May 2018

Discriminant analysis as a decision-making tool for geochemically fingerprinting sources of groundwater salinity and other work

Nathaniel Chien
Syracuse University

Abstract

The thesis presented herein is a compilation of two different research projects I have been fortunate enough to work on during my graduate career at Syracuse University. The first and most complete project is a data analysis study using linear discriminant analysis to differentiate between sources of groundwater salinity in water samples from shallow groundwater wells. It is a model validation study that builds on previous work spearheaded by the Earth Sciences Department at Syracuse University. It represents some of my best work performed at Syracuse and should be considered the bulk of my thesis submission. Due to successful publication of that research early into my graduate career I had the opportunity to work on another project. The Technical Supplement is a review of the work I have done in collaboration with The Nature Conservancy. The main project goal was to study the hydrologic effects that beaver dam analogues may have on an incised stream system and to understand the utility of drone-derived imagery for hydrologic modeling. During my time working on this project, a lot was learned about best practices and avenues of future research. To make sure that this knowledge is not lost, it is recorded here.

DISCRIMINANT ANALYSIS AS A DECISION-MAKING TOOL FOR GEOCHEMICALLY FINGERPRINTING SOURCES OF GROUNDWATER SALINITY AND OTHER WORK

by

Nathaniel Patrick Chien

B.S. College of William and Mary, 2016

THESIS
Submitted in partial fulfillment of the requirements for the degree of Master of Science in Earth Sciences

Syracuse University
May 2018

Acknowledgments

I am incredibly grateful to Syracuse University, The Department of Earth Sciences, and the EMPOWER Graduate Research Traineeship Program for providing me the opportunities that led to the included work. I want to thank my advisor, Dr. Laura Lautz, for her research guidance and career mentorship during my time in Syracuse. I want to thank my research group, Emily Baker, Amanda Schulz, and Robin Glas for their encouragement and shared commiseration. I want to thank Deanna McCay for her organization and career guidance. Lastly, to everyone else who helped me during my time in graduate school: sources of wisdom (Dr. Kelleher and Dr. Condon), the great professors (Dr. Kroll and Professor Wilson), sources of encouragement, the graduate student body, and my family. I will end with a favorite quote attributed to Borges as encouragement to future students, "My father showed me his library, which was very large, and told me to read whatever I wanted, but that if something bored me, I should immediately put it down."

Table of Contents

List of Figures

Main Text

Supporting Information

Technical Supplement

List of Tables

Main Text

Supporting Information

Technical Supplement

Main Text:

Discriminant analysis as a decision-making tool for geochemically fingerprinting sources of groundwater salinity

This section has been published as:

Chien, N.P. and L.K. Lautz. 2018. Discriminant analysis as a decision-making tool for geochemically fingerprinting sources of groundwater salinity. *Science of the Total Environment* 618: 379-387.

Abstract

Concern over contamination of groundwater resources in areas impacted by anthropogenic activities has led to an increasing number of baseline groundwater quality surveys intended to provide context for interpreting water quality data. Flexible screening tools that can parse through these large, regional datasets to identify spatial or temporal changes in water quality are becoming more important to groundwater scientists. One such tool, developed from previous work by the authors, makes use of linear discriminant analysis (LDA) to identify the most probable source of chloride salinity in groundwater samples based on their geochemical fingerprints. Here, we applied the model to a dataset of shallow groundwater with known sources of contamination compiled from two studies of groundwater quality in Illinois: Panno et al. (2005) and Hwang et al. (2015). By predicting the source of salinity in groundwater samples for which the sources of contamination are known, we validated model prediction-accuracy. Results show high classification accuracy for groundwater samples impacted by basin brines (e.g. deep saline groundwater) and road salt (>80%), with diminishing success for those impacted by organic sources of chloride, such as septic effluent and animal waste. Posterior probabilities, a statistic inherent to LDA, provide a proxy for prediction confidence that enables the model to be used for assessment and accountability measures, such as identifying parties responsible for contamination. LDA is complementary to fingerprinting using halogen ratios (e.g. Cl/Br) because it implicitly relies on halogen ratios for classification decisions while providing a clearer, more quantitative classification of contamination sources. Our model is ideal for regional assessment or initial screening of salinity sources in groundwater because it makes use of commonly measured solute concentrations in publicly available water quality databases.

1. Introduction

Identifying sources of chloride contamination in groundwater is difficult because multiple chloride sources may exist in any one area. Anthropogenic sources of chloride present a growing threat with increasing urbanization leading to greater application of road salt along with more concentrated sources of effluent from municipal wastewater and septic tanks (Kelly et al. 2012; Panno et al. 2006; Hwang et al. 2015; Kelly and Matos 2014, Thunqvist 2004). Water softening, detergents and bleach, as well as human and animal waste cause wastewater and septic effluent to be saline. Chloride also occurs naturally in the environment by way of basin brines (e.g. deep saline groundwater) that mix with valley aquifers, soil/rock-water interactions, and atmospheric deposition (Panno et al. 2005). High-volume hydraulic fracturing, used to extract natural gas trapped in low permeability shale layers, poses a more recent and contentious threat to groundwater quality. Brine contamination from hydraulic fracturing can originate from either wastewater spills during the transport and disposal of saline flowback water or the migration of brines stored in hydraulically fractured shale units to shallow aquifers (Vidic et al. 2013; Lauer et al. 2016). In addition, high density animal feeding operations can lead to elevated chloride concentrations in groundwater due to the accumulation of animal waste and subsequent seepage into the groundwater system (Becker et al. 2001, Panno et al. 2005). Other potential sources of contamination include agricultural runoff, mine drainage, salt-water intrusion, and industrial discharge (Kim et al. 2015, Panno et al. 2005).

Halite is an extremely soluble mineral. One of its constituent ions, chloride, behaves conservatively in the environment making it a useful tracer of contamination in waterways (Whittemore 1995; Panno et al. 2006). While chloride alone is not particularly toxic to humans at environmental concentrations, chloride contamination can lead to chronic toxicity for aquatic

biota and vegetation (Wilcox 1986, World Health Organization 2004; Corsi et al. 2010). Additionally, higher chloride in aquifers that serve as drinking water sources can lead to higher treatment costs or even aquifer abandonment. Identification of the source of chloride contamination is important for multiple reasons. Primarily, source identification is the first step in any future attempts at remediation. But also from a policy perspective, improved methods of contaminant identification provide a way to either enforce accountability or develop contaminant prevention measures.

Progress has been made in fingerprinting diffuse sources of salinity. Previous work has often relied on chemical characterization and comparison of contaminated waters with known contaminant sources (Panno et al. 2006; Whittemore 1995). Comparisons are made using either relatively routine water quality measurements (e.g. major ions, ionic ratios) or specialized chemical/biological measurements (e.g. isotopic ratios, trace elements, bacteria populations) (Chapman et al. 2012; Warner et al. 2012; Alawattegama et al. 2015; Lauer et al. 2016; DeSimone et al. 2009, Tran et al. 2015). Specialized measurements may provide definitive assessment of contamination, but they are rarely included in routine water quality sampling and thus present a limited opportunity for regional assessment using existing or routinely collected data. The abundance of standard groundwater solute concentrations provides an alternative, relatively inexpensive, and transferable methodology for assessing groundwater contamination.

Previous studies on identifying sources of chloride contamination have relied heavily on the chloride:bromide (Cl:Br) ratio of water samples (Whittemore 1995; Davis et al. 1998; Panno et al. 2006; Hogan et al. 2007, Mullaney et al. 2009; Katz et al. 2011; Johnson et al. 2015; Reilly et al. 2015, Hildenbrand et al. 2017). These two halides are conservative ions with soluble corresponding salts. The wide range of distinct values for Cl:Br in waters affected by different

sources of salinity make it a promising way to differentiate among contaminant sources (Whittemore 1995; Davis et al. 1998; Panno et al. 2006). Other ionic ratios have also been used with varying effectiveness. For instance, low measurements of SO_4:Cl ratios have been proposed as a way to differentiate oil/gas brines from other basin brine sources for waters with similar Cl:Br ratios (Whittemore 1995). Iodine, although not routinely measured, also has been proposed as a way to trace waters affected by basin brines due to its tendency to accumulate in marine organic matter (Lu et al. 2015). While theoretically effective, practical use of ionic ratios is limited by a reliance on subjective judgments of how to separate groundwater populations affected by different contaminants. While bivariate plots do provide effective means of visualizing certain trends (e.g. Cl:Br vs. Cl), the close coupling of different contaminant groups can lead to contradictory or equivocal determinations of salinity sources. An ideal fingerprinting tool will capture the distinct chemical signature of contaminated groundwater using readily available existing datasets and allow for quantifiable assessment of the accuracy of those predictions.

In previous studies, we applied a geochemical fingerprinting model to determine the most probable sources of salinity in groundwater samples with elevated chloride concentrations (Lautz et al. 2014, Gutchess et al. 2016). Our approach was to develop a discriminant analysis model that leveraged widely available information on solute concentrations in groundwater and contaminant sources. The machine learning classifier functioned through a process of dimensionality reduction, creating linear combinations of different solute concentrations that optimize separation of groundwater samples impacted by different contaminant sources. These separations, referred to as discriminants, became the basis for classifying high salinity groundwater samples as affected by a specific contaminant source (basin brines, road salt, septic

effluent or animal waste). In that prior work, we were able to corroborate model classifications using water quality indicators that were independent of the geochemical fingerprinting model. Groundwater samples classified as impacted by basin brines had statistically significant elevated methane concentrations relative to other groups. Similarly, we found that groundwater classified as impacted by septic effluent tended to have elevated levels of nitrate. Despite these successes, a more direct method for validating the model was needed. Although methane and nitrate concentrations provide indirect confirmation of model classification of salinity sources, groundwater samples classified in Lautz et al. (2014) did not have known sources of contamination, so there was no definitive way to validate the accuracy of model output for that dataset.

Our primary goals for this study were to validate the accuracy of the model with an alternative dataset and to demonstrate the model transferability to basins outside of the original Marcellus region. Panno et al. (2005) contains a comprehensive geochemical database on the chemistry of contaminant sources as well as groundwater samples from Illinois with known sources of contamination. By applying a recalibrated version of the Lautz et al. (2014) model to predict sources of salinity in groundwater from this new dataset, we have been able to demonstrate the accuracy of the model by comparing model classifications with known salinity sources in contaminated samples. Our results and the chemical framework that comprises them also indicate the strong potential for transferring our model to other datasets of contaminated groundwater affected by similar contaminant sources, particularly road salt and basin brines.

2. Methods

2.1 Model Framework

A brief survey of the model design will be given. For a more detailed description of the model please reference Lautz et al. (2014).

The model developed for this study is generated using the framework presented in Lautz et al. (2014), which is intended to be adaptable to different geographic regions, sources of contamination and measured solutes. The framework described here is transferable to other studies with minimal changes. Development of the geochemical fingerprinting model consists of two complementary algorithms: the creation of a synthetic high salinity training dataset and implementation of linear discriminant analysis (LDA). The synthetic high salinity dataset simulates the expected distribution of water quality data in groundwater impacted by specific sources of contamination. Water quality data in this case includes concentrations of Cl, SO4, Br, Na, K, Mg, Ca, Ba, and Sr, but the model is flexible to alternative solute choices. Sources of contamination considered in this study are road salt runoff, basin brines, and organic waste. Organic waste is defined as septic wastewater effluent or animal waste, which have similar geochemical characteristics.

Creation of the training dataset is based on a two component, conservative mixing model. One component of the mixing model is low salinity (i.e. pristine or uncontaminated) groundwater, whose geochemistry is based on groundwater samples that have chloride concentrations <20 mg/L. Low salinity is meant to be groundwater only affected by atmospheric deposition (Panno et al. 2006). Field observations of low salinity groundwater are provided as model input data by the user. Using the mean and covariance matrix of the observed log-transformed solute concentrations for low salinity groundwater, a larger (n = 3000) vector of low-salinity groundwater information are generated using a multivariate lognormal random

number generator. The second component of the mixing model is the salinity source, representing a random sample from one of a collection of chloride contaminant sources (in this case: basin brines, road salt, and organic waste). Observations of contaminant source chemistry are provided as input data by the user. Then, using the mean and covariance matrix of the observed log-transformed solute concentrations for the various end-members, larger vectors (n=3000) of contaminant end-members are generated using a multivariate lognormal random number generator.

From these two populations the chemical composition of synthetic high salinity groundwater samples ($GW_{high,i,j}$) were computed as:

$$GW_{high,i,j} = n_{i,j}\left(source_{i,j}\right) + \left(1 - n_{i,j}\right)GW_{low,i} \quad (1)$$

where $n_{i,j}$ is the percent of end-member j in sample i, $source_{i,j}$ is a vector of the solute concentrations in a randomly selected sample i of the end-member j, and $GW_{low,i}$ is a vector of the solute concentrations in a randomly selected low salinity groundwater sample i. Sample selection was random, with replacement. Mixing percentages $n_{i,j}$ were randomly varied following a uniform distribution and were capped at 20% since well water is unlikely to be comprised of greater than 20% septic effluent or animal waste (Lautz et al. 2014). A synthetic high salinity dataset was created for each separate end member using this method. Since this is a 2-component mixing model, multiple contamination sources were not considered. While real aquifers may consist of multiple contamination sources, model output (e.g. posterior probabilities), elaborated on later, can help assess whether this is a relevant concern for specific samples.

The synthetic high salinity datasets were combined into one complete dataset for model training. LDA determines the optimal combinations of solute concentrations that most effectively separate the synthetic high salinity data by source of contamination (Figure 1). Output from LDA was a set of k - 1 linear discriminant classifiers, where k is the number of classification groups (in this case, k = 3 contaminant sources). Backward feature selection sequentially removed solutes that decreased the classification accuracy of the model on the training dataset. 10-fold cross validation was used to assess model accuracy. Following model training, discriminant scores were computed for the observed samples with unknown sources of salinity, using the discriminant classifiers trained on the synthetic data. The discriminant scores were used to classify the observed high salinity dataset by the most probable source of salinity (Figure 1). Prior probabilities of each contaminated groundwater population, a parameter required to classify samples, were set equal. Although prior probabilities do influence LDA classifications, they are generally unknowable beforehand. However, if information is available to inform prior probability values, they can easily be adjusted in the model.

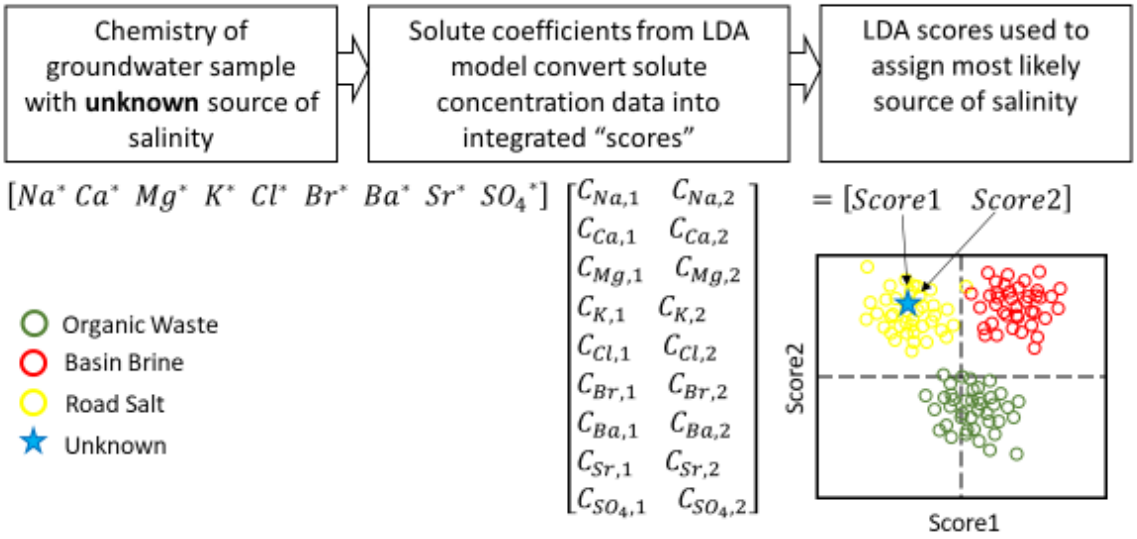


*Figure 1: Sequence of steps used to identify the most likely source of salinity in an unknown groundwater sample using the LDA model. In this case, the unknown would be classified as impacted by road salt. *Solute concentrations are log-transformed and normalized before LDA scores are calculated.*

2.2 Validation Dataset

The primary source of validation data was a geochemical database of water quality constituents found in Panno et al. (2005). An extensive collection of solute and groundwater quality information is found within this database for a variety of water sources throughout Illinois. Descriptions of sample sources and collection techniques are provided, making delineation of contaminant source samples and contaminated groundwater samples relatively clear and straightforward. A second smaller dataset of groundwater quality in Illinois was taken from a more recent publication by a subset of the same researchers, Hwang et al. (2015). Redundancies between the two datasets were removed. A complete listing of validation data can be found in Table S3.

In some cases, the latter paper had samples without a clearly defined source of contamination. These samples were generically labeled as either *Rural* or *Urban*. The author's analysis of *Urban* samples found a strong relationship between urban development and excessive application of deicing agents. Therefore, we concluded that those samples labeled as *Urban* could be considered contaminated by road salt. *Rural* samples were removed, except for ones with low salinity, because the source of high salinity in rural samples was ambiguous.

One of the most important considerations in using the new datasets for model development and validation is the designation of a given water sample as a source of contamination (input data for the mixing model) versus affected by contamination (validation data). Generally, distinguishing these sets is straightforward. Shallow groundwater samples can almost always be considered as observed groundwater and used in the validation dataset. Known samples of animal leachates, concentrated saline road salt solutions and septic effluent can be considered contamination sources. Similarly, produced water from an oil/gas well or a sample defined as formation water/basin brine would be designated as a contaminant source. However, some samples fall into a grey area between source of contamination and affected by contamination.

This concern was particularly relevant for those samples contaminated by basin brines in the Panno et al. (2005) database. For instance, one sample collected from a residential well had a chloride concentration of 17,000 ppm, a clear example of a groundwater sample that had gone from being "affected by contamination" to being more indicative of a "source of contamination". This sample was removed from the validation dataset because classifying it as affected by basin brines is unnecessary. Besides this residential well, groundwater samples contaminated with basin brines were only used if they were clearly defined as from a monitoring

or residential well. For clear examples of what samples were chosen as contaminants versus contaminated, please reference the *Supporting Information*.

2.3 Model Validation

The flexibility of our model allows for different numbers and types of end member compositions. The model was originally structured this way to maintain transferability among datasets having different measured water quality constituents, or in areas having different sources of contamination. For instance, while saltwater intrusion would be an important contaminant source to consider for coastal aquifer systems, this source can be neglected in Illinois. Here, observed groundwater data points were classified using three primary sources of contamination: basin brines, road salt, and organic waste. The organic waste dataset contains groundwater samples contaminated by private septic tanks as well as samples contaminated by animal waste leachate; for a complete discussion of this combination, please reference the *Supporting Information*. Generally, these three contaminant sources are some of the most common forms of shallow groundwater contamination and can be found in almost any geographic area within the United States and other densely populated regions (Panno et al. 2005; Mullaney et al. 2009; DeSimone et al. 2009).

While multiple trials with different combinations of solutes and end-members were explored, the model presented here consisted of the following solutes: Na, Cl, Br, K, Mg, Ca, Ba, Sr, SO_4. Although SO_4 may be significantly more reactive than the other solutes, it was included due to the abundance of SO_4 concentration measurements in groundwater quality data and its potential to differentiate animal waste from other contaminant sources (Whittemore 1995; Reilly et al. 2015). Additionally, linkages between high sulfate concentrations and basin brine

contamination have been observed in aquifers overlying geologic units actively involved in unconventional oil and gas extractions (Hildenbrand et al. 2017). Limited availability of nitrogen species in end-member datasets as well as their non-conservative nature in the environment prevented their use in the mixing model.

3. Results & Discussion

3.1 Classification Accuracy

Assessment of the accuracy of our geochemical fingerprinting model was done by comparing the model predictions of salinity sources with the known sources of contamination provided with the validation dataset. We know of few, if any, geochemical fingerprinting models that have been tested after their initial development to see how well they perform on independent, validation data. Confusion matrices for both the synthetic (i.e. training) and validation data are shown in Tables 1 and 2, respectively. The confusion matrix for the synthetic data reflects classification accuracy for the model training data. The training dataset is used to develop the model classifiers and, as a result, is not a wholly independent check of prediction-accuracy, although 10-fold cross-validation was used to develop the confusion matrix. In contrast, the confusion matrix for the validation data reflects classifications independent of data used in model development.

The diagonals of the matrices represent accurately classified samples for each contaminant source. The values populating the remainder of the table indicate which contaminant sources were commonly "confused" for one another. For both the synthetic and validation data matrices, basin brine had the highest classification accuracy with diminishing accuracy for road salt and organic waste.

Synthetic Training Data

Known Group	Predicted Group		
	basin brine	road salt	organic waste
basin brine	**96.2 (2840)**	1.1 (32)	2.7 (79)
road salt	5.1 (144)	**85.7 (2434)**	9.3 (263)
organic waste	1.8 (28)	6.1 (95)	**92.1 (1442)**

Table 1: Confusion matrix showing model classification accuracy for the synthetic training data. Numbers outside of parentheses are the percentage of samples classified correctly of the total samples contaminated by a single source (designated by the first column). Numbers within parentheses indicate total number of samples in each respective cell. Bold numbers along diagonals emphasize correct classifications. The synthetic data include 2951 basin brine, 2841 road salt, and 1565 organic waste contaminated samples. The synthetic-groundwater dataset size varies by contaminant due to the capped mixing percentage.

Validation Data

Known Group	Predicted Group		
	basin brine	road salt	organic waste
basin brine	**100 (7)**	-	-
road salt	0	**88.2 (15)**	11.8 (2)
organic waste	35.3 (6)	11.8 (2)	**52.9 (9)**

Table 2: Confusion matrix showing model classification accuracy for the validation data. Numbers outside of parentheses are the percentage of samples classified correctly of the total samples contaminated by a single source (designated by the first column). Numbers within parentheses indicate total number of samples in each respective cell. Bold numbers along diagonals emphasize correct classifications. The validation data include 7 basin brine, 17 road salt, and 17 organic waste contaminated samples.

All 7 samples known to be contaminated by basin brine were classified correctly, with a similarly high (96.2%) classification accuracy for the synthetic training data. The relatively high classification accuracy of basin brine samples illustrates that basin brine has a distinct chemical signature that LDA is able to distinguish from other contaminant sources. Road salt samples also had very high classification accuracy for the validation data with only 2 misclassifications out of a total of 17 samples. Samples contaminated by organic waste appear

relatively distinct from those contaminated by road salt or formation water, with a 92.1% classification accuracy for the organic waste training data. But, for the validation data, classification accuracy for samples contaminated by organic waste is the lowest of the three end-members.

While there were no false negative basin brine classifications for the model, there were six false positive basin brine classifications. That is, six samples not contaminated by basin brines were classified as such. This has important relevance to any kind of accountability measures. While groundwater contaminated by basin brines is unlikely to be misclassified, there is a chance that it will be misidentified. There is potential to differentiate false positives from true positives by looking at additional chemical parameters. For instance, all of the six false positives were actually contaminated by organic waste. On average, these samples all had significantly higher total nitrogen concentration (58.5 ppm) than the true basin brine samples (2.4 ppm) (Fig. S1). But, nitrogen concentrations are not universally high in misidentified samples, showing such information does not provide a definitive test for false positives.

If, when using this model, the issue of false positives is raised, group prior probabilities can be adjusted to lower the likelihood of any given sample being classified as basin brine. Prior probabilities represent the probability that a given sample, without any chemical information, will fall into a given class. Lowering the prior probability of basin brine is equivalent to lowering the likelihood that there is basin brine contamination in the study area. The result is a model less likely to classify an unknown as contaminated by basin brine. In the case of our model, lowering the prior probability of basin brines to 10% reduced the number of false positives for basin brine from 6 to 4. Further, lowering the prior probability to 4% reduced that to only 2 false positives. But, both these cases resulted in a groundwater sample actually contaminated by basin brine

getting misclassified as road salt (a false negative). Although we can test the effect of adjusting prior probabilities for our validation dataset, it is unlikely that information on prior probabilities would be available in other studies. While modifying prior probabilities slightly improved model performance in this case, prior probabilities should not be adjusted without strong justification.

Consideration of model-independent parameters, such as total nitrogen concentrations, or the model parameters specific to LDA, like prior probabilities, can be used to compare alternate model structures and view results from many different sides. However, whenever testing a dataset under different parameter combinations, reporting both manipulated- and default-parameter results is imperative for fair and transparent analysis. Additionally, LDA has other built-in metrics for examining the validity of a given classification. These will be explored in the following sections.

3.2 Posterior Probabilities

A limitation of many geochemical fingerprinting tools is the inability to quantify the level of confidence in a given prediction of a contaminant source. This limitation arises from the fact that models developed using, for example, bivariate plots of elements and elemental ratios (e.g. Cl:Br vs Cl plots) do not define boundaries between contaminant groups using numerical functions, but rather outlines of fields on the plot that are drawn empirically. In contrast, LDA classifications of contaminant sources are based on posterior probabilities; calculations of the likelihood that a sample falls within a given class (e.g. impacted by basin brines, road salt, or organic waste). While prior probabilities are input by the user, posterior probabilities are model output that is calculated as a function of the LDA scores (see later discussion) and prior probabilities. For every sample, LDA calculates the posterior probability of membership in each of the different

classes. The class with the highest posterior probability then becomes the group assigned to the sample. Comparisons of classification accuracy for different ranges of posterior probabilities reveal that these probabilities are useful gauges of prediction confidence (Fig. 2).

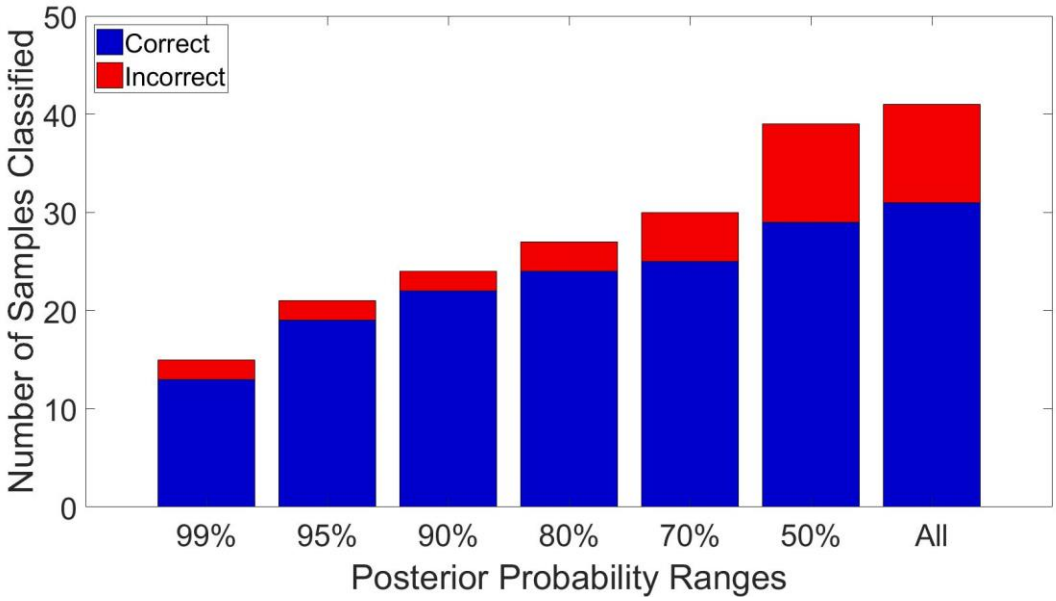


Figure 2. Classification accuracy for different posterior probability ranges. Each bar represents the total number of samples that have posterior probabilities greater than the x-axes percentage. The binary color scale delineates how many samples were classified correctly within each group.

For classifications with posterior probabilities greater than 90%, 95% and 99% there were 2 misclassifications and 22, 19 and 13 correct classifications, respectively. One additional misclassification and two additional correct classifications were added when the threshold was dropped to 80%. Limiting model results to those with high posterior probabilities (e.g. >90%) reduces the number of misclassifications, but also reduces the number of samples that are classified correctly because many correct classifications have posterior probabilities lower than 90%. So, there is a tradeoff in defining an acceptable limit for posterior probabilities. An optimal threshold value for posterior probabilities would be sufficiently high to minimize

misclassifications, but also sufficiently low to avoid limiting the number of samples classified by the model. Regardless of the threshold selected, it is important to understand the degree to which posterior probabilities reflect true model uncertainty.

Besides the 95% and 99% posterior probability ranges, all the other thresholds had posterior probabilities that closely resemble prediction confidence. For instance, 22-of-24 samples (92%) were classified correctly for probabilities greater than 90% and 24-of-27 (89%) for probabilities greater than 80%. These two posterior probability thresholds are well within the actual classification accuracy for samples within those ranges. Of course, low sample size limits the probabilities from being exact replicas of each other. But one would predict, as is the case for any collection of probabilistic events, that as sample size approaches infinity these two numbers would converge.

Posterior probabilities tend to act as conservative estimates of classification accuracy. For instance, the >50% range had the highest number of misclassifications of the ranges presented, 10-of-39 samples. This threshold provides a good cutoff for the utility of these probabilities. At this cutoff, of the set of samples with posterior probabilities greater than 50%, fewer than 50% of samples were misclassified. Below this cutoff, a single contaminant stops becoming the most likely source of contamination. Both samples below the 50% cutoff were incorrectly classified. For any group of classification results, it is unlikely for there to be more misclassifications than the 50, 70, 80 and 90% posterior probability range dictates. In cases where policy recommendations or accountability considerations need to be made, posterior probabilities provide an effective means of demonstrating the prediction confidence of any given LDA classification. While higher probabilities are desired, lowering the probability threshold

increases the number of classifications considered with comparatively little reduction in classification accuracy.

3.3 Linear Discriminants

Although LDA address limitations of other geochemical fingerprinting tools, the linear discriminants used for determining the most probable salinity source are less intuitive to understand than concentrations of individual solutes, or solute ratios. By examining the origin of the linear discriminants, we can gain better intuition for why the model is successful and what geochemical information drives model accuracy. Analysis of the linear discriminants (i.e. discriminant-scores), which are the variables responsible for classification, provides more information on the success of the model and allow us to visualize sources of misclassification.

Linear discriminants are linear combinations of the original chemical dataset that effectively reduce dataset dimensionality while optimizing the separation between classes. In the case of our model, the linear discriminant scores are computed as the sum of standardized, log-transformed solute concentrations that have been multiplied by a coefficient that is unique for each solute, while the classes represent the source of groundwater salinity. When the discriminant scores are plotted on a simple bivariate plot, the amount of overlap between data clouds reflects ambiguity of classification.

The discriminant scores divide the samples impacted by the three contaminants in two-dimensional space (Fig. 3). Score-1 (S1) is responsible for dividing the formation water samples ($S1<0$) from the road salt samples ($S1>0$). There is positive correlation between S1 and basin brine classifications, and negative correlation between S1 and road salt classification. Few training data points fall in the incorrect quadrant, reflecting the strong separation and high

classification accuracy for basin brine and road salt samples. Tables 1 & 2 reinforces the success of this separation. Only 176 of 5450 synthetic samples (3%) and 0-of-22 observed samples of basin brine or road salt contaminated groundwater were confused for each other. Score-2 (S2) separates the road salt (S2>0) from the organic waste class (S2<0). There is a negative correlation between score-2 and organic waste samples and a positive correlation between score-2 and road salt samples. While there is a clear separation between populations, neither axis acts as a definitive separator.

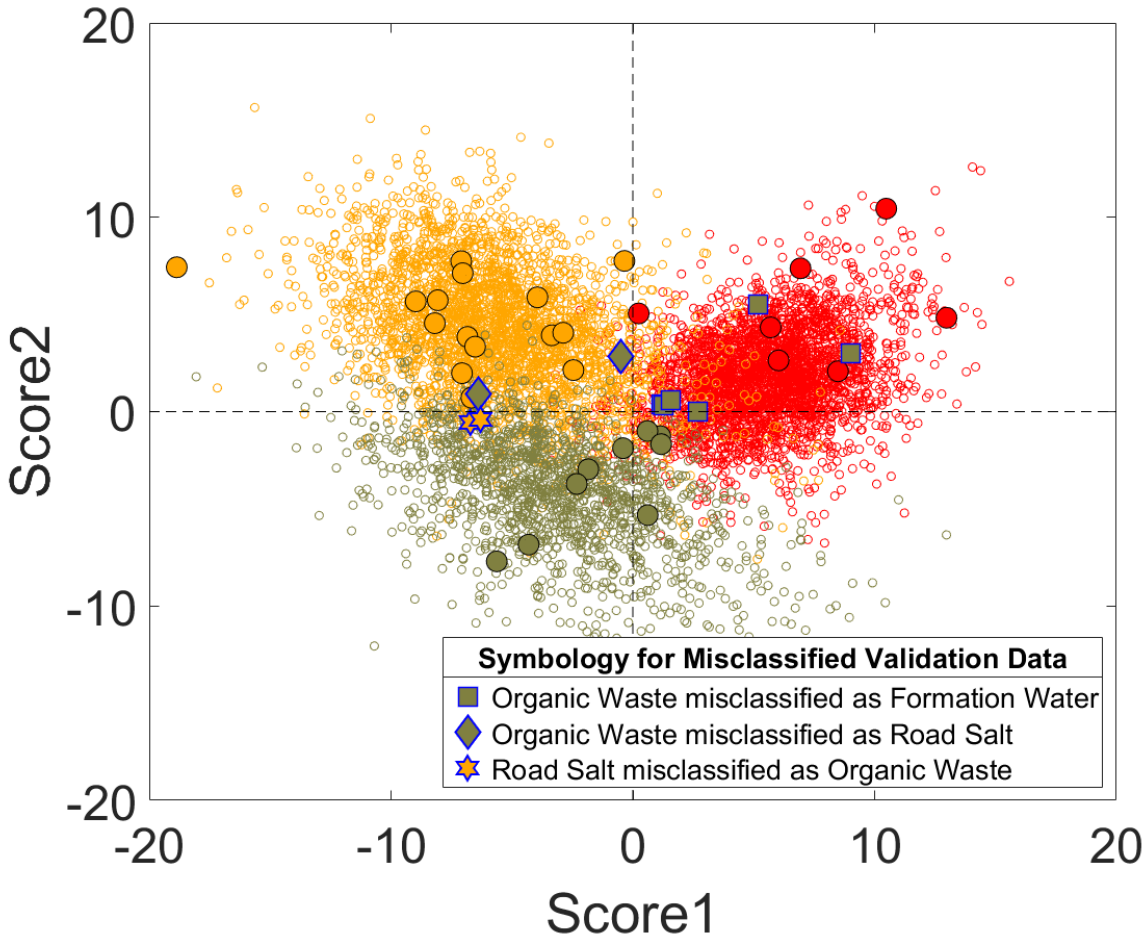


Figure 3. LDA scores for validation data (filled symbols) and synthetic training data (small open circles). For data symbolized as circles, colors represent known sources of contamination:

formation water (red), road salt (yellow), or organic waste (brown). The shapes that are not circles are misclassifications of validation data. Each shape represents a different type of misclassification delineated in the legend.

Figure 3 is a visualization of the confusion matrix for observed groundwater samples, in addition to the training data. Inherently, it is also a visualization of posterior probabilities. Samples that plot closer to the axes have lower posterior probabilities and therefore more closely resemble multiple contaminants. For model classification, only three of six possible sources of confusion were observed for the organic waste model. Road salt was confused as organic waste (2 samples) and organic waste was confused as either road salt (2) or formation water (6). Of the ten misclassifications, only two plot far from the axes and have high posterior probabilities as a result. This is corroborated by Figure 2 where two misclassifications were seen at the higher posterior probability range. All other misclassified samples plotted at the boundaries of synthetic groundwater clouds, near the axes. Either method, graphical or numerical (direct calculation of posterior probabilities), can be used to better understand the uncertainty of a classification.

3.4 Chemical Fingerprinting Contaminated Groundwater

Prior to this study, the most prominent method for differentiating groundwater contaminants involved examining variations in the chloride:bromide ratios (Whittemore 1995; Davis et al. 1998; Panno et al. 2006; Mullaney et al. 2009; Katz et al. 2011; Johnson et al. 2015; Reilly et al. 2015, Hildenbrand et al. 2017). Bromide and chloride are conservative ions that behave similarly in natural waters. However, their concentrations vary depending on their source. Basin brines have high bromide/chloride values relative to dissolved halite or modern oceans (Davis et al. 1998). During halite precipitation, bromide is excluded from the crystal lattice. This leads to enrichment of bromide in residual waters nearing the end of halite

precipitation such as in modern basin brines (Davis et al. 1998; Whittemore 1995). Basin brines, with a long residence time in groundwater, are enriched in bromide compared to groundwater that is saline due to halite dissolution (e.g. through contamination with road salt runoff). Ultimately, the distinctive ranges of bromide:chloride ratios in groundwater from different sources have made it the primary mechanism for identifying the origin of different high salinity groundwater samples (Davis et al. 1998).

Plots of chloride vs. chloride/bromide ratios can be used to qualitatively differentiate groups of groundwater samples (Whittemore 1995; Davis et al. 1998; Panno et al. 2006; Mullaney et al. 2009; Katz et al. 2011; Johnson et al. 2015; Reilly et al. 2015). These figures have been relatively useful in prior studies to map out major trends in groundwater chemistry from different sources. However, there remains a high level of ambiguity in making any kind of authoritative conclusion on the source of a contaminated groundwater sample from these plots, particularly at moderate salinities (e.g. 20-100ppm Cl).

Our use of LDA is founded on the same principle as plots of the chloride:bromide ratio. In some respects, LDA is an elaboration on this simple relationship, incorporating additional relationships that may exist between solutes. The chemical relationships that control how LDA separates contaminant groups can be determined in a number of different ways. One of the most straightforward approaches is to look at the linear combinations themselves, that is, the coefficients used to combine solute chemistry into an algebraic expression. Figure 4 visualizes these coefficients and Table 3 provides a summary. Direct comparison of the magnitude of these coefficients yields information on the relative importance of different solutes in constructing scores.

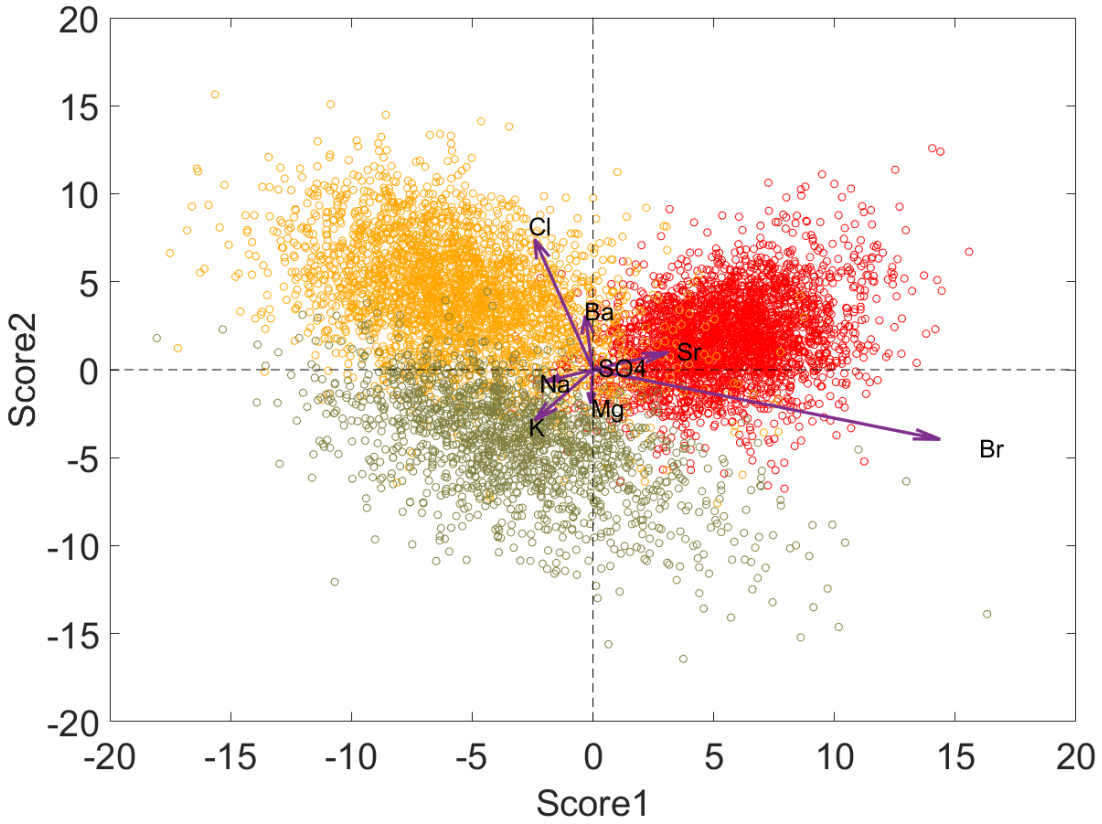


Figure 4. Vectors representing the relative weight of the solute coefficients used to compute linear discriminants. All vectors originate from the origin and point towards their weighted score-1, score-2 coefficients. Coefficients were multiplied by 2.5 for visual clarity. Label locations represent the endpoints of their respective weighted (x,y)-pair. E.g. 6.39 is the coefficient for the bromide term of score-1 and so appears as 16.0 on this figure.

Solute	Score-1 Coefficient	Score-2 Coefficient
Na	-0.89	-0.30
K	-1.07	-1.28
Mg	-0.04	-0.85
Cl	-1.07	3.29
Br	6.39	-1.76
Sr	1.39	0.44
Ba	-0.14	1.36
SO_4	0.09	0.07

Table 3. LDA coefficients that are multiplied by each solute concentration and summed for the 2 different scores. These coefficients provide a proxy for the influence of various solutes. See Figure 1 and 4 for further clarity.

The emphasis that LDA places on chloride and bromide is represented by the weights of their coefficients (Table 3). Chloride concentrations are weighted by 3.29 for score-2 and bromide by 6.39 for score-1 (Fig. 4). The least important solutes can be easily distinguished this way too. Sulfate for instance has coefficients below 0.10 and therefore is minimally involved in data transformation. The relative weight of the solute vectors can be used to determine which constituents are most important for fingerprinting contaminant sources, and which can be ignored with minimal impact on model results. This is an advantage of LDA because groundwater data available in existing public databases are missing some constituents, and therefore may be of different utility in the model. Also, these results can be used to prioritize the solute concentrations to be measured in future baseline water quality data acquisition.

The relative magnitudes of the different solute coefficients (Table 3), as well as the direction and length of the vectors in Fig. 4, provide information on the chemical nature of the different contaminated populations. For example, score-2 values are strongly and positively influenced by chloride concentrations. Low chloride concentrations contribute to low score-2 values, and classification of a sample as impacted by organic waste. This reflects the overall lower salinity of the organic waste end-member, relative to road salt runoff or basin brines. We simply do not expect groundwater contaminated by organic waste to have very high salinity because the organic waste that is the source of the contamination (e.g. septic effluent) does not have sufficiently high chloride concentrations to elevate the groundwater salinity above a few hundred milligrams per liter. In contrast, road salt runoff and basin brines can readily increase the salinity of fresh groundwater, even at low mixing percentages. Score-2 values are also negatively influenced by potassium concentrations. This reflects the fact that potassium is a nutrient commonly put into solution during the decomposition of organic matter. As a result,

high potassium concentrations at low chloride concentrations (or high K:Cl ratios) result in classification of a sample as impacted by organic waste.

Score-1 values, which separate road salt and basin brine classes, are strongly influenced by bromide concentrations (Fig. 4). High bromide concentrations relative to chloride (or high Br:Cl ratios) result in classification of a sample as impacted by basin brine. These relationships reflect the differences between road salt and basin brine geochemistry that are the result of the processes that drive their formation. Basin brines are the residual water remaining following evaporation and the resultant precipitation of halite. Since bromide is excluded from the halite crystal lattice, it is enriched in residual waters and basin brines. The consistency between our empirical model results (e.g. relative magnitude and sign of score coefficients) and what we expect for the chemical differences between end-members provides additional credibility to our model results.

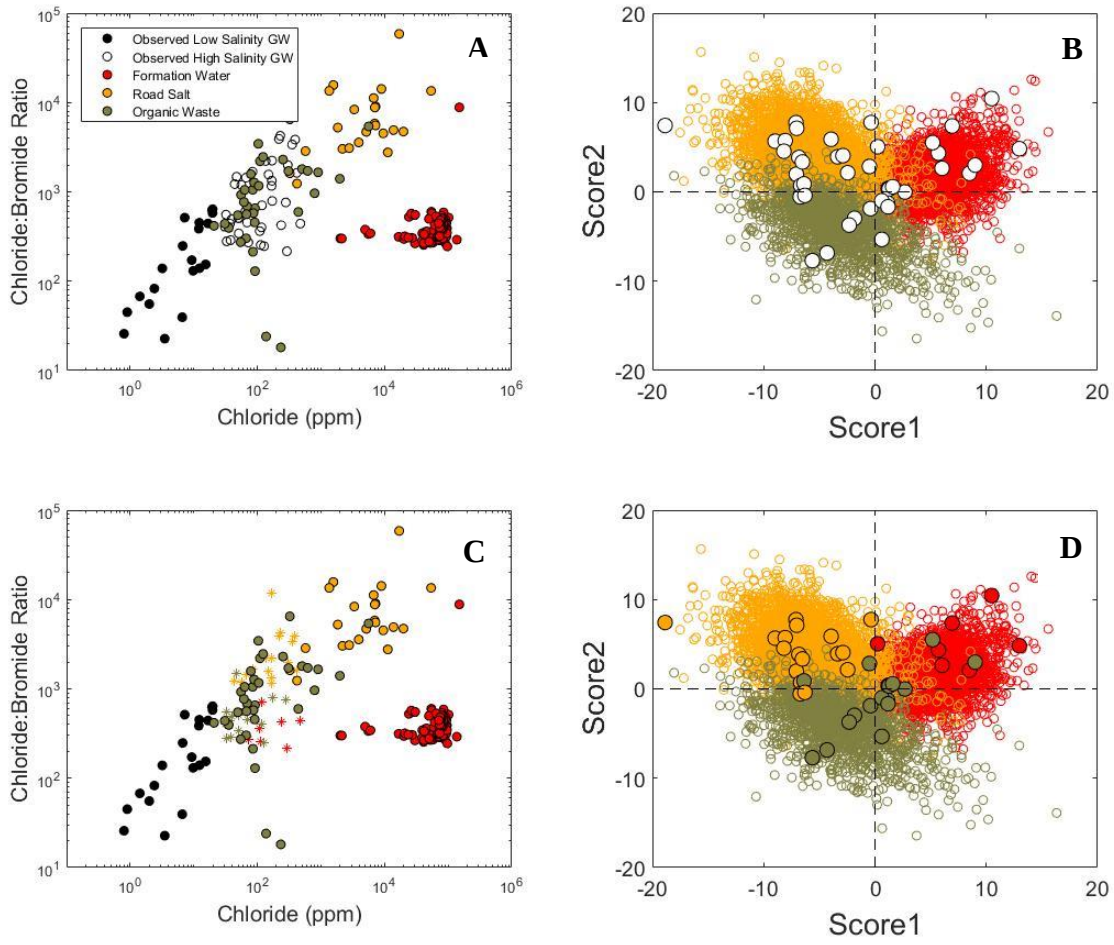


Figure 5. A) Chloride/bromide versus chloride plot for observed end-member samples used to create the synthetic dataset as well as groundwater samples (known sources of contamination are not demarcated). B) LDA scores for observed and synthetic groundwater samples, without the known sources of contamination. C) Chloride/bromide versus chloride for observed end-member samples, used to create the synthetic dataset along with groundwater samples (stars) colored with their known source of contamination. D) LDA scores for observed and synthetic groundwater samples, with colors representing known sources of contamination.

Given the clear importance of chloride and bromide in calculating the LDA scores, one question that might be raised is: how is LDA an improvement on previous chemical fingerprinting that only uses chloride and bromide concentrations? Figure 5 compares the two methods. Figures 5A and 5B both present the scenario typical to groundwater contamination

investigations. Groundwater samples have been collected and their chemistry has been determined but their source of contamination is unknown (open white circles). In order to utilize the Cl:Br ratios in Figure 5A, hypothetical mixing lines could be drawn connecting end-member populations to pristine groundwater (black dots). From there, samples that fall close to a certain mixing line will be delineated as contaminated by that end-member. Which end-member/mixing line a sample belongs to becomes more unclear the closer it plots to the pristine groundwater samples. In contrast, the groundwater samples as well as samples from prior studies (LDA is a machine learning algorithm and thus can continue to update itself as new data representing different populations is added) can be input into the LDA framework presented above. This is an automatic process that will return both scores for each sample and posterior probabilities for each contaminant. Classification confidence can be assessed in-bulk by removing posterior probabilities below a certain threshold or assessed for individual samples. A qualitative comparison of the two methods is seen in Figures 5C and 5D. There is no way to show quantitative results of this methodological comparison primarily because the former method, making use of chloride and bromide exclusively, does not provide unequivocal classification of unknowns, and therefore has the potential to yield multiple classifications for unknowns.

LDA is an improvement on graphical techniques for two reasons. For one, LDA provides a clear quantitative, and thus transferable, calculation of posterior probabilities to evaluate classification accuracy. This is in contrast to graphical techniques that are more subjective. The second improvement is that graphical analysis is limited to two dimensions whereas LDA can summarize the role of many constituents and therefore make use of many-dimensional solute datasets.

4. Conclusion

The model proposed above that makes use of LDA and natural variations in solute concentrations has clear implications for discriminating contaminants in groundwater. Between the work done in model validation (this paper) and our previous work (Lautz et al. 2014), methods have proven transferable across groundwater regions. The benefit of using our fingerprinting implementation is the clear and unambiguous use of chemical ratios (such as chloride:bromide). Whereas examining the range of chloride:bromide graphically can be equivocal for identifying the contaminant sources in groundwater, LDA returns calculated probabilities of class membership. Such probabilities tend to be conservative estimates of actual classification accuracy and give users a gauge on the general utility of any given classifications. The flexibility of the model framework, in terms of which solutes to input and which contaminants are of import, makes large-scale assessment of routine groundwater quality data highly feasible for a range of different contamination and data availability scenarios.

Going forward, proper use of this model is contingent on understanding the limitations of the approach. The current configuration of the model is transferable to other regions with few, if any, changes. However, the model does assume that no other contaminants are present besides those included, and assumes the observed data input to the model (e.g. low salinity groundwater samples, end-member samples) are representative of the regional low salinity groundwater chemistry and end-member chloride sources. If another potential source of chloride, such as seawater intrusion, is present in a study region, it must be included as a distinct end-member in the model and observations of that end-member must be included as input data. Accuracy of the model with inclusion of other contaminants, like landfill leachate, has not been validated and therefore analysis of model output needs to be critically evaluated under new conditions. The

current model configuration uses 9 solute concentrations in the LDA model. More or fewer solute concentrations could easily be considered in the model, depending on data availability and the effectiveness of various solutes for discriminating between end-members. While default model parameters, such as prior probabilities and which contaminant groups to include, are based on these authors' best-judgment, ultimately, the appropriate parameter values to use for a given study area are up to the discretion of users.

Interest in water quality monitoring has been increasing in the public and scientific consciousness. With that, more data at a wider variety of sites has been increasing. Our model, which utilizes a machine learning approach to understanding large data sets, has the ability to parse through and make use of those datasets. In fact, the ideal usage of this model would be to assess regional changes in LDA classifications through time as industry practices (e.g. farming, resource extraction, de-icing application, etc.) and land-use change. Our statistical approach to fingerprinting salinity sources has a unique advantage of utilizing commonly available solute concentrations; concentrations that have been measured historically and continue to be measured in baseline/background databases. While novel tracers are ultimately useful for point specific investigations, the ease and frugality of using this model make it ideal for assessing change over a wide area and through time. The combination of our approach with new tracers may provide a more comprehensive evaluation of salinity sources in a wider range of settings than was previously available.

Acknowledgements: This material is based upon work supported by the National Science Foundation under Grant Nos. EAR-1313522 and DGE-1449617, as well as the SyracuseCoE and the Syracuse University Office of the Vice Chancellor & Provost. We thank Greg Hoke and Zunli Lu for their initial work formulating the model and their feedback on early drafts of this manuscript. We also want to thank the anonymous reviewers that provided feedback on the writing.

References

Alawattegama, S.K., T. Kondratyuk, R. Krynock, M. Bricker, J.K. Rutter, D.J. Bain, and J.F. Stolz. 2015. Well water contamination in a rural community in southwestern Pennsylvania near unconventional shale gas extraction. *Journal of Environmental Science and Health* **50**: 516–528. doi: 10.1080/10934529.2015.992684.

Becker, M.F., D. K.D. Peter, and J. Masoner. 2001. Possible sources of nitrate in ground water at swine licensed-managed feeding operations in Oklahoma. *U.S. Geological Survey Water-Resources Investigations* Report 2002-4257.

Chapman, E.C., R.C. Capo, B.W. Stewart, C.S. Kirby, R.W. Hammack, K.T. Schroeder, and H.M. Edenborn. 2012. Geochemical and strontium isotope characterization of produced waters from marcellus shale natural gas extraction. *Environmental Science and Technology* **46**: 3545–3553. doi: 10.1021/es204005g.

Corsi, S.R., D.J. Graczyk, S.W. Geis, N.L. Booth, and K.D. Richards. 2010. A fresh look at road salt: Aquatic toxicity and water-quality impacts on local, regional, and national scales. *Environmental Science and Technology* **44**: 7376–7382. doi: 10.1021/es101333u.

Davis, S., D.O. Whittemore, and J. Fabryka-Martin. 1998. Uses of Chloride/Bromide ratios in studies of potable water. *Groundwater* **36**: 338–350. doi: 10.1111/j.1745-6584.1998.tb01099.x.

DeSimone, L.A., P.A. Hamilton, and R.J. Gilliom. 2009. Quality of water from domestic wells in principal aquifers of the United States, 1991–2004—Overview of Major findings. *U.S. Geological Survey Circular 1332*. doi: 9781411323506.

Gutchess, K., L. Jin, L.K. Lautz, S.B. Shaw, X. Zhou, and Z. Lu. 2016. Chloride sources in urban and rural headwater catchments, central New York. *Science of the Total Environment* **565**: 462-472. doi: 10.1016/j.scitotenv.2016.04.181.

Hildenbrand, Z.L., D.D. Carlton Jr., J.M. Meik, J.T. Taylor, B.E. Fontenot, J.L. Walton, D. Henderson, J.B. Thacker, S. Korlie, C.J. Whyte, P.F. Hudak and K.A. Schug. 2017. A reconnaissance analysis of groundwater quality in the Eagle Ford shale region reveals two distinct bromide/chloride populations. *Science of the Total Environment* **575**: 672-680. doi: 10.1016/j.scitotenv.2016.09.070.

Hogan, J.F., F.M. Phillips, S.K. Mills, J.M. Hendrickx, J. Ruiz, J.T. Chesley and Y. Asmerom. 2007. Geologic origins of salinization in a semi-arid river: The role of sedimentary basin brines. *Geology* **35.12**: 1063-1066.

Hwang, H.H., S.V., Panno, and K.C. Hackley. 2015. Sources and changes in groundwater quality with increasing urbanization, Northeastern Illinois. *Environmental and Engineering Geoscience* **21**: 75–90. doi: 10.2113/gseegeosci.21.2.75.

Johnson, J.D., J.R. Graney, R.C. Capo, and B.W. Stewart. 2015. Identification and quantification of regional brine and road salt sources in watersheds along the New York/Pennsylvania border, USA. *Applied Geochemistry* **60:** 37–50. doi: 10.1016/j.apgeochem.2014.08.002.

Katz, B.G., S.M. Eberts, and L.J. Kauffman. 2011. Using Cl/Br ratios and other indicators to assess potential impacts on groundwater quality from septic systems: A review and examples from principal aquifers in the United States. *Journal of Hydrology* **397**: 151–166. doi: 10.1016/j.jhydrol.2010.11.017.

Kelly, T.D., G.R. Matos, comps. 2014. Historical Statistics for mineral and material commodities in the United States (2016 version). U.S. Geological Survey Data Series 140, accessed [July 25, 2017], at https://minerals.usgs.gov/minerals/pubs/historical-statistics/.

Kelly, W.R., S.V. Panno, and K.C. Hackley. 2012. Impacts of road salt runoff on water quality of the Chicago, Illinois region. *Environmental and Engineering Geoscience* **18**: 65–81. doi: 10.2113/gseegeosci.18.1.65.

Kim, H., D. Kaown, B. Mayer, J. Lee, Y. Hyun, and K. Lee. 2015. Identifying the sources of nitrate contamination of groundwater in an agricultural area (Haean basin, Korea) using isotope and microbial community analyses. *Science of the Total Environment* **533**: 566-575. doi: 10.1016/j.scitotenv.2015.06.080.

Lauer, N.E., J.S. Harkness, and A. Vengosh. 2016. Brine Spills Associated with Unconventional Oil Development in North Dakota. *Environmental Science & Technology* **50**: 5389-5397. doi: 10.1021/acs.est.5b06349.

Lautz, L.K., G.D. Hoke, Z. Lu, D.I. Siegel, K. Christian, J.D. Kessler, and N.G. Teale. 2014. Using discriminant analysis to determine sources of salinity in shallow groundwater prior to hydraulic fracturing. *Environmental Science and Technology* **48**: 9061–9069. doi: 10.1021/es502244v.

Lu, Z., S.T. Hummel, L.K. Lautz, G.D. Hoke, X. Zhou, J. Leone, and D.I. Siegel. 2015. Iodine as a sensitive tracer for detecting influence of organic-rich shale in shallow groundwater. *Applied Geochemistry* **60**: 29–36. doi: 10.1016/j.apgeochem.2014.10.019.

Mullaney, J.R., D.L. Lorenz, and A.D. Arnstson. 2009. Chloride in groundwater and surface water in areas underlain by the glacial aquifer system, northern United States. *U.S. Geological Survey Scientific Investigations* Report 2009-05086.

Panno, S.V., K.C. Hackley, H.H. Hwang, S.E. Greenberg, I.G. Krapac, and S. Landsberger. 2005. Database for the characterization and identification of sources of sodium and chloride in natural waters of Illinois. Illinois State Geological Survey Open File Series 2005–1. Champaign, Illinois.

Panno, S.V., K.C. Hackley, H.H. Hwang, S.E. Greenberg, I.G. Krapac, S. Landsberger, and D.J. O'Kelly. 2006. Characterization and identification of Na-Cl sources in ground water. *Groundwater*. **44**: 176–187. doi: 10.1111/j.1745-6584.2005.00127.x.

Reilly, D., D. Singer, A. Jefferson, and Y. Eckstein. 2015. Identification of local groundwater pollution in northeastern Pennsylvania: Marcellus flowback or not? *Environmental Earth Sciences* **73**: 8097–8109. doi: 10.1007/s12665-014-3968-0.

Thunqvist E. 2004. Regional increase of mean chloride concentration in water due to the application of deicing salt. *Science of the Total Environment* **325**: 29-37. doi: 10.1016/j.scitotenv.2003.11.020.

Tran, N.H., K.Y. Gin, and H.H. Ngo. 2015. Fecal pollution source tracking toolbox for identification, evaluation and characterization of fecal contamination in receiving urban surface waters and groundwater. *Science of the Total Environment* **538**: 38-57. doi: 10.1016/j.scitotenv.2015.07.155.

Vidic, R.D. and S.L. Brantley. 2013. Impact of Shale Gas Development on Regional Water Quality. *Science* **340**. doi: 10.1126/science.1235009.

Warner, N.R., R.B. Jackson, T.H. Darrah, S.G. Osborn, A. Down, K. Zhao, A. White, and A. Vengosh. 2012. Geochemical evidence for possible natural migration of Marcellus Formation brine to shallow aquifers in Pennsylvania. *Proceedings of the National Academy of Sciences* **109**: 11961–11966. doi: 10.1073/pnas.1121181109.

Whittemore, D.O. 1995. Geochemical differentiation of oil and gas brine from other saltwater sources contaminating water resources: Case studies from Kansas and Oklahoma. *Environmental Geosciences* **2.1**: 15-31.

Wilcox, D.A. 1986. The effects of deicing salts on vegetation in Pinhook-Bog, Indiana. *Canadian Journal of Botany* **64**: 865–874. doi: 10.1139/b86-113.

World Health Organization. 2004. Guidelines for drinking water quality. World Health Organization vol. 1. doi: 10.1525/jer.2014.9.2.1.

Supporting Information

Chien & Lautz, 2017 – Science of the Total Environment

Section 1. Combining Septic Effluent and Organic Waste

Synthetic Training Data: Full Model

Known Group	Predicted Group			
	formation water	road salt	septic effluent	animal waste
formation water	89.3 (2635)	3.4 (99)	2.6 (78)	4.7 (139)
road salt	4.6 (132)	82.0 (2330)	12.1 (345)	1.2 (34)
septic effluent	-	2.9 (35)	89.3 (1089)	7.8 (95)
animal waste	1.5 (43)	2.5 (72)	29.5 (851)	66.5 (1921)

Table S1: Confusion matrix showing model classification accuracy for the training data for the Full Model. Numbers outside of parentheses are the percentage of samples classified correctly of the total samples contaminated by a single source (designated by the first column). Numbers within parentheses indicate total number of samples in each respective cell. Bold numbers along diagonals emphasize correct classifications. The synthetic-groundwater dataset size varies by contaminant due to the capped mixing percentage.

Validation Data: Full Model

Known Group	Predicted Group			
	formation water	road salt	septic effluent	animal waste
formation water	100 (7)	-	-	-
road salt	5.9 (1)	82.4 (14)	11.8 (2)	-
septic effluent	33.3 (1)	-	66.7 (2)	-
animal waste	28.6 (4)	7.1 (1)	42.9 (6)	21.4 (3)

Table S2: Confusion matrix showing model classification accuracy for the validation data for the Full Model. Numbers outside of parentheses are the percentage of samples classified correctly of the total samples contaminated by a single source (designated by the first column). Numbers within parentheses indicate total number of samples in each respective cell. Bold numbers along diagonals emphasize correct classifications.

Septic effluent and animal waste show relatively poor classification accuracy for the validation data as well as discrepancies in classification accuracy between the synthetic and

validation data. The poor classification accuracy for septic effluent validation data can in part be attributed to the low number of samples with known septic effluent contamination (only 3). Another possible explanation for the poor classification is that these two contaminants have indistinguishable chemical fingerprints for the solutes used in our analysis. This is corroborated by the fact that for both datasets, animal waste and septic effluent are likely to be misclassified for each other. 6 of the 14 total animal waste samples were misclassified as septic effluent for the validation data and 29.5% of animal waste samples for the synthetic data. It is possible that septic effluent and animal waste are better characterized as one distinct organic-waste end-member, at least in the context of the solutes chosen and the conservative 2-component mixing model. To test this hypothesis, septic effluent and animal waste samples (end-member and observed groundwater data) were grouped together and classified under the assumption that they represent one contaminant: organic waste. The revised version of the LDA model is referred to as the "organic waste model" and is used throughout the bulk of the main text. To develop this model, animal waste and septic effluent end-member data were pooled before being input to the model. The single organic waste pool was subsequently used to generate synthetic training data representing the chemistry of groundwater contaminated by a random source of organic waste.

Section 2. Data

Table S3: Groundwater chemistry data used for validation compiled from Panno et al. (2005) and Hwang et al. (2015). "PRISTINE" is low salinity groundwater used in the mixing model. "FW" indicates groundwater contaminated by formation water. "RS" indicates groundwater contaminated by road salt. "SEP" and "ANIM" indicate septic effluent contaminated groundwater and animal waste contaminated groundwater, respectively, and were combined to create the organic waste dataset.

Sample Name	Expected Type	Source	I ppm	Na ppm	K ppm	Mg ppm	Ca ppm	Cl ppm	Br ppm	Sr ppb	Ba ppb	SO4 ppm
Aq1	PRISTINE	Panno2005	3.9	29	2	33	85	0.9	0.02	390	80	43
Aq2	PRISTINE	Panno2005	1.3	14	1	34	77	7.2	0.014	450	60	25
Aq3	PRISTINE	Panno2005	9.1	45	1	35	85	6.7	0.027	660	100	51
Aq4	PRISTINE	Panno2005	13	49	2	28	73	0.8	0.031	690	120	8.7
Aq5	PRISTINE	Panno2005	26	70	1	61	163	9.8	0.075	1110	20	406
Aq6	PRISTINE	Panno2005	2.9	7	2	51	109	12	0.031	291	810	39
Aq7	PRISTINE	Panno2005	3.7	13	10	29	67	2.4	0.029	2500	1760	2.8
Aq8	PRISTINE	Panno2005	18.7	90	3	8.8	17	2	0.036	526	23	119
Aq9	PRISTINE	Panno2005	4.7	31	5	78	130	3.2	0.023	906	590	268
Aq10	PRISTINE	Panno2005	0.6	2.3	8	43	127	3.5	0.154	643	41	91
Aq11	PRISTINE	Panno2005	28.4	119	6	11	16	6.6	0.167	499	46	1
Aq12	PRISTINE	Panno2005	3.3	12	1	29	73	9.3	0.054	141	64	0.1
H_3	PRISTINE	Hwang2015	0.7	10.7	1	29	72.6	19.7	0.034	51	17	18.4
H_9	PRISTINE	Hwang2015	0.9	37.9	1	58.7	108	12.3	0.088	95	51	34.8
H_15	PRISTINE	Hwang2015	0.8	3.1	5	49.1	95.5	16.8	0.038	77	35	64.2
H_18	PRISTINE	Hwang2015	0.9	3.2	6	42.2	91.2	12.2	0.027	57	20	56.3
H_20	PRISTINE	Hwang2015	0.7	6.1	5	49.8	93.1	19.4	0.031	80	46	36.9
H_21	PRISTINE	Hwang2015	2.9	15.3	5	38.2	61.3	1.42	0.021	2100	122	1.3
H_29	PRISTINE	Hwang2015	1.7	12	5	39	91.4	19.8	0.031	80	27	15.7
H_31	PRISTINE	Hwang2015	NaN	5.2	4	42.9	93	15.4	0.1	101	24	15.2
B2	FW	Panno2005	86.6	308	1	38	96	474	1.075	660	490	0.1
B3	FW	Panno2005	112	220	2	25	59	237	0.56	637	275	0.01

B6	FW	Panno2005	NaN	167	5	24	135	292	1.35	1280	90	87
B8	FW	Panno2005	126	117	1	37	74	72	0.265	760	560	0.6
B10	FW	Panno2005	NaN	73	1	58	127	109	0.3	307	81	129
B11	FW	Panno2005	7.8	72	1	59	129	113	0.445	298	82	134
B13	FW	Panno2005	10.3	127	6.3	30	44	118	0.167	818	361	8.7
S1	RS	Panno2005	64.8	218	3	80	166	416	0.255	152	129	46
S2	RS	Panno2005	4.9	59	2	70	156	150	0.096	107	55	28
S3	RS	Panno2005	7	224	18	47	134	392	0.102	370	100	42
S4	RS	Panno2005	70	224	19	43	126	352	0.104	310	90	42
S5	RS	Panno2005	3.6	111	7	34	92	219	0.056	149	25	30
S6	RS	Panno2005	2.3	54	4	44	100	229	0.054	115	23	35
S7	RS	Panno2005	3.3	99	4	42	101	135	0.061	184	29	14
S8	RS	Panno2005	4.2	92	4	60	122	169	0.127	127	81	26
S9	RS	Panno2005	4	59	8	63	121	167	0.143	102	55	41
S10	RS	Panno2005	4	204	7	32	73	170	0.076	69	36	48
S11	RS	Panno2005	10.9	191	7	54	128	301	0.151	140	63	42
S12	RS	Panno2005	1.7	42	5	49	100	84	0.05	124	62	34
S13	RS	Panno2005	2.7	20	5	48	101	44	0.036	203	55	137
H_14	RS	Hwang2015	4	58.6	8	62.5	121	170	0.076	102	55	41.1
H_16	RS	Hwang2015	4	204	7	31.7	73.4	167	0.0143	69	36	48.1
H_17	RS	Hwang2015	2.8	23.8	6	52.4	111	66.2	0.046	131	73	76.4
H_23	RS	Hwang2015	5.1	34.5	5	36.6	90.5	56	0.047	151	24	28.6
Em1	SEP	Panno2005	9	21	3	45	105	50.6	0.1	120	30	54
Em2	SEP	Panno2005	7	19	4	43	101	39.7	0.08	100	30	56
Em3	SEP	Panno2005	36	75	13	47	111	116	0.29	240	80	77
H_5	ANIM	Hwang2015	26	23.3	119	38.5	81.3	30.9	0.076	115	40	13.7
H_11	ANIM	Hwang2015	1	14.3	1	53	106	47.2	0.032	83	37	46
H_25	ANIM	Hwang2015	0.7	3.8	5	49.9	105	32.2	0.058	71	65	62.4
Am1	ANIM	Panno2005	21	66	2	64	205	127	0.515	530	300	59
Am2	ANIM	Panno2005	38.8	45	2	32	127	33	0.118	200	90	15
Am3	ANIM	Panno2005	17.4	65	2	20	107	38	0.132	190	110	21
Am4	ANIM	Panno2005	16.8	102	2	44	113	51	0.149	260	100	36
Am5	ANIM	Panno2005	43.3	70	2	117	275	280	0.368	349	199	52
Am6	ANIM	Panno2005	77.8	113	23	84	116	171	0.216	471	417	22
Am7	ANIM	Panno2005	1.6	15	1	70	173	50	0.098	241	128	75
Am8	ANIM	Panno2005	1.7	13	3	68	156	65	0.107	177	50	151
Am9	ANIM	Panno2005	12.6	27	6	100	192	69	0.152	200	103	236
Am10	ANIM	Panno2005	35.1	20	67	71	147	56	0.201	142	161	68
Am11	ANIM	Panno2005	21.9	18	87	49	125	37	0.084	151	48	38

Table S4: Compiled chemistry data for Illinois formation water

Sample Name	Source	I ppm	Na ppm	K ppm	Mg ppm	Ca ppm	Cl ppm	Br ppm	Sr ppb	Ba ppb	SO4 ppm
B14	Panno2005	NaN	100000	63	73	750	150000	17	NaN	38800	1900
B15	Panno2005	NaN	34500	88	924	2540	101500	197	122000	320	NaN
B16	Panno2005	NaN	44400	86	1290	4700	89600	334	200000	750	NaN
B17	Panno2005	3900	48540	280	1720	4710	85000	180	199000	550	720
B18	Panno2005	8600	50280	282	1640	5650	89000	190	242000	3130	590
Stueb_29	Stueber1991	NaN	39000	246	2110	5510	77320	249	337	4930	54
Stueb_213	Stueber1991	NaN	31500	375	1570	4020	59660	212	225	NaN	120
Stueb_31	Stueber1991	NaN	36600	228	2270	5860	73460	235	393	29300	1
Stueb_153	Stueber1991	NaN	30900	412	1700	3390	58010	215	369	66500	10
Stueb_176	Stueber1991	NaN	52100	542	1510	6600	94010	297	260	670	1647
Stueb_177	Stueber1991	NaN	51200	1270	2430	9600	97220	397	243	NaN	1190
Stueb_187	Stueber1991	NaN	38600	509	2250	6310	76690	250	393	NaN	1
Stueb_187A	Stueber1991	NaN	36400	576	2110	6000	75960	264	379	NaN	70
Stueb_190	Stueber1991	NaN	22400	148	1180	2790	42610	151	291	NaN	127
Stueb_190A	Stueber1991	NaN	16300	108	792	1800	30160	114	186	NaN	42
Stueb_34	Stueber1991	NaN	21600	436	960	2400	41310	162	177	16100	25
Stueb_2	Stueber1991	NaN	68200	1870	2390	12150	136900	470	431	1110	727
Stueb_7	Stueber1991	NaN	26700	325	1190	2900	53380	185	212	7740	72
Stueb_10	Stueber1991	NaN	30900	289	2030	4630	65840	216	932	4220	162
Stueb_35	Stueber1991	NaN	14100	141	760	980	27220	89	103	6250	6
Stueb_44	Stueber1991	NaN	26600	173	1610	3470	51960	188	577	20100	8
Stueb_45	Stueber1991	NaN	26800	159	1610	3470	52230	184	504	19400	1
Stueb_99	Stueber1991	NaN	25800	260	1710	3810	53140	197	541	18900	1
Stueb_49	Stueber1991	NaN	35900	513	2010	6230	71370	232	160	390	1706
Stueb_51	Stueber1991	NaN	30500	369	1650	4530	59750	189	177	550	934
Stueb_54	Stueber1991	NaN	35100	496	2050	5830	70580	235	215	990	925
Stueb_58	Stueber1991	NaN	25300	188	1530	3290	48530	170	193	1150	129
Stueb_58A	Stueber1991	NaN	26700	260	1560	4040	50520	186	202	NaN	787
Stueb_66	Stueber1991	NaN	22000	193	1130	3210	42910	147	145	780	649
Stueb_68	Stueber1991	NaN	25800	230	1600	5380	53930	179	172	530	1197
Stueb_69	Stueber1991	NaN	31300	278	1980	5140	62780	210	284	720	678
Stueb_70	Stueber1991	NaN	33700	274	1970	5420	67350	219	287	910	603
Stueb_72	Stueber1991	NaN	37600	355	2190	6630	76750	267	321	930	445
Stueb_75	Stueber1991	NaN	31100	299	1620	4830	62890	219	317	1370	413

Stueb_256	Stueber1991	NaN	36030	621	2490	6970	75200	255	204	NaN	1367
Stueb_257	Stueber1991	NaN	36100	608	2420	6520	74400	259	276	NaN	687
Stueb_258	Stueber1991	NaN	36950	503	2440	6380	73900	258	328	NaN	693
Stueb_259	Stueber1991	NaN	37400	663	2440	6810	76400	262	228	NaN	1133
Stueb_260	Stueber1991	NaN	37150	632	2430	6390	75300	252	345	NaN	444
Stueb_262	Stueber1991	NaN	34250	629	2380	9090	75300	230	297	NaN	93
Stueb_21	Stueber1991	NaN	33100	379	1550	4670	66850	224	183	490	1139
Stueb_76	Stueber1991	NaN	27200	322	1310	3450	53470	198	213	12400	1
Stueb_78	Stueber1991	NaN	42900	440	2110	5910	82970	276	265	700	423
Stueb_79	Stueber1991	NaN	41800	419	2070	5910	83350	274	259	630	469
Stueb_80	Stueber1991	NaN	1550	23	22	38	2020	6.7	1.9	110	195
Stueb_85	Stueber1991	NaN	1610	24	24	50	2130	7.1	2.3	100	217
Stueb_86	Stueber1991	NaN	3840	26	49	55	5570	16.5	7	4020	8
Stueb_89	Stueber1991	NaN	3310	40	72	100	4950	13.1	6.5	30700	7
Stueb_97	Stueber1991	NaN	4130	32	54	50	6010	17.5	8.3	34600	1
Stueb_110	Stueber1991	NaN	20400	249	990	2170	39210	137	145	25200	1
Stueb_110A	Stueber1991	NaN	23000	296	1130	2350	41440	156	171	NaN	1
Stueb_114	Stueber1991	NaN	27700	309	1460	3550	53340	199	231	3180	7
Stueb_117	Stueber1991	NaN	26800	322	1360	2950	50450	182	164	5800	1
Stueb_120	Stueber1991	NaN	37400	424	1990	4510	71200	244	223	4830	1
Stueb_122	Stueber1991	NaN	43500	526	2010	6030	82340	273	258	780	173
Stueb_125	Stueber1991	NaN	41000	481	1810	5380	76580	233	241	980	342
Stueb_181	Stueber1993	NaN	30900	93	790	2070	54300	90.5	103000	34100	1
Stueb_189	Stueber1993	NaN	31300	116	927	1990	55560	96.6	107000	61300	1
Stueb_208	Stueber1993	NaN	9860	53	312	397	16630	53	77000	14600	1
Stueb_209	Stueber1993	NaN	11700	51	342	476	20110	67.8	74000	13600	1
Stueb_210	Stueber1993	NaN	12000	37	351	496	20900	65.7	47000	38500	1
Stueb_173	Stueber1993	NaN	14500	54	266	551	24170	46.1	28000	16000	1
Stueb_212	Stueber1993	NaN	17300	87	352	506	28730	52.3	35000	39000	1
Stueb_214	Stueber1993	NaN	15800	93	366	491	26350	46.1	38000	44300	1
Stueb_16	Stueber1993	NaN	44100	85	1890	5200	84750	144	453000	747000	1
Stueb_8	Stueber1993	NaN	21800	130	900	2780	42290	120	145000	4010	108
Stueb_12	Stueber1993	NaN	36500	172	1090	4910	64430	160	402000	22200	17
Stueb_139	Stueber1993	NaN	13600	47	585	6410	23670	77	79000	2580	1
Stueb_140	Stueber1993	NaN	34100	166	1650	3580	65370	136	273000	103000	1
Stueb_146	Stueber1993	NaN	33300	152	969	4810	63650	138	352000	14100	1
Stueb_185	Stueber1993	NaN	21200	80	852	1730	39300	114	135000	139000	1
Stueb_196	Stueber1993	NaN	39800	192	1410	4910	79310	147	236000	79800	9
Stueb_199	Stueber1993	NaN	45300	155	1330	4360	82800	164	191000	2130	529
Stueb_206	Stueber1993	NaN	18100	61	425	3570	30330	55	71000	24300	1
Stueb_4	Stueber1993	NaN	30200	106	1260	2300	56350	165	135000	2300	1

Stueb_6	Stueber1993	NaN	44500	211	1830	5370	85490	186	375000	7590	41
Stueb_141	Stueber1993	NaN	25700	116	1010	2030	46300	128	176000	3280	2
Stueb_148	Stueber1993	NaN	35700	260	1240	3400	64650	143	214000	50000	1
Stueb_149	Stueber1993	NaN	43100	146	1590	3930	78240	171	170000	1020	52
Stueb_150	Stueber1993	NaN	41900	206	1610	4150	76440	186	236000	58200	1
Stueb_163	Stueber1993	NaN	39500	180	1610	4810	74780	175	327000	125000	1
Stueb_164	Stueber1993	NaN	44900	196	1540	5210	83740	180	282000	770	457
Stueb_182	Stueber1993	NaN	31700	145	1310	2900	59510	160	197000	12000	1
Stueb_211	Stueber1993	NaN	34400	106	1250	2520	61580	145	142000	5430	12
Stueb_14	Stueber1993	NaN	49500	307	1870	6950	95670	244	265000	580	1200
Stueb_19	Stueber1993	NaN	48700	254	1580	6410	90610	187	189000	440	1680
Stueb_143	Stueber1993	NaN	43900	171	1810	6090	84550	197	401000	1790	339
Stueb_154	Stueber1993	NaN	44300	218	1680	4660	82170	193	236000	1920	612
Stueb_156	Stueber1993	NaN	39800	303	2080	5710	78750	218	381000	1100	601
Stueb_157	Stueber1993	NaN	34500	281	1670	5300	67180	215	217000	420	1110
Stueb_161	Stueber1993	NaN	38700	267	1580	5170	74680	206	269000	950	423
Stueb_167	Stueber1993	NaN	36500	253	1890	5390	71280	227	137000	210	1960
Stueb_168	Stueber1993	NaN	36800	193	1550	6270	73050	211	209000	400	1230
Stueb_170	Stueber1993	NaN	44500	234	1590	5590	84040	203	153000	380	1590
Stueb_178	Stueber1993	NaN	40200	237	1680	5070	78330	182	361000	38300	3
Stueb_198	Stueber1993	NaN	42700	252	1700	5210	80800	178	355000	9660	139
Stueb_205	Stueber1993	NaN	44900	232	1270	4250	80690	165	212000	1140	520
Stueb_245	Stueber1993	NaN	42500	205	1790	6290	81800	209	365000	680	702
Stueb_1	Stueber1993	NaN	49800	315	2840	4780	93780	247	650000	2440	254
Stueb_15	Stueber1993	NaN	40300	241	2360	6320	79220	237	135000	380	2040
Stueb_17	Stueber1993	NaN	50000	369	2590	5570	94250	211	343000	1160	680
Stueb_22	Stueber1993	NaN	52600	NaN	2550	4460	95730	278	580000	NaN	NaN
Stueb_126	Stueber1993	NaN	39600	184	1200	3000	69360	143	199000	740	521
Stueb_127	Stueber1993	NaN	45900	274	2260	5060	85580	208	531000	1710	379
Stueb_128	Stueber1993	NaN	41200	520	1630	4540	75530	195	150000	1440	2110
Stueb_129	Stueber1993	NaN	47900	349	2120	6020	89120	196	153000	410	1510
Stueb_130	Stueber1993	NaN	41600	300	2270	5550	81630	219	450000	1050	441
Stueb_133	Stueber1993	NaN	24100	119	1040	2390	41890	125	80000	510	425
Stueb_134	Stueber1993	NaN	46800	320	1700	5360	83090	170	121000	270	1880
Stueb_135	Stueber1993	NaN	30100	151	730	2730	49900	100	66000	180	3020
Stueb_136	Stueber1993	NaN	42500	248	1640	5390	78320	155	113000	430	1900
Stueb_137	Stueber1993	NaN	28400	170	1150	2260	49310	NaN	153000	510	2550
Stueb_138	Stueber1993	NaN	39000	192	1930	3140	72410	167	684000	2410	260
Stueb_151	Stueber1993	NaN	36100	188	1330	3820	64730	122	79000	690	2670
Stueb_160	Stueber1993	NaN	42900	413	2180	3950	77580	204	228000	620	1190
Stueb_165	Stueber1993	NaN	44300	289	2750	3970	81650	204	623000	1980	180

Stueb_169	Stueber1993	NaN	38400	251	2260	5920	73190	229	128000	330	1910
Stueb_171	Stueber1993	NaN	44400	304	2630	5080	83630	188	304000	880	1340
Stueb_179	Stueber1993	NaN	40200	283	2620	3890	77660	200	471000	570	470
Stueb_194	Stueber1993	NaN	43600	248	1300	4460	78470	167	202000	750	947
Stueb_195	Stueber1993	NaN	41900	259	1710	4670	76080	179	237000	1940	344
Stueb_197	Stueber1993	NaN	41500	232	1700	3670	74940	156	261000	550	982
Stueb_202	Stueber1993	NaN	40000	244	3080	3960	73660	168	279000	700	753
Stueb_203	Stueber1993	NaN	38700	240	1940	5490	73610	185	120000	350	1810
Stueb_204	Stueber1993	NaN	39500	196	1740	3080	71140	160	338000	780	529

Table S5: Compiled road salt end-member chemistry data from:

Sample Name	Source	I ppm	Na ppm	K ppm	Mg ppm	Ca ppm	Cl ppm	Br ppm	Sr ppb	Ba ppb	SO4 ppm
RS1	Panno2005	360	35496	NaN	NaN	NaN	54015	4.002	NaN	NaN	NaN
RS2	Panno2005	37.6	6270	26	29	241	8930	0.63	1270	192	248
MDBK2_18_19	Ledford2014	6	1317.8	3.31	33	242	2172.8	0.717	NaN	NaN	NaN
Granato_Table3	Granato1996	NaN	11000	5.1	2.2	100	17000	0.29	780	NaN	22
Howard_40	Howard1993	20	20250	13	5	62.5	30860	NaN	NaN	7400	222
Howard_41	Howard1993	20	21500	11.5	6	141	31640	NaN	NaN	16000	232
Kelly_Willow_Spring	Kelly2010	38	6270	NaN	NaN	NaN	8930	0.63	NaN	NaN	NaN
Kelly_Pekin	Kelly2010	NaN	279	NaN	NaN	NaN	1570	0.1	NaN	NaN	NaN
Foos_MS3	Foos2003	NaN	1049	10	27	148	1834	0.35	NaN	NaN	193
Foos_GT13	Foos2003	NaN	230	5	19	74	420	0.34	NaN	NaN	106
NYS_1	Lautz2014	2.1	9470	NaN	NaN	26	13830	2.82	NaN	NaN	NaN
NYS_2	Lautz2014	2.2	7750	NaN	NaN	19	11250	4.08	NaN	NaN	NaN
NYS_3	Lautz2014	3.0	13390	NaN	NaN	33	19820	4.19	NaN	NaN	NaN
NYS_4	Lautz2014	10.0	6630	NaN	NaN	74	9560	2.12	NaN	NaN	NaN
TW_2	Risch2000*	38	4190	14	48	270	7000	1.20	NaN	NaN	160
TW_94D	Risch2000*	20	2560	6.9	40	200	3930	1.10	NaN	NaN	150
TW_94E	Risch2000*	15	3110	14	64	270	5160	1.10	NaN	NaN	130
MW_3	Risch2000*	14	4420	18	34	230	7230	0.80	NaN	NaN	190
MW_4	Risch2000*	17	3980	15	55	260	7030	0.80	NaN	NaN	150
MW_5	Risch2000*	17	3690	19	77	470	6730	0.60	NaN	NaN	110
MW_7	Risch2000*	24	4400	14	24	210	7200	1.30	NaN	NaN	170
MW_8	Risch2000*	4	263	6.9	52	160	572	0.20	NaN	NaN	71
MW_9	Risch2000*	9	1320	12	110	390	2780	0.90	NaN	NaN	44
MW_13	Risch2000*	14	1440	13	96	360	3340	0.40	NaN	NaN	56
MW_14	Risch2000*	5	613	8.5	66	220	1350	0.10	NaN	NaN	36

MW_15D	Risch2000*	2	128	4	34	100	323	0.20	NaN	NaN	36

Risch2000* - Data from an State University of New York, College of Environmental Science and Forestry undergraduate thesis.

Table S6: Compiled chemistry data for septic effluent

Sample Name	Source	I ppm	Na ppm	K ppm	Mg ppm	Ca ppm	Cl ppm	Br ppm	Sr ppb	Ba ppb	SO4 ppm
E1	Panno2005	7	2740	26	300	495	5620	1.04	5650	380	19
E2	Panno2005	13	184	9	53	84	308	0.18	770	40	38
E3	Panno2005	38.6	416	11	19	84	504	0.28	220	100	52
E4	Panno2005	34.2	224	12	10	33	91	0.2	70	30	38
E5	Panno2005	7.3	262	22	23	64	253	0.11	140	30	73
E6	Panno2005	11	80	345	6.5	59	92	0.71	140	40	59
E7	Panno2005	NaN	46	17	41	60	35.5	0.09	660	130	22
E8	Panno2005	NaN	191	15	6	49	49.3	0.09	2209	40	51
E9	Panno2005	NaN	98	23	22	46	91.7	0.14	170	30	92
E10	Panno2005	NaN	55	6	39	70	20.8	0.05	530	150	24
E11	Panno2005	NaN	89	17	26	67	69.1	0.1	280	60	95
E12	Panno2005	NaN	94	14	23	61	55.5	0.06	210	20	94
E13	Panno2005	NaN	115	17	24	69	84.3	0.08	210	20	130
E14	Panno2005	NaN	89	12	24	71	86.7	0.15	260	60	112
E15	Panno2005	NaN	79	18	26	68	61.5	0.08	260	50	107
E16	Panno2005	NaN	70	16	26	68	63.2	0.06	250	50	77
E17	Panno2005	NaN	69	9	24	73	79	0.05	280	70	97
E18	Panno2005	NaN	77	28	24	118	147	NaN	280	50	84
E19	Panno2005	NaN	82	18	27	70	66.5	0.22	300	50	109
E20	Panno2005	NaN	82	16	27	120	118	NaN	350	30	97
E21	Panno2005	NaN	88	16	24	67	67.5	0.12	280	40	109
E22	Panno2005	NaN	90	25	26	89	110	0.05	300	400	97
E23	Panno2005	NaN	46	257	2.3	27	88.9	0.07	110	20	84
E24	Panno2005	NaN	255	33	24	66	312	NaN	290	100	92
E25	Panno2005	NaN	88	23	26	107	30.8	0.07	340	80	26
E26	Panno2005	NaN	399	12	25	112	618	0.36	240	60	12
E27	Panno2005	NaN	124	281	7	34	123	0.05	150	50	82
E28	Panno2005	NaN	70	8	21	142	105	0.09	260	100	8.4
E29	Panno2005	NaN	393	6	7.5	17	324	0.05	NaN	NaN	48

Table S7: Compiled chemistry data for animal waste

Sample Name	Source	I ppm	Na ppm	K ppm	Mg ppm	Ca ppm	Cl ppm	Br ppm	Sr ppb	Ba ppb	SO4 ppm
A1	Panno2005	187.7	493	2760	70	343	900	0.544	820	40	797
A2	Panno2005	328.4	348	2010	3.1	34	794	0.823	87	23	1.9
A3	Panno2005	263.1	1190	4950	0.9	26	1980	1.413	73	31	47
A4	Panno2005	182.4	110	1030	95	87	440	0.739	220	60	2.1
Whitcomb_ Jul27_S1	Whitcomb2007	NaN	47.3	146.4	55.15	112.9	103.5	0.03	NaN	NaN	195.2
Whitcomb_ Jul31_S1	Whitcomb2007	NaN	60	180.1	59.66	91.5	54.6	0.2	NaN	NaN	231.02
Whitcomb _Aug1_S1	Whitcomb2007	NaN	141.1	506.6	30.49	40.06	137	5.702	NaN	NaN	303.7
Whitcomb _Aug2_S1	Whitcomb2007	NaN	73.01	208.7	64.2	132.5	85.3	0.4	NaN	NaN	225
Whitcomb _Aug3_S1	Whitcomb2007	NaN	183.1	804.6	24.98	20.92	233.88	12.8781	NaN	NaN	395.5
L1	Becker2002	NaN	250.05	1383.36	9.75	41.75	391	NaN	220	20	28
L2	Becker2002	NaN	1473.46	2587.32	3.5	22.55	2460	NaN	133	20	0.5
L3	Becker2002	NaN	168.21	453.29	28.83	55.88	200	NaN	317	21	10
L4	Becker2002	NaN	165.63	777.3	12.47	49.67	251	NaN	359	22	0.4
L5	Becker2002	NaN	175.8	861.91	11.74	42.86	301	NaN	317	24	85
L6	Becker2002	NaN	154.51	711.56	16.28	49.16	270	NaN	357	24	9
L7	Becker2002	NaN	481.43	1606.89	34.29	48.6	990	NaN	143	20	103
L8	Becker2002	NaN	241.74	961.22	55.78	109.19	470	NaN	1567	33	22
L9	Becker2002	NaN	304.71	673.86	4.68	50.23	443	NaN	640	58	8
L10	Becker2002	NaN	481.16	1949.83	11.57	48.49	1020	NaN	288	20	16
USGS_11	Becker2002	NaN	128.45	0.82	64.4	198.26	380	NaN	1086	228	126
USGS_41	Becker2002	NaN	88.43	2.33	24.3	115	173	NaN	595	112	150
USGS_42	Becker2002	NaN	30.03	1.5	22.59	176.32	163	NaN	599	814	47.4
USGS_63	Becker2002	NaN	265.26	9.7	65.23	197.61	429	NaN	3396	196	37.6
USGS_65	Becker2002	NaN	55.19	4.78	35.7	41.72	109	NaN	2059	139	27.1
USGS_66	Becker2002	NaN	345.52	7.08	18.77	70.7	347	NaN	1032	88	67.7
USGS_67	Becker2002	NaN	275.48	5.08	51.99	100.92	219	NaN	931	18	411
USGS_69	Becker2002	NaN	132.83	2.31	31.66	72.17	105	NaN	753	49	178
USGS_70	Becker2002	NaN	240.17	2.87	9.74	18.13	175	NaN	313	38	66.4
USGS_73	Becker2002	NaN	152.63	2.53	31.64	72.4	112	NaN	810	54	210
USGS_74	Becker2002	NaN	313.59	3.08	33.51	48.96	268	NaN	568	149	141
USGS_75	Becker2002	NaN	73.58	1.66	40.55	82.18	103	NaN	1238	195	44.9
USGS_76	Becker2002	NaN	206.25	2.14	24.46	51.39	137	NaN	961	105	58.7

Table S8: Compiled chemistry data for organic waste

Sample Name	Source	I ppm	Na ppm	K ppm	Mg ppm	Ca ppm	Cl ppm	Br ppm	Sr ppb	Ba ppb	SO4 ppm
E1	Panno2005	7	2740	26	300	495	5620	1.04	5650	380	19
E2	Panno2005	13	184	9	53	84	308	0.18	770	40	38
E3	Panno2005	38.6	416	11	19	84	504	0.28	220	100	52
E4	Panno2005	34.2	224	12	10	33	91	0.2	70	30	38
E5	Panno2005	7.3	262	22	23	64	253	0.11	140	30	73
E6	Panno2005	11	80	345	6.5	59	92	0.71	140	40	59
E7	Panno2005	NaN	46	17	41	60	35.5	0.09	660	130	22
E8	Panno2005	NaN	191	15	6	49	49.3	0.09	2209	40	51
E9	Panno2005	NaN	98	23	22	46	91.7	0.14	170	30	92
E10	Panno2005	NaN	55	6	39	70	20.8	0.05	530	150	24
E11	Panno2005	NaN	89	17	26	67	69.1	0.1	280	60	95
E12	Panno2005	NaN	94	14	23	61	55.5	0.06	210	20	94
E13	Panno2005	NaN	115	17	24	69	84.3	0.08	210	20	130
E14	Panno2005	NaN	89	12	24	71	86.7	0.15	260	60	112
E15	Panno2005	NaN	79	18	26	68	61.5	0.08	260	50	107
E16	Panno2005	NaN	70	16	26	68	63.2	0.06	250	50	77
E17	Panno2005	NaN	69	9	24	73	79	0.05	280	70	97
E18	Panno2005	NaN	77	28	24	118	147	NaN	280	50	84
E19	Panno2005	NaN	82	18	27	70	66.5	0.22	300	50	109
E20	Panno2005	NaN	82	16	27	120	118	NaN	350	30	97
E21	Panno2005	NaN	88	16	24	67	67.5	0.12	280	40	109
E22	Panno2005	NaN	90	25	26	89	110	0.05	300	400	97
E23	Panno2005	NaN	46	257	2.3	27	88.9	0.07	110	20	84
E24	Panno2005	NaN	255	33	24	66	312	NaN	290	100	92
E25	Panno2005	NaN	88	23	26	107	30.8	0.07	340	80	26
E26	Panno2005	NaN	399	12	25	112	618	0.36	240	60	12
E27	Panno2005	NaN	124	281	7	34	123	0.05	150	50	82
E28	Panno2005	NaN	70	8	21	142	105	0.09	260	100	8.4
E29	Panno2005	NaN	393	6	7.5	17	324	0.05	NaN	NaN	48
A1	Panno2005	187.7	493	2760	70	343	900	0.544	820	40	797
A2	Panno2005	328.4	348	2010	3.1	34	794	0.823	87	23	1.9
A3	Panno2005	263.1	1190	4950	0.9	26	1980	1.413	73	31	47
A4	Panno2005	182.4	110	1030	95	87	440	0.739	220	60	2.1
Whitcomb_	Whitcomb2007	NaN	47.3	146.4	55.15	112.9	103.5	0.03	NaN	NaN	195.2

Jul27_S1											
Whitcomb_ Jul31_S1	Whitcomb2007	NaN	60	180.1	59.66	91.5	54.6	0.2	NaN	NaN	231.02
Whitcomb_ Aug1_S1	Whitcomb2007	NaN	141.1	506.6	30.49	40.06	137	5.702	NaN	NaN	303.7
Whitcomb_ Aug2_S1	Whitcomb2007	NaN	73.01	208.7	64.2	132.5	85.3	0.4	NaN	NaN	225
Whitcomb_ Aug3_S1	Whitcomb2007	NaN	183.1	804.6	24.98	20.92	233.88	12.8781	NaN	NaN	395.5
L1	Becker2002	NaN	250.05	1383.36	9.75	41.75	391	NaN	220	20	28
L2	Becker2002	NaN	1473.46	2587.32	3.5	22.55	2460	NaN	133	20	0.5
L3	Becker2002	NaN	168.21	453.29	28.83	55.88	200	NaN	317	21	10
L4	Becker2002	NaN	165.63	777.3	12.47	49.67	251	NaN	359	22	0.4
L5	Becker2002	NaN	175.8	861.91	11.74	42.86	301	NaN	317	24	85
L6	Becker2002	NaN	154.51	711.56	16.28	49.16	270	NaN	357	24	9
L7	Becker2002	NaN	481.43	1606.89	34.29	48.6	990	NaN	143	20	103
L8	Becker2002	NaN	241.74	961.22	55.78	109.19	470	NaN	1567	33	22
L9	Becker2002	NaN	304.71	673.86	4.68	50.23	443	NaN	640	58	8
L10	Becker2002	NaN	481.16	1949.83	11.57	48.49	1020	NaN	288	20	16
USGS_11	Becker2002	NaN	128.45	0.82	64.4	198.26	380	NaN	1086	228	126
USGS_41	Becker2002	NaN	88.43	2.33	24.3	115	173	NaN	595	112	150
USGS_42	Becker2002	NaN	30.03	1.5	22.59	176.32	163	NaN	599	814	47.4
USGS_63	Becker2002	NaN	265.26	9.7	65.23	197.61	429	NaN	3396	196	37.6
USGS_65	Becker2002	NaN	55.19	4.78	35.7	41.72	109	NaN	2059	139	27.1
USGS_66	Becker2002	NaN	345.52	7.08	18.77	70.7	347	NaN	1032	88	67.7
USGS_67	Becker2002	NaN	275.48	5.08	51.99	100.92	219	NaN	931	18	411
USGS_69	Becker2002	NaN	132.83	2.31	31.66	72.17	105	NaN	753	49	178
USGS_70	Becker2002	NaN	240.17	2.87	9.74	18.13	175	NaN	313	38	66.4
USGS_73	Becker2002	NaN	152.63	2.53	31.64	72.4	112	NaN	810	54	210
USGS_74	Becker2002	NaN	313.59	3.08	33.51	48.96	268	NaN	568	149	141
USGS_75	Becker2002	NaN	73.58	1.66	40.55	82.18	103	NaN	1238	195	44.9
USGS_76	Becker2002	NaN	206.25	2.14	24.46	51.39	137	NaN	961	105	58.7

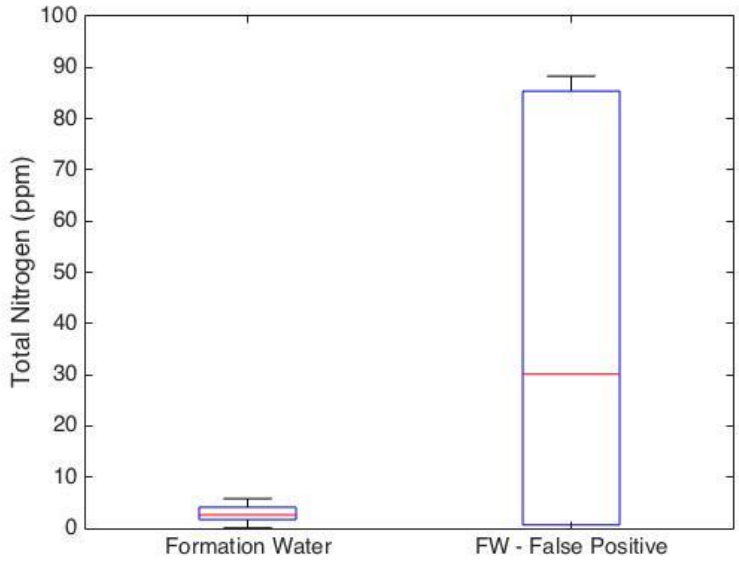
Table S9: Pristine and Contaminated Groundwater Statistics

		Mean	Median	95th Percentile	5th Percentile
Pristine Groundwater	I_ppm	6.5	2.9	26.2	0.7
	Na_ppm	28.7	13.5	91.5	3.1
	K_ppm	3.7	3.5	8.1	1.0
	Mg_ppm	39.5	38.6	61.9	10.9
	Ca_ppm	86.4	88.1	131.7	17.0
	Cl_ppm	9.1	8.3	19.7	0.9
	Br_ppm	0.1	0.0	0.2	0.0
	Sr_ppb	572.4	420.0	2120.0	56.7
	Ba_ppb	202.8	48.5	857.5	19.9
	SO4_ppm	64.9	35.9	274.9	1.0
		Mean	Median	95th Percentile	5th Percentile
Contaminated Groundwater	I_ppm	23.3	7.8	89.1	1.5
	Na_ppm	90.5	66.0	224.0	14.3
	K_ppm	11.7	5.0	67.0	1.0
	Mg_ppm	51.6	48.0	84.0	25.0
	Ca_ppm	120.1	113.0	192.0	73.0
	Cl_ppm	146.2	116.0	392.0	33.0
	Br_ppm	0.2	0.1	0.6	0.0
	Sr_ppb	268.4	177.0	760.0	71.0
	Ba_ppb	120.0	80.0	417.0	25.0
	SO4_ppm	53.7	42.0	137.0	0.6

Table S10: Contaminant End-Member Statistics

		Mean	Median	95th Percentile	5th Percentile
Basin Brine	I_ppm	6250	6250.0	8365.0	4135.0
	Na_ppm	34616.8	36500.0	49960.0	11760.0
	K_ppm	274.9274	245.0	603.2	47.6
	Mg_ppm	1550.216	1610.0	2538.0	275.2
	Ca_ppm	4245.2	4540.0	6624.0	479.0
	Cl_ppm	65605.68	73050.0	94202.0	20268.0
	Br_ppm	179.6694	186.0	273.9	46.1
	Sr_ppb	139549.1	91500.0	452550.0	145.0
	Ba_ppb	19258.82	1575.0	64160.0	324.5
	SO4_ppm	555.6803	340.5	1909.5	1.0
		Mean	Median	95th Percentile	5th Percentile
Road Salt	I_ppm	30.81326	14.5	38.0	2.1
	Na_ppm	6577.529	4085.0	21187.5	238.3
	K_ppm	11.53719	12.0	19.7	3.9
	Mg_ppm	43.22004	34.0	97.4	4.7
	Ca_ppm	186.9769	200.0	387.0	26.7
	Cl_ppm	10172.18	7015.0	31445.0	458.0
	Br_ppm	1.206744	0.8	4.1	0.1
	Sr_ppb	1025	1025.0	1245.5	804.5
	Ba_ppb	7864	7400.0	15140.0	912.8
	SO4_ppm	129.2222	140.0	234.4	33.9
		Mean	Median	95th Percentile	5th Percentile
Organic Waste	I_ppm	107.27	36.4	299.0	7.1
	Na_ppm	254.8298	141.1	493.0	47.3
	K_ppm	424.9315	18.0	2010.0	2.1
	Mg_ppm	32.46246	24.5	65.2	3.5
	Ca_ppm	84.29541	68.0	197.6	22.6
	Cl_ppm	392.1243	147.0	1020.0	49.3
	Br_ppm	0.779403	0.1	2.7	0.1
	Sr_ppb	638.0182	300.0	2104.0	103.1
	Ba_ppb	88.74545	50.0	273.6	20.0
	SO4_ppm	99.5823	73.0	303.7	2.1

Figure S1: Boxplot showing total nitrogen for the true basin brine-contaminated groundwater samples (n = 7) and the false positive formation water samples (n = 6).



References

Becker, M.F., D. K.D. Peter, and J. Masoner. 2001. Possible sources of nitrate in ground water at swine licensed-managed feeding operations in Oklahoma. *U.S. Geological Survey Water-Resources Investigations* Report 2002-4257.

Foos, A. 2003, Spatial distribution of road salt contamination of natural springs and seeps, Cuyahoga Falls, Ohio, USA. *Environmental Geolology* **44**, no. 1: 14-19.

Granato, G.E. 1996. Deicing chemicals as source of constituents of highway runoff. National Research Council: Washington D.C. Transportation Research Record 1533: 50-58.

Howard, K.W.F., and P.J. Beck. 1993. Hydrogeochemical Implications of Groundwater Contamination by Road Deicing Chemicals. *Journal Contaminant Hydrology* **12**, no. 3: 245-268.

Kelly, W.R. S.V. Panno, K.C. Hackley, H.H. Hwang, A.T. Martinsek, M. Markus. 2010. Using chloride and other ions to trace sewage and road salt in the Illinois Waterway. *Applied Geochemistry* **25**, no. 5: 661-673.

Lautz, L.K., G.D. Hoke, Z. Lu, D.I. Siegel, K. Christian, J.D. Kessler, and N.G. Teale. 2014. Using discriminant analysis to determine sources of salinity in shallow groundwater prior to hydraulic fracturing. *Environmental Science and Technology* **48**: 9061–9069. doi: 10.1021/es502244v.

Ledford, S.H., and L.K. Lautz. 2014. Floodplain connection buffers seasonal changes in urban stream water quality. *Hydrological Processes* **29**: 1002-1016. doi: 10.1002/hyp.10210.

Panno, S.V., K.C. Hackley, H.H. Hwang, S.E. Greenberg, I.G. Krapac, and S. Landsberger. 2005. Database for the characterization and identification of sources of sodium and chloride in natural waters of Illinois. Illinois State Geological Survey Open File Series 2005–1. Champaign, Illinois.

Stueber A.M. and L.M. Walter. 1991. Origin and chemical evolution of formation waters from Silurian-Devonian strata in the Illinois basin, USA. Geochemica et Cosmochimica. Acta **55**, 309-325.

Stueber, A.M., L.M. Walter, T.J. Huston, and P. Pushkar. 1993. Formation waters from Mississippian-Pennsylvanian reservoirs, Illinois basin, U.S.A.: Chemical and isotopic constraints on evolution and migration. *Geochemica et Cosmochimica Acta* **57**, no. 4: 763–784.

Whitcomb, A. 2007. Effectiveness of a vegetation buffer strip and wetland in dairy wastewater filtering. Undergraduate Thesis, State University of New York, College of Environmental Science & Forestry: Syracuse, NY.

Technical Supplement. Stream Restoration with Beaver Dam Analogues: Methods, Progress, and Future Work

1. Introduction

In the western United States, snowpack acts as a large reservoir of water that controls the annual hydrograph when snow melts during the springtime. Projected rises in global mean surface temperatures affect the timing of runoff in montane watersheds, increasing runoff during the cool season and decreasing runoff in the warm season, with peak flows from snowmelt shifting to earlier in the water year (IPCC, 2014; Barnett et al., 2005; Gleick, 1987; Adam et al., 2009). With many arid regions relying on snowmelt runoff for their water-use, populations are growing increasingly concerned about sustaining predictable, year-round water supplies (Barnett et al., 2005). With these ongoing and future changes to the hydrology of montane watersheds, creative and sustainable land management practices that can maintain annual baseflow are needed now more than ever (Lawler 2009, Rosemond and Anderson 2003).

Historically, beaver dams have played a significant ecosystem role as landscape engineers (Naiman et al. 1988). Beaver dams were once found within almost all temperate, northern-latitude, low-elevation streams from Northern Mexico to the edge of the Canadian arctic (Pollock et al., 2003; Pollock et al., 2017). Dam's control stream-systems by creating diverse hydraulic structures that lower stream power and promote wetlands with dense emergent vegetation (Gurnell, 1998; Burchsted and Daniels, 2014; Collen and Gibson, 2001). Evidence for the impact of beavers on the landscape can be found in the substantial changes to geomorphology and riparian vegetation after their removal from the vast majority of North American streams in the 19th century (Rea, 1983; Naiman et al., 1988). Currently beaver populations are rebounding after an aggressive re-introduction campaign in the 20th century (Naiman et al., 1988).

While beaver dams enact clear surficial changes to the land and water surface, they also have distinct effects on riparian hydrology. Foremost among these is the increased water retention afforded by the impoundment of water behind dams. These impoundments control the residence time of water in the watershed, attenuating the hydrograph and in many cases increasing annual baseflow (Nyssen et al., 2011; Puttock et al., 2017). Beaver dams have turned previously ephemeral streams into perennial streams and losing streams into gaining streams (Pollock et al., 2003; Majerova et al., 2015). Such changes result from increases in water surface elevation within the stream that yield increases in the near-stream water table. Further, the extra stores of water held behind dams have proven effective buffers against depletion of stored water from glaciers and snowpack, a concern for management of streams fed primarily by snowmelt (Beechie et al., 2013). Beaver dams also tend to increase the complexity of stream morphology, creating pool-riffle sequences that can reduce peak flow significantly; examples of this include a 5% reduction in peak flow from a single full beaver pond (Pollock et al., 2003; Beedle, 1991). Furthermore, these pool-riffle sequences create a diverse slow water / moving water habitat that can not only support diverse species but also increase the resilience of the system to disturbances like flooding (Collen and Gibson, 2001; Naiman et al., 1988).

More specifically to the groundwater system, beaver dams tend to increase groundwater recharge. Overbank flooding, a mechanism for replenishing groundwater in riparian areas, is promoted from water impounded behind dams (Westbrook et al., 2006). Additionally, hydraulic head created by beaver dams has been shown to increase hyporheic exchange around beaver dam complexes; by elevating near-stream head gradients, dams can increase the flux of stream water into the subsurface (Lautz and Siegel, 2006; Janzen and Westbrook, 2011). The hydrologic effects listed above combined with various water quality improvements (e.g. sediment retention,

expansion of habitat area, nutrient cycling, temperature moderation) have made beaver re-introduction a heavily endorsed stream management strategy (Lawler, 2009; Rosemond and Anderson, 2003; Pollock et al., 2017).

While beaver dams have been shown to achieve many stream restoration goals, beaver activities can conflict with human interests (Bhat et al., 1993; Jensen et al., 2001). In cases like these, restoration projects have turned toward beaver dam analogues (BDAs) to achieve restoration goals without the beavers themselves (Pollock et al., 2014; Pollock et al., 2017, Pilliod et al., 2017). BDAs mimic the natural dam quite closely. They are semi-porous allowing water, sediment, aquatic organisms, and other debris through. They are temporary structures that biodegrade when not maintained. But their primary benefit over natural dams is that they can be managed and controlled by land managers to optimize the achievement of restoration goals.

A typical BDA is made by pounding wood posts into a stream bed across a transect of the stream width (see Fig. 28 in Pollock et al., 2017). A thick assemblage of willow branches is weaved between and around the posts with soil, small rocks, and organic material packed into the sides. The structures typically remain somewhat porous to prevent the complete obstruction of water flow. BDAs have been used in a variety of research and restoration settings; however, there are few studies of their ability to replicate the hydrologic benefits listed above with most prior studies focusing on habitat restoration (Pilliod et al., 2017; Pollock et al., 2017). A potentially fruitful area of research lies in examining how BDAs replicate their natural counterparts and in quantification of the effects of small controlled dam installation on local landscape resilience to extreme climate events.

A methodology for testing the effectiveness of BDAs has yet to be proposed. Preliminary efforts rely heavily on long-term observational studies (Pollock et al., 2017; Pilliod et al., 2017).

An alternative approach is to model a previously well-studied site to forecast the future hydrologic effects of BDA installation. In general, hydrologic modeling is used to quantify flow within and out of a watershed based on a data-driven, conceptual model of the study domain. A modeling approach is suited for this project because of the ability to quantify the changes in hydrologic flow pathways caused by BDA installation, relative to baseline conditions.

While a modeling approach might be the most useful way to characterize watershed-scale hydrologic change to BDA installation, other approaches are needed to derive appropriate data and measure how BDAs cause reach-scale changes to water surface profiles, sediment aggradation, and vegetation density. Recent technological advances in small Unmanned Aircraft Systems (sUAS) have opened the door for high resolution measurement of topography and vegetation indices (Pai et al., 2017). These data products are a potential boon for those interested in understanding small scale effects of BDAs on a managed stream system.

This report explores the possibility of using both an aerial imagery and hydrologic modeling to characterize reach-scale hydrogeomorphic changes caused by BDAs. At the beginning of 2017, The Nature Conservancy reached out to Dr. Laura Lautz to collaborate on a stream restoration study. Their managed watershed in Red Canyon (Fig. 1), had experienced intense fluvial incision along many of its stream reaches. While recent drought from 2012 to 2015 followed by intense flooding in subsequent years had been posited as one cause of the incision, another potential cause was transient beaver activity. Some stream reaches in the watershed had robust beaver populations, while others had no recent beaver activity. Beaver have long been known to control near-stream morphology and hydrology (Naiman et al. 1988; Pollock et al., 2017), so their recent absence from the stream reaches was raised as a potential cause of recent fluvial incision. With this hypothesis and the growing use of BDAs in the stream

restoration community (Pilliod et al., 2017), The Nature Conservancy decided to install BDAs at Red Canyon and enlist hydrologists to study their effects.

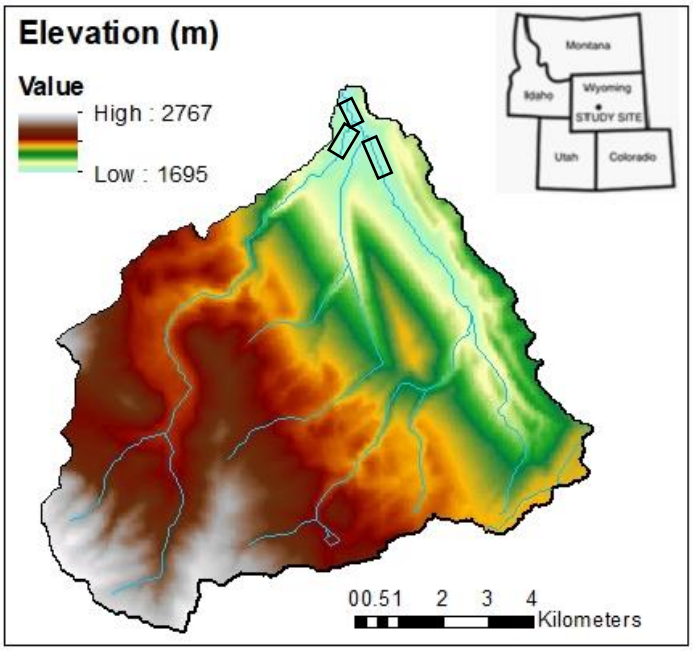


Figure 1. Topography of the Red Canyon Creek watershed. Inset map shows location relative to nearby states. Black boxes delineate the study reaches, see Fig. 2.

Initial funding for a summer intern was procured with the task of beginning data collection at Red Canyon Creek (RCC) prior to BDA installation. I traveled to RCC in August of 2017 and began that data collection, which continued into the research contained within this report. To date, pre-installation data has been collected and different modeling approaches have been explored. The report provides detailed background, methods, and future work that can be completed with either aerial imagery and/or a hydrologic modeling.

2. Aerial Imagery

A series of drone flights on August 14th and 15th were used to collect aerial imagery of three separate reaches in Red Canyon (Fig. 2). These reaches were chosen for their different histories

of recent damming; upstream and downstream RCC had very little beaver activity, while Cherry Creek had the most active beaver population. The flight details can be found in Table 1. SonyA5100 camera photos were used to calculate Digital Elevation Models (DEMs) of the study reaches. The MicaSense Rededge was used to derive Normalized Difference Vegetation Index (NDVI) values. This is one of many indices for estimating vegetation characteristics like leaf chlorophyll content, plant vigor, and depth to water (Nichols, 2000; Carroll et al., 2015). The following sections will detail the derivation of these data products and their various uses in earth systems modeling.

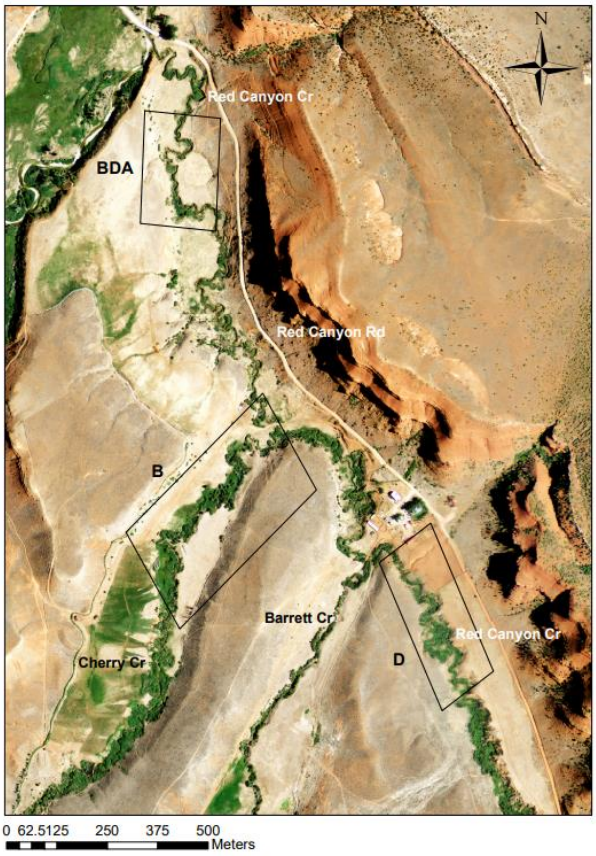


Figure 2. Approximate outline of the study reaches for the August 14-15th drone surveys. BDA is the reach where BDAs are planned to be installed; it also contains a well field. B is the downstream segment of Cherry Creek with the upstream portion containing active beaver populations. D is a degraded reach where heavy stream incision has occurred with no beaver activity present.

2.1 Digital Elevation Models using Photogrammetry

Photogrammetry is the process of making accurate spatial measurements with photography. From an Earth systems perspective, photogrammetry using aerial photography can yield DEMs and bathymetric models (Jensen, 2007). The traditional photogrammetry approach is to reconstruct a surface from overlapping images taken from multiple vantage points. In order to do this, a network of targets with known 3-dimensional positions must be defined beforehand. In contrast, Structure-from-Motion (SfM) photogrammetry calculates the geometry of the surface using a moving sensor that automatically determines camera position and orientation with respect to the target (Westoby et al., 2012; Micheletti et al., 2015). For this project, an SfM approach was taken to calculate DEMs for the three study reaches in RCC.

Imagery used to execute SfM photogrammetry came from the August drone flights. AirCTEMPS, a research group out of Oregon State University and The University of Nevada, coordinated the flight details along with Dr. Christa Kelleher (Table 1). Approximately 15 Ground Control Points (GCPs) were established at each site and surveyed relative to a local benchmark using a total station. These GCPs aid in georeferencing the SfM imagery. We used 12-in diameter bucket lids marked with black electrical tape to make an easily identifiable cross-hair at the lids center. Future work should use larger diameter lids or a pattern with greater contrast (e.g. orange lid with black cross-hair); the small, mostly white lids were easily confused with rocks during SfM georeferencing. It is also advisable to carefully map out the location of the GCPs on a to-scale map of each reach, such that they are more easily found in the visual images during data processing. Lastly, additional GCPs may aid in DEM accuracy (Woodget et al., 2015).

Camera	**Sony A5100**			**MicaSense Rededge**		
Reach	DRCC	CC	URCC	DRCC	CC	URCC
Altitude (m)	100			90		
Speed (m/s)	8			7		
Photo Overlap (%)	80			75		
Photos	145	350	175	180	450	230
Flight Time (min)	9	22	11	8	19	11
Ground Control Points	13	12	12	-		

Table 1. Flight information and camera details for the different study reaches. DRCC is the Downstream Red Canyon Creek site, CC is the Cherry Creek site and URCC is the Upstream Red Canyon Creek site.

A variety of SfM algorithms and workflows exist for deriving DEMs from aerial photographs, typically based on specific software packages. Our implementation relied on Agisoft Photoscan and a workflow developed by the AirCTEMPs team; a brief overview of that workflow will be outlined here but for more in-depth discussion see the Agisoft Photoscan Documentation and accessory AirCTEMPs documentation (Agisoft, 2018). Two major steps need to be completed before generation of a DEM: camera alignment and generation of a dense point cloud. Initial camera alignment matches key points on different photographs and uses them first to refine camera calibration parameters and then to define camera position for each photograph. It is almost entirely automatic, requiring little user input.

Between camera alignment and dense point cloud generation are a series of intermediary steps that require user intervention. The most important is georeferencing images to GCPs. These GCPs are used for setting up a coordinate system, optimizing photo alignment, and measuring distances within the photo-collage. As such, the locations of these GCPs cannot act as an independent check on the accuracy of the DEM. After this step, dense point cloud generation can begin. This initial data extraction step requires a few input parameters such as choosing Dense Cloud 'Quality'. For this work, the recommendations for appropriate parameters values from the AirCTEMPs documentation were followed; however, it is possible to create higher quality

products using other parameter sets (with increased computation time). Depending on the parameters chosen, this step takes anywhere from less than a day to a few weeks. Afterwards, generation of the DEM is a straightforward process that can be completed in Agisoft itself or in a GIS.

While the overall process of DEM extraction is almost entirely automated using Agisoft Photoscan, three additional notes of advice for future use will be included here. First, GCP identification in Agisoft can be quite difficult. Without prior aerial imagery of GCP locations, the best way to find GCPs is quickly scrolling through all the flight paths in a separate photo software to acquaint oneself with the flight path taken by the drone and the location of the GCPs during various flight lines. Second, incorrect identification of a GCP can degrade the quality of the final DEM and even prevent dense point cloud generation. To prevent this, users should keep track of which photos have georeferenced images and verify that they are correctly georeferenced multiple times. Lastly, processing runs especially for dense point cloud generation take a long time. Plan accordingly by double-checking user input to avoid errors and associated reruns and by understanding what each step should generate (i.e. verifying that part of the workflow is necessary for the project aims). With these in mind, it should be noted that Agisoft Photoscan is generally very user-friendly.

Currently, there is no standard method to independently validate the accuracy of the SfM-produced DEMs, and this should be a goal of future work. One method used by Woodget et al. (2015) was to independently survey over 500 validation points within their study reach and compare the surveyed values to the DEM produced values. Their mean error in areas without vegetation or other obstructions ranged from 0.004 to 0.111 meters. Their work simultaneously shows the potential for sub-meter scale DEM accuracy and persistent SfM-related systematic

error. One example of SfM system error is surface doming, where image accuracy deteriorates away from the DEM center (James and Robson, 2014; Pai, 2017; Woodget et al., 2015). Ongoing research is being conducted into minimizing this systematic error and should be an area of interest for future users of SfM-derived DEMs.

While the SfM technique is not free of all systematic error, there are still important applications of high-resolution DEMs that deserve attention. Geomorphologists interested in characterizing specific landform shapes likely comprise the most active users (Javernick et al., 2014; Westoby et al., 2012; Woodget et al., 2015). For instance, Westoby et al. (2012) in an attempt to demonstrate the utility of the SfM approach for small-scale glacial landform reconstruction, used the approach to develop DEMs of a moraine complex in Nepal and a glacially sculpted bedrock ridge in Wales. More relevant to hydrologists, SfM-derived DEMs have been used to reconstruct reach-scale bathymetry with questionable success (Woodget et al., 2015). Pai et al. (2017), in collaboration with the AirCTEMPS team used similar methods to our own to estimate water surface elevations. Their work paired both aerial and land-based data collection methods to estimate groundwater shortcutting near meander bends.

For this work, DEMs have been created for the three study reaches (Fig. 3). These DEMs include Digital Terrain Models (DTMs) and Digital Surface Models (DSM). In theory, the former defines the bare land surface excluding surface objects like vegetation, while the latter includes objects like vegetation. Many of the research articles referenced previously do not clearly delineate between which model they are using. While Agisoft Photoscan allows one to export both surface and terrain models, the process the software uses for determining the difference is not clearly outlined in the manual. The process roughly described in the manual has the user define a cell size and the point within that cell with the lowest estimated elevations

becomes the estimate of that cell's height, potentially a biased average. Default parameter values were chosen for this workflow such that cell size was defined with respect to the largest area within the model domain that does not contain any ground points. After performing this process to create a DSM and DTM for the upper RCC reach, a difference raster (DTM elevation subtracted from DSM elevation) was created to visualize the differences (Fig. 4). Qualitatively the process seems to work quite well; the difference raster in Fig. 4c shows a clear outline of the buildings and an outline of the dense near-stream vegetation. However, concerns still exist about the accuracy of these DEMs which will be discussed further.

To analyze the accuracy of the DEMs, elevations for the downstream RCC DEMs were compared with elevations of the GCPs. Since these GCPs were used to construct the DEM, this is not an independent check on model accuracy. Theoretically these values should be the same; however, they were not (Table 2). The mean absolute error of the DSM was 0.21-m and 0.51-m for the DTM. The DSM performed poorly when GCPs were located near dense vegetation (TP3 and P4) or on hillslopes (CP3 and CP4). The DTM performed poorly near dense vegetation (TP4, P15, P4, and TP3); however, in these areas, residuals were negative. One explanation for deteriorating DEM quality near dense vegetation is that cameras taken at different angles will estimate different elevation values depending on whether vegetation is within the camera viewpoint. Contradictory estimates will lead to deteriorating measurement quality. In general, the poor DTM quality relative to the DSM is likely due to point selection. Less points are being used to create the DTM and when multiple points are available, the DTM automatically picks the lowest point for a given area. This will favor outliers and potential measurement errors over a more weighted average approach. More work is needed to understand the optimal locations for GCPs.

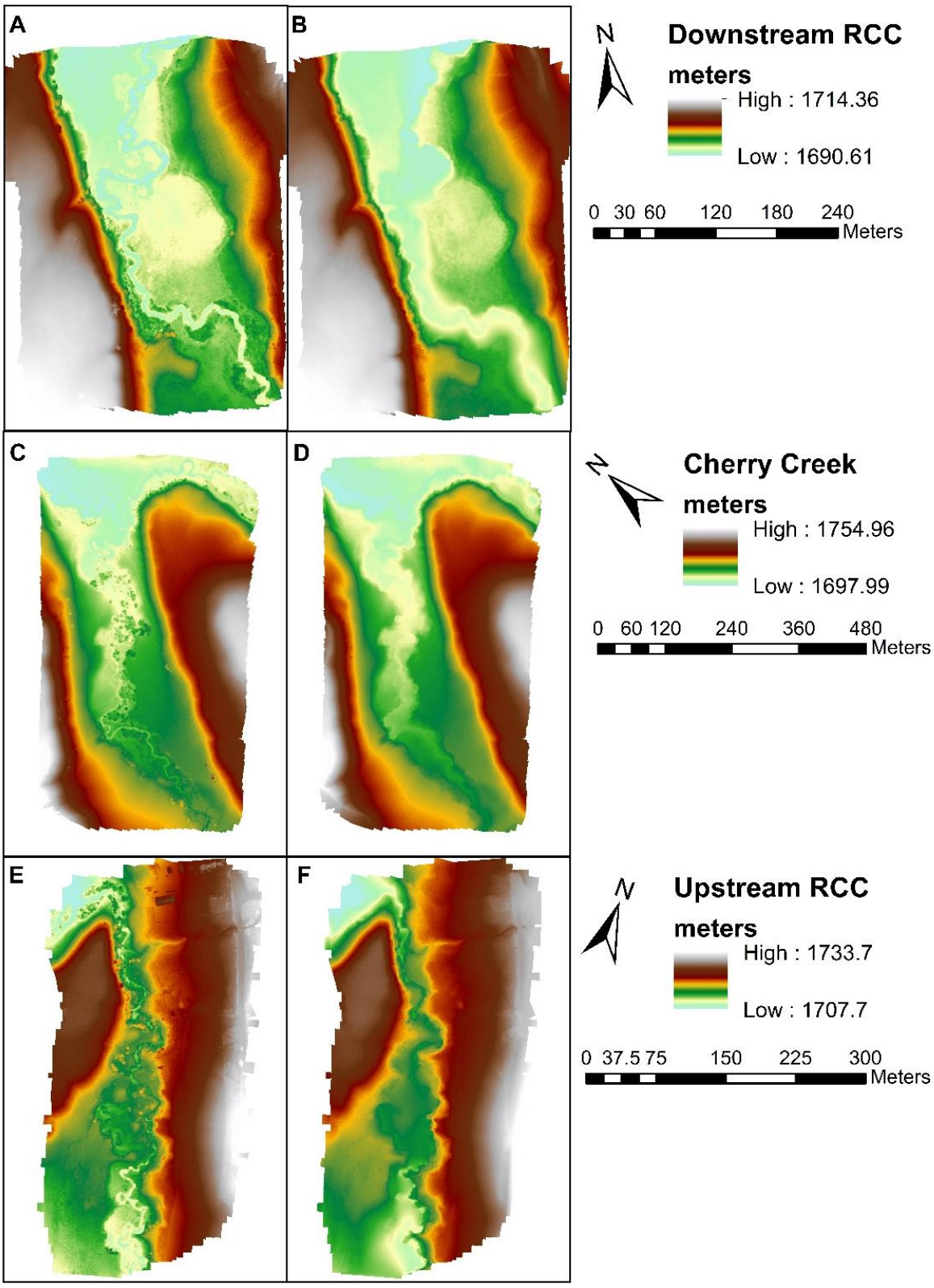


Figure 3. DEMs for the three study reaches, derived from SfM image processing. (A) Downstream RCC DSM, (B) downstream RCC DTM, (C) Cherry Creek DSM, (D) Cherry creek DTM, (E) Upstream RCC DSM, and (F) upstream RCC DTM.

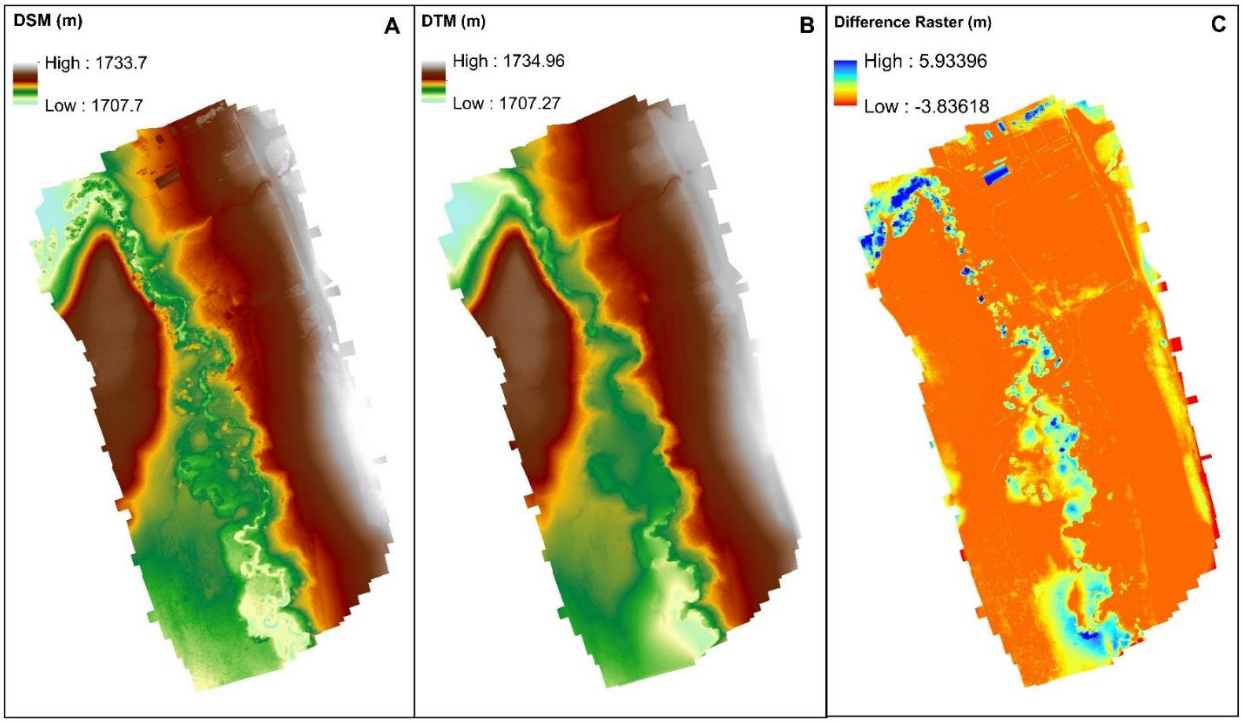


Figure 4. Upper Red Canyon Creek DEMs. a) Digital Surface Model (DSM) depicting land surface with buildings and vegetation. b) Digital Terrain Model (DTM) depicting bare land surface. c) Difference raster of DSM elevation minus DTM elevation. Negative values are related to DEMs not overlapping near raster edges.

CP_ID	Survey_elev(m)	DSM_elev(m)	DTM_elev(m)	DSM_error(m)	DTM_error(m)	DSM_erSQR	DTM_erSQR	DSM_Abs(m)	DTM_Abs(m)
TP4	1694.10	1694.11	1692.47	0.01	-1.63	0.00	2.66	0.01	1.63
P16	1694.71	1694.73	1694.61	0.02	-0.10	0.00	0.01	0.02	0.10
CP1	1698.44	1698.35	1698.34	-0.09	-0.10	0.01	0.01	0.09	0.10
P15	1694.53	1694.73	1692.76	0.20	-1.77	0.04	3.12	0.20	1.77
CP3	1700.93	1701.28	1701.29	0.35	0.36	0.12	0.13	0.35	0.36
P14	1695.04	1695.03	1695.02	-0.01	-0.02	0.00	0.00	0.01	0.02
P13	1694.87	1694.99	1694.98	0.12	0.11	0.01	0.01	0.12	0.11
CP4	1698.35	1698.69	1698.68	0.34	0.33	0.11	0.11	0.34	0.33
P5	1694.96	1695.07	1695.07	0.11	0.11	0.01	0.01	0.11	0.11
P4	1695.09	1695.64	1694.57	0.55	-0.52	0.31	0.27	0.55	0.52
CP2	1698.41	1698.53	1698.52	0.12	0.11	0.01	0.01	0.12	0.11
TP1	1695.56	1695.67	1695.67	0.11	0.11	0.01	0.01	0.11	0.11
TP3	1695.59	1696.25	1694.16	0.66	-1.43	0.43	2.05	0.66	1.43
				AVG		**RMSE**		**Mean Abs Error**	
				0.19	-0.34	0.29	0.80	0.21	0.51

Table 2. Residuals between ground control points and Structure from Motion-derived DEMs.

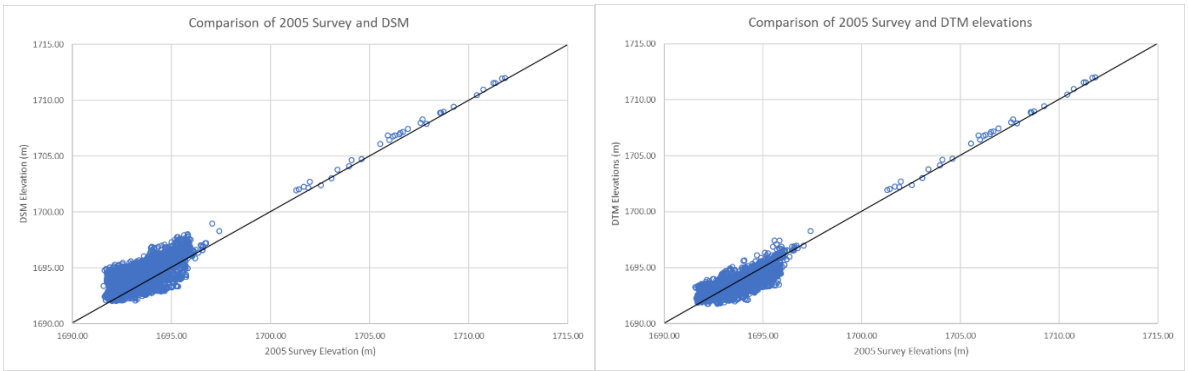


Figure 5. Comparison of 2005 survey elevations with the DSM (left) and DTM (right).

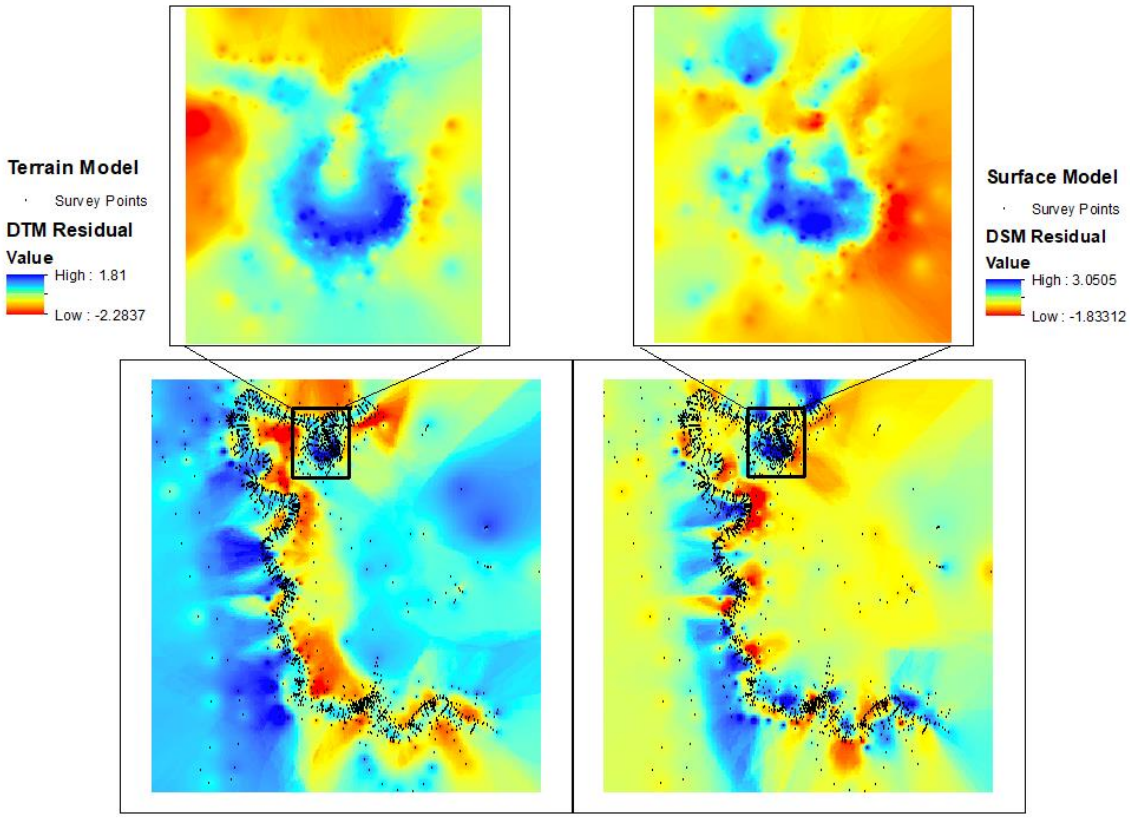


Figure 6. Residuals between the 2005 survey and the SfM DEMs interpolated using inverse-distance-weighting method. Left panels show the terrain model and right panels show the surface model. Inset maps show the oxbow without survey points.

To further interrogate DEM accuracy, residuals were calculated between the downstream RCC DEMs and a survey done in 2005 (Fig. 5). These survey points were taken mostly along the stream bed and near the stream. Since this survey occurred over a decade prior to the drone

flight, there is expected to be significant error related to erosion and deposition; however, this analysis is still valuable in comparing differences between the two DEMs. The DTM and DSM tended to align better at points of higher elevation, likely due to the paucity of vegetation in those areas. There was slightly more variation in DSM residuals than in DTM residuals (standard deviation of 0.857-m and 0.664-m, respectively). To visualize the spatial distribution of these residuals, the points were interpolated, through an inverse distance weighting method, over the model domain (Fig. 6). While there is bias in the interpolation, since survey points cluster in the stream, the highlighted oxbow area has a relatively even distribution of points. For both models, residuals tended to be higher in areas of dense vegetation. Both models fluctuate in accuracy along the stream reach (which will be discussed later). These residual maps in combination with the errors relative to the GCPs, highlight the imperfect nature of interpreting DTM as bare land surface and DSM as vegetation height.

In addition to the DEMs, water surface elevation was derived for the downstream RCC reach. These profiles are valuable for beaver dam related studies because, if SfM imagery is collected with high enough frequency through time, they provide a way of directly relating how the areal extent of surface water changes with the construction and destruction of beaver dams. Hafen (2017) attempted to estimate these changes with 10-m resolution DEMs. The alternative approach of measuring changes directly with high-resolution SfM-derived DEMs may improve estimates of ponded water volume.

Two methods were explored to determine best practices for extracting water surfaces with SfM-derived DEMs. One method followed the methodology of Pai et al. (2017); i.e. the thalweg was manually delineated in a GIS and the DEM was sampled along that transect at 0.5-m intervals (Fig. 7). Results show significant amounts of noise, likely related to the difficulty of

consistently determining whether a given surface is the channel bed or water surface. The close coupling of DTM and DSM estimates, except at seemingly random locations, underscore this point. An alternative method was undertaken to filter out some of the noise from these DEMs (Fig. 8). These water surface elevation values were calculated by sampling a water surface-only version of the DEM (again, manually delineated) at 4-meter intervals with a circular 2-meter buffer at each point, some points with significant deviation in mean-values from adjacent points were removed. Noise still exists but the buffering method gives a more gradual and realistic decline in stream elevation.

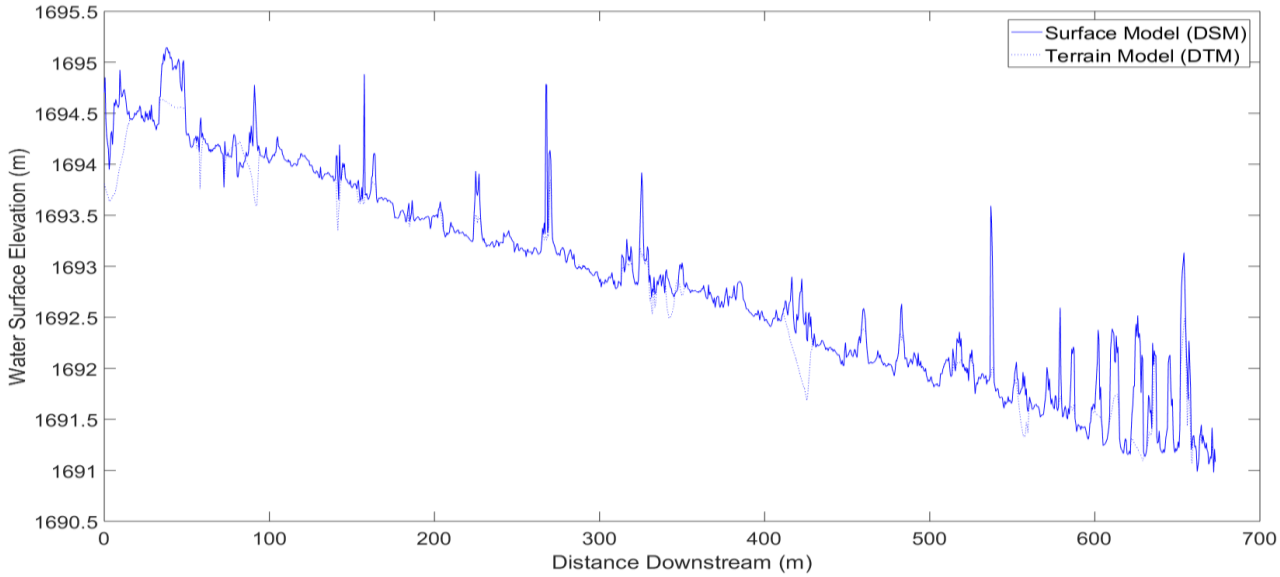


Figure 7. Water surface elevation profiles for the DEM and DTM of the downstream Red Canyon Creek stream reach for the thalweg of the creek (no data excluded).

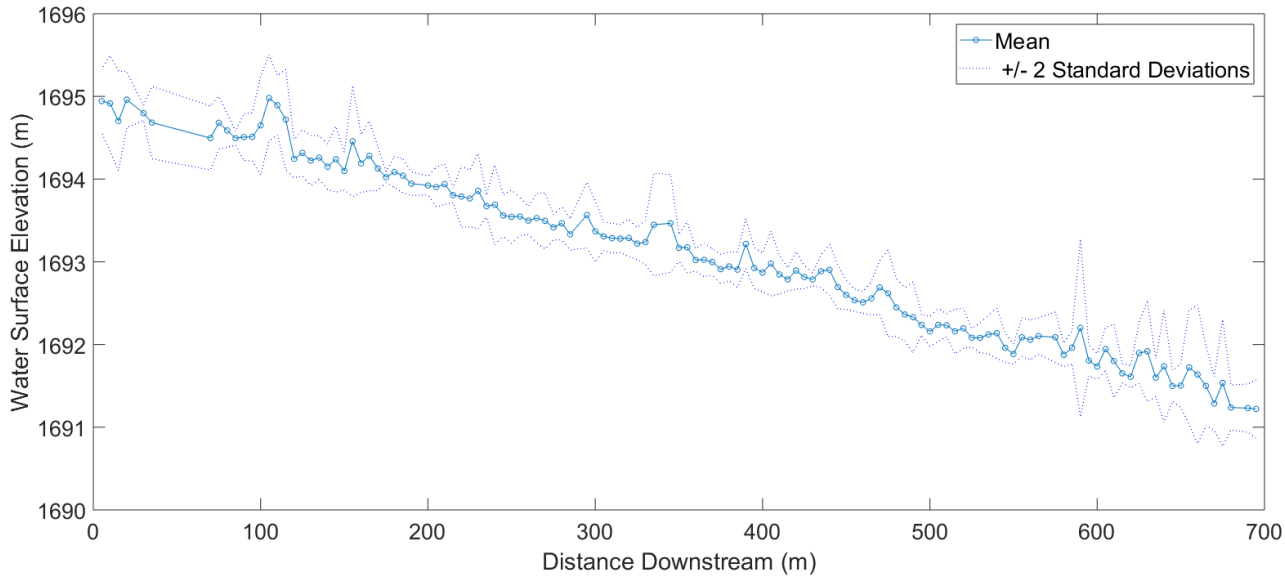


Figure 8. Water surface elevation profile of the downstream Red Canyon Creek stream reach using the buffer method discussed in the text.

It is unclear if SfM can differentiate between subsurface bedforms and water surface elevation, especially with visible water surface glare and/or vegetation casting shadows on to the water surface (Woodget et al., 2015; Pai et al., 2017). Woodget et al. (2015) used SfM to derive bedforms underneath the water surface; they found decreasing accuracy of elevation measurements with increasing water depth. In a similar vein, Pai et al. (2017) used SfM to derive water surface elevations but found significant noise along the reach profile. It is not clear whether these researchers used different methods to measure these different surfaces; however, it is clear that both their measurements had noise related to the feature they were not measuring. Those interested in sub 0.1-m-scale resolution water surface profiles should be cautious when interpreting SfM derived DEMs, especially at water depths above 0.5-m.

Ultimately, DEMs produced from SfM can be used to track landscape changes through time. Validation points like well casings and other immovable landscape features could provide an independent check on DEM accuracy. Beyond that, work done so far on water surface profiles

and streambed bathymetry is preliminary and should be interpreted as such in this author's opinion. At the very least, these DEMs provide a model of what the near-stream land surface looks like.

2.2 Normalized Difference Vegetation Index

Multispectral aerial imagery can be used to calculate vegetation indices. These indices are useful in comparing how vegetation varies spatially or temporally. The most commonly used index is NDVI, which can be calculated from multispectral camera photos, like those obtained from a MicaSense Rededge camera. NDVI is a measure of the reflectance of red and near-infrared wavelengths of the electromagnetic spectrum and can be used to compare the relative density and vigor of vegetation in an image (Tucker, 1979). Values close to 1.0 indicate high plant vigor and lower values close to or lower than 0.0 indicate the opposite. For the three study reaches, NDVI values for the August drone flight can be seen in Fig. 9. Other indices like NDRE exist but provide similar assessments as NDVI and so will not be discussed in this section.

NDVI has been used in hydrologic modeling studies typically to get at evapotranspiration rates and also to more generally track vegetation changes (Nichols, 2000; Carroll et al., 2015; Devitt et al., 2011; Huntington et al., 2017; Carroll et al., 2017). The index cannot provide a direct estimate of ET, instead it provides one of many inputs into an ET calculation. For instance, Nichols (2000) used the Modified Soil-Adjusted Vegetation Index (an index that includes NDVI as a term) to estimate plant cover. From there, an ET rate was calculated based on percent plant cover. Other uses include estimates of groundwater dependent ecosystems (Huntington et al., 2016) and as an independent evaluation of depth to water estimates in groundwater flow models (Carroll et al., 2015; Carroll et al., 2016). The widespread availability of NDVI from Landsat

imagery make it useful for assessing regional vegetation differences and the ease of calculating it from sUAS-derived multispectral imagery make it useful for reach-scale assessment.

NDVI and similar vegetation indices are of particular use for BDA studies because of the importance land managers place in restoring near-stream vegetation. Theoretically, any small dam that lifts the near-stream water surface will help promote denser vegetation growth by increasing the availability of subsurface water. Additionally, more vegetation could help reintroduce beavers so that they either build more dams or manage existing BDAs. The land manager at Red Canyon planted willow trees, a source of food and dam building material for beaver, for those purposes.

One downside in using NDVI is that it is a relative measure of plant vigor (Jensen, 2000). It is sensitive to factors such as cloud cover, plant shading, and image resolution (Jensen, 2000; Pai et al., 2017). Linking NDVI estimates to a more objective measure, like canopy height, could help in overcoming this downside. For instance, linking a difference raster (that is effectively estimating canopy height) with an overlapping map of NDVI (Fig. 10). Both maps highlight the dense near-stream vegetation (reeds and willows). However, where NDVI shows higher values in the floodplain due to grasses, the difference raster does not record the presence of the sparse grasses. Future work should move towards finding empirical relationships between these variables.

High-resolution multispectral imagery has the potential to analyze the effects that changes in small surface water impoundments have on vegetation. The methods outlined above that combine vegetation indices derived from multispectral imagery with DEMs is one way to make use of both data products. There do seem to be nearly endless ways for estimating

important hydrologic variables like evapotranspiration with these indices, so careful validation of any empirical relationships is an important future step.

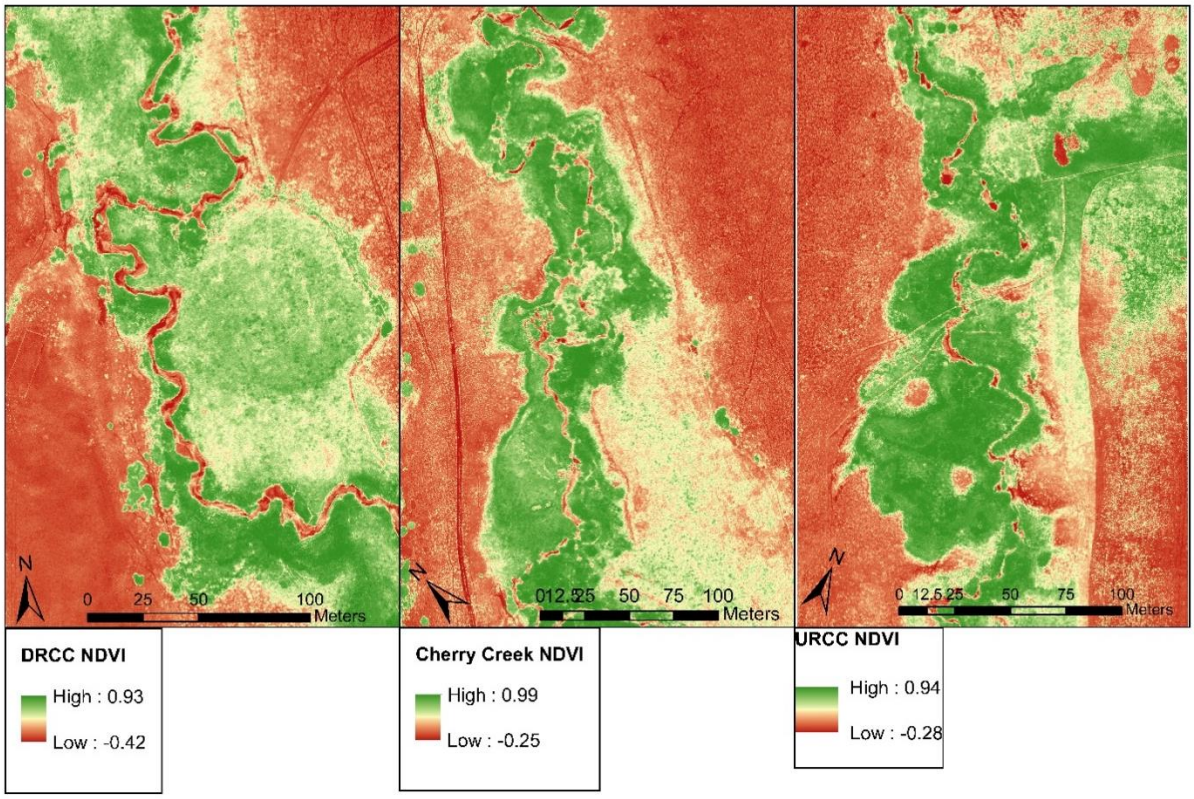


Figure 9. Normalized Difference Vegetation Index (NDVI) maps for the three stream reaches.

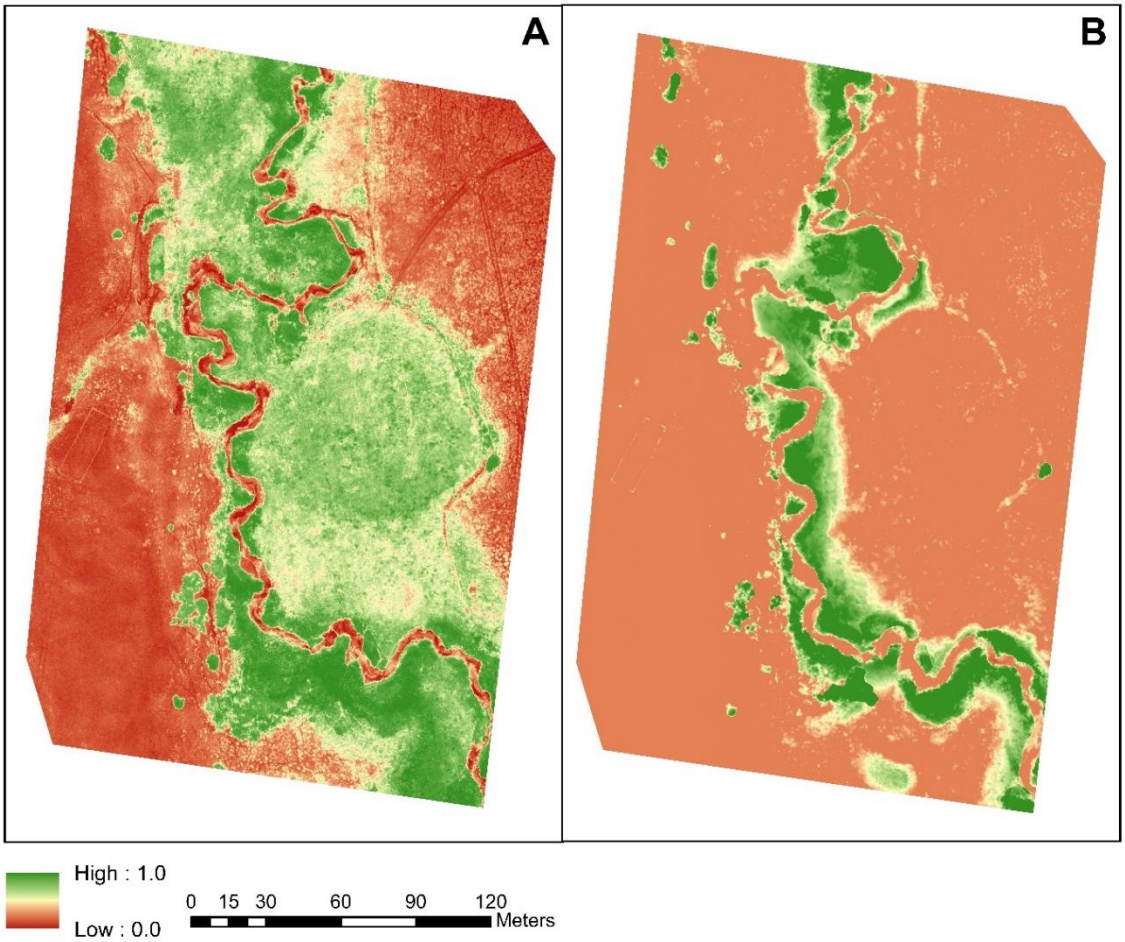


Figure 10. Downstream RCC (A) NDVI and (B) a normalized difference raster. The difference between DSM and DTM models was normalized by the maximum elevation. Both maps have the same scale.

3. Hydrologic Modeling

3.1 Introduction

How do beaver dam analogues regulate hydrologic processes along a stream reach? On the surface, this question is easily answered by looking at previous work studying hydrologic changes caused by natural beaver dams (e.g. Nyssen et al., 2011; Puttock et al., 2017; Majerova et al., 2015; Westbrook et al., 2006). However, while these studies provide a suite of hypotheses for how hydrologic processes will change with increased small damming, they do not get at the

management side of BDA stream restoration. That is, land managers can choose where to install BDAs and thus which areas get flooded. The more relevant question is then, does this added control actually matter for those concerned about volumetric water losses out of a managed watershed? Hydrologic research into BDAs can help answer that more pertinent question.

After initial data collection and processing, a hydrologic modeling approach was chosen to determine how specific BDA management strategies alter hydrologic processes in Red Canyon. Construction of a simplified conceptual hydrologic model was thought to be an initial step that could inform future data collection. Multiple different modeling approaches were explored. These included a groundwater flow-only approach with MODFLOW, a coupled surface water-groundwater approach with GSFLOW, and a separate coupled approach using HEC-RAS to calculate floodplain inundation. The next few sections will review what was learned from exploring these different approaches and provide guidance for future work.

3.2 Reach Scale Modeling

One of the main research questions related to BDA hydrologic changes is whether these structures promote groundwater storage during spring snowmelt and increase stream discharge during low flow periods, similar to natural beaver dams (Puttock et al., 2017; Majerova et al., 2015). This concept can be framed as a groundwater question: how do seasonally induced surface water head changes alter annual groundwater storage? By using a groundwater flow model like the USGS's Modular Groundwater Flow Model (MODFLOW) and altering boundary conditions, one could analyze the groundwater system response to surface water head changes. Initial exploration of this approach led to several conclusions regarding the applicability of a

groundwater flow modeling approach to simulating system response to BDA installation and subsequent stream changes.

Preliminary groundwater flow modeling was done for two steady state cases, representing the BDA and non-BDA cases (Fig. 11). A 2x2 meter grid with two model layers delineated from soil borings was chosen for cell discretization. The east/west bounding hillslopes were no-flow boundary conditions with specified flux in and out of the domain controlled by the hydraulic gradient across the central floodplain, similar to Lautz et al. (2006). For the stream boundary, constant head cell values were assigned to cells that intersected the location of the river. For the non-BDA case, constant head cell values were taken from a water surface profile derived from the SfM DEM. For the BDA case, the same downstream head value was chosen with 0.35-meter elevation jumps at each proposed BDA location.

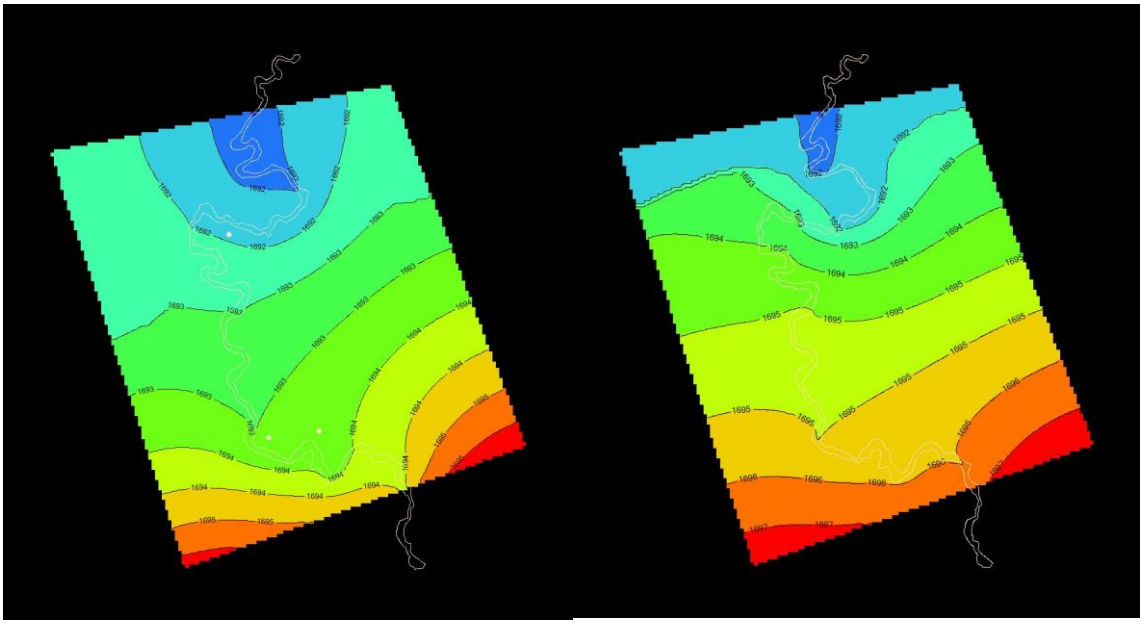


Figure 11. Preliminary groundwater flow domain at the RCC reach where BDAs installation is planned. Results show the water table without dam analogues (left) and with dam analogues (right).

The above conceptualization was a highly simplified representation of the system, built to do exploratory modeling with MODFLOW, and determine model limitations for determining the effects of BDAs on local hydrology. Results from this model unsurprisingly showed increased water storage in the elevated water surface (BDA) domain (Fig. 11). The main assumption that this model run made was in designation of the stream channel boundary condition. Interactions between landscape morphology and dam height were largely being ignored; i.e., some aspects of the ponded water behind dams, a potential source of groundwater recharge (Westbrook et al., 2006), were being ignored. The effect of that ignored water volume would be largely dependent on the areal extent of the surface water and the intersection of the new water surface elevation with the adjacent land surface; such that, the water table near an area like the ox-bow (see inset in Fig. 6) would be dramatically altered by the presence of a dam. This assumption was partly pragmatic, higher-resolution cell discretization (<1 m) was causing the MODFLOW Graphical User Interface (GUI) software to crash. But it was also due to potential software-incompatibility to the problem at hand. There is no straightforward method to integrate high resolution surface topography with the MODFLOW models using the MODFLOW GUI.

A likely reason for this software deficiency is the unlikelihood of having high-resolution (sub tens-of-cm scale) subsurface information to pair with surface information. The subsurface is largely unconstrained, except at discrete boreholes; there is an asymmetry between what we know at the surface and in the subsurface. However, it still should be possible to incorporate the best available surface water knowledge to aid in making subsurface predictions. A few related MODFLOW packages do this by simulating more complex surface water groundwater interactions. The most basic of these is the Streamflow Routing Package (Prudic et al., 2004). For this method, every cell that is intersected by the stream is assigned a reach ID (see Fig. 3 in

Prudic et al., 2004). The average depth of the reach is calculated at the reach midpoint using the streambed geometry and one of five separate depth equations. This depth is used along with the Darcy equation to calculate water loss between the stream and the aquifer. Of all the built-in head-dependent flux boundary packages, the streamflow routing package may be the best MODFLOW package for simulating complex streamflow and groundwater interactions like those caused by BDAs.

The main limitation in using the streamflow routing package is that it still does not take into account complex stream geometries. For instance, it is not meant to be used for wide channel streams where grid cell resolution is smaller than stream width (Prudic et al., 2004; Ou et al., 2013). Again, the complex geometry, derived from SfM and needed to understand BDA ponding, would be ignored using this built-in MODFLOW boundary condition. An approach that considers both changing streamflow conditions and reach morphology is needed.

To do this, multiple alternative modeling platforms were investigated. MODFLOW based platforms like DAFLOW and MODBRANCH (Jobson and Harbaugh, 1999; Swain and Wexler, 1996), while more useful for complex groundwater-surface water studies than the Streamflow Routing Package (SFR), ultimately run into similar problems with representing high-resolution stream geometry. The most complete MODFLOW-based package found was one developed by Ou et al. (2013) that improves on the streamflow routing package. The package was developed to account for streambed heterogeneity and channels wider than grid cell-size, current limitations of the SFR package. Going forward this looks like a well-suited package for effectively and efficiently coupling MODFLOW groundwater simulations with a surface water body. Although this package has strong potential for this application, it is not publicly available and those interested in using it need to reach out to the developer listed in the paper.

An alternative approach to simulate a transient surface water flow system in three dimensions is HEC-RAS, the U.S. Army Corp of Engineers River Analysis System (U.S. Army Corps of Engineers, 2016). It is a software package that allows users to perform one-dimensional steady flow as well as one and two-dimensional unsteady flow and sediment transport simulations. The outputs of a 1-D steady flow model, 1-dimensional river profiles and 2-dimensional cross section profiles, could theoretically be used to interpolate a changing 3-dimensional surface water body. HEC-RAS has been used previously to simulate hyporheic exchange from in-stream geomorphic structures in conjunction with MODFLOW (Hester and Doyle, 2008; Endreny et al., 2011). In those studies, output from 1-dimensional steady state flow simulations were input as constant head values into a vertical 2-dimensional MODFLOW profile model. These studies provide a starting point for simulating BDA-induced flooding but are significantly simpler than simulations of 3-dimensional transient flow, the goal of this project.

HEC-RAS is a particularly appealing modeling platform for this project because of its use of geometric data. With the ArcGIS add-in 'HEC-geoRAS' (U.S. Army Corps of Engineers, 2011), one can use the sUAS derived DEMs to efficiently delineate cross-sections and stream channel profiles and import them into the main program (Fig. 12). Having been designed for simulating both natural and engineered flow systems (like culverts) it has built-in methods for adding BDA-like structures to stream profiles. The thorough documentation, plentiful on-line resources, and ubiquity in the civil engineering discipline ensure a rich support system.

One consideration to make when choosing to use HEC-RAS is whether to simulate steady state or transient flow. For scenario testing with the BDA and non-BDA groundwater flow models, a transient simulation would be preferred. However, in HEC-RAS, transient simulations tend to be more unstable and require additional parameters that may be poorly

defined for this system. When a test-run of a transient simulation was performed for this project, it crashed due to accumulated error on May 5th (the hydrograph had been gradually increasing due to spring snowmelt). It appears that incorrect delineation of the flood plain was causing water to pond up at the edges of each cross-section (Fig. 13). With a record of daily average flows, a steady state simulation for each day might be an alternative way to integrate HEC-RAS with groundwater flow models. That way, flow from the previous day could be used as the initial boundaries of upstream and downstream head, with a new daily flow increasing/decreasing the stream profile accordingly.

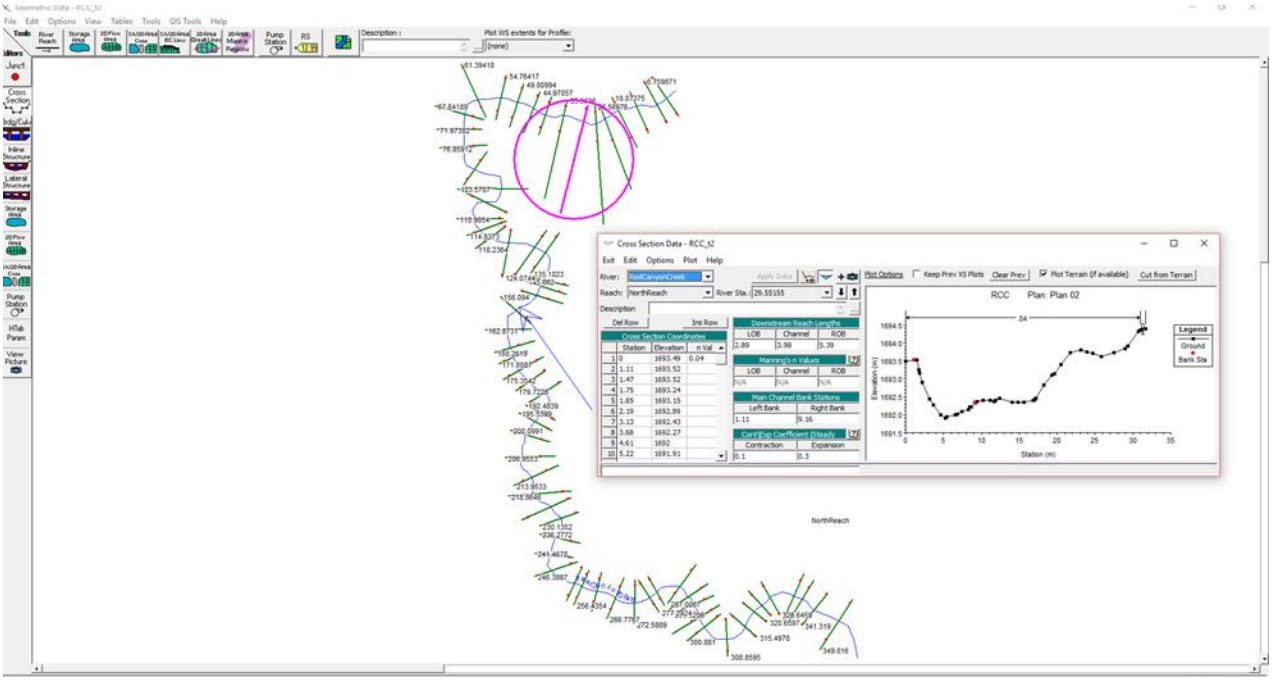


Figure 12. Example screenshot of using the Geometric Data Editor in HEC-RAS for the downstream Red Canyon Creek reach. Inset shows the pink-highlighted downstream cross-section from the left to the right bank, near the ox-bow.

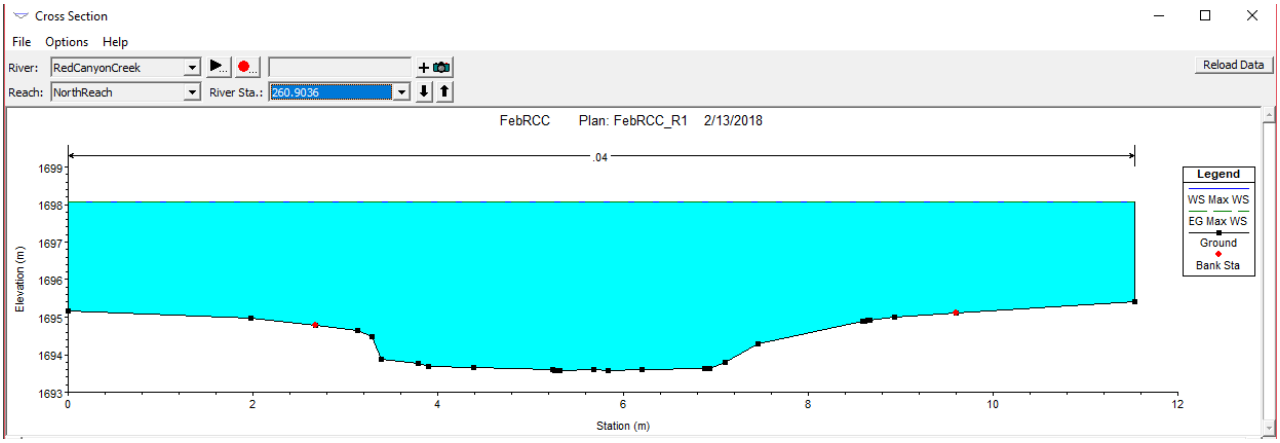


Figure 13. Example of a poorly delineated HEC-RAS cross-section. In an unsteady flow run, water could not dissipate into the floodplain and unrealistically accumulated along the banks, causing model instability and subsequent abortion of the model run.

After reviewing the literature on simulating surface water-groundwater interactions, an approach similar to Ou et al. (2013) may be the best way to incorporate surface water bodies into a MODFLOW model of BDAs. While HEC-RAS also seems to have a lot of potential, the dearth of literature using it for 3-dimensional groundwater flow models and the loose coupling between HEC-RAS and MODFLOW mean it may present significant challenges in application. Still either of these approaches have one additional, potentially important simplification; they ignore the unsaturated zone. Other modeling systems that incorporate additional flow processes, like unsaturated zone flow or quick flow may be of interest. Some of these modeling systems focus on the watershed scale and examine how land-use change (which BDA-installation effectively is) will alter hydrology on that scale, as demonstrated in the next section.

3.3 Watershed Scale Modeling with GSFLOW

GSFLOW is a coupled surface water and groundwater flow model that integrates the USGS Modular Groundwater Flow Model (MODFLOW) and the USGS Precipitation-Runoff Modeling System (PRMS) (Markstrom et al., 2008). GSFLOW was designed to incorporate land

surface and subsurface saturated and unsaturated zone flow. Primarily used to evaluate land-use and climate change, GSFLOW is potentially a powerful tool for understanding the interrelated watershed dynamics that result from beaver dams and BDAs. It utilizes numerically efficient algorithms that operate on a daily time step and has been used on models ranging from a few square kilometers to several thousand square kilometers. An example of GSFLOW input and a project framework for GSFLOW modeling of BDAs is shown in Fig. 14.

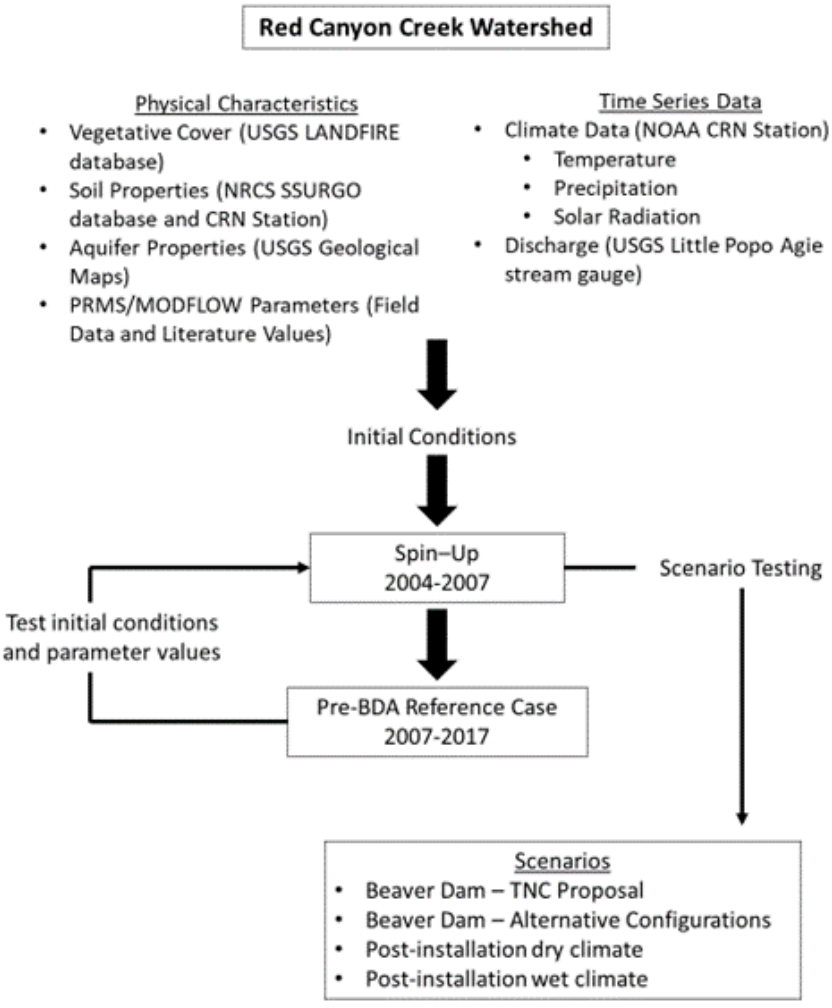


Figure 14. Proposed model design workflow. Data types and sources are placed in the top bulleted lists. Model calibration and scenario testing is presented at the bottom.

Past studies using GSFLOW share many objectives with those of BDA stream restoration studies, either through their focus on ecohydrology or their focus on climate driven hydrologic change. Within the former category, Essaid and Hill (2014) studied how water transmission in montane meadows is altered after stream channel incision. Their use of GSFLOW allowed for a direct comparison of how different hydraulic pathways (e.g. streamflow, evapotranspiration, saturated storage, etc.) responded under different incision scenarios (see Fig. 10 in Essaid and Hill, 2014). Their modeling efforts revealed that stream incision's effect on the subsurface system and its flow processes is analogous to groundwater withdrawal. Addressing both ecohydrology and hydrologic change, Carroll et al. (2016) evaluated the effect of both changing Pinyon-Juniper tree cover and projected mid-21st-century temperature changes on the Porter Canyon Experimental Watershed in central Nevada. They compared the differences between four different vegetation-climate simulations. Results indicated that small-scale (<0.5 km^2) removal of Pinyon-Juniper tree cover led to "undetectable changes in watershed-scale annual [water] yield estimates." This information is invaluable for land-managers interested in the hydrologic effect of specific management strategies. Lastly, Hunt et al. (2016) used GSFLOW paired with a stream temperature model to evaluate how a Wisconsin watershed would respond to future climate change scenarios. These three studies focused on land-use change and/or climate and used the GSFLOW modeling framework to forecast hydrologic changes. All three studies highlight the value of modeling all components of the hydrologic cycle and the potential of coupled surface water-groundwater models to evaluate the impacts of BDAs.

GSFLOW is well-suited for using remote sensing to aid with the development and calibration of hydrologic models. Many of the data products required to develop a conceptual model of a watershed in GSFLOW can be derived from remote sensing. For example, the DEMs

and landcover data required to delineate hydrologic response units (smallest spatial unit of a GSFLOW model with shared physical characteristics) can be derived from remotely sensed data. The U.S. governments LANDFIRE dataset is one such data repository for this type of information. In addition, previous GSFLOW studies have used remotely sensed data products as independent checks on model accuracy. Carroll et al. (2016) used average monthly NDVI values from Landsat imagery as an independent check on modeled estimates of depth to water (see Fig. 8 in Carroll et al., 2016) and found statistically significant correlation between these values. While this application is not wholly satisfying as a quantitative assessment of model performance, it does provide an additional metric for model validation. Tian et al. (2015) took a more nuanced approach, using an ET model based on remotely sensed MODIS and Landsat TM data products to cross-check their model results. The rapid development of remotely sensed data acquisition technologies in recent years is expected to lead to greater use of these data products in the future (Tian et al., 2015).

While GSFLOW provides a potentially novel way to estimate how land-use changes will affect watershed-scale hydrology, there are still a number of complicating factors with incorporating BDAs into this modeling system. Primary among these is the scale of potential BDA effects. Preliminary efforts made by Hafen (2017) implementing BDAs in a watershed scale groundwater flow model showed that even with a watershed at 100% dam capacity (defined as the density of dams that any given stream reach can support, based on available vegetation, geomorphic characteristics, and hydrology), damming would only be able to account for 3% of the water lost from snowpack in a warming climate. BDAs and natural beaver dams cannot be expected to be vast stores of water relative to the entire watershed, but tend to have greater impact on riparian ecosystems (Hafen, 2017; Majerova et al., 2015; Nyssen et al., 2011;

Puttock et al., 2017). With all the potential uncertainty involved in a watershed scale model, the small-scale effects of BDAs may be masked with such a complex modeling system like GSFLOW. Additional difficulties are added with representing high-resolution sUAS-derived imagery in a regional watershed scale model. However, as shown by other studies looking at smaller-scale land-use changes, if scale problems can be resolved and the appropriate data products procured, a GSFLOW model would add significantly to the literature on BDA stream restoration.

3.4 Concluding Remarks

A number of complicating factors exist in modeling BDA-induced groundwater storage using groundwater flow modeling. Possibly the most important is accurate linkage of surface water body shape to subsurface finite difference cells in a groundwater model. Once this problem is overcome, either by using HEC-RAS or some other surface water boundary condition package, subsurface groundwater storage could hypothetically be calculated on an annual timescale. From there, other interesting hydrologic process interactions could be studied.

4. Project Synthesis

While the goal of creating a well-calibrated, fully-functioning groundwater flow model to study beaver dam analogues along Red Canyon Creek with high-resolution aerial photography was not accomplished, much was learned from initial investigation of the utility and practicality of a number of approaches. The findings and recommendations presented in this report lay the foundation for continuation of BDA hydrologic modeling. In particular, the guidance presented

here should facilitate incorporation of Structure-from-Motion photogrammetry data products into hydrologic models.

In conclusion, while high-resolution aerial photography is a novel data acquisition technique with lots of potential, caution should be taking in unnecessarily relying on its use when other parts of a model are not well-constrained. And while modeling BDAs has many difficulties, as outlined above (though not insurmountable), one of the unexamined ones in this report is the lack of field data for how flow changes through a BDA. With those cautions in mind, a modeling framework that provides input on how best to install BDAs for achieving hydrologic goals, would be a valuable research goal of future work.

References

Adam, J.C., A.F. Hamlet, and D.P. Lettenmaier. 2009. Implications of global climate change for snowmelt hydrology in the twenty-first century. *Hydrological Processes* 23, no. 7: 962-972.

Agisoft LLC. Photoscan User Manual: Professional Edition, Version 1.4. http://www.agisoft.com/downloads/user-manuals/.

Barnett, T.P., J.C. Adam, and D.P. Lettenmaier. 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* 438, 303-309.

Beechie, T., H. Imaki, J. Greene, A. Wade, H. Wu, G. Pess, P. Roni, J. Kimball, J. Stanford, P. Kiffney, and N. Mantua. 2013. Restoring salmon habitat for a changing climate. *River Research and Applications* 29, no. 8: 939–960.

Beedle, D. 1991. Physical dimensions and hydrologic effects of beaver ponds on Kuiu Island in southeast Alaska. *MS Thesis*. Oregon State University, Corvallis.

Bhat, M.G., R.G. Huffaker, and S.M. Lenhart. 1993. Controlling forest damage by dispersive beaver populations: centralized optimal management strategy. *Ecological Applications* 3, no. 3: 518–530.

Burchsted, D. and M.D. Daniels. 2014. Classification of the alterations of beaver dams to headwater streams in northeastern Connecticut, USA. *Geomorphology* 205, 36–50. doi: 10.1016/j.geomorph.2012.12.029.

Carroll, R.W., G.M. Pohll, C.G. Morton, and J.L. Huntington. 2015. *Journal of the American Water Resources Association* 51(4). http://doi.org/10.1111/jawr.12285

Carroll, R.W., J.L. Huntington, K.A. Snyder, R.G. Niswonger, C. Morton, and T.K. Stringham. 2017. Evaluating mountain meadow groundwater response to Pinyon-Juniper and temperature in a great basin watershed. *Ecohydrology* 10, 1-18.

Collen, P. and R.J. Gibson. 2001. The general ecology of beavers (Castor spp.), as related to their influence on stream ecosystems and riparian habitats, and the subsequent effects on fish – a review. *Reviews in Fish Biology and Fisheries* 10, 439–461.

Devitt, D.A., L.F. Fenstermaker, M.H. Young, B. Conrad, M. Baghzouz, and B.M. Bird. 2011. Evapotranspiration of mixed shrub communities in phreatophytic zones of the Great Basin region of Nevada (USA). *Ecohydrology* 4(6): 807-822. http://doi.org/10.1002/eco.169.

Endreny, T., L.K. Lautz, D.I. Siegel. 2011. Hyporheic flow path response to hydraulic jumps at river steps: Hydrostatic model simulations. *Water Resources Research* 47, W02518. http://doi.org/10.1029/2010WR010014.

Essaid, H.I. and B.R. Hill. 2014. Watershed-scale modeling of streamflow change in incised montane meadows. *Water Resources Research* 50, 3: 2657-2678.

Gleick, P.H. 1987. Regional hydrologic consequences of increases in atmospheric CO2 and other trace gases. *Climatic Change* 10, no. 2: 137-160.

Gurnell, A. M. (1998) 'The hydrogeomorphological effects of beaver dam-building activity. *Progress in Physical Geography* 22, no. 2: 167–189.

Hafen, K. 2017. To What Extent Might Beaver Dam Building Buffer Water Storage Losses Associated with a Declining Snowpack? *MS Thesis*. Utah State University, Logan.

Hester, E.T. and M.W. Doyle. 2008. In-stream geomorphic structures as drivers of hyporheic exchange, *Water Resources Research* 44: W03417. http://doi.org/10.1029/2009WR005810.

Hunt, R.J., S.M. Westenbroek, J.F. Walker, W.R. Selbig, R.S Regan, A.T. Leaf and D.A. Saad. 2016. Simulation of climate change effects on streamflow, groundwater, and stream

temperature using GSFLOW and SNTEMP in the Black Earth Creek Watershed, Wisconsin: U.S. Geological Survey Scientific Investigations Report 2016-5091, 117 p., http://dx.doi.org/10.3133/sir20165091.

Huntington, J., K. McGwire, C. Morton, K. Snyder, S. Peterson, T. Erickson, R. Niswonger, R. Carroll, G. Smith, and R. Allen. 2016. *Remote Sensing of Environment* 185: 186-197. http://doi.org/10.1016/j.rse.2016.07.004.

IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Core Writing Team, R.K. Pachauri and L.A. Meyer*. http://doi.org/10.1017/CBO9781107415324.004.

Jaevernick, L., J. Brasington, and B. Caruso. 2014. Modeling the topography of shallow braided rivers using Structure-from-Motion photogrammetry. *Geomorphology* 213: 166-182. http://doi.org/10.1016/j.geomorph.2014.01.006.

James, M.K. and S. Robson. 2014. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surface Processes and Landforms* 39: 1413-1420. http://doi.org/10.1002/esp.3609.

Janzen, K. and C.J. Westbrook. 2011. Hyporheic flows along a channeled peatland: influence of beaver dams. *Canadian Water Resources Journal* 36, no. 4: 331-347.

Jensen, P.G., P.D. Curtis, M.E. Lehnert, and D.L. Hamelin. 2001. Habitat and structural factors influencing beaver interference with highway culverts. *Wildlife Society Bulletin* 29, no. 2: 654–664.

Jensen, J.R. 2000. Remote Sensing of the Environment: An Earth Resource Perspective. *Prentice Hall*, 2nd Edition.

Jobson, H.E. and A.W. Harbaugh. 1999. Modifications to the diffusion analogy surface-water flow model (DAFLOW) for coupling to the modular finite-difference ground-water flow model (MODFLOW). *U.S. Geological Survey Open-File Report* 99-217. https://pubs.er.usgs.gov/publication/ofr99217.

Lautz, L.K., and D.I. Siegel. 2006. Modeling surface and ground water mixing in the hyporheic zone using MODFLOW and MT3D. *Advances in Water Resources* 29 no. 11: 1618-1633, doi: 10.1016/j.advwatres.2005.12.003.

Lawler, J.J. 2009. Climate change adaptation strategies for resource management and conservation planning. *Annals of the New York Academy of Sciences* 1162, 79-98.

Majerova, M., B. Neilson, N. Schmadel, J. Wheaton, and C. Snow. 2015. Impacts of beaver dams on hydrologic and temperature regimes in a mountain stream. *Hydrology and Earth System Sciences Discussions* 12, 839-878.

Markstrom, S.L., R.G. Niswonger, R.S. Regan, D.E. Prudic, and P.M Barlow. 2008. GSFLOW-Coupled Ground-water and Surface-water FLOW model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005): U.S. Geological Survey Techniques and Methods 6-D1, 240 p.

Micheletti, N., J.H. Chandler, S.N. Lane. 2015. Structure from Motion (SfM) Photogrammetry. *Geomorphological Techniques* Chap. 2, Sec. 2.2. https://dspace.lboro.ac.uk/2134/17493.

Naiman, R.J., C.A. Johnston, and J.C. Kelley. 1988. Alteration of North American Streams by Beaver. *Bioscience* 38, no. 11: 753–762.

Nichols, W.D. 2000. Regional Ground-Water Evapotranspiration and Ground-Water Budgets, Great Basin, Nevada. *U.S. Geological Survey Professional Paper* 1628.

Nyssen, J., J. Pontzeele, P. Billi. 2011. Effect of beaver dams on the hydrology of small mountain streams: example from the Chevral in the Ourthe Orientale basin, Ardennes, Belgium. *Journal of Hydrology* 402, no. 1: 92-102.

Ou, G. X. Chen, A. Kilic, S. Bartelt-Hunt, Y. Li, and A. Samal. 2013. Development of a cross-section based streamflow routing package for MODFLOW. Environmental Modelling & Software 50: 132-143. http://doi.org/10.1016/j.envsoft.2013.09.012.

Pai, H., H.F. Malenda, M.A. Briggs, K. Singha, R.Gonzalez-Pinzon, M.N. Gooseff, S.W. Tyler, and the AirCTEMPS Team. 2017. Potential for Small Unmanned Aircraft Systems Applications for Identifying Groundwater-Surface Water Exchange in a Meandering River Reach. *Geophysical Research Letters* 44, 11,868–11,877. https://doi.org/10.1002/2017GL075836.

Pilliod, D.S., A.T. Rohde, S. Charnley, R.R. Davee, J.B. Dunham, H. Gosnell, G.E. Grant, M.B Hausner, J.L Huntington, and C. Nash. 2017. Survey of Beaver-related Restoration Practices in Rangeland Streams of the Western USA. *Environmental management*, 1-11.

Pollock, M.M., M. Heim, and D. Wener. 2003. Hydrologic and geomorphic effects of beaver dams and their influence on fishes. *American Fisheries Society Symposium* 37, 213-233.

Pollock, M.M., T.J. Beechie, J. M. Wheaton, C. E. Jordan, N. Bouwes, N. Weber and C. Volk. 2014. Using Beaver Dams to Restore Incised Stream Ecosystems. *BioScience* 64, 4: 279-290. doi: 10.1093/biosci/biu036.

Pollock, M.M., G.M. Lewallen, K. Woodruff, C.E. Jordan and J.M. Castro (Editors). 2017. The Beaver Restoration Guidebook: Working with Beaver to Restore Streams, Wetlands, and Floodplains. Version 2.0. *United States Fish and Wildlife Service*. Portland, Oregon. 219 p. Online at: http://www.fws.gov/oregonfwo/promo.cfm?id=177175812.

Prudic, D.E., L.F. Konikow, and E.R. Banta. 2004. A New Streamflow-Routing (SFR1) Package to Simulate Stream-Aquifer Interaction with MODFLOW-2000. *U.S. Geological Survey Open-File Report* 2004-1042. https://pubs.usgs.gov/of/2004/1042/ofr2004-1042.pdf.

Puttock, A., H.A. Graham, A.M. Cunliffe, M. Elliott, and R.E. Brazier. 2017. Eurasian beaver activity increases water storage, attenuates flow and mitigates diffuse pollution from intensively-managed grasslands. *Science of the Total Environment* 576, 430-443.

Rea, A.M. 1983. Once a River: Bird Life and Habitat Change on the Middle Gila. University of Arizona Press, Tucson, AZ.

Rosemond, A., and C. Anderson. 2003. Engineering role models: do non-human species have the answers? *Ecological Engineering* 20, 379-387.

Swain, E.D. and E.J. Wexler. 1996. A Coupled Surface-Water and Ground-Water Flow Model (MODBRANCH) For Simulation of Stream-Aquifer Interaction. *U.S.Geological Survey, Techniques of Water-Resources Investigations* Book 6, Ch. A6. https://pubs.usgs.gov/twri/twri6a6/index.html.

Tian, Y., Y. Zheng, C. Zheng, H. Xiao, W. Fan, S. Zou, B. Wu, Y. Yao, A. Zhang, and J. Liu. 2015. Exploring scale dependent ecohydrological responses in a large endorheic river basin through integrated surface water-groundwater modeling. *Water Resources Research* 51, 4065–4085. http://doi:10.1002/2015WR016881.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8(2): 127-150.

U.S. Army Corps of Engineers (USACE). 2011. HEC-GeoRAS: GIS Tools for Support of HEC-RAS using ArcGIS®, v. 4.3.93. *Institute for Water Resources, Hydrologic Engineering*

Center, Davis, California. http://www.hec.usace.army.mil/software/hec-georas/documentation/HEC-GeoRAS_43_Users_Manual.pdf.

U.S. Army Corps of Engineers (USACE). 2016. HEC-RAS River Analysis System, Hydraulic Reference Manual, v. 5.0. *Institute for Water Resources, Hydrologic Engineering Center, Davis, California*. http://www.hec.usace.army.mil/software/hec-ras/documentation/HEC-RAS%205.0%20Reference%20Manual.pdf.

Westbrook, C.J., D.J. Cooper, and B.W. Baker. 2006. Beaver dams and overbank floods influence groundwater – surface water interactions of a Rocky Mountain riparian area. *Water Resources Research* 42, 1-12. doi: 10.1029/2005WR004560.

Westoby, M.J., J. Brasington, N.F. Glasser, M.J. Hambrey, J.M. Reynolds. 2012. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology 179, 300-314. https://doi.org/10.1016/j.geomorph.2012.08.021.

Woodget, A.S., P.E. Carbonneau, F. Visser, and I.P. Maddock. 2015. Quantifying submerged fluvial topography using hyperspatial resolution UAS imagery and structure from motion photogrammetry. *Earth Surface Processes and Landforms* 40, 1-18. doi: 10.1002/esp.3613.

Vita

Name of Author: Nathaniel Chien

Degrees Awarded: Bachelor of Science, 2016, The College of William and Mary