James Madison University

JMU Scholarly Commons

5-8-2020

A novel approach to updating municipal tax parcel impervious surface calculations

Patrick D. Muradaz
James Madison University

A Novel Approach to Updating Municipal Tax Parcel Impervious Surface Calculations

_____

An Honors College Project Presented to

the Faculty of the Undergraduate

College of Integrated Science and Engineering

James Madison University

_____

by Patrick David Muradaz

May 2020

Accepted by the faculty of the Department of Geographic Science, James Madison University, in partial fulfillment of the requirements for the Honors College.

FACULTY COMMITTEE: HONORS COLLEGE APPROVAL:

Project Advisor: Zachary J. Bortolot, Ph.D.   Bradley R. Newcomer, Ph.D.,
Professor, Geographic Science   Dean, Honors College

Reader: Helmut Kraenzle, Ph.D.
Professor, Geographic Science

Reader: Nathan R. Sprague, Ph.D.
Professor, Computer Science

PUBLIC PRESENTATION

Due to COVID-19 complications this work is not required to be publicly presented.

Table of Contents:

List of Figures:

Procedure for Implementation:

Project Results:

Purpose and Intent:

This project was focused on creating a procedure for updating municipal tax parcel impervious surface estimates. The goal was to devise and carry out a procedure for updating impervious surface maps that any municipality with access to aerial imagery and a past impervious surface map, could easily implement to save the update effort time and money. This project was undertaken as a case study with the City of Harrisonburg to implement and demonstrate the procedure.

Having accurate impervious surface calculations is important to many municipalities because higher impervious surface density leads to higher volumes of surface rainwater runoff. Many municipalities must deal with this runoff through the establishment and maintenance of drainage infrastructure. For example, according to the Municipal Separate Storm Sewer System (MS4) Permit Program site on the Harrisonburg website "The City is required to develop, implement, and enforce a stormwater management program". To help offset the added cost of this infrastructure many municipalities, including Harrisonburg, impose taxes and fees on privately owned impervious surfaces such as driveways and patios. As of 2017 the City of Harrisonburg has imposed an annual fee of $6.00 for every 500 square ft of impervious area in a tax parcel. This money is used specifically to pay for the City's projects and programs to reduce polluted runoff (Stormwater Utility, 2017). Currently, in order for Harrisonburg to calculate tax parcel impervious surface density, aerial images must be manually digitized or mapped automatically using predictive models. These methods of impervious surface calculations can cost municipalities thousands of dollars and/or hundreds of man hours. According to Rebecca

Stimson at the City of Harrisonburg, in recent years the City has used LiDAR data to help with map digitizing at a cost of tens of thousands of dollars.

The specific objective of this thesis was to create a procedure that can be implemented by most GIS professionals, and will be able to flag tax parcels whose impervious surface density has likely increased and may need to be updated in the current map. This was accomplished by establishing a statistical relationship between current aerial imagery attributes, such as texture, and past impervious surface density calculations. This method is unusual in that it does not use both past imagery and tax parcel imperviousness labels to train a regression model that is then applied to current imagery. Rather, it uses current imagery and past labels to train the model that is then applied to the same current imagery. This provides good results because only a very small portion of the tax parcels will have any change at all in permeability (given the earlier permeability labels are relatively recent to the imagery used, in this case two years, which is Harrisonburg's standard update window). Thus, the vast majority of the imagery will fit the tax parcel labels and the model is not as reliant on perfect fitting of imagery and labels. The largest advantage to using this method is that it requires only one dataset of aerial imagery and thus requires less legwork to implement, making it easier for others to implement on their own. A disadvantage to this method is that the produced model likely has a higher rate of false positive predictions than a model fit using the traditional approach.

Tax parcels that showed a significant negative difference between current prediction and past calculation (predicted imperviousness was greater) were flagged to be looked at by a person who can then update the impervious surface calculation. Looking specifically at parcels where imperviousness is predicted to have increased is beneficial as property owners who have removed impervious surface will notify the city so as to save money on taxes (thus parcels with

5

decreased imperviousness need not be identified). This procedure will be useful to municipalities because it offers an alternative to the costly and time-consuming methods currently available. Although the City of Harrisonburg was used as a case study to create the procedure, other municipalities that have access to high resolution aerial imagery and pre-existing impervious surface maps should be able to tailor this method for their own use.

Abstract:

Accurate impervious surface calculations are important to many municipalities due to the high volumes of surface rainwater runoff caused by high impervious surface density. Municipalities must deal with this runoff through the establishment and maintenance of drainage facilities. To help offset the added cost of these facilities, many municipalities impose taxes and fees on privately owned impervious surfaces such as homes, driveways, and patios. Currently, in order for a city like Harrisonburg to calculate tax parcel impervious surface density, aerial images must be manually digitized or mapped using computer-based classification techniques using predictive models. These methods of impervious surface calculations can cost municipalities thousands of dollars and/or hundreds of person hours. The purpose of this project is to devise and test a novel procedure for updating municipal tax parcel impervious surface calculations in an effort to help reduce the time and cost currently required to perform these updates. The procedure that was devised establishes a statistical relationship between current aerial imagery attributes (such as mean pixel value, texture, and homogeneity) and past impervious surface calculations. Tax parcels where actual impervious surface density differed greatly from the model's expected density were flagged to be looked at manually for possible update. This procedure is useful to municipalities because it cuts costs and labor time compared to other methods currently available. Although this project is a case study for the City of Harrisonburg, this procedure could potentially be tailored to other municipalities that have access to high resolution aerial imagery and pre-existing impervious surface maps.

Research Context:

A literature review was conducted prior to the beginning of work on the procedure and was used to establish the validity of the core principle of the thesis, that imagery attributes can be used to build a model to predict imperviousness within a bounded area, and to guide preliminary plans for carrying out the project. Prior to conducting research, the plan was to use supervised classification to identify impervious surfaces in imagery. However, while conducting this research the plan evolved. While supervised classification methods are used often in this field and are typically able to classify impervious surfaces, often with high accuracy; these methods have already been researched extensively and this project seeks to add to the field by pursuing a different approach. Furthermore, supervised classifications require a significant amount of training data and would likely be difficult to encapsulate in a way that could be quickly/easily carried out by others. It seemed as though parcels that require updating could be identified faster by most GIS professionals, and with less incurred time and cost, by using the methodology that was used in this project.

The mapping of impervious surfaces has a long history and has relied on many methods from field surveying to manually digitizing using aerial photography (Weng, 2012). This review will focus specifically on the methods that have been used to improve the accuracy of impervious surface estimating and mapping since the use of aerial imagery became popular in the 1970s and 1980s (Slonecker et al., 2001). A discussion of these methods will help contextualize the project and will provide a background to the discussion of how the project adds a new perspective to the problem of impervious surface estimating and mapping.

There are a wide range of methods for impervious surface estimating and mapping that have been used since the 1970s. This review highlights four of the more recent and relevant methods that are discussed in a number of recent papers (Wang, et al., 2018), (Wei & Blaschke, 2018), and (Weng, 2012). These methods include: regression analysis, per-pixel classification, image classification, and object-based classification.

Regression analysis has been used to categorize super-pixel groups and individual pixels based on linear relationships between predictive variables gathered from the pixels and the likelihood that those pixels or pixel groups are of impervious surfaces (Bauer *et al*., 2007; Weng, 2012;). Predictive variables include pixel color and value and texture. Impervious surfaces tend to be darker and less green (Bauer *et al.*, 2007), and to have a smoother texture (Weng, 2012). One classification method works by employing a classification and regression tree (CART) to classify pixels at an individual level and use those classified pixels to recursively classify larger and larger image segments (Weng, 2012). At the pixel level, an individual pixel's spectral variables are used to predict the likelihood of that pixel being of an impervious surface based on the linear relationship between other pixel variables and their likelihood of being of impervious surfaces. This linear relationship is established using high-resolution aerial imagery and pre-classified data (Weng, 2012). An issue with this method is that CART algorithms typically overestimate impervious surfaces in low-density urban areas while underestimating them in high-density urban areas (Wang, *et al*., 2018).

Per-pixel classification has, until recently, been the prevailing method for impervious surface mapping (Weng, 2012). There are two methods of per-pixel classification: Hard and soft classification. Hard classification is the assignment of individual pixels to exclusive land cover categories based on the spectral properties, such as color, of the individual pixels. The issue with

assigning pixels to only one land category each is that it ignores the fact that, at certain resolutions, some pixels constitute a sum of multiple land cover categories (Strahler, Woodcock, & Smith, 1986). One solution to this is to employ a soft classification method where pixels are assigned a degree of membership of each of the possible land cover categories. For example, a pixel on the edge of a riverbank could be classified as 50% water and 50% bare soil. The issue with both of these methods is that they do not actually identify or quantify image objects, they simply categorize discrete pixels (Weng, 2012). For reference, an image object is any object that can be identified in imagery, for example a house or a tree.

Some image classification methods are able to consider not only individual pixels but also patterns, such as texture and shape, found in adjacent pixels and in the image as a whole (Sharma & Sarkar, 1998). This is useful because this helps to reduce the issue of mixed pixels (Weng, 2012). This type of image classification is typically carried out using Artificial Neural Networks or Support Vector Machines (Weng, 2012) & (Tuia, et al., 2009). These algorithms, whether they are supervised or unsupervised, require a large amount of quality training data in order to accurately classify impervious surfaces (Tuia, et al., 2009). This approach does not segment the image into discreet objects and therefore suffers from some of the same limitations as per-pixel classification (Weng, 2012).

Object-based classification is able to segment and classify discrete objects in imagery by incorporating spectral data, shape, texture, and context characteristics, as well as neighbor pixel and super- or multi-pixel relationships (Weng, 2012). Object-based classification is currently employed by algorithms that function autonomously or semi-autonomously (Wei & Blaschke, 2018). Currently, object-based classification is used commercially by companies such as Trimble, ESRI, and Sanborn. This type of classification requires a large amount of quality

training data and according to those I spoke to at the city of Harrisonburg often still must be quality checked by a human.

Many impervious surface mapping methods being worked on recently incorporate more than just spectral aerial imagery (Weng, 2012) to estimate impervious surface density. Such methods include utilizing LiDAR and spatial data (Luo, et al., 2018) & (Lee, Kim, & Lee, 2018), as well as temporal data (Tewkesbury, et al., 2015), which look at change over time. These methods are popular due to their ability to significantly improve the performance of impervious surface estimation algorithms (Weng, 2012). One of the drawbacks associated with incorporating LiDAR or spatial data into estimation algorithms is the cost associated with collecting these data. Furthermore, Luo, et al. (2018) point out that LiDAR data gathered at a different date than spectral data can impact estimations if the difference is not dealt with.

The research that was most beneficial to the project was that which focused on aerial imagery feature extraction, including methods for extracting spectral attributes such as minimum and maximum pixel values, mean values, and texture, as well as, spatial attributes such as size and shape of super-pixel groups. Extracting these kinds of attributes is of significant importance because these attributes have been shown to relate to impervious surface density (Bauer *et al*, 2007; Weng, 2012). Körting et al. (2018) developed the Geographic Data Mining Analyst (GeoDMA) software package which is specifically designed to extract the imagery attributes necessary to carry out this project.

Procedure for Implementation:

       The procedure outlined in this section should be able to be carried out by any GIS professional in the timeframe of about 1-2 working weeks. Through following this procedure, hundreds of tax parcels can be easily identified for updating at a cost of $0 on top of normal employee compensation (assuming the municipality has free access to aerial imagery).

       The following software is required to implement the procedure. Software that was used during the completion of the project is indicated by underlining.

- ESRI ArcMap, ArcGIS Pro, or other GIS software (e.g. QGIS)

- Excel

- GeoDMA (download link and instructions found in *Resources*)

- SPSS or custom script written in R (code and instructions found in *Resources*)

Steps:

1. Obtain necessary data:

       This process requires tax parcel and impervious surface data in vector format and aerial imagery in raster format. The tax parcel data should include percent impervious surface as an attribute for each parcel. If not, a separate impervious surface layer giving the percent impervious surface for each parcel will be needed.

       For this project, the impervious surface and tax parcel map layers were obtained as separate shapefiles and were created for the 2016 tax map. Aerial imagery for the project consisted of tiles from the Virginia Base Map Program (VBMP) for the year 2018, which can be downloaded for free. A link to the download site can be found in *Resources* for Virginia municipalities.
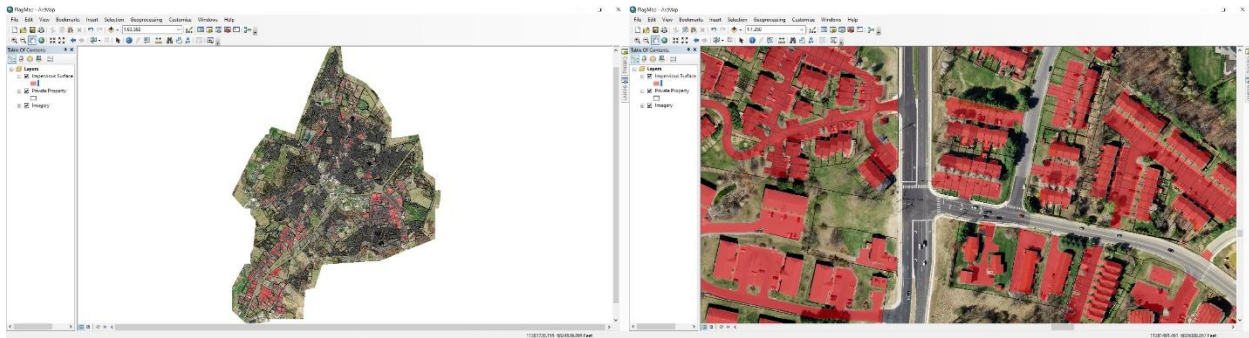
Figure 1.1 Initial data. The left image contains imagery and vector data of the whole of Harrisonburg after preprocessing (discussed in the next two steps). The right image contains a zoomed in sample of a Harrisonburg neighborhood. The black lined polygons are tax parcels and the red polygons are 2016 impervious surfaces.

2. Preprocess imagery:

        Imagery from VMBP comes in small tiles that need to be mosaicked together. This can be done in ArcMap using the 'Mosaic To New Raster' tool. It should be noted that GeoDMA runs into problems when working with very large images. To solve this, the imagery may need to be resampled to a lower resolution and/or split up. It seems that raster files under 1GB are able to be processed by GeoDMA without error.

        For this project, the imagery was resampled from .25 ft to 1ft resolution using the nearest neighbor ArcMap resample function. The city image was then divided into four tiles to be run through GeoDMA separately. If you do divide the image into smaller tiles, it is important that you pay attention to parcels that are on the edge of tiles to avoid having imagery under only part of them. I recommend making tiles that overlap each other somewhat to avoid this problem.

        Throughout imagery preprocessing it is important to ensure that the imagery remains in the same projection as the vector data. These data need to overlay each other for GeoDMA to work correctly and mismatched projections can cause errors or incorrect output.

3. Preprocess vector data

GeoDMA works by attaching extracted imagery attributes from raster imagery to overlaying vector data. The overlaying vector data used in this process is the tax parcel layer. Before running GeoDMA it is important to follow the steps below to ensure that the tax parcel data is clean and contains all necessary attributes (an ID attribute to be used for a future join and an attribute for percent impervious).

Ensure that any features that have geometry but lack any values in the attribute table are removed. It is recommended that the vector data include only taxable parcels as this will reduce the number of polygons that GeoDMA needs to extract to and will yield better end results as it narrows the number of parcels that may need updates. For this project, government, religious, and education institution-owned properties were excluded from the dataset.

Before running GeoDMA, the tax parcel data must have an attribute containing the percent of impervious surface in the parcel. For this project, a spatial join was used to join the impervious surface layer to the tax parcel layer. Once joined, it was possible to calculate the impervious surface percentage of each tax parcel by dividing the area attribute from the impervious surface layer by the area attribute from tax parcel layer.

Once the tax parcel layer has been cleaned, formatted, and has the percent impervious attribute; imagery attributes can be extracted to the tax parcel attribute table using GeoDMA.

4. Extract imagery features:

This step has been split into multiple parts as the process may be unfamiliar to some GIS professionals.

15

a. <u>Create working databases:</u>

Open TerraView, by running the 'terraview' application in the GeoDMA download folder.

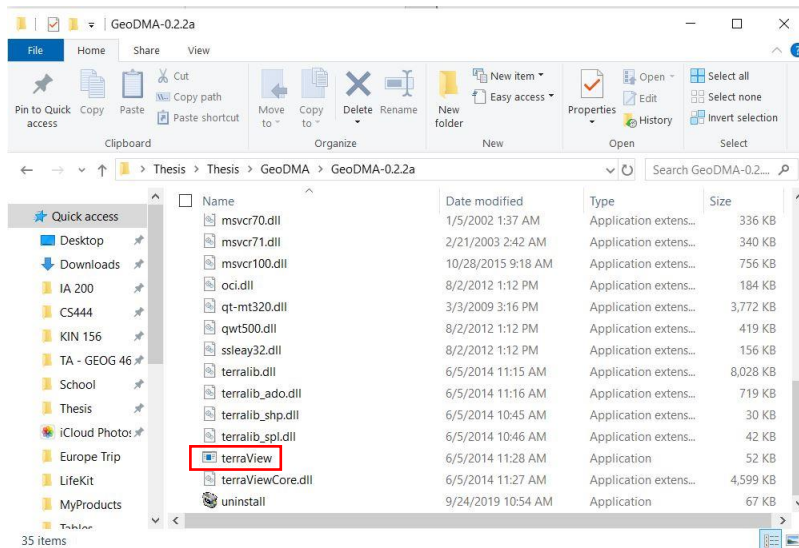Figure 1.2 Locating Terraview application

Create a new database to house the imagery and vector data:

1. Select the 'Database' icon in the upper left corner
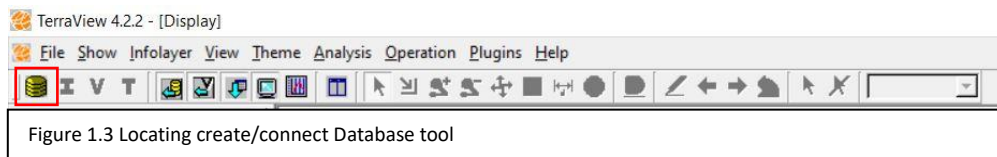
Figure 1.3 Locating create/connect Database tool

2. Within the database window select 'Create'

3. Set Database Type to 'Access'

4. Select 'Directory…' and browse to a location to save the database

5. Give the database a name and select 'OK'.

Figure 1.4 Database create/connect window
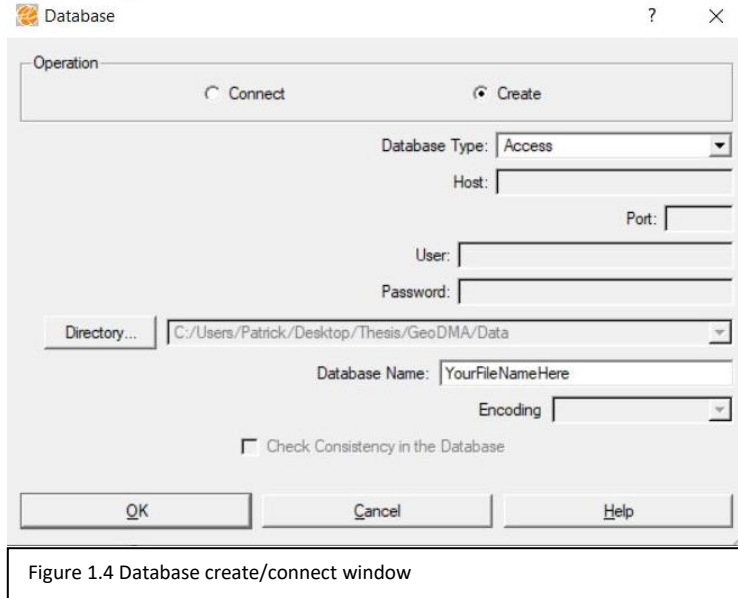
Follow these steps for each database that is to be created. A database should be created for each imagery/vector data section.

b. <u>Import data to databases:</u>

To import vector data:

1. Select the 'Import Data' icon in the upper left corner

2. Within the import window browse to select the data to import and select 'OK'

Figure 1.5 Vector data import window

To import imagery data:

1. Select 'File' > 'Simple Raster Import'

2. Within the import window browse to select the data to import and select 'OK'
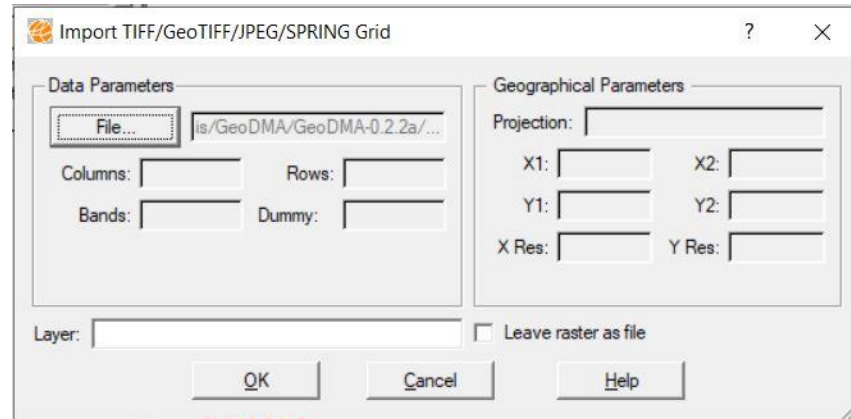
Figure 1.6 Raster data import window

c. Run GeoDMA plugin

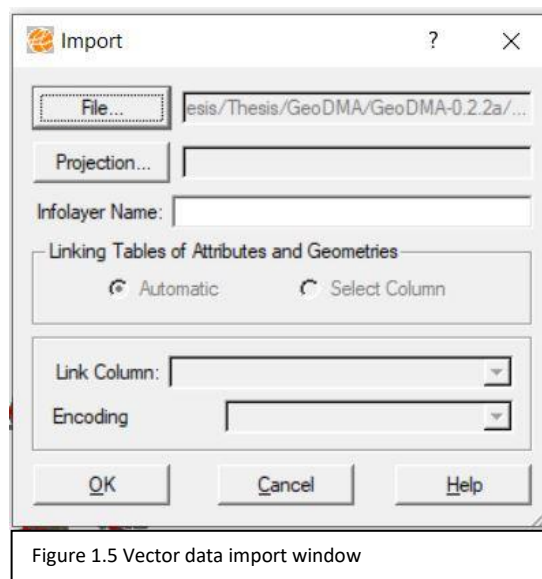 Before running GeoDMA ensure that all data is in the same view under 'Views/Themes' and that the vector data overlays the imagery (this can be checked selecting 'Draw' in the center right of the tool ribbon.)
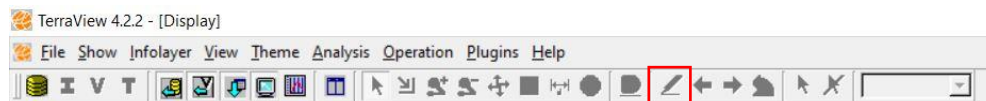
Figure 1.7 Locating Draw tool

 If the vector data does not overlay the imagery then the projections may differ which will cause imagery feature extraction to fail or work incorrectly.
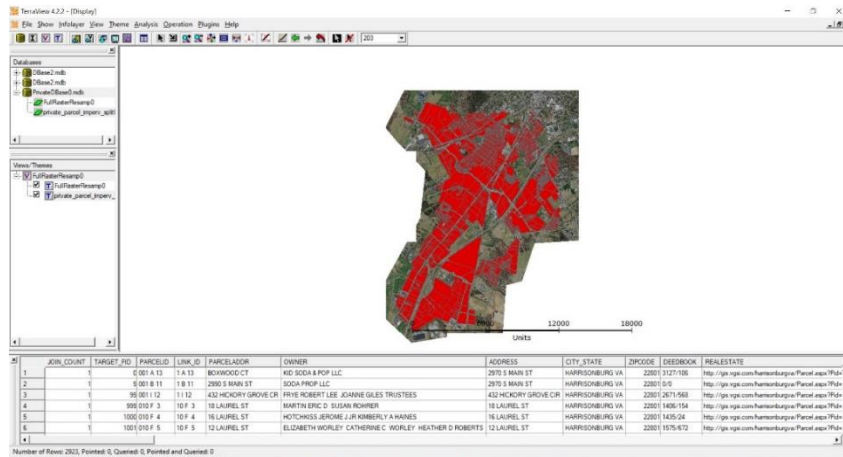
Figure 1.8 Terraview display. Note: this is an example of the imagery and vector data in the same projection.

Once the data has been checked, run GeoDMA by selecting 'Plugins' > 'GeoDMA'.

Within the plugin window:

1. Under the 'Input' tab, ensure the proper layers are in the proper fields (vector data in *Polygons* fields and imagery in *Raster Data* field) and select 'OK'
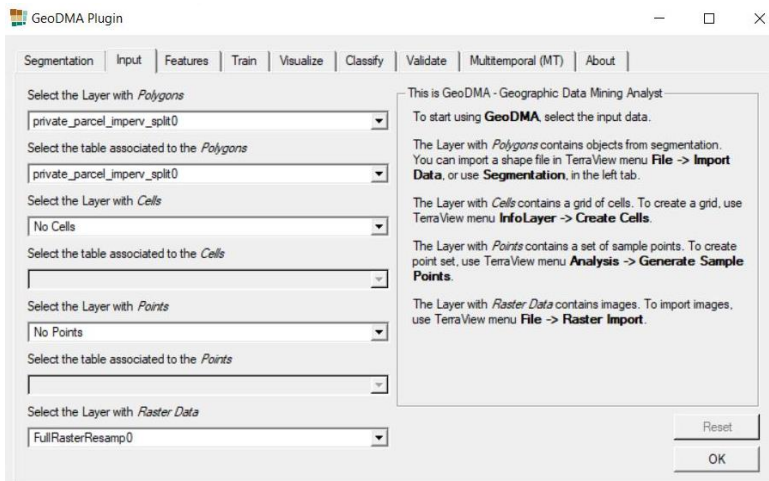
Figure 1.9 GeoDMA Input window

2. Under the 'Features' tab, select the desired features to extract. For this procedure to work any selected features must be spectral combined with

polygons, denoted with the prefix 'rp_' (for this project all spectral/spatial

features were used)

3. Once the desired features are selected, click 'Extract'
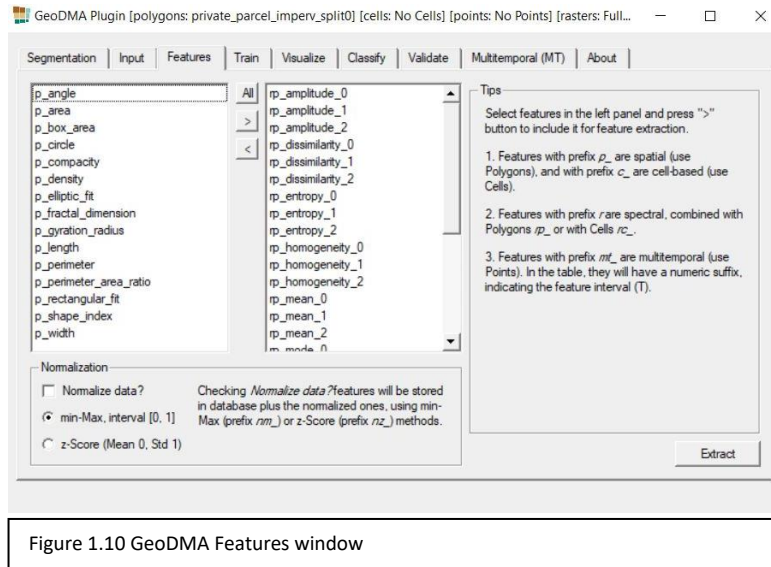
Figure 1.10 GeoDMA Features window

d. Export table

After following the steps given above, the extracted variables from the

imagery will be added as attributes to the vector data attribute table. In order to

move forward in the process, the attribute table must be exported to an excel file.

1. Under the 'Databases' table of contents right click the vector data and select

'Export layer…'
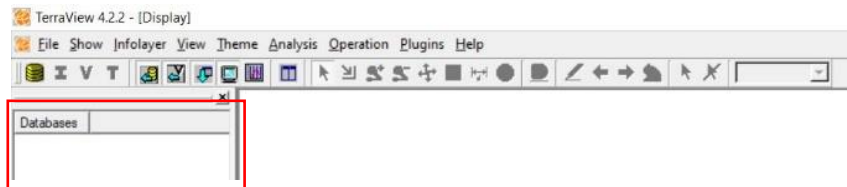
Figure 1.11 Locating Databases table of contents

2. Within the Export layer menu, under 'Format' select 'Shapefile', under

'Output' save the file to a suitable location. Select 'OK'
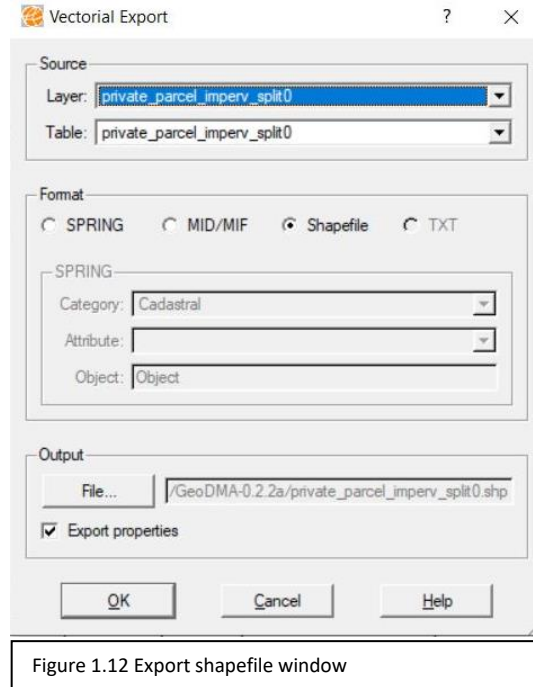
Figure 1.12 Export shapefile window

3. Inside ArcCatalog open the tool search menu and look up the 'Table to Excel' tool and select it

4. Inside the tool menu set the input table to the previously exported shapefile and use the output excel file menu to save the excel file to a suitable location. Select 'OK' when satisfied

Once the above process has been followed for all imagery sections the data is ready to be merged in preparation for regression.

5. Merge data for regression:

To perform the stepwise regression, all of the data must be in the same table. For R script compatibility, the table must have all independent variables (imagery features) and the dependent variable (percent imperviousness) adjacent to each other. Take the exported tables from each imagery section (output from the steps above) and merge them into a single table, the order does not matter. Check the final table for any errors that may

have gone unnoticed. Once the data has been checked and merged into a single table stepwise regression can be performed.
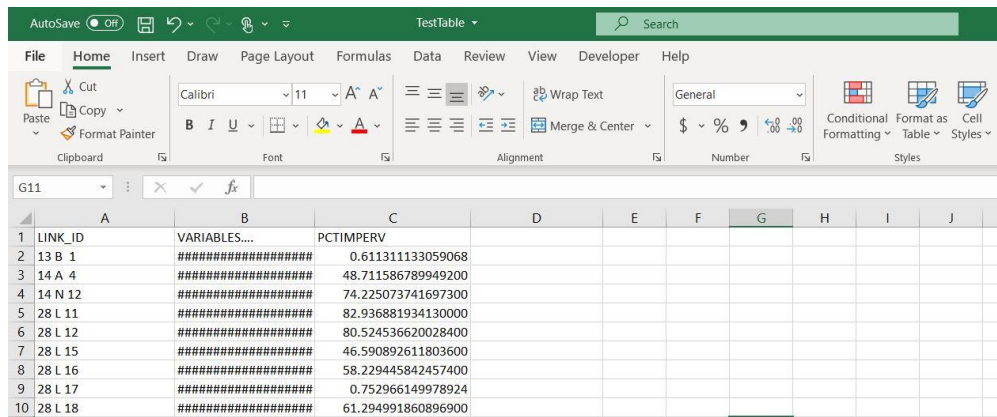
6. Run stepwise regression:

    a. Regression through SPSS:

        Because of the wealth of information on this topic that already exists, this paper will not provide a guide for running stepwise regression through SPSS. The following online tutorial may be helpful: https://www.spss-tutorials.com/stepwise-regression-in-spss-example/. Alternatively, the author of this paper and/or his advisor may be contacted for assistance. Contact info can be found in the *Resources* section.

    b. Regression through custom R script:

        To use the custom R script, follow the instructions outlined in *Resources* to download and install R and to run the script in the R scripting window. The input data required is a list of independent variables (imagery features) and the dependent variable (percent imperviousness) for each tax parcel imported from an Excel file.

Figure 1.13 sample input excel data

Follow instructions in step 5 to ensure the Excel data are formatted correctly.

The output of the regression is a list of predictions calculated using a formula of the form: $C + M_1F_1 + M_2F_2 + \ldots + M_nF_n$ where C is a constant determined by the regression, M is the coefficient determined by the regression, F is the imagery feature attribute, and n is the total number of features that were extracted for each polygon based on the imagery. This list will be saved by R as a .csv file. Each line in the file will correspond to a single parcel, and they will be in the same order that the tax parcels were arranged in the Excel file that was input into R.

Figure 1.14 Sample R script output
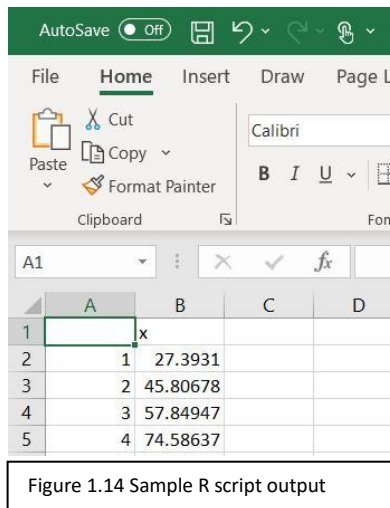
To merge everything into one Excel sheet, simply copy the output predictions from the .csv and paste them into the Excel file.

Figure 1.15 Sample excel table merge

7. <u>Use model to flag parcels:</u>

In order to flag parcels, the regression predictions will need to be compared to the values calculated using the past vector data to create a difference field in the Excel spreadsheet. To create the difference field, either subtract the predictions from the past values or vice versa. Once the difference field is created, the Excel file can be joined with the tax parcel spatial data. Once the join is complete, sort the table so that the tax parcels where the predicted imperviousness values are much greater than the past imperviousness values are at the top (if the difference field was created by subtracting the model prediction from the past value sort ascending, if vice versa sort descending). Starting at the top, select tax parcels in the table to flag them in the map so they can be visually inspected. Examples can be found on the following page.

24
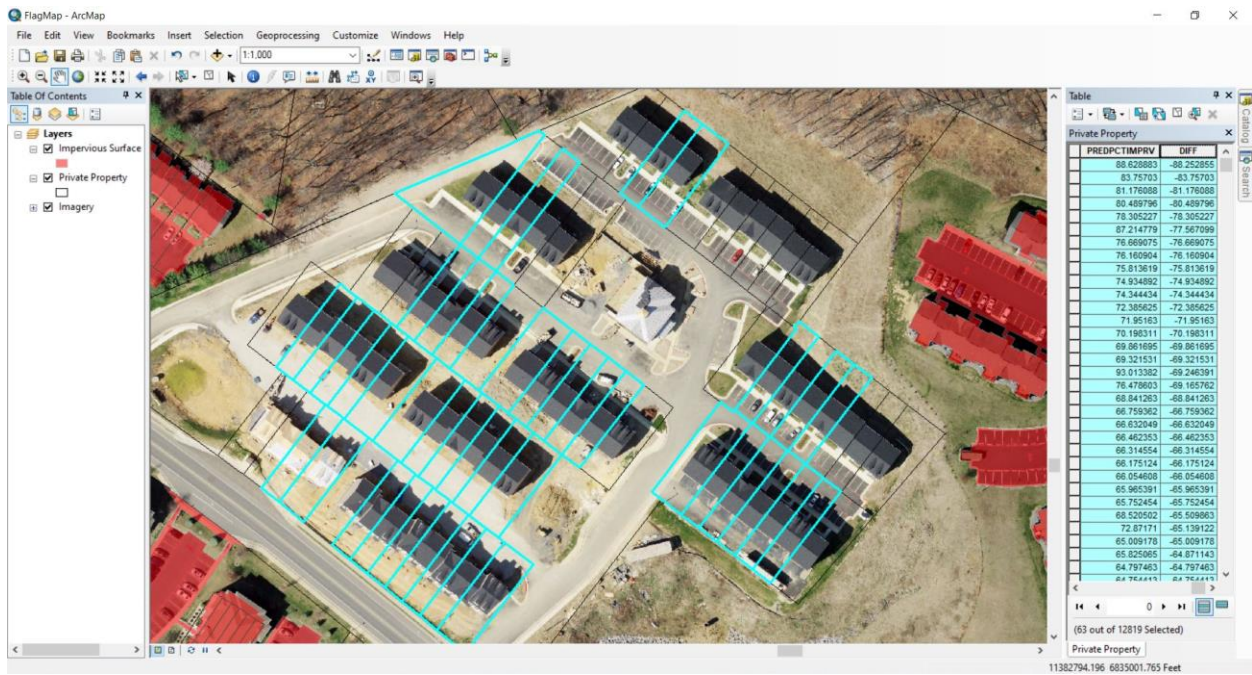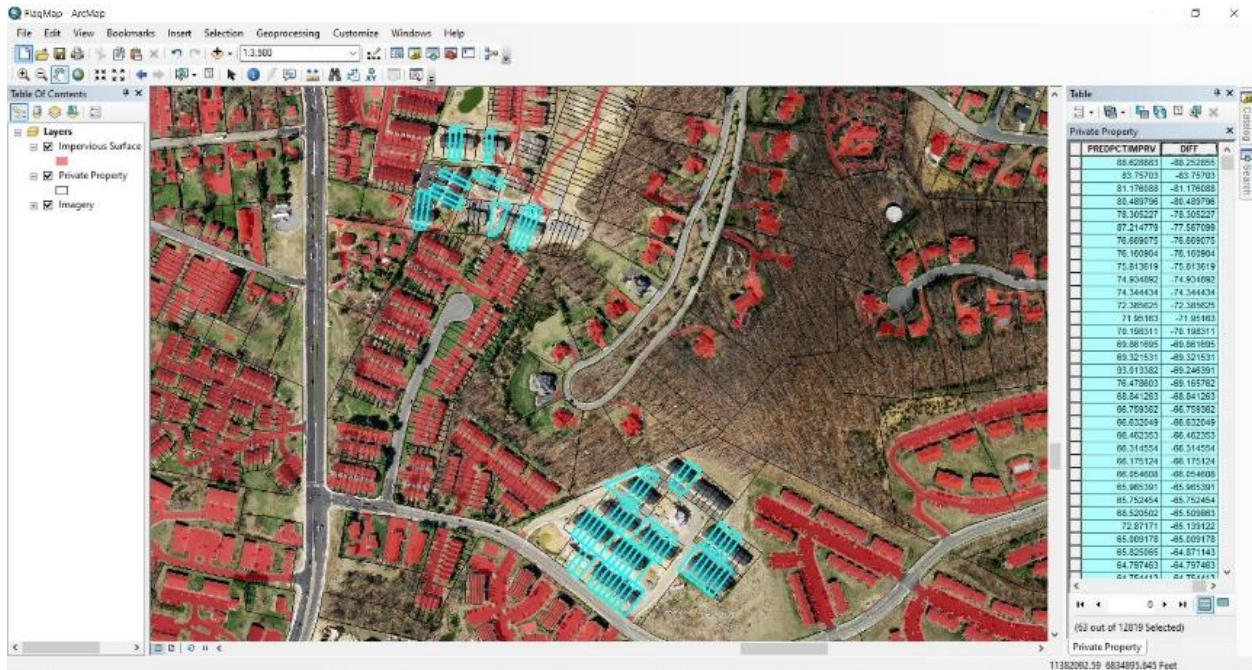
Figure 1.16 Example of flagged parcels. In these images the top 1% of tax parcels (63 parcels) have been flagged based on the difference value. The top image shows two newly developed housing communities where a majority of the new buildings have been flagged. The bottom image shows a zoomed in photo of the southern community. Black lined polygons are tax parcels, red polygons are 2016 impervious surface, and blue lined polygons are flagged tax parcels.

Project Results:

The data used for this project were obtained from two sources. The tax parcel and impervious surface shapefiles were supplied by the City of Harrisonburg and came from their 2016 impervious surface tax map. The aerial imagery was downloaded from the Virginia Base Map Program (VBMP). The VBMP supplies the most recent free aerial imagery of Virginia for download. At the time of completing the project, the downloadable imagery available for Harrisonburg was collected in the year 2018 at a resolution of .25 ft which was resampled to 1ft.

The following features were extracted from GeoDMA to be used during the stepwise regression. Each feature was extracted in the three color bands (red, green, and blue) of the imagery.

- Amplitude: The maximum pixel value minus the minimum pixel value within a tax parcel

- Dissimilarity: Measurement of difference between Grey Level Co-occurrence Matrix (GLCM) elements within a tax parcel (a high value means high contrast)

- Entropy: Measurement of disorder within a tax parcel (when the parcel is not uniform, many GLCM elements have small values, resulting in large entropy)

- Homogeneity: Measurement of similarity between GLCM elements within a tax parcel (a high value means low contrast)

- Mean: The average pixel value within a tax parcel

- Mode: The most occurring pixel value within a tax parcel

- Ratio: This feature is not given a description in the GeoDMA documentation

- Standard Deviation: the sum of standard deviations from the mean of all pixels within a tax parcel

26

- Sum: The sum of all pixel values within a tax parcel

The following features were found to be significant by the stepwise regression (features were added to the model if their p-value was below 0.05 and were removed if their p-value became higher than 0.05):

- Amplitude in green and blue bands

- Dissimilarity in red band

- Entropy in red and blue bands

- Homogeneity in all bands

- Mean in all bands

- Mode in green and blue bands

- Ratio in green and blue bands

- Sum in all bands

The multiple R^2 of the final model was 0.6726 while the adjusted R^2 was 0.6721. So, the independent variables (extracted imagery features) were able to account for ~67% of the variability in the dependent variable (percent imperviousness).

To assess the accuracy of the model and predictions, individual tax parcels were looked at and classified as a true or false positive based on whether the model correctly or incorrectly identified the parcel as needing updating. The following table shows the number of true and false positives for each percent of tax parcels that were flagged.

Parcels Flagged %	Parcels Flagged Total	True Positives at Step	False Positives at Step	True Positives Total	False Positives Total	False Positive Rate at Step	False Positive Rate Total
1%	63	57	6	57	6	9.5%	9.5%
2%	125	47	15	104	21	24.2%	16.8%
3%	188	45	18	149	39	28.6%	20.7%
4%	251	45	18	194	57	28.6%	22.7%
5%	314	27	36	221	93	57.1%	29.6%
6%	376	15	47	236	140	75.8%	37.2%
7%	439	15	48	251	188	76.2%	42.8%

Figure 2.1 Project results table. Note: Only parcels that had a higher 2018 prediction than 2016 estimate were included in the final total as these are the only tax parcels where the city would see an increase in tax revenue from updating. Parcels that had a decrease in impervious surface should be updated at the owners' notification. Therefore, even though there are ~12,000 private tax parcels in Harrisonburg there are only ~6,300 that are predicted to have increased in imperiousness so 1% of these parcels is ~63 tax parcels.

After discussing with the GIS administrator and Environmental Compliance Manager at the City of Harrisonburg these results show that the procedure is beneficial but not perfect. The key metric is the False Positive Rate at Step. This shows that the average increase in false positive percentage for each percent of parcels flagged is 10.88%. So, for every new set of ~63 tax parcels, an additional 10% of them will be false positives, on top of the current percent. This means that of the next ~63 tax parcels flagged after 7%, likely less than 10 will truly need updating. Because of this, the model is useful for identifying tax parcels that need updating up to about the limits that we tested depending on ones' tolerance for searching through false positive results.

Regardless of the imperfection, this is beneficial to Harrisonburg because the parcels that have the newest impervious surface (new homes, buildings, parking lots) are found closer to the top and the parcels that are found and flagged later have relatively less new impervious surface (new shed, walkway, small driveway). Therefore, the model is a good way for Harrisonburg to quickly identify a large amount of new impervious surface.

Project Experience:

Working on this project over the last year and a half has been an overwhelmingly positive experience. Through designing this procedure, I have dealt with numerous obstacles and setbacks, one of which I thought would be fatal to the project. I strengthened my technical skills in using mapping software and coding that will be helpful to me later in my professional and academic careers. Finally, I grew both as a student and professional and gained valuable experience that will help me succeed in my future endeavors.

Throughout the project there were several significant setbacks that necessitated the use of several skills that I have learned and developed while at JMU. Overcoming these challenges was a significant part of the learning process and helped strengthen my skills in diagnosing technical issues with mapping software, dealing with stressful situations quickly and efficiently, and, very importantly, coordinating with multiple stakeholders in person and remotely.

One of the earliest obstacles was an issue with GeoDMA that led the program to crash with every attempt to extract imagery features. The problem was that the file size of the imagery that I was attempting to extract features from was far too large for GeoDMA to handle. I was able to resolve this issue by resampling the imagery to a lower resolution of 0.5 meters per pixel and splitting it into four tiles. I found that imagery files under a gigabyte could be handled by GeoDMA.

A later, more stressful setback involved an inability to join the impervious and tax parcel layers. This was especially stressful due to the fact that for several days I could not figure out why the join was failing. The reason the join was failing was due to some impervious surface features that had geometry but no attributes, and the fact that many impervious surface features

did not perfectly overlap their corresponding tax parcel features. The solution, I found, was to clip the impervious surface layer to remove any overlap and then delete the erroneous features. Doing this allowed the join to work however it did not work perfectly.

The most difficult and stressful setback was the discovery of errors in the spatial join. Just before this was discovered I had already run the stepwise regression and believed that the project was about complete. When the errors were found I had extreme difficulty finding what caused them. It was at this time that my advisor and I made the decision to remove all non-taxable parcels from the tax parcel layer and re-run the spatial join. After removing all land owned by churches, schools, and the state/city governments I re-ran the spatial join. After surveying the resulting joined layer, I found no more errors. This was an especially stressful setback because it took almost two weeks to resolve and involved a significant investment of my time. The largest lesson I learned from this setback is that in the face of stressful situations the best course of action is to remain calm and work as hard as necessary to resolve the issue.

The final major obstacle was that I could not meet with the people at the City of Harrisonburg that I needed to due to COVID-19. The plan was to meet with the key people at the City's GIS office so that I could demo the final product and discuss with everyone the procedure and benefits of the project. The workaround was the coordination of separate meetings where I discussed the project in terms of each person's expertise. This solution ended up working out well and gave me the clarity I needed to complete the project.

Working on this project strengthened my technical skills in working with mapping software like ArcMap, programming in the statistical language R, building a GIS project workflow, and specifying detailed instructions to carry out that workflow. All of these skills will serve me well and help me to succeed in my future pursuits. I also learned how to use the feature

extraction tool GeoDMA which while potentially less useful to my future endeavors was still fun and valuable.

I also strengthened my soft skills in managing setbacks in a stressful, time sensitive environment, navigating self-set and structured deadlines, communicating with stakeholders and an advisor/mentor, and managing my own learning by carrying this project to fruition largely on my own. After graduation I am transitioning immediately into the working world and I am sure that many of these skills, especially communicating with stakeholders, will be beneficial while I am in industry. I do plan on returning to academia in 2022 by pursuing a master's degree in Artificial Intelligence. Many of these skills will be hugely beneficial to me in that pursuit as a large part of my studies will revolve around a research project which I will design with an advisor. This Honors Thesis has given me an excellent tool set to use when I move on in the world after leaving JMU.

Possible Improvements:

After having the procedure reviewed by friends, family, and a GIS professional, many improvements were made to the procedural outline to improve clarity, readability, and detail. I believe that the procedure works well and, after including these improvements, has been outlined effectively. A possible change to the procedure that may yield better results is outlined below.

It would be interesting to see if a model built using aerial imagery from the same year as the past impervious surface map would, in fact, yield a better model and more accurate predictions than using aerial imagery from several years later. If the imagery were to match the vector data perfectly then the imagery features may be able to create a better model. This model could then be used in conjunction with current imagery to create predictions for updating the map. These predictions could lead to a lower false positive rate. The downside to this method is it requires significantly more work as well as two sets of imagery. After the completion of this project I may look into using this method.

My hope is that this procedure will be used by many municipalities across the country. Through continued use new people will be able to contribute new ideas and bring up new concerns. With the input of others, the process can be improved upon and can provide better assistance to those municipalities that use the methodology. I will continue to offer my time and assistance to promote the use of this procedure, keep the process up to date, and provide guidance and feedback to any groups that implement the process. Contact information for myself and my advising professor can be found in the *Resources* section.

Conclusions:

The goal of this project was to devise, carryout, and outline a procedure for updating impervious surface maps that any municipality, with access to aerial imagery and a past impervious surface map, could implement to save the update effort time and money. After speaking to key people at the City of Harrisonburg at the conclusion of the project, that goal has been met. While the city uses building permits to update its impervious surface tax billing system, this procedure will be helpful to update their tax map.

Those I spoke with at the city of Harrisonburg believe this procedure will save the city money, by allowing them to update their maps on a consistent basis without purchasing expensive LiDAR data, and time, by highlighting tax parcels with a high likelihood of needing updating. Saving time and money will allow the city to put those valuable resources towards other beneficial projects. Over time, as more municipalities hopefully make use of this procedure, the process will be able to save those municipalities time and money as well.

R, free language and environment for statistical computing and graphics, download link:
http://lib.stat.cmu.edu/R/CRAN/
Download and installation instructions:
- Select download R 3.6.3 link on webpage and select a location to save the installer
- Run the installer and follow prompts to install the software (default settings are fine)
- R can be run by searching for R x64 in computer applications and running the app

Custom R script for stepwise regression (R is required to run this script):
Instructions for running the script below:
- Once R has been downloaded, run R using the above instructions
- In the R environment, open a new script by selecting 'File' > 'New script'
- Copy the code provided in the box below (runs through multiple pages) and paste into the script window
- To run, select all code in the script window and press 'ctrl+r'
- Follow the instructions in the provided code prior to running

```
## This is a script for running stepwise regression using data from excel.
## This script can handle many independent variables.
## During program run, users will be prompted for input four times.
## For details, please read instructions below labeled IMPORTANT.

## IMPORTANT: The packages required for this script are: 'openxlsx',
## 'dplyr', 'pillar'. If you already have any of these packages,
## comment out the 'install' line(s) below before running the script.
## Any mirror should work for install.

install.packages("openxlsx")
install.packages("dplyr")
install.packages("pillar")
install.packages("svDialogs")

library("openxlsx")      # package used to read excel file
library(MASS)       # package used to run stepwise regression
library(dplyr)            # package used to convert dataframe rows to numeric
library(svDialogs)  # package used to gather user input

# read in excel file and user input for cols parameter and excel file
cols_start <- dlgInput("Enter start column as number:", Sys.info()["user"])$res
cols_end <- dlgInput("Enter end column as number:", Sys.info()["user"])$res
in_data <- read.xlsx(file.choose(), sheet = 1, cols = cols_start:cols_end)

# convert dataframe rows to numeric type
new_data <- mutate_all(in_data, function(x) as.numeric(as.character(x)))

# ask for name of DV do be used by the program
depend_var <- dlgInput("Enter dependant variable as it appears in excel doc:",
Sys.info()["user"])$res

# create general model including all variables for stepwise regression
# if the 'variable lenghts differ' error appears note:
```

```
# this should not impact algorithm performance
full_model <- lm(depend_var ~., new_data)

# run stepwise regression
step_model <- stepAIC(full_model, direction = "both", trace = FALSE)

summary(step_model) # summary output from stepwise regression

## The following code uses the regression coefficients and the
## dataframe observations for associated variables to calculate predicted
## impervious surface density for each tax parcel. These predictions
## are then output to a .csv file. Read instructions below.

# save coefficients from regression for use in calculating predictions
coefs <- coef(step.model)

# save names of variables corresponding to coefficients for use in
# removing unimportant data from dataframe
coef_names <- names(coefs)[-1]

# remove unimportant data from dataframe
newdata <- data[coef_names]

preds <- c()                # vector of predictions (output as .csv)
xlength <- NCOL(newdata)    # number of columns in the dataframe
                                  # used to bound for loop running individual
                                  # predictions
ylength <- NROW(newdata)    # number of rows in the dataframe used to
                                  # bound for loop running total predictions

# outer for loop runs once for each tax parcel in the dataframe
for (y in 1:ylength)
{
row <- newdata[y,]  # gets the row of variables for the current tax parcel
pred <- coefs[[1]]  # incorporates the constant into the prediction calculation

# inner for loop runs once for each variable in the tax parcel, this is where
# the prediction calculation happens
for (x in 1:xlength)
{
pred <- pred + (row[[x]] * coefs[[x + 1]])     # adds the product of the current
                                                 # variable and coefficient to
the
                                                 # current total
}
pred <- c(pred)          # set the prediction for the current tax parcel to a
vector
preds <- append(preds, pred, after = y) # add the current prediction to the list
}

# IMPORTANT: at runtime choose a place to save the .csv file, give it a
# name - which must end in '.csv' - and select 'Open'. On the
# prompt: 'This file doesn't exist. Create the file?' select 'Yes'

# write predictions as output to a user specified csv file
write.csv(preds, file.choose())
```

GeoDMA, free geographic/imagery feature extraction and data mining tool, download link:
https://sourceforge.net/projects/geodma/
Download instructions:

- Select the download link on the webpage and select a location to save the installer
- Run the installer and follow prompts to install the software
- GeoDMA is a plugin for TerraView. TerraView can be opened by running the 'terraview' application in the installed 'GeoDMA' folder

VBMP, Virginia Base Map Program free imagery download link:
https://vgin.maps.arcgis.com/apps/Viewer/index.html?appid=cbe6a0c1b2c440168e228ee33b89c
b38

Contact info:

Author: Patrick Muradaz – muradapd@dukes.jmu.edu – patrickmz27@gmail.com
Advising Professor: Dr. Zachary Bortolot – bortolzj@jmu.edu

Bibliography:

Alan H. Strahler, Curtis E. Woodcock, & James A. Smith. (1986). On the nature of models in remote sensing. Remote Sensing of Environment. Volume 20, Pages 121-139, DOI: https://doi.org/10.1016/0034-4257(86)90018-0

Andrew P. Tewkesbury, Alexis J. Comber, Nicholas J. Tate, Alistair Lamb, & Peter F. Fisher. (2015). A critical synthesis of remotely sensed optical image change detection techniques, Remote Sensing and Environment, Volume 160, Pages 1-14. DOI: https://doi.org/10.1016/j.rse.2015.01.006

Cholyoung Lee, Kyehyun Kim, & Hyuk Lee. (2018). GIS based optimal impervious surface map generation using various spatial data for urban nonpoint source management, Journal of Environmental Management, Volume 206, Pages 587-601, DOI: https://doi.org/10.1016/j.jenvman.2017.10.076

Chunzhu Wei, & Thomas Blaschke. (2018). Pixel-Wise vs. Object-Based Impervious Surface Analysis from Remote Sensing: Correlations with Land Surface Temperature and Population Density. Urban Science, volume 2, DOI:10.3390/urbansci2010002

Devis Tuia, Frederic Ratle, Fabio Pacifici, Mikhail F. Kanevski, & William J. Emery. (2009). Active Learning Methods for Remote Sensing Image Classification. IEEE Transactions on Geoscience and Remote Sensing, Volume 47, Pages 2218-2232. DOI: 10.1109/TGRS.2008.2010404

E. Terrence Slonecker , David B. Jennings & Donald Garofalo. (2001). Remote sensing of impervious surfaces: A review, Remote Sensing Reviews, 20:3, 227-255, DOI: 10.1080/02757250109532436

Hui Luo, Le Wang, Chen Wu, & Lei Zhang. (2018). An Improved Method for Impervious Surface Mapping Incorporating LiDAR Data and High-Resolution Imagery at Different Acquisition Times, Remote Sensing, Volume 10, Page 1349. DOI: 10.3390/rs10091349

Jin Wang, Zhifeng Wu, Changshan Wu, Zheng Cao, Wei Fan, & Paolo Tarolli. (2018). Improving impervious surface estimation: an integrated method of classification and regression trees (CART) and linear spectral mixture analysis (LSMA) based on error analysis. GIScience & Remote Sensing, Volume 55, Number 4, Pages 583-603. DOI: 10.1080/15481603.2017.1417690

K.M.S. Sharma & A. Sarkar. (1998). A Modified Contextual Classification Technique for Remote Sensing Data. Photogrammetric Engineering and Remote Sensing, Volume 64, Pages 273-280

Marvin E. Bauer, Brian C. Loffelholz, & Bruce Wilson. "Estimating and Mapping Impervious Surface Area by Regression Analysis of Landsat Imagery", *Remote Sensing of Impervious Surfaces,* 1st edition, CRC Press, 2007. Preprint of chapter.

Quang-Thanh Bui. Manh Pham Van, Nguyen Thi Thuy Hang, Quoc-Huy Nguyen, Nguyen Xuan Linh, Pham Minh Hai, Tran Anh Tuan & Pham Van Cu, (2018) Hybrid model to optimize object-based land cover classification by meta-heuristic algorithm: an example for supporting urban management in Ha Noi, Viet Nam, International Journal of Digital Earth, DOI: 10.1080/17538947.2018.1542039

Qihao Weng. (2012) Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends, Remote Sensing of Environment, Volume 117, Pages 34-49, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2011.02.030.
Stormwater Utility Information. (2017, April 17). City of Harrisonburg, VA. Retrieved from: www.harrisonburgva.gov/stormwater-utility

Thales Körting, Leila Fonseca, & Gilberto Câmara. (2013). GeoDMA–geographic data mining analyst. Computers & Geosciences. 57. 133–145. 10.1016/j.cageo.2013.02.007.