2014-07-10

Source Apportionment of Wastewater Using Bayesian Analysis of Fluorescence Spectroscopy

Daniel B. Blake
Brigham Young University - Provo

Source Apportionment of Wastewater Using Bayesian Analysis of Fluorescence Spectroscopy

Daniel B. Blake

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

M. Brett Borup, Chair
William F. Christensen
Gustavious P. Williams
A. Woodruff Miller

Department of Civil and Environmental Engineering

Brigham Young University

July 2014

ABSTRACT

Source Apportionment of Wastewater Using Bayesian Analysis
of Fluorescence Spectroscopy

Daniel B. Blake
Department of Civil and Environmental Engineering, BYU
Master of Science

This research uses Bayesian analysis of fluorescence spectroscopy results to determine if wastewater from the Heber Valley Special Service District (HVSSD) lagoons in Midway, UT has seeped into the adjacent Provo River. This flow cannot be directly measured, but it is possible to use fluorescence spectroscopy to determine if there is seepage into the river.

Fluorescence spectroscopy results of water samples obtained from HVSSD lagoons and from upstream and downstream in the Provo River were used to conduct this statistical analysis. The fluorescence 'fingerprints' for the upstream and lagoon samples were used to deconvolute the two sources in a downstream sample in a manner similar to the tools and methods discussed in the literature and used for source apportionment of air pollutants. The Bayesian statistical method employed presents a novel way of conducting source apportionment and identifying the existence of pollution.

This research demonstrates that coupling fluorescence spectroscopy with Bayesian statistical methods allows researchers to determine the degree to which a water source has been contaminated by a pollution source. This research has applications in determining the affect sanitary wastewater lagoons and other lagoons have on an adjacent river due to groundwater seepage. The method used can be applied in scenarios where direct collection of hydrogeologic data is not possible. This research demonstrates that the Bayesian chemical mass balance model presented is a viable method of performing source apportionment.

Keywords: Fluorescence spectroscopy, Bayesian analysis, source apportionment, excitation emission, wastewater treatment, lagoon wastewater treatment

ACKNOWLEDGEMENTS

I wish to thank Dr. M. Brett Borup for his advice and guidance to this research and report, and for his role in mentoring me and teaching me about water and wastewater treatment processes. I would also like to thank Dr. William F. Christensen for his expertise on Bayesian analysis and for drafting the programming code used in this research. I also wish to express my gratitude to Scott Wright, manager of the Heber Valley Special Service District in Midway, UT, for his assistance and for allowing me to take samples at the facility.

I also express gratitude to Stephen A. Turner, manager of the Marlborough, MA water treatment plant, for his mentoring and encouragement, and for his role in being primarily responsible for getting me interested in water and wastewater treatment processes. I also express gratitude to Jody St. George and M. Joseph Geary, who were similarly influential in helping build my appreciation for the water and wastewater treatment industries.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Problem Statement

Fluorescence spectroscopy has been employed to characterize water quality in many different types of applications. Fluorescence spectroscopy methods, coupled with statistical analysis, demonstrate the potential to allow researchers to determine the degree to which a water source has been contaminated by a pollution source such as a wastewater treatment plant.

The Heber Valley Special Service District (HVSSD) in Midway, UT operates a series of wastewater lagoons. The lagoons are in close proximity to the Provo River, and it is thought that wastewater from the lagoons has the potential to seep into the river. This flow can not be directly measured, but it may be possible to use fluorescence spectroscopy to determine if there is seepage into the river. The purpose of this research is to determine if fluorescence spectroscopy methods and Bayesian statistical analysis can identify the amount of wastewater seepage from the HVSSD lagoons into the Provo River. Bayesian inference methods will be employed utilizing a computer program written in the R language to analyze the data collected.

1.2 Gaps in Research

Previous research has used a variety of methods to perform source apportionment. Chemical mass balance (CMB) methods have been used extensively (Miller, Friedlander et al. 1972, Friedlander 1973, Christensen and Gunst 2004, Massoudieh and Kayhanian 2013), but

these studies require a knowledge of the elemental profiles of the sources (Massoudieh and Kayhanian 2013). Another similar method uses a simple ratio of peaks within the EEMs (Coble 1996, Yan, Li et al. 2000). Another type of method uses specific statistical methods to perform source apportionment; in some cases the elemental profiles of the sources may be unknown. These statistical methods include principal component analysis (PCA) (Miller, Friedlander et al. 1972, Persson and Wedborg 2001), positive matrix factorization (PMF) (Ramadan, Eickhout et al. 2003, Soonthornnonda and Christensen 2008, Karanasiou, Siskos et al. 2009), partial least squares regression (PLS) (Persson and Wedborg 2001, Hall, Clow et al. 2005), and Bayesian inference (Christensen and Gunst 2004, Lingwall, Christensen et al. 2008, Massoudieh and Kayhanian 2013).

The previously listed studies include both air and water pollution source apportionment studies. Some attempt to apportion aerosol particulates, while others track pollution sources such as wastewater effluent, roadway runoff, and ship ballast water through water bodies. Many of these statistical methods attempt to account for measurement errors and variability within data sets.

These methods are promising methods to determine if a body of water has been unknowingly contaminated by a wastewater source such as an aerated lagoon. However, none specifically addresses this. Some studies have attempted to determine the environmental effect that wastewater lagoons, from swine or poultry operations, have on the groundwater and adjacent bodies of water (Sloan, Gilliam et al. 1999, Karr, Showers et al. 2001, Israel, Showers et al. 2005, Reichard and Brown 2009). These studies measure either the amount of nitrate-nitrogen or different isotopes of nitrogen to trace pollution from lagoons through groundwater and neighboring bodies of water. These studies, however, do not investigate the potential pollution

that sanitary wastewater lagoons have on bodies of water. A study published by Utah State University investigated the effect that the HVSSD and dairy farms located in Heber Valley have on nutrient levels in the groundwater (Jepson, McLean et al. 1991). This study concluded that there is potential for groundwater nutrient pollution and pollution of nearby Deer Creek Reservoir from the HVSSD and dairy farms. This study did not, however, specifically estimate the influence the HVSSD lagoons have on the Provo River; it only surmised their effect on the groundwater quality and nutrient levels.

To our knowledge, there have been no studies which seek to determine the degree to which a water source has been contaminated by a pollution source such as a wastewater lagoon.

1.3 **Objectives**

This research uses Bayesian analysis of fluorescence spectroscopy results in an attempt to determine if wastewater from the Heber Valley Special Service District lagoons has seeped into the adjacent Provo River. This research employed the relatively inexpensive measurements of fluorescence in water samples obtained from the HVSSD lagoons and from upstream and downstream in the Provo River to conduct the statistical analysis. The CMB model using Bayesian statistical methods presents a novel way of conducting source apportionment and identifying the existence of pollution.

2 BACKGROUND FOR RESEARCH

2.1 Fluorescence

Fluorescence spectroscopy has been employed to characterize water quality in many different types of applications. Organic matter in water exists in various states, including dissolved, colloidal and particulate states. Dissolved organic matter (DOM) is the most studied fraction, and fluoresces at various wavelengths (Coble 1996, Hudson, Baker et al. 2007). Many studies demonstrate that DOM has an intrinsic fluorescence (Lochmueller and Saavedra 1986, Coble 1996, Baker 2001, Hudson, Baker et al. 2007). Fluorescence occurs as an electron in a compound returns to its original energy state after having been excited to a higher energy level by the absorption of energy, in this case, from light (Hudson, Baker et al. 2007). These fluorescent properties are due to the presence of chromophores (particles that absorb light) and fluorophores (chromophores that absorb and then emit light at different wavelengths). The fluorescent fraction of DOM, or FDOM, is comprised of these compounds that emit light when excited (Mopper, Feng et al. 1996, Hudson, Baker et al. 2007). The intrinsic fluorescence of DOM has been investigated and studied extensively (Coble 1996, Patel-Sorrentino, Mounier et al. 2002, Kowalczuk, Durako et al. 2009, Murphy 2010). The most commonly studied FDOM components include humic acids and amino acids in proteins and peptides. Humic acids are produced from the decomposition of natural plant material by biological and chemical processes in both terrestrial and aquatic environments (Baker 2001, Hudson, Baker et al. 2007, Ghervase,

Carstea et al. 2010). FDOM exhibits characteristic peaks when subjected to fluorescence spectroscopy which correspond to amino acids such as tryptophan, fulvic acid, and tyrosine (Coble 1996, Baker 2001, Baker 2004, Hudson, Baker et al. 2007). Due to this characteristic, FDOM can be used as an indicator of the presence of organic matter from wastewater treatment processes.

Fluorescence spectroscopy has been used to characterize the quality of natural water bodies and to track anthropogenic pollution across a body of water (Stedmon, Markager et al. 2003, Hall, Clow et al. 2005, Hudson, Baker et al. 2007, Guo, Xu et al. 2010). In this way fluorescence spectroscopy operates as a fingerprint technique, allowing researchers to track the pollution through the ecosystem (Yan, Li et al. 2000, Baker 2001, Stedmon, Markager et al. 2003, Hall, Clow et al. 2005, Ghervase, Carstea et al. 2010).

Fluorescence spectroscopy generates excitation-emission matrices (EEMs) as excitation scans at specific wavelengths are sent through a sample of water, and emission wavelengths are subsequently recorded for each excitation scan along with intensity measurements for the emission scans. This produces a matrix of data which researchers have studied and identified specific areas that correspond to FDOM components (Lochmueller and Saavedra 1986, Coble 1996).

2.2 **Source Apportionment**

The chemical mass balance (CMB) model is a method of identifying the sources of observed or measured pollutants in a sample. CMB modeling has been used in air quality studies, surface water studies, and wastewater studies to apportion observed pollutants to their sources (Winchester and Nifong 1971, Friedlander 1973). The chemical mass balance model is based on the principle of the conservation of mass, where the measured pollutants in a recipient sample are

attributed to the sum of the elemental contributions from the sources (Miller, Friedlander et al. 1972, Christensen and Gunst 2004). Fluorescence spectroscopy data act as one means of estimating the concentration of FDOM in a sample, and can be used as a fingerprint for pollution sources, allowing researchers to track nutrient pollution through an ecosystem (Yan, Li et al. 2000, Hall, Clow et al. 2005). In this research, the fluorescence spectroscopy data are used as a means of estimating the concentration of FDOM in a sample, and the results are used in CMB modeling.

The chemical mass balance (CMB) method allows researchers to apportion pollutants to the sources from which they originate. The CMB method was used in the early 1970s by different researchers to study the contribution of pollution sources to the studied environment. Winchester and Nifong investigated the chemical compositions of air pollution around Lake Michigan, USA, in an effort to determine a relationship between air pollution and water pollution (Winchester and Nifong 1971). Miller, Friendlander and Hidy investigated the aerosol composition of the atmosphere in the Los Angeles basin of California, USA. In this study, the researchers measured the chemical compositions of aerosol particulates and then used linear algebraic equations to determine the source contributions to the aerosol from sea salt, soil dust, automobile emission, and fly ash (Miller, Friedlander et al. 1972, Friedlander 1973). These early studies established the framework for later CMB studies.

Water quality studies commonly employ a variety of different statistical methods in performing CMB analyses. Multivariate analyses are commonly used, and include parallel factor analysis (PARAFAC) (Stedmon, Markager et al. 2003, Hall, Clow et al. 2005, Kowalczuk, Durako et al. 2009, Guo, Xu et al. 2010), PCA (Persson and Wedborg 2001, Boehme, Coble et al. 2004), and PLS (Persson and Wedborg 2001, Hall, Clow et al. 2005). Multivariate analyses

are commonly employed, and allow researchers to identify patterns and structures within EEM data (Persson and Wedborg 2001). Multivariate analyses are able to decompose EEMs into different independent fluorescent components, where researchers are able to identify specific portions of the EEM data which correspond to physical parameters such as proteins and amino acids (Stedmon, Markager et al. 2003).

These statistical techniques allow for tracking of nutrient pollution through an ecosystem, and improve the ability to use fluorescence spectroscopy data as a fingerprinting technique.

The mass balance model has several inherent assumptions which are fundamental to its application in this research. These assumptions must hold in order for the results of the model to be valid. One assumption is that the fluorescent surface of the EEM fingerprint is unaltered over time. A second underlying assumption is that the only sources which contribute to the downstream sample are the HVSSD lagoons and the upstream river. If these mass balance assumptions are violated, then the estimates of the CMB are untenable.

2.3 **Proof of Technique**

This Bayesian statistical method has been tested and validated with other data collected as part of a previous research study (Ferrell 2009). The results of this model application will be presented at the International Water Association's 2014 World Water Congress & Exhibition in September 2014.

The application of the Bayesian statistical model using data from the previous study demonstrated the validity of the model. For this previous study, grab samples were collected from the Provo and Spanish Fork, UT water reclamation facility effluents, as well as upstream and downstream of the associated discharge points. Samples were collected over a period of

several months in 2008. In addition to these grab samples, the flow rate at the upstream sampling point was measured, and the flow rate of the discharge was recorded. This provided data necessary for a mass flow balance.

The main goal of the previous research "…was to use a mass balance with the flow rate and fluorescence to relate upstream, discharge, and downstream fluorescence as quantified by using EEMs. The overall question is whether or not fluorescence can be predicted in the same manner that a contaminant concentration can be estimated by incorporating a mass balance" (Ferrell 2009).

This research attempted to use simple fluorescence spectroscopy numbers to conduct a chemical mass balance. The data for this research are well suited for a chemical mass balance utilizing the Bayesian statistical method discussed in Section 3.4.1. The grab samples collected from the discharge, and upstream and downstream of the discharge location allow for a direct chemical mass balance, and can be compared with the measured volumetric flow rates of the same sampling locations. The wastewater source is a direct discharge, allowing for a direct chemical mass balance. Comparing the results of the Bayesian statistical model and the measured flow rates verify the validity of the statistical model.

The fluorescence spectroscopy data were analyzed using the Bayesian statistical method presented in Section 3.4.1. Figure 2-1 on the following page displays the measured volumetric flow rate percentage contributions for the data from the Provo wastewater treatment plant. The upstream and discharge measured flow rates are calculated as percentage contributions of the downstream flow rate.

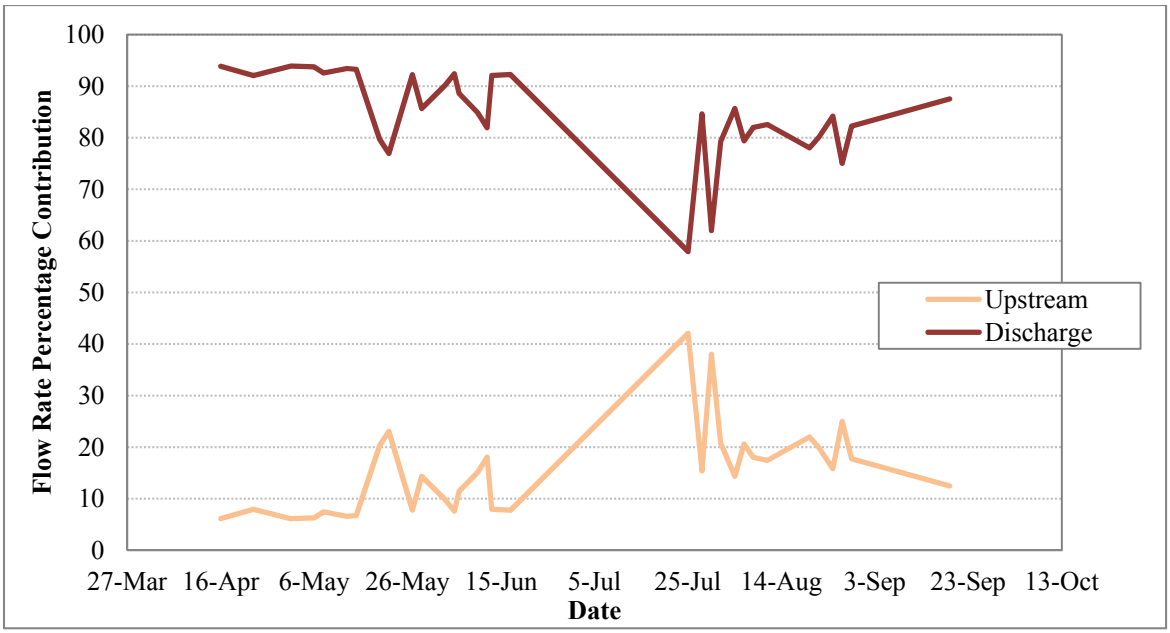


Figure 2-1. Provo measured flow rate percentage contributions.

Figure 2-2 on the following page displays the calculated mean percentage contributions and 95% credible intervals for the Provo data. These results were calculated utilizing the CMB method outlined in this research.

Figure 2-3 on the following page displays the calculated and measured discharge contributions for the Provo wastewater treatment plant data. The measured discharge generally stays within the 95% credible interval of the calculated discharge.

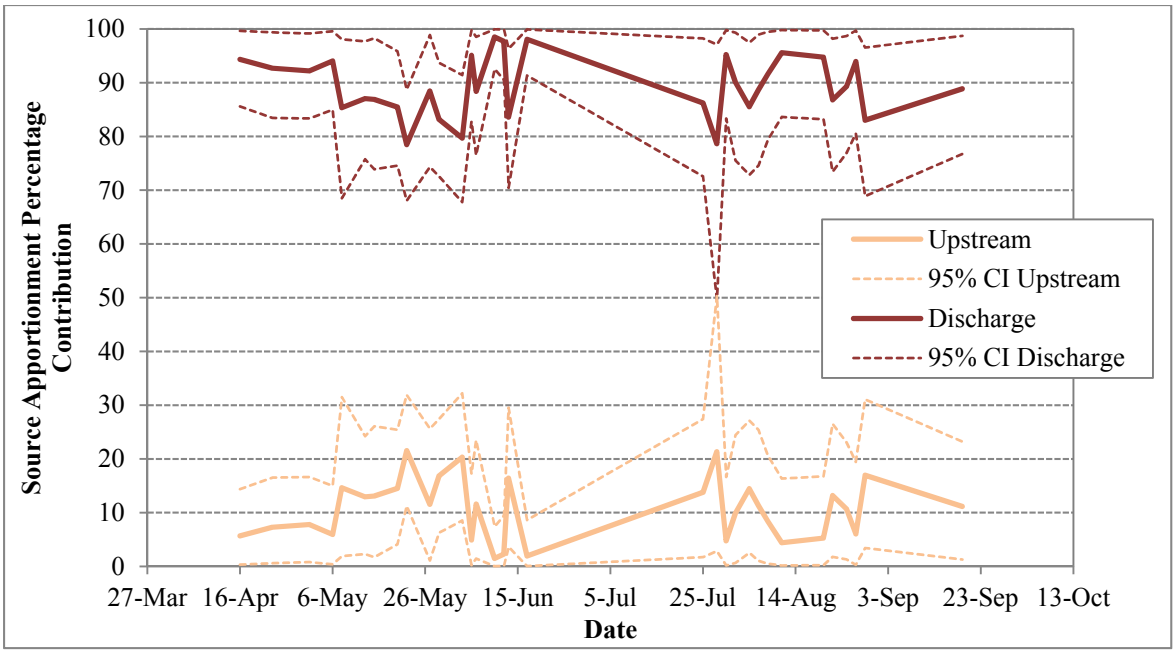


Figure 2-2. Provo Bayesian mean contribution and credible intervals.

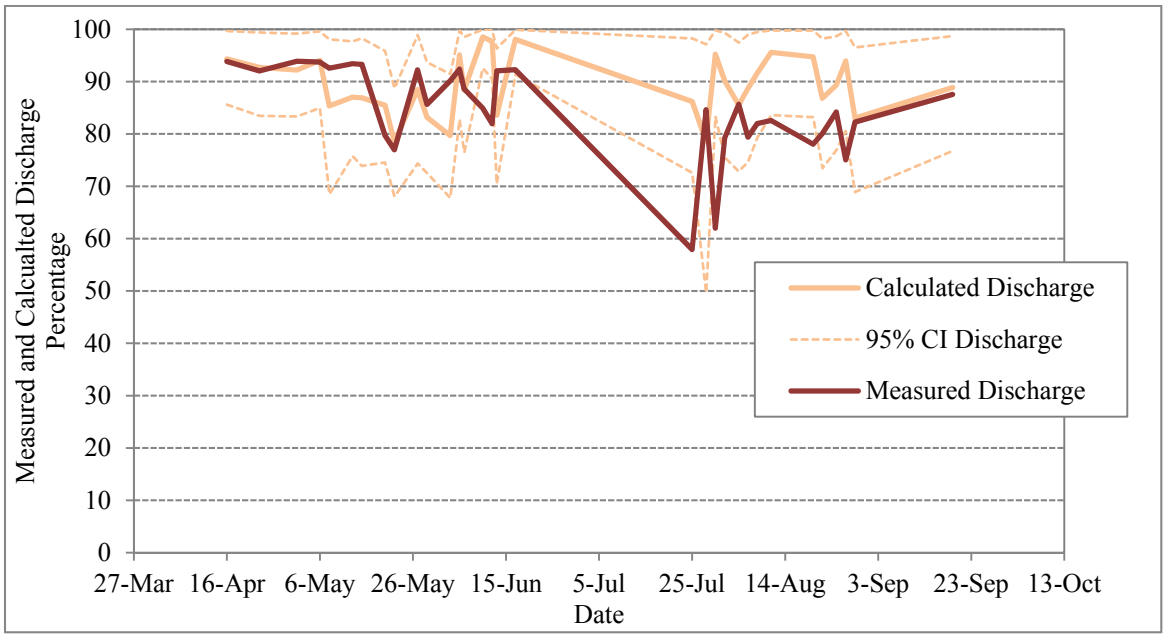


Figure 2-3. Provo measured and calculated discharge contribution.

Figure 2-4 below displays the measured volumetric flow rates of the Spanish Fork wastewater treatment plant discharge and receiving body of water.

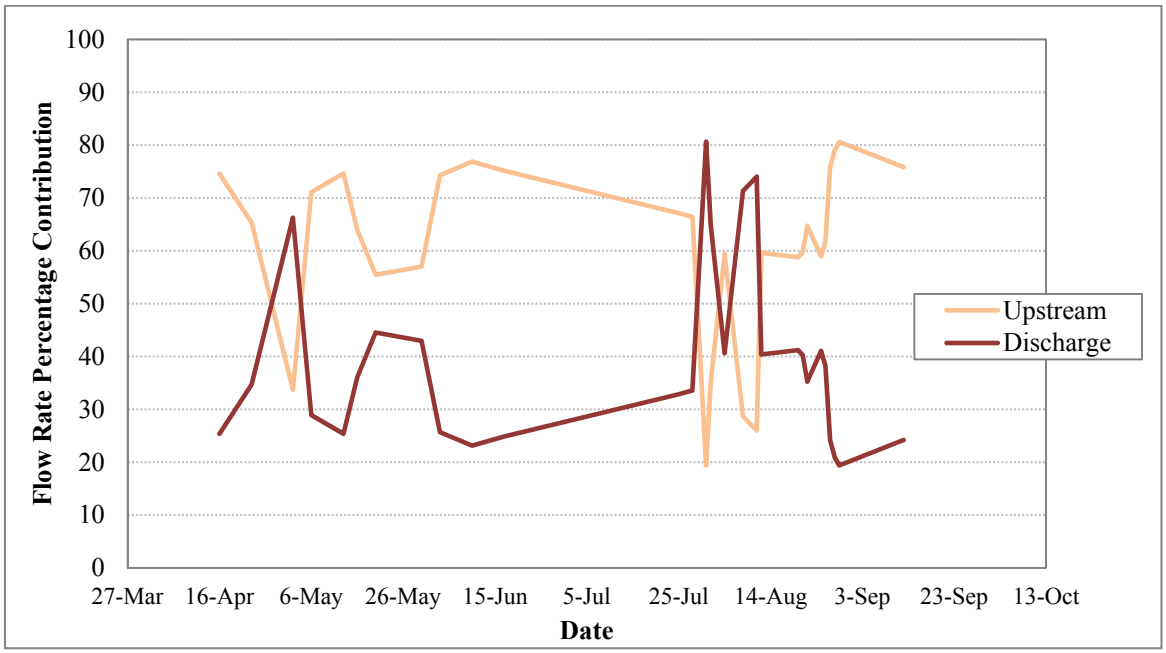


Figure 2-4. Spanish Fork measured flow rate percentage contribution.

Figure 2-5 on the following page displays the calculated mean percentage contribution and 95% credible intervals for the Spanish Fork data.

Figure 2-6 on the following page displays the calculated and measured discharge contributions for the Spanish Fork wastewater treatment plant data. The measured discharge generally stays within the 95% credible interval of the calculated discharge.

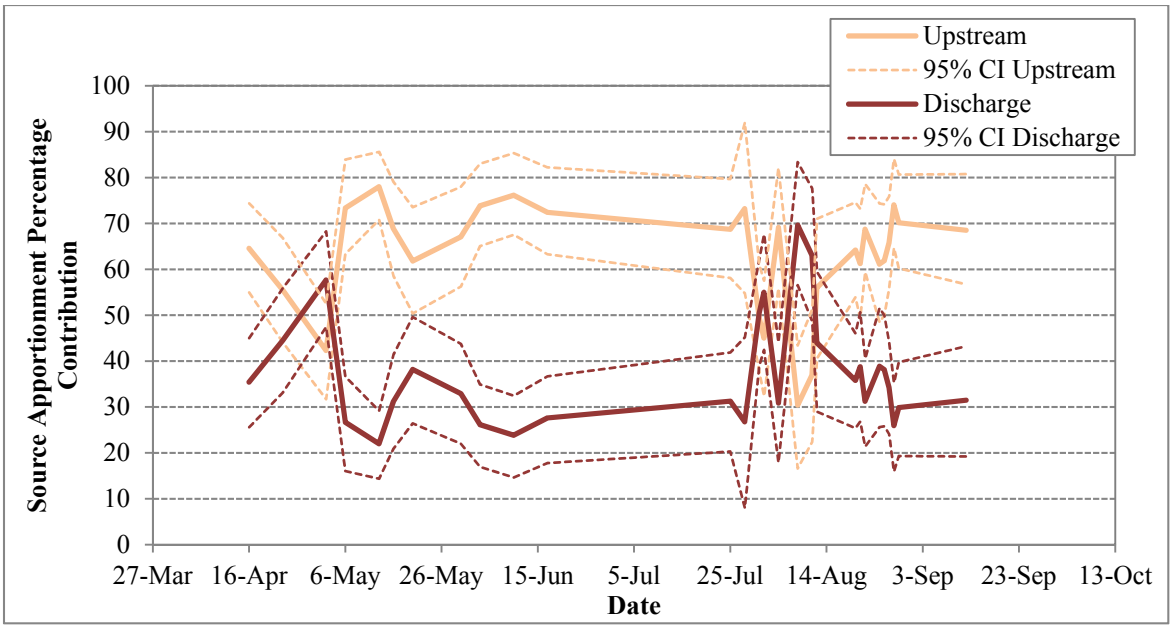


Figure 2-5. Spanish Fork Bayesian mean contribution and credible intervals.

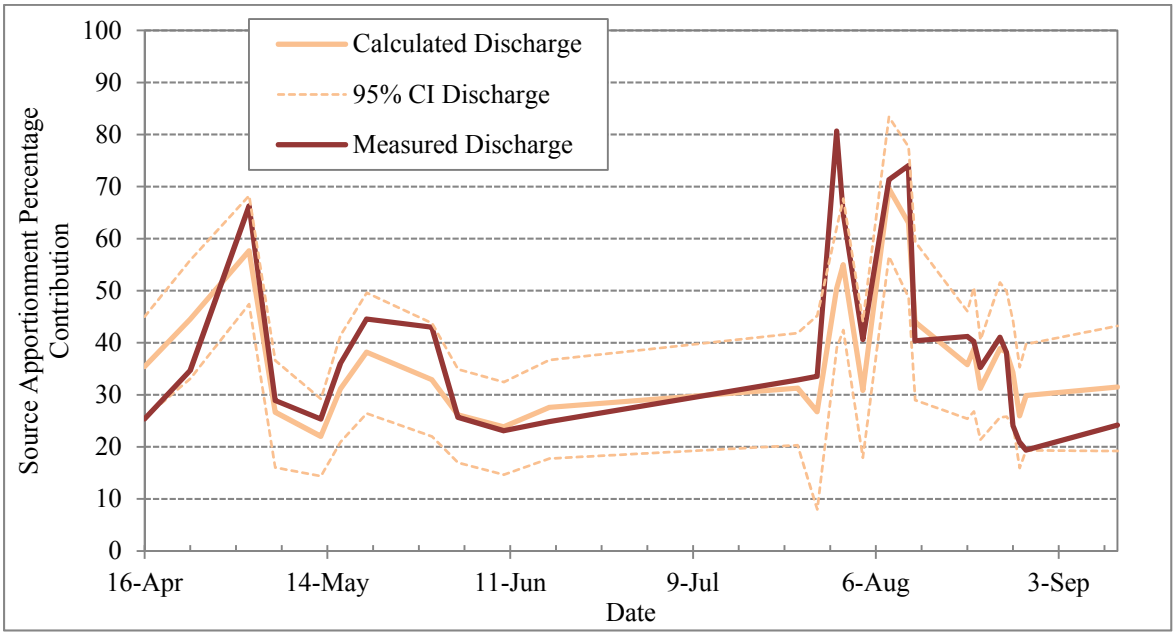


Figure 2-6. Spanish Fork measured and calculated discharge contribution.

The close similarities of each set of graphs demonstrates the ability of the Bayesian method to determine source contribution of a known pollutant source, in this case, to determine

the effect a wastewater treatment plant discharge has on the receiving body of water. The results of the CMB model correlate well with measured flow rates. Both the Provo data and Spanish Fork data exhibit temporal variability in measured flow rates. The Spanish Fork data exhibits a larger temporal variation in flow rates than the Provo data. Nevertheless, the Bayesian analysis results are similar to the measured flow rates in its fluctuations, demonstrating that the Bayesian statistical method is effective in performing source apportionment.

This Bayesian statistical method has been tested and validated with data from a previous study. Given this tested ability of the Bayesian model to perform source apportionment from a direct discharge, this research intends to determine if fluorescence spectroscopy methods and Bayesian statistical analysis can accurately identify the amount of wastewater seepage from the HVSSD lagoons into the Provo River.

2.4 **Heber Valley Special Service District**

The Heber Valley Special Service District treats wastewater from Heber, UT and Midway, UT. Both municipalities are located in the Heber Valley of Utah, in Wasatch County. Wasatch County created the Heber Valley Special Service District in 1976, and the lagoon treatment system went into operation in 1981. Around 2000 the plant was reaching its hydraulic capacity, and the following year an additional treatment cell was constructed. Figure 2-7 on the following page displays the layout of the HVSSD lagoons prior to the construction of the additional treatment cell. The last constructed cell is located in the bottom left of the figure.

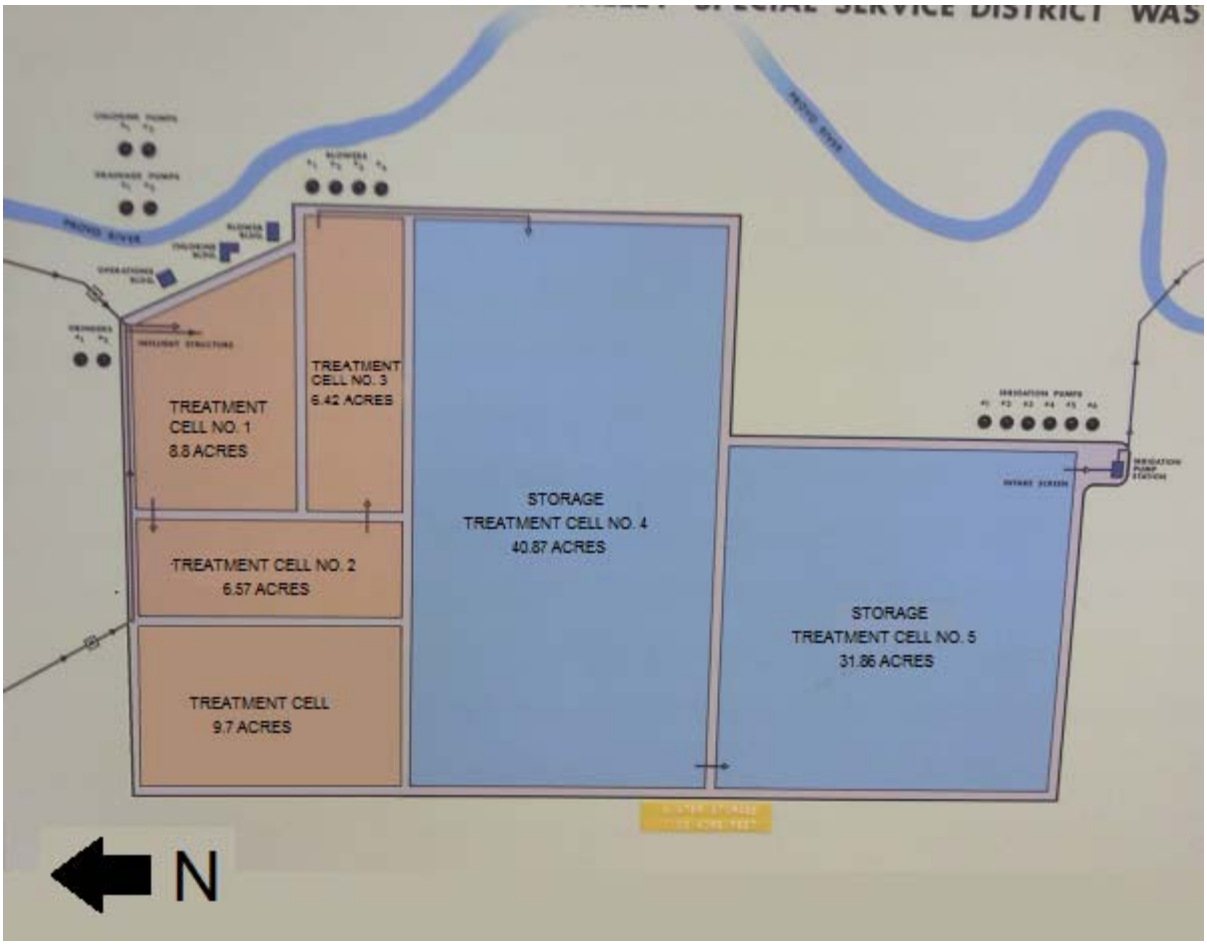


Figure 2-7. Schematic of HVSSD lagoons (District 2000).

The schematic layout of the lagoons is displayed above in Figure 2-7. Lagoon 1, also referred to as Treatment Cell No. 1, is 8.8 acres in size. Raw wastewater first flows into lagoon 1, and from there it flows into cell No. 2, with a surface area of 6.57 acres. Flow then proceeds to cell No. 3, with a surface area of 6.42 acres, and into the newest treatment cell, with a surface area of 9.7 acres. There are two storage lagoons, used during winter months – cell No. 4 and cell No. 5, colored in blue, with surface areas of 40.87 and 31.86 acres respectively.

During summer months, treated wastewater is supplied to surrounding farmlands for irrigation. Approximately 60% of the flow is used for irrigation. The other 40% is sent to a rapid

infiltration basin (RIB), a rock-filled pit where the treated wastewater seeps into the ground. The RIB is located northwest of the lagoons, further from the Provo River than any of the lagoons treatment cells. Groundwater infiltration from the RIBs will travel away from the Provo River in a southwest direction. During the winter months, the flow to the RIBs is reduced to around 40%, and the other 60% is sent to the winter storage lagoons.

In October 2013, a newly constructed activated sludge process was put on line. With this new process on line, during winter months, approximately 60% of the flow is sent to the activated sludge system, and the other 40% is sent to the lagoons. During the summer months, the district expects to send more incoming wastewater flow to the lagoons. The activated sludge system has a capacity of 2.5 MGD, and the effluent from the system is sent to the RIBs.

The Provo River is located directly east of the lagoons, less than 200 feet away, and runs from north to south. The close proximity of the lagoons to the river is suspect, and it is thought that wastewater from the lagoons has the potential to seep through the ground and into the river, thus polluting the river. Untreated or partially treated wastewater entering a natural body of water would have an adverse impact on the natural environment. Polluting wastewater could increase the levels of nutrients in the river, introduce trace metals into the ecosystem, and would negatively affect the local ecosystem.

A white paper published by the Utah State Division of Water Quality, Department of Environmental Quality states that “Because they hold large amounts of water, lagoons are at risk for leaking excessively, causing groundwater contamination.” (Quality 2014). The Heber Valley Special Service District has a ‘general permit for land disposal of municipal wastewater’ issued by the Utah DEQ. This permit states that “There shall be no discharge to Waters of the State except as provided for in paragraphs b.” Paragraphs b states that “The discharge of water from

emergency overflow systems shall occur only as a result of equipment failure and the need to protect the plant from flooding and/or to prevent severe property damage and will be allowed only if the facility has been properly operated and maintained. If such a discharge occurs, whenever possible the permittee shall dispose of the overflow on land to avoid any potential impacts on receiving waters." (Quality 2011) This permit from the state directs the Heber Valley Special Service District to not discharge to any water body. As noted in the white paper, wastewater from lagoons can leak into the ground, causing groundwater contamination and, if located in close proximity to a river, contamination of the river. Though the lagoons are considered total containment, state specifications for the design of wastewater lagoons allow for some seepage, declaring that seepage loss should not exceed 6500 gal/ac/day (Rules 2014). The design of the HVSSD lagoons did not include water tight liners (Jepson, McLean et al. 1991).

A 2004 Water Quality Implementation Report prepared by Psomas, an engineering firm in Salt Lake City, UT, for The Wasatch County Council, recommended that seepage and return flows from the HVSSD lagoons be examined. This report acknowledged that doing so would provide information on possible nutrient pollution of the Provo River. The report states that "…the lagoons are considered 60% depletion by the State Engineer's Office, but the actual ratio might be on the order of 40 to 45%" (Eckhoff, Boyd et al. 2004). This report estimates that 55% of the total phosphorus loading on Deer Creek Reservoir comes from the Provo River. Quantifying the effect of the HVSSD lagoons on the Provo River would aid in better understanding nutrient pollution of the aquatic ecosystem of the area.

2.4.1 HVSSD Hydrogeologic Study

The Utah Department of Environmental Quality (DEQ) Division of Water Quality (DWQ) required the HVSSD to complete a hydrogeologic site characterization before approval

of its plant upgrades in 2011 (District 2011). The DWQ desired to ensure that the HVSSD facilities were not contaminating groundwater and ultimately Deer Creek Reservoir. The HVSSD contracted Horrocks Engineers, Inc. of Heber, UT, which subcontracted Sunrise Engineering of Draper, UT, to conduct the hydrogeologic site characterization. The final report was submitted to the HVSSD in October 15, 2013 (Yang 2013).

The report details the findings of the hydrogeologic site characterization, groundwater flow monitoring, and water quality monitoring at the lagoons and rapid infiltration basin (RIB). The report demonstrates that general groundwater flow in the area of the HVSSD facilities is west-southwest, flowing away from the Provo River. However, the report details the finding that there is groundwater mounding associated with the two southern lagoon cells, where groundwater levels have risen by about 2 feet or more. This groundwater mounding causes a local reversal of groundwater flow, flowing toward the Provo River with an estimated hydraulic gradient of about 1% or 0.01. The report states "this local reversal of groundwater flow direction does not change the general groundwater flow direction at the lagoon and RIB site which is towards the west-southwest" (Yang 2013). The report concludes that the RIB site would not contribute groundwater flow to the Provo River. The finding of groundwater mounding, however, indicates that seepage from the lagoons may infiltrate into the Provo River. Figure 2-8 displays the potentiometric surface of the groundwater in the vicinity of the lagoons. The groundwater surface map is inconclusive, however, and conclusive results cannot be drawn from it. The northern section of the map indicates that there is no seepage into the river; rather, it indicates that the groundwater table draws water away from the river. The southern section of the map is inconclusive in determining the groundwater flow direction, and definitive statements about flow therefore cannot be made.

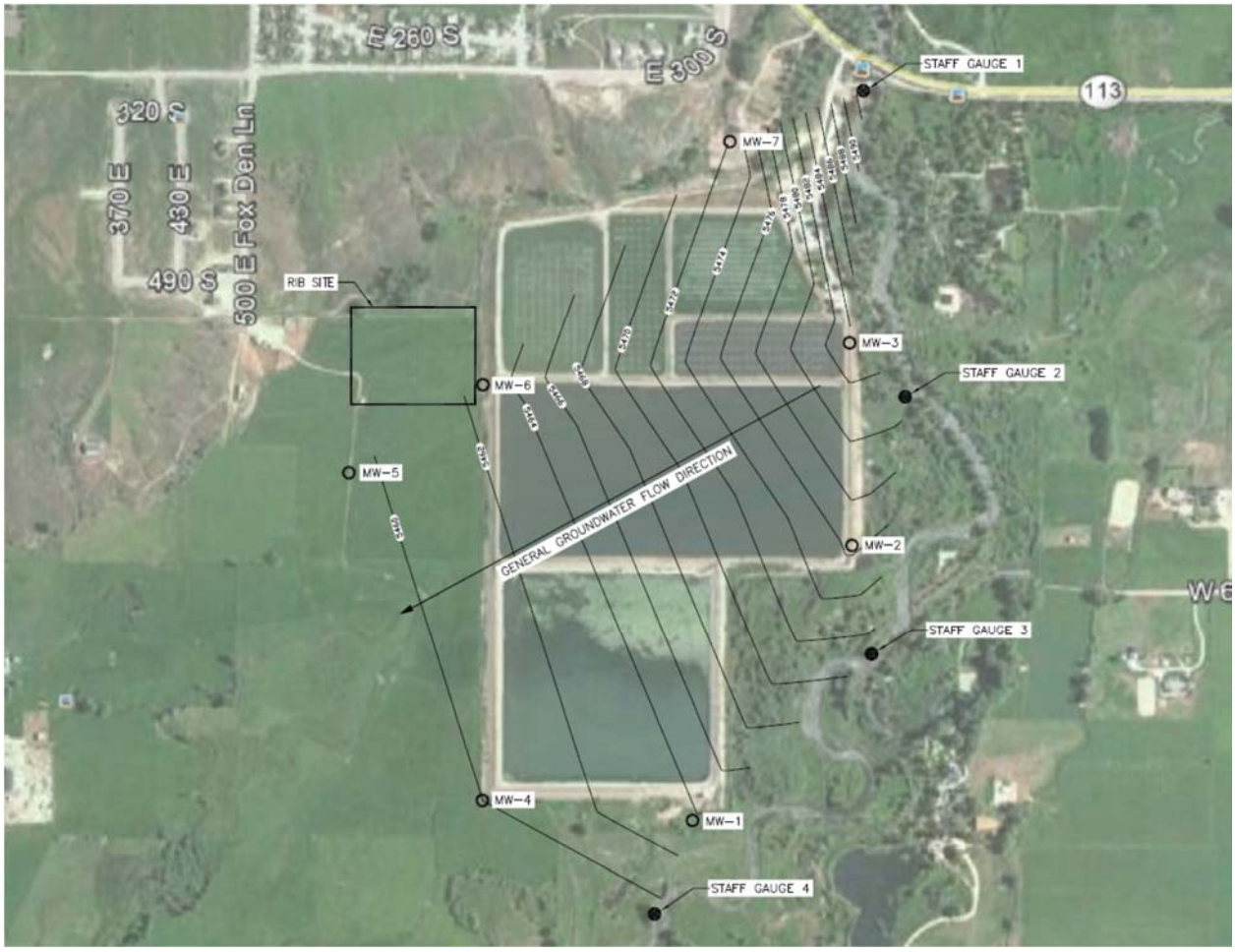


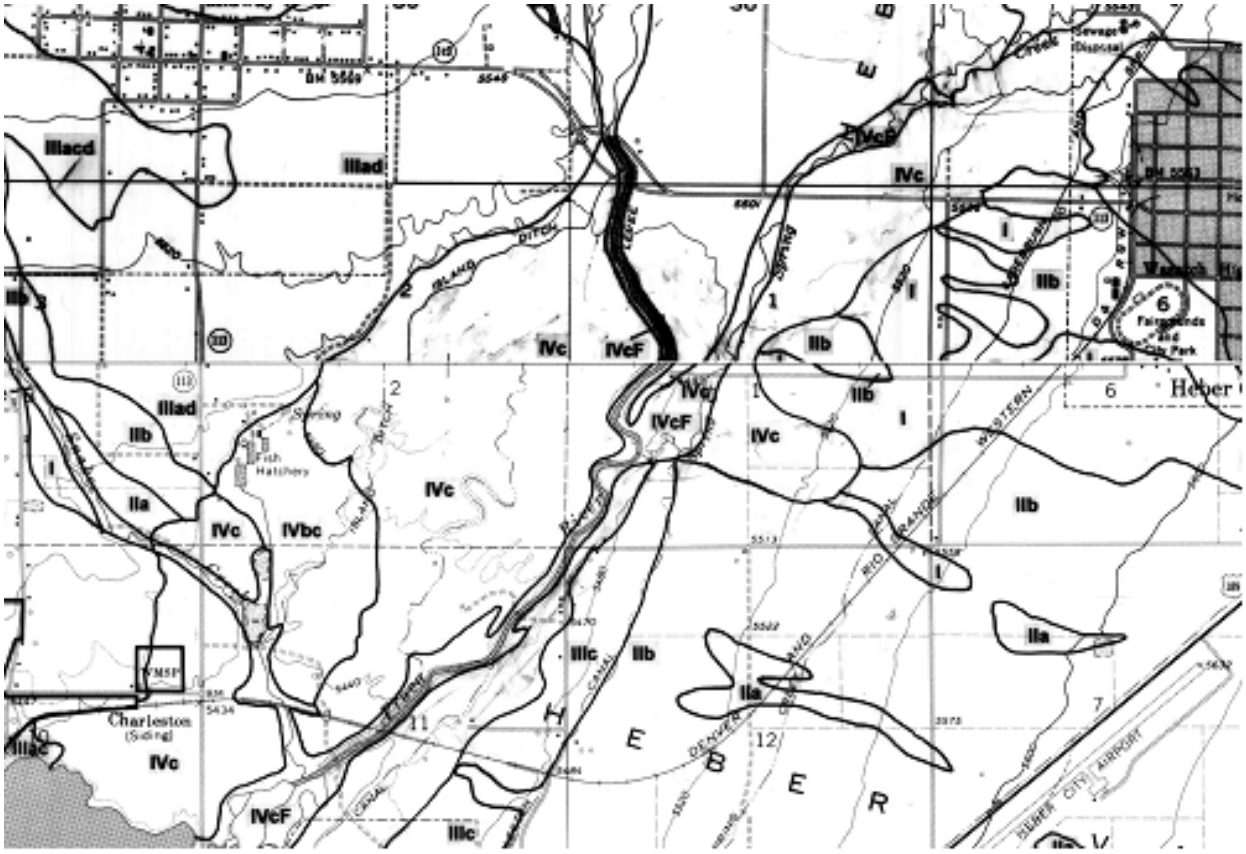
Figure 2-8. Potentiometric surface of groundwater in vicinity of lagoons.

The results of the groundwater monitoring around the HVSSD lagoons are inconclusive in determining if seepage from the lagoons enters the Provo River. The report by Sunrise Engineering estimates the amount of phosphorus that lagoon seepage could contribute to the Provo River, but does not estimate the amount of flow or DOM that would seep into the rivers.

2.4.2 Suitability of Lagoon Location

A folio of maps was published by the Utah Geological Survey in 1995 and was intended to display the spatial suitability of areas for septic-tank soil-absorption systems in western Wasatch County. These maps display areas that are classified on a spectrum from being 'generally suitable' to 'generally unsuitable'. The area where the HVSSD is located is classified as IVc, meaning it is generally unsuitable and in an area where the depth to the shallowest expected water table is 0-5 feet (Hylland 1995). Figure 2-9 below displays the map of the area around the HVSSD with its unsuitable nature for septic systems similarly explained. Figure 2-10 below displays the suitability of septic systems in the same area, with an aerial photograph superimposed over the map to illustrate the location of the HVSSD lagoons. Both figures are taken from the 1995 report by Michael D. Hylland (Hylland 1995).

The hydrogeologic site characterization study of the area completed in 2013 indicates that groundwater from the RIBs would flow away from the Provo River, and also concluded that the area south of the lagoons has the highest water table level, with groundwater levels in that area being closest to the ground surface (Yang 2013). This is evidence that seepage from the lagoons is causing a rise in groundwater levels near the southern storage lagoons in the vicinity of the Provo River, and indicates that there is potential for seepage into the Provo River.



SCALE 1:24 000

1 0 1 MILE

1000 0 1000 2000 3000 4000 5000 6000 7000 FEET

1 .5 0 1 KILOMETER

CONTOUR INTERVAL 40 FEET
NATIONAL GEODETIC VERTICAL DATUM OF 1929

EXPLANATION

Suitability:

I	Generally suitable
II	Generally suitable but locally unsuitable
III	Generally unsuitable but locally suitable
IV	Generally unsuitable

Qualifiers:

a	Slow percolation rate (greater than 60 minutes per inch)
b	Fast percolation rate (less than 4 minutes per inch)
c	Depth to shallowest expected water table 0-5 feet
d	Depth to bedrock (including tufa in Midway area) 0-5 feet
e	Slope steeper than 25 percent

Figure 2-9. Suitability of septic systems in area around HVSSD lagoons.

SCALE 1:24 000

1 0 1 MILE

1000 0 1000 2000 3000 4000 5000 6000 7000 FEET

1 .5 0 1 KILOMETER

CONTOUR INTERVAL 40 FEET
NATIONAL GEODETIC VERTICAL DATUM OF 1929

EXPLANATION

Suitability:

I	Generally suitable
II	Generally suitable but locally unsuitable
III	Generally unsuitable but locally suitable
IV	Generally unsuitable

Qualifiers:

a	Slow percolation rate (greater than 60 minutes per inch)
b	Fast percolation rate (less than 4 minutes per inch)
c	Depth to shallowest expected water table 0-5 feet
d	Depth to bedrock (including tufa in Midway area) 0-5 feet
e	Slope steeper than 25 percent

Figure 2-10. Suitability of septic systems in area around HVSSD lagoons with lagoons superimposed.

3 MATERIALS AND METHODS

3.1 Sampling Locations

Samples were collected over an eight month period, from August 2013 to May 2014. Grab samples were taken from Lagoon 3, the final wastewater lagoon at the HVSSD. Samples were also collected from the Provo River upstream and downstream of the lagoon. The upstream sampling point is located at a fishing access spot on South Ryan Lane off of W Midway Lane, Heber City, UT, at 40° 30’ 33.96”, -111° 27’ 2.54”. The downstream sampling point is located approximately 700 feet off the road from W 1200 S, Heber City, UT, at 40° 29’ 32.26”, -111° 27’ 13.93”. Figure 3-1 below displays the sampling points.

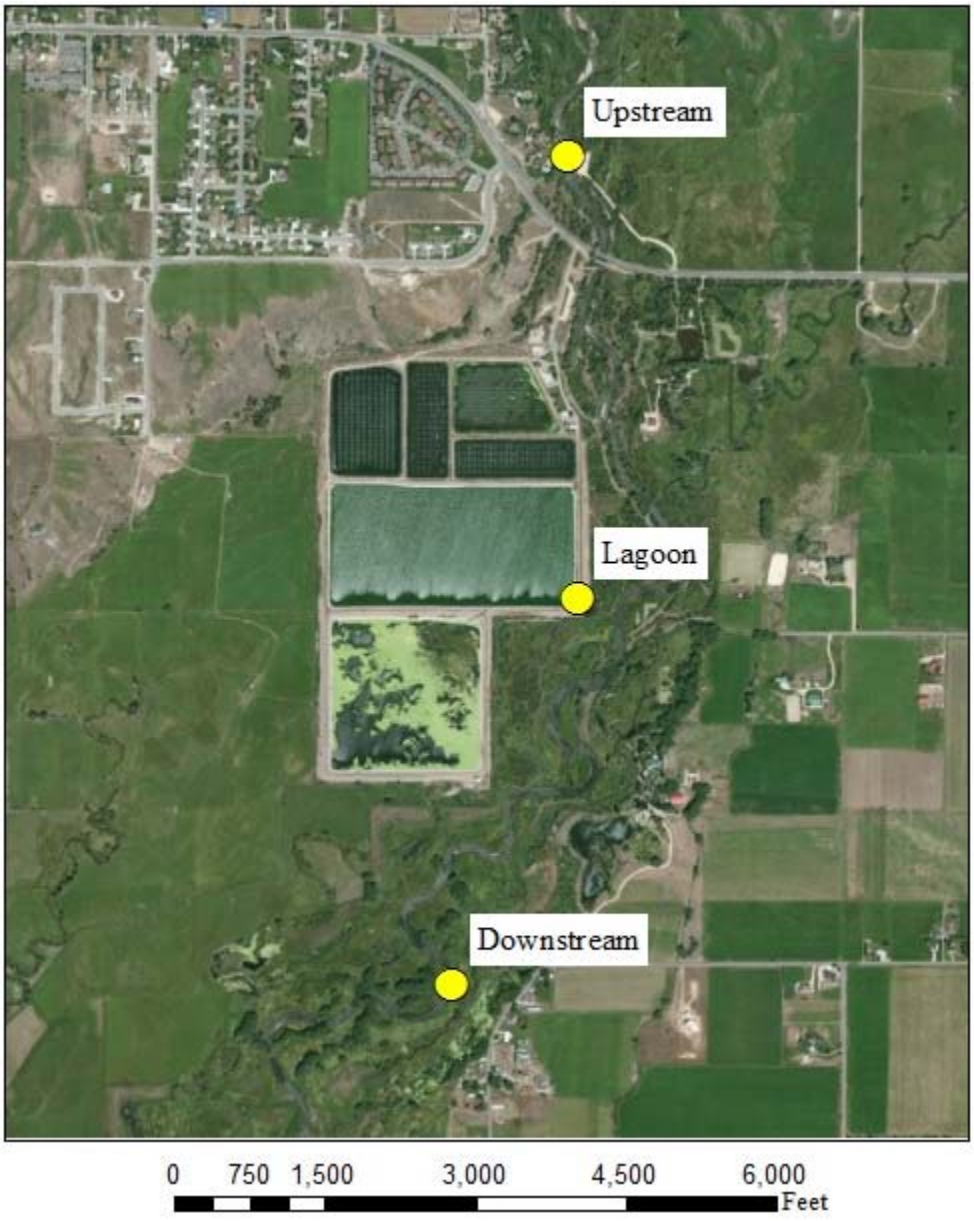


Figure 3-1. Map of Sampling Points.

Figure 3-2 on the following page is a photograph of the downstream sampling point, with the view facing upstream.

Figure 3-2. Downstream sampling point, view upstream

3.2 Testing Procedures

A total of 22 samples were collected over a period from August 2013 to May 2014. Samples were analyzed using a PerkinElmer LS55 fluorimeter. The water samples were filtered with a Whatman brand medium flow rate cellulose filter paper, which has a pore size of 11 μm. The beakers and cuvettes used were rinsed carefully to ensure that there was no contamination from other samples. After filtering, the cuvette was filled with sample and placed in the fluorimeter. A total of 30 scans were performed on each sample. The scans consisted of a range

of excitation wavelengths from 250-400nm in 5nm increments, and fluorescence intensity was measured at corresponding emission wavelengths from 300-500nm. Each scan generated one data file, for a total of 30 files for each sample. The samples from Lagoon 3 were diluted ½ so that the fluorescence intensity would be less than the maximum measurable limit of 1000.

3.3 **Analysis Methods**

The set of 30 scans from each water sample was condensed into one file, and EEM graphs were automatically generated. Figure 3-3, Figure 3-4 and Figure 3-5 below are a set of representative EEM graphs, with upstream, lagoon, and downstream samples taken August 30, 2013.

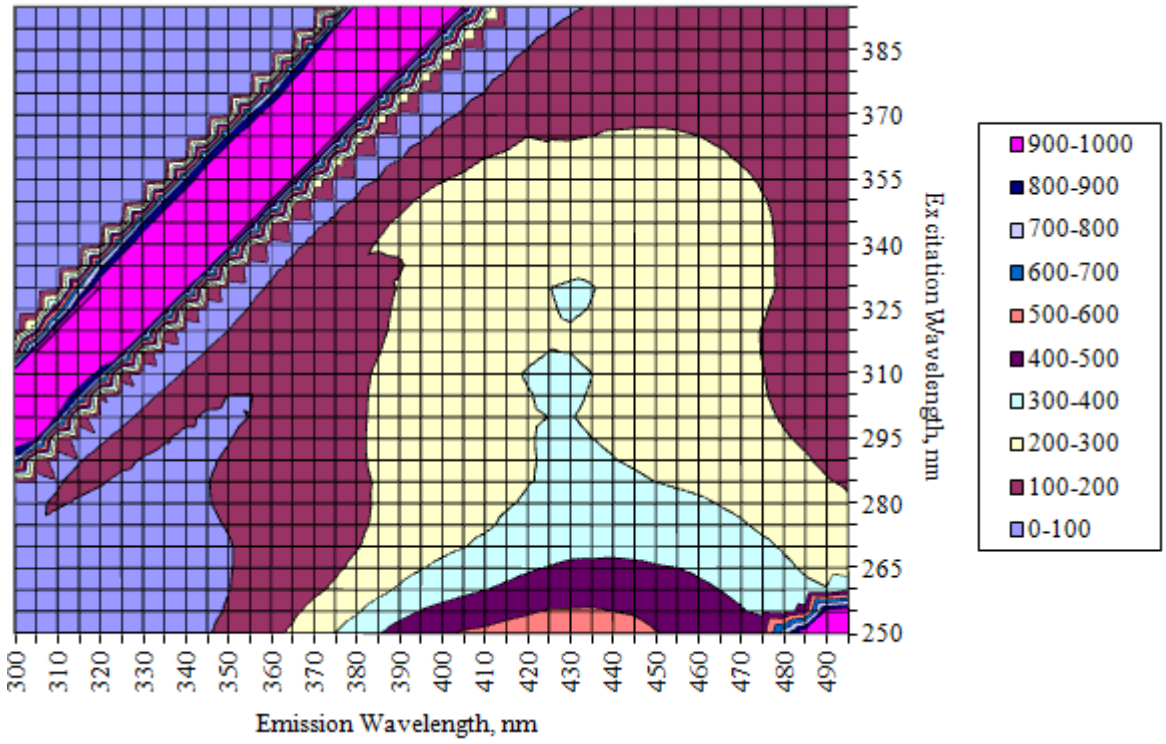


Figure 3-3. Representative upstream sample EEM.

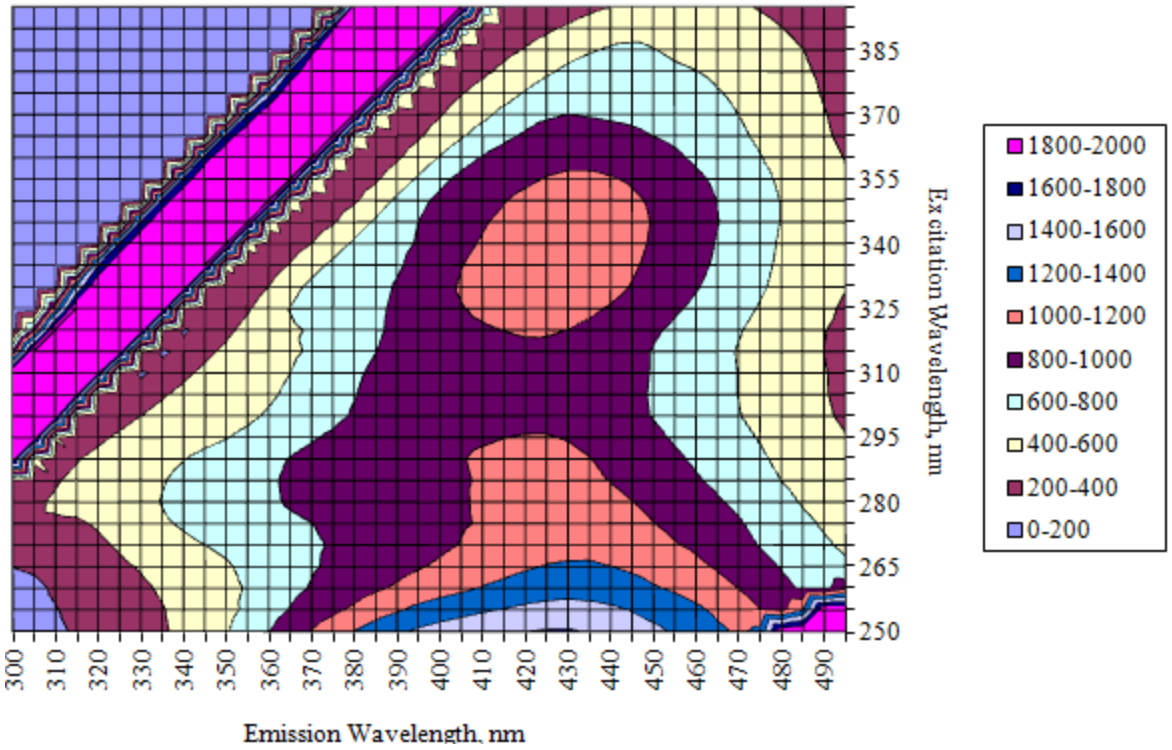


Figure 3-4. Representative lagoon sample EEM.

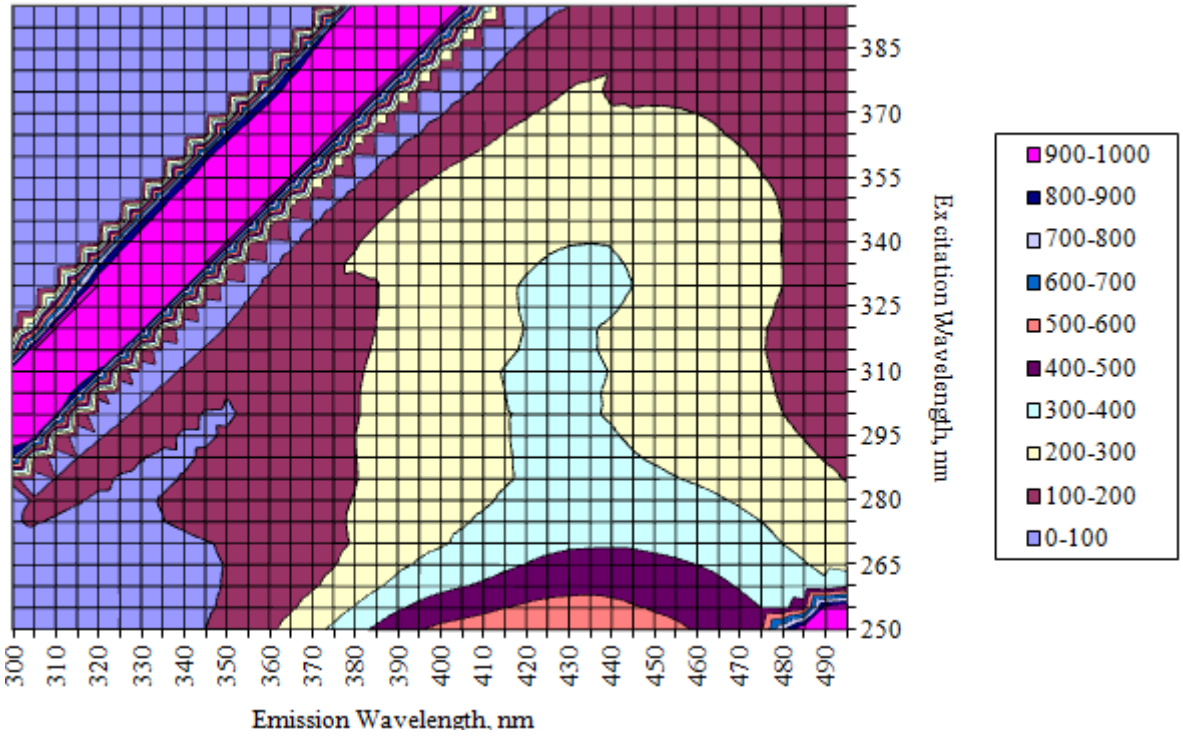


Figure 3-5. Representative downstream sample EEM.

3.4 **Bayesian Analysis**

For this analysis, 14 data points were selected for inclusion in the Bayesian analysis program. The points used in this analysis include wavelengths previously identified with the fluorescence of particular FDOM fractions including tryptophan, fulvic acid, and humic acid. Additionally, other wavelengths close to and within the observed EEM peaks (the fluorescent regions) were included in the analysis. Table 3-1 below displays the wavelengths chosen for inclusion in the statistical analysis, and if applicable, their component as identified in the listed study.

Table 3-1. Wavelengths included in statistical analysis.

Excitation	Emission	Component	Study
260	380	Humic acid	(Coble 1996)
275	310	Tyrosine	(Coble 1996)
275	340	Tryptophan	(Coble 1996)
280	350		
280	355	Tryptophan	(Hudson, Baker et al. 2008)
280	360		
330	420		
330	430		
330	440		
340	420		
340	430	Fulvic acid	(Hudson, Baker et al. 2008)
350	420	Humic acid	(Coble 1996)
350	430		
350	440		

A code was written in R language for the Bayesian analysis by Dr. William Christensen of the Department of Statistics at Brigham Young University. The code may be found in the appendix. The statistical program R, developed by the R Development Core Team, is a free software programming language which provides statistical and graphical techniques for data analysis. R is used widely by statisticians (Venables and Smith 2014). The R program used uses Markov Chain Monte Carlo (MCMC) algorithms to determine the posterior distributions. These posteriors indicate the amount of influence the upstream river and the HVSSD lagoons have on the downstream location of the Provo River. The posterior distribution is generally represented with a histogram or a density curve, and is referred to as a posterior.

Bayesian inference methods require iterative calculations and the incorporation of many parameters during these iterations. MCMC simulations carry out Bayesian analysis, and employ one of two main algorithms, the Gibbs Sampler and the Metropolis-Hastings algorithm. This research used a program called JAGS, which uses the Gibbs Sampler algorithm to perform MCMC simulations (Plummer 2013).

This Bayesian methodology provides a complete distribution for the source contribution, as opposed to a single point estimate as other statistical methods might do (Lingwall, Christensen et al. 2008, Massoudieh and Kayhanian 2013). Still, point estimates can be obtained from these posterior distributions and taken to be the median value of the distribution of the upstream contribution.

3.4.1 Bayesian Model

The Bayesian model used in this research is based on the CMB form. As discussed earlier in Section 2.2, the CMB model is used to perform source apportionment. Fluorescence spectroscopy data act as a fingerprint for pollution sources, allowing researchers to track nutrient

pollution through an ecosystem. Applying a CMB model using Bayesian statistics allows for the determination of the influence or source contribution the HVSSD lagoons have on the Provo River. As mentioned previously in Section 3.4, the Bayesian method generates a posterior probability density function, as opposed to a single value of source contribution. This is an advantage of utilizing a Bayesian approach.

The Bayesian hierarchical model employed in this research is used to fit a regression model subject to the constraints that the fractional contribution of upstream and wastewater sources (P_1 and P_2) are bounded between 0 and 1 and sum to unity. A gamma distributed model for the downstream fluorescence is used, and is shown below in Equation (1).

$$\boldsymbol{y_i \sim Gamma\left(mean = \mu = \lambda_{i1}P_1 + \lambda_{i2}P_2, variance = \sigma_i{}^2\right), i = 1, \dots, 14} \quad (1)$$

Where, for the model above, the following notation is used

y_i: true downstream profile at point i

λ_{i1}: fluorescence measurement at point i at the upstream sampling location

λ_{i2}: fluorescence measurement at point i at lagoon cell 3

P_1: true proportion of upstream contribution

P_2: true proportion of lagoon cell 3 contribution

$\sigma_i{}^2$: error variance

Generally the Gamma distribution uses parameters α and β such that the mean of a gamma random variable is $\frac{\alpha}{\beta}$ and variance is $\frac{\alpha}{\beta^2}$, where α is the shape parameter and β is the rate parameter. Given this, Equation (1) can be rewritten as:

$$Gamma(\alpha = \frac{\mu^2}{\sigma^2},\ \beta = \frac{\mu}{\sigma^2}) \quad (2)$$

Equation (1) essentially states that the true y, or downstream profile, equals the sum of the true proportion of upstream contribution and the true proportion of lagoon cell 3 contributions. The model constrains the estimate of P_1 by assigning a Beta prior distribution to

P_1, with P_2 equal to 1-P_1. The variance σ_i^2 is defined by the observed variability of the i^{th} fluorescence measurement drawn from the downstream location over the sampling period. The parameters y_i, P_1, and P_2 all have distributions, as opposed to treating them as fixed unknown values. This allows each parameter or variable to be estimated with a measure of uncertainty, which strengthens the validity of the results.

In MCMC methods, the iteratively calculated posterior distributions converge after many iterations, so it is standard practice to disregard the Markov chain until each parameter has clearly converged. This is checked with the trace plot, which displays the spread of the posterior distribution across draws from the Markov chain. The initial part of the chain (before convergence) is called a burn-in period. In this model, after a burn-in of 2000 draws, which was more than adequate for the posterior distribution to converge, 10000 draws of the posterior distribution are retained for data analysis. Figure 3-6 on the following page displays a trace plot for the post burn-in period. The medians for these distributions are reported as the point estimates for P_1 and P_2.

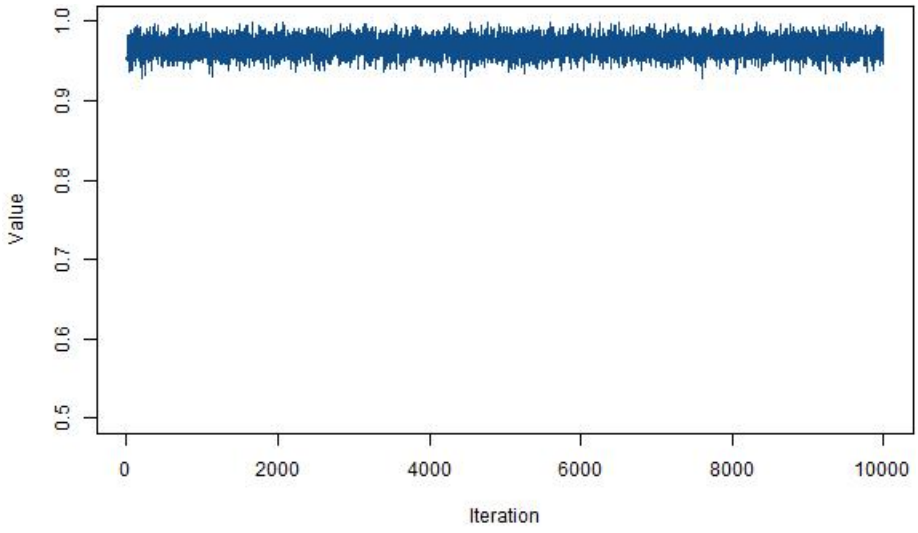


Figure 3-6. Example of trace plot for P_1.

3.4.2 Advantages of Bayesian Approach

A Bayesian approach offers certain advantages over other statistical methods which are desirable in this research. Bayesian statistics allows for the incorporation of constraints directly into the prior distributions (Lingwall, Christensen et al. 2008, Massoudieh, Gellis et al. 2013, Massoudieh and Kayhanian 2013). The prior distribution can incorporate *a priori* information about the source and recipient samples' compositions (Sharifi, Haghshenas et al. 2013).

Bayesian statistics also allows for constrained estimation, or constrained optimization. It is required that the fractional contributions of upstream and wastewater sources sums to unity. That is, the downstream sampling point is comprised of flow from upstream and of seepage from the lagoons. A Bayesian method allows for the constraint of the estimated contribution from upstream and lagoon to be non-negative and to sum to one. This is a feature missing in standard CMB methods where contributions can fall outside of [0,1] and will not generally sum to 1 (de Vos 2004, Lingwall, Christensen et al. 2008, Massoudieh and Kayhanian 2013).

The Bayesian analysis is particularly robust given that each parameter has its own uncertainty that is reflected in the parameter's probability density function (PDF). Bayesian inference allows for the incorporation of uncertainties in the measurements of the source and recipient samples' elemental profiles, which strengthens the validity of the results.

3.4.3 Assumptions of Mass Balance Model

One assumption of the model is that the intensity of the EEM measurement is linearly related to the DOM concentration. That is, an EEM with higher intensities has a higher DOM content. Different studies conclude that fluorescence intensity is linearly related to DOM (Westerhoff, Chen et al. 2001, Kowalczuk, Durako et al. 2009), and log-linearly related (Mopper, Feng et al. 1996).

Another assumption of the mass balance analysis is that the fluorescent characteristics of the DOM, or the fluorescent surface of the EEM, do not change over time. In order for the Bayesian model to be applicable, the fluorescence surface should not change as wastewater seeps from the lagoon into the Provo River. If the fluorescent surface does not change, then the model will predict the correct source contributions. However, if the fluorescent surface does change, then the CMB is not able to accurately predict source contribution.

A third assumption of the Bayesian model is that any change in the DOM content of the downstream sample is due to influence of the HVSSD lagoons. This assumes that the only contributions of DOM to the downstream sample are the upstream Provo River and seepage from the lagoons. This excludes any contribution from other sources such as surface runoff or organic matter produced by aquatic organisms.

When any of these assumptions are violated, the estimates from mass balance analyses will be unjustifiable. Conversely, when the results of mass balance analyses yield physically untenable estimates, one can conclude that one or more mass balance assumptions are unjustified and/or that measurements are unreliable.

4 RESULTS

4.1 R Program Results

Results of the Bayesian analysis demonstrate that seepage from the HVSSD lagoons affects the Provo River. Table 4-1 below displays the median point estimates taken from the posterior distributions of the 22 samples, calculated using the R statistical program. This median point estimate is equivalent to the source contribution of each element (upstream, lagoon) to the composition of the downstream profile.

The average of the lagoon contribution point estimates is 3.6%, and the average upstream contribution is 96.4%. Table 4-2 displays the median point estimates of lagoon contribution and associated 95% credible interval limits.

Table 4-3 displays the median point estimates of upstream contribution and associated 95% credible interval limits.

Table 4-1. Median point estimates of source contribution.

Sample	Median Lagoon Contribution, %	Median Upstream Contribution, %
1	3.15	96.9
2	1.75	98.2
3	0.37	99.6
4	2.24	97.8
5	6.34	93.7
6	11.4	88.6
7	1.93	98.1
8	2.93	97.1
9	2.47	97.5
10	1.41	98.6
11	1.41	98.6
12	2.25	97.7
13	1.45	98.5
14	1.62	98.4
15	0.81	99.2
16	0.80	99.2
17	10.3	89.7
18	2.53	97.5
19	8.02	92.0
20	6.95	93.1
21	3.62	96.4
22	5.54	94.5
Average	3.60	96.4

Table 4-2. Median contribution and CI of Lagoon.

Sample	Lower Credible Interval, 2.5%	Median Lagoon Contribution, %	Upper Credible Interval, 97.5%
1	1.02	3.15	5.25
2	0.16	1.75	4.07
3	0.01	0.37	1.77
4	0.15	2.24	5.92
5	2.03	6.34	10.7
6	0.75	11.4	31.0
7	0.10	1.93	6.04
8	0.21	2.93	7.41
9	0.13	2.47	7.26
10	0.06	1.41	5.24
11	0.06	1.41	5.24
12	0.11	2.25	7.14
13	0.06	1.45	5.67
14	0.07	1.62	5.87
15	0.03	0.81	3.64
16	0.03	0.80	3.53
17	5.05	10.3	15.1
18	0.13	2.53	7.04
19	3.86	8.02	12.0
20	2.73	6.95	11.0
21	0.44	3.62	7.40
22	1.68	5.54	9.21
Average		3.60	

Table 4-3. Median contribution and CI of Upstream.

Sample	Lower Credible Interval, 2.5%	Median Upstream Contribution, %	Upper Credible Interval, 97.5%
1	94.8	96.9	99.0
2	95.9	98.2	99.8
3	98.2	99.6	100.0
4	94.1	97.8	99.8
5	89.3	93.7	98.0
6	69.0	88.6	99.2
7	94.0	98.1	99.9
8	92.6	97.1	99.8
9	92.7	97.5	99.9
10	94.8	98.6	99.9
11	94.8	98.6	99.9
12	92.9	97.7	99.9
13	94.3	98.5	99.9
14	94.1	98.4	99.9
15	96.4	99.2	100.0
16	96.5	99.2	100.0
17	84.9	89.7	95.0
18	93.0	97.5	99.9
19	88.0	92.0	96.1
20	89.0	93.1	97.3
21	92.6	96.4	99.6
22	90.8	94.5	98.3
Average		96.4	

Figure 4-1 below displays the median source contribution of the lagoon and associated 95% credible intervals. This figure is not scaled according to time or sampling date. The x-axis displays the sample number, and is not representative of when the sample was taken.

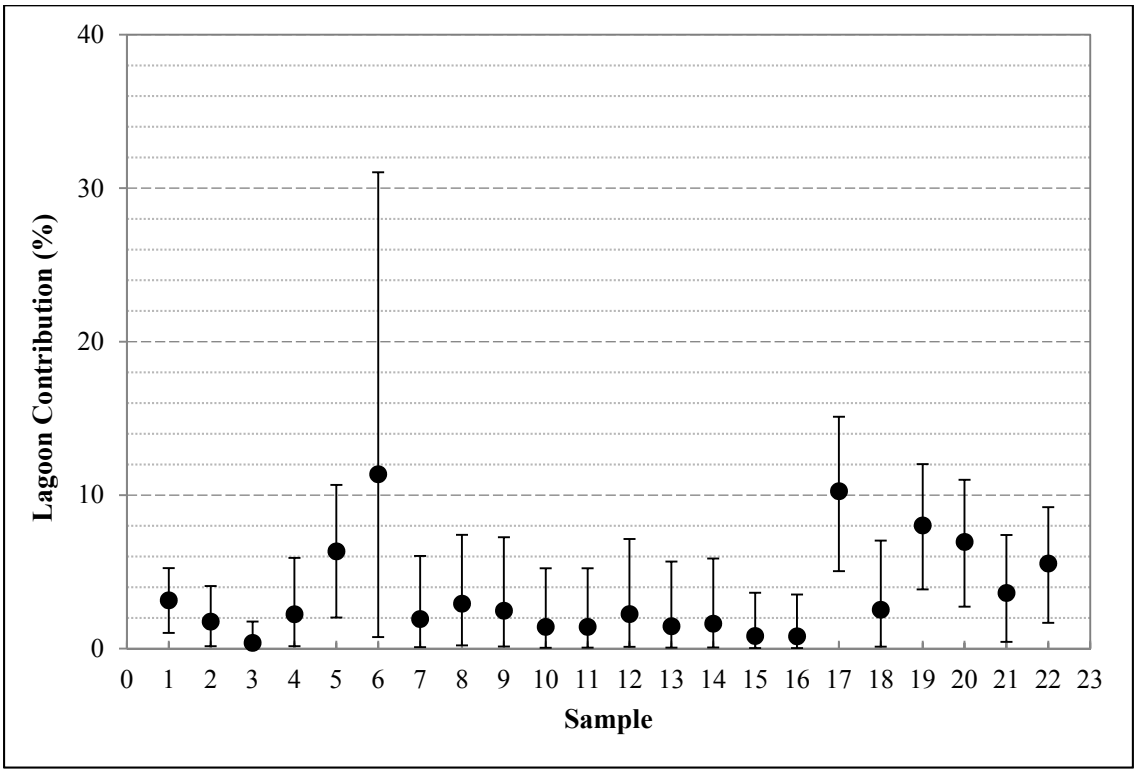


Figure 4-1. Median source contribution and CI of lagoon.

Figure 4-2 below displays the median source contribution of the upstream river and associated 95% credible intervals.

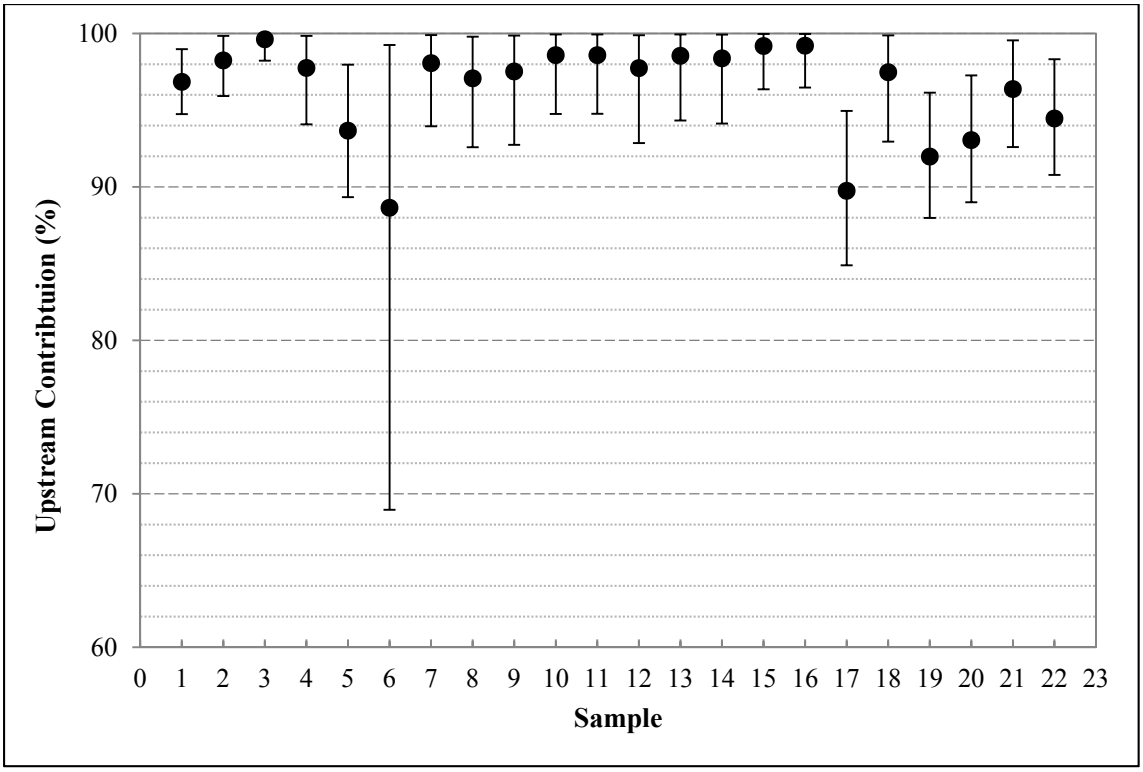


Figure 4-2. Median source contribution and CI of upstream.

The average of the median values for the estimated true fingerprint of the upstream is 96.4%, and that of the lagoon is 3.6%. This indicates that approximately 3.6% of the organic matter content of the downstream location is from a source other than the upstream location, and can be attributed to the influence of the lagoons. The average of the lower limit for the credible interval is 0.86%, and that of the upper limit is 8.07%. That is, there is a 95% probability that the percentage contribution of the lagoon to the river is between 0.86% and 8.07%.

Figure 3-6, displayed in Section 3.4.1, is a trace plot of the calculated values of P_1 for one EEM graph for 10,000 iterations, beginning after a burn-in period of 2,000 iterations and continuing to 12,000 iterations. This figure is printed again below in Figure 4-3 for reference.

The trace plot displays the 10,000 iterations drawn from the posterior distributions. This trace plot and the following histogram are for the calculated contribution of the upstream river for the sample collected on 30 August 2013.

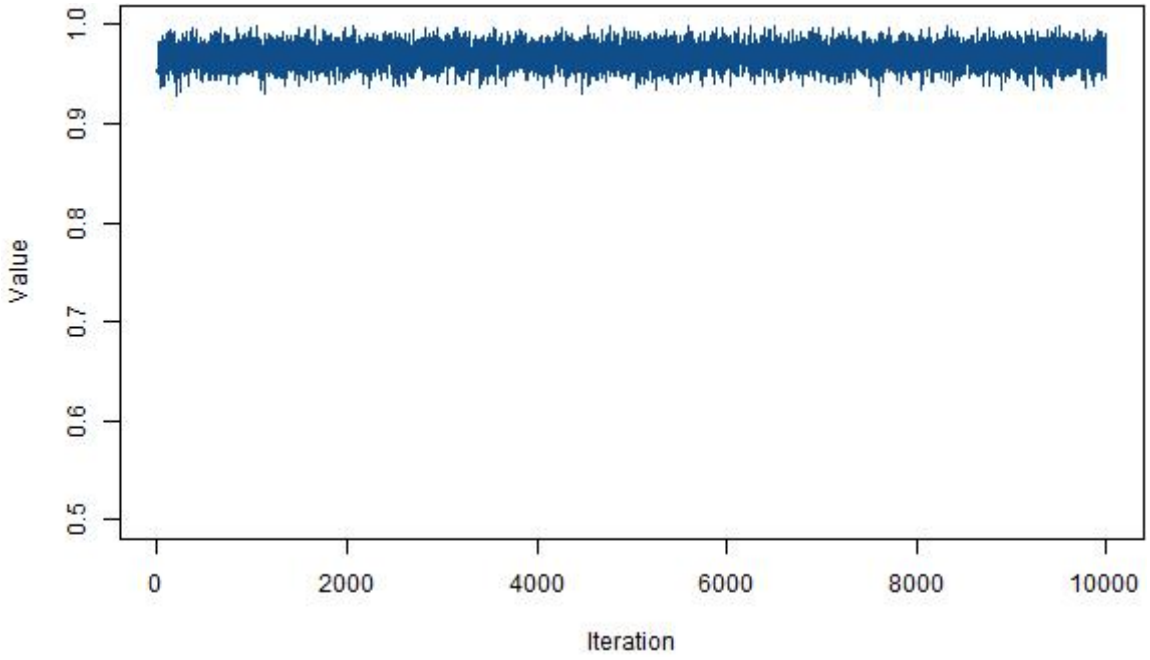


Figure 4-3. Trace plot of P_1 for 10,000 iterations.

Figure 4-4 on the following page is a histogram of P_1, or the upstream contribution, displaying the posterior distribution of the iterations.

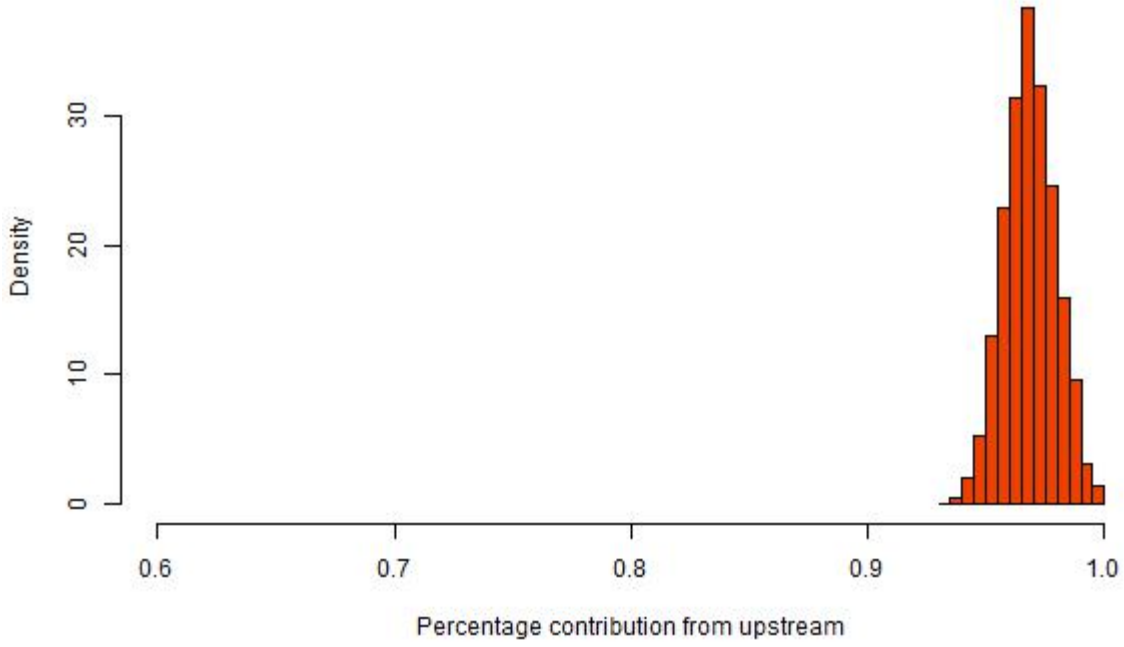


Figure 4-4. Posterior distribution histogram of P_1.

5 DISCUSSION

The Bayesian analysis calculates the percentage DOM contribution of the HVSSD lagoons to the Provo River as being 3.6%. This percentage contribution is based on a mass balance. This mass balance is different than a volume balance, where the equation is based on volumetric flow rates. Assuming that concentration of DOM is proportional to fluorescence intensity, a straightforward mass balance equation can be constructed, where the flow rate of the Provo River and measured fluorescence intensities can be used to determine an approximate seepage flow from the HVSSD lagoons. This equation would appear as below.

$$Q_U I_U + Q_L I_L = Q_D I_D \quad (3)$$

Where, for the equation above, the following notation is used

Q_U: volumetric flow rate of the upstream sampling location
I_U: fluorescence intensity of the upstream sampling location
Q_L: volumetric flow rate of lagoon cell 3 seepage
I_L: fluorescence intensity of lagoon cell 3 seepage
Q_D: volumetric flow rate of the downstream sampling location
I_D: fluorescence intensity of the downstream sampling location

Equation (3) uses fluorescence intensity in place of concentration of DOM. The volumetric flow rate of the upstream sampling location is obtained from USGS gage data for the Provo River in Heber. According to USGS data, the average flow rate over the months samples were collected is around 150 cfs, or 97 MGD.

To solve the above equation, the intensity of all 14 included wavelengths was averaged for each sample, and the average of those averaged intensities for all 22 samples was then determined. The average downstream fluorescence intensity (I_D) was 186.1, the average lagoon fluorescence intensity (I_L) was 521.9, and the average upstream fluorescence intensity (I_U) was 184.8. Taking $Q_U = 97\ MGD,\ Q_L = unknown,\ Q_D = (97 + Q_L)$, and substituting them into Equation (3) above, Q_L was calculated to be 0.375 MGD. This calculated volumetric flow rate, or contribution of the lagoon, is below the average influent rate of the HVSSD of 2.0 MGD. This simple mass balance check can be used to determine an approximate seepage flow from the HVSSD lagoons.

The Bayesian CMB model calculated an average percentage DOM contribution of the lagoon to the Provo River of 3.6%. This percentage, and the calculated volumetric flow rate of 0.38 MGD obtained as described above, are reasonable given that the average influent flow rate of the HVSSD is approximately 2.0 MGD.

This model is based on a few assumptions as described previously. If these assumptions are not met then the CMB model is violated and would yield untenable results. One assumption is that the fluorescent surface of the EEM is not changed over time. The other assumption is that any change in the DOM content of the downstream sample is due to the influence of the HVSSD lagoons. This assumption excludes any external sources of DOM, such as those from surface runoff.

This research does not violate either of those two assumptions, so those are not investigated further. It is notable, however, that though the results are justifiable, it is possible that the underlying assumptions of the CMB model are still not completely met.

This Bayesian CMB model appears to be a viable method for estimating the influence of a sewage lagoon on an adjacent river. There is some change in the DOM content between the upstream and downstream locations, and this is most likely from seepage from the HVSSD lagoons.

5.1 **Considerations**

One factor to consider is the dilution of the Lagoon samples. The Lagoon 3 samples were diluted ½ so that the measured intensity would not exceed the measurable limit of 1000. The dilution was performed with utmost care to ensure that contamination of the sample was avoided and that accurate dilution was performed. Nevertheless, this process of dilution could result in EEMs not truly representative of Lagoon 3 fluorescence.

Another factor to consider is the sampling frequency. Samples were not taken on a consistent frequency, and therefore the sampling dates vary across the span of several months. This, however, does not detract from the ability of the Bayesian analysis to calculate source contribution.

6 CONCLUSIONS

The purpose of this research was to determine if fluorescence spectroscopy methods and Bayesian statistical analysis could identify the amount of wastewater seepage from the HVSSD lagoons into the Provo River. This research indicates that this method is viable in performing source apportionment, and can allow researchers to effectively determine the degree to which a water source has been contaminated by a pollution source.

This research has application in determining the affect sanitary wastewater lagoons, swine or poultry operation lagoons, and other sewage lagoons have on an adjacent river due to groundwater seepage. The method used can be applied in scenarios where direct collection of hydrogeologic data is not possible, allowing researchers to determine the presence of unknown pollution using water samples collected from the water bodies in question.

The Bayesian statistical methods employed in this research demonstrate the ability to determine the degree to which a water source has been contaminated by a pollution source such as a wastewater lagoon, and to effectively perform source apportionment without requiring flow measurements. Bayesian analysis of EEMs is capable of using fluorescence spectroscopy, used as an indicator of organic matter, to perform source apportionment. Additionally, as discussed previously, a Bayesian approach offers certain advantages over other statistical methods which are used in source apportionment studies.

This research demonstrates the validity of the Bayesian CMB model in determining the influence of a sewage lagoon on an adjacent river. This research also demonstrates that fluorescence spectroscopy methods and Bayesian statistical analysis can effectively perform source apportionment.

REFERENCES

Baker, A. (2001). "Fluorescence Excitation-Emission Matrix Characterization of Some Sewage-Impacted Rivers." Environmental Science and Technology **35**: 948-953.

Baker, A. (2001). "Fluorescence excitation-emission matrix characterization of some sewage-impacted rivers." Environmental Science & Technology **35**(5): 948-953.

Baker, A. (2004). "Protein-like fluorescence intensity as a possible tool for determining river water quality." Hydrological Processes **18**: 2927-2945.

Boehme, J., P. Coble, R. Conmy and A. Stovall-Leonard (2004). "Examining CDOM fluorescence variability using principal component analysis: seasonal and regional modeling of three-dimensional fluorescence in the Gulf of Mexico." Marine Chemistry **89**(1): 3-14.

Christensen, W. F. and R. F. Gunst (2004). "Measurement error models in chemical mass balance analysis of air quality data." Atmospheric Environment **38**(5): 733-744.

Coble, P. G. (1996). "Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy." Marine Chemistry **51**(4): 325-346.

de Vos, A. F. (2004). "A primer in Bayesian Inference." preprint.

District, Heber Valley Special Service (2000). Heber Valley Special Service District Wastewater Treatment Facilities. Midway, UT.

District, Heber Valley Special Service (2011). Water Quality Board Request for Hardship Planning Advance to Study the Potential for Phosphorus Loading to the Provo River from Rapid Infiltration Basins Authorization. Midway, UT.

Eckhoff, D., A. Boyd, R. King and B. Wheeler (2004). 2004 Water Quality Implementation Report: Deer Creek and Jordanelle Reservoirs Water Quality Management Plan for October 2002 through September 2003, The Wasatch County Council.

Ferrell, C. E. (2009). Fluorescence Spectrophotometry and the Mass Balance: Exitation-Emission Matrices. Master of Science, Brigham Young University.

Friedlander, S. K. (1973). "Chemical element balances and identification of air pollution sources." Environmental Science & Technology **7**(3): 235-240.

Ghervase, L., E. Carstea, G. Pavelescu, E. Borisova and A. Daskalova (2010). "Fluorescence evaluation of anthropogenic influence on rivers crossing Sofia." Romanian Reports in Physics **62**(1): 193-201.

Guo, W., J. Xu, J. Wang, Y. Wen, J. Zhuo and Y. Yan (2010). "Characterization of dissolved organic matter in urban sewage using excitation emission matrix fluorescence spectroscopy and parallel factor analysis." Journal of Environmental Sciences **22**(11): 1728-1734.

Hall, G. J., K. E. Clow and J. E. Kenny (2005). "Estuarial fingerprinting through multidimensional fluorescence and multivariate analysis." Environmental Science & Technology **39**(19): 7560-7567.

Hudson, N., A. Baker and D. Reynolds (2007). "Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review." River Research and Applications **23**(6): 631-649.

Hudson, N., A. Baker, D. Ward, D. M. Reynolds, C. Brunsdon, C. Carliell-Marquet and S. Browning (2008). "Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England." Science of The Total Environment **391**(1): 149-158.

Hylland, M. D. (1995). Suitability for Wastewater Disposal in Septic-Tank Soil-Absorption Systems, Western Wasatch County, Utah. Plates 3A-3D. Utah Geological Survey, Utah Geological Survey.

Israel, D. W., W. J. Showers, M. Fountain and J. Fountain (2005). "Nitrate Movement in Shallow Ground Water from Swine-Lagoon-Effluent Spray Fields Managed under Current Application Regulations Mention of trademark names does not represent an endorsement over any other products by USDA-ARS." Journal of Environmental Quality. **34**(5): 1828-1842.

Jepson, R. W., J. McLean, C. G. Clyde and S. Korom (1991). Studies Related to Nutrients Entering Groundwater from the Heber Valley Sewer Farm and Dairies. Logan, UT, Utah State University.

Karanasiou, A., P. Siskos and K. Eleftheriadis (2009). "Assessment of source apportionment by Positive Matrix Factorization analysis on fine and coarse urban aerosol size fractions." Atmospheric Environment **43**(21): 3385-3395.

Karr, J. D., W. J. Showers, J. W. Gilliam and A. S. Andres (2001). "Tracing nitrate transport and environmental impact from intensive swine farming using delta nitrogen-15." Journal of Environmental Quality **30**(4): 1163-1175.

Kowalczuk, P., M. J. Durako, H. Young, A. E. Kahn, W. J. Cooper and M. Gonsior (2009). "Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Interannual variability." Marine Chemistry **113**(3): 182-196.

Lingwall, J. W., W. F. Christensen and C. S. Reese (2008). "Dirichlet based Bayesian multivariate receptor modeling." Environmetrics **19**(6): 618-629.

Lochmueller, C. H. and S. S. Saavedra (1986). "Conformational changes in a soil fulvic acid measured by time-dependent fluorescence depolarization." Analytical Chemistry **58**(9): 1978-1981.

Massoudieh, A., A. Gellis, W. S. Banks and M. E. Wieczorek (2013). "Suspended sediment source apportionment in Chesapeake Bay watershed using Bayesian chemical mass balance receptor modeling." Hydrological Processes **27**(24): 3363-3374.

Massoudieh, A. and M. Kayhanian (2013). "Bayesian Chemical Mass Balance Method for Surface Water Contaminant Source Apportionment." Journal of Environmental Engineering **139**(2): 250-260.

Miller, M. S., S. K. Friedlander and G. M. Hidy (1972). "A chemical element balance for the Pasadena aerosol." Journal of Colloid and Interface Science **39**(1): 165-176.

Mopper, K., Z. Feng, S. B. Bentjen and R. F. Chen (1996). "Effects of cross-flow filtration on the absorption and fluorescence properties of seawater." Marine Chemistry **55**(1): 53-74.

Murphy, K. (2010). "Measurement of Dissolved Organic Matter Fluorescence in Aquatic Environments: An Interlaboratory Comparison." Environmental Science and Technology **44**: 9405-9412.

Patel-Sorrentino, N., S. Mounier and J. Y. Benaim (2002). "Excitation–emission fluorescence matrix to study pH influence on organic matter fluorescence in the Amazon basin rivers." Water Research **36**(10): 2571-2581.

Persson, T. and M. Wedborg (2001). "Multivariate evaluation of the fluorescence of aquatic organic matter." Analytica Chimica Acta **434**(2): 179-192.

Plummer, M. (2013) "JAGS Version 3.4.0 user manual."

Quality, Division of Water (2014). Information regarding a proposed Operating Permit Program for Total Containment Lagoons, Utah Department of Environmental Quality.

Quality, State of Utah Divison of Water (2011). General Permit for Land Disposal of Municipal Wastewater. Department of Environmental Quality. Salt Lake City, Utah.

Ramadan, Z., B. Eickhout, X.-H. Song, L. Buydens and P. K. Hopke (2003). "Comparison of positive matrix factorization and multilinear engine for the source apportionment of particulate pollutants." Chemometrics and Intelligent Laboratory Systems **66**(1): 15-28.

Reichard, J. S. and C. M. Brown (2009). "Detecting groundwater contamination of a river in Georgia, USA using baseflow sampling." Hydrogeology Journal **17**(3): 735-747.

Rules, Utah Divison of Administrative (2014). Design Requirements for Wastewater Collection, Treatment and Disposal Systems. Utah Division of Administrative Rules. Salt Lake City, UT. **R317-3**.

Sharifi, S., M. M. Haghshenas, T. Deksissa, P. Green, W. Hare and A. Massoudieh (2013). "Storm Water Pollution Source Identification in Washington, DC, Using Bayesian Chemical Mass Balance Modeling." Journal of Environmental Engineering.

Sloan, A., J. Gilliam, J. Parsons, R. Mikkelsen and R. Riley (1999). "Groundwater nitrate depletion in a swine lagoon effluent-irrigated pasture and adjacent riparian zone." Journal of soil and water conservation **54**(4): 651-656.

Soonthornnonda, P. and E. R. Christensen (2008). "Source apportionment of pollutants and flows of combined sewer wastewater." Water Research **42**(8): 1989-1998.

Stedmon, C. A., S. Markager and R. Bro (2003). "Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy." Marine Chemistry **82**(3): 239-254.

Venables, W. N. and D. M. Smith (2014) "An Introduction to R: Notes on R, A Programming Environment for Data Analysis and Graphics."

Westerhoff, P., W. Chen and M. Esparza (2001). "Fluorescence analysis of a standard fulvic acid and tertiary treated wastewater." Journal of Environmental Quality **30**(6): 2037-2046.

Winchester, J. W. and G. D. Nifong (1971). "Water pollution in Lake Michigan by trace elements from pollution aerosol fallout." Water, Air, and Soil Pollution **1**(1): 50-64.

Yan, Y., H. Li and M. L. Myrick (2000). "Fluorescence Fingerprint of Waters: Excitation-Emission Matrix Spectroscopy as a Tracking Tool." Applied Spectroscopy **54**(10): 1539-1542.

Yang, D. (2013). Final Summary Report Hydrogeologic Site Characterization and Groundwater Monitoring Proposed Rapid Infiltration Basins. Draper, UT, Sunrise Engineering.

APPENDIX A. DATA

A.1 Data

Table A1. Fluorescence intensity measurements for downstream samples.

		Fluorescence intensity													
		Excitation wavelength, Emission wavelength													
Location	Date	280, 355	340, 430	330, 420	350, 420	330, 430	350, 430	330, 440	350, 440	275, 340	275, 310	260, 380	340, 420	280, 350	280, 360
Downstream	8/30/13	136.9	298.2	304.5	252.7	316.5	274.4	307.9	279.0	108.2	91.6	254.1	282.0	128.7	149.1
Downstream	9/4/13	125.7	279.6	283.8	232.0	298.2	254.8	292.9	260.5	96.7	88.9	263.2	262.2	116.8	137.8
Downstream	2/19/14	302.3	215.0	255.9	179.5	252.8	190.6	235.8	190.0	119.9	2.4	214.5	207.6	242.4	358.1
Downstream	2/19/14	289.1	213.6	255.7	179.0	250.9	190.1	235.2	190.2	114.8	2.1	206.0	206.5	233.1	342.7
Downstream	2/25/14	272.3	222.1	269.1	182.5	264.0	194.4	246.9	194.5	110.1	2.3	194.8	214.9	219.2	322.8
Downstream	2/25/14	250.9	205.6	244.7	169.4	242.5	181.8	225.2	183.4	100.1	2.1	180.0	198.4	203.4	296.5
Downstream	2/28/14	239.8	194.6	228.7	161.2	226.2	173.4	212.0	174.3	95.0	2.1	180.5	186.8	192.6	286.6
Downstream	2/28/14	229.9	195.5	226.0	165.1	225.9	176.6	210.1	176.7	90.6	2.1	170.4	188.5	183.2	272.2
Downstream	3/21/14	258.0	177.0	214.5	144.8	210.2	155.0	197.2	155.0	103.3	2.3	192.6	170.7	206.1	307.6
Downstream	3/21/14	262.8	174.7	215.2	144.9	210.3	153.2	195.7	153.7	104.9	2.2	197.2	169.4	210.8	309.8
Downstream	3/24/14	244.3	167.7	205.6	139.3	201.9	147.9	188.1	149.1	98.0	2.0	183.0	162.9	194.2	290.0
Downstream	3/24/14	234.1	165.3	200.1	135.9	197.9	144.7	185.1	146.5	92.8	2.1	175.1	159.5	186.6	277.1
Downstream	3/25/14	213.9	160.7	191.3	132.7	189.0	141.1	177.0	142.9	84.5	1.8	156.7	155.2	170.0	251.8
Downstream	3/25/14	206.8	161.8	191.6	133.7	189.9	142.7	177.7	143.4	81.1	1.7	154.3	156.0	166.5	246.4
Downstream	4/1/14	221.8	167.0	196.7	136.9	194.1	146.0	183.8	147.4	87.5	1.8	160.6	160.9	175.8	263.4
Downstream	4/1/14	214.9	171.8	199.6	140.9	198.4	150.7	186.8	153.3	84.6	1.8	158.5	164.3	171.4	254.1
Downstream	4/8/14	251.9	196.3	244.3	158.0	241.5	170.0	225.0	172.2	99.4	2.0	186.3	187.2	203.2	300.1
Downstream	4/8/14	226.3	164.8	202.7	134.3	201.1	144.6	187.8	145.8	88.7	1.9	165.7	158.1	181.0	268.8
Downstream	5/16/14	249.5	229.8	292.9	183.7	289.3	196.5	268.3	201.4	98.0	2.2	189.7	222.0	200.9	299.8
Downstream	5/16/14	240.0	248.4	272.9	205.6	272.4	219.4	256.1	221.0	94.4	2.1	179.6	238.2	192.7	287.9
Downstream	5/19/14	256.0	195.3	230.4	163.7	228.4	174.7	214.5	176.5	99.8	2.3	190.2	187.9	202.7	305.3
Downstream	5/19/14	252.9	207.8	243.3	175.4	240.2	186.3	226.1	187.5	101.3	2.3	191.2	201.3	202.7	303.8

Table A2. Fluorescence intensity measurements for Lagoon 3 samples.

		Fluorescence intensity													
		Excitation wavelength, Emission wavelength													
Location	Date	280, 355	340, 430	330, 420	350, 420	330, 430	350, 430	330, 440	350, 440	275, 340	275, 310	260, 380	340, 420	280, 350	280, 360
Lagoon 3	8/30/13	753.4	1148.5	1113.6	1047.5	1121.7	1109.6	1048.9	1072.7	619.2	374.8	909.3	1114.0	723.4	783.1
Lagoon 3	9/4/13	699.9	985.4	950.8	903.9	961.1	961.9	898.0	943.9	565.0	347.5	857.8	950.4	673.2	728.3
Lagoon 3	2/19/14	591.7	784.7	783.6	748.3	765.1	791.8	701.0	771.9	238.9	6.6	420.9	772.2	475.6	698.5
Lagoon 3	2/19/14	597.2	754.8	763.8	721.9	737.8	757.1	679.0	739.1	241.9	6.9	425.4	743.5	476.3	701.3
Lagoon 3	2/25/14	518.9	724.3	717.5	694.5	698.1	732.4	642.3	714.4	213.7	6.0	374.5	714.5	422.2	611.0
Lagoon 3	2/25/14	280.1	260.8	258.4	278.2	254.2	275.7	244.7	264.5	128.3	11.5	350.3	260.9	225.9	343.5
Lagoon 3	2/28/14	452.9	637.9	633.2	613.2	618.1	649.5	570.4	631.4	185.7	5.9	339.7	624.2	365.7	538.2
Lagoon 3	2/28/14	471.0	640.2	637.1	612.3	620.5	647.4	570.9	630.8	189.8	5.3	351.8	626.5	377.9	553.2
Lagoon 3	3/21/14	463.0	563.0	570.5	532.9	560.3	565.3	519.4	562.8	187.6	5.3	346.9	549.3	370.4	546.6
Lagoon 3	3/21/14	436.0	554.4	556.8	526.2	546.7	558.4	509.7	546.0	179.4	5.3	327.8	540.2	351.1	511.7
Lagoon 3	3/24/14	453.4	525.2	536.7	494.1	527.5	523.0	489.5	513.7	184.9	5.1	341.0	511.3	366.6	542.9
Lagoon 3	3/24/14	441.0	536.5	546.3	509.4	537.2	538.4	499.9	531.2	179.7	4.9	329.0	523.5	351.2	521.4
Lagoon 3	3/25/14	406.8	534.7	537.3	508.8	528.3	539.3	488.7	532.3	161.3	4.9	304.2	520.4	325.2	478.8
Lagoon 3	3/25/14	396.1	529.3	532.1	505.8	522.0	537.4	485.2	526.8	160.2	4.9	295.2	516.3	319.4	469.7
Lagoon 3	4/1/14	467.2	567.5	571.3	539.1	562.5	570.2	523.0	560.8	189.7	5.4	348.5	553.8	374.5	547.3
Lagoon 3	4/1/14	440.1	573.4	573.6	548.9	565.9	584.6	525.1	571.8	177.5	5.5	322.4	560.7	356.8	521.1
Lagoon 3	4/8/14	468.2	602.2	600.9	581.8	593.9	609.9	547.9	599.6	195.0	6.1	355.8	591.1	376.9	552.3
Lagoon 3	4/8/14	490.8	571.2	581.9	546.2	570.0	575.1	528.1	564.0	199.9	5.9	363.2	558.5	394.9	579.2
Lagoon 3	5/16/14	476.5	784.8	750.7	753.3	748.9	804.2	700.0	791.1	193.0	6.9	365.2	757.9	384.2	567.0
Lagoon 3	5/16/14	445.0	797.7	751.6	769.6	751.9	825.6	703.8	816.8	184.4	7.0	339.3	768.4	357.2	530.3
Lagoon 3	5/19/14	478.2	795.9	750.9	768.1	754.5	819.7	704.5	811.8	194.4	7.3	366.9	767.1	380.8	564.0
Lagoon 3	5/19/14	436.8	801.8	750.8	774.6	747.9	832.4	702.8	821.3	179.1	7.2	335.3	769.6	350.3	515.7

Table A3. Fluorescence intensity measurements for upstream samples.

		Fluorescence intensity													
		Excitation wavelength, Emission wavelength													
Location	Date	280, 355	340, 430	330, 420	350, 420	330, 430	350, 430	330, 440	350, 440	275, 340	275, 310	260, 380	340, 420	280, 350	280, 360
Upstream	8/30/13	115.2	285.7	290.0	236.6	302.5	259.3	296.8	263.8	86.5	79.5	243.9	267.3	105.9	128.1
Upstream	9/4/13	114.3	286.3	290.3	236.7	304.8	260.7	300.4	267.2	85.5	79.4	241.8	265.2	104.9	126.3
Upstream	2/19/14	276.7	276.3	322.9	239.1	317.2	254.8	292.6	252.9	108.7	2.3	197.5	269.9	221.0	326.8
Upstream	2/19/14	274.0	215.4	256.6	180.6	253.1	191.9	234.9	191.4	110.0	2.2	195.9	208.9	222.1	327.4
Upstream	2/25/14	258.2	194.9	231.4	161.9	229.8	172.2	213.1	173.5	103.7	2.1	181.9	187.7	207.9	305.8
Upstream	2/25/14	265.6	196.1	232.8	162.5	230.1	173.3	213.5	173.8	105.9	2.1	188.3	188.8	213.0	311.4
Upstream	2/28/14	258.2	194.9	231.4	161.9	229.8	172.2	213.1	173.5	103.7	2.1	181.9	187.7	207.9	305.8
Upstream	2/28/14	222.9	192.8	223.5	161.6	223.2	172.9	207.4	172.5	87.5	1.7	165.5	185.2	179.0	263.4
Upstream	3/21/14	236.4	188.0	221.1	157.5	218.7	167.1	205.1	168.0	93.1	2.1	174.6	181.9	188.1	280.8
Upstream	3/21/14	259.4	194.3	229.9	163.2	226.2	174.8	211.3	174.9	103.0	2.1	195.0	187.4	209.0	311.5
Upstream	3/24/14	243.1	188.7	226.9	154.5	226.2	165.2	211.4	167.0	97.1	2.0	180.5	182.7	194.2	288.8
Upstream	3/24/14	240.3	167.9	206.6	138.5	202.7	147.0	190.2	148.8	95.2	2.2	178.9	162.0	192.4	284.3
Upstream	3/25/14	233.1	171.5	205.6	140.7	204.1	152.3	189.2	153.3	92.1	2.0	168.7	165.4	186.3	276.8
Upstream	3/25/14	236.0	166.9	204.3	137.4	201.4	147.2	186.8	149.0	94.7	1.9	174.5	161.6	189.1	280.7
Upstream	4/1/14	247.8	187.3	223.6	155.9	222.8	167.6	208.0	167.8	98.9	2.1	184.0	180.2	197.4	293.1
Upstream	4/1/14	246.7	195.4	223.9	163.2	222.4	175.1	209.5	176.5	98.6	1.9	184.2	186.7	200.1	292.9
Upstream	4/8/14	215.9	163.3	193.5	134.2	192.0	144.2	179.6	145.2	86.0	1.9	159.2	155.8	173.3	257.4
Upstream	4/8/14	225.6	165.9	198.1	137.5	197.7	147.4	186.4	148.1	88.8	1.8	166.2	159.7	178.8	265.1
Upstream	5/16/14	248.7	185.0	222.7	153.8	220.0	164.5	206.0	166.2	97.7	2.2	185.4	179.5	198.0	295.4
Upstream	5/16/14	238.2	203.3	238.2	172.0	235.7	183.2	220.1	185.4	93.6	2.2	178.7	196.6	188.8	283.4
Upstream	5/19/14	259.9	176.3	215.0	144.5	213.1	154.2	198.6	155.7	102.4	2.3	196.1	169.4	207.1	308.5
Upstream	5/19/14	253.8	179.3	214.4	147.5	212.6	158.5	199.2	158.6	99.7	2.4	187.5	171.6	202.1	299.0

A.2 Program Code for Bayesian Analysis

```
setwd("C:/Users/Owner/Documents/BYU/Research/Thesis/Fingerprinti
ng")
library("coda")
library("R2WinBUGS")
library("R2jags")

X <- read.csv("ThesisData2.csv",header=FALSE)

PRup <- as.matrix(X[45:66,])
Lagoon <- as.matrix(X[23:44,])
PRdown <- as.matrix(X[1:22,])

for(i in 1:22){

##  BELOW IS THE ANALYSIS FOR LAGOON DAY #n
day <- i

C <- PRdown
VC <- matrix(NA,22,14)
for (i in 1:22) VC[i,] <- apply(C,2,sd)

A <- matrix(NA,14,2)
A[,1] <- PRup[day,]
A[,2] <- Lagoon[day,]
VA <- matrix(NA,14,2)
VA[,1] <- apply(PRup,2,sd)/sqrt(nrow(A))
VA[,2] <- apply(Lagoon,2,sd)/sqrt(nrow(A))



y <- c(PRdown[day,])

x1 <- c(A[,1])
x2 <- c(A[,2])
uy <- c(VC[1,])
ux1 <- c(VA[,1])
ux2 <- c(VA[,2])

linreg <- function()
{
  for (i in 1:14)
  {
     y[i] ~ dgamma(a[i],b[i]);
```

```
    a[i] <- (mu[i])^2/(uy[i])^2;
    b[i] <- mu[i]/(uy[i])^2;
    mu[i] <- p*lam1[i] + (1-p)*lam2[i];
    lam1[i] ~ dgamma((x1[i])^2/(ux1[i])^2, x1[i]/(ux1[i])^2);
    lam2[i] ~ dgamma((x2[i])^2/(ux2[i])^2, x2[i]/(ux2[i])^2);
  }
    p ~ dbeta(1, 1);
}

filename <- file.path(".",'linreg.bug')
write.model(linreg,filename)

data <- c('y','x1','x2','uy','ux1','ux2')
parameters <- c('p','lam1','lam2')

linreg.sim <-
jags(data,inits=NULL,parameters,model.file='linreg.bug',

n.iter=12000,n.burnin=2000,n.chains=1,n.thin=1,working.directory
=".",

  )
names(linreg.sim)
names(linreg.sim$BUGSoutput$sims.list)

#Histogram of p,  Posterior dist'n for p (the prop for upstream
histgram <- hist(linreg.sim$BUGSoutput$sims.list$p, main="",
col="orangered2", xlab="Percentage", ylab="Density",
xlim=c(0.6,1))  # Posterior dist'n for p (the prop for upstream)
#Save the graph
dev.copy(jpeg,file=paste("c:/users/owner/documents/byu/research/
thesis/fingerprinting/data/histogram", day,
".jpeg"),width=500,height=250)#width=500,height=250, OR
width=500,height=250
dev.off()

#Posterior for p
post <- density(linreg.sim$BUGSoutput$sims.list$p)
plot(post, main="", xlab="Percentage", ylab="Density",
xlim=c(0.6, 1))#, ylim=c(0, 40))
polygon(post, col="chartreuse4", border="black")
#Save the graph
dev.copy(jpeg,file=paste("c:/users/owner/documents/byu/research/
thesis/fingerprinting/data/posterior", day,
".jpeg"),width=500,height=250)
dev.off()
```

```
#Trace plot for P
plot(linreg.sim$BUGSoutput$sims.list$p, type="l",
col="dodgerblue4", xlab="Iteration", ylab="Value",
ylim=c(0.5,1))  # Trace plot for p
#Save the graph
dev.copy(jpeg,file=paste("c:/users/owner/documents/byu/research/
thesis/fingerprinting/data/traceplot", day,
".jpeg"),width=500,height=250)
dev.off()
plot(linreg.sim$BUGSoutput$sims.list$lam1[,1], type="l")  #
Trace plot for the first element
    # of lam1 (i.e., the first component of the estimated true
fingerprint for upstream)



pmed <- median(linreg.sim$BUGSoutput$sims.list$p)  # Posterior
median (estimate) for p
write.table(pmed,
file=paste("c:/users/owner/documents/byu/research/thesis/fingerp
rinting/data/median", day, ".txt", sep=""))

# 95% Credible interval for p
lower <- quantile(linreg.sim$BUGSoutput$sims.list$p, .025)  #
Lower limit for credible interval
write.table(lower,
file=paste("c:/users/owner/documents/byu/research/thesis/fingerp
rinting/data/lower", day, ".txt", sep=""))

upper <- quantile(linreg.sim$BUGSoutput$sims.list$p, .975)  #
Upper limit for credible interval
write.table(upper,
file=paste("c:/users/owner/documents/byu/research/thesis/fingerp
rinting/data/upper", day, ".txt", sep=""))

}
```