# IS EVERYTHING ALRIGHT WITH OUR WATER? DISTRIBUTION OF CANCER AND NITRATE CONCENTRATION IN WISCONSIN WELLS

Jason Heutmaker

[redacted]

[redacted]

USA

Running title: Cancer Distribution Based on Nitrate Concentration

[1]Received _____ Accepted _____

Correspondence: jheutmaker@wisc.edu

*Abstract*

There has been a steady increase in nitrate levels within water in many locations globally which has been largely attributed to applications of fertilizer and manure in agricultural areas. Currently, the laws that have been set in place limit the amount of nitrate in drinking water supplies in order to protect against specific instances of infant methemoglobinemia. However, risk of specific cancers and birth defects were not considered when creating these limits. Countless studies have been done to show that there is a direct correlation between drinking water with higher levels of nitrate and various kinds of cancer and other adverse health outcomes. In an attempt to see if there is such a correlation in Wisconsin a python program was developed to compare instances of cancer and nitrate concentrations in well water.

*Introduction*

What exactly is "nitrate"? Nitrate is something that forms when ammonia and other high concentrations of nitrogen sources combine with oxygen to create a water soluble-molecule. For nitrate to be considered safe within groundwater it generally must be below two parts per million. However, since the early 20$^{th}$ century humans have increased the rate at which nitrogen is placed back into the land at almost twice the rate than ever before through the production and application of fertilizers, fossil fuels, and nitrogen fixing crops like soybeans. With about fifty-percent of all nitrogen that is applied to agricultural land draining into and contaminating surface as well as groundwater we are experiencing increases of overall nitrate concentrations in public drinking water including well water. Back in 2012 a survey taking a look at the drinking water systems within Wisconsin found that a staggering forty-seven systems had samples that went well past the state nitrate limit which had increases from only fourteen back in 1999. This means that wells which currently provide water to over 1/3$^{rd}$ of residencies in Wisconsin are at risk,

potentially with the inhabitants being unaware. It's our hope that the simple, easy to use, and informative application that was created can help researchers and GIS specialists make decisions and implement policies to make the public more aware of the potential threat to themselves and their loved ones.

*Libraries and Code*

In order to create a program that is simple, gets the job done, and is informative requires code that works and makes sense for the users it's been made for. With the thought process that this program would be utilized by individuals and specialists who already understand what OLS (Ordinary Least Squares) and IDW (Inverse Distance Weighted) are more background code and libraries were used than what would normally seem necessary. A full list of the libraries used is as follows: OS, ArcPy, Sys, ArcPy Env, ArcPy.sa *,Tkinter, Tkinter *, Tkinter- tkk and messagebox, PIL- ImageTk and Image, Glob, and Webbrowser. The basic functionality of each library and why it was chosen: 1) OS: Is required for all interactions with the program and the operating system. Without it the program would not run. 2) ArcPy: Is required for all interactions with ArcMap and ArcPro. 3) Sys: Provides information about constants, functions, and methods that are needed for the program to run smoothly. 4) Arcpy.sa *: Provides spatial analyst classes that were not included in ArcPy. 5) Tkinter: Is the backbone for the TK GUI toolkit. 6) Tkinter *: Imports every exposed object into the GUI. 7) Tkinter tkk & messagebox: Is a submodule that allows for message boxes to appear from the GUI and replaces certain elements of Tkinter with ttk. 8) PIL ImageTk and Image: Another submodule that allows for better image use within Tkinter. 9) Glob: Finds pathways based on patterns and readily allows them to be deleted and avoid overlap. 10) Webbrowser: Allows for easy display of web-based documents and activating hyperlinks. Example 1: A perfect section of the code using multiple libraries (OS- Red, Glob- Blue, ArcPy- Green, Tkinter- Black, Tkinter MessageBox- Pink) in conjunction with one another is shown below.

```
try:
    directory='C:/GEOG777/Cancer_Analysis/Images'
    os.chdir(directory)
    files=glob.glob('County_*')
    for filename in files:
        os.unlink(filename)

    #Morans I Tool
    moransI = arcpy.SpatialAutocorrelation_stats(r"C:\GEOG777\Cancer_Analysis\outputFiles\olsResults.shp","Residual",
                                                 "GENERATE_REPORT", "INVERSE_DISTANCE", "EUCLIDEAN_DISTANCE", "NONE", "","")

    #Setting all buttons to normal
    displayOLS.config(state=NORMAL)
    webButton.config(state=NORMAL)
    countyCheckbutton.config(state=NORMAL)

    #Let user know that process is complete.
    messagebox.showinfo("Final Process Completed", "Final Process Completed, select the results you'd like to review below!")
except:
    #An error to inform the user that an issue occurred with running Morans I;
    messagebox.showwarning("Couldn't Delete", "Check to make sure you can delete old images or if the Morans I configuration is correct.")
    print("Check to make sure you can delete old images or if the Morans I configuration is correct..")
```

Each of these libraries plays an important role in creating, displaying, and keeping the number of files produced to a minimum amount. Almost as important is their ability to function with Python 3.6 and ArcPro.

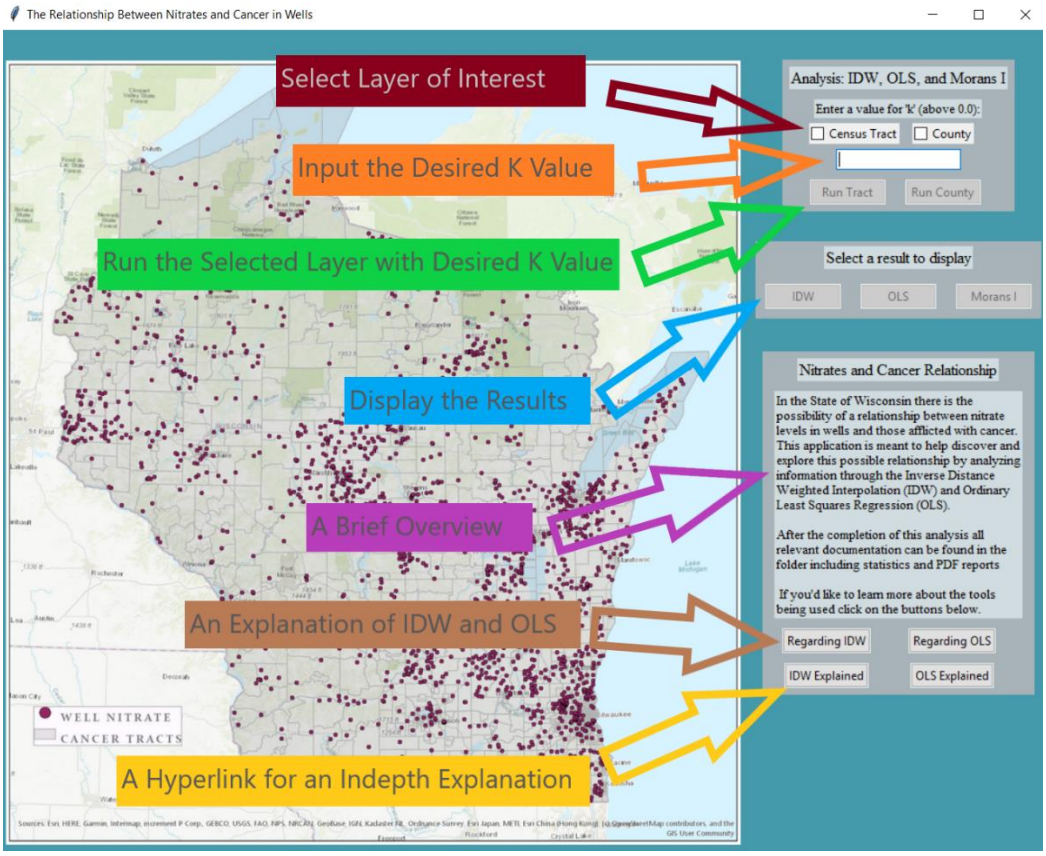
The code was constructed in conjunction with ArcPro instead of ArcMap for a few keys reasons: Python 3.6 allows for better processing speeds and compatibility with other software, ArcPro is becoming a more dominate force in more fields over its predecessor ArcMap, and because ArcPro has more capabilities than ArcMap which is limited to the Python 2.7 version. Since those who utilize ArcPro are more likely to be experienced users and may potentially wish to edit it for their own purposes a systematic approach of “Try”’ and “Except” was implemented in the Python code. Should a user edit the code in a fashion that makes it unable to correctly create an IDW or OLS display an error message informing the user will let them know approximately where in the code the likely error occurred. Example 2: The program will first attempt to run the “Try” portion of the code (circled in red). Should either one of the two processes highlighted (directory or moransI) fail to complete than the “Except” portion of the code will run instead (circled in blue). Where it will output a message telling the user that a document was unable to be deleted/replaced or that Morans I was unable to be processed.

```
try:
    directory='C:/GEOG777/Cancer_Analysis/Images'
    os.chdir(directory)
    files=glob.glob('County_*')
    for filename in files:
        os.unlink(filename)

    #Morans I Tool
    moransI = arcpy.SpatialAutocorrelation_stats(r"C:\GEOG777\Cancer_Analysis\outputFiles\olsResults.shp","Residual",
                                                 "GENERATE_REPORT", "INVERSE_DISTANCE", "EUCLIDEAN_DISTANCE", "NONE", "","")

    #Setting all buttons to normal
    displayOLS.config(state=NORMAL)
    webButton.config(state=NORMAL)
    countyCheckbutton.config(state=NORMAL)

    #Let user know that process is complete.
    messagebox.showinfo("Final Process Completed", "Final Process Completed, select the results you'd like to review below!")

except:
    #An error to inform the user that an issue occurred with running Morans I;
    messagebox.showwarning("Couldn't Delete", "Check to make sure you can delete old images or if the Morans I configuration is correct.")
    print("Check to make sure you can delete old images or if the Morans I configuration is correct..")
```

This type of formatting helps protect potentially developers from accidentally making a piece of the code inoperable. While it takes up more space and creates a slightly longer run time it can save massive amounts of time and frustration in the long run.

*Visual Display*

Just as if not more important than good code and its documentation is a visual display geared towards the type of user in mind. The main aspects of the GUI are broken down and explained below:



Overall Presentation: Normally, an explanation of the project at hand and a closer look into the processes being used (labeled as a Brief Overview, An Explanation of IDW and OLS, and A Hyperlink for an In-depth Explanation) would be placed at the top of the page for the user to gain some insight into what they were using. However, since this program was created with experienced users in mind these details, while still important, are placed below so as not to detract from the main purpose of program and allow experienced users to skip things they are already familiar with.

Select Layer of Interest: This section allows the user to select whether or not they would like to use the Census Tract or the County Outline for the state of Wisconsin. When pressed it grays out the unselected one and will on allow for a display of the selected layer.

Input Desired K Value: Allows the user to place any number larger than 0 into the program and run.

Run Selected Layer with Desired K Value: As the name suggests this button runs the main process.

Display the Results: A myriad of results are able to be displayed such as IDW, OLS, and Morans I.

A Brief Overview: Similar to the introduction of this paper, a brief overview of the program and project is provided.

An Explanation of IDW and OLS: A pop-up message is displayed giving a brief reminder as to what each thing means and how it can be used.

A Hyperlink for an In-Depth Explanation: An internet browser is opened that displays in-depth information about each of the main processes just in case the user needs a helping hand.

*Analysis Methods*

There are three main analysis methods that are utilized in this program: IDW (Inverse Distance Weighted), OLS (Ordinary Least Squares) and Morans I. IDW is the most widely used deterministic interpolation tool and predicts a value for any unmeasured location by measuring the values that surround the prediction point. It assumes that each point has an influence that in turn diminishes with distance and in turn weighs the points closer to the location greater than those that are further away.

For our project a map is outputted showing the IDW for the entire area in and surrounding Wisconsin. OLS is a form of simple linear regression that uses the strategy of attempting to create a “straight line” which matches as closely as possible to our actual data points. It is extremely popular since the outputs of the regression are unbiased estimators of the real values even though it is not the only optimization strategy out there. For our project a map is outputted showing the OLS strictly in relation to either the Census Tracts or the County Outlines of Wisconsin. Morans I is another correlation coefficient that measures the overall spatial autocorrelation of our data set except in this case it is multidirectional and multidimensional. It uses weighted matrix to place variables into three separate categories: clustering, random, and dispersion. In this case a map is not outputted, but instead a hyperlink that displays the different p-values, z-scores, and most likely category for the data being presented.

*Data Analysis and Discussion*

Two key layers are the cornerstone of the data being looked at, the census tracts and the county outlines, each of them expressing very different things. The census tracts showed an extreme case of clustering where an increase in nitrates near wells heavily correlated to the cancer rates around it. The county outline on the other hand leaned to an almost perfect randomness with very little overlay between nitrate concentration and cancer. In this particular case we are more inclined to believe that the census tracts yields a more accurate result than the county outline. This is due to the fact that area we are looking, the single state of Wisconsin, is relatively too small to find an accurate representation from such large county outlines. In certain instances, such as diseases like cancer, it is better practice to have small areas to compare against each other than relatively large ones. There are of course still flaws in the general details and map being displayed by the census tracts as well including all cancer types being accounted for instead of cancers specifically related to increased nitrate consumption. That being said, the map draws a clear picture of the drastic state of Wisconsin’s drinking water and the potential health crises currently at play in the state.