Using Big Data to Evaluate Agro-environmental Policies

John Antle, Susan Capalbo, and Laurie Houston

JEL Classifications: D8, H4, H7, Q1, Q5
Keywords: Big Data, Environmental Policy, Farm-level Management Decisions, Landscape Management Decisions, Private-public Partnerships

Sustainable management of agricultural landscapes is a goal widely shared by farmers, policy makers, and the general public. The aim is to maintain and improve food availability and quality while also maintaining and enhancing the natural resource base. These goals are evident in calls for “climate smart agriculture,” “sustainable intensification,” managing “agro-ecosystems” to enhance “ecosystem services,” and land use policies calling for “land sparing or land sharing” (Power, 2010; Phalan et al., 2011; Garnett et al., 2013; The World Bank, 2011 and 2014). To achieve such goals, we need better information, tools, opportunities, and incentives to make decisions that maintain and enhance water and air quality, soil health, biodiversity, and human quality of life now and in the future.

The advent of mobile computing and communication devices has enhanced our ability to make these informed decisions. However, making informed decisions is still not easy. Critical information may be missing, consequences may not be readily identifiable, or there may be too much information to process efficiently and effectively. The relationship between management decisions and desired outcomes is complex and requires coordination among land managers, public institutions, private sector leaders, and others in society. Recognizing these challenges, both governmental and non-governmental organizations in the United States have established an array of data, knowledge and institutional arrangements that together constitute an “infrastructure” that supports management of agricultural landscapes. Over time, this infrastructure has evolved along with public policy towards what we will describe as “science-based policy”—that is, policy designed to achieve the goal of sustainably managing agricultural landscapes as efficiently and effectively as possible given the best-available science and technology. As growing demands are placed on agricultural ecosystems and landscapes, an infrastructure is needed for supporting a comprehensive approach to data, knowledge, and its use for sustainable landscape management.

The data infrastructure could be substantially improved by exploiting emerging technologies together with new approaches to data acquisition and utilization. The infrastructure should bring together the advances in agricultural sciences with the rapidly growing capabilities of data acquisition technologies, as illustrated by “big data” and “open data” initiatives (International Life Sciences Institute, 2013; National Science Foundation, 2013; U.S. Chamber of Commerce Foundation, 2014). The increasing utilization of precision farming and mobile technologies, together with improvements in data management software, offers expanding opportunities for an integrated data infrastructure that links farm-level management decisions to site-specific bio-physical data and analytical tools to improve on-farm management and ensure food quality. This has been recently referred to as “prescriptive farming” and is seen by the industry as an opportunity to better deal with the inherent risk in farming (U.S. Chamber of Commerce, 2014). This data can also be integrated with public data at the landscape scale for research and policy analysis. Analytical tools using data at the landscape scale could provide the improved understanding needed to support science-based policy and sustainable management.
Much of this growing volume of new data is private—for example, information about where and when agricultural operations occur, and their impacts—but there is also a growing amount of public data, such as remotely-sensed satellite data. A critical feature of the new infrastructure is that it must be able to measure, store, manage, and integrate both private and public data in ways that respect the privacy of individuals while enabling diverse stakeholders to benefit from improved information and analyses. We envisage private-public partnerships that could ease the burden of reporting, support the development and testing of data systems that improve farm-level management and food quality, and contribute to the goals of science-based policy.

Science-Based Policy and Sustainable Landscape Management

A large and growing body of scientific knowledge from agricultural, environmental, economic, and social science disciplines exists as a foundation on which a science-based policy for agriculture can be improved upon. It starts with the idea that agriculture is a “managed ecosystem” (Antle et al., 2001; Antle and Capalbo, 2002; Swinton et al., 2007). The scientific literature has established that farmers’ land management decisions affect biological and physical systems through a number of mechanisms. Some effects, such as changes in soil productivity, may be limited to the land owned by the farmer; others, such as runoff into surface waters, may appear offsite. A key insight from this body of scientific literature is that agricultural productivity depends upon and plays a key role in providing a set of “ecosystem services” such as the provision of clean water and maintenance of biodiversity (Reid et al., 2005).

Agricultural policies have two primary goals, one is to manage agricultural landscapes and the other is to improve the economic well-being of agricultural households, usually through subsidies or assistance programs. Resource efficiency and the distributional effects of policies are important and need to be taken into account in designing science-based policies. Indeed, there are inevitably trade-offs among the various private and public goals related to the management of agricultural landscapes. A fundamental role for the infrastructure needed to support science-based policy is to improve our understanding of trade-offs so that stakeholders can make informed choices among policy alternatives and their likely impacts.

Assessing Synergies and Tradeoffs among Private and Public Goals

Economics provides analytical frameworks to evaluate the need for policy interventions. The standard economic framework is “benefit-cost analysis.” Such analysis combines private outcomes such as farm income, with the value of “non-market” outcomes such as maintaining water quality and biodiversity, to determine the management strategy that yield preferred outcomes for society. To implement this benefit-cost framework both quantities and values of marketed goods and non-marketed goods are needed.

While it is straightforward to measure and value market outcomes such as the price of corn, it is difficult to quantify and value non-market outcomes. For example, contamination of water by nutrients such as nitrates may have adverse impacts on human health. It may be possible to estimate the magnitude of these effects, but it is difficult to attach a monetary value to health effects. Similarly, ecosystem services such as biodiversity are difficult to both quantify and value in monetary terms. For these reasons, strict application of the “benefit-cost analysis” approach to the design of science-based policies faces serious challenges.

An alternative to benefit-cost analysis is what we refer to as “policy tradeoff analysis” (Crissman, Antle, and Capalbo, 1998; Antle, Stoorvogel, and Valdivia, 2014). Rather than attempting to attach monetary values to ecosystem services, the tradeoff analysis approach defines a set of quantifiable economic, environmental and social “indicators” that can be used to assess the status of the agricultural landscape and outcomes associated with it. Alternative policies are evaluated in terms of the interactions among these indicators. In this approach, there is not one “solution” or best policy because different stakeholders may value tradeoffs between outcomes differently. However, the tradeoff analysis approach has the virtue of providing stakeholders with the information they need to make these value judgments.

Tools suitable for policy tradeoff analysis already are in use in some aspects of agricultural policy design. Many types of indicators have been developed for policy analysis (Bates and Scarlett, 2013). Various measures of farm household well-being are used, such as farm income and its distribution among geographic regions and among different types of farms (Antle and Huston, 2013). Measures of environmental outcomes and ecosystem services are available from direct measurements and from models, including soil quality and productivity, air and water quantity and quality, greenhouse gas emissions, and wildlife habitat. For example, the U.S. Department of Agriculture (USDA) has constructed an “environmental benefits index” to assist in the design and implementation of the Conservation Reserve Program that combines a number of different environmental indicators into a summary measure (USDA-ERS, 2006a, b, and c). These tools require the collection and analysis of data from many sources.
Data and Analytical Tools to Improve On-Farm Decisions and Science-Based Agricultural Landscape Management

The U.S. agricultural sector is becoming data-rich due to advances in mobile measurement and other technologies, but needs better data management and analytic capabilities to make use of this volume of data. Designing an infrastructure that gathers these data in one place could simplify data gathering for decision tools that improve on-farm economic and environmental performance, and enhance landscape scale modeling and analysis for improved landscape-scale management policies.

Figure 1 provides an overview of the features of farm-level data and decision tools, as well as landscape-scale data and analytical tools that support science-based policy, along with their interrelationships. While farm-level decision making and landscape-scale analysis have different purposes, they each depend on private farm-specific management data, as well as publicly available data such as weather and soil types, and prices and other publicly available economic data. A key question for the design of the agricultural data infrastructure is how both private and public data can be collected, managed, and utilized efficiently and securely.

Many farm-level data and decision tools from private and public sources are currently in use, and are evolving rapidly (see Box 1). The left-hand side of Figure 1 presents the generic structure of these tools, the data they use as inputs, and the outputs that are generated. There are also many landscape-scale models (see Box 2). The right hand side of Figure 1 shows the general structure of the data and models needed to carry out landscape-scale research and policy tradeoff analysis. A key feature of these tools is that they use both public data for prices, weather forecast, and policy information as well as private site and farm-specific input use data to generate detailed information and outcome-based data that are useful for both farm-level management decisions and landscape-scale policy decisions.

There are three broad categories of regional data: publicly available biophysical data, including down-scaled

Box 1: Examples of Farm-Level Analysis and Decision Tools

Decision Support for Agro-Technology Transfer (DSSAT) is a crop simulation model where simulated yields are based on site-specific daily weather data, soil characteristics, and crop management activities. It is used to evaluate how changes in crop characteristics, management, and environmental conditions may impact crop yields.

AgBizLogic™—formerly AgTools™—is a suite of programs such as AgBizProfit™, AgBizLease™, and AgBizFinance™, developed by Oregon State University to assess operational investment choices and management alternatives, based on an individual farm operation’s input costs and financial information.

Integrated Farming SystemsSM was developed by Monsanto that provides farmers with field-by-field recommendations for ways to increase yield, optimize inputs and enhance sustainability. It combines historical yield data, satellite imagery, field specific soil and moisture information, and plant varieties to make customized variable rate seeding prescriptions for individual fields, thus maximizing the yield potential by field.

Pioneer Field360TM Select Software was developed by DuPont Pioneer, which combines current and historical field data with real-time agronomic and weather information to help growers make informed management decisions. Growers can take notes and photos with GPS tags to track field agronomic status and have the option to share information in real time with DuPont agronomists, confidentially, for improved management recommendations.

Smartphone Apps for Agriculture

There are also myriad new smartphone apps for agriculture, ranging from farm management to commodity pricing. The FieldScout GreenIndex+ app from Spectrum Technologies, Inc. is an example of how smartphone apps can provide farmers with immediate decision management tools. With FieldScout GreenIndex+, growers take a picture of their crop, and the app computes nitrogen application rate recommendations based on a color index. This provides growers with a low-cost method for managing in-season fertility which can improve yields, lower nitrogen costs, and increase profits. All results are georeferenced, logged, and can be emailed for archiving or further analysis.
decisions, based on the latest field data from their individual fields. This involves incorporating analytical data on soil types, fungicide application timing, weather, and climate and soils data; publicly available economic data, including prices and policy information; and the confidential site- and farm-specific data obtained from producer- and industry-generated databases. Landscape management and policy tradeoff analysis models require spatially and temporally explicit data that are statistically representative of the farms and landscapes in a geographic region in order to provide reliable information about economic and environmental impacts and tradeoffs. Such data are not typically available for most of the United States. As a result, implementation of these models relies on the publicly available information on land management collected periodically through mailed questionnaires or enumerator interviews (see Box 3).

Currently available data are inadequate for various reasons. Many of
these data are collected with samples that are not statistically representative of relevant regions or populations for landscape-scale analysis; many data are not spatially or temporally explicit, are only available or released after substantial aggregation, thus limiting their usefulness, and are often available with long time lags between when the land management decisions are made, the data are collected, and when they become available for research or policy purposes. Longitudinal data such as representative samples of farms that provide data for the same farms over time are particularly important for policy research. Currently, none of the data available for research or policy analysis are longitudinal, except for the agricultural census which is conducted at five-year intervals and provides limited information about many dimensions of farm management needed for effective landscape-scale policy analysis.

Considerations for Design and Implementation of a Data Infrastructure

Two kinds of strategies could be used to create a new data infrastructure, a voluntary system or a mandatory system. A voluntary system is likely to be more politically and socially acceptable, and can generate quality data if participants are motivated to provide accurate information. Clearly identifying the mutual benefits (see Box 4), or value of involvement, to all parties will facilitate participation in such a project. This value could be a quid pro quo in the form of providing management tools and data that improve a farm’s economic and environmental performance, and also provide data valuable for product quality certification or regulatory. Another approach could be to provide financial compensation for the participants’ time. The in-kind and monetary compensation approaches could also be combined in various ways. The costs of a voluntary system could be covered, at least in part, by reducing the use of more costly paper-based survey instruments and enumerator interviews.

There would be various challenges to the implementation of a voluntary approach. First, it may not be possible to achieve the needed statistical representation of all regions and farm types needed for research and policy analysis. One way to ensure adequate representation would be to combine a voluntary system with monetary compensation for participation. Another strategy would be to require participants in voluntary government subsidy, conservation, or environmental payment programs to participate in the data system.

Privacy and security concerns are another consideration. These concerns have been the subject of recent discussions among farmers and commodity organizations as they explore the use of new technologies and big-data analytics (AGree, 2014; Cisco, 2014; Mercier, 2015). Some producers may worry about the ownership and control of their data, who will have access to it, and how it will be used. In response to these concerns the American Farm Bureau brought together a consortium of farmer organizations and agriculture data technology providers and developed data privacy and security principles to help ensure that data shared with Big Data services providers will not be misused (Plume, 2014). Though these principles are nonbinding, evidence from online financial transaction systems now in widespread use, as well as new agricultural data initiatives, suggest that data can be securely transmitted and stored electronically.

A critical issue with site-specific data is that the identity of the data source can sometimes be inferred from the location associated with the data, either because there are a small number of respondents in a spatial unit, or because spatial coordinates are associated with data. This concern can be dealt with in a variety of ways. For example, once spatially-explicit data are recorded and integrated into a database, identities of the decision maker and precise locations do not need to be known to be useful for most research and policy analysis purposes. These kinds of procedures are currently being used with confidential data such as USDA’s Agricultural Census, USDA’s Agricultural Resource Management Survey (ARMS), and the access of ARMS through the National Opinion Research Center Data Enclave, as well as with non-agricultural data such as the Census of Manufactures data collected and maintained by the Department of Commerce.
A Path Forward: Private-Public Partnerships for Better Data

We envisage innovative private-public partnerships to advance the development of a new data infrastructure for agriculture in which individuals voluntarily share and use information that is valuable for them and that can be used to promote the public interest. Much of the data needed for this new system are already being collected by individuals, the federal government and private companies, and innovative initiatives are demonstrating the feasibility of acquiring, storing and using data securely and efficiently (Box 3). Currently, various private and public entities are simultaneously engaged in development of technology and software for collecting and storing data, and for developing analytical tools. Several new startups have even developed computer systems that will enable farmers to market their data (Bunge, 2015). One of the greatest challenges is determining how data that are being collected both privately and publicly can be better coordinated to lower costs, improve quality, and more efficiently meet both private and public needs. One solution appears to be a partnership among the various organizations that have a mutual interest in assuring that the data are obtained efficiently and used appropriately for both private and public purposes. In effect, there is the need to create a “pre-competitive space” for the development of data and analytical tools that is built on the recognition that there are important public-good attributes of both data and analytical tools.

A private-public partnership for a new data infrastructure could be supported by various stakeholder organizations, including producer and industry organizations, agricultural commodity organizations, the International Life Science Institute’s Center for Integrated Modeling of Sustainable Agriculture and Nutrition Security, technology firms such as Google, Facebook, and Twitter, and also charitable organizations with an interest in agriculture such as the Bill and Melinda Gates Foundation. Governmental organizations also should be involved, including USDA’s National Institute for Food and Agriculture which funds a competitive research grant program, and research organizations promoting better public data such as the Agricultural Model Inter-Comparison and Improvement Project (AgMIP). A critical issue is how long-term funding for the creation and maintenance of the data and knowledge infrastructure will be achieved. While short-term research funding can make an important contribution, on-going support will need to be provided to create and maintain the data system.

A coordinated pilot program funded through a private-public partnership could develop and test innovative approaches to incentivize data sharing and facilitate data acquisition, management, storage and utilization. Public-domain software such as AgBizLogic™ could be made available to producers with support from agricultural extension organizations and agri-business firms. This software can be linked to a cloud-based data retrieval and storage system, such as the one being developed by private data programs like On-Farm Network and public ones like AgMIP.

For More Information


The authors thank AGree for organizing and supporting the preparation of this paper, and the Oregon Agricultural Experiment Station, the Center of Agricultural and Environmental Policy at Oregon State University, and the Regional Approaches to Climate Change in Pacific Northwest Agriculture Project supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2011-68002-30191

John Antle (john.ante@oregonstate.edu) is Professor, Applied Economics, Oregon State University, Corvallis, OR.

Susan Capalbo (susan.capalbo@oregonstate.edu) is Head and Professor, Applied Economics, Oregon State University, Corvallis, OR.

Laurie Houston (Laurie.houston@oregonstate.edu) is Senior Faculty Research Assistant, Applied Economics, Oregon State University, Corvallis, OR.