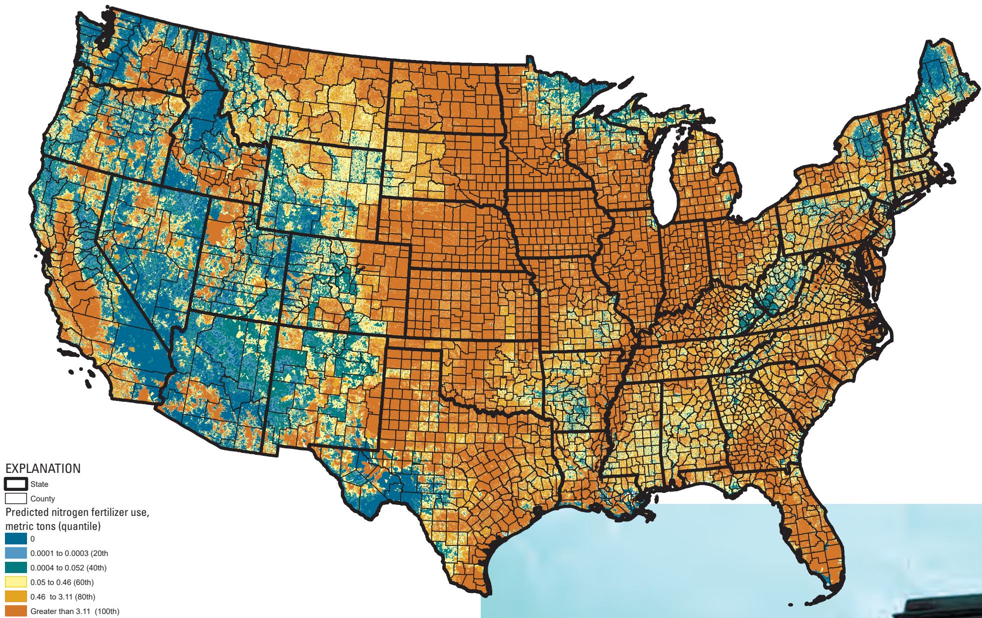
National Water Quality Program

Catchment-Level Estimates of Nitrogen and Phosphorus Agricultural Use from Commercial Fertilizer Sales for the Conterminous United States, 2012



Scientific Investigations Report 2018–5145

U.S. Department of the Interior
U.S. Geological Survey

Cover photographs:

Top: Map showing predicted 2012 nitrogen fertilizer use at the catchment scale (mertic tons), unconditional predictions.

Bottom: Photograph by Lynn Betts, U.S. Department of Agriculture, Natural Resources Conservation Service.

Catchment-Level Estimates of Nitrogen and Phosphorus Agricultural Use from Commercial Fertilizer Sales for the Conterminous United States, 2012

By Jana S. Stewart, Gregory E. Schwarz, John W. Brakebill, and
Stephen D. Preston

National Water Quality Program

Scientific Investigations Report 2018–5145

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DAVID BERNHARDT, Acting Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2019

Suggested citation:
Stewart, J.S., Schwarz, G.E., Brakebill, J.W., and Preston, S.D., 2019, Catchment-level estimates of nitrogen and phosphorus agricultural use from commercial fertilizer sales for the conterminous United States, 2012: U.S. Geological Survey Scientific Investigations Report 2018–5145, 52 p., https://doi.org/10.3133/sir20185145.

ISSN 2328–0328 (online)

Foreword

Sustaining the quality of the Nation's water resources and the health of our diverse ecosystems depends on the availability of sound water-resources data and information to develop effective, science-based policies. Effective management of water resources also brings more certainty and efficiency to important economic sectors. Taken together, these actions lead to immediate and long-term economic, social, and environmental benefits that make a difference to the lives of the almost 400 million people projected to live in the United States by 2050. (https://water.usgs.gov/nawqa/applications/).

In 1991, Congress established the National Water-Quality Assessment (NAWQA) to address where, when, why, and how the Nation's water quality has changed, or is likely to change in the future, in response to human activities and natural factors. Since 1991, NAWQA has been a leading source of scientific data and knowledge used by national, regional, state, and local agencies to develop science-based policies and management strategies to improve and protect water resources used for drinking water, recreation, irrigation, energy development, and ecosystem needs. Plans for the third decade of NAWQA (2013–23) address priority water-quality issues and science needs identified by NAWQA stakeholders (such as the Advisory Committee on Water Information) and the National Research Council. The plans are designed to meet increasing challenges related to population growth, increasing needs for clean water, and changing land use and weather patterns.

Federal, state, and local agencies have invested billions of dollars to reduce the amount of pollution entering rivers and streams that millions of Americans rely on for a variety of water needs and biota rely on for habitat. Understanding the sources and transport of pollution is crucial for designing strategies to improve water quality. Studies have indicated that estimates of total nitrogen and total phosphorus commercial fertilizer use are empirically important for estimating water-quality conditions in streams using models such as the U.S. Geological Survey's (USGS) SPAtially Referenced Regressions On Watershed attributes model know as SPARROW (Preston and others, 2011). This report describes the methods and subsequent results of two models developed for estimating elemental nitrogen and phosphorus commercial fertilizer use on agricultural lands for the conterminous United States at the National Hydrography DatasetPlus (NHDPlus) catchment/county scale for the year 2012. The results of these models will prove useful for any water-quality models that estimate total-nitrogen and total-phosphorus loads to streams and for other studies needing fertilizer use estimates related to agricultural cropping practices in the United States.

The authors hope this publication will provide insights and information to meet water resource needs and will foster increased citizen awareness and involvement in the protection and restoration of the Nation's waters. The information in this report is intended primarily for those interested or involved in resource management and protection, conservation, regulation, and policymaking at the regional and national level.

Dr. Donald W. Cline
Associate Director for Water
U.S. Geological Survey

Acknowledgments

The authors thank Dr. Larry Oldham, Mississippi State University, for his assistance and feedback on model development.

The authors thank Michael Wieczorek, Andrew LaMotte, Shannon Jackson, Sylvia Terziotti, and Nancy Baker for the development and support of providing model input data. The authors thank Michael Woodside, Neil Dubrovsky, and Anna Garcia for assistance and feedback in model development and David Anning, David Saad, Daniel Wise, Scott Ator, Anne Hoos, and Victor Roland for reviewing model results. The authors also thank and acknowledge the reviews provided by Nancy Baker, James Falcone, and Karen Ryberg.

Contents

Figures

Tables

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Area	
square meter (m^2)	10.7639	square foot (ft^2)
square kilometer (km^2)	247.1	acre
	Mass	
kilogram per year (kg/yr)	2.205	pound avoirdupois per year (lb/yr)
metric ton (t)	2204.62	pound, avoirdupois (lb)
millimeter per year (mm/yr)	0.0393701	inch per year (in/yr)
	Application rate	
metric ton per square kilometer (t/km^2)	8.92179122	pound per acre (lb/acre)

Datum

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Abbreviations

AAPFCO	Association of American Plant Food Control Officials
BLUE	Best Linear Unbiased Estimator
COA	Census of Agriculture
CDL	Cropland Data Layer
IPNI	International Plant Nutrition Institute
NAWQA	National Water-Quality Assessment
NHDPlus	National Hydrography Data Plus
NHDPlusV2	National Hydrography Data Plus version 2
P_2O_5	phosphate
R^2	coefficient of determination
RMSE	root mean square error
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey

Catchment-Level Estimates of Nitrogen and Phosphorus Agricultural Use from Commercial Fertilizer Sales for the Conterminous United States, 2012

By Jana S. Stewart, Gregory E. Schwarz, John W. Brakebill, Stephen D. Preston

Abstract

Nutrient inputs from commercial agricultural fertilizer, particularly nitrogen and phosphorus, are important factors contributing to the degradation of surface-water quality and the alteration of aquatic ecosystems. Despite this importance, information about the application of fertilizer to agricultural land is not available in a consistent manner across the United States at a scale useful for regional water-quality assessment. To address this need, an approach is developed to relate commercial fertilizer sales to a set of explanatory variables using spatially referenced modeling methods. Spatially referenced modeling in this study refers to statistically relating fertilizer use, estimated from commercial fertilizer sales data, to spatially referenced data on watershed attributes. Separate models for nitrogen and phosphorus are developed to estimate elemental fertilizer use on agricultural lands for the conterminous United States at the National Hydrography Dataset Plus (NHDPlus) catchment scale for the year 2012. The approach builds on earlier efforts that use Association of American Plant Food Control Officials data on fertilizer sales to provide county-level estimates of nitrogen and phosphorus fertilizer use. The spatially referenced method improves on these efforts by allowing for varying nitrogen to phosphorus ratios at the catchment scale and expanding the set of variables used to allocate county-level sales data to the catchment scale. The models include catchment-level factors that are either primary determinants of fertilizer use, such as the acreage of different crop types, or measures reflecting the intensity of use, such as climate. Explanatory variables available only at the county scale, such as U.S. Department of Agriculture Census of Agriculture estimates of fertilizer expenditures, are included to improve the model predictions of elemental use. The nitrogen and phosphorus models explain more than 90 percent of the variation in elemental use at the state level, and the statistical approach allows for the estimation of uncertainty of predicted use in each catchment. The spatial patterns of model predictions reflect known agricultural cropping practices across the United States that transcend political boundaries, despite the county/state orientation of the fertilizer sales information. The results are expected to be useful for a variety of water-quality assessments that are intended to estimate nitrogen and phosphorus loads to streams.

Introduction

Fertilizer is a principal source of nutrient contamination in U.S. streams and rivers, leading to eutrophic conditions that can affect the ecological health of aquatic environments, including the streams and estuaries into which they discharge (Alexander and others, 2008). To better understand the ecological consequences of this source of nutrients, models of water quality are created that relate the locational intensity of the fertilizer source to water-quality conditions at specific monitoring locations in the stream network. The identification of spatial referencing as an important feature for improving model accuracy (Smith and others, 1997) demonstrates a need to incorporate into these models highly spatially resolved fertilizer use information.

A principal source of information on fertilizer use is contained in the yearly estimates of county and state fertilizer sales, distinguished by elemental composition, produced by the Association of American Plant Food Control Officials (AAPFCO) (Association of American Plant Food Control Officials, 2015). Use of these data presents several complications. First, the data pertain to sales rather than use, with the understanding that sales made at a given location are commonly transported across county and state borders to the location of use. Second, the data are not uniformly available at a common scale, with many states, principally in the western United States, having only state-level information.

Third, most states reporting county sales data have some sales that are not assigned to any county; in some states, this class of sales is large relative to total state sales. Fourth, to make the data useful in a water-quality model, applying additional methods is necessary to allocate state- or county-level fertilizer use to a much finer spatial scale. In a study by Ruddy and others (2006), modified by Gronberg and Spahr (2012), a method was developed to use the AAPFCO annual fertilizer sales data in combination with U.S. Department of Agriculture (USDA) Census of Agriculture (COA) fertilizer expenditure data (U.S. Department of Agriculture, National Agricultural Statistics Service, 2004) to obtain county-level estimates of fertilizer use for nitrogen and phosphorus elemental components. This approach employed the following two fundamental assumptions: (1) at the state scale, fertilizer sales are equivalent to use, and (2) the ratio of nitrogen to phosphorus is constant across all counties in a state. Further allocation of the Gronberg and Spahr (2012) county estimates to individual catchments has been achieved using 30-meter land use information contained in the 2006 National Land Cover Dataset (Wieczorek and others, 2018; Fry and others, 2011). The estimates from Ruddy and others (2006) have been determined to be empirically important in total nitrogen and total phosphorus water-quality models (Alexander and others, 2008).

The method described in this study uses spatially referenced statistical modeling methods to analyze the AAPFCO sales data to produce estimates of use at the catchment scale for the year 2012. Spatially referenced modeling refers to statistically relating observations, in this case, fertilizer sales, to spatially referenced data on watershed attributes (Smith and others, 1997). In so doing, the method directly relates the scale at which fertilizer data are compiled to the scale at which these data are employed in water-quality models. Multiple sources of ancillary information are incorporated into the analysis, including the USDA COA county fertilizer expenditure data, detailed 30-meter cropping pattern information, climate, and agricultural practices information (LaMotte, 2015; U.S. Department of Agriculture, National Agricultural Statistics Service, 2013; Wieczorek and others, 2018; Gronberg and Arnold, 2017a, Gronberg and Arnold, 2017b). A useful feature of the statistical approach is that it is possible to assess the uncertainty of the estimates, and to make use of observed fertilizer sales at larger scales, considered concordant with use, to improve the estimates of use at the smaller catchment scale, by relating fertilizer sales to a set of explanatory variables. For this study, catchment refers to the spatial unit defined by the local drainage area to each individual stream segment as defined by the National Hydrography Dataset Plus version 2 (NHDPlusV2) dataset (McKay and others, 2012; Moore and Dewald, 2016).

As with the approach by Ruddy and others (2006), the encompassing assumption used to implement the spatially referenced modeling approach is that at larger scales (such as states) fertilizer sales estimates reflect actual fertilizer application to the land surface, what the authors refer to as fertilizer “use”. Under this assumption and with the proviso that the model is specified to reflect factors affecting use rather than factors affecting sales, the estimation of the model with fertilizer sales data will yield a predictive model for use. In accordance with this understanding, the determinants of use are based primarily on the spatial extent of various crop types within a given region. The spatial extent and pattern of where fertilizer application, “use”, occurs on the land surface is determined by the crops grown on that land and information on cropping patterns. Thus, the specific crop type and area within a catchment is used in this study to define the spatial extent of fertilizer use and the type and amount of fertilizer applied within a catchment. These crop variables are referred to as extensive factors in this report. Crop nutrient requirements also vary based on geographically specific natural and anthropogenic factors that may influence the amount and intensity of fertilizer application. These factors may include fertilizer expenditures in the region, the prevalence of animal manure, and other climate conditions and are referred to as intensive factors in this study. In forming these relations, spatial referencing implies the interaction of the intensive factors, factors that do not depend on spatial extent (such as climate) with the extensive factors (such as cropland), is at the highest possible spatial resolution of the model—the intersection of individual stream segment catchment areas and counties.

Two approaches to obtaining predictions are taken. First, is the unconditional approach where fertilizer use at the catchment scale is predicted based solely on the explanatory variables employed in the statistical analysis, with a correction for retransformation bias because of model error. Second, is the conditional approach where the observed fertilizer sales at the state level are apportioned to individual catchments based on the catchment’s share of state-level use as determined from the unconditional predictions. The final model predictions reflect relatively consistent elemental composition use patterns that transcend political boundaries, despite the county/state orientation of the fertilizer information. Because the method does not assume a fixed relation between nitrogen and phosphorus use, models for each being derived separately, predictions of the ratio of nitrogen to phosphorus use reproduce recognized patterns in farming practices.

Model results are evaluated in four ways. First, 2012 predictions derived from these models are compared to those generated by applying the Gronberg and Spahr (2012) method for the same 2012 period (Brakebill and Gronberg, 2017). Second, model predictions are compared to 2012 county-level estimates of nitrogen and phosphorus input, generated from AAPFCO fertilizer sales data by the International Plant Nutrition Institute (IPNI) (International Plant Nutrition Institute, 2012). Third, predictions are compared to USDA Economic Research Service survey data of reported nitrogen and phosphorus fertilizer use on corn and soybeans (U.S. Department of Agriculture, Economic Research Service, 2016) in selected states. Fourth, model predictions are compared to 2012 estimates of fertilizer use for Mississippi, generated from USDA, Farm Service Agency 2012 crop acreage data and Mississippi Extension fertilizer recommendations (U.S. Department of Agriculture, Farm Service Agency, 2013; Oldham, 2012).

Purpose and Scope

The purpose of this report is to describe the methods and subsequent results of two models developed for estimating elemental nitrogen and phosphorus commercial fertilizer use on agricultural lands for the conterminous United States at the catchment scale for the year 2012. The overall objective of this approach is to improve upon techniques currently used to estimate commercial fertilizer use on agricultural lands, including the estimated ratio of nitrogen to phosphorus applied to these lands (Gronberg and Spahr, 2012; Brakebill and Gronberg, 2017). The approach relates reported fertilizer sales to a set of explanatory variables using spatially referenced statistical modeling methods. The estimation of the nitrogen and phosphorus fertilizer-use models are described, and maps showing the predicted use at the catchment scale are presented. As validation of the approach, the nitrogen and phosphorus use predictions are compared to other estimates of fertilizer use at county and state scales.

Methods

Approach

Information on fertilizer use, the mass of nutrients (nitrogen or phosphorus) applied to agricultural crops across the conterminous United States, is largely unavailable at a scale needed for water-quality assessment. Previous studies were completed to address this issue and have determined that county-level estimates of fertilizer use (nitrogen and phosphorus) can be obtained, using existing datasets of county-level sales and county-level fertilizer expenditures (Ruddy and others, 2006; Gronberg and Spahr, 2012). In those studies, county-level fertilizer expenditure information (the annual dollar amount expended on commercial fertilizer products) was used to allocate county-level fertilizer sales (annual tonnage of elemental nitrogen and phosphorus fertilizer sold) (Association of American Plant Food Control Officials, 2015) to individual counties, with the assumption that at the state scale, fertilizer sales are equivalent to use.

The method described in this study builds upon the earlier studies, with similar assumptions, but uses spatially referenced statistical modeling methods to relate fertilizer sales to a set of explanatory variables (such as crop type, crop acreage, fertilizer expenditures, climate factors, and agricultural practices) that are primary determinants of where and how much fertilizer mass is being used across the conterminous United States. The authors of this report propose (1) that the explanatory information will help allocate county-level fertilizer sales to the catchment level (despite exports of fertilizer mass across county lines) and (2) will provide a basis for fertilizer use estimates to vary spatially across crop types (extensive factors) because of varying nutrient requirements and natural and anthropogenic conditions (intensity factors) that may affect the amount of fertilizer application. The spatially referenced modeling method also provides a means to allocate fertilizer use at finer scale (catchment/county units) than earlier county-level studies, allows for variation of the nitrogen phosphorus ratio by catchment, and provides a means to quantify statistical significance and uncertainty in the estimates.

The fundamental spatial unit for modeling is the catchment/county (the intersection of NHDPlusV2 catchment and county boundaries) by which all explanatory variables are calculated from catchment-level and county-level datasets (table 1) (Moore and Dewald, 2016; LaMotte, 2016). The approach assumes that sales at some aggregate scale (such as state scale) approximate fertilizer use at that scale, and the model is specified with predictor variables that are assumed to be independent of net fertilizer exports. Under these assumptions, a fertilizer sales model estimated with county-level data will predict fertilizer use.

Table 1. Model input variables for estimating 2012 nitrogen and phosphorus fertilizer-use models.

[kg/yr, kilogram per year; PRISM, Parameter-elevation Regression on Independent Slopes Model; NHDPlus, National Hydrography Dataset Plus; mm/yr, millimeter per year; m^2, square meter; NA, not applicable]

Variable name	Variable type	Description	Units	Source citation
		Variable category—Agricultural practices		
Nitrogen from manure	Intensive	2012 mean annual nitrogen from animal manure. Computed from county estimates of animal populations.	kg/yr	Gronberg and Arnold, 2017
Phosphorus from manure	Intensive	2012 mean annual phosphorus from animal manure. Computed from county estimates of animal populations.	kg/yr	Gronberg and Arnold, 2017
		Variable category—Climate		
Precipitation	Intensive	2012 average annual precipitation in mm/yr based on PRISM attributed to NHDPlus version 2.1 reach catchments.	mm/yr	Wieczorek and others, 2018
Actual evapotranspiration	Intensive	Average annual actual evapotranspiration (2000–14) in mm/yr attributed to NHDPlus version 2.1 reach catchments.	mm/yr	Wieczorek and others, 2018
		Variable category—Cropland		
Crop group	Extensive	Area of crops grown based on the 2012 Cropland Data Layer, aggregated into five crop groups (appendix 2).	m^2	U.S. Department of Agriculture, National Agricultural Statistics Service , 2013
		Variable category—Fertilizer		
County farm fertilizer expenditures	Intensive	U.S. Department of Agriculture Census of Agriculture 2012 county fertilizer expenditures. County estimates of dollar amounts spent on farm fertilizer products; replacement values for 2012 missing data from 2009 and 2002 reported expenditures.	dollar/year	LaMotte, 2015
Fertilizer price index	NA	U.S. Department of Agriculture Economic Research Service national price index used to normalize fertilizer expenditures from different years to an equivalent 1992 expenditure.	1992 = 100	U.S. Department of Agriculture, Economic Research Service, 2016a
County fertilizer sales	Dependent	Association of American Plant Food Control Officials 2012 annual fertilizer sales data. Farm fertilizer products reported sold by nutrient content. Reported by individual states, in most cases at the county level.	kg/yr	Association of American Plant Food Control Officials, 2015

Spatial Framework

The NHDPlusV2 (McKay and others, 2012) is a digital network of streams developed at 1:100000 scale with associated catchments. Catchments are polygons defined by 30-meter digital elevation data that represent areas draining each stream segment (Moore and Dewald, 2016). The fundamental spatial unit of the modeling approach, which is the unit by which all explanatory data are defined, is the spatial intersection of the NHDPlusV2 catchment boundaries and county boundary polygons (LaMotte, 2016), referred to as catchment/county units (Moore and Dewald, 2016; LaMotte, 2016). Numerical values representing each specific model variable are assigned to these units. This unit of analysis permits aggregation to either counties—the smallest regional unit for fertilizer sales data or catchments—the fundamental spatial unit for water-quality models. Approximately 2.65 million NHDPlusV2 catchments are across the conterminous United States, which upon intersection with counties results in nearly 2.88 million catchment/county units (fig. 1).

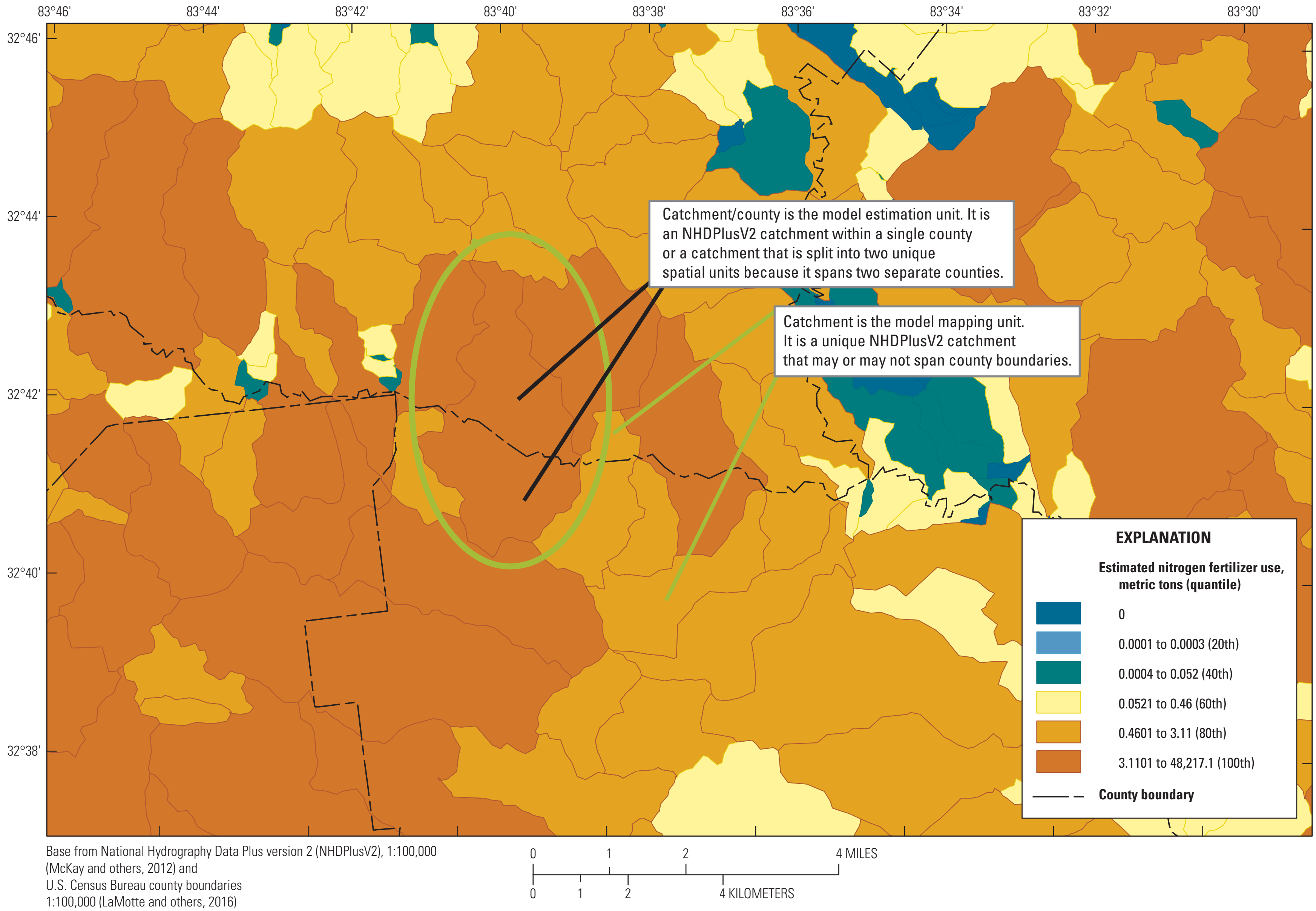


Figure 1. Catchment/county spatial unit used for estimating 2012 nitrogen and phosphorus fertilizer-use models.

Source Data

Development of the nitrogen and phosphorus fertilizer models utilizes six input datasets including (1) state reported commercial farm-fertilizer sales data, (2) reported commercial farm-fertilizer expenditure data, (3) crop-specific land cover, (4) mean annual precipitation, (5) evapotranspiration, and (6) nitrogen and phosphorus estimates from animal manure. The source data used in this study are listed in table 1. Variables serving as intensive variables in the model are logarithm transformed. To accommodate this transformation, any zero values for these variables are substituted with the corresponding minimum, nonzero value for that same variable. The dataset used to estimate catchment-level nitrogen and phosphorus fertilizer-use models is available for download as a U.S. Geological Survey (USGS) data release from the USGS ScienceBase Catalog (Stewart and others, 2019a).

Fertilizer Sales

The principal source of fertilizer information is the AAPFCO commercial fertilizer sales data for 2012 (Association of American Plant Food Control Officials, 2015). These data are available for purchase, by elemental composition, for all states in the conterminous United States and by county for many states. During our analysis, the data for Wyoming consistently generated anomalous results and are deemed unreliable and excluded from model estimation.

All fertilizer sold in a year is assumed to be applied in that same year. Negative sales in the data, which indicate fertilizer returns and account for less than 1 percent of total nitrogen or phosphorus elemental sales, are set to zero. Fertilizer sales data are converted from metric tons of product to metric tons of elemental nitrogen and phosphorus based on chemical composition data reported for each product. Each state reports the expected use of individual fertilizer products as being for farm or nonfarm use; only products identified as being for farm use are used in this study. Most states with county-level sales information include a component of sales that is not assigned to any county, with the share of this component to total state sales displaying considerable variation across states.

To address this complication, the fertilizer sales data are adjusted as follows, prior to analysis. For states with more than 20 percent of elemental mass nitrogen plus phosphorus total sales of unknown county origin (coded as 998), the county attributed sales data are deemed unreliable, and all county-level sales numbers, for nitrogen and phosphorus, are aggregated to form a state-level reporting-unit sales estimate (figs. 2–3, respectively). This criterion affected the aggregation of county-level information for seven states. For the remaining states (those with less than 20 percent of total sales unassigned to any county), sales data for the county-level reporting unit is retained. In the case of counties with no reported sales, the zero-sales county data are aggregated into a single multicounty reporting unit within the state. If the state had unassigned sales, those sales are distributed to the reporting counties and multicounty units according to each unit's share of cropland and USDA COA fertilizer expenditure (U.S. Department of Agriculture, National Agricultural Statistics Service, 2013; LaMotte, 2015) using a signal extraction method described in appendix 1. Counties (or multicounty) with no sales after allocation of county-unreported sales are excluded from model estimation.

Fertilizer Expenditures

Fertilizer expenditures, reported by the USDA COA (LaMotte, 2015), present the dollar amount expended on commercial farm fertilizer products and soil conditioners by county in the year 2012. For 43 counties where census reporting requirements prevent the reporting of 2012 expenditures, the most recent year of expenditures prior to 2012 are used (34 counties use 2007 expenditures and 9 counties use 2002 expenditures). To ensure comparability of expenditure estimates across multiple years, all expenditures are normalized to a common base year, 1992, using the fertilizer price index (U.S. Department of Agriculture, Economic Research Service, 2016). An additional 52 counties contain little cropland and report no expenditures for any of the agricultural census, in which case expenditures are set to zero.

Cropland Data Layer

The spatial extent and distribution of crops differ across the United States, as do their nutrient (such as nitrogen and phosphorus) requirements. Understanding the location and spatial extent of crop types is important for understanding the amount and type of fertilizer application on those lands. The Cropland Data Layer (CDL) 2012 (fig. 4), maps parcels of land at 30-meter resolution classified into 1 of 108 different cultivated crop types, using a decision tree classifier in conjunction with medium resolution satellite imagery from a variety of sensors. These data are spatially intersected with the boundaries of the catchment/county data layer to determine the area of each crop grown within each catchment/county unit in the conterminous United States (U.S. Department of Agriculture, National Agricultural Statistics Service, 2013;

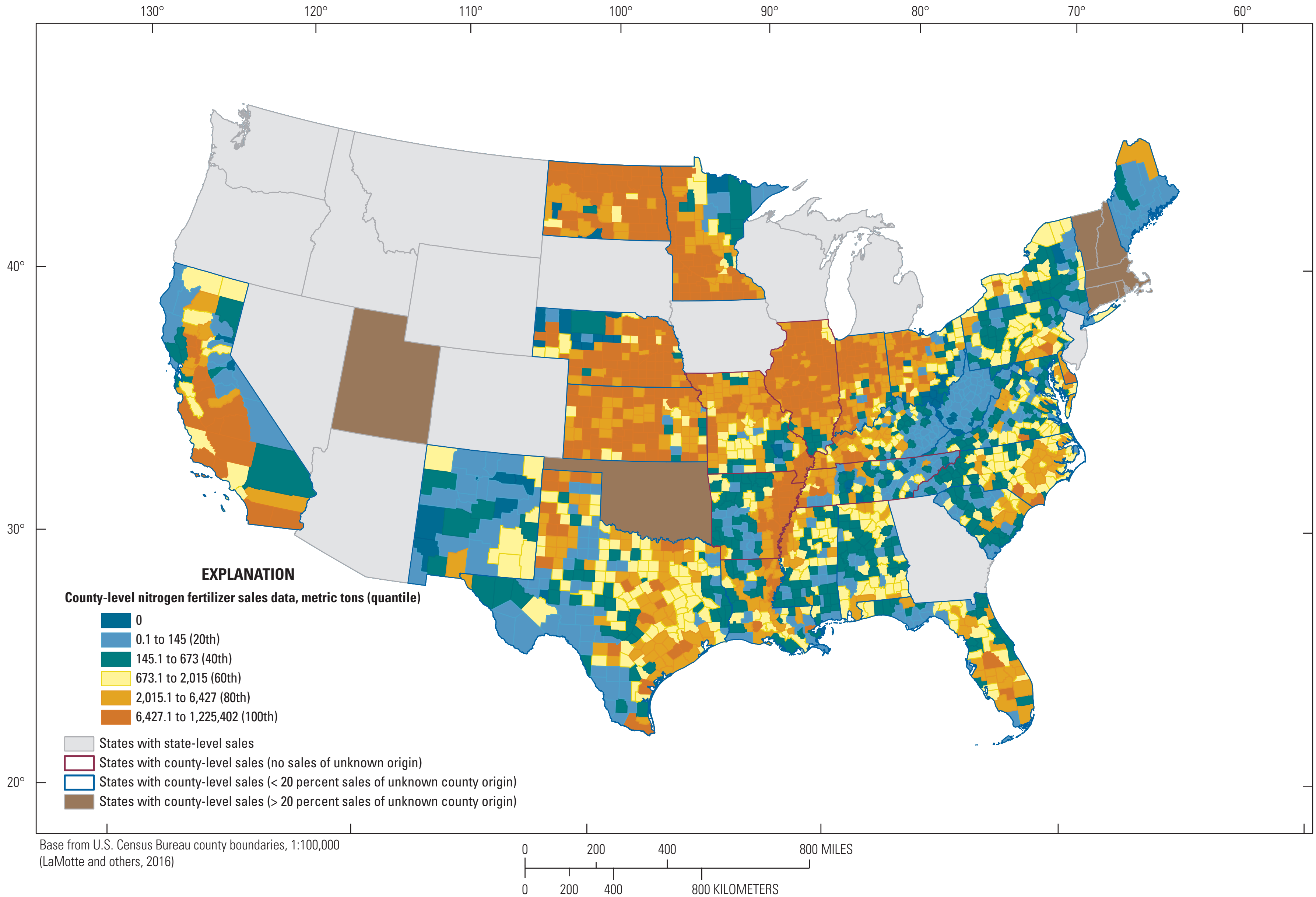
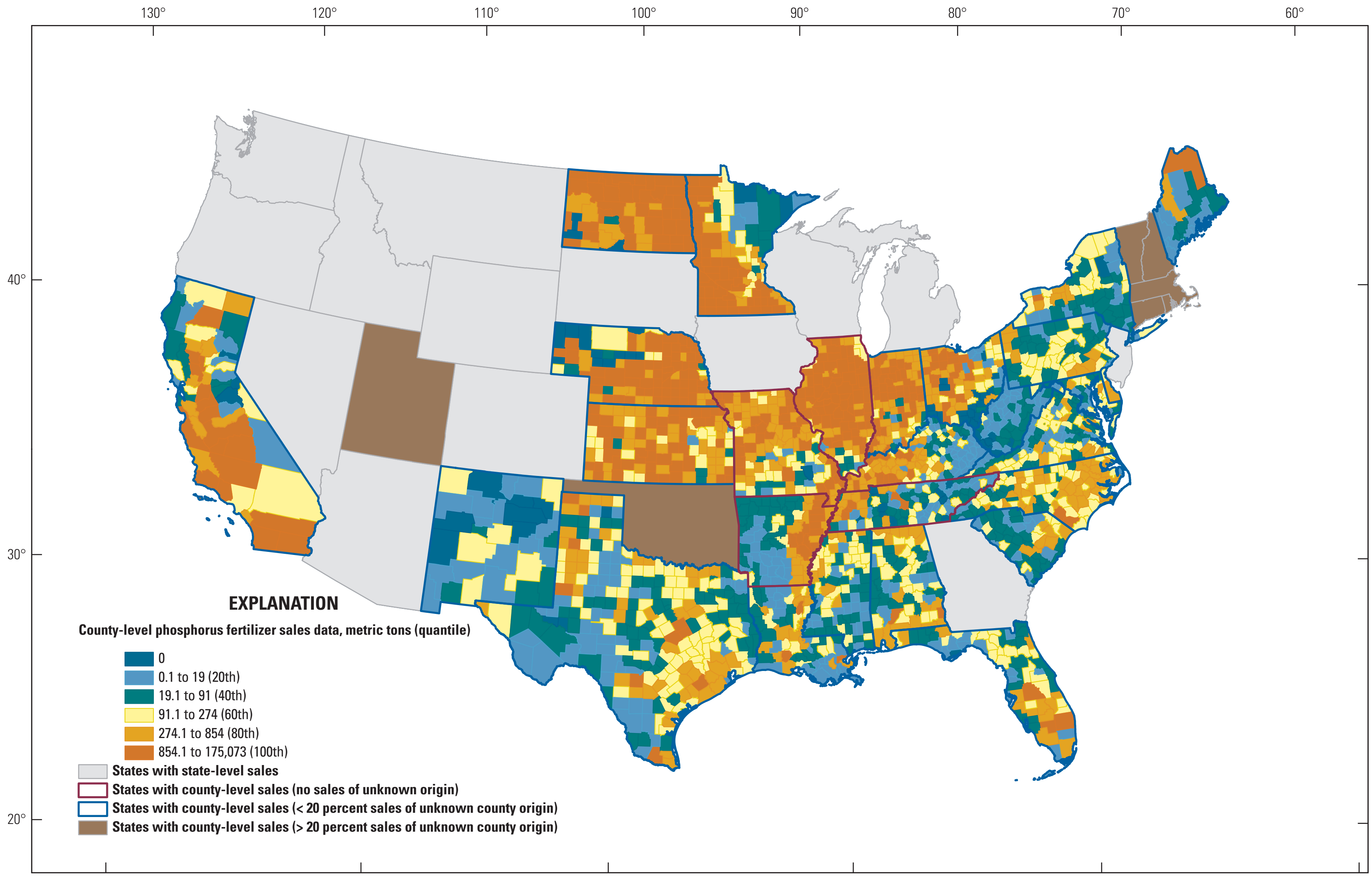


Figure 2. County-level 2012 nitrogen fertilizer sales data used to estimate nitrogen fertilizer use (Association of American Plant Food Control Officials, 2015).



Base from U.S. Census Bureau county boundaries, 1:100,000
(LaMotte and others, 2016)

0 200 400 800 MILES
0 200 400 800 KILOMETERS

Figure 3. County-level 2012 phosphorus fertilizer sales data used to estimate phosphorus fertilizer use (Association of American Plant Food Control Officials, 2015).

Wieczorek and others, 2018). Crops of similar types or with similar characteristics are grouped into the following five crop groups: (1) corn (corn, sweetcorn, popcorn, and double crops with corn), (2) grass-pasture-hay (grassland, pastureland, and nonalfalfa hay), (3) miscellaneous (81 fruit, vegetable, and other miscellaneous agricultural crops), (4) nitrogen fixing (soybeans and nitrogen-fixing cover crops of alfalfa, clover, and vetch), and (5) small grains (barley, durum wheat, millet, oats, rye, sorghum, speltz, spring wheat, triticale, winter wheat, and small grain double crops). Nitrogen-fixing chick peas, dry beans, lentils, peas, and peanuts were included with other miscellaneous crops and accounted for less than 1 percent of all cropland. Fallow cropland was not included in the groupings and accounted for less than 4 percent of total cropland. Specific groupings of crop types are described in table 2.1, and the prevalence of the crop groups across the conterminous United States is shown in figure 4. The groupings are sufficiently broad that errors in classification among the groupings are relatively minor, with overall classification accuracy among agricultural land of nearly 94 *percent*, as based on the CDL error matrices provided with the data (table 3.1). Evidence indicates, however, that the CDL classification is not highly accurate in differentiating agriculture from nonagriculture, particularly because of commission and omission errors associated with the classification of grass-pasture-hay (table 3.1).

Climate

Climate (such as precipitation, temperature, evapotranspiration) plays an important role in crop production influencing plant growth, nutrient uptake, soil microorganisms, and other factors, thereby influencing the type and amount of commercial fertilizer needed for crops. Two climatic data layers are used in the models as intensive variables—2012 mean annual precipitation and a long-term average of actual evapotranspiration (2000–14). Precipitation is derived from monthly rainfall estimates at various locations throughout the country, generalized to the land surface, and averaged across all months as one of the products of the Parameter-elevation Regression on Independent Slopes Model (PRISM) dataset (McCabe and Wolock, 2011; Wieczorek and others, 2018; Wolock and McCabe, 2018). Average annual actual evapotranspiration, in millimeters per year, is based on monthly estimates for 4-kilometer grid cells, derived from a water balance model (McCabe and Wolock, 2011; Wolock and McCabe, 2018), averaged during 2000–14, and apportioned to individual catchments (Wieczorek and others, 2018).

Nitrogen and Phosphorus From Manure

Animal manure is a recognized source of nitrogen and phosphorus and often utilized as a fertilizer source for pastureland, cropland, and hay production, thereby reducing the need for commercial fertilizer on some agricultural lands. County estimates of nitrogen and phosphorus from farm animal manure are derived from county animal population inventories compiled from the 2012 COA (LaMotte, 2015) and expected nutrient content in manure from each animal type (Ruddy and others, 2006). County-level farm animal population inventories are organized by animal categories of cattle, hogs, poultry, and other animals. The populations are multiplied by animal-specific estimates of nitrogen and phosphorus content in manure and summed by element to estimate total nitrogen and phosphorus from manure (Gronberg and Arnold, 2017a, Gronberg and Arnold, 2017b). County nutrient estimates are normalized by land area and apportioned to catchments based on the area of each county within a given catchment.

Estimation Methodology

The conceptual model of elemental fertilizer use is based on its observed relation with factors representing the extensivity and intensivity of use. In the model, the extensivity of use within a catchment/county is related to the land areas of a broad set of crop types that approximately scale with the extent of that type within each spatial unit. The intensity of use for each crop type within a spatial unit is related to several ancillary factors such as climate, the availability of commercial fertilizer substitutes like manure, and observed expenditures for total fertilizer. Relations between use and the crop type and intensity variables are quantified by a set of nonlinear model coefficients that must be statistically estimated. The coefficients are estimated through the application of nonlinear least-squares methods to minimize the sum of squared differences between observed county or state elemental sales and the predicted elemental use in each catchment/county, summed to the same level of aggregation as the sales observation (county or state). The model is spatially referenced because the interaction of variables explaining fertilizer use occurs at the fine scale of a catchment/county, whereas the data representing the dependent variable are at the coarser scale of a county or state. Differences between actual fertilizer use and fertilizer sales within a county or state, because of import or export, are quantified as uncertainty. Under specific statistical assumptions regarding these net exports, predictions from the model can be expected to correspond to elemental use, despite estimation with elemental sales data, and estimates can be made regarding the uncertainty of use.

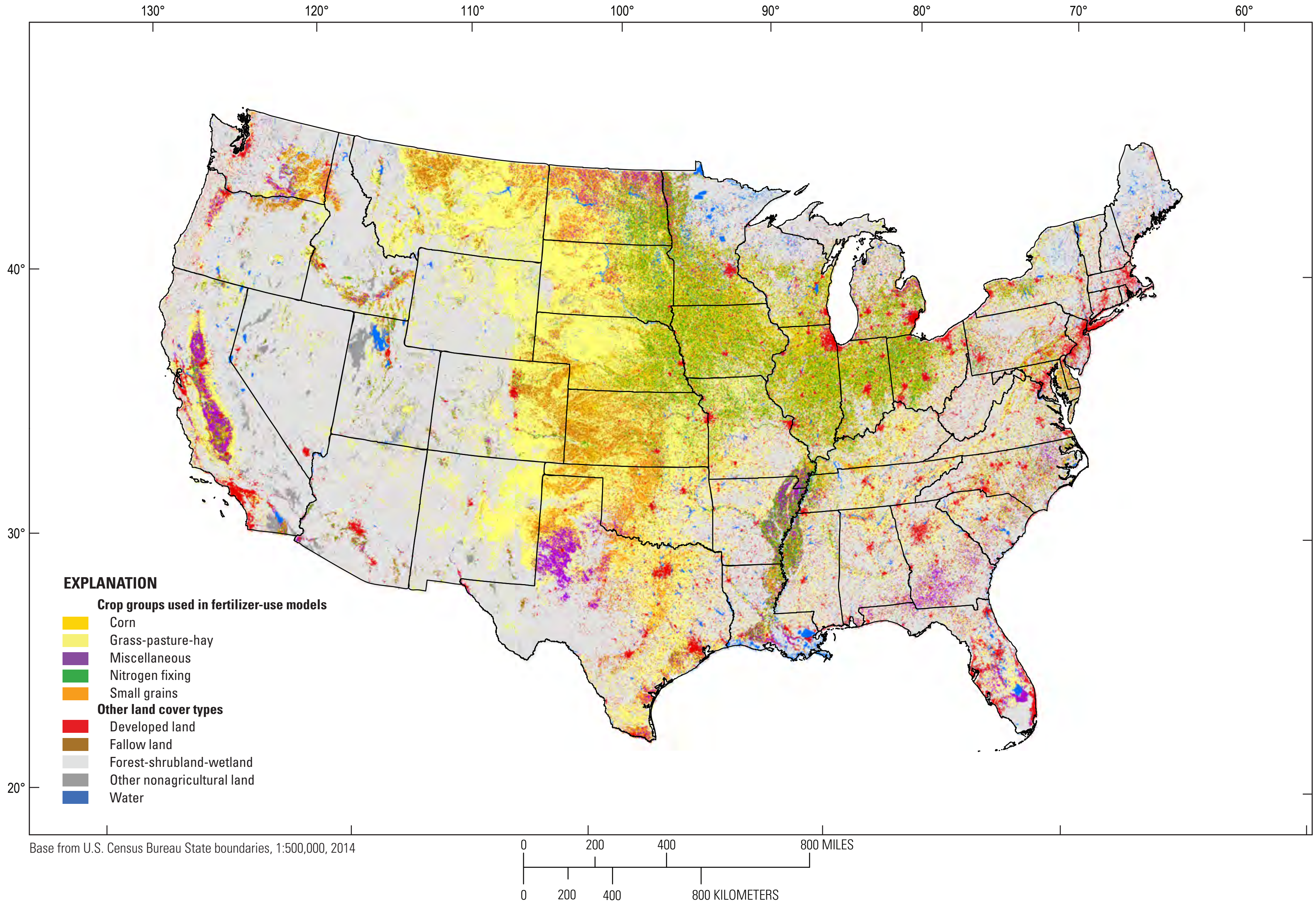


Figure 4. Cropland Data Layer 2012 crop groups used for estimating 2012 nitrogen and phosphorus fertilizer-use models.

Elemental fertilizer (either nitrogen or phosphorus) applied to a given crop-group k in catchment/county unit j, denoted W_{jk}, is assumed to be proportional to the area of the crop group, A_{jk}, with the rate of application specified to depend on a vector of spatially varying intensity variables, denoted $\boldsymbol{Z}_{jk}$, according to

$$W_{jk} = \alpha_k A_{jk} \exp\left(\boldsymbol{\beta}_k' \boldsymbol{Z}_{jk}\right) \tag{1}$$

where

- α_k is a crop-group-specific application factor;
- A_{jk} is the area of the crop group;
- $\boldsymbol{\beta}_k$ is a crop-group-specific vector of coefficients, having the same number of elements as the vector $\boldsymbol{Z}_{jk}$, which mediate the effects of the intensity variables; and
- $\boldsymbol{Z}_{jk}$ is the vector of spatially varying intensity variables.

To improve the interpretability of the crop-group-specific application factors, the intensity variables are specified as deviations from their mean values. If the intensity variables are approximately normally distributed, then their transformation as deviations from a mean imply the application factor can be interpreted as the approximate median mass per unit area of elemental fertilizer applied to the given crop group.

Let F_i^u represent the elemental fertilizer use for spatial reporting unit i, either a state, county, or multicounty reporting unit. Let J_i represent the set of catchment/counties, j, that comprise reporting unit i, and let there be K identified crop groups, the collection of which comprises all agricultural area. The formulated model for reporting unit i is

$$\ln F_i^u = \ln\left(\sum\nolimits_{j \in J_i} \sum\nolimits_{k=1}^{K} \alpha_k A_{jk} \exp\left(\boldsymbol{\beta}_k' \boldsymbol{Z}_{jk}\right)\right) + u_i \tag{2}$$

where

- J_i is the set of catchment/counties, j, that comprise reporting unit i;
- K is the number of crop groups that comprise all agricultural area;
- α_k is a crop-group-specific application factor;
- A_{jk} is the area of the crop group;
- $\boldsymbol{\beta}_k$ is a crop-group-specific vector of coefficients, having the same number of elements as the vector $\boldsymbol{Z}_{jk}$, which mediate the effects of the intensity variables;
- $\boldsymbol{Z}_{jk}$ is the vector of spatially varying intensity variables; and
- u_i is a model residual, possibly spatially correlated, having a mean of zero and a heteroscedastic variance $\sigma_{u_i}^2$.

The model described by equation 2 adheres to mass-balance constraints in that total fertilizer application for a reporting unit is simply a summation of the application at each fundamental spatial unit, j, and crop group, k. Moreover, fertilizer application scales directly with agricultural area in that a doubling of all A_{jk} causes a doubling of F_i^u. The model formulation can be characterized as spatially referenced because the interaction of the intensity variables with the crop areas occurs at the smallest spatial scale of the model, the catchment/county scale.

The elemental composition of fertilizer use is unknown; therefore, the model in equation 2, is impossible to estimate directly. The available data for model estimation relate to the elemental composition of fertilizer sales, F_i^s, with the relation between sales and use described by

$$\ln F_i^s = \ln F_i^u + x_i \tag{3}$$

where

- F_i^u is the logarithm of elemental fertilizer use for spatial reporting unit t, either a state, county, or multicounty reporting unit; and
- x_i is the logarithm of a proportional net-export factor, assumed to have a mean of zero and heteroscedastic variance, σ_{xi}^2, whereby if sales in reporting unit i exceed use then x_i is positive, and is negative otherwise.

An assumption of the analysis is that the variance of x_i goes to zero at the state level of aggregation. This assumption is consistent with the assumption employed in other studies (Ruddy and others, 2006; International Plant Nutrition Institute, 2012) that at the state scale, fertilizer sales approximate use.

If F_i^u is independent of x_i, then estimation of equation 2 using F_i^s instead of F_i^u does not lead to inconsistency in the estimated coefficients, although for smaller reporting units the residuals will tend to have a variance that exceeds σ_i^2 because of the additional variability induced by the variability of η_i. Because net exports are assumed to go to zero, the variance of residuals computed at the state level should better approximate σ_i^2. Similarly, if the CDL data contain classification error, then this error could also cause the variability of county-based residuals to overstate the variance of the fertilizer use residual. If the spatial correlation of this error is not too great, the error can be assumed to diminish with reporting-unit size, making the residual variance computed at larger scales essentially free of this bias.

The residual obtained in model estimation, denoted e_i, is a combination of the use residual, u_i, and the net-export residual, x_i,

$$e_i = u_i + x_i \tag{4}$$

where

u_i is the use residual for reporting unit i; and
x_i is the net-export residual.

The analysis allows for spatial correlation of the residual terms, u_i and x_i, according to multicomponent processes. Without significant loss of generality, the fundamental spatial unit for correlation is assumed to be the county, indexed by c. The county use residual, u_c, is assumed to consist of two components—a component that is common to all use residuals in the same state, denoted $\eta_{s(c)}$, where $s(c)$ refers to the state containing county c and a local component, ϵ_c. The state component is assumed to be independent across states, with a zero mean and variance of σ^2_η. The local component has a zero mean and variance of σ^2_ϵ and has a correlation of ρ_ϵ with neighboring counties within the same state, but is independent across nonneighboring counties.

A complication in empirically estimating the spatial correlation structure of the use residual is that the residual can only be observed at the state level; residuals at the county level are obscured by the net-export residual. However, it is possible to infer the correlation structure using only the state-level residuals. Fertilizer use for state s, F^u_s, is given by the sum of county use across all counties, $C(s)$, in the state. In the logarithm space used for model estimation,

$$\ln F^u_s = \ln\left(\sum\nolimits_{c\epsilon C(s)} F^u_c e^{u_c}\right) \approx \ln\left(\sum\nolimits_{c\epsilon C(s)} F^u_c\right) + \sum\nolimits_{c\epsilon C(s)} w_c u_c \tag{5}$$

where

$C(s)$ is the set of all counties in state s;
F^u_c is the fertilizer use for county c;
e^{u_c} is the estimated residual for county c;
w_c is the county-c share of state-s elemental fertilizer use ($w_c = F^u_c/F^u_s$); and
u_c is the use residual for county c.

The approximation in equation 5 is based on a first-order Taylor series expansion of $\ln F^u_s$, with respect to u_c evaluated at zero, for all $c \in C(s)$. Thus, the state-level model residual, e_s, which is assumed to reflect the state-level residual for use, is given by

$$e_s = \sum\nolimits_{c\in C(s)} w_c u_c = \eta_s + \sum\nolimits_{c\in C(s)} w_c \varepsilon_c \tag{6}$$

If the state residual is squared, the expectation is given by

$$E\left[e^2_s\right] = \sigma^2_\eta + \sigma^2_\varepsilon \sum\nolimits_{c\in C(s)} w^2_c + \sigma^2_\varepsilon \rho_\varepsilon \sum\nolimits_{c\in C(s)} w_c \sum\nolimits_{d\in D(c)} w_d \tag{7}$$

where

$C(s)$ is the set of all counties in state s;
w_c is the county-c share of state-s elemental fertilizer use ($w_c = F^u_c/F^u_s$);
$D(c)$ is the set of all counties, indexed by d, that are in the same state and share a border with county c; and
w_d is the share of state-elemental sales for a set of counties that share a border with county c.

In the empirical analysis, the set of county neighbors is based on the 1:2,000,000-scale county spatial coverage (Lanfear, 1994).

Equation 7 can be estimated for the parameters $\{\sigma^2_\eta, \sigma^2_\epsilon, \rho_\epsilon\}$ using nonlinear least squares applied to the regression of the squared state residuals on state values of the two model-estimated sums, $\sum_{c\epsilon C(s)} w^2_c$ and $\sum_{c\epsilon C(s)} w_c \sum_{d\epsilon D(s)} w_d$, with an intercept (which serves as the estimate of σ^2_η, where the nonlinear regression is constrained to have nonnegative values for the variance terms (σ^2_η and σ^2_ϵ) and the correlation term, ρ_ϵ, is constrained to be between zero and one.

The county net-export residual, x_c, has no state component because net exports are assumed to be zero at the state level of aggregation. The county component of the net-export residual has a variance σ^2_x and is assumed to be correlated with the nearest neighbor, with correlation coefficient ρ_x.

The model described in equation 2 is estimated using nonlinear least squares to obtain estimates of the coefficients $\boldsymbol{\theta} = (\alpha_k, \boldsymbol{\beta}_k)$, $k = 1, \ldots, K$. In estimation, the crop-group parameters, α_k, are constrained to be nonnegative, causing the statistical significance of these parameters to be evaluated using one-sided significance tests. Although the $\boldsymbol{\beta}_k$ can vary with crop group, for most intensive variables the coefficients are restricted to be equal across all groups, with one notable exception discussed in the "Results" section.

The existence of spatial correlation in the county residuals complicates the determination of the estimated coefficient covariance matrix. Generally, spatial correlation implies the standard estimates of coefficient variances are downward biased. To correct for this bias, an alternative estimator of the covariance matrix is adopted. The estimator accounts for covariance in the residuals and is asymptotically consistent (Amemiya, 1985). With observations ordered by state, the alternative covariance matrix, $V[\hat{\boldsymbol{\theta}}]$, takes the form

$$V\left[\hat{\boldsymbol{\theta}}\right] = \left(\boldsymbol{G}'_{\boldsymbol{\theta}} \boldsymbol{G}_{\boldsymbol{\theta}}\right)^{-1} \boldsymbol{G}'_{\boldsymbol{\theta}} \Omega \boldsymbol{G}_{\boldsymbol{\theta}} \left(\boldsymbol{G}'_{\boldsymbol{\theta}} \boldsymbol{G}_{\boldsymbol{\theta}}\right)^{-1} \tag{8}$$

where

$\boldsymbol{G}_{\boldsymbol{\theta}}$ is the matrix of gradients of the model (the partial derivatives of the right-hand side of equation 2 with respect to the coefficient vector, $\boldsymbol{\theta}$); and
$\boldsymbol{\Omega}$ is the covariance matrix for reporting-unit residuals.

Because observations are assumed to be independent across state borders, $\mathbf{\Omega}$ is a block diagonal matrix, with state blocks, $\mathbf{\Omega}_s$, having dimension determined by the number of reporting units, n_s, in each state s. For states with no county observations, $\mathbf{\Omega}_s$ is a single element having the value given by the right-hand side of equation 7; for states with county-level reporting units, the state block takes the form

$$\mathbf{\Omega}_s = \sigma_\eta^2 \boldsymbol{i}_{n_s} \boldsymbol{i}_{n_s}' + \left(\sigma_\varepsilon^2 + \sigma_x^2\right)\mathbf{I}_{n_s} + \left(\rho_\varepsilon \sigma_\varepsilon^2 + \rho_x \sigma_x^2\right)\mathbf{M}_s \quad (9)$$

where

$\mathbf{i}_{n_s}$ is a n_s-element vector of ones;
$\mathbf{I}_{n_s}$ is the $n_s \times n_s$ identity matrix; and
$\mathbf{M}_s$ is an $n_s \times n_s$ matrix with element $\mathbf{M}_{sc_1c_2}$ equal to one if counties c_1 and c_2 are neighbors, and equal to zero otherwise.

For states with no county observations, $\mathbf{\Omega}_s$ is evaluated by the estimation of equation 7. For states with county observations, note that the variance of the county residuals is $[\sigma_\eta^2] + \sigma_\epsilon^2 + \sigma_x^2$ and the variance of the difference, $(e_{c_1} - e_{c_2})$, where counties c_1 and c_2 are neighbors, is $2(\sigma_\epsilon^2 + \sigma_x^2 - (\rho_\epsilon \sigma_\epsilon^2 + \rho_x \sigma_x^2))$. Therefore, equation 9 can be rewritten as follows:

$$\mathbf{\Omega}_s = \sigma_\eta^2\left(\boldsymbol{i}_{n_s}\boldsymbol{i}_{n_s}' - \mathbf{I}_{n_s} - \mathbf{M}_s\right) + V\left[e_c\right]\left(\mathbf{I}_{n_s} + \mathbf{M}_s\right) - V\left[e_{c_1} - e_{c_2}\right]\mathbf{M}_s / 2 \quad (10)$$

where

$\mathbf{i}_{n_s}$ is defined in equation 9;
$\mathbf{I}_{n_s}$ is defined in equation 9;
$\mathbf{M}_s$ is defined in equation 9;
$V[e_c]$ is the variance of the county residuals; and
$V[e_{c_1} - e_{c_2}]$ is the variance of the difference in neighboring county (within the same state) residuals.

Because the grouped-county observations have no definitive neighbors, these observations are excluded from the calculation of $V[e_{c_1} - e_{c_2}]$. Given the estimate of σ_η^2 from the regression in equation 7, all terms in equation 10 are estimated allowing for the corrected estimate of the coefficient covariance matrix through equation 8.

Prediction Methodology

The estimated models are used to generate two types of predictions—unconditional prediction, whereby the model is the sole basis for the predicted use, with a correction for retransformation bias and conditional prediction, whereby sales information at the state level is used to modify the unconditional predictions generated by the model. Both methods of prediction invoke the assumption that fertilizer sales and use are equivalent at the state level of aggregation. Predictions of elemental fertilizer use using the model are initially made at the finest spatial scale, the catchment/county, which are then aggregated to the catchment scale. Because the estimation of the model uses statistical methods, to generate measures of prediction uncertainty is possible for all predictions.

Unconditional Prediction

In deriving the unconditional and conditional predictions a preliminary estimate is made based on equation 1, summed for all crop types, with coefficients evaluated at their estimated values (denoted by the accent “^”) and without an accounting for the model error estimated in equation 2. Accordingly, the preliminary prediction of fertilizer use for catchment/county j, $\hat{W}_j$, is given by

$$\hat{W}_j = \sum_{k=1}^{K} \hat{\alpha}_k A_{jk} \exp\left(\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_{jk}\right) \quad (11)$$

where

K identifies crop groups, the collection of which comprises all agricultural area;
α_k is a crop-group-specific application factor;
A_{jk} is the area of the crop group;
$\boldsymbol{\beta}_k$ is a crop-group-specific vector of coefficients; and
$\boldsymbol{X}_{jk}$ is the intensity variable for catchment/county unit j, associated with crop group k.

Because nonlinear least squares (under standard assumptions) results in consistent estimates of the coefficients (the standard error of the coefficients goes to zero as the sample size goes to infinity) and because of the continuity of equation 11 with respect to the coefficients, the preliminary prediction also is consistent (Amemiya, 1985).

As indicated in equation 2, fertilizer use also depends on a residual term, u, which is additive in logarithm space and multiplicative in real space, with a distribution that is independent of scale, location, and the values of the explanatory variables. The expectation of the exponential transform of this residual is likely greater than one due to skew, implying the preliminary prediction given in equation 4 is biased downwards.

To correct this bias, the preliminary prediction is scaled by a retransformation-bias correction factor, denoted $\hat{e^u}$, representing the expectation of the exponential transform of the fertilizer use residuals. If the use residual is normally distributed, the appropriate correction factor is given by $\exp((\sigma_\eta^2 + \sigma_\epsilon^2)/2)$, the parameters of which can be consistently estimated from the state-level squared-residual, nonlinear least-squares regression described in equation 7. However, as explained in the results, to obtain a logically consistency estimate for σ_ϵ^2 is not possible.

Therefore, σ_ϵ^2 is set to zero, making the state-level residuals a direct estimate of the state-level component of the use residual, η, without the confounding variation from a possibly spatially correlated, county-level component. Under these conditions, the state residuals are homoscedastic, and a smearing estimator approach can be used to estimate the retransformation-bias correction factor. This estimate, which does not require the residuals to have a known distribution, is given by

$$\hat{e^u} = S^{-1}\sum_{s=1}^{S}\exp\left(\hat{e}_s\right) \tag{12}$$

where

S is the total number of states for which state-level residuals, $\hat{e}_s$, are available.

Given the preliminary prediction for catchment/county j, $\hat{W}_j$, and the national estimate of the retransformation-bias correction factor, $\hat{e^\epsilon}$, the unconditional prediction of fertilizer use for catchment t, $\hat{F}^u_{tu}$, is obtained by product according to

$$\hat{F}^u_{tU} = \sum_{j\in J(t)}\hat{W}_j\,\hat{e^u} \tag{13}$$

where

$J(t)$ is the set of catchment/counties contained in catchment t;

$\hat{W}_j$ is the preliminary prediction for catchment/county j; and

$\hat{e^u}$ is the retransformation-bias correction factor.

Conditional Prediction

The conditional prediction of fertilizer use, for catchment/county j derives from a slightly alternative formulation of the fertilizer-use model,

$$F_j^u = \left(\sum_{k=1}^{K}\alpha_k A_{jk}\exp\left(\beta_k'\boldsymbol{X}_{jk}\right)\right)\exp\left(\eta_{s(j)} + \varepsilon_{c(j)}\right) = W_j\left(\theta\right)m_j \tag{14}$$

where

$W_j(\theta)$ is the model component of the prediction, the estimation of which is the preliminary prediction indicated in equation 11; and

m_j is a multiplicative-residual component composed of the exponentiated sum of the state- and county-component residuals, $\eta_{s(j)}$ and $\epsilon_{c(j)}$, corresponding to the state, $s(j)$, and county, $c(j)$, in which catchment/county j resides.

The conditional prediction replaces W_j with its consistent estimate, $\hat{W}_j$, and the multiplicative factor, m_j, with its expectation given sample data, $E[m_j|\boldsymbol{F}]$, where $\boldsymbol{F}$ is a vector of the observed elemental fertilizer sales data used to estimate the model. Given the assumption that residuals are independent across states, only sales within the same state as catchment/county j are relevant, implying $E[m_j|\boldsymbol{F}] = E[m_j|\boldsymbol{F}_{s(j)}]$, where $\boldsymbol{F}_{s(j)}$ is the vector of elemental fertilizer sales for catchment/county state, $s(j)$.

In general, evaluating the expectation $E[m_j|\boldsymbol{F}_{s(j)}]$ requires specifying explicit distributions for each of the residual terms, η, ϵ, and x, these distributions reflecting the assumed spatial correlation properties described previously. One way to simplify the analysis is to restrict the evaluation to functions that are linear in the dependent variable and select the precise linear relation to meet unbiased criteria and minimize the variance of the prediction error. Such an estimator is said to be a Best Linear Unbiased Estimator (BLUE) (Amemiya, 1985).

Let $\boldsymbol{W}_s$ be an $N_s \times 1$ vector of the preliminary predictions aggregated to the N_s reporting units in state s, and define $\tilde{\mathbf{m}}_s$ as the $N_s \times 1$ aggregate sales multiplicative factor,

$$\tilde{\boldsymbol{m}}_s = D\left(\boldsymbol{W}_s\right)^{-1}\boldsymbol{F}_s \tag{15}$$

where

$D(x)$ is a diagonal matrix with vector x along the diagonal;

$\boldsymbol{W}_s$ is an $N_s \times 1$ vector of the preliminary predictions aggregated to the N_s reporting units in state s; and

$\boldsymbol{F}_{s(j)}$ is the vector of elemental fertilizer sales for catchment/county state, s.

For states with county-level reporting units, N_s is the number of counties (or counties and combined-counties), and $\tilde{\mathbf{m}}_s$ is a vector of exponentiated sums of the county- and state-level error components η_s, ϵ_c, and x_c. If state s has no county reporting units, then $\tilde{\mathbf{m}}_s$ is a scaler equal to $\exp(e_s)$, where e_s is approximately equal to the relation indicated in equation 6.

In appendix 4, the BLUE estimator for m_j, defined as $\hat{m}_j$, takes the form

$$\hat{m}_j = \mu + \Sigma_{js(j)}D\left(\boldsymbol{W}_{s(j)}\right)\left(D\left(\boldsymbol{W}_{s(j)}\right)\Sigma_{s(j)}D\left(\boldsymbol{W}_{s(j)}\right)\right)^{-1}\left(\boldsymbol{F}_{s(j)} - \tilde{\mu}\boldsymbol{W}_{s(j)}\right) \tag{16}$$

where

μ is the unconditional expectation of m_j;

$\tilde{\mu}$ is the unconditional expectation of the aggregate sales multiplicative factor, $\tilde{\mathbf{m}}_s$ (if $\tilde{\mathbf{m}}_s$ is a vector the assumptions imply it has the same unconditional expectation across all reporting units);

$\sum_{js(j)}$ is $1 \times N_{s(j)}$ a row vector of the covariance between m_j and each of the $N_{s(j)}$ reporting-unit real-space multiplicative residuals, $\tilde{\mathbf{m}}_s$;

$D(\mathbf{W}_{s(j)})$ is a $N_{s(j)} \times N_{s(j)}$ diagonal matrix with diagonal elements given by the $N_{s(j)} \times 1$ vector $\boldsymbol{W}_{s(j)}$ representing the preliminary predictions aggregated to the $N_{s(j)}$ reporting units in state $s(j)$;

$\sum_{s(j)}$ is the $N_{s(j)} \times N_{s(j)}$ covariance matrix between the state's $N_{s(j)}$ reporting unit real-space, zero-mean multiplicative residuals, $\tilde{\mathbf{m}}_s$; and

$\boldsymbol{F}_{s(j)}$ is the vector of elemental fertilizer sales for catchment/county state, $s(j)$.

Appendix 4 describes the evaluation of the covariance matrices under the assumption that the error components in logarithm space are each normally distributed. As remarked with the unconditional predictions, an acceptable variance was possible to obtain only for the state component, the county-level variance component being set to zero. This result greatly simplifies the evaluation of the BLUE estimator for m_j. Under the assumption that the use residual consists only of a state-level component and given the previous assumption that the state-level residual excludes the net-export component, then the state-level use residual is fully revealed by the state-level model residual. That is, from equation 6, if ϵ_c has no variance, then $e_s = \eta_s$ and m_j can be estimated perfectly by knowing the state value of fertilizer sales, F_s. In terms of equation 16, with $\sigma^2_\epsilon = 0$ and $\boldsymbol{F}_{s(j)}$ set to the state-level reporting unit for all states, then $\tilde{\mu} = \mu$ and $\sum_{js(j)} = \sum_{(j)}$, all vector terms are scalers, and the determination of $\hat{m}_j$ simplifies to

$$\hat{m}_j = \boldsymbol{F}_{s(j)} / \boldsymbol{W}_{s(j)} \tag{17}$$

where

$\boldsymbol{F}_{s(j)}$ is defined in equation 16; and

$\boldsymbol{W}_{s(j)}$ is defined in equation 16.

Because some catchments straddle state borders, the BLUE conditional estimate for elemental fertilizer use in catchment t, $\hat{F}^u_{tC}$, is a combination of the normalized state-level sales of the respective states comprising the catchment. Accordingly, the estimate is

$$\hat{F}^u_{tC} = \Sigma_{j \in J(t)} \hat{W}_j \boldsymbol{F}_{s(j)} / \hat{\boldsymbol{W}}_{s(j)} \tag{18}$$

where

$J(t)$ is the set of catchment/counties contained in catchment t;

$\hat{W}_j$ is the preliminary prediction of fertilizer use, for catchment/county j;

$\boldsymbol{F}_{s(j)}$ is defined in equation 16; and

$\boldsymbol{W}_{s(j)}$ is defined in equation 16.

Effectively, the BLUE conditional estimates represent an allocation of the state-level fertilizer sales according to the share of model-predicted use in each catchment/county. In this regard, the conditional estimates are like the estimates generated by Brakebill and Gronberg (2017), the difference being the method used to derive the share allocations.

Prediction Uncertainty

An advantage of the statistical approach to estimating fertilizer use is the ability to assess the uncertainty of the estimate. For the unconditional predictions, the two sources of uncertainty are as follows: (1) uncertainty because of finite random sampling, which is manifest in the estimated model coefficients, these coefficients being based on a finite sample that is randomly drawn; and (2) model uncertainty, which is the uncertainty arising from the model residual. The consistency of the estimated coefficients implies that uncertainty because of random sampling goes to zero as the sample size goes to infinity; uncertainty because of model error does not go to zero in large samples.

Because of the nonlinear specification of the model coefficients, uncertainty is assessed using a parametric bootstrap method, as described in Schwarz and others (2006). The nonlinear least-squares estimates of the coefficients and their covariance matrix as defined by equation 8 are used to randomly generate 200 sets of coefficients assuming a normal multivariate distribution. The use of a normal distribution is justified because it is the large-sample limiting distribution for nonlinear least-squares estimation (Amemiya, 1985).

Let $\hat{\theta}^{(r)}$ denote the r-th bootstrap iteration randomly generated coefficient vector and let $\hat{W}^{(r)}_j$ be the r-th iteration realization of the preliminary prediction for catchment/county j. To obtain bootstrap evaluations of the retransformation-bias correction factor, the $\hat{W}^{(r)}_j$ are aggregated to the state level, log transformed, and subtracted from log-transformed state fertilizer sales to obtain bootstrap-iteration evaluations of state-level residuals. Using the smearing estimator defined in equation 12, the residuals are used to compute bootstrap iteration estimates of the retransformation bias factor, $e^{\hat{u}(r)}$. Accordingly, the r-th bootstrap iteration unconditional estimate of fertilizer use for catchment t is

$$\hat{F}^{u(r)}_{tU} = \Sigma_{j \in J(t)} \hat{W}^{(r)}_j e^{\hat{u}(r)} \tag{19}$$

where

$J(t)$ is the set of catchment/counties contained in catchment t;

$\hat{W}^{(r)}_j$ is the r-th iteration realization of the preliminary prediction for catchment/county j; and

$e^{\hat{u}(r)}$ is the retransformation bias factor.

As described in Schwarz and others (2006), the coefficient of variation of unconditional estimated fertilizer use for catchment t, expressed as a percent, is given by

$$COV\left[\hat{F}_{tU}^{u}\right] = 100\sqrt{\frac{V\left[e^{u}\right]}{\hat{e}^{u^{2}}} + \frac{V\left[\hat{F}_{tU}^{u\,(r)}\right]}{\hat{F}_{tU}^{u\;2}}} \tag{20}$$

where

$\hat{F}_{tU}^{u}$ is defined in equation 13;

$V[e^{u}]$ is the variance of exponentiated state-level residuals, $\exp(\hat{e}_s)$, from the original nonlinear least-squares model estimates;

$\hat{e}^{u}$ is given by equation 12; and

$V[\hat{F}_{tU}^{u\,(r)}]$ is the variance over bootstrap iterations, r, of the bias-corrected unconditional estimate of fertilizer use for catchment t, $\hat{F}_{tU}^{u\,(r)}$, as given in equation 19.

Prediction uncertainty for the conditional estimates is simpler to evaluate because the model error for use, u, is assumed to be identified by the state-level fertilizer sales, which is known and not subject to uncertainty. From equation 20, the first term in the sum under the radical is zero. The only source of uncertainty, therefore, is due to random sampling, which affects the modeled shares used to allocate state-level sales to individual catchments. Therefore, the coefficient of variation for the conditional elemental fertilizer use estimate for catchment t, expressed as a percent, is

$$COV\left[\hat{F}_{tC}^{u}\right] = 100\sqrt{\frac{V\left[\hat{F}_{tC}^{u\,(r)}\right]}{\hat{F}_{tC}^{u\;2}}} \tag{21}$$

where

$\hat{F}_{tC}^{u}$ is given by equation 18; and

$V[\hat{F}_{tC}^{u\,(r)}]$ is the variance over bootstrap iterations, r, of the catchment-aggregated, conditional estimates, $\Sigma_{j\in J(t)} \hat{W}_{j}^{(r)} F_{s(j)} / \hat{W}_{s(j)}^{(r)}$.

Results

Model Estimation

The specifications of the nitrogen and phosphorus fertilizer-use models include the same extensive variables and nominally similar intensive variables. The extensive variables consist of a delineation of the agricultural area within each catchment/county reporting unit into five distinct crop groups. These crop groups include corn, grass-pasture-hay, miscellaneous, nitrogen-fixing, and small grains (fig. 4, table 2.1). These variables are not subject to any mathematical transformation in order to maintain a strict scaling of fertilizer use with agricultural cropping area. Because of this specification and the mean difference form specified for the intensive variables, the coefficients associated with the extensive variables can be interpreted as the elemental mass of fertilizer application per unit area of the given crop group.

The coefficients for the extensive variables are estimated with the constraint that the estimated value must be nonnegative. The constraint implies the significance of the coefficient is determined by applying a one-sided t-test, and the reported *p*-values are one-sided—meaning the *p*-values are one-half the value of a standard, two-sided *p*-value.

The intensive factors included in the models are derived from a set of five variables—fertilizer expenditures per unit of cropland area; mean precipitation; mean actual evapotranspiration; the elemental mass of manure per unit of cropland area, which is the nitrogen mass for the nitrogen fertilizer-use model; and the phosphorus mass for the phosphorus fertilizer-use model. Each of the intensive variables is included in the model after transformation via the natural logarithm function and subsequent differencing from each of the intensive variables mean logarithm-transformed value. This transformation implies the estimated coefficients can be interpreted as the rate of change of fertilizer use, per rate of change of the intensive variable. A physical interpretation of the coefficients is complicated by the inclusion of the expenditure variable, which likely varies in response to variations in the other intensive variables. Thus, for example, the precipitation coefficient measures the effect on elemental fertilizer use holding total expenditure for fertilizer fixed, without allowing for the effect of varying precipitation on expenditures.

All intensive variables, except for manure, are specified in the model to interact with the extensive variables in an equivalent way, implying $\boldsymbol{\beta}_k$ does not vary by source, k. The exception is the manure variable that likely has a differential effect on delivery depending on the extensive variable with which it interacts. Manure often serves as a direct substitute for commercial fertilizer for grass-pasture-hay crops and is less of a substitute for other crop groups. Consequently, the specification of the intensive variables has manure included twice—first as an interaction with all crop groups and second as an interaction with grass-pasture-hay only. This specification implies the coefficient associated with manure interacting with grass-pasture-hay represents a differential effect of manure on commercial fertilizer use, as compared to the full effect determined by the coefficient of manure interacting with all crop groups.

The decision to use similar specifications for both the nitrogen and phosphorus fertilizer-use models is based on evaluations of preliminary models that consistently demonstrate the statistical significance of like variables in the two models. An effort was made to incorporate variables having a special influence on the phosphorus model, through variables reflecting natural concentrations of phosphorus in the soil, but such efforts failed to discern statistical significance either regionally or nationally. The nearly identical specification of the two models (identical except for the elemental content of the manure variable) implies spatial variation in the predicted ratio of nitrogen to phosphorus fertilizer content in use is largely a consequence of different valuations of the mediating coefficients estimated for the models.

Nitrogen

The estimation results for the nitrogen fertilizer-use model are listed in table 2, with the standard errors of the coefficients reflecting the correction described by equation 8. All but one coefficient in the model is significant at the 0.05 level (the significance of extensive variable coefficients is reported by a one-sided *p*-value; the significance of the intensive variables is a standard two-sided *p*-value). The two crop groups having the largest model coefficients are small grains (2.04) and corn (1.54). The smaller coefficient for corn may in part, be a result of crop rotation, whereby the nitrogen demands for growing corn are partially supplied by rotation with nitrogen-fixing crops in previous years, a factor that is not accounted for in the model. The coefficient for nitrogen-fixing crops is unexpectedly larger than some of the other crop groups, possibly because of misclassification errors for this crop group, which is the second least accurate after grass-pasture-hay (table 3.1).

The grass-pasture-hay coefficient has the smallest standard error. Grass-pasture-hay accounts for more than 50 percent of all cropland in the conterminous United States and occurs with the greatest variation of any crop group, although the rate of application of fertilizer to this crop group is probably the most variable, as indicated by relating the absolute magnitude of model residuals to crop group areas. This enhanced variability associated with grass-pasture-hay is likely a consequence of misclassification error, which is most prevalent for this crop group.

Three intensive variables—the logarithms of fertilizer expenditures per unit cropland, actual evapotranspiration, and nitrogen from manure per unit cropland as interacted with all crops—have a positive coefficient, indicating an enhancement of fertilizer use; and, two intensive variables—precipitation and manure interacting with grass-pasture-hay—have a negative coefficient, indicating suppressed use. The coefficient on manure interacting with all crops is the only coefficient that is not statistically significant at the 0.05 level.

The overall effect of manure interacting with grass-pasture-hay is given by the sum of the two manure coefficients, which at –0.072 (0.067 plus –0.139) is negative and statistically significant (the two-sided *p*-value equals 0.026). The negative effect of manure interacting with grass-pasture-hay is consistent with the interpretation that manure is a substitute for commercial fertilizer for application to that crop group. The positive sign of the coefficient for manure interacting with all crops is consistent with this interpretation, given the presence of expenditures in the model. With expenditures on fertilizer held fixed, a decrease in commercial fertilizer applied to grass-pasture-hay implies an increase in fertilizer applied to all other crops, which the model results confirm.

The nitrogen fertilizer-use model explains 74 percent of the variation in the logarithm of total nitrogen sold at the county level, and the root mean square residual (RMSE) for the log-transformed residuals is 1.20. The reporting-unit residuals are likely more variable than the residuals associated with use because of net exportation of fertilizer sales across county boundaries, and the fit of the model as it pertains to use is better represented by the residuals computed at the state level. This assumption is supported by the fit statistics computed from state-level residuals that indicate the model explains 93 percent of the logarithm of nitrogen fertilizer use. The RMSE for the use log-transformed residuals is 0.473, a much better fit than indicated by the reporting-unit residuals (note, the residual for Wyoming is excluded from this evaluation as the data for that State are deemed unreliable). The state-RMSE implies, as an approximation, that predictions of fertilizer use within any given catchment will be within 47 percent of the true use for a one standard-deviation error (Schwarz and others, 2006).

Studentized residuals are the residuals divided (or scaled) by the reporting-unit RMSE and are mapped for the state- and county-level reporting units (fig. 5), with a negative residual indicating overprediction and a positive residual indicating underprediction. The largest residual for state-level reporting units is for Wyoming, where the positive residual indicates that actual sales greatly overstate likely use, and where fertilizer sales data are deemed unreliable—although Wyoming is mostly pasture, the grass-pasture-hay crop group is known to be poorly classified (table 3.1). For county-level reporting units, the largest residuals reside in Texas and Tennessee, where positive and negative residuals are observed, and in the isolated cases of Baltimore City, Maryland, and Elliott County, Kentucky. The export and import of fertilizer across county borders can partially explain these anomalies, although reporting errors are also a possible cause.

Table 2. Model coefficient estimates and statistics for the 2012 nitrogen and phosphorus fertilizer-use models of the conterminous United States.

[Standard errors of model estimates are based on the coefficient covariance matrix with the correction for spatial correlation (see eq. 8). Reported *p*-values for crop-group variables are one sided probabilities from the t-distribution, all other *p*-values are two-sided. Fertilizer use R^2 and RMSE are based on logarithm transforms of state-level aggregations of actual and predicted use. t/km^2, metric ton per square kilometer; <, less than; km^2, square kilometer; NA, not applicable; R^2, coefficient of determination for model estimated in logarithm space; RMSE, root mean square error computed from residuals in logarithm space]

Variables or statistic	Model coefficient units	Nitrogen			Phosphorus		
		Estimate of model coefficient	Standard error of model coefficient	Probability level (*p*-value)	Estimate of model coefficient	Standard error of model coefficient	Probability level (*p*-value)
Crop (extensive) variables							
Corn crops	t/km^2	1.54	0.749	0.020	0.107	0.078	0.085
Grass-pasture-hay crops	t/km^2	1.07	0.161	<0.0001	0.205	0.029	<0.0001
Miscellaneous crops	t/km^2	1.05	0.460	0.011	0.044	0.028	0.062
Nitrogen-fixing crops	t/km^2	1.09	0.612	0.037	0.210	0.074	0.002
Small grain crops	t/km^2	2.04	0.926	0.014	0.117	0.069	0.045
Intensive variables							
Log expenditures (times $1000) per km^2 cropland (2012) in 1992 dollars	Unitless	0.961	0.061	<0.0001	1.07	0.058	0.000
Log precipitation (2012)	Unitless	–0.761	0.315	0.016	–0.610	0.313	0.052
Log actual evapotranspiration (2000–14 average)	Unitless	1.35	0.513	0.008	1.01	0.514	0.050
Log nitrogen from manure (interacted with all crops) per km^2 cropland	Unitless	0.067	0.047	0.153	NA	NA	NA
Log nitrogen from manure (interacted with pasture only) per km^2 cropland	Unitless	–0.139	0.063	0.026	NA	NA	NA
Log phosphorus from manure (interacted with all crops) per km^2 cropland	Unitless	NA	NA	NA	0.134	0.053	0.011
Log phosphorus from manure (interacted with pasture only) per km^2 cropland	Unitless	NA	NA	NA	–0.260	0.069	<0.0001
Model diagnostics							
Reporting unit (state/county)							
Estimated model R^2	NA	0.738	NA	NA	0.725	NA	NA
Estimated model RMSE	NA	1.200	NA	NA	1.221	NA	NA
Number of sites (model)	NA	2,097	NA	NA	2,090	NA	NA
State							
Fertilizer use R^2	NA	0.927	NA	NA	0.929	NA	NA
Fertilizer use RMSE	NA	0.473	NA	NA	0.473	NA	NA
Number of sites (state level)	NA	47	NA	NA	46	NA	NA

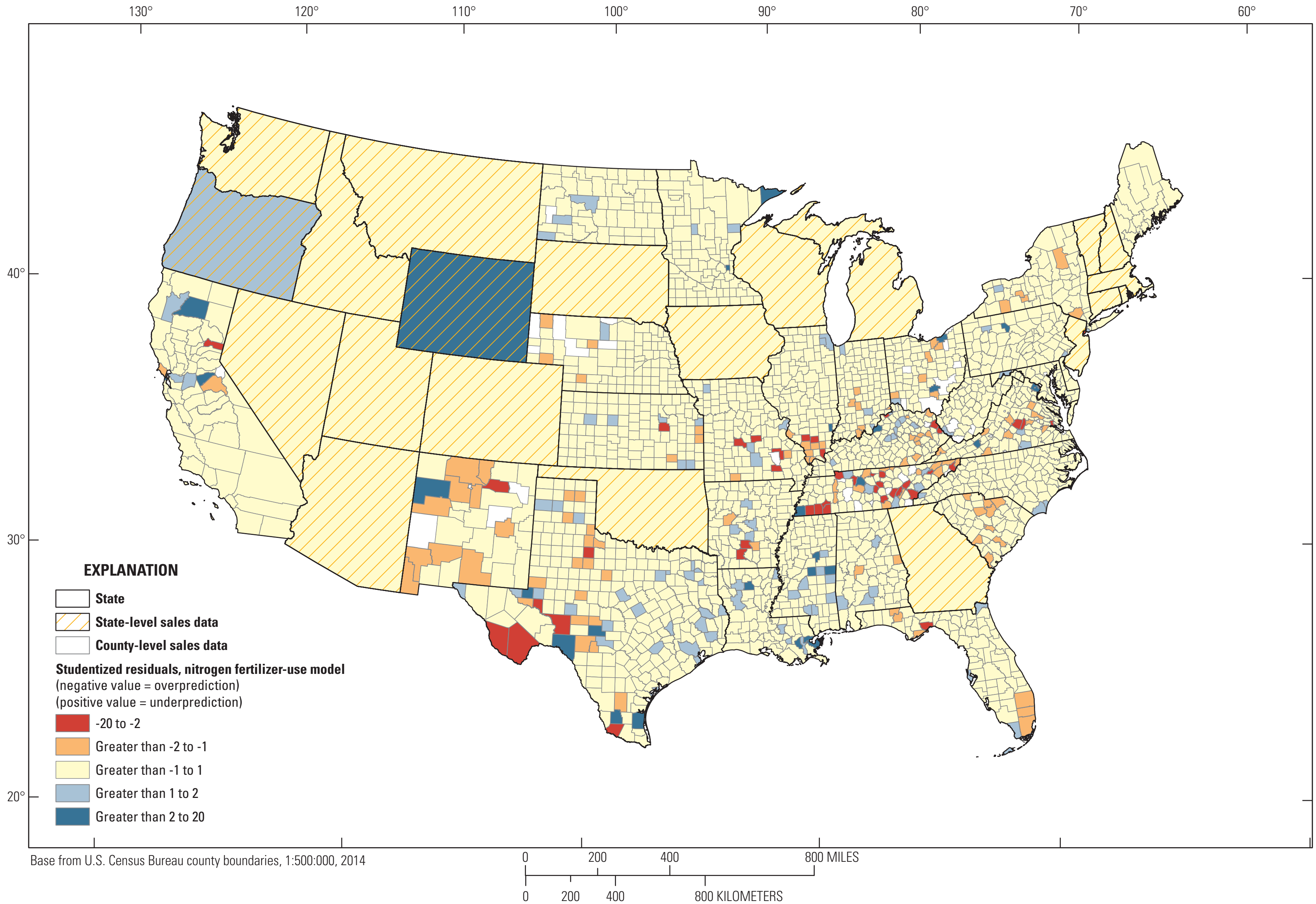


Figure 5. Studentized residuals for 2012 nitrogen fertilizer-use model.

Scatter plots relating observed nitrogen fertilizer sales to predicted use at the county and state scales are shown in figures 6 and 7, respectively. State-level reporting units are also displayed (yellow dots) in the county-level plot in figure 6. County-level reporting units in Mississippi are also displayed (black dots) in figures 6 and 7 as part of the model evaluation, as described in the “Model Evaluation” section of this report. In both figures, because the observations are plotted on a logarithmic scale, predicted use is computed prior to applying the retransformation bias correction factor. Figure 6 shows that the relation between sales and predicted use is much more variable for lower values, with variability declining as predicted use increases—significantly so for observations reported at the state level. Figure 6 also shows a bias whereby predicted use tends to exceed sales for lower mass, a result that is likely due to counties with less agriculture tending to have a less-developed fertilizer distribution network, implying more use is met by imports of sales made in neighboring counties. The state-level relation between use and sales, displayed in figure 7, shows even variability for the full range of use, with considerably less variability than is evident at the county level. Again, state-level variability supports the assumption that sales approximate use at the state scale.

Phosphorus

Estimation results for the phosphorus fertilizer-use model are listed in table 2. All coefficients are statistically significant at the 0.10 level and most are significant at the 0.05 level. The two crop groups with the largest model coefficients are nitrogen-fixing crops (0.210) and grass-pasture-hay (0.205). Phosphorus enhances nitrogen fixation and nitrogen-fixing plants (alfalfa, soybeans, and other legumes) generally require more phosphorus than grasses for growth and development. Phosphorus is not nearly as important for the growth of corn, as reflected by the model coefficients.

The nitrogen fertilizer-use model crop coefficients are approximately 5 to 24 times higher than the phosphorus model coefficients. Presumably, this is a consequence of greater mass plant uptake of nitrogen as compared to phosphorus. Nitrogen plays a key role in photosynthesis and the formation of protein. Moreover, the greater solubility of nitrogen implies more nitrogen mass can be lost through infiltration through the soil, necessitating a need for more replacement after each growing season. Additionally, phosphorus in fertilizer is more expensive per unit mass than nitrogen (U.S. Department of Agriculture, Economic Research Service, 2016) and farmers rely on soil and tissue testing to economize on phosphate (P_2O_5) fertilizer application.

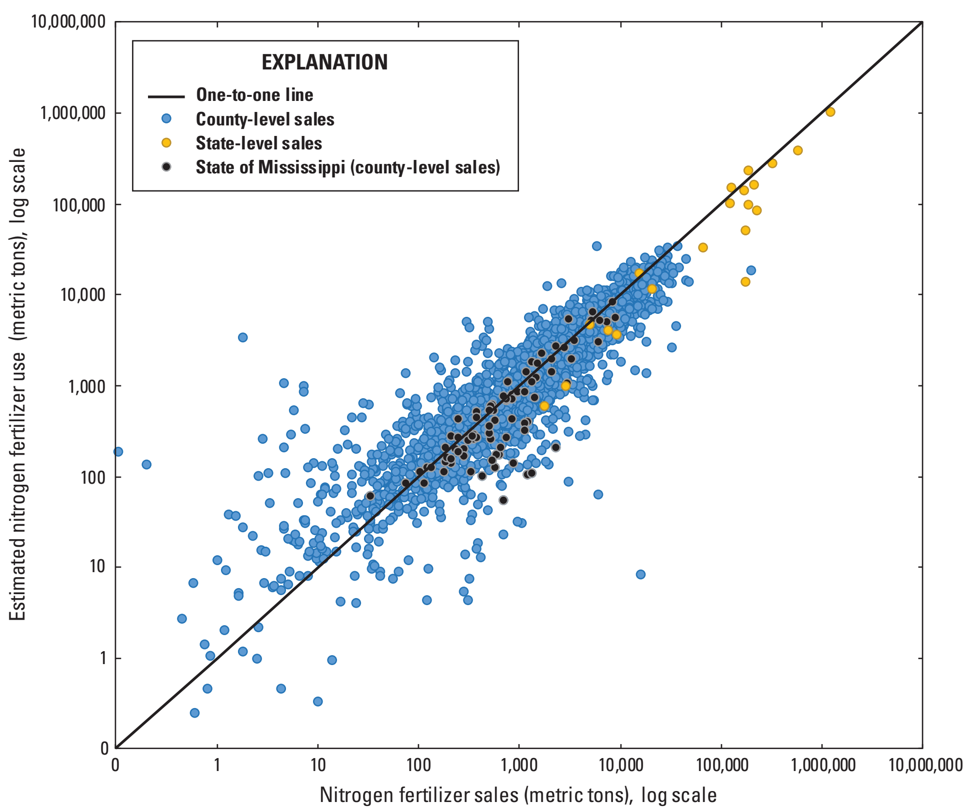


Figure 6. Observed 2012 nitrogen fertilizer sales compared to estimated 2012 nitrogen fertilizer use summarized at the county level.

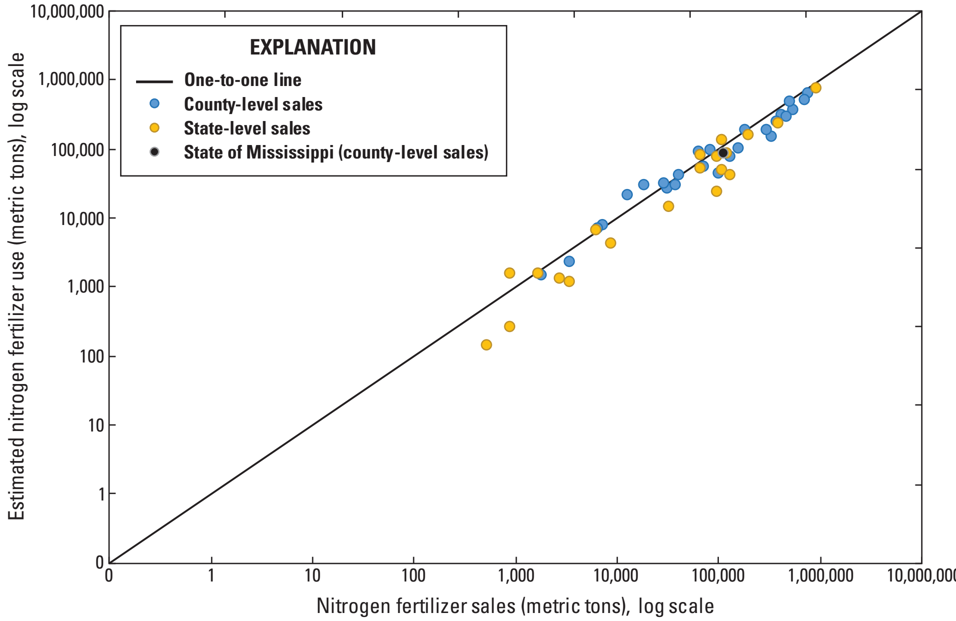


Figure 7. Observed 2012 nitrogen fertilizer sales compared to estimated 2012 nitrogen fertilizer use summarized at the state level.

The coefficients associated with the intensive variables have values that conform to those for the nitrogen model. The combined effect of phosphorus manure interacting with grass-pasture-hay is negative at –0.126 (coefficient for phosphorus from manure without interaction equals 0.134 and coefficient for phosphorus from manure interacting with grass-pasture-hay equals –0.260; the two-sided *p*-value is 0.0004), supporting the finding from the nitrogen fertilizer-use model that manure is a substitute for commercial fertilizer for that crop group. In magnitude, the total effect is nearly twice that for nitrogen, suggesting that manure is a better substitute for commercial fertilizer for phosphorus than for nitrogen. This result may be due to large sources of phosphorus coming from poultry manure, a major source of manure in some regions of the country.

The model does not include any variable that accounts for natural sources of phosphorus occurring in the soil and attempts to include such a variable were unsuccessful. The lack of such a variable does not indicate that the model is blind to this factor. The phosphorus content of fertilizer has a significant effect on fertilizer price; therefore, variations in phosphorus fertilizer application possibly are reflected in fertilizer expenditures. The presumption, therefore, is that given the inclusion of fertilizer expenditures as a determinant of phosphorus fertilizer use, the inclusion of variables related to natural fertilizer content of the soil fail to improve model fit. This result could be a consequence of noise confounding the signal contained in the natural fertilizer variables. It could also be a consequence of past agricultural activity depleting soils of their natural phosphorus content or to the availability of phosphorus in the soil in a form for plant uptake, or both.

The phosphorus fertilizer-use model explains 72 percent of the variation in the logarithm of total elemental phosphorus sold, and the RMSE for the residuals in logarithm space is 1.22, a fit that is similar to the nitrogen model. The fit statistics computed from state-level residuals (without application of the retransformation bias correction factor) have a coefficient of determination (R^2) of 93 percent and an RMSE of 0.47, indicating that the method determines nearly identical fit statistics for both the nitrogen and phosphorus fertilizer-use models. As with nitrogen, the residual for Wyoming is excluded from the determination of state-level statistics. Also excluded is Nevada, which has no phosphorus sales.

Phosphorus model studentized residuals (residuals normalized by the reporting-unit RMSE) for the state- and county-level reporting units are mapped in figure 8. The location of extreme residuals, those having absolute studentized values exceeding one, is nearly the same as the nitrogen case (fig. 5), implying the errors in predicting fertilizer use relate more to total fertilizer application than to errors specific to elemental composition. This observation is consistent with the notion that errors in the models arise primarily from idiosyncrasies in the distribution system leading to a greater range in net exports and misclassification of cropping patterns inducing common errors in both models.

Patterns contained in the scatterplot of predicted nitrogen use relative to nitrogen sales at the state- and county-level reporting units (fig. 6) are apparent in the corresponding plot for phosphorus, shown as figure 9, although county-level observations display a slightly greater variability for phosphorus. The scatterplot of phosphorus predicted use relative to sales exclusively at the state level (fig. 10) shows a better fit than for nitrogen (fig. 7), particularly in the higher-use states. Together, these results imply a slightly greater tendency for exporting and importing fertilizers with high phosphorus content.

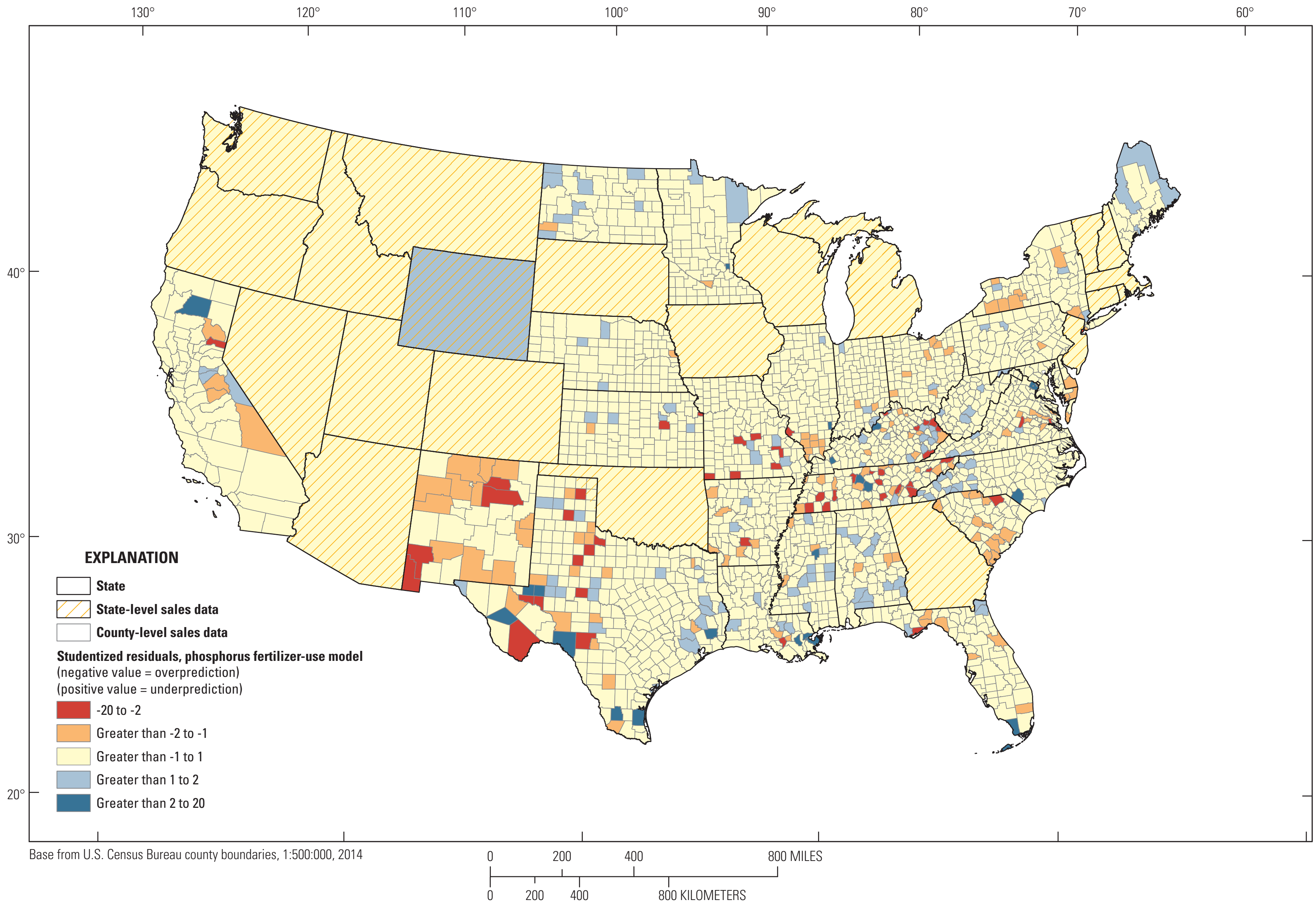


Figure 8. Studentized residuals for 2012 phosphorus fertilizer-use model.

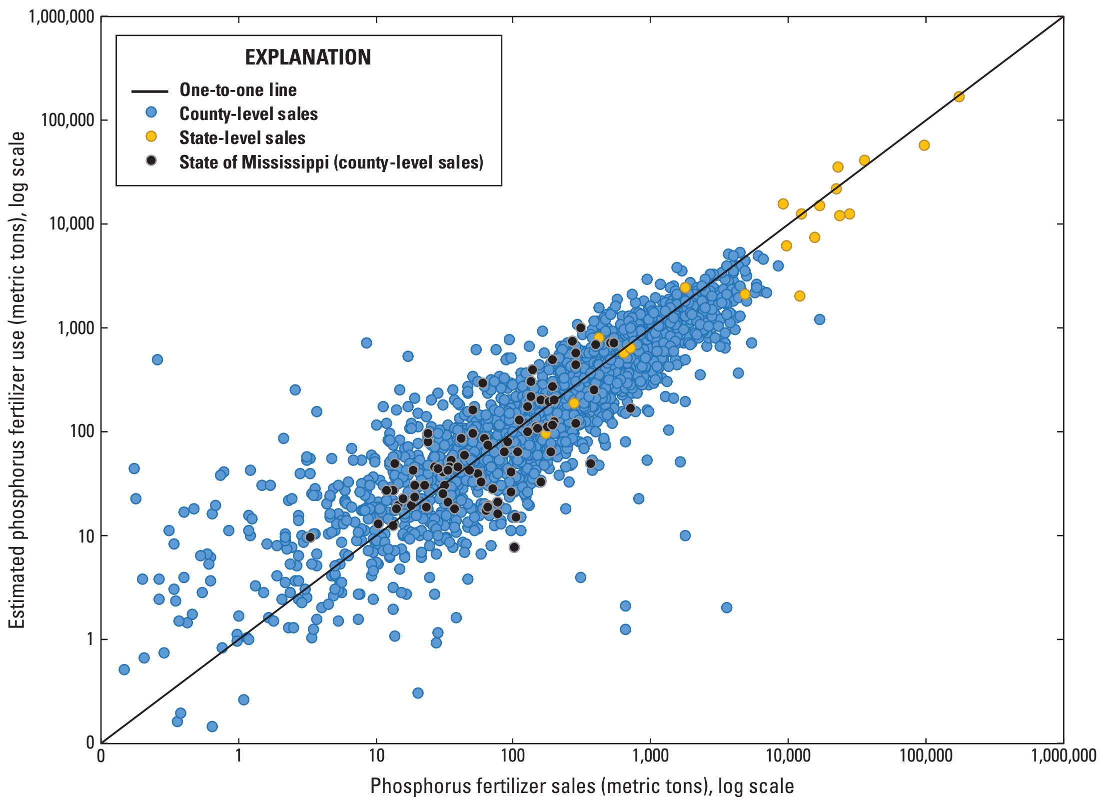


Figure 9. Observed 2012 phosphorus fertilizer sales compared to estimated 2012 phosphorus fertilizer use summarized at the county level.

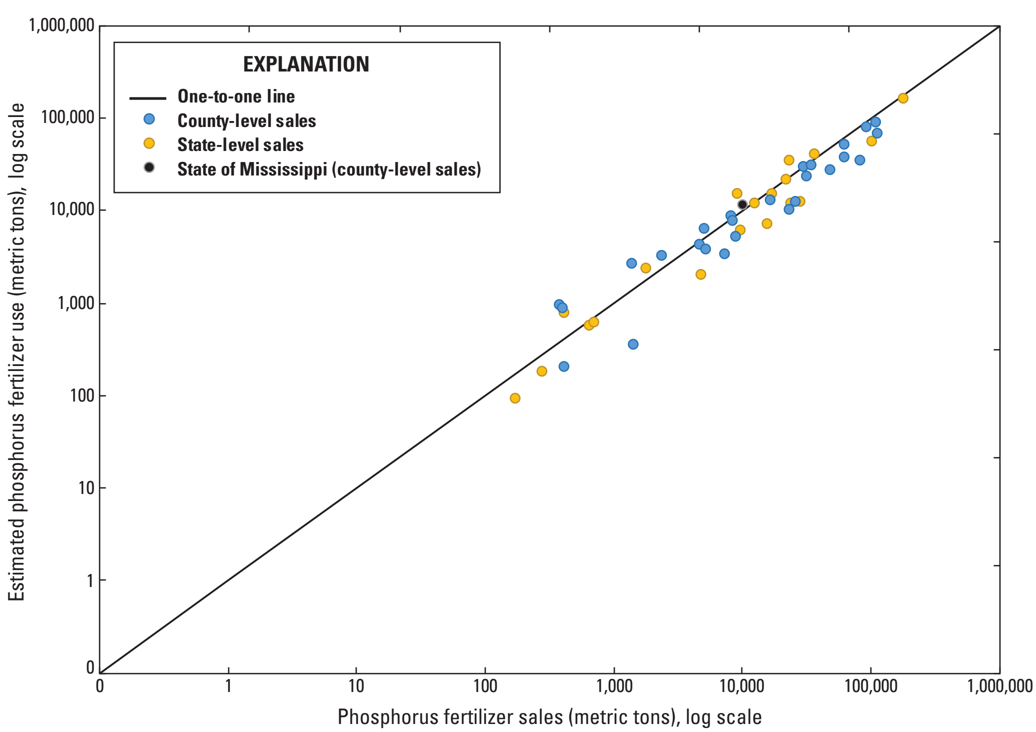


Figure 10. Observed 2012 phosphorus fertilizer sales compared to estimated 2012 phosphorus fertilizer use summarized at the state level.

Residual Analysis

The correction to the covariance matrix, described by equation 8, requires an estimate of the residual covariance matrix, **Ω**. The estimates of the residual variances and correlations used to evaluate this matrix are listed in table 3.

Attempts to estimate the fertilizer use error-components model based on equation 7 resulted in several inconsistencies. The nonlinear regression for squared, nitrogen model residuals yielded an insignificant state-component variance (σ^2_η), a significant county-component variance (σ^2_ϵ), and a county-component spatial correlation (ρ_ϵ) of 1; however, a spatial correlation of 1 implies the county component is perfectly correlated. The county component is observationally equivalent to a simpler covariance structure consisting of just a state-level component—the state-level component is inconsistent with the insignificant result for the state-component variance. The nonlinear regression for squared, phosphorus model residuals gave significant state- and county- component variances and a county-component spatial correlation of zero; however, the summed values of the state- and county- component variances greatly exceeds the variance of the county residuals from the phosphorus model, implying the net-export variance is negative—an impossible result.

Because of these inconsistencies, and given that county use is not observable, the decision was made to suppress the county component to the use residual and base all variance and covariance on the state component, the implication being that residual use throughout a state is perfectly correlated but uncorrelated across states. This assumption may overstate the degree to which use is correlated across substate units within the same state. Moreover, counter to the assumption, if a component of the state residual is independent across substate units, then the variability of the use residual at the substate level is understated by the variability of the state-level residual.

The first three rows of table 3 reflect the assumed use residual covariance structure, with a positive state-component variance (σ^2_η) and the county-component variance (σ^2_ϵ) set to zero. With no county-component variance, the county-component spatial correlation (ρ_ϵ) indicated in the third row among neighboring counties becomes irrelevant. The fourth and fifth rows of table 3 give the estimates of the variance of the county-level residuals, $V[e_c]$, and the variance of the difference in the residuals among neighboring counties in the same state, $V[e_{c_1} - e_{c_2}]$, for those states having county-level reporting units. Counties that are merged because of no sales are included in the variance of county-level residuals but excluded from the variance of the difference in neighboring county residuals.

The last two rows of table 3 give the values of the variance and spatial correlation among neighboring counties of the county-level net-export residual component, *x*, as derived from the estimates of σ^2_η and the two variances, $V[e_c]$ and $V[e_{c_1} - e_{c_2}]$. The net-export residual component is more than five times the magnitude of the use component, implying the fertilizer sales data are noisy indicators of fertilizer use at the county level, as expected. The spatial correlation of net exports among neighboring counties is expected to be negative, because a county with significant positive net exports is likely to locate next to a county with large imports (that is, negative net exports). The finding of a positive spatial correlation implies exports must be balanced by imports across a larger spatial extent than neighboring counties. The fact that the state-level residuals have a much smaller variance than county-level residuals implies that a negative correlation must exist at some scale below the state level.

Table 3. Residual variance and covariance estimates for the nitrogen and phosphorus fertilizer models.

[Note that σ_η^2 and σ_ε^2 are the variances of the state- and county-component use residuals; ρ_ε is the correlation coefficient between the county-level components; the variances $V[e_c]$ and $V[e_{c1} - e_{c2}]$ are computed without centering; $V[e_{c1} - e_{c2}]$ excludes grouped-county observations; and σ_x^2 and ρ_x are derived from other estimates in the table using equations described in the estimation methods section. <, less than; NA, not available]

Parameter	Description	Nitrogen			Phosphorus		
		Estimate	Standard error	Probability level (*p*-value)	Estimate	Standard error	Probability level (*p*-value)
σ_η^2	Variance of state-component use residual	0.223	0.049	<0.0001	0.224	0.039	<0.0001
σ_ε^2	Variance of county-component use residual	0.000	NA	NA	0.000	NA	NA
ρ_ε	Correlation among neighboring county use residuals	NA	NA	NA	NA	NA	NA
$V[e_c]$	Variance of county residuals	1.434	0.298	<0.0001	1.488	0.286	<0.0001
$V[e_{c1} - e_{c2}]/2$	Variance of difference of neighboring county residuals (halved)	1.060	0.093	<0.0001	1.118	0.090	<0.0001
σ_x^2	Variance of county net export residual	1.211	NA	NA	1.264	NA	NA
ρ_x	Correlation among neighboring county net export residuals	0.309	NA	NA	0.293	NA	NA

Model Predictions

The spatial patterns of predicted nitrogen and phosphorus fertilizer use in mass units, shown in figures 11 and 12, respectively, are based on the unconditional predictions. For reference, states with county-level reporting units have bolded black borders. The fertilizer-use patterns for nitrogen and phosphorus reflect known agricultural cropping patterns and practices. The maps show the pattern of high use largely mimics the pattern of cropland shown in figure 4, although high use is prevalent in the coastal region of the Southeast, Florida, and isolated areas in the Southwest despite the presence of extensive nonagricultural land—a possible consequence of double cropping. Nitrogen and phosphorus applications are lower relative to cropland occurrence in portions of states in the Southeast where poultry production is common (U.S. Department of Agriculture, National Agricultural Statistics Service, 2015), including eastern Arkansas, northern Louisiana, southeastern Mississippi, Alabama, and northern Georgia.

Nitrogen and phosphorus fertilizer use expressed per unit area of cropland, based on unconditional predictions, are shown in figures 13 and 14, respectively. Unlike the maps of fertilizer use in mass, both figures display a pronounced affinity for political boundaries, a consequence of normalization of use by catchment-level agricultural area advancing the less-resolved county fertilizer expenditures variable as the preeminent determinant of concentrated use. The most intensive use of nitrogen (fig. 13) occurs in areas dominated by crops that require high application rates (U.S. Department of Agriculture, Economic Research Service, 2013). High application rates are predicted for corn (Iowa, Illinois, Indiana, western Ohio, southern Minnesota, eastern North and South Dakota, and eastern Nebraska), rice (southeastern Missouri, eastern Arkansas, northwestern Mississippi, and the gulf coast of Louisiana and Texas), cotton (the coastal plain of North and South Carolina and southern Georgia), and vegetables (southern Florida and central California) (U.S. Department of Agriculture, Office of the Chief Economist, 2014). Lower nitrogen application rates are predicted in the semiarid, West particularly the Plains states where grassland and pasture predominate.

The highest application rates of phosphorus fertilizer (fig. 14) closely follow the pattern for nitrogen, with some differences derived from the availability of natural phosphorus, such as the Delta region of northwestern Mississippi or in areas where poultry production is common and phosphorus-rich manure is made readily available (Oldham, 2012; U.S. Department of Agriculture, National Agricultural Statistics Service, 2015). The model may respond to the availability of natural phosphorus indirectly through an adjustment of fertilizer expenditures, despite the absence of natural phosphorus as an explicit model variable.

Previous studies combining fertilizer sales with Census expenditure data (Ruddy and others, 2006; Gronberg and Spahr, 2012) constrain the ratio of nitrogen to phosphorus to be constant for all areas within a state. Because the present study derives nitrogen and phosphorus use separately, with different model coefficients and slightly different data, the ratio of nitrogen to phosphorus use is not constrained. The ratio derived from unconditional predictions is displayed in figure 15 to better highlight the modeled determinants of variability. The ratio of nitrogen to phosphorus fertilizer use is generally highest in the Plains states, between the Mississippi River and Rocky Mountains, where nitrogen-fixing crops, such as soybeans, are uncommon (U.S. Department of Agriculture, Office of the Chief Economist, 2014). The ratio is also elevated in Florida and Southeastern states. Lower values of the ratio are observed in the corn-soybean crop-growing region of the Midwest (Iowa, Illinois, Indiana, western Ohio, southern Minnesota, eastern North and South Dakota, and Nebraska) where crop rotation with nitrogen-fixing crops is a common practice to supplement application of commercial fertilizer.

Prediction Uncertainty Estimates

The coefficient of variation, expressed as a percent of the fertilizer use prediction, is derived for each catchment according to equations 20 and 21 using a parametric bootstrap analysis based on 200 randomly generated sets of coefficients. A summary of the estimates for two-digit hydrologic regions (U.S. Geological Survey and others, 2017) is given in table 4.

The average coefficient of variation for nitrogen use is greater than that for phosphorus use in all cases listed in table 4, although the difference is small. The average coefficient of variation for conditional predictions is significantly smaller than that for unconditional predictions. This result is expected because the conditional predictions, unlike the unconditional predictions, have no model error—a consequence of assuming the use residuals consist only of a state-level component, a component that is known with certainty because of the observation of state sales and the presumed equivalence of sales and use at that level of aggregation.

The average coefficient of variation for unconditional predictions in all catchments exhibits a considerable range in values across regions, with the maximum value among regions being approximately twice the minimum value. Much of this variation is due to the prevalence within a region of catchments with no agricultural land, which the model predicts with certainty to have no fertilizer use. As listed on the right-hand side of table 4, the average coefficient of variation among catchments with positive use is significantly larger and less variable across regions.

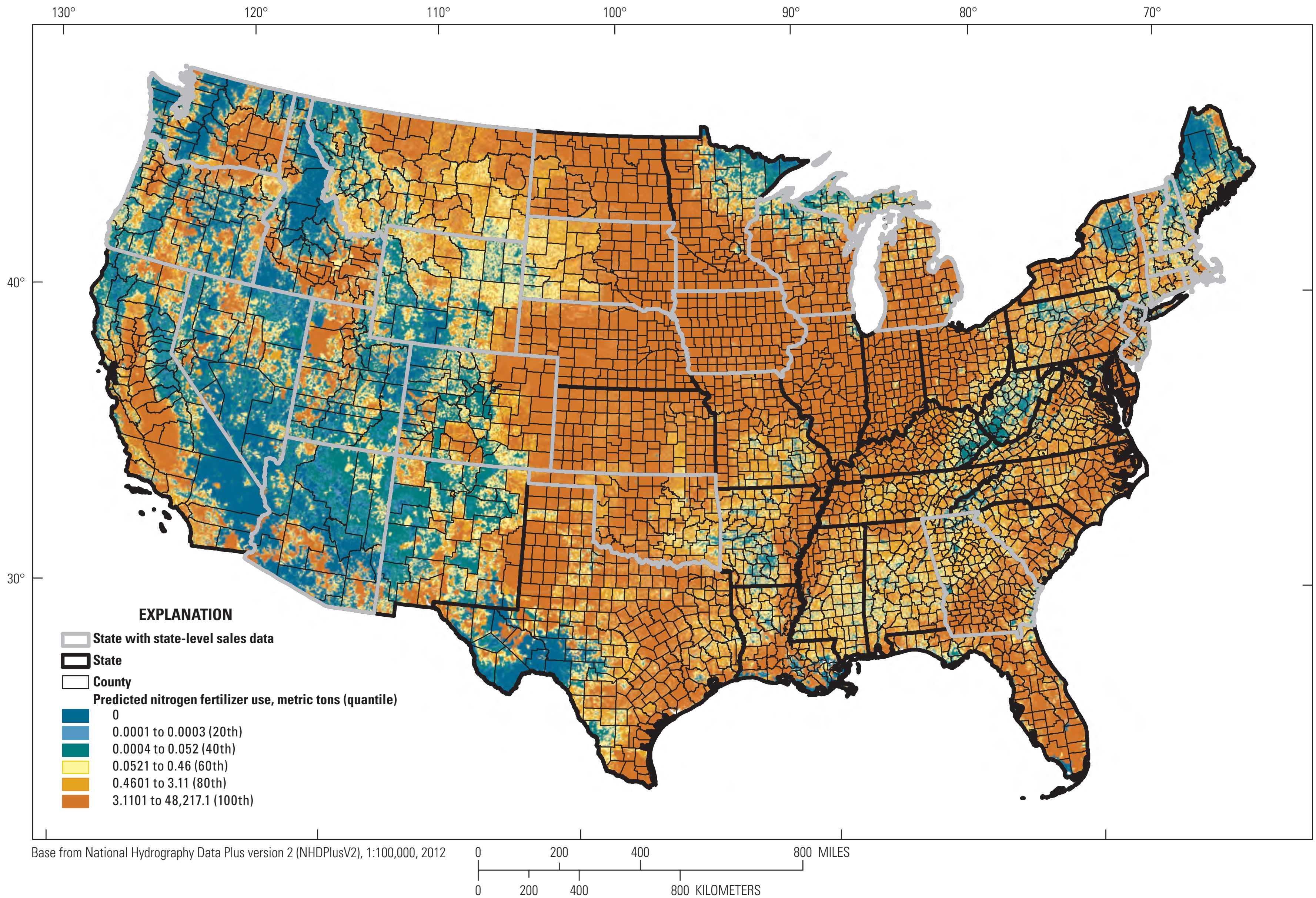


Figure 11. Predicted 2012 nitrogen fertilizer use at the catchment scale (metric tons), unconditional predictions.

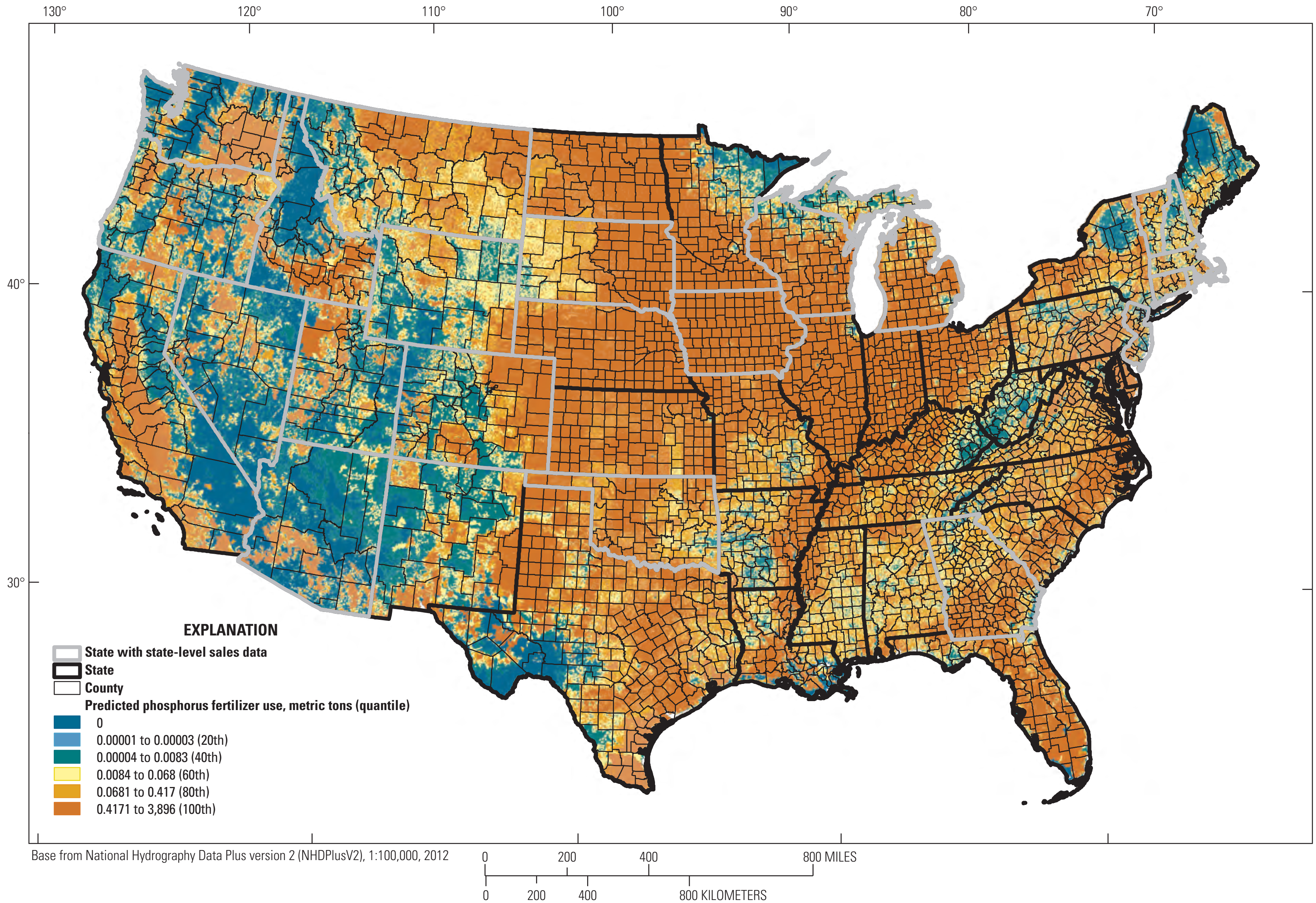


Figure 12. Predicted 2012 phosphorus fertilizer use at the catchment scale (metric tons), unconditional predictions.

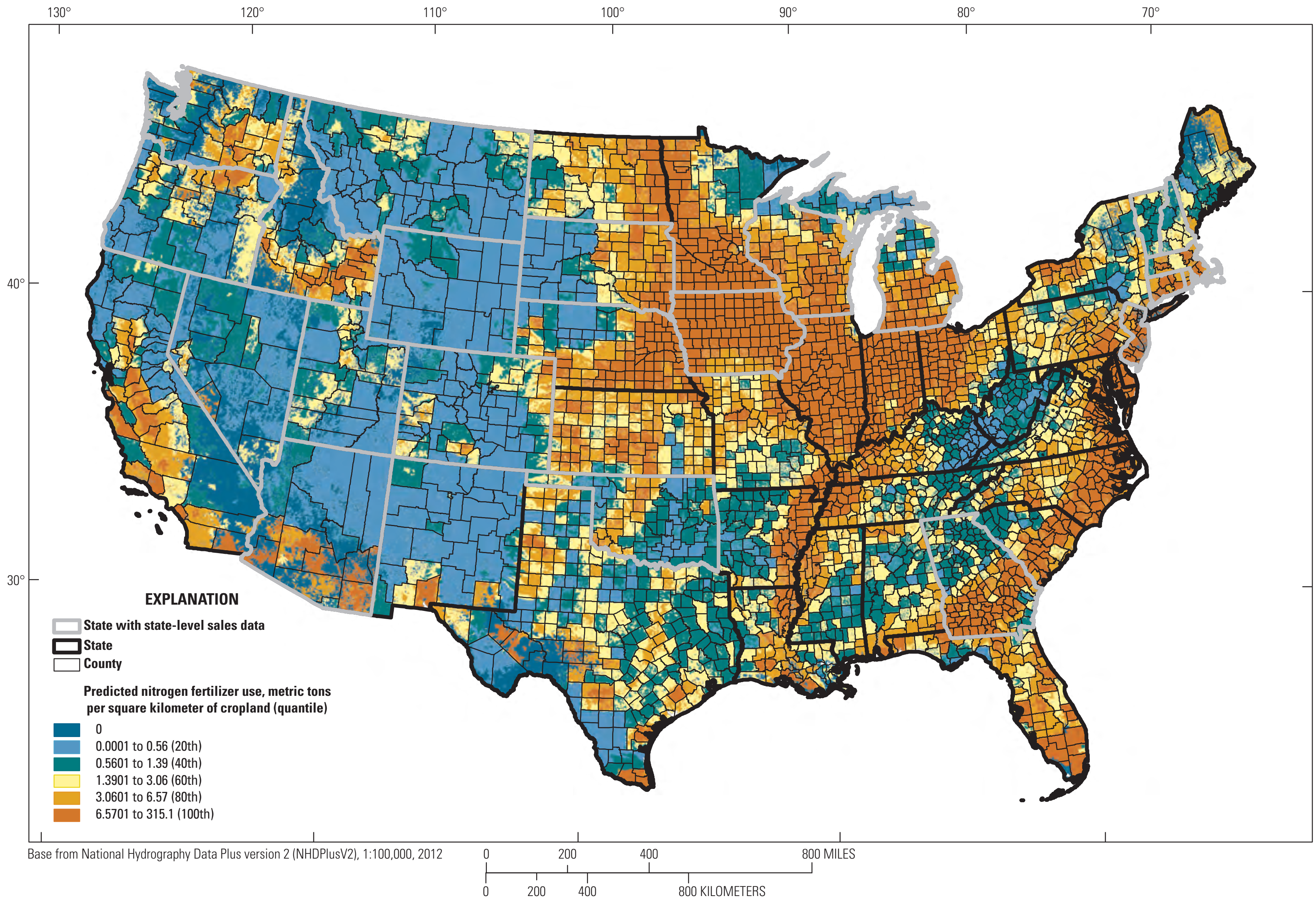


Figure 13. Predicted 2012 nitrogen fertilizer use at the catchment scale (metric tons per square kilometer of cropland), unconditional predictions.

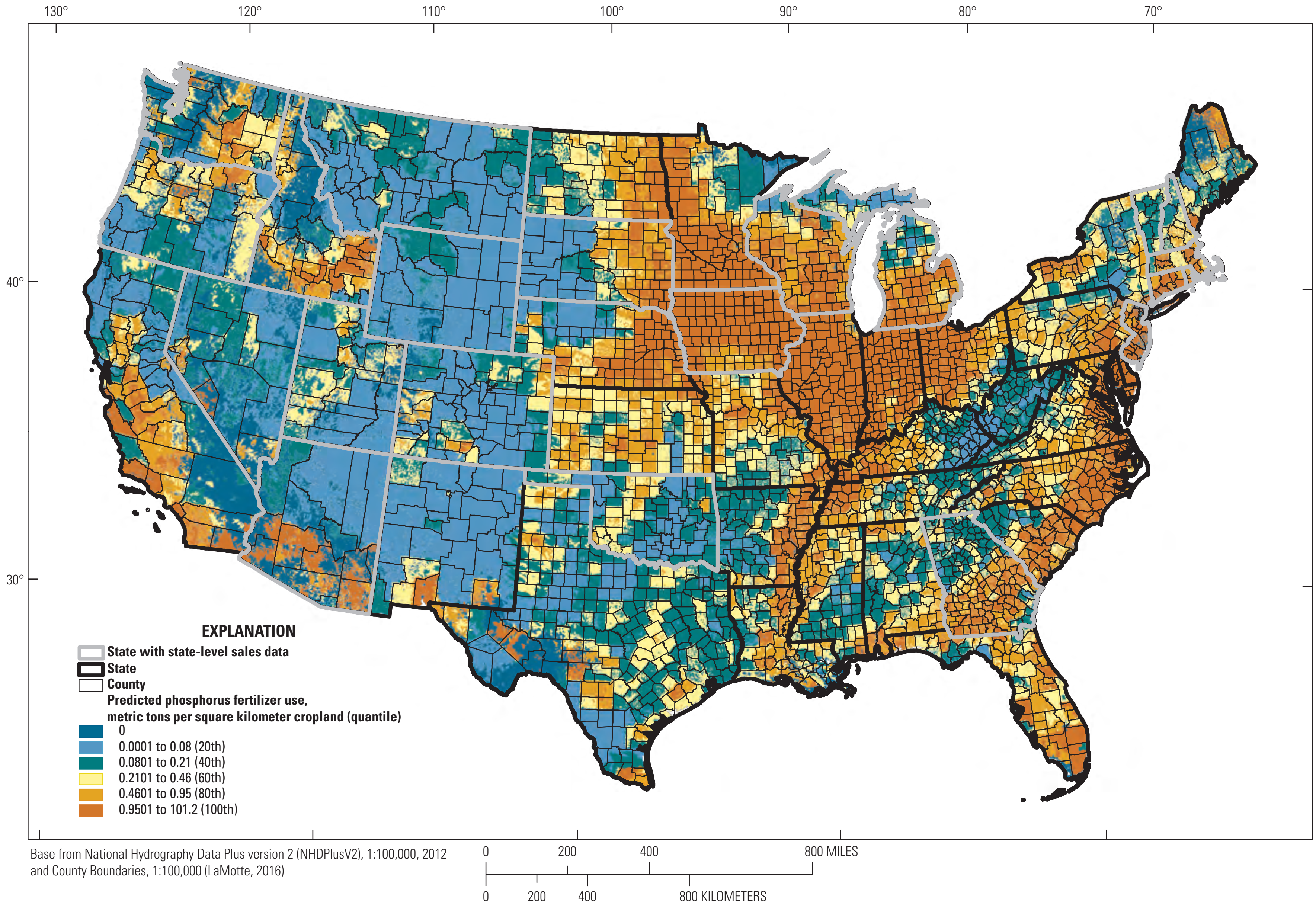


Figure 14. Predicted 2012 phosphorus fertilizer use at the catchment scale (metric tons per square kilometer of cropland), unconditional predictions.

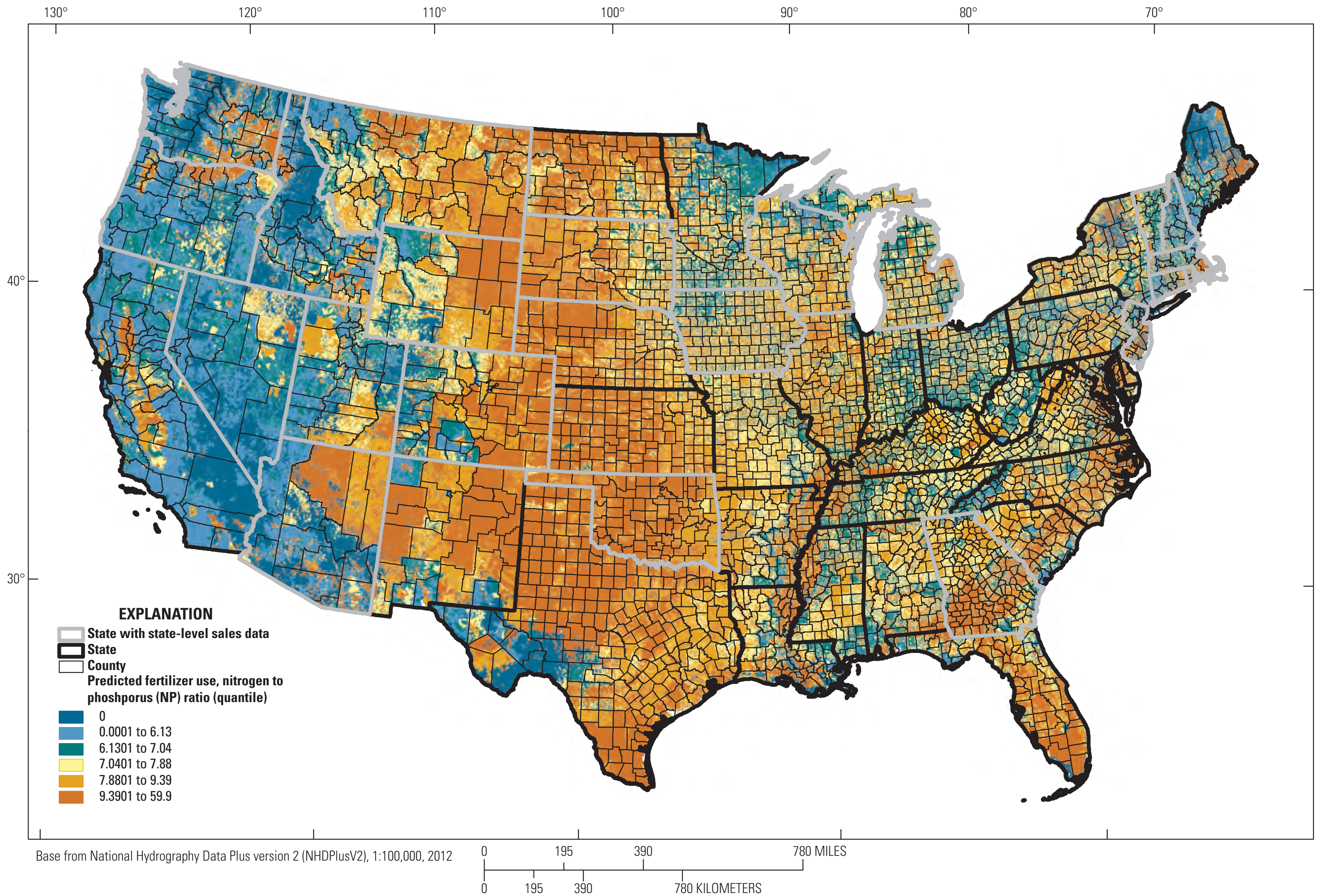


Figure 15. Ratio of predicted 2012 nitrogen fertilizer use to predicted 2012 phosphorus fertilizer use, unconditional predictions.

Table 4. Average of catchment predicted nitrogen and phosphorus fertilizer use coefficient of variation, by hydrologic region and conterminous United States.

[COV, coefficient of variation expressed as a percent of predicted use]

Hydrologic region (code)	All catchments				Catchments with positive use			
	Nitrogen		Phosphorus		Nitrogen		Phosphorus	
	Unconditional prediction COV (percent)	Conditional prediction COV (percent)	Unconditional prediction COV (percent)	Conditional prediction COV (percent)	Unconditional prediction COV (percent)	Conditional prediction COV (percent)	Unconditional prediction COV (percent)	Conditional prediction COV (percent)
New England (01)	33.9	12.0	33.8	10.8	53.3	18.8	53.2	16.9
Mid-Atlantic (02)	38.8	14.4	37.9	13.9	51.6	19.1	50.5	18.5
South Atlantic-Gulf (03)	43.9	18.3	42.0	15.0	51.4	21.5	49.3	17.6
Great Lakes (04)	42.5	15.9	41.0	15.2	50.9	19.1	49.1	18.2
Ohio (05)	46.7	16.3	44.6	14.8	50.6	17.6	48.3	16.0
Tennessee (06)	44.4	14.7	42.6	12.6	49.6	16.4	47.6	14.1
Upper Mississippi (07)	48.6	16.8	46.4	16.3	52.3	18.1	49.9	17.5
Lower Mississippi (08)	42.1	17.7	40.3	16.5	54.1	22.8	51.8	21.2
Souris-Red-Rainy (09)	40.1	20.0	38.6	19.2	53.6	26.8	51.7	25.7
Missouri (10)	51.6	20.6	50.0	18.8	53.1	21.2	51.5	19.3
Arkansas-White-Red (11)	49.5	23.4	47.9	21.5	52.6	24.9	50.8	22.8
Texas-Gulf (12)	46.7	24.4	45.5	23.3	51.0	26.7	49.7	25.4
Rio Grande (13)	42.6	18.2	41.9	17.5	56.1	24.0	55.2	23.0
Upper Colorado (14)	38.6	20.8	37.7	18.2	56.4	30.2	55.1	26.5
Lower Colorado (15)	26.2	14.4	26.1	16.1	59.9	33.0	59.5	36.4
Great Basin (16)	28.7	10.4	27.9	13.7	56.1	20.7	54.5	21.3
Pacific Northwest (17)	34.3	18.9	33.9	16.8	59.8	33.2	59.1	29.3
California (18)	46.7	22.9	46.5	21.1	62.9	30.9	62.6	28.2
Conterminous United States	43.4	18.4	42.1	17.0	53.6	22.7	52.0	20.8

Among catchments with positive fertilizer use, the greatest unconditional prediction uncertainty, for nitrogen and phosphorus, is in the West (regions 13–18). This result is likely a consequence of the absence of county-level sales data in these regions, causing the cropping patterns of the West to be underrepresented in the regression models. The uncertainty for conditional predictions has a similar pattern, although the average uncertainty estimate for phosphorus use in the Great Basin region is suspect; Nevada has no phosphorus sales and consequently no conditional predictions are generated for Nevada's catchments. The same result is true for Wyoming, which has no conditional predictions because of suspect sales data; however, Wyoming comprises a smaller portion of the Missouri Basin, its predominant region, than does Nevada of the Great Basin.

Model Evaluation

Model results are evaluated visually by comparing scatterplots of the predictions of elemental fertilizer use to estimates obtained from fertilizer sales data using other methods and to estimates based on direct survey data compiled for selected crop groups and subregions of the Nation in scatterplots. Model evaluation datasets are also compared statistically by calculating a measure of relative bias and correlation (table 5). For relative bias, the average of the differences between the natural logarithms of the paired evaluation and prediction values is computed. A negative relative bias indicates the model predictions from this study are high relative to the evaluation estimates. For small relative bias, the relative bias statistic is approximately equal to the difference between the two compared estimates, normalized by this study's model prediction. Also reported is a *p*-value to evaluate the statistical significance of the relative bias. This statistic indicates little evidence of bias in any of the evaluation comparisons. Correlation is estimated using the Pearson correlation statistic and is applied to the paired natural logarithms of the evaluation and prediction values.

The first comparison relates the unconditional predictions aggregated at the county level to 2012 county predictions from Brakebill and Gronberg (2017) using the Gronberg and Spahr (2012) method. The plot for nitrogen (fig. 16) shows consistent agreement between the two estimates across the full range of predictions, with a Pearson correlation coefficient of 0.983 (table 5) for the paired natural logarithms of the evaluation and prediction values. The comparison for phosphorus (fig. 17) is similar to nitrogen, although slightly less interrelated with a correlation coefficient of 0.979.

Table 5. Statistical summary of model evaluation comparisons for 2012 nitrogen and phosphorus fertilizer-use models.

[N is number of paired values; relative bias is the average of the natural logarithms of the evaluation data minus the average of the natural logarithms of the paired prediction values; *p*-value pertains to the statistical significance of the relative bias estimate; correlation is the Pearson correlation coefficient of the paired natural logarithms of the evaluation and prediction values]

Prediction type	Evaluation dataset	Nitrogen				Phosphorus			
		N	Relative bias		Correlation	N	Relative bias		Correlation
			Estimate	*p*-value			Estimate	*p*-value	
2012 unconditional prediction of fertilizer use	2012 estimates of fertilizer use, Gronberg and Spahr method[1]	2,100	–0.034	0.919	0.983	2,100	0.049	0.890	0.979
2012 conditional prediction of fertilizer use	2012 estimates of fertilizer use, Gronberg and Spahr method[1]	2,100	0.126	0.606	0.992	2,085	0.074	0.736	0.993
2012 unconditional prediction of fertilizer use	2012 estimates of fertilizer use by International Plant Nutrition Institute[2]	3,067	–0.002	0.998	0.894	3,067	–0.006	0.994	0.899
2012 unconditional prediction of fertilizer use	2010 survey data of fertilizer use on corn and soybeans[3]	14	0.173	0.540	0.944	14	0.129	0.734	0.896
2012 unconditional prediction of fertilizer use	2012 estimates of fertilizer use for Mississippi based on Mississippi Extension fertilizer recommendations[4]	82	–0.309	0.698	0.875	82	0.177	0.834	0.843

[1]Gronberg and Spahr (2012).

[2]International Plant Nutrition Institute (2012).

[3]U.S. Department of Agriculture, Economic Research Service (2016a).

[4]Oldham (2012); U.S. Department of Agriculture, Farm Service Agency (2013).

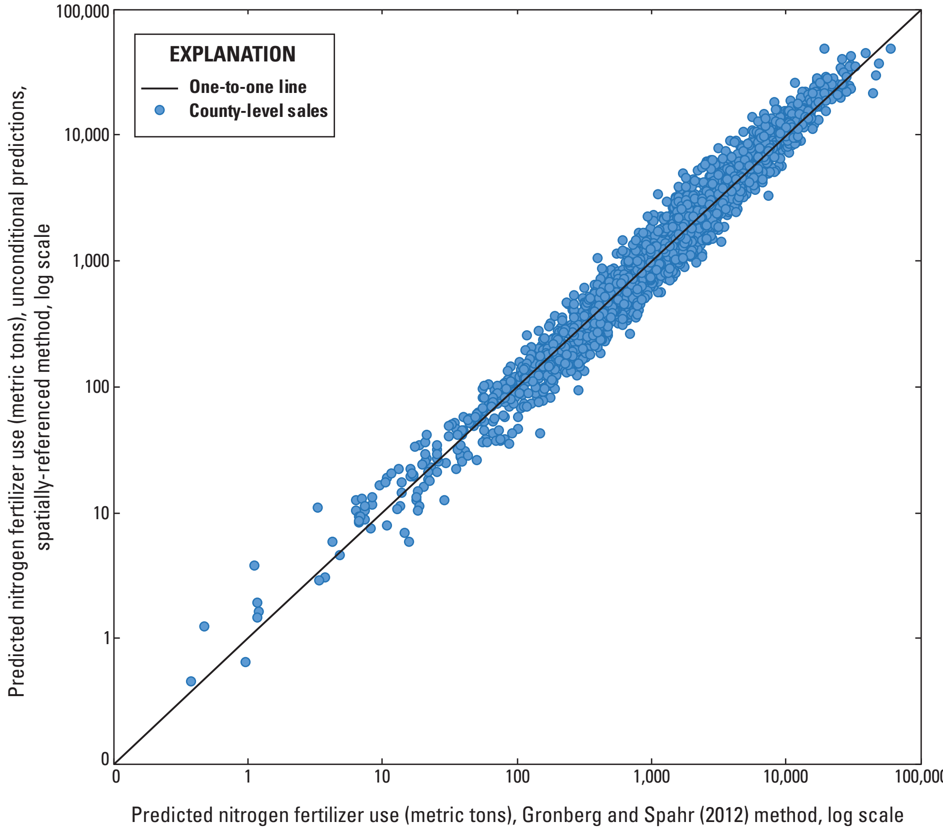


Figure 16. Predicted 2012 nitrogen fertilizer use from the unconditional, spatially referenced method aggregated to the county level compared to 2012 county predictions from Brakebill and Gronberg (2017) using the Gronberg and Spahr (2012) method.

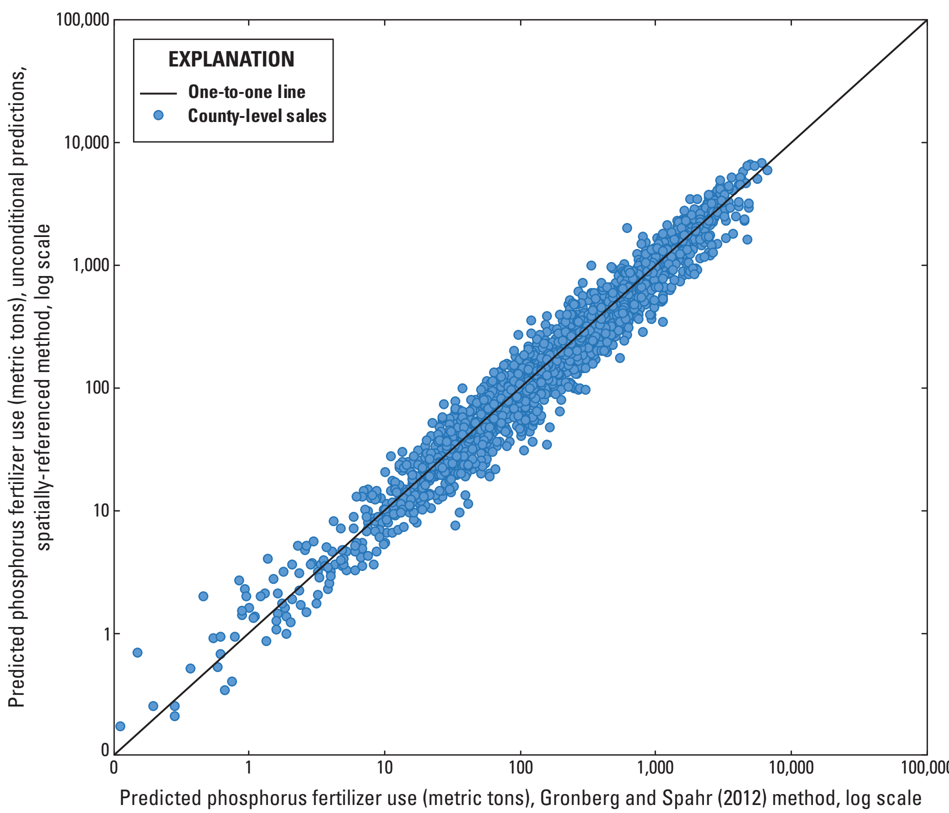


Figure 17. Predicted 2012 phosphorus fertilizer use from the unconditional, spatially referenced method aggregated to the county level compared to 2012 county predictions from Brakebill and Gronberg (2017) using the Gronberg and Spahr (2012) method.

The conditional predictions of elemental fertilizer use are compared to the Brakebill and Gronberg (2017) 2012 county predictions, at the county level, based on the Gronberg and Spahr (2012) method in figures 18 and 19. The figures show a closer relation with Brakebill and Gronberg than observed for the unconditional predictions in the respective figures 16 and 17, with a correlation coefficient of 0.992 for nitrogen and 0.993 for phosphorus (table 5). The conditional adjustment is included in this report for comparison, given the adjustment is most similar to predictions generated by Brakebill and Gronberg (2017). The implication of these figures is that if the spatially referenced predictions of fertilizer use are to outperform the Brakebill and Gronberg predictions in a water-quality model, the unconditional predictions are more likely to do so because they are more flexible in allowing for net export of sales across state borders. Nevertheless, the spatially referenced models incorporate many more subcounty resolution variables than the Gronberg and Spahr method in determining fertilizer use at the catchment scale, making the comparisons in figures 18 and 19 inconclusive with regards to the relative performance of the predictions in a water-quality model. Future testing in SPAtially Referenced Regressions On Watershed attributes models know as SPARROW, could help further evaluate the unconditional compared to the conditional predictions.

The IPNI computes county estimates of elemental fertilizer use based on AAPFCO elemental fertilizer sales data (International Plant Nutrition Institute, 2012). The approach of the IPNI is to interpret fertilizer sales as fertilizer use for counties where sales estimates are deemed reliable. For states with no county-level sales data, the IPNI computes county estimates of use from state sales apportioned by the county share of 2012 USDA COA expenditures. The county sales per unit of cropland area are then located at the centroid of each county and statistically smoothed across the conterminous United States. A national map of fertilizer use is generated by assigning the smoothed fertilizer sales per unit area to classified cropland area using the CDL. The comparisons between the nitrogen and phosphorus unconditional predictions of 2012 county fertilizer use to the IPNI 2012 county estimates are shown in figures 20 and 21, respectively. The comparisons indicate generally greater scatter than the Brakebill and Gronberg predictions (2017) with a correlation coefficient of 0.894 for nitrogen and 0.899 for phosphorus (table 5). This result may largely be due to the presence of significant imports and exports in county sales data that are not directly addressed by the IPNI method.

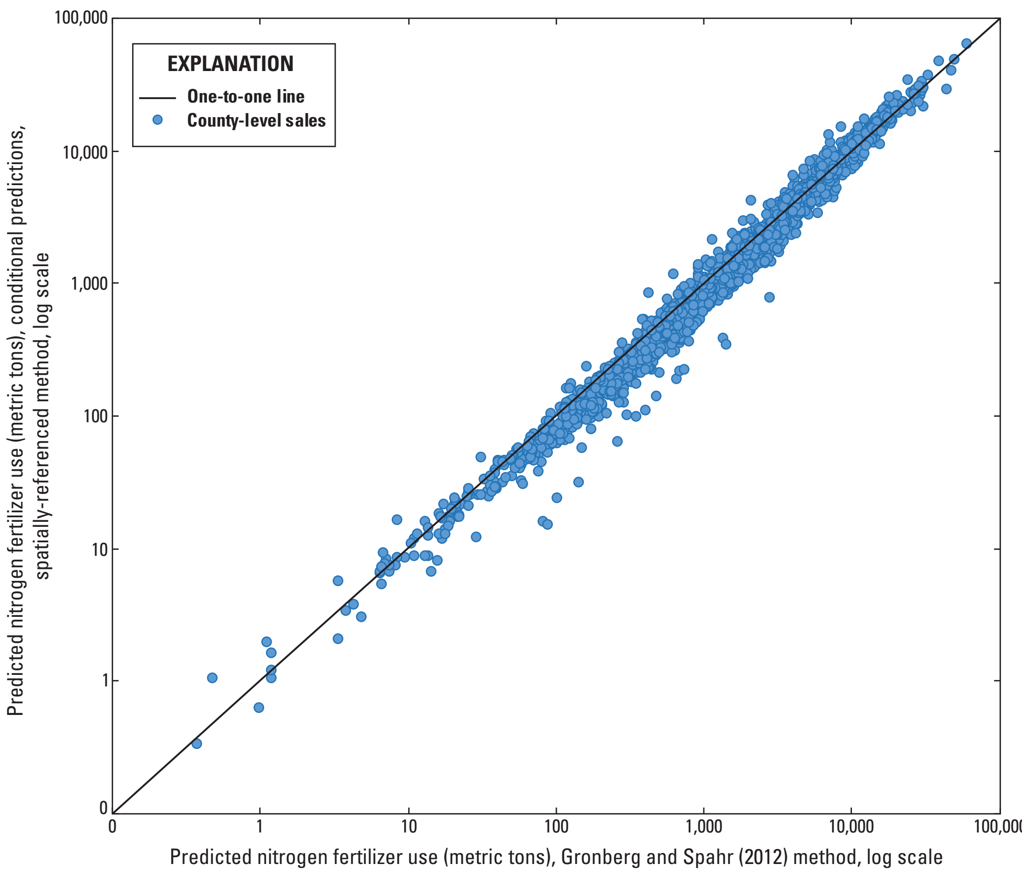


Figure 18. Predicted 2012 nitrogen fertilizer use from the conditional, spatially referenced method aggregated to the county level compared to 2012 county predictions from Brakebill and Gronberg (2017) using the Gronberg and Spahr (2012) method.

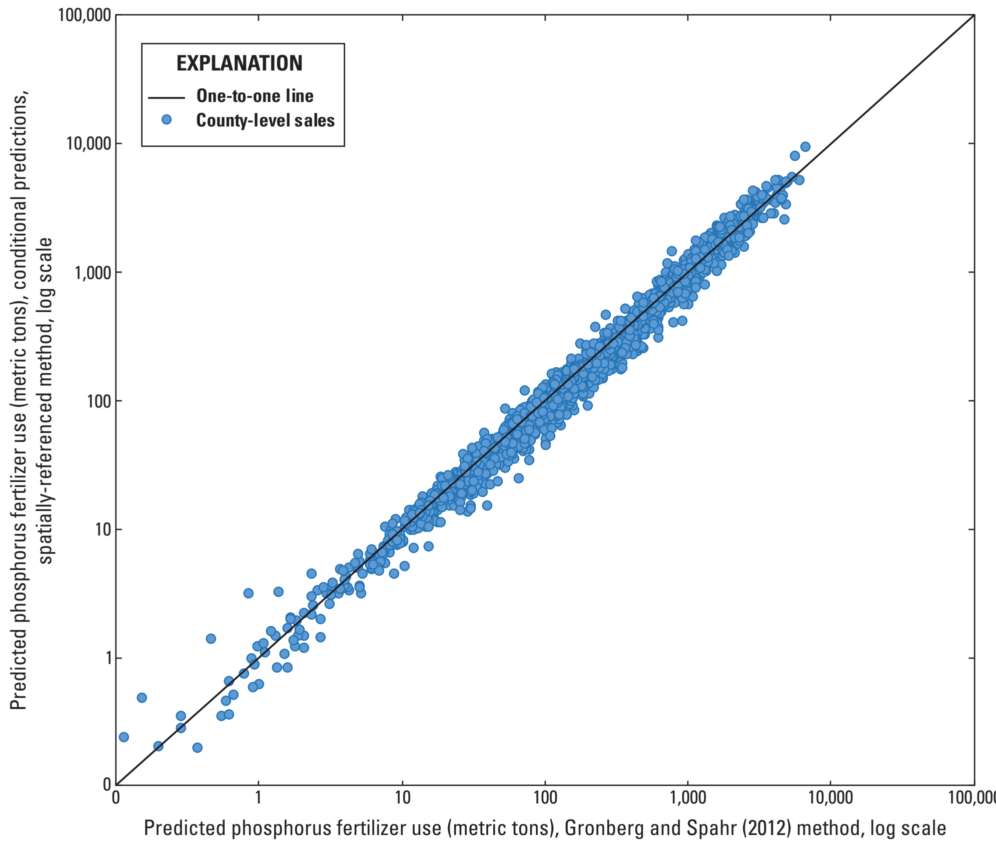


Figure 19. Predicted 2012 phosphorus fertilizer use from the conditional, spatially referenced method aggregated to the county level compared to 2012 county predictions from Brakebill and Gronberg (2017) using the Gronberg and Spahr (2012) method.

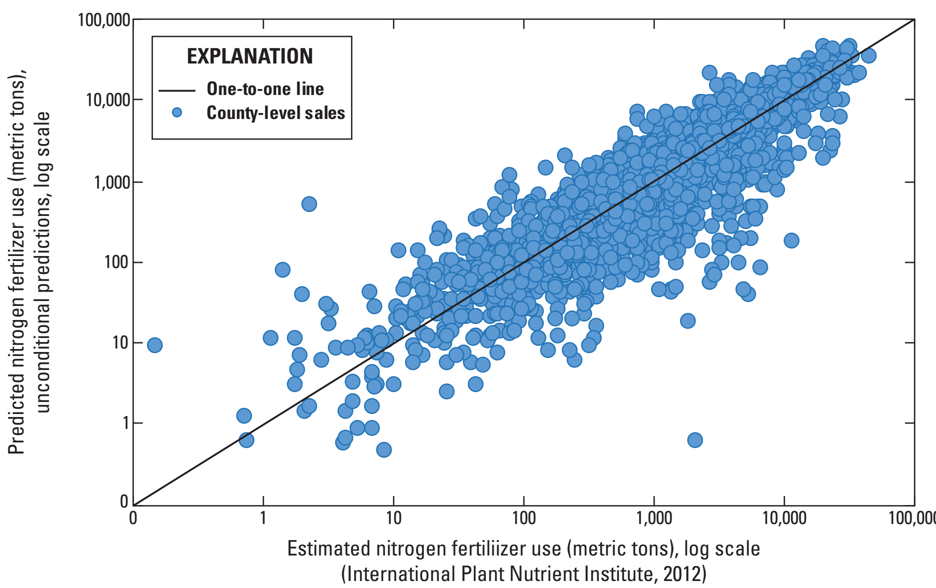


Figure 20. Predicted 2012 nitrogen fertilizer use from the unconditional, spatially referenced method aggregated to the county level compared to 2012 county estimates produced by the International Plant Nutrition Institute (2012).

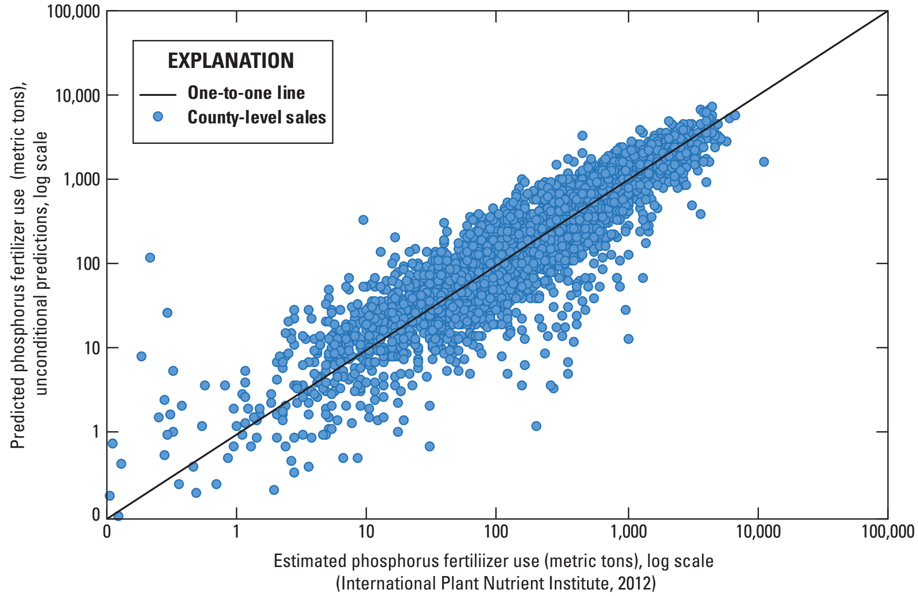


Figure 21. Predicted 2012 phosphorus fertilizer use from the unconditional, spatially referenced method aggregated to the county level compared to 2012 county estimates produced by the International Plant Nutrition Institute (2012).

The nitrogen unconditional predictions (fig. 22), with a Pearson correlation coefficient of 0.944, conform better to the USDA Economic Research Service (2016) estimates than the phosphorus predictions (fig. 23), with correlation of only 0.896, but both comparisons indicate a reasonable relation (table 5). The stronger relation of the unconditional predictions for nitrogen use may partly be explained by the wide range of use among the USDA selected states; however, evidence indicates that the unconditional predictions overestimate nitrogen use, as compared to the USDA Economic Research Service estimates in the largest using states. Phosphorus does not display such a pattern in bias.

Lastly, model predictions are compared to 2012 estimates of total nitrogen and P_2O_5 fertilizer applied in Mississippi counties generated from USDA, Farm Service Agency 2012 crop acreage data and Mississippi State University Extension Service fertilizer recommendations (U.S. Department of Agriculture, Farm Service Agency, 2013; Oldham, 2012). Planted crop acreage is determined for all crops by county using USDA, Farm Service Agency 2012 data. Fertilizer recommendations for nitrogen and P_2O_5 on 13 major crops are determined based on Mississippi State University Extension Service fertilizer recommendations (Oldham, 2012; Larry Oldham, written commun., 2017). The 13 crops account for at least 93 percent of all crop acreage within a given county (U.S. Department of Agriculture, Farm Service Agency, 2013). Fertilizer recommendations for each crop are determined based on soil test-based fertilizer recommendations. The recommendations are based on the amount of P_2O_5 measured in the soil compared to the amount needed by a specific crop. The P_2O_5 recommendations for soils testing medium for P_2O_5 are used in this comparison, as listed in Oldham (2012) with the exception for the Delta region of Mississippi. For the 18 counties that compose the Delta region of Mississippi, where plant-available phosphorus in the soil is high, the need for P_2O_5 from commercial fertilizer sources is lower, so the Mississippi State University Extension Service P_2O_5 recommendation (based on medium soil testing values) is reduced to one-half of the recommendation used in non-Delta counties, for comparison with model predictions. (Oldham, 2012; Larry Oldham, written commun., 2017) (table 5.1). According to soil testing, soils of the Delta region are generally high in plant-available phosphorus, so P_2O_5 fertilization is likely less than the “medium soil test” recommendations in that region than other parts of the State (Oldham, 2012; Larry Oldham, written commun., 2017). The estimated nitrogen and P_2O_5 applied to each crop is determined by multiplying the recommended application rate by the individual crop acreage for each county. For comparison, phosphorus model unconditional predictions are converted to P_2O_5, and then nitrogen and P_2O_5 are summed at the county level. Both estimates are reported in metric units.

The comparison between the unconditional, spatially referenced model predictions and the recommendation-based Mississippi estimates for nitrogen and P_2O_5 fertilizer mass applied among Mississippi counties is shown in figures 24 and 25, respectively. The Pearson correlation coefficient is 0.875 for nitrogen and 0.843 for P_2O_5 (table 5). For phosphorus, evidence indicates an upward relative bias in the unconditional predictions, particularly for the higher use counties. Given the somewhat subjective adjustment made to P_2O_5 requirements for the Delta counties, which also tend to be the larger use counties, some decline in agreement can be expected; although evidence also indicates a consistent upward relative bias among the non-Delta counties.

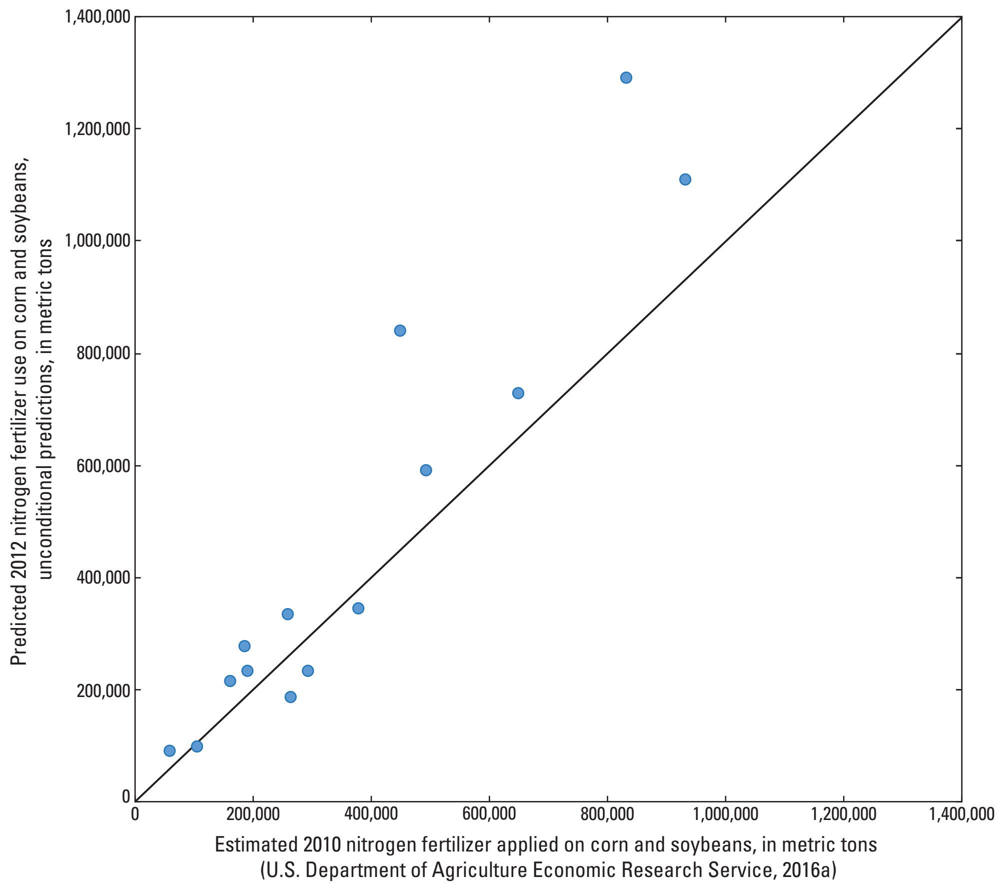


Figure 22. U.S. Department of Agriculture, Economic Research Service survey data of 2010 nitrogen fertilizer use on corn and soybeans compared to unconditional predictions of 2012 nitrogen fertilizer use on corn and soybeans (metric tons) for selected states.

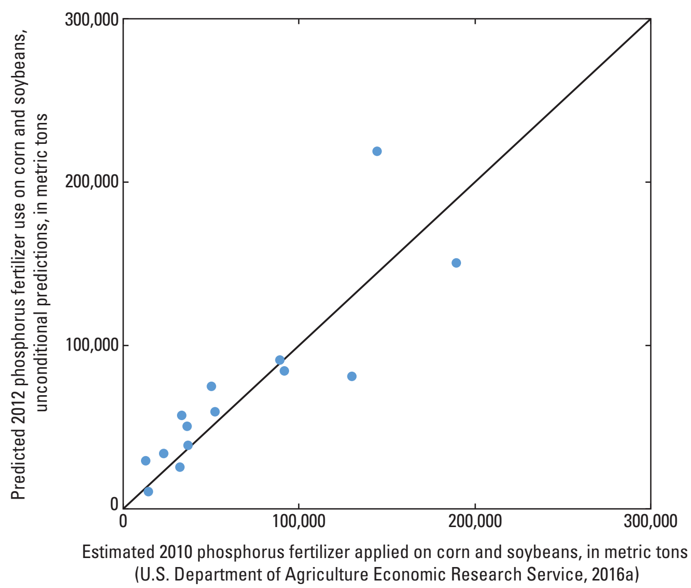


Figure 23. U.S. Department of Agriculture, Economic Research Service survey data of 2010 phosphorus fertilizer use on corn and soybeans compared to unconditional predictions of 2012 phosphorus fertilizer use on corn and soybeans (metric tons) for selected states.

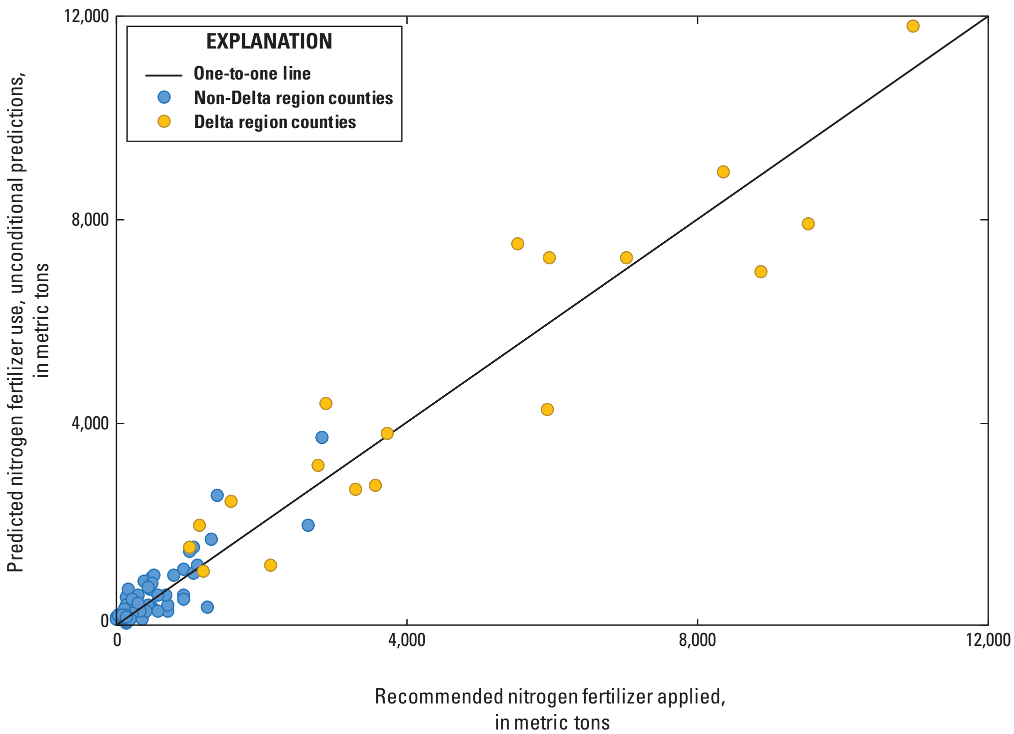


Figure 24. Recommendation-based nitrogen fertilizer estimates applied to agricultural crops in Mississippi counties compared to unconditional predictions, 2012 nitrogen fertilizer use (metric tons).

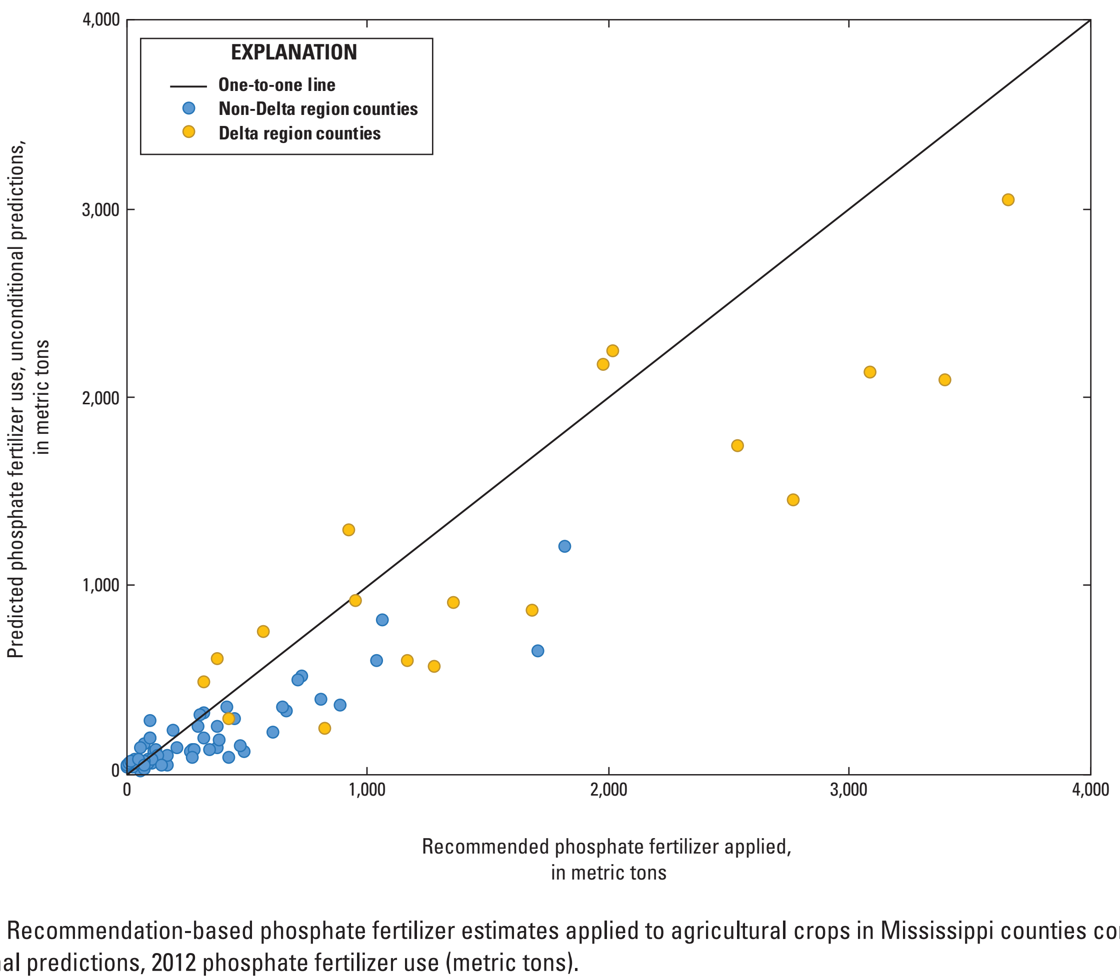
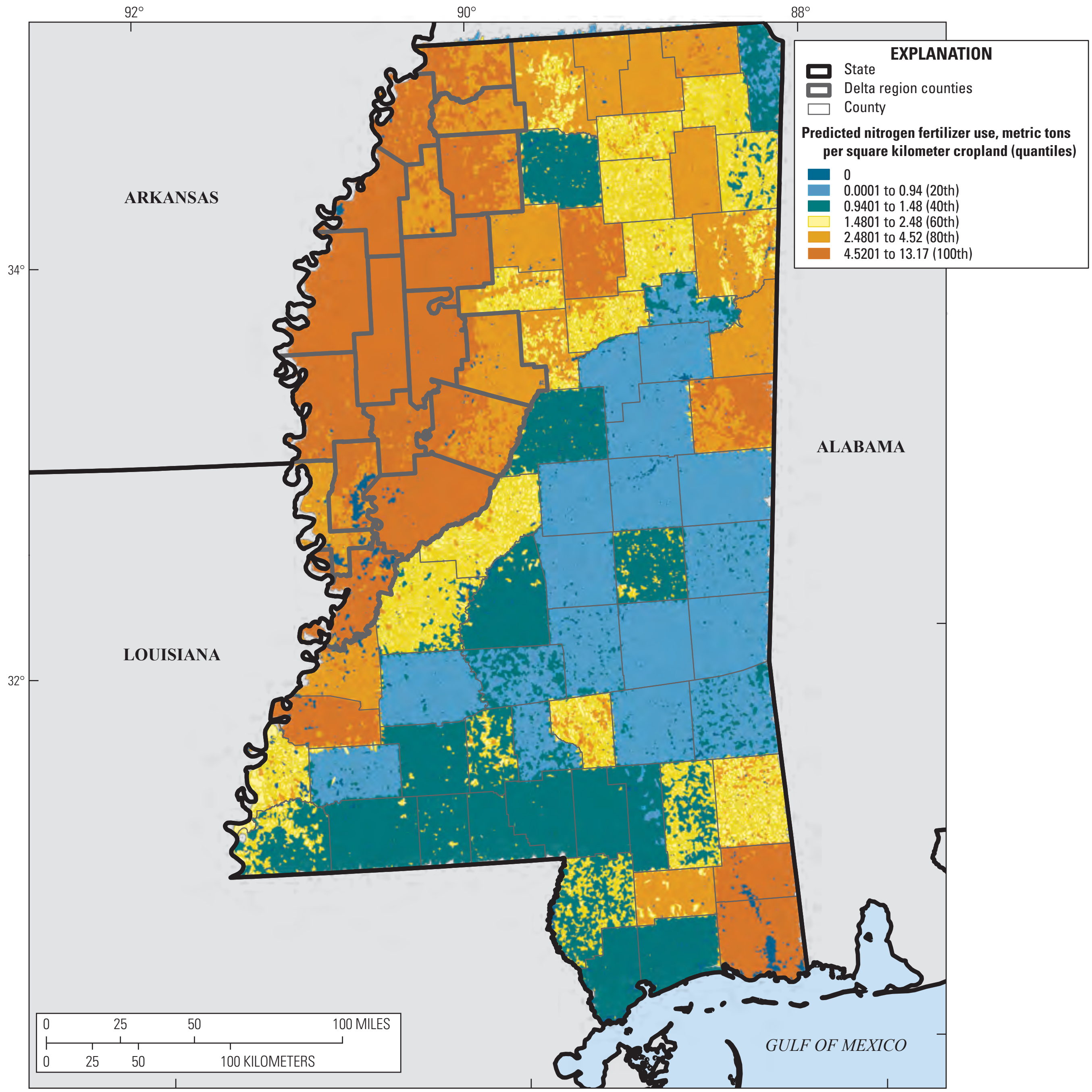


Figure 25. Recommendation-based phosphate fertilizer estimates applied to agricultural crops in Mississippi counties compared to unconditional predictions, 2012 phosphate fertilizer use (metric tons).

Maps of the unconditional predictions of nitrogen and P_2O_5 application rates for catchments in Mississippi are shown in figures 26 and 27, respectively. The predicted nitrogen to phosphorus ratio in Mississippi catchments is shown in figure 28. The highest rates of application, for both elemental components, are in counties of the Delta (counties with widened borders) and southeastern Mississippi regions, which are both major agricultural crop-growing regions of the State. Most of the non-Delta counties have lower estimates of nitrogen and P_2O_5 use, but higher nitrogen to phosphorus ratios. This result is consistent with known agricultural practices in those regions, where poultry and hog industries are prevalent, making manure a significant alternative to commercial fertilizer.



Base from National Hydrography Data Plus version 2 (NHDPlusV2), 1:100,000, 2012 County Boundaries, 1:100,000 (LaMotte, 2016)

Figure 26. Predicted 2012 nitrogen fertilizer use at the catchment scale, metric tons per square kilometer of cropland in the State of Mississippi, based on unconditional predictions from the spatially referenced model.

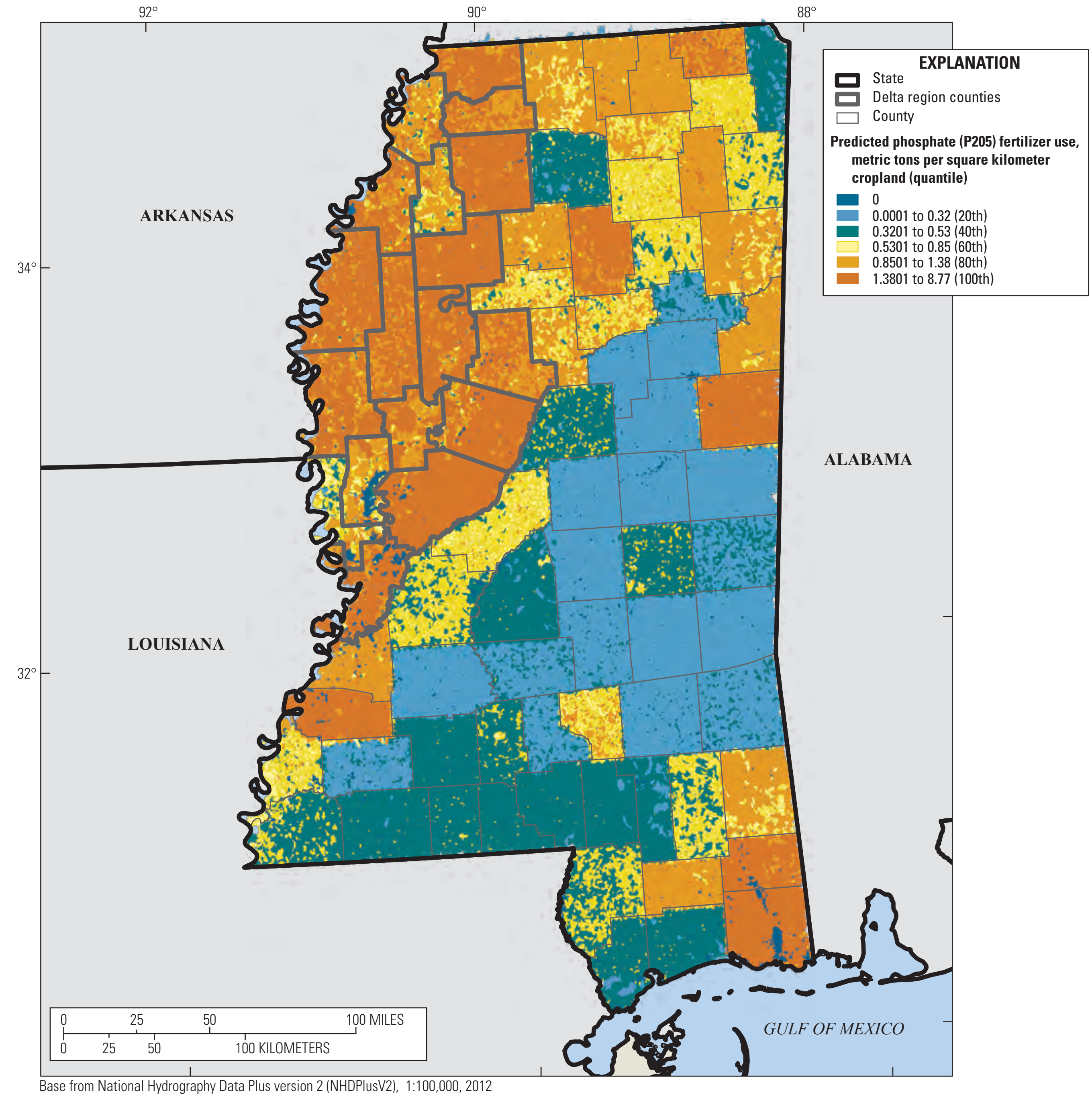


Figure 27. Predicted 2012 phosphate fertilizer use at the catchment scale, metric tons per square kilometer of cropland in the State of Mississippi, based on unconditional predictions from the spatially referenced model.

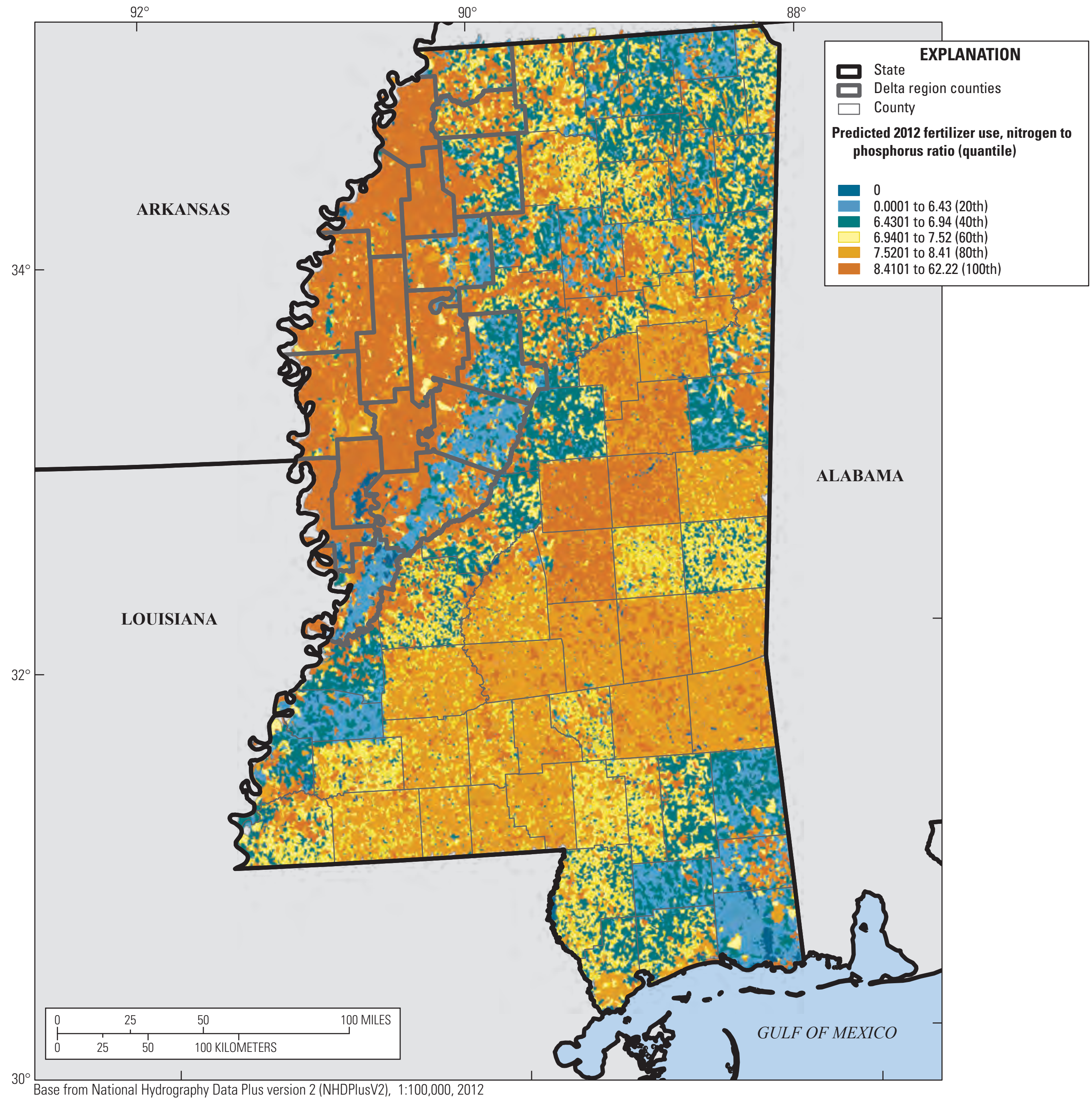


Figure 28. Predicted 2012 fertilizer use, nitrogen to phosphorus ratio for catchments in the State of Mississippi, based on unconditional predictions from the spatially referenced model.

Summary and Conclusions

Elemental nitrogen and phosphorus fertilizer use on agricultural cropland is estimated for the conterminous United States for the year 2012 at the National Hydrography Dataset Plus version 2.0 (NHDPlusV2) catchment scale. An approach is developed that uses spatially referenced statistical modeling methods to relate the Association of American Plant Food Control Officials commercial fertilizer sales data to a set of explanatory variables to produce separate estimates of nitrogen and phosphorus fertilizer use. The spatially referenced method improves upon earlier techniques by allowing for varying nitrogen to phosphorus ratios at the catchment scale, rather than assuming the nitrogen to phosphorus ratio was constant across all counties in a state and by expanding the set of variables used to allocate county-level sales data to the catchment scale.

The nitrogen model explains 74 percent and the phosphorus model 72 percent of the variation in total nitrogen or phosphorus tonnage sold at the county level, and both models, based on state-level results, are estimated to explain more than 90 percent of the variation in fertilizer use. The planted areas of five major crop groups are determined to be statistically significant determinants of fertilizer use for nitrogen and phosphorus components. Similar explanatory variables were used in both models to determine the intensity with which elemental fertilizer is applied to the crops. The common factors, including climate variables and area-normalized fertilizer expenditures, had consistent results in both models. The respective area-normalized manure variables demonstrated the importance of manure as a substitute for commercial fertilizer, particularly for the grass-pasture-hay crop group as reflected by model coefficients.

The spatial patterns of predicted nitrogen and phosphorus fertilizer use reflect known agricultural cropping patterns and practices across the United States. Total nitrogen and phosphorus tonnage use is predicted to be higher in states located in the eastern one-half of the United States and in major crop growing regions of the West. For nitrogen, the highest use (per unit area of cropland) is predicted to be in areas dominated by crops that require high application rates of nitrogen (such as corn, rice, cotton, and vegetables), and lower nitrogen use per unit area is predicted to be where grassland and pasture predominate. The highest rates of phosphorus fertilizer use, per unit area, closely follow the pattern of high nitrogen use, except for areas where soils are naturally rich in phosphorus, such as the Delta region of northwestern Mississippi or in areas where poultry production is common and phosphorus-rich manure available. Unlike previous studies, the nitrogen to phosphorus ratio in this study is not constant and reproduces recognized patterns in agricultural cropping practices. Highest values of the ratio are in the Plains states where nitrogen-fixing crops are not commonly grown. Lower ratios are in the corn-soybean crop-growing region of the Midwest where crop rotation with nitrogen-fixing crops is a common practice to supplement the nitrogen content in soil.

The estimates of this study compare favorably with other estimates of nitrogen and phosphorus use that are based on earlier U.S. Geological Survey modeling techniques and International Plant Nutrition Institute methods using 2012 Association of American Plant Food Control Officials data. Reasonable comparisons are obtained relating model predictions to application rates derived from independent sources consisting of U.S. Department of Agriculture, Economic Research Service survey data for corn and soybeans and recommendations from the Mississippi State University Extension Service.

The primary products of this study are separate predictions of nitrogen and phosphorus fertilizer use at the NHDPlusV2 catchment scale. This information will prove useful for water-quality models that estimate total-nitrogen and total-phosphorus loads to streams and for other studies requiring fertilizer-use estimates related to agricultural cropping practices in the United States. The statistical methodology used to derive the estimates allows for the determination of uncertainty of the estimates, which has not previously been assessed by other methods. The unconditional predictions of nitrogen and phosphorus fertilizer use at the NHDPlusV2 catchment scale for the conterminous United States are available for download as a U.S. Geological Survey data release (Stewart and others, 2019b).

Determining fertilizer use for other years should be possible by extending this method, for years in which Association of American Plant Food Control Officials data are available. It should be possible to extend the method to determine fertilizer application rates for other years in which Association of American Plant Food Control Officials data are available. The main complication in doing so is finding reasonable substitutes for the Cropland Data Layer (CDL) cropping data. A shortcoming of the CDL is the classification accuracy for pasture/grassland, a principle crop determining the extent of fertilizer use. A revised CDL that reduces this error could yield significant improvements in fertilizer use estimates. Alternatively, the method could be generalized to include lands that are classified as close substitutes to pasture/grassland according to the CDL error matrices. Inclusion of these lands as an additional source may improve model fit and correct possible biases in the estimated application factor for pasture/grassland. Another shortcoming of CDL is that the CDL only goes back consistently to 2009, limiting the ability to create a historical time series.

References Cited

Alexander, R.B., Smith, R.A., Schwarz, G.E., Boyer, E.W., Nolan, J.V., and Brakebill, J.W., 2008, Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin: Environmental Science & Technology, v. 42, no. 3, p. 822–830, https://doi.org/10.1021/es0716103.

Amemiya, T., 1985, Advanced econometrics: Cambridge, Mass., Harvard University Press, 521 p.

Association of American Plant Food Control Officials, 2015, Commercial fertilizer: available for purchase online at http://www.aapfco.org/publications.html.

Brakebill, J.W., and Gronberg, J.M., 2017, County-level estimates of nitrogen and phosphorus from commercial fertilizer for the conterminous United States, 1987–2012: U.S. Geological Survey data release, https://doi.org/10.5066/F7H41PKX.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J., 2011, Completion of the 2006 National Land Cover Database for the conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 77, no. 9, p. 858–864.

Gronberg, J.M., and Arnold, T.L., 2017a, County-level estimates of nitrogen and phosphorus from animal manure for the conterminous United States, 2007 and 2012: U.S. Geological Survey Open-File Report 2017–1021, 6 p., https://doi.org/10.3133/ofr20171021.

Gronberg, J.M., and Arnold, T.L., 2017b, County-level estimates of nitrogen and phosphorus from animal manure (2007 and 2012) and 30-meter-resolution grid of counties (2010) for the conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/F7X34VMZ.

Gronberg, J.M., and Spahr, N.E., 2012, County-level estimates of nitrogen and phosphorus from commercial fertilizer for the conterminous United States, 1987–2006: U.S. Geological Survey Scientific Investigations Report 2012–5207, 20 p., accessed June 12, 2018, at https://pubs.usgs.gov/sir/2012/5207/.

International Plant Nutrition Institute, 2012, A Nutrient Use Information System (NuGIS) for the U.S.: Norcross, Ga., January 12, 2012, accessed October 24, 2017, at http://www.ipni.net/nugis.

LaMotte, A.E., 2015, Selected items from the Census of Agriculture at the county level for the conterminous United States, 1950–2012: U.S. Geological Survey data release, https://dx.doi.org/10.5066/F7H13016.

LaMotte, A.E., 2016, County boundaries for selected items from the Census of Agriculture, 1950–2012 (COA_STCOFIPS): U.S. Geological Survey data release, https://dx.doi.org/10.5066/F7G44NCN.

Lanfear, K.J., 1994, 1:2,000,000-scale Counties of the United States, edition 2.3.1: U.S. Geological Survey web page, accessed June 12, 2018, at https://water.usgs.gov/GIS/metadata/usgswrd/XML/county2m.xml.

McCabe, G.J., and Wolock, D.M., 2011, Independent effects of temperature and precipitation on modeled runoff in the conterminous United States: Water Resources Research, v. 47, no. 11, accessed June 12, 2018, at https://doi.org/10.1029/2011WR010630.

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., 2012, NHDPlus Version 2—User guide, accessed June 12, 2018, at http://www.horizon-systems.com/NHDPlus/NHDPlusV2_documentation.php.

Moore, R.B., and Dewald, T.G., 2016, The road to NHDPlus—Advancements in digital stream networks and associated catchments: Journal of the American Water Resources Association, v. 52, no. 4, p. 890–900, accessed June 12, 2018, at https://doi.org/10.1111/1752-1688.12389.

Oldham, L., 2012, Nutrient management guidelines for agronomic crops grown in Mississippi: Mississippi State University Extension Service Publication 2647, 55 p., accessed June 12, 2018, at https://mssoy.org/wp-content/uploads/2012/03/NutrientManagementGuidelinesforMississippiCrops.pdf.

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: Journal of the American Water Resources Association, v. 47, no. 5, p. 891–915, accessed June 12, 2018, at https://doi.org/10.1111/j.1752-1688.2011.00577.x.

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: U.S. Geological Survey Scientific Investigations Report 2006–5012, 17 p., accessed June 12, 2018, at https://pubs.usgs.gov/sir/2006/5012/.

Sargent, T.J., 1979, Macroeconomic theory: New York, Academic Press, 404 p.

Schwarz, G.E., Hoos, A.B., Alexander, R.B., and Smith, R.A., 2006, The SPARROW surface water-quality model—Theory, application and user documentation: U.S. Geological Survey Techniques and Methods, book 6, chap. B3, 248 p., accessed June 12, 2018, at https://pubs.usgs.gov/tm/2006/tm6b3.

Smith, R.A., Schwarz, G.E., and Alexander, R.E., 1997, Regional interpretation of water-quality monitoring data: Water Resources Research, v. 33, no. 12, p. 2781–2798, accessed June 12, 2018, at https://doi.org/10.1029/97WR02171.

Stewart, J.S., Schwarz, G.E., Brakebill, J.W., and Preston, S.D., 2019a, Dataset used for estimating catchment-level nitrogen and phosphorus fertilizer use from commercial fertilizer sales data for the conterminous United States, 2012: U.S. Geological Survey data release, https://doi.org/10.5066/F73J3C50.

Stewart, J.S., Schwarz, G.E., Brakebill, J.W., and Preston, S.D., 2019b, Catchment-level predictions of nitrogen and phosphorus fertilizer use from commercial fertilizer sales data for the conterminous United States, 2012: U.S. Geological Survey data release, https://doi.org/10.5066/F7CZ36F4.

U.S. Department of Agriculture, Economic Research Service, 2016, Fertilizer use and price, updated February 21, 2018, accessed December 1, 2016, at https://www.ers.usda.gov/data-products/fertilizer-use-and-price.aspx.

U.S. Department of Agriculture, Economic Research Service, 2013, ARMS Farm Financial and Crop Production Practices, accessed September 12, 2017, at https://www.ers.usda.gov/data-products/arms-farm-financial-and-crop-production-practices/.

U.S. Department of Agriculture, Farm Service Agency, 2013, 2012 crop acreage data as of January 2013, accessed June 12, 2018, at https://www.fsa.usda.gov/news-room/efoia/electronic-reading-room/frequently-requested-information/crop-acreage-data/index.

U.S. Department of Agriculture, National Agricultural Statistics Service, 2015, 2012 Census of Agriculture highlights—Poultry and egg production: U.S. Department of Agriculture, ACH12–18, accessed June 12, 2018, at https://www.agcensus.usda.gov/Publications/2012/Online_Resources/Highlights/Poultry/Poultry_and_Egg_Production.pdf.

U.S. Department of Agriculture, National Agricultural Statistics Service, 2013, 2012 Cropland Data layer, accessed June 12, 2018, at https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php.

U.S. Department of Agriculture, National Agricultural Statistics Service, 2004, Census of agriculture historical archive, 2002 census publications, volume 1—Geographic area series: Washington, D.C., U.S. Department of Agriculture, accessed June 12, 2018, at http://agcensus.mannlib.cornell.edu/AgCensus/censusParts.do?year=2002.

U.S. Department of Agriculture, Office of the Chief Economist, 2014, Major world crop areas and climate profiles (MWCAP), North America, accessed September 2017, at https://www.usda.gov/oce/weather/pubs/Other/MWCACP/namerica.htm.

U.S. Geological Survey, U.S. Department of Agriculture Natural Resource Conservation Service, U.S. Environmental Protection Agency, and other Federal, State, and local partners, 2017, The Watershed Boundary Dataset (WBD), accessed September 6, 2017, at https://nhd.usgs.gov/wbd.html.

Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2018, Select attributes for NHDPlus version 2.1 reach catchments and modified network routed upstream watersheds for the conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/F7765D7V.

Wolock, D.M., and McCabe, G.J., 2018, Water balance model inputs and outputs for the conterminous United States, 1900–2015: U.S. Geological Survey data release, https://doi.org/10.5066/F71V5CWN.

Appendix 1. Description of the Method Used to Assign County Origin of Sales For Reported Fertilizer Sales With No County Identifier

The Association of American Plant Food Control Officials sales data include a locational category corresponding to an unknown county of origin. A signal extraction method is implemented to address this issue, allowing sales to be fully attributed to a county, or multicounty representing the grouping of counties with no sales information. The method makes use of the Census of Agriculture county fertilizer expenditure data, or county cropland data if expenditure data are not available. The methodology apportions to counties the total fertilizer sales having no assigned county, without regard to elemental composition. The elemental allocation of unassigned nitrogen and phosphorus sales is based on the total fertilizer sales apportionment.

For county *c* with reported sales, the true share of state total fertilizer sales, S_c, can be expressed in at least the following two ways: (1) in terms of the state share of expenditures, S_c, plus a random term representing net exports, u_c, assumed to have zero mean and constant variance, s_u^2; or (2) in terms of the observed (but incomplete) share of state sales, S_{0c}, plus a simple expenditure-based apportionment of the unassigned-county share of state sales, S^*, plus a zero-mean residual term having a variance that scales with S^*. Thus, the two equations describing the unknown county *c* true share of state sales are,

$$S_c = s_c + u_c \text{ and} \tag{1–1}$$

$$S_c = S_{0c} + s_c S^* + \sqrt{S^*} e_c \tag{1–2}$$

where

- S_c is the unknown county *c* state share of expenditures;
- u_c is a random term representing net exports;
- S_{0c} is observed (but incomplete) share of state sales;
- S^* is a simple expenditure-based apportionment of the unassigned-county share of state sales; and
- e_c is the normalized residual for equation 1–2 having mean zero and variance σ_ϵ^2.

If a county has no fertilizer expenditure estimate, then s_c is set equal to the county share of cropland.

Solving the two equations for true sales share, S_c, gives a relation among observable shares for county expenditures, s_c, reported sales, S_{0c}, and the share of unassigned sales, S^*, in terms of the unobserved residuals, u_i and $\sqrt{S^* e_c}$,

$$S_{0c} + s_c\left(S^* - 1\right) = u_c - \sqrt{S^* e_c} \tag{1–3}$$

where

- u_c is a random term representing net exports;
- S^* is a simple expenditure-based apportionment of the unassigned-county share of state sales; and
- e_c is the normalized residual for equation 1–2 having mean zero and variance σ_e^2.

The minimum variance unbiased linear estimate of u_c, call it $\hat{u}_c$, is given by (Sargent, 1979) as follows:

$$\hat{u}_c = k\left(S_{0c} + s_c\left(S^* - 1\right)\right) \tag{1–4}$$

where

- k is $\sigma_u^2 / \left(\sigma_u^2 + S^* \sigma_e^2\right)$;
- S_{0c} is observed (but incomplete) share of state sales;
- s_c is set equal to the county share of cropland, if a county has no fertilizer expenditure estimate; and
- S^* is a simple expenditure-based apportionment of the unassigned-county share of state sales.

To obtain an estimate of $\hat{u}_c$ requires a state estimate of *k*, which requires estimates of σ_u^2 and σ_e^2. To obtain such estimates, the observable variance is compiled across counties among states with county-level sales data of the quantity $S_{0c} + s_c(S^* - 1)$, a separate value for each state. A linear regression is conducted of these state-level variances on state values of the share of sales classified as having an unassigned county, S^*. According to equation 1–3, the state-level variance of $S_{0c} + s_c(S^* - 1)$ equals $\sigma_u^2 + S^*\sigma_e^2$, implying the intercept of the linear regression is an estimate σ_u^2 and the coefficient associated with S^* is an estimate of σ_e^2.

The results of the regression are given in table 1.1, which indicate that both coefficients are marginally statistically significant, although overall model fit is poor. The results indicate that greater variability is associated with e_c as compared to the net-export residual, u_c.

Given estimates of σ_u^2 and σ_e^2, which gives a state estimate of k, equation 1–4 can be evaluated for every county and subsequently substituted into equation 1–1 to obtain an estimate of county sales inclusive of unassigned sales, $\hat{S}_c$. Let p_c represent the proportion of state county-unassigned sales assigned to county (or multicounty) c. This proportion is computed according to

$$p_c = \frac{\hat{S}_c - S_{0c}}{S^*} = \frac{(1-k)(s_c - s_{0c}) + k s_c S^*}{S^*} \qquad (1\text{–}5)$$

where

k is $\sigma_u^2 / \left(\sigma_u^2 + S^* \sigma_e^2\right)$;

s_c is the county-c share of state expenditures, set equal to the county share of cropland, if a county has no fertilizer expenditure estimate;

S_{0c} is observed (but incomplete) share of state sales; and

S^* is a simple expenditure-based apportionment of the unassigned-county share of state sales.

Because this apportion factor can be either positive or negative, the factor is constrained to be nonnegative then factors are renormalized so that the sum of p_c across all counties in the state is 1. Thus, the final proportion used to assign county-unassigned sales to specific counties is given by

$$\bar{p}_c = \max\left(p_c, 0\right) / \Sigma_i \max\left(p_c, 0\right) \qquad (1\text{–}6)$$

where

p_c is the proportion of state county-unassigned sales assigned to county (or multicounty) c.

This same proportion is applied to assign county-unassigned sales of nitrogen and phosphorus to specific counties.

Table 1.1. Results of the regression of state estimates of the variance across counties of the quantity defined by the reported county share of total state fertilizer sales plus the product of the county share of state fertilizer expenditures and the state share of county assigned sales ($S_{0c} + s_c(S^* - 1)$) on the state share of sales unassigned to a county (S^*) where * represents an individual state. The intercept of this regression is an estimate of the variance of net exports, σ_u^2, and the slope coefficient associated with S^* is an estimate of the variance σ_e^2.

[t value, test statistic that is the ratio of the parameter estimate to its standard error; >, greater than; NA, not applicable]

Variable	Parameter estimate	Standard error	t value	Probability > absolute value of t value
Intercept	0.00024	0.00016	1.55	0.1359
Unassigned county sales share (S*)	0.00369	0.00232	1.59	0.1271
Model diagnostics				
Root mean square error	0.0006072	NA	NA	NA
Coefficent of determination	0.1025	NA	NA	NA
Number of observations	24	NA	NA	NA

Appendix 2. Summary of Information For Cropland Data Layer 2012 Crops by Crop Group Used in the 2012 Nitrogen and Phosphorus Fertilizer-Use Models (U.S. Department of Agriculture, National Agricultural Statistics Service, 2013)

Table 2.1. Summary of information for Cropland Data Layer 2012 crops by crop group used in the 2012 nitrogen and phosphorus fertilizer-use models (U.S. Department of Agriculture, National Agricultural Statistics Service, 2013).

Cropland Data Layer 2012, crop(s)	Cropland Data Layer 2012, crop code	Percent of total cropland
Corn crops		
Corn	1	13.241
Double crop—barley and corn	237	0.005
Double crop—corn and soybeans	241	0.002
Double crop—oats and corn	226	0.016
Double crop—winter wheat and corn	225	0.058
Popcorn or ornamental corn	13	0.017
Sweet corn	12	0.042
Grass-pasture-hay crops		
Grass and pasture	176	53.596
Other hay and nonalfalfa hay	37	3.329
Miscellaneous crops		
Almonds	75	0.164
Apples	68	0.062
Apricots	223	0.001
Asparagus	207	0.003
Blueberries	242	0.013
Broccoli	214	0.002
Buckwheat	39	0.003
Cabbage	243	0.003
Camelina	38	0.001
Caneberries	55	0.002
Canola	31	0.237
Cantaloupes	209	0.003
Carrots	206	0.006
Cauliflower	244	0.000
Celery	245	0.000
Cherries	66	0.028
Chick peas	51	0.000
Christmas trees	70	0.009
Citrus	72	0.020
Cotton	2	1.835
Cranberries	250	0.005
Miscellaneous crops—Continued		
Cucumbers	50	0.005
Double crop—lettuce and cantaloupe	231	0.001
Double crop—lettuce and cotton	232	0.001
Double crop—lettuce and barley	233	0.000
Double crop—lettuce and durum wheat	230	0.006
Double crop—soybeans and cotton	239	0.001
Double crop—winter wheat and cotton	238	0.045
Dry beans	42	0.244
Eggplants	248	0.000
Flaxseed	32	0.040
Garlic	208	0.002
Gourds	249	0.000
Grapes	69	0.161
Greens	219	0.002
Herbs	57	0.015
Honeydew melons	213	0.001
Hops	56	0.004
Lentils	52	0.054
Lettuce	227	0.004
Mint	14	0.001
Miscellaneous vegetables and fruits	47	0.007
Mustard	35	0.005
Nectarines	218	0.000
Olives	211	0.006
Onions	49	0.020
Oranges	212	0.143
Other crops	44	0.024
Other tree crops	71	0.010
Peaches	67	0.007
Peanuts	10	0.231

Table 2.1. Summary of information for Cropland Data Layer 2012 crops by crop group used in the 2012 nitrogen and phosphorus fertilizer-use models (U.S. Department of Agriculture, National Agricultural Statistics Service, 2013).—Continued

Cropland Data Layer 2012, crop(s)	Cropland Data Layer 2012, crop code	Percent of total cropland
Miscellaneous crops—Continued		
Pears	77	0.004
Peas	53	0.108
Pecans	74	0.056
Peppers	216	0.003
Pistachios	204	0.029
Plums	220	0.008
Pomegranates	217	0.003
Potatoes	43	0.155
Prunes	210	0.000
Pumpkins	229	0.003
Radishes	246	0.001
Rape seed	34	0.000
Rice	3	0.371
Safflower	33	0.021
Sod and grass seed	59	0.111
Squash	222	0.003
Strawberries	221	0.006
Sugarbeets	41	0.173
Sugarcane	45	0.143
Sunflower	6	0.224
Sweet potatoes	46	0.012
Switchgrass	60	0.001
Tobacco	11	0.016
Tomatoes	54	0.050
Turnips	247	0.000
Walnuts	76	0.048
Watermelons	48	0.005

Cropland Data Layer 2012, crop(s)	Cropland Data Layer 2012, crop code	Percent of total cropland
Nitrogen-fixing crops		
Alfalfa	36	2.273
Clover and wildflowers	58	0.020
Soybeans	5	9.720
Vetch	224	0.001
Small grain crops		
Barley	21	0.403
Double crop—barley and sorghum	235	0.002
Double crop—barley and soybeans	254	0.014
Double crop—durum wheat and sorghum	234	0.001
Double crop—soybeans and oats	240	0.002
Double crop—winter wheat and sorghum	236	0.054
Double crop—winter wheat and soybeans	26	0.739
Durum wheat	22	0.261
Millet	29	0.065
Oats	28	0.180
Other small grains	25	0.001
Rye	27	0.063
Speltz	30	0.000
Sorghum	4	0.875
Spring wheat	23	1.721
Triticale	205	0.022
Winter wheat	24	4.861

Appendix 3. Accuracy of Crop-Group Classification in the 2012 Cropland Data Layer (U.S. Department of Agriculture, National Agricultural Statistics Service, 2013)

Table 3.1. Accuracy of crop-group classification in the 2012 Cropland Data Layer (U.S. Department of Agriculture, National Agricultural Statistics Service, 2013).

[Accuracy is assessed by aggregation of Cropland Data Layer error matrices for individual states (or group of states), and aggregation of individual classes into the crop groups defined in appendix 2; the nonagriculture class represents all classes in the Cropland Data Layer not included in the classes appearing in appendix 2. Producer accuracy is given by the number of cases with correctly classified reference classes (the diagonal elements of the error matrix) divided by the sum of cases having the given reference class (the column sum); user accuracy is the number of cases with correctly classified reference classes divided by the sum of cases of the given classification (the row sum). Overall accuracy (reported in the row labeled "Sum") is the sum of the diagonal elements divided by the sum of all cases. NA, not applicable]

Classified classes	Reference classes						
	Corn	Grass-pasture-hay	Nitrogen-fixing	Small grains	Other	Nonagriculture	Sum
Corn	38,246,605	295,169	892,885	296,300	159,137	359,300	40,249,396
Grass-pasture-hay	555,077	24,974,287	1,541,971	1,149,371	379,012	70,302,329	98,902,047
Nitrogen-fixing	811,486	961,188	31,457,247	452,519	316,645	673,648	34,672,733
Small grains	217,325	307,184	356,315	35,457,949	454,375	709,349	37,502,497
Other	85,143	94,378	231,492	337,649	15,160,464	215,813	16,124,939
Nonagriculture	902,737	80,264,218	1,652,383	1,884,057	749,366	106,401,146	191,853,907
Sum	40,818,373	106,896,424	36,132,293	39,577,845	17,218,999	178,661,585	419,305,519

Classified classes	Accuracy (percent)			
	All classes		Agriculture only	
	Producer	User	Producer	User
Corn	93.7	95.0	95.8	95.9
Grass-pasture-hay	23.4	25.3	93.8	87.3
Nitrogen-fixing	87.1	90.7	91.2	92.5
Small grains	89.6	94.5	94.1	96.4
Other	88.0	94.0	92.1	95.3
Nonagriculture	59.6	55.5	NA	NA
Sum	60.0		93.6	

Appendix 4. Derivation of the Best Linear Unbiased (Conditional) Estimates of Fertilizer Use

The Best Linear Unbiased Estimator (BLUE) prediction of fertilizer use, as described in equation 16, in the main section of this report, requires an estimate of the covariance matrices $\sum_{js(j)}$ and $\sum_{s(j)}$; the covariances between the catchment/county j multiplier, m_j; and the vector of multipliers associated with the reporting units of state $s(j)$, $\tilde{\boldsymbol{m}}_{s(j)}$. The following describes how these covariances are evaluated given estimates of the use and net-export residual components variances and covariances. Also, derived is a proof that the prediction equation 16 for these multipliers is BLUE.

The covariance matrix, $\sum_{s(j)}$, represents the covariances of the vector $\tilde{\boldsymbol{m}}_{s(j)} - \tilde{\mu}_{s(j)}\mathbf{i}$, where $\mathbf{i}$ is a vector of ones and $\tilde{\mu}_{s(j)}$ is a scaler equal to the unconditional expectation of the elements of $\tilde{\boldsymbol{m}}_{s(j)}$, which under the assumptions of the analysis are all equal to the same value. If state $s(j)$ consists of county-level reporting units, then the county-c element of $\tilde{\boldsymbol{m}}_{s(j)}$ is $\exp(\eta_{s(j)} + \epsilon_c + x_c)$, where the three error components are jointly independent and have the following assumed properties: $\eta_{s(j)}$ is a state-specific use error component, having variance σ^2_η, which is common to all catchments and counties within state $s(j)$ but independent across states; ϵ_c is a county-specific use error component, having variance σ^2_ϵ, which is common to all catchments within the same county, independent across nonneighboring counties or counties in different states, but has correlation ρ_ϵ with counties and their catchments that neighbor county c; and x_c is a county-specific net-export error component, having variance σ^2_x, which is common to all catchments within the same county, independent across nonneighboring counties or counties in different states, but has ρ_x correlation with counties and their catchments that neighbor county c. Conversely, if state $s(j)$ consists of only a state-level observation, then $\tilde{\boldsymbol{m}}_{s(j)}$ is a scaler and approximately equals (see equation 6 in the main text of this report) $\exp(\eta_{s(j)} + \sum_{c\epsilon C(s(j))} w_c \epsilon_c)$, where is the county-$c$ share of state $s(j)$ fertilizer use. This specification reflects the assumption that the net-export component is not present at the state level.

Under the assumption of normality, the unconditional expectation of $\tilde{\boldsymbol{m}}_{s(j)}$ in the case of county-level reporting units is

$$\tilde{\mu}_{s(j)} = \exp\left(\left(\sigma^2_\eta + \sigma^2_\varepsilon + \sigma^2_x\right)/2\right) \quad (4\text{–}1)$$

which is the same for all states with county-level reporting units.

For a state-level reporting unit, the assumption of normality implies the unconditional expectation of $\tilde{\boldsymbol{m}}_{s(j)}$ is

$$\tilde{\mu}_{s(j)} = \exp\left(\left(\sigma^2_\eta + \sigma^2_\varepsilon \Sigma_{c_1 \in C(s(j))}\left(w^2_{c_1} + \rho_\varepsilon w_{c_1} \Sigma_{c_2 \in M(c_1) \cap C(s(j)) w_{c_2}}\right)\right)/2\right) \quad (4\text{–}2)$$

where

$M(c_1)$ is the set of neighboring counties to county c_1 in the same state.

This evaluation varies across states with different shares and county neighbor configuration.

If state s(j) has county-level observations, then the (a, b) element of $\sum_{s(j)}$ takes the form

$$\Sigma_{s(j)a,b} = \tilde{\mu}^2_{s(j)}\left(\exp\left(\sigma^2_\eta + \delta_{a=b}\left(\sigma^2_\varepsilon + \sigma^2_x\right) + \delta_{c_b \in M(c_\alpha)}\left(\rho_\varepsilon \sigma^2_x + \rho_x \sigma^2_x\right)\right) - 1\right) \quad (4\text{–}3)$$

where

$\tilde{\mu}_{s(j)}$ is given by equation 4–1;

$\delta_{\alpha=b}$ is 1 if the element is on the diagonal (both indices refer to the same county), and 0 otherwise; and

$\delta_{c \in M(c_\alpha)}$ is 1 if the county referenced by index b, c_b, is a neighbor of the county referenced by the index α, c_α, and 0 otherwise.

Conversely, if state $s(j)$ has only a state-level observation, then $\mathbf{\Sigma}_{s(j)}$ consists of a single element and

$$\Sigma_{s(j)} = \tilde{\mu}^2_{s(j)}\left(\tilde{\mu}^2_{s(j)} - 1\right) \quad (4\text{–}4)$$

where

$\tilde{\mu}_{s(j)}$ is given by equation 4–2.

The evaluation of the row vector $\mathbf{\Sigma}_{js(j)}$ can similarly be characterized in terms of δ conditional factors, where the a-th element of the vector is given by

$$\Sigma_{js(j)\alpha} = \mu\tilde{\mu}_{s(j)}\left(\exp\left(\sigma^2_\eta + \delta_{c_\alpha = c(j)}\sigma^2_\varepsilon + \delta_{c_\alpha \in M(c(j) \cap C(s(j))) \rho_\varepsilon \sigma^2_\varepsilon}\right) - 1\right) \quad (4\text{–}5)$$

where

$\tilde{\mu}_{s(j)}$ is given by equation 4–1;

$\delta_{c_\alpha = c(j)}$ is equal to 1 if element α refers to the same county in which catchment/county j is located, and 0 otherwise; and

$\delta_{c_\alpha \in M(c(j))}$ is 1 if element α refers to a neighboring county to that which contains catchment/county j, and 0 otherwise.

The situation where state $s(j)$ has only a state-level observation implies the covariance matrix is a scaler and is evaluated as

$$
\begin{aligned}
\Sigma_{js(j)} &= E\left[\exp\left(2\eta_{s(j)} + \in_{c(j)} + \Sigma_{c\in C(s(j))} w_c \in_c\right)\right] - \mu\tilde{\mu}_{s(j)} \\
\Sigma_{js(j)} &= E\left[\exp\left(2\eta_{s(j)} + \Sigma_{c\in C(s(j))} w_c^{*} \in_c\right)\right] - \mu\tilde{\mu}_{s(j)} \\
\Sigma_{js(j)} &= \exp\left(2\sigma_\eta^2 + \left(\sigma_\in^2 \Sigma_{c\in C(s(j))} w_c^{*2} + \rho_\in \sigma_\in^2 \Sigma_{c_1\in C(s(j))} w_{c_1}^{*} \Sigma_{c_2\in M(c_1)\cap C(s(j)) w_{c_2}^{*}}\right)/2\right) - \mu\tilde{\mu}_{s(j)} \\
\Sigma_{js(j)} &= \exp\begin{pmatrix} 2\sigma_\eta^2 + \sigma_\in^2\left(1 + \Sigma_{c_1\in C(s(j))}\left(w_{c_1}^2 + \rho_\in w_{c_1} \Sigma_{c_2\in M(c_1)\cap C(s(j))} w_{c_2}\right)\right)/2 + w_{c(j)} + \\ \rho_\in \Sigma_{c\in M(c(j)\cap C(s(j)) w_c)} \end{pmatrix} \\
\Sigma_{js(j)} &= \mu\tilde{\mu}_{s(j)}\left(\exp\left(\sigma_\eta^2 + \sigma_\in^2\left(w_{c(j)} + \rho_\in \Sigma_{c\in M(c(j))\cap C(s(j))} w_c\right)\right) - 1\right)
\end{aligned}
\tag{4–6}
$$

where

$\tilde{\mu}_{s(j)}$ is given by equation 4–2; and
w_c^* is equal to $w_c + 1$ if $c = c(j)$ and equals w_c otherwise.

The proof that $\hat{m}_j$ defined by equation 16 (in the main text of this report) is BLUE is as follows. Let the general linear unbiased estimator of m_j be given by modifying the prediction equation 16,

$$\hat{m}_j = \mu + \left(\boldsymbol{B} + \Sigma_{js(j)} D\left(\boldsymbol{W}_{s(j)}\right)\left(D\left(\boldsymbol{W}_{s(j)}\right)\Sigma_{s(j)} D\left(\boldsymbol{W}_{s(j)}\right)\right)^{-1}\right)\left(\boldsymbol{F}_{s(j)} - \tilde{\mu}\boldsymbol{W}_{s(j)}\right) \tag{4–7}$$

where

$\boldsymbol{B}$ is an arbitrary row vector.

To simplify notation, let $\boldsymbol{\Sigma}_j \equiv \boldsymbol{\Sigma}_{js(s)}$, $\boldsymbol{\Sigma}_s \equiv \boldsymbol{\Sigma}_{s(j)}$, $\tilde{\boldsymbol{m}}_s = \tilde{\boldsymbol{m}}_{s(j)}$, and $D_w \equiv D(\boldsymbol{W}_{s(j)})$. The difference between the true multiplier, m_j, and the predicted value is

$$m_j - \hat{m}_j = m_j - \mu - \left(\boldsymbol{B} + \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1}\right) D_w \left(\tilde{\boldsymbol{m}}_s - \tilde{\mu}\mathbf{i}\right) \tag{4–8}$$

where

$D(\boldsymbol{W}_{s(j)})\tilde{\boldsymbol{m}}_{s(j)}$ is substituted for $\boldsymbol{F}_{s(j)}$.

The prediction variance is given by

$$
\begin{aligned}
V\left[\hat{m}_j\right] &= E\left[\left(m_j - \mu\right)^2\right] - 2\left(\boldsymbol{B} + \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1}\right) D_w E\left[\left(m_j - \mu\right)\left(\tilde{\boldsymbol{m}}_s - \tilde{\mu}\mathbf{i}\right)\right] + \\
&\quad \left(\boldsymbol{B} + \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1}\right) D_w E\left[\left(\tilde{\boldsymbol{m}}_s - \tilde{\mu}\mathbf{i}\right)\left(\tilde{\boldsymbol{m}}_s - \tilde{\mu}\mathbf{i}\right)'\right] D_w \\
&\quad \left(\boldsymbol{B} + \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1}\right)' \\
V\left[\hat{m}_j\right] &= \sigma_m^2 - 2\left(\boldsymbol{B} + \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1}\right) D_w \Sigma_j{}' + \\
&\quad \left(\boldsymbol{B} + \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1}\right) D_w \Sigma_s D_w \left(\boldsymbol{B} + \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1}\right)' \\
V\left[\hat{m}_j\right] &= \sigma_m^2 - \Sigma_j D_w \left(D_w \Sigma_s D_w\right)^{-1} D_w \Sigma_j' + \boldsymbol{B} D_w \Sigma_s D_w \boldsymbol{B}'
\end{aligned}
\tag{4–9}
$$

where

σ_m^2 is the unconditional variance of the multiplier, m_j.

The last term in the last line of equation 4–9 is the positive-definite matrix (in this case, a scaler), $\boldsymbol{B}D_w\boldsymbol{\Sigma}_s D_w\boldsymbol{B}'$, and is strictly greater than zero if any element of $\boldsymbol{B}$ is nonzero. Therefore, because $\boldsymbol{B}$ is an arbitrary vector, to minimize prediction error variance $\boldsymbol{B}$ should be set to a vector of zeros, implying prediction equation 16 (in the main text of this report) results in the minimum prediction uncertainty, making it BLUE.

Appendix 5. Mississippi State University Extension Service Fertilizer Recommendations For Selected Crops Used to Estimate Total Nitrogen and Phosphate Fertilizer Applied to Crops in Mississippi Counties (Oldham, 2012)

Table 5.1. Mississippi State University Extension Service fertilizer recommendations for selected crops used to estimate total nitrogen and phosphate fertilizer applied to crops in Mississippi counties (Oldham, 2012).

Crops	Fertilizer recommendations (pounds per acre)		
	Nitrogen	Phosphate	
	All counties	Delta region counties	Non-Delta region counties
Corn—irrigated	112	80	40
Corn—Sorghum	180	50	25
Cotton	100	40	20
Grass	80	30	15
Mixed forage	80	30	15
Peanuts	20	60	30
Rice	170	30	15
Rye	90	45	22.5
Sorghum	100	45	22.5
Soybeans—small grains	100	80	40
Soybeans	20	30	15
Sweet potatoes	35	60	30
Wheat	100	45	22.5

Publishing support provided by the U.S. Geological Survey
Science Publishing Network, Madison, Pembroke, and Sacramento
Publishing Service Center

For more information concerning the research in this report, contact the
Director, Upper Midwest Water Science Center
U.S. Geological Survey
8505 Research Way
Middleton, Wisconsin 53562
(608) 828–9901
https://wi.water.usgs.gov

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20185145

USGS

Stewart and others—**Catchment-Level Estimates of Nitrogen and Phosphorus Agricultural Use from Commercial Fertilizer Sales for the United States, 2012**—SIR 2018–5145