QUANTIFYING THE "NUTRIENT LANDSCAPE" IN THE GREAT LAKES REGION: MAPPING AND ANALYZING NUTRIENT SOURCES AND GROUNDWATER NITRATE

By

Quercus Florence Hamlin

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Environmental Geosciences—Master of Science

2020

ABSTRACT

QUANTIFYING THE "NUTRIENT LANDSCAPE" IN THE GREAT LAKES REGION: MAPPING AND ANALYZING NUTRIENT SOURCES AND GROUNDWATER NITRATE

By

Quercus Florence Hamlin

Since the mid-19th century, the rise of industrial agriculture and growing population has significantly altered nutrient cycling. These changes are from multiple sources, such as chemical fertilizers, livestock waste, and human waste. Excess nutrients have led to a suite of water quality problems that damage human and animal health, ecology, and economics. In this thesis, I begin to quantify the "Nutrient Landscape", a term I use to refer to the set of processes and properties that drive cycling of nitrogen and phosphorus throughout our modern environment.

To understand the "Nutrient Landscape", I first develop algorithms utilizing broadly available data to estimate nutrient inputs from seven distinct sources across the U.S. portion of the Laurentian Great Lakes Basin at 30 meter resolution. Chapter I's mapping effort, referred to as the Spatially Explicit Nutrient Source Estimate map (SENSEmap), provides new information for management and modeling, as well as a classification system to categorize watersheds based on their nutrient source composition. Second, I examine the groundwater component of the "Nutrient Landscape" by exploring a dataset of over 300,000 nitrate samples from drinking water wells using Classification and Regression Tree (CART) analysis to determine drivers of elevated concentration. This analysis revealed high nitrate concentrations result from a combination of hazardous land use and vulnerable geology. The data products and findings in this thesis provide a quantitative framework for informing management strategies and driving the next generation of nutrient modeling.

ACKNOWLEDGEMENTS

The work involved in my Master's thesis spans a four year period with support from people all along the way. This thesis represents more than a body of research, but a transformative path from which I found purpose and friendship. I want to acknowledge the tremendous support from my friends, family, and lab, to which there is significant overlap.

First and foremost, the kindness, positivity, and support from my home in the Hydrolab got me to where I am. Starting as an undergraduate in the lab in 2015, I quickly gained the responsibility of what would make up my first chapter and inspire me to pursue graduate degrees in water quality. My work on SENSEmap started with the support of my supervisors and mentors, Drs. Anthony Kendall and Sherry Martin. Early in my lab career, they trusted me with the responsibility and challenge of an intense coding and GIS project. Anthony and Sherry's mentorship shaped the scientist I am today from an early place. Their support while I was an undergraduate continued into my graduate career starting at Syracuse University and later welcoming my return to MSU. Additionally, their friendship made it all a humble, warm experience.

I thank my committee, Drs. David Hyndman, Anthony Kendall, and Lisa Tiemman for their guidance and insight. I thank my primary advisor Dr. David Hyndman for his optimism – often a double-edged sword for his students, but overall a positive trait. Particularly at the end of my Master's degree during the COVID-19 Pandemic, he was flexible and supportive while working to dispel my negativity about my work. I am excited to continue working with Dave throughout my PhD. It seems so long ago, but as an undergraduate I was intimidated by him, unsure what to expect as we began the paper process.

The friends I made in the Hydrolab brought joy to graduate school and my life – these are friendships I will cherish throughout my life. I thank my officemates Brent Heerspink, Bailey Hannah, Luwen Wan, Jake Roush, and Chanse Ford for the great atmosphere in the "Boat Room" as well as my other friends and labmates: Alex Kuhl, Ally Brady, Behnaz Mirzendehdel, Ben McCarthy, Erin Haacker, Jake Stid, Jeremy Rapp, Jill Deines, Joe Lee-Cullin (honorary Hydrolabmate), Kyle Cole, Leanne Hancock, and Mohammed Rahman. Some special shoutouts: At the end of my Master's, during mandatory quarantine, I thank Brent and Alex for their daily Zoom companionship that ensured I completed my thesis on time and stayed sane. This crisis brought us closer and without our friendship I am not sure I would have made it through. I thank Jeremy for his friendship since I started in the lab with his humor, intellectual banter, and empathy as a friend and colleague. To Luwen and Ally, thank you for your positivity and light as dear friends. Thank you to my student coauthors on the SENSEmap manuscript, Henry Whitenack, Jake Roush, and Bailey Hannah – those revisions were a marathon. Thank you to Kyle and Yvonne, the undergraduates I am mentoring, for giving me the joy of teaching and letting me be a part of your growth.

Thank you as well to the Department of Earth and Environmental Sciences staff who make the logistics of academia run: Ami McMurphy, Pam Robinson, Brittany Walter, Elizabeth McElroy, and Judi Smelser.

Outside of the university, I thank my friends, particularly Briar Fraleigh, Kenzie Eddy, and Julia Cousino for getting together for all the brunch dates, concerts, and late nights. There would be no success without the wonderful vacations taken to see Nico Cabanayan, Bev Yockelson, Tom Quinn, and Cynthia Santos – my long-distance, wonderful friends who have always let me crash on couches and helped revive me. You all let me ramble about my science

and the frustrations inherent to graduate school. I can't wait to see you again when travel becomes possible again.

Finally, I thank my parents, Fern and Chris Hamlin, for your support throughout my journey. To my dad – thanks for the bike rides, women's basketball fanaticism, and helping me move so many times. You also got me into science with my first research job at Notre Dame with Dr. Jason McLachlan, Jill Deines, and Jody Peters. This start pushed me forward. The last thank you, a critical one, goes to my therapist for the last two years at the DBT Institute of Michigan. The skills I learned and support received were life changing, allowing me to grow into a life I want to have.

Thank you all. It has been a wild ride and I am excited for what comes next.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

CART	Classification and Regression Trees
CDL	Cropland Data Layer
EPA	Environmental Protection Agency
LP	Michigan's Lower Peninsula
N	Nitrogen
NANI	Net Anthropogenic Nitrogen Index
NIL	Nutrient Input Landscape
NLCD	National Land Cover Database
NO_3-N	Nitrate as Nitrogen
P	Phosphorus
SENSEmap	Spatially Explicit Nutrient Source Estimate map
SPARROW	SPAtially Referenced Regressions on Watershed attributes
TN	Total Nitrogen
TP	Total Phosphorus
USDA	United States Department of Agriculture
USGS	United States Geological Survey
US-GLB	United States Great Lakes Basin
WWTP	Wastewater Treatment Plant

CHAPTER 1:
QUANTIFYING LANDSCAPE NUTRIENT INPUTS WITH SPATIALLY EXPLICIT NUTRIENT SOURCE ESTIMATE MAPS

Abstract

Nutrient management is an essential part of watershed planning worldwide to protect water resources from both widespread landscape inputs of nutrients (N and P) and point source emissions. To provide information to regional watershed planners and better understand nutrient sources, we developed the Spatially Explicit Nutrient Source Estimate Map (SENSEmap) to quantify individual sources of N and P at their entry points in the landscape. We modeled seven sources of N and six sources of P across the United States Great Lakes Basin at 30m resolution: atmospheric deposition, septic systems, chemical non-agricultural fertilizer, chemical agricultural fertilizer, manure, nitrogen fixation, and point sources. By modeling these sources, we provide a more detailed view of nutrient inputs to the landscape beyond what would be possible from land use alone. We found 71% and 88% of N and P, respectively, came from agricultural sources. The nature of agricultural nutrient inputs varied significantly across the basin, as relative contributions of chemical agricultural fertilizers, manure, and N fixation changed according to diverse land use practices regionally. We then applied k-means cluster analysis and identified nine Nutrient Input Landscapes (NIL) with N and P source characteristics, grouped into intensive agricultural, urban, and rural landscapes. These NILs can offer insights into landscape variability that land use data alone cannot; within agricultural NILs, application of chemical fertilizer and manure varied greatly, but land uses were similar. These NILs can provide a framework for broadly categorizing watersheds that may prove useful to both ecological and management practices.

1. Introduction

Agricultural intensification, rising population, and increased meat consumption have led to a massive increase in landscape nutrient inputs worldwide. Anthropogenic nitrogen and phosphorus inputs dominate global nitrogen (N) and phosphorus (P) cycles, and drive eutrophication of both inland and coastal waters (Howarth, 2008; Smith et al., 1999; Anderson, 2002). Beyond fundamentally altering ecosystems, these high loads of N and P have fueled harmful algal blooms (HABs), which directly harm human and ecosystem health, and bring negative economic consequences (Paerl & Otten, 2013; Carmichael, et al. 2001; Dodds et al. 2009; Brooks et al. 2016). The need to manage anthropogenic inputs of nutrients has become clear, however a lack of detailed information regarding where and how nutrients are applied hinders both science and management.

Nutrients are applied to the landscape through a host of natural and anthropogenic mechanisms, few of which can be effectively measured, and even fewer of which are actively monitored. Further complicating the issue is that where present, federal and provincial level statues exempt non-point source inputs from reporting requirements. For instance, the US Clean Water Act tightly regulates point source loads from wastewater treatment plants and other facilities but allows unreported nutrient applications for agricultural and municipal land management (Clean Water Act, 1972). Lacking detailed data, scientists and watershed managers commonly turn to nutrient modeling as a primary tool to understand and quantify nutrient inputs, especially to aid in the management of Total Maximum Daily Loads (TMDL), a part of the US EPA's Clean Watershed Act (US EPA, 1991).

Since nutrients are poorly monitored, modeling has become a major tool to understand non-point source nutrient fluxes within watersheds. Land use is commonly used to determine

nutrient loads, as distinct sources can be estimated by land uses. An "export coefficient" based on land use may be used, or land use may be an input within a more mechanistic transport framework. USGS SPAtially Referenced Regressions on Watershed attributes (SPARROW) uses a variety of watershed characteristics to fit a regression model calibrated with measured in stream loads and transported through a stream network (Smith et al., 1997; Alexander et al., 2004). Other methods include the Net Anthropogenic Nitrogen Index (NANI) which correlates riverine N flux with four variables: synthetic fertilizer, agricultural N fixation, atmospheric deposition, and net movement of human and animal food and has recently been applied to all counties in the United States (Swaney et al., 2018; Howarth et al., 2012; Howarth et al., 1996). Nutrient sources and land use are also inputs to process-based models. For example, the ELEMeNT (Exploration of Long-tErM Nutrient Trajectories) modeling approach uses land use and nitrogen surplus calculations (atmospheric deposition, biological nitrogen fixation, chemical fertilizer, and manure) within a legacy and transport framework (Van Meter et al., 2017). A commonly used process-based model, SWAT (Soil and Water Assessment Tool) represents different nutrient source input rates through parametrized land use types (Arnold et al., 1998).

However, these approaches often make broad assumptions about the loading rate from a single land use. This has been identified as a concern in regional USGS SPARROW models (Preston et al., 2011), and becomes especially problematic in agricultural landscapes where a variety of crops are grown and fertilized with different amounts and types of fertilizer. This is an issue for quantifying the environmental effects of different fertilizing systems, as manure and chemical fertilizers have different nitrous oxide emissions (a potent greenhouse gas) and nitrate leaching to groundwater (Basso & Ritchie, 2005; Tuomisto et al., 2012; Charles et al., 2017). Additionally, local data about septic systems and point sources have been shown to improve

modeling when added to coarser national data (Scown et al., 2017). Approaches that use explicit sources, like NANI, generally estimate loads at large watershed or county scales and are used to predict riverine outputs (Han & Allan, 2008; Hong et al., 2013; Goyette et al., 2016), rather than informing specific changes in local management.

Here we describe the first version of the Spatially Explicit Nutrient Source Estimate Map (SENSEmap) product, constructed using an approach that seeks to avoid compromises frequently made between spatial extent, resolution, and detail of nutrient specificity. We then demonstrate the approach for the U.S. portion of the Laurentian Great Lakes Basin (US-GLB), a region that contains both inland lakes and Great Lakes representing relatively pristine to highly eutrophic conditions, and where HABs are increasingly inflicting ecological, human health, and economic harm. With SENSEmap-US-GLB, we estimate landscape nutrient inputs for seven N and six P sources at 30 m resolution. Sources include atmospheric deposition, septic systems, chemical non-agricultural fertilizer, chemical agricultural fertilizer, manure, nitrogen fixation from legumes, and point sources.

SENSEmap-US-GLB was developed from and expands upon previous research in Michigan's Lower Peninsula where six nutrient sources were quantified and transported using a statistical model to estimate riverine exports (Luscz et al., 2015; 2017). Highlights of improvements to the method include updating and improving all nutrient source modeling methods, adding N Fixation, and upscaling our analysis from Michigan's Lower Peninsula to all eight Great Lakes States. SENSEmap outputs include pixel-based nutrient inputs aggregated to the HUC12 watershed scale. We also identify nine distinct Nutrient Input Landscapes (NILs), which are sets of watersheds most similar in their quantity and composition of nutrient sources.

These NILs reduce a complex set of 13 spatially varying nutrient sources into a straightforward framework that can be linked to other indicators of water quality and ecological condition.

2. Materials and Methods

2.1 Study Area

The North American Laurentian Great Lakes Basin (GLB) consists of five basins surrounding Lakes Superior, Michigan, Huron, Erie, and Ontario, covering 580,000 square kilometers in the United States and Canada. HABs are found throughout the world and across the American Great Lakes region, most notably in the large annual blooms in western Lake Erie (Figure 1.1). The HABs in Lake Erie have been particularly dangerous due to toxins in cyanobacteria that cause neurological and liver damage to humans and animals, resulting in Toledo's 2014 drinking water ban (Carmichael et al., 2001; Michalak, et al. 2013; Fitzsimmons, 2014). Intensive use of chemical agricultural fertilizer and manure are major causes of modern high nutrient loads (Withers et al., 2014). Although Lake Erie is a prominent example, eutrophication and HABs are a problem at all scales, and problems as large as those found in Lake Erie result from mismanagement of landscapes in the contributing watershed.

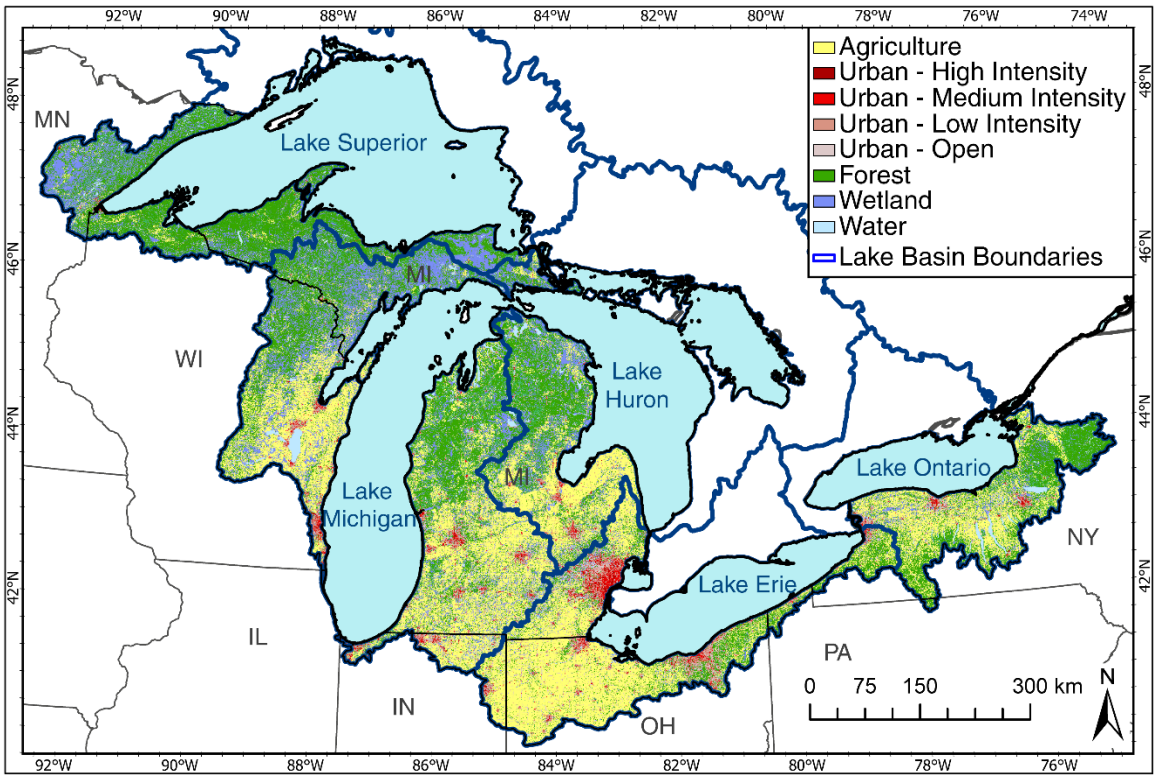

Figure 1.1 SENSEmap study area. Land use/land cover in the U.S. Great Lakes Basin from NLCD 2011. Nonurban areas use the Anderson level 1 description, while urban areas are differentiated using Anderson level 2 description.

This study describes the development of SENSEmap-US-GLB, which quantifies average annual nutrient inputs for the 2008-2015 period over the United States Great Lakes Basin (US-GLB) portion of eight states: Indiana, Illinois, Michigan, Minnesota, New York, Ohio, Pennsylvania, and Wisconsin. This expansion of nutrient input maps was possible due to readily available data in the US, including census, agricultural census, and land use datasets. Future work may extend these maps to the Canadian portion of the basin and develop temporal estimates of landscape nutrient inputs.

Climatologically, the US-GLB is temperate and sub-humid, and includes Koeppen-Geiger zones Dfa and Dfb (hot-summer and warm-summer humid continental climates, respectively). Annual average (1981 - 2010) temperatures range from 3 - 10° C (PRISM).

Annual precipitation is distributed relatively uniformly between 700 and 1000 mm, with local areas receiving up to 1500 mm/yr. A significant annual snowpack accumulates during the winter months, melting intermittently during the season in the southern part of the basin and often persisting until early spring in the northern parts of the basin.

The US-GLB exhibits a north-south gradient of land use: in the north, typically above 43 to 45 degrees N, population density is low and land use/cover (LULC) is primarily forest and wetlands, with some low intensity agriculture (Figure 1.1). In the central and southern basin, intensive agriculture and urban centers dominate. Ohio, Indiana, and southern Michigan primarily grow corn and soybean in rotations with a single annual growing season. Alfalfa and other hay crops are also grown as a primary crop in northern regions. Winter wheat and other small grains like rye and barley are also common. Specialty crops include potatoes, apples, grapes, blueberries, dry beans, cherries, squash, sugar beets, asparagus, carrots, tomatoes, and cucumbers. Livestock agriculture is also common, primarily chickens, dairy cows, cattle, hogs, chickens, and turkeys. Wisconsin notably has intensive dairy operations (USDA Ag Census, 2012).

The Great Lakes Basin contains Detroit, Michigan; Milwaukee, Wisconsin; Cleveland, Ohio; and Buffalo, New York, as well as other mid-size cities and many rural agricultural towns. Population density is higher in the southern and central GLB, as there are more major cities and agricultural towns. Although cities and many towns have wastewater treatment plants, septic systems are still heavily used within suburbs, small towns, and rural populations. Septic systems are a dispersed, but commonly neglected, nutrient source that loads directly to groundwater.

The last glacial period and subsequent melt deposited coarse-textured sediments across much of the basin. Exceptions to this are areas of lake plain clays that formed in pro-glacial

lakes, now covering some of the areas adjacent to lakes Huron and Erie. As a result, soil textures vary from coarse sands and gravels to fine textured soils with greater than 90% clay particles. This variability in soil textures is relevant to what crops are grown, how much fertilizer is applied, and eventually how much of those applied nutrients are used by crops versus those that are mobilized via runoff, erosion, or deep percolation into groundwater systems.

2.2 Data Sources

SENSEmap quantifies nutrient inputs at their origin on the landscape at a 30 m resolution to match the resolution of the National Land Cover Dataset (NLCD, Homer et al., 2015). The model uses data from 2008 to 2015 to produce average annual nutrient inputs estimates for this timespan. Those data sources are listed in Table 1.1 and are described in further detail for each nutrient source in the following section. Core datasets for SENSEmap include the US Census (population and households), USDA Agricultural Census (livestock, crops), USDA Cropland Data Layer (remote sensing product locating major crops), and the NLCD (land use/land cover).

Table 1.1 Summary of SENSEmap-US-GLB data sources. Data sources, time period, and resolution for each nutrient source.

Description	Dataset(s)	Source(s)	Year(s)	Resolution	Used For
Land Cover, Imperviousness	National Land Cover Dataset (NLCD)	Multi-Resolution Land Characteristics Consortium (MRLC)	2011	30 m	All
N Atmospheric Deposition	Total Deposition Maps (Total N, wet+dry)	National Atmospheric Deposition Program (NADP) Schwede & Lear (2014)	2008-2015	4134.354 m	Atm. Dep.
Soil Properties	SSURGO	National Resources Conservation Service	n/a	Vector/30 m	N Fixation, Chem. Ag. Fertilizer, Manure
Crop Locations	Cropland Data Layer (CDL)	USDA NASS	2008-2015	30 m	N Fixation, Chem. Ag. Fertilizer, Manure
Crop Yields	Agricultural Census	USDA NASS	2007, 2012	County	N Fixation, Chem. Ag. Fertilizer, Manure
Chemical Fertilizer Quantities	County level chemical fertilizer	USGS: Brakebill & Gronberg (2017), Gronberg & Spahr (2012)	2008-2012	County	Chem. Ag., Chem. NonAg. Fertilizer
Remotely-sensed Greenness	LANDSAT	NASA, Accessed via Google Earth Engine	2008-2015	30 m	N Fixation
Animal/Farm Inventories	Agricultural Census	USDA NASS	2012	County	Manure
Confined Animal Feeding Operations	State Databases			Point	Manure
Incorporated Areas, Census Blocks	TIGER Line Files	US Census	2010	Polygon	Septic
Drinking Water Well Locations	State Databases			Point	Septic
Regulated Point Sources	National Pollutant Discharge Elimination System (NPDES) Reports	USEPA Discharge Monitoring Report	2008-2015	Point	Septic, Point Sources
Household Information	US Census	US Census, IPUMS	2010	Tabular	Septic, Chem. NonAg Fertilizer
Golf Course Locations	Original	Google Earth imagery	2008-2015	Polygon	Chem. NonAg Fertilizer

2.3 Mapping Nutrient Sources

Seven nutrient sources are included in SENSEmap as illustrated in Figure 1.2, which shows the different landscape processes associated with each source. SENSEmap's high spatial resolution allows these processes to be captured at their origin, rather than averaged over a watershed or county. SENSEmap's individual source inputs are data-driven and vary at 30 meter resolution. Quantification and spatial location of each input source is done via one of four sets of methods: 1) summarizing reported loads at known locations for point sources, 2) interpolation of observed data for atmospheric deposition, 3) spatial disaggregation of county-level estimated inputs for chemical agricultural and non-agricultural fertilizers, and 4) summarizing, locating, and disaggregation of "population"-based inputs, from a variety of coarser scales for CAFOs, septic systems, non-CAFO manure, and N fixation.

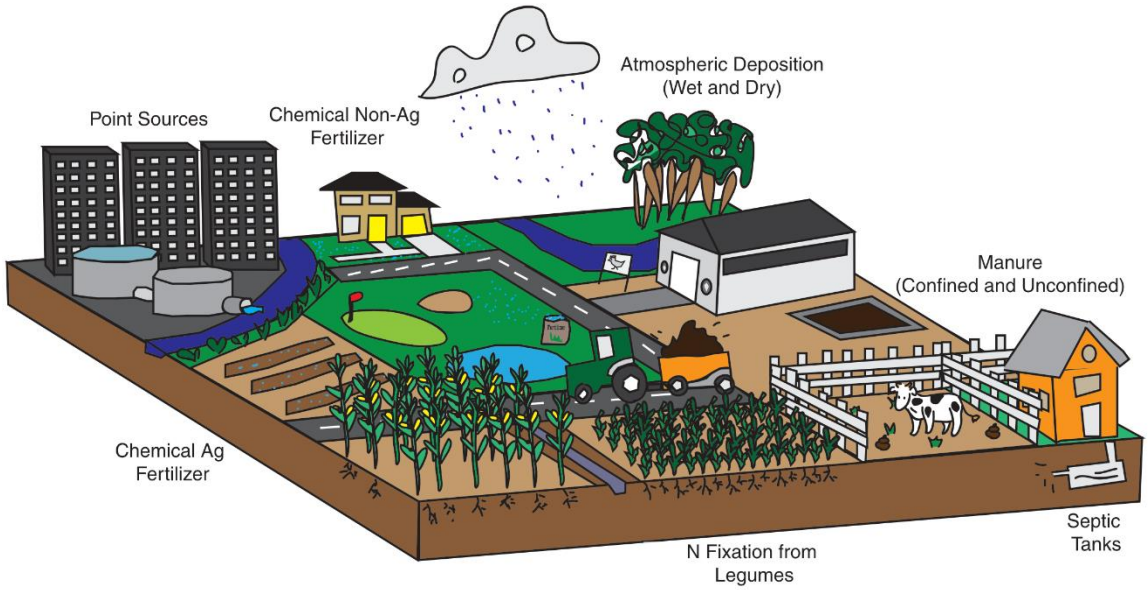

Figure 1.2 Schematic diagram of SENSEmap nutrient sources on the landscape. Sources pictured include (clockwise from upper left): point sources, chemical non-agricultural fertilizer, atmospheric deposition (wet and dry), manure (confined and unconfined), septic tanks, N fixation from legumes, and chemical agricultural fertilizer.

For both N and P, the individual SENSEmap sources are quantified as mass of Total N (TN as kg) or Total P (TP as kg) across all organic and inorganic forms. Each of the individual sources consists of multiple species, which are summed according to source-specific procedures. For all non-point sources, inputs are reported as annual area-specific inputs (kg-N/ha/yr or kg-P/ha/yr). Point source loads are quantified as annual mass fluxes (kg-N/yr or kg-P/yr).

For nutrient accounting, we considered all significant sources of nutrients added directly to the landscape. For example, all nutrients from animal excretion could become applied manure (although some N is lost to the atmosphere prior to application, as described below), even though some of those nutrients likely came from harvested N and P within the Great Lakes Basin. By contrast, properly-maintained septic systems are regularly pumped out, thus removing a significant quantity of the nutrients that are then commonly land-applied. Here we omit land application of septage and assume that all human-excreted nutrients deposited into septic tanks are applied at the septic location.

In the follow sub-sections, we describe key details about the seven nutrient sources. Much more detailed information, including flow charts illustrating the complete workflow for each source are included in the Appendix Figures A1.3-A1.10.

2.3.1 Atmospheric Deposition

Atmospheric loading of nitrogen and phosphorus occurs through both wet and dry deposition and can be the primary loading mechanism for either nutrient in remote areas. The main components of phosphorus emitted to the atmosphere are dust from soils, marine aerosols, volcanic ash, biomass burning, and combustion of oil and coal. Agricultural activity and phosphate manufacturing and mining may contribute to higher rates of localized deposition in some areas. Fossil fuel combustion and high intensity agriculture are the main sources of

nitrogen to the atmosphere. Wet deposition is measured by collecting precipitation in a sampler and analyzing for nitrogen and phosphorus compounds. The volume of collected precipitation is then multiplied by the analyzed concentrations of nitrogen and phosphorus compounds and corrected for sampler cross-sectional area. Dry deposition is more challenging to directly measure, so rates are estimated by modeling particle deposition velocities of measured atmospheric concentrations of compounds (USEPA, 2010).

Wet deposition rates are generally driven by large scale processes and can be significantly correlated between sites separated by large distances (Anderson and Downing, 2006): therefore, all sites selected were within 700 km of the Great Lakes Basin. Large phosphorus particles can move over short distances while finer dust may be able to travel over thousands of kilometers (Tipping et al., 2014) and a 700 km buffer was chosen for atmospheric phosphorus analysis as well.

Atmospheric phosphorus concentrations are much lower relative to other macronutrients making accurate detection difficult (Mahowald, 2008). As a result, atmospheric monitoring is limited and nutrient accounting models rarely include it within their budgets (David & Gentry, 2000). Luscz et al. (2015) used four sites from a single monitoring network to interpolate P deposition. To expand this analysis, we identified all comparable P deposition data through an extensive literature review (see Appendix Text 1.1 for a detailed description of this review process). Through this review, we identified 23 papers and 1 monitoring network containing data from 98 sites reporting total wet and dry TP deposition within a 700 km radius of the GLB from 1970 to 2011 (see complete references in Table A1.1). We assumed that P deposition rates have remained largely stable through time. This assumption is consistent with a review by Tipping et al. (2014) who found that P deposition rates at sites with ten or more years of continuous data

shows that significant systematic changes are uncommon. This is supported by our longest continuous set of atmospheric phosphorus data collection which showed no significant trend in annual deposition (Eimers et al., 2009). In contrast, both emission and subsequent deposition rates of atmospheric nitrogen during the past four decades largely declined on national and regional scales (Monks et al., 2009; Kothawala et al., 2011).

Due to the limited available data points and their inherent distribution across the landscape, available phosphorus data was interpolated using kriging to develop spatially distributed loading estimates. IDW and kriging interpolation methods were tested to find the most appropriate model for both nutrients. Interpolation methods were tested and selected based upon the parameters which best minimized the root mean square error (see Text A1.1). Phosphorus is mostly collected through bulk precipitation methods (both dry and wet fallout) and as such loading estimates were modeled one time as total phosphorus.

SENSEmap uses the National Atmospheric Deposition Program (NADP) Total Deposition Science Committee (TDEP) product for total nitrogen deposition. TDEP estimates atmospheric nitrogen deposition rates using a hybrid approach combining modeled concentrations with measured concentrations (Schwede and Lear, 2014). The approach estimates dry concentrations using data from the Clean Air Status and Trends Network (CASTNET), the Ammonia Monitoring Network (AMoN) and the SouthEastern Aerosol Research and Characterization (SEARCH) network combined with modeled concentrations from the Community Multiscale Air Quality (CMAQ) model. The dry deposition value estimates are then combined with wet deposition values from the National Trends Network (NTN) to develop total deposition values of nitrogen including organic nitrogen. The TDEP model is useful for regional, broad-scale deposition mapping (Bytnerowicz et al., 2016; Sicard et al., 2016). SENSEmap uses

a mean of 2008-2015 values from gridded TDEP maps to capture average annual total nitrogen deposition for that timeframe (ftp.epa.gov/castnet/tdep).

2.3.2 Septic Systems

Onsite wastewater treatment systems, known as septic tanks, are widely used in rural and suburban areas. They can be considered point sources on small scales, but on a regional scale they are best represented as a non-point sources due to unknown locations, high densities, and small loads. While septic tanks are an urban source of nutrients, they are a direct source of nutrients to groundwater and therefore transport differs significantly from surface-applied urban nutrients such as non-agricultural chemical fertilizer.

Our septic source model consists of three elements: 1) creating an exclusion mask defined by service area boundaries of wastewater treatment plants (WWTP) within which no septic systems are placed, 2) overlaying an inclusion mask based on land use and appropriate set back distances, and 3) calculating septic system numbers and loading rates from US Census data and literature sources. Determining exclusion areas required extensive data synthesis to model WWTP service areas, as these are not publicly available at the scale of the GLB. Drinking water wells, population density, and distance to WWTPs were used to classify service areas. Inclusion areas were selected based on land use, using primarily urban, non-WWTP service areas. Within each Census block in inclusion areas, one septic system was assumed per household unit. These septic systems were first placed adjacent to well locations, and then randomly within the primary inclusion mask of each Census block if there were more reported household units than known wells. Septic systems were placed at densities less than 1 per 675 square meters (equivalent to ⅙ acre, a normal household lot size in the US) and were secondarily applied to non-urban areas near roads as needed. See Text A1.2 for more information.

Septic nutrient discharge was assumed to occur at the septic system location and vary across Census blocks according to average household size and vacancy. We did not consider pumping for off-site disposal of septage solids--an omission that would not change total loaded nutrients within a watershed but would affect the spatial placement. For specific nutrient rates, we used the EPA's Onsite Wastewater Treatment Systems Manual estimate of 4.1 kg/year/person nitrogen and 1 kg/year/person phosphorus (USEPA, 2002). These values were then multiplied by average household size at the Census block level and reduced to account for household units that are designated seasonal or vacant. Within each Census block, the final per-septic load was applied to all placed septic systems. These loading rates were then rasterized and summed at the 30 m cell size.

2.3.3 Chemical Non-agricultural Fertilizer

Non-agricultural chemical fertilizers, which are primarily applied to lawns and golf courses, are major sources of nutrients in urban areas. We developed a two-part approach to spatially disaggregate county-level estimates of total non-agricultural chemical fertilizer sales from 2008-2012, where 2012 was the latest available year (Brakebill & Gronberg, 2017), to the 30 meter scale. Within each county, fertilizer applications were placed on golf courses, followed by distribution of the remaining nutrients to suburban and urban lawn areas.

Golf courses, which generally have large and frequent fertilizer applications, were first identified using state-specific golf course business directories. Golf courses were first batch identified based on two general golf business directories with golf course addresses in each GLB state. When available, additional state and county specific golf course business directories were used to find courses that may not have been identified in the first sweep. To create a complete golf course inventory, we further located golf courses not included in business directories

through systematic visual inspection of 2008 - 2015 aerial imagery across the Great Lakes Basin. The golf courses listed in the business directories were then address-matched in ArcGIS to locate them. For each golf course, polygons were manually delineated for each course. We analyzed the average proportion of fairway and green to total golf course area (which includes cart paths, sand traps, roughs, and water features) and estimated the fertilizable area to be 85% for this region. We then randomly applied nutrients at local (Midwest) golf course fertilizer application rates (N: 117 kg/ha/yr; P: 10.7 kg/ha/yr) estimated by a survey from the Environmental Institute for Golf to golf course raster cells (GCSAA, 2009) until the total fertilized area matched the estimated area within each the course.

Following golf course applications, remaining nutrients from the county-level sales data (Brakebill & Gronberg, 2017) was applied to urban lawns. As in Ruddy et al. (2006), we assumed that areas with higher population density are more likely to apply lawn fertilizer. In other words, more lawns and businesses choose to fertilize when there are more people nearby. In aggregate, this increases fertilizer application rates per Census tract area; here we assumed that rates increase linearly to a maximum of 4996 kg N/km^2 within a tract at a population density of 700 persons/km^2. This method was also used in Luscz et al. (2015). Remaining county fertilizer was distributed among tracts and placed randomly within urban cells between 10 and 100 m from roads, assuming that fertilizer would be primarily applied to lawns near roads. Rates within each selected cell were given by average recommended residential application rates for low-maintenance established lawns given by the University of Minnesota Extension (N: 73.2 kg/ha/yr, P: 10.7 kg/ha), and sufficient cells within each tract were selected to match the tract-level application rate (Rosen et al., 2015). Finally, total cell input amounts were corrected based on the cell percent perviousness (NLCD 2011, Homer et al. 2015).

2.3.4 Point Sources

Point source data was acquired through EPA's Discharge Monitoring Report (DMR) Pollutant Loading Tool (US EPA, 2017), which combines National Pollutant Discharge Elimination System (NPDES) permit data with DMR data to describe the nature and amount of discharge from each point source. NPDES observes and regulates effluent information for facilities that discharge to surface waters in the United States. Depending on the facility and its permit, reporting parameters, techniques, and frequency can vary, making the data difficult to use without harmonization. As such, we downloaded data for each state within the Great Lakes Basin for each year from 2008 to 2015. Depending on the facility, nitrogen or phosphorus is occasionally reported as something other than simply N or P (such as NH_3, NO_3-N, and PO_4). In these situations, we converted the reported value based on mass balance equations, so all data were represented as either nitrogen or phosphorus on a mass basis. Most sources did not report every year, so to arrive at a representative value for 2008-2015 we calculated a geometric mean for each point source across the eight years.

2.3.5 Agricultural Fertilizer Sources

The majority of the nutrient demand for agricultural crops are met by applications of manure and chemical agricultural fertilizers. Crop-specific fertilizer demand was only used for N, with values scaled by intensity and area normalized using recommended rates for Michigan (Warncke et al., 2004; Warncke & Dahl, 2003 – see Table A1.3 for values). P requirements are commonly determined based on local soil tests and desired crop yield rather than simply crop type, so we did not include a P crop demand variable within the fertilizer demand model (Warncke et al., 2004). We assumed that if a local source of manure is available, farmers will choose manure over chemical agricultural fertilizers due to lower costs, or the need to spread

manure associated with larger animal feeding operations, thus leading to high rates of nutrient input (Long et al., 2018). The process for determining which cells would have manure applied are described in the manure section. Since farmers often apply additional N fertilizer to manured fields, we applied N chemical agricultural fertilizer to pixels where manure had not met the pixel's N demand (Long et al., 2018). The cells outside of manure spreading areas then received chemical agricultural fertilizer as described below.

Quantities of manure and chemical agriculture N fertilizer applications were both based on a per-crop pixel fertilizer demand model. This was first parameterized using county-level data, then applied at the pixel scale to disaggregate county-level fertilizer totals. The model considered four county-level averaged variables: latitude, soil texture, cropland area, and crop-specific nutrient demand. The county level was chosen to match the resolution of USGS fertilizer use data (Brakebill & Gronberg, 2017). We used gradient boosted regression trees (BRT) in *scikit-learn*, a machine learning technique that creates many regression trees (models that separate predictors by binary splits) and has better predicting power at extremes than CART models (Elith et al., 2008; Pedregosa et al., 2011). The BRT models produced good fits to county data (N: $R^2 = 0.93$; P: $R^2 = 0.87$), suggesting that the model could handle a wide range of predictor values at the cell level. We then applied the BRT using per-pixel values for the four variables to predict per-pixel fertilizer demand, which was then used as a starting estimate for manure and chemical fertilizer applications. Additional information on the methods used to model fertilizer demand is provided in Text A1.3, including additional details on our BRT application.

2.3.5.1 Manure

We used multiple data sources to estimate manure application areas with animal-specific and nutrient-specific loading rates. This procedure consists of four steps: 1) determine a complete animal inventory including counts, types, and confinement status for all farms; 2) calculate how much manure is produced at each farm, quantify N volatilization during storage, and determine the total acreage to which manure should be applied; 3) place the farms within each county; and 4) create a buffered area around each farm upon which the manure will be spread.

Key details about each of these steps are discussed below, with additional information in Text A1.4. Animal inventories and confinement status for all farms were estimated using a combination of USDA Ag Census data and state-level records. These records do not allow us to precisely determine the size of each farm, but rather to designate appropriate numbers of farms in different size intervals within each county. Manure is then produced by each animal at specific loading rates, either provided by literature or estimated via an empirical relationship based on average animal size (for less common animal types). Some N is lost to the atmosphere during manure storage, which is estimated for each farm and animal type separately. An additional small fixed ratio of manure is lost from CAFOs due to incomplete recovery. Farms are then placed within each county, at actual known locations for large Confined Animal Feeding Operations and pseudo-randomly within agricultural areas of each county for smaller farms. Finally, a buffer is iteratively created around each farm to which manure is applied on both row crop and grassland land use types. Within each buffer, manure is applied at relative rates given by the BRT fertilizer demand model, adjusted to match total farm nutrient production following losses.

2.3.5.2 Chemical Agricultural Fertilizer

Following manure application, chemical agricultural fertilizer was applied to all non-manured cells at the N and P determined rates from the BRT demand model. Manured cells that had not met N demand received supplemental N fertilizer. Brakebill and Gronberg (2017) calculated chemical agricultural fertilizer use by county based on observed values and a fertilizer spending model for states without records up to 2012. We used the average value from 2008-2012 as our observed county fertilizer load and scaled non-manured fertilized pixels within each county to match observed county values. Additional information on chemical agricultural fertilizer methodology can be found Text A1.1.5.

2.3.6 Nitrogen Fixation

Symbiotic nitrogen fixation is an important component of nitrogen budgets in agricultural regions. Legumes such as beans, hay, and clover host microbes within their root systems that fix nitrogen that becomes incorporated both within plant biomass and surrounding soil at rates varying by crop yield, soil, and climatic conditions (Goolsby et al., 1999; Barry et al., 1993; Meisinger & Randall, 1991). In regions with significant legume cultivation, omitting N fixation from budgets would significantly underestimate agricultural N source inputs to the landscape. Because of this, we included four nitrogen fixing crops that are commonly grown in the GLB: soybeans, dry beans, alfalfa, and non-alfalfa hay. Natural N fixation was not included, as high natural N fixing biomes (savannah, grassland) are not present across significant areas of the GLB, while forested biomes almost entirely recycle N (Cleveland et al., 2013).

Here we calculate nitrogen fixation as a function of crop type, yield, soil conditions, and local fertilization rates following methods developed in other studies. In the literature, both area-based and yield-based calculations have been used to estimate N fixation rates. Area-based

calculations use a fixed rate per unit area of cropland (Han & Allan, 2008; Boyer et al., 2002). Yield-based estimates can be based on either aboveground-only or total system biomass (Meisinger & Randall, 1991). Above-ground yield-based estimates rely on percentages of grain to total plant mass, protein content in grain, and the proportion of above-ground accumulated N harvested (nitrogen harvest index) (David & Gentry, 2000; David et al., 1997). These authors used a constant rate of nitrogen from fixation based on data from Illinois for extents as broad as the Mississippi River Basin, irrespective of yield variability (David et al. 2010). As relatively cool temperatures and shorter growing seasons limit yields in much of the US-GLB, we chose to account for heterogeneity in crop N fixation using a yield-dependent approach following the method of Han and Allan (2008), which was developed from Meisinger and Randall's (1991) work. This provides estimates of fixed nitrogen mass per unit crop corrected for typical crop-specific moisture rates (i.e., the ratio of harvested dry biomass to total harvest mass) based on a range of published field studies. Meisinger and Randall (1991) also provide guidelines to estimate the percent of total plant nitrogen from fixation based on soil nitrogen availability and other nitrogen inputs such as fertilization.

We consider N fixation from soybeans, dry beans, alfalfa, and non-alfalfa hay, which are the primary N fixing crops grown in the US-GLB region. Details of the N fixation model, including how we statistically estimated yield using soil type and remotely-sensed greenness, are provided in Text A1.1.6. To account for crop rotation, especially the common corn-soy rotation, we adjusted per-cell values based on how frequently the cell was planted in soy between 2008 and 2015 using the CDL. As a satellite product, the CDL contains inaccuracies and errors in area due to mixed pixels (cells with one class that may have multiple classes within them). To correct for area inaccuracy, we used ancillary data from the USDA Ag Census 2012 to adjust total

cultivated areas, as recommended by Lark et al. (2017). Using this yield-based method, our N fixation values varied based on yield, soil, N applications, and percent organic matter. As a result of crop-rotation based averaging across the 2008-2015 model period, final values for average annual N fixation are lower than fixation during a single fixing crop year. The resulting model considers more heterogeneous landscape features and integrates multiple remote sensing and spatially explicit datasets.

2.4 Analytical Methods

After producing the 30 m rasters for each source, we analyzed these rasters individually and in their spatially aggregated forms to identify patterns and compare to other data. Three scales of aggregation were used: Great Lakes catchment (US-side only), Hydrologic Unit Code 8-digit (HUC8), and HUC 12-digit (HUC12). Within those catchments, a variety of aggregate quantities were calculated, including totals of each source, totals across N and P, and ratios of N:P inputs within aggregation features. Finally, we used k-means clustering to identify watersheds with similar nutrient input profiles, which we call Nutrient Input Landscapes (NILs).

SENSEmap results were summarized at the HUC12 watershed level, which average 79 km^2 in the US-GLB. By aggregating to a fine watershed scale, we can understand local patterns in nutrient inputs at a scale relevant to management. Additionally, aggregating removes any potential privacy issues that may exist with 30 m outputs (i.e., farms). Furthermore, HUC12 is the spatial framework within the NOAA Great Lakes Basin Tipping Point Planner, a multi-model decision support system to aid watershed planners (tippingpointplanner.org). The Tipping Point Planner is being used to help Great Lakes communities understand the health of their watersheds.

2.4.1 k-means Clustering and Nutrient Input Landscapes

We grouped the resultant 3627 HUC12 watersheds based on their nutrient input composition using k-means clustering to identify similar watersheds across the US-GLB. K-means clustering is a commonly used clustering algorithm that minimizes the sum of squared error among a specified number of groups (Jain, 2010). We explored the effect of different numbers of groups (k) and different ways of explaining a watershed's nutrient loads (rate of input by source (kg/ha/yr), total input from source (kg), and percent of total nutrient input by source). The goal was to find physically meaningful patterns in the landscape, rather than fit a specific predictive model. Our final model included 9 groups (k=9) generated from 13 variables (% of total nutrient input by source for all N and P) described watersheds in the most physically meaningful way and produced the highest mean silhouette score (measure of similarity of k-groups) (mean silhouette score = 0.39; silhouette plot and further cluster information found in Text A1.2). We labelled nine Nutrient Input Landscapes post-hoc based on the composition of sources, land use, and input rates within the groups (See section 3.3).

2.4.2 Model Comparisons

We compared SENSEmap nutrient inputs to three other products: 1) The USGS SPAtially Referenced Regressions on Watershed attributes (SPARROW) model (Smith et al., 1997), 2) Net Anthropogenic Nitrogen Index (NANI) inputs from Swaney et al., (2018), and 3) 5 km annual fertilizer maps from Cao et al. (2018). SPARROW provides both a nutrient inputs database, in some ways comparable to the SENSEmap product here, along with regressions of nutrient uptake within watersheds. We compared chemical agricultural fertilizer, chemical non-agricultural fertilizer, and total manure to SPARROW inputs from Robertson and Saad's (2011) SPARROW model including the Great Lakes Basin using data processed by Wieczorek and

Lamotte (2011). We summarized these sources from SENSEmap at the HUC8 level to be consistent with SPARROW output summaries. NANI from Swaney et al. (2018) quantifies key nitrogen inputs and exports at county level across the continental US. We summarized SENSEmap at the county level and normalized to county area to compare to chemical agricultural fertilizer, total manure, and N fixation. Cao et al. (2018) produced 5 km resolution maps of nitrogen fertilizer applications from 1850 to 2015 that depict crop-specific rates, application timing, and ammonium-N and nitrate-N proportions. SENSEmap was aggregated to 5 km cells using mean input intensity and compared to the average value of Cao et al. (2018) between 2008-2012 to match SENSEmap's fertilizer time period. Comparisons are discussed in Section 3.4 and additional details in Text A1.4.

2.4.3 Sensitivity of Chemical Agricultural Fertilizer

We performed a sensitivity analysis of SENSEmap chemical agricultural fertilizer inputs to better understand the role of this important nutrient source in controlling broader landscape patterns. For this analysis, county level totals from Brakebill & Gronberg (2017) were adjusted to create fourteen new simulations addressing three different types of uncertainty: systematic bias in 1) positive or 2) negative direction, and 3) random uncertainty. Following the methodology described above to produce the chemical agricultural fertilizer values, these county-level values were used to adjust pixel level predictions. For each of the fourteen simulations, N and P were aggregated at the HUC12 watershed level with variation between runs quantified via coefficient of variation. Watersheds in each simulation were then assigned to Nutrient Input Landscapes using the k-means clusters defined from the primary run. Changes between Nutrient Input Landscape classifications for each of the simulations and the primary run

were then tabulated. Further description of sensitivity analysis methods can be found in A1.5 and discussion of the sensitivity results can be found in Section 3.5.

3 Results and Discussion

3.1 SENSEmap-US-GLB

The primary SENSEmap-US-GLB product is composed of seven N sources and six P sources, with each N source input map shown in Figure 1.3 (P sources are visually similar in distribution, see Figure A1.13). Generated at 30 m resolution, the full variability within each map is difficult to visualize in printed form. Within Figure 1.3, we provide an inset for each source showing an enlarged area surrounding Toledo in western Ohio to better illustrate the high-resolution 30 m outputs. The full raster datasets, along with the HUC12 and HUC8 summaries, are available for download from Hydroshare (https://doi.org/10.4211/hs.1a116e5460e24177999c7bd6f8292421).

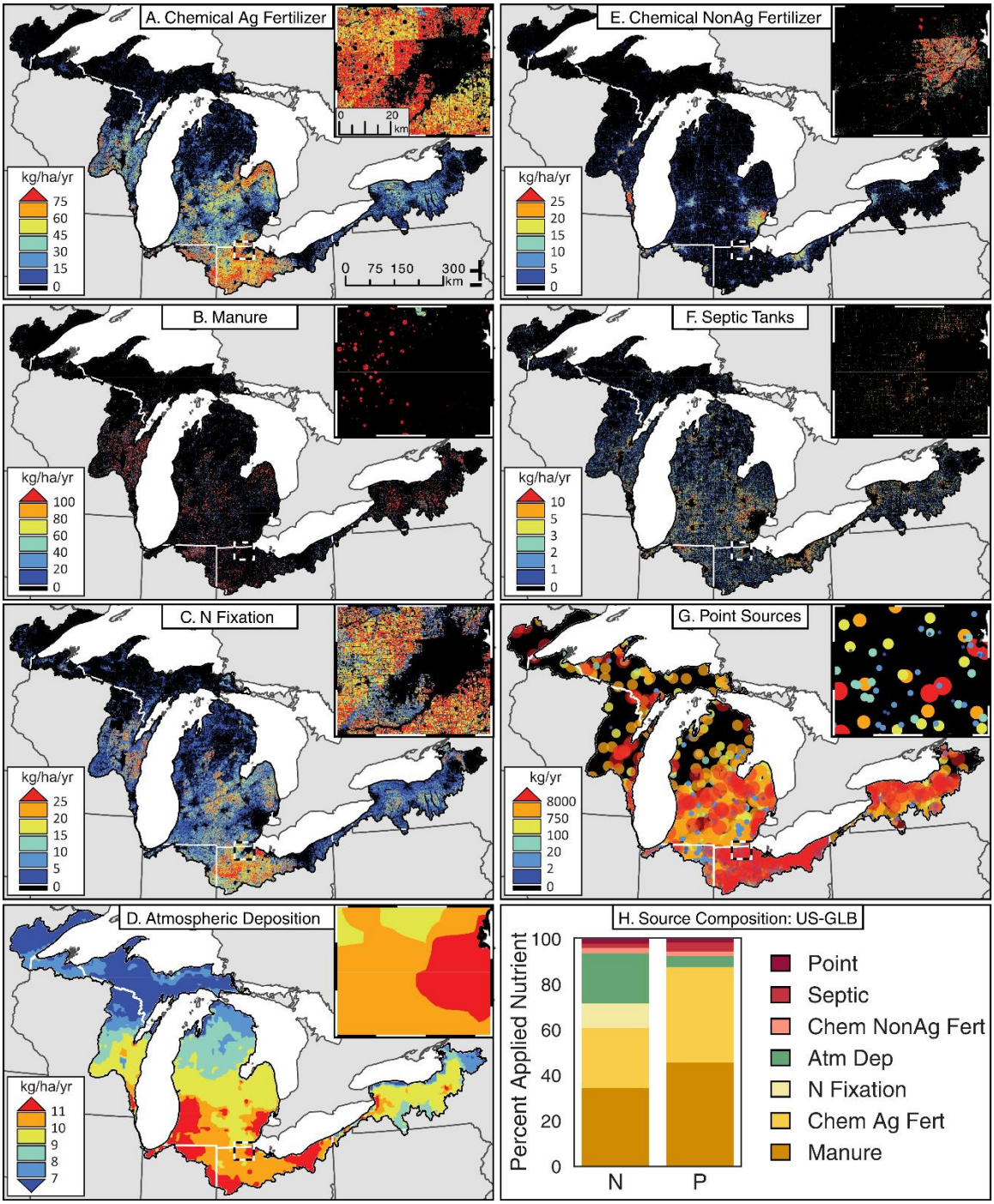

Figure 1.3 SENSEmap nitrogen sources for the U.S. Great Lakes Basin. The color bar breaks were selected to highlight variability within each nutrient source at this map scale. Refer to Table A1.5 for pixel-level distributions. At the top right of each subfigure is a zoom box, with location identified in (A). (A) Chemical agricultural (Ag) fertilizer, (B) Manure, (C) N fixation, (D) Atmospheric deposition, (E) Chemical nonagricultural (NonAg) fertilizer, (F) Septic tanks, (G) Point sources, and (H) Proportion of sources to total inputs of each nutrient within the U.S. Great Lakes Basin (US-GLB).

Agricultural sources (chemical agricultural fertilizer, manure, and N fixation) dominate nutrient inputs at the scale of the US-GLB. Chemical agricultural fertilizer makes up 34% of N and 45% of P inputs, while manure contributes 27% of N and 42% of P (See Figure 1.3H). Across the GLB, chemical agricultural fertilizer use is most intense in western Ohio and eastern Indiana, as well as in Michigan's thumb region. While agriculture is also common in the southern lower peninsula of Michigan, Wisconsin, and New York, nutrient input intensities are lower in these areas. Nitrogen fixation shows similar patterns as chemical agricultural fertilizer due to the common practice of corn-soy rotations. Nitrogen fixation makes up 11% of N inputs in the US-GLB, making it an important source to estimate in nutrient budgets. The landscape patterns of manure appear differently than chemical agricultural fertilizer and N fixation due to model assumptions that manure is spread close to the farms and CAFOs where it is generated. This creates circular pockets of high intensity manure loads as is apparent in the inset manure map. Although there are farms and CAFOs in Ohio, Indiana, and Michigan, more livestock operations are present in Wisconsin and New York causing higher intensity manure inputs.

Urban sources are generally smaller contributors to total inputs but can be major sources in localized areas. Chemical non-agricultural fertilizer and septic tanks represent similar proportions of nutrient input in the US-GLB (chemical non-ag. fertilizer N: 2%, P: 2%; septic N: 2%, P: 2%) but have very different landscape patterns. Cities serviced by WWTPs do not have septic system inputs, but nearby suburbs and rural areas have more densely populated areas, leading to higher septic input intensities. Although septic systems do not amount to a high proportion of total nutrient input within the US-GLB, they load distinctly differently than surface inputs, supplying nutrients directly into groundwater pathways. Furthermore, septic systems are often present in areas lacking other sources (except for atmospheric deposition), and they tend to

cluster near water bodies where their nutrient loads may contribute to lake eutrophication (Beal et al., 2005; Holman et al., 2008). Chemical non-agricultural fertilizer shows an opposite landscape pattern to septic: proximal suburbs and urban cores have high intensity input due to extensive lawn and golf course fertilizer application. Similarly to septic systems, chemical non-agricultural fertilizer can be an intense local nutrient input, even if its relative proportion is low at the US-GLB scale. Point sources are highest in urban areas with large or multiple WWTPs, with reporting requirements varying by state. As is the case with septic systems, point sources generally load directly to streams. Thus, fewer mechanisms for nutrient removal exist, and transport times are shorter--magnifying the impact of this relatively small proportion of total inputs on in-stream loads and deliveries to the Great Lakes.

Atmospheric deposition follows a gradient based on population and agricultural LULC. The highest rates of atmospheric deposition are found in heavily farmed and more densely populated areas in the southern basin, while lower values are found in the more sparsely populated and forested northern basin. However, in these northern areas, atmospheric deposition may be the dominant nutrient input source. For N, atmospheric deposition is a significant source across the US-GLB, accounting for 22% of total inputs, while just 5% of applied P comes from atmospheric deposition.

SENSEmap-US-GLB's non-point sources (atmospheric deposition, septic systems, chemical non-agricultural fertilizer, chemical agricultural fertilizer, manure, nitrogen fixation from legumes) were combined into maps of total non-point N and P (Figure 1.4). Figure 1.4 is classified in quantiles to show the distribution of input across the US-GLB, with each color representing 20% of cells within the domain. N and P show similar trends in the highest loading quantiles (red and orange) across the intensely farmed southern US-GLB. However, N and P

show differences in the northern basin. While broadly similar, patterns in low loading quantiles in Michigan's upper peninsula vary from N to P; in N, there is a distinct input increase from west to east, whereas P shows a less distinct pattern and more variability. This is driven by the greater relative importance of atmospheric deposition to landscape N inputs than for P. Also, cells with higher quantiles of N inputs are concentrated toward to the lower portion of Michigan's lower peninsula than for P.

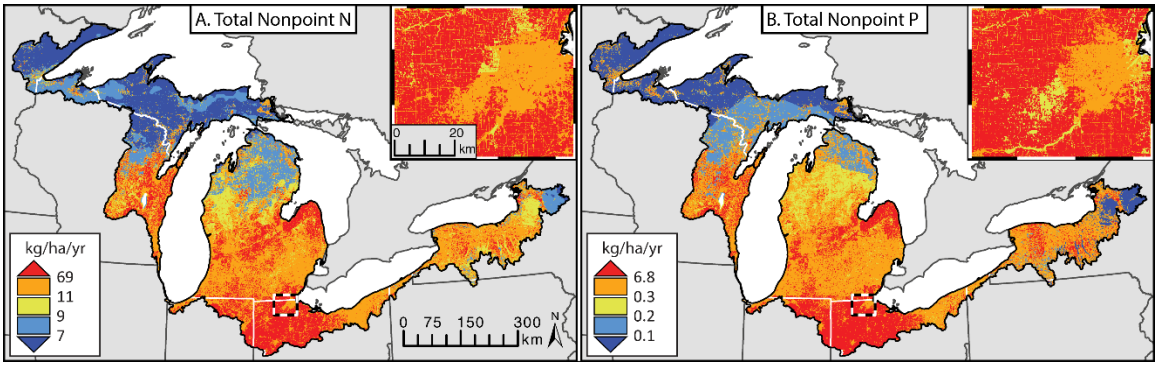

Figure 1.4 Total nonpoint N and P. (A) Total nonpoint nitrogen and (B) Total nonpoint phosphorus. The color breaks are classified to assign ~20% of pixels to each nutrient input rate bin. Point sources are excluded because they are applied as a mass flux, rather than an area-specific mass flux of the nonpoint sources.

We analyzed nutrient source compositions (Figure 1.5A-B) and input intensity (Figure 1.5C) at individual lake basin scales to better understand the heterogeneity across the US-GLB. Such metrics can be useful to understand the relative possible nutrient sources to a lake and the highest possible loading. Recall, these sources and input rates refer to the landscape inputs and are not estimates of a total load to the lakes. Strikingly, landscape's that drain to Lakes Erie, Huron, Michigan, and Ontario receive ~70-75% N and ~80-90% P input from agricultural sources (chemical ag fertilizer, manure, N fixation). Non-agricultural human sources (point, septic, chemical non-ag fertilizer) make up less than 10% of lake basin inputs except P inputs to Lake Superior, which receives 11% loading from septic and an additional 10% of loading from point and chemical non-ag fertilizer combined. Lake Superior receives a more balanced P input

composition than any other lake, but has an atmospheric deposition-dominated N load. However, as seen in Figure 1.5C, Lake Superior's input rates are very small compared to other more agricultural and populated lake basins. Despite similar source compositions in all lake basins (excluding Superior), Lake Erie has a catchment loading rate ~1.5x higher than Lakes Huron, Michigan, and Ontario. This high input rate and intense agricultural contribution supports the accepted understanding of Lake Erie's HAB problem. It is notable that in the 45 years since the passage of the Clean Water Act, point source loads have dropped sufficiently that they no longer dominate nutrient inputs. In contrast to the success of reducing point sources, non-point source inputs remain high as they are not regulated in a similar manner. An additional concern with high input rates of non-point sources is the presence of land use legacies that can result in decades of delay before improvements to surface water concentrations are observed (Meals et al., 2010, Martin et al., 2011, 2017; Verhougstraete, et al. 2015; Ray et al. 2012; Van Meter & Basu, 2015).

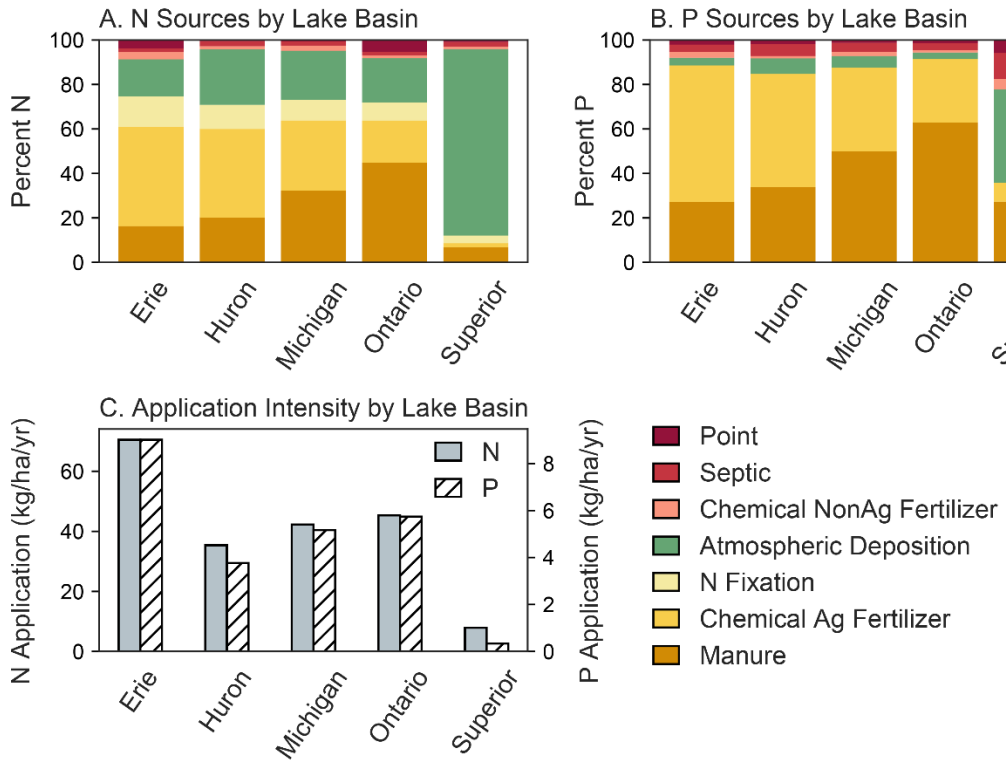

Figure 1.5 Summaries of sources and application intensity at the lake basin scale. (A) Proportion of N input from each source by lake basin, (B) Proportion of P input from each source by lake basin, and (C) Input intensities by lake basin (area normalized by the land area of the U.S. lake basin.

3.2 N:P Ratios

Nitrogen to phosphorus (N:P) ratios can be used to help understand nutrient inputs and their potential ecological effects. The Redfield Ratio is a ratio of N:P typical of aquatic life and is commonly near 16:1 in natural waters. High N:P ratios within the water column are associated with oligotrophic, non-anthropogenic impacted lakes; whereas, low N:P ratios are associated with mesotrophic to eutrophic lakes with higher anthropogenic inputs (Downing & McCauley, 1992). For comparison, we calculated N:P ratios of nutrient inputs at the HUC12 watershed scale to provide a view of N:P inputs in the US-GLB watersheds (Figure 1.6A). When viewed this way, the northern basin is highly N-enriched (and thus P-poor), with a 51:1 ratio of nutrient inputs to Lake Superior's US basin (Figure 1.6B). This is consistent with studies that have

described Lake Superior as an extreme, low P environment (Sterner, 2011). Lake Superior's N source is largely from atmospheric deposition (Figure 1.5A) and shows a much higher N:P landscape input ratio than the other Great Lakes (Figure 1.5C). Generally, and perhaps surprisingly, more southern watersheds have input ratios within range of the Redfield Ratio (pale yellow), with urban and low agricultural intensity areas being more N-enriched than intensive agricultural areas. Few watersheds had lower ratios than 12:1, and those that did were frequently high manure watersheds. These values provide a first look at nutrients relative to one another the lake basin scale; however, due to varied transport rates of different sources, these values are not representative of loads to the lake.

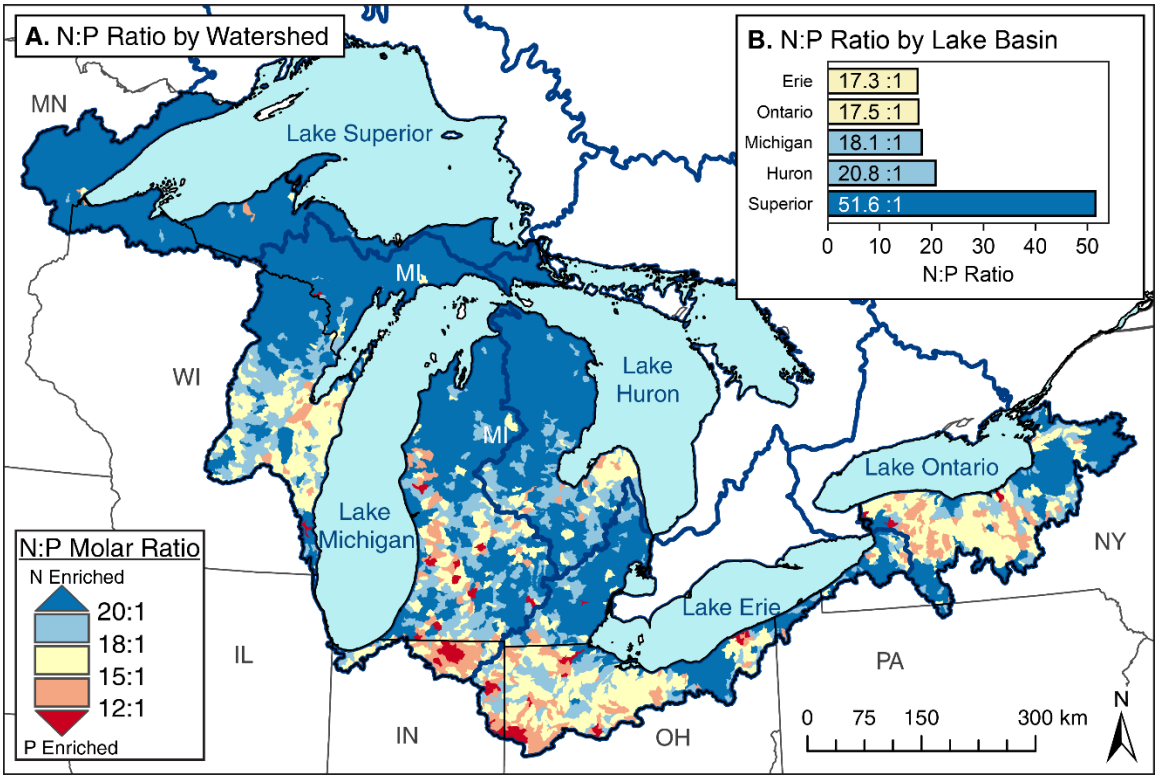

Figure 1.6 SENSEmap N:P Ratios. (A) Map of N:P input ratio by HUC12 watershed and (B) Bar chart of N:P input ratio by US portion of each lake basin.

3.3 Nutrient Input Landscapes

SENSEmap-US-GLB at its finest scale shows 30 m cell values of nutrient inputs; however nutrient inputs are a result of wider landscape processes that are managed at watershed scales. By zooming out to the HUC12 watershed scale, nutrient input patterns become more readily visible. Following k-means clustering (section 2.4.1), nine Nutrient Input Landscapes (hereafter, “landscapes”) were classified, and then labelled after clustering according to the nutrient and land use characteristics of each cluster. Those 9 landscapes of the US-GLB can then be grouped into three broad categories: intensive agricultural, urban, and rural landscapes (Figure 1.7). Names for each of the landscapes are provided in the Figure 1.7 legend and landscape definitions visualized through watershed source distributions are shown in Figure 1.8. For summary statistics of each landscape definition, see Table A1.1.4.

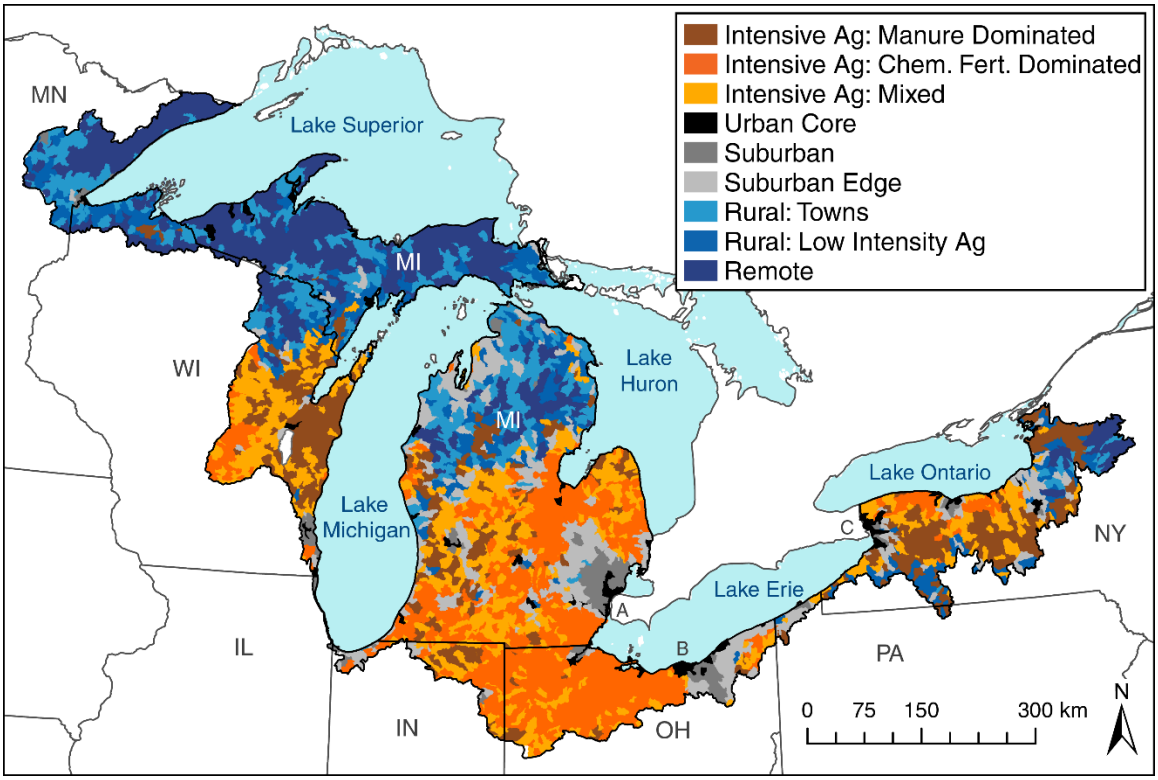

Figure 1.7 Nutrient Input Landscapes. Nutrient Input Landscapes determined from *k*-means clusters and labeled post hoc according to cluster characteristics. Letters indicate major cities, A: Detroit, MI; B: Cleveland, OH; C: Buffalo, NY. Landscapes, as described in the legend, include three intensive agricultural groups (manure dominated, chemical agricultural fertilizer dominated, and mixed), three urban groups (urban core, suburban, and suburban edge), and three rural groups (towns, low intensity agriculture, and remote). Landscapes are colored to be identifiable by group, with intensive agricultural landscapes in warm colors, urban landscapes in greyscales, and rural landscapes in blue.

The intensive agricultural group contains three separate landscapes, clustered based on proportions of chemical agricultural fertilizer versus manure: Manure Dominated, Chemical Fertilizer Dominated, and Mixed (Figure 1.8 A-C). These landscapes are found primarily in the south and central US-GLB (Figure 1.7). Large areas in Wisconsin and New York as well as multiple smaller areas in Michigan are dominated by manure, whereas northern Indiana and Ohio and much of the central Lower Peninsula of Michigan are dominated by chemical agricultural fertilizer. Agricultural HUC12 watersheds at the periphery of the Manure and Chemical Fertilizer Dominated landscapes tend to fall into the Mixed class.

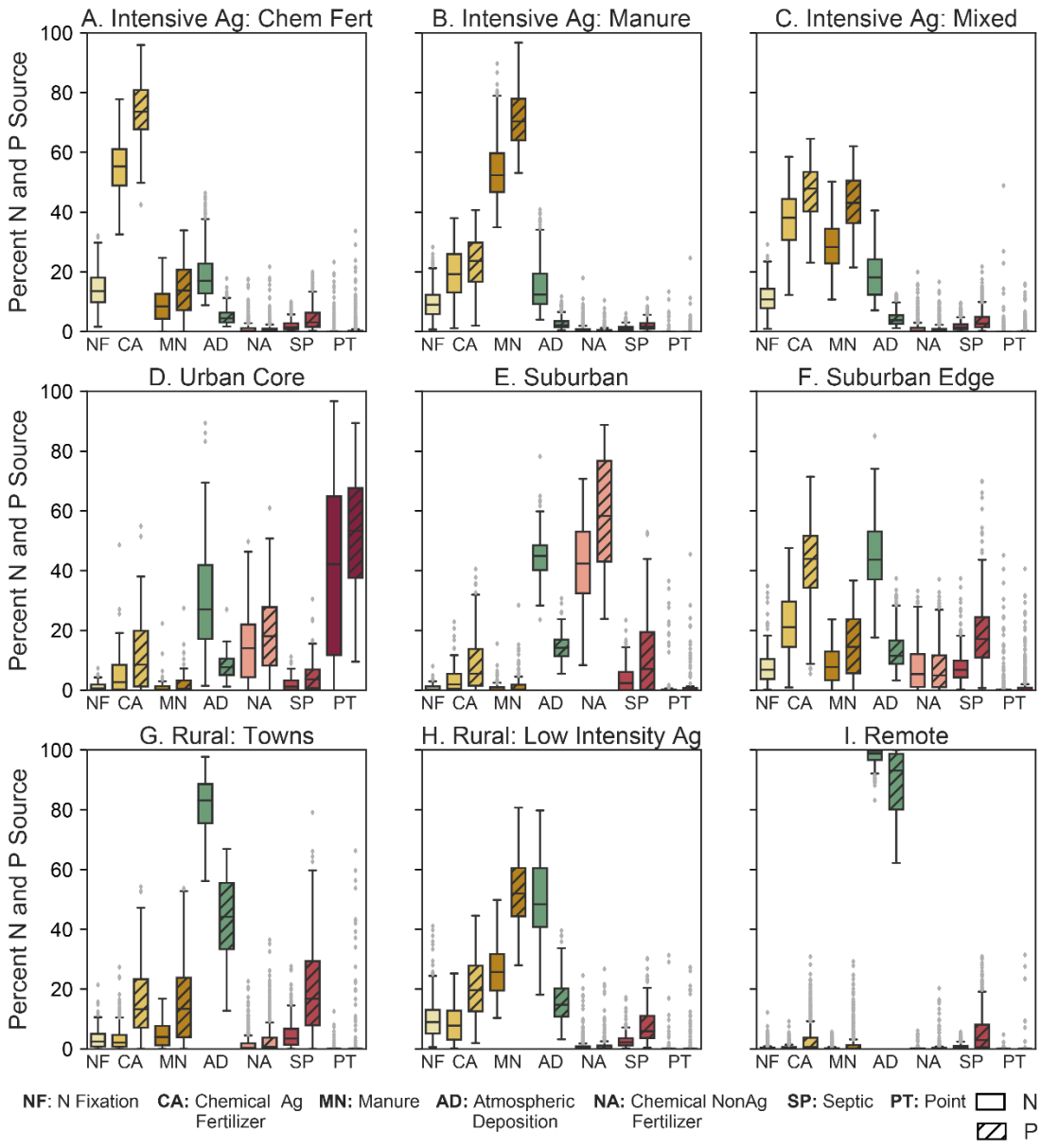

Figure 1.8 Distribution of N and P source percent within each Nutrient Input Landscape. (A) Intensive Ag: Chemical Fertilizer Dominated, (B) Intensive Ag: Manure Dominated, (C) Intensive Ag: Mixed, (D) Urban Core, (E) Suburban, (F) Suburban Edge, (G) Rural: Towns, (H) Rural: Low Intensity Agriculture, (I) Remote. Legend: NF: N fixation, CA: Chemical agricultural fertilizer, MN: Manure, AD: Atmospheric deposition, NA: Chemical nonagricultural fertilizer, SP: Septic, PT: Point. Hatched boxes: Phosphorus, unhatched boxes: Nitrogen.

More densely populated areas of cities and suburbs classify into Urban Core, Suburban, and Suburban Edge landscapes. Urban Core watersheds have large contributions from point sources (Figure 1.8D), meaning in general that they are located in urban watersheds with major WWTPs. Suburban watersheds are characterized by high chemical non-agricultural fertilizer due to lawns and golf courses (Figure 1.8E). Suburban edge watersheds have both rural and suburban populations, leading to a mixture of agricultural sources, chemical non-agricultural fertilizer, and septic tanks (Figure 1.8F). The radiating spatial pattern of an Urban Core downtown transitioning to Suburban, Suburban Edge and beyond can be seen in both Detroit in southeast Michigan and Cleveland in northeast Ohio, as well as around Buffalo in New York (municipalities labeled on Figure 1.7 caption).

Rural landscapes are most abundant in the northern tier of the US-GLB and in less populated portions of New York state (Figure 1.7). These landscapes are classed into: Towns, Low Intensity Agricultural, and Remote. Nutrient inputs in remote regions come almost entirely from atmospheric deposition, whereas rural towns and rural low intensity agriculture have appreciable fractions of anthropogenic sources including septic/non-ag fertilizer and chemical fertilizer/manure sources respectively (Figure 1.8G-I).

Nutrient inputs and loading are often linked to LULC because it is a proxy for management practices and population dynamics, however LULC is insufficient to characterize the source composition of nutrients. Since SENSEmap is source-specific, it allows a more detailed interpretation of land uses that affect nutrient sources than LULC alone. In Intensive Agricultural landscapes, distributions of nutrient source (Figure 1.8A-C) vary significantly by source despite similar LULC (Figure 1.9A-C) (See Figure A1.12 for LULC for all landscapes). Similar to the intensive ag landscapes, Urban Core and Suburban landscapes have similar LULC

signatures but receive nutrients from very different sources, point sources and chemical non-agricultural fertilizer, respectively (Figure 1.8D-F, Figure A1.12 D-F). Without source specificity, unique characteristics of nutrient source and composition cannot be determined. Our Nutrient Input Landscapes thus illuminate additional patterns in the landscape not visible with LULC alone.

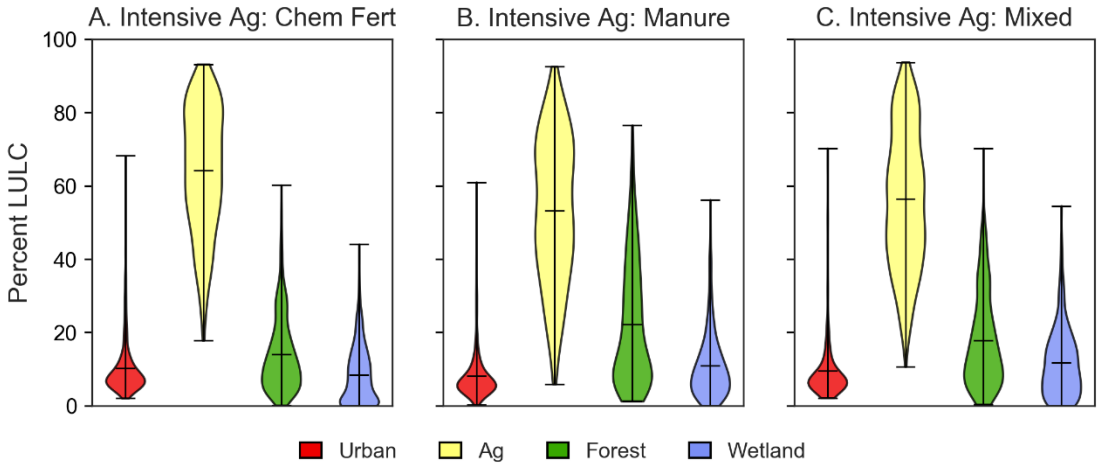

Figure 1.9 Distributions of LULC within three intensive agricultural Nutrient Input Landscapes. Distributions of LULC are shown as violin plots, where filled area represents probability density to provide additional information on the underlying distribution. Note that while source distributions for these three classes (Figures 1.8A – 1.8C) differed greatly, their LULC distributions are strikingly similar. (A) Intensive Agriculture: Chemical Fertilizer Dominated, (B) Intensive Agriculture: Manure Dominated, (C) Intensive Agriculture: Mixed.

The additional complexity of the landscape becomes clear when LULC maps are juxtaposed with different views offered by SENSEmap and Nutrient Input Landscapes (Figure 1.10). For instance, variations in total non-point P (Figure 1.9B) within land classified simply as agriculture have P inputs from 5 to 10 kg/ha/yr depending on location and crop practices. Beyond just a single nutrient map, N:P ratios at the HUC12 level (Figure 1.10C) help illustrate variable N:P input relationships over a small area. While similarly classified, agricultural watersheds with low ratios are likely hotspots for P loading to the western Lake Erie Basin, exacerbating algal bloom issues there. These watersheds can then become key focus areas for

management action to reduce P deliveries to the Lake. The four intensive ag watersheds in Figure 1.10D are all P-enriched relative to their largely chemically-fertilized neighbors. While some of this may be due to assumptions within SENSEmap (i.e. fields receive either manure or chemical fertilizer, not both), manure is highly P enriched due to significant losses in N between animal emissions and field inputs. Thus, these maps serve to highlight the benefits that may be achieved through policies demanding robust design and enforcement of nutrient management plans for livestock operations. Beyond simple input differences, the pathways and rates of nutrient transport and uptake differ across sources. By being space- and source- specific, SENSEmap provides the detailed view of the landscape needed by local planners to better manage their watersheds.

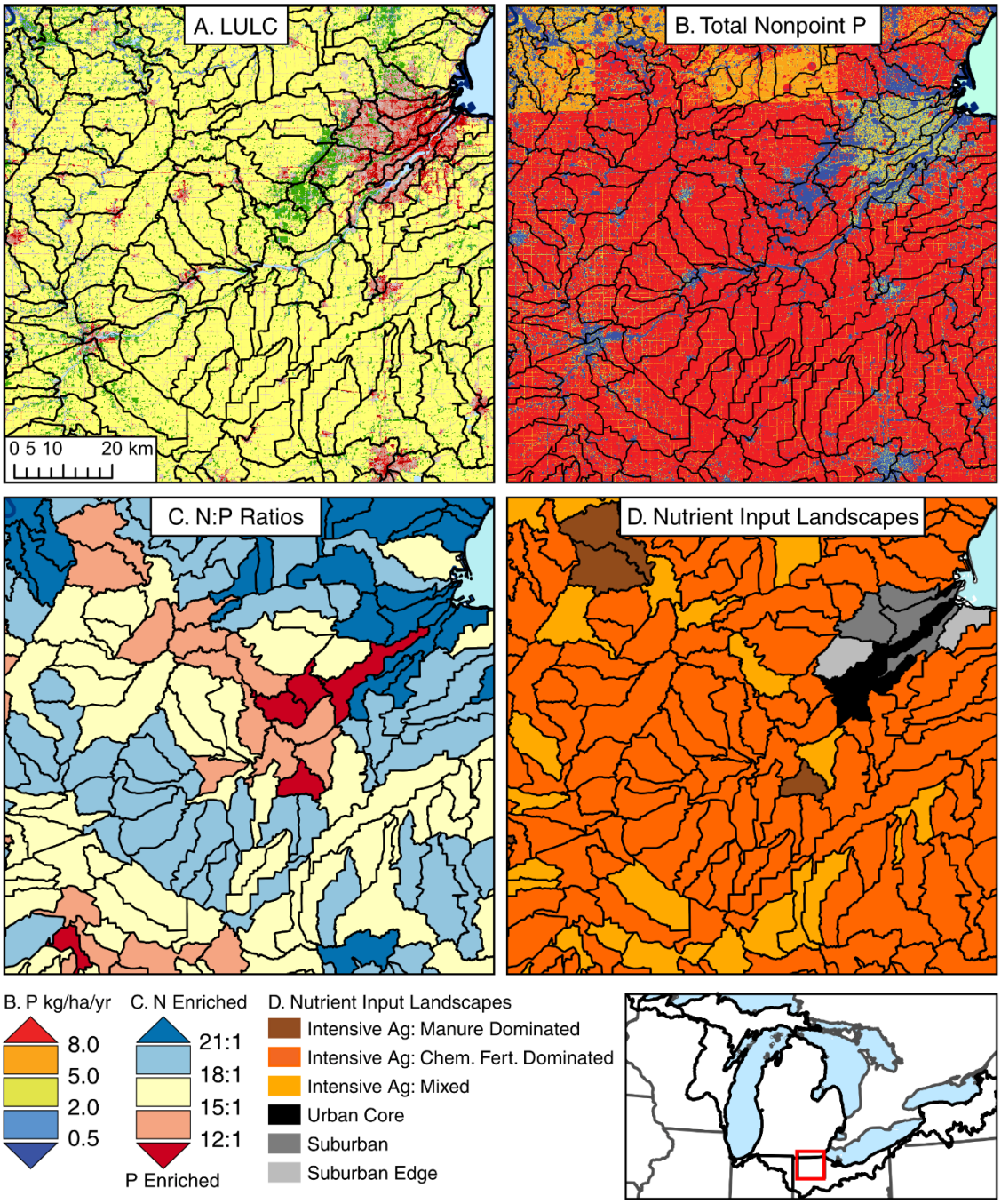

Figure 1.10 Comparison Panel of LULC, SENSEmap TP, N:P ratios, and NILs. Four views of a portion of the US-GLB surrounding Toledo, OH (red box on locator map). (A) LULC; see Figure 1.3 for caption and legend, (B) Total nonpoint P, (C) N:P ratio at HUC12 watershed level, and (D) Nutrient Input Landscapes.

3.4 Comparisons to Other Products

To better understand SENSEmap's performance, we compared SENSEmap to other existing nutrient products. No other products were found to compare at the same resolution and for all nutrient sources; however, we selected USGS SPARROW (Robertson & Saad, 2011), the Net Anthropogenic Nitrogen Index (NANI) (Swaney et al., 2018), and Cao et al's. (2018) N chemical agricultural fertilizer product. Here, we summarize the results of these comparisons. Additional information can be found in Text A1.4.1, A1.4.2, and A1.4.3.

We compared the landscape inputs within a subset of the SENSEmap sources to those from the USGS SPARROW model (Robertson & Saad, 2011; Wieczorek & Lamotte, 2011). Note that these products are not equivalent: SENSEmap calculates discrete nutrient inputs in kg/ha/yr, whereas USGS SPARROW uses both LULC area and nutrient masses to statistically compute source-specific fluxes. In particular, SPARROW designates urban inputs by LULC, rather than by urban sources (i.e., septic and non-agricultural chemical fertilizer). Differences between mean HUC8 N and P landscape input rates were small, with SENSEmap-US-GLB showing slightly higher rates for N and slightly lower for P (TN: 2.1 +/- 8.5 kg/ha/yr, TP: -0.1 +/- 1.3 kg/ha/yr; positive values indicate SENSEmap was higher). These differences could be partly due to different data sources and timeframes. For source specific comparisons and data differences, see Text A1.4.1.

The Net Anthropogenic Nitrogen Index focuses on mass fluxes of key nitrogen sources to explain nitrogen dynamics and in-stream concentrations. Swaney et al. (2018) calculated N fixation, total fertilizer, agricultural fertilizer, manure, and atmospheric deposition at the county level across the United States. Unlike SENSEmap, NANI includes nitrogen exports via food and feed, whereas SENSEmap describes only nutrient inputs. We compared 2012 N input values for

manure, chemical agricultural fertilizer, and N fixation at the county-level using rates per total county area. We found that N intensity of manure and chemical agricultural fertilizer were similar, with SENSEmap showing consistently higher manure and lower chemical agricultural fertilizer values, while N fixation had more variation (mean +/- standard deviation, positive values denote SENSEmap had higher values: Manure N: 3.8 +/- 6.2 kg/ha/yr, Chemical Agricultural Fertilizer N: -2.7 +/- 5.6 kg/ha/yr, N Fixation: 7.1 +/- 15.9 kg/ha/yr). Additional information is provided in Text A1.4.2.

Cao et al.'s (2018) recently developed 5 km resolution maps of nitrogen fertilizer inputs, which are directly comparable (though at a coarser resolution) to the chemical agricultural fertilizer N source layer from SENSEmap. Following aggregation of SENSEmap to 5 km resolution, we found that in general the fertilizer estimates had good agreement (1.5 kg/ha +/- 9 kg/ha). At low nutrient level (< 20 kg/ha), there was no net bias between the two products (Figure A1.17), while SENSEmap had slightly higher estimates in higher nutrient pixels. Differences were more pronounced in intensive agricultural areas; in high manure areas SENSEmap tended to have lower chemical agricultural fertilizer values, while the reverse was true in non-manured areas (Figure A.16). As a result of this source-specific bias between products, differences were not spatially uniform (Figure A1.16). For additional methodology and figures, see A1.4.3.

3.5 Sensitivity to Chemical Agricultural Fertilizer

SENSEmap is driven by external products to quantify chemical agricultural and chemical non-agricultural fertilizer. Brakebill & Gronberg (2017) have recently updated work by Ruddy et al. (2006) and Gronberg & Spahrg (2012) to model county level N and P fertilizer kg per year. However, the Brakebill & Gronberg (2017) product is modelled based on fertilizer expenditures

and there are few other fertilizer products to compare to. This reliance on a single product methodology introduces uncertainty into nutrient budgets based on these values, such as SENSEmap. Given the dominance of the chemical agricultural fertilizer source across much of the GLB, a sensitivity analysis of SENSEmap to uncertainty within the county-level Brakebill & Gronberg (2017) product is warranted.

TN and TP were largely unaffected at the HUC12 scale (coefficient of variation (CV, or standard deviation divided by mean): TN: 2.3%, TP: 5.1%), while as expected total chemical agricultural fertilizer (kg) was more affected (CV: chem. ag fertilizer kg N: 11.4%, P: 17.0%). Landscape level patterns, as described by Nutrient Input Landscapes, remained largely the same. 85% of HUC12 watersheds were always assigned the same landscape, regardless of change in chemical agricultural fertilizer. Among the 15% of watersheds with at least one simulation producing a different landscape, changes were primarily between Intensive Ag: Chemical Agricultural Fertilizer Dominated and Intensive Ag: Mixed. Suburban Edge, the least cohesive cluster, also saw significant numbers of changes. Overall, broad patterns in the landscape were not sensitive to specific values of total chemical agricultural fertilizer at the county level. Additional information can be found in Text A1.5.

4 Conclusions

SENSEmap provides a spatially- and source- explicit description of landscape nutrient inputs in a format that should be useful to both managers and scientists. By quantifying seven sources of nutrients, we found agricultural sources (Chemical Ag Fertilizer, Manure, N Fixation) dominate the US-GLB, accounting for 71% of N and 88% of P. We can use the insights from SENSEmap-US-GLB to provide managers with detailed estimates of nutrient inputs. As part of the Great Lakes Basin Tipping Point Planner, we are cultivating connections with watershed managers.

To summarize nutrient source budgets and find patterns in inputs, we used k-means clustering to describe nine Nutrient Input Landscapes. Notably, land use composition did not vary significantly between three intensive agricultural landscapes, suggesting that land use alone does not capture the heterogeneity and detail in nutrient sources. Individual nutrient sources (specifically manure and chemical agricultural fertilizer) behave differently in terms of speciation and transport, making it important to recognize that a total input from agriculture can have significantly differences based on nutrient source. The concept of our Nutrient Input Landscapes can serve a number of other purposes, such as analyzing different management strategies across similarly composed watersheds to better understand their transferability and developing metamodels for nutrient modeling.

Quantifying uncertainty in a product with no ground truthing is difficult, but uncertainty arises from errors in classification in remote sensing products (NLCD, CDL), pseudo-random placement (septic, non-CAFO manure farms), literature compilation of P deposition, and statistical models used to approximate spatial heterogeneity (chemical agricultural fertilizer, N fixation). Because of the assumptions and pseudo-random placement within many algorithms,

we do not claim that SENSEmap inputs are precise at the 30 m scale; rather, they are more accurate at aggregated watershed scales that are better representative of spatial uncertainties. We also recognize that this approach does not account for every nutrient source; future work will add other nutrient sources that are regionally-minor but locally-important, such as land applied septage, land-applied secondary treated wastewater, commercial septic systems, urban animal wastes, and aquaculture.

Opportunities for future work with the SENSEmap framework are broad. We plan to investigate different spatial extents and use both statistical and process-based modeling to explore surface and subsurface nutrient transport. We are currently applying the SENSEmap framework to the Canadian GLB, and the methods developed here can also be expanded to any part of the continental United States. Currently, SENSEmap-US-GLB is being used as a driving variable to determine causes of coastal wetland invasion within the region. Additionally, we are processing SENSEmap-US-GLB source inputs with a statistical and geospatial nutrient transport model, similar to the methods used in Luscz et al. (2017). The SENSEmap-US-GLB products can also be used as inputs to a range of crop, hydrology, and biogeochemical models.

Acknowledgments

This chapter was coauthored by Anthony Kendall, Sherry Martin, Henry Whitenack, Jacob Roush, Bailey Hannah, and David Hyndman. We thank Emily Luscz for developing the initial framework for the conceptual model and Jillian Deines for assistance processing remote sensing inputs. The authors have no competing interests to declare. SENSEmap-US-GLB is available for download on Hydroshare at 30 m, HUC12 and HUC 8 scale (https://doi.org/10.4211/hs.1a116e5460e24177999c7bd6f8292421). Funding for this work was provided by USDA-NIFA grant 2015-68007-2313, NOAA grant NA12OAR4320071, NASA

grants 80NSSC17K0262 and NNX11AC72G. Quercus Hamlin was partially supported by the NSF Graduate Research Fellowship Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of USDA NIFA, NASA, NSF, or NOAA.

APPENDIX

Text A 1.1 Source Methodology

The following sections contain additional methodology for nutrient source modeling processes.

Text A1.1.1 Atmospheric Deposition of Phosphorus

Compared to nitrogen, atmospheric phosphorus deposition is more difficult to measure and lacks the extent of monitoring networks that are available for nitrogen. To overcome this limitation, we conducted an extensive literature review to gather reporting data on annual atmospheric phosphorus deposition rates. We first conducted an exhaustive search, then limited the papers based on the following criteria: sites could be a maximum of 700 kilometers distance from the Great Lakes basin, a year or more of data must have been collected, and the study must have used acceptable collection methods, as discussed below. Data collection methods varied, and studies were excluded for reasons such as basing annual load estimates on 5-months of collection, collections based off of ships, and using volunteer collected samples.

Each site selected for our modeling had to be reviewed and selected individually, a process which often involved nutrient species and unit conversions, averaging of loading rates across multiple years and removal of outliers on a case by case basis. To standardize the model all units were converted to kg/ha/yr. Variations of units were common, including “thousandths of a ton/day/mi^2” and units of time, weight or area had to be carefully converted and double checked. A number of papers also reported values in phosphate (PO4) and were converted to phosphorus (P) for our model. Site locations often lacked specific latitude and longitude coordinates. In these cases, site locations were estimated based on georeferencing of map images from the papers. The map scale at which images were georeferenced varied.

A number of papers had multiple collection sites but did not report individual collector deposition rates, instead opting to average site deposition values across study. These sites would have the same values as other sites from the same paper. With individual sites rates unavailable,

“duplicate” sites were denoted, unaltered and interpolated as is. Clustered sites were considered to be points 5 miles away or less from each other and may have been from the same or different studies. In total there were 7 clusters that were each were examined very carefully; means were established for each cluster, and individual sites within the clusters were closely examined to see if their values coincided with nearby sites in the same cluster. Consideration was made whether to average nearby sites together to achieve "area averages", however this was not done.

In total, 98 sites were input to the ArcGIS kriging toolset to develop spatially distributed loading estimates for both atmospheric nitrogen and phosphorus. Other modeling techniques (IDW) were also tested. In ArcGIS Desktop, ESRI’s ‘Compare’ function under the Geostatistical Analyst was used to compare and adjust models and variables until we were confident in the accuracy of our model. Model optimization was based on minimizing the root mean square error. Wet and dry nitrogen were modeled separately using ordinary kriging, with no transform, no trend removal, and parameters non-optimized settings. Their outputs were summed together to develop a total nitrogen load estimate. One model for total phosphorus (both dry and wet deposition) was generated using ordinary kriging; log-transform; no trend removal; hole-effect model; parameters optimized settings. Optimized versus non-optimized in this context refers to whether the Optimize button in the Geostatistical Analyst was used. We found that we were able to best minimize the root mean error using non-optimized settings for nitrogen and optimized settings for phosphorus.

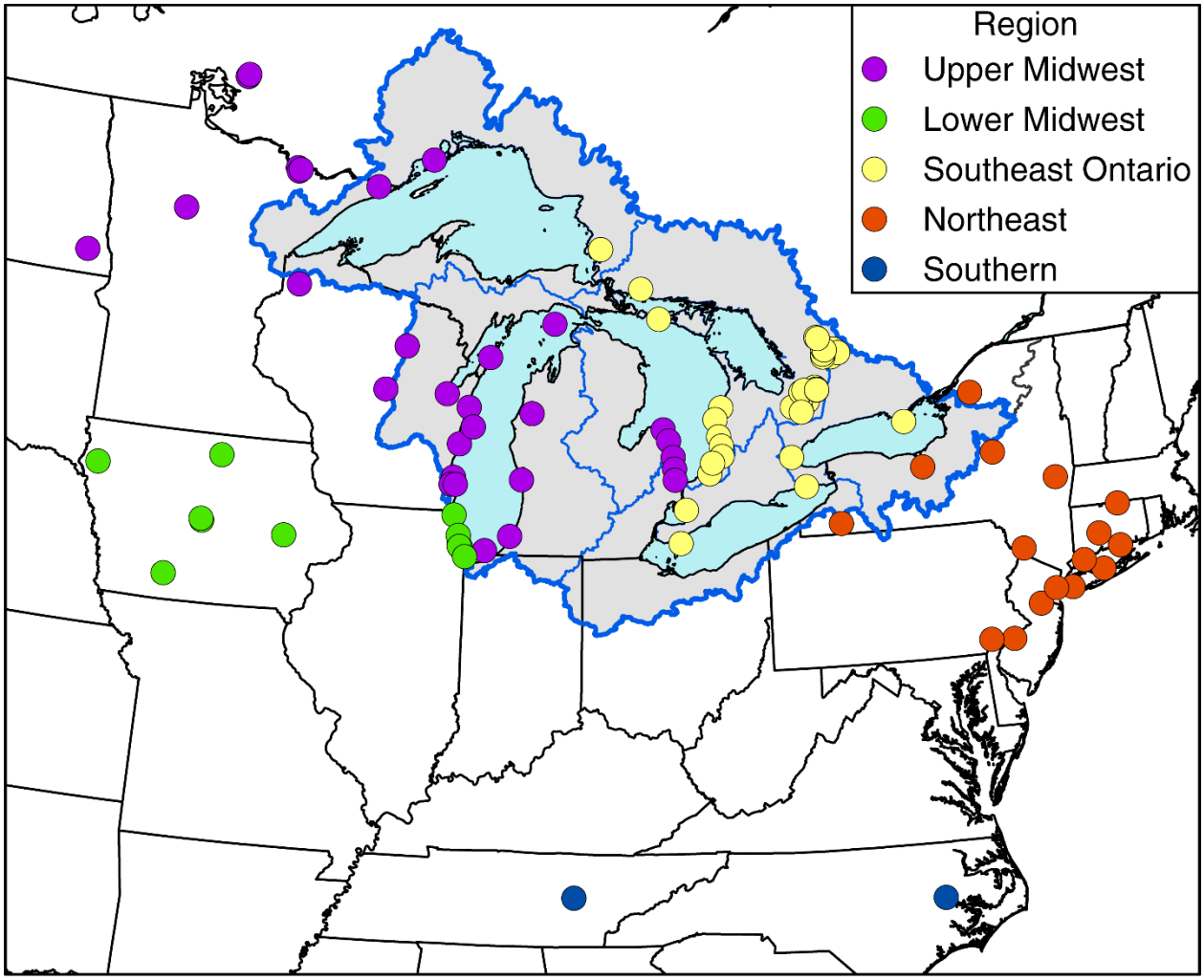

Figure A1.1 Atmospheric deposition P site map by region.

Table A1.1 Summary of phosphorus atmospheric deposition sites.

Regions	Range	Mean	Median	Number of Sites	Number of Papers	States Included
Upper Midwest	0.07 - 0.51	0.24	0.23	32	9[a]	MI, MN, ND, ON, WI
Lower Midwest	0.25 - 0.56	0.43	0.41	12	3[b]	IL, IA
Southeast Ontario	0.09 - 0.56	0.33	0.35	36	8[c]	ON
Northeast	0.04 - 0.29	0.11	0.09	16	3[d]	CT, NJ, NY, PA, RI
Southern	0.18 - 0.49	0.33	0.33	2	2[e]	NC, TN

[a]*(Bradley, 2011), (Delumyea and Petel, 1978), (Eisenriech et al., 1977), (Linsey et al., 1987), (Munger, 1982), (Robertson et al., 2009), (Schindler et al., 1971), (Schindler et al., 1973), (Schindler et al., 1976), (Wright, 1976)*

[b]*(Anderson and Downing, 2006), (Eisenriech et al., 1977), (Murphy, 1974)*

[c]*(Bradley, 2011), (Brown et al., 2011), (Foster, 1974), (Linsey et al., 1987), (Schindler and Nighswander, 1970), (Winter et al., 2002), (Winter et al., 2007)*

[d]*(Koelliker et al., 2003), (Pearson and Fisher, 1983), (Yang et al., 1996)*

[e]*(Brinson et al., 1980), (Swank and Henderson, 1976)*

Text A1.1.2 Septic - Exclusion and Inclusion Masks

Our exclusion mask, given by the service areas of WWTPs throughout the study area, was determined based on drinking water well data, US Census data, and WWTP locations as reported in the Clean Watershed Needs Act database (US Census, 2010; CWNS, 2012). Because no comprehensive statewide data were available for WWTP service areas, we developed an algorithm to infer their locations from these three datasets. The automated wastewater treatment plant (WWTP) service area delineation model was developed as an improvement to the Luscz et al. (2015) septic placement model, which required hand delineation of WWTP service boundaries. The larger spatial extent necessitated a more efficient, systematic, and reproducible method to estimate boundaries.

To facilitate WWTP service area classification, we compiled a point dataset of drinking water wells in all eight Great Lakes states. The algorithm uses these along with Census data to compute two quantities at the Census block level: the ratio of corrected drinking water well counts to household unit and population density. Individual states collect and maintain well records; however, the quality and completeness of digitized records varies. After compiling and removing non-domestic classified wells, we compared ratios of wells to household units in blocks unlikely to be in service areas (not in a Census designated place, greater than 5 household units, greater than 1 well) to come up with an approximated "undercount" of wells. Our assumption was that a fixed proportion of wells would be missed in each county because of incomplete digitization efforts at the county level, and by using only blocks highly likely to be outside of WWTP service areas, we could find this proportion of missing wells. For example, if a rural Census block had 10 household units and 8 wells, we could assume an undercount of 2, or 20%. We would then divide the 8 wells by 80% to add 2 wells and raise the well number to 10, matching household units in a block highly likely to be outside a WWTP service area. We

applied a fixed undercount proportion at each county. The county undercount value was chosen by taking a state-calibrated percentile of all block undercounts within a county. This state threshold percentile (85-95%) was set to the 90th percentile where possible; however, Indiana had to be lowered to the 85th because of excess misclassified agricultural wells that led to 90th percentile values over 1. We chose a high percentile to select a ratio that was reflective of areas most likely to be unserved, thus adding a conservative amount of wells. At the county level, all blocks' well counts were divided by the undercount well:household ratio to add the corrected number of wells.

Block classification was based on corrected well to household unit ratios, population density (block group level), and block membership to a Census designated place. Census designated places are named loci of population but may or may not be incorporated (US Census, 2010). Variables were calibrated based on multiple runs, qualitative assessment of spatial distributions, comparison to existing service area maps, and confirmation of existing wastewater services for small, low density areas. Calibrations were performed at the state level to account for variation in drinking water well data collection (Table A1.1). The conditions (Figure A1.2) followed the assumption that a block with higher density, low well to household unit ratio (given by Table A1.3), and inside a CDP would be within a service area. Additional conditions to account for small, low density areas and densely populated areas with an inflated ratio due to low household unit number were also included. After initial classification, the 97th percentile of density in serviced blocks was calculated and used to reclassify blocks that had higher ratios than the standard ratio threshold.

After initial classification, blocks were dissolved. Dissolved, in the GIS sense, refers to removing boundaries of subregions (Census blocks) with common characteristics (WWTP

service area classification) to create larger, contiguous regions. This created a layer of service areas, non-service areas, and unclassified blocks. To correct for small polygons likely misclassified (identifiable by being surrounded by a different class), we used buffered WWTP points from Clean Watershed Needs Survey data (CWNS 2012). Buffer size was variable based on the population receiving treatment and determined by qualitative assessment of distance from WWTP to plausible boundaries in each state. Because WWTPs are not always centrally located, buffer sizes erred on the large side. Buffer population ranges and corresponding size were determined for each state. To improve the service areas, polygons under 2 km^2 were "flipped" from service area to unserviced based on if they fell within a WWTP buffer. For example, if a 1 km^2 polygon was originally classified as unserviced but was within a WWTP buffer, it would be changed to serviced. The flipped polygons were dissolved again to create final service areas.

After excluding WWTP service areas, primary and secondary inclusion masks for septic placement were defined. The primary inclusion mask represents urban and suburban areas, whereas the secondary inclusion mask proved necessary to account for septic systems in rural settings. The primary inclusion mask includes all areas with NLCD cells classified as urban, over 15 meters away from water features, and between 10 and 60 meters from roads. These setbacks were selected to represent a reasonable set of requirements for riparian sanitary codes, distances to public infrastructure, and under the assumption that in most cases, residences are built relatively close to roads. The secondary inclusion mask included all non-urban cells that met the riparian and road set back requirements.

Table A1.2 Calibrated values for WWTP and Septic conditions. See Figure A1.2 for conditions.

State	Well : Household Ratio			Population Density (Pop/km^2)		
	Low	Standard	High	Low	Standard	High
IN	0.08	0.12	0.5	20	220	500
IL	0.1	0.1	0.5	20	300	400
MI	0.05	0.1	0.5	18	135	400
MN	0.1	0.1	0.5	20	250	400
NY	0.1	0.1	0.5	20	220	400
OH	0.1	0.12	0.5	20	230	400
PA	0.1	0.1	0.5	12	300	500
WI	0.08	0.12	0.5	30	300	400

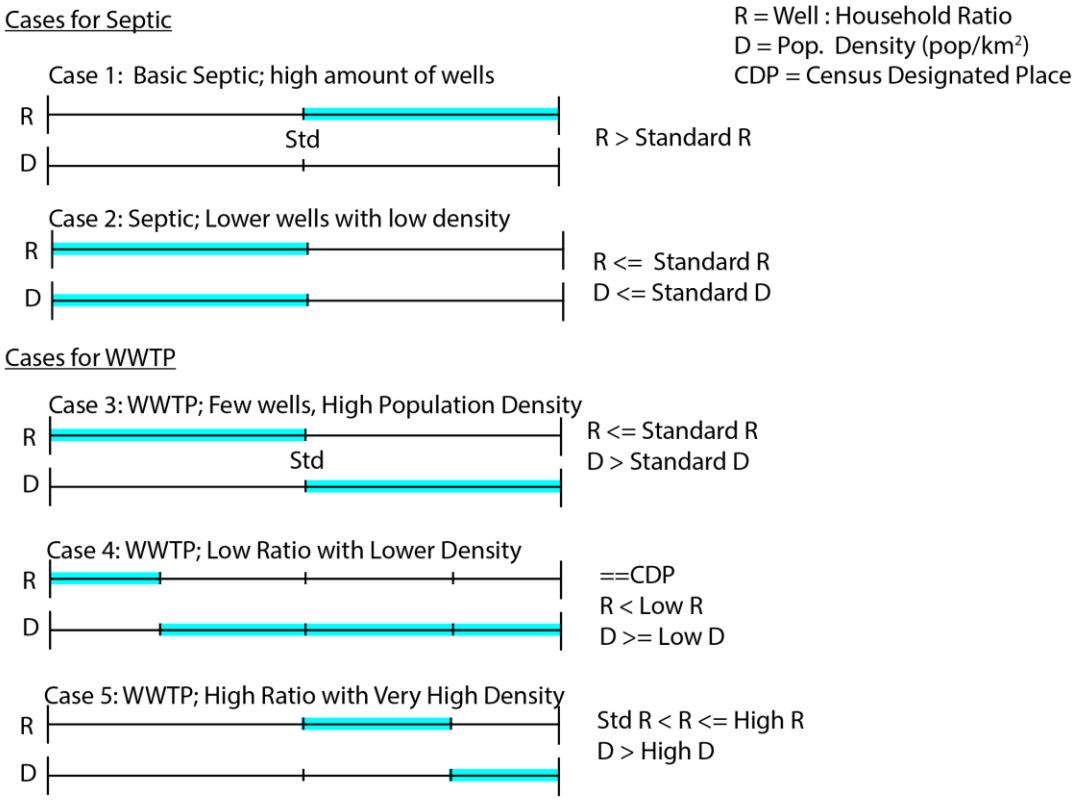

Figure A1.2 Logical statements used to classify blocks as WWTP service area.

Text A1.1.3 Fertilizer Demand Model

The basis for the spatial arrangement of both manure and chemical agricultural fertilizer models is crop- and location-specific fertilizer demand. In reality, this is an incredibly complex function of environment, management, policy, economic, and individual considerations—the modeling of which is beyond the scope of this effort. Instead, we chose a more straightforward approach of fitting a machine learning model of fertilizer nutrients as determined from fertilizer sales data at the county-level (Brakebill & Gronberg, 2017) created using Boosted Regression Trees (BRT) to county-level crop, soil, and geographic variables. We used the guiding principles that fertilizer is applied based on type of crop, as well as soil texture, regional agricultural intensity, and climate. To vary fertilizer application within a county, we fit a boosted regression tree model that used crop type, soil texture, total cropland, and latitude to the average rate of chemical agricultural fertilizer spread over county cropland (kg/ha N and P). This model produces a relative demand of fertilizer across crop types, specific to each location across the Great Lakes Basin. This demand model, applied at the pixel-level, is then used as a means to spatially distribute crop nutrient applications within manured areas and across chemically fertilized areas of each county.

Boosted regression trees (BRT) is a method that combines classification trees and boosting. Classification trees take any number of variables and breaks data into mutually exclusive groups that explain the most variance between groups (Brieman et al., 1984). The "tree" is built by continuous binary splitting until the most variance can be explained in the datasets by these logical rules. The boosting algorithm combines many individual models to increase the analysis's predictive power (Quinlan, 1996; Friedman et al., 2000). The final model used by BRT utilizes a cumulative regression that forms through recursive trees and calculated

residuals (Elith et al., 2008). Independent variables are ranked on their relative influence, allowing for a proportional measure of variable importance.

We used BRT to quantify crop-specific nutrient demand at the county level in response to geography, soil, and crop types, with input variables described in section A1.1.3.1 below. BRT has a number of input parameter that control aspects of the algorithm. Our implementation, conducted using Python's *scikit-learn* version 0.19.1, used largely default parameters. We did, however specify 100 boosting stages and required a minimum of 5 samples per leaf (each split produces two leaves) within each classification tree.

The county-model was fit with variables that are also observed at the pixel level. Ideally, we could fit a pixel-level model to observed fertilizer rates in fields; however, this was not possible. Thus, the model is fit to observed county-level variables, including Brakebill and Gronberg (2017)'s county fertilizer rates. We then applied the county-level BRT to predict fertilizer application rate in every cropland pixel in the GLB. At the pixel level, crop demand was calculated based on the average demand hectares (described below) over an 8-year crop rotation, described by CDL data from 2008-2015. The total hectares of cropland within a county was applied to all pixels within the county as an additional measure of regional agricultural intensity. Latitude and soil texture were native to 30 m resolution. Applying this model created variation within applications within a county. Finally, we adjusted all pixel rates within each county by a debiasing factor to match the observed total fertilizer at the county level.

In the future, we hope to improve this method with point observations for validation, regionally-specific crop recommended rates, and rules for smoothing rates within fields.

Text A1.1.3.1 Fertilizer Demand Model Variables

1. *Crop-specific demand*: Crop types were categorized into four levels of demand, based on Michigan recommended rates given in Warncke (2004). Table A1.3 provides the input

fertilizer rates and their demand classes. The median fertilizer rate specified by Warncke (2004) for each demand group was used to create a scaled coefficient applied as a "weight" to area. "Fertilizer demand hectares" were calculated by multiplying crop areas by their demand class weight. Thus, 100 hectares of corn would be weighted to more fertilizer demand hectares than 100 hectares of a lesser fertilized crop like wheat. Ag Census 2012 Crop harvested area for all crops with greater than 1000 ha in the US-GLB was converted to demand hectares and area normalized (USDA, 2012).

2. *Latitude*: We chose latitude as a variable that describes both regional practice and climate, as in the US-GLB, there is a north-south gradient in both climate and agricultural intensity. Southern parts of the US-GLB like Ohio and Indiana are intensely farmed and more suitable for high yields, thus resulting in higher fertilizer application rates, whereas central and northern counties produce lower yields and apply fertilizer at lower rates.

3. *Soil Texture*: Soil texture was represented via mapunit-, horizon- and component-weighted average percent sand and percent clay from SSURGO for depths 0-50 cm. These were processed from the gSSURGO dataset (gSSURGO 2019).

4. *Total Cropland*: Total area of cropland was used as a variable to describe agriculture intensity, following the idea of agriculturally intense counties applying fertilizer at higher rates. We used the 2012 Ag Census total cropland area at the county level.

Table A1.3 Nitrogen fertilizer rates. Rates derived from Warncke et al. (2004) and Warncke & Dahl (2003). Raw rates were not applied in SENSEmap.

Fertilizer Demand Group	Crop	N Fertilizer kg/ha/yr
High	Potatoes	203
	Onions	180
	Sod/Grass Seed	162
	Cabbage	158
	Corn	143
	Sweet Corn, Popcorn	133
	Carrots	124
Medium High	Tomatoes	111
	Peppers	111
	Greens	111
	Watermelons	111
	Sugarbeets	106
	Sunflower	89
	Herbs	89
	Squash	89
	Pumpkins	89
	Mint	89
	Cherries	84
	Peaches	84
	Asparagus	79
	Christmas Trees	72
Medium Low	Sorghum	67
	Spring and Winter Wheat	67
	Millet	67
	Speltz	67
	Grapes	67
	Cranberries	67
	Triticale	62
	Canola	57
	Apples	57
	Radishes	57
Low	Blueberries	49
	Dry Beans	44
	Peas	44
	Barley	27
	Oats	27
	Rye	22
No Fertilizer	Switchgrass, Soybeans, Alfalfa, Other Hay, Clover,	0

Text A1.1.4 Manure

The animal inventory was built from two sources: 1) the 2012 USDA Agricultural Census that reports county level information for all states on the number and type of farms, and 2) separate state-level records of confined animal feeding operations (CAFOs), which only include addresses and inventories for farms with animals counts above a certain size (CAFO designations vary across both animals and states). The USDA Agricultural Census includes all farms, including those regulated as CAFOs. However, the Census does not disclose animal counts for all farms, especially small operations, out of privacy concerns. To estimate animal counts for undisclosed farms, the number of animals in CAFOs were subtracted from the USDA Census population to give an average herd size of non-CAFO farms, by animal type and county. Herd sizes were used to label farms as pasture operations or confined operations based on a threshold confinement level given by Kellogg et al (2000). The confinement threshold limit differs by animal; for example, dairy cows are always confined, while hogs are only confined if the head count is greater than 450 (Kellogg et al., 2000).

For each farm in the animal inventory, a nutrient excretion rate and average county-level application rate was then estimated. Nutrient loads for each animal type in each county were calculated based on estimated nutrient excretion rates per animal (NRCS, 2008). Not all animal types in the USDA Ag Census have an excretion rate described by the NRCS. For animals with no nutrient excretion information, data on excretion rates was used to create a linear model to estimate excretion rate as a function of animal weight, following the method used by Luscz et al (2015). Manure is typically stored prior to application, and N is volatilized and lost to the atmosphere via denitrification during this time. We used per-animal empirical ratios of N to P for applied manure (Table 6 and 7, from Lorimor et al., 2008) to quantify denitrification for each farm, based on the original N:P ratio. County-average manure application rates were then

calculated by dividing the spreadable manure acreage from the Ag Census by total county animal excretion in terms of kg P.

We then randomly placed individual non-CAFO farms reported in the USDA Ag Census into agricultural areas, defined using a combination of the 2011 NLCD and CDL. We assumed that manure fertilizer would be used in the area surrounding farms, rather than transported to distant fields, as farmers would minimize transportation expenses. Farms were only placed in cells where the surrounding area could support the required area to spread manure, calculated using the proportion of that farm's load to the total county load with the reported spreadable manure acres.

To prevent spreading manure from confined farms on pasture areas for unconfined animals, we created exclusion buffers for pasture animal farms. These were created according to the farm's required spreadable area. For confined farms, including CAFOs, manure spreading (inclusion) buffers were created according to estimates of the required spreading area. To account for overlapping initial buffers and areas where the buffers intersect roads and urban areas, buffer sizes increased iteratively until 96% of farms met their required area for manure spreading. Overlapping buffers were merged and dissolved after each iteration of the buffer size, and re-clipped to include only agricultural land.

Each confined farm is assumed to lose a portion of manure during transport and storage. This portion, given by Kellogg (2000), varies by animal species and is applied as a single point at the location of each confined farm. Manure loads were applied within buffers at the rate of P fertilizer demand given by the fertilizer demand model. Demand was summed and adjusted by the factor to match observed farm P loads. N applications were determined by the N:P ratio after N loss for each farm. Due to manure being significantly P-enriched after N loss, we applied at

the crop's desired P rate in order to not over-fertilize with N. If N need was not met, additional N chemical agricultural fertilizer was applied.

Text A1.1.5 Chemical Agricultural Fertilizer

Chemical agricultural fertilizers are applied following manure separately for two cases: 1) areas within a manure buffer that did not receive adequate N fertilizer from manure alone, and 2) cropland areas outside of a manure buffer. For the first case, applied manure was subtracted from N demand from the county-adjusted BRT prediction at each pixel and any remaining demand was satisfied using chemical agricultural fertilizer. The remaining cropland cells received N and P fertilizer at the rate predicted by the relative demand model, applied based on the 8-year rotation for each pixel. We then adjusted all pixels within a county by the factor needed to match the observed county fertilizer from Brakebill & Gronberg (2017). Since the county level fertilizer is derived from purchasing data, intercounty transport (i.e. a farmer purchases fertilizer in one county and applies it in a neighboring county) can create unrealistic TN and TP values at the county level. To assure realistic values, we put upper and lower bounds on the adjustment factors used to assure aggregate inputs matched the county-level data in order to keep pixels within reasonable fertilizer rates. As a result, occasionally inputs of fertilizers did not sum to the observed county value. In particular, we specified that pixel values for N could not be adjusted by a multiplicative factor less than 0.2 or greater than 5, which represent the 3rd and 95th percentile of county level adjustment factors. P had a minimum factor of 0.1 (5th percentile) and no maximum factor because the crop specific demand model for P tended to overestimate actual fertilizer applications.

Text A1.1.6 Nitrogen Fixation

To calculate nitrogen fixation from legumes, we used the Cropland Data Layer (CDL, USDA NASS, 2008-2015) to identify cells containing leguminous crops and grasses (here soy,

dry beans, alfalfa, and other hay). We then applied a fixation rate to all legume and grass cells based on soil inorganic N, applied manure fertilizer, and spatially disaggregated yield.

Measurements of yield are available at the county level in the in the USDA Agricultural Census and Survey. To calculate yield at the 30 m cell resolution, we analyzed the relationship between county yield from the 2012 Ag Censuses and county mean vegetation indices, plant available water, and latitude at the included cells. Greener fields and higher plant available water are associated with higher yields and are available at high spatial resolutions (Courault et al., 2016). Latitude was included due to a north-south gradient in climatic change and suitability of fields for soybeans and hay. The southern parts of the basin in Ohio and southern Michigan are more heavily planted with soybeans, while alfalfa and other hay are more associated with northern latitudes in northern Michigan, Wisconsin, Minnesota, and New York (CDL). Using these relationships allowed for spatially disaggregating coarser county yield data to individual cells within a county.

We tested three vegetation indices: greenness index (GI), normalized difference vegetation index (NDVI), and enhanced vegetation index (EVI). Landsat 5, 7, and 8 images (native 30 m resolution) were compiled and cleaned in Google Earth Engine over the 2008-2015 time series used for placement. Cells were only included if the CDL crop type matched the identified crop type in each year. We ran versions of each model with all CDL cells over the 8 year period and with only cells with the correct crop more frequent than one year to adjust for random error in the CDL. Median and mean of annual maximum values were tested in the regression. Plant available water (PAW) was estimated using SSURGO textural classes mapped to hydrologic properties using the ROSETTA database (Schaap et al. 2001), and calculated as

field capacity minus wilting point for the depths 0-50 cm. Cell row in our model grid was used as a proxy for latitude.

We built models using both multiple linear regression and CART (Classification and Regression Trees). CART using NDVI median of annual maximum, latitude, and PAW produced the best model for soybeans ($R^2 = 0.75$, $p < 0.001$). Multiple linear regression using EVI median of annual maximum and latitude was the most successful model for dry beans ($R^2 = 0.53$, $p < 0.001$). Unsurprisingly, there was little variation in R^2 values (< .05 difference) across models due to similarity between vegetation indices. Once yield was calculated at all cells, rates were debiased by multiplying the observed Census and Survey average yield divided by average county calculated yields. This assured calculated mean county yield was equal to observed county yield. Alfalfa and non-alfalfa hay did not produce satisfactory relationships: the models could not predict high yields and coefficients of determination were between 0.1 and 0.4. We chose to assign county yields to all cells within a county and not spatially disaggregate the values.

To calculate N fixation, we first determined available soil inorganic N. We followed the method described in Goolsby et al. (1999) and used in Han and Allan (2008). Using SSURGO data converted to raster and resampled to our model grid as previously described, we calculated the mass of organic matter in the upper 50 cm of soil (to capture the majority of root zone activity) with the product of bulk density, volume, and percent organic matter. We followed literature (Han & Allan 2008; Goolsby et al. 1999) and used soil N content of 3% of organic matter mineralization at a rate of 2% per year in cultivated soil (Gentry et al., 1998; Stevenson, 1994). Using mineralized N, applied manure, and applied chemical fertilizer; we calculated percent of plant N fixed using Meisinger and Randall's (1991) formula (Eq 1).

$$N\ fixation\ (\frac{kg}{ha}) = Yield\ (\frac{kg}{ha}) * \%\ Plant\ N\ Content * \%\ N\ from\ Fixation \quad (1)$$

Meisinger and Randall (1991) provide a table of common values of nitrogen content based on crop and moisture rate. To pick a rate for non-alfalfa hay, we determined the types of commonly grown hay in the GLB through contact with the Michigan and Minnesota agricultural extensions. We averaged rates for the major types of hay (clover, birdsfoot trefoil, orchard grass and bluegrass) from Meisinger and Randall (1991) and used a rate of 0.02% N from fixation (Warncke et al., 2004; Kaiser et al., 2011). Rates for soy, dry beans, and alfalfa were 0.055%, 0.036%, and 0.0235% respectively (Meisinger & Randall, 1991). We calculated N fixation (kg/ha/yr) in each cell as the product of yield (kg/ha/yr), N content, and percent N from fixation (Eq 1).

Since our model is not a specific year, but rather a "circa" 2010 timeframe, we included all cells in the Cropland Data Layer (CDL) that had been any N fixing crop between 2008 and 2015. By using multiple years of data, we allowed for the effect of a corn-soy rotation. Although we use all cells, we corrected calculated fixed nitrogen to take into account frequency and yearly area. To do this, we adjusted each cell to a proportion of total yearly fixation based on the frequency of being identified as the assigned crop (e.g., 5-year soybean cell received the predicted fixation multiplied by ⅝). After correcting for frequency, we corrected the calculated fixation based on area using the USDA Agricultural Census 2012 values, as suggested as a best practice for CDL use in Lark et al. (2017). By adjusting by observed area divided by CDL area, we assured that we did not over-account for the usual planted area of each crop. The final cell values are abstract, in that they do not represent an actual yearly load at the point, but when aggregated to scales of analysis they provide representative total loads.

Text A1.1.7 Source Methodology Flow Charts

The following figures contain flow charts for each SENSEmap source methodology. Additional description of terminology and variables for each chart can be found in the main document Methods Section 1.2.3, and in the Text S1.1 above.

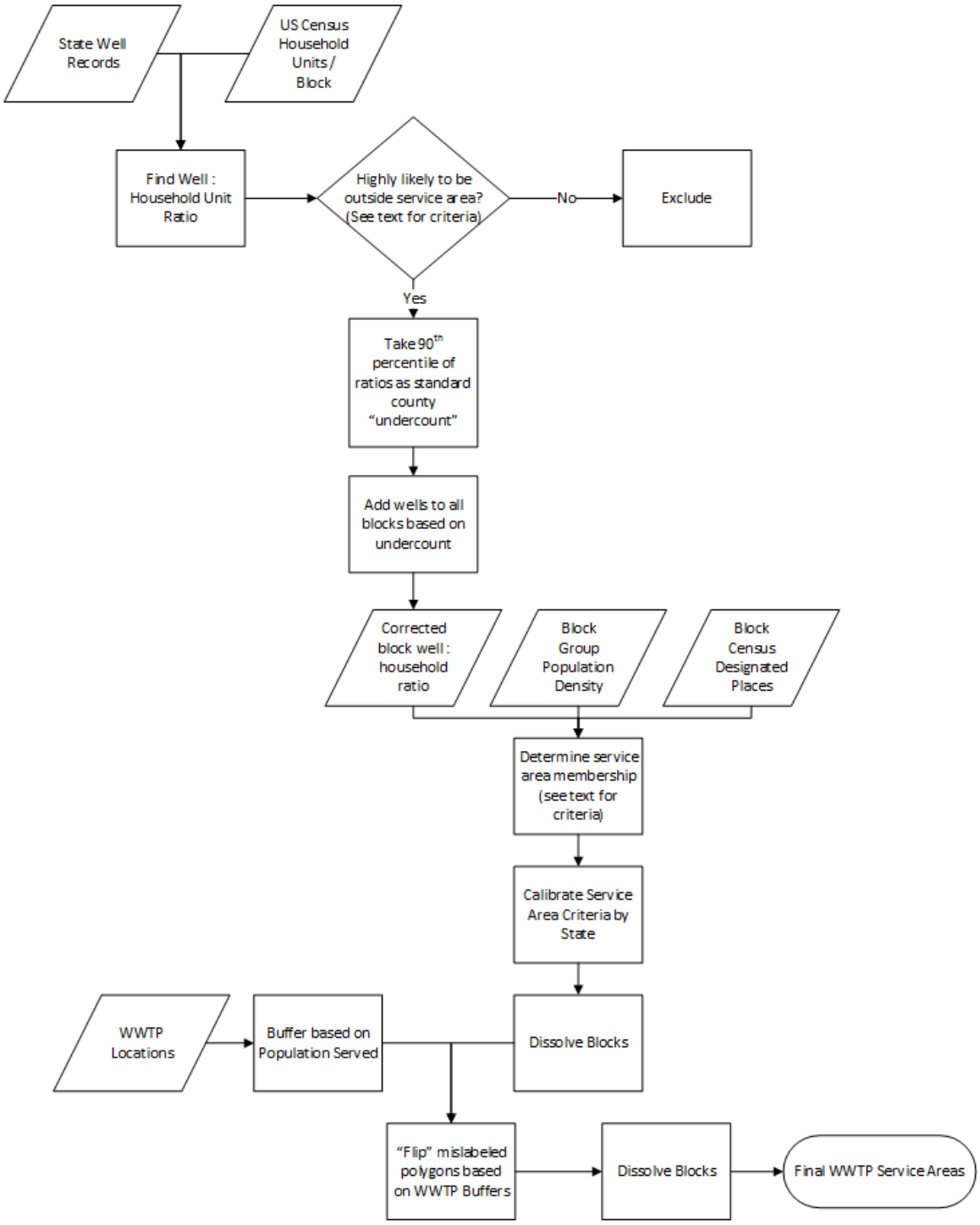

Figure A1.3 Septic WWTP Delineation flowchart.

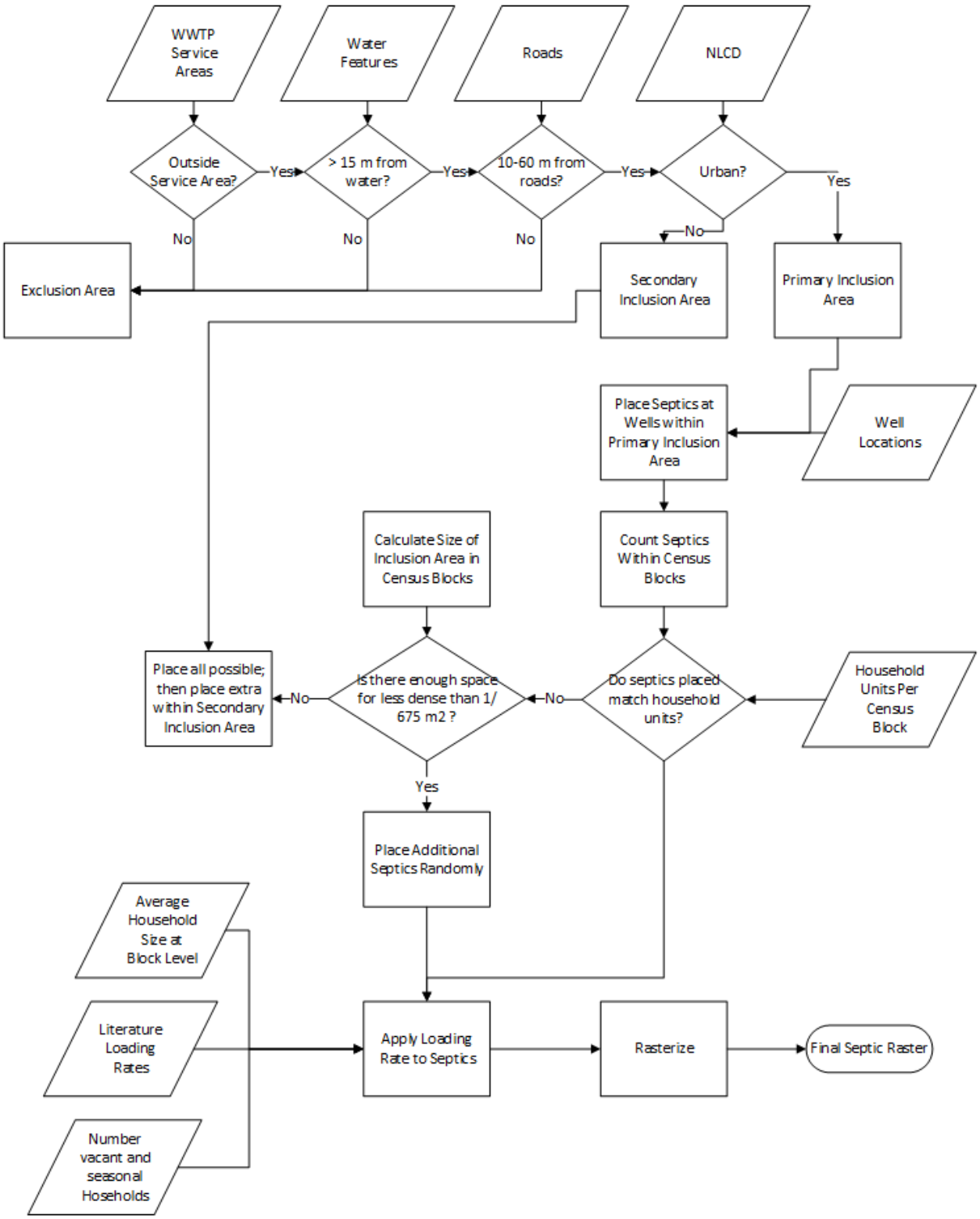

Figure A1.4 Septic Placement flowchart

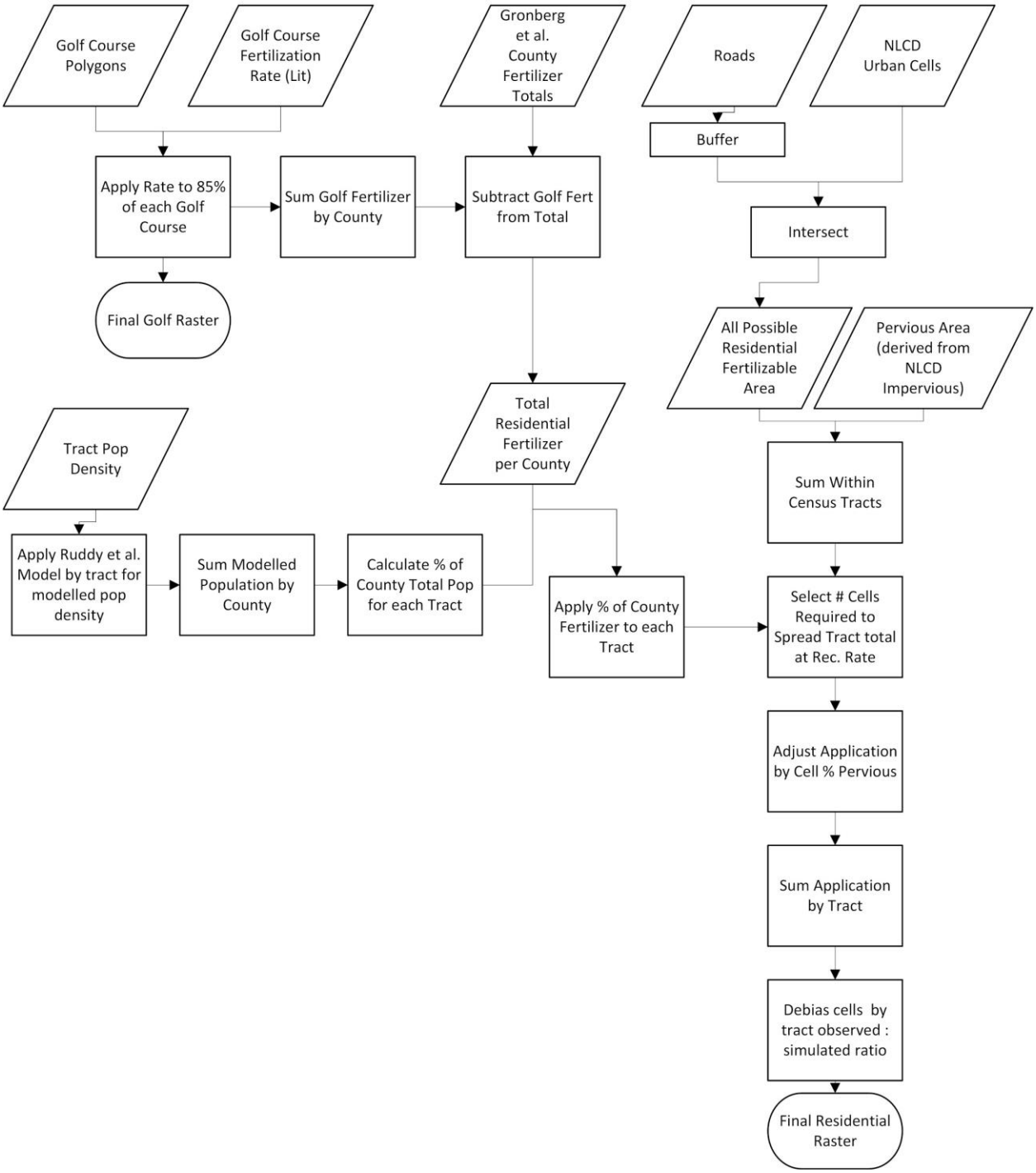

Figure A1.5 Chemical Non-Agricultural Fertilizer flowchart.

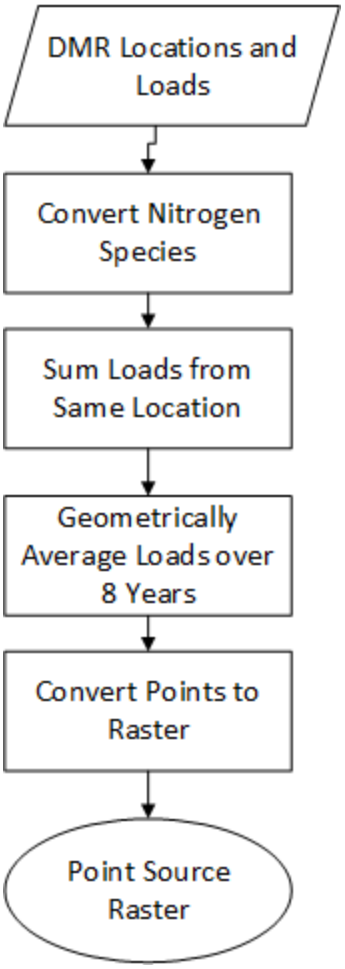

Figure A1.6 Point Sources flowchart.

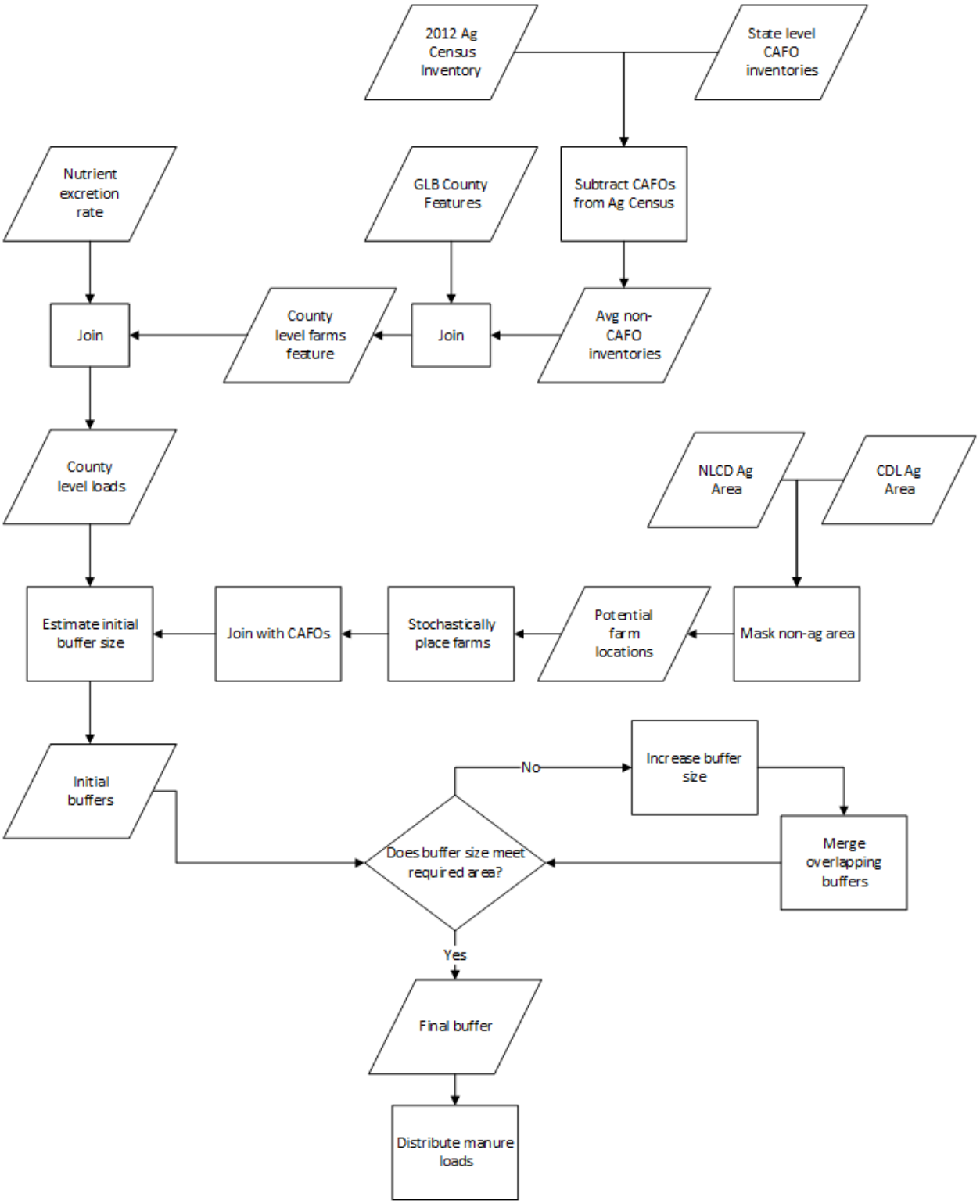

Figure A1.7 Manure flowchart.

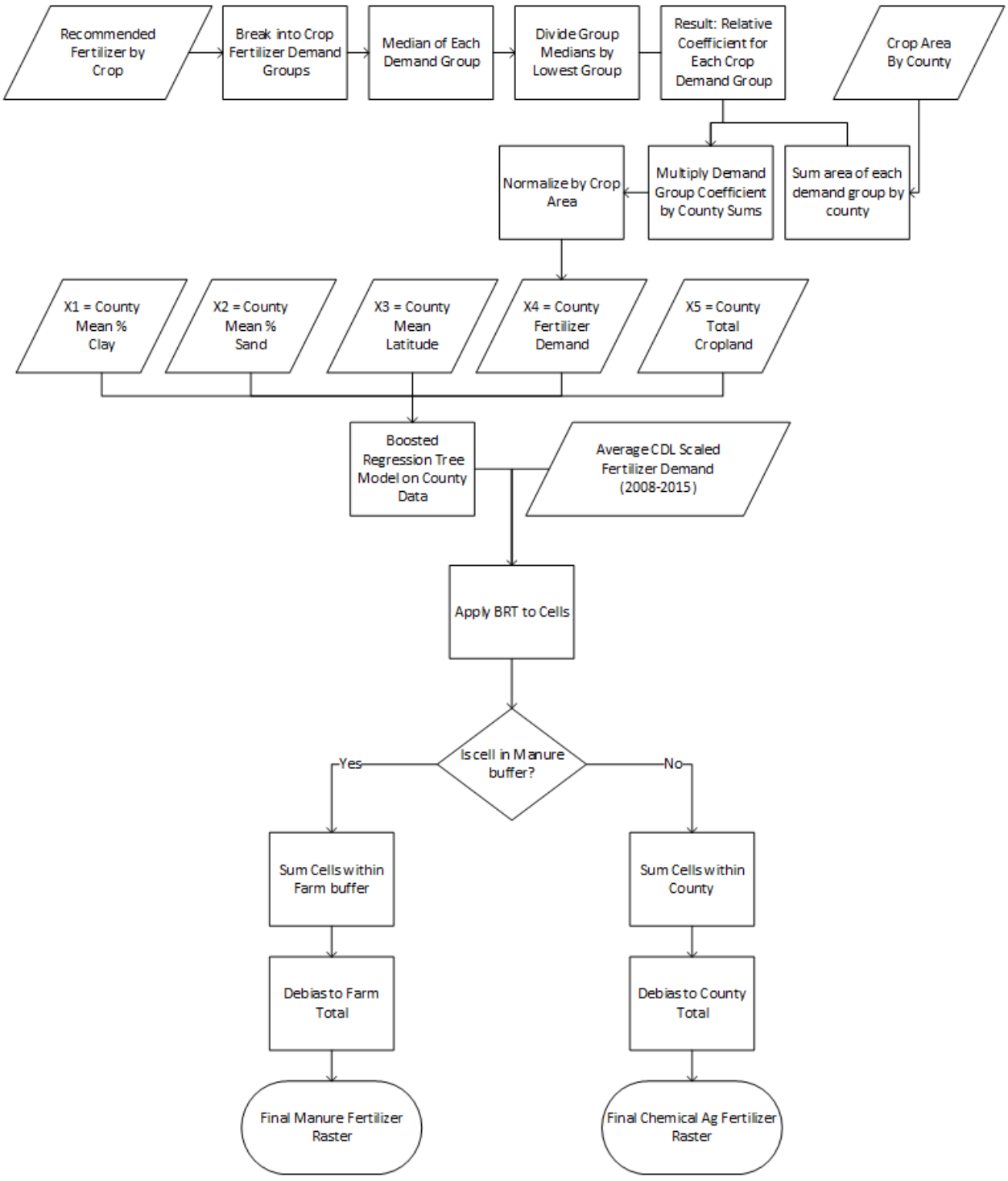

Figure A1.8 Chemical Agricultural Fertilizer Placement flowchart.

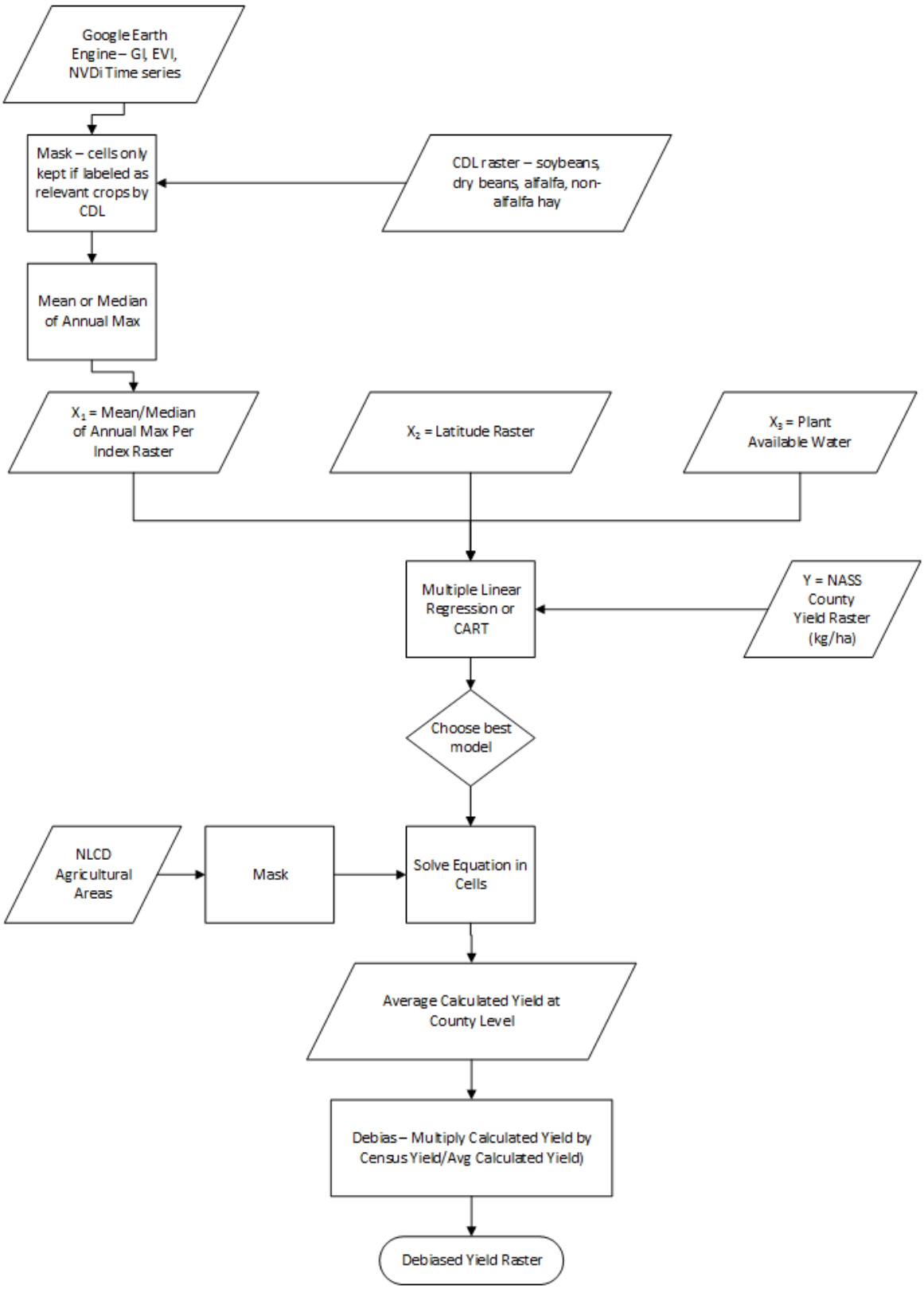

Figure A1.9 N Fixation Yield Calculation flowchart

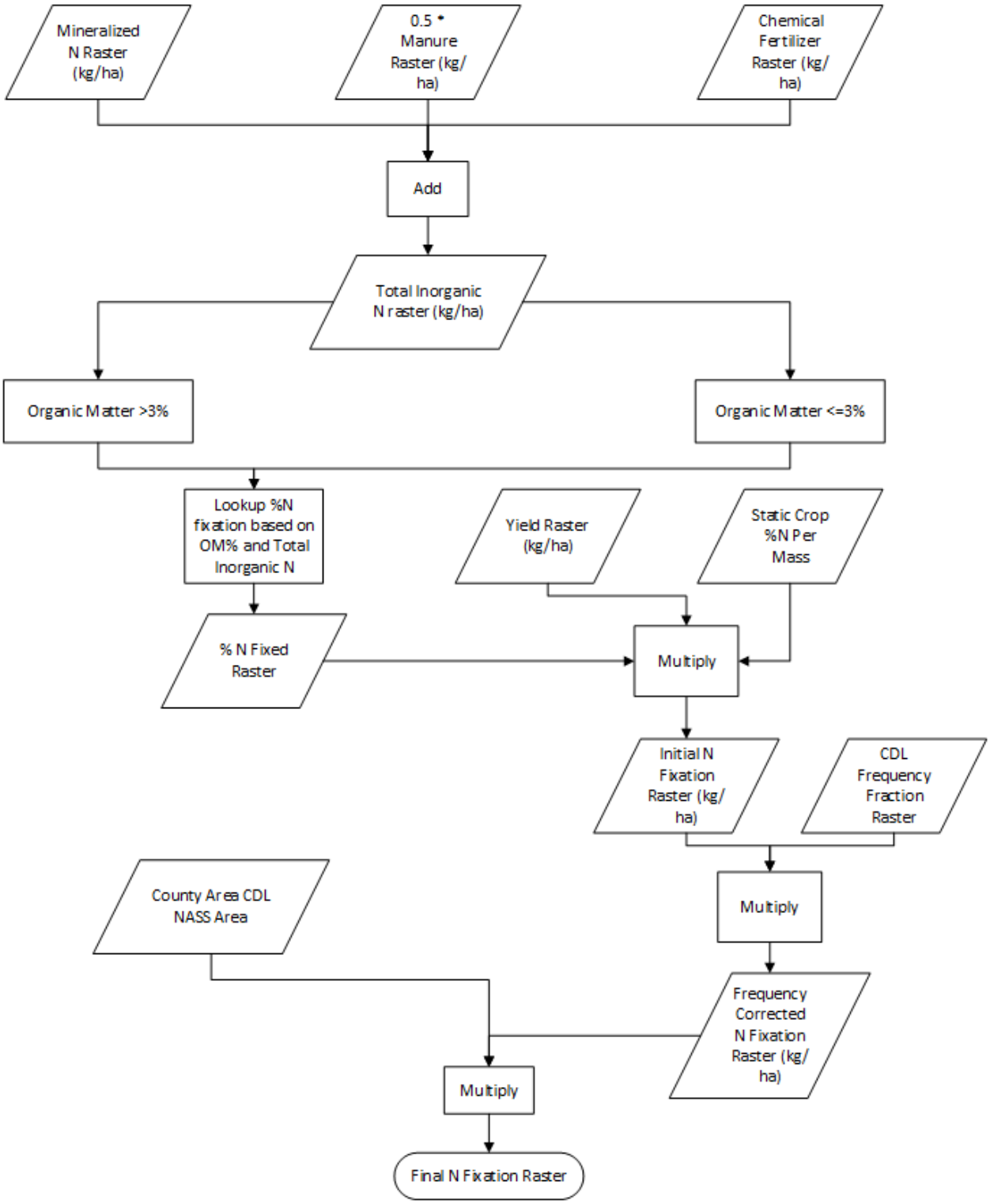

Figure A1.10 N Fixation Calculation flowchart.

Text A1.2 K-means clustering of watersheds to produce Nutrient Input Landscapes

Data clustering, or cluster analysis, uses algorithms to find underlying, natural groups or patterns in a *representation* of objects based on *similarities* (Jain, 2010). The term representation refers to how objects within a dataset are described – for example, we could cluster watersheds based on land use, population, flow, or nutrients and we would find different groups based on their similarity. Additionally, the way we quantified the variables of choice could influence the clusters – in that using total area of each land use compared instead of percent land use could create different groups. Similarity is a broad term and how similarity is determined is decided by the clustering algorithm used (Jain, 2010).

In our analysis, we selected k-means clustering to determine if there were natural groupings of watersheds based on their relative composition of nutrient sources. K-means clustering is one of the oldest, simplest, and widely used clustering algorithms, dating back in multiple fields to the 1950s and 1960s (Jain, 2010). K-means creates clusters for a set of n-dimensional points into an a priori determined *K* number of clusters where the squared error of a point and the mean cluster centroid is minimized (Jain, 2010).

We experimented with different *representations* of watershed nutrients by using rates of nutrient input sources in kg/ha, including total nitrogen and phosphorus, and using percent nutrient input by nutrient source. We experimented with the value of *k*, which determines the number of clusters, as well as running multiple times with different initialization random seeds. Silhouette coefficients are used in clustering analysis to describe how cohesive clusters are. Silhouette coefficients are calculated by comparing the distance of a sample to the mean distance to other samples in the same clusters compared to the mean distance of the next nearest cluster. Figure A1.11 visualizes the silhouette coefficients from our Nutrient Input Landscapes. We found nine clusters best physically represented the data. At fewer than nine clusters, our point

source-dominated “urban core” cluster (or landscape) was lost, due to its small size. By using our clusters, we provide a way to categorize watersheds based on the 13 variables of N and P sources.

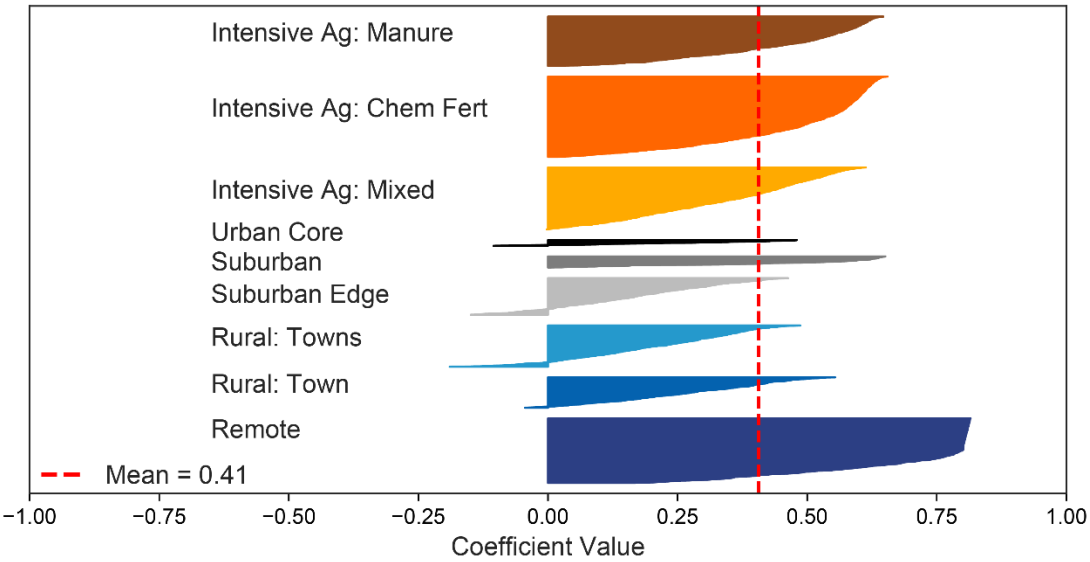

Figure A1.11 Silhouette plot for Nutrient Input Landscapes silhouette coefficients. Silhouette coefficients represent how well a sample fits into its cluster and range from -1 to 1, where negative values indicate a sample should have been classified in a different cluster. Each watershed is represented by a line, sorted by cluster, thus creating different thickness of each cluster representative of how many watersheds are in a cluster. For example, Urban Core is the smallest cluster and Remote and Intensive Ag: Chemical Fertilizer are the largest. The line extends in the x direction to its silhouette coefficient.

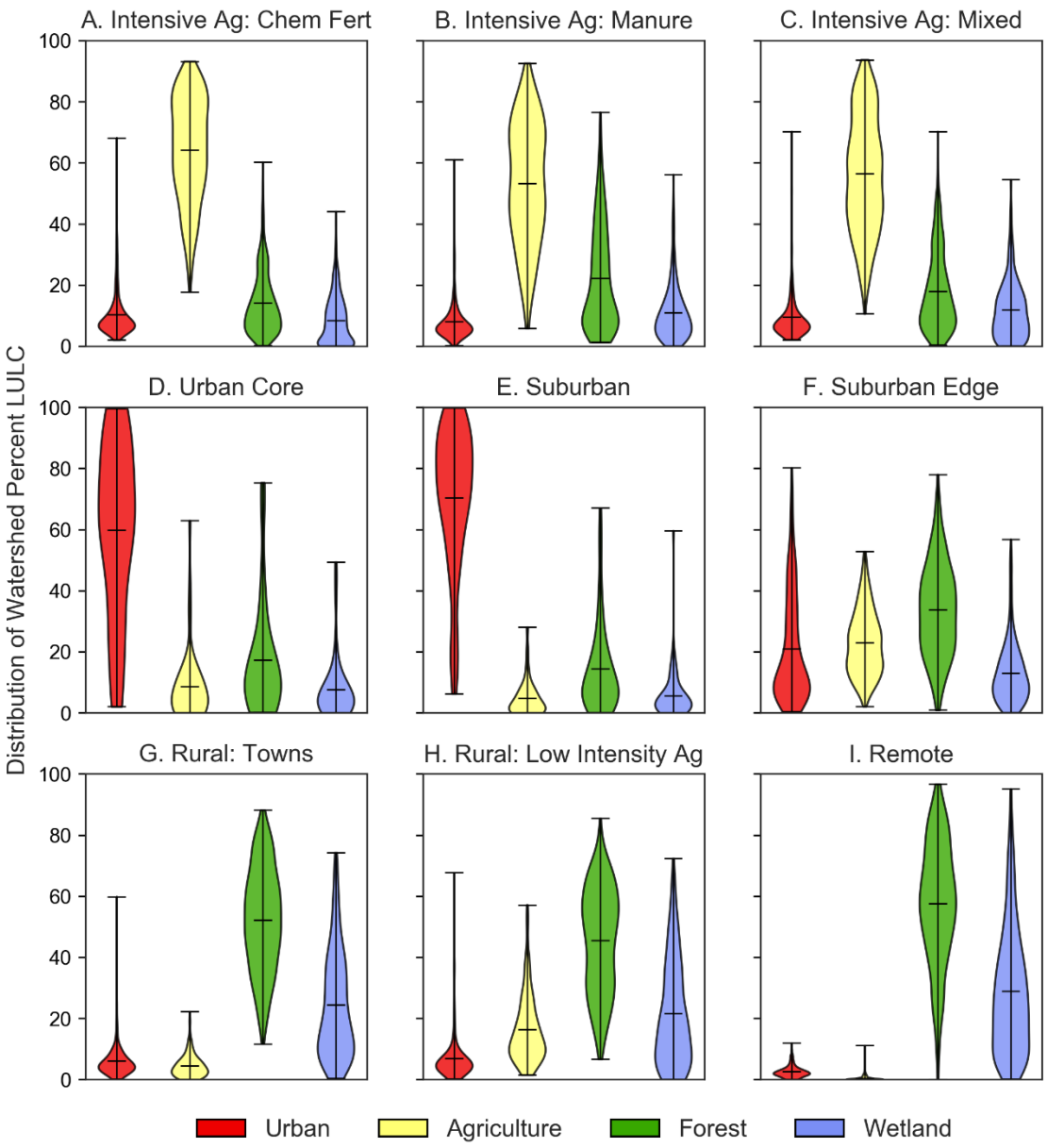

Figure A1.12 Proportion of LULC within all Nutrient Input Landscapes. The violin represents the distribution of LULUC proportions for each watershed. For example, a hash mark at ~60% for agriculture in (A) means the median percentage agriculture in Intensive Ag: Chem Fert landscapes was 60%.

Table A1.4 Summary statistics for Nutrient Input Landscapes. All 13 variables define a single landscape.

	N (%)													
	Chem NonAg Fertilizer		Atmospheric Deposition		Septic		Point		Manure		Chem Ag Fertilizer		N Fix	
Nutrient Input Landscape	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Intensive Ag: Mixed	1.2	2.3	18.9	7.5	1.9	1.7	0.4	2.7	28.8	7.4	37.6	9.0	11.2	4.5
Intensive Ag: Manure	0.8	1.6	14.9	7.7	1.2	0.9	0.2	1.0	54.0	10.1	19.4	8.4	9.5	4.7
Intensive Ag: Chem Fertilizer	1.1	2.0	18.8	7.5	2.0	1.7	0.5	2.0	8.7	5.3	54.9	8.5	14.0	5.6
Urban Core	16.2	13.1	31.6	21.0	2.2	2.6	41.4	28.9	1.5	3.5	6.0	8.7	1.3	1.7
Suburban	42.7	13.0	44.8	8.7	4.1	5.0	2.8	7.7	1.0	2.1	3.7	4.7	0.8	1.3
Suburban Edge	7.8	7.8	45.1	11.7	7.8	5.4	1.5	5.5	8.4	5.8	21.7	9.8	7.7	5.2
Remote	0.2	0.6	97.8	2.5	0.7	1.0	0.0	0.2	0.3	0.8	0.5	0.9	0.4	1.0
Rural: Low Intensity Ag	1.4	3.2	50.5	12.9	2.8	2.5	0.6	3.1	25.9	8.1	8.5	6.1	10.3	7.2
Rural: Towns	2.0	3.8	81.7	8.9	4.8	4.6	0.2	1.0	4.7	4.1	3.4	3.8	3.3	3.2
Nutrient Input Landscape	***P (%)***													
Intensive Ag: Mixed	1.0	1.9	4.3	2.3	3.8	3.4	0.6	1.8	43.2	8.9	47.1	8.4		
Intensive Ag: Manure	0.5	1.0	2.8	2.0	2.2	1.7	0.3	1.5	71.4	9.2	22.8	8.5		
Intensive Ag: Chem Fertilizer	1.0	2.0	5.0	2.4	4.4	3.7	1.0	3.1	14.2	8.4	74.5	9.3		
Urban Core	18.9	14.1	8.3	4.6	5.4	6.6	52.3	20.6	2.8	5.0	12.3	13.4		
Suburban	58.8	18.6	14.6	4.6	11.8	13.7	3.3	8.2	2.2	4.3	9.4	10.1		
Suburban Edge	7.7	8.1	13.3	6.3	19.0	11.7	2.2	5.6	15.2	10.0	42.6	12.8		
Remote	0.5	1.9	89.0	10.9	5.4	6.4	0.2	2.2	1.8	4.3	3.1	5.0		
Rural: Low Intensity Ag	1.5	3.4	16.2	7.6	7.8	5.7	1.1	4.2	52.6	11.1	20.7	10.1		
Rural: Towns	3.5	6.1	43.7	13.3	19.9	14.9	1.9	8.2	15.2	12.7	15.7	11.4		

Text A1.3 Results Figures and Tables

The following pages contain additional results related figures and tables.

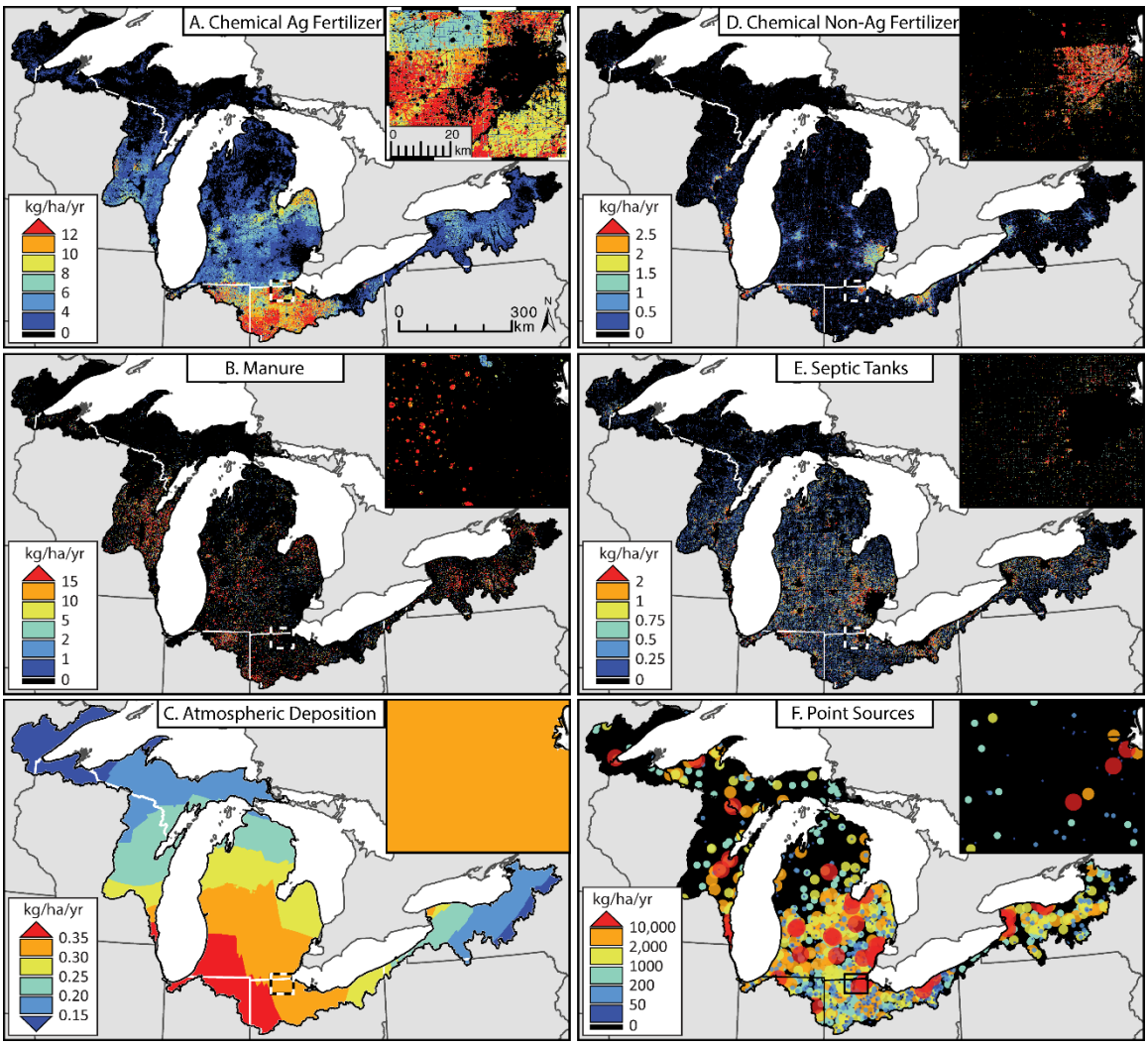

Figure A1.13 SENSEmap phosphorus sources. Color breaks are not necessarily representative of individual pixel values due to visual resampling. Colors are meant to show physically meaningful patterns in the data visible at this map scale. Pixel level distributions of sources can be found in Table A1.4.

Table A1.5 Pixel level quantile values for N nonpoint source maps. Zeroes are excluded.

	N (kg/ha/yr)					
Percentile	Chem NonAg Fertilizer	Septic	Atmospheric Deposition	Chem Ag Fertilizer	Manure	Nitrogen Fixation
0	0.59	11.4	3.6	0.01	0.74	0.01
20	28.4	93.2	6.9	40.4	250.4	4.4
40	41.6	112.1	8.5	56.1	346.6	11.9
60	51.7	125.0	10.0	69.9	451.7	20.6
80	67.8	144.1	11.4	87.5	615.7	30.0
100	167.6	1000.0	22.1	477.1	915.6	301.7

Table A1.6 Pixel level quantile values for P nonpoint source maps. Zeroes are excluded.

	P (kg/ha/yr)				
Percentile	Chem NonAg Fertilizer	Septic	Atmospheric Deposition	Chem Ag Fertilizer	Manure
0	0.08	2.7	0.11	0.10	0.12
20	4.1	22.5	0.17	4.6	46.0
40	6.0	27.0	0.21	6.1	62.3
60	7.6	30.1	0.28	7.9	83.4
80	10.3	34.7	0.33	11.4	119.4
100	28.3	200.0	0.43	27.9	198.4

Text A1.4 Comparisons to other Products

The following sections detail methodology and results from comparisons made between SENSEmap and other nutrient products.

Text A1.4.1 SPARROW

Table A1.7 shows mean and standard deviation of absolute difference between nutrient inputs at the HUC8 level in units kg/ha/yr, where positive values indicate SENSEmap had higher values than SPARROW. SENSEmap and SPARROW have a few key differences in data sources and temporal resolution. SENSEmap-GLB uses the 2012 Agricultural Census for manure calculations and Brakebill & Gronberg (2017) data for chemical fertilizer (2008-2012), whereas SPARROW uses earlier data from Ruddy et al. (2006) for both manure and chemical fertilizer. SENSEmap also uses more contemporary state-level CAFO datasets, which may have raised the total confined manure quantities. Additionally, each model used different approaches to spatially disaggregate datasets and distribute them across watersheds, thus creating additional variance between the two datasets.

Table A1.7 Summary statistics comparing SENSEmap and SPARROW. Mean and standard deviation of difference between SENSEmap and SPARROW. Positive values indicate SENSEmap had higher values than SPARROW.

SPARROW	Source	Difference (kg/ha/yr)
N	TN	2.1 +/- 8.5
	Total Manure	4.9 +/- 8.6
	Chem NonAg Fert.	0.1 +/- 1.1
	Chem Ag Fert	-2.8 +/- 5.8
P	TP	-0.1 +/- 1.2
	Total Manure	0.6 +/- 1.3
	Chem NonAg Fert.	-0.1 +/- 0.3
	Chem Ag Fert	-0.7 +/- 1.1

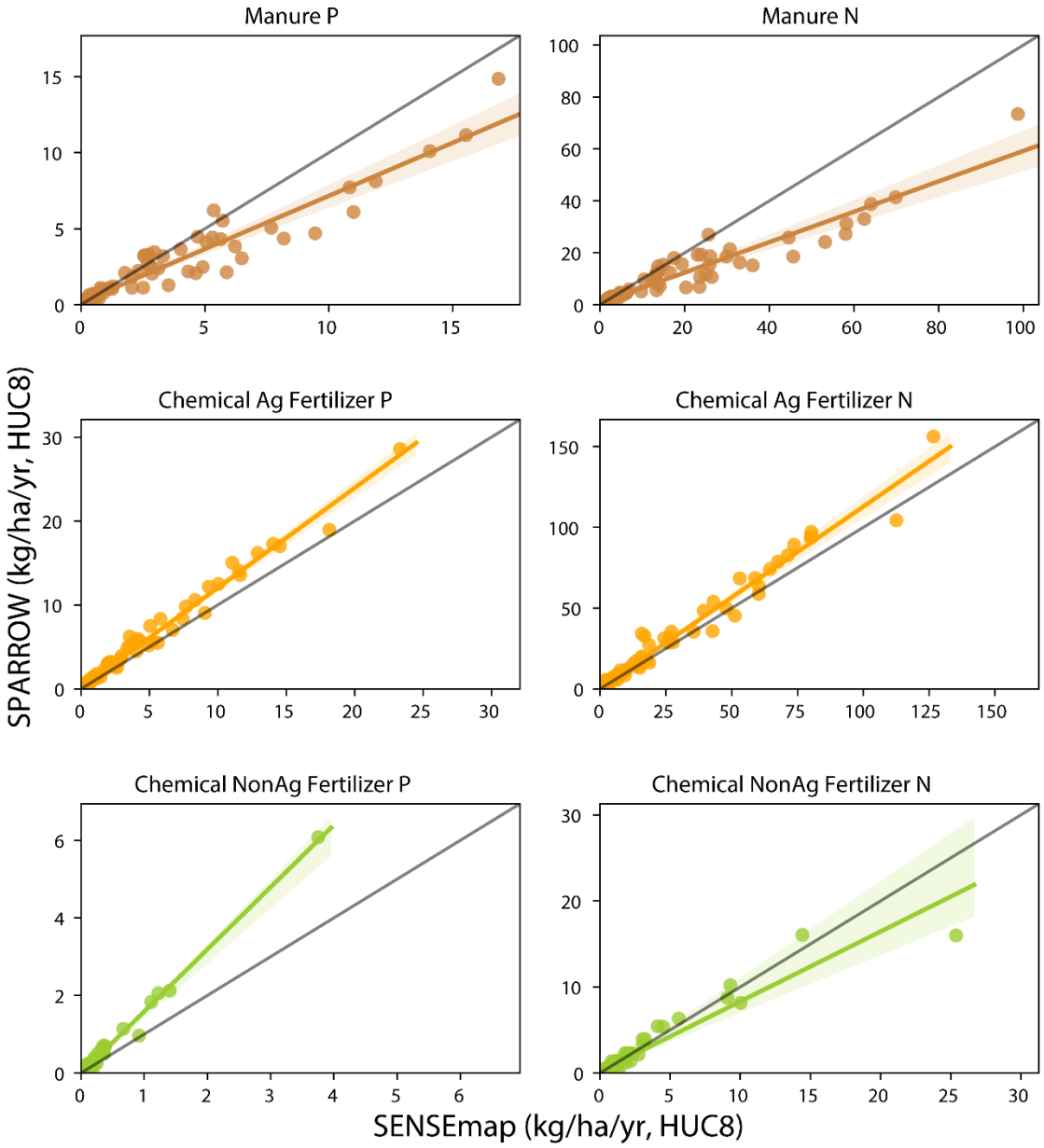

Figure A1.14 HUC8 level comparison of SPARROW and SENSEmap. kg/ha/yr. Grey line indicates 1:1.

Text A1.4.2 Net Anthropogenic Nitrogen Index (NANI) in Swaney et al. (2018)

SENSEmap was compared to the Net Anthropogenic Nitrogen Index (NANI) as presented in Swaney et al. (2018). N inputs in kg/ha/yr were calculated at the county-level using total county area. 2012 inputs were compared due to SENSEmap using the 2012 Agricultural Census and fertilizer data ending in 2012. Figure A1.15 shows bivariate plots for chemical agricultural fertilizer, manure, and N fixation.

Table A1.8 Summary statistics comparing SENSEmap and NANI. Mean and standard deviation of difference between SENSEmap and NANI (Swaney et al., 2018) at county level. Positive values indicate SENSEmap was higher than NANI.

NANI	Source	Difference (kg/ha/yr)
N	Total Manure	3.8 +/- 6.2
	Chemical Agricultural Fertilizer	-2.7 +/- 5.6
	N Fixation	7.1 +/- 15.9

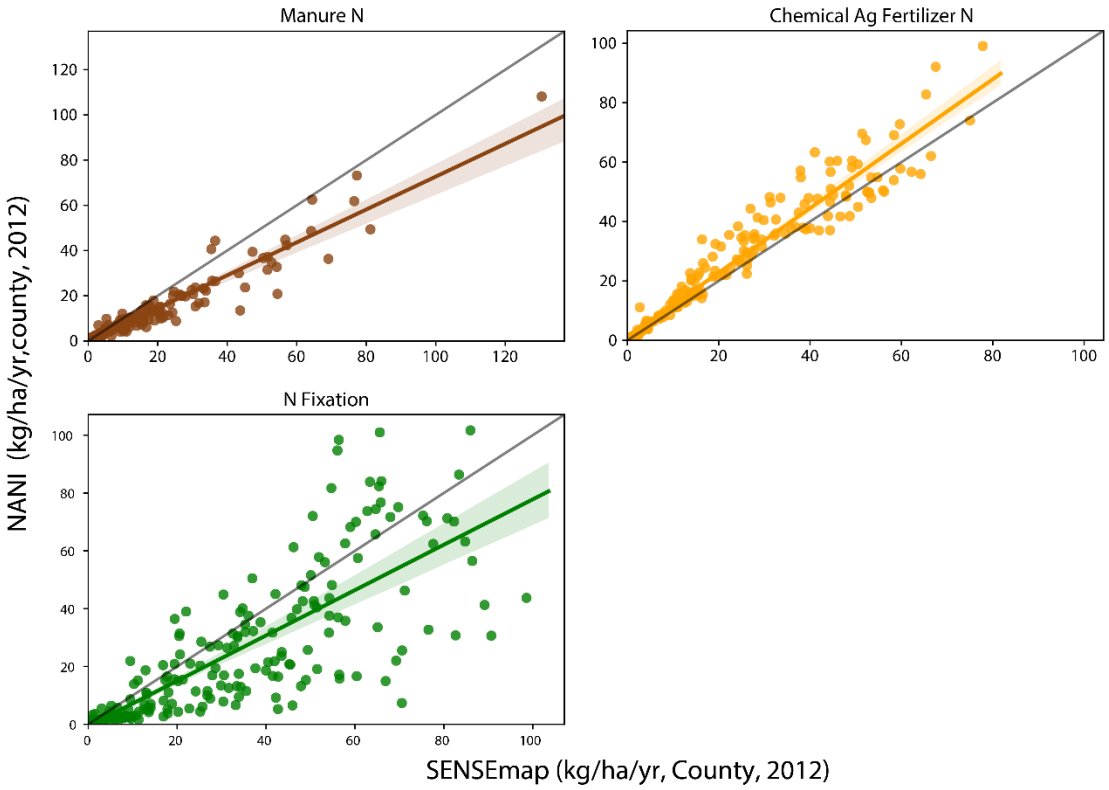

Figure A1.15 County level comparisons between SENSEmap and NANI. kg/ha/yr. Grey line indicates 1:1.

Text A1.4.3 Chemical Agricultural Fertilizer – Cao et al. (2018)

We compared SENSEmap's chemical agricultural fertilizer N to Cao et al.'s (2018) 5 km N fertilizer product. SENSEmap was resampled to 5 km resolution with mean fertilizer intensity to match Cao et al. (2018). SENSEmap generally had lower fertilizer intensities in areas with higher manure content (Wisconsin, northeastern Indiana, western Michigan) and higher fertilizer intensities in intensely farmed areas like Ohio and the thumb of Michigan (Figure A1.16).

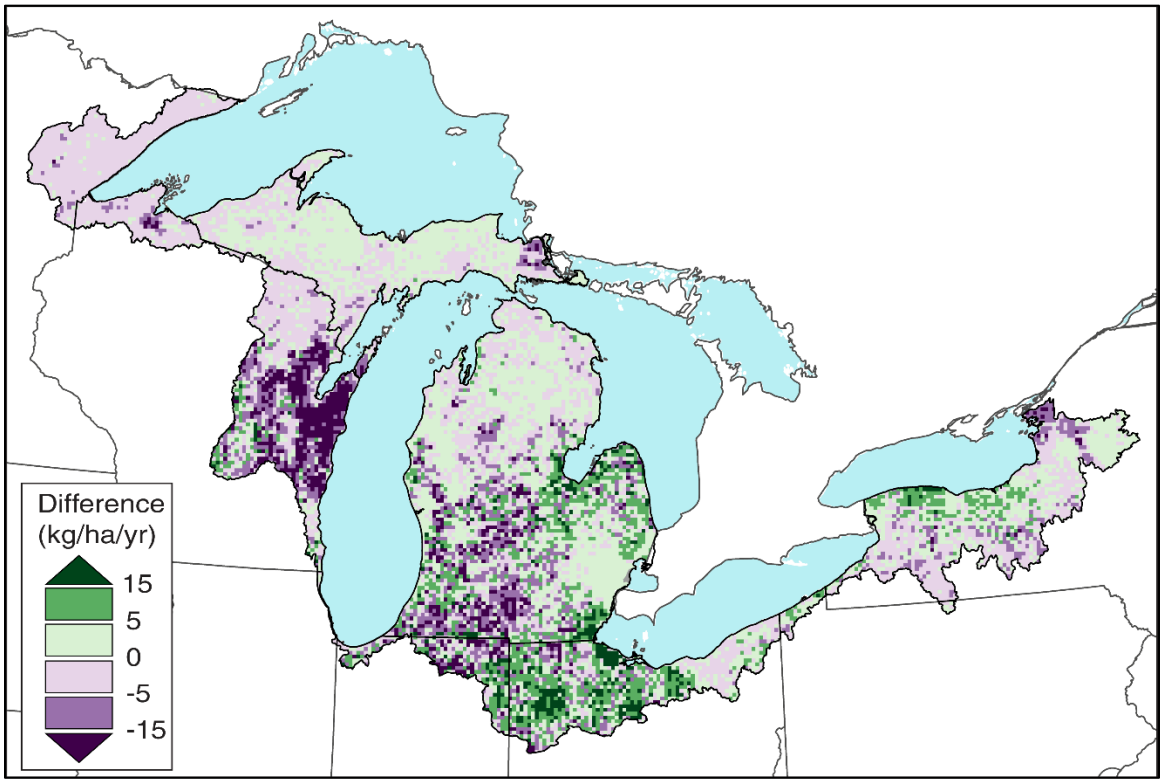

Figure A1.16 Difference map between SENSEmap and Cao et al. (2018) N fertilizer. Positive difference indicates SENSEmap > Cao et al. (2018).

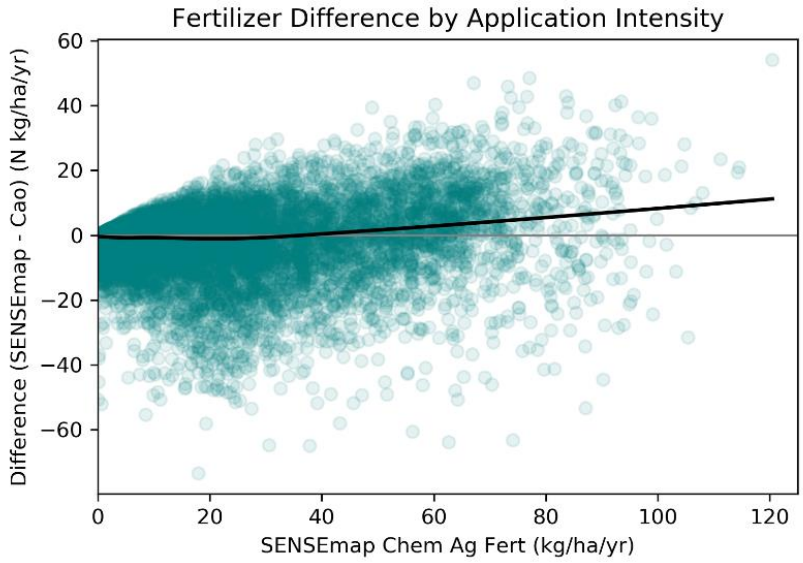

Figure A1.17 Pixel level differences by SENSEmap fertilizer application intensity. The difference between Cao et al. (2018) and SENSEmap is plotted on the y-axis, with positive values indicating SENSEmap had higher fertilizer application and negative values indicating Cao et al. (2018) was higher. The grey line marks zero and the black line is a fitted lowess curve.

Text A1.5 Sensitivity to Chemical Agricultural Fertilizer

We performed a sensitivity analysis based on the "observed" value of total kg N and P of chemical agricultural fertilizer at the county level. Our final model uses the average value from 2008-2012 from Brakebill & Gronberg's (2017) county annual fertilizer product. In order to test SENSEmap's sensitivity to this total fertilizer, we ran 14 new simulations with different values of chemical agricultural fertilizer at the county level. County-level values are used within SENSEmap to adjust pixel level predictions by the ratio needed to match simulated values to observed county N and P. However, these values are not "observed" and are also model outputs from the USGS's fertilizer modeling based on expenditures.

We produced 14 new simulations of pixel level chemical agricultural fertilizer N and P. New county fertilizer values were chosen based on the following methods. First, mean and standard deviation of N and P kg was calculated for both 2000-2012 and 2008-2012 in each county. One standard deviation was added and subtracted to each mean per county, creating 4 runs where values were systematically shifted positively or negatively. The other method created random noise by using a random adjustment uniformly between +/- 1 standard deviation per county, with 5 runs using the 2000-2012 distribution and 5 runs using 2008-2012 distribution. This resulted in 14 new simulations of county-level chemical agricultural fertilizer.

After pixels were adjusted for each simulation, HUC12 watershed aggregations were performed. Coefficient of variation (standard deviation/mean) was calculated across all simulations per watershed for chemical agricultural fertilizer N and P (kg/yr) and total N and P (kg/yr) to describe the width of the distribution across runs.

Finally, to assess changes in broader nutrient patterns, we applied our Nutrient Input Landscape cluster definitions to each sensitivity simulation. K-means clusters produce a centroid value for each cluster and distance is computed to find which cluster a sample belongs in. Thus,

we can apply clusters trained on a different dataset (i.e. primary SENSEmap) to predict which cluster any watershed would fall in. If the product is not sensitive to these changes in chemical agricultural fertilizer, Nutrient Input Landscape should always be the same. 85% of watersheds remained consistent through all sensitivity simulations.

CHAPTER 2:
CONNECTING LANDSCAPE CHARACTERITICS TO GROUNDWATER NITRATE CONCENTRATION

Abstract

Nitrate in groundwater has become a growing concern due to recent studies suggesting that even low concentrations can lead to elevated cancer risk. Data on groundwater nitrate concentrations is not widely available due to varying regulations and sampling limitations. Here, we utilize a dataset of 300,000+ nitrate samples collected across Michigan's Lower Peninsula since the 1980s to better understand drivers of groundwater nitrate concentrations. We focus on the 2006-2015 period and evaluate the probability of exceeding 0.4 mg/L and 2 mg/L concentrations using interpolated nitrate concentrations. Classification and Regression Tree (CART) analysis was used with watershed-level summaries of concentration exceedance probabilities with a suite of potential driver variables, including land use, geologic attributes, soil characteristics, and nitrogen loading. CART divides a dataset into groups that minimize within-group variance based on the highest performing split in driver variables. CART explained 43.2% of variation in the >0.4 mg/L case with only six terminal groups, which is a strong performance for a small final tree and a complex dataset. Aquifer recharge was identified across all analyses as the most important initial variable to separate low and high concentration watersheds. A combination of deep soil variables (texture and saturated hydraulic conductivity) and land use variables further separated low, medium, and high probability groups. Our findings suggest that sufficient recharge is necessary to mobilize nitrogen, and that even forested or low intensity agricultural areas can load nitrate to aquifers if high recharge and vulnerable geologic features are present. These findings can improve identification of high nitrate risk areas and determine how and where management strategies will affect aquifers based on climatologic and geologic vulnerability.

1 Introduction

Nitrogen cycling has been significantly altered by anthropogenic activities with the rise of industrial agriculture and growing population (Keeler et al., 2016; Vitousek et al., 1997). Nitrogen sources including chemical fertilizer, manure, and human waste have created a suite of water quality challenges that affect the environment, human health, and economics. Nitrogen loading to groundwater can threaten human health when ingested via drinking water. As early as the 1930s, high concentrations of nitrate in drinking water from wells in agricultural areas was shown to cause infant methemoglobinemia, also known as blue baby syndrome, a fatal condition that renders hemoglobin unable to carry oxygen throughout the body (Ward et al., 2005). These high nitrate loads were a result of increasing prevalence of chemical fertilizers and resulted in public health regulations limiting drinking water to less than 10 mg/L NO_3-N (Walton, 1951).

In recent years, researchers have begun to link nitrate in drinking water to higher prevalence of multiple cancers and birth defects, such as colorectal cancer, thyroid cancer, and adverse pregnancy outcomes (Manassaram et al., 2006; Ward et al., 2005, 2018). In contrast to infant methemoglobinemia, these risks occur at much lower nitrate concentrations in drinking water. Ward et al. (2018) reviewed epidemiology literature and found many studies with increased risk of various cancers associated with exposure to drinking water nitrate at concentrations as low as 1.5 mg/L NO_3-N, with multiple studies finding heightened risk around 2 mg/L NO_3-N (Schullehner et al, 2018; Temkin et al., 2019) . As these health risks become quantified, it is increasingly important to understand what environmental conditions lead to high groundwater nitrate concentrations, and through both this and more extensive sampling identify communities at risk. Additionally, nitrate in groundwater is a consistent and long-term source of

nitrogen to streams and lakes, making it important to understand when constructing nutrient budgets and addressing eutrophication (Benettin et al., 2016; Vero et al., 2017).

Limited data on groundwater nitrate is available due to varying regulations combined with the effort, cost, and privacy concerns related to sampling wells. There can be highly heterogeneous patterns of nitrate concentrations in the subsurface due to variable loading pulses, geologic characteristics, and the laminar flow characteristics of groundwater (Canter, 1997). Thus, extensive sampling would be required to effectively map areas. These challenges have led researchers to turn to a variety of index-based, statistical, and process-based modeling methods to both understand drivers of groundwater nitrate concentration and predict nitrate concentration at unsampled locations.

Recently, machine learning methods like classification and regression trees (Burow et al., 2010), boosted regression trees (Motevalli et al., 2019; Nolan, Fienen, & Lorenz, 2015; Ransom et al., 2017), and random forest (Messier et al., 2019; Wheeler et al., 2015) have been used to link large datasets of nitrate samples to driver variables to predict concentrations at non-sampled locations. In addition to producing non-linear predictions, these tree-based algorithms assess the importance of different variables in describing concentrations, which can help decipher the underlying conditions that lead to elevated nitrate concentrations in a region. Large datasets of well samples are frequently used to train the models with related variables summarized within circular buffers. These studies have ranged from continental to small watershed scales and have had varying predictive success. Studies have reported the most significant driver variables falling into categories related to nitrogen inputs (Nolan et al., 2014; Nolan & Hitt, 2006), geologic properties (Barzegar et al., 2018; Messier et al., 2019; Motevalli et al., 2019), redox conditions (Ransom et al., 2017), and well depth (Wheeler et al., 2015).

Here, we analyze drivers of groundwater nitrate concentration with classification and regression tree analysis (CART) using an extensive dataset of nitrate measurements from drinking water wells across Michigan's Lower Peninsula (LP) in the Midwest United States. Over 300,000 samples from 76,724 unique wells were used to characterize nitrate concentration across the region and interpolated using kriging. Probabilities of exceedance for two concentration levels were calculated from kriging and aggregated to a small surface-watershed (HUC12) scale to analyze using CART with 29 physical and management variables (USGS, 2013). Unlike other machine learning studies published to date, we utilize the high sampling density in our dataset to interpolate and generate training data using watershed summaries instead of circular buffers surrounding sparse individual well points. The objectives of this work are to:

1. Map nitrate concentration in Michigan's Lower Peninsula during 2006-2015 using over 76,000 unique wells with more than 300,000 samples.
2. Assess drivers of nitrate concentration using landscape characteristics at small watershed level with CART.

With this extensive dataset, we provide maps of nitrate concentration in groundwater and link them to the physical processes and land use management. By understanding the drivers of groundwater nitrate concentrations, we can inform management strategies and quantify nitrate exposure.

2 Study Area

Michigan's Lower Peninsula (hereafter LP) is located in the Midwestern United States and bordered by four of five Laurentian Great Lakes (Figure 2.1B). The state of Michigan includes the largely forested Upper Peninsula, however it was not included in this analysis due to limited data.

Land use within the LP is highly variable, with urban, intensive agriculture, and large swaths of forests and wetlands throughout the state (Homer et al., 2015). There is a significant north-south divide in land use, where major urban areas and extensive agriculture lie south of 44º N (Figure 2.1A). Metropolitan areas such as Detroit and Grand Rapids are located on the southeastern and western central regions, home to most of Michigan's over 9 million residents (U.S. Bureau, 2020). Much of the land area of the southern LP is under agricultural management, including corn-soy rotations, wheat, and hay in major row- and field- cropped areas and orchards along the coast of Lake Michigan (USDA Ag Census, 2012). North of 44º N, the LP is largely forested and sparsely inhabited.

The Quaternary period deposited much of the glacial and post-glacial alluvial geology making up aquifers commonly used for private drinking water wells in the LP. Coarse-textured glacial deposits cover much of the LP, with the exception of increased clay and silt in the central and eastern LP, which was formerly a lakebed. Quaternary glacial drift aquifers are used for private drinking water wells for much of the LP due to sandy, high conductivity sediments. This aquifer is deepest in the northwest aquifer thickness exceeds 100 meters (Figure A2.1, Soller and Garrity 2018). Soil saturated hydraulic conductivity (K_{Sat}) ranges from 3-550 mm/hr with lower values in the south central and eastern former lake beds, and higher conductivity across the western and northern LP (Figure 2.1D). These patterns correspond with soil texture (Figure

A2.2). Bedrock wells are used where glacial drift is limited (Figure A2.1, <5 meters). Aquifer recharge, an important characteristic for groundwater nitrate analysis, exhibits an east-west gradient due to a combination of highly conductive soils and high precipitation in the west (Figure 2.1C). The western half of the LP has increased precipitation in the form of lake effect precipitation due to its proximity to Lake Michigan. Annual precipitation for the LP ranges from approximately 700-1100 mm/yr (PRISM Climate Group, 2012).

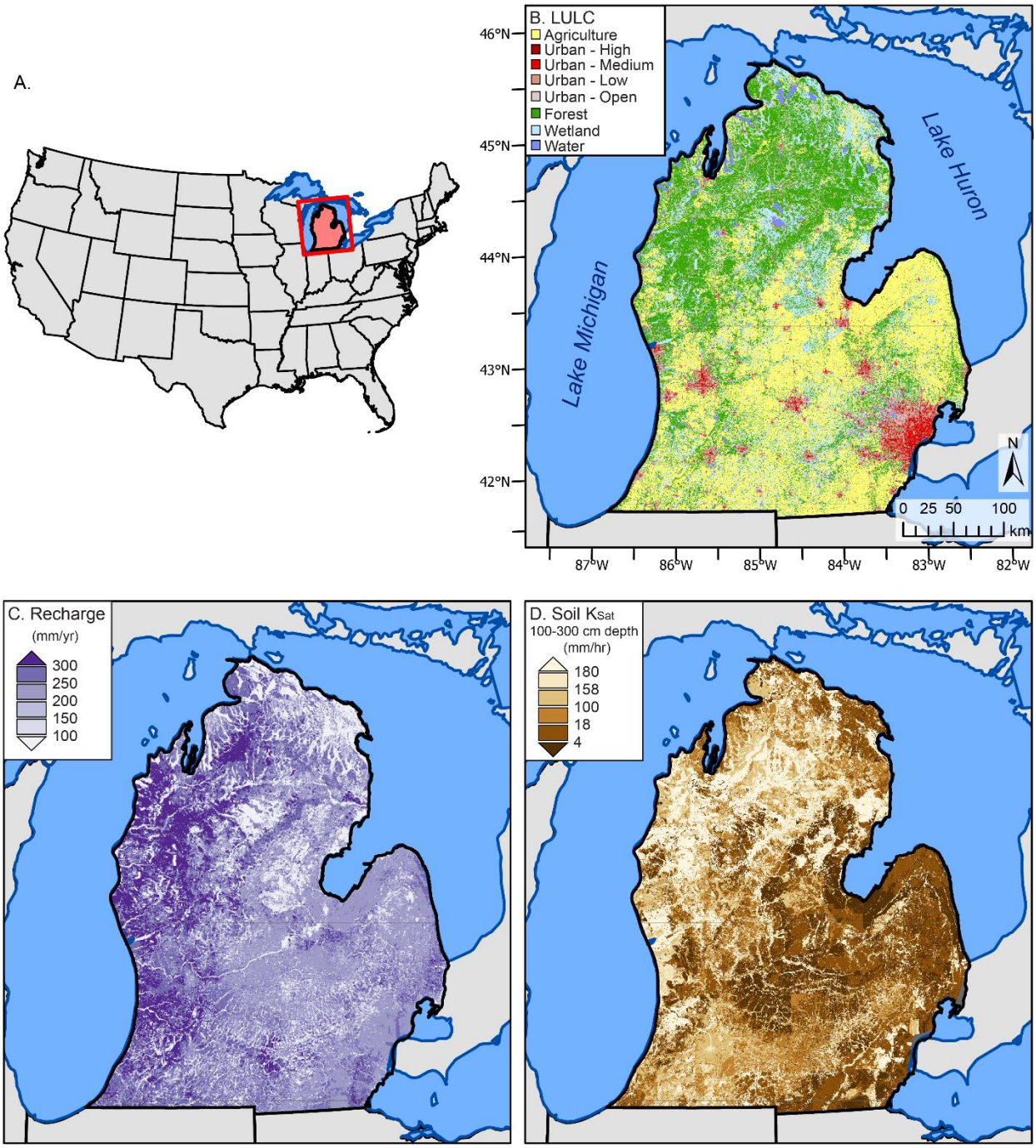

Figure 2.1 Groundwater nitrate study area. Michigan's Lower Peninsula (LP) with key environmental variables shown. A) Locator map of Continental United States with study area highlighted in red, B) Land use/land cover map modified from NLCD 2011 (Homer et al., 2015), C) Modeled map of yearly aquifer recharge (See Section 2.3.6), D) Map of soil saturated hydraulic conductivity for 100-300 cm depth from SSURGO with quantile breaks (NRCS, 2017).

3 Data

The following subsections briefly describe the data sources used for this study. Table 2.1 summarizes these sources.

Table 2.1. Summary of data sources. Detailed variables found in each group can be found in Table A2.1.

Data Type	Data Source	Author	Time	Spatial Resolution
Nitrate	Well Chemistry	MI EGLE	2006-2015	Well Points
Well Geology	Wellogic	MI EGLE	2006-2015	Well Points
Groundwater N	SENSEmap	Hamlin et al (2020), Wan et al (In prep)	2008-2015	120 m
Soil Properties	gSSURGO	USDA NRCS		30 m
Aquifer K_{Sat}	Wellogic	Wellogic, Farrand & Bell (1982)		120 m
Recharge	*Modeled*	Hyndman et al. (2007); Wan et al. (In prep)	2010	120 m
Land Use	NLCD	Homer et al. (2015)	2011	30 m

3.1 Well Chemistry

Well chemistry data was retrieved via a Freedom of Information Act request from the Michigan Department of Environment, Great Lakes, and Energy (EGLE) in December 2019. This database included over 3.6 million samples of various chemicals from drinking water wells across the state dating from 1984 to 2019. Well chemistry data was provided as two categories of wells, described here as “public” and “private”. Public wells include public supply wells, industrial wells, and wells belonging to businesses or organizations. Private wells, which are those used in private residences, were sampled when drilled, generally resulting in a single chemistry sample per well, whereas public wells are routinely monitored depending on their category of use.

3.2 Wellogic

EGLE provides a publicly available database of drinking water well information digitized from drillers' boring logs called Wellogic, which contains all wells drilled since 1996 and partial records of wells drilled prior to 1996 due to variability in county-by-county archival digitization efforts. Information includes date of drilling, screening intervals, and aquifer properties. In this study, we utilized both the spatial location of the wells, as well as information on which aquifer each well was screened within.

3.3 Nitrogen Loads

The Spatially Explicit Nutrient Source Estimate map (SENSEmap) was used to quantify total nitrogen inputs to groundwater (Hamlin et al., 2020; Wan et al., In Prep). SENSEmap quantifies point sources and six non-point sources of nitrogen inputs to the landscape in Hamlin et al. (2020) by synthesizing literature, remote sensing products, government records (ie., US Census, US Agricultural Census), and modeling products (ie. Estimates of county level fertilizer loads). Sources include atmospheric deposition, chemical agricultural fertilizer, manure, chemical non-agricultural fertilizer, septic tanks, N fixation from legumes, and point sources. Wan et al. (In prep) implements a statistical transport model to quantify how much nitrogen survives to the Great Lakes coastline after attenuation along surface and groundwater pathways. This model was updated from previous work by Luscz et al. (2017). In this study, we use two model outputs: the total nitrogen (TN) load to groundwater per year from all surface-applied sources (kg/ha/yr) and the TN load from septic tanks to groundwater (kg/ha/yr). Both estimates are representative of an average year during 2008-2015. See Figure A2.3 for the map of TN loads to groundwater.

3.4 Land Use/Land Cover

Land Use/Land Cover (LULC) variables were summarized from the 2011 National Land Cover Database, shown in Figure 2.1 (Homer et al., 2015). Both individual land use classes as reported in NLCD and aggregated land use classes (ie., Total urban, total forest, total agriculture) were used within analysis. Land use per category was tabulated and normalized to the proportion of the watershed.

3.5 Soil

Soil variables were extracted from gSSURGO and processed using ROSETTA (NRCS, 2017; Schaap et al., 2001). This study used data from soil layer 1 (0-5 cm) and soil layer 4 (100-300 cm). Variables included were texture by percent (Figure A2.2) and saturated hydraulic conductivity (K_{Sat}) (Figure 2.1D).

3.6 Aquifer Recharge

Aquifer recharge was calculated using outputs from the Landscape Hydrology Model (LHM) and statistical relationships derived from those results. Modeling was performed over a 28-year period for an approximately 20,000 km^2 within the Muskegon HUC8 watershed in west-central LP (Hyndman, Kendall, & Welty, 2007; Kendall, 2009). This watershed has diverse geologic and land use characteristics reasonably representative of conditions found across the remainder of the LP. Within each of five major land use classes (urban, agricultural, grass, deciduous, coniferous), linear regressions were developed to quantify the fraction of annual precipitation that has been simulated to become aquifer recharge as a function of soil hydraulic conductivity.

3.7 Aquifer Saturated Hydraulic Conductivity

Aquifer saturated hydraulic conductivity (K_{Sat}) was estimated by the Michigan State University Remote Sensing & GIS using well log descriptions of sediment texture and pump tests to determine K values and was reported in the Michigan Wellogic Database (*Michigan Waterwells for WELLOGIC dataset*, 2019). To extend these well-based measurements to the rest of Michigan, we used the geometric mean of well-based K for each Quaternary geologic polygon (Farrand & Bell, 1982). Polygons without wells used the average K for other polygons of similar geologic class. See Figure A2.4 for this map of aquifer conductivity.

4 Methods

Utilizing the large well chemistry dataset collected, analyzed, and stored by a variety of organizations and individuals resulted in the need for extensive pre-processing and quality assurance/quality control (QA/QC) protocols. For more details, see section A2.1. Briefly, we used the following major steps to prepare well chemistry data for use: 1) Geocode data, or identify spatial coordinates for each point, 2) Perform QA/QC on geocoded addresses, 3) Perform QA/QC on concentration measurements, and 4) Join well chemistry data to well geologic data from Wellogic. At this time, only nitrate measurements were used, while future work will include processing sulfate and nitrite for use in this analysis. The 2006-2015 period was used in this analysis to incorporate a large volume of data and match the time period of a key analysis dataset, SENSEmap, which represents the average nitrogen inputs expected within 2008-2015.

4.1 Kriging

Kriging, a geostatistical interpolation method, was used to create two distinct map types: 1) a continuous surface of groundwater nitrate concentrations, and 2) a similar surface of the probability for exceedance of different nitrate concentrations. Kriging fits a function to the spatial autocorrelation (here, semivariogram) calculated from known points in a geographic dataset, assuming that some variation between points is due partially to randomness and to the distance between points (Bailey & Gatrell, 1995). Depending on the data, the weight given to each point near a prediction cell varies by the shape of the chosen semivariogram function. Due to the large spatial extent, variable sampling density, and variation in concentration measurements, we used Empirical Bayesian Kriging (EBK) in ArcGIS Pro 2.5 to automate the kriging process (ESRI, 2020a). This method iteratively produces semivariograms for subsets of

the data to tailor the kriging function to each neighborhood of points, thus removing the user's manual control of fitting a semivariogram function to the entire dataset (ESRI, 2020b; Hussain, Pilz, & Spoeck, 2010).

Probability kriging is a method that computes the probability of a cell's kriging prediction to exceed a given threshold. This provided additional information beyond kriging predictions. The probability kriging option within EBK was used to test four thresholds of exceedance in this analysis: 0.4 (the most common detect limit, or lowest detectable concentration of nitrate by the machine), 2, 5, and 10 mg/L NO_3-N. 2 mg/L NO_3-N was chosen to be representative of new health literature indicating increased risk at values from >2 mg/L NO_3-N (Schullehner et al., 2018; Temkin et al., 2019). 5 mg/L NO_3-N was chosen as a mid-level and 10 mg/L NO_3-N was chosen as the EPA's maximum contaminant limit (MCL) for public drinking water. Results are only be presented here for exceedance of 0.4 mg/L and 2 mg/L NO_3-N.

Although kriging produces a prediction at every cell within the study area, error rises with distance from known data points. We only included kriging results within 3 km of a sample point to eliminate areas where kriging does not have enough data, and predictions approach the dataset mean rather than being influenced by nearby points. This threshold was chosen because as it is the length of correlation, or range, found from a preliminary simple kriging of the entire dataset. Essentially, these 3 km buffered inclusion masks allow for the benefits of kriging to extrapolate and spatially debias the point data, without over-extending the kriged extrapolations to areas distant from sampling locations.

4.2 CART

Classification and Regression Trees (CART) analysis was used to explore nonlinear statistical relationships between groundwater nitrate concentration and potential driver variables (De'Ath & Fabricius, 2000). CART performs a nested hierarchical series of "splits" to a dataset using a single response variable and a suite of driver variables as inputs. CART starts with the entire dataset and breaks the response variable into two groups based on thresholds in the driver variable that minimizes within-group variance. Each resultant group is then split again (the split is referred to as a node) based on whichever driver threshold next minimizes variance. These recursive splits create the inverted "tree" shape familiar to decision trees. CART stops splitting groups apart when the "pruning" criteria is reached. Here, we specified that if an individual split does not improve the performance (here measured using the complexity parameter, otherwise known as proportion reduction of error, PRE) of the CART by at least 3%, the group will not be split further. The final CART results include the optimal decision tree and terminal groups, a measure of response explained (PRE), and a list of "competitor" and "surrogate" options for each split. PRE behaves similarly to the coefficient of determination (R^2) used in linear regression, ranging from 0 to 1 and corresponding to the percent of dataset variance explained. Competitor splits are defined for each node and would divide the data into groups with a similar PRE, but typically create different groups than the optimal split. Surrogate splits are other driver thresholds that would split the group in a similar manner to the optimal split. Analysis of competitor and surrogate splits can provide extended understanding of how the driver variables are related.

We summarized driver and response data at the watershed level using the USGS Watershed Boundary Dataset Hydrologic Unit Code system; specifically at the HUC12 scale

which includes watersheds for small streams with approximately 80 km^2 area (USGS, 2013). Within watersheds, summarizes were compiled using medians for all variables except land use and soil texture, which were reported as areal proportions summing to one. CART analysis was performed for all exceedance thresholds for both Quaternary and Bedrock aquifers, producing eight CART trees. Median probability of threshold exceedance in each watershed was used as the response variable. For sensitivity analysis, mean concentration from all wells in watersheds with over 20 wells was used as a response variable in a separate CART analysis. Driver variables included representatives from all variables in Table 2.1. Detailed descriptions of all 29 variables included can be found in Table A2.1.

5 Results and Discussion

Groundwater nitrate concentrations can first be viewed as a map of well points and concentrations for each aquifer after QA/QC and geocoding (Figure 2.2). Concentration breaks were determined based on physically meaningful thresholds described in Section 4.1. Quaternary aquifer wells are most densely located in the western and north central areas of the state, as well as a strip in the eastern portion, directly outside Detroit (Figure 2.2A-3) that generally corresponds to glacial sediment features greater than 50 meters thick (Figure A1.1). The western portion of the state consistently had areas of elevated nitrate concentration, including in the northwestern region where land cover is primarily forest (Figure 2.1A). The areas of highest concentration samples are north of Grand Rapids (Figure 2.2A-1) and south of Kalamazoo (Figure 2.2A-2), both corresponding with intensive agricultural areas. 26% of Quaternary wells have detectable nitrate (>=0.4 mg/L NO_3-N), with the distribution shown in legend of Figure 2.2A. Bedrock wells are primarily found in the eastern and south-central portions of the state, and only 6% of these wells had detectable concentrations (>=0.4 mg/L, Figure 2.2B legend). Bedrock wells in the most southern central area are associated with higher nitrate concentrations, likely due to their placement in an intensely agricultural region (Figure 2.1A). The remaining analysis focuses on wells in the Quaternary glacial deposit aquifers, due to their wide use for drinking water and closer relationship with surficial nitrogen loading processes (Figure 2.2A).

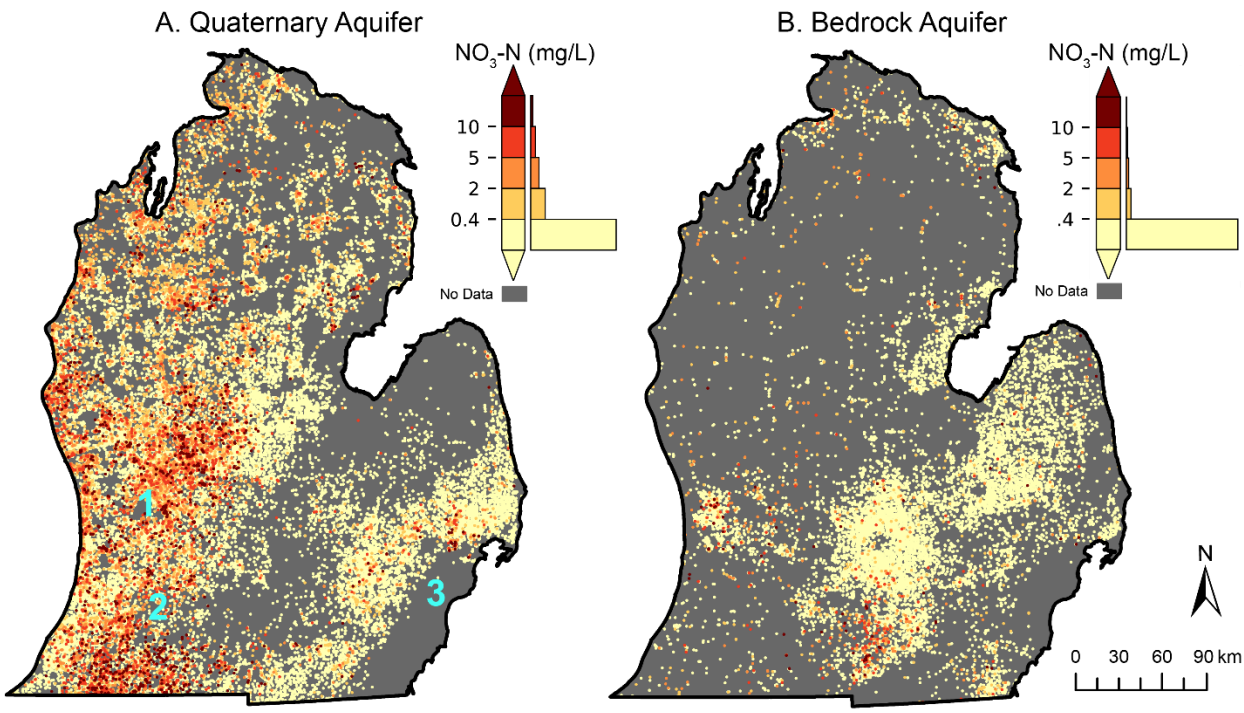

Figure 2.2 Well-level nitrate concentration maps. Locations and concentration values of wells sampled in 2006-2015 in the A) Quaternary Aquifer and B) Bedrock Aquifer. Legends show the distribution within each bin normalized to 100%. Highest concentration point values are plotted on top. Blue numerals in Figure 2.2A denote general locations of major cities: 1) Grand Rapids, 2) Kalamazoo, 3) Detroit.

Kriging provides both a smooth surface of predicted concentrations (Figure 2.3A) and probability layers for exceeding selected thresholds, here >0.4 mg/L (Figure 2.3B) and >2 mg/L NO_3- N (Figure 2.3D). The probability of exceedance can be understood conceptually through the lens of sampling within a watershed. A new sample collected within a given watershed would have an expected probability of exceeding 0.4 mg/L NO_3-N shown in Figure 2.3C. Thus, localized areas within the watershed will be both more and less likely to exhibit nitrate exceedance than the watershed median.

Although the general pattern of higher nitrate values in the western portion of the region remains throughout all maps in Figure 2.3, the nuance of simply having detectable nitrate to having likelihood of health concern can be seen in the variation in pattern in Figure 2.3B and

2.3D. Much of the western half of the LP has median probabilities of exceedance of 0.4 mg/L NO_3-N over 0.5, visible in both continuous (Figure 2.3B) and watershed summarized (Figure 2.3C) views. In contract, probability of exceeding 2 mg/L NO_3-N is less than 0.25 for most of the study region (Figure 2.3D), excluding pockets within the western LP with higher values. These pockets are more easily identified as the medium reds in Figure 2.3E. This distinction is also temporally relevant: although this provides recent nitrate concentrations, areas with detectable nitrate may exceed 2 mg/L if nitrogen continues to be loaded in excess levels to what is taken up by plants.

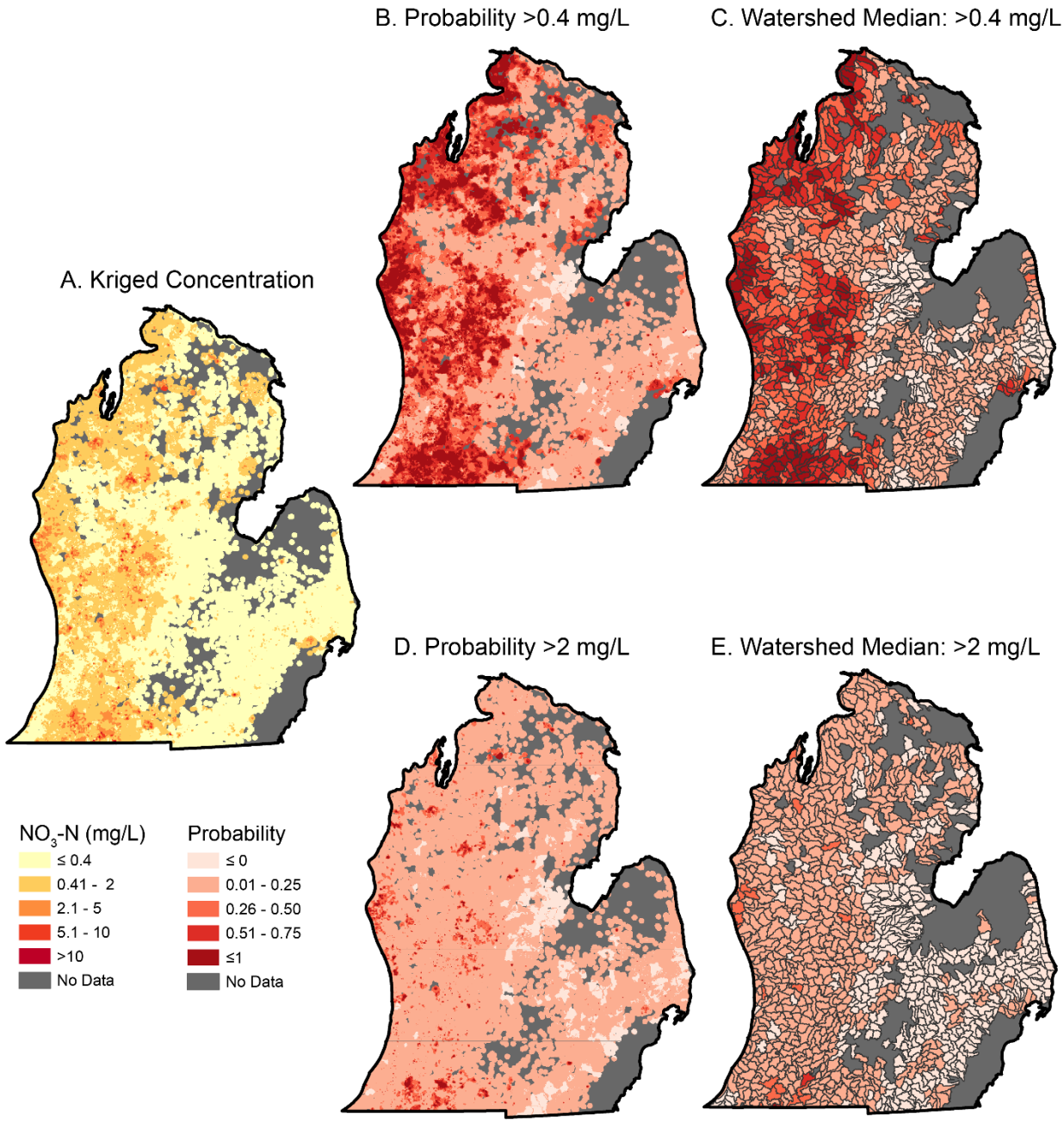

Figure 2.3 Nitrate kriging results. Quaternary aquifer (2006-2015) kriged map, exceedance probabilities, and watershed (HUC12) summarized median exceedance probabilities. Empirical Bayesian Kriging produces a predicted concentration at each cell and probability of exceedance of selected thresholds. Both predictions and probability layers are clipped to areas presumed to have enough data to be confident, here chosen by the correlation length of 3 km buffers around each well. Probability exceedance maps are then summarized to the HUC12 watersheds if there were nitrate concentration predictions for greater than 66% of the area. A) Predicted NO_3-N concentration clipped to areas of confidence. B) Probability of exceeding 0.4 mg/L, or the detection limit. C) Median probability of exceeding 0.4 mg/L at HUC12 watershed level. D)

Figure 2.3 (cont'd) Probability of exceeding 2 mg/L NO_3-N. E) Median probability of exceeding 2 mg/L at HUC12 watershed level.

5.2 Interpreting CART

The benefits of CART for this study are: 1) Identifying the most descriptive variables leading nitrate concentrations and 2) Grouping similarly behaving watersheds. Additional analysis of competitor and surrogate splits can further identify correlations between driver variables and determine breaks in the dataset. Often, spatial trends emerge in these groups due to the inherent spatial nature of geographic analysis – geology and climate influence land use, population density, and nitrogen inputs. Here, we present CART analysis for two cases: Quaternary aquifer exceedance of 0.4 mg/L NO_3-N (Figure 2.4) and Quaternary aquifer exceedance of 2 mg/L NO_3-N (Figure 2.5). Results are discussed by examining each split in the decision tree (Figure 2.4A, Figure 2.5A), the distribution of terminal CART groups (Figure 2.4B, 2.5B), and the geographic distribution of exceedance probabilities within each terminal CART group (Figure 2.4C, Figure 2.5C).

Variables referred to in this section fall into four groups determined by physical understanding and analysis of competitor and surrogate splits: 1) recharge, 2) geologic, 3) land use/land cover (LULC), 4) nitrogen inputs. Recharge is isolated to its own category due to its specific nature as a "combination" variable: it is influenced by precipitation (climate), land use (ability of water to permeate soil), and geology (soil characteristics, unsaturated zone travel time). Geologic variables include soil textures, soil saturated hydraulic conductivity, and aquifer saturated hydraulic conductivity. LULC describe the land use composition of a watershed – e.g., how much is forested, urban, or agricultural. Nitrogen inputs are described by total nitrogen loads to groundwater and loads from septic systems, which are another variable combining

driving factors: nitrogen loading is determined by specific management (not visible solely by land use), geologic conditions, and precipitation driving transport.

5.2.1 Probability of Exceeding 0.4 mg/L

CART for detectable nitrate resulted in six groups of watersheds representing low, medium, and high probability and are described by a combination of recharge, soil, and land use (Figure 2.4A). Low probability, where it would be unlikely to detect nitrate in groundwater, is described by Group X1 (mean probability of exceedance = 0.082) and is explained by recharge < 250 mm/yr and deep soil (100-300 cm) with a sand texture < 65%, paired with a surrogate split of deep soil clay texture > 11.5% (Figure 2.4A). This soil textural split corresponds with sandy loam. Group X1's distribution is centered closely around its mean, suggesting that identifying low recharge areas with soil with less sand than sandy loam will isolate the least nitrate-susceptible areas. In contrast, high probability Groups X5 (mean = 0.51) and X6 (mean = 0.53) are found on both sides of the recharge split and are described by a combination of geologic and land use variables. For Group X5, on the lower recharge side (<250 mm/yr), deep soil is sandy (>65%) and has a high proportion of land in crop agriculture (>40%). Although Group X6 falls on the high recharge side, it similarly is described by a geologic variable (high deep soil K_{Sat}; >3.6e-6 m/s) and a land use variable (lower proportion of forest; <= 57%). Medium risk groups (X2, mean = 0.21; X3, mean = 0.25; X4, mean = 0.3) are found on the "margins" of nitrate-favorable geology and land use – Group X3 has high sand (>65%) but low crops (<40%), while Group X4 has low deep soil KSat (<3.6e-6 m/sec) and Group X5 has high deep soil K_{Sat} (13 mm/hr) and high forest (>57%). Note that for simply detecting nitrate, groups of high and medium risk fall on both sides of the recharge threshold but are differentiated based on a combination of geologic and land use variables.

Mapping the groups (Figure 2.4B) reveals a strong east-west divide, where almost the entire western half of the LP is at medium or high probability of detectable nitrate. There is not a strong north-south divide in high probability Groups X5 and X6, which can be found in any latitude in the western LP. Lacking this latitudinal divide drives some of the variability within Groups X5 and X6 (Figure 2.4B) – land use in northwestern LP has more forest, smaller towns, and lacks intensive corn-soy agriculture found in much of the southwest LP.

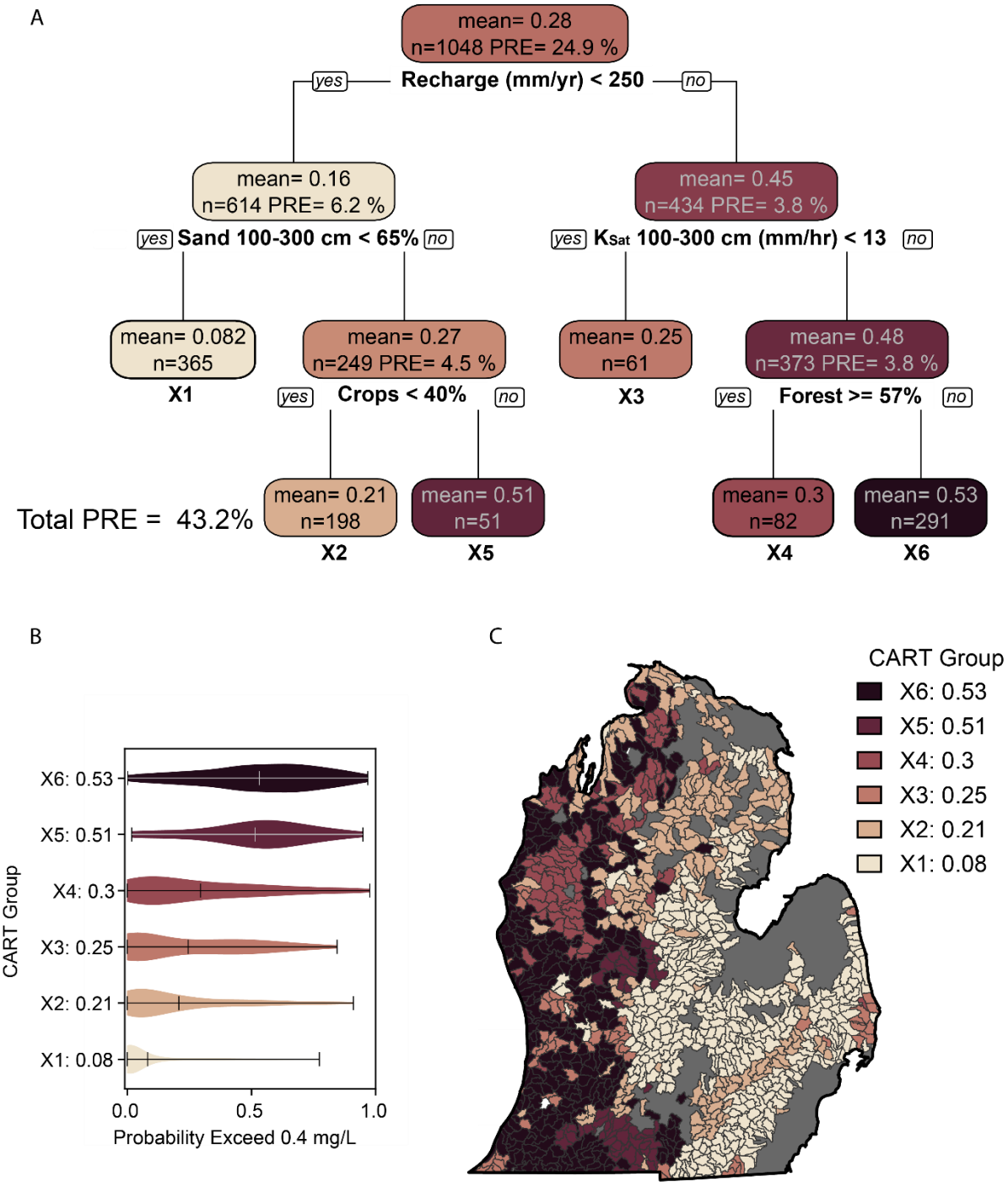

Figure 2.4 CART results for >0.4 mg/L NO_3-N. Results from CART analysis for median probability of exceeding 0.4 mg/L, or nitrate detection limit within a HUC12 watershed. A) Decision tree identifying each split within the data. Each node contains the group mean probability of exceedance, the total watersheds in the group (*n*), and the proportion response explained (PRE), a metric that describes the performance of the split relative to the entire dataset. Nodes are colored based on their mean probability, with higher probabilities in darker colors.

Figure 2.4 (cont'd) The final CART groups are numbered from low to high concentration. B) The final CART groups are mapped by HUC12. Colors correspond and groups are consistent with those in A, with higher probabilities in darker colors. C) Violin plots for each final group. Violin plots display the range and mean as a line with a probability density function fit to show the "shape" of the distribution – thicker areas contain more samples.

5.2.2 Probability of Exceeding 2 mg/L

The basic tenets of a combination of recharge, soil, and land use remains intact for identifying the probability of exceeding human health risk thresholds, 2 mg/L NO_3-N, in watersheds. Here (Figure 2.5A), recharge again creates a significant break in the data: 624 watersheds are isolated based solely on having low aquifer recharge at an almost identical threshold as detectable NO_3-N (0.4 mg/L: <250 mm/yr; 2 mg/L: <251 mm/yr). The remaining watersheds are split first based on a relatively rare occurrence which includes the 31 watersheds with almost no mixed forest (< 0.3%), with surrogate splits highlighting other non-agricultural or urban land cover. These outliers are then divided into the highest and lowest groups (Y1, mean = 0.014 and Y7, mean = 0.27) based on deep soil K_{Sat} (threshold: 15 mm/hr). Moving to the left of Mixed Forest, to watersheds that have mixed forest, the groups are split by a series of land use variables (Crops < 34%, Grassland < 3.82%) and finally a soil variable (deep soil clay >= 6%). Here, the story is less clear and the groups of medium risk (Y5) and high risk (Y6) are both small and have wide distributions (Figure 2.5B), with only 17 and 24 watersheds, respectively. Generally, those watersheds with higher probabilities of exceedance have more agricultural land, and if soil is introduced there is extremely low clay (<5.5%) with a surrogate split of high sand content (>90.5%). Spatially, a different pattern emerges – much of the northwestern LP is no longer in a medium or high probability category and pockets of such in the southwestern LP become more isolated. It is important to note the relative level of difficulty in explaining the

probability of exceeding 2 mg/L because the overall probabilities at the watershed level are much lower (Figure 2.3E) and few have high values.

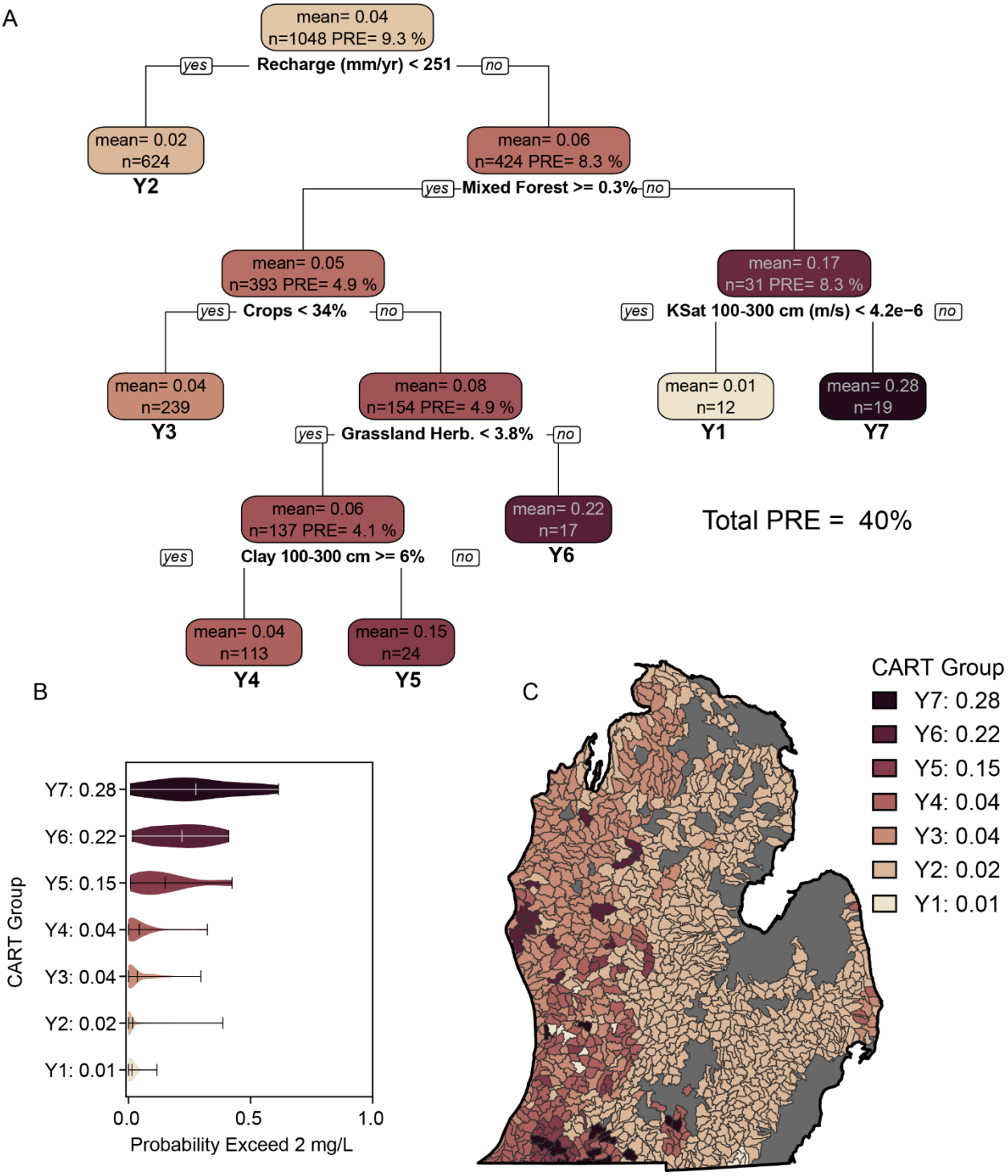

Figure 2.5 CART results for >2 mg/L NO_3-N. Results from CART analysis for median probability of exceeding 2 mg/L, or detecting nitrate within a HUC12 watershed. A) Decision tree identifying each split within the data. Each node contains the group mean probability of

Figure 2.5 (cont'd) exceedance, the total watersheds in the group (*n*), and the proportion response explained (PRE), a metric that describes the performance of the split relative to the entire dataset. Nodes are colored based on their mean probability, with higher probabilities in darker colors. The final CART groups are numbered from low to high concentration. B) The final CART groups are mapped by HUC12. Colors correspond and groups are consistent with those in A, with higher probabilities in darker colors. C) Violin plots for each final group. Violin plots display the range and mean as a line with a probability density function fit to show the "shape" of the distribution – thicker areas contain more samples.

5.2.3 Sensitivity

We performed CART analysis on a response variable based solely on point data to confirm that kriging methods produce comparable results. This CART reasonably predicts watershed-averaged well point concentrations in Quaternary aquifers (Figure A2.5). This response variable includes spatial biases in sampling that we believe were reduced by kriging. Regardless, driver variables had similar splitting thresholds. Recharge isolated the lowest concentration watersheds at a threshold of 242 mm/year. The group of outliers identified by having no mixed forest in the 2 mg/L CART were also identified and further split by a similar soil K_{Sat} value (12 mm/hr in soil depth 0-20 cm). Forest proportion of 47% and crop proportion of 50% appeared and split logically (ie., higher crop proportion was associated with higher nitrate concentrations). Aquifer K_{Sat}, which was not a top performing driver in the exceedance CARTs, appeared and sensibly split away a low concentration group based on low conductivity. The CART also performed similarly to the exceedance probability CARTs (PRE=40.3% compared to 0.4 mg/L: 43.2%, 2 mg/L: 40%). The similarity of these results, along with the benefits of the kriging, suggest that our primary analyses are both robust and preferred to this similar watershed point averaging for our dataset.

5.3 Understanding Drivers of Groundwater Nitrate Concentration

5.3.1 Recharge

Aquifer recharge was consistently the highest performing initial split at a threshold of 242-251 mm/yr. This threshold creates a visible split in the LP (Figure 2.6A) consistent with the general pattern of the presence of groundwater nitrate (Figure 2.6B). There were no comparable surrogate splits; however, soil texture variables were almost as effective as competitors. The descriptive ability of recharge in LP groundwater nitrate may be due to its multifaceted dimensions as a transport variable – combining precipitation and geology. Nitrate would be unexpected in an aquifer underlying soils that are not sufficiently permeable; not only must nitrogen be applied, regardless of land use, but the resulting nitrate must be mobile within the unsaturated and saturated zones. However, recharge never appeared outside of the initial split in any CARTs, nor were there any equally performing variable in surrogate or competitor analysis. Recharge is a modeled variable and not generally readily available for groundwater nitrate analysis elsewhere, suggesting the importance of modeling and the need for better quantified recharge maps. After identifying high recharge areas, the nuance in groups was described by additional geologic and land use variables.

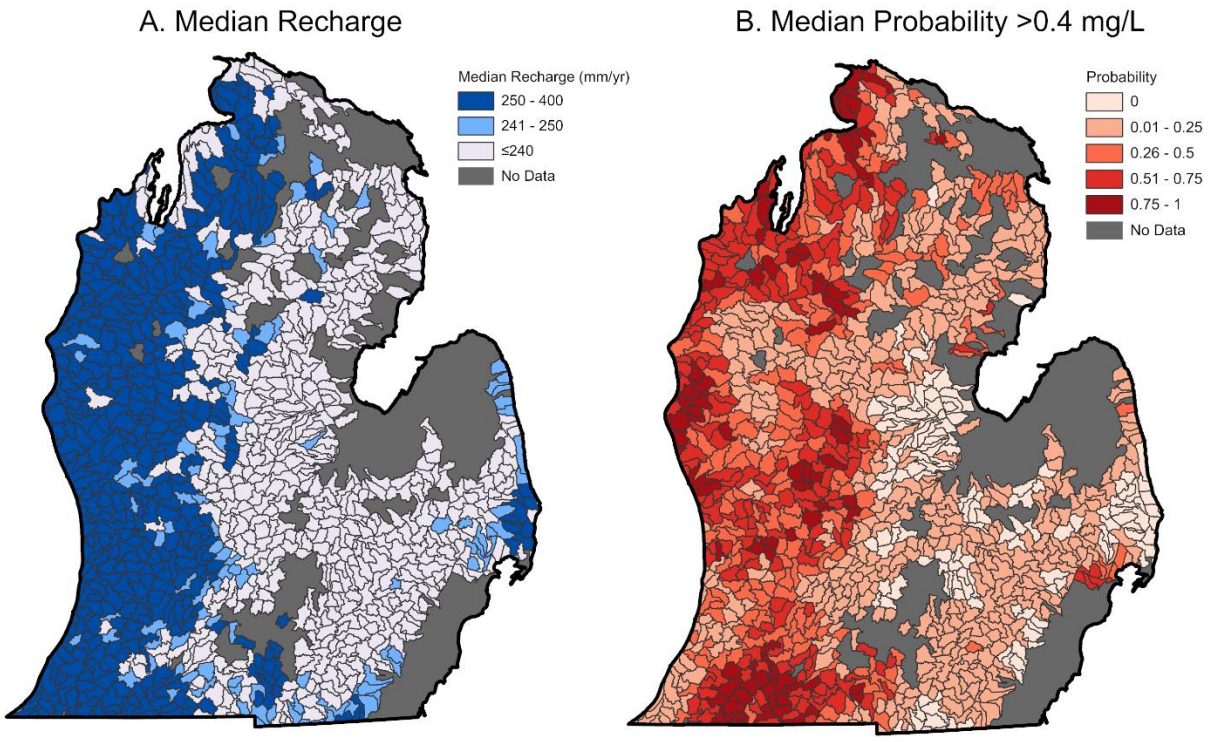

Figure 2.6 Aquifer recharge compared to probability of exceeding 0.4 mg/L NO_3-N. A) HUC12 summarized median aquifer recharge colored by CART significant splits. >250 mm/yr was the consistent split for Quaternary aquifers. 241-250 mm/yr represents the additional watersheds included at the lower Bedrock CART recharge split. B) HUC12 summarized median probability of exceeding 0.4 mg/L, replicated from Figure 2.3C. Side by side, the figures show the broad consistent pattern of high recharge and elevated detectable nitrate probabilities.

5.3.2 Vulnerable Geology and Hazardous Land Use

Following recharge (a transport-related variables) comes the factors of geology and land use. Agricultural land and high nitrogen inputs are found across the LP, but detecting groundwater nitrate is improbable in the eastern parts of the state despite extensive row- and field- crop agriculture. In contrast, even heavily forested areas in northwestern LP, such as the Manistee National Forest, fall into medium detection probability Group X4 (Figure 2.4C). Surrogate splits often group land use variables when they appear in CART and similarly group soil and geology. Although this analysis does not show the full distributions of each variable within a group, future work will bring together the broader “definitions” of exceedance probability groups. As it stands, simply having non-vulnerable geology will usually isolate the

lowest probability classes. Groups with vulnerable geology combined with hazardous land use result in the highest probabilities.

The breaks in variables identified in CART analysis allow the description of what constitutes “vulnerable” geology and “hazardous” land use. Vulnerable geology in the LP consists of sandy (>65%), low clay (<6%) soils at 100-300 cm depth with K_{Sat} greater than ~14 mm/hr. Soil texture and hydraulic conductivity are connected variables, with coarser soils having increased conductivity. Aquifer K_{Sat} did not appear as top CART selection, but surrogate splits showed a level of 11-13 m/day was associated with soil variables. Hazardous land use includes watersheds with > 30-40% crop land use and minimal natural land use. Although forest and agriculture both appeared as deciding land uses, they did not act in binary, meaning that a split at 40% agriculture did not correspond to a surrogate split of 60% forest due to the complexity of other urban and wetland land use.

Notably, nitrogen inputs to groundwater or inputs from septic did not appear in CART analysis for >0.4 mg/L or >2 mg/L. Occasionally, these variables acted as surrogates with land use variables (TN and crops; septic and urban), but were not as effective at improving the model performance and so were not selected as the split. Nitrogen inputs to groundwater, like recharge, is a combination variable affected by land use and management, recharge, and geologic conditions. The lack of nitrogen loading as a dominant variable in this analysis may be due to modeling parameters or the lack of distinction between nitrogen sources. TN, without chemical speciation, may not be effective to describe different behavior of nitrogen sources in variable subsurface environments.

6 Conclusion

This analysis provides new views and methods to describe drivers of groundwater nitrate concentrations in Michigan's Lower Peninsula using an extensive well dataset, kriging, and CART. Sources of uncertainty include well nitrate measurements, kriging methods, modeled inputs, and variables not included in the current analysis. These sources of uncertainty also provide avenues of future work and improvements to this study.

While there are significant benefits of using a spatially and temporally extensive dataset, the breadth introduces limitations and uncertainty. Many practitioners retrieved, analyzed, and digitized samples across time and space. Water table elevation, recent precipitation conditions, and season of sample were not included in this analysis, meaning spatially proximal samples may not reflect the same physical conditions. Though, it is likely that this variation is dampened by the presence of samples from a wide range of environments. Additionally, we selected a ten-year period of samples to increase our spatial coverage and sampling density to improve kriging. Despite the variable conditions and uncertainty in such a large dataset, we see benefit in including more data and allowing kriging to account for uncertainty between nearby data points.

This analysis did not include some variables found to be significant in other studies, notably well depth and a redox condition variable. Future work will include sulfate concentration in the watershed from the well chemistry database as a proxy for redox conditions, as dissolved oxygen is unavailable in this dataset. Well depth, screening interval, and static water level were not included due to insufficient data within the Wellogic database and difficulty summarizing this data at the watershed level. Although wells were separated by aquifer type, it is expected that shallower wells (with shallow screening interval) are more susceptible to elevated nitrate concentration.

This extensive dataset from state-collected samples allows for a statistically powerful analysis over a wide range of land use and geologic conditions. Future work may include increased driver variables, alternative tree-based machine learning methods, extended analysis of the distributions of other driver variables in each CART group, and an application of quantifying nitrate exposure hotspots based on the relationship between elevated nitrate concentration wells, population density, public supply well service, and continued nitrogen loading areas. This increased understanding of how vulnerable geology is affected by a variety of hazardous land uses can be combined with land use legacies and future projections to identify areas that are of concern or worth protecting.

This study utilizes an extensive database of drinking water nitrate well measurements across Michigan's Lower Peninsula in tandem with a suite of environmental variables to identify drivers of groundwater nitrate concentration using CART analysis. Aquifer recharge was isolated as an overarching "first cut" variable that can quickly eliminate many watersheds from concern of nitrate concentration. Watersheds with sufficient recharge were separated based on a combination of vulnerable geology and hazardous land use. Our analysis identified sandy, high hydraulic conductivity soils as "vulnerable" geologic conditions and heavy agricultural activity as "hazardous" land use. However, even more-forested, sparsely-populated watersheds were found to have medium likelihood of detectable nitrate concentrations if they were in high recharge, geologically vulnerable areas. These findings can be used to identify areas of concern for elevated nitrate concentration and drive modeling efforts using land use legacies, future land use projections, and process-based modeling to aid in management strategy.

Acknowledgements

This chapter was coauthored by Anthony Kendall, Sherry Martin, and David Hyndman. We thank Yvonne Krimmel and Bailey Hannah for assistance in processing data and providing feedback on the manuscript. Data was provided from the Michigan Department of Environment, Great Lakes, and Energy from FOIAA 5371-19. Funding was provided by NOAA Grant NA12OAR4320071 and NASA Grant NNX11AC72G.

APPENDIX

Text A2.1 Methods: Geocoding

Well chemistry data was not provided with latitude-longitude coordinates for use in spatial analysis. However, addresses for each sample were provided. Because wells had been sampled and data recorded by many people over a long time span, there were significant inconsistencies in how addresses were entered. Wells were also frequently given a street address of "Well #1" or "Pumphouse" and only identifiable by an owner. Before we were able to use GIS geocoding software to convert address to coordinates, we filtered samples based on addresses and eliminated wells without street addresses. Although private wells were usually only sampled once, public wells were often sampled multiple times with inconsistently entered addresses. OpenRefine (https://openrefine.org/), a software for filtering and parsing text in messy datasets, was used to create unique address IDs for samples that had similar addresses so that only one spatial point would be created per well. Due to the size of the dataset and our limitations in human resources, we were not able to utilize all samples or manually find addresses that were not easily identifiable. Geocoding was performed in ArcGIS Pro 2.2 using the ArcGIS World Geocoding Service. This geocoding algorithm includes a score which demonstrates how confident it is that the address was correctly matched. We selected a minimum score of 95/100 based on a qualitative analysis of sampled addresses with lower scores. Lower scores were often missing a street address (placed at a town center) or were moved to different street numbers than the input address had.

Text A2.2 Methods: Wellogic Join

To classify samples aquifer source, data from the well properties database, Wellogic, was needed. Public wells are identified by a WSSN code for their owner; however, many owners have multiple wells. Public well owners with a single well were joined to the chemistry database based on WSSN. All other wells within the database were joined based on a spatial join operation performed in ArcGIS Pro 2.5. Wells in the chemistry dataset were joined to the nearest Wellogic well within 1 km. If there was not a well within 1 km, the wells could not be joined. addresses were available in the Wellogic databases, differences in address nomenclature and format required an unrealistic amount of pre-processing work to join wells based on address.

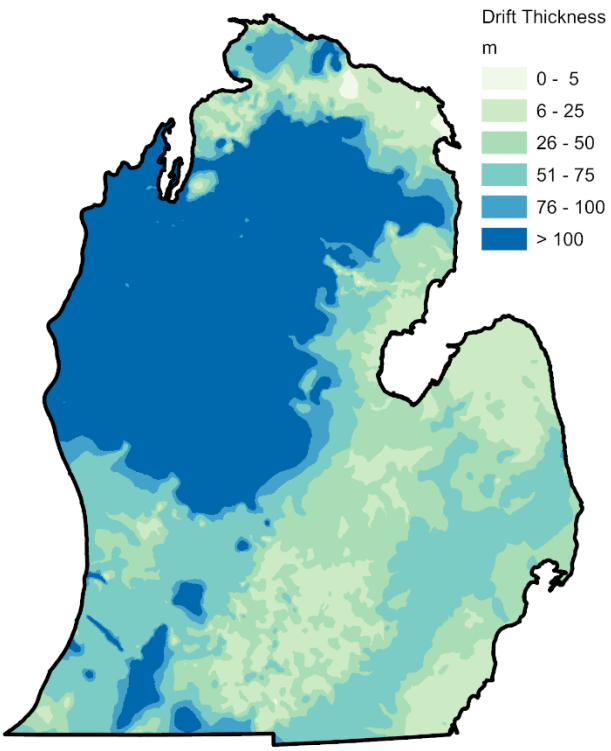

Figure A2.1 Map of thickness of Quaternary glacial sediment. From Soller & Garrity (2018).

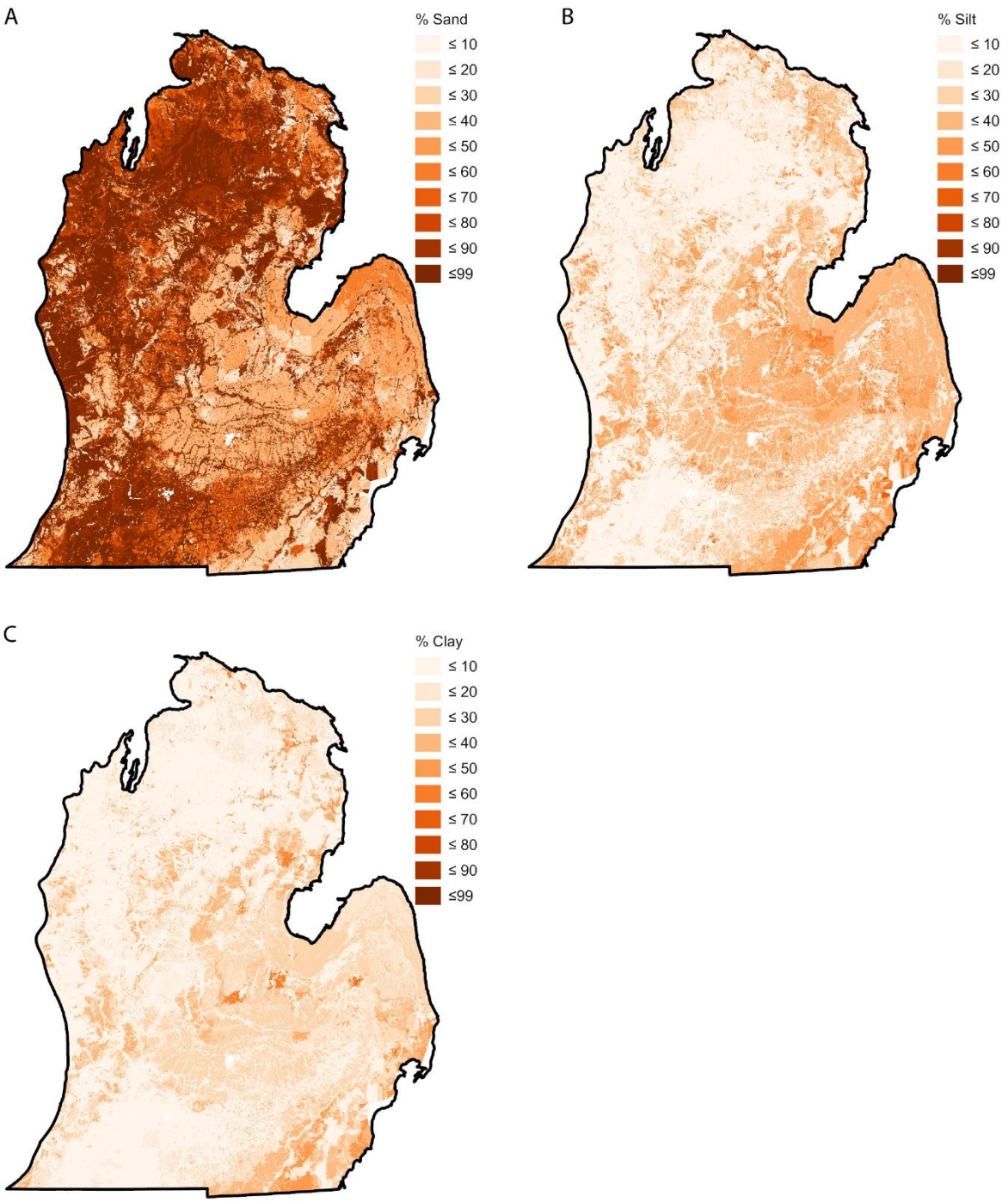

Figure A2.2 Maps of soil textures for soil depths 100-300 cm. Fom NRCS (2017) A) % Sand, B) % Silt, C) %Clay

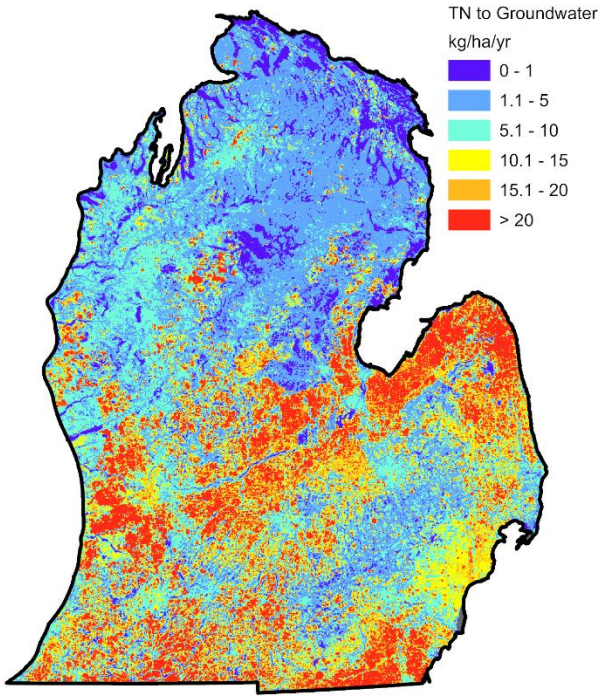

Figure A2.3 Map of nitrogen loads to groundwater. TN loads to groundwater. (Hamlin et al., 2020; Wan et al., In Prep)

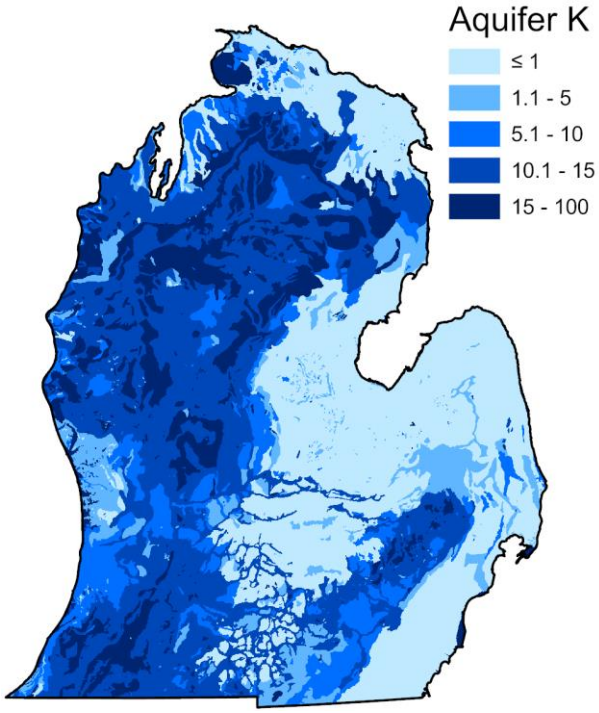

Figure A2.4 Map of aquifer saturated conductivity. Calculated using Wellogic and Quaternary geology (Farrand & Bell, 1982)

Table A2.1 Complete set of variables used within CART. Variables are grouped by variable category.

Category	Variable	Definition	Unit
Nitrogen	TN_GW_kghayr_median	Total Nitrogen Load to Groundwater, HUC12 median	kg/ha/yr
	TN_Sep_kghayr_HUC12_md	Total Nitrogen Load from Septic Tanks, HUC12 median	kg/ha/yr
Aquifer	aq_ksat_mday_median	Aquifer Ksat, HUC12 Median	m/day
	recharge_mmyr_median	Aquifer Recharge, HUC12 median	mm/yr
	WellProp_Drift	Proportion of HUC12 Wells in Quaternary Aquifer	-
Soil	sandtotal_r_lay1_median	Soil 0-20 cm % Sand, HUC12 median	%
	sandtotal_r_lay4_median	Soil 100-300 cm % Sand, HUC12 median	%
	claytotal_r_lay1_median	Soil 0-20 cm % Clay, HUC12 median	%
	claytotal_r_lay4_median	Soil 100-300 cm % Clay, HUC12 median	%
	ksat_soil_lay1_median	Soil 0-20 cm Ksat, HUC12 median	mm/hr
	ksat_soil_lay4_median	Soil 100-300 cm Ksat, HUC12 median	mm/hr
LULC	OpenWater_per	Open Water %, HUC12	%
	DevOpen_per	Developed - Open %, HUC12	%
	DevLow_per	Developed - Low Intensity %, HUC12	%
	DevMed_per	Developed - Medium Intensity %, HUC12	%
	DevHigh_per	Developed - High Intensity %, HUC12	%
	Barren_per	Barren %, HUC12	%
	DecidFor_per	Forest - Deciduous %, HUC12	%
	EvergreenFor_per	Forest - Evergreen %, HUC12	%
	MixedFor_per	Forest - Mixed %, HUC12	%
	ShrubScrub_per	Shrub/Scrub %, HUC12	%
	GrasslandHerb_per	Grassland Herbaceous %, HUC12	%
	PastureHay_per	Pasture/Hay %, HUC12	%
	Crops_per	Cultivated Crop %, HUC12	%
	WoodyWetlands_per	Woody Wetlands %, HUC12	%
	HerbWetlands_per	Herbaceous Wetlands %, HUC12	%
	forest_per	Aggregated Forest %, HUC12	%
	urban_per	Aggregated Urban %, HUC12	%
	wetland_per	Aggregated Wetland %, HUC12	%

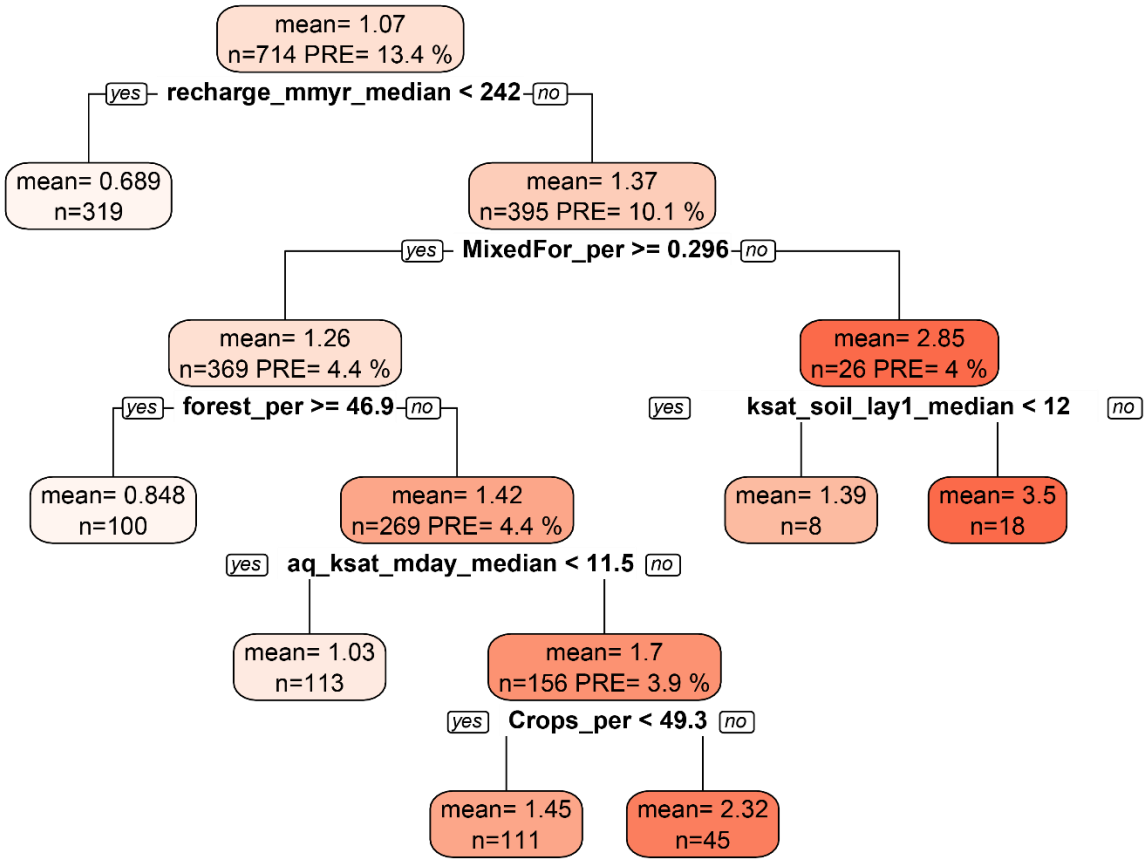

Figure A2.5 CART decision tree for sensitivity analysis predicting mean NO_3-N concentration. Watersheds included must have >= 20 wells in Quaternary aquifer. Variable names can be referenced in Table A2.1.

REFERENCES

REFERENCES

Alexander, R. B., Smith, R. A., & Schwarz, G. E. (2004). Estimates of diffuse phosphorus sources in surface waters of the United States using a spatially referenced watershed model. *Water Science and Technology*, *49*(3), 1–10.

Anderson, D. M., Gilbert, P. M., & Burkholder, J. M. (2002). Harmful Algal Blooms and Eutrophication Nutrient Sources, Composition, and Consequences. *Estuaries*, *25*(4b), 704–726. https://doi.org/10.1007/BF02804901

Anderson, K. A., & Downing, J. A. (2006). Dry and wet atmospheric deposition of nitrogen, phosphorus and silicon in an agricultural region. *Water, Air, and Soil Pollution*, *176*(1–4), 351–374. https://doi.org/10.1007/s11270-006-9172-4

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area Hydrologic Modeling and Assessment Part I: Model development. *Journal of the American Water Resources Association*, *34*(February), 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Bailey, T. C., & Gatrell, A. C. (1996). Interactive spatial data analysis. *Computers & Geosciences*, 22(8), 953-954. https://doi.org/10.1016/s0098-3004(96)80468-7

Barry, D. A. J., Goorahoo, D., & Goss, M. J. (1993). Estimation of Nitrate Concentrations in Groundwater Using a Whole Farm Nitrogen Budget. *Journal of Environmental Quality*, *22*(4), 767–775. https://doi.org/10.2134/jeq1993.00472425002200040019x

Barzegar, R., Moghaddam, A. A., Deo, R., Fijani, E., & Tziritis, E. (2018). Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. *Science of the Total Environment*, *621*, 697–712. https://doi.org/10.1016/j.scitotenv.2017.11.185

Basso, B., & Ritchie, J. T. (2005). Impact of compost, manure and inorganic fertilizer on nitrate leaching and yield for a 6-year maize-alfalfa rotation in Michigan. *Agriculture, Ecosystems and Environment*, *108*(4), 329–341. https://doi.org/10.1016/j.agee.2005.01.011

Beal, C. D., Gardner, E. A., & Menzies, N. W. (2005). Process, performance, and pollution potential: A review of septic tank-soil absorption systems. *Australian Journal of Soil Research*, *43*(7), 781–802. https://doi.org/10.1071/SR05018

Benettin, P., Fovet, O., van der Velde, Y., Ruiz, L., Hrachowitz, M., Howden, N. J. K., … Wade, A. J. (2016). Transit times-the link between hydrology and water quality at the catchment scale. *Wiley Interdisciplinary Reviews: Water*, *3*(5), 629–657. https://doi.org/10.1002/wat2.1155

Boyer, E. W., Goodale, C. L., Jaworski, N. A., & Howarth, R. W. (2002). Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern U.S.A. *Biogeochemistry*, *57*(1), 137–169. https://doi.org/10.1023/A:1015709302073

Brakebill, J., & Gronberg, J. (2017). County-Level Estimates of Nitrogen And Phosphorus from Commercial Fertilizer for the Conterminous United States, 1987-2012. https://doi.org/doi.org/10.5066/F7H41PKX

Brieman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. New York: Routledge. https://doi.org/10.1201/9781315139470

Brinson, M. M., Bradshaw, H. D., Holmes, R. N., & Elkins, J. B. (1980). Litterfall, Stemflow, and Throughfall Nutrient Fluxes in an Alluvial Swamp Forest. *Ecology*, *61*(4), 827–835. https://doi.org/10.2307/1936753

Brooks, B. W., Lazorchak, J. M., Howard, M. D. A., Johnson, M. V. V., Morton, S. L., Perkins, D. A. K., … Steevens, J. A. (2016). Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environmental Toxicology and Chemistry*, *35*(1), 6–13. https://doi.org/10.1002/etc.3220

Brown, L. J., Taleban, V., Gharabaghi, B., & Weiss, L. (2011). Seasonal and spatial distribution patterns of atmospheric phosphorus deposition to Lake Simcoe, ON. *Journal of Great Lakes Research*, *37*(SUPPL. 3), 15–25. https://doi.org/10.1016/j.jglr.2010.09.008

Burow, K. R., Nolan, B. T., Rupert, M. G., & Dubrovsky, N. M. (2010). Nitrate in groundwater of the United States, 1991-2003. *Environmental Science and Technology*, *44*(13), 4988–4997. https://doi.org/10.1021/es100546y

Bytnerowicz, A., Johnson, R. F., Zhang, L., Jenerette, G. D., Fenn, M. E., Schilling, S. L., & Gonzalez-Fernandez, I. (2015). An empirical inferential method of estimating nitrogen deposition to Mediterranean-type ecosystems: The San Bernardino mountains case study. *Environmental Pollution*, *203*(2), 69–88. https://doi.org/10.1016/j.envpol.2015.03.028

Canter, L. W., & Canter, L. W. (2019). Influence of Subsurface Processes. In *Nitrates in Groundwater*. https://doi.org/10.1201/9780203745793-2

Cao, P., Lu, C., & Yu, Z. (2018). Historical nitrogen fertilizer use in agricultural ecosystems of the contiguous United States during 1850-2015: Application rate, timing, and fertilizer types. *Earth System Science Data*, *10*(2), 969–984. https://doi.org/10.5194/essd-10-969-2018

Carmichael, W. W., Azevedo, S. M. F. O., An, J. S., Molica, R. J. R., Jochimsen, E. M., Lau, S., … Eaglesham, G. K. (2001). Human fatalities form cyanobacteria: Chemical and biological evidence for cyanotoxins. *Environmental Health Perspectives*, *109*(7), 663–668. https://doi.org/10.1289/ehp.01109663

Charles, A., Rochette, P., Whalen, J. K., Angers, D. A., Chantigny, M. H., & Bertrand, N. (2017). Global nitrous oxide emission factors from agricultural soils after addition of organic amendments: A meta-analysis. *Agriculture, Ecosystems and Environment*, *236*(3), 88–98. https://doi.org/10.1016/j.agee.2016.11.021

Cleveland, C. C., Houlton, B. Z., Smith, W. K., Marklein, A. R., Reed, S. C., Parton, W., … Running, S. W. (2013). Patterns of new versus recycled primary production in the terrestrial biosphere. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(31), 12733–12737. https://doi.org/10.1073/pnas.1302768110

Courault, D., Demarez, V., Guérif, M., Le Page, M., Simonneaux, V., Ferrant, S., & Veloso, A. (2016). Contribution of Remote Sensing for Crop and Water Monitoring. In N. Baghdadi & M. Zribi (Eds.), *Land Surface Remote Sensing in Agriculture and Forest* (pp. 113–177). Elsevier. https://doi.org/10.1016/B978-1-78548-103-1.50004-2

David, M. B., Drinkwater, L. E., & Mcisaac, G. F. (2010). Sources of Nitrate Yields in the Mississippi River Basin. *Journal of Environmental Quality*, *39*(5), 1657–1667. https://doi.org/10.2134/jeq2010.0115

David, M. B., & Gentry, L. E. (2000). Anthropogenic Inputs of Nitrogen and Phosphorus and Riverine Export for Illinois, USA. *Journal of Environmental Quality*, *29*(2), 494–508. https://doi.org/10.2134/jeq2000.00472425002900020018x

David, M. B., Gentry, L. E., Kovacic, D. A., & Smith, K. M. (1997). Nitrogen balance in and export from an agricultural watershed. *Journal of Environmental Quality*, *26*(4), 1038–1048. https://doi.org/10.2134/jeq1997.00472425002600040015x

De'Ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, *81*(11), 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Delumyea, R. G., & Petel, R. L. (1978). Wet and dry deposition of phosphorus into Lake Huron. *Water, Air, and Soil Pollution*, *10*(2), 187–198. https://doi.org/10.1007/BF00464714

Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., … Thornbrugh, D. J. (2009). Policy Analysis: Eutrophication of U . S . Freshwaters : Damages. *Environmental Science & Technology*, *43*(1), 12–19. https://doi.org/10.1021/es801217q

Downing, J. A., & Mccauley, E. (1992). The nitrogen phosphorous relationship in lakes. *Limnology and Oceanography*, *37*(37), 936–945.

Eimers, M. C., Watmough, S. A., Paterson, A. M., Dillon, P. J., & Yao, H. (2009). Long-term declines in phosphorus export from forested catchments in south-central Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(10), 1682–1692. https://doi.org/10.1139/F09-101

Eisenreich, S. J., Emmling, P. J., & Beeton, A. M. (1977). Atmospheric Loading of Phosphorus and Other Chemicals to Lake Michigan. *Journal of Great Lakes Research*, *3*(3–4), 291–304. https://doi.org/10.1016/S0380-1330(77)72261-0

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Environment and Climate Change Canada (2016). Great Lakes Precipitation Network. https://doi.org/10.17616/R3J349

ESRI. (2020). ArcGIS Pro Version 2.5.

ESRI. (2020). What is Empirical Bayesian kriging? Retrieved March 5, 2020, from https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm

Farrand, W., & Bell, D. (1982). 1982 Quaternary Geology. Dept. of Geological Sciences, Univeristy of Michigan. Geologic Survey Division, MDEQ. Division Geographic Information Services Unit, Resource Mapping and Aerial Photography, MDNR. Retrieved from http://www.dnr.state.mi.us/spatialdatalibrary/metadata/quaternary_geology.htm

Fitzsimmons, E. G. (2014, August 3). Tap Water Ban for Toledo Residents. *The New York Times*.

Foster, N. W. (1974). Annual Macroelement Transfer From Pinus Banksiana Lamb. Forest to Soil. *Canadian Journal of Forest Research*, *4*(4), 470–476. https://doi.org/10.1139/x74-069

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*, *28*(2), 337–407. https://doi.org/10.1214/aos/1016218223

GCSAA. (2009). Golf course environment profile: volume III summary nutrient use and management on U.S. golf courses.

Gentry, L. E., David, M. B., Smith, K. M., & Kovacic, D. A. (1998). Nitrogen cycling and tile drainage nitrate loss in a corn/soybean watershed. *Agriculture, Ecosystems and Environment*, *68*(1–2), 85–97. https://doi.org/10.1016/S0167-8809(97)00139-4

Goolsby, D. A., Battaglin, W. A., Lawrence, G. B., Artz, R. S., Aulenbach, B. T., Hooper, R. P., … Stensland, G. J. (1999). Flux and Sources of Nutrients in the Mississippi-Atchafalaya River Basin: Topic 3 Report for the Integrated Assessment of Hypoxia in the Gulf of Mexico. Silver Spring, MD: NOAA Coastal Ocean Program.

Goyette, J., Bennett, E. M., Howarth, R. W., & Maranger, R. (2016). Changes in anthropogenic nirogen and phosphorus inputs to the St. Lawrence sub-basin over 110 years and impacts on

riverine export. *Global Biogeochemical Cycles*, 1–15. https://doi.org/10.1002/2016GB005384.Received

Gronberg, J. A. M., & Spahr, N. E. (2012). *County-Level Estimates of Nitrogen and Phosphorus from Commercial Fertilizer for the Conterminous United States, 1987-2006.* Reston, VA. Retrieved from https://pubs.usgs.gov/sir/2012/5207/pdf/sir20125207.pdf

Hamlin, Q. F., Kendall, A. D., Martin, S. L., Whitenack, H. D., Roush, J. A., Hannah, B. A., & Hyndman, D. W. (2020). Quantifying Landscape Nutrient Inputs With Spatially Explicit Nutrient Source Estimate Maps. *Journal of Geophysical Research: Biogeosciences*, *125*(2), 1–24. https://doi.org/10.1029/2019JG005134

Hamlin, Q. F., Kendall, A. D., Martin, S. L., Whitenack, H. D., Roush, J. A., Hannah, B. A., & Hyndman, D. W. (2020). SENSEmap-USGLB: Nitrogen and Phosphorus Inputs. *Hydroshare*. https://doi.org/https://doi.org/10.4211/hs.1a116e5460e24177999c7bd6f8292421

Han, H., & Allan, J. D. (2008). Estimation of nitrogen inputs to catchments: Comparison of methods and consequences for riverine export prediction. *Biogeochemistry*, *91*(2–3), 177–199. https://doi.org/10.1007/s10533-008-9279-3

Hargan, K. E., Paterson, A. M., & Dillon, P. J. (2011). A total phosphorus budget for the Lake of the Woods and the Rainy River catchment. *Journal of Great Lakes Research*. https://doi.org/10.1016/j.jglr.2011.09.001

Holman, I. P., Whelan, M. J., Howden, N. J. K., Bellamy, P. H., Willby, N. J., Rivas-Casado, M., & McConvey, P. (2008). Phosphorus in groundwater - an overlooked contributor to eutrophication? *Hydrological Processes*, *22*, 5121–5127. https://doi.org/10.1002/hyp.7198

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., … Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, *81*(5), 345–354. https://doi.org/10.14358/PERS.81.5.345

Hong, B., Swaney, D. P., & Howarth, R. W. (2013). Estimating net anthropogenic nitrogen inputs to U.S. watersheds: Comparison of methodologies. *Environmental Science and Technology*. https://doi.org/10.1021/es303437c

Howarth, R. W., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., … Zhao-Liang, Z. (1996). Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences. *Biogeochemistry*, *35*, 75–139.

Howarth, R. W. (2008). Coastal nitrogen pollution: A review of sources and trends globally and regionally. *Harmful Algae*, *8*(1), 14–20. https://doi.org/10.1016/j.hal.2008.08.015

Howarth, R., Swaney, D., Billen, G., Garnier, J., Hong, B., Humborg, C., … Marino, R. (2012). Nitrogen fluxes from the landscape are controlled by net anthropogenic nitrogen inputs and by climate. *Frontiers in Ecology and the Environment*. https://doi.org/10.1890/100178

Hussain, I., Pilz, J., & Spoeck, G. (2010). Hierarchical Bayesian space-time interpolation versus spatio-temporal BME approach. *Advances in Geosciences, 25*(March), 97–102. https://doi.org/10.5194/adgeo-25-97-2010

Hyndman, D. W., Kendall, A. D., & Welty, N. R. H. (2007). Evaluating temporal and spatial variations in recharge and streamflow using the integrated landscape hydrology model (ILHM). *Geophysical Monograph Series*, *171*, 121–141. https://doi.org/10.1029/171GM11

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Kaiser, D. E., Lamb, J. A., & Eliason, R. (2011). Fertilizer guidelines for agronomic crops in Minnesota. *University of Minnesota Extension*, 1–44. Retrieved from www.extension.umn.edu/nutrient-management/%0ABU-06240-S

Keeler, B. L., Isbell, F., Hill, J. D., Tessum, C. W., Marshall, J. D., Gourevitch, J. D., & Polasky, S. (2016). The social costs of nitrogen. *Science Advances*, (October). https://doi.org/10.1126/sciadv.1600219

Kellogg, R. L., Lander, C. H., Moffitt, D. C., & Gollehon, N. (2000). Manure nutrients relative to the capacity of cropland and pastureland to assimilate nutrients: spatial and temporal trends for the United States. *U.S. Department of Agriculture: Natural Resources Conserve Service.*

Kendall, A. (2009). Predicting the impacts of land use and climate on regional-scale hydrologic fluxes. Michigan State University.

Koelliker, Y., Totten, L. A., Gigliotti, C. L., Offenberg, J. H., Reinfelder, J. R., Zhuang, Y., & Eisenreich, S. J. (2004). Atmospheric wet deposition of total phosphorus in New Jersey. *Water, Air, and Soil Pollution*, *154*(1–4), 139–150. https://doi.org/10.1023/B:WATE.0000022952.12577.c5

Kothawala, D. N., Watmough, S. A., Futter, M. N., Zhang, L., & Dillon, P. J. (2011). Stream Nitrate Responds Rapidly to Decreasing Nitrate Deposition. *Ecosystems*, *14*(2), 274–286. https://doi.org/10.1007/s10021-011-9422-1

Lark, T. J., Mueller, R. M., Johnson, D. M., & Gibbs, H. K. (2017). Measuring land-use and land-cover change using the U.S. department of agriculture's cropland data layer: Cautions and recommendations. *International Journal of Applied Earth Observation and Geoinformation*, *62*(June), 224–235. https://doi.org/10.1016/j.jag.2017.06.007

Linsey, C. A., Schindler, D. W., & Stainton, M. P. (1987). Atmospheric Deposition of Nutrients and Major Ions at the Experimental Lakes Area in Northwestern Ontario, 1970 to 1982. *Can. J. Fish. Aquat. Sci.*, *44*, 206–214.

Long, C. M., Muenich, R. L., Kalcic, M. M., & Scavia, D. (2018). Use of manure nutrients from concentrated animal feeding operations. *Journal of Great Lakes Research*, *44*(2), 245–252. https://doi.org/10.1016/j.jglr.2018.01.006

Lorimor, J., Powers, W., & Sutton, A. (2008). Manure Characteristics. *Mwps-18*, 1–24.

Luscz, E. C., Kendall, A. D., & Hyndman, D. W. (2017). A spatially explicit statistical model to quantify nutrient sources, pathways, and delivery at the regional scale. *Biogeochemistry*, *133*(1), 37–57. https://doi.org/10.1007/s10533-017-0305-1

Luscz, E. C., Kendall, A. D., & Hyndman, D. W. (2015). High resolution spatially explicit nutrient source models for the Lower Peninsula of Michigan. *Journal of Great Lakes Research*, *41*, 618–629. https://doi.org/10.1016/j.jglr.2015.02.004

Mahowald, N., Jickells, T. D., Baker, A. R., Artaxo, P., Benitez-Nelson, C. R., Bergametti, G., … Tsukuda, S. (2008). Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. *Global Biogeochemical Cycles*, *22*(4), 1–19. https://doi.org/10.1029/2008GB003240

Martin, S. L., Hayes, D. B., Kendall, A. D., & Hyndman, D. W. (2017). The land-use legacy effect: Towards a mechanistic understanding of time-lagged water quality responses to land use/cover. *Science of the Total Environment*, *579*, 1794–1803. https://doi.org/10.1016/j.scitotenv.2016.11.158

Martin, S. L., Hayes, D. B., Rutledge, D. T., & Hyndman, D. W. (2011). The land-use legacy effect: Adding temporal context to lake chemistry. *Limnology and Oceanography*, *56*(6), 2362–2370. https://doi.org/10.4319/lo.2011.56.6.2362

Meals, D. W., Dressing, S. A., & Davenport, T. E. (2010). Lag Time in Water Quality Response to Best Management Practices: A Review. *Journal of Environment Quality*, *39*(1), 85. https://doi.org/10.2134/jeq2009.0108

Meisinger, J. J., & Randall, G. W. (1991). Estimating Nitrogen Budgets for Soil-Crop Systems. In R. F. Follett, D. R. Keeney, & R. M. Cruse (Eds.), *Managing nitrogen for groundwater quality and farm profitability* (pp. 85–124). Madison, WI: Soil Science Society of America. https://doi.org/10.2136/1991.managingnitrogen.c5

Messier, K. P., Wheeler, D. C., Flory, A. R., Jones, R. R., Patel, D., Nolan, B. T., & Ward, M. H. (2019). Modeling groundwater nitrate exposure in private wells of North Carolina for the Agricultural Health Study. *Science of the Total Environment*, *655*, 512–519. https://doi.org/10.1016/j.scitotenv.2018.11.022

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., … Zagorski, M. A. (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences*, *201216006*. https://doi.org/10.1073/pnas.1216006110

Michigan Waterwells for WELLOGIC dataset. (2019). Lansing, MI: Michigan State University Remote Sensing and GIS Research and Outreach Services. Retrieved November 1, 2019 from http://gis-michigan.opendata.arcgis.com/datasets/water-wells-southwest-michigan

Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., Akimoto, H., … von Glasow, R. (2009). Atmospheric composition change - global and regional air quality. *Atmospheric Environment*, *43*(33), 5268–5350. https://doi.org/10.1016/j.atmosenv.2009.08.021

Motevalli, A., Naghibi, S. A., Hashemi, H., Berndtsson, R., Pradhan, B., & Gholami, V. (2019). Inverse method using boosted regression tree and k-nearest neighbor to quantify effects of point and non-point source nitrate pollution in groundwater. *Journal of Cleaner Production*, *228*, 1248–1263. https://doi.org/10.1016/j.jclepro.2019.04.293

Munger, J. W. (1982). Chemistry of atmospheric precipitation in the north-central united states: Influence of sulfate, nitrate, ammonia and calcareous soil particulates. *Atmospheric Environment (1967)*, *16*(7), 1633–1645. https://doi.org/10.1016/0004-6981(82)90258-X

Murphy, T. J. (1974). Sources of phosphorus inputs from the atmosphere and their significance to oligotrophic lakes. *WRC Research Report*. Retrieved from http://web.extension.illinois.edu/iwrc/pdf/92.pdf

Nolan, B. T., Fienen, M. N., & Lorenz, D. L. (2015). A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *Journal of Hydrology*, *531*, 902–911. https://doi.org/10.1016/j.jhydrol.2015.10.025

Paerl, H. W., & Otten, T. G. (2013). Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology*, *65*(4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y

Pearson, F. J., & Fisher, D. W. (1983). Chemical Composition of Atmospheric Precipitation in the Northeastern United States. Retrieved from https://pubs.usgs.gov/wsp/1535p/report.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

PRISM (2012). PRISM Climate Group, Oregon State University. http://www.prism.oregonstate.edu/

Preston, S. D., Alexander, R. B., Schwarz, G. E., & Crawford, C. G. (2011). Factors Affecting Stream Nutrient Loads: A Synthesis of Regional SPARROW Model Results for the

Continental United States. *Journal of the American Water Resources Association*, *47*(5), 891–915. https://doi.org/10.1111/j.1752-1688.2011.00577.x

Quinlan, J. (1996). Bagging, Boosting, and C4.5. *Proceedings Thirteenth American Association for Artificial Intelligence National Conference on Artificial Intelligence*, 725–730.

Ransom, K. M., Nolan, B. T., A. Traum, J., Faunt, C. C., Bell, A. M., Gronberg, J. A. M., … Harter, T. (2017). A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. *Science of the Total Environment*, *601–602*, 1160–1172. https://doi.org/10.1016/j.scitotenv.2017.05.192

Ray, D., Pijanowski, B. C., Kendall, A. D., & Hyndman, D. W. (2012). Coupling land use and groundwater models to map land use legacies: Assessment of model uncertainties relevant to land use planning. *Applied Geography*, *34*, 356–370. https://doi.org/10.1016/j.apgeog.2012.01.002

Robertson, D. M., & Saad, D. A. (2011). Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimted Using SPARROW Watershed Models. *Journal of the American Water Resources Association*, *47*(5), 1011–1033. https://doi.org/10.1111/j.1752-1688.2011.00574.x

Robertson, D. M., Rose, W. J., & Fitzpatrick, F. A. (2009). Water Quality and Hydrology of Silver Lake, Barron County, Wisconsin, With Special Emphasis on Responses of a Terminal Lake to Changes in Phosphorus Loading and Water Level, 50.

Rosen, C. J., Horgan, B. P., & Mugaas, R. J. (2015). Fertilizing lawns. Retrieved May 5, 2015, from http://www.extension.umn.edu:80/garden/yard-garden/lawns/fertilizing-lawns/

Ruddy, B. C., Lorenz, D. L., & Mueller, D. K. (2006). County-Level Estimates of Nutrient Inputs to the Land Surface of the Conterminous United States, 1982 – 2001. Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/sir20065012

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Paca, J., & Sobek, M. (2019). *IPUMS USA: Version 9.0: Households*. Minneapolis, MN: IPUMS. https://doi.org/https://doi.org/10.18128/D010.V9.0

Schaap, M. G., Leij, F. J., & Van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters wit hierarchical pedotransfer functions. *Journal of Hydrology*, *25*(3–4), 163–176. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000007

Scheider, W. A., Snyder, W. R., & Clark, B. (1979). Deposition of nutrients and major ions by precipitation in south-central Ontario. *Water, Air, and Soil Pollution.* https://doi.org/10.1007/BF01047121

Schindler, D. W., Armstrong, F. A. J., Holmgren, S. K., & Brunskill, G. J. (1971). Eutrophication of Lake 227, Experimental Lakes Area, Northwestern Ontario, by Addition

of Phosphate and Nitrate. *Journal of the Fisheries Research Board of Canada*, *28*(11), 1763–1782. https://doi.org/10.1139/f71-261

Schindler, D. W., Kling, H., Schmidt, R. V, Prokopowich, J., Frost, V. E., Reid, R. A., & Capel, M. (1973). Eutrophication of Lake 227 by Addition of Phosphate and Nitrate: the Second, Third, and Fourth Years of Enrichment, 1970, 1971, and 1972. *Journal of the Fisheries Research Board of Canada*, *30*(10), 1415–1440. https://doi.org/10.1139/f73-233

Schindler, D. W., Newbury, R. W., Beaty, K. G., & Campbell, P. (1976). Natural Water and Chemical Budgets for a Small Precambrian Lake Basin in Central Canada. *Journal of Fisheries Research Board of Canada*, *33*(11), 2526–2543. https://doi.org/10.1139/f76-297

Schindler, D. W., & Nighswander, J. E. (1970). Nutrient supply and primary production in clear Lake, Eastern Ontario. *Journal of the Fisheries Research Board of Canada*, *27*(11), 2009–2036. https://doi.org/https://doi.org/10.1139/f70-226

Schullehner, J., Hansen, B., Thygesen, M., Pedersen, C. B., & Sigsgaard, T. (2018). Nitrate in drinking water and colorectal cancer risk: A nationwide population-based cohort study. *International Journal of Cancer*, *143*(1), 73–79. https://doi.org/10.1002/ijc.31306

Schwede, D. B., & Lear, G. G. (2014). A novel hybrid approach for estimating total deposition in the United States. *Atmospheric Environment*, *92*, 207–220. https://doi.org/10.1016/j.atmosenv.2014.04.008

Scown, M. W., McManus, M. G., Carson, J. H., & Nietch, C. T. (2017). Improving Predictive Models of In-Stream Phosphorus Concentration Based on Nationally-Available Spatial Data Coverages. *Journal of the American Water Resources Association*, *53*(4), 944–960. https://doi.org/10.1111/1752-1688.12543

Sicard, P., Augustaitis, A., Belyazid, S., Calfapietra, C., de Marco, A., Fenn, M., … Paoletti, E. (2016). Global topics and novel approaches in the study of air pollution, climate change and forest ecosystems. *Environmental Pollution*, *213*, 977–987. https://doi.org/10.1016/j.envpol.2016.01.075

Smith, R. A., Schwarz, G. E., & Alexander, R. B. (1997). Regional interpretation of water-quality monitoring data. *Water Resources Research*, *33*(12), 2781–2798.

Smith, V. H., Tilman, G. D., & Nekola, J. C. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution*, *100*, 179–196. https://doi.org/10.1016/S0269-7491(99)00091-3

Soller, D., & Garrity, C. (2018). Quaternary sediment thickness and bedrock topography of the glaciated United States east of the Rocky Mountains. U.S Geological Survey Scientific Investigations Map 3392. https://doi.org/10.3133/sim3392

Sterner, R. (2011). C:N:P stoichiometry in Lake Superior: freshwater sea as end member. *Inland Waters, 1*(1), 29–46. https://doi.org/10.5268/IW-1.1.365

Stevenson, F. J. (1994). *Humus Chemistry: Genesis, Composition, Reactions*. John Wiley & Sons, Inc.

Swaney, D. P., Howarth, R. W., & Hong, B. (2018). Nitrogen use efficiency and crop production: Patterns of regional variation in the United States, 1987–2012. *Science of the Total Environment, 635*, 498–511. https://doi.org/10.1016/j.scitotenv.2018.04.027

Swank, W. T., & Henderson, G. S. (1976). Atmospheric Input of Some Cations and Anions to Forest Ecosystems in North-Carolina and Tennessee. *Water Resources Research, 12*(3), 541–546.

Temkin, A., Evans, S., Manidis, T., Campbell, C., & Naidenko, O. V. (2019). Exposure-based assessment and economic valuation of adverse birth outcomes and cancer risk due to nitrate in United States drinking water. *Environmental Research*, (December 2018), 108442. https://doi.org/10.1016/j.envres.2019.04.009

Tipping, E., Benham, S., Boyle, J. F., Crow, P., Davies, J., Fischer, U., … Toberman, H. (2014). Atmospheric deposition of phosphorus to land and freshwater. *Environmental Sciences: Processes and Impacts*, *16*(7), 1608–1617. https://doi.org/10.1039/c3em00641g

Tuomisto, H. L., Hodge, I. D., Riordan, P., & Macdonald, D. W. (2012). Does organic farming reduce environmental impacts? - A meta-analysis of European research. *Journal of Environmental Management*, *112*(834), 309–320. https://doi.org/10.1016/j.jenvman.2012.08.018

US Clean Water Act (1972). 33 U.S.C. §§1251‑1387.

US Census Bureau. (2010). US 2010 Census.

US Census Bureau. (2020). U.S. Census Michigan QuickFacts. Retrieved March 6, 2020, from https://www.census.gov/quickfacts/MI

US Department of Agriculture National Agricultural Statistics Service. (2012). Census of Agriculture.

US Department of Agriculture National Agricultural Statistics Service. (2015). Cropland Data Layer 2008-2015. Washington, D.C.: USDA-NASS. Retrieved from https://nassgeodata.gmu.edu/CropScape/

US Department of Agriculture Natural Resources Conservation Service (NRCS). (2008). Agricultural Waste Characteristics.

US Department of Agriculture Natural Resources Conservation Service (NRCS). (2017). Gridded Soil Survey Geographic (gSSURGO) Database: User Guide, version 2.2.

US Environmental Protection Agency. (2010). Clean Air Status and Trends Network (CASTNET). Retrieved from https://www.epa.gov/castnet

US Environmental Protection Agency. (2012). Clean Watershed Needs Survey (CWNS) - 2012 Report and Data. Retrieved from https://www.epa.gov/cwns

US Environmental Protection Agency. (2012). Clean Watershed Needs Survey (CWNS) - 2012 Report and Data. Retrieved from https://www.epa.gov/cwns

US Environmental Protection Agency. (2017). Discharge Monitoring Report (DMR) Pollutant Loading Tools. Retrieved July 10, 2017, from https://cfpub.epa.gov/dmr/

US Environmental Protection Agency. (1991). Guidance for Water Quality-based Decisions: The TMDL Process, *440/4-91–0*.

US Environmental Protection Agency. (2002). Onsite wastewater treatment systems manual.

US Geological Survey, US Department of Agriculture, Natural Resources Conservation Service. (2013). Federal standards and procedures for the National Watershed Boundary Dataset (WBD). *Techniques and Methods 11-A3*. USGS. https://doi.org/10.3133/tm11A34

Van Meter, K. J., Basu, N. B., & Van Cappellen, P. (2017). Two centuries of nitrogen dynamics: Legacy sources and sinks in the Mississippi and Susquehanna River Basins. *Global Biogeochemical Cycles*, *31*(1), 2–23. https://doi.org/10.1002/2016GB005498

Van Meter, K. J., & Basu, N. B. (2015). Catchment legacies and time lags: A parsimonious watershed model to predict the effects of legacy storage on nitrogen export. *PLoS ONE*, *10*(5), 1–22. https://doi.org/10.1371/journal.pone.0125971

Verhougstraete, M. P., Martin, S. L., Kendall, A. D., Hyndman, D. W., & Rose, J. B. (2015). Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale. *Proceedings of the National Academy of Sciences*, *112*(33), 10419–10424. https://doi.org/10.1073/pnas.1415836112

Vero, S. E., Basu, N. B., Van Meter, K., Richards, K. G., Mellander, P.-E., Healy, M. G., & Fenton, O. (2017). Review: the environmental status and implications of the nitrate time lag in Europe and North America. *Hydrogeology Journal*, 7–22. https://doi.org/10.1007/s10040-017-1650-9

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., … Tilman, D. G. (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecological Applications*. https://doi.org/10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Walton, G. (1951). Survey of literature relating to infant methemoglobinemia due to nitrate-contaminated water. *American Journal of Public Health*. https://doi.org/10.2105/ajph.41.8_pt_1.986

Wan, L., Kendall, A., Martin, S., Hamlin, Q., & Hyndman, D. (n.d.). Quantifying Nutrient Loads to the Great Lakes Cosatline with a Spatially Explicit Nutrient Transport Model.

Ward, M. H., deKok, T. M., Levallois, P., Brender, J., Gulis, G., Nolan, B. T., & VanDerslice, J. (2005). Workgroup report: Drinking-water nitrate and health - Recent findings and research needs. *Environmental Health Perspectives*, *113*(11), 1607–1614. https://doi.org/10.1289/ehp.8043

Ward, M. H., Jones, R. R., Brender, J. D., de Kok, T. M., Weyer, P. J., Nolan, B. T., … van Breda, S. G. (2018). Drinking water nitrate and human health: An updated review. *International Journal of Environmental Research and Public Health*, *15*(7), 1–31. https://doi.org/10.3390/ijerph15071557

Warncke, D. D., & Dahl, J. (2003). *Nutrient Recommendations for Vegetable Crops Grown in Michigan*. Retrieved from https://msu.edu/~warncke/Nutrient Management Info 3.1 Nutrient Recommendations for Vegetable Crops Grown in Michigan The Structure.pdf

Warncke, D., & Dahl, J. (2004). Nutrient Recommendations for Field Crops in Michigan (E2904). *Management*. Michigan State University Extension. Retrieved from https://msu.edu/~warncke/Nutrient Management Info 3.1 Nutrient Recommendations for Vegetable Crops Grown in Michigan The Structure.pdf

Warncke, D., Dahl, J., & Jacobs, L. (2004). Nutrient Recommendations for Field Crops in Michigan. *Extension Bulletin, E2904*(May).

Wheeler, D. C., Nolan, B. T., Flory, A. R., DellaValle, C. T., & Ward, M. H. (2015). Modeling groundwater nitrate concentrations in private wells in Iowa. *Science of the Total Environment*, *536*, 481–488. https://doi.org/10.1016/j.scitotenv.2015.07.080

Wieczorek, M. E., & Lamotte, A. E. (2011). Attributes for MRB_E2RF1 Catchments by Major River Basins in the Conterminous United Stages (U.S. Geological Survey Digital Data Series DS-491). Retrieved from http://water.usgs.gov/nawqa/modeling/rf1attributes.html

Winter, J. G., Catherine Eimers, M., Dillon, P. J., Scott, L. D., Scheider, W. A., & Willox, C. C. (2007). Phosphorus Inputs to Lake Simcoe from 1990 to 2003: Declines in Tributary Loads and Observations on Lake Water Quality. *Journal of Great Lakes Research*, *33*(1), 46–61. https://doi.org/10.3394/0380-1330(2007)33

Winter, J. G., Dillon, P. J., Futter, M. N., Nicholls, K. H., Scheider, W. A., & Scott, L. D. (2002). Total phosphorus budgets and nitrogen loads: Lake Simcoe, Ontario (1990 to 1998). *Journal of Great Lakes Research*, *28*(3), 301–314. https://doi.org/10.1016/S0380-1330(02)70586-8

Withers, P. J. A., Neal, C., Jarvie, H. P., & Doody, D. G. (2014). Agriculture and eutrophication: Where do we go from here? *Sustainability (Switzerland)*, *6*(9), 5853–5875. https://doi.org/10.3390/su6095853

Wright, R. F. (1976). The Impact of Forest Fire on the Nutrient Influxes to Small Lakes in Northeastern Minnesota. *Ecology*, *57*(4), 649–663. https://doi.org/10.2307/1936180

Yang, X., Miller, D. R., Xu, X., Yang, L. H., Chen, H. M., & Nikolaidis, N. P. (1996). Spatial and temporal variations of atmospheric deposition in interior and coastal Connecticut. *Atmospheric Environment*, *30*(22), 3801–3810. https://doi.org/10.1016/1352-2310(96)00094-5