Predicting Water Quality Parameters and Investigating the Impacts of Rainfall on Bacterial Concentrations in Arizona Surface Waters

by

Andrew Hale Buell

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved November 2018 by the
Graduate Supervisory Committee:

Peter Fox, Co-Chair
Morteza Abbaszadegan, Co-Chair
Absar Alum

ARIZONA STATE UNIVERSITY

December 2018

ABSTRACT

One of the two objectives of this dissertation is an investigation into the possible correlation between rainfall events and increased levels of E. *coli* and *Mycobacterium* using an existing data set. The literature states that levels of microbial concentrations do increase after rainfall events, but there are no studies to indicate this correlation applies in any Arizona water systems. The data analyzed for the bacterial concentrations project suggested the possibility of a correlation along one river but it is not conclusive to state that any correlation exists between rainfall events and the microbial concentration for many other sites included in the analysis. This is most likely due to the highly engineered water delivery systems that are not directly impacted.

The secondary objective was to determine if there are environmental variables collected from an ongoing project which would be a good candidate for making predictions about any of the project data parameters. Of the 79 possible opportunities for the model to accurately predict the dependent variable, it showed strong statistical favorability as well as experimentally favorable results towards Dissolved Organic Carbon as the best dependent variable from the data set, resulting in an accuracy of 41%. This is relevant since Dissolved Organic Carbon is one of the most important water quality parameters of concern for drinking water treatment plants where disinfection by-products are a limiting factor. The need for further analysis and additional data collection is an obvious result from both studies. The use of hydrograph data instead of rainfall would be a logical new direction for the heavily engineered water delivery systems.

ACKNOWLEDGMENTS

I would like to thank the many people who have made my research possible. I would like to thank the Salt River Project and Central Arizona Project for their partnership, help, and support. Thank you, Jeff Conyers and Brian Morrehead, at SRP for your help with field sampling and access to Granite Reef Dam. Thank you to Henry Sanger as the USGS for collecting the water temperature data. I would especially like to thank the members of my committee: Dr. Peter Fox, Dr. Morteza Abbaszadegan, and Dr. Absar Alum. I am truly grateful for your belief in me as well as your support. I literally would not have made it this far without you. I would like to thank The City of Anthem, City of Chandler, City of Glendale, City of Mesa, City of Peoria, City of Phoenix, and City of Tempe for their partnership and support of the Regional Water Quality Sampling Project. I would like to thank staff at Arizona State University for their support: Stan Klonowski and Ray Murdock. Finally, I would like to thank my family and friends for their love and unwavering support through graduate school and beyond. I could not have accomplished anything without all of you. You have forever and for better altered the course of my life.

TABLE OF CONTENTS

APPENDIX Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

LITERATURE REVIEW

Introduction

Water quality is an ever changing and dynamic process which is influenced by many different factors / processes and many regulations to sustain its safety for consumers. Long gone are the days, (for most people on Earth) when a thirsty person would walk to a river to fetch water. The World Health Organization (WHO) states that approximately 780 million people in the world drink unsafe water (Mattioli, Pickering, Gilsdorf, Davis, & Boehm, 2013), which in the context of inadequate sanitation and poor hygiene, causes diarrhea and dehydration leading to 1.2 million deaths annually (Mattioli et al., 2013)

That's not to say that even though those who are fortunate enough to live in a region where safe water is readily obtainable for various applications, the development of infrastructure as well as the processes of treating water have its limitations and drawbacks. Technological developments are impacting both the developed and the undeveloped countries in the ways they obtain freshwater and the way it is treated. Many methods include physical means of purification, but also there are many chemical methods for treating water. Aside from the treatment aspect, there is also a need to monitor the water quality. Despite humankinds' best efforts to remove bacteria and viruses and other harmful organisms, today's technology is still far from being able to produce a perfectly safe drinking water. The following section presents a review of

parameters used for water quality monitoring and their significance. These parameters have been monitored in the Regional Water Quality (RWQ) project.

Temperature

Temperature, like many essential chemical elements and organic substrates, is always a potentially limiting factor. Temperature should be viewed as a consistent and, interactive factor, because it influences all chemical as well as biochemical processes. All bacteria have minimum, maximum, and optimal temperature characteristics. (Pomeroy & Wiebe, 2001)

Dissolved Organic Carbon (DOC)

There are several forms of carbon in surface waters and the need to differentiate between them is important in the treatment processes. Autochthonous DOC, which is produced inside the lake by the degradation of phytoplankton and other photosynthetic organisms, does not absorb light and consists mainly of non-humic substances (Bertilsson & Jones, 2003) that are easily utilized by microorganisms (Thurman, 1985). The presence of organic matter in our drinking water at high levels is problematic due to the chemical interactions which can occur in the final step of the treatment processes with the addition of chlorine. If there is too much organic matter in the water after treatment, the addition of chlorine will cause elevated levels of disinfection by-products (DPBs). The need to prevent or minimize the formation of DPBs are primarily due to the public health concerns associated with their production which promotes the potential for cancer,

reproductive and or developmental effects (Richardson, Plewa, Wagner, Schoeny, & DeMarini, 2007).

UV-254

UV254 is an important surrogate for Total Organic Carbon (TOC) analysis while monitoring the water treatment processes as it is an ideal indicator of the concentration of dissolved organic matter. This form of organic carbon is a great predictor of TOC because of its molecular structure which contains unsaturated double or triple bonds that are easily able to absorb ultraviolet light. Allochthonous DOC originates primarily from vascular plants and soil organic matter of the catchment area. It consists mainly of humic substances, is refractory to decomposition, absorbs light, and is colored brownish (Thurman, 1985). DOC is optimal for detecting humic substances in surface waters as they are the primary form of organic matter (Edzwald, Becker, & Wattier, 1985).

Total Dissolved Nitrogen

Total Dissolved Nitrogen (TDN) consists of two fractions: an inorganic fraction which is composed of ammonium (NH^{4+}), nitrate (NO_3^-), and nitrite (NO^{3-}) and an organic fraction i.e., dissolved organic nitrogen (DON), which is an unknown configuration but may include amino acids, proteins, urea, and humic and fulvic acids (Bronk, Lomas, Glibert, Schukert, & Sanderson, 2000). For the purposes of this project, TDN is defined as the fraction of all dissolved organic nitrogen that passes through a Whatman GF/C 0.45micron filter.

E. *coli*

The ability of E. *coli* to survive in the environment is influenced by many factors such as: pH, salinity, stream bed resuspension, sunlight intensity, and temperature (Francy, Gifford, & Darner, 2003). For the purposes of the RWQ project, none of these are measured, except for temperature and there is only data starting from January 2018. While the data is being collected, it should be noted that all samples are processed in a well-lit area or come into contact with direct, artificial light. This is an important distinction to make as light or more specifically, sunlight has an impact on E. *coli* survivability (Blaustein, Pachepsky, Hill, Shelton, & Whelan, 2013).

Total Coliforms

Isolating human pathogens from environmental samples is expensive, and time consuming. Fecal indicator organisms (FIOs) are typically used for water quality assurances and are the basis of implementable and enforceable federal and state regulations for recreational water in the United States (United States Environmental Protection Agency (US EPA), 2004). The invariable control of the sanitary parameters, determining water quality, is used worldwide to monitor and control the quality and safety of various types of water reservoirs, and for prevention of illnesses caused by the polluted water. (Todorov, Iliev, & Trifonova, 2012).

The occurrence data from total coliforms and E. *coli* monitoring studies need the logarithmic transformation of total concentrations per 100 ml asthis has the effect of increasing normality in the dependent variables (D. Kay & McDonald, 1983).

Mycobacterium

Since 1882, the genus *Mycobacterium* have been believed to be dominated by the *Mycobacterium* tuberculosis as the only clinically significant species. It was not until the 1950's that other strains of acid-fast bacilli were cultured from pathological materials leading some researchers to believe that other strains were also of clinical significance. The nontuberculous mycobacteria (NTM) include those *Mycobacterium* species that are not members of the *Mycobacterium* tuberculosis complex (Covert, Rodgers, Reyes, Jr, & Stelma, 1999).

The genus *Mycobacterium* includes approximatly 200 species which differ with respect to their ecologies and pathogenicity (Tortoli, 2014). *Mycobacterium* are typucally resistant to chlorination, so issues related to public health are important due to the fact that disinfection by chlorination or chloramines is the primary mode here in the US for water treatment (Gebert et al., 2018). One set of species of *Mycobacterium* are members of the *M. avium* complex (MAC) which are considered to be opportunistic human pathogens which can infect the lungs, producing cough, fatigue, weight loss, low-grade fever, and night sweats similar to *M. tuberculosis* (Lechevallier, 2006). Due to the health issues associated with *M. avium*, members of the MAC and other mycobacteria have been sought after and recovered from natural surface waters (Von Reyn et al., 1993) and

drinking water systems (Carson, Petersen, Favero, & Aguero, 1978) throughout the United States (Falkinham, Norton, & Mark, 2001). The members of MAC can be transmitted through inhalation or ingestion of contaminated water, soil, or other materials. Evidence for environmental transmission, especially in immunocompromised individuals include (1) the frequency of gastrointestinal colonization increases as the stage immunodeficiency virus (HIV) advances in humans, (2) higher frequency of isolation of MAC from the gut than from the respiratory tract, and (3) gastrointestinal symptoms (e.g., nausea, vomiting, and diarrhea) (Lechevallier, 2006).

CHAPTER 2

REGIONAL WATER QUALITY DATA COLLECTION

Summary

Arizona State University (ASU) has been working with regional water providers (Salt River Project (SRP), Central Arizona Project (CAP)) and metropolitan Phoenix cities since 1998 on algae-related issues affecting drinking water supplies, treatment, and distribution. The results have improved the understanding of taste and odor (T&O) occurrence, control, and treatment, improved the understanding of dissolved organic and algae dynamics, and initiated a forum to discuss and address regional water quality issues. The monitoring benefits local Water Treatment Plants (WTPs) by optimizing ongoing operations (i.e., reducing operating costs), improving the quality of municipal water for consumers, facilitating long-term water quality planning, and providing information on potentially future-regulated compounds. ASU has been monitoring water quality in terminal reservoirs (Lake Pleasant, Saguaro Lake, and Bartlett Lake) continuously from 1998 to the present for algae-related constituents (taste and odors, and more recently cyanotoxins), nutrients, and disinfection by-product precursors (i.e., total and dissolved organic carbon and organic nitrogen). Additional monitoring has been conducted in the SRP and CAP canal systems and in water treatment plants in Phoenix, Tempe and Peoria. During this work the Valley has been in a prolonged drought and recently one above average wet year, and this data provides important baseline data for development of new or expanded WTPs and management of existing WTPs in the future.

The current work has improved the understanding of T&O sources and treatment, but additional research and monitoring into the future is necessary.

Reservoir monitoring is conducted once per month at Bartlett Lake, Saguaro Lake, and Lake Pleasant, and quarterly at Roosevelt, Apache, and Canyon Lakes. Samples are depth integrated in the epilimnion and hypolimnion. CAP will collect samples from Lake Pleasant. SRP will collect samples from Bartlett, Saguaro, Roosevelt, Apache, and Canyon Lakes (at no cost to ASU or cities). Field measurements for temperature with depth will also be collected. River samples (Salt River below Saguaro Lake @ Blue Point Bridge and Verde River at the Beeline Highway) will be collected once per month. Samples will be analyzed for carbon (TOC/DOC), total nitrogen, total phosphorous, arsenic, conductance and T&O compounds (2-methylisoborneol (MIB), Geosmin, Cyclocitrol). The purpose of the lake sampling is to provide early warning information on potentially large changes in water quality – due to algae production, lake destratification, and forest fire or other runoff events. Additional monthly sampling will be coordinated with USGS (Salt River above Roosevelt, Verde River at Tangle) and CAP (Lake Havasu).

Canal monitoring is conducted once per month (January through June) and twice per month as needed during periods of higher T&O production (i.e., July-December). Field measurements for temperature and pH will be made. Sampling will include the CAP, Arizona, and South canals at multiple locations. Monthly samples will be analyzed for carbon (TOC/DOC), total nitrogen, arsenic, conductance, and T&O compounds (MIB, Geosmin, Cyclocitrol). Bi-weekly samples will be analyzed only for T&O

compounds. The purpose of the canal sampling is to identify hot-spots of T&O production and to make recommendations to the cities/SRP/CAP to perform some type of treatment (brushing, copper, etc.). Additional canal sampling will be scheduled to further identify canal hot spots or to provide more frequent process control information.

WTP raw and finished water is collected once per month (January through June) and twice per month as needed during periods of higher T&O production (e.g., July-December). WTP sampling will be conducted at two Tempe WTPs, one Peoria WTP, Glendale WTPs and other selected WTPs. Monthly samples will be analyzed for carbon (TOC/DOC), total nitrogen, arsenic, conductance, and T&O compounds (MIB, Geosmin, Cyclocitrol). Bi-weekly samples will be analyzed only for T&O compounds. The purpose of the WTP sampling is to provide continued evaluation of water quality produced at the WTPs.

Introduction

The regional water quality sampling project began in 1999 with an EPA order for various municipalities in the Phoenix metropolitan area to address taste and odor issues that were plaguing the regional water supply. The compounds of concern were identified as 2-methylisoborneol (MIB) which causes as musty odor, and Geosmin which causes an earthy taste. Both compounds are released by algae that grows in the reservoirs and canals that compose the region's surface water supply, especially during summer when abundant sunshine and warm temperatures promote algal blooms. Complaints from water utility users were so numerous that the issue had to be addressed. Although these compounds do not cause any health problems, they create foul odor and taste which

causes users to lose faith in their water utilities and question the reliability of the treatment process. Both chemicals can be detected in concentrations as low as 10 ng/L and during peak summer algal blooms this concentration is regularly exceeded.

The project has been ongoing since 1999, providing valuable data to the region's water utilities about the water supply. The project data has been used in a variety of other projects examining topics ranging from disinfection by-product formation to invasive species infestation. At Arizona State University, the project is run by Dr. Peter Fox and Dr. Morteza Abbaszadegan.

What Parameters Do We Test?

As part of the Regional Water Quality Sampling Project the following tests are performed:

- UV-254 Absorbance to look for natural organic matter
- DOC (Dissolved Organic Carbon)
- E. *coli* and fecal coliform count
- *Mycobacterium* count
- TDN (Total Dissolved Nitrogen)
- MIB (2-methylisoborneol) and Geosmin
- Turbidity and Conductivity
- Trace metals using Inductively Coupled Plasma – Mass Spectroscopy (ICP-MS). This is done during quarterly sampling.

- Dissolved Oxygen and Temperature

Materials and Methods

Ashing Glassware/DOC Vials/Filters

Prior to use in sample analysis, the materials used for carbon and organic procedures, were ashed to remove any trace carbon present. This was completed by placing the materials into a furnace and heating to 600 °C, which burns off any carbon present. This ensures that the carbon detected during total organic carbon analysis or UV-254 analysis was introduced from the environment and not some other source of contamination. The objects that were ashed include 40 mL vials, DOC vials for the TOC analyzer, and glass filters (GF/F Whatman™, 25 mm diameter CAT No. 1825-025) for sample processing. The procedure to ash is briefly described below:

1. Wrap the materials in aluminum foil. Make sure that everything is completely covered so that it will not become contaminated after removal from the furnace. Filters can be ashed by placing 50 or so in an envelope made by folding aluminum foil (they do not have to be individually placed).
2. Placed wrapped materials into the furnace at 600°C until cycle is complete. Ensure that the door is able to close without obstruction. Close the door and lock into place by turning the handle.
3. Press the run button once that would prompt a program pop up. Press the run button again resulting in a click. The furnace will automatically heat to 600 °C

and cool off. Do not open the furnace door if the internal temperature is more than 100°C. This will result in glassware shattering due to rapid temperature change.

4. Once materials have cooled bring them back to the lab and place them in a safe place.

Autoclaving Bottles for Microbial Samples

Bottles used for microbial samples must be autoclaved prior to use in order to sterilize them. This ensures that all microbial colonies detected during analyses are actually coming from the sample and not just from contamination in the bottles. The procedure for autoclaving the bottles is briefly described below:

1. Make sure that all the material being autoclaved is autoclavable. Non-autoclavable bottles will melt in autoclave, possibly damaging it.
2. Loosen caps on the bottles so that they are only lightly engaged (tightened a quarter turn). If the caps are too tight during autoclaving, the change in pressure will result in rupturing of the bottles.
3. Place a strip of autoclave tape over the top of the bottle/cap to ensure that the cap does not fall off the bottle. The white stripes on autoclave tape will turn black upon completion. The intense black color will verify that the temperature required for sterilization was reached during autoclaving.
4. Check to make sure autoclave (HICLAVE^{TM} HVE-50) is at proper water levels. There should be water inside the bottom of the autoclave and the steam exhaust trap tank should have a level between low and high.

5. Place bottles in the basket and place inside autoclave. Secure the lid by locking the latch.
6. Select solid mode and ensure that temperature is set for 121° C and time is set for 15 minutes. Autoclaving takes approximately 1.5 h.
7. Remove objects from autoclave and secure lids. Store in a place where no one will mistake the identity of sterilized bottles.

Preparation of Microbial Media

Two types of media were prepared – Brilliance and 7H11 for the detection of coliform / E. *coli* and *Mycobacterium*, respectively. Each plate should have 15-20 mL of media in it to avoid drying out during incubation.

Procedure to Prepare Brilliance Media

1. Calculate the amount of media required for the analysis. You can prepare media sufficient for up to 2 months of your experimental needs. Each site where microbial measurements are taken, needs its individual plate.
2. Measure out the required volume of water. You will need to use DI water.
3. Pour water into a flask, place on hotplate, and begin heating. Add a magnetic stirrer and mix.
4. Weigh out the media powder (CM1046, OXOID) needed for the volume of media (the side of the bottle has the required ratio of media to water).
5. Add the media powder and cover the flask with aluminum foil.
6. Once the solution starts to boil and turns transparent turn off the heat.

7. Cool the media down to 50° C (approximately when you can touch the glass without burning your hand).
8. Turn on a Bunsen burner and allow it to run. Using a 25 mL serological pipette add 15 mL of liquid media in each plate. Allow media to cool on biosafety hood.
9. Once media is cool and has solidified place the plates in a bag and label with your name, media type and date. Store lid side down in the refrigerator until used.

Procedure to Prepare 7H11 Media

1. Calculate the batch size of the media you will be making. You can make media sufficient for up to 2 months of analytical needs at a time. Prepared one plate for every 2 sampling sites.
2. Measure out the required volume of water. You will need to use DI water.
3. Pour water into a flask place on hotplate and begin heating. Add a magnetic stirrer and mix.
4. Weigh out the media powder (M0428-500g, Fluka) needed for the batch (the side of the bottle has the required ratio of media to water).
5. Along with required media powder, add 1 mL of 50% glyceryl solution for every 100 mL of solution and cover the flask with aluminum foil.
6. Once the solution starts to boil and turns transparent turn off the heat.
7. Load in autoclave the boiled media along with another flask filled with 100 mL DI water. Autoclave for 15 min liquid default cycle.
8. After autoclave cycle is complete place it on a mixer. Add 10 mL of OADC growth supplement (0678-1VL, Sigma-Aldrich) (kept in the fridge) for every 100

mL of solution. Transfer 10 mL of autoclaved DI water into 1 vial of Panta BBL Antibiotic (245114, Becton, Dickinson and Company) (also kept in fridge) and mix thoroughly. Add this solution to 100 mL of autoclaved media and mix thoroughly.

9. Turn on a Bunsen burner and allow it to run. Using a 25 mL serological pipette add 15 mL of liquid media in each plate. Allow media to cool on biosafety hood.
10. Once media is cool and has solidified place the plates in a bag and label with your name, media type and date. Store lid side down in the refrigerator until used.

Membrane Filtration for E. *coli* and Fecal Coliforms

Within 24 hours of sample collection, samples need to be analyzed using membrane filtration technique. Samples should be refrigerated from collection until membrane filtration. The membrane filtration procedure is briefly described below:

1. Arrange all the supplies required for membrane filtration which include: forceps, Bunsen burner and propane tank, Millipore membrane filters (EZHAWG474, Millipore), ethyl alcohol, lighter, autoclaved filtration cups.
2. Check the level of the pump oil and fill to the line if necessary.
3. Remove the cap from one of the vacuums on the filtration unit. Spray with ethyl alcohol solution and light with Bunsen burner. This will sterilize the surface of filter assembly head.
4. Flame the forceps in the Bunsen burner. Pick up a Millipore filter and place on the vacuum.

5. Place a filtration cup over the vacuum. Pour the sample up to 100 mL line on the filtration cup.
6. Turn on the vacuum pump and twist the valve to open the vacuum to filter the samples.
7. Once all of the water sample has passed through membrane and drained into reservoir flask close the valve and shut off the pump.
8. Flame the forceps, remove the filtration cup, and pick up the filter from the vacuum. Place on top of the Brilliance media plate (hatched side up) being careful not to trap any air bubbles under the filter.
9. Repeat the process until all samples have been filtered. Place all plates in the incubator lid side down. Incubate 18-24 hours at 37℃ and count. Coliform colonies are purple/pink while E. *coli* colonies are blue. Be sure to dump the filtered water in the flask after you perform membrane filtration as part of clean up. If you cannot count the samples right away place them in the fridge lid side down. You have a few days before the plates will no longer be countable.

Membrane Filtration for *Mycobacterium*

Samples must be analyzed within 2 weeks of collection (*Mycobacterium* is a much hardier organism than coliforms). Samples should be refrigerated from collection until analyzed using membrane filtration. The procedure for membrane filtration is described below:

1. Arrange all the supplies required for membrane filtration (forceps, Bunsen burner and propane, Millipore membrane filters, ethyl alcohol, lighter, autoclaved filtration cups).
2. Check the level of the pump oil and fill to the line if necessary.
3. Remove the cap from one of the vacuums on the filtration unit. Spray with 70% ethyl alcohol solution and light with Bunsen burner. This will sterilize the surface.
4. Flame the forceps in the Bunsen burner. Pick up a Millipore filter and place on the vacuum.
5. Place a filtration cup over the vacuum. Pour the sample up to 100 mL line on the filtration cup.
6. Turn on the vacuum pump and twist the valve to open the vacuum to filter the water samples. Drain the sample down to approximately 5 mL and close the valve (shut off the pump).
7. Add 5 mL of 4% NaOH solution. Cover the filtration cup with aluminum foil and let it sit for 30 minutes.
8. After 30 minutes add 5 mL of 3% HCl solution. Allow the sample to sit for 1 minute to neutralize. Turn on the pump and let sample completely pass through membrane and drain the filtrate to reservoir flask.
9. Once all of the sample has drained into reservoir flask close the valve and shut off the pump.

10. Flame the forceps, remove the filtration cup, and pick up the filter from the vacuum. Place on top of the 7H11 media plate (hatched side up) being careful not to trap any air bubbles under the filter.
11. Repeat the process until all samples have been filtered. Place all plates in the incubator lid side down. Incubate for 2 weeks and count.

For cost saving, generally 2 membrane filters are placed on each of the 7H11 media plate. After 2 week incubation, if plates cannot be counted right away place them in the fridge lid side down. However, within a few days the plates will no longer be countable.

Sample Filtration for UV254/DOC

Collected samples are filtered through ashed 0.45-micron glass filters (WHA1825025, Whatman) to remove particles to ensure that the only carbon remaining is dissolved carbon (the difference between DOC and TOC). The procedure for sample filtration is provided below:

1. Make sure that enough number of ashed 0.45-micron glass filters are available before filtration
2. Using a gloved hand place 1 filter inside the filter tip for the 60 mL syringe.
3. Attach the filter tip to the 60 mL syringe once the plunger has been removed.
4. Pour 20-30 mL of your sample from the 250 mL amber bottle into the syringe, swirl and dump out to rinse the syringe.

5. Fill the 60 mL syringe all the way full of your sample. Place the syringe over the 40 mL vial, insert the plunger and flush sample through the filter down to the 40 mL mark on the syringe.
6. Dump the vial out to rinse it.
7. Filter the remaining 40 mL from the syringe into the 40 mL vial and cap.
8. Move on to the next sample.

UV254

To run UV254 follow this procedure:

1. Turn on the Hach DR 5000 by flicking the power switch back panel on the upper left side.
2. System will perform self-check
3. Select single wavelength, confirm that 254 is the wavelength selected
4. Insert cuvette filled with nanopure and select zero
5. Dump the nanopure and fill with your sample then press read, it will display the absorbance value.
6. Record the absorbance value, dump your sample and refill the cuvette with the next sample.
7. Once all samples have been run flick the power switch to turn off the machine.

In between samples you may spill on the cuvette. Use a kimwipe to clean off the surface of the quartz cuvette without scratching it. Make sure that you are using a quartz cuvette. The plastic cuvettes will not give you accurate results.

Dissolved Organic Carbon (DOC)

Once you have filtered your samples into the 40 mL amber vials you can run DOC (Dissolved Organic Carbon) using a Total Organic Carbon Analyzer (Shimadzu TOC-500A). The TOC-5000A was operated using the following procedure:

Turning on and preparing the instrument

1. Turn on TOC analyzer (push button on front panel). Both TOC and TN will turn on.
2. Click on TOC-Control L icon
3. Click on Sample Table Editor
4. Log in
5. From Sample Table panel (on left), select New, System TOC or TOC/TN and press OK
6. Select 'Connect' from upper horizontal bar
7. Check the following
    a. TOC module: Continuous bubbles in IC reaction vessel
    b. TOC module: Humidifier (in right front) – if low, add DI before the run (do not fill above max line).
    c. TOC module: Drain pot (in left back, black lid) – Make sure the vessel is completely filled
    d. TOC module: Halogen tube filled with 0.05 M HCl and bubbling
    e. TOC module: Dilution water bottle filled with acidified nanopure water
    f. Hazardous waste container (behind gas tank) should not be full

8. Select ‘Monitor’ from upper horizontal bar
    a. Wait for TOC, TN, and ASI tabs to have all green check marks
    b. Make sure Supply Gas Pressure is ~200 kPa and Carrier Gas Flow is 150 mL/min
9. Fill rinse vessel with acidified nanopure water (plastic container to the left of the ASI-L). Make sure tubing extends to the bottom of container.
10. Check HCl and H3PO4 containers between TOC analyzer and Autosampler (Fill according to the instructions for standards if necessary).

Total Dissolved Nitrogen (TDN)

The Shimadzu TOC-L with TN module converts all nitrogen compounds to NO at 720 ºC. The instrument uses an auto-sampler to automatically add a small amount of acid to ~50 µL aliquot of sample and inject it onto a platinum catalyst inside the heated combustion chamber. After reaction with ozone, the quantitation is by chemiluminescence. A 1000 mg/L nitrogen stock standard is prepared from ammonium sulfate and potassium nitrate. The instrument automatically calibrates from a single 10 mg/L N solution to establish a multiple point calibration curve from 0.2 – 10 mg/L N. The instrument will automatically dilute (or injects less sample aliquot) off-scale peaks, enabling quantitation up to 500 mg/L. Concentrations higher than 500 mg/L N are diluted manually. Total analysis time, per injection, is 2 – 5 minutes. The Method Detection Limit (MDL) is 0.05 mg/L N (*WHITEPAPER Introducing a New ASTM Method for the Determination of Total Nitrogen, and TKN by Calculation, in Water Samples Analytical and Measuring Instruments*, n.d.).

Conclusion

The Regional Water Quality project has been ongoing for many years. Its continued support and funding shows that the project is not only important or relevant but is a useful tool for water treatment operators to keep track of current trends regarding the status of water inflows and their respective treatment processes. The long-term continuation of the project has resulted in the generation of a sizeable dataset which may as an entire set, be useful to analyzing for probable trends and make predictions (Chapter 2). Such analyses can help to better understand the hydrologic cycle as well as the trophic structures of reservoirs in the greater Phoenix metropolitan area that affect the water quality.

CHAPTER 3

PREDICTIVE MODELING

Summary

Starting in June 2014 the Regional Water Quality project began to collect microbial data to monitor the occurrence of Total coliforms, E. *coli* and *Mycobacterium*. The model generated using the Statistical Package for Social Sciences (SPSS) includes microbial data as a parameter, from the corresponding sampling times frame the non-microbial data such as DOC, TDN, UV254, will be used in this analysis. At the time of initial microbial data collection, no surface water temperatures were taken. Due to the influence temperature has on microbiological systems, it was deemed necessary to obtain temperature data to be able to include in the model as a parameter. Surface water temperatures were obtained by contacting the United States Geological Survey (USGS) for temperature data regarding the gage located just below Stewart Mountain Dam, which is upstream of the SRP canal surface water entry system known as Granite Reef Diversion Dam. This gage was used as for its relative closeness to the sampling sites.

To determine the accuracy of the model, the data collected from June 2014 thru December 2016, was used to make predictions the values of all the parameters for the period starting from January 2018 thru June 2018. To obtain a best fit style model, each iteration was evaluated using statistical analysis to determine the best dependent variable. Since the number of known parameters is limited, and the data is based on environmental monitoring, the number of variables needed to create a functioning model is hard to predict, so all model variations will still include variables that are deemed statically

insignificant at the $p<0.005$ level. All iterations of the equation were used to make predictions about the 2018 data collected and a percent difference was calculated to determine model accuracy. Data for 2017 was not used due to a difference in water release from one of the dams. Normally water is released from the bottom of the dam, while for most of 2017, water from this particular dam was released via an overflow spillway causing unusually high levels of MIB/Geosmin.

Introduction

There is a large data set on the historical quality of water in Central Arizona. Sometimes collecting data is not always possible due to canal maintenance or when treatment plants shut down for routine maintenance. Two of the explanatory variables are closely related, i.e., UV and DOC, so there is a higher chance of multicollinearity. The model will be used to makes predictions about future parameter values and not to make sense of the data in a way which explains trends or make physical sense of the output; therefore, multicollinearity is acceptable (Chris A. Mack, 2016). The objective of this study was to determine the utility of various parameters in accurately determining the water quality parameters in future.

Materials and Methods

To test the accuracy of the different models, the equations were input into Excel and then used to make predictions from January 2018 through July 2018. The predicted numbers were checked against the actual number and a percent difference was calculated to determine model accuracy. Certain sampling sites, but not any one in particular, are

missing data points, so only data sites that have all model parameters are included in the results. That is, if there is any data missing from the table, then the predicted value, regardless of the accuracy, is not being included in the accuracy results. In addition, to determine the amount of influence DOC and UV-254 had on the model, independent of one another, each variable was removed from the model to see how much of an impact on accuracy each had when not used together.

Water Temperature Data

The data used to generate the SPSS model for which prediction are made, was the seven-day average surface water temperature as reported by USGS gaging station which is located along the Salt River just below the Stewart Mountain Dam. The gauge of interest (See appendix A for image location) is at a location nearest the start of surface water entry into the SRP canal systems known as Granite Reef Diversion Dam. USGS gaging station located after Stewart Mountain Dam was chosen due to its relative distance to the nearest sampling site located at Granite Reef Diversion Dam, where the surface water is channeled into two canals. However, the temperature data in the SPSS generated equation used to make predictions for the 2018 test year, is data collected from each site where microbial samples are required using a FLUKE 62 MAX IR Thermometer. The IR thermometer has a +/-0.02 accuracy at 15 feet or greater. The water sampler will stand as close to the canal as possible, point the thermometer at the water so that their arm is approximately parallel to the canal wall and hold the trigger until the

digital display stops on a value. These measurements are presented in a table located in appendix A.

SPSS Input Procedure

Open SPSS and from the top ribbon, click on the Analyze tab, then select Regression, then Linear. Once in the Linear regression program input menu, select the Dependent variable from variable list and then pressing the arrow button pointing to the dependent input box. For the Independent variables, follow the same procedure only press the arrow button leading to the independent box.

Next, while still in the input menu, click on the Statistics tab to open the next dialogue box. From the appropriate boxes, select estimates, Confidence intervals (leave at 95%), Model fit, Descriptives, Part and partial correlations, Collinearity diagnostics, Durbin-Watson and casewise diagnostics (which is used to determine any outlier data) Leave the standard deviations set to 3. Once all these are selected, press continue. Next clicking on the plots tab, under standardized Residual plots, click Histogram, Normal Probability plot and produce all partial plots. Click continue. Under the save tab, click the following boxes in all selection categories. Unstandardized, Studentized, Studentized deleted, cooks, leverage values. Click continue. Next click OK on the main menu to run the analysis. The SPSS output file will contain the detailed statistical information used to determine if the regression model generated is a good fit for the data. The following parameters are what will be used to determine the efficacy of the model.

SPSS Procedure to Check Model Goodness

The following procedure was used from several sites including the Laerd Statistics website on linear regression analysis as well as University of California, Los Angeles (UCLA) Institute for Digital Research and Education.

$R^2 or\ adjusted\ R^2$:

This parameter is a numerical representation of how well the model fits the data. The R-squared statistic indicates the percentage of the variance in the dependent variable that the independent variables collectively explain. R-squared measures the strength of the relationship between your model and the dependent variable on a 0 – 100% scale.

p-value:

The p-value is compared to some alpha level in testing the null hypothesis that all of the model coefficients are 0. Usually this value should be less than or equal to 0.05

Outlier data:

When the casewise diagnostics box is checked in the SPSS Linear regression dialogue box, the program when preforming the analysis will note any iteration where the standardized residual is greater than +/- the number of standard deviations set in the dialogue box. This is seen in the casewise diagnostics output box, after the program has been run as the form of the residual. That is, SPSS makes a prediction, takes the difference from the actual input value and creates a residual. If this residual is +/- the set number of standard deviations as denoted in the program, it is labeled an outlier and must

be dealt with. In the event that none of the data meets the criteria set in the dialogue box, then this will not be seen in the output file.

Additionally, descriptive statistics will run on the SPSS created variable labeled SDR_1(Studentized Deleted Residuals) to determine the Standard deviation. Both ends of the data will be checked to see if they also met the +/- 3 Standard deviations. The standard deviation of the SDR_1= some value. 3 times this value will yield you limit and should be used to check against the highest and lowest SDR_1 value in the program after the initial analysis is performed. If there are data which do not meet this requirement, they are labeled outlier and will need to be dealt with.

Results and Discussion

The model developed using SPSS was imported into an excel sheet to calculate the predicted values. The model data ends in December 2016 and is used to make predictions for January through June of 2018. In Excel, the theoretical values were set next to the actual value and a percent error was calculated using the following formula:

$$Percent\ Difference = \frac{Actual - Theoretical}{\frac{(Actual + Theoretical)}{2}} * 100$$

The negative sign was left in the analysis to indicate whether the expected values was above or below the theoretical. If the value is negative, then the model predicted a value which was higher than the actual. A positive value means the model predicted a value which was less than the actual.

There are a total of 91 possible opportunities (13 sites, for 7 months) for the model to accurately predict the value of the dependent variable. This number will be adjusted to account for times where there is no data for the model to make predictions

about. This is seen in the spreadsheet as "#VALUE!". For these cases there is either no monthly data or one of the predictor variables is missing and the equation generates an output of "#VALUE!". This occurs a total of 12 times and is removed from the potential number of times the model can make an accurate prediction. This yields a potential of 79 possible chances to have an accurate model.

Once the data was input into Excel, threshold values were assigned based on the likelihood of the output being relevant to a plant operator. That is, the data was separated into three categories based on their percentage of accuracy. Category one(C1): $\pm 10\%$, category two(C2): between $\pm 10 - 15\%$, and category three(C3): greater than $\pm 15\%$. These categories were applied to all possible iterations and were tallied to determine which parameter yielded the most desirable results, with C1 being the optimal or most desirable outcome and C3 being the least desirable. When the model was to include a trial run where DOC and UV-254 were both included as dependent and independent variables, the scenario in which the dependent variable was DOC had an accuracy of 41% in C1 and 41% in C3. The least desirable trial had TDN being the dependent variable with a C1 of 16% and a C3 of 84%. When DOC and UV-254 were included separately in the model, where DOC was dependent and UV-254 was not included as an independent variable, and also the other way around, the model predicted DOC with a C1 value of 5% and a C3 value of 91%. Alternatively, UV-254 produced a C1 value of 30% and a C3 value of 57%. The model increased in accuracy by 2% over the previous trial where UV-254 was the dependent and DOC was included as an independent, with a C1 of 28%.

The data output given the discussed input parameters yielded a statistically desirable model using the discussed output analysis procedure, but when applied to an actual dataset, failed to preform or accurately predict output parameters which may not be useful to water treatment operators. This is most likely due to the randomness associated with the environmental monitoring data, which makes it difficult to make predictions for. Another possibility is that there are not enough variables to produce a viable model. A similar paper on E. *coli* prediction used as many as 20 variables, many of which were hydrologic variables (D. Kay & McDonald, 1983). Surface waters contain many nutrients and phytoplankton which die and produce organic matter which feed the bacterial population, and ultimately helps to increase variability of organic loading in the sample (Bertilsson & Jones, 2003). Every attempt is made to make sure the water is cooled during transport, this is not always the case. The ability to control environmental conditions which promote optimal survivability varies while in the field. This may result in data that is not as accurate as if the data collected were entirely from a controlled environment.

The possibility of having too many data points is known to be just as detrimental to model success as having too few (Francy et al., 2003). It is possible that occurrence of multicollinearity is due to inclusion of two interdependent variables: DOC and UV-254, which was overlooked and is playing a factor. The decision to leave the dependent variables in the model was due to the poor statistical output generated by SPSS via their exclusion. The inclusion of these two dependent variables was allowed since the model was not being used to analyze the data for any significance, only for predictions (Chris A.

Mack, 2016). The exact degree of permissible multicollinearity has not been adequately defined in the literature (D. Kay & McDonald, 1983). Figures one and two seen below summarizes the results of each category and its respective accuracy for each dependent model variation.

Table1: DOC/UV-254 dependent variable iterations

	Category 1(C1)	Category 2(C2)	Category 3(C3)	% correct C1	% correct C2	% correct C3
UVA	22	13	44	0.28	0.16	0.56
DOC	32	15	32	0.41	0.19	0.41
TDN	13	0	66	0.16	0.00	0.84
LnEcoli	13	10	56	0.16	0.13	0.71
TempC	19	20	40	0.24	0.25	0.51

Table 2:DOC/UV-254 removed variable iterations

	Category 1(C1)	Category 2(C2)	Category 3(C3)	% correct C1	% correct C2	% correct C3
UVA	24	10	45	0.30	0.13	0.57
DOC	4	3	72	0.05	0.04	0.91
TDN	13	0	66	0.16	0.00	0.84
LnEcoli	13	10	56	0.16	0.13	0.71
TempC	19	20	40	0.24	0.25	0.51

Conclusion

Overall the data provided in the model, from a statistical point of view, yielded a positive result as seen by the satisfactory residuals plot as well as the other statistical checks. However, when applied to actual data, the model failed to perform in every category for every model iteration except for the category 3 (C3) which was the largest category for output results and included every possible answer greater than 15%. Because of this, the model can be considered to be non-functional. Additional work is required to enhance the predictive capability of the equations by considering more variables such as hydrologic variables as well as more spatially relevant temperature data.

CHAPTER 4

RAINFALL IMPACT ON BACTERIAL CONCENTRATIONS

Summary

The objective of this study was to determine if there is an increase in microbial loading in select Arizona surface waters after a rainfall event. No special collections or measurements were taken. The microbial data used for analysis was obtained from an existing data set and the rainfall data was collected from the Salt River Project (SRP). To see if there is any impact on bacteria levels after a rainfall event, the rainfall data was temporally aligned with the sampling dates from when the water was collected to test for *Mycobacterium* and E. *coli.* Graphs were created (See appendix A) to visually represent the two data sets and then a visual analysis was performed. Data suggests correlations between the two data sets for E. *coli* at one testing site, but the data is not conclusive to examine if any correlation exists between rainfall events and the microbial concentration for many other sites.

Introduction

There are many factors which affect run off. Some are characteristic of the storm itself which includes the type of precipitation such as rain or snowfall, the intensity of the storm, and the duration of the storm, while others are characteristic of the land the water is flowing over. These include but are not limited to: land elevation, soil type and vegetation or land cover type. During a rainfall event, the soil, if the intensity and duration of the storm is strong enough and long enough, will become saturated and will

no longer be able to absorb water. The water that touches a saturated surface turns into what is known as storm runoff. Water will always, unless moved by humans, flow from higher elevation to lower elevation. While this water is traveling, there can be many things it can encounter. Water is a powerful force. Two feet of flowing water can move smaller cars and is capable of moving large vehicles.

Stormwater will most likely end up in one of several places, which includes lakes, rivers and canals. These are water bodies used as sources of drinking water in the greater Phoenix metropolitan area. With these flowing waters comes all the things that it picks up along the way including chemicals from parking lots or spills, nutrients from fertilizer from agricultural land or feces from animals and humans, some of which will end up in the surface waters we play in and ultimately consume at the tap.

Although the impact of contamination has not been quantified along the Arizona or South canals, there is at least one known site along the Arizona canal where an irrigation channel surrounding agricultural land has water discharging directly into the canal (See appendix A for image location). In addition, along the Verde River, known to the Regional Water Quality Project as sampling location "R25" (See appendix A for image location) there is at least one site which has been observed to have cattle grazing. Grazing next to the river has the potential for the runoff to carry the feces of cattle, wild horses, or humans.

The contamination of surface waters with pathogenic microorganisms transported from fields which have been treated with livestock slurries and manure, is a serious environmental concern because it may lead to human exposure to such micro-organisms

via several routes: drinking water (Ongerth and Stibbs, 1987; Hansen and Ongerth, 1991; Poulton et al., 1991; Skerrett and Holland, 2000); bathing waters (Geldreich, 1996; Wyer et al., 1996; Baudart et al., 2000); and water used for the irrigation of ready to eat foods (Tyrrel, 1999).

Materials and Methods

The values in the data columns are calculated from the monthly average for each location. The Arizona Canal average reported in the rainfall graph is the average of all six sampling locations along the Arizona Canal. The data reported for the Verde River was collected in the same manner only the Verde River average consists of only one sampling point. Since the data set being analyzed includes multiple water sources, locations were chosen to best represent an inflow of rainfall in to the system from each source. The two rainfall gauges selected for analysis and data collection are situated along two different rivers. One is the Salt River System and the other is along the Verde River. These two rives converge on one another approximately 4 miles upstream of granite reef diversion dam, which is where the water is diverted into two canals used for conveyance throughout the valley.

Rainfall data was collected form the SRP website which shows all SRP gages with real time data as well as some historical data. Rainfall data is not of archival quality so only 2 years of data are able to be collected at any one time. The timeframe for the rainfall data from the specified gages begins on August 21st, 2016 and ends on August 20th, 2017. When collected from SRP, the data is represented as cumulative rainfall

which was then converted into a hyetograph of rainfall events rather than continuous rainfall so that the data could be plotted against microbial data for the recorded rainfall events. This conversion was made by subtracting the current days cumulative rainfall from the previous days cumulative rainfall which yielded the rainfall for the day of interest. The historical microbial data collected by the RWQ was analyzed for collection dates that were nearest the recorded rainfall events as possible. Also, only rainfall events of over 0.5 inches were to be included as any less would likely not result in runoff into any streams or reservoirs. Although it should be noted that just the amount of recorded precipitation is not always a good indicator of the type of storm which occurred. Another important factor that would be important to note would be the intensity of the event. A flash storm may be more likely to produce runoff than a slow storm which yielded equivalent rainfall.

Results and Discussion

Once both data sets were compiled, they were plotted against each other to observe for any trends. The graph for the Verde River seems to make a reasonably compelling argument that there may be some correlation between rainfall events and increased microbial activity. However, looking at the AZ Canal graph, the correlation seems much weaker. This may be due to dilution as the AZ canal graph is water from two sources whereas the Verde River graph is only from one source. Literature has shown a relationship between rainfall events and increased microbial loading of reservoirs as the results of one study show that peak loadings of bacterial concentration at a reservoir outlet are likely to occur when a period of heavy rainfall causes rapid filling of a depleted

reservoir (David Kay & McDonald, 1980). Another study whose investigation parameters among others also included Total Coliforms and E. *coli* reported that concentrations of these microbes increased considerably during intense runoff events (Kristemann et al., 2002).

Conclusion

The study graphs show the possibility of correlation at one site but are not conclusive enough to suggest a correlation among the other sites. More information will need to be gathered to conclusively say that there is a correlation between rainfall and increased microbial activity. With this small-scale data collection, some key variables were not accounted for such as: Time of concentration, or the amount of time it takes for a water molecule to travel from the inlet to the out let of a water shed, as well as the lifespan or decay rate of the microbes being investigated

The most important factor to be considered is that the sampling date needs to be carefully aligned after the rainfall event such that the sampling location is receiving the runoff from the storm. Sampling too early or too late would yield inconclusive data which is most likely included in this analysis.

CHAPTER 5

CONCLUSION

Predictive modeling

There is not any one conclusive resolution to any of the issues presented in this paper. Rather, there is an emphasis on the need for more research into technical areas of water treatment. The possibility of data collection being useful will only increase as more and more data is collected. This also as can be seen in the predictive modeling portion, as it is entirely possible that with more variables, the better the model will be able to predict. The decision to include or exclude certain variables was due to the number of available data sets and or the certainty of the quantitative analysis. For example, the decision to use E. *coli* vs Total Coliforms is due to the number of coliforms present in the water, vs the number of E. *coli* present. Both are quantitative values, but the room for error in terms of the correct amount being counted is much lower with E. *coli* as these values are typically on orders of magnitudes less than Total Coliform. This means that the colony counts are more accurate as they are easier to count. Also due to the nature of culturing or activating bacteria which can be a difficult process, it is possible that not all bacteria present in the water sample were able to grow on the media. If there was an error in handling the samples, the nutrient content of the medium was not correct, or the incubation or transportation temperatures were not stable, all of these could result in inactivation issues or premature bacterial death. This can most likely be seen in the model as the model the number or times prediction was useful was far less than those where it was useful.

Of all the variables selected for analysis, the most useful was when DOC was the independent variable. Of the 79 possible chances for the model to make an accurate prediction under any possible dependent variable variation, when DOC was the dependent variable, the theoretical output was within 10% of the target or actual DOC concentration 41% of the time. This scenario is almost twice as high as all other dependent variable scenarios. Overall, the model was overwhelmingly out of the greater than 15% range for every scenario by up to 4 times as much.

The statistical output presented by SPSS as well as the analytical application of the model suggests that with more modifications or the addition of variables, there is the possibility to increase model functionality. More hydrologic variables may help to account for important factors such as rainfall intensity, hydrographs, as well as land cover type. The inclusion of information about the land cover will help to present a better picture of how the rain actually flows overland into the river or stream.

Rainfall and Bacterial Concentrations

With the limited amount of data available for this analysis there are not many conclusions to be made. Elevated levels of colonies during or after a storm is supported by the literature (Kristemann et al., 2002). However, On the Verde River graph data from August 2016 through January 2017 suggests that some correlation may be present, while February 2017 through August 2017 does not support this theory as there are elevated numbers of *Mycobacterium* colonies with little to no precipitation. It is possible that these are the months that SRP switches to ground water. Due to the difficulty in culturing

bacteria, there is a significant number of sampling days with no data. Between the two graphs, for the same rainfall event, there are significant differences in the average colonies detected between the two source waters. Since AZ Canal contains both Salt and Verde River water it's possible that a dilution is happening, but an individual sample from just the Salt River would be necessary to better determine if this is happening or not.

A better analysis would be to independently collect water from each river instead of one sample being from an induvial source and the other sample being a mixture which includes water from the first sample. Collecting water samples immediately after a rain fall even would also help to ensure that the microbial transport into the water is being collected as the time of concentration is very quick. The probability that any microbes transported into the river, still being present at that location several days after a rainfall event would most likely not be very high.

The lack of historical or archival quality rainfall data for this investigation it the largest obstacle to overcome. Rainfall data is not kept longer than 2 years by SRP due to the difficulties in maintaining the gages. Also, the need to collect samples immediately after a rainfall event ends and the overland flow stops would help to ensure that the samples collected contain water which has carried nutrients and bacteria over land, into the lake or stream. The method outlined in this paper does the exact opposite. The contradictory methods are acceptable for now because this was an exploratory investigation into the possibility of a correlation using an existing data set. The fact that a correlation was not found does not mean one does not exist. Despite what the literature

has stated about increased microbial loads after rainfall (Kristemann et al., 2002), for this dataset there has not been enough analysis to determine whether or not there is any correlation between rainfall and increased levels of *Mycobacterium* or E. *coli*.

REFERENCES

Baudart, J., Grabulos, J., Barusseau, J.-P., Lebaron, P., 2000. Salmonella spp. and fecal *coli*form loads in coastal waters from a point vs. non-point source of pollution. Journal of Environmental Quality 29, 241–250.

Bertilsson, S., & Jones, J. B. (2003). Aquatic Ecosystems. *Aquatic Ecosystems*, 3–24. https://doi.org/10.1016/B978-012256371-3/50002-0

Blaustein, R. A., Pachepsky, Y., Hill, R. L., Shelton, D. R., & Whelan, G. (2013). Escherichia coli survival in waters: Temperature dependence. *Water Research*, *47*(2), 569–578. https://doi.org/10.1016/j.watres.2012.10.027

Bronk, D. A., Lomas, M. W., Glibert, P. M., Schukert, K. J., & Sanderson, M. P. (2000). Total dissolved nitrogen analysis: Comparisons between the persulfate, UV and high temperature oxidation methods. *Marine Chemistry*. https://doi.org/10.1016/S0304-4203(99)00103-6

Carson, L. A., Petersen, N. J., Favero, M. S., & Aguero, S. M. (1978). Growth characteristics of atypical mycobacteria in water and their comparative resistance to disinfectants. *Applied and Environmental Microbiology*, *36*(6), 839–846. https://doi.org/10.1086/430500

Chris A. Mack. (2016). *Lecture51 (Data2Decision) Addressing Multicollinearity - YouTube*. Austin. Retrieved from https://www.youtube.com/watch?v=5NVcmLZCGOg

Covert, T. C., Rodgers, M. R., Reyes, A. L., Jr, G. N. S., & Stelma, G. N. (1999). Occurrence of Nontuberculous Mycobacteria in Environmental Samples Occurrence of Nontuberculous Mycobacteria in Environmental Samples, *65*(6), 2492–2496.

Edzwald, J. K., Becker, W. C., & Wattier, K. L. (1985). Surrogate parameters for monitoring organic matter and THM precursors. *Journal / American Water Works Association*, *77*(4), 122–131. https://doi.org/10.1002/j.1551-8833.1985.tb05521.x

Falkinham, J. O., Norton, C. D., & Mark, W. (2001). Factors Influencing Numbers of Mycobacterium avium , Mycobacterium intracellulare , and Other Mycobacteria in Drinking Water Distribution Systems Factors Influencing Numbers of Mycobacterium avium , Mycobacterium intracellulare , and Other Mycobacteria in. *Applied and Environmental Microbiology*, *67*(3), 1225–1231. https://doi.org/10.1128/AEM.67.3.1225

Francy, D. S., Gifford, A. M., & Darner, R. A. (2003). Escherichia coli at Ohio Bathing Beaches - Distribution, Sources, Wastewater Indicators, and Predictive Modeling, 47.

Gebert, M. J., Delgado-baquerizo, M., Oliverio, A. M., Webster, T. M., Nichols, L. M., Honda, J. R., … Fierer, N. (2018). Ecological analyses of mycobacteria in showerhead biofilms and their relevance to human health. *BioRxiv*, *9*(5), 366088.

https://doi.org/10.1101/366088

Kay, D., & McDonald, A. (1980). Reduction of coliform bacteria in two upland reservoirs: The significance of distance decay relationships. *Water Research*, *14*(4), 305–318. https://doi.org/10.1016/0043-1354(80)90076-7

Kay, D., & McDonald, A. (1983). Predicting coliform concentrations in upland impoundments: Design and calibration of a multivariate model. *Applied and Environmental Microbiology*, *46*(3), 611–618.

Kristemann, T., Claßen, T., Koch, C., Dangendorf, F., Gebel, J., Vacata, V., … Fischeder, R. (2002). Microbial Load of Drinking Water Reservoir Tributaries during Extreme Rainfall and Runoff Microbial Load of Drinking Water Reservoir Tributaries during Extreme Rainfall and Runoff. *Applied and Environmental Microbiology*, *68*(5), 2188–2197. https://doi.org/10.1128/AEM.68.5.2188

Lechevallier, M. W. (2006). Complex, (1), 125–130.

Mattioli, M. C., Pickering, A. J., Gilsdorf, R. J., Davis, J., & Boehm, A. B. (2013). Hands and water as vectors of diarrheal pathogens in Bagamoyo, Tanzania. *Environmental Science and Technology*, *47*(1), 355–363. https://doi.org/10.1021/es303878d

Pomeroy, L. R., & Wiebe, W. J. (2001). Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquatic Microbial Ecology*, *23*(2), 187–204. https://doi.org/10.3354/ame023187

Richardson, S. D., Plewa, M. J., Wagner, E. D., Schoeny, R., & DeMarini, D. M. (2007). Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. *Mutation Research - Reviews in Mutation Research*, *636*(1–3), 178–242. https://doi.org/10.1016/j.mrrev.2007.09.001

Todorov, O. Y., Iliev, I. I., & Trifonova, S. K. (2012). Study on the Total Coliforms Count and Coli Titter in the Waters of Kardzhali Reservoir, Bulgaria, *4*(2), 15–23. Retrieved from http://eb.bio.uni-plovdiv.bg

Tortoli, E. (2014). Microbiological Features and Clinical Relevance of New Species of the Genus Mycobacterium, *27*(4), 727–752. https://doi.org/10.1128/CMR.00035-14

Von Reyn, C. F., Waddell, R. D., Eaton, T., Arbeit, R. D., Maslow, J. N., Barber, T. W., … Falkinham, J. O. (1993). Isolation of Mycobacterium avium complex from water in the United States, Finland, Zaire, and Kenya. *Journal of Clinical Microbiology*, *31*(12), 3227–3230. https://doi.org/10.1016/j.trf.2016.10.002

WHITEPAPER Introducing a New ASTM Method for the Determination of Total Nitrogen, and TKN by Calculation, in Water Samples Analytical and Measuring Instruments. (n.d.). Retrieved from https://www.epw.senate.gov/water.pdf,

Wyer, M.D., Kay, D., Dawson, H.M., Jackson, G.F., Jones, F., Yeo, J., Whittle, J., 1996. Delivery of microbial indicator organisms to coastal waters from catchment sources. Water Science and Technology 33 (2), 37–50.

APENDIX A

FIGURES AND TABLES

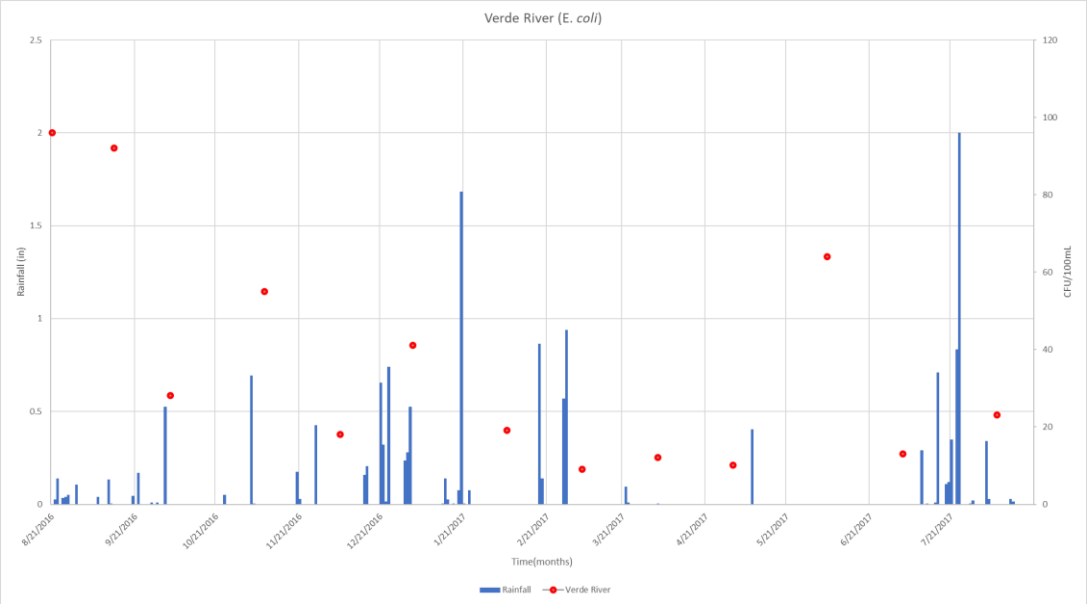


figure 1: E. *coli concentrations along the Verde river*

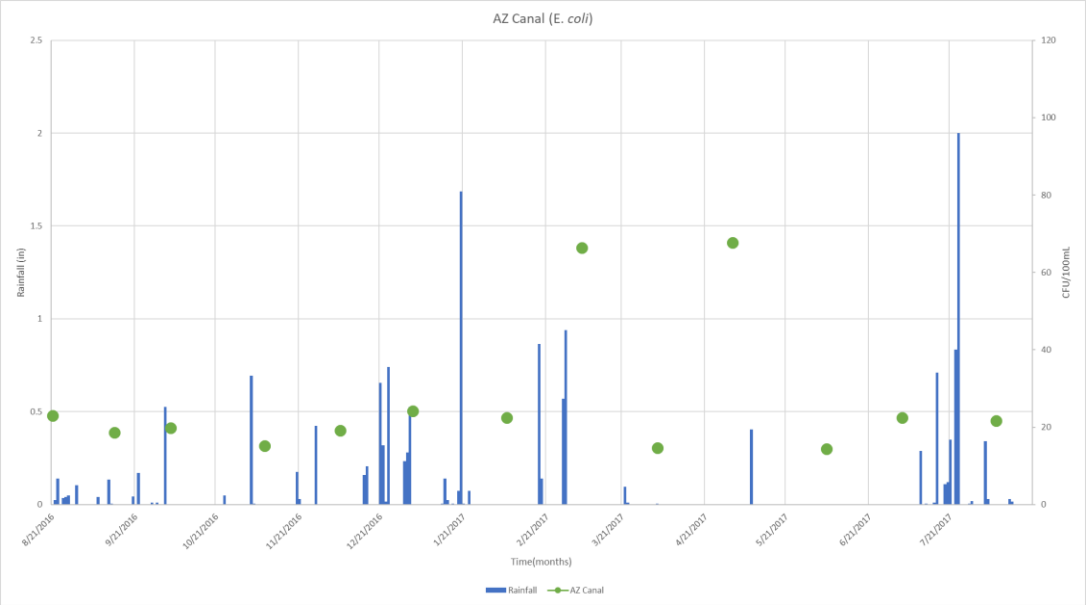


figure 2: E. *coli concentrations along az canal*

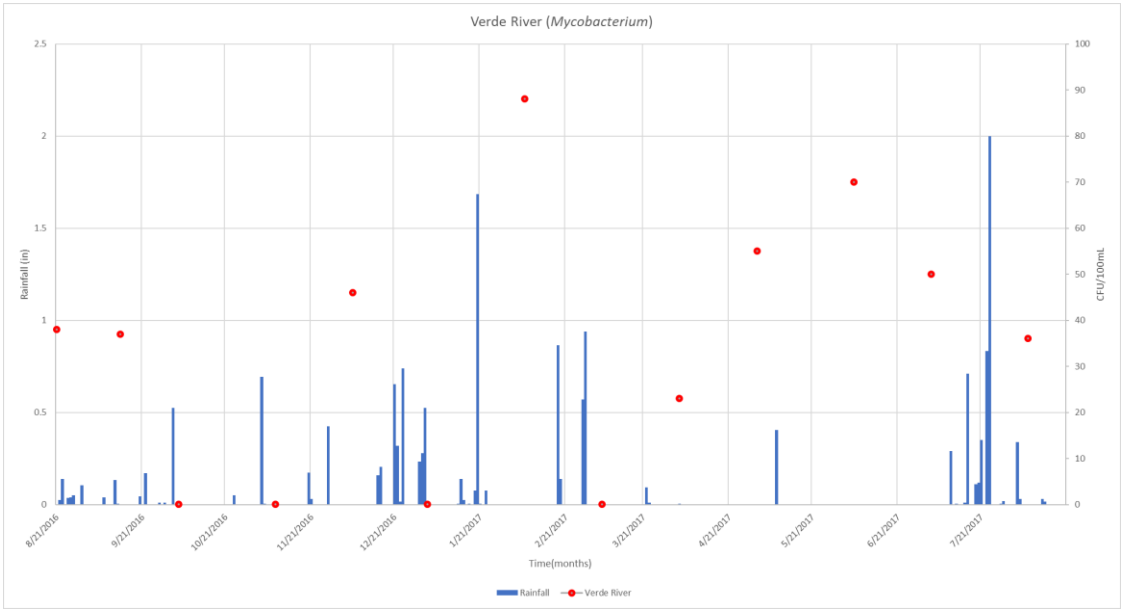


figure 3: Mycobacterium concentrations along the verde river

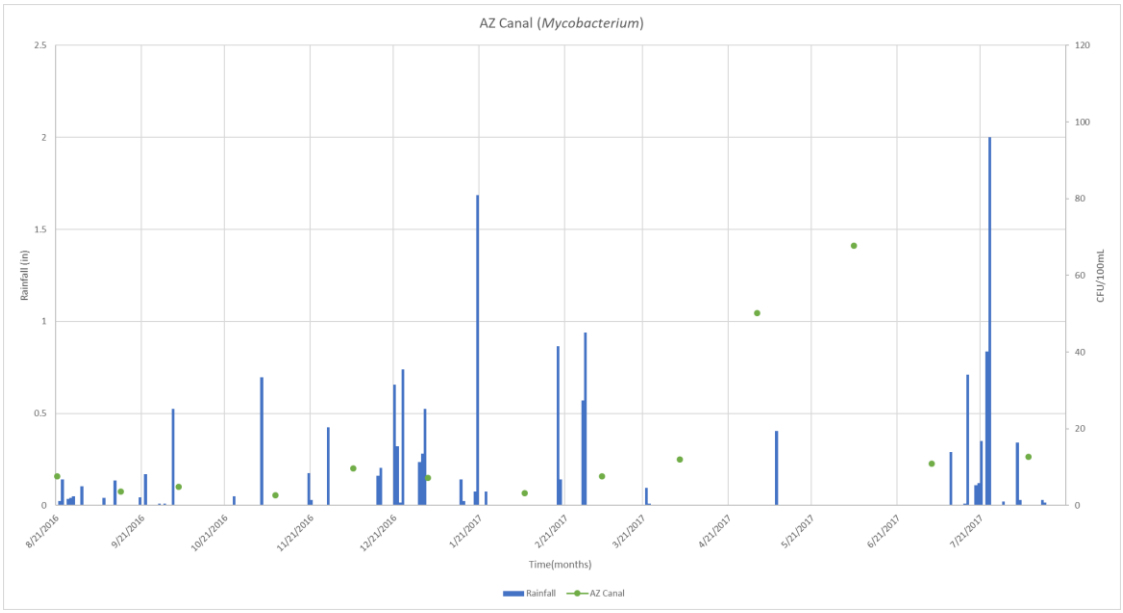


figure 4:Mycobacterium concentrations along az canal

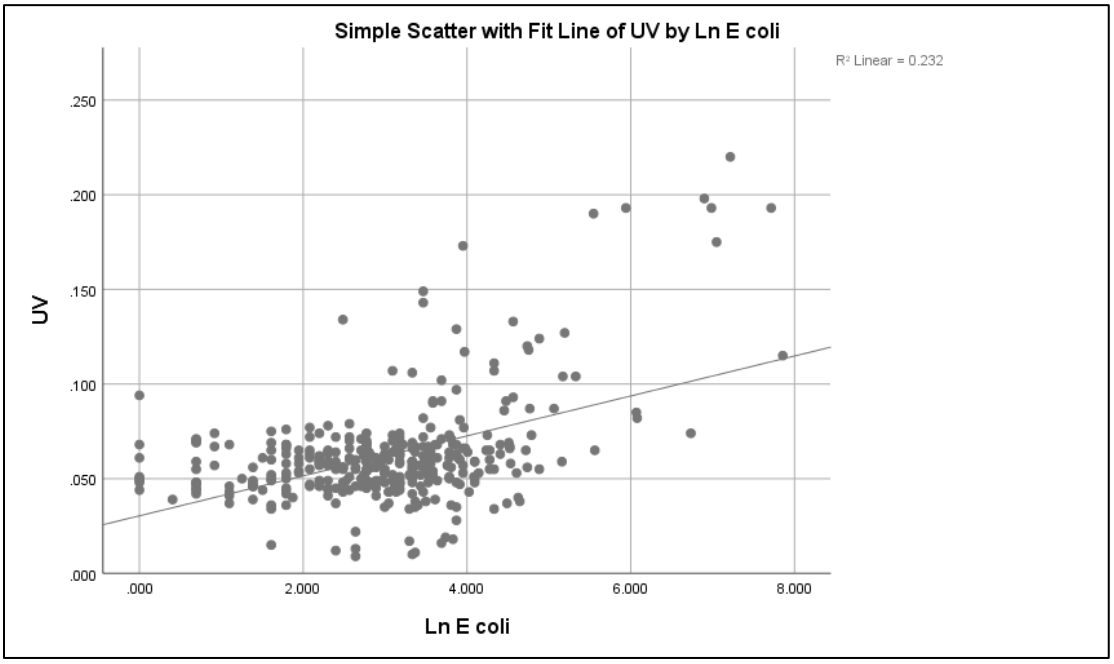


figure 5:UV-254 by ln E.*coli*

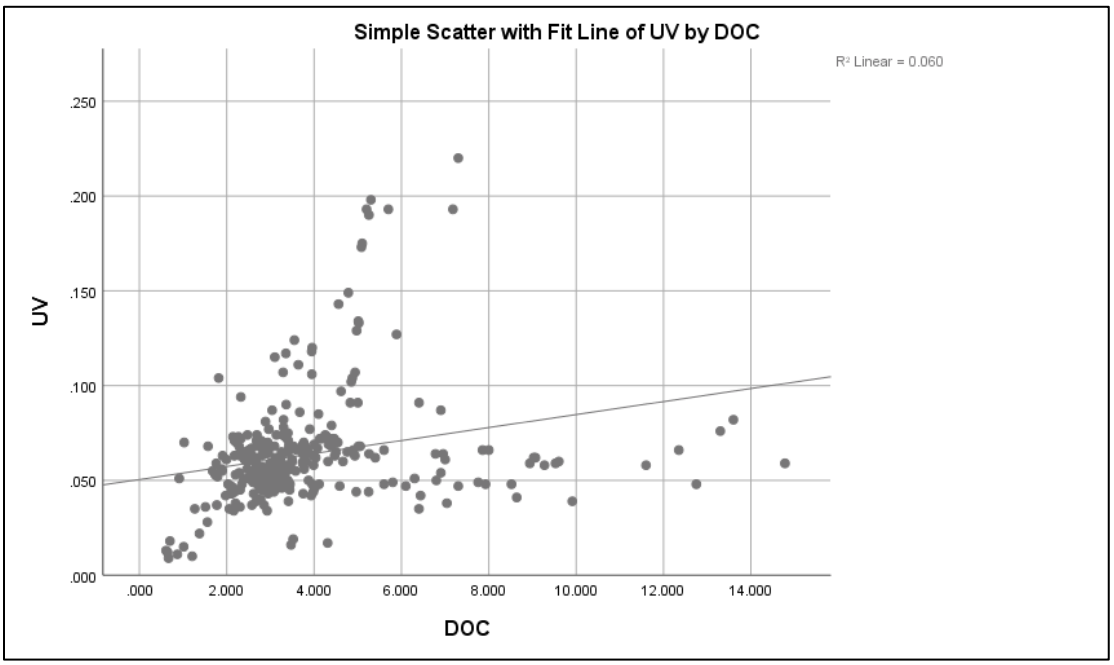


figure 6:UV-254 by DOC

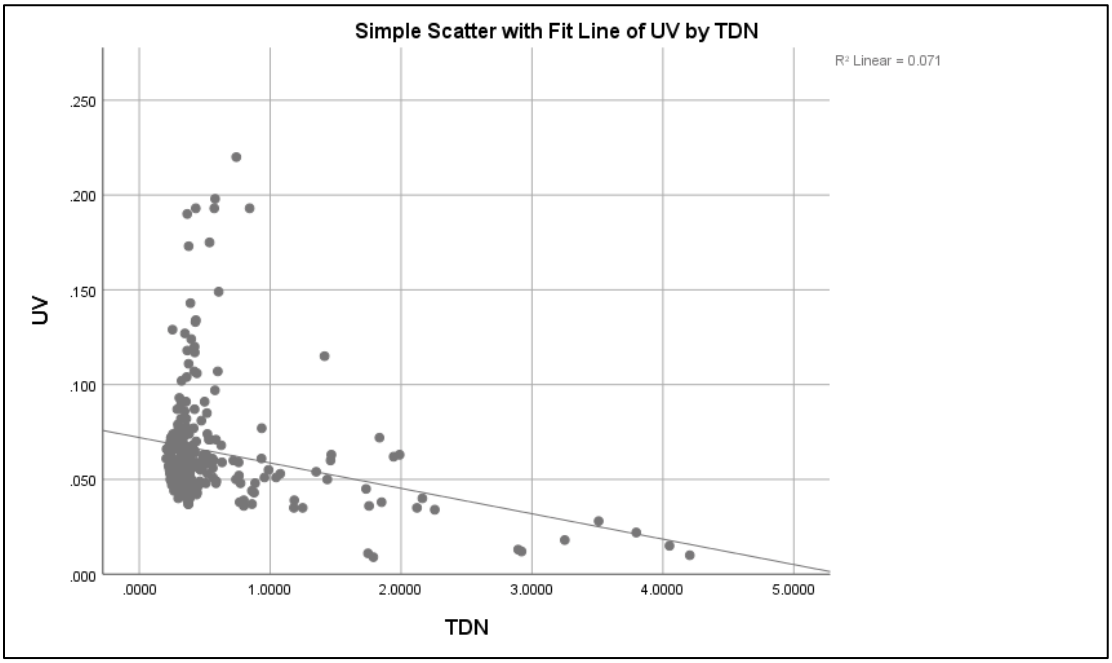


figure 7:UV-254 by TDN

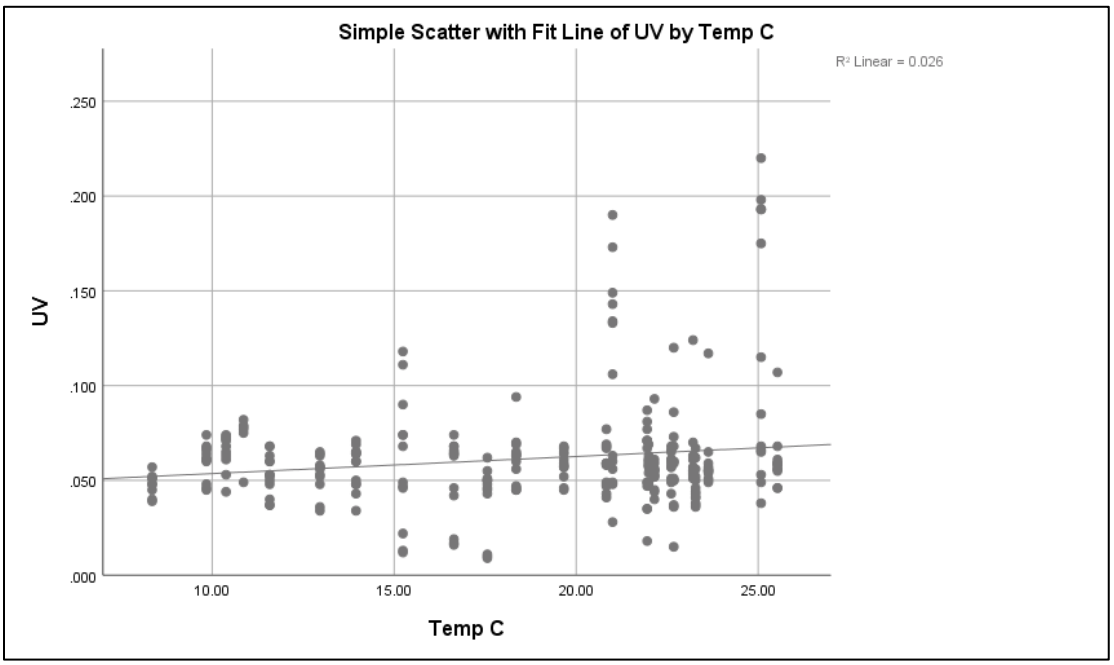


figure 8:UV by TEMP C

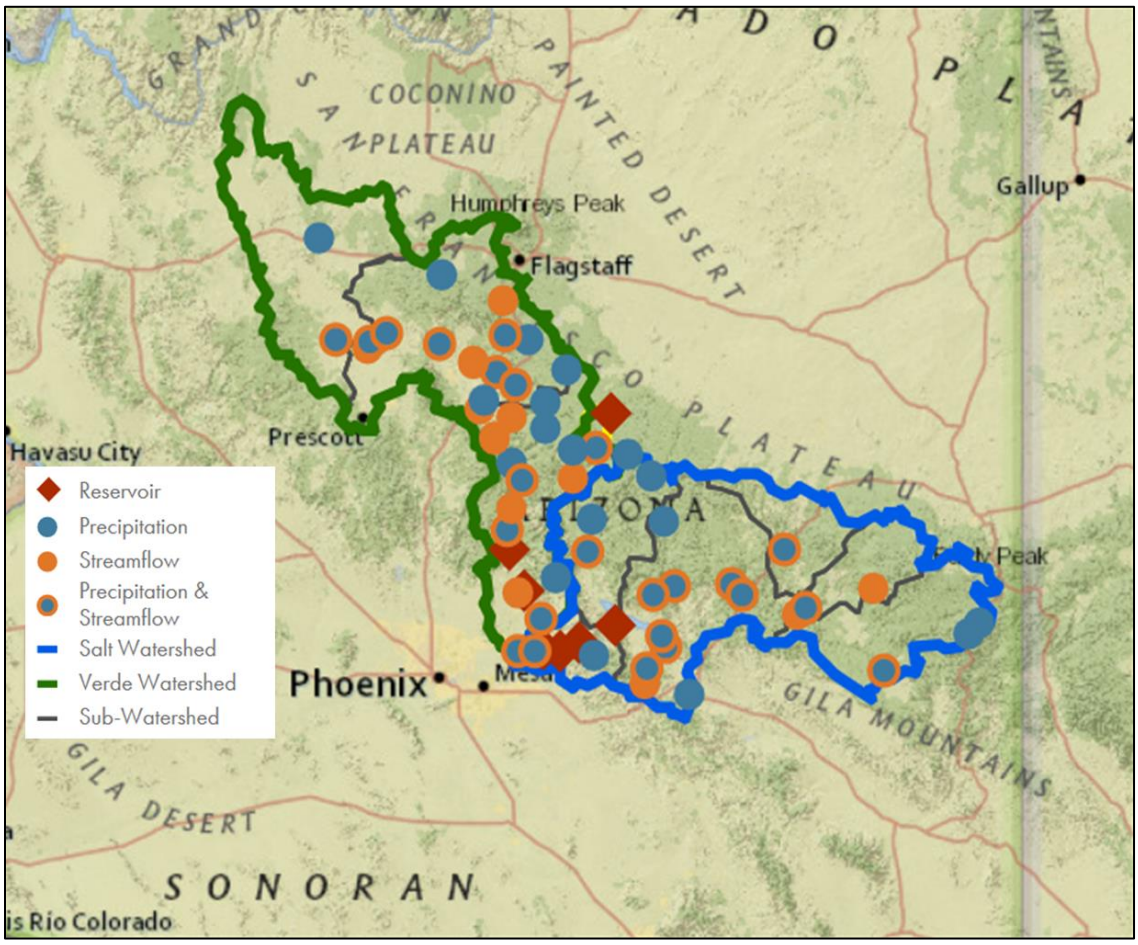


figure 9: SRP watershed map

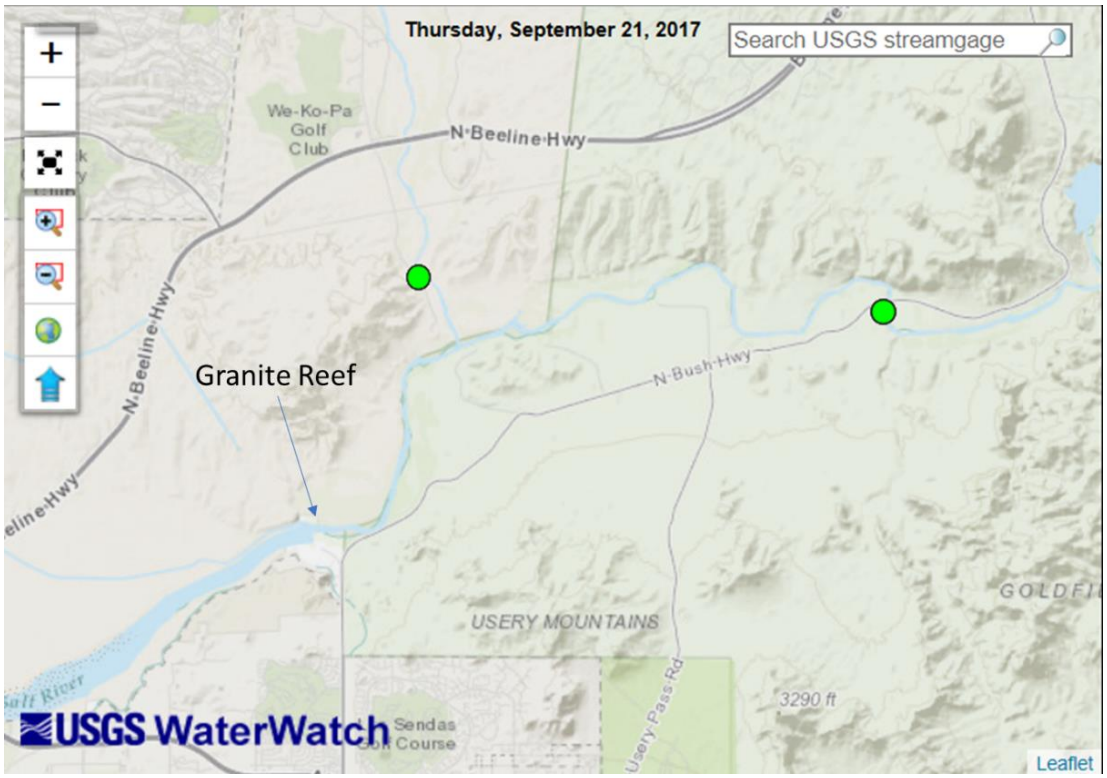


figure 10: rain gage locations

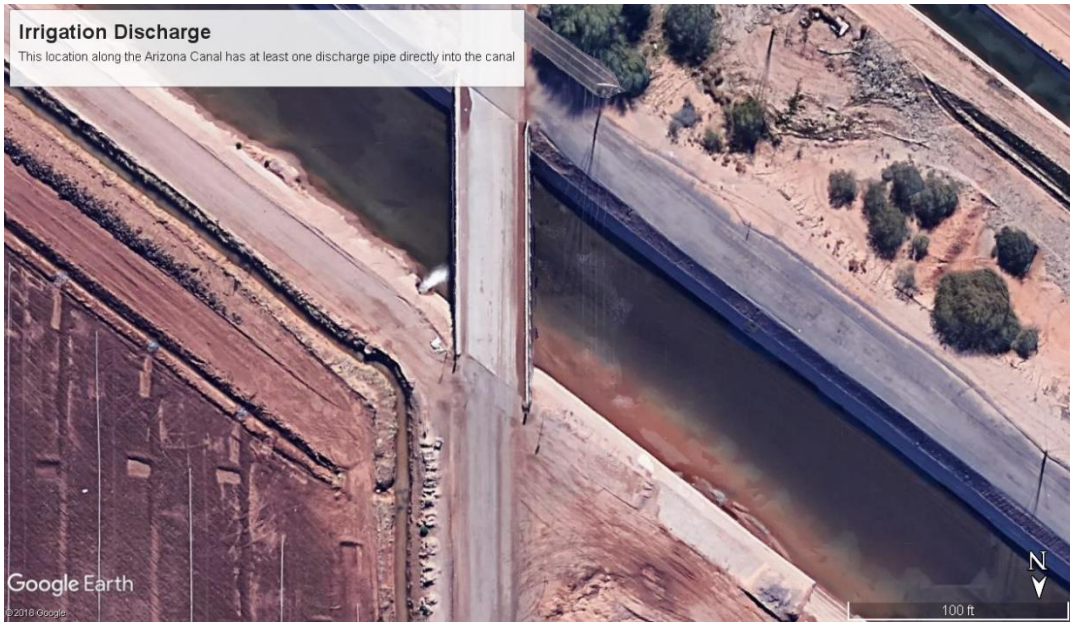


figure 11: discharge into az canal

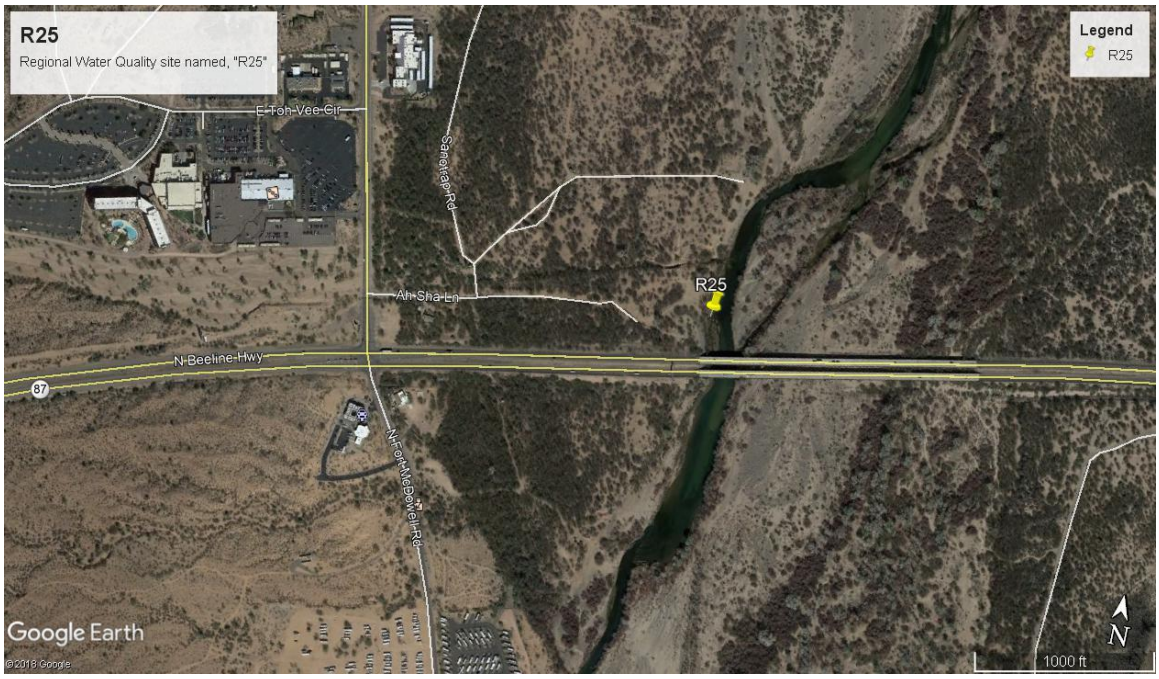


Figure 12: R25 location along the Verde River

Site Date	1/8/1018	2/5/1018	3/5/2018	4/2/2018	5/7/2018	6/4/2018	7/9/2018
R10	8	9.4	12.1	12.5	16.5	17.7	21.5
R11	Offline	Offline	11.3	Offline	21.5	23.3	20
SOCA	12.3	11	8.5	12.8	17.1	17	20.8
R12	Offline	10.5	11.3	13.3	21.4	23.1	21.3
R13	Offline	11.1	11.1	12.8	18.7	18.6	21.3
HWY 87	Offline	11	11.1	14	18.6	18.4	21.2
R25	11.7	9	9.9	15.9	21	21.7	26.2
GR(N)	Offline	Offline	Offline	22.6	23	27.6	31.3
GR(O)	Offline	Offline	Offline	22.3	23.2	27.7	30
GL(N)	Offline	Offline	Offline	19.7	26.3	25.9	29.5
GL(O)	Offline	Offline	Offline	19.9	25.9	25.6	28.3
AN(N)	12.6	14.1	12	25.1		19.7	23.8
AN(O)	13.3	13.9	15.1	NA	22.1	-	22
R3	11.1	10.2	9.8	15	20.6	13.3	14.9
UH(N)	15.5	15.8	15.9	21.1	24.3	28.6	28.9
UH(O)	13.3	14	14.1	20.1	22.3	26	22.8
CENT	12.3	14.1	14	17	22.4	24.8	26.2
24TH(N)	Offline	13.1	15	16.1	22.9	26	26.2
24TH(O)	Offline	Offline	16.6	19.6	23.2	25.7	27.5
56TH	Offline	19	14.3	16	21.8	24.4	26.5
PIMA	Offline	12.1	15.1	16.5	23.7	23.3	26.1
NP(N)	Offline	Offline	12.3	16.8	20.8	25	28.2
NP(O)	Offline	Offline	Offline	27.6	28	37.8	38.6
STP(N)	18.9	18.5	-	24.1	24.8	26.5	29.1
STP(O)	16.5	17.5	-	22.7	24.6	26	29.1
CH(N)	Offline	15	-	20.7	24.5	28.1	31.8
CH(O)	Offline	17	-	19.6	26.1	26.8	29
MOC	14.3	13.3	18.9	20.6	24.8	26.3	29.6
HOC	12.3	14.1	15.5	19.3	21.1	20.8	24.3
HTC	12.2	14	16	19.5	22.4	20.6	

Table 3: Temperature data used in the SPSS model

APPENDIX B

SPSS OUTPUT FILE: DOC

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT DOC
  /METHOD=ENTER LnEcoli TDN TempC UV
  /PARTIALPLOT ALL
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3)
  /SAVE PRED COOK LEVER SRESID SDRESID.
```

Regression

Notes

Output Created		18-OCT-2018 15:43:26
Comments		
Input	Data	C:\Users\Owner\Google Drive\DOC & UV254\SPSS\NEW\USE THIS MODEL _UPDATED.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any variable used.

Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT DOC /METHOD=ENTER LnEcoli TDN TempC UV /PARTIALPLOT ALL /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID) /CASEWISE PLOT(ZRESID) OUTLIERS(3) /SAVE PRED COOK LEVER SRESID SDRESID.
Resources	Processor Time	00:00:01.06
	Elapsed Time	00:00:00.94
	Memory Required	5072 bytes
	Additional Memory Required for Residual Plots	1368 bytes
Variables Created or Modified	PRE_5	Unstandardized Predicted Value
	SRE_5	Studentized Residual
	SDR_5	Studentized Deleted Residual
	COO_5	Cook's Distance
	LEV_5	Centered Leverage Value

Descriptive Statistics

	Mean	Std. Deviation	N
DOC	3.043520833333335	.983349420846436	144
Ln E coli	3.191573789175535	1.393359439555727	144
TDN	.677414	.7543728	144
Temp C	18.0419	6.13349	144
UV	.06699	.037550	144

Correlations

		DOC	Ln E coli	TDN	Temp C	UV
Pearson Correlation	DOC	1.000	.453	-.513	.319	.819
	Ln E coli	.453	1.000	.015	.309	.606
	TDN	-.513	.015	1.000	-.085	-.297
	Temp C	.319	.309	-.085	1.000	.235
	UV	.819	.606	-.297	.235	1.000
Sig. (1-tailed)	DOC	.	.000	.000	.000	.000
	Ln E coli	.000	.	.428	.000	.000
	TDN	.000	.428	.	.154	.000
	Temp C	.000	.000	.154	.	.002
	UV	.000	.000	.000	.002	.
N	DOC	144	144	144	144	144
	Ln E coli	144	144	144	144	144
	TDN	144	144	144	144	144
	Temp C	144	144	144	144	144
	UV	144	144	144	144	144

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	UV, Temp C, TDN, Ln E coli[b]	.	Enter

a. Dependent Variable: DOC

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.876[a]	.767	.760	.4819133965 16060	1.623

a. Predictors: (Constant), UV, Temp C, TDN, Ln E coli
b. Dependent Variable: DOC

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	105.996	4	26.499	114.102	.000[b]
	Residual	32.281	139	.232		
	Total	138.278	143			

a. Dependent Variable: DOC
b. Predictors: (Constant), UV, Temp C, TDN, Ln E coli

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.705	.149		11.460	.000
	Ln E coli	-.010	.039	-.014	-.250	.803
	TDN	-.378	.058	-.290	-6.513	.000
	Temp C	.021	.007	.131	3.028	.003
	UV	18.607	1.464	.711	12.713	.000

Coefficients[a]

Model		95.0% Confidence Interval for B		Correlations			Collinearity Statistics
		Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance
1	(Constant)	1.411	1.999				
	Ln E coli	-.086	.067	.453	-.021	-.010	.560
	TDN	-.493	-.263	-.513	-.484	-.267	.846
	Temp C	.007	.035	.319	.249	.124	.896
	UV	15.713	21.500	.819	.733	.521	.538

Coefficients[a]

Model		Collinearity Statistics
		VIF
1	(Constant)	
	Ln E coli	1.786
	TDN	1.182
	Temp C	1.116
	UV	1.860

a. Dependent Variable: DOC

Coefficient Correlations[a]

Model			UV	Temp C	TDN	Ln E coli
1	Correlations	UV	1.000	-.029	.382	-.616
		Temp C	-.029	1.000	.076	-.227
		TDN	.382	.076	1.000	-.267
		Ln E coli	-.616	-.227	-.267	1.000
	Covariances	UV	2.142	.000	.032	-.035
		Temp C	.000	4.820E-5	3.075E-5	-6.088E-5
		TDN	.032	3.075E-5	.003	-.001
		Ln E coli	-.035	-6.088E-5	-.001	.001

a. Dependent Variable: DOC

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	Ln E coli	TDN	Temp C
1	1	4.163	1.000	.00	.00	.01	.00
	2	.586	2.665	.00	.00	.65	.00
	3	.140	5.457	.08	.06	.14	.25
	4	.064	8.090	.08	.90	.07	.00
	5	.048	9.335	.84	.03	.13	.74

Collinearity Diagnostics[a]

Model	Dimension	Variance Proportions UV
1	1	.01
	2	.03
	3	.30
	4	.56
	5	.10

a. Dependent Variable: DOC

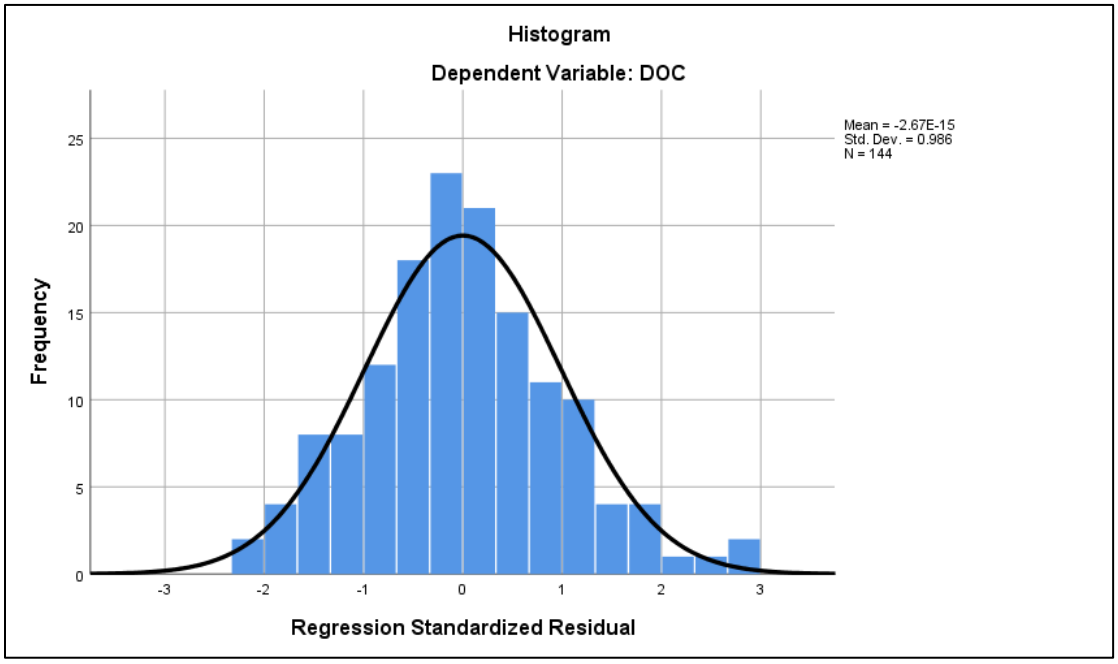
Residuals Statistics[a]

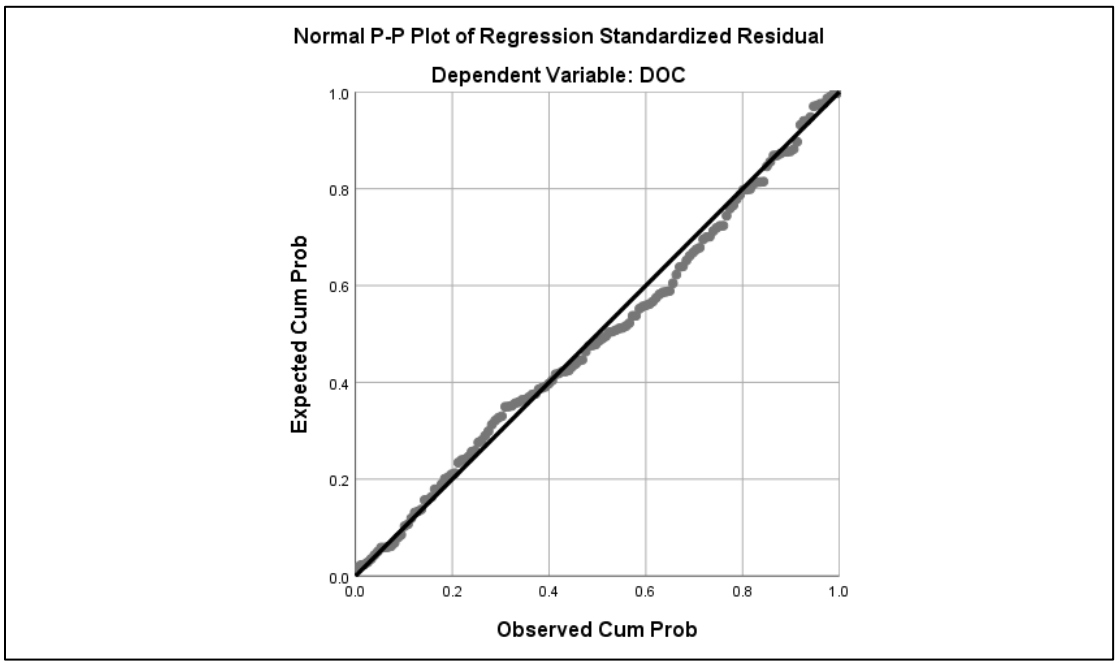
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.637379527091980	5.975498676300049	3.043520833333335	.860948198271977	144
Std. Predicted Value	-2.795	3.406	.000	1.000	144
Standard Error of Predicted Value	.050	.205	.085	.030	144
Adjusted Predicted Value	.527941942214966	5.766294002532959	3.042491595221673	.860285693333945	144
Residual	-1.005667090415955	1.330499053001404	-.000000000000001	.475125544852358	144

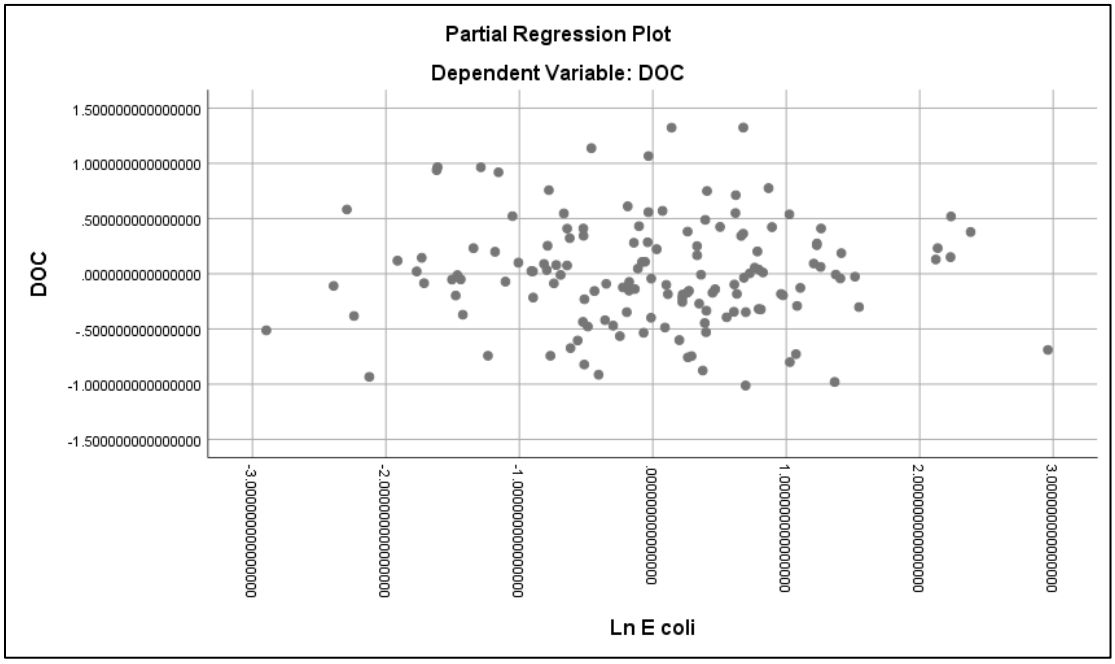
Std. Residual	-2.087	2.761	.000	.986	144
Stud. Residual	-2.110	2.958	.001	1.005	144
Deleted Residual	-1.027806162834168	1.533706068992615	.001029238111659	.494237519401751	144
Stud. Deleted Residual	-2.137	3.044	.002	1.013	144
Mahal. Distance	.528	24.760	3.972	4.159	144
Cook's Distance	.000	.276	.008	.025	144
Centered Leverage Value	.004	.173	.028	.029	144

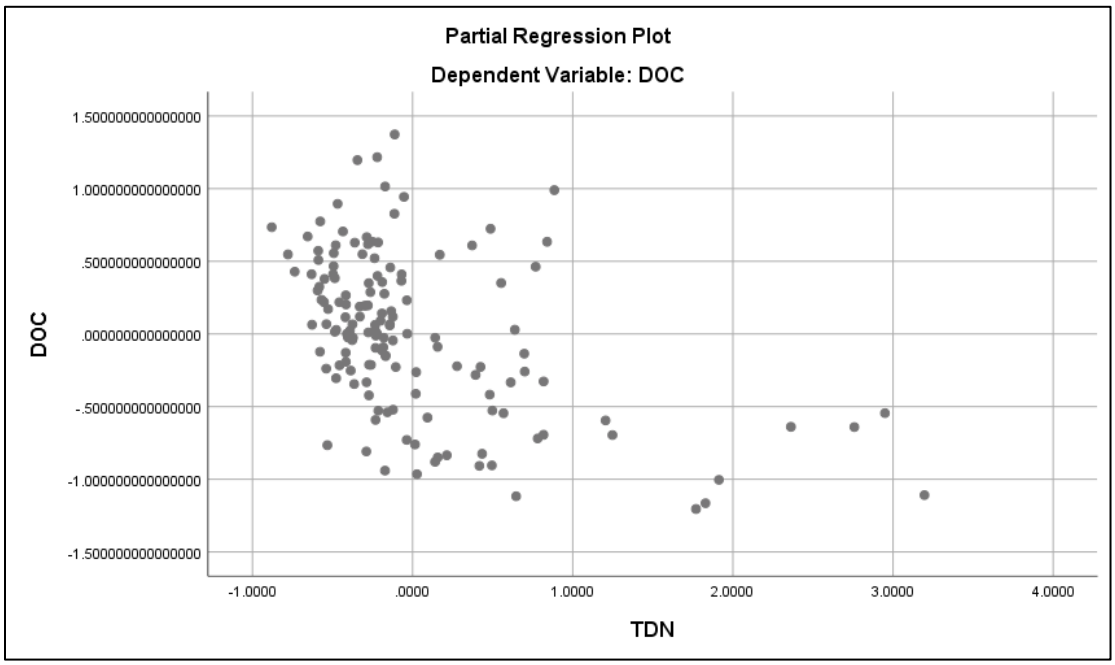
a. Dependent Variable: DOC

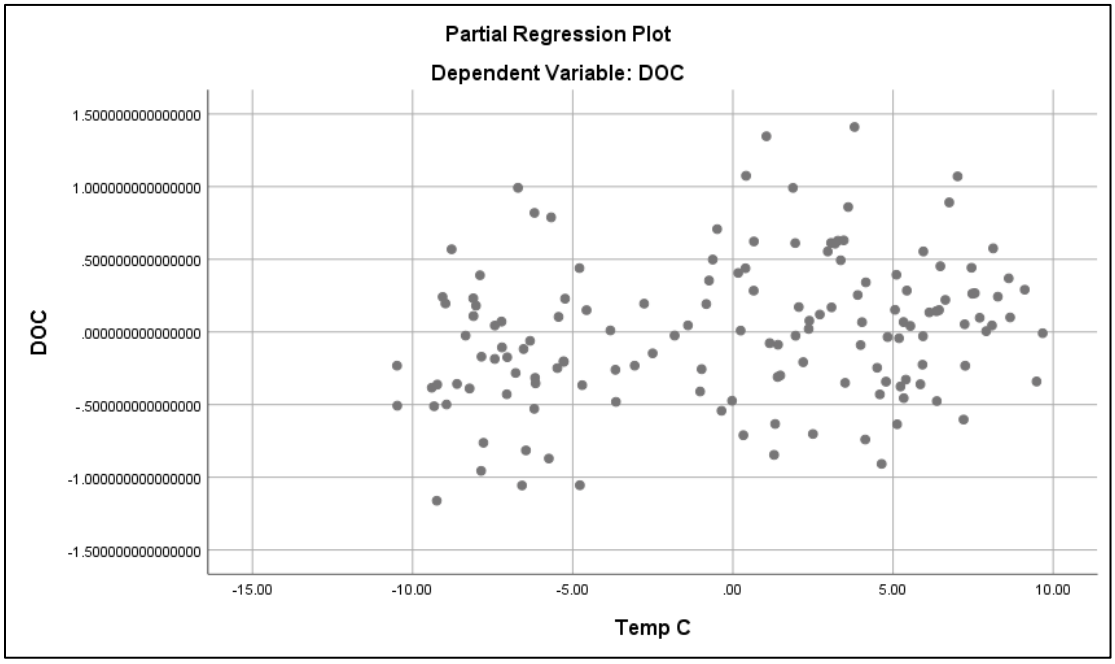
Charts

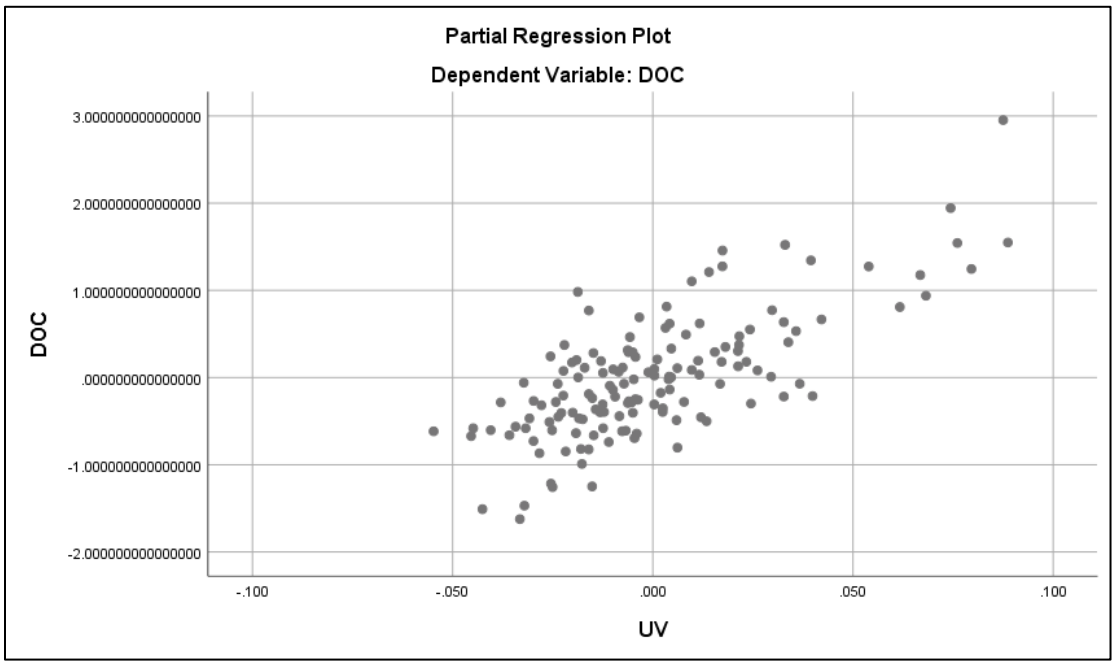



Normal P-P Plot of Regression Standardized Residual
Dependent Variable: DOC
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Partial Regression Plot
Dependent Variable: DOC
DOC
Ln E coli
1.500000000000000
1.000000000000000
.500000000000000
.000000000000000
-.500000000000000
-1.000000000000000
-1.500000000000000
-3.000000000000000
-2.000000000000000
-1.000000000000000
.000000000000000
1.000000000000000
2.000000000000000
3.000000000000000


Partial Regression Plot
Dependent Variable: DOC
DOC
1.500000000000000
1.000000000000000
.500000000000000
.000000000000000
-.500000000000000
-1.000000000000000
-1.500000000000000
-1.0000
.0000
1.0000
2.0000
3.0000
4.0000
TDN


Partial Regression Plot
Dependent Variable: DOC
DOC
1.500000000000000
1.000000000000000
.500000000000000
.000000000000000
-.500000000000000
-1.000000000000000
-1.500000000000000
-15.00
-10.00
-5.00
.00
5.00
10.00
Temp C


Partial Regression Plot
Dependent Variable: DOC
3.000000000000000
2.000000000000000
1.000000000000000
.000000000000000
-1.000000000000000
-2.000000000000000
DOC
-.100
-.050
.000
.050
.100
UV

APPENDIX C

SPSS OUTPUT FILE: TDN

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT TDN
  /METHOD=ENTER TempC UV DOC LnEcoli
  /PARTIALPLOT ALL
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3)
  /SAVE PRED COOK LEVER SRESID SDRESID.
```

Regression

Notes

Output Created		18-OCT-2018 15:45:02
Comments		
Input	Data	C:\Users\Owner\Google Drive\DOC & UV254\SPSS\NEW\USE THIS MODEL _UPDATED.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.

	Cases Used	Statistics are based on cases with no missing values for any variable used.
Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT TDN /METHOD=ENTER TempC UV DOC LnEcoli /PARTIALPLOT ALL /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID) /CASEWISE PLOT(ZRESID) OUTLIERS(3) /SAVE PRED COOK LEVER SRESID SDRESID.
Resources	Processor Time	00:00:01.16
	Elapsed Time	00:00:00.86
	Memory Required	5472 bytes
	Additional Memory Required for Residual Plots	1368 bytes

Variables Created or Modified	PRE_7	Unstandardized Predicted Value
	SRE_7	Studentized Residual
	SDR_7	Studentized Deleted Residual
	COO_7	Cook's Distance
	LEV_7	Centered Leverage Value

Descriptive Statistics

	Mean	Std. Deviation	N
TDN	.677414	.7543728	144
Temp C	18.0419	6.13349	144
UV	.06699	.037550	144
DOC	3.043520833333335	.9833494208 46436	144
Ln E coli	3.191573789175535	1.393359439 555727	144

Correlations

		TDN	Temp C	UV	DOC	Ln E coli
Pearson Correlation	TDN	1.000	-.085	-.297	-.513	.015
	Temp C	-.085	1.000	.235	.319	.309
	UV	-.297	.235	1.000	.819	.606
	DOC	-.513	.319	.819	1.000	.453
	Ln E coli	.015	.309	.606	.453	1.000
Sig. (1-tailed)	TDN	.	.154	.000	.000	.428
	Temp C	.154	.	.002	.000	.000
	UV	.000	.002	.	.000	.000
	DOC	.000	.000	.000	.	.000
	Ln E coli	.428	.000	.000	.000	.
N	TDN	144	144	144	144	144
	Temp C	144	144	144	144	144

	UV	144	144	144	144	144
	DOC	144	144	144	144	144
	Ln E coli	144	144	144	144	144

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Ln E coli, Temp C, DOC, UV[b]	.	Enter

a. Dependent Variable: TDN

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.593[a]	.352	.333	.6160664	1.495

a. Predictors: (Constant), Ln E coli, Temp C, DOC, UV

b. Dependent Variable: TDN

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28.622	4	7.156	18.853	.000[b]
	Residual	52.756	139	.380		
	Total	81.378	143			

a. Dependent Variable: TDN

b. Predictors: (Constant), Ln E coli, Temp C, DOC, UV

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1.758	.219		8.013	.000
	Temp C	.006	.009	.049	.657	.512
	UV	4.124	2.729	.205	1.511	.133
	DOC	-.618	.095	-.806	-6.513	.000
	Ln E coli	.130	.048	.241	2.707	.008

Coefficients[a]

Model		95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance
1	(Constant)	1.324	2.192				
	Temp C	-.012	.024	-.085	.056	.045	.843
	UV	-1.272	9.520	-.297	.127	.103	.253
	DOC	-.806	-.431	-.513	-.484	-.445	.305
	Ln E coli	.035	.226	.015	.224	.185	.589

Coefficients[a]

Model		Collinearity Statistics VIF
1	(Constant)	
	Temp C	1.186
	UV	3.957
	DOC	3.282
	Ln E coli	1.697

a. Dependent Variable: TDN

Coefficient Correlations[a]

Model			Ln E coli	Temp C	DOC	UV
1	Correlations	Ln E coli	1.000	-.244	.152	-.479
		Temp C	-.244	1.000	-.254	.157

		DOC	.152	-.254	1.000	-.774
		UV	-.479	.157	-.774	1.000
	Covariances	Ln E coli	.002	.000	.001	-.063
		Temp C	.000	8.370E-5	.000	.004
		DOC	.001	.000	.009	-.200
		UV	-.063	.004	-.200	7.449

a. Dependent Variable: TDN

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	Temp C	UV	DOC
1	1	4.678	1.000	.00	.00	.00	.00
	2	.166	5.307	.06	.15	.18	.00
	3	.083	7.490	.01	.00	.05	.06
	4	.054	9.283	.43	.82	.03	.01
	5	.018	16.096	.49	.02	.74	.93

Collinearity Diagnostics[a]

Model	Dimension	Variance Proportions Ln E coli
1	1	.00
	2	.01
	3	.84
	4	.01
	5	.14

a. Dependent Variable: TDN

Casewise Diagnostics[a]

Case Number	Std. Residual	TDN	Predicted Value	Residual
36	4.075	4.0500	1.539758	2.5102424
41	4.241	4.2050	1.592510	2.6124902

42	3.193	3.5090	1.541958	1.9670421
49	3.834	3.7960	1.433812	2.3621885

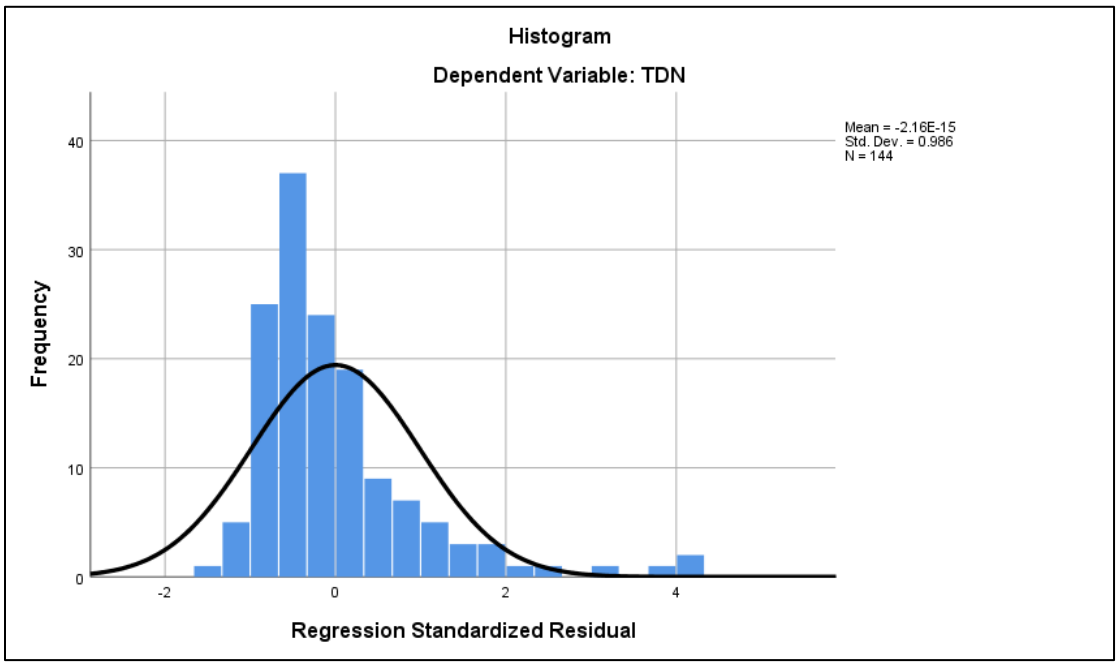
a. Dependent Variable: TDN

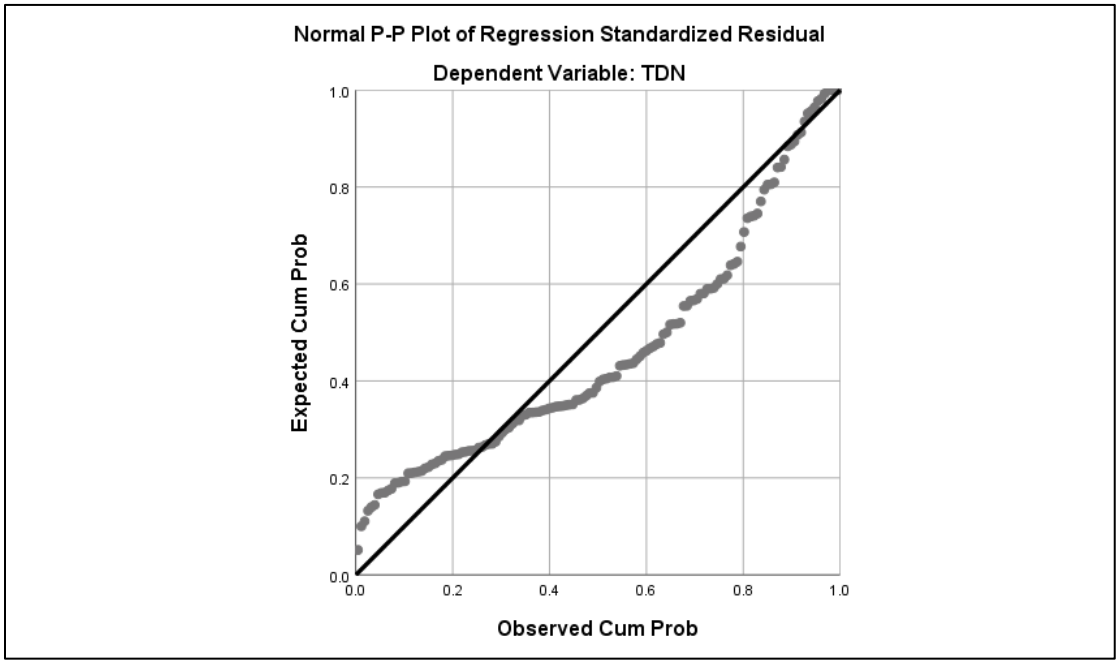
Residuals Statistics[a]

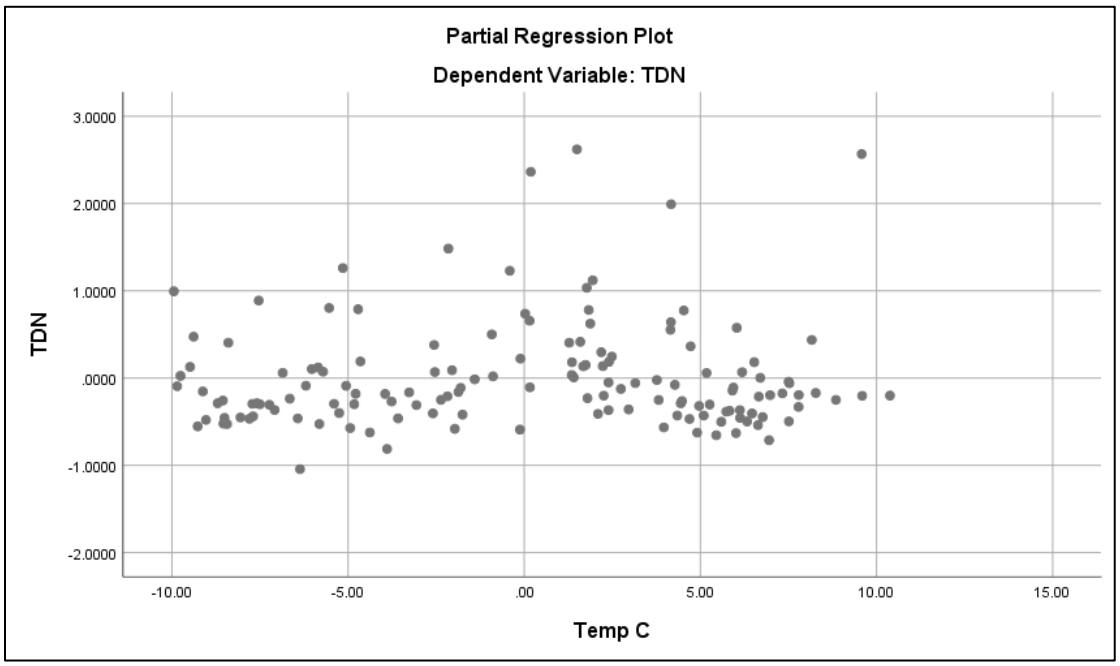
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-.755762	1.957893	.677414	.4473891	144
Std. Predicted Value	-3.203	2.862	.000	1.000	144
Standard Error of Predicted Value	.057	.237	.110	.032	144
Adjusted Predicted Value	-1.016347	1.854156	.671781	.4522408	144
Residual	-1.0055923	2.6124902	.0000000	.6073889	144
Std. Residual	-1.632	4.241	.000	.986	144
Stud. Residual	-1.668	4.332	.004	1.009	144
Deleted Residual	-1.0501482	2.7259972	.0056327	.6368607	144
Stud. Deleted Residual	-1.679	4.641	.012	1.038	144
Mahal. Distance	.242	20.211	3.972	3.181	144
Cook's Distance	.000	.241	.010	.033	144
Centered Leverage Value	.002	.141	.028	.022	144

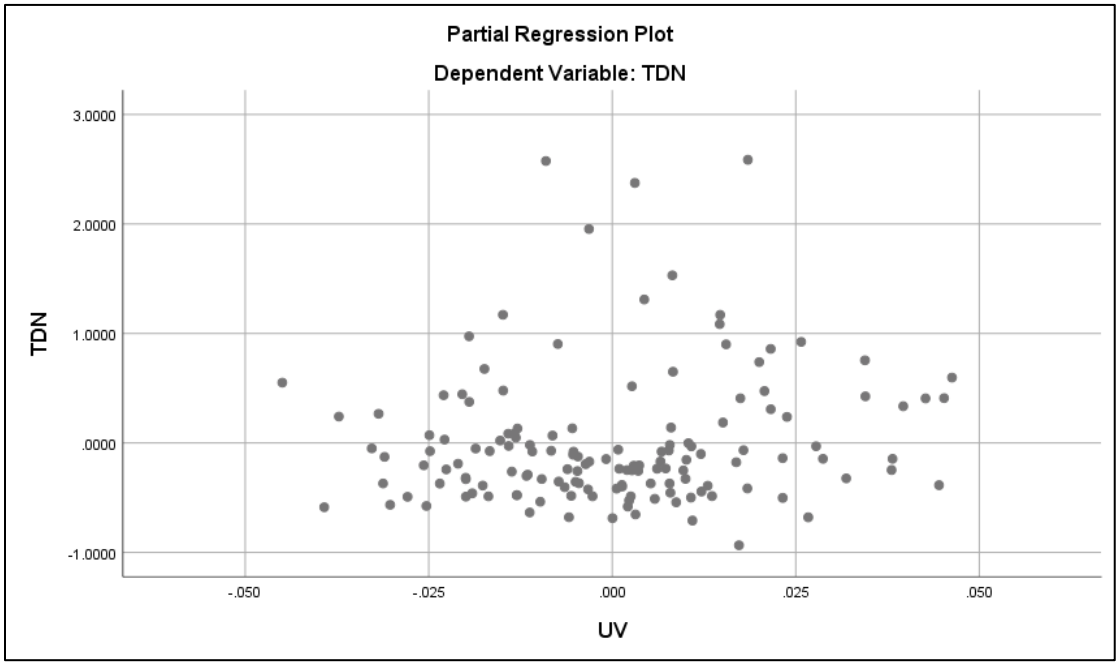
a. Dependent Variable: TDN

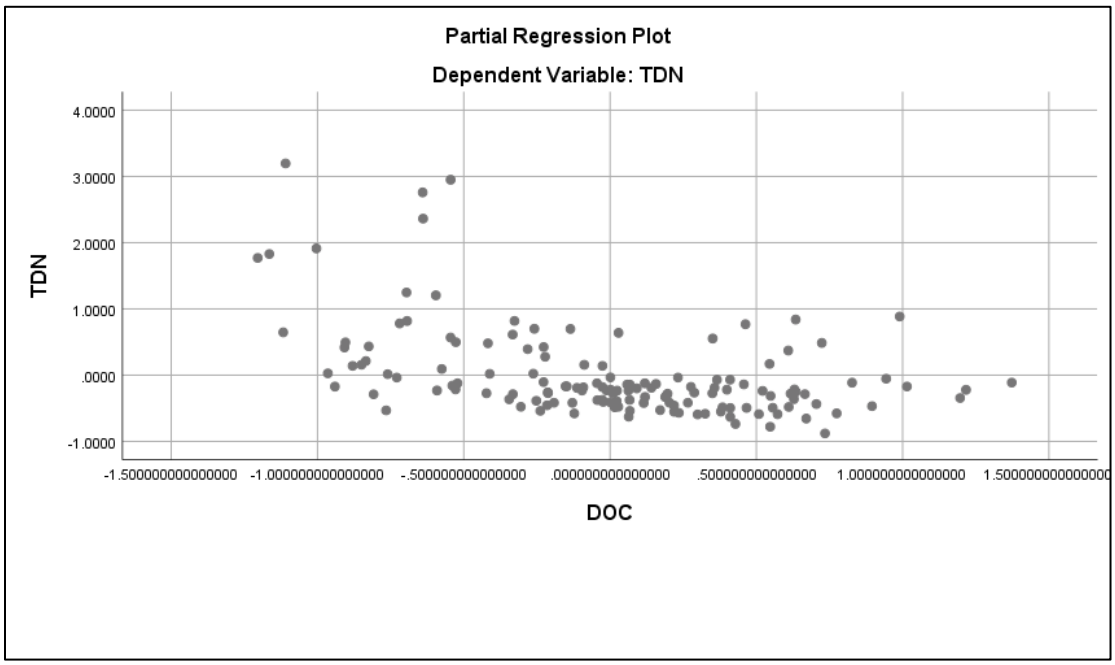
Charts

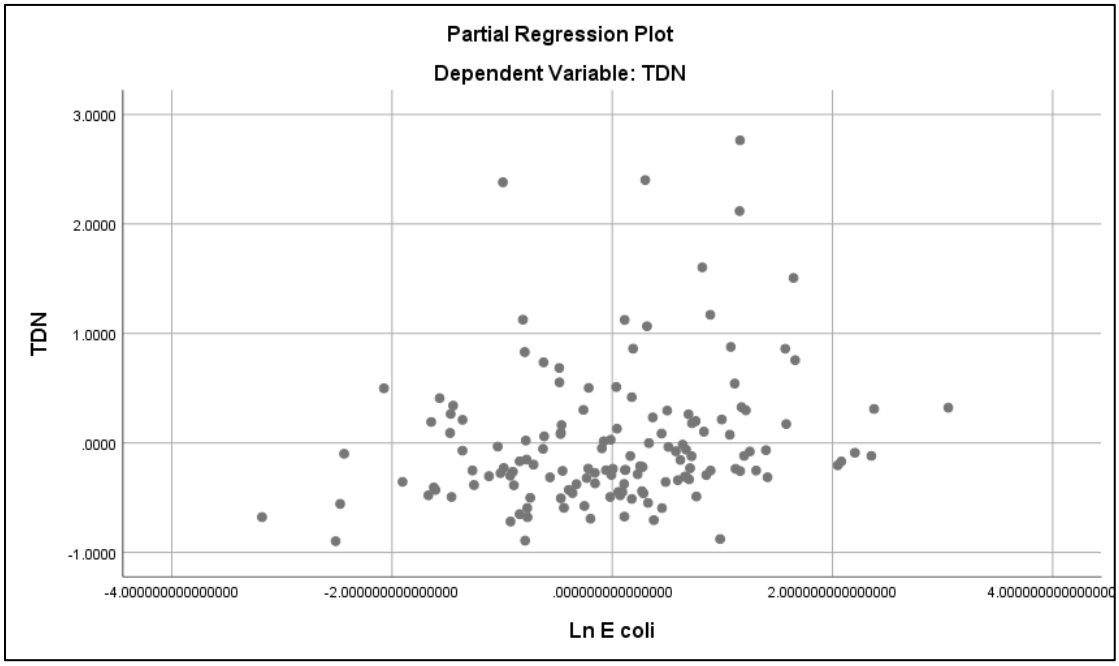

Histogram
Dependent Variable: TDN
Mean = -2.16E-15
Std. Dev. = 0.986
N = 144
Frequency
40
30
20
10
0
-2
0
2
4
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: TDN
Expected Cum Prob
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob


Partial Regression Plot
Dependent Variable: TDN
TDN
3.0000
2.0000
1.0000
.0000
-1.0000
-2.0000
-10.00
-5.00
.00
5.00
10.00
15.00
Temp C


Partial Regression Plot
Dependent Variable: TDN
TDN
3.0000
2.0000
1.0000
.0000
-1.0000
-.050
-.025
.000
.025
.050
UV


Partial Regression Plot
Dependent Variable: TDN
TDN
4.0000
3.0000
2.0000
1.0000
.0000
-1.0000
-1.500000000000000
-1.000000000000000
-.500000000000000
.000000000000000
.500000000000000
1.000000000000000
1.500000000000000
DOC


Partial Regression Plot
Dependent Variable: TDN
TDN
3.0000
2.0000
1.0000
.0000
-1.0000
-4.000000000000000
-2.000000000000000
.000000000000000
2.000000000000000
4.000000000000000
Ln E coli

APPENDIX D

SPSS OUTPUT FILE: UV-254

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT UV
  /METHOD=ENTER LnEcoli TDN TempC DOC
  /PARTIALPLOT ALL
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3)
  /SAVE PRED COOK LEVER SRESID SDRESID.
```

Regression

Notes

Output Created		18-OCT-2018 15:42:44
Comments		
Input	Data	C:\Users\Owner\Google Drive\DOC & UV254\SPSS\NEW\USE THIS MODEL _UPDATED.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any variable used.

Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT UV /METHOD=ENTER LnEcoli TDN TempC DOC /PARTIALPLOT ALL /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID) /CASEWISE PLOT(ZRESID) OUTLIERS(3) /SAVE PRED COOK LEVER SRESID SDRESID.
Resources	Processor Time	00:00:01.11
	Elapsed Time	00:00:00.88
	Memory Required	4864 bytes
	Additional Memory Required for Residual Plots	1368 bytes
Variables Created or Modified	PRE_4	Unstandardized Predicted Value
	SRE_4	Studentized Residual

	SDR_4	Studentized Deleted Residual
	COO_4	Cook's Distance
	LEV_4	Centered Leverage Value

Descriptive Statistics

	Mean	Std. Deviation	N
UV	.06699	.037550	144
Ln E coli	3.191573789175535	1.393359439555727	144
TDN	.677414	.7543728	144
Temp C	18.0419	6.13349	144
DOC	3.043520833333335	.9833494208 46436	144

Correlations

		UV	Ln E coli	TDN	Temp C	DOC
Pearson Correlation	UV	1.000	.606	-.297	.235	.819
	Ln E coli	.606	1.000	.015	.309	.453
	TDN	-.297	.015	1.000	-.085	-.513
	Temp C	.235	.309	-.085	1.000	.319
	DOC	.819	.453	-.513	.319	1.000
Sig. (1-tailed)	UV	.	.000	.000	.002	.000
	Ln E coli	.000	.	.428	.000	.000
	TDN	.000	.428	.	.154	.000
	Temp C	.002	.000	.154	.	.000
	DOC	.000	.000	.000	.000	.
N	UV	144	144	144	144	144
	Ln E coli	144	144	144	144	144
	TDN	144	144	144	144	144
	Temp C	144	144	144	144	144
	DOC	144	144	144	144	144

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	DOC, Temp C, Ln E coli, TDN[b]	.	Enter

a. Dependent Variable: UV
b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.867[a]	.751	.744	.018991	1.520

a. Predictors: (Constant), DOC, Temp C, Ln E coli, TDN
b. Dependent Variable: UV

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.151	4	.038	105.018	.000[b]
	Residual	.050	139	.000		
	Total	.202	143			

a. Dependent Variable: UV
b. Predictors: (Constant), DOC, Temp C, Ln E coli, TDN

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.039	.007		-5.173	.000
	Ln E coli	.008	.001	.290	5.689	.000

	TDN	.004	.003	.079	1.511	.133
	Temp C	-.001	.000	-.089	-1.947	.054
	DOC	.029	.002	.757	12.713	.000

Coefficients[a]

Model		95.0% Confidence Interval for B		Correlations			Collinearity Statistics
		Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance
1	(Constant)	-.054	-.024				
	Ln E coli	.005	.011	.606	.435	.241	.690
	TDN	-.001	.009	-.297	.127	.064	.659
	Temp C	-.001	.000	.235	-.163	-.082	.863
	DOC	.024	.033	.819	.733	.538	.505

Coefficients[a]

Model		Collinearity Statistics VIF
1	(Constant)	
	Ln E coli	1.449
	TDN	1.518
	Temp C	1.158
	DOC	1.981

a. Dependent Variable: UV

Coefficient Correlations[a]

Model			DOC	Temp C	Ln E coli	TDN
1	Correlations	DOC	1.000	-.193	-.486	.579
		Temp C	-.193	1.000	-.173	-.036
		Ln E coli	-.486	-.173	1.000	-.312
		TDN	.579	-.036	-.312	1.000
	Covariances	DOC	5.166E-6	-1.220E-7	-1.515E-6	3.411E-6
		Temp C	-1.220E-7	7.766E-8	-6.633E-8	-2.577E-8
		Ln E coli	-1.515E-6	-6.633E-8	1.883E-6	-1.112E-6
		TDN	3.411E-6	-2.577E-8	-1.112E-6	6.726E-6

a. Dependent Variable: UV

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	Ln E coli	TDN	Temp C
1	1	4.245	1.000	.00	.01	.01	.00
	2	.568	2.733	.00	.00	.54	.00
	3	.099	6.533	.04	.75	.00	.20
	4	.061	8.330	.14	.10	.01	.76
	5	.026	12.732	.81	.15	.44	.03

Collinearity Diagnostics[a]

Model	Dimension	Variance Proportions DOC
1	1	.00
	2	.01
	3	.00
	4	.18
	5	.81

a. Dependent Variable: UV

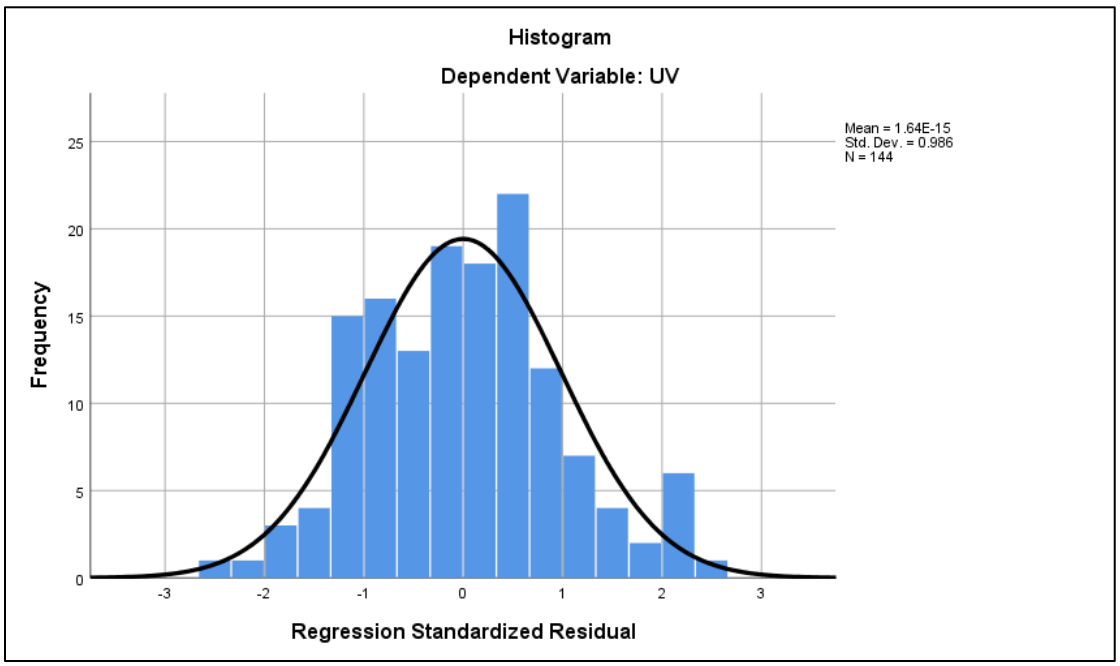
Residuals Statistics[a]

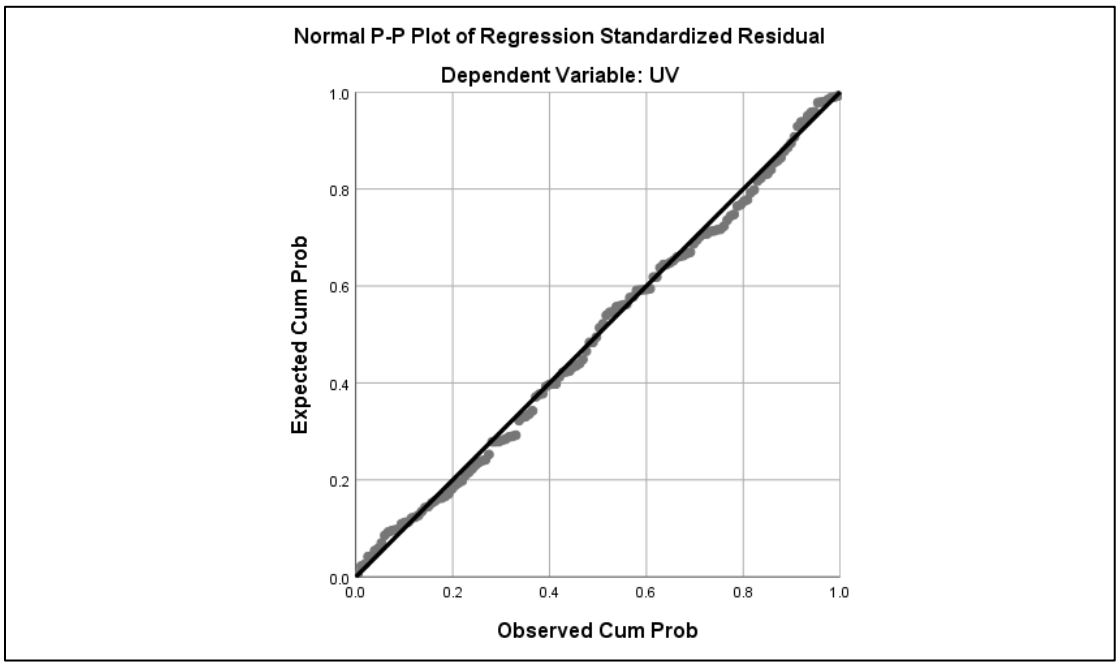
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-.00138	.21781	.06699	.032549	144
Std. Predicted Value	-2.100	4.634	.000	1.000	144
Standard Error of Predicted Value	.002	.008	.003	.001	144
Adjusted Predicted Value	-.00198	.21730	.06690	.032407	144
Residual	-.047108	.046027	.000000	.018723	144
Std. Residual	-2.481	2.424	.000	.986	144
Stud. Residual	-2.513	2.468	.002	1.004	144

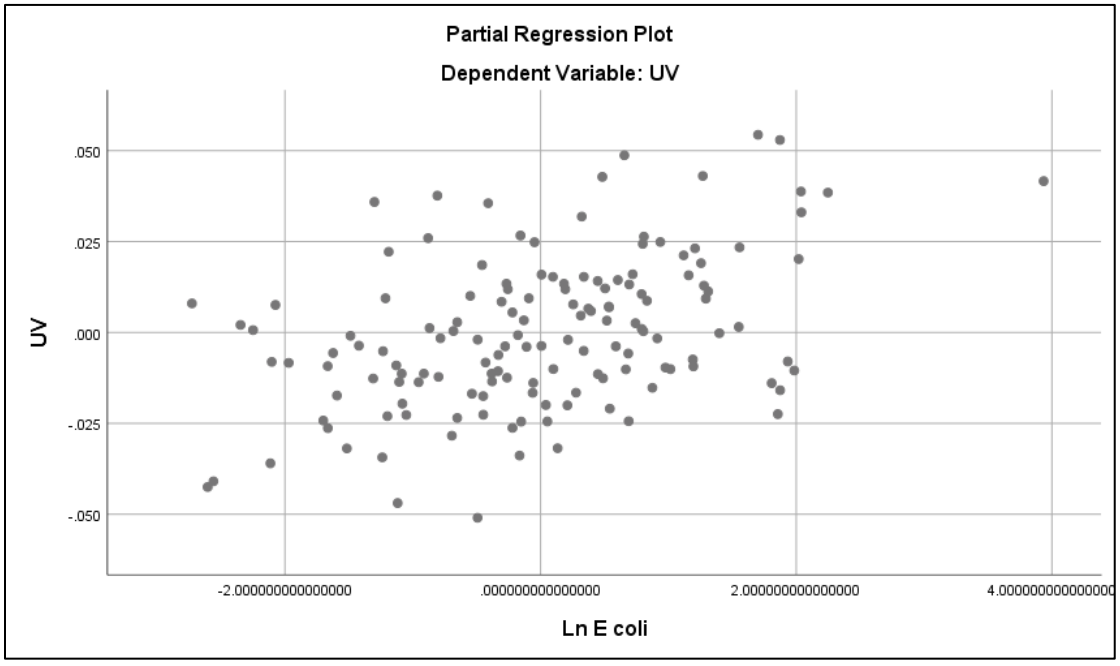
Deleted Residual	-.048339	.047745	.000087	.019414	144
Stud. Deleted Residual	-2.563	2.515	.003	1.010	144
Mahal. Distance	.200	26.270	3.972	4.268	144
Cook's Distance	.000	.076	.007	.013	144
Centered Leverage Value	.001	.184	.028	.030	144

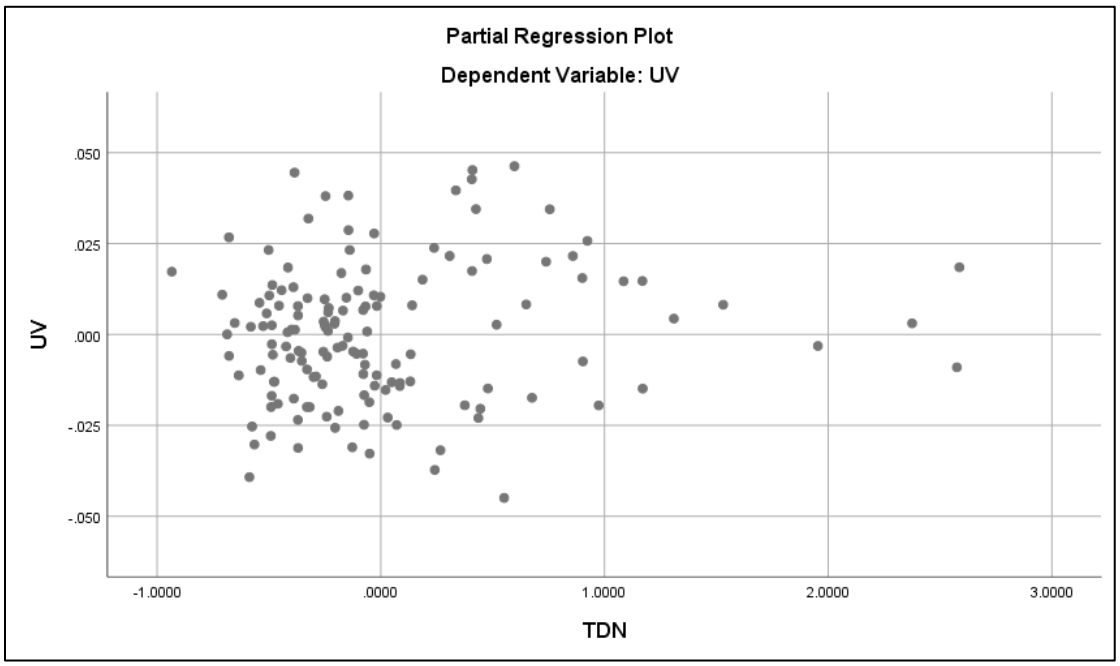
a. Dependent Variable: UV

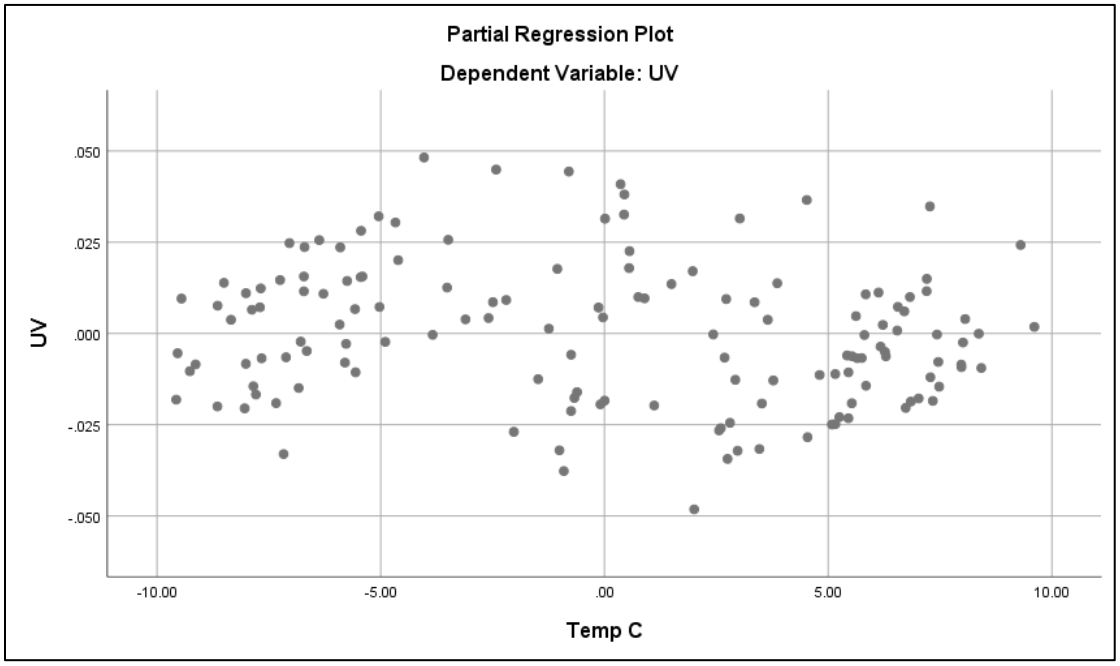
Charts

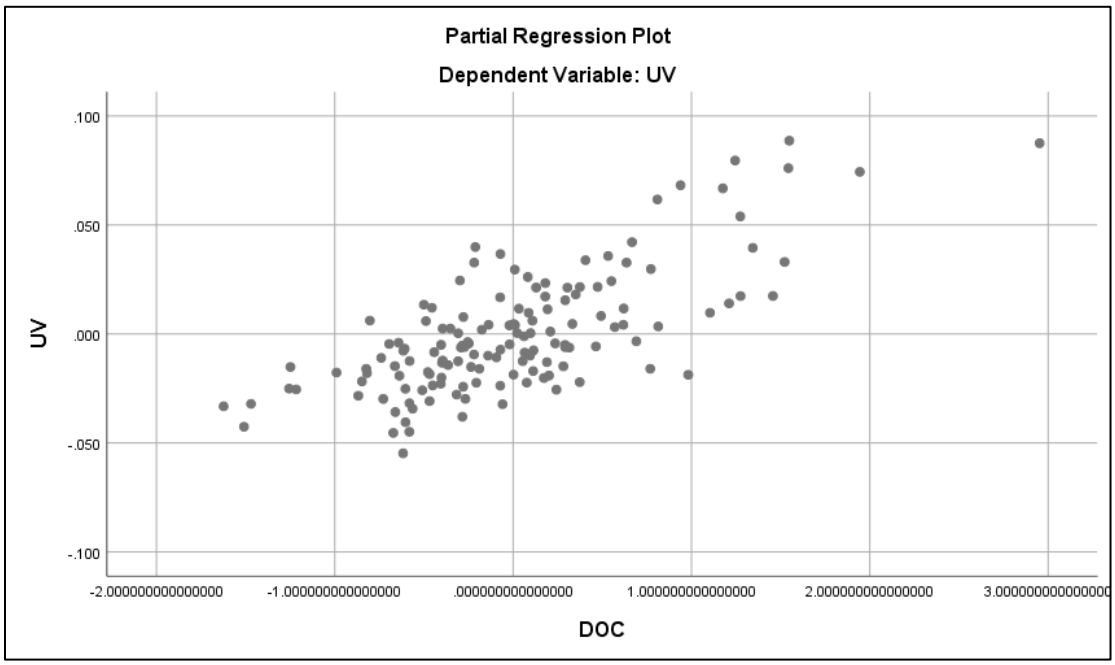



Normal P-P Plot of Regression Standardized Residual
Dependent Variable: UV
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Partial Regression Plot
Dependent Variable: UV
UV
Ln E coli
.050
.025
.000
-.025
-.050
-2.000000000000000
.000000000000000
2.000000000000000
4.000000000000000


Partial Regression Plot
Dependent Variable: UV
UV
.050
.025
.000
-.025
-.050
-1.0000
.0000
1.0000
2.0000
3.0000
TDN


Partial Regression Plot
Dependent Variable: UV
UV
.050
.025
.000
-.025
-.050
-10.00
-5.00
.00
5.00
10.00
Temp C


Partial Regression Plot
Dependent Variable: UV
.100
.050
.000
-.050
-.100
UV
-2.000000000000000
-1.000000000000000
.000000000000000
1.000000000000000
2.000000000000000
3.000000000000000
DOC

APPENDIX E

SPSS OUTPUT FILE: LN E. *coli*

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT LnEcoli
  /METHOD=ENTER TDN TempC UV DOC
  /PARTIALPLOT ALL
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3)
  /SAVE PRED COOK LEVER SRESID SDRESID.
```

Regression

Notes

Output Created		18-OCT-2018 15:44:19
Comments		
Input	Data	C:\Users\Owner\Google Drive\DOC & UV254\SPSS\NEW\USE THIS MODEL _UPDATED.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any variable used.

Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL ZPP /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT LnEcoli /METHOD=ENTER TDN TempC UV DOC /PARTIALPLOT ALL /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID) /CASEWISE PLOT(ZRESID) OUTLIERS(3) /SAVE PRED COOK LEVER SRESID SDRESID.
Resources	Processor Time	00:00:01.25
	Elapsed Time	00:00:00.83
	Memory Required	5264 bytes
	Additional Memory Required for Residual Plots	1368 bytes
Variables Created or Modified	PRE_6	Unstandardized Predicted Value
	SRE_6	Studentized Residual

	SDR_6	Studentized Deleted Residual
	COO_6	Cook's Distance
	LEV_6	Centered Leverage Value

Descriptive Statistics

	Mean	Std. Deviation	N
Ln E coli	3.191573789175535	1.393359439555727	144
TDN	.677414	.7543728	144
Temp C	18.0419	6.13349	144
UV	.06699	.037550	144
DOC	3.043520833333335	.9833494208 46436	144

Correlations

		Ln E coli	TDN	Temp C	UV	DOC
Pearson Correlation	Ln E coli	1.000	.015	.309	.606	.453
	TDN	.015	1.000	-.085	-.297	-.513
	Temp C	.309	-.085	1.000	.235	.319
	UV	.606	-.297	.235	1.000	.819
	DOC	.453	-.513	.319	.819	1.000
Sig. (1-tailed)	Ln E coli	.	.428	.000	.000	.000
	TDN	.428	.	.154	.000	.000
	Temp C	.000	.154	.	.002	.000
	UV	.000	.000	.002	.	.000
	DOC	.000	.000	.000	.000	.
N	Ln E coli	144	144	144	144	144
	TDN	144	144	144	144	144
	Temp C	144	144	144	144	144
	UV	144	144	144	144	144
	DOC	144	144	144	144	144

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	DOC, Temp C, TDN, UV[b]	.	Enter

a. Dependent Variable: Ln E coli
b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.664[a]	.440	.424	1.057230551103024	1.848

a. Predictors: (Constant), DOC, Temp C, TDN, UV
b. Dependent Variable: Ln E coli

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	122.262	4	30.566	27.346	.000[b]
	Residual	155.365	139	1.118		
	Total	277.627	143			

a. Dependent Variable: Ln E coli
b. Predictors: (Constant), DOC, Temp C, TDN, UV

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.700	.451		1.551	.123
	TDN	.384	.142	.208	2.707	.008
	Temp C	.042	.015	.184	2.722	.007

	UV	24.195	4.253	.652	5.689	.000
	DOC	-.047	.186	-.033	-.250	.803

Coefficients[a]

Model		95.0% Confidence Interval for B		Correlations			Collinearity Statistics
		Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance
1	(Constant)	-.192	1.592				
	TDN	.104	.665	.015	.224	.172	.682
	Temp C	.011	.072	.309	.225	.173	.885
	UV	15.786	32.603	.606	.435	.361	.307
	DOC	-.414	.321	.453	-.021	-.016	.234

Coefficients[a]

Model		Collinearity Statistics VIF
1	(Constant)	
	TDN	1.465
	Temp C	1.130
	UV	3.262
	DOC	4.282

a. Dependent Variable: Ln E coli

Coefficient Correlations[a]

Model			DOC	Temp C	TDN	UV
1	Correlations	DOC	1.000	-.250	.501	-.804
		Temp C	-.250	1.000	-.112	.074
		TDN	.501	-.112	1.000	-.256
		UV	-.804	.074	-.256	1.000
	Covariances	DOC	.035	-.001	.013	-.636
		Temp C	-.001	.000	.000	.005
		TDN	.013	.000	.020	-.154
		UV	-.636	.005	-.154	18.085

a. Dependent Variable: Ln E coli

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	TDN	Temp C	U V
1	1	4.172	1.000	.00	.01	.00	.00
	2	.630	2.573	.00	.49	.00	.01
	3	.131	5.654	.03	.16	.24	.28
	4	.053	8.914	.31	.03	.75	.09
	5	.014	17.168	.66	.32	.01	.62

Collinearity Diagnostics[a]

Model	Dimension	Variance Proportions DOC
1	1	.00
	2	.00
	3	.00
	4	.03
	5	.96

a. Dependent Variable: Ln E coli

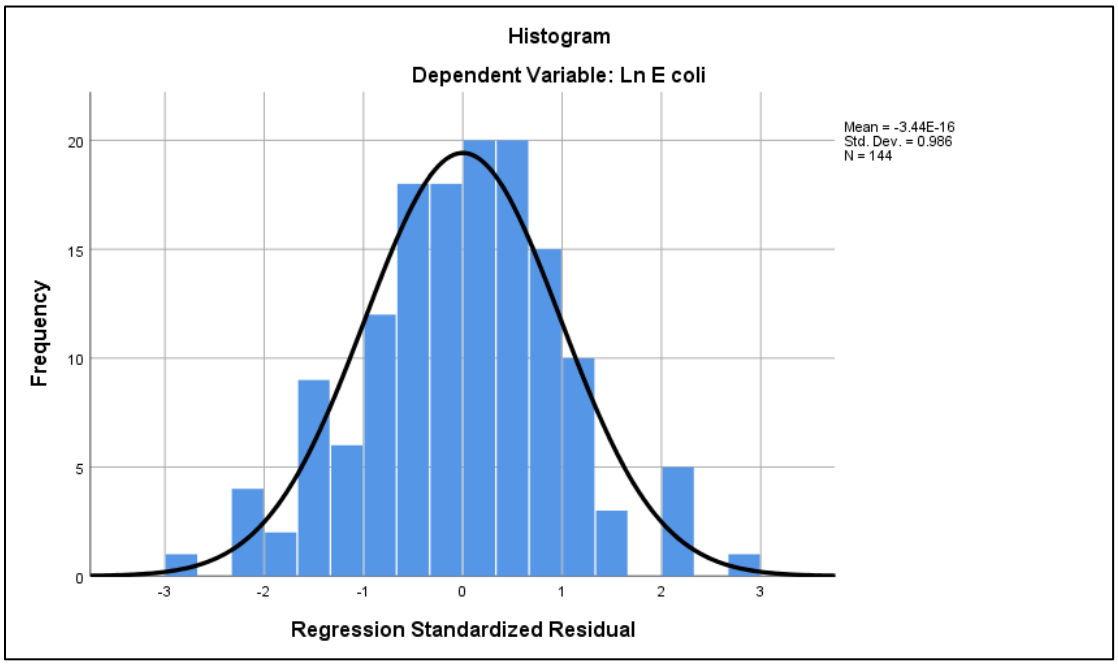
Residuals Statistics[a]

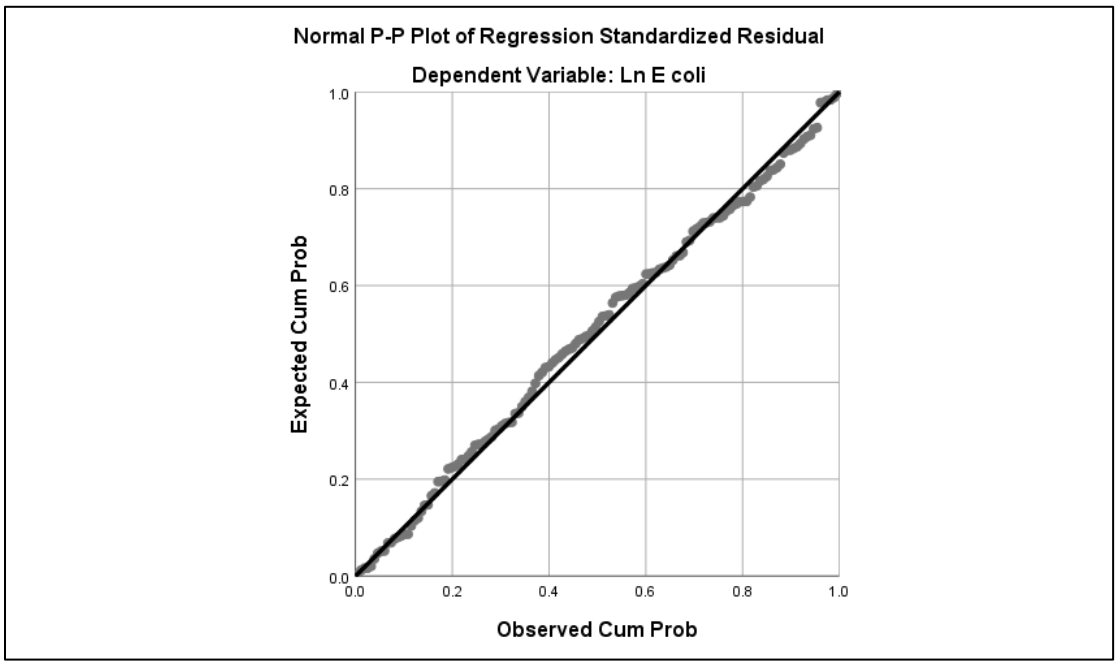
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.974081516265869	7.013273715972900	3.191573789175535	.924650985252131	144
Std. Predicted Value	-1.317	4.133	.000	1.000	144
Standard Error of Predicted Value	.104	.461	.186	.065	144
Adjusted Predicted Value	2.005201816558838	6.965748786926270	3.193540842647253	.921814226324575	144

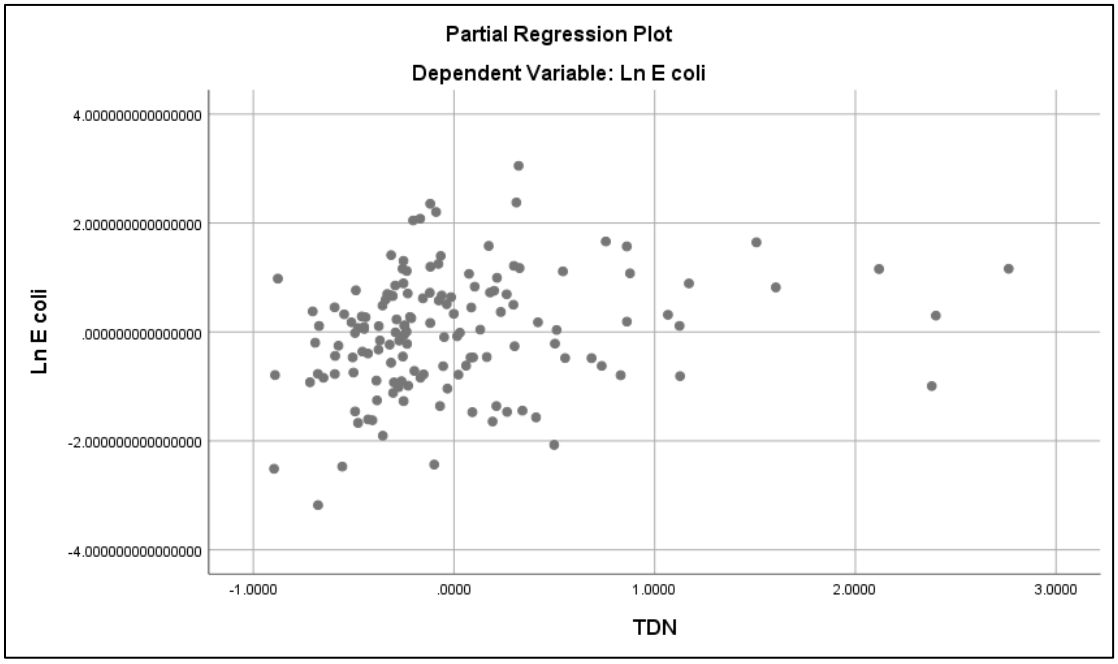
Residual	-2.920381069183350	2.928116083145142	.000000000000000	1.042339236175690	144
Std. Residual	-2.762	2.770	.000	.986	144
Stud. Residual	-2.800	2.846	-.001	1.004	144
Deleted Residual	-3.000510692596436	3.091455936431885	-.001967053471720	1.081211340367546	144
Stud. Deleted Residual	-2.872	2.922	-.001	1.012	144
Mahal. Distance	.389	26.246	3.972	4.224	144
Cook's Distance	.000	.144	.008	.016	144
Centered Leverage Value	.003	.184	.028	.030	144

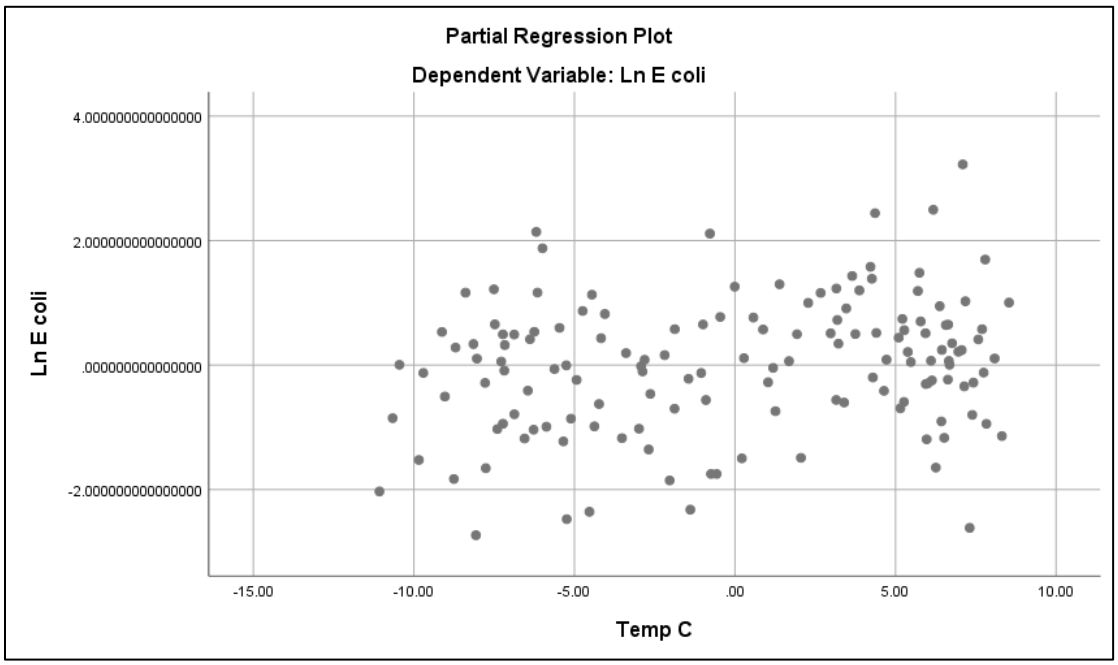
a. Dependent Variable: Ln E coli

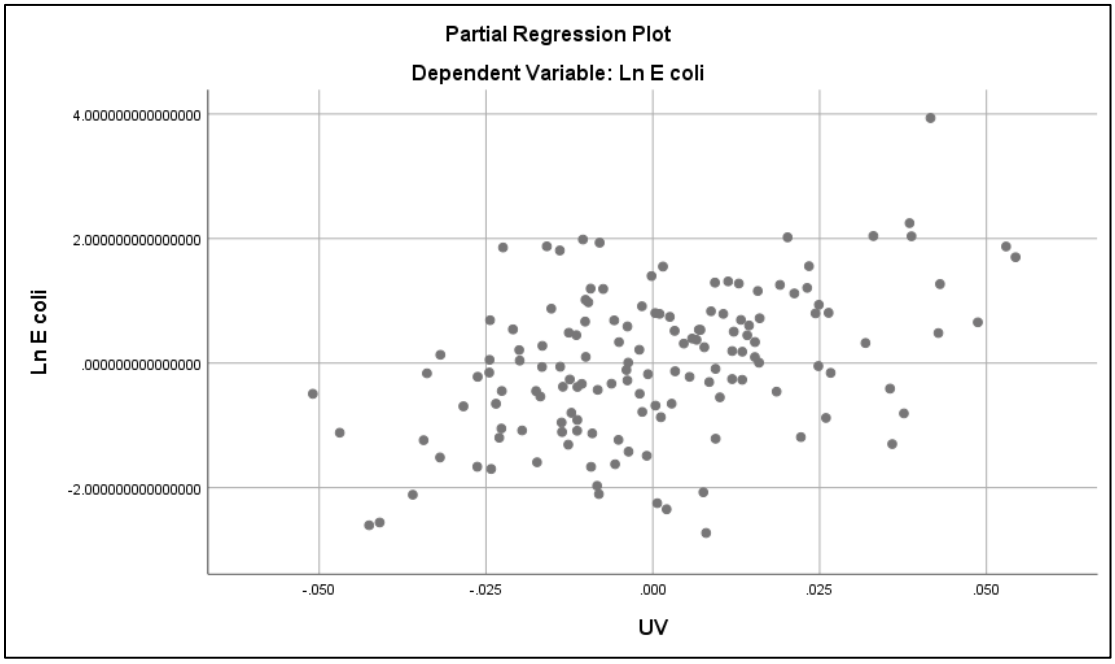
Charts

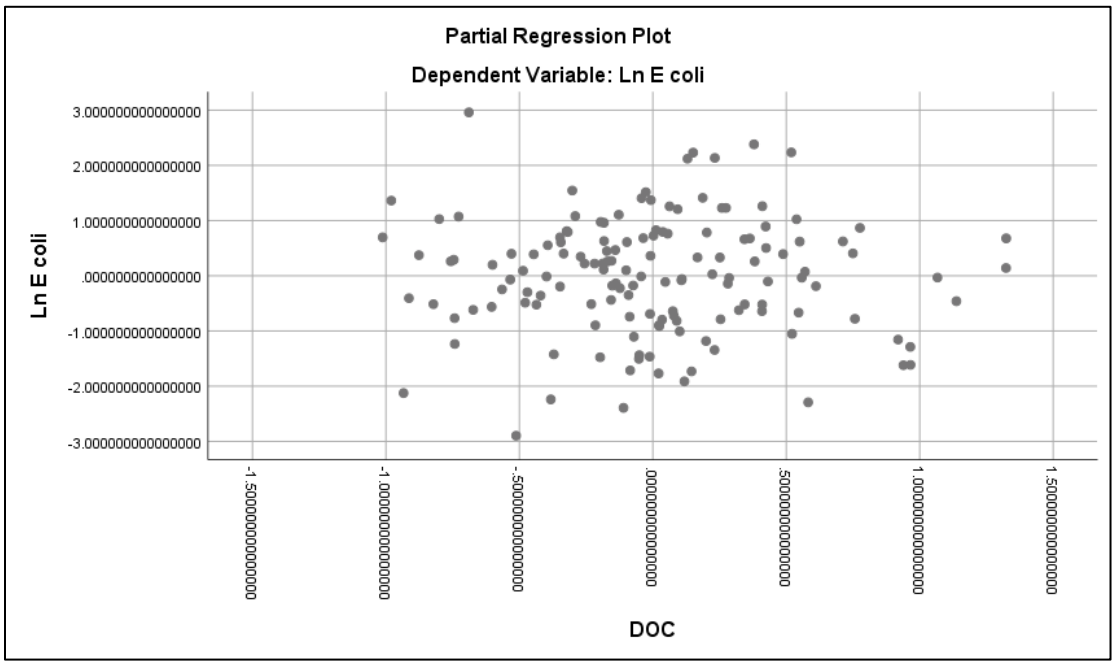



Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Ln E coli
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Partial Regression Plot
Dependent Variable: Ln E coli
Ln E coli
TDN
4.000000000000000
2.000000000000000
.000000000000000
-2.000000000000000
-4.000000000000000
-1.0000
.0000
1.0000
2.0000
3.0000


Partial Regression Plot
Dependent Variable: Ln E coli
Ln E coli
4.000000000000000
2.000000000000000
.000000000000000
-2.000000000000000
-15.00
-10.00
-5.00
.00
5.00
10.00
Temp C


Partial Regression Plot
Dependent Variable: Ln E coli
Ln E coli
4.000000000000000
2.000000000000000
.000000000000000
-2.000000000000000
-.050
-.025
.000
.025
.050
UV


Partial Regression Plot
Dependent Variable: Ln E coli
Ln E coli
3.000000000000000
2.000000000000000
1.000000000000000
.000000000000000
-1.000000000000000
-2.000000000000000
-3.000000000000000
-1.500000000000000
-1.000000000000000
-.500000000000000
.000000000000000
.500000000000000
1.000000000000000
1.500000000000000
DOC