Chronic pollution problems are impairing water quality throughout the United States. In every state and for every major category of water (e.g., rivers and stream, lakes, wetlands), monitoring data show that serious, long term water quality problems are limiting the use of these waters for human consumption, recreation, wildlife, fishing, and other uses. Although monitoring is not comprehensive (see fig. 1), a large share of water bodies that have been assessed have water quality problems. Of the 42 percent of lakes that have been assessed, for example, more than 65 percent are impaired for at least one designated use (fig. 2).

Agriculture is a major source of sediment, nutrient, pathogen, and pesticide loads. Where water quality has been assessed, EPA data show that runoff from farms is the primary cause of water quality problems in 11 percent of rivers and streams, 9 percent of lakes, 16 percent of the bays and estuaries (fig. 2). Agriculture has been identified as a major contributor to pollution problems in specific resources of national importance including the Chesapeake Bay (USDA-NRCS, 2011) and the zone of hypoxic waters in the Gulf of Mexico (Goolsby et al., 1999). It is unlikely that water quality problems in these and many
other areas of the U.S. can be solved in the absence of better environmental performance from agriculture (Ribaudo and Johansson).

In the past 20 years, agricultural conservation programs have shifted from a narrow focus on controlling soil erosion to preserve agricultural productivity to a broader set of environmental objectives—including a major emphasis on water quality. Soil erosion is still a major concern because of water quality damage from sediment. Loss of nutrient and pathogens to water are also major concerns for agricultural water quality programs. In the Conservation Reserve Program (CRP), which provides annual payments to farmers who agree to retire cropland from production for 10 or more years, this broader focus has yielded a significant increase in the water quality benefits without increasing program costs (Feather et al. 1999). While the CRP has been the single largest U.S. conservation program since the mid 1980s, U.S. spending on incentive payments for conservation on working farmlands has increased dramatically, from less than $200 million in 2000 to nearly $2 billion in 2010. These programs encourage adoption of practices that address water quality and other environmental issues on land in crop production or grazing. Overall, the U.S. spent $5.5 billion on conservation payment programs in 2010.

A portion of ERS research is concerned with the cost-effectiveness of USDA conservation programs. From a purely environmental point of view, a cost-effective program would achieve water quality goals at the least possible cost to society. Cost-effective incentives would encourage adoption of practices that yield a high level of water quality improvement relative to implementation cost. In other words, programs would target producers and practices that would yield the biggest environmental bang for the conservation buck.

Analyzing Agricultural Water Quality Programs

While cost-effectiveness is simple in concept, it can be very difficult to achieve in practice. To understand the costs and environmental effects of a specific program, researchers must understand (1) farmer response to conservation program incentives for the adoption of conservation practices (also referred to as best management practices) and (2) how adoption of these practices would affect pollutant runoff from fields and, ultimately, water quality (Smith and Weinberg, 2004). We focus on the adoption
of management practices such as conservation tillage and nutrient management because they are increasingly important in USDA conservation programs. Conservation tillage involves reducing or eliminating tillage to keep crop residue on the soil surface slowing runoff and reducing erosion and sediment loss. Nutrient management means applying fertilizer or manure in amounts and at times that meet crop needs while minimizing nutrient loss to the environment.

Assembling the data needed to analyze the role of conservation programs in conservation practice adoption or the effect of these practices on the environment is a major challenge. As we will show, a complete dataset would draw variables from a number of disparate sources that can be difficult to link. Omitted or poorly measured variables could, of course, result in biased estimates of critical parameters such as the producer’s response to the offer of subsidy payments. Aggregated data (e.g., county averages) are sometimes substituted when farm- or field-specific data cannot be obtained. In analyzing agricultural water quality problems, however, evidence suggests that aggregated data is a poor substitute for field-specific data. Sediment and nutrients, for example, tend to originate in fields that are particularly vulnerable to soil erosion and rainfall runoff. When site-specific data on soil erosion became available 25 years ago, for example, researchers found that 53 percent of soil erosion occurred on 11 percent of cropland that was extremely vulnerable to erosion because of soil, topography, and climate conditions (AFT, 1984). Because soil characteristics and topography can vary widely within small geographic areas (such as counties), critical variation is often lost in aggregated data (Lambert et al., 2007).

While cost is a key factor in conservation practice adoption decisions, farm- or field-specific data on conservation practice adoption costs is seldom available. Biophysical data on soil characteristics, topography and climate, however, can serve as proxies for variation in conservation practice adoption cost and have been used frequently to help explain conservation practice adoption (e.g., Wu et al., 2004; Soule, 2001). Consider the example of conservation tillage (CT). On one hand, reducing or eliminating tillage can reduce production costs by saving on labor, fuel, and machinery and many farmers have adopted CT because it is profitable (Hopkins and Johansson, 2004). In warmer, dryer climates on well drained soils, moreover, conservation tillage can conserve moisture for crop growth possibly increasing
yields. In other conditions, however, CT may reduce crop yields or increase pest control costs (Sandretto and Payne, 2006). In wetter, cooler climates or on soils that tend to dry slowly, leaving residue on the soil surface can delay planting and make young crops more vulnerable to a late spring freeze. In either case, crop yields will likely be reduced.

Disaggregated data on farming operations and farmers can also be important in explaining practice adoption. While farms of all types have adopted well established practices like conservation tillage, practices that are more complex and difficult to apply appear to be adopted more frequently on larger farms where one or more people are full-time farmers (Lambert et al. 2006). For example, data from USDA field-level surveys of corn, soybean, and cotton farms show that large farms are more likely to adopt nutrient management practices when compared to smaller farms. The reason for these differences is not known, although large farms may have more to gain through careful management of production inputs, such as fertilizer, and are better able to successfully implement management intensive practices. Because most U.S. counties contain a mix of large and small farms, data aggregation can also mask important differences that could lead to differences in conservation practice adoption.

Farm and field-specific data is also important in defining the effect of conservation programs on practice adoption. Program incentives tend to be farm- and field-specific because conservation programs are flexibly designed to allow program managers to address local environmental concerns using practices that are appropriate to the biophysical characteristics of specific fields and acceptable to the producer. That is, the definition of a “best” management practice can vary, even from field-to-field within a farm. For example, controlling runoff and soil erosion on steep slopes may require relatively expensive structural practices to carry rainfall runoff from fields. On more gently sloping land, conservation tillage may be sufficient to hold soil erosion to acceptable levels. To some extent, conservation program incentives reflect these differences. That is, the level of payment incentive received for a specific practice and the likelihood of receiving any payment for a specific practice may vary with field-specific factors.

---

1 Structural soil conservation practices involve the construction of physical barriers to slow runoff and/or remove it from the field while minimizing the loss of soil depth and sediment from the field. Terraces and grassed waterways are common examples.
Practice adoption, of course, does not necessarily mean that water quality improvements will follow. Linking practice adoption and water quality is difficult because of the non-point source nature of most agricultural pollution problems. Because they are not generated at a limited number of specific locations, runoff of sediment, nutrients, pesticides, and other pollutants can be monitored only with considerable effort and expense. In lieu of monitoring, simulation models have been used extensively to link farm practices (e.g., fertilizer use) to downstream consequences (e.g., nutrient concentrations) (Smith and Weinberg, 2004). Examples include the Soil and Water Assessment Tool (SWAT) developed primarily by USDA researchers (Neitsch et al.) and the spatially referenced regression on water quality (SPARROW) model developed by U.S. Geological Survey researchers (Smith et al., 1997). These models require information on production practices and the biophysical environment—some of the same information needed to estimate models of producer response to incentive payments.

**Agricultural Data and Water Quality Analysis**

At first glance, U.S. analysts appear to be awash in data for water quality analysis (Table 1). And, in many respects that is true. The Federal government takes a full census of agricultural producers every five years, providing extensive data on land use, crop and livestock production, production expenses, government payments, producer demographic characteristics, and other aspects of agricultural production. In the intervening years, USDA surveys a nationally representative sample of farms (the Agricultural Resources Management Survey), obtaining much of the same data, and more. Also under the aegis of ARMS, crop-specific, field-level surveys (known as phase II of ARMS) provide a wealth of data on production practices, including tillage, nutrient management, pest management, irrigation management, and selected conservation practices. Regarding nutrient management, for example, farmers are asked to list all fertilizer applications giving the approximate date, type of fertilizer used, method of application, and the amount applied. Farmers are also asked about manure applications, soil tests, precision farming techniques and other practices rounding out a detailed picture of nutrient management. Major commodity crops are surveyed on rotating basis so that most crops (e.g., corn, wheat) are surveyed every 5-7 years.
In recent years, rapid growth in the availability and accuracy of data derived from remote sensing and the digitization of USDA soil surveys (the Soil Survey Geographic (SSURGO) database) have dramatically increased the amount of biophysical data available to researchers. If the exact location of a field is known, information on land use and cropping, soil quality, topography, and proximity to streams, rivers, and lakes can be obtained. The development of agricultural productivity indices (e.g., Dobos et al., 2008) based on these biophysical data have further enhanced their value in policy research. Temporally disaggregated and spatially referenced data on temperature and rainfall, available from 1950 to the present, allows creation a wide range of climate and drought measures for a wide range of spatial and temporal scales.

Finally, the Federal government also holds a great deal of administrative data collected from farmers in the process of delivering conservation programs and a host of other Federal agricultural benefits including farm income support, subsidized crop insurance, and disaster assistance. In terms of conservation programs, these data offer information on cost-sharing and incentive payments, types of practices adopted, and conservation program priorities. Local information is critical because some programs, including EQIP, allow state and county officials significant latitude in prioritizing local expenditures. Data on priorities, together with information on biophysical circumstances can help researchers understand program incentives for specific farms and fields.

Combined together, cost and production, biophysical, and administrative data would represent a rich source of information, allowing researchers to accurately define the specific context in which farmers make decisions about conservation practice adoption. At present, however, biophysical and administrative data cannot be readily merged with data from on-going surveys, including ARMS. We discuss barriers to data integration in the next section.

The paucity of fully integrated data has been a limiting factor in research on conservation practice adoption and the effect of conservation programs. While there is a large literature on conservation practice adoption, very few articles actually address the role of Federal cost-sharing and incentive payments. Cooper and Keim (1996) and Lohr and Park (1995), for example, base their studies on stated
responses to hypothetical payments for conservation practices. The empirical analysis in Lichtenberg (2004) is based on cost-sharing for structural soil conservation practices provided by the state of Maryland. Many other studies (e.g., Wu et al., 2004) use simulation models to estimate the effect of payments which lower costs for production systems that include conservation practices. These studies consider the role of hypothetical payments in leveraging practice adoption rather than payments actually offered by the Federal government. While these studies are valuable, they do not necessarily yield information on the role of existing programs in the adoption of conservation practices.

ERS researchers have investigated the bias that results from use of aggregated data on soil biophysical condition in place of site-specific data. A special version of the ARMS field-level (phase II) survey, administered to a subsample of the wheat and corn farms in the 2004 and 2005 survey samples, respectively, was designed to include site-specific biophysical information. Each surveyed field contained a National Resources Inventory (NRI) point (USDA-NRCS, 2009). The NRI contains extensive data on land use, land quality, land condition, and other factors for more than 800,000 points of non-federal land at five year intervals beginning in 1982. At each of these points in time, estimates of annual soil erosion and contributing factors are available for each NRI point.2

Using these data, ERS researchers explored the effect of using county or watershed averages for key biophysical variables on estimates of the probability that farmers will adopt soil conservation plans3. Soil erosion tends to be concentrated on relatively small area of land that is particularly vulnerable to soil erosion. When erosion rates are averaged to the county or watershed, however, soil erosion (and the potential for sediment runoff) is significantly underestimated for fields where erosion rates are high (i.e.,

---

2 The survey was part of the Conservation Effects Assessment Program (CEAP), a one-time effort carried out largely by the USDA Natural Resources Conservation Service (NRCS) to quantify the (1) environmental effect of conservation practices that are currently in use and (2) need for additional conservation treatment (USDA-NRCS, 2010). Data for the analysis was obtained by surveying farmers about production, input use, and practices for fields that contain NRI points, effectively creating a dataset with both survey and biophysical information gleaned from NRI and from other sources.

3 A soil conservation plan is a collection of practices designed to reduce soil erosion. In most cases, plans are designed to reduce erosion to 5 tons per acre per year. Highly erodible land was removed from the sample because these fields are required to have conservation plans to be eligible for farm income support and other agricultural payments.
the fields of greatest policy concern). The special survey data allowed researchers to quantify the extent of information lost in aggregated data (fig. 3). The county and watershed averages approximate the onsite values for only 54 and 43 percent of fields, respectively (Lambert et al, 2007). Performance is better for county data, in general, because counties are smaller than watersheds, although the difference appears to be slight (fig. 3).

Using average data as a proxy for site-specific data can also bias regression results. Lambert et al. (2007) devised logistic regression models of soil conservation plan adoption as a function of past soil erosion using site-specific, county average, and watershed average data. The use of county and watershed average data led to a significant over estimate of the effect of past soil erosion on soil conservation plan adoption. That is, the use of county- or watershed- average data could prompt researchers to conclude that farmers are more likely to take action to control soil erosion than they really are, based on the level of soil erosion occurring on their farms. Use of county-average socio-economic data (e.g., farm size, operator age) also changed the results, leading to a modest underestimate of the effect of past soil erosion on soil conservation plan adoption.

Data Integration

Data integration is the process of combining data from different sources to increase opportunities for policy research and analysis. For the analysis of U.S. agricultural water quality programs (and conservation programs in general), we seek to combine data from farm- and field-level surveys, administrative data on conservation programs, and biophysical information such as soil productivity and topography. While the special ARMS survey data demonstrated that estimation results based on site-specific biophysical data can differ significantly from estimates based on aggregated data, it is neither feasible nor necessary to concentrate survey data collection on farms and fields containing NRI points. Tracts of land (primary sampling units) containing NRI points account for about 3 percent of non-federal land in the conterminous U.S. Moreover, because USDA surveys are already long, they cannot be expanded to include an extensive set of questions on the biophysical context or program participation without imposing unacceptable burdens on survey respondents or severely limiting other lines of
questioning. Integration of existing datasets may be the only way to obtain data that represents practice adoption decisions and the biophysical and policy context in which they are made.

Although the promise of data integration is substantial, there are a number of barriers to its realization. Successful data integration may require small changes in the way all three types of data are collected and managed. Some of the barriers to data integration include (1) a lack of identifiers suitable for linking, particularly the lack accurate geo-referencing on field-level data, (2) concerns about increasing the risk of disclosure for confidential survey data, (3) “informed consent” requirements that could mean agencies would need to revise notices about uses of collected data so program applicants and survey respondents are informed about possible plans to link administrative and survey data, and (4) the fact that survey and administrative data are often collected at different spatial scales.

In a limited way, existing surveys are used to obtain biophysical and program participation information. In the ARMS field-level (phase II) surveys producers are asked whether the surveyed field (or a part of it) has been designated as “highly erodible” land or wetland and what conservation practices are in use on the field. In the 2009 (wheat) and 2010 (corn) surveys, farmers were also asked to give the year in which practices were adopted or installed, and whether a conservation program payment was received.

Biophysical information could be obtained from other sources, if the field location is known. Unfortunately, the latitude and longitude (geo-referencing) provided with the ARMS field-level (phase II) data is not accurate enough to facilitate a solid link to biophysical information such as soil-specific productivity indicators. Currently, ARMS enumerators ask producers to mark the location of surveyed fields on a highway map. The marks are then used to estimate the latitude and longitude of the fields. In a 2004 USDA study that tested the accuracy of this procedure, ARMS enumerators in Washington State collected the highway map marks but also used handheld GPS receivers to identify the actual location of surveyed fields. The GPS data indicated that geo-references collected using the highway map method were, on average, two miles from the actual field.
While accurate geo-referencing of field-level data is feasible, collecting accurate geo-references would increase the cost of data collection (for GPS receivers or other equipment), extend the interview time (to go to the field and take the GPS reading), and could increase the risk of disclosure for farm-specific data which is provided voluntarily by farmers with the promise of confidentiality. When the exact location of the field is known, there is some concern that individual farmers could be identified by people with local knowledge or through links to other USDA data. Careful handling of geo-references can minimize this risk. At ERS, for example, access to sensitive geo-reference data is restricted to a handful of individuals who make data links on behalf of the research staff, reducing the likelihood of disclosure.

Linking survey and administrative data can be somewhat more challenging. In addition to protecting the confidentiality of survey data, the use of administrative data can be limited by “informed consent” requirements. Individuals who provide data to the Federal government in the process of applying for program benefits must be told about how that information may be used. Usually, that involves (1) a public notice that a specific type of data is being collected and stored (e.g., data on conservation program contracts), how it will be used, and the general categories of users (including whether the data will be shared outside the agency) and (2) a notice of routine uses provided at the point of data collection (on the program application form, for example). The level of specificity needed to allow the use of data for research purposes and the linking of two or more administrative databases or administrative and survey data is not necessarily clear. Several Federal statutes are relevant in any given situation, some provisions of which can be (and have been) interpreted differently by different Federal agencies. At present, the legal situation is not entirely resolved, but it is possible that data use disclosure statements will need to be altered to allow the linking of survey and administrative data.

A second issue is the scale at which administrative data is maintained. In some USDA conservation programs, a farmer may have only one contract at a time, although the contract may involve multiple fields and multiple practices, making it difficult to determine which practices where applied to the surveyed field. In the Federal crop insurance program, for example, yield histories could provide
valuable information on land productivity and the variability of yields. These data, however, are identified to crop insurance “units” which are typically larger than a single field and can encompass an entire farm. One way to improve consistency across agencies and facilitate data integration is to use a common unit of observation. The USDA Common Land Unit (CLU) is a digitized set of boundaries that roughly approximate field boundaries. Increasingly, CLUs are being used as a tool for the collection, storage, and analysis of program information within USDA. The CLU could become the common unit by which all administrative data could be linked with other administrative and survey data, assuming that the location of surveyed fields can be accurately determined.

Given the existing systems for collecting and storing administrative data, small changes to field-level survey instruments could improve the quality of links to administrative data. Asking producers to provide the USDA farm and tract numbers for the surveyed field and/or the identification numbers for conservation program contracts could improve linkages. Field-specific questions about conservation program participation and payments, some of which are already a part of recent questionnaires, may also help sort out which contract provisions are applicable to surveyed fields.

Conclusion

Analysis of agricultural water quality programs is complex and data intensive and, in the U.S., analysis has been hampered by data limitations. While there is a wealth of data on practice use, biophysical conditions, and conservation programs, combining these sources into a single coherent dataset has not yet been accomplished. Although it is tempting to use aggregated data to proxy for variables that are not available at a farm or field scale, the evidence suggests that the use of aggregated data, particularly biophysical data, can lead to biased and misleading results. Of course, clarity about the incentives offered by conservation programs—the best source of which is contract data—is critical to determining their role in practice adoption. Data integration is next step in improving water quality analysis.

Despite the promise of data integration in avoiding omitted variables and aggregation bias, it may also raise additional questions about the statistical properties of combined datasets. While the ARMS-CEAP data is one of the first datasets to fully integrate survey and biophysical data, the sampling of farms
and fields was restricted those containing NRI points. It is difficult to say exactly how this limitation may have affected survey statistical properties. Because many existing sets of biophysical data have national or near-national coverage, the sampling of farms and fields is not restricted. Nonetheless, there may be statistical issues. Data collected through remote sensing, for example, may be subject to interpretation errors or limited by cloud cover at crucial points in the growing season. Likewise, administrative data is collected for program implementation and may not be retained beyond the current fiscal year or the end of the contract. Administrative data may also reflect local differences in program interpretation, which can be difficult to decipher in the absence of knowledge about local practices. Finally, even with unique identifiers designed to facilitate data linking, exact linkages are not always possible. If a higher level of data integration is achieved, identifying and quantifying the statistical effect of these types of errors may be the next step in the quest for meaningful water quality policy analysis.
Figure 1  Proportion of U.S. Waters Assessed for Use Impairment, by Type of Water

Source: ERS analysis of Environmental Protection Agency data (reporting date varies by State but is between 2002 to 2010)
Figure 2  Source of Water Quality Impairment for Assessed Waters, by Type of Water

Source: ERS analysis of Environmental Protection Agency data (reporting date varies by State but is between 2002 to 2010)
Soil loss (tons/acre/year)

0.20 0.40 0.60 0.80 1.00

0 5 10 15 20 25 30

Empirical density

Figure 3. Empirical Density Function, Soil Loss per Acre, for Site-Specific Measures and Measures Aggregated to the County and Watershed Level

Source: Lambert et al., 2007
Table 1. Selected USDA Data and Data Collection Efforts

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
<th>Resolution</th>
<th>USDA Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Census and Survey Data</strong></td>
<td>Agriculture Census</td>
<td>Complete enumeration of farms, completed once every 5 years; land use, production and sales, producer demographics, production expenses, government payments,</td>
<td>Farm</td>
</tr>
<tr>
<td></td>
<td>Agricultural Resources Management Survey (ARMS)</td>
<td>Annual survey of farms; production and sales, producer demographics, production expenses, government; Production practices in phase II survey</td>
<td>Farm and field</td>
</tr>
<tr>
<td><strong>Crop Production</strong></td>
<td>Production, acreage, and yields for crops</td>
<td>County</td>
<td>NASS</td>
</tr>
<tr>
<td></td>
<td>National Resources Inventory (NRI)</td>
<td>Data based on site visits and aerial photography at 5 year intervals; Land use, structural conservation practices, soil erosion.</td>
<td>Point of land</td>
</tr>
<tr>
<td><strong>Biophysical Data</strong></td>
<td>Cropland Data Layer</td>
<td>Land use, interpreted from satellite imagery</td>
<td>50 meter grid</td>
</tr>
<tr>
<td></td>
<td>Soil Survey Geographic Database (SURGO)</td>
<td>Detailed data on soil properties based on county soil survey data</td>
<td>sub-field</td>
</tr>
<tr>
<td>conservation program contract data</td>
<td>Information on practices, payments, and ranking index scores</td>
<td>farm/ field</td>
<td>FSA/ NRCS</td>
</tr>
<tr>
<td><strong>Administrative Data</strong></td>
<td>Crop Insurance Contract data</td>
<td>Yield histories, type of coverage, premiums paid, indemnities</td>
<td>crop insurance &quot;unit&quot;</td>
</tr>
</tbody>
</table>

Source: Economic Research Service
References


