Audit Quality: Detection of Material Misstatement

Megan Jones

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
https://researchrepository.wvu.edu/etd/5906

This Dissertation is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Graduate Theses, Dissertations, and Problem Reports by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
Modeling Phosphorus Environmental Thresholds on a Watershed Scale

Sara L. Jones

Thesis submitted to the
Davis College of Agriculture, Natural Resources, and Design
at West Virginia University

In partial fulfillment of the requirements for the degree of

Master of Science in Agronomy

James A. Thompson, Ph.D., Chair
Louis M. McDonald, Ph.D.
Michael P Strager, Ph.D.

Division of Plant and Soil Sciences

Morgantown, West Virginia

2015

keywords: soil, phosphorus, change point, spatial modeling
ABSTRACT

Modeling Phosphorus Environmental Thresholds on a Watershed Scale

Sara L. Jones

Anthropogenic eutrophication of surface waterways is a major environmental concern as it leads to the decline of aquatic resources and degradation of water. Phosphorus is the most limiting nutrient in freshwater systems and, therefore, an important factor in the control of eutrophication. Identifying the natural P thresholds in soils using the relationship between soil P and potential P losses can help determine areas with a high risk of P loss to waterways via leaching. Twenty-three sites covering a range of land uses were sampled by the USDA-NRCS in the Potomac Watershed in the northeast United States. Four pedons described and sampled at each site were used in this analysis. Samples from horizons in the plow layer (0 to 15 cm) were analyzed and Mehlich 3 P and water extractable P determined. A threshold in the relationship between CaCl$_2$ P and Mehlich 3 P was identified using segmented linear regression (change point analysis) for each land use type. The change point identified for forest lands was 42 mg P kg$^{-1}$, and for agricultural lands it was 64 mg P kg$^{-1}$. For the pooled land uses, the change point was 66 mg P kg$^{-1}$. The change point relates to the P level in soil above which an increase in soil test P is expected to lead to a large rise in P loss to waterways. Using environmental covariates such as soil pH, parent material, and elevation, a prediction surface was generated to predict Mehlich 3 P and Mehlich 3 P sorption ratio levels using various kriging techniques. The modeled surface can be used by land managers as additional information to help prevent P losses that lead to eutrophication.
**List of Figures**

1.1 The P Cycle 43
1.2 Effect of pH on P speciation 44
1.3 Correlation between P saturation measurements using Mehlich 3 and ammonium oxalate extractions 46
1.4 Parts of the semivariogram 47
2.1 The Potomac watershed 49
2.2 Location of the study area 50
3.1 Correlation between Mehlich 3 P and environmental covariates 54
3.2 Correlation between CaCl$_2$ P and environmental covariates 55
3.3 Change point analysis for forest land use 56
3.4 Change point analysis for agricultural land use 57
3.5 Change point analysis for pooled land uses 58
3.6 Linear regression analysis for forest land use 59
3.7 Linear regression analysis for agricultural land use 60
3.8 Linear regression analysis for pooled land uses 61
3.9 Correlation between M3-PSR and environmental covariates 63
3.10 Relationship between M3-PSR and Mehlich 3 P 64
3.11 Relationship between CaCl$_2$ P and Mehlich 3 P 65
3.12 Mehlich 3 P regression tree analysis 66
3.13 Relationship between observed Mehlich 3 P values and expected Mehlich 3 P values 67
3.14 Relationship between observed M3-PSR values and expected M3-PSR values 68
3.15 Regression tree analysis of M3-PSR 69
3.16 Histogram of Mehlich 3 P 71
3.17 Histogram of logarithmic transformed Mehlich 3 P 72
3.18 Histogram of M3-PSR 73
3.19 Histogram of logarithmic transformed M3-PSR 74
3.20 Normal Q/Q plot of Mehlich 3 P 75
3.21 Normal Q/Q plot of logarithmic transformed Mehlich 3 P 76
3.22 Normal Q/Q plot of M3-PSR 77
| 3.23 | Normal Q/Q plot of logarithmic transformed M3-PSR | 78 |
| 3.24 | Mehlich 3 P prediction surface from co-kriging | 80 |
| 3.25 | Mehlich 3 P prediction surface from ordinary kriging | 81 |
| 3.26 | Distribution of the probability that a site will exceed the change point 42 mg P kg\(^{-1}\) | 82 |
| 3.27 | Distribution of the probability that a site will exceed the change point 64 mg P kg\(^{-1}\) | 83 |
| 3.28 | Distribution of the probability that a site will exceed the change point 66 mg P kg\(^{-1}\) | 84 |
| 3.29 | M3-PSR prediction surface | 85 |
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Summary of published change point analysis</td>
<td>45</td>
</tr>
<tr>
<td>1.2</td>
<td>STP environmental covariates</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>Five number summary of agricultural sites</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>Five number summary of forested sites</td>
<td>52</td>
</tr>
<tr>
<td>3.3</td>
<td>Five number summary of all sites</td>
<td>53</td>
</tr>
<tr>
<td>3.4</td>
<td>Summary of regression models used to determine relationship between Mehlich 3 P and CaCl$_2$ P</td>
<td>62</td>
</tr>
<tr>
<td>3.5</td>
<td>Moran’s I Index analysis output</td>
<td>70</td>
</tr>
<tr>
<td>3.6</td>
<td>Summary of interpolation methods used in study</td>
<td>79</td>
</tr>
</tbody>
</table>
# Table of Contents

Chapter 1. Introduction ................................................................................................................................. 7

1.1 Phosphorus .......................................................................................................................................... 7

1.2 Environmental Thresholds .................................................................................................................. 9

1.3 Degree of Phosphorus Saturation ...................................................................................................... 11

1.4 Mapping Soil Phosphorus ................................................................................................................. 11

   1.4.1 \textit{scorpan} Modeling ...................................................................................................................... 12

   1.4.2 Kriging ....................................................................................................................................... 13

   1.4.3 Co-Kriging ................................................................................................................................. 14

   1.4.4 Indicator Kriging ........................................................................................................................ 15

   1.4.5 Regression Tree Modeling ......................................................................................................... 15

Chapter 2. Materials and Methods .............................................................................................................. 17

2.1 Study Area and Sample Collection ................................................................................................... 17

2.2 Determination of Soil Chemical Properties ...................................................................................... 18

2.3 Covariates Influencing STP .............................................................................................................. 19

2.4 Statistical Analyses ........................................................................................................................... 21

2.5 Kriging .............................................................................................................................................. 23

Chapter 3. Results and Discussion .............................................................................................................. 24

3.1 Soil Chemical Properties ................................................................................................................... 24

3.2 Change Point Analysis ...................................................................................................................... 25

3.3 Phosphorus Sorption Ratio ................................................................................................................ 27

3.4 Regression Tree Analysis .................................................................................................................. 29

3.5 Geospatial Analysis .......................................................................................................................... 30

Chapter 4. Conclusions ............................................................................................................................... 35

Chapter 5. Implications of Mapping Soil Phosphorus ................................................................................ 38

   5.1 Soil Phosphorus ............................................................................................................................... 38

   5.2 Modeling Soil Phosphorus ............................................................................................................. 39

   5.3 Future Analysis ............................................................................................................................... 41

Tables and Figures ...................................................................................................................................... 43

References ................................................................................................................................................... 86
Chapter 1. Introduction

1.1 Phosphorus

Phosphorus (P) is an essential nutrient for crops. Plants use P to build and maintain cell membranes and DNA (Gruek et al., 2000). About 90% of the P in the soil-plant-animal system is found in the soil (Ozanne, 1980). Within the soil, P can be separated into three fractions: an organic fraction, a soluble inorganic fraction that can be absorbed by plants, and an insoluble inorganic fraction (Ozanne, 1980). Up to 65% of the P in soils may be stored in the organic matter fraction, which is derived mainly from plant residues (Frossard et al., 1995). The highest ratios of organic P to inorganic P occur in the surface horizons, with a decrease down the profile (Anderson, 1980).

In agricultural soils, the amount of P in the soil solution ranges from 0.01 to 3.0 mg P L\(^{-1}\) (Frossard et al., 2000). This represents only a small fraction of plant needs. The remainder must be obtained through a variety of abiotic and biotic processes that transform P into a form available to plants (Figure 1.1). The major processes involved in P transformation in soil are precipitation-dissolution and adsorption-desorption. These processes control the abiotic transformation of P between the solid phase and the soil solution. Controlling the biotic transformations of P between inorganic and organic forms are immobilization and mineralization processes.

Inorganic P is made unavailable to plants by binding with aluminum (Al), calcium (Ca) and iron (Fe) to form insoluble compounds. The immobilization of inorganic P and the unavailability of organic P contribute to the low levels of P in the soil solution. In fact, only about 20% of the P fertilizer applied is taken up by the first year crop (Vance, 2001). When fertilizers containing P are added to the soil and dissolved by water, a fraction of the P will be released into the soil solution, readily available for plant uptake.

Inorganic P also enters the soil solution via dissolution or desorption from soil solids. The ionic form is controlled by the pH (Pierzynski et al., 1994). Most agricultural crops require a pH between 5.5 and 7.5 for optimal growth. At this pH range, P is found in the monovalent (H\(_2\)PO\(_4\)) or divalent (HPO\(_4^{2-}\))
form (Figure 1.2), both of which can easily be taken up by plants. Inorganic P that is not available for plant uptake is susceptible to loss via leaching or runoff.

In the soil, P is frequently found in an insoluble form, making it unavailable for plant uptake and limiting the amount of P that plants can use. In addition, P is quickly fixed into unavailable forms when it is added as an amendment to the soil (Sanyal and DeDatta, 1991). For this reason, it has become common for farmers to apply more P than necessary to meet the needs of the crops (Gbruek et al., 2000). This excess in applied P can end up in water systems via runoff, leaching, and erosion.

P leaching is often overlooked as an important environmental factor due to the sorption capacity of soils for P (Pierzynski et al., 1994). However, preferential flow has been shown to promote translocation of P to subsurface horizons and possibly to shallow groundwater (Addiscott et al., 2000; Harman et al., 2013; Sims et al., 1998; Sinaj et al., 2002). In contrast, leaching of nitrogen in the form of nitrate is viewed as a critical environmental issue. This is primarily due to the fact that the only negative environmental effect of P is eutrophication while nitrate can be hazardous to human and animal health in addition to contributing to eutrophication (McLaren and Cameron, 1996).

Water quality concerns arise because only a small amount of excess P can lead to eutrophication. Eutrophication is defined as the process by which a body of water acquires a high concentration of nutrients, especially phosphates and nitrates (Art, 1993). The introduction of excess nutrients fuels the rapid growth of aquatic plants and algae. As these plants die and decompose, the dissolved oxygen in the water is depleted (hypoxia). This leads to the degradation of water for human consumption and recreation health as well as the decline of aquatic communities (Kleinman et al., 2011). Phosphorus (P) loss into surface waterways exacerbates the problem of eutrophication. Often, poor water quality limits water uses like supporting human, plant, and animal life and to providing a source of recreation (USEPA, 1996).

Sharpley et al. (1996) determined that the concentrations of P that lead to eutrophication ranges from 0.01 to 0.03 mg L\(^{-1}\). This low eutrophication threshold means that even small amounts of excess P in the soil can lead to large problems in the watershed. For example, a rain event that results in surface and
subsurface water flows of 10 cm ha\(^{-1}\) only needs to dissolve 10 g of P per ha to result in a concentration of 0.01 mg L\(^{-1}\), a threshold for eutrophication (Taylor and Kilmer, 1980).

In order to control the level of eutrophication in surface waters nationwide, the United States Environmental Protection Agency (USEPA) has set water quality guidelines required by the Clean Water Act. In streams, an upper limit of 0.1 mg P L\(^{-1}\) is suggested while this limit in lakes is 0.05 mg P L\(^{-1}\) (USEPA, 1986). Above these limits, water quality is susceptible to eutrophication and hypoxia that can be lethal to multiple aquatic organisms. Developing P concentration guidelines in soils is more difficult due to the fact that very little is known about the paths the P takes in the soil. Even though P concentrations in soils exceed the water quality guidelines, the P can be altered or adsorbed before it reaches surface waters (Maguire and Sims, 2002a). Therefore, it is important to understand the soil interactions of P in order to set guidelines to protect water quality.

**1.2 Environmental Thresholds**

While there has been much study on the movement of P via surface pathways, little attention has been given to P losses via subsurface pathways and processes (Maguire and Sims, 2002a). It has been shown that certain environmental relationships are well correlated to P concentrations in runoff waters and subsurface drainage (Heckrath et al., 1995; Kleinman et al., 2000; Pote et al., 1999).

The relationship between the total amount of P in the soil and the amount of plant available P can be used to identify certain thresholds. The thresholds in these relationships can be identified using segmented linear regression. These thresholds are referred to as *change points* (Table 1.1). Above the change point, an increase in soil test P (STP) will lead to a large rise in P concentrations in runoff. Below the change point, additional input of P will not cause a significant change in the P concentrations in effluent (Kleinman et al., 2000).

Heckrath et al. (1995) first studied the existence of this threshold by relating STP to P concentrations found in pipe drains containing water flow from 20 plots treated with fertilizer. They used a Split-Line Model to determine the change point for different drainage events. Hesketh and Brookes (2000) also identified a change point in the relationship between STP (using an Olsen extraction) and
calcium chloride (CaCl$_2$) extractable P (CaCl$_2$ P). They concluded that, using the change point, the measurements of STP could be used to indicate the potential risk of P leaching from the soil into water.

A change point has also been identified in the relationship between CaCl$_2$ P and STP by Kleinman et al. (2000). STP was determined using two different agronomic soil tests, Morgan’s P and Mehlich 3 P. The change point in both relationships occurred at a CaCl$_2$ P value of 0.9 mg kg$^{-1}$. Kleinman et al. (2000) suggest that this consistency indicates that the change point is related to the behavior of soil P chemistry, regardless of the P extraction method. A change point was also identified in the relationship between CaCl$_2$ P and P sorption saturation at 0.9 mg kg$^{-1}$.

McDowell et al. (2001) determined that the change point is inversely related to the pH. This was attributed to the greater P sorptive potential at lower pH. Specifically, P is bound to Al and Fe oxides in the soil more tightly at lower pH due to the lower amount of charge and electrostatic potential sites for P sorption (Sanyal et al., 1993). This reduces the potential for P loss, which decreases the change point value.

A change point can also be determined from the relationship between STP and water-extractable P (McDowell and Sharpley, 2001). This change point is different from the change point determined using the relationship between STP and CaCl$_2$ P on the same soil. This implies that water extractable P (WEP) and CaCl$_2$ P extract P from two different pools. The WEP measurement is well correlated to the amount of dissolved P in runoff (Pote et al., 1996). WEP is mainly derived from the pool of plant-available P (Luscombe et al., 1979) while CaCl$_2$ P is derived from soil solution P (McDowell et al., 2001).

This change point threshold is important in the management of P as it represents the amount of P that can be applied to soil before significant environmental impacts occur. The change point generally occurs at a STP level about twice what is needed for optimal plant growth (McDowell et al., 2001). Keeping soil P levels below the change point can prevent losses of P that can lead to the degradation of waterways without sacrificing the needs of the plant. In addition, economic losses can be prevented by efficiently managing P to reduce excess application (Sharpley et al., 1996).
1.3 Degree of Phosphorus Saturation

Mechlich 3 extractable P, Al, and Fe (P$_{M3}$, Al$_{M3}$, and Fe$_{M3}$) concentrations have been used to determine the degree of P saturation (DPS) (Khiari et al., 2000; Kleinman and Sharpley, 2002; Maguire and Sims, 2002b; Onweremadu, 2007; Zhang et al., 2005). DPS is defined as the ratio of ammonium oxalate extractable P (P$_{ox}$) to the total phosphorus sorption capacity determined using ammonium oxalate extractable Al and Fe (Al$_{ox}$ and Fe$_{ox}$) (van der Zee and van Riemsdjik, 1988). DPS is commonly estimated using the equation:

\[
\text{DPS}_{\text{ox}}(\%) = \frac{P_{\text{ox}}}{[\text{Al}_{\text{ox}} + \text{Fe}_{\text{ox}}]} \times 100
\]  

where P$_{ox}$, Al$_{ox}$, and Fe$_{ox}$ are in mmol kg$^{-1}$ (Beck et al., 2004). Using P$_{M3}$, Al$_{M3}$, and Fe$_{M3}$, a P saturation ratio (M3-PSR) was determined from the equation described by Maguire and Sims (2002b):

\[
\text{M3-PSR} = \frac{P_{M3}}{[\text{Al}_{M3} + \text{Fe}_{M3}]} \times 100
\]  

where P$_{M3}$, Fe$_{M3}$, and Al$_{M3}$ are in ppm P. The M3-PSR and DPS$_{ox}$ have been determined to be well correlated (Figure 1.3). This is significant because the Mehlich 3 extraction is an inexpensive, routine procedure for many soil testing labs (Maguire and Sims, 2002b).

M3-PSR has been used to identify certain environmental thresholds. Maguire and Sims (2002b) determined that M3-PSR was a good predictor of P losses in leachate. Using a M3-PSR test to determine P saturation is preferable because it makes use of data that is readily available from soil testing labs that use Mehlich 3 as a standard procedure (Kleinman and Sharpley, 2002). No additional testing is required because P, Al, and Fe can all be analyzed simultaneously using inductively coupled plasma (ICP)-atomic emission spectroscopy (AES), reducing analysis costs and processing times.

1.4 Mapping Soil Phosphorus

As P management becomes more important, knowledge of the P spatial distribution becomes valuable. Delineating areas most susceptible to P loss will make it easier focus conservation efforts on critical locations. This would make P management more time and cost efficient. Mapping the change
point at the watershed scale would enable land managers to better conserve their land and protect the environment by focusing efforts on highly susceptible areas.

The goal of this effort is not to create a map of P that can be used to determine the P at a specific site. A map of this type on a regional scale would not be accurate due to the dynamic nature of P. In order to accurately determine P at a single location, a site soil test is needed (Lark and Ferguson, 2004). Even a map of P distribution at a field scale may be inaccurate due to the soil variability that is often unrelated to observable variations in soil properties (Phillips, 2001). Creating a map that indicates these high risk areas would be helpful. This map could be used by land managers to better understand the P distribution across a landscape and to identify areas susceptible to P loss via leaching.

Understanding the distribution of P is helpful because it can be used to identify areas where P application should be avoided as well as areas were additional P application presents fewer risks. Determining areas of potential P application is helpful as certain guidelines exist limiting the amount of manure that can be applied to land (USEPA, 2011). For producers with excess manure that cannot be applied to their land due to these guidelines, alternative application sites must be identified. Developing these large scale nutrient plans can help producers avoid P losses and comply with federal regulations.

1.4.1 **scorpan Modeling**

A model for soil formation was developed by Jenny (1941). Jenny’s descriptive model for soil development includes climate, organisms, topography, parent material, and time as the basic soil forming factors. Using Jenny’s formula to quantify the relationship between soil and spatially referenced factors, the **scorpan** model was specified by McBratney et al. (2003). This model considers seven prediction factors: \( s \) (soil, other properties of the soil), \( c \) (climate), \( o \) (organisms), \( r \) (relief, or landscape attributes), \( p \) (parent material), \( a \) (age), and \( n \) (space, or spatial position). The **scorpan** model is similar to Jenny’s model with the addition of the space and soil factors. The addition of the soil factor is necessary because soil can be predicted from its properties (McBratney et al., 2003).
Using the *scorpan* model as a theoretical basis, spatial modeling techniques can be used to develop digital soil maps (Grimm et al., 2009; Kumar et al., 2012; Malone et al., 2014). Using *scorpan* to guide the selection of environmental covariates expected to be related to observed soil spatial variability, spatial data can be used to predict soil properties over an area. Some of the common spatial modeling techniques include linear regression, kriging, and classification and regression trees, as well as combinations of these techniques (McBratney et al., 2003).

### 1.4.2 Kriging

Kriging is a spatial interpolation method that is based on the idea that each sampled location is a random variable with a probability distribution at each sampling site. All of the site distributions are assumed to share some common properties. The kriging tool in ERSI’s ArcMap fits a mathematical function to all of the points within a specified radius to determine the output value for each location.

Using the weights of the surrounding measured values, kriging derives a prediction for an unsampled location using the formula:

$$
\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)
$$

where $Z(s_i)$ is the measured value at the $i$th location, $\lambda_i$ is the unknown weight for the measured value at the $i$th location, $s_0$ is the prediction location, and $N$ is the number of measured values. In ordinary kriging, the weight depends on the model fitted to the measured values, the distance between the measured values and the predicted value, and the spatial relationships between the measured values around the predicted location (i.e. spatial autocorrelation between the measured points).

The regionalized variable theory is the basis of kriging. A regionalized variable is a function that is defined over space and time. At a local level, the variable is random, but at a regional scale is structured in a way that the spatial distribution varies from place to place with a measurable continuity (Roye et al., 1981). The two conditions of the intrinsic hypothesis of regionalized variable theory are (1) the expected values at any two points separated by distance, $h$, is zero and (2) the variance of the difference between...
two points depends on the distance between the points and not the position of the points (McBratney and Webster, 1986). The difference in variance between the two points is given by the equation:

\[
\gamma(h) = \frac{1}{2} [z(x) - z(x + h)]^2
\]

(4)

where the \(z(x)\) is the value at a specific location and \(z(x + h)\) is the value at a specific location some distance, \(h\), away from \(x\). The semivariance, \(\gamma(h)\), is a quantification of the degree of dissimilarity. Using the only this formula to develop a prediction surface is often referred to as ordinary kriging (OK).

The semivariogram relates the distance between points to the semivariance. Semivariograms are used in kriging to illustrate the spatial autocorrelation and to interpolate the value of a property at an unsampled location (Curran, 1988). The semivariogram is the single most important tool in geostatistical applications to soil (McBratney and Webster, 1986). Choosing an appropriate semivariogram requires an understanding of the semivariogram and its limitations.

The three characteristics used to describe a semivariogram are the sill, the range, and the nugget (Figure 1.4). The range is the distance where the model first flattens. Above the range, the sample location pairs are not spatially correlated. Sample location pairs separated by distances closer than the range are spatially autocorrelated. The value at which the model reaches the range is referred to as the sill. Finally, the nugget is a value of the y-axis intercept. Theoretically, the intercept should be at \((0,0)\), i.e., points separated by distance zero, are the same point. The nugget is attributed to measurement error, variation over distances shorter than the sampling interval, and variation that is not spatial correlated (Oliver and Webster, 1990). Understanding how semivariograms influence the prediction surface allows the user more control over the kriging process.

### 1.4.3 Co-Kriging

A logical extension of kriging uses multiple spatially interdependent ancillary variables to generate a better prediction surface. Co-kriging (CK) combines sampled locations with fully sampled variables (Odeh et al., 1995). Many of these variables are inexpensive to measure and can reduce the
number of observations of an expensive variable (Knotters et al, 1995). Co-kriging is also more powerful for predicting values when limited data is available (Vauclin et al., 1983). The additional data input can make the co-kriging model cumbersome (McBratney et al., 2003) and increase processing times. The final output of co-kriging is a weighted average of the available data (McBratney and Webster, 1983).

### 1.4.4 Indicator Kriging

One form of kriging that is used to determine probability is indicator kriging (IK). This form of kriging uses conventional ordinary kriging techniques applied to data that has been transformed into a binary indicator variable. This form of kriging produces an output that estimates the probability that a site will exceed a certain threshold. This has been used by Lark and Ferguson (2004) to determine the probability that a site exceeds certain P recommendation thresholds. Many properties in soil have thresholds that favor the use of IK, such as nutrients (von Steiger et al., 1996; Webster and Oliver, 1989) and salinity (Eldeiry and Garcia, 2011; Wood et al., 1990).

IK uses a threshold (in this case the change point value) to transform data into a binary variable. Where a one represents a value over the threshold and a zero represents a value below the threshold. The transformed dataset can then be used to predict values at other, unsampled locations. The predicted values are the probability that the location values are greater that the threshold (Journel, 1988).

### 1.4.5 Regression Tree Modeling

Using nonlinear models is the preferred method for determining parameter relationship in soils (Kumar et al., 2012). This is due to the fact that linear models are not able to account for the spatially varying relationships in variables (Bates and Watts, 1988). In soils, many parameters are spatially variable and cannot be described using a linear model. One nonlinear model often used in soil modeling is the regression tree model. Regression trees are used to predict by identifying optimum break points in predictor variables and separating the dependent variable into groups so that these groups are as homogenous as possible (Breiman, 1984). A regression tree contains multiple splits that are determined by optimum break points in the dataset to achieve minimum in-group variability and maximum inter-
group variability. The resulting tree contains a certain number of terminal nodes at which data can no longer be split into groups.

Regression trees have been used to predict continuous soil attributes like total organic carbon, (McKenzie and Ryan, 1999), sand content (Pachepsky et al., 2001), silt content (Park and Vlek, 2002), and cation exchange capacity (CEC) (Bishop and McBratney, 2001). Regression tree analysis is able to deal with nonlinearity of data, incorporate categorical variables, and requires no assumption about the data (Ryan et al., 2000). Some of the limitations of regression trees include subjectivity in choosing the tree size and the discrete output (the output contains a finite number of values equal to the number of terminal nodes) (McBratney et al., 2003).

In this study, our hypothesis was that the change point analysis can be used to explain P environmental thresholds throughout a watershed. The objectives of this research are to (1) develop a relationship between STP and CaCl₂ extractable P (CaCl₂ P) for each land type within the watershed to determine the change point for each land use type in a watershed and (2) use scorpan modeling to predict STP for the entire watershed and use these values to predict the amount of P that can be added to the soil before loss of P based on the change point determined for each land use type.
Chapter 2. Materials and Methods

2.1 Study Area and Sample Collection

The Potomac watershed is a 9 million acre watershed of the Potomac River (Figure 2.1). This basin covers the Washington D.C. area as well as parts of Maryland, Pennsylvania, Virginia, and West Virginia. In addition, this watershed covers parts of six MLRA: 127, 130A, 136, 147, 148, and 149A. This watershed is of importance because it drains directly into the Chesapeake Bay. Eutrophication has become a substantial problem in the Chesapeake Bay over the last quarter century, affecting the entire ecosystem in the Bay (Boesch et al., 2001). Four sub-watersheds of the Potomac watershed including the (1) North Branch Potomac, (2) Cacapon-Town, (3) Conococheague-Opequon, and (4) Middle Potomac-Catoctin watersheds were used in this study. Collectively, these watersheds cover 3.8 million acres.

Within the sampling area, 23 sites were sampled for the Rapid Carbon Assessment (RaCA) project (Figure 2.2). The RaCA project was developed by the United States Department of Agriculture (USDA)-Natural Resources Conservation Service (NRCS) to determine the soil organic carbon stock across the conterminous United States at a single point in time (USDA, 2013). Sampling sites were randomly located based on multi-level stratified sampling. Sampling points were first stratified by RaCA region, which were based on major land resource areas (MLRA). There were 17 RaCA regions within the conterminous US, each managed by a regional office. Sampling sites were further stratified by soils and land use. The land use classes that were the basis of stratification were coordinated with National Resources Inventory (NRI) definitions. The exact procedure used to place each point is not publicly available due to the confidentiality requirements of the NRI program.

After each site was located, soil samples and site information were collected. Each site was treated as a plot and five pedons were located in each plot. One pedon was located in the center and four pedons were located 30 m away from the center pedon in each cardinal direction. Each pedon was described individually to 50 cm or to a root limiting layer, such as a fragipan or bedrock. Samples of a known volume were taken from each horizon and sent to the regional RaCA office for analysis.
At the regional office, samples were air-dried and sieved to less than 2 mm. For purposes of the RaCA project, mineral samples were scanned with a Visible Near Infrared (VNIR) spectrometer. All of the organic horizons along with the mineral samples from the center pedon were sent to the Kellogg Soil Survey Lab (KSSL) at the National Soil Survey Center in Lincoln, NE. Samples that were analyzed at the regional offices were stored at the regional office after analysis. This study utilized the RaCA samples located at the RaCA Region 13 office in Morgantown, WV.

There were 23 plots sampled within the study area. Without the central pedon from each site, which was sent to the KSSL for analysis, a total of 92 pedons were available for analysis in this project. The surface horizons of the soil contribute the most P to surface waters (Hesketh and Brookes, 2000); therefore, only the horizons in the upper 15 cm are included in analysis. This excludes the organic horizons that were sent to the KSSL. One sampled pedon was incomplete, missing horizons above 20 cm. Therefore, this pedon was not used in analysis. In total, about 240 samples were available from horizons wholly or partially in the upper 15 cm of the 92 pedons sampled.

2.2 Determination of Soil Chemical Properties

The quantity of P in the soil was measured using a Mehlich 3 extraction. Mehlich 3 extractable P is the standard STP method in most of the northeastern United States (McDowell et al., 2001). Soil extractions with Mehlich 3 were completed by shaking 2 g soil with 20 mL Mehlich 3 extracting solution (0.2M CH$_3$COOH, 0.25M NH$_4$NO$_3$, 0.015M NH$_4$F, 0.013M HNO$_3$, and 0.001M EDTA) and shaking for 5 min on a reciprocal shaker. Extracts were then filtered through Whatman No. 42 filter paper and analyzed using ICP-AES. ICP-AES analysis was used to determine the quantity of P in the extract as well as the quantity of Al and Fe.

The intensity of P in the soil was represented by CaCl$_2$ P. CaCl$_2$ P is correlated to the amount of dissolved P in runoff (Soltanpour et al., 1974). Determination of CaCl$_2$ P is a simple procedure developed by Luscombe et al. (1979). The procedure has been modified to allow for a soil to solution ratio of 1:5 since this ratio is most appropriate to predict the change point in STP (McDowell and Sharpley, 2001).
CaCl$_2$ P was determined by shaking 4 g of soil with 20 mL of 0.01M CaCl$_2$ on a reciprocal shaker for 1 hour. Extracts were also analyzed by ICP-AES after filtering through Whatman No. 42 filter paper until clear.

All extractions were done in duplicate and multiple quality controls were implemented. A blank for each extractant (CaCl$_2$ and Mehlich 3) was incorporated into ICP-AES runs. These blanks contained 0 ppm P. The blanks served as quality control measure and were also used to determine the detection limit of the instrument. Standards were also incorporated into the ICP-AES run. These standards contained known concentrations of P ranging from 0.5 ppm to 10 ppm. These samples were used to insure the accuracy of the instrument. Finally, a set of benchmark soils was included in every set of extractions. These benchmarks soils contained known Mehlich 3 P and CaCl$_2$ P concentrations. These samples were included as a quality control measure for the extraction procedure. These three quality control measures helped insure the accuracy of the chemical analysis.

### 2.3 Covariates Influencing STP

Factors that influence STP include pH, organic matter content, Fe and Al oxide content, and mineralogy (Barrow, 1984; Frossard et al., 2000; Sanyal and DeDatta, 1991). Data regarding a number of these variables can be accessed using USDA-NRCS soils information. Most of the soils in the US have been mapped by the USDA-NRCS, and these maps are compiled in the Soil Survey Geographic (SSURGO) database (Soil Survey Division Staff, 1993). The SSURGO data, which uses a vector data model, has been converted to a raster data model, known as Gridded SSURGO (gSSURGO). Each grid cell represents the soil survey map unit that is present in the landscape. gSSURGO data also contains data tables with information about the soils in each map unit, including the particle size distribution, pH, organic matter content, and cation exchange capacity (CEC) for each soil series.

Elevation data in raster format is available from the United States Geologic Survey (USGS) National Elevation Dataset (NED). This raster is a 10-meter digital elevation model (DEM) created from a variety of sources including cartographic contours (Gesch et al., 2009). This data are available for the
entire U.S. from the USGS. Multiple geomorphology layers including surface roughness, moisture, slope position, and compound topographic index (CTI) were created using the Geomorphology and Gradient Metrics Toolbox for ESRI ArcGIS (Evans et al., 2014). Using the DEM as an input this toolbox was able to create multiple layers to better describe the geomorphology at a site. The CTI is a steady state wetness index that is a function of the slope and the upstream catchment area per unit width orthogonal to the flow direction (Gesssler et al., 1995). In particular, Moore et al. (1993) determined that the CTI was well correlated with STP concentrations (R= 0.53) across a single hillslope.

The slope position layer is a function of elevation. It is calculated by subtracting the focalmean raster from the original elevation. The slope position is a measurement of where a point is on a slope, i.e., if a point is at the bottom, the middle, or the top of a slope. This parameter influences many soil properties due to its effect on water flow and soil movement. Points at the bottom of slopes act as a constructive environment, accumulating organic material, water flow, and nutrients from upslope.

The surface roughness parameter is calculated by taking the sum of the absolute value of the difference in elevation between a cell and its neighbors (Riley et al., 1999). This layer is a function of the complexity of the surface and can influence water flow and accumulation. Roughness is a useful parameter for evaluating the flow pathways of a system.

Climate data including precipitation and temperature was developed and published by the Oregon State University’s PRISM Climate Group. Climate grids are developed using DEM as input variables. Elevation is assumed to be the most important factor in the distribution of climate variables (PRISM Climate Group, 2004). 30-year normals for precipitation and mean temperature were analyzed in this study. These datasets are the official spatial climate data of the USDA.

Land cover information is available from the National Land Cover Database 2011 (NLCD 2011). This dataset is developed using decision tree classification of Landsat satellite data (Jin et al., 2013). The
dataset is available for the entire U.S. at a 30m resolution. Data are grouped into 16 land cover classes. The most recent version of this dataset is based on satellite imagery from 2011.

Geology information was obtained from the USGS. Digital geologic maps are available in vector format for all 50 states, Puerto Rico, and Washington, D.C. The attribute of interest in this dataset was the dominant lithology of the map unit. These polygon units were converted to raster format. In the event that a polygon contains more than one dominant lithology, the lithology with the greatest maximum area within the output cell is considered the value of the output. The geology is a proxy for soil parent material.

Parent material is also correlated to the gamma-ray radiation of earth surfaces (Cook et al., 1996). Gridded data developed using aerial sensing for the radiation emanating from the earth’s surface is used to estimate the abundance of Potassium (K), Thorium (Th), and Uranium (U). This data is published by the USGS and covers the conterminous U.S. as well of parts of Canada and Alaska.

gSSURGO data along with geomorphology layer and additional information about parent material (from geologic maps), and other environmental covariates including precipitation, landscape position, and elevation can be used to define the influential soil properties. Using the statistical software R it is possible to determine significant correlations between soil properties and the Mehlich 3 P and the M3-PSR using a regression tree analysis. These covariates and their sources are summarized in Table 1.2.

2.4 Statistical Analyses
Duplicate extractions were averaged, giving one Mehlich 3 P and one CaCl$_2$ P measurement for each horizon in the upper 15 cm of the mineral surface. A depth weighted average was then taken for each pedon. Using a depth weight average to aggregate the horizon data insured that the final pedon measurement was representative of the P leaching contributions from each horizon.

The relationship between the quantity and intensity of the P in the soil is determined by plotting the Mehlich 3 P against the CaCl$_2$ P. A split-line model describes this relationship and divides the curve
into two linear sections. Split-line model is executed in R using the statistical package “segmented” (Muggeo, 2008). The algorithm used by this package is an iterative procedure. This means that linear model fitting is repeated until there is a statistically significant slope change found. The parameterization used to model this segmented linear relationship between the response and the variable $Z$ fits the terms:

$$aZ + \beta (Z-\psi),$$  \hspace{1cm} (5)

Where $\psi$ is the break point, $a$ is the slope of the left line, and $\beta$ is the difference in slopes. Therefore $(a+\beta)$ is the slope of the right line. If a break point exists, then $\beta$ does not equal 0. At each iteration, the two lines are fitted until no further improvement in the break point estimate (Muggeo, 2003). When a break point is not found, it is still possible that a break point exists, but isn’t detected in the configuration of the data.

The significant change in slope occurs at the change point (Heckrath et al., 1995). The change point was determined independently for each land use within the watershed (forestland and agricultural lands) as well as for all samples within the study area. The change point represents a significant environmental threshold where adding additional P will have a substantial impact on the ability of the soil to retain P. Keeping STP levels below the change point is expected to reduce the amount of P in surface waterways, increasing the overall health of the water.

A linear regression will also be performed on the data to determine if there is a simple linear relationship between the Mehlich 3 P data and the CaCl$_2$ P data. The existence of this linear relationship could be used to determine the strength of the segmented linear relationship. Linear regression uses used to predict values using the formula:

$$y_i = \alpha + \beta x_i + \epsilon_i$$ \hspace{1cm} (6)

Where $\beta$ is the slope of the regression line (called the regression coefficient), $\epsilon_i$ summarizes the error, and $\alpha$ is the point where regression line intersects with the x-axis.
2.5 Kriging

A variety of OK, IK, and CK methods were used to create predication surfaces. These methods were chosen due to the distribution of the data and the error results of the methods. IK was used to determine the probability that a location exceed the change point. Three separate IK surfaces were generated using the change point for forested land uses, agricultural land uses, and pooled land uses. OK was used to model the M3-PSR and Mehlich 3 P throughout the watershed. Finally, CK was also used to model Mehlich 3 P, using CEC and geology as covariates. These methods were chosen based on the distribution of the data. It is suggested that a density of at least 1 observation per 25,000 acres is adequate for making point predictions via scorpion kriging (GlobalSoilMap Science Committee, 2013). This dataset has about 6 points per 25,000 acres. The data was also well distributed throughout the watershed; this leads to more accurate results because there is adequate coverage.
Chapter 3. Results and Discussion

3.1 Soil Chemical Properties

The soils tested had a wide range of Mehlich 3 P concentrations (Tables 3.1-3.3). The agricultural sites generally had more STP than the forested areas. The higher levels of soil P are likely due to application of P in agricultural lands in the form of inorganic P fertilizer as well as manures and poultry litter. Without any knowledge of site management practices, it is difficult to draw more definitive conclusions on how land use impacts the STP levels. The distribution of CaCl$_2$ P in the soil was similar to Mehlich 3 P. There was a wide range of values with the agricultural sites having a higher CaCl$_2$ P on average (Table 3.1).

Another trend that becomes apparent is the higher amounts of Fe and Al in the forest soils. This is most likely a result of the lower pH in forest environments. Lime is often added to agricultural soil to raise the pH and make essential nutrients available. At higher pH, Al and Fe form stable compound with hydroxyl ions (OH$^-$) in solution, making them harder to extract using a Mehlich 3. In addition, there may be more Fe due to the lower removal rates of Fe by plants because there is little to no harvesting in these systems that can interrupt nutrient cycling. An additional measurement of soil pH taken on all of the samples in this study would elucidate the relationship between P concentrations and land use.

There were few linear relationships between the Mehlich 3 P results and the environmental covariates used in the geostatistical analysis (Figure 3.1). There is a trend in the slope position and the Mehlich 3 P (Figure 3.1j). Points on the landscape in a summit, shoulder, or footslope position have lower levels of STP. This may be due to loss of P due to erosion, which has a greater potential on sloped areas. There is also a relationship between elevation and STP. Higher elevations have lower Mehlich 3 P values (Figure 3.1h). This is most likely a function of land use. Areas at lower elevations are more likely to be agricultural land uses where P additions are common. Finally there is a relationship between the abundance of K (from gamma radiometric data) and Mehlich 3 levels. K abundance is related to parent
material. This implies that parent material influences STP. The amount of P minerals that a lithology contains can cause different P levels in soils (Cathcart, 1980).

Many of these trends were also apparent in CaCl₂ P (Figure 3.2). However, there was a better relationship between CaCl₂ P and sand content (Figure 3.2c). The amount of CaCl₂ P in the soil decreased as the sand content increased. This may be a proxy for CEC as sandy soils generally have a lower CEC. Soils with a lower CEC have less potential for P retention resulting in lower CaCl₂ P levels.

### 3.2 Change Point Analysis

The change point was calculated for each land use separately and pooled land uses. The value of the change point in the forested sites was 42 mg P kg⁻¹ (Figure 3.3). This means that when soils have STP levels above 42 mg P kg⁻¹, CaCl₂ P increases at a greater rate per unit increase of STP than below this threshold. The lack of points above the change point indicates that this analysis may contain some inaccuracy. In theory, it is possible to determine the change point from four data points, two on each side of the change point. McDowell et al. (2001) performed a change point analysis on their data and then randomly removed data points from the analysis in order to determine how many points are needed to accurately predict the change point. They determined that in order to determine the change point, within 40% of the determined STP value, at least eight data points are necessary, four on each side of the change point. This implies that the change point for forested soils in this study may require the inclusion of additional data points in order to produce a more accurate change point result. Few published works have looked at the change point of forested land. More research of P dynamics in forested systems is necessary to accurately describe the change point in forests.

In agricultural lands, the change point value was 64 mg P kg⁻¹ (Figure 3.4). This number is significant because it lies above the optimum STP levels for agronomic crops. For most states that provide agronomic recommendations based on Mehlich 3 P, including Pennsylvania and Delaware, Mehlich 3 P values above 50 mg P kg⁻¹ have no plant yield response (Beegle, 1999; Kleinman et al., 2000). The optimum range for adequate yields is 30-50 mg P kg⁻¹ (Beegle et al., 2014). The change point
determined in this analysis is comparable to previously published change points (Table 1.1). For studies in agricultural lands that also used Mehlich 3 P and CaCl$_2$ P as change point analysis factors, the change point was in the range of 20 to 175 mg P kg$^{-1}$.

There has been evidence that the change point dynamics differ within the agricultural land use category. McDowell and Condron (1999) showed that there was a difference between the change points in arable and grassland management. Soils collected for the RaCA project and used in this study were classified as cropland and pastureland. Because of the low number of samples classified as cropland, these two land uses were pooled together into agricultural lands. If more data points were available, it may be useful to separate these two land uses and perform change point analysis on the groups separately.

When the land uses were pooled together, the change point was 66 mg P kg$^{-1}$ (Figure 3.5). Pooling land uses should produce a less accurate result because of the different P loss dynamics of land uses (McDowell and Condron, 1999). It has been shown that catchments with greater forested land uses generally have less P concentrations in streams (Scott et al., 2011). The change point determined by pooling land uses is higher than the individual change point values under different land uses. It was expected that the pooled change point would be in between the two individual land uses. The additional data points at lower Mehlich 3 P concentrations as well as the lack of forest data points at higher Mehlich 3 P concentrations may contribute to the higher pooled change point value. This result may reinforce the necessity of exploring the P dynamics in forested systems.

Previous change point analysis studies group datasets by soil type or by land use. It may be useful to perform an alternative change point analysis on data grouped by soil type if the data was available. The result of such an analysis would help determine if the change point is a function of land use or an inherent property of the soil. Without the soil type data we are only able to conclude that a different change point was determined for each land use, suggesting that land use does play a role on the P leaching loss from a soil.
A linear regression was performed on the Mehlich 3 P and the CaCl₂ P data for forested land uses, agricultural land uses, and pooled land uses (Figures 3.6-3.8). It was determined that there was no significant difference in the regression methods. (Table 3.4). This result means that the segmented linear model explains the relationship in the data as well as a simple linear model. This may be due to the fact that the Mehlich 3 P levels of the sampled soils, especially agricultural soils, are much lower than expected. For example, it is known that in the eastern panhandle of West Virginia, an area that is covered by the study area, STP levels are commonly much higher than the optimum range (Sutton and Cox, 2012). Because these high STP soils were not captured in the testing, it is possible that there is a more significant change point at a higher STP level. Many published change point analysis studies find a change point in Mehlich 3 P and CaCl₂ P at a higher STP level than found in this analysis (Table 1.1). Kleinman et al. (2000) found a change point at 141 mg P kg⁻¹ using Mehlich 3 P and CaCl₂ P data determined using agricultural soils from the Delaware River watershed in NY and PA. This value is more than double the change point value determined for agricultural soils in the study area (64 mg P kg⁻¹).

Adding additional data points that are in a higher Mehlich 3 P range may increase the accuracy of the change point analysis result. This could be accomplished by treating soils samples with P and performing additional Mehlich 3 P and CaCl₂ P analysis on the samples or resampling the watershed to capture the known high STP values. The change point values determined in this study using segmented linear regression can still be used with spatial modeling techniques to visualize the distribution of P throughout the watershed.

### 3.3 Phosphorus Sorption Ratio

The capacity of a soil to sorb P is a function of the Al and Fe concentrations. Al and Fe oxides account for a majority of the P retention reactions in the soil. This analysis is commonly performed using acid ammonium oxalate extractable Al, Fe, and P. The ratio determined using an acid ammonium oxalate extraction is well correlated to the ratio determined using a Mehlich 3 extraction (Khiari et al., 2000; Kleinman and Sharpley, 2002; Maguire and Sims, 2002b; Onweremadu, 2007; Zhang et al., 2005). The
M3-PSR determined using the soils in this watershed had a wide range. On average, the agricultural lands had a higher M3-PSR than the forested lands. This is similar to the results from the Mehlich 3 extractions.

There is a positive relationship between the M3-PSR and Mehlich 3 P as well as the M3-PSR and CaCl$_2$P (Figures 3.4 and 3.5). There is often a change point found in the relationship between the degree of P saturation and the environmental P concentration (Ilg et al., 2005; Kleinman et al., 2000; Maguire and Sims, 2002b; Wang et al., 2012). Maguire and Sims (2002b) compared the relationship between M3-PSR and leachate P concentration to the relationship between oxalate extractable P saturation and leachate P concentrations and determined that the M3-PSR was better at predicting P losses in leachate. We found no change point relationship in the data.

The M3-PSR analysis showed that a majority of the soils were not saturated with P. The soil with the highest M3-PSR was 28% saturated with P (Table 3.3). Based on the equation developed by Kleinman and Sharpley (2002) where the M3-PSR was about 70% of the DPS$_{ox}$ (Figure 1.3) this value is equivalent to about 40% saturation. This demonstrates that there is much more potential for P to react with Al and Fe to retain P in these soils. Sims et al. (2002) suggest that a 0.10-0.15 M3-PSR value be used to identify soils at risk of detrimental P losses. Most of the soils in this study were well below this range. Only 11% of soils analyzed were between or above this range. This is most likely due to high levels of Al and Fe that were found in the soil, which is probably a function of parent materials and land use. Untested chemical properties like pH, organic matter content, and Ca content have an impact on P retention. In the future, testing these properties of soils along with the quantity and intensity of P in the soil would give more insight into the interactions in the soil in an effort to better explain the dynamics of P.

The M3-PSR also had relationships with a few environmental covariates listed in Table 1.2. The mean annual precipitation was negatively correlated with M3-PSR. Areas with lower precipitation have higher levels of M3-PSR (Figure 3.9f). The precipitation values were derived mainly from elevation data.
Generally, areas with higher elevation have more precipitation. Because most agricultural land uses are found at lower elevations, areas at higher elevations will most likely receive less applications of P. In addition, other M3-PSR relationships followed the same trends as the relationships between M3P and the environmental covariates described in Section 3.1.

M3-PSR is often useful in predicting P losses in leachate (Ilg et al., 2005; Kleinman and Sharpley, 2002; Wang et al., 2011). A segmented linear relationship between an environmental P measurement and M3-PSR was also identified by Wang et al. (2011). While there was no change point between the M3-PSR values and the Mehlich 3 P values in this study, the relationship was linear ($r^2 = 0.80$). The lack of change point may be due to the high number of low M3-PSR values. It is possible that the change point was not apparent because of the lack of data points with a high M3-PSR. The linear relationship between the M3-PSR values and the Mehlich 3 P values was stronger than the linear relationship between the M3-PSR values and the CaCl$_2$ P values ($r^2 = 0.50$). This may be due to the fact that the Mehlich 3 P value was a variable in the calculation of M3-PSR.

### 3.4 Regression Tree Analysis

Regression tree analysis provides predicted values based on the fitted tree model. The difference between the actual value and the predicted value, the residual, can be used to evaluate the accuracy of the prediction model. Each point will have a residual ($e$) where $e = y - \hat{y}$ when $y$ represents the actual value and $\hat{y}$ represents the predicted value. The best fitting model will have a low root mean square error (RMSE) ($\text{RMSE} = \sqrt{\frac{\sum (y - \hat{y})^2}{n}}$; where $n$ = the number of samples).

Using regression tree analysis is it possible to determine the most influential factors contributing to the Mehlich 3 P concentrations. In this case, environmental covariates summarized in Table 3.5 were used to predict the Mehlich 3 P concentrations. The most influential values in this dataset were surface roughness, geology, and CEC (Figure 3.12).
The predicted values produced from the tree model were compared with the actual Mehlich 3 P concentrations determined using the Mehlich 3 extraction. The RMSE for this model was 14 mg P kg\(^{-1}\) and the predicted values were correlated to the observed values \((r^2 = 0.58)\) (Figure 3.13). This demonstrates that the tree model fits the data very well.

A tree model was also used to model the M3-PSR data. The predicted M3-PSR ratios were compared with the measured M3-PSR ratios and the fit of the model was evaluated using the RMSE parameter. The RMSE for the M3-PSR data was 2.9% and the predicted values were correlated to the observed values \((r^2 = 0.67)\) (Figure 3.14). The most important variables in this model were sand content, K abundance, and moisture (Figure 3.15).

### 3.5 Geospatial Analysis

Geospatial analysis is commonly used to predict soil properties (Kumar et al., 2012; Knotters et al., 1995; Malone et al., 2009; Odeh et al., 1995). Extensive soil sampling at every location within a study area is costly and time consuming. The ability to predict soil properties using easily obtained data as inputs has made soil mapping more efficient and effective.

Spatial autocorrelation can be quantified using the Moran’s I Index. This index value is a measurement of the spatial autocorrelation to determine if the pattern is clustered, random, or dispersed. The Moran’s I Index was determined for the Mehlich 3 P data and results indicated that the points were clustered \((I= 0.51, Z= 6.2)\) (Table 3.5). The spatial autocorrelation in this data means that there is a correlation between degree of similarity and proximity. An analysis of the spatial autocorrelation using Moran’s I Index was also performed for the M3-PSR data (Table 3.5). The data was also determined to be clustered \((I = 0.72, Z= 8.8)\). Kriging interpolation methods take this correlation into account by using the relationship as a weight to make more accurate predictions.

The Mehlich 3 P data was highly skewed to the right (Figure 3.16). Most interpolation methods work best with normally distributed data. In order to make a more accurate surface, it is necessary to
transform the data to make it normal. For this dataset, a log transformation was used to stabilize the
variance so that the mean and median were approximately equal (Figure 3.17). M3-PSR values were also
skewed to the right (Figure 3.18). A log transformation was necessary on this dataset as well (Figure
3.19). Kriging of the transformed data was used to generate an accurate prediction surface free of bias due
to data distribution.

Another measurement of data normality is a normal Q/Q plot. Plotting the k\textsuperscript{th} smallest observed
value against the k\textsuperscript{th} smallest predicted value should produce a straight line if the data is normally
distributed (Dalgaard, 2008). The Mehlich 3 P data and the M3-PSR data were evaluated for normality
using normal Q/Q plots. The normal Q/Q plot of the Mehlich 3 data showed that the data was not
normally distributed (Figure 3.20) because the data follows a shape different from a straight line. After a
logarithmic transformation was done on the data, the normal Q/Q plot (Figure 3.21) shows that the points
follow a relatively straight line.

The values of M3-PSR were also analyzed for normality using a normal Q/Q plot. The M3-PSR
data was not normal as the Q/Q plot did not produce a straight line (Figure 3.22). A logarithmic
transformation of the M3-PSR data produced a normal Q/Q plot where deviations from a straight line
were minimal (Figure 3.23). These results show that the logarithmic transformation of the Mehlich 3 P
data and the M3-PSR data produced approximately normally distributed data appropriate for kriging
spatial interpolation methods.

CK was used as the interpolation method to determine the amount of Mehlich 3 P in the soil
throughout the watershed (Figure 3.24). Regression tree analysis was used to determine the most
important variables (Section 3.4). Two of these important covariates, geology and CEC, were used to as
ancillary variables in co-kriging to develop a prediction surface for the entire watershed. This
interpolation method had a RMSE of 17 mg P kg\textsuperscript{-1} (Table 3.6). When the Mehlich 3 P values were
interpolated using OK, the results did not improve (RMSE= 18 mg P kg\textsuperscript{-1}). Both of these values are below
the median of the Mehlich 3 P data (Table 3.3). This means that the predictions are relatively accurate; they may contain some inaccuracies but are useful for comparison purposes (Cattle et al., 2002). The maps produced from CK and OK (Figures 3.24-3.25) show areas where high P values are likely to occur. At a watershed scale this information is only useful in delineating areas that may need further investigation. To accurately determine the amount of P in soil at a certain location, it is necessary to analyze the soil from that location. Webster and McBratney (1987) determined that local maximum and minima of STP occur at an average spacing between 200-250 m. Because of this inherent variability, it may not be useful to use prediction surfaces of STP to determine the amount of STP at a specific location.

There were explainable trends apparent in the CK and OK models. The highest predicted locations were located in the eastern panhandle of West Virginia. This area has a high density of farms (USDA, 2007), where P is often applied to the soil in the form of inorganic fertilizer and manures. It is likely that these areas will have STP levels at or near the agronomic optimum, $50 \text{ mg P kg}^{-1}$ (Sims et al., 2001). Areas with the lowest predicted P are located near urban areas like Harrisburg, Pennsylvania; Hagerstown, Maryland; Frederick, Maryland; and Martinsburg, West Virginia. Urban areas will have little to no input of P. Soils in these areas had low STP values, driving the predication model to estimate surrounding values lower.

It should be noted that predicted values in the eastern panhandle of West Virginia were much lower than known values (Sutton and Cox, 2012). This may be due to the fact that points used in the prediction surface were taken for farmlands involved in conservation planning. Agricultural lands were chosen from cooperating farmer’s land. These farmers are working with government agencies to conserve farmland and reduce their environmental impact via nutrient management planning. It is more likely that farms enrolled in these programs are following P application guidelines. Farmers not enrolled in these programs have no incentive to follow guidelines and may be applying P in excess of recommended rates in order to produces the highest yields. This bias is one downside of using previously collected RaCA soil samples. Using a more representative sampling throughout the watershed in order to capture non-
cooperating farms may yield more accurate results. Overall, the RaCA data is still very useful for this study. A prediction surface was generated that portrayed the distribution of P throughout the watershed. Because the model does not contain known high Mehlich 3 P values, the modeled trends should not be used to make absolute management decisions, but rather to identify areas where further investigation is necessary.

The STP values predicted also correlated with the abundance of U determined by gamma radiation. The abundance was not used a predictor variable in any model. But, because gamma radiation corresponds with parent material the correlation to STP values demonstrates the influence of soil forming factors on STP. P can enter the soil by weathering of primary minerals like apatite (Carthcart, 1980). Parent materials with high levels of apatite will inherently have more soil P than soils with lower amounts of apatite. Another dynamic between P and parent material comes from the process of P sorption onto clays. P can bind tightly to clays making the nutrient unavailable for plant uptake (Sanyal and DeDatta, 1991). Areas with higher clay contents will have more P retention, leading to lower STP values.

IK was used to determine if soils in the watershed exceed the environmental threshold called the change point. The change point is a threshold where, above it an increase in soil test P (STP) will lead to a large rise in P concentrations in runoff. Below the change point, additional input of P will not cause a significant change in the P concentrations in effluent (Kleinman et al., 2000). Change point analysis was performed separately for forest lands, agricultural lands, and then pooled for both land uses within the watershed. The change point identified for forest lands was 42 mg P kg\(^{-1}\). For agricultural lands, the change point was 64 mg P kg\(^{-1}\). For the pooled land uses, the change point was 66 mg P kg\(^{-1}\) (Figures 3.3-3.5).

IK was performed on the entire watershed for both land uses and the pooled land use (Figures 3.26-3.28). The probability of exceeding the change point can then be overlaid with the land use data to determine where there is a probability of exceeding the change point in that land use (Figures 3.26 and
The IK prediction surfaces were all reasonable because all of the RMSE values were less than 1 (Table 3.6) (Eldeiry and Garcia, 2011).

Trends in IK outputs are well explained by the clay distribution. Areas with higher clay contents have a lower probability of exceeding the change point. This may be due to the sorption potential of clays. The clay content of a soil may also influence the land use, explaining the trends in the IK probability out. Certain land uses may not be successful on clayey soils. For example, agricultural practices may fail on soils with high clay content due to the high water holding capacity of clays that can lead to improper drainage and ponding (Kern, 1995).

OK was used to determine the M3-PSR in the entire watershed. The final map produced was relatively accurate (RMSE= 2.9%) and shows the predictions of M3-PSR throughout all of the watersheds in the study area (Figure 3.29). The RMSE is higher than expected, falling between the median and the third quartile (Table 3.3) implying that there are some inaccuracies; most likely due to the variable nature of STP. This result is not accurate enough to predict M3-PSR at a specific location, but is still useful for evaluating trends (Cattle et al., 2002).

The M3-PSR prediction surface follows the same trends as the Mehlich 3 P prediction surface. This may be due to the fact that Mehlich 3 P is used in the calculation of M3-PSR. The addition of Mehlich 3 Fe and Mehlich 3 Al in the calculation increases the influence of parent material because the amount of Al and Fe present in the soil is influenced mainly by parent material (Kabata-Pendias, 2010). The land use is expected to have an influence over M3-PSR distribution due to its influence on Mehlich 3 Fe and Mehlich 3 Al. There was no trend apparent in the final predicted distribution. This may be due to the scale of the study area. It is possible that the interactions between land use and M3-PSR are not captured in the model due to the variable nature of land use.
Chapter 4. Conclusions

P losses into surface waterways and the subsequent degradation it causes are becoming an increasing concern worldwide. In fact, it is one of the major challenges facing agriculture today. Producers rely on the addition on supplemental nutrients via fertilization to insure adequate yields. Without the additional input of fertilizer, many yields would not be productive enough to maintain a livelihood. Agriculture has been identified as one of the major contributors of P to surface waterways (Kleinman et al., 2011). The addition of P in excess of the change point could lead to P losses that would impact yields.

In addition, in areas with high concentrations of animal production, manure is often applied in excess of crop needs. In these cases it is common to dispose of manure to satisfy certain application limits (USEPA, 2011). Finding areas that are less susceptible to P losses would be useful to determine areas where excess manure can be applied safely. By identifying areas susceptible to P loss, it is possible to manage land in an environmentally sound manner.

Our results showed that it is possible to identify environmental thresholds using a Mehlich 3 extraction and a 0.01M CaCl₂ extraction for forested and agricultural land uses. These change points were all at or above the agronomic optimum of 30-50 mg P kg⁻¹. This implies that most soils are not rapidly releasing P into the environment. But, with application of additional P above the agronomic optimum, a problem may arise.

Soil testing alone will not be able to accurately quantify P surface and subsurface losses via leaching because it cannot adequately describe P transport processes like subsurface preferential flow. The assessment of P loss risk should be site-specific because it needs to take certain information like hydrology and P application method into account. While soil testing may exclude some important pathways, it is an efficient and cost effective way to describe P losses in soil. Many soil testing labs use Mehlich 3 extractions to determine nutrient concentrations. Coupling these routine analyses with a 0.01M CaCl₂ extraction would provide details about the potential for P loss with minimal additional effort.
The binary nature of the change point (i.e., a site is either above the change point or not) makes it less appealing to incorporate into nutrient management systems. Wang et al. (2011) suggest using a gradation in the change point system that would offer more continuous variation that reflects the uncertainty of P loss in the field. Many other factors are included in P movement that are not included into change point analysis. Including these factors like hydrology and P sorption-desorption characteristics could produce better estimate of P leaching loss. It may also reduce the possibility of incorrectly assuming that an increase in STP leads to an increase in P loss.

We found that multiple sites did not behave this way; the P loss at one site was not always greater than the P loss of a site with a lower STP. For example, one point had a CaCl$_2$ P value of 3.6 mg P L$^{-1}$ and a Mehlich 3 P value of 44 mg P kg$^{-1}$. It would be expected that another site with a Mehlich 3 P value of 64 mg P kg$^{-1}$ would have a higher CaCl$_2$ P values compared to the actual CaCl$_2$ P value at that site (0.88 mg P L$^{-1}$). Many confounding factors may contribute to this behavior, highlighting the need to accurately characterize P transport processes in soils. Change point analysis is used to shed light onto these processes but it is not a complete answer.

Using spatial interpolation methods makes soil mapping easy and cost effective. With the increasing availability of spatial data on important environmental covariates, it is more ideal to use these methods for mapping compared with tradition transect oriented soil mapping. Understanding the various types of interpolation methods makes it easier to choose an appropriate method that will produce accurate results. Some of the important factors for choosing an interpolation method include the distribution of data, the spatial correlation of data, and the use of the final product.

Mapping the distribution of nutrients at a watershed scale is most useful in identifying areas with high potential to lose P or high potential to retain P. Kriged surfaces of M3-PSR and Mehlich 3 P should not be used to determine the value at a certain location in the study area but rather to identify trends and determine areas where values are likely to occur. For example, using the M3-PSR surface, an area can be
identified that may have a high P retention. This area could be used to make use of excess manure, but site specific M3-PSR analysis should be performed before any management decisions are made. In addition, the low Mehlich 3 P levels of soils used in the study are not indicative of actual M3P values. It is known that some areas, like the eastern panhandle of West Virginia have much higher STP levels than measured. This reinforces the concept that the modeled surface should only be used to identify trends in STP.

The lack of published work on modeling the M3-PSR implies that little is known about the distribution and dynamics of this soil property. A more in depth look at the characteristics of M3-PSR and how land use impacts M3-PSR may be useful to land managers. Creating a repeatable model to identify areas of high P retention potential would help make environmentally sound determinations about application rates.

IK is very useful for mapping soil properties that have inherent thresholds. For these properties, determining the probability that a site will exceed a threshold is more useful than determining the actual value at a site. For soil P, identifying areas that are at high risk of exceed the change point threshold should be evaluated by land managers in order to avoid the environmental impacts of P loss.

The interpolation methods outlined above are repeatable for most of the U.S. All of the spatial data used in this analysis is readily available for the entire contiguous U.S. In addition, RaCA samples for the contiguous U.S. are available. Producing the maps outlined in this study for the entire U.S. will be helpful for land managers and decision makers. But, mapping soil P over a large area has some limitations. When these limitations are acknowledged, the maps become more meaningful and useful. Land managers should understand how to use these maps in order to take advantage of this cost effective and efficient mapping technique.
Chapter 5. Implications of Mapping Soil Phosphorus

5.1 Soil Phosphorus

Phosphorus (P) is an essential nutrient for many crops. Plants use P to build and maintain cell membranes and DNA (Gbruek et al., 2000). P applications are often necessary to supply growing plants with adequate levels of P. Only about 20% of the P fertilizer applied is taken up by the first year crop (Vance, 2001). This leads to over application of P and subsequent buildup of P in the soil. The excess P in soil can be lost to surface waterways via leaching and runoff.

Water quality concerns arise because only a small amount of excess P can lead to eutrophication. Eutrophication is defined as the process by which a body of water acquires a high concentration of nutrients, especially phosphates and nitrates (Art, 1993). The introduction of excess nutrients fuels the rapid growth of aquatic plants. As these plants die and decompose, the dissolved oxygen in the water is depleted (hypoxia). This leads to the degradation of water for human consumption and recreation health as well as the decline of aquatic communities (Kleinman et al., 2011).

Sharpley et al. (1996) determined that the concentrations of P that lead to eutrophication ranges from 0.01 to 0.03 mg L⁻¹. This low eutrophication threshold means that even small amount of excess P in the soil can lead to large problems in the watershed. For example, a rain event that results in surface and subsurface water flows of 10 cm ha⁻¹ only needs to dissolve 10 g of P per ha to result in a concentration of 0.01 mg L⁻¹, a threshold for eutrophication (Taylor and Kilmer, 1980).

The relationship between the total amount of P in the soil and the amount of plant available P can be used to identify certain thresholds. The thresholds in these relationships can be identified using segmented linear regression. These thresholds are referred to as change points (Table 1.1). Above the change point, an increase in soil test P (STP) will lead to a large rise in P concentrations in runoff. Below the change point, additional input of P will not cause a significant change in the P concentrations in effluent (Kleinman et al., 2000). Keeping soils at or below the change point can help reduce the amount of P in waterways therefore avoid eutrophication and other environmental impacts.
Our results showed that it is possible to identify environmental thresholds using a Mehlich 3 extraction and a 0.01M CaCl₂ extraction. Three change points were determined; one for each land use within the study area and one for the pooled samples representing both land uses. The change point identified for forest lands was 42 mg P kg⁻¹. For agricultural lands, the change point was 64 mg P kg⁻¹. For the pooled land uses, the change point was 66 mg P kg⁻¹ (Figures 3.3-3.5). For most states that provide agronomic recommendations based on Mehlich 3 P, including Pennsylvania and Delaware, Mehlich 3 P values above 50 mg P kg⁻¹ have no plant yield response (Beegle, 1999; Kleinman et al., 2000). The calculated change points were all at or above the agronomic optimum of 50 mg P kg⁻¹. This implies that most soils have the potential to rapidly release P, especially in agricultural areas where additional P inputs are common. Reducing the agronomic optimum levels to below the change point may help mitigate P losses via leaching.

A determination of the percentage P saturation (M3-PSR) of soil was reached by finding the ratio of P to Al and Fe (from Mehlich 3 analysis). The analysis showed that a majority of the soils were not saturated with P. The soil with the highest M3-PSR is 28% saturated with P. This demonstrates that there is a potential for soils to react with Al and Fe to retain P. Sims et al. (2002) suggest that a 0.10 to 0.15 M3-PSR value be used to identify soils at risk of detrimental P losses. Most of the soils in this study were well below this range. Only 11% of soils analyzed were between or above this range. This is most likely due to high levels of Al and Fe that were found in the soil, which is most likely a function of parent material and land use.

5.2 Modeling Soil Phosphorus
The ability to visualize soil P distribution at a regional scale is helpful for keeping soil P levels below these environmental thresholds. Digital soil mapping makes it easy to predict the distribution of P. One of the fundamental theories of digital soil mapping is the model for soil formation developed by Jenny in 1941. Jenny’s model for soil development included climate, organisms, topography, parent material, and time as the basic soil forming factors. Using Jenny’s formula to describe the relationship
between soil and spatially referenced factors, the *scorpan* model was developed (McBratney et al., 2003). This model considers seven prediction factors: $s$ (soil, other properties of the soil), $c$ (climate), $o$ (organisms), $r$ (topography, landscape attributes), $p$ (parent material), $a$ (age), and $n$ (space, spatial position). The *scorpan* model is similar to Jenny’s model with the addition of the space and soil factors. The addition of the soil factor is necessary because soil can be predicted from its properties (McBratney et al., 2003).

Using the *scorpan* model as a theoretical basis, spatial modeling techniques can be used to develop digital soil maps (Grimm et al., 2009; Kumar et al., 2012; Malone et al., 2014). Using *scorpan* to guide the selection of environmental covariates expected to be related to observed soil spatial variability, spatial data can be used to predict soil properties over an area. Some of the common spatial modeling techniques include linear regression, kriging, and classification and regression trees, as well as combinations of these techniques (McBratney et al., 2003).

Using OK, CK, and IK multiple maps were created to fully explain the distribution of soil P. Lark and Ferguson (2004) determined that IK was the preferred method to predict P distribution due to the high variability of P in the landscape and the useful of knowing the probability that P exceeds certain thresholds.

Areas with a high probability of being at or near the change point were identified. Efforts to reduce the loss of P into the environment can be focused on these areas where the rapid release of P into the water is more likely. In addition, a map of the M3-PSR was developed to indicate the distribution of this characteristic. M3-PSR is related to the amount of P that a soil can retain. Areas where P retention is high can act as a sink to prevent excess P release into waterways.

Delineating areas most susceptible to P loss will make it easier focus conservation efforts on critical locations. This would make P management more time and cost efficient. Mapping the change point at the watershed scale would enable land managers to better conserve their land and protect the environment by focusing efforts on highly susceptible areas.
The goal of this effort is not to create a map of P that can be used to determine the P at a specific site. A map of this type on a regional scale would not be accurate due to the dynamic nature of P. In order to accurately determine P at a single location, a site soil test is needed (Lark and Ferguson, 2004). Even a map of P distribution at a field scale may be inaccurate due to the soil variability that is often unrelated to observable variations in soil properties (Cambardella et al., 1994; Phillips, 2001). Creating a map that indicates these high risk areas would be helpful. This map could be used by land managers to better understand the P distribution across a landscape and to identify area where susceptible to P loss via leaching.

Understanding the distribution of P is helpful as it can be used to identify areas where P application should be avoided as well as areas were additional P application presents no risks. Determining areas of potential P application is helpful as certain guidelines exist limiting the amount of manure that can be applied to land (USEPA, 2011). For producers with excess manure that cannot be applied to their land due to these guidelines, alternative application sites must be identified. Developing these large scale nutrient plans can help avoid P losses and help producers comply with federal regulations.

5.3 Future Analysis

The methods previously presented in Chapter 2 for predicting the Mehlich 3 P levels, the M3-PSR, and the probability of a soil exceeding a change point were developed in a repeatable manner. All of the spatial data used in the analysis is readily available for the entire contiguous U.S. In addition, RaCA samples for the contiguous U.S. are available from regional NRCS offices. Producing the maps outlined in this study for the entire U.S. will be helpful for land managers and decision makers. But, mapping soil P trends over a large area has some limitations. When these limitations are acknowledged, the maps become more meaningful and useful. Land managers should understand how to use these maps in order to take advantage of this cost effective and efficient mapping technique.
Additional analysis of RaCA samples would also be appropriate. A measurement of pH, calcium concentration, and organic matter content would help elucidate some of the relationships between land use and change point. These tests are straightforward and would give useful insights into P dynamics. Understanding how P dynamics differ by land use via additional investigation into the factors that have the most influence on change point would help make nutrient recommendations more accurate. In addition, field validation of model would give them more reliability. Samples collected from the study area and analyzed for P, Al, and Fe could be compared to the predicted values to validate the accuracy of the model.
The P cycle adapted from Hyland et al., 2005. P in soil is split into three fractions: an organic fraction, an insoluble inorganic fraction, and a soluble inorganic fraction. Only a small portion of P, the soluble inorganic fraction, is available to plants.
Figure 1.2. Forms of soil P by pH. P is available to plants in the form of $\text{H}_2\text{PO}_4^-$ and $\text{HPO}_4^{2-}$ (adapted from Pierzynski et al., 1994). Other forms of soil P are susceptible to loss via leaching and runoff.
Table 1.1. Summary of change point analysis by authors. Different factors are used to determine the relationship between STP and plant available P. All change point analysis has been done on agricultural soils.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Factors to Define the Relationship between STP and P Loss</th>
<th>Location</th>
<th>Study Area (acres)</th>
<th>Number of Soil Samples</th>
<th>Change Point Value (mg kg⁻¹ STP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heckrath et al., 1995</td>
<td>Olsen P vs DRP</td>
<td>U.K.</td>
<td>9</td>
<td>95</td>
<td>10-119</td>
</tr>
<tr>
<td>Hesketh and Brookes, 2000</td>
<td>Olsen P vs. CaCl₂P</td>
<td>U.K.</td>
<td>9</td>
<td>19</td>
<td>56-57</td>
</tr>
<tr>
<td>Kleinman et al., 2000</td>
<td>Mehlich 3 P /Morgan P vs. CaCl₂P</td>
<td>U.S.</td>
<td>1,529,600</td>
<td>59</td>
<td>16-144</td>
</tr>
<tr>
<td>McDowell and Sharpley, 2001</td>
<td>Mehlich 3 P /Olsen P vs. CaCl₂P/WEP</td>
<td>U.K. and U.S.</td>
<td>121</td>
<td>83</td>
<td>33-190</td>
</tr>
<tr>
<td>McDowell et al., 2001</td>
<td>Mehlich 3 P vs. CaCl₂P</td>
<td>U.K., N.Z., and U.S.</td>
<td>variable</td>
<td>variable</td>
<td>20-175</td>
</tr>
</tbody>
</table>
Figure 1.3. The relationship between M3-PSR and ammonium extractable degree of P saturation (from Kleinman and Sharpley, 2002). This relationship is well correlated using the relationship $y = 0.70x$. The high correlation means that Mehlich 3 Fe, Al, and P can be used to determine the P saturation ratio.
Figure 1.4. An example semivariogram showing the ranges, the sill, and the nugget. These three characteristics of a semivariograms can be used to choose the most appropriate kriging model.
Table 1.2. Environmental covariates used in regression tree analysis. These covariates were chosen based on their influence on STP (Section 1.5). All data are available from a variety of sources and cover the contiguous US.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>texture</td>
<td>gSSURGO&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>pH</td>
<td>gSSURGO</td>
</tr>
<tr>
<td>organic matter content</td>
<td>gSSURGO</td>
</tr>
<tr>
<td>cation exchange capacity</td>
<td>gSSURGO</td>
</tr>
<tr>
<td>land cover</td>
<td>NLCD&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>precipitation</td>
<td>PRISM&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>temperature</td>
<td>PRISM</td>
</tr>
<tr>
<td>parent material</td>
<td>USGS geologic maps</td>
</tr>
<tr>
<td>gamma-ray radiation</td>
<td>USGS aerial sensing</td>
</tr>
<tr>
<td>landscape position</td>
<td>DEM&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>elevation</td>
<td>DEM</td>
</tr>
</tbody>
</table>

<sup>a</sup> gSoil Survey Geographic Database (distributed by the USDA-NRCS)

<sup>b</sup> National Land Cover Database 2011 (distributed by the USGS)

<sup>c</sup> Parameter-elevation Regressions in Independent Slopes Model (distributed by NRCS National Water and Climate Center (NWCC))

<sup>d</sup> Digital elevation model (distributed by the USGS)
Figure 2.1. Location of the Potomac watershed in the Eastern United States. This watershed covers 9 million acres and drains directly into the Chesapeake Bay.
Figure 2.2. Location of the study area and sampling sites within the Potomac Watershed. The study area contains 23 sites sampled by the NRCS for use in quantifying soil organic carbon.
Table 3.1. Five number summary of results for agricultural sites. The median of Mehlich 3 P concentrations falls within the agronomic optimum levels (Section 3.2).

<table>
<thead>
<tr>
<th>Agricultural Sites</th>
<th>Mehlich 3 P (mg kg⁻¹)</th>
<th>CaCl₂ P (mg P L⁻¹)</th>
<th>Mehlich 3 Al (mg kg⁻¹)</th>
<th>Mehlich 3 Fe (mg kg⁻¹)</th>
<th>M3-PSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>2.3</td>
<td>0.20</td>
<td>160</td>
<td>25</td>
<td>0.60</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>10</td>
<td>0.88</td>
<td>350</td>
<td>76</td>
<td>1.7</td>
</tr>
<tr>
<td>Median</td>
<td>30</td>
<td>1.3</td>
<td>430</td>
<td>99</td>
<td>5.7</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>47</td>
<td>2.2</td>
<td>510</td>
<td>130</td>
<td>8.9</td>
</tr>
<tr>
<td>Maximum</td>
<td>110</td>
<td>7.7</td>
<td>619</td>
<td>410</td>
<td>28</td>
</tr>
</tbody>
</table>

n= 40
Table 3.2. Five number summary of results for forest sites. The Mehlich 3 P and CaCl\(_2\) P levels are lower than the agricultural land use. On average, the Al and Fe levels are higher than the agricultural land uses.

<table>
<thead>
<tr>
<th>Forest Sites</th>
<th>Mehlich 3 P (mg kg(^{-1}))</th>
<th>CaCl(_2) P (mg P L(^{-1}))</th>
<th>Mehlich 3 Al (mg kg(^{-1}))</th>
<th>Mehlich 3 Fe (mg kg(^{-1}))</th>
<th>PSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>1.2</td>
<td>0.22</td>
<td>290</td>
<td>82</td>
<td>0.08</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>8.5</td>
<td>0.89</td>
<td>940</td>
<td>150</td>
<td>0.58</td>
</tr>
<tr>
<td>Median</td>
<td>17</td>
<td>1.4</td>
<td>1100</td>
<td>260</td>
<td>1.3</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>27</td>
<td>1.9</td>
<td>1200</td>
<td>330</td>
<td>2.5</td>
</tr>
<tr>
<td>Maximum</td>
<td>51</td>
<td>5.6</td>
<td>1500</td>
<td>570</td>
<td>4.5</td>
</tr>
</tbody>
</table>

n= 51
Table 3.3. Five number summary of results for all sites. The median Mehlich 3 P level is below the agronomic optimum (Section 3.2).

<table>
<thead>
<tr>
<th>All Sites</th>
<th>Mehlich 3 P (mg kg(^{-1}))</th>
<th>CaCl(_2) P (mg P L(^{-1}))</th>
<th>Mehlich 3 Al (mg kg(^{-1}))</th>
<th>Mehlich 3 Fe (mg kg(^{-1}))</th>
<th>PSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>2.3</td>
<td>0.20</td>
<td>160</td>
<td>25</td>
<td>0.35</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>10</td>
<td>0.89</td>
<td>440</td>
<td>99</td>
<td>1.2</td>
</tr>
<tr>
<td>Median</td>
<td>24</td>
<td>1.4</td>
<td>590</td>
<td>150</td>
<td>2.3</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>34</td>
<td>2.2</td>
<td>1100</td>
<td>290</td>
<td>4.6</td>
</tr>
<tr>
<td>Maximum</td>
<td>110</td>
<td>7.7</td>
<td>1500</td>
<td>570</td>
<td>28</td>
</tr>
</tbody>
</table>

\(n=91\)
Figure 3.1. Correlation between Mehlich 3 P and the environmental covariates described in Section 1.5. Mehlich 3 P has a relationship with multiple factors including CEC, elevation, slope position, and K abundance.
Figure 3.2. Correlation between CaCl₂ P and the environmental covariates described in Section 1.5. CaCl₂ P followed many of the same trends as Mehlich 3 P having a relationship with multiple factors including CEC, elevation, slope position, and K abundance.
Figure 3.3. The relationship between Mehlich 3 P and CaCl₂ P in forest land use. A change point is located at the significant slope break, 42 mg P kg⁻¹, fitted using a split-line model.
Figure 3.4. The relationship between Mehlich 3 P and CaCl$_2$ P in agricultural land use. A change point is located at the significant slope break, 64 mg P kg$^{-1}$, fitted using a split-line model.
Figure 3.5. The relationship between Mehlich 3 P and CaCl\(_2\) P in pooled land use. A change point is located at the significant slope break, 66 mg P kg\(^{-1}\), fitted using a split-line model. Agricultural sites are represented by a blue circle while forested sites are represented by a green triangle.
Figure 3.6. The relationship between Mehlich 3 P and CaCl\(_2\) P in forest soils described by a simple linear regression. There is a positive relationship between the variables (R\(^2\) = 0.34).
Figure 3.7. The relationship between Mehlich 3 P and CaCl\textsubscript{2} P in agricultural soils described by a simple linear regression. There is a positive relationship between the variables ($R^2 = 0.62$).
Figure 3.8. The relationship between Mehlich 3 P and CaCl$_2$ P for pooled land uses described by a simple linear regression. There is a positive relationship between the variables ($R^2 = 0.51$).
Table 3.4 Summary of how well the different regression models fit the data. RMSE is used as a measurement of accuracy. In this case the RMSE for the three different data sets were equal, meaning that there was no significant difference in the regression models.

<table>
<thead>
<tr>
<th>Regression Method</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forested Land Use</td>
</tr>
<tr>
<td>Segmented Linear Regression</td>
<td>42</td>
</tr>
<tr>
<td>Simple Linear Regression</td>
<td>42</td>
</tr>
</tbody>
</table>
Figure 3.9. Correlation between M3-PSR and the environmental covariates described in Section 1.5. M3-PSR has a relationship with multiple factors including CEC, precipitation, elevation, and slope position.
Figure 3.10. Relationship between the M3-PSR and Mehlich 3 P concentration. The variables are correlated ($r^2= 0.80$). This is most likely due to the use of Mehlich 3 P in the determination of M3-PSR.
Figure 3.11. Relationship between 0.01M CaCl$_2$ P and M3-PSR. The variables are correlated ($r^2 = 0.50$).
Figure 3.12. Regression tree output of Mehlich 3 P concentrations using environmental covariates (RMSE=14 mg P kg$^{-1}$). Important variables have longer tree legs representative of the inter-group variance. The important covariates determined were CEC, geology, surface roughness, and slope position.
Figure 3.13. Regression of Mehlich 3 P values predicted from regression tree and observed values. The predicted values and the observed values are correlated ($r^2 = 0.58$), meaning that the regression tree model did a well explaining the Mehlich 3 P data.
Figure 3.14. Regression of M3-PSR values predicted from regression tree analysis and observed M3-PSR values. The predicted values and the observed values are correlated ($r^2 = 0.67$), meaning that the regression tree model did a well explaining the M3-PSR data.
Figure 3.14. Regression tree output using M3-PSR data and environmental covariates (RMSE = 2.9%). Important variables have longer tree legs representative of the inter-group variance. The important covariates determined were sand, K abundance, and moisture.
Table 3.5. Moran’s I index analysis output for Mehlich 3 P and M3-PSR data. Points were determined to be spatially autocorrelated.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Moran's Index</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mehlich 3 P</td>
<td>0.51</td>
<td>6.2</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>M3-PSR</td>
<td>0.72</td>
<td>8.8</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>
Figure 3.16. Distribution of raw Mehlich 3 P data. The mean is higher than the median, indicating that the data is skewed to the right.
Figure 3.17. Distribution of logarithmic transformed Mehlich 3 P values. The mean and median are approximately equal, showing that the transformation normalized the data.
Figure 3.18. Distribution of the raw M3-PSR. The mean is double the median, indicating that the data is highly skewed to the right.
Figure 3.19. Distribution of logarithmic transformed Mehlich 3 P values. The mean and median are approximately equal, showing that the transformation normalized the data.
Figure 3.20. The normal Q/Q plot of the Mehlich 3 P values. The values do not follow a straight line meaning that the data deviates from normal.
Figure 3.21. The normal Q/Q plot of the logarithmic transformed Mehlich 3 P values. The values follow a relatively straight line meaning that the data is approximately normal.
Figure 3.22. The normal Q/Q plot of the M3-PSR values. The values do not follow a straight line meaning that the data deviates from normal.
Figure 3.23. The normal Q/Q plot of the logarithmic transformed M3-PSR values. The values follow a relatively straight line meaning that the data is approximately normal.
Table 3.6. Summary of interpolation methods used in study. Methods with lower with lower RMSEs are more accurate. Only RMSE determined using the same prediction parameters can be compared.

<table>
<thead>
<tr>
<th>Prediction Parameter</th>
<th>Interpolation Method</th>
<th>Covariates</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mehlich 3 P</td>
<td>Ordinary Kriging</td>
<td>---</td>
<td>18 mg P kg⁻¹</td>
</tr>
<tr>
<td>Mehlich 3 P</td>
<td>Co-Kriging</td>
<td>CEC, Geology</td>
<td>17 mg P kg⁻¹</td>
</tr>
<tr>
<td>42 mg P kg⁻¹ change point threshold</td>
<td>Indicator Kriging</td>
<td>---</td>
<td>0.35</td>
</tr>
<tr>
<td>64 mg P kg⁻¹ change point threshold</td>
<td>Indicator Kriging</td>
<td>---</td>
<td>0.18</td>
</tr>
<tr>
<td>66 mg P kg⁻¹ change point threshold</td>
<td>Indicator Kriging</td>
<td>---</td>
<td>0.15</td>
</tr>
<tr>
<td>M3-PSR</td>
<td>Ordinary Kriging</td>
<td>---</td>
<td>2.9%</td>
</tr>
</tbody>
</table>
Figure 3.24. Mehlich 3 P value predictions throughout the watershed using co-kriging as the interpolation method with geology and CEC as the covariates. The classes were determined based on agronomic optimum values. The optimum P value ranges from 30-50 mg P kg\(^{-1}\).
Figure 3.25. Mehlich 3 P value predictions throughout the watershed using ordinary kriging as the interpolation method. The classes were determined based on agronomic optimum values. The optimum P value ranges from 30-50 mg P kg$^{-1}$. 

81
Figure 3.26. The probability that forested soils will exceed the change point identified for forest land (42 mg P kg$^{-1}$). The classes of low probability, medium probability and high probability were classified using a quantile method.
Figure 3.27. The probability that agricultural soil will exceed the change point for agricultural lands (64 mg P kg$^{-1}$). The classes of low probability, medium probability and high probability were classified using a quantile method.
Figure 3.28. The IK output for the study area. The output is the probability that a location will exceed the change point determined for the pooled land uses soils, 66 mg P kg\(^{-1}\). The classes of low probability, medium probability and high probability were classified using a quantile method.
Figure 3.29. M3-PSR prediction surface for soils within the watershed of interest. Data is broken into classes based on a quantile classification where low contains sites with 0.36-1.6%, the medium class contains sites between 1.6-6.7%, and the high class contains sites between 6.7-28%. Only sites in the high class are at risk of detrimental P loss via leaching.
References


