Applied Economic Perspectives and Policy (2018) volume 40, number 1, pp. 79–96.
doi:10.1093/aepp/ppx056

Submitted Article

Big Data in Agriculture: A Challenge for the Future

Keith H. Coble*, Ashok K. Mishra, Shannon Ferrell, and Terry Griffin

Keith H. Coble is the Giles Distinguished Professor, Department of Agricultural Economics, Mississippi State University. Ashok Mishra is the Kemper and Ethel Marley Foundation Chair in Food Management, W.P Carey Morrison School of Agribusiness, Arizona State University. Shannon Ferrell is an associate professor, Department of Agricultural Economics, Oklahoma State University. Terry Griffin is an assistant professor, Department of Agricultural Economics, Kansas State University.
*Correspondence to be sent to: coble@agecon.msstate.edu.

Submitted 27 March 2017; editorial decision 21 October 2017.

Abstract *This article examines the challenge and opportunities of Big Data, and concludes that these technologies will lead to relevant analysis at every stage of the agricultural value chain. Big Data is defined by several characteristics beyond size, particularly, the volume, velocity, variety, and veracity of the data. We discuss a set of analytical techniques that are increasingly relevant to our profession as one addresses these issues. Ultimately, we resolve that agricultural and applied economists are uniquely positioned to contribute to the research and outreach agenda on Big Data. We believe there are relevant policy, farm management, supply chain, consumer demand, and sustainability issues where our profession can make major contributions. The authors are thankful to the anonymous reviewers and editor Craig Gundersen for helpful comments. Support was provided by the Mississippi Agricultural and Forestry Experiment Station Special Research Initiative.*

Key words: Big Data, precision agriculture, analytical methods.

JEL codes: K11, Q12, Q16, Q18.

A variety of indicators suggest that the availability of sensors, mapping technology, and tracking technologies have changed many farming systems and the management of the food system as it flows from producers to consumers. Big Data has significant potential to address the issues of modern societies, including the needs of consumers, financial analysts, marketing agents, producers, and decision makers. While some of these information technologies have been available for some time, adoption surveys such as

Griffin et al. (2017), Schimmelpfennig (2016), Erickson and Widmar (2015), and Hennessy, Läpple, and Moran (2016) suggest continued increased rates of adoption of the various forms of these technologies.

Dyer (2016) suggests we have moved to an informational revolution in the agricultural sector. In many cases, sensor technology and data analytics from other industries are now applied to agricultural applications. Robert Fraley, Chief Technology Officer of Monsanto, has stated that "Monsanto executives are seeking to reposition the company as a business built on data science and services, as well as its traditional chemicals, seeds and genetic traits operations". The $930 million acquisition of Climate Corp in 2013 by Monsanto evidences this trend (Upbin 2013). The AgFunder Agtech Investing Report (AgFunder 2017) identifies approximately $1.4 billion of total investments in two categories in 2015, including robotics, mechanization, and other hardware, along with farm management software, sensing, and Internet of Things (IoT).

A variety of technological advances have created the opportunities of Big Data (Sonka 2015). In many cases, computational capacity both in terms of speed and volume allows for novel analyses previously not possible. First, it is now possible to conduct analysis on large volumes of data (such as weather data) and use it for actionable decision-making. Interestingly, data from multiple sources including public data, machine and sensor data, and other privately- held data are often integrated. In some applications "macro" level analysis is possible that aggregates data to provide useful industry- or market-level analysis. Conversely, data can affordably be obtained and utilized at a "micro" scale. In this case, management can occur at a site-specific or unit level such as sub-field areas. This may mean site-specific fertilization in crop agriculture, or tracking of the cuts from a beef carcass to final consumer. These trends lead to numerous discussions of the present and future impact of "Big Data," but to date those discussions have lacked a clear definition of what Big Data means. Coble et al. (2016) suggest that it refers to "large, diverse, complex, longitudinal, and/or distributed data sets generated from click streams, email, instruments, Internet transactions, satellites, sensors, video, and/or all other digital sources available today and in the future." Stubbs suggests the term *big data* as it is applied to agriculture is less about the size of the data and more about the combination of technology and advanced analytics that creates a new way of processing information in a way that is more useful and timely. Coble et al. (2016) support this approach by defining the data in terms of **volume, velocity, variety,** and **veracity,** with "volume" referring to the size of the data, "velocity" measuring the flow of data, "variety" reflecting the frequent lack of structure or design to the data, and finally "veracity" reflecting the accuracy and credibility of the data.

Information technologies provide new and useful data for decision making and analysis, therefore they naturally align with the skills and interests of applied economists. In fact, pockets of this type of economic analysis have existed for some time. For example, the use of retail scanner data (Capps 1989) has largely met the Big Data definition provided earlier. Other areas legitimately claiming the Big Data label include some large-scale ecological models, certain government-collected survey and government program data. However, we sense that many new challenges are ahead. Here are a few issues that appear imminent.

- Farm management, long a mainstay of the agricultural economics profession, has been a relatively dormant area for research in recent decades. With precision agriculture advances and adoption, new opportunities and requests are likely to confront our profession. Liu, Swinton, and Miller (2006) provide a useful case study of how precision agriculture poses new and relevant questions to our profession. In the present issue, Featherstone (2018) also provides a useful forward-looking discussion of these issues.
- Food scanner and similar data may be classified as Big Data, given their volume, velocity, and variety. Scanner data have a wide variety of applications, including research projects, program evaluations, regulatory impact analysis, and data products. Real-time store scanner data can be used to study healthy diets (Kuchler, Tegene, and Harris 2005). Food scanner data has been linked to USDA nutritional data and the USDA's National Household Food Acquisition and Purchase Survey (FoodAPS) to provide information about the food environment, such as prices and offerings at stores where the consumers did not go shopping.
- The use of geo-spatial techniques could improve modeling of crop yield and, by extension, pricing of crop insurance products (Ker and Coble 2003; Ozaki, Ghosh, and Goodwin 2008; Annan et al. 2014; Woodard and Verteramo-Chiu 2017). We expect the profession to find use of these techniques across many sub-disciplines including environmental economics (e.g., determining demand for non-production services using commercial satellite imagery data).
- Food and agricultural policy analysis is likely to evolve with new data and analytical techniques. Environmental management at a micro- and macro-scale will be enhanced. For example, nitrogen management at the sub-field level or for a major watershed will be possible with precision in places that it was not possible before. Another example is the use of low - cost commercial satellite imagery and Big Data to develop daily models of non-market values of wildernesses.
- The role of program data and government data collection will be changed in fundamental ways to reflect new data sources and analytics. In some instances, digital agriculture may allow enhanced analysis of government program data (Woodard 2016). Tack et al. (2017) discuss potential competition between public surveys and private data. This discussion is illustrated by the USDA National Agricultural Statistical Service requesting a National Academies of Sciences, Engineering, and Medicine panel review of yield and cash rent estimation methods used by the USDA. A crucial question addressed in this report was the integration of survey data with government program data and models based on imagery. (National Academies of Sciences, Engineering, and Medicine 2017).

Much of the useful Big Data produced today is in the hands of the private sector. Coble et al. (2016) suggest that the landscape of public and private farm data is likely to change and access to data for research will be a critical issue. In many cases the data are held by a variety of firms ranging from small individual farms to large corporate input suppliers. Tremblay (2017) argues that agricultural research must reach beyond Fisher's experimental design and utilize analytical techniques capable of learning from the machine and sensor data, that is, it must rely upon observation farm production data along with data from controlled experiments.

Economic Research Priorities for Agricultural Big Data

Within the changing agricultural management landscape created by advancements in Big Data, what areas should be prioritized for economic research? Numerous opportunities present themselves, ranging from farm-level to societal benefits.

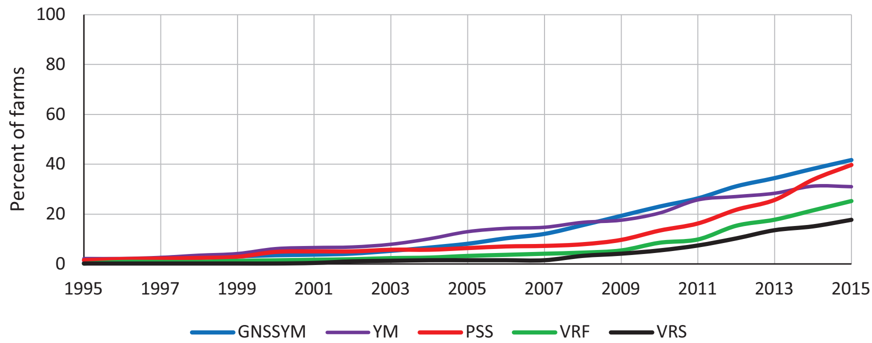
Precision Agriculture and Farm Management

To understand the connection between "Small Data" and "Big Data" in agriculture, it is useful to discuss the prevalence of farm-level sensors and other precision agriculture technology (figure 1). Perhaps ironically, the evolution and revolution in agricultural Big Data comes from the expansion of "Small Data" in agriculture; that is, the remarkable growth in producers' ability to collect data pertaining only to their own operation through the growth of techniques and technologies such as grid soil sampling, telematics systems for farm equipment, Global Navigation Satellite Systems (GNSS), farm aerial imagery acquired via small unmanned aerial systems (sUAS), and the like. In simplest terms, farms use "Small Data" when data are isolated to the fields where the data originated. Farmers who use information technology to conduct their own on-farm experiments, document yield penalties from poor drainage, or negotiate crop share agreements are using data that is considered "small." Producer adoption of these information technologies has increased dramatically in recent years (Griffin et al. 2017), giving rise to a profusion of agricultural data heretofore unseen (Erickson and Widmar 2015).

The new abundance of field-level information provided by these technologies could improve the ability of producers to make profit-maximizing decisions benefitting the producer operating the field, that is, "Small Data" (Griffin et al. 2017). However, pooling the datasets of hundreds or thousands of fields could hold a much greater potential value both to individual producers and the agricultural industry as a whole. Agricultural Big Data— farm data that has been combined into an aggregate form— has the potential to reveal undiscovered insights. Currently, only limited quantitative evidence exists regarding the value of assembling data from precision agriculture technology into a community; however, indirect evidence suggests that farm data has economic value.

One conceptual example of farm-level decision making is analysis based upon product-by-environment-by-management scenarios, or the so-called GxExM relationship. Historically, agricultural research focused on the interaction between inputs (such as a specific grain variety) and the environment (such as the presence of a given profile of soil nutrients and expected precipitation). One could refer to this as Genetics by Environment ("G x E") analysis. This approach generally excluded farmers' management practice variables in the analyses, in part because focus was still on the one-field-at-a-time paradigm such that the specific farmer's management practices were held constant for that field. Big Data's inclusion of outcomes from differing management strategies, from numerous fields employing a variety of inputs and environmental conditions could enable evaluation of the producers' management decisions as a variable as well, creating genetics by environment by management ("G x E x M") analyses where "genetics" loosely represents any product or system. Traditional agricultural research has

Figure 1 Proportion of Kansas farms using precision agriculture technology (N = 455)



focused on the phenotypic product-by-environment interaction rather than including the farmer and their management practices as a variable in the analyses. The utilization of farm data originating from precision agricultural technology is guiding decisions not only at the farm level but also for the manufacturers of inputs and equipment. This possibility opens innumerable avenues for research on the impacts of management practices on production outcomes, and could profoundly impact the sub-discipline of farm management. Farmers are but one of the many players attempting to benefit from Big Data. The marginal benefit differs not only for each population of players, but also differs along its lifecycle. The economics of networks, that is, network externalities, describe how individuals benefit from participation in a community or network (Varian 1999). The data from farms aggregated into the community are more valuable than data from any one farm would be individually. Given the network effects, the value of the data community is a function of the number of members of the system, and the data service provider enjoys much greater benefits than any other groups in the long run. However, in the short run, data service providers are likely to entice farms to join their network at least up to the point where a critical number of farms have joined (Varian 1999).

When farm data are aggregated into a community, the secondary uses of the data have a greater value than the summation of the initial uses of that data (Mayer-Schönberger and Cukier 2014). The distinction between Small Data and Big Data can be made clear by examining how the data fits into the initial or primary use of data versus the re-use or secondary use of that same data. For example, the initial uses of yield monitors may include documenting yields near drainage structures, while farm-level data on soil nutrient testing and subsequent as-applied variable rate fertility information are used by the farmer to fine-tune sub-field production. In the aggregate these same data— site-specific yield and soil test plus as-applied fertility— combined with similar data from thousands of other farmers provide insights into nutrient run-off. It is the re-use of farm data that gives rise to Big Data and the ability to assess environmental issues.

At the current position along the lifecycle of Big Data, data service providers strive to entice a critical mass of farmers to submit farm data so that the repository is replete (Coble et al. 2016). This is in part due to the fact that the value of a farm data community eventually depends on the number of farms and acres in the system, that is, the size of the network. Early in the farm data community lifecycle, data service providers may entice farmers, especially given the nonhomogeneous characteristics of farm data and farmers. This lack of homogeneity may result from farms with varying levels of data quality, for example, some farmers are known to calibrate yield monitors properly while other farmers may not correctly tag corn hybrids to fields. Further, some farms may be able to provide quality data from substantially larger acreages while other farms may have limited acreage that precision agriculture sensors were utilized. In addition to quantity and quality concerns, some farms may be perceived as local leaders. When these local leaders join the data community, other farmers are likely to follow. However, it should be noted that only a few exceptions fit the above criteria; and the overwhelming majority of farms are likely to voluntarily join the system with most even paying a fee. Essentially, data service providers are vying to become what is expected to be a natural monopoly. In the long run, the group that controls the data system enjoys the majority of the value (Mayer-Schönberger and Cukier 2014). Therefore, the next wave of farm management education is likely to focus on farm data issues and whether farms should relinquish control of farm data to third parties.

Policy and Legal implications

As mentioned above, Big Data has the potential to expand and deepen the tools for evaluation of farm-level decisions; by the same token, it could also expand the ability to evaluate the effect of policy interventions on the agricultural macroeconomy. In the long term, the growth of Big Data may give rise to new models for the evaluation of policy shocks to economic systems, but in the near term, the availability of larger and potentially more robust datasets may increase the accuracy of existing model outputs.

While Big Data eventually may impact the ability to evaluate policy decisions, its growth necessitates policy decisions today. Big Data carries the ability for potentially market-distorting actions as discussed below, but before one can have Big Data, individual producers must be willing to share their data. Concerns about data ownership and protections against both deliberate and inadvertent data disclosure abound among producers. Currently, there is no federal legislation protecting farm data like there is for health data (such as HIPAA) or personal financial data (FCRA 1970). Significant discussion on these points have led to several public dialogues calling for both public and private policies regarding farm data protections, such as the "Privacy and Security Principles for Farm Data" coordinated by the American Farm Bureau Federation (American Farm Bureau Federation 2017). Federal policymakers have taken note of these issues as well (House Agriculture Committee 2015). While these policy discussions continue, there has been no action at the federal level regarding farm data protections.

Until the data privacy issues are resolved, Big Data systems are reliant upon farmers both trusting data aggregators and sharing farm-level data for use in the aggregate. Farmers have typically readily shared their farm production and financial data, including geo-referenced farm data (Griffin, Reichlin, and Small 2008) with trusted partners such as universities; however, existing transfer systems are time-consuming and inefficient. Both parties would benefit from an improved system of transferring data, preferably wirelessly and in real-time. Farmers are being incentivized to share farm data via low-cost or "freemium" models, and some services are providing rudimentary comparative analysis, that is, agronomic and financial benchmarking, in exchange for providing data. Given that the agricultural industry is currently in the infancy of Big Data services, it is expected that farmers must be enticed to join data systems and share farm data. In the longer run, it is expected that farmers will freely join and even pay to participate in Big Data services; however, it is unclear when a critical mass of farms and acreage will enroll.

Economic theory applied to networks suggests that when a critical mass of users, that is, farms or acreage, join the system, the membership will exponentially increase. However, until critical mass is achieved, the growth of data services is expected to be slow. At least one example of farmers being paid for data exists; in 2016, Farmobile guaranteed their customers in southern Minnesota that they would receive at least $2 per acre (Farmobile 2016). In 2017, the company expanded similar offering to other customers, but at $1 per acre. During the infancy of Big Data, incentives such as this may be relatively more common than at any other time along the life cycle. Once a critical mass of acreage exists with one data company, farmers are expected to freely join that company and submit data from their farm. Essentially, farmers are poised to share farm data with third parties, especially if clear benefit-cost analyses indicate some perceived tangible or even intangible advantage.

Asymmetric Information Implications

The Holy Grail for market participants is to obtain perfect information as soon as it is knowable, and preferably before it is knowable to others. While Big Data has a long, long way to go before achieving this, bigger steps toward that goal are being taken faster than ever before. Thus, a significant concern with aggregating agricultural data is whether— either legitimately or not— a small number of market participants (or a single actor) could gain access to information sufficient to move (or even manipulate) markets faster than, or to the exclusion of, other market participants. While there are numerous rules in place to deal with a broad range of market-manipulating activities, none of these current rules contemplate the type of actions that could take place with a sufficiently large aggregated dataset. Currently, there are various rules restricting insider trading (see 17 C.F.R. §1.59(a); 17 C.F.R. § 1.3(ee)), and government employees are prohibited from using data for financial gain that has not been disseminated to the public (7 U.S.C. §6c(a)(3)). However, there are no rules governing "very good market information" such as that which could be obtained through completely legal means by aggregating sufficient telematics data (as an example). As a result, research on the potential market effects of growing market asymmetries that could be triggered by growing Big Data aggregations and the implications of policies restricting the use of aggregated data in commodity market transactions could do much to inform the development of law in the arena.

A farmer's decision to join a farm data network is likely a function of how they perceive their data and its value. Farmers who view farm data as an intangible resource may fear that relinquishing data may reduce their local negotiation power with landowners, retailers, and other service providers (Griffin et al. 2016). Further, some farmers may opt not to participate in data communities for fear that others may disproportionately benefit from their participation or the data that they bring into the system.

Sustainability and Traceability

While Big Data holds the promise of numerous economic benefits, it also creates opportunities for environmental benefits. Agricultural pollutant runoff has been a source of growing concern for water quality in the Chesapeake Bay, the Gulf of Mexico, and the Great Lakes. Traditionally agricultural runoff has been difficult to regulate by virtue of the "non-point" nature of such runoff, and the fact it is not directly regulated under the Clean Water Act's National Pollutant Discharge Elimination System (NPDES; 40 C.F.R. § 122.3). Historically, nutrient runoff was addressed through voluntary programs through the Clean Water Act's non-point source management program, which provides funding to state programs aimed at reducing nutrient releases in agricultural storm water runoff (33 U.S.C. § 1329). However, in some circumstances (such as with pollution concerns in the Chesapeake Bay), a "total maximum daily load" or TMDL may be imposed with the effect of requiring states to develop enforceable nutrient management plans (33 U.S.C. § 1313(d)). Both "small ag data" and Big Data have prospective roles to play in helping address nutrient runoff concerns. The increased adoption of precision agricultural tools at the farm level holds the potential to actually decrease nutrient application by matching nutrient inputs more closely to plant needs; at the regional level, this could reduce overall nutrient loading to sensitive waterways. The "as-applied" maps generated from precision agriculture tools could also facilitate farmers' ability to demonstrate compliance with nutrient management plans by showing the specific amount and location of nutrient applications (though this raises separate issues of sensor calibration and accuracy; Sisung 2016). Big Data tools could also significantly advance the tools used to manage nutrient concerns at the regional level through improved evaluation of policy tools instruments and modeling of nutrient management strategies such as nutrient "cap and trade" systems.

Concerns about food safety and consumer desires for more information about the sourcing of their food could also be addressed through small and Big Data as well. Telematics systems from the tractor to the retail center create the possibility of complete "farm to fork" tracking of foodstuffs which would enable disease traceability, while metadata collected along the distribution chain could be used to provide support for source verification and compliance with any number of production practice requirements.

From this discussion, it can be seen that there are numerous potential economic research questions to be answered through the application of precision agriculture and Big Data tools, and as the power of those tools grow, so will the calls for agricultural economists to respond to these and other questions. But how will those tools actually help find answers? To unlock the potential of evidence-based decision-making, entities or organizations need to convert the high volume, high frequency, and diverse data into meaningful insights. In this process, Labrinidis and Jagadish (2012) note that the extraction of insights can be broken down into two stages, namely, *data management* and *analytics*. Data management, on the one hand, includes process and supporting technologies to acquire and store data. Data is then prepared, transformed, and retrieved for analysis. Diebold (2012) notes that Big Data can lead to much stronger conclusions for data-mining applications. On the other hand, analytics refers to techniques that can be used to analyze and acquire information or intelligence from Big Data. Several Big Data techniques can be used to analyze both structured and unstructured data. These include (*a*) text analytics; (*b*) audio analytics; (*c*) social media analytics; (*d*) video analytics; and (*e*) predictive analytics. In applied economics, the main focus is on predictive analytics.

Research Methods Using Big Data

Predictive analytics includes a variety of techniques or procedures that can predict the future outcome based on either historical and/or current data. For example, we can predict consumers' buying habits based on what they buy, when they buy, and even what they are writing about the product that they bought on social media. One of the hallmarks of predictive analytics is seeking to uncover patterns and relationships in data; tools for accomplishing this can be subdivided into two groups. First, techniques such as moving averages attempt to discover historical patterns in the outcome variables and then predict the future. Second is the regression analysis, which is well known in our profession.

Recall that all precision agriculture is based on statistical methods, and the statistical methods behind these methods may not apply to the problems being addressed by Big Data. There are several reasons for this. For example, conventional statistical methods are based on statistical significance, where results from a small sample, (obtained from the population) are compared to examine the significance of particular relationships, and the conclusions are generalized with respect to the entire population. However, in the case of Big Data, which are massive in size, the "sample" may actually represent the majority of, or the entire population; that is. the sample size equals "all" (Mayer-Schönberger and Cukier 2014). Therefore, any statistical significance test is not relevant to Big Data, especially those tests aimed at samples from a population. Finally, Fan, Han, and Lui (2014) point out that Big Data has heterogeneity, noise accumulation, spurious correlations, and incidental endogeneity. In other words, the underlying concept of Big Data relies relatively more on correlation and less on causation than the theory-based science upon which agricultural economics analyses have largely been based.

Big Data is heterogeneous because it represents information from different sub-populations and from different sources. The sheer size of Big Data helps us in modeling heterogeneity and requires sophisticated statistical techniques. Since estimation of predictive models using Big Data often involves the simultaneous estimation of several parameters, it may give rise to accumulated error terms. As a result, the true effect of variables may be masked. In their study, Fan and Lv (2008), through simulation modeling, show that the correlation between independent variables tends to increase with the size of the dataset. Therefore, in Big Data analysis, because of high dimensionality, we may see some variables that should not be in the model (unrelated) may be correlated. Finally, recall that in regression modeling, we assume exogeneity—the error term is independent of the predictors or the explanatory variables. The assumption of exogeneity is usually met in small samples, but incidental endogeneity is commonly present in Big Data.

Machine Learning

Machine learning, a branch of computer science and one of the major areas of artificial intelligence, can be used to construct algorithms to exploit the potential value of Big Data.[1] Note that for machines to become intelligent like humans, they must learn like humans; human minds learn from past data and experiences and then applies this learning to future decisions. Machine learning is a two-step process. First, the machine has to learn the input data; secondly, the machine has to interpret it and analyze the input and output data to create machine algorithms. The algorithms can then construct a system model, which is used to predict future values. Machine learning methods are more flexible than conventional statistical methods because they do not rely on user-specified models. Instead, they self-improvise using the available volume of data.

There are three types of machine learning algorithms: Supervised learning (SL): If the output variables are provided, then the learning becomes supervised. In SL, the algorithm is given some training examples and the machine studies input and corresponding outputs.[2] Therefore, popular SL algorithms include artificial neural networks (Kaul et al. 2005; Uno 2005; Chen and Mcnairn 2006; Khoshnevisan et al. 2014), decision trees (Veenadhari, Mishra, and Singh 2011), K-means clustering (Shawe-Taylor and Cristianini 2004), support vector machines (Radhika and Shashi 2009), and Bayesian networks (Bakker and Heskes 2003).[3,4] The artificial neural network (ANN) algorithm has been widely used in the agricultural field. ANN is an interconnected set of inputs and output units where weight is associated with each connection (see Drummond, Sudduth, and Birrell 2008).[5] The ANN has an advantage over multiple regression because ANN can select an independent variable in the data, learn complex relationships, and does not place strict requirements *a priori* on a functional a functional form. The neural network can discover more complex variables.

The second type of algorithm is unsupervised learning (UL): In UL, the algorithm is not provided with outputs and learning helps us find interesting information about our dataset solely looking at its features alone. Popular UL algorithms are self-organizing maps (SOM), partial based clustering, hierarchical clustering, K-means clustering, COBWEB, and density-based spatial clustering.[6,7] To date, these techniques have rarely been used in agriculture and economics field.

The third type of algorithm is reinforcement learning (RL): With RL, the learning process works on the principle of feedback. The notion is that every

[1]*Applications of machine learning are multi-disciplinary.*

[2]*See Mucherino et al. 2009.*

[3]*See Cheng and Titterington (1994) and Warner and Misra (1996). On one hand, Cheng and Titterington (1994) have reviewed the artificial neural network (ANN) methodology. On the other hand, Warner ad Misra (1996) emphasize understanding ANN as a statistical tool. The accuracy of ANN increases with the volume of data. The advantages of the ANN is that:* (a) *ANN are capable of adopting their complexity without knowing the underlying principles;* (b) *ANN can derive relationships between input and output on any process.*

[4]*Bayesian networks focus on two issues: estimating the conditional probability tables from training data when the structure of the network is known;and learning a network's structure from training data.*

[5]*The ANN can be used in flood forecasting, modeling rainfall, and run-off relationships.*

[6]*See Moshou et al. 2006.*

[7]*The COBWEB is an incremental and unsupervised clustering algorithm that produces a hierarchy of classes:its incremental nature allows clustering of new data without having to repeat the existing clustering. See Fisher's Cobweb (1987).*

action has an impact on the system; the impact or information is then reported back to the algorithm. Consequently, the algorithm modifies its behavior. Popular algorithms include genetic algorithms, and Markov decision algorithms (e.g., Matis, Birkett, and Bourreaux 1989; Jain and Ramasubramalliall 1998; Osman, Inglada, and Dejoux 2015).

An example of machine learning in agricultural economics is the prediction of farmland values. Academic research and at least one commercial offering has focused on predicting the value that a parcel of land will be sold for using current and historical land sales, soil characteristics, climatic and weather data, cropping systems, remotely sensed imagery, potential of urban sprawl (Livanis et al. 2006; Castle, Wu, and Weber 2011), and the general economic situation including commodity prices and interest rates (Irwin and Sanders 2011). The Big Data implication of predicted farmland values is to determine if expected sales prices are over- or under-valued. A commercial example is Granular AcreValue from DuPont.

The commonly used models are linear regression models (Shibayama 1991); polynomial regression models (Wilcox et al. 2001), and nonlinear regression models (House 1979). Variable selection for the models can be based on several methods including stepwise regression, principal component regression, Bayesian information criterion, Akaike information criterion, and partial least squares (for details, see Castle, Qin, and Reed 2009). Varian (2014) show that the classical multivariate regression model can be used to predict the outcome variable using predictor variables and adding a penalty term to the classical minimization of the sum of squared residual—a technique called *elastic net regression* (ENR). The complexity in numbers and size of the predictors coming from Big Data tend to shrink the least squares coefficients to zero, which can make ENR an attractive technique for working with such datasets. The researchers can choose the coefficients in ENR. In the case of both ENR and least absolute shrinkage and selection operator (LASSO), some of the variables are set to be exactly zero—leading to computation efficiency, feasibility, and providing good predictions (Varian 2014).

Spike and Slab Regression Analysis

Another regression technique useful for Big Data is spike and slab regression. This is a Bayesian technique, originally coined by Mitchell and Beauchamp (1988), which refers to a type of prior probability distribution ("prior") used for the regression coefficients in linear regression models.[8] Note that the use of a normal prior was instrumental in facilitating efficient Gibbs sampling of the posterior; this, in turn, made the spike and slab variable selection method computationally attractive. In 2010, Ishwaran and Rao (2010) developed a generalized ridge regression (GRR), which possesses unique advantages in high-dimensional correlated settings to estimate the model; the weighted GRR is more effective than other tools in many circumstances.

A technique related, but not identical to, the spike-and-slab method is Bayesian moving averaging (BMA). Bayesian methods are becoming

[8]*It is assumed that the regression coefficients were mutually independent with a two-point mixture distribution made up of a uniform flat distribution (the slab) and a degenerate distribution at zero (the spike).*

increasingly popular as frameworks for model selection and forecasting tools. In some cases, analysts ignore the uncertainty in model selection, resulting in overconfident inferences and decisions that are riskier. The BMA techniques are designed to account for this uncertainty. By averaging over several different competing models, BMA incorporates the model uncertainty into the parameters and predictions (see Jacobs et al. 1991). Zhou et al. (2012) have proposed model selection and comparison in the case of BMA. These authors constructed posterior probabilities properties and model parameters based on sequential Monte Carlo sampling, and used these properties to compare different models. A final model is obtained as a weighted average of all models, where the weight of each model is its posterior probability. Varian (2014) concludes that Bayesian techniques are computationally efficient and preferred to exhaustive searches. Finally, using Big Data, Ley and Steel (2009) have compared LASSO, Bayesian model averaging, and spike-and-slab methods to show which variables are important predictors of economic growth.

Time Series Analysis Using Big Data

Time series forecasting is a model used to predict future values based on previously observed values. The time series analysis is important for crop forecasting, stock prices, price movement, and futures and options. There are various types of time series analysis methods, including parametric or non-parametric, frequency domain and time domain, and linear, univariate, and multivariate. Note that frequency domain analysis includes spectral analysis and wavelet analysis; time domain includes auto-correlation and cross-correlation. Parametric approaches include autoregressive or moving average models; non-parametric approaches include covariance or spectrum and usually focus on a smooth spectral density. The Bayesian Structural Time Series (BSTS) model works well for handling the variable selection problem in the case of time series analysis. Banbura, Giannone, and Reichlin (2011) introduced "*nowcasting*" as a term in econometric time series analysis, which refers to forecasting a current value instead of the future value.[9] The *nowcasting* model has two components, namely a general trend, and seasonal pattern in the data. In the case of Big Data, where the number of potential predictors in the regression model is large (often larger than the number of observations available to fit the model), a Markov chain Monte Carlo (MCMC) sampling algorithm can be used to simulate from the posterior distribution. Finally, one can use Bayesian model averaging to smooth the predictions over a large number of potential models.

Applications in Agriculture and Applied Economics

Several applications of the above-mentioned techniques can be used to enhance the productivity of farms along with reducing their use of inputs.

Weather forecasting. Environmental factors like weather influence crop growth and development as well as recreational demand for both agricultural and non-agricultural lands. Production agriculture has spatial yield

[9]*Banbure et al. (2011) conclude that a good or effective nowcasting model should consider both past behavior of the series and easily observed contemporaneous signals. Now casting is a contraction term for* now *and* forecasting *(Giannone, Reichlin, and Small 2008).*

variability, partly because of spatial variability in soil properties and interactions with the weather, which is also spatially varied. Machine learning techniques like Support Vector Machines (see Vapnik 1998) can be used to predict the weather for farmers to aid in their decision-making (see Agrawal and Mehta 2007 and Radhika and Shashi 2009).

Crop yield prediction and crop selection. Machine learning provides many effective algorithms, which can identify input and output relationships in crop selection and yield prediction. Popular techniques such as artificial neural networks, K-nearest neighbors, and decision trees have proven to be effective in crop selection, which is based on various factors like climate, soils, natural calamities, famine, and other inputs. Using several soil characteristics (e.g., topsoil depth, phosphorous, potassium, salt, organic matter, and magnesium saturation) as input and artificial neural networks, Drummond, Suddeth, and Birrell (2008) accurately predicted corn and soybean yields.

Irrigation systems. Agriculture consumes a major portion of world's fresh water. Variability in rainfall, climate change, and dropping of the water table in developing countries is alarming. Using smart irrigation systems and data collected by sensors can be used to make better decisions regarding water usage. Several studies using artificial neural network algorithms have been able to predict accurate water levels and rainfall runoffs (Ashaary, Ishak, and Ku-Mahamud 2015; Chakravarti, Joshi, and Panjiar 2015).

Crop disease prediction. Early crop disease detection can be accomplished through machine learning.[10] In their study, Drummond, Suddeth, and Birrell 2008 note that ANN could be helpful in predicting pest attacks in advance. Such models deal well with noisy and multi-faceted data and account for wide ranges of possible factors (e.g., historical data, satellite/sensor data, field conditions, images of leaves) to effectively learn and predict crop diseases. In 2010 Rumpf et al., used Support Vector Machines to develop early crop disease detection algorithms.

Agricultural policy and trade. A large quantity of data on production output of crops, changes in input costs, market demand and supply, market price trends, cultivation costs, wages, transportation costs, and marketing costs could be used by ANN algorithms to predict support prices for farmers by governments in both developed and developing countries. For instance, Big Data can be beneficial when simulating agricultural policy impacts. The application to the Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP) model in the European Union illustrates the capability for assessing policy impacts at the farm level (Louhichi et al. 2015).

There is one more critical application—or, rather, implication—of Big Data for the agricultural and applied economics profession. All the challenges discussed here beg to address the question of whether agricultural economics departments should provide more graduate student instruction for Big Data issues. Supplementing the traditional analytical tools our students learn in classrooms and during their thesis research with computer science and non-traditional statistics may increase the rate at which agricultural economists can make meaningful contributions not only in applied economics but other disciplines. This may mean that the departments of agricultural economics dedicated to providing research and education on Big Data employ non-economist faculty in their ranks to provide specific

[10] *Factors like soil quality, crop rotation cycle, and seed quality can help detect crop diseases.*

expertise. At the 2017 AAEA Symposium on Big Data, graduate students were challenged to consider how they could replace themselves with an algorithm in lieu of physically interacting with data. The challenge was extended to include the students considering how to identify outliers without ever having the opportunity to "see" the data, but rather to build models to anticipate erroneous data, flag it for omission, and continue the analysis. Three reasons exist for the need to automate analytic processes: data may be considered confidential such that no set of eyes can be on the data; replacing human capital with an algorithm substantially lowers per unit costs of analysis; and there are likely not enough analysts available to meet the demand for analysis in the future.

Conclusions

Given our assessment of the needs and opportunities arising from the Big Data expansion, we come to a few significant conclusions for our profession and those who draw upon our work. First, there is a unique and important role for agricultural and applied economists in this changing technological environment. We see an opportunity for our profession to stand at the hub of work within multi-disciplinary teams. Our profession is trained to handle and draw valid inferences from non-experimental data. Further, most applied economists are trained and comfortable with unstructured, messy data. Many in our discipline have already engaged in some form of big data analysis and we understand the important distinctions between causation and simple predictive models. Having noted some comparative advantages of our profession, we also challenge agricultural and applied economists to prepare a next generation of our profession with training in geo-spatial analysis and analytical techniques described in this paper. Furthermore, we need to be the champions for the merit of research with these type of data, and advocate for non-experimental data access and research funding.

We perceive an important role for academic researchers and land grant personnel in this venue. First, there is a need for basic and applied multidisciplinary research that provides objective third-party analysis. Ground truthing seed varieties may morph into ground truthing software and other roles. There is also a clear role for extension to help train and educate producers and agri-business professionals how to manage new tools and data. Clearly, educational topics like data ownership and evaluation of precision agriculture investment will be in demand.

Finally, we have touched upon several looming policy issues, which is not surprising as many policy debates are stimulated by technological change. First, there is room for discussion regarding data ownership of these data. The returns and development of these technologies depend on the ownership rules in place. Second, we find that infrastructure needs such as rural broadband are potentially limiting the use of these technologies, as rural broadband access provides a critical bridge between small data and Big Data. Thus, to the extent that access to these technology provides a comparative advantage to certain areas, largely rural areas are disadvantaged. Third, we perceive opportunities and threats to public objective data collection and government program data. Ultimately, we advocate for a reimagining of agricultural data collection such that the greatest synergism can be obtained from integrating private data, government program data, and specific data collection surveys meant to complement other available tools.

References

AgFunder. 2017. AgTech Investing Report Year in Review 2016. Available at: https://research.agfunder.com/2016/AgFunder-Agtech-Investing-Report-2016.pdf.

Agrawal, R., and S.C. Mehta. 2007. Weather Based Forecasting of Crop Yields, Pests, and Diseases - IASRI Models. *Journal of Indian Society Agricultural Statistics* 62 (2): 1–12.

American Farm Bureau Federation. *Privacy and Security Principles for Farm Data.* Available at: http://www.fb.org/issues/technology/data-privacy/privacy-and-security-principles-for-farm-data/.

Annan, F., J.B. Tack, A. Harri, and K.H. Coble. 2014. Spatial Pattern of Yield Distributions: Implications for Crop Insurance. *American Journal of Agricultural Economics* 96 (1): 253–68.

Ashaary, N., W. Ishak, and K. Ku-Mahamud. 2015. Neural Network Application in the Change of Reservoir Water Level Stage Forecasting. *Indian Journal of Science and Technology* 8 (13): 1–6.

Bakker, B., and T. Heskes. 2003. Task Clustering and Gating for Bayesian Multitasking Learning. *Journal of Machine Learning Research* 4 (5): 83–99.

Banbura, M., D. Giannone, and L. Reichlin. 2011. Nowcasting. In *Oxford Handbook of Economic Forecasting*, eds. M.P. Clements and D.F. Hendry, 193–224. Oxford University Press.

Capps, O. 1989. Utilizing Scanner Data to Estimate Retail Demand Functions for Meat Products. *American Journal of Agricultural Economics* 71 (3): 750–60.

Castle, J., X. Qin, and R. Reed. 2009. How to Pick the Best Regression Equation: A Review and Comparison of Model Selection Algorithms. Working Paper No. 13/2009 Department of Economics and Finance College of Business and Economics University of Canterbury.

Castle, E., J.J. Wu, and B. Weber. 2011. Place Orientation and Rural-Urban Interdependence. *Applied Economic Perspectives and Policy* 33 (2): 179–204.

Chakravarti, A., N. Joshi, and H. Panjiar. 2015. Rainfall Runoff Analysis Using the Artificial Neural Network. *Indian Journal of Science and Technology* 8 (14): 1–7.

Chen, C., and H. Mcnairn. 2006. A Neural Network Integrated Approach for Rice Crop Monitoring. *International Journal of Remote Sensing* 27 (7): 1367–93.

Cheng, B., and D.M. Titterington. 1994. Neural Networks: A Review from Statistical Perspective: Rejoinder. *Statistical Science* 9 (1): 49–54.

Co, H., and R. Boosarawongse. 2007. Forecasting Thailand's Rice Export: Statistical Techniques vs. Artificial Neural Networks. *Computers and Industrial Engineering* 53 (4): 610–27.

Coble, K., T.W. Griffin, M. Ahearn, S. Ferrell., J. McFadden, S. Sonka, and J. Fulton. 2016. Advancing U.S. Agricultural Competitiveness with Big Data and Agricultural Economic Market Information, Analysis, and Research (No. 249847). Washington DC: Council on Food, Agricultural, and Resource Economics.

Diebold, F.X. 2012. A Personal Perspective on the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline, Second Version. University of Pennsylvania, Penn Institute for Economic Research, Working Paper No. 13-003.

Dyer, J. 2016. The Data Farm: An Investigation of the Implications of Collecting Data on the Farm. Taunton, Somerset: Nuffield Australia Project No 1506.

Erickson, B., and D.A. Widmar. 2015. Precision Agricultural Services Dealership Survey Results. West Lafayette, IN: Purdue University. Available at: http://agribusiness.purdue.edu/files/resources/2015–crop–life–purdue–precision–dealer–survey.pdf.

Fair Credit Reporting Act (FCRA). 1970. 15 U.S. Code §§ 1681. Washington DC: U.S. Congress.

Fan, J., F. Han, and H. Liu. 2014. Challenges of Big Data Analysis. *National Science Review* 1 (2): 293–314.

Fan, J., and J. Lv. 2008. Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (5): 849–911.

Farmobile. 2016. Farmobile Announces Data Store, Guarantees Minnesota Farmers at Least $2 per Acre for Electronic Field Records. Available at: https://www.farmobile.com/static/www/docs/Farmobile–Data–Store–Press–Release.pdf.

Featherstone, A.M. 2018. The Farm Economy: Future Research and Education Priorities. *Applied Economic Perspectives and Policy* 40 (1): 136–54.

Fisher, D.H. 1987. Knowledge Acquisition Via Incremental Conceptual Clustering. *Machine Learning* 2: 139–72.

Giannone, D., L. Reichlin, and D. Small. 2008. Nowcasting: The Real-time Informational Content of Macroeconomic Data. *Journal of Monetary Economics* 55 (4): 665–76.

Griffin, T.W., C.L. Dobbins, T.J. Vyn, R.J.G.M. Florax, and J.M. Lowenberg-DeBoer. 2008. Spatial Analysis of Yield Monitor Data: Case Studies of On-farm Trials and Farm Management Decision Making. *Precision Agriculture* 9 (5): 269–83.

Griffin, T.W., T.B. Mark, S. Ferrell, T. Janzen, G. Ibendahl, J.D. Bennett, J.L. Maurer, and A. Shanoyan. 2016. Big Data Considerations for Rural Property Professionals. *Journal of American Society of Farm Managers and Rural Appraisers* 79: 167–80.

Griffin, T.W., N.J. Miller, J. Bergtold, A. Shanoyan, A. Sharda, and I.A. Ciampitti. 2017. Farm's Sequence of Adoption of Information-Intensive Precision Agricultural Technology. *Applied Engineering in Agriculture* 33 (4): 521–7.

Health Insurance Portability and Accountability Act. 42 U.S. Code §§ 201 *et seq.*, and 45 C.F.R Parts 160 and Part 164. Washington DC: U.S. Congress.

Hennessy, T., D. Läpple, and B. Moran. 2016. The Digital Divide in Farming: A Problem of Access or Engagement? *Applied Economic Perspectives and Policy* 38 (3): 474–91.

House, C.C. 1979. Forecasting Corn Yields: A Comparison Study Using 1977 Missouri Data. U.S. Department of Agriculture, Economics, Statistics and Cooperatives Service, Statistical Research Division. June 1979, 66.

Irwin, S., and D. Sanders. 2011. Index Funds, Financialization, and Commodity Futures Markets. *Applied Economic Perspectives and Policy* 33 (1): 1–31.

Ishwaran, H., U.B. Kogalur, and J.S. Rao. 2010. Spikeslab: Prediction and Variable Selection Using Spike and Slab Regression. *The R Journal* 2 (2): 68–73.

Ishwaran, H., J.S. Rao, and U.B. Kogalur 2013. Spikeslab: Prediction and Variable Selection Using Spike and Slab Regression. *R Package Version 1.1.5.*

Jacobs, R.A., M.I. Jordan, S.J. Nowlan, and G.E.Hinton. 1991. Adaptive Mixtures of Local Experts. *Neural Computation* 3 (1): 79–87.

Jain, R., and V. Ramasubramalliall. 1998. Forecasting of Crop Yields Using Second Order Markov Chains. *Journal of the Indian Society of Agricultural Statistics* 51: 61–72.

Kaul, M., L. Robert, H. Hill, and C. Walthall. 2005. Artificial Neural Networks for Corn and Soybean Yield Prediction. *Agricultural System* 85 (1): 1–18.

Ker, A., and K. Coble. 2003. Modeling Conditional Yield Densities. *American Journal of Agricultural Economics* 85 (2): 291–304.

Khoshnevisan, B., S. Rafiee, M. Omid, H. Mousazadeh, and M.A. Rajaeifar. 2014. Application of Artificial Neural Networks for Prediction of Output Energy and GHG Emissions in Potato Production in Iran. *Agricultural Systems* 123: 120–27.

Kuchler, F., A. Tegene, and J.M. Harris. 2005. Taxing Snack Foods: Manipulating Diet Quality or Financing Information Programs? *Applied Economic Perspectives and Policy* 27 (1): 4–20.

Labrinidis, A., and H.V. Jagadish. 2012. Challenges and Opportunities with Big Data. *Proceedings of the VLDB Endowment* 5 (12): 2032–3.

Ley, E., and M.K. Steel. 2009. On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression. *Journal of Applied Econometrics* 24 (4): 651–74.

Liu, Y., S.M. Swinton, and N.R. Miller. 2006. Is Site-specific Yield Response Consistent Over Time? Does It Pay? *American Journal of Agricultural Economics* 88 (2): 471–83.

Livanis, G., C.B. Moss, V. Brennan, and R. Nehring. 2006. Urban Sprawl and Farmland Prices. *American Journal of Agricultural Economics* 88 (4): 915–29.

Louhichi, K., P. Ciaian, M. Espinosa, L. Colen, A. Perni, and S. Gomez y Paloma. 2015. EU-wide Individual Farm Model for CAP Analysis (IFM-CAP): Application to Crop Diversification Policy. European Commission, Joint Research Center, Sevilla, Spain. Available at: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92574/jrcreport_jrc92574.pdf.

Matis, J.H., T. Birkett, and D. Boudreaux. 1989. An Application of the Markov Chain Approach to Forecasting Cotton Yields from Surveys. *Agricultural Systems* 29 (4): 357–70.

Mayer-Schönberger, V., and K. Cukier. 2014. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin Harcourt Publishing Company.

Mitchell, T., and J. Beauchamp. 1998. Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association* 83: 1023–36.

Moshou, D., C. Bravo, S. Wahlen, J. West, A. McCartney, J. Baerdemaeker, and H. Ramon. 2006. Simultaneous Identification of Plant Stresses and Diseases in Arable Crops Using Proximal Optical Sensing and Self-organising Maps. *Precision Agriculture* 7 (3): 149–64.

Mucherino, A., P. Papajorgji, and M. Paradalos. 2009. A Survey of Data Mining Techniques Applied to Agriculture. *Operational Research* 9 (2): 121–40.

National Academies of Sciences, Engineering, and Medicine. 2017. Improving Crop Estimates by Integrating Multiple Data Sources. Washington DC: The National Academies Press.

Osman J., J. Inglada, and J.F. Dejoux. 2015. Assessment of a Markov Logic Model of Crop Rotations for Early Crop Mapping. *Computers and Electronics in Agriculture* 113: 234–43.

Ozaki, V.A., S.K. Ghosh, and B.K. Goodwin. 2008. Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts. *American Journal of Agricultural Economics* 90 (4): 951–61.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Radhika, Y., and M. Shashi. 2009. Atmospheric Temperature Prediction Using Support Vector Machines. *International Journal of Computer Theory and Engineering* 1 (1): 55–9.

Rumpf, T., A. Mahlein, U. Steiner, E. Oerke, H. Dehne, and L. Lumer. 2010. Early Detection and Classification of Plant Diseases with Support Vector Machines Based on Hyperspectral Reflectance. *Computers and Electronics in Agriculture* 74 (1): 91–9.

Schimmelpfennig, D., and R. Ebel. 2016. Sequential Adoption and Cost Savings from Precision Agriculture. *Journal of Agricultural and Resource Economics* 41 (1): 97–115.

Sisung, T. 2016. Soil Testing and Nutrient Application Practices of Agricultural Retailers in the Great Lakes Region. Master of Agribusiness Thesis. Department of Agricultural Economics, Kansas State University.

Sonka, S. 2015. Big Data: From Hype to Agricultural Tool. *Farm Policy Journal* 12 (1): 1–9.

Stubbs, M. 2016. Big Data in U.S. Agriculture. Washington DC: Congressional Research Service, Report R44331.

Tack, J., K.H. Coble, R. Johansson, A. Harri, and B. Barnett. 2017. The Potential Implications of "Big Ag Data" for USDA Forecasts. Available at: https://ssrn.com/abstract=2909215.

Tremblay, N. 2017. Confronting the Challenges of Big Data for Precision Agriculture. Presentation at International Society of Precision Agriculture Annual Meetings, February 17. SaintJean-sur-Richeliueu, Quebec.

Uno, Y. 2005. Artificial Neural Networks to Predict Corn Yield from Compact Airborne Spectrographic Imager Data. *Computers and Electronics in Agriculture* 47 (2): 149–61.

Upbin, B. 2013. Monsanto Buys Climate Corp for $930 Million. *Forbes* (online edition), October 2, 2013. Available at: https://www.forbes.com/sites/bruceupbin/2013/10/02/monsanto-buys-climate-corp-for-930-million/#5cb3cf2f177a.

U.S. House of Representatives Committee on Agriculture, Big Data and Agriculture: Innovations and Implications. 2015. Washington DC: Committee Hearing Proceedings.

Vapnik, N.V. 1998. *Statistical Learning Theory*. New York: Wiley.

Varian, H. 2014. Big Data: New Trick for Econometrics. *Journal of Economic Perspectives* 28 (2): 3–28.

Varian, H.R. 1999. Market Structure in the Network Age. Understanding the Digital Economy Conference. Washington DC: U.S. Department of Commerce.

Veenadhari, S., B. Mishra, and C.D. Singh. 2011. Soybean Productivity Modelling Using Decision Tree Algorithms. *International Journal of Computer Applications*. 27 (7): 11–15.

Warner, B., and M. Misra. 1996. Understanding Neural Networks as Statistical Tools. *The American Statistician* 50 (4): 284–93.

Wilcox, A., N.H. Perry, N.D. Boatman, and K. Chaney. 2000. Factors Affecting the Yield of Winter Cereals in Crop Margins. *Journal of Agricultural Science* 135 (4): 335–46.

Woodard, J.D. 2016. Data Science and Management for Large Scale Empirical Applications in Agricultural and Applied Economics Research. *Applied Economic Perspectives and Policy* 38 (3): 373–88.

Woodard, J.D., and L.J. Verteramo-Chiu. 2017. Efficiency Impacts of Utilizing Soil Data in the Pricing of the Federal Crop Insurance Program. *American Journal of Agricultural Economics* 99 (3): 757–72.

Zhou, Y., M.A. Johansen, and J.A.D. Aston. 2012. Bayesian Model Comparison Via Path-sampling Sequential Monte Carlo. Proceedings of the IEEE Workshop on Statistical Signal Processing.