communication.

This new and improved data platform is the subject of the present paper. As a data warehouse, this platform provides more than 28 million items describing source water quality and additional environmental projects in Southeast Michigan. With this platform, extensive statistical and research analyses can be completed to assess the environmental condition and trends within the Huron-to-Erie corridor. In particular, this paper provides several example studies using data provided by the platform, through consideration of e coli, DO, and pH observations along the corridor and connected urban beaches. The examples provided here provide only an introduction to the many possible applications.

The closures of recreational beaches because of the surge of fecal pollution in water occurs occasionally throughout the country. e coli, as a fecal indicator organism (FIO) is monitored to ensure water quality meets health standards and associated legislation (Oliver, Porter et al. 2015). Elevated e coli identified in these water resources comes from a variety of sources, including sewage overflows, agricultural runoff, and urban storm water (McLellan 2004). Storm water and combined sewer overflows are the most significant polluting sources causing beach closures or advisories across the city and region (Dorfman, Stoner et al. 2007). The e coli in storm water derived from human and animal sources may run into nearby water bodies, resulting in increased bacterial loads and threating the health of swimmers upon contamination of recreational beaches.

The near-shore area is an extremely variable system. DO and pH are greatly affected by fluctuations of water temperature, salinity, carbon dioxide (CO_2) exchange and mix, and biogeochemical process (Frieder, Nam et al. 2012). The increased level of temperature and salinity reduces the surface oxygen concentrations, while the uptake of atmospheric CO_2 will result in the increase of water acidity (Frieder, Nam et al. 2012), and concomitant decrease in water pH (Doney, Fabry et al. 2009). DO and pH are related to each other since deoxygenation and acidification are tightly linked through mutual transformation between inorganic carbon and organic matter (Cai, Hu et al. 2011, Hofmann, Peltzer et al. 2011). To better understand this highly variable system, we used the data monitoring platform to conducted an analysis of the relationship between DO and pH along the Huron-to-Erie corridor.

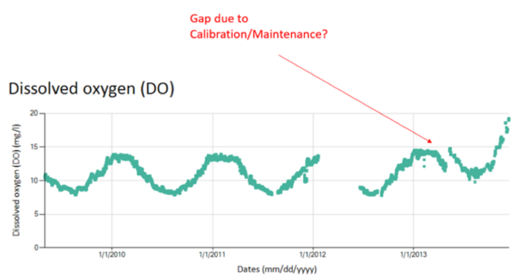
As a basal resource in aquatic ecosystems, algae concentrations naturally fluctuate seasonally. However, cyanobacteria, commonly known as BGA, can proliferate into harmful algal blooms caused by elevated nutrient levels. BGA produce cyanotoxins, which can kill fish, negatively impact tourism, and threaten drinking water and safe body contact water recreation. Over the last decade, algal blooms have become a pervasive problem in the western basin of Lake Erie (Smith, King et al. 2015). According to NOAA, algal blooms develop in summer months and peak in September due to nutrient runoff from agriculture and urban areas, sunlight and warm water temperature (https://www.nasa.gov/image-feature/algae-bloom-in-lake-st-clair, Jan 21, 2018). In 2013 and 2014, Lake Erie suffered from large harmful algal blooms that cut off drinking water in some municipalities in Ohio (https://coastalscience.noaa.gov/news/habs/large-summer-harmful-algal-bloom-predicted-lake-erie/, Jan 21, 2018). Using data from the Drinking Water Monitoring Network we will investigate the BGA variation and its related parameters, including its different manifestations between Lake Huron and Lake Erie.

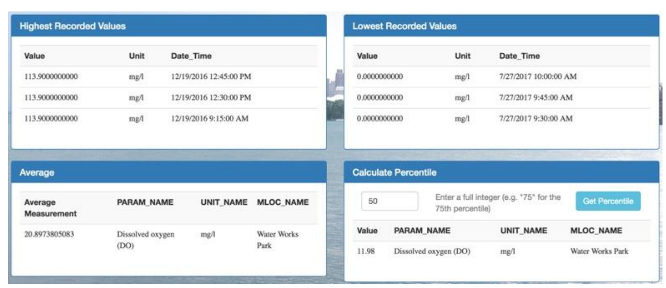
The remainder of the paper is organized as follows. Section 2 and 3 describes the Drinking Water Monitoring Network, as well as the collected data and statistical methods that are used for analysis. Section 4 provides the example application results and various interpretations of those results. Section 5 concludes this paper and considers the future development of this platform.

2 Methods

2.1 Cloud-based Platform

Motivated by the fact that many organizations collect a wealth of environmental data in this region, we were driven to establish a home-grown cloud-based platform which could store data from all sources and support data sharing and collaboration. With a REST API, the platform collects data from providers automatically and greatly reduces the tedious work of collecting data using other methods, allowing greater time for scientific research. Moreover, the platform provides online statistical tools to allow researchers to gain a general understanding of the data distribution before the export of data for further analysis. For instance, users are able to plot time-series data for a specific parameter and potentially locate a time segment of interest for further analysis. Also, irregular data (figure 2(a)) can be identified for later consideration. Percentiles and averages can be explored in the browser allowing users to get value ranges for parameters (figure 2(b)). The platform also provides an interactive map to assist the user in the selection of monitoring data locations for analysis.





(a) Plot graph

(b) get statistic

Fig. 2 Online statistical tools on the monitoring platform

The architecture of the platform is shown in figure 3. Millions of water quality sample measurements are collected by dispersed monitoring equipment and then dispatched to different environmental organizations and departments of local governments and in some cases, non-governmental organizations (NGOs). The source water monitoring data is provided in real-time to municipalities and water treatment plant operators through a secure portal, WQDataLive. This is the system used for continuous source water quality, spill reporting and other threats to drinking water. Our platform lies between WQDataLive and the community, providing the only public-facing view of the measurements taken by the Drinking Water Monitoring Network. Data users can easily access the monitoring data and meta-data by straightforward queries on the platform.

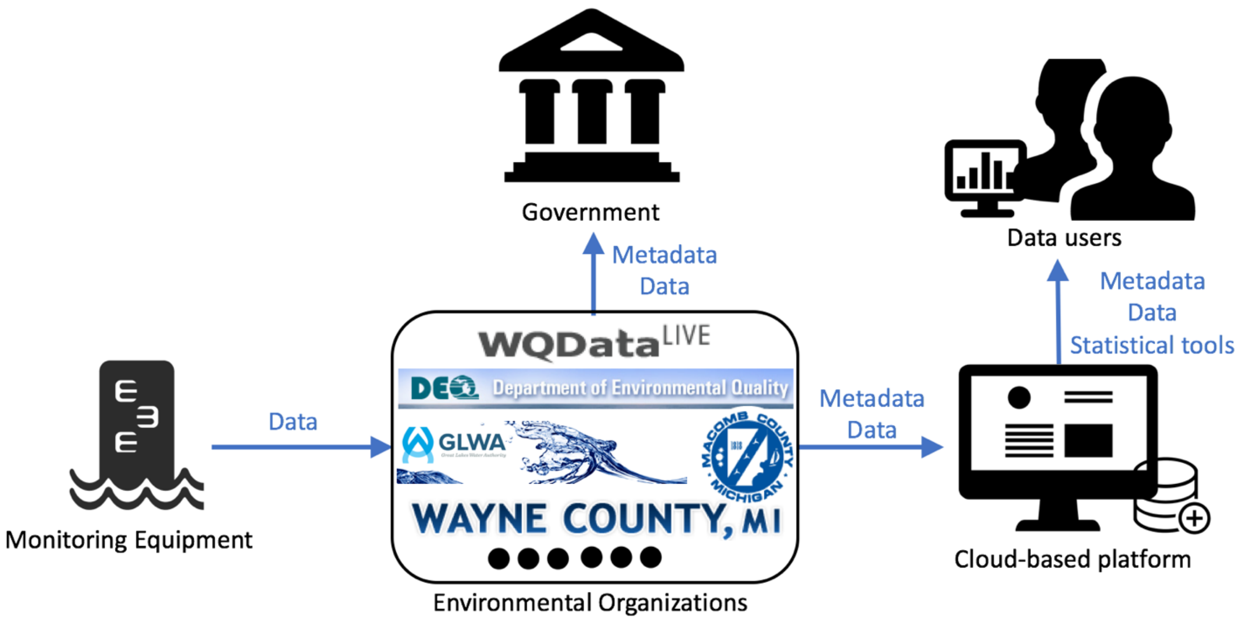


Fig. 3 The architecture of the water quality platform.

Specifically, this platform collects data from the following agencies to provide an integrative data warehouse for the Southeast Michigan environment:

- *WQData Live* (https://wqdatalive.com/, Jan 21, 2018) is an on-line environmental data viewer that provide real-time data to authorized visitors. It provides all drinking water monitoring data from the Huron-to-Erie corridor with updates every 10 minutes.
- *Macomb County Health Department* (http://health.macombgov.org/Health-DataStatistics, Jan 21, 2018) is the county agency which updates environmental data in the form of annual reports. It manages a wide range of data, including beach monitoring, lake assessment, surface water tests and similar across Macomb County.

- *BeachGuard System* (http://www.deq.state.mi.us/beach/Default.aspx, Jan 21, 2018) is provided by Michigan Department of Environmental Quality as a public resource. It archives water quality sampling measurements of Michigan beaches, with updates every day from April to September, and is a primary communication route for beach closures and advisories.
- Great Lakes Water Authority (GLWA) (http://www.glwater.org/, Jan 21, 2018) is a regional water and sewer authority that provides both raw water and tap water sampling results over the past 16 years, with updates each month.

As a data warehouse storing more than 27 million water quality measurements from around 750 monitoring locations and 13 projects throughout Southeast Michigan, a comprehensive yet flexible database design was essential to ensure efficiencies of the platform. The entity relationship diagram for the database is shown in figure 4. Key features in the diagram include:

- One-to-one mapping of monitoring location and project: Exact geographical position is specified by latitude and longitude. Project and location are tightly linked here to identify sampling results measured in same location belonging to separate projects.
- Standardized parameter names: Different agencies may use different terminology to represent the same parameter. Before migrating to the database, each parameter is denoted by the appropriate name in the Substance Registry Service (SRS) (https://ofmpub.epa.gov/sor_internet/registry/substreg/home/overview/home.do, Jan 21, 2018).
- Unified monitoring results: Measurement results supplied by data providers are reformatted to predefined format in the database. A Python script is used to produce the required format from the initial one, and ensures the integrity of the original information.
- Full data description: The platform has metadata for all the critical details associated with the data measurement that can impact the interpretation of the value, including the source water-body, geographic location, measuring season, hydrologic unit, and others.

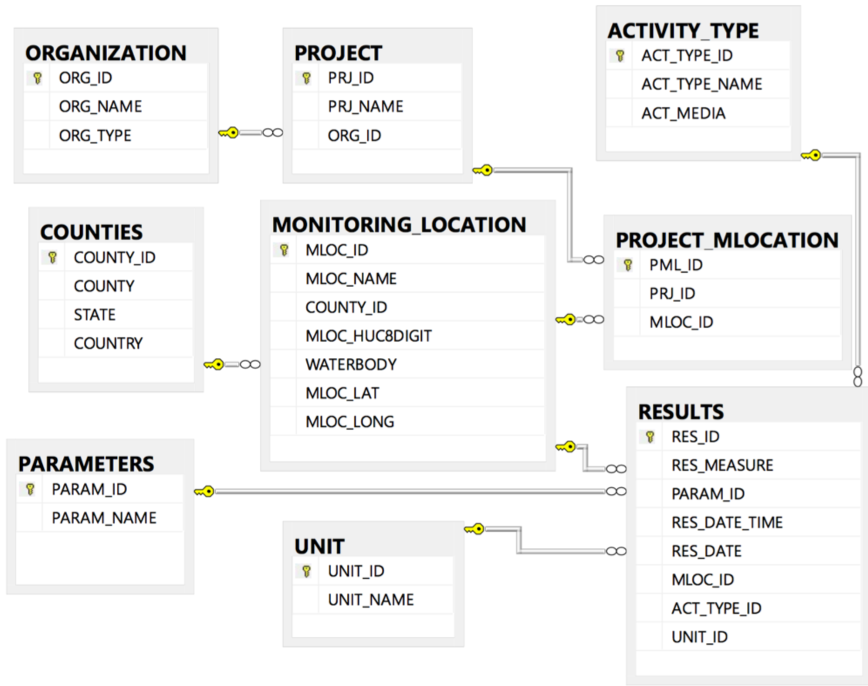


Fig. 4 Entity relationship diagram of conceptual data model for the water quality platform.

2.2 Data Source

Example applications of the data platform are provided to demonstrate ways in which the data from this platform may be leveraged to better understand the behavior of water resources within this geographic region and its associated watershed. Several parameters are addressed in these example applications, including e coli, DO, pH and BGA. A summary of the sample data used in these applications is provided in table 1. All of the data were retrieved through our cloud-based platform, except for the precipitation data which was obtained directly from NOAA. All sample sites are located within or adjacent to the Huron-to-Erie corridor, with different sites providing different types of data. The DO and pH characterization is provided by measurements from an array of 14 source water intake points along the U.S. side of the corridor (figure 1) which have been transmitting data to the platform from 2007 to present. For the beach e coli concentration values, data from recreational beaches in the corridor vicinity are available from 2004 to 2017 (marked in figure 5). The precipitation data is from a single gage site located, on average, 3.6 miles from monitored recreational beaches. BGA data is available from three of the drinking water intake points: Marysville, Monroe and Toledo. Marysville and Monroe are at the upstream and downstream ends of the corridor, respectively, while Toledo is along the northwestern shore of Lake Erie.

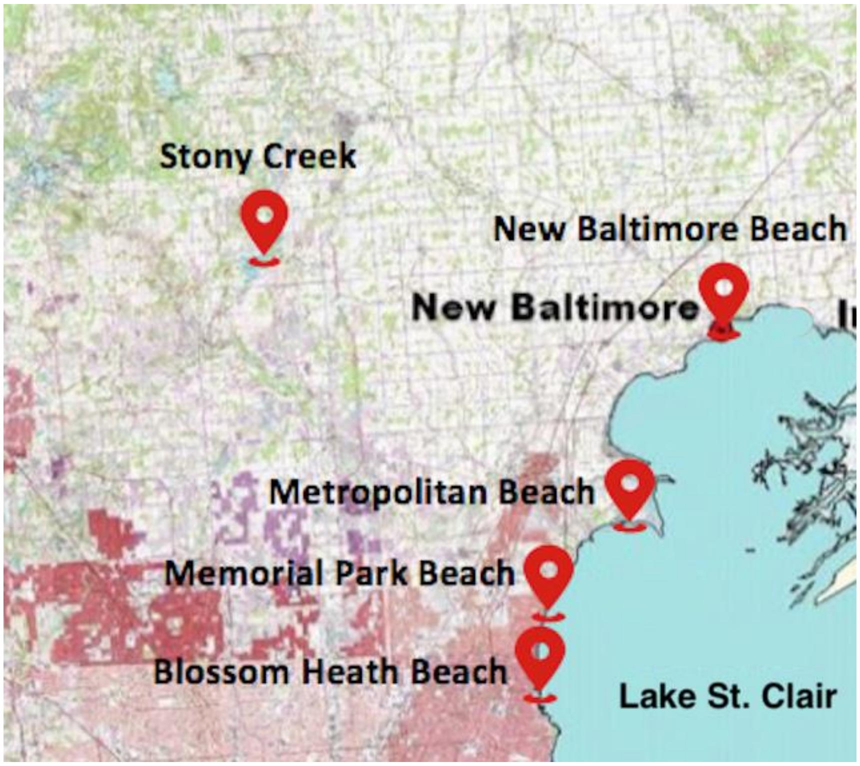


Fig. 5 Geographical Distribution of Study Beaches.

Table 1 Summary of sample data.

Data	Recorded frequency	Monitoring location	Source
BGA	Hourly	Marysville, Monroe	Water quality data platform (http://waterdatadetroit.azurewebsites.net/)
		Toledo	Great lakes observing system (GLOS) (http://habs.glos.us)
DO	Hourly	Port Huron, Marysville, St. Clair, East China, Marine City, Algonac, Ira Township, New Baltimore, Mt. Clemens, Grosse Pointe Farms, Water Works Park, Wyandotte, Southwest, Monroe	Water quality data platform
pH	Hourly	Port Huron, Marysville, St. Clair, East China, Marine City, Algonac, Ira Township, New Baltimore, Mt. Clemens, Grosse Pointe Farms, Water Works Park, Wyandotte, Southwest, Monroe	Water quality data platform

Data	Recorded frequency	Monitoring location	Source
E coli	Daily in wet season	New Baltimore Beach, Metropolitan Beach, Memorial Park Beach, Blossom Heath Beach, Stony Creek	Water quality data platform
precipitation	Hourly	Washington, St. Clair Shores, New Baltimore	National oceanic and atmospheric administration (NOAA) (https://www.ncdc.noaa.gov/)

The e coli measurement used standard county Health Department methods - a membrane filter procedure which relies on colony counts (USEPA 2000). Other parameters such as DO, pH, Turbidity, and BGA are collected by YSI multi-parameter sondes at the drinking water treatment plants and are reflective of the water quality at the intake point.

3 Theory

3.1 Pearson Correlation Coefficient

In this paper, we use Pearson Correlation Coefficient (Benesty, Chen et al. 2009) to demonstrate the relationship between two parameters.

An absolute value of Pearson correlation coefficient greater than 0.8 suggests that the two parameters are highly correlated; when the value is between 0.5 and 0.8, the two parameters have significant correlation; when the value is between 0.3 and 0.5, the two parameters exhibit low correlation; and when the value is lower than 0.3, the two parameters are not linearly correlated. In section~\ref{result}, the Pearson correlation coefficient is applied to represent the correlation between two parameters, such as e coli and antecedent precipitation, DO and pH. For these applications, correlation coefficient values above 0.5 are considered to reflect significant correlation.

In addition, we use the best fit line through a scatter plot of two parameters to provide a measure of the linearity of the relationship. The applicability of the straight line relationship is quantified through the Pearson correlation coefficient.

3.2 Seasonal Influences

The analysis also considered seasonal influences on parameter relationships. During the spring and summer (May to October), water runoff and quality is greatly influenced by land cover, while in the fall and winter seasons (November to April), land cover has less effect because of limited runoff (Pratt and Chang 2012). While the other parameters are available year round, the e coli data is only available for the spring and summer seasons. It is because e coli data is collected as part of the regional beach health monitoring program. This program does not continue through the other two seasons, as humans are not using the water for recreational purpose during those reasons.

4 Results and Discussion

4.1 Correlation between E coli and Antecedent Precipitation Amount

As an application of our database for better understanding of the e coli variation in several swimming beaches, the effects of antecedent rainfall on e coli is investigated. The observation sites for e coli are clustered around Lake St. Clair, and Stony Creek Lake which is geographically 47 miles from Lake St. Clair (figure 5). e coli concentration are significantly higher in the beaches adjacent to Lake St. Clair as compared to the Stony Creek Lake measurements. e coli concentration in beaches adjacent to Lake St. Clair show a pattern of increasing concentration with decrease in latitude(figure 6).

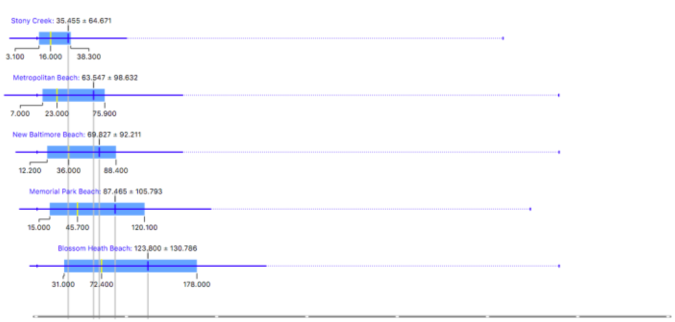


Fig.6 Quartile graph of e coli in the studying area.

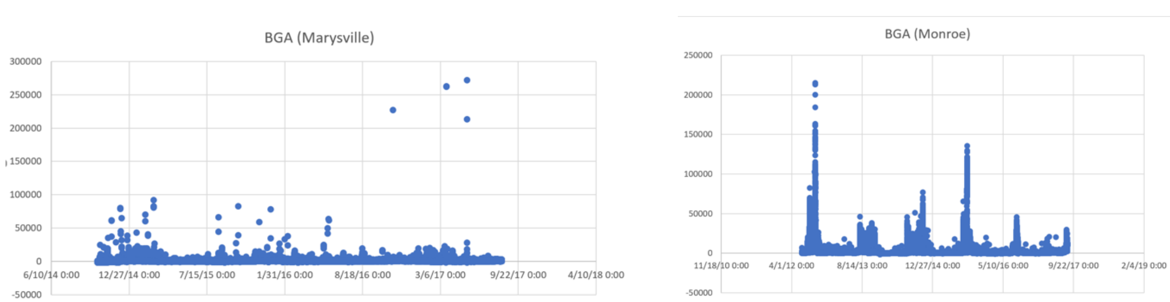
To understand the relationship between e coli and antecedent precipitation, table 2 compares the correlation results of e coli concentration with one previous day's rainfall, as well as the summation of precipitation from the two days antecedent to the e coli measurement. Although the number of observations vary from site to site, a significant positive correlation is observed between e coli and the one-day antecedent precipitation.

Table 2 Correlation between e coli and antecedent rainfall.

	New Baltimore Beach		Metropolitan Beach		Memorial Park Beach		Blossom Heath Beach	
	Corr.	No.	Corr.	No.	Corr.	No.	Corr.	No.
1 next day	0.49	48	0.50	29	0.64	21	0.67	8
2 next day	0.40	88	0.35	25	0.21	25	0.40	10

4.2 BGA Related Analytics

Seasonal variation of BGA measurements were compared between the upstream and downstream reaches of the Huron-to-Erie corridor. The measured values for Marysville, located near Lake Huron at the upstream end of the St. Clair River, do not exhibit any obvious seasonal pattern of BGA (figure 7(a)). However, at the downstream end of the corridor, in Monroe, the measurements follow a seasonal pattern (figure 7(b)) with highest values in late summer. These findings are consistent with the higher frequency of algal blooms in Lake Erie each summer. Additional analytics are performed using two Ohio sites close to this focus area of algal blooms in western Lake Erie, Toledo and Oregon. Data for these latter two sites are maintained at the GLOS platform (http://habs.glos.us/stations/?id=tolcrib&variable=ysi_blue_green_algae, Jan 21, 2018).



(a) Marysville

(b) Monroe

Fig. 7 BGA distribution on different locations.

The BGA readings at all four locations were placed in consistent units, with the resulting time series graphs for BGA shown in figure 8. The three sampling locations in western Lake Erie (namely Monroe, Toledo, and Oregon) exhibit peak values of BGA in late summer - and in all cases the summer readings at those three sites exceed the Marysville values. The Marysville values for the period Winter 2015 to late Spring 2016 are unreasonably high and spurious. There are suspect and this analysis suggests the data should be used with caution.

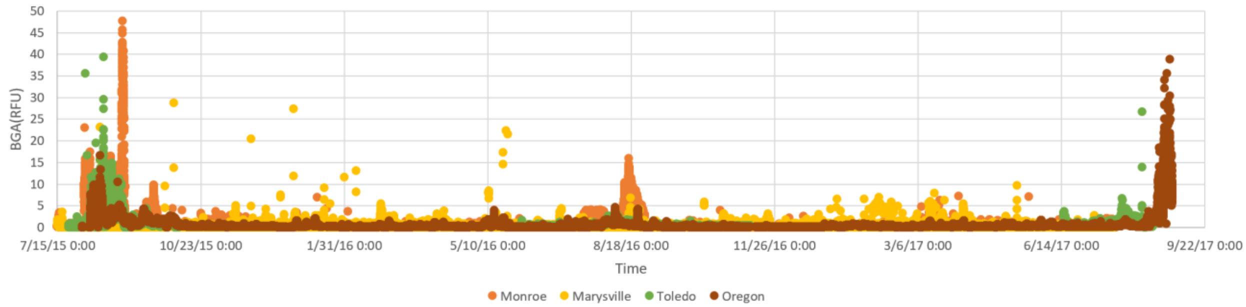
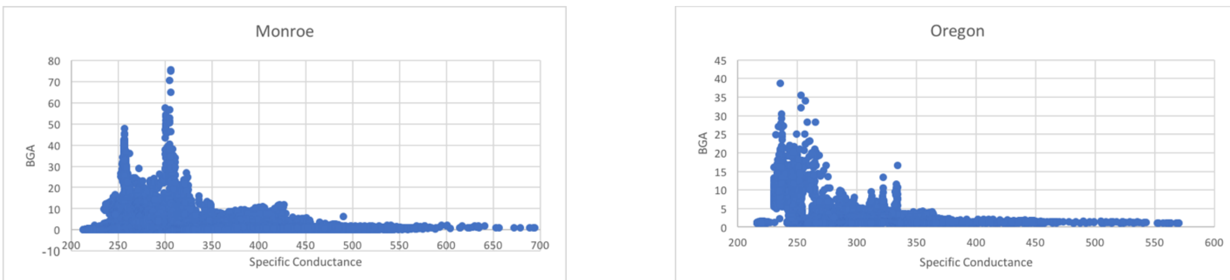


Fig. 8 BGA time series graph.

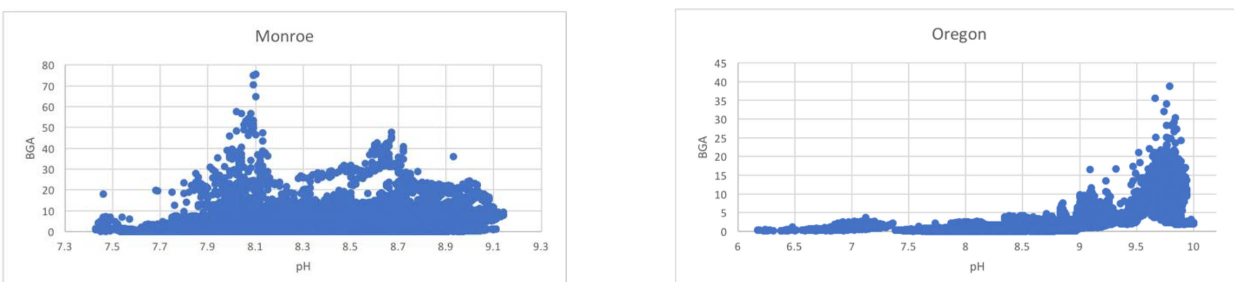
The relationship between BGA and Specific Conductance, pH, and Turbidity was evaluated as part of this investigation (figure 9). Although each of these parameters has seasonal patterns, none of the evaluated relationships are statistically significant (correlation result less than 0.5 in all cases), except BGA and pH relationship in Oregon. However, there are yet some interesting points to pursue in the future. For example, the data suggests that Specific Conductance values between 200 and 350 are most conducive to large BGA measurements. In terms of pH, the Monroe measurements suggests that pH values around 8 are associated with growth periods for BGA, while both the Monroe and Toledo measurements indicate that increasing values of pH are associated with higher BGA values. In addition, at both locations, significant presence of BGA is associated with turbidity values less than 200.

4.3 Correlation between DO and pH

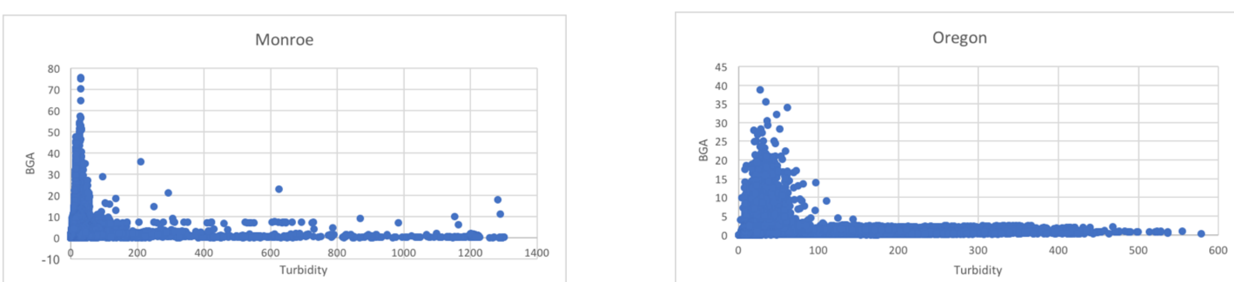
Multiple studies have considered the relationship between DO and pH in the past. In northern Australian mangrove waterways, pH and DO measurements exhibit consistently high correlation ($r^2 \geq 0.8$) (Boto and Bunt 1981). Frieder *et al.* studied DO and pH in a nearshore kelp forest in California, and found they are strongly positively correlated (Frieder, Nam et al. 2012). In the present analysis, the DO and pH data are considered for the 2015-16 time period. Five of the 14 Huron-to-Erie WTPs are chosen for this analysis as these five sites (Marysville, Marine City, Algonac, Water Works Park, and Monroe) have a full data set for this period. Figure 10 displays the DO measurements, with low values in the summer to fall seasons, and high values in the winter to spring season. Figure 11 provides the pH measurement, with a pattern generally opposite to that exhibited by the DO measurements, although much less defined. This counter-pattern shown by the DO-pH relationship does not exhibit high correlation; as an example, note the data from New Baltimore in figure 12. Upon further evaluation, a month by month correlation is uncovered. Figure 13 shows the correlation of pH and DO in May, June, July, and August for New Baltimore Beach, with correlation coefficient values of 0.95, 0.87, 0.93, and 0.96 respectively. This observation is consistent with the works conducted by (Boto and Bunt 1981, Frieder, Nam et al. 2012), in which the analyses were conducted for discrete time periods and focused on the correlation as affected by environmental factors.



(a) BGA and Specific Conductance(Corr. are 0.11, and 0.26 respectively)



(b) BGA and pH(Corr. are 0.22, and 0.62 respectively)



(c) BGA and Turbidity(Corr. are 0.04, and 0.07 respectively)

Fig. 9 BGA related analysis.

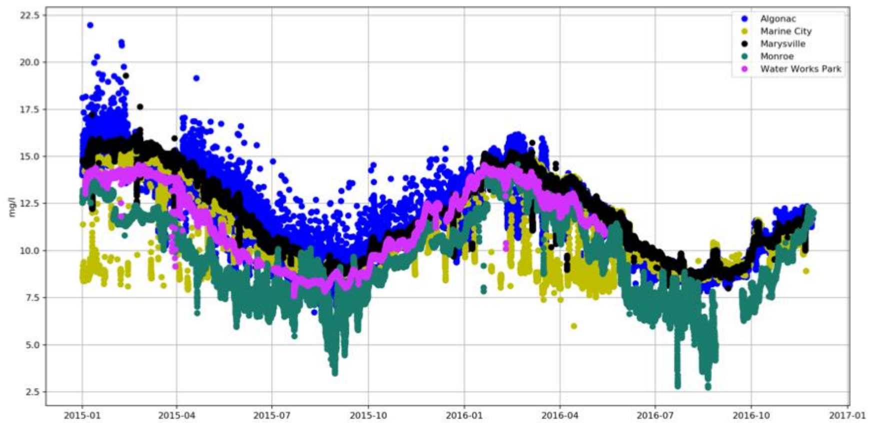


Fig. 10 Time series graph of DO (2015-2016).

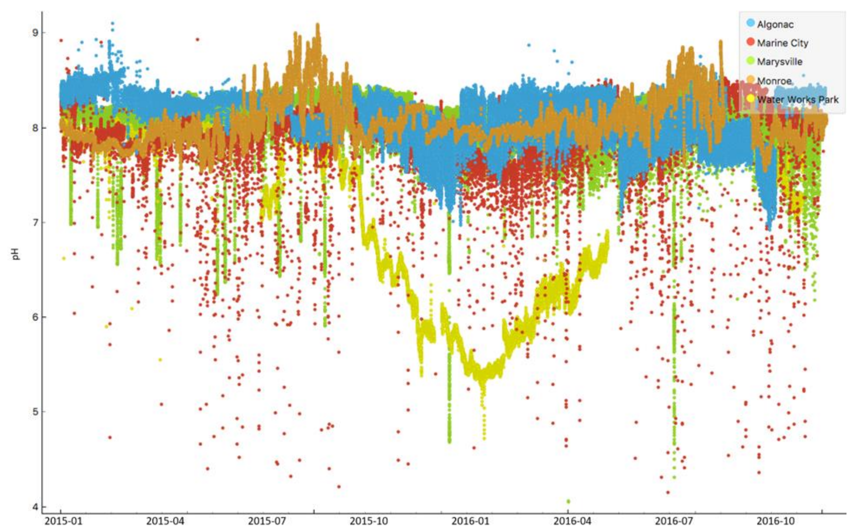


Fig. 11 Time series graph of pH (2015-2016).

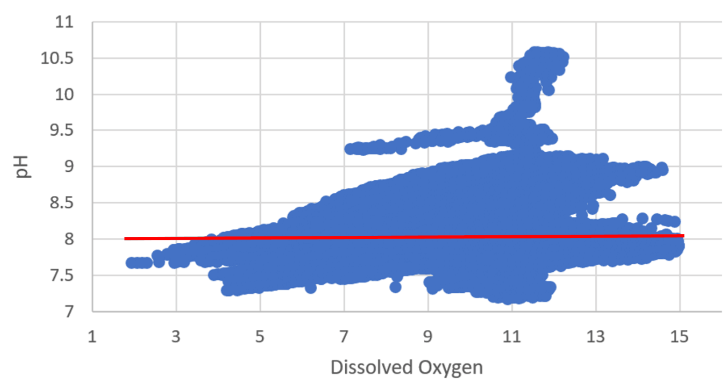
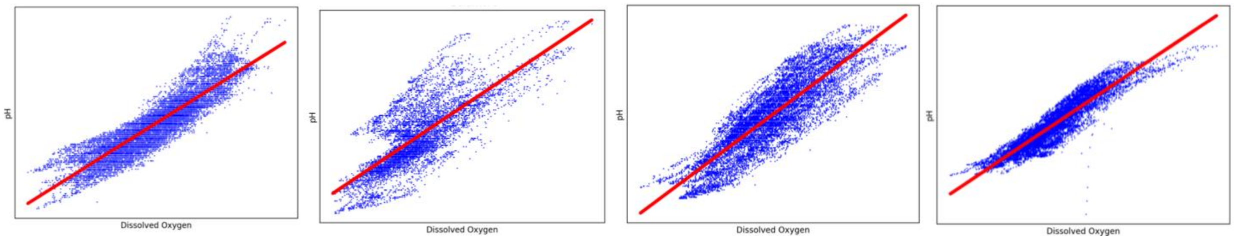


Fig. 12 Relationship between DO and pH in New Baltimore over five years (red line is the best fit line).



(a) May (b) June (c) July (d) August

Fig. 13 Correlation between DO and pH in New Baltimore on a monthly basis (red line is the best fit line).

4.4 Global Relevance

While the present application provides a significant demonstration of multiple platform water quality data use in the US, it is important to note that there is much international progress in this area as well. As a particular example, environmental data collection is being collected outside of utility and research programs in many countries. Information and Communication Technologies (ICTs) is an emerging field, which integrates multiple forms of technology to collect, access, and store data. itis being increasingly used internationally by citizen groups in environmental monitoring (Gouveia and Fonseca 2008), in particular to manage water resources (Fraternali, Castelletti et al. 2012, Mongi and Meinhardt 2016). Even more so, it has allowed citizens to get involved with water governance (Pahl-Wostl, Craps et al. 2007, Hernández-Mora, Cabello et al. 2015). The expanse of this technology has been fairly limited to regions of Africa and Europe; however, this use of technology can easily be expanded to the United States, particularly the Laurentian Great Lakes, which would benefit from expanded water resource data collection.

5 Conclusion

In this paper, we introduce the Huron-to-Erie water quality data platform by demonstrating the mass-oriented, user-friendly and cloud-based characteristics. We also use several example applications of the platform through consideration of e coli, DO, pH, and BGA detection along the Huron-to-Erie corridor.

The example applications suggest some interesting points deserving further study. A significant positive correlation is observed between e coli and one day antecedent precipitation during wet season around recreational beaches. Although the correlations of BGA and Specific

Conductance, BGA and pH, BGA and Turbidity are not statistically significant, the analysis has provided further understanding of the aquatic environments more conducive to algal blooms. In addition, we observe a significant correlation between DO and pH when investigating the data on a monthly basis.

The purpose of this paper was to introduce the availability of the data repository and examples of the types of applications that can be completed using the data and the tools associated with the data platform. Future papers will provide more in-depth and complex data analytics. in the meantime, the data archived in this platform continues to grow as does its value for environmental studies within the Huron-to-Erie corridor of the Laurentian Great Lakes. Seamlessly linking this data platform with others at the upstream and downstream boundary (e.g., the HABS data in Lake Erie) will further expand the utility of this system.

Acknowledgements This work was supported by the Erb Family Foundation and Wayne State University. We acknowledge the contributions of Kelly Karl, Bill Parkus, and Amy Mangus of SEMCOG. In addition, we appreciate the input of Fondriest Environmental and the water utilities along the Huron to Erie corridor.

References

Benesty, J., et al. (2009). Pearson correlation coefficient. Noise reduction in speech processing, Springer**:** 1-4.

Boto, K. G. and J. S. Bunt (1981). "Dissolved oxygen and pH relationships in northern Australian mangrove waterways." Limnology and Oceanography **26**(6): 1176-1178.

Cai, W.-J., et al. (2011). "Acidification of subsurface coastal waters enhanced by eutrophication." Nature Geoscience **4**(11): 766.

Doney, S. C., et al. (2009). "Ocean acidification: the other CO2 problem."

Dorfman, M. H., et al. (2007). Testing the waters: a guide to water quality at vacation beaches, Natural Resources Defense Council New York, NY, USA.

Fraternali, P., et al. (2012). "Putting humans in the loop: Social computing for Water Resources Management." Environmental Modelling & Software **37**: 68-77.

Frieder, C., et al. (2012). "High temporal and spatial variability of dissolved oxygen and pH in a nearshore California kelp forest." Biogeosciences **9**(10): 3917.

Gouveia, C. and A. Fonseca (2008). "New approaches to environmental monitoring: the use of ICT to explore volunteered geographic information." GeoJournal **72**(3-4): 185-197.

Hernández-Mora, N., et al. (2015). "Networked water citizen organizations in Spain: Potential for transformation of existing power structures in water management." Water Alternatives **8**(2).

Hofmann, A. F., et al. (2011). "Hypoxia by degrees: Establishing definitions for a changing ocean." Deep Sea Research Part I: Oceanographic Research Papers **58**(12): 1212-1226.

McLellan, S. L. (2004). "Genetic diversity of Escherichia coli isolated from urban rivers and beach water." Applied and environmental microbiology **70**(8): 4658-4665.

Mongi, H. and M. Meinhardt (2016). "Integrated ICTs for Water Basins Management in Southern Africa: Systematic Review and Meta-analyses for Perceived Relevance Criteria." AGRIS on-line Papers in Economics and Informatics **8**(2): 103.

Oliver, D. M., et al. (2015). "Impact of low intensity summer rainfall on E. coli-discharge event dynamics with reference to sample acquisition and storage." Environmental monitoring and assessment **187**(7): 426.

Pahl-Wostl, C., et al. (2007). "Social learning and water resources management." Ecology and society **12**(2).

Pratt, B. and H. Chang (2012). "Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales." Journal of hazardous materials **209**: 48-58.

Smith, D. R., et al. (2015). "Surface runoff and tile drainage transport of phosphorus in the midwestern United States." Journal of Environmental Quality **44**(2): 495-502.

USEPA (2000). Improved enumeration methods for recreational water quality indicators: enterococci and Escherichia coli, US Environmental Protection Agency Office of Water Washington DC.

Appendix

Given the difference of both data source platform and the sonde model, BGA data associated with two units, cell/ml and RFU. Marysville, fortunately, provides both units at the same time. To make the comparison consistency, the quantitative relationship between the two units is set up according to the BGA information provided by Marysville. Let x represent the BGA value in RFU units, and y represent the BGA value in cell/ml units:

$$y = 2838.3x + 2.8356$$

Using this formula, the BGA values for Monroe are successfully represented by RFU (figure 14).

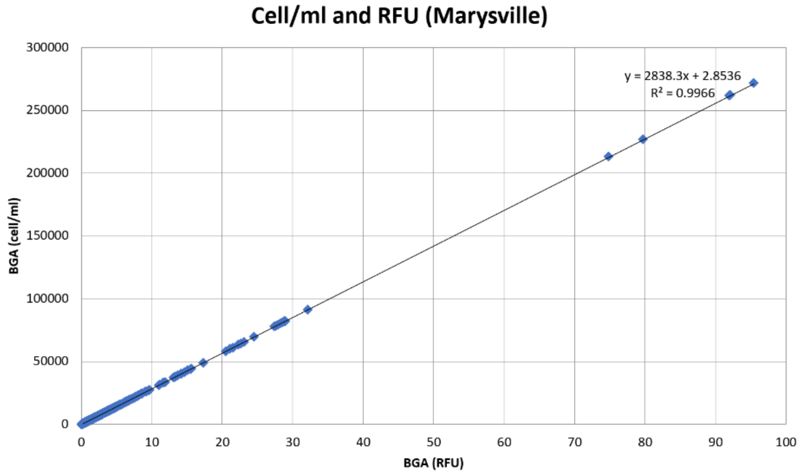


Fig. 14 The relationship between cell/ml and RFU of BGA.