12-2019

Potable Water Leakage Prediction and Detection using Geospatial Analysis

Jacob Tittle
East Tennessee State University

Potable Water Leakage Prediction and Detection using Geospatial Analysis

---

A thesis

presented to

the faculty of the Department of Geosciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Geosciences

---

by

Jacob Tittle

December 2019

Dr. Eileen G. Ernenwein, Chair

Dr. Andrew Joyner

Dr. Ingrid Luffman

Python, Kriging, Pressure, Water GIS, Water Leakage, Leakage Detection

ABSTRACT

Potable Water Leakage Prediction and Detection using Geospatial Analysis

by

Jacob Tittle

Due to increasing water treatment costs and conservation needs, traditional water loss analysis and acoustic leak detection methods are becoming heavily scrutinized by water utilities. This study explores water loss in Johnson City, Tennessee and how geospatial data analysis techniques improve water loss mitigation. This project uses sample water system pressure data and ordinary kriging spatial interpolation methods to identify leakage areas for further investigation. Analysis of existing geographic information system (GIS) water utility datasets with interpolated hydraulic grade values at sample water pressure points produce manageable survey areas that pinpoint areas with possible water leakage. Field detection methods, including ground-penetrating radar (GPR) and traditional acoustic methods, are employed to verify leakage predictions. Ten leakage areas are identified and verified using traditional acoustic detection methods, work order research, and GPR. The resulting data show that spatial analysis coupled with geospatial analysis of field pressure information improves water loss mitigation.

ACKNOWLEDGMENTS

I thank my committee chair, Dr. Eileen Ernenwein, for her research assistance, and meaningful input during the revision process. I also thank Dr. Ingrid Luffman and Dr. Andrew Joyner for serving on my committee and providing input. Each of my committee members have guided my continuing education in meaningful ways that I cannot describe.

I thank the City of Johnson City Water and Sewer Services Department for the opportunity to use their water system as the subject of this project. The staff members supported me through field data collections, patiently answered water system questions, and allowed access to data. Without their help and compliance, this project would not be possible.

Lastly and most importantly, I thank Alana Claxton for her support through the entire project. We went back to school together, studied together, and finished together. Without Alana, this project would have never materialized.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Water is a fundamental building block in support of human populations and the efficient distribution of water is crucial to maintenance of sustainable communities. The treatment and distribution systems responsible for supplying water serve as the arteries and veins of communities and the sustainable development of a city must include the sustainable use of water in an era of rapid urbanization (Xu et al. 2014). Due to aging infrastructure and rising potable water treatment costs within water utilities across the globe, water loss/leakage has become an increasingly critical topic. Until the 1990s, community water utilities had no way to quantify potable water loss within their systems (Frauendorfer and Liemberger 2010). As water conservation needs increase and geospatial technology continues to advance, public and private water utilities are looking at ways to mitigate water loss, ranging from hydraulic modeling to improved methods in field leak detection (e.g. pressure monitoring, acoustic leak detection, and consumption monitoring). The goal of this project is to help reduce water loss in Johnson City, Tennessee's water distribution system by detecting areas of potable water leakage using a combination of fire hydrant water pressure measurements and spatial interpolation methods to identify water leakage areas. After leakage areas are identified, field detection methods including ground-penetrating radar (GPR) and traditional acoustic leak detection are used to test leakage areas and pinpoint leaks.

Within recent years, quantifying water leakage has improved due to the creation of the Infrastructure Leakage Index (ILI). ILI defines a community water utility's non-revenue water and identifies real losses, or water loss occurring by leakage before the point of sale (Winarni 2009). The ILI is calculated by dividing a water distribution system's Current Annual Real Losses (CARL) by the Unavoidable Annual Real Losses (UARL), which represents the lowest achievable

annual real losses for a well-managed distribution system (Samir et al. 2017). Recently, states within the U.S., including Tennessee, began requiring water utilities to report ILI information on an annual basis using water audit tools supplied by the American Water Works Association (AWWA). Other states are also in various stages of ILI report adoption and look to reduce water leakage (Hodgkins et al. 2016).

Current approaches to water leakage monitoring include visual monitoring for signs of leakage, noise monitoring, and flow and pressure monitoring (Xu et al. 2014). Visual monitoring includes identifying visible clues of leaking water such as surface water and anomalous vegetation growth and is usually reported by the public. Noise or acoustic monitoring has long been used as a leakage detection method and identifies water leakage by using acoustic equipment to capture noise created by leaking pipes underground. The third method of flow and pressure monitoring acknowledges that leakage can cause change in hydraulic characteristics, which causes pressure to decrease and flow to increase (Xu et al. 2014). This study focuses on flow and pressure monitoring and uses pressure monitoring techniques to direct field monitoring efforts using both acoustic sounding and GPR.

Using GIS and hydraulic modeling techniques, preliminary spatial analysis has improved efficiency in leak detection by discovering areas of water loss before field leak surveys are conducted. Water distribution systems are composed of large complex pipe networks and breaking these networks into smaller areas enables improvements in leakage management efforts. Efficiently detecting and defining these smaller areas known as District Metered Areas (DMAs) and measuring incoming/outgoing flows have enabled researchers to identify high leakage areas (Scibetta et al. 2013). Projects like the Italian H20LEAK project was successful in mitigating

water loss by using GIS to help identify and partition water networks into DMAs and discover localized leakage by monitoring pressure and flow values (Candelieri and Messina 2012).

Dating back to the 1980s in the United Kingdom, many large water utilities adopted the practice of creating DMAs to quantify and manage water leakage, but the process becomes complex in large urbanized areas (Savic and Ferrari 2014). When creating DMAs, large water system networks are split into independent sub-networks, making water system management easier (Izquierdo et al. 2009). After dividing the water system into multiple DMAs, measuring incoming and outgoing flows for each DMA allows for the quantification of water losses (Scibetta et al. 2013). Once DMA analysis verifies high water loss sub-systems, hydraulic grade elevations converted from water hydrant flow information or other known pressure points compared to water tank elevations should indicate high demand areas (Walski 1983). These high demand areas may indicate water leakage, thus creating smaller targeted areas for field leak detection surveys.

This project uses spatial analysis of hydrant pressure tests to guide water leak detection surveys. Using Johnson City's existing DMA boundary information and water network GIS data coupled with water pressure test information, the sample area shown in Figure 1 illustrates the 1838A DMA and serves as the project focus area. Using past water consumption information, City of Johnson City Water Department personnel identified 1838A as the utility's DMA with the highest water loss. This has been determined by metering all water flowing into the DMA and comparing the measured volume of water to customer consumption and accounting for any known loss through fire hydrant flows or construction.

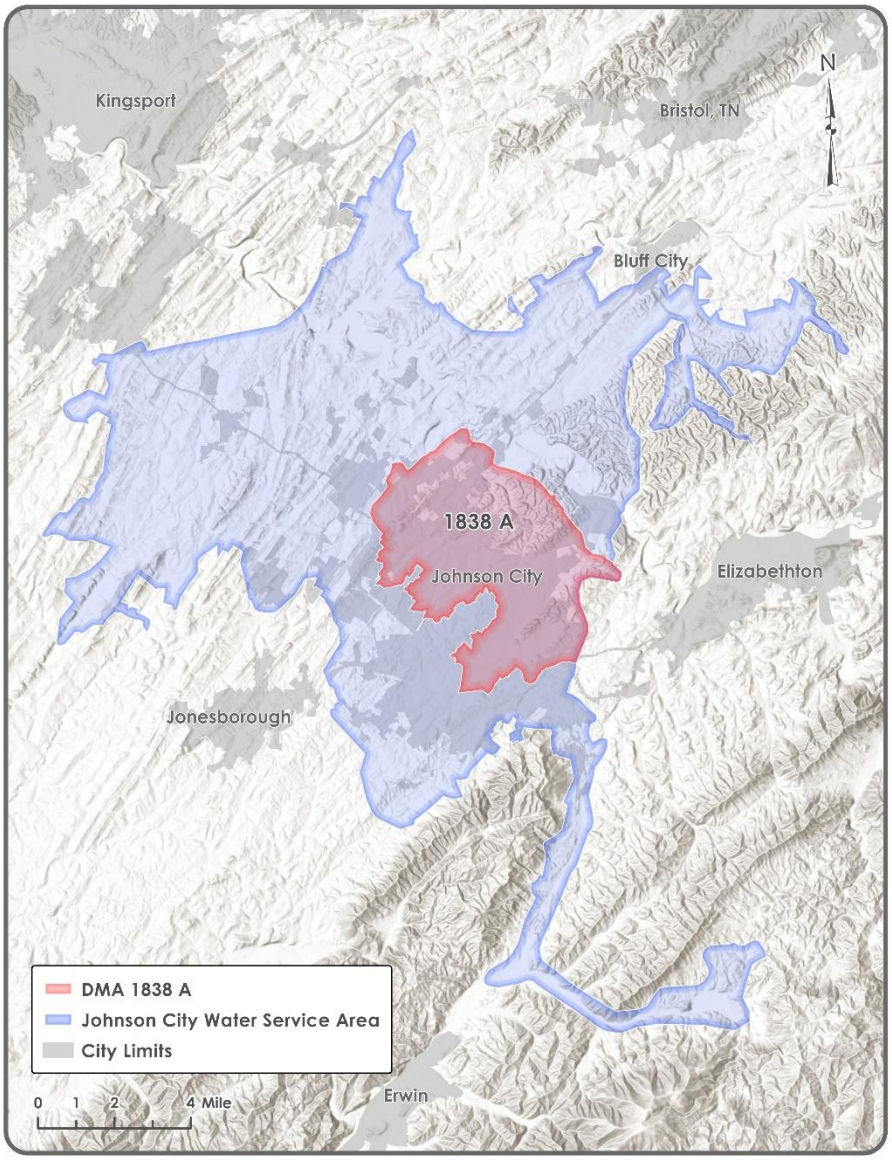


Figure 1. Johnson City Water System and 1838A Leakage Study Area

Kriging Interpolation and Hydraulic Grade

Predicting areas with water leakage requires interpolating hydraulic grade points taken from the field pressure test data. The hydraulic grade surface is then used to identify zones of hydraulic grade values lower than the local water tank elevations. A hydraulic grade value below local water tank elevation indicates high demand and possible leakage, although exceptions to that rule may occur (Wu et al. 2010). Areas illustrating hydraulic grade values below water tank

elevations must also be analyzed for specialized conditions involving pressure reducing valves or pumped areas exhibiting high pressures above water tank elevation. These areas are scrutinized for localized dips in hydraulic grade, thus also predicting water leakage.

Ordinary prediction kriging is an appropriate interpolation method due to its ability to be flexible in dealing with slight departures from the initial data assumptions such as data normality, trends, and spatial autocorrelation along with its reputation as the "work-horse" of geostatistics (Oliver and Webster 2014). To use ordinary kriging, the initial data must conform to a normal distribution, be spatially autocorrelated, and not exhibit any overarching global trends (Scheeres 2016). After ensuring the data meets ordinary kriging assumptions, model parameters are chosen to create a semivariogram, which can be plotted as a graph showing variance of measured distances between all sampled pair locations (Scheeres 2016). Due to the local nature of kriging, a minimum and maximum number of neighbors is selected as well as dividing each neighborhood into octants if points are unevenly scattered (Webster and Oliver 2007). Lastly, kriging models use cross-validation to check for appropriate fit by omitting each point from the data and predicted its value by ordinary kriging with the proposed model (Oliver and Webster 2014). Although a goal of the project is to automate leakage surface production, the interpolation process must be developed first to establish the appropriate tool settings. If the kriging process is to be replicated in other areas of the water system, a unique model must be developed per individual area.

Process Automation

The data collection, post-processing, and analysis was achieved through a mixture of open source and proprietary geospatial technologies including data collection interfaces in the Cartegraph work order management software, Esri's ArcGIS Pro geospatial software, and the Python programming language. The general process flow involved data collection by field

personnel in Cartegraph, automated data cleanup and manipulation using Python, then interpolating pressure test point data using kriging tools within ESRI's software platform. Although the process is described in this study using the previously mentioned tools, other technologies would also be able to replicate the process. The main function of the work order software is to provide a mobile solution to capture static pressure information in the field. Any other field solution would work if able to collect text and numeric information and provide it in a tabular format over the web. Other geospatial platforms could also be used including QGIS or any GIS software containing interpolation capabilities and providing access through a Python library.

The Python component of this project was used to automate daily data imports and maintenance required to dependably produce leakage area results. Since the project is applied to a real-world water distribution system, data were generated daily, creating the need to automate labor intensive tasks. To do this, the pandas and geopandas Python libraries were used to handle most of the data management by performing data imports, data cleanup, data table merges, and hydraulic grade calculations. This was accomplished by importing regularly generated pressure information into pandas data frames and hydrant locational information into geopandas spatially enabled geodataframes. The term data frame refers to an object designed to align structured data into rows and columns within these Python libraries (Harrison and Prentiss 2016). The pandas library gets its name from the term "panel data," which refers to three dimensional datasets found in statistics and is ideal for working with tabular data because of an ability to organize structured data (McKinney 2010).

Any data frame created with geopandas is known as a geodataframe. Geodataframes are generally similar to pandas data frames, but have an added geometry column containing locational information used in a GIS. The geopandas library also uses the shapely Python library and enables

users to perform GIS functions within Python, without requiring a geospatial database (Geopandas 2019). In addition to using pandas and geopandas, the resulting script was developed using multiple other libraries including datetime, which is used to manage date information within the pressure data, and fiona, a file handling package that can work with geodatabases. Esri's arcpy library was also used to handle the project's interpolation process and requires licensing, which was covered by using an enterprise license agreement. Each of the Python libraries are open source except for arcpy, which can be substituted with the QGIS library if an open source GIS platform is used.

The Python language coupled with the pandas library has recently been used to perform data analysis within water leakage studies and help manage water utilities. One UK study employed the pandas library as a proof of concept to discover leakage using nighttime-flow time series data and to predict household water consumption (Wills et al. 2017). Also, the widely used open source EPANET water modeling software has incorporated pandas data frames for time series analysis in the Water Network Tool for Resilience (WNTR) Python package, which is a new tool to model water hydraulic characteristics and simulate water utility disaster scenarios (Klisel et al. 2018). The geopandas library has not yet been explored in the same manner as pandas within the water industry. This study utilizes pandas and geopandas to manage locational data in relation to water leakage, helping to develop points to feed into the kriging model.

Field Leak Detection Techniques

Physical water loss is the product of numerous leak events that consist of reported leaks, unreported leaks, and background losses with flow rates too low for detection by traditional equipment (Samir et al. 2017). After the initial pressure testing and creation of a leakage prediction surface, high proabability leakage sites may be investigated using either acoustic detection

methods or GPR surveys. A third investigation technique involves past work order repair research in case a repair has already been performed. This step is not technically a field method but a necessary step to confirm leakage predicted by older pressure test data.

The most common field method for locating water leakage utilizes acoustic technologies to detect sounds created by leaking pipes. Acoustic leak location detects water leakage sounds caused by velocity of flow and the size and condition of the pipe opening. Also, different detectable sounds are made depending on whether the leak is discharging under water or into the air (Babbitt et al. 1920). Currently, acoustic water leakage field techniques miss significant leakage occurring in plastic piping (Hunaidi et al. 2000). Acoustic leak detection methods typically work well in metallic piping, but leaks in plastic piping remain difficult to locate due to the inability of sound to travel through plastic materials (Hunaidi et al. 2000). Additional technologies have been developed to advance traditional acoustic leak detection, but problems with background noise within heavily populated areas and issues associated with plastic piping persist (Hunaidi 2012).

Another field solution, GPR, is currently the most common geophysical survey technique employed to map underground utilities. In addition to utility location, GPR surveys extend to water leak detection. Instead of detecting leakage by audible noise, GPR can detect leakage by water's reflection, signal attenuation, or underground voids created by leaking water. The method reflects radio waves off objects below the surface. The waves act as a digital tape measure, allowing the surveyor to estimate the sub-surface object's depth (Witten 2006). Depending on soil conditions, GPR can detect buried pipe, increased soil moisture, and voids created around piping due to water leakage (Hunaidi 1998).

Water System and Project Background

The study focuses on Johnson City, Tennessee's public water utility, which owns and manages over 950 miles of water mains consisting of pipe varying in material and age. Johnson City manages its water distribution system by splitting the system into multiple DMAs. The water department had previously analyzed the amount of water supplied to each area and sold to customers over a multiyear period, calculating water loss amounts per DMA. The 1838A DMA consistently reported the highest water leakage in Johnson City's water distribution system and was picked to be the project focus area.

The chosen 1838A DMA contains 247 miles of water mains and currently serves 33 percent of the utility's 43,000 water accounts. The zone boundaries are established using strategically closed water valves known as isolation valves. These valves are commonly used to isolate the flow of water to turn off a portion of a system and may be intentionally kept closed to control area boundaries (Walski et al. 2003). Also, 1838A contains a number of pressure reducing valves, which serve to protect areas prone to damage caused by high pressures (Walski et al. 2003). Water flowing into 1838A is monitored through metering known entry points and compared with water sold to help quantify leakage.

The 1838A DMA also contains some of the oldest water distribution infrastructure within the city, containing pipe installation dates ranging from the 1890s through the present day. The area's majority of pipe materials include cast iron, ductile iron, polyvinyl chloride (PVC), and galvanized iron. The cast iron water mains are the oldest pipe materials within the DMA, while ductile iron and PVC are the youngest pipes in the ground. Large diameter, equal to or greater than six inches, cast iron pipe within the DMA present a significant amount of water leakage due to age and a tendency towards corrosion. Cast-iron pipe, the oldest water mains in most water systems,

usually develop corrosion pitting along pipe segments that cause breakage and are difficult to predict (Rajani and Makar 2000). Also, small-diameter galvanized water mains installed within the mid-twentieth century are largely identified as a source of water leakage by the water utility. Each year, a portion of these galvanized water mains are scheduled for replacement. Other pipe materials found in smaller quantities include copper, concrete, and high-density polyethylene (HDPE), along with a portion of unknown materials. Figure 2 shows 1838A's material quantities reported in miles of pipe.

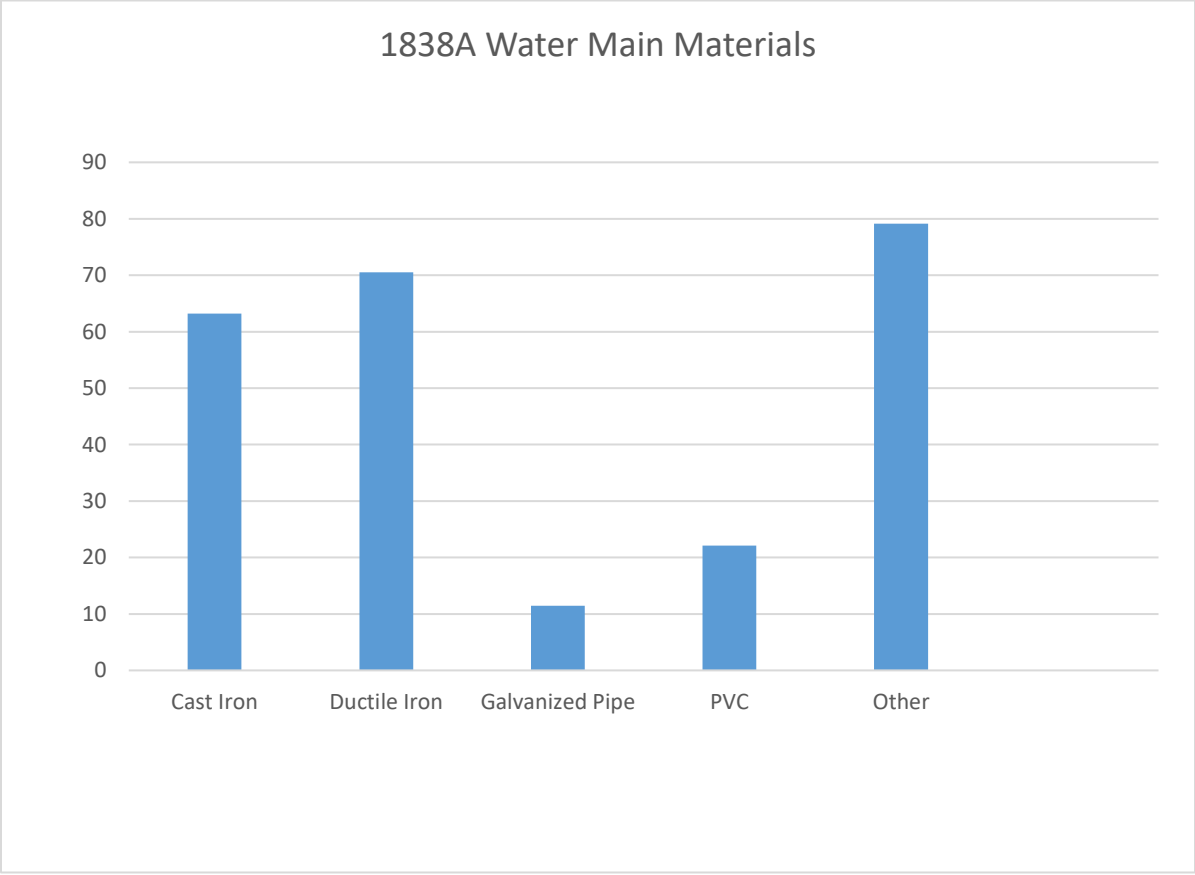


Figure 2. 1838A Water Pipe Material Percentages by Mileage

In addition to containing the oldest pipe materials in the water system, 1838A encompasses Johnson City's downtown area and is in close proximity to the city's largest stormwater

infrastructure. The downtown area's stormwater system includes multiple streams, causing water leakage to not always surface, going to stormwater instead. This creates difficulty in detecting water leakage due to non-surfacing water leaks and inhibits traditional acoustic leak detection techniques due to the typical sound pollution associated with densely populated downtown areas.

As for storage infrastructure, the 1838A DMA contains three water tanks that serve as a guideline for expected hydraulic grade. The tanks, named Carter Hill, Masters Knob and Tannery Knob, all have a top elevation of 1838 feet above sea level and the number 1838 in within the DMA's name indicates the top elevation of each water tank in the zone. The hilly terrain that comprises the DMA also presents the need for several water booster pumps and pressure reducing valves (PRVs). Homes sitting at higher elevations within the DMA rely on pumps to maintain water pressure while PRVs reduce pressure in areas experiencing high water pressure that may damage residential plumbing and burst pipes (Walski et al. 2003). Areas around these features exhibit hydraulic grade values differing from the typically observed values.

CHAPTER 2

METHODS

The process developed for this project includes multiple steps that require field data collection, data analysis, and field verification. The workflow used to generate areas exhibiting possible water leakage is described in the following sections. The first section briefly describes the initial pressure dataset and additional field pressure tests. The second section explains the overall GIS workflow developed to process pressure test data and produce a leakage prediction surface. Although incorporated into the GIS workflow, the kriging surface model development is described separately and includes selected parameters specific to 1838A. Another section incorporates the previous processes into an automated script. Lastly, field techniques used to verify predicted leakage areas are described. Figure 3 illustrates the general workflow, which breaks the process into data collection, data management and analysis, interpretation, and field verification.

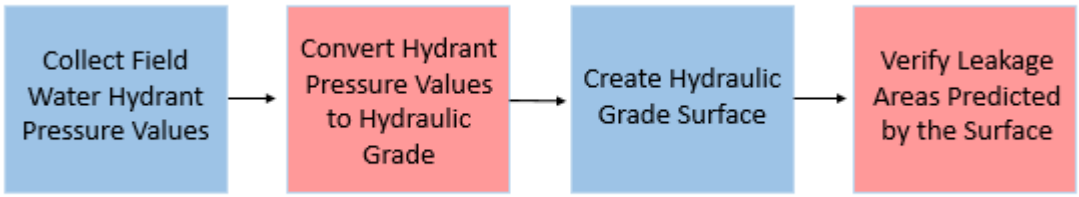


Figure 3. General Leakage Detection Workflow

Data Collection

The initial focus on the 1838A DMA uses the Johnson City's water distribution Esri feature dataset obtained by request from the Johnson City Water/Sewer utility. The dataset contains multiple feature classes consisting of location and attribute information for water mains, hydrants, water meters, tanks, system valves, and DMA boundaries. In addition to existing GIS data, water

hydrant flow history information was obtained from the Johnson City Fire Department's work order records. The flow history reflected the most readily available data containing hydrant pressure readings taken during routine fire hydrant flow inspections between 2015 and 2017 and provided the baseline pressure test information. Additional pressure test data taken from water hydrants were acquired through the water utility's Cartegraph work order system, which is dynamic and was later incorporated into the Python automation. The water utility's pressure tests replace fire hydrant pressure test information as newer data are collected.

Overall GIS Workflow

The GIS data management and analysis workflow used to generate areas exhibiting possible water leakage is described in the following section. The general steps are illustrated within the flowchart in Figure 4, which supplies a road map to reproduce results in similar water distribution systems or DMAs. The green polygons represent data inputs and blue/red squares represent processes performed within ArcGIS. When performed in order, the workflow produces an interpolated surface of hydraulic grade elevations and identifies possible water leakage areas. The entire GIS workflow illustrated in Figure 4 was automated with a script written in the Python programming language and utilized multiple Python libraries to perform the necessary data cleanup, manipulation, and interpolation steps described later in this section.

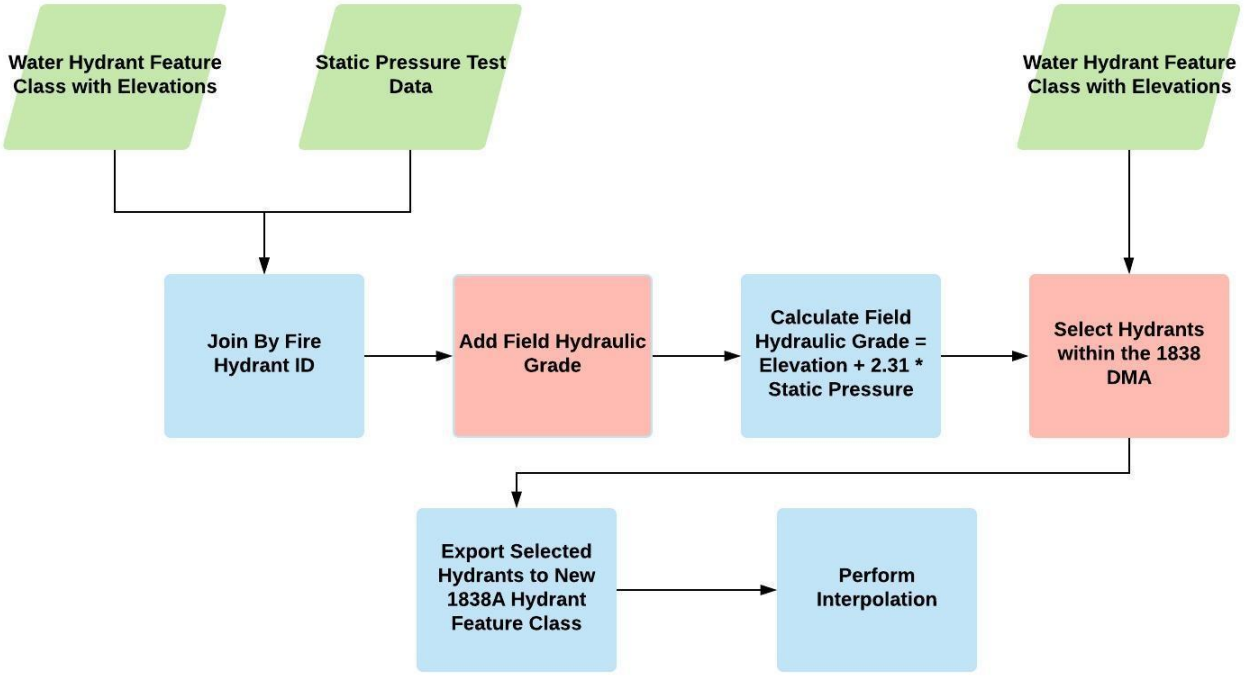


Figure 4. GIS Workflow for Water Hydrant Static Pressure Interpolation

Hydrant pressure test data were first imported into a table and joined to the water hydrant feature class using the fire department's hydrant identification number. Null values and pressures above 300 pounds per square inch, which reflect any data entry errors, were removed. Each hydrant required an elevation value, so elevations of all public water hydrants were either previously surveyed by traditional field methods or extracted from the 2015 Tennessee LIDAR dataset generated digital terrain model (DTM). A new field named HydroGrade within the hydrant feature class was then created to record hydraulic grade calculations.

At each pressure test site, the hydraulic grade was calculated for each hydrant using the elevation added to the product of a 2.31 unit conversion factor and the flow data's static pressure value. The 2.31 conversion factor represents 2.31 feet for every pound per square inch of pressure

(Walski et al. 2003). Shown in the equation (1) below, hydraulic grade (H) equals the elevation (z) plus the 2.31 conversion factor multiplied by the pressure (p).

$$H = z + 2.31p \quad (1)$$

Ideally, if no leakage is present, hydraulic grade values would equal the elevation of the water tank supplying water to the DMA. The 1838A DMA is part of a larger water pressure zone that is supplied by multiple tanks with an elevation at the top of each tank averaging 1838 feet above sea level. The 1019 hydrants with completed hydraulic grade calculations within the 1838A DMA were then selected and exported into a new feature class. Figure 5 illustrates the hydrant locations within the 1838A DMA and area water mains.

The initial 1838A DMA boundary polygon includes a large area with no hydrants available for static pressure testing. This area and a small area in the southwest were removed because interpolated values there would be invalid. Along with the hydrant test location points, Figure 5 also shows the trimmed DMA zone after redefinition. This step prepared the data for interpolation and reduced the possibility of false leakage predictions in large areas lacking pressure test information.

After importing pressure tests, data cleanup, and calculating hydraulic grade values at each hydrant pressure test site, the resulting pressure test point locations were fed into an ordinary kriging interpolation model using ArcGIS Pro's Geostatistical Wizard toolset. The kriging model, developed to be replicated in an automated workflow, is used to specify customized parameters appropriate to hydrants within the 1838A DMA and creates a geostatistical layer within ArcGIS Pro. After the layer was produced, the surface was clipped to the boundary of 1838A. This process is further explained in the kriging development section and script development sections.

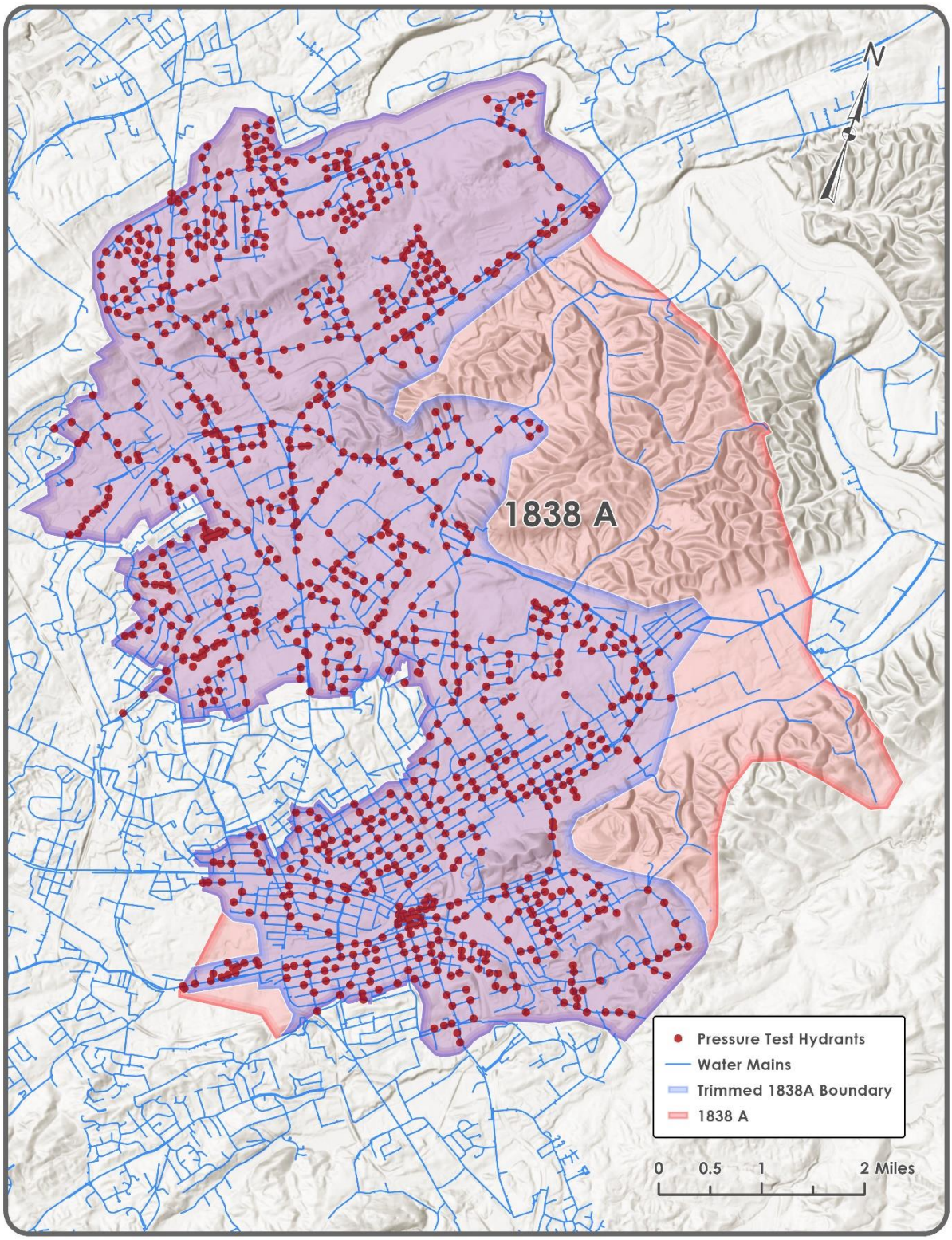


Figure 5. Hydrants Located Within the Redefined 1838A Boundary

Kriging Model Development

The pressure test points converted to hydraulic grade were analyzed for ordinary kriging model suitability by checking for normal distribution, spatial autocorrelation, and any overarching global trends. The histogram illustrated in Figure 6 shows that the data are generally normally distributed. The spatial autocorrelation report in Figure 7 shows a 0.4105 Moran's I value along with a z-score of 29.9673 and pseudo p-score of 0.0000. These numbers indicate that there is less than one percent likelihood that the data are random in space and suggests spatial autocorrelation. A test for overarching global trends is illustrated in Figure 8. Although showing a slight increase, the plotted polynomial trend lines are relatively flat, suggesting no overarching global trends. If overarching global trends were present the line would reflect a more dramatic curve. The graph in Figure 8 illustrates the lack of a global trend within the test data and serves to reinforce the kriging method choice for the purposed of the water leakage project.

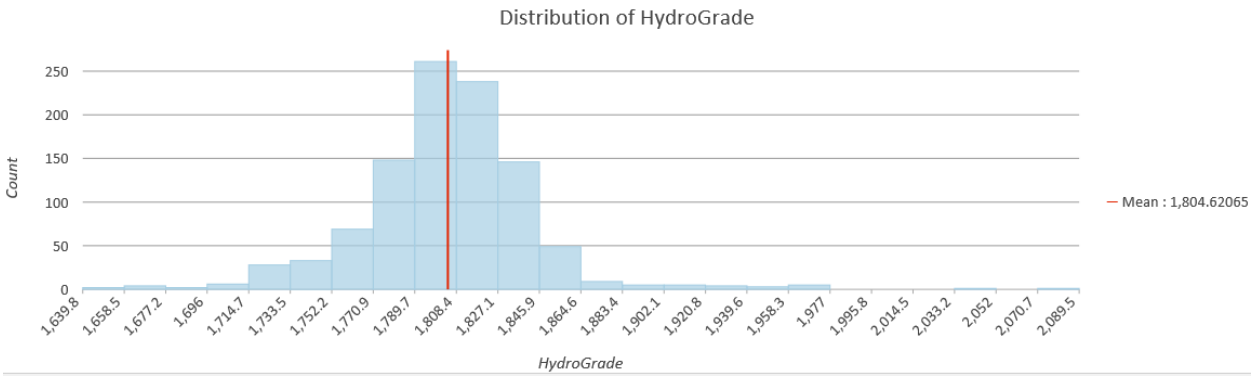
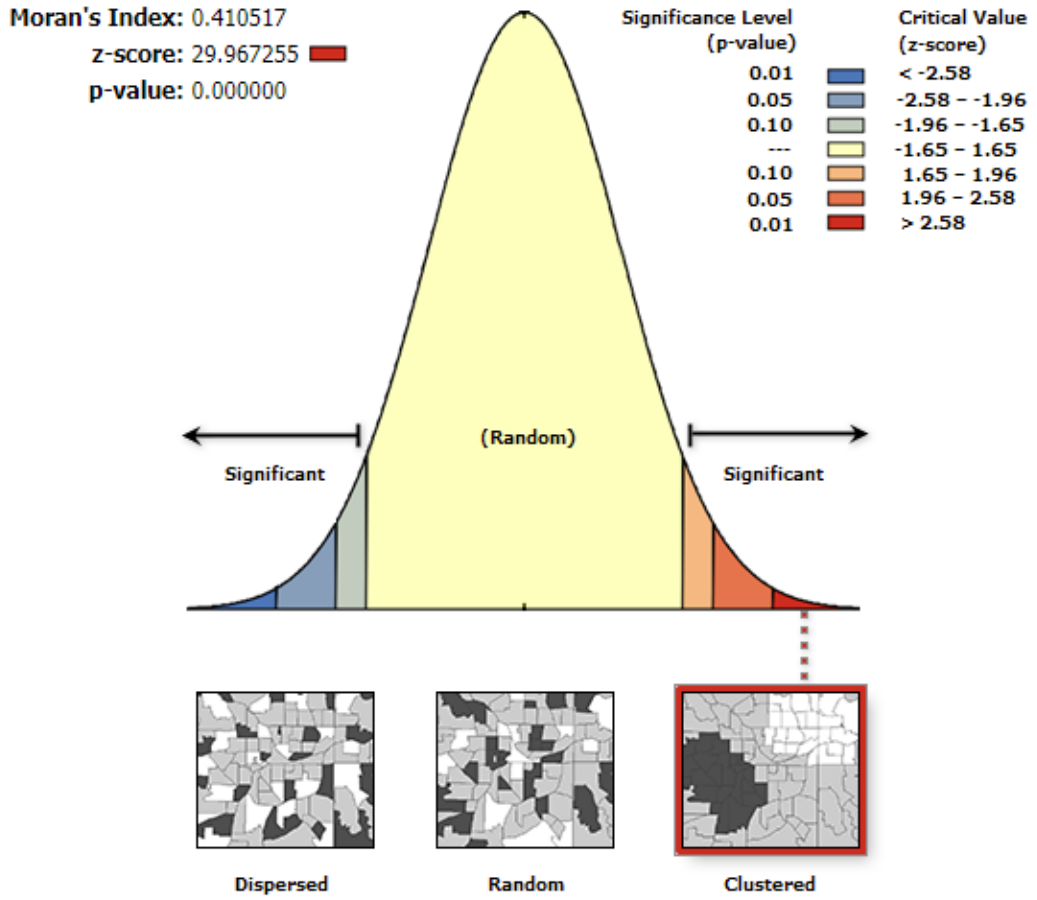


Figure 6. HydroGrade Suggested Normal Distribution Illustrated in ArcGIS Pro

Spatial Autocorrelation Report



Given the z-score of 29.967255, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index:	0.410517
Expected Index:	-0.000982
Variance:	0.000189
z-score:	29.967255
p-value:	0.000000

Figure 7. Spatial Autocorrelation Report in ArcGIS

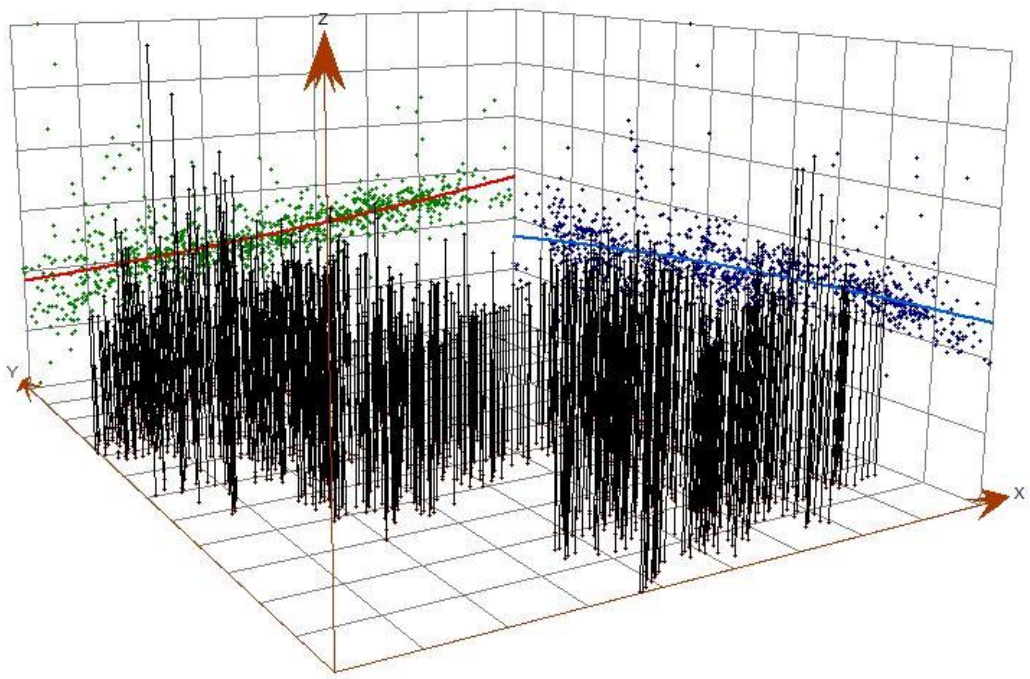


Figure 8. Analysis Plot Used to Detect Global Trends

After the data were determined to be a good fit for ordinary kriging, the hydrant pressure test data were plugged into the Geostatistical Wizard to create an appropriate model. Parameters were selected to create a semivariogram including a nugget or error value, number of lags, and lag size. Within ArcGIS, this was achieved using the optimization tool, which picks lag size and nugget values that result in the lowest mean standard error (Esri 2019). Figure 9 displays the semivariogram graph and optimized parameters. For the model displayed in Figure 9, the number of lags was set to 12, the lag size was 393.9094 feet and the nugget value was 335.8350. Other values in the optimized model are also displayed in Figure 9 including the model nugget measurement error, model parameter, major range, and partial sill number.

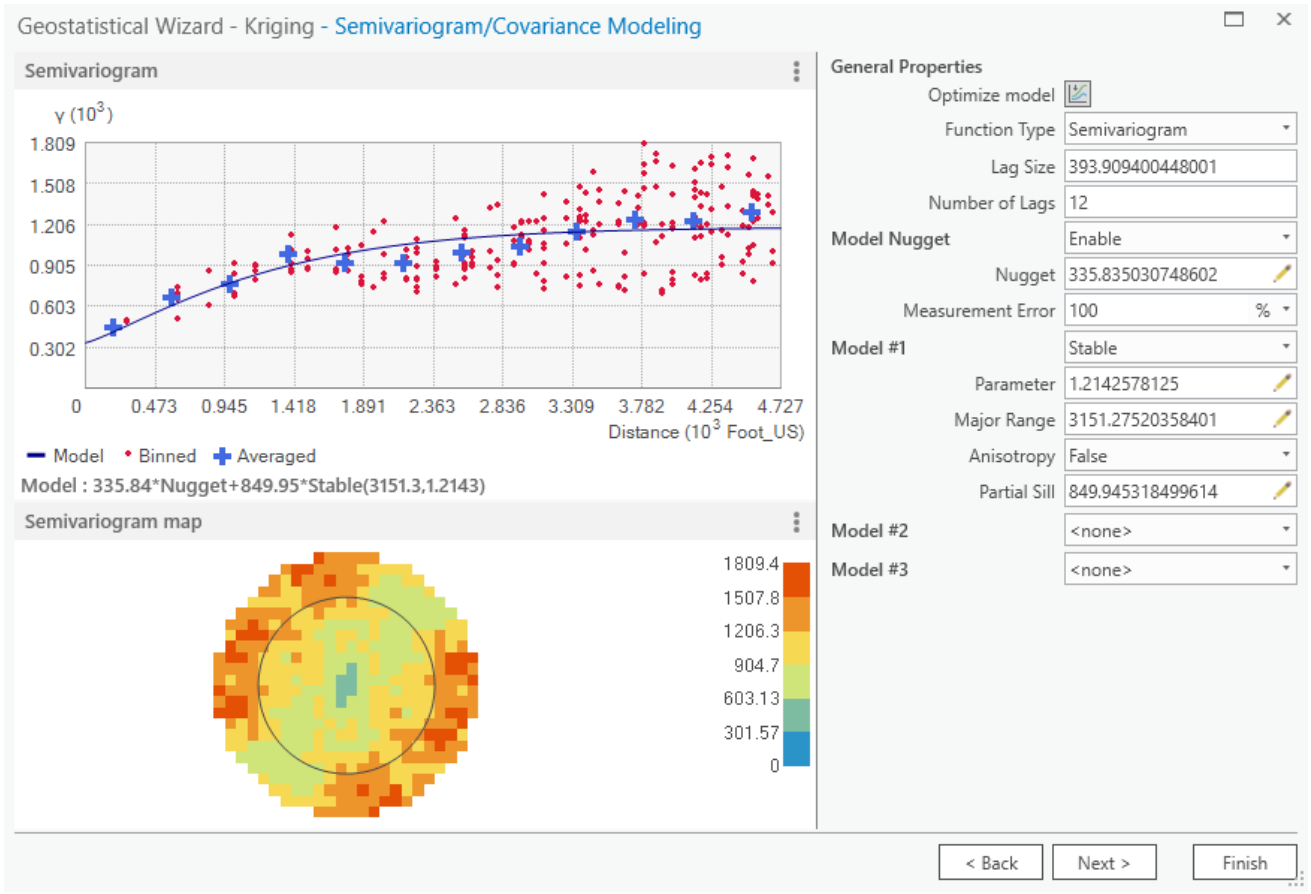


Figure 9. Kriging Semivariogram Modeling Values and Graphs

The next step to develop the model specified attributes relating to the kriging search neighborhood and included setting the neighborhood type, sector type, and the maximum and minimum number of neighbors. For this model, the standard neighborhood type was selected. As recommended by Oliver and Webster's 2014 article about computing variograms and ordinary kriging, the maximum neighborhood was set to 25 neighbors, the minimum was specified as 7 neighbors, and the sector type was set to 8 sectors (Oliver and Webster 2014). The searching neighborhood settings are illustrated in Figure 10 as well as a map displaying the sample kriging surface with an example neighborhood broken into 8 sectors.

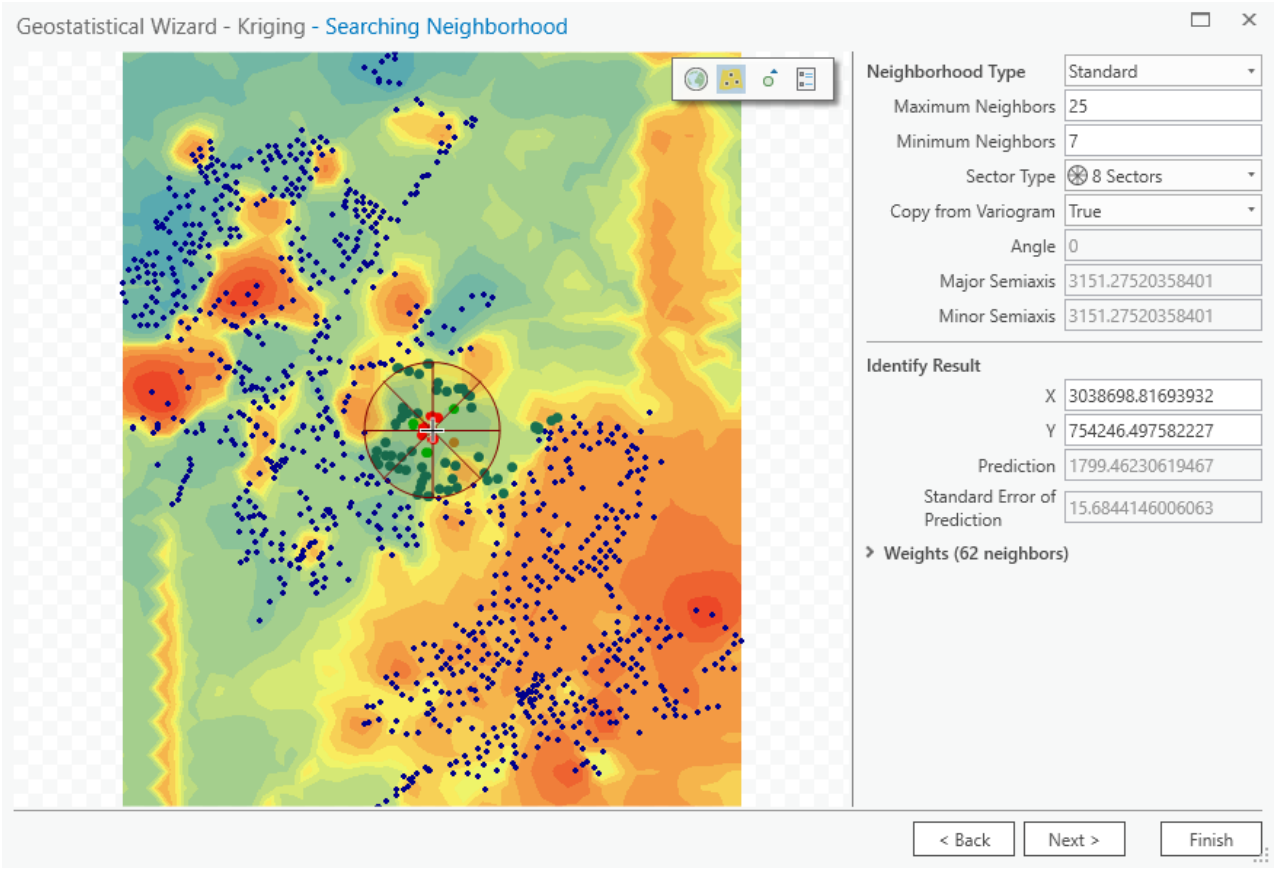


Figure 10. Search Neighborhood and Sector Type

Figure 11 displays the cross-validation report generated to check the model validity. The left side displays a graph based on predicted error and the right portion reports the data summary and a record table of all the points. The summary includes multiple values including three that are used to judge validity of the kriging model. These are mean error, mean standardized error, and the root mean square standardized error. Both the mean error and mean standardized error are close to zero, which is an indication that the model is accurate. The root mean square standardized error is close to 1, which also indicates an appropriate ordinary kriging model. A root mean square standardized error over 1 underestimates prediction variabilty, while a value under 1 overestimates variability (Oliver and Webster 2014). Also, the root mean square error was 27.5515 and the average standard error was 24.9850. After model development, the chosen parameters were

exported to an xml document for later incorporation in the Python script, allowing automation and replication of the ordinary kriging process. Figure 12 displays a sample kriging surface using the 1838A model.

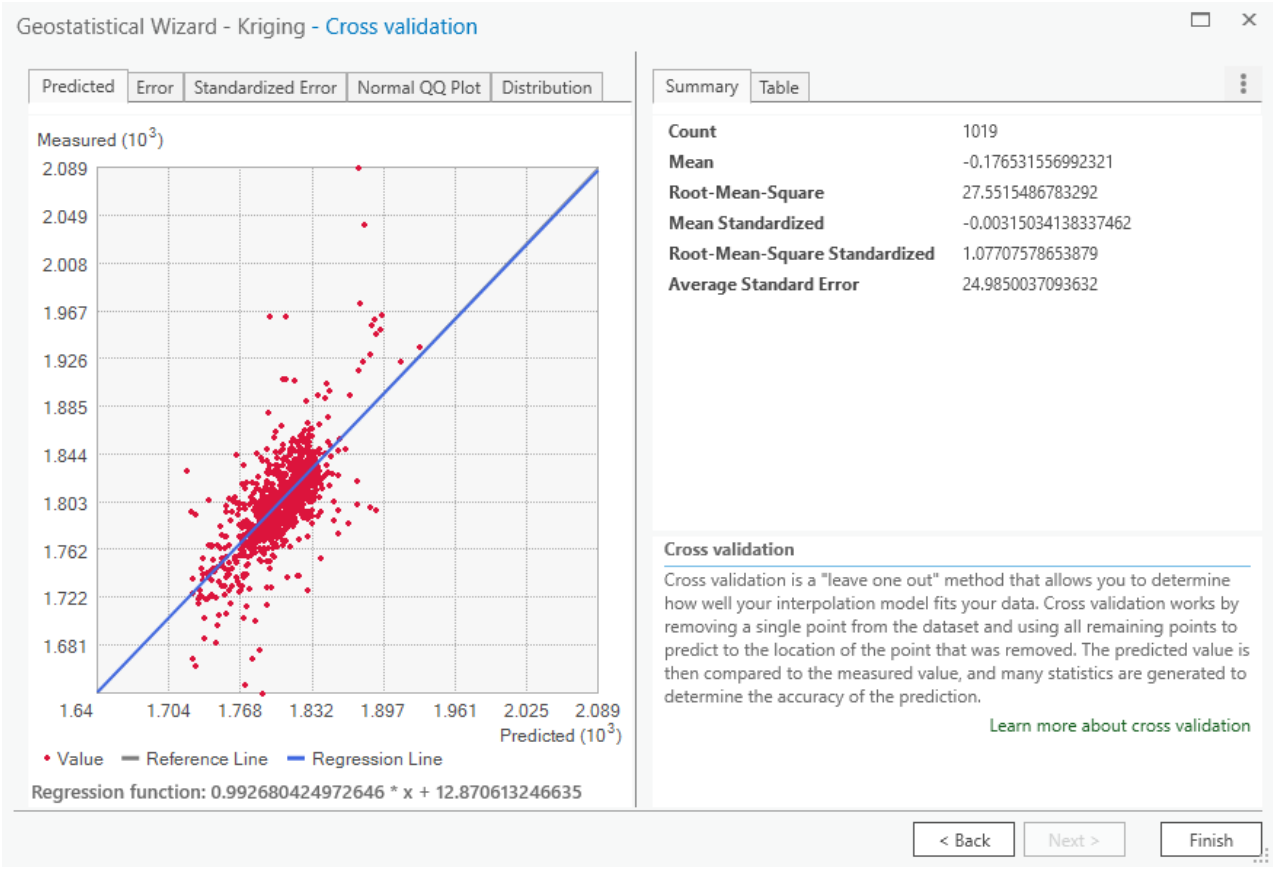


Figure 11. Kriging Cross Validation

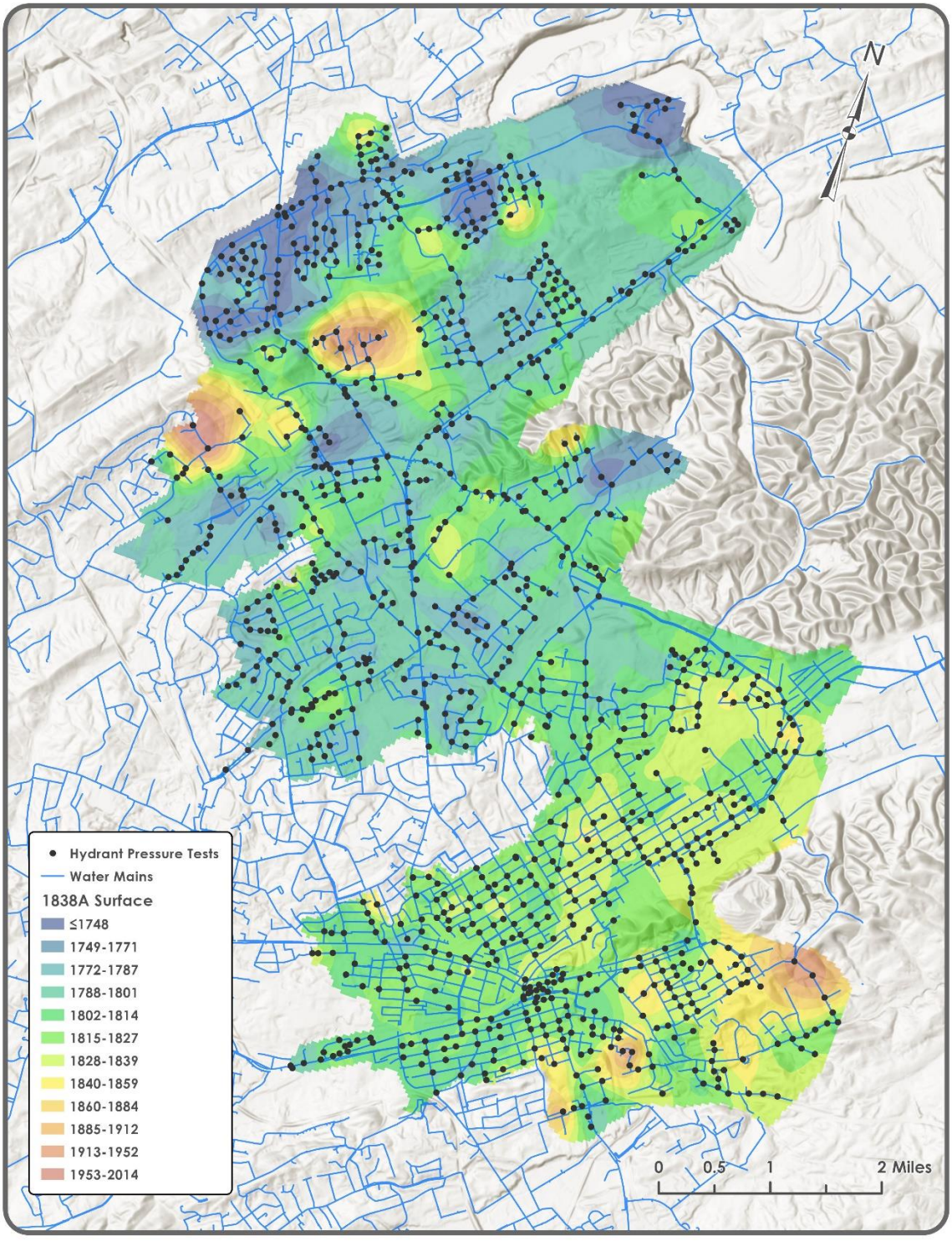


Figure 12. Hydrant Pressure Interpolation Reclassified to Show Hydraulic Grade Values

Model Python Script Development

The pressure test update processing is an ongoing program at the water department, so an automated script was developed to capture additional hydrant pressure tests daily. This places information indicating large areas of water leakage in front of decision-makers in a timely manner so that large water breaks can be identified and repaired. The script was written in Python 3.6 and developed using Jupyter Notebooks in conjunction with ArcGIS Pro to document each step and support replication in other water systems. Each cell within the Jupyter Notebook contains individual process code for steps within the script. The code can be referenced within Appendix A to follow along with the process steps. The flowchart in Figure 13 describes the general automated steps.

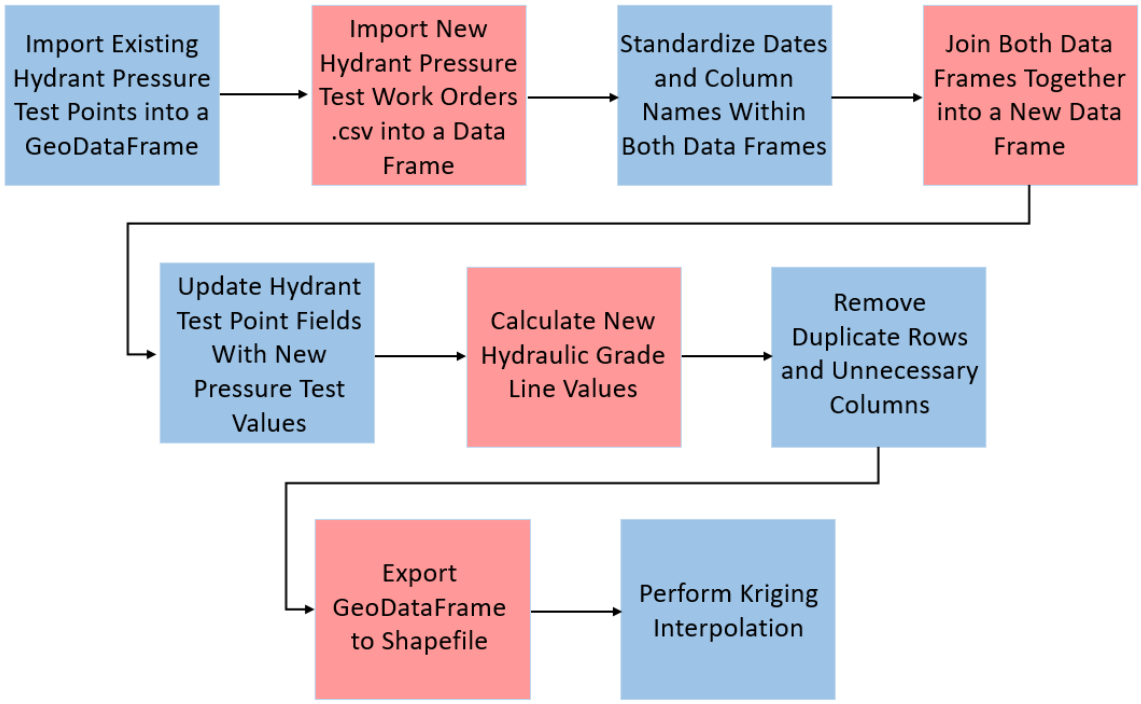


Figure 13. General Python Script Workflow

The resulting Jupyter Notebook required a custom Python virtual environment created to manage custom Python libraries, including pandas, geopandas, numpy, arcpy, matplotlib, shapely, and fiona. The script also utilized stock Python packages (e.g. os, csv, and datetime). The virtual environment was created with Anaconda Navigator, which managed custom library installations through the Anaconda command prompt and was also used to install the Jupyter Notebook application. The custom virtual environment also only referenced the Python version 3.6 included with ArcGIS Pro, so that any older versions of Python previously installed were not used. This installation step allowed portability when duplicated on additional computers and ensured separation between previously installed Python versions.

The script begins by importing all the necessary libraries including arcpy, pandas, geopandas, and others. The complete list of needed Python libraries is shown in Table 1. The arcpy, pandas, and geopandas libraries handle the bulk of the computation while datetime and fiona are used to access and write to file format types as well as standardize any dates within the data.

Table 1. Python Libraries Imported into the Automated Script

Package Name	Script Role Description
arcpy	Used to perform the kriging interpolation process
csv	Used to open and read pressure record csv files
datetime	Used to manipulate date information within the data
fiona	Used to read layers within a geodatabase
geopandas	Used to manipulate water hydrant data frames
matplotlib	Used to test plot resulting pressure test locations
numpy	Used by pandas for array computations
os	Used to name resulting shapefiles
pandas	Used to manipulate pressure test results data frames
shapely	Used by geopandas for geometry operations
sys	Used to name resulting files

The next few cells set up the project's workspace and variables used throughout the project. The code defines multiple variables containing file locations for a pressure test csv file, pressure test site file containing existing water hydrants, and a polygon file containing the DMA boundary. The water hydrants and DMA polygons are located within a geodatabase and serve as the project workspace. An output raster location and name variable are also created as well as variables for the location of a kriging model .xml file and name of the kriging geostatistical layer created later in the script. Each variable is called later in the script and used in multiple process cells. Fiona's listlayers function is then called to list all the layers within the project geodatabase containing the spatial information for the 1838A test hydrants and the DMA boundary. Geopandas converts the attribute table from a geodatabase feature class to a geodataframe using the read_file function. This imports the test hydrants layer as a data frame, using the fiona generated list to call the appropriate geodatabase layer by position in the layer list. The same read_file function then opens the hydrant pressure test csv file to import data into a second data frame. Both data frames contain a date column containing date information about when the pressure test was completed. The pandas to_datetime function runs on both columns to standardize the date format. The result is two data frames representing the physical hydrant locations with existing elevation information and existing pressure information, and another data frame containing updated pressure test results.

The two data frames must be manipulated so their columns match in name and data type to update new pressure information. A critical column in both data frames contains a unique identifier for each water hydrant and is stored in two fields called "FACILITYID" in the water hydrant data frame and "Asset" in the updated pressure test data frame. These columns record the same information and act as a common item to join the data frames later in the script. Without cleanup, the Asset column contains the needed hydrant identifier info, but contains a string before each id.

This is remedied by adding a new “FACILITYID” column to the pressure test data frame and populating it with the Asset field’s information minus the first characters in the string using a string slicing command. The resulting column contains water hydrant identifiers without the extra string characters and can now serve as the common column to merge the two data frames.

Additional functions performed on the data frame include removing spaces from any column names, changing any static pressure columns to a numeric data type, and removing any records containing static pressure information that is incorrect or zero values. Initially, geopandas imports the given field names from Cartegraph as column names. These field names contain spaces and can cause difficulty with some Python functionality, so the spaces are replaced with underscores using the Python replace command. Also, the static pressure values are imported as text and are converted to the numeric data type using the pandas to_numeric function. This permits numeric calculations later in the script when the data are converted to hydraulic grade. The zero values and values above 300 indicate errors within the pressure test data and can be removed from the data frame, since those readings are not typical of Johnson City’s water system.

After formatting the new pressure test data frame, the two data frames are joined using the pandas merge function and specifying the FACILITYID column as the common item or key field. As a result, a new pandas data frame is created, but still maintains the geometry column necessary for a later re-import into geopandas. The new data frame contains only records matching the data frame containing hydrants in the 1838A boundary. Within the next cell of code, new static pressure values are updated in the original StaticPressure column from the testSites data frame and new dates are updated in the DateCollected column. Any new static pressures and dates overwrite the original data to update the data frame with new hydrant pressure tests. The new hydraulic grade calculations are also performed and stored in the HydroGrade column.

Further manipulation of the new pandas data frame includes deleting any unnecessary columns using the del command after the new hydraulic grade values have been created. This includes an extra geometry column created by the pressure test data frame. The newly created geometry_x field, which contains the hydrant location information, is then converted back to the original geometry column name. After all the field names are set, the sort_values and drop_duplicates functions are used within pandas to remove duplicate values. The sort_values function is used to sort the records by the DataCollected field in ascending order. The drop_duplicates function is then called to remove records containing duplicate facility IDs previously sorted. The "keep" parameter is set to 'first', since the DateCollected is sorted as text and shows later dates before older dates when sorted. The resulting data frame contains updated pressure tests with newly calculated hydraulic grade values that do not contain duplicates. The next step within the process takes this new data frame and updates the values in the existing hydrant test sites.

The merged data frame containing water hydrants with newly acquired pressure test information and hydraulic grade calculations must check back in with the original hydrant data frame to update any new information. This is accomplished by first setting the FACILITYID field in both the merged data frame and the original hydrant data frame as the index using the pandas set_index function. After that, using the pandas update command on the hydrant data frame and specifying the merged data frame as the update source, new hydrant pressure test information and hydraulic grade calculations are applied to the water hydrants geodataframe. The indexes are reset using the reset_index command and the hydrant data frame can now be prepared for the kriging interpolation process.

Exporting the newly updated hydrant information is accomplished by first converting the updated hydrant data frame to a geodataframe using the geopandas GeoDataFrame function, specifying the geometry field as the column containing the locational information. Any hydrants containing null values in the HydroGrade column are dropped by exporting to a new geodataframe called “NewGdf” using the notnull() command on the HydroGrade column. Another .astype(str) function is performed on the DateCollected column to ensure dates are recorded as strings. This step permits export of the geodataframe to a shapefile without error later in the script. Another cell within the notebook is added containing code to plot the hydrant locations using the matplotlib plot function and project the geodataframe to the NAD83 Tennessee State Plane projection. Figure 14 illustrates the expected plot for the 1838A DMA.

After inspection of the plotted hydrant locations containing pressure information, the geopandas to_file function is called to convert the NewGdf geodataframe to an Esri shapefile. The DateCollected column data type is also changed to a string using the .astype command. The shapefile name is specified by a variable containing the file location and current date, then exported using the to_file command.

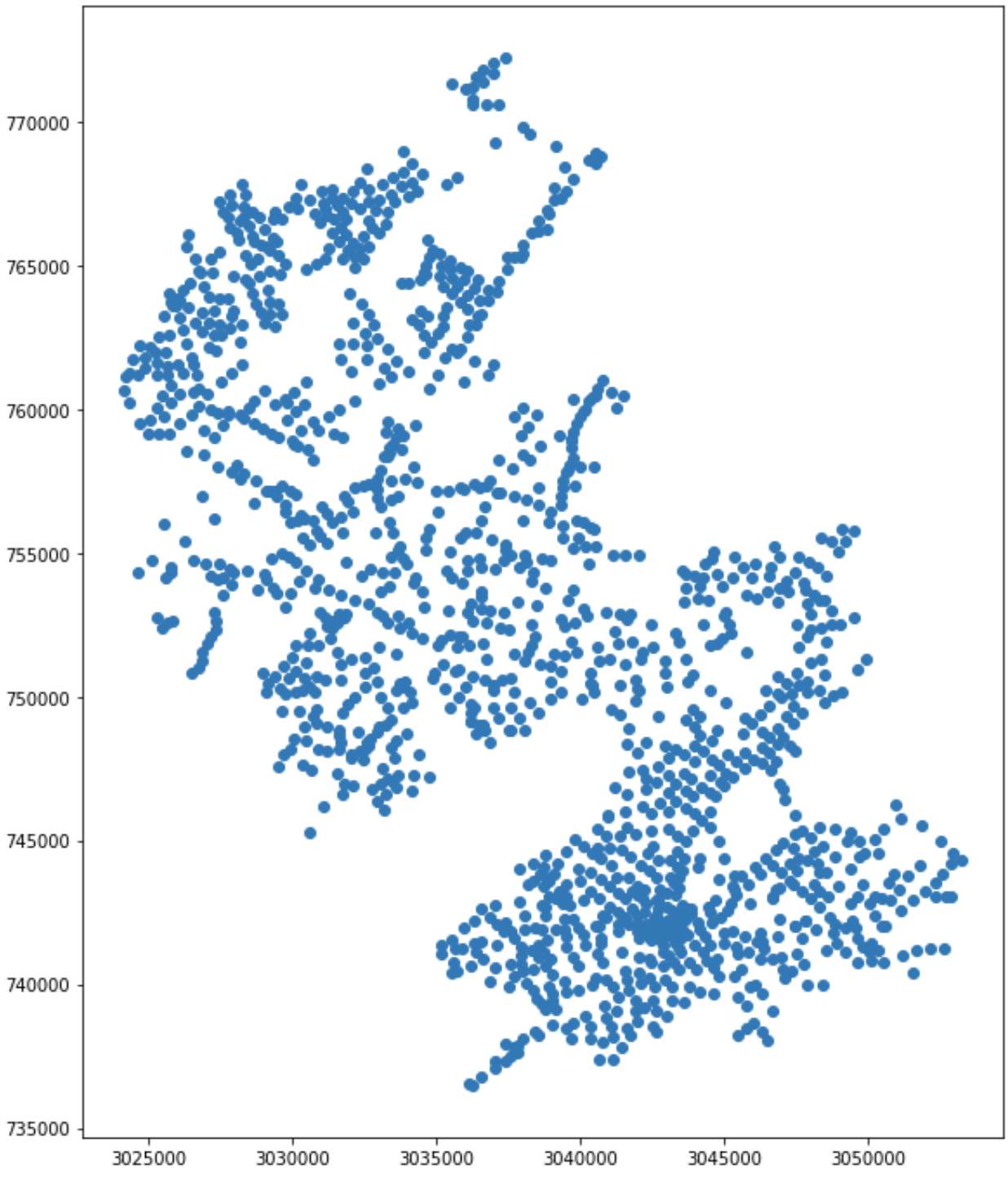


Figure 14. Plotted Hydrant Pressure Test Sites in NAD83 TN State Plane

The last portion of code within the script uses the arcpy library to create an interpolated surface from the updated pressure test sites. The tools used within the arcpy library are part of the Spatial Analyst and Geostatistical Analyst software extensions and require Esri licensing. Before using arcpy, the required Python virtual environment must be set up and have an active ArcGIS license. Once established, the arcpy.CheckOutExtension command is used to check out the needed Esri extensions. This allows users to use the extension tools with an appropriate license. The code then calls arcpy's Create Geostatistical Layer tool to build a kriging layer using three variables including the pressure test site shapefile, kriging model file, and output layer name. Since a kriging model has already been established for the 1838A DMA, the earlier kriging model xml information is used as a variable within Esri's Create Geostatistical Layer geoprocessing tool. The resulting layer is then clipped and exported as a clipped raster. The last cell of code within the notebook calls the arcpy.CheckInExtension to check the two Esri extension licenses back in. This is only important if licensing is shared with other users.

The completed code was developed in Jupyter Notebooks to document each step thoroughly for replication in other DMAs in the Johnson City water system and allows replication in other systems. The code is also stored as a script in a Python .py file which allows the process to run as an automation on a regular basis.

Field Testing Process

The kriging prediction surface indicated multiple areas predicting hydraulic grade values less than 1838 as well as areas exhibiting higher pressures than expected. Ten sites were selected from the resulting leak prediction surface for field investigation using either acoustic detection methods, GPR survey, or past work order repair research. After work order research, traditional acoustic leak surveys were performed by water department personnel to verify water leakage and

identify areas to repair water mains. The chosen locations all contained metallic water mains, making acoustic leak surveys the appropriate choice of inspection technique. To ensure data validity, area hydrants were re-tested to ensure existing low pressures before each survey.

Although most field verification was performed by traditional acoustic leak detection methods, they are not always able to detect water leakage in non-metallic pipe. To help discover leakage, ground-penetrating radar (GPR) survey techniques were also selected to identify water leakage. Initially, GPR was chosen as a supplementary survey method to detect leakage where acoustic techniques were ineffective due to plastic water mains or areas with high levels of noise. Since all areas selected for field detection surveys contained metallic pipe, GPR was only used at one site to verify the ability to detect leaking water.

CHAPTER 3

RESULTS

Pressure Test Results

Using 1,019 hydrant pressure tests within the 1838A DMA, 87 percent fell below the expected 1838 hydraulic grade value. The mean hydraulic grade value was 1804.63 and the median value was 1805. The minimum hydraulic grade value was 1640 and the maximum value was 2089. Figure 15 displays hydraulic grade values with a centerline drawn on the mean value. The data conforms to a normal distribution as shown in Figure 15 and does not follow any global spatial trends. Additionally, a test for spatial autocorrelation resulted in a 0.4105 Moran's I value along with a z-score of 29.9672 and p-score of 0.0000. These numbers indicate that there is less than one percent likelihood that the data are random and suggests spatial autocorrelation.

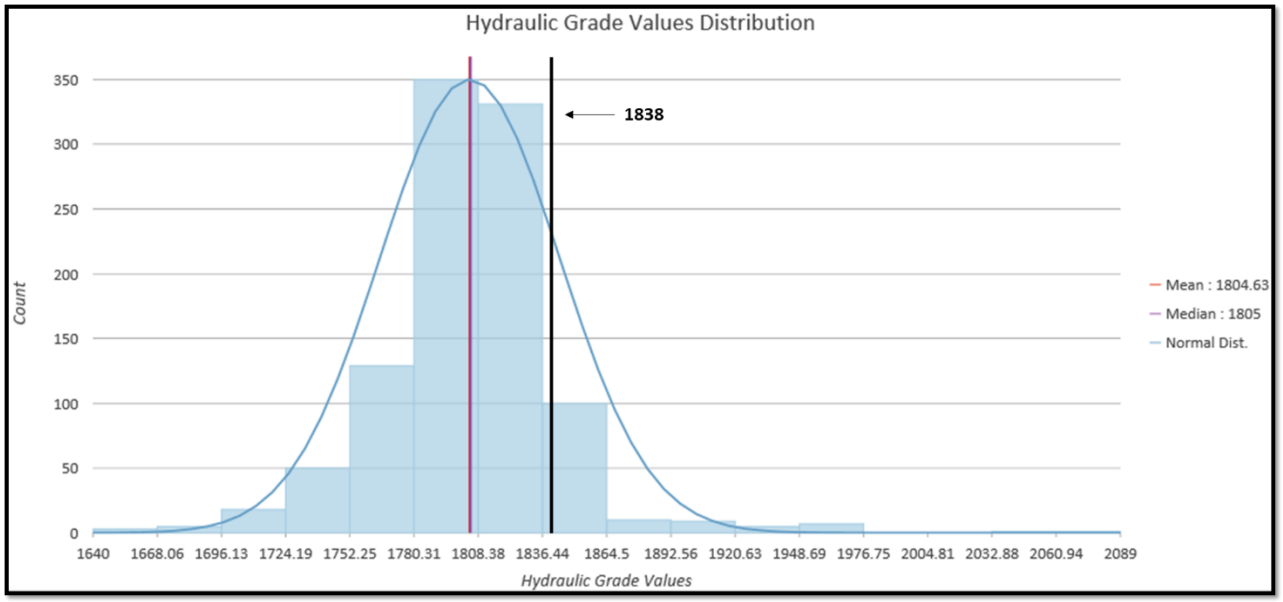


Figure 15. Histogram of Hydraulic Grade Values

Figure 16 displays the sample hydrant pressure values distributed throughout 1838A and illustrate a concentration of hydraulic grade values closest to 1838 within the DMA's southern

area. These southern pressure test areas also contain multiple values spiking above 1838, which make up 13 percent of all hydrant pressure tests. The lowest hydraulic grade values ranging from 1640 through 1774 are predominantly located within the northern portion of 1838A and consist of values well below 1838.

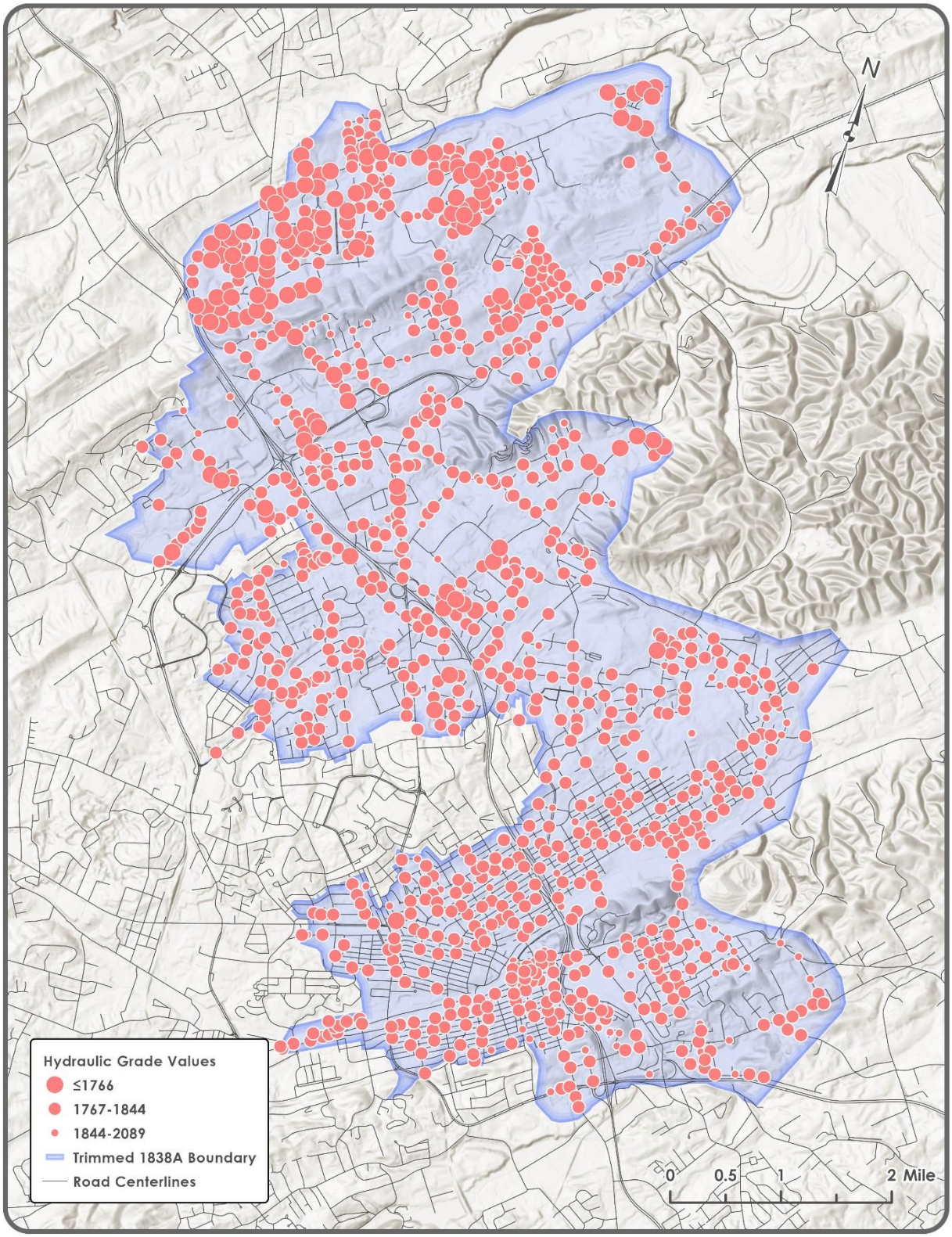


Figure 16. Pressure Test Hydraulic Grade Value Distribution

Kriging Model Results

The resulting kriging interpolation model used Esri's Geostatistical Wizard and reports statistically necessary values to indicate the model's quality of fit within a cross-validation data summary. These include three values used to judge the validity of the model and include the mean error (0.1765), mean standardized error (-0.0031), and root mean square standardized error(1.0771). The mean error and the mean standardized error are both acceptably close to zero. The root mean square standardized error is close to 1 and ideal. These values are illustrated in Figure 11 within the previous methods chapter.

After running the model with 1,019 hydrant pressure test sites, the interpolated surface results are illustrated in Figure 17 and display predicted hydraulic grade values within the 1838A DMA. The surface was reclassified into twelve classes using natural breaks and symbolized. Six classes illustrate hydraulic grade values lower than 1838, leaving another six to display values equal to and above 1838. Additionally, an error surface was created to illustrate areas with insufficient data to predict leakage using the model. The error suface is displayed in Figure 18.

Verification Findings

Illustrated in Figure 17, the resulting raster surface shows multiple areas predicting hydraulic grade values less than 1838, especially within the northern portion of the 1838A DMA. Figure 17 also shows multiple areas exhibiting high pressure. A selection of ten high and low hydraulic grade value areas was investigated during the project and are referenced as their street name location or the associated subdivision name if located in residential areas. Each investigated area is labeled in Figure 17 and described in Table 2. Of the ten areas, three sites predicted large water main breaks, one area resulted in a hydrant leak, three sites were affected by either a pump

or PRV, and another area discovered a water lateral leak. The remaining two sites were inconclusive and need further investigation.

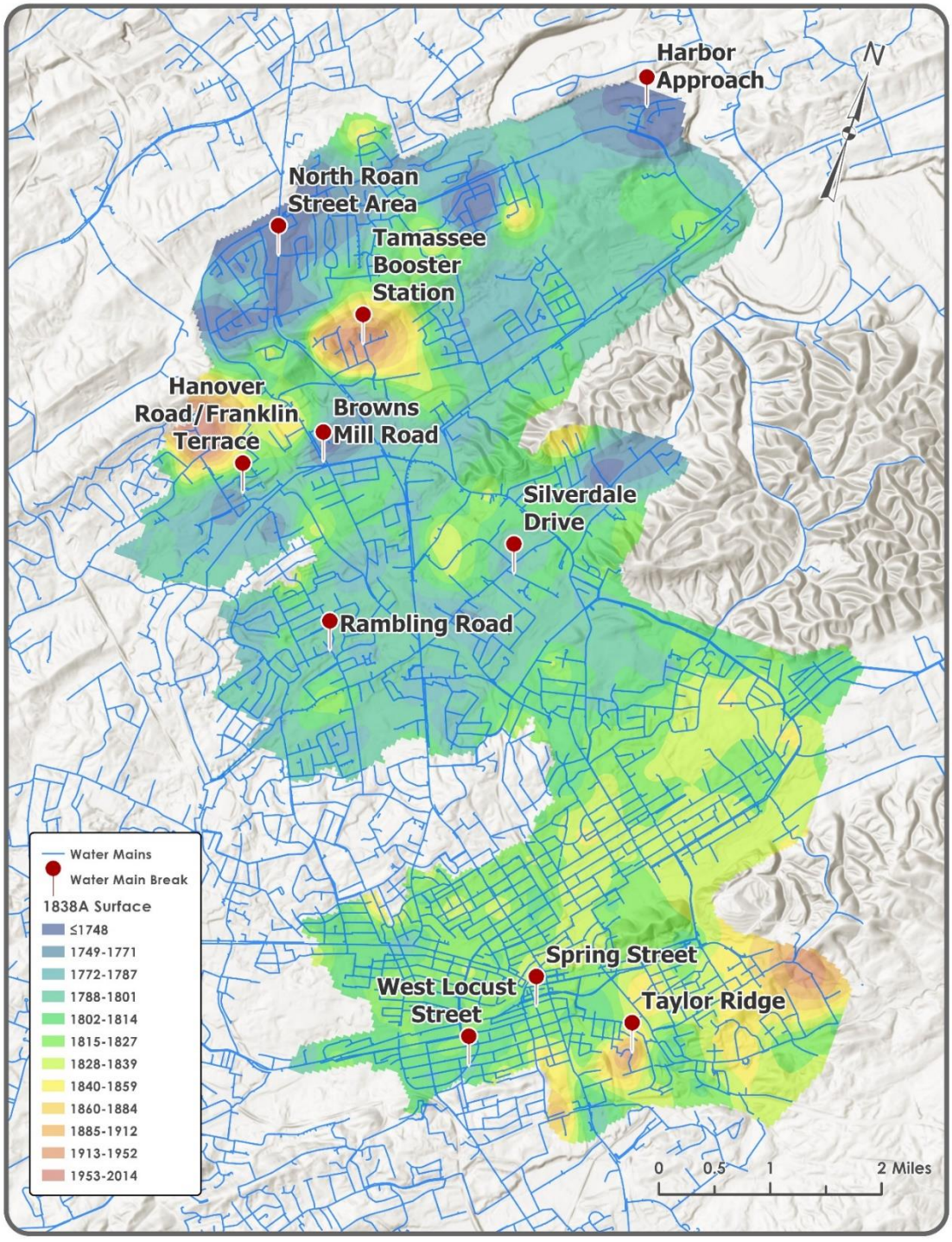


Figure 17. Hydrant Pressure Field Investigation Areas

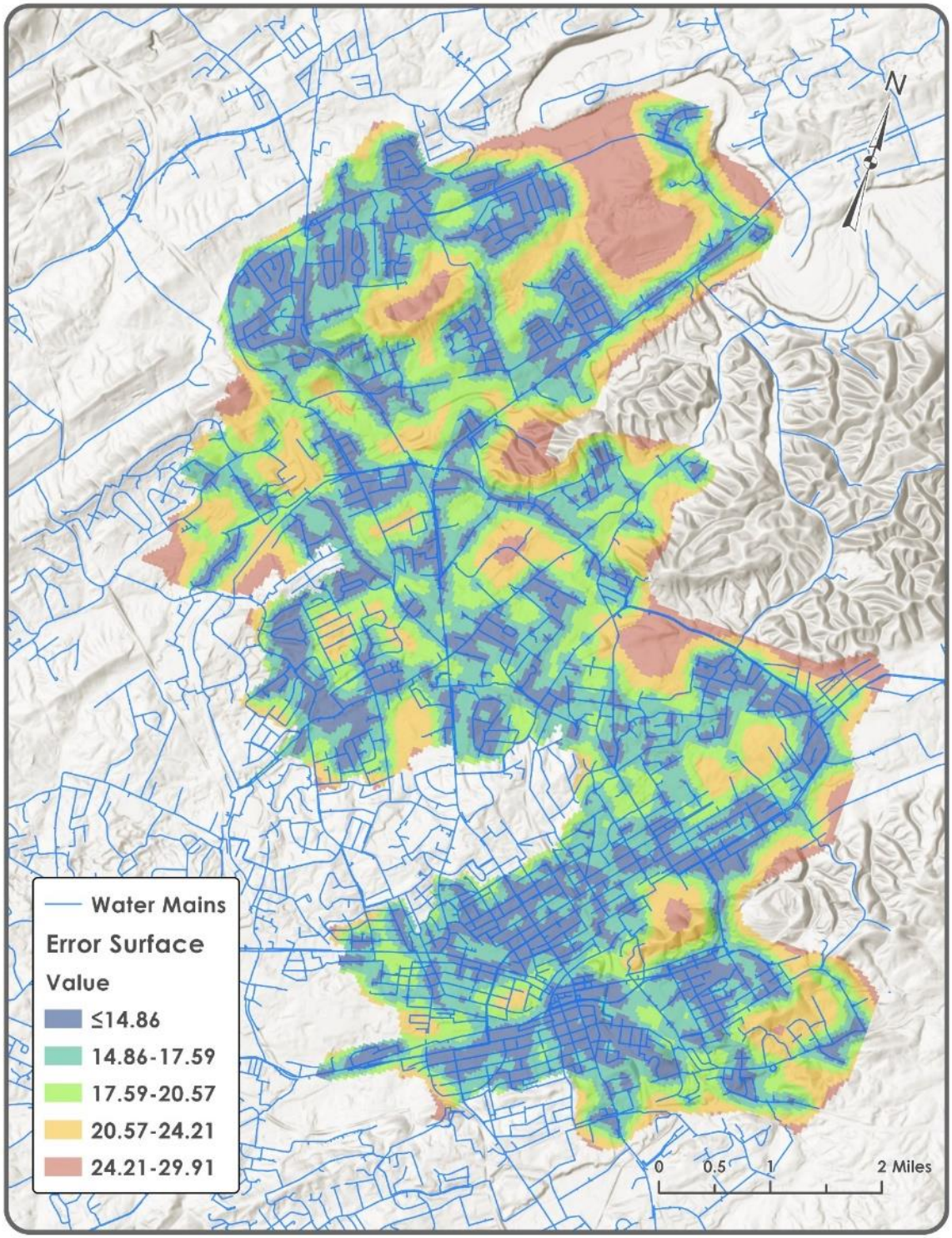


Figure 18. Kriging Error Surface

Table 2. Field Investigation Areas and Results

Site	Hydraulic Grade Values	Result
Harbor Approach	1677 - 1795	Pressures influenced by a PRV
Spring Street	1803 - 1838	Water lateral leak
North Roan Street	1683 - 1801	Investigation ongoing
Browns Mill Road	1709 - 1806	Large water main break
Rambling Road	1773	Water hydrant leak
Silverdale Drive	1747 - 1798	Large water main break
West Locust Street	1784 - 1789	Inconclusive
Tamassee area	1898 - 1960	Pressures influenced by a pump
Hanover Road	1730 - 1782	Large water main break
Taylor Ridge	1916 - 1955	Pressures influenced by a pump

The Silverdale Drive area reported multiple hydraulic grade values ranging from 1747 to 1798 and is illustrated in Figure 19. After further investigation by the city's leak detection personnel using acoustic leak detection technology, leakage was verified and repaired within the zone identified by the interpolation process. Additional low pressures detected on Hanover Road resulted in a 12-inch waterline break and is illustrated in Figure 20. This leakage area was first thought to be the result of suction pressure created by a water pump at the edge of the 1838A zone but was later found to be a broken pipe when leak detection field crews investigated the area. Low pressures were detected on Hanover Road, but the leaking water main was located on Franklin Terrace. Both areas contain cast iron water mains, making acoustic leak detection possible.

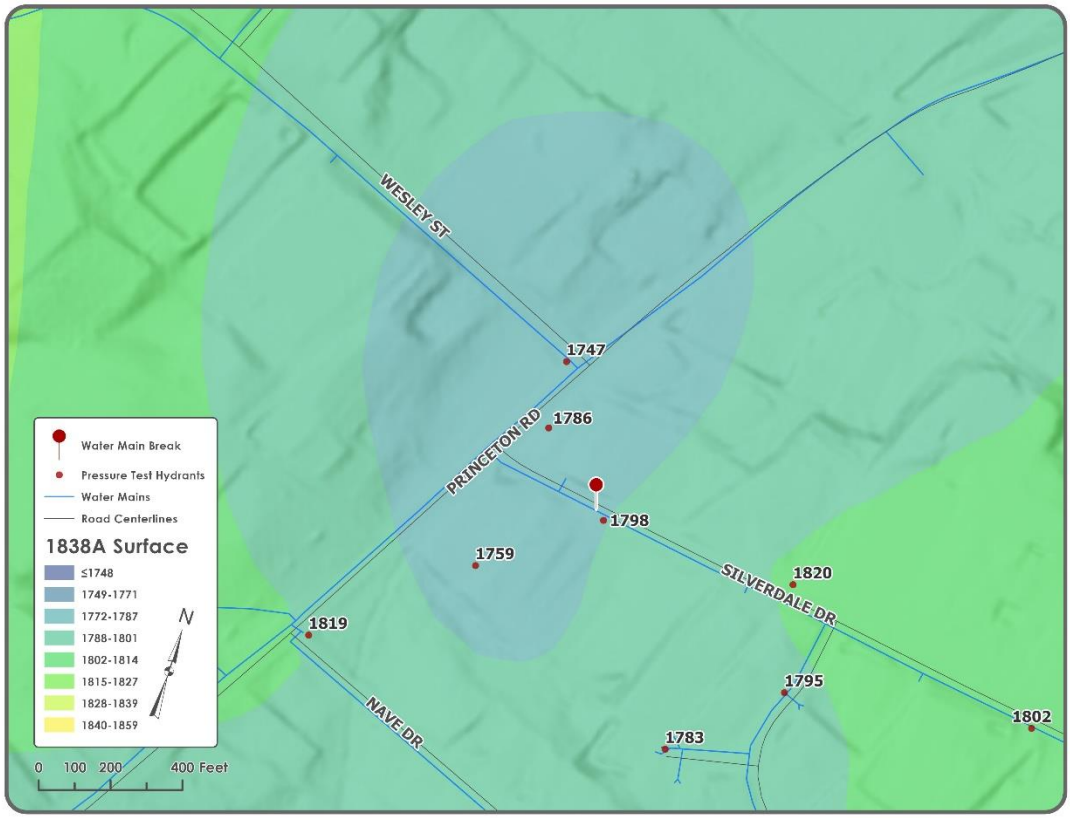


Figure 19. Silverdale Drive Water Leakage Area

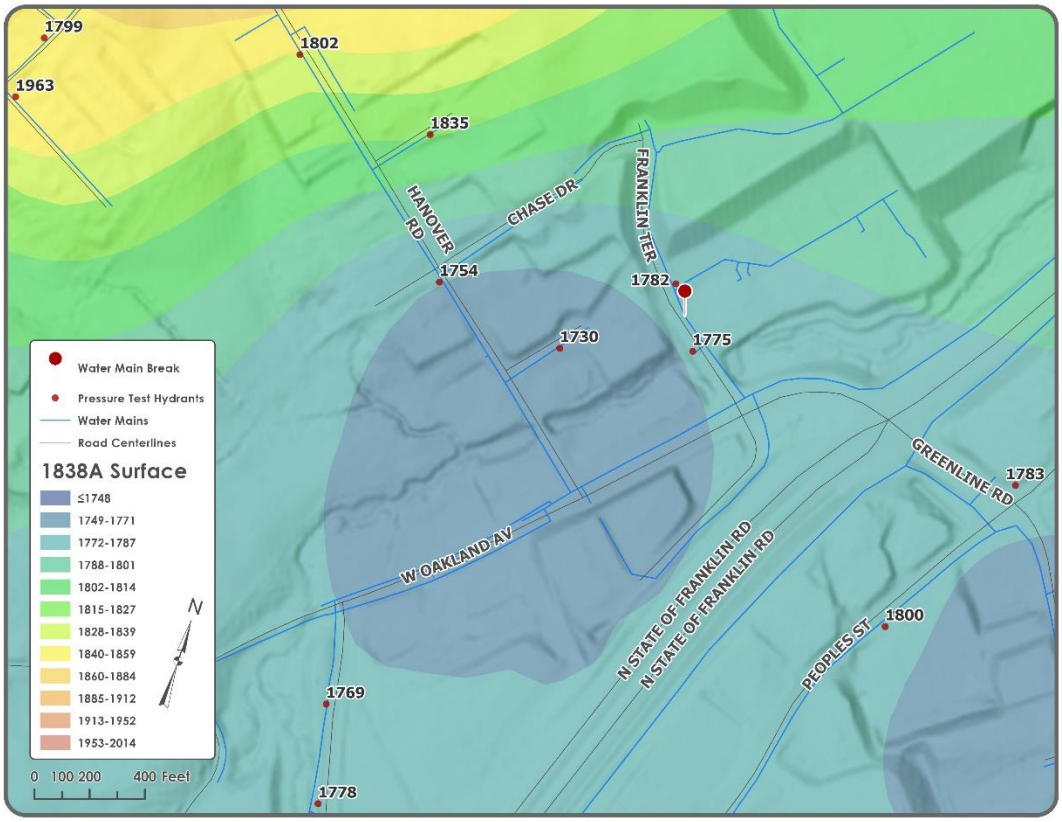


Figure 20. Hanover Road Area Water Leakage

An additional area indicating possible water leakage included the Browns Mill Road Area and produced low pressures as early as 2015, displaying hydrants with hydraulic grade values ranging from 1709 to 1806. The intersection of Browns Mill Road and North State of Franklin Road contains multiple large diameter water mains supplying water to many large businesses in the area. In September 2018 and during this study, a 10-inch ductile iron water main break running through the intersection resulted in customer water outages for half of a day. Once the broken pipe was excavated and repaired, multiple holes indicated pipe leakage for an extended time. Although the kriging surface illustrates low hydraulic pressure as early as 2015, water did not surface at that time. The hydraulic grade calculations produced from hydrant pressure readings and the resulting kriging interpolation model is displayed in Figure 21.

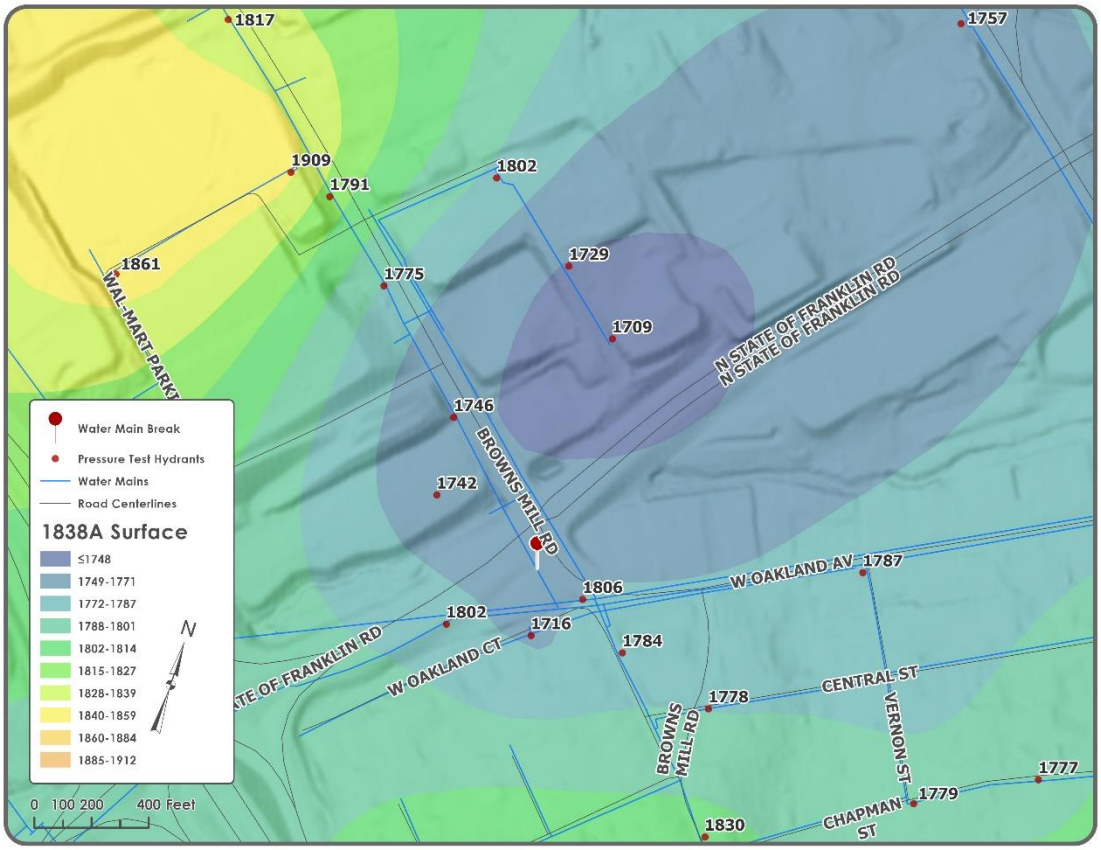


Figure 21. Browns Mill Road Intersection Water Leakage Area

Another leakage area was identified around the Sundale/Town Acres portion at the intersection of Rambling Road and Sundale Circle. Displayed in Figure 22, this leak was next to a water hydrant and identified by pressure information collected in 2015. After referencing past work order repair records and consulting with Johnson City Water personnel, the repair was determined to be made before this study.

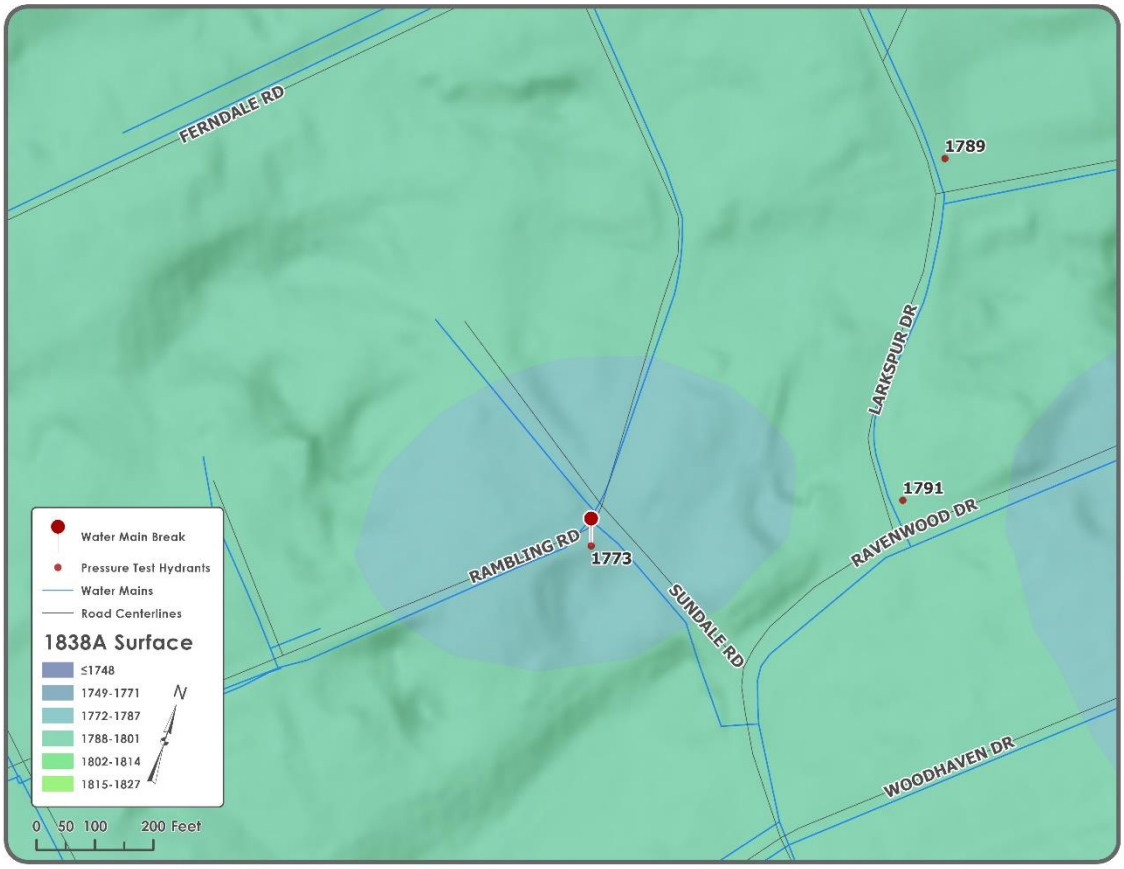


Figure 22. Rambling Road Water Hydrant Leak

Other low hydraulic pressure areas include the Harbor Approach area shown in Figure 23, the North Roan Street Area, and West Locust Street. Harbor Approach's low-pressure values were determined to be the result of a pressure-reducing valve located south of that neighborhood. The North Roan Street and West Locust Street areas are stilling under investigation. Both areas exhibited low hydrant pressure tests, which are currently scheduled to be re-tested.

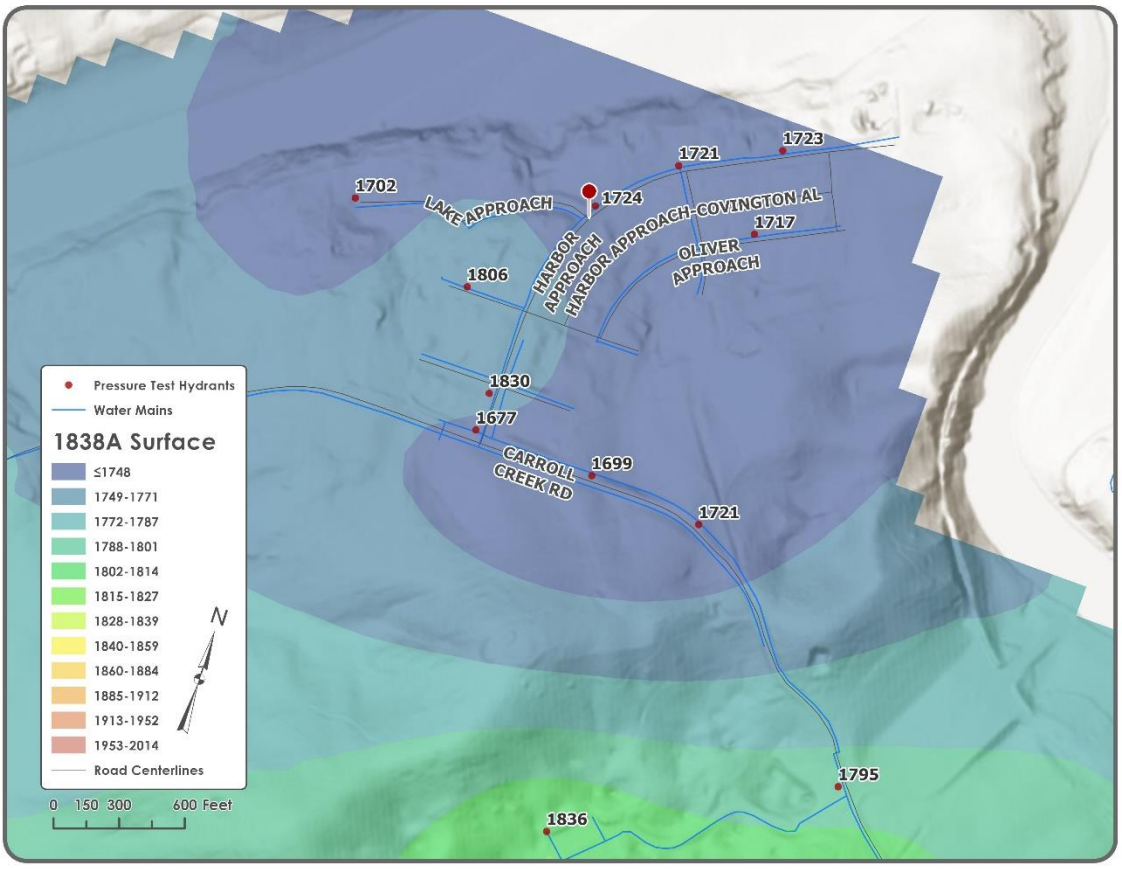


Figure 23. Harbor Approach High Pressure Area

The Tamassee, Hanover Road, and Taylor Ridge areas all displayed hydraulic grade values well above 1838. In Figure 24, the map containing the Tamassee area included values as high as 1960. Both Hanover Road and Taylor Ridge displayed similar results with values above 1900. These three areas are all located at higher elevations than surrounding neighborhoods and in near water pumping stations.

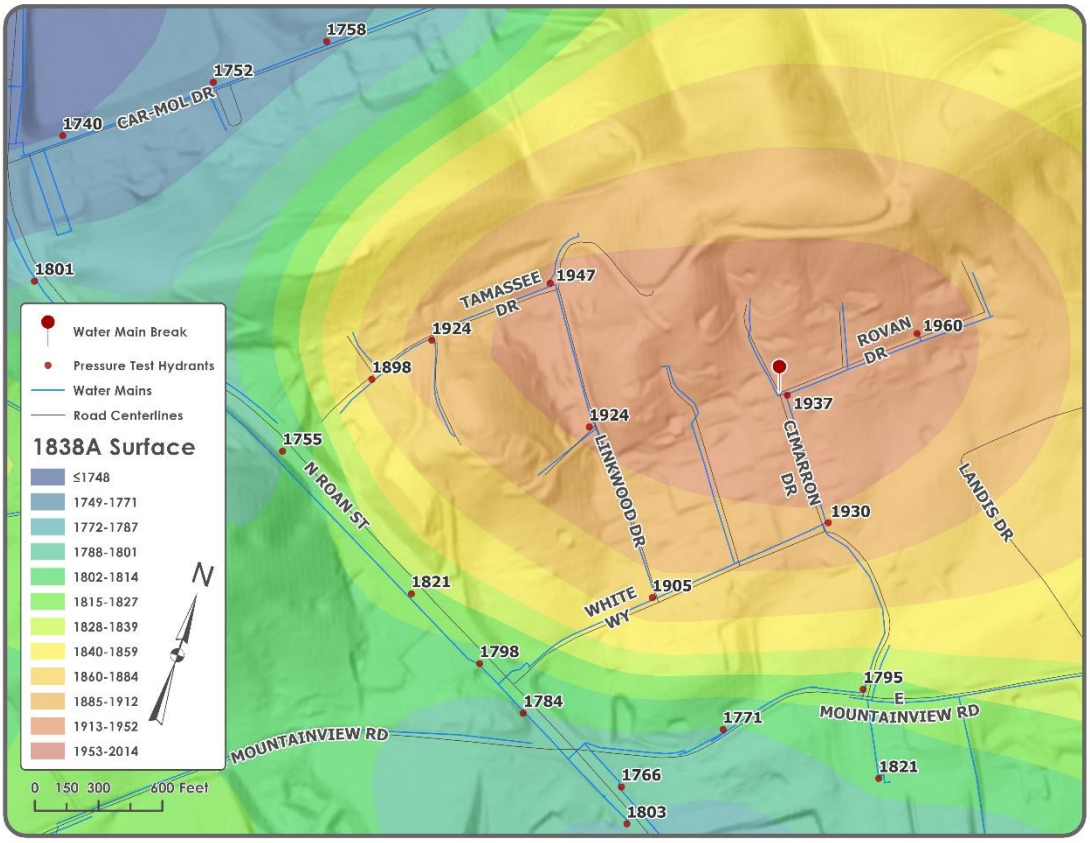


Figure 24. Tamassee High Pressure Area

All selected investigations in the field were conducted using traditional leak detection equipment. One individual location at Spring Street did not display prominently within the interpolated surface but did produce values lower that 1838 and was requested by the water department as a candidate for GPR inspection. The resulting images in Figure 25 display evidence of water leakage and were later found to be a leaking water lateral. A six-inch diameter cast iron water main was detected with GPR along with patch material from a previous repair.

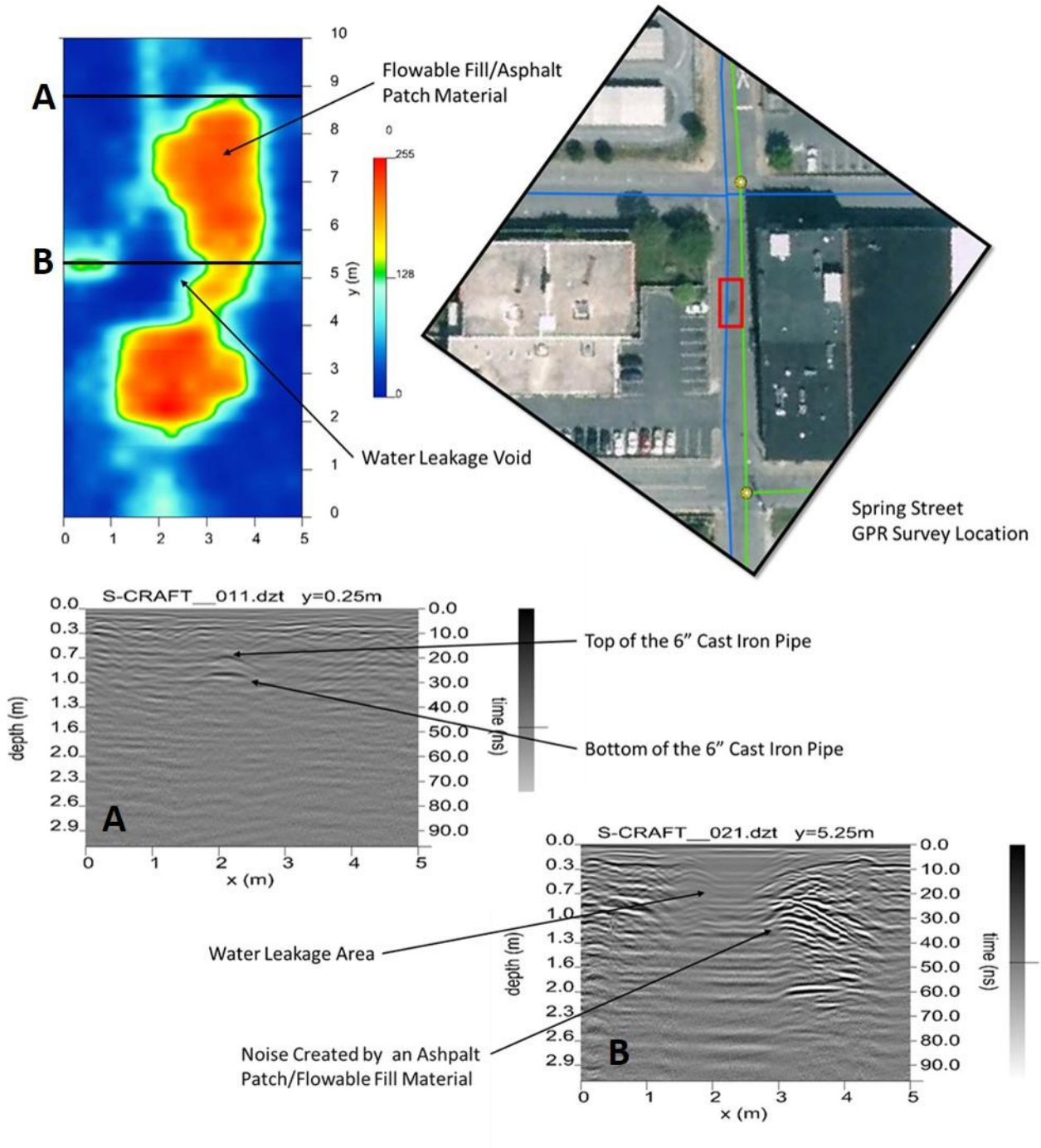


Figure 25. GPR Sample Survey Used to Identify Water Leakage around Spring Street

Process Automation

To update the leakage prediction surface results quickly, the data management and analysis process was automated using the Python programming language. Each day, an exported csv was obtained from the water departments work order system which documented any new hydrant pressure tests. These new tests were imported into the existing data and used to update the existing surface. The resulting Python script runs daily and is illustrated as a Jupyter Notebook within Appendix A.

CHAPTER 4

DISCUSSION

The water leakage project's key components included collecting pressure data within the water distribution system, converting pressure results into hydraulic grade values, creating an interpolated surface based on hydraulic grade values, identifying potential leaks based on that surface, and testing them. The results indicate successes in water main leak detection and provided key indicators describing water system pressure conditions using a kriging interpolation model to predict changes in hydraulic gradient. The pressure conditions discovered included high pressures created by water pump stations as well as low pressure areas created by pressure reducing valves. Additionally, the bulk of data manipulation and analysis was automated using a script developed with the Python programming language.

Data Collection Process

The importance of accurate and varied field data is critical to the success of water distribution system optimization efforts including leakage detection. The field pressure test data for this project was initially taken from fire hydrant flows conducted between 2015 and 2017 as performed by Johnson City's fire department. Although providing sufficiently distributed coverage across the 1838A DMA, the data's age did not reflect current and changing conditions within the water distribution system. To capture updated information, hydrant pressure test readings collected by water department personnel provided new data as it was recorded using a geospatially enabled work order management system. These data collection efforts were already part of the water department's current water quality and leak detection efforts, so no additional personnel were needed. Access to the work order system allowed incorporation of new hydrant pressure test data in a timely fashion by incorporating exported work orders into updated prediction surfaces.

Kriging Model Development

When developing the ordinary kriging prediction surface model, the cross-validation report generated by the Geostatistical Wizard toolset supported the selection of ordinary kriging model parameters. The reported values for the mean error, mean standardized error, and square standardized error were all included and helped develop an appropriate kriging model. For the 1838A DMA hydrant test points, both the mean error and mean standardized error were close to zero, which supported the model fit. Also, the root mean square standardized error was close to 1 and the best kriging model cross-validation test indicator, meaning that the model was an appropriate fit (Oliver and Webster 2014).

The resulting model parameters were then incorporated into the process automation script as a parameter within the arcpy geoprocessing tool. To replicate these results within another DMA in Johnson City or other water system, a separate ordinary kriging model must be developed, checking the respective cross-validation results for an appropriate fit. Also, initial data points should be checked for normal distribution, the absence of any overarching global trends, and spatial autocorrelation (Sheeres 2016). If the data points do not reflect these conditions, steps should be taken to normalize the data or remove any trends. Otherwise, a different interpolation method should be investigated.

Hydraulic Grade Surface Predictions and Investigations

After developing the hydraulic grade surface model, multiple variations in hydraulic gradient were identified throughout 1838A that suggested high water demand and possible leakage. Any predicted area with a hydraulic grade value lower than the expected supply tank's top elevation suggests high demand and possible water leakage (Walski 1983). The 1838A hydraulic grade surface predicted hydraulic gradient values in those areas ranging from 1748 to 2014. The

1838A DMA's expected hydraulic gradient was 1838. Multiple hydrants within the DMA provided hydraulic grade values well below 1838 and was reflected in the prediction surface. After running the model multiple times, water department personnel elected to field verify results in a few selected areas that were reported previously as results.

Using a combination of field investigations and research of previous work orders, four areas investigated for low hydraulic grade values were false positives. All four had documented water leakage or water main breaks that had since been repaired, but were manifest in the hydraulic grade surface from outdated pressure test data. Browns Mill Road, Silverdale Drive, and Hanover Road areas contained water main breaks with hydraulic gradient values well below 1838 since 2015. These areas experienced a period of water leakage leading up to a water main break event and subsequent repair. The Rambling Road water leakage area was the result of a broken hydrant lateral. Past records within the work order management system identified the previous repair and leak history from 2015, the same year as the pressure test. After identifying these four sites, water department personnel re-tested the hydrants and reported hydraulic gradient values above 1800. The new values were used to regenerate the hydraulic grade surface.

Additional areas resulting in spikes in hydraulic gradient were found to be caused by pumps supplying water to areas at high elevations within the DMA. These pumps add energy in the form of increased hydraulic grade to overcome pressure losses experienced with physical changes in elevation (Walski et al. 2003). This was evident in the Tamassee and Taylor Ridge areas, which are within the 1838A DMA. Currently, twelve water booster stations are located within the 1838A DMA and are used to maintain water pressure to homes and businesses. Pumped areas with spikes in hydraulic gradient need further investigation, since excess and frequent variations in pressure are directly related to new leaks and water main breaks (Lambert 2000).

Another two areas investigated for water leakage after reporting low pressures returned inconclusive results. They are the North Roan Street area and West Locust Street and these are still under investigation. Leak detection crews re-tested hydrants and surveyed the area around West Locust Street, finding no leaks capable of affecting pressure. The North Roan Street area lies at the boundary of another DMA and may be influenced by pressure reduction due to an open isolation valve, which can be used to block water and control pressure zone boundaries (Walski et al. 2003).

Field Leak Detection

The field leak detection efforts were used to validate drops in hydraulic gradient in the previously discussed areas plotted on top of in the interpolated surface in Figure 3.3. By using a pressure-dependent leak detection method, leakage hotspots were able to direct field investigations on focused areas instead of random or regular sounding surveys. This method attempted to discover unreported leaks and breaks, which is one of three categories of water leakage according to the Bursts and Background Estimates (BABE) philosophy to approaching water system leakage. The other two categories of BABE include reported leaks or breaks, and background losses with flow rates too low to be detected by traditional means (Samir et al. 2017).

Traditional acoustic noise monitoring is still the primary field leak detection method used by Johnson City's water system personnel. This method detects noise propagated along pipes and the ground that is generated as water passes through a hole or fractures in pipes as well as passing through substances outside pipes (Xu et al. 2014). Each selected low-pressure area generated by the interpolated hydraulic gradient surface was inspected using acoustic leak detection survey methods. GPR was also used on the Spring Street water leak to validate results. Although both acoustic and GPR detected water leakage, GPR created a more detailed view of conditions below the surface as illustrated in Figure 3.10. The Spring Street water leak was a service line connecting

to an older six-inch cast iron water main that was easily detectable by both methods. GPR would likely prove itself more valuable for leak detection in plastic water mains, since there is evidence that acoustic techniques struggle to detect leakage in plastic water mains where leaks typically create less sound in plastic material and lower pressure conditions (Wu et al. 2010). Additional study is reccommended to verify these claims.

Hidden Benefits

The hydraulic grade surface illustrates realistic conditions in the field including pumping pressure effects and conditions created by PRVs. The northern region of 1838A contains multiple areas exhibiting these scenarios. Multiple pumps within the DMA increase pressures to reach residential areas in higher elevations, but also increase pressures to levels above 1838. These areas can be seen in Figure 3.3. Low predicted values include the Harbor Approach residential area in the northeastern portion of 1838A where hydraulic gradient values are below 1800 due to a PRV put in place to reduce damage to plumbing systems within private homes. These areas where pressure is regulated to control leakage are downstream of a PRV and known as DPAs, or discrete pressure areas (Sage 2014). The hydraulic grade surface clearly detected the Harbor Approach DPA and can be viewed in Figure 3.8. This may also be the case in the North Roan Street area, but further investigation is needed. Continuous updates to the surface as new pressure test data are collected will detect these pressure anomalies and help future decision making in regard to water system optimization.

Process Automation

In addition to developing the prediction surface and investigating possible leakage areas, a large portion of the project was devoted to developing an automated script to quickly update the prediction surface as new hydrant pressure test results became available. The pandas and

geopandas libraries used within the resulting Python script handled a large amount of data management and analysis. Developed in conjunction with Jupyter Notebooks, the pandas and geopandas tools calculated hydraulic gradient values, updated timestamp information, and joined hydrant test work orders with existing GIS data. This was possible due to geopandas' inherent GeoSeries object, which preserved each hydrant's spatial location within a GeoDataFrame as a spatial attribute in the geometry column (GeoPandas 2019).

Process Improvements

Improvements to the pressure test process would include pressure tests on additional asset types such as meters and water blowoff valves. The ability to collect more pressure tests in a shorter time period would also help detect water main leaks and breaks as they occur. The Johnson City Water/Sewer department has indicated that it will invest in a future advanced metering infrastructure (AMI) system, which would greatly improve pressure testing capabilities by recording pressures at water meters and transmitting data across cellular networks. Instead of only gathering hydrant pressure information when staff members are working in the field, water meters will have the ability to send real time pressure information. This would exponentially increase the data to be processed daily and may require running the process on a more powerful computer. Johnson City's current water hydrant inventory contains 3,700 hydrants, which limits the number of pressure test sites. Implementing AMI would add roughly 43,000 water meters to the process, thus drastically increasing the ability to detect leakage. Being able to gather meter pressures quickly would increase water leakage detection times and decrease repair times.

Another improvement involves imported csv files converted to pandas data frames obtained from the work order system, which captured daily hydrant pressure tests and served to update existing test point hydraulic gradient values. This process allowed daily updates to the prediction

surface, but still required manual file updates to the script. Future utilization of the work order system's application programming interface (API) would eliminate the need to import a csv into the script, thus enabling the Python program to run as a scheduled task. The data frame information would be retrieved by an API request instead, thus eliminating the need to export a csv file out of Cartegraph and further automating the process.

Lastly, further development of a GIS-based hydraulic model would allow the water department to improve DMA management. For example, the system could be monitored for errors such as accidentally closed valves (when valves were temporarily closed for repair and not reopened after). In addition, the lower average hydraulic grade values experienced within the northern areas of 1838A and higher values in the south suggest that the DMA should be split into two or more zones. Currently, 1838A contains over 14,000 customers, which is well over the suggested optimal range of 350 to 5,000 water users (Scibetta et al. 2013). A well-developed hydraulic model would allow users to simulate multiple real-world water demand scenarios and better plan DMA boundaries (Samir et al. 2017).

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

Conclusion

This project used water pressure is an indicator of water leakage as described in other research. Not only will hydraulic gradient checks throughout a DMA identify water leakage areas, pressure checks conducted at water hydrant locations translated into an interpolated surface produce an effective tool to predict water leakage. The chosen kriging interpolation method not only creates hotspots indicating water loss, but also provides an indication of real-world pressure conditions and overall water system health. The pressure testing and kriging interpolation process easily identifies areas experiencing high pressures created by water pumps boosting pressure and low-pressure areas established by pressure reducing valves.

Pressure test information is expensive when collected by pressure monitoring equipment and is limited by the number of purchased monitors throughout a system. In populated areas, water hydrants are already distributed throughout a system as a function of public safety and provide multiple test sites that are ideal in developing an interpolated prediction surface. In addition to regular hydrant maintenance and water quality testing by water distribution professionals, fire department personnel regularly inspect and flow-test hydrants. These maintenance activities by multiple departments create the possibility of multiple pressure test data collection efforts used to detect water leakage. Combined with the previously described kriging model and easily configured Python automation, existing maintenance activity data are employed to detect water leaks.

Leak detection through interpolated pressure testing helps to reduce costs by discovering large water leaks that create enough demand to lower hydraulic gradient values throughout a water system. This is especially true for large water leaks that do not surface, run for long periods of

time, and usually cost large amounts of money. Using the previously described methods within this project and within DMAs helps to isolate systems into zones and prioritize leak repair activities. In Johnson City, multiple leaks were discovered using these methods in the 1838A DMA.

Recommendations

Improvements to the current process include collecting pressure tests at additional water infrastructure assets such as water meters. Water meter pressure monitoring would greatly increase the number of pressure test locations. Also, increased pressure test frequency would provide water leakage indicators in a timely manner. This would increase chances of detecting large water leaks as they occur. When scrutinizing process speed and data management, the incorporation of an API within the automated Python script would allow pressure test updates to be scheduled at regular intervals and not rely on manual work order system exported files. Water pressure test information could be directly fed into a pandas data frame through the API instead of reading manually created csv files.

When investigating leakage within DMAs, the optimal number of water customers is between 350 and 5,000 (Scibetta et al. 2013). The 1838A DMA contains over 14,000 water customers and should be broken into three or more DMAs. Additional DMA boundaries could be developed through incorporating a hydraulic model using existing water system GIS information. Hydraulic models help to investigate existing and planned water system conditions by modeling possible scenarios in valving, pumping, and pressure reduction (Walski et al. 2003). Adoption of a hydraulic model would also help to investigate the observed average 1805 hydraulic gradient values within 1838A.

Further use of alternative field leak detection survey methods to the traditional acoustic techniques is also recommended. Johnson City's water system contains many miles of plastic

water mains that create difficult leak detection conditions for acoustic location devices. The tested GPR method proved capable in detecting water leakage in cast iron water mains as well as supplying a detailed view of underground conditions surrounding water mains. Further testing in plastic water mains may prove helpful in detecting water leakage that is not audible due to material conditions.

REFERENCES

Babbitt HE, Amsbary, FC, Gwinn DR. 1920. The detection of leaks in underground pipes. Journal of the American Water Works Association. *7*(4): 589–595.

Candelieri A, Messina E. 2012. Sectorization and analytical leaks localization in the h2oleak project: clustering-based services for supporting water distribution networks management. Environmental Engineering & Management Journal (EEMJ). 11(5): 953–962.

Esri. 2019. Parameter optimization. Esri [Internet]. [cited 2019 Aug 23]. Available from: https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/parameter-optimization.htm

Frauendorfer R, Liemberger R. 2010. The issues and challenges of reducing non-revenue water. Asian Development Bank. 49 p.

GeoPandas. 2019. GeoPandas 0.5.0 documentation. GeoPandas [Internet]. [cited 2019 Jun 29]. Available from: http://geopandas.org/

Harrison M, Prentiss M. 2016. Learning the pandas library: python tools for data munging, analysis, and visual. 1st ed.. USA: CreateSpace Independent Publishing Platform.

Hodgins M, Sturm R, Gasner K. 2016. Managing water loss control. Opflow Online. 42(4): 10–14. https://doi.org/10.5991/OPF.2016.42.0019

Hunaidi O. 1998. Ground-penetrating radar for detection of leaks in buried plastic water distribution pipes. Seventh International Conference of Ground-Penetrating Radar: Lawrence, Kansas.

Hunaidi O, Chu W, Wang A, Guan W. 2000. Detecting leaks in plastic pipes. American Water Works Association. Journal; Denver, 92(2): 82.

Hunaidi O. 2012. Acoustic leak detection survey strategies for water distribution pipes. Construction Technology Update, 79: 1–5.

Izquierdo J, Herrera M, Montalvo I, Pérez-García R. 2009. Division of water supply systems into district metered areas using a multi-agent based approach. Software and Data Technologies, 167–180. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20116-5_13

Klisel KA, Murray R, Haxton T. 2018. An overview of the water network tool for resilience (wntr). International WDSA/CCWI 2018 Joint Conference, Kingston, Ontario, Canada.

Lambert A. 2000. What do we know about pressure leakage relationships in distribution systems? IWA Conference Proceedings; May 2000. Brno, Czech Republic

McKinney WG. 2010. Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference; 2010.

Oliver MA, Webster R. 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. CATENA. 113: 56–69. https://doi.org/10.1016/j.catena.2013.09.006

Rajani B, Makar J. 2000. A methodology to estimate remaining service life of grey cast iron water mains. Canadian Journal of Civil Engineering. 27(6): 1259-1272.

Sage P. 2014. Practical methods to obtain improved outputs from water network modelling optimization. Procedia Engineering. 70: 1450–1459. https://doi.org/10.1016/j.proeng.2014.02.160

Samir N, Kansoh R, Elbarki W, Fleifle A. 2017. Pressure control for minimizing leakage in water distribution systems. Alexandria Engineering Journal. 56(4): 601–612. https://doi.org/10.1016/j.aej.2017.07.008

Savić D, Ferrari G. 2014. Design and performance of district metering areas in water distribution Systems. Procedia Engineering. 89: 1136–1143. https://doi.org/10.1016/j.proeng.2014.11.236

Scheeres A. 2016. Kriging: spatial interpolation in desktop gis. Asavea [Internet]. [cited 2019 Aug 6]. Available from: https://www.azavea.com/blog/2016/09/12/kriging-spatial-interpolation-desktop-gis/

Scibetta M, Boano F, Revelli R, Ridolfi L. 2013. Community detection as a tool for complex pipe network clustering. EPL (Europhysics Letters). 103(4): 48001. https://doi.org/10.1209/0295-5075/103/48001

Walski TM, Chase DV, Savic DA, Grayman W, Beckwith S, Koelle E. 2003. Advanced Water Distribution Modeling and Management. 1st ed. Haestead Press.

Walski TM. 1983. Technique for calibrating network models. Journal of Water Resources Planning and Management. 109(4): 360–372. https://doi.org/10.1061/(ASCE)0733-9496(1983)109:4(360)

Webster R, Oliver MA. 2007. Geostatistics for environmental scientists. 2nd ed. Chichester (West Sussex): John Wiley and Sons Ltd.

Wills P, Memon FA, Savic D. 2017. November. High-resolution domestic water consumption data – scope for leakage management and demand prediction. Retrieved from https://ore.exeter.ac.uk/repository/handle/10871/30367

Winarni W. 2009. Infrastructure leakage index (ILI) as water losses indicator. Civil Engineering Dimension. 11(2): 126-134.

Witten A. 2006. Ground penetrating radar: a true wave-based technique. Handbook of Geophysics and Archaeology. 214–58. London: Equinox.

Wu ZY, Sage, P, Turtle D. 2010. Pressure-dependent leak detection model and its application to a district water system. Journal of Water Resources Planning & Management. 136(1): 116–128. https://doi.org/10.1061/(ASCE)0733-9496(2010)136:1(116)

Xu Q, Liu R, Chen Q, Li R. 2014. Review on water leakage control in distribution networks and the associated environmental benefits. Journal of Environmental Sciences, 26(5): 955–961. https://doi.org/10.1016/S1001-0742(13)60569-0

APPENDIX: Process Automation Python Code

https://github.com/jktittle/Water-Leakage-Surface-Updates/blob/master/WaterLeakagePressureUpdates.ipynb

Script to Update Hydraulic Grade Values and Interpolate Sample Points Using Ordinary Kriging

Introduction

The Python component of this project was used to automate daily data imports and maintenance required to dependably produce leakage area results. Since the project is applied to a real-world water distribution system, data were generated daily, creating the need to automate labor intensive tasks. To do this, the pandas and geopandas Python libraries were used to handle most of the data management by performing data imports, data cleanup, data table merges, and hydraulic grade calculations. This was accomplished by importing regularly generated pressure information into pandas data frames and hydrant locational information into a geopandas spatially enabled geodataframes.

The static pressure update processing is an ongoing program at the water department, so an automated script was developed to capture additional hydrant pressure tests daily. This placed information indicating large areas exhibiting water leakage in front of decision-makers in a timely manner so that large water breaks can be identified and repaired. The script was written in Python and developed using Jupyter Notebooks in conjunction with ArcGIS to document each step and support replication in other water systems.

```
#Import the following python libraries
import sys, os, csv, fiona, datetime, arcpy
import pandas as pd
import numpy as np
import geopandas as gp
from geopandas import GeoSeries, GeoDataFrame
from shapely.geometry import Point
import matplotlib.pyplot as plt
from arcpy import env
from arcpy.sa import *
```

```
#Set the environment workspace and overwrite settings
arcpy.env.overwriteOutput = True
arcpy.env.workspace = "C:\\StaticPressureProcess\\StaticPressureData.gdb"
```

```
#Create variables for the pressure update csv, pressure test point file, and pr
essure zone polygon file
pressureUpdateFile = "C:\\StaticPressureProcess\\TasksExport.csv"
pressurePoint = "C:\\StaticPressureProcess\\StaticPressureData.gdb\\PZ1838A_Pre
ssureTestPnts"
```

```
pressureZone = "C:\\StaticPressureProcess\\StaticPressureData.gdb\\PZ1838A_Redefined"
outRaster = "C:\\StaticPressureProcess\\StaticPressureData.gdb\\LeakSurface_" + datetime.date.today().strftime("%m%d%Y")
clippedRaster = "C:\\StaticPressureProcess\\StaticPressureData.gdb\\ClippedSurface_" + datetime.date.today().strftime("%m%d%Y")
redefined1838aDma = "C:\\StaticPressureProcess\\DMA1838A.shp"
geoStatModel = "C:\\StaticPressureProcess\\OrdinaryKrigingModel_1838A_TheBest.xml"
geoStatLayer = "KrigingOutLayer"


#Use the fiona library to list all layers within the StaticPressureData geodatabase.
#The list will be used to reference the layer imported with geopandas
fiona.listlayers("C:\\StaticPressureProcess\\StaticPressureData.gdb")


#Import the PZ1838_PressureTestPnts feature class as a geodataframe.
#The layer parameter is taken from the fiona generated list position of the desired geodatabase feature class.
testSites = gp.read_file("C:\\StaticPressureProcess\\StaticPressureData.gdb",driver='FileGDB', layer=3)
testSites


#Standardize the DateCollected column
testSites['DateCollected']=pd.to_datetime(testSites['DateCollected'])
testSites


#Import the pressure updates csv into a pandas data frame
staticUpdates = gp.read_file( "C:\\StaticPressureProcess\\TasksExport.csv")
staticUpdates


#Add the FACILITYID column and slice the text to only contain hydrant identifiers
staticUpdates['FACILITYID'] = staticUpdates.Asset.str[14:]

#Replace spaces with underscores
staticUpdates.columns = staticUpdates.columns.str.replace(' ', '_').str.replace('(', '').str.replace(')', '')

#Convert the Static_Pressure column to numericvalues
staticUpdates['Static_Pressure']=pd.to_numeric(staticUpdates.Static_Pressure)
staticUpdates
```

```
#Standardize the Actual_Stop_Date column
staticUpdates['Actual_Stop_Date']=pd.to_datetime(staticUpdates['Actual_Stop_Dat
e'])
staticUpdates


#Find and remove all rows with a Static_Pressure value equal to zero
zeroStaticP = staticUpdates[ staticUpdates['Static_Pressure'] == 0 ].index
staticUpdates.drop(zeroStaticP , inplace=True)

#Find and remove all rows with a Static_Pressure value greater than 200
zeroStaticP = staticUpdates[ staticUpdates['Static_Pressure'] > 300 ].index
staticUpdates.drop(zeroStaticP , inplace=True)

staticUpdates


#Join the staticUpdates data frame to the testSites data frame using the FACILI
TYID field
#This creates a new data frame that contains the static pressure updates to app
ly to the 1838A test hydrants
mergedPressureInfo = testSites.merge(staticUpdates, on='FACILITYID')
mergedPressureInfo


#Update new StaticPressure column
mergedPressureInfo.StaticPressure = mergedPressureInfo.Static_Pressure

#Recalculate the Hydrograde column
mergedPressureInfo.HydroGrade = mergedPressureInfo.Elevation + 2.31 * mergedPre
ssureInfo.StaticPressure

#Update the DateCollected column with new dates from the Actual_Stop_Date colum
n
mergedPressureInfo.DateCollected = mergedPressureInfo.Actual_Stop_Date
mergedPressureInfo


#Remove unneeded fields from the data frame
del mergedPressureInfo['Task_ID']
del mergedPressureInfo['Asset']
del mergedPressureInfo['Activity']
del mergedPressureInfo['Static_Pressure']
del mergedPressureInfo['Actual_Stop_Date']
```

```
del mergedPressureInfo['geometry_y']

#Rename the geometry column
mergedPressureInfo.rename(columns={"geometry_x":"geometry"}, inplace=True)
mergedPressureInfo


#Remove duplicate values
mergedPressureInfo = mergedPressureInfo.sort_values('DateCollected',ascending=T
rue)
mergedPressureInfo = mergedPressureInfo.drop_duplicates(subset='FACILITYID', ke
ep='first')
mergedPressureInfo = mergedPressureInfo.sort_values('FACILITYID',ascending=True
)
mergedPressureInfo
```

Update testSite values with the new static pressure test values and export to a shapefile

```
#Set the testSites index to the FACILITYID column
testSites = testSites.set_index('FACILITYID')


testSites


#Set the mergedPressureInfo data frame index to the FACILITYID column
mergedPressureInfo = mergedPressureInfo.set_index('FACILITYID')


mergedPressureInfo


#Run the update function on the testSites data frame
testSites.update(mergedPressureInfo)


#Reset the indexes
testSites.reset_index(inplace=True)


testSites


#Convert the merged data frame to a GeoDataFrame and remove null HydroGrade Val
ues
```

```
updatedGdf = gp.GeoDataFrame(testSites, geometry='geometry')
NewGdf = updatedGdf[updatedGdf.HydroGrade.notnull()]
NewGdf


#Convert the DateCollected column to string values in order to export to shapef
ile
NewGdf['DateCollected']=NewGdf['DateCollected'].astype(str)


#Set the new geodataframe's projection and plot the new pressure tests within 1
838A
NewGdf.crs = {"init":"epsg:2274"}
updatedGdf.plot(figsize=(12,12));


##Create new shapefile name and export the geodataframe to new shapefile
shpFileName = r"C:\StaticPressureProcess\UpdatedStaticPressureTests_" + datetim
e.date.today().strftime("%m%d%Y") + ".shp"


NewGdf.to_file(shpFileName)
```

Run the ordinary kriging model on the updated hydrant pressure points

Run the Kriging interpolation using the pressure point layer. This step creates a Geostatistical Layer using tools from Geostatistical Analyst. The tool uses an existing Geostatistical layer as a model source to duplicate its parameters and should be stored in the project workspace.

```
#Check out the ESRI Spatial and Geostatistical Analyst Extensions
arcpy.CheckOutExtension("Spatial")
arcpy.CheckOutExtension("GeoStats")

krigingInLayer = shpFileName + " X=Shape Y=Shape F1=HydroGrade"


#arcpy.GACreateGeostatisticalLayer_ga(in_ga_model_source, in_datasets, out_laye
r)
arcpy.GACreateGeostatisticalLayer_ga(geoStatModel, krigingInLayer, geoStatLayer
)
```

#Export Geostatistical layer to a raster arcpy.GALayerToRasters_ga(geoStatLayer, outRaster)

```
#Clip the interpolation surface to the desired polygon boundary layer
```

```
arcpy.Clip_management(outRaster, "#",clippedRaster, redefined1838aDma,"0","
ClippingGeometry")


#Check back in the ESRI Spatial and Geostatistical Analyst Extensions
arcpy.CheckInExtension("Spatial")
arcpy.CheckInExtension("GeoStats")


print("Completed Script")
```

VITA

JACOB TITTLE

Education: M.S. Geosciences, East Tennessee State University;
Johnson City, Tennessee 2019

B.S. Geosciences, Tennessee Technological University;
Cookeville, Tennessee 2006

Professional Experience: Geospatial Coordinator, City of Johnson City; Johnson City,
Tennessee, July 2019 – Present

GIS Analyst, City of Johnson City; Johnson City,
Tennessee, July 2015 – June 2019

GIS Database Specialist, City of Johnson City; Johnson City,
Tennessee, June 2008 – July 2015

GIS Technician, Anderson and Associates, Inc.; Johnson City,
Tennessee, August 2006 – May 2008

GIS/GPS Technician, Anderson and Associates, Inc.; Johnson City,
Tennessee, May 2005 – December 2005