Differential effects of land use on nutrient concentrations in streams of Pennsylvania

John W Clune1,*, J Kent Crawford2, William T Chappell3 and Elizabeth W Boyer1,*

1 Pennsylvania State University, Department of Ecosystem Science and Management, University Park, PA 16802, United States of America
2 Environmental Scientist, 1115 Stonegate Road, Hummelstown, PA 17036, United States of America
3 Massachusetts College of Liberal Arts, Department of Environmental Studies, North Adams, MA 01247, United States of America

* Authors to whom any correspondence should be addressed.

E-mail: ewb100@psu.edu and juc71@psu.edu

Keywords: water quality, streams, land use, Pennsylvania, nitrogen, phosphorus

Abstract
Nutrient pollution of surface waters is a widespread problem, calling for regional assessments of water quality conditions. In this study, we quantified long-term median nutrient concentrations of total nitrogen (TN) and total phosphorus (TP) in streams and rivers of Pennsylvania and explored relationships between stream nutrient concentrations and the land use of their watersheds. Our analysis is based on a synthesis of monitoring data from multiple agencies that included records of nutrient concentrations observed between 2000–2019. Across the state, stream nutrient concentrations observed in predominantly undeveloped areas (e.g., forests, shrubs, and grasslands) have median concentration values of 0.42 mg l$^{-1}$ for TN and 0.011 mg l$^{-1}$ for TP, reflecting background concentrations for minimally impacted watersheds. Median stream concentrations of TN in agricultural areas are about eleven times higher than in undeveloped areas; and are about five times higher in developed areas than in undeveloped areas. Median stream concentrations of TP in developed areas have about eight times higher concentrations than undeveloped areas; and are about four times higher in agricultural areas than in undeveloped areas. Concentrations of TN and TP increased substantially as the combined percentage of agricultural and developed land use increased. Fragmented data storage practices (e.g. incomplete metadata, ambiguous site names, and missing coordinates) and inconsistencies in monitoring protocols (e.g., differences in constituents measured, parameter names, and measurement methods) made leveraging the secondary use of multiple sources of data challenging. Nonetheless, our integrated dataset is robust, represents the best data available, and provides a new window into the nutrient status of Pennsylvania’s surface waters. The long-term median nutrient concentrations reveal the magnitude of variability in TN and TP concentrations across the state’s diverse environmental settings of land use, physiography, and geology. This information is useful for interpreting additional monitoring data, informing evaluation of water quality conditions, and guiding watershed management.

Introduction
The primary nutrients nitrogen (N) and phosphorus (P) are essential to life and necessary for plant growth, but excess loadings can cause eutrophication of surface waters. Nutrient pollution of rivers, lakes, and streams is a widespread water quality problem in the United States (Stets et al. 2020), and over two-thirds of the nation’s coastal bays and estuaries are moderately to severely impaired from nutrient enrichment (Bricker et al. 2007). Excess loadings of N and P have led to adverse environmental impacts in many surface waters, including exceedances of drinking water quality standards and impairments of water quality for aquatic life (Driscoll et al. 2003, Galloway et al. 2004, Kemp et al. 2005, USEPA 2011, Clune and Cravotta 2020). Economic impacts of nutrient pollution in the environment can be substantial, due to increased water treatment costs of removing nitrate from drinking water supplies, removing nutrients from wastewaters, and decreased revenue from the
tourism, fishing, and shellfish industries due to hypoxia, algal blooms, or toxins (Driscoll et al 2003, USEPA 2015).

Inputs of reactive forms of N and P to the environment stem from human activities associated with food production via agriculture, wastewater discharge, and energy consumption through the combustion of fossil fuels and biofuels (Galloway et al 2004, USEPA 2011). Nutrient inputs have generally increased as populations have increased—where development, deforestation, and land cultivation have contributed human waste and agricultural runoff to surface waters (Erisman et al 2008, Sutton et al 2011). In the United States, efforts to mitigate nutrient pollution problems have long been underway and are still a priority (Boyer and Howarth 2002, USEPA 2011). Voluntary bans on the use of phosphate in detergents in the 1990s led to lower phosphorus levels in streams, leaving wastewater and agricultural inputs as the important sources of phosphorus loadings (Litke 1999). The US Clean Air Act Amendments of 1990 and other policies focusing on reductions of atmospheric emissions have decreased atmospheric N deposition inputs to watersheds, and advances in water treatment technologies have reduced nutrient effluent from wastewater treatment plants (Linker et al 2013).

Despite longstanding policy and management efforts, nutrients remain an important water quality problem in Pennsylvania. Statewide assessments of waterways, under section 303(d) of the US Clean Water Act, reveal that nutrients are currently the fifth leading cause of legal water quality impairment in Pennsylvania streams that do not meet criteria for their designated uses (PA DEP 2018). In 2010, the US Environmental Protection Agency established the nation’s largest watershed-scale Total Maximum Daily Load management plan toward the goal of restoring clean water in the Chesapeake Bay. This plan’s success relies heavily on reducing the flow of nutrients from surface waters in Pennsylvania, which contribute 44% of the N loads and 24% of the P loads delivered downstream to the Bay (USEPA 2010). Efforts to stem nutrient pollution from Pennsylvania to the Chesapeake Bay have been extensive in investment and scope. However, they have not yet achieved the goals set for loadings delivered downstream to the Bay and have been insufficient for the state to meet its legal pollution reduction commitments (USEPA 2019). Nutrients are also a significant concern in the Delaware River basin of Pennsylvania, where tributaries such as the Schuylkill River have the highest mean-annual N and P yields throughout the northeast United States (Ator 2019). In the Mississippi River basin, streams that originate in the Ohio River headwaters in Pennsylvania contribute 2% of the N and P flux delivered downstream to the Gulf of Mexico (Alexander et al 2008).

Direct sources of nutrients to the landscape in Pennsylvania (figure 1) include animal manure, synthetic fertilizer application, biological nitrogen fixation, atmospheric deposition, effluent from septic systems, and point sources from wastewater treatment plants (Boyer et al 2002). Inputs of N and P from manure and fertilizer applications are among the largest sources statewide and are more heavily concentrated in the southeast than other regions of the state (figure 2). Agricultural nutrients have been a continuing source of nutrient pollution in Pennsylvania despite substantial declines in agricultural land area since the 1950s (LaMotte 2015).

A comprehensive understanding of nutrient concentrations in surface waters is needed to inform water quality evaluation and guide watershed management in Pennsylvania. Recent advances in the public availability of monitoring data offer new opportunities to compile data statewide, facilitating new insights to water–quality conditions. The objective of this study was to quantify long-term median nutrient concentrations of N and P in streams and rivers of Pennsylvania. Our analysis is based on a synthesis of monitoring data from multiple agencies that included records of nutrient concentrations observed over the past two decades from 2000–2019. We consider relationships between stream nutrient concentrations and characteristics of their watersheds (land use, physiography, and geology), providing insights into the variability of nutrient conditions among the state’s different environmental settings.

Methods

Study area
Pennsylvania covers an area of 116,083 km², with approximately 137,029 km of rivers and streams (PADEP 2018), and a population of 12.8 million people (US Census Bureau 2018). The streams of Pennsylvania drain to the Susquehanna, Ohio, Delaware, Potomac, Genesee River basins or to Lake Erie. The state’s landscape is heterogeneous in terms of environmental settings. Major physiographic provinces include the Appalachian Plateaus, Ridge and Valley, Piedmont, New England, Central Lowlands, and Coastal Plain (figure 3(a)). The bedrock geology is comprised mainly of siliciclastic (sandstone, siltstone, mudstone and shales), crystalline (schist and other metamorphic and igneous rocks), carbonate (limestone, dolomite) and sand/gravel (glacial outwash, coastal plain) formations (figure 3(b)). The land use is predominantly undeveloped (54.2%) with lesser amounts of agricultural (23.5%) and developed (15.5%) lands (figure 3(c)) (USDA 2018). These varying environmental settings influence the water quality observed in streams and rivers across Pennsylvania.
Figure 1. Inputs of nutrients to the landscape in Pennsylvania for (a) nitrogen and (b) phosphorus by percentage. Sources shown include manure, synthetic fertilizers, wet atmospheric deposition, wastewater point sources, and septic effluent. Based on data from Brakebill and Gronberg 2017, Gronberg and Arnold 2017, LaMotte 2018, Skinner and Maupin 2019, Ruddy et al 2006.

Figure 2. Nutrient inputs (kilograms/year) from fertilizer and manure across the counties of Pennsylvania, for: (a) nitrogen and (b) phosphorus. Based on 2012 estimates for manure (Gronberg and Arnold 2017) and fertilizer (Brakebill and Gronberg 2017).
Water quality data synthesis

We assembled a comprehensive dataset of total nitrogen (TN, expressed in units of mg N l\(^{-1}\)) and total phosphorus (TP, expressed in units of mg P l\(^{-1}\)) of available concentrations observed for streams and rivers of Pennsylvania from 2000–2019, by screening and combining water quality data from multiple agencies. Data sources included the US, Geological Survey (USGS), Susquehanna River Basin Commission (SRBC), Pennsylvania Department of Environmental Protection (PA DEP), Ohio Environmental Protection Agency (OEPA), New York State Department of Environmental Conservation (NYDEC) and Stroud Water Research Center obtained from the multi-agency Water Quality Portal (WQP) (NWQMC 2020). Data from the WQP was retrieved using the dataRetrieval R package (Hirsch and DeCicco 2015, Hirsch et al 2015a, 2015b). Concentration values were selected based on parameter descriptors for TN and TP, with fractions reported as either total, whole, or unfiltered. We also included TN values that were calculated as the sum of nitrogen in mixed forms, from measurements on unfiltered water samples of nitrate-nitrogen, nitrite-nitrogen, ammonia-nitrogen, and organically bonded nitrogen.

Duplicate data records were removed from the dataset. Raw data records that had important metadata missing (e.g., coordinates or sampling date), or with reported missing TN or TP values were not used. Outliers that were outside six standard deviations from the mean were also removed (Oelsner et al 2017). Censored data observations (non-detects) having concentration values less than the reporting limit were set to the detection...
limit and medians or percentiles above the highest reporting level were used to present descriptive statistics (Ator and Denver 2012). Results that were labeled as non-detect where TN > 0.25 mg l$^{-1}$ or TP > 0.10 mg l$^{-1}$ were considered erroneous and removed from the dataset. A total of 55,437 useable nutrient concentration values were available, observed at 1307 stream monitoring sites. Not every site had data for the full 20-year period, but each site had some amount of data within that period. For each site, a discrete median concentration for each season for each year from 2000 to 2019 was calculated from the raw nutrient concentration data, and these values were used as the aggregate dataset for analysis, following methods recommended by the US Environmental Protection Agency (USEPA, variously dated). The use of median-seasonal concentrations minimizes bias from outliers or undue influence by a small number of monitoring sites. Medians were calculated from at least three observations at a site during a season over the period of record. If only two observations were available for a site, the average was computed; and if only one observation was available, the single true value was used. This approach preserves the spatial extent of observations and the range of streams sampled among the land use and geological settings in Pennsylvania (figure 4). Eliminating data (e.g. requiring 3 or more observations to be included) would greatly reduce the spatial extent of monitoring site coverage, biasing the data analysis toward information from only frequently sampled sites, such as large rivers and locations routinely monitored due to pollution issues.

Using the available data, we calculated 27,683 median-seasonal concentration values based on the observations at the individual stream monitoring sites, with 13,861 median-seasonal concentration values for TN and 13,822 median-seasonal concentration values for TP. This aggregate dataset of median-seasonal concentrations was then used to calculate the overall medians for the period of record (2000–2019), as well as for each season within that period. Seasonal timeframe classifications were based on those used for ambient water quality criteria recommendations for nutrient ecoregions of Pennsylvania (USEPA, variously dated). To simplify terminology, the use of nutrient concentration in this paper refers to the aggregate values of median-seasonal concentration as described. Summary and frequency distributions of the median values for TN and TP were developed to provide regional-scale knowledge of nutrient concentrations across land use settings, physiographic provinces, and geological settings of Pennsylvania.
Watershed characteristics
Characteristics including drainage area, land use, geology, and physiography were based on the entire watershed area delineated upstream of each monitoring site location. Watershed boundaries sometimes extended outside of Pennsylvania’s borders to the watershed divide, and smaller watersheds were sometimes nested within larger watersheds. These attributes were calculated from geospatial data using geographic information system, statistical, and data analysis software (ESRI 2016, TIBCO 2016, R Core Team 2017, JMP 2018). Python programming language code was developed to automate the geospatial calculations for watershed delineations and percent land use and geology for the large number of sampling sites, using the USGS StreamStats application programming interface service and ArcGIS software (ESRI 2016, PSF 2019, USGS 2019).

Land use types were classified as undeveloped, agricultural, developed, or mixed using geospatial information from the 2011 National Land Cover Database (USGS 2014), which represents a mid-point of our study period. Undeveloped areas are minimally impacted lands such as forests, shrubs, and grassland. Agricultural lands are areas with cultivated crops or pasture. Developed areas include urban and suburban centers, but also high, medium, and low intensity development. Mixed land use was designated for all other combinations of developed, agricultural, and undeveloped land. The dominant land use type for the upland watershed draining to each stream sampling site was assigned based on criteria developed by Dubrovsky et al. (2010), but modified to further classify agricultural areas as watersheds having >45 percent agricultural land and ≤25 percent developed land. Physiographic provinces were assigned to each watershed based on geospatial data from PA DCNR (2008). Major geological types were assigned to each watershed based on geospatial data from Horton (2017). The primary geological setting for each site was defined as siliciclastic or crystalline if >75% of the respective geology encompassed the watershed, whereas sites were assigned carbonate classification if carbonate geology was >25%.

Results
Spatial patterns
A summary of nutrient concentrations for physiographic provinces, geological settings, and land use coverages across the entire period and each season for Pennsylvania are presented in table 1. The size of the sample population of median concentrations generally reflects the proportion of area for each setting represented in Pennsylvania. For instance, the Central Lowlands physiographic province has only 102 TN concentrations and 100 TP concentrations because this setting occupies only a very narrow strip in the northwest portion of the state. The four major geological settings with substantial observations were used for comparison, and excluded the sand/gravel setting with too few observations for analysis. A large population of concentration values was available for all land use types.

Variation with land use
Total nitrogen and total phosphorus concentrations in streams showed marked differences among land use (table 1, figure 5). For total nitrogen (figure 5(a)), median concentrations were highest in streams draining landscapes with agricultural land use (4.60 mg l⁻¹), followed by developed (2.20 mg l⁻¹), mixed (1.03 mg l⁻¹), and undeveloped (0.42 mg l⁻¹). Concentrations of TN were about 11 times greater in agricultural areas than undeveloped areas, and about 5 times greater in developed than undeveloped areas. For total phosphorus (figure 5(b)), median concentrations were highest in streams draining landscapes with developed land use (0.090 mg l⁻¹), followed by agricultural (0.042 mg l⁻¹), mixed (0.030 mg l⁻¹), and undeveloped (0.011 mg l⁻¹). Concentrations of TP were about 8 times greater in developed areas than undeveloped areas, and about 4 times greater in agricultural than undeveloped areas. Total nitrogen and total phosphorus median concentrations in streams varied seasonally among land use categories (figure 6). Regardless of the season, land use in order of highest to lowest TN concentrations was ranked as: agricultural > developed > mixed > undeveloped. Similarly, regardless of the season, land use in order of highest to lowest TP concentrations was ranked as: developed > agricultural > mixed > undeveloped.

Variation with geology
Over the period of record (table 1), median concentrations for TN were highest in crystalline (4.30 mg l⁻¹) and carbonate (4.00 mg l⁻¹) settings compared to mixed (2.78 mg l⁻¹) and siliciclastic (0.70 mg l⁻¹) rock types. Median concentrations for TP were highest in mixed geological settings (0.109) followed by crystalline (0.040 mg l⁻¹), carbonate (0.039 mg l⁻¹), and siliciclastic (0.020) rock types. These results are consistent with the notion that there is generally higher delivery of nutrients to streams from higher permeability areas, where intrinsic permeability values for major rock types typically rank as fractured crystalline > carbonates > siliciclastic sandstone > shale > unfractured crystalline (Hornberger et al 2014). Median concentration values of TN and TP in streams varied greatly across

---

*References were not included in the text as they are not relevant to the content.*
different geological settings crossed with land use (figure 7), though there is an unbalanced number of observations available in each category (see table 1) and no explicit information on the intensity of land use in each category, limiting the ability to identify geological versus land-use controls.

### Discussion

**Differential effects of land use**

Concentrations of nutrients in stream and river networks have been well-studied at regional and national scales (Langland et al. 2001, Smith et al. 2003, Mueller and Spahr 2006, Dubrovsky et al. 2010, Moore et al. 2011, Preston et al. 2011, Saad et al. 2011). This analysis is unique in presenting a comprehensive nutrient water quality dataset for Pennsylvania, providing new insights into how water quality conditions statewide are greatly influenced by land use (see table 1).

Median concentrations of TN observed in Pennsylvania for agricultural (4.60 mg l⁻¹), developed (2.20 mg l⁻¹), mixed (1.03 mg l⁻¹), and undeveloped (0.42 mg l⁻¹) land use categories were comparable to median concentrations observed nationally as reported by Dubrovsky et al. (2010) for agricultural (3.8 mg l⁻¹), urban (1.5 mg l⁻¹), mixed (1.3 mg l⁻¹), and undeveloped (0.50 mg l⁻¹) land uses. Agricultural inputs from fertilizer and manure provide large N and P sources to the Pennsylvania landscape, especially in the southeastern region (see figure 2). Excess nutrient inputs—along with traditional agricultural practices such as tillage, tile
drains, and ditches—have been shown to facilitate the transport of nutrients to streams, contributing to elevated nutrient concentrations in agricultural settings (Zhao et al. 2001, Clune and Denver 2012). The median concentrations of TN for agricultural settings may be higher than national results because productive farmlands in Pennsylvania are typically confined to valleys with carbonate lithology that is permeable, and therefore more vulnerable to nitrate contamination (Lizarraga 1997, Miller et al. 1997, Clune and Cravotta 2020). Though our dataset does not allow for statistical analyses of seasonal differences or temporal trends, other studies have shown that TN concentrations issuing from streams in vegetated areas can be markedly less in the spring or summer than in the fall or winter, due to nutrient uptake by plants during the growing season (Lee et al. 2012, Clune and Denver 2012, Baker et al. 2014). Higher TN concentrations observed in this study for carbonate settings overlain by agricultural land (see figure 7) have also been found among regional models used to help focus the location of restoration efforts (Ator et al. 2011).

Median concentrations of TP observed by land use category in Pennsylvania were highest in developed (0.090 mg L⁻¹) followed by agricultural (0.042 mg L⁻¹), mixed (0.030 mg L⁻¹), and undeveloped (0.011) land use settings. In the national scale study, concentrations were highest in agricultural (0.26 mg L⁻¹), urban (0.25 mg L⁻¹) settings, and lower in mixed (0.13 mg L⁻¹) and undeveloped (0.04 mg L⁻¹) settings (Dubrovsky et al. 2010). In recent decades, areas of Pennsylvania within the Chesapeake Bay watershed have been showing declining trends in TP concentrations even as orthophosphate concentrations have been increasing (Fanelli et al. 2019). Seasonally, TP concentrations are often largest in the summer (see figure 6), perhaps because of the strong influence of wastewater effluent as a percentage of total flow when streams are at their annual minimums (Land et al. 1998, Dubrovsky et al. 2010). The wider range of TP concentrations in developed areas has been attributed to the effects of fluctuations in runoff and soil-bound phosphorus from impervious surfaces (e.g. lawns, golf courses, etc), degrading streams, construction sites, and unsewered development (Carpenter et al. 1998).
Overall, land use strongly influences nutrient concentrations in surface waters (see table 1 and figure 5). Median stream concentrations of TN are about 11 times higher in agricultural areas than in undeveloped areas; and are about 5 times higher in developed areas than in undeveloped areas. Median stream concentrations of TP are about 8 times higher in developed areas than in undeveloped areas; and are about 4 times higher in agricultural areas than in undeveloped areas. Concentrations of TN and TP increased according to the combined percentage of agricultural and developed land (figure 8). This highlights that nutrient water quality conditions can degrade as natural ecosystems are converted to agriculture or become urbanized. The median concentration values where the combined percentage of agricultural and developed land was less than 20 percent (0.42 mg l$^{-1}$ for TN 0.011 mg l$^{-1}$ for TP) can be taken to represent background nutrient concentrations for minimally impacted and predominately undeveloped watersheds. Biological and habitat integrity have been shown to be of consistently good quality at sites with watersheds having greater than 80% of undeveloped (and mostly forested) land use (Wang et al 1997).

**Challenges and recommendations for use of multi-agency data**

Our analysis is based on a synthesis of stream monitoring data (observed between 2000–2019) from multiple agencies. Sufficient nutrient water quality data exist for most areas of the state, allowing us to identify undeveloped (background) or reference nutrient stream conditions for most combinations of land use and

---

**Figure 6.** Median concentrations from 2000–2019 observed at stream monitoring stations in Pennsylvania for (a) total nitrogen and (b) total phosphorus—across land use settings by season.
geological settings (see table 1). However, there was limited data for nutrients in streams in particular locations such as carbonate geology having predominately undeveloped land use. One area where this may become important is southeastern Pennsylvania, where agriculture has intensified and expanded, and additional monitoring data in this region may aid conservation efforts.

It is important to note that significant amounts of the raw data records for Pennsylvania assembled from the multi-agency legacy databases were unusable in this study. Approximately 58% of nutrient concentration data had to be removed from our analysis due to data storage practices. Duplication of data among and within data sources was a significant issue. The Water Quality Portal notes that duplication of records stems, in part, from problems with the fact that the site identifiers from various contributing organizations have not been harmonized (NWQMC 2020). There were also many metadata issues, with improper labeling of waterbodies, incorrect units, zero or negative concentrations, and erroneous data and outliers not quality assured or reviewed. Information about the monitoring sites such as station names and descriptions often lacked important details such as stream name or location. Incorrect coordinates made it difficult and time-consuming to perform

![Figure 7. Median concentrations from 2000–2019 observed at stream monitoring stations in Pennsylvania for (a) total nitrogen and (b) total phosphorus—across geological crossed with land use settings.](image)
geospatial analysis of watershed boundaries and develop percent geology and land use areas. Information about individual samples collected at monitoring sites often lacked adequate recording of the date and time of sampling, or sample collection methods. Information about specific water quality characteristics often had inconsistencies among parameter names, chemical forms, units, remarks, and other notations important for data synthesis and reuse.

Our efforts at compilation and analysis of available data statewide highlight the broader challenges in leveraging data from multiple independent sources. Many individual data records of nutrient concentration observations cannot reliably be used in subsequent studies due to the fragmented data storage issues (e.g. the duplicate records or incomplete metadata) or cannot be easily integrated due to inconsistencies in monitoring protocols (e.g., differences in specific parameters measured, parameter naming conventions, or measurement methods used) (Sprague et al. 2017). Additionally, some water quality observations have limited use in future studies due to sampling limitations, for example, where co-located streamflow data are not available to facilitate analyses of flow-weighted concentrations or loads, or where the temporal frequency of sampling is inadequate to facilitate analyses of long-term statistics or trends. It has been estimated that the collection and storage of ambiguous water quality data has cost the United States $6.8 to $19 billion dollars (Sprague et al 2017).

To ensure wise investments in environmental monitoring and maximize the potential for environmental data analyses, agencies involved in water quality data collection in Pennsylvania would benefit from following contemporary guidance for data collection, storage, and access to better enable the merging and reuse of datasets (NWQMC 2006). Coordination among such agencies could aid with watershed planning and could reduce costs associated with water quality monitoring. Similarly, collaborative efforts toward shared and reliable datasets across agencies have the potential to improve the scientific basis for decision making. To facilitate this, we suggest the formation of a statewide interagency committee on water information. Such a committee could bring together stakeholders and serve an advisory role for sharing recommended sampling and metadata protocols and develop a plan to resolve issues with the synthesis of legacy data. An updated comprehensive dataset and automated data retrieval methods could be used as the foundation for modern data analytics web tools for data visualization and access for scientists and resource managers toward decision making.

Figure 8. Tukey boxplot of median concentrations from 2000–2019 observed at stream monitoring stations in Pennsylvania for (a) total nitrogen and (b) total phosphorus. Nutrient concentrations in streams increase according to the combined percentage of agricultural and developed land uses.
Conclusions

We quantified long-term median concentrations of nutrients in Pennsylvania’s streams and rivers using the best available data sampled at monitoring sites from 2000 to 2019. Our results reveal the magnitude and variability of nutrient concentrations across the state’s diverse environmental settings. Land use strongly influenced nutrient water quality conditions. During all seasons, TN and TP concentrations were higher in streams in agricultural, developed, or mixed land use watersheds than in undeveloped watersheds. Elevated nutrient concentrations were evident in streams draining watersheds with predominantly agricultural and developed land uses. The median nutrient concentration values presented here (see table 1) provide new baseline data and a window into stream water quality conditions statewide. This information is useful for interpreting additional monitoring data, informing water quality evaluation, and guiding watershed management.

Comprehensive datasets of water quality observations are essential for understanding the status and trends of nutrient concentrations in surface waters and highlighting progress toward achieving water quality goals. Such data can help understand water quality trends, changes in nutrient use efficiency (e.g., through advances in precision agriculture or crop genetics), and the effectiveness of best management practices; and help identify locations for mitigation efforts. The use of legacy and future stream monitoring data will continue to be essential for environmental assessment, and our work highlights the need for consistent sampling protocols and open data sharing.

Acknowledgments

This work was partially funded with support to EWB from the National Institute of Food and Agriculture regional research programs, under Project # PEN04730 and Accession # 1022594. We thank the College of Agricultural Sciences at Penn State University for support to JC. We sincerely appreciate the cooperative effort of the US Geological Survey, the US Environmental Protection Agency, and the National Water Quality Monitoring Council to integrate and serve publicly available water quality data collected by federal, state, tribal, and local agencies via the collaborative Water Quality Portal.

Conflict of interest

The authors express no conflict of interest.

ORCID iDs

John W Clune @ https://orcid.org/0000-0002-3563-1975
Elizabeth W Boyer @ https://orcid.org/0000-0003-4369-4201

References

Alexander R B, Smith R A, Schwarz G E, Boyer E W, Nolan J V and Brakebill J W 2008 Differences in sources and recent trends in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi and Atchafalaya River Basins Environmental Science & Technology 42 822–30

Ator S W 2019 Spatially referenced models of streamflow and nitrogen, phosphorus, and suspended-sediment loads in streams of the northeastern United States USGS Scientific Investigations Report 2019 – 5118

Ator S W, Brakebill J W and Blomquist J D 2011 Sources, fate, and transport of nitrogen and phosphorus in the Chesapeake Bay watershed—an empirical model: US Geological Survey Scientific Investigations Report 2011 – 5167, 27 p, (https://pubs.usgs.gov/sir/2011/5167/)

Ator S W and Denver J M 2012 Estimating contributions of nitrate and herbicides from groundwater to headwater streams, Northern Atlantic Coastal Plain, United States J. Am. Water Resour. Assoc. 48 1075–90

Baker R J, Wieben C M, Lathrop R G and Nicholson R S 2014 Concentrations, loads, and yields of total nitrogen and total phosphorus in the Barnegat Bay–Little Egg Harbor watershed, New Jersey, 1989–2011, at multiple spatial scales USGS Scientific Investigations Report 2014 – 5072

Boyer E W, Goodale C L, Jaworski N A and Howarth R W 2002 Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern USA Biogeochemistry 57 137–69

Boyer E W and Howarth R W (ed) 2002 The Nitrogen Cycle at Regional to Global Scales. (Netherlands: Springer Publications) p 519

Brakebill J W and Gronberg J A M 2017 County-level Estimates of Nitrogen and Phosphorus from Commercial Fertilizer for the Contaminous United States, 1987–2012. (Reston, VA: US Geological Survey) (https://doi.org/10.5066/FLHF41PKX)

Bricker S, Longstaff B, Dennison W, Jones A, Boicourt K, Wicks C and Woermer J 2007 Effects of nutrient enrichment in the nation’s estuaries—A decade of change. NOAA Coastal Ocean Program Decision Analysis Series No. 26. (https://repository.library.noaa.gov/view/noaa/17779)

Carpenter S R, Caraco N F, Correll D L, Howarth R W, Sharples A N and Smith V H 1998 Non-point pollution of surface waters with phosphorus and nitrogen Ecological Applications 8 559–68
Clune J W and Cravotta C A III 2020 Groundwater quality in relation to drinking water health standards and geochemical characteristics for 54 domestic wells in Clinton County, Pennsylvania, 2017 USGS Scientific Investigations Report 2020–5022

Clune J W and Denver J M 2012 Residence time, chemical and isotopic analysis of nitrate in the groundwater and surface water of a small agricultural watershed in the Coastal Plain, Bucks Branch, Sussex County, Delaware USGS Scientific Investigations Report 2012–5235 (http://pubs.usgs.gov/sir/2012/5235)

Driscoll C et al 2003 Nitrogen pollution: sources and consequences in the US northeast Environment 45 8–22

Dubrovsky N M et al 2010 The quality of our nation’s waters-Nutrients in the nation’s streams and groundwater, 1992–2004. USGS Circular No. 1350 (http://pubs.usgs.gov/circ/1350/)

Erisman J W, Sutton M A, Galloway J, Klimont Z and Winiswarter W 2008 How a century of ammonia synthesis changed the world Nat. Geosci. 1 636–9

ESRI 2016 ArcGIS Desktop (Version 10.4.1) [Software]. (Redland, CA: Environmental Systems Research Institute)

Fanelli R M, Blomquist J D and Hirsch R M 2019 Point sources and agricultural practices control spatial-temporal patterns of orthophosphate in tributaries to Chesapeake Bay Sci. Total Environ. 652 422–33

Galloway J N et al 2004 Nitrogen cycles: past, present, and future Biogeochemistry 70 153–226

Grönborg J M and Arnold T L 2017 County-level estimates of nitrogen and phosphorus from animal manure for the conterminous United States, 2007 and 2012 USGS Open-File Report 2017–1021

Hirsch R M and De Cicco L A 2015 User Guide to Exploration and Graphics for River Trends (EGRest) and dataRetrieval: R Packages for Hydrologic Data (version 2.0, February 2015). (Reston, VA: US Geological Survey Techniques and Methods) Book 4, Chap. A10 (https://doi.org/10.3133/tm4A10)

Hirsch R M, De Cicco L A and Lorenz D L 2015a dataRetrieval Reference Manual. (Reston, VA: US Geological Survey) (https://cran.r-project.org/web/packages/dataRetrieval/dataRetrieval.pdf)

Hirsch R M, De Cicco L A and Lorenz D L 2015b dataRetrieval: Retrieval Functions for USGS and EPA Hydrologic and Water Quality Data. (US Geological Survey) (https://crdr.io/cran/dataRetrieval/)

Hornberger G M, Wilberg P L, Raffensperger J P and D’Odorico P 2014 Elements of Physical Hydrology 2nd edn (Baltimore, MD: Johns Hopkins University Press) p 388

Horton J D 2017 The State Geologic Map Compilation (SGMC) Geodatabase of the Conterminous United States (ver. 1.1, August 2017) (Reston, VA: US Geological Survey data release (https://doi.org/10.5066/F7WH2N65)

JMP 2018 JMP® Software (Version 14) [Software]. (Cary, NC: SAS Institute Inc)

Kemp W M et al 2005 Eutrophication of Chesapeake Bay: historical trends and ecological interactions Marine Ecology Progress Series 303 1–29

LaMotte A E 2015 Selected Items from the Census of Agriculture at the County Level for the Conterminous United States, 1950–2012. (US Geological Survey) (https://doi.org/10.5066/F7H13016)

LaMotte A E 2018 Estimated Nitrogen from Septic for the Conterminous United States, 2010. (Reston, VA: US Geological Survey) (https://doi.org/10.5066/F9Q7GIS7)

Land J F, Moring J B, Van Metre P C, Reutter D C, Mahler B J, Shipp A A and Ulery R L 1998 Water Quality in the Trinity River Basin, Texas, 1992–95 (Reston, VA: US Geological Survey Circular) 1171 (http://pubs.usgs.gov/circ/1171/)

Langland M J, Edwards R E, Sprague L A and Yochum S E 2001 Summary of trends and status analysis for flow, nutrients, and sediments at selected nontidal sites, Chesapeake Bay basin, 1985–99 USGS Open-File Report 2001–73 49

Lee K E, Lorenz D L, Petersen J C and Greene J B 2012 Seasonal patterns in nutrients, carbon, and algal responses in wadeable streams within three geographically distinct areas of the United States, 2007–08 USGS Scientific Investigations Report 2012–5086. (Reston, VA: USGS Publications Warehouse) (https://doi.org/10.3133/sir20125086)

Linker I C, Dennis R, Shenk G W, Batuik R A, Grimm J and Wang P 2013 Computing atmospheric nutrient loads to the Chesapeake Bay watershed and tidal waters JAWRA J. Am. Water Resour. Assoc. 49 1025–41

Litke D W 1999 Review of phosphorus control measures in the United States and their effects on water quality USGS Water-Resources Investigations Report 99–4007

Lizzaraga J S 1997 Estimation and Analysis of Nutrient and suspended-Sediment Loads at Selected Sites in the Potomac River Basin, 1993–95. (Reston, VA: USGS Water-Resources Investigations Report) 97–4154

Miller C V, Denis J M, Atri S W and Brakelhill J W 1997 Nutrients in streams during baseflow in selected environmental settings of the Potomac River basin J. Am. Water Resour. Assoc. 33 1135–71

Moore R B, Johnston C M, Smith R A and Milstead B 2011 Source and delivery of nutrients to receiving waters in the northeastern and Mid-Atlantic regions of the United States J. Am. Water Resour. Assoc. 47 965–90

Mueller D K and Spahr N E 2006 Nutrients in streams and rivers across the nation—1992–2001 USGS Scientific Investigations Report 2006–5107 (http://pubs.usgs.gov/sir/2006/5107/)

NWQM 2006 National Water Quality Monitoring Council. Water Quality Data Elements: A user guide. (https://acwi.gov/methods/pubs/wqde_pubs/wqde_trno3.pdf)

NWQM 2020 National Water Quality Monitoring Council. Water Quality Portal (WQP). (https://waterqualitydata.us/)

Oelner G P et al 2017 Water-quality trends in the nation’s rivers and streams, 1972–2012—Data preparation, statistical methods, and trend results USGS Scientific Investigations Report 2017–5006 (http://pubs.er.usgs.gov/publication/sir175006)

PA DCNR 2008 Physiographic sections of Pennsylvania: Pennsylvania Bureau of Topographic and Geologic Survey (Harrisburg, PA: Dept. of Conservation and Natural Resources) (http://pasda.psu.edu/ci/DataSummary.aspx?dataset=1154)

PA DEP 2018 2018 Final Pennsylvania integrated water quality monitoring and assessment report (Clean Water Act, section 305(b) report and 303(d) list). (Harrisburg, PA: Pennsylvania Department of Environmental Protection) (https://dep/Publication/sir/2018_integrated_report/index.html)

Preston S D, Alexander R B, Schwarz G E and Crawford C G 2011 Factors affecting stream nutrient loads: a synthesis of regional SPARROW model results for the continental United States J. Am. Water Resour. Assoc. 47 891–915

PSF 2019 Python Language Reference (Version 3.8.1) [Software]. (https://python.org/)

R Core Team 2017 R: A Language and Environment for Statistical Computing Software. (R Foundation for Statistical Computing) (https://cran.r-project.org/)

Ruddy B C, Lorenz D L and Mueller D K 2006 County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982–2001 USGS Scientific Investigations Report 2006–5012 (http://pubs.usgs.gov/sir/2006/5012/)

Saad A, Schwarz G E, Robertson D M and Booth N L 2011 A multi-agency nutrient dataset used to estimate loads, improve monitoring design, and calibrate regional nutrient SPARROW models J. Am. Water Resour. Assoc. 47 933–49

Skinner K D and Maupin M A 2019 Point-source nutrient loads to streams of the conterminous United States, 2012 USGS Data Series 1101
Smith R A, Alexander R B and Schwartz G E 2003 Natural background concentrations of nutrients in streams and rivers of the conterminous United States Environmental Science & Technology 37 3039–47
Sprague L A, Oelsner G P and Argue D M 2017 Challenges with secondary use of multi-source water-quality data in the United States Water Res. 110 252–61
Stets E G, Sprague L A, Oelsner G P, Johnson H M, Murphy J C, Ryberg K, Vecchia A V, Zueilig R E, Falcone J A and Riskin M L. 2020 Environmental Science & Technology 54 4336–43
Sutton M A, Oenema O, Erisman J W, Leip A, van Grinsven H and Winiwarter W 2011 Too much of a good thing Nature 472 159–61
TIBCO 2016 Spotfire Desktop (Version 7.5.0) [Software]. (Palo Alto, CA: TIBCO Software Inc.)
U.S. Census Bureau 2018 Annual Estimates of the resident population: April 1, 2010 to July 1, 2018. (Suitland, MD: US Census Bureau Population Division) (https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF)
USEPA variously dated. Ecoregional nutrient criteria for rivers & streams documents. (https://epa.gov/nutrient-policy-data/ecoregional-nutrient-criteria-rivers-streams-documents)
USEPA 2010 Chesapeake Bay Total Maximum Daily Load Document. (Annapolis, MD: US Environmental Protection Agency, Chesapeake Bay Program Office) (https://epa.gov/chesapeake-bay-tmdl/chesapeake-bay-tmdl-document)
USEPA 2011 Reactive nitrogen in the United States: An analysis of inputs, flows, consequences, and management options. Washington, DC: Science Advisory Board, Report EPA-SAB-11-013, 172p
USEPA 2015 A compilation of cost data associated with the impacts and control of nutrient pollution. Washington, DC: US Environmental Protection Agency, EPA-820-F-15-096
USEPA 2019 EPA Final Evaluation of 2016-2017 Milestone and Midpoint Progress and 2018-2019 Milestone Commitments in the Chesapeake Bay Watershed. (Washington, DC: US Environmental Protection Agency) (https://epa.gov/chesapeake-bay-tmdl/epa-final-evaluation-2016-2017-milestone-and-midpoint-progress-and-2018-2019)
USGS 2014 National Land Cover Database, 2011 Land Cover (2011 Edition, amended 2014) - National Geospatial Data Asset (NGDA) Land Use Land Cover (Reston, VA: US Geological Survey) (https://sciencebase.gov/catalog/item/5760129de4b04f417c2c28b7)
USGS 2019 StreamStats Services. (Reston, VA: US Geological Survey) (https://streamstats.usgs.gov/docs/streamstatsservices/#/)
Wang L, Lyons J, Kanehl P and Gatti R 1997 Influences of watershed land use on habitat quality and biotic integrity in Wisconsin Streams Fisheries 22 6–12
Zhao S L, Gupta S C, Huggins D R and Moncrief J F 2001 Tillage and nutrient source effects on surface and subsurface water quality at corn planting. J. Environ. Qual. 30 998–1008